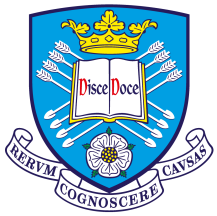
Mechanisms Determining the Dynamic Localisation of Ribosomes in Neurons



The University Of Sheffield.

A thesis presented for the degree of

Doctor of Philosophy

University of Sheffield

Ashleigh Frances Davey

July 2023

Abstract

Local translation of mRNA in neuronal processes is essential for neuronal development and function. Recent research has shown that local translation also occurs in mature axons and defects in this essential protein synthesis are associated with the development of neurodegenerative disease. However, the mechanism behind ribosome translocation from the cell body into the distal axon is unknown. This work details the creation of several knock-in CRISPR constructs which introduce a HaloTag into the terminal regions of several ribosomal proteins. Using these constructs in PC-12 and primary neuronal lines, the axonal transport of endogenous ribosomes was visualised using live cell imaging. Analysis of transport events determined that cell-body derived ribosomes are transported into the distal axon at speeds consistent with fast axonal transport. The bi-directional nature of axonal ribosome transport also indicates the involvement of kinesin and dynein motors, whilst the co-localisation of ribosome and Lysotracker Green labelled particle transport implies that the movement of axonal ribosomes may be coupled to lysosome/endosome trafficking. RNA-Seq analysis of ribosome associated gene expression in brain regions also provided evidence of neuronal ribosome specialisation, which may be important for ribosome selection, transport and local translation in the distal axon. This project has delivered the first microscopic evidence of axonal ribosome transport in a neuronal model system. Future work using these constructs may assist in understanding the underlying pathologies of several major neurodegenerative diseases.

Declaration

I declare that I, Ashleigh Frances Davey, confirm that this Thesis is my own work. I am aware of the University's guidance on the use of Unfair Means. The research presented in this Thesis was carried out from October 2019 to July 2023 at the University of Sheffield, under the supervision of Dr Alison Twelvetrees (Department of Neuroscience) and Dr Emma Thomson (School of Biosciences). This work has not previously been presented for an award at this, or any other, university.

Acknowledgements

It is truth universally acknowledged that good science cannot be done in a vacuum; a scientist is only as strong as the lab, friends, and community that supports them. Therefore, it is with tremendous gratitude that I would like to acknowledge those that made this work possible. From the typical stressors of PhD life, to a world-ending global pandemic, completing this PhD was challenging experience that I couldn't have done alone.

Firstly, I'd like to thank my supervisors, Dr Alison Twelvetrees and Dr Emma Thomson. This PhD couldn't have been completed without the excellent supervision I received and I consider myself lucky to have both your guidance during these last few years. Ali, you made me a better scientist, your support, advice, and encouragement made my time in the lab a pleasure. Although I may not be at the bench anymore, I'll always remember your motto - "Should that be on ice?". Yes, it's always yes. Emma, your enthusiasm for science is infectious, thank you for your support and kind words - I'll always be team ribosome!

None of this work could have been done without the infinite support of my lab group. Working in the Twelvetrees lab for the last few years (almost five!) has been a delight, and it has been an honour to watch the group grow and succeed. In particular, I'd like to thank Evie and Emma for their incredible support, your enthusiasm for my ribosome facts and western blot skills has always brought me joy. You guys make coming to work fun and my PhD wouldn't have been the same without you. Similarly, I'd like to thank the Thomson and Bose labs for your support and friendship, especially over those "unprecedented" Covid years. Special thanks

goes to Katie, Nic and Pete - writing club has been a source of comfort over these last few months, and whilst I won't miss the early mornings, the camaraderie kept me going.

Next, I'd like to thank my family and friends for their continued support. Mum and Nat, thank you for everything, your sage advice, love, and enthusiasm for my science was infinitely valuable (I'll cherish the letters!). Matt, Charlotte, Bobby, Marine, Sam, and Caitlin, I greatly appreciate your friendship over the last few years, and for putting up with my absence for the last few months. Here's to more D&D adventures! I'd also like to thank Eve and Anna, I really value our long chats over too much coffee, your encouragement and kind words mean more than you know. Finally, to Callum, Kat, Chef de Party and all the spooky friends I've made over the years, thanks for reminding me that there's more to life than work. Go have fun, the science will be there when you come back.

Of course, I couldn't have made it to this point without my two biggest supporters, Dave and Pumpkin. To the cutest cat I know, thank you for being the sweetest, loudest, tiniest monster it has been my pleasure to live with - thesis writing wouldn't have been the same without you sitting under my desk and intermittently biting my feet when I haven't given you enough attention. Dave, you've been my biggest cheerleader throughout everything. You've brought me hundreds of cups of tea, celebrated my wins, consoled me during my lows, and believed in me more than I ever could. My love and gratitude for you is infinite.

Lastly, I'd like to thank the many benefactors which supported me in my academic journey, without whom I'd have likely not made it to University education. Your incredible generosity funded scholarships which provided the means for me to pursue research, allowing me to develop the passion and scientific curiosity that has served me throughout my PhD. It is my hope that one day I will be able to pay this kindness forward in my own life.

To Dave,

For your unwavering support, despite still not knowing what a ribosome is.

I couldn't have done it without you.

Contents

1 Introduction **23**

1.1 The Necessity of Axonal Transport in Axons 23

1.2 Local Translation in Neurites Supports Neuronal Function 28

1.2.1 The Varied Roles of Axonal Local Translation 34

1.2.2 Defective Local Translation is Implicated in Neurodegenerative Disease 36

1.3 Ribosomes Are an Essential Component of the Translational Machinery 39

1.3.1 Ribosome Biogenesis is a Complex and Highly Conserved Pathway 45

1.3.2 Specialised Ribosomes 53

1.4 Investigating Ribosome Localisation in Neurons 59

1.5 Axonal Ribosome Localisation Hypotheses 60

1.5.1 The Glial Origin Hypothesis 60

1.5.2 The Hitchhiker Hypothesis 62

1.6 Thesis Aims 64

2 Materials and Methods **66**

2.1 Tissue Culture and Transfection 66

2.1.1 Cell Culture of PC-12 Line 66

2.1.2 Electroporation of PC-12 Lines and Primary Neurons 67

2.1.3 Differentiation of PC-12 Cell Line 67

2.1.4 Preparation of Primary Cortical Neurons 68

2.1.5 Culture of Primary Cortical Neurons 68
2.1.6 Plating Cells onto Acid-Washed Poly-L-Lysine Treated Coverslips 69
2.1.7 Transfection of Primary Cortical Neurons 70
2.1.7.1 Transfection Using Lipofectamine 2000 70
2.1.7.2 Transfection Using Lipofectamine 3000 72
2.2 Molecular Biology 72
2.2.1 Constructs 72
2.2.2 Bacterial Culture and Plasmid DNA Preparation 72
2.2.2.1 Preparation of Overnight Bacterial Cultures and Agar Plates 72
2.2.2.2 Mini-preparation of Plasmid DNA 73
2.2.2.3 Midi-preparation of Plasmid DNA 73
2.2.2.4 Maxi-preparation of Plasmid DNA 74
2.2.2.5 Plasmid DNA Concentration via Precipitation 76
2.2.2.6 Bacterial Transformation Using Plasmid DNA 76
2.2.3 Genomic DNA Preparation 77
2.2.4 DNA Amplification 77
2.2.4.1 Polymerase Chain Reaction 77
2.2.4.2 DNA Gel Electrophoresis and Imaging 78
2.2.5 Restriction Digest of DNA 79
2.2.6 DNA Extraction from Agarose Gel 80
2.2.7 DNA Ligation 80
2.2.8 Sanger Sequencing of DNA 80
2.3 Gene Editing 82
2.3.1 Guide RNA Design 82
2.3.2 Determining Guide RNA Cutting Efficiency 83
2.3.2.1 Guide RNA Insertion into PX458/pORANGE 83
2.3.2.2 T7 Endonuclease I Assay and ICE Analysis 84

2.3.3 pORANGE Plasmid Creation 87
2.3.4 Fluorescence-Activated Cell Sorting of Edited Cells 91
2.4 Protein Biochemistry 92
2.4.1 SDS-PAGE and Western Blotting 92
2.4.1.1 Western Blot Sample Preparation 92
2.4.1.2 SDS-PAGE 92
2.4.1.3 Transfer of PAGE Gels to PVDF Membrane 93
2.4.1.4 Fluorescent Western Blotting 94
2.4.2 Chemiluminescent Western Blotting 94
2.4.3 In-gel Fluorescence Assays 97
2.5 RNA Biology 97
2.5.1 RNA Extraction from Cell Lines 97
2.5.1.1 Preparing Cells for RNA Extraction 97
2.5.1.2 Trizol/Chloroform RNA Extraction 98
2.5.2 Quantification and Integrity Analysis of RNA 98
2.5.3 Polysome Profiling 99
2.6 Next Generation Sequencing 101
2.6.1 mRNA-Seq Library Preparation 101
2.6.2 mRNA-Seq Read Quality Analysis, Alignment, and Expression Analysis 102
2.7 GTEx Database Analysis 102
2.8 Imaging and Microscopy 103
2.8.1 Acid-washing Coverslips 103
2.8.2 Paraformaldehyde Fixation 103
2.8.3 Immunofluorescence Assays 103
2.8.4 Confocal Microscopy 105
2.8.5 Live Cell Imaging 106
2.8.6 Kymograph Creation 108

3 Creation of CRISPR Constructs and Cell Lines 109

3.1 Introduction . . . 109

3.1.1 Overview of ORANGE CRISPR System . . . 109

3.1.2 pORANGE Cloning Vector Design and Usage . . . 110

3.1.3 Tagging Ribosomal Proteins With HaloTag . . . 114

3.1.4 Use of the PC-12 Cell Line . . . 116

3.2 Results . . . 118

3.2.1 Selection of Ribosomal Protein Tagging Targets . . . 118

3.2.2 Design of Guide RNAs Targeting Ribosomal Proteins . . . 122

3.2.3 CRISPR Donor DNA Design . . . 124

3.2.4 Determination of Guide RNA Cutting Efficiency . . . 124

3.2.4.1 Qualitative Editing Analysis Using T7E1 Assay . . . 126

3.2.4.2 Quantitative Editing Analysis Using ICE Analysis . 130

3.2.5 Creation of pORANGE CRISPR Plasmids . . . 141

3.2.6 FACS Analysis and Creation of PC-12 Knock-In Cell Populations . . . 142

3.3 Discussion . . . 153

3.3.1 Limitations of the ORANGE System . . . 153

3.3.2 CRISPR Scores May Be a Poor Indicator of Guide RNA Editing Efficiency . . . 154

3.3.3 Quantification of Cutting Efficiency Using the T7E1 Assay . . 155

3.3.4 Cutting Efficiency May Be Higher Than ICE Analysis Determined . . . 156

3.3.5 Considerations for FACS Assays . . . 157

3.3.6 CRISPR-Edited PC-12 Cell Lines Need Further Validation To Ensure Correct HaloTag Insertion . . . 159

4 Validation of ORANGE-Edited PC-12 Lines 160

4.1 Introduction . . . 160

4.2 Results . . . 162

4.2.1 Validation of Positive Control Plasmid pORANGE GFP-Actb KI . . . 162
4.2.2 Validation of pORANGE CRISPR Plasmids in PC-12 Cell Lines 166
4.2.2.1 Determining HaloTag Presence Via In-Gel Fluorescence . . . 166
4.2.2.2 RPS3-Halo Cells Lose HaloTag Expression After Several Weeks in Culture . . . 169
4.2.2.3 Verifying the Site of HaloTag Insertion Via Sanger Sequencing . . . 174
4.2.2.4 Validating Ribosomal Protein Tagging by Western Blot . . . 175
4.2.2.5 HaloPROTAC3 Degradation Confirms HaloTag Expression in PC-12 Halo Lines . . . 180
4.2.2.6 Polysome Profiling of Tagged Ribosomes . . . 184
4.3 Discussion . . . 195
4.3.1 The RPS3-Halo Line Does Not Stably Express HaloTag . . . 197
4.3.2 The Identity of the Tagged Protein in the RPL22-Halo Line . 197
4.3.3 PC-12 Cells Express Large Amounts of 40S Subunits . . . 199

5 Visualising Ribosome Transport In Processes Using Live Cell Imaging 201

5.1 Introduction . . . 201
5.2 Results . . . 203
5.2.1 Optimisation of Primary Cortical Neuron Transfection . . . 203
5.2.2 In-Gel Fluorescence and Western Blot Detection of ORANGE CRISPR-Edited Primary Neurons . . . 206
5.2.3 Live Cell Imaging of ORANGE CRISPR Edited Primary Neurons . . . 211
5.2.4 Live Cell Imaging of pORANGE-RPL4-Halo Differentiated PC-12 Cells . . . 218

5.3 Discussion . . . 236
5.3.1 The Knock-In Efficiency of ORANGE Constructs is Extremely Low Using Lipofection . . . 236
5.3.2 Ribosomes Undergo Fast Transport in PC-12 Processes . . . 238
5.3.3 The PC-12 Line as a Neuronal Model System . . . 240

6 Expression of Ribosome-Associated Genes in Brain Regions 241
6.1 Introduction . . . 241
6.1.1 The GTEx Database . . . 243
6.2 Results . . . 244
6.2.1 Most Ribosome Protein Genes Are Downregulated In Brain Regions . . . 244
6.2.2 Ribosome Protein Genes Show Differential Exon Expression in Brain Regions . . . 248
6.2.3 snoRNA Analysis Provides Evidence for Brain-specific rRNA Modification . . . 253
6.2.4 Detection of Brain-Elevated Ribosome Protein Genes by mRNA-Seq Analysis . . . 258
6.3 Discussion . . . 271
6.3.1 Limitations of RNA-Seq Datasets . . . 271
6.3.2 Further RNA-Seq Evidence of Heterogeneous Ribosome Composition in the Nervous System . . . 272
6.3.3 Subsequent Experiments To Prove Neuronal Ribosome Specialisation . . . 274

7 Discussion 277
7.1 Summary of Findings . . . 277
7.2 Considerations for Creating ORANGE Constructs . . . 278
7.3 Building a Model for Ribosome Transport in Axons . . . 280
7.4 Future Work . . . 282

8 Appendix **316**

List of Figures

1.1 Basic Overview of Neuronal Structural Classification 24

1.2 Schematic of Axonal Local Translation in a Typical Vertebrate Multipolar Neuron 33

1.3 Structure of the 80S Human Ribosome 41

1.4 Overview of Ribosome Biogenesis in Higher Eukaryotes 46

1.5 Simplified Structures of C/D and H/ACA snoRNAs 49

3.1 Creation of ORANGE Constructs Using pORANGE 112

3.2 Plasmid Maps of ORANGE CRISPR Plasmids 113

3.3 Crystal Structure of HaloTag Bound to TMR Ligand 115

3.4 Differentiated PC-12 Cells 117

3.5 CRISPR Protein Targets on Eukaryotic 80S Ribosome 120

3.6 PX458 Plasmid Map 127

3.7 T7E1 Assay for Determining Guide RNA Cutting Efficiency 128

3.8 GFP Western Blot of T7E1 Samples 131

3.9 RPL4 Guide 1 ICE CRISPR Analysis for Cutting Efficiency 133

3.10 RPL4 Guide 2 ICE CRISPR Analysis for Cutting Efficiency 134

3.11 RPL22 Guide 2 ICE CRISPR Analysis for Cutting Efficiency 136

3.12 RPS2 Guide 1 ICE CRISPR Analysis for Cutting Efficiency 137

3.13 RPS2 Guide 2 ICE CRISPR Analysis for Cutting Efficiency 138

3.14 RPS3 Guide 1 ICE CRISPR Analysis for Cutting Efficiency 139

3.15 Schematic of Cell Sorting Using FACS 143

3.16 FACS Analysis of pORANGE-RPL4-Halo Transfected PC-12 Cells Labelled with Halo-TMR 148
3.17 FACS Analysis of pORANGE-RPL22-Halo Transfected PC-12 Cells Labelled with Halo-TMR 149
3.18 FACS Analysis of pORANGE-RPS2-Halo Transfected PC-12 Cells Labelled with Halo-TMR 150
3.19 FACS Analysis of pORANGE-RPS3-Halo Transfected PC-12 Cells Labelled with Halo-TMR 151
3.20 FACS Analysis of pORANGE-GFP-ACTB Transfected PC-12 Cells . 152

4.1 PC-12 GFP-Actb Edited Line Fluorescent Western Blot 164
4.2 Immunofluorescence and Confocal Imaging of GFP-ACTB PC-12 Cell Line 165
4.3 Detection of Halo-Tagged Protein Via In Gel Fluorescence 168
4.4 FACS Re-Sort of RPS3-Halo PC-12 Line Labelled with Halo-TMR . . 170
4.5 Second FACS Analysis of pORANGE-RPS3-Halo Transfected PC-12 Cells Labelled with Halo-TMR 171
4.6 In-Gel Fluorescence Assay of Second RPS3-Halo Line Labelled with Halo-TMR 172
4.7 FACS Re-Sort of Second RPS3-Halo PC-12 Line Labelled with Halo-TMR 173
4.8 PCR of PC-12 Halo Line Donor DNA Insertion Region 176
4.9 Sanger Sequencing of RPL4-Halo and RPS2-Halo Insertion Sites . . . 177
4.10 RPL4-Halo PC-12 Fluorescent Western Blot 181
4.11 RPL22-Halo PC-12 Fluorescent Western Blot 182
4.12 RPS2-Halo PC-12 Fluorescent Western Blot 183
4.13 PC-12 Halo Line HaloPROTAC3 Treatment Western Blot 185
4.14 Overview of Polysome Isolation Via Sucrose Sedimentation 187
4.15 Polysome Profiling of WT PC-12 Cells 189
4.16 Polysome Profiling of RPL4-Halo PC-12 Cells 191

4.17 Polysome Profiling of RPS2-Halo PC-12 Cells 193
4.18 Summary of PC-12 Halo Line Validation Results 196

5.1 Schematic of TIRF and HiLo Microscopy 204
5.2 Cortical Neuron Transfection Optimisation Using Lipofectamine Reagents 208
5.3 In-Gel Fluorescence Analysis of Cortical Neuron Transfection with pORANGE-Halo Plasmids 213
5.4 Detection of pHAAP1 Transfection in Cortical Neurons and HEK293FT Cells 215
5.5 Cortical Neuron Transfection With mEGFP-N1 Using Lipofectmine 2000 215
5.6 Live Cell Imaging of pHAAP1 Transfected Cortical Neurons Labelled with Halo-TMR 216
5.7 Live Cell Imaging of pORANGE-GFP-Actb Transfected Cortical Neurons 219
5.8 Live Cell Imaging of pORANGE-RPL4-Halo Transfected Cortical Neurons Labelled with JFX650-Halo 221
5.9 Differentiated RPL4-Halo PC-12 Immunofluorescence Assay 222
5.10 Live Cell Imaging of RPL4-Halo PC-12 Line Labelled with Halo-TMR 226
5.11 Live Cell Imaging of RPL4-Halo PC-12 Cells Using 0.05 nM JFX650-Halo 228
5.12 Live Cell Imaging of RPL4-Halo PC-12 Line Using JFX650-Halo and LysoTracker Green 231
5.13 Single Channel Controls for L4-Halo PC-12 Imaging Using JFX650-Halo and LysoTracker Green 232
5.14 Fluorescence Spectra of JFX650 and LysoTracker Green 233
5.15 Live Cell Imaging of RPS2-Halo PC-12 Cells Using JFX650-Halo . . . 236

6.1 Gene Expression of Ribosome Protein Genes in Human Tissues . . . 248
6.2 Gene Exon Expression Analysis - Consistent Expression Examples . . 250

6.3 Gene Exon Expression Analysis - Differential Expression Examples . 252
6.4 Gene Expression Analysis of SNORA Genes 261
6.5 Gene Expression Analysis of SNORD Genes 263
6.6 Gene Expression Analysis of SCARNA Genes 265
6.7 18S Pseudouridylation Sites Mapped on the 80S Ribosome 268
6.8 18S U1368 is in Close Proximity to Potential Binding Site RPS28 K16 269
6.9 MA Plot of Differential Gene Expression in DIV 0 and DIV 14 Primary Neurons 270

8.1 R Script for ICE Analysis 319
8.2 R Script for FACS Analysis 326
8.3 Loading Controls for Chapter 4 Western Blots 327
8.4 Polysome Profiling Western Blots 328
8.5 R Script for Ribosome Protein and snoRNA Gene Expression Analysis 333
8.6 Exon Expression Heatmaps for Ribosome Protein Genes 344
8.7 R Script for Exon Expression Heatmaps for Ribosome Protein Genes 348

List of Tables

2.1 Neuronal Dissection and Culture Media Composition 70

2.2 Cell Densities and Reagent Amounts Used for Lipofectamine 2000 Transfection 72

2.3 Constructs Used for Cloning, CRISPR Validation, and Transfection . 75

2.4 Composition of Bacterial Culture Solutions 77

2.5 PCR Reaction Mixture Composition 78

2.6 General PCR Protocol 78

2.7 Restriction Enzymes Used in Cloning 79

2.8 Primers Used for Sanger Sequencing 81

2.9 Forward and Reverse Guide RNA Oligos for ORANGE System 82

2.10 Guide RNA Oligo Annealing Reaction Mixture 83

2.11 Guide RNA Duplex Formation Protocol 83

2.12 Guide RNA Oligo Ligation Reaction 84

2.13 Primers Used to Amplify the Genomic Region of Interest for T7E1 Assay 86

2.14 PCR protocol for T7E1 Assay DNA Amplification 87

2.15 PCR Protocol for T7E1 Assay Heteroduplex Formation 87

2.16 Primers for Donor DNA Insertion Into pORANGE 90

2.17 Buffers and Solutions for SDS-PAGE, Transfer and Western Blot . . . 95

2.18 Primary Antibodies Used in Western Blot Assays 96

2.19 Secondary Fluorescent Antibodies for Western Blot assays 96

2.20 Polysome Profiling Solution Compositions 101

2.21 Immunofluorescence Assay Solutions 104
2.22 Primary Antibodies used in Immunofluorescence Assays 105
2.23 Secondary Antibodies Used in Immunofluorescence Assays 105
2.24 Kymograph Velocity Line Measurements 108

3.1 Ribosomal Protein Biogenesis Stage and Exposed Regions 121
3.2 CRISPR Scores for ORANGE Guide RNAs 123
3.3 Guide RNA Locations and Integration Sites 124
3.4 Rationale for Guide RNA Selection 141

6.1 Differentially Expressed Ribosome Protein Isoforms in Brain Regions 255
6.2 Expression of Brain-Elevated snoRNA Genes and Host Genes 266
6.3 The Modification Sites of Brain-Elevated snoRNA Genes 267

8.1 Human Ribosomal Protein Genes Used in Gene Expression Analysis . 329
8.2 snoRNA Genes Used in Gene Expression Analysis 349

Abbreviations

AD Alzheimer's disease

AIS axonal initial segment

ALS amyotrophic lateral sclerosis

DAPI 4',6-diamidino-2-phenylindole

ddH_2O double-distilled H_2O

DIV days in vitro

DMEM Dulbecco's Modified Eagle Medium

DSB double-strand break

EDTA ethylenediaminetetraacetic acid

EGFP enhanced green fluorescent protein

EM electron microscopy

FACS Fluorescence Activated Cell Sorting

FBS foetal bovine serum

FISH fluorescence in situ hybridisation

FOV field of view

FRAP fluorescence recovery after photobleaching

FSC forward scatter

GFP Green Fluorescent Protein

GTEx Genotype-Tissue Expression

HA hemagglutinin

HDR homology-directed repair

HiLo Highly Inclined and Laminated Optical sheet

mRNA messenger RNA

NGF nerve growth factor

NHEJ non-homologous end joining

NOR nuclear organiser regions

ORANGE Open Resource for the Application of Neuronal Genome Editing

P-S penicillin-streptomycin

PAM protospacer adjacent motif

PBS phosphate buffered saline

PCR polymerase chain reaction

PFA paraformaldehyde

PIC preinitiation complex

PLL poly-L-lysine

PolI RNA Polymerase I

PROTAC Proteolysis targeting chimera

RCF relative centrifugal force

rDNA ribosomal DNA

RGC retinal ganglion cell

RPM revolutions per minute

rRNA ribosomal RNA

scaRNA small Cajal-body specific RNA

SMA spinal muscular atrophy

SMN1 survival motor neuron 1

snoRNA small nucleolar RNA

snoRNP small nucleolar ribonucleoprotein

SSC side scatter

T7EI T7 endonuclease I

TBST Tris-buffered saline with Tween 20 detergent

TCA trichloroacetic acid

TIRF total internal reflection fluorescence

TMR tetramethylrhodamine

TPM Transcripts per Million

TRAP translating ribosome affinity purification

tRNA transfer RNA

UTR untranslated region

WT wild type

Chapter 1

Introduction

1.1 The Necessity of Axonal Transport in Axons

Neurons are specialised cells responsible for the transmission of electrical signals, known as action potentials, which are essential for human life. Their function is assisted by the distinct and polarised morphology of neurons, which are composed of a cell body (or "soma") and a variable number of neuronal processes containing up to 99% of the cell's cytoplasm (Holt et al. 2019). Neuronal processes can largely be categorised into two groups; dendrites (responsible for receiving incoming action potentials), and axons (responsible for the propagation of action potentials to other neurons in the network). These processes come into contact to form structures known as synapses - small junctions between dendritic and axonal terminals - which are the site of signal transfer between neurons. Notably, there exists many types of neuron which differ in structure and polarity (Figure 1.1). In vertebrates, multipolar neurons are the most abundant neuron type in the central nervous system (Jennes 2017), consisting of multiple dendritic processes and a single axonal outgrowth which can grow to be up to one meter long in humans (Prokop 2020).

Neuronal function is supported by long-distance trafficking of essential cellular cargo into neuronal processes (reviewed by Guedes-Dias and Holzbaur 2019). In neurons, most protein required for cellular function is synthesised by ribosomes in the cell body. Due to the large spatial separation of the neuronal processes from

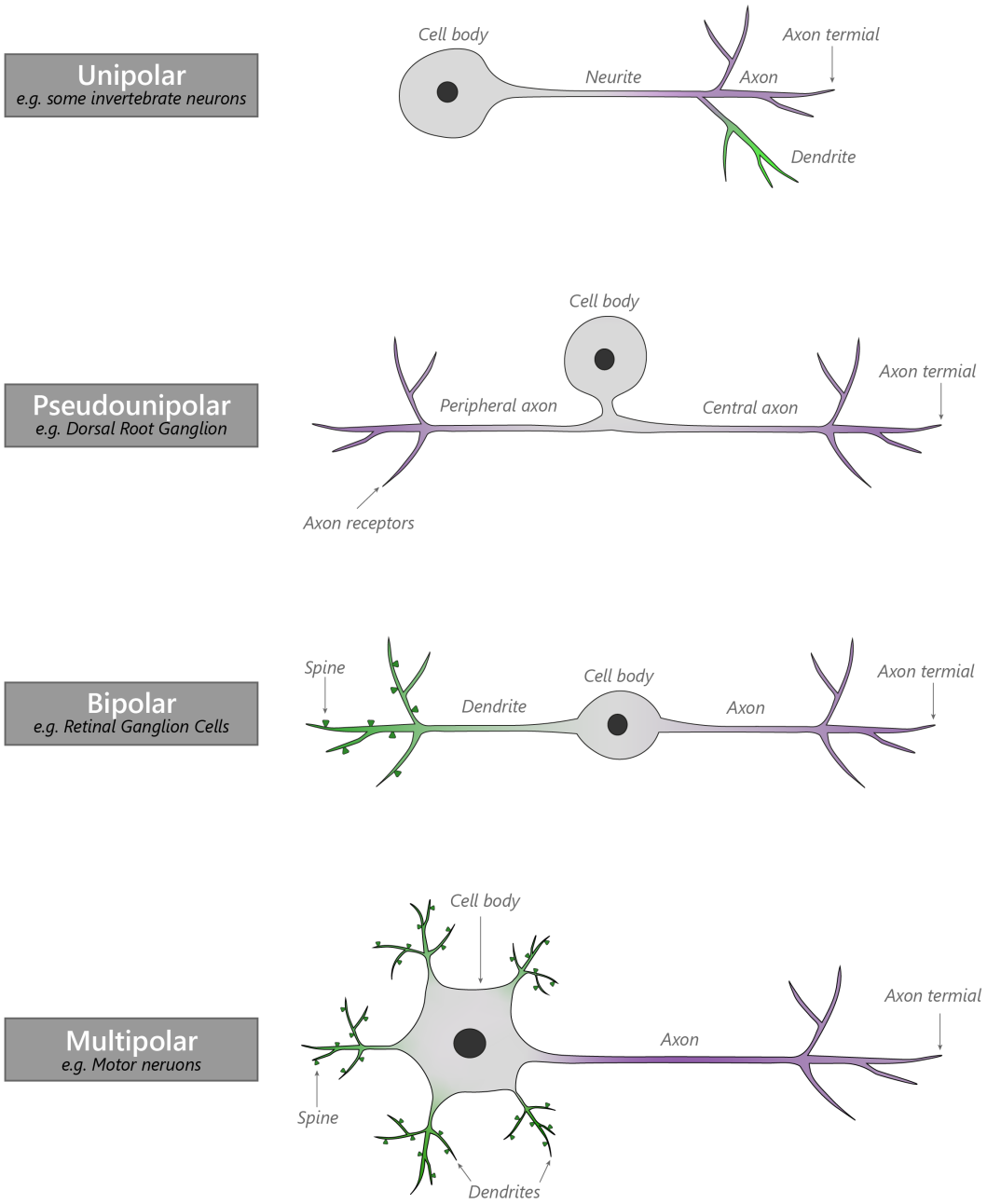


Figure 1.1: Basic Overview of Neuronal Structural Classification.

Neurons can largely be classified into 4 groups according to their polarity. Unipolar neurons have a single neurite extending from the cell body and are most commonly found in the central nervous system of invertebrates (such as *Aplysia*). Pseudounipolar neurons are sensory neurons found in the peripheral nervous system, characterised by a single axon which bifurcates into two branches. Bipolar neurons are also sensory neurons which have both a dendrite and an axon extending from the cell body. Multipolar neurons have an axon and two or more dendrites extending from their cell body and are the most abundant neuron type in vertebrates.

their cell body, this protein must be actively transported to the correct location to support neuronal function. Delivery of cellular cargo into the distal processes cannot be achieved through diffusion alone (Scott et al. 2011). For example, in a one meter long axon it would take approximately 40 years for a ribosome to diffuse from the cell body to the axon terminal (using a simplified equation where time in seconds = (distance in cm)^2 / (2 x diffusion coefficient in cm^2 per second) (Schavemaker et al. 2018). This calculation uses prokaryotic 70S ribosome diffusion coefficient of 0.04 µm^2/second (Bakshi et al. 2012) and is therefore likely to be an underestimate for 80S eukaryotic ribosome diffusion. Additionally, in most vertebrate neurons, the cell body and axon are connected by a structure known as the axonal initial segment (AIS) which is important for the generation of action potentials. The AIS can act as selectivity filter preventing passive diffusion of proteins and lipids into the axonal cytoplasm (reviewed in Leterrier 2018). Therefore, the intracellular transport of cargo into axons and dendrites requires the activity of specialised transport machinery - including microtubules, motor proteins, and their adaptors. Transport of cellular cargo is especially important in axons, where defects in the transport machinery have been implicated in neurodegenerative diseases (reviewed in De Vos and Hafezparast 2017; Sleigh et al. 2019; Guillaud et al. 2020).

Microtubules are an essential part of the intracellular transport machinery, acting as molecular tracks for the movement of motor proteins and playing an important role in many neuronal functions - including cargo trafficking, migration, neuronal development and signalling (reviewed in Kapitein and Hoogenraad 2015). Composed of α-tubulin and β-tubulin heterodimers which bind in a head-to-tail fashion to form a linear protofilament, a microtubule is formed from 13 protofilaments which arrange to form a hollow tube structure. In neurons, microtubules also define the polarisation of the cell, which is a key factor in the direction of cargo transport. Microtubules are comprised of a fast-growing end with a high rate of tubulin polymerisation (the plus end), and a more stable, slow-growing end (the minus end) (reviewed in Guedes-Dias and Holzbaur 2019). In axons, arrays of microtubules are found to be unidirectional

in nature, with microtubule plus ends orientated away from the cell body (Burton and Paige 1981). The mechanism of motor protein transport relies on recognition of plus and minus ends, resulting in the directed movement of motor proteins from the cell body to the axon terminal (anterograde transport) or vice versa (retrograde transport). These cytoskeletal elements work in unison to delivery cellular cargo to distal processes, supporting neuronal activity and survival.

Axonal transport of cellular cargo is conducted by motor proteins, such as kinesin and dynein (reviewed by Maday et al. 2014). These molecular machines are a key component of the cytoskeleton and rely upon microtubules and adaptor proteins to bind and transport a range of cargo in a processive, ATP-dependent manner. Different cellular cargos are transported at different speed within neuronal processes. It has been shown that smaller cargo (such as soluble proteins and cytoskeletal components) are transported at speeds of <0.1 µm/s - defined as "slow" axonal transport. Slow transport can be further divided into two subtypes, slow component A (SCa) and component B (SCb) (reviewed in Roy 2014). At speeds of 0.2–1 mm/day, SCa is the slowest form of transport responsible for the movement of tubulin and neurofilaments, whereas actin and soluble proteins move at 1–10 mm/day by SCb transport. In contrast, larger cargo (organelles, vesicles and mRNA) can be transported at speeds of up to 0.6 - 5 µm/s, commonly referred to as "fast" axonal transport.

Anterograde transport in neurons is carried out by the motor protein kinesin (Hirokawa et al. 1991). The human kinesin superfamily is diverse, with 14 distinct families which vary in structure, cargo-preferences, and tissue expression (Nobutaka Hirokawa et al. 2009). However, most anterograde transport in neurons is mediated by Kinesin-1, a processive motor which hydrolyses ATP to travel in regular 8 nm steps along microtubule filaments towards their plus ends (Svoboda et al. 1993). The heterodimeric nature of kinesin enables its transport capabilities. Kinesin-1 microtubule binding and walking is conducted by its motor domain, encoded by the *KIF5A*, *KIF5B* and *KIF5C* heavy chain genes. This domain is connected by

a flexible stalk region to a cargo-binding tail domain, which associates with kinesin light chain (KLC) proteins. A wide variety of cargo is known to undergo kinesin-mediated transport, including proteins, vesicles, RNA, and organelles. The speed of this transport is also well characterised, with velocities ranging from 0.5 – 1 µm/s in axons (Maday et al. 2014).

In addition to Kinesin-1, there are several other processive kinesins which are responsible for long-distance cargo transport in neurons (Maday et al. 2014). For example, Kinesin-2 is a kinesin motor involved in a diverse range of intracellular transport processes in various cell types, including neurons. Studies into Kinesin-2 mediated transport (reviewed in Scholey 2013) have shown that this motor is responsible for the anterograde movement of fodrin-bound vesicles (Takeda et al. 2000) and the enzyme choline acetyltransferase (Ray et al. 1999), enabling the processes of neurite outgrowth and neurotransmission respectively. Intriguingly, the Kinesin-2 motor is variable in structure, with both heterotrimeric and homodimeric motors observed which show different transport qualities (Scholey 2013). In addition to Kinesin-2, long-distance axonal transport also relies on the activity of Kinesin-3 (reviewed in Siddiqui and Straube 2020). Unlike Kinesin-2, Kinesin-3 is predominantly involved in neuronal cargo transport. Upon cargo binding, kinesin-3 motors dimerise to form a highly processive motor (Soppina et al. 2014) where it has been shown to deliver various cargos, including synaptic vesicle precursors (Okada et al. 1995) and dense core vesicles (Lipka et al. 2016).

In contrast, transport of materials from the axon to the cell body (retrograde transport) is mediated by the motor protein dynein (reviewed in Reck-Peterson et al. 2018). Cytoplasmic dynein (otherwise known as Dynein-1) is well characterised due to its role in both mitotic spindle formation and axonal transport. Dynein-1 is composed of a homodimer of heavy chains, with the C-terminal regions (the motor domains) forming a ring-like structure made of six AAA+ domains. The N-terminus of the heavy chains is the tail domain, which when associated with dynein intermediate and light chains, is responsible for cargo binding. Like Kinesin-

1, Dynein-1 also moves in 8 nm steps along microtubules in an ATP-dependent manner, although dynein's step size has been shown to vary depending on its load (Mallik et al. 2004). Dynein-1 requires an additional accessory factor, dynactin, to enable processive cargo transport (Schlager et al. 2014), which binds to Dynein-1 alongside activation factors (such as BICD2) that provide cargo binding specificity (reviewed in Tempes et al. 2020). Interestingly, single molecule assays have shown that dynein movement is not unidirectional. Instead, the motor is capable of moving sideways and backwards on the microtubule, with similar transport velocities to kinesin (0.5 – 1 µm/s) (Maday et al. 2014).

1.2 Local Translation in Neurites Supports Neuronal Function

Although most protein required for cellular function is synthesised in the cell body, a multitude of essential proteins are also translated locally in neuronal processes (reviewed in Biever, Donlin-Asp, et al. 2019). Local translation in neuronal processes is an active area of research, providing insights into how the proteins expressed by neurons at a given time (the proteome) is dynamically regulated in response to environmental cues. Recent efforts to understand local translation in neurons have predominantly focussed on dendrites and spines, revealing a fundamental requirement for local translation to maintain and regulate the postsynaptic proteome (reviewed in Holt et al. 2019). However, advancements in gene editing and imaging techniques have demonstrated that local translation also occurs within mature axons, allowing stricter spatio-temporal regulation of essential presynaptic proteins (Hafner et al. 2019). In addition to synaptic function, local translation in neurites has also been implicated in several major neuronal functions, including neuronal development, synaptic function and injury response (Figure 1.2, see Section 1.2.1).

Despite mounting evidence that axonal local translation is critical for neuronal survival, the ability for adult axons to synthesise proteins locally was historically

controversial. The lack of research into axonal translation can be attributed to early axonal transport experiments in the 1970s, where analysis of giant squid neuronal axoplasm incorrectly determined that no ribosomal RNA (rRNA) (and thus no ribosomes) was present (Lasek, Dabrowski, et al. 1973). Therefore, it was incorrectly assumed that protein synthesis in axons was minimal or non-existent. This notion prevailed for several decades despite some evidence to the contrary, which garnered little attention from the scientific community. The rRNA content of giant squid axoplasm was re-visited several years later, leading to the discovery that several species of rRNA could be detected, albeit in small amounts (Giuditta, Cupello, et al. 1980). To this end, it is likely that axonal translation research has been significantly hindered due to low ribosome availability which requires more sensitive approaches, especially when compared to ribosome-abundant dendrites. Additionally, studying local translation in axon terminals presents unique challenges that are technically demanding to overcome, with axon size and close association with the post-synapse being a limiting factor (Biever, Donlin-Asp, et al. 2019). Research into axonal translation has only recently started to progress, although no direct evidence of ribosome transport to mature axon terminals currently exists.

Initial evidence of local translation in neurites dates back to the 1960s, where biochemical assays investigating acetylcholinesterase generation in neurons illustrated the possibility of local protein synthesis. Using a simple experiment, researchers treated cats with an acetylcholinesterase-inhibiting compound, disopropyl-fluorophosphate in order to reduce the levels of the enzyme within neurons and observed the pattern of recovery over several days (Koenig 1965). Considering the accepted theory of neuronal protein production at the time, it was expected that newly synthesised acetylcholinesterase would appear as a gradient, concentrated within the cell body with lower amounts extending into the processes. In contrast, the restoration was instead observed to be uniform along the nerve, indicating local protein synthesis was occurring within neurites. Although these experiments provided exciting preliminary data of local protein production, many researchers

remained sceptical claiming there may be possible interference of protein synthesis from the cell body (Holt et al. 2019). Subsequent experiments utilised more sophisticated labelling techniques, often isolating the axon to illustrate local translation activity. One such landmark study came several years later, again utilising isolated giant squid axons to analyse axonal protein synthesis (Giuditta, Dettbarn, et al. 1968). *Giuditta et al.* removed axons from their cell body and measured the incorporation of radioactive amino acids into the protein fraction of the axoplasm. They found that a significant amount of radioactive protein was synthesised after several hours, with the production of labelled proteins also being strongly hindered by protein synthesis inhibitors (including puromycin and cycloheximide). Despite this evidence supporting axonal translation it remained a controversial topic, with many disregarding the aforementioned study as an outlier that does not reflect most neuronal populations (Rangaraju et al. 2017).

Subsequent work on neuronal local translation predominantly focussed on the presence of translational machinery within processes (reviewed in Holt et al. 2019). Ribosomes localised in neuronal processes carry out local protein synthesis, translating a messenger RNA (mRNA) population in response to extracellular stimuli. Initial work on mRNA localisation in dendrites used the newly devised fluorescence in situ hybridisation (FISH) technique, whereby analysis of rat brain slices identified mRNAs coding Ca2+/calmodulin-dependent protein kinase type II (CaM-KII) subunits in dendrites (Burgin et al. 1990). This discovery pointed to local CaM-KII synthesis in postsynaptic compartments. Since this initial finding, mRNA transcripts for several important postsynaptic proteins have been discovered in neuronal processes, including the microtubule-stabilizing protein MAP-2 (Kleiman et al. 1990). Efforts to understand the diversity and abundance of mRNA transcripts present in neuronal processes was first conducted on rat hippocampal neurites (Miyashiro et al. 1994), establishing that a large number of mRNA transcripts are localised to dendrites. With the development of microarray technology, approximately 100 neurite-localised transcripts were identified (Poon et al. 2006) using the mechanical

separation of the neuronal processes to prevent somatic contamination. In particular, the authors noted the dendritic transcripts identified were predominantly associated with translational machinery and the RNA-binding protein, Staufen.

Further complexity of mRNAs localised to neuropils (regions of the central nervous system with high synaptic density, containing both unmyelinated axons, dendrites and glia) was identified by *Cajigas et al.* Deep sequencing of hippocampal synaptic neuropils uncovered 2,550 local transcripts; coding for receptors, signalling molecules and scaffolds which are important in synaptic function (Cajigas et al. 2012). The sheer diversity and abundance of mRNA in neuropil processes highlighted the possibility that local translation is a fundamental process in neurons, and may not be limited to neural plasticity as previously assumed (Biever, Donlin-Asp, et al. 2019). Research into the structure of neuropil-localised mRNA transcripts has also identified features within their 3' and 5' untranslated region (UTR) which influence their trafficking into processes. Localisation is thought to mostly be mediated by 3' UTR motifs which are essential for RNA-binding protein recognition and transport (Eom et al. 2003), although 5' UTR elements may also play a key role in the sorting and localisation of some transcripts (Colliva and Tongiorgi 2021). Several mRNA cis-regulatory elements have been identified and characterised, and it is likely that multiple elements work in tandem to generate mRNA localisation (Mendonsa et al. 2023). In support of this, studies on hippocampal mRNAs in mature neurons uncovered that somatic and neuropil-enriched mRNA transcripts also have significantly longer 3' UTR isoforms than their non-localised counterparts, increasing the number of novel regulatory elements present (Tushev et al. 2018). Neuropil-derived mRNAs also showed increased levels of enrichment for proteins associated with synapse and dendritic spine function - including CamKIIa. This, in combination with longer half-lives and increased transcript stability (Tushev et al. 2018), provides a strong indication that this mRNA population has an essential and broad functional role within neurons.

Similar to dendrites, mature axons also have a wide and varied collection of

mRNA available for local translation (reviewed in Das et al. 2021). The presence of mRNA in axons was first discovered via axoplasm translation studies in squid axons (Antonio Giuditta and Santella 1986), although no evidence of active translation of axonal mRNA was observed at this time. Additionally, the necessity of local translation in *Aplysia* neurons for synaptic plasticity also provided evidence that critical protein synthesis occurs in the presynaptic compartment (Martin et al. 1997). Advances in ribosome tagging strategies have enabled the translatome (all mRNA transcripts being translated) of mature axons of the retinal ganglion cell (RGC) to be studied via translating ribosome affinity purification (TRAP) (Shigeoka, H. Jung, et al. 2016; Shigeoka, J. Jung, et al. 2018). The novel axon-TRAP-RiboTag mouse utilised a previous RiboTag mouse (Sanz, Yang, et al. 2009) and was engineered to tag ribosomes *in vivo* through the addition of a hemagglutinin (HA) tag to the C-terminal of the large ribosomal subunit protein, RPL22. Expression of this modified allele is controlled by the cre-recombinase promoter, allowing transient and tissue-specific expression. The translatome of the RGC axons was studied during cell development until maturation. Two distinct classes of translatome - developmental stage-dependent and constitutive - were identified. *Hafner et al.* adopted a different methodology to analyse the mRNA population in presynaptic terminals, using FISH and expansion microscopy to study presynaptic and postsynaptic compartments in mouse brain slices (Hafner et al. 2019). Metabolic labelling experiments in excitatory boutons selected using fluorescent synaptosome labelling showed more than 30% were actively translating transcripts. Furthermore, purification and mRNA sequencing of pre-synaptic nerve terminals showed approximately 450 mRNA transcripts were enriched in the pre-synapse, including proteins involved in neurotransmitter release. Both studies indicate a significant abundance of active zone synaptic proteins transcripts in axons.

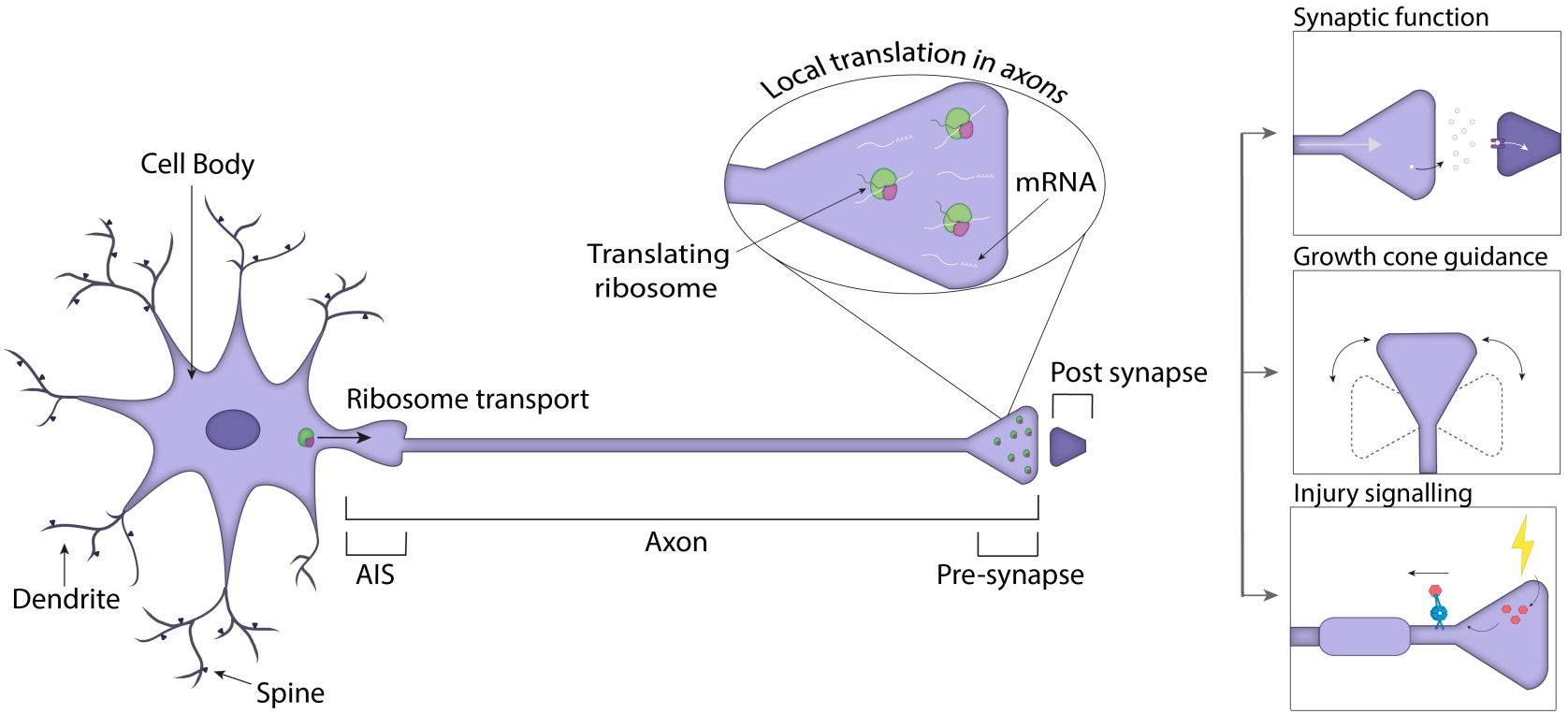


Figure 1.2: Schematic of Axonal Local Translation in a Typical Vertebrate Multipolar Neuron.

Local Translation is carried out in axons, which supports a wide range of neuronal functions. mRNA transcripts are localised to axons and pre-synaptic terminals where they are translated by ribosomes to support the local proteome, which is essential for neuronal development, synaptic function, and injury signalling.

1.2.1 The Varied Roles of Axonal Local Translation

Axonal translation is an essential and often overlooked process, despite its involvement in many neuronal processes (L. Li et al. 2021). Historically, axonal translation was thought to be restricted to the early axon during neuronal cell development, which may primarily be due to the lack of polyribosome detection in mature axons. Before the use of advanced ribosome imaging techniques, clusters of ribosomes were first identified in the early axon of embryonic mice dorsal root neuroblasts (Tennyson 1970). Following this, polyribosomes were also detected in the growth cones of embryonic cortical neuronal cultures, in close association with microtubules (G. J. Bassell et al. 1998). For a considerable period, however, ribosomes and rRNA were not detected in adult axons in EM studies (Bartlett and Banker 1984), and were assumed to not be present at significant levels (Biever, Donlin-Asp, et al. 2019). (See Section 1.4 for a discussion as to why axonal ribosomes are difficult to detect). However, recent research has associated the locally translated proteome of axons with a variety of roles, including neuronal development (reviewed in Agrawal and Welshhans 2021), synapse formation and maintenance (reviewed in Holt et al. 2019), injury response (reviewed in Koley et al. 2019), and neuronal survival (reviewed in Cioni et al. 2018).

Local translation has been associated with axonal development for some time. One well studied example is axonal pathfinding, whereby the end of the developing axon (referred to as the growth cone) navigates to form neuronal connections in response to environmental guidance cues (Agrawal and Welshhans 2021). Early experiments by *Campbell and Holt* determined that inhibition of local translation and protein synthesis in growth cones prevents the turning response in chemotrophic gradients of netrin-1 and Sema3A (Campbell and C. E. Holt 2001). By treating *Xenopus* retinal neuron growth cones with protein synthesis inhibitors (anisomycin and cycloheximide), growth cone collapse by Sema3A was prevented and repulsive turning was abolished. A similar response was also observed to netrin-1, where both attractive and repulsive movements were blocked and growth cones travelled with

no directional preference. This work illustrates the requirement for local translation in axon terminals to respond to environmental cues and rapidly alter the local proteome. Further work determined that the asymmetrical local synthesis of β-actin on the near side of a turning growth cone generates an attractive turning response (Leung et al. 2006). Although the observed turning response in this study was netrin-1 dependent, there is also evidence of similar asymmetrical synthesis of β-actin influenced by brain-derived neurotrophic factor (Yao et al. 2006). In addition to axonal pathfinding, local translation has also been associated with other developmental processes, including axonal branching (Wong et al. 2017) and elongation (Hengst et al. 2009).

As previously established, many essential transcripts are localised to the presynapse which undergo local translation within the axon (Shigeoka, H. Jung, et al. 2016; Hafner et al. 2019). Collectively, there is now strong evidence to support the essential role of local translation in the formation, regulation and maintenance of the presynapse. Synapse formation (a process known as synaptogenesis) has been determined to be reliant on the localisation and translation of neurotransmitter sensorin mRNA in *Aplysia* sensory neurons (Lyles et al. 2006). After *Aplysia* neurons make contact with a motor neuron, sensorin mRNA (and subsequently, sensorin protein) was found to rapidly accumulate in pre-synaptic compartments. By decreasing the concentration of sensorin mRNA using dsRNA, the authors also confirmed that synaptogenesis was inhibited. These findings are supported by work on SNAP25, which has been shown to be synthesised locally for formation of presynaptic terminals in hippocampal neurons (Batista et al. 2017). Subsequent work on synaptic local translation also delivered evidence that presynaptic protein synthesis supports proper synaptic function. Specifically, β-catenin mRNA and ribosomes were discovered to associate with newly formed presynaptic terminals in adult hippocampal neurons (Taylor et al. 2013). Translation of β-catenin mRNA was implicated in the regulation of presynaptic vesicles dynamics, with axonal translation of β-catenin linked to a reduction of synaptic vesicle exocytosis. The regulation of synaptic ac-

tivity is also closely linked to synaptic plasticity, where the requirements for axonal translation have been well characterised (Holt et al. 2019).

Injury response is another major role for axonal translation. Following injury, response signals must be transported retrogradely to the cell body in order for a transcriptional response to be carried out (Koley et al. 2019). In a landmark study, *Hanz et al.* showed that nuclear import factors importin α and β are present in axons, and importin β levels increased significantly in the distal axon following nerve lesion injury (Hanz et al. 2003). This allows the formation of a complex containing a nuclear localisation signal which is retrogradely transported to the cell body. Subsequent research on importin β found that its 3' UTR targets the localisation of the mRNA to the axon, with axonal knockout affecting cell body response to injury signalling and delaying recovery *in vivo* following sciatic nerve crush (Perry et al. 2012). The requirement for local axonal translation for several injury response proteins is now established, including STAT3 (Ben-Yaakov et al. 2012), RanBP1 (Yudin et al. 2008) and the nuclear receptor PPARγ (Lezana et al. 2016).

1.2.2 Defective Local Translation is Implicated in Neurodegenerative Disease

A number of major neurodegenerative disorders have been associated with defective axonal local translation, including amyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA), and Alzheimer's disease (AD). As the connection between defective axonal transport machinery and neurodegenerative disease is well established (reviewed in Sleigh et al. 2019; De Vos and Hafezparast 2017), the mislocalisation of translational machinery (such as mRNA and ribosomes) within axons can clearly result in neuronal dysfunction. To this end, several mechanisms have been elucidated which link defective axonal mRNA transport with reduced neuronal function and survival (Fernandopulle et al. 2021; Khalil et al. 2018). Most of these disease-causing pathways involve aberrant transport of mRNA binding proteins and

their associated mRNA transcripts, resulting in impaired local protein synthesis in the distal axon.

ALS is a common neurodegenerative disorder that affects both the upper and lower motor neurons. It is characterised by progressive muscular atrophy which results in paralysis and death, typically due to respiratory failure. Although most ALS cases are sporadic in nature, it is thought that 10% of cases are familial (Khalil et al. 2018), with mutations in key genes being inherited and leading to the onset of disease. Most ALS research predominantly focuses on identifying these genes and understanding how they contribute to disease pathology. One key contributor to ALS development is the TAR DNA-binding protein gene, which produces the DNA/RNA binding protein TDP-43. Aggregates containing TDP-43 are a classic hallmark of ALS, with approximately 97% of patients observed to have cellular aggregates post-mortem (Suk and Rousseaux 2020). TDP-43 protein has a variety of functions within healthy cells, mainly involving regulation of RNA processing (including mRNA translation, transcript splicing, and mRNA localisation). Notably, TDP-43 is known to assist in the transport of mRNA molecules in neuronal processes (Alami et al. 2014), associating with complexes known as RNA granules (see Section 1.5.2 for a thorough discussion of RNA granule-based transport). Composed of mRNA transcripts, RNA-binding proteins, and ribosomes, RNA granules are actively transported along microtubules by motor proteins until they reach their cellular destination. RNA-binding proteins, such as TDP-43, are thought to play an important role in granule formation and preventing active translation by ribosomes while transport is occurring (Gamarra et al. 2021).

One important target of TDP-43-mediated transport is futsch mRNA, encoding futsch protein which is required for synapse formation at neuromuscular junctions in *Drosophila* (Coyne et al. 2014). By studying TDP-43 in *Drosophila* ALS models, researchers discovered that the amounts of futsch mRNA (and protein) was significantly reduced at neuromuscular junctions. Additionally, *Coyne et al.* determined that TDP-43 has a regulatory function, whereby TDP-43 association re-

sults in futsch mRNA populations shifting from active polysome translation to non-translating RNA granule formation. The mammalian homologue of futsch, MAP1B, also showed similar localisation patterns in the spinal cord of ALS patients. Therefore, it is possible that the mislocalisation of MAP1B mRNA in the distal axon causes cytoskeletal defects which may result in impaired neuronal function and disease (Gamarra et al. 2021). Additionally, as MAP1B is has an important function in microtubule stability, it has been suggested that defective MAP1B mRNA transport into axons has a deleterious effect on the axonal transport and translation of other synaptic mRNAs (Khalil et al. 2018). This mechanism explains one possible role that local translation plays in ALS development, however, several other proteins (including FUS and C9orf72) have also been implicated in axonal translation defects and impaired synaptic function in ALS patients (Nagano and Araki 2021a).

Similarly to ALS, impaired RNA-binding protein function and mRNA localisation has also been implicated in SMA pathology. SMA, a progressive neurological disorder of the spinal motor neurons, is the second most common autosomal recessive disease with an incidence of up to 1:10,000. However, unlike ALS, the cause of SMA is mostly be attributed to pathogenic mutations in a single gene - survival motor neuron 1 (SMN1) (Gamarra et al. 2021). This results in decreased expression of SMN protein, a vital component of RNA granules requiring the activity of a neuron-specific RNA binding protein, HuD (Hao le et al. 2017). Reduced RNA granule assembly impacts the translocation of several important mRNA populations, including β-actin (Rossoll et al. 2003) and GAP43 (Fallini et al. 2016). Both proteins are essential for the maintenance of the cytoskeleton in growth cones, therefore, the mislocalisation of these transcripts adversely affects synaptic function and axonal development.

Lastly, there is a strong case for the role of defective axonal translation in the development of AD (reviewed in Gamarra et al. 2021). As the leading cause of dementia, AD is a well-characterised neurodegenerative disease where the development of β-amyloid (Aβ) plaques is a classic hallmark of its pathology. One form of

β-amyloid, $A\beta_{1\text{-}42}$, is generally considered to be the main aggregate of AD plaques (Nagano and Araki 2021b). Treatment with $A\beta_{1\text{-}42}$ in mouse brains stimulated the axonal translation of several proteins, including the transcription factor ATF4 (Baleriola et al. 2014). Synthesised ATF4 is retrogradely transported to the cell body, where it then induces specific changes in gene expression which result in neuronal death. $A\beta_{1\text{-}42}$ has also been shown to contribute to the abnormal localisation and translation of Tau mRNA in dendrites (C. Li and Götz 2017), leading to hyperphosphorylated Tau aggregates - another classic hallmark of AD pathology.

These examples of pathological impaired local translation highlights the requirement for the axon to dynamically change its local proteome to support a range of cellular roles. Although the pathways presented herein mainly involve aberrant mRNA localisation, there is evidence that impaired transport of ribosomes into axons may also contribute to neuronal dysfunction. For example, mouse models of SMA have shown a 27% depletion in axonal ribosomes compared to WT neurons (Bernabò et al. 2017). Despite this promising data, the precise mechanisms of ribosome transport to the distal axon are still unknown. Further advancements in determining this transport pathway require a deeper understanding of the eukaryotic ribosome structure and function, and importantly, how their composition can be manipulated to enable visualisation of ribosome translocation *in vivo*.

1.3 Ribosomes Are an Essential Component of the Translational Machinery

The translation of mRNA transcripts into protein is carried out by ribosomes, large, complex molecular machines composed of ribosomal proteins and rRNA (Figure 1.3). The human 80S ribosome contains approximately 80 ribosomal proteins, which alongside 4 rRNAs, creates two subunits which differ in size and composition. The small (40S) subunit is formed from 33 ribosomal proteins and the 18S rRNA, whereas the large (60S) subunit contains several rRNAs (5S, 5.2S and 28S) and 47 ribosomal

proteins (Khatter, Myasnikov, Natchiar, et al. 2015). When actively translating mRNA, the two subunits transiently join together to form the 4.3 MDa 80S ribosome. With mammalian cells containing up to 10 million ribosomes per cell (Raveh et al. 2016), the ribosome is one of the most abundant and essential components of all eukaryotic organisms. This is reflected in the RNA composition of the average mammalian cell, where it is estimated that 80 - 90% of an average cell's total RNA is rRNA, whereas mRNA totals just 3 - 7% (Palazzo and Lee 2015).

Despite its critical role in all cellular life, the ribosome was only discovered around 70 years ago, by George Palade (Palade 1955). Palade utilised electron microscopy (EM), a relatively advanced technique for the time, to investigate the cellular contents of rat and chicken tissue sections. He observed an abundance of small, roughly spherical particulates within the cytoplasm - which he thereafter refers to as "granules". Although generally small in size (approximately 100 to 150 Å), it was noted that the size of granules varied, with larger structures of approximately 200 - 300 Å observed in striated muscle tissue. Granules were found in all 40 cell types analysed, although their numbers were found to vary greatly between different cell types. He also noted that the difference in granule density was seemingly associated with the proliferation rate of the cell type. Importantly, Palade found that in some cell types the granules were predominantly located on the membrane of the endoplasmic reticulum, in addition to those found existing freely in the cytoplasm. This paper on cellular granules was the first extensive exploration of ribosomes within mammalian cells, which corroborated earlier biochemical work also describing the presence of small particulates, or "microsomes", within cells.

Following this discovery, interest and research on ribosome composition and function increased rapidly (reviewed in Genuth and Barna 2018). At a seminal research symposium of the Biophysical Society in 1958, current research on microsome particles was collated and discussed at length, allowing several common findings to come to light (Roberts 1958). Notably, the particles were determined to be a mixture of both protein and RNA, with Roberts commenting that most scientists present

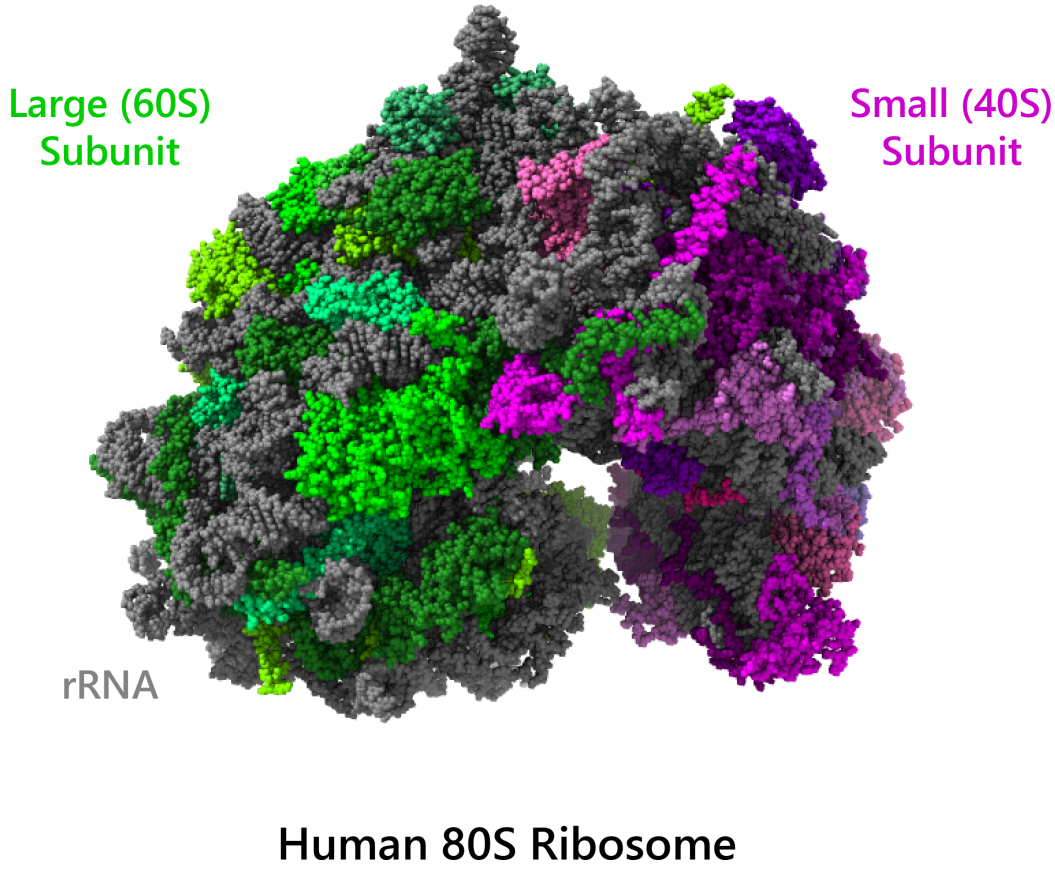


Figure 1.3: Structure of the 80S Human Ribosome.

The human 80S ribosome structure was obtained from Protein Data Bank (PDB ID: 4UG0, DOI:10.2210/pdb4UG0/pdb) (Khatter, Myasnikov, Natchiar, et al. 2015) and modelled using UCSF ChimeraX, version 1.4 (Pettersen et al. 2021). The large (60S) subunit is formed from 47 large ribosomal proteins (green) in conjunction with the 5S, 5.2S, and 28S rRNAs (grey). This subunit transiently joins with the small (40S) subunit, which is composed of 33 small ribosomal proteins (pink) and the 18S rRNA (grey).

maintained "faith that the particles are an important part of the machinery for protein synthesis." This meeting also highlighted issues with the name "microsome particles", as this term caused confusion amongst attending researchers due to their differing definitions of microsome. Instead, the name "ribosome" was then suggested for the particles, which was accepted by the attending scientists as it effectively described the ribonucleoprotein nature of the particles whilst also having a "pleasant sound".

Initial efforts to decipher the structure of the ribosome focussed on prokaryotic ribosomes, perhaps owing to their smaller size, simpler composition, and easier purification. Early efforts to resolve the 70S ribosome were assisted by advances in X-ray crystallography, allowing the structures of the small (30S) and large (50S) subunits to be resolved in bacteria (Ban et al. 2000; Schluenzen et al. 2000). This gave insights into the number of ribosomal proteins and rRNAs present, their size, and the complex nature by which they assemble and stabilise together. Work on the eukaryotic 80S began to advance swiftly soon after. In a landmark 2001 study, a cryo-EM map of the 80S ribosome was calculated at a 15Å resolution (Spahn et al. 2001), using cryo-EM data of ribosomes isolated from *S. cerevisiae* (Beckmann et al. 2001). The core of the 80S ribosome was determined to show high similarity to the bacterial 70S ribosomes that had previously been resolved, indicating a strong level of conservation between prokaryotic and eukaryotic ribosomes. However, this work also highlighted an additional level of complexity in the structure of 80S ribosomes, with *S. cerevisiae* ribosomes being significantly larger, having additional surface-exposed ribosomal proteins, longer rRNA, and increased RNA-RNA intersubunit bonds. Subsequent cryo-EM studies helped place several more ribosomal proteins in the 80S model (reviewed in Wilson and Doudna Cate 2012), before the first complete *S. cerevisiae* 80S crystal structure was delivered in 2011 (Ben-Shem et al. 2011).

Though much of the subsequent early work on eukaryotic ribosome structure was originally conducted in *S. cerevisiae* (reviewed in Baßler and Hurt 2019), the highly conserved nature of the 80S eukaryotic ribosome ensures that most of findings

directly translate to human ribosomes - although some differences in rRNA and ribosome protein composition are present. The human ribosome is, unsurprisingly, more complex than its yeast counterpart. The first robust evidence of this was provided by *Katter et al.* who developed a protocol for the purification and crystallisation of large quantities of homogeneous 80S ribosome populations from HeLa cells (Khatter, Myasnikov, Mastio, et al. 2014). The first near-atomic structure of the human 80S ribosome was published in Nature one year later, which used single-particle cryo-EM to generate a high-resolution structure (Khatter, Myasnikov, Natchiar, et al. 2015). The authors noted several differences between the human and *S. cerevisiae* ribosome, most significantly, that the human ribosome contains one extra ribosomal protein - RPL28 - which is not found in its *S. cerevisiae* counterpart.

The intricate composition of the ribosome enables the ribosome to carry out translation. All prokaryotic and eukaryotic ribosomes contain similar structural features that enable mRNA transcripts to be translated by the ribosome for nascent peptide generation. One of the most essential structural components of the 80S ribosome are the 40S and 60S subunits, which are responsible for mRNA codon decoding and catalysing peptide bond formation respectively (Gregory et al. 2019). Both subunits carry out their functional role simultaneously, assisted by a number of protein factors which interact with ribosomes, mRNA, or transfer RNA (tRNA) to regulate translation and allow the addition of amino acids to the peptide chain. In eukaryotes, the process of protein translation can largely be separated into three distinct steps; translation initiation, elongation of the peptide chain, and termination of translation. Understanding the stages of translation is helpful when discussing the mechanisms of ribosome specialisation (see Section 1.3.2) and axonal ribosome transport (see Section 1.5.2).

Translation initiation (reviewed by Merrick and Pavitt 2018) is the first step of translation, whereby the mRNA start codon is detected and the 80S ribosome is assembled. The process begins with the formation of the ternary complex - composed of initiator Met-tRNAi (tRNA carrying the first amino acid of the peptide chain),

protein initiation factor eIF2, and GTP (Levin et al. 1973). This complex recruits the 40S subunit and the GTPase-activating protein, eIF5, facilitating the formation of the stable 43S preinitiation complex (PIC). Formation of the PIC requires several other initiation factors (eIF1, eIF1A, and eIF3) which bind to the 40S subunit, thus forcing it to adopt an open confirmation to allow ternary complex binding (Llácer et al. 2015). The PIC then binds to the m^7G capped region of an mRNA, assisted by eIF4F which disrupts the folded structure of the mRNA. After scanning along the 5' untranslated region of the mRNA in a 3' direction, the PIC eventually identifies the mRNA start codon (typically AUG), which forms a stable interaction with the Met-tRNAi anticodon. Upon AUG recognition, the PIC undergoes a structural change into a closed conformation, induced by eIF2-mediated GTP hydrolysis and phosphate release. This causes dissociation of eIF1 (Cheung et al. 2007). eIF2 and eIF5 are subsequently exchanged and the initiation complex (formerly the PIC) then interacts with an additional initiation factor, GTPase eIF5B, to recruit the 60S subunit to the complex (Pestova et al. 2000). Once again, GTP hydrolysis by eIF5B results in the removal of both eIF5B and eIF1A, creating the functional 80S ribosome which is primed and ready for translation.

The second stage of translation - elongation - concerns the addition of amino acids to the nascent peptide chain. Elongation (reviewed in Blanchet and Ranjan 2022) takes place in the highly conserved ribosome active site, formed of three tRNA binding sites; A (aminoacyl), P (peptidyl) and E (exit). After initiation, the initiator Met-tRNAi is situated in the ribosome P-site, leaving the A-site empty. First, aminoacyl-tRNAs (delivered by the elongation factor, eEF1A) enter the A-site and recognise the complimentary mRNA codon. This triggers hydrolysis of GTP by eEF1A, which is then released from the ribosome. Subsequently, peptide bond formation occurs. Within a region of the large subunit known as the peptidyl transferase center, a reaction between the amino acids in the P-site and A-site is catalysed. Following this peptide bond formation, the ribosomal subunits rotate causing the tRNAs in the A and P sites to enter a hybrid state (Moazed and Noller

1989). In this state, their anticodon regions remain in their original A and P site, with their tRNA acceptor regions being shifted in to the P and E sites respectively. Binding of eEF2 then causes ribosome translocation in a GTP-dependent manner. The ribosome is moved along the mRNA and the tRNAs return to their original conformation. The E-site now contains the deacylated tRNA, the P site contains an aminoacyl-tRNA, and the A site is again empty - ready for a new aminoacyl-tRNA to enter. This cycle of aminoacyl-tRNA entering the A-site, peptide bond formation with the P-site amino acid, and tRNA exit via the E-site continues until the stop codon is reached.

Upon recognising the stop codon, translation is terminated. Several different codons (including UGA, UAA and UAG) cause translation termination, whereby the stop codon is recognised by release factors. eRF1, which recognises all three stop codons, catalyses the hydrolysis of the polypeptidyl-tRNA bond, releasing the newly synthesised protein from the ribosome. This process also requires eRF3, which hydrolyses GTP to assist eRF1 in translation termination. Finally, the ribosome undergoes "recycling". The 40S and 60S subunits dissociate from one another, and the mRNA and tRNA are released to enable further translation within the cell.

1.3.1 Ribosome Biogenesis is a Complex and Highly Conserved Pathway

It could be argued that axonal ribosomes are made *in situ*, however, eukaryotic ribosome biogenesis is a highly conserved and tightly regulated pathway that could not occur outside of the cell body (Figure 1.4). Generally considered to be the most energy demanding process within dividing cells (K. E. Bohnsack and M. T. Bohnsack 2019), the production of nascent ribosomes is one of the most intricate and well-studied biogenesis pathways. To this end, the assembly and maturation of eukaryotic ribosomes requires the impressive co-ordination of over 200 proteins, including ribosomal proteins and assembly factors, polymerases, and non-coding RNAs (Thomson et al. 2013).

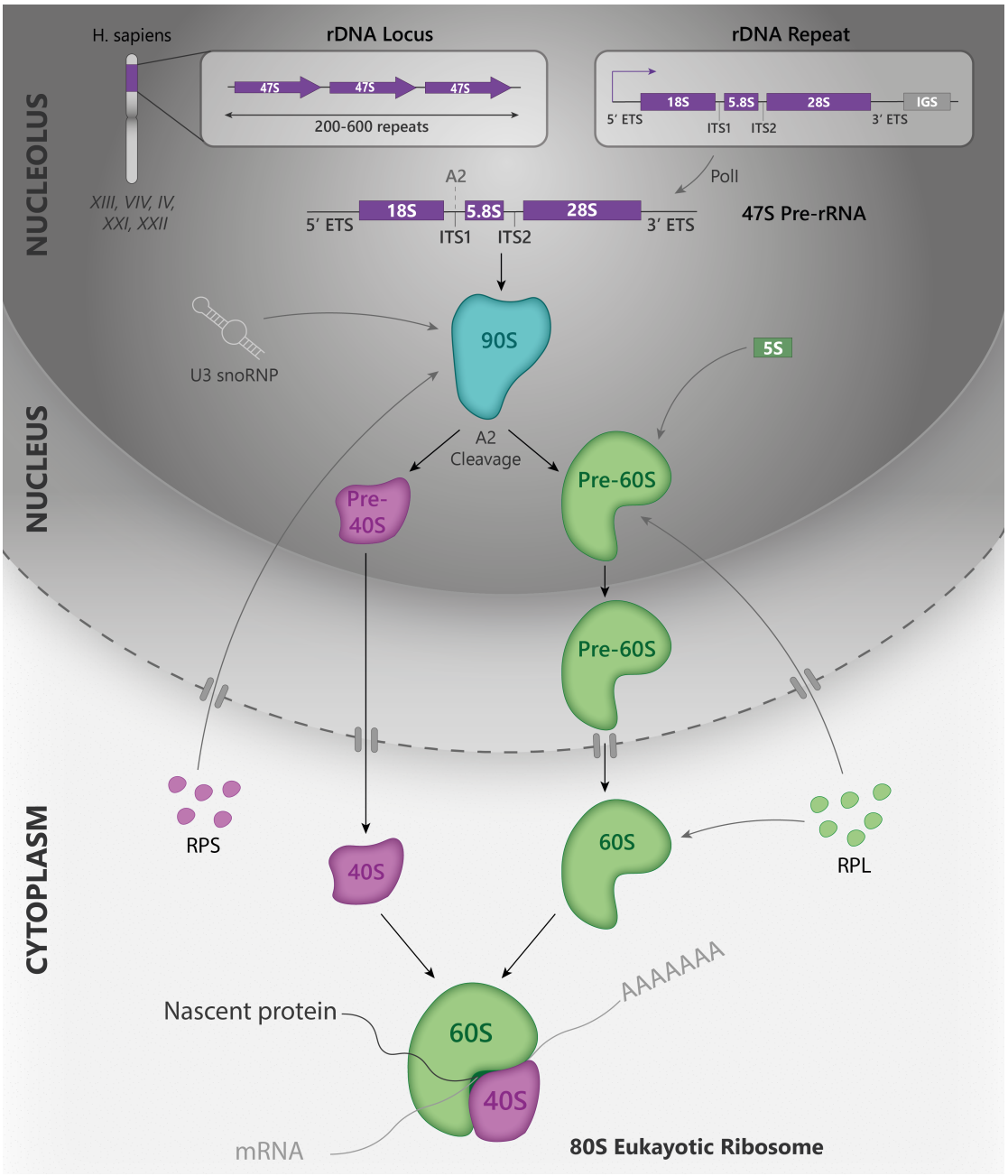


Figure 1.4: Overview of Ribosome Biogenesis in Higher Eukaryotes.

Eukaryotic ribosome biogenesis occurs predominantly in the nucleolus and involves over 200 assembly factors (not shown). Biogenesis begins with the transcription of rDNA genes, which are formed of hundreds of repeats of an rDNA locus. These repeats are translated by PolI to create the 47S pre-rRNA. Subsequent processing steps, along with the interactions of snoRNPs and ribosomal proteins, generates the 90S precursor ribosome (blue) which is further processed into pre-40S (magenta) and pre-60S (green) particles after pre-rRNA cleavage at the A_2 site. The precursor subunits undergo more processing before being exported into the cytoplasm where the final maturation steps take place.

The process of ribosome biogenesis begins in the nucleolus - a large, membraneless structure within the nucleus that contains all the necessary RNA and ribosomal protein genes required for ribosome synthesis. Present in almost all eukaryotic cells, the nucleolus is the site of ribosome RNA transcription and assembly, two processes which occur simultaneously. Ribosome biogenesis begins with the transcription of the ribosomal DNA (rDNA) to produce the rRNA precursor transcript (pre-rRNA). The rDNA genes are present as large clusters of tandem repeats within the nucleolus, located on chromosome regions known as nuclear organiser regions (NOR). The copy number of rDNA repeats varies widely between organisms, ranging from 1-15 copies in bacterial species to several thousand repeats in plants (Espejo and Plaza 2018). In human genomes around 200 rDNA copies are present, located on 5 different chromosomes (13, 14, 15, 21 and 22) (Henderson et al. 1972). Although hundreds of rDNA clusters are present within the human genome, not all of them are transcriptionally active (Grummt and Pikaard 2003).

In order to produce the rRNAs required for ribosome assembly, rDNA is transcribed by RNA Polymerase I (PolI) to create the polycistronic pre-rRNA transcript within the nucleolus - producing the 35S transcript in yeast and the 47S transcript in mammalian cells. Early work to define the mechanism of pre-rRNA transcription was carried out in *S. cerevisiae* (reviewed in Thomson et al. 2013). To begin eukaryotic rDNA transcription, a PolI initiation complex is formed from several protein factors - the core factor, upstream activating factor, and the TATA-box binding protein. After the initiation complex binds to an rDNA promoter, PolI is then recruited with the assistance of an initiation factor, Rrn3. Subsequently, transcription begins and the nascent pre-rRNA is formed. Interestingly, the transcription and modification of the pre-rRNA are coupled processes, with the 35S transcript being actively modified while it is being synthesised and folded. Modification of the pre-rRNA is essential, enabling the processing and cleavage of the transcript to generate the 5.8S, 18S and 25S (or 28S in mammals) rRNAs, which are the structural backbone of the mature ribosome. The last essential rRNA species, 5S, is transcribed by RNA

Polymerase III independently of the pre-rRNA transcript.

RNA modification is carried out by small nucleolar RNA (snoRNA) molecules, a class of short (60-300 nucleotides), non-coding RNAs which play an essential role in rRNA processing and ribosome biogenesis (reviewed in Ojha et al. 2020). Importantly, the activity of snoRNAs is thought to contribute to ribosome specialisation (see Section 1.3.2), providing a possible mechanism for neuronal ribosome heterogeneity and axonal ribosome transport (see Chapter 6). Of all the non-coding RNA species, snoRNAs are some of the most diverse and highly abundant RNA populations, thought to be expressed within all eukaryotic organisms (Kufel and Grzechnik 2019). snoRNAs associate with proteins to create small nucleolar ribonucleoprotein (snoRNP) complexes, where the snoRNA is responsible for directing the snoRNP to the target site and the protein component catalyses the modification reaction - typically either 2'-O-methylation or pseudouridylation. There are two main classes of snoRNA, classified by the presence of a highly conserved motif which enables RNA base recognition and chemical modification (Figure 1.5).

The most abundant class of snoRNA is the C/D class, which are characterised by the presence of C Box (RUGAUGA, where R is any purine nucleotide) and D Box (CUGA) motifs, located on the 5' and 3' ends of the snoRNA respectively. Interactions upstream of the C and D Box termini allows the C/D snoRNA to fold into a kink-turn structure (Watkins et al. 2000), with some C/D snoRNAs containing secondary C' and D' Box motifs located towards the middle of the loop. Recognition of the rRNA target requires a specific 7 - 21 nucleotide sequence upstream of the D Box region (Bachellerie et al. 1995), which basepairs with the rRNA modification site to direct the snoRNP protein to carry out the methylation reaction. There are several conserved enzymes which have been demonstrated to associate with C/D snoRNAs, including Fibrillarin, NOP56, NOP58, and SNU13. These proteins, in association with C/D box snoRNAs, form snoRNPs which are responsible for the 2'-O-methylation of the target rRNA.

Similarly, the H/ACA class (reviewed in Czekay and Kothe 2021) is defined as

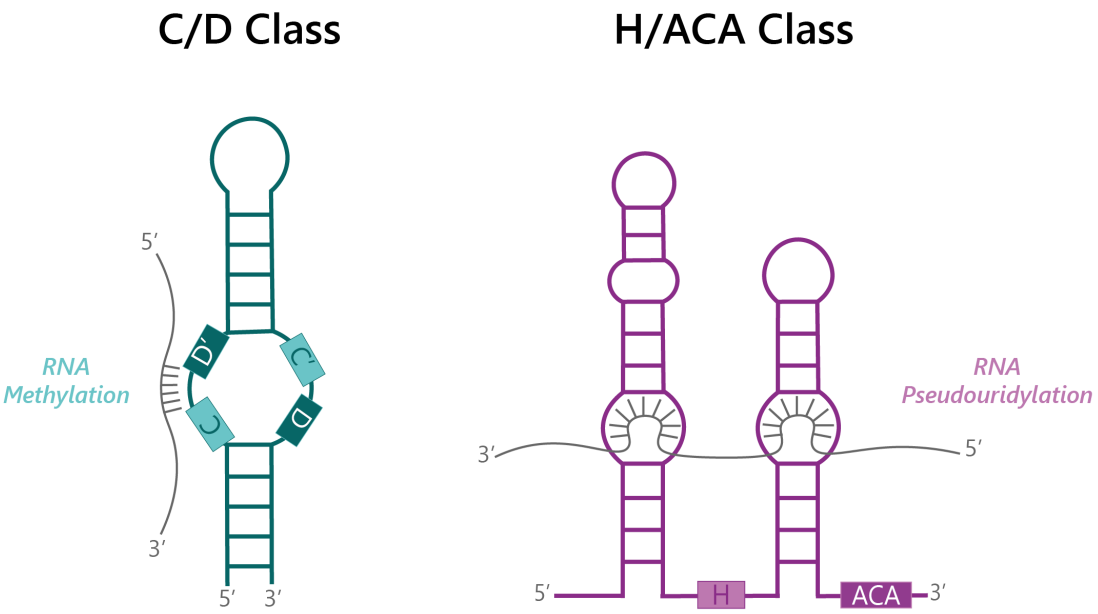


Figure 1.5: Simplified Structures of C/D and H/ACA snoRNAs.

Basic structures of the C/D and H/ACA class snoRNAs and their conserved motifs. C/D snoRNAs (green) form a kink-turn structure containing C and D box motifs. Secondary C' and D' boxes are often present within the loop. Target mRNA (grey) is recognised by a sequence upstream of the D box motifs, which then undergoes 2'-O-methylation by enzymes associated with the C/D snoRNA. H/ACA snoRNAs (purple) adopt a hairpin-hinge-hairpin-tail structure, with the H and ACA boxes located in the hinge and tail regions respectively. mRNAs (grey) are recognised by recognition loops in the hairpin strictures, which guide pseudouridylation by H/ACA snoRNA-associated enzymes. This figure was adopted from Liang et al. 2019.

snoRNAs containing the H (ANANNA, where N is any nucleotide) and ACA (ACA) box conserved motifs. H/ACA snoRNAs form a distinct secondary hairpin-hinge-hairpin-tail structure, where the H and the ACA box sequences are located in the hinge and tail regions respectively. The RNA recognition sites are located in the hairpin regions (Bortolin et al. 1999), where sequences known as recognition loops form pseudo-knots with the target site to enable chemical base modification. This reaction is catalysed by several proteins - NOP10, GAR1, NHP2, and dyskerin - which all interact with the H/ACA snoRNA to modify uridine residues to pseudouridine in a process known as pseudouridylation.

In addition to the C/D and H/ACA classes, a third class of snoRNA has also been discovered. Unlike the aforementioned snoRNAs which are nucleolus-specific, this type of snoRNA is localised to a sub-nuclear region known as the Cajal-body. Appropriately named small Cajal-body specific RNA (scaRNA), this subset of snoRNAs have similar conserved motifs to C/D and H/ACA snoRNAs, although composite scaRNAs containing features from both classes have also been identified (Jády and Kiss 2001). The modification mediated by scaRNAs is dependent on the type of motifs present (either C/D box to guide methylation, H/ACA to direct pseudouridylation, or occasionally both modifications when composite in structure). In contrast to the impressive diversity of C/D and H/ACA snoRNAs, scaRNAs number significantly less - only approximately 30 scaRNA genes are known to be expressed within human cells (Jorjani et al. 2016).

rRNA modification is not the only process which occurs while the nascent pre-rRNA is being produced. In eukaryotic cells, a multitude of ribosomal proteins join the newly synthesised pre-rRNA transcript in a processes known as transcriptional assembly, forming stable ribonucleoprotein precursor particles (occasionally referred to as the "90S preribosome"). Ribosomal proteins are translated in the cytoplasm and subsequently imported into the nucleolus, travelling through nuclear pores to bind to the nascent rRNA before transcription has terminated. Around 70 assembly factors and several snoRNAs also associate with the pre-rRNA in a hierarchical

manner to create the 90S preribosome (Kornprobst et al. 2016), playing a vital role in pre-rRNA processing and release of the mature rRNA species.

The existence of snoRNAs and their ability to direct RNA modification was established long ago (Maxwell and Fournier 1995), however, several snoRNAs have been shown to play a more direct role in rRNA processing. After rDNA transcription, the pre-rRNA transcript undergoes cleavage of external transcribed spacers at the termini (3' and 5' ETS) and internal transcribed spacers which separate the mature rRNA transcripts (ITS1 and ITS2). In addition to directing rRNA modification, a small number of essential snoRNAs are responsible for pre-rRNA cleavage and processing. One notable example is the C/D class snoRNA, U3. By depleting the levels of several snoRNAs (U3, U8 and U13) using immunoprecipitation *in vitro*, *Kaximierz et al.* illustrated that U3 snoRNP was essential for the first cleavage stage of pre-RNA processing (Kaximierz et al. 1990). Further research into 90S preribosome formation has illustrated that several snoRNPs (including U3) bind to the pre-rRNA and enable RNA cleavage at several well-defined pre-rRNA sites - A_0, A_1, and A_2. In particular, the U3 snoRNP is thought to assist with endonucleolytic cleavage at the A_1 and A_2 sites by binding to specific regions of the 35S transcript and keeping them in a premature conformation, thus playing a predominantly structural role in pre-rRNA processing (Baßler and Hurt 2019).

The 90S preribosome intermediate is subsequently disassembled upon rRNA cleavage at the A_2 site, stimulating two distinct pathways which result in the formation of the pre-40S and pre-60S precursor subunits (Thomson et al. 2013). The pre-40S subunit, already containing most of the necessary small ribosome proteins and biogenesis factors due to their incorporation into the 90S preribosome, is quickly exported into the cytoplasm where the final maturation steps take place (Thomson et al. 2013). In contrast, maturation of the pre-60S subunit is far more intricate, requiring many additional processing steps and biogenesis factors within the nucleus before cytoplasmic export (reviewed in Moraleva et al. 2022). Notably, the 5S rRNA (which is transcribed separately to the pre-rRNA) is imported into the nucleolus as

part of a ribonucleoprotein complex (Lau et al. 2023). This complex is believed to be incorporated into the pre-60S subunit at an early stage, although the exact mechanism is yet to be elucidated. After nuclear export, the precursor subunits are subject to the final maturation stages - including further rRNA processing, association of additional assembly factors, and late ribosomal protein incorporation (Baßler and Hurt 2019).

Although ribosome biogenesis is highly conserved across species, the pre-rRNA processing and maturation pathways differ slightly between *S. cerevisiae* and higher eukaryotes (reviewed in K. E. Bohnsack and M. T. Bohnsack 2019). In human cells, these differences are thought to be due to the increased ribosome size in conjunction with the tighter regulatory control required to maintain their complex biogenesis pathways (Moraleva et al. 2022). Higher snoRNA diversity is also observed in mammalian cells (Kufel and Grzechnik 2019), with studies identifying that vertebrate rRNA contains approximately double the base modifications than yeast rRNA (Piekna-Przybylska et al. 2008). Although some aspects of ribosome biogenesis are more complex in human ribosome biogenesis, many stages are functionally identical with genetic screens identifying many conserved assembly factors identified in both yeast and humans (Dörner et al. 2023).

Owing to the complexity of the process, ribosome biogenesis is strictly controlled at every stage. Regulation of this process begins with the architecture of the genome, rDNA transcription and pre-rRNA processing, and finally, the quality control checks that occur before 40S and 60S pre-subunit nuclear export. Many of these processes are regulated by several cell signalling pathways which are closely linked to cell growth (K. E. Bohnsack and M. T. Bohnsack 2019). Specifically, three mechanisms - the mTORC1 pathway, regulation of Myc, and association of noncoding RNA - have been identified as key regulators of ribosome biogenesis (reviewed in Jiao et al. 2023). Dysregulation of these pathways, and multiple other aspects of the ribosome biogenesis machinery, results in defective ribosome assembly and the development of a range of diseases - known as “ribosomopathies” (Kang et al. 2021).

To summarise, ribosome biogenesis is an incredibly intricate pathway, requiring the impressive co-ordination of hundreds of proteins - including assembly factors, polymerases, snoRNAs, and ribosomal proteins. This process occurs mainly in the nucleolus, a distinct sub-compartment of the nucleus containing the essential rDNA genes and assembly factors for rRNA synthesis and its subsequent processing. In neurons, the production of ribosomes is known to be influenced by a range of processes, including the stage of brain development (Chau et al. 2018), 5-HT induced neuronal stimulation (Hernández et al. 2009), and synaptic activation (D. Li et al. 2018). Given that ribosome biogenesis predominantly occurs in the nucleolar compartment, it is unlikely that axonal ribosomes can be produced *in situ*. Like all other eukaryotic ribosomes, they must be synthesised in the cell body and transported to the distal axon - although the mechanism of this transport is a topic of debate.

1.3.2 Specialised Ribosomes

For many decades following their discovery (Palade 1955), ribosomes were considered to be identical molecular machines with no variation in their protein or RNA components. This notion has only recently begun to be challenged, with evidence of ribosome heterogeneity being discovered across many different species (reviewed in (Gay et al. 2022)). It has been proposed that ribosome heterogeneity is an evolutionary necessity, with ribosome variation increasing in higher eukaryotes due to selective pressures and a greater need for spatio-temporal control of translation in complex cells (Gay et al. 2022; Dinman 2016). To this end, ribosomes with high heterogeneity have been shown to have different translational capabilities, with ribosomes appearing to be tailored to the protein requirements of their cellular environments. These ribosomes (otherwise known as "specialised ribosomes") can utilise a variety of different mechanisms which vary their composition to enable their specialised functions. Such variations include changes to ribosomal protein expression, rRNA modifications, and interactions with ribosome-associated factors (Joo et al. 2022).

The heterogeneous nature of ribosomes was proposed soon after their discovery.

One of the earliest theories supporting ribosome variation was delivered by Francis Crick, who postulated that each gene produces RNA which acts as the template for a unique ribosome, which in turn produces one kind of protein (Crick 1958). Termed the "One gene - one ribosome - one protein" hypothesis, this theory provided an explanation as to how genetic information could transfer from gene to ribosome to create a diverse population of proteins. Crick's hypothesis was not universally accepted, with others arguing that ribosomes are more likely to be homogenous structures which translate RNA intermediates (or "messages") without discrimination (Jacob and Monod 1961). Subsequently, the "One gene" theory was quickly discredited upon the discovery of mRNA in 1961 (Brenner et al. 1961). By analysing nascent RNA synthesis in *E. coli* following phage infection, *Brenner et al.* elegantly determined that the transfer of genetic information occurs through an unstable RNA intermediate, rather than ribosomal RNA as previously assumed. Following this landmark discovery, the concept of ribosome heterogeneity was largely abandoned for several decades (Genuth and Barna 2018).

Despite evidence of ribosome structural changes in the parasite *Plasmodium berghei* (Gunderson et al. 1987) and the slime mold *Dictyostelium discoideum* (Ramagopal 1990) at different stages of their life cycle, ribosome heterogeneity remained controversial for many years. The next major advancement in ribosome composition came in 2002, when *Mauro et al.* published "The Ribosome Filter Hypothesis" - a revolutionary theory proposing that the composition of the ribosomal subunits acts as a selectivity filter, enabling interactions between specific subsets of mRNA and the translational machinery (Mauro and Edelman 2002). Following this, in conjunction with additional evidence of ribosome structural variation, the concept of ribosome heterogeneity has become more widely accepted in recent years. In particular, advancements in proteomic techniques and sequencing has enabled the components of ribosomes to be studied in detail (Genuth and Barna 2018), highlighting the unique composition some ribosomes have adopted which enable their specialised translational roles.

Variation in ribosomal protein incorporation into the mature ribosome is perhaps one of the most obvious ways that ribosome heterogeneity can be achieved. As previously discussed, approximately 80 canonical ribosomal proteins make up the 80S eukaryotic ribosome. In addition to these core proteins, many species also express ribosomal protein paralogues, ribosomal protein variants with large regions of homology to the core protein allowing for ribosomal incorporation (Genuth and Barna 2018). The number of ribosomal protein paralogues varies widely between species, ranging from 59 paralogue pairs in yeast (Segev and Gerst 2018) to just 19 pairs in humans (Guimaraes and Zavolan 2016). One interesting example of paralogue incorporation occurs in *Drosophila melanogaster*, where testes-specific expression of an RPL22 paralogue (RPL22-like) is observed (Kearse et al. 2011). The authors note that the low sequence homology of RPL22 and RPL22-like (∼37%), coupled with RPL22-Like being present in polysomes, hints at this paralogue having a tissue-specific ribosomal function in *Drosophila*. Further work utilising RNA-Seq showed that ribosomes containing RPL22-like translate specific subsets of mRNA expressed at specific time points in spermatogenesis - suggestive of ribosome specialisation (Mageeney and Ware 2019). Additional studies indicating ribosomal protein paralogue presence in sub-populations of ribosomes have also been published for many other species, including yeast (Samir et al. 2018), mice (Sugihara et al. 2010) and humans (Kearse et al. 2011).

In addition to the incorporation of ribosome protein paralogues, the expression of the 80 canonical ribosomal proteins can also vary. This is predominantly seen as changes in their stoichiometry (i.e. the absence or addition of core ribosomal proteins) which changes the 1:1 ratio typically observed in mature ribosomes (Gay et al. 2022). Mass spectrometry has proved to be an invaluable tool for ribosome protein composition analysis in a variety of different species and cell lines. Using quantitative mass spectrometry on mouse embryonic stem cells, *Shi et al.* demonstrated that some translationally active polysomes had a 30-40% reduction in four core ribosomal proteins (Z. Shi et al. 2017). Conversely, the addition of extra copies

of a ribosomal protein has also been documented, with yeast and mouse embryonic stem cells appearing to have variable ribosome protein compositions depending on growth conditions (Slavov et al. 2015). Overall, this data provides evidence that cells can regulate their translational capabilities according to their cellular environment. Interestingly, the post-translational modification of ribosomal proteins has also been suggested to contribute to ribosomal heterogeneity (reviewed in Joo et al. 2022; Guo 2018). Although there is some evidence of this in yeast, plant and *E. coli* ribosomes, it is still unclear how these modifications may play a role in translational control (Gay et al. 2022).

The other major source of ribosome heterogeneity arises from variations in the rRNA sequence. Despite its highly conserved biogenesis pathway, rRNA allele heterogeneity has been well documented since the 1980s (Joo et al. 2022), although in-depth analysis of rRNA sequence has only recently been possible due to advances in RNA sequencing technology and bioinformatics analysis (Genuth and Barna 2018). To this end, a study analysing whole genome sequencing data from the 1000 Genomes Project and the Mouse Genomes Project provided convincing evidence that rRNA variants are present in the genomes of humans and mice respectively (Parks et al. 2018). Interestingly, these variants were found to be located in the functional centres of translating ribosomes, which the authors postulate may enable rRNA variants to influence the binding of ribosomal proteins, non-ribosomal factors, or influence ribosome conformational changes that occur in protein synthesis. Although the existence of rRNA heterogeneity is well established, the majority of our understanding of this heterogeneity and its effect on ribosomal function originates from studies in prokaryotic systems (Gay et al. 2022); the exact functional role of variant rRNA expression in eukaryotes is still yet to be determined.

In contrast, the translational effect of ribosome heterogeneity originating from rRNA modifications is more fully defined. As previously discussed, rRNA is post-transcriptionally modified by snoRNPs, with 14 distinct modifications being identified within human cells (Taoka et al. 2018). The most abundant modifications are

2'-O-methylation and pseudouridylation, which is thought to account for 90% of all rRNA modifications (Taoka et al. 2018) targeting approximately 200 rRNA modification sites identified in human cells (Sloan et al. 2017). The post-transcriptional modification of rRNA is carried out by snoRNAs (see Section 1.3), which play an essential role rRNA processing during biogenesis. However, in addition to modification at conserved sites, recent RNA sequencing studies have identified additional unique modifications within human ribosomes (Natchiar et al. 2017). Concentrated around the ribosomal functional sites, these modifications were found predominantly on the 18S and 28S, with the authors suggesting that rRNA modification may be a source of heterogeneity between cell types. In yeast, knockdown of C/D box snoRNA expression resulted in the identification of variable 2'-O-methylation sites, 70% of which are capable of being hypo-2'-methylated (Khoshnevis et al. 2022). Strikingly, hypomethylation at these sites was closely related to changes in translational output, caused by reduced binding efficiency of eIF1 to the 40S subunit. Generally, rRNA modifications are thought to stabilise RNA bond stability and modify the translational capabilities of ribosomes (Miller et al. 2023), although the functional roles of many modifications are still unknown (Gay et al. 2022).

As our understanding of ribosome specialisation grows, so too does the evidence of possible ribosome heterogeneity in neuronal cells (reviewed in Dastidar and Nair 2022, see Chapter 6). Given the need for rapid alterations of the local proteome in neurons, it stands to reason that neuronal cells would benefit from the translational control offered by specialised ribosomes. Recent advancements in mRNA sequencing techniques in neuronal processes provided the first hint of heterogeneous ribosome composition. Multiple studies reported that an unexpected abundance of ribosomal protein mRNAs were present in neuronal processes, including dendrites (Cajigas et al. 2012), axons (Briese et al. 2016; Cajigas et al. 2012; Shigeoka, H. Jung, et al. 2016), and pre-synaptic compartments (Hafner et al. 2019). Given that mRNAs must be actively transported into the distal processes from the cell body, this data provided the first indication that the neuronal translational machinery may utilise

local translation as a way to support the synthesis of essential proteins in neurites (Dastidar and Nair 2022). Moreover, utilising their novel Axon-TRAP-RiboTag transgenic mouse line, *Shigeoka et al.* determined that ribosomal protein transcripts are actively translated in the axons of retinal ganglion cells (Shigeoka, H. Jung, et al. 2016). Therefore, it was clear that the local synthesis of ribosomal proteins is a consistent, although unexplained, feature of neuronal processes.

Building on this work, studies have shown that neurite-localised ribosomes undergo local remodelling using ribosomal proteins translated in the distal processes. In a landmark study, retinal ganglion axons isolated from *Xenopus* eyes were analysed to investigate the composition and dynamics of free ribosomal proteins within axons (Shigeoka, Koppers, et al. 2019). By adopting a range of proteomic and imaging techniques, researchers demonstrated that newly synthesised ribosomal proteins dynamically join axonal ribosomes, maintaining the ribosome population and supporting local synthesis. The authors state that the incorporation of nascent ribosomal proteins may play a role in the modification or maintenance of axonal ribosomes, the latter of which may functionally repair ribosomes and extend their half lives (Farias et al. 2020). Through structural analysis of axon-enriched ribosomal protein transcripts, it was also determined that surface-exposed ribosomal proteins had a higher likelihood of being incorporated into axonal ribosomes. Similarly, dendritic ribosomes have also been shown to to undergo local remodelling with newly synthesised ribosomal proteins (Fusco et al. 2021).

Both the aforementioned studies provide strong evidence in favour of neuronal ribosome heterogeneity. The rapid exchange of ribosomal proteins observed could allow for neurite-localised ribosomes to dynamically modify their composition and translational output, possibly in response to cellular cues or neuronal stimulation. This could include paralogue incorporation, stoichiometry changes, or the addition of modified ribosomal proteins into existing neurite-localised ribosomes. Although a mechanism of neuronal heterogeneity and specialisation has been uncovered, direct evidence of ribosomes localised to neuronal compartments with variant composition

is still lacking (Dastidar and Nair 2022).

1.4 Investigating Ribosome Localisation in Neurons

The presence of ribosomes in dendrites is well established. Early microscopic analysis of primate motor neurons provided the first glimpse of dendritic ribosomes (Bodian 1965). Active translation of mRNA is usually characterised by the presence of polysomes - multiple ribosomes attached to a single transcript. Evidence of this came several years later, with EM studies confirming the presence of polysomes in dendritic spines in granule cells (Steward and Levy 1982). This has since been reproduced many times, with polysomes being identified throughout the dendrite and spines (Holt et al. 2019).

Studies identifying ribosomes in adult axons are more controversial. As previously discussed, early EM studies could not identify ribosomes within the axoplasm of giant squid neurons (Lasek, Dabrowski, et al. 1973). Difficulties in identifying ribosomes in adult axons via EM may arise due to their low abundance, with polysome formation being rare (Holt et al. 2019). Commonly, the issue of low endogenous target expression is overcome by over-expressing the fluorescently-tagged protein of interest. However, studies in *S. cerevisiae* have shown that over-expressed ribosomal proteins are not incorporated into mature ribosomes and are quickly degraded by the ubiquitin–proteasome system (Sung et al. 2016). This unique problem has only recently been addressed with the use of genetic tagging techniques, which have allowed specific ribosomal proteins (and ribosomes) to be visualised *in vivo*. Genetic manipulation, in conjunction with improved imaging methodologies, have proved to be essential in the search for ribosomes in adult axons.

As a result of these technical challenges, axonal ribosomes have only recently been observed in mature neurons. *Hafner et al.* conducted FISH experiments on rat brain slices and neurons, and reported that over 75% of pre-synaptic terminals

have translation machinery present; including ribosomes and mRNA transcripts (Hafner et al. 2019). Additionally, an explanation for the scarcity of polysomes in axons was provided by *Biever et al.* who analysed synaptic tissue using polysome profiling and ribosome footprinting, showing that both axons and dendrites have increased monosome levels (i.e. a single ribosome attached to a transcript) (Biever, Glock, et al. 2020). The elongated form of these monosomes indicate that they are translationally active, with a proportion of high abundance synaptic protein transcripts showing monosome preference. This research provides convincing evidence in support of the presence of monosomes in axon terminals, which could explain how local translation occurs, despite EM studies pointing to minimal polysome activity. Additionally, monosome translation may contribute to the diversity in the pre-synaptic mRNA transcript population and neuronal proteome (Biever, Donlin-Asp, et al. 2019). Despite the essential role of axonal ribosomes, the mechanism for their localisation is still yet to be elucidated.

1.5 Axonal Ribosome Localisation Hypotheses

1.5.1 The Glial Origin Hypothesis

In order for localized translation in axons to occur, ribosomes must likewise be localized to carry out this protein synthesis. However, the origin of axonal ribosomes is a topic of debate. One prevailing model is that ribosomes are delivered to axon terminals from a non-neuronal source, more specifically, glial cells are responsible for laterally transferring ribosomes to neurons (Koley et al. 2019). This was first proposed in 1977, with analysis of giant squid axons providing support of the glial-neuronal protein transfer hypothesis (Gainer et al. 1977; Lasek, Gainer, et al. 1977). The latter paper provided a simple experiment whereby the giant squid axon was separated from the cell body, and radioactive amino acids ([3H]leucine) were seen to be incorporated into glial cells and axoplasm (Lasek, Gainer, et al. 1977). Suggestions that axons may be synthesising the labelled protein were dismissed by the

authors, stating that ribosomal RNA and ribosomes were not present in axons at significant levels. However, the presence of polysomes in pre-synaptic compartments, discovered years later, indicated that some labelled protein could be derived from axonal translation (Twiss and Fainzilber 2009).

With research into localised translation quickly advancing, the next major update in the glial origin hypothesis came with the demonstration of a possible polyribosome transfer from Schwann cells in the Wallerian degeneration slow mouse (Wld5) (Court, Hendriks, et al. 2008). Using an enhanced green fluorescent protein (EGFP) tag on the RPL4 ribosomal protein, desomatised (i.e. cell body removed) axons and associated cells were transfected following sciatic nerve crush injury of Wld5 mice. Upon analysing the fluorescent signal, 79% of surrounding Schwann cells showed RPL4-EGFP signal which co-localised strongly with a ribosomal marker. 60% of adjacent axons also showed diffuse RPL4-EGFP signal, which the authors conclude shows polysome transfer from the schwann cells into the axoplasm of injured nerves. Further microscopic analysis indicated that ribosomes are transferred in double-membrane vesicles, with the outer and inner membranes originating from the axonal and Schwann cell membranes respectively (Court, Hendriks, et al. 2008). Morphological evidence of this process *in vivo* was provided several years later (Court, Midha, et al. 2011). An almost identical experimental method was followed, however this time ribosome transfer was seen in regenerating axons of WT mice.

The proposal that axonal ribosomes have a glial origin presents several major problems, which cannot be resolved by this model alone. *Court et al.* present a convincing argument for polyribosome transport from Schwann cells, but only in the case of neurons following injury and axonal regeneration. Similarly, a study of Multiple Sclerosis patients and mouse models also reports a glial contribution to axonal ribosome levels (Shakhbazau et al. 2016), although the mechanism for glial vesicle formation and fusion with the axon is still unknown (Sotelo et al. 2014). However, this data does not exclude the possibility of a soma-originating population of ribosomes in intact, healthy neurons. Furthermore, some neurons have unmyelinated

axons, so this model cannot explain how local axonal translation occurs in these neuron types. Additionally, the glial-origin model of polyribosome transfer cannot account for the significant monosome presence in the pre-synaptic compartment (Biever, Glock, et al. 2020). The abundance of research illustrating the requirement of axonal translation in a multitude of processes (in addition to injury response) also points to the cell body as being a major source of axonal ribosomes, though evidence of this is still scarce.

1.5.2 The Hitchhiker Hypothesis

In recent years, experimental evidence has allowed a new theory for long-distance ribosome transport to emerge. As research into mRNA localisation in neurons progressed, our understanding of their transport mechanisms also deepened - with mRNA transcripts relying upon the formation of RNA granules for their directed movement to distal processes (reviewed in Fernandopulle et al. 2021; De Graeve and Besse 2018). It has long been established that RNA-binding proteins closely associate with mRNA transcripts, which in turn assemble into larger, membrane-less RNA granule complexes (Knowles et al. 1996). Biochemical purification of these granules have shown that various components of the transport and translational machinery are incorporated - including mRNAs, cytoskeletal elements, ribosomal binding proteins, motor proteins, and ribosome components (Elvira et al. 2006). Strikingly, these granules are heterogeneous in nature, with different RNA granule compositions being observed in dendritic and axonal sub-compartments (Christie et al. 2009).

The movement of RNA granules within neurons has been extensively studied (reviewed in De Graeve and Besse 2018). The transport of RNA granules is variable within neurons, with dendritic granules reported to regularly cycle between long stationary and short motile phases (Yoon et al. 2016). mRNA imaging experiments determined that RNA granules, when motile, undergo fast, bi-directional transport which is reliant on motor protein interaction (Kanai et al. 2004; Chu et al. 2019;

Fukuda et al. 2021). Accordingly, the velocity of granule transport (0.2 – 1.5 µm/s) matches the speeds reported for fast axonal transport (Kiebler and G. Bassell 2006). RNA granule movement by motor proteins necessitates the interaction of adaptor proteins (such as KLC and BICD2), which act as molecular bridges between the motor and the RNA binding protein component of the granule.

However, in addition to direct transport of RNA granules by motor proteins, there is now evidence that granules can associate with moving organelles in order to be localised to distal processes. This process, referred to as "hitchhiking", is observed with several membrane-bound organelles in a range of species (reviewed in Salogiannis and Reck-Peterson 2017). Early examples of mRNA hitchhiking were discovered in the *Ustilago maydis* fungus (Baumann, Pohlmann, et al. 2012), where mRNA granules were shown to be co-transported with early endosomes, requiring the activity of both kinesin-3 and dynein. Studies into mammalian organelle-associated mRNA motility are only just beginning to catch up. In a landmark study investigating RNA granule dynamics in primary cortical neuron cultures, *Liao et al.* demonstrated that RNA granules associate with motile lysosomes for long-distance transport in axons (Liao et al. 2019). This interaction is mediated by a phosphoinositide-binding protein, ANXA11, which tethers the RNA granule cargo to the moving lysosome. Interestingly, the authors also note that some ANXA11 mutations are closely linked to the development of ALS, suggesting that mRNA hitchhiking transport may be a possible mechanism that is aberrant in ALS patient cells.

The role of ribosomes within neuronal RNA granules is still being defined. Purification of FMRP-containing granules from mouse brain tissues has identified polyribosomes as a core component (El Fatimy et al. 2016), which are densely packed into a morula-like structure. Therefore, *El Fatimy et al.* postulate that granules may originate from a somatic population of stalled ribosomes, which also explains the translational silencing observed with actively transported mRNAs (Krichevsky and Kosik 2001). Further structural analysis has corroborated these findings, with neu-

ronal RNA granules isolated from Wistar rat brains also containing 80S ribosomes stalled in a pre-translocation phase (Kipper et al. 2022). As previously discussed (see Section 1.3), ribosomes undergo a conformational change during translation elongation, whereby tRNAs within the ribosomal active site enter a hybrid A/P and P/E state. Using cryo-EM, *Kipper et al.* determined that the majority of neuronal RNA granule ribosomes are present in this conformation, possibly lowering the energy barrier required for translation initiation upon granule delivery to the target site. Additionally, this study also identified lysosome vesicle remnants within the purified granule samples, further substantiating the claims that neuronal RNA granules rely on lysosome hitchhiking for long-distance transport. Although there is evidence of ribosomes within purified neuronal RNA granules, there is currently no studies of ribosome transport from the cell body into distal neuronal processes.

1.6 Thesis Aims

Given the literature discussed in this Introduction, I hypothesise that the majority of axonal ribosomes originate from the cell body. Therefore, this suggests that axonal ribosomes are synthesised in the cell body and are subsequently localised into axons by the existing axonal transport machinery. Such transport of axonal ribosomes is likely to require some level of structural heterogeneity, which would allow their selection and transport through the AIS and into the axon. Overall, this research project will provide direct microscopic evidence in support of the neuronal origin of axonal ribosomes, which is currently an unexplored field. The research aims of this PhD are as follows:

1. Create CRISPR plasmids which deliver a knock-in tag into the coding sequence of ribosomal proteins.

2. Produce CRISPR-edited PC-12 cell lines to validate that the created CRISPR plasmids generate the desired knock-in effect.

3. Use the validated CRISPR plasmids in primary neuronal cultures and PC-12 cells to visualise ribosome subunit transport in axons.

4. Analyse the gene expression of ribosome-associated genes to identify genes which may contribute to neuronal ribosome heterogeneity.

Chapter 2

Materials and Methods

2.1 Tissue Culture and Transfection

2.1.1 Cell Culture of PC-12 Line

The PC-12 cell line was kindly donated to the Twelvetrees lab by Dr Elizabeth Seward (School of Biosciences, University of Sheffield). The cells were grown in suspension and maintained at a 1:2 to 1:4 subcultivation ratio. Once cell numbers reached approximately 6-8x10^6 cells in a T25 flask, cells were transferred to a 15 mL tube and spun at 400 relative centrifugal force (RCF) for 5 minutes. The old media was removed without disturbing the pellet, leaving 25% of the conditioned media. Fresh PC-12 media (Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Horse Serum, 5% foetal bovine serum (FBS), 1% GlutaMAX Supplement (Thermo Fisher Scientific, Cat: 11574466) and 100U penicillin-streptomycin (P-S)) was added to the tube and the cells were resuspended with a 5 mL stripette to ensure a single-cell suspension. The required volume of cells was added to new T25 flasks and additional fresh media was added to make up 5 mL total volume. Flasks were then incubated at 37°C under 5% CO_2 levels.

2.1.2 Electroporation of PC-12 Lines and Primary Neurons

PC-12 cells were transfected using the Neon Electroporation System (Invitrogen, Cat: MPK5000) following the Neon 100 µL Kit manufacturer's guidelines. PC-12 cells were cultured until confluent and counted using an automated cell counter (Countess 2 Automated Cell Counter, Invitrogen) and the desired number of cells ($2x10^7$ x (0.1 x number of reactions)) were removed and pelleted at 400 RCF for 5 minutes. Cells were then washed with phosphate buffered saline (PBS) and pelleted again. Subsequently, the pellet was resuspended with Resuspension Buffer R (100 µL buffer x number of reactions) and 100 µL of cells per reaction were gently mixed with 6 µg of the desired DNA plasmid. This reaction mixture was then inserted into the Neon station and electroporated at 1500V at 10 ms for 3 pulses. Transfected cells were plated into PC-12 DMEM media (see Section 2.1.1) without P-S. When culturing transfected cells for downstream applications, P-S was added directly to the plate 2 days post transfection.

For primary neuronal cultures, $2x10^6$ cells per sample were electroporated as above using Neon settings of 1300 V, 20 ms, 2 pulses. Cells were then plated into 6 cm dishes containing equilibrated BrainPhys Plating Medium.

2.1.3 Differentiation of PC-12 Cell Line

Acid-washed coverslips (Section 2.8.1) were incubated with 100 mg/mL poly-L-lysine (PLL) at 37°C with 5% CO_2 for 2 - 16 hours, before being washed twice with sterile distilled water. PC-12 cells were then plated onto coverslips in a 6 well plate, in 2 mL Opti-MEM media (Thermo Fisher Scientific, Catalogue No: 11520386) supplemented with 50 ng/mL nerve growth factor (NGF) (Proteinech Group, Catalogue No: HZ-1222). After 3 - 6 hours, 0.5% FBS was added to the media. Complete media changes occurred every 2 days for 1 week.

2.1.4 Preparation of Primary Cortical Neurons

Cortical neurons were prepared from E18 Sprague-Dawley rat embryos (dissection carried out at E18), supplied by Charles River. First, the cortices were dissected from the brain in ice-cold dissection solution (Table 2.1) and placed in a 10 cm dish on ice containing 2-3 mL of dissection solution. Cortical tissue was then transferred into a 10 mL tube and washed with 10 mL of HBSS (-/-) w/o Ca2+ or Mg2+ with Phenol Red (Thermo Fisher Scientific, Cat: 12082739). Once tissue had settled at the bottom of the tube, the HBSS (-/-) was removed and the tissue was resuspended in 5 mL HBSS (-/-) w/o Ca2+ or Mg2+ with Phenol Red. 70 µL of 2.5% trypsin stock (Fisher, Cat: 11538876) was then added to the tube, and the tube was then incubated at 37°C for 15 minutes. The tissue was gently agitated every 5 minutes during this incubation step. Subsequently, 5 mL DNAse working solution (Table 2.1) was added and the tube was gently inverted. Once the tissue had settled at the bottom, the media was aspirated.

To triturate the tissue, 5 mL triturating solution (Table 2.1) was added to the tube using a 5 mL stripette. The cortices were gently triturated up and down 15 times. The tissue was then allowed to settle for 2 minutes. After this time, the supernatant containing isolated cortical neurons was transferred to a new tube. 5 mL trituration solution was again added to the original tube and the trituration cycle was repeated 3 - 4 times until there was no visible pellet remaining. Subsequently, the cortical neurons were counted using a haemocytometer and 0.4% Trypan Blue Stain (Thermo Fisher Scientific, Cat: T10282) at a 1:1 ratio of stain to cells. Cells were immediately plated (Section 2.1.5) or electroporated and then plated (Section 2.1.2).

2.1.5 Culture of Primary Cortical Neurons

Glass coverslips were first prepared by transferring the desired number of coverslips into a glass petri dish and baking in an oven at 175°C for 2 hours. Plates and dishes for neuron plating were coated in PLL (100 mg/mL for plastic dishes or plates, 250

mg/mL for glass coverslips) overnight at 37°C with 5% CO_2. The dishes and plates were then washed twice with sterile double-distilled H_2O (ddH_2O) and dried well. Brain Phys Plating Medium (Stem Cell Technologies, Cat: 05794) containing 2% NeuroCult SM1 Neuronal Supplement (Stem Cell Technologies, Cat Cat: 05711) and 0.25% GlutaMAX Supplement (Thermo Fisher Scientific, Cat: 11574466) or neuron maintenance media (Table 2.1) was then added and equilibrated at 37°C with 5% CO_2 for 1 - 3 hours before plating. After extraction, the desired amount of cortical neurons were added drop-wise to the dish, followed by gentle agitation back and forth. The plates were sealed with micropore tape and incubated at 37°C with 5% CO_2.

For long-term cultures (7+ days), a partial media exchange was carried out every 4 - 5 days. For neurons cultured in Brain Phys media, 5 days post-extraction half the BrainPhys Plating Medium was removed and replaced with the same volume of Brain Phys Maturation Medium (Stem Cell Technologies, Cat: 05794) containing 2% NeuroCult SM1 Neuronal Supplement and 1% P-S. Subsequently, the cells underwent a 50% media change using equilibrated BrainPhys Maturation Medium every 3 - 4 days. For cells in maintenance media, every 4 - 5 days 1 - 2 mL was removed and and replaced with 2 - 3 mL fresh maintenance media. To prevent aggressive growth of glia and astrocytes, 1-β-D-Arabinofuranosylcytosine (EMD Millipore Corp. Catalogue No.: 251010-1GM) was added to the first feed (with the exception of cultures used for RNA extractions) at a final concentration of 1 µM.

2.1.6 Plating Cells onto Acid-Washed Poly-L-Lysine Treated Coverslips

13mm acid-washed coverslips were removed from ethanol storage and allowed to dry entirely. Coverslips were then transferred to the desired dish and an appropriate concentration of PLL (100 mg/mL for PC-12 and neuronal lines) was applied to the coverslip. The dish was then incubated at 37°C with 5% CO_2 for 2 - 4 hours or overnight. Before use, the PLL was aspirated and the coverslips washed twice with

Table 2.1: Neuronal Dissection and Culture Media Composition.

Component	Final Conc.	Supplier (Catalogue No.)
Dissection Solution		
HBSS (-/-)	1X	Fisher (15420614)
1M HEPES	10 mM	Fisher (11560496)
DNAse Stock Solution		
HBSS (+/+)	1X	Fisher (15420614)
DNAse	0.02%	Merck (D5025-150KU)
DNAse Working Solution		
HBSS (+/+)	1X	Fisher (15420614)
DNAse Stock Solution	0.1%	-
Trituration Solution		
HBSS (+/+)	1X	Fisher (15420614)
Albumax I	1%	Fisher (11530376)
Trypsin Inhibitor	0.05%	Merck (T9003-25MG)
DNAse Stock Solution	0.1%	-
Maintenance Media		
Neurobasal medium	1X	Gibco (21103-049)
B-27 supplement	2%	Gibco (17504044)
100X GlutaMAX	1%	Gibco (35050038)
2.5M Glucose Solution	0.6%	Sigma (G8769)
100X Pencillin Streptomycin	1%	Corning (30-001-CI)

ddH_2O. Once sufficiently dry, the coverslips were transferred into the desired plate and an appropriate media for the cell line was added. The desired amount of cells were plated, the dish gently agitated, and incubated at 37°C with 5% CO_2 overnight to allow cells to adhere.

2.1.7 Transfection of Primary Cortical Neurons

2.1.7.1 Transfection Using Lipofectamine 2000

For immunofluorescence assays, cortical neurons were plated into a 6 cm dish containing 8 PLL-coated coverslips at a density of 900,000 cells per mL. Transfections were carried out on day 4 or day 6 post-plating. Briefly, 10 mL of Neurobasal media without supplements (NB^-) and maintenance media (Table 2.1) were equilibrated in an incubator at 37°C for 1 hour before transfection. For each transfection carried out, two mixtures were prepared; a microcentrifuge tube containing 100 µL NB^-

and 1 µg maxiprep DNA, and a microcentrifuge tube containing 100 µL NB^- and 3.3 µL Lipofectamine 2000 reagent (Invitrogen, Catalogue No.: 11668-030). After incubation for 5 minutes at room temperature, the Lipofectamine 2000 dilution was added drop-wise to the DNA dilution and the mixtures were allowed to complex for 30 minutes at room temperature.

Neuron-covered coverslips were transferred into the inner wells of a 24 well plate. 1 mL ddH_2O was added to the outermost wells of the plate to help prevent evaporation in the incubator. Additionally, 3 mL of new maintenance media was combined with the conditioned media present in the 6 cm dishes, and 0.5 mL of this media was added to each well. The remaining media was equilibrated in the incubator until needed. After complexing, the DNA mixture was added to the coverslips drop-wise and the plate was agitated gently to mix. The plate was then incubated for 1 - 1.5 hours to allow the transfection to take place. Finally, the media was removed from the coverslips and replaced with 0.5 mL of the equilibrated conditioned maintenance media.

For all other assays, transfections were carried out in 10 cm dishes, 6 well plates or live cell imaging dishes. The density of cells and amount of reagents required varied according to the surface area of the culture vessel used (Table 2.2). All timing steps were as previously described. For example, for a 10 cm dish, $3x10^6$ cells were plated and cultured until days *in vitro* (DIV) 4. Dilutions containing 95.7 µL Lipofectamine 2000 and 28.9 µg DNA were prepared in 1 mL equilibrated NB^-. After complexing (see above), media was removed from the dish until approximately 3 mL media remained, and the transfection solution was added drop-wise. After 1 - 1.5 hours, the media was replaced entirely with the removed conditioned media made up to 10 mL with fresh maintenance media (Table 2.1) or BrainPhys maturation media (Section 2.1.5).

Table 2.2: Cell Densities and Reagent Amounts Used for Lipofectamine 2000 Transfection.

Culture Vessel	Cell Density	Reagent Amount	DNA Amount
10 cm dish	$3x10^6$	95.7 µL	28.9 µg
6 well plate	$0.5x10^6$	16.2 µL	4.9 µg
Live cell dish	$0.18x10^6$	5.9 µL	1.8 µg

2.1.7.2 Transfection Using Lipofectamine 3000

Transfections using the Lipofectamine 3000 Transfection Kit (Invitrogen, Catalogue No.: L3000-001) were done as per the manufacturer's protocol, using 1 µg of maxiprepped DNA. Coverslips were transferred from a 6 cm dish into a 24 well plate for transfection, as outlined previously (Section 2.1.7.1).

2.2 Molecular Biology

2.2.1 Constructs

The constructs used for cloning and CRISPR construct validation are listed in Table 2.3. pSpCas9(BB)-2A-GFP (PX458) was a generous gift from Dr Dan Bose, originally purchased from Addgene. Similarly, pET17-K560-Halo-Strep was a generous gift from Dr Emma Turner.

2.2.2 Bacterial Culture and Plasmid DNA Preparation

2.2.2.1 Preparation of Overnight Bacterial Cultures and Agar Plates

Using proper aseptic technique, the desired amount of Luria Broth (Table 2.4) was added to a sterile tube or culture flask alongside the appropriate antibiotic at a final concentration of either 100 µg/mL (carbenicillin) and 50 µg/mL (kanamycin). A scraping of a glycerol stock or a single colony from an agar plate was aseptically introduced to the broth, with the tip used being dispensed into the flask. The flask was then incubated for 12-16 hours overnight on a shaker at 37°C at 180 revolutions

per minute (RPM).

100 mL of Luria Agar (Table 2.4) was prepared in a conical flask and autoclaved. The flask was allowed to cool until hot to the touch (approximately 60°C) and the appropriate antibiotic was added at the required concentration (see above) using sterile technique. The agar was swirled to mix the antibiotic. Approximately 15 mL of agar was then poured into 10 cm sterile culture dishes, ensuring the dishes are kept close to the Bunsen burner. When the agar solidified, the lids were put on the dishes - doing this when the agar is still hot can cause the formation of condensation. After 1 hour, the plates were wrapped in Parafilm and stored at 4°C.

2.2.2.2 Mini-preparation of Plasmid DNA

Mini-preparation of plasmid DNA was done using Monarch Plasmid Miniprep Kit (New England BioLabs Inc. Cat No: T1010S). 1 mL of overnight bacterial culture was pelleted at 16,000 RCF at room temperature for 30 seconds. Cells were lysed, and the DNA bound to the column and washed following the manufacturer's instructions. To elute DNA, 30 µL 37°C ddH_2O was applied to the column and incubated for 5 minutes at room temperature, before centrifuging again for 1 minute at 16,000 RCF. The final DNA concentration was determined by measuring the absorbance at 260 nm using a NanoDrop ND-1000 Spectrophotometer (Labtech International).

2.2.2.3 Midi-preparation of Plasmid DNA

Midi-preparation of plasmid DNA was done using PureYield™ Plasmid Midiprep System (Promega Corporation). 100 mL of bacterial culture was used in the preparation, following the manufacturer's instructions. Elution of DNA was done using 37°C ddH_2O. The final DNA concentration was determined by measuring the absorbance of the samples at 260 nm on a NanoDrop ND-1000 Spectrophotometer (Labtech International).

2.2.2.4 Maxi-preparation of Plasmid DNA

Maxi-preparation of plasmid DNA was done using PureYield™ Plasmid Maxiprep System (Promega Corporation). 200-250 ml of bacterial culture was used in the preparation, following the manufacturer's instructions. Elution of DNA was done using 37°C ddH_2O. The final DNA concentration was determined by measuring the absorbance of the samples at 260 nm on a NanoDrop ND-1000 Spectrophotometer (Labtech International).

Table 2.3: Constructs Used for Cloning, CRISPR Validation, and Transfection.

Construct	Antibiotic Resistance	Source	Reference
pSpCas9(BB)-2A-GFP (PX458)	Ampicillin	Feng Zhang (Addgene plasmid: 48138)	(Ran et al. 2013)
pORANGE Cloning Vector	Ampicillin	Harold MacGillavry (Addgene plasmid: 131471)	(Willems et al. 2020)
pORANGE GFP-Actb KI	Ampicillin	Harold MacGillavry (Addgene plasmid: 131479)	(Willems et al. 2020)
PX458-RPL4-gRNA-T1	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPL4-gRNA-T2	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPL22-gRNA-T1	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPL22-gRNA-T2	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPS2-gRNA-T1	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPS2-gRNA-T2	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPS3-gRNA-T1	Ampicillin	Creation described in Section 2.3.2.1	-
PX458-RPS3-gRNA-T2	Ampicillin	Creation described in Section 2.3.2.1	-
pORANGE RPL4-HaloTag KI	Ampicillin	Creation described in Section 2.3.3	-
pORANGE RPL22-HaloTag KI	Ampicillin	Creation described in Section 2.3.3	-
pORANGE RPS2-HaloTag KI	Ampicillin	Creation described in Section 2.3.3	-
pORANGE RPS3-HaloTag KI	Ampicillin	Creation described in Section 2.3.3	-
pET17-K560-Halo-Strep	Ampicillin	Emma Turner (Twelvetrees Lab)	-
mEGFP-N1	Kanamycin	Michael Davidson (Addgene plasmid: 54767)	-
pHAAP1	Ampicillin	Thermo Fisher Scientific (Custom)	-
pN-StrepTag-HaloTag	Ampicillin	Thermo Fisher Scientific (Custom)	-

2.2.2.5 Plasmid DNA Concentration via Precipitation

After the plasmid preparation intended for Neon transfection is completed, 1/10 DNA sample volume of 3M Na-Acetate pH 5.2 and 2 volumes of ice-cold 100% ethanol was added to the sample. The sample was mixed and stored at -20°C for 2 - 6 hours or overnight to precipitate the DNA. Samples were then centrifuged at room temperature for 20 minutes at 17,000 RCF. After the supernatant was aspirated, the pellet was washed twice with room-temperature 70% ethanol and then allowed to air-dry for approximately 1 hour. 100 µL ddH_2O was applied to the tube and the sample was left at 4°C overnight to allow the DNA to dissolve into solution. The final DNA concentration was determined by measuring the absorbance of the samples at 260 nm on a NanoDrop ND-1000 Spectrophotometer (Labtech International).

2.2.2.6 Bacterial Transformation Using Plasmid DNA

One aliquot of 100 µL competent Stbl3 cells were thawed on ice for 10 minutes. When defrosted, 1-2 µL of plasmid DNA was added to the cells and mixed by flicking 5 times. The mixture was then incubated on ice for 30 minutes before being heat shocked at 42°C for 45 seconds and cooled again on ice for 2 minutes. Following this, 900 µL of room temperature Super Optimal broth with Catabolite repression media (New England Biolabs, Catalogue No: B9020S) was added to the sample and the bacteria were incubated on a shaker at 37°C, 250 RPM for 45 - 60 minutes to allow for outgrowth. The bacteria were then centrifuged at 4,500 RCF for 5 minutes and 800 µL of the supernatant removed. The pellet was resuspended in the remaining 100 µL and spread across pre-warmed selection plates containing the appropriate antibiotic (working concentrations for carbenicillin and kanamycin are 100 µg/mL and 50 µg/mL respectively). Plates were incubated at 37 °C overnight to allow colony formation.

Table 2.4: Composition of Bacterial Culture Solutions.

Solution	Component	Final Concentration
Luria Broth	Tryptone	1%
	Yeast Extract	0.5%
	NaCl	171 mM
	ddH_2O	—
Luria Agar	Tryptone	1%
	Yeast Extract	0.5%
	NaCl	171 mM
	Agar	2%
	ddH_2O	—

2.2.3 Genomic DNA Preparation

Two days post-transfection, PC-12 cells were centrifuged at 500 RCF for 5 minutes and washed with PBS. After being pelleted again, the PBS was aspirated and the pellet was resuspended in 200 µL of QuickExtract DNA Extraction Solution (Lucigen, Catalogue No: QE0905T). The solution was then vortexed for 15 seconds and heated at 65°C for 6 minutes. Following this, the samples were again vortexed for 15 seconds and then heated at 98°C for 2 minutes. The DNA concentrations were measured using the Broad Range DNA Kit (Invitrogen, Cat No: Q32850) on the Qubit 4 Fluorometer (Invitrogen, Cat No: Q33238) following the manufacturer's protocol and stored at -80°C.

2.2.4 DNA Amplification

2.2.4.1 Polymerase Chain Reaction

Forward and reverse cloning primers were designed using Benchling (Benchling Biosciences), ranging in length from 18 - 24 nucleotides with 40 - 60% G/C content unless otherwise specified. Once designs were finalised, single-stranded DNA primers were synthesised by Integrated DNA Technologies. 100 µM stocks were made using filtered ddH_2O. On ice, the reaction mixture (Table 2.5) was made up in 0.5 ml sterile microcentrifuge tubes containing the desired DNA template. Negative control

samples containing no template DNA were also set up for every assay. The samples were mixed and spun down with a benchtop microcentrifuge to ensure the reaction mixture was at the bottom of the tube.

Table 2.5: PCR Reaction Mixture Composition.

Component	Final Concentration
Template DNA	50-100 ng
Forward Primer	10 µM
Reverse Primer	10 µM
Q5 High-Fidelity 2X Master Mix	1X
ddH_2O	Up to 25 µL

Subsequently, the samples were heated using a standard polymerase chain reaction (PCR) reaction cycle (Table 2.6) with the annealing temperature and extension time varying dependent on the size of the DNA fragment being amplified. In the instance that the amplification was unsuccessful, further changes were made to the cycle including; increasing the number of amplification cycles, extension time, final extension time, and annealing temperature.

Table 2.6: General PCR Protocol.

Step	Temperature (°C)	Time (Seconds)	Cycles
Initial Denaturation	98	30	1
Denaturation	98	10	
Annealing	T_A	30	30
Extension	72	20-30 /kb	
Final extension	72	120	1
Hold	10	∞	1

2.2.4.2 DNA Gel Electrophoresis and Imaging

Agarose gels were prepared by mixing 1X lithium borate buffer (10 mM lithium acetate dihydrate and 10 mM boric acid in distilled H_2O) with the desired amount of agarose in a conical flask. All gels were made with 1% agarose unless otherwise stated. After mixing thoroughly, the flask was heated at medium heat in a microwave

until the agarose had dissolved entirely. The gel was cooled slightly by running the flask under cold water for approximately 30 seconds. SYBR Safe DNA stain (Thermo Fisher Scientific, Catalogue No: 10328162) was added to the flask at a 1X concentration and swirled to mix. The gel was poured into a gel cassette and the desired comb applied ensuring no bubbles are introduced.

To visualise DNA, samples were mixed with an appropriate amount of 6X Gel Loading Dye (New England Biolabs, Catalogue No: B7024) and loaded onto the gel. For most gels, a 1kb DNA ladder (New England Biolabs, Catalogue No: N3232S) was run alongside the samples, unless otherwise stated. The gel was run at 80V for 30 minutes and visualised on a G:Box Chemi XRQ (SynGene), blueBox Pro Transilluminator (Cambio, Catalogue No: MP-QP-1700-03), or a LI-COR XF Odessey using the 600 nm channel.

2.2.5 Restriction Digest of DNA

Restriction digests were carried out using restriction enzymes (Table 2.7) from New England Biolabs according to the manufacturer's instructions. Briefly, 1 µg of DNA was incubated with 5 µl 10X rCutSmart Buffer (New England Biolabs, Catalogue No: B6004S) and 1 µl (20 units) of enzyme, made up to 50 µL with ddH_2O. The reaction mixture was incubated at the recommended temperature for 5-15 minutes. The samples were then heat-inactivated if required. If a double-digest was carried out with two enzymes with different heat inactivation temperatures, the higher temperature was selected.

Table 2.7: Restriction Enzymes Used in Cloning.

Restriction Enzyme	Catalogue No.	Incubation	Heat Inactivation
BbsI-HF	R3539S	37°C	65°C, 20 minutes
HindIII-HF	R3104S	37°C	80°C, 20 minutes
XhoI	R0146S	37°C	65°C, 20 minutes

2.2.6 DNA Extraction from Agarose Gel

DNA extraction was done using Monarch DNA gel extraction kit (New England Biolabs, Catalogue No: T1020S). The band of interest was excised from the gel using a clean scalpel, transferred into a clean 1.5 mL microcentrifuge tube, and weighed. 4 volumes of Monarch Gel Dissolving Buffer was added to the microcentrifuge tube. The gel slice was heated at 50°C with regular agitation until fully dissolved. The sample was then loaded onto a column and spun at 16,000 RCF for 1 minute. After discarding the flow-through, 200 µL DNA Wash Buffer was then added to the column and the column was centrifuged again at 16,000 RCF for 1 minute. This wash step was repeated once again. Finally, the DNA was eluted in 6 - 10 µL ddH_2O. The DNA concentration was determined by measuring the absorbance of the samples at 260 nm on a NanoDrop ND-1000 Spectrophotometer (Labtech International).

2.2.7 DNA Ligation

Ligation of digested PCR products and plasmids was done using T4 DNA Ligase (New England Biolabs, Catalogue No: M0202S) following the manufacturer's instructions. Briefly, a 20 µL reaction mixture was set up in a 0.5 mL microcentrifuge tube on ice (2 µL 10X T4 DNA Ligase Buffer, plasmid and insert DNA at a 1:3 molar ratio, 1 µL T4 DNA ligase, and up to 20 µL ddH_2O). The reaction mixture was mixed by gentle pipetting and incubated at room temperature for 10 minutes. The reaction was then heat inactivated at 65°C for 10 minutes. The product was then stored on ice and transformed into competent cells (Section 2.2.2.6).

2.2.8 Sanger Sequencing of DNA

Sanger sequencing of DNA constructs and PCR products was done by Eurofins Scientific or Azenta Life Sciences. Primers used for sequencing are provided in Table 2.8.

Table 2.8: Primers Used for Sanger Sequencing.

Name (Direction)	**5'-3' Sequence**	**Usage**
U6 (Fwd)	GAGGGCCTATTTCCCATGATTCC	PX458/pORANGE plasmids
RPL4 T7E1 (Fwd)	CTCCCCTCCCCTGAGACAGGAATTTATGAG	RPL4 T7 Endonuclease I products
RPL22 T7E1 (Fwd)	GAGCACTGGCTTGGGTGTACTTGGTCTGAG	RPL22 T7 Endonuclease I products
RPS2-Halo (Fwd)	GTTACTGGGCTCTGATCTTG	RPS2 T7 Endonuclease I products
RPS3 T7E1 (Fwd)	TGAATAGATGCTGCTCTCCCCGCTGTCCTC	RPS3 T7 Endonuclease I products
Halo End (Fwd)	GCAACATCATCCCGCATG	pORANGE-Halo plasmids
Halo Middle (Rev)	CCAGGGCTTCGATGAAGG	pORANGE-Halo validation

2.3 Gene Editing

2.3.1 Guide RNA Design

Guide RNAs sequences targeting *Rpl4*, *Rpl22*, *Rps2* and *Rps3* were identified using the Knock-In CRISPR Tool by Benchling Biosciences. The guides were designed as described in Chapter 3 (Table 3.2) using the basic guidelines from Willems et al. 2020. For insertion into the pORANGE cloning vector and precursor PX458 plasmids, additional 3' and 5' overhangs complimentary to the Bbs1 restriction sites were added. Additionally, an extra guanine base was also inserted before the guide sequence to aid transcription from the U6 promoter present in the plasmids. Single-stranded DNA oligos (Table 2.9) were then ordered from Integrated DNA Technologies Inc. and reconstituted to 100 µM concentration using filtered ddH_2O.

Table 2.9: Forward and Reverse Guide RNA Oligos for ORANGE System.

Protein Target	Direction	5' - 3' Sequence
RPL22 Guide 1	Fwd	CACCGGCGCTACTTCCAGATCAACC
	Rev	AAACGGTTGATCTGGAAGTAGCGCC
RPL22 Guide 2	Fwd	CACCGCTTCCAGATCAACCAGGACG
	Rev	AAACCGTCCTGGTTGATCTGGAAGC
RPL4 Guide 1	Fwd	CACCGGGAGACAAAAAACCTGCAGT
	Rev	AAACACTGCAGGTTTTTTGTCTCCC
RPL4 Guide 2	Fwd	CACCGCAAACTTCGAGTGAAAAAGC
	Rev	AAACGCTTTTTCACTCGAAGTTTGC
RPS2 Guide 1	Fwd	CACCGCTCTTGTCACTAGCAAATGG
	Rev	AAACCCATTTGCTAGTGACAAGAGC
RPS2 Guide 2	Fwd	CACCGTAGCAAATGGCGGATGACGC
	Rev	AAACGCGTCATCCGCCATTTGCTAC
RPS3 Guide 1	Fwd	CACCGACTGGCTGAGGCATGGCGGG
	Rev	AAACCCCGCCATGCCTCAGCCAGTC
RPS3 Guide 2	Fwd	CACCGGCCATGCCTCAGCCAGTGCC
	Rev	AAACGGCACTGGCTGAGGCATGGCC

2.3.2 Determining Guide RNA Cutting Efficiency

2.3.2.1 Guide RNA Insertion into PX458/pORANGE

The single-stranded guide RNA oligos were annealed together using the reaction mixture and PCR times in Table 2.10 and Table 2.11 respectively. The annealed oligos were then diluted 1:200 and the ligation reaction was set up (Table 2.12). A reaction containing no annealed oligo duplex was also included as a negative control. The ligation reaction was incubated in a thermocycler at 37°C for 5 minutes and 21°C for 5 minutes for 6 cycles. Lastly, the product was transformed into competent Stbl3 cells (Section 2.2.2.6). Competent Stbl3 cells were created by members of the Twelvetrees lab.

Table 2.10: Guide RNA Oligo Annealing Reaction Mixture.

Component	Final Concentration
100 µM Forward gRNA Oligo	10 µM
100 µM Reverse gRNA Oligo	10 µM
10X T4 Ligation Buffer	1X
T4 Polynucleotide Kinase	1 unit
ddH_2O	Up to 10 µL

Table 2.11: Guide RNA Duplex Formation Protocol.

PCR Step	Temperature (°C)	Incubation Time (min)
Phosphorylate	37	30
Denature	95	5
Anneal	95 - 25 (-5°C/min)	14
Hold	10	∞

Colonies were selected from the selection plates and DNA was extracted (Section 2.2.2.2) from overnight bacterial cultures (Section 2.2.2.1) of these colonies. Subsequently, guide RNA insertion was validated using a restriction digest using BbsI-HF (Section 2.2.5). If the insert was present, the BbsI restriction sites within the plasmid were destroyed and the plasmid was not cut in the presence of BbsI enzyme. Constructs created from this protocol are listed in Table 2.3.

Table 2.12: Guide RNA Oligo Ligation Reaction.

Component	Final Concentration
Plasmid DNA	100 ng
Oligo duplex (1:200)	0.05 µM
10 mM Dithiothreitol	0.5 mM
10 mM adenosine triphosphate	0.5 mM
Bbs1-HF	1 unit
10X Cutsmart Buffer	1X
T7 Ligase	1 unit
ddH_2O	Up to 20 µL

2.3.2.2 T7 Endonuclease I Assay and ICE Analysis

The ability of guide RNAs to create cuts at the desired location was analysed using a T7 endonuclease I (T7EI) assay. Briefly, the PX458 plasmid containing the desired guide RNA was transfected into PC-12 cells (Section 2.1.2). 2 days post transfection, the cells were harvested and the DNA extracted using QuickExtract DNA Extraction Solution (Section 2.2.3).

First, the region of interest was amplified by PCR using the reaction amounts in Table 2.5 and times specified in Table 2.14. The region of interest was an approximately 400 - 500 bp section of the gene spanning 200 - 250 bp upstream and downstream from the expected Cas9 cut site. Primers were designed using Benchling and are shown in Table 2.13. 50 ng of DNA from wild type (WT) and transfected cells was used as using more than 5 µL of DNA in QuickExtract solution can inhibit the PCR reaction. An additional negative control containing no template DNA was also included in the assay. PCR products were then run on a 1% agarose gel and imaged using a G:BOX Chemi XRQ (Syngene) to confirm that a DNA band of the correct size was present.

Subsequently, two 19 µL annealing reactions (5 µL PCR product, 2 µL 10X NEBuffer 2, 12 µL ddH_2O) were made for each sample. The DNA products were then denatured and annealed using the PCR cycle in Table 2.15 to form WT/Edited DNA heteroduplexes. Digestion of the heteroduplexes was done by adding 1 µL T7EI (New England BioLabs, Catalogue No.:M0302S) to the sample and mixing well. For

negative control samples, 1 µL ddH_2O was added instead of enzyme. The samples were then spun down and incubated at 37°C for 15 minutes. Upon completion, 1.5 µL 0.25M ethylenediaminetetraacetic acid (EDTA) solution was added to stop the reaction. To allow for better separation of DNA bands, the samples were run on a 2% agarose gel alongside a 100 bp DNA ladder (New England Biolabs, Catalogue No: N3231) and imaged on a G:BOX Chemi XRQ (Syngene).

If the T7EI assay was successful, the initial DNA amplification was repeated and the product cleaned using a Monarch PCR & DNA Cleanup Kit (New England BioLabs, Catalogue No: T1030S) or QIAquick PCR Purification Kit (Qiagen, Catalogue No: 28104). The product was then sent for Sanger sequencing and the data generated was used to determine the efficiency of guide RNA cutting using the Synthego ICE CRISPR Analysis Tool (Synthego Performance Analysis, ICE Analysis. 2019. v3.0. Synthego; [2020-2022]). Algorithms to generate CRISPR editing scores are provided in Hsiau et al. 2018. Raw data from the Synthego ICR CRISPR Analysis tool was plotted in R (version 4.2.0). The code used for analysis is provided in the Appendix, Figure 8.1.

Table 2.13: Primers Used to Amplify the Genomic Region of Interest for T7E1 Assay.

Target (Direction)	5'-3' Sequence	T_A (°C)	Amplicon size (base-pairs)
RPL4 (Fwd)	CTCCCCTCCCCTGAGACAGGAATTTATGAG	62.5	544
RPL4 Rev	GCTCTTTCTTGCTCTGTAGTGCTGATG	59.8	
RPL22 Fwd	GAGCACTGGCTTGGGTGTACTTGGTCTGAG	65.4	408
RPL22 Rev	AGCCCAGTTAAACCCCAGTCCTGTCTTTCC	64.8	
RPS2 Fwd	GTATAGACGTTCGAGATCTCGCGGCT	61.8	469
RPS2 Rev	TCAGTCAAAGGGCTGCATCTCCGTG	62.8	
RPS3 Fwd	TGAATAGATGCTGCTCTCCCCGCTGTCCTC	66.0	396
RPS3 Rev	GATACCACATCCCAACTTGCATTCTCCCAC	62.6	

Table 2.14: PCR protocol for T7E1 Assay DNA Amplification.

PCR Step	Temperature (°C)	Time (Seconds)	Cycles
Initial Denaturation	98	30	1
Denaturation	98	5	40
Annealing	T_A	10	
Extension	72	20	
Final extension	72	120	1
Hold	10	∞	-

Table 2.15: PCR Protocol for T7E1 Assay Heteroduplex Formation.

Step	Temperature (°C)	Time (Seconds)
Initial Denaturation	95	300
Annealing Step 1	95 - 85 (-2°C/s)	5
Annealing Step 2	85 - 25 (-0.1°C/s)	600
Hold	10	∞

2.3.3 pORANGE Plasmid Creation

To create a functional ORANGE construct, the donor DNA must be cloned into the pORANGE cloning vector (Addgene Plasmid #131471) alongside the selected guide RNAs (Table 2.9, Section 2.3.2.1). The donor DNA insert includes the HaloTag sequence flanked by the guide RNA sequence and associated PAM site, with a HindIII and XhoI restriction sites on the 3' and 5' end respectively. As a feature of the ORANGE system, the guide RNA sequence and PAM are inserted into pORANGE as the reverse complement to ensure that the donor DNA is inserted in the correct orientation within the genome. If the donor DNA is inserted in the reverse orientation, the guide RNA and PAM are reconstituted and the Cas9 enzyme will recognise and remove the donor DNA. The restriction sites are present to allow the donor DNA to be cloned into the cloning vector using basic restriction cloning techniques.

Importantly, the donor DNA also includes small linker regions before and after the desired tag to ensure that the insertion is in frame and creates no frameshift mutations. These linker regions are composed of ~3 uncharged amino acids - glycine and serine are commonly used - with additional bases to make the insert in frame.

In the instance that the insertion is before the initiation codon (e.g. RPS2 target 1), a new initiation codon and Kozak sequence was included in the linker. Similarly, where the donor DNA was incorporated into the C terminus (RPL4, RPL22, RPS3) a stop codon was included after the HaloTag sequence. In some cases, this resulted in a slightly truncated tagged protein.

To create the donor DNA insert, forward and reverse primers were designed to amplify the HaloTag gene from a template vector, pET17-K560-Halo-Strep. The primers designed (Table 2.16) were approximately 60 bases long and contained the features necessary for the PCR amplification and insertion of the donor DNA into pORANGE. This included the necessary restriction sites; the PAM and target sequences; the linker sequence; and part of the HaloTag sequence to enable HaloTag amplification via PCR. The order and orientation of these sequences were dependent on the location of the target sequence within the genome - on the coding or complementary DNA strands.

If the target sequence was present on the coding DNA strand, the target and PAM are present in the primers in a reverse orientation. The order of the forward primer is as follows: HindIII restriction site, reverse complement PAM, reverse complement target sequence, linker region (6-9 bases coding for G or S amino acids, including a start codon and kozak sequence if required), and the first 18 bases of the HaloTag gene (not including the start codon). The reverse primer is as follows: the last 18 bases of the HaloTag sequence (including a stop codon if required), linker region, reversed PAM, reversed target sequence, and XhoI restriction site.

If the target sequence was on the complementary DNA strand, the target sequence and PAM are not reversed. The order of the forward primer is as follows: HindIII restriction site, target sequence, PAM, linker region (6-9 bases coding for G or S amino acids, including a start codon and kozak sequence if required), and the first 18 bases of the HaloTag gene (not including the start codon). The reverse primer is as follows: the last 18 bases of the HaloTag sequence (including a stop codon if required), linker region, complementary target sequence, complementary

PAM, and XhoI restriction site. For both the coding and complementary primers, a generic PAM sequence of GGG was used.

The donor insert PCR was set up using the reaction mixture and timings described previously (Table 2.5 and Table 2.6 respectively). The PCR timings were slightly amended to allow for 40 cycles of denaturation, annealing and extension. Following this, the PCR product and the pORANGE vectors with the guide RNA inserts present were digested with HindIII and XhoI restriction enzymes (Section 2.2.5). The products were resolved on an agarose gel and the DNA extracted from the gel (Section 2.2.6). Subsequently, the digested insert and plasmid were used in a ligation reaction and transformed into Stbl3 competent cells (Section 2.2.2.6), before selected colonies were grown up, miniprepped and validated by Sanger sequencing (Section 2.2.7).

Table 2.16: Primers for Donor DNA Insertion Into pORANGE.

60 bp primers were designed to create the donor DNA insert for each pORANGE-Halo plasmid. Primers were ordered from Integrated DNA technologies and reconstituted to a 100 µM concentration before use.

Target	5'-3' Sequence	Tm (°C)
RPL4 Fwd	ATAAAGCTTACCCACTGCAGGTTTTTTGTCTCCGGAGCGGCGGATCCGAAATCGGTACT	74.4
RPL4 Rev	ATACTCGAGGGAGACAAAAAACCTGCAGTGGGCCCCCTTAACCGGAAATCTCCAGAGT	74.0
RPL22 Fwd	ATAAAGCTTACCCCGTCCTGGTTGATCTGGAAGGGCTCTGGAGGATCCGAAATCGGTACT	73.9
RPL22 Rev	ATACTCGAGCTTCCAGATCAACCAGGACGGGGAGCGCCCTTAACCGGAAATCTCCAGAGT	75.5
RPS2 Fwd	ATAAAGCTTACCCCCATTTGCTAGTGACAAGAGGGGATGGGATCCGAAATCGGTACT	68.2
RPS2 Rev	ATACTCGAGCTCTTGTCACTAGCAAATGGGGGCGCCACTACCGGAAATCTCCAGAGT	69.7
RPS3 Fwd	ATAAAGCTTACTGGCTGAGGCATGGCGGGGGGGAGCGGTGGAGGATCCGAAATCGGTACT	77.4
RPS3 Rev	ATACTCGAGCCCCCCGCCATGCCTCAGCCAGTTTGAACCTTAACCGGAAATCTCCAGAGT	76.0

2.3.4 Fluorescence-Activated Cell Sorting of Edited Cells

Initial TMR-Halo labelling optimisation was carried out by Dr Smith of the Twelvetrees lab. Briefly, one week post transfection of HaloTag constructs, 100 nM Halo-TMR ligand (Promega, Catalogue No:G8251, Lot:0000400971) was added directly to cells in a 6 well plate. The plates were incubated at 37°C with 5% CO_2 for 1 hour. The cells were subsequently transferred to a tube and spun at 400 RCF for 5 minutes and the media containing Halo-TMR was aspirated. Cells were then transferred to a new 6 well plate containing fresh media and incubated for 2 - 3 hours to remove unbound Halo-TMR. Following this, cells were pelleted and washed with PBS, before being resuspended in 1 mL PBS and transported on ice to the BD FACSMelody Cell Sorter (BD Biosciences).

Cell populations to be sorted were put through a 35 µm cell strainer to ensure a single-cell suspension and then loaded onto the BD FACSMelody Cell Sorter. The data visualisation and gating was done using the associated BD FACSChorus software (BD Biosciences). The 561 nm laser and PE YG(A) filter were selected to excite the Halo-TMR ligand, and all fluorescent cells detected were sorted into a single well of a 6 well plate. For the GFP-Actin ORANGE CRISPR construct, cells were washed in PBS and taken directly to be sorted without the addition of fluorescent ligand. Fluorescence was detected using the 488 nm laser and cells were collected as above. All collected cells were incubated at 37°C with 5% CO_2 and cultured with the addition of filter-sterilised conditioned PC-12 media. Lines were expanded for several weeks until the populations were large enough for downstream applications.

The raw data from these assays was analysed and plotted in R (version 4.2.0) using RStudio (Postit Software, version 2022.12.0, build 353). The code generated is provided in the Appendix, Figure 8.2.

2.4 Protein Biochemistry

2.4.1 SDS-PAGE and Western Blotting

2.4.1.1 Western Blot Sample Preparation

Cell lysis buffer (Table 2.17) was made up on ice with protease inhibitors added - aprotinin (Melford, Catalogue No: A2301), pepstatin A (Melford, Catalogue No: P2203), leupeptin (Melford, Catalogue No: L22035-0.01), Nα-p-Tosyl-L-arginine methyl ester hydrochloride (Sigma Aldrich, Catalogue No: T4626-5G), and phenylmethylsulfonyl fluoride (Sigma, Catalogue No: P7626-5G). Cells were transferred to a 1.5 mL microcentrifuge tube and pelleted at 500 RCF for 5 minutes at 4°C. The media was aspirated, and the cells were washed in 1 mL PBS. After pelleting again as before, the PBS was removed, and the pellet resuspended in 100 µL lysis buffer. The samples were incubated on ice for 10 minutes and subsequently centrifuged at 17,000 RCF for 10 minutes at 4°C. 100 µL of supernatant was transferred to a clean 1.5 mL microcentrifuge tube and protein concentration was quantified using a NanoDrop ND-1000 Spectrophotometer (Labtech International). 50 µL 3x sample buffer (Table 2.17) was added and the samples were then heated at 100°C for 5 minutes. Lastly, the samples were cooled to room temperature and stored at -20°C.

2.4.1.2 SDS-PAGE

Acrylamide gels for SDS-PAGE were made and stored at 4°C until used. Briefly, 10% resolving gel (Table 2.17) was poured into a Invitrogen Novex Empty Gel Cassette (Fisher Fisher Scientific, Catalogue No: 11559156) and a thin layer of water was applied. The gel was allowed to set entirely, and the water was removed and replaced with 5% stacking gel (Table 2.17) to the top of the casing. The required 10 or 15 well Novex 1.5 mm comb (Thermo Fisher Scientific, Catalogue No: NC3510, NC3515) was then inserted. Care was taken to ensure no bubbles were introduced during this process. Once fully set, the gels were wrapped in wet blue Centrefeed roll and foil. The wrapped gels were stored at 4°C and used within 2 weeks.

To run a SDS-PAGE, an acrylamide gel with the white tape removed was placed into the SE300 miniVE Veritical Electrophoresis System (Hoefer) cassette and the tank was filled to the max line with 1x running buffer (Table 2.17). The comb was then removed, and the gel checked to ensure no air bubbles or defects are present within the wells. Samples to be run were heated at 95°C for 5 minutes to ensure the proteins were denatured. 5 µL of EZ-Run Prestained Protein Ladder (Thermo Fisher Scientific Inc., Catalogue No: 10638393) was added to the gel alongside the required amount of protein – most experiments used 50 - 100 µg of protein unless otherwise specified. Empty wells were filled with an equal amount of 1x sample buffer (100 µL 3x sample buffer (Table 2.17) in 200 µL ddH_2O). The gel was then run at 100V until the dye front enters the resolving gel, at which point the voltage was increased to 150V until the dye front reached the bottom of the gel. Subsequently, the gel was removed from the casing and used immediately in the transfer process unless an in-gel fluorescence assay was being carried out (see Section 2.4.3).

2.4.1.3 Transfer of PAGE Gels to PVDF Membrane

Merck Immobilon-FL PVDF membrane (Thermo Fisher Scientific, Catalogue No: 15740319) was activated using 100% methanol for 15 seconds, before being washed in ddH_2O for 2 minutes. Components of the transfer stack (sponges, filter paper, and membrane) were then soaked in transfer buffer (Table 2.17), ensuring the membrane was submerged in buffer for at least 5 minutes. After completion, the stacking gel and dye front was cut from the acrylamide gel and the miniVE blotting module (Sigma-Aldrich, Catalogue No: GE80-6418-96) was placed in the buffer black-side down. The transfer stack was prepared as follows: sponge, filter paper, acrylamide gel, membrane, filter paper (bubbles were removed by rolling the filter paper), and sponge. The complete transfer stack was shut and placed in the tank in the correct orientation. To keep the tank cool, the stack was submerged in iced, distilled H_2O. The transfer was run at 300 mA for 72 minutes. After the transfer was complete, the membrane was removed from the stack and protein was visualised using Ponceau S

stain solution (0.1% (w/v) Ponceau S, 5% glacial acetic acid). Unwanted sections of the membrane were removed using a sterile scalpel and the membrane was then washed in ddH_2O to de-stain.

2.4.1.4 Fluorescent Western Blotting

Membranes were blocked in 1% fish gelatin blocking solution (Table 2.17) with agitation at room temperature for 1 hour. Subsequently, the desired primary antibody (Table 2.18) was diluted in blocking solution (Table 2.17) and applied to the membrane overnight at 4°C. The following day, the membranes were washed 3 times for 15 minutes in approximately 5 mL of Tris-buffered saline with Tween 20 detergent (TBST) wash solution (Table 2.17), before addition of the secondary fluorescent antibody diluted in TBST wash solution. The membrane was incubated with secondary antibody (Table 2.19) on a roller for 1 hour at room temperature. Finally, the membrane was washed as before and imaged using the LI-COR Odyssey XF Imaging System (LI-COR Biosciences).

2.4.2 Chemiluminescent Western Blotting

Chemiluminescent western blots were carried out in the Thomson laboratory (School of Biosciences, University of Sheffield). First, 10% acrylamide gels (Table 2.17) were cast using glass plates and a 10 well comb was applied. SDS-PAGE was done as standard (Section 2.4.1.2), running 50 - 100 µg sample alongside 8 µL PageRuler™ prestained protein ladder (Thermo Fisher Scientific, Catalogue No: 11812124). The transfer process used Amersham Protran nitrocellulose membrane (Sigma Aldrich, Catalogue No: GE10600002), which was activated in transfer buffer for 5 minutes before the transfer. The transfer was carried out as previously described (Section 2.4.1.3) using the same timings and amperage.

The blot was then blocked with 5% milk solution (5% dried skimmed milk powder (Marvel) in 1X PBS). Primary antibodies diluted in milk solution were then applied (Table 2.18) and the blots were incubated at 4°C overnight. Blots were washed in

Table 2.17: Buffers and Solutions for SDS-PAGE, Transfer and Western Blot.

Buffer/Solution	Component	Final Concentration
Lysis Buffer (1 mL)	1M HEPES pH 7.5 0.5M EDTA 5M NaCl 20% Triton Protease Inhibitors ddH_2O	40 mM 1 mM 120 mM 0.05% 1X Up to 1 mL
10% Resolving Gel (25 mL)	30% Acrylamide 1.5M Tris-Cl pH8.8 10% SDS 10% APS TEMED ddH_2O	10% 0.375M 1% 1% 0.04% Up to 25 mL
5% Stacking Gel (8 mL)	30% Acrylamide 1.5M Tris-Cl pH6.8 10% SDS 10% APS TEMED ddH_2O	5% 0.125M 1% 1% 0.1% Up to 8 mL
10X Running Buffer (1 L)	Tris Base Glycine ddH_2O	250.1 mM 1.924 M Up to 1 L
1X Running Buffer (1 L)	10X Running Buffer 10% SDS ddH_2O	1X 0.1% Up to 1L
3X Sample Buffer (10 mL)	1M Tris pH 6.8 10% SDS 1M DTT Bromophenol Blue Glycerol ddH_2O	150 mM 2% 100 mM 0.1% 10% Up to 10 mL
Transfer Buffer (1 L)	10X Running Buffer 10% SDS ddH_2O	0.5X 0.005% Up to 1 L
1X TBST pH 7.6 (1 L)	Tris Base NaCl KCl Tween-20 ddH_2O	2.4g 8.76g 3.73g 2ml Up to 1 L
Blocking Solution	Cold water fish gelatin TBST	10% 1X
Wash Solution	BSA TBST	0.1% 1X

Table 2.18: Primary Antibodies Used in Western Blot Assays.

Abbreviations: NEB- New England Biolabs, Thermo- Thermo Fisher Scientific, Sigma - Sigma-Alrich.

Target / Clonality	Host	Conc.	Supplier	Catalogue No.	Lot No.
RPL22/Poly	Rabbit	1:500	Proteintech	5002-1-AP	00021636
RPL4/Mono	Mouse	1:10,000	Proteintech	67028-1-Ig	10009949
RPS2/Poly	Rabbit	1:1000	Thermo	PA5-99706	XK3763844A
RPS3/Mono	Mouse	1:20,000	Proteintech	66046-1-Ig	10002065
HaloTag/Poly	Rabbit	1:1000	Promega	G9281	0000460762
HaloTag/Mono	Mouse	1:1000	Promega	G921A	0000507057
GFP/Mono	Mouse	1:2500	Sigma	MAB3580	3018724
β-Actin/Mono	Mouse	1:5000	Proteintech	60008-1-Ig	10004413
β-Actin/Poly	Rabbit	1:1000	Proteintech	20536-1-AP	00104926
GAPDH/Mono	Mouse	1:5000	Proteintech	6004-1-1g	10013030
KIF5B/Poly	Rabbit	1:2000	Proteintech	21632-1-AP	00017112
SNAPtag/Poly	Rabbit	1:1000	NEB	P9310S	10061023

Table 2.19: Secondary Fluorescent Antibodies for Western Blot assays.

Fluorophore/Specificity	Conc.	Supplier	Catalogue No.	Lot No.
Alexa Fluor 680-AffiniPure/ Donkey Anti-Rabbit	1:50,000	Stratech	711-625-152-JIR-0.5mg	147633
Alexa Fluor 790-AffiniPure/ Donkey Anti-Mouse	1:50,000	Stratech	715-655-150-JIR-0.5mg	146110

wash buffer (1X PBS containing 0.1% Tween-20) using the same timings described previously (Section 2.4.1.4). Secondary antibody (HRP-conjugated Affinipure Goat Anti-Mouse IgG (H+L), Proteintech, Catalogue No: SA00001-1) was diluted to a 1:10,000 concentration in wash buffer and applied to the blots for 1 hour at room temperature with gentle agitation. Subsequently, the blots were washed again as before.

The blot was treated with Amersham ECL Western Blotting Detection Reagent (Cytiva Life Sciences, Catalogue No: RPN2109) for 3 minutes. Post incubation, the blot was then imaged on a G:Box Chemi XRQ (SynGene) using the associated

GeneSys software.

2.4.3 In-gel Fluorescence Assays

Halo-TMR ligand (Promega, Catalogue No: G8251, Lot No: 0000400971) was applied to cell lines as outlined in Section 2.3.4. After unbound Halo-TMR was washed out, cells were lysed (Section 2.4.1.1) taking care to minimise exposure to light. 50 - 100 µg cell lysate was then run on an SDS-PAGE gel (Section 2.4.1.2) with the EZ-Run Prestained protein ladder (Thermo Fisher Scientific, Catalogue No: 10638393) as standard. To prevent ligand photobleaching, the electrophoresis apparatus was placed under a box for the duration of the protocol. Once the gel front has reached the bottom of the acrylamide gel, the stacking gel and dye front were removed and the gel imaged directly on the LI-COR Odyssey XF Imaging System (LI-COR Biosciences). Halo-TMR was visualised using the 600 nm channel and the ladder was identified using the 700 nm channel.

2.5 RNA Biology

2.5.1 RNA Extraction from Cell Lines

2.5.1.1 Preparing Cells for RNA Extraction

To prepare suspension cells for RNA extraction, cells were pelleted using a suitable speed for the cell line and washed in PBS before the PBS was aspirated entirely. Cell pellets were then flash frozen in liquid nitrogen for a few seconds to allow for long-term storage before extraction. Similarly, surface-adhered cells were placed on ice and the media removed. Cells were washed twice with PBS, before being scraped into PBS and pelleted. Subsequently, the PBS was removed and the cell pellet is flash frozen until use. The numbers of cells required for extraction using TRI Reagent (Sigma-Aldrich, Catalogue No: T9424) varies by cell line, with PC-12 and cortical neuron extractions using approximately $6x10^6$ and $10\text{-}15x10^6$ cells

respectively.

2.5.1.2 Trizol/Chloroform RNA Extraction

To extract the RNA, cell pellets were thawed by hand for several minutes before being resuspended in 750 µL of TRI Reagent. The sample was then homogenised by triturating 10 times before incubation at room temperature for 5 minutes. 150 µL chloroform was then added and the samples were mixed extensively by vortexing for 15 seconds. The sample was then incubated at room temperature for 10-15 minutes and subsequently centrifuged for 15 minutes at 12,000 RCF at 4°C to allow for the separation of RNA from other cell material. The clear, aqueous phase containing RNA was transferred to a new 1.5 mL RNAse-free microcentrifuge tube (Biosphere safeseal tube, Sarstedt, Catalogue No: 72.706.200). 375 µL isopropanol was then added to the aqueous phase and the samples vortexed for 10 seconds. The samples were then incubated at room temperature for 10 minutes. To visualise the RNA pellet more clearly, 1-2 µL of glycogen was added at this step.

Next, the samples were centrifuged for 10 minutes at 12,000 RCF at 4°C and the supernatant was discarded. The RNA precipitate pellet was then washed in 500 µL 75% ethanol and mixed by vortexing. Lastly, the samples were centrifuged at 12,000 RCF for 10 minutes at 4°C and the pellet was air dried at room temperature for 5 minutes. 20 µL nuclease-free H_2O (New England BioLabs Inc, Catalogue No: B1500S) was added to reconstitute the dried RNA pellet. RNA concentrations were determined (Section 2.5.2) before the samples were stored at -80°C.

2.5.2 Quantification and Integrity Analysis of RNA

Immediately after extraction, samples were put on ice and presence of RNA was detected using a NanoDrop ND-1000 Spectrophotometer (Labtech International) on the RNA-40 setting. Integrity and quality analysis of RNA samples was carried out on an Agilent 4200 Tape Station using TapeStation Analysis Software 4.1. This was done using the High Sensitivity RNA ScreenTape kit as per the manufacturer's

instructions.

2.5.3 Polysome Profiling

Sucrose gradients were prepared using 15% and 45% sucrose solutions (Table 2.20). For each gradient, the plastic gradient tube was inserted into the Gradient Master (BioComp Instruments) tube holder, and a line was drawn around the top of the holder using a permanent marker. The 45% sucrose solution was pipetted into the bottom of the tube, until the bottom of the meniscus is aligned with the permanent marker line. 10% sucrose solution was then pipetted into the bottom of the tube using a metal needle. This sucrose solution was added gently (as to not introduce bubbles) until the meniscus of the 10% sucrose solution was aligned with the pen line. Subsequently, the tube cap was applied and any overflow sucrose solution was removed from the interior of the cap. The tubes were then inserted into the tube holder, placed on the Gradient Master platform and mixed using the correct setting (Long Sucrose w/v 10-45%, 1 step, 1 min 48 seconds, 80 degrees, speed 19). Gradients were either used immediately, or alternatively, they can be frozen at -20°C until needed.

8-10x10^6 PC-12 cells were cultured and subsequently pelleted by centrifugation at 400 RCF for 5 minutes. The media was removed and the cells were then resuspended in 1 mL PBS containing 100 µg/mL cycloheximide solution (Sigma-Aldrich, Catalogue No: C4859-1ML, Lot No: 0000182620). The cells were incubated for 2 minutes at 37°C. After incubation, the cells were pelleted again as before and the supernatant was removed. The pellets were then flash frozen in liquid nitrogen and stored at -80°C until required.

Subsequently, pellets were lysed in 400 µL chilled lysis buffer (Table 2.20), pipetting well to mix. Samples were then incubated for 30 minutes at 4°C, before being centrifuged at 10,000 RCF for 5 minutes at 4°C. The supernatant was removed and transferred into a new microcentrifuge tube. RNA concentration was then quantified using the a Nanodrop Spectrophotometer to ensure that an equal amount of

sample was loaded onto each gradient (typically 600 - 1000 µg RNA). The appropriate amount of each sample was applied to the top of a gradient, and then gradients were centrifuged at 40,000 RPM for 2.5 hours at 4°C using a standing centrifuge. The gradients were subsequently fractionated using the Gilson Fraction Collector (Gilson) and TRIAX Full Spectrum Flow Cell (BioComp Instruments), using the associated Flow Cell software (BioComp Instruments). 16 fractions (approximately 0.7 mL each) plus a dead fraction were collected per sample.

Fractions then underwent trichloroacetic acid (TCA) precipitation to concentrate the protein present within the samples. Half of each fraction was precipitated (approximately 300 µL), with the other half being stored at -20°C. 100% TCA solution was added to the sample at a final concentration of 10%. Samples were subsequently incubated on ice for 10 minutes, before being centrifuged at 4°C at 14,000 RPM. The supernatant was then removed, and the pellet washed with 1 mL of ice-cold acetone and vortexed. The samples were centrifuged again as before and the supernatant was removed. Samples were then washed in 1 mL of ice-cold 70% ethanol and vortexed. Finally, samples were centrifuged again as before and the ethanol was gently removed as to not disturb the pellet. The samples were then warmed at 37°C for 30 minutes until the pellets had dried completely. To resuspend the pellet, 10 µL of 1X NuPAGE LDS Sample Buffer (Invitrogen, Cat No: NP0007) was added to the microcentrifuge tubes and samples were incubated at room temperature for 20 minutes. The sample was then mixed by pipetting and stored at -20°C until needed.

Precipitated samples were run on a NuPAGE 4-12% Bis-Tris Midi Gel (Invitrogen, Cat No: WG1402BOX) using NuPAGE MOPS SDS 20X Running Buffer (Invitrogen, Cat No: NP0001). The full quantity of each precipitated sample was loaded alongside 6 µL of PageRuler Prestained Protein Ladder (ThermoFisher, Cat No: 26616). The gel was run at 80V until the dye front had entered the resolving gel, at which point the voltage was increased to 130V. When the dye front reached the bottom of the gel, the gel was removed from the cassette and used immediately in the transfer process. A transfer sandwich was constructed in the same manner outlined

in Section 2.4.1.3 using Amersham Protran nitrocellulose membrane (Sigma Aldrich, Catalogue No: GE10600002). The tank was filled with 1X NuPAGE Transfer Buffer (Invitrogen, Cat No: NP00061) containing 20% methanol. Membrane transfer was done overnight at 10V and 4°C.

Subsequently, the membranes were stained with ponceau stain and the excess membrane removed with a scalpel. Blots were then de-stained using TBST for several minutes. Following this, the rest of the western blot was conducted using the standard fluorescent western blot protocol (Section 2.4.1.4).

Table 2.20: Polysome Profiling Solution Compositions.

Component	Final Conc.	Supplier (Catalogue No.)
Lysis Buffer		
Tris-HCl pH 7.5	20mM	-
NaCl	150 mM	-
$MgCl_2$	15mM	-
Glycerol	8%	-
DTT	1 mM	-
100 mg/mL Cycloheximide	100 µg/mL	Sigma-Aldrich (C4859-1ML)
Protease Inhibitor Tablet	1X	Roche (04693159001)
DNAse I	200U/mL	Thermo Fisher (EN0525)
SUPERase-In	200U/mL	Thermo Fisher (AM2694)
10/45% Sucrose Solution		
Tris-HCl pH 7.5	20mM	-
NaCl	150 mM	-
$MgCl_2$	15mM	-
Sucrose	10% or 45%	-
DTT	1 mM	-
100 mg/mL Cycloheximide	100 µg/mL	Sigma-Aldrich (C4859-1ML)

2.6 Next Generation Sequencing

2.6.1 mRNA-Seq Library Preparation

mRNA sequencing was done by Novogene Co. To prepare the RNA samples for sequencing, quantification and integrity analysis was carried out as in Section 2.5.2. RNA samples were quantified as in Section 2.5.2 and samples were used if their

RNA integrity number was above 8.0. In a 1.5 mL RNAse-free microcentrifuge tube (Biosphere safeseal tube, Sarstedt, Catalogue No: 72.706.200), RNA samples were diluted with nuclease-free H_2O (New England BioLabs Inc, Catalogue No: B1500S) to a 800 ng concentration in 20 µL final volume. Samples were sent to Novogene and paired-end 150 bp eukaryotic RNA-seq (with poly(A)-selection and rRNA depletion) was carried out.

2.6.2 mRNA-Seq Read Quality Analysis, Alignment, and Expression Analysis

Quality control analysis of raw data was done using the Galaxy Europe platform (version 21.09) (Galaxy Community 2022). All analysis was done using the integrated tools within the Galaxy interface. Quality control was done using FastQC (version 0.73), where multiple reports were aggregated using MultiQC (version 1.11). Alignment was done using HISAT2 (version 2.2.1) using the Rnor_6.0/rn6 genome (RefSeq assembly accession: GCF_000001895.5) as a reference. Quality assessment was then carried out on the aligned reads using Samtools Stats (version 2.0.4), the data from this was again aggregated into a report using MultiQC. Differential expression analysis was done using Degust (version 4.1.1)(Powell 2019), with enrichment analysis carried out using GO Enrichment Analysis Tool (Ashburner et al. 2000) (The Gene Ontology Consortium 2021).

2.7 GTEx Database Analysis

Gene expression data was provided by Genotype-Tissue Expression (GTEx) Portal, using the GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2) dataset (GTEx Consortium 2020). Collection of relevant data was done using GTEx's API, Swagger. All analysis of gene expression data was done in R (version 4.2.0) using RStudio (Postit Software, version 2022.12.0, Build 353). Code used in this analysis is provided in the Appendix (Figures 8.5 and 8.7).

2.8 Imaging and Microscopy

2.8.1 Acid-washing Coverslips

13 mm coverslips were added to a 50 mL tube and submerged in a 0.25% acetic acid solution. The tube was agitated on a roller overnight and the acetic acid solution was aspirated the following day. Coverslips were then washed 3 times with 100% ethanol before being stored under 100% ethanol until use.

2.8.2 Paraformaldehyde Fixation

A coating solution (4% paraformaldehyde (PFA), 4% sucrose, 1X PBS) was made from 16% stock solution (TAAB, Catalogue No: F017) using 10X PBS as a dilutant. The addition of sucrose to the coating solution was done to assist in fixing cellular projections when imaging neurons or differentiated PC-12 cells. 1 mL of this solution was heated at 37°C while the cells were removed from the incubator and the media aspirated. The coverslips were washed gently with 1X PBS before the 4% PFA was applied to the coverslips and incubated at room temperature for 10 minutes. After fixation, the PFA was removed and the coverslips again washed in 1X PBS before use.

2.8.3 Immunofluorescence Assays

After the PFA fixation protocol, coverslips were washed by dipping the coverslips sequentially in 5 tubes containing approximately 30 mL 1X PBS using tweezers. The coverslips were then treated with permeabilisation solution (Table 2.21), floated cell-side down on a 100 µL drop of solution for 5 minutes at room temperature. Subsequently, coverslips were washed again as before, then the coverslips were floated on a 100 µL drop of block solution (Table 2.21) for 1 hour at room temperature.

After removal of the block solution, the desired primary antibody (Table 2.22) was diluted in block solution and coverslips were floated on a 50 µL drop of primary

Table 2.21: Immunofluorescence Assay Solutions.

Solution	Component	Final Concentration
Block Solution (15 mL	FBS BSA 10X PBS ddH_2O	5% 1% 1X Up to 15 mL
Permeabilisation Solution (5 mL)	20% Triton 10X PBS ddH_2O	0.1% 1X Up to 5 mL

antibody solution in the dark for 1 - 2 hours at room temperature. If a better signal to noise ratio was desired, primary incubation was done overnight at 4°C in the dark. Care was taken to ensure the coverslips were kept in a humid environment to prevent evaporation of the primary antibody solution. This was done by placing damp blue roll into the bottom of a plastic box (an empty tip box is suitable), putting the coverslips on a piece of plastic above the blue roll, and sealing the box using tape. After primary incubation, coverslips were washed in PBS. The secondary antibody (Table 2.23) was then diluted in block solution to the desired concentration and coverslips were floated on 100 µL secondary antibody in the dark for 45 minutes - 1 hour at room temperature. After this incubation, the coverslips were again washed as before.

Lastly, cell stains were applied. For 4',6-diamidino-2-phenylindole (DAPI) nuclear stain, NucBlue™ Fixed Cell ReadyProbes™ Reagent (Invitrogen, Catalogue No: R37606) was used. To create the working solution, 2 drops of reagent was mixed with 1 mL of 1X PBS as per the manufacturer's recommendation. 50 µL was applied to coverslips and incubated in the dark for 10 minutes at room temperature. Coverslips were then washed in PBS and mounted onto microscope slides using a drop of ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific, Catalogue No:P36961). The slides were left to cure in the dark overnight, before the edge of the coverslips were sealed using clear nail varnish and the slides stored in the dark at 4°C.

Table 2.22: Primary Antibodies used in immunofluorescence assays.

Target / Clonality	Host	Conc.	Supplier	Cat. No.	Lot No.
RPL4/Mono	Mouse	1:100	Proteintech	67028-1-Ig	10009949
RPS2/Mono	Rabbit	1:100	Abcam	ab155961	GR117602-5
HaloTag/Poly	Rabbit	1:500	Promega	G9281	0000460762
HaloTag/Mono	Mouse	1:200	Promega	G921A	0000507057
GFP/Mono	Mouse	1:500	Sigma Aldrich	MAB3580	3018724
GFP/Poly	Chicken	1:200	Insight Biotech.	GTX13970	821802049
β-Actin/Mono	Mouse	1:200	Proteintech	60008-1-Ig	10004413
Tau/Mono	Mouse	1:200	Proteintech	66499-1-Ig	10005161

Table 2.23: Secondary Antibodies Used in Immunofluorescence Assays.

All secondary antibodies for immunofluorescence assays were used at a concentration of 1:500.

Fluorophore/Specificity	Supplier	Cat. No.	Lot No.
Alexa Fluor Plus 488/ Donkey Anti-Mouse	Thermo/Fisher	A32766	WF319853
Alexa Fluor Plus 488/ Goat Anti-Chicken	Invitrogen	15979077	XB343360
Alexa Fluor Plus 594/ Donkey Anti-Rabbit	Invitrogen	15910767	XA336888
Alexa Fluor Plus 647/ Donkey Anti-Mouse	Thermo/Fisher	A32787	XH357261

2.8.4 Confocal Microscopy

Confocal microscopy was done using a TCS SP5 confocal (Leica Microsystems). Images were taken using the associated LAS-AF software. Lasers were selected based on the cell stain or secondary antibody used; 405 Blue Diode (excitation at 405 nm) for DAPI, Argon (excitation at 488 nm) for Alexa Fluor Plus 488 antibodies, Helium Neon 2 (excitation at 594 nm) for Alexa Fluor Plus 594 antibodies, and Helium Neon 3 (excitation at 633 nm) for Alexa Fluor Plus 647 antibodies. Objectives used include HXC PL APO 63x/1.40-0.60 Oil CS (Leica Microsystems, Catalogue No: 15506188) and HC PL APO 20x/0.70 Dry CS (Leica Microsystems, Catalogue No: 11506513).

Images were analysed using the Fiji distribution of ImageJ (version 1.54b) (Schindelin et al. 2012). For each dataset, an image was randomly selected and channel

brightness and contrast were set to a suitable level which was applied to every image in the set. Where specified, images were also cropped by selecting a 200 px by 200 px square around the region of interest and cropping using the built-in ImageJ tool.

2.8.5 Live Cell Imaging

For cortical neuron live cell imaging, Ibidi µ-Dish 35 mm Glass Bottom imaging dishes (Thistle scientific, Cat no: IB-81158) were treated with 250 mg/mL PLL overnight at 37°C. Neurons were plated at a density of $0.18x10^6$ cells per dish. At DIV 4, neurons were transfected using Lipofectamine 2000 (Section 2.1.7.1), scaling down the quantity of reagents used in proportion to the surface area of the live cell dish. Cells were then cultured for several days depending on the experiment type (2 days for over-expression, 1 week for CRISPR knock-in). On the day of imaging, cells were incubated with the desired ligand (1 hour with 50 nM Halo-TMR, 15 minutes with JFX650-Halo (Janelia)), before the media was replaced. Unbound ligand was washed out for 2+ hours at 37°C. Directly before imaging, the media was replaced with BrainPhys Imaging Optimized Medium (Stemcell Technologies, Cat:05796) containing 2% NeuroCult SM1 Neuronal Supplement (Stemcell Technologies, Cat:05711).

For PC-12 live cell imaging, Ibidi µ-Dish 35 mm Glass Bottom imaging dishes (Thistle scientific, Cat no: IB-81158) were treated with 100 mg/mL PLL overnight at 37°C. The following day, dishes were washed twice with ddH_2O before 2 mL of fresh PC-12 media was added and 7,000 cells were seeded into the dish. Cells were differentiated according to the standard protocol (Section 2.1.3). 6 days post-differentiation, cells were incubated with the desired ligand (Halo-TMR, JFX650-Halo or JF549-Halo (Promega, Cat No: GA1110) for 1 hour, before the media was replaced with fresh differentiation media (Opti-MEM (no phenol red) containing 50 ng/mL NGF and 0.5% FBS). Unbound ligand was washed out for 2+ hours at 37°C. For experiments using LysoTracker Green DND-26 (Invitrogen, Cat No: 11594976), 0.5 nM reagent was added at the bench and incubated for 1 minute at

room temperature. The cells were then gently washed in differentiation media and fresh differentiation media was added.

Live cell imaging was conducted on a custom-built single-molecule scanning TIRF/FRAP imaging system (BONSAI) built by Cairn Research. The setup uses a Ti-2 automated inverted microscope base (Nikon), ASI automated XY stage & pieco-Z, MonoLED brightfield illuminator (Cairn), iLas scanning TIRF/FRAP unit with Cairn dual-collimation optics, Cairn multi-line laserbank (405, 488, 532, 561, and 637), and a Photometrics Prime 95B sCMOS camera. TIRF and Widfield imaging were performed with the corresponding excitation laser and a TIRF Quad cube with the TRF89901-EMv2 filter set from Chroma: Excitation, ZET405/488/561/640x-V2; Dichroic, ZT405/488/561/640rpc-UF2-V2 and Emission filter ZET405/488/561/640m-V2. Additional emission clean up was provided by a Cairn Optospin filterwheel with ET445/58m, ET525/50m, ET595/44m or ET706/95m filters for 405, 488, 561 and 638 laser lines respectively.

All images were taken using MetaMorph acquisition software at 100x magnification using a Nikon CFI Apochromat TIRF 100XC oil, N.A. 1.49, W.D. 0.12 mm oil objective. Lasers were controlled using the Modular software and were selected based on the fluorescent ligand used; 405 for DAPI stain, 488 for GFP fusion proteins or LysoTracker Green stain, 561 for Halo-TMR and JF549-Halo ligands, and 638 for JFX650-Halo ligand. The maximum power for each laser line is 75 mW, 61 mW, 53 mW, and 98 mW for 405, 488, 561 and 638 laser lines respectively. Unless otherwise stated, lasers were used at 10% power and images were captured with 100 ms exposure. Additionally, cortical neuron and PC-12 TIRF acquisition used a near-TIRF excitation laser beam angle (approximately 58°) to allow for better visualisation of processes. Widefield imaging used a excitation laser beam angle of 0°.

2.8.6 Kymograph Creation

Kymographs were created using ImageJ (v.154d). First, a segmented line was drawn across the process and the line width was adjusted to its width. Subsequently, this selection was straightened and resliced with interpolation avoided. A maximum z-projection was then carried out on the video to create a kymograph. To calculate the velocity of transport events, lines were drawn on any clear diagonal lines present on the kymograph. These lines were added to the ROI manager and the width and height of the lines was calculated using the Measure tool. The slope of the lines is equal to the velocity of the particle. Therefore, the velocity of transport events is given by:

$$\textit{Velocity} = (\textit{width} \times \mu\textit{m per pixel}) \,/\, (\textit{height} \times \textit{seconds per frame})$$

Several lines were analysed per kymograph and the mean velocity was calculated. The line measurements used for each kymograph are presented in Table 2.24. On the BONSAI system, 1 pixel = 0.11µm.

Table 2.24: Kymograph Velocity Line Measurements.

Line Number	Width (px)	Height (px)	Velocity (µm/s)
TMR-Halo Labelled RPL4-Halo Cells			
Line 1	65	34	0.88
Line 2	39	15	1.19
Line 3	50	16	1.43
Line 4	44	15	1.34
JFX650-Halo and Lysotracker Labelled RPL4-Halo Cells			
Line 1	69	12	0.63
Line 2	55	18	0.34
Line 3	51	20	0.28
Line 4	49	20	0.27
Line 5	54	12	0.50
Line 6	62	11	0.62
JFX650-Halo Labelled RPS2-Halo Cells			
Line 1	29	41	0.29
Line 2	16	61	0.11
Line 3	28	21	0.54

Chapter 3

Creation of CRISPR Constructs and Cell Lines

3.1 Introduction

3.1.1 Overview of ORANGE CRISPR System

Open Resource for the Application of Neuronal Genome Editing (ORANGE) (Willems et al. 2020) is a novel CRISPR system which provides a flexible framework for tagging endogenous neuronal proteins in primary neurons. First published as a preprint in 2019, ORANGE was designed as a simple knock-in system that generates efficient CRISPR constructs using basic molecular cloning techniques. ORANGE built on the work of previous knock-in neuronal CRISPR systems to increase editing efficiency. In an early 2016 Nature article, Uemura et al. utilised a similar knock-in technique to tag β-actin with EGFP in neurons, relying on homology-directed repair (HDR) which introduces low numbers of base-pair insertions and deletions (commonly referred to as indels) into the site of integration (Uemura et al. 2016). Genome editing using HDR is a more precise method of double-strand break (DSB) repair, but a large amount of protein in neurons will remain untagged as HDR is downregulated in post-mitotic cells. Willems at al. therefore decided to use a more recently established homology-independent targeted integration (HITI) repair sys-

tem, which has been shown to offer increased knock-in efficiency compared to HDR in neurons (Suzuki et al. 2016).

Notably, Willems et al. illustrated that ORANGE delivers quick and highly accurate donor integration within neurons – a key feature of any reliable CRISPR system. Using β3-tubulin-GFP and soluble mCherry expression constructs, they co-transfected hippocampal neurons and measured the length of time until expression of the tagged protein was observed. The authors report that after 4 days post transfection 10.9% ± 0.1% labelling efficiency is achieved. Further validation of donor integration was done with genomic sequencing, showing highly efficient integration for almost every construct tested. A variable amount of base-pair indels were also observed. Despite the possibility for out-of-frame integration occurring, the authors conclude that this appears to be dependent on the target sequence selected and does not result in fluorescent signal generation.

Using this method, the authors generated a library of 38 knock-in constructs tagging endogenous neuronal proteins of interest - including commonly studied cytoskeletal, trafficking and receptor proteins. The GFP tagged β-actin construct (pORANGE GFP-Actb KI #1, Addgene Plasmid #131479, Figure 3.2B) was selected to use as a positive control for this project. As an essential component of the cytoskeleton, β-actin is ubiquitously expressed within eukaryotic cells and produces a strong signal in microscopy assays. Therefore, this construct is a useful positive control for developing a robust pipeline for CRISPR construct validation. The guide RNA sequence information for this construct is provided online (Supplementary Information S13, Willems et al. 2020).

3.1.2 pORANGE Cloning Vector Design and Usage

One of the main advantages of the ORANGE system is its elegant single-plasmid design. Instead of using a traditional two-plasmid approach to CRISPR knock-in, ORANGE utilises a single cloning vector (pORANGE Cloning Template vector, Addgene Plasmid #131471) containing both the SpCas9 enzyme and restriction sites

for target and donor DNA insertion (Figure 3.1A, Figure 3.2A). pORANGE was developed from pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene Plasmid #62988), a precursor plasmid containing many of the basic elements required for a single-plasmid CRISPR system. A schematic of ORANGE CRISPR construct generation from pORANGE is given in Figure 3.1B. Briefly, upon selection of a suitable 20 base-pair target sequence 3' of a NGG protospacer adjacent motif (PAM) site, the forward and reverse oligos are annealed and inserted into pORANGE via BbsI restriction digest and ligation. Similarly, the donor DNA insert containing the desired tag is generated via PCR and inserted using two adjacent HindIII and XhoI restriction sites. Notably, the ORANGE system preserves these restriction sites allowing for easy removal of donor DNA should a different epitope tag be desired.

Upon transfection in primary neuronal cells, the target and PAM sequences direct the SpCas9 enzyme to the desired location within the genome where it creates a DSB within the DNA. The PAM sequences flanking the donor DNA in the pORANGE construct are also recognised and cut by SpCas9. This releases the donor DNA containing the desired tag, which is then inserted into the gene at the desired location. The DSB is repaired via non-homologous end joining (NHEJ) which introduces indels at and near the cutting site. Although this error-prone repair may be deleterious and produce frameshift mutations, neurons without the correct in-frame insertion will not be visible during downstream microscopy assays as a functional fusion protein will not be expressed.

Another desirable feature of the pORANGE cloning vector is the inclusion of the reversed target sequence and PAM in the donor DNA insert - thereby promoting correct insertion of the donor DNA. In the instance the donor DNA is inserted into the genome in the wrong orientation, the target sequence is reconstituted and the donor DNA is removed by the SpCas9 enzyme. Overall, Willems et al. designed ORANGE to be a simple one-plasmid system, easily modifiable to the desired target and optimised for success in primary neurons.

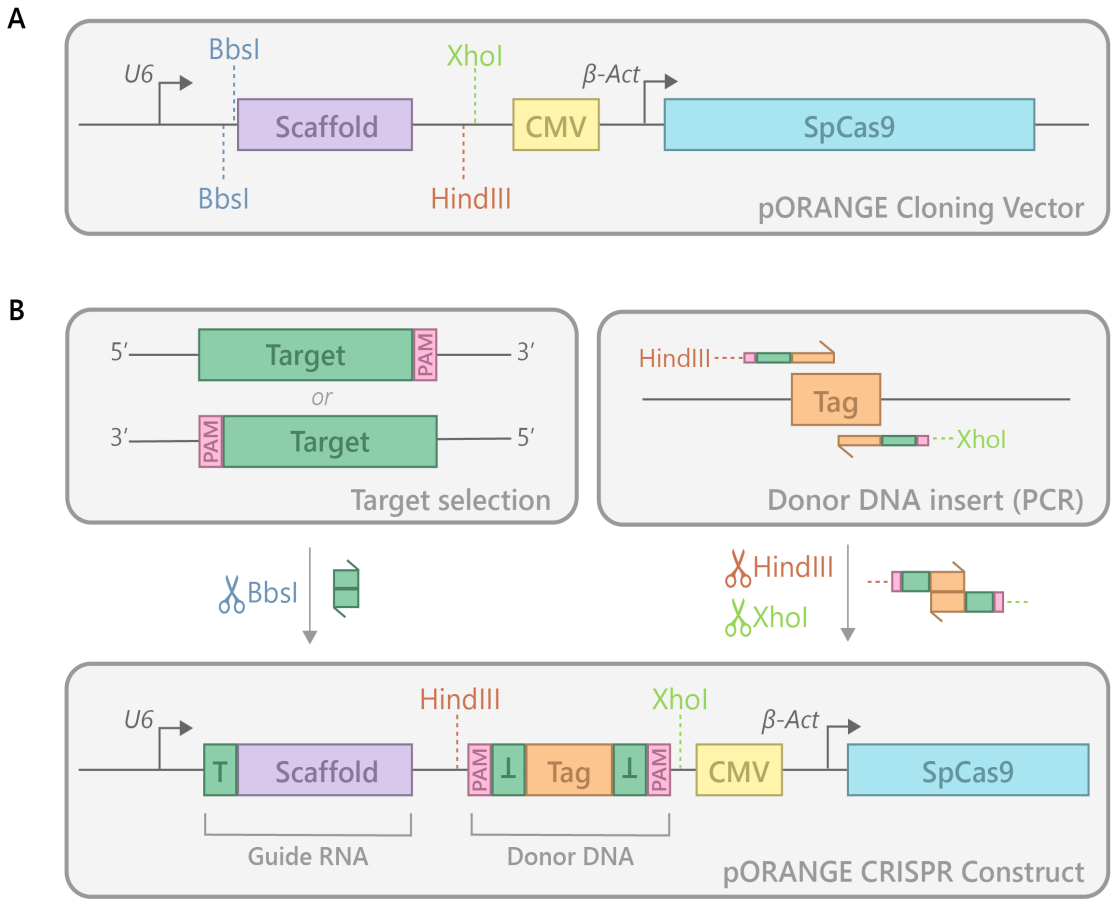


Figure 3.1: Creation of ORANGE Constructs Using pORANGE.

An Overview of the ORANGE CRISPR system. Restriction sites are denoted by a dotted line. A) Basic schematic of the pORANGE features, including relevant restriction sites BbsI, HindIII and XhoI used to create ORANGE CRISPR constructs against the gene of interest. B) Overview of target DNA and donor DNA insertion. A suitable target upstream of an 3' NGG PAM is selected on either the coding or non-coding strand. Foward and reverse oligos are synthesised and the duplex inserted using the BbsI restriction site. The donor DNA is generated using ∼60 base-pair primers via PCR and inserted using HindIII and XhoI restriction sites.

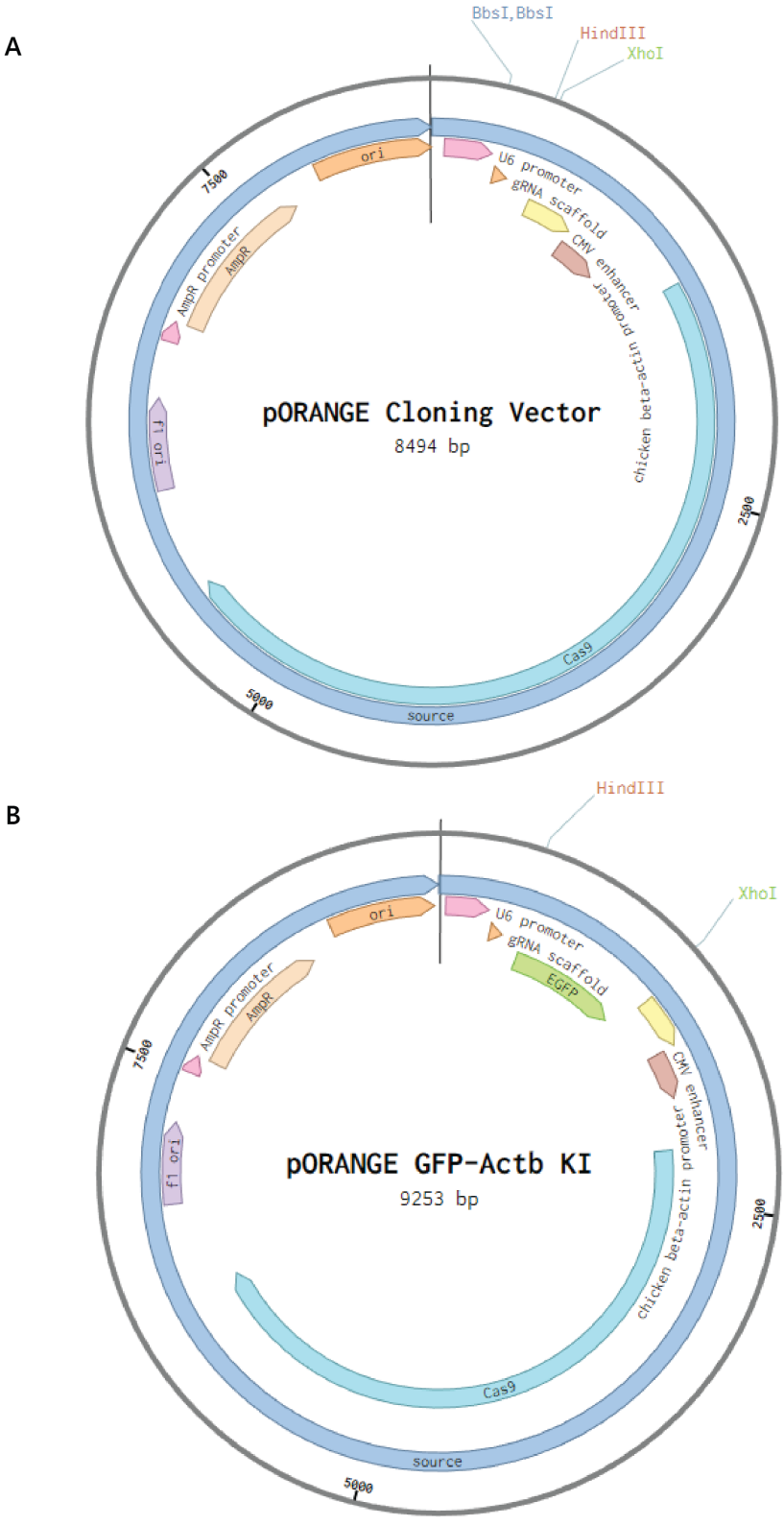


Figure 3.2: Plasmid Maps of ORANGE CRISPR Plasmids.

A) pORANGE Cloning Template Vector (Addgene Plasmid #131471) containing restriction sites for target and donor DNA insertion. B) pORANGE GFP-Actb KI plasmid (Addgene Plasmid #131479) used as a positive control for CRISPR construct validation. Plasmid map was generated using Benchling Biosciences software (2022).

3.1.3 Tagging Ribosomal Proteins With HaloTag

The selection of which donor DNA tag to use took several factors into consideration. Firstly, the tag must be able to be used in live cell imaging as visualisation of tagged ribosomes in mature axons is a primary experimental aim of this project. Therefore, genetically encoded fluorescent proteins (such as Green Fluorescent Protein (GFP)) or self-labelling enzymes were required. However, self-labelling enzymes provide several advantages over the more commonly used genetically encoded fluorescent proteins. Self-labelling enzymes have a wide variety of ligands available, offering more functionality than just fluorescence imaging alone. Commercial fluorescent ligands also have increased photostability and fluorescence intensity, making them a desirable alternative to fluorescent proteins for super-resolution microscopy. Lastly, unlike with fluorescent protein expression, there is no maturation time required before signal can be detected when using self-labelling enzymes and fluorescent ligands (Liss et al. 2015).

It was for these reasons that the self-labelling enzyme, HaloTag (Los et al. 2008), was selected. This 297 residue peptide is derived from a bacterial haloalkane dehalogenase and forms an irreversible covalent bond with synthetic HaloTag ligands (Figure 3.3). Since its development in 2008, a wide range of synthetic ligands have been designed that enable HaloTag to perform a variety of functions beyond fluorescence imaging. These include binding partner capture for protein function analysis, protein immobilisation, and protein purification. The adaptable functionality of HaloTag ensures that the same tag can be used for single molecule imaging of axonal ribosomes and any future proteomic and biochemical assays that require ribosome purification. Additionally, HaloTag has been extensively used in single-molecule microscopy assays including research investigating the translocation of components of the axonal transport machinery (Twelvetrees et al. 2016). Taken together, these considerations make HaloTag a strong choice for the endogenous tagging of ribosomal protein.

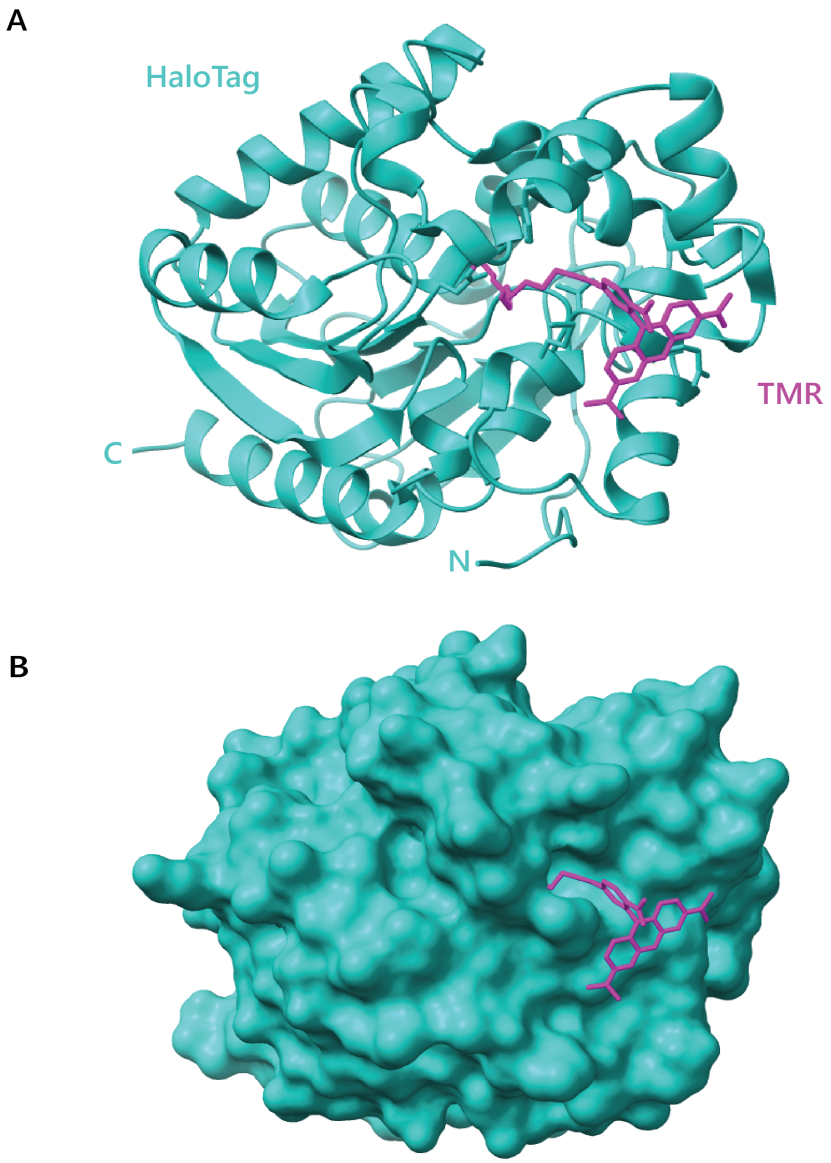


Figure 3.3: Crystal Structure of HaloTag Bound to TMR Ligand.

HaloTag Crystal structure obtained from Protein Data Bank (PDB ID: 6U32, DOI: 10.2210/pdb6U32/pdb) and modelled using UCSF ChimeraX (version 1.4, (Pettersen et al. 2021)). A) Ribbon structure of the HaloTag protein (cyan) with the tetramethylrhodamine (TMR) ligand bound (magenta). B) Surface structure of HaloTag with TMR-Halo ligand bound, highlighting the binding pocket. The HaloTag is approximately 33kDa with a diameter of approximately 58Å (distance modelled using ChimeraX).

3.1.4 Use of the PC-12 Cell Line

Performing validation of CRISPR constructs in primary neurons is challenging due to their post-mitotic nature, short lifespan when compared with immortalized cell lines, and low transfection efficiencies. To this end, the initial CRISPR efficiency assays have been carried out using the PC-12 cell line - an extensively used model system for neuroscience research. The PC-12 cell line originates from rat pheochromocytoma cells from the adrenal medulla, and is commonly used to study neuronal secretion, neurodegeneration, and neuronal inflammation (Wiatrak et al. 2020). An important feature of the PC-12 line that lends itself to neuronal research is its ability to differentiate into cells resembling sympathetic ganglion neurons (Figure 3.4). Differentiation changes the cellular morphology, induces neurite growth, synapse formation, and dopamine synthesis, and the production of neuron-specific biomarkers.

As the ORANGE CRISPR targets were designed for use in mammalian cells, this work will use rat primary neuronal cultures for CRISPR assays. Rat neuron preparations generate a large number of neurons which is beneficial for CRISPR and microscopy assays which require a high quantity of cells. Therefore, CRISPR validation assays also require a rat-derived cell line. Both this and the ability to differentiate make the PC-12 cell line a suitable candidate for early CRISPR efficiency assays to determine if the ORANGE plasmids work as intended.

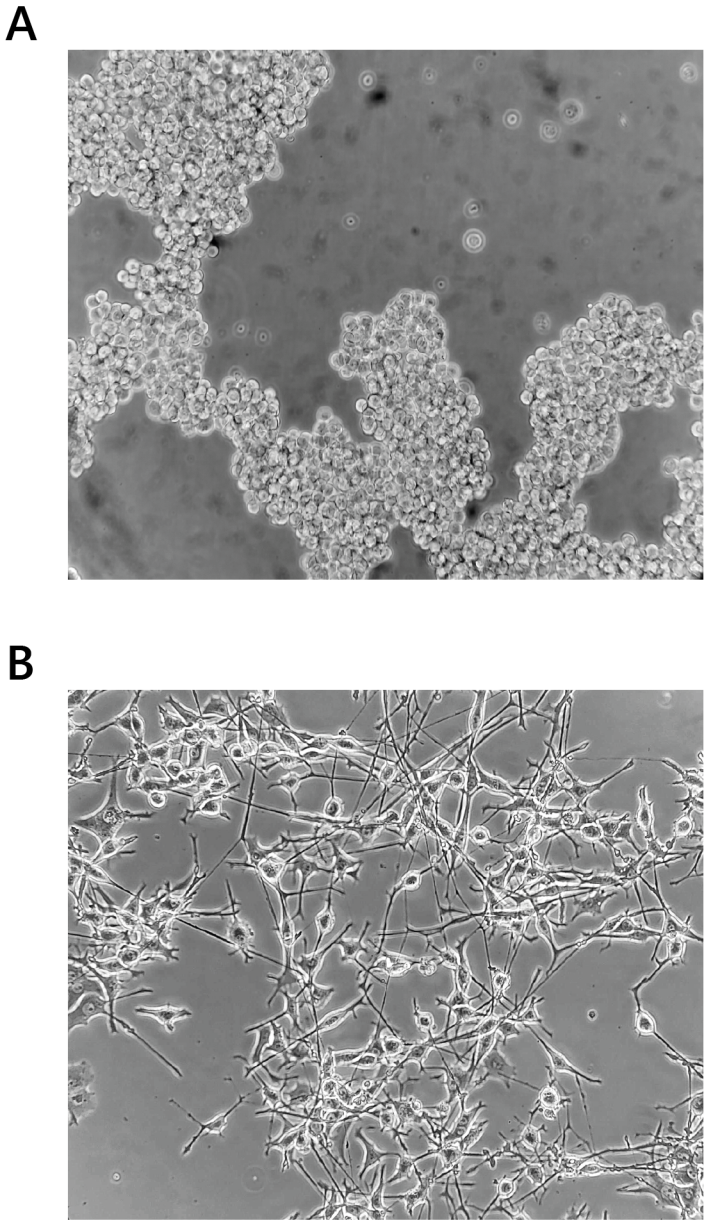


Figure 3.4: Differentiated PC-12 Cells.

A) WT PC-12 cells have a similar morphology to other cells of the adrenal gland. When grown in suspension, the irregularly shaped PC-12 cells form clusters. B) Adhered WT PC-12 cells can be differentiated into sympathetic ganglion neurons following treatment with nerve growth factor (cells presented are 4 days post the start of differentiation). Differentiated PC-12 share many features with mature ganglion neurons, including the development of neurites, functional synapses, and dopamine synthesis. Images were taken using a phone camera on a bright-field microscope at 20X magnification.

3.2 Results

3.2.1 Selection of Ribosomal Protein Tagging Targets

Several factors were considered when selecting ribosomal protein targets that have a higher likelihood of tolerating a HaloTag. As the eukaryotic ribosome is large complex of ∼80 ribosomal proteins and 4 rRNAs, it is essential that tagged ribosomal protein is incorporated into the ribosome early in the biogenesis pathway. This increases the likelihood of visualising the tagged protein being transported, as it is unknown at what stage of maturation the ribosome is in when transported in the axon. As a result, ribosomal proteins that join the nascent ribosome at the nucleolar stage are preferable. Many of the ribosomal proteins that are incorporated in the cytoplasm were therefore excluded. From the remaining candidates, a shortlist of surface localised ribosomal proteins was generated (Table 3.1).

Evaluating the experimental tagging history of the proteins generated insight into whether they would be good CRISPR targets. Tagging ribosomal proteins that have been previously shown to tolerate a tag offers a higher chance of success within this project. As previously discussed, RPL22 has been shown to tolerate epitope tagging with the creation of the Rpl22-HA RiboTag transgenic mouse line (Sanz, Bean, et al. 2019). Similarly, a "RiboTracker" tdTomato-tagged RPL4 transgenic mouse line also provided evidence that RPL4 is a good target which tolerates large genetically-encoded tags (Müller et al. 2018). Most published evidence of endogenous tagging for RPS2 and RPS3 involves the expression of fluorescent proteins (such as GFP) in yeast, drosophila and some mammalian lines (Table 3.1). Given that the protein targets have been shown to tolerate large tags, this provides some evidence that the proteins would be capable of being tagged with HaloTag.

Several other surface-exposed ribosomal proteins that have been successfully tagged were also considered - including RPL10a, RPL36, RPS17, and RPS25. However, these proteins were later determined to be unsuitable due to their lack of efficient guide RNA sequences. The remaining four targets (RPL4, RPL22, RPS2

and RPS3) were selected, ensuring that there were two protein targets for both the large (60S) and small (40S) subunits. This increased the likelihood of tagging both subunits should one protein construct fail to generate the desired CRISPR knock-in.

The tag must be inserted into the C or N terminal domain of the gene as to not interfere with the tertiary structure of the protein. Given that the HaloTag is a large 33kDa tag, CRISPR insertion of this large tag into the coding sequence could cause incorrect folding of the protein, interfere with functionality, or prevent insertion into the ribosome. Care must also be taken to ensure that no frameshift mutations are introduced post-insertion. Therefore, protein targets with surface-exposed N or C terminal regions are desirable as this will increase the likelihood of correct protein folding and function. Ribosomal protein locations were modelled (Figure 3.5) using UCSF ChimeraX (version 1.4, Pettersen et al. 2021) and the N and C terminals were visualised within the 4UG0 80S eukaryotic ribosome (Protein Data Bank DOI: 10.2210/pdb4UG0/pdb, Khatter, Myasnikov, Natchiar, et al. 2015). All proteins were determined to have surface exposed N and C terminal regions, with the exception of RPL4 which has its N terminal buried deep inside the ribosome core.

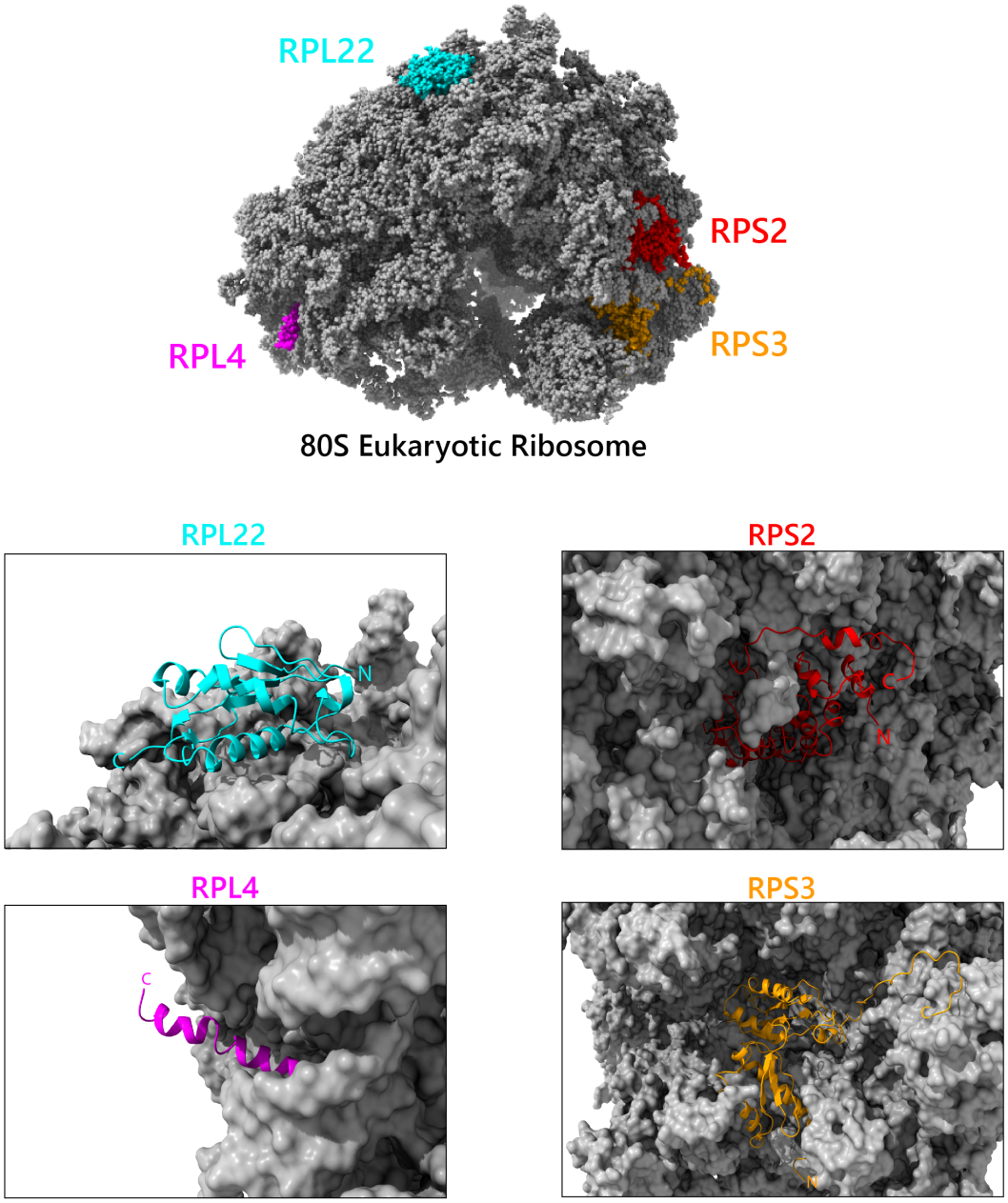


Figure 3.5: CRISPR Protein Targets on Eukaryotic 80S Ribosome.

Positions of the N and C terminal regions of ribosomal proteins RPL4 (magenta), RPL22 (blue), RPS2 (red) and RPS3 (orange) were modelled within the human 80S ribosome complex. All N and C regions were surface exposed, with the exception of the RPL4 N terminal region. Human 80S ribosome structure was obtained from Protein Data Bank (PDB ID: 4UG0, DOI:10.2210/pdb4UG0/pdb) (Khatter, Myasnikov, Natchiar, et al. 2015) and modelling using UCSF ChimeraX (version 1.4, (Pettersen et al. 2021)).

Table 3.1: Ribosomal Protein Biogenesis Stage and Exposed Regions.

Abbreviations: AAV - Adeno-associated virus, BAC - bacterial artificial chromosome, GFP - green fluorescent protein, LV - Lentiviral, RFP - Red fluorescent protein.

Protein	Incorporation	Exposed Term.	Tag	Expression	Model System	References
Selected Ribosomal Proteins for Tagging						
RPL4	Early Nucleolar	C	EGFP	LV overexpression	M. musculus	(Court, Hendriks, et al. 2008)
			tdTomato	Plasmid overexpression	M. musculus	(Müller et al. 2018)
			Flag	Plasmid overexpression	C. elegans	(Nousch 2020)
RPL22	Nucleolar	N, C	HA	KO + overexpression	M. musculus	(Sanz, Yang, et al. 2009)
			HA	AAV overexpression	M. musculus	(Sanz, Bean, et al. 2019)
			Flag	Plasmid overexpression	C. elegans	(Nousch 2020)
RPS2	Medium Nucleolar	N, C	GFP	KO + overexpression	S. cerevisiae	(Milkereit et al. 2003)
			GFP	Plasmid overexpression	D. melanogaster	(Rugjee et al. 2013)
			GFP	Plasmid overexpression	S. cerevisiae	(Black et al. 2019)
RPS3	Late Nucleolar	N, C	GFP	Plasmid overexpression	Jurkat cells	(Jang et al. 2004)
			GFP	KO + overexpression	S. cerevisiae	(Koch et al. 2012)
Unselected Ribosomal Proteins						
RPL10a	Medium Nucleolar	N, C	GFP	BAC overexpression	M. musculus	(Heiman et al. 2008)
RPL36	Late Nucleolar	N, C	RFP	Plasmid overexpression	D. melanogaster	(Rugjee et al. 2013)
RPS17	Late Nucleolar	C	-	-	-	-
RPS25	Late Nucleolar	N, C	SNAP	KO + overexpression	HAP1 cells	(Fuchs et al. 2015)

3.2.2 Design of Guide RNAs Targeting Ribosomal Proteins

Guide RNAs targeting *Rpl4*, *Rpl22*, *Rps2* and *Rps3* were designed using the Knock-In CRISPR Tool by Benchling Biosciences (2020). Two targets per ribosomal protein were selected to maximise the chances of generating a guide that cuts efficiently *in vivo*. First, the ribosome protein gene sequences from the Rnor_6.0 genome were uploaded into Benchling and the CRISPR tool was run using the first and last exons of the gene with the relevant parameters for the ORANGE system (20 nucleotides in length, SpCas9 enzyme, 3'NGG PAM site). This tool generated a list of possible guide RNA sequences with two additional CRISPR scores – the Doench On-target score (Doench et al. 2016) and the Hsu et al. MIT Off-target score (Hsu et al. 2013). The Doench score gives an indication of how well the SpCas9 enzyme will cut at the desired location. In contrast, the MIT off-target score reflects how specific the guide RNA target is when compared with the rest of the genome, and whether this sequence may produce undesirable off-target effects. Benchling provides both of these scores from 0-100, with a higher score indicating a more efficient and specific target.

CRISPR scores were taken into consideration when selecting target sequences (Table 3.2). A robust on-target and off-target score is defined by Benchling as anything above 60 and 50 respectively, therefore, sequences with high scores and close proximity to N or C terminal regions were desirable. Willems et al. published their MIT off-target scores for their constructs, many of which were below 50 (Willems et al. 2020). Despite having low scores these constructs showed high editing specificity, providing evidence that guides with MIT scores below 50 can still generate the desired effect. Additionally, the authors state that they see no correlation between the Doench on-target score and the frequency of integration events for their CRISPR constructs. To this end, protein targets with lower CRISPR scores (e.g. RPS3) were considered if suitably close to the terminal regions. The integration sites for each guide are given in Table 3.3.

Table 3.2: CRISPR Scores for ORANGE Guide RNAs.

CRISPR target sequences and scores were generated using the Benchling Biosciences CRISPR tool. The PAM sites for each target sequence are underlined. The + and - symbols denote the coding (5' - 3') and non-coding (3'-5') DNA strands respectively.

Target	5' - 3' Sequence	Strand Type	Doench (on-target) Score	MIT (off-target) Score
RPL4 Guide 1	GGAGACAAAAAACCTGCAGTAGG	+	58.65	45.56
RPL4 Guide 2	CAAACTTCGAGTGAAAAAGCTGG	+	53.10	65.00
RPL22 Guide 1	GCGCTACTTCCAGATCAACCAGG	+	55.47	54.83
RPL22 Guide 2	CTTCCAGATCAACCAGGACGAGG	+	52.00	78.90
RPS2 Guide 1	CTCTTGTCACTAGCAAATGGCGG	+	66.92	66.66
RPS2 Guide 2	TAGCAAATGGCGGATGACGCCGG	+	59.10	62.46
RPS3 Guide 1	ACTGGCTGAGGCATGGCGGGTGG	-	35.10	56.10
RPS3 Guide 2	GGCACTGGCTGAGGCATGGCGGG	-	32.70	29.40

Table 3.3: Guide RNA Locations and Integration Sites.

Target	Terminal	Guide Location	Integration Site
RPL4 Guide 1	C	Chr8:69126592, Exon 10	A380
RPL4 Guide 2	C	Chr8:69126510, Exon 10	K353
RPL22 Guide 1	C	Chr5:169512768, Exon 4	N118
RPL22 Guide 2	C	Chr5:169512774, Exon 4	D120
RPS2 Guide 1	N	Chr10:14088320, Exon 1	M1
RPS2 Guide 2	N	Chr10:14088330, Exon 1	D4
RPS3 Guide 1	C	Chr1:164436306, Exon 6	P234
RPS3 Guide 2	C	Chr1:164436315, Exon 6	A235

3.2.3 CRISPR Donor DNA Design

To successfully tag proteins with the HaloTag, the CRISPR donor DNA insert was designed following the basic guidelines provided by Willems et al. (Willems et al. 2020). Donor DNA and the necessary primers (see Chapter 2, Table 2.16) were successfully designed for most of the targets. When designing the insert for RPL22 guide 1 it was noticed that the incorporation of the guide RNA sequence before the HaloTag creates an undesired stop codon within the coding sequence. As this is a C terminal tag, this would generate a truncated protein without the HaloTag present. As a result, it was decided that only RPL22 guide 2 would be used for subsequent assays moving forward. The donor DNA inserts for each target were created via PCR (see Chapter 2, Section 2.3.3) and resolved on an agarose gel. See section 3.2.5 for further details on ORANGE construct creation.

3.2.4 Determination of Guide RNA Cutting Efficiency

It is essential that the editing efficiency of the CRISPR guides is validated to ensure that the Cas9 enzyme cuts at the desired location. This validation will be used to determine which guide RNA has the best editing efficiency before cloning the selected guides into the pORANGE cloning vector. It is essential that this validation is carried out using the PC-12 cells so that we have an indication as to whether the CRISPR will be successful before moving on to using the constructs in primary

neurons.

Initial cutting efficiency assays used the PX458 plasmid. As previously discussed, construction of the pORANGE cloning vector used pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid #62988) as a precursor. Similarly, this early guide RNA validation work used pSpCas9(BB)-2A-GFP (PX458, Addgene plasmid #48138, Figure 3.6), a plasmid with similar features to PX459. This plasmid contains the necessary Cas9 enzyme, gRNA scaffold and BbsI restriction sites required for cloning. However, there are no HindIII or XhoI restriction sites present. Instead, the plasmid contains an EGFP reporter gene which acts as a readout of transfection levels. As PX458 allows the insertion of guide RNA but no donor DNA insert, it is a suitable plasmid to use in guide cutting efficiency assays as only the guides' ability to generate Cas9-mediated cutting is analysed. Additionally, the presence of EGFP within the samples was also a useful indicator that transfections worked as expected. The single-stranded gRNA oligos were annealed and inserted into the PX458 plasmid (see Chapter 2, Section 2.3.2.1). This was successfully done using simple restriction cloning and the insertion verified by Sanger sequencing.

In order to assess gRNA cutting efficiency, two methods were used. Firstly, an endonuclease assay was carried out to give a qualitative indicator as to whether Cas9-mediated cutting is achieved. If this assay was successful, Sanger sequencing was then conducted and the results analysed to determine the cutting efficiency. To carry out these assays, the PX458-gRNA plasmids were transfected into PC-12 cells via electroporation. Following transfection of PX458-gRNA plasmids, Cas9-mediated cutting and NHEJ should occur within the genome. Subsequently, the genomic DNA was extracted and a 400 - 600 base-pair region around the cutting site was amplified via PCR (see Chapter 2, Section 2.3.2.2) for use in the cutting assays.

3.2.4.1 Qualitative Editing Analysis Using T7E1 Assay

An indication of gRNA editing efficiency was first determined using a T7 endonuclease I (T7E1) assay. T7E1 is an enzyme that recognises and cuts mismatches in double-stranded DNA. Through denaturing and re-annealing Cas9-edited and WT DNA, a WT/edited DNA heteroduplex is created. If Cas9 directed cutting has occurred, indels will have been generated around the cut site which are not present in the WT sample. Therefore, the heteroduplex will contain base-pair mismatches which are recognised and cut by T7E1. When the sample is resolved on an agarose gel, T7E1 cutting is indicated by the presence of additional non-parental bands in the transfected sample which are not present in the WT sample.

Several controls were used for the T7E1 assay to ensure that the bands seen were due to enzyme activity. First, a WT PC-12 DNA control was set up that had not been transfected with a PX458-gRNA plasmid. This WT sample was otherwise treated in the same manner, including all subsequent heteroduplex formation steps. A negative control was also included for all samples tested. These controls did not have T7E1 added after heteroduplex formation, therefore, no Cas9-mediated cutting occurring should occur. Lastly, all transfections were carried out at the same time on PC-12 cells of the same passage number to ensure that similar levels of transfection are achieved. The results for the T7E1 assay for all guides tested are shown in Figure 3.7.

Both guides for RPL4 were tested, alongside a WT sample as previously described. A clear band for the PCR product can be seen for all samples. Notably, the PCR product for RPL4 is approximately 100bp larger than the other samples. This was due to difficulties in sequencing the PCR product, so the assay was repeated with a forward primer 100bp upstream of the initial primer designed. For RPL4 guide 1, an additional band of approximately half the weight of the PCR product can be seen. Additionally, the approximately 600bp PCR parent band is also dimmer than its negative control counterpart. This is a clear indication that there is less of the uncut product present in this sample. In contrast, no additional bands

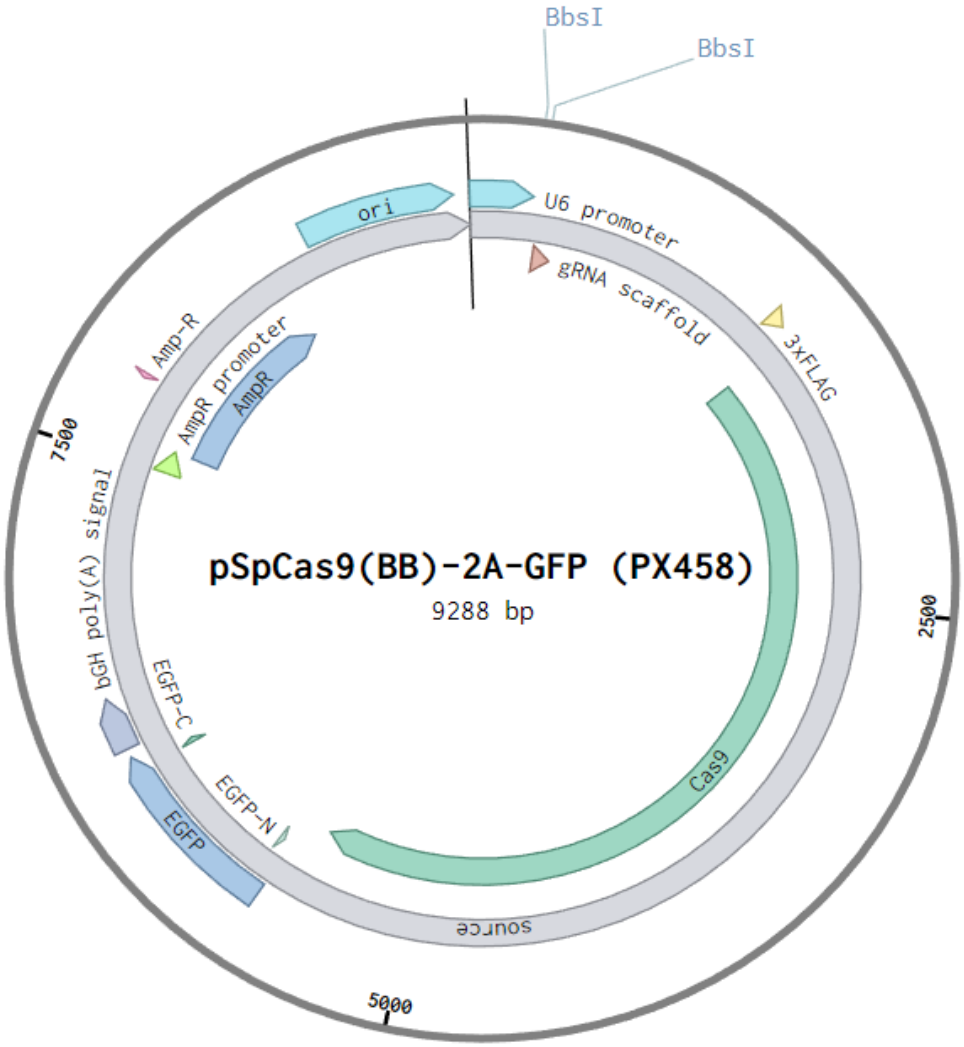


Figure 3.6: PX458 Plasmid Map.

Plasmid map of pSpCas9(BB)-2A-GFP (PX458, Addgene plasmid #48138). Plasmid map was generated using Benchling Biosciences (2022). PX458 is similar to the precursor plasmid to PX459, containing the same SpCas9 and Ampicillin genes, gRNA scaffold and BbsI restriction sites. PX458 contains an EGFP reporter gene and no HindIII / XhoI sites for donor DNA insertion.

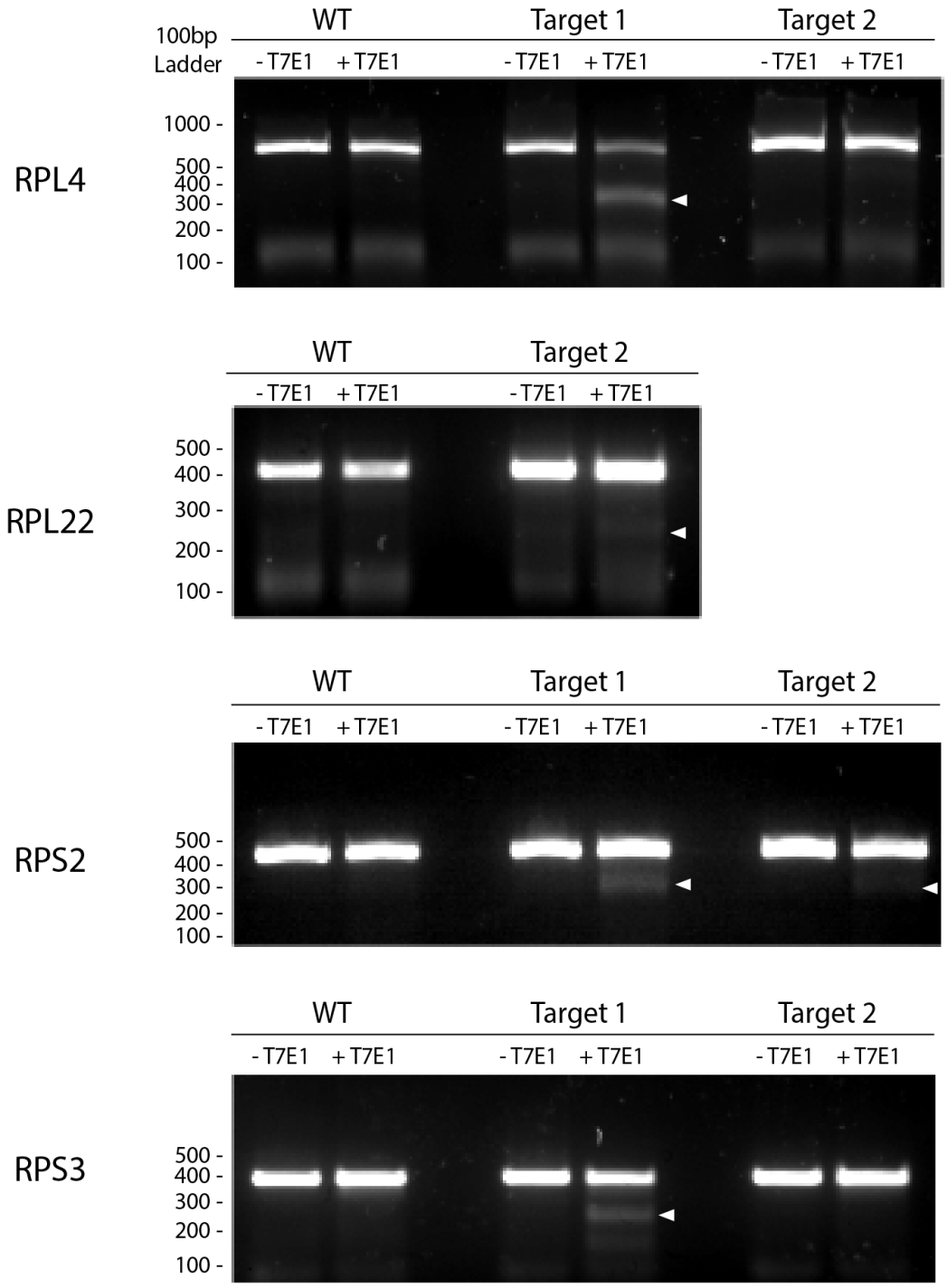


Figure 3.7: T7E1 Assay for Determining Guide RNA Cutting Efficiency.

WT/Cas9-Edited heteroduplex DNA was incubated with T7E1 enzyme, which recognises and cleaves DNA at base-pair mismatches. After incubation, DNA samples with Cas9-mediated editing contain multiple DNA fragments. Additional bands generated by T7E1 cutting are denoted with a white arrow. Samples were run on a 2% agarose gel alongside a 100bp ladder (New England Biosciences, Cat: N3231S).

were seen for guide 2. Overall, this qualitative assay appears to show Cas9-mediated cutting for RPL4 guide 1.

In the case of RPL22, only guide 2 was tested as guide 1 was already proven to be unsuitable due to stop codon generation during donor DNA insertion. A faint band at approximately 250 bp can be seen for guide 2, which is not present in the guide 2 negative control or WT samples. It also appears that the PCR parent band at approximately 450 bp is less intense than the guide 1 negative control band, but the difference is minimal. To this end, guide 2 appeared to be capable of creating Cas9 cutting, albeit at a low efficiency.

Next, both guides for RPS2 were tested. For both guides, faint bands could be seen at approximately half the size of the uncut PCR product. The band for guide 1 was brighter than guide 2, however, the PCR parent bands for both targets appeared to be very similar to their negative control counterparts. Overall, both guides appeared to generate Cas9-mediated cutting, with guide 1 exhibiting a higher efficiency than guide 2.

Lastly, guides for RPS3 were tested, with only guide 1 showing a band half the approximate size of the uncut PCR product. As with the other successful guides, the uncut PCR product band also appeared dimmer compared to the control. No additional bands were seen for guide 2. During this work, it was discovered that the RPS3 guide 2 oligos had been ordered incorrectly. Both RPS3 guides are present on the non-coding (-) strand, and as such require the sequence to be converted to 5' to 3' for synthesis. It was discovered that for guide 2 the associated coding strand sequence had been erroneously ordered. As RPS3 guide 1 showed cutting ability it was decided to continue with the guide due to time constraints around lab access caused by the pandemic.

To determine that the results seen were due to the efficiency of the selected guides, western blots probing for GFP were carried out (Figure 3.8). The PX458 plasmid contains a EGFP reporter gene, therefore, GFP levels in transfected cells should give an indication of transfection levels. This ensures that guides with low

cutting efficiency are due to poor guides and not low levels of transfection. As a control, untransfected WT cells were included which showed no GFP production as expected. Most of the PX458-gRNA samples had similar levels of transfection (Figure 3.8A and 3.8B), apart from RPL4 T2 (guide 2) and RPS3 T2 (guide 2). As previously mentioned, RPS3 guide 2 was determined to be incorrectly cloned and will not be used in future work - therefore, low transfection levels for this sample is not an issue. However, RPL4 T2 (guide 2) has much lower transfection than RPL4 T1 (guide 1). This may be due to a difference in the quality of the DNA preparation, as transfections were carried out at the same time on cells of the same passage number for all samples. As a result, it is possible that the lack of Cas9-mediated cutting for RPL4 guide 2 is due to poor transfection. Despite this, RPL4 guide 1 appeared to create cutting with high efficiency, so it was decided to move forward with guide 1 instead of repeating the T7E1 assays again for all samples. This decision was made to ensure the project progressed during periods of limited lab access, but it is acknowledged that repeating the transfections and T7E1 assays would have been preferable if more time was available.

3.2.4.2 Quantitative Editing Analysis Using ICE Analysis

The T7E1 assay provides a qualitative indication that cutting has occurred *in vivo*, but further quantitative analysis to select the best guides was needed. To do this, the genomic regions containing the cut site were amplified by PCR for all samples (with the exception of RPS3 guide 2) and the resulting PCR products were Sanger sequenced. If Cas9 editing had occurred, indels would be present within the edited samples. The .ab1 file generated from the Sanger sequencing can be used to determine what proportion of the edited sample has indels present - therefore providing an approximate percentage of cutting efficiency. This analysis was done using the Synthego ICE CRISPR Analysis Tool (Synthego Performance Analysis, ICE Analysis. 2019. v3.0. Synthego; [2021-2022]). This tool compares the .ab1 files from the unedited WT control sample and the edited samples, and when supplied with

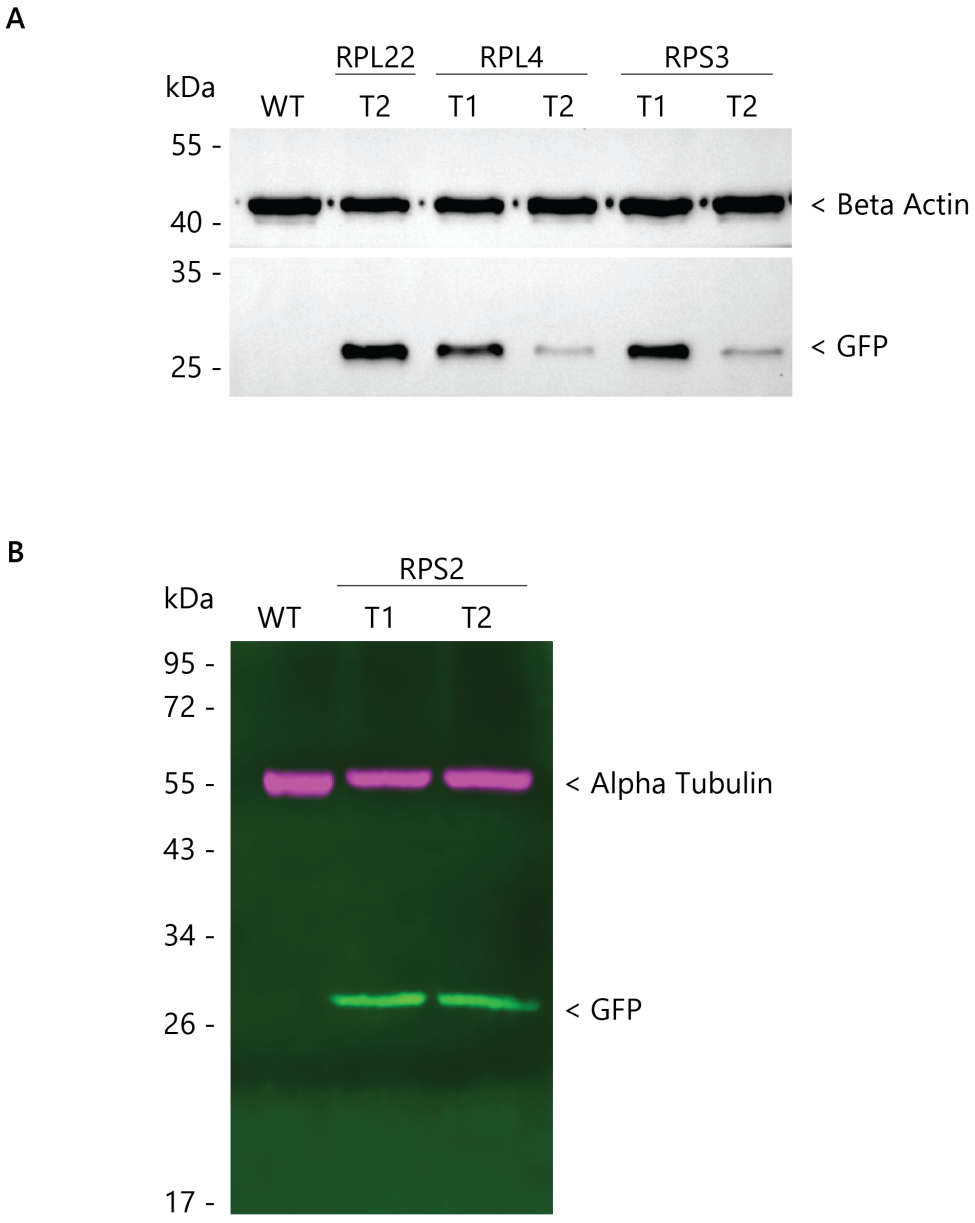


Figure 3.8: GFP Western Blot of T7E1 Samples.

Transfections for the T7E1 assay were done in duplicate. Transfected cells were lysed and analysed by western blot and probed for GFP to determine whether transfection of PX458 plasmids was similar in all samples. A beta actin or alpha tubulin loading control was included to illustrate equal protein loading. Both guides for all targets were tested, denoted by either T1 or T2. A) Chemiluminescent western blot showing the GFP expression for PX458-RPL22-gRNA, PX458-RPL4-gRNA, and PX458-RPS3-gRNA transfected PC-12 cells. This western was carried out in the Thomson lab (School of Biosciences) during the period of Covid-19 lockdown. B) Two-colour fluorescent western showing GFP expression (green) and alpha tubulin (magenta) for PX458-RPS2-gRNA.

the guide RNA sequence and PAM, uses a regression model to calculate the indel percentage present. As an indication of how well the indel data fits with the Sanger sequencing data of the edited sample, a model fit (R^2) value is also provided - ranging from 0 to 1, with 1 indicating a perfect correlation. Details of how these scores are generated is given in Conant et al. 2022.

The ICE CRISPR analysis for RPL4 guide 1 showed editing capability (Figure 3.9), as expected from the clear band in the T7E1 assay (Figure 3.7). The sequencing trace shows small peaks present after the cut site, which aren't present in the WT sample. The occasional small peak can be seen upstream of the cut site in both the edited and WT samples, but this is likely due to poor sequencing quality as a result of Sanger sequencing of short purified PCR products. The indel percentage graph shows that 16% of the edited sample contains an indel of -1 (a single base pair deletion) or +1 (a base pair insertion), as a result of Cas9 cutting. Additionally, the R^2 value is 0.98, an indication that the sequencing data correlates well with the proposed model. Lastly, the discordance graph also shows that the edited and WT samples decrease in similarity after the cut site, with the mean discordance value increasing after the cut site (Table 3.4). Overall, this analysis confirmed the results from the T7E1 analysis, validating that guide 1 creates Cas9-mediated cutting at the desired location with 16% efficiency. Given that not all cells in the sample were transfected, the true cutting efficiency may be higher than the percentage calculated by ICE analysis.

In contrast, RPL4 guide 2 showed no editing efficiency within the ICE analysis (Figure 3.10). The sequencing trace has some small peaks after the cut site, but given their low amplitude when compared to the main peak in the data, it is probable that this is indicative of a very low editing efficiency or sequencing noise. Furthermore, the indel graph illustrates that 99% of the edited sample appeared to have no indels present when compared to the WT. This is reflected in generated editing and R^2 values, which were 0 and 0.9 respectively. The discordance between the samples also decreases slightly after the cut site, which indicates no major changes in the edited

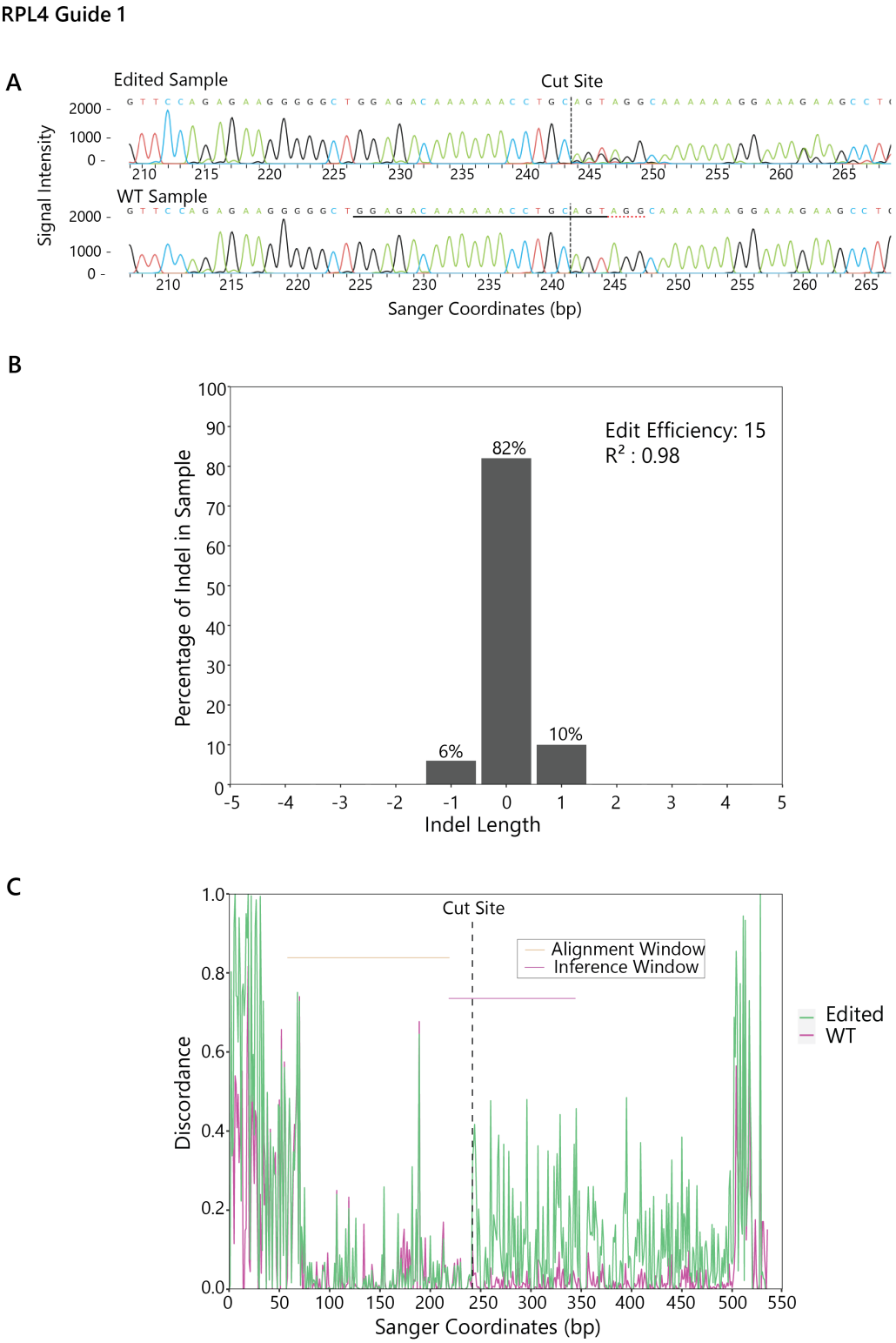


Figure 3.9: RPL4 Guide 1 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

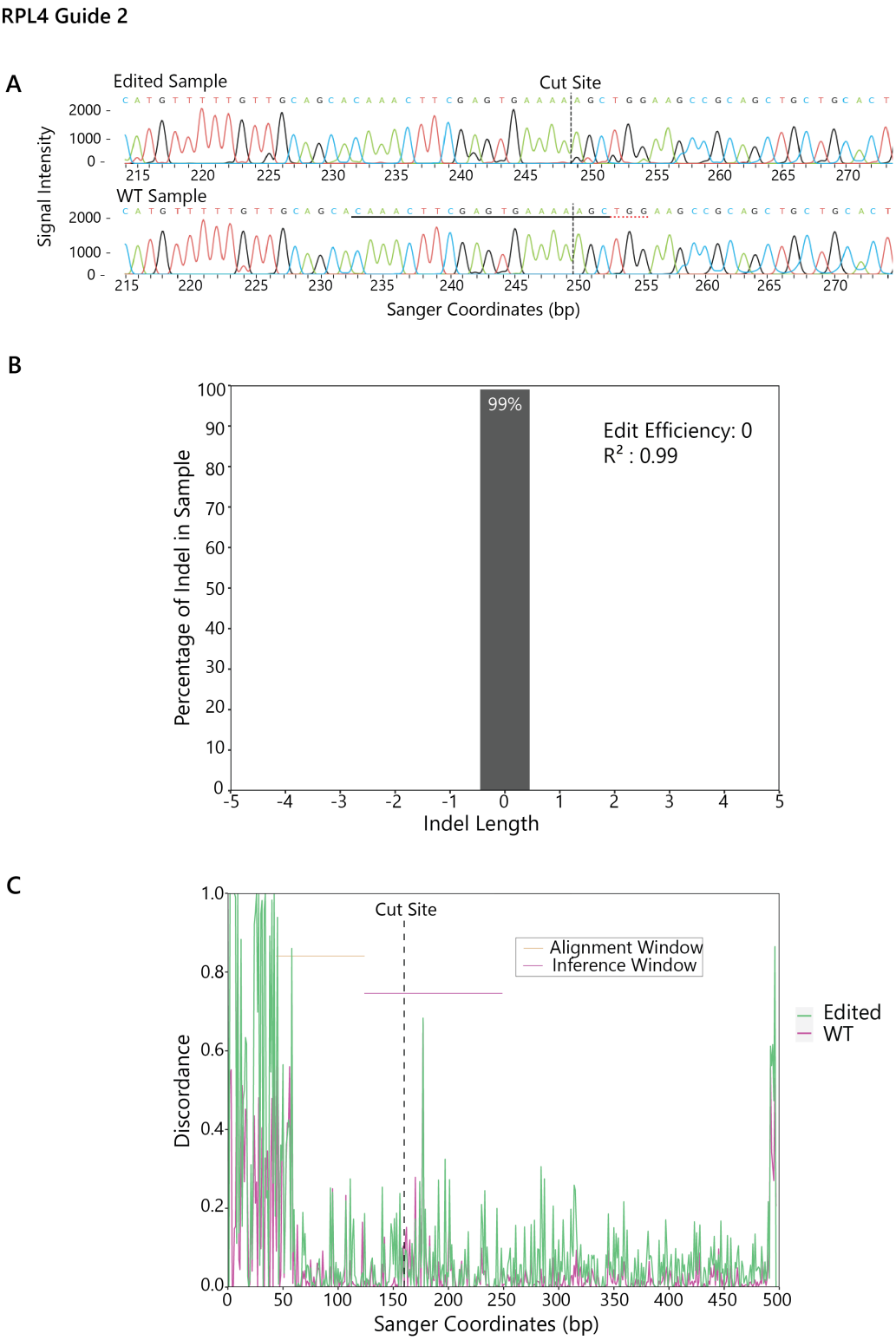


Figure 3.10: RPL4 Guide 2 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

sample as a result of Cas9-editing. This analysis validates the T7E1 asaay which also provided no evidence of Cas9 cutting from RPL4 guide 2. However, the low cutting efficiency observed is likely to be influenced by poor transfection efficiency, as shown in Figure 3.8.

As mentioned previously, the ICE analysis for RPL22 was done using Guide 2 only (Figure 3.11), which showed low levels of editing efficiency. The sequencing trace had small peaks around the cut site for the edited sample, which weren't present in the WT sample. Additionally, the indel graph highlights this small change, with only 5% of the edited sample having a 1 base pair deletion present. The correlation between the sequencing data and the proposed model was robust, with an R^2 value of 0.99. Despite the low levels of editing, the mean discordance value between the edited and WT samples increased notably from 0.051 to 0.245 after the cut site. RPL22 Guide 2 appears to be capable of Cas9-directed cutting at the correct site, albeit at a lower efficiency than RPL4.

RPS2 ICE editing analysis determined that guide 1 generates high editing efficiency (Figure 3.12). The sequencing trace showed some small peaks after the cut site for the edited sample, although signal intensity is low overall (possibly due to poor sequencing quality). The indel graph illustrates that there is the generation of -1 and +1 base pair indels, with an overall editing efficiency of 18%. It must be noted that the R^2 value is 0.97 - slightly lower than seen with other samples, but still above the 0.96 threshold considered adequate for the model to be accurate. Perhaps as another consequence of poor sequencing quality, the edited and WT data aligns less well before the cut site in the discordance trace. However, the mean discordance after the cut site does increase substantially from 0.16 to 0.57, indicating that some editing has caused subsequent changes within the edited sample genome. In summary, RPS2 guide 1 appears to generate indels with a good efficiency, though the quality of the sequencing data for this analysis is decreased compared with other samples.

In contrast, the ICE analysis for RPS2 guide 2 showed no editing present (Figure

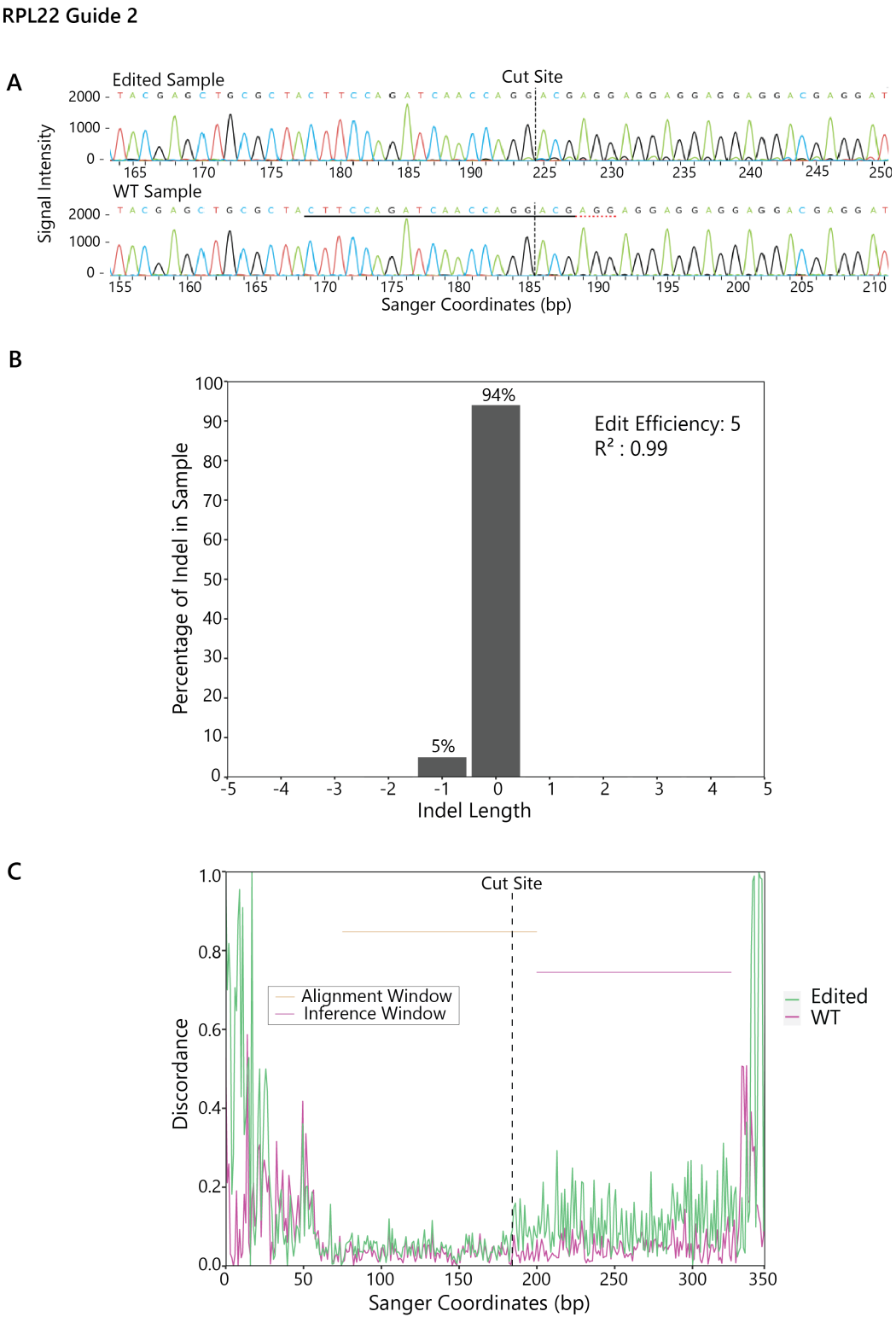


Figure 3.11: RPL22 Guide 2 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

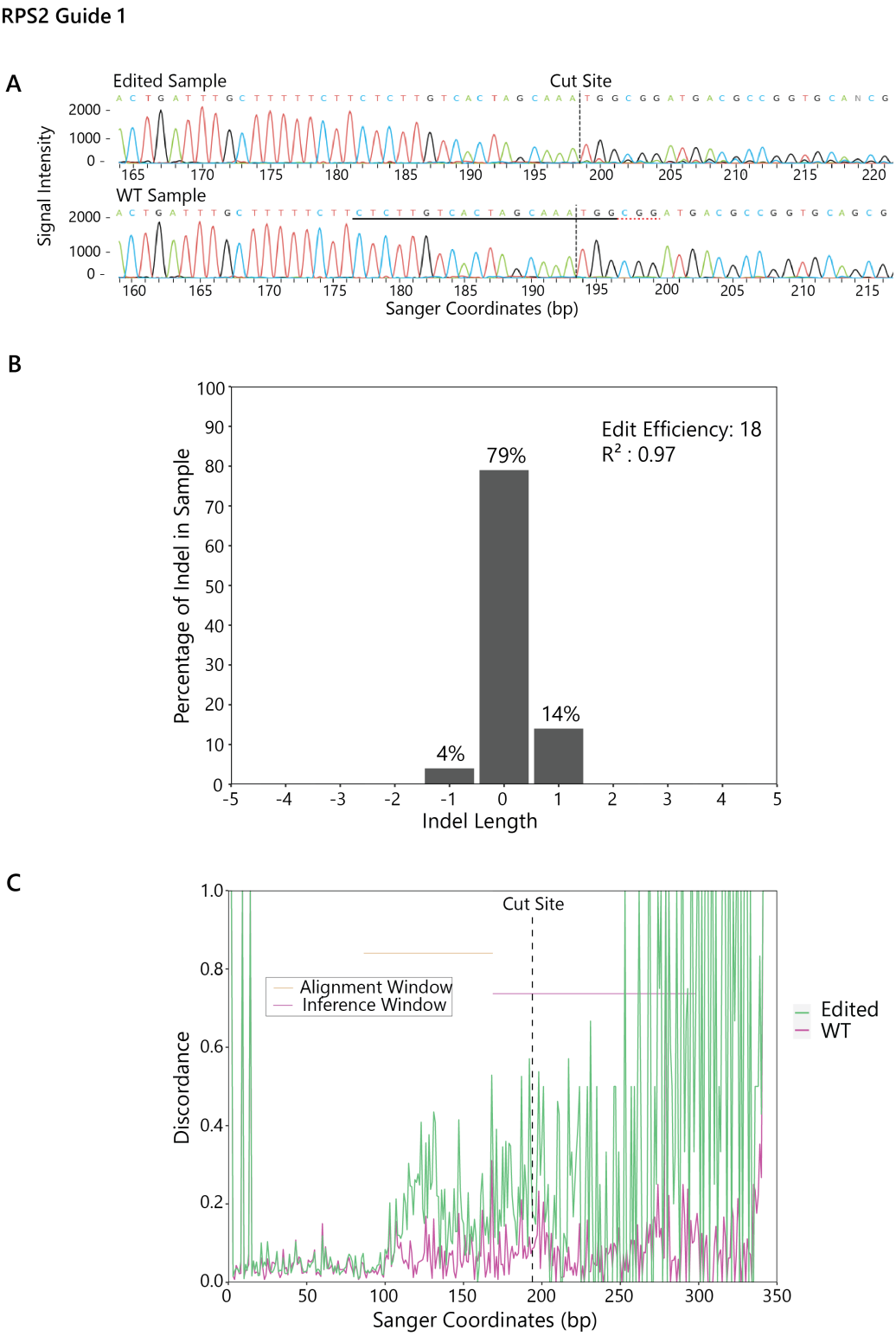


Figure 3.12: RPS2 Guide 1 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

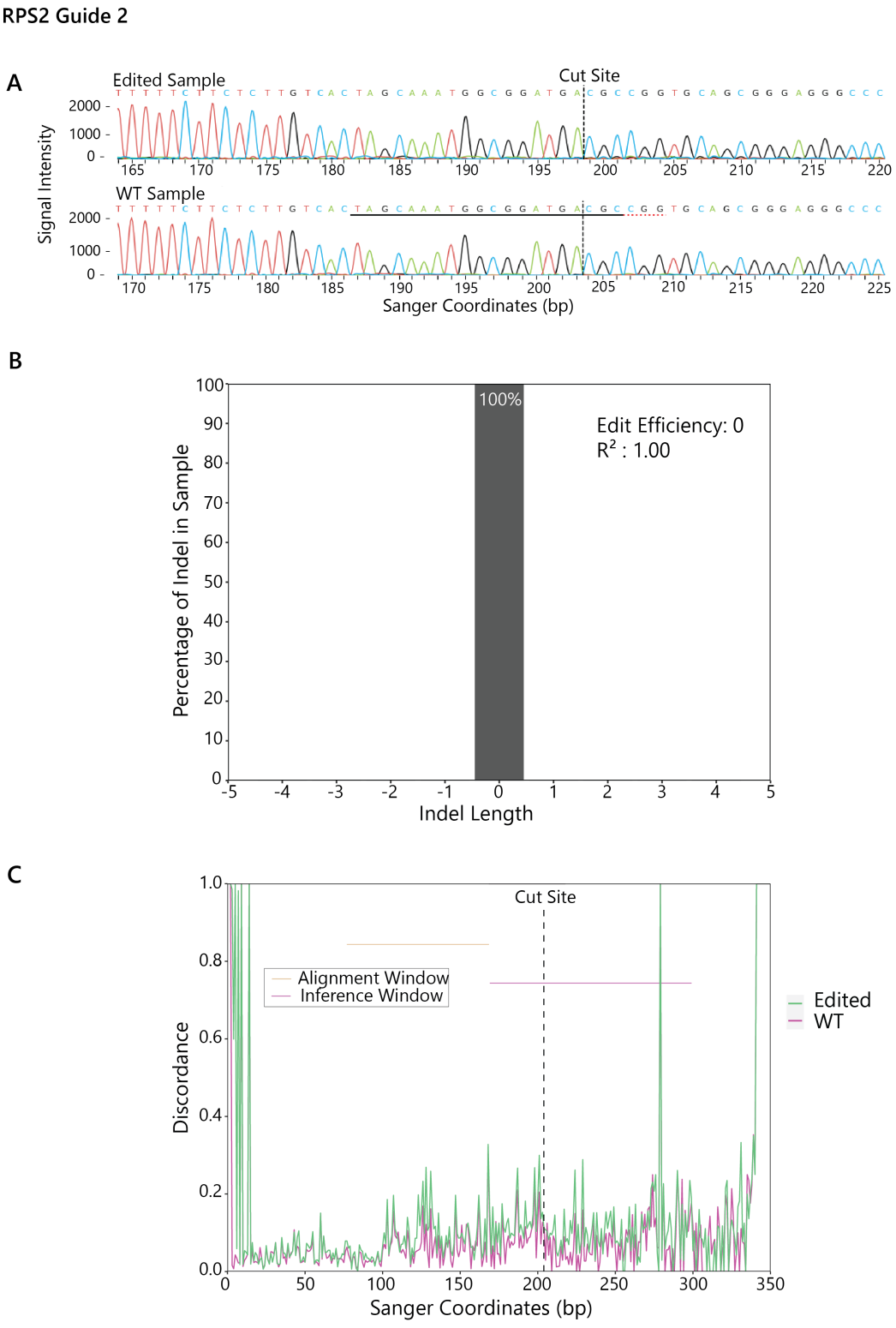


Figure 3.13: RPS2 Guide 2 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

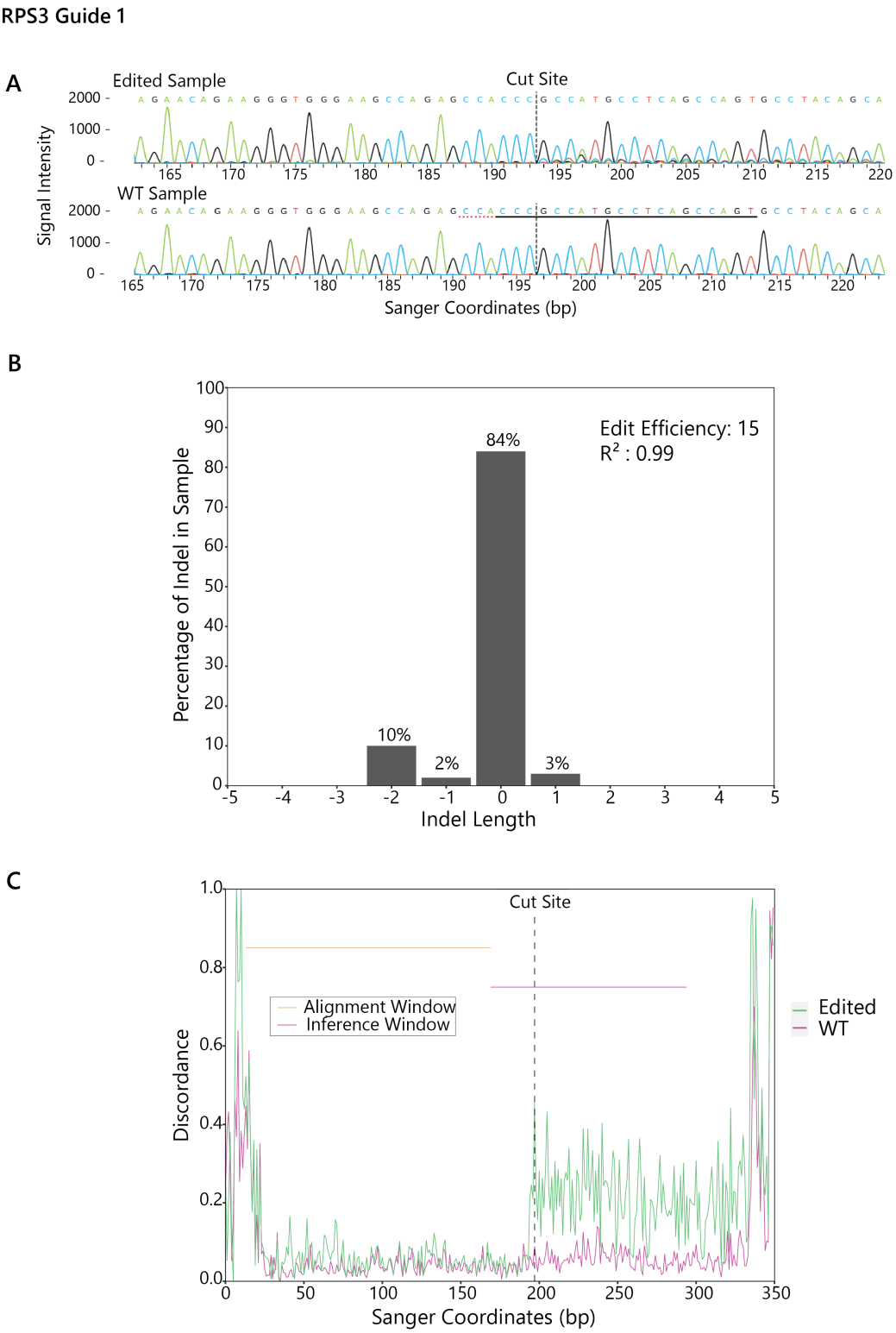


Figure 3.14: RPS3 Guide 1 ICE CRISPR Analysis for Cutting Efficiency.

A) Sequencing trace showing indel formation after the cut site. Guide RNA sequence and PAM site are underlined in black and red respectively. B) Quantification of indel generation in the edited sample. C) Quantification of discordance between WT and edited samples before and after the cut site (denoted by dotted line). Data was generated using Synthego ICE CRISPR Tool (version 3.0) and re-plotted in R (version 2.4.0). N = 1.

3.13). The sequencing trace for the edited and WT samples were almost identical. Although some small peaks could be present around the cut site, they have low signal and could be generated from base miscalls from poor sequencing quality. In addition, the indel graph shows no indel generation - 100% of the edited sample appears to be identical to the WT sequence. Despite the discordance between the samples increasing after the cut site from 0.09 to 0.37, the two traces are very similar and this rise in discordance is likely due to sequencing noise. The results of this ICE analysis were surprising, as the T7E1 assay indicated there were low levels of editing occurring due to the presence of a faint band not present in the WT or control samples. However, it is possible that the levels of cutting were low enough to not be reflected in the sequencing data and therefore not detected in the analysis - this is discussed in detail in Section 3.3.4.

Similar to RPL22, RPS3 only had the one viable guide (guide 1) to analyse (Figure 3.14). The trace showed clear signs of editing with numerous small peaks present after the cut site in the edited sample. The indel graph indicates that the target generates editing of approximately 15% of the sample, with indel length ranging from -2 to +1 base pairs. Overall, the analysis determined this guide has an editing efficiency of 15, with a high R^2 value of 0.99. The discordance trace also provides support in favour of RPS3 guide 1 being an effective target. The mean discordance after the cut site increases from 0.073 to 0.249 - similar to the difference also seen with RPL22. This ICE analysis provides clear evidence that RPS3 guide 1 creates Cas9-mediated cutting at the desired site.

The editing percentages generated by the ICE CRISPR analysis are a helpful indication as to whether the guides worked, and if so, the efficiency which could be expected. However, there are several experimental variables which could influence this analysis - which are discussed in Section 3.3. Additionally, the ICE CRISPR analysis was only performed once due to time constraints, and it is recognised that more biological repeats of this experiment would have been preferable when selecting the guides to be used in downstream cloning. Accordingly, the scores are likely to be

an estimation of the true cutting efficiency. One guide for each protein target was chosen to be cloned into the pORANGE vector, selecting the guides which showed the highest cutting efficiency (Table 3.4). To this end, the following guides were selected; RPL4 Guide 1, RPL22 Guide 2, RPS2 Guide 1 and RPS3 Guide 1.

Table 3.4: Rationale for Guide RNA Selection.

Target	Editing %	R^2	Δ Discordance	Selection Reason
RPL4 Guide 1	16	0.98	+0.06	Highest editing efficiency
RPL4 Guide 2	0	0.99	-0.03	No evidence of guide cutting
RPL22 Guide 1	NA	NA	NA	Payload stop codon incorporated
RPL22 Guide 2	5	0.99	+ 0.19	Highest efficiency, guide 1 unsuitable
RPS2 Guide 1	18	0.97	+0.41	Highest editing efficiency
RPS2 Guide 2	0	1.00	+0.28	Editing efficiency of zero
RPS3 Guide 1	15	0.99	+0.18	Highest efficiency, guide 2 unsuitable
RPS3 Guide 2	NA	NA	NA	Incorrect oligo used

3.2.5 Creation of pORANGE CRISPR Plasmids

After the cutting efficiency of each guide was determined, the selected guides were cloned into the pORANGE cloning vector using the adjacent BbsI restriction sites (see Chapter 2, Section 2.3.2.1). Successful cloning was validated by Sanger sequencing. For each construct, 60bp donor DNA primers (see Chapter 2, Table 2.16) were used to generate the donor insert by PCR. For targets on the coding strand, the forward primers included the HindIII restriction site, the reverse complement of the PAM and guide RNA sequence, a flexible linker (containing a start codon and kozak sequence if necessary), and the start of the HaloTag sequence. For targets on the non-coding strand the primer is similar, including all the same features but without the reverse complement of the guide RNA and PAM. The coding strand reverse

primer contains the last few bases of the HaloTag, a flexible linker (containing a stop codon if necessary), the reverse complement of the guide sequence and PAM, and lastly the XhoI restriction site. Again, the reverse primer for targets on the non-coding strand is the same as for the coding strand, but without the reverse complement of the guide sequence and PAM. A pCMV-K560-Halo-Strep plasmid (created by Emma Turner of the Twelvetrees lab) was used as a template for PCR amplification of the HaloTag. The donor DNA was inserted using the HindIII and XhoI restriction sites (see Chapter 2, Section 2.3.3). Successful cloning was validated by Sanger sequencing.

3.2.6 FACS Analysis and Creation of PC-12 Knock-In Cell Populations

The next stage of the CRISPR validation was to verify that the ORANGE plasmids generated are capable of inserting a HaloTag into the gene. To do this, a Fluorescence Activated Cell Sorting (FACS) assay was set up (Figure 3.15). FACS is a process whereby a population of cells is sorted into sub-populations depending on the fluorescence intensity of the cell. This assay can be used to both check for HaloTag presence and also to collect fluorescent cells to create a CRISPR-edited population of cells for further downstream validation.

To generate a population for FACS sorting, PC-12 cells were transfected with the desired ORANGE-Halo plasmid and incubated for one week to allow CRISPR knock-in to occur. Following this, the cells were treated with tetramethylrhodamine (TMR)-Halo ligand. TMR is a synthetic, cell-permeable, fluorescent ligand that forms a permanent covalent bond with the HaloTag protein. When edited cells are put through the FACS machine, this ligand is excited by the laser and emits fluorescence, which is then picked up by detectors. If the cells haven't been edited, the HaloTag (and by extension, the bound TMR ligand) is not present, and the cells emit no fluorescence. This allows selection of the CRISPR edited cells.

For these assays, three populations of cells were sorted. As a negative control,

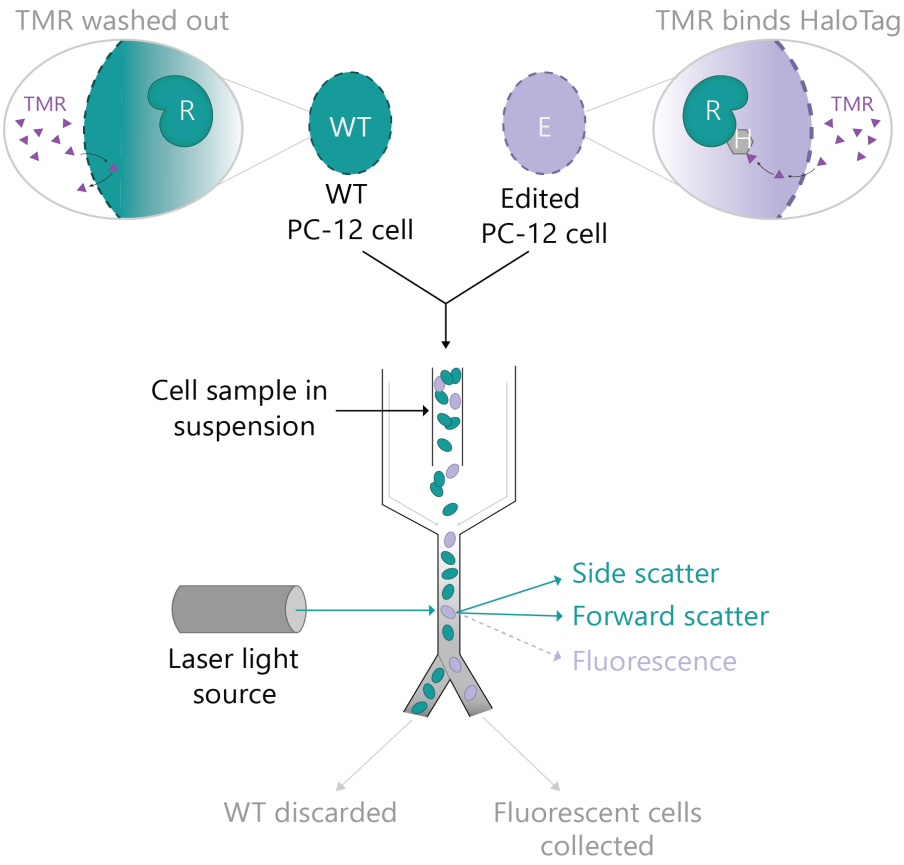


Figure 3.15: Schematic of Cell Sorting Using FACS.

Heterogeneous populations of cells in suspension are inserted into the FACS machine and are focused into a single-cell stream. Single cells pass through the laser light and the light scatter and fluorescence of the individual cell is recorded. After this point, the stream of cells is broken into droplets, with one cell per droplet. If the cell showed fluorescence over the desired intensity, the droplet is charged and deflected into a separate stream for collection.

WT PC-12 cells with no TMR treatment are used to give an indication of baseline fluorescence. In addition, WT cells that have been treated with TMR are used to demonstrate the background labelling from TMR use. These cells are subject to the same incubation times and wash steps as the edited cells, but these cells control for any unbound ligand that remains in the cells. Lastly, the population of transfected and potentially Cas9-edited cells are analysed. Each CRISPR construct was assayed in this manner alongside control samples of the same passage number.

When the cell populations are passed through the cell sorter, the hydrostatic pressure applied forces the cells into a single-cell suspension. This allows the cells to pass through the laser light as a stream, whereby each droplet contains one cell. When the laser hits the cell, forward scatter (FSC) and side scatter (SSC) light is produced. FSC gives information about the size of the cell, whereas SSC gives an indication of the cell's granularity. Given together, FSC and SSC allows us to select living cells for downstream fluorescence detection, excluding dead cells or cellular debris which may be present in the sample. This first gating step is reflected in row A of the following FACS analysis figures (3.16 - 3.19).

Next, the live cell population is gated to select only singlets (i.e single cells) as doublets or cell clumps can generate false positives and lead to the selection of non-edited cells during sorting. This involves gating the live population using information from the SSC height and width (SSC-H and SSC-W respectively). This second gating step is reflected in row B of the FACS Analysis figures (3.16 - 3.19). After the data has been gated for this selection, the information about the cells' fluorescence is recorded. A final gate can be set up to select live cells whose fluorescence is above a certain level - i.e above that of the WT TMR+ treatment sample. This SSC area (SSC-A) against TMR fluorescence intensity gate is reflected in row C of the FACS analysis figures (3.16 - 3.19).

The selected cells are then isolated into a separate stream by deflection plates, collected, and expanded for validation. The validation of these cell lines is presented in Chapter 4. Initial optimisation experiments included a single cell sort, whereby

cells were sorted into a 96 well plate to allow for clonal populations to be validated. This provides the benefit of having the cells be genetically identical, allowing the selection of a homogeneous clonal population where all ribosomal protein is tagged. However, the low density of cells (and therefore cell-derived growth factors) proved to be a limiting factor in the development of PC-12 clonal lines. After several months post sort, the clonal lines numbered only tens of cells, despite regular feeding with conditioned media. This low division rate was coupled with a slightly altered morphology, indicative of cellular stress. Consequently, I decided that the single cell sort strategy wasn't feasible with the PC-12 cell line in a reasonable time frame. As an alternative, all fluorescent cells from the FACS analysis were collected in a single well of a 6 well plate and the heterogeneous line was grown up for further validation. The implications of this are discussed in Section 3.3.

The RPL4 FACS analysis showed a clear population of fluorescent cells (Figure 3.16C). The control WT TMR+ cell population (Figure 3.16B) shows a slight shift to the right compared with the WT cell population (Figure 3.11A), which indicates that there is a small amount of background staining from the TMR ligand. The TMR fluorescence gate was therefore set to slightly above the highest WT TMR+ fluorescence level. The majority of the pORANGE-RPL4-Halo transfected cell population had similar levels of fluorescence to the WT TMR+ cells, however, 0.66% of the population had much higher levels of fluorescence. Most of the 0.66% were clustered around the 10^4 fluorescence level, over 100x higher than the average fluorescence intensity of the WT population.

Similarly, the RPL22 FACS assay also determined that 0.81% of the transfected population had been edited to express the HaloTag (Figure 3.17C). Compared to RPL4, there is a much larger increase of fluorescence intensity in the WT TMR+ population, indicating higher levels of background staining. However, this is not an issue as the gates are individually tailored to each population to ensure that only cells with higher fluorescence intensity than the WT TMR+ population are selected. Although this data indicates that the RPL22-Halo plasmid has similar

efficiency to RPL4-Halo, it must be noted that different numbers of cells may have been transfected in the RPL4-Halo and RPL22-Halo samples.

The RPS2-Halo construct appears to be the most efficiently edited target in the FACS assay, with 3.0% of the S2-Halo TMR+ cells having a fluorescence intensity above that of WT TMR+ (Figure 3.18C). This editing percentage is considerably higher than seen with the other CRISPR constructs, and is unlikely to be due to non-specific background staining, as the WT TMR+ population is similar to that of the WT population. Although it is probable that the S2-Halo construct is more efficient, the use of newly thawed cells at a low passage number in this assay may also be a factor in the higher editing efficiency. As a result of the large number of fluorescent cells, two distinct populations of cells can be visualised - a lower fluorescence population near the red gate line, and a smaller, high fluorescence population. These populations are likely to be heterogeneous and homogeneous edited cells respectively.

Lastly, the RPS3-Halo FACS assay showed a small population of fluorescent cells (Figure 3.19C), albeit at a much lower percentage than seen with the other constructs. Only 0.14% of the transfected cells showed fluorescence intensity above that of WT TMR+ population. It must be noted that the RPS3 and RPL22 FACS assays were carried out at the same time, so the same WT and WT TMR+ samples were used as controls for both assays. The first FSC-A SSC-A gating step is slightly different in the RPL22 and RPS3 analysis to ensure that the maximum number of live cells are included. However, the rest of the gating steps were identical, including the final TMR fluorescence gate.

In each assay, all cells with a fluorescent intensity above that of the WT TMR+ gating threshold were collected. The collected cells were expanded until there were enough cells to generate and propagate a heterozygous HaloTag knock-in cell line for each CRISPR construct. These cell lines then underwent more thorough analysis to determine that the HaloTag is present in the desired location and the tagged protein is incorporated into functional ribosomes (Chapter 4). This preliminary analysis in

the PC-12 cell line provided a valuable opportunity to validate the constructs before use in primary neurons (see Chapter 4, Section 4.1 for an explanation of why this is necessary). Additionally, a pORANGE-GFP-ACTB edited PC-12 line was also generated in a similar manner (Figure 3.20) for use as a positive control for CRISPR validation assays.

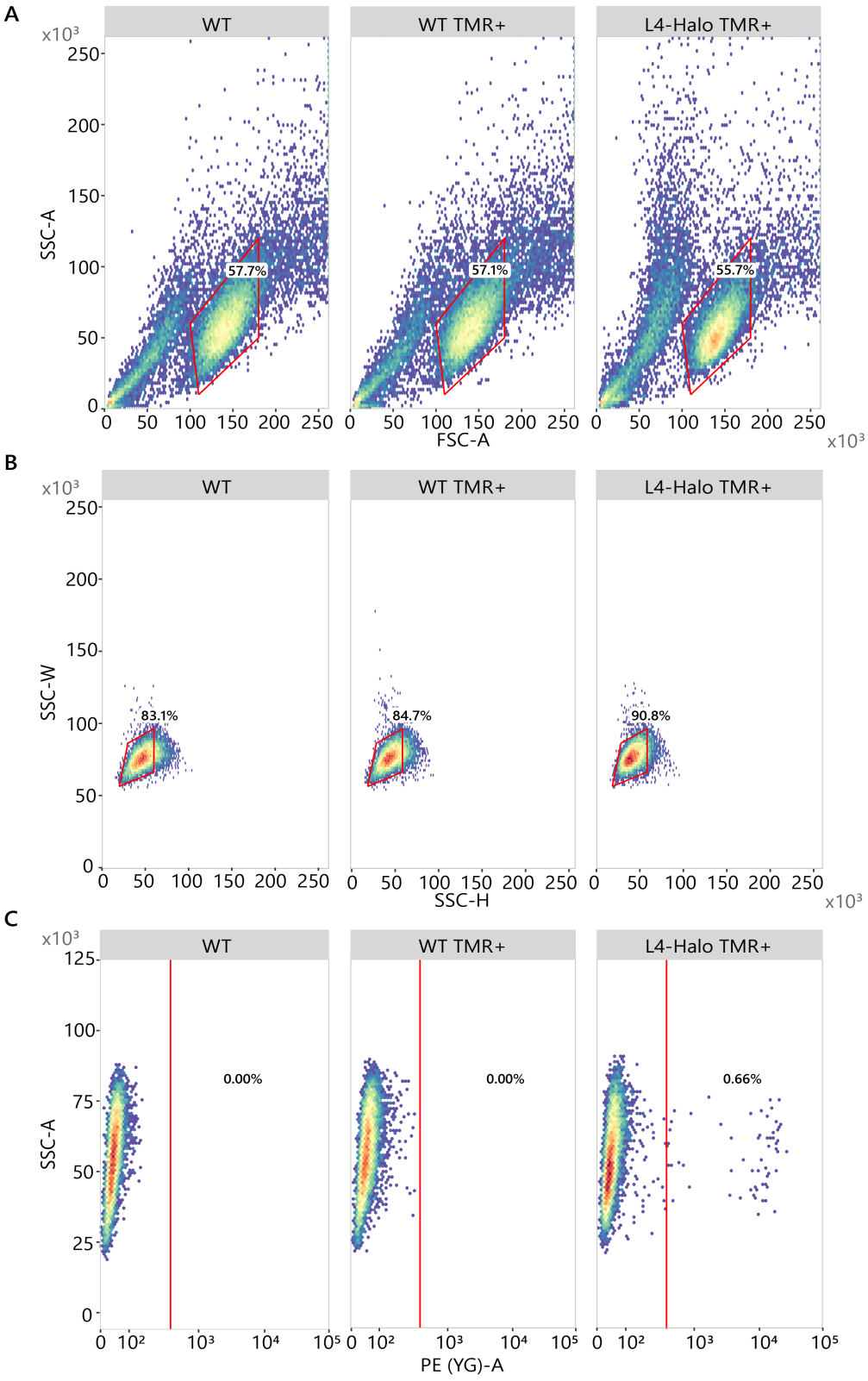


Figure 3.16: FACS Analysis of pORANGE-RPL4-Halo Transfected PC-12 Cells Labelled with Halo-TMR.

Gating (denoted using red lines) used in the FACS assay. Raw data was replotted using R (version 4.2.0). A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlet cells. C) Fluorescence intensity of live singlets.

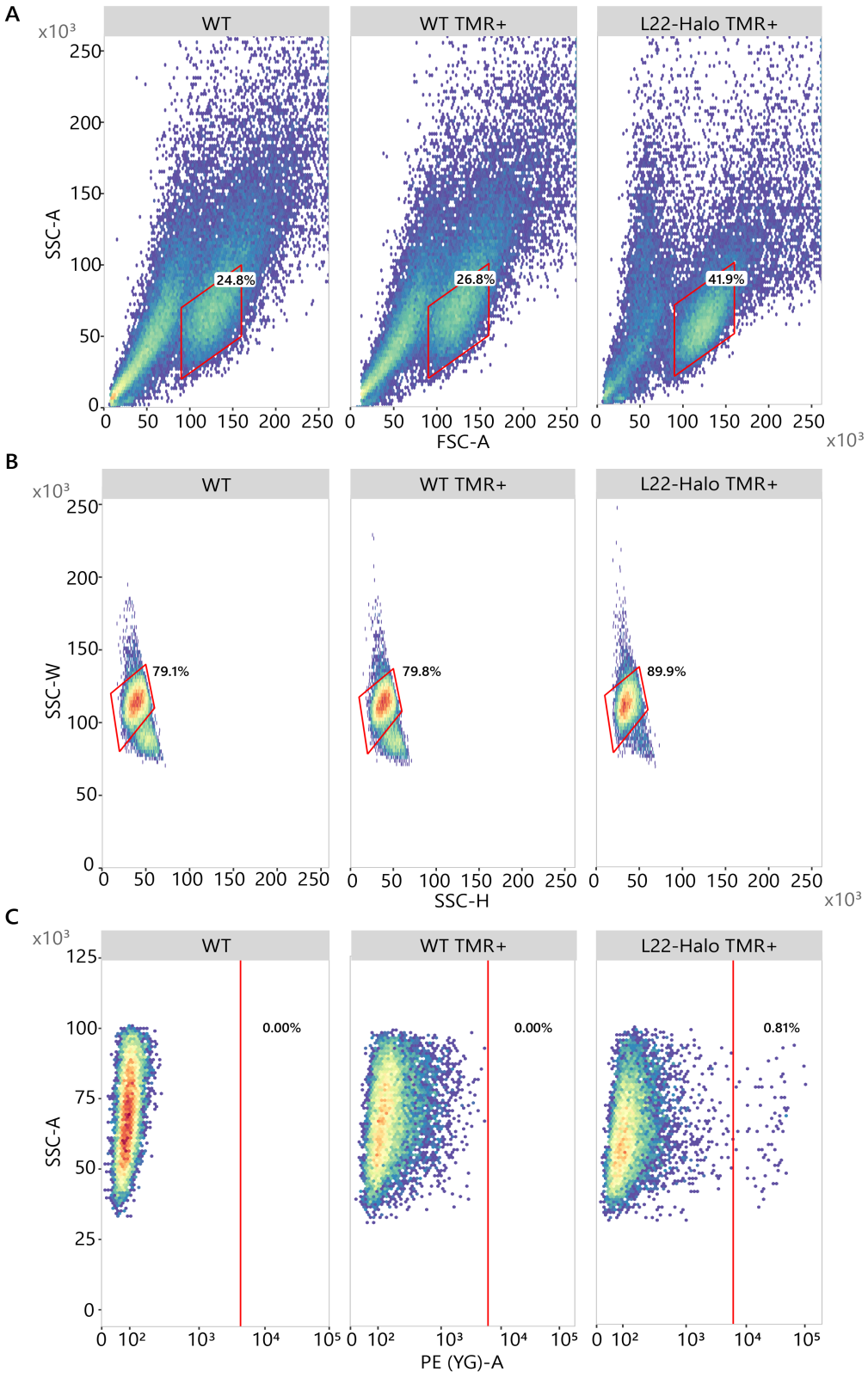


Figure 3.17: FACS Analysis of pORANGE-RPL22-Halo Transfected PC-12 Cells Labelled with Halo-TMR.

Gating (denoted using red lines) used in the FACS assay. Raw data was replotted using R (version 4.2.0). A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlet cells. C) Fluorescence intensity of live singlets.

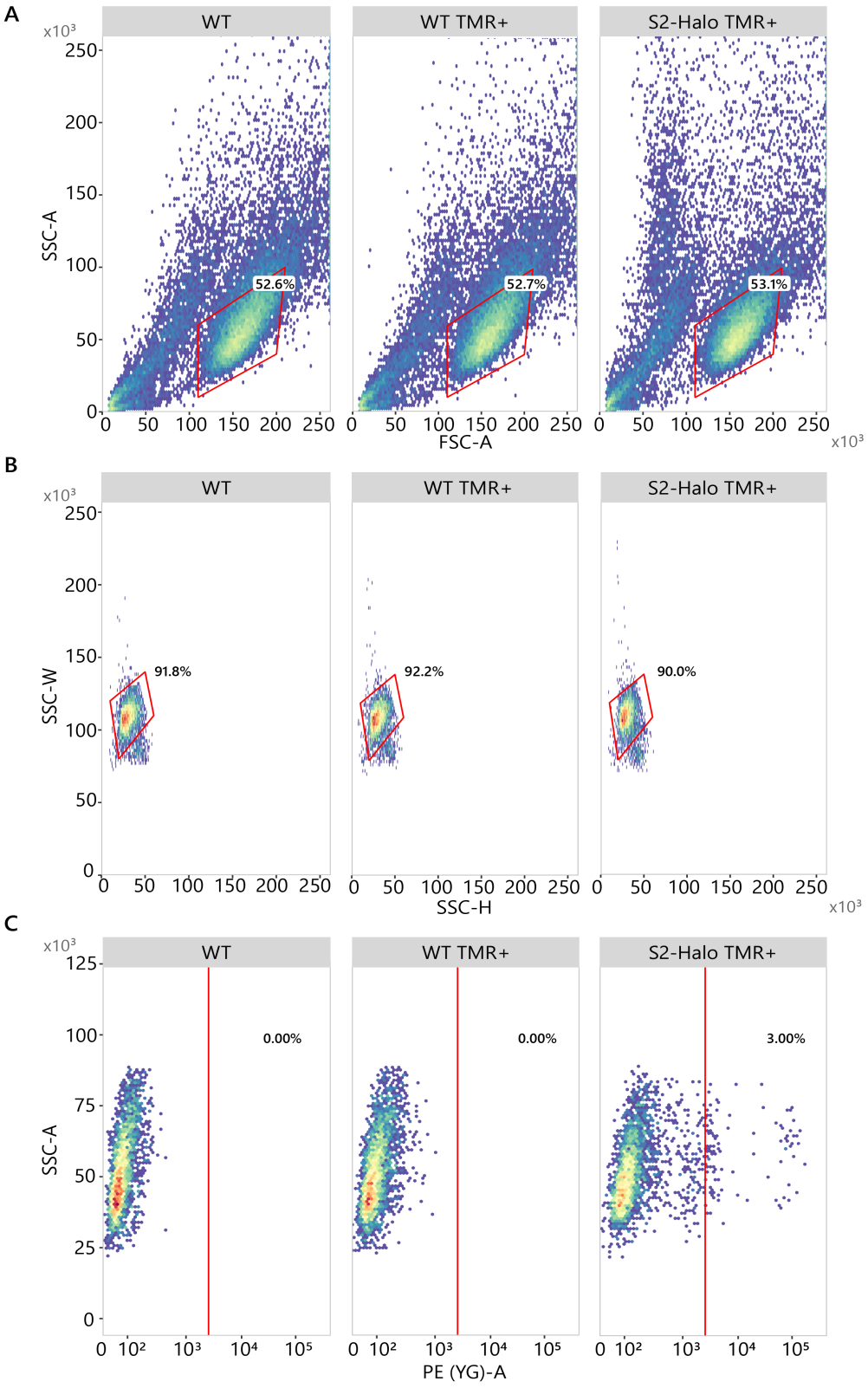


Figure 3.18: FACS Analysis of pORANGE-RPS2-Halo Transfected PC-12 Cells Labelled with Halo-TMR.

Gating (denoted using red lines) used in the FACS assay. Raw data was replotted using R (version 4.2.0). A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlet cells. C) Fluorescence intensity of live singlets.

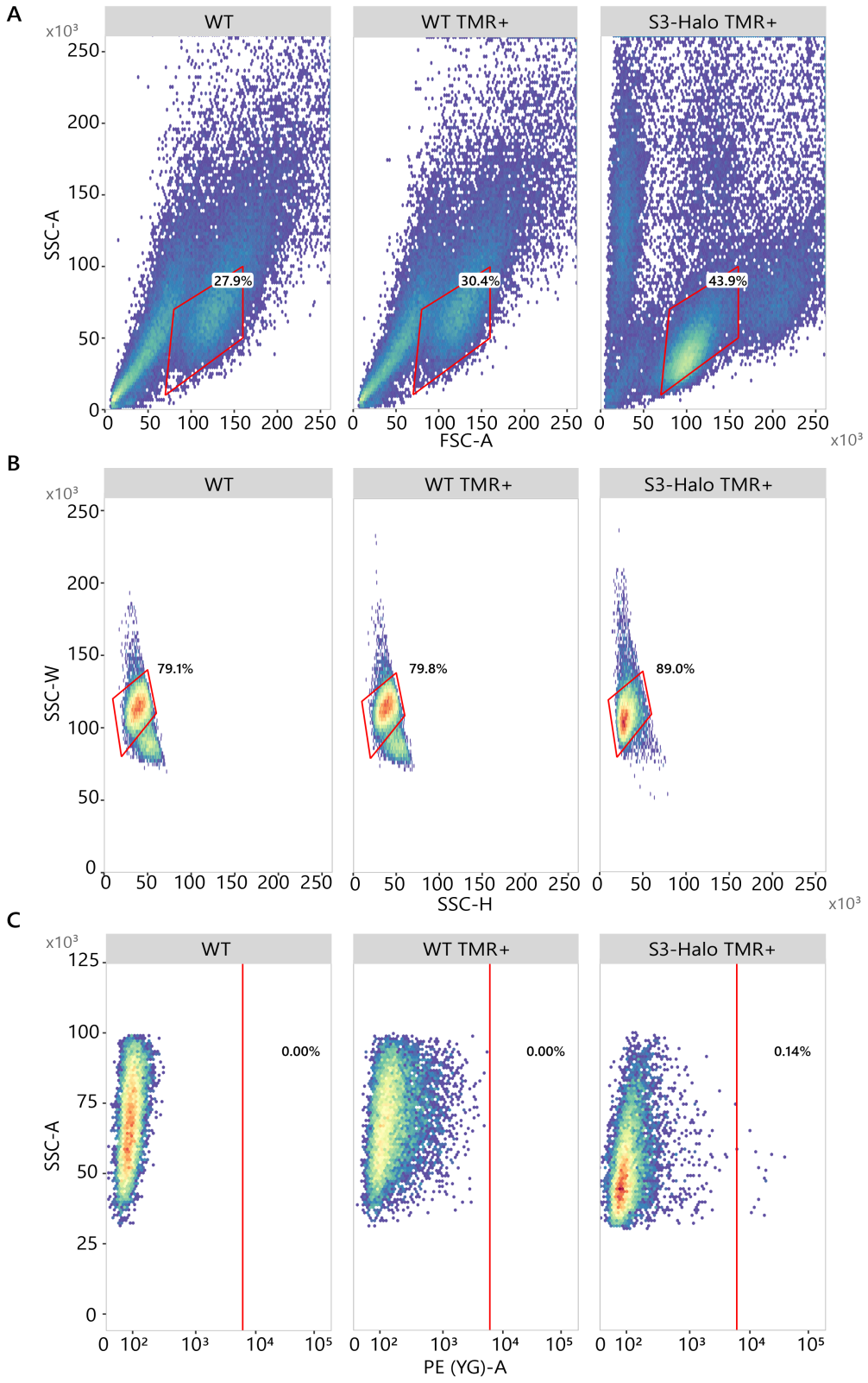


Figure 3.19: FACS Analysis of pORANGE-RPS3-Halo Transfected PC-12 Cells Labelled with Halo-TMR.

Gating (denoted using red lines) used in the FACS assay. Raw data was replotted using R (version 4.2.0). A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlet cells. C) Fluorescence intensity of live singlets.

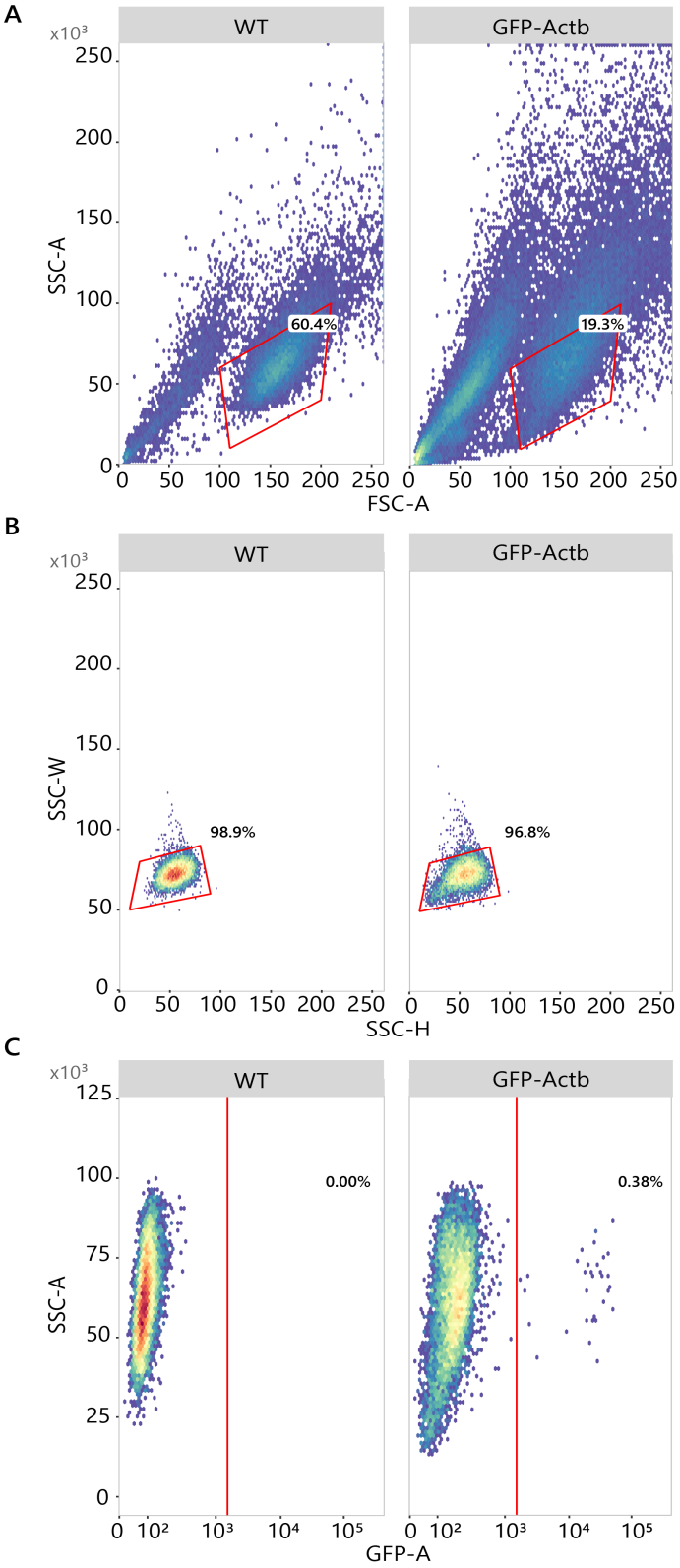


Figure 3.20: FACS Analysis of pORANGE-GFP-ACTB Transfected PC-12 Cells.

Gating (denoted using red lines) used in the FACS assay. Raw data was replotted using R (version 4.2.0). A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlets. C) GFP fluorescence intensity of live singlets.

3.3 Discussion

3.3.1 Limitations of the ORANGE System

As with all CRISPR Systems, ORANGE has several limitations that must be addressed. One of the most pressing issues with ORANGE is that there is no way to determine which step requires optimisation should no fluorescent signal be detected. There are a number of steps in the CRISPR pipeline that require optimisation - including guide RNA design and cutting, donor insertion, and transfection conditions. In this work, the PX458 cutting efficiency analysis was indispensable for ensuring that the guides were functional before use in the ORANGE system. This essential step gave confidence that the guides selected created Cas9-mediated cutting with the highest efficiency. It also enabled a visualisation of transfection rates, ensuring that any differences in cutting rate were because of guide efficiency and not poor transfection levels. Although this work mitigated these issues by using PX458, other users of the ORANGE system may encounter these difficulties if no initial guide optimisation is carried out.

Additionally, the ORANGE system, like other neuronal-based CRISPR methodologies, relies on NHEJ to repair the Cas9-mediated DSBs created during donor DNA insertion. As discussed in the introduction to this chapter, this is due to the fact that HDR is downregulated in post-mitotic cells such as primary neurons. NHEJ is an error-prone repair process resulting in the generation of indels around the donor DNA insertion site. Although this may be problematic for some downstream applications, for fluorescence imaging the generation of indels is not an issue - only cells with in-frame donor DNA insertions will be fluorescent. In this way, ORANGE naturally selects for the correct donor DNA insertion containing no deleterious indels.

3.3.2 CRISPR Scores May Be a Poor Indicator of Guide RNA Editing Efficiency

This research suggests that the CRISPR scores used to generate guide RNA targets are not an accurate reflection of guide efficiency. The Doench on-target (Doench et al. 2016) score is commonly used as an indicator of guide RNA efficiency. However, there is mounting evidence that this scoring system, whilst accurate for the system it was developed in, may not be as accurate in other CRISPR systems. In particular, the Doench on-target score for the guides selected in this work does not correlate strongly with the cutting efficiency scores generated by the ICE Analysis. One clear example of this is RPS2 guide 2, which had a Doench score similar to that of its guide 1 counterpart but produced an ICE cutting efficiency percentage of 0%, despite having similar levels of transfection (Figure 3.8). Additionally, RPS3 guide 1 had a substantially lower Doench score than RPL22 target 2 (35.1 versus 52.0), however, RPL22 guide 2 showed a much lower editing efficiency in the ICE analysis. As these samples had similar levels of transfection, it can be assumed that the cutting efficiencies of these guides aren't accurately reflected in their Doench on-target score.

This observation is discussed in the ORANGE paper. Willems et al. state that they see no clear correlation between Doench score and the frequency of donor DNA integration in the ORANGE system (Willems et al. 2020). The authors suggest that this may be due to the Doench score being developed based on knock-out efficiency, which is in contrast to the desired knock-in outcome of ORANGE. Interestingly, the converse appears to be true for the MIT CRISPR score. Willems et al. note that for ORANGE the instances of non-target protein tagging are rare, with selection of guides with high MIT scores resulting in very little off-target donor expression. However, this may not be due to MIT score alone. It was also suggested that the design of the knock-in constructs assists with lowering off-target effects, as 50% of donor insertions are in the wrong orientation resulting in Cas9-mediated donor removal. Additionally, 66% of donor insertions will be out-of-frame owing to the

triplet nature of genetic code. As a result of this, Willems. et al state they only took MIT score into account when designing CRISPR guides.

In summary, CRISPR scores are helpful when selecting possible guide RNAs, however guides with lower Doench scores may still produce the desired effect and should not be dismissed on the basis of low score alone. This also emphasises the importance of checking guide RNA cutting efficiency experimentally before selection of guides for downstream applications.

3.3.3 Quantification of Cutting Efficiency Using the T7E1 Assay

Two methods were used to determine if the guide RNAs selected generated Cas-mediated cutting - the T7E1 assay and Sanger sequencing followed by ICE analysis. Both methods had clear advantages and disadvantages. Commonly, the T7E1 assay is used as a fast, cost-effective method to determine whether CRISPR guides generate genomic DSBs. In particular, the assay is helpful when screening large numbers of guides at once, before picking successful candidates for further testing by Sanger sequencing. Although the T7E1 produced some helpful data in this research, the method required large amounts of optimisation and delivered variable results.

Issues with the T7E1 assay are well documented. The quantification of editing efficiency by T7E1 is possible by measuring the relative intensity of additional edited bands compared to the parent band. In a 2018 Nature study comparing the T7E1 assay to next generation sequencing analysis, the authors determined that the maximum NHEJ detection by the T7E1 assay plateaus at 30-40% (Sentmanat et al. 2018). Another common caveat of the T7E1 assay is issues with high background, which manifests as smears or faint bands present in the WT and negative control samples. This is likely due to the non-specific activity of the T7E1 enzyme. Therefore, the assay requires careful optimisation of DNA preparation, enzyme amount, incubation temperature and time. As a result of this high background, the bands produced from guides with very low cutting efficiencies are often not clearly visible.

This is supported by the *Sentmanat et al.* study, which determined that guides with an editing percentage of ≤10% appear inactive in the T7E1 assay (Sentmanat et al. 2018).

Overall, the T7E1 assay is a useful tool for inexpensive mass screening of guide RNAs. In this thesis, the presence of additional bands was only used as a qualitative indicator that editing had occurred. However, compared to the T7E1 assay, Sanger sequencing validation provided clearer results with less optimisation required. Whilst the T7E1 assay provided some indicator of guide cutting, the assay was a major time investment for the number of guides being tested in this work. In contrast, Sanger sequencing followed by ICE analysis involves fewer steps and requires less optimisation than the T7E1 assay. To this end, sequencing validation is generally considered a more reliable, if more expensive, method for guide RNA validation. If this process was to be repeated, guide cutting efficiency determination could be streamlined by focusing on quantification by sequencing and ICE analysis alone.

3.3.4 Cutting Efficiency May Be Higher Than ICE Analysis Determined

The ICE analysis program from Synthego (Synthego Performance Analysis, ICE Analysis. 2019. v3.0. Synthego; [2021-2022]) proved to be a useful tool in guide RNA validation, however, it is possible that extremely low cutting efficiencies were masked in this analysis. Two of the guides tested (RPL4 guide 2 and RPS2 guide 2) were determined to have an editing efficiency of 0%, despite both having good Doench on-target scores. In reality, it is possible that the editing efficiency is above zero as a number of downstream experimental factors can influence the ICE editing efficiency estimate. These factors include; poor transfection rates, quality of the DNA preparation used for transfection, low abundance of editing resulting in higher PCR amplification of WT DNA, degradation of PCR product during transport to the sequencing facility, and lower sequencing quality from purified PCR product. All of these factors must be taken into consideration before it can be determined

that a guide has a true editing efficiency of zero.

For RPL4 Guide 2 there was lower transfection efficiency than seen with other guides, which may have been a large factor in its ICE efficiency score. However, as RPL4 guide 1 was proven to cut effectively, it was decided that in the interest of time it wasn't necessary to repeat the cutting efficiency assays for guide 2. In contrast, the transfection level for RPS2 guide 2 was shown to be similar to that of RPS2 guide 1, so this is unlikely to explain the ICE editing efficiency. Additionally, the RPS2 transfection DNA preparations were made at the same time, and determined to be of a suitable concentration and purity. Interestingly, RPS2 guide 2 also appeared to work in the T7E1 assay, indicating that some Cas9-mediated cutting was taking place. A likely possibility is that a low amount of cutting took place, but there may have been contamination of the PCR product originating from the PCR cleanup step which could have negatively affected the quality of the sequencing data.

Interestingly, the published paper outlining the ICE tool shows that the ability of ICE to detect single nucleotide variants has not been validated for populations with $\leq$5% edited DNA (Conant et al. 2022). Additionally, the authors note that ICE analysis is sensitive to noise arising from poor Sanger sequencing quality. As a result, it is possible that any samples with an editing efficiency under 5% are incorrectly determined to be zero. For this reason, it is likely that RPS2 guide 2 creates Cas9-mediated cutting, however, the editing efficiency score was artificially lowered due to a variety of experimental factors in combination with the ICE tool not recognising editing levels under 5%. However, guides with an editing efficiency of $\leq$5% are unlikely to be utilised in downstream CRISPR assays. This finding may explain the discrepancies in this research between the results of the T7E1 assay and the ICE analysis for some samples.

3.3.5 Considerations for FACS Assays

The heterogeneous nature of the CRISPR-edited cell lines is a possible point of concern for downstream assays. While the validation results would be clearer with a

homogeneous population, the nature of the PC-12 cells prevented the development of single-cell clones. Single cell sorts of CRISPR-edited cells was attempted multiple times, however, the population did not reach an exponential growth phase despite regular feeding with conditioned media. Additionally, there is no evidence in the literature that single cell sorting and expansion of PC-12 cells is possible. However, as these lines were developed for CRISPR validation, the presence of WT cells producing untagged ribosomal protein should not pose a problem. As presented in the next chapter, the identification of tagged protein bands using in-gel fluorescence and western blot assays is proof of successful HaloTag insertion. Additionally, the PCR and subsequent sequencing of the genomic insertion region are also not hindered by the presence of unedited alleles. For the purposes of this research, a heterogeneous population is adequate and enabled the project to progress at a faster rate.

The percentage of edited cells collected in the FACS assays varied greatly between assays due to a number of experimental factors. The factor that influenced CRISPR editing rate the most appeared to be the age of the culture - which is unsurprising given immortalized cells of a high passage number are generally considered to be less healthy and show greater genetic instability (Hughes et al. 2007). To this end, cultures above 20-25 passages showed decreased editing levels and the FACS assays had to be repeated. Freshly thawed cells (transfection at 2 weeks post thaw) proved to deliver the highest levels of cell editing. Interestingly, when experimental factors are identical between FACS assays (i.e. cells of the same passage, using the same DNA preparation and transfection and labelling protocols) there can still some variation in the number of CRISPR-edited cells. For example, for RPS2-Halo transfected cells, 3% of the parent population had a HaloTag present (Figure 3.18), whereas a previous attempt showed 5.2% edited cells. Considering many of the variables were kept the same, this difference may be attributed to pipetting error or incorrect cell count at the time of transfection.

3.3.6 CRISPR-Edited PC-12 Cell Lines Need Further Validation To Ensure Correct HaloTag Insertion

This chapter details the design and creation of ORANGE CRISPR plasmids and subsequent development of PC-12 CRISPR-edited cell lines, completing aim one and two of this thesis. However, further work is required to validate that the HaloTag has been inserted in the correct genomic location and that tagged protein is incorporated into functional ribosomes. Chapter 4 will present the validation assays for each of the CRISPR cell lines, an essential step before use in primary neuronal cultures.

Chapter 4

Validation of ORANGE-Edited PC-12 Lines

4.1 Introduction

The ORANGE CRISPR plasmids created in Chapter 3 require further validation to ensure that the desired knock-in effect is achieved, before their transfection into primary neurons. As previously discussed, neurons are post-mitotic cells with low transfection efficiencies making them a challenging system to carry out CRISPR validation assays. However, it is essential that the ORANGE CRISPR plasmids created only tag the desired protein target and that those proteins are incorporated into functional ribosomes. This ensures that downstream live cell imaging assays visualise the movement of ribosomal protein in axons and not any other erroneously tagged endogenous protein. Therefore, validation of the effectiveness of the CRISPR plasmids is critical to the success of this project.

The CRISPR-edited PC-12 lines established in the previous chapter allow easier characterisation and circumvent the issues of CRISPR construct validation in neurons. These heterogeneous lines are composed of cells with a HaloTag sequence inserted in one or more alleles - ensuring a moderately sized population of tagged protein, and therefore, reasonably strong HaloTag expression for downstream assays.

Additionally, there is no need for repeated transfections as all cells should express the HaloTag. As a stable line, the PC-12 CRISPR cells can also easily be expanded to produce large quantities of cells for validation assays. This is in contrast to primary neurons, where tissue availability and transfection efficiencies are a limiting factor.

The overall aim of this chapter is to validate each of the ORANGE CRISPR-edited PC-12 cell lines - GFP-ACTB, RPL4-Halo, RPL22-Halo, RPS2-Halo, and RPS3-Halo. The positive control pORANGE GFP-ACTB KI #1 (Addgene Plasmid #131479) construct will be necessary in downstream imaging assays to verify the transfection and imaging protocols are working before using the experimental ORANGE ribosomal protein constructs. Therefore, validating that this construct produces the expected knock-in effect is essential. Subsequently, the PC-12 Halo lines will undergo a robust validation pipeline to validate correct knock-in and determine whether the tag affects ribosomal protein functionality.

The validation pipeline includes several assays that give confirmation that a HaloTag is present and inserted at the correct location in the desired gene (including in-gel fluorescence, western blot, and Sanger sequencing). Importantly, the ability for the tagged protein to be incorporated into functional ribosomes is determined by polysome profiling. This is an critical step as failure of tagged protein to join the nascent ribosome will prevent the analysis of ribosome translocation in axons. Additionally, free tagged ribosomal protein that is not incorporated into ribosomes is likely to localise in the nucleus and be targeted for degradation (Sung et al. 2016). After the validation assays are complete, the most efficient plasmids will be selected for use in future ribosome protein imaging experiments (see Chapter 5).

4.2 Results

4.2.1 Validation of Positive Control Plasmid pORANGE GFP-Actb KI

To determine if transfection of the pORANGE-GFP-ACTB plasmid produces GFP-tagged β-actin protein, a two-colour fluorescent western blot was carried out (Figure 4.1). Confluent T25 flasks of both WT PC-12 cells (acting as a negative control) and GFP-ACTB edited cells collected via FACS were lysed. 75 µg cell lysate was run on a 10% acrylamide gel, transferred to a PVDF membrane, and the blot was then cut at 95kDa using a sterile scalpel. The top half of the blot (with proteins above 95kDa bound) was probed with KIF5B primary antibody as a loading control. Given that the weights of the ribosomal proteins and Halo-tagged proteins being detected in the western blot assays ranges from 15kDa (RPL22) to 81kDa (RPL4-Halo), most commonly used loading controls - such as GAPDH - would not be suitable for these assays. Therefore, KIF5B (a ubiquitously expressed Kinesin-1 heavy chain isoform) was selected as a loading control. This protein should be uniformly expressed and produce a 110kDa band in all cell lines - running higher than the expected range for all endogenous ribosomal proteins or HaloTag CRISPR products. All loading controls for western blot assays can be found in the Appendix, Figure 8.2. Loading controls were checked to confirm equivalent amounts of protein in each sample.

The bottom section of the blot (with proteins below 95kDa bound) was probed with mouse anti-GFP primary antibody. Secondary antibody (Alexa Fluor 790-Affini-Pure Donkey Anti-Mouse) was then applied. The blot was imaged using the LI-COR XF odyssey imaging system, using the 800 nm channel (Figure 4.1). For the GFP-ACTB sample, a band is observed at ∼69kDa in the GFP channel (800nm, green) which is not present in the WT sample. This band is the expected weight for GFP-tagged β-actin protein. The blot was then re-probed with rabbit anti-β-actin primary antibody and Alexa Fluor 680-AffiniPure Donkey Anti-Rabbit secondary. The β-actin channel (700 nm, magenta) clearly shows a large bands at

∼42kDa for both the WT and GFP-ACTB samples, indicating the presence of WT β-actin. A possible faint β-actin-GFP band can be seen in the GFP-ACTB sample at around 69kDa (denoted by the arrow), but the multiple non-specific bands for this antibody make this difficult to clearly observe. However, as the immunogen for this antibody targets the N-terminal region of the protein, it is probable that the N-terminal GFP tag on β-actin prevents proper antibody binding and western blot detection. Despite this, the western blot provides data that the pORANGE GFP-ACTB plasmid correctly edits the rat genome to produce GFP-tagged β-actin protein.

Additionally, immunofluorescence assays of differentiated GFP-ACTB and WT PC-12 cells were also carried out (Figure 4.2). This assay acts as further validation that the β-actin protein has been GFP tagged and the tagged protein does not show abnormal distribution or accumulation within the cell. Single channel controls for each antibody used were also included to detect any signal bleed-through into other channels. For the WT cells, endogenous β-actin signal can be detected predominantly localised to the cell periphery. No significant signal for GFP primary antibody is detected, although some faint signal is observed in the β-actin channel (likely due to background antibody staining). In contrast, strong GFP signal is observed for the GFP-ACTB cells, which is shown to overlap with the β-actin signal in the composite image. The GFP-ACTB cells appear similar to WT PC-12 cells in morphology, with no obvious accumulation of GFP-tagged protein.

To conclude, the GFP-ACTB PC-12 line was validated using fluorescent western blot and immunofluorescence to determine that GFP-tagged β-actin protein is present. It can be stated with confidence that the pORANGE-GFP-ACTB KI plasmid works as intended and can be used in subsequent optimisation steps in primary neuronal cultures.

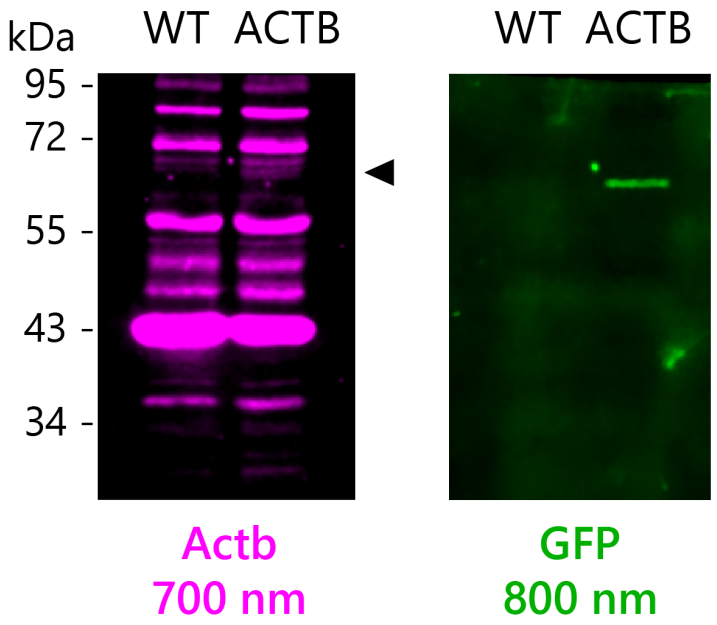


Figure 4.1: PC-12 GFP-Actb Edited Line Fluorescent Western Blot.

Cell lysate from GFP-ACTB edited PC-12 cells was analysed by two-colour fluorescent western blot to detect the presence of GFP tagged β-actin protein. A WT negative control sample was also run alongside the GFP-ACTB sample. The blot was probed with anti-β-actin and anti-GFP primary antibodies (see Chapter 2 for antibody concentrations and supplier information). Both samples show a large 42kDa band for β-actin (700 nm channel, magenta), with only the GFP-ACTB sample having an additional 69kDa band for GFP tagged β-actin (800 nm channel, green). The loading control for this western blot is provided in Appendix, Figure 8.2. Abbreviations: ACTB - GFP-ACTB PC-12 lysate.

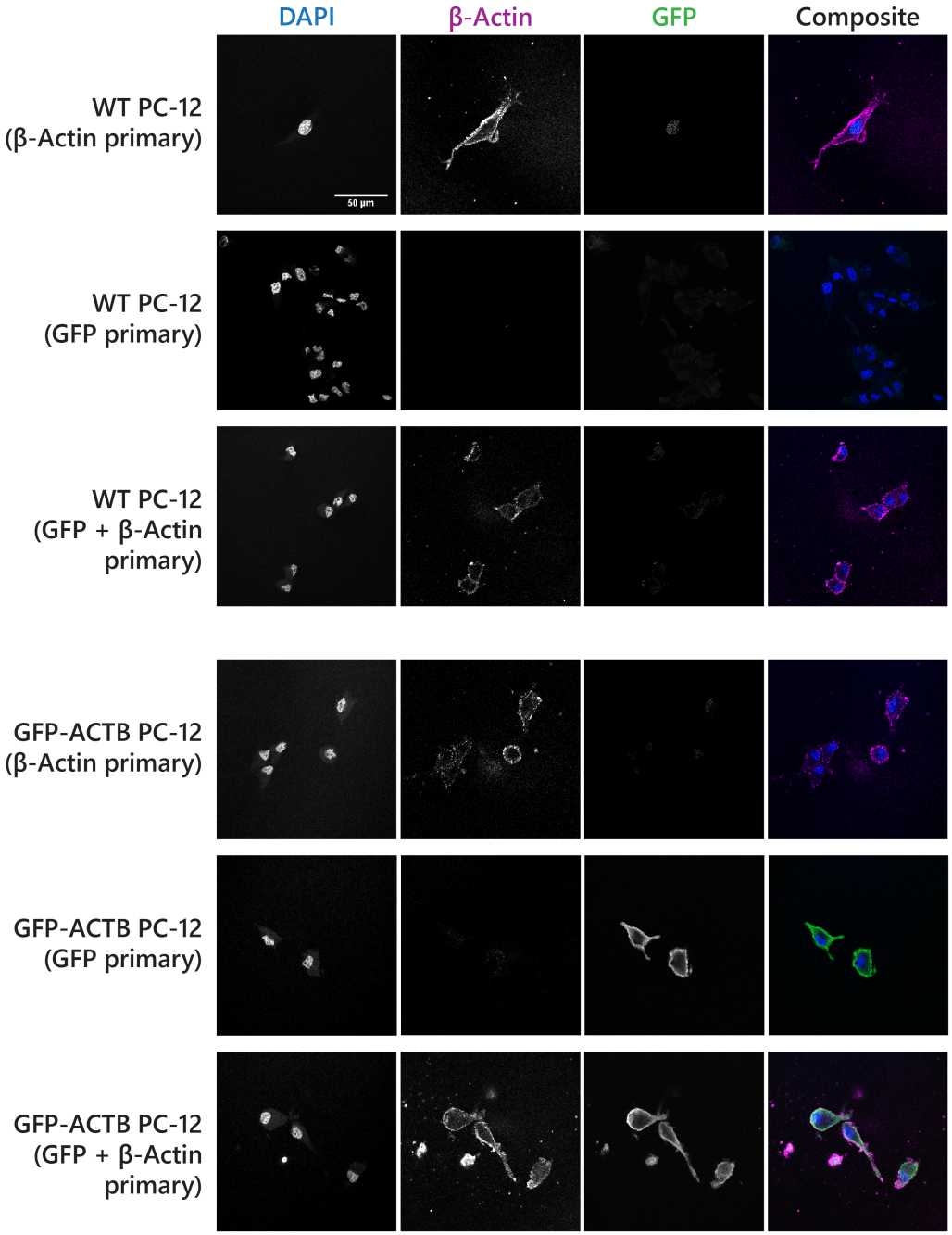


Figure 4.2: Immunofluorescence and Confocal Imaging of GFP-ACTB PC-12 Cell Line.

WT and GFP-ACTB edited cells were plated on PLL-coated coverslips and differentiated using NGF for 7 days. On day 7 post differentiation, cells were fixed and immunofluorescence was carried out with anti-mouse β-actin and anti-chicken GFP primary antibodies. Secondary antibodies used were Donkey anti-Mouse Alexa Fluor Plus 647 and Goat anti-Chicken Alexa Fluor Plus 488 respectively. Cells were imaged using a TCS SP5 confocal (Leica Microsystems) using a 63x oil objective. Images were then cropped processed using ImageJ (see Section 2.8.4).

4.2.2 Validation of pORANGE CRISPR Plasmids in PC-12 Cell Lines

4.2.2.1 Determining HaloTag Presence Via In-Gel Fluorescence

The first assay in the ORANGE Halo knock-in pipeline is in-gel fluorescence, a simple assay which detects the presence of the HaloTag within cells. Detection of the tag is done by labelling the edited cells with 100 nM Halo-TMR ligand for 1 hour, before cells are lysed and the lysate is run on a 10% acrylamide gel. As the fluorescent TMR ligand is a small molecule that remains stable in protein denaturing conditions (Los et al. 2008), the HaloTag-TMR complex can be visualised by SDS-PAGE followed by imaging using the 600 nm channel on a LI-COR imaging system. In-gel fluorescence is a quick and easy way to determine if the HaloTag is being expressed in cells, the number of different proteins which are Halo-tagged, and lastly, whether the tagged protein has the expected molecular weight.

In-gel fluorescence for each PC-12 Halo knock-in line was carried out (Figure 4.3). Three separate gels were run on different days due to expansion of the knock-in lines being done sequentially, and therefore, the assays being conducted at different times. As a negative control, each sample was run alongside a WT PC-12 cell lysate sample of the same passage number that had undergone the same Halo-TMR labelling. During the lysis and SDS-PAGE process, care was taken to ensure that light exposure to the samples was minimal to prevent signal loss due to photobleaching.

Most ORANGE Halo lines showed HaloTag expression via in-gel fluorescence. The RPL4-Halo PC-12 line gel showed one clear band at approximately 81kDa, which is the expected weight for RPL4-Halo (Figure 4.3). No additional bands were observed in either the RPL4-Halo or WT lanes. Similarly, a single band ∼81kDa was observed for RPL22-Halo (Figure 4.3). However, this band was the incorrect weight for RPL22-Halo - the expected weight for RPL22-Halo is approximately 48kDa. This was a surprising result given the strong band signal intensity that is not indicative of off-target CRISPR effects. RPS3-Halo, in contrast, had no bands present despite

successful expansion of FACS-selected cells (Figure 4.3). This was a highly unusual finding, as all cells present in this sample supposedly expressed HaloTag in one or more alleles of the RPS3 gene, thus the edited cells should produce a RPS3-Halo band at ∼59kDa. The in-gel assay was repeated with RPS3-Halo lysate, delivering the same result. The absence of a band in this assay points to a possible issue with the expression of the tag within the cell line, so further investigation was needed. Lastly, several bands were observed for the RPS2-Halo sample, including one at ∼64kDa which is the expected weight for RPS2-Halo. The additional bands at a lower weight could possibly be degradation products of the tagged protein, which would explain the similar band intensity seen with all three bands.

Overall, the in-gel fluorescence assay produced insight into the success of each ORANGE Halo plasmid. In particular, pORANGE-RPL4-Halo appeared to be creating the desired knock-in product. Other plasmids (such as RPL22) appear to show tagged protein at the incorrect weight. One possibility is that the RPL4-Halo line had been mistook for the RPL22-Halo line, or that the same sample had been loaded for both. Further work to determine the identity of this protein is required. RPS2 also appeared to work but additional assays to investigate the identity of the multiple other bands would be helpful to rule out off-target effects. Lastly, RPS3-Halo cells did not produce a band in the in-gel assay, which could possibly be explained by loss of the HaloTag after FACS selection. Subsequent FACS re-analysis of the population should provide evidence in support of this theory.

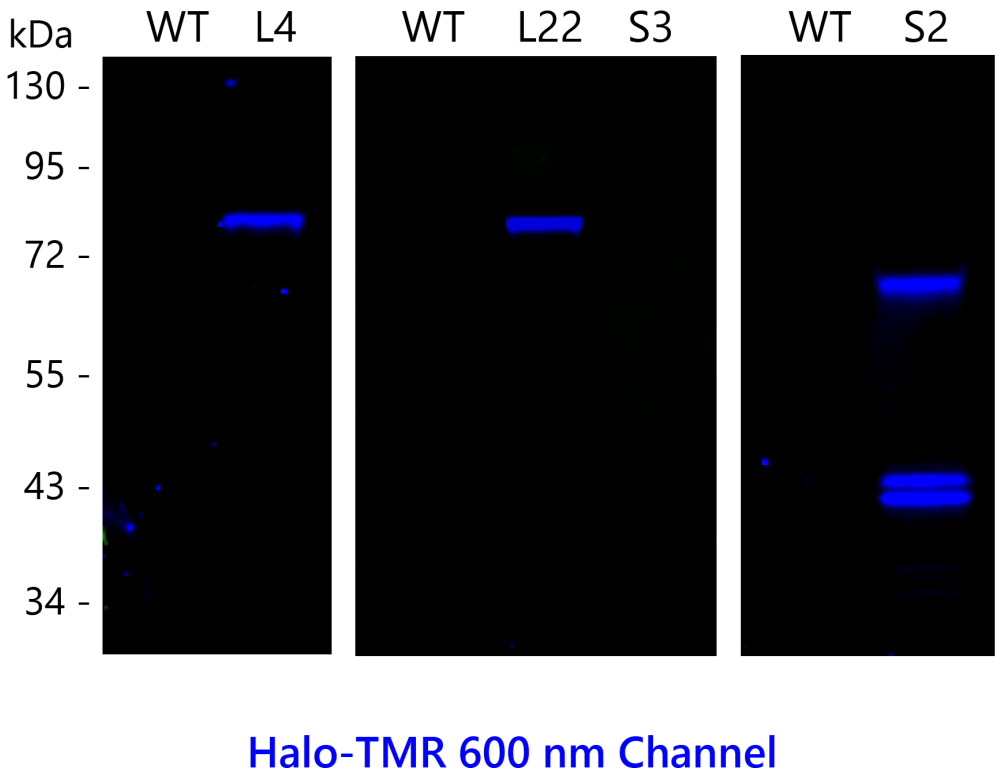


Figure 4.3: Detection of Halo-Tagged Protein Via In Gel Fluorescence.

Cell lysate from stable CRISPR-edited PC-12 Halo cells was analysed by in-gel fluorescence assay to detect the presence of the HaloTag. Cells were labelled with 100 mM Halo-TMR ligand for 1 hour before being lysed and 75 µg lysate run on a 10% acrylamide gel. Samples were protected from light during the duration of the assay. Imaging was done on the LI-COR XF Odyssey using the 600 nm channel. A WT PC-12 sample of the same passage number was run alongside each edited PC-12 sample. Each gel was run on a different day, due to the expansion of FACS cells being completed at different times. Abbreviations: L4 - RPL4-Halo, L22 - RPL22-Halo, S3 - RPS3-Halo, S2 - RPS2-Halo.

4.2.2.2 RPS3-Halo Cells Lose HaloTag Expression After Several Weeks in Culture

The RPS3-Halo PC-12 line was labelled with Halo-TMR and again examined via FACS to detect fluorescent cells which are expressing the HaloTag. As this line had been expanded from a population of highly fluorescent cells previously collected by FACS, it was expected that the RPS3-Halo line would predominantly contain cells with fluorescence intensity that was significantly higher than WT samples. FACS re-analysis (Figure 4.4) determined this was not the case. Only 0.01% of cells had a fluorescence intensity that placed them in the gated region used in the previous RPS3-Halo FACS assay (Figure 4.4C). As a result, it can be assumed that the RPS3-Halo cells do not express the HaloTag at detectable levels after several weeks post-sort.

The pORANGE-RPS3-Halo transfections and FACS selection were repeated to generate a second RPS3-Halo biological replicate (Figure 4.5). After several weeks of expansion from a population of fluorescent cells, the in-gel fluorescence assay was repeated with Halo TMR-labelled cell lysate (Figure 4.6). No bands were observed. Subsequently, the second RPS3-Halo line was labelled with Halo-TMR and re-analysed by FACS (Figure 4.7). The FACS assay determined that only 0.08% of the population had a fluorescence intensity above that of the WT TMR+ sample, therefore, it appeared that the second RPS3-Halo line was also not expressing the HaloTag at the expected level. As the creation of CRISPR-edited stable PC-12 lines was time intensive and was showing no variation in result, it was decided that a third biological repeat of RPS3-Halo was not necessary. A thorough discussion as to why the RPS3-Halo PC-12 lines do not seem to stably express the HaloTag is provided in Section 4.3. Subsequent validation of pORANGE-RPS3-Halo was stopped at this stage to focus on more promising CRISPR plasmids.

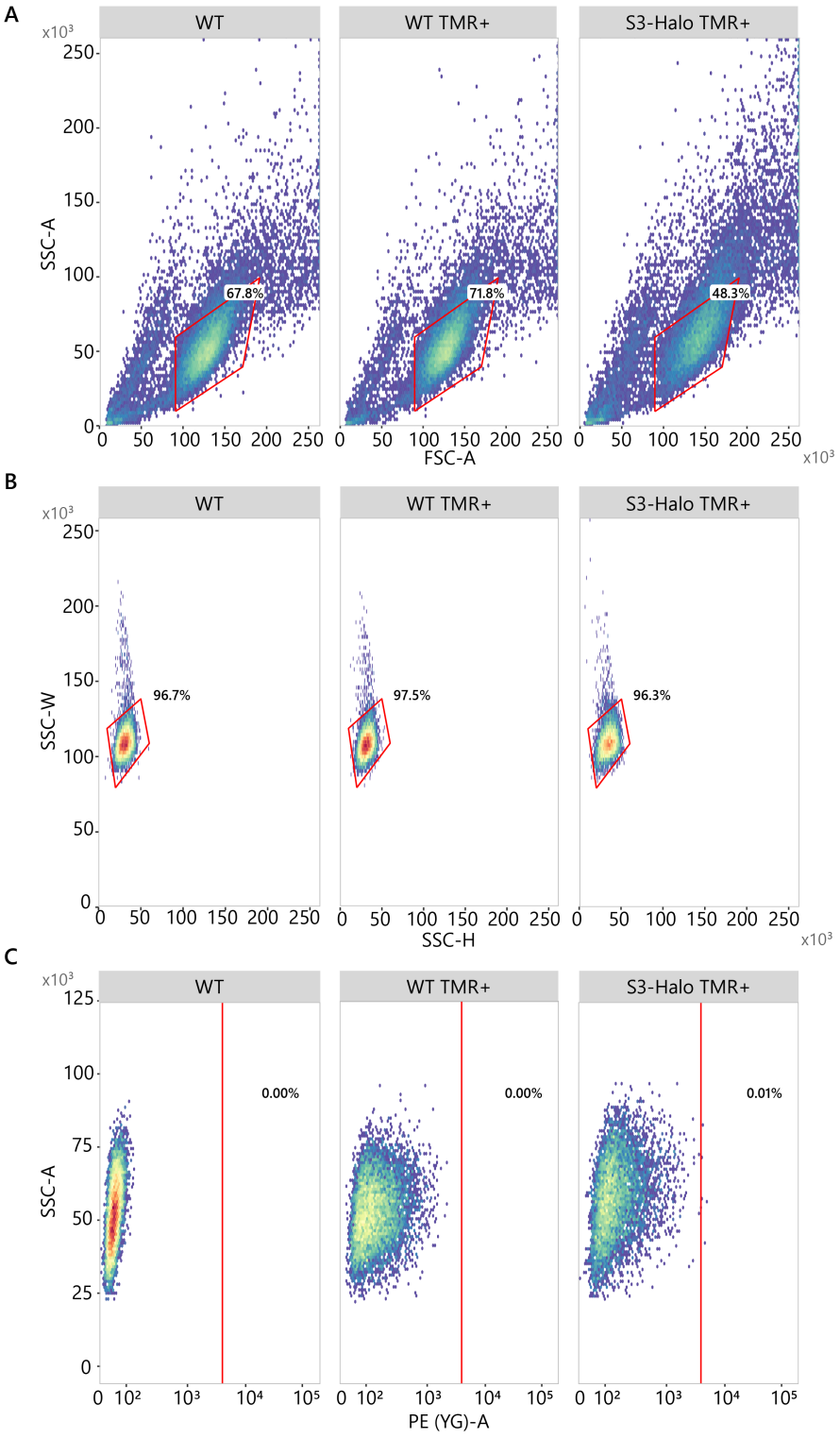


Figure 4.4: FACS Re-Sort of RPS3-Halo PC-12 Line Labelled with Halo-TMR.

The RPS3-Halo PC-12 line was labelled with Halo-TMR and FACS re-sorted (see Section 2.3.4) to determine if the cells still expressed the HaloTag. Raw data was plotted using R (version 4.2.0). Gating regions used for cell selection are denoted with red lines. A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlets. C) Fluorescence intensity of live singlets.

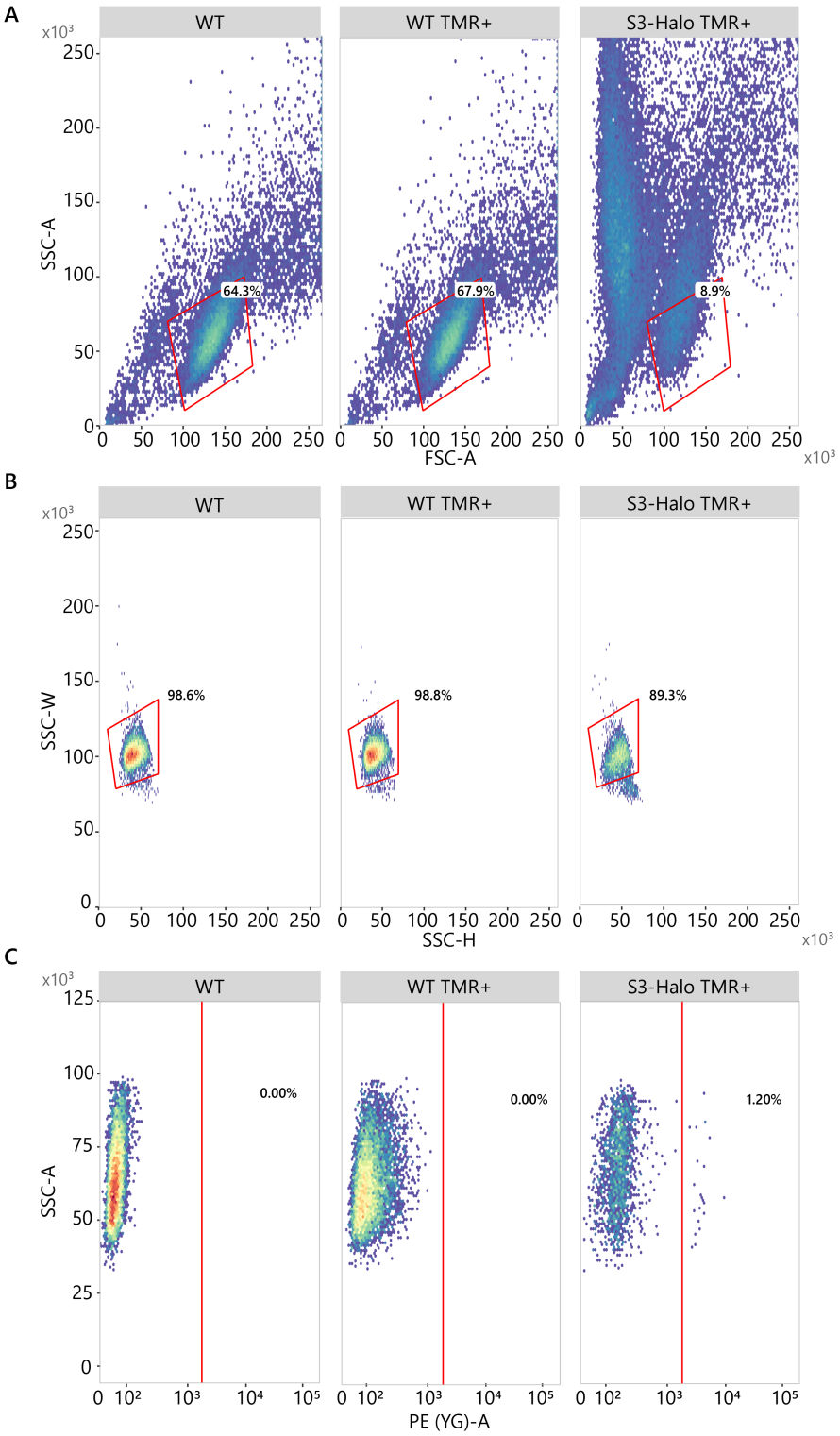


Figure 4.5: Second FACS Analysis of pORANGE-RPS3-Halo Transfected PC-12 Cells Labelled with Halo-TMR.

A second RPS3-Halo line was made by transfection with the pORANGE-RPS3-Halo plasmid, labelling with Halo-TMR and FACS sorting (see Section 2.3.4). Raw data was plotted using R (version 4.2.0). Gating regions used for cell selection are denoted with red lines. A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlets. C) Fluorescence intensity of live singlets.

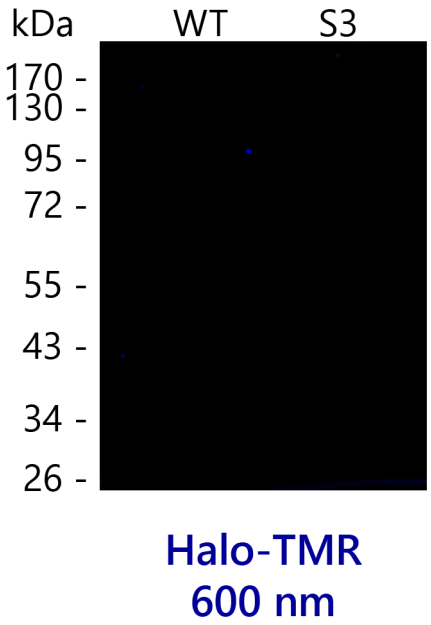


Figure 4.6: In-Gel Fluorescence Assay of Second RPS3-Halo Line Labelled with Halo-TMR.

Cell lysate from the RPS3-Halo PC-12 line was analysed by in-gel fluorescence assay to detect the presence of the HaloTag. Cells were labelled with 100 mM Halo-TMR ligand for 1 hour before being lysed and 75 µg lysate run on a 10% acrylamide gel. Samples were protected from light during the duration of the assay. Imaging was done on the LI-COR XF odyssey using the 600 nm channel. A WT PC-12 sample of the same passage number was run alongside the RPS3-Halo sample. Abbreviations: S3 - RPS3-Halo cell lysate.

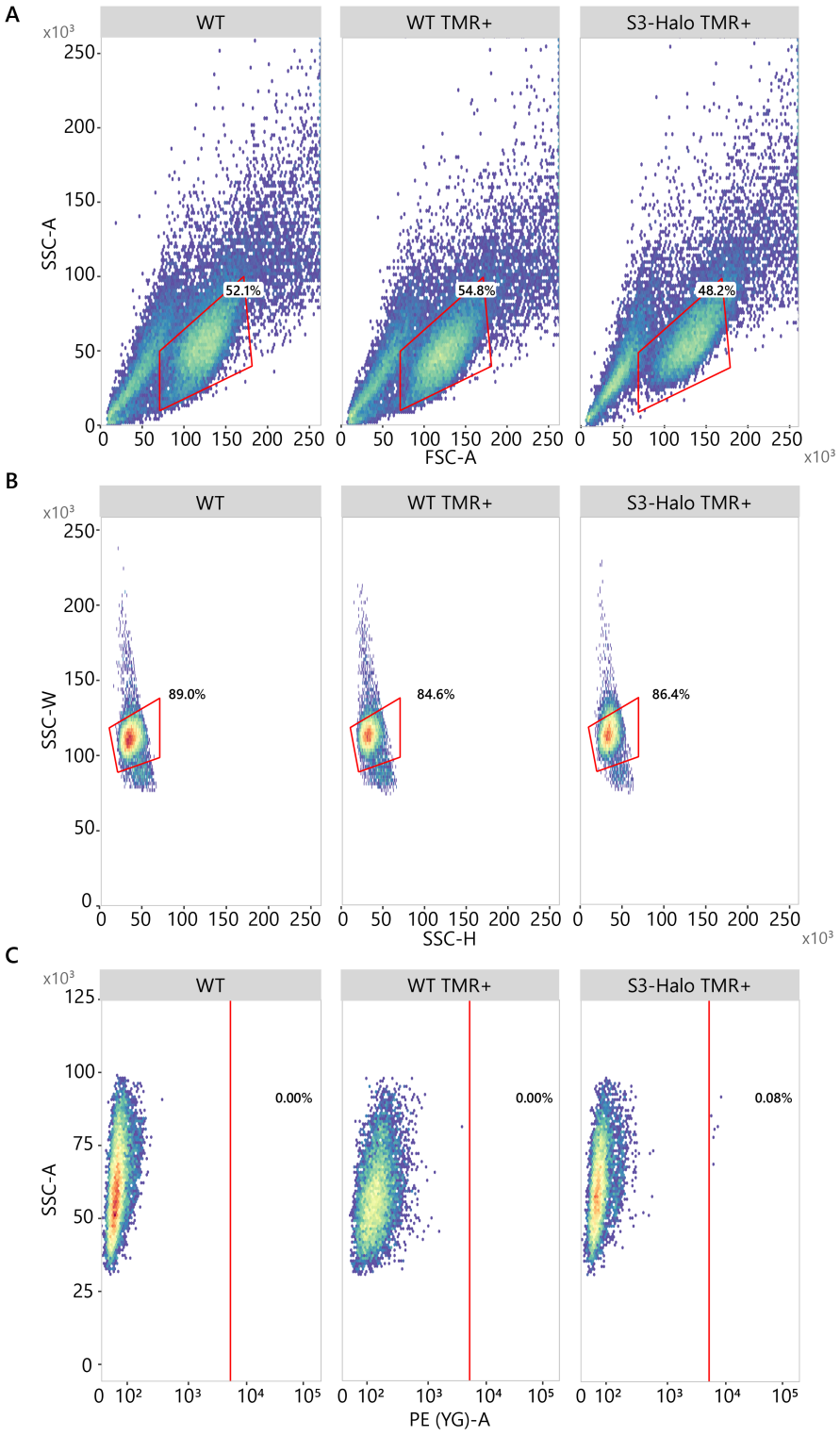


Figure 4.7: FACS Re-Sort of Second RPS3-Halo PC-12 Line Labelled with Halo-TMR.

The second RPS3-Halo PC-12 line was labelled with Halo-TMR and FACS re-sorted (see Section 2.3.4) to determine if the cells still expressed the HaloTag. Raw data was plotted using R (version 4.2.0). Gating regions used for cell selection are denoted with red lines. A) Gating to select live cells, excluding dead cells and cellular debris. B) Gating of live population to select singlets. C) Fluorescence intensity of live singlets.

4.2.2.3 Verifying the Site of HaloTag Insertion Via Sanger Sequencing

The next stage of CRISPR cell line validation was to verify that the HaloTag has been inserted into the genome at the expected location. By amplifying the genomic locus containing the insertion site by PCR, the DNA fragment can be analysed by Sanger sequencing to confirm the HaloTag is at the correct terminal region. A 500 - 600 base pair section of the RPL4, RPL22 and RPS2 genes were amplified from genomic DNA of RPL4-Halo, RPL22-Halo, RPS2-Halo and WT PC-12 lines. Specifically, the expected size for the RPL4-Halo, RPL22-Halo and RPS2-Halo products were 580bp, 544bp and 536bp respectively. The PCR was done using the forward primers designed for their T7E1 assay (Table 2.13) and a reverse primer targeting the middle of the HaloTag sequence (Halo Middle (Rev), Table 2.8). If a HaloTag is present, the PCR product generated should contain the insertion site at approximately 250 bases followed by the beginning of the HaloTag sequence. For WT cells, no PCR product is expected as there should be no reverse primer binding.

Amplification of the expected insertion site was successful using genomic DNA from both RPL4-Halo and RPS2-Halo lines (Figure 4.8). For both, a band of roughly ∼500 bp can be observed in the Halo sample lanes, with no band for the negative control or WT samples. Importantly, this gel provides clear evidence that the HaloTag is present in the RPS2 gene of RPS2-Halo cells, as previous the in-gel validation step had been inconclusive. Conversely, no PCR product could be amplified using genomic DNA extracted from RPL22-Halo cells. Several attempts were made, varying the annealing temperature and number of PCR cycles. As the forward primer used was previously successfully used in the RPL22 T7E1 assay (see Chapter 3), and the reverse HaloTag primer had been successfully used to amplify the HaloTag insertion site of RPL4-Halo and RPS2-Halo cells, this indicates that the HaloTag is not present in RPL22-Halo cells at the expected locus.

Subsequently, the PCR for RPL4-Halo and RPS2-Halo was repeated in a 50 µl volume, purified using a column-based PCR cleanup kit, and sent for Sanger sequencing. The sequencing results were then aligned with the expected knock-in

genomic sequence using Benchling (Figure 4.9). For both RPL4-Halo and RPS2-Halo samples, the sequence data matches the expected sequence and the HaloTag sequence is present at the correct locus. Notably, the quality of the sequencing data from RPS2-Halo samples was poorer than the RPL4-Halo samples, with more uncertain base calls (highlighted in yellow) and base miscalls (highlighted in red). It is possible that this was due to sample contamination from the PCR cleanup, however, the PCR and sequencing was repeated several times with the similar results in quality. Another explanation is that there may be multiple alleles of the RPS2 gene within PC-12 cells, resulting in a heterogeneous DNA mixture. Despite this disparity in sequencing quality, both RPL4-Halo and RPS2-Halo show DNA insertion in the expected location.

4.2.2.4 Validating Ribosomal Protein Tagging by Western Blot

To verify that the desired ribosome protein had been tagged, western blots using fluorescent secondary antibodies were carried out for the RPL4, RPL22 and RPS2 Halo lines. By using secondaries which are excitable at different wavelengths, multiple primary antibodies can be detected simultaneously. This allows visualisation of antibody signal overlap without needing to strip and re-probe the membrane. If the desired protein target has been successfully tagged, a band in both the HaloTag/Halo-TMR and the protein target channels at the expected weight will be observed. For all edited samples, a WT PC-12 lysate sample is also run alongside to illustrate endogenous ribosome protein expression in un-edited PC-12 cells.

First, RPL4-Halo cell lysate was examined by western blot (Figure 4.10). The blot was probed with RPL4 and HaloTag primary antibodies, taking care to ensure that each antibody used was raised in a different species to prevent cross-reactivity. In the HaloTag (700 nm, green) channel, a clear band can be seen at ∼81kDa similar to the one observed in the RPL4-Halo in-gel (Figure 4.3). A band of this weight is not present in the WT sample. The presence of this band confirms the finding of the in-gel assay, verifying that the HaloTag is expressed on a protein of roughly

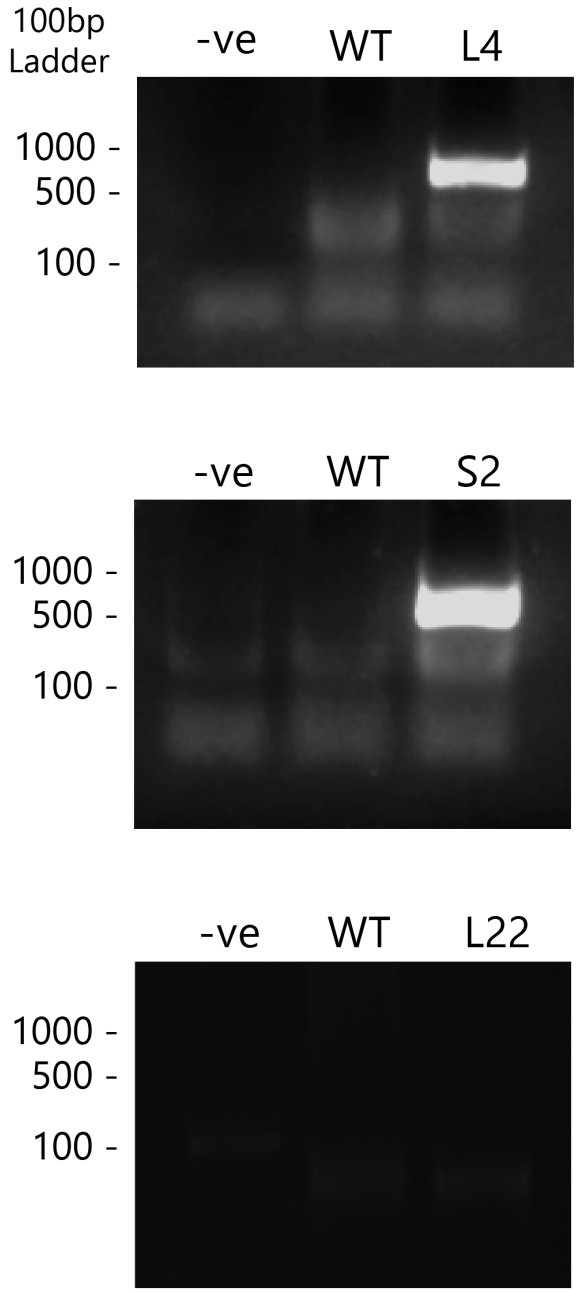


Figure 4.8: PCR of PC-12 Halo Line Donor DNA Insertion Region.

The genomic DNA from RPL4-Halo, RPS2-Halo, RPL22-Halo and WT PC-12 lines was extracted and the insertion site region was amplified by PCR. The forward primer used was the forward primer used for the T7E1 assay for that particular construct (Table 2.7). The reverse primer used was specific to the middle of the HaloTag sequence (Halo Middle (Rev)), Table 2.7). A product of 500-600 bp was observed if the HaloTag was present within the sample. A negative control reaction containing no genomic DNA template was set up for each PCR. PCR components and timings are given in Section . The DNA gel was imaged using the LI-COR imaging system using the 600 nm channel, or on a G:Box (RPL22-Halo). Abbreviations: L4 - RPL4-Halo, S2 - RPS2-Halo, L22 - RPL22-Halo.

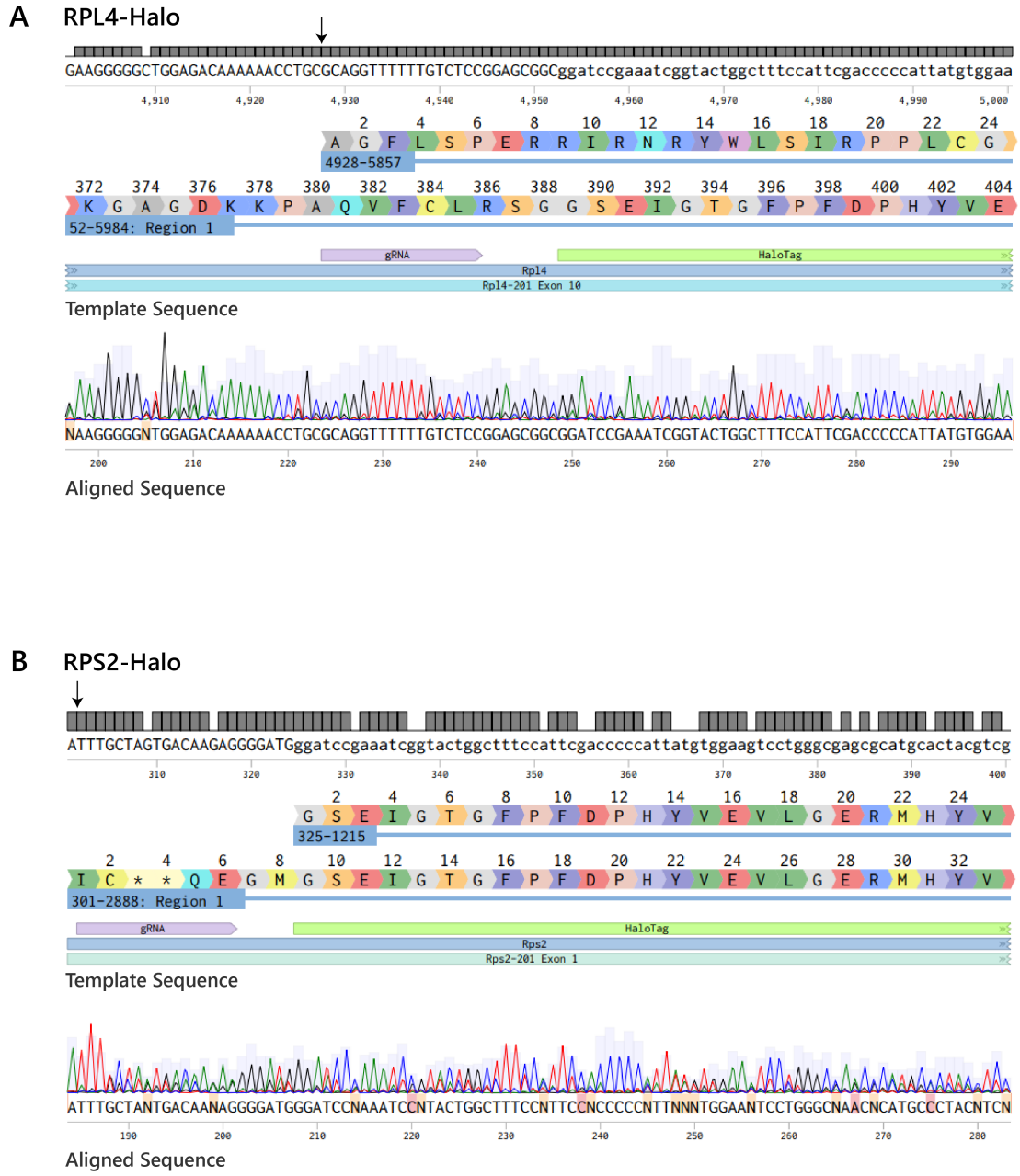


Figure 4.9: Sanger Sequencing of RPL4-Halo and RPS2-Halo Insertion Sites.

The genomic DNA from RPL4-Halo and RPS2-Halo PC-12 lines was extracted and the insertion site region was amplified by PCR (Figure 4.8). After purification, the samples were sent for Sanger Sequencing, using the forward primers previously designed for their respective T7E1 assays (Table 2.13). Sequencing data was aligned to the expected template sequence using Benchling. The SpCas9 cut site is denoted with an arrow. A) Sanger sequencing result from RPL4-Halo showing C-terminal HaloTag insertion in exon 10. B) Sanger sequencing result from RPS2-Halo showing N-terminal HaloTag insertion in exon 1.

48kDa. It must also be noted that the HaloTag primary does show some additional bands, but as these bands have lower signal intensity and are also present in the WT sample (which does not express the HaloTag), these bands are assumed to be due to non-specific binding of the HaloTag antibody. The ∼81kDa band, which has higher signal intensity, is only present in the edited sample and is therefore assumed to be the antibody interacting with HaloTag protein.

Furthermore, the RPL4 (800 nm, magenta) channel also shows a band at ∼81kDa which is also not present within the WT sample. Bands for the endogenous, untagged RPL4 protein can be seen at around 48kDa in both the WT and RPL4-Halo samples. The presence of the double endogenous band may be the result of the antibody binding to an additional RPL4 isoform expressed within rat cells. A composite image of both the 700 nm and 800 nm channel was then created, demonstrating that there is signal overlap of the ∼81kDa band from each channel. As a result, it can be concluded that RPL4 has been successfully edited to express HaloTag.

A fluorescent western blot was then conducted to determine the identity of the tagged protein expressed within the RPL22-Halo PC-12 line (Figure 4.11A). As determined in the in-gel fluorescence assay (Figure 4.3), the TMR-labelled protein did not appear to be the expected weight for RPL22-Halo protein (48kDa). To this end, a western blot to detect RPL22 protein would provide information as to whether the tagged protein is RPL22 or if RPL4-Halo had been erroneously loaded instead. Due to the lack of a suitable anti-mouse HaloTag antibody at this time, Halo-TMR signal (600 nm, blue) was used to visualise the tagged protein.

As expected, the ∼81kDa protein did not appear to be RPL22. The Halo-TMR channel had a band at ∼81kDa present in the L22-Halo sample, as observed in the previous in-gel assay. The western blot showed no overlap of RPL22 (700 nm, magenta) signal within the composite image. Notably, the endogenous RPL22 protein (15kDa) is not present on the blot due to the blot being erroneously cut at the wrong marker. However, the antibody had been previously validated in PC-12 lysate and shown to produce a band at ∼15kDa (with additional bands at ∼34kDa

and ∼55kDa which are assumed to be either non-specific interactions with other proteins or RPL22 variants). As the immunogen recognises the N-terminal region of the protein, the RPL22 antibody should be able to bind to the C-terminally tagged RPL22-Halo. Therefore, initial efforts to identify this protein suggest that the 81kDa band is not RPL22-Halo.

The next step was to determine if the ∼81kDa band was RPL4-Halo. Previous western blots (Figure 4.10) have shown that RPL4-Halo can be detected at ∼81kDa using the RPL4 antibody. To determine if the band was RPL4-Halo, the blot from Figure 4.6A was stripped and re-probed with RPL4 primary antibody to see if a RPL4 band at ∼81kDa is detected (Figure 4.11B). For both the WT and RPL22-Halo sample, only endogenous RPL4 bands were observed. This result provides evidence that the RPL4-Halo and RPL22-Halo lines were not swapped, and neither was the RPL4-Halo sample loaded onto the RPL22-Halo gel accidentally. As the identity of the tagged protein was not confirmed to be RPL22, it was decided that pORANGE-RPL4-Halo would be used to tag the 60S subunit in future live cell imaging experiments.

Finally, RPS2-Halo lysate was validated by western blot using RPS2 and HaloTag primary antibodies (Figure 4.12A). In the RPS2 channel (700 nm, magenta), the endogenous RPS2 protein bands can be seen in both the WT and RPS2-Halo samples at approximately 31kDa. A ∼64kDa band for RPS2-Halo can possibly be seen in the RPS2-Halo sample which isn't present in the WT sample. This image is presented with increased contrast in Figure 4.12B, with the band denoted by an arrow. However, this band is faint and looks similar in appearance to the multiple non-specific bands on the blot. Notably, a similar band of the same weight is also present in the HaloTag (800nm, green) channel. Two smaller bands of high signal intensity are also observed in the RPS2-Halo lane, confirming the findings of the earlier in-gel fluorescence assay (Figure 4.3). The composite image does not show clear overlap of HaloTag and RPS2 signal. This may be due to the RPS2 antibody N-terminal specific immunogen not being able to interact with the N-terminally

tagged RPS2-Halo fusion protein.

Overall, two-colour fluorescent western blots were an invaluable tool in the PC-12 Halo validation pipeline. pORANGE-RPL4-Halo appears to be working as expected and will progress to the next stages of CRISPR validation. Conversely, the western blot assays provided valuable evidence that the pORANGE-RPL22-Halo plasmid does not appear to produce the desired knock-in effect (an extensive discussion as to the possible identity of this tagged protein is provided in Section 4.3). Lastly, although there was no clear RPS2-Halo band in the RPS2-Halo western, the sequencing data provided convincing evidence that RPS2 is tagged at the correct genomic locus so this line will continue in the validation pipeline.

4.2.2.5 HaloPROTAC3 Degradation Confirms HaloTag Expression in PC-12 Halo Lines

Further validation of HaloTag expression in PC-12 Halo lines was done using a Halo-PROTAC3 assay. Proteolysis targeting chimera (PROTAC) are a class of molecules that can selectively target a protein for cellular degradation (Sakamoto et al. 2001). The two-domain structure of PROTAC ligands allows for the binding of the target by one domain and interaction with E3-ligase through the secondary PROTAC domain. This results in the selective degradation of the target protein. Recently, PROTAC ligands have been developed which recognise and degrade HaloTag fusion proteins (Buckley et al. 2015). Incubation of PC-12 Halo lines with HaloPROTAC3 ligands could therefore provide further evidence of HaloTag ribosomal protein expression.

The HaloPROTAC3 analysis was carried out using RPL4-Halo and RPS2-Halo PC-12 lines (Figure 4.13). Cell lines were incubated with 1 µM Halo-PROTAC3 ligand (Promega, Catalogue No: GA3110) for 3 hours, before the cells were lysed and 75 µg lysate was run on a 10% acrylamide gel. The gel was then transferred to PVDF membrane and probed with RPL4 and RPS2 primary antibodies simultaneously. The blot was cut at 95kDa and the 150kDa- 95kDa blot was probed with KIF5B primary to act as a loading control (see Appendix 1). Band signal intensity

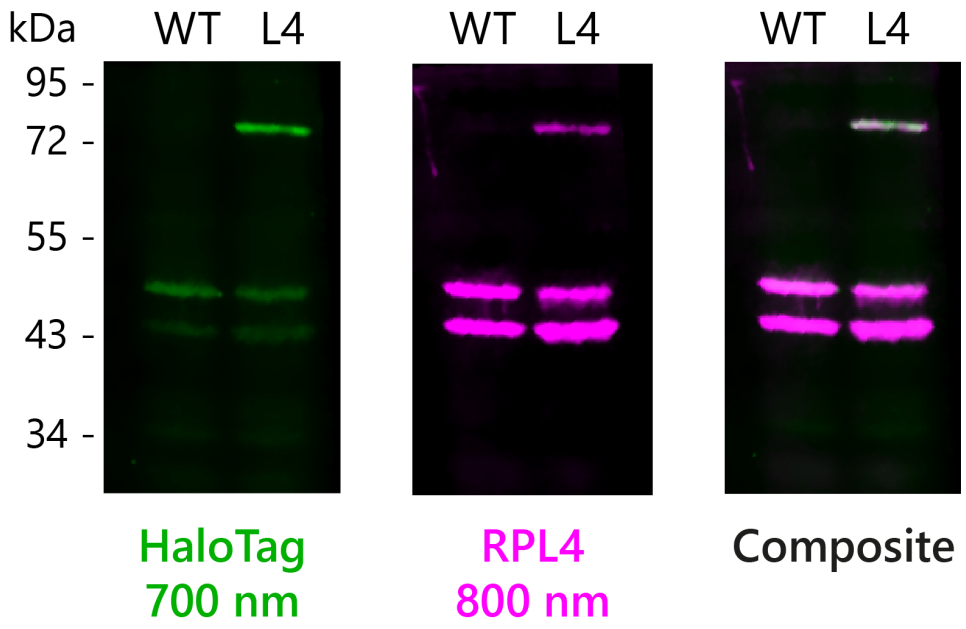


Figure 4.10: RPL4-Halo PC-12 Fluorescent Western Blot.

Cell lysate from stable RPL4-Halo edited PC-12 cells was analysed by two-colour fluorescent western blot to detect the presence of Halo-tagged RPL4 protein. 75 µg of cell lysate was run on a 10% acrylamide gel, transferred to PVDF membrane, and probed with both RPL4 and HaloTag primary antibody. The appropriate fluorescent secondary antibodies were used to ensure signal could be detected in two separate channels. Imaging was done on the LI-COR XF odyssey using the 700 nm (HaloTag) and 800 nm (RPL4) channel. A WT PC-12 sample of the same passage number was run alongside the RPL4-Halo sample. The loading control for this western blot is provided in Appendix 1. Abbreviations: L4 - RPL4-Halo.

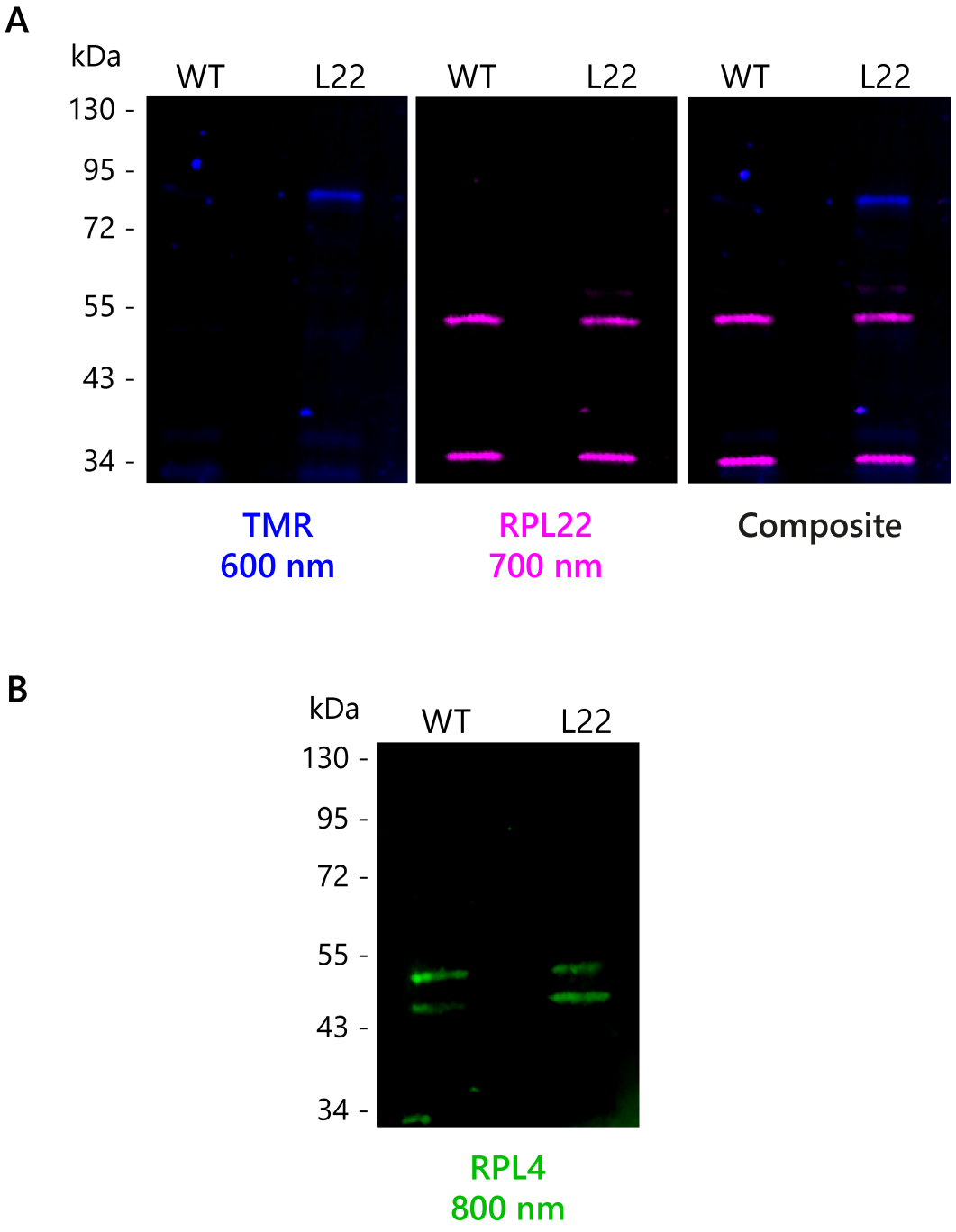


Figure 4.11: RPL22-Halo PC-12 Fluorescent Western Blot.

A) Cell lysate from stable RPL22-Halo edited PC-12 cells was analysed by two-colour fluorescent western blot to detect the presence of Halo-tagged RPL22 protein. Cells were incubated with Halo-TMR for 1 hour before lysis. 75 µg of cell lysate was run on a 10% acrylamide gel, transferred to PVDF membrane, and probed with RPL22 primary antibody. Imaging was done on the LI-COR XF odyssey using the 600 nm (TMR) and 700 nm (RPL22) channel. B) Re-probe of the blot from A using RPL4 primary antibody to determine whether the TMR-labelled protein was RPL4-Halo. A WT PC-12 sample of the same passage number was run alongside the PC-12 Halo edited samples. Abbreviations: L4 - RPL4-Halo, L22 - RPL22-Halo.

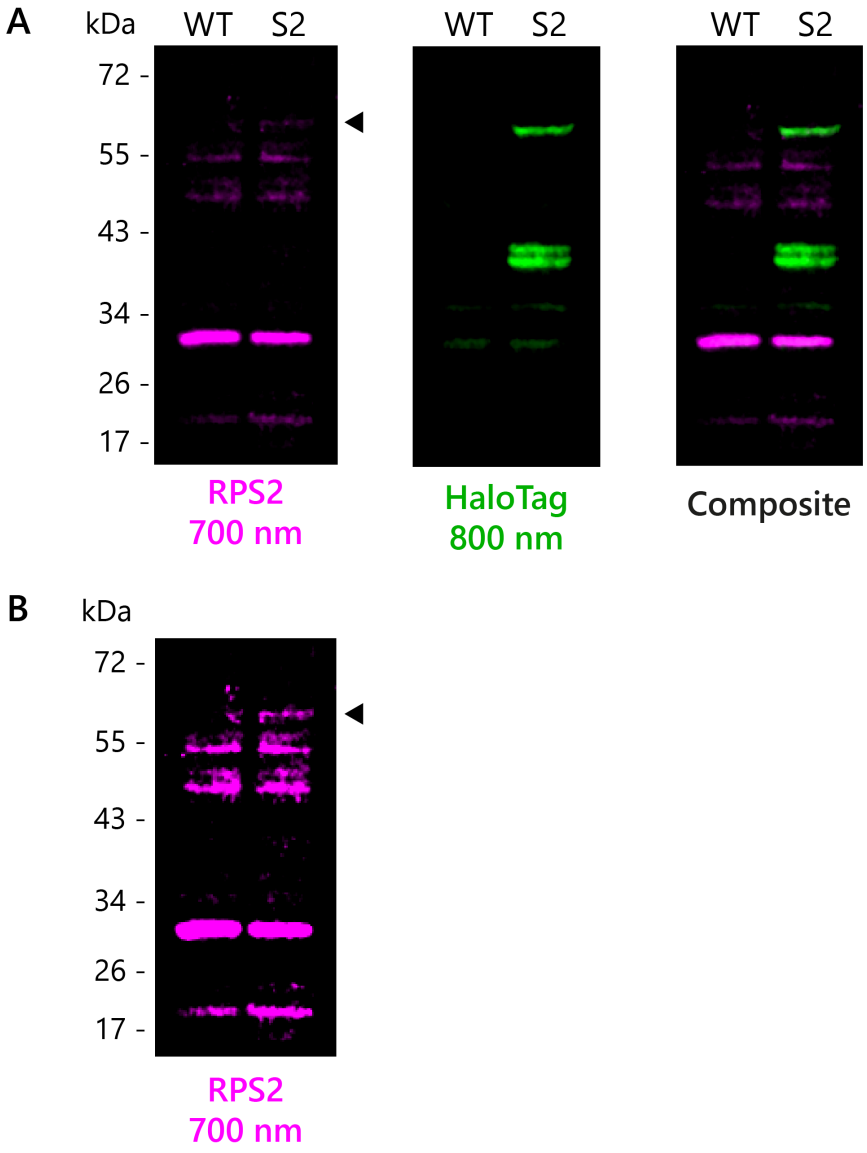


Figure 4.12: RPS2-Halo PC-12 Fluorescent Western Blot.

A) Cell lysate from stable RPS2-Halo edited PC-12 cells was analysed by two-colour fluorescent western blot to detect the presence of Halo-tagged RPS2 protein. 75 µg of cell lysate was run on a 10% acrylamide gel, transferred to PVDF membrane, and probed with both RPS2 and HaloTag primary antibody. The appropriate fluorescent secondary antibodies were used to ensure signal could be detected in two separate channels. Imaging was done on the LI-COR XF odyssey using the 700 nm (RPS2, magenta) and 800 nm (HaloTag, green) channel. A WT PC-12 sample of the same passage number was run alongside the RPS2-Halo sample. The loading control for this western blot is provided in Appendix 1. B) RPS2 channel from Figure 4.12A with boosted contrast to better visualised Halo tagged RPS2 band. Abbreviations: S2 - RPS2-Halo.

was determined using the Analysis tool in LI-COR Image Studio software. Initial optimisation of this assay included a 3 hour ligand incubation before lysis (Figure 4.13A). For the RPL4 channel (800 nm, green), bands at ∼81kDa can be detected for the RPL4-Halo samples which are not present in the WT or RPS2-Halo lanes. The signal intensity for this band decreases in the PROTAC3-treated RPL4-Halo sample (0.044 for RPL4-Halo and 0.026 in RPL4-Halo PROTAC3 samples, 40.1% decrease in signal intensity) which indicates a reduction of RPL4 Halo-tagged protein within the cell. No tagged protein can be detected in the RPS2 channel (700 nm, magenta). This is likely due to poor antibody specificity in recognising the RPS2-Halo protein - as also observed in figure 4.12.

The western blot assay was repeated with a 24 hour ligand incubation (Figure 4.13B) to determine if a bigger reduction in signal intensity could be observed. After 24 hour incubation with Halo-PROTAC3, the WT and Halo-tagged PC-12 lines appeared similar, with no noticeable cell death occurring in the Halo lines. A fluorescent western blot was done as before using RPL4 and RPS2 primaries. Overall, it was found that a longer incubation with Halo-PROTAC3 ligand increased the level of protein degradation observed in RPL4-Halo cells. In the RPL4 channel, the ∼81kDa band intensity is reduced in the RPL4-PROTAC3 sample (signal intensity of 0.040 and 0.003 in RPL4-Halo and RPL4-Halo PROTAC3 samples respectively, a 92.5% reduction in signal intensity). Additionally, a smaller band is observed at ∼34kDa which is likely to be a product of PROTAC3 degradation. As in Figure 4.13A, no RPS2-Halo bands were observed in the RPS2 channel. In summary, observation of HaloPROTAC3-induced degradation of RPL4-Halo protein is additional proof that pORANGE-RPL4-Halo generates the desired knock-in effect in PC-12 cells.

4.2.2.6 Polysome Profiling of Tagged Ribosomes

After confirming that the HaloTag is expressed on the desired protein, it must then be determined if the tagged protein is incorporated into functional ribosomes. As

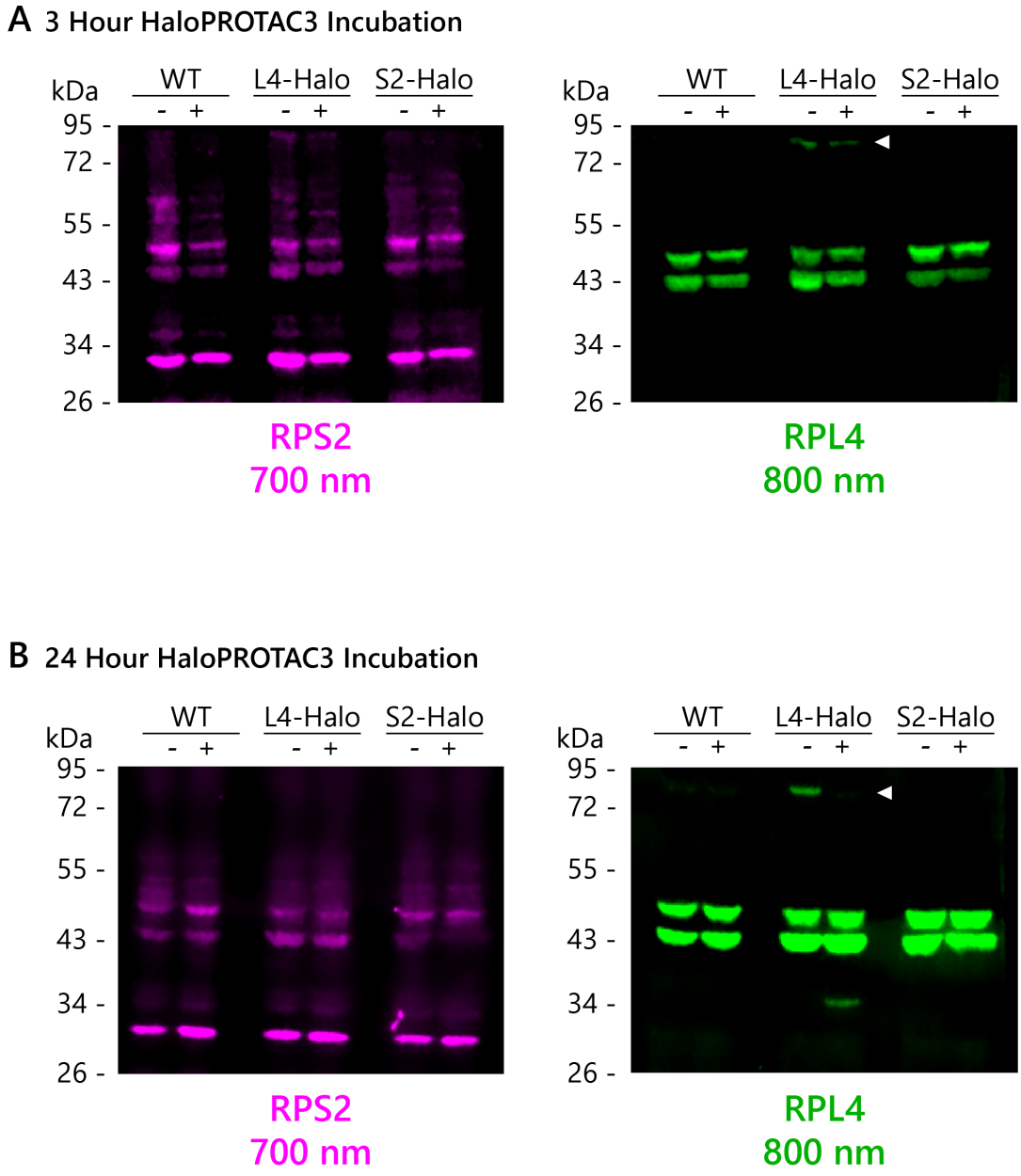


Figure 4.13: PC-12 Halo Line HaloPROTAC3 Treatment Western Blot.

Cell lysate from WT and stable PC-12 CRISPR-edited lines post Halo-PROTAC3 incubation was analysed by two-colour fluorescent western blot to determine if Halo-PROTAC3-mediated degradation of Halo-tagged protein had occured. 75 µg of cell lysate was run on a 10% acrylamide gel, transferred to PVDF membrane, and probed with both RPL4 and RPS2 primary antibody. The appropriate fluorescent secondary antibodies were used to ensure signal could be detected in two separate channels. Imaging was done on the LI-COR XF Odyssey using the 700 nm (RPS2, magenta) and 800 nm (RPL4, green) channel. A WT PC-12 sample of the same passage number was run alongside the CRISPR-edited PC-12 samples. The loading control for this western blot is provided in the Appendix. A) Western blot of samples after a 3 hour incubation with Halo-PROTAC3 ligand. B) Western blot of samples after a 24 hour incubation with Halo-PROTAC3 ligand. N = 1.

previously discussed in Chapter 1, mRNA translation is carried out by polysomes - a group of ribosomes attached to an mRNA transcript. To determine if the Halo-tagged protein is incorporated into functional ribosomes, the polysomes within a cell can be extracted and analysed to identify if the tagged protein is present. The extraction of polysomes from cell lysate is a technique known as polysome profiling, and is a commonly used tool within RNA biology to investigate the translational activity within cells.

Polysome profiling works due to the different sizes (and thus, different sedimentation coefficients) of polysomes and their components. An overview of polysome profiling is given in Figure 4.14. First, RPL4-Halo, RPS2-Halo and WT cells were treated with cycloheximide, a compound that inhibits translational elongation thus pausing the ribosome on the mRNA transcript. The cells were then lysed, releasing polysomes, single ribosomes (monosomes), ribosome subunits (60S and 40S) and free soluble proteins. These distinct populations were then separated by centrifugation for further analysis. The cell lysate was applied to a sucrose gradient, typically ranging from 10% sucrose at the top and 45% sucrose at the bottom. The different concentrations of sucrose enables the separation of the ribosome populations - with proteins (e.g. free ribosomal proteins) at the top and larger complexes (e.g. polysomes attached to mRNA) at the bottom. After centrifugation, the contents of the column was fractionated and the components of each fraction identified by optical density measurement (A260). This generated traces (Figures 4.15 - 4.17) which allow identification of polysome-containing fractions.

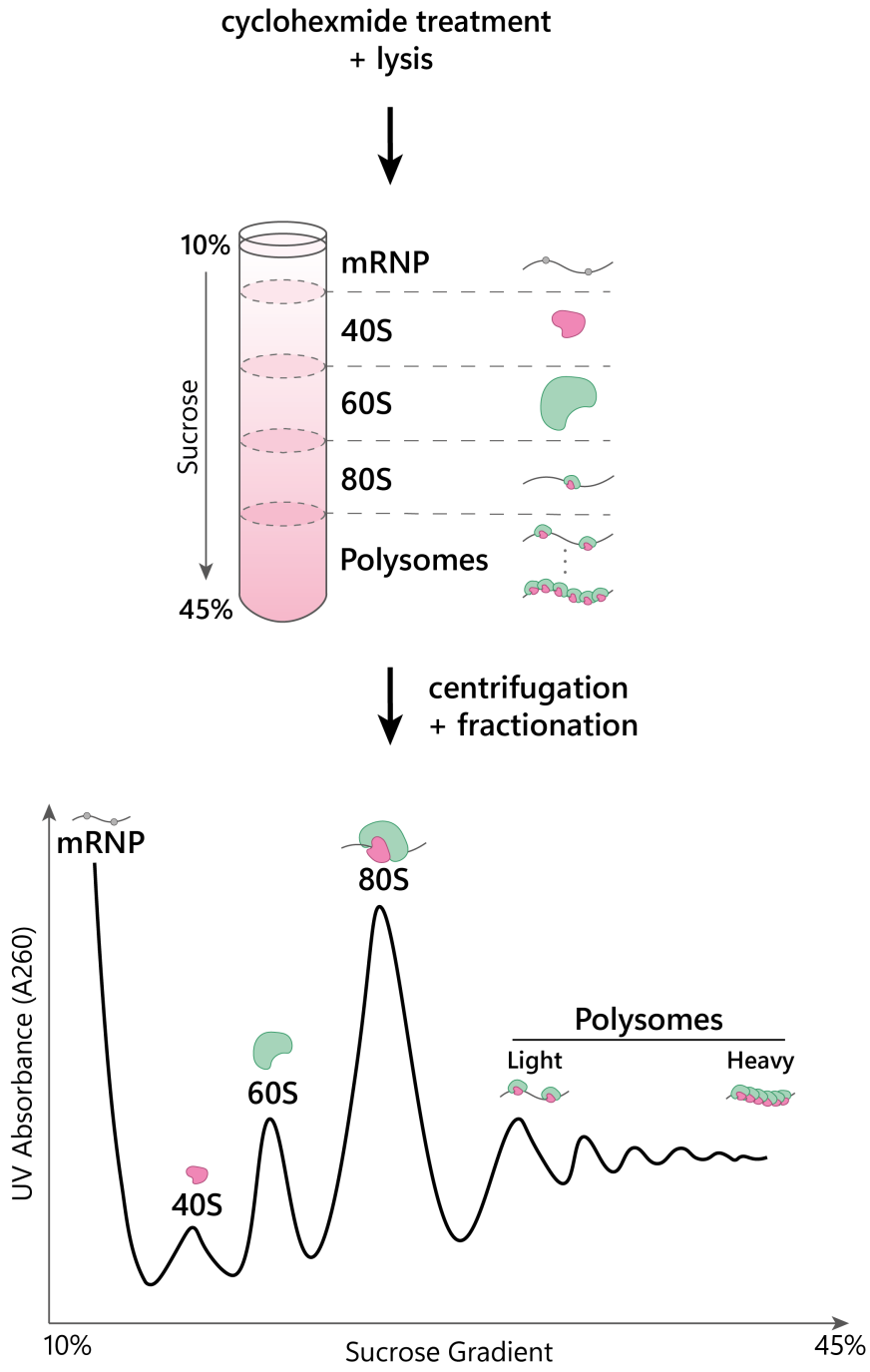


Figure 4.14: Overview of Polysome Isolation Via Sucrose Sedimentation.

Approximately 10x10^6 PC-12 cells were treated with PBS containing 100 µg/mL cycloheximide for 2 min at 37°C. The cells were pelleted and the supernatant removed, before being flash frozen in liquid nitrogen. On the day of the polysome profiling assay, the cell pellets were lysed and the lysate applied to a 10-45% sucrose gradient. The gradient was centrifuged for 2.5 hours to separate the ribosomal populations present within the samples. Fractions from the column were collected and the UV absorbance (A260) of each fraction was detected using a Triax™ Flow Cell and associated software.

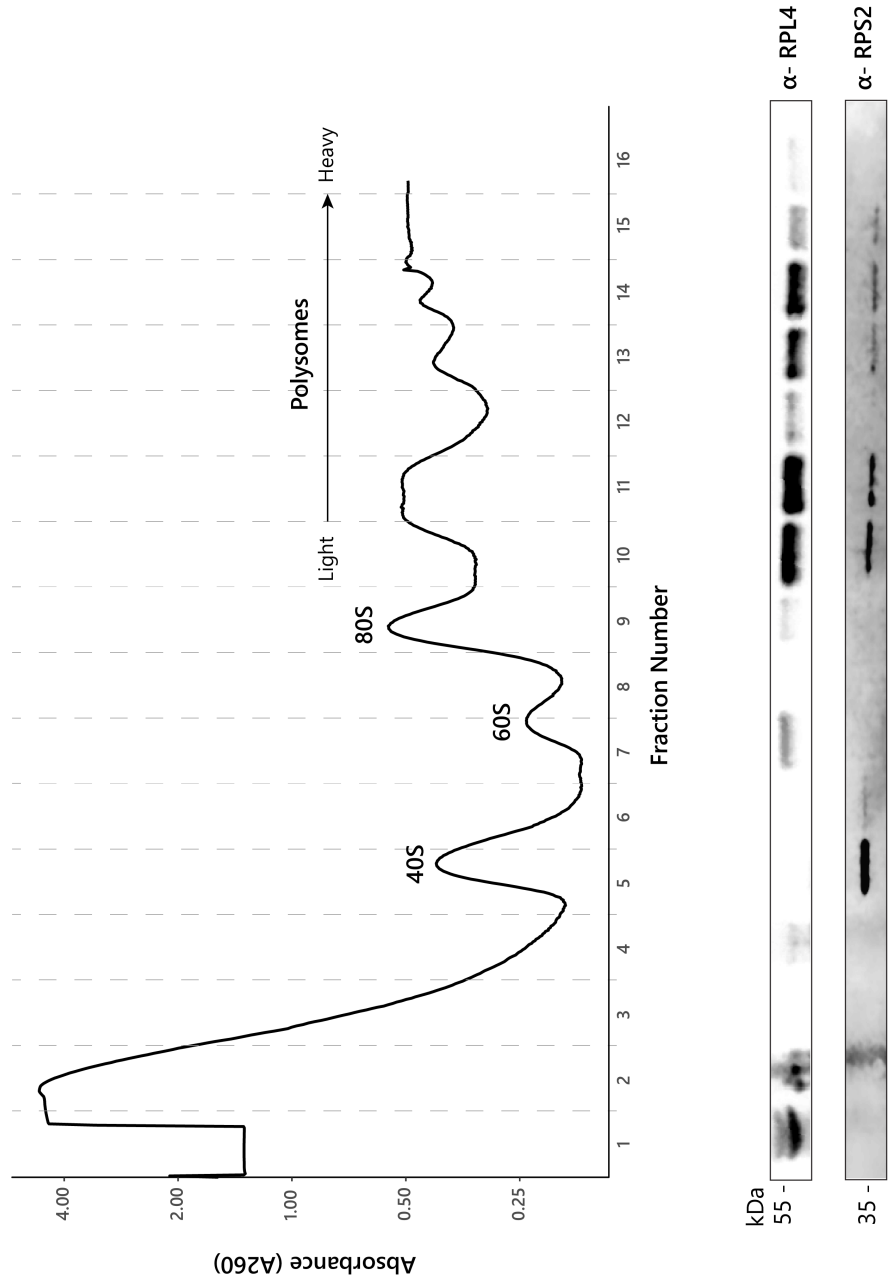

Absorbance (A260)
4.00
2.00
1.00
0.50
0.25
Fraction Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
40S
60S
80S
Polysomes
Light
Heavy
kDa
55 -
35 -
α- RPL4
α- RPS2

Figure 4.15: Polysome Profiling of WT PC-12 Cells.

8-10x10^6 WT PC-12 cells were treated with 100 µg/mL cycloheximide and lysed. The lysate was applied to a 10-45% sucrose gradient and the ribosome components separated by centrifugation. The gradient was subsequently fractionated into 16 fractions, with the absorbance at A260 measured during this process to generate the polysome trace. After protein concentration, the fractions were used in a fluorescent western blot with mouse anti-RPL4 and rabbit anti-RPS2 primary antibodies applied. For methods and antibody concentrations used, see Chapter 2. Raw data was plotted using R (version 4.2.0). N = 1.

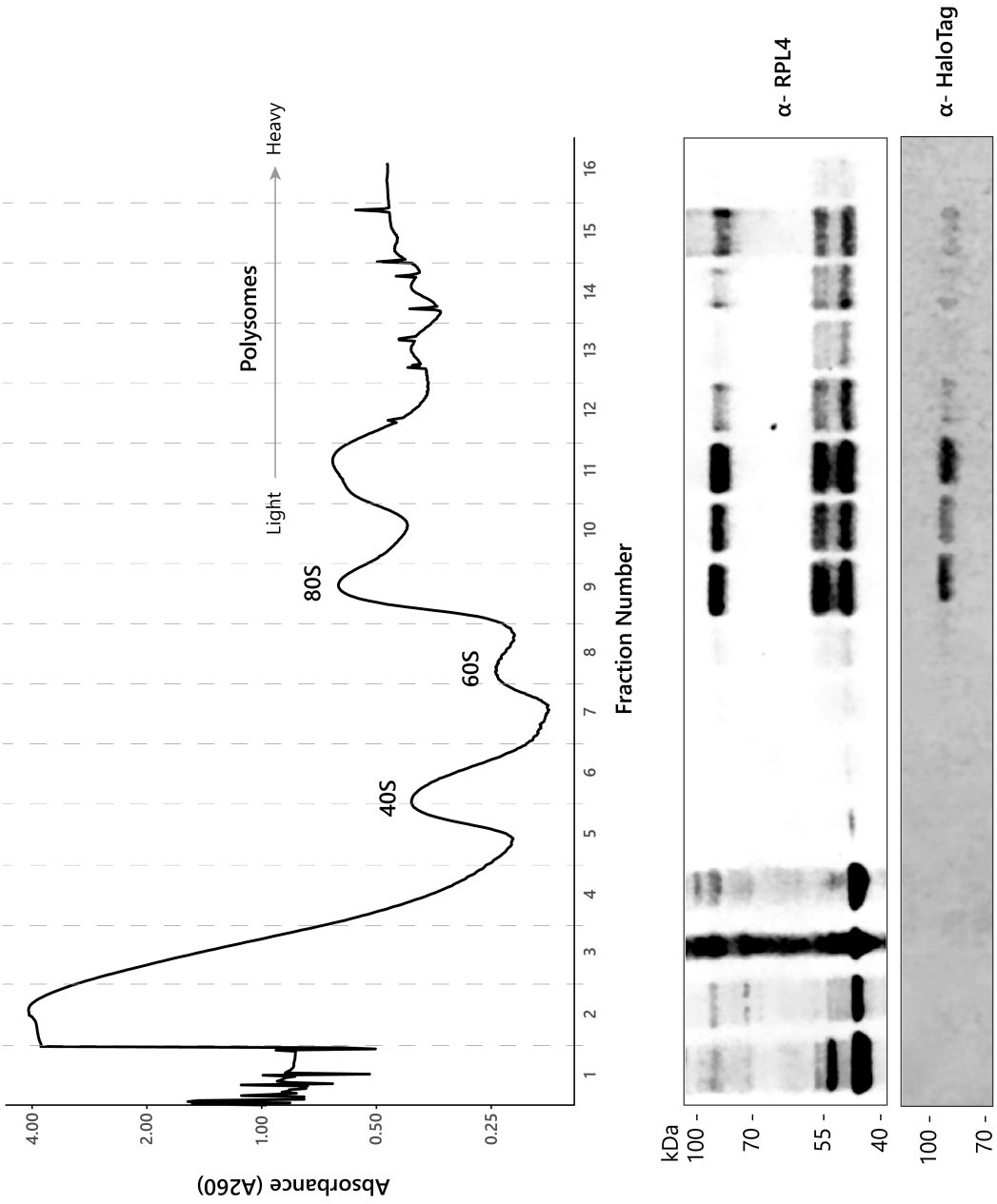

Absorbance (A260)
4.00
2.00
1.00
0.50
0.25
Light
Heavy
Polysomes
40S
60S
80S
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Fraction Number
kDa
100 -
70 -
55 -
40 -
α- RPL4
100 -
70 -
α- HaloTag

Figure 4.16: Polysome Profiling of RPL4-Halo PC-12 Cells.

8-10x10^6 RPL4-Halo PC-12 cells were treated with 100 µg/mL cycloheximide and lysed. The lysate was applied to a 10-45% sucrose gradient and the ribosome components separated by centrifugation. The gradient was subsequently fractionated into 16 fractions, with the absorbance at A260 measured during this process to generate the polysome trace. After protein concentration, the fractions were used in a fluorescent western blot with mouse anti-RPL4 and rabbit anti-HaloTag primary antibodies applied. For methods and antibody concentrations used, see Chapter 2. Raw data was plotted using R (version 4.2.0). N = 1.

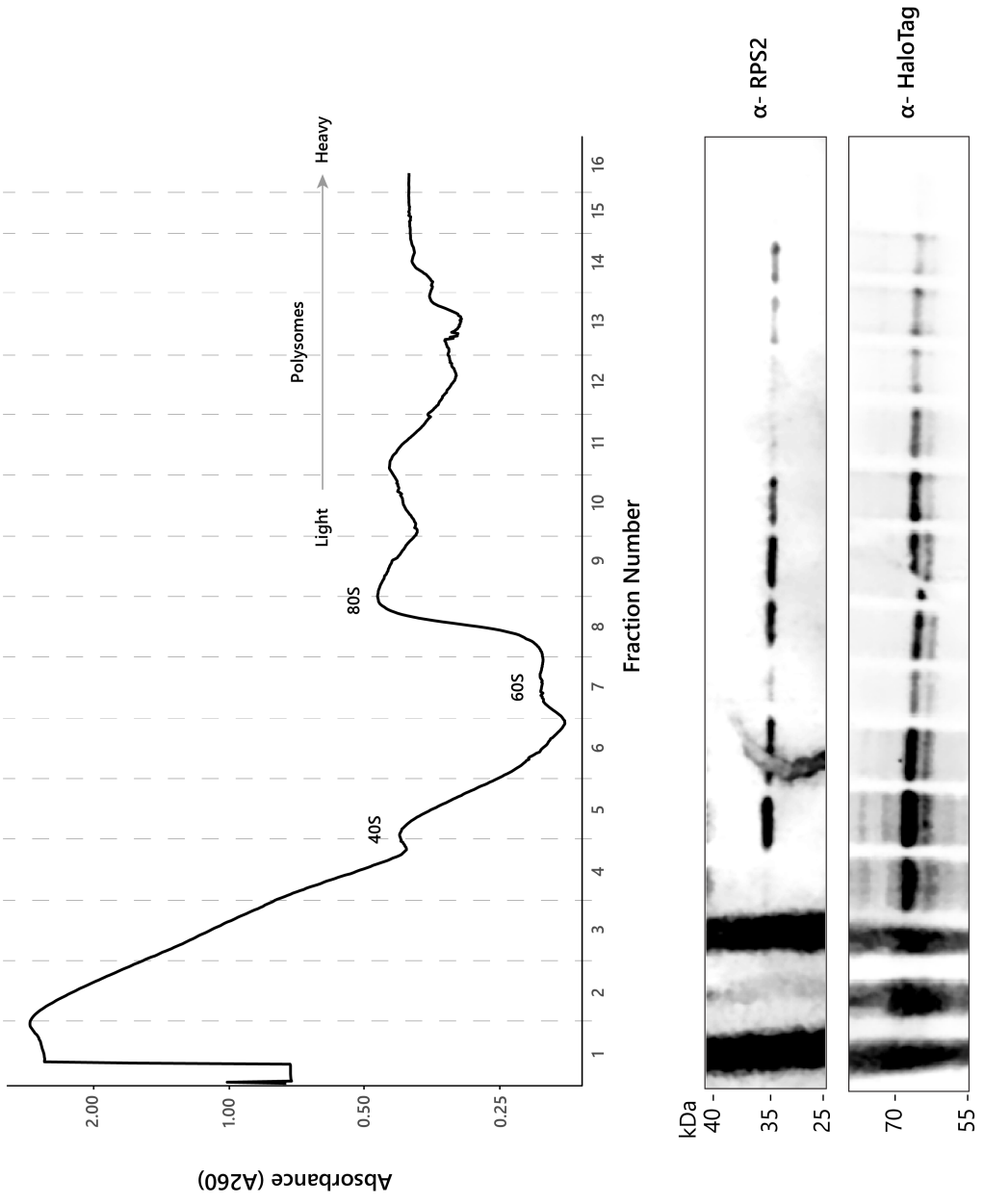

Absorbance (A260)
2.00
1.00
0.50
0.25
Fraction Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
40S
60S
80S
Light
Polysomes
Heavy
α- RPS2
α- HaloTag
kDa
40
35
25
70
55

Figure 4.17: Polysome Profiling of RPS2-Halo PC-12 Cells.

8-10x10^6 RPS2-Halo PC-12 cells were treated with 100 µg/mL cycloheximide and lysed. The lysate was applied to a 10-45% sucrose gradient and the ribosome components separated by centrifugation. The gradient was subsequently fractionated into 16 fractions, with the absorbance at A260 measured during this process to generate the polysome trace. After protein concentration, the fractions were used in a fluorescent western blot with mouse anti-HaloTag and rabbit anti-RPS2 primary antibodies applied. For methods and antibody concentrations used, see Chapter 2. Raw data was plotted using R (version 4.2.0). N = 1.

The polysome profiling traces for each sample looked very similar, which is a good indication that the HaloTag is not interfering with the translational abilities of the cells. In each trace, clear peaks can be seen for the 40S, 60S and 80S ribosome components. The first and second fractions contain an abundance of small, soluble proteins which is observed as a large peak at the beginning of the trace. The polysomes can be identified in the later fractions (fractions 11-16). Surprisingly, the 40S peak has a higher optical density than the 60S peak, which is an unusual observation as the 60S peak is typically higher due to the 60S subunits being larger in size than the 40S subunits. Therefore, it is possible that PC-12 cells have higher amounts of 40S subunits than 60S subunits. Several studies utilising polysome profiling of WT PC-12 cells also corroborate this large 40S peak observation (Koh et al. 2013; Krokowski et al. 2011). Therefore, this is possibly a feature of the PC-12 cell line and not a translational abnormality created during this experimental work.

The protein composition of the polysome fractions was then investigated by western blot. After concentrating the protein, the fractions were run on a 10% acrylamide gel and transferred to a membrane. The membrane was then probed with ribosomal protein or HaloTag primary antibody to determine whether tagged ribosomal protein can be detected in the polysome fractions. The western blot data for each fraction is given under the corresponding fraction trace in Figures 4.15 - 4.17. The WT blot was probed with both anti-RPL4 and anti-RPS2 primary antibodies to show the endogenous ribosomal protein present in WT PC-12 cells. For the Halo edited PC-12 lines, the corresponding ribosomal protein antibody and

complimentary HaloTag antibody were used to visualise both the tagged and untagged ribosomal protein. Details of the polysome profiling protocol and antibody concentrations used can be found in Chapter 2.

For the WT polysome western (Figure 4.15), bands can be observed for both the RPL4 and RPS2 protein at the expected weights. RPL4 protein is present in both the 60S subunit peak (fractions 7 and 8), monosome (fractions 9 and 10) and polysome fractions (fractions 10 - 16). Similarly, RPS2 protein was present in the 40S subunit peak (fraction 5 and 6), monosome (fractions 9 and 10) and polysome fractions (fractions 10 - 16). No RPS2 bands can be seen for fractions 9, 12 or 16, but that may be due to poor antibody signal or high background. Interestingly, the RPS2 western indicates that a large amount of endogenous RPS2 protein is present within the 40S subunit fraction, corroborating the high 40S peak observed in the polysome traces.

For the RPL4-Halo cells, the polysome profiling western demonstrated that tagged RPL4 protein is present within the polysome fractions (Figure 4.16). Two bands for endogenous RPL4 can be observed at ∼48kDa, similar to those observed in Figure 4.10. In addition, heavier RPL4 bands are present at ∼81kDa which is the expected weight for RPL4-Halo. Bands of a similar weight are also detected by the HaloTag antibody. The ∼81kDa RPL4 bands are present in fractions 8-15, although the bands for fraction 8 and 13 have low signal intensity. This western confirms that tagged protein is present in the 60S (fraction 8), monosome (fraction 9) and polysome fractions (fractions 10 - 15), indicating that RPL4-Halo is incorporated into translationally active ribosomes.

RPS2-Halo also shows incorporation into the polysome fractions (Figure 4.17). A 31kDa band for endogenous RPS2 is observed in 40S (fractions 4 and 5), monosome (fractions 8 and 9), and polysome fractions (fractions 10 - 14). A faint band for RPS2 can be seen in the 60S subunit fraction, but this may be due to poor separation of ribosome components - which is also seen in the improper separation of the 40S peak in the trace (fractions 4 and 5). No additional bands for RPS2-Halo were

observed, supporting previous findings that detection of RPS2-Halo by western blot isn't possible with the antibody used (Figure 4.12, 4.13). However, RPS2-Halo bands at the correct weight are detected by the HaloTag antibody in fractions 4 - 14. Notably, this western indicates that most RPS2-Halo protein is present in the 40S fractions (fractions 4 and 5), with smaller amounts present in the polysome fractions (fractions 10 - 14). A discussion of this result is provided in Section 4.3.

4.3 Discussion

The validation of the pORANGE CRISPR plasmids was an essential step before their use in neuronal single molecule imaging experiments. Each CRISPR-edited PC-12 line underwent a robust pipeline of assays to confirm HaloTag expression in the desired protein, at the expected genomic locus, with incorporation of the tagged protein into functional ribosomes. A summary of the results of this chapter are given in Figure 4.18. The RPL4-Halo line delivered a positive result for all validation assays. The pORANGE-RPS2-Halo construct also generates RPS2-Halo protein at the desired genomic locus, which is also incorporated into functional ribosomes. Therefore, these ORANGE constructs will be transfected into rat primary neurons for live cell imaging of ribosome transport. Unfortunately, all other PC-12 Halo lines did not show the intended knock-in result and their respective pORANGE plasmids will not be used in downstream microscopy assays.

Notably, the PROTAC3-Halo and polysome profiling assays were repeated only once, with one biological replicate (N = 1). This was due to time constraints and a shortage of samples in which to repeat assays. Taking this into consideration, these assays would benefit from further experimental repeats to improve the accuracy of the data. In particular, the polysome profiling assays require ideally 3 biological repeats before any firm conclusions are drawn about the amounts of tagged protein in each fraction. However, given the strong signal intensity of the polysome fraction bands in the western blots, it can be assumed that some amount of tagged protein is successfully incorporated into functional ribosomes in PC-12 cells.

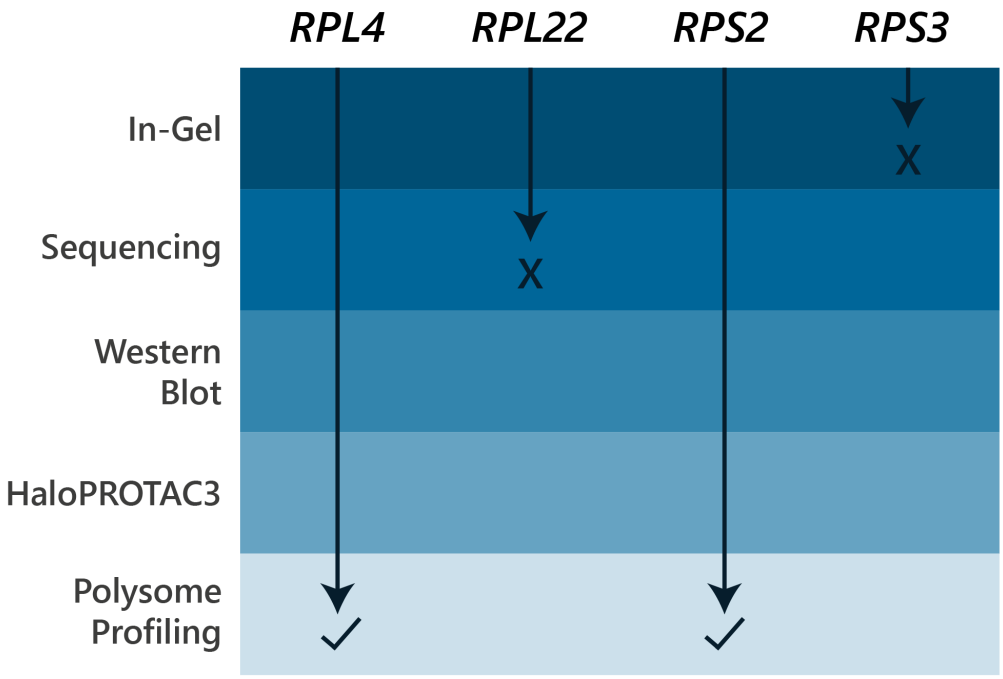


Figure 4.18: Summary of PC-12 Halo Line Validation Results.

An overview of the validation assays conducted for each PC-12 Halo line, with an indication as to which stage in the pipeline the lines were considered to not deliver the desired knock-in effect (if applicable). Abbreviations: L4 - RPL4-Halo, L22 - RPL22-Halo, S2 - RPS2-Halo, S3 - RPS3-Halo.

4.3.1 The RPS3-Halo Line Does Not Stably Express HaloTag

RPS3-Halo failed the validation at the earliest stage, with no bands in the in-gel fluorescence assay indicating a lack of Halo-TMR protein labelling. Given that all cells should have the donor DNA present in at least one RPS3 allele after FACS selection, this points to issues with the cell expansion or expression of the tag in the cell line. It is possible that there was an error with the FACS selection process which caused large numbers of WT cells to be collected. However, this is unlikely as loss of the HaloTag was also seen in the second biological repeat. A more probable hypothesis is that the tagged RPS3 protein production was detrimental to the cell, creating a selective pressure that reduced the number of cells with the knock-in tag over time.

One possible mechanism could be the death of homozygous knock-in cells due to the ribosomal incorporation of tagged RPS3 protein, resulting in non-functional ribosomes and therefore a critical reduction in translation. Additionally, it is possible that a small number of unedited cells - possibly with fluorescence intensity above the gating threshold due to non-specific Halo-TMR staining - were collected during FACS. Due to the RPS3 HaloTag possibly being detrimental, this may have a negative effect on growth rate and unedited cells may divide faster than CRISPR-edited cells. This would lead to unedited cells outnumbering the edited cells during the expansion phase, despite the FACS assay initially collecting a majority population of highly fluorescent CRISPR-edited cells.

4.3.2 The Identity of the Tagged Protein in the RPL22-Halo Line

The identity of the HaloTag expressing protein within the RPL22-Halo PC-12 line also warrants further discussion. There are several possible theories that explain the presence of the ∼81kDa band seen in the in-gel fluorescence assay. The first, and most straightforward explanation, is that a different protein was erroneously targeted and CRISPR-edited to express the HaloTag. In support of this argument,

RPL22-Halo (calculated to be ∼48kDa) would be significantly smaller than the 81kDa band observed. Additionally, the HaloTag primary antibody signal did not overlap with RPL22 antibody signal in the fluorescent western, providing further confirmation that this tagged protein does not appear to be the canonical RPL22.

Conversely, there are several counterarguments to this theory. Off-target effects for the RPL22 guide RNA selected are incredibly unlikely, with RPL22 having the highest specificity score (Hsu et al. 2013) of all guides designed (78.9 out of 100). Five genes were predicted to contain a DNA sequence similar enough to the guide RNA to generate an off-target effect. However, these genes have a low likelihood (specificity scores <0.4) of being edited, with each gene having 4 base mismatches including at least 1 mismatch in the guide RNA seed region - the 8-10 bases upstream of the PAM site which are crucial for Cas9 genomic target recognition. Deep sequencing studies investigating Cas9 off-target effects have shown that 2 or more mismatches in the guide RNA sequence significantly decrease the likelihood of Cas9 cleavage in off-target genomic regions (Cho et al. 2014). Furthermore, even if off-target effects had produced the 81kDa tagged protein, RPL22 is very likely to have also been edited and observed in the in-gel fluorescence assay.

Alternatively, we must then examine the possibility that the 81kDa tagged protein is RPL22, albeit manifesting in an irregular form. Given that the 20 base pair target sequence is present in the *Rpl22* gene in the Rnor_6 rat genome, and that the ORANGE plasmid has been designed with no obvious errors, it is clear that the tagged protein is not RPL22 in the conventional sense. One possibility is that an alternative isoform has been tagged instead. However, upon further inspection of the rat RPL22 isoforms and similar paralogs (such as RPL22L1), no translated proteins larger than the canonical 148 amino acid RPL22 are discovered. Therefore, it is unlikely that a larger RPL22-related protein has been CRISPR-edited. Another theory is that PC-12 cells produce a mutant form of RPL22 that is larger than the canonical 15kDa RPL22 protein, which may explain why the RPL22 primary antibody did not recognise the tagged protein.

Additional assays could help to determine the identity of the Halo-tagged protein in the RPL22-Halo line. In particular, the HaloTag expressing protein could be purified from RPL22-Halo cells using Halo-Trap agarose beads, and the tagged protein subsequently identified by mass spectrometry. Although verifying the identity of the tagged protein is an interesting prospect, it would not significantly contribute to the output of this project. The pORANGE-RPL22-Halo plasmid does not appear to produce the desired knock-in product and thus will not be used for future imaging assays. However, this does not pose a problem to the success of this project. pORANGE-RPL4-Halo has been shown to accurately tag the 60S ribosome subunit and multiple CRISPR plasmids targeting the same subunit are not likely to be required.

4.3.3 PC-12 Cells Express Large Amounts of 40S Subunits

One notable finding of the PC-12 Halo line polysome profiling was that PC-12 cells appear to produce more 40S than 60S subunits. This was determined by the polysome traces (Figures 4.15 - 4.17), whereby the 60S fractions are expected to have a higher UV absorbance than the 40S fractions due to their increased size and larger amount of RNA. A possible explanation is that PC-12 cells have a variation or defect in their ribosome biogenesis pathway which leads to more 40S or less 60S subunits being produced. Although there is no published evidence to support this theory in PC-12 cells, reduction of 60S production has been shown to cause an accumulation of 40S subunits in yeast, perturbing the 1:1 40S:60S ratio that is normally observed (Gregory et al. 2019).

Additionally, the large amounts of RPS2-Halo protein present within the 40S fractions may prove to be problematic in downstream PC-12 imaging assays. The polysome profiling assay determined that a large quantity of tagged RPS2 protein is present in the 40S subunits, with less in the monosome and polysome fractions (Figure 4.17). Therefore, live cell imaging of ribosome transport (i.e. monosomes and polysomes) may be more challenging. However, polysome profiling of rat cortical

neurons does not indicate an abnormal 40S subunit population (Biever, Glock, et al. 2020), so this is unlikely to be an issue when imaging CRISPR-edited primary neurons.

Chapter 5

Visualising Ribosome Transport In Processes Using Live Cell Imaging

5.1 Introduction

Using the ORANGE plasmids validated in the previous chapter, the pORANGE-RPL4-Halo and the pORANGE-RPS2-Halo constructs will be used in live cell imaging experiments to observe the movement of ribosomes in axons. Polysome profiling illustrated that the Halo-tagged RPL4 protein is predominantly incorporated in 80S and polysome fractions (see Chapter 4), therefore, any movement events visualised in axons are unlikely to be free RPL4 protein or individual 40S subunits. In contrast, the RPS2-Halo polysome profiling identified an abundance of tagged free RPS2 protein and 40S subunits, which may result in individual transport events being harder to identify. Additionally, transfections using the positive control ORANGE plasmid (pORANGE GFP-Actb KI #1, Addgene Plasmid #131479) will provide a helpful indication as to whether the ORANGE system is working as expected in cultured neurons.

In order to maximise the likelihood of ORANGE CRISPR-editing, the transfection protocols must be carefully optimised. Generally, the transfection efficiency of primary neuronal cultures is low, although the transfection method selected can

greatly impact transfection success and neuronal health (Alabdullah et al. 2019). Herein, a variety of transfection protocols will be assayed to determine which delivers the highest transfection efficiency whilst also preserving the health of the culture. Initial transfections of rat primary cortical neurons will necessitate a common and easily observable marker (such as GFP) to quickly visualise the transfection efficiency. Notably, the knock-in efficiency of the pORANGE-RPL4-Halo plasmid was calculated at <1% for a heterogeneous culture of electroporated PC-12 cells, which are more easily transfected than neuronal cultures (Covello et al. 2014). Therefore, the low transfection efficiency of primary neurons is expected to be a limiting factor for CRISPR-edited ribosome transport imaging.

Transfected cells will be imaged on a custom-built imaging setup, designed specifically for live cell single molecule imaging. This imaging setup (BONSAI) utilises total internal reflection fluorescence (TIRF) microscopy (Axelrod 1981), a technique which delivers a high signal to noise ratio enabling the visualisation of individual fluorescently-tagged molecules in cultured cells. TIRF microscopy is achieved by firing the laser light at the sample at a specific angle - the incident angle - which hits the sample and is internally reflected back into the coverslip (Figure 5.1). This total internal reflection is due to the different refractive indices of the glass coverslip and the aqueous sample. When the laser light is reflected, an evanescent field is generated in the sample close to the surface of the coverslip. The evanescent field excites any fluorophores in a thin section of the sample (100 - 200 nm above the surface), with the field exponentially decreasing in intensity as the distance from the slide surface increases (Mattheyses and Daniel Axelrod 2006). As a result, TIRF illuminates a narrower Z section of the sample compared to other microscopy techniques (such as confocal microscopy) (Fish 2022), generating a higher signal to noise ratio which is beneficial for live cell imaging.

As TIRF generates an imaging depth of 100 - 200 nm, it is predominantly suited to imaging molecules on the cell membrane. Therefore, imaging of neuronal processes will require the use of near-TIRF or Highly Inclined and Laminated Optical

sheet (HiLo) microscopy (Tokunaga et al. 2008). By changing the excitation laser beam angle to below that of the incidence angle, a thin optical sheet is generated within the sample (Figure 5.1). Like TIRF, this excites fluorophores in a thin z-section of the sample - thus creating lower background noise and better signal-to-noise ratio than some other microscopy techniques. However, the imaging depth for HiLo is larger than TIRF (∼10 µm), making it better suited for imaging molecules further away from the cell surface.

The aim of this chapter is to image axonal ribosome transport in CRISPR-edited primary neuronal cultures and PC-12 Halo lines using the BONSAI live cell imaging system. These transport events can then be quantified to determine the speed and direction of axonal transport. Using this microscopy data, details of the transport mechanism can then be inferred - such as motor protein involvement - which will help build a model for the axonal translocation of cell body-derived ribosomes.

5.2 Results

5.2.1 Optimisation of Primary Cortical Neuron Transfection

The first step of primary neuronal culture imaging optimisation focussed on testing different transfection conditions to determine which protocol generates the highest transfection efficiencies. As transfection with HaloTag constructs may require further labelling optimisation, a vector containing enhanced green fluorescent protein (EGFP) was used in initial transfection assays as it provided several advantages. Importantly, no secondary fluorescent labelling is required with GFP and its variants, therefore, transfected cells should be easily identified within the culture. EGFP is also a commonly used reporter gene for neuronal transfection due to its homogeneous distribution throughout the cytoplasm (Malinow et al. 2010), allowing neuronal processes to be clearly visualised. To this end, rat cortical neurons were transfected with mEGFP-N1 with different protocols to determine which conditions generated

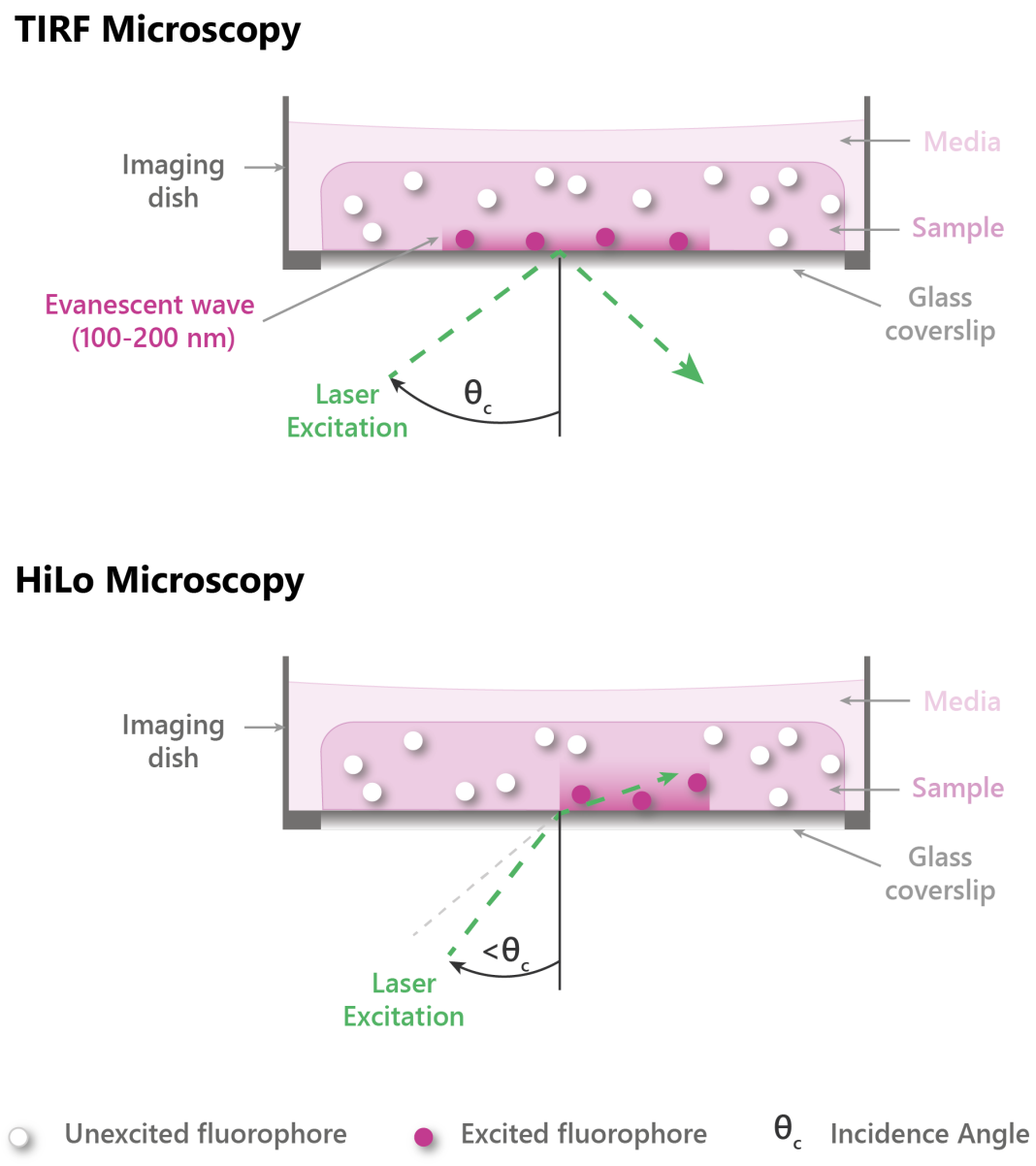


Figure 5.1: Schematic of TIRF and HiLo Microscopy.

A) TIRF microscopy excites fluorophores in close proximity (<200 nm) to the interface of the glass coverslip, with other fluorophores within the sample remaining unexcited. This greatly lowers background fluorescence, improves resolution, and lowers sample phototoxicity (Fish 2022) B) HiLo microscopy uses a excitation laser beam at an angle less than the incidence angle, resulting in an increased penetration depth compared to TIRF microscopy.

the highest number of transfected cells.

The first variable tested was the age of the culture at the time of transfection, whereby transfection at two different time points were assayed - days in vitro (DIV) 4 and DIV 7. Transfections were done using Lipofectamine 2000, a commonly used transfection reagent which has been previously shown to be compatible with ORANGE CRISPR imaging (Willems et al. 2020). Using the reagent amounts and timings from *Willems et al.* (See section 2.1.7.1), rat primary cortical neurons were transfected at either DIV 4 or DIV 7 and were fixed 48 hours post-transfection. Subsequently, an immunofluorescence assay was carried out using anti-GFP primary antibody to amplify the signal from transfected cells for easier detection. DAPI stain was also applied so that the total number of cells could be counted.

Overall, similar levels of transfection were observed for both the DIV 4 and DIV 7 cultures (Figure 5.2A). At a 20x magnification, 6 field of view (FOV) images were randomly captured per condition, using DAPI signal (blue) to locate regions containing a moderate density of cells. EGFP transfected cells (green) could be detected for both DIV 4 and DIV 7-transfected cultures. The transformation efficiency for each image was calculated by dividing the number of EGFP-transfected cells by the number of nuclei (Figure 5.2B). Transfected cells were identified manually by counting the number of cell bodies with intact processes expressing EGFP. Overall, DIV 4 transfected cultures had a higher transfection efficiency per FOV (mean= 1.81%), compared to the DIV 7 culture (mean= 1.34%). Both cultures appear visually healthy before fixation, with cell death due to Lipofectamine 2000 toxicity appearing to be minimal. Overall, this data illustrates that transfection at DIV 4 appears to be more favourable with Lipofectamine 2000. However, as this experiment features data from only one biological replicate (N = 1), no firm conclusions can be drawn.

Additionally, the effect of neuronal culture age on transfection was also tested with Lipofectamine 3000 reagent. DIV 4 and DIV 7 rat primary cortical cultures were transfected with Lipofectamine 3000 reagent following the manufacturer's in-

structions (see section 2.1.7.2). The cells were fixed after 48 hours and underwent the same immunofluorescence and imaging protocol as in Figure 5.2A. For both DIV 4 and DIV 7 cultures, large aggregates of cells were commonly observed (a representative image is provided in Figure 5.2C). Although GFP-transfected cells were detected, the EGFP-positive cells were often present within cellular aggregates which made identifying individual neurons and their axons difficult. Additionally, the overall health of the cultures was poorer compared to cultures transfected with Lipofectamine 2000, which may indicate that Lipofectamine 3000 reagent is more toxic to neuronal cultures. Given the negative effects of Lipofectamine 3000 on culture health and morphology, Lipofectmine 2000 was determined to be the preferable transfection reagent for future imaging experiments.

5.2.2 In-Gel Fluorescence and Western Blot Detection of ORANGE CRISPR-Edited Primary Neurons

Before conducting live cell imaging experiments, it was helpful to determine whether the current transfection conditions could produce enough CRISPR-edited cells to generate signal on a gel. This would then give an indication of the number of cells required for onward mass spectrometry experiments. Given a neuronal transfection efficiency of 1.81% and a CRISPR-editing efficiency of approximately 1% (from FACS assays in transfected PC-12 cells), we would expect to generate approximately 543 CRISPR-edited neurons from the transfection of 3 million neurons. To this end, in gel fluorescence assays of ORANGE-transfected neurons would provide valuable information for the feasibility of downstream proteomic assays whilst also validating that the ORANGE constructs worked in primary neuronal cultures.

Primary neuronal cultures were transfected with ORANGE Halo plasmids and assayed by in-gel fluorescence to determine if tagged ribosomal protein was being expressed. 3 million cortical neurons were plated into PLL-coated 10 cm dishes, with 2 dishes plated per transfection condition (WT, pORANGE-RPL4-Halo, and pORANGE-RPS2-Halo). Using the aforementioned Lipofectamine 2000 transfection

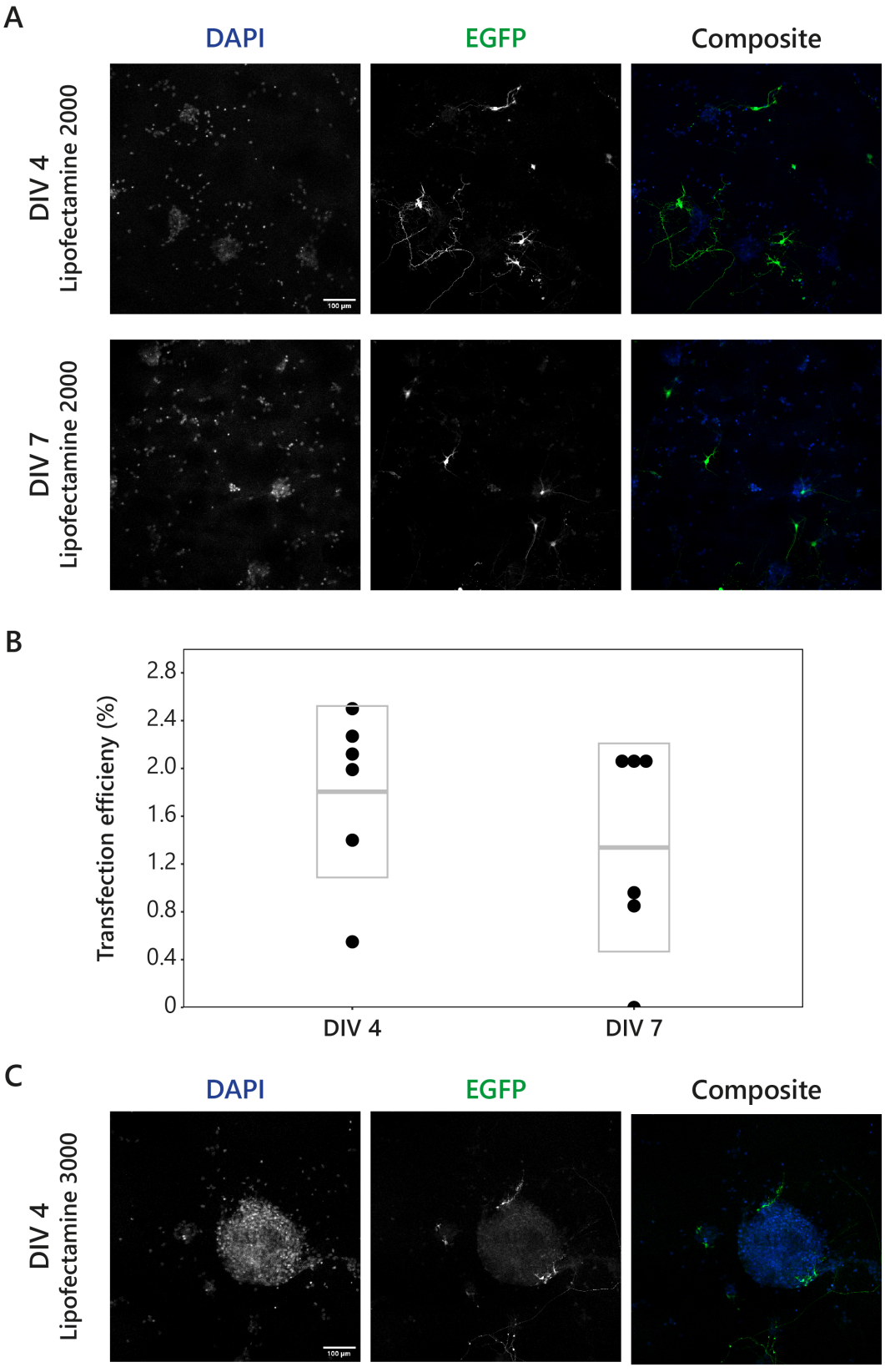

A
DAPI
EGFP
Composite
DIV 4
Lipofectamine 2000
DIV 7
Lipofectamine 2000
100 µm
B
Transfection efficieny (%)
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0
DIV 4
DIV 7
C
DAPI
EGFP
Composite
DIV 4
Lipofectamine 3000
100 µm

Figure 5.2: Cortical Neuron Transfection Optimisation Using Lipofectamine Reagents.

A) Representative immunofluorescence images of rat primary cortical neurons transfected using Lipofectamine 2000 at DIV 4 and DIV 7. Cells were transfected with mEGFP-N1 (Addgene Plasmid #54767), fixed, and probed with α-GFP primary antibody (green). Random sampling was carried out using DAPI signal (blue) to detect cells, with 6 images taken per condition. Images were taken on an SP5 Leica Confocal at 20x magnification. B) Quantification of GFP-transfected neurons per field of view (FOV) for Lipofectamine 2000 transfected primary cortical cultures. Transformation efficiency was calculated by first detecting the number of nuclei per FOV using the ImageJ Threshold and Analyse Particles tools. The number of transfected cells was then divided by the total number of nuclei to generate a percentage of transfected cells in the population. The box plot (grey) denotes the standard deviation of the data (DIV 4= 0.65%, DIV 7= 0.80%), with means represented with the middle bar (DIV 4= 1.81%, DIV 7= 1.34%). Data was plotted using R (version 2.4.0). N = 1. C) Representative image of Lipofectamine 3000 transfected neurons (example presented is DIV 4, 1.5µl Lipofectamine 3000). Abnormal aggregations of neurons were see with all conditions tested with this reagent. Images were taken on a SP5 Leica Confocal at 20x magnification.

protocol as a starting point, the reagent volumes were scaled up to reflect the larger surface area (see Section 2.1.7.1). All transfection timings were kept consistent. Rat cortical neurons were transfected with pORANGE-RPL4-Halo or pORANGE-RPS2-Halo constructs and cultured for 1 week post-transfection to allow donor DNA integration to occur (Willems et al. 2020). The cells were then labelled with Halo-TMR at a final concentration of 100 nM for 1 hour. Following this, the neurons were washed quickly with PBS, scraped into PBS, pelleted and lysed (see Section 2.4.1.1). The lysate was then used in an in-gel fluorescence assay as standard (see Section 2.4.3), with detection of Halo-TMR labelled proteins using the LI-COR XF Odyssey.

No bands were observed in pORANGE-RPL4-Halo or pORANGE-RPS2-Halo transfected neuronal lysate (Figure 5.3A). The gel was subsequently treated with Coomassie stain to visualise protein within the gel, illustrating that a moderate concentration of protein was present and loaded in equal amounts across the samples. Notably, the protein concentration of the cortical neuron samples was low (3-4 mg/mL), therefore, the maximum amount of sample was loaded to increase the likelihood of a band being present (40 µL loaded, approximately 120 µg protein per

sample). Subsequently, the lysis protocol underwent several changes to maximise protein concentration - cells were lysed directly into 250 µL lysis buffer without the PBS wash. Additionally, to check the labelling was working as expected, RPL4-Halo and RPS2-Halo PC-12 samples were labelled and lysed alongside the cortical neurons to act as a positive control. The in-gel fluorescence assay for the PC-12 Halo samples worked as expected, with clear bands observed for the RPL4-Halo and RPS2-Halo fusion proteins (Figure 5.3B). However, as before, no bands are observed for pORANGE-RPL4-Halo or pORANGE-RPS2-Halo transfected neurons, and subsequent Coomassie staining illustrated that protein was present in all samples. This assay determined that the TMR-Halo used was capable of labelling HaloTag proteins and there were no issues with the cell lysis buffer components.

The method of transfection was then considered. Nucleofection of neurons is generally considered to deliver higher transfection efficiencies, with some studies reporting up to 85-95% transfection efficiency in primary neurons (Alabdullah et al. 2019). Therefore, to determine whether changing the transfection method could assist in detecting CRISPR-edited protein, cortical neurons were transfected using the Neon electroporation system. Like other transfection systems, the Neon electroporation parameters must be carefully optimised according to the cell type used. Accordingly, settings suitable for transfection of rat primary neurons were selected from the manufacturer's guidelines (Thermo Fisher Scientific Inc. 2011). As three possible settings were provided, further analysis of published studies identified one setting (1300 V, 20 ms, 2 pulses) as providing the best transfection efficiency with the highest cell viability in DRG neurons (Yu et al. 2015). As tissue was limited, these parameters were a good starting point to determine whether nucleofection could generate tagged protein detectable by in-gel analysis. 6 million neurons per condition were transfected using the Neon system post-dissection (see Section 2.1.2), which were then plated in 6 cm dishes (2 million neurons per dish) and cultured for 1 week. Labelling with Halo-TMR was done as standard, before the cells were lysed and run on an SDS-PAGE gel. As before, no bands were detected in the in-gel

fluorescence assay (Figure 5.3C), despite the Coomassie staining indicating that a high concentration of protein was loaded for each sample.

Subsequently, the ability for transfection products to be visualised via in-gel fluorescence was tested with a vector containing the HaloTag. Therefore, the pHAAP1 plasmid (a vector containing a HaloTag-SNAPtag fusion protein) was selected, as it had previously been used in successful transfections in the lab. It was assumed that pHAAP1 should produce a band as production of the HaloTag is not limited to successful knock-in events in pHAAP1-transfected cells. Transfection of pORANGE-RPL4-Halo and pORANGE-RPS2-Halo was also repeated using new DNA maxipreparations. The same transfection and labelling procedure was carried out in 10 cm dishes, followed by cell lysis and SDS-PAGE protocols. Strikingly, no bands were observed in the in-gel fluorescence assay (Figure 5.4A). However, it must be noted that the lysis step yielded a low concentration of protein in the samples ($\sim$ 1 mg/mL, approximately 40 µg loaded per sample). This was likely due to poor health of the cultures as the rat brain dissection protocol was still being optimised at this stage. Therefore, it is possible that the lack of bands was due to not enough protein being loaded.

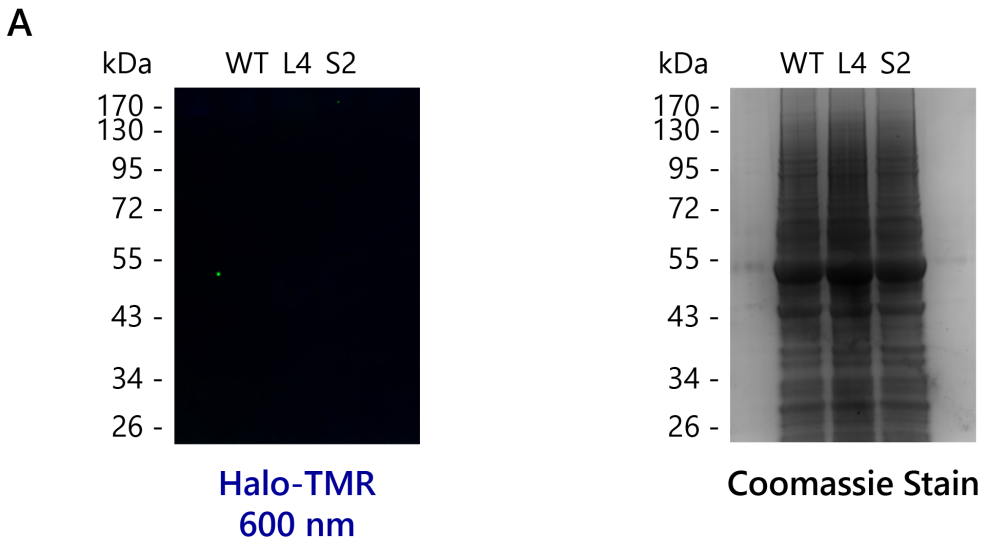
To check that there were no issues with our Lipofectamine 2000 or DNA preparations, a transfection using the same reagents and protocol was carried out by Dr Smith of the Twelvetrees Lab. Using HEK293FT cells plated in a 6-well plate, Dr Smith scaled down the reagents appropriately and transfected the cells with both pHAAP1 and pN-StrepTag-HaloTag (a plasmid encoding HaloTag-StrepTag). The transfected cells were lysed using the standard lysis buffer 2 days post-transfection. Dr Smith used the cell lysate in a western blot (Section 2.4.1.4), probing with both SNAPTag and HaloTag antibodies (Figure 5.4B). Upon imaging, the blots showed the expected bands - pHAAP1 transfected samples had a co-localised band in the HaloTag channel (green, 800 nm) and the SNAPTag channel (magenta, 700 nm) at $\sim$52kDa for HaloTag-SNAP expression. pN-StrepTag-HaloTag transfected samples showed band in the HaloTag channel at $\sim$34kDa for Strep-Halo. This western blot

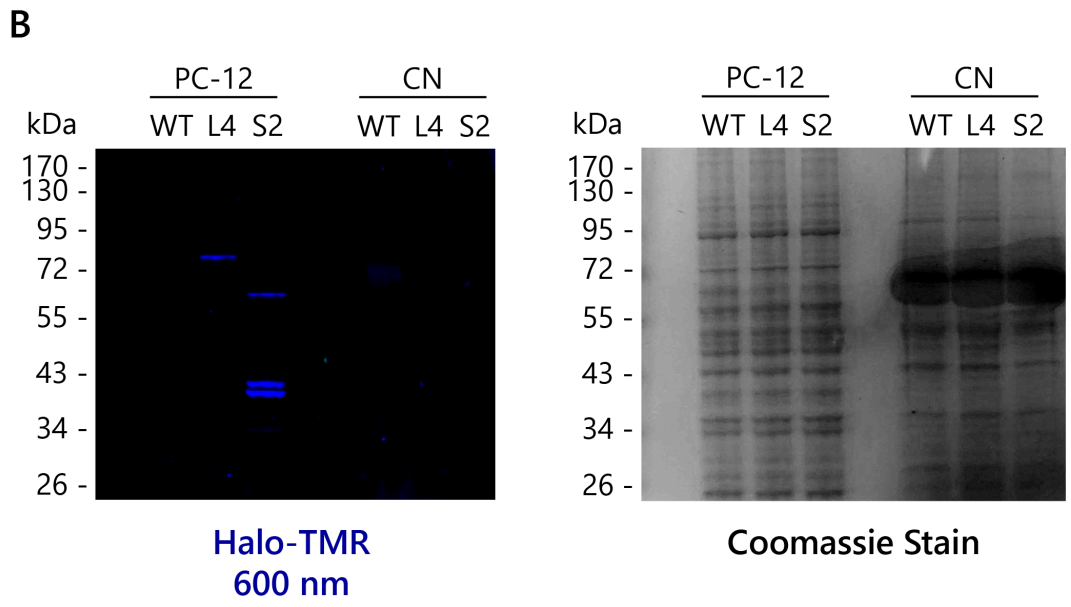
showed that there were no obvious problems with the reagents used in the cortical neuron transfections.

Since transfection was possible with the current protocol and reagents used, the transfections were repeated to replicate the format of Dr Smith's transfections. This assay utilised a plasmid that has been previously successfully transected into cortical neurons (mEGFP-N1) to determine whether transfection was more successful in smaller volumes. Cortical neurons were seeded into a 6 well plate and transfected using the same reagents and timings as before (Section 2.1.7.1). As EGFP is expressed rapidly after transfection, the cells were lysed after 2 days and a western blot was done as standard. No bands were observed (Figure 5.5). Given the difficulties visualising transfection in cortical neuron samples by in-gel fluorescence or western blot, it was proposed that identification of exogenous or CRISPR-mediated protein expression may not be possible using these methods without further optimisation. Therefore, subsequent assays focussed on live cell imaging experiments, as it was possible that CRISPR-edited neurons were present in transfected cultures - albeit at extremely low levels.

5.2.3 Live Cell Imaging of ORANGE CRISPR Edited Primary Neurons

Following on from the conclusions drawn from the western blot and in-gel fluorescence assays, transfected cortical neurons were imaged using live cell near-TIRF microscopy to see if transfected neurons could be observed. Rat primary cortical neuron cultures were seeded at a high density into live cell imaging dishes and transfected with pHAAP1. Transfections were done using Lipofectamine 2000 as before, with the same timings and reagents scaled down to account for the smaller surface area of the live cell dishes. Two days post-transfection, the cells were labelled with Halo-TMR (see Section 2.8.5) for 1 hour. The media was then changed to wash out unbound ligand. Before imaging, the media was replaced with imaging media (Section 2.8.5). Cells were imaged on the custom-build BONSAI imaging setup using


A
kDa
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Halo-TMR
600 nm
kDa
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Coomassie Stain


B
PC-12
CN
kDa
WT L4 S2
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Halo-TMR
600 nm
PC-12
CN
kDa
WT L4 S2
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Coomassie Stain

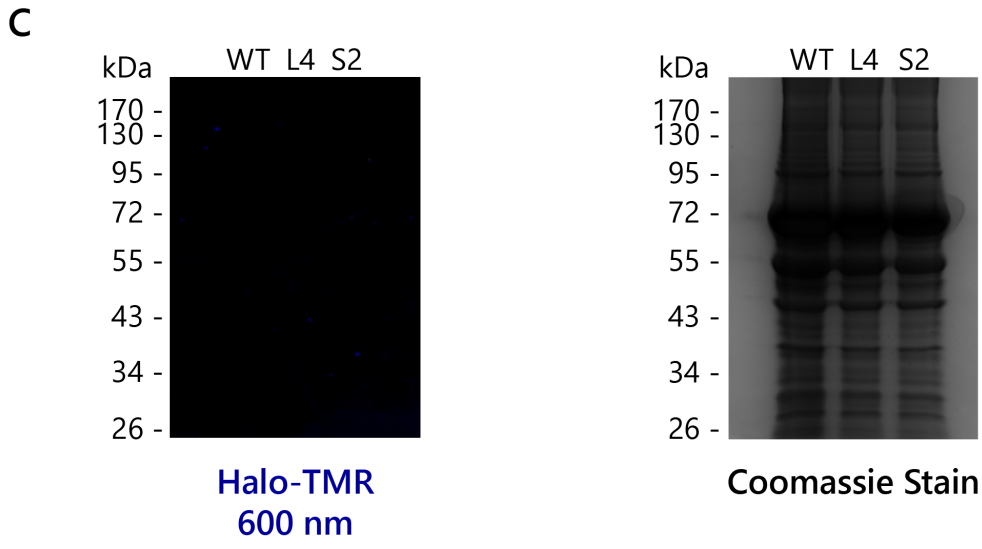

C
kDa
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Halo-TMR
600 nm
kDa
WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Coomassie Stain

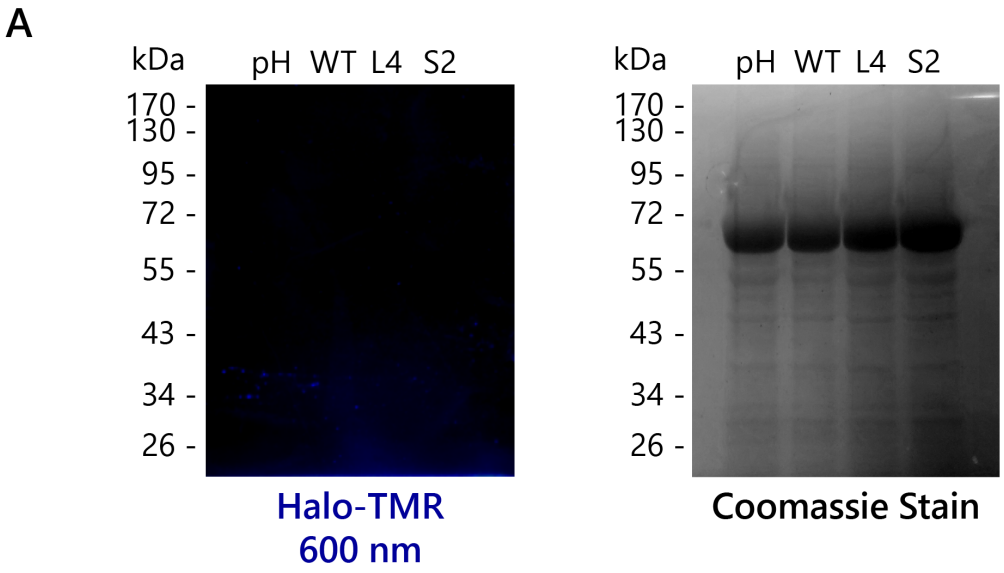
Figure 5.3: In-Gel Fluorescence Analysis of Cortical Neuron Transfection with pORANGE-Halo Plasmids.

Primary rat cortical neuron cultures were transfected with Lipofectamine 2000 and cultured for 1 week. Cultures were labelled with 100 nM Halo-TMR for 1 hour, before the cells were lysed and the lysate run on an 10% acrylamide SDS-PAGE gel. WT samples of the same number DIV were also prepared using the same protocol and run alongside the transfected samples. Gels were imaged using the LI-COR XF Odyssey using the 600 nm channel. Subsequently gels were then treated with Coomassie stain to visualise protein concentration and imaged on a G:Box (Syngene). A) In-gel fluorescence assay of cortical neurons transfected with pORANGE-RPL4-Halo and pORANGE-RPS2-Halo using Lipofectamine 2000. Neurons were plated into a 10 cm dish at a density of $3x10^6$ cells per dish. B) In-gel fluorescence assay to detect transfection of cortical neurons pORANGE-RPL4-Halo and pORANGE-RPS2-Halo using Lipofectamine 2000, run alongside TMR-labelled PC-12 Halo lysate. Neurons were plated into a 10 cm dish at a density of $3x10^6$ cells per dish before transfection. PC-12 Halo lines were grown in suspension, approximately $3\text{-}4x10^6$ cells were labelled and lysed. C) In Gel fluorescence assay to detect transfection of cortical neurons pORANGE-RPL4-Halo and pORANGE-RPS2-Halo using the Neon electroporation system. $6x10^6$ cells were electroporated per condition, with cells being subsequently plated at a density of $2x10^6$ cells in 6 cm dishes. Cells were then cultured, labelled, and lysed as previously described. Abbreviations: L4 - pORANGE-RPL4-Halo transfected cells, S2 - pORANGE-RPS2-Halo transfected cells, CN - cortical neurons.

the 561 nm laser.

Transfected neurons could clearly be observed within the culture (Figure 5.6). In a tile scan of 13,200 μm^2, 2 transfected neurons were identified (Figure 5.6A). The presence of pHAAP1-positive cells confirms that the transfection and TMR-Halo labelling protocols work as expected in cortical neuronal cultures. Areas of high cell density were also labelled by Halo-TMR ligand, however, this was not an issue as signal from background staining was lower than that of the transfected neurons. Additionally, close up imaging of a transfected neuron (Figure 5.6B) illustrates that Halo-TMR labelling generates a strong signal that can be visualised in neuronal processes as well as cell bodies, if the labelled protein is present within neurites. This data was a promising sign that it may be possible to identify ORANGE-edited cortical neurons using near-TIRF microscopy.

To determine whether the ORANGE CRISPR system was working as intended in rat primary cortical cultures, the pORANGE GFP-ACTB KI#1 (Addgene Plasmid #131479, Willems et al. 2020) plasmid was transfected into neurons as a positive


A
kDa
pH WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Halo-TMR
600 nm
kDa
pH WT L4 S2
170 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
Coomassie Stain

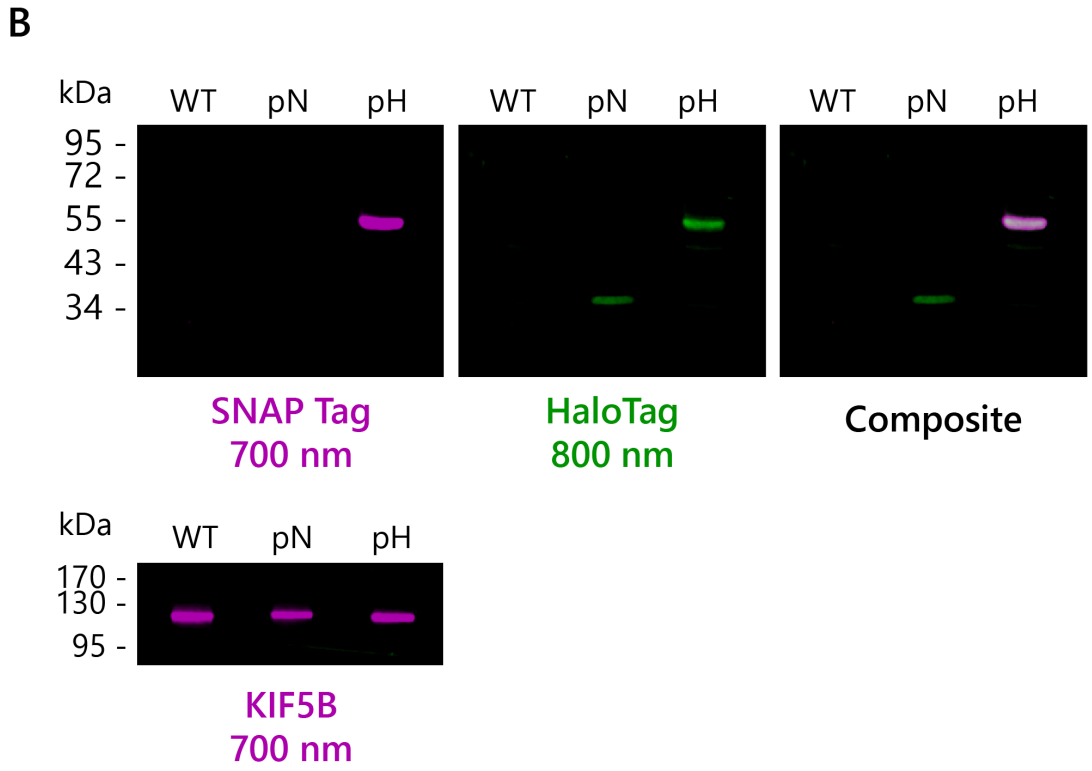

B
kDa
WT pN pH
WT pN pH
WT pN pH
95 -
72 -
55 -
43 -
34 -
SNAP Tag
700 nm
HaloTag
800 nm
Composite
kDa
WT pN pH
170 -
130 -
95 -
KIF5B
700 nm

Figure 5.4: Detection of pHAAP1 Transfection in Cortical Neurons and HEK293FT Cells.

A) Primary rat cortical neurons were plated into a 10 cm dish at a density of 3x10^6 cells per dish and transfected with pORANGE-RPL4-Halo, pORANGE-RPS2-Halo and pHAAP1 using Lipofectamine 2000. Cells were cultured for 1 week then labelled with 100 nM Halo-TMR for 1 hour, before the cells were lysed and the lysate run on an 10% acrylamide SDS-PAGE gel. WT samples of the same number DIV were also prepared using the same protocol. Gels were imaged using the LI-COR XF Odyssey using the 600 nm channel. Subsequently, gels were then treated with Coomassie stain to visualise protein concentration and imaged on a G:Box (Syngene). B) Dr Smith transfected HEK293FT cells with pHAAP1 and pN-StrepTag-HaloTag using Lipofectamine 2000. 48 hours post-transfection, cells were lysed and used in a fluorescent western as standard. Primary antibodies used were rabbit α-SNAPtag, mouse α-HaloTag and rabbit α-KIF5B as a loading control. Blots were imaged using the LI-COR XF Odyssey using the 700 nm (SNAP, KIF5B, magenta) and 800 nm (HaloTag, green) channels. Abbreviations: L4 - pORANGE-RPL4-Halo transfected cells, S2 - pORANGE-RPS2-Halo transfected cells, pH - pHAAP1 transfected cells, pN - pN-StrepTag-HaloTag transfected cells.



Figure 5.5: Cortical Neuron Transfection With mEGFP-N1 Using Lipofectamine 2000.

Neurons were plated into a 6 well plate at a density of 0.5x10^6 cells per well. Primary rat cortical neuron cultures were transfected with mEGFP-N1 using Lipofectamine 2000. 48 hours post-transfection, cells were lysed and the lysate run on an 10% acrylamide SDS-PAGE gel. WT samples of the same DIV number were also prepared using the same protocol and run alongside the transfected samples. A fluorescent western blot was done as standard using mouse α-GFP primary antibody. The blot was imaged using the LI-COR XF Odyssey using the 800 nm channel. Abbreviations: GFP - mEGFP-N1 transfected cortical neurons.

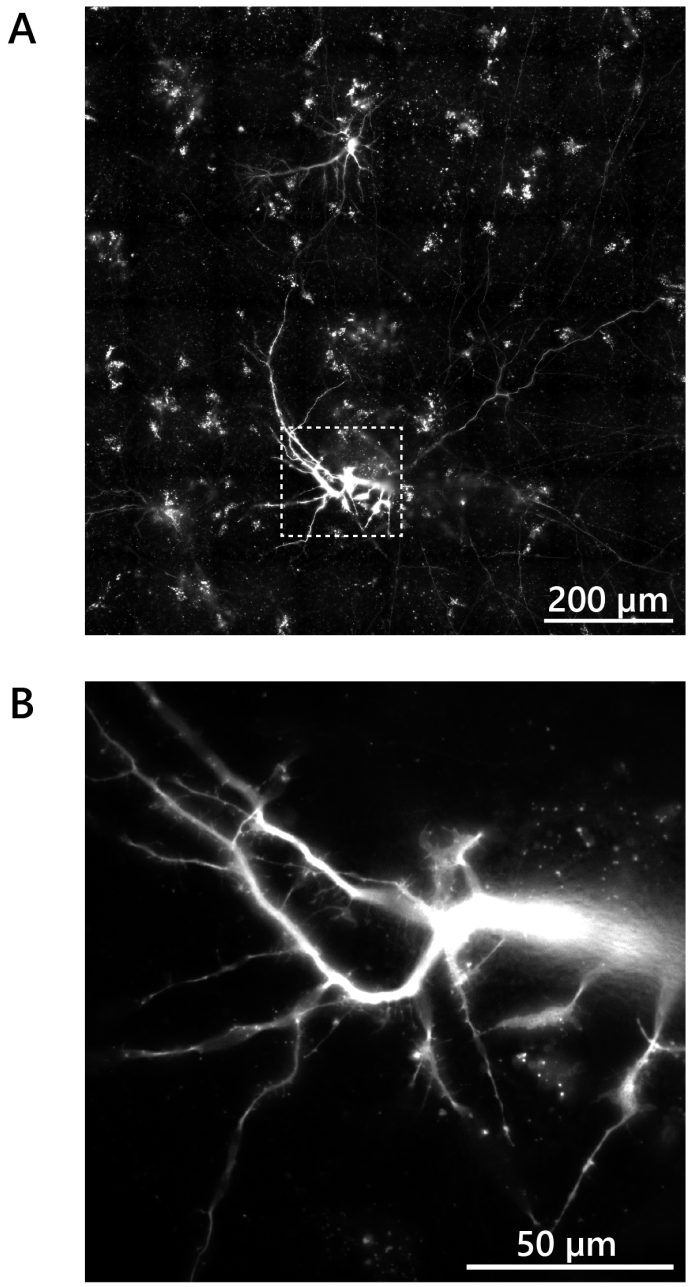


Figure 5.6: Live Cell Imaging of pHAAP1 Transfected Cortical Neurons Labelled with Halo-TMR.

Live cell imaging of rat primary cortical neurons transfected with pHAAP1 was done on BONSAI using the TIRF 561 setting and 100x magnification. Cells were prepared for imaging as outlined in Section 2.5.8, and labelled with 50 nM Halo-TMR for 1 hour. The 561 laser was used at 10% power (max power = 53 mW). Images were prepared using ImageJ. A) Tilescan image showing transfected neuron, with fluorescence above background levels. B) Closeup image of the cell in the boxed region in 5.6A, captured with 500 ms exposure.

control. As low transfection efficiency was expected, cortical neurons were seeded into live cell imaging dishes at a high density (see Section 2.8.5). Using the BONSAI imaging system on the TIRF 488 setting, neurons expressing GFP-tagged β-actin were successfully identified (Figure 5.7A). The GFP-tagged β-actin is highly expressed within the cell body, with uniform distribution of the signal in neuronal processes as is expected for β-actin neuronal localisation. Areas of high cell density showed low levels of auto-fluorescence in the 488 channel, which are also present in the untransfected WT cortical neuron negative control (Figure 5.7B).

A small region of the transfected neuron then underwent fluorescence recovery after photobleaching (FRAP) analysis to determine the health of the culture and see whether the movement of GFP-tagged β-actin could be detected. Using the 488 laser at 20% power, a process was photobleached and imaged to illustrate the movement of GFP-tagged β-actin into the bleached region (Figure 5.7C). GFP signal can be seen to return to the process over a period of approximately 20 seconds, with GFP-tagged β-actin appearing to be transported in a predominantly anterograde fashion (the bottom of images in Figure 5.7C is the direction of the cell body). This finding is in line with previous findings that β-actin is transported by slow axonal transport in an anterograde manner (Chakrabarty et al. 2019). Overall, cortical neurons can be successfully transfected with pORANGE GFP-ACTB KI#1 plasmid, with the desired knock-in effect. However, it must be noted that the low transfection efficiency, coupled with CRISPR editing efficiency of the plasmid (Willems et al. 2020), resulted in very few edited neurons. Across two biological replicates, 14 live cell dishes were imaged on the BONSAI system, culminating in the detection of only 2 GFP-tagged β-actin neurons. With a plating density of 0.18×10^6, the CRISPR editing efficiency observed in these assays is 0.0001%, although low transfection efficiency is likely to also be a factor.

Subsequently, rat cortical neurons were transfected with pORANGE-RPL4-Halo. A similar protocol was used as in Figure 5.7, with the notable addition of a labelling step. One week post-transfection, neurons were incubated with JFX650-Halo flu-

orescent ligand (Janelia), for 15 minutes at a final concentration of 50 nM. After incubation, the media was replaced and the ligand was washed out for an additional 15 minutes. Neurons were then imaged using the TIRF 632 setting on the BONSAI imaging setup. Fluorescent JFX650-labelled cells could be observed within the cultures at a low frequency (Figure 5.8A, 5.8B). Upon closer inspection, these labelled cells lacked the distinct morphology of mature cortical neurons (Figure 5.8A, 5.8B, closeup image). However, the presence of processes indicates that the cells could be CRISPR-edited glia. Overall, no CRISPR-edited neurons were observed in these imaging assays (10 live cell imaging dishes transfected, seeded at a $0.18x10^6$ density, across 2 biological repeats).

Notably, this assay had high background levels of fluorescence. Imaging of WT untransfected cortical neuronal cultures determined that this signal was not from autofluorescence of the cells (Figure 5.8C). However, when WT untransfected cells are labelled with JFX650 following the same protocol, a similar background fluorescence signal is observed (Figure 5.8D). Although areas of high cell density showed moderate fluorescent signal, labelled cells could clearly be observed due to their higher fluorescence intensity. Future imaging experiments using high densities of cells may need further optimisation when using the JFX650 ligand to improve the signal to noise ratio.

5.2.4 Live Cell Imaging of pORANGE-RPL4-Halo Differentiated PC-12 Cells

As no RPL4-Halo positive neurons were identified in the previous primary neuronal culture transfections, live cell imaging was subsequently conducted with differentiated PC-12 Halo lines. As previously discussed (see Chapter 3), PC-12 cells can be differentiated into a model of sympathetic ganglion neurons after treatment with NGF. Upon differentiation, PC-12 cells grow axon-like processes expressing axonal and synaptic markers such as GAP-43 and synapsin-1 (Hu et al. 2018). Additionally, the previous validation of these cell lines (see Chapter 4) indicates that the

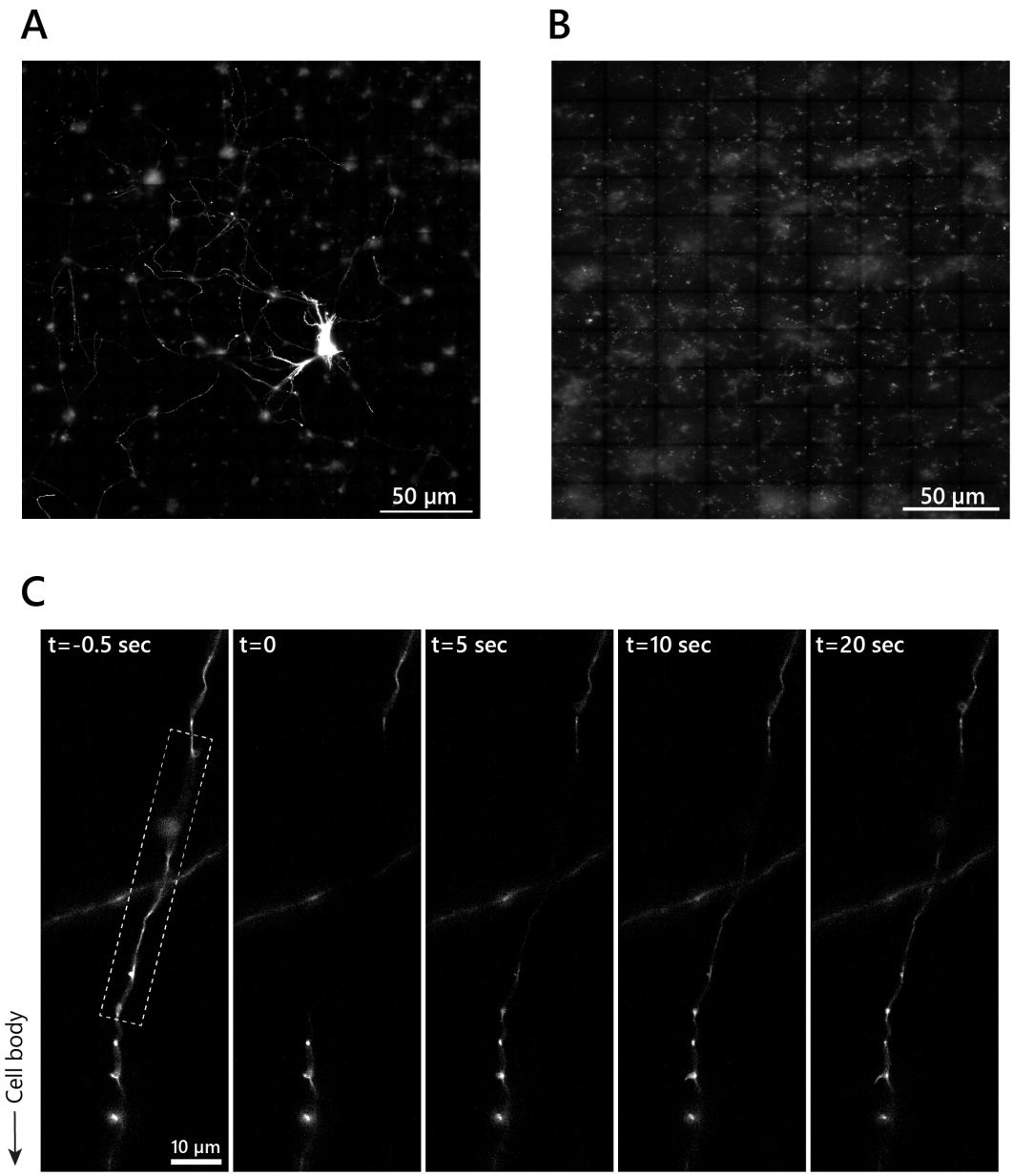


Figure 5.7: Live Cell Imaging of pORANGE-GFP-ACTB Transfected Cortical Neurons.

Live cell imaging of rat primary cortical neurons transfected with pORANGE-GFP-ACTB #1 was done on BONSAI using the TIRF 488 setting and 100x magnification. Cells were prepared for imaging as outlined in Section 2.5.8. The 488 laser was used at 10% power (max power = 61 mW). Images were prepared using ImageJ. N = 2. A) Tilescan image showing a transfected neuron, with fluorescence above background levels. B) Tilescan of unlabelled WT cortical neurons to demonstrate low background autofluorescence with the 488 laser. C) Timelapse of FRAP, where t=0 is the first frame after the photobleaching step, taken at 100 ms exposure. The region denoted by the box in the first image represents the photobleached region. Photobleaching was done using the 488 laser at 20% power (max power = 61 mW) with 1 repetition.

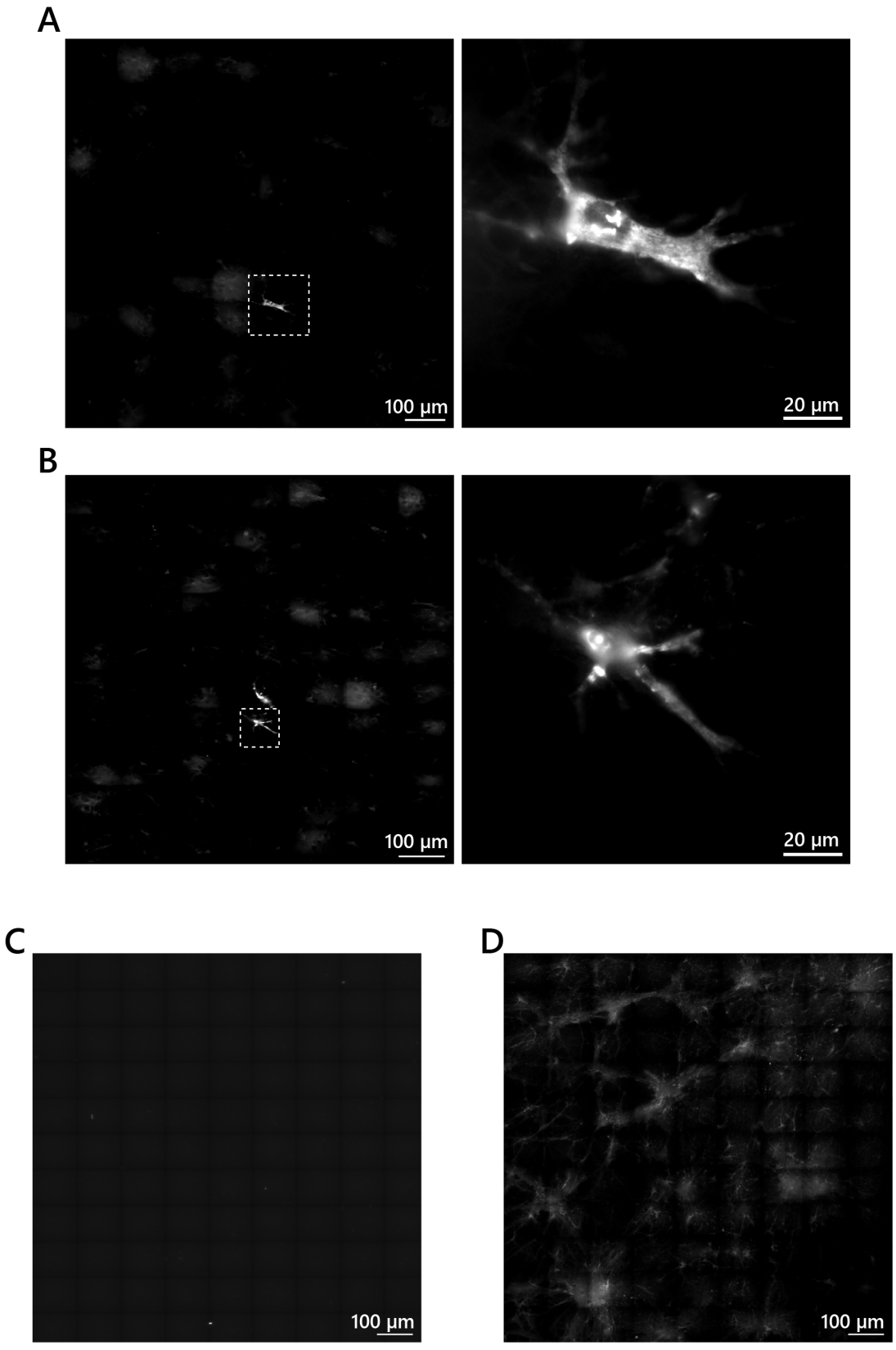

A
100 µm
20 µm
B
100 µm
20 µm
C
100 µm
D
100 µm

Figure 5.8: Live Cell Imaging of pORANGE-RPL4-Halo Transfected Cortical Neurons Labelled with JFX650-Halo.

Live cell imaging of rat primary cortical neurons transfected with pORANGE-RPL4-Halo was done on BONSAI using the TIRF 632 setting and 100x magnification. The 638 laser was used at 10% power (max power = 98 mW). Images were prepared using ImageJ. Cells were prepared for imaging as outlined in Section 2.5.8, and labelled with 50 nM JFX650-Halo for 15 minutes. N = 1. A) Tilescan image showing a possible transfected cell, with fluorescence above background levels. Second image is a closeup of the cell in the boxed region in 5.8A, captured with 500 ms exposure. B) Tilescan image showing two possible transfected cells, with fluorescence above background levels. Second image is a closeup of the cell in the boxed region in 5.8B, captured with 500 ms exposure. C) Tilescan of unlabelled transfected cortical neurons to demonstrate low background autofluorescence. D) Tilescan of JFX650-Halo labelled WT cortical neurons to demonstrate that areas of high cellular density are stained with JFX650.

RPL4 and RPS2 PC-12 Halo lines express a moderate quantity of tagged ribosomal protein which is incorporated into monosomes and polysomes. Therefore, live cell imaging of these lines may provide useful information about ribosome movement in neuronal model systems.

Before live cell imaging was conducted, it was useful to determine how RPL4 was localised in differentiated PC-12 cells to ensure that tagged protein could be detected in PC-12 processes. Therefore, an immunofluorescence assay using differentiated RPL4-Halo PC-12 cells was carried out. WT and RPL4-Halo PC-12 cells were differentiated for 1 week (see Section 2.1.3) and fixed, before being used in an immunofluorescence assay using anti-RPL4 and anti-HaloTag antibodies (see Section 2.8.3). DAPI stain was also applied to visualise the nuclei. Subsequently, the fixed cells were imaged on BONSAI using the 405-Widefield (DAPI stain), 488-Widefield (RPL4 primary) and 561-Widefield (HaloTag primary) settings.

Both WT and RPL4-Halo cells treated with both antibodies showed strong signal for RPL4, present in both the cell bodies and the processes (Figure 5.9). As expected, only RPL4-Halo cells had a clear signal from the HaloTag primary, which is also observed within the PC-12 processes. The single channel controls show low signal background staining with the HaloTag primary in WT samples. Overall, this assay illustrates that Halo-tagged RPL4 protein is present within PC-12 processes

and should be detectable using live cell imaging on the BONSAI system.

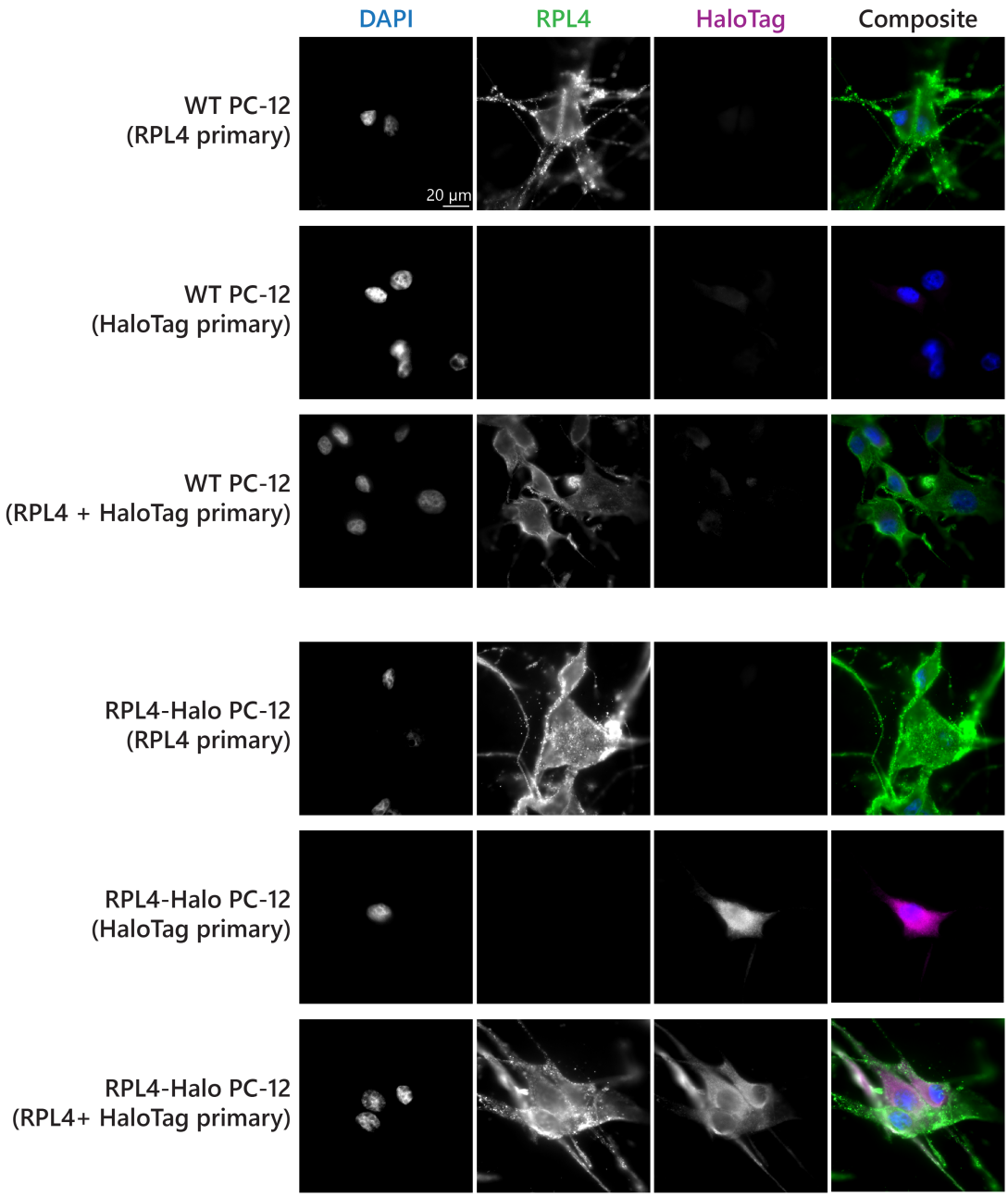


Figure 5.9: Differentiated RPL4-Halo PC-12 Immunofluorescence Assay.

WT and RPL4-Halo cells were plated on acid-washed 13mm converslips in a 6-well plate at a density of 18,200 cells per well. Cells were differentiated for 1 week using NGF. Immunofluorescence was carried out as standard using mouse anti-RPL4 and rabbit anti-HaloTag primary antibodies. Secondary antibodies used were donkey anti-mouse Alexa Fluor Plus 488 and donkey anti-rabbit Alexa Fluor Plus 594. For suppliers and concentrations used, see Table 2.20. Imaging was done on BONSAI using widefield settings at 100x magnification, using the 405 (DAPI, blue), 488 (RPL4, green) and 561 (HaloTag, magenta) lasers.

For live cell imaging, PC-12 Halo cells were plated into Ibidi µ-Dish 35 mm glass bottom imaging dishes at a low density (7,000 cells/dish) and differentiated using NGF. On day 6 of differentiation, the cells were labelled with 50 nM Halo-TMR for 1 hour. The media was then replaced and the unbound ligand was washed out for approximately 2 hours. As the differentiation media does not contain phenol red, replacement with imaging media was not required. Labelled cells were imaged on BONSAI using the TIRF 561 setting. Further details on the BONSAI system and live cell imaging can be found in Section 2.8.5.

Early RPL4-Halo PC-12 imaging experiments identified the movement of Halo-tagged particles in PC-12 processes (Figure 5.10). These particles were likely to be ribosomes, as the polysome profiling experiments (see Chapter 4) determined that RPL4-Halo protein was predominantly present in the monosome and polysome fractions. TMR-labelled particles in processes could clearly be observed in all cells (Figure 5.10A), although the signal intensity and amounts of labelled ribosomes varied from cell to cell - possibly due to the heterogeneous nature of the line. Although some bright particles could be observed, identifying the movement of single particles was challenging. Therefore, a small region of the process was photobleached so tagged ribosomes could be observed clearly (Figure 5.10B). In the seconds that followed, ribosomes were seen to move across the photobleached region, in both an anterograde and retrograde direction.

Particle velocities were quantified with the creation of a kymograph (Figure 5.10C). Each frame of the video is presented as a line in the image, with particle displacement and time on the x and y axis respectively. After the photobleaching, TMR-labelled particles can be seen traversing the bleached region (represented by black diagonal lines across the white central region). In contrast, particles that are not moving are displayed as black, vertical lines that show no change in displacement. The velocity of the transport events can be quantified from the kymograph by measuring the slope of the diagonal lines (see Section 2.8.6 for more details, Table 2.22 shows data used for all kymograph quantification). Overall, the parti-

cles were found to have an average velocity of 1.21 µm/second (calculated based on Figure 5.10C, based on 4 transport events in one process, across one biological repeat), which aligns with the recorded velocity for fast axonal transport (Maday et al. 2014).

Optimising the labelling and imaging techniques to enable clearer detection of labelled single particles was a priority. Therefore, subsequent RPL4-Halo imaging assays utilised Janelia Fluor fluorescent ligands, TMR analogues which are brighter and more photostable (Grimm et al. 2017). By labelling the cells with 50 nM JF549 and 0.05 nM JFX650 (a 100 fold difference in concentration), processes could be detected using the JF549 ligand and single particles imaged using the lower concentration of JFX650 ligand. Additionally, the imaging settings were also changed to stream (whereby data from the camera is saved straight onto the RAM, thus a faster mode of imaging) to determine whether this improved the detection of individual transport events.

Labelled ribosomes were again observed in PC-12 processes (Figure 5.11A). This assay utilised a variation of stream mode - livestream - which captures the last several seconds of data from the camera before the photobleaching program begins. Livestream mode was selected to capture an infrequent transport event, whilst also demonstrating that the most particles in the processes remained stationary. Although single particles were more easily observable than when cells were labelled with Halo-TMR alone, the number of transport events appeared to be greatly reduced. To this end, FRAP analysis illustrated that no particles moved into the photobleached area during the duration of the video (Figure 5.11B). However, the low signal intensity of the JFX650 (and therefore, lower signal to noise ratio) also made visualisation of transport more challenging.

The movement of particles in the FRAP assay timelapse was illustrated in a kymograph (Figure 5.11C). Several short transport events were detected (see boxes), however, most other particles remain stationary. No clear movement across the photobleached region is observed. One possible explanation for this is that ribo-

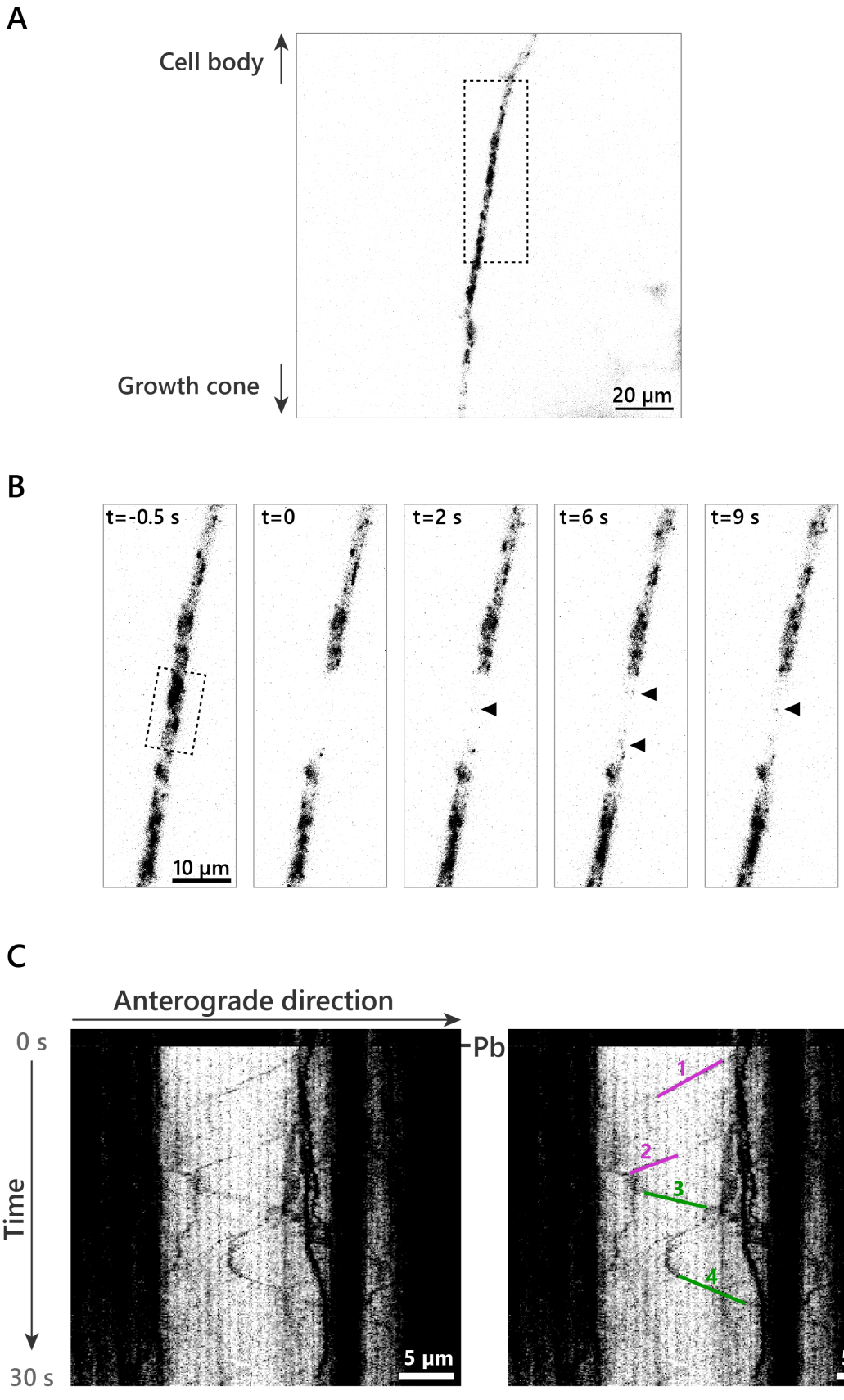

A
Cell body
Growth cone
20 µm
B
t=-0.5 s
t=0
t=2 s
t=6 s
t=9 s
10 µm
C
Anterograde direction
0 s
Time
30 s
Pb
1
2
3
4
5 µm
5 µm

Figure 5.10: Live Cell Imaging of RPL4-Halo PC-12 Line Labelled with Halo-TMR.

Live cell imaging of differentiated RPL4-Halo PC-12 cells was done on BONSAI using the TIRF 561 setting, 50 ms exposure, and 100x magnification. Images were prepared using ImageJ. Cells were prepared for imaging as outlined in Section 2.5.8, using 50 nM JFX650-Halo. N = 1. A) The middle of a PC-12 process showing TMR-labelled particles. B) Timelapse of FRAP, where t=0 is the frame after the photobleach step ended. FRAP was done using the 561 laser at 10% power (max power = 53 mW) and 1 repetition. Photobleached region is denoted by the white box in the first image. Images are a magnification of the boxed region in 5.15A. Times were rounded to the nearest second. C) Kymograph of particle movement during FRAP assay. The photobleaching event is denoted by "Pb". The second image is an example of an annotated kymograph for particle velocity calculations. Anterograde and retrograde transport events are shown by green and magenta lines respectively. Data for kymograph velocity calculations is provided in Table 2.22. Frame rate = 0.24 seconds/frame.

some transport events within processes are relatively infrequent, with most particles remaining stationary. Therefore, by using a low concentration of JFX650, less ribosomes were labelled and thus the frequency of observable transport events decreased. However, JFX650-labelled ribosomes appeared brighter than Halo-TMR labelled ribosomes, therefore, a higher concentration of this ligand was used for all subsequent live cell imaging assays.

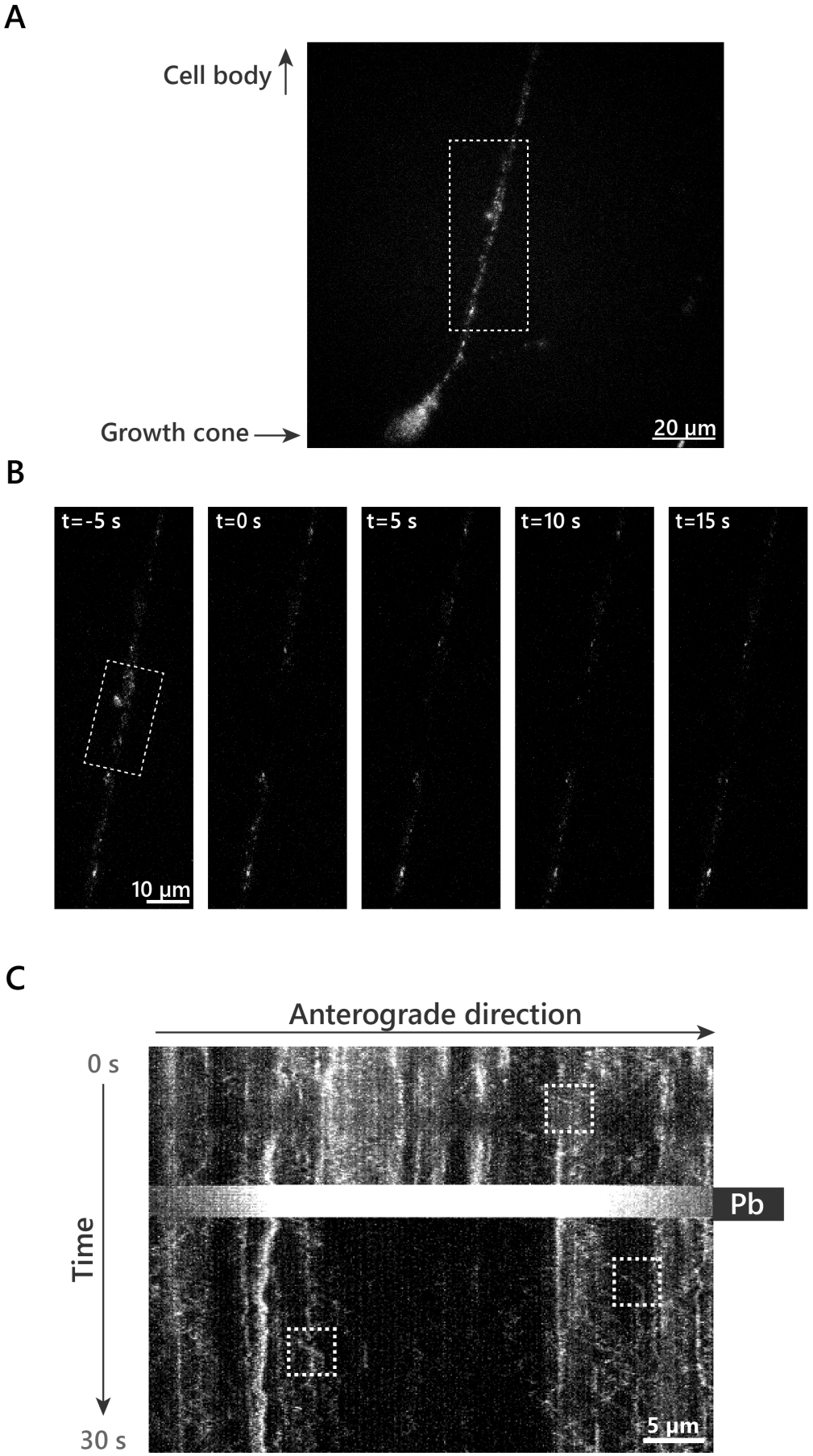

A
Cell body
Growth cone
20 µm
B
t=-5 s
t=0 s
t=5 s
t=10 s
t=15 s
10 µm
C
Anterograde direction
0 s
Time
30 s
Pb
5 µm

Figure 5.11: Live Cell Imaging of RPL4-Halo PC-12 Cells Using 0.05 nM JFX650-Halo.

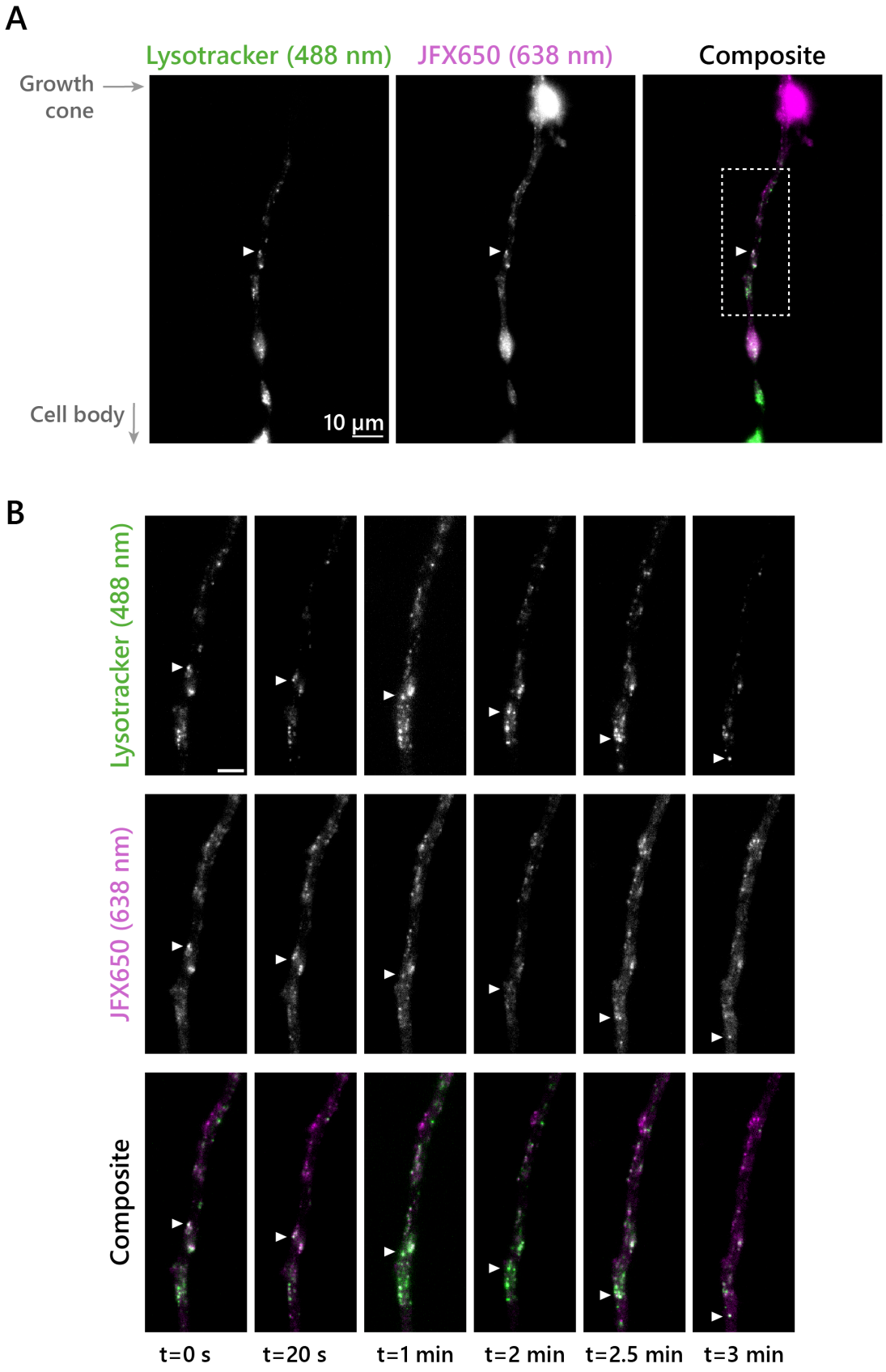
Live cell imaging of differentiated RPL4-Halo PC-12 cells was done on BONSAI using the TIRF 638 setting, 100 ms exposure, and 100x magnification. Images were prepared using ImageJ. Cells were prepared for imaging as outlined in Section 2.5.8, using 50 nM JF549 to locate processes and imaged using JFX650-Halo and the 638 channel. Laser power was 10% for the 638 laser (max power = 98 mW). N = 1. A) The end of a PC-12 process showing the growth cone. B) Timelapse of FRAP, where t=0 is the frame after the photobleach step ended. FRAP was done using the 638 laser at 10% power (max power = 98 mW) and 10 repetitions. Photobleached region is denoted by the white box in the first image. Images are a magnification of the boxed region in 5.15A. Times were rounded to the nearest second. C) Kymograph of particle movement during FRAP assay. The photobleaching event is denoted by "Pb". Possible transport events are denoted by white boxes. Frame rate = 0.11 seconds/frame.

RPL4-Halo live cell imaging was repeated with 50 nM JFX650 ligand to better capture ribosome movement. In addition, the movement of lysosomes/endosomes was also visualised using LysoTracker Green DND-26, a reagent which stains acidic compartments within cells. As the literature surrounding neuronal RNA granule transport points to the possibility of ribosomal hitchhiking on lysosomes (see Chapter 1), this experiment would provide evidence in favour of this transport model, should the transport events co-localise. Differentiated cells were labelled with 50 nM JFX650 for 1 hour (the low density of cells ensures that background staining due to JFX650 is minimal) and ligand was washed out as standard. 0.5 nM LysoTracker was then applied to the cells for 1 minute. The media was replaced and the labelled cells were imaged on BONSAI using the 488 TIRF and 638 TIRF settings.

Individual ribosome-containing particles and LysoTracker-stained particles were identified in PC-12 processes, with the signal for many JFX650 and LysoTracker particles appearing to co-localise (Figure 5.12A). A region of the timelapse is presented in Figure 5.12B, where selected frames demonstrate the movement of a ribosome/lysosome particle (see arrow) down the axon across a 3 minute period. This particle is present within both 488 nm and 638 nm channels and overlap of signal is demonstrated in the composite image. This was just one of many transport events observed, although stationary particles present in only the JFX650 or LysoTracker

channels were also seen. A kymograph for each channel was then created (Figure 5.12C). Both the LysoTracker and JFX650 kymographs are similar, although differences between the two are present. There are many transport events in both kymographs, with the direction of transport appearing to be mostly anterograde. However, a small number of retrograde events were also observed. Interestingly, the average speed of transport for JFX650-labelled ribosomes was 0.4 µm/second (calculated based on Figure 5.12C, based on 6 transport events in one process, across one biological repeat), which is lower than the calculated speed for TMR-labelled ribosomes (1.21 µm/second, Figure 5.10). However, as velocity calculations are from just one biological replicate per experiment, this difference could be attributed to the health of the cells on the day of imaging. For a detailed discussion about this difference in transport velocity, see Section 5.3.

Overall, this data suggests that endosomes/lysosomes may be involved in the transport of ribosomes within PC-12 processes. To demonstrate that the signal overlap was not due to bleed-through of the LysoTracker channel into the JFX650 channel (or vice versa), single channel controls were set up (Figure 5.13). For each channel, either LysoTracker Green or JFX650-Halo was applied to differentiated RPL4-Halo cells as per the standard labelling protocol. Cells were then imaged with both the 488 and 638 channels. For the LysoTracker only control, a small region of very low signal can be seen in the 638 channel. For the JFX650 only control, no signal is seen in the 488 channel. Additionally, the fluorescence spectra of both LysoTracker green and JFX650 ligand is provided in Figure 5.14. This figure illustrates that there is very little spectral overlap between the two ligands, therefore, bleed-through of signal is unlikely. Moreover, each channel was imaged separately (i.e. both lasers were never active at the same time), which also ensures that signal bleed-through is minimal.


A
Lysotracker (488 nm)
JFX650 (638 nm)
Composite
Growth cone
Cell body
10 µm
B
Lysotracker (488 nm)
JFX650 (638 nm)
Composite
t=0 s
t=20 s
t=1 min
t=2 min
t=2.5 min
t=3 min

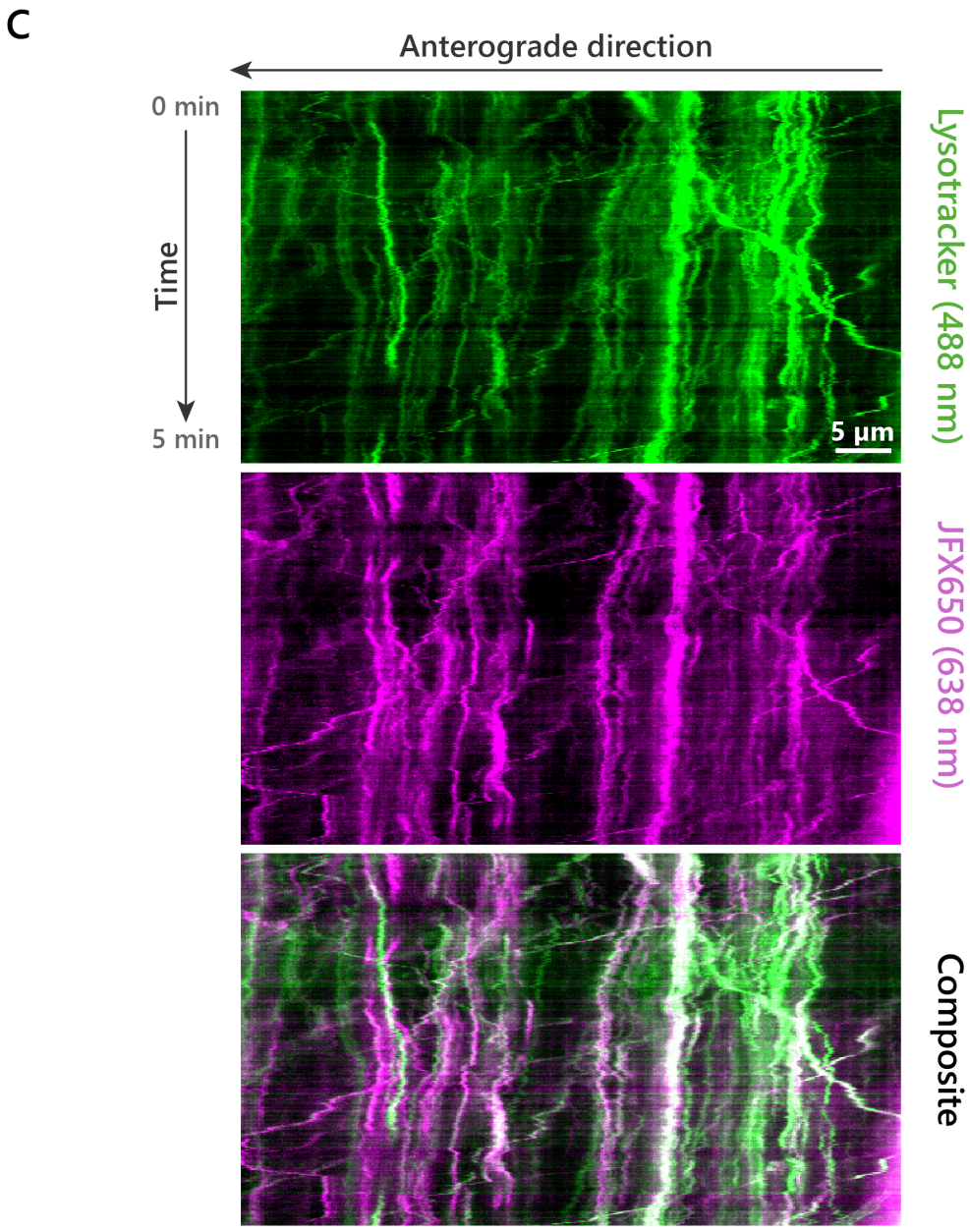


Figure 5.12: Live Cell Imaging of RPL4-Halo PC-12 Line Using JFX650-Halo and LysoTracker Green.

Live cell imaging of differentiated RPL4-Halo PC-12 cells was done on BONSAI by TIRF using the 488 and 638 lasers, at 100 ms exposure, and 100x magnification. Images were prepared from videos using ImageJ. Cells were prepared for imaging as outlined in Section 2.5.8, using 50 nM JFX650-Halo and 0.5 nM LysoTracker Green. Laser power was kept consistent at 7% for 488 and 10% for 638 (max power = 61 mW and 98 mW for 488 and 638 lasers respectively). N = 1. A) The end of a PC-12 process showing the growth cone, where labelled ribosomes and lysosomes/endosomes can be clearly observed. B) Timelapse of particle movement over 3 minutes. The anterograde movement of one particle is denoted by the white arrow. Images are a magnification of the boxed region in 5.15A. Times were rounded to the nearest ten seconds. Scale bar = 10 µM. C) Kymograph of particle movement during timelapse. Data for kymograph velocity calculations is provided in Table 2.22. Frame rate = 1.00 seconds/frame.

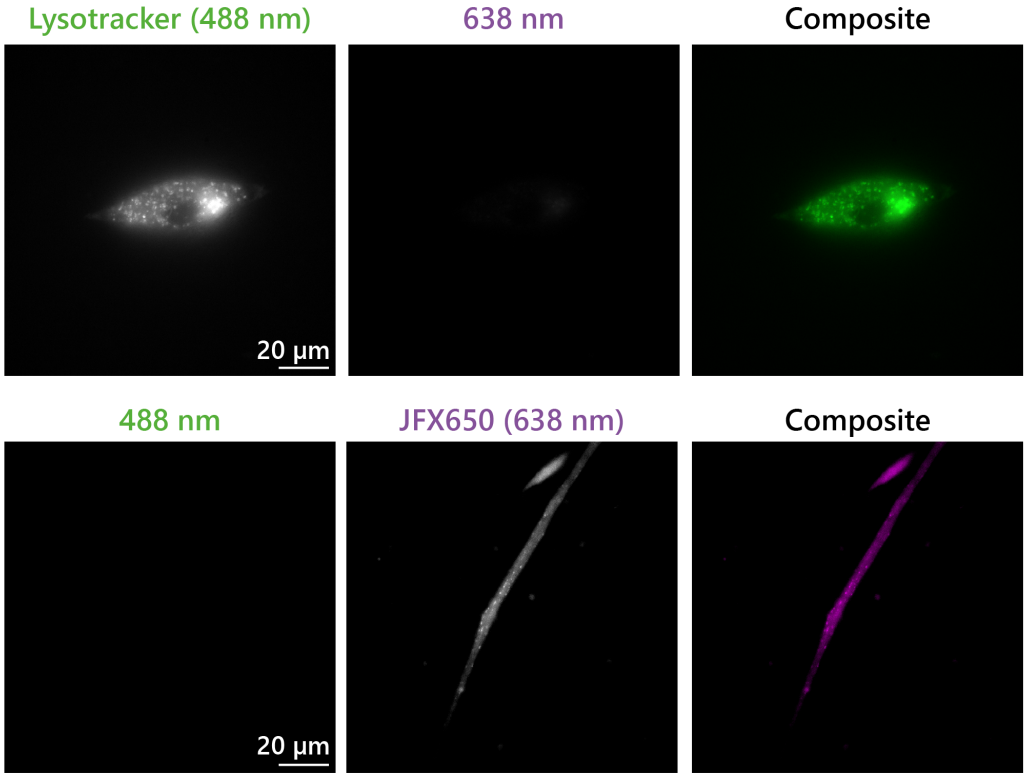


Figure 5.13: Single Channel Controls for L4-Halo PC-12 Imaging Using JFX650-Halo and LysoTracker Green.

L4-Halo PC-12 cells were differentiated as per the standard protocol and labelled with either 50 nM JFX650-Halo or 0.5 nM LysoTracker Green. Controls were then imaged using BONSAI using the 488 (LysoTracker, green) and 638 laser (JFX650, magenta) at 100 ms exposure. Laser power was kept consistent at 7% for 488 and 10% for 638. The same brightness and contrast levels were set for both channels using ImageJ. Minimal amounts of signal bleed-through is observed.

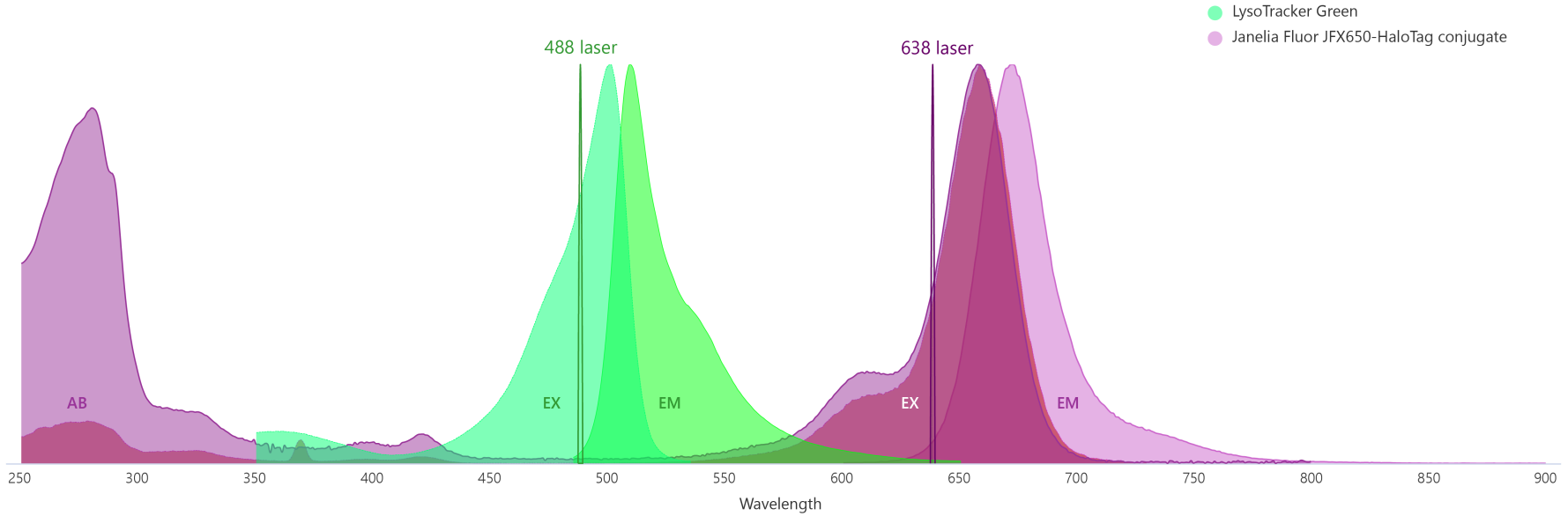


Figure 5.14: Fluorescence Spectra of JFX650 and LysoTracker Green.

This fluorescence spectra was generated using FPbase (Lambert 2019) and modified using Adobe Illustrator. The LysoTracker Green (green) is excited by the 488 laser and shows peak emission at approximately 510 nm. In contrast, the JFX650-Halo ligand (magenta) is excited by the 638 laser and shows peak emission at approximately 670 nm. There is very little overlap between the two fluorescence spectra. Abbreviations: EM - emission, EX - excitation, AB - Absorbance.

Additionally, live cell imaging was also carried out using RPS2-Halo PC-12 cells. In contrast to RPL4-Halo cells, a large amount of Halo-tagged RPS2 protein is present within the free 40S subunit fractions of RPS2-Halo cells. Therefore, imaging of this line would provide a useful comparison between cells where the tagged protein is mostly incorporated within ribosomes (RPL4-Halo), and cells where the tagged protein is present in free subunits as well as ribosomes (RPS2-Halo). PC-12 RPS2-Halo cells were plated, differentiated, and labelled with 50 nM JFX650 as previously described. Labelled cells were imaged on BONSAI using the timelapse mode on the TIRF 638 setting.

RPS2-Halo cells had a very diffuse JFX650 signal compared to RPL4-Halo cells (Figure 5.15A). Although some large, bright stationary particles could be detected in the processes, individual puncta were infrequently observed. To determine if transport events were occurring, the axon was photobleached using the 638 laser at 20% power (Figure 5.15B). Unlike in RPL4-Halo cells, the JFX650 signal returned to the photobleached area rapidly, with no clear transport events being identified. However, when presented as a kymograph, some infrequent transport events after the photobleaching are present (Figure 5.15C). Analysis of this kymograph determined that the particle movement had an average velocity of 0.3 µm/second (calculated based on Figure 5.15C, based on 3 transport events in one process, across one biological repeat), which is similar to the velocity previously observed for JFX650-labelled RPL4-Halo (0.4 µm/second). Due to the diffuse nature of RPS2-Halo, the pORANGE-RPS2-Halo CRISPR construct is more likely suited for other proteomic and biochemical assays (such as ribosome purification), as individual transport events are difficult to capture at a good resolution.

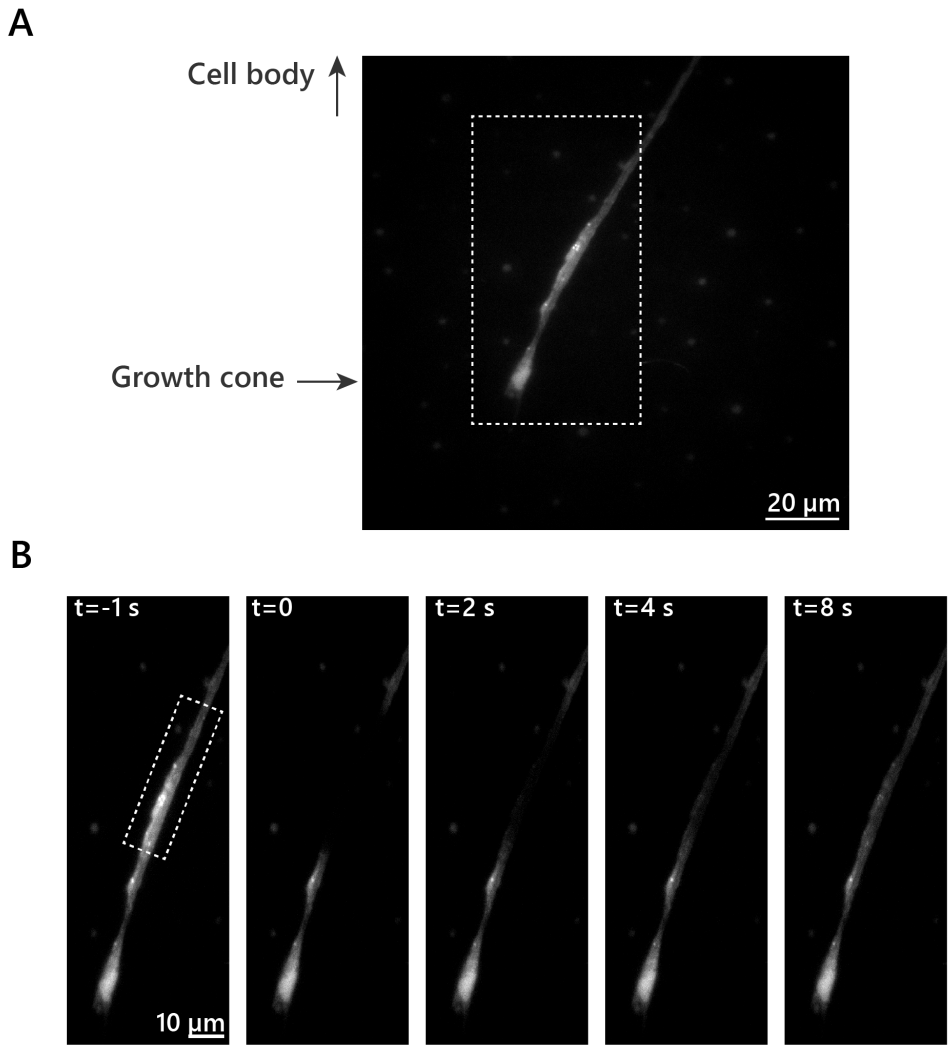

A
Cell body
Growth cone
20 µm
B
t=-1 s
t=0
t=2 s
t=4 s
t=8 s
10 µm
C
Anterograde direction
0 s
Pb
Time
90 s

Figure 5.15: Live Cell Imaging of RPS2-Halo PC-12 Line Using JFX650-Halo.

Live cell imaging of differentiated RPS2-Halo PC-12 cells was done on BONSAI using the TIRF 638 setting, 100 ms exposure, and 100x magnification. Images were prepared from videos using ImageJ. Cells were prepared for imaging as outlined in Section 2.5.8, using 50 nM JFX650-Halo. N = 1. A) The end of a PC-12 process showing the growth cone. Some bright JFX650-labelled particles can be observed in the process. B) Timelapse of FRAP, where t=0 is the frame after the photobleach step ended. FRAP was done using the 638 laser at 20% power (max power = 98 mW) and 10 repetitions. Photobleached region is denoted by the white box in the first image. Images are a magnification of the boxed region in 5.15A. Times were rounded to the nearest second. C) Kymograph of particle movement during FRAP assay. The photobleaching event is denoted by "Pb". Data for kymograph velocity calculations is provided in Table 2.22. Frame rate = 0.27 seconds/frame.

5.3 Discussion

5.3.1 The Knock-In Efficiency of ORANGE Constructs is Extremely Low Using Lipofection

Identifying CRISPR-edited cortical neurons in live cell imaging assays was challenging. The lack of CRISPR-edited neurons is likely caused by the low transfection and CRISPR editing efficiencies. In particular, the low number of pORANGE-GFP-ACTB transfected neurons was surprising. However, upon revisiting the literature detailing the plasmid's creation, it was noticed that two GFP-ACTB constructs had been produced with differing knock-in efficiencies (Willems et al. 2020). The authors noticed that their pORANGE-GFP-ACTB #1 plasmid had a low expression of the tagged protein in transfected cells (0.42% ± 0.09%), in comparison to the other ORANGE CRISPR plasmids designed. Therefore, another GFP-ACTB construct was produced (pORANGE GFP-ACTB KI #2, Addgene Plasmid #139666) using an alternative PAM and guide sequence in the same target exon. Strikingly, this construct was determined to have a knock-in efficiency of 7.4% ± 1.1%. This considerable increase in knock-in efficiency is surprising given that the GFP-ACTB #2 guide sequence selected is in close proximity to GFP-ACTB #1 guide sequence (5 bp between the guide integration sites).

As this project begun in 2019, using data and constructs provided in the ORANGE preprint manuscript (Willems et al. 2019), we were unaware of the extremely low knock-in efficiency of the pORANGE-GFP-ACTB #1 construct. Taking the transfection efficiency calculated using Lipofectamine 2000 transfection in DIV 4 neuronal cultures (mean=1.81%, Figure 5.2), approximately 3,258 neurons are expected to be transfected out of a population of $0.18x10^6$. This number is an estimate, as transfection efficiency is affected by several experimental variables (including DNA plasmid used, health of the culture, etc). Given the knock-in efficiency for the pORANGE-GFP-ACTB plasmid in transfected cells is 0.42% on average, we would therefore expect approximately 14 GFP-tagged β-actin positive cells per dish. However, this number is greater than the number of GFP-tagged β-actin neurons observed (≤1 per dish). This indicates that the transfection efficiency of the live cell imaging assays with Lipofectamine 2000 is likely to be less than 1.81%, or that more CRISPR-edited neurons were present that went undetected.

A straightforward way to increase the number of CRISPR-edited cells is to increase the transfection efficiency of the ORANGE constructs in primary neuronal cultures. The primary method of transfection (Lipofectamine 2000) was selected as it is a commonly used method for neuronal transfection which has been successfully used within the Twelvetrees lab using non-CRISPR constructs. In addition to lipofection, *Willems et al.* have demonstrated that the ORANGE system is also compatible with electroporation and viral delivery methods (Willems et al. 2020). Both the aforementioned DNA delivery methods are generally considered to have higher average transfection efficiencies than Lipofectamine 2000 (Alabdullah et al. 2019), although no data is provided on the transformation efficiencies of the DNA delivery protocols used in the ORANGE paper (Willems et al. 2020). To this end, a thorough investigation of a range of electroporation and viral methods would be beneficial when optimising ORANGE plasmid transfection, which in turn would increase the number of CRISPR-edited neurons for imaging assays.

5.3.2 Ribosomes Undergo Fast Transport in PC-12 Processes

The chapter presents the first microscopic data of endogenous ribosome transport in PC-12 processes. Ribosome transport events were bi-directional in nature, indicating the involvement of kinesin and dynein motors. This is supported by the literature, where evidence of neuronal RNA granule transport mediated by motor proteins is growing (reviewed in Abouward and Schiavo 2021). For example, a recent study in DRG axons showed that RNA granules containing the RNA binding protein, SFPQ, are actively transported by Kinesin-1 KIF5A motors in association with KCL1 (Fukuda et al. 2021). In addition, Kinesin-2 motors have also been implicated in the transport of β-actin and β2B-tubulin mRNA-containing RNA granules (Baumann, Komissarov, et al. 2020).

In this work, ribosomes were transported with similar velocities to those documented for fast axonal transport - approximately 0.5 – 1 µm/second (Maday et al. 2014). Specifically, quantification of kymograph transport events determined that ribosome velocity ranged from of 0.3 µm/second (RPS2-Halo, Figure 5.15) to 1.2 µm/second (RPL4-Halo, TMR-Halo labelling, Figure 5.10). Interestingly, for RPL4-Halo processes labelled with JFX650 (Figure 5.12), the kymograph calculations (see Table 2.22) showed that smaller particles had a faster velocity (0.58 µm/second) than larger particles (0.3 µm/second). This difference may reflect the transport of monosomes and polysomes respectively. Similar speeds have been documented for the movement of RNA granules in neurons, ranging from 0.1 µm/second in primary cortical neurons (Knowles et al. 1996) up to 3 µm/second in DRG neurons (Fukuda et al. 2021).

The difference in average velocity between Halo-TMR and JFX650 labelled ribosomes (TMR = 1.2 µm/second, JFX650 = 0.4 µm/second) in RPL4-Halo processes is also intriguing. However, there are several possible explanations. Firstly, as these experiments have an N of 1, this variation can be explained by differences in neuronal culture health on the day of imaging. This highlights a limitation of this work

- more biological repeats are required before the velocity of ribosome transport in this system can be confidently calculated. Additionally, it is likely that the average transport velocity varies from cell-to-cell, dependent on the cell's translational requirements. For example, oxidative stress caused by the accumulation of reactive oxygen species has been shown to inhibit axonal transport in rat hippocampal neurons (Fang et al. 2012). This also highlights that longer imaging times (where fluorophore photobleaching releases reactive oxygen species) may negatively impact the rates and velocity of ribosome transport in processes. Finally, the JFX650-labelled imaging was done in close proximity to the growth cone (Figure 5.12), in comparison to the TMR-labelled imaging which captured a region in the middle of the process (Figure 5.10). Therefore, it is possible that ribosome transport slows down when approaching process terminals.

There are several additional PC-12 imaging experiments which would provide further insights into the mechanisms of ribosome transport. Firstly, there are a variety of live cell stains which could be used to provide further details about how ribosomes are localised to the processes. By treating cells with a tubulin probe in conjunction with a Halo ligand, it could be determined if ribosome transport occurs along microtubules. This would provide evidence that ribosome translocation is motor protein dependent, as kinesin and dynein require microtubules for their processive movement (see Chapter 1). Secondly, treatment of PC-12 Halo cells with drugs that inhibit ribosome biogenesis (such as Leptomycin B) could be used to determine the effect on Halo-tagged ribosome protein localisation. Leptomycin B prevents ribosome biogenesis by blocking the nuclear export of ribosomal subunits into the cytoplasm (Thomas and Kutay 2003), therefore, we would expect to see a reduction in HaloTag signal after treatment. This would further demonstrate that these live cell imaging experiments are detecting the movement of ribosomes.

5.3.3 The PC-12 Line as a Neuronal Model System

The PC-12 line is a helpful model system which has provided convincing data in favour of the RNA granule model of ribosomal transport in neurons. As previously detailed in Chapter 3, the PC-12 line can be differentiated into non-dividing cells resembling sympathetic ganglion neurons which grow processes. These processes share some similarities to neuronal axons (reviewed in Wiatrak et al. 2020), including neurotransmitter release, expression of axonal markers (such as Tau and GAP-43) and the formation of synapse-like structures (Jeon et al. 2010). These qualities make it likely that ribosome transport in axons is similar to the directed transport observed in PC-12 processes. However, it must be acknowledged that neuronal model systems are not a perfect substitute for primary neurons. Therefore, the Halo PC-12 line imaging experiments must be replicated in primary neurons before any firm conclusions about axonal ribosome transport can be drawn.

Chapter 6

Expression of Ribosome-Associated Genes in Brain Regions

6.1 Introduction

The heterogeneous nature of ribosomes is now a widely accepted concept in the RNA biology field, however, recognition of the variation in ribosome structure took the scientific community several decades (Genuth and Barna 2018; Chaillou 2019). Since their discovery in the 1950's (Palade 1955), ribosomes have largely been considered to be uniform structures with little variation in composition. One large influence in this was the discovery of mRNA and its role in protein translation, with early viral translation experiments concluding that ribosomes are "non-specialised structures" (Brenner et al. 1961) which do not appear to exert any regulatory control on protein synthesis. Following this, evidence of ribosomes with different ribosomal protein compositions was discovered in several model organisms - but this data often proved controversial and difficult to replicate (Genuth and Barna 2018).

Concrete evidence of ribosomes with distinct compositions has been increasing rapidly over recent years, owing to developments in proteomic techniques, ribosome

profiling and sequencing technologies (Gay et al. 2022). These advancements have allowed us to investigate the composition of the ribosome on the molecular level, identifying differences in ribosome protein composition and rRNA sequence. To this end, ribosome heterogeneity can largely be separated into three categories; ribosomal protein composition (including their stoichiometry, post-translational modifications, paralogs and isoforms), the transcription and modifications of rRNA variants, and interaction with ribosome-associated factors (Joo et al. 2022; Guo 2018). Together, these compositional changes can create highly heterogeneous ribosomes - otherwise known as "specialised ribosomes" - which fulfil a distinct translational requirement within the cell (reviewed in Gay et al. 2022, see Chapter 1).

The cellular role of specialised ribosomes is varied and still not fully understood (Joo et al. 2022). A specialised ribosome is defined as a ribosome that is functionally distinct from those in other tissues or cells, therefore, playing a "specialised" translational role within that location. At a minimum, all ribosomes must be capable of synthesising protein, yet the catalytic abilities of ribosomes can differ and greatly impact their translational outcomes (Ferretti and Karbstein 2019). There are several mechanisms ribosomes employ to achieve this, such as preferential translation of specific mRNA populations (Z. Shi et al. 2017), control of translation initiation (Landry et al. 2009), and regulation of translation fidelity (McMahon et al. 2019).

Despite the significant advancements in the ribosome specialisation field, robust evidence of neuronal ribosome heterogeneity and specialisation is still lacking. However, developments in our understanding of neuronal local translation in neurites has provided plausibility to the theory that axons contain specialised ribosomes (reviewed in Norris et al. 2021). Although purification and analysis of axonal ribosomes is experimentally difficult, bioinformatic analysis of existing RNA-Seq databases could provide insight into ribosome heterogeneity. In particular, RNA expression analysis may identify gene candidates with notable brain expression, providing a starting point for future proteomic and sequencing work of axonal ribosomes.

The aim of this chapter is to investigate the gene expression levels of both ribo-

some proteins and snoRNA genes to determine if there are brain-specific ribosome protein compositions or rRNA modifications. If present, these changes may provide evidence for structural changes or modifications to neuronal ribosomes which may be indicative of ribosome specialisation in brain tissue. Furthermore, investigation of changes in ribosome-associated genes may provide insight into the mechanisms behind axonal ribosome translocation in neurons.

6.1.1 The GTEx Database

The expression data used in this analysis was provided by the GTEx database V8.0 (GTEx Consortium 2020). This data was gathered as part of the GTEx project, an effort to create a dataset of tissue-specific mRNA expression levels for investigating the link between human genetic variation and gene expression and regulation. Notably, the total RNA extracted from tissue samples undergoes poly(A) selection before sequencing, a common RNA-Seq methodology which enriches for RNA species containing poly(A) tails - such as mRNA and some small RNAs (including snoRNAs). The GTEx database is composed of data from 948 male and female donors of varying ages, providing gene expression information from 54 non-diseased tissues including 13 distinct brain regions.

The GTEx dataset was selected for this gene expression analysis as it fulfilled several important criteria. GTEx is one of the largest public RNA-seq datasets (Suntsova et al. 2019), contributing data for many published research papers since its conception in 2015. Importantly, GTEx contains data from a large number of individuals where the cause of death was predominantly either traumatic injury or heart disease. Donors that have died from neurological disorders may show gene expression changes which could impact the results of this analysis. Although GTEx includes samples which originate from donors who died from neurological causes, this group makes up less than 3% of the total donor population. Furthermore, the GTEx database contains data from a large number of tissue sites and transformed cells, allowing for comparison between many brain and non-brain tissues.

6.2 Results

6.2.1 Most Ribosome Protein Genes Are Downregulated In Brain Regions

The GTEx database contains RNA read count data from tissue samples taken from donors after death. The gene expression data was generated via RNA-seq utilising poly-A+ selection and Illumina sequencing technology (Illumina, Inc 2014). The sequencing reads were then aligned to the human GRCh38/hg38 reference genome and gene annotation was done using GENCODE v26. Expression levels of genes were normalised to give the transcript count per million transcripts (Transcripts per Million (TPM)) to ensure that comparison between tissues is possible. Further technical details for RNA library generation and sequencing are published online (GTEx Consortium 2021).

Some of the 54 tissue sites present in the GTEx database were excluded from this analysis. All transformed cell lines (fibroblasts and EBV-transformed lymphocytes) were excluded, as experimental immortalised lines are not naturally occurring within the healthy human body and therefore outside the scope of this analysis. In addition, data from the testis and ovary were also excluded. These tissues are sites of extremely high cellular division resulting in increased expression of most ribosome-associated genes. Therefore, inclusion of these tissues prevented expression patterns being detected in all other tissue sites. Conversely, several tissue sites were included despite low numbers of samples (i.e. bladder, cervix, Fallopian tube, and kidney medulla). Low sample number may cause outliers to have a larger affect on averaging, artificially increasing the average of the dataset. However, as the median value is taken for each gene in each tissue site, the effect of outliers is minimal.

The gene expression of 89 ribosome subunit protein genes was analysed across 54 tissues to determine whether ribosome proteins had differential expression in brain regions. A list comprising of 79 canonical ribosome protein genes of the small (40S) and large (60S) ribosome subunits alongside all known ribosome protein paralogs

was generated (see Appendix 1, Table 8.1) for analysis. All mitochondrial ribosome protein genes and ribosome gene pseudogenes were excluded from this analysis as they are unlikely to significantly contribute to ribosome heterogeneity in neurons. Ribosome protein gene symbols and chromosome locations were provided by the HUGO Gene Nomenclature Committee (Ribosome Protein gene group, group number 1054). Additional gene details such as Genecode ID and Entrez Gene were provided by GTEx Portal (V8). This list was used in conjunction with the GTEx API (Swagger) to pull relevant gene expression data for analysis.

All data was normalised to show the expression level as a proportion of the maximum gene expression in all tissue sites. The commonly used min-max normalisation method $((x - x_{(min)})/ (x_{(max)} - x_{(min)}))$ was selected for this analysis as it allows for easier visualisation whilst preserving the relationships between the data. After normalisation, data representing the lowest gene expression in a gene is transformed to zero (dark blue). In contrast, data representing the highest gene expression for a gene adopts the value of 1 (white). This results in a data range of 0-1, allowing each data point to be easily plotted regardless of variation within the gene expression values for each gene analysed.

The median TPM for 89 ribosome-associated genes in 50 human tissue sites was plotted as a heat map (Figure 6.1). Overall, the brain regions showed lower expression of most ribosome protein genes compared to most other tissue sites in the body - illustrated by the dark blue regions on the heat map. This result was expected as brain regions are sites of decreased cell division. Similarly, the heart, kidney and liver tissues also had lower expression of ribosome protein genes. The highest expression levels were seen in the cervical tissues, skin tissues, the uterus and vagina which are all tissues with high cellular turnover. This initial analysis acted as proof of concept that accurate visualisation of ribosome genes across human tissues was possible using this methodology.

Interestingly, three ribosome protein paralog genes showed tissue-specific expression (where the majority of the gene expression is observed in a single tissue). One

example is RPL3L which shows a predominantly muscle-specific pattern of expression - although low levels of expression are also observed in heart tissues. The ribosome protein paralog RPL7L1 had the highest gene expression levels in the cerebellar hemisphere (a median expression of 50.21 TPM). This is a surprising result given the low expression of most other ribosome protein genes, which may point to this paralog being an important component of neuronal ribosomes. Similarly, RPL26L1 also showed brain-elevated expression (median of 19.9 TPM in BA9), however, the highest gene expression for this paralog was identified in the adrenal gland (median of 28.49 TPM). Overall, this data suggests that paralogs may contribute to ribosome heterogeneity, with RPL7L1 and RPL26L1 in particular playing a possible role in neuronal ribosome specialisation.

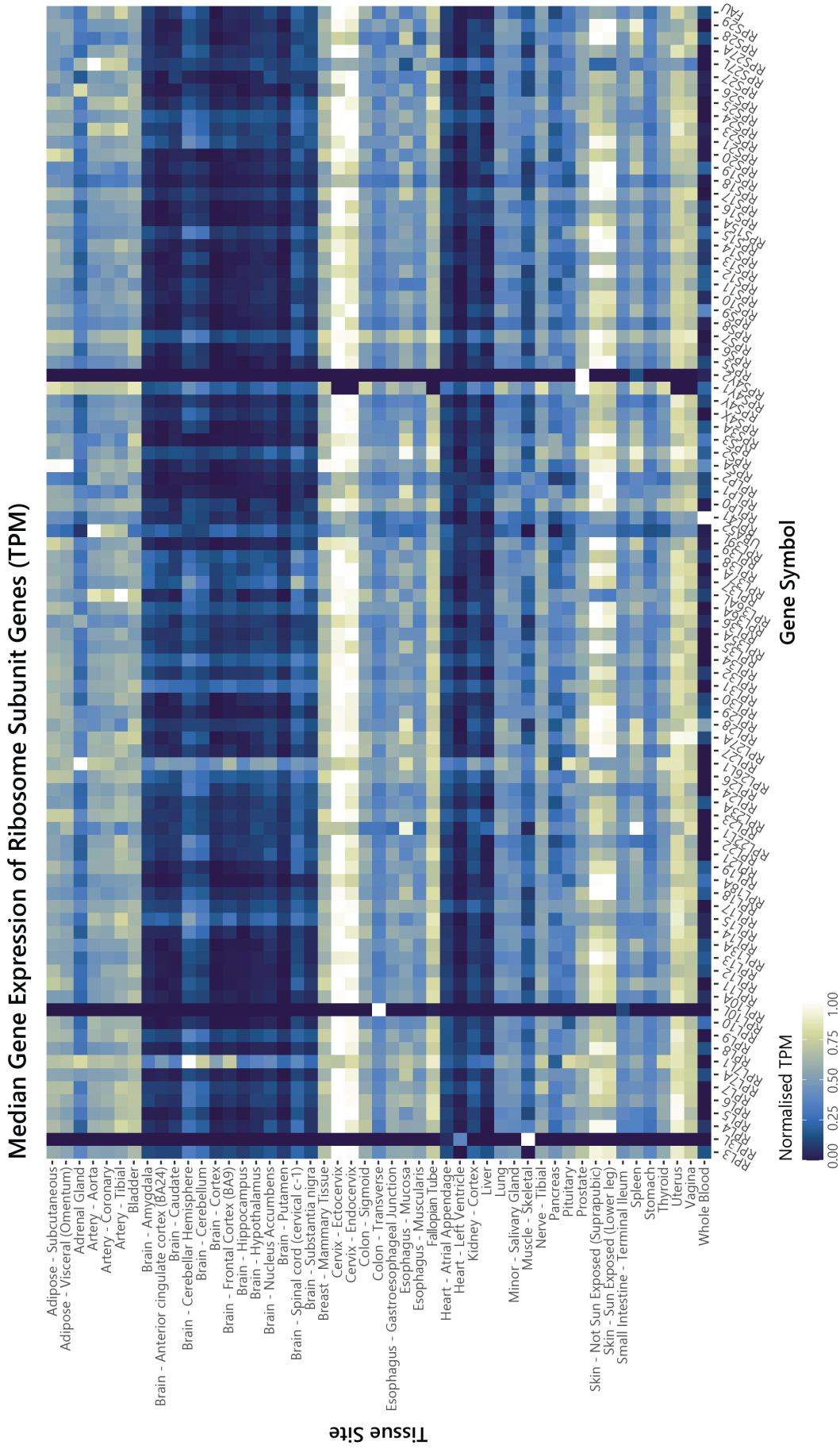

Median Gene Expression of Ribosome Subunit Genes (TPM)
Tissue Site
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cerebellum
Brain - Cortex
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood
Gene Symbol
Normalised TPM
0.00
0.25
0.50
0.75
1.00

Figure 6.1: Gene Expression of Ribosome Protein Genes in Human Tissue.

Heatmap visualising the gene expression of 89 ribosome protein genes across 50 tissue sites. Gene expression data was provided by the GTEx Portal V8. Raw data was analysed and plotted using R (version 2.4.0). Gene expression levels are given in transcripts per million (TPM), with the median TPM plotted for every ribosome protein gene for each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from 0 to 1. The code used to generate this heatmap is provided in the Appendix, Figure 8.5.

6.2.2 Ribosome Protein Genes Show Differential Exon Expression in Brain Regions

Subsequently, the exon expression of the ribosome protein genes were analysed. Exon expression analysis allows identification of upregulated, downregulated or missing exons, which can give an indication as to whether specific ribosome protein isoforms are present within brain regions. For each of the 89 ribosome protein genes previously analysed, the median expression (expressed in TPM) for each exon present was determined for each tissue site. These values were then normalised using min-max normalisation to transform each data point as a proportion of the total expression for that exon - ranging between 0 and 1. In the instance that an exon has a median TPM value of 0 for a tissue site, it was assumed that the exon is unlikely to be expressed at significant levels within that tissue. All ribosome protein exon expression heat maps are provided in the Appendix, Figure 8.6.

Interestingly, the exon expression of ribosome protein genes could be classified as either consistent or differential. Heat maps showing consistent exon expression (Figure 6.2) had little variation in expression levels across all exons present, often with lower expression levels seen in all brain regions. From the 89 genes analysed, 42 were identified to have consistent expression of all exons in brain regions - 47% of the total ribosome protein gene group. Additionally, consistent expression was seen with both 40S and 60S proteins and did not appear to be more prevalent in either subunit.

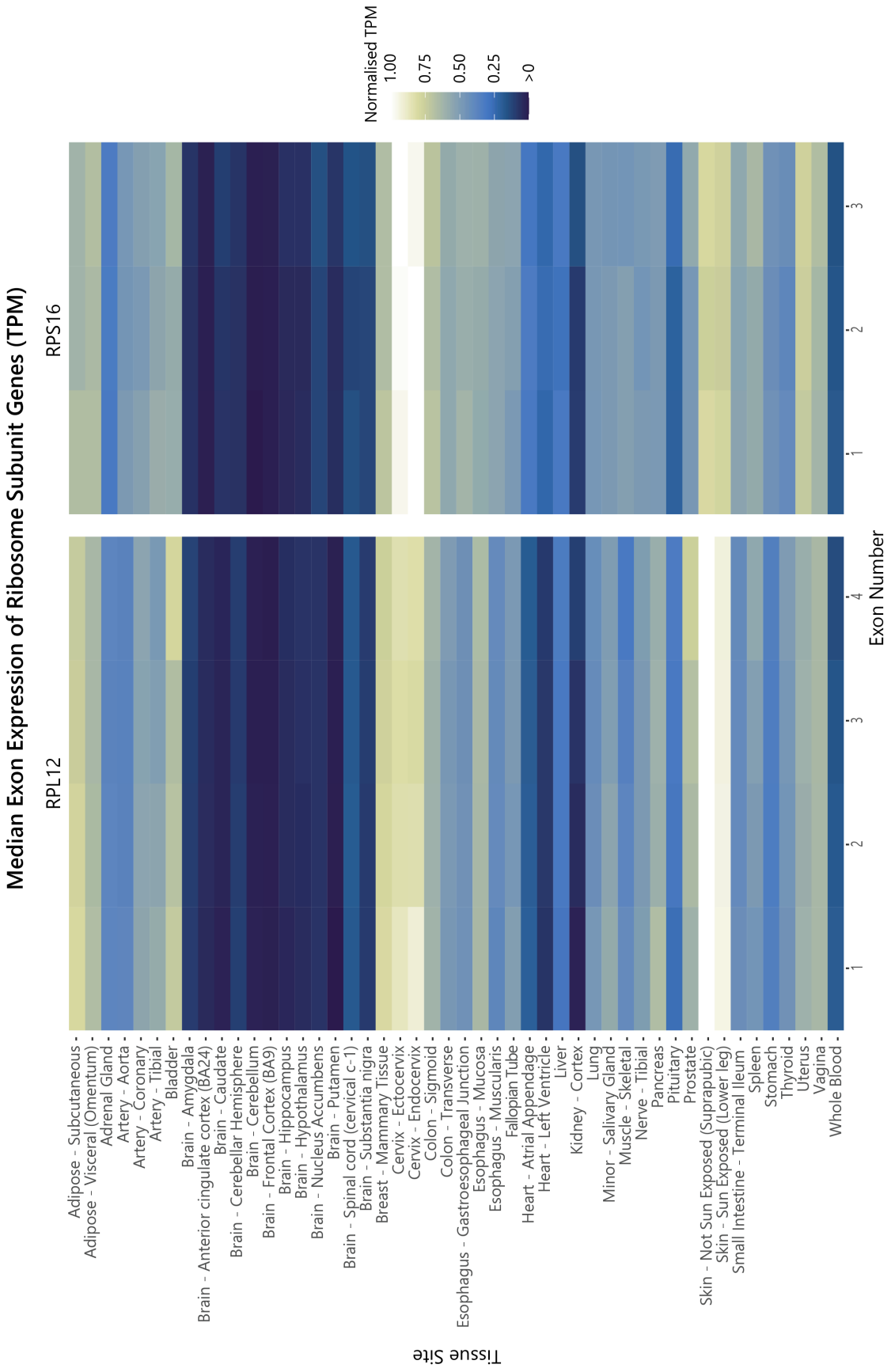

Median Exon Expression of Ribosome Subunit Genes (TPM)
RPL12
RPS16
Normalised TPM
1.00
0.75
0.50
0.25
>0
Tissue Site
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cerebellum
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Liver
Kidney - Cortex
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood
1
2
3
4
Exon Number

Figure 6.2: Gene Exon Expression Analysis - Consistent Expression Examples.

Heatmaps visualising consistent exon expression of selected ribosome protein genes. Gene exon expression data was provided by the GTEx Portal V8. Raw data was analysed and plotted using R (version 2.4.0). Expression levels for each exon are given in transcripts per million (TPM) with the median TPM calculated for every gene in each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from 0 to 1. The data was filtered so that median expression levels of zero were excluded from this analysis. Tissues with a median TPM of 0 for every exon were not plotted (i.e. Brain - Cortex). The analysis was carried out to produce a heatmap for each of the 89 ribosome protein genes. One example of consistent exon expression is provided for both the 40S and 60S subunits (RPS16 and RPL12 respectively). See the Appendix for all exon expression heatmaps (Figure 8.6) and the code used to generate them (Figure 8.7).

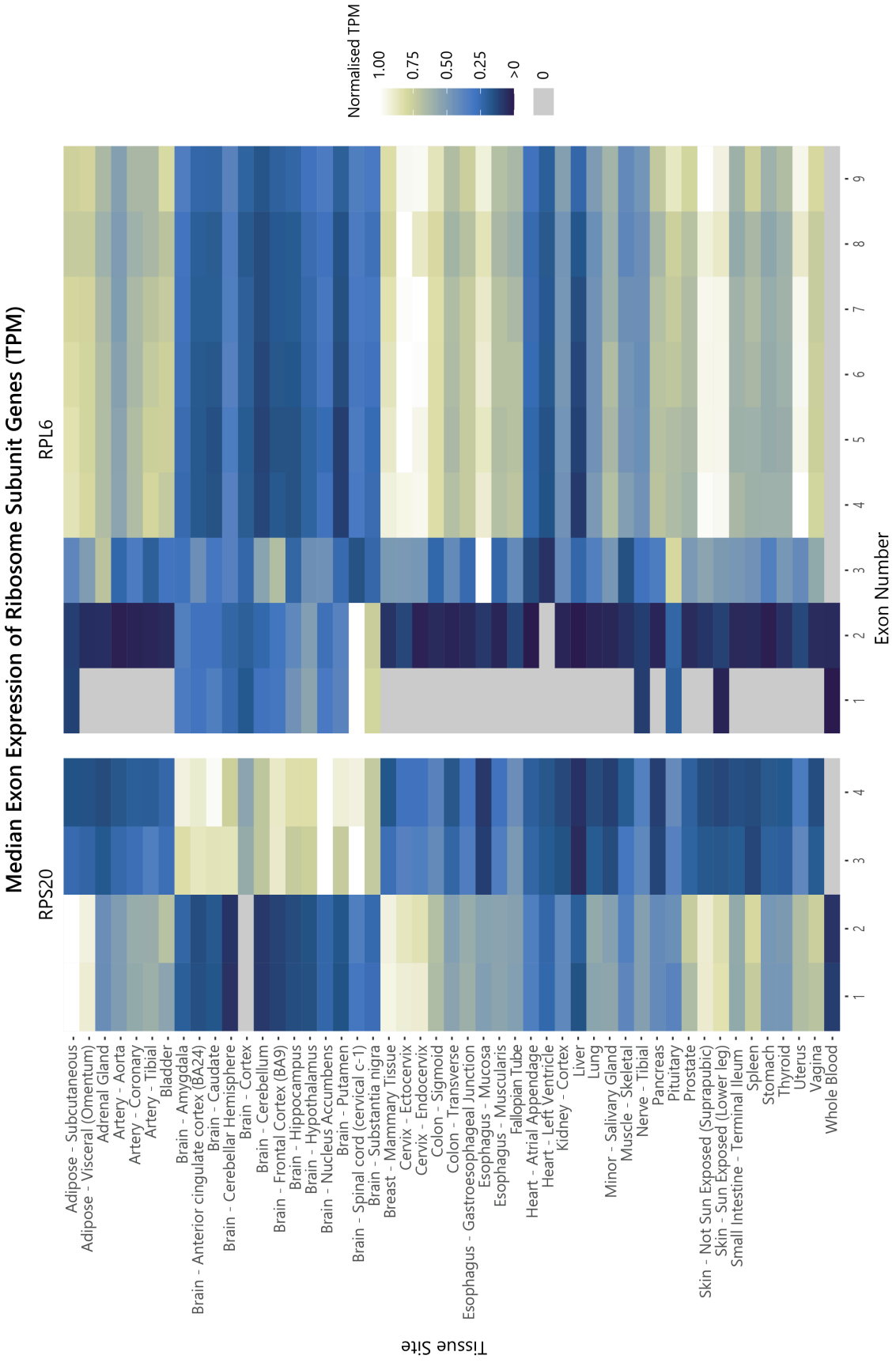

Median Exon Expression of Ribosome Subunit Genes (TPM)
RPS20
RPL6
Normalised TPM
1.00
0.75
0.50
0.25
>0
0
Exon Number
Tissue Site
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cortex
Brain - Cerebellum
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood

Figure 6.3: Gene Exon Expression Analysis - Differential Expression Examples.

Heatmaps visualising differential exon expression of selected ribosome protein genes. Gene exon expression data was provided by the GTEx Portal V8. Expression levels for each exon are given in transcripts per million (TPM) with the median TPM calculated for every gene in each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from 0 to 1. The data was filtered so that median expression levels of zero were excluded from this analysis. Tissues with a median TPM of 0 for every exon were not plotted. One example of differential exon expression is provided for both the 40S and 60S subunits (RPS20 and RPL6 respectively). See the Appendix for all exon expression heatmaps (Figure 8.6) and the code used to generate them (Figure 8.7).

In contrast, several ribosome protein genes showed differential exon expression within brain regions (Figure 6.3), whereby exons were upregulated, downregulated or missing entirely within brain regions. In the examples provided, RPS20 shows upregulation of exons 3 and 4, with the white and yellow colours indicating that the highest expression levels for these exons are present only within brain regions. Similarly, RPL6 exon 1 and 2 appear to be mostly expressed within brain regions and a low number of non-brain tissues, with highest expression in the spinal cord region. Overall, over half (53%) of the differential exon expression heat maps showed possible differential expression of ribosome protein genes in brain regions. A shortlist of possible hits was created which were then analysed using GTEx's Isoform Expression tool to identify isoforms which show brain tissue upregulation (Table 6.1).

Of the 47 genes that showed possible differential exon expression, 22 genes were found to have the highest expression of ribosome protein isoforms in brain regions. The majority of the isoforms identified were determined to be protein-coding by Ensembl (Cunningham et al. 2022), with 4 isoforms being non-protein coding or subject to cellular decay. Moreover, the isoforms identified were predominantly 60S ribosome protein genes - only 5 of the 22 isoforms identified were 40S ribosome proteins. This small number may be explained by the smaller number of protein components of the 40S subunit. Despite the large number of differentially expressed isoforms identified, no isoforms showed brain-specific expression. Interestingly, like the previously identified brain-upregulated paralog RPL7L1, the highest isoform

expression levels were commonly seen in the cerebellar hemisphere region.

In summary, 22 ribosome protein genes have isoforms which show upregulated expression within brain regions. The expression of brain-elevated ribosome protein isoforms is generally low (<10 TPM) but there are several exceptions to this pattern. Two notable examples are isoforms for RPL13 and RPL28, which have higher median TPM values of 109 and 70.1 respectively. Therefore, it is possible that some isoforms identified are expressed highly enough to be incorporated into ribosomes. As ribosome protein structure is a major factor of ribosome specialisation, the isoforms and paralogs identified may contribute to neuronal ribosome heterogeneity and possibly neuronal ribosome axonal transport. However, it is still unclear as to how significant this contribution may be or what functionality their incorporation could provide.

6.2.3 snoRNA Analysis Provides Evidence for Brain-specific rRNA Modification

The final GTEx gene expression analysis investigated snoRNA expression in brain regions as an indicator for brain-specific RNA modifications. As previously described in Chapter 1, snoRNAs can be divided into three main classes; H/ACA Box (encoded by SNORA genes), C/D Box (encoded by SNORD genes) and small Cajal body-specific snoRNAs (commonly referred to as scaRNAs, encoded by SCARNA genes). Some scaRNAs are hybrid snoRNAs formed of both a H/ACA and a C/D box domain, thus capable of interacting with enzymes to create both pseydouridylation and methylation of RNA targets. Each of the three classes were analysed separately, including snoRNA genes present in the GTEx database and excluding any genes which did not have an associated Entrez Gene ID (see the Appendix, Table 8.2 for a full list of snoRNA genes analysed). Analysis was carried out in the same manner as before, with the median TPM for each tissue site calculated and then normalised using min-max normalisation. In total, expression levels of 425 snoRNA genes were analysed including 137 SNORA genes (Figure 6.4), 266 SNORD genes (Figure 6.5),

and 22 SCARNA genes (Figure 6.6).

The snoRNA analysis identified 13 snoRNA genes which displayed the highest median TPM within brain regions (Table 6.2). Most snoRNA genes identified are expressed at low levels (<5 TPM) compared to the canonical ribosome protein genes. However, low expression (<1 TPM) of tissue-enriched snoRNAs in brain tissue has been previously documented (Fafard-Couture et al. 2021). As snoRNAs are thought to play more of a regulatory function within the cell, the low TPM value for many snoRNAs does not discount them from playing an important role within translational control. Several snoRNAs had a median TPM of 0 in all non-brain tissues and were therefore assumed to not be expressed.

Surprisingly, snoRNA upregulation in brain tissues was not found to perfectly correlate to host gene upregulation. snoRNAs genes are commonly present in the intronic regions of other host genes (Dieci et al. 2009), with the expression of snoRNAs often thought to be coupled to the transcription of their host. Therefore, it was expected that the host genes of the brain-upregulated snoRNAs would also show elevated expression in brain tissues. Although many snoRNAs identified follow this trend, 38% snoRNAs had a host gene that had lower expression in brain regions compared to other tissues. For example, SNORA8's host gene, TAF1D, had the highest expression in the Ectocervix (55.11 TPM) despite SNORA8 having the highest expression in the BA9 brain region (TAF1D expression of 21.49 TPM). This finding is corroborated by recent transcriptome studies that have also observed an "un-coupling" of host RNA transcription and snoRNA abundance (Boivin et al. 2018).

Table 6.1: Differentially Expressed Ribosome Protein Isoforms in Brain Regions.

Genes with heatmaps that showed differential exon expression were analysed further to detect genes with notable expression of ribosome protein isoforms in brain regions. The identification of isoforms was done using GTEx's isoform tool. Isoforms were then determined to be protein coding using Ensembl(Cunningham et al. 2022). Abbreviations: Adrenal - Adrenal Gland, Art. Aor. - Artery (Aorta), Art. Tib. - Artery (Tibial), Cere. - Cerebellum, Ecto. - Ectocervix, Endo. - Endocervix, Esoph. - Esophagus, Fallopian - Fallopian Tube, Hemi. - Cerebellar Hemisphere, Hippo. - Hippocampus, Muscle - Skeletal Muscle, Nuc. - Nucleus accumbens, Skin Exp. - Skin Sun Exposed, Skin No. - Skin Not Sun Exposed, SpCo. - Spinal Cord, Subcut. - Adipose (Subcutaneous).

Gene	Brain-upregulated Isoforms	Highest Brain Expression (TPM)	Highest Non-Brain Expression (TPM)	Protein Coding?
RPL3	ENST00000465618.5	1.41 (BA9)	1.31 (Adrenal)	N
	ENST00000402527.5	3.61 (Hemi.)	1.60 (Endo.)	Y
RPL6	ENST00000551291.6	1.28 (SpCo.)	0.61 (Pituitary)	Y
	ENST00000549847.5	7.73 (BA9)	2.03 (Subcut.)	Y
RPL7L1	ENST00000493763.5	25.30 (Hemi.)	19.00 (Uterus)	Y
RPL11	ENST00000467075.1	1.52 (Hemi.)	1.15 (Fallopian)	N
RPL13	ENST00000472354.1	109.00 (Hemi.)	78.00 (Pituitary)	Y
	ENST00000311528.9	3.24 (Hemi.)	2.52 (Muscle)	Y
RPL14	ENST00000338970.10	21.00 (Hemi.)	13.10 (Muscle)	Y
RPL15	ENST00000456530.6	5.80 (Hemi.)	1.89 (Uterus)	Y
RPL18A	ENST00000597648.5	1.53 (BA24)	0.78 (Ecto.)	Y
RPL22	ENST00000465335.1	2.45 (Hemi.)	0.60 (Adrenal)	Y
RPL23	ENST00000470646.1	18.90 (Hemi.)	11.30 (Medulla)	Y
RPL26	ENST00000584164.5	35.40 (BA9)	7.13 (Pituitary)	Y
	ENST00000582556.5	9.89 (BA9)	2.98 (Esoph.)	Y
RPL27	ENST00000589037.5	4.59 (Hemi.)	2.39 (Spleen)	Y
RPL28	ENST00000560055.5	70.10 (Nuc.)	32.20 (Skin Exp.)	Y
RPL29	ENST00000492277.5	41.00 (Cere.)	13.40 (Pituitary)	Y

Gene	Brain-upregulated Isoforms	Highest Brain Expression (TPM)	Highest Non-Brain Expression (TPM)	Protein Coding?
RPL30	ENST00000517805.1	5.04 (Hemi.)	3.32 (Skin No.)	N
RPL31	ENST00000409038.5	3.86 (Cere.)	2.09 (Ecto.)	Y
RPL37	ENST00000511787.1	3.98 (Hemi.)	1.60 (Endo.)	N
	ENST00000274242.9	43.4 (SpCo.)	20.40 (Endo.)	Y
RPS7	ENST00000462576.5	10.50 (Hemi.)	8.79 (Fallopian)	Y
RPS10	ENST00000621356.2	18.60 (Cortex)	11.60 (Pituitary)	Y
RPS20	ENST00000519807.5	12.40 (SpCo.)	6.52 (Pituitary)	Y
RPS23	ENST00000296674.12	44.00 (Hemi.)	35.3 (Art. Tib.)	Y
RPS29	ENST00000396020.7	3.70 (Hemi.)	2.26 (Art. Aor.)	Y

Notably, over two thirds of the snoRNA genes identified were of the H/ACA class. The predicted and known modification sites of the snoRNA genes were identified using the Human snoRNA Altas (Jorjani et al. 2016). 77% of the identified snoRNA genes modify the RNA by psuedouridylation, with the majority of these snoRNAs predicted to target the 18S rRNA. This result was unexpected given that the 28S rRNA is a common target of snoRNAs pseudouridylation (Natchiar et al. 2017), yet only one of the snoRNAs identified predicted a 28S pseudouridylation. To this end, it is possible that the identified 18S modifications play a role in neuronal heterogeneity, especially considering several of the identified snoRNAs (including SNORA70J, SNORA14A, SNORD114-20, and SNORD115-45) appear to be brain-specific.

The identified 18S pseudouridylation sites are modelled on the human 80S ribosome (Protein Data Bank ID: 4UG0) to determine whether the sites are in close proximity to one another (Figure 6.7). Modelling was done using UCSF ChimeraX, with the modification sites highlighted in magenta. Most of the target sites are situated near the central region of the 18S rRNA with only a few target sites towards the exterior regions of the rRNA. Surface modelling of the 18S indicates that some target sites are surface-exposed on the mature ribosome, possibly enabling other proteins and co-factors to interact with the modified residues.

Intriguingly, one 18S modification site is in close proximity to a possible RNA-binding protein binding site. In 2022, *Kipper et al.* identified a potential binding site of the RNA-binding protein G3BP2 near the E site on the 40S subunit (Kipper et al. 2022). Using cross-linking coupled mass-spectrometry, they determined that there is an interaction between G3BP2 (residue 248 or 332) and RPS28 (residue 16). Additionally, Kipper et al. showed that G3BP2 is a major component of cerebral cortex-derived RNA granules, linking G3BP2 interaction to ribosome translocation in neurons. Using distance modelling in ChimeraX, it was determined that the G3BP2 binding site - RPS28 K16 - is 19.2Å away from the 18S U1368 modification site. This short distance may enable the pseudouridylated-residue to stabilise the interaction between G3BP2 and RPS28.

Overall, analysis of snoRNA genes and the 18S modification sites has provided evidence in support of neuronal ribosome specialisation. In particular, the localisation of the 18S pseudouridylation sites hints at enabling two major functions; stabilising both the rRNA conformation and ribosome-protein interactions. The clustering of modification sites within the internal core of the 18S rRNA may allow for RNA-RNA stabilisation, whilst the presence of surface-exposed modification sites points to the ability of modified residues to interact with ribosome binding proteins and co-factors. This may assist with the stabilisation of RNA-protein binding, providing a potential mechanism for adapter-ribosome interaction and enabling directed transport of axonal ribosomes.

6.2.4 Detection of Brain-Elevated Ribosome Protein Genes by mRNA-Seq Analysis

Given that the bioinformatic analysis in this chapter provides evidence in favour of ribosome specialisation in brain regions, it was intriguing to see whether these results can be replicated experimentally. Therefore, RNA was extracted from rat primary cortical neurons (see Section 2.5.1) and mRNA-seq was conducted to determine if any of the brain-upregulated ribosome associated genes could be identified. As the mRNA-seq service (provided by Novogene) could not accurately detect mRNAs under 200bp, this analysis was done primarily to identify ribosomal protein genes (such as the brain-upregulated paralogs RPL7L1 and RPL26L1). Additionally, to determine whether expression of ribosomal protein genes changes over time, RNA samples from DIV 0 and DIV 14 cultures were analysed to detect differential gene expression.

Using the raw read count data (see Section 2.6 for library preparation, quality analysis and expression analysis protocols), differential expression of genes between DIV 0 and DIV 14 neurons was analysed using Degust software (Powell 2019) (Figure 6.9). Overall, ribosomal protein genes were downregulated in DIV 14 neurons, which is reflective of the decreased translational requirements in mature neurons

compared to DIV 0 cultures. Conversely, subsequent GO term analysis of the DIV 14 genes shows that the upregulated genes in this sample are associated with neuronal function, including; synaptic signalling, synapse assembly, and propagation of action potentials. Notably, expression data for neither RPL7L1 or RPL26L1 could be identified within the dataset. This may indicate that the system and protocol used in this assay requires further optimisation, perhaps by increasing the RNA quantity or using a more sensitive sequencing methodology. Importantly, only two biological replicates were used in this mRNA-seq analysis (N = 2). Repeating this work with more biological replicates is therefore necessary before any firm conclusions can be drawn about gene expression patterns in DIV0 and DIV14 cortical neuron cultures.

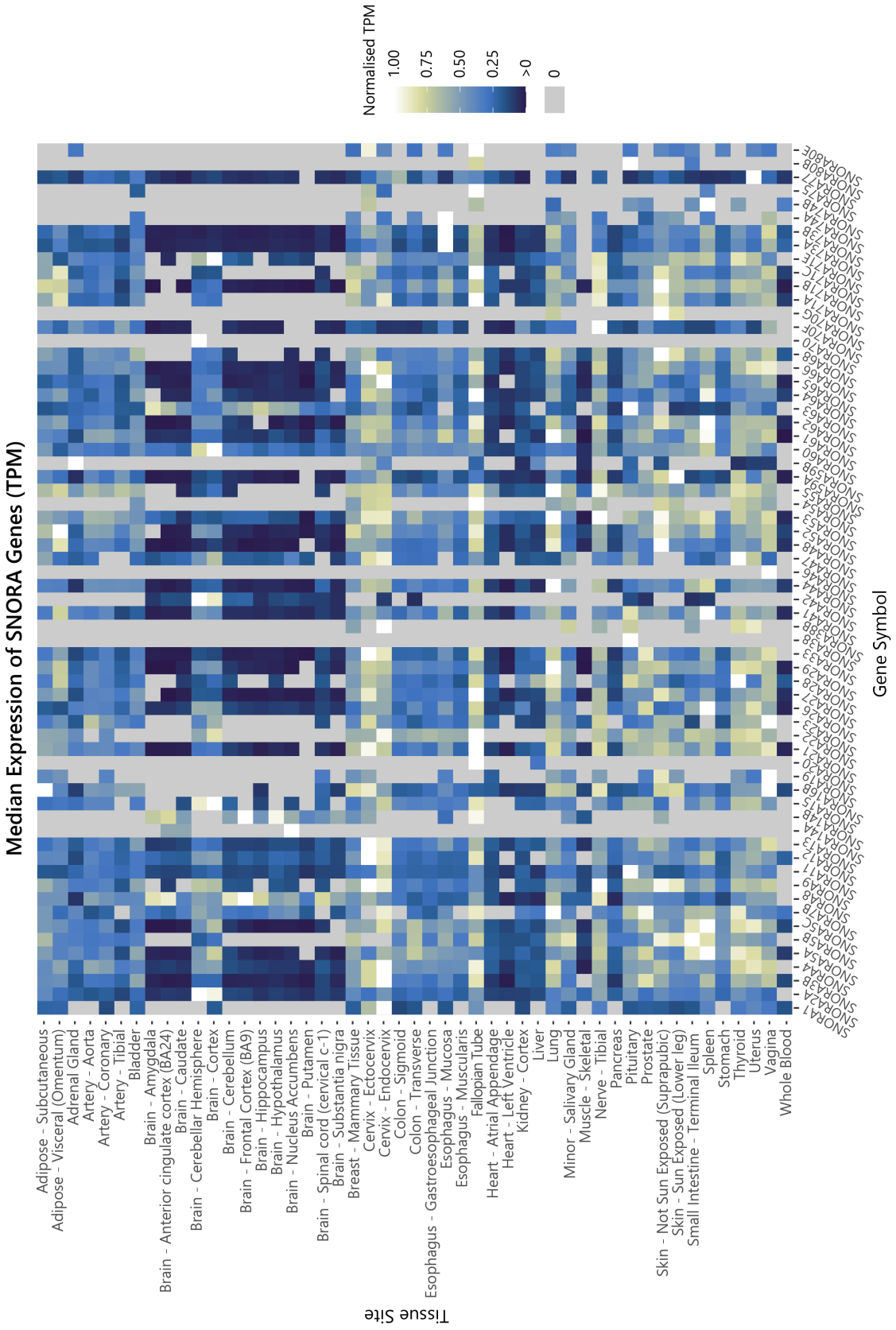

Median Expression of SNORA Genes (TPM)
Normalised TPM
1.00
0.75
0.50
0.25
>0
0
Tissue Site
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cortex
Brain - Cerebellum
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood
Gene Symbol

Figure 6.4: Gene Expression Analysis of SNORA Genes.

Heatmap visualising the gene expression of 137 SNORA genes across 50 tissue sites. Gene expression data was provided by the GTEx Portal V8. Expression levels for each gene are given in transcripts per million (TPM) with the median TPM calculated for every gene in each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from >0 to 1. Data points for genes with a median TPM of 0 are illustrated in grey and assumed to not be expressed at significant levels within that tissue. See the Appendix for all snoRNA genes analysed (Table 8.2) and the code used to perform gene expression analysis (Figure 8.5).

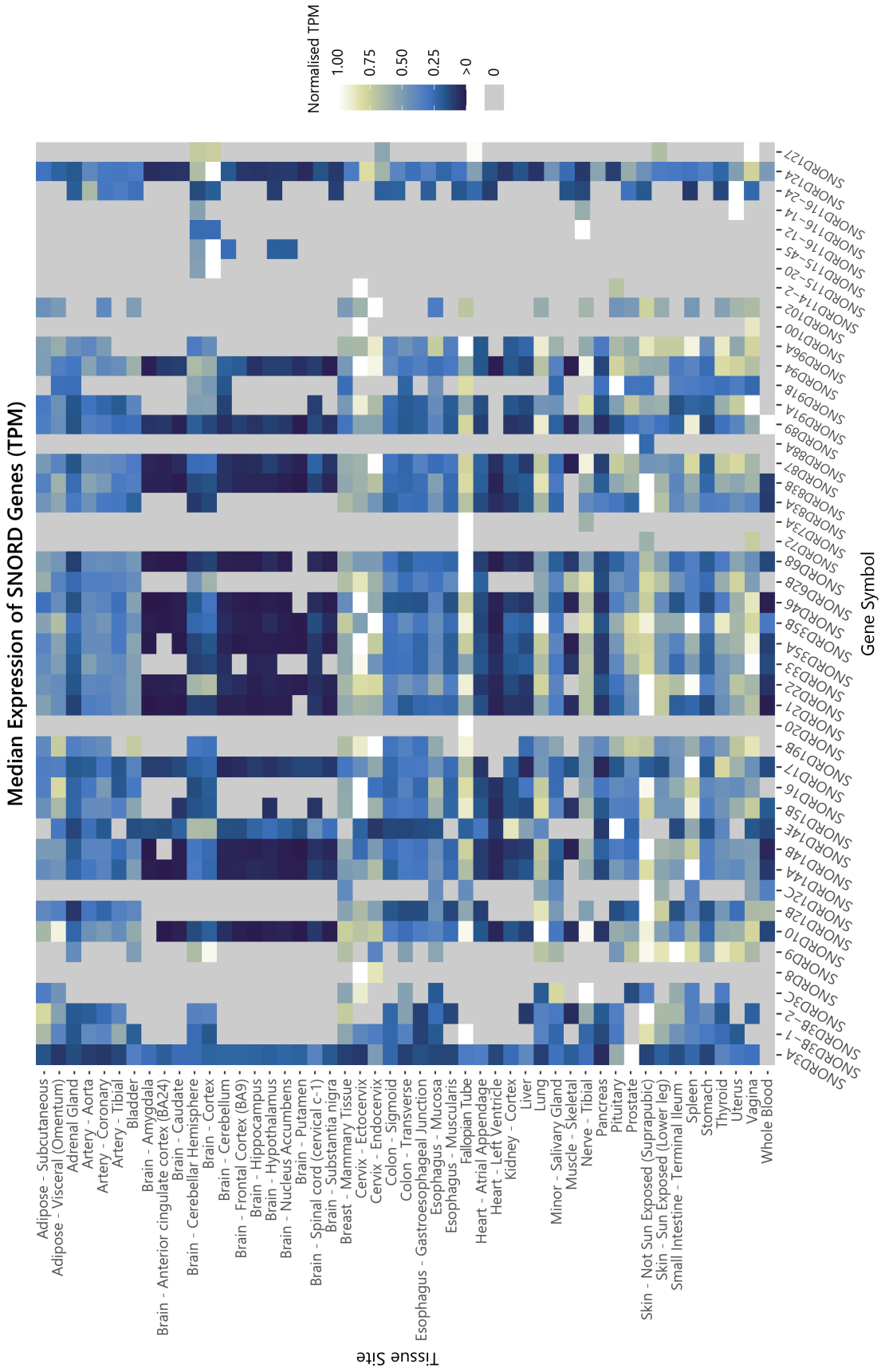

Normalised TPM
1.00
0.75
0.50
0.25
>0
0
Median Expression of SNORD Genes (TPM)
Tissue Site
Gene Symbol
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cortex
Brain - Cerebellum
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood
SNORD34
SNORD38A
SNORD38-1
SNORD38-2
SNORD3C
SNORD8
SNORD9
SNORD10
SNORD12B
SNORD12C
SNORD14A
SNORD14B
SNORD14C
SNORD15A
SNORD15B
SNORD16
SNORD17
SNORD19B
SNORD20
SNORD21
SNORD22
SNORD33
SNORD35A
SNORD35B
SNORD46
SNORD62B
SNORD68
SNORD72
SNORD73A
SNORD83A
SNORD83B
SNORD87
SNORD88A
SNORD89
SNORD91A
SNORD91B
SNORD94
SNORD96A
SNORD100
SNORD102
SNORD114-2
SNORD115-20
SNORD115-45
SNORD116-12
SNORD116-14
SNORD116-24
SNORD124
SNORD127

Figure 6.5: Gene Expression Analysis of SNORD Genes.

Heatmap visualising the gene expression of 266 SNORD genes across 50 tissue sites. Gene expression data was provided by the GTEx Portal V8. Expression levels for each gene are given in transcripts per million (TPM) with the median TPM calculated for every gene in each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from >0 to 1. Data points for genes with a median TPM of 0 are illustrated in grey and assumed to not be expressed at significant levels within that tissue. See the Appendix for all snoRNA genes analysed (Table 8.2) and the code used to perform gene expression analysis (Figure 8.5).

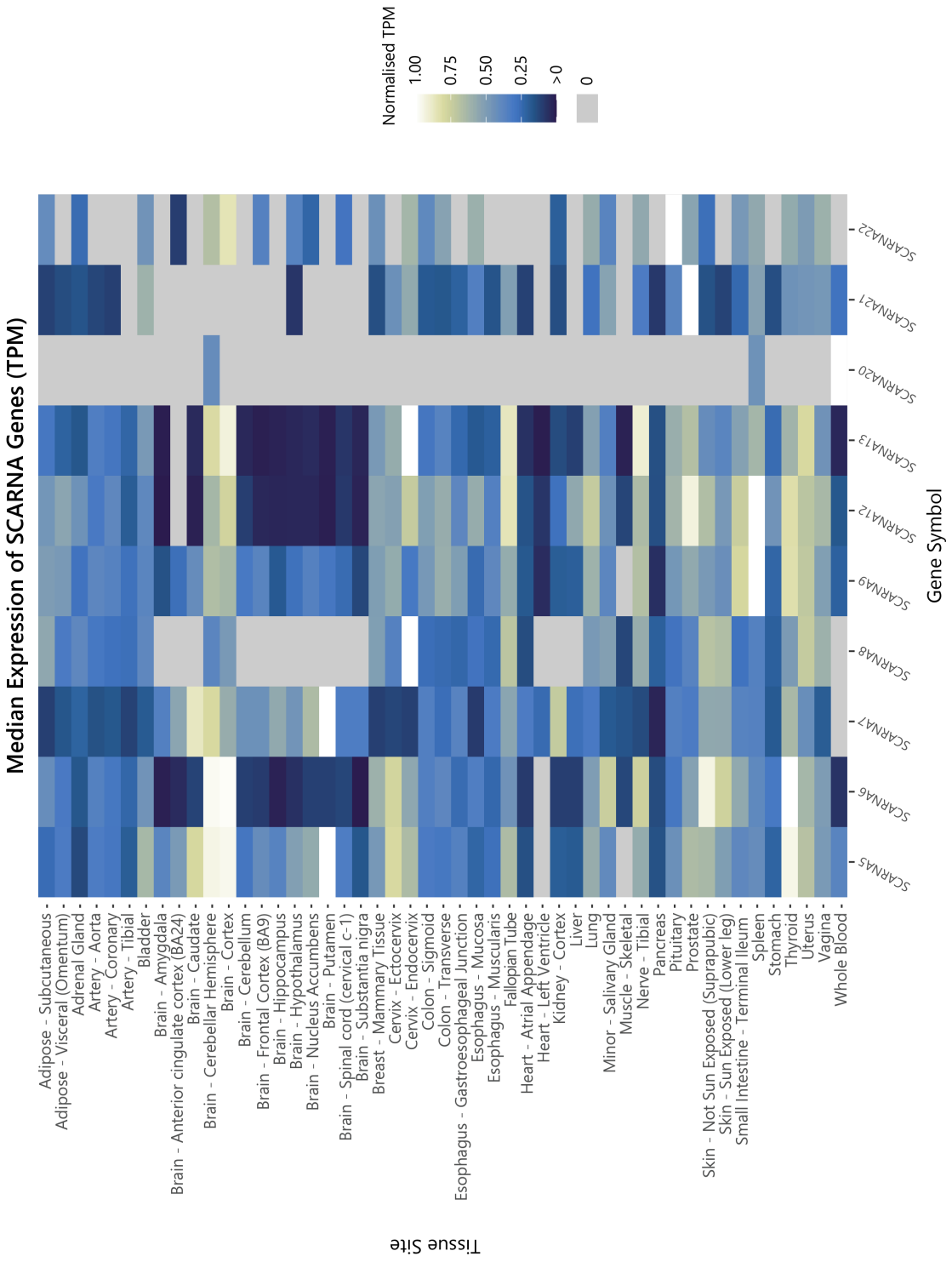

Normalised TPM
1.00
0.75
0.50
0.25
>0
0
Median Expression of SCARNA Genes (TPM)
Gene Symbol
SCARNA22
SCARNA21
SCARNA20
SCARNA13
SCARNA12
SCARNA9
SCARNA8
SCARNA7
SCARNA6
SCARNA5
Tissue Site
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate
Brain - Cerebellar Hemisphere
Brain - Cortex
Brain - Cerebellum
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus Accumbens
Brain - Putamen
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor - Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Thyroid
Uterus
Vagina
Whole Blood

Figure 6.6: Gene Expression Analysis of SCARNA Genes.

Heatmap visualising the gene expression of 22 SCARNA genes across 50 tissue sites. Gene expression data was provided by the GTEx Portal V8. Expression levels for each gene are given in transcripts per million (TPM) with the median TPM calculated for every gene in each tissue. All data was normalised to provide TPM as a proportion of the maximum TPM for each gene, ranging from >0 to 1. Data points for genes with a median TPM of 0 are illustrated in grey and assumed to not be expressed at significant levels within that tissue. See the Appendix for all snoRNA genes analysed (Table 8.2) and the code used to perform gene expression analysis (Figure 8.5).

Table 6.2: Expression of Brain-Elevated snoRNA Genes and Host Genes.

snoRNA heatmaps were analysed to identify genes with the highest median TPM value in brain regions. The highest non-brain tissue expression of the gene is also provided for comparison. snoRNA and snoRNA host genes were provided by the GTEx Portal V8 (GTEx Consortium 2020). Abbreviations: Adrenal - Adrenal Gland, Art. Tib. - Artery (Tibial), Cere. - Cerebellum, Ecto. - Ectocervix, Endo. - Endocervix, Fallopian - Fallopian Tube, Hemi. - Cerebellar Hemisphere, Nuc. - Nucleus accumbens, Tib. - Tibial Nerve.

Gene	Highest Brain Expression (TPM)	Highest Non-Brain Expression (TPM)	Host Gene	Highest Brain Host Expression (TPM)	Highest Non-Brain Host Expression (TPM)
SNORA2A	2.42 (Hemi.)	1.35 (Tib.)	KANSL2	36.45 (Hemi.)	34.64 (Endo.)
SNORA8	50.96 (BA9)	36.29 (Lung)	TAF1D	26.39 (Cere.)	55.11 (Ecto.)
SNORA14A	0.25 (Nuc.)	0 (All)	POR	21.33 (Cere.)	164.10 (Adrenal)
SNORA14B	0.49 (BA9)	0.47 (Fallopian)	TOMM20	191.90 (BA9)	170.40 (Art. Tib.)
SNORA15	1.17 (Cere.)	0.79 (Fallopian)	CCT6A	125.10 (Hemi.)	123.20 (Uterus)
SNORA42	3.2 (Hemi.)	0.48 (Pituitary)	KIAA0907	62.80 (Hemi.)	74.47 (Endo.)
SNORA60	2.96 (Cere.)	2.43 (Kid. Med.)	SNHG11	22.34 (Cere.)	29.11 (Tib.)
SNORA70J	0.2 (Hemi.)	0 (All)	RPSAP71	0.43 (Cere.)	0 (All)
SNORA71E	1.28 (Cere.)	1.1 (Tib.)	SNHG11	22.34 (Cere.)	29.11 (Tib.)
SNORD115-20	0.68 (Cere.)	0 (All)	SNHG14	32.35 (Cere.)	27.89 (Thyroid)
SNORD115-45	1.27 (Cere.)	0 (All)	SNHG14	32.35 (Cere.)	27.89 (Thyroid)
SNORD124	4.46 (Cere.)	3.66 (Ecto.)	MED24	72.97 (Cere.)	62.94 (Uterus)
SCARNA5	0.66 (Putamen)	0.63 (Thyroid)	ATG16L1	56.59 (Hemi.)	35.80 (Thyroid)

Table 6.3: The Modification Sites of Brain-Elevated snoRNA Genes and Modification Sites.

snoRNA genes and associated Genecode IDs were provided by the GTEx Portal V8 (GTEx Consortium 2020). Additional snoRNA details including type and target site were provided by the Human snoRNA Atlas (Jorjani et al. 2016). The first target site prediction listed refers to the 5' antisense element. Where one or more target sites are provided, these reflect targets for the 3' antisense element or additional hybrid scaRNA elements respectively. Target site abbreviations match those provided on the Human snoRNA Atlas; R - reported target prediction, C - conserved target prediction, H - best human prediction.

Gene	Gencode ID	Type	Modification	Target Site
SNORA2A	ENSG00000206612.1	H/ACA	pseudouridylation	RH:28S-4263, RC:28S-4282
SNORA8	ENSG00000207304.1	H/ACA	pseudouridylation	RCH:18S-1056, RCH:18S-1081
SNORA14A	ENSG00000201643.1	H/ACA	pseudouridylation	CH:18S-681, RCH:18S-966
SNORA14B	ENSG00000207181.1	H/ACA	pseudouridylation	CH:18S-681, RCH:18S-966
SNORA15	ENSG00000207062.1	H/ACA	pseudouridylation	CH:18S-1368
SNORA42	ENSG00000207217.1	H/ACA	pseudouridylation	RC:18S-572, RCH:18S-109
SNORA60	ENSG00000199266.1	H/ACA	pseudouridylation	H:U11.1-97, C:18S-406
SNORA70J	ENSG00000206909.1	H/ACA	pseudouridylation	C:18S-1232, H:U4.2-40
SNORA71E	ENSG00000274309.1	H/ACA	pseudouridylation	H:18S-1377, H:U4atac.1-19
SNORD115-20	ENSG00000201969.1	C/D	methylation	Unknown
SNORD115-45	ENSG00000212380.1	C/D	methylation	Unknown
SNORD124	ENSG00000238793.1	C/D	methylation	CH:U4atac.1-7
SCARNA5	ENSG00000252010.1	hybrid	both	RCH:U5.1-39, C:18S-595, C:18S-1530, RCH:U4.1-65

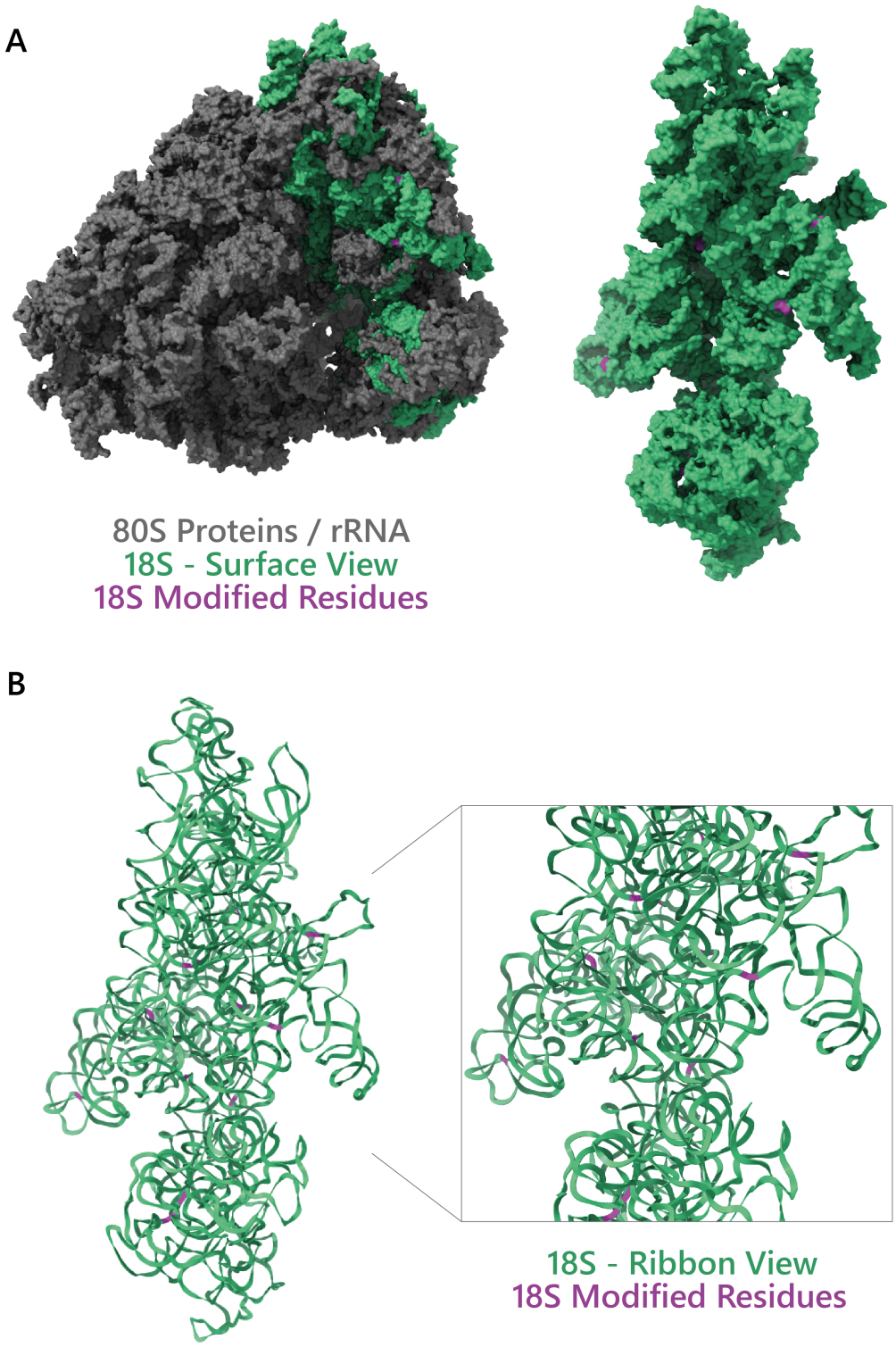


Figure 6.7: 18S Pseudouridylation Sites Mapped onto the 80S Ribosome.

The 18S pseudouridylation modification sites identified were mapped onto the 80S human ribosome (Protein Data Bank ID: 4UG0) using UCSF Chimera X, version 1.5 (Pettersen et al. 2021). A) Surface view of the 80S ribosome, including the 18S rRNA (green), modification sites (magenta), and all other ribosome proteins and rRNA (grey). B) Ribbon view of the isolated 18S rRNA (green) with a closeup of the modification sites (magenta) showing clustering of the sites towards the central region of the rRNA.

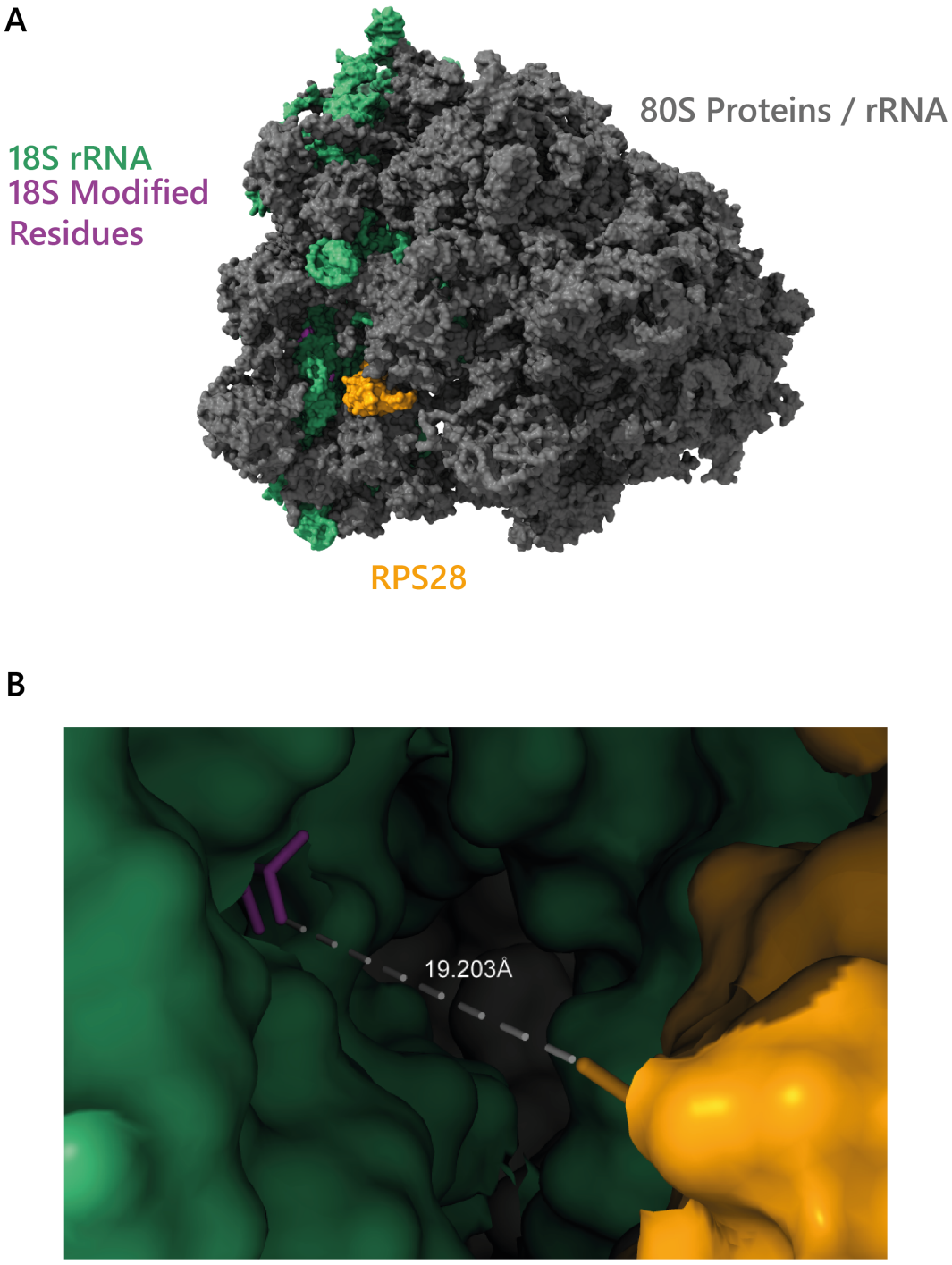


Figure 6.8: 18S U1368 is in Close Proximity to Potential Binding Site RPS28 K16.

Several 18S pseudouridylation modification sites are surface-exposed, with one site in close proximity to a potential G3BP2 binding site (Kipper et al. 2022). Computational modelling was done using the 4UG0 80S human ribosome (Khatter, Myasnikov, Natchiar, et al. 2015) using UCSF Chimera X, version 1.5 (Pettersen et al. 2021). A) Surface view of the 80S ribosome, including the 18S rRNA (green), modification sites (magenta), RPS28 (yellow) and all other proteins and rRNA (grey). B) Distance modelling between the two closest atoms of 18S U1368 (magenta) and RPS28 K16 (yellow).

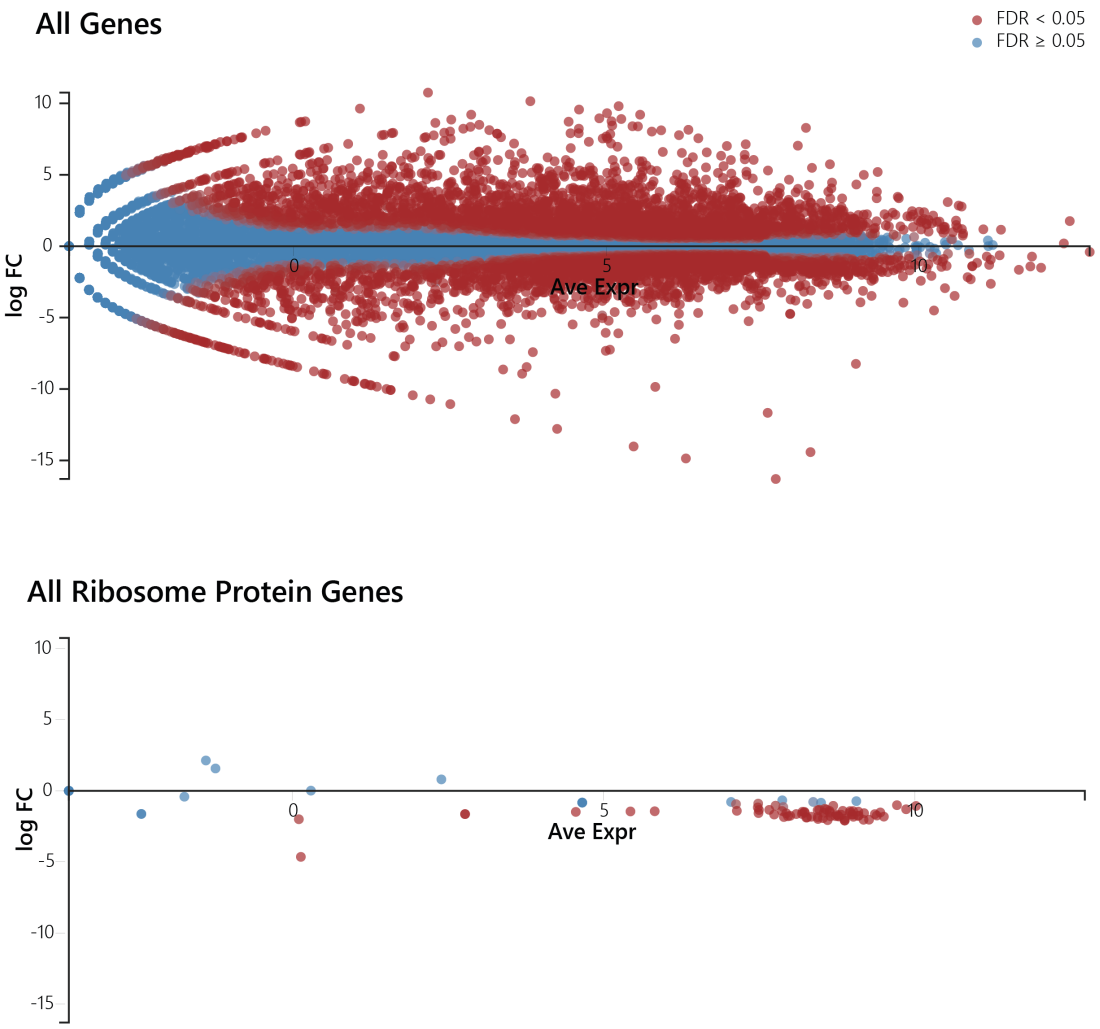


Figure 6.9: MA Plot of Differential Gene Expression in DIV 0 and DIV 14 Primary Neurons.

RNA was extracted from DIV 0 or DIV 14 rat primary cortical neurons and analysed by mRNA-seq (see Section 2.6 for further details on next generation sequencing library preparation, quality analysis, and expression analysis). For each condition, RNA samples from two biological replicates were analysed. Differential gene analysis was done using read count data using Degust (version 4.1.1)(Powell 2019). As per the standard Degust settings, genes with a false discovery rate (FDR) of below 0.05 are considered significantly differentially expressed and denoted by a red circle. Genes with an expression FDR of 0.05 and above are considered not significantly differentially expressed and denoted by a blue circle. Here, all genes with a FDR of 1 or below are included. N = 2.

6.3 Discussion

6.3.1 Limitations of RNA-Seq Datasets

Before drawing conclusions from analysis of RNA-seq data, we must first acknowledge the common limitations of such studies. RNA-seq data is subject to bias and contamination which can cause subtle data changes leading to inaccurate interpretations of results. This is largely due to the complex workflow of RNA sequencing which involves RNA extraction from samples, library preparation, sequencing and imaging, and subsequent read data processing (H. Shi et al. 2021). At each step, the specific populations of RNA reads can be over or underrepresented. For example, GC content bias is a common issue that can affect library preparation for some next-generation sequencing platforms (Benjamini and Speed 2012). As GC-rich sequences are less likely to be amplified by PCR, fewer reads are mapped to GC-rich areas of the genome. As a result, RNA segments with these base compositions are commonly under-represented in Illumina read data (Aird et al. 2011).

However, careful consideration was taken to minimise experimental bias and error when collecting the GTEx data. Firstly, RNA was predominantly extracted from samples using Qiagen column-based miRNA kits (Null et al. 2015) which enhance extraction of small RNAs (<200 nucleotides), avoiding over-representation of long mRNAs within the library. RNA sequencing was then carried out using the Illumina HiSeq 2500 (GTEx Consortium 2020), a platform that was determined to have one of the lowest error rates (approximately 0.1%) out of all the Illumina platforms (Stoler and Nekrutenko 2021). Lastly, read data underwent robust quality control checks to exclude poor quality reads, outliers, and samples from donors with cytogenetic abnormalities (GTEx Consortium 2020). Despite this, some level of bias is likely to be present within the GTEx dataset as it is unfeasible to control every experimental factor when conducting RNA-Seq assays (H. Shi et al. 2021).

One major limitation of GTEX expression analysis is RNA sequencing contamination, whereby sequencing data for one sample is reflected in the data for another

sample. In a recent study, it was determined that low-level sequencing contamination could affect up to 40% of GTEx's samples (Nieuwenhuis et al. 2020). Through analysis of several highly expressed pancreas-enriched genes, the study found that there was unexpected low-level expression of pancreas-specific genes in a range of other tissues. Samples with data contamination were found to strongly correlate to the date of sample sequencing. The authors conclude that contamination was possibly as a result of library preparation error or the sequencing reaction itself. In the analysis of ribosomal proteins and snoRNAs, it is unlikely that sequencing contamination greatly affects the conclusions herein. The variation observed by Nieuwenhuis et al. affected highly expressed genes with tissue-specific expression. This analysis identified genes with the highest TPM levels in brain regions, so sequencing contamination producing low TPM expression in other tissues should have a minimal effect on these results.

In summary, RNA-seq datasets can provide useful insight into gene expression but commonly suffer from sequencing bias and contamination. It is therefore useful to recognise the limitations of both RNA-Seq and gene expression databases before making generalisations about gene expression patterns - especially when the variation between tissues is small. Additionally, it must be noted that mRNA expression is not necessarily perfectly correlated to protein production (Khan et al. 2013; Edfors et al. 2016) so additional proteomics work is essential to validate claims made by RNA-Seq analysis.

6.3.2 Further RNA-Seq Evidence of Heterogeneous Ribosome Composition in the Nervous System

This chapter provides data in support of ribosome heterogeneity in brain regions, observing some changes in ribosome protein expression and rRNA modification which may allow for specialised functionality in neurons. Specifically, it was determined that some ribosome protein paralogs and isoforms show highest RNA expression levels within brain tissues, in addition to expression of several upregulated and

brain-specific snoRNAs. This data contributes to a wider body of experimental and bioinformatic analysis which proposes that some neuronal ribosomes are compositionally distinct (see Chapter 1 for a thorough examination of these studies).

In a 2016 study investigating the existence of the "modular ribosome" across species, analysis of multiple RNA-Seq datasets also produced computational evidence in favour of ribosome heterogeneity in the nervous system (Yadav et al. 2016). *Yadav et al.* discuss ribosome adaptation which allow organisms to thrive in specific environments and contexts - such as thermophiles whose ribosomes can withstand extreme temperatures. Therefore, the authors postulate that these differing ribosomal protein compositions across species could form "modules" which work together to allow the ribosome to adapt to environmental cues. By examining ribosome protein RNA-seq data from ENCODE and GTEx databases, they used a novel ranking system to group ribosomal proteins with similar expression patterns across tissues, identifying modules of ribosomal proteins with similar expression in different tissues.

The data from this study corroborated some of the findings from the GTEx analysis presented in this Thesis. Overall, *Yadav et al.* determined that tissues of the nervous system showed lower expression of ribosome protein mRNA expression when compared to most other tissues in humans and mice. Although expression of ribosomal proteins is reduced, the analysis also determined that the nervous system utilises a distinct set of ribosomal proteins when compared to other tissues. The authors also state that the more recently evolved ribosomal protein paralogs - including RPL7L1 - showed low expression and high tissue-specificity compared to other ribosomal proteins. These findings support the data herein, although further experimental work is required to validate the presence of neuronal ribosome heterogeneity.

6.3.3 Subsequent Experiments To Prove Neuronal Ribosome Specialisation

Several elements of the GTEx analysis could be optimised to generate more potential hits for future proteomics work. One clear example is the inclusion of any ribosome protein or snoRNA gene which shows brain expression above the median TPM value for all tissues. The analysis presented here identified genes with the highest expression in brain regions. This simplified methodology was chosen as it made brain-elevated transcripts easier to detect and also selected genes that had the highest chance to play a significant role in neuronal heterogeneity. In contrast, identification of all genes with above average expression could produce more hits with potential effects on ribosome composition. In addition, splicing analysis of ribosome-associated transcripts is also of interest. Similar to the exon expression data, analysis of alternative splicing data could provide another method of differential gene expression, with splicing regulating gene expression and transcript stability. Although junction expression data is provided in the GTEx portal, this was not investigated due the complex computational requirements of the analysis. However, future work into splicing analysis could provide key details into differential expression of genes in brain regions.

The GTEx dataset also provides expression data for the Tibial nerve, a section of the sciatic nerve composed of a group of axons surrounded by connective tissue. As a component of the peripheral nervous system, it was assumed that this tissue could provide some valuable insights into axonal ribosome protein expression. However, an in-depth analysis of Tibial nerve RNA-Seq data was not conducted for several reasons. The initial ribosome protein analysis (Figure 6.1) indicates that the Tibial nerve has much higher expression of ribosome proteins compared to brain regions. This was an unusual result, as the non-dividing axons of the Tibial nerve are expected to have low levels of ribosomal protein expression compared to other non-brain tissues. Additionally, the Tibial nerve appears to have elevated expression of some snoRNAs in this analysis, including SNORD3C and SNORA70F. As this

tissue predominantly contains axons, expression of snoRNAs (which show nuclear localisation) is expected to be minimal. This gene expression may originate from non-neuronal cells present in the Tibial nerve, therefore, the Tibial nerve RNA-Seq data is unlikely to provide neuron-specific information about axonal ribosomal protein gene expression.

Further analysis of ribosome-associated factors within brain regions could provide additional evidence of neuronal ribosome diversity. As previously stated, ribosome heterogeneity can be divided into three subcategories; ribosomal protein composition, rRNA and its modifications, and interaction with ribosome-associated factors. To simplify analysis this work focussed on the first two categories, as these involve known constituents which could be easily examined using existing gene lists. However, interaction with ribosome-associated factors should not be overlooked as a major source of translational regulation and ribosome heterogeneity (Joo et al. 2022). A study utilising a novel ribosome affinity purification and mass spectrometry identified hundreds of ribosome-associated proteins with a variety of cellular roles, allowing the creation of a "ribo-interactome" in mouse embryonic stem cells (Simsek et al. 2017). Data from these ribosome mass spectrometry studies could provide an excellent starting point when curating a list of ribosome-associated factors for gene expression analysis.

Future proteomic studies of neuronal ribosomes could provide more insight into ribosome heterogeneity in the nervous system. In particular, the purification of axonal ribosomes using the HaloTag from CRISPR-edited neurons would allow comparison between the ribosome structure of axonal versus their soma-derived counterparts. Further analysis of axonal ribosomes by mass spectrometry would deliver invaluable understanding of ribosomal protein stoichiometry, paralog presence, and isoform expression. The paralogs and isoforms identified from this GTEx analysis are of particular interest, especially since differential expression of ribosomal protein variants may explain how local remodelling of axonal ribosomes is possible (Shigeoka, Koppers, et al. 2019; Fusco et al. 2021). In particular, the presence of the paralog

RPL7L1 within axonal ribosomes would be intriguing, as computational studies have linked this paralog to an axonal form of the Charcot-Marie-Tooth disease (Type 2N) (Bis-Brewer et al. 2019).

Finally, developments in Nanopore long-read sequencing allows detection of many RNA modifications, including methylation and pseudouridylation (reviewed in White and Hesselberth 2022). Sequencing of axonal ribosome rRNA, and in particular 18S rRNA pseudouridylation detection, could provide insight into unique modifications present on axonal ribosomes that allow them to be transported into axons to carry out local translation. Nanopore can also be used for transcriptome studies, detecting both protin-coding mRNAs and smaller non-coding RNAs - such as snoRNAs (Parker et al. 2020). This analysis identified several brain-specific snoRNAs (for example, SNORD115-45) which validates claims that the brain expresses several highly-tissue specific snoRNAs including those of the SNORD115 family (Fafard-Couture et al. 2021). Additional nanopore analysis of Halo-tagged axonal ribosomes and their modifications would therefore substantiate this RNA-Seq analysis and significantly contribute to our understanding of neuronal ribosome heterogeneity.

Chapter 7

Discussion

7.1 Summary of Findings

The work in this thesis documents the creation and use of an exciting tool to visualise the transport of endogenous ribosomes within live cells. This is achieved by using a knock-in CRISPR system, ORANGE, to introduce a self-labelling enzyme into the coding region of ribosomal proteins. Four ribosomal protein targets were chosen (RPL4, RPL22, RPS2 and RPS3), with two guide RNA sequences designed for each protein (see Chapter 3). Subsequently, these guides were cloned into the PX458 plasmid and guide cutting efficiency analysis was conducted to determine which guides were the most efficient at creating Cas9-mediated cutting. Using the selected guides, a pORANGE-Halo plasmid was created for each ribosomal protein target. By electroporating PC-12 cells, stable CRISPR lines were generated which allowed knock-in validation (see Chapter 4). Of the four constructs created, pORANGE-RPL4-Halo and pORANGE-RPS2-Halo showed successful HaloTag expression at the desired genomic locus. Additionally, it was determined that the Halo-tagged protein was incorporated into functional ribosomes in both cell lines.

These validated ORANGE constructs, alongside the positive control pORANGE-GFP-ACTB #1 plasmid, were then transfected into rat primary cortical neuron cultures for live cell imaging assays (see Chapter 5). A low number of GFP-ACTB positive neurons were imaged using TIRF microscopy, highlighting the need for

further optimisation of neuronal transfection methods. However, analysis of the validated PC-12 RPL4-Halo line showed the directed transport of ribosomes in PC-12 processes. This movement was bi-directional, with velocities which align with fast axonal transport. Additionally, ribosome transport appears to co-localise with the movement of lysosomes/endosomes, which supports the hitchhiker theory of neuronal RNA granule transport (see Section 1.5.2). This data is the first microscopic evidence of cell body-derived ribosomes being actively transported in a neuronal model system.

Lastly, expression analysis of ribosome-associated genes was carried out to determine whether any genes were differentially expressed within the nervous system (see Chapter 6). This work aimed to identify genes which may contribute to neuronal ribosome heterogeneity as structural changes may provide insights into axonal ribosome translocation. Analysis of ribosome protein genes determined that two paralogs - RPL7L1 and RPL26L1 - had brain-elevated expression, and a number of upregulated ribosomal protein isoforms were also identified. In addition, 13 snoRNAs also showed upregulated expression in brain regions. Most of the snoRNAs identified directed the pseudouridylation of 18S rRNA, with one site in close proximity to a possible RNA binding protein interaction site on the ribosome. This gene expression data indicates that there is some level of heterogeneity in neuronal ribosomes which may enable their axonal transport, but further proteomic work is needed.

7.2 Considerations for Creating ORANGE Constructs

The ORANGE system is a straightforward and relatively fast way to generate knock-in CRISPR constructs. To generate effective guide RNA sequences, the Doench on-target and the MIT off-target scores are often taken into consideration, allowing the selection of an efficient and specific guide. Interestingly, *Willems et al.* found

no strong correlation between Doench score and knock-in efficiency (Willems et al. 2020). For example, the authors create three GFP-GluN1 (Glutamate receptor NMDA 1) constructs which varied in knock-in efficiency. These constructs contained different guide and PAM sites which were all in close proximity. The third construct had the highest knock-in efficiency (5.6% ± %0.4) compared to the other guides (#1= 0.43% ± %0.04 and #2= 3.0% ± %0.7) despite having the lowest Doench score of just 13%. For comparison, constructs #1 and #2 had scores of 56% and 46% respectively. Taking this in account, creating multiple constructs using different guide RNAs (ignoring their Doench score) is likely to result in successful knock-in.

Selection of an appropriate tag is also an important consideration. As previously discussed, HaloTag was selected for this project as it allowed for increased functionality and had been previously used for imaging experiments in the Twelvetrees lab. However, future ORANGE experiments may benefit from the use of a smaller tag. At 33kDa, the HaloTag is a large self-labelling enzyme, especially in comparison to the ribosomal proteins being tagged (the smallest protein, RPL22, is just 15kDa). It is generally accepted that large protein tags may interfere with native protein folding and function (Suyama et al. 2023). Therefore, using a smaller self labelling enzyme - such as the 19.4kDa SNAP-tag (Keppler et al. 2003) - may increase fusion protein expression.

Whilst a valuable tool when tagging endogenous ribosomes, the production of stable CRISPR cell lines may not be a necessary step for most CRISPR targets. In this work, PC-12 Halo lines were made to verify tagged protein incorporation into ribosomes via polysome profiling, which requires a large amount of CRISPR-edited cells. It also made other validation assays (such as western blots) easier, as the amount of edited protein material was not a limiting factor and multiple transfections were not necessary. However, generating stable CRISPR lines is time consuming. For well-expressed neuronal protein targets, electroporation followed by DNA sequencing and imaging may be adequate for CRISPR construct validation - as was done for the original ORANGE constructs (Willems et al. 2020). However,

for proteins which are part of a large complex or have low endogenous expression in neurons, the production of stable CRISPR lines in neuronal model systems (such as PC-12) may be a valuable tool for validation assays before use in primary cultures.

7.3 Building a Model for Ribosome Transport in Axons

In this thesis, the data indicates that ribosomes are transported by fast axonal transport, moving at speeds ranging from 0.3 - 1.2 µm/second. This aligns with many other studies investigating neurons RNA granule movement (see Section 5.3.2), although these studies predominantly track RNA granule transport through fluorescently labelled RNA binding proteins instead of direct visualisation of ribosomes. Axonal transport is carried out by kinesin and dynein motors, and there is evidence that kinesin-1 and kinesin-2 play a role in neuronal RNA granule transport (Fukuda et al. 2021; Baumann, Komissarov, et al. 2020). However, it is possible that other kinesins (such as the neuron-specific kinesin-3) may also be involved. In summary, the data presented herein supports the RNA granule theory of ribosome transport in neurons, although transport of non-granule ribosomes may also be occurring.

Further work is needed to elucidate the ribosome transport mechanism. In particular, how ribosomes are selected from the cell body population and are packaged into RNA granules is an intriguing question. However, after thoroughly reviewing the literature surrounding ribosome specialisation and neuronal RNA granules, I propose two possible explanations. The first theory is that axonal ribosomes are differentially modified by snoRNAs, which allows RNA binding proteins to interact with the ribosome, leading to RNA granule formation. Recent research has illustrated that certain rRNA modifications may influence the binding of ribosomal factors (Khoshnevis et al. 2022) to mature ribosomes. Additionally, cross-linking studies have determined that the RNA granule-associated protein, G3BP2, interacts with ribosomal protein RPS28 (Kipper et al. 2022) which is in close proximity

to a brain-elevated snoRNA modification site. Taken together, it is possible that specific rRNA modifications mediate interactions between heterogeneous axonal ribosomes and RNA granule-associated RNA binding proteins.

Alternatively, another possible explanation is that specific mRNA elements promote the recruitment of axonal ribosomes and neuronal RNA granule formation. As previously discussed in Chapter 1, neuronal mRNA isoforms exist which differ in their 3' UTR sequence, with certain isoforms showing enrichment in specific neuronal compartments (Tushev et al. 2018). In addition, elements in the 5' UTR region have also been implicated in mRNA sorting and distribution in neurons (Colliva and Tongiorgi 2021). Studies into ribosome specialisation have indicated that heterogeneous ribosomes can preferentially translate distinct sub-populations of mRNAs, controlled by specific regulatory elements in the 5' UTR of the transcript (Z. Shi et al. 2017). As neuronal RNA granules are thought to contain mRNA transcripts with stalled ribosomes (Kipper et al. 2022), some neuronal mRNA isoforms may contain UTR elements which promotes their translation by heterogeneous axonal ribosomes. These polysome-associated transcripts could then be incorporated into RNA granules and transported into distal processes.

The data in this thesis demonstrates that ribosomes are created in the cell body and are actively transported into processes, also providing support for the "hitchhiker" mechanism for long-distance axonal transport. Notably, this work does not disprove the glial origin hypothesis. However, it does provide evidence that a large population of ribosomes in processes are generated in the cell body and transported using the existing transport machinery. It could be that both of these mechanisms are occurring in certain types of neurons, with Schwann cell transfer helping to supplement the ribosome population and support the local axonal proteome post-injury. Additionally, the work in this thesis also doesn't disprove the local ribosome remodelling observed in neuronal processes. Ribosome remodelling within processes may be a complementary mechanism to ribosome transport, essentially increasing the half life of axonal ribosomes and reducing the need (and energy demands) of

active ribosome transport by motor proteins. It appears likely that all three mechanisms are essential for ribosome localisation and maintenance, providing the means for rapid proteome adjustment to response to the neuronal activity and injury.

7.4 Future Work

The pORANGE Halo plasmids are a versatile tool that can offer valuable insights into the ribosome transport mechanism in neurons. For example, imaging assays using transfected CRISPR-edited cells could provide useful information about the speed of ribosome transport in processes. In a straightforward experiment, live cell imaging could be used to determine whether the speed of ribosome transport varies between axons and dendrites, or if the type of neuron affects transport velocity. The proportion of motile versus stationary ribosomes could also be calculated in individual neurons. Using this measurement, it would be interesting to determine how ribosome transport is affected by neuronal activity, such as stimulation or injury response. Similarly, axon-targeted treatment with protein inhibitors (such as puromycin) could be used to determine if ribosome transport is correlated with the rate of protein synthesis in the distal axon.

More in-depth proteomic studies could reveal further details about axonal ribosome translocation. In particular, isolation and purification of axonal ribosomes is possible using the HaloTag, which can then be used in pull down and mass spectrometry experiments. This could identify essential factors required for RNA granule assembly, whilst also providing more information about the types of RNA binding proteins and kinesins associated with axonal ribosome transport. Compositional analysis of cell body, axonal, and dendritic ribosomes could also generate findings in support of neuronal specialisation. Should a key component of the axonal ribosome transport mechanism be identified, subsequent work involving knock-down of this factor could investigate the effect of reduced axonal ribosome transport on presynaptic function. For example, synaptic vesicle recycling could be studied using styryl FM dyes which are reversibly incorporated into lipid membranes (Gaffield and

Betz 2006). Thus, fluorescence intensity signal can give an indication of exocytosis levels which is expected to decrease with reduced axonal ribosome transport.

Finally, visualisation of endogenous ribosomes in processes could be an excellent tool in neurodegenerative disease research. As previously discussed (see Chapter 1), a number of neurological disorders have been implicated in defective local translation, including SMA, ALS, and AD. Most of these conditions show evidence of axonal mRNA mislocalisation, which negatively impacts synaptic function. Other than one study which reported a decrease of axonal ribosomes in SMA mouse models (Bernabò et al. 2017), axonal ribosomes and their impact on local translation remains a largely unexplored field. Using disease models (such as induced pluripotent stem cell lines), defects in axonal ribosome localisation and transport could be identified through endogenous ribosome tagging. Additionally, understanding the basis of axonal ribosome transport may assist in future studies into novel therapeutics for disorders caused by impaired axonal transport and defective local translation of essential mRNA transcripts. In summary, the ORANGE Halo constructs created in this PhD can be used to answer a range of important biological questions, developing our understanding of the molecular mechanisms behind long-distance ribosome transport and how this is implicated in the pathology of neurodegenerative disease.

Bibliography

Abouward, Reem and Giampietro Schiavo (2021). "Walking the line: mechanisms underlying directional mRNA transport and localisation in neurons and beyond". *Cellular and molecular life sciences: CMLS* 78.6, pp. 2665–2681. ISSN: 1420-682X, 1420-9071. DOI: 10.1007/s00018-020-03724-3.

Agrawal, Manasi and Kristy Welshhans (2021). "Local Translation Across Neural Development: A Focus on Radial Glial Cells, Axons, and Synaptogenesis". *Frontiers in molecular neuroscience* 14, p. 717170. ISSN: 1662-5099. DOI: 10.3389/fnmol.2021.717170.

Aird, Daniel et al. (2011). "Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries". *Genome biology* 12.2, R18. ISSN: 1465-6906. DOI: 10.1186/gb-2011-12-2-r18.

Alabdullah, Abeer A et al. (2019). "Estimating transfection efficiency in differentiated and undifferentiated neural cells". *BMC research notes* 12.1, p. 225. ISSN: 1756-0500. DOI: 10.1186/s13104-019-4249-5.

Alami, Nael H et al. (2014). "Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations". *Neuron* 81.3, pp. 536–543. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2013.12.018.

Ashburner, Michael et al. (2000). "Gene Ontology: tool for the unification of biology". *Nature genetics* 25.1, pp. 25–29. ISSN: 1061-4036. DOI: 10.1038/75556.

Axelrod, D (1981). "Cell-substrate contacts illuminated by total internal reflection fluorescence". *The Journal of cell biology* 89.1, pp. 141–145. ISSN: 0021-9525. DOI: 10.1083/jcb.89.1.141.

Bachellerie, J P et al. (1995). "Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA". *Trends in biochemical sciences* 20.7, pp. 261–264. ISSN: 0968-0004. DOI: 10.1016/s0968-0004(00)89039-8.

Bakshi, Somenath et al. (2012). "Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells". *Molecular microbiology* 85.1, pp. 21–38. ISSN: 0950-382X, 1365-2958. DOI: 10.1111/j.1365-2958.2012.08081.x.

Baleriola, Jimena et al. (2014). "Axonally synthesized ATF4 transmits a neurodegenerative signal across brain regions". *Cell* 158.5, pp. 1159–1172. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2014.07.001.

Ban, N et al. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution". *Science* 289.5481, pp. 905–920. ISSN: 0036-8075. DOI: 10.1126/science.289.5481.905.

Bartlett, W P and G A Banker (1984). "An electron microscopic study of the development of axons and dendrites by hippocampal neurons in culture. II. Synaptic relationships". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 4.8, pp. 1954–1965. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.04-08-01954.1984.

Bassell, G J et al. (1998). "Sorting of beta-actin mRNA and protein to neurites and growth cones in culture". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 18.1, pp. 251–265. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.18-01-00251.1998.

Baßler, Jochen and Ed Hurt (2019). "Eukaryotic Ribosome Assembly". *Annual review of biochemistry* 88, pp. 281–306. ISSN: 0066-4154, 1545-4509. DOI: 10.1146/annurev-biochem-013118-110817.

Batista, Andreia F R et al. (2017). "Intra-axonal Synthesis of SNAP25 Is Required for the Formation of Presynaptic Terminals". *Cell reports* 20.13, pp. 3085–3098. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2017.08.097.

Baumann, Sebastian, Artem Komissarov, et al. (2020). "A reconstituted mammalian APC-kinesin complex selectively transports defined packages of axonal

mRNAs". *Science advances* 6.11, eaaz1588. ISSN: 2375-2548. DOI: 10.1126/sciadv.aaz1588.

Baumann, Sebastian, Thomas Pohlmann, et al. (2012). "Kinesin-3 and dynein mediate microtubule-dependent co-transport of mRNPs and endosomes". *Journal of cell science* 125.Pt 11, pp. 2740–2752. ISSN: 0021-9533, 1477-9137. DOI: 10.1242/jcs.101212.

Beckmann, R et al. (2001). "Architecture of the protein-conducting channel associated with the translating 80S ribosome". *Cell* 107.3, pp. 361–372. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(01)00541-4.

Ben-Shem, Adam et al. (2011). "The structure of the eukaryotic ribosome at 3.0 Å resolution". *Science* 334.6062, pp. 1524–1529. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1212642.

Ben-Yaakov, Keren et al. (2012). "Axonal transcription factors signal retrogradely in lesioned peripheral nerve". *The EMBO journal* 31.6, pp. 1350–1363. ISSN: 0261-4189, 1460-2075. DOI: 10.1038/emboj.2011.494.

Benjamini, Yuval and Terence P Speed (2012). "Summarizing and correcting the GC content bias in high-throughput sequencing". *Nucleic acids research* 40.10, e72. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gks001.

Bernabò, Paola et al. (2017). "In Vivo Translatome Profiling in Spinal Muscular Atrophy Reveals a Role for SMN Protein in Ribosome Biology". *Cell reports* 21.4, pp. 953–965. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2017.10.010.

Biever, Anne, Paul G Donlin-Asp, et al. (2019). "Local translation in neuronal processes". *Current opinion in neurobiology* 57, pp. 141–148. ISSN: 0959-4388, 1873-6882. DOI: 10.1016/j.conb.2019.02.008.

Biever, Anne, Caspar Glock, et al. (2020). "Monosomes actively translate synaptic mRNAs in neuronal processes". *Science* 367.6477. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aay4991.

Bis-Brewer, Dana M et al. (2019). "A network biology approach to unraveling inherited axonopathies". *Scientific reports* 9.1, p. 1692. ISSN: 2045-2322. DOI: 10.1038/s41598-018-37119-z.

Black, Joshua J et al. (2019). "Tsr4 Is a Cytoplasmic Chaperone for the Ribosomal Protein Rps2 in Saccharomyces cerevisiae". *Molecular and cellular biology* 39.17. ISSN: 0270-7306, 1098-5549. DOI: 10.1128/MCB.00094-19.

Blanchet, Sandra and Namit Ranjan (2022). "Translation Phases in Eukaryotes". *Ribosome Biogenesis: Methods and Protocols*. Ed. by Karl-Dieter Entian. New York, NY: Springer US, pp. 217–228. ISBN: 9781071625019. DOI: 10.1007/978-1-0716-2501-9_13.

Bodian, D (1965). "A SUGGESTIVE RELATIONSHIP OF NERVE CELL RNA WITH SPECIFIC SYNAPTIC SITES". *Proceedings of the National Academy of Sciences of the United States of America* 53, pp. 418–425. ISSN: 0027-8424. DOI: 10.1073/pnas.53.2.418.

Bohnsack, Katherine E and Markus T Bohnsack (2019). "Uncovering the assembly pathway of human ribosomes and its emerging links to disease". *The EMBO journal* 38.13, e100278. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.2018100278.

Boivin, Vincent et al. (2018). "Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes". *RNA* 24.7, pp. 950–965. ISSN: 1355-8382, 1469-9001. DOI: 10.1261/rna.064493.117.

Bortolin, M L et al. (1999). "Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs". *The EMBO journal* 18.2, pp. 457–469. ISSN: 0261-4189. DOI: 10.1093/emboj/18.2.457.

Brenner, S et al. (1961). "An unstable intermediate carrying information from genes to ribosomes for protein synthesis". *Nature* 190, pp. 576–581. ISSN: 0028-0836. DOI: 10.1038/190576a0.

Briese, Michael et al. (2016). "Whole transcriptome profiling reveals the RNA content of motor axons". *Nucleic acids research* 44.4, e33. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv1027.

Buckley, Dennis L et al. (2015). "HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins". *ACS chemical biology* 10.8, pp. 1831–1837. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/acschembio.5b00442.

Burgin, K E et al. (1990). "In situ hybridization histochemistry of Ca2+/calmodulin-dependent protein kinase in developing rat brain". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 10.6, pp. 1788–1798. ISSN: 0270-6474.

Burton, P R and J L Paige (1981). "Polarity of axoplasmic microtubules in the olfactory nerve of the frog". *Proceedings of the National Academy of Sciences of the United States of America* 78.5, pp. 3269–3273. ISSN: 0027-8424. DOI: 10.1073/pnas.78.5.3269.

Cajigas, Iván J et al. (2012). "The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging". *Neuron* 74.3, pp. 453–466. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2012.02.036.

Campbell, D S and C E Holt (2001). "Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation". *Neuron* 32.6, pp. 1013–1026. ISSN: 0896-6273. DOI: 10.1016/s0896-6273(01)00551-7.

Chaillou, Thomas (2019). "Ribosome specialization and its potential role in the control of protein translation and skeletal muscle size". *Journal of applied physiology* 127.2, pp. 599–607. ISSN: 0021-8987. DOI: 10.1152/japplphysiol.00946.2018.

Chakrabarty, Nilaj et al. (2019). "Processive flow by biased polymerization mediates the slow axonal transport of actin". *The Journal of cell biology* 218.1, pp. 112–124. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.201711022.

Chau, Kevin F et al. (2018). "Downregulation of ribosome biogenesis during early forebrain development". *eLife* 7. ISSN: 2050-084X. DOI: 10.7554/eLife.36998.

Cheung, Yuen-Nei et al. (2007). "Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo". *Genes & development* 21.10, pp. 1217–1230. ISSN: 0890-9369. DOI: 10.1101/gad.1528307.

Cho, Seung Woo et al. (2014). "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases". *Genome research* 24.1, pp. 132–141. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.162339.113.

Christie, Sean B et al. (2009). "The FXG: a presynaptic fragile X granule expressed in a subset of developing brain circuits". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 29.5, pp. 1514–1524. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3937-08.2009.

Chu, Jen-Fei et al. (2019). "TDP-43 Regulates Coupled Dendritic mRNA Transport-Translation Processes in Co-operation with FMRP and Staufen1". *Cell reports* 29.10, 3118–3133.e6. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2019.10.061.

Cioni, Jean-Michel et al. (2018). "Molecular control of local translation in axon development and maintenance". *Current opinion in neurobiology* 51, pp. 86–94. ISSN: 0959-4388, 1873-6882. DOI: 10.1016/j.conb.2018.02.025.

Colliva, Andrea and Enrico Tongiorgi (2021). "Distinct role of 5'UTR sequences in dendritic trafficking of BDNF mRNA: additional mechanisms for the BDNF splice variants spatial code". *Molecular brain* 14.1, p. 10. ISSN: 1756-6606. DOI: 10.1186/s13041-020-00680-8.

Conant, David et al. (2022). "Inference of CRISPR Edits from Sanger Trace Data". *The CRISPR journal* 5.1, pp. 123–130. ISSN: 2573-1602, 2573-1599. DOI: 10.1089/crispr.2021.0113.

Court, Felipe A, William T J Hendriks, et al. (2008). "Schwann cell to axon transfer of ribosomes: toward a novel understanding of the role of glia in the nervous system". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 28.43, pp. 11024–11029. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2429-08.2008.

Court, Felipe A, Rajiv Midha, et al. (2011). "Morphological evidence for a transport of ribosomes from Schwann cells to regenerating axons". *Glia* 59.10, pp. 1529–1539. ISSN: 0894-1491, 1098-1136. DOI: 10.1002/glia.21196.

Covello, Giuseppina et al. (2014). "An electroporation protocol for efficient DNA transfection in PC12 cells". *Cytotechnology* 66.4, pp. 543–553. ISSN: 0920-9069. DOI: 10.1007/s10616-013-9608-9.

Coyne, Alyssa N et al. (2014). "Futsch/MAP1B mRNA is a translational target of TDP-43 and is neuroprotective in a Drosophila model of amyotrophic lateral sclerosis". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 34.48, pp. 15962–15974. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2526-14.2014.

Crick, F H (1958). "On protein synthesis". *Symposia of the Society for Experimental Biology* 12, pp. 138–163. ISSN: 0081-1386.

Cunningham, Fiona et al. (2022). "Ensembl 2022". *Nucleic acids research* 50.D1, pp. D988–D995. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab1049.

Czekay, Dominic P and Ute Kothe (2021). "H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes". *Frontiers in microbiology* 12, p. 654370. ISSN: 1664-302X. DOI: 10.3389/fmicb.2021.654370.

Das, Sulagna et al. (2021). "Intracellular mRNA transport and localized translation". *Nature reviews. Molecular cell biology* 22.7, pp. 483–504. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/s41580-021-00356-8.

Dastidar, Sudhriti Ghosh and Deepak Nair (2022). "A Ribosomal Perspective on Neuronal Local Protein Synthesis". *Frontiers in molecular neuroscience* 15, p. 823135. ISSN: 1662-5099. DOI: 10.3389/fnmol.2022.823135.

De Graeve, Fabienne and Florence Besse (2018). "Neuronal RNP granules: from physiological to pathological assemblies". *Biological chemistry* 399.7, pp. 623–635. ISSN: 1431-6730, 1437-4315. DOI: 10.1515/hsz-2018-0141.

De Vos, Kurt J and Majid Hafezparast (2017). "Neurobiology of axonal transport defects in motor neuron diseases: Opportunities for translational research?" *Neurobiology of disease* 105, pp. 283–299. ISSN: 0969-9961, 1095-953X. DOI: 10.1016/j.nbd.2017.02.004.

Dieci, Giorgio et al. (2009). "Eukaryotic snoRNAs: a paradigm for gene expression flexibility". *Genomics* 94.2, pp. 83–88. ISSN: 0888-7543, 1089-8646. DOI: 10.1016/j.ygeno.2009.05.002.

Dinman, Jonathan D (2016). "Pathways to Specialized Ribosomes: The Brussels Lecture". *Journal of molecular biology* 428.10 Pt B, pp. 2186–2194. ISSN: 0022-2836, 1089-8638. DOI: 10.1016/j.jmb.2015.12.021.

Doench, John G et al. (2016). "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9". *Nature biotechnology* 34.2, pp. 184–191. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.3437.

Dörner, Kerstin et al. (2023). "Ribosome biogenesis factors-from names to functions". *The EMBO journal* 42.7, e112699. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.2022112699.

Edfors, Fredrik et al. (2016). "Gene-specific correlation of RNA and protein levels in human cells and tissues". *Molecular systems biology* 12.10, p. 883. ISSN: 1744-4292. DOI: 10.15252/msb.20167144.

El Fatimy, Rachid et al. (2016). "Tracking the Fragile X Mental Retardation Protein in a Highly Ordered Neuronal RiboNucleoParticles Population: A Link between Stalled Polyribosomes and RNA Granules". *PLoS genetics* 12.7, e1006192. ISSN: 1553-7390, 1553-7404. DOI: 10.1371/journal.pgen.1006192.

Elvira, George et al. (2006). "Characterization of an RNA granule from developing brain". *Molecular & cellular proteomics: MCP* 5.4, pp. 635–651. ISSN: 1535-9476. DOI: 10.1074/mcp.M500255-MCP200.

Eom, Taesun et al. (2003). "Localization of a beta-actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses". *The Journal of neuroscience: the official journal of*

the Society for Neuroscience 23.32, pp. 10433–10444. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.23-32-10433.2003.

Espejo, Romilio T and Nicolás Plaza (2018). "Multiple Ribosomal RNA Operons in Bacteria; Their Concerted Evolution and Potential Consequences on the Rate of Evolution of Their 16S rRNA". *Frontiers in microbiology* 9, p. 1232. ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.01232.

Fafard-Couture, Étienne et al. (2021). "Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships". *Genome biology* 22.1, p. 172. ISSN: 1465-6906. DOI: 10.1186/s13059-021-02391-2.

Fallini, Claudia et al. (2016). "Deficiency of the Survival of Motor Neuron Protein Impairs mRNA Localization and Local Translation in the Growth Cone of Motor Neurons". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 36.13, pp. 3811–3820. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2396-15.2016.

Fang, Cheng et al. (2012). "Oxidative stress inhibits axonal transport: implications for neurodegenerative diseases". *Molecular neurodegeneration* 7, p. 29. ISSN: 1750-1326. DOI: 10.1186/1750-1326-7-29.

Farias, Joaquina et al. (2020). "Axon microdissection and transcriptome profiling reveals the in vivo RNA content of fully differentiated myelinated motor axons". *RNA* 26.5, pp. 595–612. ISSN: 1355-8382, 1469-9001. DOI: 10.1261/rna.073700.119.

Fernandopulle, Michael S et al. (2021). "RNA transport and local translation in neurodevelopmental and neurodegenerative disease". *Nature neuroscience* 24.5, pp. 622–632. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/s41593-020-00785-2.

Ferretti, Max B and Katrin Karbstein (2019). "Does functional specialization of ribosomes really exist?" *RNA* 25.5, pp. 521–538. ISSN: 1355-8382, 1469-9001. DOI: 10.1261/rna.069823.118.

Fish, Kenneth N (2022). "Total Internal Reflection Fluorescence (TIRF) Microscopy". *Current protocols* 2.8, e517. ISSN: 2691-1299. DOI: 10.1002/cpz1.517.

Fuchs, Gabriele et al. (2015). "Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site". *Proceedings of the National Academy of Sciences of the United States of America* 112.2, pp. 319–325. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1421328111.

Fukuda, Yusuke et al. (2021). "Binding and transport of SFPQ-RNA granules by KIF5A/KLC1 motors promotes axon survival". *The Journal of cell biology* 220.1. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.202005051.

Fusco, Claudia M et al. (2021). "Neuronal ribosomes exhibit dynamic and context-dependent exchange of ribosomal proteins". *Nature communications* 12.1, p. 6127. ISSN: 2041-1723. DOI: 10.1038/s41467-021-26365-x.

Gaffield, Michael A and William J Betz (2006). "Imaging synaptic vesicle exocytosis and endocytosis with FM dyes". *Nature protocols* 1.6, pp. 2916–2921. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/nprot.2006.476.

Gainer, H et al. (1977). "Evidence for the glia-neuron protein transfer hypothesis from intracellular perfusion studies of squid giant axons". *The Journal of cell biology* 74.2, pp. 524–530. ISSN: 0021-9525. DOI: 10.1083/jcb.74.2.524.

Galaxy Community (2022). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update". *Nucleic acids research* 50.W1, W345–51. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac247.

Gamarra, Maria et al. (2021). "Local Translation in Nervous System Pathologies". *Frontiers in integrative neuroscience* 15, p. 689208. ISSN: 1662-5145. DOI: 10.3389/fnint.2021.689208.

Gay, David M et al. (2022). "Translational control through ribosome heterogeneity and functional specialization". *Trends in biochemical sciences* 47.1, pp. 66–81. ISSN: 0968-0004. DOI: 10.1016/j.tibs.2021.07.001.

Genuth, Naomi R and Maria Barna (2018). "The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life". *Molecular cell* 71.3, pp. 364–374. ISSN: 1097-2765, 1097-4164. DOI: 10.1016/j.molcel.2018.07.018.

Giuditta, A, A Cupello, et al. (1980). "Ribosomal RNA in the axoplasm of the squid giant axon". *Journal of neurochemistry* 34.6, pp. 1757–1760. ISSN: 0022-3042. DOI: 10.1111/j.1471-4159.1980.tb11271.x.

Giuditta, A, W D Dettbarn, et al. (1968). "Protein synthesis in the isolated giant axon of the squid". *Proceedings of the National Academy of Sciences of the United States of America* 59.4, pp. 1284–1287. ISSN: 0027-8424. DOI: 10.1073/pnas.59.4.1284.

Giuditta, Antonio and Luigia Santella (1986). "MESSENGER RNA IN SQUID AXOPLASM". *Neurochemistry international* 8.3, pp. 435–442. ISSN: 0197-0186.

Gregory, Brian et al. (2019). "The small and large ribosomal subunits depend on each other for stability and accumulation". *Life science alliance* 2.2. ISSN: 2575-1077. DOI: 10.26508/lsa.201800150.

Grimm, Jonathan B et al. (2017). "Synthesis of Janelia Fluor HaloTag and SNAP-Tag Ligands and Their Use in Cellular Imaging Experiments". *Super-Resolution Microscopy: Methods and Protocols*. Ed. by Holger Erfle. New York, NY: Springer New York, pp. 179–188. ISBN: 9781493972654. DOI: 10.1007/978-1-4939-7265-4_15.

Grummt, Ingrid and Craig S Pikaard (2003). "Epigenetic silencing of RNA polymerase I transcription". *Nature reviews. Molecular cell biology* 4.8, pp. 641–649. ISSN: 1471-0072. DOI: 10.1038/nrm1171.

GTEx Consortium (2020). "The GTEx Consortium atlas of genetic regulatory effects across human tissues". *Science* 369.6509, pp. 1318–1330. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaz1776.

— (2021). *Laboratory Methods*. https://gtexportal.org/home/methods. Accessed: 2023-4-6.

Guedes-Dias, Pedro and Erika L F Holzbaur (2019). "Axonal transport: Driving synaptic function". *Science* 366.6462. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaw9997.

Guillaud, Laurent et al. (2020). "Anterograde Axonal Transport in Neuronal Homeostasis and Disease". *Frontiers in molecular neuroscience* 13, p. 556175. ISSN: 1662-5099. DOI: 10.3389/fnmol.2020.556175.

Guimaraes, Joao C and Mihaela Zavolan (2016). "Patterns of ribosomal protein expression specify normal and malignant human cells". *Genome biology* 17.1, p. 236. ISSN: 1465-6906. DOI: 10.1186/s13059-016-1104-z.

Gunderson, J H et al. (1987). "Structurally distinct, stage-specific ribosomes occur in Plasmodium". *Science* 238.4829, pp. 933–937. ISSN: 0036-8075. DOI: 10.1126/science.3672135.

Guo, Huili (2018). "Specialized ribosomes and the control of translation". *Biochemical Society transactions* 46.4, pp. 855–869. ISSN: 0300-5127, 1470-8752. DOI: 10.1042/BST20160426.

Hafner, Anne-Sophie et al. (2019). "Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments". *Science* 364.6441. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau3644.

Hanz, Shlomit et al. (2003). "Axoplasmic importins enable retrograde injury signaling in lesioned nerve". *Neuron* 40.6, pp. 1095–1104. ISSN: 0896-6273. DOI: 10.1016/s0896-6273(03)00770-0.

Hao le, Thi et al. (2017). "HuD and the Survival Motor Neuron Protein Interact in Motoneurons and Are Essential for Motoneuron Development, Function, and mRNA Regulation". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 37.48, pp. 11559–11571. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1528-17.2017.

Heiman, Myriam et al. (2008). "A translational profiling approach for the molecular characterization of CNS cell types". *Cell* 135.4, pp. 738–748. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2008.10.028.

Henderson, A S et al. (1972). "Location of ribosomal DNA in the human chromosome complement". *Proceedings of the National Academy of Sciences of the United*

States of America 69.11, pp. 3394–3398. ISSN: 0027-8424. DOI: 10.1073/pnas.69.11.3394.

Hengst, Ulrich et al. (2009). "Axonal elongation triggered by stimulus-induced local translation of a polarity complex protein". *Nature cell biology* 11.8, pp. 1024–1030. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb1916.

Hernández, A Iván et al. (2009). "Poly-(ADP-ribose) polymerase-1 is necessary for long-term facilitation in Aplysia". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 29.30, pp. 9553–9562. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1512-09.2009.

Hirokawa, N et al. (1991). "Kinesin associates with anterogradely transported membranous organelles in vivo". *The Journal of cell biology* 114.2, pp. 295–302. ISSN: 0021-9525. DOI: 10.1083/jcb.114.2.295.

Hirokawa, Nobutaka et al. (2009). "Kinesin superfamily motor proteins and intracellular transport". *Nature reviews. Molecular cell biology* 10.10, pp. 682–696. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm2774.

Holt et al. (2019). "Local translation in neurons: visualization and function". *Nature structural & molecular biology* 26.7, pp. 557–566. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/s41594-019-0263-5.

Hsiau, Tim et al. (2018). "Inference of CRISPR Edits from Sanger Trace Data". *bioRxiv*, p. 251082. DOI: 10.1101/251082.

Hsu, Patrick D et al. (2013). "DNA targeting specificity of RNA-guided Cas9 nucleases". *Nature biotechnology* 31.9, pp. 827–832. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.2647.

Hu, Rendong et al. (2018). "A novel method of neural differentiation of PC12 cells by using Opti-MEM as a basic induction medium". *International journal of molecular medicine* 41.1, pp. 195–201. ISSN: 1107-3756, 1791-244X. DOI: 10.3892/ijmm.2017.3195.

Hughes, Peyton et al. (2007). "The costs of using unauthenticated, over-passaged cell lines: how much more data do we need?" *BioTechniques* 43.5, 575, 577–8, 581–2 passim. ISSN: 0736-6205. DOI: 10.2144/000112598.

Illumina, Inc (2014). *TruSeq™ RNA and DNA Library Preparation Kits v2*. https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_truseq_sample_prep_kits.pdf.

Jacob, F and J Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins". *Journal of molecular biology* 3, pp. 318–356. ISSN: 0022-2836. DOI: 10.1016/s0022-2836(61)80072-7.

Jády, B E and T Kiss (2001). "A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA". *The EMBO journal* 20.3, pp. 541–551. ISSN: 0261-4189. DOI: 10.1093/emboj/20.3.541.

Jang, Chang-Young et al. (2004). "RpS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis". *FEBS letters* 560.1-3, pp. 81–85. ISSN: 0014-5793. DOI: 10.1016/S0014-5793(04)00074-2.

Jennes, L (2017). "Chapter 1 - Cytology of the Central Nervous System". *Conn's Translational Neuroscience*. Ed. by P Michael Conn. San Diego: Academic Press, pp. 1–10. ISBN: 9780128023815. DOI: 10.1016/B978-0-12-802381-5.00001-4.

Jeon, Chan-Young et al. (2010). "Neurites from PC12 cells are connected to each other by synapse-like structures". *Synapse* 64.10, pp. 765–772. ISSN: 0887-4476, 1098-2396. DOI: 10.1002/syn.20789.

Jiao, Lijuan et al. (2023). "Ribosome biogenesis in disease: new players and therapeutic targets". *Signal transduction and targeted therapy* 8.1, p. 15. ISSN: 2059-3635. DOI: 10.1038/s41392-022-01285-4.

Joo, Minju et al. (2022). "Specialised ribosomes as versatile regulators of gene expression". *RNA biology* 19.1, pp. 1103–1114. ISSN: 1547-6286, 1555-8584. DOI: 10.1080/15476286.2022.2135299.

Jorjani, Hadi et al. (2016). "An updated human snoRNAome". *Nucleic acids research* 44.11, pp. 5068–5082. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkw386.

Kanai, Yoshimitsu et al. (2004). "Kinesin transports RNA: isolation and characterization of an RNA-transporting granule". *Neuron* 43.4, pp. 513–525. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2004.07.022.

Kang, Jian et al. (2021). "Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy". *Signal transduction and targeted therapy* 6.1, p. 323. ISSN: 2059-3635. DOI: 10.1038/s41392-021-00728-8.

Kapitein, Lukas C and Casper C Hoogenraad (2015). "Building the Neuronal Microtubule Cytoskeleton". *Neuron* 87.3, pp. 492–506. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2015.05.046.

Kaximierz, Susan Kass: et al. (1990). "The U3 Small Nucleolar Ribonucleuprotein Functions in the First Step of Preribosomal RNA Processing". *Cell* 60, p. 697408. ISSN: 0092-8674.

Kearse, Michael G et al. (2011). "Expression of ribosomal protein L22e family members in Drosophila melanogaster: rpL22-like is differentially expressed and alternatively spliced". *Nucleic acids research* 39.7, pp. 2701–2716. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkq1218.

Keppler, Antje et al. (2003). "A general method for the covalent labeling of fusion proteins with small molecules in vivo". *Nature biotechnology* 21.1, pp. 86–89. ISSN: 1087-0156. DOI: 10.1038/nbt765.

Khalil, Bilal et al. (2018). "mRNP assembly, axonal transport, and local translation in neurodegenerative diseases". *Brain research* 1693.Pt A, pp. 75–91. ISSN: 0006-8993, 1872-6240. DOI: 10.1016/j.brainres.2018.02.018.

Khan, Zia et al. (2013). "Primate transcript and protein expression levels evolve under compensatory selection pressures". *Science* 342.6162, pp. 1100–1104. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1242379.

Khatter, Heena, Alexander G Myasnikov, Leslie Mastio, et al. (2014). "Purification, characterization and crystallization of the human 80S ribosome". *Nucleic acids research* 42.6, e49. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkt1404.

Khatter, Heena, Alexander G Myasnikov, S Kundhavai Natchiar, et al. (2015). "Structure of the human 80S ribosome". *Nature* 520.7549, pp. 640–645. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14427.

Khoshnevis, Sohail et al. (2022). "Ribosomal RNA 2'-O-methylations regulate translation by impacting ribosome dynamics". *Proceedings of the National Academy of Sciences of the United States of America* 119.12, e2117334119. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2117334119.

Kiebler, M and G Bassell (2006). "Neuronal RNA granules: movers and makers". *Neuron* 51.6, pp. 685–690. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.08.021.

Kipper, Kalle et al. (2022). "Neuronal RNA granules are ribosome complexes stalled at the pre-translocation state". *Journal of molecular biology* 434.20, p. 167801. ISSN: 0022-2836, 1089-8638. DOI: 10.1016/j.jmb.2022.167801.

Kleiman, R et al. (1990). "Differential subcellular localization of particular mRNAs in hippocampal neurons in culture". *Neuron* 5.6, pp. 821–830. ISSN: 0896-6273. DOI: 10.1016/0896-6273(90)90341-c.

Knowles, R B et al. (1996). "Translocation of RNA granules in living neurons". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 16.24, pp. 7812–7820. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.16-24-07812.1996.

Koch, Barbara et al. (2012). "Yar1 protects the ribosomal protein Rps3 from aggregation". *The Journal of biological chemistry* 287.26, pp. 21806–21815. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M112.365791.

Koenig, E (1965). "SYNTHETIC MECHANISMS IN THE AXON. I. LOCAL AXONAL SYNTHESIS OF ACETYLCHOLINESTERASE". *Journal of neurochemistry* 12, pp. 343–355. ISSN: 0022-3042. DOI: 10.1111/j.1471-4159.1965.tb04235.x.

Koh, Dora C et al. (2013). "Physical evidence supporting a ribosomal shunting mechanism of translation initiation for BACE1 mRNA". *Translation (Austin, Tex.)* 1.1, e24400. ISSN: 2169-074X, 2169-0731. DOI: 10.4161/trla.24400.

Koley, Sandip et al. (2019). "Translating regeneration: Local protein synthesis in the neuronal injury response". *Neuroscience research* 139, pp. 26–36. ISSN: 0168-0102, 1872-8111. DOI: 10.1016/j.neures.2018.10.003.

Kornprobst, Markus et al. (2016). "Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome". *Cell* 166.2, pp. 380–393. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.06.014.

Krichevsky, A M and K S Kosik (2001). "Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation". *Neuron* 32.4, pp. 683–696. ISSN: 0896-6273. DOI: 10.1016/s0896-6273(01)00508-6.

Krokowski, Dawid et al. (2011). "Characterization of hibernating ribosomes in mammalian cells". *Cell cycle* 10.16, pp. 2691–2702. ISSN: 1538-4101, 1551-4005. DOI: 10.4161/cc.10.16.16844.

Kufel, Joanna and Pawel Grzechnik (2019). "Small Nucleolar RNAs Tell a Different Tale". *Trends in genetics: TIG* 35.2, pp. 104–117. ISSN: 0168-9525. DOI: 10.1016/j.tig.2018.11.005.

Lambert, Talley J (2019). "FPbase: a community-editable fluorescent protein database". *Nature methods* 16.4, pp. 277–278. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-019-0352-8.

Landry, Dori M et al. (2009). "RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs". *Genes & development* 23.23, pp. 2753–2764. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.1832209.

Lasek, R J, C Dabrowski, et al. (1973). "Analysis of axoplasmic RNA from invertebrate giant axons". *Nature: New biology* 244.136, pp. 162–165. ISSN: 0090-0028. DOI: 10.1038/newbio244162a0.

Lasek, R J, H Gainer, et al. (1977). "Cell-to-cell transfer of glial proteins to the squid giant axon. The glia-neuron protein trnasfer hypothesis". *The Journal of cell biology* 74.2, pp. 501–523. ISSN: 0021-9525. DOI: 10.1083/jcb.74.2.501.

Lau, Benjamin et al. (2023). "Mechanism of 5S RNP recruitment and helicase-surveilled rRNA maturation during pre-60S biogenesis". *EMBO reports*, e56910. ISSN: 1469-221X, 1469-3178. DOI: 10.15252/embr.202356910.

Leterrier, Christophe (2018). "The Axon Initial Segment: An Updated Viewpoint". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 38.9, pp. 2135–2145. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1922-17.2018.

Leung, Kin-Mei et al. (2006). "Asymmetrical beta-actin mRNA translation in growth cones mediates attractive turning to netrin-1". *Nature neuroscience* 9.10, pp. 1247–1256. ISSN: 1097-6256. DOI: 10.1038/nn1775.

Levin, D H et al. (1973). "Protein initiation in eukaryotes: formation and function of a ternary complex composed of a partially purified ribosomal factor, methionyl transfer RNA, and guanosine triphosphate". *Proceedings of the National Academy of Sciences of the United States of America* 70.1, pp. 41–45. ISSN: 0027-8424. DOI: 10.1073/pnas.70.1.41.

Lezana, Juan Pablo et al. (2016). "Axonal PPARγ promotes neuronal regeneration after injury". *Developmental neurobiology* 76.6, pp. 688–701. ISSN: 1932-8451, 1932-846X. DOI: 10.1002/dneu.22353.

Li, Chuanzhou and Jürgen Götz (2017). "Somatodendritic accumulation of Tau in Alzheimer's disease is promoted by Fyn-mediated local protein translation". *The EMBO journal* 36.21, pp. 3120–3138. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201797724.

Li, Dingfeng et al. (2018). "Activity dependent LoNA regulates translation by coordinating rRNA transcription and methylation". *Nature communications* 9.1, p. 1726. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04072-4.

Li, Lichao et al. (2021). "Axonal mRNA localization and translation: local events with broad roles". *Cellular and molecular life sciences: CMLS* 78.23, pp. 7379–7395. ISSN: 1420-682X, 1420-9071. DOI: 10.1007/s00018-021-03995-4.

Liang, Junnan et al. (2019). "Small Nucleolar RNAs: Insight Into Their Function in Cancer". *Frontiers in oncology* 9, p. 587. ISSN: 2234-943X. DOI: 10.3389/fonc.2019.00587.

Liao, Ya-Cheng et al. (2019). "RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether". *Cell* 179.1, 147–164.e20. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2019.08.050.

Lipka, Joanna et al. (2016). "Microtubule-binding protein doublecortin-like kinase 1 (DCLK1) guides kinesin-3-mediated cargo transport to dendrites". *The EMBO journal* 35.3, pp. 302–318. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201592929.

Liss, Viktoria et al. (2015). "Self-labelling enzymes as universal tags for fluorescence microscopy, super-resolution microscopy and electron microscopy". *Scientific reports* 5, p. 17740. ISSN: 2045-2322. DOI: 10.1038/srep17740.

Llácer, Jose L et al. (2015). "Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex". *Molecular cell* 59.3, pp. 399–412. ISSN: 1097-2765, 1097-4164. DOI: 10.1016/j.molcel.2015.06.033.

Los, Georgyi V et al. (2008). "HaloTag: a novel protein labeling technology for cell imaging and protein analysis". *ACS chemical biology* 3.6, pp. 373–382. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/cb800025k.

Lyles, Vlasta et al. (2006). "Synapse formation and mRNA localization in cultured Aplysia neurons". *Neuron* 49.3, pp. 349–356. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2005.12.029.

Maday, Sandra et al. (2014). "Axonal transport: cargo-specific mechanisms of motility and regulation". *Neuron* 84.2, pp. 292–309. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2014.10.019.

Mageeney, Catherine M and Vassie C Ware (2019). "Specialized eRpL22 paralogue-specific ribosomes regulate specific mRNA translation in spermatogenesis in Drosophila melanogaster". *Molecular biology of the cell* 30.17, pp. 2240–2253. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E19-02-0086.

Malinow, Roberto et al. (2010). "Introduction of green fluorescent protein (GFP) into hippocampal neurons through viral infection". *Cold Spring Harbor protocols* 2010.4, db.prot5406. ISSN: 1940-3402, 1559-6095. DOI: 10.1101/pdb.prot5406.

Mallik, Roop et al. (2004). "Cytoplasmic dynein functions as a gear in response to load". *Nature* 427.6975, pp. 649–652. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature02293.

Martin, K C et al. (1997). "Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage". *Cell* 91.7, pp. 927–938. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(00)80484-5.

Mattheyses, Alexa L and Daniel Axelrod (2006). "Direct measurement of the evanescent field profile produced by objective-based total internal reflection fluorescence". *Journal of biomedical optics* 11.1, p. 014006. ISSN: 1083-3668. DOI: 10.1117/1.2161018.

Mauro, Vincent P and Gerald M Edelman (2002). "The ribosome filter hypothesis". *Proceedings of the National Academy of Sciences of the United States of America* 99.19, pp. 12031–12036. ISSN: 0027-8424. DOI: 10.1073/pnas.192442499.

Maxwell, E S and M J Fournier (1995). "The small nucleolar RNAs". *Annual review of biochemistry* 64, pp. 897–934. ISSN: 0066-4154. DOI: 10.1146/annurev.bi.64.070195.004341.

McMahon, M et al. (2019). "A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS". *eLife* 8. ISSN: 2050-084X. DOI: 10.7554/eLife.48847.

Mendonsa, Samantha et al. (2023). "Massively parallel identification of mRNA localization elements in primary cortical neurons". *Nature neuroscience* 26.3, pp. 394–405. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/s41593-022-01243-x.

Merrick, William C and Graham D Pavitt (2018). "Protein Synthesis Initiation in Eukaryotic Cells". *Cold Spring Harbor perspectives in biology* 10.12. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a033092.

Milkereit, Philipp et al. (2003). "A Noc Complex Specifically Involved in the Formation and Nuclear Export of Ribosomal 40 S Subunits*". *The Journal of biological chemistry* 278.6, pp. 4072–4081. ISSN: 0021-9258. DOI: 10.1074/jbc.M208898200.

Miller, Sarah C et al. (2023). "Specialized Ribosomes in Health and Disease". *International journal of molecular sciences* 24.7. ISSN: 1422-0067. DOI: 10.3390/ijms24076334.

Miyashiro, K et al. (1994). "On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning". *Proceedings of the National Academy of Sciences of the United States of America* 91.23, pp. 10800–10804. ISSN: 0027-8424. DOI: 10.1073/pnas.91.23.10800.

Moazed, D and H F Noller (1989). "Intermediate states in the movement of transfer RNA in the ribosome". *Nature* 342.6246, pp. 142–148. ISSN: 0028-0836. DOI: 10.1038/342142a0.

Moraleva, A A et al. (2022). "Eukaryotic Ribosome Biogenesis: The 60S Subunit". *Acta naturae* 14.2, pp. 39–49. ISSN: 2075-8251. DOI: 10.32607/actanaturae.11541.

Müller, Kerstin et al. (2018). "A predominantly glial origin of axonal ribosomes after nerve injury". *Glia* 66.8, pp. 1591–1610. ISSN: 0894-1491, 1098-1136. DOI: 10.1002/glia.23327.

Nagano, Seiichi and Toshiyuki Araki (2021a). *Axonal Transport and Local Translation of mRNA in Amyotrophic Lateral Sclerosis*. Exon Publications. DOI: 10.36255/exonpublications.amyotrophiclateralsclerosis.axonaltransport.2021.

— (2021b). “Axonal Transport and Local Translation of mRNA in Neurodegenerative Diseases”. *Frontiers in molecular neuroscience* 14, p. 697973. ISSN: 1662-5099. DOI: 10.3389/fnmol.2021.697973.

Natchiar, S Kundhavai et al. (2017). “Visualization of chemical modifications in the human 80S ribosome structure”. *Nature* 551.7681, pp. 472–477. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature24482.

Nieuwenhuis, Tim O et al. (2020). “Consistent RNA sequencing contamination in GTEx and other data sets”. *Nature communications* 11.1, p. 1933. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15821-9.

Norris, Karl et al. (2021). “Ribosome heterogeneity and specialization in development”. *Wiley interdisciplinary reviews. RNA* 12.4, e1644. ISSN: 1757-7004, 1757-7012. DOI: 10.1002/wrna.1644.

Nousch, Marco (2020). “RPL-4 and RPL-9 Mediated Ribosome Purifications Facilitate the Efficient Analysis of Gene Expression in Caenorhabditis elegans Germ Cells”. *G3* 10.11, pp. 4063–4069. ISSN: 2160-1836. DOI: 10.1534/g3.120.401644.

Null, Null et al. (2015). “The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans”. *Science* 348.6235, pp. 648–660. ISSN: 0036-8075. DOI: 10.1126/science.1262110. eprint: https://www.science.org/doi/pdf/10.1126/science.1262110.

Ojha, Sandeep et al. (2020). “snoRNPs: Functions in Ribosome Biogenesis”. *Biomolecules* 10.5. ISSN: 2218-273X. DOI: 10.3390/biom10050783.

Okada, Y et al. (1995). “The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors”. *Cell* 81.5, pp. 769–780. ISSN: 0092-8674. DOI: 10.1016/0092-8674(95)90538-3.

Palade, G E (1955). “A small particulate component of the cytoplasm”. *The Journal of biophysical and biochemical cytology* 1.1, pp. 59–68. ISSN: 0095-9901. DOI: 10.1083/jcb.1.1.59.

Palazzo, Alexander F and Eliza S Lee (2015). "Non-coding RNA: what is functional and what is junk?" *Frontiers in genetics* 6, p. 2. ISSN: 1664-8021. DOI: 10.3389/fgene.2015.00002.

Parker, Matthew T et al. (2020). "Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification". *eLife* 9. ISSN: 2050-084X. DOI: 10.7554/eLife.49658.

Parks, Matthew M et al. (2018). "Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression". *Science advances* 4.2, eaao0665. ISSN: 2375-2548. DOI: 10.1126/sciadv.aao0665.

Perry, Rotem Ben-Tov et al. (2012). "Subcellular knockout of importin β1 perturbs axonal retrograde signaling". *Neuron* 75.2, pp. 294–305. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2012.05.033.

Pestova, T V et al. (2000). "The joining of ribosomal subunits in eukaryotes requires eIF5B". *Nature* 403.6767, pp. 332–335. ISSN: 0028-0836. DOI: 10.1038/35002118.

Pettersen, Eric F et al. (2021). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers". *Protein science: a publication of the Protein Society* 30.1, pp. 70–82. ISSN: 0961-8368, 1469-896X. DOI: 10.1002/pro.3943.

Piekna-Przybylska, Dorota et al. (2008). "The 3D rRNA modification maps database: with interactive tools for ribosome analysis". *Nucleic acids research* 36.Database issue, pp. D178–83. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkm855.

Poon, Michael M et al. (2006). "Identification of process-localized mRNAs from cultured rodent hippocampal neurons". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 26.51, pp. 13390–13399. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3432-06.2006.

Powell, David (2019). *Degust: interactive RNA-seq analysis*. DOI: 10.5281/zenodo.3501067.

Prokop, Andreas (2020). "Cytoskeletal organization of axons in vertebrates and invertebrates". *The Journal of cell biology* 219.7. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.201912081.

Ramagopal, S (1990). "Induction of cell-specific ribosomal proteins in aggregation-competent nonmorphogenetic Dictyostelium discoideum". *Biochemistry and cell biology = Biochimie et biologie cellulaire* 68.11, pp. 1281–1287. ISSN: 0829-8211. DOI: 10.1139/o90-190.

Ran, F Ann et al. (2013). "Genome engineering using the CRISPR-Cas9 system". *Nature protocols* 8.11, pp. 2281–2308. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/nprot.2013.143.

Rangaraju, Vidhya et al. (2017). "Local translation in neuronal compartments: how local is local?" *EMBO reports* 18.5, pp. 693–711. ISSN: 1469-221X, 1469-3178. DOI: 10.15252/embr.201744045.

Raveh, Alon et al. (2016). "A model for competition for ribosomes in the cell". *Journal of the Royal Society, Interface / the Royal Society* 13.116. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2015.1062.

Ray, K et al. (1999). "Kinesin-II is required for axonal transport of choline acetyltransferase in Drosophila". *The Journal of cell biology* 147.3, pp. 507–518. ISSN: 0021-9525. DOI: 10.1083/jcb.147.3.507.

Reck-Peterson, Samara L et al. (2018). "The cytoplasmic dynein transport machinery and its many cargoes". *Nature reviews. Molecular cell biology* 19.6, pp. 382–398. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/s41580-018-0004-3.

Roberts, Richard B (1958). *Microsomal particles and protein synthesis; papers presented at the First Symposium of the Biophysical Society, at the Massachusetts Institute of Technology, Cambridge, February 5, 6, and 8, 1958*. New York, Published on behalf of the Washington Academy of Sciences, Washington, D.C., by Pergamon Press, 1958, p. 188.

Rossoll, Wilfried et al. (2003). "Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth

cones of motoneurons". *The Journal of cell biology* 163.4, pp. 801–812. ISSN: 0021-9525. DOI: 10.1083/jcb.200304128.

Roy, Subhojit (2014). "Seeing the unseen: the hidden world of slow axonal transport". *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry* 20.1, pp. 71–81. ISSN: 1073-8584, 1089-4098. DOI: 10.1177/1073858413498306.

Rugjee, Kushal Nivriti et al. (2013). "Fluorescent protein tagging confirms the presence of ribosomal proteins at Drosophila polytene chromosomes". *PeerJ* 1, e15. ISSN: 2167-8359. DOI: 10.7717/peerj.15.

Sakamoto, K M et al. (2001). "Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation". *Proceedings of the National Academy of Sciences of the United States of America* 98.15, pp. 8554–8559. ISSN: 0027-8424. DOI: 10.1073/pnas.141230798.

Salogiannis, John and Samara L Reck-Peterson (2017). "Hitchhiking: A Non-Canonical Mode of Microtubule-Based Transport". *Trends in cell biology* 27.2, pp. 141–150. ISSN: 0962-8924, 1879-3088. DOI: 10.1016/j.tcb.2016.09.005.

Samir, Parimal et al. (2018). "Identification of Changing Ribosome Protein Compositions using Mass Spectrometry". *Proteomics* 18.20, e1800217. ISSN: 1615-9853, 1615-9861. DOI: 10.1002/pmic.201800217.

Sanz, Elisenda, Jonathan C Bean, et al. (2019). "RiboTag: Ribosomal Tagging Strategy to Analyze Cell-Type-Specific mRNA Expression In Vivo". *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]* 88.1, e77. ISSN: 1934-8584, 1934-8576. DOI: 10.1002/cpns.77.

Sanz, Elisenda, Linghai Yang, et al. (2009). "Cell-type-specific isolation of ribosome-associated mRNA from complex tissues". *Proceedings of the National Academy of Sciences of the United States of America* 106.33, pp. 13939–13944. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0907143106.

Schavemaker, Paul E et al. (2018). "How Important Is Protein Diffusion in Prokaryotes?" *Frontiers in molecular biosciences* 5, p. 93. ISSN: 2296-889X. DOI: 10.3389/fmolb.2018.00093.

Schindelin, Johannes et al. (2012). "Fiji: an open-source platform for biological-image analysis". *Nature methods* 9.7, pp. 676–682. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2019.

Schlager, Max A et al. (2014). "In vitro reconstitution of a highly processive recombinant human dynein complex". *The EMBO journal* 33.17, pp. 1855–1868. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201488792.

Schluenzen, F et al. (2000). "Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution". *Cell* 102.5, pp. 615–623. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(00)00084-2.

Scholey, Jonathan M (2013). "Kinesin-2: a family of heterotrimeric and homodimeric motors with diverse intracellular transport functions". *Annual review of cell and developmental biology* 29, pp. 443–469. ISSN: 1081-0706, 1530-8995. DOI: 10.1146/annurev-cellbio-101512-122335.

Scott, David A et al. (2011). "Mechanistic logic underlying the axonal transport of cytosolic proteins". *Neuron* 70.3, pp. 441–454. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2011.03.022.

Segev, Nadav and Jeffrey E Gerst (2018). "Specialized ribosomes and specific ribosomal protein paralogs control translation of mitochondrial proteins". *The Journal of cell biology* 217.1, pp. 117–126. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.201706059.

Sentmanat, Monica F et al. (2018). "A Survey of Validation Strategies for CRISPR-Cas9 Editing". *Scientific reports* 8.1, p. 888. ISSN: 2045-2322. DOI: 10.1038/s41598-018-19441-8.

Shakhbazau, Antos et al. (2016). "Demyelination induces transport of ribosome-containing vesicles from glia to axons: evidence from animal models and MS

patient brains". *Molecular biology reports* 43.6, pp. 495–507. ISSN: 0301-4851, 1573-4978. DOI: 10.1007/s11033-016-3990-2.

Shi, Huajuan et al. (2021). "Bias in RNA-seq Library Preparation: Current Challenges and Solutions". *BioMed research international* 2021, p. 6647597. ISSN: 2314-6133, 2314-6141. DOI: 10.1155/2021/6647597.

Shi, Zhen et al. (2017). "Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide". *Molecular cell* 67.1, 71–83.e7. ISSN: 1097-2765, 1097-4164. DOI: 10.1016/j.molcel.2017.05.021.

Shigeoka, Toshiaki, Hosung Jung, et al. (2016). "Dynamic Axonal Translation in Developing and Mature Visual Circuits". *Cell* 166.1, pp. 181–192. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.05.029.

Shigeoka, Toshiaki, Jane Jung, et al. (2018). "Axon-TRAP-RiboTag: Affinity Purification of Translated mRNAs from Neuronal Axons in Mouse In Vivo". *RNA Detection: Methods and Protocols*. Ed. by Imre Gaspar. New York, NY: Springer New York, pp. 85–94. ISBN: 9781493972135. DOI: 10.1007/978-1-4939-7213-5_5.

Shigeoka, Toshiaki, Max Koppers, et al. (2019). "On-Site Ribosome Remodeling by Locally Synthesized Ribosomal Proteins in Axons". *Cell reports* 29.11, 3605–3619.e10. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2019.11.025.

Siddiqui, Nida and Anne Straube (2020). *The Kinesin-3 Family*. CRC Press. DOI: 10.1201/9780429491559-4.

Simsek, Deniz et al. (2017). "The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity". *Cell* 169.6, 1051–1065.e18. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2017.05.022.

Slavov, Nikolai et al. (2015). "Differential Stoichiometry among Core Ribosomal Proteins". *Cell reports* 13.5, pp. 865–873. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.09.056.

Sleigh, James N et al. (2019). "Axonal transport and neurological disease". *Nature reviews. Neurology* 15.12, pp. 691–703. ISSN: 1759-4758, 1759-4766. DOI: 10.1038/s41582-019-0257-2.

Sloan, Katherine E et al. (2017). "Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function". *RNA biology* 14.9, pp. 1138–1152. ISSN: 1547-6286, 1555-8584. DOI: 10.1080/15476286.2016.1259781.

Soppina, Virupakshi et al. (2014). "Dimerization of mammalian kinesin-3 motors results in superprocessive motion". *Proceedings of the National Academy of Sciences of the United States of America* 111.15, pp. 5562–5567. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1400759111.

Sotelo, José Roberto et al. (2014). "Glia to axon RNA transfer". *Developmental neurobiology* 74.3, pp. 292–302. ISSN: 1932-8451, 1932-846X. DOI: 10.1002/dneu.22125.

Spahn, C M et al. (2001). "Structure of the 80S ribosome from Saccharomyces cerevisiae–tRNA-ribosome and subunit-subunit interactions". *Cell* 107.3, pp. 373–386. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(01)00539-6.

Steward, O and W B Levy (1982). "Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 2.3, pp. 284–291. ISSN: 0270-6474.

Stoler, Nicholas and Anton Nekrutenko (2021). "Sequencing error profiles of Illumina sequencing instruments". *NAR genomics and bioinformatics* 3.1, lqab019. ISSN: 2631-9268. DOI: 10.1093/nargab/lqab019.

Sugihara, Yoshihiko et al. (2010). "Proteomic analysis of rodent ribosomes revealed heterogeneity including ribosomal proteins L10-like, L22-like 1, and L39-like". *Journal of proteome research* 9.3, pp. 1351–1366. ISSN: 1535-3893, 1535-3907. DOI: 10.1021/pr9008964.

Suk, Terry R and Maxime W C Rousseaux (2020). "The role of TDP-43 mislocalization in amyotrophic lateral sclerosis". *Molecular neurodegeneration* 15.1, p. 45. ISSN: 1750-1326. DOI: 10.1186/s13024-020-00397-1.

Sung, Min-Kyung et al. (2016). "Ribosomal proteins produced in excess are degraded by the ubiquitin-proteasome system". *Molecular biology of the cell* 27.17, pp. 2642–2652. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E16-05-0290.

Suntsova, Maria et al. (2019). "Atlas of RNA sequencing profiles for normal human tissues". *Scientific data* 6.1, p. 36. ISSN: 2052-4463. DOI: 10.1038/s41597-019-0043-4.

Suyama, Alexa et al. (2023). "Orthogonal Versatile Interacting Peptide Tags for Imaging Cellular Proteins". *Biochemistry* 62.11, pp. 1735–1743. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/acs.biochem.2c00712.

Svoboda, K et al. (1993). "Direct observation of kinesin stepping by optical trapping interferometry". *Nature* 365.6448, pp. 721–727. ISSN: 0028-0836. DOI: 10.1038/365721a0.

Takeda, S et al. (2000). "Kinesin superfamily protein 3 (KIF3) motor transports fodrin-associating vesicles important for neurite building". *The Journal of cell biology* 148.6, pp. 1255–1265. ISSN: 0021-9525. DOI: 10.1083/jcb.148.6.1255.

Taoka, Masato et al. (2018). "Landscape of the complete RNA chemical modifications in the human 80S ribosome". *Nucleic acids research* 46.18, pp. 9289–9298. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky811.

Taylor, Anne Marion et al. (2013). "Axonal translation of β-catenin regulates synaptic vesicle dynamics". *The Journal of neuroscience: the official journal of the Society for Neuroscience* 33.13, pp. 5584–5589. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2944-12.2013.

Tempes, Aleksandra et al. (2020). "Role of dynein-dynactin complex, kinesins, motor adaptors, and their phosphorylation in dendritogenesis". *Journal of neurochemistry* 155.1, pp. 10–28. ISSN: 0022-3042, 1471-4159. DOI: 10.1111/jnc.15010.

Tennyson, V M (1970). "The fine structure of the axon and growth cone of the dorsal root neuroblast of the rabbit embryo". *The Journal of cell biology* 44.1, pp. 62–79. ISSN: 0021-9525. DOI: 10.1083/jcb.44.1.62.

The Gene Ontology Consortium (2021). "The Gene Ontology resource: enriching a GOld mine". *Nucleic acids research* 49.D1, p. D325. ISSN: 0305-1048. DOI: 10.1093/nar/gkaa1113.

Thermo Fisher Scientific Inc. (2011). *Transfecting Neural Cells Using the Neon® Transfection System - UK*. https://www.thermofisher.com/uk/en/home/references/protocols/neurobiology/neurobiology-protocols/transfecting-neural-cells-using-the-neon-transfection-system.html. Accessed: 2023-6-7.

Thomas, Franziska and Ulrike Kutay (2003). "Biogenesis and nuclear export of ribosomal subunits in higher eukaryotes depend on the CRM1 export pathway". *Journal of cell science* 116.Pt 12, pp. 2409–2419. ISSN: 0021-9533. DOI: 10.1242/jcs.00464.

Thomson, Emma et al. (2013). "Eukaryotic ribosome biogenesis at a glance". *Journal of cell science* 126.Pt 21, pp. 4815–4821. ISSN: 0021-9533, 1477-9137. DOI: 10.1242/jcs.111948.

Tokunaga, Makio et al. (2008). "Highly inclined thin illumination enables clear single-molecule imaging in cells". *Nature methods* 5.2, pp. 159–161. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth1171.

Tushev, Georgi et al. (2018). "Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments". *Neuron* 98.3, 495–511.e6. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2018.03.030.

Twelvetrees, Alison E et al. (2016). "The Dynamic Localization of Cytoplasmic Dynein in Neurons Is Driven by Kinesin-1". *Neuron* 90.5, pp. 1000–1015. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2016.04.046.

Twiss, Jeffery L and Mike Fainzilber (2009). "Ribosomes in axons–scrounging from the neighbors?" *Trends in cell biology* 19.5, pp. 236–243. ISSN: 0962-8924, 1879-3088. DOI: 10.1016/j.tcb.2009.02.007.

Watkins, N J et al. (2000). "A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP". *Cell* 103.3, pp. 457–466. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(00)00137-9.

White, Laura K and Jay R Hesselberth (2022). "Modification mapping by nanopore sequencing". *Frontiers in genetics* 13, p. 1037134. ISSN: 1664-8021. DOI: 10.3389/fgene.2022.1037134.

Wiatrak, Benita et al. (2020). "PC12 Cell Line: Cell Types, Coating of Culture Vessels, Differentiation and Other Culture Conditions". *Cells* 9.4. ISSN: 2073-4409. DOI: 10.3390/cells9040958.

Willems, Jelmer et al. (2020). "ORANGE: A CRISPR/Cas9-based genome editing toolbox for epitope tagging of endogenous proteins in neurons". *PLoS biology* 18.4, e3000665. ISSN: 1544-9173, 1545-7885. DOI: 10.1371/journal.pbio.3000665.

Wilson, Daniel N and Jamie H Doudna Cate (2012). "The structure and function of the eukaryotic ribosome". *Cold Spring Harbor perspectives in biology* 4.5. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a011536.

Wong, H H-W et al. (2017). "RNA Docking and Local Translation Regulate Site-Specific Axon Remodeling In Vivo". *Neuron* 95.4, 852–868.e8. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2017.07.016.

Yadav, Anupama et al. (2016). "The Modular Adaptive Ribosome". *PloS one* 11.11, e0166021. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0166021.

Yao, Jiaqi et al. (2006). "An essential role for beta-actin mRNA localization and translation in Ca2+-dependent growth cone guidance". *Nature neuroscience* 9.10, pp. 1265–1273. ISSN: 1097-6256. DOI: 10.1038/nn1773.

Yoon, Young J et al. (2016). "Glutamate-induced RNA localization and translation in neurons". *Proceedings of the National Academy of Sciences of the United States*

of America 113.44, E6877–E6886. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1614267113.

Yu, Lingli et al. (2015). "Highly efficient method for gene delivery into mouse dorsal root ganglia neurons". *Frontiers in molecular neuroscience* 8, p. 2. ISSN: 1662-5099. DOI: 10.3389/fnmol.2015.00002.

Yudin, Dmitry et al. (2008). "Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve". *Neuron* 59.2, pp. 241–252. ISSN: 0896-6273, 1097-4199. DOI: 10.1016/j.neuron.2008.05.029.

Chapter 8

Appendix

ICE CRISPR Analysis Code

The results below are generated from an R script.

```
# R Script to re-plot raw data from Synthego ICE CRISPR Analysis.
#This is example code used to plot data for RPS3 Guide 1.

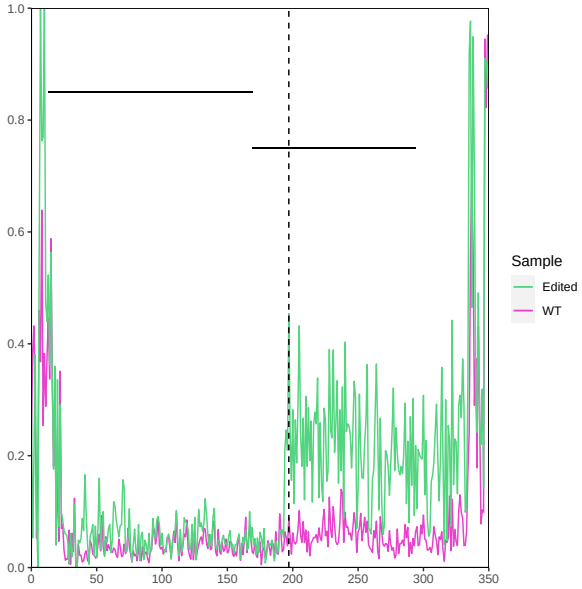
#Load library (install packages first)
library("ggplot2")
library("jsonlite")
library("tidyverse")

#Read in json file + convert to dataframe
df <- jsonlite::fromJSON("~/plot.json", simplifyDataFrame = TRUE)
df2 <- as.data.frame(df)

## Warning in (function (..., row.names = NULL, check.rows = FALSE, check.names = TRUE, :  row
names were found from a short variable and have been discarded

#df2 - check dataframe

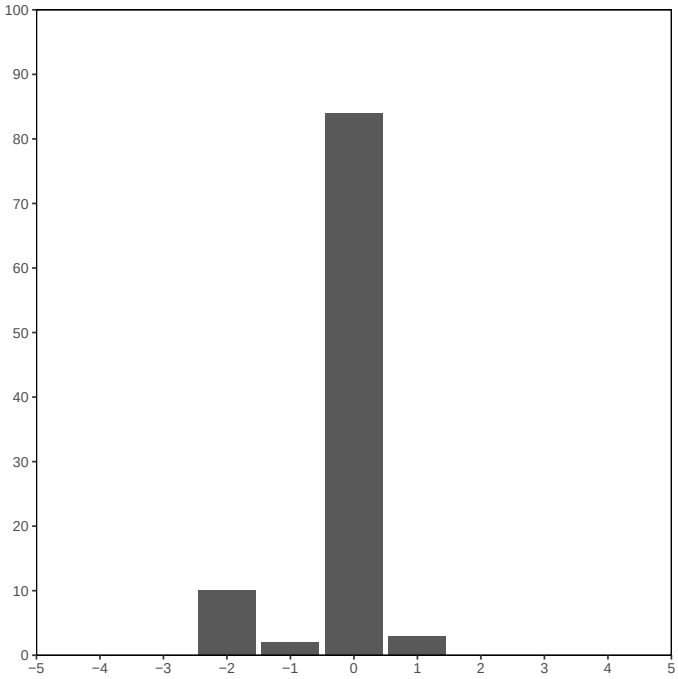
#plot discordance data
p1 <- ggplot(df2, aes(x = discord_plot.bp))+
  geom_line(aes(y=discord_plot.control_discord, color="Magenta"))+
  geom_line(aes(y=discord_plot.edited_discord, color="Green"))+
  scale_color_manual(name="Sample", values = c("#53D47F","#E642CB"), labels = c("Edited","WT"))+
  geom_vline(xintercept=197, linetype="dashed", show.legend = FALSE)+ #cutsite line
  geom_segment(aes(x = 13, y = 0.85, xend = 169, yend = 0.85))+ #alignment line
  geom_segment(aes(x = 169, y = 0.75, xend = 294, yend = 0.75))+ #inference line
  scale_x_continuous(breaks= c(0,50,100,150,200,250,300,350),limits=c(0,350),expand = c(0, 0)) +
  scale_y_continuous(breaks= c(0,0.2,0.4,0.6,0.8,1.0), expand = c(0, 0))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(fill=NA, colour="black"),
        axis.title.x = element_blank(),
        axis.title.y = element_blank())
p1 #print p1
```




```
#generate indel dataframe
indel_percentage <- df2[1,10:55]
indel_percentage_list <- as.list(indel_percentage)
indel_length <- c(-30:15)
indel_percent <- c(unlist(indel_percentage_list))
indel_percent <- as.integer(indel_percent)
indel_df2 <- cbind.data.frame(indel_length,indel_percent)

#plot indel graph
p2 <- ggplot(indel_df2, aes(x=indel_length, y=indel_percent))+
  geom_bar(stat="identity")+
  scale_x_continuous(breaks=c(-5:5),limits=c(-5,5),expand = c(0, 0)) +
  scale_y_continuous(breaks= c(0,10,20,30,40,50,60,70,80,90,100),limits=c(0,100),expand = c(0, 0))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(fill=NA, colour="black"),
        axis.title.x = element_blank(),
        axis.title.y = element_blank())
p2 #print p2

## Warning:  Removed 35 rows containing missing values (position_stack).
## Warning:  Removed 2 rows containing missing values (geom_bar).
```



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
```

```
## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.utf8  LC_CTYPE=English_United Kingdom.utf8
## [3] LC_MONETARY=English_United Kingdom.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.39      forcats_0.5.1   stringr_1.4.0   dplyr_1.0.9     purrr_0.3.4
##  [6] readr_2.1.2     tidyr_1.2.0     tibble_3.1.7    tidyverse_1.3.1 jsonlite_1.8.0
## [11] ggplot2_3.3.6
##
## loaded via a namespace (and not attached):
##  [1] tinytex_0.38        xfun_0.31           tidyselect_1.1.2    haven_2.5.0
##  [5] colorspace_2.0-3    vctrs_0.4.1         generics_0.1.2      stats4_4.2.0
##  [9] utf8_1.2.2          flowCore_2.8.0      rlang_1.0.2         pillar_1.7.0
## [13] glue_1.6.2          withr_2.5.0         DBI_1.1.2           BiocGenerics_0.42.0
## [17] dbplyr_2.1.1        modelr_0.1.8        readxl_1.4.0        matrixStats_0.62.0
## [21] lifecycle_1.0.1     RProtoBufLib_2.8.0  munsell_0.5.0       gtable_0.3.0
## [25] cytolib_2.8.0       cellranger_1.1.0    rvest_1.0.2         evaluate_0.16
```

```
## [29] Biobase_2.56.0      tzdb_0.3.0          fansi_1.0.3         highr_0.9
## [33] broom_0.8.0         Rcpp_1.0.8.3        scales_1.2.0        backports_1.4.1
## [37] S4Vectors_0.34.0    RcppParallel_5.1.5  farver_2.1.0        fs_1.5.2
## [41] digest_0.6.29       hms_1.1.1           stringi_1.7.6       grid_4.2.0
## [45] cli_3.3.0           tools_4.2.0         magrittr_2.0.3      crayon_1.5.1
## [49] pkgconfig_2.0.3     ellipsis_0.3.2      xml2_1.3.3          reprex_2.0.1
## [53] lubridate_1.8.0     assertthat_0.2.1    httr_1.4.3          rstudioapi_0.13
## [57] R6_2.5.1            compiler_4.2.0

Sys.time()

## [1] "2022-08-15 15:56:03 BST"
```

Figure 8.1: R Script for ICE Analysis.

Code used to re-plot data generated by ICE CRISPR analysis (see Chapter 3). Data was re-plotted using R (version 4.2.0).

FACS Analysis Code

The results below are generated from an R script.

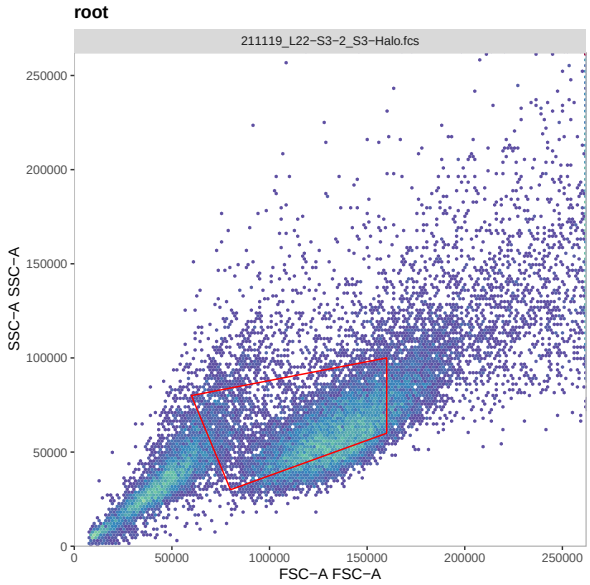
```
# R version 4.2.0

# Load packages
library(flowCore)
library(ggcyto)
library(tidyverse)
library(knitr)
library(forcats)

# Read in .fcs files
fs <- flowCore::read.flowSet(path = "C:\\Users\\mdp19afd\\Documents", pattern = "*.fcs")
pData(fs)$sample <- gsub("211119_L22-S3-2_(.*).fcs","\\1",sampleNames(fs)) #rename samples
pData(fs)$exp_date <- gsub("_.*", "", sampleNames(fs))
gs <- GatingSet(fs) #Allows gates to be added to flowSet


#Gating data FSC and SSC-A
g.live <- polygonGate(filterId = "Live","FSC-A"=c(8e4,16e4,16e4,6e4),"SSC-A"=c(3e4,6e4,10e4,8e4)) # defi
ggcyto(gs[[1]],aes(x="FSC-A",y="SSC-A"),subset="root")+
  geom_hex(bins = 150)+geom_gate(g.live)+
  ggcyto_par_set(limits = "instrument")+ # check gate
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0))+
  theme(panel.grid = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        panel.border = element_rect(fill=NA, colour="grey"),
        panel.background = element_blank())

## Coordinate system already present.  Adding new coordinate system, which will replace the
existing one.
```




```
add(gs,g.live) # add gate to GatingSet

## Warning:  'add.default' is deprecated.
## Use 'gs_pop_add' instead.
## See help("Deprecated")

## [1] 2

recompute(gs) # recompute GatingSet

## done!

#Plot all .fcs files in one plot
g.live <- polygonGate(filterId = "Live","FSC-A"=c(7e4,18e4,17e4,7e4),"SSC-A"=c(1e4,4e4,10e4,5e4)) # def
add(gs,g.live) # add gate to GatingSet

## Warning:  'add.default' is deprecated.
## Use 'gs_pop_add' instead.
## See help("Deprecated")
## Error in .cpp_addGate(ptr, sn, filterObject, parent, name):  Live already exists!

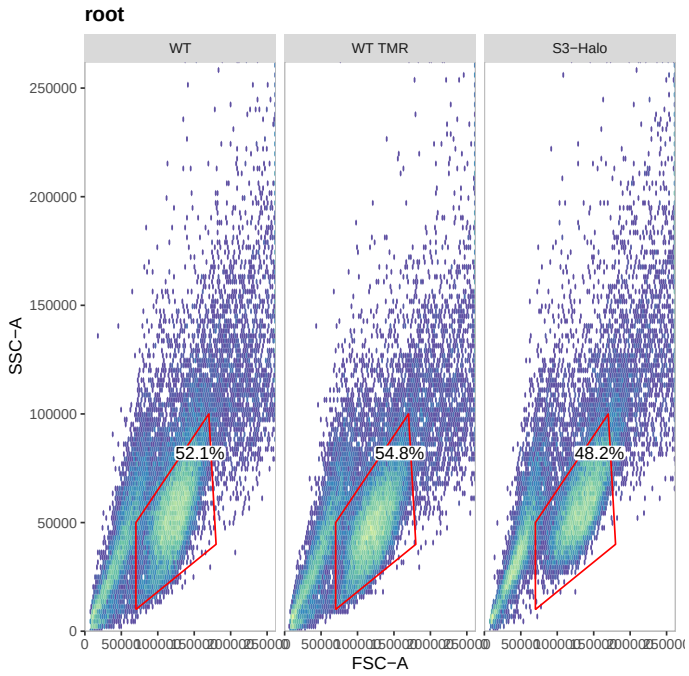
recompute(gs) # recompute GatingSet

## done!

ggcyto(gs,aes(x="FSC-A",y="SSC-A"),subset="root")+
  geom_hex(bins = 100)+
  geom_gate(g.live)+
  ggcyto_par_set(limits = "instrument")+
  geom_stats(adjust = 0.8)+
  labs_cyto("channel")+
  scale_x_continuous(expand = c(0, 0)) +
```

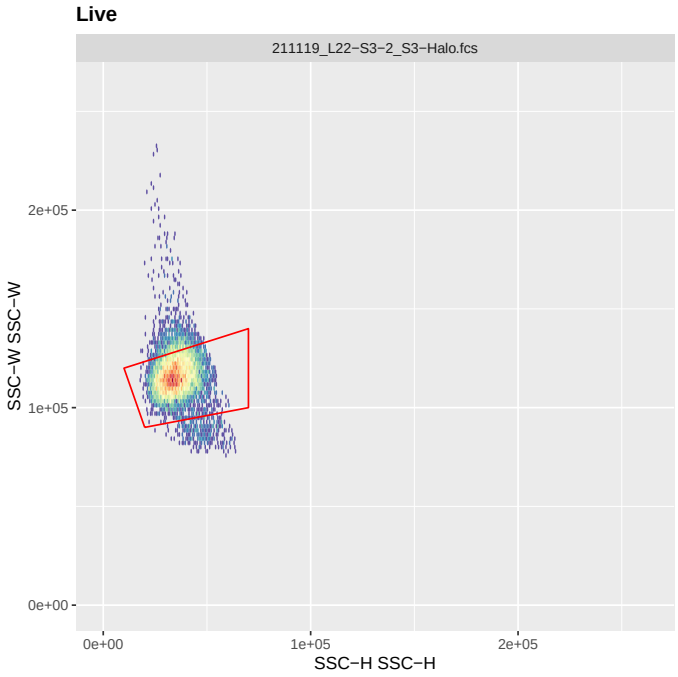
```
  scale_y_continuous(expand = c(0, 0))+
  facet_wrap(~factor(sample, levels=c('WT','WT TMR','S3-Halo')))+ #re-order plots
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_rect(fill=NA, colour="grey"),
        panel.background = element_blank())
```

```
## Coordinate system already present.  Adding new coordinate system, which will replace the
existing one.
```




```
#Gating SSC-H and SSC-W
g.singlets <- polygonGate(filterId = "Singlets","SSC-H"=c(2e4,7e4,7e4,1e4),"SSC-W"=c(9e4,10e4,14e4,12e4)
ggcyto(gs[[1]],aes(x="SSC-H",y="SSC-W"),subset="Live")+
  geom_hex(bins = 64)+
  geom_gate(g.singlets)+
  ggcyto_par_set(limits = "instrument") # check gate
```

```
## Coordinate system already present.  Adding new coordinate system, which will replace the
existing one.
```




```
add(gs,g.singlets, parent="Live") # add gate to GatingSet

## Warning:  'add.default' is deprecated.
## Use 'gs_pop_add' instead.
## See help("Deprecated")

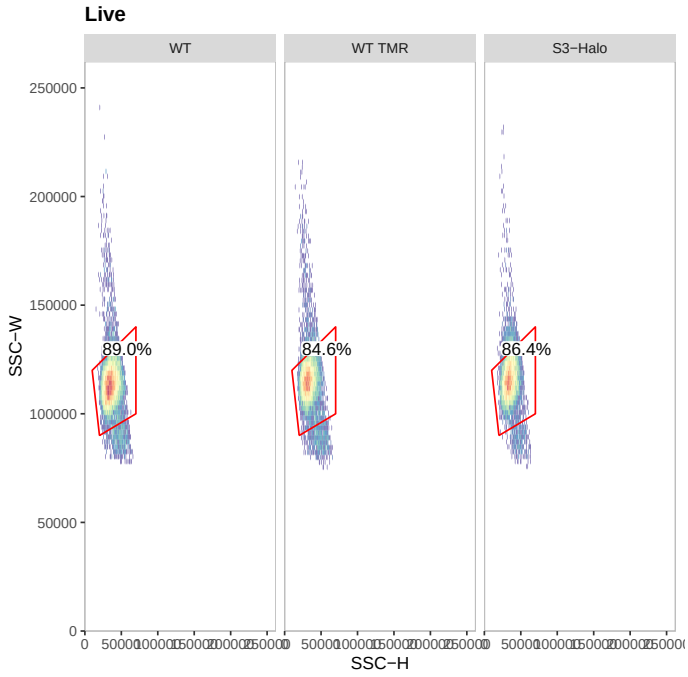
## [1] 3

recompute(gs) # recompute GatingSet

## done!

#Plot all files SSC-H and SSC-W Gating
ggcyto(gs,aes(x="SSC-H",y="SSC-W"),subset="Live")+
  geom_hex(bins = 64)+
  geom_gate(g.singlets)+
  ggcyto_par_set(limits = "instrument")+
  geom_stats(adjust = 0.8)+
  labs_cyto("channel")+
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0))+
  facet_wrap(~factor(sample, levels=c('WT','WT TMR','S3-Halo')))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_rect(fill=NA, colour="grey"),
        panel.background = element_blank())

## Coordinate system already present.  Adding new coordinate system, which will replace the
existing one.
```




```
add(gs,g.singlets, parent="Live") # add gate to GatingSet

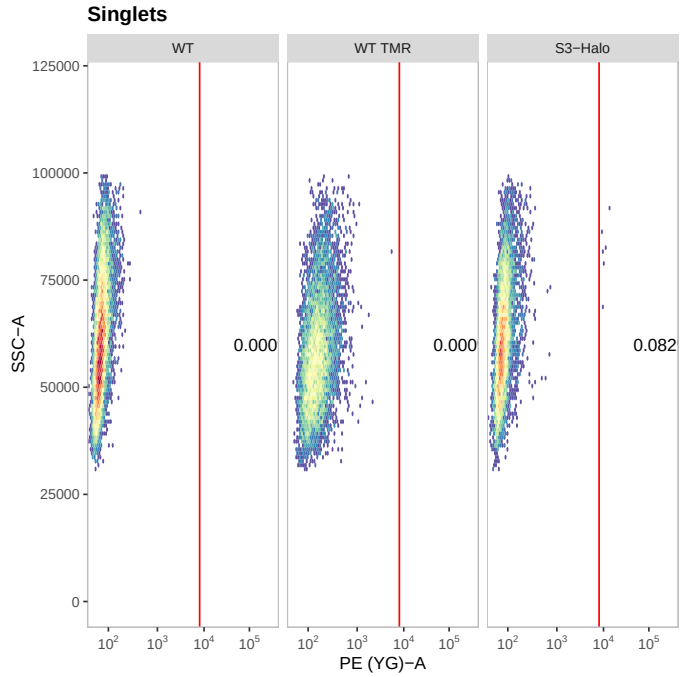
## Warning:  'add.default' is deprecated.
## Use 'gs_pop_add' instead.
## See help("Deprecated")
## Error in .cpp_addGate(ptr, sn, filterObject, parent, name):  Singlets already exists!

recompute(gs) # recompute GatingSet

## done!

#Plot all TMR graphs
g.tmr <- rectangleGate(filterId="TMR+","PE (YG)-A"=c(8000, Inf)) # set gate
ggcyto(gs,aes(x="PE (YG)-A",y="SSC-A"),subset="Singlets")+
  geom_hex(bins = 64)+
  geom_gate(g.tmr)+
  ggcyto_par_set(limits = list(x = c(0,400000), y = c(0, 120000)))+
  scale_x_flowjo_biexp(expand = c(0, 0))+
  geom_stats(type="percent", digits=2)+
  labs_cyto("channel")+
  facet_wrap(~factor(sample, levels=c('WT','WT TMR','S3-Halo')))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_rect(fill=NA, colour="grey"),
        panel.background = element_blank())

## Coordinate system already present.  Adding new coordinate system, which will replace the
existing one.
## Warning:  Removed 1 rows containing missing values (geom_segment).
## Warning:  Removed 1 rows containing missing values (geom_segment).
## Removed 1 rows containing missing values (geom_segment).
```



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
```

```
## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.utf8  LC_CTYPE=English_United Kingdom.utf8
## [3] LC_MONETARY=English_United Kingdom.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.39               ggcyto_1.24.0            flowWorkspace_4.8.0
##  [4] ncdfFlow_2.42.0          BH_1.78.0-0              RcppArmadillo_0.11.0.0.0
##  [7] flowCore_2.8.0           rlist_0.4.6.2            scico_1.3.1
## [10] forcats_0.5.1            stringr_1.4.0            dplyr_1.0.9
## [13] purrr_0.3.4              readr_2.1.2              tibble_3.1.7
## [16] tidyverse_1.3.2          tidyr_1.2.0              jsonlite_1.8.0
## [19] rvest_1.0.2              ggplot2_3.3.6
##
## loaded via a namespace (and not attached):
##  [1] matrixStats_0.62.0    fs_1.5.2              lubridate_1.8.0
##  [4] RColorBrewer_1.1-3    httr_1.4.3            Rgraphviz_2.40.0
##  [7] tools_4.2.0           backports_1.4.1       utf8_1.2.2
```

```
## [10] R6_2.5.1              DBI_1.1.2             BiocGenerics_0.42.0
## [13] colorspace_2.0-3      withr_2.5.0           tidyselect_1.1.2
## [16] gridExtra_2.3         curl_4.3.2            compiler_4.2.0
## [19] graph_1.74.0          cli_3.3.0             Biobase_2.56.0
## [22] xml2_1.3.3            labeling_0.4.2        scales_1.2.0
## [25] hexbin_1.28.2         digest_0.6.29         base64enc_0.1-3
## [28] jpeg_0.1-9            pkgconfig_2.0.3       highr_0.9
## [31] dbplyr_2.1.1          rlang_1.0.2           readxl_1.4.0
## [34] rstudioapi_0.13       farver_2.1.0          generics_0.1.2
## [37] googlesheets4_1.0.0   magrittr_2.0.3        RProtoBufLib_2.8.0
## [40] aws.ec2metadata_0.2.0 Rcpp_1.0.8.3          munsell_0.5.0
## [43] S4Vectors_0.34.0      fansi_1.0.3           lifecycle_1.0.1
```

Figure 8.2: R Script for FACS Analysis.

Code used to re-plot data generated by FACS Chorus software (see Chapter 3). Data was re-plotted using R (version 4.2.0).

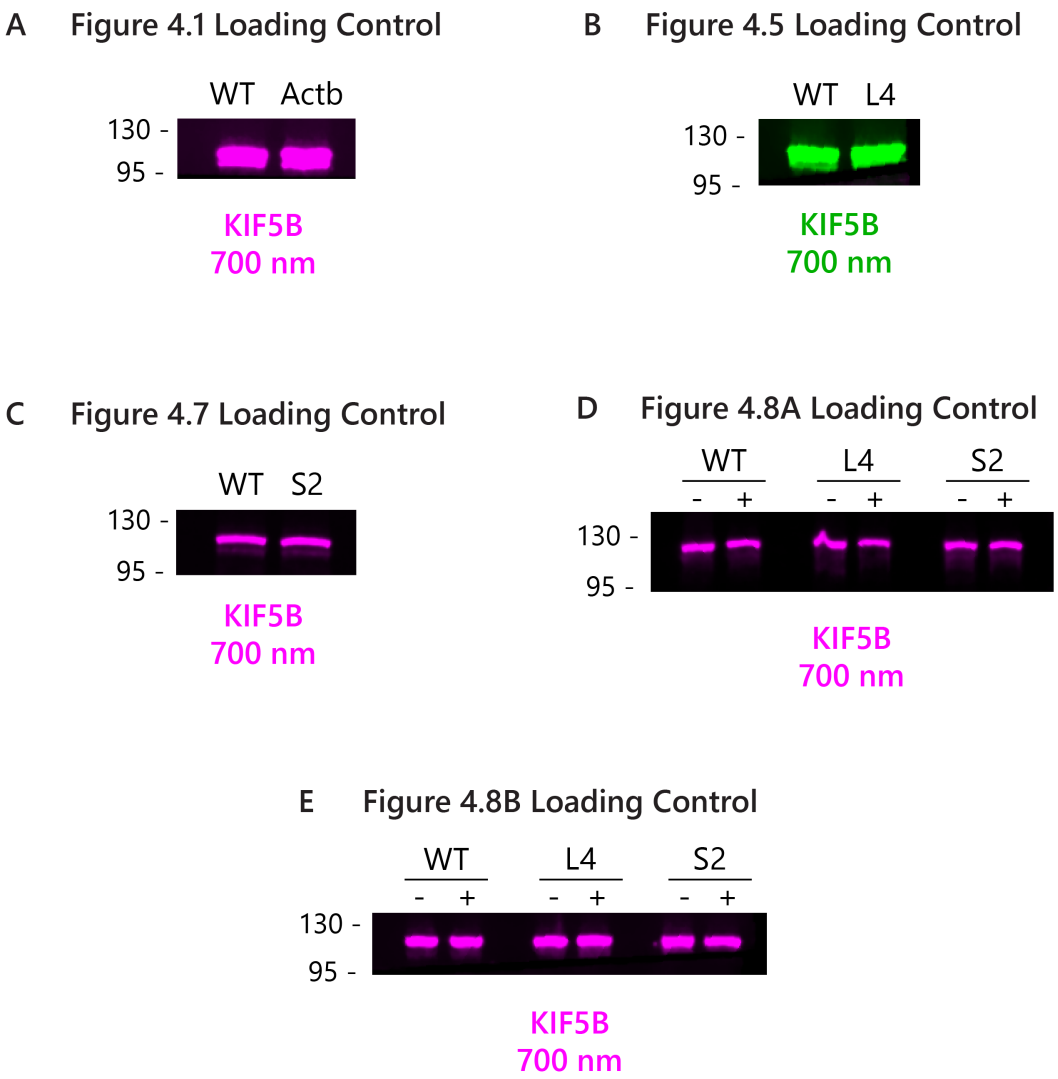


Figure 8.3: Loading Controls for Chapter 4 Western Blots.

To ensure equal sample loading for western blots, KIF5B primary antibody was used as a loading control. The antibody was used at a 1:1000 dilution. Abbreviations: Actb-GFP-Actb PC-12 Lysate, L4- RPL4-Halo PC-12 Lysate, S2- RPS2-Halo RPC-12 Lysate.

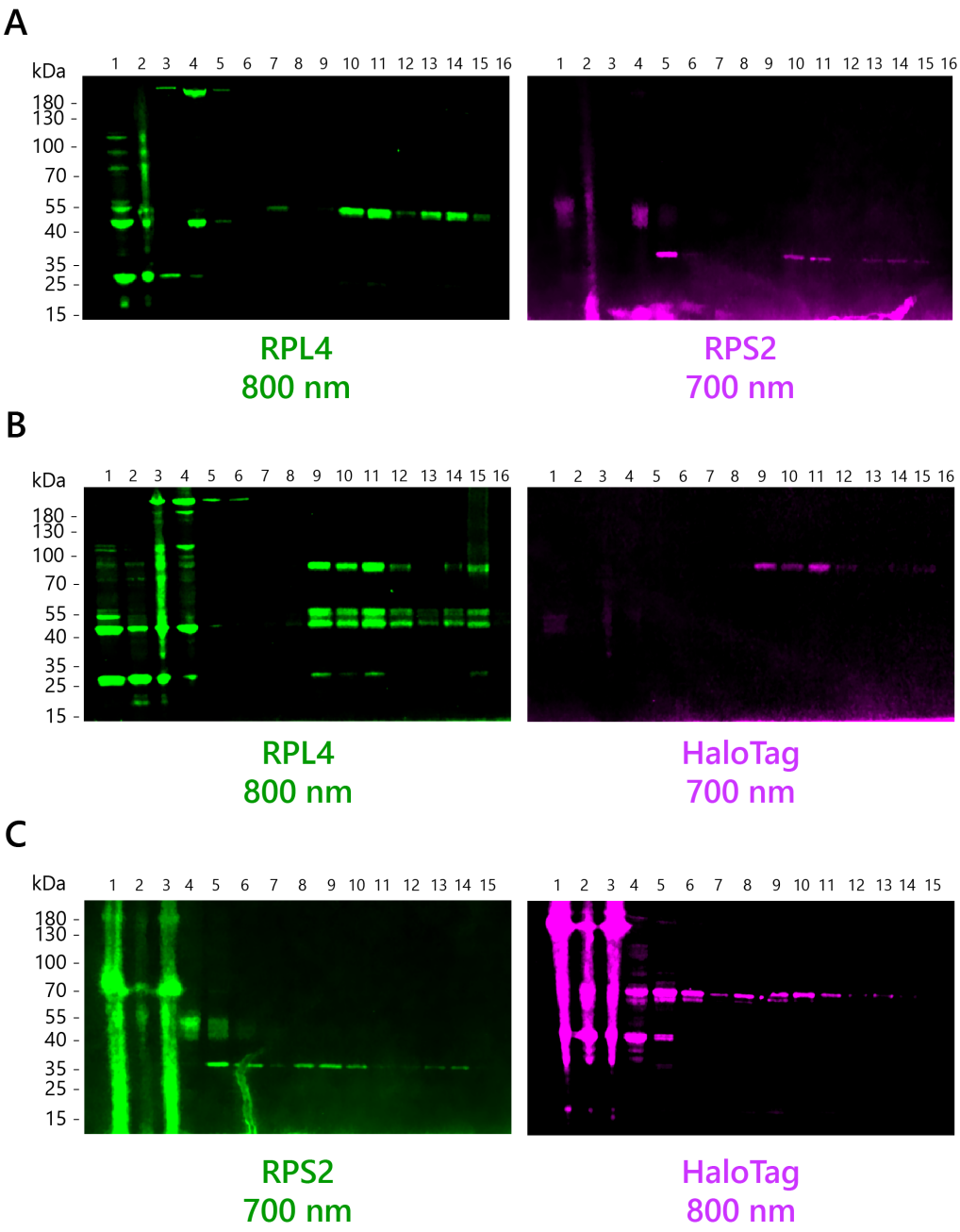


Figure 8.4: Polysome Profiling Western Blots.

Polysome profiling was conducted as detailed in Chapter 2. WT, RPL4-Halo, and RPS2-Halo blots were used in a fluorescent western, proved simultaneously for ribosomal protein or HaloTag primary antibodies. Details of antibody suppliers and concentrations used is present in Chapter 2. Imaging was done using a LI-COR XF Odyssey.

Table 8.1: Human Ribosomal Protein Genes Used in Gene Expression Analysis.

Gene Symbol	Gencode ID	Entrez Gene	Genomic Location
60S Subunit Genes			
RPL3	ENSG00000100316.15	6122	22q13.1
RPL3L	ENSG00000140986.7	6123	16p13.3
RPL4	ENSG00000174444.14	6124	15q22.31
RPL5	ENSG00000122406.12	6125	1p22.1
RPL6	ENSG00000089009.15	6128	12q24.13
RPL7	ENSG00000147604.13	6129	8q21.11
RPL7A	ENSG00000148303.16	6130	9q34.2
RPL7L1	ENSG00000146223.14	285855	6p21.1
RPL8	ENSG00000161016.17	6132	8q24.3
RPL9	ENSG00000163682.15	6133	4p14
RPL10	ENSG00000147403.16	6134	Xq28
RPL10A	ENSG00000198755.10	4736	6p21.31
RPL10L	ENSG00000165496.4	140801	14q21.2
RPL11	ENSG00000142676.12	6135	1p36.11
RPL12	ENSG00000197958.12	6136	9q33.3
RPL13	ENSG00000167526.13	6137	16q24.3
RPL13A	ENSG00000142541.16	23521	19q13.33
RPL14	ENSG00000188846.13	9045	3p22.1
RPL15	ENSG00000174748.18	6138	3p24.2
RPL17	ENSG00000265681.7	6139	18q21.1
RPL18	ENSG00000063177.12	6141	19q13.33
RPL18A	ENSG00000105640.12	6142	19p13.11
RPL19	ENSG00000108298.9	6143	17q12
RPL21	ENSG00000122026.10	6144	13q12.2
RPL22	ENSG00000116251.9	6146	1p36.31
RPL22L1	ENSG00000163584.17	200916	3q26.2
RPL23	ENSG00000125691.12	9349	17q12
RPL23A	ENSG00000198242.13	6147	17q11.2
RPL24	ENSG00000114391.12	6152	3q12.3
RPL26	ENSG00000161970.12	6154	17p13.1
RPL26L1	ENSG00000037241.7	51121	5q35.1
RPL27	ENSG00000131469.12	6155	17q21.31
RPL27A	ENSG00000166441.12	6157	11p15.4
RPL28	ENSG00000108107.14	6158	19q13.42
RPL29	ENSG00000162244.10	6159	3p21.2
RPL30	ENSG00000156482.10	6156	8q22.2
RPL31	ENSG00000071082.10	6160	2q11.2
RPL32	ENSG00000144713.12	6161	3p25.2
RPL34	ENSG00000109475.16	6164	4q25
RPL35	ENSG00000136942.14	11224	9q33.3
RPL35A	ENSG00000182899.14	6165	3q29
RPL36	ENSG00000130255.12	25873	19p13.3
RPL36A	ENSG00000241343.9	6173	Xq22.1

Gene Symbol	Gencode ID	Entrez Gene	Genomic Location
RPL36AL	ENSG00000165502.6	6166	14q21.3
RPL37	ENSG00000145592.13	6167	5p13.1
RPL37A	ENSG00000197756.9	6168	2q35
RPL38	ENSG00000172809.12	6169	17q25.1
RPL39	ENSG00000198918.7	6170	Xq24
RPL39L	ENSG00000163923.9	116832	3q27.3
UBA52	ENSG00000221983.7	7311	19p13.1-p12
RPL41	ENSG00000229117.8	6171	12q13
RPLP0	ENSG00000089157.15	6175	12q24.23
RPLP1	ENSG00000137818.11	6176	15q23
RPLP2	ENSG00000177600.8	6181	11p15.5
40S Subunit Genes			
RPSA	ENSG00000168028.13	3921	3p22.1
RPS2	ENSG00000140988.15	6187	16p13.3
RPS3	ENSG00000149273.14	6188	11q13.4
RPS3A	ENSG00000145425.9	6189	4q31.3
RPS4X	ENSG00000198034.10	6191	Xq13.1
RPS4Y1	ENSG00000129824.15	6192	Yp11.2
RPS4Y2	ENSG00000280969.1	140032	Yq11.223
RPS5	ENSG00000083845.8	6193	19q13.43
RPS6	ENSG00000137154.12	6194	9p22.1
RPS7	ENSG00000171863.12	6201	2p25.3
RPS8	ENSG00000142937.11	6202	1p34.1
RPS9	ENSG00000170889.13	6203	19q13.42
RPS10	ENSG00000124614.13	6204	6p21.31
RPS11	ENSG00000142534.6	6205	19q13.3
RPS12	ENSG00000112306.7	6206	6q23.2
RPS13	ENSG00000110700.6	6207	11p15.1
RPS14	ENSG00000164587.11	6208	5q33.1
RPS15	ENSG00000115268.9	6209	19p13.3
RPS15A	ENSG00000134419.15	6210	16p12.3
RPS16	ENSG00000105193.8	6217	19q13.2
RPS17	ENSG00000182774.10	6218	15q25.2
RPS18	ENSG00000231500.6	6222	6p21.32
RPS19	ENSG00000105372.6	6223	19q13.2
RPS20	ENSG00000008988.9	6224	8q12.1
RPS21	ENSG00000171858.17	6227	20q13.33
RPS23	ENSG00000186468.12	6228	5q14.2
RPS24	ENSG00000138326.18	6229	10q22.3
RPS25	ENSG00000118181.10	6230	11q23.3
RPS26	ENSG00000197728.9	6231	12q13.2
RPS27	ENSG00000177954.11	6232	1q21.3
RPS27L	ENSG00000185088.13	6232	15q22.2
RPS27A	ENSG00000143947.13	6233	15q22.2
RPS28	ENSG00000233927.4	6234	19p13.2
RPS29	ENSG00000213741.9	6235	14q21.3
FAU	ENSG00000149806.10	2197	11q13.1

Ribosome Protein and snoRNA Gene Expression Analysis Code

June 23, 2023

The results below are generated from an R script.

```r
# 07/05/20 Data Normalised and cleaned up. Ovary + testes removed

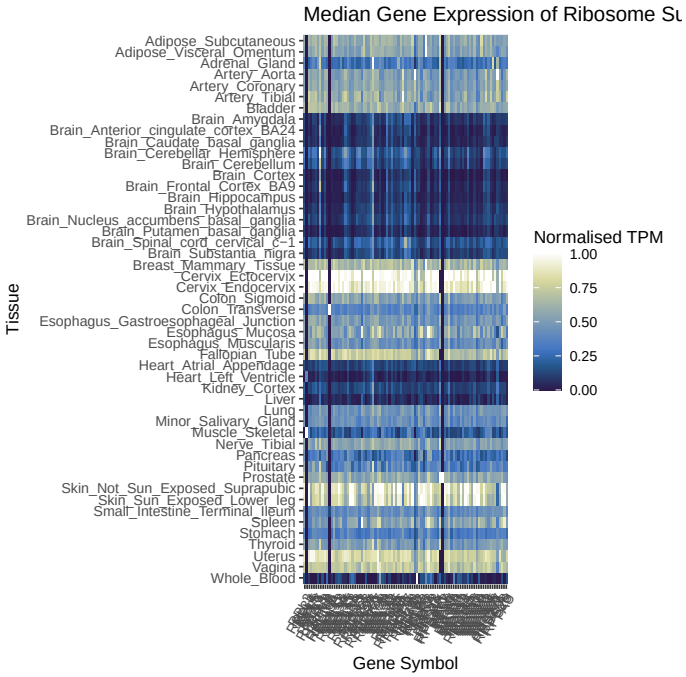
# Load in libraries
library("tidyverse")
library("rvest")
library("jsonlite")
library("tidyverse")
library("scico")
library("rlist")

# Read in gene .csv, transform into list
proteins<- read.csv("Ribosomal protein on Gtex Final.csv")
protId<- proteins[[1]]
protSym<- proteins[[5]]
protIdList<- paste(as.character(protId),collapse="%2C") #gets protein id in url format


#Get data from GETEx
url <- paste("https://gtexportal.org/rest/v1/expression/geneExpression?datasetId=gtex_v8&gencodeId=",pro
df <- fromJSON(url(url), simplifyDataFrame = TRUE) #url() used because url too long for jsonlite
df <-unnest(data=df$geneExpression, cols=c(data))
df <- mutate(.data=df, tissueSiteDetailId=factor(tissueSiteDetailId,levels=rev(sort(unique(tissueSiteDet
df <- subset(df, select = c("data", "tissueSiteDetailId","geneSymbol"))
agg <- aggregate(.~geneSymbol+tissueSiteDetailId, df, median)  #group dataframe and find mean
agg$data[is.na(agg$data)] <- 0  #remove NaN
agg$geneSymbol<- factor(agg$geneSymbol, levels = protSym)


# Normalise data
normdf<- data.frame() #create max and min df
for (i in unique(agg$geneSymbol)){
  max<- max(agg$data[which(agg$geneSymbol==i)])
  min<- min(agg$data[which(agg$geneSymbol==i)])
  output<- data.frame(geneSymbol=i, max=max, min=min)
  normdf<- rbind(normdf, output)
}
agg$normdata <- 0 #create empty column for norm data (outside of for loop)
for (i in 1:nrow(agg)) { #cycle through rows, normalise data, append agg column
  row <- agg[i,]
  normdata <- (row$data - normdf$min[which(normdf$geneSymbol==row$geneSymbol)]) / (normdf$max[which(norm
  agg$normdata[i]<- normdata
}
```

```
#Plot
ggplot(agg, aes(x = geneSymbol, y = tissueSiteDetailId)) +
  geom_raster(aes(fill=normdata)) +
  scale_fill_scico(palette="davos", direction=1, limits=c(0,1), name="Normalised TPM")+
  labs(y="Tissue", x = "Gene Symbol") +
  ggtitle(paste("Median Gene Expression of Ribosome Subunit Genes (TPM)"))+
  theme(panel.grid = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        panel.border = element_rect(fill=NA, color=NA),
        axis.text.x = element_text(angle = 60, hjust = 1))
```



The R session information (including the OS info, R version and all packages used):

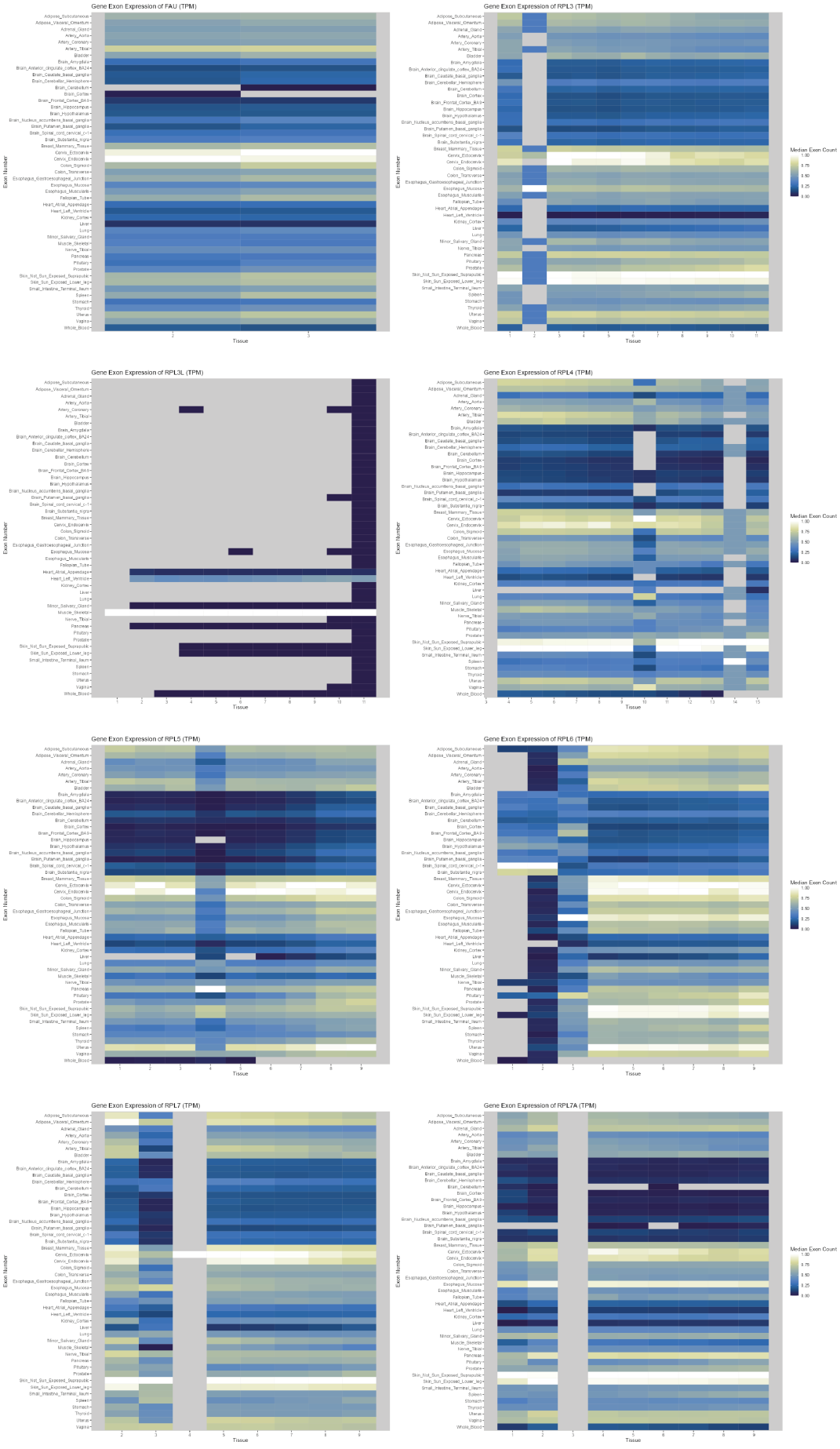
```
sessionInfo()

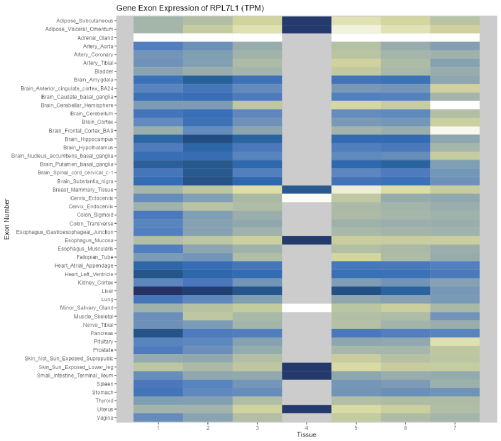
## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.utf8  LC_CTYPE=English_United Kingdom.utf8
## [3] LC_MONETARY=English_United Kingdom.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
```

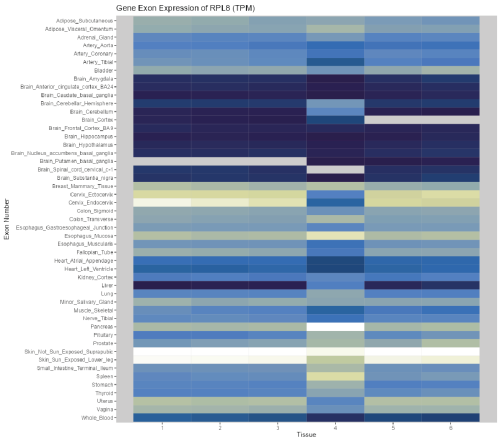
```
## other attached packages:
##  [1] knitr_1.39               ggcyto_1.24.0            flowWorkspace_4.8.0
##  [4] ncdfFlow_2.42.0          BH_1.78.0-0              RcppArmadillo_0.11.0.0.0
##  [7] flowCore_2.8.0           rlist_0.4.6.2            scico_1.3.1
## [10] forcats_0.5.1            stringr_1.4.0            dplyr_1.0.9
## [13] purrr_0.3.4              readr_2.1.2              tibble_3.1.7
## [16] tidyverse_1.3.2          tidyr_1.2.0              jsonlite_1.8.0
## [19] rvest_1.0.2              ggplot2_3.3.6
##
## loaded via a namespace (and not attached):
##  [1] matrixStats_0.62.0    fs_1.5.2              lubridate_1.8.0
##  [4] RColorBrewer_1.1-3    httr_1.4.3            Rgraphviz_2.40.0
```

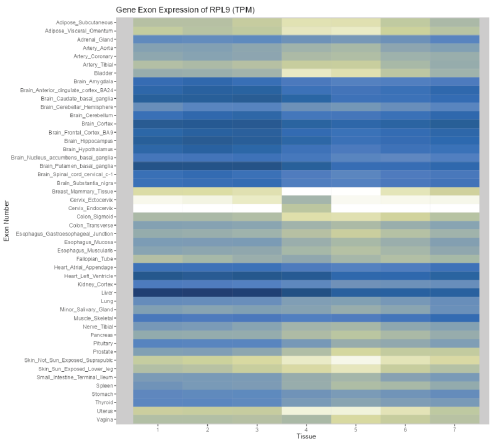
Figure 8.5: R Script for Ribosome Protein and snoRNA Gene Expression Analysis.

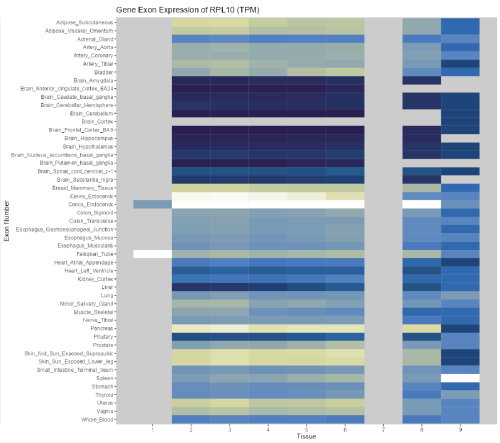
Code used to re-plot data generated by FACS Chorus software (see Chapter 6). Data was sourced from GTEx and plotted using R (version 4.2.0). This script plots the 89 ribosomal protein genes, but the same code was used for snoRNA analysis by reading in a different .csv file containing snoRNA gene details.

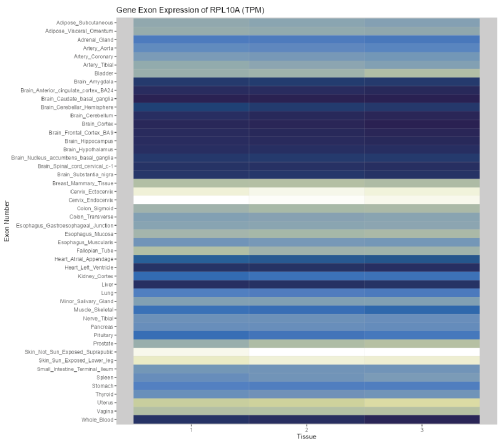

Gene Exon Expression of FAU (TPM)
Gene Exon Expression of RPL3 (TPM)
Gene Exon Expression of RPL3L (TPM)
Gene Exon Expression of RPL4 (TPM)
Gene Exon Expression of RPL5 (TPM)
Gene Exon Expression of RPL6 (TPM)
Gene Exon Expression of RPL7 (TPM)
Gene Exon Expression of RPL7A (TPM)
Exon Number
Tissue
Median Exon Count

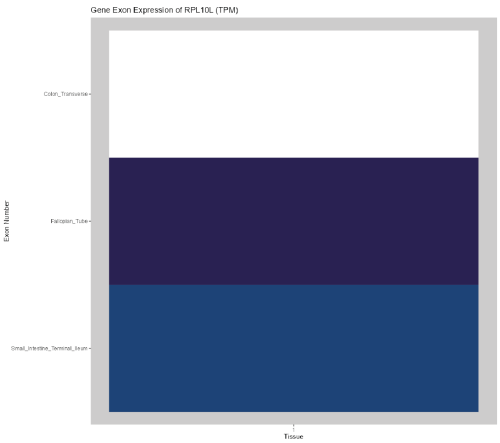

Gene Exon Expression of RPL7L1 (TPM)
Exon Number
Tissue

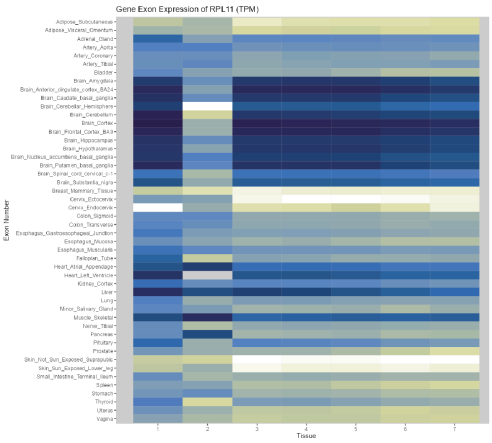

Gene Exon Expression of RPL8 (TPM)
Exon Number
Tissue
Median Exon Count

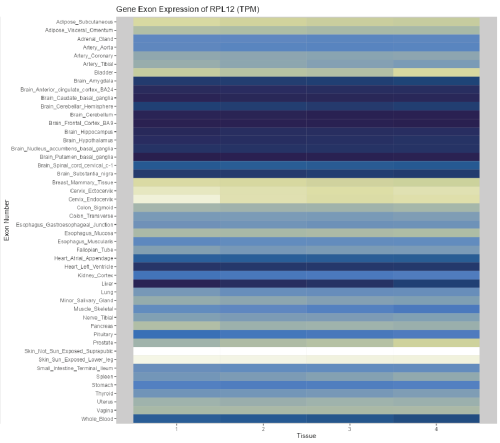

Gene Exon Expression of RPL9 (TPM)
Exon Number
Tissue

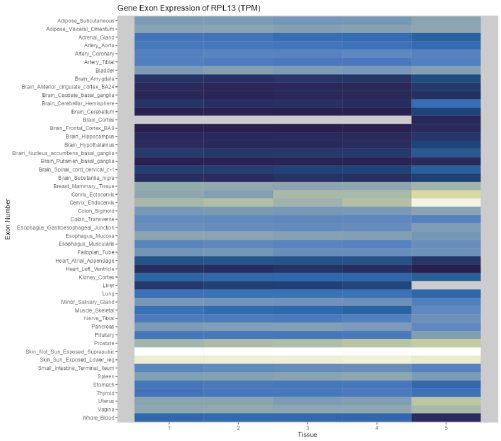

Gene Exon Expression of RPL10 (TPM)
Exon Number
Tissue
Median Exon Count

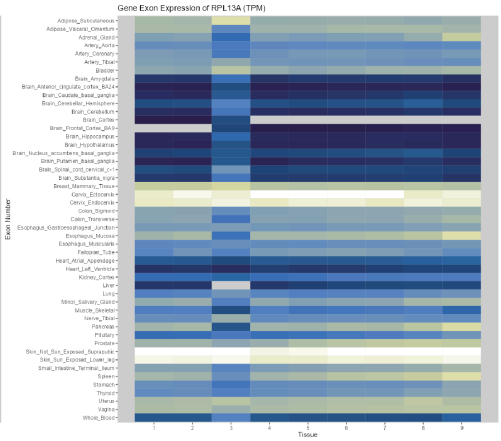

Gene Exon Expression of RPL10A (TPM)
Exon Number
Tissue

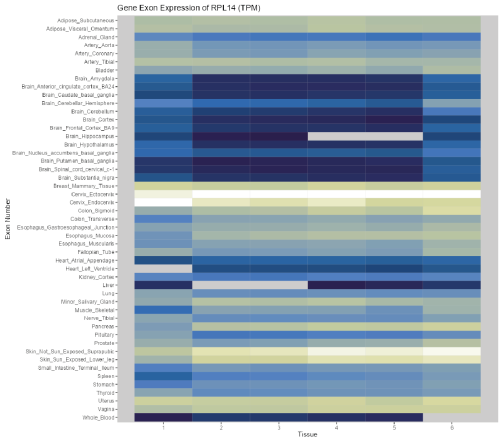

Gene Exon Expression of RPL10L (TPM)
Colon_Transverse
Fallopian_Tube
Small_Intestine_Terminal_Ileum
Exon Number
Tissue
Median Exon Count

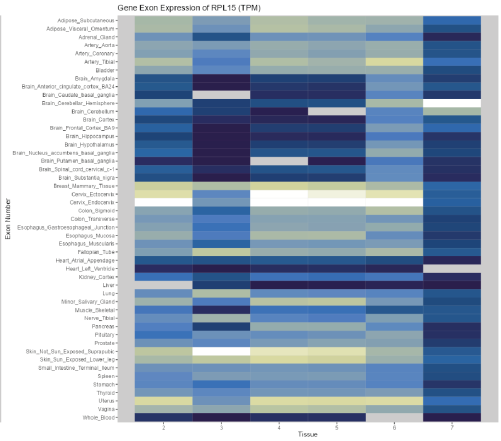

Gene Exon Expression of RPL11 (TPM)
Exon Number
Tissue

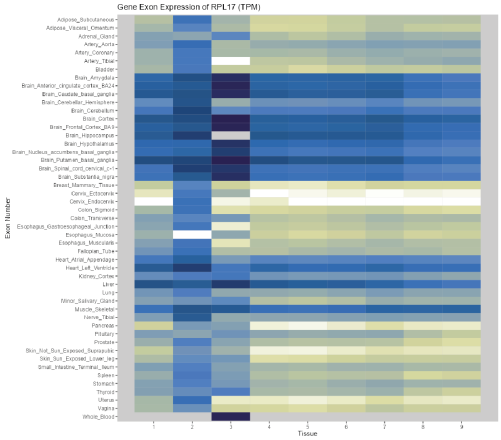

Gene Exon Expression of RPL12 (TPM)
Exon Number
Tissue
Median Exon Count

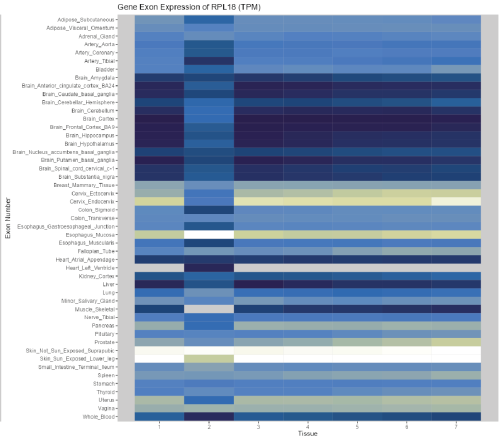

Gene Exon Expression of RPL13 (TPM)
Exon Number
Tissue

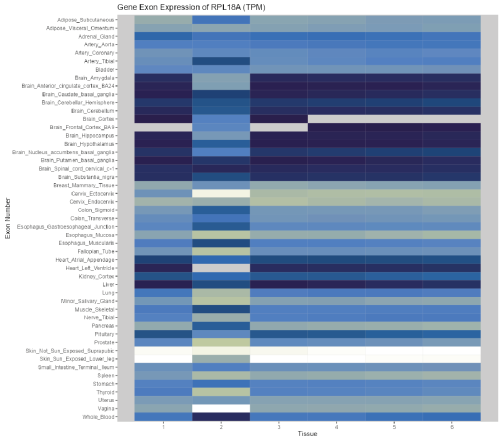

Gene Exon Expression of RPL13A (TPM)
Exon Number
Tissue
Median Exon Count

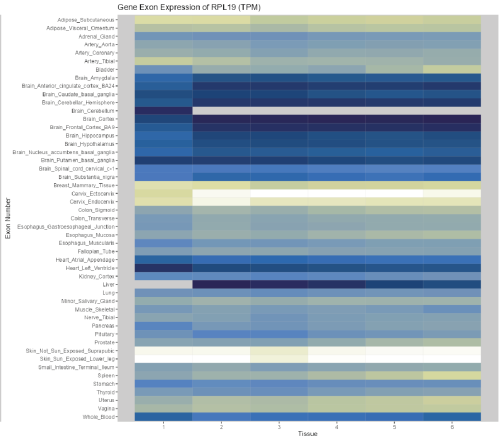

Gene Exon Expression of RPL14 (TPM)
Exon Number
Tissue

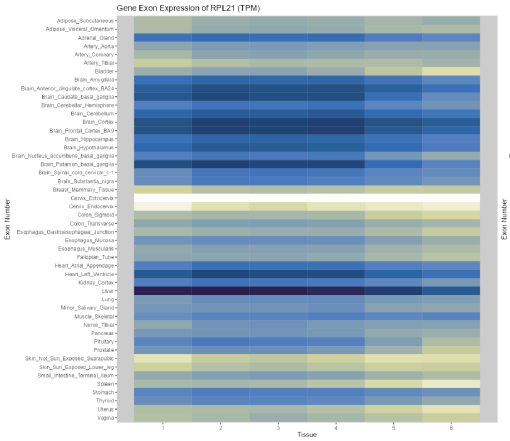

Gene Exon Expression of RPL15 (TPM)
Exon Number
Tissue
Median Exon Count

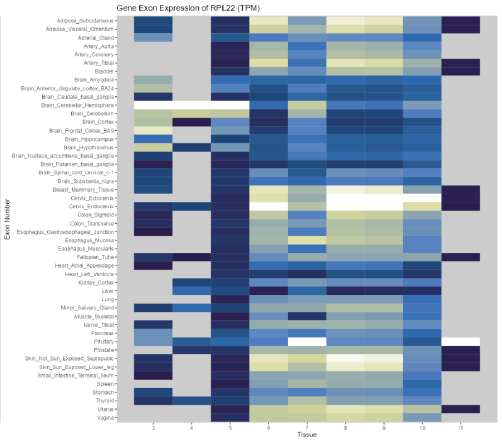

Gene Exon Expression of RPL17 (TPM)
Exon Number
Tissue

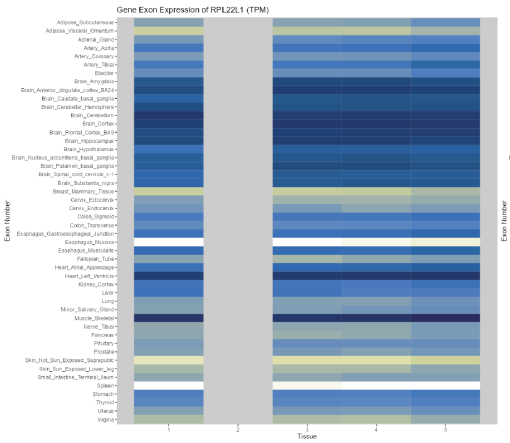

Gene Exon Expression of RPL18 (TPM)
Exon Number
Tissue
Median Exon Count

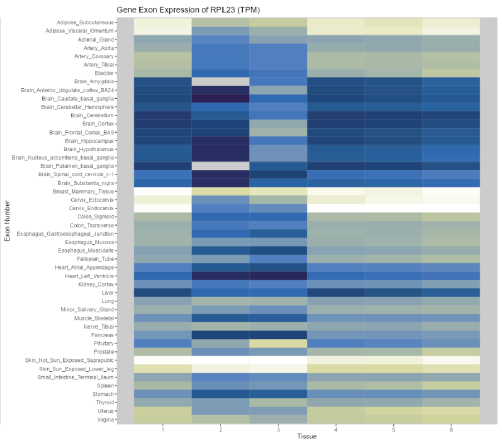

Gene Exon Expression of RPL18A (TPM)
Exon Number
Tissue

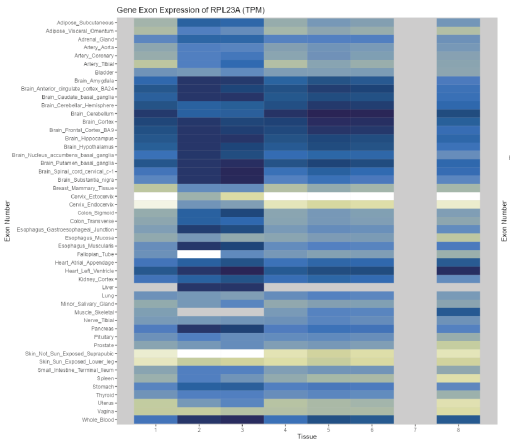

Gene Exon Expression of RPL19 (TPM)
Exon Number
Tissue
Median Exon Count

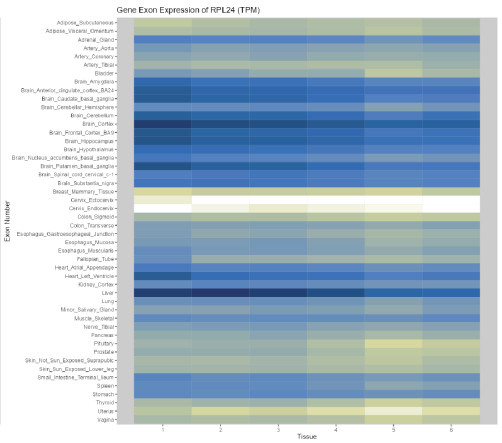

Gene Exon Expression of RPL21 (TPM)
Exon Number
Tissue

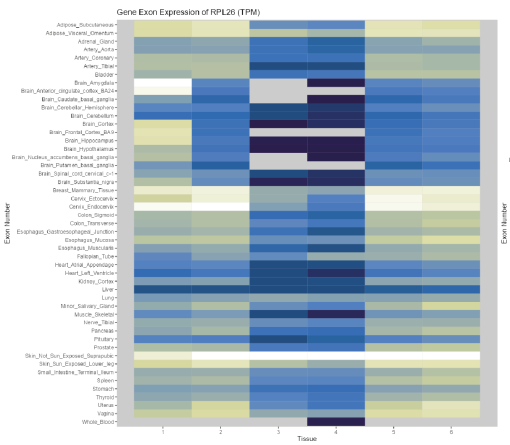

Gene Exon Expression of RPL22 (TPM)
Exon Number
Tissue
Median Exon Count

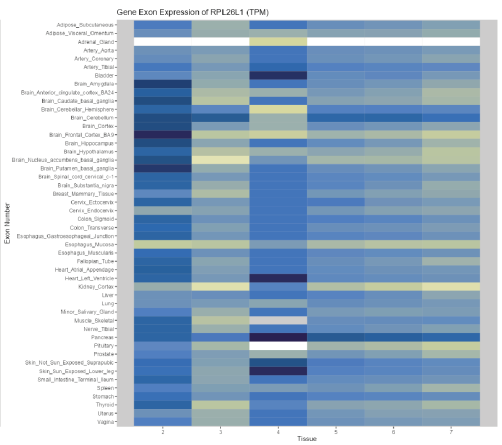

Gene Exon Expression of RPL22L1 (TPM)
Exon Number
Tissue

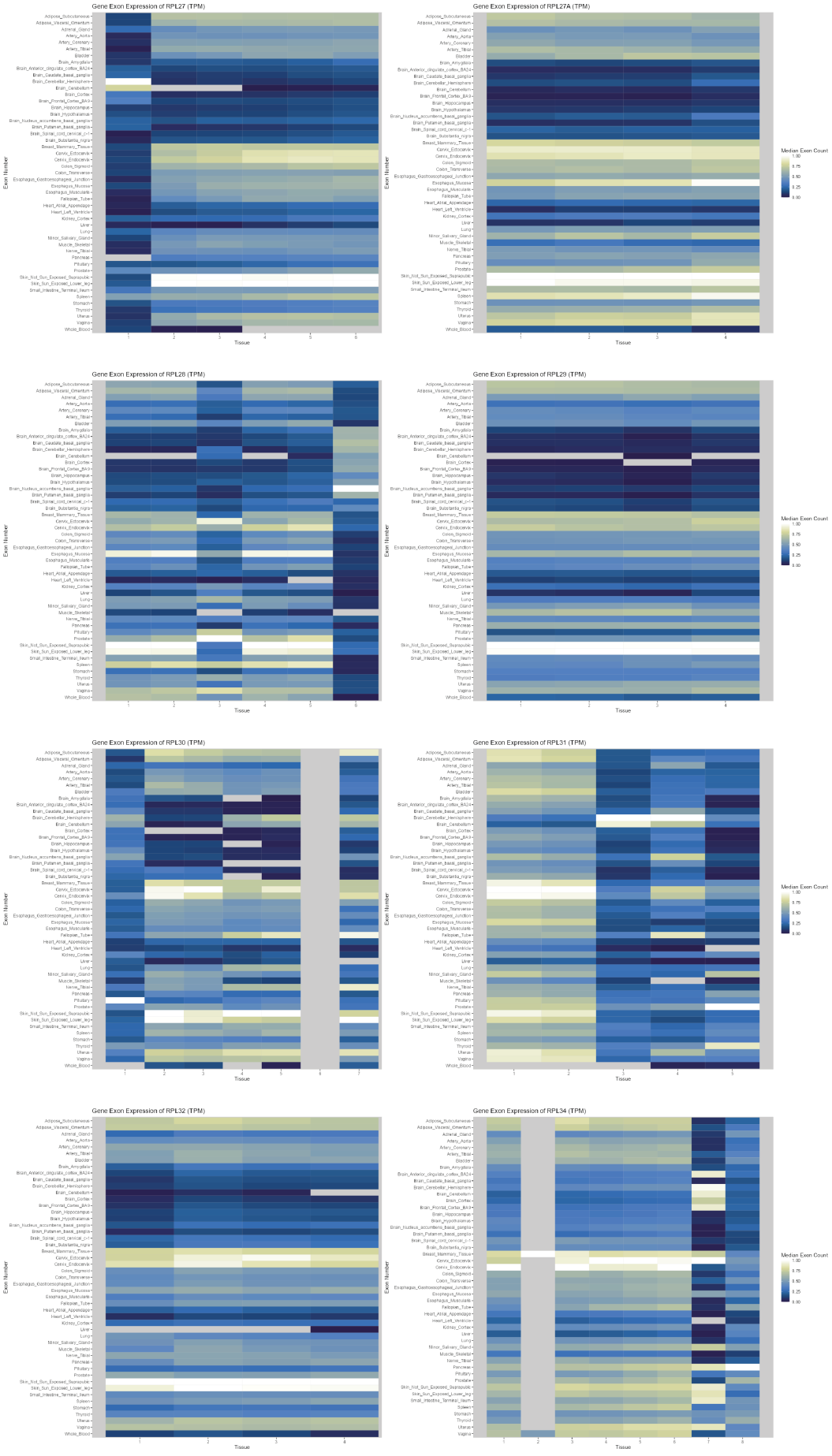

Gene Exon Expression of RPL23 (TPM)
Exon Number
Tissue
Median Exon Count

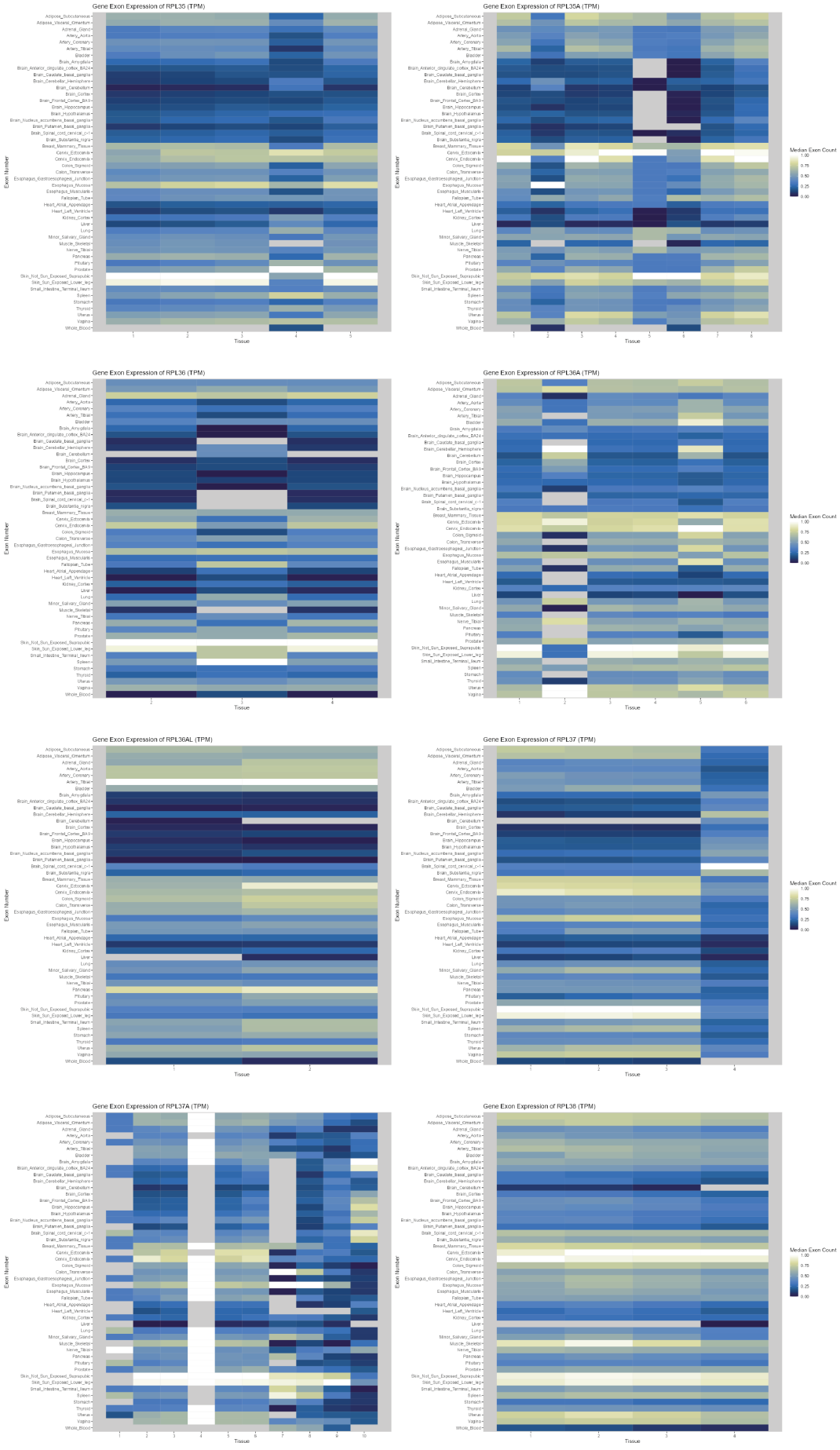

Gene Exon Expression of RPL23A (TPM)
Exon Number
Tissue

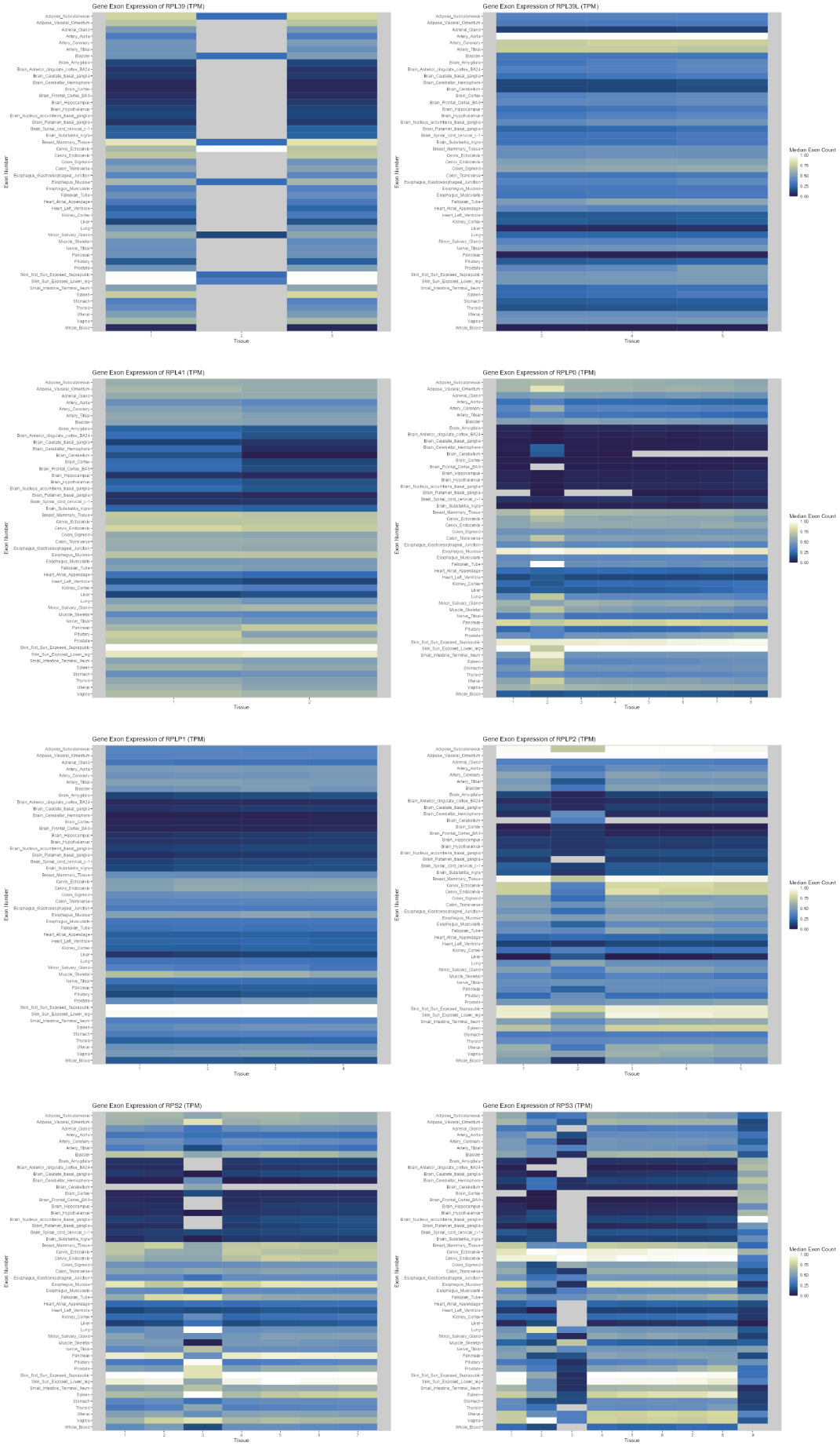

Gene Exon Expression of RPL24 (TPM)
Exon Number
Tissue
Median Exon Count

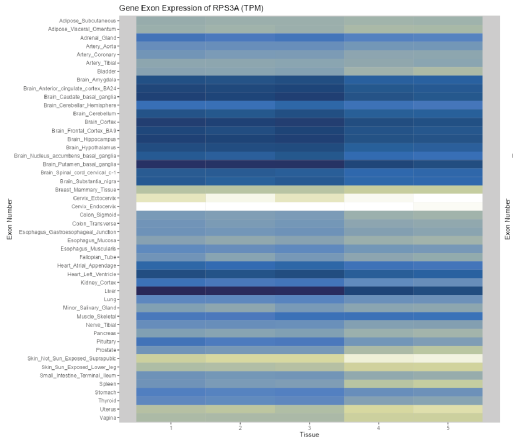

Gene Exon Expression of RPL26 (TPM)
Exon Number
Tissue

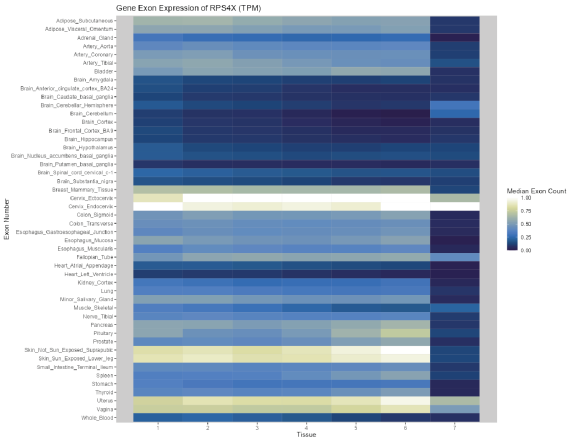

Gene Exon Expression of RPL26L1 (TPM)
Exon Number
Tissue
Median Exon Count

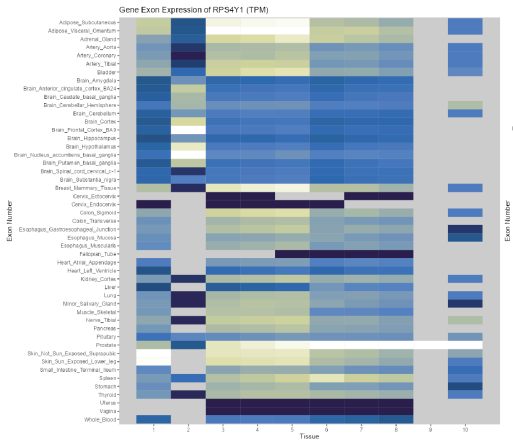

Gene Exon Expression of RPL27 (TPM)
Gene Exon Expression of RPL27A (TPM)
Gene Exon Expression of RPL28 (TPM)
Gene Exon Expression of RPL29 (TPM)
Gene Exon Expression of RPL30 (TPM)
Gene Exon Expression of RPL31 (TPM)
Gene Exon Expression of RPL32 (TPM)
Gene Exon Expression of RPL34 (TPM)
Exon Number
Tissue
Median Exon Count

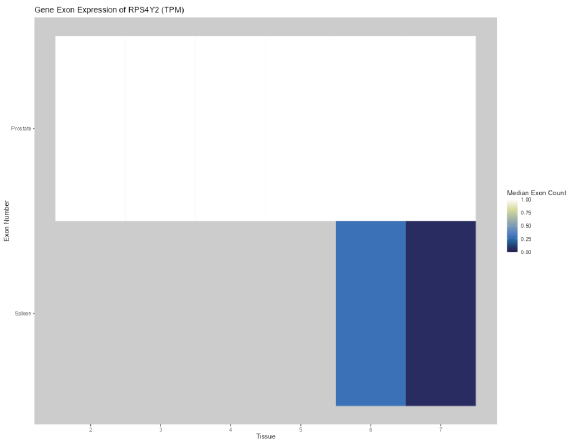

Gene Exon Expression of RPL35 (TPM)
Gene Exon Expression of RPL35A (TPM)
Gene Exon Expression of RPL36 (TPM)
Gene Exon Expression of RPL36A (TPM)
Gene Exon Expression of RPL36AL (TPM)
Gene Exon Expression of RPL37 (TPM)
Gene Exon Expression of RPL37A (TPM)
Gene Exon Expression of RPL38 (TPM)
Exon Number
Tissue
Median Exon Count

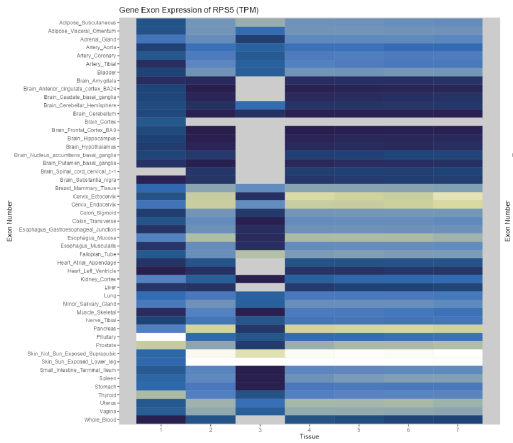

Gene Exon Expression of RPL39 (TPM)
Gene Exon Expression of RPL39L (TPM)
Gene Exon Expression of RPL41 (TPM)
Gene Exon Expression of RPLP0 (TPM)
Gene Exon Expression of RPLP1 (TPM)
Gene Exon Expression of RPLP2 (TPM)
Gene Exon Expression of RPS2 (TPM)
Gene Exon Expression of RPS3 (TPM)
Exon Number
Tissue
Median Exon Count

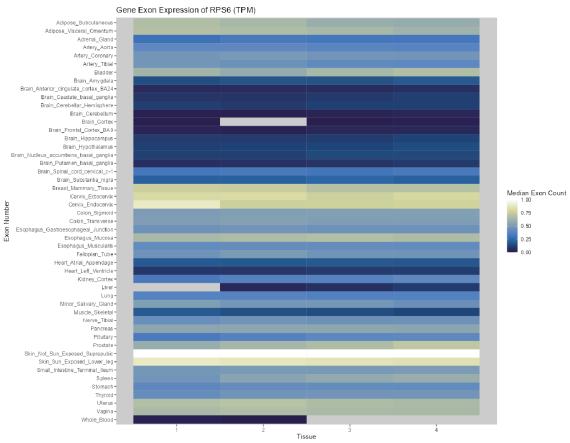

Gene Exon Expression of RPS3A (TPM)
Exon Number
Tissue

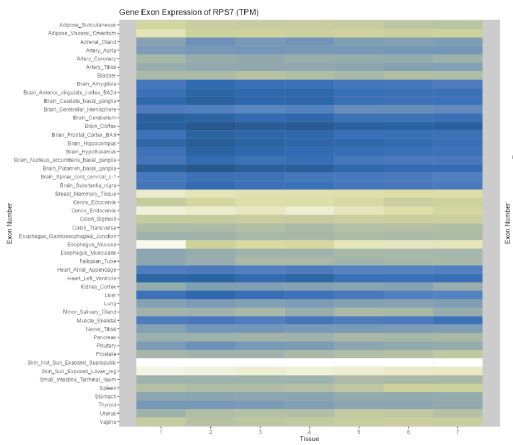

Gene Exon Expression of RPS4X (TPM)
Exon Number
Tissue
Median Exon Count

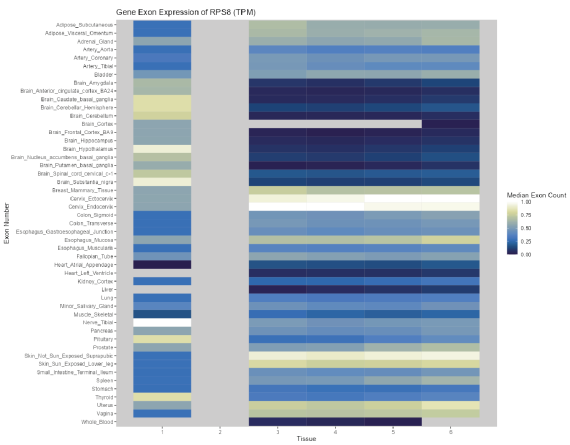

Gene Exon Expression of RPS4Y1 (TPM)
Exon Number
Tissue

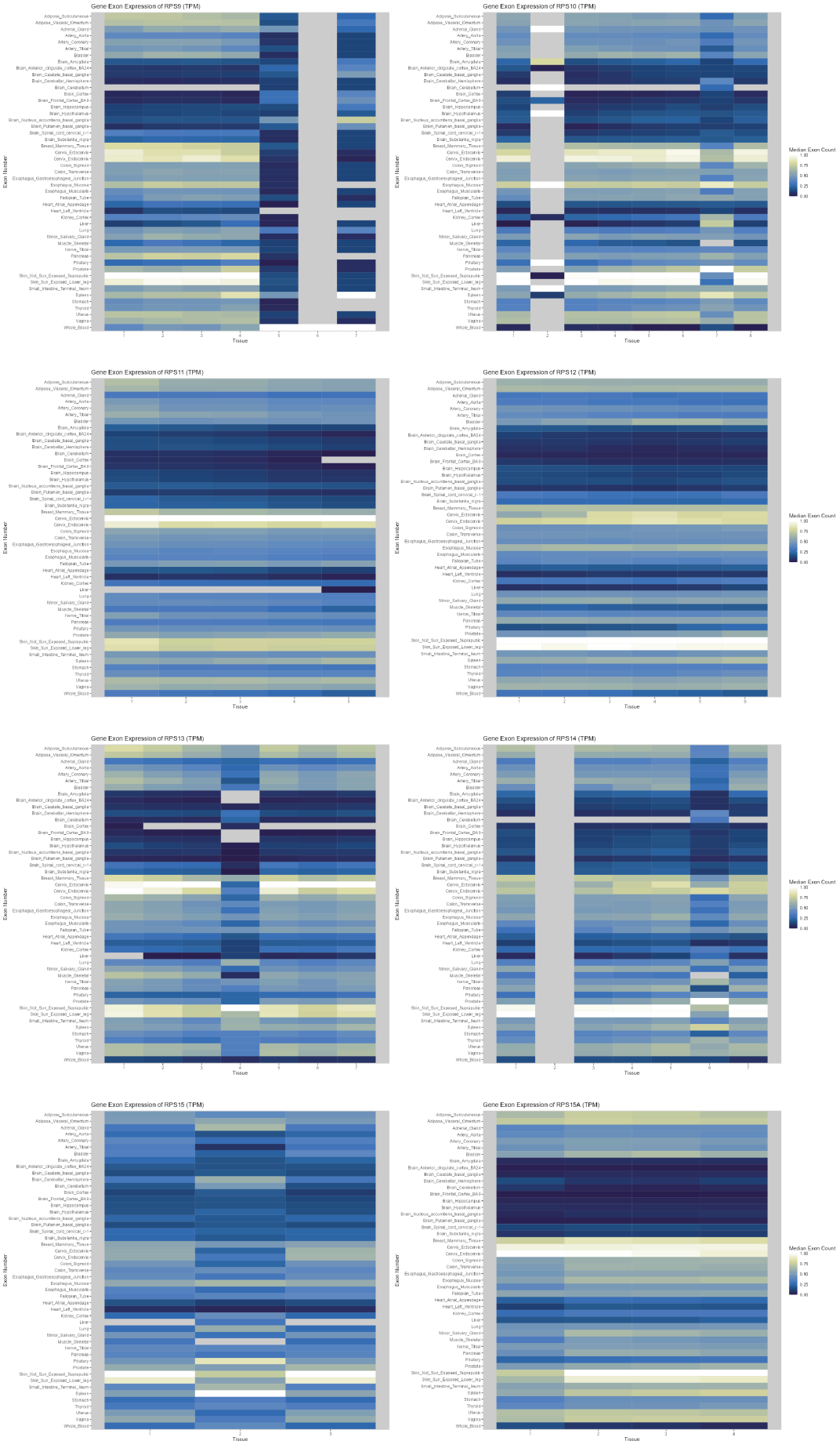

Gene Exon Expression of RPS4Y2 (TPM)
Exon Number
Tissue
Median Exon Count

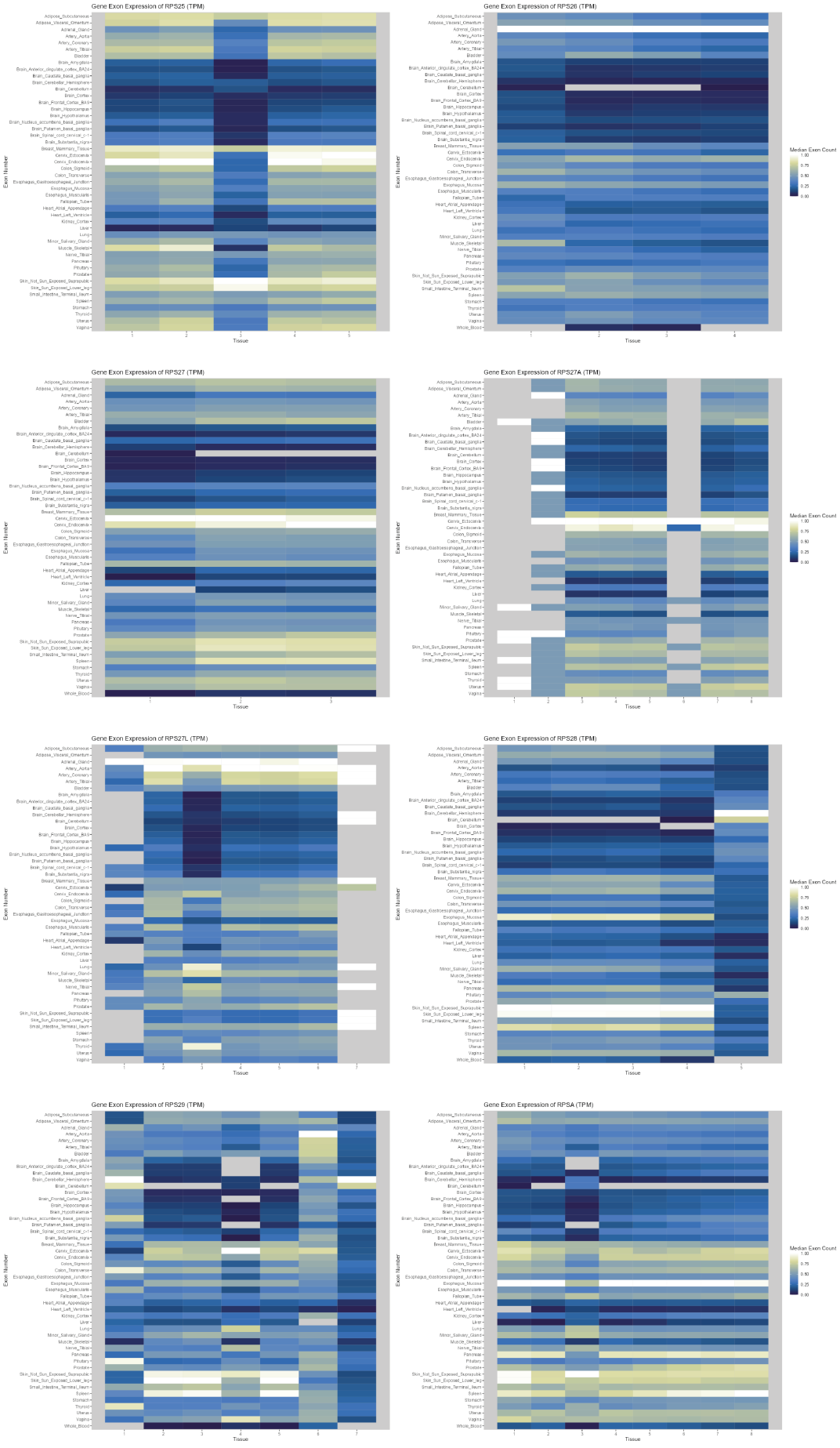

Gene Exon Expression of RPS5 (TPM)
Exon Number
Tissue


Gene Exon Expression of RPS6 (TPM)
Exon Number
Tissue
Median Exon Count


Gene Exon Expression of RPS7 (TPM)
Exon Number
Tissue


Gene Exon Expression of RPS8 (TPM)
Exon Number
Tissue
Median Exon Count


Gene Exon Expression of RPS9 (TPM)
Gene Exon Expression of RPS10 (TPM)
Gene Exon Expression of RPS11 (TPM)
Gene Exon Expression of RPS12 (TPM)
Gene Exon Expression of RPS13 (TPM)
Gene Exon Expression of RPS14 (TPM)
Gene Exon Expression of RPS15 (TPM)
Gene Exon Expression of RPS15A (TPM)
Exon Number
Tissue
Median Exon Count


Gene Exon Expression of RPS25 (TPM)
Gene Exon Expression of RPS26 (TPM)
Gene Exon Expression of RPS27 (TPM)
Gene Exon Expression of RPS27A (TPM)
Gene Exon Expression of RPS27L (TPM)
Gene Exon Expression of RPS28 (TPM)
Gene Exon Expression of RPS29 (TPM)
Gene Exon Expression of RPSA (TPM)
Exon Number
Tissue
Median Exon Count

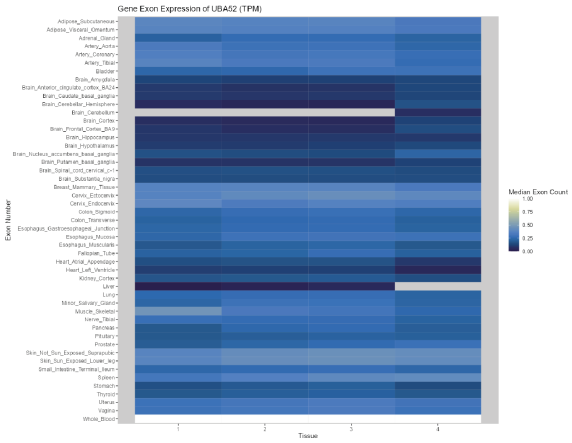


Figure 8.6: Exon Expression Heatmaps for Ribosome Protein Genes.

Exon Expression heatmaps were generated for the 89 ribosomal protein genes analysed (see Table 8.1). Heatmaps were created using R (version 4.2.0).

Exon Expression Analysis Code

*

June 23, 2023

The results below are generated from an R script.

```
# 07/05/20 Removed ovary and testes.
# 11/05/20 Automated plot generation.
# 20/07/20 Remove background, fixed y scales to go up by 1.

# Load in libraries

library("ggplot2")
library("rvest")
library("jsonlite")
library("tidyr")
library("tidyverse")
library("scico")
library("rlist")


# Read in list of genes, must be <150
proteins<- read.csv("Ribosomal protein on Gtex Final.csv")
protId<- proteins[[1]]
protIdList<- paste(as.character(protId),collapse="%2C")


# Web scrape data + make dataframe

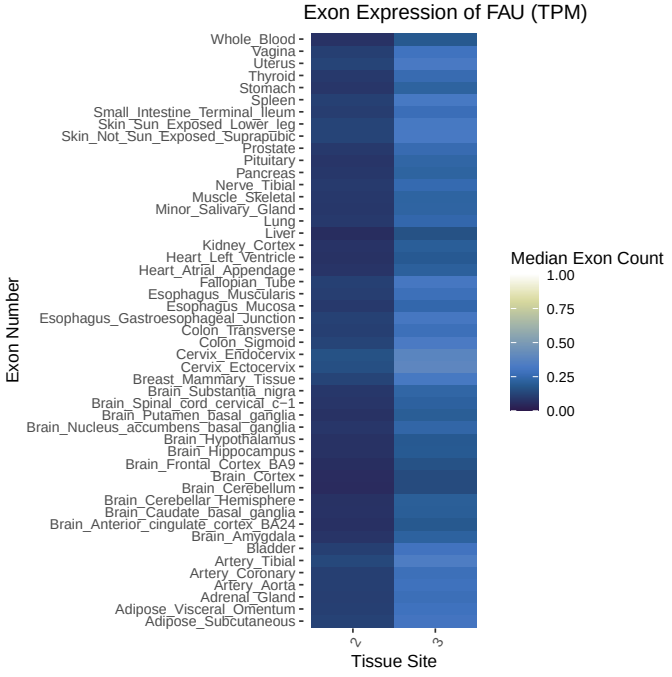
url <- paste("https://gtexportal.org/rest/v1/expression/medianExonExpression?datasetId=gtex_v8&gencodeId
df <- fromJSON(url(url), simplifyDataFrame = TRUE)
df <- unnest(data= df$medianExonExpression, cols=c()) %>%
  subset(select= c("median","tissueSiteDetailId","geneSymbol","exonId")) %>%
  mutate(exonId= str_sub(exonId, -2,-1)) %>%
  mutate(exonId= str_replace(exonId,"_","")) %>%
  mutate(exonId = as.numeric(exonId))


# Find max and min for exons
newdf<-data.frame(exon=integer(), min=double(), max=double())
for (exon in unique(df$exonId)){
  newdata <- subset(df, df$exonId==exon, select= c(median, exonId))
  max <- max(newdata$median, na.rm = TRUE)
  min <- min(newdata$median, na.rm = TRUE)
  newMax <- max - min
  output<- data.frame(exon=exon, min=min, max=max, new_max=newMax)
  newdf<- rbind(newdf, output)
}
```

*This report is automatically generated with the R package **knitr** (version 1.39).

```
#Normalise medians using min and max
for (exon in unique(df$exonId)){
  exonMin <- newdf[newdf$exon==exon, "min"]
  exonNewMax <- newdf[newdf$exon==exon, "new_max"]
  df$median[which(df$exonId==exon)] = (df$median[which(df$exonId==exon)] - exonMin) / exonNewMax
}
df$median[is.na(df$median)] <- 0  #remove NaN


#plot
ggplot(df[which(df$geneSymbol=="FAU" & df$median>0),], aes(y= tissueSiteDetailId, x=exonId)) +
         scale_x_continuous(breaks = seq(1, unique(exon), by = 1), labels = seq(1, unique(exon), by = 1)
         geom_tile(aes(fill=median)) +
         theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
         scale_fill_scico(palette="davos", direction=1, name="Median Exon Count", limits=c(0, 1))+
         labs(y="Exon Number", x = "Tissue Site") +
         ggtitle("Exon Expression of FAU (TPM)")+
         theme(panel.background = element_blank())
```




```
# Save plots
ggsave("test.png", width=13, height=10, dpi=100)


# Automated analysis

for (gene in protId){
url <- paste("https://gtexportal.org/rest/v1/expression/medianExonExpression?datasetId=gtex_v8&gencodeId
df <- fromJSON(url(url), simplifyDataFrame = TRUE)
df <- unnest(data= df$medianExonExpression, cols=c()) %>%
```

```
  subset(select= c("median","tissueSiteDetailId","geneSymbol","exonId")) %>%
  mutate(exonId= str_sub(exonId, -2,-1)) %>%
  mutate(tissueSiteDetailId=factor(tissueSiteDetailId,levels=rev(sort(unique(tissueSiteDetailId))))) %>%
  mutate(exonId= str_replace(exonId,"_","")) %>%
  mutate(exonId = as.numeric(exonId))

newdf<-data.frame(exon=integer(), min=double(), max=double())
for (exon in unique(df$exonId)){
  newdata <- subset(df, df$exonId==exon, select= c(median, exonId))
  max <- max(newdata$median, na.rm = TRUE)
  min <- min(newdata$median, na.rm = TRUE)
  newMax <- max - min
  output<- data.frame(exon=exon, min=min, max=max, new_max=newMax)
  newdf<- rbind(newdf, output)
}

for (exon in unique(df$exonId)){
  exonMin <- newdf[newdf$exon==exon, "min"]
  exonNewMax <- newdf[newdf$exon==exon, "new_max"]
  df$median[which(df$exonId==exon)] = (df$median[which(df$exonId==exon)] - exonMin) / exonNewMax
}
df$median[is.na(df$median)] <- 0  #remove NaN

symbol<- proteins[proteins$Gencode.ID == gene, 5]

# First line commented so pdf doesn't contain all graphs
#ggplot(df[which(df$median>0),], aes(y = tissueSiteDetailId, x=exonId)) +
  scale_x_continuous(breaks = seq(1, unique(exon), by = 1), labels = seq(1, unique(exon), by = 1), minor
  geom_tile(aes(fill=median)) +
  scale_fill_scico(palette="davos", direction=1, name="Median Exon Count", limits=c(0, 1))+
  labs(y="Exon Number", x = "Tissue") +
  ggtitle(paste("Gene Exon Expression of",symbol,"(TPM)",sep=" "))+
  theme(panel.grid = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        panel.background = element_rect(fill="gray80", color="gray80"),
        axis.text.x = element_text())

ggsave(paste("Exon Analysis",symbol,".png", sep=" "), width=13, height=10, dpi=100)
}

## Error:
## !  Cannot add ggproto objects together.  Did you forget to add this object to a ggplot object?
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
```

```
## locale:
## [1] LC_COLLATE=English_United Kingdom.utf8  LC_CTYPE=English_United Kingdom.utf8
## [3] LC_MONETARY=English_United Kingdom.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.39               ggcyto_1.24.0            flowWorkspace_4.8.0
##  [4] ncdfFlow_2.42.0          BH_1.78.0-0              RcppArmadillo_0.11.0.0.0
##  [7] flowCore_2.8.0           rlist_0.4.6.2            scico_1.3.1
## [10] forcats_0.5.1            stringr_1.4.0            dplyr_1.0.9
## [13] purrr_0.3.4              readr_2.1.2              tibble_3.1.7
## [16] tidyverse_1.3.2          tidyr_1.2.0              jsonlite_1.8.0
## [19] rvest_1.0.2              ggplot2_3.3.6
##
## loaded via a namespace (and not attached):
##  [1] matrixStats_0.62.0    fs_1.5.2              lubridate_1.8.0
##  [4] RColorBrewer_1.1-3    httr_1.4.3            Rgraphviz_2.40.0
##  [7] tools_4.2.0           backports_1.4.1       utf8_1.2.2
## [10] R6_2.5.1              DBI_1.1.2             BiocGenerics_0.42.0
## [13] colorspace_2.0-3      withr_2.5.0           tidyselect_1.1.2
## [16] gridExtra_2.3         curl_4.3.2            compiler_4.2.0
## [19] textshaping_0.3.6     graph_1.74.0          cli_3.3.0
## [22] Biobase_2.56.0        xml2_1.3.3            labeling_0.4.2
## [25] scales_1.2.0          hexbin_1.28.2         systemfonts_1.0.4
## [28] digest_0.6.29         base64enc_0.1-3       jpeg_0.1-9
## [31] pkgconfig_2.0.3       highr_0.9             dbplyr_2.1.1
## [34] rlang_1.0.2           readxl_1.4.0          rstudioapi_0.13
## [37] farver_2.1.0          generics_0.1.2        googlesheets4_1.0.0
## [40] magrittr_2.0.3        RProtoBufLib_2.8.0    aws.ec2metadata_0.2.0
## [43] Rcpp_1.0.8.3          munsell_0.5.0         S4Vectors_0.34.0
## [46] fansi_1.0.3           lifecycle_1.0.1       stringi_1.7.6
## [49] zlibbioc_1.42.0       plyr_1.8.7            grid_4.2.0
## [52] crayon_1.5.1          lattice_0.20-45       haven_2.5.0
## [55] hms_1.1.1             pillar_1.7.0          stats4_4.2.0
## [58] reprex_2.0.1          XML_3.99-0.9          glue_1.6.2
## [61] evaluate_0.16         latticeExtra_0.6-29   data.table_1.14.2
## [64] RcppParallel_5.1.5    modelr_0.1.8          png_0.1-7
## [67] vctrs_0.4.1           tzdb_0.3.0            cellranger_1.1.0
## [70] gtable_0.3.0          aws.s3_0.3.21         assertthat_0.2.1
## [73] xfun_0.31             broom_0.8.0           ragg_1.2.2
## [76] googledrive_2.0.0     gargle_1.2.0          tinytex_0.38
## [79] cytolib_2.8.0         aws.signature_0.6.0   ellipsis_0.3.2

Sys.time()
```

Figure 8.7: R Script for Exon Expression Heatmaps for Ribosome Protein Genes.

Exon Expression heatmaps were generated for the 89 ribosomal protein genes analysed (see Table 8.1). Heatmaps were created using R (version 4.2.0).

Table 8.2: snoRNA Genes Used in Gene Expression Analysis.

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORA1	ENSG00000199405.1	109617001	chr8:55902723-55902855:-
SNORA1	ENSG00000206834.1	677792	chr11:93732004-93732133:-
SNORA2A	ENSG00000206612.1	677793	chr12:48656648-48656782:-
SNORA2B	ENSG00000207313.1	677794	chr12:48667457-48667593:-
SNORA2C	ENSG00000221491.2	677815	chr12:48654382-48654518:-
SNORA3	ENSG00000221719.1	109616997	chr16:2796408-2796532:-
SNORA3A	ENSG00000200983.1	619562	chr11:8684227-8684356:+
SNORA3B	ENSG00000212607.1	677826	chr11:8685439-8685569:+
SNORA4	ENSG00000263776.1	619568	chr3:186787612-186787749:+
SNORA5A	ENSG00000206838.1	654319	chr7:45104349-45104482:-
SNORA5B	ENSG00000200656.1	677795	chr7:45105968-45106099:-
SNORA5C	ENSG00000201772.1	677796	chr7:45104906-45105042:-
SNORA6	ENSG00000206760.1	574040	chr3:39408389-39408539:+
SNORA7A	ENSG00000207496.1	619563	chr3:12840312-12840450:-
SNORA7B	ENSG00000207088.1	677797	chr3:129397210-129397348:-
SNORA8	ENSG00000207304.1	654320	chr11:93732361-93732499:-
SNORA9	ENSG00000206897.1	109616977	chr12:123616708-123616840:+
SNORA9	ENSG00000277184.1	677798	chr7:44985378-44985510:-
SNORA10	ENSG00000206811.1	574042	chr16:1962334-1962466:-
SNORA11	ENSG00000221164.1	109617000	chr10:73126080-73126207:-
SNORA11	ENSG00000221716.1	677799	chrX:54814370-54814497:+
SNORA11E	ENSG00000221705.1	101340250	chrX:52063347-52063474:-
SNORA11D	ENSG00000221475.1	100124541	chrX:52190621-52190748:+
SNORA11C	ENSG00000221459.1	100124540	chrX:47388649-47388777:+
SNORA11B	ENSG00000221102.1	100124539	chr14:91126425-91126553:+
SNORA12	ENSG00000212464.1	677800	chr10:100237156-100237302:-
SNORA13	ENSG00000238363.1	654322	chr5:112161485-112161617:+
SNORA14A	ENSG00000201643.1	677801	chr7:75943782-75943916:+
SNORA14B	ENSG00000207181.1	677802	chr1:235127803-235127937:-
SNORA15	ENSG00000207168.1	677803	chr7:56060470-56060602:+
SNORA15	ENSG00000207062.1	109616962	chr7:65070538-65070672:+
SNORA15	ENSG00000206785.1	109616962	chr7:65760052-65760186:+

continued . . .

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORA16A	ENSG00000280498.1	692073	chr1:28580920-28581054:-
SNORA16B	ENSG00000201544.1	692157	chr1:212352816-212352950:+
SNORA17A	ENSG00000274998.1	677804	chr9:136726747-136726879:-
SNORA17B	ENSG00000280496.1	677824	chr9:136726104-136726239:-
SNORA18	ENSG00000201544.1	692157	chr1:212352816-212352950:+
SNORA19	ENSG00000207468.1	641451	chr10:119060011-119060138:-
SNORA20	ENSG00000207392.1	677806	chr6:159780250-159780381:-
SNORA20	ENSG00000206947.1	109616963	chr7:39329003-39329134:-
SNORA21	ENSG00000252699.1	109617002	chr17:38851524-38851659:-
SNORA21	ENSG00000199293.1	619505	chr17:38852863-38852994:-
SNORA22	ENSG00000206603.1	109616964	chr7:56055365-56055502:+
SNORA22	ENSG00000207344.1	109616965	chr7:65065999-65066132:+
SNORA22	ENSG00000206634.1	677807	chr7:65755526-65755659:+
SNORA23	ENSG00000201998.1	677808	chr11:9428773-9428954:+
SNORA24	ENSG00000275994.1	677809	chr4:118279190-118279320:+
SNORA25	ENSG00000202377.1	109623459	chr7:115581315-115581442:-
SNORA25	ENSG00000207112.1	684959	chr11:93730513-93730646:-
SNORA26	ENSG00000212588.1	677810	chr4:52713249-52713370:+
SNORA27	ENSG00000207051.1	619499	chr13:27255401-27255526:+
SNORA28	ENSG00000272533.1	677811	chr14:103337849-103337974:+
SNORA29	ENSG00000206910.1	677812	chr6:159785594-159785733:-
SNORA30	ENSG00000206755.1	677813	chr16:30710537-30710665:+
SNORA30	ENSG00000202189.1	109616992	chr9:20786927-20787055:+
SNORA31	ENSG00000199477.1	677814	chr13:45337480-45337609:-
SNORA31	ENSG00000253051.1	109616966	chr13:45336314-45336447:-
SNORA32	ENSG00000206799.1	692063	chr11:93730979-93731099:-
SNORA33	ENSG00000200534.1	594839	chr6:132817219-132817348:+
SNORA35	ENSG00000208839.1	677816	chrX:114630797-114630924:+
SNORA35	ENSG00000271907.1	109617004	chrX:115126316-115126444:-
SNORA36A	ENSG00000206948.1	677817	chrX:154768528-154768659:+
SNORA36B	ENSG00000222370.1	677818	chr1:220200546-220200676:-
SNORA36C	ENSG00000207016.1	100124535	chr2:69520043-69520174:-
SNORA37	ENSG00000207233.1	677819	chr18:54222284-54222412:-

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORA38	ENSG00000200816.1	677820	chr6:31623079-31623210:+
SNORA38B	ENSG00000200394.1	100124536	chr17:67740669-67740799:+
SNORA40	ENSG00000210825.1	677822	chr11:93735111-93735236:-
SNORA40	ENSG00000212587.1	109616968	chr6:110848546-110848673:+
SNORA41	ENSG00000207406.1	619569	chr2:206162228-206162359:+
SNORA42	ENSG00000207217.1	109616999	chr7:6016877-6017011:+
SNORA44	ENSG00000273544.1	677825	chr1:28580381-28580512:-
SNORA46	ENSG00000207493.1	677827	chr16:58548499-58548633:-
SNORA47	ENSG00000238961.1	677828	chr5:77080434-77080571:-
SNORA48	ENSG00000209582.1	652965	chr17:7574713-7574847:+
SNORA49	ENSG00000208892.1	677829	chr12:132031224-132031359:+
SNORA50	ENSG00000220986.1	109616958	chr5:69160806-69160939:+
SNORA50	ENSG00000281295.1	109616998	chr22:33704786-33704922:-
SNORA50A	ENSG00000281910.1	677830	chr16:58559796-58559931:-
SNORA50C	ENSG00000277887.1	677842	chr17:64146339-64146471:+
SNORA51	ENSG00000271798.1	677831	chr20:2655067-2655198:+
SNORA52	ENSG00000199785.1	619565	chr11:811681-811814:+
SNORA53	ENSG00000212443.1	677832	chr12:98599635-98599883:+
SNORA54	ENSG00000207008.1	677833	chr11:2963771-2963893:-
SNORA55	ENSG00000201457.1	677834	chr1:39567374-39567508:-
SNORA56	ENSG00000206693.1	677835	chrX:154774998-154775126:+
SNORA57	ENSG00000206597.1	692158	chr11:62665422-62665570:+
SNORA58	ENSG00000249020.1	677836	chr3:131479097-131479233:-
SNORA59A	ENSG00000239149.1	677885	chr1:12507246-12507397:+
SNORA59B	ENSG00000266079.4	677882	chr17:19557211-19557911:+
SNORA60	ENSG00000199266.1	677837	chr20:38449370-38449503:+
SNORA61	ENSG00000278274.1	677838	chr1:28579764-28579893:-
SNORA62	ENSG00000202363.1	6044	chr3:39411054-39411206:+
SNORA63	ENSG00000201448.1	109616970	chr1:36418450-36418578:-
SNORA63	ENSG00000201229.1	109616971	chr3:183451857-183451988:+
SNORA63	ENSG00000200320.1	6043	chr3:186787300-186787431:+
SNORA64	ENSG00000207405.1	26784	chr16:1962973-1963106:-
SNORA64	ENSG00000207187.1	109616960	chr2:30187434-30187566:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORA65	ENSG00000201302.1	26783	chr9:127448501-127448630:-
SNORA66	ENSG00000207523.1	26782	chr1:92840719-92840851:+
SNORA67	ENSG00000277985.1	26781	chr17:7577955-7578091:+
SNORA68	ENSG00000207166.1	26780	chr19:17862588-17862720:+
SNORA68	ENSG00000201388.1	109616972	chr19:32608337-32608469:-
SNORA69	ENSG00000206622.1	26779	chrX:119787353-119787484:-
SNORA70	ENSG00000207274.1	109617008	chr2:214846947-214847081:+
SNORA70	ENSG00000207165.1	26778	chrX:154400281-154400415:+
SNORA70	ENSG00000206909.1	109616973	chr5:171366118-171366252:+
SNORA70B	ENSG00000206937.1	100124537	chr2:61417244-61417378:-
SNORA70C	ENSG00000207268.1	100124538	chr9:117181066-117181200:-
SNORA70D	ENSG00000263666.1	100379141	chr16:71698567-71698701:-
SNORA70E	ENSG00000207221.1	100379250	chr11:83041464-83041598:-
SNORA70F	ENSG00000206869.1	100337591	chr2:164687643-164687777:-
SNORA70G	ENSG00000206650.1	100379132	chr12:68627234-68627375:-
SNORA71A	ENSG00000225091.3	26777	chr20:38427309-38427442:-
SNORA71B	ENSG00000235408.5	26776	chr20:38425083-38425354:-
SNORA71C	ENSG00000201512.1	677839	chr20:38429670-38429803:-
SNORA71D	ENSG00000200354.1	677840	chr20:38433865-38433998:-
SNORA71E	ENSG00000274309.1	677821	chr20:38448084-38448215:+
SNORA72	ENSG00000207067.1	26775	chr8:98042086-98042217:-
SNORA73A	ENSG00000274266.1	6080	chr1:28507366-28507571:+
SNORA73B	ENSG00000200087.1	26768	chr1:28508559-28508762:+
SNORA74	ENSG00000265733.1	109616974	chr10:45984610-45984810:+
SNORA74A	ENSG00000200959.1	26821	chr5:139278781-139278978:+
SNORA74B	ENSG00000212402.1	677841	chr5:173020728-173020928:+
SNORA75	ENSG00000206885.1	654321	chr2:231455800-231455936:-
SNORA75	ENSG00000206780.1	109616994	chr4:17320746-17320882:-
SNORA76	ENSG00000253007.2	109616998	chr22:33704786-33704920:-
SNORA76	ENSG00000280894.1	109616958	chr5:69160808-69160939:+
SNORA77	ENSG00000221643.1	677843	chr1:203729581-203729705:+
SNORA78	ENSG00000273587.1	677844	chr16:1965184-1965310:+
SNORA79	ENSG00000221303.1	677845	chr14:81202695-81202834:-

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORA80A	ENSG00000200792.1	677846	chr21:32377187-32377322:-
SNORA80B	ENSG00000206633.1	100302743	chr2:10446714-10446849:-
SNORA80E	ENSG00000207475.1	677823	chr1:155919909-155920045:-
SNORA81	ENSG00000221420.2	677847	chr3:186786675-186786852:+
SNORA84	ENSG00000239183.1	100124534	chr9:92292461-92292593:-
SNORA108	ENSG00000252284.1	106635543	chr18:58600631-58600698:+
SNORD1A	ENSG00000278261.1	677848	chr17:76561634-76561705:+
SNORD1B	ENSG00000199961.1	677849	chr17:76561109-76561192:+
SNORD1C	ENSG00000274091.1	677850	chr17:76558791-76558868:+
SNORD2	ENSG00000238942.1	619567	chr3:186784796-186784864:+
SNORD3A	ENSG00000263934.4	780851	chr17:19188016-19188714:+
SNORD3B-1	ENSG00000265185.5	26851	chr17:19061912-19062669:+
SNORD3B-2	ENSG00000262074.6	780852	chr17:19063346-19064136:-
SNORD3C	ENSG00000264940.4	780853	chr17:19189665-19190245:-
SNORD3D	ENSG00000281000.1	780854	chr17:19112420-19112636:-
SNORD5	ENSG00000239195.1	692072	chr11:93733228-93733300:-
SNORD6	ENSG00000202314.1	692075	chr11:93731502-93731574:-
SNORD7	ENSG00000207297.1	692076	chr17:35573657-35573753:+
SNORD8	ENSG00000200785.1	319103	chr14:21397292-21397401:-
SNORD9	ENSG00000199436.1	692053	chr14:21392150-21392253:-
SNORD10	ENSG00000238917.1	652966	chr17:7576811-7576952:+
SNORD11	ENSG00000238317.2	692058	chr2:202293051-202293134:+
SNORD118	ENSG00000200463.1	727676	chr17:8173454-8173587:-
SNORD11B	ENSG00000271852.1	100113392	chr2:202291332-202291421:+
SNORD12	ENSG00000212304.1	692057	chr20:49280683-49280772:+
SNORD12B	ENSG00000222365.1	100113393	chr20:49280319-49280409:+
SNORD12C	ENSG00000209042.1	26765	chr20:49278940-49279028:+
SNORD13	ENSG00000239039.1	692084	chr8:33513475-33513578:+
SNORD14A	ENSG00000272034.1	26822	chr11:17074654-17074744:-
SNORD14B	ENSG00000201403.1	85388	chr11:17075779-17075868:-
SNORD14C	ENSG00000202252.1	85389	chr11:123059335-123059422:-
SNORD14D	ENSG00000207118.1	85390	chr11:123058909-123058995:-
SNORD14E	ENSG00000200879.1	85391	chr11:123058077-123058161:-

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD15A	ENSG00000206941.1	6079	chr11:75400391-75400538:+
SNORD15B	ENSG00000207445.1	114599	chr11:75404421-75404566:+
SNORD16	ENSG00000199673.1	595097	chr15:66502812-66502910:-
SNORD17	ENSG00000212232.1	692086	chr20:17962710-17962946:-
SNORD18A	ENSG00000200623.1	595098	chr15:66503243-66503314:-
SNORD18B	ENSG00000202529.1	595099	chr15:66502020-66502091:-
SNORD18C	ENSG00000199574.1	595100	chr15:66501250-66501318:-
SNORD19	ENSG00000212493.1	692089	chr3:52689248-52689315:+
SNORD19B	ENSG00000238862.1	100113381	chr3:52690744-52690827:+
SNORD20	ENSG00000207280.1	6082	chr2:231456444-231456523:-
SNORD21	ENSG00000206680.1	6083	chr1:92837289-92837383:+
SNORD22	ENSG00000277194.1	9304	chr11:62852910-62853035:-
SNORD24	ENSG00000206611.1	26820	chr9:133349396-133349470:+
SNORD25	ENSG00000275043.1	9303	chr11:62855564-62855632:-
SNORD26	ENSG00000276788.1	9302	chr11:62855292-62855366:-
SNORD27	ENSG00000275996.1	9301	chr11:62855012-62855083:-
SNORD28	ENSG00000212295.1	109617011	chr6:156378750-156378824:+
SNORD28	ENSG00000274544.1	9300	chr11:62854621-62854695:-
SNORD30	ENSG00000277846.1	9299	chr11:62853663-62853732:-
SNORD31	ENSG00000201847.1	109616978	chr13:107320895-107320963:-
SNORD32A	ENSG00000201675.1	26819	chr19:49489965-49490048:+
SNORD32B	ENSG00000201330.1	692092	chr6:29582249-29582332:+
SNORD33	ENSG00000199631.1	26818	chr19:49490615-49490699:+
SNORD34	ENSG00000202503.1	26817	chr19:49490904-49490974:+
SNORD35A	ENSG00000200259.1	26816	chr19:49491175-49491260:+
SNORD35B	ENSG00000200530.1	84546	chr19:49497720-49497806:+
SNORD36A	ENSG00000199744.1	26815	chr9:133350456-133350528:+
SNORD36B	ENSG00000200831.1	26814	chr9:133350095-133350168:+
SNORD36C	ENSG00000252542.1	26813	chr9:133350847-133350912:+
SNORD37	ENSG00000206775.1	26812	chr19:3982507-3982572:-
SNORD38	ENSG00000207199.1	109616979	chr8:39018615-39018683:-
SNORD38A	ENSG00000202031.1	94162	chr1:44777843-44777912:+
SNORD38B	ENSG00000281859.1	94163	chr1:44778390-44778458:+

continued . . .

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD41	ENSG00000209702.1	26810	chr19:12706449-12706518:-
SNORD42A	ENSG00000238649.1	26809	chr17:28723429-28723492:+
SNORD42B	ENSG00000238423.1	26808	chr17:28720550-28720616:+
SNORD43	ENSG00000263764.1	26807	chr22:39319050-39319113:-
SNORD45A	ENSG00000207241.1	26805	chr1:75787889-75787972:+
SNORD45B	ENSG00000201487.1	26804	chr1:75789477-75789548:+
SNORD45C	ENSG00000206620.1	692085	chr1:75787072-75787150:+
SNORD46	ENSG00000200913.1	94161	chr1:44776490-44776593:+
SNORD48	ENSG00000201823.1	26801	chr6:31835263-31835326:+
SNORD49A	ENSG00000277370.1	26800	chr17:16440036-16440106:+
SNORD49B	ENSG00000277108.1	692087	chr17:16439505-16439576:+
SNORD4A	ENSG00000238578.1	26773	chr17:28722582-28722653:+
SNORD4B	ENSG00000238597.1	26772	chr17:28723682-28723753:+
SNORD50B	ENSG00000275072.1	692088	chr6:85677589-85677658:-
SNORD51	ENSG00000207047.2	26798	chr2:206161878-206161957:+
SNORD52	ENSG00000201754.1	26797	chr6:31837076-31837142:+
SNORD53	ENSG00000265145.1	26796	chr2:28927067-28927144:+
SNORD53-SNORD92	ENSG00000265706.1	109617012	chr2:28927983-28928060:+
SNORD54	ENSG00000238650.1	26795	chr8:56073835-56073901:-
SNORD55	ENSG00000264294.1	26811	chr1:44775864-44775943:+
SNORD56	ENSG00000229686.1	26793	chr20:2656624-2656694:+
SNORD56B	ENSG00000207444.1	319139	chr14:71398337-71398407:+
SNORD57	ENSG00000226572.1	26792	chr20:2656939-2657010:+
SNORD58A	ENSG00000206602.1	26791	chr18:49491283-49491347:-
SNORD58B	ENSG00000271982.1	26790	chr18:49491664-49491729:-
SNORD58C	ENSG00000202093.1	100124516	chr18:49489245-49489308:-
SNORD59A	ENSG00000207031.1	26789	chr12:56645027-56645101:-
SNORD60	ENSG00000206630.1	26788	chr16:2155023-2155105:-
SNORD61	ENSG00000206979.1	26787	chrX:136879199-136879271:-
SNORD62A	ENSG00000235284.1	26786	chr9:131485665-131485750:+
SNORD62B	ENSG00000231587.1	692093	chr9:131490486-131490571:+
SNORD63	ENSG00000206989.1	26785	chr5:138561043-138561110:-
SNORD64	ENSG00000276610.1	347686	chr15:24985100-24985166:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD65	ENSG00000238936.1	109616980	chr8:41426655-41426727:+
SNORD65	ENSG00000277512.1	692106	chr17:16441226-16441298:+
SNORD66	ENSG00000212158.1	692107	chr3:184325696-184325771:+
SNORD67	ENSG00000212135.1	692108	chr11:46762389-46762499:-
SNORD68	ENSG00000200084.1	606500	chr16:89561434-89561517:+
SNORD69	ENSG00000212452.1	692109	chr3:52692736-52692812:+
SNORD70	ENSG00000212534.1	692110	chr2:202276431-202276518:+
SNORD71	ENSG00000223224.1	692111	chr16:71758402-71758487:-
SNORD72	ENSG00000212296.1	619564	chr5:40832656-40832735:-
SNORD73A	ENSG00000208797.1	8944	chr4:151103827-151103891:+
SNORD73B	ENSG00000201264.1	114655	chr4:151102057-151102128:+
SNORD74	ENSG00000201025.1	109617014	chr21:16284696-16284768:-
SNORD77	ENSG00000212414.1	109616982	chr8:100004583-100004649:+
SNORD82	ENSG00000202400.1	25826	chr2:231460371-231460440:-
SNORD83A	ENSG00000209482.1	116937	chr22:39315213-39315307:-
SNORD83B	ENSG00000209480.1	116938	chr22:39313819-39313911:-
SNORD84	ENSG00000265236.1	692199	chr6:31541101-31541178:-
SNORD86	ENSG00000212498.1	692201	chr20:2656097-2656182:+
SNORD87	ENSG00000254341.2	641648	chr8:66922467-66922555:-
SNORD88A	ENSG00000221241.1	692202	chr19:50799442-50799532:-
SNORD88B	ENSG00000221381.1	692203	chr19:50799032-50799122:-
SNORD88C	ENSG00000220988.1	692204	chr19:50802328-50802418:-
SNORD89	ENSG00000212283.1	692205	chr2:101272936-101273049:-
SNORD90	ENSG00000212447.1	692206	chr9:122880213-122880323:-
SNORD91A	ENSG00000212163.5	692207	chr17:2330180-2330370:-
SNORD91B	ENSG00000275084.3	692208	chr17:2329016-2329237:-
SNORD92	ENSG00000264994.1	692209	chr2:28913664-28913748:+
SNORD93	ENSG00000221740.1	692210	chr7:22856613-22856686:+
SNORD94	ENSG00000208772.1	692225	chr2:86135870-86136006:+
SNORD95	ENSG00000264549.1	619570	chr5:181243312-181243379:-
SNORD96A	ENSG00000272296.1	619571	chr5:181241814-181241892:-
SNORD96B	ENSG00000208883.1	692226	chrX:110224985-110225063:-
SNORD97	ENSG00000238622.1	692223	chr11:10801467-10801608:-

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD98	ENSG00000283551.1	692211	chr10:68755172-68755238:+
SNORD99	ENSG00000221539.1	692212	chr1:28578749-28578822:-
SNORD100	ENSG00000221500.1	594838	chr6:132816802-132816877:+
SNORD101	ENSG00000206754.1	594837	chr6:132815307-132815379:+
SNORD102	ENSG00000207500.1	26771	chr13:27255064-27255135:+
SNORD104	ENSG00000199753.1	692227	chr17:64146083-64146152:+
SNORD105	ENSG00000209645.1	692229	chr19:10107651-10107735:+
SNORD105B	ENSG00000238531.1	100113382	chr19:10109757-10109835:+
SNORD107	ENSG00000276314.1	91380	chr15:24981994-24982068:+
SNORD108	ENSG00000239014.1	338427	chr15:24986925-24986995:+
SNORD109A	ENSG00000274640.1	338428	chr15:25041974-25042040:+
SNORD109B	ENSG00000239169.1	338429	chr15:25278343-25278409:+
SNORD110	ENSG00000221116.1	692213	chr20:2654212-2654286:+
SNORD111	ENSG00000221066.1	692214	chr16:70538005-70538098:+
SNORD111B	ENSG00000221514.1	100113402	chr16:70529509-70529588:+
SNORD112	ENSG00000275662.1	692215	chr14:100897920-100897996:+
SNORD113-1	ENSG00000202191.1	767561	chr14:100924824-100924892:+
SNORD113-2	ENSG00000212384.1	767562	chr14:100927342-100927413:+
SNORD113-3	ENSG00000201700.3	767563	chr14:100929919-100929990:+
SNORD113-4	ENSG00000201672.1	767564	chr14:100936491-100936565:+
SNORD113-5	ENSG00000272474.1	767565	chr14:100938187-100938264:+
SNORD113-6	ENSG00000200215.3	767566	chr14:100939556-100939630:+
SNORD113-7	ENSG00000200632.1	767567	chr14:100941126-100941202:+
SNORD113-8	ENSG00000200367.1	767568	chr14:100943451-100943524:+
SNORD113-9	ENSG00000201950.1	767569	chr14:100945649-100945720:+
SNORD114-1	ENSG00000199575.1	767577	chr14:100949833-100949904:+
SNORD114-2	ENSG00000200823.1	767578	chr14:100951856-100951933:+
SNORD114-3	ENSG00000201839.1	767579	chr14:100953349-100953423:+
SNORD114-4	ENSG00000200832.1	767580	chr14:100954374-100954448:+
SNORD114-5	ENSG00000199798.1	767581	chr14:100955370-100955439:+
SNORD114-6	ENSG00000201263.1	767582	chr14:100957166-100957237:+
SNORD114-7	ENSG00000199390.1	767583	chr14:100963054-100963130:+
SNORD114-9	ENSG00000201240.1	767585	chr14:100966029-100966100:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD114-10	ENSG00000200279.1	767588	chr14:100967052-100967123:+
SNORD114-11	ENSG00000200608.1	767589	chr14:100968111-100968185:+
SNORD114-12	ENSG00000202270.1	767590	chr14:100968948-100969022:+
SNORD114-13	ENSG00000201247.1	767591	chr14:100969879-100969952:+
SNORD114-14	ENSG00000199593.1	767592	chr14:100972103-100972177:+
SNORD114-15	ENSG00000201557.1	767593	chr14:100972670-100972741:+
SNORD114-16	ENSG00000199914.1	767594	chr14:100973595-100973664:+
SNORD114-17	ENSG00000201569.1	767595	chr14:100974806-100974880:+
SNORD114-18	ENSG00000202142.1	767596	chr14:100975825-100975896:+
SNORD114-19	ENSG00000199942.1	767597	chr14:100976477-100976551:+
SNORD114-20	ENSG00000202048.1	767598	chr14:100981004-100981075:+
SNORD114-21	ENSG00000272344.1	767599	chr14:100981975-100982046:+
SNORD114-22	ENSG00000202293.1	767600	chr14:100982926-100982997:+
SNORD114-23	ENSG00000200406.1	767603	chr14:100983876-100983947:+
SNORD114-24	ENSG00000201899.1	767604	chr14:100984777-100984848:+
SNORD114-25	ENSG00000200612.1	767605	chr14:100986057-100986128:+
SNORD114-26	ENSG00000200413.1	767606	chr14:100987046-100987117:+
SNORD114-27	ENSG00000200636.1	767608	chr14:100988161-100988230:+
SNORD114-28	ENSG00000200480.1	767609	chr14:100989130-100989201:+
SNORD114-29	ENSG00000201689.1	767610	chr14:100990091-100990160:+
SNORD114-30	ENSG00000201318.1	767611	chr14:100991919-100991990:+
SNORD114-31	ENSG00000200089.1	767612	chr14:100993236-100993310:+
SNORD115-1	ENSG00000201831.1	338433	chr15:25170723-25170804:+
SNORD115-2	ENSG00000199712.1	100033437	chr15:25172635-25172716:+
SNORD115-3	ENSG00000199970.1	100033440	chr15:25174927-25175008:+
SNORD115-4	ENSG00000200680.1	100033441	chr15:25176832-25176913:+
SNORD115-5	ENSG00000200503.1	100033442	chr15:25178738-25178819:+
SNORD115-6	ENSG00000200812.1	100033443	chr15:25180497-25180578:+
SNORD115-7	ENSG00000278089.1	100033444	chr15:25182385-25182466:+
SNORD115-8	ENSG00000200726.1	100033445	chr15:25184306-25184387:+
SNORD115-9	ENSG00000199782.1	100033446	chr15:25185631-25185712:+
SNORD115-10	ENSG00000201943.1	100033447	chr15:25187536-25187616:+
SNORD115-11	ENSG00000200486.1	100033448	chr15:25189414-25189495:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD115-12	ENSG00000199453.1	100033449	chr15:25191416-25191497:+
SNORD115-13	ENSG00000273835.1	100033450	chr15:25193321-25193402:+
SNORD115-14	ENSG00000199960.1	100033451	chr15:25194921-25195001:+
SNORD115-15	ENSG00000201679.1	100033453	chr15:25197576-25197656:+
SNORD115-16	ENSG00000200757.1	100033454	chr15:25199448-25199529:+
SNORD115-17	ENSG00000201482.1	100033455	chr15:25201323-25201404:+
SNORD115-18	ENSG00000200163.1	100033456	chr15:25203227-25203308:+
SNORD115-19	ENSG00000199968.1	100033458	chr15:25204357-25204438:+
SNORD115-20	ENSG00000201969.1	100033460	chr15:25206262-25206343:+
SNORD115-21	ENSG00000199833.1	100033603	chr15:25208083-25208163:+
SNORD115-22	ENSG00000201326.1	100033799	chr15:25209918-25209999:+
SNORD115-23	ENSG00000201331.1	100033800	chr15:25211796-25211877:+
SNORD115-24	ENSG00000200398.1	100036563	chr15:25213648-25213729:+
SNORD115-25	ENSG00000199489.1	100033801	chr15:25215541-25215622:+
SNORD115-26	ENSG00000275524.1	100033802	chr15:25218617-25218698:+
SNORD115-27	ENSG00000201300.1	100036564	chr15:25220497-25220578:+
SNORD115-28	ENSG00000200801.1	100036565	chr15:25222348-25222428:+
SNORD115-29	ENSG00000199704.1	100033803	chr15:25223246-25223327:+
SNORD115-30	ENSG00000200987.1	100033804	chr15:25225203-25225284:+
SNORD115-31	ENSG00000202188.1	100033805	chr15:25227109-25227190:+
SNORD115-32	ENSG00000200949.1	100033806	chr15:25228967-25229048:+
SNORD115-33	ENSG00000200593.1	100033807	chr15:25230838-25230919:+
SNORD115-34	ENSG00000199311.1	100033808	chr15:25232387-25232468:+
SNORD115-35	ENSG00000201992.1	100033809	chr15:25234247-25234328:+
SNORD115-36	ENSG00000202499.1	100033810	chr15:25236085-25236166:+
SNORD115-37	ENSG00000200638.1	100033811	chr15:25237986-25238067:+
SNORD115-38	ENSG00000201907.1	100033812	chr15:25239838-25239919:+
SNORD115-39	ENSG00000200564.1	100033813	chr15:25241746-25241827:+
SNORD115-40	ENSG00000272460.1	100033814	chr15:25243614-25243695:+
SNORD115-41	ENSG00000200478.1	100033815	chr15:25245478-25245559:+
SNORD115-42	ENSG00000201143.1	100033816	chr15:25247345-25247426:+
SNORD115-43	ENSG00000202373.1	100033817	chr15:25249198-25249279:+
SNORD115-44	ENSG00000202261.1	100033818	chr15:25250859-25250940:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD115-45	ENSG00000212380.1	100036566	chr15:25264510-25264590:+
SNORD115-46	ENSG00000276844.1	100873857	chr15:25266591-25266661:+
SNORD115-48	ENSG00000201634.1	100033822	chr15:25269783-25269858:+
SNORD116-1	ENSG00000207063.1	100033413	chr15:25051477-25051571:+
SNORD116-2	ENSG00000207001.1	100033414	chr15:25054210-25054304:+
SNORD116-3	ENSG00000207014.1	100033415	chr15:25056860-25056954:+
SNORD116-4	ENSG00000275529.1	100033416	chr15:25059538-25059633:+
SNORD116-5	ENSG00000207191.1	100033417	chr15:25062333-25062427:+
SNORD116-6	ENSG00000207442.1	100033418	chr15:25065026-25065121:+
SNORD116-7	ENSG00000207133.1	100033419	chr15:25067788-25067882:+
SNORD116-8	ENSG00000207093.1	100033420	chr15:25070432-25070526:+
SNORD116-9	ENSG00000206727.1	100033421	chr15:25073107-25073201:+
SNORD116-10	ENSG00000200661.1	100033422	chr15:25074114-25074215:+
SNORD116-11	ENSG00000206609.1	100033423	chr15:25075929-25076020:+
SNORD116-12	ENSG00000207197.1	100033424	chr15:25077051-25077142:+
SNORD116-13	ENSG00000207137.1	100033425	chr15:25079058-25079149:+
SNORD116-14	ENSG00000206621.1	100033426	chr15:25080142-25080233:+
SNORD116-15	ENSG00000207174.1	100033427	chr15:25081287-25081378:+
SNORD116-16	ENSG00000207263.1	100033428	chr15:25082768-25082859:+
SNORD116-17	ENSG00000206656.1	100033429	chr15:25083588-25083679:+
SNORD116-18	ENSG00000206688.1	100033430	chr15:25085385-25085476:+
SNORD116-19	ENSG00000207460.1	727708	chr15:25086527-25086618:+
SNORD116-20	ENSG00000278715.1	100033431	chr15:25087662-25087753:+
SNORD116-21	ENSG00000277785.1	100033432	chr15:25088804-25088895:+
SNORD116-22	ENSG00000275127.1	100033433	chr15:25089923-25090014:+
SNORD116-23	ENSG00000207375.1	100033434	chr15:25091786-25091877:+
SNORD116-24	ENSG00000207279.1	100033435	chr15:25094037-25094128:+
SNORD116-25	ENSG00000252326.1	100033436	chr15:25097663-25097754:+
SNORD116-26	ENSG00000251815.1	100033438	chr15:25099499-25099594:+
SNORD116-27	ENSG00000251896.1	100033439	chr15:25101575-25101666:+
SNORD116-28	ENSG00000278123.1	100033820	chr15:25104642-25104732:+
SNORD116-29	ENSG00000207245.1	100033821	chr15:25106521-25106603:+
SNORD116-30	ENSG00000252277.1	100873856	chr15:25108268-25108352:+

continued ...

Gene Symbol	Gencode ID	Entrez ID	Genomic Location
SNORD117	ENSG00000201785.1	692233	chr6:31536374-31536449:-
SNORD121A	ENSG00000238886.1	100113379	chr9:33952769-33952850:-
SNORD121B	ENSG00000238300.1	101340252	chr9:33934296-33934376:-
SNORD123	ENSG00000239112.1	100113384	chr5:9548836-9548905:+
SNORD124	ENSG00000238793.1	101340251	chr17:40027542-40027645:-
SNORD125	ENSG00000239127.1	100113380	chr22:29333163-29333258:-
SNORD126	ENSG00000238344.1	100113391	chr14:20326450-20326526:-
SNORD127	ENSG00000239043.1	100113389	chr14:45110883-45110968:+
SCARNA1	ENSG00000252947.1	677774	chr1:27834401-27834566:+
SCARNA2	ENSG00000278249.1	677766	chr1:109100193-109100612:+
SCARNA3	ENSG00000252906.1	677679	chr1:175968398-175968540:-
SCARNA4	ENSG00000281394.1	677771	chr1:155925958-155926086:-
SCARNA5	ENSG00000252010.1	677775	chr2:233275727-233276002:+
SCARNA6	ENSG00000251791.1	677772	chr2:233288676-233288940:+
SCARNA7	ENSG00000238741.1	677767	chr3:160514907-160515236:-
SCARNA8	ENSG00000251733.1	677776	chr9:19063656-19063786:-
SCARNA9	ENSG00000254911.3	619383	chr11:93721513-93721865:+
SCARNA10	ENSG00000239002.3	692148	chr12:6510222-6510551:+
SCARNA11	ENSG00000251898.1	677780	chr12:6581474-6581609:-
SCARNA12	ENSG00000238795.1	677777	chr12:6967337-6967606:-
SCARNA13	ENSG00000252481.1	677768	chr14:95533355-95533629:-
SCARNA14	ENSG00000252712.1	692149	chr15:66347207-66347342:-
SCARNA15	ENSG00000277864.1	677778	chr15:82755945-82756071:+
SCARNA16	ENSG00000275143.1	677781	chr17:77089307-77089493:+
SCARNA18	ENSG00000238835.1	677765	chr5:83064204-83064337:-
SCARNA18B	ENSG00000238754.1	107397392	chr1:193057281-193057415:-
SCARNA20	ENSG00000252577.1	677681	chr17:60231516-60231646:-
SCARNA21	ENSG00000252835.1	677763	chr17:7906122-7906260:+
SCARNA22	ENSG00000249784.1	677770	chr4:1974636-1974760:+
SCARNA23	ENSG00000251869.1	677773	chrX:24744441-24744570:+