Translating ribosome affinity purification from *C9orf72*-associated amyotrophic lateral sclerosis patient-derived iPSC motor neurons



Chaitra Sathyaprakash

St Cross College

Nuffield Department of Clinical Neuroscience

University of Oxford

Supervised by Prof Kevin Talbot and Dr Ruxandra Dafinca

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2019

“Sometimes fate is like a small sandstorm that keeps changing direction. You change direction, but the sandstorm chases you. You turn again, but the storm adjusts. Over and over you play this out, like some ominous dance with death just before dawn. Why? Because this storm isn’t something that blew in from far away, something that has nothing to do with you. Something inside you. So all you can do is give in to it, step right inside the storm, closing your eyes and plugging up your ears so the sand doesn’t get in, and walk through it, step by step. There’s no sun there, no moon, no direction, no sense of time. Just fine white sand swirling up into the sky like pulverised bones. That’s the kind of sandstorm you need to imagine.

And you really will have to make it through that violent, metaphysical, symbolic storm. No matter how metaphysical or symbolic it might be, make no mistake about it: it will cut through flesh like a thousand razor blades. People will bleed there, and you will bleed too. Hot, red blood. You’ll catch that blood in your hands, your own blood and the blood of others.

And once the storm is over you won’t remember how you made it through, how you managed to survive. You won’t even be sure, in fact, whether the storm is really over. But one thing is certain. When you come out of the storm you won’t be the same person who walked in. That’s what this storm’s all about.”

Haruki Murakami

Abstract

The *C9orf72* hexanucleotide repeat expansion $[GGGGCC]_n$ is the most common cause of familial amyotrophic lateral sclerosis. The endogenous role of the C9orf72 protein has not been fully characterised, nor how the presence of the repeat expansion affects downstream mechanisms. Numerous cellular and molecular pathologies have been characterised in patient-derived induced pluripotent stem cell (iPSC) motor neuron cultures. This includes the accumulation of nuclear RNA foci and cytoplasmic dipeptide repeat proteins, altered calcium homeostasis, accumulation of apoptotic factors, nucleocytoplasmic defects and reduced excitability. Targeted antisense oligonucleotide knockdown of mutant *C9orf72* mRNA transcripts mitigates some of these pathological phenotypes, though whether gene expression is restored to normal is unknown. Transcriptomic analysis of iPSC-motor neurons is challenging due to multiple sources of experimental variance, such as the presence of up to 30% of cells in culture being of unknown identity. The aim of this project is to establish a translating ribosome affinity purification (TRAP) method to isolate an enriched population of RNA from human iPSC-motor neurons. This will help achieve a more accurate analysis of the translatome profile of this model. Differential gene expression analysis between control and *C9orf72* ALS motor neurons will be performed to identify patient motor neuron-specific changes in gene expression.

Declaration

Unless otherwise stated, all the work presented in this thesis is the work of the author and conducted at the Nuffield Department of Clinical Neuroscience, University of Oxford. This work has not been submitted for any other degrees at any other University or educational institute.

Acknowledgements

I would like to thank Prof Kevin Talbot and Dr Ruxandra Dafinca for their fantastic supervision over the 4-year course of this DPhil project. Without their guidance, the completion of this project would not have been possible. Prof Talbot's experience and expertise in this field has been an indispensable source for learning. Dr Dafinca's day-to-day supervision in the laboratory has taught me the crucial skills necessary for my future ambitions. Their patience, and in particular, choice in the freedom I have been given to direct the course of my project was one of the best aspects of my experience at Oxford. I would like to thank the Talbot and Turner research group members, who have provided an enriching and supportive environment for a young researcher to learn and grow.

I would like to thank the members of the Computational Genome: Analysis and Training (CGAT) team, especially Dr David Sims and Dr Charlotte George, and Dr Jakub Scaber (Talbot group), for their invaluable expertise in providing instruction for me to perform the RNA sequencing analysis for this project.

I would like to thank my family, who have been a crucial source of comfort throughout this experience, as well as the inspirations for my decision to lead a career in science.

Finally, I would like to thank all the patients and participants from the motor neuron disease clinic, at the John Radcliffe Hospital, the Motor Neuron Disease Association (Barry Bastin Studentship) and All About Alie charitable organisations, without whom this research would not be possible.

Publications arising from this work

Sathyaprakash C, Scaber J, Candalija A, Barbagallo P, Dafinca R, Talbot K

Translating ribosome affinity purification (TRAP) in human-derived induced pluripotent stem cell (iPSC) cultures as a tool for understanding the pathogenesis of motor neuron disorders. Manuscript in preparation.

Table of Contents

Abstract **3**

Declaration **4**

Acknowledgements **5**

Publications arising from this work **6**

List of Figures **12**

List of Tables **15**

List of Abbreviations **16**

1 Introduction **20**

1.1 Motor neurone diseases *20*

1.2 ALS: a heterogenous, motor neurodegenerative disease *22*

1.2.1 There is no effective treatment for ALS 23

1.3 The genetics of ALS *24*

1.3.1 Superoxide Dismutase 1 24

1.3.2 TAR DNA binding protein 27

1.3.3 Fused-in-sarcoma 29

1.4 The C9orf72 hexanucleotide repeat expansion is the most common gene mutation linked to familial cases of ALS and FTD *29*

1.4.1 Identification and population genetics 29

1.4.2 C9orf72 protein has been implicated in vesicular transport and autophagy pathways 31

1.5 The C9orf72 HRE is thought to cause toxic gain-of-function and loss-of-function pathways leading to motor neuron degeneration *34*

1.5.1 Nuclear RNA foci and sequestered RNA-binding proteins 34

1.5.2 Repeat-associated non-AUG translation produces dipeptide repeat proteins from the C9orf72 HRE 40
1.5.3 Nucleocytoplasmic transport dysfunction 47
1.5.4 Aberrant calcium homeostasis and glutamate-induced toxicity 54
1.5.5 TDP-43 mislocalisation 55
1.5.6 Loss-of-function and endosomal pathway defects 57
1.5.7 Spectrum with frontotemporal dementia (FTD) 61
1.6 C9orf72 ALS has been characterised in a number of versatile models 62
1.6.1 Mouse models of *C9orf72* ALS 63
1.6.2 Patient post-mortem tissue as a validation tool 65
1.7 Induced pluripotent stem cell-derived motor neurons from patients recapitulate several pathological phenotypes of C9orf72 ALS 66
1.8 RNA sequencing of motor neurons is crucial to understanding C9orf72 ALS 69
1.8.1 Methods for studying neurons in isolation 71
1.8.2 Translating ribosome affinity purification 75
1.9 Aims 79
2 Materials and Methods 84
2.1 Molecular cloning of translating ribosome affinity purification constructs and validation of expression 84
2.1.1 Plasmids and primers 84
2.1.2 Patient-derived fibroblast lines and induced pluripotent stem cell induction 87
2.1.3 Molecular cloning 87
2.1.4 HEK293-T cell culture 89
2.1.5 Sanger sequencing 90
2.1.6 Immunocytochemistry 90
2.1.7 SDS-PAGE and Immunoblotting 91
2.2 Optimisation of translating ribosome affinity purification from HEK293-T cells and iPSC-derived motor neurons 93
2.2.1 Lentiviral particle generation, titration and transduction 93
2.2.2 iPSC-motor neuron differentiation 94
2.2.3 Reverse transcription, quantitative PCR (qRT-PCR) 98

2.2.4 Microscopy 98
2.2.5 Translating ribosome affinity purification 98
2.2.6 Quality control using RNA Integrity Number (RIN) analysis 101
2.2.7 Quant-iT RiboGreen Assay 101
2.3 RNA sequencing and data analysis 101
2.3.1 Addition of *sequin* spike-ins 102
2.3.2 RNA sequencing library preparation and platform 102
2.3.3 Bioinformatic analysis, processing and differential gene expression analysis of translatome data 102
3 Molecular cloning of a *translating ribosome affinity purification* (TRAP) lentiviral expression vector 106
3.1 Introduction 106
3.2 A lentiviral TRAP vector for targeted expression in human ChAT-expressing motor neurons in culture 107
3.2.1 TRAP vector starting commercial constructs and design conception 107
3.2.2 Starting commercial vectors produce expected band sizes following restriction digest reactions 111
3.2.3 Starting commercial vectors are expressed in HEK293-T cells and human iPSC-motor neurons 111
3.3 Cloning Strategy 116
3.4 TRAP vector efficiently co-expresses eGFP and tagged RPL22 in target cells .. 126
3.5 Conclusion 134
4 Optimisation of translating ribosome affinity purification from an iPSC-derived motor neuron model of patient *C9orf72* amyotrophic lateral sclerosis 135
4.1 Introduction 135
4.2 High co-expression of TRAP vector components and ChAT-expressing iPSC-derived motor neurons was achieved 137

4.2.1 Optimal lentiviral transduction concentration varies across differentiated iPSC-motor neuron lines 139
4.2.2 Optimisation of transduction conditions resulted in minimal cell death 147
4.3 Optimisation of translating ribosome affinity purification from transduced cell cultures 153
4.3.1 Harvesting and handling techniques dramatically affect quality and yield of RNA 157
4.3.2 Varied TRAP vector transduction efficiency rate results in different yields of RNA 158
4.3.3 Dissociation and purification of mRNA from magnetic beads was optimised for maximum output 169
4.4 RNA sequencing of TRAP-RNA from C9orf72 ALS patient iPSC-motor neurons 170
5 Bioinformatic analysis of RNA sequencing translatome data 177
5.1 Introduction 177
5.2 TRAP and transcriptome input samples have distinctly different quality across all raw and mapped reads 177
5.3 TRAP samples are enriched with mature mRNA species compared to controls. 185
5.4 Differential expression analysis 189
5.4.1 Introduction 189
5.4.2 *Sequin* normalisation for the TRAP and transcriptome datasets is challenging 189
5.4.3 Pulldown of transcripts between individual samples are consistent 194
5.4.4 Exploratory data analysis shows distinct clustering of samples 200
5.4.5 TRAP-RNA is enriched for neuronal markers in the motor neuron translatome 204
5.4.6 *C9orf72* patient motor neurons have several differentially expressed genes compared with age-matched controls 210
5.4.7 DNA-binding, transcriptional activity and intracellular vesicle-associated genes are significantly enriched in the *C9orf72* iPSC-motor neuron translatome 214
5.5 Conclusions 220

6 Discussion 221

6.1 A TRAP vector for human translatome analysis: main conclusions 221

6.1.1 Mature mRNA was successfully purified from iPSC-motor neurons 221
6.1.2 Comparison of TRAP and transcriptome profiles is challenging 224

6.2 DE analysis between C9orf72 patient and control translatomes reveal disease-relevant targets 227

6.2.1 Genes associated with synaptic transmission are misregulated 227
6.2.2 Discrepancies with other *C9orf72* ALS research 230

6.3 iPSC-motor neurons are a challenging model for transcriptomic analysis 232

6.3.1 Increased sample sizes and better controls would improve power 233
6.3.2 Improvements to the protocol may improve RNA yields 234

6.4 TRAP has wide potential for other human models and co-cultures 234

6.4.1 TRAP as a tool in drug screening 235

6.5 Future directions 236

6.6 Closing statements 237

7 References 238

8 Appendix 259

List of Figures

Figure 1.1. Translating ribosome affinity purification. 82

Figure 2.1. Differentiation of iPSCs to motor neurons. 97

Figure 3.1. The lentiviral TRAP vector was cloned from commercially bought expression constructs containing the desired cDNA elements for TRAP vectors. 109

Figure 3.2. Restriction digest reactions and immunofluorescence was used to confirm the sizes and expression of the primary constructs to be used for cloning of the TRAP vector. 113

Figure 3.3. TRAP vector cloning strategy. 117

Figure 3.4. Cloning of the TRAP construct (Figure 3.3, steps 1-4). 120

Figure 3.5. Continued cloning of the TRAP construct (Figure 3.3, steps 5-8). ...123

Figure 3.6. Immunostaining of HEK293-T cells and iPSC-motor neurons following lentiviral transduction TRAP constructs. 127

Figure 3.7. Immunoblotting analysis of protein isolated from mature iPSC-motor neurons transduced with TRAP vectors. 132

Figure 4.1. Lentiviral transduction efficiency of RChAT:TRAP and MChAT:TRAP are slightly different, as shown by expression of reporters. 136

Figure 4.2. Workflow to generate lentiviral particles from cloned TRAP vectors using 3rd generation lentiviral plasmids. 138

Figure 4.3. Immunofluorescence confocal images of iPSC-motor neurons treated with various concentrations of RChAT:TRAP lentivirus to optimise transduction. 141

Figure 4.4. Live fluorescence imaging of iPSC-motor neurons and HEK293-T cells following hUbC:TRAP or RChAT:TRAP lentivirus transduction. 149

Figure 4.5. RNA integrity number (RIN) analysis of iPSC-motor neuron RNA samples pulled down using TRAP protocol from Lesiak et al., 2015. 155

Figure 4.6. Optimisation of washing buffer of beads interacting with antibody-ribosomal complexes in HEK293-T cells.....159

Figure 4.7. Antibody optimisation against molecular tag used for ribosomal pulldown in TRAP protocol.....164

Figure 4.8. The final optimised TRAP protocol was carried out on control, patient and edited iPSC-motor neurons.....174

Figure 5.1. Quality control of raw reads from TRAP and transcriptome input fractions following RNA sequencing using FastQC.....179

Figure 5.2. Assessment of reads mapping to hg38 annotated human genome shows an over-representation of reads in chromosome 21. 183

Figure 5.3. Quality control of reads mapping to hg38 annotated human genome. 186

Figure 5.4. Sample spike-in analysis to normalise RNA sequencing dataset. ...192

Figure 5.5. The normalised counts of all detected genes from individual TRAP samples demonstrates the reproducibility of counts obtained per gene. 196

Figure 5.6. Exploratory differential gene expression analysis of the translatome. 202

Figure 5.7. Normalised gene expression transcripts per million (TPM) counts of marker genes from DE analysis. 207

Figure 5.8. Normalised gene expression in transcripts per million (TPM) counts of genes of interest from DE analysis, grouped by cellular function. 212

Figure 5.9. Fast gene set enrichment analysis (FGSEA) and leading edge analysis of the top 500 most significant genes in the C9orf72 translatome. 216

Figure S8.1. Alignment of predicted promoter regions, MChAT and RChAT for TRAP construct to the human ChAT locus. 259

Figure S8.2. Overlap of total genes or transcripts detected in RNA sequencing experiment between TRAP and transcriptome input sample groups. 261

Figure S8.3. The top 250 most significant genes from differential expression analysis comparing *C9orf72*-TRAP and Control-TRAP samples were used to explore the gene ontology (GO) terms associated with them. 265

List of Tables

Table 2.1. Commercially bought plasmids used in the cloning of TRAP constructs. 84

Table 2.2. List of all primers used in the cloning of the TRAP constructs and for screening colonies transformed with successful ligations. 85

Table 2.3. List of all primers used in Sanger sequencing experiments of the TRAP constructs following molecular cloning of the regions of interest. 86

Table 2.4. List of all antibodies used for either immunocytochemistry or western blotting, with the optimised dilutions used in all experiments, unless otherwise specified. Bolded c-myc antibody was used for TRAP of RNA in final protocol. ...92

Table 4.1. All TRAP experiments performed to obtain TRAP and corresponding transcriptome input samples for RNA sequencing, using the final optimised protocol. IF = input fraction. Samples submitted for sequencing are in bold. 171

Table S8.1. 50 genes with highest log2 fold change in expression, upregulated in differential expression analysis, comparing *C9orf72*-TRAP and Control-TRAP samples. 263

Table S8.2. 50 genes with lowest log2 fold change in expression, downregulated in differential expression analysis, comparing *C9orf72*-TRAP and Control-TRAP samples. 264

Table S8.3. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for molecular function. 268

Table S8.4. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for biological pathway. 269

Table S8.5. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for cellular components. 272

List of Abbreviations

Central nervous system (CNS)

Amyotrophic lateral sclerosis (ALS)

Frontotemporal dementia (FTD)

Neuromuscular junction (NMJ)

Endoplasmic reticulum (ER)

Superoxide dismutase 1 (*SOD1*)

TAR DNA binding protein (*TARDBP*)

Fused-in-sarcoma (*FUS*)

Optineurin (*OPTN*)

Hexanucleotide repeat expansion (HRE)

Repeat-associated non-AUG (RAN)

Dipeptide repeat (DPR)

Differentially expressed in normal and neoplasia (DENN)

Ribonucleic acid (RNA)

Deoxyribonucleic acid (DNA)

Messenger RNA (mRNA)

Ribosomal RNA (rRNA)

Transfer RNA (tRNA)

Micro RNA (miRNA)

Short interfering RNA (siRNA)

Untranslated Region (UTR)

Guanine Exchange Factor (GEF)

Unc-51-like kinase 1 (ULK1)

1A/1B-light chain 3-II (LC3-II)

Translating ribosome affinity purification (TRAP)

Laser-capture microdissection (LCM)

Fluorescence-activated cell sorting (FACS)

Magnetic-activated cell sorting (MACS)

Polymerase chain reaction (PCR)

Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Fluorescence in situ hybridization (FISH)

Genome-wide association study (GWAS)

Induced pluripotent stem cell (iPSC)

Smoothened agonist (SAG)

Antisense oligonucleotide (ASO)

Locked nucleic acid (LNA)

Clustered regularly interspaced short palindromic repeats (CRISPR)

Green fluorescent protein (GFP)

Enhanced green fluorescent protein (eGFP)

Red fluorescent protein (RFP)

Adeno-associated virus (AAV)

Multiplicity of infection (MOI)

Bacterial artificial chromosome (BAC)

Enzyme-linked immunosorbent assay (ELISA)

RNA interference (RNAi)

Phosphate buffered saline (PBS)

Rho-associated, coiled-coil containing protein kinase inhibitor (ROCKi)

Glutathione S-transferases (GST)

Long terminal repeats (LTR)

Liquid-liquid phase separation (LLPS)

Unfolded protein response (UPR)

RNA integrity number (RIN)

Gene transfer format (GTF)

Transcripts per million (TPM)

Open reading frames (ORF)

Principal component analysis (PCA)

Differentially Expressed (DE)

Fast gene set enrichment analysis (FGSEA)

Choline acetyl transferase (ChAT)

Islet 1 (*ISL1*)

Neurofilament heavy (*NEFH*)

Beta-III tubulin (*TUBB3*)

Motor neuron and pancreas homeobox 1 (*MNX1*)

Homeobox 9 (HB9)

Oligodendrocyte transcription factor 2 (*OLIG2*)

Glyceraldehyde-3-Phosphate Dehydrogenase (*GAPDH*)

Beta actin (*ACTB*)

Topoisomerase 1 (*TOP1*)

ATP synthase subunit beta (*ATP5B*)

Acetylcholine receptor (AChR)

α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)

N-methyl-D-aspartate (*NMDA*)

Synaptotagmin-9 (*SYT9*)

Synaptotagmin-11 (*SYT11*)

Syntaxin-6 (*STX6*)

Syntaxin-8 (*STX8*)

Calretinin 2 (*CALB2*)

Calmodulin 1 (*CALM1*)

Calmodulin 3 (*CALM3*)

Valosin-containing protein (*VCP*)

Charged multivesicular body protein 2B (*CHMP2B*)

Microtubule-associated protein tau (*MAPT*)

Granulin (*GRN*)

Glutamate Ionotropic Receptor AMPA Type Subunit 1 (*GRIA1*)

Glutamate Ionotropic Receptor AMPA Type Subunit 4 (*GRIA4*)

Glutamate Metabotropic Receptor 4 (*GRM4*)

Calmodulin (CaM)

Calmodulin-dependent kinase (CaMK)

1 Introduction

1.1 Motor neurone diseases

The human central nervous system (CNS) is composed of billions of neurons of diverse types, each with potentially unique functions, forming trillions of synaptic connections (Kandel et al., 2000). The brain is a highly dense organ, mostly composed of neurons and their supporting cells, known as glia. Neurons are organised in highly dense tissue, forming nuclei of cell bodies, with axons projecting both locally, and to remote areas of the brain and toward peripheral junctions with muscle and skin (Kandel et al., 2000). Immune cells can also be recruited to the CNS upon injury, infection or other trauma via blood vessels, which form a complex vascular network. Collectively this organ monitors, responds and modulates every process occurring in the rest of our bodies, including triggers from our environment. It is also the centre of cognitive processing. Though still being studied to greater and greater detail, the role of the various parts of the CNS have been well characterised. For example, the cerebellum has a role in fine motor control, motor learning and cognition, the hypothalamus in hormonal regulation via the pituitary glands, the hippocampus in memory and the spinal cord in motor reflexes (Kandel et al., 2000). The complexity of this system is increased by the communication of these processes to other parts of the brain. The onset of an injury, trauma or neurodegenerative disease can therefore have massive detrimental effects. The loss of specific subtypes of neurons, via injury or disease, is often accompanied by specific and characteristic pathological phenotypes reflecting the underlying biology of the disease process.

Motor neurons are a neuronal subtype found in the spinal cord, brain stem motor nuclei, and the motor cortex, which communicate to coordinate our voluntary and reflexive movements (Kandel et al., 2000). Spinal motor neurons, amongst the largest cell type in mammals, develop from the mesodermal layer of an embryo, growing over long distances *in utero* to form synaptic connections with muscle fibres. Given that the cell bodies of spinal motor neurons begin at the spinal cord, and extend to the furthest muscle in the legs, it will be evident that they have axons which can extend up to a meter, and are thought to be maintained from birth to adulthood, without replacement (Kandel et al., 2000). The activation of muscle fibres is controlled by the release of neurotransmitters at neuromuscular junctions (NMJ). A motor neuron and its connection with muscles are together known as the motor unit. Voluntary movement is specifically controlled by motor neurons which synapse onto skeletal muscles, in contrast to cardiac or smooth muscle, which are involved in rhythmic heart beat control or gut movement, respectively (Kandel et al., 2000). Diseases of the motor unit are therefore characterised by loss of neurons or muscle. Loss of muscle tissue due to a primary myopathic disease does not lead to atrophy of neuronal tissue. In contrast, neurogenic diseases affecting the motor neuron first, ultimately lead to muscle atrophy. Though easily classifiable through post-mortem tissue examination, both of these disease types are initially characterised by muscle weakness (Kandel et al., 2000; van Es et al., 2017). Historically this led to clinical uncertainty in distinguishing neurogenic and muscular disease, which has largely been resolved with a greater understanding of the taxonomy of neurological disease and modern diagnostic methods, principally electromyography (Kandel et al., 2000). Certain clinical features can imply the loss

of solely upper or lower motor neurons, or both, as is typically the case for the most common type of motor neuron disease, amyotrophic lateral sclerosis (ALS).

1.2 ALS: a heterogenous, motor neurodegenerative disease

ALS is a rapidly progressive neurodegenerative disease, where combined upper and lower motor neurons loss leads to a loss of function, increasing paralysis, and eventually death from loss of ability to breathe (Tandan and Bradley, 1985). The median age of symptom onset is 65 years, with some cases occurring as early as 20 years of age. In over 90% of cases ALS occurs without a family history of a similar disorder (Kiernan et al., 2011). A significant proportion of the remaining 10% of cases can be linked to various autosomal dominant gene mutations. One of the largest challenges in this area of research is characterising the mechanisms whereby genetic mutations lead to motor neuron death through pathological cellular pathways. This is especially challenging as individuals with a mutation live with it from conception, though the disease does not manifest until much later in life. Collectively, over 60% of genetically determined ALS cases are accounted for by mutations in superoxide dismutase 1 (*SOD1*), TAR DNA binding protein (*TARDBP*), fused-in-sarcoma (*FUS*) and *C9orf72* (Talbot and Ansorge, 2006; Scaber and Talbot, 2016; Talbot et al., 2018). An interesting aspect of this is that all of these mutations apparently cause the same range of clinical phenotypes. Although most ALS cases at autopsy show TDP-43 pathology, *SOD1* and *FUS* cases are pathologically distinct. Understanding how each mutation leads to motor neuron degeneration, identifying any commonalities, either in specific pathology cellular pathways, or a particular age-related vulnerability specific to motor neurons in ALS is crucial to finding potential treatments.

1.2.1 There is no effective treatment for ALS

Researchers have tried to find common pathological mechanisms, which may lead to a unified treatment. More recent studies have shown that though the mutated gene products may behave differently, they may have detrimental effects on the same kind of pathways, which become defective and lead to cell death in the same way by end stage of the disease. The first approved drug for ALS was riluzole, a glutamate receptor inhibitor, which is thought to inhibit receptors which cause glutamate-induced excitotoxicity in motor neurons (Lacomblez et al., 1996). This drug was approved before most genetic mutations linked to ALS were discovered, suggesting a universal underlying pathology was present. To investigate this, Bursch et al., 2019 used iPSC-motor neurons from patients with *C9orf72*, TDP-43, *FUS* or *SOD1* mutations to examine responses to calcium modulation and glutamate receptor stimulation, following targeted agonist treatments. Each mutation showed a distinct difference in receptor subunit expression and calcium dynamics, suggesting that though there is a shared pathology, the specific pathway mechanics vary in subtle ways, which are unlikely to be sufficiently rescued with one kind of drug. Another study, looked at a cohort of iPSCs from sporadic ALS patients, using an automated system to culture and differentiate motor neurons from 32 individuals (Fujimori et al., 2018). This study identified multiple phenotypes in their patient motor neurons and used an automated drug screen to find hits which corrected these cellular readouts. Ropinirole was identified as a potential treatment for sporadic cases of ALS, though strikingly not for *SOD1* ALS, in whose iPSC-motor neuron cultures, pathological phenotypes where not corrected. Further investigation of this drug in animal models or in clinical trials has not yet been published.

1.3 The genetics of ALS

Despite only making up 10% of all ALS cases, inherited mutations that have been shown to cause ALS are numerous. Currently, over 20 genes have been implicated in familial ALS as reviewed in Volk et al., 2018. The majority of these are rare, making up less than 1% of all cases, worldwide. The most well characterised mutations are also the most common.

1.3.1 Superoxide Dismutase 1

Superoxide dismutase 1 (*SOD1*) is a metalloprotein, with a function in the conversion of toxic oxygen free radicals to oxygen and hydrogen peroxide (Rosen et al., 1993). It was originally found to be associated with familial ALS via autosomal inheritance in a linkage study (Siddique et al., 1991). 23 families' DNA were extracted, digested using restriction endonucleases and run on a gel to identify common markers. These were narrowed down to 4 loci on the long arm of chromosome 21 (Siddique et al., 1991). These were further analysed in the families, narrowing down to the locus, D21S223. In a follow up study, polymerase chain reaction (PCR) was used to identify the latter as *SOD1* and various mutations associated with 13 families, 2 of which were unrelated, but had the same mutation (Rosen et al., 1993). The number of known validated *SOD1* mutations has now increased to above 180 (Maurel et al., 2018), with a mixture of point substitutions or deletions, and exon truncations (Turner and Talbot, 2008). In the following years, mouse models for *SOD1* mutations were rapidly generated and used for pathological analysis of *SOD1* and its potential role in causing neurodegeneration in spinal motor neurons. A potential mechanism through loss-of-function of *SOD1*

protein was deemed unlikely from several studies, as the knockout mouse model did not develop any fatal motor defects, though a number of other pathological phenotypes were noted, including a reduction in fertility of the females and increased rate of ageing, light sensitivity degeneration (resulting in the formation of cataracts and macular degeneration) and death at two years from tumours (Turner and Talbot, 2008). Further investigation of the knockout model found significant denervation and muscle atrophy by 18 months in these mice, though not directly causing death. It was hypothesised that *SOD1* protein has a role in several cellular functions, thus mutations may be causing a partial loss-of-function, which in turn could confer motor neuron sensitivity. The original authors of the *SOD1* discovery shortly after published a study where several mice with the identified familial *SOD1* mutations were generated, which displayed a fatal motor phenotype, strongly resembling ALS with lower motor neuron defects. This strongly suggested a toxic gain-of-function mechanism, though the model invited technical criticism due to the inconsistency of transgene copy numbers and gene expression across models, making it difficult to normalise and compare pathophysiological results. Furthermore, the onset of disease phenotype was at the 3-month stage, which is comparatively too early for an ageing neurodegeneration model. Nevertheless, more than a dozen ALS models of *SOD1* mutations have been generated to date, and thoroughly characterised in terms of neuropathology, behavioural studies relating to motor coordination and cognitive impairment, molecular pathways and RNA sequencing studies (Gurney et al., 1994; Bruijn et al., 1997; Sun et al., 2015). Since the emergence of induced pluripotent stem cell (iPSC)-motor neuron differentiation models, some *SOD1* mutation-linked pathologies have also been characterised in human cell cultures (Chen et al., 2014), and have opened up an avenue to study

the role of endogenous human SOD1 (Bursch et al., 2019; Wong and Venkatachalam, 2019). For instance, Chen et al., 2014 identified that the well-known ALS pathology of the accumulation of neurofilament chains (Hirano et al., 1984) and SOD1 protein, in affected tissue in ALS, can also be observed in iPSC-motor neurons from a patient with the $SOD1^{D90D}$ mutation, though not in non-motor neurons or wildtype controls. These iPSC-motor neurons also exhibited neurite swelling, neurite degradation and increased levels of the apoptotic marker, cleaved caspase-3. Homologous recombination was used to generate iPSCs which expressed neurofilament-light chains under a tetracycline-inducible CAG promoter, which reversed this phenotype. This pathology was further linked to the increased binding of neurofilament-light mRNA to mutant SOD1 via the 3'-untranslated region (3'UTR). The polymerization of neurofilament chains is known to be a controlled process, where the balance between neurofilament-light, -medium and -heavy proteins is crucial. The sequestration of one of these components' mRNA results in a loss of this balance, as a consequence of fewer light chains being translated, thus causing the accumulation of the other chains (Chen et al., 2014). This study was crucial to linking a well-known neuropathology in SOD1 ALS patients and transgenic mouse models, to a human iPSC-motor neuron model, showing that a high copy number of mutant SOD1 is not required to trigger axonal degradation as in mouse models. It highlighted a potential point of therapeutic intervention, suggesting blocking the interaction between mutant SOD1 and neurofilament-light mRNA. Finally, while glia are thought to play a role in contributing to pathological mechanisms in ALS, Chen et al., 2014 showed that in this case the phenotype can manifest in their absence, implying a cell autonomous mechanism. However, the paper did not identify why neurofilament accumulation is specific to iPSC-motor

neurons, a factor that remains unknown. Several other toxic gain-of-function mechanisms have been linked to *SOD1* ALS, likely triggered by the various mutations, which may induce codon shifts through their deletions, truncated peptides or altered sequence from base substitutions (Turner and Talbot, 2008). These mechanisms include defects in the endoplasmic reticulum and mitochondria (Higgins et al., 2003; Kikuchi et al., 2006; Magrané et al., 2014). It is thought that increased accumulation of mutant SOD1 protein over time leads to impairment of core cellular pathways, and subsequent atrophy of neuronal tissue. RNA sequencing of iPSC-motor neurons from patients with the $SOD1^{A4V}$ mutation showed perturbations in mitochondrial pathways, though not in $SOD1^{A272C}$ patients (Wong and Venkatachalam, 2019). The latter showed an upregulation of microtubule transport pathway, which does not correlate with the known pathology of axonal degradation, though this may be because the RNA was harvested at a stage of the motor neuron culture where this phenotype is absent.

1.3.2 TAR DNA binding protein

The *TARDBP* gene encodes a protein called TDP-43, which has a general function in RNA processing, with roles in DNA binding, messenger RNA (mRNA) splicing and RNA maturation. It was originally identified as relevant to ALS in 2006 as component of the cytoplasmic ubiquitin-positive granules associated with the vast majority of ALS cases (Neumann et al., 2006). TDP-43 has predominant nuclear localisation in healthy cells. Thus, this key pathological phenotype, seen in over 90% of TDP-43 and *C9orf72* mutation-associated, and sporadic cases of ALS, a cytoplasmic mislocalisation of TDP-43 implies a loss-of-function of the nuclear RNA-binding protein. Its increased recruitment to stress granules also suggests a

potential toxic gain-of-function mechanism. Interestingly, ALS cases with a *SOD1* or *FUS* mutation are negative for TDP-43-positive aggregates (Mackenzie et al., 2007).

TDP-43 is a component of stress granules, which recruit protein and RNA from the local environment in the cytoplasm, sequestering them, and thus preventing their endogenous functions in times of cellular stress (McDonald et al., 2011; Chou et al., 2018; Mann et al., 2019). It has been speculated, but remains unproven, that these may seed TDP-43 aggregates in ALS. The components of stress granules have been widely studied to attempt to identify key cellular pathways that have been affected by this pathological process. Stress granules are dynamic non-membrane bound organelles with a role in buffering proteins in response to toxic signals triggering autophagy pathways to recycle and conserve proteins (Anderson and Kedersha, 2008). TDP-43 is 414 amino acids in length, containing two RNA recognition motifs from the N-terminal domain and one glycine-rich domain at the C-terminus, where the majority of the dozens of mutations implicated in TDP-43-linked ALS lie (Kuo et al., 2009). This is also known as a low complexity domain, and is the part of the protein which forms non-specific interactions in stress granules. Rarer cases, where mutations exist in the RNA recognition motifs have also been found. It is thought that TDP-43 mutations disturb the local conformation of the protein, leading to nuclear mislocalisation and aggregation at ubiquitin-positive stress granules. Different proteins have been found to co-localise with these granules, including different patterns in different parts of the brain in ALS patient post-mortem tissue (Lagier-Tourenne et al., 2010). Strikingly, TDP-43 mislocalisation is a characteristic pathology seen in sporadic ALS cases and

patients carrying a *C9orf72* hexanucleotide repeat expansion. Initial studies showed that cytoplasmic granules in *C9orf72* ALS patient tissue did not co-localise with TDP-43 (Al-Sarraj et al., 2011; Troakes et al., 2012). However, later studies showed a potential role for TDP-43 in *C9orf72*-related disease (Mackenzie et al., 2015; Davidson et al., 2016). Specific examples of these have been described below (section 1.5.5).

1.3.3 Fused-in-sarcoma

The *FUS* gene encodes an RNA-binding protein, originally found to be linked to a UK-based kindred, in a lower motor neuron predominant ALS subtype (Vance et al., 2009). Similar to TDP-43, *FUS* was shown to accumulate in cytoplasmic granules. *FUS* is thought to have a role in transcription, processing and splicing of RNA, suggesting mechanistic similarities to TDP-43 in ALS. However, like *SOD1* cases of ALS, patients with *FUS* mutations have not been found to contain TDP-43-positive, ubiquitin-positive, tau-negative cytoplasmic inclusions, suggesting a distinct pathophysiology.

1.4 The *C9orf72* hexanucleotide repeat expansion is the most common gene mutation linked to familial cases of ALS and FTD

1.4.1 Identification and population genetics

The commonest mutation linked to ALS is a hexanucleotide ($[GGGGCC]_n$) repeat expansion (HRE) in intron 1a of the *C9orf72* gene. It was discovered through a combined positional cloning and genome-wide association (GWAS) strategy from different cohorts including a crucial Finnish population of ALS patients (DeJesus-

Hernandez et al., 2011; Renton et al., 2011). The *C9orf72* mutation makes up the most significant proportion of familial ALS cases (~40%) in European and North American populations, though it appears to be virtually non-existent in East Asian populations. These 6 nucleotides have been found to repeat hundreds to thousands of times in patients, though a threshold for disease manifestation has not been determined. Furthermore, studies observing age of onset, disease progression and disease severity to the number of hexanucleotide repeats have not found any significant correlations (van Blitterswijk et al., 2013; Nordin et al., 2014). Examination of post-mortem human tissue using Southern blot analysis showed that the number of repeats in intron 1a of *C9orf72* varies across neuronal and non-neuronal tissue between 18 different patients (Nordin et al., 2014). Variation in repeat length within groups of similar tissue, such as different brain regions, compared to kidney, or various muscles for example, were closer. However repeat lengths from individual patients in the same tissue were highly variable. An overall pattern appeared to suggest a greater number of repeats in neural tissue compared with non-neuronal, though blood appeared to be an anomaly to this rule, being generally more similar to neural tissue (Nordin et al., 2014). This may suggest a potential reason for neuronal-specificity in disease, though mechanisms of somatic instability in mature or developing motor neurons are unknown. Additionally, a study has also found that the *C9orf72* HRE appears in a broader phenotypic spectrum of neurodegenerative conditions other than pure ALS or frontotemporal dementia (FTD), such as Huntington's and Alzheimer's like syndromes at a significantly higher rate than across the national average in a UK-based cohort (Beck et al., 2013). This finding has prompted researchers in the field to wonder whether a true overarching mechanistic link between wider neurodegenerative disease may exist,

and if more studies to find common molecular pathways should be conducted. This topic remains controversial, and no evidence yet exists to suggest a stronger link between *C9orf72* and other neurodegenerative diseases as strong as the evidence between ALS and FTD.

1.4.2 C9orf72 protein has been implicated in vesicular transport and autophagy pathways

Three spliced transcript isoforms arising from the *C9orf72* locus have been identified. The total levels of *C9orf72* mRNA are reduced in post-mortem patient neuronal tissue, including in the cerebellum, motor cortex, spinal cord and iPSC-derived neurons (DeJesus-Hernandez et al., 2011; Donnelly et al., 2013). The most common isoform does not contain intron 1a in its immature RNA transcripts. The two that do are generally less abundant, but also downregulated in disease, compared with age-matched controls and non-*C9orf72* forms of ALS. Distinctive protein functions of these isoforms are currently unclear. Motif scanning and sequence homology comparison of the exon regions of *C9orf72* with known motifs suggest the presence of a *differentially expressed in normal and neoplasia* (DENN)-like domain (Levine et al., 2013), which has a function as guanine exchange factors (GEF) in regulators of members of the Rab family of GTPases (Marat et al., 2011). With around 70 members, this protein family is involved in the trafficking of vesicles between membrane-bound organelles and the cell membrane (Zerial and McBride, 2001). Through the exchange of GDP to GTP, switching from an inactive to active state, respectively, Rab proteins can recruit vesicles via SNARE proteins to mediate their fusion at various organelles. Through this pathway, the transport of cargo between the endoplasmic reticulum, various layers of the Golgi body, early and late

endosomes, and lysosome is maintained (Zerial and McBride, 2001). Therefore, the role of GEFs as mediators of Rab protein function is crucial. The primary sequence of C9orf72 does not match significantly with any other known family of proteins, but the predicted secondary and tertiary conformation at the N-terminal region of its DENN-like domain, comprising of alternate beta sheets and alpha coils (known as a *longin* domain), is also found in SNARE proteins, which have a role in vesicle fusion at the synaptic membrane (Levine et al., 2013). C9orf72 has been well established as a binding partner of the Rab family of GTPases, involved in several processes, including exosome secretion and endosomal trafficking pathways (Farg et al., 2014; Aoki et al., 2017). A potential role for the C9orf72 and Rab proteins in ALS pathology is discussed in section 1.5.6.

The first paper to show an endogenous function for C9orf72 protein in autophagy was Webster et al., 2016, where its interaction with Rab1a and a protein called Unc-51-like kinase 1 (ULK1) was demonstrated. ULK1 is a known member of the initiation complex in autophagy, which is involved in the generation of autophagosomes. This study showed that the knockdown of *C9orf72* mRNA transcripts in human immortalised cell lines and rat primary neurons via siRNA or miRNA, respectively, resulted in a depletion in autophagy marker, 1A/1B light chain 3-II (LC3-II). Conversely, overexpression of C9orf72 protein increased levels of LC3-II, though not when the ULK1-initiation complex was inhibited, suggesting a link between the two. Overexpression of long or short isoform of C9orf72-c-myc or C9orf72-FLAG in HEK293 cells, followed by immunoprecipitation revealed that ULK1, and other initiation complex proteins were pulled down with both isoforms (Webster et al., 2016). This was further validated in a proximity ligation assay. It

was observed that ULK1 phosphorylation levels were not affected by the knockdown of *C9orf72*, though cellular localisation was, suggesting a potential role in trafficking of the autophagy initiation complex. Rab1a is a known regulator of autophagy in mammalian cells, and was further confirmed in this study through cellular knockdown, resulting in inhibition of initiation complex translocation to phagosomes. Confirmation of an interaction between C9orf72 and Rab1a suggested a role of the former as a GEF of this Rab protein and subsequent trafficking of the ULK1 autophagy initiation complex (Webster et al., 2016). Subsequent studies have replicated the finding that C9orf72 clearly has a role in autophagy (Amick et al., 2016; Sullivan et al., 2016; Yang et al., 2016). However, there is a discrepancy over whether C9orf72 activates or inhibits the process, and how this occurs. Yang et al., 2016 showed that a protein called SMCR8, which also contains a predicted DENN-like domain, interacts with C9orf72 and ULK1. When SMCR8 was knocked out in a mouse model, autophagy was compromised. This study also showed that C9orf72 can interact with several Rab proteins, including Rab39B, though not Rab1a. Further, it was shown that only the long isoform of C9orf72 interacted with Rab39B, though Rab1a also co-immunoprecipitated with the short isoform. Furthermore, Yang et al., 2016 showed that Rab39B did not pull down with SMCR8, though Sellier et al., 2016 showed the converse. It should be noted that these studies used different models to validate their results suggesting a reason for different results from immunoprecipitation experiments. Furthermore, other studies have shown that C9orf72 can potentially interact with many members of the Rab family of GTPases (Farg et al., 2014; Aoki et al., 2017; Corbier and Sellier, 2017), implying that it may have many roles as GEFs, and interactions with other proteins such as SMCR8, and the metabolic condition of a cell may affect this.

1.5 The *C9orf72* HRE is thought to cause toxic *gain-of-function* and *loss-of-function* pathways leading to motor neuron degeneration

An ongoing debate in this field is whether the *C9orf72* HRE mutation causes motor neuron degeneration resulting from RNA or protein-related toxicity, or through loss-of-function by downregulation of the C9orf72 protein, whose specific function is still unclear. A number of studies have strived to find a prevailing pathological pathway in ALS which causes motor neuron defects, and may be targeted by pharmacological means, or by altering gene expression using antisense oligonucleotides, as a potential means for therapy. Targeted knockdown of C9orf72 HRE-containing mRNA is currently undergoing phase I of clinical trials.

1.5.1 Nuclear RNA foci and sequestered RNA-binding proteins

The first cellular pathology identified in the tissue of patients carrying the *C9orf72* HRE mutation was the accumulation of transcripts containing the mutant repeat in the nucleus. Fluorescence in situ hybridization (FISH) using molecular probes targeting the repeat showed small puncta in the nucleus of post-mortem ALS patient tissue. This pathology was evident in the spinal cord, frontal cortex and cerebellum (DeJesus-Hernandez et al., 2011; Gendron et al., 2013; Lagier-Tourenne et al., 2013). A potential pathological mechanism associated with nuclear RNA accumulation in patients carrying the *C9orf72* repeat expansion was originally hypothesised and demonstrated by Lagier-Tourenne et al., 2013. Patient fibroblasts, immortalized lymphocyte lines and CNS tissue from post-mortem patients with the *C9orf72* repeat, were treated with repeat-targeted fluorescent DNA probes,

showing the presence of nuclear foci in these cells (Lagier-Tourenne et al., 2013). This study identified the presence of both sense and antisense RNA foci in neuronal and glial cells in patients. The presence of antisense transcripts have been further validated in other studies (Cooper-Knock et al., 2015b). Despite being a motor neuron disease, these foci were seen in cortical, hippocampal and cerebellar neurons (Lagier-Tourenne et al., 2013). The same pathology was soon after identified in iPSCs, and differentiated neurons, derived from a patient diagnosed with frontotemporal dementia, and a related pre-symptomatic carrier (Almeida et al., 2013). RNA foci, comprising the *C9orf72* intronic repeat mRNA, is the most common pathology in post-mortem brains of ALS patients, and has been robustly identified in cellular and animal models of *C9orf72* ALS, using FISH (Almeida et al., 2013; Lee et al., 2013; Sareen et al., 2013; Zu et al., 2013). The HRE was predicted to form known tertiary structures known as G-quadruplexes, due to the presence of four consecutive guanine residues interacting via *Hoogsteen* hydrogen bonds in a planar conformation, to form multiple stacks (Haeusler et al., 2014). Absorbance patterns using circular dichroism spectroscopy on repeat-containing DNA, and RNA transcripts containing short and longer repeat expansions, generated *in vitro* in physiological osmotic conditions supported this hypothesis (Haeusler et al., 2014). Pulldown of repeat transcripts overexpressed in HEK293-T cells via a streptavidin-conjugated repeat oligonucleotide, and analysis of co-immunoprecipitated proteins via western blotting showed preferential binding of RNA-binding proteins hnRNPs U, F and, nucleolin and RPL7 (Haeusler et al., 2014). Specific pulldown of RNA via nucleolin following stable isotope labelling by amino acids in cell culture and analysis using mass spectroscopy suggested the presence of cytosine-rich hairpin loops, representing the presence of antisense repeat transcripts (Haeusler et al.,

2014). Nuclear RNA foci may comprise of both sense or antisense mRNA forms of the transcript (Gendron et al., 2013), and preferentially interact with various RNA-binding proteins (Haeusler et al., 2014). Further studies also identified proteins that interact with the immature *C9orf72* transcript. Initial characterisation of RNA foci and RNA-binding protein colocalization was examined in Sareen et al., 2013 where a panel of common proteins were co-stained with FISH. RNA-binding proteins, hnRNPA1 and Pur-α showed positive colocalization in iPSC-derived motor neurons from four *C9orf72* ALS patients. However, RNA sequencing analysis of these motor neurons compared with four control lines did not reveal these proteins to be dysregulated. This may imply that sequestered proteins are not being transcribed further following the protein is sequestered at nuclear foci. Three of the patient lines had around 800 intron 1a repeats; which were revealed to cluster closely in hierarchical clustering analysis, while the patient line with only 70 repeats was more distinctive, suggesting a repeat-length mechanistic link (Sareen et al., 2013). Larger studies looking at multiple patients with a wide range of repeat lengths have not been done, which could clarify this pattern. It may also be important to conduct studies by grouping patients from the same families to examine the potential importance of any shared inherited risk factors. Another study performed RNA immunoprecipitation by incubating the nuclear lysate of SH-SY5Y neuroblastoma line cells with biotin-labelled $[GGGGCC]_{72}$ (Lee et al., 2013). Heterogenous nuclear ribonucleoprotein, hnRNPH1, was pulled down, and also found to colocalize with nuclear foci in cultures where $[GGGGCC]_{72}$ was overexpressed using immunostaining (Lee et al., 2013). Immunostaining also showed that serine-arginine-rich splicing factor proteins, SRSF1 and SC35 also colocalize with HRE foci, but do not directly coimmunoprecipitate with *C9orf72* repeat RNA. Finally, a

study where tagged $[GGGGCC]_5$ was incubated with the cell lysates of SH-SY5Y cultures, followed by coimmunoprecipitation, corroborated the findings from Lee et al., 2013 by identifying the pulldown and colocalization of hnRNPs, *ALYREF* and SRSF2 proteins with *C9orf72* HRE RNA (Cooper-Knock et al., 2014). Arguably the latter study does not represent a pathological HRE, which are generally at least 100 in patients, but does validate that specific proteins involved in nuclear transport bind to *C9orf72* transcripts with different HRE lengths, which is valuable to understand the mechanism of transcript nuclear export (further discussed in section 1.5.3).

The interaction of these disease-specific RNA foci with multiple nuclear proteins immediately suggests a mechanism of pathology, whereby sequestered RNA-binding proteins are unable to carry out endogenous functions, potentially leading to defects in key cellular functions. The presence of the *C9orf72* HRE in nuclear foci, and ALS pathology of mislocalised TDP-43 and *FUS* RNA-binding proteins suggests a mechanistic link in these clinically identical diseases, as well as other neurodegenerative diseases. Myotonic dystrophia and Fragile X-associated tremor ataxia syndrome also show accumulation of repeat-containing transcripts, and sequester RNA-binding proteins through non-specific binding (Polymenidou et al., 2012). Non-specific binding to RNA in myotonic dystrophy causes the sequestration of muscleblind protein and global splicing misregulation in muscle tissue, which can be rescued through overexpression of muscleblind (Miller et al., 2000; Kanadia et al., 2011). The group of interactors that have emerged in other *C9orf72* ALS studies (as given above) appear to be less clear and reproducible. Sequestration of these proteins is thought to prevent them from carrying out their endogenous function, and have effects further downstream, though there is little direct evidence for this.

Sequestered RNA-binding proteins implicated in multiple functions may have more diverse pathological effects which may be challenging to treat therapeutically. The specificity or reproducibility of such processes are unknown, especially in relation to *C9orf72* mutant repeat length. The reliability of the sequestration of the same RNA-binding proteins to pathological foci is not clear, therefore, the pathways disrupted in all patient motor neurons cannot necessarily be predicted. This may be dependent on other factors, such as the number of repeats, local formation of secondary structures, and potentially subtle differences between patient gene isoforms. It would be interesting to pull down the repeat-containing RNA and potential protein interactors directly from patient-derived iPSC-motor neurons, however this would be extremely technically challenging. Furthermore, it has been shown that compensatory mechanisms are likely to be in place to increase transcription of genes, whose mRNA are sequestered. Cooper-Knock et al., 2015a performed RNA sequencing on motor neurons from post-mortem human C9orf72 ALS patients and controls, using laser-capture microdissection. In contrast to Sareen et al., 2013, network analysis of the differentially expressed genes showed the enrichment of genes involved in splicing by more than 50%. Several of these also overlapped with proteins known to be sequestered at RNA foci.

Together these studies do not suggest a converging pathway to target for therapeutic strategies, especially since each independent study appears to show a unique group of proteins affected by the repeat. The nuclear proteins implicated in non-specific sequestration have roles in mRNA transcription, processing, splicing and nuclear transport. Misregulation of RNAs may not only affect coding-transcripts, but also ribosomal RNAs (rRNA), micro RNA (miRNA) and other species involved

in RNA regulation. An overall change to *C9orf72* ALS patient transcriptome is predicted. Furthermore, the proteome will also be affected as a result of defects in RNA processing and transport. Antisense oligonucleotide (ASO) intervention, targeting the *C9orf72* repeat in many *C9orf72* ALS models have shown that the RNA foci phenotype can be ameliorated (Donnelly et al., 2013; Lagier-Tourenne et al., 2013; Sareen et al., 2013; O'Rourke et al., 2015; Zhang et al., 2015; Jiang et al., 2016). Though a convenient readout and early target in the cellular pathway, this has not been shown to slow down or reverse motor phenotypes in a mammalian model. There are also several challenges associated in ASO delivery in patients, including crossing the human blood brain barrier, specific targeting to motor neurons and cortical neurons, and optimising stability of the ASO through various chemistries to allow delivery to the nucleus. It is therefore important to also look for other pharmacological targets. To better understand how the *C9orf72* repeat may affect a motor neuron transcriptional profile, further pathological characteristics resulting from the mutation should be understood.

1.5.1.1 The *C9orf72* HRE affects RNA processing and alternative splicing

Experiments investigating the structure of *C9orf72* HRE DNA have suggested that complex tertiary structures may form at the chromosomal locus (Haeusler et al., 2014). The transcriptional process of *C9orf72* may be affected by a large repeat sequence due to changes in methylation at the gene locus compared with the wildtype loci (Gijselinck et al., 2015). Also, local secondary and tertiary structures of the chromosomal DNA may be altered causing unpredictable effects on the recruitment of transcription factors or other machinery to the site. The transcribed immature mRNA may alter splicing efficiency and other pre-processing

mechanisms. These immature transcripts are translocated to the cytoplasm via the nuclear pore complex (further discussion in section 1.5.3), finally producing disease-specific repeat-associated non-AUG (RAN) translated peptides. For a more detailed discussion of this please see section 1.5.2.

In addition to the misprocessing and translocation of *C9orf72* transcripts, it has also been found that transcripts of proteins found to co-localise at RNA foci are upregulated in expression in *C9orf72* ALS patient motor neurons (Cooper-Knock et al., 2015a). Additionally, this correlated with an increased discrepancy in the splicing profile between *C9orf72* ALS, and non-*C9orf72* ALS and control motor neurons. There was no significant difference in the overall number of splicing events between the different genotypes in this study, but the levels of exon inclusion or exclusion varied considerably. The splicing profiles between slower and fast progressing *C9orf72* ALS patient groups were also examined. This analysis was performed by independent comparison of random splicing events within groups to give an unbiased output, showing a potential conservation of splicing events depending on disease severity (Cooper-Knock et al., 2015a). The alleviation of RNA foci following antisense oligonucleotide-targeted knockdown of the *C9orf72* HRE has been shown in several studies (Donnelly et al., 2013; Lagier-Tourenne et al., 2013), but whether this restores altered splicing events is not clear.

1.5.2 Repeat-associated non-AUG translation produces dipeptide repeat proteins from the C9orf72 HRE

Repeat-associated non-AUG (RAN) translation is a disease-specific, non-canonical form of ribosomal translation that is initiated in the absence of a start codon (AUG)

and without the recognition of a 5' cap (Kearse and Todd, 2014). Messenger RNA containing repeated nucleotide regions are translated to produce potentially toxic peptides. The absence of a start codon allows translation to begin at codons shifted along by one base on the mRNA, producing peptides from three reading frames. This can occur on both the sense and antisense mRNA molecule. The occurrence of RAN translation is thought to be dependent on repeat length, though the precise threshold is unclear (Wojciechowska et al., 2014). The presence of RAN translated dipeptides being part of the cellular pathology of *C9orf72*-related diseases was hypothesised due to a previous similar observation in microsatellite expansion-causing diseases including Huntington's disease (MacDonald et al., 1993) and various subtypes of spinocerebellar ataxia (Lindblad et al., 1996; Mann et al., 2013). In the latter, the presence of varying lengths of $[CAG]_n$ repeats results in toxic poly-glutamine chains being translated (Lindblad et al., 1996). However, an important distinction between the two diseases is that the number of repeats in spinocerebellar ataxia is positively correlated with disease severity (Lindblad et al., 1996), though no correlation exists in *C9orf72* ALS between size of repeat and disease progression, onset age or emergence of symptoms (Mann et al., 2013; van Blitterswijk et al., 2013). In the $[GGGGCC]_n$ sense strand repeat, three dipeptide codon patterns emerge, if RAN translation takes place. These dipeptide repeat (DPR) proteins are poly-glycine-proline (poly-GP), poly-glycine-alanine (poly-GA) and poly-glycine-arginine (poly-GR) (Mann et al., 2013). From the $[CCCCGG]_n$ antisense strand, a further unique poly-alanine-proline (poly-AP) and poly-proline-arginine (poly-PR) can occur, as well as reverse translated poly-GP. The presence of these dipeptides in human post-mortem tissue was detected in Mann et al., 2013 using in-house antibodies raised against poly-GR, -PR, -GP and -AP dipeptides.

They stained positive at the cytoplasm of neurons in the cerebellum and hippocampus, in a granular-like pattern (Mann et al., 2013). This staining was found to have a similar pattern to p62 staining, a major player in autophagy, and potential neuroprotective response to DPR protein expression (Mann et al., 2013). Further studies of tissue from *C9orf72* ALS and FTLD patients have corroborated these findings (Ash et al., 2013; Mori et al., 2013a; Davidson et al., 2014; Mackenzie et al., 2015). The role of DPRs as a trigger for disease, or simply as the result of the presence of the *C9orf72* hexanucleotide expansion, is not yet fully established in human neuropathology. The discovery of these DPRs has led to a surge in the ALS community to find a link to pathological molecular pathways, which results in neurodegeneration. To this end, a number of animal models and human cell culture methods have been utilised, which each have their own advantages and drawbacks. Furthermore, in contrast to the *C9orf72* RNA-related pathology of nuclear foci and sequestration of RNA-binding proteins, DPRs do not have a direct link to effects on gene expression in ALS. However, more recent research studies imply pathological effects for DPRs in several levels of cellular function, which may contribute to overall changes in gene expression through feedback loops. A comprehensive understanding of DPR expression, cellular localisation and linked pathways is necessary to understand how the peptides could alter gene expression.

The first question asked in *C9orf72* ALS was whether the mutation causes a loss-of-function or toxic gain-of-function-driven pathology. Published work supports both hypotheses, however there has generally been a stronger argument for toxic gain-of-function. As mentioned, knockout of the mouse *C9orf72* orthologue gene does not result in any motor deficits (Sudria-Lopez et al., 2016), however overexpression

of poly-PR_n in *Drosophila* eyes, causes a severe toxic phenotype (Wen et al., 2014; Freibaum et al., 2015; Zhang et al., 2015). However, evidence from post-mortem tissue, showing positive immunostaining of cerebellum and hippocampus (Ash et al., 2013; Mann et al., 2013; Mori et al., 2013a) for certain DPRs suggests that even neurons unaffected in disease express the apparently toxic peptides. A transgenic mouse model was generated by inserting a *bacterial artificial chromosome* (BAC) (153.2 kb) containing the human *C9orf72* locus into the target mouse genome by homologous recombination, through pronuclear injections into female mouse oocytes (Peters et al., 2015). The fragment was taken from a BAC library generated from a *C9orf72* ALS patient with 500 HRE repeats. This mouse showed evidence of both RNA foci (FISH) and cytoplasmic DPRs (enzyme-linked immunosorbent assay, ELISA) (Peters et al., 2015). However, this mouse did not have any phenotypes affecting motor function, behaviour or memory (Peters et al., 2015). Assuming that *C9orf72* has a similar role in mouse as in humans, this suggests that the presence of these molecular traits is insufficient to inducing targeted neurodegeneration in motor neurons by endogenous expression. This contrasts with a separate study where P0 mice injected with adeno-associated virus (AAV), packaging a vector encoding human $[GGGGCC]_{66}$ (or $[GGGGCC]_2$ as a control), under a constitutive promoter (Chew et al., 2015). In this case, in addition to foci and DPRs, phospho-TDP-43 inclusions were seen. The mice also displayed evidence of overall neuronal loss in the cortex and motor cortex, along with slight motor function loss by 6 months of age (Chew et al., 2015). While striking, this paper represents an unnatural overexpression model, also excluding the majority of the *C9orf72* gene locus, whose regulation and local methylation may play a role in disease onset and progression. Similar studies choosing to make use of

overexpression models have an inherent problem in providing evidence or validation of their findings in a more disease-relevant model, making conclusions challenging to interpret. Importantly, these two mouse models have striking differences in DPR pathology. The BAC mouse begins to show evidence of poly-GP staining at around 24 months (Peters et al., 2015), while the mouse injected with AAVs to overexpress $[GGGGCC]_{66}$ already displays poly-GA, poly-GP and poly-GR by 6 months (Chew et al., 2015). This suggests a crucial role for DPRs in the onset of ALS pathology over an RNA toxic gain-of-function mechanism.

The biochemical interactions between the different types of DPRs may be crucial to understanding their effect on other proteins or cellular machinery as a result of their prevailing charge or hydrophobicity. The more hydrophobic poly-GA, poly-GP and poly-AP should be more prone to aggregation, reducing the surface area of amino acids to the hydrophilic cellular environment. On the other hand, arginine-containing peptides have often been associated with increased toxicity due to their negative charge, which can alter the interactions between other proteins, and can have downstream effects on pathways relying on protein-protein interactions. Investigation of protein aggregation in *C9orf72* ALS may be crucial to understanding the disease mechanism, especially since other major neurodegenerative diseases, including Alzheimer's and Parkinson's disease have a robust protein aggregation pathological phenotype. This pathology has immediate implications for activity linked to the autophagy and unfolded protein response (UPR) pathways, in recycling and disposal of superfluous protein, as has been thoroughly investigated in these fields (Lynch-Day et al., 2012; Uddin et al., 2018).

The molecular characteristics of DPRs have been studied by examining the dynamics of the aggregates formed by the dipeptides by liquid droplet phase separation experiments, and examining the proteins interacting with these aggregates (Lee et al., 2016). This study was crucial in linking DPRs to several potential pathways that could be disrupted in *C9orf72* ALS (Lee et al., 2016). Each of the five DPRs was overexpressed in HEK293-T cells ($[DPR]_{50}$-GFP) to pulldown bound proteins and perform liquid chromatography tandem mass spectrometry, to obtain the independent interactomes. Green fluorescent protein (GFP)-only expression and pulldown was used as a control for background noise. Strikingly, the arginine-containing DPRs pulled down the most interactors, with no significant difference for poly-GP, poly-GA or poly-PA from the GFP-only pulldown (Lee et al., 2016). A large number of proteins associated with RNA-binding protein complexes, spliceosome, nuclear pore complex and granules were identified. An RNA interference (RNAi) screen of these targets was performed on *Drosophila* expressing poly-GR_{50} under an eye-specific promoter, GMR-GAL4, which presented a severe rough eye phenotype. Several of the interactome hits were modifiers of the phenotype. In addition, the function of these interactors suggests specific cellular localisation of the DPRs. Immunostaining of cellular organelles, including the Golgi body, nucleus and nucleolus and visualisation of overexpressed DPRs with wither GFP or mCherry tag was done using super resolution microscopy (Lee et al., 2016). Poly-GA, poly-GR and poly-PR were shown to form cytoplasmic aggregates, with the latter two also showing nuclear immunostaining. Specific staining of the fibrillar centre, dense fibrillar centre and granular component of the nucleolus allowed the authors to see that poly-GR partially localised at the fibrillar centre and dense fibrillar centre (Lee et al., 2016). The nucleolus is a distinct

organelle involved in ribosome biogenesis, and crucial for downstream translation; also suggested in Kwon et al., 2014. It should be noted that these experiments were carried out in HeLa cells, and though highly conserved, certain pathways may have slight differences in regulation in neurons, which should be further examined. A time-lapse photobleaching experiment was conducted to examine the dynamics of each DPR species (Lee et al., 2016). The arginine-containing peptides showed slower, but eventual, reformation of aggregates following bleaching, compared to GFP. However, recovery of non-arginine-containing peptides was not observed and thus suggested to the less dynamic (Lee et al., 2016). This supports the theory that the more hydrophobic amino acids aggregate naturally, but charged compounds actively interact with other proteins. The titration of synthetically-produced poly-DPRs was done to determine the effect of on the liquid-liquid phase separation (LLPS) of the granular component protein, NPM1 (Lee et al., 2016). NPM1 usually undergoes LLPS with other binding partners, such as SURF6. However, addition of increasing concentration of arginine-containing DPRs interfered with this process. The other DPRs did not have the effect of causing LLPS (Lee et al., 2016). Finally, the DPRs were examined for colocalization with stress granule markers (TDP-43 and eIF4G), for which poly-GR was positive, and a puromycin incorporation assay was used to show decreased translation in poly-DPR-expressing HeLa cells (Lee et al., 2016). Poly-PR was implicated to have a role in stress granules in a separate study (Boeynaems et al., 2017).

The majority of studies published to date show that arginine-containing DPRs are likely to play a crucial role in gain-of-function toxicity in *C9orf72* ALS, implying that the remaining three are bystanders and functionally benign. Aggregates of poly-GA

have been seen in clinical samples (Mori et al., 2013a) and models of DPR overexpression (Lee et al., 2016). Some studies have found a role in ALS pathology for poly-GA in a mammalian animal model (Zhang et al., 2016) (section 1.5.3).

Further roles of *C9orf72*-induced DPRs in ALS are discussed below, in a function-specific context.

1.5.3 Nucleocytoplasmic transport dysfunction

The presence of apparently toxic DPRs clearly has distinct gain-of-function effects on affected cells in *C9orf72* models of ALS. Their presence itself implies the translation of immature transcripts retaining the intron 1a, which must follow the transport of these transcripts from the nucleus to the cytoplasm. As previously mentioned, a mechanism of altered transcript transport has been proposed in (Cooper-Knock et al., 2014). Nuclear adaptor proteins ALYREF and SRSF1 have been identified as interactors of $[GGGGCC]_n$ and antisense, $[CCCCGG]_n$-containing RNA foci (Lee et al., 2013; Cooper-Knock et al., 2014; Hautbergue et al., 2017). *ALYREF* knockdown in fly and *C. elegans* models results in some defects in nuclear transport and development, respectively (Longman et al., 2003; Berson et al., 2019). However, knockdown in human cells is thought to be moderately redundant (Katahira et al., 2009). Similarly, the serine-arginine splicing factor family of proteins collectively moderate the transport of various mRNAs, and loss of one is thought to generally be balanced by another (Müller-McNicoll et al., 2016). Collectively with these data, Hautbergue et al., 2017 hypothesised that with the knockdown of *ALYREF* or *SRSF1*, which appears to be specific for HRE-transcript transcript nuclear export, the remaining proteins from the serine-arginine splicing

factor family will compensate for the transport of mature mRNA, but block mutant transcripts. A *Drosophila* model expressing $[GGGGCC]_{36}$ was used in this study. Flies expressing the HRE in the eyes were crossed with RNAi strains for the two given genes (fly orthologues), to observe alterations to the rough eye phenotype induced by the HRE (Hautbergue et al., 2017). Targeted RNA interference of SRSF1 with two different strains partially rescues the phenotype. Targeted HRE expression in *Drosophila* motor neurons resulted in motor defects, but SRSF1 RNAi also partially rescued this phenotype (Hautbergue et al., 2017). Immunostaining of protein suspension from flies expressing $[GGGGCC]_{36}$ by the dot-blot method showed the expression of poly-GP DPR, and depletion when RNAi of SRSF1 was performed, suggesting inhibition of transcript transport, resulting in reduced RAN translation to generate the toxic DPRs. To demonstrate that transport was specific to repeat-containing transcripts, flies expressing DPRs via alternative codons were cross with RNAi strains and did not show correction of the rough eye phenotype (Hautbergue et al., 2017). To assess the role of SRSF1 in a human context, mouse primary motor neuron and iAstrocyte co-culture (derived from *C9orf72* patient fibroblasts) was used to knockdown the human transcript via lentiviral miRNA expression in the astrocytes. Quantification showed increased motor neuron survival showed with SRSRF1 knockdown (Hautbergue et al., 2017). Furthermore, while previously the iAstrocytes stained for both nuclear and cytoplasmic *C9orf72* HRE foci (done using FISH), after SRSF1 knockdown, the ratio of the number of nuclear to cytoplasmic foci increased. It should be noted that this study performed knockdown in astrocyte co-cultures as no motor neuron survival is present in iNeuron cultures alone. It is now clear how SRSF1 would affect survival in motor neurons that also have the *C9orf72* mutation.

These authors previously performed a number of studies to understand the nature of families of proteins involved in nuclear export of RNA. From this, NXF1 is also required to form the protein complex with splice factors and mRNAs and promote nuclear export (Viphakone et al., 2012). NXF1 is also known to interact with SRSF1 via a particular chain of amino acids (Tintaru et al., 2007). Investigation of the depletion of SRSF1 from mouse N2A neuroblastomas with HRE overexpression showed that DPRs could not form in the cytoplasm, then further overexpression with a SRSF1 mutant, not containing the crucial linker amino acids showed that DPR expression could still not be restored. This demonstrated a crucial interaction between SRSF1 and NXF1 for *C9orf72* repeat transcript export, thus showing that sequestration of RNA binding proteins may trigger downstream toxic gain-of-function processes (Hautbergue et al., 2017). The majority of the work demonstrating the role of SRSF1 and interacting proteins has been shown in non-*C9orf72* ALS patient models. Future experiments, including those of other groups, may wish to confirm the co-staining of SRSF1 in nuclear RNA foci of iPSC-derived motor neurons as well as post-mortem tissue. This pathway of *C9orf72* transcripts containing the HRE is a crucial pathway to understand, as it could provide a point of therapeutic intervention as a means of therapy, preventing the RAN translation of potentially toxic DPRs, whose own pathological effects on nucleocytoplasmic transport is discussed further below.

One of the first pathways shown to have a defect as a result of the $[GGGGCC]_n$ repeat and DPRs was nucleocytoplasmic transport. Three studies published within a year of each other, used either *Drosophila* or *Saccharomyces cerevisiae* (yeast)

to perform screening studies (Freibaum et al., 2015; Jovičić et al., 2015; Zhang et al., 2015), and examine this hypothesis in depth. The overexpression of the *C9orf72* repeat (8-58 repeats) in *Drosophila* studies, were used to test the effect on the rough eye phenotype following knockdowns via RNA interference or genetic crosses with flies with chromosomal deficiencies (Freibaum et al., 2015; Zhang et al., 2015). These studies used the GMR-GAL4 or OK371-GAL4 driver to overexpress $[GGGGCC]_n$ in the eye or motor neurons, respectively. A number of enhancers of the rough eye phenotype were found to be associated with nuclear pore complex, specifically RNA export and protein import (Freibaum et al., 2015). Specifically, a number of Nup proteins involved in nuclear protein import were shown to enhance eye degradation. Evidence of nuclear RNA retention in *C9orf72* patient-derived iPSC-neurons, but not fibroblasts, suggested a role for RNA export proteins implicated as enhancers (Freibaum et al., 2015). The parallel *Drosophila* study used crosses with various RNAi homologues to screen for enhancers and suppressors of the fly eye phenotype (Zhang et al., 2015). The most significant hit was RanGAP (human orthologue, RanGAP1), which when knocked-out in the screen, worsened the phenotype. RanGAP is localised in the cytoplasm, with a role in active transport of protein import or export, via activation of Ran GTPase (GTP to GDP conversion). Knockdown also enhanced eye degeneration phenotype in an age-dependant manner (Zhang et al., 2015). Flies with motor neuron-specific knockdown (OK371-GAL4 driver) showed locomotor function defects and decreased neuromuscular junctions, including decreased neurotransmitter release potency, which could not be reversed with RanGAP overexpression (Zhang et al., 2015). Furthermore, the localisation of Ran protein, which is normally nuclear (dependant on active transport), was shown to have increased cytoplasmic

retention in *C9orf72* iPSC-neurons compared with controls (Zhang et al., 2015). Together, these papers suggested a novel role for defective nucleocytoplasmic transport, linked to the *C9orf72* repeat. These studies have drawbacks in that they rely on over-expression of the *C9orf72* repeat, as well as partial validation in non-motor neuron lines. However, the evidence of the mislocalisation of TDP-43, along with Ran in cortical neurons from an FTD mouse model, and use of ASOs targeting the *C9orf72* repeat to show rescue of Ran nuclear localisation in patient iPSC-neurons, suggests an impairment of nucleocytoplasmic transport, which is likely to be important in overall pathology (Zhang et al., 2015).

The standalone yeast screening paper published with these two *Drosophila* papers also performed an unbiased screen to look for targets altering toxicity induced by DPR expression (Jovičić et al., 2015). In contrast to the fly work, this study overexpressed variants of the DPRs under a galactose-inducible promoter using codons other than those contained in the repeat (100 amino acids in length each), in order to examine the specific effect of the apparently toxic peptides, omitting the effect of repeat-containing RNAs from the study (Jovičić et al., 2015). While, low levels of the DPRs could be detected in Freibaum et al., 2015 and Zhang et al., 2015, it was not firmly established whether the repeat RNA or peptides were causing the transport defects. Yeast growth assays showed a stronger toxicity for the arginine-containing peptides, as consistent with previous studies (Kwon et al., 2014; Mizielinska et al., 2014; Wen et al., 2014; Lee et al., 2016). Poly-PR showed the strongest toxicity, and was used for further assays. The first test to screen enhancing or suppressing yeast growth was done by transforming plasmids encoding 5,500 genes into yeast colonies, already expressing poly-PR. The second

yeast was done by transforming yeast strains with non-lethal gene knockouts with poly-PR. Instances improving yeast growth were investigated further. A number of proteins involved in nuclear protein import were identified, though these did not fall under similar gene families as in the *Drosophila* work. Importin-1 (karyopherin β2 homologue in yeast), involved in nuclear import of ALS-linked protein, FUS, was identified as suppressor of toxicity when overexpressed. In addition, an number of proteins involved in ribosome biogenesis, RNA binding and transcription were selected as significant hits in the yeast growth phenotype (Jovičić et al., 2015). A protein involved in the import of serine-arginine domain-containing proteins into the nucleus, Mtr10p, was found to strongly increase yeast growth if overexpressed. Finally, in consistence with Freibaum et al., 2015, a protein involved in active nuclear protein transport, Ran guanine-nucleotide exchange factor, RCC1 (Jovičić et al., 2015). However, yeast growth rate was reduced in contrast to RanGAP, whose overexpression rescued the rough eye phenotype. Validation of RCC1 expression in *C9orf72* ALS patient iNeurons showed that the protein was depleted from the nucleus compared with age-matched controls (Jovičić et al., 2015). Although, RanGAP1 staining was not different between patients and controls, in contrast to Freibaum et al., 2015. These studies together point to a defect in nucleocytoplasmic transport due to pathological features seen in *C9orf72* ALS, but the proteins directly causing toxicity is not clear.

Poly-GA pathology has also been linked with the nuclear transport of TDP-43, whose pathology is seen in more than 90% of familial and sporadic ALS cases. One study co-transfected HeLa cells with a combination of the TDP-43 nuclear localisation signal sequence (with an RFP reporter), and various poly-DPR

reporters, combined with cellular fractionation to show that poly-GA disrupts the localisation of TDP-43 (Khosravi et al., 2017). Another study showed the sequestration of several crucial nucleocytoplasmic pore complex proteins at TDP-43-positive granules, using proximity ligation assays and co-immunoprecipitation assays from N2A neuroblastomas (Chou et al., 2018). These results were further validated using immunohistological studies in the post-mortem tissue of patients with phoso-TDP-43 inclusions. This shows a potential link between two of the most abundant ALS mutations, showing how the toxic gain-of-function phenotype as a result of the *C9orf72* HRE results in the mislocalisation of TDP-43, but leading to its accumulation in the cytoplasm and depletion from the nucleus, where its function is required. This would disrupt RNA editing events, some of which may be linked to the *C9orf72* transcripts themselves, which accumulate into foci and themselves sequester other RNA binding proteins. Finally, Zhang et al., 2016 generated a mouse expressing poly-GA^{50} to demonstrate the presence of ubiquitin-positive aggregates, accompanied by significant neuronal loss and a defective motor phenotype. Cortex neuronal inclusions were positive for poly-GA, and overlapped with proteins involved in proteasomal degradation, HR23A and HR23B, as seen by immunohistological staining. These proteins are usually not sequestered, and this phenotype was accompanied by an increase in apoptotic markers in primary mouse neurons. This was subsequently rescued by the exogenous expression of the two proteins, suggesting that they contribute to the overall toxicity triggered by poly-GA (Zhang et al., 2016).

1.5.4 Aberrant calcium homeostasis and glutamate-induced toxicity

Pathologies associated with *C9orf72* ALS that have been less explored relate to aberrant calcium regulation and endoplasmic reticulum (ER) function. One of the first studies to examine the pathways of calcium dysregulation in human iPSC-motor neurons, showed increased levels of free calcium ions in the cytoplasm of *C9orf72* ALS patient neurons (Dafinca et al., 2016). Thapsigargin was used to block an ER calcium uptake protein, and the iPSC-motor neurons were treated with a dye to stain for calcium ions, and live imaging was used to analyse the dye intensity. The study also examined a key protein involved in calcium regulation, Bcl-2, which was found to have reduced levels in patient neurons, using immunoblotting. As sustained, high cytoplasmic levels of calcium are crucial to triggering apoptosis, these changes to protein expression in patient motor neurons may be a crucial threshold for cell death in ALS. However, this phenotype is likely to occur in later stages of ALS, as it directly precedes the activation of factors which cause mitochondrial membrane permeability to release cytochrome C, in the final stages of apoptosis. The patient iPSC-motor neurons from this study also had elevated levels of cytochrome C, cleaved caspase-3 and p62-positive granules.

A separate perspective on calcium regulation stems from a study examining the expression of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors in *C9orf72* ALS patient iPSC-motor neurons compared with isogenic edited controls (Selvaraj et al., 2018). This study shows a difference in the expression of subunits of AMPA receptors in their neurons as the cultures age. RNA sequencing and qRT-PCR validation showed an increased expression of GluA1 subunit in the mutant compared with edited and age-matched controls. This switch

leads to the expression of more calcium-permeable AMPA receptors, suggesting a mechanism for glutamate excitotoxicity. This is in agreement with Bursch et al., 2019, which also saw that GluA1 levels were increased, compared with controls, though not as significantly.

1.5.5 TDP-43 mislocalisation

Although no cases of *C9orf72* ALS that also have a TDP-43 mutation have been noted, the presence of TDP-43-positive stress granules in 90% of ALS and FTD patient brain histology studies, shows that this protein is likely to play a greater role outside of patients carrying a mutation. One mouse model for the *C9orf72* mutation detected mislocalised TDP-43 in the mouse frontal cortex and hippocampus at 6 months, in conjunction with NeuN-positive neuronal loss in the same regions (Chew et al., 2015). However, no significant loss of spinal motor neurons or motor deficits could be identified in this model. As previously described (section 1.5.3), proteomic analysis of cells with TDP-43 pathology showed a difference in expression of several proteins vital for nucleocytoplasmic transport via the nuclear pore complex, particular the Nup family of protein importers (Chou et al., 2018). A potential mechanism linking TDP-43 with *C9orf72* repeat-containing transcripts was suggested by Cooper-Knock et al., 2015a, who found a correlation between the presence of nuclear antisense foci and depletion of TDP-43 from the nucleus. This study did not give evidence for a mechanism of action. A later study by a separate group, used a *C9orf72* ALS mouse model containing $[GGGGCC]_{149}$, with several typical pathological phenotypes, to show the presence of both sense and antisense RNA foci and an age-linked increasing accumulation of RAN translation DPRs (Chew et al., 2019). These mice also had phospho-TDP-43 inclusions and

mislocalised RanGAP1, which was previously been identified in a *Drosophila* screen as an enhancers of the rough eye phenotype when knocked down (Zhang et al., 2015). This, along with Khosravi et al., 2017, suggested a further link between the pathologies where the gain-of-function effect from the *C9orf72* HRE resulted in the mislocalisation of TDP-43, which would have further consequences on its RNA binding functions. A clinical histological study of patients diagnosed with either ALS or FTLD from the same family, found that cases with the *C9orf72* mutation had widespread DPR pathology in the hippocampus, cerebral cortex and cerebellum (regions unaffected in disease) (Baborie et al., 2015). They found that family members who died earlier from their disease did not have TDP-43 granules, while those who died later had sparse, but some evidence of this pathology. The authors hypothesise that *toxic* DPRs must form before a TDP-43 phenotype can be observed. Finally, a recent study used a *Drosophila* model to show a toxic phenotype with the expression of DPRs under non-canonical codons, and linked this to the emergence of a TDP-43 phenotype (*TBPH*; fly orthologue), as well as the mislocalisation of karyopherin-α, an importin-like protein (Solomon et al., 2018). Flies expressing poly-GA and poly-GP developed a motor phenotype later than those expressing poly-GR (early onset and fast progression). TBPH mislocalisation only occurred in flies which were expressing the DPRs for longer, suggesting mislocalisation was required before neurodegeneration occurred. This was further verified by expressing *TBPH*, with a mutation in the nuclear localisation sequence, in the flies, finding that they developed a motor phenotype in a dose-dependent manner. This model was also used to show karyopherin-α mislocalisation to the cytoplasm, which did not necessarily occur with only DPR expression, suggesting a trigger for the loss-of-function of multiple proteins as a direct result of TBPH in the

cytoplasm. A link between TDP-43 and karyopherin-α has been previously shown in (Freibaum et al., 2015).

1.5.6 Loss-of-function and endosomal pathway defects

A loss-of-function mechanism for disease has been a recurring argument for the cause of the *C9orf72* ALS disease subtype, despite seemingly stronger evidence for a toxic gain-of-function mechanism. The mouse *C9orf72* orthologue transcript is enriched in neurons known to degenerate in ALS and FTD (Suzuki et al., 2013), suggesting a tissue specificity in a mammalian system. However, knockout of the *C9orf72* orthologue gene in a mouse model is not fatal, nor leads to a motor phenotype (Sudria-Lopez et al., 2016), though it does in a zebrafish model (Ciura et al., 2013). This suggests that loss-of-function is unlikely to be the sole mechanism of disease.

Attempts to assess the function of *C9orf72* by generating in-house antibodies and determining cellular localisation in *C9orf72* ALS, ALS/ FTD and sporadic ALS post-mortem tissue, have been done (Xiao et al., 2015). The protein functions of the known long and short isoform transcripts of *C9orf72* are unknown, thus short synthetic peptides unique to each isoform were generated as the antigens to raise polyclonal antibodies against, in rabbit (Xiao et al., 2015). Although useful to distinguish between the protein isoforms, generating antibodies against a truncated peptide may prove to generate unreliable antibodies, which may not recognise full-length, folded C9orf72 proteins. Immunoblotting of protein from post-mortem tissue was used to show that higher levels of the long isoform are present in the frontal cortex of non-*C9orf72*-related cases of ALS compared with *C9orf72* ALS patients

(Xiao et al., 2015). In contrast, stronger bands for the short isoform were seen in the temporal and frontal cortex in *C9orf72* ALS patients. No differences in band intensity were seen in either isoform in the motor cortex, lumbar spinal cord or cerebellum, despite being key regions where molecular pathology for *C9orf72* ALS are seen. This study also examined *C9orf72* isoform cellular localisation using immunostaining, and showed cytoplasmic staining of the long isoform and perinuclear for the short, suggesting different functions. There were distinct staining patterns for the long isoform in the cerebellum compared with spinal cord. The latter only had a diffuse staining pattern, while the cerebellum also contained granules. Interestingly, this staining was seen in both *C9orf72* ALS and non-*C9orf72* ALS tissue. Quantification of staining intensity in ALS tissue nuclei, compared with controls, was significantly reduced. Due to previous studies proposing *C9orf72* as having a role in autophagy, various antibodies for endosomal and lysosomal markers were used to co-stain with the long isoform antibody, though no overlap was observed. Some correlation studies of TDP-43 staining with *C9orf72* showed that the mislocalisation of the former in ALS tissue occurred when short *C9orf72* staining became stronger at the perinucleus and depleted from the nucleus (Xiao et al., 2015). Finally, this paper showed that both synthetic peptide isoforms of *C9orf72* coimmunoprecipitated with nucleocytoplasmic transport proteins, Importin-β1 and RanGTPase, from overexpression in mouse N2A neuroblastoma cells. These results have yet to be replicated in a human model, from the full length proteins. This study shows a clear difference in cellular localisation for the two protein variants of *C9orf72* in ALS patient tissue, however their functions could not be confirmed.

The function of the C9orf72 protein product, if at all important, has remained elusive (Levine et al., 2013). A general decrease in *C9orf72* transcripts has been noted in patient cells, but a linked loss-of-function mutation-driven phenotype has not been convincingly validated. The C9orf72 protein has been suggested to have a role in endosomal pathways, potentially interacting with the Rab family of GTPases (Levine et al., 2013; Aoki et al., 2017), to which *C9orf72* has been predicted contain a related DENN-like domain (Levine et al., 2013). Proteins containing DENN-like domains have previously been suggested to be key regulators of Rab proteins (Marat et al., 2011). However, the lack of a robust antibody has made studying the function of C9orf72 protein challenging. The first study to provide evidence for this link to a pathway showed that *C9orf72* colocalised with Rab5, 7 and 11 in murine Neuro2A cells, SH-SY5Y cells, mouse primary cortical neurons and in post-mortem human spinal cord (Farg et al., 2014). This was done using immunostaining of fixed cells or tissue. This study also performed short interfering RNA (siRNA) knockdown of *C9orf72* to show the inhibition of Shiga toxin (conjugated to a dye) uptake from the cell membrane to the Golgi body using GM-130 staining. Various members of the Rab family are known to be involved in different molecular pathways associated with metabolism, autophagy and, vesicle transport and secretion, specifically as a tether molecule between these processes (Chua and Tang, 2015). For instance, Rab5 has been implicated in trafficking of early endosome components and recruitment of epidermal growth factor receptors (EGFRs) from the plasma membrane (Chen et al., 2009). Rab7 has a role in the degradation of EGFRs by transporting them from the late endosome to the lysosome (Ceresa and Bahr, 2006). Conversely, Rab11 has been previously shown to play a role in the transport of EGFRs back to the plasma membrane, via endosomes where proteins are repaired

and recycled (Cullis et al., 2002). Rab proteins have several modulating proteins involved in GTP/ GDP exchange, the alter their activation states. Many of these are known to contain DENN-like domains (Marat et al., 2011). These modulating proteins can be shared between different members of the Rab family (Chua and Tang, 2015). Under endogenous conditions in mammalian cells, more than 60 Rab family members are known to have unique cellular localisations, implying specific functions (Farg et al., 2014). The degree of redundancy between family members is not known, thus the effect of a mutation in a regulating protein such as *C9orf72* may not be as dramatic as being the sole cause of motor neuron degeneration. However, functional roles of *C9orf72* should be understood in detail to provide more knowledge regarding the regulation of the gene, its regulatory effect on other proteins and pathways to target for drug intervention.

Aoki et al., 2017 showed that *C9orf72* preferentially co-immunoprecipitated with Rab7L1 (Rab29) by co-overexpressing various members of the human GST-tagged Rab family of GTPases and *C9orf72*-c-myc-FLAG, via plasmid transfection in the SH-SY5Y neuroblastoma cell line. The results were in contradiction with Farg et al., 2014, despite using the same antibody in experiments, where Rab11, 5 and 7 showed weak or no pulldown with *C9orf72*. Rab7L1 is known to have a role in late endosomal transport of exosomes, which were shown to have decreased in secretion from cells with either Rab7L1 or *C9orf72* knockdown by siRNA (Aoki et al., 2017). This was measured by nanoparticle tracking analysis and immunoblotting analysis of common exosome membrane markers, Alix and Tsg101. This suggested a cooperative role for the two proteins in this pathway, as well as a loss-of-function pathology. This pathology was also seen in fibroblasts grown from

skin biopsies of ALS patients with a *C9orf72* mutation (Aoki et al., 2017). ASO knockdown of repeat-containing transcripts using locked nucleic acid (LNA) gapmers rescued this secretion phenotype (Aoki et al., 2017). However, the cargo of Rab7L1-*C9orf72*-related exosomes, or how this decreased secretion may affect local molecular signalling with other cell types in the spinal cord is unknown. Proteomic analysis of exosomes purified from a human model, such as *C9orf72* ALS patient iPSC-derived motor neurons, may provide important clues to the loss-of-function effect of the HRE mutation.

1.5.7 Spectrum with frontotemporal dementia (FTD)

It has been long known that FTD has a similar pathological phenotypes in diseased tissue, in the form of ubiquitinated aggregates and shared genetic mutations (Morita et al., 2006; Neumann et al., 2006). Many cases of patients are known, who develop both the motor defects characteristic of ALS, and the cognitive and behavioural traits seen in FTD, suggesting a shared mechanism of disease (Lomen-Hoerth et al., 2002). FTD is distinguished by atrophy of the frontal lobe of the cortex. When the *C9orf72* HRE mutation was discovered, it was quickly established to be inherited by FTD patients as well (DeJesus-Hernandez et al., 2011). As with ALS, FTD is known to have gene mutations which only cause dementia-related symptoms, but some mutations have been identified in patients who develop both diseases. The most common of these is *C9orf72*, which occurs in around 10% of FTD patients, similar to ALS. This shared mutation has led to the hypothesis that similar cellular pathways are perturbed, which lead to a targeted degeneration of neurons (Ling et al., 2013). It is still not clear why this spectrum is seen in patients who carry the *C9orf72* mutation, rather than causing one disease. It has been

hypothesised that additional genetic or environmental factors are required, such as a family history where other neurodegenerative diseases have been recorded. A common perturbed pathway seen in Alzheimer's and Parkinson's disease, along with ALS and FTD is defects in the autophagy pathway, though this is thought to be a late-stage phenotype. Potentially, patients with non-lethal gene variants which determine a certain disadvantage for coping with cellular stress have a reduced chance of coping with adult-onset diseases which become more challenging for ageing the nervous system to deal with. Nevertheless, this key link between ALS and FTD means that any key advances made in the respective fields are likely to be highly useful in progressing towards treatments in the other.

1.6 *C9orf72* ALS has been characterised in a number of versatile models

A wide variety of models have been used to try and understand the molecular mechanisms and clinical features associated with *C9orf72* ALS. Though not perfect, each in its own way can be highly useful in finding an answer for specific question, and results from models are usually strongest when validated by another. This is especially true when studying neurodegenerative diseases, where obtaining CNS tissue from patients while still alive is not possible due to inaccessibility and danger of exacerbating the disease. The familial component of *C9orf72* ALS has led to several genetically modified animals and cell lines to serve as a model to investigate now the mutation in question causes neurodegeneration in the spinal cord and the occurrence of pathological cellular pathways, respectively.

Common models, which are easy to manipulate genetically, and provide a platform for rapid data output include zebrafish (*Danio rerio*) (Swinnen et al., 2018),

Drosophila melanogaster (Tran et al., 2015; Boeynaems et al., 2016) and *Caenorhabditis elegans* (*C. elegans*). The majority of gene families, even in invertebrates, are functionally conserved up to higher mammals. This makes these models useful for the genetic modification of orthologue genes to the human counterpart, thought to be important in ALS. Specifically, hundreds to thousands of *Drosophila* and *C. elegans* strains can be generated by a relatively small number of people and analysed for different phenotypes. This includes motor phenotype screening, cellular pathology, survival assays, and treatments to test the rescue of pathological phenotypes. However, a major drawback of these models is their significantly diminished CNS complexity compared with higher mammals. Replicating the conditions which initiate a disease cascade as seemingly complex as *C9orf72* ALS is highly challenging. This has increased the need for a mammalian model, which can corroborate the results found in experiments from invertebrates and zebrafish.

1.6.1 Mouse models of *C9orf72* ALS

Historically, genetically modified mice have frequently been the model of choice while studying human gene mutations in disease and efficacy of potential treatments. This is due to the strong similarity of gene function, the presence of gene orthologues and around 85% protein coding gene homology between humans and rodents. Nearly all ALS-linked mutations have been modelled in rodents, usually replicating tissue pathology seen in patients and also exhibiting defective motor and behavioural phenotypes. Though rarer cases of younger adult onset are seen, a crucial aspect of ALS is that it is widely seen as an ageing disease, and several established mouse models represent extreme overexpression of certain

mutated genes resulting in mice that are days or weeks old displaying pathological phenotypes. The *SOD1* and TDP-43 fields particularly have widely used rodent models of this kind to form conclusions about the disease. While powerful in understanding how the respective mutated proteins interact, localise and affect surrounding tissue, ALS is an insidious disease, manifesting in later life. This suggests that these overly toxic models are unlikely to represent cellular pathways which may serve as better pharmacological targets.

Several mouse models for *C9orf72* ALS have been published, generated by different techniques (Chew et al., 2015; O'Rourke et al., 2015; Peters et al., 2015; Jiang et al., 2016; Liu et al., 2016; Sudria-Lopez et al., 2016). The knockout of the mouse *C9orf72* orthologue gene (Sudria-Lopez et al., 2016), generation of human *C9orf72* BAC transgenic mice (O'Rourke et al., 2015; Peters et al., 2015; Liu et al., 2016) and injection of AAVs with the *C9orf72* HRE (Chew et al., 2015). Each were generated with specific questions in mind. The knockout mouse, which does not develop any typical ALS phenotypes, highlights that the *C9orf72* HRE is unlikely to be a loss-of-function mutation. The models incorporating the HRE are trying to answer whether the mutation alone is enough to induce motor and pathological defects. While several show typical *C9orf72* ALS neuropathology, including nuclear RNA foci and DPRs, only one displays any motor phenotype (Liu et al., 2016). Additionally, neuronal loss in brain regions unaffected in human patients, including cerebellar Purkinje neurons (Chew et al., 2015) and hippocampal neurons (Jiang et al., 2016) were seen in these models. The model published by Liu et al., 2016 was most representative of the human disease, showing mice with severe and slow progressing disease. This study used pronuclear injection to create a stable mouse

line from an expression construct cloned from a patient BAC library, with the full human *C9orf72* gene locus ($[GGGGCC]_{500}$) including regulatory and promoter regions. This mouse developed gait abnormalities, anxiety, cerebellar neuron loss, significant spinal cord neuron degeneration and muscle denervation. It also displayed typical cellular pathologies, including mislocalised, TDP-43-positive granules. Interestingly, every mouse which inherited the HRE in a litter eventually presented motor deficit symptoms without exception. Though not fully investigated, it is not clear whether every human who carries a large *C9orf72* HRE will develop ALS, thus making it likely for genetic background to play a role in manifestation of the disease. This cannot be easily studied in mice, which are generally thoroughly inbred, and often contain numerous genetic modifications in order to maintain a steady genetic background for a controlled experiment. This is highly useful for simplistic study of the effect of the *C9orf72* mutation, but will likely lose its power when drug screens are performed on such mice. Patients often respond in very different ways to different therapeutic inventions, which is increasingly pushing the argument for clinical trials where drugs have been pre-screened on human cellular models first (personalised medicine). Distinctly, a model missing from this set is the overexpression of DPRs from alternative codons to test whether the presence of these peptides is enough for ALS-like symptoms to manifest, as it has been in other models (Hautbergue et al., 2017). This would be a crucial experiment to understand ALS disease progression in mammals.

1.6.2 Patient post-mortem tissue as a validation tool

The brains and spinal cord of patients who have donated their tissue to scientific research has been the most common tool for the analysis of neurons affected by

disease, particularly in identifying end-stage pathology (Baborie et al., 2015; Mackenzie et al., 2015). Analysis of post-mortem tissue can be highly instructive when grouping patients and performing analysis on their pathology in several ways, such as by disease progression, age of symptom emergence, genetics, and cellular pathology (van Blitterswijk et al., 2013, 2015). Analysis of patients according to what kind of treatments they were administered during their disease, and end stage analysis could also be beneficial to understanding drug efficacy. This medium is often used to validate novel protein or RNA expression changes identified from high throughput RNA sequencing, proteomics and other screens (Haeusler et al., 2014; Selvaraj et al., 2018; Shaw et al., 2018). Post-mortem tissue is also commonly used to perform many of these high throughput experiments, though this is often challenging due to the complexity of brain tissue, and difficulty in extracting the target cell type in question (Lobsiger et al., 2007). Furthermore, depending on the preservation technique used for the tissue, purification of RNA and protein in high yields and sufficient quality is often tough. At present, the characteristics of *C9orf72* ALS tissue has been studied at great depth, but one of the least understood aspects of the disease is the mechanism of disease progression at a cellular level. Particularly, the mechanisms occurring at a pre-symptomatic stage are missing from our knowledge, but post-mortem tissue analysis is not suitable for this.

1.7 Induced pluripotent stem cell-derived motor neurons from patients recapitulate several pathological phenotypes of *C9orf72* ALS

An increasing number of studies in the *C9orf72* ALS field have shifted to studying cells directly from patients, especially iPSC-derived motor neurons and iNeurons. These studies either perform induction of patient skin biopsy-derived fibroblasts to

pluripotency using lentiviral transduction of Yamanaka transcription factor plasmids (Takahashi and Yamanaka, 2006) followed by neuronal differentiation, or perform direct induction to neural progenitors (Meyer et al., 2014). There are many advantages of this technique. Firstly, motor neurons from the patient, with all their genetic background, may be studied. Motor neurons can be obtained from iPSCs or fibroblasts in a relatively short amount of time, compared with mouse models, and also comparatively more cheaply (though the cost rapidly increases with the number of lines used). Patient-derived iPSC-motor neurons have been shown to express the same molecular markers characteristic of human spinal motor neurons, including homeobox 9 (HB9), islet 1, choline acetyl transferase (ChAT) and heavy neuronal filaments, beta-III tubulin and SMI-32. As a cell culture platform, they can be easily used in a variety of drug treatments and manipulated using genetic editing strategies such as, clustered regularly interspaced short palindromic repeats (CRISPR)/ Cas9 cleavage, and transcript knockdowns, using siRNA or ASOs. Most importantly, iPSC-motor neurons recapitulate several pathological phenotypes relating to *C9orf72* ALS (described in section 1.5), which have also been identified in patient post-mortem tissue. This includes the differentiation protocol optimised in this group (Dafinca et al., 2016). Specifically, nuclear RNA foci (sense and antisense), DPRs, defects in calcium homeostasis and increased apoptotic factors. Crucially, these neuronal cultures are also electrophysiologically active and can be stimulated with glutamate to fire action potentials, recorded using whole electrode voltage clamp (Dafinca et al., 2016). Patient-derived iPSC-motor neurons are be stimulated as well as the controls, though if the latter are exposed to glutamate for excessive amounts of time, they take longer to recover to resting membrane potential.

Despite the many benefits, some disadvantages also exist for this model. Firstly, though some differentiation protocols can generate up to 90% neurons expressing motor neurons markers, this process can be highly variable. Several protocols for the differentiation of motor neurons from iPSCs have been published, and often exclusively used by the group which performed the optimisation. This increases the variability of data in the field, and often novel results reported in one batch of iPSC-motor neurons cannot be replicated in another group's hands. Also, some iPSC cohorts use different cells to induce pluripotency. Blood and fibroblasts are most commonly used as the starting cell type. This could have unexpected effects on the condition of the derived iPSCs. For instance, the methylation state of iPSCs and subsequent differentiated motor neurons, compared with actual aged motor neurons in a human spinal cord is unknown. These epigenetic traits may be crucial to identifying the molecular mechanisms involved in a disease, like *C9orf72* ALS, which has a strong RNA aspect to the pathology. Some studies have argued that differentiated cells bear strong molecular resemblance to stem cells, and are thus *unaged*. There is also uncertainty as to the somatic stability of the *C9orf72* HRE, where one study showed that the same individual had repeats of different lengths in different tissue. Therefore, iPSCs derived from skin or blood may not represent the actual number of repeats in the motor neurons of the affected individual. Finally, motor neurons cultured in a dish are not truly representative of the complex network of cells present in the spinal cord, where oligodendrocytes, astrocytes and other glia are crucial for providing metabolic and structural support to motor neurons. However, in order to understand how a mutation like the *C9orf72* HRE leads to neurodegeneration, it is vital to understand the earlier pathological cellular

pathways which slowly manifest into a late-onset disease. Differentiated *human* motor neurons are among the best models available today that allow this.

1.8 RNA sequencing of motor neurons is crucial to understanding *C9orf72* ALS

Despite the many models and tools available to perform genetic manipulation of animal models recapitulating the motor phenotypes seen in *C9orf72* ALS, a key fact remains that the prevailing molecular pathways and gene expression changes that occur to trigger symptoms in patients are largely unknown. Mutations known to be associated with ALS are present from conception, yet do not manifest until adulthood. We do not know if there are a series of environmental effects which cause a cellular vulnerability that cannot be dealt with by mutation carries, or if varying genetic backgrounds play a role in tolerance for the mutation until ageing leads to its breakdown. We do not know if one or multiple failings of cellular pathways must occur for motor neurons to start degenerating. Finally, the reason for motor neurons having a particular susceptibility to the mutations in question (or frontal lobe neurons in FTD) is largely unknown. An important step to understanding this process is to perform transcriptional and proteomic profiling of motor neurons, preferably from a model realistically representing the disease. Various methods for global or targeted gene sequencing have long been available, though as discussed ideal models are challenging to obtain. Several prior studies have used next generation sequencing methods on cells and tissue from ALS patients to characterise changes in molecular pathology, which result in the disease. A main drawback of these methods were often the lack of targeting to a specific cell type, or use of a cellular model which is arguably irrelevant to ALS. Prudencio et al., 2015 extracted RNA from post-mortem patient cerebellum and frontal cortex, both of

which exhibit evidence of *C9orf72*-linked pathology in the form of TDP-43 and DPR-positive granules, and nuclear *C9orf72* transcript-positive foci in ALS, and degeneration in the case of ALS/ FTD patients. Extracting motor neuron RNA from ALS patients is highly challenging, as by end-stage atrophy of spinal cord tissue is very advanced. Therefore, the decision to harvest RNA from preserved tissue which display key pathologies is logical. On the other, this is unlikely to help answer the question of selective motor neuron vulnerability. It is unknown whether neurons from other regions of the brain would start to degenerate if the patient were to live longer, despite their early tolerance, but potential to present pathology, may suggest that they represent an early cellular state of ALS. This study compared the RNA profiles of 8 patients with the *C9orf72* mutation, 10 sporadic patients and 8 controls (individuals without neurodegeneration). It was recorded that patients from each group had similar disease progression, with both lower and upper spinal cord neurons affected. While a number of differentially expressed genes associated with the UPR, vesicular transport and intracellular transport were identified, no further validation in a motor neuron model was done (Prudencio et al., 2015).

Understanding the global gene expression changes that occur directly due to the presence of the *C9orf72* mutation is key to understanding why the HRE affects motor neurons uniquely. Furthermore, studying these cells from pre-symptomatic to symptomatic stage is crucial to understanding the molecular changes which occur as a result of the mutation, and are linked to the many pathological phenotypes seen *in vitro* and *in vivo*. However, the complexity of the mammalian brain makes it challenging to study these populations of neurons in isolation due to their inaccessibility and the complex nature of the CNS. Additionally, most motor

neuron differentiation culture protocols have a significant contamination of non-motor neuronal neurons, glia and cells of unknown lineage which will cloud a technique as sensitive as RNA sequencing, where background transcripts affect the significance of gene changes occurring in the motor neuron population. In an age where RNA and protein profiling are increasingly a crucial point of validation in biological studies, methods need to be developed which can be employed to perform sequencing on neuronal subpopulations of interest, while improving the drawbacks to existing techniques. In order to get around this, a number of laboratory and analysis techniques have been developed.

1.8.1 Methods for studying neurons in isolation

The isolation of neuronal subtypes from frozen of fixed tissue from rodent, or human post-mortem samples would be highly useful for the purification of RNA to perform RNA sequencing and analyse the transcriptional profiles of cells at a particular stage of disease. As described, the density and complexity of brain matter make this process very challenging.

1.8.1.1 Laser-capture microdissection

A technique known as laser-capture microdissection (LCM), coupled with single cell RNA sequencing kits (*LCM-seq*) (Nichterwitz et al., 2016), has been developed to manually extract neurons from snap-frozen cubes of tissue, which have been treated with fast-penetrating stain. Neurons are identified by morphology under a microscope and cut from the tissue using a laser. The cell is immediately lysed and purified RNA is quantified for amplification using the SMARTer Seq2 kit (Picelli et al., 2014) for single cell library preparation. This technique has improved on

previous LCM protocols, where hundreds to thousands of cells were required as enough material for a bulk RNA sequencing experiment (Lobsiger et al., 2007; Saxena et al., 2009). This increased the potential for contamination from surrounding cells in the tissue sample. The authors were able to show that a distinct transcriptional profile for motor neurons harvested using LCM-seq, from mouse pups. They compared the profiles of RNA from 1, 2, 5, 10, 30, 50 and 120 cells (Nichterwitz et al., 2016). The fewer the number of cells used, the fewer the number of glial markers that could be detected, and stronger signals for motor neuron markers including ChAT, HB9, Islet1 and 2 and SMI-32, were seen. While a crucial results in this field, LCM is still not the ideal technique for studying neurons as the technique can only chiefly isolate cell matter around the cell body, which is the large and most easily identifiable part of a neuron. The axon, branched dendrites and terminal synaptic regions will all be lost, though interesting populations of RNA are likely to be localised at these regions. It is known that translation of mRNA occurs locally at the synapse, which could be crucial for determining changes to neuronal function at the neuromuscular junction, or other synapses.

1.8.1.2 Cell sorting by marker detection

Cell types are commonly identified by the expression of a particular combination of genes, including cell surface proteins, proteins with a determining function of the cell and transcription factors. The expression of these markers is determined by the developmental lineage of a cell type. Mature motor neurons are distinguished by their expression of ChAT, which is crucial for the modulation of the neurotransmitter, acetyl choline release at the synapse, Islet 1 and 2 (transcription factors with a role in early patterning of motor neuron differentiation, by modulating the expression of

multiple genes) (Liang et al., 2011) and several neurofilament proteins such as Beta-III tubulin and SMI-32. Fluorescence-activated cell sorting (FACS) is a technique that takes advantage of this feature by using fluorescent antibodies to recognise a cell surface marker and sort the cells based on the wavelength of excitation, by the FACS instrument. Similarly, magnetic-activated cell sorting (MACS) (Welzel et al., 2015), uses antibodies conjugated to tiny magnetic particles to detect and sort cells based on magnetic attraction. FACS is a highly efficient, well characterised technique in many fields, especially haematology and immunology, where separating subpopulations of cells is of major interest. A previous study used FACS to sort and perform RNA sequencing on iPSC-motor neurons from ALS patients by, firstly overexpressing red fluorescent protein (RFP) under the HB9 promoter, and sorting based on the co-expression of RFP and cell surface marker, $p75^{NTR}$ (Toli et al., 2015). This study successfully obtained a pure population of motor neurons by pre-sorting early motor neurons which were expressing the early motor neuron marker, HB9, and replating the cells after the FACS experiment. Though a highly efficacious method for producing a cleaner differentiation, sorting maturing neurons and successfully re-establishing them in culture is very technically challenging. On the other hand, for well-established protocols, FACS could be optimised to obtain multiple cell subtypes, such as neurons and glia from a co-culture. Nevertheless, as in laser-capture microdissection, a physiological barrier is present in neurons. The dissociation of neurological tissue into individual cells, while trying to maintain their structural integrity, keeping dendrites, cell bodies, axons and synaptic junctions intact to obtain a whole cell, before sorting and purifying RNA from them, is highly challenging. Dissociation of the brain tissue or mature cell cultures often causes loss of neuronal architecture and vast amounts of

cell death, and also results in contamination in the case of tissue. Following this process, the neurons must be passed through the FACS instrument before being pelleted for RNA extraction. This is likely to induce stress that could potentially affect the transcriptome of the neurons, in addition to the stress of losing its processes. The development of techniques, which are relatively stress-free is consequently vital to the advancement of the field.

1.8.1.3 Bioinformatic analysis of mixed species cell cultures

The relevance of cell cultures to study complex niches where neurons, glia, muscle and immune cells cooperate to maintain the optimal conditions for function and metabolism has frequently been debated. As such, co-cultures are more commonly being sought as an improved model of neurodegenerative diseases where supporting astrocytes have been shown to potentially play a significant role in disease progression and manifestation. For example, a study using a co-culture of iNeurons and iAstrocytes, directly derived from ALS patient fibroblasts, showed that astrocytes with a mutant *C9orf72* HRE, cultured with control motor neurons, resulted in reduced survival of the latter in a non-cell autonomous fashion (Meyer et al., 2014). While analysis using immunostaining and imaging of neuronal survival can be easily performed in such cultures, the more intriguing changes associated with gene expression changes in neurons, as a result of toxicity cannot be performed. Co-cultures have also been developed where the different cell types are separated by pores or a mesh, but this less representative of the endogenous niche as the cells are not in close contact. Recently, coding pipelines have published which are written to perform analysis of RNA sequencing data containing reads from more than one species (Qiu et al., 2018). Using this method a co-culture of

neurons and glia from different species can be maintained, but ultimately RNA is purified in bulk, and RNA sequencing is done under one library preparation. The output reads are then separated by species by mapping to the genomes of the respective species used. The tool, *Sargasso*, has been developed to be sensitive enough to distinguish between the mouse and rat genome (Qiu et al., 2018). Though this method completely eliminates the concern of stressing live cells during the process of sorting them as with FACS, the issue remains that any contaminating cells would still bias any high throughput sequencing data.

1.8.2 Translating ribosome affinity purification

Translating ribosome affinity purification (TRAP) was originally conceived in 2008 as a technique to dissect the RNA profiles of specific cellular subtypes, followed by a detailed methods paper in 2014 (Heiman et al., 2008, 2014). The concept of the technique relies on the expression of a tagged large ribosomal subunit protein under the promoter of specific cellular marker. These exogenous tagged ribosomal proteins will incorporate into the translation pathway by incorporating into polysome complexes, and provide a point of pulldown of RNA, via the ribosomes that they are interacting with. Messenger RNA is purified by lysis of total tissue or cellular matter, and subsequent immunoprecipitation of ribosomes via the molecular tag. These mRNA represent the profile of genes being expressed at a particular moment in the target cell type, enriched for the neuronal marker of interest, referred to as the translatome. This translatome profile excludes quiescent RNA that is not being actively translated, including immature transcripts, microRNA and other non-protein coding region-derived RNA species. TRAP has most commonly been applied in mouse models to isolate mRNA from a subpopulation of cells, expressing a specific

maker from larger, complex brain tissue. High throughout sequencing of this RNA results in the generation of a translatome profile, which allows the analysis of the target cell type in disease conditions or in response to drug treatments. This technique is particularly powerful as a tool for investigating disease models often affecting a particular cell type, such as motor neurons in ALS (MacNair et al., 2015; Sun et al., 2015) or spinal muscular atrophy (Mattedi et al., 2017). Multiple cell types may also be examined in parallel on the same model using different back crosses with BAC:TRAP mice with different marker promoters (Heiman et al., 2008) to examine the profiles of astrocytes or oligodendrocytes compared to the gene expression of affected motor neurons in disease (Sun et al., 2015). In addition to the CNS, TRAP has also been applied to models in order to explore the altered translatome of cell types during embryonic or larval development (Tryon et al., 2013; Hupe et al., 2014).

At present TRAP has mostly been applied to transgenic animal models, including mice (Doyle et al., 2008; Heiman et al., 2008; Drane et al., 2014; Hupe et al., 2014; MacNair et al., 2015; Nectow et al., 2015), zebrafish (Tryon et al., 2013), *Drosophila melanogaster* (Bertin et al., 2015; Huang et al., 2015) and *C. elegans* (Gracida and Calarco, 2017). There are also some recent examples of TRAP being applied to plant species (Collum et al., 2019; Tian et al., 2019). The application of TRAP in animals has mostly been to isolate specific neuronal populations from the brain or subpopulations during embryonic development. A great benefit of TRAP is that the cells are effectively pre-sorted, expressing the tagged ribosomal protein in the desired target cell type. This causes relatively little stress to the cells as the targeted expression of the ribosomal protein can enter endogenous cellular pathways (Sanz et al., 2009). There is also evidence that ribosomal proteins regulate their own

expression and have significant redundancy with other family members (O'Leary et al., 2013). This would allow unhindered expression of the exogenous tagged protein without disturbing normal cellular function. The tissue must be harvested, transferred to a stabilising buffer and homogenized, before proceeding to ribosomal pulldown. Cell populations not expressing the ribosomal protein will be removed during a wash stage. An advantage is also the ability to target relatively low populations of neurons but achieve a better extraction of high integrity RNA than with traditional methods. TRAP has been applied in a number of ways to animal models, including BAC mice (readily expressing tagged ribosomal proteins under a target promoter) (Doyle et al., 2008; Heiman et al., 2008), tetracycline-inducible systems (Drane et al., 2014), adeno-associated virus (AAV) injections (Nectow et al., 2015) and *in utero* electroporation (Huang et al., 2018). One study also performed TRAP from primary mouse cortical neurons, following the transduction of an adenovirus with a packaged *RiboTag* (Lesiak et al., 2015). However, to date no studies have been published showing TRAP extraction of RNA from a human iPSC-motor neuron model, giving a translatome profile for human motor neurons in *C9orf72* ALS.

TRAP has been very successful when performed in mouse models. The existence of several BAC drivers with eGFP fused with ribosomal protein, RPL10a, has meant that a wide variety of neuronal and glial RNA populations have been isolated and sequenced (Doyle et al., 2008). Two studies have examined motor neuron and glial subpopulations using TRAP in ALS mouse models, confirming that the extracted RNA pools are enriched for specific markers, and have unique translatome profiles (MacNair et al., 2015; Sun et al., 2015). They identified novel gene expression

changes previously unknown to occur in these mutations. Sun et al., 2015 performed TRAP from motor neurons (ChAT-bacTRAP), astrocytes (Aldh1l1-bacTRAP) and oligodendrocytes (Cnp1-bacTRAP) from the *loxSOD1*^{G37R} mouse spinal cord. The RNA sequencing analysis, of significantly differentially expressed genes, of the TRAP RNA was compared with between the different samples, showing significant enrichment of the known respective markers from glia and motor neurons. Additionally, heat shock proteins, HSF1 and HSF2, associated with the UPR and *CHOP* (endoplasmic reticulum stress transcription factor) were shown to be uniquely upregulated in motor neurons. These results were confirmed by qRT-PCR, in addition to the misregulation of other proteins known to be involved in the UPR pathway. MacNair et al., 2015 performed microarray analysis on RNA collected via TRAP from TDP-43^{A315T} (Swarup et al., 2011) mice at the age of 5 months (pre-symptomatic) and 10 months (symptomatic), along with age-matched non-transgenic control littermates. This mouse relative early cognitive and motor defects, and is one of the more severe mouse models. The authors found no significant differences in differential gene expression analysis between TDP-43^{A315T} and control mice at the pre-symptomatic time point, though by 10 months 28 genes were differentially misregulated. This is relatively few, especially considering the false discovery rate (FDR), q-value was reduced to 0.1, rather than the conventional 0.05. These genes had cellular processes linked to RNA metabolic processes, immune response and cell cycle regulation (MacNair et al., 2015). Immunofluorescence was used to identify these changes in harvested tissue. In particular, two genes with more than a two-fold increase in expression were found to potentially be crucial to TDP-43 induced ALS. These genes, *DDX58* and *MTHFSD*, are RNA binding proteins. *MTHFSD* appears to be particularly significant

as it was found to co-localise with TDP-43 and G3BP in cytoplasmic *stress* granules. This is another mislocalisation phenotype which has resulted in the loss of an RNA binding protein from the nucleus. In contrast to this study, Sun et al., 2015 found 260 gene expression changes in motor neurons in a severe *SOD1*^{G37R} mouse, though relatively few in oligodendrocytes (23) and astrocytes (108). This striking contrast in the number of differentially expression genes detected may be accounted for due to the amount of input (total whole tissue) RNA submitted for sequencing, representing the transcriptome. Sun et al., 2015 used up to 500 ng, while MacNair et al., 2015 used 40 ng.

1.9 Aims

The differentiation protocol used in this group can achieve up to 80% purity of mature motor neurons (Dafinca et al., 2016), though this varies from culture to culture. To gain deeper insight into the effect of the *C9orf72* HRE has on motor neuron biology, this DPhil project aims to perform RNA sequencing on a line of *C9orf72* ALS patient-derived iPSC-motor neurons. The presence of non-neuronal cells which do not differentiate is often seen, which continue to proliferate after mature motor neurons develop. The ratio of motor neurons to non-neuronal cells in culture varies between lines and passage number. This factor often restricts the amount of time that the motor neurons can be maintained in culture, and the possibility of ageing them further for later harvests is challenging. To consistently obtain mRNA from patient-derived iPSC-motor neurons, a TRAP method has been developed to apply to this iPSC-motor neuron model (Figure 1.1).

1. The first aim of this project was to design and use molecular cloning to generate a human lentiviral TRAP construct to use in mature, electrophysiologically active motor neurons in culture. This relies on a transgenic, ubiquitous, large ribosomal protein, RPL22, to pull down actively translating ribosomes via a c-myc-FLAG-tag, from a select population of cells to isolate the mRNA bound to them (Figure 1.1). The transgene is delivered via lentiviral particles to the motor neurons, and is under the control of a motor neuron-specific promoter.

2. The second aim was to optimise lentiviral transduction of the construct into our cultures, and perform RNA extraction via ribosomal pulldown. This research group works with iPSC lines derived from a number of patients with ALS mutations, with particular focus on *C9orf72* HRE carriers. Work from a former DPhil student project resulted in the generation of a CRISPR/Cas9-edited line where the majority of the HRE was removed, leaving an isogenic control line for one patient line. We also have access to age-matched control lines. Following optimisation of TRAP from these lines, RNA samples were planned to be collected from this patient line, its edited isogenic line and two age-matched controls, over three differentiations. An input fraction of each sample was also isolated, representing the total RNA from a culture, i.e. the transcriptome, when sequenced. These samples were then submitted for next generation RNA sequencing, following quality control steps.

3. The third aim was to perform bioinformatic analysis of the raw reads from RNA sequencing to obtain a counts table, and perform differential expression analysis between TRAP and transcriptome input samples. The success of the TRAP

experiment was be assessed in this analysis. Furthermore, differential expression analysis between *C9orf72* ALS patient and control samples was done to look for a pathological effect of the mutation. The overarching aim is to assess whether motor neurons from patients can be studied in isolation using this method, and if TRAP improves the transcriptional signal of *C9orf72* ALS iPSC-motor neurons.

Figure 1.1. Translating ribosome affinity purification.

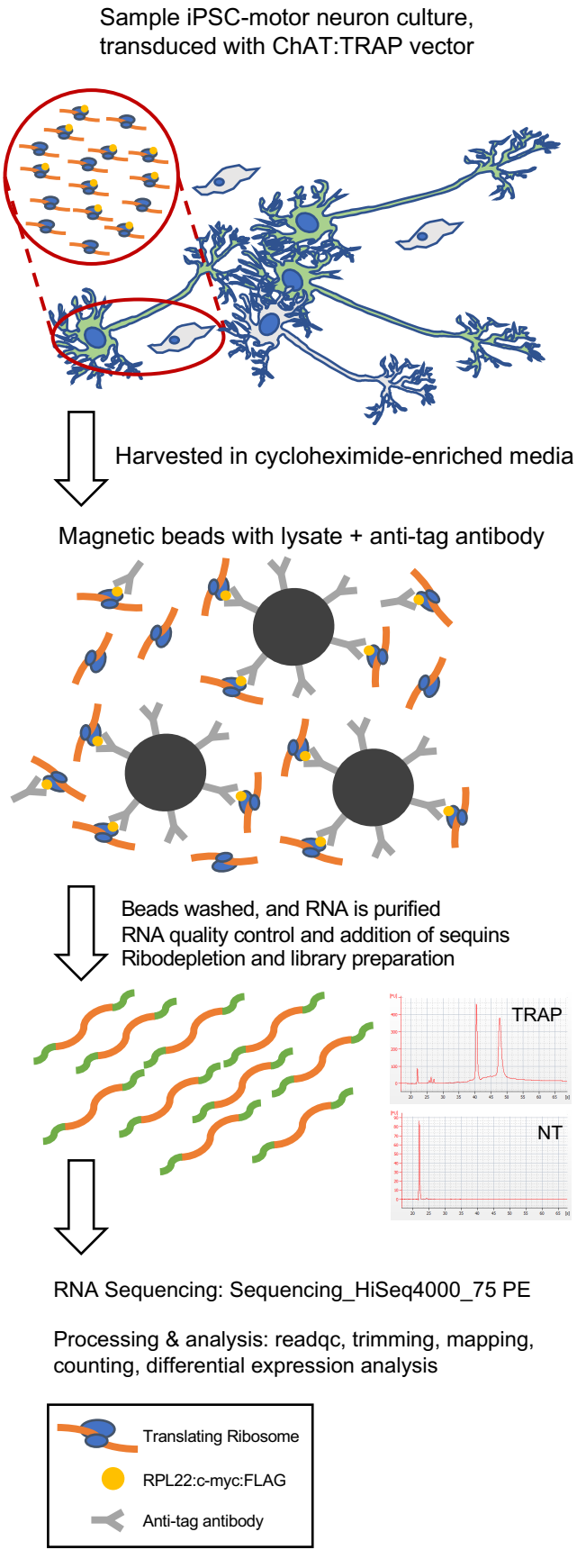


Figure 1.1. Schematic diagram of the TRAP method. A mixed culture of motor neurons and non-neuronal cells is transduced with a lentiviral construct, expressing a tagged-ribosomal protein, under a motor neuron-specific promoter. Using immunoprecipitation, the tagged ribosome complexes are harvested and purified, giving a motor neuron-enriched pool of RNA. This can be checked for quality, quantified and prepared for RNA sequencing analysis. Data from this experiment is then processed and analysed using differential expression analysis.

2 Materials and Methods

2.1 Molecular cloning of translating ribosome affinity purification constructs and validation of expression

2.1.1 Plasmids and primers

Table 2.1 gives the names of the original commercial expression plasmids and the contained transgene components using in the molecular cloning of the TRAP vectors.

Table 2.1. Commercially bought plasmids used in the cloning of TRAP constructs.

Plasmid Backbone	Transgene	Company	Schematic
pCMV6-Entry	RPL22:c-myc:FLAG	Origene	Figure 3.1b.
pEZX-LvPF02	MChAT:GFP	Genecopoeia	Figure 3.1c.
pEZX-LvPF02	RChAT:GFP	Genecopoeia	Figure 3.1c.
pUltra	hUbC:eGFP:P2A:T2A	Addgene	Figure 3.1d.

Table 2.2 gives a complete list of primers, and their sequences, used in the for the molecular cloning of the TRAP vectors. These include the primer pair used to amplify the RPL22:c-myc:FLAG insert from the pCMV6-Entry plasmid, including XbaI and BamHI restriction digest sites to create sticky ends for ligation, primers for amplifying ChAT promoters from the pEZX-LvPF02 plasmid with PacI and AgeI restriction sites, junction primers for colony screening and primers targeting ChAT

and RPL22:c-myc:FLAG for qRT-PCR. Red highlights are restriction site sequences introduced into inserts for ligation into the pUltra lenti-backbone. The blue highlight in RPL22:c-myc: FLAG_XbaI_BamHI FWD is a KOSAK sequence, introduced to improve recruitment of ribosomes to RPL22:c-myc:FLAG transcripts.

Table 2.2. List of all primers used in the cloning of the TRAP constructs and for screening colonies transformed with successful ligations.

Primer	Direction	Sequence (5' - 3')
RPL22:c-myc:FLAG XbaI_BamHI	FWD	acgattctagagccaccatggctcctgtgaaaaagc
	REV	atagtaggatcccttatcgtcgtcatccttgt
5' Junction RPL22 Insert-pUltra	FWD	tcaaacaagcaggtgacgtgg
	REV	tcggatgtcacggtgatcttg
3' Junction RPL22 Insert-pUltra	FWD	ctggattacaaggatgacgacga
	REV	gcattaaagcagcgtatccacat
MChAT_PacI_AgeI	FWD	cctccgttaattaataacgactcactataggg
	REV	ataataaccggtcccttccccagctccccgcc
RChAT_PacI_AgeI	FWD	cctccgttaattaataacgactcactatagg
	REV	aatactaccggtgtgactgtggctggggtcct
5' Junction MChAT-TRAP Insert-pUltra	FWD	cgccaattctgcagacaaat
	REV	cacactcgcatctctccctt
3' Junction MChAT-TRAP Insert-pUltra	FWD	gcagtgaatattgatcccag
	REV	cggtggtgcagatgaactt
5' Junction RChAT-TRAP Insert-pUltra	FWD	cgccaattctgcagacaaat
	REV	gaccttgggcaagtcacagc
3' Junction RChAT-TRAP Insert-pUltra	FWD	ccgtaatctcaacaacacaat
	REV	cggtggtgcagatgaactt
ChAT_qPCR	FWD	cgtggacaacatcagatcg
	REV	atggccatgactgtgtatgc
RPL22:c-myc:FLAG_qPCR	FWD	tcttgattgcacccaccct
	REV	atcgtcgtcatccttgtaatcc

Table 2.3 gives the total list of primers used for Sanger sequencing of complete TRAP products to confirm the final insert orientation, promoter regions and plasmid backbone.

Table 2.3. List of all primers used in Sanger sequencing experiments of the TRAP constructs following molecular cloning of the regions of interest.

Sequencing Experiment	Sequencing Primer	Direction	Sequence (5' - 3')
RPL22:c-myc:FLAG insert in pUltra backbone	SEQ_pUltra-RPL22_1	FWD	ggggttgctctggaaaactc
	SEQ_pUltra-RPL22_1	REV	actccaactagcattccaagg
	SEQ_pUltra-RPL22_2	FWD	agataaatgggcaagtttgtgga
	SEQ_pUltra-RPL22_2	REV	tctttcccctgcactgtacc
	SEQ_pUltra-RPL22_3	FWD	cgtgtcggctccagatct
	SEQ_pUltra-RPL22_3	REV	atccacaaacaagaaccgcg
	SEQ_pUltra-RPL22_4	FWD	caagggctgtagtctgggtc
	SEQ_pUltra-RPL22_4	REV	gccccaccctgcattataag
	SEQ_pUltra-RPL22_5	FWD	tctggtgaggggagggataa
	SEQ_pUltra-RPL22_5	REV	gaacttcagggtcagcttgc
	SEQ_pUltra-RPL22_6	FWD	cacatgaagcagcacgactt
	SEQ_pUltra-RPL22_6	REV	gcttctcgttggggtctttg
	SEQ_pUltra-RPL22_7	FWD	tcaaacaagcaggtgacgtg
	SEQ_pUltra-RPL22_7	REV	caactacgcgcaaccagt
	SEQ_pUltra-RPL22_8	FWD	tgacgacgataagggatccg
	SEQ_pUltra-RPL22_8	REV	gcaatgccccaaccagtg
	SEQ_pUltra-RPL22_9	FWD	acaattccgtggtgttgtcg
	SEQ_pUltra-RPL22_9	REV	cagccatggaaaggacgtc
MChAT Promoter region sequencing primers	M-CHAT_SEQ1	FWD	ttctgcagacaaatggcagt
	M-CHAT_SEQ2	FWD	cacaccagcctcagagct
	M-CHAT_SEQ3	FWD	gtcgcctgtgtgttagctg
	M-CHAT_SEQ4	FWD	gtgactgggaaatgctgagc
	M-CHAT_SEQ5	FWD	gaggtgaagttcgagggc
	M-CHAT_SEQ1	REV	ggaggtgggtctgaaacgat
	M-CHAT_SEQ2	REV	ggtctatcctgggctgagg
	M-CHAT_SEQ3	REV	gctgggactaaggatgggaa
	M-CHAT_SEQ4	REV	catccaggccagtctcaaac
	M-CHAT_SEQ5	REV	aagtcgtgctgcttcatgtg
RChAT Promoter region sequencing primers	R-CHAT_SEQ1	FWD	gattagtgaacggatcggca
	R-CHAT_SEQ2	FWD	tctcacctcaacctgaacca
	R-CHAT_SEQ3	FWD	aaggaaggtgagtctccctg
	R-CHAT_SEQ4	FWD	aagggcatcgacttcaagga
	R-CHAT_SEQ1	REV	ggcctttcgtcagtacatgc
	R-CHAT_SEQ2	REV	ctagctggctgatccctctg
	R-CHAT_SEQ3	REV	cgttgaccctggatggca
	R-CHAT_SEQ4	REV	gcatggcggacttgaagaa
	R-CHAT_SEQ5	REV	cacgtcacctgcttgtttga
TRAP Lentiviral plasmid backbone sequencing primers (including hUbC promoter region)	TRAP_plasmid_SEQ1	FWD	acaataaacgggtctctctggt
	TRAP_plasmid_SEQ2	FWD	ggacagctacaaccatccct
	TRAP_plasmid_SEQ3	FWD	cctggctgtggaaagatacc
	TRAP_plasmid_SEQ4	FWD	cttccgcatcttcgccttc
	TRAP_plasmid_SEQ5	FWD	cctgtagcggcgcattaag
	TRAP_plasmid_SEQ6	FWD	ccgctcatgagacaataaccc
	TRAP_plasmid_SEQ7	FWD	aagggagaaaggcggacag
	TRAP_plasmid_SEQ8	FWD	tcactcattaggcaccccag
	TRAP_plasmid_SEQ1	REV	tccaatccatgtcagacccg
	TRAP_plasmid_SEQ2	REV	ccctggccttaaccgaattt
	TRAP_plasmid_SEQ3	REV	aattgtctggcctgtaccgt
	TRAP_plasmid_SEQ4	REV	tggtgaatatccctgcctaact
	TRAP_plasmid_SEQ5	REV	ggaattgtcagtgcccaaca
	TRAP_plasmid_SEQ6	REV	tccacactgactaaaagggtc
	TRAP_plasmid_SEQ7	REV	tgcaaggcgattaagttggg
	TRAP_plasmid_SEQ8	REV	tgcttttctgtgactggtgag
	TRAP_plasmid_SEQ9	REV	tccatagttgcctgactccc
	TRAP_plasmid_SEQ10	REV	aacccggtaagacacgactt
	TRAP_plasmid_SEQ11	REV	ctcttccgcttcctcgct
	TRAP_plasmid_SEQ12	REV	tccagcttttgttccctttagtg

2.1.2 Patient-derived fibroblast lines and induced pluripotent stem cell induction

Skin biopsies were taken from control and *C9orf72* ALS or ALS/ FTD patient donors in the Oxford Motor Neuron Disease Clinic run by Prof Kevin Talbot, University of Oxford. Collection and expansion of the fibroblasts were performed by qualified members of the clinic. Induction to pluripotency, via the Yamanaka method (Takahashi and Yamanaka, 2006) was performed at the Dunn School of Pathology, University of Oxford, by the group of Dr Sally Cowley at the Stem Bank facility (as described in Dafinca et al., 2016). CRISPR/Cas9 edited isogenic lines were generated via the double nickase strategy and positive colony picking by Dr Nidaa Ababneh, Talbot group, Nuffield Department of Clinical Neuroscience, University of Oxford. The lines used in this project were two healthy age-matched controls (SFC840 and SFC180), two clones from one patient with a C9orf72 HRE $[GGGGCC]_{1000}$ (OXC90202 and OXC90203) and their edited isogenic lines (59_1, 59_2 and 59_3), where the *C9orf72* HRE was removed leaving $[GGGGCC]_2$ in intron1a of the gene.

2.1.3 Molecular cloning

All expression vectors obtained from commercial companies were expanded by bacterial heat shock transfection into Stbl3® OneShot competent *Escherichia coli* (*E. coli*) (Life Technologies). All plasmids were purified from 4 mL LB broth + 34 µg/ mL chloramphenicol or 100 µg/ mL ampicillin, from overnight cultures (PureLink® Quick Plasmid DNA Miniprep Kits, Life Technologies). Plasmid sizes were confirmed by using restriction enzyme (RE) digests and checking expected band sizes by agarose gel (1%) electrophoresis. All restriction enzymes were obtained

from New England Biolabs (NEB). cDNA (<1 µg) was incubated in the appropriate enzyme(s) and buffer, at 37 °C, between 15 minutes and 1 hour. All reactions were inactivated at 65 °C, 20 minutes. Larger overnight cultures (250-300 mL LB broth) were used to purify higher concentrations of plasmid to use for cloning (Endotoxin-free plasmid DNA purification Xtra Midi EF, Nucleobond®).

Polymerase chain reaction (PCR) (Q5 Hot Start High-Fidelity DNA Polymerase, NEB) was used to generate copies of the gene or promoter of interest from their original constructs using primers with overhanging regions, containing new restriction sites and at least 4 random bases (Table 2.2), with 50-100 ng of template DNA. Master mixes were pipetted according to manufacturer's instructions (primers 10 µM) and PCR was performed over 35 cycles. Annealing temperature varied by primer pair. Cloned fragments were isolated by running a 2% agarose gel at 75 V, 3 hours and performing gel extraction (Qiaquick Gel Extraction, QIAGEN) to purify the amplified DNA from the predicted band size. All cDNA species for cloning had unique restriction sites at the 5' and 3' ends with sticky ends to ensure correct orientation of all inserts into backbones and efficient ligation. Phosphatase reactions (Antarctic Phosphatase, NEB) were setup for cDNA templates extracted from bacteria to prevent re-ligation following a RE digest. Ligation reactions were set up at various ratios of insert: template (1:1, 3:1, 6:1 and 10:1) and incubated with T4 DNA Ligase (Promega) for either 1 hour, room temperature or overnight at 16 °C. Inactivation of ligation reactions was done at 65 °C, 10 minutes. Ligation products were transformed into Stbl3® OneShot competent *E. coli* (2-5 µL), spread onto Agar + Ampicillin (100 µg/ mL) plates and incubated overnight at 37 °C.

Colony PCR was used to screen for colonies transformed with successful ligations. Colonies were picked and dissolved in 10 µL of distilled H_2O in 96 well plates. cDNA was exposed by heating plate to 95 °C on a thermal cycler. This was used as the DNA template component of a PCR master mix (MegaMix-Blue, Microzone). Primers were designed to anneal at the junctions (Table 2.2) of the ligation site, and size of product depended on whether the ligation was successful or not. PCR was run according to manufacturer's instructions.

2.1.4 HEK293-T cell culture

HEK293-T cells with a modified large T antigen was obtained from ATCC® CRL-3216™. HEK293-T cells were cultured on T25, T75 or T175 flasks (Corning®) at 37 °C. Cells were cultured in DMEM high glucose, sodium pyruvate (ThermoFisher Scientific; 11995065) enriched with fetal bovine serum (ThermoFisher Scientific; 10270106) and Antibiotic-Antimycotic (10,000 units/ mL penicillin, 10,000 µg/ mL streptomycin, and 25 µg/ mL Amphotericin B) (ThermoFisher Scientific, 15240062). Cultures were passaged at approximately 80-90% confluency. Media was aspirated, and cells were briefly washed with phosphate buffered saline (PBS). 0.05% trypsin-EDTA in PBS (0.5% Trypsin-EDTA 10X; ThermoFisher Scientific, 15400054) was added to the cells (T25-500 µL, T75-2 mL, T175-4 mL) and incubated at 37 °C for 3-4 minutes. Media was added to the cells following trypsinization in a 2:1 ratio to inhibit further enzymatic reaction. The cell suspension was collected in a 50 mL falcon tube and centrifuged at 500 g, 3 minutes, room temperature. Supernatant was discarded, and cell pellet was resuspended in media and passaged as desired. General seeding density was 40%. Cells were counted using a manual haemocytometer for controlled experiments.

HEK293-T cells were used to generate 3rd generation lentiviral particles by co-transfection with VSV-G, Gag-Pol and Rev plasmids, and the lentiviral cDNA constructs; MChAT:TRAP, RChAT:TRAP or hUbC:TRAP.

2.1.5 Sanger sequencing

The Sanger sequencing facility at the Weatherall Institute of Molecular Medicine, Oxford, was used to sequence constructs. All sequencing primers were designed using the *Primer3* tool and assessed using standard software to view *.ad1* files, by aligning expected sequences with results. See primers in Table 2.3.

2.1.6 Immunocytochemistry

Fixation of cells seeded on coverslips was done by washing gently with warmed PBS and incubating with 4% PFA for 15 minutes, room temperature. Cells were blocked in 10% normal donkey serum in 0.2% Triton/ PBS for 1 hour, room temperature. Primary (Table 2.4) and secondary antibody (1:500) incubation was done in 1% normal donkey serum/ 0.1% Triton/ PBS, overnight at 4 °C for the former and 1 hour, room temperature, for the latter. All secondary antibodies were obtained from ThermoFisher Scientific. DAPI incubation (1:2000) in 0.1% Triton/ PBS was done at room temperature, 3-5 minutes. Coverslips were mounted onto slides using DAKO mounting medium (Agilent), and allowed to dry before viewing. A list of all antibodies used in this project have been given along with their working dilutions in Table 2.4.

2.1.7 SDS-PAGE and Immunoblotting

iPSC-motor neurons were harvested for protein by washing cells cultured in 6-well plates in PBS once and scraping. The harvested cells were centrifuged (2000g, 10 minutes, 4 °C), and the pellets were lysed with RIPA buffer (Sigma-Aldrich) enriched with a cOmplete protease inhibitor cocktail (Roche). The cell suspension was sonicated using a motor pestle with 10, 1 second bursts and incubated on ice for 30 minutes. A bicinchoninic acid (BCA) assay (Sigma-Aldrich) was used to quantify the protein samples. 20 µg per sample was run on a NuPAGE Bis-Tris 4%-12% gel (Invitrogen), at 80 V for 10-15 minutes followed by 100-120 V for 1.5 – 2 hours. Protein was transferred to a PVDF membrane using an iBlot® 7-minute transfer system (Sigma-Aldrich). Membranes were blocked overnight, 4ºC in 4% milk/ TBS-T. Primary antibody incubation (Table 2.4) was done in 1% milk/ TBS-T, overnight at 4ºC. Horseradish peroxidase-conjugated secondary antibody incubation (1:10,000) was done in 1% milk/ TBS-T at room temperature, 1 hour. Membranes were incubated briefly with ECL reagent (Amersham Pharmacia Biotech) to visualise the protein bands via exposure to photographic film. Protein band quantification was done using the ImageJ software.

Table 2.4. List of all antibodies used for either immunocytochemistry or western blotting, with the optimised dilutions used in all experiments, unless otherwise specified. Bolded c-myc antibody was used for TRAP of RNA in final protocol.

			Dilution	
Target	**Species**	**Company**	**Immunocytochemistry**	**Western Blot**
GFP	Rabbit	Life Technologies	1:500	-
RPL22	Goat	Novus Biologicals	-	1:500
RPL22	Rabbit	ProteinTech	-	1:500
ChAT	Goat	Millipore	1:100	-
SMI-32	Mouse	Abcam	1:1000	-
c-myc	Mouse	Sigma-Aldrich	1:1000	-
DDK	Rabbit	Origene	-	1:500
DDK	Mouse	Origene	1:1000	1:500
Actin	Mouse	Abcam	-	1:10,000

2.2 Optimisation of translating ribosome affinity purification from HEK293-T cells and iPSC-derived motor neurons

2.2.1 Lentiviral particle generation, titration and transduction

HEK293-T cells were transfected with plasmids encoding VSVG, Gag-Pol, REV and the packaging plasmid of choice, using PEImax reagent (Polyplus-Transfection), to produce 3rd generation lentiviruses. HEK293-T cells were grown in DMEM + high glucose + pyruvate media (ThermoFisher Scientific) with 10% foetal bovine serum (ThermoFisher Scientific) and 1% Penicillin-Streptomycin-Amphotericin B (ThermoFisher Scientific). 12 flasks of cells at 80-90% confluency were used to generate one lentiviral species. The media was replaced 24 hours following transfection. After a further 48 hours, the media of replicate transfections were pooled and filtered through a nitrocellulose-free 0.45 um filter (Millipore). LentiXTM concentrator reagent (Clontech) was mixed with pooled lentivirus media in a 1:3 ratio and incubated overnight at 4 °C. The lentiviral pellet was isolated by centrifugation at 1500 g, 45 minutes, 4 °C. Pellet was resuspended in 1:100 PBS of total media pooled and stored at -80 °C. 5 µL of concentrated media was diluted in 495 µL of media for titration.

Lentiviral RNA was purified and titrated using the Lenti-X™ qRT-PCR Titration Kit (Takara). RNA was extracted and a DNase I reaction was performed to remove contaminants. Quantitative-RT-PCR, with primers targeting the LTR regions of the viral RNA were used to amplify the RNA products and detected using ROX reference dye LSR. Serial dilutions of a known standard RNA mixture were used to construct a standard curve and used to estimate the concentrations of own lentiviral

RNA. These concentrations were used to estimate multiplicity of infection (MOI) due to variability of the iPSC-motor neuron differentiation.

Cells to be transduced (HEK293-T cells or iPSC-motor neurons) were seeded in 6-well plates, 24-well plates on coverslips or 10 cm petri dishes, following counting by haemocytometer or via an automated instrument. HEK293-T cells were seeded at 30-40% confluency and neurons according to the differentiation protocol. Lentiviral transduction was performed by enriching nutrient media (enriched with ROCKi) with calculated volume of lentivirus and immediately adding cells by gentle pipetting using a stripette and returned to 37 °C. Incubation was done from 4 hours to overnight. The cells were washed once with PBS and media was replaced. ROCKi was kept in media for at least 24 hours following transduction.

2.2.2 iPSC-motor neuron differentiation

Differentiation of motor neurons from control and *C9orf72* ALS/ FTD patient-derived induced pluripotent stem cells (iPSC) was followed as in Dafinca et al., 2016. See Figure 2.1 for a summary of this protocol. Induced pluripotent stem cells, generated by Prof Sally Cowley, Dunn School of Pathology, University of Oxford, were expanded as monolayer cultures in mTeSR growth media (STEMCELL™ Technologies), supplemented with 1% Antibiotic-Antimycotic, and Rho-associated, coiled-coil containing protein kinase inhibitor (ROCKi) (Y-27632 dihydrochloride, Tocris Biosciences) on thawing from liquid nitrogen storage.

All iPSC vials, stored at liquid nitrogen phase, were transferred to dry ice prior to rapid thawing. Vials were thawed in 37 °C water bath for maximum 1 minute before rapidly thawing via addition and gentle pipetting of warmed PBS. Thawed cells were

diluted in 7-10 mL of PBS and centrifuged at 200 g, 5 minutes, room temperature. Supernatant was discarded, and cells were resuspended in mTeSR™ growth media, enriched with ROCKi and Antibiotic-Antimycotic. iPSCs from one vial were seeded onto 6-well plate wells coated with Geltrex™ (ThermoFisher Scientific) in Knock-out™ DMEM (ThermoFisher Scientific). ROCKi was removed from media 24 hours post-thawing. iPSC cultures were passaged at no more than 80% confluency by 0.5 mM EDTA in PBS. Cultures were washed and incubated with EDTA for 5 minutes, 37 °C. EDTA was carefully aspirated and cells were detached by pipetting warmed mTeSR + Antibiotic-Antimycotic onto cells and replating onto Geltrex-coated wells.

Differentiation protocol is initiated when iPSCs are 60-80% confluent, depending on the growth rate of cultures, such that small gaps will fill by proliferating cells early in the protocol. This monolayer motor neuron differentiation protocol lasts up to 30 days. The protocol involves a series of enriched growth media changes and passages (see summary in Figure 2.1). **Day 0-1:** DMEM-F12, HEPES (ThermoFisher Scientific) and Neurobasal (ThermoFisher Scientific) (mixed 1:1) enriched with 2% B27 (ThermoFisher Scientific), 1% N2 (ThermoFisher Scientific), 1% antibiotic-antimycotic (ThermoFisher Scientific), 1x 2-mercaptoethanol (ThermoFisher Scientific), 0.5 µM ascorbic acid (Sigma-Aldrich), 3 µM Chir99021 (Tocris Biosciences), 1 µM Compound C (Merck-Millipore). **Day 2-3:** DMEM-F12 HEPES and Neurobasal enriched with B27, N2, antibiotic-antimycotic, 1 µM retinoic acid, 2-mercaptoethanol, ascorbic acid, Chir99021, Compound C. **Day 4-8:** DMEM-F12 HEPES and Neurobasal enriched with B27, N2, antibiotic-antimycotic, retinoic acid, 2-mercaptoethanol (ThermoFisher Scientific), 500 nM SAG (Tocris Biosciences), ascorbic acid. **Day 9-17:** DMEM-F12, HEPES and Neurobasal

enriched with B27, N2, antibiotic-antimycotic, retinoic acid, 2-mercaptoethanol, 0.1% ROCKi, SAG, ascorbic acid. **Day 18-30:** DMEM-F12, HEPES and Neurobasal enriched with B27, N2, antibiotic-antimycotic, retinoic acid, 2-mercaptoethanol, ROCKi, SAG, ascorbic acid, 0.5 µg/ mL laminin (ThermoFisher Scientific), 10 ng/ mL BDNF (ThermoFisher Scientific), 10 ng/ mL GDNF (ThermoFisher Scientific), 10 µM DAPT (Tocris Biosciences). The iPSCs are induced to the initial differentiation media on day 0 and cannot be passaged until day 9. On day 9, each well may be split up to 1:3 into Geltrex-coated wells and allowed to grow to full confluency. At this stage the cells are neural progenitors and may be indefinitely passaged and maintained in day 9 media for expansion. For final motor neuron induction, the cells are split at least 1:6 onto Geltrex-coated wells or coverslips in day 18 media. The cells are counted at this stage if needed for lentiviral transduction. Mature motor neurons are harvested from day 25-30.

Figure 2.1. Differentiation of iPSCs to motor neurons.

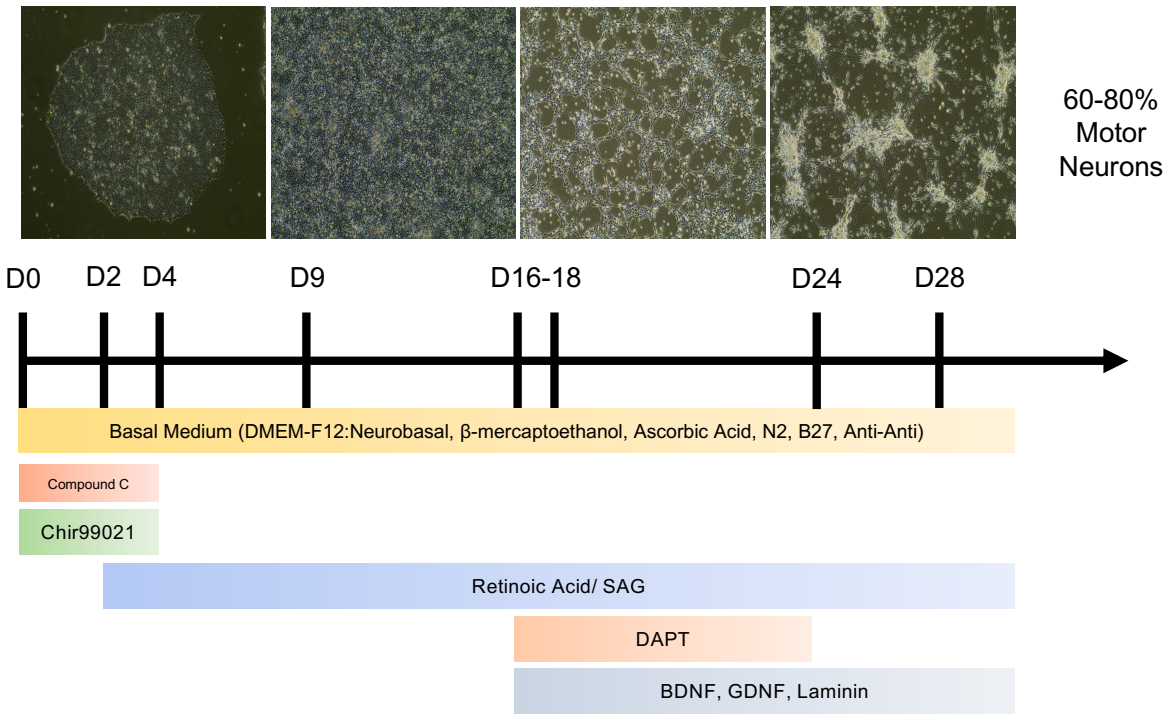


Figure 2.1. Schematic summarising the differentiation protocol to obtain motor neurons expressing mature markers, from human induced pluripotent stem cells, as published in Dafinca et al., 2016.

2.2.3 Reverse transcription, quantitative PCR (qRT-PCR)

List of primers used for qRT-PCR can be seen in Table 2.2. Reverse transcription, or cDNA synthesis was carried using SuperScript™ III First-Strand Synthesis SuperMix (Invitrogen) with 10-500 ng of total RNA. Messenger RNA were selected for using poly(A) tail enrichment (Oligo(dT) priming). Fast SYBR™ Green Master Mix (Applied Biosystems™) and FastStart SYBR Green Master (Roche) reagents were used in reactions according to manufacturer's instructions. 200-300 nM of primer pairs used in master mix. 1-2 µL of cDNA synthesis reaction was used in qRT-PCR reactions, depending on amount of starting RNA used. Generally, the annealing temperature of primers was optimised to 58-60 °C, 30 seconds and 40-45 cycles were run.

2.2.4 Microscopy

Confocal microscopy was used to image fixed and fluorescent antibody-stained cells. Cells were imaged at x40 or x63 magnification. Exposures were adjusted to negative controls to minimise autofluorescence signals. Experimental slides and coverslips were blinded for imaging and analysis.

2.2.5 Translating ribosome affinity purification

This protocol was optimised for human iPSC-derived motor neuron differentiation cultures, using previously published TRAP protocols as a starting point (see Chapter 4). Control and *C9orf72* ALS patient iPSC-motor neurons were transduced with lentiviral particles (concentration = 500-1000 particles/ cell), packaged with TRAP vector and incubated at 37 °C for 4 hours. ROCKi-enriched media was used to maintain culture viability during and after lentiviral transduction for up 48 hours.

Cultures were harvested 96 hours post-lentiviral transduction. Prior to harvesting cultures were incubated with Day 18+ motor neuron media enriched with 100 μg/ mL cycloheximide for 1 hour, 37 °C. The cultures were washed with PBS enriched with 100 μg/ mL cycloheximide and all liquid was completely aspirated. 300 μL of homogenization buffer (50 mM Tris-HCl, 100 mM KCl, 12 mM $MgCl_2$, 1% NP-40, 1 mM DTT, **1x Protease inhibitor cocktail (Roche), 200 U/ mL RNasin (Promega), 100 μg/ mL cycloheximide**, 1 mg/ ml heparin (Sigma); pH adjusted to 7.4) was pipetted per well to cover all cells. Buffers were made beforehand, with pH adjustments and stored at 4 °C. Bolded reagents should be added on the same day as experiment. Cells were immediately scraped and collected in 2 mL tubes (pre-chilled, DNase/ RNase-free quality). 6 wells were pooled per tube (1.8 mL total). Cells were lysed by gently pipetting up and down 10-20 times. Tubes were immediately incubated on ice for the duration of harvesting all cultures. Tubes were centrifuged at 12,000 g, 4 °C, 10 minutes. The supernatant was transferred into a new, pre-chilled 2 mL tube; 180 μL (10%) of the supernatant was taken as a transcriptome input fraction and stored at 4 °C on ice until RNA extraction was completed. Antibody for the TRAP tag, c-myc or FLAG, was added to the remaining 90% lysate fraction, thoroughly mixed by pipetting and incubated at 4 °C, rotating for 2 hours. Meanwhile, 20-40 μL Dynabeads™ Protein G beads were equilibrated by resuspension in the homogenization buffer, twice. Following the 2 hours of incubation with antibody, all the lysate was added to the equilibrated protein beads and mixed until no bead aggregates remained. This is crucial to avoid pulling down non-specific protein and RNA complexes. Beads and lysate were incubated, rotating at 4 °C for 1 hour. This suspension was then spun down at 100 g, 5 seconds to pulldown all liquid from the cap. A magnetic rack was used to attract the beads

from the suspension. Lysate was pipetted out and discarded. Beads were washed 3 times, 5 minutes at 4 °C in wash buffer (50 mM Tris-HCl, 100 mM KCl, 12 mM $MgCl_2$, 1% NP-40, **1x Protease inhibitor cocktail (Roche), 200 U/ mL RNasin, 100 µg/ mL cycloheximide**; pH adjusted to 7.4). Bolded reagents were added on the day of the experiment, while the wash buffer can be made up in advance and stored at 4 °C. After the final wash, all liquid should be aspirated, and beads were resuspended in 150-300 µL TRIzol™ reagent. Transcriptome input fractions were also made up to final resuspension volume of pulldown samples in TRIzol™. All samples were then mixed 1:1 with 100% molecular grade ethanol and standard RNA extraction protocol via Direct-zol™ Microprep RNA kit (Zymo Research). Samples were passed through Microprep columns and spun at 16,000 g, room temperature. To perform a DNase on-column digest, the column was washed with RNA Wash Buffer and, the enzyme and buffer master mix were made up and pipetted directly onto the filter membrane of the columns and incubated at room temperature for 15 minutes. Two further washes were performed using *Direct-zol*™ RNA Pre-wash Buffer. These are done at 16,000 g, room temperature, 30 seconds. A final wash with RNA Wash Buffer was done at 16,000 g, room temperature, 2 minutes. RNA was eluted with 20 µL DNase/ RNase-free water and passed through the column twice to increase yield (16,000 g, room temperature, 1 minute). 2 µL of sample was aliquoted for quality control and concentration measurement and stored at -80 °C along with remaining sample. Samples at all steps are maintained in DNase/ RNase-free tubes and at 4 °C or on ice.

2.2.6 Quality control using RNA Integrity Number (RIN) analysis

The Bioanalyzer method (Agilent 2100 Bioanalyzer) was used to calculate the RNA integrity numbers (RIN) of RNA obtained via the TRAP method and transcriptome input fractions. This was performed according to the manufacturer's instructions using both the Pico and Nano kits (Agilent) due to differing stock concentrations of RNA from TRAP and transcriptome input samples.

2.2.7 Quant-iT RiboGreen Assay

The concentration of RNA samples to be sent for Illumina RNA sequencing obtained via TRAP and through standard RNA purification for transcriptome input fractions, were measured using the RiboGreen assay (Quant-iT™ RiboGreen® RNA Reagent, Invitrogen). Low input concentrations of TRAP RNA meant that spectrophotometry using a Nanodrop could not be used to accurately determine sample concentrations. Furthermore, bioanalyzer chips that were used to determine RIN values would not have been reliable to measure concentration as TRAP and transcriptome input RNA were measured on Pico and Nano chips respectively, due to different ranges of concentration. This assay allows accurate measurement of RNA species in the range 20 pg/ µL to 1 ng/ µL, comparing signals from a dye to a standard curve on one 96 well plate, whereas concentrations from bioanalyzer chips would vary between Pico and Nano chips and depend on gel quality. This could bias the RNA sequencing experiment.

2.3 RNA sequencing and data analysis

2.3.1 Addition of *sequin* spike-ins

Artificial spike-in RNA mixtures (*sequins*) were added to all RNA sequencing samples for downstream normalisation according to the manufacturer's instructions (https://www.sequin.xyz). Mixture A was added to age-matched control-TRAP and the respective transcriptome input samples. Mixture B was added to patient *C9orf72*-TRAP and respective transcriptome input samples.

2.3.2 RNA sequencing library preparation and platform

Repeated quality control, library preparation of samples and running of sequencing experiment was carried out by the Wellcome Sanger Institute, Wellcome Genome Campus, University of Oxford. The ribodepletion method, using a Ribozero kit was employed, with stranded library preparation for all samples using the TruSeq Stranded mRNA kit. Paired-end (2x 75 bp) RNA sequencing was performed on the HiSeq4000 platform over 75 cycles for amplification and detection of barcoded fragments in samples, multiplexed over three lanes.

2.3.3 Bioinformatic analysis, processing and differential gene expression analysis of translatome data

Bioinformatic analysis of the all presented RNA sequencing data was done using Python pipelines (*CGATPipelines*) written and regulated by the Computational Genomics: Analysis and Training (CGAT) team, Weatherall Institute of Molecular Medicine, University of Oxford. Raw paired-end reads in *fastq* file format from TRAP RNA sequencing experiment were received from the Oxford Wellcome Centre for Human Genomics and copied to the remote CGAT server. The source code of these pipelines may be viewed at https://github.com/cgat-developers/cgat-apps.

2.3.3.1 Processing of raw reads using online tools

CGATpipelines readqc was used to validate the quality of reads in FastQC (Babraham Institute), including phred scores across reads, per sample, duplication rates, % GC content and levels adaptor sequences. No trimming methods were applied to the dataset.

2.3.3.2 Mapping and quality control

All *fastq* files that passed quality control of raw reads were passed to *CGATpipelines mapping*. The STAR tool was used to map reads to an indexed human genome (*fasta* format) and annotated human genome, generated from Ensembl build 87 for the human genome in a gene transfer format (GTF) file. This GTF file was modified to include *sequin* information, as obtained from manufacturer's website (https://www.sequin.xyz) (performed by Dr Jakub Scaber). Further manual modifications were made to the GTF file to include TRAP vector transcript read including eGFP:P2A:RPL22:c-myc:FLAG. The binary output of a sequence alignment map (SAM) file, known as BAM file, was passed to *CGATpipelines bamstats* to obtain detailed information regarding mapping quality of raw reads to the human genome. The pipeline allows the options to run *Samtools* (Li et al., 2009), *Picard Tools* (http://broadinstitute.github.io/picard/), a Java tool written by the Broad Institute of MIT and Harvard and BamStats, a script written by CGAT. Collectively these tools allow the extraction of mapping statistics to regions of a genome by RNA species, alignment and coverage statistics to annotated contigs, information on duplication and unmapped reads. The Integrated Genome Browser was used to upload BAM files and visualise individual genomic loci,

including housekeeping genes to confirm uniform coverage of mapped reads in each sample.

2.3.3.3 Quantification of counts

CGATpipelines rnaseqdiffexpression was used to run *featureCounts* (Liao et al., 2014) (*Subreads* package) with BAM file inputs for quantification of reads mapped to protein-coding regions using annotated human genome GTF file. *Salmon* (Patro et al., 2017) tool was run to obtain transcripts per million (TPM) counts, with untrimmed, paired-end *fastq* file inputs. An indexed genome for *Salmon* was generated by the pipeline and modified to include *sequins* artificial chromosome features and TRAP vector transcript.

2.3.3.4 Exploratory and differential expression analysis

Counts tables and *quant* file outputs from gene and transcript quantification tools, respectively were imported into R for differential expression analysis. *Tximport* (Soneson et al., 2019) was used to import TPM counts from the *Salmon* output. *DESeq2* (Love et al., 2014) and dependant packages were used to plot and perform exploratory and differential analysis of translatome data. Translatome data was normalised across dataset after excluding outlier and transcriptome input samples. Group-wise comparisons with appropriate thresholds were applied to obtain log2 fold changes with adjusted p-values (padj), contrasting control-TRAP and *C9orf72* ALS-TRAP samples. Normalisation of the gene set was done by fitting a generalised linear model (GLM) to the data taking into account, batch effects occurring from RNA sequencing lane, *sequin* calibration and condition group.

2.3.3.5 Spike-in Analysis

Raw TPM counts of *sequin* transcript isoforms were imported into the *ANAQUIN* package for grouped sample analysis. Correlation of differentially expressed *sequins* and normalised translatome data was performed using *DESeq2* to calculate log2 fold changes, and plotting against the expected log2 fold change values for *sequins* between groups with spike-ins from mixture A and B.

2.3.3.6 Batch correction using *sequins*

Variation captured by *sequin* counts across samples was corrected using the *RUVseq* package (Risso et al., 2014).

2.3.3.7 Gene ontology analysis

Online tools such as *Gorilla* and *REVIGO* were used to input genes from statistically significant gene lists after differentially expression analysis and post-hoc filtering. Total detected genes from the appropriate gene list were used as a background for analysis. Gene ontology databases of annotated human biological pathways, gene molecular function and cellular components were used to explore the data and search for enriched genes in own data set. Fast gene set enrichment analysis (FGSEA) was then used to look for enriched pathways and gene networks with the added component of log2 fold change and adjusted p-value in the enrichment analysis.

3 Molecular cloning of a *translating ribosome affinity purification* (TRAP) lentiviral expression vector

3.1 Introduction

The TRAP method used in this project is currently the only known example of the system being implemented in a human iPSC-derived culture. As such, a construct from human cDNA elements needed to be cloned to generate a human expression vector within a lentiviral backbone. TRAP was originally applied in a mouse model to harvest RNA from neurons with a specific molecular marker in brain tissue (Doyle et al., 2008; Heiman et al., 2008). It is generally challenging to obtain RNA of acceptable quality for RNA sequencing analysis from specific neuronal cell types in tissue without significant degradation of both tissue and RNA. The mouse was genetically modified by means of integrating a BAC at the mouse *ChAT* locus. Subsequently, harvested tissue already contained target neurons expressing tagged ribosomal proteins, which could be immediately extracted without the need to dissect, or homogenize and sort the neurons. Our research group uses an ALS patient-derived iPSC-motor neuron model, comprising of potentially unlimited lines originating from different patients with their various ALS-linked mutations, including *C9orf72*. These patients will have a host of differential background genes which may influence the onset of disease. In future this lab aims to generate isogenic CRISPR/Cas9 lines for as many of these lines as possible. These will provide a control giving a representation of the sole effect of the mutation. The generation of a TRAP construct, which could be applied broadly across this model without the need to generate a stable clone for each new line was sought. This chapter describes the design of the lentiviral, human TRAP construct used to pulldown

mature RNA from *C9orf72* ALS patient-derived iPSC-motor neurons for next generation RNA sequencing.

3.2 A lentiviral TRAP vector for targeted expression in human ChAT-expressing motor neurons in culture

3.2.1 TRAP vector starting commercial constructs and design conception

The TRAP construct was designed based on the *RiboTag* used in Lesiak et al., 2015 to isolate mRNA from mouse primary neurons. The large ribosomal subunit protein, RPL22 to be used for mRNA pulldown is fused to a c-myc-FLAG tag (Figure 3.1a). RPL22:c-myc:FLAG is bicistronically expressed with enhanced green fluorescent protein (eGFP) (Figure 3.1a). The two open reading frames (ORF) are separated by a P2A, *cis-acting hydrolase element* (CHYSEL) peptide cleavage sequence (GDVEENPGP) (Figure 3.1a). This sequence can be cleaved between the N-terminal glycine and proline amino acids by ubiquitous proteases in the target cells (Figure 3.1a). Consequently, a separate eGFP reporter is simultaneously produced with RPL22:c-myc:FLAG, facilitating the identification of cells which are successfully expressing the lentiviral construct. Previous TRAP techniques have used systems where the reporter is fused with the ribosomal proteins. This serves to report the expression and localisation of the exogenous protein, but also used as the tag to pull down RNA. As a large protein, eGFP could potentially interfere with the cellular localisation of RPL22, which is thought to have a crucial role in early ribosomal synthesis in the nucleolus, as well as the efficiency of its incorporation into polysomal complexes. Thus, the given design of bicistronic gene expression was used, with a smaller c-myc:FLAG tag fused to RPL22. The ChAT promoter was

chosen to drive expression of the TRAP vector, as it is robustly expressed in the motor neurons of our iPSC-derived cultures. The human *ChAT* locus is known to contain several promoter and regulatory regions, interspaced across the gene (Figure S8.1), which collectively would be too large to clone into a lentiviral vector backbone, which has a maximum size of 10 kb, for efficient uptake and integration into a host's genome. These different regulatory regions are known to modulate the expression of various ChAT transcript isoforms, all of which code for the same choline acetyltransferase protein (Naciff et al., 1999). The efficiency of the various promoter regions has not been characterised in human iPSC-motor neurons, however a former study has identified the M and R isoforms as the most active promoters in human spinal cord and motor cortex (Trifonov et al., 2009), hence these were chosen as the promoters of the TRAP constructs. The commercial supplier used bioinformatic analysis to predict the sequences of these promoters, and the regions they map to in the human ChAT locus. This has been given in Figure S8.1. In total, three TRAP constructs were cloned (Figure 3.1a), one control vector containing a ubiquitous mammalian promoter, hUbC, and two with ChAT promoters, M and R, which will be used to pulldown mRNA species only from ChAT-positive motor neurons in differentiated cultures. The ChAT promoter with the best efficiency will be used to pulldown mRNA for the final RNA sequencing experiment. More than one ChAT:TRAP vector was cloned in case one did not work, or was inefficient. The vector containing the hUbC promoter was used for optimisation of the TRAP method in HEK293-T cells.

Figure 3.1. The lentiviral TRAP vector was cloned from commercially bought expression constructs containing the desired cDNA elements for TRAP vectors.

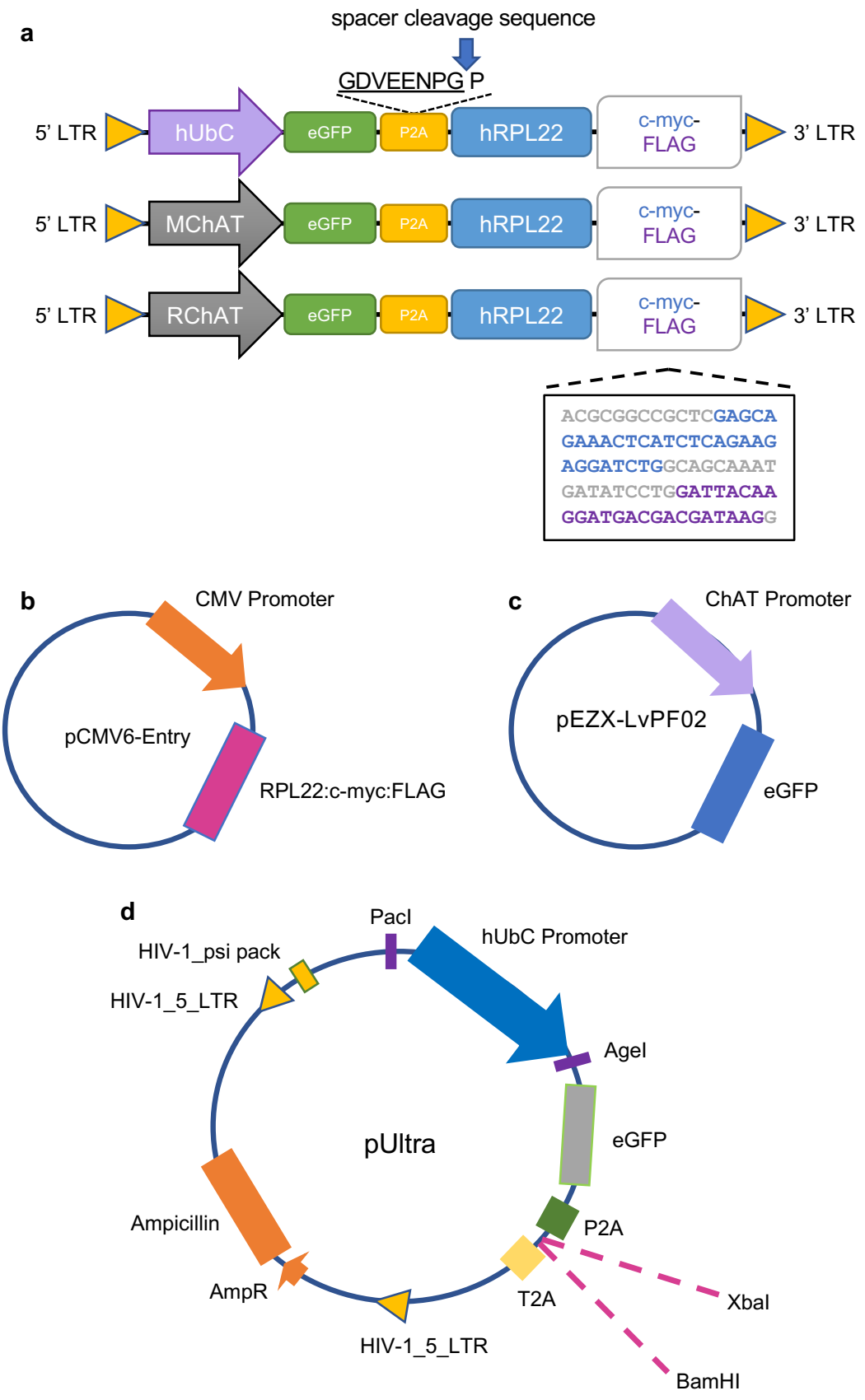


Figure 3.1. Commercially bought expression constructs containing the TRAP elements were used to clone lentiviral ChAT:TRAP vectors.

(a) The final constructs are designed to contain a predicted isoform of human choline acetyl transferase (ChAT) promoter, R or M. One open reading frame (ORF) contains cDNA encoding enhanced green fluorescent protein and RPL22 fused to c-myc and FLAG (FLAG) tags. These two regions are separated by a cis-acting hydrolase elements (CHYSEL), which contains an endogenous protease cleavage site.
(b) pCMV6-Entry contains RPL22-myc-FLAG under a constitutive promoter, CMV.
(c) pEZX-LvPF02 contains a ChAT promoter with an eGFP reporter. The ChAT promoter sequences are estimated through bioinformatic analysis. The human isoforms RChAT and MChAT have previously been found to have specific expression in human motor cortex and spinal cord, through post-mortem tissue staining.
(d) pUltra is a lenti-vector which was used as the backbone of the TRAP construct, with all the elements required to be packed into lentiviral particles (LTR domains). It contains a ubiquitous, mammalian promoter, hUbC, eGFP reporter and two CHYSEL sequences for bicistronic expression of multiple genes in the same ORF. Restriction sites labelled were used to clone in elements for TRAP construct.

3.2.2 Starting commercial vectors produce expected band sizes following restriction digest reactions

The constructs used to clone the TRAP vector were: *pCMV6-Entry* containing RPL22:c-myc:FLAG, under the ubiquitous human cytomegalovirus (CMV) promoter, *pUltra*, the final lentiviral backbone, readily containing eGFP and CHYSEL sequences, under the ubiquitous human ubiquitin C (hUbC) promoter, and *pEZX-LvPF02*, containing either the M- or RChAT promoter isoforms expressing an eGFP reporter (Figure 3.1b-d). Prior to beginning the cloning, preliminary experiments confirmed that all the commercial constructs produced DNA fragments of expected sizes following double and single restriction digest reactions (Figure 3.2a-c). These were estimated using the plasmid + insert sequences provided by the companies, and using online virtual digest tools.

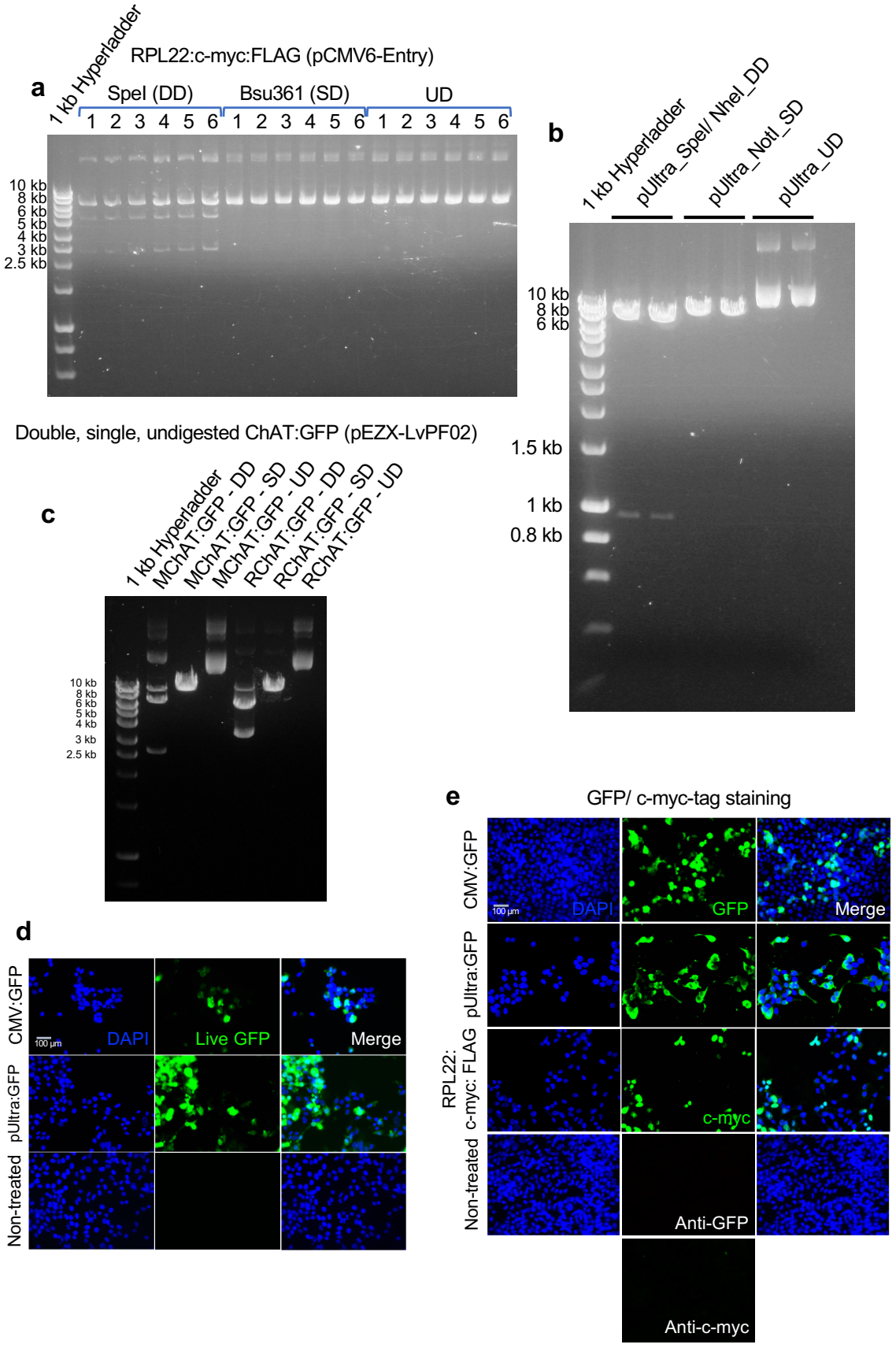
3.2.3 Starting commercial vectors are expressed in HEK293-T cells and human iPSC-motor neurons

Next, it was important to confirm that each construct contained expressible genes, promoters and tags, as desired in the final product. HEK293-T cells were transfected with pUltra and pCMV6-Entry starting plasmids, and CMV:GFP as a positive control (Figure 3.2d-e). The pUltra plasmid was confirmed to express eGFP under the hUbC promoter (Figure 3.2e). This was also true for live GFP expression (Figure 3.2d). Cells transfected with pCMV6-Entry (RPL22:c-myc:FLAG) were confirmed to express the FLAG (FLAG) tag (Figure 3.2e). ChAT:pEZX-LvPF02 vectors were packaged into lentiviruses (with CMV:GFP as a positive control), due to low efficiency of plasmid uptake into motor neurons by lipofection. Control (SFC840) and patient (C90202) iPSC-derived motor neurons (Day 28) were

transduced with these lentiviruses (Figure 3.2f-g). GFP-positive cells were seen to co-express with ChAT-positive and SMI-32-positive cells, though not with non-neuronal, undifferentiated cells. GFP in CMV:GFP-transduced cells expressed in both motor neurons and cells of unknown identity.

With confirmation that these plasmids efficiently express their transgenes and that the pEZX-LvPF02 vectors have active ChAT promoters in mature ChAT-positive iPSC-motor neurons, cloning of the TRAP construct could proceed.

Figure 3.2. Restriction digest reactions and immunofluorescence was used to confirm the sizes and expression of the primary constructs to be used for cloning of the TRAP vector.



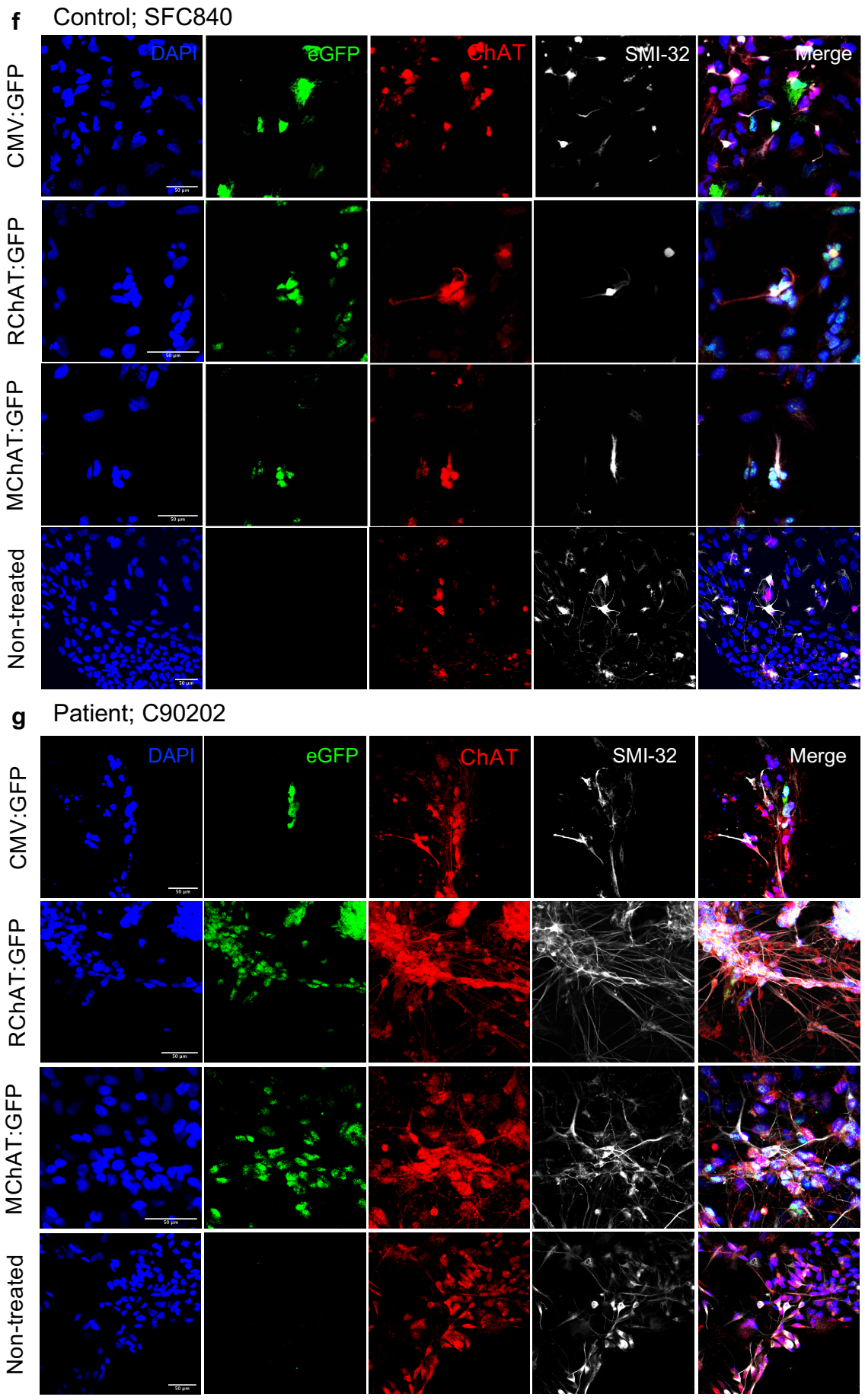

f
Control; SFC840
DAPI
eGFP
ChAT
SMI-32
Merge
CMV:GFP
RChAT:GFP
MChAT:GFP
Non-treated
g
Patient; C90202
DAPI
eGFP
ChAT
SMI-32
Merge
CMV:GFP
RChAT:GFP
MChAT:GFP
Non-treated

Figure 3.2. Confirmation of correct band sizes run on agarose gels following restriction digests of primary construct to be used for cloning of the TRAP vector, and expression in HEK293-T cells and iPSC-motor neurons via lipofection and lentiviral transduction, respectively, to confirm they can all actively express their elements.

(a) RPL22:c-myc:FLAG:pCMV6-Entry produced bands of size 4353 and 2406 bp from a double digest (DD; SpeI) and 6759 bp via a single digest (Bsu361) and undigested samples.
(b) pUltra produced bands of size 7380 and 947 bp from a double digest (DD; SpeI/NheI) and 8327 bp via a single digest (SD; NotI) and undigested (UD) samples.
(c) MChAT and RChAT:GFP (pEZX-LvPF02) constructs confirmed to digest into expected sizes: MChAT – 6044/ 3323 bp – MluI (DD), EcoRI (SD) and UD were 9367 bp. RChAT – 6694/ 2566 bp – MluI (DD), EcoRI (SD) and UD were 9260 bp.
(d) Transfection of the starter construct, pUltra:GFP, into HEK293-T cells showed expression of live GFP, with CMV:GFP as a known positive control. Scale bar = 100 µm
(e) Transfection of the starter constructs in HEK293-T cells and staining of fixed cells with anti-GFP and anti-c-myc confirmed expression of pUltra and RPL22:c-myc:FLAG under the hUbC reporter (pLentiC), respectively. Scale bar = 100 µm
(f) Lentiviral transduction of control (SFC840) iPSC-motor neurons (day 28) with fixed volume of ChAT:GFP and immunostaining of fixed motor neurons with GFP, ChAT and SMI-32.
(g) Lentiviral transduction of patient (C90202) iPSC-motor neurons (day 28) with fixed volume of ChAT:GFP and immunostaining of fixed motor neurons with GFP, ChAT and SMI-32. Scale bar = 20 µm

3.3 Cloning Strategy

A schematic of the cloning strategy for the TRAP construct has been outlined in Figure 3.3. Primers were designed (Table 2.2) to amplify RPL22:c-myc:FLAG from pCMV6-Entry via PCR (RPL22:c-myc:FLAG XbaI_BamHI) and introduce new restriction enzyme sites at the 5' and 3' ends of the PCR product, to be used as the insert in a ligation reaction with the lentiviral backbone, pUltra. The forward primer for this reaction contained a Kozak consensus sequence (GCCACC), as highlighted in blue, to increase ribosome binding to the RPL22:c-myc:FLAG ORF, hence enhancing translation of the transgene (Table 2.2). These primers anneal to the two ends of the RPL22:c-myc:FLAG ORF, with overhanging ends for the first two cycles of PCR, producing an insert of 503 bp (Figure 3.4a), which was purified via gel extraction. The RPL22:c-myc:FLAG insert and pUltra backbone were digested with XbaI/ BamHI restriction enzymes and mixed in 3:1 and 6:1 ratios for ligation. The reactions were transformed into competent *E. coli* and grown overnight on agar. *Colony whole insert PCR* (Table 2.2; 5' Junction RPL22 Insert-pUltra, 3' Junction RPL22 Insert-pUltra) was performed on the transformed colonies to screen for successful ligation reactions (Figure 3.4b). Successful ligations produced PCR products 730 bp in length, while pUltra templates resulted in 250 bp bands (Figure 3.4b). Plasmid DNA was isolated from successful colonies and restriction digests were performed to confirm that predicted band sizes would occur for double and single digests (Figure 3.4c). Sanger sequencing performed on the regions covering the hUbC promoter, eGFP, P2A and RPL22:c-myc:FLAG elements, confirmed that no unexpected mutations had been introduced (Figure 3.4d). The completion of this step refers to step 4 of the cloning strategy (Figure 3.3), producing the mammalian

Figure 3.3. TRAP vector cloning strategy.

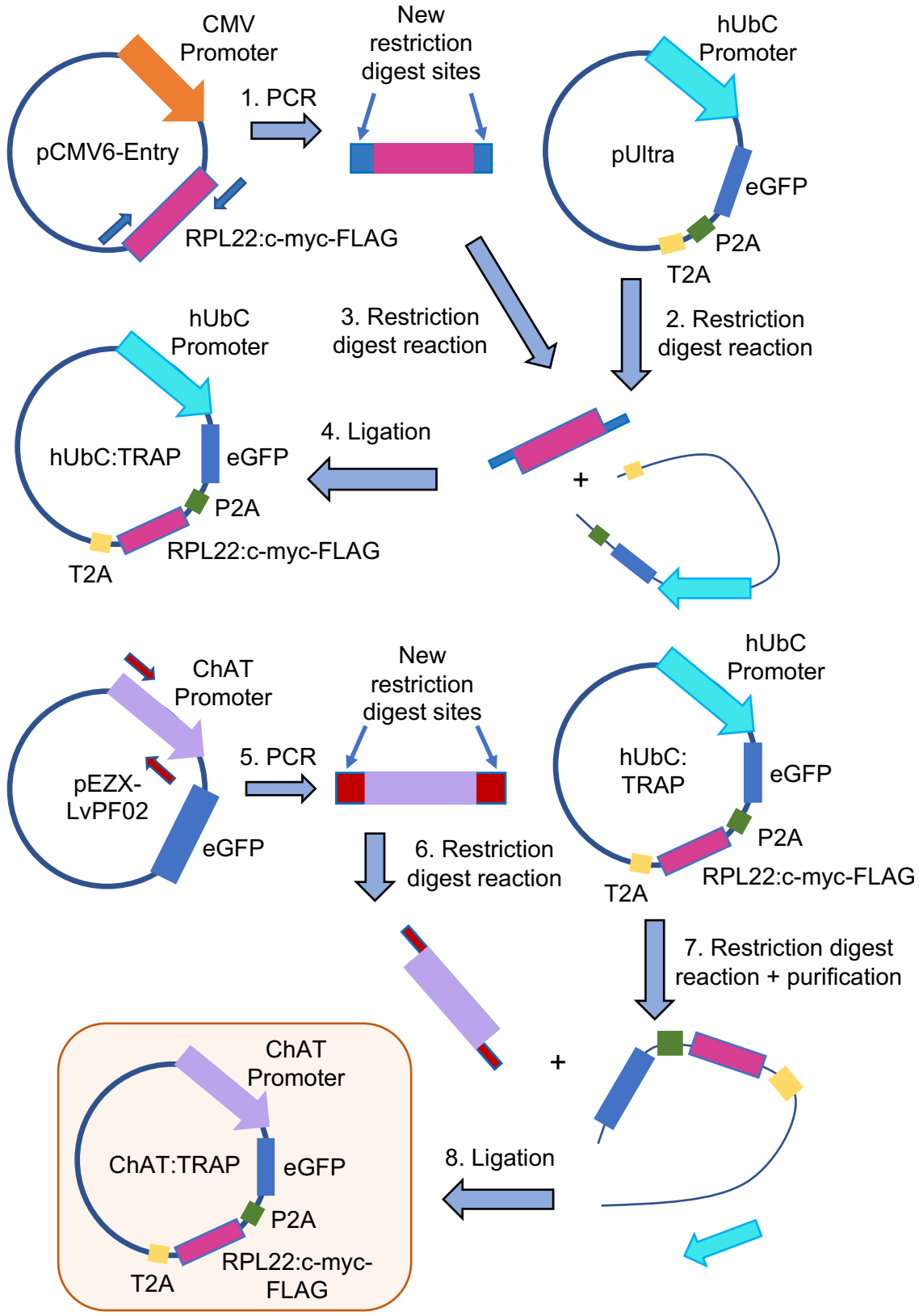


Figure 3.3. Schematic of TRAP vector cloning strategy.
Cloning strategy aimed to subclone RPL22:c-myc:FLAG from its original expression vector, while simultaneously introducing new restriction sites and ligating it into pUltra, lentiviral backbone. This results in hUbC:TRAP, the control TRAP construct. The hUbC promoter would be cut from this construct allowing the subcloned ChAT promoter to ligate in its place.

TRAP construct, hUbC:TRAP. This construct could still express eGFP and RPL22:c-myc:FLAG, as seen through transfection into HEK293-T cells via lipofection (Figure 3.4e).

Figure 3.4. Cloning of the TRAP construct (Figure 3.3, steps 1-4).

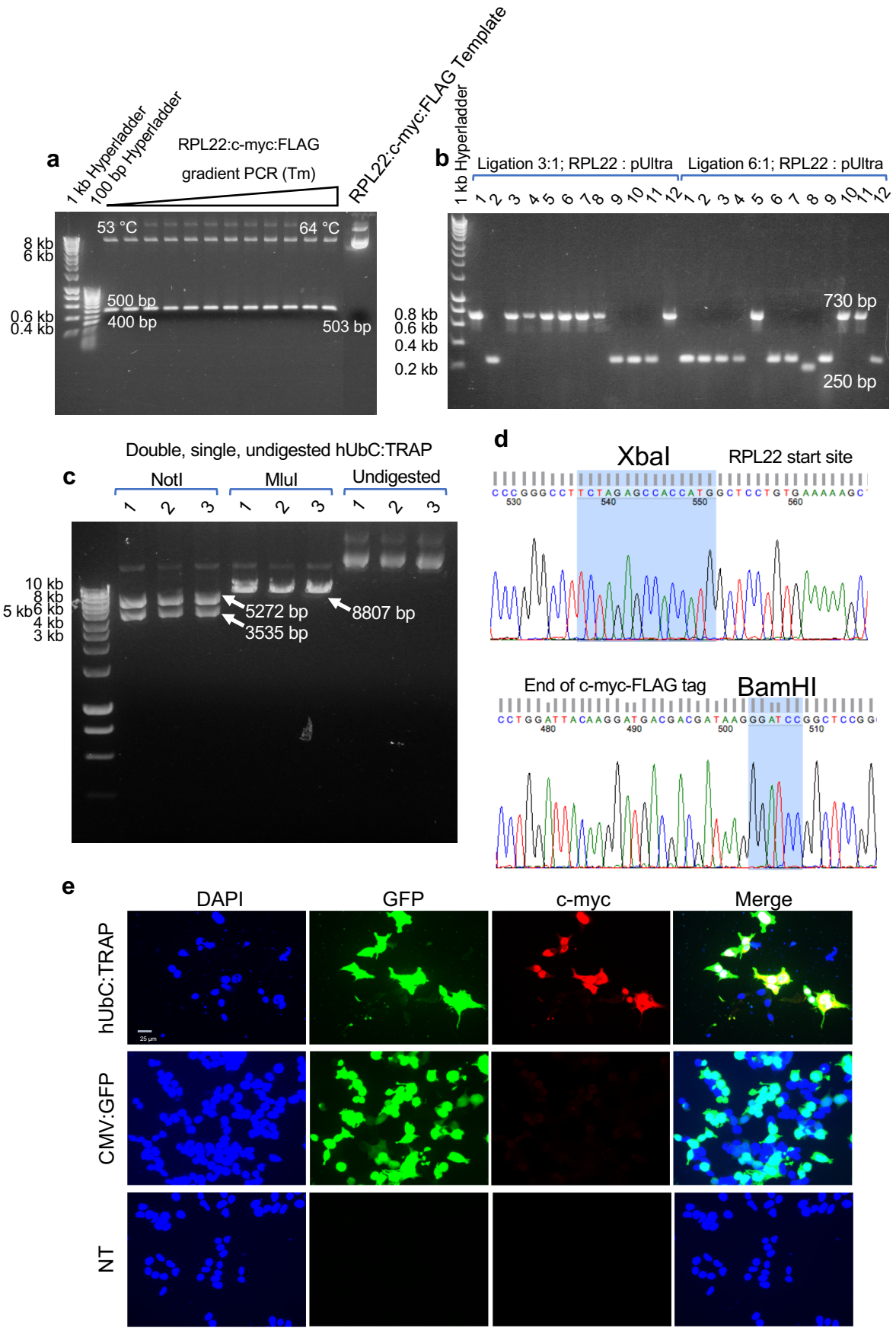


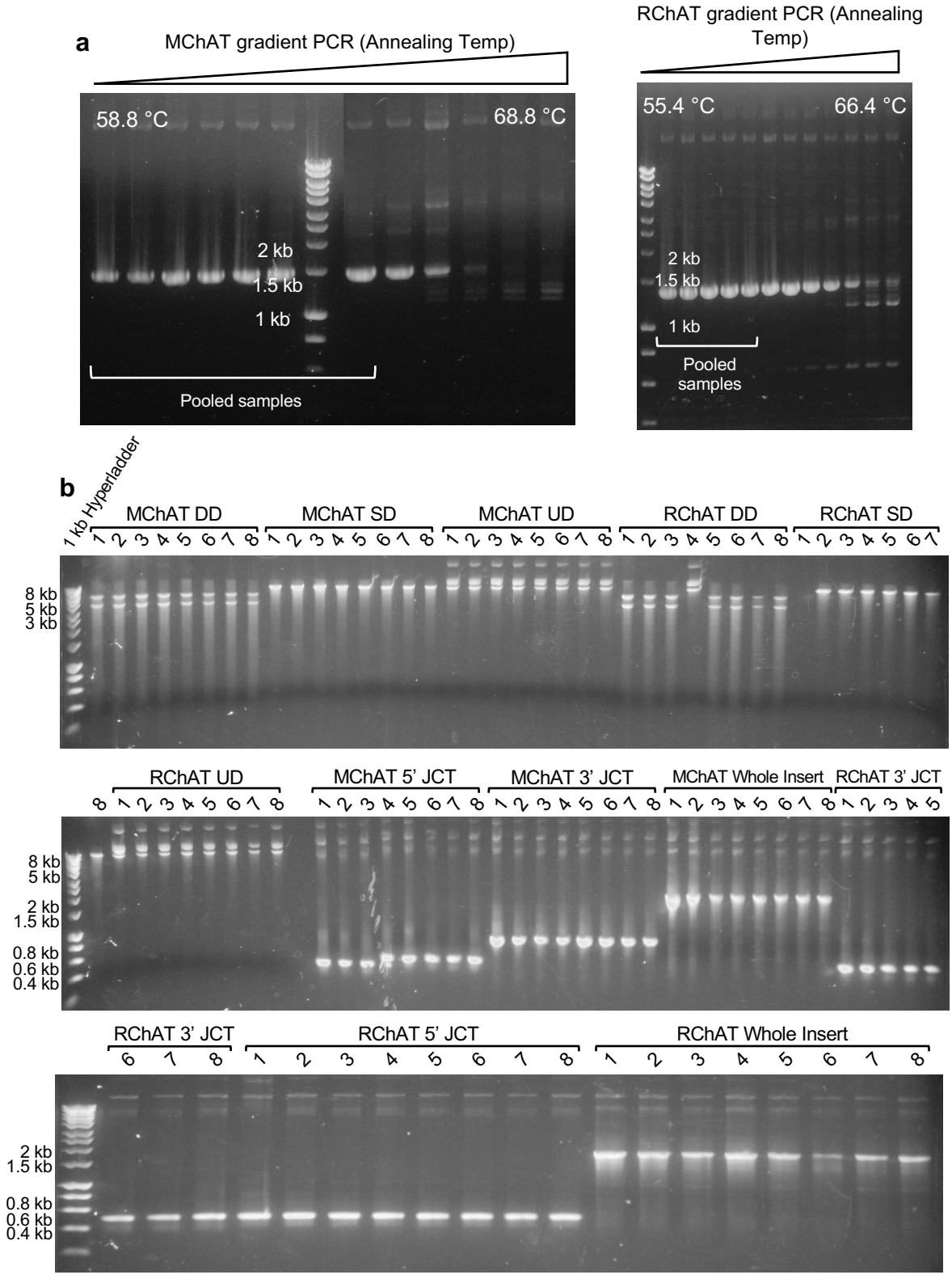
Figure 3.4. Agarose gel runs of restriction digests and PCR reactions confirming the successful intermediate steps of the TRAP vector cloning (steps 1-4, Figure 1d).

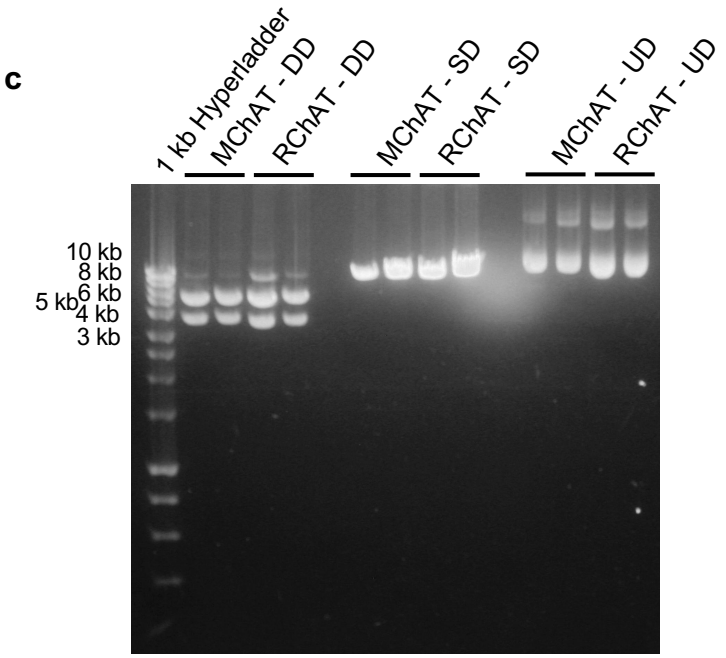
(a) A temperature gradient PCR reaction was run (annealing temperatures between 53 and 64°C) to amplify RPL22:c-myc:FLAG (503 bp) from its original construct (6739 bp).
(b) Transformed colonies of RPL22:c-myc:FLAG and pUltra ligation reactions were picked and run in colony PCR reactions using junction primers (Table 1) to screen for successful (730 bp) and unsuccessful (250 bp) ligations.
(c) Successful ligations were purified and digested with restriction enzymes to confirm expected band sizes are produced for control TRAP vector, hUbC:eGFP:RPL22:c-myc:FLAG (hUbC:TRAP): 5272/ 3535 bp (NotI, DD) and 8807 bp for SD and UD.
(d) Sanger sequencing of hUbC:TRAP at the regions of 5' and 3' inserted RPL22:c-myc:FLAG showed successful introduction of new restriction sites XbaI and BamHI (remaining insert regions was also sequenced and confirmed not to contain any mutations – data not shown).
(e) Transfection of hUbC:TRAP into HEK293-T cells confirmed that the FLAG tag and eGFP reporter are expressed. Scale bar = 25 μm

The second goal of the cloning was to remove the hUbC promoter from hUbC:TRAP and replace it with two ChAT promoter isoforms, M and R (Figure 3.3; step 5 onwards) in parallel. As before, overhanging primers (Table 2.2; MChAT_PacI_AgeI, RChAT_PacI_AgeI) containing restriction site sequences were used to subclone each ChAT promoter from its original construct via temperature gradient PCR (Figure 3.5a). Each PCR product was extracted from reactions without any non-specific bands. Simultaneously, a PacI/ AgeI double restriction digest was performed on hUbC:TRAP to remove the hUbC promoter from the lentiviral backbone and extracted after running it on an agarose gel (not shown). Following quantification, the ChAT promoters and extracted backbone were mixed in 3:1 and 6:1 ratios for ligation reactions. The reactions were transformed into competent bacteria, and spread onto agar plates. Junction PCR, whole insert PCR, and restriction digests were performed on the plasmids extracted from 4 mL overnight cultures to confirm expected band sizes (Figure 3.5b). Finally, a restriction digest was carried out on the final constructs extracted from 300 mL cultures, which was used for all future experiments (Figure 3.5c).

Sanger sequencing was performed covering the whole MChAT- and RChAT:TRAP constructs (Figure 3.5d) (Table 2.3). No mutations were found at the cloning restriction sites or transgenes. With the completion of this sequencing, the TRAP vector cloning was complete, and ready to be tested in iPSC-motor neuron cultures.

Figure 3.5. Continued cloning of the TRAP construct (Figure 3.3, steps 5-8).




c
1 kb Hyperladder
MChAT - DD
RChAT - DD
MChAT - SD
RChAT - SD
MChAT - UD
RChAT - UD
10 kb
8 kb
6 kb
5 kb
4 kb
3 kb

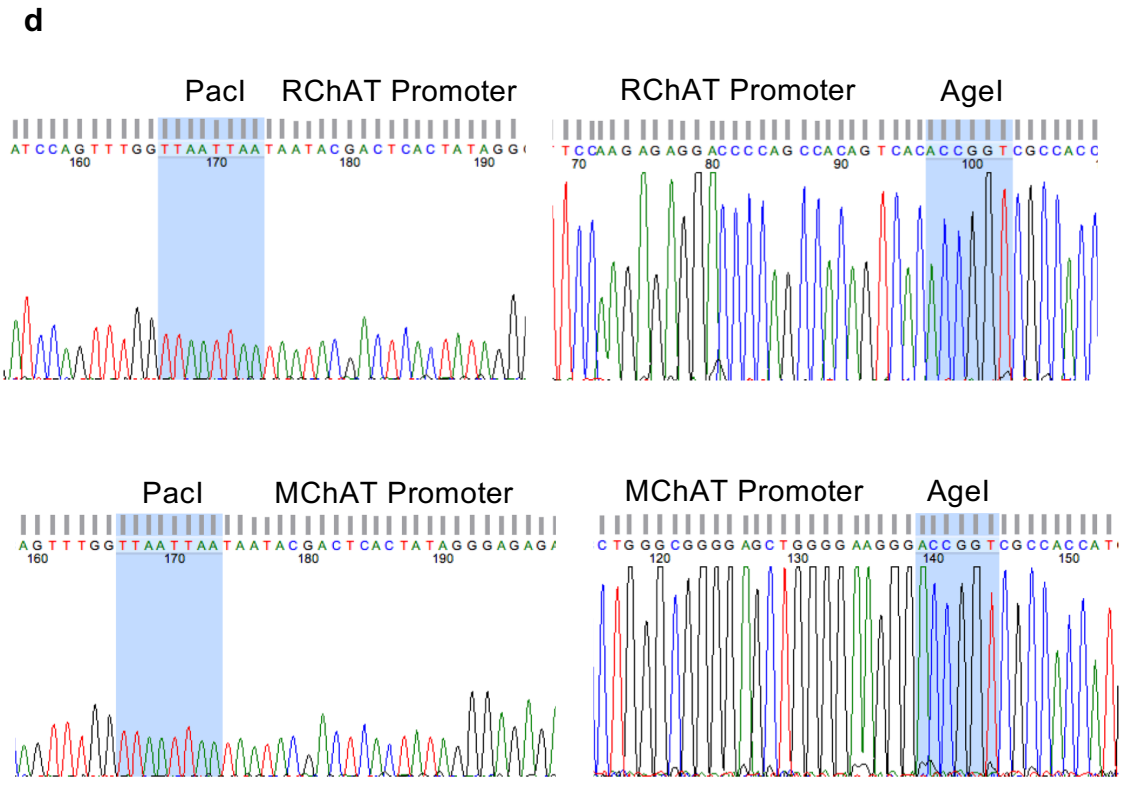

d
PacI
RChAT Promoter
RChAT Promoter
AgeI
PacI
MChAT Promoter
MChAT Promoter
AgeI

Figure 3.5. Continued cloning of the TRAP construct (Figure 3.3, steps 5-8), intermediate products from each step is identified used agarose gel electrophoresis.

(a) MChAT (1479 bp) and RChAT (1372 bp) promoters were amplified from their constructs via temperature gradient PCR (annealing temperatures of 58.8 to 68.8 °C and 55.4 to 66.4 °C, respectively), clean runs were pooled to purify via gel extraction.

(b) Confirmation of successful ligation of MChAT and RChAT promoter into control TRAP vector, after cutting out hUbC promoter was done using restriction enzyme digests, junction PCR of the 3' and 5' ends of the insert, and also the whole insert. Expected band sizes: MChAT:TRAP: 3'Junction = 513 bp, 5'Junction = 796 bp, whole insert junction (1820 bp (successful), 367 bp (unsuccessful)). RChAT:TRAP: 3'Junction = 493 bp, 5'Junction = 494 bp, whole insert junction (1713 bp (successful), 367 bp (unsuccessful)). Restriction digests were also performed (see (c) for sizes)

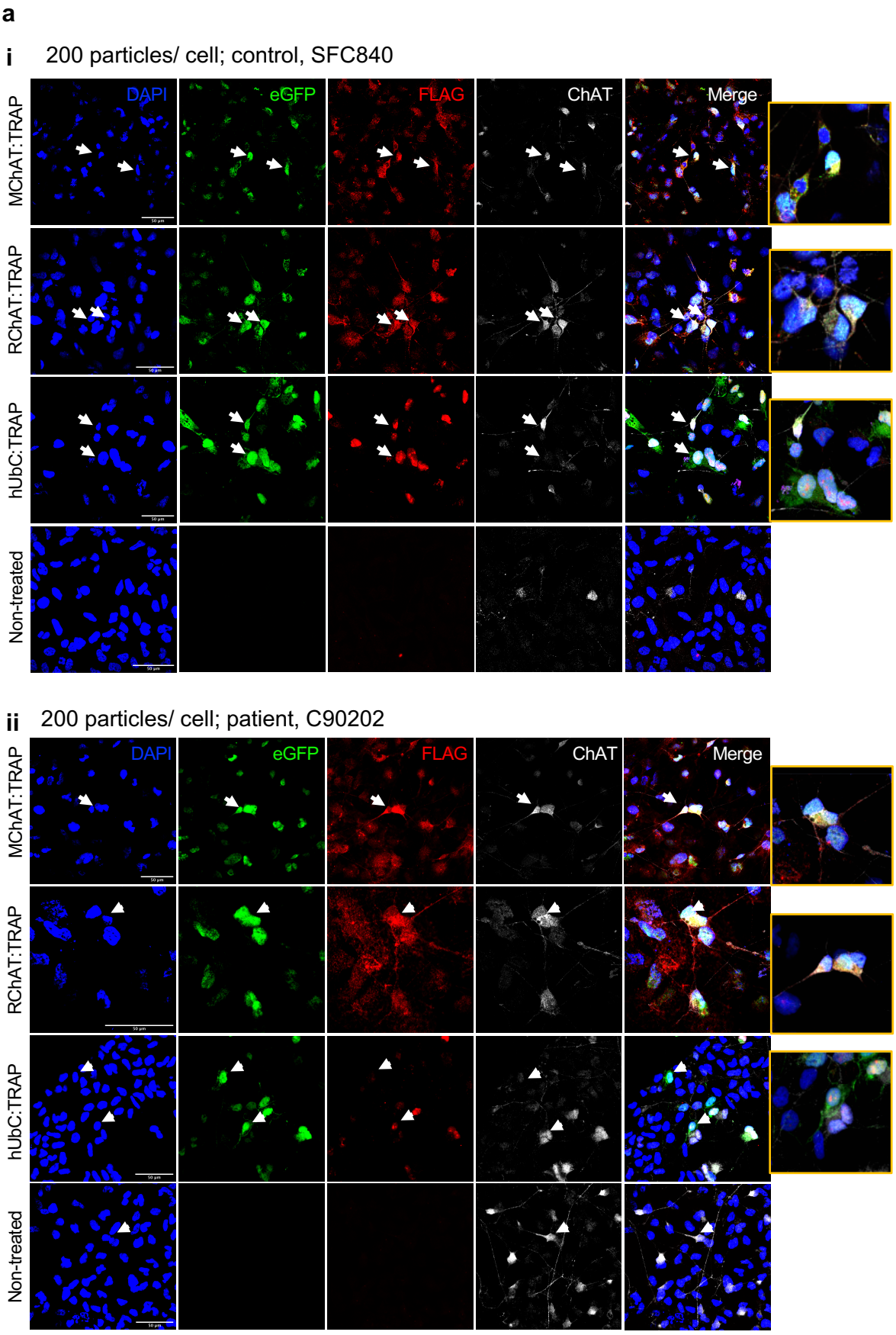
(c) Restriction digest of purified ChAT:TRAP vectors: MChAT = 5272/ 3631 bp (NotI, DD) and 8903 bp for SD (BamHI) and UD. RChAT = 5272/ 3524 bp (NotI, DD) and 8796 bp for SD (BamHI) and UD.

(d) Sequencing data shown is the 5' and 3' boundaries of the ligation sites. No mutations were seen here. Sanger sequencing of the whole plasmid, including all major regions, such as the LTR domains was done (data not shown).

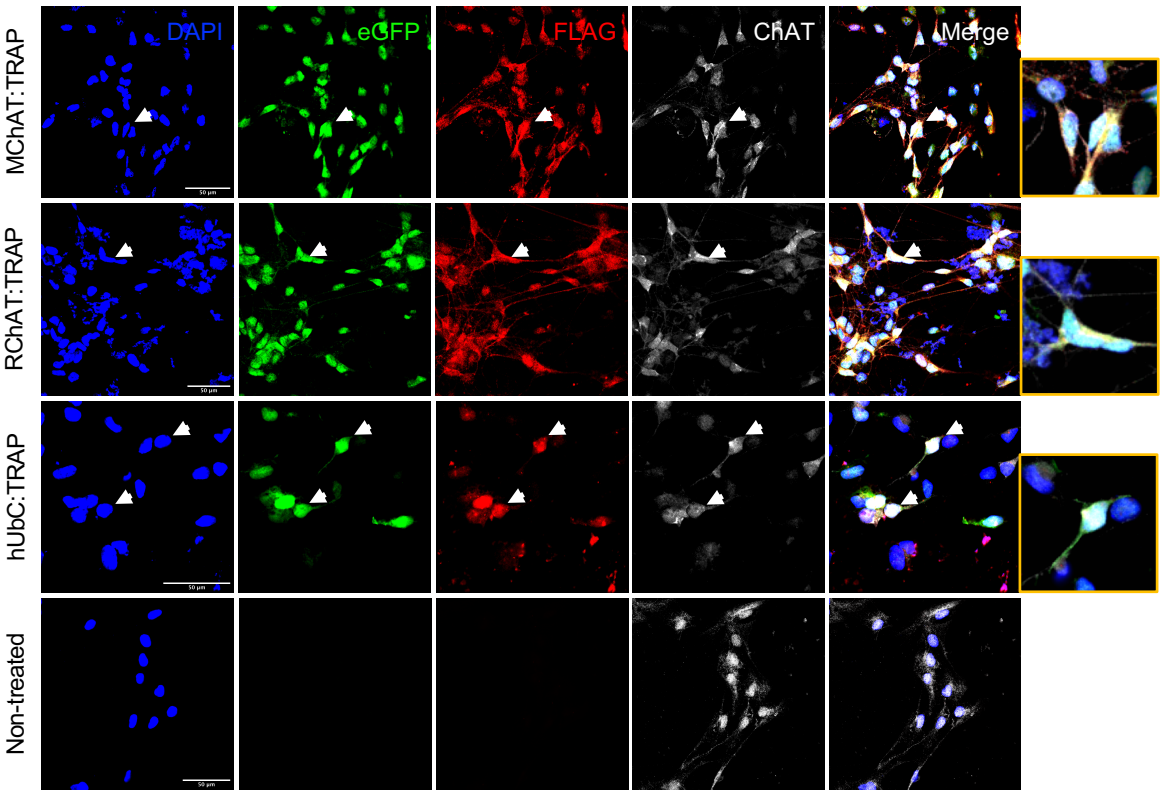
3.4 TRAP vector efficiently co-expresses eGFP and tagged RPL22 in target cells

A control (SFC840) and *C9orf72* ALS/ FTD patient-derived (C90202) iPSC-motor neuron line were transduced on day 27 of differentiation with MChAT:TRAP, RChAT:TRAP and hUbC:TRAP lentivirus, with a concentration of 200 or 500 particles/ cells, and analysed 72 hours later (Day 30) (Figure 3.6a). GFP-positive cells are present in both ChAT:TRAP-transduced cells and overlap with FLAG-expressing cells, in only ChAT-positive neurons. The proportion of cells expressing GFP and FLAG appears to have increased in cells transduced at 500 particles/ cell compared with 200 particles/ cell. Cultures transduced with hUbC:TRAP, expressing GFP also co-express with FLAG in all cell types, including non-ChAT-positive cells. Notably, the pattern of FLAG staining shows localisation in the nucleus and cytoplasm. This is consistent with studies implicating RPL22 in early ribosome synthesis in the nucleolus, and a role in translation. This suggests that the transgenic RPL22:c-myc:FLAG is entering endogenous pathways, and more likely to form complexes with ribosomes, which is crucial for TRAP. Furthermore, the pattern of GFP staining is distinct from FLAG, showing that the two proteins are differentially localised. This signifies that the P2A CHYSEL sites have successfully been cleaved by endogenous proteases. There were no obvious signs of cellular toxicity, such as mass cell death, by increasing the lentiviral concentration. Lentiviral transduction optimisation and assessment of associated toxicity in iPSC-motor neurons is explored further in chapter 4.

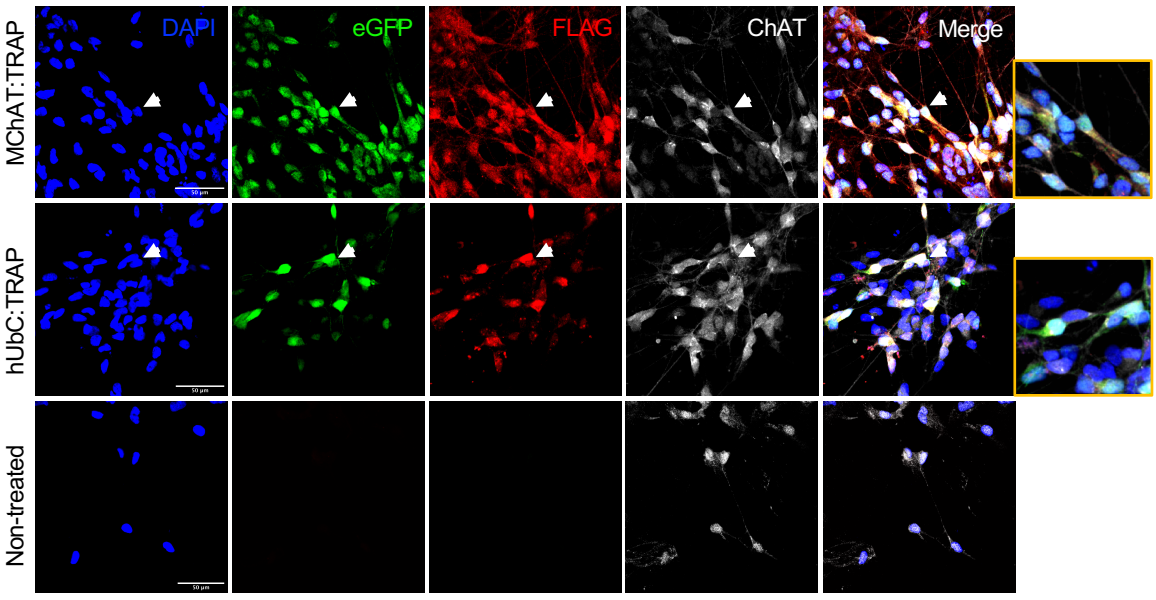
Figure 3.6. Immunostaining of HEK293-T cells and iPSC-motor neurons following lentiviral transduction TRAP constructs.



iii 500 particles/ cell; control SFC840



iv 500 particles/ cell; patient C90202



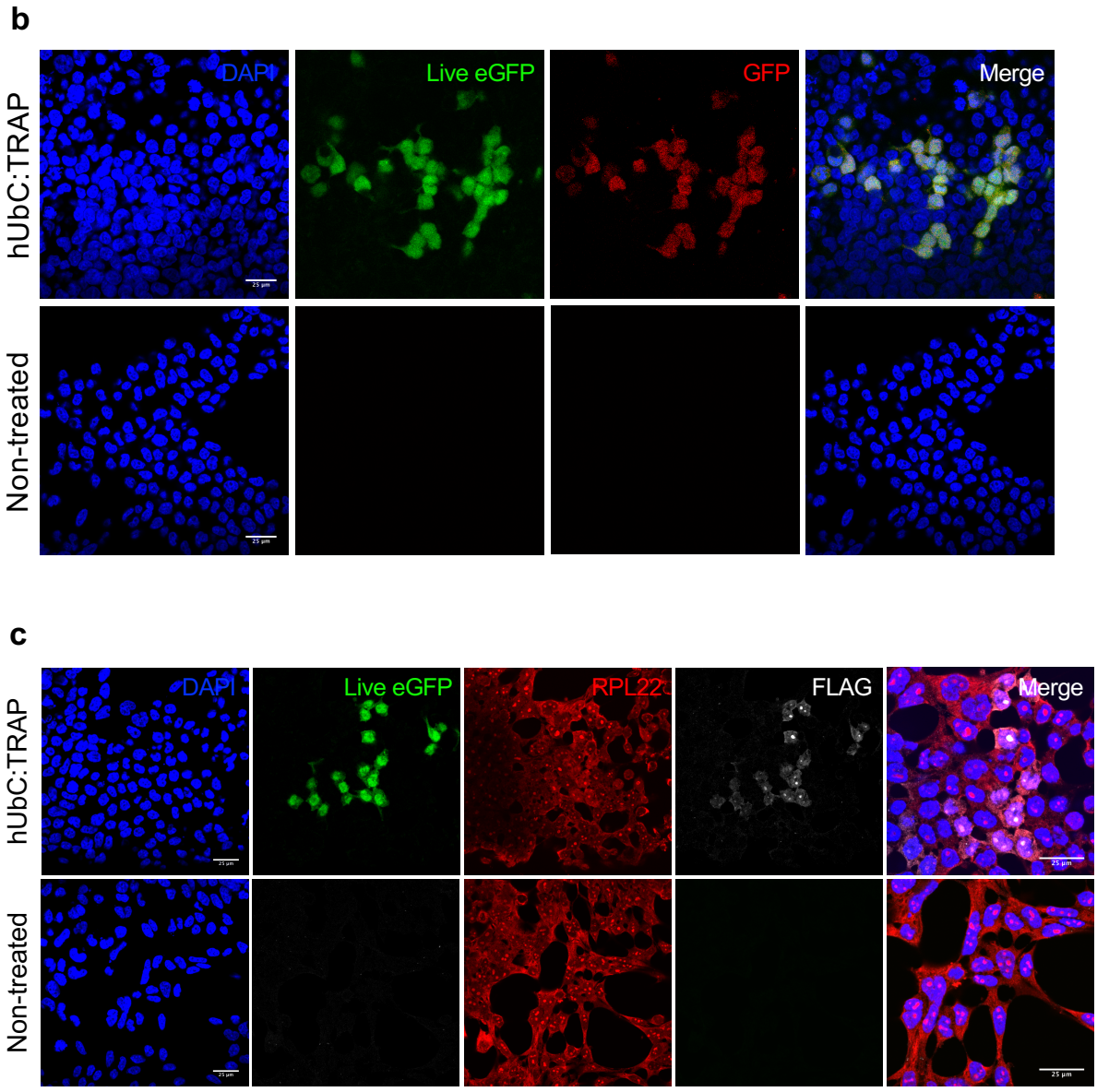


Figure 3.6. Immunostaining of HEK293-T cells and iPSC-motor neurons following lentiviral transduction of hUbC-, MChAT- or RChAT:TRAP constructs. Particles packaged with hUbC:TRAP, and non-transduced samples were used as controls.

(a) Concentrations of 200 and 500 particles/ cell were used to transduce iPSC-motor neuron. Cells were stained with anti-GFP and anti-FLAG (FLAG) to show co-expression of these transgenes in ChAT-positive motor neurons. Scale bar = 50 µm

(b) Co-fluorescence of live GFP in HEK293-T cells transduced with hUbC:TRAP and anti-GFP staining reveals co-expression, and no false signals. Scale bar = 25 µm

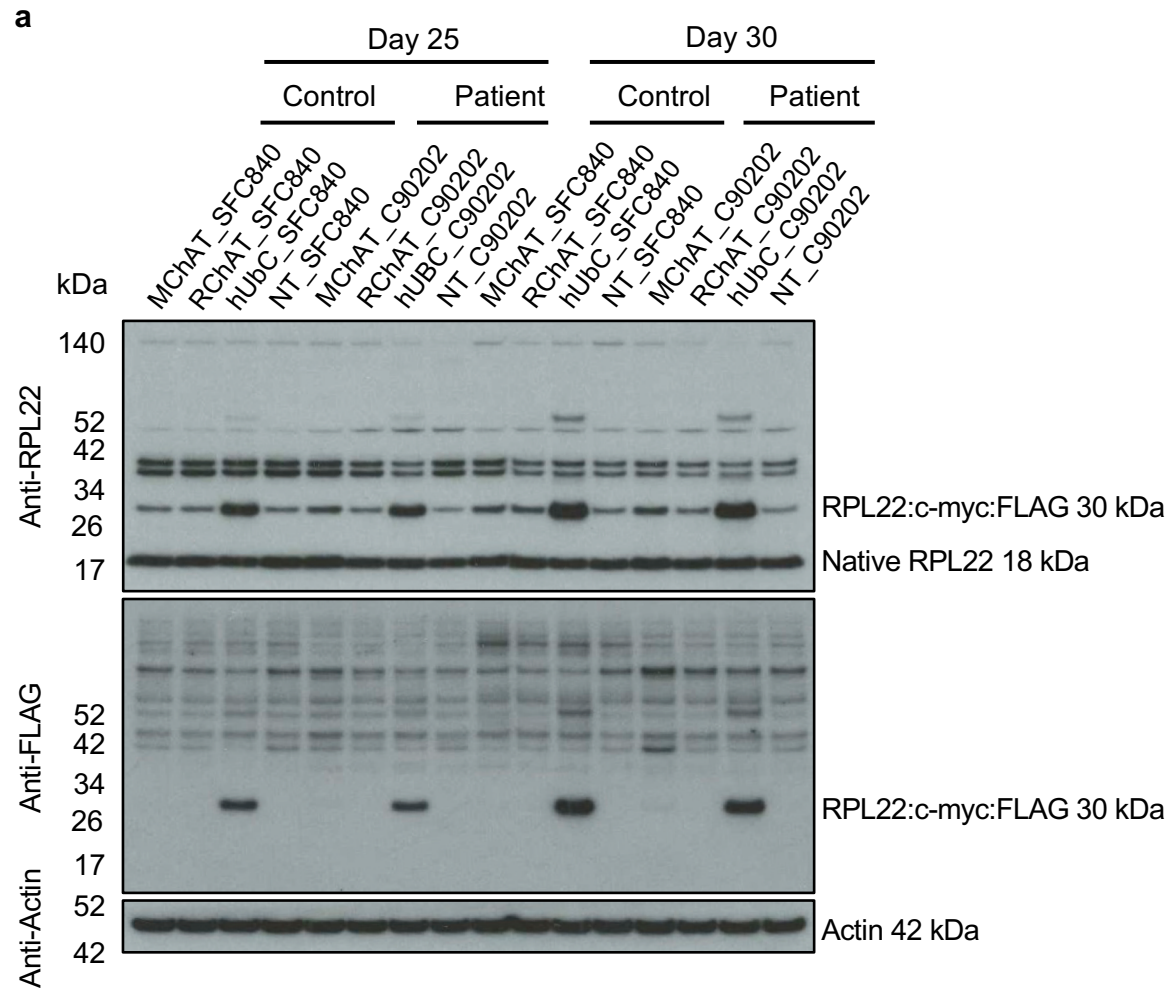
(c) Immunostaining of HEK293-T cells transduced with hUbC:TRAP lentivirus showed that FLAG tag expression overlapped with live GFP. Anti-RPL22 was expected to stain ubiquitously, but FLAG staining pattern in transduced cells matched that of non-transduced cells. This suggests that RPL22:c-myc:FLAG is localising to cellular compartments with the endogenous protein. Scale bar = 25 µm

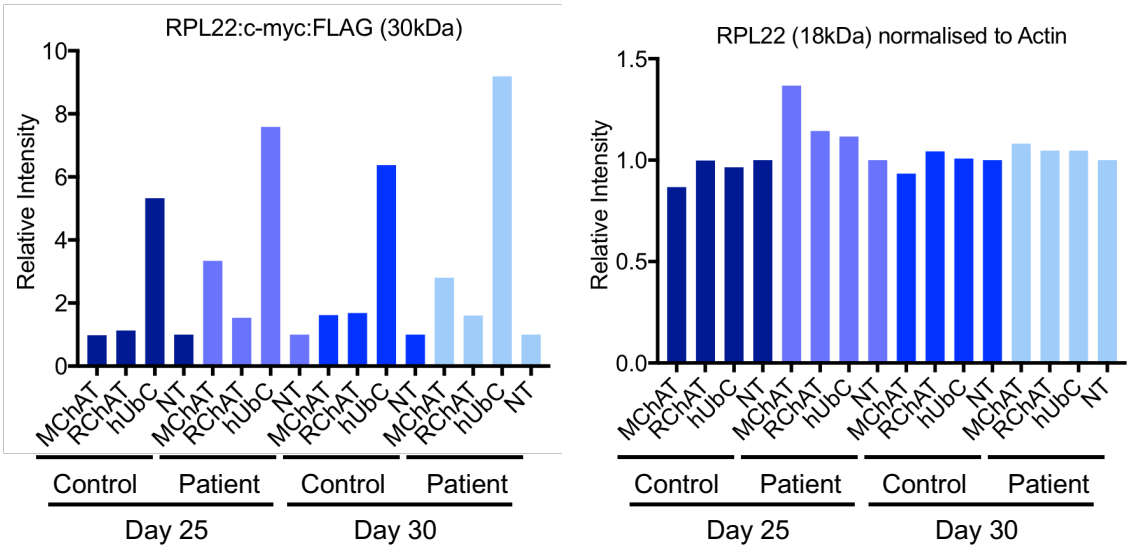
Several steps of the TRAP protocol were optimised using HEK293-T cells, therefore these cultures were transduced with hUbC:TRAP lentivirus to assess expression of the vector under the mammalian promoter (Figure 3.6b-c). Live eGFP co-expression with anti-GFP staining was clear with no visible artefacts, confirming the specificity of the antibody used for this model (Figure 3.6b). The expression of c-myc tag overlapped with cells also expressing live GFP, though they did not colocalise, similar to expression in iPSC-motor neurons, confirming the success of the peptide cleavage sequence (Figure 3.6c). Staining pattern of overexpressed c-myc also matches that of endogenous RPL22.

Immunoblotting of protein extracted from lentivirus-transduced iPSC-motor neurons on day 25 and day 30 of differentiation, showed a number of significant bands for RPL22 staining. RPL22 protein is approximately 15 kDa in size, and has been observed to be detected between 17-18 kDa by immunoblotting (as reported by the antibody supplier). Bands of similar intensity were seen in all samples for RPL22 antibody staining at ~18 kDa, which was estimated to be the native RPL22 protein (Figure 3.7a). This is quantified in Figure 3.7b. RPL22:c-myc:FLAG is estimated to be ~18 kDa. This is not much larger than the endogenous protein, and thus expected to run at a similar rate to the native protein. However, staining of the membrane with FLAG showed a set of bands for hUbC:TRAP-transduced cultures. These bands were at approximately 30 kDa on the membrane, but overlapped with the “RPL22:c-myc:FLAG 30 kDa” bands, seen with RPL22 staining. The strongest band for RPL22:c-myc:FLAG occurs for hUbC:TRAP-transduced cells, and to a lesser intensity in M-, RChAT:TRAP- transduced cells and non-transduced cells (Figure 3.7a-b). FLAG staining also showed a number of larger bands, none of which overlapped with native RPL22 staining. This implies that the 30 kDa RPL22

band is likely to be the tagged protein, although no bands could be detected in the lanes for M- and RChAT:TRAP-treated cells (Figure 3.7a-b). GFP staining was not performed in here, though GFP:P2A:RPL22:c-myc:FLAG is estimated to be 48 kDa. This analysis would reveal the ratio of exogenous protein which has been cleaved at the P2A bicistronic cleavage site, to non-cleaved protein.

Figure 3.7. Immunoblotting analysis of protein isolated from mature iPSC-motor neurons transduced with TRAP vectors.





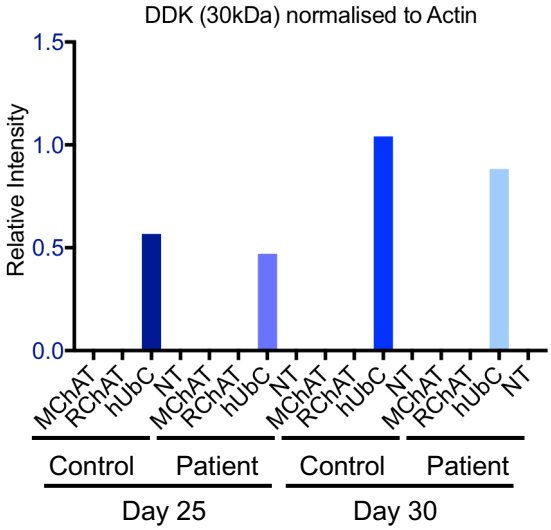


Figure 3.7. Western blot analysis of protein isolated from mature iPSC-motor neurons transduced with MChAT:TRAP and RChAT:TRAP.

(a) The membrane was probed with rabbit anti-RPL22 to show the existence of two species of RPL22; endogenous and FLAG-tagged protein. The same membrane was stripped and probed mouse anti-FLAG (FLAG). In this case only one significant band is expected, representing the FLAG-tagged RPL22 transgenic protein, though not the endogenous protein.

(b) Quantification of the bands shown in (a). Actin was used as a housekeeping gene for all normalisations.

3.5 Conclusion

With the evidence that the cloned human TRAP vector expresses in the models of interest we wish to investigate, in an expected way, the implementation and optimisation of the TRAP protocol for purifying mRNA-associated with translating ribosomes was proceeded to.

4 Optimisation of translating ribosome affinity purification from an iPSC-derived motor neuron model of patient *C9orf72* amyotrophic lateral sclerosis

4.1 Introduction

Following cloning of the TRAP vectors, and confirmation that they are able to express their encoded proteins in both a mammalian, HEK293-T immortalised line, and iPSC-derived motor neurons, focus was shifted to optimising lentiviral transduction and mRNA-pulldown via ChAT promoter-induced expression of RPL22:c-myc:FLAG. Several challenges needed to be overcome to produce a significant transduction yield, and subsequently an optimal protocol to purify enough high quality, ChAT-enriched mRNA for RNA sequencing. It should be noted that little difference in expression of TRAP vectors under MChAT and RChAT promoters was observed from preliminary data (Figure 4.1). RChAT promoter-containing vector expressed eGFP slightly more efficiently in iPSC-motor neurons, than MChAT, when transduced at 1000 particles/ cell (Figure 4.1b), though there was no difference with 3000 particles/ cell (Figure 4.1b). Only the RChAT-containing vector was used in all further experiments to optimise the TRAP protocol and to pulldown mRNA for RNA sequencing. Henceforth, the full TRAP constructs will be described as hUbC:TRAP and RChAT:TRAP.

Figure 4.1. Lentiviral transduction efficiency of RChAT:TRAP and MChAT:TRAP are slightly different, as shown by expression of reporters.

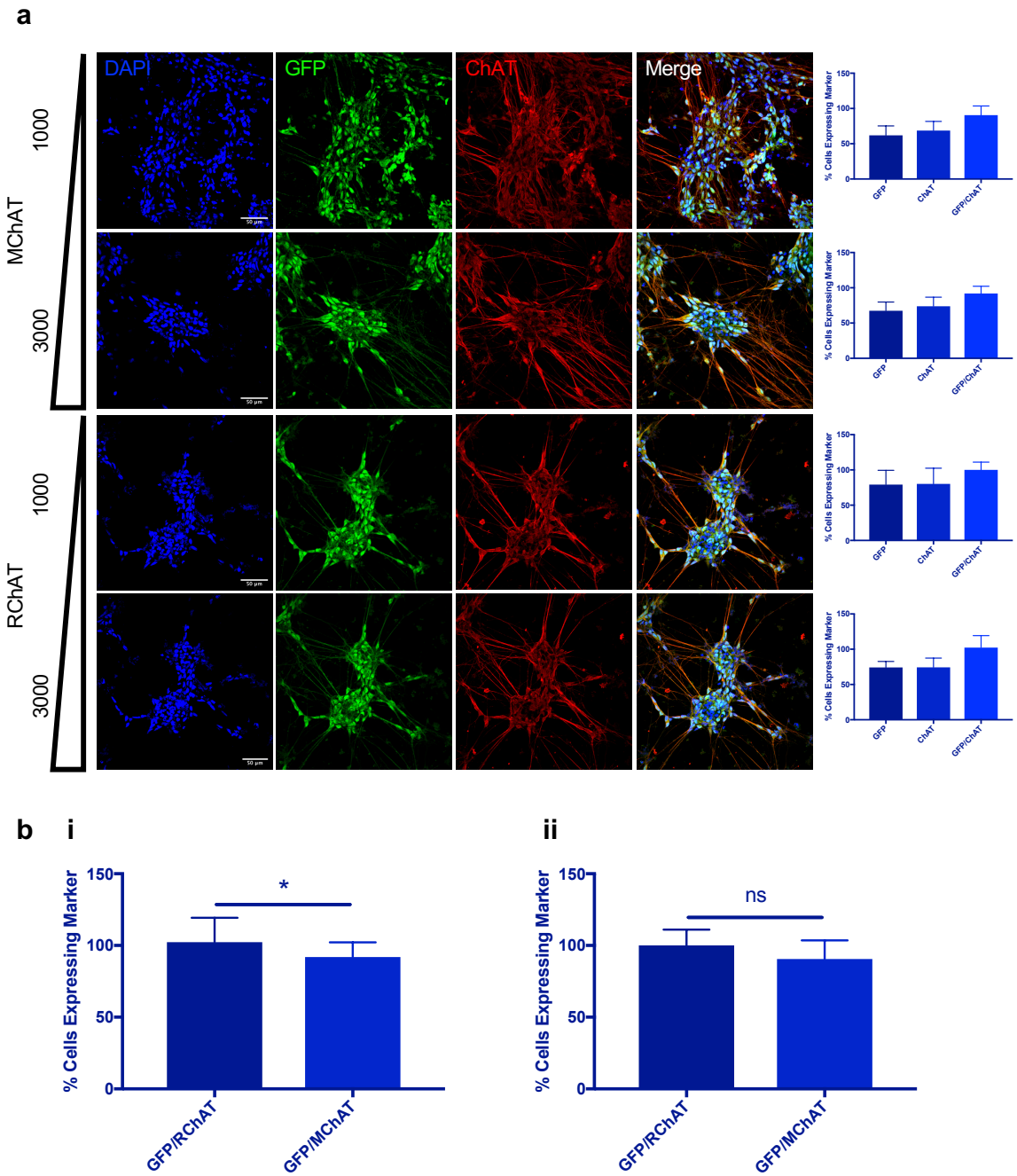


Figure 4.1. Comparison of the expression of RChAT:TRAP and MChAT:TRAP following transduction of lentivirus.

(a) Immunostaining of iPSC-motor neurons on day 28 of the protocol, transduced with 1000 or 3000 particles/ cell of ChAT:TRAP vector lentivirus with either the RChAT or MChAT promoter isoform. Scale bar = 50 μm

(b) Quantification and comparison of transduction efficiencies at two concentrations between the two ChAT:TRAP vectors shows that RChAT:TRAP vector is more significantly expressed at 1000 particles/ cell (i), though no difference is seen at 3000 particles/ cell (ii). For every condition, at least 500 cells were counted, in >5 fields across 1-2 coverslips (technical replicates). Biological replicate n = 1; Student's T-test, unpaired.

4.2 High co-expression of TRAP vector components and ChAT-expressing iPSC-derived motor neurons was achieved

Third generation lentivirus particles were generated in-house via transfection of 3 plasmids (Gag/ pol, REV and VSV-G) with a TRAP vector into HEK293-T cells, and purified from conditioned media (Figure 4.2). Titration was performed by extracting lentiviral RNA and performing qRT-PCR by targeting the 5' and 3' long terminal repeats (LTR). Further titration by performing transduction of various concentrations of lentivirus and determining expression efficiency could not be performed reliably for each new production of lentivirus as each batch would need to be titrated on a new differentiation of iPSC-motor neurons. This step would be highly time consuming, with each motor neuron differentiation taking over 30 days, and unlikely to produce consistent motor neuron yields. Consequently, the qRT-PCR-determined concentration of lentivirus was kept consistent, with cultures treated the same to reduce experimental variation. Thus, lentiviral particle transductions will be referred to in terms of concentration (particles/ cell) used to treat cultures, rather than multiplicity of infection (MOI).

Figure 4.2. Workflow to generate lentiviral particles from cloned TRAP vectors using 3rd generation lentiviral plasmids.

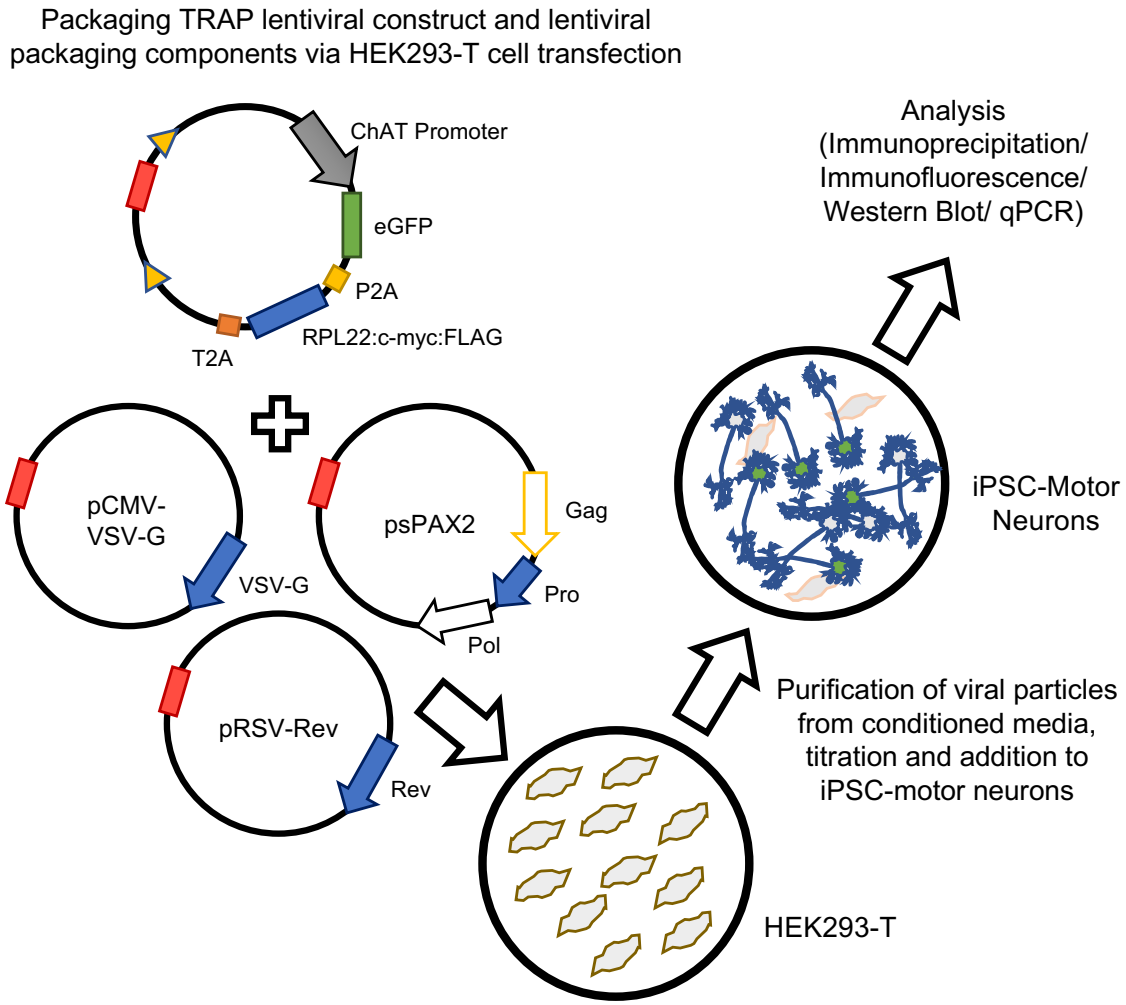


Figure 4.2. Schematic of the workflow to generate lentiviral particles from previously cloned TRAP vectors using 3rd generation lentiviral plasmids encoding proteins Gag/ Pol, VSV-G and Rev, by co-transfecting HEK293-T cells. Collection of conditioned media is done 72 hours post-transfection followed by purification and titration of the particles. Motor neurons derived from human iPSCs can then be transduced for several downstream applications, including TRAP.

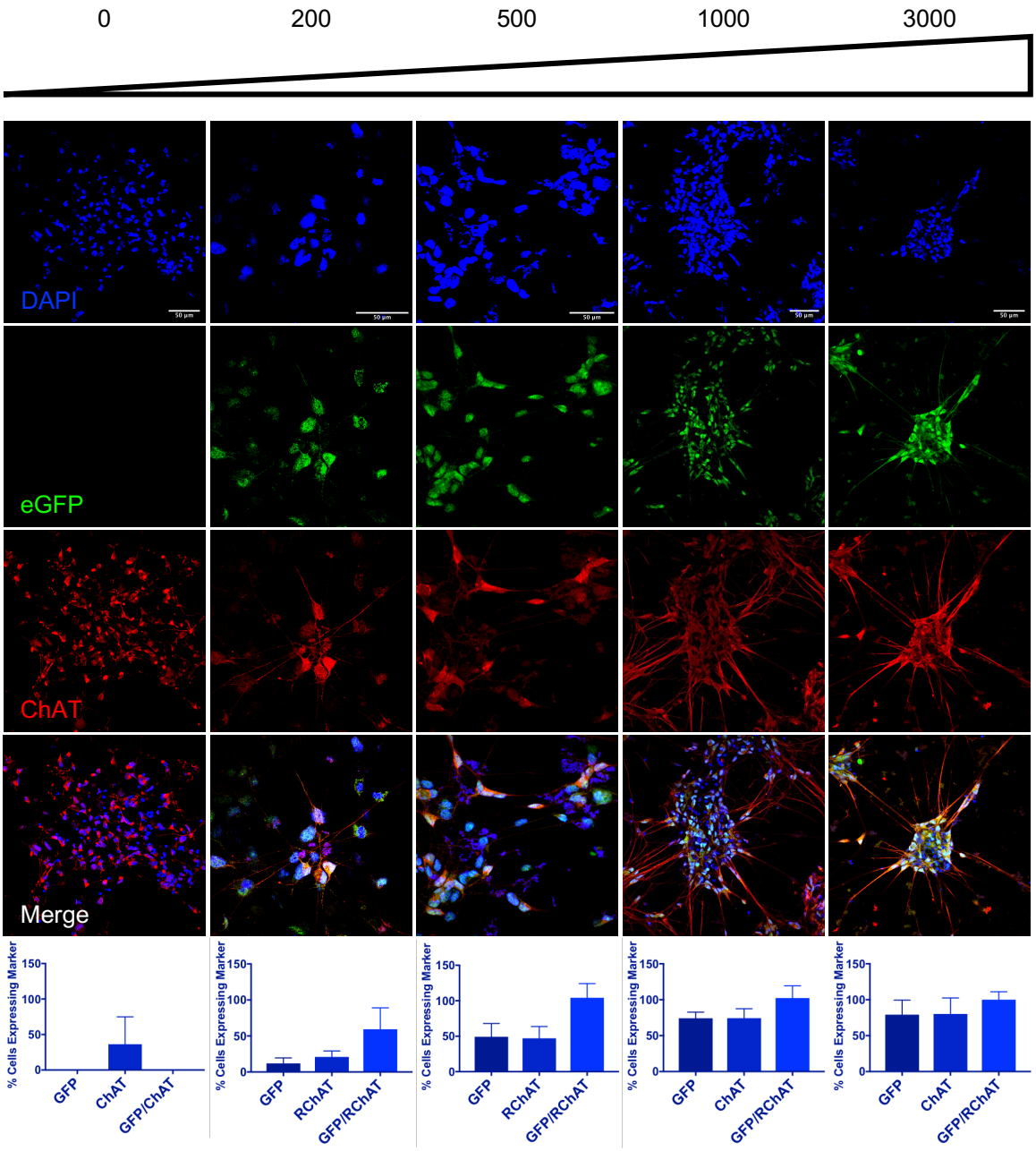
4.2.1 Optimal lentiviral transduction concentration varies across differentiated iPSC-motor neuron lines

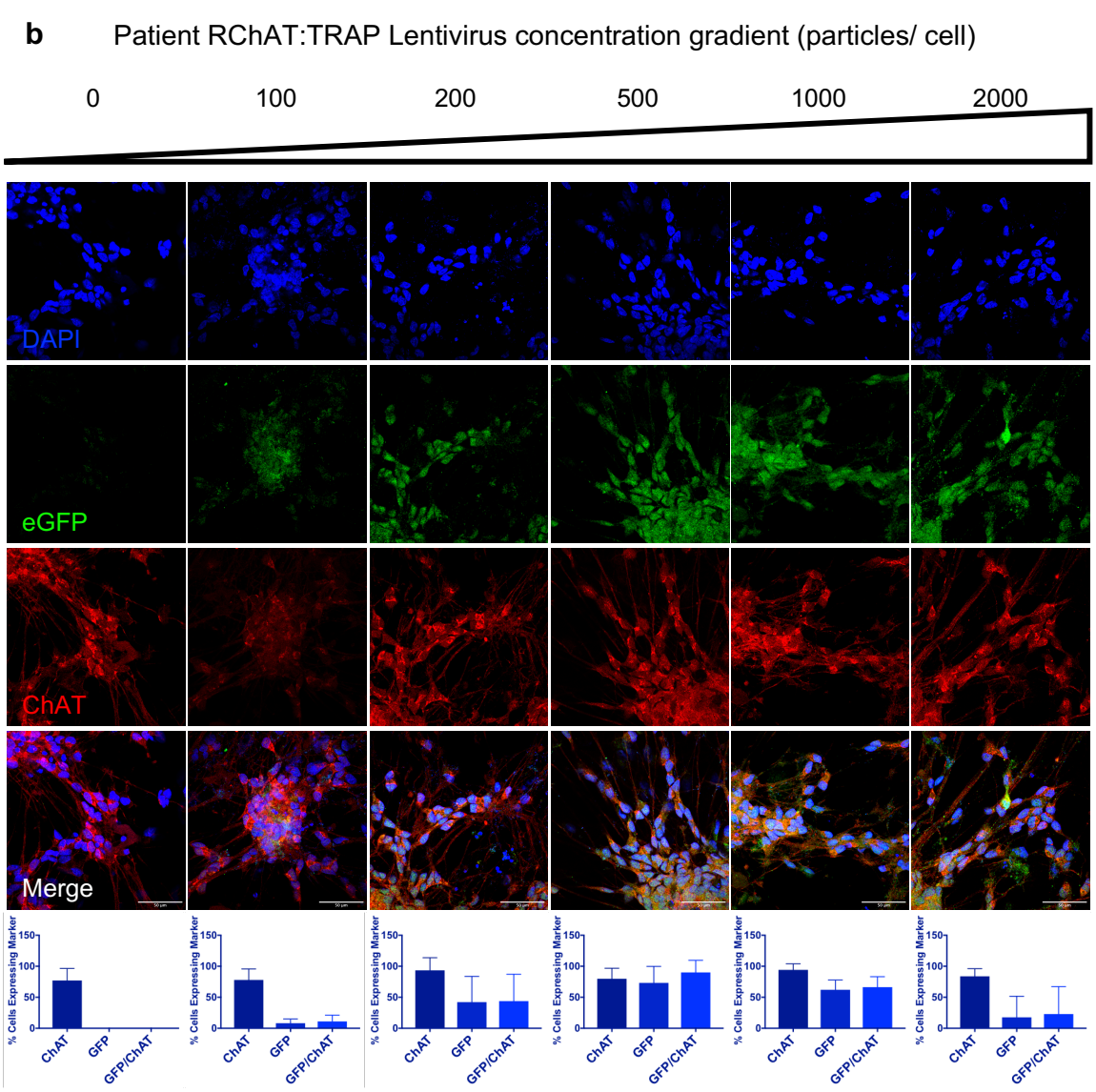
Concentrations from 100 to 10,000 particles/ cell were tested on patient, age-matched control and CRISPR/Cas9-edited isogenic lines (Figure 4.3). Concentrations above 2000 particles/ cell were not tolerated in most cases by all cultures, leading to cell death in less than 48 hours of treatment. In these cultures, neurons were found to have lifted off the plate surface. Concentrations below 2000 particles/ cell were further examined for live eGFP expression using live microscopy at regular timepoints of 24, 48, 72 and 96 hours, and subsequently analysed by confocal microscopy imaging, following fixation and immunostaining. At least 72 hours were required for expression of transduced TRAP vectors in iPSC-motor neuron cultures. Live eGFP fluorescence was not visible in all cultures transduced with RChAT:TRAP (even 96 hours post-transduction), though fixation and immunostaining revealed successful transductions. This is suspected to be because the ChAT promoter is not as strong as a ubiquitous mammalian promoter to result in strong overexpression. Transduction of the hUbC:TRAP construct into HEK293-T cell and iPSC-motor neuron cultures gave stronger expression and allowed better assessment of expression via live expression imaging (Figure 4.4). This can also be attributed to the non-ChAT-positive cells, including non-neuronal cells, also taking up and expressing the vector. Rare successful live imaging of motor neuron cultures transduced with 1000 particles/ cell and 8000 particles/ cell using hUbC:TRAP lentivirus are given here (Figure 4.4). 500-1000 particles/ cell were found to be the most efficient and reliable concentrations in obtaining up to 100% co-expression with ChAT marker in mature age-matched control and patient motor neurons on day 28 of the differentiation protocol, transduction on day 24, 96

hours previously (Figure 4.3a-b). Staining of cleaved caspase 3 in cultures treated with 1000 particles/ cell for 4 hours, and fixed after 96 hours showed that there was no difference in this apoptotic marker compared to non-transduced iPSC-motor neurons (Figure 4.3f). This suggests that any initial toxicity that may have been induced by lentiviral transduction is returned to basal levels by the time of harvest. This should control for any significant RNA sequencing hits which could have potentially been attributed to experimental variance from transduction. Given that this is the highest concentration of lentivirus that can be tolerated by the cultures, it was decided to proceed with these transduction conditions in the final TRAP protocol.

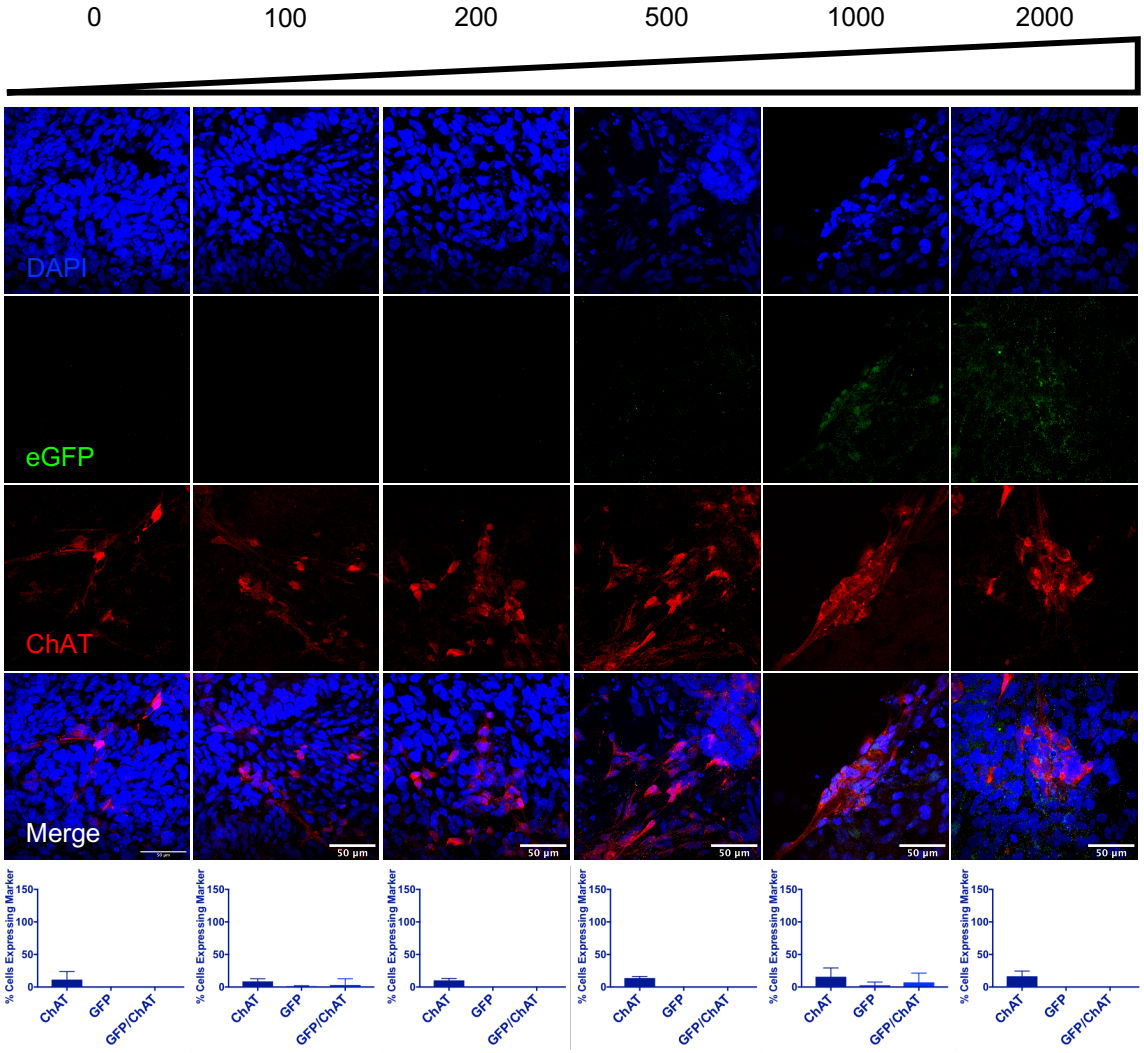
Figure 4.3. Immunofluorescence confocal images of iPSC-motor neurons treated with various concentrations of RChAT:TRAP lentivirus to optimise transduction.

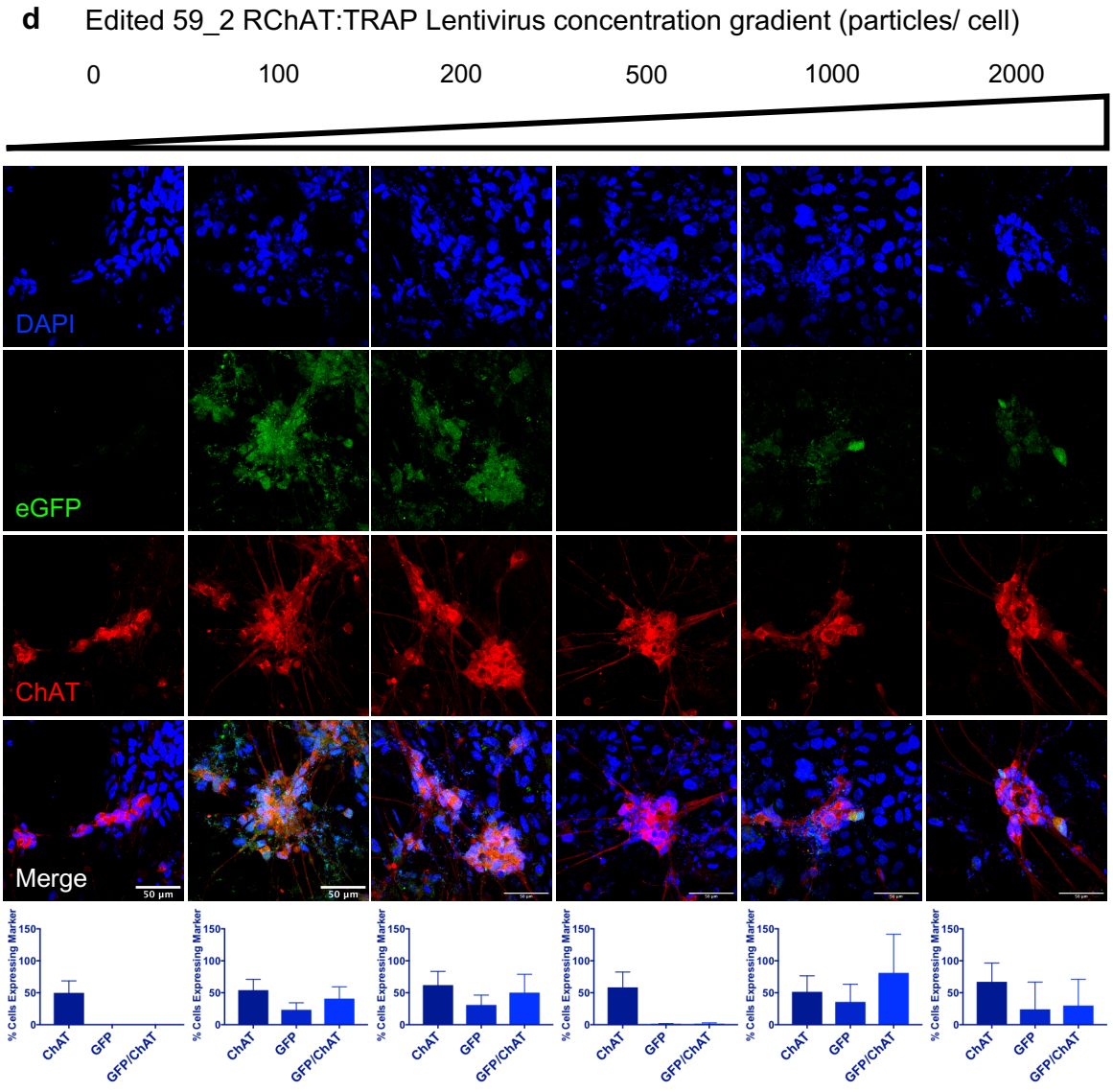
a SFC840 Control RChAT:TRAP Lentivirus concentration gradient (particles/ cell)



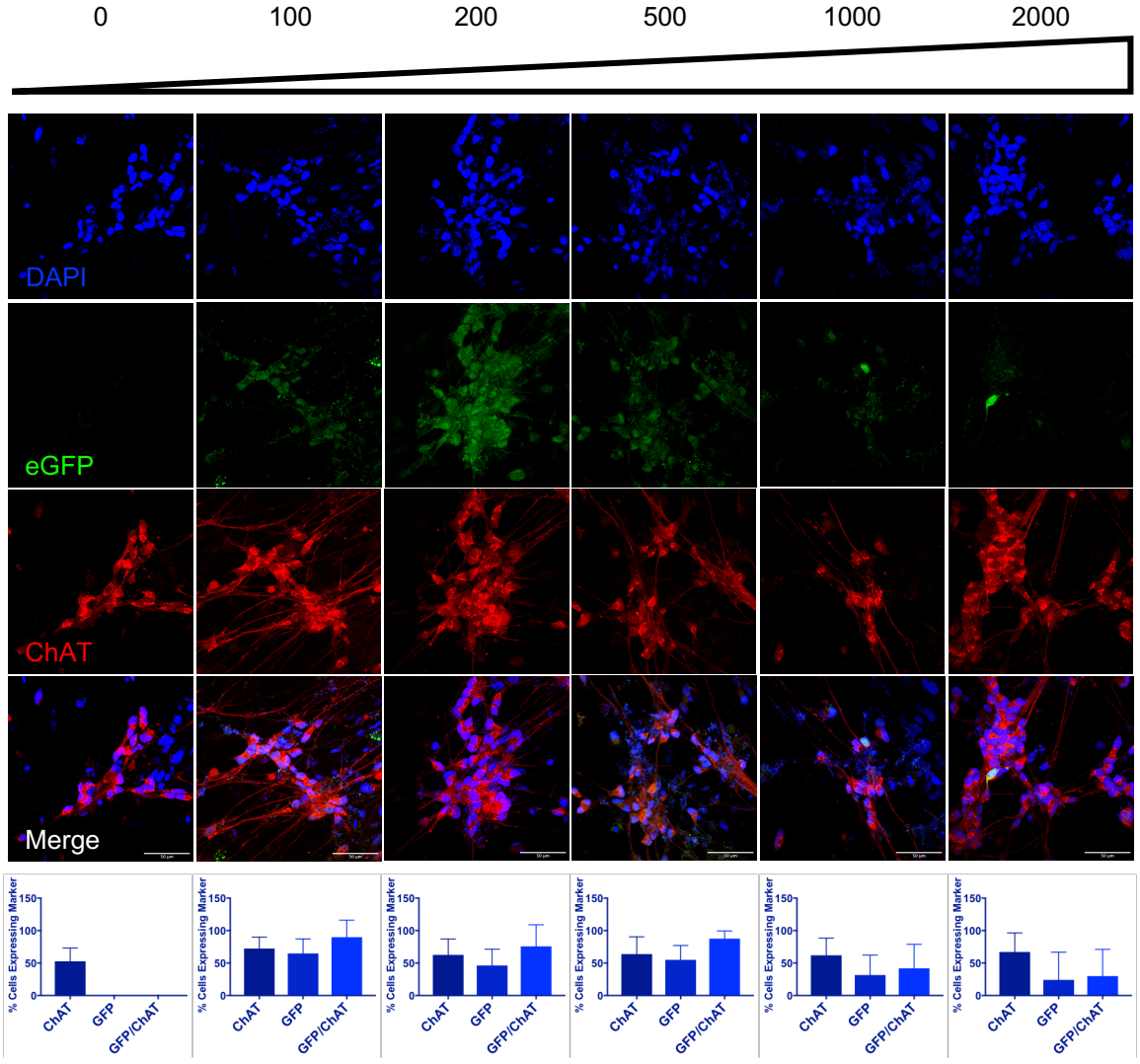

b
Patient RChAT:TRAP Lentivirus concentration gradient (particles/ cell)
0
100
200
500
1000
2000
DAPI
eGFP
ChAT
Merge
% Cells Expressing Marker
ChAT
GFP
GFP/ChAT

c Edited 59_1 RChAT:TRAP Lentivirus concentration gradient (particles/ cell)




d
Edited 59_2 RChAT:TRAP Lentivirus concentration gradient (particles/ cell)
0
100
200
500
1000
2000
DAPI
eGFP
ChAT
Merge
50 µm
% Cells Expressing Marker
ChAT
GFP
GFP/ChAT

e Edited 59_3 RChAT:TRAP Lentivirus concentration gradient (particles/ cell)



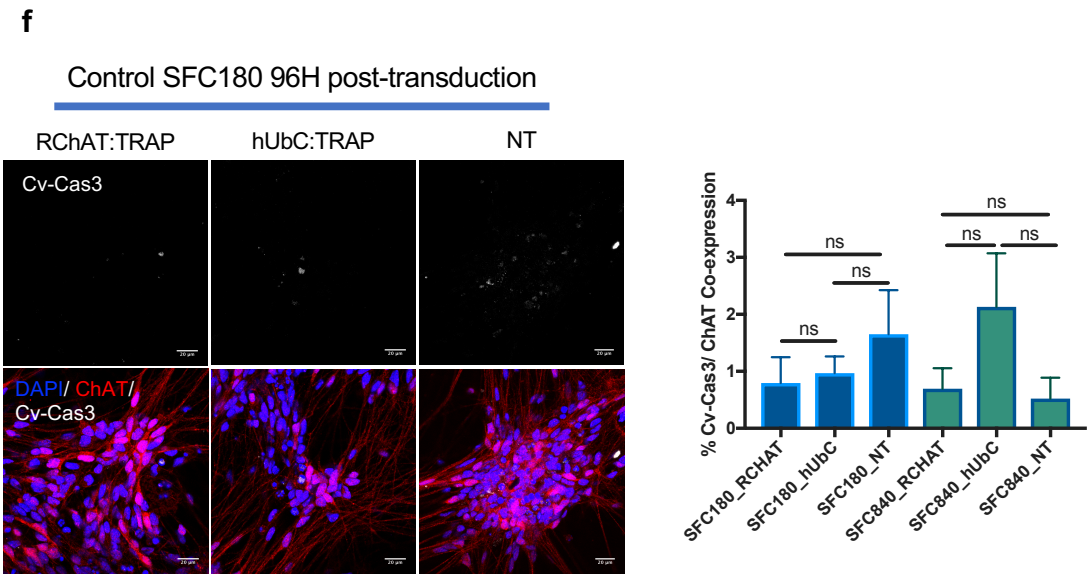


Figure 4.3. Immunofluorescence confocal images of iPSC-motor neurons treated with various concentrations of RChAT:TRAP lentivirus. This was done in order to optimise the conditions to obtain cultures co-expressing ChAT and RChAT:TRAP vector eGFP reporter. Representative images give cultures treated for 4 hours (day 24) with lentivirus and harvested 96 hours later (day 28) for analysis.

(a) Age-matched control treated with lentivirus concentrations up to 3000 particles/ cell. Concentrations up to 10,000 particles/ cell were tried, though no surviving cultures were found above the highest given here. Biological replicate n = 3. Scale bar = 50 µm
(b) C9orf72 patient-derived line, using two clones from the same patient (C90202 and C90203). Biological replicate n = 2. Scale bar = 50 µm
(c) Edited line, 59_1. Biological replicate n = 1. Scale bar = 50 µm
(d) Edited line, 59_2. Biological replicate n = 1. Scale bar = 50 µm
(e) Edited line, 59_3. Biological replicate n = 1. Scale bar = 50 µm
(f) Cleaved caspase-3 staining and quantification of age-matched control iPSC-motor neurons treated with RChAT:TRAP and hUbC:TRAP lentivirus. Biological replicate n = 1. Scale bar = 20 µm
(g) For every biological replicate, at least 500 cells were counted, in >5 fields across 1-2 coverslips (technical replicates).

Lentiviral transductions were found to be less well tolerated by the edited lines. A gradient of RChAT:TRAP lentivirus treatments for the edited isogenic lines are given (Figure 4.3c-e). Quantification of motor neuron markers showed that ChAT-positive cells were generally fewer in these lines, as seen in the non-transduced (NT) controls. Furthermore, optimal eGFP expression could be seen at transduction concentrations as little as 100 particles/ cell (Figure 4.3d). Between 100-200 particles/ cell were used to transduce edited iPSC-motor neuron cultures, which were subsequently used to pull down RNA via TRAP. Further discussion on RNA purification from these cultures is given later.

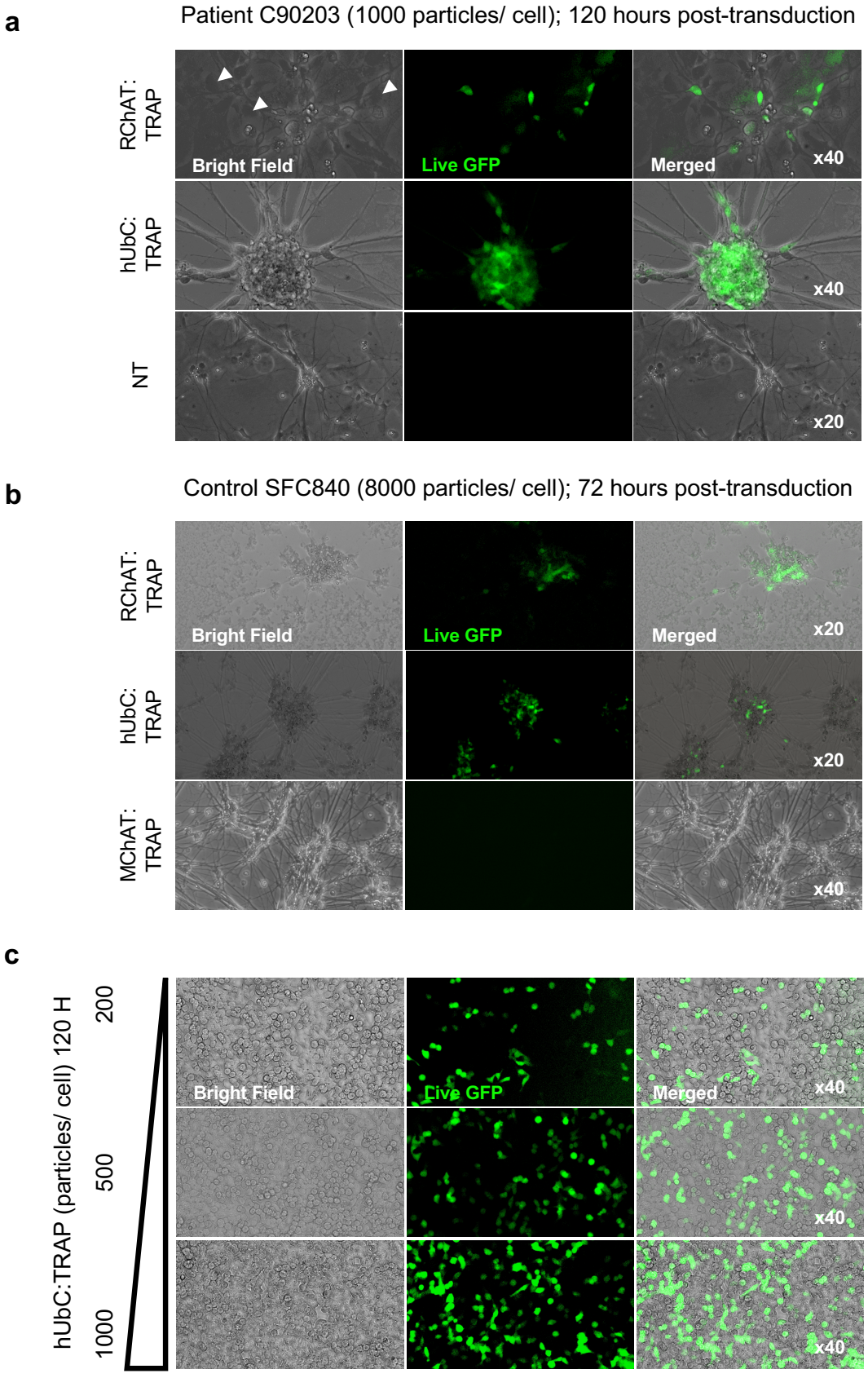
4.2.2 Optimisation of transduction conditions resulted in minimal cell death

Polybrene reagent is commonly used while transducing neuronal cultures with lentivirus due to their limited propensity to bind the exogenous protein sheaths and take up the packaged RNA. However, this was omitted from the final protocol as control non-transduced cultures, enriched only with polybrene had increased cell death compared with those also containing lentivirus. An explanation for this could not be found. In addition, to minimise cell death, cell culture media was enriched with ROCKi during lentivirus transduction, and for at least 24 hours after.

The optimal days for lentiviral transduction and harvest were carefully considered to maximise motor neuron yield, TRAP vector transduction efficiency, motor neuron culture maturity and reproducibility. Lentiviral transduction results in permanent integration of the packaged construct into the host cell's genome via the LTR regions, thus consideration was given to performing transduction at a stage of the differentiation when cells are likely to be more likely to take up lentiviral constructs,

for example at the iPSC stage, prior to introducing any growth factors for differentiation. This method was ultimately unjustified as the best course for optimisation of transduction, as the percentage of transduced cells expressed the TRAP vector could not be validated until the cultures matured, more than three weeks following transduction. Preliminary experiments suggested that the overall quality of the cultures was reduced with such early transductions, with an increased proportion of non-neuronal cells at early time points in the differentiation, compared with non-transduced cultures. Quantifiable evidence of this growth cannot be shown as no molecular marker for these cells is known. Non-neuronal cells are generally identified by their morphology (Figure 4.4a). The possibility of performing multiple transductions on the same cells throughout the differentiation was excluded because of similar observations.

Figure 4.4. Live fluorescence imaging of iPSC-motor neurons and HEK293-T cells following hUbC:TRAP or RChAT:TRAP lentivirus transduction.



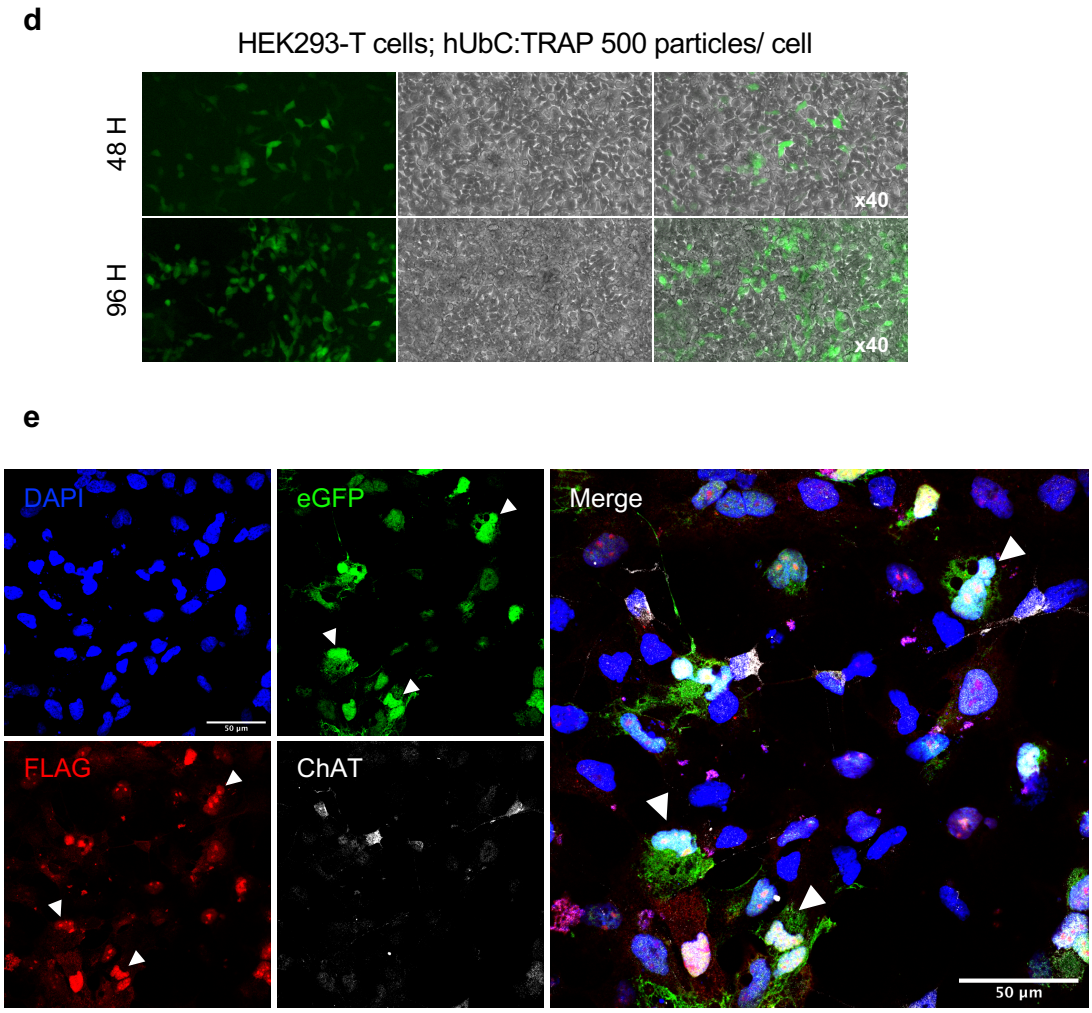


Figure 4.4. Differences in eGFP reporter expression following hUbC:TRAP or RChAT:TRAP lentivirus transduction can be visualised via live fluorescence imaging of iPSC-motor neurons and HEK293-T cells. Regions of increased lentiviral-induced toxicity results in the proliferation of non-neuronal cells. Relative magnification of eGFP live images are given in each case.

(a) Optimised transduction of RChAT:TRAP-packaged lentivirus (1000 particles/ cell) into patient iPSC-motor neurons is challenging to visualise using live fluorescence microscopy, though TRAP vector under the ubiquitous hUbC promoter gives a stronger signal, 120 hours post-transduction. Non-neuronal cells are highlighted here with white arrows.

(b) Higher concentrations of lentiviral transduction (8000 particles/ cell) into control iPSC-motor neurons gives a higher reporter signal, though significantly increased cell death, as can be seen by cellular debris. Toxicity and transduction efficiency may vary depending on the lentivirus used as seen by treatment with hUbC:TRAP and MChAT:TRAP lentivirus under identical conditions.

(c) HEK293-T cells can be transduced with hUbC:TRAP lentivirus at relatively low concentrations giving good transduction efficiency, and low visible toxicity from higher concentrations up to 120 H post-transduction, until over-confluence of the cell culture dish.
(d) HEK293-T cells transduced at 500 particles/ cell, harvested 96 H post-transduction, were used in TRAP pulldown protocol optimisation experiments.
(e) Immunostaining of fixed day 25 control SFC840 iPSC-motor neurons, transduced with hUbC:TRAP lentivirus (200 particles/ cell; 72 H) shows that regions of increased neuronal death have cells of unknown identity. These cells increase proliferation as differentiations mature, and motor neurons exit cell cycle and stop dividing. Scale bar = 50 µm.

As previously mentioned, this iPSC-motor neuron differentiation protocol can robustly produce cultures between 60-80% positive for specific markers, including ChAT, beta-III tubulin and Islet 1. However, a successful lentiviral transduction is dependent on a number of factors, which increases the challenge of consistently obtaining a high yield of ChAT-positive motor neurons expressing the TRAP vector transgenes. Fully differentiated motor neurons have a limited tropism for 3rd generation lentiviral particles, due to the down regulation of the glycoproteins necessary to interact with VSV-G envelope protein, making up the sheath of the lentivirus particles, enabling delivery of a packaged plasmid into a host cell. Also, as the motor neurons mature, their cell bodies cluster into bundles, with axons extending to other bundles of cell bodies. This tight conformation limits available surface area, and thus significantly decreases transduction efficiency. Therefore, transduction of the cultures needs to be done while the cells are still spread out. An increase in non-neuronal cell proliferation rate after lentiviral transduction was also observed. As the motor neurons mature and exit the cell cycle, competition with non-neuronal cells for growth factors and nutrients increases. This increases the rate of motor neuron cell death and causes them to lift from cell culture dishes. To optimise the balance between expression of TRAP vector, with enough time for transgenes to express and tagged-RPL22 to enter the translation pathway, and to prevent increased mitotic cell growth, cultures were transduced on day 24 and harvested after 96 hours. In this group's motor neuron differentiation protocol, final passaging of the cultures and introduction of DAPT, BDNF and GDNF is done between day 18-20. Following this step, the cells are passaged harshly to allow space for axonal growth and to form cell body bundles. Generally, by day 26-30 these bundles are too tightly packed for lentiviral transduction to be efficient (Figure

2.1). Transductions were performed on day 24 when the neurons are still wide spread enough to allow for the best transduction (Figure 2.1).

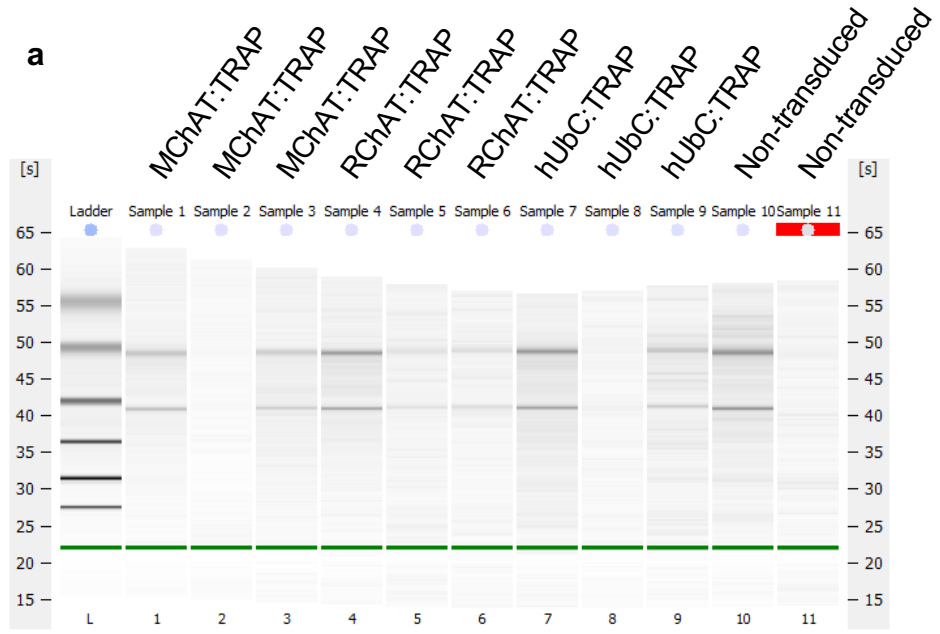
Due to the lengthy time required to complete the differentiation protocol of motor neurons from the iPSC stage (30 days), the HEK293-T cell line was used to optimise much of the TRAP protocol. The high turnover of results from an immortalised cell line were essential to validate the many steps of the protocol, particularly in terms of RNA quality control. Live eGFP fluorescence was reliably observed in HEK293-T cells, transduced with the mammalian TRAP vector from 48 hours onwards, and up to 8 days post-transduction (Figure 4.4c-d).

4.3 Optimisation of translating ribosome affinity purification from transduced cell cultures

The TRAP protocol has been widely applied and published in the context of mouse models, with only one study to date using primary rat cortical neurons (Lesiak et al., 2015). The protocol from the latter was initially used to obtain RNA via tagged-ribosome pulldowns. RIN analysis of these samples (Figure 4.5) showed a high degree of RNA degradation, with RIN values ranging from 1 to 8. Measurement of sample concentrations of pulldowns from approximately two million cells gave yields around 100 pg/ µL, which is too low for bulk RNA sequencing, without introducing an additional PCR amplification step during library preparation. RIN analysis and concentration measurement of negative control samples, non-transduced motor neurons, in the same pulldown experiment showed similar results (Figure 4.5). This suggested that the pulldown from transduced samples were likely

to have been non-specific background from magnetic beads, in addition to increased degradation, potentially from excess overnight incubation or poor handling of samples (Lesiak et al., 2015). Consequently, a number of steps of the TRAP protocol needed to be optimised in order to improve the overall yield and RNA quality. These steps have been described in detail below.

Figure 4.5. RNA integrity number (RIN) analysis of iPSC-motor neuron RNA samples pulled down using TRAP protocol from Lesiak et al., 2015.



b

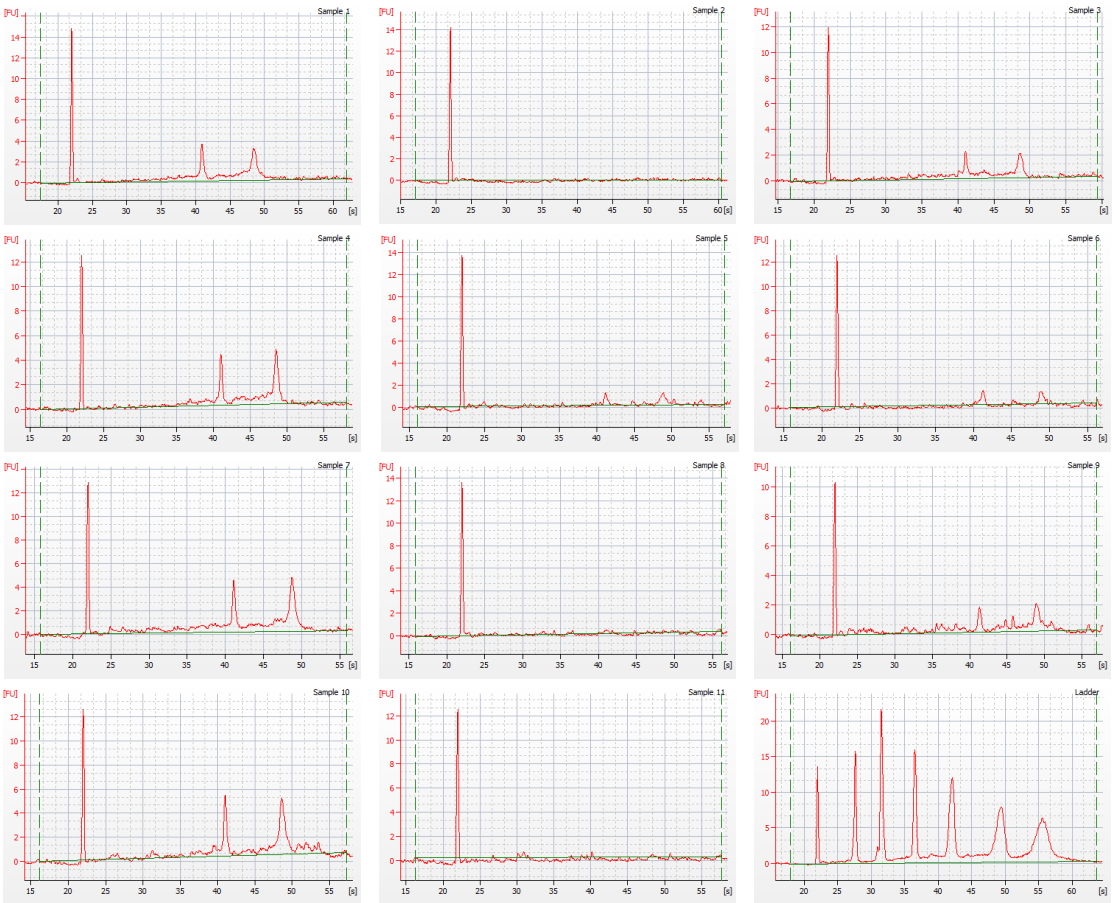


Figure 4.5. RNA integrity number (RIN) analysis of iPSC-motor neuron RNA samples pulled down using TRAP protocol from Lesiak et al., 2015.

(a) The Agilent bioanalyzer kit was used to assess the separation of RNA fragments in each sample using electrophoretic separation followed by laser fluorescence for band visualisation.

(b) Electropherograms, generated by the Bioanalyzer software, are given for each sample. Low yields and severe RNA degradation can be seen in samples obtained using the original protocol, including in non-transduced negative controls.

4.3.1 Harvesting and handling techniques dramatically affect quality and yield of RNA

Cell cultures were incubated with growth media enriched with cycloheximide to inhibit translation and prevent the degradation of polysomal complexes during harvesting of iPSC-motor neuron lysates. Incubations of up to 1 hour could be tolerated by the cultures without inducing visible cell death. All buffers (adapted from Lesiak et al., 2015) used throughout the TRAP protocol were supplemented with cycloheximide to maintain ribosome-mRNA complexes. It was found that this step is crucial to obtaining decent yields of TRAP-RNA. As previously recommended in most published TRAP protocols (Heiman et al., 2014; Lesiak et al., 2015), all buffers were freshly supplemented with RNase/ DNase-inhibitors and protease inhibitors to reduce degradation of ribosomes, mRNA and RPL22:c-myc:FLAG protein.

The nature of TRAP requires lysate from tissue or cells to remain in suspension for extended periods of time, exposed to endogenous ribonucleases and potential contaminants from the environment. To prevent RNA degradation from these sources, the samples must be handled very carefully from start to finish, in addition to avoiding disruption of the precarious polysome complexes. Following translation inhibition, cells were harvested by scraping from cell culture dishes and transferred immediately to pre-chilled RNase/ DNase-free 2 mL centrifuge tubes. These cell suspensions were lysed by gentle pipetting and incubating on ice to allow chemical degradation of cell matter by the detergents enriched in the homogenization buffer. This was in contrast to the recommended step (Lesiak et al., 2015) of lysing the cells via mechanical perturbation with motorised pestles. It was found that any

harsh treatment of suspensions containing tagged-ribosome-mRNA complexes could lead to decreased yields of RNA. Tubes containing samples were always kept on ice or at 4 °C to prevent heat-induced RNA degradation. The use of pre-chilled tubes and buffers at all stages of the protocol is recommended. If possible, working in a 4 °C environment (cold room) is suggested if frequently handling the samples for transferring and washing, and especially if working with several samples at once. Following lysis, samples were centrifuged at 12,000 g for 10 minutes at 4 °C to pellet out dead cell matter. The supernatants were transferred to fresh tubes, ready for ribosomal pulldown. It should be noted that samples with a large number of starting cells may find it beneficial to include an extra centrifugation step of 2000 g to remove the largest cell matter, prior to the 12,000 g spin to avoid dislodging the pellets and contaminating the sample. This is usually done by eye.

4.3.2 Varied TRAP vector transduction efficiency rate results in different yields of RNA

The pulldown of ribosomal complexes via RPL22:c-myc:FLAG to obtain mRNA requires optimisation of the conjugating antibody, incubation with magnetic beads, as well as incubation times, washing steps and handling of the samples. The original TRAP protocol (Heiman et al., 2014) and subsequent uses of the technique have employed overnight incubation of the tissue or cell lysate with antibody-conjugated beads. The Lesiak et al., 2015 protocol incubates antibody with the cell lysate for 4 hours at 4 °C prior to overnight incubation with magnetic beads, followed by washing with a high salt buffer to reduce non-specific interactions. Repeated attempts using this method did not precipitate good quality RNA, or a high yield

(Figure 4.6). Subsequently, a number of parameters were optimised as described below.

Figure 4.6. Optimisation of washing buffer of beads interacting with antibody-ribosomal complexes in HEK293-T cells.

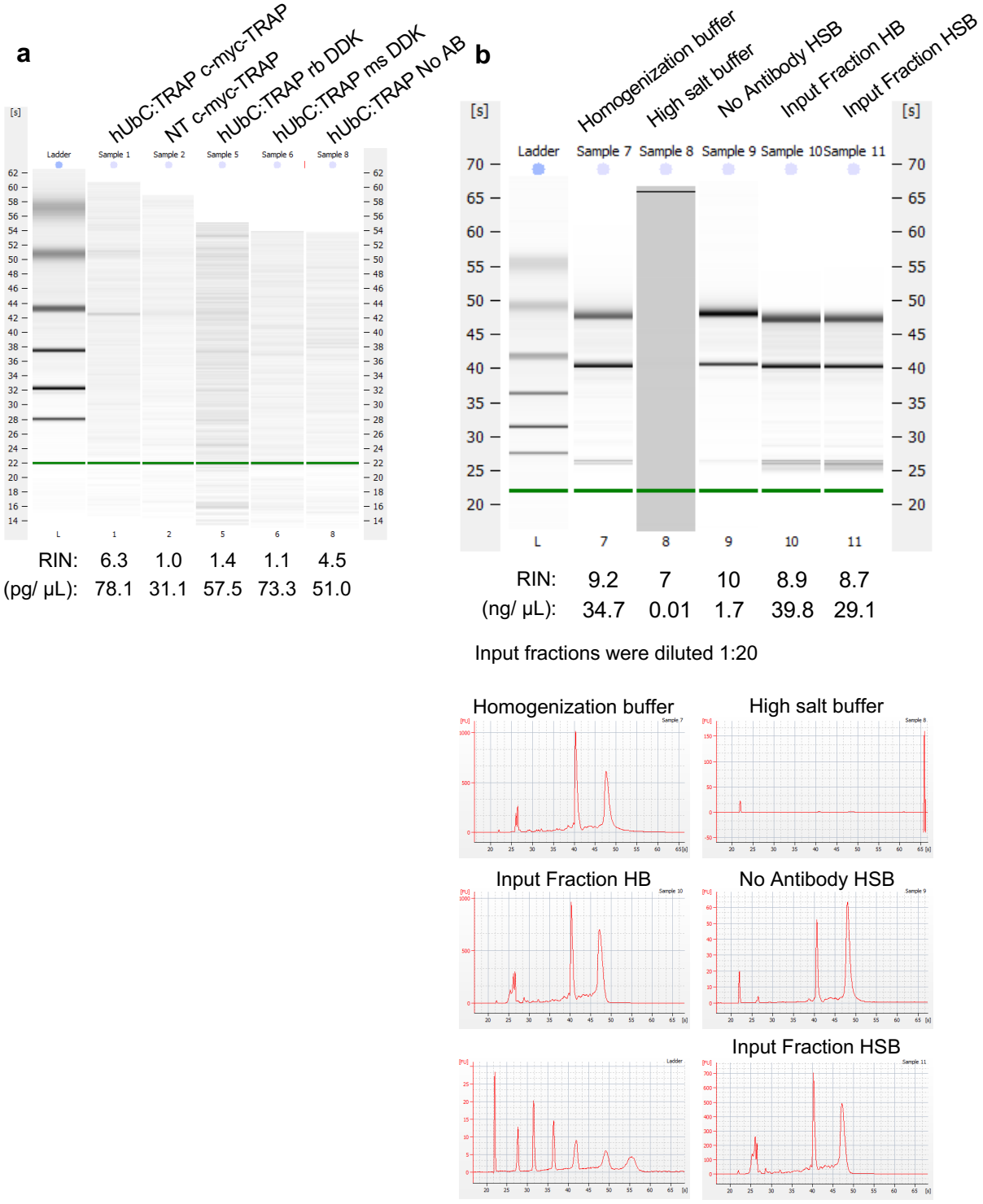


Figure 4.6. Optimisation of washing buffer of beads interacting with antibody-ribosomal complexes in HEK293-T cells.

(a) Repeated RNA pulldown experiments following the TRAP protocol from Lesiak et al., 2015, using different antibodies including anti-c-myc, pre-conjugated to magnetic beads (c-myc-TRAP), and polyclonal and monoclonal antibodies against FLAG antigen TRAP protocol was performed on HEK293-T cells transduced with hUbC:TRAP lentivirus. Neither high quality or yield of RNA could be obtained, as seen from RIN analysis. NT = non-transduced, AB = antibody.

(b) Validation of the pulldown of any specific or non-specific RNA was done by omitting the use of high salt wash buffer (HSB) (300 mM KCl) to determine whether any RNA, could be obtained from lysates. Beads were washed with non-supplemented homogenization buffer (HB), and samples validated via RIN analysis. This TRAP experiment used mouse c-myc antibody at 1:200 dilution.

4.3.2.1 Modification of washing steps lead to purification of high quality RNA

Repeated use of the Lesiak et al., 2015 protocol, using various antibodies against FLAG and c-myc tags, including commercial magnetic beads pre-conjugated to c-myc antibody (c-myc-TRAP) was done, but failed to produced good results (Figure 4.6a). In order to investigate whether any material, specific or non-specific could be obtained from the antibodies or Protein G magnetic beads, a combination of reducing incubation times (section 4.3.2.3) and washing beads with non-supplemented homogenization buffer was done to investigate this (Figure 4.6b). Following TRAP from human cell cultures, prior to performing RIN analysis, the concentration of each sample was measured using spectrophotometry using a Nanodrop (not given here). It was found that washing with high salt buffer only precipitated ~10 pg/ µL of RNA, showing less RNA pulldown than a negative control omitting antibody from the cell lysate (Figure 4.6b). On the other hand, washing with homogenization buffer and RNA dissociation pulled down 34.7 ng/ µL of RNA (Figure 4.6b). The equivalent negative control without antibody incubation with the lysate, showed relatively low non-specific binding at 4.3 ng/ µL, comparable with that of the high salt buffer negative control, which measured at 3.7 ng/ µL using spectrophotometry. This protocol also showed a marked increase in RIN quality with that of the Lesiak et al., 2015 protocol, suggesting that reduced incubation times reduced RNA degradation (described in section 4.3.2.3). Minimal background RNA was seen from pulldowns using homogenization buffer to wash beads, however the amount of this background was highly variable across experiments and was increased in TRAP RNA from iPSC-motor neurons. The enrichment of motor neuron-specific pools of RNA, compared with non-specific RNA, was further validated following antibody optimisation (section 4.3.2.2).

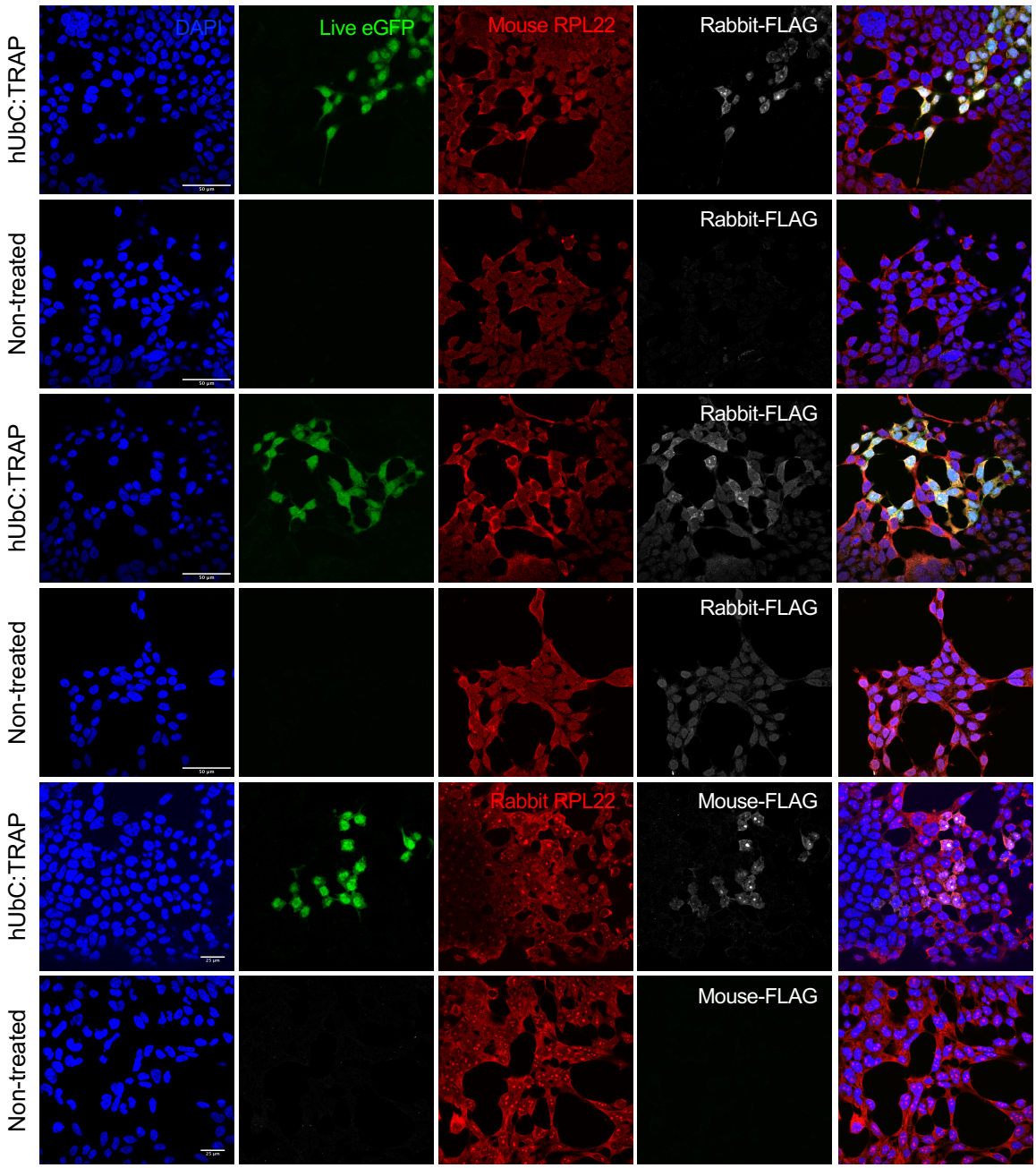
4.3.2.2 Monoclonal c-myc antibody immunoprecipitated the highest yields of ChAT-enriched RNA pools

The TRAP construct encodes a large ribosomal protein, RPL22, fused to c-myc and FLAG (FLAG) tags. Previous TRAP systems have often used GFP as a tag (Heiman et al., 2008; Tryon et al., 2013; MacNair et al., 2015) and reporter for successful TRAP vector expression. The potential cellular mislocalisation of tagged-RPL22 and its ability to integrate into early large ribosome complexes was a concern, so a TRAP construct with a smaller fused tag was designed. The two tags, c-myc and FLAG also increase the potential antibodies to test for optimal pulldown. To reduce immunoprecipitation of non-specific binding partners, visual analysis of the background seen from immunostaining using a variety of commercial antibodies against c-myc or FLAG/ FLAG were tested in HEK293-T cells transduced with hUbC:TRAP lentivirus (Figure 4.7a). Specific staining was examined in cells co-expressed live eGFP reporter. As expected, the monoclonal mouse antibodies showed far less background staining than the polyclonal rabbit antibodies (Figure 4.7a). All antibodies were tested in a TRAP pulldown experiment using hUbC:TRAP-transduced HEK293-T cells. Significantly less RNA was obtained from anti-FLAG rabbit antibody compared with mouse anti-FLAG and anti-c-myc antibodies (Figure 4.7b). Notably, mouse c-myc antibody gave the greatest yield from the same number of cells compared with anti-FLAG antibodies (Figure 4.7b). As previously mentioned, background RNA was often measured from pulldowns using these antibodies with non-transduced HEK293-T cell lysates. To elucidate whether this higher yield was due to additional background or truly being enriched with actively translating mRNA, qRT-PCR was performed to quantify RPL22:c-myc:FLAG transcripts (Figure 4.7c-d). TRAP samples using mouse FLAG and c-

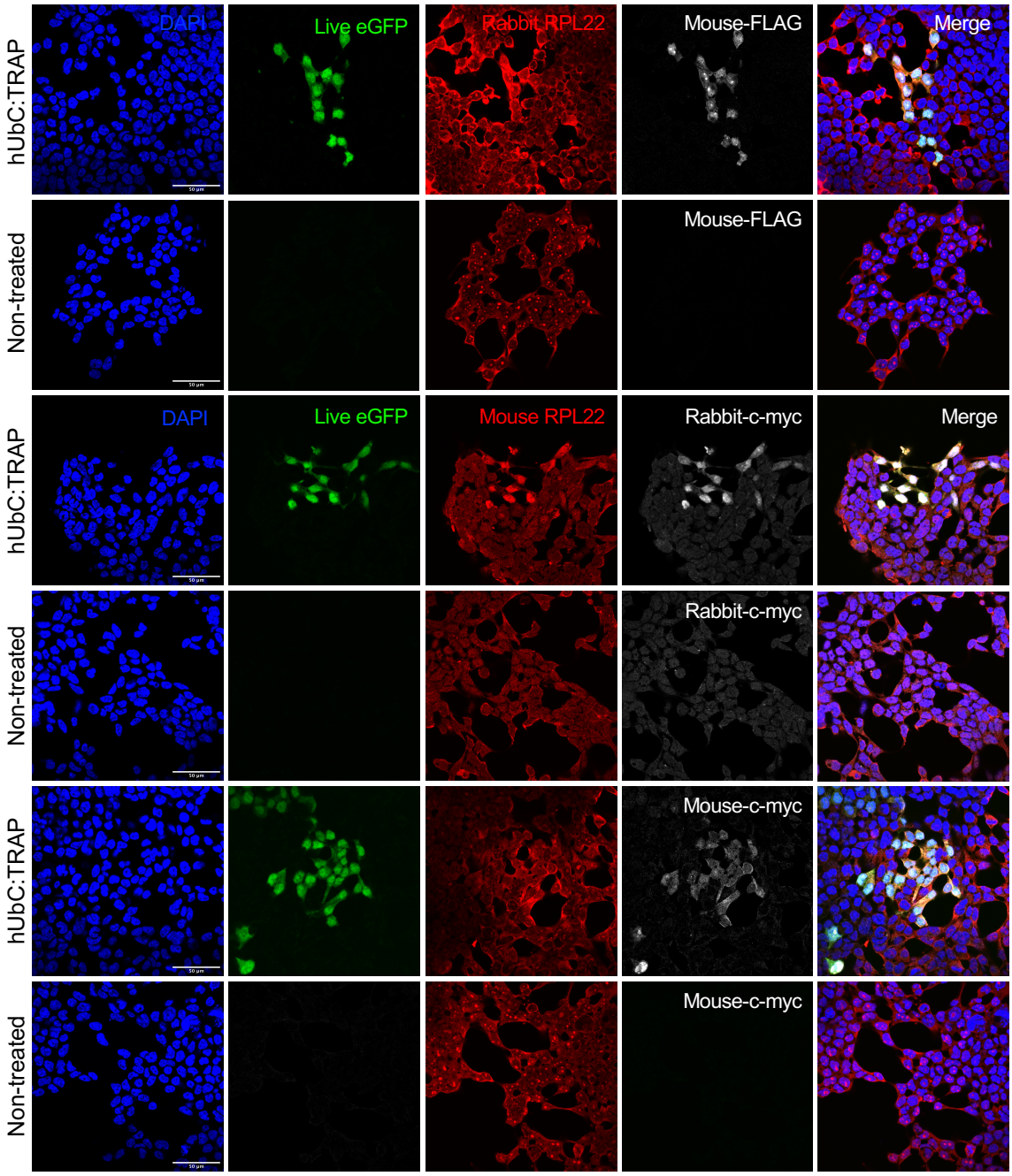
myc antibodies were both significantly enriched with the TRAP transcript compared with non-specific RNA from non-transduced cells (Figure 4.7c). However, only TRAP experiments using monoclonal mouse antibody consistently gave high quality RNA that was enriched compared with its corresponding transcriptome input fraction (Figure 4.7d). Together with the observation that this antibody was found to bind to a target in hUbC:TRAP-transduced HEK293-T cells that overlaps with cells also expressing live eGFP reporter (Figure 4.7a), and high yields of RPL22:c-myc:FLAG-enriched pools of RNA (Figure 4.7b-d), mouse anti-c-myc (Sigma-Aldrich) was chosen as the optimised antibody for the final TRAP protocol.

Figure 4.7. Antibody optimisation against molecular tag used for ribosomal pulldown in TRAP protocol.

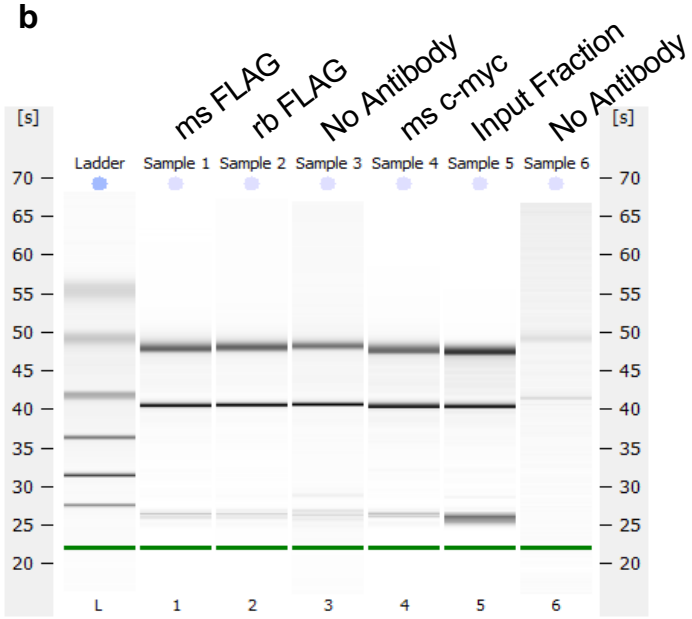
a



a continued


hUbC:TRAP
DAPI
Live eGFP
Rabbit RPL22
Mouse-FLAG
Merge
Non-treated
Mouse-FLAG
hUbC:TRAP
DAPI
Live eGFP
Mouse RPL22
Rabbit-c-myc
Merge
Non-treated
Rabbit-c-myc
hUbC:TRAP
Mouse-c-myc
Non-treated
Mouse-c-myc

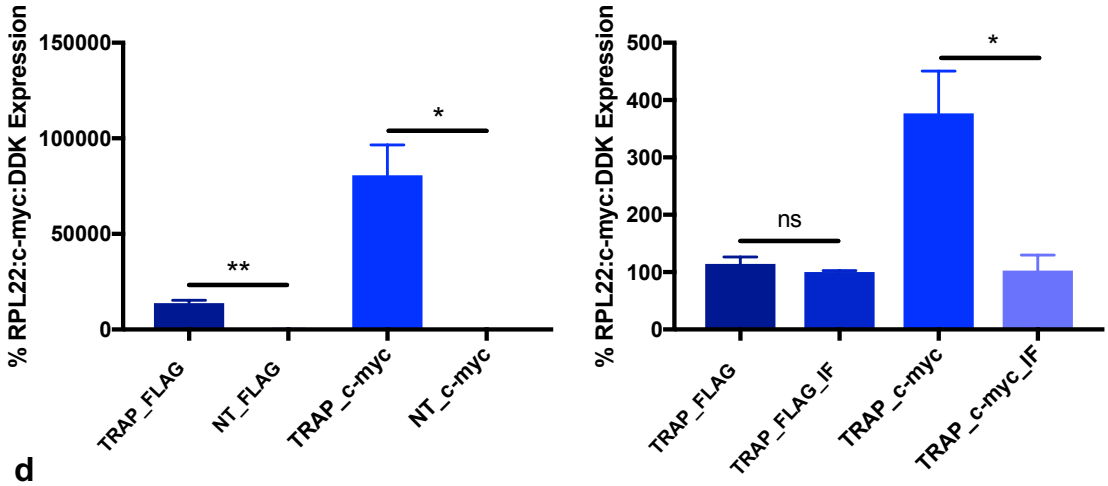
b



RIN:	9.6	9.8	10	9.4	9.1	10
(ng/ µL):	18.7	4.1	0.7	40.3	27.1	0.00

*Input fractions were diluted 1:20

c



d

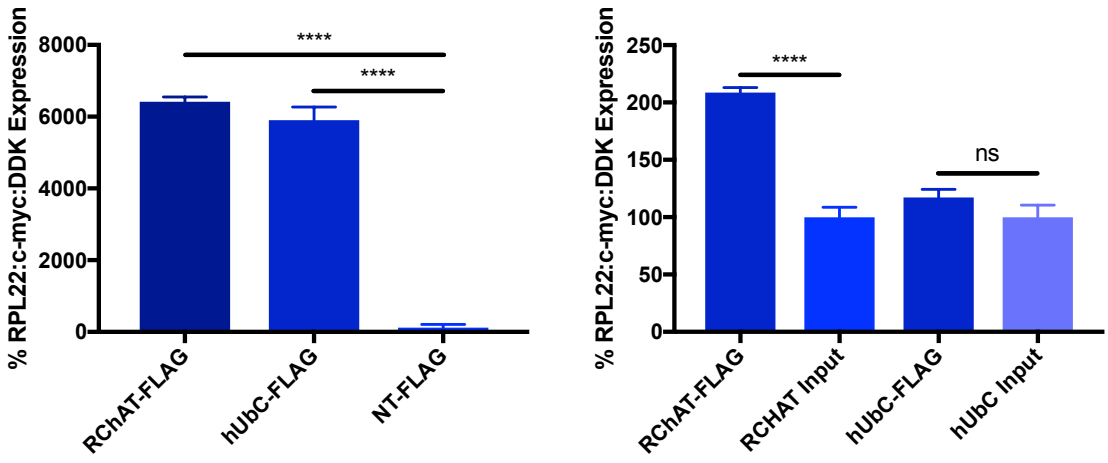


Figure 4.7. Optimisation of antibody against molecular tag used for ribosomal pulldown in TRAP protocol, using HEK293-T cells, transduced with hUbC:TRAP lentivirus, and patient C90203 iPSC-motor neurons.

(a) Immunostaining of fixed HEK293-T cells to assess the differences in non-specific background protein binding of various antibodies. Commercially available polyclonal and monoclonal antibodies, against FLAG/ FLAG and c-myc were tested.

(b) Antibodies with the least visible non-specific staining were used in TRAP of RNA from HEK293-T cell cultures transduced with hUbC:TRAP lentivirus. RIN analysis shows that decent yields are produced compared with samples where antibody has not been added, showing that minimal background RNA is being pulled down.

(c) Quantitative RT-PCR was performed to examine the levels of RPL22:c-myc:FLAG transcript in the TRAP-RNA pools from HEK293-T cells transduced with hUbC:TRAP, compared with non-specific RNA from non-transduced (NT) controls or transcriptome (input, IF) fractions. Biological replicate n = 1, technical replicates n = 3; unpaired students t-test, SD < 0.05* or 0.005**.

(d) Quantitative RT-PCR was performed to examine the levels of RPL22:c-myc:FLAG transcript in the TRAP-RNA pools from iPSC-motor neurons transduced with RChAT:TRAP and hUbC:TRAP, compared with non-specific RNA from non-transduced (NT) controls or transcriptome (input, IF) fractions. Biological replicate n = 1, technical replicates n = 3; unpaired students t-test, SD < 0.0005****.

4.3.2.3 Incubation times were optimised to obtain high quality RNA with the least background

The protocol in Lesiak et al., 2015 performed pulldowns by incubating anti-tag antibody with harvested cell lysate, followed by incubation with Protein G magnetic beads. This method was investigated, as well as the conjugation of magnetic beads with antibodies prior to incubation with cell lysate. Commercially bought magnetic beads readily conjugated to c-myc antibodies were also used (Figure 4.7b). The latter was done in order to reduce the time of cell lysate exposure. The former method was ultimately employed, with which the best RNA purification yields were achieved. Incubation times of antibody with cell lysate, from 30 minutes to four hours at 4 °C was attempted (not shown). Differences in pulldown between two and four hours were not seen, thus the former was adapted for the final protocol. Incubation of lysate and antibody with magnetic beads was tested at 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours and overnight (not shown). Reasonable RNA pulldown was seen at 1 hour, and used in the final protocol. The volume of magnetic beads used per sample was further optimised. Lesiak et al., 2015 suggests using 200 µL of Protein G beads per 450,000 cells. This ratio was initially employed, using approximately 2 million cells/ sample, from iPSC-motor neuron cultures. It was found that increased amounts of non-specific RNA pulled down using this method. Commercial Protein G beads are designed to preferentially bind antibodies and are suggested by the manufacturer (ThermoFisher) to be highly efficient at binding all IgG interactors in a sample in 10 to 40 minutes, using 20-100 µL of beads. Examining the overall yield of motor neurons from iPSC differentiations by eye, the magnetic bead volume was adjusted between 20-50 µL. For HEK293-T optimisations, 40 µL of beads were used with ~8.8 million cells input. RNA yields

from iPSC-motor neurons were increased by pooling ~12 million cells per sample (one 6-well plate).

A number of optimisation steps were taken in the immunoprecipitation of tagged ribosomes from human cells transduced with lentiviral TRAP vector. Incubation time of antibody with harvested cellular lysate was optimised to 2 hours at 4 °C. Along with antibody species (Figure 4.7a), the incubation time of the antibody-lysate suspension with Protein G magnetic beads was optimised to 1 hour at 4 °C. Bead suspension volumes were altered to prevent non-specific binding based on observations of the number of iPSC-motor neurons between differentiations and cell lines. The beads were found to have very highly specific immunoprecipitated RNA, such that overnight incubation with beads and high salt buffer were not required to dissociate non-specific interactions (Figure 4.6b; Figure 4.7c-d). Collectively these alternations to the Lesiak et al., 2015 protocol were done to obtain the best possible pulldown of actively translating mRNA from human iPSC-motor neurons.

4.3.3 Dissociation and purification of mRNA from magnetic beads was optimised for maximum output

A brief comparison of the reagent to dissociate mRNA bound to immunoprecipitated ribosomes was done. Lesiak et al., 2015 suggests supplementing the RNA lysis buffer from a standard RNA purification kit with 1% betamercaptoethanol. This was compared with phenol reagent (TRIzol) extraction and found that the latter to be more effective at overall RNA extraction. Thus, TRIzol was mixed with the magnetic beads after washing steps, followed by ethanol RNA precipitation. A compatible

RNA column extraction kit was used to load the eluate and on-column DNase digestion was performed to remove any potential contaminants. Manufacturer's instructions were followed, and RNA was eluted in RNase/ DNase-free water.

Collectively these alterations to published TRAP protocols have been adapted to human-derived iPSC-motor neurons (fully detailed in chapter 2, section 2.2.5), and final RNA extractions for an RNA sequencing experiment could be carried out.

4.4 RNA sequencing of TRAP-RNA from *C9orf72* ALS patient iPSC-motor neurons

The original aim was to extract TRAP-RNA from iPSC-motor neurons, from *C9orf72* ALS patients, age-matched controls and isogenic CRISPR/Cas9-lines where the *C9orf72* repeat had been edited out. Multiple differentiations were carried out to achieve this (Table 4.1). Due to a variability in the cell lines which were successfully carried through to the end of the differentiation and the failure of several of the latter experiments to extract enough RNA to submit for sequencing, only five RNA samples from 2 clone lines of a patient over 3 differentiations, and 2 age-matched controls over 3 differentiations (Table 4.1), were obtained to submit for RNA sequencing. The edited lines especially posed a challenge due to the overall low quality and yield of differentiation cultures, despite optimising their lentiviral transductions (Figure 4.3c-e).

Table 4.1. All TRAP experiments performed to obtain TRAP and corresponding transcriptome input samples for RNA sequencing, using the final optimised protocol. IF = input fraction. Samples submitted for sequencing are in bold.

Sample Number	Line	Phenotype	Treatment	Date	Replicate	Differentiation	RIN	Conc (ng/ µL)
1	**SFC180**	**Control**	**RChAT**	**23_04_18**	**1**	**1**	**N/A**	**48.43**
2	**SFC180_IF**	**Control**	**RChAT**	**23_04_18**	**1**	**1**	**9**	**47.04**
3	**SFC840**	**Control**	**RChAT**	**23_04_18**	**1**	**1**	**N/A**	**71.1**
4	**SFC840_IF**	**Control**	**RChAT**	**23_04_18**	**1**	**1**	**7.7**	**43.61**
5	**SFC180**	**Control**	**RChAT**	**27_04_18**	**1**	**2**	**9.5**	**5.39**
6	**SFC180_IF**	**Control**	**RChAT**	**27_04_18**	**1**	**2**	**N/A**	**37.02**
7	SFC840	Control	RChAT	27_04_18	1	2	2.9	0.36
8	SFC840_IF	Control	RChAT	27_04_18	1	2	8.9	33.53
9	C902_02	Patient	RChAT	04_08_18	1	3	10	4.83
10	**C902_02**	**Patient**	**RChAT**	**04_08_18**	**2**	**3**	**9.7**	**8.35**
11	**C902_03**	**Patient**	**RChAT**	**04_08_18**	**1**	**3**	**9.8**	**10.03**
12	C902_03	Patient	RChAT	04_08_18	2	3	9.7	8.34
13	C902_03	Patient	RChAT	04_08_18	3	3	9.9	5.21
14	C902_02_IF	Patient	RChAT	04_08_18	1	3	10	189.35
15	**C902_02_IF**	**Patient**	**RChAT**	**04_08_18**	**2**	**3**	**10**	**147.65**
16	**C902_03_IF**	**Patient**	**RChAT**	**04_08_18**	**1**	**3**	**10**	**197.21**
17	C902_03_IF	Patient	RChAT	04_08_18	2	3	10	205.54
18	C902_03_IF	Patient	RChAT	04_08_18	3	3	10	285.38
19	C902_03	Patient	RChAT	06_08_18	1	3	9.6	12.11
20	**C902_03**	**Patient**	**RChAT**	**06_08_18**	**2**	**4**	**9.6**	**13.81**
21	C902_03	Patient	RChAT	06_08_18	3	4	9.7	10.51
22	**SFC840**	**Control**	**RChAT**	**06_08_18**	**1**	**4**	**9.7**	**11.27**
23	SFC840	Control	RChAT	06_08_18	2	4	10	6.91
24	**SFC840**	**Control**	**RChAT**	**06_08_18**	**3**	**4**	**9.4**	**7.64**
25	C902_03_IF	Patient	RChAT	06_08_18	1	4	10	120.10
26	**C902_03_IF**	**Patient**	**RChAT**	**06_08_18**	**2**	**4**	**10**	**191.14**
27	C902_03_IF	Patient	RChAT	06_08_18	3	4	10	136.62
28	**SFC840_IF**	**Control**	**RChAT**	**06_08_18**	**1**	**4**	**10**	**171.11**
29	SFC840_IF	Control	RChAT	06_08_18	2	4	10	162.11
30	**SFC840_IF**	**Control**	**RChAT**	**06_08_18**	**3**	**4**	**10**	**207.27**
31	59_2	CRISPR	RChAT	17_08_18	1	5	9.5	4.80
32	**C902_02**	**Patient**	**RChAT**	**17_08_18**	**1**	**5**	**9.8**	**6.16**
33	**C902_03**	**Patient**	**RChAT**	**17_08_18**	**1**	**5**	**9.7**	**6.75**
34	59_2_IF	CRISPR	RChAT	17_08_18	1	5	10	174.89
35	**C902_02_IF**	**Patient**	**RChAT**	**17_08_18**	**1**	**5**	**10**	**144.20**
36	**C902_03_IF**	**Patient**	**RChAT**	**17_08_18**	**1**	**5**	**10**	**201.59**
37	SFC840	Control	RChAT	18_08_18	1	6	10	0.05
38	59_1	CRISPR	RChAT	18_08_18	1	6	9.8	0.07
39	59_1	CRISPR	RChAT	18_08_18	2	6	1	0.04
40	59_2	CRISPR	RChAT	18_08_18	1	6	9	0.09
41	SFC840_IF	Control	RChAT	18_08_18	1	6	8.3	10.99
42	59_1_IF	CRISPR	RChAT	18_08_18	1	6	5	20.56
43	59_1_IF	CRISPR	RChAT	18_08_18	2	6	9.9	31.62
44	59_2_IF	CRISPR	RChAT	18_08_18	1	6	10	16.85

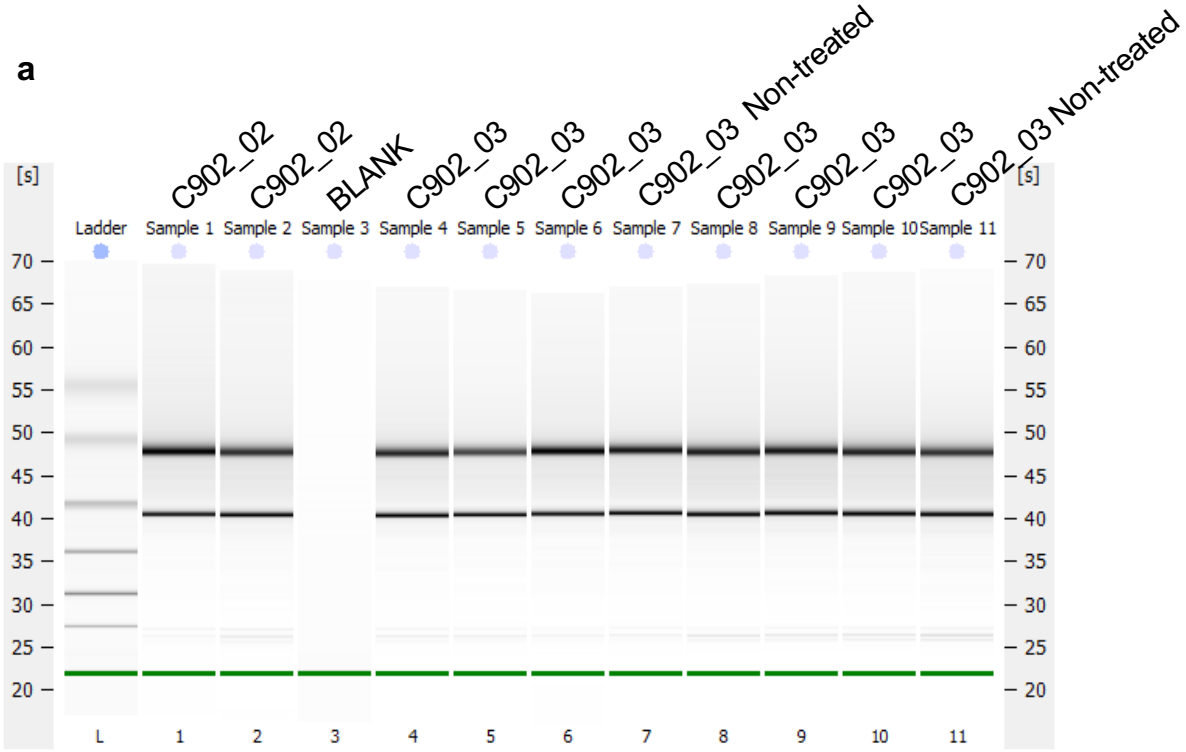
Table 4.1. continued.

Sample Number	Line	Phenotype	Treatment	Date	Replicate	Differentiation	RIN	Conc (ng/ µL)
45	SFC180	CRISPR	RChAT	19_08_18	1	7	8.6	0.08
46	SFC180_IF	CRISPR	RChAT	19_08_18	1	7	N/A	5.64
47	59_1	CRISPR	RChAT	12_10_18	1	8	7.4	0.05
48	59_1	CRISPR	RChAT	12_10_18	2	8	5	0.04
49	59_1	CRISPR	RChAT	12_10_18	3	8	1	0.10
50	59_1	CRISPR	RChAT	12_10_18	1	8	1.2	0.07
51	59_1_IF	CRISPR	RChAT	12_10_18	1	8	9.8	150.47
52	59_1_IF	CRISPR	RChAT	12_10_18	2	8	N/A	104.32
53	59_1_IF	CRISPR	RChAT	12_10_18	3	8	9.9	126.64
54	59_1_IF	CRISPR	RChAT	12_10_18	3	8	10	62.62
55	C902_02	Patient	RChAT	14_10_18	1	9	9.8	0.03
56	C902_02	Patient	RChAT	14_10_18	2	9	5.3	0.25
57	C902_02	Patient	RChAT	14_10_18	3	9	1	0.02
58	C902_02_IF	Patient	RChAT	14_10_18	1	9	N/A	166.52
59	C902_02_IF	Patient	RChAT	14_10_18	2	9	10	110.67
60	C902_02_IF	Patient	RChAT	14_10_18	3	9	1	97.58
61	SFC180	Control	RChAT	14_10_18	1	9	8	0.06
62	SFC180	Control	RChAT	14_10_18	2	9	9	0.06
63	SFC180_IF	Control	RChAT	14_10_18	1	9	N/A	125.44
64	SFC180_IF	Control	RChAT	14_10_18	2	9	10	164.28
65	59_1	CRISPR	RChAT	14_10_18	1	9	6.6	0.11
66	59_2	CRISPR	RChAT	14_10_18	1	9	7.7	0.02
67	59_1_IF	CRISPR	RChAT	14_10_18	1	9	10	154.72
68	59_2_IF	CRISPR	RChAT	14_10_18	1	9	10	222.11
69	59_1	CRISPR	RChAT	22_10_18	1	10	1	0.02
70	59_1	CRISPR	RChAT	22_10_18	2	10	1.1	0.01
71	59_1	CRISPR	RChAT	22_10_18	3	10	1	0.03
72	59_2	CRISPR	RChAT	22_10_18	1	10	9.6	0.02
73	59_2	CRISPR	RChAT	22_10_18	2	10	8.1	0.02
74	59_2	CRISPR	RChAT	22_10_18	3	10	1.1	3.15
75	59_1_IF	CRISPR	RChAT	22_10_18	1	10	10	50.04
76	59_1_IF	CRISPR	RChAT	22_10_18	2	10	10	50.08
77	59_1_IF	CRISPR	RChAT	22_10_18	3	10	10	54.43
78	59_2_IF	CRISPR	RChAT	22_10_18	1	10	10	45.05
79	59_2_IF	CRISPR	RChAT	22_10_18	2	10	10	45.96
80	59_2_IF	CRISPR	RChAT	22_10_18	3	10	10	60.57

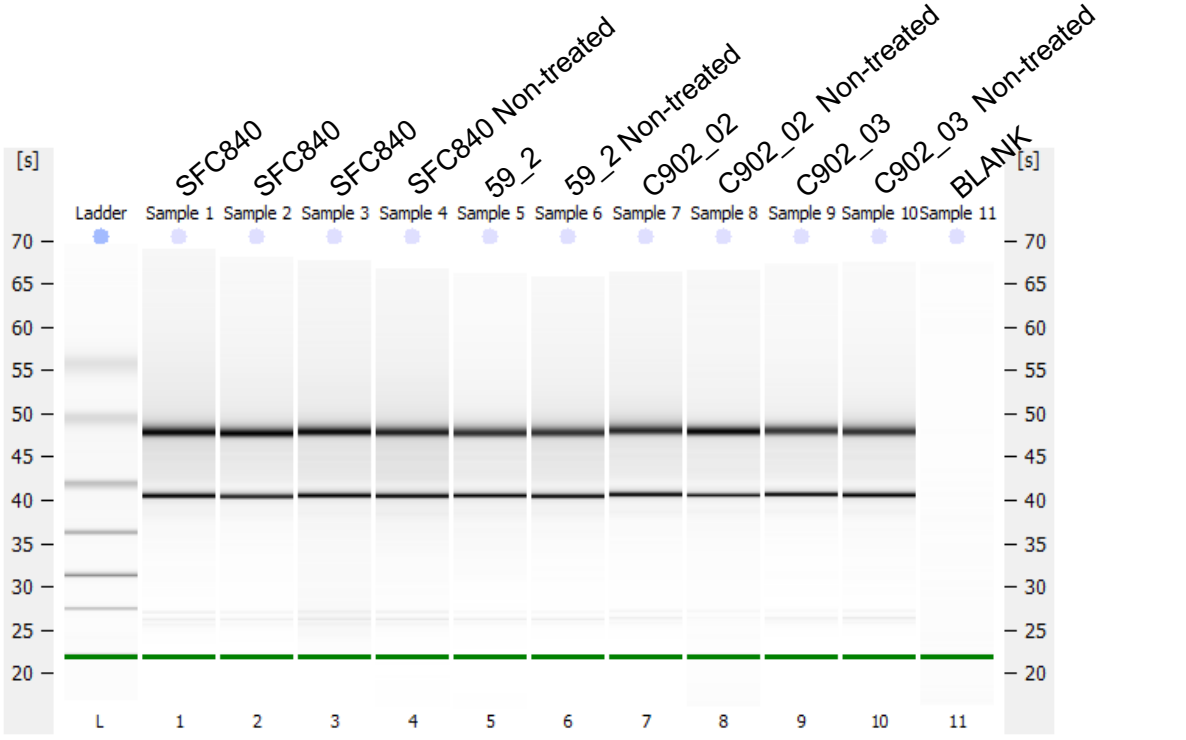
Quality control via RIN analysis was performed on of all the samples (Figure 4.8a). A brief examination of the enrichment of RPL22:c-myc:FLAG (Figure 4.8b) and ChAT (Figure 4.8c) transcripts, compared with their transcriptome input fractions in control lines was done. Interestingly, various magnitudes of enrichment were seen in the different samples, potentially highlighting the extent of variation and yield of motor neurons in separate cultures. However, this may be explained by a source variation presented by the TRAP experiment itself, originating from the lentiviral transduction of the TRAP vector or pulldown efficiency. TRAP samples and their respective transcriptome input fractions were submitted for stranded, paired-end Illumina next generation RNA sequencing, with ribodepletion during library preparation, at the Wellcome Oxford Genomics Centre for Human Genetics.

Figure 4.8. The final optimised TRAP protocol was carried out on control, patient and edited iPSC-motor neurons.

a

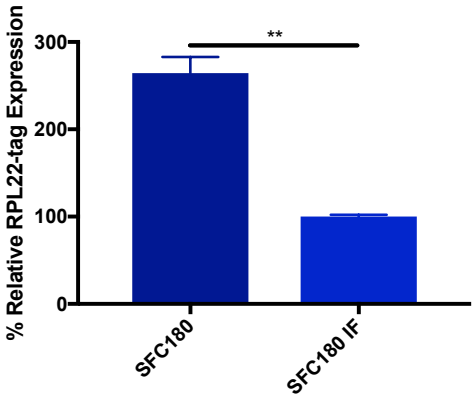


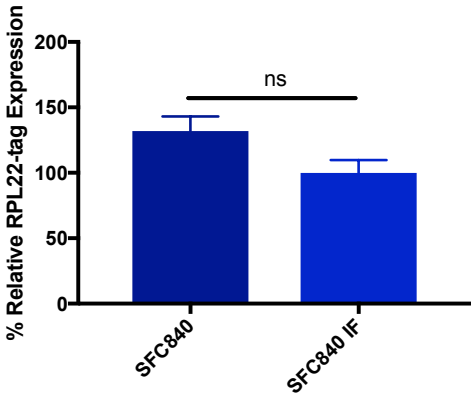
RIN:	10	9.7	-	9.8	9.7	9.9	9.8	9.6	9.6	9.7	9.1
(ng/ µL):	8.74	12.78	-	11.57	10.01	7.81	5.36	12.68	14.64	13.02	12.18

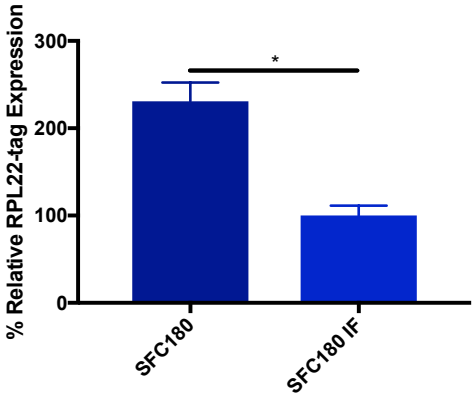


RIN:	9.7	10	9.4	9.3	9.5	9.4	9.8	10	9.7	9.6	-
(ng/ µL):	11.27	6.91	7.64	7.38	4.8	8.59	6.16	1.58	6.75	6.18	-

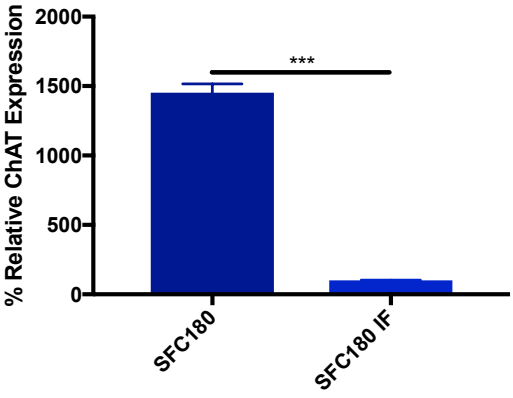
b

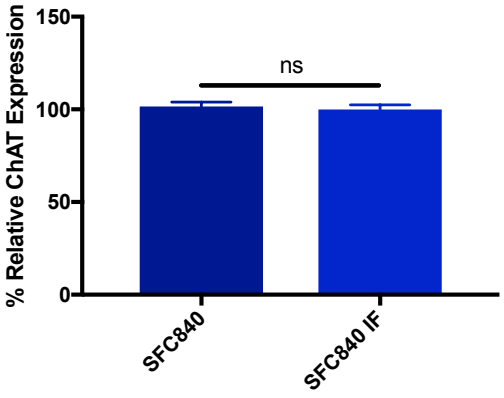

% Relative RPL22-tag Expression
300
200
100
0
**
SFC180
SFC180 IF


% Relative RPL22-tag Expression
200
150
100
50
0
ns
SFC840
SFC840 IF


% Relative RPL22-tag Expression
300
200
100
0
*
SFC180
SFC180 IF

c


% Relative ChAT Expression
2000
1500
1000
500
0
***
SFC180
SFC180 IF


% Relative ChAT Expression
150
100
50
0
ns
SFC840
SFC840 IF

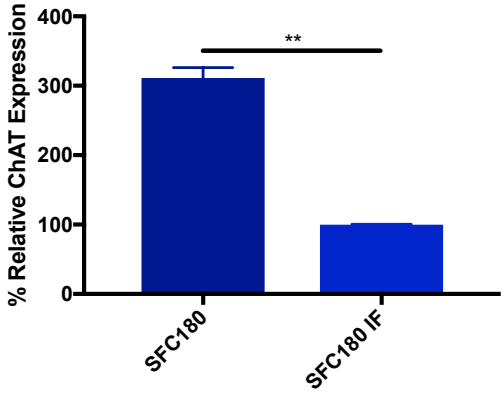

% Relative ChAT Expression
400
300
200
100
0
**
SFC180
SFC180 IF

Figure 4.8. The final optimised TRAP protocol was carried out on control, patient and edited iPSC-motor neurons.

(a) RNA quality was validated using a Bioanalyzer instrument. Increased background was seen to be pulled down from these cultures

(b) qRT-PCR was used to assess the relative number of RPL22:c-myc:FLAG transcripts in TRAP-RNA and respective transcriptome (input, IF) fractions, in control iPSC-motor neurons cultures from 3 differentiations.

(c) qRT-PCR was used to assess the relative number of ChAT transcripts in TRAP-RNA and respective transcriptome (input, IF) fractions, in control iPSC-motor neurons cultures from 3 differentiations.

5 Bioinformatic analysis of RNA sequencing translatome data

5.1 Introduction

Following RNA sequencing the raw, paired-end (2x 75 bp) reads from each submitted sample were returned to us for processing and analysis. All RNA sequencing analysis was done by the author of this thesis, with use of tools from various sources, coded into pipelines, written by the Computational Genomics: Analysis and Training (CGAT) group, University of Oxford.

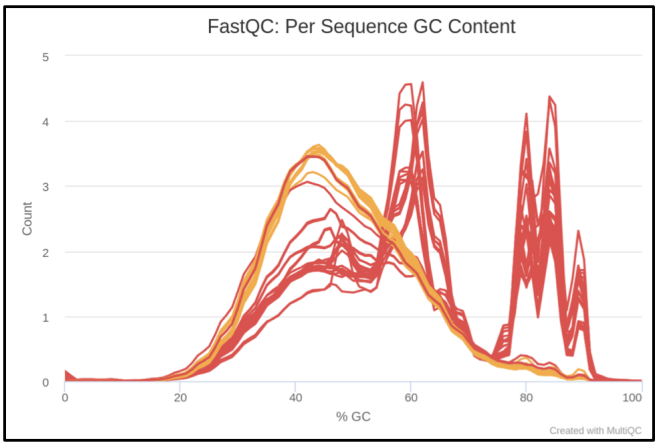
5.2 TRAP and transcriptome input samples have distinctly different quality across all raw and mapped reads

Raw reads were passed to FastQC, a tool to visualise read quality and content across all samples, by plotting their associated phred scores and extracting base per sequence content data. Parameters that can be visualised include per sequence guanine-cytosine (GC) nucleotide content, overall base content, adaptor content and sequence duplication rate estimates. Most of the samples were observed to satisfy the quality control thresholds; passing for read quality and adaptor content, negating the need to trim any reads, thus increasing overall mapping quality. However, an unusual pattern in % GC content (Figure 5.1a-b) and duplication rate (Figure 5.1b) was seen, where these parameters were higher than expected in the transcriptome input samples, and the paired-end reads from two TRAP samples, SFC180 and SFC840 from the first differentiation performed (see Table 4.1). These parameters were also closely correlated (Figure 5.1b). Investigation into this feature of the data was further examined following the mapping of raw *fasta* file reads to an annotated human genome database, modified

to include an artificial *sequin* contig and the over-expressed RChAT:TRAP vector sequences. It should also be noted that the TRAP samples had the highest percentage of uniquely mapped reads to the human genome (~85%), excluding two samples from the SFC180 and SFC840 control lines, which are more similar to their transcriptome input fractions (Figure 5.1c). The transcriptome input samples contained far more duplicate reads. All of the samples had minimal artefactual reads, and less than 5% unmapped reads for most samples.

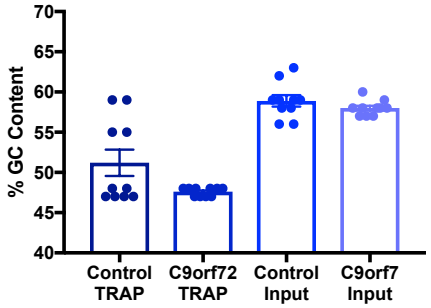
Figure 5.1. Quality control of raw reads from TRAP and transcriptome input fractions following RNA sequencing using FastQC.

a

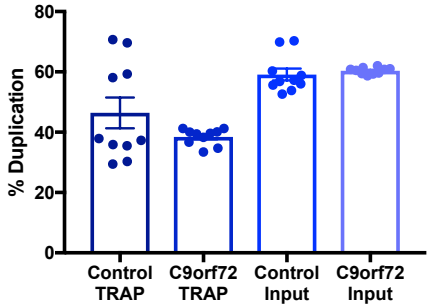


b

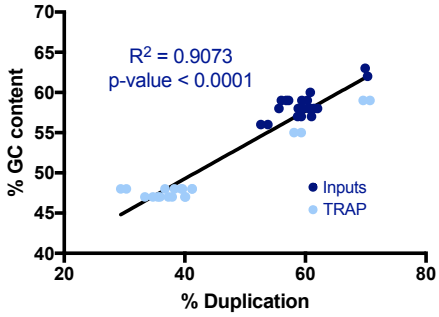
i



ii



iii



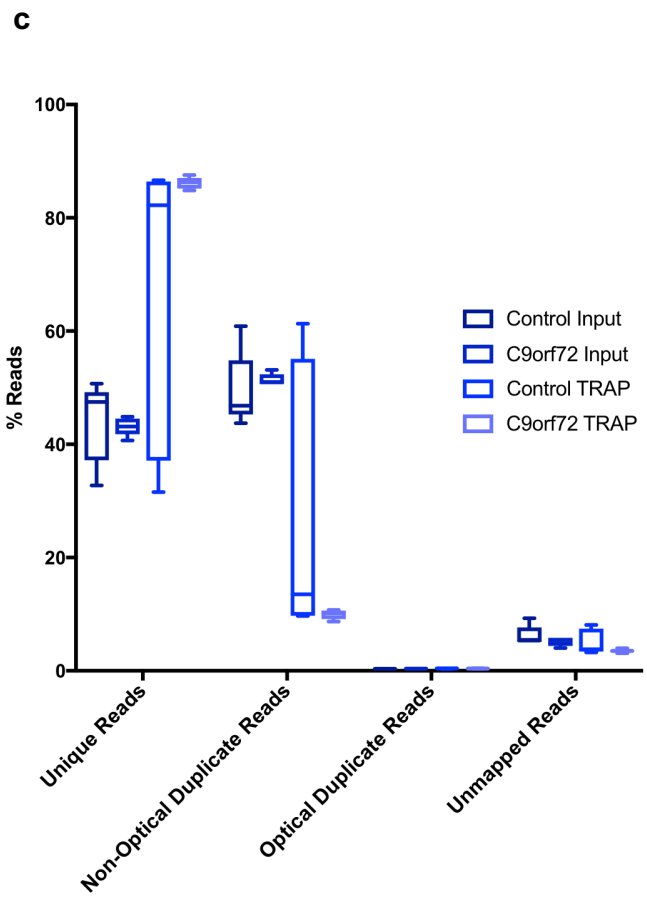


Figure 5.1. Quality control of raw reads from TRAP and transcriptome input fractions following RNA sequencing using FastQC tool.

(a) Distribution of % GC content of raw reads from TRAP and transcriptome (Input) samples submitted for RNA sequencing. A batch of samples were passably evenly contained an even % GC content across all reads, but another significant group of samples were found have a significant number of reads with up to 90 % GC content.

(b) The % guanine-cytosine (GC) content (i) in raw paired reads and % duplication rate (ii) from RNA sequencing experiment is unevenly distributed between samples, as calculated by FastQC software. These two features were plotted against each other and found to be closely correlated (iii).

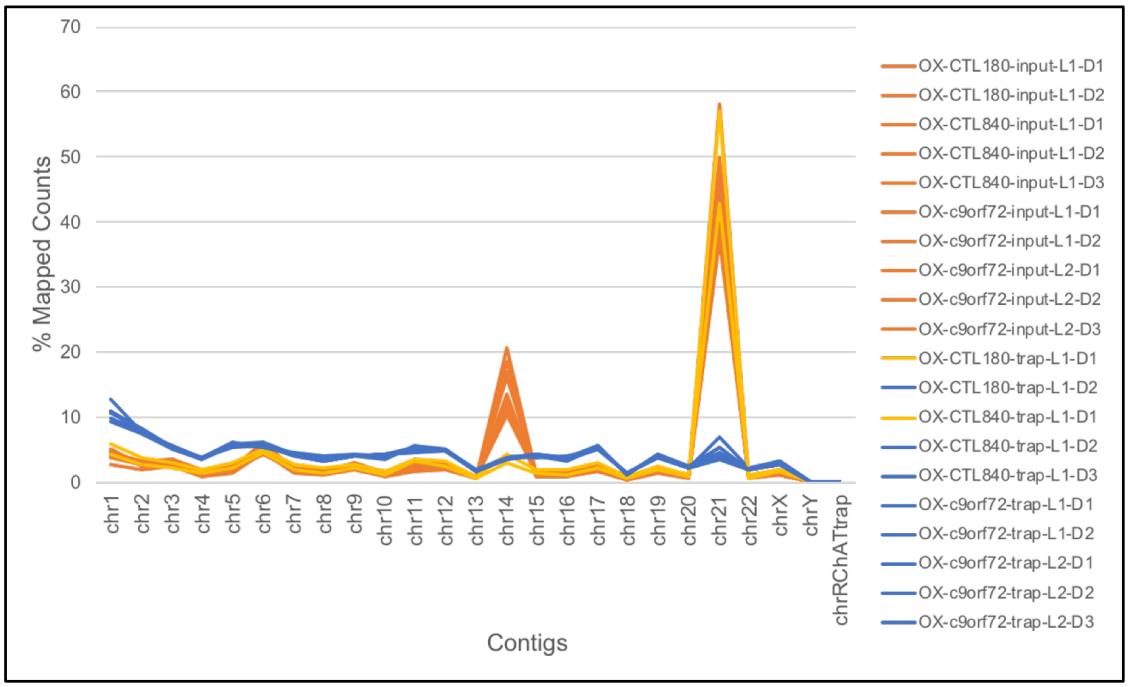
(c) Picard tools were used to plot the proportion of unique reads, non-optical duplicates, optical duplicates and unmapped reads that were mapped to the human genome, in each sample group. The transcriptome input and TRAP samples have only ~50% unique reads in inputs and two Control-TRAP outlier samples, compared with >80% in all other TRAP samples.

Mapping analysis revealed differences in read depth between transcriptome input and TRAP samples. Most of the TRAP samples had between 26.8 million to 34.4 million uniquely mapped reads. Transcriptome input fractions, and two TRAP samples, OX-SFC180-trap-L1-D1 and OX-SFC840-trap-L1-D1, had between 17 million to 22.4 million uniquely mapped reads. Examination of read coverage across annotated human contigs showed that transcriptome input samples, and the previously seen control-TRAP outliers, contained a high proportion of reads which mapped to chromosomes 14 and 21 (Figure 5.2a). The latter accounts for ~40% of total reads in transcriptome input samples, effectively decreasing the read depth of other contigs, compared with TRAP samples (Figure 5.2a-b). This has potential implications for downstream analysis, making it challenging to normalise transcriptome and TRAP data together. To understand where this difference in mapping originated from, the genome features accounting for the most mapped reads on chromosome 21 for transcriptome input samples where plotted, normalised against the total reads (Figure 5.2b). The two control transcriptome input fractions previously seen as outliers, OX-SFC180-trap-L1-D1 and OX-SFC840-trap-L1-D1, were consistently seen to have similar overall characteristics to transcriptome samples (Figure 5.2). Though the reason for this is not clear, due to their ambiguity, they were removed from downstream differential expression analysis. Generalisations made to TRAP samples henceforth will not include these two samples, unless otherwise stated. The highest proportion of reads found to map to chromosome 21 can be attributed to intergenic reads, which are significantly depleted in TRAP samples (Figure 5.2b). The next most enriched RNA species in transcriptome input samples were short non-coding RNAs and ribosomal RNA (rRNA), which were also depleted in TRAP samples (Figure 5.2b). This depletion is

expected of successful TRAP samples, which should be highly enriched with mature mRNA, with some rRNA contamination due to pulldown of polysome complexes. Particularly high numbers of rRNA reads are expected to map to chromosome 14 and 21 as there are known to encode for a high proportion of ribosomal DNA regions, though does not account for the intergenic reads (Henderson et al., 1972). Subsequent examination of reads mapping to protein coding regions on chromosome 21 show that TRAP samples are significantly enriched for these (Figure 5.2b). These high intergenic reads may be particularly enriched in transcriptome input samples due to the chosen library preparation method to enrich for mRNA by using the ribodepletion protocol. Although it can be decisively shown that TRAP samples are far more enriched with mRNA or protein-coding RNA, the resulting decreased coverage in transcriptome input samples presents a challenge for downstream analysis. Furthermore, the increased library complexity of the transcriptome input samples also presents a potential source of variation during RNA sequencing experiment. The traditional method of Poly A tail selection was not chosen as a previous bulk RNA sequencing experiment conducted in this lab group had a distinct bias for three prime reads (Scaber et al., unpublished). A study comparing these two protocols in identical bulk samples for RNA sequencing experiments showed similar results and overall coverage, though ribodepletion was slightly more varied (Kumar et al., 2017). With this in mind, future TRAP experiments may benefit from Poly A tail-selection, however the impact on samples which are prepared differently such as, TRAP and transcriptome input samples is not easily predictable.

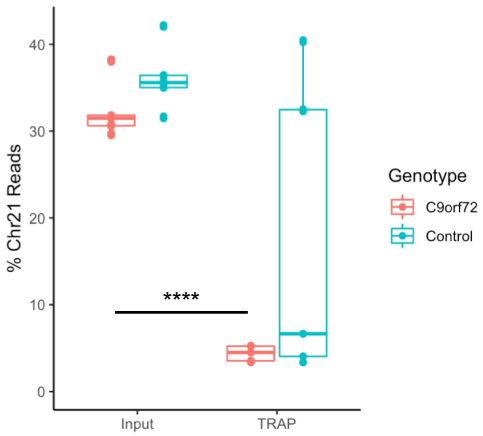
Figure 5.2. Assessment of reads mapping to hg38 annotated human genome shows an over-representation of reads in chromosome 21.

a

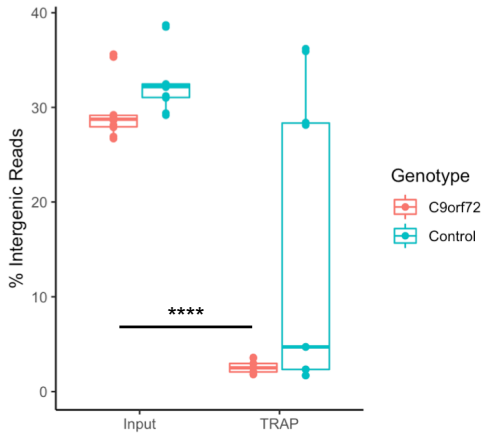


b

i



ii



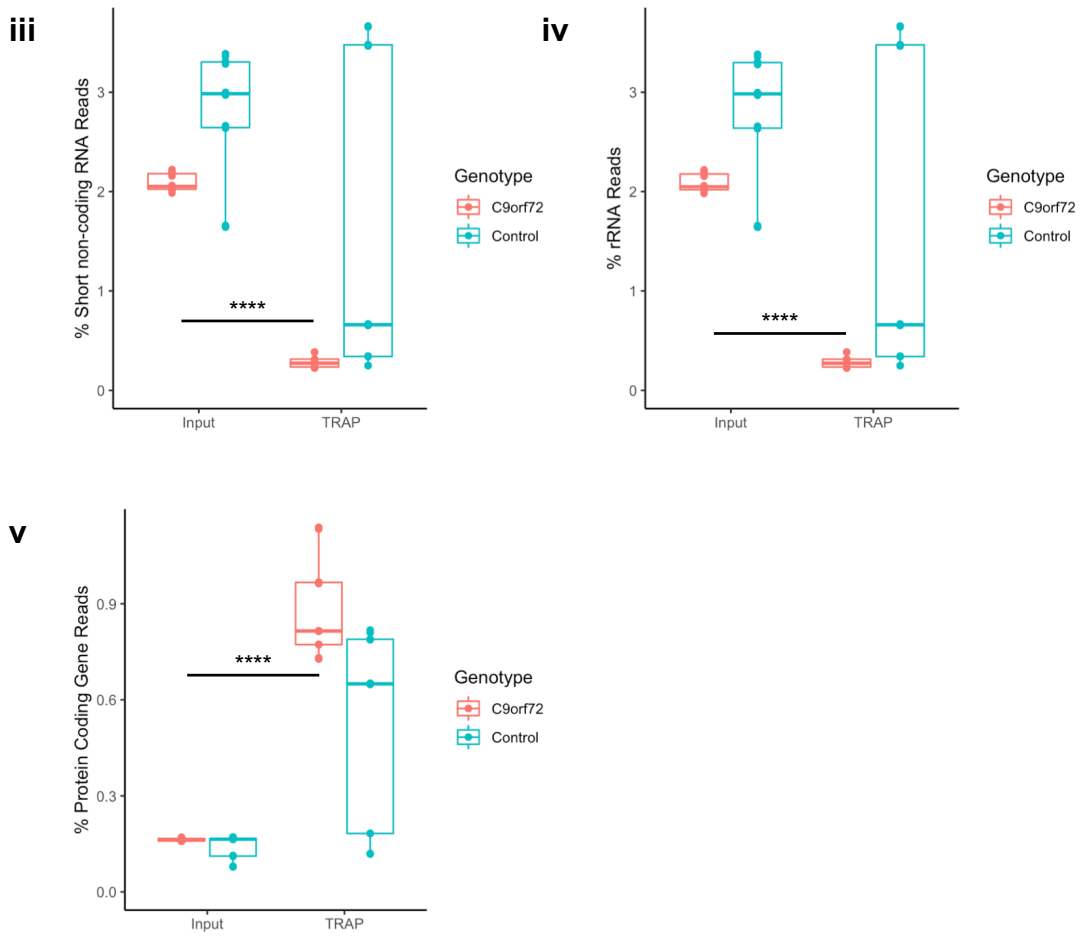


Figure 5.2. Assessment of reads mapping to hg38 annotated human genome shows an over-representation of reads in chromosome 21.

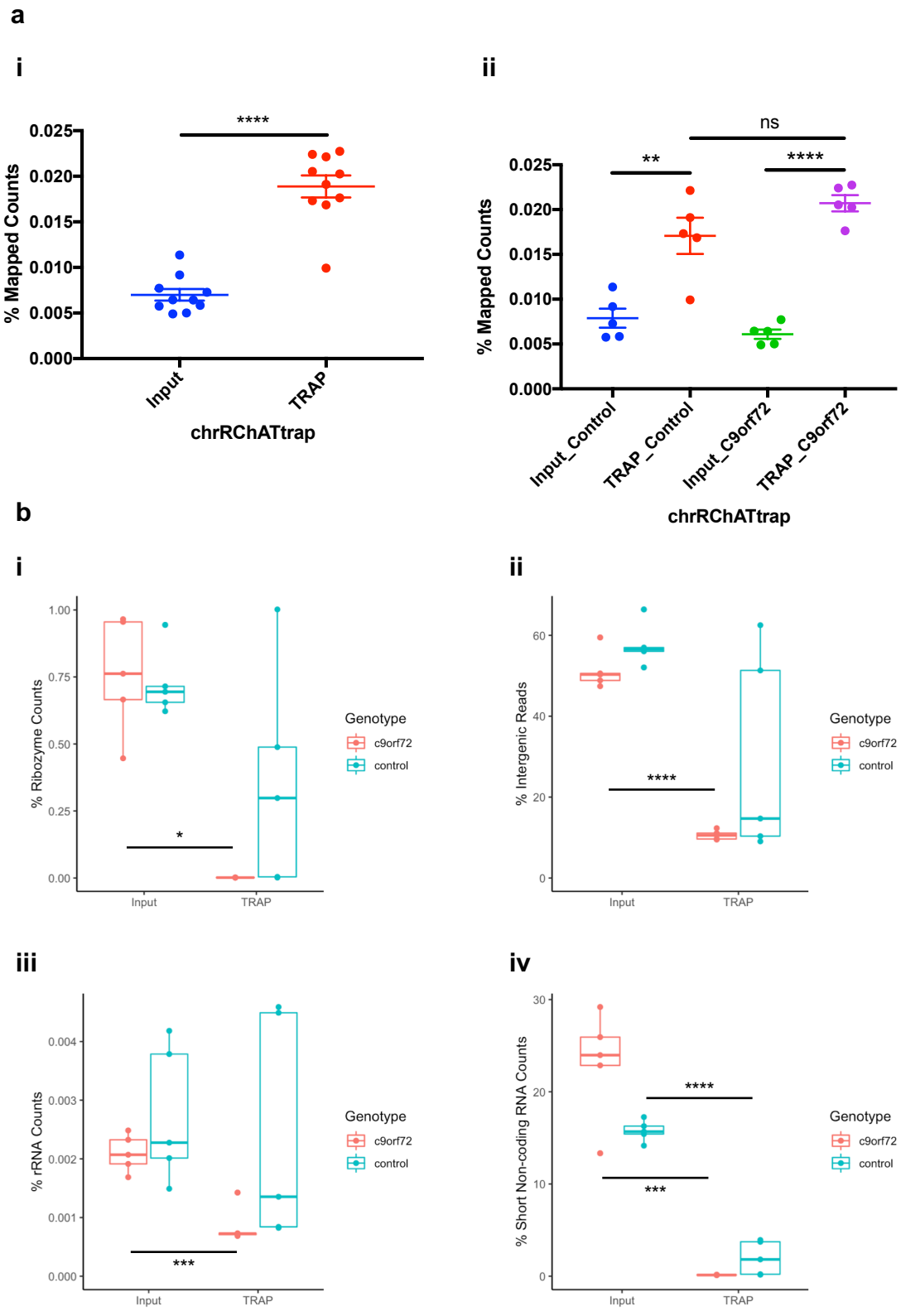
(a) Percentage of mapped reads to annotated contigs (chromosomes) across TRAP and transcriptome input samples from patients and controls show a significant increased number of reads for transcriptome input samples, and two control TRAP outliers, to chromosome 21 and chromosome 14. Other contigs are relatively more closely matched in coverage, though transcriptome input samples, and two control TRAP outliers have moderately lower reads in other chromosomes.

(b) An increased proportion of reads have been mapped to chromosome 21 in all transcriptome input samples compared to most TRAP samples (i), calculated against the total number of reads detected from each sample. The greatest proportion of reads on chromosome 21 map to intergenic (ii), short-noncoding (iii) and rRNA (iv)-coding regions. Reads from TRAP samples (except two control samples) that mapped to protein coding regions on chromosome 21 are significantly increased compared with transcriptome (Input) samples (v). Unpaired, two-tailed Student's T-test, assuming unequal standard deviations (Welch's correction). P-value < 0.05*, 0.001**, 0.005***, 0.0001****

5.3 TRAP samples are enriched with mature mRNA species compared to controls

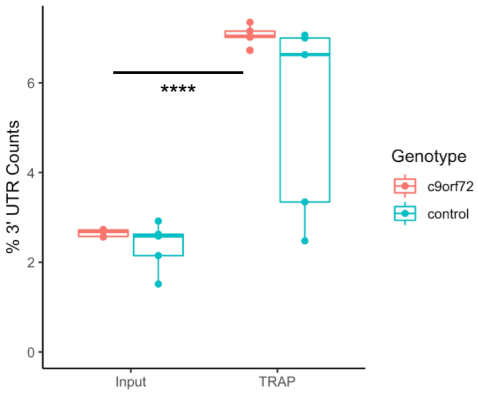
To confirm overall enrichment of mature mRNA in TRAP samples, the mapping of reads across the whole genome was investigated between sample groups (Figure 5.3). The percentage of mapped reads to the artificial RChAT:TRAP transcript in TRAP samples showed that these transcripts are significantly enriched compared to the transcriptomes, apart from OX-SFC840-trap-L1-D1 control sample (Figure 5.3a). There was also no significant difference in % mapped reads to the TRAP vector between control and *C9orf72* TRAP groups (Figure 5.3a), suggesting similar level of expression following lentiviral transduction. Percentage of reads mapping to RNA species, which should be depleted in TRAP samples were plotted, including ribozyme sequences, intergenic reads, general rRNA and short non-coding RNA species (Figure 5.3b). In all cases TRAP samples were depleted of such reads compared with transcriptome input samples, aside from the same aforementioned outliers. Conversely, reads mapping to three-prime untranslated and protein-coding regions were significantly enriched in successful TRAP samples (Figure 5.3c). Venn diagrams to summarise the average proportion of various RNA species across the sample groups are given in Figure 5.3d. Brief examination of the coverage of standard housekeeping genes across sample was done using an interactive genome browser (not shown here). Similar coverage of exons was seen within sample groups for these genes.

Figure 5.3. Quality control of reads mapping to hg38 annotated human genome.

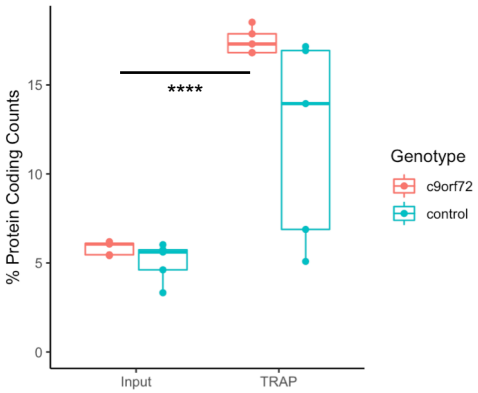


c

i


% 3' UTR Counts
0
2
4
6
****
Input
TRAP
Genotype
c9orf72
control

ii


% Protein Coding Counts
0
5
10
15
****
Input
TRAP
Genotype
c9orf72
control

d

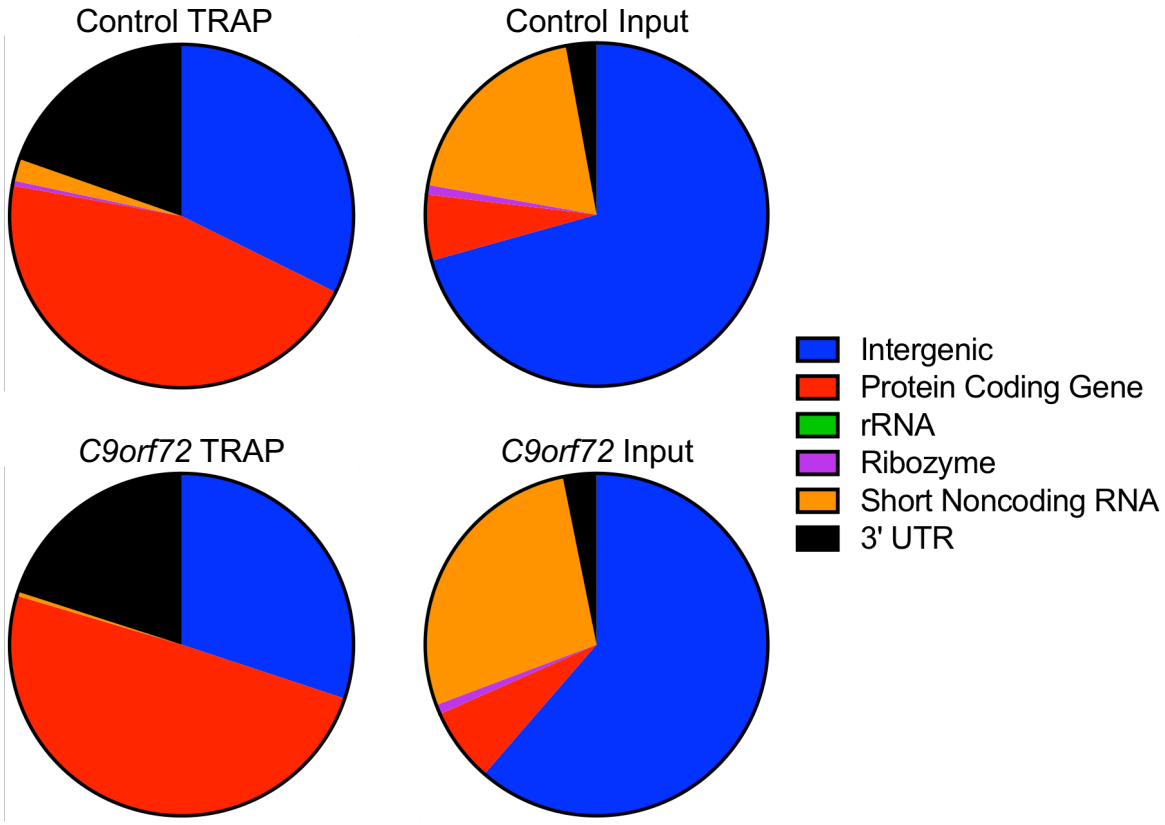

Control TRAP
Control Input
C9orf72 TRAP
C9orf72 Input
Intergenic
Protein Coding Gene
rRNA
Ribozyme
Short Noncoding RNA
3' UTR

Figure 5.3. Quality control of reads mapping to hg38 annotated human genome.

(a) The percentage of mapped reads to artificial RChAT:TRAP vector contig between TRAP and transcriptome (Input) (i) samples and across internal groups of patient and controls (ii).Unpaired, two-tailed Student's T-test, assuming unequal standard deviations (Welch's correction). P-value < 0.001**, 0.0001****

(b) The percentage of reads mapping to non-coding regions of the annotated human genome. Reads mapping to ribozyme (i), intergenic (ii), rRNA (iii) and short non-coding RNAs (iv) were depleted in TRAP samples compared to transcriptome (Input) samples. Unpaired, non-parametric two-tailed Student's T-test, assuming equal standard deviations where relevant. P-value < 0.05*, 0.005***, 0.0001****

(c) The percentage of reads mapping to 3' untranslated regions (UTR) regions of transcripts and protein-coding regions are significantly enriched in TRAP samples. Unpaired, non-parametric two-tailed Student's T-test, assuming equal standard deviations where relevant. P-value < 0.05*, 0.005***, 0.0001****

(d) Proportion of significant RNA species detected within samples groups. Outliers, OX-SFC840-trap-L1-D1 and OX-SFC180-trap-L1-D1 in Control-TRAP were excluded from plots.

With the confirmation of good mapping quality within TRAP and transcriptome input groups, and the enrichment of mature mRNA in TRAP samples, differential expression analysis could be explored, following quantification of mapped reads to gene loci.

5.4 Differential expression analysis

5.4.1 Introduction

Following the initial steps to establish the quality of raw and mapped reads of the RNA sequencing samples, the aforementioned outlier samples from the control-TRAP group were removed. A raw counts table (output from *Salmon*) was imported into R to perform differential expression analysis via the *DESeq2* package (Love et al., 2014). A generalised linear model (GLM) was built for the dataset, including information regarding sample differentiation date, experimental lane and phenotype condition. *DESeq2* can hypothetically correct for library size internally, allowing for the differences in genome coverage seen for transcriptome input samples to be accounted for, as protein-coding genes were filtered for in quantification process. This was further investigated by using spike-in controls to normalise the RNA sequencing dataset. Normalised counts obtained using this method were used in further analysis, as described below.

5.4.2 *Sequin* normalisation for the TRAP and transcriptome datasets is challenging

Synthetic RNA isoforms, known as *sequins*, were spiked into all TRAP and transcriptome input samples prior to library preparation, to act as an internal control

while correcting for technical biases within the RNA sequencing dataset. Quantification of these spike-ins was possible due to the presence of an annotated artificial genome added to the human genome file during the mapping and counting of reads from the experiment. The averaged raw counts (transcripts per million, TPM) per isoform detected within each sample group correlated well with the input concentration of *sequins* (Figure 5.4a). One of two mixtures of *sequins* can be used to spike a given sample. Each mixture contains a set of artificial RNA isoforms at different concentrations. Two groups of samples spiked with different mixtures of *sequins* can theoretically be used to calibrate differentially expressed genes. Differential expression analysis was done using sample groups, *C9orf72*-TRAP, control-TRAP, *C9orf72*-input and Control-input, differentiation date and RNA sequencing lane to construct a generalised linear model to normalise the data. From this the observed log2 fold change values (p-adjusted < 0.05) for *sequins* between *C9orf72*-TRAP and control-TRAP groups were obtained and plotted against the expected log2 fold change values (Figure 5.4bi). A poor correlation was seen (R^2 = 0.401), suggesting an amount of technical variability between samples. In order to correct for this, the *sequin* raw counts predicted to give a log fold change of zero, were used as a control against the raw counts table comprising all detected human genes within sample groups. However, this did not appear to solve the inherent variability present in the dataset (Figure 5.4bii). A poorer correlation between the expected and observed log2 fold changes (R^2 = 0.114) was seen. Various attempts to efficiently correct the variation in the data were made (not shown here). Models incorporating only differentiation dates and *sequin* correction, lane and *sequin* correction, including several others were run. An issue while trying to correlate expected and observed values of log2 fold change for *sequins*, was

obtaining enough significant differentially expressed values to correlate. This can be seen in Figure 5.4bi, where very few data points were obtained. Generally, this improved when *sequin* normalisation was used, though perfect expected log fold change were not obtained. In the final analysis, *sequins* were used to normalise the data, and only the most crucial factors to the experiment (lane and condition) were included in the GLM. Differentiation was omitted from the model because there is likely to be more biological significance in this batch effect, which may be lost if used to correct the data.

Figure 5.4. Sample spike-in analysis to normalise RNA sequencing dataset.

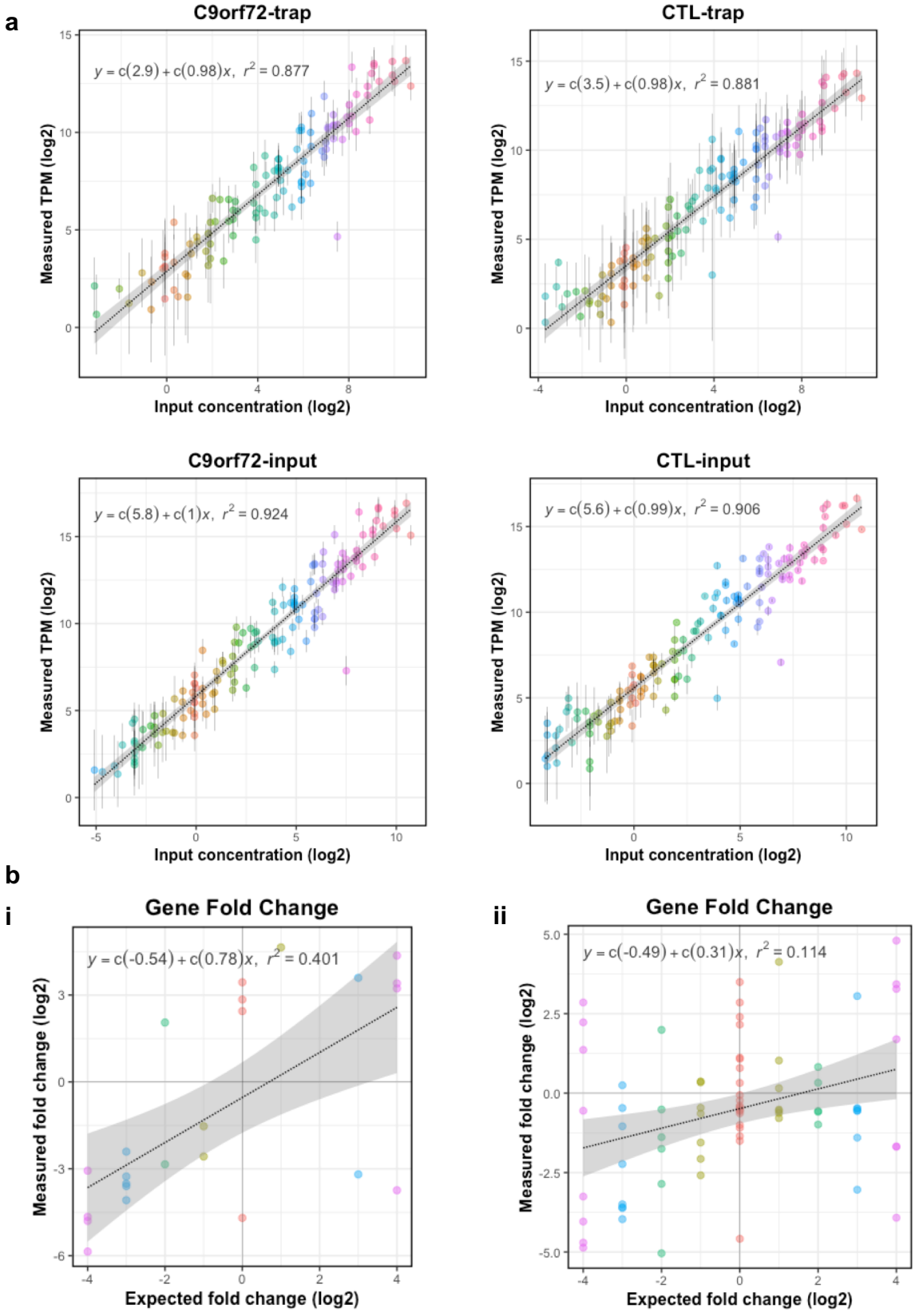


Figure 5.4. Sample spike-in analysis to normalise RNA sequencing dataset.

(a) The unnormalised counts (transcripts per million, TPM) of synthetic *sequin* transcript isoforms from each sample submitted for sequencing were plotted against the known concentrations from their respective spike-in mixtures. This allows estimation of the quality and reliability of library preparation and experiment across flow cell lanes. Strong R^2 values were seen for all samples.

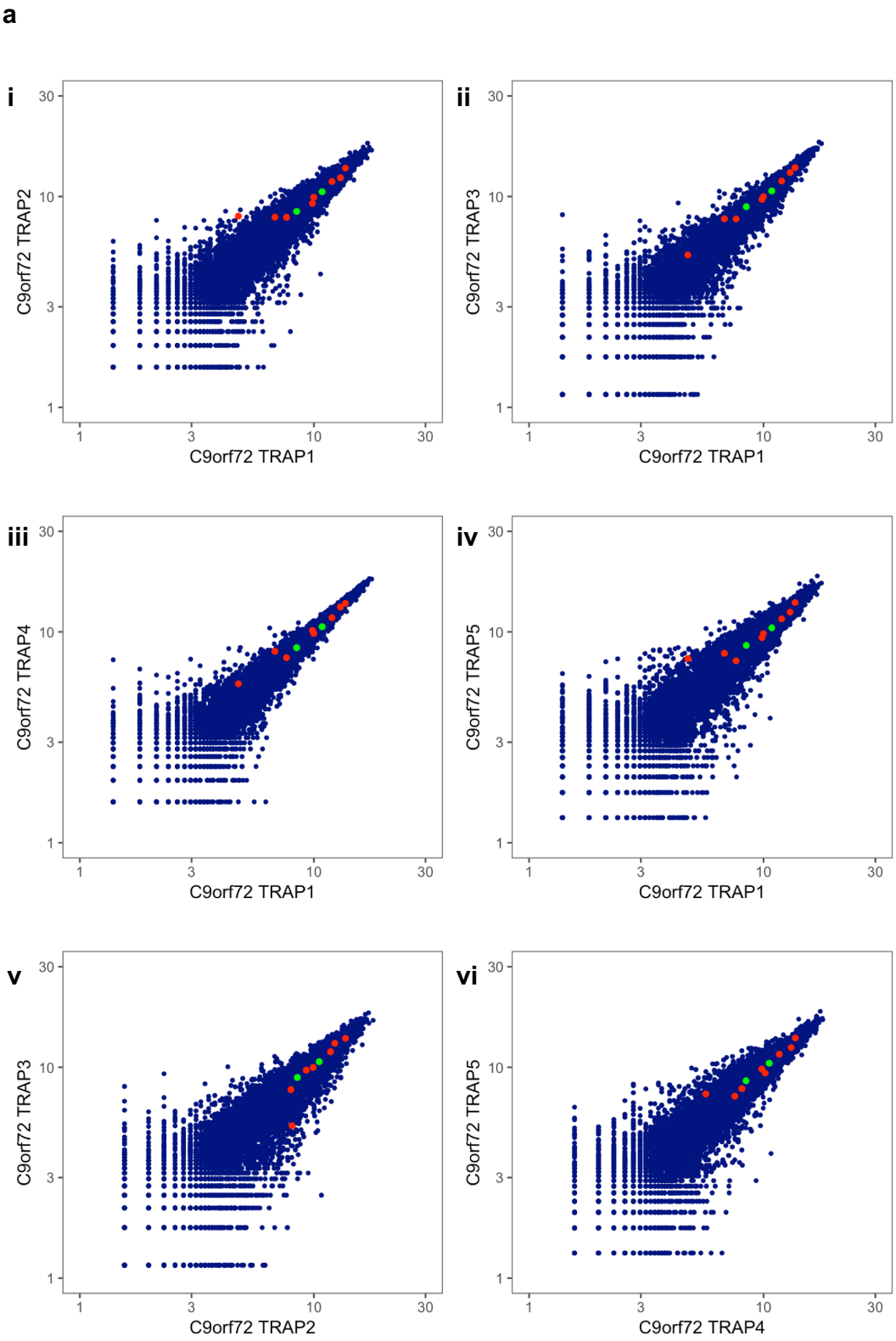
(b) The statistically significant log2 fold change values derived from differential expression analysis between C9orf72-TRAP and Control-TRAP samples, were plotted against the expected log2 fold changes between two different *sequin* mixtures. The values calculated from a generalised linear model, taking into account lane and experimental condition (phenotype and sample type) did not give a strong correlation $R^2 = 0.401$ (i). Counts table was imported to *RUVSeq* in R to attempt to correct for this variation by using *sequin* counts with a predicted zero log fold change as a control in the data (ii). This did not improve the correlation ($R^2 = 0.114$).

5.4.3 Pulldown of transcripts between individual samples are consistent

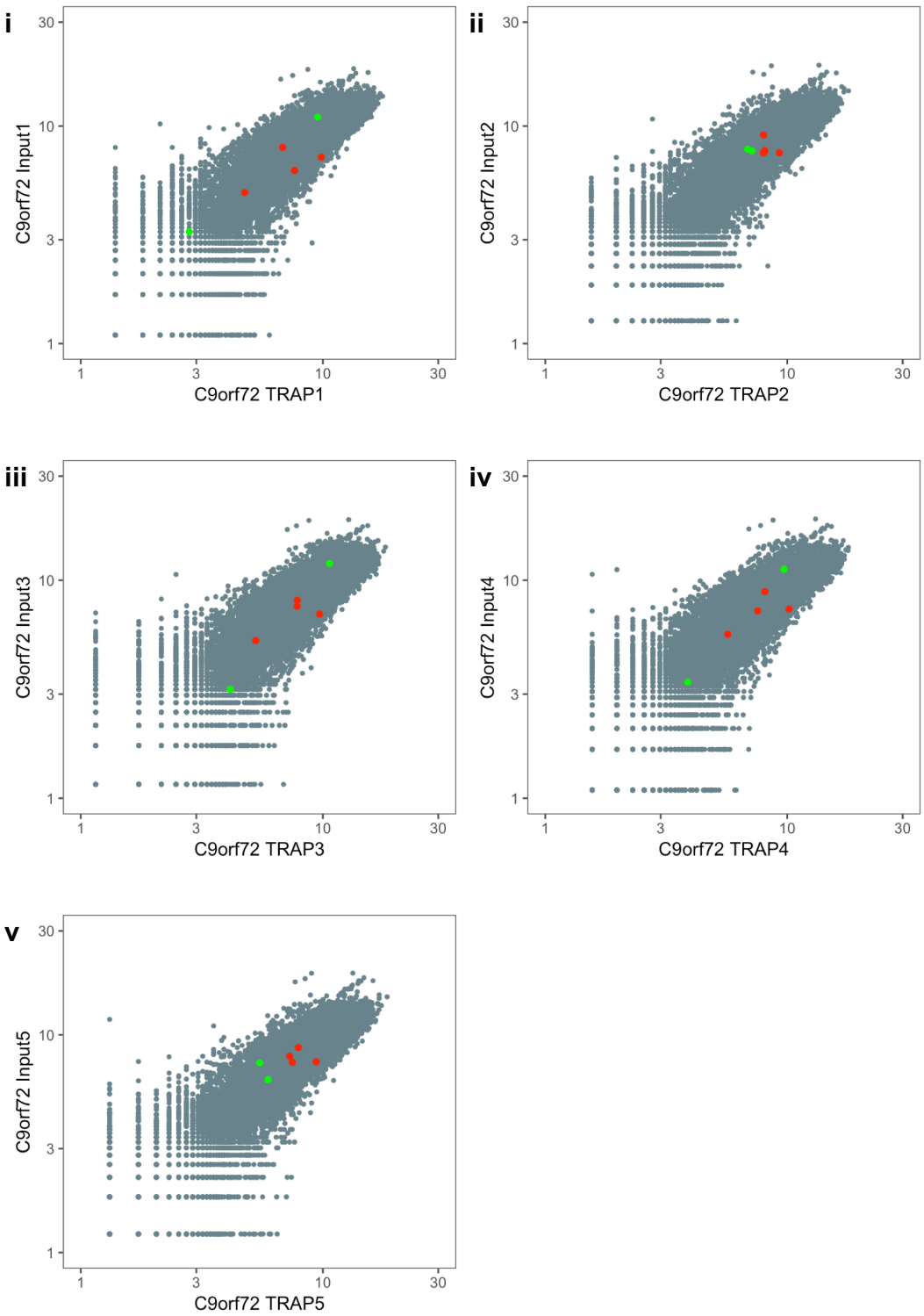
A potential challenge of using translatome data to interpret gene expression changes in a complex model such as human-derived iPSC-differentiated neuronal cultures, is the variability that is likely to arise from experimental conditions. To examine the consistency of transcript pulldown, the normalised gene counts between individual samples were plotted against each other (Figure 5.5). A strong positive correlation is seen between *C9orf72* patient TRAP samples (Figure 5.5a), especially with greater measured counts, with greater variability at lower counts. This is likely due to the high incidence of false positives and variation for lower counts, most of which at filtered for the final differential expression analysis (threshold > 10 counts). Encouragingly, there are generally a proportionally low number of outliers. Similar correlations were plotted for control lines from two individuals, SFC180 and SFC840. Positive correlations were seen for all samples (Figure 5.5c), with most reliable pulldowns seen from cultures originating from the same control line (Figure 5.5ciii). Transcript counts between SFC180 and SFC840 are less reliable than plots between patient lines, which were all from the same individual. Red points highlight housekeeping genes *GAPDH*, *ACTB*, topoisomerase-1 (*TOP1*) and ATP synthase subunit beta (*ATP5B*) (Rydbirk et al., 2016), and motor neuron markers *ChAT*, Islet-1 (*ISL1*), beta-III tubulin (*TUBB3*) and neurofilament heavy marker (*NEFH*). Green points highlight *C9orf72* and *TARDBP* normalised transcript counts. These transcripts pulldown consistently between samples, demonstrating the reliability of the technique. The relationships between TRAP samples and their corresponding transcriptome input fractions were also examined (Figure 5.5b, d) to demonstrate the shift in RNA expression profiles between the entire transcript pool and actively translated mRNA from ChAT-

enriched motor neurons. Here, the red points are highlighting motor neuron markers: *ChAT*, *ISL1*, *TUBB3* and *NEFH*. The green points represent neural progenitor markers *MNX1* and *OLIG2*, which would be expected to represent a higher proportion in transcriptome input fractions. Plots which do not show this shift in markers between TRAP and transcriptome input samples, such as in Figure 5.5b, potentially suggest cultures with a high proportion of motor neurons to begin with. Finally, the total detected mRNA species both at gene and transcript level was examined in all samples, and analysed for overlap to determine if the population genes being detected are similar (Figure S8.2). Across all the samples, almost 20,000 genes and 145,000 transcripts were detected. Significant overlap in genes between the *C9orf72*-TRAP and Control-TRAP samples were detected, with only 1.5-2.5% unique genes. This was slightly increased at transcript level, with ~10% unique genes in each group, but 90% overlap. Notably, TRAP samples from *C9orf72* ALS or control iPSC-motor neurons overlapped with each other more strongly than with the transcriptome input fractions (Figure S8.2).

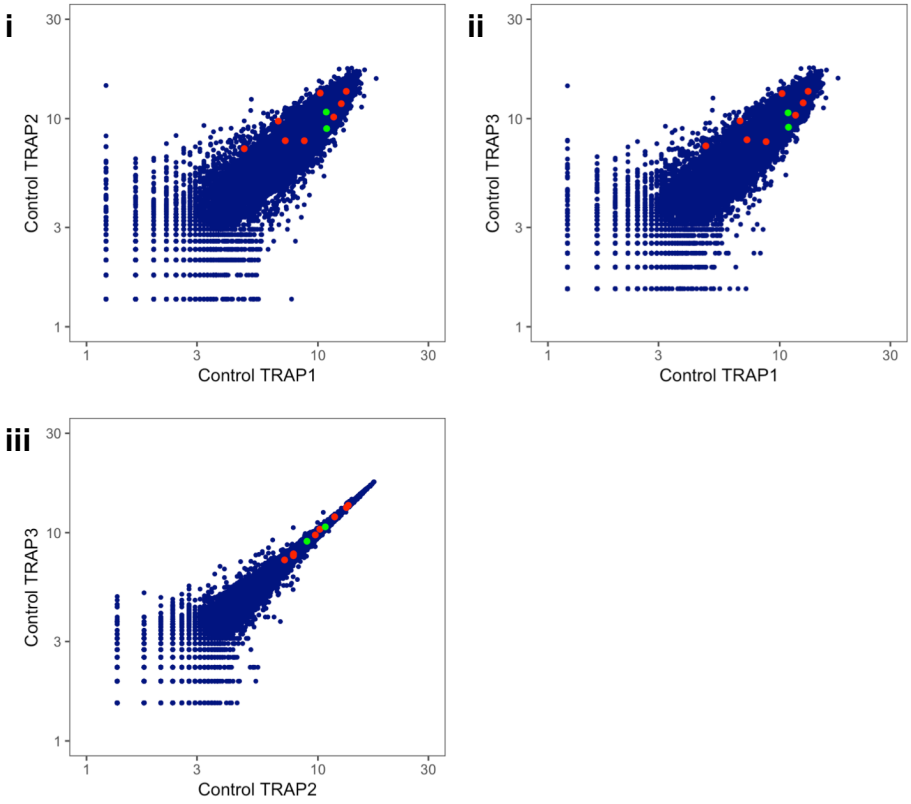
Figure 5.5. The normalised counts of all detected genes from individual TRAP samples demonstrates the reproducibility of counts obtained per gene.



b


i
C9orf72 Input1
C9orf72 TRAP1
ii
C9orf72 Input2
C9orf72 TRAP2
iii
C9orf72 Input3
C9orf72 TRAP3
iv
C9orf72 Input4
C9orf72 TRAP4
v
C9orf72 Input5
C9orf72 TRAP5
1
3
10
30

c


i
Control TRAP2
Control TRAP1
ii
Control TRAP3
Control TRAP1
iii
Control TRAP3
Control TRAP2
1
3
10
30

d

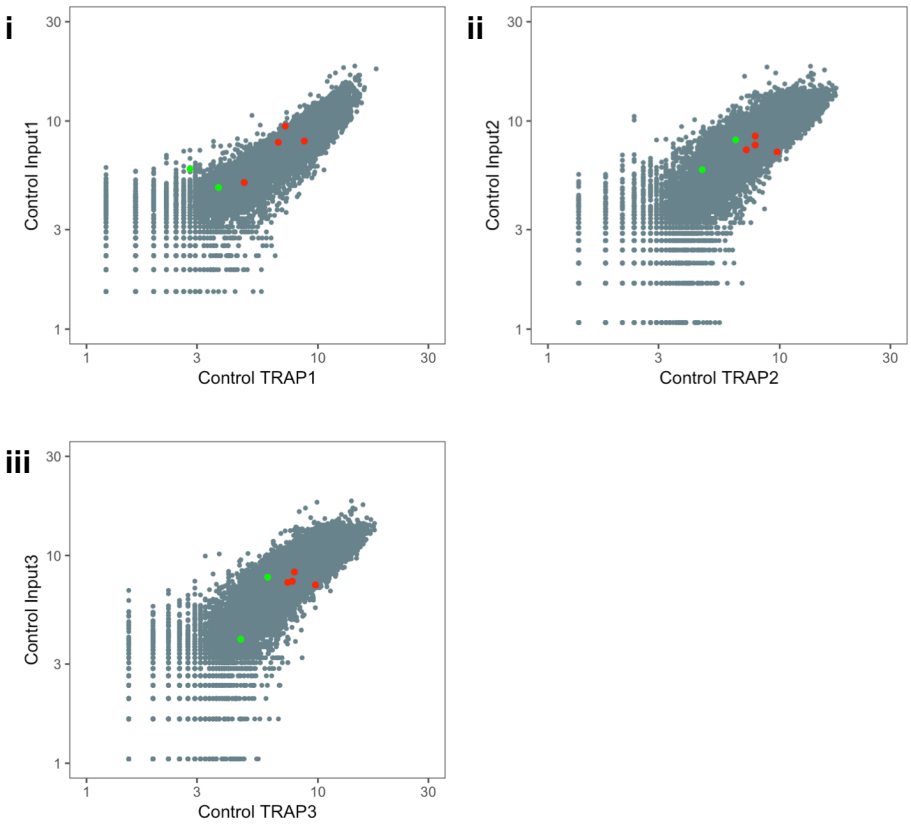

i
Control Input1
Control TRAP1
ii
Control Input2
Control TRAP2
iii
Control Input3
Control TRAP3
1
3
10
30

Figure 5.5. The normalised counts of all detected genes from individual TRAP samples, from the same group, were plotted against each other (log2 scale) to examine the potential reproducibility of counts obtained per gene, as a measure of the reliability of this TRAP method.

(a) Patient TRAP sample normalised counts were extracted and plotted to look for reproducibility between samples (log2 scale). C9orf72 TRAP1 = OXC90202; C9orf72 TRAP2 = OXC90202; C9orf72 TRAP3 = OXC90203; C9orf72 TRAP4 = OXC90203; C9orf72 TRAP5 = OXC90203. Red = *GAPDH*, *ACTB*, *TOP1*, *ATP5B*, *ChAT*, *ISL1*, *TUBB3*, *NEFH*; green = *C9orf72*, *TARDBP*

(b) Normalised counts from individual patient samples were plotted against their respective input fractions, to highlight the difference in transcriptome input and translatome profiles for individual genes (log2 scale). Red = *ChAT*, *ISL1*, *TUBB3*, *NEFH*; green = *MNX1*, *OLIG2*

(c) Age-matched control TRAP sample normalised counts were extracted and plotted to look for reproducibility between samples (log2 scale). Control TRAP1 = SFC180; Control TRAP2 = SFC840; Control TRAP3 = SFC840. Red = *GAPDH*, *ACTB*, *TOP1*, *ATP5B*, *ChAT*, *ISL1*, *TUBB3*, *NEFH*; green = *C9orf72*, *TARDBP*

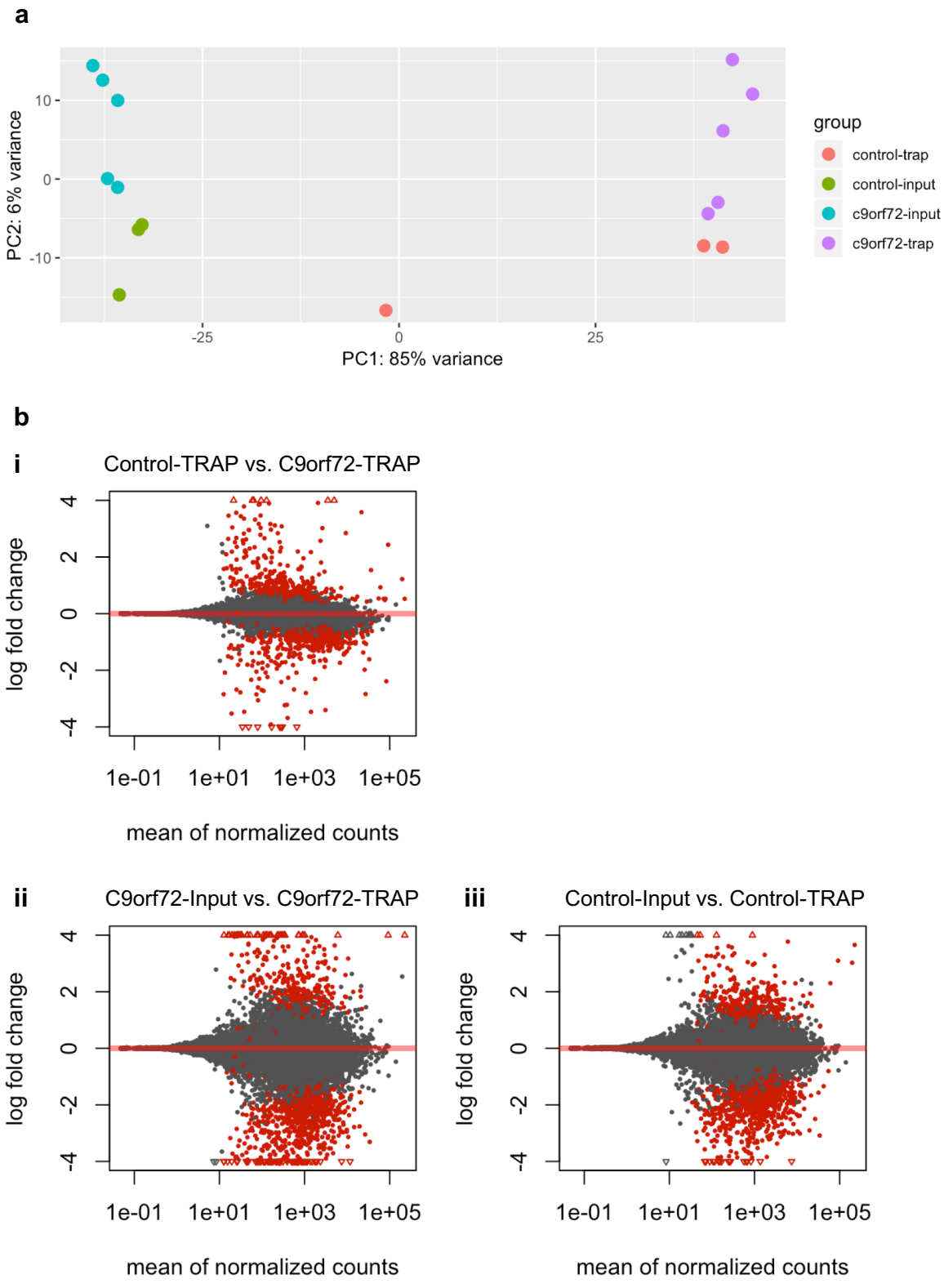
(d) Normalised counts from individual control samples were plotted against their respective input fractions, to highlight the difference in transcriptome input and translatome profiles for individual genes (log2 scale). Red = *ChAT*, *ISL1*, *TUBB3*, *NEFH*; green = *MNX1*, *OLIG2*

5.4.4 Exploratory data analysis shows distinct clustering of samples

Variance stabilising transformation was used to condense the variance of the dataset for principal component analysis (PCA). The PCA plot shows the clustering of samples along the first principal component (PC1) by transcriptome input and TRAP samples, and then further by phenotype, *C9orf72* mutation and controls, in PC2 (Figure 5.6a). This suggests the potential impact of the TRAP method in pulling out distinct translatome profiles to the degree of diminishing the effect of the mutation itself compared with age-matched controls. It should also be noted that control-TRAP samples from the SFC840 line cluster closer to patient-TRAP samples (Figure 5.6a). Specific analysis to obtain statistically significant (adjusted p-value < 0.05) differentially expressed genes between *C9orf72* and control TRAP samples showed that 478 genes were upregulated, and 285 genes were downregulated (Figure 5.6b). Differential gene expression between TRAP and transcriptome input samples (Figure 5.6b) had several overlapping genes (Figure 5.6c), suggesting a strong reproducibility of genes expressed at basal level, enriched in motor neurons. However, many unique genes were also detected from *C9orf72* and control iPSC-motor neurons. 645 genes are unique to the *C9orf72*-TRAP and control-TRAP comparison. Differential expression analysis between the *C9orf72*-Input and Control-Input transcriptome samples was also performed, but yielded very few genes (not shown here). Specifically, only 25 genes were upregulated, and 52 genes were downregulated (p-value < 0.05), with no log2 fold change thresholds. No targets of interest stood out from this list of significant genes, and the list was also too small to perform further downstream analysis, such as pathway or network analysis. This is likely due to both a poor read depth in these

samples and a lower sample size. As a result, no further in depth analysis of the transcriptome is discussed here.

Figure 5.6. Exploratory differential gene expression analysis of the translatome.



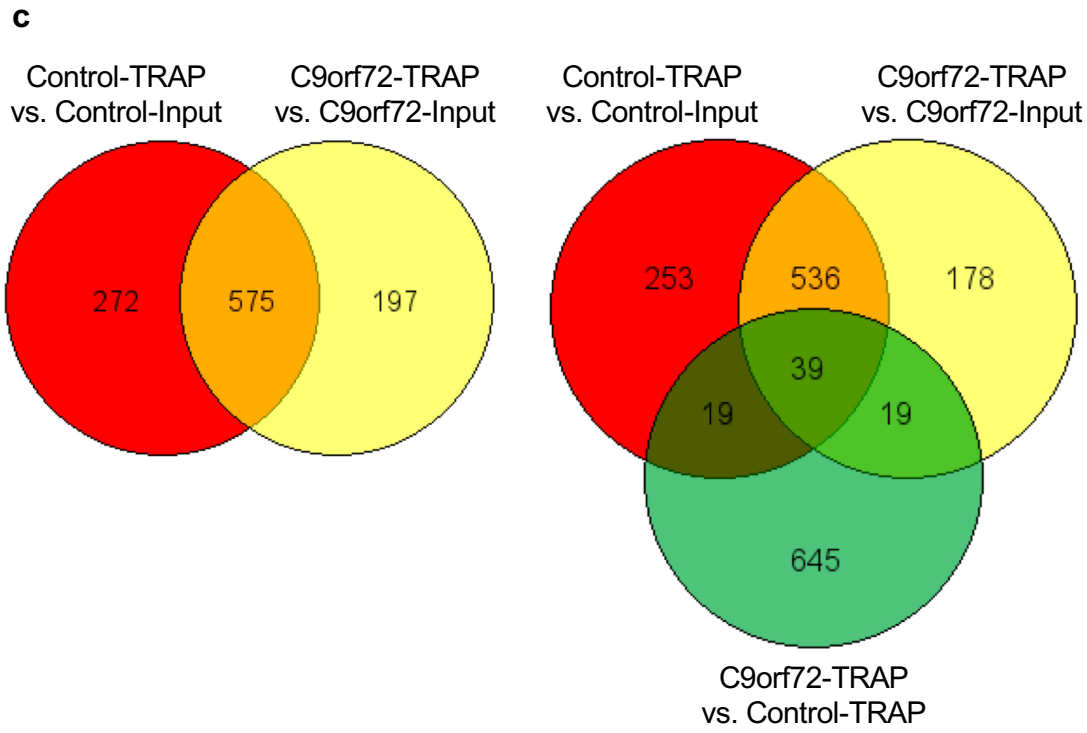


Figure 5.6. Exploratory differential gene expression analysis of the translatome.

(a) Principal component analysis of normalised translatome dataset, following variance stabilising transformation was plotted, resolving samples by TRAP experiment and phenotype (C9orf72 vs. control) by the first and second principal components, respectively.

(b) MA plots of normalised gene expression versus the log2 fold change of C9orf72-TRAP vs. Control-TRAP (i), C9orf72-TRAP vs. C9orf72-Input (ii) and Control-TRAP vs. Control-Input.

(c) All significant (padj < 0.05) differentially expressed genes were used to examine the similarity of three interesting group-wise comparisons. TRAP vs. input sample comparisons have a significant gene enrichment crossover (i). C9orf72-TRAP and Control-TRAP differential expression contrast shows a large number of unique differentially expressed genes (ii).

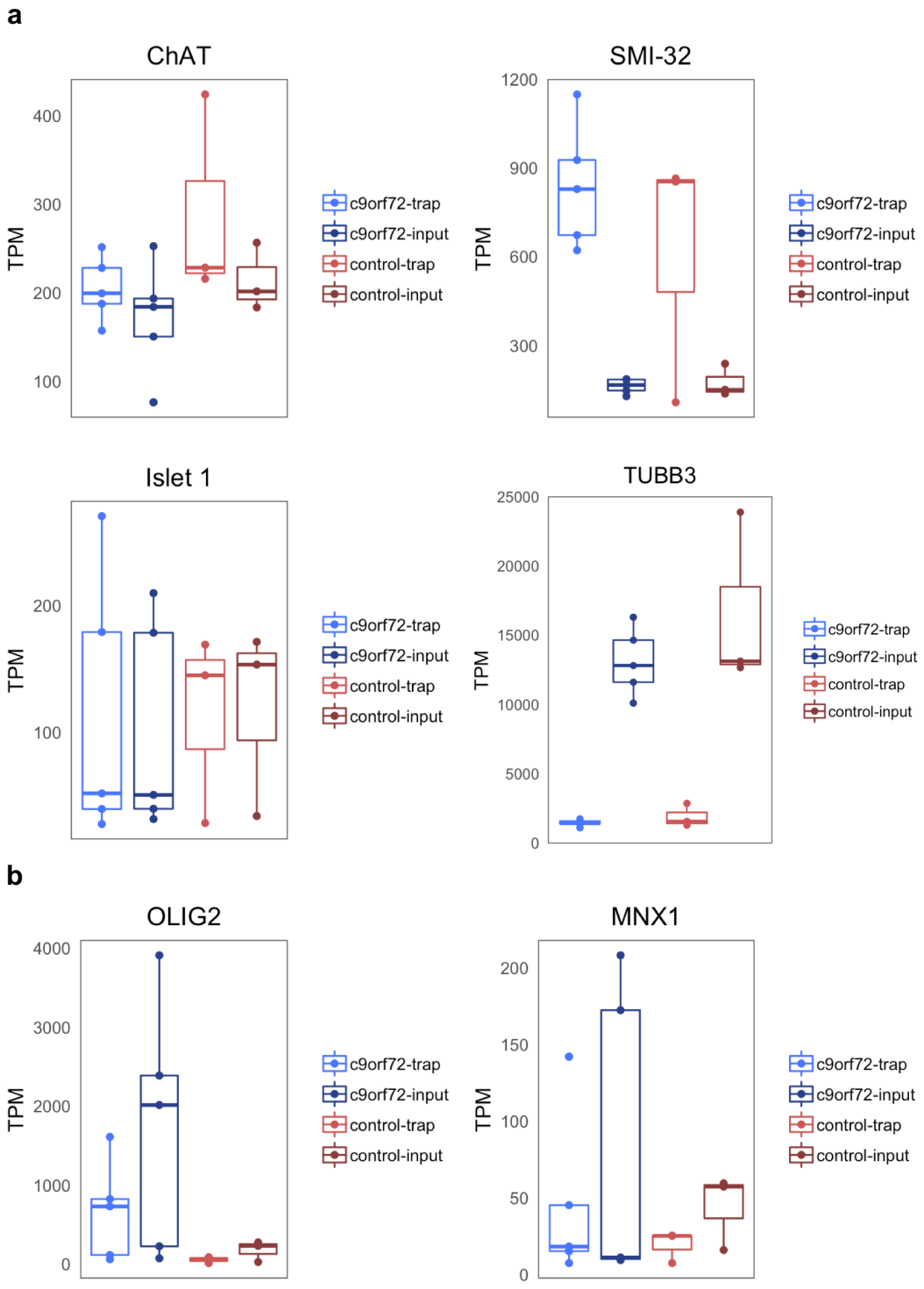
5.4.5 TRAP-RNA is enriched for neuronal markers in the motor neuron translatome

With the result that a large number of genes are differentially expressed between motor neurons with the *C9orf72* mutation and age-matched controls, further analysis to examine the expression of genes of interest in motor neurons (Figure 5.7a-b) and ALS was done (Figure 5.7c). *ChAT* transcripts are significantly (padj < 0.032) enriched in *C9orf72*-TRAP samples compared with transcriptome input fractions (Figure 5.7a). Control-TRAP sample enrichment for *ChAT* were not statistically significant, though the overall pattern suggests enrichment compared with transcriptome inputs (Figure 5.7a). There was no difference in expression for *ChAT* between the TRAP sample groups. Another marker for mature neurons, *NEFH* (SMI-32) had increased expression in TRAP samples compared with transcriptome input samples as well (Figure 5.7a). Islet 1 expression was generally the same across all sample types. Unexpectedly, *TUBB3* was significantly upregulated in transcriptome input samples compared to TRAP samples, suggesting a culture-wide expression, rather than being specific to motor neurons (Figure 5.7a). *OLIG2*, a crucial marker during motor neuron differentiation is expected to be enriched in progenitors and downregulated in mature motor neurons. In this dataset, there is no difference in *OLIG2* expression between TRAP and transcriptome input samples overall, but a sample pairwise pattern is suggested (Figure 5.7b). However, the marker is upregulated in *C9orf72*-TRAP samples compared with age-matched controls (padj < 0.048). Finally, the motor neuron marker *MNX1* has been shown to express transiently in the iPSC-motor neuron cultures grown in this lab, downregulating towards the final days of the differentiation. Thus, a downregulation of *MNX1* was expected in TRAP samples.

Although this result is not statistically significant, the pattern of TPM counts in the sample groups does suggest this (Figure 5.7b). Examination of genes found to be strongly linked to ALS, though not with *C9orf72* pathology, such as *FUS* and *OPTN*, were not statistically significant, though appeared to be enriched in the motor neuron population (Figure 5.7c.). The number of total *C9orf72* transcripts appeared to be slightly decreased in *C9orf72*-TRAP samples compared with the other groups. RNA-binding protein, TDP-43 protein has been implicated in over 90% of ALS cases, familial or sporadic. Transcript levels in iPSC-motor neurons from this patient do not appear to differ in TRAP-enriched or total pools of RNA. There is also no difference between the patient or control (Figure 5.7c). This is consistent with previous results found for TDP-43 expression using immunoblotting for this patient line. *FUS* and *OPTN* proteins appeared to be somewhat enriched in TRAP samples, suggesting mature motor neuron-specific activity, though did not have any significant difference between *C9orf72* and controls. The patient who donated the *C9orf72*-cell line used in this experiment was diagnosed with both ALS and FTD. Therefore, genes linked to FTD were also examined. Ubiquilin 2 (*UBQLN2*) which, along with *C9orf72*, has been implicated as a gene involved in both ALS and FTD, was found to be downregulated overall in TRAP samples ($padj < 0.009$), though no difference between the *C9orf72*- and control-TRAP samples could be seen (Figure 5.7c). A striking feature of this result was the consistency of normalised TPM counts within the samples. The translatome is predictably more stable than a transcriptome profile as it has fewer points of regulation. This was thought to cause fewer issues with reproducibility, especially in an iPSC differentiation model, which is prone to wide experimental variation. However, a wide range of expression was seen within sample groups despite this. Expression of genes thought to be solely involved in

the manifestation of FTD: Valosin-containing protein (*VCP*), charged multivesicular body protein 2B (*CHMP2B*), microtubule-associated protein tau (*MAPT*) and granulin (*GRN*) were examined. None showed any differential expression between *C9orf72* and control TRAP samples.

Figure 5.7. Normalised gene expression transcripts per million (TPM) counts of marker genes from DE analysis.



c

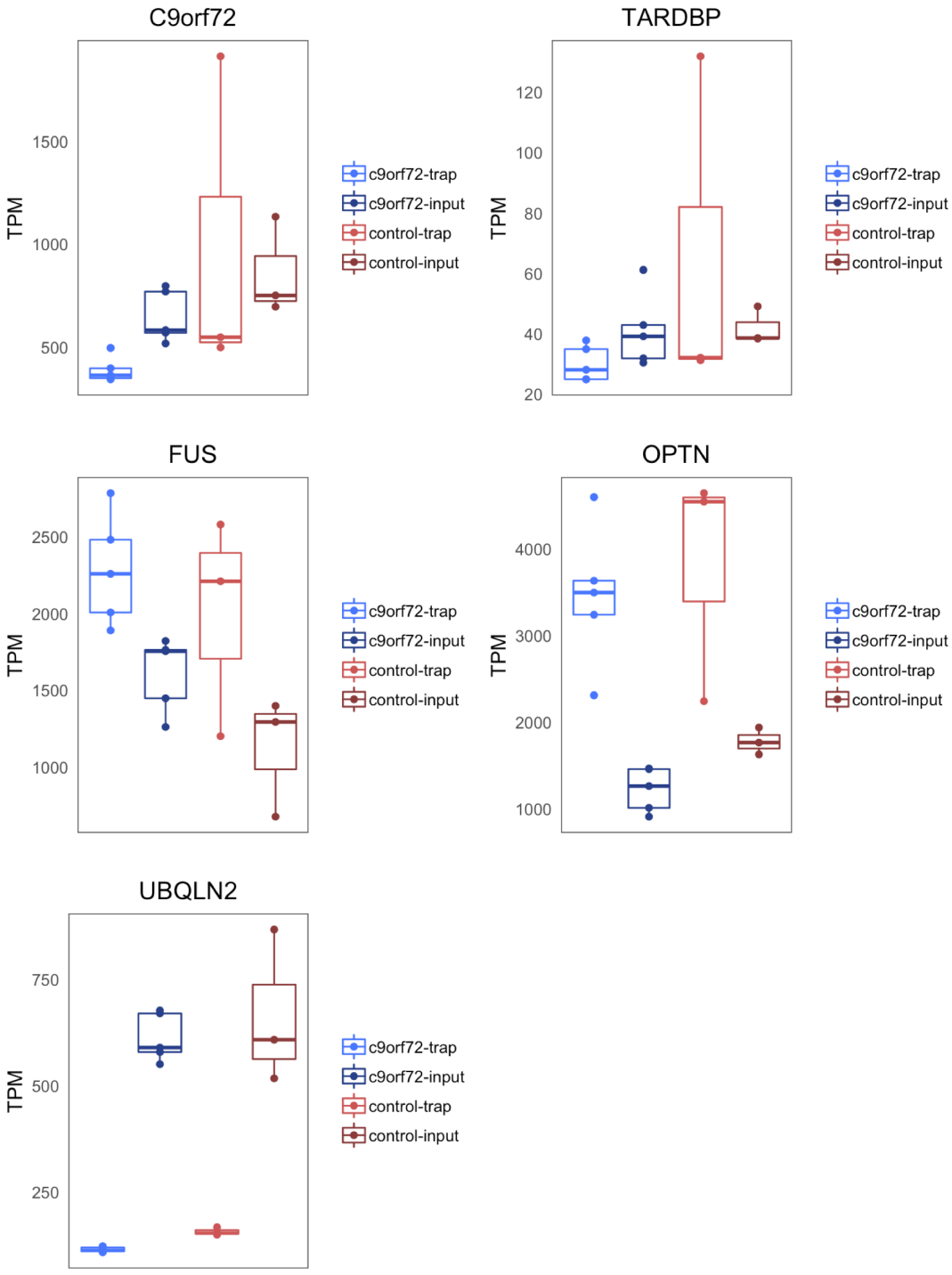

C9orf72
TPM
1500
1000
500
c9orf72-trap
c9orf72-input
control-trap
control-input
TARDBP
TPM
120
100
80
60
40
20
c9orf72-trap
c9orf72-input
control-trap
control-input
FUS
TPM
2500
2000
1500
1000
c9orf72-trap
c9orf72-input
control-trap
control-input
OPTN
TPM
4000
3000
2000
1000
c9orf72-trap
c9orf72-input
control-trap
control-input
UBQLN2
TPM
750
500
250
c9orf72-trap
c9orf72-input
control-trap
control-input

Figure 5.7. Normalised gene expression transcripts per million (TPM) counts of marker genes from DE analysis.

(a) Key mature motor neuron markers. *ChAT* is enriched in C9orf72-TRAP, compared with *C9ORF72* (padj < 0.032). *TUBB3* (beta-III tubulin) is significantly enriched in input samples (padj < 2.37E-05).
(b) Differentiation markers. *OLIG2* was differentially expressed (padj < 0.048) in C9orf72-TRAP compared to Control-TRAP samples.
(c) ALS-linked genes across sample groups. Ubiquilin 2 (*UBQLN2*) is depleted in TRAP samples (padj < 0.009). Other comparisons are not statistically significant, but are shown to visualise the patterns of gene expression within the data.

The Wald Test was used to test for significant differences in expression in a normalised counts table from *DESeq2*.

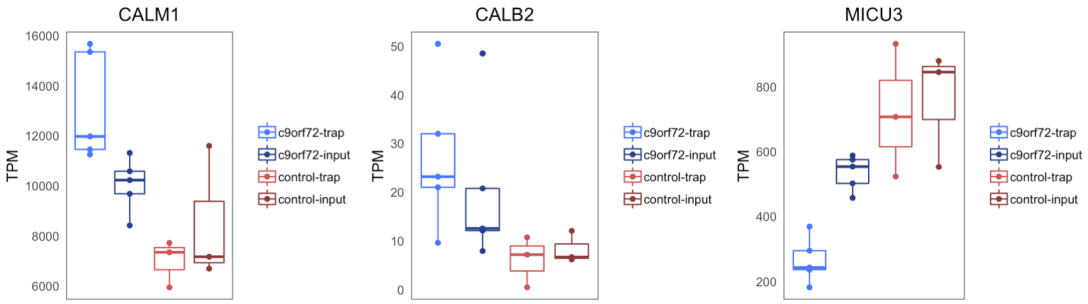
5.4.6 *C9orf72* patient motor neurons have several differentially expressed genes compared with age-matched controls

With minimal thresholds, 722 genes were significantly (padj < 0.05) differentially expressed between *C9orf72*-TRAP and control-TRAP samples. With a log2 fold threshold of < -1.5 or > 1.5 included in the Wald test, 71 (38 upregulated and 33 downregulated) genes were significantly differentially expressed. Close examination of the total misregulated gene families showed several interesting gene families to consider in further investigations. 30 members of the zinc finger family of proteins were misregulated. 7 members of the TBC1 domain family of proteins were upregulated. Genes associated with synaptic activity, syntaxin-6 (*STX6*), syntaxin-8 (*STX8*), synaptotagmin-11 (*SYT11*) and synaptotagmin-9 (*SYT9*) were upregulated. Several solute carrier genes, including *SLC17A6* and *SLC17A7*, which are involved the trafficking of glutamate-loaded vesicles, were downregulated. A significant misregulation of ribosomal and mitochondrial ribosomal genes was seen. Whether this is linked to a translation defect or to the exogenous expression of RPL22:c-myc:FLAG is not clear. Several different members of this family of genes exist, and it is well known that ribosomal proteins regulate the expression of other members. The Rab family of GTPases have previously been implicated in neurodegenerative diseases, including the downregulation of Rab7L1 leading to defects in exosome release from patients with the *C9orf72* mutation (Aoki et al., 2017). While expression of this gene was not altered, 6 members of the family were misregulated. A gene with the most significant increased expression, appearing in all versions of the analysis was NPIPA8, a nuclear pore complex-interacting protein. Various genes for ion channel subunits were altered, including calcium and potassium (*CACNA1E*, *CACNA1I*,

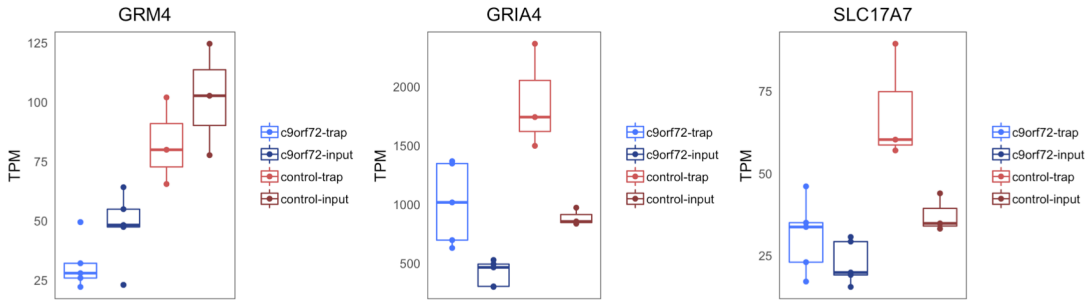
KCNA2, *KCNG1*, *KCNJ5*) and glutamate receptor (*GRIA1*, *GRIA4* and *GRM4*). Genes linked to calcium homeostasis were seen, including calretinin 2 (*CALB2*), calmodulin 1 and 3 (*CALM1*, *CALM3*). Finally, a large number of transcription factors and RNA processing genes were differentially expressed in *C9orf72*-TRAP samples compared to controls. A list of the 50 most upregulated and 50 most downregulated DE genes can be found in the appendix (Table S8.1, S8.2.). With more stringent thresholds applied to the DE analysis (log2 fold change < -1.5 or > 1.5), genes that persist in DE include *SLC17A6*, *NPIPA8*, *ZNF558* and a number of transcription factors. The normalised expression of some of these genes of interest have been given in Figure 5.8. With these gene groups in mind, significant differentially expressed genes were used to search for enriched cellular pathways and specific molecular functions.

Figure 5.8. Normalised gene expression in transcripts per million (TPM) counts of genes of interest from DE analysis, grouped by cellular function.

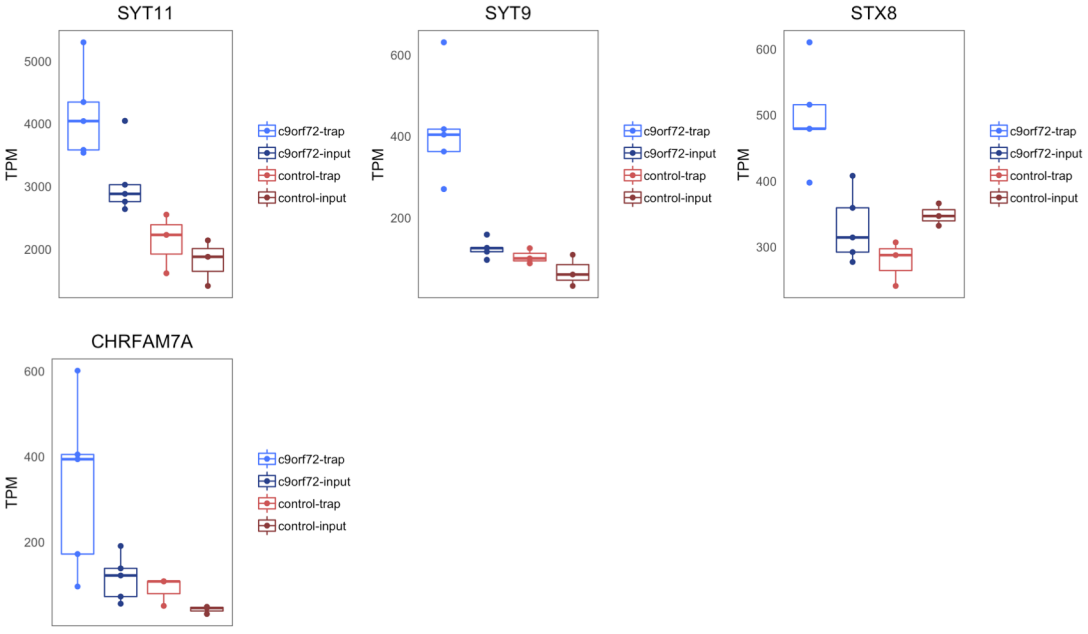
Calcium homeostasis



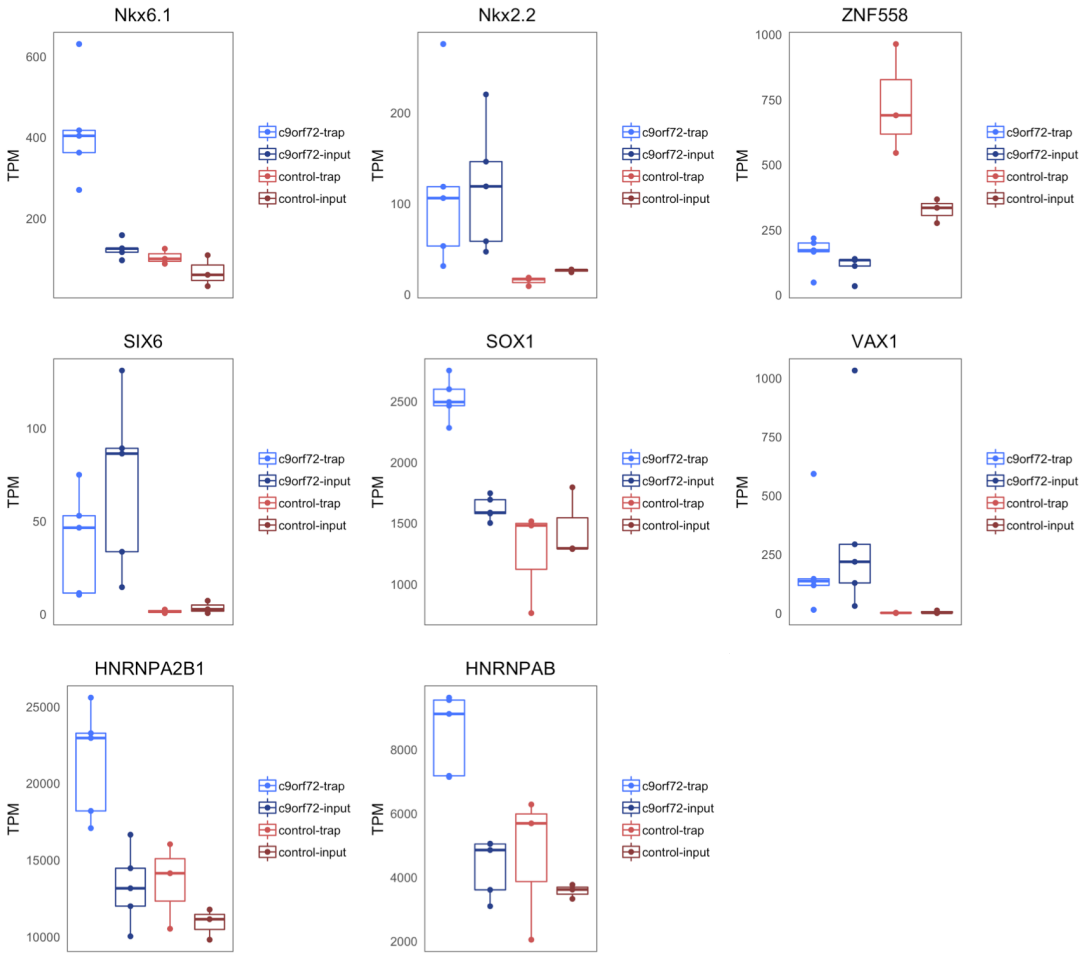
Glutamate receptors



Synaptic activity



Transcription factors and RNA processing



Nuclear pore

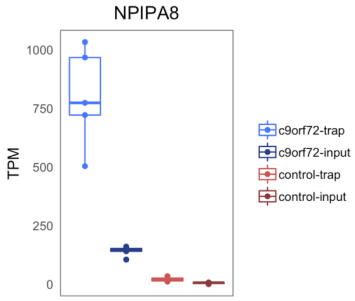


Figure 5.8. Normalised gene expression in transcripts per million (TPM) counts of significant genes of interest from DE analysis, grouped by cellular function. Padj < 0.05. Wald Test was used to test for significant differences in expression in a normalised counts table from *DESeq2*.

5.4.7 DNA-binding, transcriptional activity and intracellular vesicle-associated genes are significantly enriched in the *C9orf72* iPSC-motor neuron translatome

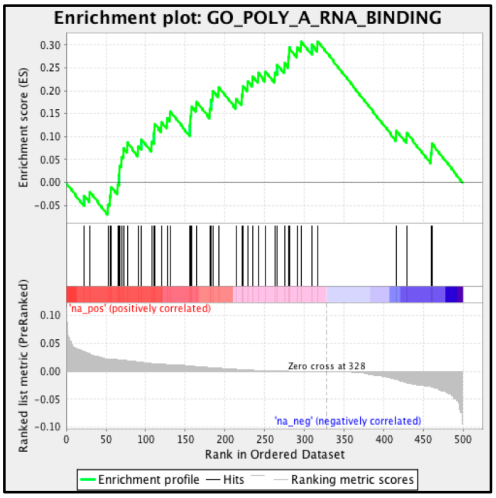
In order to extract more information about the most significant cellular pathways perturbed in motor neurons with the *C9orf72* mutation, the most significant genes from differential expression analysis were used in pathway analysis. The top 250 most significant DE genes were used to search for annotated pathways from gene ontology (GO) (Figure S8.3). Significant terms (FDR < 0.25) for molecular function (MF) obtained, mostly associated with transcription activity, comprising the various transcription factors found in analysis (Figure S8.3a). To get a clearer perspective on how the differential expression of the genes linked to these pathways are affecting cellular activity, fast gene set enrichment analysis (FGSEA) was performed. This analysis not only links genes to known pathways, but also considers its normalised expression and statistical significance (log2 fold change * padj). From this we can also retrieve information regarding the up- or downregulation of a particular pathway from subsequent leading edge analysis (Figure 5.9). MF GO terms suggest an increase in RNA and chromatin binding gene activity (Figure 5.9a). Pathway analysis for biological pathways (BP) suggested an accumulation of misregulated genes in pathways linked mitochondrial electron transport chain and unusually, forebrain brain development (Figure S8.3b). The latter is estimated to have arisen due to the upregulation of transcription factors such as *VAX1*, *SIX3* and *SOX1*, which have been implicated in the differentiation of retinal and forebrain neurons. FGSEA eliminated these terms and instead showed a significant increase in regulation of pathways for cell cycle and mitosis (Figure 5.9b). This is unexpected since the TRAP-RNA from iPSC-motor neurons was targeting mature neurons, which should have exited the cell cycle. This may

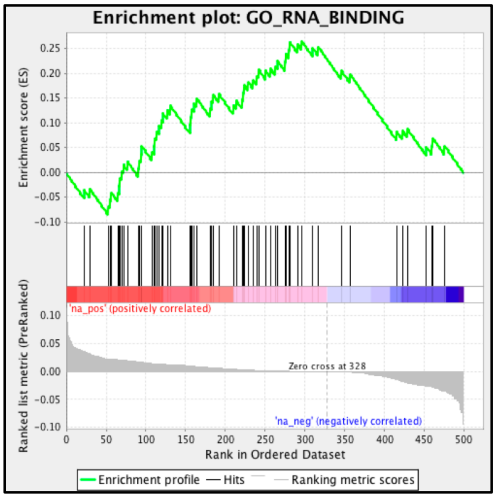
have implications for the true nature of our *C9orf72*-ChAT-expressing motor neurons in culture, though conclusions cannot be made without in-depth validation of these gene expression changes via immunostaining, qRT-PCR and immunoblotting. Inspection of genes enriched in specific cellular components (CC) (Figure S8.3c) suggested the misregulation of GO pathways localised at the mitochondrial membrane and other organelle membranes. There was also an enrichment for synaptic vesicle membrane, suggesting a role for genes *SLC17A6*, *SYT9* and *SYT11* (Figure S8.3c). FGSEA did not include any terms linked with synaptic activity, instead giving a number of terms linked to cytoskeleton and intracellular vesicles, the latter of which was downregulated (Figure 5.9c). A full list of genes associated with the GO terms derived from FGSEA can be found in Tables S8.3, S8.4 and S8.5.

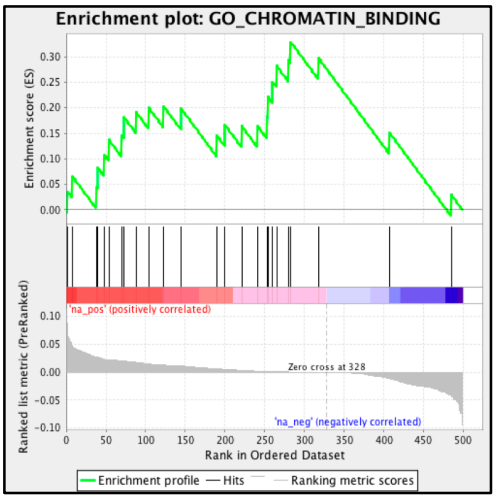
Figure 5.9. Fast gene set enrichment analysis (FGSEA) and leading edge analysis of the top 500 most significant genes in the C9orf72 translatome.

a

	Molecular Function GO Terms	# members	# members in signal	FDR q-value	FWER p-value
1	Poly A RNA binding	47	43	0	0.007
2	RNA binding	63	52	0	0.005
3	Chromatin binding	25	22	0.013	0.052

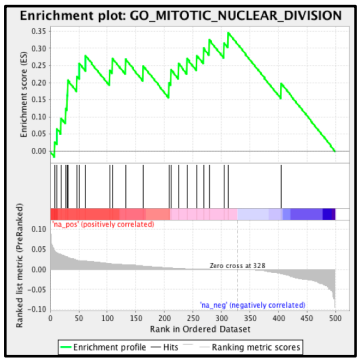


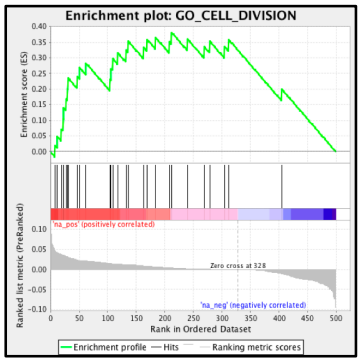


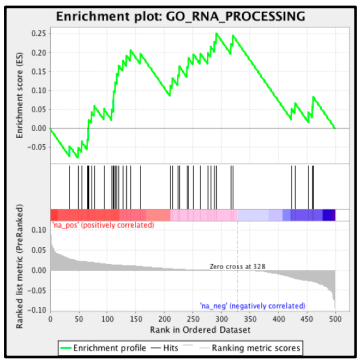


b

	Biological Processes GO Terms	# members	# members in signal	FDR q-value	FWER p-value
1	RNA processing	38	31	0.168	0.862
2	Organelle assembly	30	28	0.064	0.475
3	Cellular macromolecular complex assembly	29	26	0.195	0.938
4	DNA metabolic process	22	18	0.208	0.935
5	Cytoskeleton organization	36	23	0.069	0.139
6	Microtubule based process	25	24	0.063	0.272
7	Microtubule cytoskeleton organization	23	22	0.073	0.408
8	Cell division	28	22	0.073	0.073
9	Mitotic cell cycle	36	33	0.101	0.266
10	Cell cycle process	44	39	0.076	0.267
11	Cell cycle	52	45	0.066	0.334
12	Mitotic nuclear division	23	22	0.066	0.454
13	Chromosome organization	41	37	0.068	0.422
14	Peptidyl lysine modification	15	14	0.167	0.836
15	Peptidyl amino acid modification	23	23	0.242	0.978

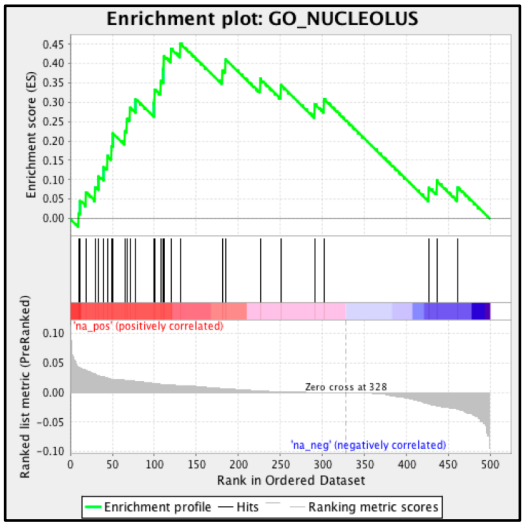






c

	Cellular Components GO Terms	# members	# members in signal	FDR q-value	FWER p-value
1	Ribonucleoprotein complex	40	33	0.159	0.938
2	Nucleolus	29	20	0	0
3	Transferase complex	15	10	0.023	0.207
4	Catalytic complex	26	15	0.025	0.175
5	Nucleoplasm part	25	21	0.079	0.641
6	Chromosome	36	27	0.092	0.727
7	Membrane region	32	25	0.094	0.141
8	Intracellular vesicle	29	23	0.145	0.111
9	Spindle	22	20	0.027	0.155
10	Cytoskeleton	71	46	0.005	0.019
11	Microtubule cytoskeleton	50	45	0.005	0.012
12	Cytoskeletal part	55	44	0.022	0.106
13	Microtubule organizing centre	35	31	0.022	0.179
14	Centrosome	31	27	0.059	0.49
15	Cilium	18	11	0.091	0.75
16	Microtubule	21	17	0.146	0.911
17	Cell projection part	32	26	0.189	0.973



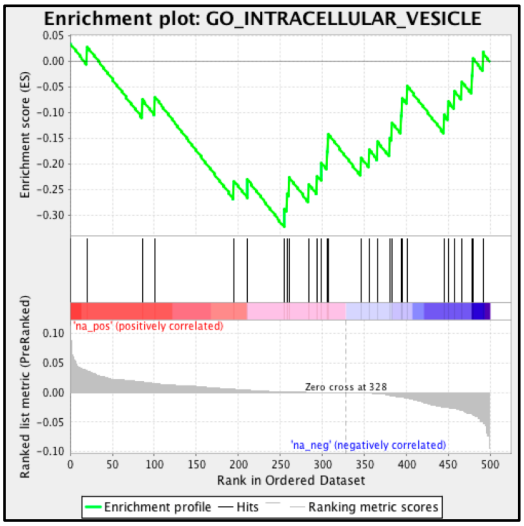


Figure 5.9. The top 500 most significant genes in the C9orf72 translatome (padj < 0.05) were used to explore the gene ontology (GO) terms using fast gene set enrichment analysis (FGSEA) and leading edge analysis. The enrichment of specific genes GO terms (FDR < 0.25) are given as enrichment plots. See Tables S8.3, S8.4 and S8.5 for full genes lists of associated genes. The colouring refers to how the expression of a gene has an effect on a particular pathway that it is involved in. Red = positive phenotype, blue = negative phenotype.

(a) Genes enriched to annotated known molecular functions (MF).
(b) Genes enriched to annotated known biological pathways (BP).
(c) Genes enriched to annotated known cellular components (CC).

5.5 Conclusions

Next generation sequencing of TRAP-RNA obtained from human iPSC-neuronal cultures has previously never been done, and thus is a unique approach to studying the perturbed cellular pathways in ALS. The analysis of this RNA sequencing data was expected to be challenging due to several potential sources of experimental variation which could result in biases in the data. In order to help remove some of this bias, the use of synthetic spike-in RNA was employed, and proved highly useful in data normalisation (Figure 5.4). Examination of the data at different points in the analysis showed that TRAP samples were enriched with mature RNA species (Figure 5.3), and *ChAT* (Figure 5.7). It was possible to estimate the relative consistency of RNA pulldown between samples (Figure 5.5), which was shown to be reasonable. However, the need to discard two age-matched controls due to questionable sample quality made this analysis more difficult to verify. Furthermore, despite the separate clustering of sample groups from principal component analysis (Figure 5.6a), some doubt may be cast on the close proximity of clustering of SFC840 controls to *C9orf72* patient-TRAP samples. Only the use of further controls, including CRISPR/Cas9-edited lines, can solve whether this is due to a lack of sample number, which considerably reduces the power of this experiment, or if at translatome level these iPSC-motor neurons are not very different without additional perturbation. Despite these observations, there are several differentially expressed genes between the *C9orf72*-TRAP and control-TRAP translatomes. A number of genes have been isolated, along with enriched functions from pathway analysis for future validation.

6 Discussion

ALS is late-onset neurodegenerative disease with a complex pathology and a diverse genetic and sporadic aetiology (Kiernan et al., 2011). Clinical diagnosis does not allow a simple stratification by genetics. No effective treatments have yet been approved, largely due to a lack of basic mechanistic understanding of the molecular pathways which drive disease. The most common mutation linked to ALS, the *C9orf72* HRE (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Gijselinck et al., 2012), overlaps with the clinical syndrome of FTD (Ling et al., 2013). In order to understand how early global gene changes in the affected cell type triggers these phenotypes, RNA sequencing must be performed in a human model. The increasing use of patient-derived iPSC-neuronal models to illustrate robust molecular phenotypes that have also been validated in mouse and human post-mortem tissue, makes this one of the best tools for research of a human disease available. However, significant experimental variance has led to the need for improved techniques for RNA isolation for sensitive high throughput techniques. This DPhil project aimed to generate and optimise a lentiviral TRAP vector to harvest mature motor neuron-enriched RNA to characterise translatome profiles from our *C9orf72* ALS human cultures.

6.1 A TRAP vector for human translatome analysis: main conclusions

6.1.1 Mature mRNA was successfully purified from iPSC-motor neurons

The aim to use molecular cloning to generate a TRAP vector, containing a ubiquitous large ribosomal protein, RPL22 (Bortoluzzi et al., 2001; Sanz et al., 2009; Lesiak et al., 2015), fused with a molecular tag and a bicistronic eGFP reporter

under a ChAT promoter was successfully achieved (Figure 3.1, 3.5, 3.6). Restriction digest reactions and Sanger sequencing of the vector showed that the expected insertion of the functional cDNA components into a lentiviral backbone was successful. Immunostaining of transduced iPSC-motor neurons with the TRAP vector showed the expected localisation of RPL22:c-myc:FLAG in the nucleolus and cytoplasm, with ubiquitous cellular expression of the eGFP reporter. This showed both the successful expression and targeting of the ribosomal protein to endogenous pathways for ribosome manufacture (Shu-Nu et al., 2000), and the cleavage of the CHYSEL site to separate eGFP from RPL22:c-myc:FLAG. Subsequently, the vector was used to optimise lentiviral transduction into iPSC-motor neurons, at a later stage of the differentiation protocol. Up to 100% co-expression of the eGFP reporter and ChAT-expressing motor neurons was achieved in age-matched control and patient motor neurons at day 28 of the protocol, with no significant cell death (Figure 4.3). However, the isogenic lines were more sensitive to transduction, and better expression of the TRAP vector was achieved at lower concentrations of lentivirus (Figure 4.3). Multiple steps of the protocol, originally adapted from Heiman et al., 2014 and Lesiak et al., 2015, were optimised to obtain RNA with a RIN value over 9, and with enough yields (40 ng) to perform bulk RNA sequencing, without any PCR amplification steps. This final protocol could not obtain enough RNA from the isogenic lines, forcing them to be omitted from the experiment. The final RNA sequencing experiment consisted of control and patient samples from five differentiations (Table 4.1). Future repetitions of this experiment should include these CRISPR/Cas9 controls, by increasing the total number of cells used for RNA extraction.

Along with the TRAP RNA samples, the same amount of the respective transcriptome input samples were submitted for library preparation and sequencing. The sequencing facility performed a ribodepletion protocol on the samples followed by another step of quality control and RNA quantification. The facility reported back that the TRAP samples appeared not to have been depleted of any significant amounts of rRNA, while transcriptome input samples were depleted (data not shown). Total bulk RNA preparations from cells consists of a mixture of rRNA, transfer RNA (tRNA), non-coding RNA species with regulatory roles and a small percentage of protein-coding mRNA. Of this pool, rRNA comprises around 80% of total RNA. Ribodepletion or polyA tail selection is performed to exclude rRNA which would otherwise reduce the sequencing depth of protein coding mRNA, which is of greater interest for a transcriptomics experiment. The lack of depletion in the TRAP samples showed that selection as a result of ribosomal pulldown had occurred, giving proof of the success of the experiment. Furthermore, examination of the various species of RNA detected in the total raw reads showed a significant depletion of rRNA and intergenic reads in TRAP samples, while enriched for protein coding, 3'-UTR sequences and reads mapping to exogenous TRAP vector transcripts (Figure 5.3). The two control TRAP samples which appeared to resemble their transcriptome inputs were excluded from differential expression analysis; no reason could be found for this occurrence. The reproducibility of TRAP from a neuronal differentiation model, was confirmed using correlation plots of normalised gene counts between individual samples (Figure 5.5). The consistency between housekeeping genes and *C9orf72* transcript pulldown shows the specificity of the pulldown with each experiment. This was particularly striking with control samples where pulldown was most consistent between samples from the

same control line, compared with a line from a different individual. This stresses the importance of using CRISPR/Cas9 edited lines as a true control of identifying mutation-specific changes in the translatome. Similar plots between TRAP and transcriptome input samples were done to examine enrichment of markers, though this was only partially seen.

Collectively these data gave strong evidence to support the success of purifying mature mRNA from iPSC-motor neuron cultures with mixed background cells compared with total transcriptome input RNA from the same cultures.

6.1.2 Comparison of TRAP and transcriptome profiles is challenging

Differential expression analysis, comparing TRAP and transcriptome input samples was performed to investigate the enrichment of specific motor neuron markers. The accuracy of the RNA sequencing experiment was assessed by the use of internal synthetic spike-ins (Hardwick et al., 2016). The apparent bias, as demonstrated by a lack of correlation between the expected and observed gene fold change in for the *sequins*, was somewhat corrected for (Figure 5.4) (Wong et al., 2017). Overall, the generalised linear model (GLM) constructed using DESeq2 (Love et al., 2014), accounted for the *sequins* correction, lane distribution of the samples and samples type (i.e. *C9orf72*-TRAP, control-TRAP, *C9orf72*-input or control-input). A GLM is modelled on a binomial distribution, which is thought to best represent the distribution of gene counts and gene-level variants which may be detected in a given RNA sequencing experiment (Love et al., 2014). This model takes an average of the number of gene counts per sample in a group and adjusts the fit of the model based on variance or dispersion factors. This was an imperfect model due to the

missing criteria of differentiation date, as being a potential source of experimental variation. This was chosen to be omitted from the final analysis to avoid over-correcting the data, where lane and *sequins* correction were already crucial. Altering the fit of a GLM with too many factors is often likely to adjust the normalised count estimates for genes, which may not give a true result, and alter the significance of a comparative test. Arguably, the differentiation factor, as a source of variance is being corrected for by the TRAP method itself, and thus is less consequential than *sequins*, which can be visibly seen to have disparity within the data. Further, the arrangement of samples within the lanes was not ideal, as some samples from the same group were not well distributed between different experimental lanes. This may cause a bias in the data where the comparison of sample groups which were processed in separate flow cell reactions directly causes differential expression, which cannot be detected unless varied. This should be altered in all future RNA sequencing plans. However, the correction of lanes was crucial to account for the differences in sequencing depth seen between TRAP and transcriptome input samples. As described, due to different results from the ribodepletion library preparation, the RNA pools sequenced for each sample type were distinctly different. The sequencing of the transcriptome input fractions was crucial for this experiment. As a novel technique, TRAP of RNA from iPSC-motor neurons needed to establish that mature mRNA were enriched from total RNA. Using differential expression analysis, it could be shown that ChAT and SMI-32 markers were enriched in patient and control TRAP samples (Figure 5.7). However, the enrichment of other markers, *TUBB3*, *MNX1* and *ISL1* were not seen. A comprehensive study of the markers expressed in this model, using immunocytochemistry, shows that Islet 1 expression does not always co-express

with ChAT-positive neurons (Dafinca et al., 2016). This suggests a reason for the broad range of counts detected for Islet 1 in different samples, including the equivalent transcriptome inputs (Figure 5.7). Though *MNX1* (HB9) is a robust marker for motor neurons, our cultures seem to downregulate the gene by day 28 of the differentiation, peaking at around day 24 (unpublished). TUBB3 is depleted in the TRAP samples, suggesting a high expression of neurons in our culture without motor neuron identity, as a pan-neuronal marker. Additionally, *OLIG2* is expected to be downregulated in mature motor neurons, and this is not significant on average, but potentially is significant if compared pairwise across samples (Figure 5.7). Finally, differences seen from the analysis of the percentage of reads mapping to various transcriptomic RNAs (Figure 5.2, 5.3), % GC content and duplication rate, suggest an increased library complexity of the transcriptome input samples. These patterns in the data may also support the hypothesis that transcriptome input samples are not the ideal control for quantifying enrichment. A better comparison may be obtained by purifying TRAP RNA following transduction with the hUbC:TRAP vector. This would allow sequencing of the translatome of all the cells in culture, giving a control at the point of library preparation and negating the need for immense sequencing depth correction during analysis. Transcriptome input RNA from total brain homogenate tissue in mouse TRAP experiments have previously been performed, though how these studies corrected for these differences was not described (Doyle et al., 2008). The enrichment of *ChAT* and RPL22:c-myc:FLAG transcripts were validated prior to sequencing (Figure 4.8). Further validation of significant hits has not been done in the course of this DPhil project, but should be carried out for future studies to confirm the accuracy of this analysis. Quantitative-RT-PCR should be used for this.

6.2 DE analysis between *C9orf72* patient and control translatomes reveal disease-relevant targets

6.2.1 Genes associated with synaptic transmission are misregulated

Examination of the 722 differentially expressed genes between the *C9orf72*-TRAP and control-TRAP samples revealed a number of hits with potential links to pathways which have been linked to *C9orf72* ALS, and should be validated *in vitro*. In particular, misregulated genes associated with calcium modulation and homeostasis, expression of glutamate receptor subunits and synaptic transmission were recognised as points for further examination (Figure 5.8). Calcium concentration between intra- and extracellular regions has a ratio of around 1:1000 to 1:10,000. This gradient must be cautiously maintained to ensure the ability of a cell to respond to transient calcium ion signals (Kirberger et al., 2017), regulated by several proteins, including the calmodulin and calretinin family of calcium binding proteins (Kirberger et al., 2017). In neurons these influxes of calcium from the extracellular space are often triggered by action potentials or ion channel receptor agonists. The influx of calcium activates signal transduction pathways, where the binding of free calcium ions to proteins, such as calmodulin causes this protein to alter its conformation and allow its binding to protein kinases, like calmodulin-dependant protein kinases (CaMK) (Kirberger et al., 2017). CaMKs are then activated, and via phosphorylation reactions alters the activity of further downstream proteins, eventually resulting in modulatory effects on transcription factors, which may alter gene expression in response to the original trigger (Corcoran and Means, 2001). Therefore, the maintenance of the required levels of calmodulin is crucial to maintaining this balance in a highly optimised transduction

pathway (Kool et al., 2019). *CALM1* and *CALM3* have increased expression in *C9orf72*-TRAP samples. This may suggest that during the receipt of a Ca^{2+} signal, an increased amount of CaMK proteins are activated, potentially triggering an exaggerated inhibition or activation of downstream pathways, such as neurotransmitter release at the synapse (Colbran and Brown, 2004). Calmodulin and calretinin (*CALB2*) (also upregulated in translatome data) are well known to localise at the synapse, with roles in modulating NMDA receptors and calcium-gated voltage channels, respectively (Camp and Wijesinghe, 2009; Iacobucci and Popescu, 2017). A study has shown that calretinin is downregulated in a *SOD1*^{G93A} model of ALS, opposing this result (Clark et al., 2017). However, a recent paper has also shown that although similar pathways may be misregulated due to different ALS mutations (Bursch et al., 2019), the specific targets affected vary considerably, potentially making this finding significant as *C9orf72*-specific defect. Finally, the downregulation of MICU3, a mitochondrial calcium buffering uniporter was decreased in expression, suggesting a deficit in removal of free calcium from the cytoplasm, consistent with calcium imaging studies in *C9orf72* ALS patient iPSC-motor neurons, by this group (Dafinca et al., 2016). Furthermore, differential expression of other MICU proteins were not seen. MICU3 is known to be a neuronal tissue-specific protein, substantiating the motor neuron specificity using TRAP (Patron et al., 2019).

The expression of glutamate receptor subunit transcript, *GRIA4*, coding for a glutamate receptor AMPA (GluA4) receptor subunit was significantly reduced along with glutamate metabotropic receptor 4 (*GRM4*) (Figure 5.8). These genes play a role in synaptic transmission, suggesting that these neurons may present a reduced

capacity for signal transduction via these receptors. This was also accompanied by a downregulation of glutamate vesicle transporters. Expression of various glutamate receptor subunits in *C9orf72* ALS iPSC-motor neurons has previously been characterised (Selvaraj et al., 2018). This study found different glutamate subunit transcripts to be upregulated in their cultures, hypothesising a mechanism for glutamate excitotoxicity. Though contradictory, it is not straightforward to compare these studies as they use different protocols for differentiation and harvest RNA at different stages. These results may represent the RNA profile of motor neurons at different relative ages, potentially suggesting a significant shift in receptor expression during disease progression. This requires further characterisation of alterations in glutamate receptor expression in our model. Increased expression of genes, *SYT9* and *SYT11*, *STX8* and dupαt (*CHRFAM7A*), with direct roles at the synapse is seen (Figure 5.8) (Shin et al., 2004; Wang et al., 2015). *CHRFAM7A* encodes an acetylcholine receptor (AChR) subunit, which interacts with the receptor at the synaptic membrane, causing inhibition of function, as a mechanism for functional regulation (Sinkus et al., 2016). This increase in dupαt could be construed as a mechanism for reduced synaptic transmission of these motor neurons.

The synaptotagmin family of proteins have a role in synaptic vesicular trafficking, co-localising to the SNARE complex (Adolfsen and Littleton, 2001). They can be grouped as being modulated by calcium or not. Syt-11 is thought to have a link to Parkinson's disease, as an interactor of *parkin* (Wang et al., 2018). It has also been found as an inhibitor of endocytosis in neurons, by interacting with clathrin-mediated vesicles (Wang et al., 2015). This increase in *SYT11* was also seen in other RNA

sequencing data from this lab; from bulk and FACS-sorted *C9orf72* ALS iPSC-motor neurons and from bulk spinal muscular atrophy data (unpublished). Unlike Syt-11, Syt-9 is mediated by Ca^{2+} (Shin et al., 2004), suggesting its differential expression may be linked to the calcium phenotypes already seen for this model (Dafinca et al., 2016). Finally, *STX8*, another player in SNARE complex dynamics was upregulated in *C9orf72*-TRAP samples. A role for this protein has been implicated in cystic fibrosis, though not neurodegeneration, with a function in early endosomal trafficking (Sabirzhanova et al., 2018).

As some of the most significant hits found from the differential expression analysis between *C9orf72*-TRAP and control-TRAP translatomes, the above genes are suggested as targets for *in vitro* validation by qRT-PCR in TRAP samples from the C90202 and C90203 patient lines that were used in this project. Examination in other *C9orf72* ALS iPSC lines used by this group is also suggested, prior to performing further high throughput analysis. Further discussion of biological validation studies for this project are given in section 6.5.

6.2.2 Discrepancies with other *C9orf72* ALS research

The existence of a vast amount of prior research in the *C9orf72* field provided a good starting point for looking for significant hits within the final translatome data. However, differentially expressed targets matching patterns seen in the literature for *C9orf72* RNA sequencing data were mostly lacking. No differentially expressed targets consistent with nucleocytoplasmic transport defects, such as RanGAP1, or other nuclear pore complex members from other studies were found. The only protein with a link was *NPIPA8*, but no literature regarding the function of this gene

exists. A slight decrease in the expression of total *C9orf72* transcripts was observed, consistent with prior work, though this was not significant (DeJesus-Hernandez et al., 2011). Examination of the Rab family of GTPases showed a misregulation for 6 transcripts, though none were the subtypes previously implicated in ALS (Chua and Tang, 2015; Aoki et al., 2017).

One of the most significant GO terms revealed through fast gene set enrichment analysis was *RNA binding* (Figure 5.9). This is correlated with the large number of transcription factors that were differentially expressed (Table S8.3). A subset are given in Figure 5.8. This is expected in *C9orf72* ALS, where the non-specific sequestration of RNA-binding proteins, such as hnRNPA, are likely to have downstream effects on the expression their various targets (D'Ambrogio et al., 2009; Mori et al., 2013b; Cooper-Knock et al., 2014; Romano et al., 2014). Two unusual transcription factors seen were, SOX1 and VAX1, which have roles in early development of cortical neurons and specification of forebrain neurons, respectively (Hallonet et al., 1999; Elkouris et al., 2012). The presence of this factor in our patient cultures should be examined. The expression of lineage-determining transcription factors in differentiating cellular cultures may not be as strictly regulated as in developing tissue, but their higher expression in patient neurons may suggest a slight difference in age or rate of differentiation compared with control motor neurons.

Some GO terms, like *Spindle* and *Mitotic cell cycle* suggest that these patient ChAT-positive motor neurons are not fully developed. Further optimisation of our motor neuron differentiation protocol to obtain longer-living neurons may improve this

feature. The presence of these GO terms may also suggest an inherent stress in the patient cultures, such as re-entry into cell cycle for apoptosis, which stimulates the expression of genes involved in such processes. This is a cell culture-specific issue, but it is overwhelmingly significant compared with other GO terms, which still appear to be too subtle, but may be more crucial for elucidating the underlying *C9orf72* ALS cellular translatome.

6.3 iPSC-motor neurons are a challenging model for transcriptomic analysis

The differentiation protocol optimised by the Oxford group, relies on monolayer expansion of neuronal progenitors, induced to this lineage from iPSCs and finally passaged with the addition of retinoic acid, SAG and DAPT-enriched growth media to induce the emergence of mature motor neurons by 25-30 days of culture. The iPSCs are derived from different individuals including, ALS patients, age-matched control donors and CRISPR/Cas9-edited lines with the *C9orf72* HRE removed. Each of these lines have been observed to differentiate with modest variation resulting in different rates of proliferation at iPSC and early neuronal progenitor stages. This includes the quality between passages and potential somatic mutations arising from overuse of a batch. The TRAP vector method was developed as a means of removing this variation and to control for differences in transient gene expression, which is negated by a translatome profile, rather than total gene expression, which is likely to be more unstable. The translatome only captures actively translating ribosomes, with some contamination from inactive ribosomes, which may sequestered at stress granules or P-bodies. On the other hand, the transcriptome has several more points for variation, including active transcription, similarly sequestered RNAs, small non-coding RNAs and rRNAs. This hypothetical

reduced complexity is somewhat captured in our translatome, as can be seen from the PCA plot showing clear distinctions between sample groups, but still shows variation between samples harvest on different dates and cultures.

6.3.1 Increased sample sizes and better controls would improve power

This translatome data is unlikely be strong enough to use for comprehensive analysis of the *C9orf72* ALS RNA profile. The lack of CRISPR/Cas9 controls and more patient lines reduces the power of a group-wise comparison in DE analysis. Repetition with more patient lines and isogenic controls is necessary. The use of age-matched controls is no longer justified as genetic background is likely to play a significant role in affecting the culture and translatome. This is immediately seen from principal component analysis of our data where the control sample from SFC180 is closer in overall profile to the transcriptome input fractions than the SFC840-TRAP samples (Figure 5.6). This is especially striking as there is a clear grouping of the samples based on the effect of the TRAP experiment itself in PC1, followed by genotype in PC2. The power of this RNA sequencing experiment was also reduced by the need to discard two control samples. Far less variability is seen in TRAP experiments performed in mice, where littermates are almost clones of each other, controlling for variability in CNS development, and later experimental variance (Doyle et al., 2008; Sun et al., 2015). An overall low success rate was achieved in obtaining enough, high quality RNA from the final TRAP protocol in this project (Table 4.1). Additional optimisation can be performed to improve this.

6.3.2 Improvements to the protocol may improve RNA yields

In accordance with results observed from this project, the following to the alterations to the TRAP protocol are suggested. Firstly, doubling the total number of cells for TRAP pulldown is recommended, in addition to improving the rate of successful lentiviral transduction. Generating stable iPSC lines with the TRAP construct genetically engineered into the host genome would improve consistently, but greatly reduces the versatility of the technique. Secondly, optimising the ribosomal protein best suited for TRAP harvests from human motor neurons could be done. Screens examining the expression of RPL22 as the ideal pulldown tether from polysomes have not been confirmed, rather estimated as a good candidate based on previous studies (Lesiak et al., 2015). Here, the localisation of RPL22 in expected cellular compartments is seen, though polysome isolation via sucrose gradient centrifugation and subsequent staining of RPL22:c-myc:FLAG was not done due to insufficient yields from cultures, giving no absorbance signals for polysomes (data not shown) (Sanz et al., 2009; Faye et al., 2014). Repetition of experiment, possibly in HEK293-T cells, using hUbC:TRAP lentivirus, could be done to confirm incorporation of tagged ribosomal protein into polysomes.

6.4 TRAP has wide potential for other human models and co-cultures

The motor neurons grown in this lab do not show any differences in ability to differentiate between genotype (Dafinca et al., 2016). Without additional stress applied, they do not show denervation or branching defect phenotypes, however they can only be kept in culture for up to 30 days. Improving the longevity of the cultures may allow us to characterise such phenotypes better. Human iPSC differentiation models exist where authors have successfully shown that it is

possible to age a culture over time with differences in control and patients (Fujimori et al., 2018). Time course experiments using TRAP at pre- and post-defect stages may provide clearer insight into the translatome changes. Furthermore, evidence of glia being an important contributing factor in ALS is becoming more apparent (Meyer et al., 2014; Lall and Baloh, 2017; Varcianna et al., 2019). Using co-cultures with glia and motor neurons containing an ALS mutation, and performing TRAP in parallel could be a powerful way of determining the order of molecular pathway changes which leads to specific cellular phenotypes.

6.4.1 TRAP as a tool in drug screening

Determining the subtle differences between patient and control motor neurons in culture is quite challenging, especially due to similarities in basal translatome profiles. Treatment of cultures with drugs could be highly useful in validating short term perturbations to patient iPSC-motor neurons. Cell culture treatments are much more likely to have a dramatic effect on the translatome, which responds immediately to environmental stimuli. Using ASOs, targeting the *C9orf72* transcript, would be a highly useful tool in combination with TRAP, to validate whether the knockdown of HRE-containing transcripts would be a viable method for *C9orf72* ALS treatment. TRAP could be used to examine how quickly the translatome is altered by downstream effects of *C9orf72*-knockdown, and whether this is mirrored in the translatome profile of isogenic controls. This can be done over a time course following treatment to examine how long drug effects are sustained. In such studies, keeping the consistency of experimental conditions is crucial to validation with RNA sequencing.

6.5 Future directions

Future work stemming from this project should initially focus on replicating the RNA sequencing experiment, with the suggested modifications and validation to generate a more powerful dataset. More in depth analysis of translatome data should include examination of alternatively spliced transcript variants in *C9orf72* patients. This will provide a better platform for comparison with other publicly available data concerning motor neuron gene expression. It is also of strong interest to perform proteomic data analysis of iPSC-motor neurons in order to correlate differences with bulk RNA sequencing data readily available in our group, and with TRAP translatome data. These datasets together will provide a strong method of observing differences in gene expression from immature to mature RNA transcripts and at protein level, allowing us to make estimates about regulatory pathways that are crucial in *C9orf72* ALS. This may also provide clues about the best kind of pathways to target in a potential therapy, with our human motor neurons as a powerful tool for validation. Further biological validation of the hits found in the translatome dataset may be performed in our iPSC-motor neurons. Several of the targets have known functions in calcium regulation. For example, overexpression of the calcium modulator, MICU3, in patient lines could be done to examine potential correction of calcium homeostasis defects previously found in this group (Dafinca et al., 2016). In addition, corrections of glutamate receptor subunit expression in mature iPSC-motor neurons could be done to look at the effect of responses to glutamate using electrophysiology techniques. These overexpression or knockdown studies may be validated by qRT-PCR and western blotting. The available *C9orf72* mouse models may also be utilised to examine potential similaries in gene expression changes using a combination of histological staining

and western blotting. This may give us an idea of whether the targets determined from the human translatome correlate with a model with a controlled genetic background. It may also signify the importance of varied genetic backgrounds in causing the disease. This may be more clearly investigated from more powerful RNA sequencing experiments using larger sample sizes of patient lines with the C9orf72 HRE. Finally, the validation of targets in post-mortem human spinal cord and motor cortex tissue would crucial to determining altered gene expression in affected individuals. Immunoblotting and immunohistochemistry in brain regions unaffected in ALS could also be performed to examine whether these changes are disease-specific, or an artefact of ageing.

6.6 Closing statements

This DPhil project has generated a versatile TRAP vector to purify RNA from human ChAT-enriched iPSC-motor neurons, a *C9orf72* ALS model recapitulating many molecular and cellular pathological phenotypes observed through numerous studies. The optimisation of this technique and validation via an RNA sequencing experiment has shown its potential as a powerful tool in several areas of neurodegeneration research.

7 References

Adolfsen B, Littleton JT (2001) Genetic and molecular analysis of the synaptotagmin family. Cell Mol Life Sci 58:393–402.

Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, Rogelj B, Al-Chalabi A, Hortobágyi T, Shaw CE (2011) P62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta Neuropathol 122:691–702.

Almeida S et al. (2013) Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol 126:385–399.

Amick J, Roczniak-Ferguson A, Ferguson SM (2016) C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. Mol Biol Cell 27:3040–3051.

Anderson P, Kedersha N (2008) Stress granules: the Tao of RNA triage. Trends Biochem Sci 33:141–150.

Aoki Y et al. (2017) C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. Brain 140:887–897.

Ash PEA, Bieniek KF, Gendron TF, Caulfield T, Lin W-L, DeJesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW, Rademakers R, Boylan KB, Dickson DW, Petrucelli L (2013) Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. Neuron 77:639–646.

Baborie A, Griffiths TD, Jaros E, Perry R, McKeith IG, Burn DJ, Masuda-Suzukake M, Hasegawa M, Rollinson S, Pickering-Brown S, Robinson AC, Davidson YS, Mann DMA (2015) Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. Neuropathol Appl Neurobiol 41:601–612.

Beck J et al. (2013) Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population.

Am J Hum Genet 92:345–353.

Berson A, Goodman LD, Sartoris AN, Otte CG, Aykit JA, Lee VMY, Trojanowski JQ, Bonini NM (2019) Drosophila Ref1/ALYREF regulates transcription and toxicity associated with ALS/FTD disease etiologies. Acta Neuropathol Commun 7:65.

Bertin B, Renaud Y, Aradhya R, Jagla K, Junion G (2015) TRAP-rc, Translating Ribosome Affinity Purification from Rare Cell Populations of Drosophila Embryos. J Vis Exp:1–6.

Boeynaems S et al. (2017) Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell 65:1044-1055.e5.

Boeynaems S, Bogaert E, Michiels E, Gijselinck I, Sieben A, Verstrepen KJ, Callaerts P, Rousseau F, Schymkowitz J (2016) Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS / FTD. Sci Rep:7–14.

Bortoluzzi S, D'Alessi F, Romualdi C, Danieli G a (2001) Differential expression of genes coding for ribosomal proteins in different human tissues. Bioinformatics 17:1152–1157.

Bruijn LI, Becher MW, Lee MK, Anderson KL, Jenkins NA, Copeland NG, Sisodia SS, Rothstein JD, Borchelt DR, Price DL, Cleveland DW (1997) ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions. Neuron 18:327–338.

Bursch F, Kalmbach N, Naujock M, Staege S, Eggenschwiler R, Abo-Rady M, Japtok J, Guo W, Hensel N, Reinhardt P, Boeckers TM, Cantz T, Sterneckert J, Van Den Bosch L, Hermann A, Petri S, Wegner F (2019) Altered calcium dynamics and glutamate receptor properties in iPSC derived motor neurons from ALS patients with C9orf72, FUS, SOD1 or TDP43 mutations. Hum Mol Genet:1–52.

Camp AJ, Wijesinghe R (2009) Calretinin: Modulator of neuronal excitability. Int J Biochem Cell Biol 41:2118–2121.

Ceresa BP, Bahr SJ (2006) rab7 activity affects epidermal growth factor: Epidermal growth factor receptor degradation by regulating endocytic trafficking from the late endosome. J Biol Chem 281:1099–1106.

Chen H, Qian K, Du Z, Cao J, Petersen A, Liu H, Blackbourn LW, Huang CL, Errigo A, Yin Y,

Lu J, Ayala M, Zhang SC (2014) Modeling ALS with iPSCs reveals that mutant SOD1 misregulates neurofilament balance in motor neurons. Cell Stem Cell 14:796–809.

Chen PI, Kong C, Su X, Stahl PD (2009) Rab5 isoforms differentially regulate the trafficking and degradation of epidermal growth factor receptors. J Biol Chem 284:30328–30338.

Chew J et al. (2015) C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science (80-) 348:1151–1154.

Chew J, Cook C, Gendron TF, Jansen-West K, del Rosso G, Daughrity LM, Castanedes-Casey M, Kurti A, Stankowski JN, Disney MD, Rothstein JD, Dickson DW, Fryer JD, Zhang Y-J, Petrucelli L (2019) Aberrant deposition of stress granule-resident proteins linked to C9orf72-associated TDP-43 proteinopathy. Mol Neurodegener 14:9.

Chou CC et al. (2018) TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. Nat Neurosci 21:228–239.

Chua CEL, Tang BL (2015) Role of Rab GTPases and their interacting proteins in mediating metabolic signalling and regulation. Cell Mol Life Sci 72:2289–2304.

Ciura S, Lattante S, Le Ber I, Latouche M, Tostivint H, Brice A, Kabashi E (2013) Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. Ann Neurol 74:180–187.

Clark RM, Blizzard CA, Young KM, King AE, Dickson TC (2017) Calretinin and Neuropeptide Y interneurons are differentially altered in the motor cortex of the SOD1G93A mouse model of ALS. Sci Rep 7:44461.

Colbran RJ, Brown AM (2004) Calcium/calmodulin-dependent protein kinase II and synaptic plasticity. Curr Opin Neurobiol 14:318–327.

Collum TD, Lutton E, Raines CD, Dardick C, Culver JN (2019) Identification of phloem-associated translatome alterations during leaf development in Prunus domestica L. Hortic Res 6.

Cooper-Knock J, Bury JJ, Heath PR, Wyles M, Higginbottom A, Gelsthorpe C, Highley JR, Hautbergue G, Rattray M, Kirby J, Shaw PJ (2015a) C9ORF72 GGGGCC expanded repeats produce splicing dysregulation which correlates with disease severity in

amyotrophic lateral sclerosis. PLoS One 10:1–17.

Cooper-Knock J, Higginbottom A, Stopford MJ, Highley JR, Ince PG, Wharton SB, Pickering-Brown S, Kirby J, Hautbergue GM, Shaw PJ (2015b) Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. Acta Neuropathol 130:63–75.

Cooper-Knock J, Walsh MJ, Higginbottom A, Highley JR, Dickman MJ, Edbauer D, Ince PG, Wharton SB, Wilson S a., Kirby J, Hautbergue GM, Shaw PJ (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain 137:2040–2051.

Corbier C, Sellier C (2017) C9ORF72 is a GDP/GTP exchange factor for Rab8 and Rab39 and regulates autophagy. Small GTPases 8:181–186.

Corcoran EE, Means AR (2001) Defining Ca2+/Calmodulin-dependent Protein Kinase Cascades in Transcriptional Regulation. J Biol Chem 276:2975–2978.

Cullis DN, Philip B, Baleja JD, Feig LA (2002) Rab11-FIP2, an adaptor protein connecting cellular components involved in internalization and recycling of epidermal growth factor receptors. J Biol Chem 277:49158–49166.

D'Ambrogio A, Buratti E, Stuani C, Guarnaccia C, Romano M, Ayala YM, Baralle FE (2009) Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. Nucleic Acids Res 37:4116–4126.

Dafinca R, Scaber J, Ababneh N, Lalic T, Weir G, Christian H, Vowles J, Douglas AGL, Fletcher-Jones A, Browne C, Nakanishi M, Turner MR, Wade-Martins R, Cowley SA, Talbot K (2016) C9orf72 Hexanucleotide Expansions are Associated with Altered ER Calcium Homeostasis and Stress Granule Formation in iPSC-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Stem Cells.

Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, Masuda-Suzukake M, Suzuki G, Nonaka T, Shi J, Tian J, Hamdalla H, Ealing J, Richardson A, Jones M, Pickering-Brown S, Snowden JS, Hasegawa M, Mann DMA (2016) Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with

expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol Appl Neurobiol 42:242–254.

Davidson YS, Barker H, Robinson AC, Thompson JC, Harris J, Troakes C, Smith B, Al-Saraj S, Shaw C, Rollinson S, Masuda-Suzukake M, Hasegawa M, Pickering-Brown S, Snowden JS, Mann DM (2014) Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol Commun 2:70.

DeJesus-Hernandez M et al. (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron 72:245–256.

Donnelly CJ et al. (2013) RNA Toxicity from the ALS / FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron 80:415–428.

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N (2008) Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. Cell 135:749–762.

Drane L, Ainsley J a., Mayford MR, Reijmers LG (2014) A transgenic mouse line for collecting ribosome-bound mRNA using the tetracycline transactivator system. Front Mol Neurosci 7:1–10.

Elkouris M, Balaskas N, Poulou M, Politis PK, BE P, Malas S, D T, DE R (2012) Sox1 Maintains the Undifferentiated State of Cortical Neural Progenitor Cells via the Suppression of Prox1-Mediated Cell Cycle Exit and Neurogenesis. In: Stem Cells, pp 591–598.

Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RAK, Levina V, Halloran M a., Gleeson P a., Blair IP, Soo KY, King AE, Atkin JD (2014) C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum Mol Genet 23:3579–3595.

Faye MD, Graber TE, Holcik M (2014) Assessment of selective mRNA translation in

mammalian cells by polysome profiling. J Vis Exp:e52295.

Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee K-H, Badders N, Valentine M, Miller BL, Wong PC, Petrucelli L, Kim HJ, Gao F-B, Taylor JP (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature.

Fujimori K, Ishikawa M, Otomo A, Atsuta N, Nakamura R, Akiyama T, Hadano S, Aoki M, Saya H, Sobue G, Okano H (2018) Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. Nat Med 24.

Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE a, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J, Cosio DM, Van Blitterswijk M, Lee WC, Rademakers R, Boylan KB, Dickson DW, Petrucelli L (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol 126:829–844.

Gijselinck I et al. (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. Lancet Neurol 11:54–65.

Gijselinck I et al. (2015) The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry:1–13.

Gracida X, Calarco JA (2017) Cell type-specific transcriptome profiling in C. elegans using the Translating Ribosome Affinity Purification technique. Methods 126:130–137.

Gurney ME, Pu H, Chiu AY, Dal Canto MC, Polchow C, Alexander D, Caliendo J, Hentati A, Kwon Y, Siddique T (1994) Motor neuron degeneration in mice that express a human Cu, Zn Superoxide Dismutase Mutation. Science (80-) 264:1772.

Haeusler AR, Donnelly CJ, Periz G, Simko E a J, Shaw PG, Kim M-S, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 507:195–200.

Hallonet M, Hollemann T, Pieler T, Gruss P (1999) Vax1, a novel homeobox-containing gene, directs development of the basal forebrain and visual system. Genes Dev:3106–3114.

Hardwick SA, Chen WY, Wong T, Deveson IW, Blackburn J, Andersen SB, Nielsen LK, Mattick

JS, Mercer TR (2016) Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13:792–798.

Hautbergue GM et al. (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. Nat Commun 8:1–18.

Heiman M, Kulicke R, Fenster RJ, Greengard P, Heintz N (2014) Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). Nat Protoc 9:1282–1291.

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan D a., Surmeier DJ, Greengard P, Heintz N (2008) A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. Cell 135:738–748.

Henderson AS, Warburton D, Atwood KC (1972) Location of ribosomal DNA in the human chromosome complement. Proc Natl Acad Sci U S A 69:3394–3398.

Higgins CMJ, Jung C, Xu Z (2003) ALS-associated mutant SOD1G93A causes mitochondrial vacuolation by expansion of the intermembrane space and by involvement of SOD1 aggregation and peroxisomes. BMC Neurosci 14:1–14.

Hirano A, Nakano I, Kurland LT, Mulder DW, Holley PW, Saccomanno G (1984) Fine Structural Study of Neurofibrillary Changes in a Family with Amyotrophic Lateral Sclerosis. J Neuropathol Exp Neurol 43:471–480.

Huang T, Nguyen LH, Lin T V., Gong X, Zhang L, Kim GB, Sarkisian MR, Breunig JJ, Bordey A (2018) In utero electroporation-based translating ribosome affinity purification identifies age-dependent mRNA expression in cortical pyramidal neurons. Neurosci Res.

Huang Y, Ng FS, Jackson FR (2015) Comparison of Larval and Adult Drosophila Astrocytes Reveals Stage-Specific Gene Expression Profiles. Genes, Genomes, Genet 5:551–558.

Hupe M, Li MX, Gertow Gillner K, Adams RH, Stenman JM (2014) Evaluation of TRAP-sequencing technology with a versatile conditional mouse model. Nucleic Acids Res 42:1–11.

Iacobucci GJ, Popescu GK (2017) Resident Calmodulin Primes NMDA Receptors for Ca2+-

Dependent Inactivation. Biophys J 113:2236–2248.

Jiang J, Zhu Q, Gendron TF, Rigo F, Cleveland DW, Jiang J, Zhu Q, Gendron TF, Saberi S, Mcalonis-downes M, Seelman A (2016) Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron 90:535–550.

Jovičić A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, Paul JW, Sun S, Herdy JR, Bieri G, Kramer NJ, Gage FH, Van Den Bosch L, Robberecht W, Gitler AD (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci 18:1226–1229.

Kanadia RN, Johnstone KA, Mankodi A, Timmers AM, Hauswirth WW (2011) A Muscleblind Knockout Model for Myotonic Dystrophy. Science (80-) 1978:10–13.

Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ (2000) Principles of Neural Science (Kandel ER, Schwartz JH, Jessell TM, Mack S, Dodd J, Butler J, Lebowitz H, Dahlgren S, Siegel E, Ackerman J, Cuddihy J, eds)., 4th ed. The McGraw-Hill Companies.

Katahira J, Inoue H, Hurt E, Yoneda Y (2009) Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. EMBO J 28:556–567.

Kearse MG, Todd PK (2014) Repeat-Associated Non-AUG Translation and Its Impact in Neurodegenerative Disease. Neurotherapeutics 11:721–731.

Khosravi B, Hartmann H, May S, Möhl C, Ederle H, Michaelsen M, Schludi MH, Dormann D, Edbauer D (2017) Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. Hum Mol Genet 26:790–800.

Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, Burrell JR, Zoing MC (2011) Amyotrophic lateral sclerosis. Lancet 377:942–955.

Kikuchi H, Almer G, Yamashita S, Guégan C, Nagai M, Xu Z, Sosunov AA, McKhann GM, Przedborski S (2006) Spinal cord endoplasmic reticulum stress associated with a microsomal accumulation of mutant superoxide dismutase-1 in an ALS model. Proc Natl Acad Sci U S A 103:6025–6030.

Kirberger M, Gorkhali R, Salarian M, Yang J (2017) Calmodulin (CALM1). Springer Sci Encycl Signal Mol:1–10.

Kool MJ, Proietti Onori M, Borgesius NZ, van de Bree JE, Elgersma-Hooisma M, Nio E, Bezstarosti K, Buitendijk GHS, Aghadavoud Jolfaei M, Demmers JAA, Elgersma Y, van Woerden GM (2019) CAMK2-Dependent Signaling in Neurons Is Essential for Survival. J Neurosci 39:5424–5439.

Kumar A, Kankainen M, Parsons A, Kallioniemi O, Mattila P, Heckman CA (2017) The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. BMC Genomics 18:1–13.

Kuo PH, Doudeva LG, Wang YT, Shen CKJ, Yuan HS (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. Nucleic Acids Res 37:1799–1808.

Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL (2014) Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science (80-) 345:1254917.

Lacomblez L, Bensimon G, Leigh PN, Guillet P, Meininger V (1996) Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Lancet 347:1425–1431.

Lagier-Tourenne C et al. (2013) Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. PNAS 110:E4530-9.

Lagier-Tourenne C, Polymenidou M, Cleveland DW (2010) TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. Hum Mol Genet 19:46–64.

Lall D, Baloh RH (2017) Microglia and C9orf72 in neuroinflammation and ALS and frontotemporal dementia. J Clin Invest 127:3250–3258.

Lee KH et al. (2016) C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell 167:774-788.e17.

Lee YB et al. (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep 5:1178–1186.

Lesiak AJ, Brodsky M, Neumaier JF (2015) RiboTag is a flexible tool for measuring the translational state of targeted cells in heterogeneous cell cultures. Biotechniques 317.

Levine TP, Daniels RD, Gatta AT, Wong LH, Hayes MJ (2013) The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics 29:499–503.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liang X, Song MR, Xu ZG, Lanuza GM, Liu Y, Zhuang T, Chen Y, Pfaff SL, Evans SM, Sun Y (2011) Isl1 Is required for multiple aspects of motor neuron development. Mol Cell Neurosci 47:215–222.

Liao Y, Smyth GK, Shi W (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

Lindblad K, Savontaus ML, Stevanin G, Holmberg M, Digre K, Zander C, Ehrsson H, David G, Benomar A, Nikoskelainen E, Trottier Y, Holmgren G, Ptacek LJ, Anttinen A, Brice A, Schalling M (1996) An expanded CAG repeat sequence in spinocerebellar ataxia type 7. Genome Res 6:965–971.

Ling SC, Polymenidou M, Cleveland DW (2013) Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. Neuron 79:416–438.

Liu Y, Pattamatta A, Zu T, Reid T, Bardhi O, Borchelt DR, Yachnis AT, Ranum LPW (2016) C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron 90:521–534.

Lobsiger CS, Boillee S, Cleveland DW (2007) Toxicity from different SOD1 mutants dysregulates the complement system and the neuronal regenerative response in ALS motor neurons. Proc Natl Acad Sci 104:7319–7326.

Lomen-Hoerth C, Anderson T, Miller B (2002) The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. Neurology 59:1077–1079.

Longman D, Johnstone IL, Cáceres JF (2003) The REF/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. RNA 9:881–891.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for

RNA-seq data with DESeq2. Genome Biol 15:1–21.

Lynch-Day M, Mao K, Wang K, Zhao M, Klionsky D (2012) Role of Autophagy in Parkinson's Disease. Cold Spring Harb Perspect Med 2:1–13.

MacDonald ME, Bates GP, Buckler AJ, Altherr M, Tagle D, Snell R (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell 72:971–983.

Mackenzie IR a et al. (2007) Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann Neurol 61:427–434.

Mackenzie IRA, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, Edbauer D, Kremmer E, Prudlo J, Troost D, Neumann M (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol 130:845–861.

MacNair L, Xiao S, Miletic D, Ghani M, Julien J-P, Keith J, Zinman L, Rogaeva E, Robertson J (2015) MTHFSD and DDX58 are novel RNA-binding proteins abnormally regulated in amyotrophic lateral sclerosis. Brain:awv308.

Magrané J, Cortez C, Gan WB, Manfredi G (2014) Abnormal mitochondrial transport and morphology are common pathological denominators in SOD1 and TDP43 ALS mouse models. Hum Mol Genet 23:1413–1424.

Mann DM, Robinson A, Rollinson S, Bennion Callister J, Thompson JC, Snowden JS, Gendron T, Petrucelli L, Masuda-Suzukake M, Hasegawa M, Davidson Y, Pickering-Brown S (2013) Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol Commun 1:1.

Mann JR, Gleixner AM, Mauna JC, Gomes E, DeChellis-Marks MR, Needham PG, Copley KE, Hurtle B, Portz B, Pyles NJ, Guo L, Calder CB, Wills ZP, Pandey UB, Kofler JK, Brodsky JL, Thathiah A, Shorter J, Donnelly CJ (2019) RNA Binding Antagonizes

Neurotoxic Phase Transitions of TDP-43. Neuron 102:321-338.e8.

Marat AL, Dokainish H, McPherson PS (2011) DENN domain proteins: Regulators of Rab GTPases. J Biol Chem 286:13791–13800.

Mattedi F, Potrich V, Quattrone A, Shorrock HK, Zhou H, Groen EJN, Zuccotti P, Newbery HJ, Gillingwater TH, Muntoni F, Perenthaler E, Viero G, Tebaldi T, Bernabò P, Lane FM (2017) In Vivo Translatome Profiling in Spinal Muscular Atrophy Reveals a Role for SMN Protein in Ribosome Biology. Cell Rep 21:953–965.

Maurel C, Dangoumau A, Marouillat S, Brulard C, Chami A, Hergesheimer R, Corcia P, Blasco H, Andres CR, Vourc'h P (2018) Causative Genes in Amyotrophic Lateral Sclerosis and Protein Degradation Pathways: a Link to Neurodegeneration. Mol Neurobiol 55:6480–6499.

McDonald KK, Aulas A, Destroismaisons L, Pickles S, Beleac E, Camu W, Rouleau GA, Velde C Vande (2011) TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum Mol Genet 20:1400–1410.

Meyer K, Ferraiuolo L, Miranda CJ, Likhite S, McElroy S, Renusch S, Ditsworth D, Lagier-Tourenne C, Smith RA, Ravits J, Burghes AH, Shaw PJ, Cleveland DW, Kolb SJ, Kaspar BK (2014) Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. Proc Natl Acad Sci 111:829–832.

Miller JW, Urbinati CR, Teng-umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS (2000) Recruitment of human muscleblindproteins to (CUG)n expansions associatedwith myotonic dystrophy. EMBO J 19:4439–4448.

Mizielinska S, Mizielinska S, Grönke S, Niccoli T, Ridler CE, Woollacott IOC, Pietrzyk J, Cleverley K, Nicoll AJ, Fratta P, Fisher EMC, Partridge L, Isaacs AM (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science (80-) 345:1192–1194.

Mori K, Arzberger T, Grässer FA, Gijselinck I, May S, Rentzsch K, Weng SM, Schludi MH, Van Der Zee J, Cruts M, Van Broeckhoven C, Kremmer E, Kretzschmar HA, Haass C,

Edbauer D (2013a) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol 126:881–893.

Mori K, Lammich S, Mackenzie IR a, Forné I, Zilow S, Kretzschmar H, Edbauer D, Janssens J, Kleinberger G, Cruts M, Herms J, Neumann M, Van Broeckhoven C, Arzberger T, Haass C (2013b) HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol 125:413–423.

Morita M, Al-Chalabi A, Andersen PM, Hosler B, Sapp P, Englund E, Mitchell JE, Habgood JJ, De Belleroche J, Xi J, Jongjaroenprasert W, Horvitz HR, Gunnarsson LG, Brown RH (2006) A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology 66:839–844.

Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. Genes Dev 30:553–566.

Naciff JM, Behbehani MM, Misawa H, Dedman JR (1999) Identification and transgenic analysis of a murine promoter that targets cholinergic neuron expression. J Neurochem 72:17–28.

Nectow AR, Ekstrand MI, Friedman JM (2015) Molecular characterization of neuronal cell types based on patterns of projection with Retro-TRAP. Nat Protoc 10:1319–1327.

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar H a, Trojanowski JQ, Lee VM-Y (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science (80-) 314:130–133.

Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storvall H, Cao M, Sandberg R, Deng Q, Hedlund E (2016) Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat Commun 7:1–11.

Nordin A, Akimoto C, Wuolikainen A, Alstermark H, Jonsson P, Birve A, Marklund SL, Graffmo KS, Forsberg K, Brännström T, Andersen PM (2014) Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. Hum Mol Genet 24:3133–3142.

O'Leary MN, Schreiber KH, Zhang Y, Duc ACE, Rao S, Hale JS, Academia EC, Shah SR, Morton JF, Holstein CA, Martin DB, Kaeberlein M, Ladiges WC, Fink PJ, MacKay VL, Wiest DL, Kennedy BK (2013) The Ribosomal Protein Rpl22 Controls Ribosome Composition by Directly Repressing Expression of Its Own Paralog, Rpl22l1. PLoS Genet 9.

O'Rourke JG et al. (2015) C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. Neuron 88:892–901.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods 14:417–419.

Patron M, Granatiero V, Espino J, Rizzuto R, De Stefani D (2019) MICU3 is a tissue-specific enhancer of mitochondrial calcium uptake. Cell Death Differ 26:179–195.

Peters OM, Toro G, Tran H, Mueller C, Robert H, Peters OM, Cabrera GT, Tran H, Gendron TF, Mckeon JE, Metterville J (2015) Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. Neuron 88:902–909.

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9:171–181.

Polymenidou M, Lagier-Tourenne C, Hutt KR, Bennett CF, Cleveland DW, Yeo GW (2012) Misregulated RNA processing in amyotrophic lateral sclerosis. Brain Res 1462:3–15.

Prudencio M et al. (2015) Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat Neurosci 18:1175–1182.

Qiu J, Dando O, Baxter PS, Hasel P, Heron S, Simpson TI, Hardingham GE (2018) Mixed-species RNA-seq for elucidation of non-cell-autonomous control of gene transcription.

Nat Protoc 13:2176–2199.

Renton AE et al. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72:257–268.

Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32:896–902.

Romano M, Buratti E, Romano G, Klima R, Del Bel Belluz L, Stuani C, Baralle F, Feiguin F (2014) Evolutionarily conserved heterogeneous nuclear ribonucleoprotein (hnRNP) A/B proteins functionally interact with human and Drosophila TAR DNA-binding protein 43 (TDP-43). J Biol Chem 289:7121–7130.

Rosen DR et al. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362:59–62.

Rydbirk R, Folke J, Winge K, Aznar S, Pakkenberg B, Brudek T (2016) Assessment of brain reference genes for RT-qPCR studies in neurodegenerative diseases. Sci Rep 6:37116.

Sabirzhanova I, Boinot C, Guggino WB, Cebotaru L (2018) Syntaxin 8 and the endoplasmic reticulum processing of Δf508-CFTR. Cell Physiol Biochem 51:1489–1499.

Sanz E, Yang L, Su T, Morris DR, McKnight GS, Amieux PS (2009) Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. PNAS 106:13939–13944.

Sareen D et al. (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci Transl Med 5:208ra149.

Saxena S, Cabuy E, Caroni P (2009) A role for motoneuron subtype-selective ER stress in disease manifestations of FALS mice. Nat Neurosci 12:627–636.

Scaber J, Talbot K (2016) What is the role of TDP-43 in C9orf72-related amyotrophic lateral sclerosis and frontemporal dementia? Brain 19:1175–1187.

Sellier C, Campanari M, Julie Corbier C, Gaucherot A, Kolb-Cheynel I, Oulad-Abdelghani M, Ruffenach F, Page A, Ciura S, Kabashi E, Charlet-Berguerand N (2016) Loss of C9 ORF 72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death . EMBO J 35:1276–1297.

Selvaraj BT et al. (2018) C9ORF72 repeat expansion causes vulnerability of motor neurons

to Ca2+-permeable AMPA receptor-mediated excitotoxicity. Nat Commun 9:347.

Shaw MP, Higginbottom A, McGown A, Castelli LM, James E, Hautbergue GM, Shaw PJ, Ramesh TM (2018) Stable transgenic C9orf72 zebrafish model key aspects of the ALS/FTD phenotype and reveal novel pathological features. Acta Neuropathol Commun 6:125.

Shin OH, Maximov A, Lim BK, Rizo J, Südhof TC (2004) Unexpected Ca2+-binding properties of synaptotagmin 9. Proc Natl Acad Sci U S A 101:2554–2559.

Shu-Nu C, Lin CH, Lin A (2000) An acidic amino acid cluster regulates the nucleolar localization and ribosome assembly of human ribosomal protein L22. FEBS Lett 484:22–28.

Siddique T, Figlewigz DA, Pericak-Vance MA, Haines JL, Rouleau G, Jeffers AJ, Sapp P, Hung W-Y, Bebout J, McKenna-Yasek D, Deng G, Horvitz HR, Gusella JF, Brown RH, Roses AD (1991) Linkage of a Gene Causing Familial Amyotrophic Lateral Sclerosis to Chromosome 21 and Evidence of Genetic-Locus Heterogeneity. N Engl J Med 324:1381–1384.

Sinkus M, Graw S, Freedman R, Ross R, Lester H, Leonard S (2016) The Human CHRNA7 and CHRFAM7A Genes: A Review of the Genetics, Regulation, and Function. Neuropharmacology 118:6072–6078.

Solomon DA et al. (2018) A feedback loop between dipeptide-repeat protein, TDP-43 and karyopherin-α mediates C9orf72-related neurodegeneration. Brain:2908–2924.

Soneson C, Love MI, Robinson MD (2019) Differential analyses for RNA-seq : transcript-level estimates improve gene-level inferences. F1000 Res:1–18.

Sudria-Lopez E, Koppers M, de Wit M, van der Meer C, Westeneng H-J, Zundel CAC, Youssef SA, Harkema L, de Bruin A, Veldink JH, van den Berg LH, Pasterkamp RJ (2016) Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol 132:145–147.

Sullivan PM, Zhou X, Robins AM, Paushter DH, Kim D, Smolka MB, Hu F (2016) The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate

the autophagy-lysosome pathway. Acta Neuropathol Commun 4:51.

Sun S, Sun Y, Ling S, Ferraiuolo L, Mcalonis-downes M, Zou Y (2015) Translational profiling identifies a cascade of damage initiated in motor neurons and spreading to glia in mutant SOD1-mediated ALS. PNAS.

Suzuki N, Maroof AM, Merkle FT, Koszka K, Intoh A, Armstrong I, Moccia R, Davis-Dusenbery BN, Eggan K (2013) The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. Nat Neurosci 16:1725–1727.

Swarup V, Phaneuf D, Bareil C, Robertson J, Rouleau GA, Kriz J, Julien JP (2011) Pathological hallmarks of amyotrophic lateral sclerosis/frontotemporal lobar degeneration in transgenic mice produced with TDP-43 genomic fragments. Brain 134:2610–2626.

Swinnen B, Abreu AB, Gendron TF, Boeynaems S, Bogaert E, Nuyts R, Timmers M, Scheveneels W, Hersmus N, Wang J, Mizielinska S, Isaacs AM, Petrucelli L, Lemmens R, Damme P Van, Bosch L Van Den, Robberecht W (2018) A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. Acta Neuropathol.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131:861–872.

Takahashi K, Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126:663–676.

Talbot K, Ansorge O (2006) Recent advances in the genetics of amyotrophic lateral sclerosis and frontotemporal dementia: common pathways in neurodegenerative disease. Hum Mol Genet 15 Spec No:R182-7.

Talbot K, Feneberg E, Scaber J, Thompson AG, Turner MR (2018) Amyotrophic lateral sclerosis: the complex path to precision medicine. J Neurol 265:2454–2462.

Tandan R, Bradley WG (1985) Amyotrophic lateral sclerosis: Part 1. Clinical features, pathology, and ethical issues in management. Ann Neurol 18:271–280.

Tian C, Wang Y, Yu H, He J, Wang J, Shi B, Du Q, Provart NJ, Meyerowitz EM, Jiao Y (2019)

A gene expression map of shoot domains reveals regulatory mechanisms. Nat Commun 10:1–12.

Tintaru AM, Hautbergue GM, Hounslow AM, Hung ML, Lian LY, Craven JC, Wilson SA (2007) Structural and functional analysis of RNA and TAP binding to SF2/ASF. EMBO Rep 8:756–762.

Toli D, Buttigieg D, Blanchard S, Lemonnier T, Lamotte d'Incamps B, Bellouze S, Baillat G, Bohl D, Haase G (2015) Modeling amyotrophic lateral sclerosis in pure human iPSc-derived motor neurons isolated by a novel FACS double selection technique. Neurobiol Dis 82:269–280.

Tran H, Almeida S, Moore J, Petrucelli L, Nickerson JAA, Petrucelli L, Weng Z, Gao F, Gendron TF, Chalasani U, Lu Y, Du X, Nickerson JAA, Petrucelli L, Weng Z, Gao F (2015) Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron 87:1207–1214.

Trifonov S, Houtani T, Hamada S, Kase M, Maruyama M, Sugimoto T (2009) In situ hybridization study of the distribution of choline acetyltransferase mRNA and its splice variants in the mouse brain and spinal cord. Neuroscience 159:344–357.

Troakes C, Maekawa S, Wijesekera L, Rogelj B, Siklós L, Bell C, Smith B, Newhouse S, Vance C, Johnson L, Hortobágyi T, Shatunov A, Al-Chalabi A, Leigh N, Shaw CE, King A, Al-Sarraj S (2012) An MND/ALS phenotype associated with C9orf72 repeat expansion: Abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. Neuropathology 32:505–514.

Tryon RC, Pisat N, Johnson SL, Dougherty JD (2013) Development of Translating Ribosome Affinity Purification for Zebrafish. Genesis 192:187–192.

Turner BJ, Talbot K (2008) Transgenics, toxicity and therapeutics in rodent models of mutant SOD1-mediated familial ALS. Prog Neurobiol 85:94–134.

Uddin MS, Stachowiak A, Al Mamun A, Tzvetkov NT, Takeda S, Atanasov AG, Bergantin LB, Abdel-Daim MM, Stankiewicz AM (2018) Autophagy and Alzheimer's disease: From molecular mechanisms to therapeutic implications. Front Aging Neurosci 10:1–18.

van Blitterswijk M et al. (2013) Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol 12:978–988.

van Blitterswijk M et al. (2015) Novel clinical associations with specific C9ORF72 transcripts in patients with repeat expansions in C9ORF72. Acta Neuropathol 130:863–876.

van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, van den Berg LH (2017) Amyotrophic lateral sclerosis. Lancet 390:2084–2098.

Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingham J, Williams KL, Blair IP, Nicholson G, de Belleroche J, Gallo J-M, Miller CC, Shaw CE (2009) Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science (80-):1208–1211.

Varcianna A, Myszczynska MA, Castelli LM, O'Neill B, Kim Y, Talbot J, Nyberg S, Nyamali I, Heath PR, Stopford MJ, Hautbergue GM, Ferraiuolo L (2019) Micro-RNAs secreted through astrocyte-derived extracellular vesicles cause neuronal network degeneration in C9orf72 ALS. EBioMedicine 40:626–635.

Viphakone N, Hautbergue GM, Walsh M, Chang C Te, Holland A, Folco EG, Reed R, Wilson SA (2012) TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat Commun 3:1006–1011.

Volk AE, Weishaupt JH, Andersen PM, Ludolph AC, Kubisch C (2018) Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. Medizinische Genet 30:252–258.

Wang C, Kang X, Zhou L, Chai Z, Wu Q, Huang R, Xu H, Hu M, Sun X, Sun S, Li J, Jiao R, Zuo P, Zheng L, Yue Z, Zhou Z (2018) Synaptotagmin-11 is a critical mediator of parkin-linked neurotoxicity and Parkinson's disease-like pathology. Nat Commun 9:1–14.

Wang C, Wang Y, Hu M, Chai Z, Wu Q, Huang R, Han W, Zhang CX, Zhou Z (2015) Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. EMBO Rep 17:47–63.

Webster CP, Smith EF, Bauer CS, Moller A, Guillaume M, Ferraiuolo L, Myszczynska MA, Higginbottom A, Walsh MJ, Whitworth J, Kaspar BK, Meyer K, Shaw PJ, Grierson AJ,

Vos KJ De, Webster CP (2016) The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J:1–21.

Welzel G, Seitz D, Schuster S (2015) Magnetic-activated cell sorting (MACS) can be used as a large-scale method for establishing zebrafish neuronal cell cultures. Sci Rep 5:1–7.

Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y, Lin S, Shneider NA, Monaghan J, Pandey UB, Pasinelli P, Ichida JK, Trotti D (2014) Antisense Proline-Arginine RAN Dipeptides Linked to C9ORF72-ALS/FTD Form Toxic Nuclear Aggregates that Initiate In Vitro and In Vivo Neuronal Death. Neuron 84:1213–1225.

Wojciechowska M, Olejniczak M, Galka-Marciniak P, Jazurek M, Krzyzosiak WJ (2014) RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic Acids Res 42:11849–11864.

Wong C-O, Venkatachalam K (2019) Motor neurons from ALS patients with mutations in C9ORF72 and SOD1 exhibit distinct transcriptional landscapes. Hum Mol Genet 28:2799–2810.

Wong T, Deveson IW, Hardwick SA, Mercer TR (2017) ANAQUIN: A software toolkit for the analysis of spike-in controls for next generation sequencing. Bioinformatics 33:1723–1724.

Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, Keith J, Zinman L, Rogaeva E, Robertson J (2015) Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann Neurol 78:568–583.

Yang M, Liang C, Swaminathan K, Herrlinger S, Lai F, Shiekhattar R, Chen J-F (2016) A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Neurosci Lett:1–17.

Zerial M, McBride H (2001) Rab proteins as membrane organizers. Nat Rev Mol Cell Biol 2:107–117.

Zhang K et al. (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature.

Zhang Y-J et al. (2016) C9ORF72 poly(GA) aggregates sequester and impair HR23 and

nucleocytoplasmic transport proteins. Nat Neurosci 19.

Zu T, Liu Y, Bañez-coronel M, Reid T, Pletnikova O, Lewis J (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. PNAS.

8 Appendix

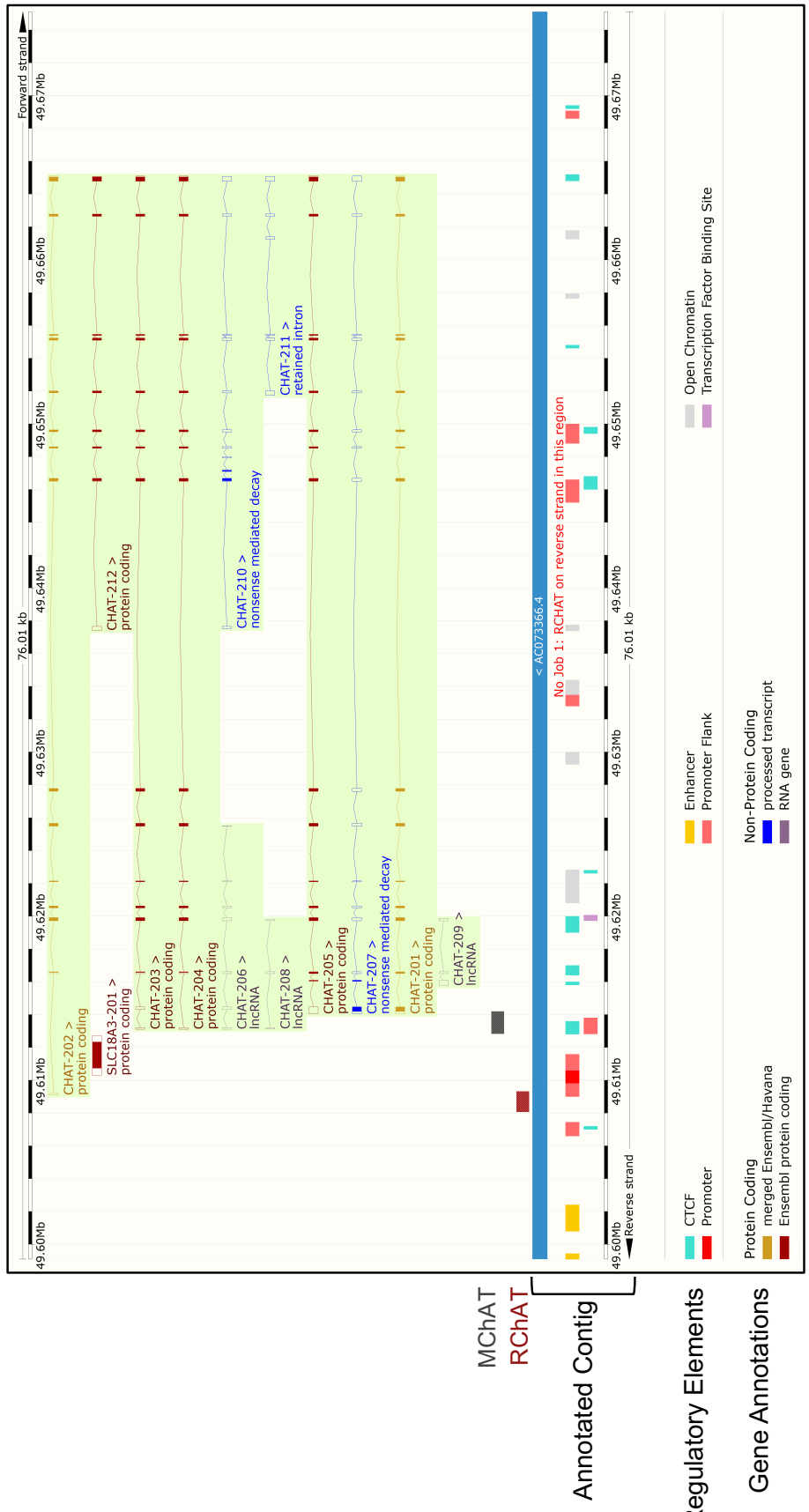
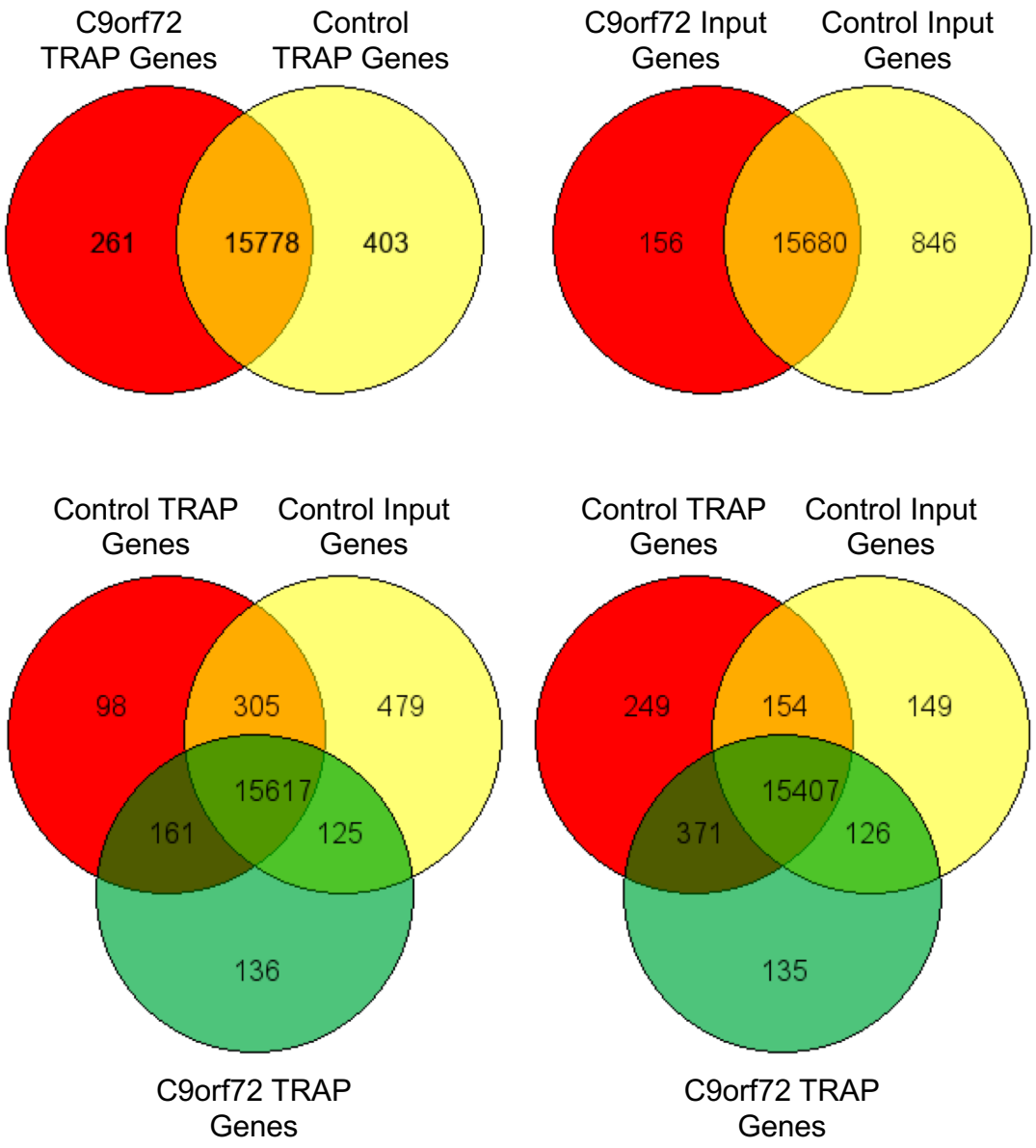


Figure S8.1. Alignment of predicted promoter regions, MChAT and RChAT for TRAP construct to the human ChAT locus.

Figure S8.1. Alignment of predicted promoter regions, MChAT and RChAT for TRAP construct to the human ChAT locus.
The sequences of MChAT and RChAT promoter isoforms, as given by Genecopoeia, were submitted to a Basic Local Alignment Search Tool (BLAST) search on www.ensembl.org. The alignment was done on the GRCh38 human genome build. The sequences were identified as aligning to the human ChAT locus, as given above in an automatically generated annotated diagram. Both isoforms overlap with the regulatory/ promoter regions at the human ChAT locus.

Figure S8.2. Overlap of total genes or transcripts detected in RNA sequencing experiment between TRAP and transcriptome input sample groups.

a



b

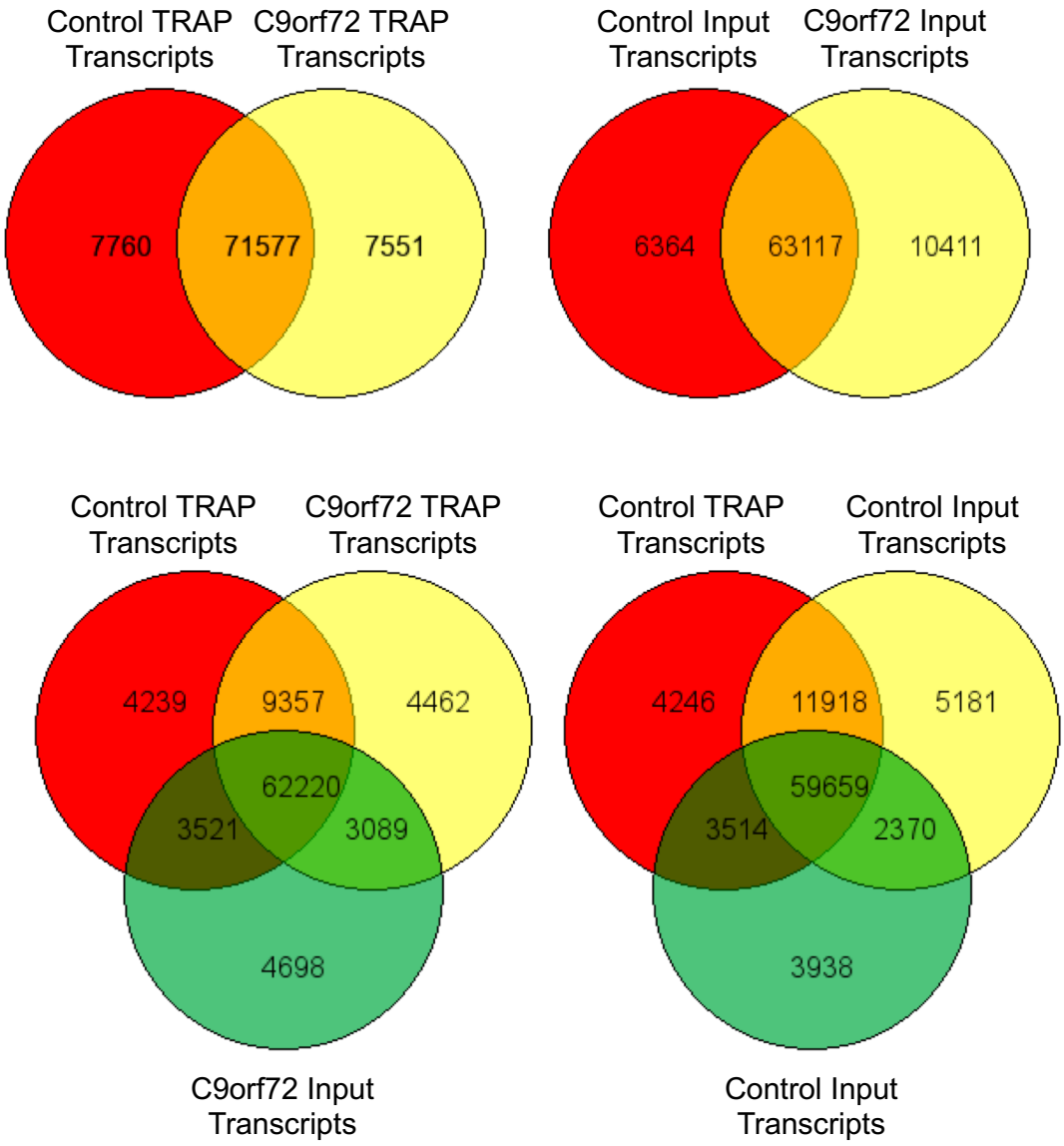


Figure S8.2. Venn diagrams showing the overlap of total genes or transcripts detected in RNA sequencing experiment between TRAP and transcriptome input sample groups. Count tables of total detected genes (a) or transcripts (b) with TRAP and transcriptome input sample groups were filtered for rows with at least 10 or more counts, and sequins were redacted.

Table S8.1. 50 genes with highest log2 fold change in expression, upregulated in differential expression analysis, comparing *C9orf72*-TRAP and Control-TRAP samples.

C9orf72-TRAP vs. Control-TRAP Top log2 Fold Change Genes				
Gene Name	**Ensembl**	**base Mean**	**log2 Fold Change**	**p-adj**
TBC1D3E	ENSG00000278599	264.53	10.33	1.40E-18
TBC1D3K	ENSG00000273513	48.65	9.63	2.60E-10
VAX1	ENSG00000148704	170.15	8.50	1.22E-04
FEZF1	ENSG00000128610	41.35	8.44	4.84E-03
NR2E1	ENSG00000112333	312.47	7.27	9.75E-05
LHX2	ENSG00000106689	45.80	7.19	1.13E-04
FOXG1	ENSG00000176165	111.71	7.16	1.26E-02
SIX3	ENSG00000138083	58.56	7.15	6.59E-05
RAX	ENSG00000134438	52.26	7.12	3.74E-02
FEZF2	ENSG00000153266	168.78	6.79	5.70E-09
TPTE	ENSG00000274391	141.15	6.62	4.30E-02
SIX6	ENSG00000184302	34.79	6.00	2.22E-05
ZNF248	ENSG00000198105	265.80	5.64	1.59E-12
TBC1D3C	ENSG00000278299	79.59	5.25	9.62E-13
NPIPA8	ENSG00000214940	298.74	5.20	2.59E-27
CHST9	ENSG00000154080	659.25	4.42	1.91E-09
FERD3L	ENSG00000146618	19.29	4.23	9.18E-03
OLIG2	ENSG00000205927	790.98	4.20	4.79E-02
PTPRN2	ENSG00000155093	399.73	4.03	7.09E-06
F8A3	ENSG00000277150	15.82	4.03	3.74E-02
ANKRD63	ENSG00000230778	38.20	3.94	1.35E-03
TEK	ENSG00000120156	43.29	3.91	1.29E-02
WNT7B	ENSG00000188064	51.66	3.82	2.97E-02
SOX14	ENSG00000168875	25.19	3.59	2.43E-02
ITPR3	ENSG00000096433	844.61	3.53	3.73E-02
NKX2-2	ENSG00000125820	80.69	3.40	3.03E-04
EIF2S3L	ENSG00000180574	12.59	3.36	8.80E-03
CTXN3	ENSG00000205279	108.77	3.33	5.92E-03
AJAP1	ENSG00000196581	342.76	3.30	1.01E-21
CBS	ENSG00000274276	1074.82	3.26	6.12E-04
GOLGA8K	ENSG00000249931	98.90	3.12	2.48E-03
TUBB8	ENSG00000261456	80.47	3.10	2.83E-03
FAM50B	ENSG00000145945	30.97	3.07	2.08E-04
CALB2	ENSG00000172137	17.11	3.01	3.52E-02
GOLGA8T	ENSG00000261247	77.42	2.92	4.86E-02
STEAP1B	ENSG00000105889	13.47	2.86	3.92E-02
RSPH10B2	ENSG00000169402	134.38	2.83	1.96E-02
ATP1A2	ENSG00000018625	56.08	2.80	9.03E-03
SOX1	ENSG00000182968	565.80	2.79	4.11E-10
EFCC1	ENSG00000114654	284.06	2.66	7.50E-05
CCDC37	ENSG00000163885	58.77	2.66	3.66E-02
ELF4	ENSG00000102034	413.52	2.62	3.53E-03
TCERG1L	ENSG00000176769	224.00	2.60	1.72E-02
PCDH18	ENSG00000189184	1772.11	2.50	3.25E-05
STON1-GTF2A1L	ENSG00000068781	122.13	2.49	3.19E-03
MMRN1	ENSG00000138722	86.12	2.49	3.25E-05
CCDC169-SOHLH2	ENSG00000250709	102.63	2.48	2.15E-02
NUTM2E	ENSG00000228570	27.50	2.45	3.86E-02
DNAI1	ENSG00000122735	222.05	2.38	2.72E-02
TCF7L2	ENSG00000148737	644.81	2.33	3.71E-04

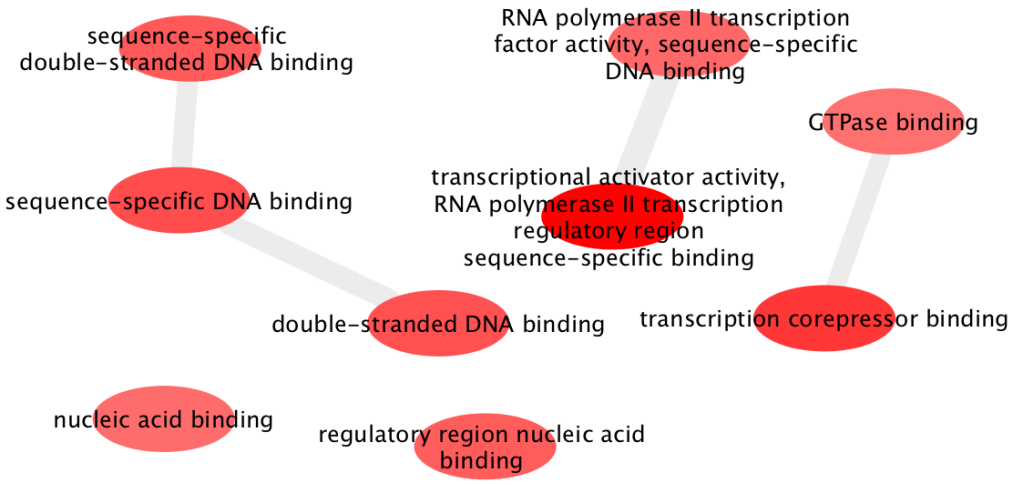
Table S8.2. 50 genes with lowest log2 fold change in expression, downregulated in differential expression analysis, comparing *C9orf72*-TRAP and Control-TRAP samples.

C9orf72-TRAP vs. Control-TRAP Bottom log2 Fold Change Genes				
Gene Name	**Ensembl**	**base Mean**	**log2 Fold Change**	**p-adj**
HIST1H3D	ENSG00000197409	34.09	-14.96	2.37E-03
NA	ENSG00000283299	25.18	-12.57	4.79E-02
VIP	ENSG00000146469	18.71	-6.49	1.44E-02
CYP24A1	ENSG00000019186	59.27	-6.03	9.29E-03
HAND2	ENSG00000164107	128.06	-5.75	1.29E-02
KIAA0226L	ENSG00000102445	95.12	-4.46	1.22E-04
MINOS1-NBL1	ENSG00000270136	37.62	-4.44	3.52E-02
PAX7	ENSG00000009709	87.92	-4.39	1.80E-03
FOXD3	ENSG00000187140	31.85	-4.30	1.13E-02
POU4F1	ENSG00000152192	84.05	-4.13	7.24E-06
SLC17A6	ENSG00000091664	16.41	-3.86	5.14E-04
TLX2	ENSG00000115297	201.43	-3.72	5.82E-04
TLR7	ENSG00000196664	37.68	-3.69	1.17E-04
INSC	ENSG00000188487	48.74	-3.57	1.13E-02
TSPYL5	ENSG00000180543	84.09	-3.54	3.91E-04
POSTN	ENSG00000133110	289.40	-3.52	1.08E-03
MT-ND1	ENSG00000198888	20.04	-3.47	1.77E-03
SHOX2	ENSG00000168779	162.94	-3.43	2.56E-03
MT-CO3	ENSG00000198938	92.95	-3.39	6.26E-04
HR	ENSG00000168453	24.84	-3.32	9.32E-03
TBX18	ENSG00000112837	16.07	-3.26	7.76E-03
MT-ND5	ENSG00000198786	49.73	-3.25	2.27E-04
TAC3	ENSG00000166863	35.41	-3.14	8.73E-03
HMX1	ENSG00000215612	92.03	-3.13	9.41E-04
MT-ND4	ENSG00000198886	37.47	-3.13	2.52E-04
ANKRD20A4	ENSG00000172014	234.41	-3.11	6.15E-05
DRGX	ENSG00000165606	114.11	-3.10	5.65E-04
TFAP2A	ENSG00000137203	36.59	-3.02	1.36E-02
CDH19	ENSG00000071991	25.59	-3.01	2.30E-02
MT-CYB	ENSG00000198727	42.37	-3.00	2.04E-03
MT-ND2	ENSG00000198763	23.16	-2.94	4.93E-03
HTR3A	ENSG00000166736	39.77	-2.87	1.73E-02
ZNF559	ENSG00000188321	260.52	-2.83	6.72E-04
UNCX	ENSG00000164853	51.16	-2.77	1.06E-02
HEYL	ENSG00000163909	151.33	-2.73	1.08E-02
MT-ATP6	ENSG00000198899	19.58	-2.72	1.79E-02
MTRNR2L12	ENSG00000269028	112.87	-2.70	3.49E-02
MPZ	ENSG00000158887	41.97	-2.69	1.65E-02
ADGRF4	ENSG00000153294	42.81	-2.64	2.80E-02
PPEF2	ENSG00000156194	92204.16	-2.59	9.13E-07
MT-CO2	ENSG00000198712	40.25	-2.59	4.14E-03
MYO7B	ENSG00000169994	14.39	-2.58	1.55E-02
MT-CO1	ENSG00000198804	266.31	-2.57	5.38E-03
ERBB3	ENSG00000065361	66.89	-2.55	9.06E-05
CAT	ENSG00000121691	169.51	-2.54	1.53E-02
SEMA3G	ENSG00000010319	37.94	-2.50	2.12E-02
PAMR1	ENSG00000149090	164.15	-2.48	4.05E-02
DQX1	ENSG00000144045	112.91	-2.47	8.35E-03
TBX1	ENSG00000184058	25.29	-2.46	2.76E-02
ZNF558	ENSG00000167785	282.40	-2.44	8.91E-09

Figure S8.3. The top 250 most significant genes from differential expression analysis comparing *C9orf72*-TRAP and Control-TRAP samples were used to explore the gene ontology (GO) terms associated with them.

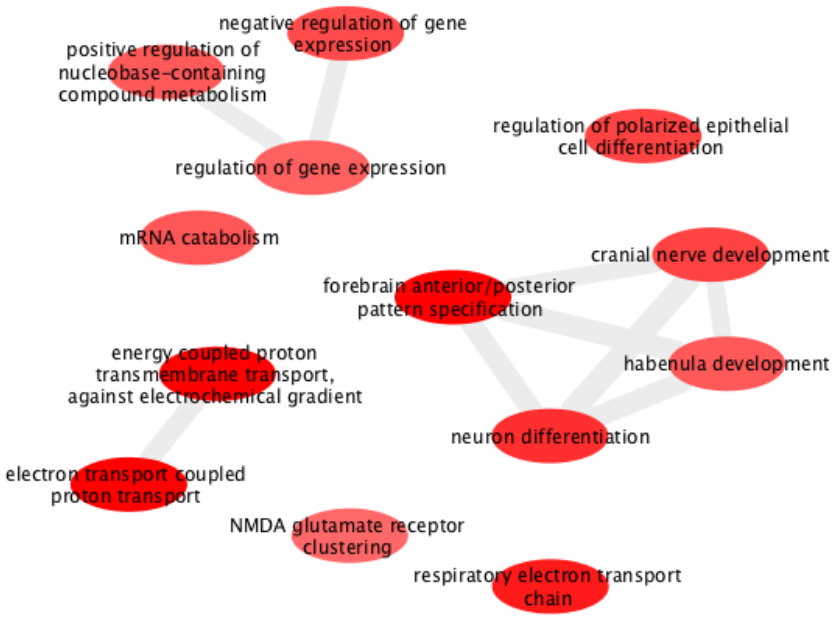
a

GO term MF; description – 250 most significant genes (padj < 0.05)	p-value	FDR q-value
GO:0001228; DNA-binding transcription activator activity, RNA polymerase II-specific	2.05E-06	9.29E-03
GO:0001222; transcription corepressor binding	3.37E-05	7.65E-02
GO:0000976; transcription regulatory region sequence-specific DNA binding	7.90E-05	1.19E-01
GO:0043565; sequence-specific DNA binding	9.89E-05	1.12E-01
GO:0003690; double-stranded DNA binding	1.27E-04	1.16E-01
GO:1990837; sequence-specific double-stranded DNA binding	1.89E-04	1.43E-01
GO:0044212; transcription regulatory region DNA binding	2.16E-04	1.40E-01
GO:0001067; regulatory region nucleic acid binding	2.20E-04	1.25E-01
GO:0000977; RNA polymerase II regulatory region sequence-specific DNA binding	2.43E-04	1.22E-01
GO:0001012; RNA polymerase II regulatory region DNA binding	2.72E-04	1.23E-01
GO:0000981; DNA-binding transcription factor activity, RNA polymerase II-specific	4.21E-04	1.74E-01
GO:0003676; nucleic acid binding	5.06E-04	1.91E-01
GO:0051020; GTPase binding	6.31E-04	2.20E-01



b

GO term BP; description – 250 most significant genes (padj < 0.05)	P-value	FDR q-value
GO:0015988; energy coupled proton transmembrane transport, against electrochemical gradient	8.55E-06	1.31E-01
GO:0015990; electron transport coupled proton transport	8.55E-06	6.53E-02
GO:0021797; forebrain anterior/posterior pattern specification	8.55E-06	4.35E-02
GO:0022904; respiratory electron transport chain	2.81E-05	1.07E-01
GO:0030182; neuron differentiation	7.31E-05	2.23E-01



c

GO term CC; description – 250 most significant genes (padj < 0.05)	P-value	FDR q-value
GO:0044424; intracellular part	3.40E-06	6.62E-03
GO:0098803; respiratory chain complex	6.27E-06	6.12E-03
GO:0005743; mitochondrial inner membrane	9.67E-05	6.29E-02
GO:0019866; organelle inner membrane	2.89E-04	1.41E-01
GO:0070069; cytochrome complex	5.09E-04	1.98E-01
GO:0030285; integral component of synaptic vesicle membrane	6.60E-04	2.15E-01
GO:0098800; inner mitochondrial membrane protein complex	9.72E-04	2.71E-01

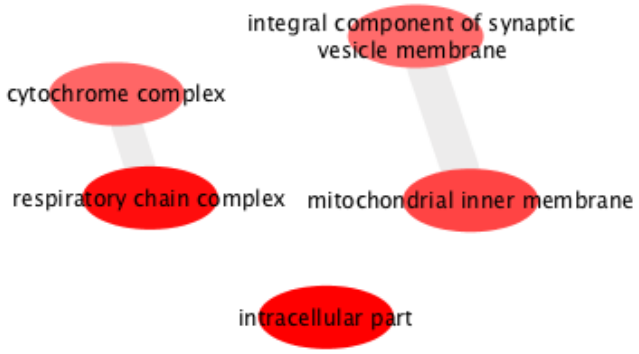


Figure S8.3. The top 250 most significant genes (padj < 0.05) from differential expression analysis comparing C9orf72-TRAP and Control-TRAP samples were used to explore the gene ontology (GO) terms associated with them. GO terms for molecular function are listed with the likelihood of enrichment and false discovery rate (FDR). False discovery rates (FDR) of 0.25 were kept as threshold for positive results, where at least a 3 in 4 probability of a term being correct is estimated, rather than occurring by chance, as calculated by the Benjamini-Hochberg procedure. The relationship between these terms are given in terms of a network for genes associated with a molecular function (MF) (a), biological pathway (BP) (b) or cellular component (CC) (c).

Table S8.3. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for molecular function.

FGSEA: Gene Ontology Molecular Function		
Chromatin binding	**Poly A RNA binding**	**RNA binding**
CENPF	HDGF	HDGF
HMGN2	HMGB3	HMGB3
JARID2	HMGN2	HMGN2
NKX2-2	HNRNPAB	HNRNPAB
NKX6-1	RPL3	POLR2D
SMARCB1	RPL19	PPP5C
SP3	RPL23A	QARS
WHSC1	RPS8	RPL3
PCGF2	RPS25	RPL19
NELFE	RPS27	RPL23A
SOX14	SNRPD2	RPL38
LHX2	CCT3	RPL41
SETDB1	NELFE	RPS8
RUVBL2	EIF3G	RPS25
VAX1	H1FX	SNRPD2
MBD3	QKI	CCT3
PHF12	RBM8A	NELFE
NCOA5	HNRNPDL	EIF3G
PYGO2	DDX39A	H1FX
PHF21B	RPL35	LRRFIP1
RNF168	PDAP1	QKI
APITD1	RALY	RBM8A
	ZC3H7B	HNRNPDL
	ZNF346	DDX39A
	LSM3	RPL35
	RRP7A	PDAP1
	RBM15B	RALY
	MRPS7	ZC3H7B
	NUSAP1	ZNF346
	SF3B6	LSM3
	ALKBH5	RRP7A
	MRPL20	RBM15B
	NCOA5	MRPS7
	NUCKS1	NUSAP1
	MRPS15	SF3B6
	RBM42	LSM7
	ISG20L2	ALKBH5
	ZCCHC7	MRPL20
	C7orf50	RBM23
	CCDC124	RNPC3
	C1orf131	NCOA5
	RC3H1	NUCKS1
	MEX3D	MRPS15
		RBM42
		ISG20L2
		ZCCHC7
		C7orf50
		CCDC124
		C1orf131
		DIS3L2
		RC3H1
		KHDRBS2

Table S8.4. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for biological pathway.

FGSEA: Gene Ontology Biological Pathway						
Cell cycle process	**Cell cycle**	**Cell division**	**Cellular macromolecular complex assembly**	**Chromosome organization**	**Cytoskeleton organization**	**DNA metabolic process**
BIRC5	BIRC5	CENPF	CENPF	BIRC5	CETN3	CBS
CALM1	CALM1	CETN3	HIP1	CENPF	EML1	CENPF
CALM3	CALM3	GNAI2	POLR2D	HMGB3	RANBP1	HMGB3
CENPF	CENPF	NEK4	RPL3	JARID2	ZNF135	POLR2D
CETN3	CETN3	TIMELESS	RPL23A	MYB	NCK2	POLR2F
EML1	EML1	RAB11FIP3	RPL38	PHF2	TRIP10	POLR2I
ENSA	ENSA	CKAP5	RPS27	PPP5C	CKAP5	PPP5C
RAB8A	GNAI2	ARHGEF11	SNRPD2	RPS27	ARHGEF11	PTMS
MYBL2	RAB8A	MAEA	SOX12	SMARCB1	IQSEC1	SMARCB1
PPP5C	MYBL2	SPG20	HIST2H2BE	SOX1	MAEA	TCF7L2
PRCC	PPP5C	ANAPC15	NCK2	TCF7L2	TTLL5	WHSC1
PSMC5	PRCC	VPS4A	EIF3G	VPS72	DNAH1	TIMELESS
PSMD3	PSMC5	NUSAP1	H1FX	WHSC1	DNAI1	HERC2
RANBP1	PSMD3	SPDL1	DNAH1	PCGF2	VPS4A	RUVBL2
RPS27	RANBP1	CDCA8	DNAI1	HIST2H2BE	SH3KBP1	COPS6
NEK4	RPS27	CENPJ	LSM3	PER2	NUSAP1	EHMT1
TCF7L2	SMARCB1	CEP63	RRP7A	H1FX	SPDL1	RNF168
TFDP1	NEK4	PSRC1	MRPS7	CDYL	MAP7D1	APITD1
TIMELESS	TCF7L2	KNSTRN	NDE1	CKAP5	CENPJ	
MAPKAPK2	TFDP1	CCDC124	MRPL20	SETDB1	PSRC1	
RAB11FIP3	TIMELESS	HAUS1	CENPJ	RUVBL2	HAUS1	
CKAP5	MAPKAPK2	APITD1	PREX1	COPS6	FGD4	
ARHGEF11	RAB11FIP3		NDUFAF6	VPS4A	SRGAP2C	
NEK6	RASSF2		RC3H1	NUSAP1		
ANAPC15	CKAP5		APITD1	MBD3		
VPS4A	ARHGEF11		HIST2H2BF	NDE1		
NUSAP1	MAEA			SPDL1		
NDE1	NEK6			CDCA8		
SPDL1	ANAPC15			EHMT1		
CDCA8	VPS4A			KAT8		
CENPJ	GMNN			PSRC1		
INTS3	NUSAP1			KNSTRN		
CEP63	NDE1			DIS3L2		
PSRC1	SPDL1			RNF168		
KNSTRN	CDCA8			KANSL1		
GADD45GIP1	CENPJ			APITD1		
HAUS1	INTS3			HIST2H2BF		
DIS3L2	CEP63					
APITD1	PSRC1					
	KNSTRN					
	GADD45GIP1					
	CCDC124					
	HAUS1					
	DIS3L2					
	APITD1					

Table S8.4. continued.

FGSEA: Gene Ontology Biological Pathway							
Microtubule based process	**Microtubule cytoskeleton organization**	**Mitotic cell cycle**	**Mitotic nuclear division**	**Organelle assembly**	**Peptidyl amino acid modification**	**Peptidyl lysine modification**	**RNA processing**
BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	POLR2D
CETN3	CETN3	CALM1	CENPF	CENPF	DHPS	DHPS	POLR2F
EML1	EML1	CALM3	CETN3	RAB8A	ERBB3	SP3	POLR2I
MYBL2	MYBL2	CENPF	ENSA	MYBL2	RELN	WHSC1	RPL3
RANBP1	RANBP1	CETN3	MYBL2	RPL3	SP3	PCGF2	RPL19
CKAP5	CKAP5	EML1	PPP5C	RPL23A	TEK	HERC2	RPL23A
NEK6	NEK6	ENSA	NEK4	RPL38	WHSC1	CDYL	RPL38
TTLL5	TTLL5	RAB8A	TIMELESS	RPS27	PCGF2	SETDB1	RPL41
SRGAP2	SRGAP2	MYBL2	CKAP5	NEK6	HERC2	RUVBL2	RPS8
DNAH1	DNAH1	PPP5C	NEK6	CEP162	MAPKAPK2	MBD3	RPS25
DNAI1	DNAI1	PRCC	ANAPC15	DZIP1	CDYL	CDCA8	SNRPD2
VPS4A	VPS4A	PSMC5	VPS4A	CLUAP1	SETDB1	EHMT1	BUD31
NUSAP1	NUSAP1	PSMD3	NUSAP1	TTLL5	NEK6	KAT8	QKI
SPA17	NDE1	RANBP1	NDE1	DNAH1	RUVBL2	RNF168	RBM8A
NDE1	SPDL1	NEK4	SPDL1	DNAI1	TTLL5		HNRNPDL
SPDL1	MAP7D1	TFDP1	CDCA8	INTU	MBD3		DDX39A
MAP7D1	CENPJ	TIMELESS	CEP63	VPS4A	CDCA8		RPP38
CENPJ	CEP63	CKAP5	PSRC1	LSM3	AGBL5		RPL35
CEP63	PSRC1	NEK6	KNSTRN	RRP7A	EHMT1		RALY
PSRC1	KNSTRN	ANAPC15	HAUS1	MRPS7	KAT8		LSM3
KNSTRN	HAUS1	VPS4A	DIS3L2	AHI1	RNF168		RRP7A
HAUS1	SRGAP2C	NUSAP1	APITD1	MRPL20	KANSL1		SSU72
TUBB8		NDE1		CENPJ	ZDHHC21		RBM15B
SRGAP2C		SPDL1		CEP63			SF3B6
		CDCA8		FAM161A			LSM7
		CENPJ		CEP89			ALKBH5
		INTS3		HAUS1			RBM23
		CEP63		RC3H1			RNPC3
		PSRC1					INTS3
		KNSTRN					ISG20L2
		HAUS1					DIS3L2
		DIS3L2					
		APITD1					

Table S8.4. continued 2.

FGSEA: Gene Ontology Biological Pathway							
Microtubule based process	**Microtubule cytoskeleton organization**	**Mitotic cell cycle**	**Mitotic nuclear division**	**Organelle assembly**	**Peptidyl amino acid modification**	**Peptidyl lysine modification**	**RNA processing**
BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	BIRC5	POLR2D
CETN3	CETN3	CALM1	CENPF	CENPF	DHPS	DHPS	POLR2F
EML1	EML1	CALM3	CETN3	RAB8A	ERBB3	SP3	POLR2I
MYBL2	MYBL2	CENPF	ENSA	MYBL2	RELN	WHSC1	RPL3
RANBP1	RANBP1	CETN3	MYBL2	RPL3	SP3	PCGF2	RPL19
CKAP5	CKAP5	EML1	PPP5C	RPL23A	TEK	HERC2	RPL23A
NEK6	NEK6	ENSA	NEK4	RPL38	WHSC1	CDYL	RPL38
TTLL5	TTLL5	RAB8A	TIMELESS	RPS27	PCGF2	SETDB1	RPL41
SRGAP2	SRGAP2	MYBL2	CKAP5	NEK6	HERC2	RUVBL2	RPS8
DNAH1	DNAH1	PPP5C	NEK6	CEP162	MAPKAPK2	MBD3	RPS25
DNAI1	DNAI1	PRCC	ANAPC15	DZIP1	CDYL	CDCA8	SNRPD2
VPS4A	VPS4A	PSMC5	VPS4A	CLUAP1	SETDB1	EHMT1	BUD31
NUSAP1	NUSAP1	PSMD3	NUSAP1	TTLL5	NEK6	KAT8	QKI
SPA17	NDE1	RANBP1	NDE1	DNAH1	RUVBL2	RNF168	RBM8A
NDE1	SPDL1	NEK4	SPDL1	DNAI1	TTLL5		HNRNPDL
SPDL1	MAP7D1	TFDP1	CDCA8	INTU	MBD3		DDX39A
MAP7D1	CENPJ	TIMELESS	CEP63	VPS4A	CDCA8		RPP38
CENPJ	CEP63	CKAP5	PSRC1	LSM3	AGBL5		RPL35
CEP63	PSRC1	NEK6	KNSTRN	RRP7A	EHMT1		RALY
PSRC1	KNSTRN	ANAPC15	HAUS1	MRPS7	KAT8		LSM3
KNSTRN	HAUS1	VPS4A	DIS3L2	AHI1	RNF168		RRP7A
HAUS1	SRGAP2C	NUSAP1	APITD1	MRPL20	KANSL1		SSU72
TUBB8		NDE1		CENPJ	ZDHHC21		RBM15B
SRGAP2C		SPDL1		CEP63			SF3B6
		CDCA8		FAM161A			LSM7
		CENPJ		CEP89			ALKBH5
		INTS3		HAUS1			RBM23
		CEP63		RC3H1			RNPC3
		PSRC1					INTS3
		KNSTRN					ISG20L2
		HAUS1					DIS3L2
		DIS3L2					
		APITD1					

Table S8.5. List of genes associated with fast gene set enrichment analysis (FGSEA) gene ontology terms for cellular components.omponents.

FGSEA: Gene Ontology Cellular Components				
Catalytic complex	**Cell projection part**	**Centrosome**	**Chromosome**	**Cilium**
ATP1A2	ATP1A2	BIRC5	BIRC5	CENPF
JARID2	ATP2B4	CALM1	CENPF	RAB8A
POLR2D	CALM1	CALM3	HMGB3	NEK4
POLR2I	CALM3	CENPF	HMGN2	CEP162
SALL2	CENPF	CETN3	MYBL2	DZIP1
SNRPD2	GRIA1	GNAI2	PHF2	CLUAP1
PCGF2	RAB8A	RAB8A	SMARCB1	TTLL5
DCAF5	PDE9A	NUBP1	TCF7L2	IQCE
RPP38	PPP5C	RANBP1	WHSC1	DNAH1
RUVBL2	PTPRN2	HERC2	PCGF2	DNAI1
ANAPC15	PEX19	MAPKAPK2	HIST2H2BE	CEP89
DNAH1	NCK2	RAB11FIP3	BUD31	
DNAI1	CEP162	CKAP5	TIMELESS	
KAT8	DZIP1	CEP162	H1FX	
RNF168	CLUAP1	DZIP1	SETDB1	
	SRGAP2	CLUAP1	RUVBL2	
	DNAH1	TTLL5	NUSAP1	
	DNAI1	VPS4A	MBD3	
	ARHGEF4	CCHCR1	SPDL1	
	PLEKHO1	AHI1	CDCA8	
	SPA17	NDE1	PHF12	
	AHI1	CENPJ	EHMT1	
	PREX1	RFWD2	KAT8	
	FAM161A	CEP63	KNSTRN	
	CEP89	FAM161A	RNF168	
	SPATA13	CEP89	APITD1	
		HAUS1	HIST2H2BF	

Table S8.5. continued.

FGSEA: Gene Ontology Cellular Components				
Cytoskeletal part	**Cytoskeleton**	**Intracellular vesicle**	**Membrane region**	**Microtubule cytoskeleton**
BIRC5	ARHGDIA	CD59	ARF4	BIRC5
CALM1	CALM1	COPB1	ATP2B4	CALM1
CALM3	CENPF	GCH1	CASP3	CALM3
CENPF	CETN3	GRIA1	ERBB3	CENPF
CETN3	EML1	GRM4	GRIA1	CETN3
EML1	GNAI2	MME	GRM4	EML1
GNAI2	HIP1	PTPRN2	HTR3A	GNAI2
RAB8A	MCF2	TIMP3	MPZ	RAB8A
NUBP1	RAB8A	ZP3	PDE9A	NUBP1
RANBP1	RANBP1	DYSF	PEX19	RANBP1
SH3GL1	SH3GL1	HAX1	ZP3	CCT3
TEK	TEK	MMRN1	LAMTOR3	LRP8
CCT3	LRP8	SYT11	ADGRG2	TIMELESS
LRP8	TIMELESS	SRGAP2	SRGAP2	HERC2
HERC2	HERC2	RHOQ	RHOQ	MAPKAPK2
MAPKAPK2	LRRFIP1	LAMP5	LAMP5	RAB11FIP3
RAB11FIP3	TRIP10	NPC1L1	NPC1L1	CKAP5
CKAP5	RAB11FIP3	TLR7	ARHGEF4	MAEA
MAEA	CKAP5	PI4K2A	PLLP	NEK6
NEK6	MAEA	SLC17A6	PLEKHO1	CEP162
CEP162	CEP162	CRISPLD2	PI4K2A	DZIP1
DZIP1	DZIP1	ARRDC4	AJAP1	CLUAP1
CLUAP1	CLUAP1	SYT9	CPEB1	TTLL5
TTLL5	TTLL5		SLC29A4	DNAH1
DNAH1	MPRIP		RGMB	DNAI1
DNAI1	SEPT8			VPS4A
VPS4A	DNAH1			NUSAP1
NUSAP1	DNAI1			CCHCR1
CCHCR1	VPS4A			AHI1
SPDL1	SH3KBP1			NDE1
CDCA8	NUSAP1			SPDL1
MAP7D1	CCHCR1			CDCA8
CENPJ	SPDL1			MAP7D1
AGBL5	CDCA8			CENPJ
RFWD2	MAP7D1			AGBL5
CEP63	CENPJ			RFWD2
FAM161A	NCOA5			CEP63
PSRC1	CEP63			FAM161A
CEP89	PSRC1			PSRC1
KNSTRN	CEP89			CEP89
CCDC124	KNSTRN			KNSTRN
HAUS1	CCDC124			CCDC124
CCDC96	HAUS1			HAUS1
TUBB8	FGD4			CCDC96
	CCDC96			TUBB8
	TUBB8			

Table S8.5. continued 2.

FGSEA: Gene Ontology Cellular Components						
Microtubule organizing center	**Microtubule**	**Nucleolus**	**Nucleoplasm part**	**Ribonucleoprotein complex**	**Spindle**	**Transferase complex**
BIRC5	BIRC5	CETN3	CDK8	HNRNPAB	BIRC5	JARID2
CALM1	CALM1	RAB8A	ELF4	POLR2D	CALM1	POLR2D
CALM3	CALM3	MXI1	JARID2	RPL3	CALM3	POLR2I
CENPF	EML1	PHF2	POLR2D	RPL19	CENPF	SNRPD2
CETN3	CCT3	POLR2I	POLR2F	RPL23A	EML1	PCGF2
GNAI2	CKAP5	RPL3	POLR2I	RPL38	CKAP5	DCAF5
RAB8A	NEK6	RPL19	SALL2	RPL41	MAEA	RUVBL2
NUBP1	CEP162	RPL23A	SP3	RPS8	NEK6	ANAPC15
RANBP1	TTLL5	RPS25	TCF7L2	RPS25	CEP162	KAT8
HERC2	DNAH1	SMARCB1	PCGF2	RPS27	VPS4A	RNF168
MAPKAPK2	DNAI1	PER2	NELFE	SNRPD2	NUSAP1	
RAB11FIP3	NUSAP1	TIMELESS	RBM8A	EIF3G	SPDL1	
CKAP5	CENPJ	RPP38	NEK6	BUD31	CDCA8	
NEK6	PSRC1	RBM15B	RUVBL2	RBM8A	MAP7D1	
CEP162	KNSTRN	NUSAP1	MBD3	HNRNPDL	AGBL5	
DZIP1	HAUS1	CDCA8	ALKBH5	DDX39A	CEP63	
CLUAP1	TUBB8	RNF213	PHF12	RPP38	PSRC1	
TTLL5		ISG20L2	RFWD2	RUVBL2	CEP89	
VPS4A		ZCCHC7	INTS3	RPL35	KNSTRN	
CCHCR1		CRACR2A	KAT8	RALY	HAUS1	
AHI1			APITD1	LSM3		
NDE1				RRP7A		
SPDL1				MRPS7		
CENPJ				SF3B6		
RFWD2				LSM7		
CEP63				MRPL20		
FAM161A				RNPC3		
CEP89				MRPS15		
CCDC124				MRPS34		
HAUS1				SNRNP25		
CCDC96				GADD45GIP1		
				DIS3L2		
				RC3H1		