Systems view of neuronal mRNA localisation -
comparative analysis and determinants of the sub-cellular neuronal transcriptome

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Nicolai von Kügelgen

August, 2022

The following work was performed from September 2017 until February 2022 under the supervision of Dr. Marina Chekulaeva at the Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, Hannoversche Str. 28, 10115 Berlin-Mitte.

1st reviewer: Dr. Marina Chekulaeva
Berlin Institute for Medical Systems Biology (BIMSB)
Max-Delbrück-Centrum für Molekulare Medizin
Hannoversche Str. 28, 10115 Berlin

2nd reviewer: Prof. Dr. Irmtraud Meyer
Freie Universität Berlin, Special Professor
Berlin Institute for Medical Systems Biology (BIMSB)
Max-Delbrück-Centrum für Molekulare Medizin
Hannoversche Str. 28, 10115 Berlin

Date of the defense: 10.03.2023

Acknowledgements

Science is a team effort. As such, even though I present my own work here, it would never have been possible without leaning on previous works or support by colleagues and friends. A lot of thanks are due to my supervisor Marina Chekulaeva and all members of her group, past and present, with whom I've enjoyed working: David van den Bruck, Alessandra Zappulo, Camilla Ciolli Mattioli, Didrik Olofsson, Gerard Arrey, Cora Zwiesigk, Rosa Einenkel, Nevena Mateva, Lucija Bujanic, Vsevolod 'Seva' Cherepanov, Lina Hellwig, Anna Kalinina, Gabrielle Capin, Nadja Zerna, Esteban Peguero Sánchez, Inga Lödige, Katarzyna 'Kasia' Ludwik, Christine Römer, and of course especially my fellow PhD students Guli-Nuer Imincan, Samantha Mendonsa, Sayaka Dantsuji, Artem Baranovskii and Erik Becher.
Furthermore, I am grateful to my additional thesis advisory committee members Markus Landthaler, Altuna Akalin and Irmtraud Meyer, who also served as my secondary supervisor with FU, for providing additional perspective and guidance throughout my PhD and to the group members from Altuna who answered all my bioinformatics questions, especially Bora Uyar and Vedran Franke. Of course, research also goes hand in hand with collaboration and in this vein I have to thank Igor Ulitsky and his student Maya Ron for help with the Nzip project, Cedric Bardy for providing data files and the RNA-Seq community for driving efforts to make raw data generally available. The local community from BIMSB and MDC has too many people to name explicitly, but did prove to be an excellent environment for my PhD time. Finally, I am very grateful to all family and friends who had my back in these eventful and often challenging years.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. I also confirm that my dissertation has not been accepted or rejected in any previous doctoral procedure.

Nicolai von Kügelgen

27.03.2023

List of abbreviations

A2RE	A2 response element	ALS	amyotrophic lateral sclerosis
AP	action potential	ARE	AU rich element
BH	Benjamini-Hochberg	CDE	constitutive decay element
CLIP	crosslinking immuno-precipitation	CPE	cytoplasmic polyadenylation element
DE	differential expresison	DIV*X*	*X* days in vitro
DNA	deoxyribonucleic acid	DOX	doxycycline
cDNA	complementary DNA	ECM	extracellular matrix
ER	endoplasmic reticulum	ERCC	External RNA Controls Consortium
FDR	false discovery rate	GEO	Gene Expression Omnibus
GLM	generalised liner model	GO	gene ontology
hiPSC	human induced pluripotent stem cell	iBAQ	intensity based absolute quantification
IM	induction media	iMN	induced motor neurons
LFQ	label free quantification	log2fc	log2 fold change
MM	motor neuron media	MPRA	massively parallel reporter assay
NCBI	National Center for Biotechnology Information	NGS	next generation sequencing
NIL	NGN2, ISL1 and LHX3	NMJ	neuromuscular junction
Nzip	neuronal zipcode identification protocol	PCA	principle component analysis
PCN	primary cortical neurons	PCR	polymerase chain reaction
RBP	RNA binding protein	RNA-Seq	RNA sequencing
RNA	ribonucleic acid	RNP	ribonucleoprotein particle
mRNA	messenger RNA	RUV	removal of unwanted variation
miRNA	micro RNA	TPM	transcripts per million
UMI	unique molecular identifier	UTR	untranslated region
WT	wild type		

Contents

1A Summary

Control over the intracellular localisation of RNA is an important aspect of post-transcriptional regulation, especially for highly polarised cells like neurons. The presence of specific transcripts in axons and dendrites, together neurites, is determined by *cis*-active elements called zipcodes and ultimately allows neurons to locally synthesise required proteins and adapt quickly to cues of the local environment. This capability is important for the correct function of synapse remodelling and memory formation and a disruption of RNA localisation in neurons has been associated with several neurodegenerative diseases.

Using RNA sequencing a vast number of transcripts can be detected in neurites of neuronal model systems. With over 7500 transcripts even the neurite core transcriptome, which I summarised from the published datasets generated in the last decade, contains at least half or a third of the full neuronal transcriptome. Whether all of these transcripts should be considered localised to neurites or if this designation is better determined by differential expression analysis between compartments is still difficult to answer. For one thing, no strong overlap of transcripts with localisation based on significant enrichment in many individual datasets exists and also my integrated analysis utilising batch correction did only generate a relatively small set of differentially expressed genes, which however tend to have more conserved enrichment. Secondly, several transcripts that are generally considered as classical localised transcripts, like the Actb mRNA, are not relatively enriched in neurites, even if they are strongly expressed there.

Relying on a set of transcripts with consistent neurite enrichment based on datasets from primary murine neurons I designed Nzip, a massively parallel reporter assay (MPRA) aimed at the identification of unknown zipcodes. Based on 16 candidate sequences determined from the first experiment, it was possible to identify 2 new zipcode motifs utilising a secondary library with a mutational analysis approach: the let-7 miRNA seed sequence CUACCUC and an $(AU)_n$ repeat motif. The compartmental quantification of miRNAs and associated protein machinery indicates a stronger activity of let-7 in soma, providing a potential mechanism for its zipcode activity. Additionally, also the $(AU)_n$ motif is also associated with lower read counts in soma and several identified binding proteins have known effects on RNA stability, indicating that it likely also affects RNA localisation through stability regulation.

Building on my observations that assignment of RNA localisation state based on either detection or enrichment in neurites both is problematic and that Nzip mainly identified motifs conferring neurite enrichment by RNA stability, I argue that a clear distinction between localised and not-localised transcripts may not be an accurate description of the biological system. Instead, zipcodes likely affect the probability of a given transcript to reach neurites and there may also be different mechanisms that affect the tendency for localisation as measured by enrichment or detection. Whether this is a more accurate description of RNA localisation mechanics as well as the exact functions of the zipcodes I identified

should be further investigated in future studies.

As a second part of my work I have contributed to studying the human neurodegenerative disease amyotrophic lateral sclerosis (ALS). This affliction is mainly characterised by the degradation of motor neurons usually starting at the synapses between axons and skeletal muscle, the neuromuscular junction (NMJ), and in many cases is known to be caused by mutations in several RNA binding proteins affecting RNA localisation. Among these is the FUS protein, whose mutations often disrupt its exclusive nuclear localisation and thus can lead both to a loss-of-function as well as a toxic gain-of-function trough availability of new RNA targets in the cytoplasm.

To study a disease like ALS a cellular model system for human neurons is needed, which replicates the relevant molecular signatures of affected motor neurons, specifically including the axonal containing neurite compartment. I have characterised the transcriptome and proteome of induced motor neurons (iMN) generated by expression of NGN2, ISL1 and LHX3 transcription factors. This system showed expected expression of marker genes throughout motor neuron differentiation as well as proper specification of neurite compartment and similarity with signatures of electrophysiological maturity.

Using the iMN model system I performed investigative analysis for the effect of ALS patient derived FUS mutations on the proteome and transcriptome, specifically including effects pronounced in the neurite compartment. With this I identified many differentially expressed genes already associated with ALS or FUS mutations, which, however, span a very wide field of functional associations and to my understanding are more likely linked to a disruption of normal FUS activity. However, I also observed a more consistent and rarely reported pattern of down-regulation of genes building the extracellular matrix around the NMJ, which was specifically notable in the neurites of cells with cytoplasmic localised P525L FUS. Additionally, I found a very similar pattern of down-regulation in neurites for genes passing the secretory pathway, known target transcripts of FUS, as well as those with a G-quadruplex motif, which has been identified as a potential binding site for both FUS and other ALS associated RBPs. This highlights a potential toxic gain-of-function for FUS as well as a particular pathway which may be important in the axonal degeneration in ALS. Validation of this observation including any potential significance of an overlap between the affected gene groups I identified should be the focus of further work.

1B Zusammenfassung

Die Kontrolle über die intrazelluläre Lokalisation von RNA is ein wichtiger Aspekt der post-transkriptionalen Regulation, insbesondere für stark polarisierte Zellen wie Neuronen. Die Präsenz von bestimmten RNA Molekülen in Axonen oder Dendriten, zusammen Neuriten, wird durch *cis*-aktive Elemente, sogenannte 'Zipcodes', bestimmt und erlaubt ultimativ, dass Neuronen schnell auf Reize reagieren und benötigte Proteine dort lokal synthetisieren können. Diese Fähigkeit ist essentiell, damit die Remodellierung von Synapsen und das Bilden von Erinnerungen korrekt funktioniert, weiterhin wurde die Störung von RNA-Lokalisation mit der Pathophysiologie mehrerer neurodegenerativer Erkrankungen in Zusammenhang gebracht.

Durch RNA Sequenzierung kann eine große Anzahl von Transkripten in Neuriten von Modellsystemen für Neuronen identifiziert werden. Mit über 7500 verschiedenen Transkripten enthält selbst das Kerntranskriptom, das ich aus den veröffentlichten Datensätzen der letzten Jahre zusammengestellt habe, nahezu ein Drittel oder die Hälfte des kompletten neuronalen Transkriptoms. Ob all diese Transkripte tatsächlich als lokalisiert angesehen werden sollten oder ob diese Bezeichnung besser durch differenzielle Expression zwischen Zellfraktionen bestimmt werden sollte, ist schwer zu beurteilen. Einerseits gibt es keine klare Übereinstimmung von Transkripten mit differenzieller Lokalisation zwischen einzelnen Datensätzen und auch mit meiner integrativen Datenanalyse inklusive Batch-Korrektur konnte ich nur ein relativ kleines Set an differenziell exprimierten Transkripten identifizieren. Zum anderen gibt es einige Transkripte, die im allgemeinen als lokalisiert angesehen werden, wie beispielsweise die Actb mRNA, die allerdings trotzdem keine relative erhöhte Expression in Neuriten aufzeigen, auch wenn sie dort stark exprimiert sind.

Mit einem Set von in primären murinen Neuronen konsistent Neurit-angereicherten Transkripten habe ich 'Nzip' konstruiert, ein massiv paralles Reporter Experiment mit dem Ziel unbekannte Zipcodes zu identifizieren. Mit 16 potentiellen Sequenzen, die in einem ersten Experiment entdeckt wurden und einem Mutations-Analyse Ansatz war es möglich 2 neue Zipcodes zu bestimmen: die let-7 miRNA Zielsequenz CUACCUC und ein $(AU)_n$ Motiv. Die Zellfraktion-spezifische Quantifizierung von miRNAs und ihrer assoziierten Proteinmaschinerie legt zusätzlich nahe, dass die Aktivität von let-7 im Zellkörper stärker ist, was einen potentiellen Mechanismus für diese Zipcode Aktivität liefert. Weiterhin wurde auch für das $(AU)_n$ Motiv eine ähnliche Reduzierung von RNA Molekülen in Zellkörper festgestellt und einige der dazu identifizierten Bindeproteine können die Stabilität von RNA beeinflussen, sodass dieses Motif RNA Lokalisation vermutlich ebenfalls über Stabilitätsregulation kontrolliert.

Aufbauend auf meinen Beobachtungen, dass die Designation des RNA Lokalisationsstatus basierend auf entweder Detektion oder Anreicherung in Neuriten problematisch ist, und dass Nzip primär Motive identifiziert hat, die RNA Anreicherung durch Stabilität kontrollieren, bin ich der Auffassung, dass

eine klare Unterscheidung zwischen lokalisierten und nicht lokalisierten Transkripten nicht unbedingt einer akkuraten Beschreibung entspricht. Stattdessen halte ich es für zutreffender, dass Zipcodes die Wahrscheinlichkeit beeinflussen, dass eines gegebenes Transkript den Zellkörper verlässt, und dass potentiell unterschiedliche Mechanismen die Präsenz und Anreicherung von Transkripten in Neuriten kontrollieren. Ob dieses Modell von RNA Lokalisation korrekt ist sowie die exakte Funktionsweise der entdeckten Zipcodes, muss allerdings noch durch weitere Untersuchungen kontrolliert werden.

Der zweiten Teil meiner Arbeit widmet sich der Erforschung der neurodegenerativen Erkrankung Amyotrophe Lateralsklerose (ALS). Diese zeichnet sich vor allem dadurch aus, dass Motorneuronen beginnend an den Synapsen zwischen Axonen und Muskeln, der neuromuskulären Endplatte (NME), degenerieren. Weiterhin gibt es RNA bindende Proteine, deren Mutationen die Krankheit auslösen können. Unter diesen befindet sich das FUS Protein, dessen Mutation oft seine Lokalisation im Zellkern beeinträchtigt und damit sowohl zu funktionalem Verlust als auch zu neuen toxischen Funktionen durch das Binden anderer RNA Moleküle im Zytoplasma führen kann.

Um eine Krankheit wie ALS zu erforschen ist ein zelluläres Modellsystem für humane Neuronen erforderlich, das relevante molekulare Signaturen der betroffenen Motorneuronen und insbesondere der Axon- und Neurit-Fraktion abbildet. Ich habe das Transkriptom und Proteom von induzierten Motorneuronen (iMN) charakterisiert, die durch die Expression von NGN2, ISL1 und LHX3 generiert werden. Dieses System zeigt die während der Motorneuron-Differenzierung erwarteten Expressionsmuster sowie klar spezifizierte Neurite und Ähnlichkeit zu elektrisch aktiven Signaturen.

Mit dem iMN System habe ich eine erste Untersuchung der Effekte von FUS Mutationen aus ALS Patienten auf das Proteom und Transkriptom und insbesondere die Neurit Fraktion durchgeführt. Dabei habe ich einige differentiell exprimierte Gene gefunden, die bereits mit ALS oder FUS Mutationen assoziert wurden, allerdings auch ein sehr breites funtionales Spektrum umfassen und daher vermutlich mit der Störung von normaler FUS Funktion zusammenhängen. Zusätzlich habe ich aber auch eine konsistente und bisher weniger beachtete Expressionsreduktion von Genen der extrazellulären Matrix nahe der NME beobachtet, die insbesondere in Neuriten von Zellen mit zytoplasmatischem P525L FUS präsent ist. Weiterhin habe ich ähnliche Muster mit reduzierter Expression von den Genen gefunden, die den Sekretionsweg passieren, bekannte Bindeziele von FUS sind, oder ein G-Quadruplex Motiv besitzen, wobei letzteres als potentielle Bindestelle für FUS und andere ALS assoziierte Proteine identifiziert wurde. Auch wenn diese Beobachtungen einen potentiellen toxischen Funktionsgewinn für FUS darstellen und einen bestimmten molekularen Pfad hervorheben, der wichtig für den Verlauf von ALS sein könnte, müssen sie noch durch weitere Studien verifiziert werden, da insbesondere die Signifikanz einer Überschneidung der von mir identifizierten betroffenen Gen-Gruppen bisher nicht klar ist .

2 Introduction

The distribution of individual RNA molecules to specific places within a cell, or the localisation of RNA, plays an important role in many cellular functions given that it affects several aspects of RNA biology ranging from accessibility of the RNA to locally available factors to the origin point of translation and new proteins (Chin & Lécuyer, 2017; Lécuyer et al., 2007). It is therefore not surprising that specific control of RNA localisation is a common occurrence not only in all types of eukaryotes (Das et al., 2021; Engel et al., 2020) but even in bacteria (Nevo-Dinur et al., 2011). Of course, the well studied nuclear export of mRNA and non-coding transcripts is the most common occurrence of RNA transport across cellular compartments, even if this aspect of RNA biology is generally considered as a part of transcription and RNA maturation rather than RNA localisation (Rodriguez et al., 2004; Williams et al., 2018). However, the nuclear localisation of both coding and non-coding transcripts, mostly through retention in the nucleus, has recently also emerged as an area of research interest in RNA localisation (Halpern et al., 2015; Lubelsky & Ulitsky, 2018).

While sub-cellular localisation of RNA to other parts of the cell has been studied for a long time, many aspects of it are still not fully understood. Especially in animals, RNA localisation occurs throughout different stages of development from embryonic patterning controlled by specific localisation of maternal mRNA in the oocyte and asymmetric cell division (Medioni et al., 2012), over localisation of transcripts to cellular organelles (Fazal et al., 2019; Kraut-Cohen et al., 2013) to RNA localisation in axons and dendrites of neurons (Costa et al., 2021; Engel et al., 2020). Based on the studies of all these systems several mechanisms that control RNA localisation have been discovered and can be summarised into three broad categories. First, RNA can be enriched at specific points in the cell by anchoring the transcripts there. This mechanism is best described in the oocytes and embryos of flies (Trcek & Lehmann, 2019), but also occurs in others systems and can rely both on passive diffusion and active transport (Becalska & Gavis, 2009; Forrest & Gavis, 2003). Secondly, the protection from degradation or preferential degradation of RNA in a specific compartment can also lead to local enrichment of transcripts. However, this mechanism also has so far mainly been shown to be utilised in developing *drosophila melanogaster* (L. Chen et al., 2014) with some evidence of nonsense mediated decay controlling transcripts in neuronal outgrowths (Colak et al., 2013; Notaras et al., 2020) and also still depends on additional transport mechanisms for transcripts to reach their destination. And finally the third mechanism to achieve RNA localisation is the active transport of transcripts usually along the actin or microtubule cytoskeleton (de Heredia & Jansen, 2004), or possibly alongside vesicles (Cioni et al., 2019; Liao et al., 2019). Since such transport of RNA has to be mediated by motor proteins, it occurs through the transport of ribonucleoprotein particles (RNPs). The amount of transcripts an average RNP contains is still a matter of debate with evidence for both multiplexing of different transcripts (Gao et al., 2008) as well as claims of only single

molecule transcript transport (Batish et al., 2012). Additionally, it is generally assumed that transcripts inside transport RNPs are translationaly repressed (Besse & Ephrussi, 2008; Pimentel & Boccaccio, 2014; Wells, 2006), which has specifically been shown for RNPs incorporating fragile X mental retardation protein (FMRP) (Darnell et al., 2011) or pumilio (PUM) protein (Zahr et al., 2018).
Overall there are several mechanisms and many more individual factors that influence RNA localisation in the cytoplasm and together they all contribute to a system that can be quite complex even for individual transcripts, but is of course not the same between different ones. While RNA localisation patterns play important roles in multiple cell types (Das et al., 2021; Engel et al., 2020), some of the few well described localisation machineries have been shown to work similarly across different cellular systems. One of the best examples for this is the IGF2BP1 (also known as ZBP1 or IMP1) mediated localisation of β-actin (Actb) mRNA to both axonal growth cones (Bassell et al., 1998; Leung et al., 2006) and dendrites (Tiruchinapalli et al., 2003) as well as the leading edges of fibroblasts, where it was first identified as a localised RNA (Kislauskis et al., 1994; Ross et al., 1997).

2.1 Neurons as experimental model systems for sub-cellular RNA localisation

Neurons have emerged as the cellular model that is most often used to study RNA localisation, not only because their large size and complex morphology allows for comparatively easy separation of the cell body and neuronal outgrowths, but also because disruption of RNA localisation can strongly impact neuronal function and cause diseases (Mofatteh, 2021). The first experiments and observations suggesting local translation and therefore localised RNA at axons and dendritic synapses were made already over 40 years ago based on the detection of localised polysomes and translation activity (Steward & Levy, 1982; Tobias & Koenig, 1975). The first clear evidence for local presence of RNA itself came with the discovery of Map2 and Camk2a transcripts in dendrites (Burgin et al., 1990; Garner et al., 1988) followed by that of β-actin mRNA in axonal growth cones (Bassell et al., 1998; Leung et al., 2006). Experiments with neurons of the sea slug *aplysia* also showed that local translation in axonal pre-synapses can be activated by stimulation and that there exists a larger set of transcripts localised to neurites (Martin et al., 1997; Moccia et al., 2003). Initially, the numbers of different transcripts identified in neurites by these and other, mainly microarray based, studies only reached a few hundred (Lein et al., 2007; Moccia et al., 2003; Poon et al., 2006; A. M. Taylor et al., 2009; Zhong et al., 2006), but with the emergence of next generation sequencing and therefore increasingly more sensitive detection capabilities this figure soon increased first to the often cited figure of approximately 2500 transcripts (Cajigas et al., 2012; Gumy et al., 2011) and then to above ten thousand in more recent publications (Briese et al., 2016; Maciel et al., 2018; Minis et al., 2014; Rotem et al., 2017; Taliaferro et al., 2016; Zappulo et al., 2017). Similarly, these studies started focusing more on transcripts with relative enrichment in the neurite compartment and

their numbers as opposed to the absolute number of transcripts detectable at all.
Building on a large body of research, neuronal RNA localisation has been studied in several species and different neuronal subtypes, mainly cortical, hippocampal and motor neurons, either derived from primary sources or generated from stem cells. Additionally, several different techniques for the separation of neurites have been established, the most common being separation outgrowths alongside porous filter membranes (Poon et al., 2006; Zappulo et al., 2017) or the groves of microfluidic chambers (Saal et al., 2014; A. M. Taylor et al., 2009), which are only a few μm wide and are supposed to exclude the slightly larger dendrites and allow harvesting of only axons. Other approaches have analysed RNA of the neuropil from hippocampal CA1 stratum radiatum, an anatomic regions, which lacks neuronal cell bodies and mainly consist of dendrites, mixed with some axons and non-neuronal cells (Cajigas et al., 2012; Tushev et al., 2018; Zhong et al., 2006). Regardless of the approach, very few clear differences between RNA localisation to axon and dendrites have been described so far, with the exception of few individual transcripts with specific localisation, like the exclusively dendritic Map2 transcript (Garner et al., 1988). Also, since no single technique allows for the separation of both axons and dendrites, no direct comparison of the transcripts in these compartments has been possible so far. However, recent advances in proximity labelling of RNA at specific subcellular sites based on localised marker proteins might allow such studies in the future (Fazal et al., 2019). While axons and dendrites do have important functional and molecular differences, especially at the pre- and post-synapse, in my work I am focusing on RNA generally localised to all neuronal outgrowths, summarised under the term neurites, which is the most easily obtained cell fraction and encompasses all available experimental models.
Apart from bulk analysis of RNA from neurons physically separated into soma and neurites, several additional approaches have allowed further insights into the organisation of the neuronal subcellular transcriptome. Some investigations into smaller sub-compartments of neurons, like axonal growth cones or nanobiopsies of dendritic cytosol, reported only a small subset of more specific transcripts (Poulopoulos et al., 2019; Tóth et al., 2018), whereas studies of synaptosomes, which incorporate only slightly larger cell fractions, allowed the analysis of responses across nearly the whole transcriptome (B. J. Chen et al., 2017; Most et al., 2015).

Of course, the interest in localised RNA in neurons is strongly based on the expectation that this presence of transcripts is functionally relevant and allows localised generation of new proteins. Indeed, it has been shown that local translation in neurites and especially near synapses is important for neuronal functions including synaptic plasticity, memory formation and neural regeneration, but also axonal branching and even retrograde signalling to the nucleus (Cioni et al., 2018; Holt & Schuman, 2013; Tasdemir-Yilmaz & Segal, 2016). Direct evidence of local translation in neurites and these functions

comes from several studies, that used sequencing of RNA fragments covered by the ribosome during active translation (Ribo-Seq) or at least ribosome bound transcripts (Ainsley et al., 2014; Ouwenga et al., 2017; Shigeoka et al., 2016; Zappulo et al., 2017). The considerable overlap between actively translated transcript sets and generally localised transcripts (von Kügelgen & Chekulaeva, 2020) as well as indication that a large fraction of the local proteome is derived from local translation (Zappulo et al., 2017), indicate that the comparatively easier study of the localised transcriptome is a good proxy for the contribution of local translation to local protein functions.

2.2 Zipcodes and RBPs - elements controlling RNA localisation

The control over which RNA is localised to a specific compartment in a cell is generally encoded by *cis*-active elements, which are also called zipcodes, like the first localisation element identified in the β-actin mRNA (Kislauskis et al., 1994). Zipcodes, like many other *cis*-active elements, are most often found in the 3'UTR of mRNAs and usually allow the binding of associated *trans*-factors and thereby a connection to the localisation machinery (Andreassi & Riccio, 2009; Gomes et al., 2014). These *trans*-factors are most often RNA binding proteins (RBPs), which are able to bind many different transcripts with similar motifs. While considerable effort has been undertaken towards the characterisation of RBP sequence preferences (Dominguez et al., 2018; Ray et al., 2013), the resulting motifs are often short or promiscuous and therefore yield a very wide prediction of potential transcript targets. Large numbers of RBP targets were also reported by studies directly measuring the interaction of an RBP with RNA using crosslinking immunoprecipitation (CLIP), like FMRP which can bind hundreds of different transcripts (Darnell et al., 2011). Furthermore, transcripts can have multiple binding sites for the same or different RBP as well as other *trans*-factors, leading to a complex interplay of different regulation factors that can both increase binding specificity but also lead to competition for target binding (Gomes et al., 2014; Iadevaia & Gerber, 2015).
Finally, even in the case of well studied binding motifs like the 54nt zipcode of β-actin mRNA bound by IGF2BP1 (Kislauskis et al., 1994), orthogonal binding sites in other target transcripts like Gap43, which has no sequence homology to β-actin, can also lead to RNA localisation (Donnelly et al., 2011). Additionally, the binding site for IGF2BP1 in the β-actin mRNA also overlaps with a binding site for ELAVL4 (also known as HUD; Kim et al., 2015), another RBP which has been implicated in the localisation of β-actin and Gap43 transcripts (Yoo et al., 2013).

Similarly to the β-actin zipcode, most other RNA localisation elements described so far were all identified in case studies of individual genes, most of which were known to be localised, had localised proteins products or are known targets of localised RBPs. The non-coding regions of transcripts from

Arc (Kobayashi et al., 2005), Bc1 (Muslimov et al., 2006), Bdnf (An et al., 2008; Oe & Yoneda, 2010), Camk2a (Blichenberg et al., 2001; Huang et al., 2003; Mori et al., 2000), and Map2 (Blichenberg et al., 1999) were all studied using reporter assays, which allowed determination of certain sequence fragments with the capability to drive localisation towards neurites. However, many of the localisation elements identified by these approaches are not clearly defined as a specific sequence motif and only describe a broad region of the transcript 3'UTR, which confers zipcode activity.

Only in few cases it was possible to identify a sequence motif or a *trans*-acting RBP, which binds a specific element: the A2 response element (A2RE) originally identified as a zipcode in Mbp transcripts (Ainger et al., 1997) also provides localisation activity for the Map2a or reporter transcripts in neurons and is dependent on the heterogeneous nuclear ribonucleoprotein (hnRNP) A2 (Shan et al., 2003). Furthermore, it is also present in the Bdnf, Camk2a, Arc, and Bc1 transcripts where it similarly can drive localisation, which can also be mediated by hnRNP A/B (Raju et al., 2011). While the neuritic localisation of Bc1 has also been shown to be dependent on hnRNP A2, the direct binding is based on a structural motif overlapping the A2RE, where the contribution of the primary sequence is not clear, but the interaction is linked to activity correlated intracellular calcium levels (Muslimov et al., 2006; Muslimov et al., 2014). Another defined sequence motif identified as a zipcode in the Camk2a 3'UTR is the cytoplasmic poly-adenylation element (CPE), with its known binding protein CPEB (Huang et al., 2003). While no further reports on its effect on RNA localisation have been made, there is extensive knowledge about the effect of the CPE on translational control (Ivshina et al., 2014), a level of regulation that has also be reported for the A2RE (Kosturko et al., 2006). As such, it is not clear whether the effect of CPE on localisation is specific to the Camk2a transcript or also present in other transcripts, however there are many transcripts potentially regulated by it, as a CLIP study of the *drosophila* ortholog *orb* identified over 3000 targets of CPEB (Stepien et al., 2016).

Another more recently discovered zipcode is the structural RNA G-quadruplex motif that induces neurite localisation in Camk2a and Psd95 transcripts, but is also present in others known to be localised like Bdnf and Shank1 (Subramanian et al., 2011; Y. Zhang et al., 2014). The same studies identifying the zipcode activity of the RNA G-quadruplex have also shown that it is a binding site for FMRP, a neuronal protein with many functions and strong influence on translation or neuronal and synaptic transcripts as well as neurological disorders (Mofatteh, 2021; Thelen & Kye, 2020). Furthermore, several other RBPs like FUS and TDP43, which are also linked to both RNA transport and neurodegenerative diseases, have now been shown to bind the same motif (Imperatore et al., 2020; Ishiguro et al., 2016). In addition to their functions in RNA localisation these RBPs share other functional capabilities like control of translation (Efimova et al., 2017; Z. Li et al., 2001) and interaction with complexes that mediate

miRNA dependent silencing of transcripts (Kawahara & Mieda-Sato, 2012; Muddashetty et al., 2011).

Similarly, further proteins that mainly regulate other aspects of RNA biology, most often translation, have also been implicated to affect RNA localisation. This includes the Staufen proteins, which among other functions are also involved in the regulation of stress granules, that can globally inhibit translation (Thomas et al., 2009) and control localisation of their own transcript both during development of *drosophila* oocytes and in asymmetric neuroblast differentiation (P. Li et al., 1997; Vessey et al., 2012). Additionally, it was shown that Staufen associates with many RNA granules (Furic et al., 2008; Heraud-Farlow et al., 2013), including actively transported granules (Jeong et al., 2007) and also affects synaptic plasticity and long-term potentiation in the brain (Lebeau et al., 2008). Similarly, the Pumillo proteins, which are also required for neuronal development, have as well been implicated in mRNA transport and formation of stress granules (Menon et al., 2004; Zahr et al., 2018).

In summary, the complexity of well studied individual factors that control RNA localisation and often other aspects of RNA biology can not be understated and many of these factors also interact with one another, further complicating the prediction of their exact effects in RNA localisation. This pattern of complex interactions governing localisation of individual transcripts of concurrently controlled regulons likely extends from the known examples into a much larger field of unknown ones (Costa et al., 2021; Turner-Bridger et al., 2020). The interplay of several factors, often with both different target transcripts and potentially using orthogonal binding motifs, also highlights that many RBPs may not strictly bind to a specific sequence motif, but rather employ structural and context clues to achieve combinatorial control of transcript target sets (Dominguez et al., 2018).

2.3 Massively parallel reporter assays

While several zipcodes and RNA localisation factors are quite well described, the number of known zipcodes and associated localisation factors is relatively small compared to the complexity and extent of the neurite localised transcriptome. Therefore, it is generally expected that several unknown zipcode motifs and associated RPBs exist which control localisation of the many transcripts that have not been studied in depth (Gomes et al., 2014; Kar et al., 2018).

One emerging approach to uncover unknown control elements like zipcodes combines both the advantages of next generation sequencing techniques and the principle of reporter assays already used in early studies of RNA localisation elements. A sequencing based approach requires the generation of distinct RNA pools, which is of course ideal in the neuronal system where neurites and soma can be physically separated and also allows the analysis of a very large number of individual reporters at the same time. Therefore, these approaches, usually called massively parallel reporter assays (MPRAs), have

allowed the discovery of control elements for stability and translation control of RNAs (Oikonomou et al., 2014; Rabani et al., 2017; Yartseva et al., 2017; Zhao et al., 2014) as well as nuclear RNA localisation (Lubelsky & Ulitsky, 2018; Shukla et al., 2018).

Several technical aspects of the MPRA approach as well as choices in assay design can have strong impacts on the obtained results, which is especially important to consider with an assay for a new model system like neuronal RNA localisation. Foremost is the selection of the set of genes or sequences to be analysed in the MPRA. Different strategies have been employed to select these sequence pools: the most common approach is informed manual or guided selection of differently sized sets of biologically relevant genes (Lubelsky & Ulitsky, 2018; Rabani et al., 2017; Shukla et al., 2018; Yartseva et al., 2017), but other approaches include the pre-screening of randomised oligos (Wissink et al., 2016), selection of conversed sequences (Oikonomou et al., 2014) or analysis of different mutations of a single sequence to obtain detailed insights (Zhao et al., 2014). Since *cis*-active elements, and especially localisation associated ones like zipcodes, are most often located in the 3'UTR of a transcript (Andreassi & Riccio, 2009; Gomes et al., 2014), it is reasonable to include only this part of the transcript sequences in the assay.

While it would be possible to build a reporter system for the whole 3'UTR of a large set of genes, this is technically quite challenging due to the very wide range of sequence lengths and would also not even provide any information about the specific position of any potential zipcodes. Therefore, the genes selected for analysis in an MPRA are usually subdivided into many equally sized sequence fragments, which not only reduces technical noise in the reporter libraries, but also aids with streamlining analysis. This subdivision can be done in two different ways: either by selecting specific, potentially overlapping, tiles with a defined offset along each sequence (Lubelsky & Ulitsky, 2018), or by using random fragmentation and size selection of a sequence pool (Yartseva et al., 2017). While the latter offers potentially more distinct fragments and therefore a theoretically higher positional resolution for identifying active elements, it also more heavily depends on high quality technical performance and readouts to achieve this. Additionally, pre-defined sets of sequences allow for easier analysis and quality control with RNA sequencing, since one can check for exact sequence matches.

Whichever set of genes and reporter sequences are chosen and generated for an MPRA, they have to be converted into an actual reporter library pool, that can be used in the desired model system. While some model systems like developing embryos allow direct injection of RNA (Yartseva et al., 2017), this is not possible in neurons, where a transfection system that also works well with the non-proliferative cells has to be used. Such a system is usually based on an initial plasmid library pool, in which additional experimental parameters can be fine tuned to the model system, like utilisation of a neuron specific promoter. Furthermore, the source of neurons, or any other cells, is also an important parameter, not

only because reliance on an established and well described system like primary neurons further cultured *in vitro* determines the data quality obtainable from the assay, but also because the initial selection of genes with potential active elements is usually built on knowledge that also derives from the same or at least commonly used model systems. Finally, the reporter RNA in the cellular system has to be separated into distinct functional groups material for RNA sequencing library generation. In the case of zipcodes for neuronal RNA localisation, this can be achieved through the same physical separation procedures that allow identification and quantification of localised transcripts in neurites.

2.4 RNA-Seq: high throughput data and their analysis

Regardless of whether one wants to study transcriptomics, perform an MRPA or study other aspects of RNA biology, the most common method to identify and quantify many different RNA molecules in one experiments is RNA sequencing (RNA-Seq). This approach relies on reverse transcription of RNA into cDNA, which can be identified and quantified using next generation sequencing (NGS) machines. While NGS was originally developed as a tool to identify the primary sequence of genomic DNA fragments, in modern transcriptomics it is mainly used to quantify the number of reads associated with a specific transcript or gene and therefore also allows differential expression analysis (Stark et al., 2019).

A similar analysis approach was already possible based on RNA microarrays before the emergence of sequencing. However, microarrays are limited to the detection and quantification of known and predefined transcript sequences and are also limited by the dynamic range of the probe scanner and have higher noise levels than RNA-seq, where the dynamic range is also only limited by sequencing depth (Wang et al., 2009; Wilhelm et al., 2008). Similarly to microarrays, other high throughput methods, like mass spectrometry protein measurements, can usually also provide only intensity based readouts. Nonetheless, modern shotgun proteomics are similarly to RNA-Seq in their ability to quantify the whole proteome in a single experiment, relying on the identification of peptides fragments to obtain summarised quantification of individual proteins (Tyanova et al., 2016).

All high-throughput methods have technical biases and uncertainties, which need to be accounted for in the analysis to achieve proper quantification of molecules. In RNA-Seq based transcriptomics one major source of such uncertainty are over-amplified cDNA duplicates generated during PCR based library construction. The most common approach to overcome this issue is the introduction of unique molecular identifiers (UMIs), which are short randomised sequences tags introduced early on in the cDNA or library preparation. The UMIs are then part of the sequenced reads and allow the distinction between artificially duplicated cDNA fragments and true biological signal from multiple RNA fragments with the same sequence (Kivioja et al., 2012).

Many different approaches or methods built on RNA sequencing exist, like different transcript

selection procedures for mRNA (oligo-dT selection) or the total transcriptome (depletion of ribosomal RNA). However, regardless of the approach or specific question, the basic steps in analysis of RNA-Seq data are always similar: first read sequences need to be mapped to the genome or transcriptome to subsequently allow summarisation of gene or transcript counts. These counts then have to be normalised and potentially filtered before differential expression analysis can be performed (Conesa et al., 2016). Genome alignment of RNA-Seq reads to specific positions is performed with tools like STAR, that perform a full sequence based alignment of reads (Dobin et al., 2013), but require a secondary step to obtain the counts of reads mapped to each gene body. Alternative tools like salmon have emerged more recently, which use k-mer based matching of reads to known transcript sequences to directly perform quantification (Patro et al., 2017). This approach is computationally much faster than other methods and always includes partially estimated counts of individual overlapping transcript isoforms, but also looses some accuracy for very small or lowly expressed RNAs (Wu et al., 2018) and only allows quantification and no other analysis like identification of novel transcripts or splice sites (Stark et al., 2019).

The normalisation of RNA-Seq gene or transcript abundances, which are generally correlated to expression levels, needs to account for the fact that read counts are derived from individually identified molecules. This is also the biggest difference between RNA-Seq and other high throughput methods like mass spectrometry or microarrays, which produce intensity measurements. The distribution of these counts is intrinsically biased to samples with a higher sequencing depth and, in the case of mRNA or total RNA sequencing, but not necessarily for other protocols, also towards the transcript length, since random fragmentation produces more readable fragments from longer transcripts (Conesa et al., 2016; Stark et al., 2019). Normalisation for these two parameters does, depending on the order of normalisation, generated either RPKM (reads per kilobase per million) or TPM (transcripts per million) values (Wagner et al., 2012). While these values give good indications of relative transcript expression levels within or, in the case of TPM, also between samples, they are not suited for differential expression analysis, since the relative expression levels are not always directly comparable between samples (Bullard et al., 2010; Dillies et al., 2013; Wagner et al., 2012).

2.5 Differential expression analysis and integration of multiple datasets

After counts from RNA-Seq have been generated and quality controlled, depending on the tools utilised, either raw or properly normalized values can used for differential expression (DE) analysis, which allows the calculation of relative expression changes between two experimental conditions expressed as log2 fold change values. Some of the most popular tools for DE analysis are DESeq2 (Love et al., 2014), which was specifically developed for RNA-Seq data, and limma (Ritchie et al., 2015), which was originally developed for microarray data, but can be used on any kind of expression data. DEseq2 operates

on raw counts, since it performs it's own normalisation, which includes calculation of size factors controlling sequencing depth of samples that are modelled based on a negative binomial distribution of the underlying count values. This approach allows for a robust estimation of baseline expression levels, even with a relatively small number of samples and also for genes with low counts. However, it also assumes that genes with similar expression levels also have similar dispersion, in order to allow correction of the dispersion measured for individual genes from few samples by the observed average of similarly expressed genes (Love et al., 2014).

In order to determine and enumerate expression changes between different samples, both DESeq2 and other DE tools utilise a generalised linear model (GLM) that allows fitting of coefficients for user defined experimental variables. These coefficients are derived from a statistical model provided for any set of samples, which can accommodate anything from a simple comparison between two conditions to a design with several co-factors and experimental variables. DESeq2 controls for the strong effect of very low expression on log2 fold change values by correction of initial values with a secondary GLM fit and enables testing for statistical significance with the Wald test (Love et al., 2014). Both this test for significance and the GLM are build with the assumption that the majority of analysed genes does not differ between conditions and changes occur equally likely in either direction, i.e. that log2 fold change values are centred around zero with a normal distribution. Other tools for DE analysis generally share most assumptions made by DESeq2, but differ in the exact algorithms used for individual steps of the analysis. The limma tool for example, also is based on fitting of a GLM and groups similarly expressed genes by variability to increase the statistical accuracy of log2 fold change values, but does not itself contain steps for normalisation, which makes it also applicable to DE analysis of data types other than RNA-Seq (Ritchie et al., 2015). A GLM supported analysis of multiple samples and effects at the same time provides several benefits, but also makes proper model design and normalisation of read counts more difficult. The sharing of information in a larger set of samples and consideration of known covariates can increase the statistical power of the model, even if DE is only calculated for a given subset of conditions (Ritchie et al., 2015). Additionally, multiple comparisons can potentially be done in the same singular modelling step, which improves both the runtime and comparability of results within a given experiment.

As large sets of biological samples can often not be generated and processed at the same time, the emergence of processing correlated batch effects is a common issue in DE analysis, but can be somewhat mitigated by appropriate study design, even though additional correction may still be necessary (Leek et al., 2010). This problem can also be overcome by using batch correction approaches, several of which were developed for genomics or microarray studies before the wide spread usage of RNA-Seq. If the batch variables are known, for example from different processing times of samples, then it is

possible to correct systematic differences in expression values using empirical Bayes models. This approach is similar to GLM based determination of fold differences between conditions, as it also allows borrowing of information from several genes to estimate and consecutively remove a global effect of a certain batch variable on expression (Johnson et al., 2007). Since this approach can only adjust for known batch variables, another method called surrogate variable analysis (SVA) was developed, which allows batch correction after determination of unexplained singular vectors, or principal components, within the residual expression variation not correlated to known parameters of interest (Leek & Storey, 2007). Another approach to determine batch factors is to focus on a set of negative control genes like ubiquitously expressed house keeping genes or, especially for RNA-Seq, exogenous spike-in controls. This approach minimizes the risk that any known or determined batch factors correlate with a biological signal of interest, which could get lost during batch correction, and has been implemented as the remove unwanted variation (RUV) method by Risso et al. (2014). It has also been shown that batch correction is highly beneficial for DE analysis, especially for large collections of data, and can even be applied without explicit knowledge about negative control genes by including all genes as a proxy (RUVs approach) (Peixoto et al., 2015; Risso et al., 2014).

Another possible complication for DE analysis is breaking of assumptions in the used models. While this is very unlikely to occur during analysis of the full transcriptome, it is possible in an analysis with a reduced number of genes or under circumstances where DE is very common. Such conditions are most likely within in a reporter assay context, where several analysed transcripts share a similar sequence and the experimental design may intentionally produce many positive results. Additionally, MPRAs or other reporter libraries also have further deviation from standard transcriptome sequencing, like a more homogeneous read length distribution and often either specifically designed fragments that can be mapped to a known reference set or unknown fragments for which count bins over genomic or 3'UTR regions need to be generated. Therefore analysis of such data often uses only a very simple approach that only includes basic sequencing depth normalisation, generation of log ratios between conditions of interest and potentially statistical test for the significance of differences (Lubelsky & Ulitsky, 2018; Oikonomou et al., 2014; Yartseva et al., 2017), even if more sophisticated analysis including model fitting are possible if one ascertains that all model assumptions are met (Rabani et al., 2017).

2.6 The importance of localised RNA in disease - ALS and the FUS protein

While basic research leading to a better understanding of how RNA localisation in general and within neurons in particular is an important goal in itself, there is also considerable interest in understanding its importance in neurodegenerative diseases. Several RNA binding proteins that have been tightly linked to different diseases, like SMN to spinal muscular atrophy, FMRP to fragile X syndrome as well as FUS and TDP43 to frontotemporal dementia and amyotrophic lateral sclerosis (ALS), have indeed also been associated with local RNA localisation (Thelen & Kye, 2020). Correspondingly, disruption of RNA localisation has also been identified has a hallmark of several neurodegenerative diseases including ALS, but it is so far not clearly understood, whether this is a cause of or a symptom underlying the molecular disease pathways (Mofatteh, 2021).

ALS in particular is an incurable disease that ultimately leads to death about 2-5 years after diagnosis due to the irreversible loss of motor neurons. While the exact cause and progression of the disease are not fully understood, one common theory is that the degradation starts at the synapses between motor neurons and muscles, the neuromuscular junction (NMJ), and slowly progresses from there with a 'dying-back' phenomenon (Dadon-Nachum et al., 2011). The exact mechanisms of how this degradation and loss of motor neurons occurs involves many cellular pathways in a complicated network and as such is still a matter of ongoing research. A good overview of the molecular pathology of ALS has been given by J. P. Taylor et al. (2016). As ALS can occur both with and without a familiar genetic background, a considerable amount of research into mutations that are most often associated with especially the familiar origin of the disease has been conducted. The most important of these proteins include SOD1, a mitochondrial enzyme for which most mutations gain a toxic function, C9ORF, a gene in which CGG-repeat expansion leads to ALS, as well as the RBPs TDP43 and FUS, which both lose nuclear localisation and become part of inclusion bodies in ALS (J. P. Taylor et al., 2016). This aggregation as well as disrupted degradation of proteins, specifically including TDP43, is also one of the most common hallmarks of most forms of sporadic and familiar ALS. Additionally, protein aggregation and miss-folding, which also affects other ALS associated proteins like SOD1 and FUS, as well as generally increased ER stress has also been reported as a common occurrence in ALS. Furthermore, the molecular pathology of ALS also involves disruption of axonal transport, homeostasis, and the local transcriptome and translatome (Suzuki et al., 2020), as well as contributions from non-neuronal cells, especially glia, towards the disease (Haidet-Phillips et al., 2011). A final mechanistic link between the genes, whose mutations are associated with ALS, which has so far not been investigated in depth, is the structural G-quadruplex motif, which can also act as a zipcode in neurons (Maltby et al., 2020; Subramanian et al., 2011). Not only can both FUS and TDP43 bind this motif (Imperatore et al., 2020; Ishiguro et al., 2016), it can also be formed by the hexanucleotide repeats within C9ORF (Haeusler et al., 2014).

In order to study ALS one either needs to rely on mouse models that can replicate the human disease, or use *in vitro* generated human motor neurons as a model system. As importance of disease related processes inside axons are reinforced by the fact that the dying back phenomenon starts there, an *in vitro* model system that allows separation of soma and neurites is a promising approach towards this particular field of study. The generation of motor neurons can either be achieved by differentiation of stem cells after natural activation of signalling pathways through small molecules (Amoroso et al., 2013) or by induced expression of neurogenic transcription factors (Hester et al., 2011; Mazzoni et al., 2013). Both of these approaches can yield motor neurons, however, induced differentiation is often possible in a shorter time frame, achieves higher differentiation efficiency and also yields more reproducible results, even if the resulting cells may not fully recapitulate specific motor neuron subtypes formed *in vivo* (Davis-Dusenbery et al., 2014). In addition to their ease of handling, *in vitro* generated motor neurons have the added benefit that ALS associated mutations can easily be introduced in a wild type cell line, or corrected in a phenotypic cell line.

Of particular interest for basic research into the molecular mechanics of ALS is the FUS protein, which is not only the second most common cause of familiar ALS with around 5% of all such cases (Kwiatkowski et al., 2009), but also has been associated with frontotemporal dementia (FTD), another neurodegenerative disease that has molecular similarities to ALS, but affects the brain rather than motor neurons (J. P. Taylor et al., 2016). Furthermore, FUS can affect both RNA transport and local translation in axons, where the presence of mutant protein also impedes synaptic activity (López-Erauskin et al., 2018). Additionally, most ALS associated mutations of FUS occur in its 15th exon, which codes for its nuclear localisation sequence (NLS) and a considerable portion of its neurotoxic potential has been ascribed to a gain-of-function related to an increased cytoplasmic localisation of FUS (Efimova et al., 2017). These aspects indicate that mutant FUS protein most likely induces ALS by a toxic gain-of-function, which is assumed to be linked to both its propensity towards formation of aggregates (Sun et al., 2011) and its ability to bind cytoplasmic transcripts that are normally not available to it (Hoell et al., 2011). Nevertheless, FUS also serves a plethora of other functions, including DNA repair, transcriptional regulation, splicing, and RNA transport (Efimova et al., 2017), whose impediment through mutations may also contribute to cellular ALS mechanisms and thereby complicate identification of individual pathways contributing to disease etiology.

2.7 Aims of my thesis

The work I have performed for this thesis all revolves around and tries to improve our understanding of the particulars, mechanisms and implications of RNA localisation in neurons.

Towards this end I aim to leverage newly generated and available public RNA-Seq data and integrate them into an encompassing and robust summary of a systems level view of localised transcripts based on both presence and enrichment in neurites. Further, I use this knowledge to design an MPRA to identify so far unknown zipcode elements within the 3'UTR space of commonly neurite localised transcripts. Using, first, the combination a of data driven selection procedure and curation for reliable signals and, second, a mutation analysis approach I want to identify the sequences identity and potential mechanisms of novel candidate zipcode elements.

Additionally, I apply omics analysis of RNA and protein expression in a motor neuron model with the aim of gaining further insights into potential disease mechanisms of ALS. For this I first need to ascertain the usability of an induced motor neuron (iMN) model generated from controlled differentiation of human induced pluripotent stem cells (hiPSC) by generating a detailed characterisation including the local transcriptome and proteome. Using differential expression analysis between different patient derived FUS mutant and specific isogenic control motor neurons I then aim to identify individual or groups of genes that show potentially compartment specific expression changes linkable to the ALS associated mutations.

3 Material and methods

3.1 Comparison of RNA-Seq data from neuronal compartments

All experimental procedures regarding the generation of RNA-Seq data from PCN described in this section 3.1 were performed by Sayaka Dantsuji.

3.1.1 PCN generation and cell culture

Primary neurons were dissected from cortex of E14 mouse embryos and cultured in neurobasal A media supplemented with B27 (1:50), Glutamax (1:100) and penicillin-streptomycin (1:1000). For subsequent separation of neurons, cells were grown on microporous membrane filter insets, which also allowed cells to be grown in co-culture with astrocytes, which were as well obtained from mouse embryos and placed in the wells before culture of neurons began. Medium was changed regularly and after 9 days *in vitro* (DIV9) cells were fixed with 100% methanol prior to compartment separation and RNA extraction.

3.1.2 Compartment separation and generation of RNA-Seq libraries

For separation of neuronal compartments, first the soma growing on top of the filter insets were collected with cold PBS and transferred into Trizol (Thermo Fischer) after removal of PBS. The filters were cleaned using cotton swabs and then also transferred to Trizol (see Ludwik et al. (2019) for a detailed protocol). RNA was extracted from Trizol using chloroform/isopropanol extraction according to manufacturers recommendations. Integrity of collected RNA samples was analysed using Bioanalyser and efficiency of compartment separation was ascertained based on qPCR.
RNA sequencing libraries were generated in triplicates from 120 ng of total RNA using the TruSeq Stranded Total RNA library kit (Illumina, RS-122-2201) according to manufacturers protocol. The final libraries were sequenced using an Illumina NextSeq500 machine with 151 single end read cycles.

3.1.3 RNA-Seq analysis using PiGx

For publicly available RNA-Seq data from compartment separated neurons, I obtained raw data deposited with the NCBI GEO repository from the following studies: Briese et al. (2016), Cajigas et al. (2012), Ciolli Mattioli et al. (2019), Farris et al. (2019), Middleton et al. (2019), Minis et al. (2014), Nijssen et al. (2018), Poulopoulos et al. (2019), Taliaferro et al. (2016), Tóth et al. (2018), Tushev et al. (2018), and Zappulo et al. (2017). For datasets without available fastq data, I used the un-normalised counts provided in the supplementary files of the respective studies (Maciel et al. (2018) and Rotem et al. (2017)). For processing of fastq files, I used the PiGx RNA-Seq pipeline (version 0.0.10) (Wurmus et al., 2018), which included quality and adapter trimming performed with `trim_galore`, genome mapping with `STAR`

(Dobin et al., 2013) and read counting with R (`GenomicAlignemts::summarizeOverlaps`). The necessary genome and transcriptome files were downloaded from ENSEMBL: Mus musculus, GRCm38, version 96; Rattus norvegicus, 6.0, version 98; Homo sapiens, GRCh38, version 97. Genes from human or rat datasets were assigned to orthologous mouse genes based on biomaRt annotation (Durinck et al., 2009) with exclusion of many-to-many orthologs and genes without any orthologs. In the case of multiple human or rat genes orthologous to a single mouse gene, I only assigned the highest expressed gene as an ortholog to mouse. To obtain TPM values, I normalised read counts of all samples to average transcript length based on respective genome annotation and to total reads per sample. Then I averaged TPM values for datasets and compartments and only retained genes with an average TPM > 1 in a given dataset for comparison.

Additionally RNA-Seq libraries from PCN were generated by Sayaka Dantsuji. I processed these data as described above using PiGx, however the data also showed evidence of DNA contamination so I only used genes with htseq-based log2 read counts ratios of exon/intron > 2.5 and sense/antisense strand > 2 for any further analysis of this dataset.

3.1.4 Differential expression analysis

I performed differential expression analysis for samples from neurite and soma compartments individually within each dataset on raw counts of all genes with average TPM > 0 (across all samples) using DESeq2 (Love et al., 2014). Genes with significant (adj. p-value < 0.05) enrichment with log2 fold change (log2fc) at least 0.5 I considered as localised (for either compartment).

In order to perform differential expression analysis on all datasets at the same time, I removed unwanted variation between the datasets using the RUVs approach (Risso et al., 2014). Using kBET, a k-nearest neighbour test method for batch effects (Büttner et al., 2019), I determined k=14 to be best number of factors of unwanted variation to remove, as it showed the smallest difference between expected and observed batch effect. I then used DESeq2 with a model considering all 14 numerical factors as determined by RUVs as well as the dataset-identity as covariates to determine the differential expression between neurites and soma for all datasets.

In addition to differential expression I also extracted variance stabilised counts from the DESeq2 model and used those for principle component analysis.

3.2 Neuronal zipcode identfication protocol (Nzip)

All experimental procedures described in this section 3.2 were performed by colleagues from the Chekulaeva lab and MDC: Sayaka Dantsuji generated all Nzip library pools, obtained PCN and performed sequencing of Nzip libraries (3.2.2-3) as well as small RNA libraries (3.2.6); Marina Chekulaeva performed RNA affinity pulldown with help from Nadja Zerna (3.2.7); Samantha Mendonsa generated PCN protein lysate for mass spectrometry (3.2.7); and Marieluise Kirchner (from the lab of Philipp Mertins) performed mass spectrometry measurements and raw data processing (3.2.7-8). Additional analysis of miRNAs counts in small RNA-Seq data and of Nzip mutation libraries was performed by collaborators from the Weizman institute (Maya Ron, Igor Ulitsky).

3.2.1 Design of Nzip reporter library

For the selection of genes with a potentially unknown zipcode in the 3'UTR , I used neurite and soma compartment separated RNA-Seq data both from the PCN data generated in our group and from other published datasets derived from primary neurons (Briese et al., 2016; Middleton et al., 2019; Minis et al., 2014; Rotem et al., 2017; Taliaferro et al., 2016; Tushev et al., 2018). I considered genes localised to neurites in a given dataset if they had log2fc >0 with an adj. p-value <0.05. For the selection across datasets I also calculated the average and median of all significant log2fc values across the datasets and classified genes as unanimously neurite localised if no significant enrichment in soma (log2fc <0, adj. p-value <0.05) existed. Then I selected genes with both consistent and sufficiently strong neurite localisation by taking those that showed: (1) neurite localisation in at least 4 datasets with a median log2fc >1 and either unanimous neurite enrichment or an average log2fc >1; or (2) neurite localisation in at least 5 datasets with median log2fc >0 and additionally either mean or median log2fc >1 or unanimous neurite localisation. To ensure compatibility of this resulting Nzip selection with our PCN model system I then removed all genes that did not have a significant neurite enrichment in our PCN data (log2fc >0, adj. p-value <0.05). Further manual curation of this gene list was performed to: (1) ensure inclusion of genes with described zipcodes or neurite localisation elements: Actb (Kim et al., 2015; Kislauskis et al., 1994), Arc (Kobayashi et al., 2005), Bc1 (Muslimov et al., 2006), Bdnf (An et al., 2008; Oe & Yoneda, 2010), Camk2a (Blichenberg et al., 2001; Huang et al., 2003; Mori et al., 2000; Subramanian et al., 2011), Cdc42 (Ciolli Mattioli et al., 2019), and Map2 (Blichenberg et al., 1999)); (2) add genes with strong or consistent neurite localisation in other non-primary datasets, either published (Taliaferro et al., 2016; von Kügelgen & Chekulaeva, 2020; Zappulo et al., 2017) or in-house: Rab13, Net1, Hmgn5, 2410006H16Rik, Pfdn5, Tagln2, Pfdn1, Cryab, Rpl14, Eef1b2, and Eef1a1; (3) remove genes with functions in nucleus or translation (Pola1 - DNA polymerase, Ezh2 - Polycomb complex, Smc4 - Chromosome structure, Cenpb - Centromere structure, Ncl - nucleolus, Pink1 - mitochondrial damage response; and

(4) restrict the number of ribosomal proteins in list of enriched transcripts to a smaller subset, also with sufficiently long 3'UTR to use for tiling (Rplp2, Rpl12, Rpl39, Rpl37, Rps28, Rpsa, Rps24, Rps23, Rps18).

I used the biomaRt ENSEMBL interface (Durinck et al., 2009) to download the 3'UTR sequences of all transcript isoforms annotated to these genes. For all different isoforms of the same gene I merged the overlapping sequences to obtain continuous unique sequences of each. For genes with multiple sequences obtained in this manner I relied on transcript annotation as well as genome browser tracks from the PCN data to select a single expressed and functional 3'UTR sequence. In the case of Cflar and Cdc42 I included two different 3'UTR sequences in the selection each derived from alternative last exons. For the genes Hdac5 and Arhgap11a I obtained different but partially overlapping or heavily repetitive 3'UTR sequences, which I manually merged into a single one containing all unique sequence elements. This resulted in a final set of 99 3'UTR sequences derived from 97 different genes (Table S3).
For each of these sequences I generated 75 nt or 100 nt tiles for sequences with total length above 500 nt with 15 or 25 nt offsets so that each sequence was covered from start (5') to end (3'). For the last remaining part of each sequence I either extended the last tile to a maximum size of 80 nt or 110 nt respectively or added an additional shortened tile. Additionally I added 5 scrambled (randomised) tiles for each of the first tile of Camk2a, Actb and Bc1. The final set encompassed 4813 tiles and was ordered including 3' and 5' cloning adapters.

3.2.2 Generation of Nzip reporter library pools

The synthesized oligo tile pool was PCR amplified based on the cloning adapters and then cloned into the 3'UTR of a synapsin promoter driven GFP cassette in a plamsid allowing for lentiviral packaging (Addgene #20945).
Lentiviral particles were generated in HEK 293T cells and concentrated using Lenti-X concentrator (631232 Takara Bio) before application to cortical neurons at around DIV5.

3.2.3 Generation of Nzip sequencing libraries

RNA material from PCN transfected with the Nzip reporter library was collected as described above (see 3.1.2) in two initial experiments at DIV14 or in one (confirmation) experiment at DIV9. For the two initial experiments neurons were inhibited with 50 µM AP5 and 10 µM NBQX for 16h before harvesting of RNA material. In one of these experiments the neurons were depolarized after the silencing treatment with 55 mM KCl for 1 h. Additionally, the material obtained from soma compartment of these experiments was further separated using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific) to obtain RNA from these compartments.

For preparation of Nzip sequencing libraries 600 ng total RNA treated with DNAse and a custom protocol utilising Superscript III (Thermo Fischer) was used with specific first and second strand primers designed to only bind and amplify the common adapter sequence of all tiles and add UMI sequences with staggered lenghts to the 3' start of each sequence fragment (Table S1). Final amplification of the library was performed using NEBNext High-Fidelity 2X PCR Master Mix and added the binding sequences for Illumina sequencing and barcoding to the libraries. The final libraries were sequenced after purification and dilution using an Illumina NextSeq 500 sequencer with either paired-end 76 nt reads (initial experiments) or single-end 151 nt reads (confirmation experiment). All experiments were performed in triplicates.

3.2.4 Nzip reporter library data analysis

To obtain read and UMI counts from the Nzip libraries I build a data processing workflow utilising umi_tools (Smith et al., 2017) for identification and error correction of UMIs and salmon (Patro et al., 2017) for mapping reads to the full sequence set of all tiles. First I extracted the UMIs encoded in the second strand library generation primers (Table S1) using umi_tools with the following regular expression pattern: `"(?P<umi_1>.9)(?P<umi_2>(CA|AGC|TCGC|GGCTC|CCAGCACCA)?)(?P<discard_1>GCCATAA)s<=1"`. After using the PiGx pipeline to obtain quality trimmed reads and building a salmon index of all tile sequences including the cloning adapters, I mapped reads to the tiles using salmon with the options `--validateMappings -z --recoverOrphans --skipQuant` (salmon version 1.1). I further processed the obtained bam files using bash scripts relying on `samtools` and awk to pad the UMI sequences for each read, so that all UMIs had the same length needed for umi_tools processing, and to assign a new attribute flag to the primary mapped location of each read. Next, I used `umi_tools group` to add error corrected UMIs and their group identifiers to the bam file entries, before I finally used an `awk|sort|uniq -c` based bash pipe to count the occurrence of all unique read mapping states separated by error corrected UMI groups, relative tile position and read length. Finally, I summarised count tables for all samples using custom R scripts to obtain the complete UMI counts for each tile. Reads not mapping to the tile start were almost non existent and I did not remove reads not spanning a full tile (or without a mapped read mate) to retain comparability of my pipeline with that from our collaborators.

After initial tests to rule out negative influence of different tile mapping states, as well as assertion that modelling of size factors worked similarly as on transcriptomic data, I proceeded with differential expression analysis using DEseq2 (Love et al., 2014). For this I calculated enrichment of tiles between neurite compartment and either soma or cytoplasm, from all three replicate samples. Only in the case

of the initial experiment with depolarized PCN I removed one sample due to bad comparability with replicates.

From each experiment I defined peaks of overlapping neurite enriched tiles by selecting all groups of consecutive tiles that showed a neurite enrichment with log2 fold change >0.5 and had at least one tile with significant adj. p-value ($p < 0.05$) and one with log2 fold change >0.75.

3.2.5 Design of Nzip mutation library

Using preliminary analysis of the initial Nzip experiments provided by collaborators (Igor Ulitsky, Maya Ron) I determined neurite enriched tile groups with candidate zipcodes. For this I relied on the neurite/cytoplasm log2 ratios as well as nucleus/cytoplasm ratios or on p-values I obtained from running DEseq on the raw counts of neurite and cytoplasm samples. I considered a group of tiles as a peak if (1) at least 2 consecutive tiles had a median neurite/cytoplasm ratio >0.5 and nucleus/cytoplasm ratio between -3 and 0.5, while at least one tile in a given group had a neurite/cytoplasm ratio >0.75 or if (2) at least 3 consecutive tiles had either a mean neurite/cytoplasm ratio >0.5 or a mean neurite/cytoplasm ratio >0 with a significant adj. p-value (<0.05), while at least two tiles in a given group had a significant adj. p-value and a neurite/cytoplasm ratio >0.75. Based on these peaks from either or preferentially both of the initial Nzip experiments, the following 16 tiles were chosen to be further studied by mutational analysis: Bdnf tile 56, Cald1 tile 58, Camk2n1 tile 12, Cox5b tiles 6&7, Golim4 tile 56, Kif1c tile 80, Mcf2l tile 7, Msn tile 48, Ndufa2 tiles 11&12, Rassf3 tile 91, Rps23 tiles 11&12, and Utrn tile 61. (see also Table S4).

For each of these tiles I generated: (1) every possible single point mutation, (2) transversion mutations in non-overlapping 2, 5 and 10 nt windows each, so that every position was covered by one window of each size, and (3) three scrambled (randomised) control tiles. All of these tiles together with non-mutated WT tiles were combined into the second Nzip mutation library and also ordered with added 3' and 5' cloning adapters.

The mutation reporter library and viral vectors were generated in the same manner as before and sequencing libraries were obtained following the same protocol as the confirmation experiment. Processing and generation of tile counts for these libraries was performed by collaborators.

3.2.6 miRNAs quantification from small RNA-Seq

Small RNA-Seq sequencing was performed by extracting short RNAs from 500 ng of total RNA from PCN compartments separated as described before and then using 100 ng of the shorter RNAs as input for TruSeq Stranded Total RNA library kit (Illumina, RS-122-2201) according to manufacturers instructions. Triplicate libraries were sequenced on Illumina NextSeq 500 machine with 151 nt single end reads.

Identification and counting of miRNA sequences and sequence families was performed by collaborators. Due to no known normalisation markers with similar relative expression in the small RNA-Seq reads, no differential expression analysis between miRNAs and no relative expression strength of mRNA and miRNAs could be performed.

3.2.7 RNA affinity pulldown and mass spectrometry

For affinity pulldown of proteins binding the $(AU)_8$ element, DNA constructs with carrying $(AU)_8$-boxB or mutated-$(AU)_8$-boxB were generated from synthetic oligos corresponding to Rassf3 tile 91 or a mutated version, where $(AT)_8$ was replaced with GTACATACATGTACAT, by annealing and cloning into a vector containing boxB sites. RNA probes for the pulldown essay were generated using T3 Megascript *in vitro* transcription kit (Thermo AM1338) according to the manufacturer's recommendations. RNA affinity pulldown was performed following a modified gRNA chromatography protocol (Chekulaeva et al., 2006; Czaplinski et al., 2005): first GST-lambdaN fusion peptide was immobilized on Glutathione-Sepharose 4B (Amersham, 17075601); then beads were incubated with first 25 pmol of either $(AU)_8$-boxB or mutated-$(AU)_8$-boxB RNA and second with 3 mg of protein lysate prepared from P0 mouse brain, with in-between washing steps; and after final washing proteins were eluted from the beads with 0.15 µg RNAse A and recovered by centrifugation after precipitation from the eluate.

To obtain total protein lysate from neurite and soma compartment of PCN the cells were grown and separated as before, but instead of using Trizol for lysis and sample collection buffer (8 M UREA, 0.1 M Tris-HCl pH7.5) was used instead.

Both eluate from RNA affinity pulldown and total protein lysate from PCN compartments were further processed using in solution digest with trypsin and desalted peptides were then analysed with liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Q Exactive HF-X mass spectrometer coupled to an Easy nLC 1200 system (Thermo Scientific). All samples were measured in triplicates.

3.2.8 Mass spectrometry data analysis

Raw mass spectrometry data was processed using MaxQuant software (1.6.3.4) (Tyanova et al., 2016) with false discovery rate (FDR) for peptide identification set at 1%. Identified proteins were filtered to exclude reverse database hits, potential contaminants, and proteins only identified by site.
Analysis of protein enrichment in the eluates of RNA affinity pulldown from intact and mutated $(AU)_8$ element was performed by Marieluise Kirchner: first LFQ intensity values were used to remove proteins

not present in at least one eluate group with a minimum value of 3, before performing imputation for missing values using a random noise distribution to simulate the detection limit of the mass spectrometer. Proteins with significantly different presence between the two eluates were determined using Student's Two sample t-test with equal variance and permutation based FDR, which was corrected fo multiple testing by the Benjamini-Hochberg (BH) approach.

For analysis of mass spectrometry data from neurite and soma compartments of PCN, I used the iBAQ values from the preprocessed MaxQuant output and the DEP R package (X. Zhang et al., 2018). First, I removed all proteins, which were detected in fewer than half of all samples irregardless of compartments, and the performed imputation using random draws from a Gaussian distribution centred on a minimal expression values (DEP MinProb algorithm). For normalisation I used the housekeeping gene GAPDH as a reference and for each compartment divided all iBAQ values by the median GAPDH expression. Finally, I performed differential expression analysis between compartments based on a generalised linear model by using the DEP implementation of limma and added p-value correction with the BH approach.

3.3 Characterisation of iMN and analysis of iMN with FUS mutantions

Experiments for generation of iMN derived data (3.3.1-3) were performed by Katarzyna Ludwik (wild type iMN) and Samantha Mendonsa (FUS mutant iMN) with help of Tommaso Mari for mass spectrometry processing and anaylsis (3.3.5).

3.3.1 Generation of hiPSC NIL lines

Human iPSC lines were derived from patient fibroblasts by the MDC stem cell score (Dr. Diecke) or Applied StemCell, Inc. (USA, California). Fibroblasts originated from patients with either no known ALS background (wild type (WT)); a FUS P525L mutation, but no diagnosed ALS; or a FUS R244RR mutation and ALS disease. For each of the hiPSC lines carrying FUS mutations an isogenic control line with a corrected FUS gene was generated using the CRISPR/Cas9 system by Axol Bioscience Ltd (UK). An expression cassette for doxycycline induction of NIL (NGN2, ISL1, LHX3) factors, designed by Fernandopulle et al. (2018) (Addgene plasmid #105841), was stably inserted into the CLYBL locus using TALEN mediated insertion (TALEN's from Addgene plasmids #62197 and #62196).

3.3.2 Cell culture and differentiation of hiPSC

For general cell culture hiPSC lines were maintained in E8 media. Before the start of differentiation of hiPSC into iMN, at day 0, the cells were passaged with Accutase and plated on Geltrex coated dishes

in E8 media with Rock inhibitor. At the start of differentiation, day 1, cells were changed to induction media (IM), which includes doxycycline. At day 3, cells were replated into IM supplemented with FuDR onto Geltrex coated 6-well dishes for total cell material or filter membrane inserts for separation of compartments. From day 4 on differentiating cells were maintained in motor neuron media (MM) with supplements. Following that media was partially exchanged every 2-3 days, after day 9 MM without B27 and N2 supplements was used.

E8 media:

- DMEM Nutirent Mix F12 (500 ml, Thermo Fischer)
- Sodium Bicarbonate Sol (3.6 ml, Thermo Fischer)
- Stem cell supplements (provided by MDC stem cell core facility)

Induction media (IM):

- DMEM/F12 w/ HEPES (Thermo Fischer)
- N2 supplement (1:100, Thermo Fischer)
- Non-Essential Amino Acids (1:100, Sigma)
- Glutamax (1:100, Thermo Fischer)
- ROCKi (10 µM) (1:1000, Selleck Chem)
- Doxycycline (2 µg/ml) 1:1500
- Compound E (0.1 µM) 1:10000

Motor neuron media (MM):

- Neurobasal medium (Thermo Fischer)
- B27 supplement (1:50, Thermo Fischer)
- N2 supplement (1:100, Thermo Fischer)
- Non-Essential Amino Acids (1:100, Sigma)
- Glutamax (1:100, Thermo Fischer)
- Laminin (1 µg/ml) (1:1000, Sigma Aldrich)
- 10 ng/ml each BDNF, CNTF, GDNF

3.3.3 Generation of iMN sequencing libraries and proteomics samples

Throughout the differentiation of WT iMN samples from total cells were collected on days 1, 2, 4, 7, 14, and 21. RNA samples were collected in Trizol and processed according to manufacturers recommendations and protein samples were collected in 8 M UREA, 0.1 M Tris-HCl (pH7.5) buffer. For total cell samples of FUS mutant iMN and compartment separation of all lines samples were collected at day 21 of the differentiation. Separation of neurite and soma compartments was performed in the same manner as described above for PCN, only that two membrane filters were combined for a single neurite sample.
For generation of RNA sequencing libraries 100 ng of total RNA from total cells, neurite or soma compartment with added ERCC RNA spike-in mix (Ambion) were used as input for Truseq stranded mRNA library prep kit (Illumina 20020594). All libraries were prepared in triplicate and sequenced on an Illumina NextSeq 500 machine with single-end 151 single end read cycles.
For mass spectrometry triplicate samples were processed for in-solution protein digestion (Mertins et al., 2018). Then digests were acidified and centrifuged to remove the precipitated urea, before the resulting peptides were de-salted via stop-and-go extraction (Rappsilber et al., 2007). Finally, samples were loaded on acetonitrile and formic acid activated C18 material (3M Empore) and washed twice

with trifluoroacetic acid and formic acid before elution. For mass spectrometry approximately 1 µg of peptides per sample was online-separated on an EASY-nLC 1200 (Thermo Fisher Scientific) and acquired on a Q-Exactive HFx (Thermo Fisher Scientific).

3.3.4 RNA-Seq analysis for iMN

I processed the RNA-Seq data from all iMN samples with PiGx as described above (3.1.3), except that I included the codon optimised LHX3 sequence of the NIL cassette in the transcriptome annotation for analysis of total cell samples from the differentiation time course. Also, I used the transcriptome quantification output of salmon (Patro et al., 2017) for all downstream analysis.
For data from the WT differentiation I removed one of three replicates each from the days 2, 4 and 21 due to low quality and similarity to the remaining replicates. Furthermore, I only retained genes detected with a TPM value of at least 1 in 2 or more replicates for analysis and visualisation. I calculated the compartment enrichment between neurites and soma using the DESeq2 (Love et al., 2014) implementation within the PiGx environment and again filtered genes with the same criteria.
For analysis of the iMN with mutated or corrected FUS genes I used tximport (Soneson et al., 2015) and DEseq2 (Love et al., 2014) to perform differential expression analysis between individual compartments of FUS mutant lines and their respective isogenic controls for all comparisons in parallel to benefit from normalisation across a multitude of samples. This analysis also included mutant and control samples from a third patient, which was excluded from the final results because chromosomal aberrations were detected in one of these hiPSC lines.

3.3.5 Mass spectrometry analysis for iMN

All raw mass spectrometry data was processed with MaxQuant (Tyanova et al., 2016) version 1.6.3.4 using the MaxLFQ quantification method (Cox et al., 2014). Proteins were identified from the human uniprot databases (Jan 2020) with a 5% FDR cutoff and hits from the reverse database, only identified by modified site or with less than two peptides were removed. The protein data was the filtered so that only proteins detected either in all replicates of one time point for the WT differentiation data or in at least two out of three replicates for other experiments were retained. Missing values were the imputed by randomly selecting replacement values from a normal distribution with 30% the standard deviation of known values and shifted downwards by 1.8 standard deviation units (Hein et al., 2015). Imputation and normalisation of LFQ values was performed across timepoints for WT differentiation data and within compartments (neurite, soma and total) for all other experiments.
For analysis of neurite/soma protein enrichment in the WT line a t-test was used to evaluate whether differences between (log scale) LFQ values were significant. For differential expression between FUS

mutant lines and isogenic controls I used the limma R package (Ritchie et al., 2015) to fit a linear model and obtain log2 fold change and p-values for each pair of patient line with control and compartment. Finally, I mapped uniprot identifiers to Ensembl gene ids and resolved many-to-one mapping conflicts by only taking values from the protein detected in the most samples using highest overall detection as a tie breaker.

3.4 Celltype deconvolution for iMN

Raw fastq files from sequencing libraries of individual neurons with accompanying action potential (AP) type classification based on electrophysiology measurements were provided by Cedric Bardy et al. (2016). I processed these files using PiGx and imported raw counts from salmon using tximport. Then I used the Seurat R package (Stuart et al., 2019) to process (CreateSeuratObject, parameters: min.cells = 3) and normalise counts. For this I only used cells with AP type 1-5 and grouped the types 1-3 together. Finally I determined differentially expressed marker genes between AP type groups (1-2-3, 4 and 5) using the FindAllMarkers function (parameters: logfc.threshold = 0.25 and min.pct = 0.25).
I used all genes identified this way as signature markers for celltype deconvolution of total cell time points day 4 to day 21 with CIBERSORTx (Non-default signature matrix settings: 25-300 barcode genes; single cell min. expression 1, replicates 0 and sampling 0; 500 permutations used for statistical analysis in cell fraction imputation) (Newman et al., 2019).

3.4.1 Functional enrichment analysis

I performed gene ontology (GO) term enrichment analysis of functional terms among genes with significant enrichment in specific compartments or differential expression between FUS mutant iMN lines and controls using the gprofiler2 R package (Kolberg et al., 2020) with default settings except for setting a custom background of all detected genes.
For analysis of the WT line iMN I used proteins and transcripts enriched in either neurite or soma compartment (adj. p-value < 0.05 and log2fc > 1 or < -1) for GO term analysis (using the gene ontology release from 2020-06). I filtered the terms enriched in the groups of localised genes for overlap between the analysis of both localised proteins and transcripts, a maximum of 1000 annotated genes and at least 25 genes overlapping with the localised groups before applying an additional filter based on graph of GO term relationships to focus remove enriched but functionally related terms (see below).

In case of the FUS mutant lines I performed GO analysis for all proteins differentially expressed in any compartment of either FUS mutation separately for up and down regulated proteins (using the gene ontology release from 2022-01). Then I filtered for enriched GO terms with at least 2 differentially expressed proteins and further with the graph based approach (see below), before taking the 5 most

significant terms from each GO domain, direction (up/down), compartment and FUS mutation line.

For filtering based on the graph structure of the GO terms I relied on the parent and daughter relationships of only the enriched terms from a given analysis (including direction, compartment and mutation line): within one analysis from all enriched terms I selected only those terms that had either no direct daughter terms enriched or no direct parental terms enriched. This approach removes terms which are neither start nor end nodes in a connected subtree of functionally enriched terms.

3.4.2 Analysis of genes disrupted by FUS mutations

Based on the differential expression analysis between FUS mutant iMN lines and their isogenic controls for both RNA and protein data (see above), I assigned a differential expression score for each gene. This score represents the sum of instances in which there is significant (adj. p-value < 0.05) differential expression with a log2fc > 1.5 counted as one (1) score point and log2fc > 1 (but ≤ 1.5) counted one half (½) score points. Points from differential expression in all compartments, data modalities (RNA and protein) and FUS mutations (P525L and R244RR) were combined together.

To compare the log2 fold change values of transcripts with different features, I first assigned different sequence based and functional annotation to genes: for direct binding by either wild type or mutant (cytoplasmic localised R521G or R521H) FUS I relied on CLIP hits as determined by Hoell et al. (2011). The presence of a G-quadruplex motif I determined by searching for the presence of sequence motif $GGGN_{0-6}GGGN_{0-6}GGGN_{0-6}GGG$ (Subramanian et al., 2011) in the most expressed transcript (based on salmon quantification) of each gene. I obtained information about the presence of a signal peptide in a given gene from the signalp database (Armenteros et al., 2019) accessible by biomaRt (Durinck et al., 2009). For annotation of final localisation of gene products I relied on GO annotation (GO 2022/01 release mapped to ensembl gene ids by uniprot identifiers): plasma membrane (PM) components (combined from the GO terms: integral component of plasma membrane, GO:0005887; intrinsic component of plasma membrane, GO:0031226; anchored component of plasma membrane, GO:0046658; extrinsic component of plasma membrane, GO:0019897; and spanning component of plasma membrane, GO:0044214), extracellular (GO:0005576), and extracellular matrix (ECM, GO:0031012). To test for an effect of these gene features on differential expression between FUS mutant iMN and controls within one compartment, I used t-tests to compare the log2fc values of genes with a certain feature to those without the feature. For an overlap of the G-quadruplex motif with GO annotation features or for the ECM sub-group of extracellular proteins I compared against all genes that had both features (all ECM annotated genes were also annotated as extracellular) and compared them against the broader group of GO annotated genes. Finally, I adjusted all calculated p-values using the BH approach.

4 Results

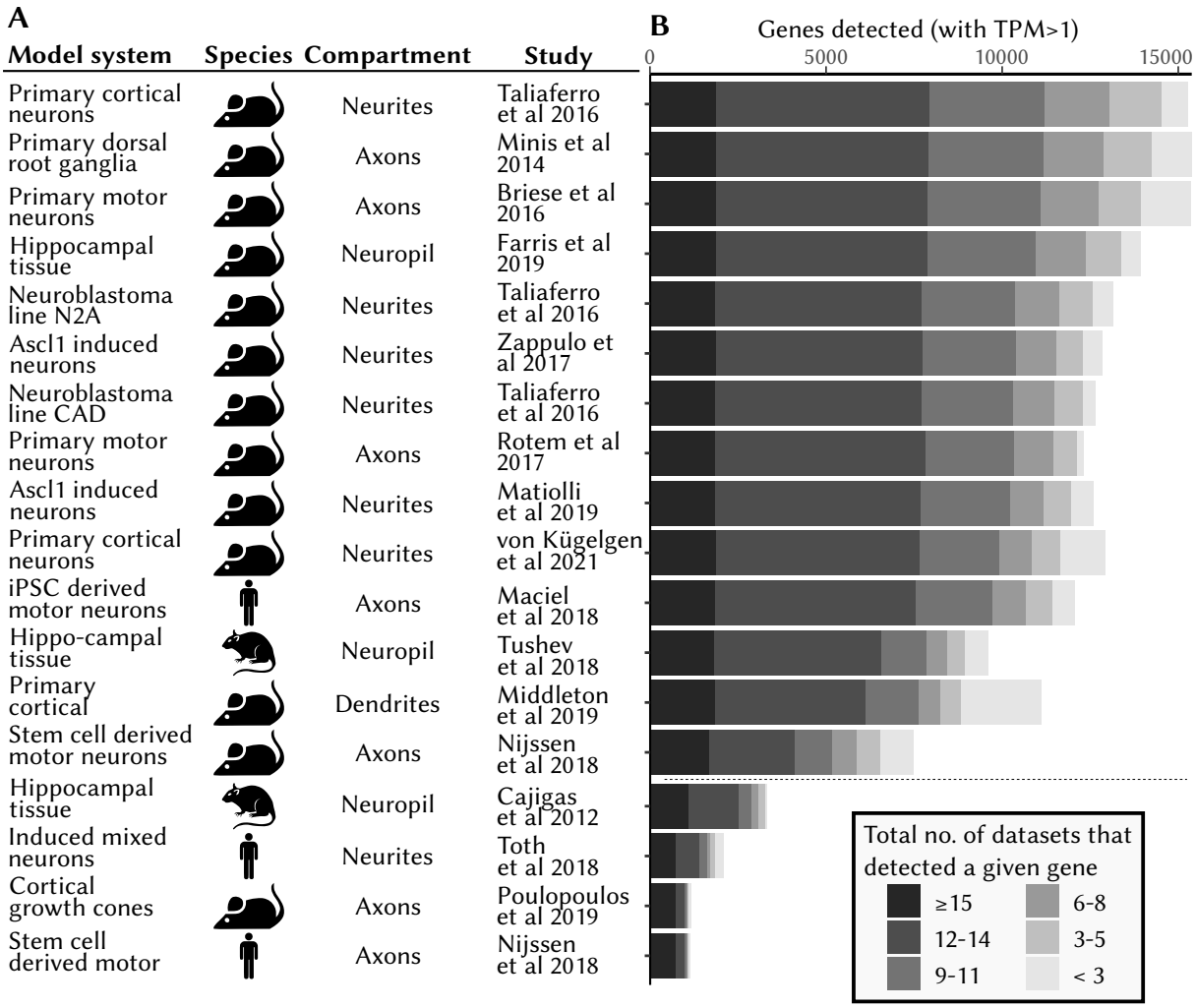
The work I am presenting in my thesis is grouped into three parts: first, the comparison of different RNA-Seq datasets to describe the composition of a core transcriptome present in neurites and to determine transcripts commonly localised to neurites across different model systems. Second, the analysis of the Nzip MPRA to uncover novel zipcode element candidates and identify their sequence identity and potential modes of action. And finally, the characterisation of human motor neurons induced from expression of transcription factors NGN2, ISL1 and LHX3 as well as the analysis of transcriptome and proteome of such iMN carrying ALS associated FUS mutations to identify genes involved in the molecular disease etiology.

4.1 Commonly neurite localized mRNAs

4.1.1 The core neurite transcriptome

The study of subcellular localization of mRNA in neurons is a wide research field with many different contributors spanning both diverging motivations and approaches. While several previous works have stated similarities among studies, no encompassing overview or comparison of the available data had been performed until I addressed this issue in one of my previous works and provided a first comprehensive resource including most published datasets available at that time (von Kügelgen & Chekulaeva, 2020). Observable trends, which individually have also been mentioned by previous studies, but generally apply across all datasets include the common neurite localisation of mRNAs encoding ribosomal proteins, cytoskeleton associated proteins and proteins with functions in mitochondria (Turner-Bridger et al., 2020). Here I am presenting an analysis building on this work, which excludes microarray data, but does include additional RNA-Seq datasets, especially one from mouse primary cortical neurons (PCN), which was generated by colleagues and analysed by me (von Kügelgen et al., 2021). This strand specific RNA-Seq dataset had several genes with comparatively few but strand unspecific reads, indicating weak background contamination with genomic DNA. Therefore, I filtered genes based on the ratios of exon/intron and sense/antisense reads to remove those not distinguishable from contaminating noise before further analysis (Figure S1).

Figure 1 : Core neurite transcriptome of published RNA-Seq datasets.
(A) Table listing the datasets and studies included in analysis of a core neurite transcriptome. Different neuronal models, compartments and species represented by each dataset are also listed. (B) Barplot showing the number of genes detected (with TPM >1) in neurites in each dataset. Detected genes are grouped by the total number of datasets in which they were detected as indicated by the hue of the bars. A dashed line separates datasets with high coverage from those that were excluded from comparative analysis. A similar version of this analysis has been published in von Kügelgen and Chekulaeva, 2020. (C,D) Principle component analysis of normalised and batch corrected (RUV-seq) expression values in neurite (C) and soma (D) compartments from high coverage datasets. The neuronal cell type used in each dataset is indicated by color and source of neuronal cells by shapes of the dots representing each dataset. The PCN data from von Kügelgen et al., 2021 was obtained by Sayaka Dantsuji and all other data is publicly available.


A
Model system
Species
Compartment
Study
Primary cortical neurons
Neurites
Taliaferro et al 2016
Primary dorsal root ganglia
Axons
Minis et al 2014
Primary motor neurons
Axons
Briese et al 2016
Hippocampal tissue
Neuropil
Farris et al 2019
Neuroblastoma line N2A
Neurites
Taliaferro et al 2016
Ascl1 induced neurons
Neurites
Zappulo et al 2017
Neuroblastoma line CAD
Neurites
Taliaferro et al 2016
Primary motor neurons
Axons
Rotem et al 2017
Ascl1 induced neurons
Neurites
Matiolli et al 2019
Primary cortical neurons
Neurites
von Kügelgen et al 2021
iPSC derived motor neurons
Axons
Maciel et al 2018
Hippo-campal tissue
Neuropil
Tushev et al 2018
Primary cortical
Dendrites
Middleton et al 2019
Stem cell derived motor neurons
Axons
Nijssen et al 2018
Hippocampal tissue
Neuropil
Cajigas et al 2012
Induced mixed neurons
Neurites
Toth et al 2018
Cortical growth cones
Axons
Poulopoulos et al 2019
Stem cell derived motor
Axons
Nijssen et al 2018
B
Genes detected (with TPM>1)
0
5000
10000
15000
Total no. of datasets that detected a given gene
≥15
12-14
9-11
6-8
3-5
< 3
C
Neuron source
in vitro generated
Primary
D
Cell type
cortical neuron
hippocampal
DRG
motor neuron
neuroblastoma
Neurite expr. - PC2 (17.2%)
Neurite expr. - PC1 (29.6%)
Tushev
Middleton
von Kügelgen
Farris
Rotem
Maciel
Taliaferro (PCN)
Briese
Nijssen
Matiolli
Minis
Taliaferro (CAD)
Zappulo
Taliaferro (N2A)
Soma expr. - PC2 (17.7%)
Soma expr. - PC1 (20.9%)
Briese
Maciel
Farris
Rotem
von Kügelgen
Minis
Taliaferro (CAD)
Taliaferro (N2A)
Zappulo
Middleton
Taliaferro (PCN)
Matiolli
Tushev
Nijssen

The collection of RNA-Seq datasets from separated neuronal compartments, that I am comparing to determine a core transcriptome generally present in the neurite compartment, encompasses 18 dataset in total and includes: 6 from primary murine neurons (Briese et al., 2016; Middleton et al., 2019; Minis et al., 2014; Rotem et al., 2017; Taliaferro et al., 2016; von Kügelgen et al., 2021), 3 from hippocampal tissue slices (rat or mouse) (Cajigas et al., 2012; Farris et al., 2019; Tushev et al., 2018), 1 from sorted growth cones from mouse brain (Poulopoulos et al., 2019), 6 from *in vitro* murine or human stem cell derived neurons (Ciolli Mattioli et al., 2019; Maciel et al., 2018; Nijssen et al., 2018; Tóth et al., 2018; Zappulo et al., 2017) and 2 from murine neuroblastoma lines (Taliaferro et al., 2016) (Figure 1A). The number of transcripts detected in neuronal outgrowths (classified as either neurites, axons or neuropil by the respective studies) ranges from around 1,000 in only a few datasets with comparatively low detection limits to well over 10,000 or even 15,000 in most datasets. Furthermore, a large number of transcripts (approximately 7500) is detected in at least three quarters of the datasets (12/18). This set of transcripts outlines the size and complexity of the core neurite transcriptome generally present across most datasets and also identifies the subset of RNA-Seq datasets with high enough coverage capture the majority of this core neurite transcriptome to allow comparative and quantitative analysis (Figure 1A, above dotted line).

As in my previous work (von Kügelgen & Chekulaeva, 2020), I compared the expression signatures of high coverage datasets using principle component analysis (PCA). However, the comparison presented here utilises an integrative analysis with a combined differential expression model as well as a batch correction approach to reduce the effects of noise within individual datasets. For batch correction I identified and removed unwanted variation not associated with compartments using the RUVs approach (Risso et al., 2014), so that differential expression analysis with DESeq2 can directly apply corrections to values from all datasets. While no clear clustering of datasets by the neuronal cell types can be observed in the PCA, there is preferential separation between datasets derived from primary and those from *in vitro* neurons both for expression in neurites (respective PC2) and in soma (respective PC1). Notably, the differences within primary datasets span at least double the range of the ones from induced neurons on the corresponding separating PCs, even as both groups are separated from one another. These observations generally replicate the results I obtained before, however it is striking that the PCA based on the batch corrected model shows no strong separation of any individual dataset based on potentially biasing features, which I did observe when not applying batch correction (see Figure 2A,B from von Kügelgen and Chekulaeva, 2020).

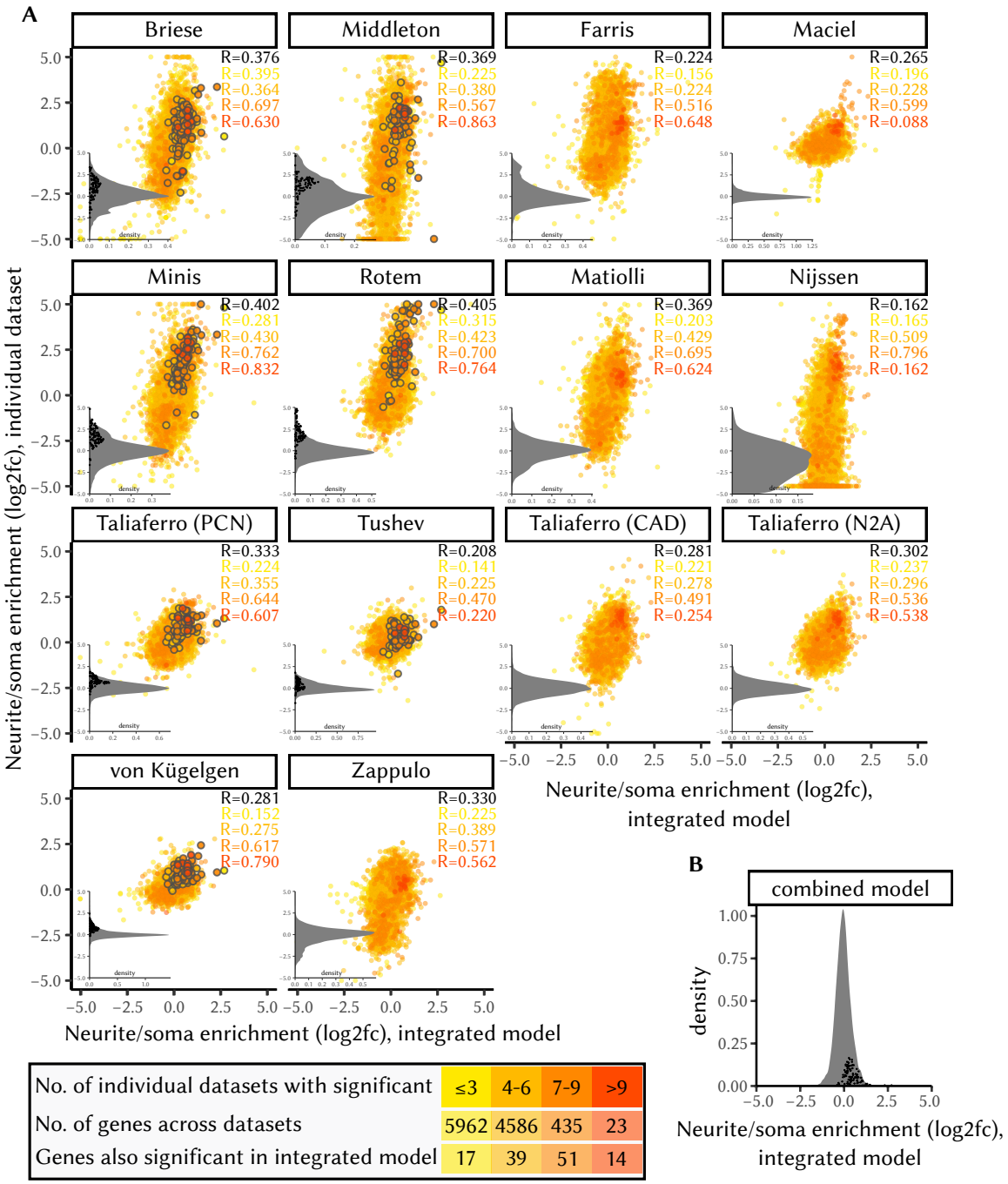


Figure 2: RNA enrichment in neurites across different datasets.
(A) Scatterplots showing the neurite/soma enrichment (log2 fold change) from the batch corrected model (RUV-Seq) of all high coverage datasets (x-axis, see also Fig. 1) as well as the enrichment based on individual datasets (y-axis, values above 5 or below -5 were restricted to the area, only genes with average TPM > 10 across all datasets are used). Genes that were chosen for inclusion in the Nzip MPRA based on a subset of datasets are marked with grey borders (see also Table S3). The hue of the dots represents the number of individual datasets in which a gene is significantly neurite enriched (absolute log2fc > 0.5, p.adj < 0.05). The legend for these colours also includes the number of genes enriched in individual datasets or in the combined model. The Pearson correlation coefficient (R) for the comparison of individual datasets to the integrated model also shown for all genes (black) and the different groupings based on significant enrichment across datasets. The inset graphs and panel (B) display the density distribution of log2 fold change values from (A) individual datasets or (B) the integrated model. The enrichment values of genes included in the Nzip MPRA are indicated by small black dots inside the density distribution of respective datasets or the integrated model. The PCN data from von Kügelgen et al., 2021 was obtained by Sayaka Dantsuji and all other data is publicly available.

4.1.2 Integrated model for RNA enrichment

In addition to the analysis of expression values from transcripts detectable in neurites, RNA-Seq datasets of separated compartments also allow to calculate the enrichment on transcripts in either compartment using differential expression analysis. In my previous work, the neurite enrichment of transcripts was only summarised based on the number of individual datasets that reported them as such. However, the batch corrected integrated model also allows to determine a single neurite/soma enrichment log2 fold change value on the basis of all datasets together. While the enrichment values from some individual datasets show comparatively wide distributions or tails from larger groups of genes enriched in one particular compartment (Figure 2A, insets), this is not the case for the values calculated with the integrated model. Additionally, there is no observable trend for stronger or weaker correlation of the combined model with individual datasets with wide or narrow distributions and with or without tails.

While a comparison of the integrative analysis with the enrichment values calculated from individual datasets indicates on overall agreement between all datasets, the correlation in any individual dataset in not very strong if based on all analysed genes (Figure 2A, black $R \lesssim 0.4$). However, transcripts with significant neurite enrichment in at least 7 datasets show a much higher correlation (orange/red $R \approx 0.5 - 0.8$) for most of the datasets. Additionally, a much higher percentage of genes with significant enrichment in at least 7 datasets is also significantly enriched based in the integrated model and respectively $2/3^{rd}$ of the neurite enriched genes from this model also show neurite enrichment in at least 7 individual datasets (Table 1).

These genes that show significant neurite enrichment in the combined model include several coding for ribosomal and mitochondrial proteins (Table 1 and S2), groups which both have been noted before in my (von Kügelgen & Chekulaeva, 2020) as well as other work (Turner-Bridger et al., 2020). Several other enriched transcripts belong to groups among which others have also been reported as neurite enriched before including cytoskeleton organisation cytoskeleton organisation and associated transport (Cgnl1, Cdc42bpg, Nes, Kif1c, Kif1a, Kif5c; Gumy et al., 2011; Turner-Bridger et al., 2020; von Kügelgen and Chekulaeva, 2020) or membrane trafficking (Rab13, Cep128; Costa et al., 2021). However, the combined analysis of the datasets also uncovers the neurite enrichment of transcripts coding for genes with nuclear function (Hmgn5, Trir), or others with elusive but still important functions in neurites (Ybx1; Lyabin et al., 2014; Nijssen et al., 2018; and Fth1/Ftl1, Reinert et al., 2019).

Table 1: Genes with significant (adj. p-value < 0.05) enrichment in neurites based on the integrated model of all high sensitivity datasets. Columns list gene names, number of individual datasets with significant enrichment in neurites, neurite/soma log2 fold change (log2fc) from the combined model, adjusted p-value from the combined model and mean expression. Only genes that are not ribosomal proteins, encoded in mitochondria and are not unnamed (Gm) are shown sorted by descending neurite/some log2fc values. A list of all genes with significant enrichment in either neurite or soma can be found in the supplementary table S2.

Gene name	Datasets with neurite enrichment	Neurite/Soma log2fc	Adj. p-value	Mean expression (TPM)
Rab13	8	2.322	0.0020	94.9
Sh3pxd2a	8	2.138	0.0025	41.7
Synpo	1	1.996	0.0350	36.0
Cep128	8	1.702	0.0087	29.0
Cgnl1	6	1.696	0.0001	11.5
Kif1c	8	1.485	0.0468	114.7
Cdc42bpg	8	1.481	0.0105	14.1
Nes	7	1.348	0.0218	655.7
Gab2	4	1.315	0.0423	66.6
Pabpc4	7	1.296	0.0019	82.5
Arhgap11a	8	1.250	0.0085	65.2
Hmgn5	5	1.238	0.0059	22.0
Kif1a	7	0.947	0.0119	445.6
Fth1	10	0.936	0.0106	944.8
Kif5c	6	0.917	0.0020	403.0
Eef1b2	9	0.804	0.0020	613.0
Trir	2	0.773	0.0420	66.5
Ftl1	10	0.756	0.0058	1,971.6
Dynll2	4	0.755	0.0266	261.9
Cox6a1	11	0.742	0.0066	431.3
Ndufa1	8	0.727	0.0117	127.4
Txn1	6	0.672	0.0166	238.7
Ybx1	11	0.666	0.0415	1,836.2
Ftl1-ps1	10	0.645	0.0356	1,421.1
Pomp	4	0.597	0.0089	84.6
Cox8a	10	0.594	0.0327	480.4
Cox7c	9	0.564	0.0307	348.2
Ndufs5	7	0.516	0.0194	221.5

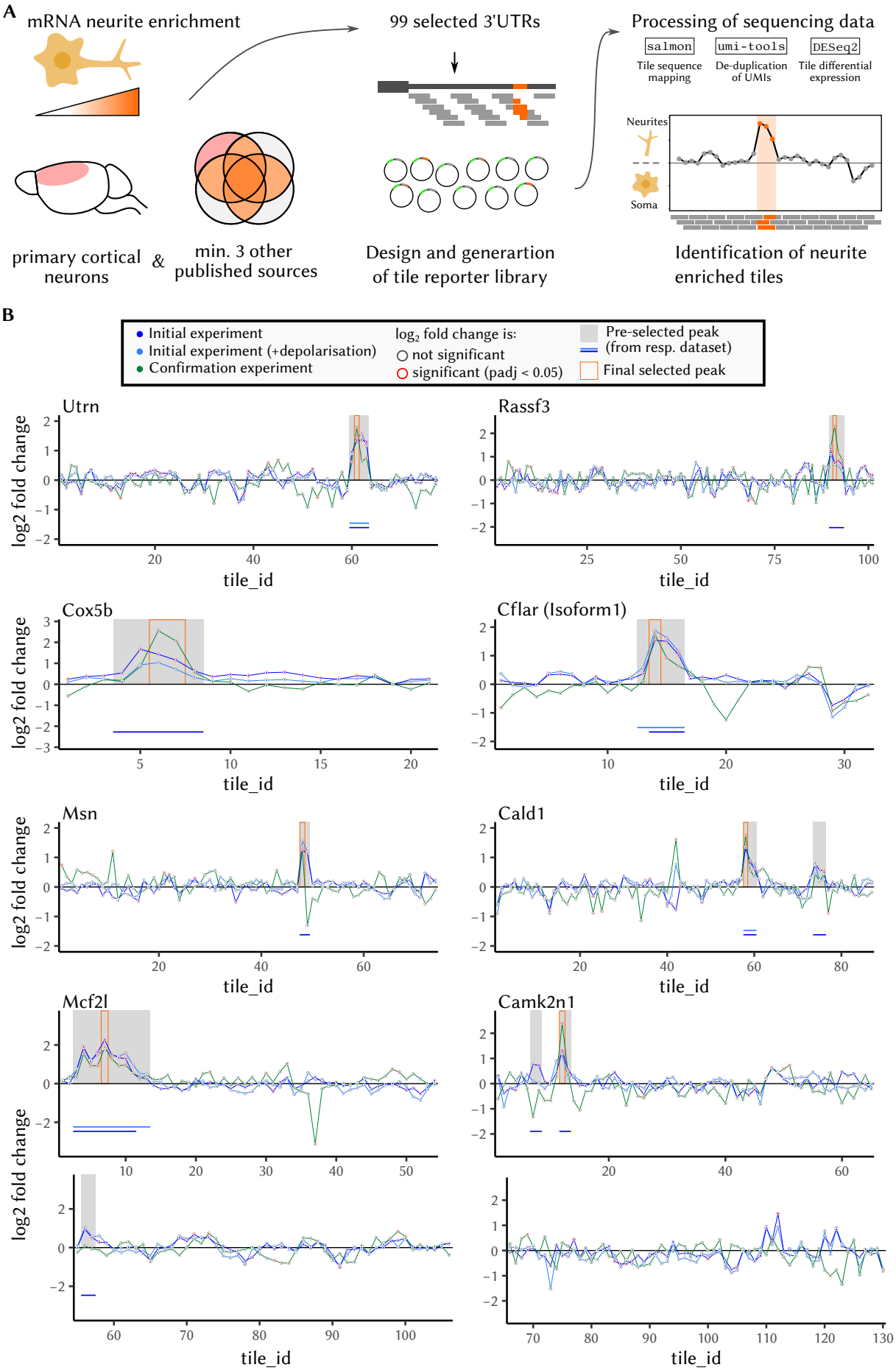
4.2 Neuronal zipcode localisation protocol (Nzip)

With the Nzip MPRA our group designed an approach to uncover novel zipcode sequences by investigating the potential of the 3'UTR region fragments of selected neurite enriched transcripts. The initial attempts at such an MPRA made by colleagues of my group were based on random fragmentation of 3'UTR amplicons (Yartseva et al., 2017) and transcripts selected from only a single dataset (Zappulo et al., 2017). However, this resulted in uneven coverage of reads and a low signal to noise ration for neurite enrichment. Therefore, switching to a design based on defined, overlapping, and *in vitro* generated tiles as well as a gene selection based on multiple data source, allowed me to aim for cleaner readouts. This was combined with depolarisation of neurons expressing the MPRA library, since this can also affect RNA localisation (Mofatteh, 2020). Additionally, the soma compartment was further separated into nuclei and cytoplasm as part of a collaboration project for the analysis of nuclear localised RNA. Later an additional experiment without these further parameters was added to confirm results in the baseline context of neurite/soma enrichment.

4.2.1 Gene selection and identification of candidate regions

To select transcripts with potential zipcodes to be investigated with Nzip, I relied both on the RNA-Seq enrichment data from the PCN model system, which I analysed for these experiments (Figure S1), and on public data from other, mostly primary, neuronal model systems. Focusing on genes with significant enrichment in our PCN system as well as a majority of these public datasets (Briese et al., 2016; Middleton et al., 2019; Minis et al., 2014; Rotem et al., 2017; Taliaferro et al., 2016; Tushev et al., 2018), I generated, after additional manual curation, a list of 99 transcripts to be included in the Nzip MPRA (Table S3 and Figure 2, highlighted dots). This list includes many genes of the functional classes with many enriched transcripts noted above, but due to differences in the analysis approach and included datasets does not have strong overlaps with the exact set of genes designated as commonly neurite enriched by my integrated model.

Figure 3 : Nzip MPRA indentifies neurite localisation sequences.
A) Scheme portraying the design as well as experimental and analysis strategy for the Nzip MPRA. First, genes with neurite localisation in both the PCN data as well as at least 3 other published datasets were chosen (see Table S3). The 3'UTR sequences of these genes were divided into 75 or 100 nt tiles overlapping by 15/25 nt each. All tiles were generated in vitro and cloned into a lentiviral reporter library. Reporter specific RNA-seq libraries where generated from neuronal compartments of library transfected PCN and after mapping of reads to tile sequences neurite/soma ratios were calculated from UMI counts for each tile.
B) Graphs displaying the neurite/soma enrichment (y-axis, log2 fold change) across the 3'UTR of selected genes fo the Nzip MPRA. The dots represent enrichment of individual tiles, while colours indicate from which experiment a value originates (initial, initial with depolarisation or confirmation); the borders around each dot indicate whether the corresponding enrichment value is significant (p.adj < 0.05, red border) or not (grey border). Sets of consecutive tiles with sufficient enrichment to meet threshold criteria (see methods) were preselected (grey highlight boxes), the coloured bars below these boxes designate from which experiment the pre-selection was done. The tiles manually selected for further investigation in the mutation library are marked with an orange border. A similar version of this analysis has been published in von Kügelgen et al., 2021 and data was obtained by Sayaka Dantsuji.


A
mRNA neurite enrichment
99 selected 3'UTRs
Processing of sequencing data
salmon
umi-tools
DESeq2
Tile sequence mapping
De-duplication of UMIs
Tile differential expression
Neurites
Soma
primary cortical neurons
&
min. 3 other published sources
Design and generartion of tile reporter library
Identification of neurite enriched tiles
B
Initial experiment
Initial experiment (+depolarisation)
Confirmation experiment
log2 fold change is:
not significant
significant (padj < 0.05)
Pre-selected peak (from resp. dataset)
Final selected peak
Utrn
Rassf3
Cox5b
Cflar (Isoform1)
Msn
Cald1
Mcf2l
Camk2n1
log2 fold change
tile_id

After gene selection, collection of 3'UTR sequences, and design of the tiled Nzip reporter library, the 3'UTR tiles were cloned into a plasmid construct, transfected into the PCN model system and after separation of compartments and generation of libraries sequenced (Figure 3A). Due to adapters surrounding the tile inserts, only reads for the tiles were generated, which I then mapped to the full sequence collection of tiles with salmon before de-duplication with umi_tools and finally calculated enrichment of tiles in compartments with DESeq2. Based on this analysis I designated groups of at least 2 partially overlapping neurite localised tiles with sufficiently strong localisation as peaks (log2fc of all tiles >0.5, with at least one >0.75 and one significant). The neurite enrichment of all tiles from several genes is shown in Figure 3B with the pre-selected peaks highlighted (grey background boxes), while the full list of all 54 peaks I could identify across all 3 Nzip experiments is given in Table S5.
No zipcode candidates were found that showed strong neurite enrichment only in the Nzip experiment including depolarisation treatment. Instead, the two experiments, which were performed in parallel to one another, had very similar results, especially including neurite enrichment for all designated zipcode candidate peaks. Additionally, the later added experiment confirmed the neurite enrichment of these peaks, even if the similarity of log2 fold change values for other tiles was not as high.

4.2.2 Mutational analysis of candidate regions

In order to identify the exact *cis*-active zipcode sequence elements a mutational analysis approach was used. Selection of 90-100nt peaks with zipcode candidates was based on the neurite/cytoplasm ratios of the initial two experiments, derived from preliminary analysis performed by collaborators. My peak selection criteria were similar to the final analysis, however, p-values and log2 ration were initially not directly linked, so I used them to get two selection sets of peaks. Additionally, tiles with positive enrichment in nucleus compared to cytoplasm or with overly strong de-enrichment were not considered for peaks. This resulted in a final and manually curated selection yielded 16 genes, for which one or two central tiles of the main peaks containing a candidate zipcode were chosen. The neurite enrichment of all tiles from the following Nzip included 3'UTR regions are exemplarily shown for all 3 Nzip experiments in Figure 3B: Cald1, Camk2n1, Cflar (isoform 1), Cox5b, Mcf2l, Msn, Rassf3 and Utrn. The complete list of genes from which tiles from selected peaks were chosen for further analysis includes Bdnf, Cdc42 (isoform 2), Cflar (isoform 2), Golim4, Kif1c, Map2, Ndufa2 and Rps23 and is listed in Table S4.

As the introduction of mutations into zipcode impacts their functionality, I generated all possible point mutations for each candidate sequence. Since multiple neighbouring or larger mutations can help with the exact definition and sequence specificity of the *cis*-active elements, I also added transversions

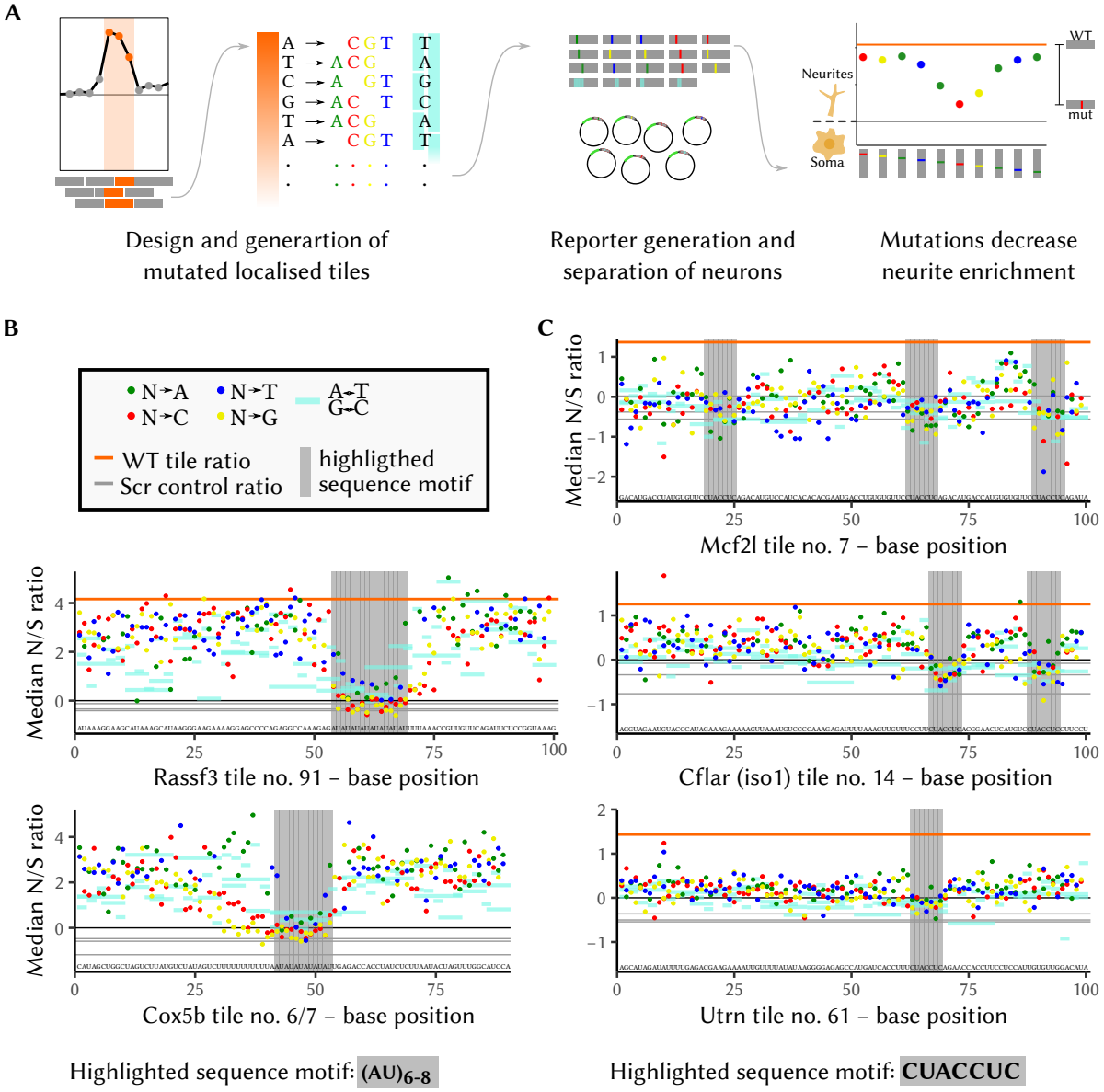


Figure 4: Nzip mutation libraries uncover zipcode sequence identity.
A) Scheme explaining the design strategy for mutational analysis using the Nzip MPRA approach. For neurite enriched candidate tiles from the Nzip MPRA all possible single point mutations as well as base transversions in 2, 5 and 10nt windows are created. After generation and analysis of the mutation library those tiles that show reduced neurite localisation compared to the WT sequence indicate the disturbation of an important zipcode sequence element.
B,C) Graphs displaying median neurite/soma ratios (n=3) for all mutations of exemplarily chosen candidate tiles. The new identity of each mutation is indicated by color of each dot (green - A, red – C, yellow – G, blue - T) or bar (cyan – transversion). Horizontal lines indicate the neurite/soma ratio of the WT tile (orange) or controls with a scrambled sequence (grey). Sequence stretches with a common motif that showed reduction of neurite/soma ratio in mutations are highlighted by grey background boxes, the sequences are (B) $(AU)_{6-8}$ and (C) CUACCUC. A similar version of this analysis has been published in von Kügelgen et al., 2021, data was obtained by Sayaka Dantusji and processed by Maya Ron and Igor Ulitsky.

(A↔T, C↔G) in 2, 5 and 10nt windows covering the sequence without overlaps to overall generate a second Nzip mutation library (Figure 4A). From this second Nzip MPRA library, containing all mutated candidate tiles as well as the original sequences and controls, the neurite enrichment of all mutated tiles was determined by collaborators,with experience in this approach (Lubelsky & Ulitsky, 2018).

The resulting traces of neurite/soma log2 ratios revealed that most mutations in the selected candidate tiles showed neurite enrichment similarly to the non-mutated tiles. Especially the tiles of Rassf3 (no. 91) and Cox5b (no. 6,7) show strong enrichment in neurites (median N/S log2 ratio ≈ 2 – 3) similarly to the WT tiles from either experiment (mutation library ≈ 4, initial libraries ≈ 2). However, any mutations in an AU-repeat sequence as well as the randomised scramble control sequences show no enrichment in neurites at all (Figure 4B). Additionally, the introduction of additional A's into U-stretches flanking the AU-repeat in the Cox5b tile even led to an increase in neurite enrichment (Figure 4B, green dots) and mutations to G or C in the repeat lead to a slightly stronger reduction of enrichment. For the candidate tiles of Mcf2l, Cflar (isoform 1) and Utrn neurite enrichment of most mutated tiles is somewhat weaker than for the WT tile, however all 3 carry one or more copies of a common CUACCUC sequence motif for which neurite/soma log2fc values are 0 or lower, corresponding to scramble controls (Figure 4C). In this way both the $(AU)_n$-repeat sequence as well as the CUACCUC motif are identified as potential zipcodes.

4.2.3 let-7 miRNA seed as a zipcode

Interestingly, the CUACCUC motif could be identified as an exact match to the seed sequence of the let-7 micro RNA. Since the primary effect of miRNAs binding their matching seed sequence is the degradation of their binding target, the destabilisation of tiles carrying an intact let-7 motif can be expected. Comparison of the mean count values from tiles carrying an intact or mutated let-7 seed sequenced in both compartments shows not only the expected strong reduction of RNA levels for tiles with an intact let-7 seed, but also indicates that, contrarily to tiles with a mutated seed, counts for tiles with the let-7 seed are significantly lower in soma compared to neurites (Figure 5A). Such preferential destabilisation in one subcellular compartment may induce RNA localisation and thereby provide a potential mechanism that allows the let-7 seed to act as a zipcode. With reason to assume that the let-7 seed is indeed a functional zipcode, we also confirmed that transcripts with one or more let-7 binding sites are indeed more likely to show neurite enrichment compared to those without one (Figure 5B).
Because localisation by targeted degradation requires stronger activity in one compartment the distribution of miRNA and their associated degradation machinery across the PCN model system was investigated. Based on small RNA-Seq data from PCN, we identified let-7 as the most expressed miRNA in both the neurite and soma compartment by a far margin (Figure 5C), even if relative enrichment could not be determined due to missing normalisation controls. Furthermore, using mass spectrometry data from compartment separated PCN, I was able to determine that the AGO and TNRC6 proteins, which facilitate miRNA mediated RNA degradation, are indeed enriched in the soma compartment (Figure 5D).

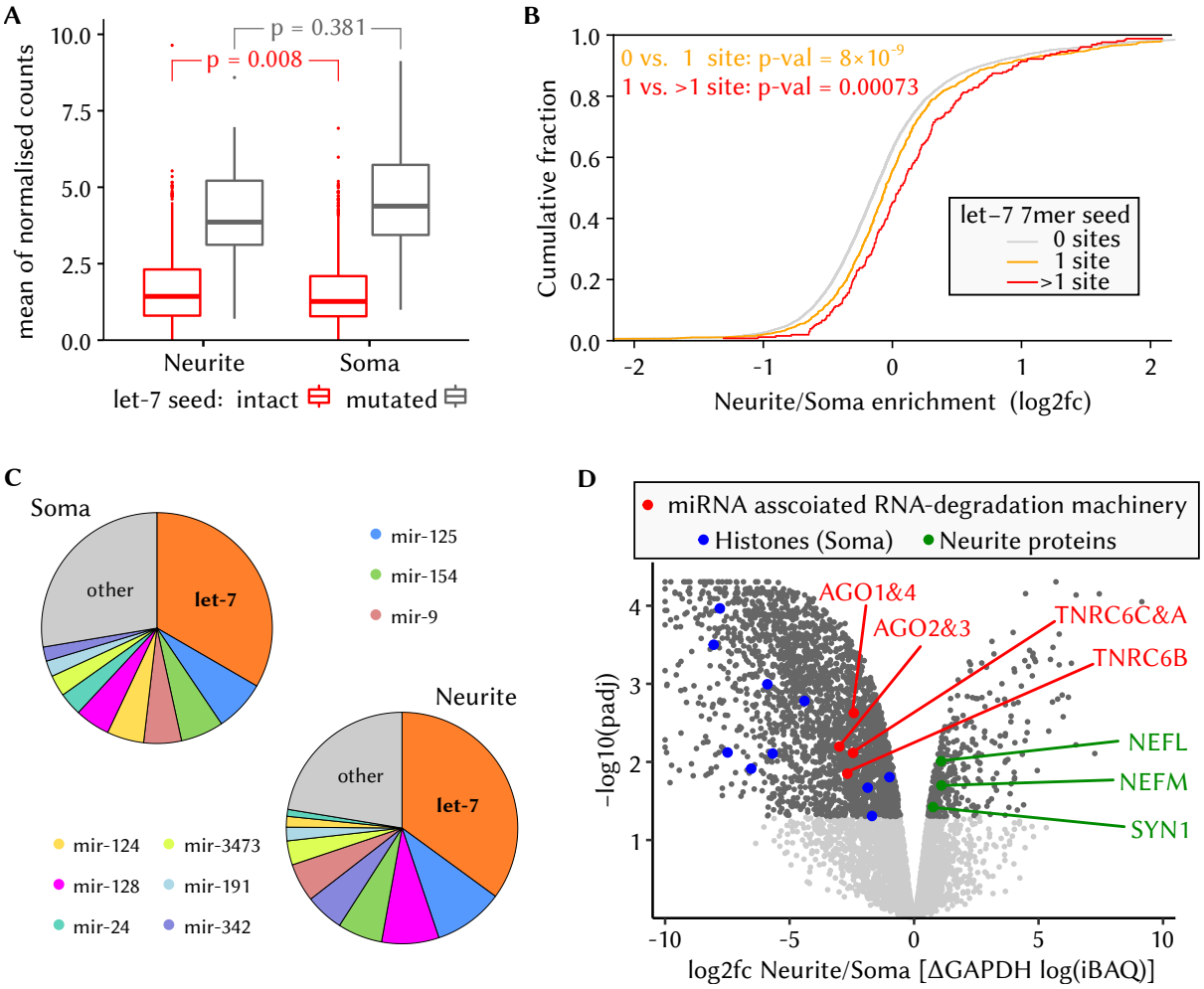


Figure 5: let-7 miRNA is active in neurons and may act as a zipcode through preferential RNA degradation in soma. (A) Boxplots displaying the normalised mean counts for tiles from the mutation library with an intact (red) or mutated (grey) let-7 seed site for each compartment. The p-values (two-sided t-test) for comparison between the compartments within each group of tiles are shown as well. (B) Cumulative distribution plot (ECDF) of transcript neurite/soma log2 fold change values stratified by the number of let-7 seed sites in the 3'UTR. The p-value for comparison of the distributions for different number of let-7 sites using the Wilcoxon rank sum test is shown as well. Only the most abundant transcript isoform for each gene was considered for analysis. (C) Pie charts displaying the relative abundance of different miRNA species in both soma and neurite compartment of PCN, based on counts from small-RNA-Seq data. The proportions of the top10 expressed miRNAs are designated by colours as shown. (D) Volcano plot displaying the mass spectrometry derived neurite/soma enrichment of proteins in PCN (log2 fold change of GAPDH-normalised log(iBAQ) values, x-axis) against the negative log-transformed p-values (BH adjusted, y-axis). Histone proteins with nuclear localisation are marked blue, proteins with prevalent localisation to neurites (NEFL, NEFM, SYN1) are marked and labelled green, and proteins that facilitate the miRNA mediated mRNA degradation are marked and labelled in red. This analysis has already been published in von Kügelgen et al., 2021. Data for A-C was obtained by Sayaka Dantsuji and processed by Maya Ron and Igor Ulitsky, data for D was obtained by Samantha Mendonsa and processed by Marieluise Kirchner.

4.2.4 $(AU)_n$ zipcode activity and binding proteins

For the AU-repeat zipcode no *trans*-factor with binding specificity is currently known. Since endogenous transcripts with long AU repeats of at least 6 repetitions also show a tendency for increased neurite enrichment (Figure 6A), effects of the motif presence in tiles on counts in was analysed similarly to the effect of the let-7 seed: an intact $(AU)_{\geq 5}$ repeat strongly reduced mean read counts from tiles indicating a destabilising effect and the reduction was also higher in soma compartment (Figure 6B).

To identify proteins that can bind the AU-repeat zipcode and facilitate the localisation, an RNA affinity pulldown experiment was performed. The $(AU)_8$ repeat of the selected Rassf3 candidate tile as well as a mutated version without the repeat motif were cloned into a bait construct with a boxB element for binding to GST-tagged lambdaN protein (Czaplinski et al., 2005). Using this setup, proteins from mouse brain lysate were captured and quantified using mass spectrometry and I identified several proteins, that showed preferential binding to the bait containing the $(AU)_8$ element. Among those proteins are 5, which also have known roles in RNA stability control: Elavl2,3 and 4 (HuB,C and D) are known to bind AU rich elements (ARE) in the 3'UTR of neuronal transcripts and thereby stabilise their targets (Allen et al., 2013; Ratti et al., 2006); Hbs1l is part of the no-go decay pathway (Doma & Parker, 2006), which can lead to mRNA degradation (Pisareva et al., 2011); and Rc3h2 (Roquin-2) is an RBP with specificity for a stem loop in the constitutive decay element (CDE) and thus leads to degradation of its targets by the Ccr4/Not complex (Leppek et al., 2013).

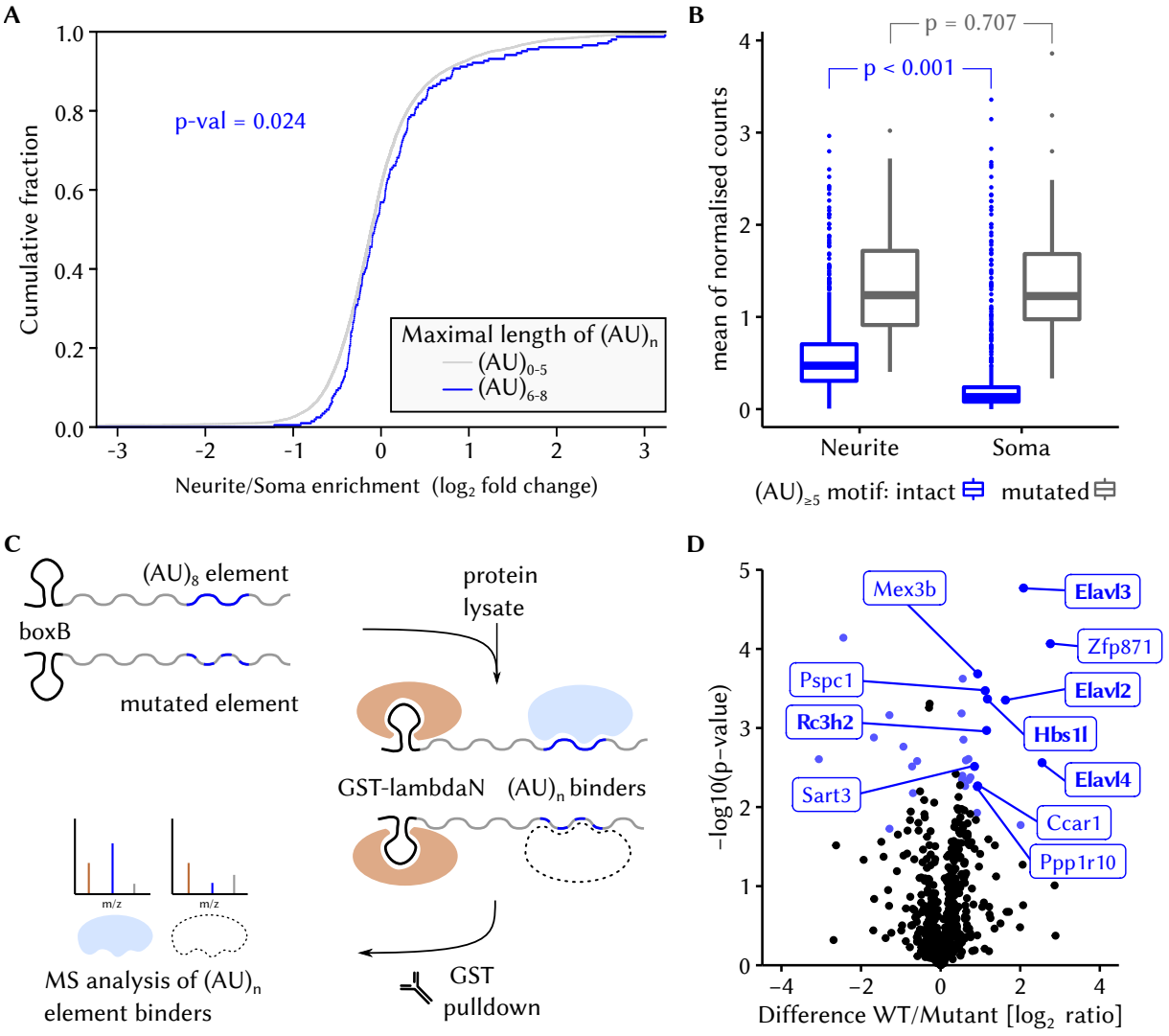


Figure 6: Potential effects and effectors of the $(AU)_n$ zipcode motif.
(A) Cumulative distribution plot of transcript neurite/soma log2 fold change values stratified by the longest AU repeat in the 3'UTR. The p-value for comparison of the two distributions using the Wilcoxon rank sum test is shown as well. Only the most abundant transcript isoform for each gene was considered for analysis. (B) Boxplots displaying the normalised mean counts for tiles from the mutation library with an intact (blue) or mutated (grey) $(AU)_{\geq 5}$ element from each compartment. The p-values (two-sided t-test) for comparison between the compartments within each group of tiles are shown as well. (C) RNA affinity pulldown schematic. RNA with both a boxB and either an $(AU)_8$ containing 3'UTR tile from Rassf3 or a mutated version without the $(AU)_8$ are combined with mouse brain protein lysate. After proteins with specificity for the $(AU)_8$ element are bound to the RNA, probe complexes are recovered via glutathione-S column and the GST-lamdaN-boxB unit and finally identified using mass spectrometry. (D) Volcano plot displaying enrichment in pulldown of the WT $(AU)_8$ or the mutated probe (x-axis) and negative log-transformed p-values. Proteins with significant adj. p-values (p.adj < 0.1) are coloured in light blue, proteins with significant enrichment (p.adj < 0.05 & log2fc > 1) are coloured dark blue and labelled. Proteins of specific interest with function in RNA stability control (Elavl2,3&4, Hbs1l, Rc3h2) (Allen et al., 2013; Leppek et al., 2013; Pisareva et al., 2011; Ratti et al., 2006) have highlighted labels. This analysis has already been published in von Kügelgen et al., 2021. Data for A-B was obtained by Sayaka Dantsuji and processed by Maya Ron and Igor Ulitsky, data for D was obtained by Marina Chekulaeva and processed by Marieluise Kirchner.

4.3 Human induced motor neuron model for the study of FUS-mutant neurons

The study of human neurons and neuronal diseases like ALS is challenging not only because of their complexity, but also because the non-proliferative nature of neurons prevents effective use of cell lines and other sources of cells are scarce. Therefore, the *in vitro* generation of neurons from human stem cells or hiPSCs is a promising approach to reach insight into the molecular processes in human neurons. In order to be able to rely on the analysis of *in vitro* generated motor neurons, it is important to check that the transcriptome and proteome of these cells recapitulates the known core characteristics of motor neurons.

4.3.1 Characterisation of motor neurons generated from hiPSCs

Differentiation of stem cells into motor neurons can be achieved either by activation of the endogenous expression programs through the exposure of cells to growth factors that activate lineage specific transcription programs or through the controlled expression of transcription factors that were made inducible by gene editing. For human motor neurons a protocol that relies on doxycycline (DOX) induced expression of the transcription factors NGN2, ISL1 and LHX3 (NIL) has been developed, which drives cells into a natural progression to a mature motor neuron fate (Fernandopulle et al., 2018; Mazzoni et al., 2013) (Figure 7A). However, no extensive characterisation of the transcriptome and proteome of iMN derived from this protocol exists so far.

The differentiation of iMN based on the NIL protocol takes 21 days and throughout this time course both protein and RNA expression was measured using mass spectrometry and RNA-Seq methods respectively. PCA shows that the overall expression signatures continuously change throughput the differentiation (Figure S2A,B). To asses the characteristics of the cells at different time points I investigated the relative expression levels of selected marker genes for stem cells, motor neuron precursors, neuronal cells and mature motor neurons at each measured time point (Figure 7B). The expression of stem cell markers that maintain pluripotency (NANOG, OCT4) rapidly decreases after differentiation is initiated by DOX addition. Notably, their protein expression is abolished faster than respective RNA expression levels. Early lineage markers of motor neuron precursor stages (OLIG1, OLIG2, SOX3) are expressed together with the DOX induced transcription factors NGN2, ISL1 and LHX3. Indeed their expression levels all subside, when DOX is removed and the cells reach an early motor neuron stage from which they continue to differentiate. HB9 is a marker of early postmitotic motor neurons (Arber et al., 1999) and peaks in expression during differentiation of the NIL-iMN at day 4 and 7. Furthermore, HB9 RNA expression can also be detected shortly before and continuously after these time points albeit at lower levels. Both HB9 and LHX3 are transcription factors that *in vivo* are primarily expressed in

developing spinal motor neurons of the median motor column (Dasen et al., 2008). Another transcription factor affecting the development of motor neuron lineages is FOXP1, which is usually expressed in neurons of the lateral motor column that are negative for LHX3 and HB9. Therefore the co-expression of HB9, LHX3 and FOXP1 in the NIL-iMN is an unusual combination of motor neuron lineage markers, which is further accompanied by a complete lack of expression of HOX genes that *in vivo* affect both rostro-caudal and motor column patterning (Dasen et al., 2008) (data not shown).

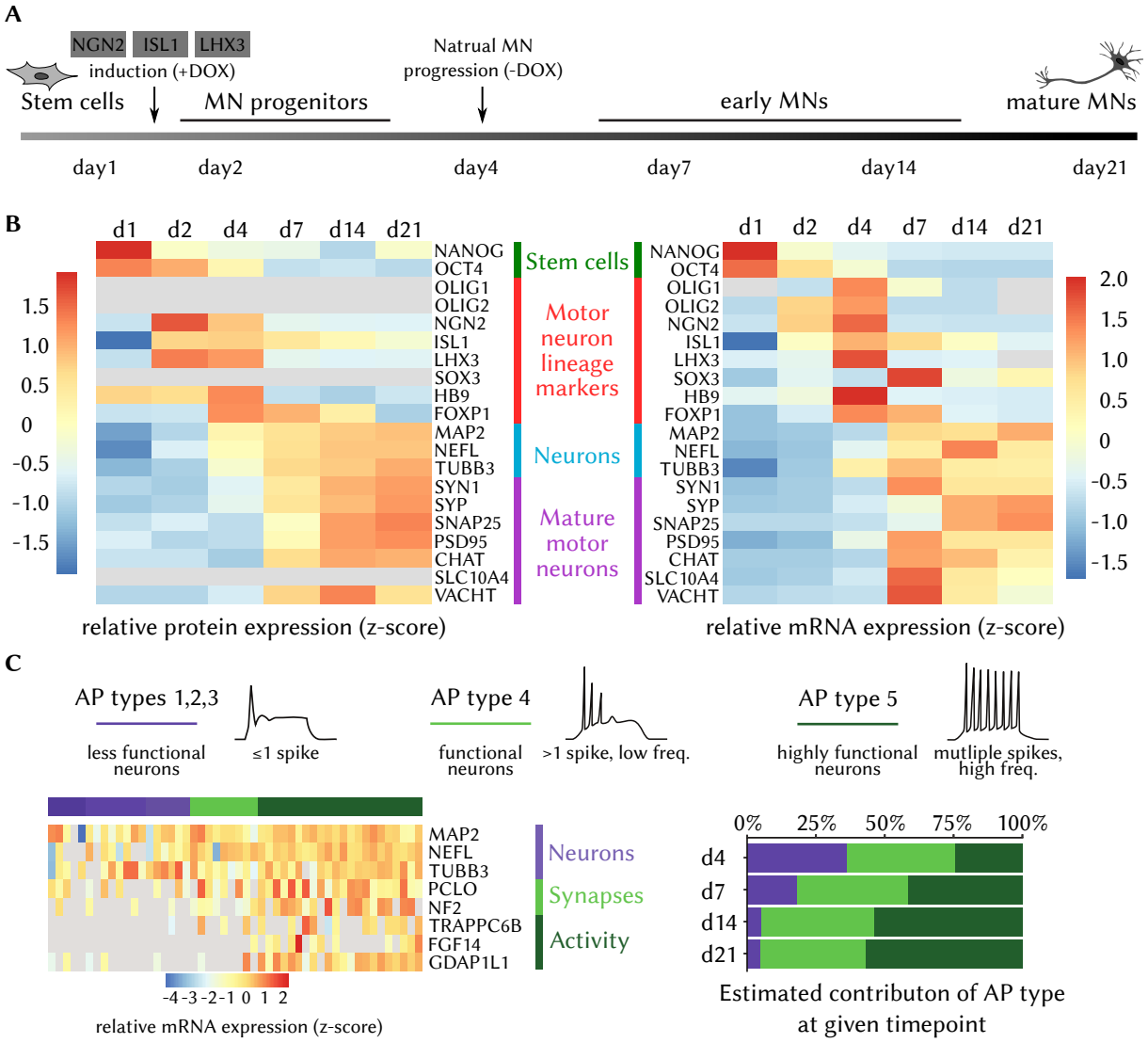


Figure 7: Characterisation of NIL induced human motor neuron differentiation based on transcriptome and proteome. (A) Scheme displaying the differentiation protocol for induced motor neurons (iMN) from human induced pluripotent stem cells (hiPCS) using doxycycline inducible expression NGN2, ISL1 and LHX3 over the course of 21 days. (B) Heatmaps of z-score normalised mean protein (left, log2 LFQ values) and RNA (right, log2 TPM values) expression of different marker genes at the indicated timepoints of the differentiation protocol. The displayed genes are grouped into markers of stem cells, motor neuron precursors, neuronal cells and mature motor neurons. (C) Celltype deconvolution of differentiation timepoints to different action potential types. Single cell sequencing data from patch clamped neurons links different action potential (AP) types (provided by Bardy et al. (Bardy et al., 2016), schematic adapted from Fig1D) with RNA expression markers for electric activity (heatmap, left). Cell type deconvolution of neuronal progenitor and neuronal stages with CIBERSORTx (Newman et al., 2019) estimates percentages of different AP types at different stages of differentiation (bar plot, right). iMN data was obtained by Katarzyna Ludwik.

Together with the motor neuron lineage marker HB9 other neuronal genes (MAP2, NEFL (Neurofilament), TUBB3) are also detectable from day 4 on and reach mostly stable levels by day7. Similarly, general synaptic proteins (SYN1, SYP, SNAP25, PSD95) and motor neurons markers of the cholinergic system (CHAT, SLC10A4, VACHT) increase in epxression from day 7 throughout the remainder of the differentiation time course. Again, the expression level increase at the mRNA level precedes that at the protein level.

The most important hallmark of neuronal maturity is the ability to conduct electric signals through continuous de- and re-polarisation. Using a single cell RNA-Seq dataset for which electric action potentials (APs) were measured and classified in individual neurons, the transcriptome signatures associated with neurons that show a functional AP type classification can be established (Bardy et al., 2016). Using these transcriptome signatures I performed cell type deconvolution with CIBERSORTx (Newman et al., 2019) to estimate the potential for electric activity at different stages of our motor neurons and precursors (Figure 7C). This indicates that from day 14 onwards a large majority of the NIL-iMN are likely capable of conferring action potentials.

4.3.2 Transcriptome and proteome of iMN sub-compartments

In addition to characterisation of iMN throughout the differentiation time course the subcellular transcriptome and proteome in soma and neurites of mature iMN was also analysed and I could see clear distinction between the compartments based on principle component analysis (Figure S2C,D). Specifically enriched transcripts and proteins included nuclear splicing factors (SRSF1-3) with respective somatic localisation and, for neurite enrichment, commonly neurite enriched genes (RAB13, KIF1C, KIF1A, KIF5C, NES; Table 1), synaptic markers (SYN1, PSD95) and neurite abundant cytoskeleton proteins (NEFL, NEFM, GAP43) (Figure 8A,B). Both the neurite RNA expression and neurite/soma enrichment values from iMN show positive correlation with values from the integrated analysis of other neuronal datasets (Figure 8C,D). This correlation is generally higher for genes with a strongly significant p-value for enrichment, which also coincides with a higher overall expression in iMN. For the neurite/soma enrichment of proteins in iMN no encompassing comparative reference data is available, so instead I used functional enrichment analysis to determine the association of neurite enriched proteins with annotated functions. Terms significantly associated with localised proteins that are also enriched in neurite localised transcripts are mostly associated with functions at the synapse, cytoskeleton or in outgrowth development, while metabolic and basic molecular functions are more common in soma (Figure 8E).

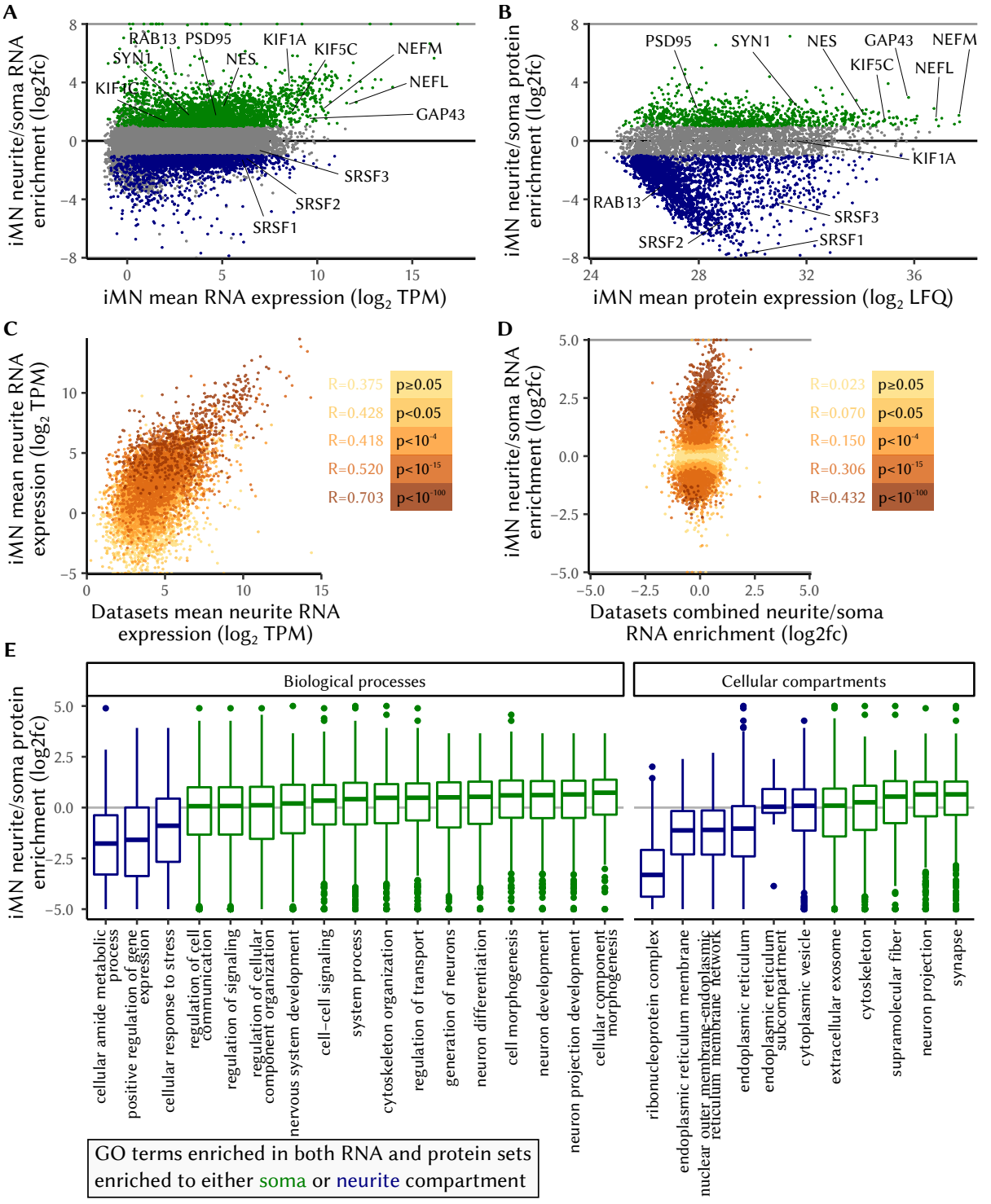


Figure 8: iMN display typical subcellular proteome and transcriptome.
(A,B) MA-plots displaying the relationships and distributions of mean expression and neurite/soma enrichment (log2fc) values for (A) RNA and (B) proteins. Proteins and RNA are coloured based on significant (p.adj < 0.05) enrichment (absolute log2fc > 1) in either compartment (neurites: green, soma: blue). Values exceeding the y-axis were restricted to the displayed values and overlap with the grey borders. Markers for soma (nuclear: SRSF1,2,3) and neurites (commonly enriched: RAB13, KIF1C, KIF1A, KIF5C, NES; cytoskeleton: NEFL,NEFM,GAP43; synapses: SYN1, PSD95) are labelled. (C,D) Correlation of (C) mean RNA expression and (D) neurite/soma enrichment (log2fc) between iMN (y-axis) and other neuronal datasets (see figures 1 and 2; x-axis). Genes are coloured by brackets of adj. p-values from iMN. Pearson correlation coefficient for all values (black) or by adj. p-value (by colour) are given as well. In (D) values exceeding the y-axis were restricted to the displayed values and overlap with the grey borders. (E) GO enrichment analysis was performed for proteins and RNA localised to either compartment and enriched terms were filtered for overlap between both analyses, a size no more than 1000 annotated genes and overlapping with at least 25 localised proteins. Finally all terms with directly related enriched terms in the GO graph relationship were removed. The box plots show the neurite/soma enrichment for all proteins annotated to the final filtered terms. Data was obtained by Katarzyna Ludwik and partially processed by Tommaso Mari.

4.3.3 Differential expression in iMN compartments due to mutant FUS protein

ALS is a neurodegenerative disease, in which dying of motor neurons starts at the tips of their axons all the while many molecular mechanisms including RNA metabolism and localisation are disrupted by the disease (J. P. Taylor et al., 2016). The use of the NIL-iMN system coupled with separation of subcellular compartments is a promising approach to identify potential disease mechanisms of ALS given the high importance of molecular processes in the axons. Therefore the differentiation protocol based on NGN2, ISL1 and LHX3 induction was used to create iMN from hiPSC lines with ALS associated mutations as well as from isogenic control lines. These hiPSC lines derived from one ALS patient carrying a FUS R244RR mutation and another patient not diagnosed with the disease, but carrying the P525L mutation that is known to usually cause ALS (Efimova et al., 2017). From these lines the transcriptome and proteome of totals cells as well as separated soma and neurites were measured by RNA-Seq and mass spectrometry. Based on principle component analysis all samples of the same compartment clearly show overall similarities, while the RNA expression signatures in neurites also somewhat differs between the hiPSC donor origins (Figure S3).

To compare the iMN of either FUS mutations to their isogenic controls I performed differential expression analysis in each of the three compartments for both the protein and RNA expression data. Based on the combined magnitude of significant differential expression I highlighted the top 15 genes with the strongest changes across transcriptome and proteome in each compartment and FUS mutation (Figure 9). To identify genes with consistent changes across all data modalities and compartments I also assigned a score to each gene based on the significance and magnitude of the differential expression between FUS mutant and control samples. This score is the sum of 0.5 or 1 points given for significant differential expression on RNA or protein level in any compartment of either FUS mutation (adj. p-value <0.05 and log2 fold change >1 or >1.5 respectively). All genes with a score >2 are listed in Table S6, corresponding to strong differential expression in at least 3 independent measurements or weaker differences in at least 5.

Among the genes that show both a high differential expression score and a strong differential expression in individual compartments are GPR50 and PCSK1, which both show a lower expression upon FUS mutation, especially in neurites. GPR50 is an orphan G protein-coupled receptor (GPCR) that can inhibit the melatonin receptor MT_1 (Levoye et al., 2006) and has also been reported to affect the outgrowth of neurites in the hypothalamus (Grünewald et al., 2009). PCSK1 is an essential protein for the regulation of neuroendocrine signalling, as it activates extracellular or plasma membrane bound hormones and neuropeptides by cleavage. Previous studies have also identified down regulation of PCSK1 as associated with ALS and attributed this to reduced glucose availability through non-cell

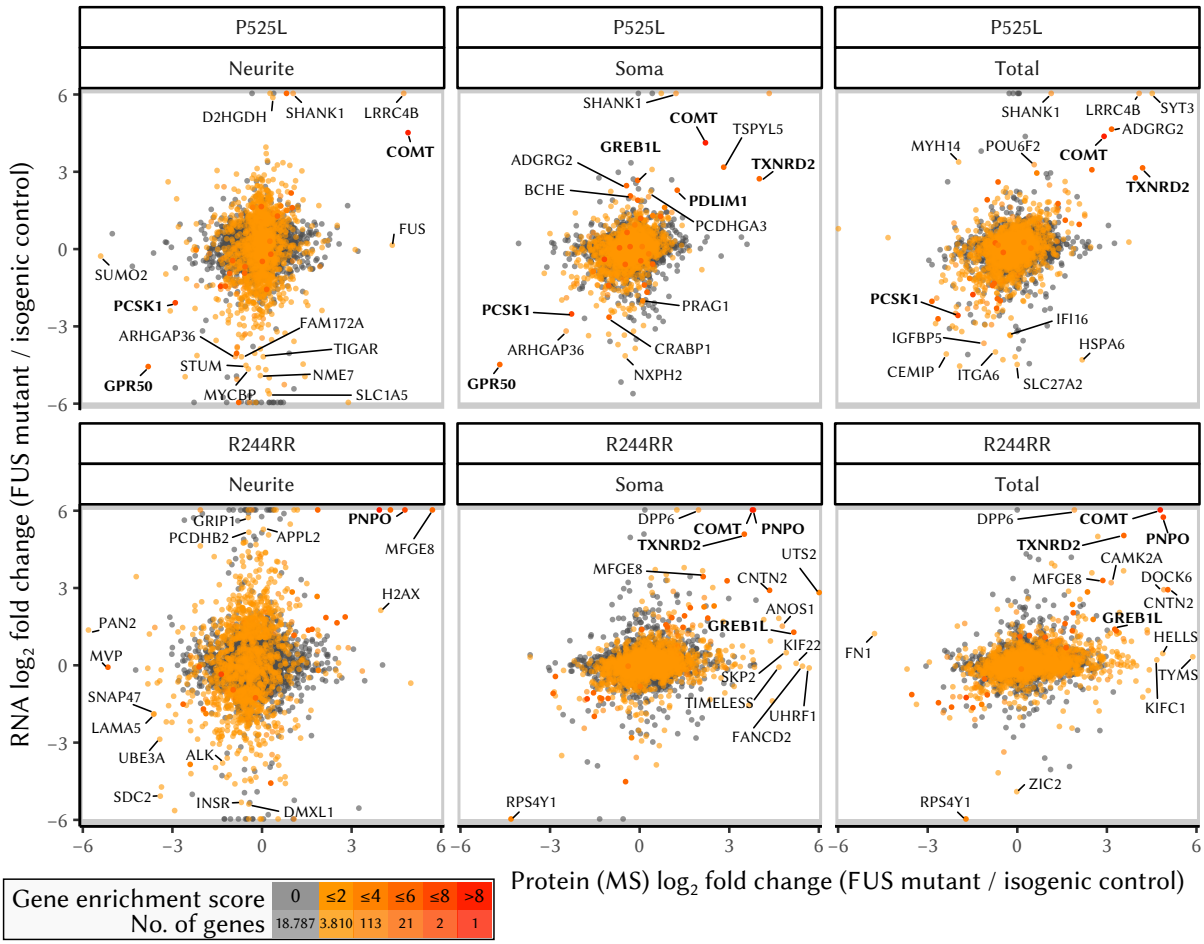


Figure 9: Effect of ALS associated FUS mutations on iMN proteome and trancriptome.
The differential expression on protein (x-axis) and RNA (y-axis) level between iMN with ALS associated FUS mutants and their isogenic controls are shown for two different mutations (P525L and R244RR) and three different compartments (neurites, soma, total cell). Enrichment values were restricted to the displayed values (-6 to 6) in this plot (higher values overlap the grey borders). Genes are coloured based on an enrichment score summarising the number of significant changes with a minimal magnitude across each cell line, compartment and expression modality (score increased by 0.5 or 1 for each significant change with log2 fold change >1 or >1.5 respectively). The total number of genes in different brackets of the enrichment score are shown in the colour legend. The 15 genes with the highest sum of significant absolute enrichment values in each panel are labelled; genes with an enrichment score of at least 6 are labelled in bold. Data was obtained by Samantha Mendonsa and partially processed by Tommaso Mari.

autonomous reduction of insulin secretion (Lederer et al., 2007). Other genes that show lower expression upon FUS mutations, especially in neurite compartment, and have potentially important functions in neurons or ALS include LAMA5, a neuron specific laminin gene involved in the formation of the neuromuscular junction (Maselli et al., 2018; Nishimune et al., 2008) and SNAP47, a membrane protein facilitating the fusion of vesicles with the plasma membrane, but does not result in neurotransmitter release (Arendt et al., 2015).

Genes with a high differential expression score that are also strongly up-regulated across both mutation lines include COMT, which is up-regulated in neurites and other compartments, as well as TXNRD2, GREB1L, and PDLIM1, which are only differentially expressed in soma and total cells. Among these, the transcripts of COMT, TXNRD2, and GREB1L are all direct targets of FUS based on CLIP data from

the wild type protein (see Table S6; Hoell et al., 2011). The enzyme COMT inactivates catecholamine neurotransmitters and chemically related hormones by methylation. While the protein itself has no known connection to ALS, its catalysed reaction depends on the the co-factor S-adenosly methionine (SAM), which can only be recovered by regeneration of methionine from homocysteine by the rate limiting MTHFR enzyme. Mutations in the enzymes of this pathway as well as increased homocysteine levels have been linked to neurotoxicity and neurodegenerative disease (Bhatia & Singh, 2015; McCaddon et al., 2002). TXNRD2 is a mitochondrial reductase that disposes of reactive oxygen species (Prasad et al., 2014) and regulates overall redox metabolism, which is commonly disturbed in ALS and specifically an isozyme of TXNRD2 has been linked to the disease (Mitchell et al., 2009). The last of the up-regulated direct FUS tragets, GREB1L, is a membrane protein with important functions in the development of sensory neurons of the inner ear as well as other tissues (Schrauwen et al., 2020). Another gene with consistent up-regulation, though only very strongly in soma of the P525L mutant iMN, is the cytoskeleton adaptor PDLIM1, which regulates neurite extension at growth cones (Ohno et al., 2009), is present at synapses (Ríos et al., 2020), and has also been predicted to be relevant for ALS based on feature extraction from gene expression data (Taguchi et al., 2015). Furthermore two other genes, LRRC4B and SHANK1, which functions at the synapse, also show much stronger expression especially in neurites of iMN with the P525L FUS mutation. Both of these genes play roles in the structural organisation of the synapse, either by mediating contacts with other cells (LRRC4B) or by affecting synaptic structure and organisation (SHANK1). Interestingly, several other genes with high differential expression in R244RR mutant iMN or with high scores (Table S6) have similar functions: both MFGE8 and TAGLN2 mediate cell-cell interactions and TAGLN2, KIF1C, and FCHSD1 all interact with the cytoskeleton. Finally, another gene with a high DE score, but only strong up-regulation in the R244RR iMN is PNPO, an enzyme involved in the metabolic cycle that produces hydrogen peroxide and regenerates the cofactor pyridoxal phosphate (PLP), which has various functions including the synthesis of several neurotransmitters (Wilson et al., 2019).

4.3.4 Functional associations of differential genes in FUS mutant iMN

In order to identify groups of proteins with similar functions I performed GO term enrichment analysis for all proteins up- or down-regulated between either of the mutant FUS iMN and their controls in any compartment. The top terms from the cellular compartment domain with the most significant enrichment are listed in figure 10A. Proteins that have lower expression in neurites or total iMN of mutant FUS are enriched for presence or function the basement membrane and other large extracellular complexes, especially including laminin. In addition to these the term 'neuromuscular junction' was

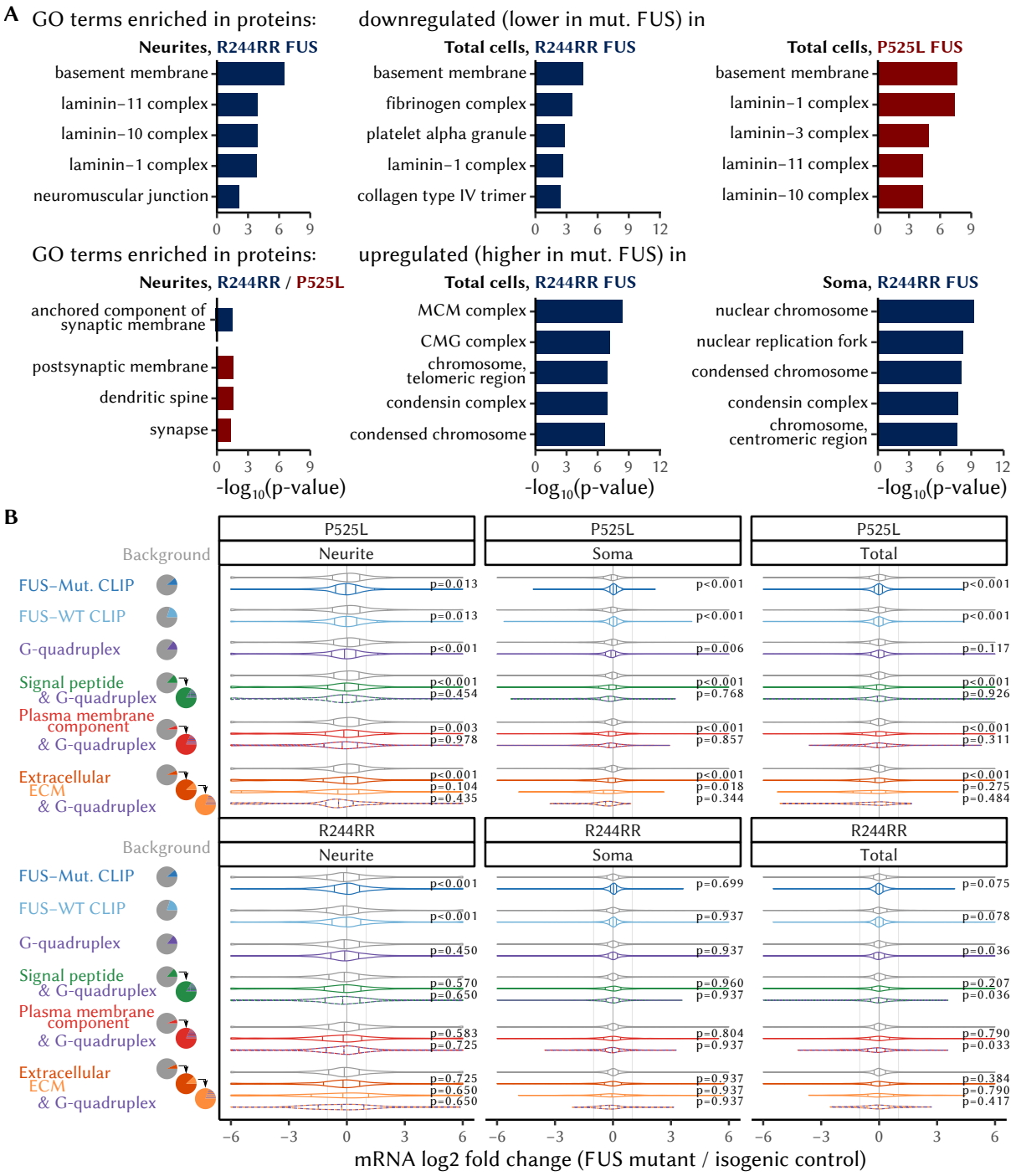


Figure 10: Functional analysis of differential expression between FUS mutants and wild type iMN.
(A) GO analysis of proteins with differential expression between FUS mutant and isogenic control iMN. The 5 terms from the GO cellular compartments domain (y-axis) with lowest p-value (highest negative log-transformed p-value, x-axis) are shown for each set of compartment, mutation and direction of differential expression. Before top5 selection, GO terms enriched in the initial analysis (gProfiler2) were removed if they had only one overlapping differential protein and then those that had enriched directly related terms above or below them were removed. (B) log2 fold change density distributions (horizontal violin plots) for different subsets of transcripts. The differential expression values between FUS mutant and isogenic controls for different mutations and compartments are divided into different groups based on features present on the RNA (G-quadruplex motif in the strongest expressed transcript isoform; CLIP peaks from mutant FUS or CLIP peaks from WT FUS (Hoell et al., 2011) or protein level (signal peptide from signalp-database (Armenteros et al., 2019; Durinck et al., 2009), plasma membrane component (GO:0005887, GO:0031226, GO:0046658, GO:0019897, GO:0044214), extracellular region (GO:0005576) or extracellular matrix (ECM, GO:0031012) GO annotation). Each set of values is compared to the remaining background or encompassing set and BH corrected p-values from two-sided t-test of all comparisons are shown as well. Data was obtained by Samantha Mendonsa and partially processed by Tommaso Mari.

also enriched due to lower expression of SLC5A7, EPHA4, LAMA5, EPHA7, SERPINE2, and LAMB2 in neurites of the R244RR FUS mutant. Up-regulated proteins have enriched terms for presence at synapses or synaptic membranes in neurites in both FUS mutants or for nuclear and DNA-associated complexes in soma and total cells of iMN R244RR FUS.

Because FUS is a known RBP, I compared the distribution of differential expression values of genes that have been identified as targets of mutant or WT FUS (Hoell et al., 2011) with the remaining background set of all other genes. I also performed such comparisons for genes whose transcripts have a G-quadruplex motif, which is a zipcode motif (Maltby et al., 2020; Subramanian et al., 2011) that has been reported to be bound by FUS (Imperatore et al., 2020), those with a signal peptide (signalP database, accessed via biomart (Armenteros et al., 2019; Durinck et al., 2009)) or those that are components of the plasma membrane, extracellular region or extracellular matrix based on gene ontology annotation. Furthermore, for the last 3 groups I added comparisons of subsets of these genes with those that also carry the G-quadruplex motif (Figure 10B).
The trends of RNA log2 fold change values I observed from these selected gene groups differ between iMN with the P525L or R244RR FUS mutations. In case of the P525L mutation transcripts that can be bound by FUS directly have reduced expression in neurites, but are centred around log2fc 0 in soma or total cells, where the remaining transcripts show an overall lower expression compared to the background of all other genes. For the R244RR mutation no significant changes from background can be observed in soma or total cells, while direct FUS targets in neurites are also zero-centred the remaining transcripts show an overall reduction in the context of this mutation. Genes with the G-quadruplex motif, signal peptides or protein function at the plasma membrane or extracellular regions show overall similar changes between FUS mutant iMN and controls: with the P525L mutation their expression is reduced compared to the background in every compartment, while no significant changes can be observed with the R244RR mutation. Further more, genes that are part of the extracellular matrix show an even stronger reduction than other extracellular genes with the P525L mutation. Finally, those genes that have annotation for a signal peptide or specific localised protein function and also carry the G-quadruplex motif show no difference to ones without the motif in all compartments of the P525L mutant iMN. With the R244RR mutation however, the genes with both of these two features do show reduced expression compared to genes that have the localised function but not the structural RNA motif, but this effect is only significant in total cells.

5 Discussion

5.1 The transcriptome of neurites

Together with the first discovery of mRNA in the processes of neurons many new questions emerged: is mRNA being locally translated? How many different transcripts are present in neurites? How is the localisation of transcripts controlled? And, what is the importance of localised mRNA in health and disease? While the occurrence of local translation in neurites is now indisputable (Besse & Ephrussi, 2008; Costa et al., 2021; Holt & Schuman, 2013; Shigeoka et al., 2016), the full extent and the percentage of localised and locally translated RNA species is not yet exactly determined, even if it has indeed been shown that local translation is major determinant of protein localisation in neurites (Zappulo et al., 2017). However, few of the remaining questions have conclusive answers so far. Using my research I hope to answer at least some of these open questions.

The first microarray based transcriptomics studies of neurites reported only a few hundred localised transcripts (Poon et al., 2006). However, with continuously improving sensitivity of transcriptomics methods, especially in RNA sequencing approaches, the number of detectable transcripts quickly rose thousands into the ten thousands. Based on my comparison of most of the available public RNA-Seq datasets, the true extent and complexity of transcripts present in neurites ranges from a minimal core neurite transcriptome of around 7500 transcripts present in most neurites to potentially twice that number.

Whether this number is accurate or if the differences between measurements of neurites from different neurons are actually correct is difficult to asses. Generally one can not fully exclude the possibility that the separation of neurites and soma is not clean enough to remove all contribution of the somatic transcriptome to the RNA material obtained for neurites, especially given the sensitivity of modern sequencing approaches. However, the transcripts detection cutoff of TPM > 1 that I used in my analysis is generally assumed to be a reasonable lower limit for truly expressed genes (Wagner et al., 2013) and at least in the data that I processed myself this also correlates well with the actual detection limit as determined by ERCC spike-in controls. Additionally my comparison encompasses datasets using several different techniques of separating neurites and soma, ranging from reliance on brain tissue organisation (neuropil) to microfluidic chambers channels with only a few microns width that are supposed to allow only growth of axons and I did not observe any indication that differences between these methods affect the number of detectable transcripts or the similarity between datasets positively or negatively.

One likely explanation for differences between neurites from different datasets is the existence of cell type specific neurite transcriptomes. However, I was only able to identify differences between neurites from primary or *in vitro* generated neurons with a PCA based analysis. Given that I observed

considerable variance between total expression values in neurites, especially among datasets from primary sources, it could also be possible that specifically designed analysis approach focusing on genes with variance corrected for cell types or species would allow a clearer distinction of neurite transcriptomes. In this line, controlling for the baseline expression of a core neurite transcriptome and focusing only the remaining or deviating genes, which are possibly unique to specific cell types, may prove a promising approach to identify cell type specific signatures. However, the differences between neurites from primary and *in vitro* sources might still confound such an analysis if they are not carefully controlled for.

Of course, it is also possible that there are indeed no identifiable differences across neurite transcriptomes from different cell types, even if some sources of variance clearly exist. While *in vitro* generated neurons may simply lack some characteristics of cultured primary or *in vivo* ones, there is also the possibility that the transcriptome in neurites is not homogeneous, but instead has further local differences of smaller subunits, which could differently affecting datasets. One possible indication for this is that studies focussing on sub-compartments like axonal growth cones (Poulopoulos et al., 2019) or other small neurite volumes obtained from nano-biopsies (Tóth et al., 2018) detect a much smaller number transcripts ($\approx 1100 - 3700$). While these approaches need to work with very low amounts of RNA and therefore likely have lower detection capabilities than broader RNA-Seq, it does pose the question of how different transcript sub-populations are distributed along neurites. It has also been proposed in the so called 'sushi belt model' that neurons dynamically shift the distribution of mRNA and RNPs dependent on the need of local translation at individual synapses (Doyle & Kiebler, 2011), which would provide an explanation for differing transcriptomes of smaller neurite sections.

5.2 RNA enrichment in neurites

Apart from the high detection sensitivity, one major advantage of RNA sequencing methods for the study of RNA localisation in subcellular compartments is the possibility to determine highly accurate relative expression level differences of individual transcripts between compartments. This allows one not only to determine which transcripts are very or commonly abundant in neurites, but also to designate localised RNA based on their relative enrichment in the neurite compartment. The tools used to determine this enrichment are the same used for calculation of differential expression in other circumstances and generally assume that the majority of transcripts does not show strong differences, whereas a few select ones do. This assumption might be counter-intuitive to the traditional view that RNA localisation to neurites is dependent on a machinery specifically dedicated to all or at least most localised transcripts, since the DE approach assumes that most transcripts, while not enriched, are at present to a similar degree in the compared groups. However, nearly all studies utilising RNA-Seq methods have used such

a differential expression approach and none saw reason to abandon it or use a different model due to sparser distribution of transcript counts from neurites. Indeed, for most datasets in my comparison I could observe a approximately normal distribution of log2 fold change values for neurite/soma ratios. While this is a strong indication that the default differential expression approach is indeed suitable for RNA-Seq data from separated compartments, it does not mean the majority of transcripts is present at similar absolute amounts in neurites and soma. Since the sequencing libraries are generated from the same amount of total RNA, the log2 fold change values for compartment localisation have to be interpreted as relative values between two equal sized *in vitro* samples and not within a neuron or culture. Accordingly, the total amount of RNA obtainable from neurites, while variable and dependent on the experimental setup, is around 5 to 10-fold lower than that available from soma. Furthermore, the used models may also shift the real distribution of log2 fold change values towards the assumed expectation and thereby mask smaller deviations that could very well exist.

There is a discrepancy between a traditional view on RNA enrichment, which would predict fewer transcripts present or enriched in neurites, and the actual observations, which show a tendency towards equal relative expression. This is in addition to what I have already noted in my previous work: mRNA that are often considered neurite markers or 'gold standard' localised transcripts, are not actually enriched in neurites (von Kügelgen & Chekulaeva, 2020). This can be explained by the fact that nearly all of these marker transcripts were identified by mRNA localisation studies utilising imaging approaches, which allowed the identification of transcripts that are very abundant in neurites, but not necessarily also enriched. One good example for this is the mRNA for β-Actin, which carries the first identified zipcode motif (Kislauskis et al., 1994) and, while being one of the most abundant transcripts in neurites, is not consistently enriched in neurites, but rather equally distributed (von Kügelgen & Chekulaeva, 2020). While enrichment of Actb mRNA at growth cones and other sites is functionally important for the growth of the actin cytoskeleton, there are several of these points across a cell, so a relative enrichment should maybe not be expected. This is contrasted by transcripts of other genes, with functions that are concentrated in neurites or synapses: both Camk2a mRNA, which is known to locally translated upon synaptic activity (Bagni et al., 2000; Dahm et al., 2007; Miller et al., 2002), and Bdnf mRNA, which has been known to be regulated by neuronal activity (An et al., 2008; Tongiorgi et al., 1997), are enriched in neurites of several datasets.

Based on these general assumptions, RNA localisation defined by differential expression of RNA between neurites and soma is only expected for transcripts with a functional role in neurites, since a cellular mechanism is needed to generate such an asymmetric distribution. Combined with the fact that diffusion or experimental inaccuracies might lead to spurious detection of transcripts in proximal

neurites this implied functional relevance of relative RNA enrichment has often been used as the primary determinant for RNA localisation in sequencing based studies, especially with larger numbers of transcripts detected in neurites over all. However, the term RNA localisation also often implies strong or consistent presence of a transcript in neurites, like it is the case several marker transcripts. Additionally, some experimental contexts, like imaging based studies, may assay RNA localisation without taking relative enrichment into consideration at all.

Even more than expression levels of transcripts their relative enrichment values from different studies have rarely been compared across a multitude of datasets before. Even though I relied on analysis of all raw data together, coupled with a batch correction approach to achieve harmonised integration of the enrichment values from different sources, my analysis does not show very strong correlation of the integrated model with any one individual dataset. However, there is a much stronger agreement for genes that are significantly neurite enriched also individually in multiple datasets. This indicates that neurite enrichment based on differential expression calculated from a default number of 3 biological replicates is likely influenced by many confounding factors and sources of noise and even an integrative approach, that should allow a systems view on neurite enrichment, still suffers from this noise between datasets. Differences in the neurite localised transcriptome of different neuronal cell types have been known to exist for a long time, but also effects from different *in vitro* culture and experimental conditions across the available datasets can not be excluded (Steward, 1997; Turner-Bridger et al., 2020). Without several independent samples or even datasets for each of these variables it is therefore difficult to properly adjust the model for all these factors in an integrated analysis approach. Therefore, I may have only been able to find consistent localisation for those transcripts that are neurite enriched with few exceptions in the currently available datasets. Interestingly, my approach identified a few genes with overly strong soma enrichment values (log2fc < -15, Table S2), which seem to represent transcripts exclusively detected in soma compartment of datasets, but never in neurites. In individual datasets, such an occurrence is often not statistically significant without a larger number of replicates, highlighting another strength of the integrated analysis.

As with any data analysis approach, the interpretation of numerical values or ranks for genes or transcripts, which anyway represent an aggregation from many different molecules, is difficult. Even if significant neurite enrichment in my integrated analysis does correlate with a higher number of enrichments across individual datasets, high log2fc values for neurite enrichment from either integrated or individual dataset analysis do not show an increased rate of clear correlation with significant neurite enrichment. This can be expected for individual studies where confounding factors are hard to control. However, it also means that even with an integrative approach it is ill advisable to focus on a certain number genes with the highest enrichment values, since those may well not be the most reliable or most

consistently enriched genes. Therefore, the need for independent and, ideally, experimental validation of neurite localisation of individual transcripts will continue to exists, while further improvements of the RNA-Seq analysis used to label transcripts as localised can also be made. For example, independent assessment of detection thresholds for true biological transcript expression in neurites and soma, may allow a better classification and normalisation of transcript levels in neurites. Furthermore, the differential expression models used to determine relative neurite enrichment could potentially be expanded to account for the lower expression signal to noise ratios of neurites and the higher likelihood of transcripts not being present there at all.

5.3 Neuronal zipcode identification protocol

Even though the question of exactly how to define a transcripts localisation state in neurites may require a clearer definition of RNA localisation and potentially additional research, the amount of neurite localised transcripts by far outnumbers the transcripts with known zipcode elements. While some examples of zipcodes or neurite localisation elements have been very well described, for most localised transcripts no specific molecular machineries controlling their transport or enrichment are known or understood. To tackle this problem the Nzip MPRA was designed such that specified 3'UTR pieces are inserted into a repoter construct to assay their influence of RNA enrichment in neurites. This approach offers several advantages, but also some limitations: while the use of *in silico* designed sequence fragments offers a lower resolution for the exact position of any zipcodes candidates identified, it is also much easier to process both experimentally and computationally. Compared to random fragmentation of full 3'UTR sequences, the limited number of individual designed fragment makes experimental dropouts much more unlikely and does not necessitate extensive coverage controls. Additionally, the mapping and quantification of reads from known and clearly defined sequences is much easier than mapping fragments of endogenous sequences back to their original position in the genome and averaging counts over windows. Independently of the tile design approach, the reliance on RNA-Seq means that RNA localisation of the reporter has to be quantified based on neurite enrichment values.

In my selection of 3'UTR sequences to include into the Nzip MPRA I was careful to specifically take those genes, for which I had already determined neurite localisation in our PCN model system based on RNA-Seq data as well as other datasets, to ensure biological validity and robustness of the reporter signal. However, this selection was based on an earlier version of my comparison of neurite localised transcripts across datasets, so it is possible that a selection procedure based on more datasets and also including a batch correction approach could produce a better gene selection from which even more candidate zipcodes might be identified.

While incremental improvements in the determination of the core neurite localised gene set can likely always be made and utilised, the Nzip library we did use proved successful as I was able to identify 54 candidate zipcode regions within 33 out of the 99 3'UTR sequences included in the Nzip library, of which several were also identified in multiple experiments. My peak selection criteria included a requirement of at least two neighbouring neurite enriched tiles, which ensures a sequence overlap of at least 75 nt and therefore reduces the likelihood of false positive candidates. This approach has been proven successful for the analysis of localisation elements before (Lubelsky & Ulitsky, 2018) and I further relied on the significance test from the statistical model, which is the most common approach to select candidate tiles from an MPRA (Mikl et al., 2021; Rabani et al., 2017; Shukla et al., 2018). Alternative analysis approaches focusing on read counts within smaller non-overlapping windows across multiple tiles exist and can also identify larger regions spanning multiple tiles (Shukla et al., 2018). However, the setup for Nzip is not well suited to investigate candidate regions with different sizes, so it proved easier to rely on the fixed tiles, which also makes experimental validation of selected candidates easier, since they can be amplified directly from the MPRA pool.

Keeping in mind that the enrichment based analysis of localisation in an MPRA like Nzip also affects how zipcodes can be identified, one should also remember that RNA localisation can not only be defined by relative enrichment. Indeed, among several of the transcripts known to be localised, only some were included in the Nzip library due to commonly detected neurite enrichment (Camk2a, Bdnf), while other's were only included as manually selected additions (Actb). However, even the known zipcodes of neurite enriched Bdnf and Camk2a transcripts were identified using imaging analysis (An et al., 2008; Huang et al., 2003; Subramanian et al., 2011) and I could not validate the same active elements in tiles overlapping them using Nzip. Given the difficulty of replicating previous results not based on RNA enrichment in neurites, it was also uncertain whether the common approach of sequence motif enrichment analysis within the zipcode candidates sequences would have reliable results. While this analysis is nearly always possible computationally, the results can be heavily impacted by irregularities and unmet assumptions in the input sequence set, which can make interpretation difficult. I chose not to rely on prediction of relevant sequences from motif analysis, because I had no way of mitigating the potential impact of several such factors: sequences overrepresented from overlapping tiles, structural motifs not easily detected, multiple motifs within candidate regions contributing differently to enrichment strength and significance, and possibly present localisation elements not contributing to enrichment in the MPRA context.

Instead of a computational approach to uncover a sequence identity for the zipcode candidates,

I designed the mutation library based on 16 of the candidates chosen from the analysis of the initial experiments. Inclusion of all possible point mutations for both maximal positional accuracy and sensitivity to sequence features allowed the identification of two zipcodes motifs: CUACCUC and $(AU)_n$. Additionally, I used mutations across larger stretches in order to allow the detection of zipcodes that are resistant to single nucleotide changes, which could include binding sites for RBPs with promiscuous binding. However, no motifs with a clear reduction of localisation effects only due to longer mutations could be identified. In addition to the identified motifs, the behaviour of the control tiles included in the Nzip mutation library allow some relevant insights: nearly all the control tiles with randomised sequences show a neurite log2fc of around 0, so an equal distribution between neurite and soma compartment. While this is of course expected from the statistical model, which assumes a equal distribution for most tiles and normalises read counts accordingly, it does also show that even exogenous GFP transcripts with random 3'UTR sequences will be present and detectable in neurites equally to soma. Since the Nzip system used the neuron specific synapsin1 promoter to avoid overexpression of the reporter, an unspecific baseline localisation of these GFP coding transcripts has to occur. While this replicates the behaviour of most endogenous transcripts in DE analysis, it also hightlights one shortcoming of relative enrichment measurements: contrary to house keeping genes in whole transcriptome measurements, the Nzip MPRA sequencing libraries do not contain any reads for which a clearly established equal ratio exists. This means that one can not exclude the possibility of count normalisation impacting the log2 fold change ratios, even if the relations between them are not affected by any bias.

Contrary to the randomised control tiles that show no enrichment to either neurite or soma compartment, the positive controls without mutations in the sequence of the selected zipcode candidate tiles, almost all show a stronger neurite enrichment than any of the tiles with single point mutations introduced to them. While mutations enhancing a neurite localisation signal should not be expected, they can occur for example by extending an $(AU)_n$ motif if it has nearby A or T stretches. However, if all mutations are assumed neutral, even a baseline variability of random measurement fluctuations should mean that most mutations are similar to and potentially even higher than the non-mutated control. Since all tiles within in Nzip were generated and treated in a single pool, no characteristic exists that should separate the non-mutated tiles from the mutated ones. Therefore, the fact that such difference is observable can be taken as an indication that not just the identified zipcode motifs themselves, but also their whole sequence context strongly contribute to the observed neurite enrichment of Nzip tiles. This could mean the identified motifs interact with sequence context clues like base distributions or some undetected structural features, possibly outside of the tile sequence itself.

5.4 Identified zipcode motifs

Using the Nzip mutation library two zipcode motifs that also correlate with localisation of endogenous transcripts could be identified: both the let-7 miRNA binding site and the $(AU)_n$ repeat control localisation most likely through affecting stability of the mRNA. The reduction of read counts observed in soma strongly indicates that preferential degradation of let-7 targets in soma induces their relative neurite enrichment. Since let-7 is the most abundant miRNA in both neurites and soma, the mRNA localisation is likely driven by the somatic enrichment of proteins involved in the miRNA degradation pathway or potentially by other factors controlling the activity of this pathway in a compartment specific manner. Furthermore, it is also possible that let-7 was the only miRNA for which an effect on localisation could be identified, because it is the only one with strong enough expression levels and therefore, potentially, the only one to cause a measurable effect on localisation. Indeed, the neurite enrichment of non-mutated let-7 target tiles in the Nzip mutation analysis was only around 2-fold ($\log2fc \approx 1$), which is just around the level for which a clear effect can be ascertained or even validated and much lower than that of tiles with the $(AU)_n$ motif.

The effect of mutations in the $(AU)_n$ stretches are a very strong match to the expectation of a mutation zipcode in the candidate tiles. Similarly to tiles with a let-7 binding site, tiles with the intact $(AU)_n$ motif showed much a lower amount read counts, especially in soma, indicating that this motif might also affect localisation through changes in mRNA stability. Compared to let-7, however, both the neurite enrichment and the change in read counts is much stronger, which could be linked to more asymmetric stability control or to additional features of this yet unidentified pathway. Several RNA binding proteins with the potential to bind the $(AU)_n$ motif in the Rassf3 candidate tile could be identified and more than one of those can affect the stability of its targets. Among these the Elval2,3 and 4 (also known as HuB, HuC and HuD) are known both to play important roles during the differentiation of neurons (Deschênes-Furry et al., 2003; Ratti et al., 2006) and to bind several neurite localised transcripts such as Bdnf and Gap43 (Allen et al., 2013; Beckel-Mitchener et al., 2002). However, these proteins also generally bind all AU-rich elements (ARE) and should not be overly sensitive to A↔T mutations within the $(AU)_n$ motif. Furthermore, they are known to stabilise their mRNA targets, rather than the destabilisation observed for the $(AU)_n$ zipcode motif (Allen et al., 2013; Beckel-Mitchener et al., 2002; Deschênes-Furry et al., 2003). Therefore, it is unlikely the Elavl proteins directly mediate the zipcode activity of the $(AU)_n$ motif, even if they probably do bind it and might thereby regulate its activity as a zipcode. While little is known about the Hbsl1 protein, except that it leads to degradation of transcripts with stalled ribosomes in the no-go decay pathway (Doma & Parker, 2006; Pisareva et al., 2011), initial experiments carried out by my colleagues indicate that it does indeed affect $(AU)_n$ mediated localisation

(von Kügelgen et al., 2021). Whether Hbs1l-linked destabilisation of $(AU)_n$ bearing transcripts truly occurs and if it would also be linked to translation, or potentially the common pattern of translation repression during mRNA transport (Pimentel & Boccaccio, 2014; Wells, 2006), is still unknown. Finally, Rc3h2, or Roquin-2, is a protein that binds the so called constitutive decay element (CDE) and thereby leads to the degradation of its tagrets via the Ccr4-Not complex. The minimal functional feature of the CDE is a stemloop, rich in U-A pairs with 3 free nucleotides that are most often UGU, but can also be UAU (Leppek et al., 2013). Given that a sufficiently long (AU)-repeat could quite easily assume such a stem loop structure, it is very plausible that the $(AU)_n$ zipcode mediated degradation is at least partially facilitated by roquin-2. However, as roquin-2 also has a secondary binding site for dsRNA as well as an E3 ubiquitin ligase function (Tan et al., 2014; Q. Zhang et al., 2015) and has even been shown to bind and stabilise some miRNAs (Srivastava et al., 2015), it is difficult to speculate on possible ways, in which it interacts with other factors that might also affect neurite enrichment of its targets.

Without further experiments it is not possible to say how exactly the RBPs with the ability to bind $(AU)_n$ and affect stability contribute to its function as a zipcode. It might even be possible that control of RNA degradation and localisation are independent effects of the RBPs, even though this seems unlikely as stronger somatic degradation was observed. Since it is often the case that many different RBPs bind a single transcript, it is very well possible that multiple factors also use the same binding site (Iadevaia & Gerber, 2015). Indeed, it was also shown that both in the case of the β-Actin mRNA, whose localisation is mainly facilitated by ZBP1, and with an ARE element in Gap43 3'UTR the Elavl4 protein (HuD) can bind the same zipcode region and potentially affect neurite localisation (Kim et al., 2015; Yoo et al., 2013). The potential interplay of multiple *trans*-factors binding the $(AU)_n$ zipcode motifs are nearly impossible to predict, but could provide complex regulation or even enhancement of the localisation effect leading to stronger enrichment than caused through degradation or differential let-7 activity alone.

5.5 Model of neurite localisation control

Similarly to other studies investigating RNA localisation signals in neurons, Nzip is built based on the understanding that the localisation of transcripts is controlled by many different factors each with their own binding site, target spectrum and one of several potential mechanisms (Turner-Bridger et al., 2020). The results I obtained from Nzip are mostly in accordance with these assumptions, as I was able to identify a relatively large number of potential zipcode candidate regions across many of the studied transcripts. Additionally, it is also generally assumed that most transcripts can have multiple zipcodes working together to control localisation and while I only found one potential localisation element for most transcripts included in the assay, it is not really clear how many one should expect to find. It has

also been suggested that, especially in the complex control networks utilised by neurons, many RBPs only show relatively weak binding specificity for individual transcripts to allow faster scanning and collaborative post-transcriptional control in combination with other RBPs or effector protein (Achsel & Bagni, 2016). This theory also fits the findings of a recent study by Mikl et al. (2021), which also utilised an MPRA to look for zipcode elements and raises the point that the effect of many RBPs on localisation is not strong enough to induce localisation by themselves even if they do provide a potentially necessary contribution to that effect. These effects of course also explain why the binding sites of such RBPs are very hard to identify as zipcodes using Nzip or similar MPRAs: since individually assayed tiles will most often carry only one RBP site or zipcode element, they need to be exceptionally strong to be biologically active. Considering the potential extent of zipcodes or RBP binding sites with weak signal strength and the general experimental variability inherit to any MPRA study, the results obtained with the first Nzip library are therefore overall very satisfactory.

A striking observation from my results is that both of the zipcodes discovered using the Nzip mutation library seem to drive localisation by inducing compartment specific increased degradation in soma. This raises the questions whether this mechanism of inducing neurite enrichment is potentially more prevalent in neurons or at least in the PCN model. For most other neuronal zipcodes mechanisms described so far, it is assumed if not shown that they link their bound transcripts as RNPs to the transport machinery (Song et al., 2015; Turner-Bridger et al., 2020), while RNA localisation through local degradation has mainly been described in systems such as *drosophila* oocytes. Another possibility that can not be excluded is of course, that the Nzip system itself may be biased towards detecting zipcodes utilizing a degradation based localisation mechanism.

Without any other explanation for the observed Nzip results, it is interesting to consider whether a potential bias for detecting stability induced localisation might be caused by reliance of enrichment as a measure of localisation. One possible explanation for such a connection could be the existence of different functional classes of zipcodes, that cause RNA localisation which either manifests through relative enrichment across a whole compartment or through transport induced presence at a specific site. Indeed, if this were the case one should expect localisation induced by differential stability to be more easily detectable by enrichment analysis, compared to other mechanisms, that increase the rate of active transport towards specific sites, like the β-Actin mRNA accumulating at growth cones, but do not necessarily affect the ratio of transcript between compartments.

Different classes of zipcodes that control distinct modes of RNA localisation do not neatly fit into a traditionally two-sided model of RNA localisation, in which subcellular localisation to a compartment like neurites has only a single state: either transcripts are actively localised somewhere or they are not

and mostly remain in soma. Based on this view, both the enrichment strength and relative localisation frequency of individual transcripts are then secondary measures further describing the already assigned localisation state. However, such a view of RNA localisation also does not accommodate the fact that the number of transcripts consistently detectable in neurites is very high: with at leas 7.500 and up to 15.000 transcripts, nearly a quarter or even half of the genome might be in some way localised neurites. Considering that often only a part of the whole genome is transcribed in a given cell, this means that, based on the number of transcripts detected by RNA-Seq measurements, a large majority of neuronal transcripts should be considered localised to neurites in a certain fashion. Furthermore, the enrichment analysis across datasets identifies only relatively few transcripts as enriched in soma and even fewer with exclusively somatic expression, which many assume to be the default state of any unspecific transcript and control tiles with randomised sequence inserts were also not soma enriched in Nzip.

To provide an explanation for the lacking overlap between gold standard localisation markers and commonly enriched RNAs as well as the presence of a very large number of transcripts in neurites and possibly the existence of different classes of zipcodes, I propose to use a more elaborate model for describing RNA localisation in neurons. If one assumes that the neurite localisation of any given transcript is always possible with a certain probabilistic distribution, the detection of most transcripts in neurites would be expected, potentially even that of all transcripts not specifically retained in soma. However, only transcripts for which mechanisms exist that increase the probability or extend of localisation into neurites, can become relatively enriched in neurites. Additionally, separate mechanisms could facilitate the transport or otherwise induced accumulation of specific transport to specific sites in either neurites or soma, without directly affecting the general propensity towards general enrichment.
This view is supported, and indeed also partially inspired, by a recently published theoretical model of RNA and protein localisation by Fonkeu et al. (2019), which relies on production, diffusion, transport and degradation rates of transcripts and proteins. The model presented by these authors uses such values reported for the strongly localised Camk2a mRNA with the use of probability distributions derived from solved differential equations. Due to the nature of this mathematical derivation, the model generates and predicts continuous density distributions values for RNA localisation along dendrites, but does in fact conclude that RNA concentrations are most likely continuously decreasing with distance in dendrites, but are still sufficient to achieve steady distribution of proteins due to local translation. However, if one extends the model to other genes, it still implies that most transcripts can reach at least proximal neurites, even with lower transport speeds. Furthermore, Fonkeu et al. (2019) also note that both changes in mRNA transport rates and in mRNA stability are poised to have strong effects on localisation of both RNAs and proteins. These two features of a transcript can of course strongly be influenced by zipcodes

whose associated *trans*-factors could connect transcripts to transport machinery to improve median transport speeds or affect degradation rates directly. Compared to degradation, the RBP facilitated connection to transport complexes might be weaker and also reversible, so that the effects of stability changes on RNA localisation may be easier to detect with enrichment based analysis, especially if the stability is affected disproportionally across compartments.

Finally, this model view of RNA localisation I am proposing does not only have to explain observations from experiments, but also match the biological reality and needs of neurons. One often mentioned requirement for directly controlled transport dependent localisation of RNAs in neurons is the length of human axons, which can reach up to nearly 1 meter. This could be taken as a potential counterargument for an RNA localisation model based on enrichment distributions reaching an equilibrium state, in which RNA levels likely decrease with distance. However, one has to keep in mind that neuronal RNA localisation is an evolutionary conserved process and indeed first evidence of functional mRNA in synapses was detected in the invertebrate sea slug *Aplysia* (Martin et al., 1997; Moccia et al., 2003). Also more recent work on the subcellular transcriptome of neurons from *C. elegans* showed typical agreement with singular studies from mammalian systems (Arey et al., 2019). This strongly indicates that the mechanisms for RNA localisation in neurons are evolutionary far older than the emergence of tetrapods or mammals with neurons that would require special mechanisms to accommodate for their extraordinary length. Additionally, one should keep in mind that all large RNA-Seq datasets generated so far use either *in vitro* cultured neurons or hippocampal tissue slices, neither of which allow for neurites longer than a few hundred microns. Therefore, it could well be the case that additional cellular systems exist that ensure proper localisation of transcripts along very long axons, which have not yet been identified, or that only select transcripts do reach such distances, where the theoretical models might not apply accurately.

5.6 Characterisation of NIL induced iMN

One large motivation for uncovering zipcodes and how they function is to deepen our understanding of how RNA localisation works. This is meaningful, because RNA localisation plays an important role for the health and function of neurons and many neurodegenerative diseases have been associated with disruption of normal RNA localisation. While it is not yet known whether this disruption is cause or effect of other facets of the diseases, studying them with an eye on RNA localisation is still worthwhile.

To properly study the role of RNA localisation in human neurodegenerative diseases like ALS, one needs a good model system of human motor neurons that also allows investigation of RNA localisation. In order to set up such a system, colleagues from the Chekulaeva lab established a differentiation protocol

of hiPSCs using doxycycline inducible expression of neurogenic transcription factors NGN2, ISL1 and LHX3. My analysis of the transcriptome and proteome of these iMN throughout their differentiation in this protocol revealed that they do express the expected marker genes of developing (HB9) and mature (CHAT, VACHT) motor neurons. However, the markers that usually distinguish different populations of the spinal motor columns are either not expressed at all (HOX genes) or have combinations not observed *in vivo* like the co-expression of LHX3 and FOXP1 (Dasen et al., 2008; Davis-Dusenbery et al., 2014). This is not overly surprising, given that the spatial patterning of developing motor neurons depends on tightly controlled gradients of retinoic acid and sonic hedgehog, which are not used at all in the NIL-iMN differentiation protocol. Nonetheless, the iMN express the typical markers of only spinal motor neurons (HB9, FOXP1) and not those of cranial motor neurons (PHOX2A) (Mazzoni et al., 2013), even if one has to assume that they either form a mixed population of different subtypes or represent a spatial expression pattern not present for spinal motor neurons developing *in vivo*.

Further on, the iMN from the NIL protocol reach maturity after already 21 days of differentiation and cell type deconvolution against single cells with electrophysiology measurements suggests that a majority of the iMN has the capability for electric signal transduction. However, this correlation based prediction still has to be confirmed with experiments to fully characterise the iMN system. Given that the purpose of the iMN system is the analysis of bulk samples traditional patch clamping assays may not be able to give a full overview over whole iMN cultures. Instead, other approaches like multi electrode arrays (Kapucu et al., 2022) or imaging based analysis of dyes sensitive to intracellular calcium levels (Farsi & Woehler, 2017) would be more suited to provide a more reliable and statistically interpretable overview of the electrophysiologic capabilities of the iMN bulk culture.

Analysis of the enrichment of transcripts in the subcellular transcriptome of either neurites or soma from iMN mostly showed the expected results: genes with previously described localisation, specific function, or localisation were found enriched in their respective compartments. Additionally, the correlation of expression and enrichment values with those from the integrated analysis of many datasets is on a similar level to correlation of individual studies included in the integrated model. For the comparison of proteomics data only very few possible reference datasets exist (Zappulo et al., 2017), but none from human or primary neurons. Instead, I assessed functional associations for localised proteins using GO analysis, where I found expected enrichment of terms for synapses and neuronal projections in neurites. Additionally, the overall log2 fold change distribution of all proteins annotated to terms enriched in neurite or soma localised genes followed the same pattern, indicating a clear preference of localisation according to function.

5.7 Effects of FUS mutations on subcellular transcriptome and proteome of iMN

The iMN system allows for the study of both healthy and disease model human motor neurons in parallel, which is not possible with other systems. Mutations in the FUS gene are known to cause ALS, and while many hallmarks of this disease have been described well, including the dying back of axons starting the NMJ, the exact molecular mechanism through which FUS or other proteins contribute to disease etiology are still a matter of active research.

My analysis of data from compartment separated iMN with different FUS mutations and their isogenic controls might uncover new insights specifically in neurites. Given the combination of multiple compartments and FUS mutations, any approach with the gaol of identifying genes with a potential influence of the disease, needs to integrate the different layers of data. To achieve this several approaches are possible, however I chose the simplest one and considered mainly those genes with overlapping differential expression between FUS mutant and control iMN across RNA and protein measurements from different compartments and mutations. More sophisticated models for the calculation of differential expression, that allow the consideration of interaction or dependency of differential expression on compartments or mutations directly can be constructed. However, formulating such a model also requires a specific hypothesis regarding the type of the interaction and is therefore not well suited for initial explorative analysis. Additionally, protein and RNA expression data can not easily be used as input into one singular analysis, especially for differential expression with methods like DESeq2, since they normally use different identifiers and more importantly have different numerical scales. However, changes in proteins levels are often easier to interpret in terms of their functional implications, while it is also known that RBPs like FUS affect the RNA homeostasis in cells. Therefore, I feel confident that genes with an overlapping differential expression across both different data types and biological models of ALS associated FUS mutations is a good approach to identify genes important in the disease progression.

Based on GO term enrichment analysis I identified distinct functional terms associated with both down and up regulated genes in FUS mutant iMN. Contrary to genes with lower expression in iMN with mutant FUS, which are mainly associated with laminins and the basement membrane, the functional associations of up-regulated genes are quite diverse. My analysis indicates synaptic function for some genes in neurites, but also chromosome and DNA replication associated proteins for soma and total cells. The latter group is unexpected as neurons should have no need for nuclear DNA replication, however it is also not represented by any of the individual up-regulated genes with high enrichment scores. Given that several other up-regulated proteins with high scores are direct targets of FUS, it is also possible

that these results stem from indirect consequences of the FUS mutations on gene expression or from off-target effects introduced during the generation of the different hiPSC lines. Interestingly, uniquely in neurites of the P525L iMN the FUS protein itself, but not its mRNA, is also up-regulated (Figure 9, upper left panel). However, this can most likely by explained by the loss of nuclear FUS localisation in this particular mutation and does not necessarily imply changes in FUS expression.

First among the up-regulated genes with synaptic function is COMT, an enzyme that inactivates catecholamine neurotransmitters like adrenaline by methylation. While motor neurons are not known to utilise these neurotransmitters, changes in catecholamine metabolising enzymes have previously been associated with ALS (Ekblom et al., 1993) and by-products of oxidative catecholamine catabolism, which is not mutually exclusive with methylation inactivation, have been implicated as relevant neurotoxic agents (Salauze et al., 2005). Interestingly, another gene related to neurotransmitters, the glutamate transporter SLC17A6, better known as VGLUT2, is also up-regulated (see table S6), especially in the context of the R244RR mutation. Increased glutamate levels can induce excitotoxicity in motor neurons, which is known to often occur in ALS (Foran & Trotti, 2009). Additionally, it was shown that reduced VGLUT2 levels improve motor neuron health in ALS, even though this does not halt disease progression and was so far mainly associated with transport level dependent glutamate release by other cell types (Schütz, 2005; Wootz et al., 2010).

Other functional associations of the diverse group of up-regulated and partially synapse localised genes include regulation of neurite growth, cell-cell contacts, the cytoskeleton and metabolism. Additionally, several individual up-regulated genes have mutations linked to ALS or affect some of common disease associated phenotypes. This diversity of up-regulated genes and ALS associated mechanisms highlights the multifaceted nature and complexity of FUS functionality especially in ALS progression, but also makes distinction between causative effects relevant for disease etiology and indirect or disease supporting effects harder. Interestingly, several of the most differential up-regulated transcripts can be bound by wild type FUS (Table S6), which makes it likely that most of these genes are perturbed due to irregularities of their usual regulation by FUS. Therefore, I think it is more likely that most of these diverse effects do not directly contribute coherently to disease etiology, but are rather indirect effects of FUS mutation and likely contribute to ALS progression and cellular stress levels through a perturbation of normal functions.

For the genes down regulated in FUS mutant iMN, I identified mainly laminins and other associations with the basement membrane as enriched functional associations. This was particularly evident in the neurite compartment containing the axons, where also proteins with association to the NMJ were less expressed than in control iMN. There is clear evidence that a proper build-up of the basal lamina and extracellular matrix around the NMJ is crucial for its maturation and synaptic function

(Heikkinen et al., 2020). Among the proteins with lower expression in FUS mutant iMN is also the neuron specific laminin 5 protein, which is particularly important for both maturation of the NMJ and neuromuscular signal transmission (Maselli et al., 2018; Nishimune et al., 2008). Interestingly, among the proteins with reduced expression both the ones with a strong score across all analyses (GPR50, PCSK1, LAMA5, SNAP47) and the remaining ones associated with the NMJ (SLC5A7, EPHA4, EPHA7, SERPINE2, LAMB2) are all either transmembrane proteins located in the plasma membrane or otherwise associated with secretory vesicles. Furthermore, aberrant splicing of transcripts coding for proteins with functions in cell-matrix adhesion, specifically of transmembrane or secreted ones, has previously also been associated with ALS (Rabin et al., 2010). Therefore, I believe that there exists a defect in the FUS mutant iMN that, especially in neurites, reduces expression of proteins which have to cross the secretory pathway, including several with important functions for the maturation of the NMJ. Indeed, the presence and function of an axonal secretory pathway composed of both endoplasmic reticulum and structures with golgi function has been uncovered over the last years (Luarte et al., 2018; Merianda et al., 2009). It is therefore possible that specific mechanisms linked to FUS or other ALS associated proteins might interfere with a neurite specific secretory pathway, which could offer an explanation of how ALS afflicted motor neurons are not able to maintain activity of NMJs and thereafter start degrading.

Based on this theory I also investigated the distributions of differential expression values from transcripts coding for proteins travelling through different stages of the secretory pathway. In addition, I performed the same analysis for direct mRNA targets of mutant and wild type FUS as well as transcripts with a G-quadruplex motif, which is a connecting point of several ALS associated proteins and can be bound by both FUS and TDP43 (Imperatore et al., 2020; Ishiguro et al., 2016). This analysis revealed that, indeed, the transcripts which code for proteins of the secretory pathway or those that carry a G-quadruplex motif show a trend of lower expression in iMN with the FUS P525L mutations. This shift of log2fc values is very similar to that on transcripts, which are known binding targets of wild type or mutant FUS protein, further indicating that disturbance of these gene sets might indeed be linked to FUS mutations and potentially play an important role in the disease progression of ALS. Interestingly, the difference I observed between these specific gene sets and their background is not only stronger in neurites, but also barely visible for the R244RR iMN and only significant for the P525L mutation, which abolishes the nuclear localisation of FUS. Therefore, the mutation specific reduced expression of these genes could be a toxic gain-of-function of FUS localised to the cytoplasm. While it is relatively clear that FUS could affect G-quadruplex carrying transcripts by direct interaction, there is so far not much evidence or any explanation of how it could directly interfere with transcripts designated for passage through the secretory pathway. While I observed marginally stronger log2fc shifts for transcripts with

both a G-quadruplex and coding for genes passing the secretory pathway, this effect was not significant. However, even though I did not test for this specifically, it might nonetheless be possible that there exists a big enough overlap between these two groups, which would lead to FUS affecting them in a similar way. Another possibility is that FUS interferes with neurite localisation or local translation of all these transcripts: since the G-quadruplex motif is known to act as a zipcode (Maltby et al., 2020; Subramanian et al., 2011), binding by FUS could block this function. Furthermore, transport to neurites and local translation of transcripts is known to occur in association with lysosomal and endosome vesicles (Cioni et al., 2019; Liao et al., 2019). Interestingly, the study by Liao et al. (2019) not only showed that disruption of ANXA11 mediated tethering between RNA and lysosome is linked to ALS, but also that FUS is another protein able to interact both with liquid phase separated RNA and with this type of lysosome. This provides another potential mechanism of how FUS can disrupt certain subsets of the local transcriptome and translatome specifically linked to the secretory pathway.

Taken together my exploratory analysis of the effects of FUS mutations show that there are many effects on the transcriptome and proteome which can be grouped and preliminarily attributed to distinct modi of FUS actions. Most striking in my eyes is the disruption of extracellular matrix and other genes that have to pass the secretory pathway, since they are not only more strongly affected in neurites, but also only affected by the cytoplasmic localised FUS mutant, which makes this most likely a toxic gain of function effect. However, this theory still requires considerable further investigations, since I can not clearly say whether there exists and overlap of FUS affected transcripts with the G-quadruplex motif and those coding for proteins of the secretory pathway. Additionally, the reduced expression of these genes in neurites could be affected at different stages of RNA localisation and local translation. While effects of FUS on transport and translation of RNA have been described before (Yasuda et al., 2013), potential effects on local translation of proteins that need to pass the secretory pathway are much harder to address since the possibility of local translation at the axonal endoplasmic reticulum (ER) is still controversial (González et al., 2018). Still, the dysfunction of local translation into an axonal ER might prove to be an explanation to the elusive etiology of ALS, not only because it is an so far understudied phenomenon, but also because the contact between ER and mitochondria or ER calcium storage could provide further links to other hallmark features of ALS like redox stress and glutamate excitotoxicity.

5.8 Conclusions & Outlook

In my work I have shown that the number of transcripts consistently detectable in neurites is very large and nearly in the same order of magnitude as the total transcriptome. However, modern definition of RNA localisation also relies on relative enrichment of transcripts in neurites, which encompasses a much smaller set of transcripts that is distinct from many top expressed neurite markers. Furthermore, the analysis of the Nzip MPRA led to the identification of two novel zipcode elements, the let-7 miRNA seed sequence and an $(AU)_n$ repeat sequence, both of which most likely lead to enrichment of transcripts in neurites by somatic RNA destabilisation. Contrarily, using enrichment based analysis I could not observe strong localisation signals for several known zipcodes included in the Nzip MPRA. These two key observations lead me to the hypothesis that the neuronal RNA localisation machinery might separately control the presence and enrichment of transcripts in neurites and that consequently zipcodes may affect either or both of these measures independently.
Stratification of localised RNA into groups of enriched transcripts and strongly expressed ones may allow for easier discovery of sequence motifs that are enriched in those distinct groups, but not all localised transcripts. Additionally, motif discovery for RNA localisation could likely be improved if a baseline probability of transcript presence in neurites is accounted for, since many tools available from genomics assume that motifs derive from a selection driven by a single factor. Furthermore, experiments with the aim to validate zipcode activity affecting two different measures of RNA localisation separately, should not only uncover whether there indeed exist distinct mechanisms to control enrichment and presence of RNA in neurites, but also improve sequence or motif based predictions about the compartmental distribution of a given transcript. Thus knowledge about zipcodes and their potential effects on transport speeds or stability could be integrated with a stochastic model of RNA localisation, in which any transcripts distribution depends on its probability to reach an equilibrium with a given percentage of molecules in neurites.

As the second part of my work I analysed transcriptomic and proteomic data from human NIL induced motor neurons first to characterise this model system, which proofed to show expected motor neuron and maturity hallmarks as well as similarities to other neuronal model systems including, compartment specific expression signatures. Building on this, I assessed the differential expression induced by ALS associated FUS mutations to identify sets of genes with strong responses across compartments and data modalities. This uncovered several different up-regulated genes with partially direct binding targets of FUS and often known ALS associations, as well as a larger trend of reduced expression of secretory pathway targeted genes especially in neurites of iMN with cytoplasmic P525L-FUS, specifically including extracellular matrix proteins important for the formation and function of the NMJ. A disruption of local

translation at the ER in axons could provide a mechanism for a toxic gain of function of FUS but also other ALS related genes, since it also provides possible links to other processes that are known to be disrupted in ALS. Additionally, I observed that transcripts with the G-quadruplex motif, which can be linked to ALS associated RBPs FUS and TDP43 as well as C9ORF repeats, show similar expression changes in the context of the cytoplasmic FUS mutant.

Building on these insights, further studies to determine the relevance of the axonal ER and nearby RNA clusters therefore seem to be a promising next step. This should include the determination of whether transport of RNA, local translation or export of proteins from a locally present ER in axons is most affected. Utilisation of the already existing data might allow targeted identification of proteins known to pass the secretory pathway, which are affected specifically on the level of localised transcripts or proteins. Furthermore, proximity labelling of RNA and proteins nearby the ER or other vesicular structures specifically in neurites as well as detection of newly translated transmembrane or secreted proteins in neurites could allow a better distinction of the exact genes processed through a local secretory pathway.

6 Bibliography

Achsel, T., & Bagni, C. (2016). Cooperativity in RNA–protein interactions: the complex is more than the sum of its partners. *Current Opinion in Neurobiology*, *39*, 146–151. https://doi.org/10.1016/j.conb.2016.06.007

Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E., & Carson, J. H. (1997). Transport and localization elements in myelin basic protein mRNA. *The Journal of cell biology*, *138*, 1077–1087. https://doi.org/10.1083/jcb.138.5.1077

Ainsley, J. A., Drane, L., Jacobs, J., Kittelberger, K. A., & Reijmers, L. G. (2014). Functionally diverse dendritic mRNAs rapidly associate with ribosomes following a novel experience. *Nature Communications*, *5*, 1–11. https://doi.org/10.1038/ncomms5510

Allen, M., Bird, C., Feng, W., Liu, G., Li, W., Perrone-Bizzozero, N. I., & Feng, Y. (2013). HuD Promotes BDNF Expression in Brain Neurons via Selective Stabilization of the BDNF Long 3′UTR mRNA. *PLoS ONE*, *8*, e55718. https://doi.org/10.1371/journal.pone.0055718

Amoroso, M. W., Croft, G. F., Williams, D. J., O'Keeffe, S., Carrasco, M. A., Davis, A. R., Roybon, L., Oakley, D. H., Maniatis, T., Henderson, C. E., & Wichterle, H. (2013). Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0906-12.2013

An, J. J., Gharami, K., Liao, G. Y., Woo, N. H., Lau, A. G., Vanevski, F., Torre, E. R., Jones, K. R., Feng, Y., Lu, B., & Xu, B. (2008). Distinct Role of Long 3' UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. *Cell*. https://doi.org/10.1016/j.cell.2008.05.045

Andreassi, C., & Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends in Cell Biology*, *19*, 465–474. https://doi.org/10.1016/j.tcb.2009.06.001

Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T. M., & Sockanathan, S. (1999). Requirement for the Homeobox Gene Hb9 in the Consolidation of Motor Neuron Identity. *Neuron*, *23*, 659–674. https://doi.org/10.1016/S0896-6273(01)80026-X

Arendt, K. L., Zhang, Y., Jurado, S., Malenka, R. C., Südhof, T. C., & Chen, L. (2015). Retinoic Acid and LTP Recruit Postsynaptic AMPA Receptors Using Distinct SNARE-Dependent Mechanisms. *Neuron*, *86*, 442–456. https://doi.org/10.1016/j.neuron.2015.03.009

Arey, R. N., Kaletsky, R., & Murphy, C. T. (2019). Nervous system-wide profiling of presynaptic mRNAs reveals regulators of associative memory. *Scientific Reports*, *9*, 20314. https://doi.org/10.1038/s41598-019-56908-8

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Bagni, C., Mannucci, L., Dotti, C. G., & Amaldi, F. (2000). Chemical Stimulation of Synaptosomes Modulates α-Ca^{2+} /Calmodulin-Dependent Protein Kinase II mRNA Association to Polysomes. *The Journal of Neuroscience*, *20*, RC76–RC76. https://doi.org/10.1523/JNEUROSCI.20-10-j0004.2000

Bardy, C., Hurk, M. V. D., Kakaradov, B., Erwin, J. A., Jaeger, B. N., Hernandez, R. V., Eames, T., Paucar, A. A., Gorris, M., Marchand, C., Jappelli, R., Barron, J., Bryant, A. K., Kellogg, M., Lasken, R. S., Rutten, B. P., Steinbusch, H. W., Yeo, G. W., & Gage, F. H. (2016). Predicting the functional states of human iPSC-derived neurons with single-cell RNA-seq and electrophysiology. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2016.158

Bassell, G. J., Zhang, H., Byrd, A. L., Femino, A. M., Singer, R. H., Taneja, K. L., Lifshitz, L. M., Herman, I. M., & Kosik, K. S. (1998). Sorting of β-actin mRNA and protein to neurites and growth cones in culture. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.18-01-00251.1998

Batish, M., van den Bogaard, P., Kramer, F. R., & Tyagi, S. (2012). Neuronal mRNAs travel singly into dendrites. *Proceedings of the National Academy of Sciences*, *109*, 4645–4650. https://doi.org/10.1073/pnas.1111226109

Becalska, A. N., & Gavis, E. R. (2009). Lighting up mRNA localization in *Drosophila* oogenesis. *Development*, *136*, 2493–2503. https://doi.org/10.1242/dev.032391

Beckel-Mitchener, A. C., Miera, A., Keller, R., & Perrone-Bizzozero, N. I. (2002). Poly(A) Tail Length-dependent Stabilization of GAP-43 mRNA by the RNA-binding Protein HuD. *Journal of Biological Chemistry*, *277*, 27996–28002. https://doi.org/10.1074/jbc.M201982200

Besse, F., & Ephrussi, A. (2008). Translational control of localized mRNAs: Restricting protein synthesis in space and time. *Nature Reviews Molecular Cell Biology*, *9*, 971–980. https://doi.org/10.1038/nrm2548

Bhatia, P., & Singh, N. (2015). Homocysteine excess: delineating the possible mechanism of neurotoxicity and depression. *Fundamental & Clinical Pharmacology*, *29*, 522–528. https://doi.org/10.1111/fcp.12145

Blichenberg, A., Rehbein, M., Müller, R., Garner, C. C., Richter, D., & Kindler, S. (2001). Identification of a cis-acting dendritic targeting element in the mRNA encoding the alpha subunit of CA2+/calmodul-independent protein kinase II. *European Journal of Neuroscience*, *13*, 1881–1888. https://doi.org/10.1046/j.0953-816X.2001.01565.x

Blichenberg, A., Schwanke, B., Rehbein, M., Garner, C. C., Richter, D., & Kindler, S. (1999). Identification of a *cis* -Acting Dendritic Targeting Element in MAP2 mRNAs. *The Journal of Neuroscience*, *19*, 8818–8829. https://doi.org/10.1523/JNEUROSCI.19-20-08818.1999

Briese, M., Saal, L., Appenzeller, S., Moradi, M., Baluapuri, A., & Sendtner, M. (2016). Whole transcriptome profiling reveals the RNA content of motor axons. *Nucleic Acids Research*, *44*, e33–e33. https://doi.org/10.1093/nar/gkv1027

Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*, *11*, 94. https://doi.org/10.1186/1471-2105-11-94

Burgin, K., Waxham, M., Rickling, S., Westgate, S., Mobley, W., & Kelly, P. (1990). In situ hybridization histochemistry of Ca2+/calmodulin-dependent protein kinase in developing rat brain. *The Journal of Neuroscience*, *10*, 1788–1798. https://doi.org/10.1523/JNEUROSCI.10-06-01788.1990

Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A., & Theis, F. J. (2019). A test metric for assessing single-cell RNA-seq batch correction. *Nature Methods*, *16*, 43–49. https://doi.org/10.1038/s41592-018-0254-1

Cajigas, I. J., Tushev, G., Will, T. J., Dieck, S. T., Fuerst, N., & Schuman, E. M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron*, *74*, 453–466. https://doi.org/10.1016/j.neuron.2012.02.036

Chekulaeva, M., Hentze, M. W., & Ephrussi, A. (2006). Bruno Acts as a Dual Repressor of oskar Translation, Promoting mRNA Oligomerization and Formation of Silencing Particles. *Cell*, *124*, 521–533. https://doi.org/10.1016/j.cell.2006.01.031

Chen, B. J., Ueberham, U., Mills, J. D., Kirazov, L., Kirazov, E., Knobloch, M., Bochmann, J., Jendrek, R., Takenaka, K., Bliim, N., Arendt, T., & Janitz, M. (2017). RNA sequencing reveals pronounced changes in the noncoding transcriptome of aging synaptosomes. *Neurobiology of Aging*, *56*, 67–77. https://doi.org/10.1016/j.neurobiolaging.2017.04.005

Chen, L., Dumelie, J. G., Li, X., Cheng, M. H., Yang, Z., Laver, J. D., Siddiqui, N. U., Westwood, J. T., Morris, Q., Lipshitz, H. D., & Smibert, C. A. (2014). Global regulation of mRNA translation and stability in the early Drosophila embryo by the Smaug RNA-binding protein. *Genome Biology*, *15*, R4. https://doi.org/10.1186/gb-2014-15-1-r4

Chin, A., & Lécuyer, E. (2017). RNA localization: Making its way to the center stage. *Biochimica et Biophysica Acta - General Subjects*, *1861*, 2956–2970. https://doi.org/10.1016/j.bbagen.2017.06.011

Ciolli Mattioli, C., Rom, A., Franke, V., Imami, K., Arrey, G., Terne, M., Woehler, A., Akalin, A., Ulitsky, I., & Chekulaeva, M. (2019). Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. *Nucleic Acids Research*, *47*, 2560–2573. https://doi.org/10.1093/nar/gky1270

Cioni, J.-M., Koppers, M., & Holt, C. E. (2018). Molecular control of local translation in axon development and maintenance. *Current Opinion in Neurobiology*, *51*, 86–94. https://doi.org/10.1016/j.conb.2018.02.025

Cioni, J.-M., Lin, J. Q., Holtermann, A. V., Koppers, M., Jakobs, M. A., Azizi, A., Turner-Bridger, B., Shigeoka, T., Franze, K., Harris, W. A., & Holt, C. E. (2019). Late Endosomes Act as mRNA Translation Platforms and Sustain Mitochondria in Axons. *Cell*, *176*, 56–72.e15. https://doi.org/10.1016/j.cell.2018.11.030

Colak, D., Ji, S.-J., Porse, B. T., & Jaffrey, S. R. (2013). Regulation of Axon Guidance by Compartmentalized Nonsense-Mediated mRNA Decay. *Cell*, *153*, 1252–1265. https://doi.org/10.1016/j.cell.2013.04.056

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13. https://doi.org/10.1186/s13059-016-0881-8

Costa, I. D., Buchanan, C. N., Zdradzinski, M. D., Sahoo, P. K., Smith, T. P., Thames, E., Kar, A. N., & Twiss, J. L. (2021). The functional organization of axonal mRNA transport and translation. *Nature Reviews Neuroscience*, *22*, 77–91. https://doi.org/10.1038/s41583-020-00407-7

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*, *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591

Czaplinski, K., Köcher, T., Schelder, M., Segref, A., Wilm, M., & Mattaj, I. W. (2005). Identification of 40LoVe, a Xenopus hnRNP D Family Protein Involved in Localizing a TGF-β-Related mRNA during Oogenesis. *Developmental Cell*, *8*, 505–515. https://doi.org/10.1016/j.devcel.2005.01.012

Dadon-Nachum, M., Melamed, E., & Offen, D. (2011). The "Dying-Back" Phenomenon of Motor Neurons in ALS. *Journal of Molecular Neuroscience*, *43*, 470–477. https://doi.org/10.1007/s12031-010-9467-1

Dahm, R., Kiebler, M., & Macchi, P. (2007). RNA localisation in the nervous system. *Seminars in Cell & Developmental Biology*, *18*, 216–223. https://doi.org/10.1016/j.semcdb.2007.01.009

Darnell, J. C., Driesche, S. J. V., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D., & Darnell, R. B. (2011). FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell*, *146*, 247–261. https://doi.org/10.1016/j.cell.2011.06.013

Das, S., Vera, M., Gandin, V., Singer, R. H., & Tutucci, E. (2021). Intracellular mRNA transport and localized translation. *Nature Reviews Molecular Cell Biology*, *22*, 483–504. https://doi.org/10.1038/s41580-021-00356-8

Dasen, J. S., Camilli, A. D., Wang, B., Tucker, P. W., & Jessell, T. M. (2008). Hox Repertoires for Motor Neuron Diversity and Connectivity Gated by a Single Accessory Factor, FoxP1. *Cell*, *134*, 304–316. https://doi.org/10.1016/j.cell.2008.06.019

Davis-Dusenbery, B. N., Williams, L. A., Klim, J. R., & Eggan, K. (2014). How to make spinal motor neurons. *Development*, *141*, 491–501. https://doi.org/10.1242/dev.097410

de Heredia, M. L., & Jansen, R.-P. (2004). mRNA localization and the cytoskeleton. *Current Opinion in Cell Biology*, *16*, 80–85. https://doi.org/10.1016/j.ceb.2003.11.002

Deschênes-Furry, J., Bélanger, G., Perrone-Bizzozero, N., & Jasmin, B. J. (2003). Post-transcriptional Regulation of Acetylcholinesterase mRNAs in Nerve Growth Factor-treated PC12 Cells by the RNA-binding Protein HuD. *Journal of Biological Chemistry*, *278*, 5710–5717. https://doi.org/10.1074/jbc.M209383200

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloe, D., Gall, C. L., Schaeffer, B., Crom, S. L., Guedj, M., & Jaffrezic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, *14*, 671–683. https://doi.org/10.1093/bib/bbs046

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, *440*, 561–564. https://doi.org/10.1038/nature04530

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Nostrand, E. L. V., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Burge, C. B. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2018.05.001

Donnelly, C. J., Willis, D. E., Xu, M., Tep, C., Jiang, C., Yoo, S., Schanen, N. C., Kirn-Safran, C. B., van Minnen, J., English, A., Yoon, S. O., Bassell, G. J., & Twiss, J. L. (2011). Limited availability of ZBP1 restricts axonal mRNA localization and nerve regeneration capacity. *The EMBO Journal*, *30*, 4665–4677. https://doi.org/10.1038/emboj.2011.347

Doyle, M., & Kiebler, M. A. (2011). Mechanisms of dendritic mRNA transport and its role in synaptic tagging. *The EMBO Journal*, *30*, 3540–3552. https://doi.org/10.1038/emboj.2011.278

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*, 1184–1191. https://doi.org/10.1038/nprot.2009.97

Efimova, A. D., Ovchinnikov, R. K., Roman, A. Y., Maltsev, A. V., Grigoriev, V. V., Kovrazhkina, E. A., & Skvortsova, V. I. (2017). The FUS protein: Physiological functions and a role in amyotrophic lateral sclerosis. *Molecular Biology*, *51*, 341–351. https://doi.org/10.1134/S0026893317020091

Ekblom, J., Aquilonius, S., & Jossan, S. (1993). Differential Increases in Catecholamine Metabolizing Enzymes in Amyotrophic Lateral Sclerosis. *Experimental Neurology*, *123*, 289–294. https://doi.org/10.1006/exnr.1993.1161

Engel, K. L., Arora, A., Goering, R., Lo, H.-Y. G., & Taliaferro, J. M. (2020). Mechanisms and consequences of subcellular RNA localization across diverse cell types. *Traffic*, *21*, 404–418. https://doi.org/10.1111/tra.12730

Farris, S., Ward, J. M., Carstens, K. E., Samadi, M., Wang, Y., & Dudek, S. M. (2019). Hippocampal subregions express distinct dendritic transcriptomes that reveal differences in mitochondrial function in CA2. *Cell reports*, *29*, 522–539. https://doi.org/10.1016/j.celrep.2019.08.093

Farsi, Z., & Woehler, A. (2017). *Imaging Activity-Dependent Signaling Dynamics at the Neuronal Synapse Using FRET-Based Biosensors*. https://doi.org/10.1007/978-1-4939-6688-2_18

Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., Chang, H. Y., & Ting, A. Y. (2019). Atlas of subcellular RNA localization revealed by APEX-seq. *Cell*.

Fernandopulle, M. S., Prestil, R., Grunseich, C., Wang, C., Gan, L., & Ward, M. E. (2018). Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology*, *79*, e51. https://doi.org/10.1002/cpcb.51

Fonkeu, Y., Kraynyukova, N., Hafner, A.-S., Kochen, L., Sartori, F., Schuman, E. M., & Tchumatchenko, T. (2019). How mRNA Localization and Protein Synthesis Sites Influence Dendritic Protein Distribution and Dynamics. *Neuron*, *103*, 1109–1122.e7. https://doi.org/10.1016/j.neuron.2019.06.022

Foran, E., & Trotti, D. (2009). Glutamate Transporters and the Excitotoxic Path to Motor Neuron Degeneration in Amyotrophic Lateral Sclerosis. *Antioxidants & Redox Signaling*, *11*, 1587–1602. https://doi.org/10.1089/ars.2009.2444

Forrest, K. M., & Gavis, E. R. (2003). Live Imaging of Endogenous RNA Reveals a Diffusion and Entrapment Mechanism for nanos mRNA Localization in Drosophila. *Current Biology*, *13*, 1159–1168. https://doi.org/10.1016/S0960-9822(03)00451-2

Furic, L., Maher-Laporte, M., & DesGroseillers, L. (2008). A genome-wide approach identifies distinct but overlapping subsets of cellular mRNAs associated with Staufen1- and Staufen2-containing ribonucleoprotein complexes. *RNA*, *14*, 324–335. https://doi.org/10.1261/rna.720308

Gao, Y., Tatavarty, V., Korza, G., Levin, M. K., & Carson, J. H. (2008). Multiplexed Dendritic Targeting of α Calcium Calmodulin-dependent Protein Kinase II, Neurogranin, and Activity-regulated Cytoskeleton-associated Protein RNAs by the A2 Pathway. *Molecular Biology of the Cell*, *19*, 2311–2327. https://doi.org/10.1091/mbc.e07-09-0914

Garner, C. C., Tucker, R. P., & Matus, A. (1988). Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites. *Nature*, *336*, 674–677. https://doi.org/10.1038/336674a0

Gomes, C., Merianda, T. T., Lee, S. J., Yoo, S., & Twiss, J. L. (2014). Molecular determinants of the axonal mRNA transcriptome. *Developmental Neurobiology*, *74*, 218–232. https://doi.org/10.1002/dneu.22123

González, C., Cornejo, V. H., & Couve, A. (2018). Golgi bypass for local delivery of axonal proteins, fact or fiction? *Current Opinion in Cell Biology*, *53*, 9–14. https://doi.org/10.1016/j.ceb.2018.03.010

Grünewald, E., Kinnell, H. L., Porteous, D. J., & Thomson, P. A. (2009). GPR50 interacts with neuronal NOGO-A and affects neurite outgrowth. *Molecular and Cellular Neuroscience*, *42*, 363–371. https://doi.org/10.1016/j.mcn.2009.08.007

Gumy, L. F., Yeo, G. S. H., Tung, Y.-C. L. C. L., Zivraj, K. H., Willis, D., Coppola, G., Lam, B. Y. H., Twiss, J. L., Holt, C. E., & Fawcett, J. W. (2011). Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. *Rna*, *17*, 85–98. https://doi.org/10.1261/rna.2386111

Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M.-S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., & Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, *507*, 195–200. https://doi.org/10.1038/nature13124

Haidet-Phillips, A. M., Hester, M. E., Miranda, C. J., Meyer, K., Braun, L., Frakes, A., Song, S., Likhite, S., Murtha, M. J., Foust, K. D., Rao, M., Eagle, A., Kammesheidt, A., Christensen, A., Mendell, J. R., Burghes, A. H. M., & Kaspar, B. K. (2011). Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nature Biotechnology*, *29*, 824–828. https://doi.org/10.1038/nbt.1957

Halpern, K. B., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., & Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. *Cell Reports*, *13*, 2653–2662. https://doi.org/10.1016/j.celrep.2015.11.036

Heikkinen, A., Härönen, H., Norman, O., & Pihlajaniemi, T. (2020). Collagen XIII and Other ECM Components in the Assembly and Disease of the Neuromuscular Junction. *The Anatomical Record*, *303*, 1653–1663. https://doi.org/10.1002/ar.24092

Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A., & Mann, M. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell*, *163*, 712–723. https://doi.org/10.1016/j.cell.2015.09.053

Heraud-Farlow, J. E., Sharangdhar, T., Li, X., Pfeifer, P., Tauber, S., Orozco, D., Hörmann, A., Thomas, S., Bakosova, A., Farlow, A. R., Edbauer, D., Lipshitz, H. D., Morris, Q. D., Bilban, M., Doyle, M., & Kiebler, M. A. (2013). Staufen2 Regulates Neuronal Target RNAs. *Cell Reports*, *5*, 1511–1518. https://doi.org/10.1016/j.celrep.2013.11.039

Hester, M. E., Murtha, M. J., Song, S., Rao, M., Miranda, C. J., Meyer, K., Tian, J., Boulting, G., Schaffer, D. V., Zhu, M. X., Pfaff, S. L., Gage, F. H., & Kaspar, B. K. (2011). Rapid and Efficient Generation of Functional Motor Neurons From Human Pluripotent Stem Cells Using Gene Delivered Transcription Factor Codes. *Molecular Therapy*, *19*, 1905–1912. https://doi.org/10.1038/mt.2011.135

Hoell, J. I., Larsson, E., Runge, S., Nusbaum, J. D., Duggimpudi, S., Farazi, T. A., Hafner, M., Borkhardt, A., Sander, C., & Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. *Nature structural & molecular biology*, *18*, 1428–31. https://doi.org/10.1038/nsmb.2163

Holt, C. E., & Schuman, E. M. (2013). The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*, *80*, 648–657. https://doi.org/10.1016/j.neuron.2013.10.036

Huang, Y. S., Carson, J. H., Barbarese, E., & Richter, J. D. (2003). Facilitation of dendritic mRNA transport by CPEB. *Genes and Development*. https://doi.org/10.1101/gad.1053003

Iadevaia, V., & Gerber, A. P. (2015). Combinatorial control of mRNA fates by RNA-binding proteins and non-coding RNAs. *Biomolecules*. https://doi.org/10.3390/biom5042207

Imperatore, J. A., McAninch, D. S., Valdez-Sinon, A. N., Bassell, G. J., & Mihailescu, M. R. (2020). FUS Recognizes G Quadruplex Structures Within Neuronal mRNAs. *Frontiers in Molecular Biosciences*, *7*. https://doi.org/10.3389/fmolb.2020.00006

Ishiguro, A., Kimura, N., Watanabe, Y., Watanabe, S., & Ishihama, A. (2016). TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes to Cells*, *21*, 466–481. https://doi.org/10.1111/gtc.12352

Ivshina, M., Lasko, P., & Richter, J. D. (2014). Cytoplasmic Polyadenylation Element Binding Proteins in Development, Health, and Disease. *Annual Review of Cell and Developmental Biology*, *30*, 393–415. https://doi.org/10.1146/annurev-cellbio-101011-155831

Jeong, J.-H., Nam, Y.-J., Kim, S.-Y., Kim, E.-G., Jeong, J., & Kim, H. K. (2007). The transport of Staufen2-containing ribonucleoprotein complexes involves kinesin motor protein and is modulated by mitogen-activated protein kinase pathway. *Journal of Neurochemistry*, *102*, 2073–2084. https://doi.org/10.1111/j.1471-4159.2007.04697.x

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kapucu, F. E., Vinogradov, A., Hyvärinen, T., Ylä-Outinen, L., & Narkilahti, S. (2022). Comparative microelectrode array data of the functional development of hPSC-derived and rat neuronal networks. *Scientific Data*, *9*, 120. https://doi.org/10.1038/s41597-022-01242-4

Kar, A. N., Lee, S. J., & Twiss, J. L. (2018). Expanding Axonal Transcriptome Brings New Functions for Axonally Synthesized Proteins in Health and Disease. *The Neuroscientist*, *24*, 111–129. https://doi.org/10.1177/1073858417712668

Kawahara, Y., & Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences*, *109*, 3347–3352. https://doi.org/10.1073/pnas.1112427109

Kim, H. H., Lee, S. J., Gardiner, A. S., Perrone-Bizzozero, N. I., & Yoo, S. (2015). Different motif requirements for the localization zipcode element of β-actin mRNA binding by HuD and ZBP1. *Nucleic Acids Research*, *43*, 7432–7446. https://doi.org/10.1093/nar/gkv699

Kislauskis, E. H., Zhu, X., & Singer, R. H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *The Journal of cell biology*, *127*, 441–451. https://doi.org/10.1083/jcb.127.2.441

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, *9*, 72–74. https://doi.org/10.1038/nmeth.1778

Kobayashi, H., Yamamoto, S., Maruo, T., & Murakami, F. (2005). Identification of a *cis* -acting element required for dendritic targeting of activity-regulated cytoskeleton-associated protein mRNA. *European Journal of Neuroscience*, *22*, 2977–2984. https://doi.org/10.1111/j.1460-9568.2005.04508.x

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., & Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, *9*. https://doi.org/10.12688/f1000research.24956.2

Kosturko, L. D., Maggipinto, M. J., Korza, G., Lee, J. W., Carson, J. H., & Barbarese, E. (2006). Heterogeneous Nuclear Ribonucleoprotein (hnRNP) E1 Binds to hnRNP A2 and Inhibits Translation of A2 Response Element mRNAs. *Molecular Biology of the Cell*, *17*, 3521–3533. https://doi.org/10.1091/mbc.e05-10-0946

Kraut-Cohen, J., Afanasieva, E., Haim-Vilmovsky, L., Slobodin, B., Yosef, I., Bibi, E., & Gerst, J. E. (2013). Translation- and SRP-independent mRNA targeting to the endoplasmic reticulum in the yeast *Saccharomyces cerevisiae*. *Molecular Biology of the Cell*, *24*, 3069–3084. https://doi.org/10.1091/mbc.e13-01-0038

Kwiatkowski, T. J., Bosco, D. A., LeClerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., de Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., … Brown, R. H. (2009). Mutations in the *FUS/TLS* Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science*, *323*, 1205–1208. https://doi.org/10.1126/science.1166066

Lebeau, G., Maher-Laporte, M., Topolnik, L., Laurent, C. E., Sossin, W., DesGroseillers, L., & Lacaille, J.-C. (2008). Staufen1 Regulation of Protein Synthesis-Dependent Long-Term Potentiation and

Synaptic Function in Hippocampal Pyramidal Cells. *Molecular and Cellular Biology*, *28*, 2896–2907. https://doi.org/10.1128/MCB.01844-07

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., & Krause, H. M. (2007). Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell*, *131*, 174–187. https://doi.org/10.1016/j.cell.2007.08.003

Lederer, C. W., Torrisi, A., Pantelidou, M., Santama, N., & Cavallaro, S. (2007). Pathways and genes differentially expressed in the motor cortex of patients with sporadic amyotrophic lateral sclerosis. *BMC Genomics*, *8*, 26. https://doi.org/10.1186/1471-2164-8-26

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, *11*, 733–739. https://doi.org/10.1038/nrg2825

Leek, J. T., & Storey, J. D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genetics*, *3*, e161. https://doi.org/10.1371/journal.pgen.0030161

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T.-M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*, 168–176. https://doi.org/10.1038/nature05453

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M. C., & Stoecklin, G. (2013). Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. *Cell*, *153*, 869–881. https://doi.org/10.1016/j.cell.2013.04.016

Leung, K.-M., van Horck, F. P., Lin, A. C., Allison, R., Standart, N., & Holt, C. E. (2006). Asymmetrical β-actin mRNA translation in growth cones mediates attractive turning to netrin-1. *Nature Neuroscience*, *9*, 1247–1256. https://doi.org/10.1038/nn1775

Levoye, A., Dam, J., Ayoub, M. A., Guillaume, J.-L., Couturier, C., Delagrange, P., & Jockers, R. (2006). The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. *The EMBO Journal*, *25*, 3012–3023. https://doi.org/10.1038/sj.emboj.7601193

Li, P., Yang, X., Wasser, M., Cai, Y., & Chia, W. (1997). Inscuteable and Staufen Mediate Asymmetric Localization and Segregation of prospero RNA during Drosophila Neuroblast Cell Divisions. *Cell*, *90*, 437–447. https://doi.org/10.1016/S0092-8674(00)80504-8

Li, Z., Zhang, Y., Ku, L., Wilkinson, K. D., Warren, S. T., & Feng, Y. (2001). The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Research*, *29*, 2276–2283. https://doi.org/10.1093/nar/29.11.2276

Liao, Y.-C., Fernandopulle, M. S., Wang, G., Choi, H., Hao, L., Drerup, C. M., Patel, R., Qamar, S., Nixon-Abell, J., Shen, Y., Meadows, W., Vendruscolo, M., Knowles, T. P., Nelson, M., Czekalska, M. A., Musteikyte, G., Gachechiladze, M. A., Stephens, C. A., Pasolli, H. A., … Ward, M. E. (2019). RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether. *Cell*, *179*, 147–164.e20. https://doi.org/10.1016/j.cell.2019.08.050

López-Erauskin, J., Tadokoro, T., Baughn, M. W., Myers, B., McAlonis-Downes, M., Chillon-Marinas, C., Asiaban, J. N., Artates, J., Bui, A. T., Vetto, A. P., Lee, S. K., Le, A. V., Sun, Y., Jambeau, M., Boubaker, J., Swing, D., Qiu, J., Hicks, G. G., Ouyang, Z., … Cruz, S. D. (2018). ALS/FTD-Linked Mutation in FUS Suppresses Intra-axonal Protein Synthesis and Drives Disease Without Nuclear Loss-of-Function of FUS. *Neuron*, *100*, 816–830.e7. https://doi.org/10.1016/j.neuron.2018.09.044

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*, 550.

Luarte, A., Cornejo, V. H., Bertin, F., Gallardo, J., & Couve, A. (2018). The axonal endoplasmic reticulum: One organelle-many functions in development, maintenance, and plasticity. *Developmental Neurobiology*, *78*, 181–208. https://doi.org/10.1002/dneu.22560

Lubelsky, Y., & Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*. https://doi.org/10.1038/nature25757

Ludwik, K., von Kuegelgen, N., & Chekulaeva, M. (2019). Genome-wide analysis of RNA and protein localization and local translation in mESC-derived neurons. *Methods*, *162-163*. https://doi.org/10.1016/j.ymeth.2019.02.002

Lyabin, D. N., Eliseeva, I. A., & Ovchinnikov, L. P. (2014). YB-1 protein: functions and regulation. *Wiley Interdisciplinary Reviews: RNA*, *5*, 95–110. https://doi.org/10.1002/wrna.1200

Maciel, R., Bis, D. M., Rebelo, A. P., Saghira, C., Züchner, S., & Saporta, M. A. (2018). The human motor neuron axonal transcriptome is enriched for transcripts related to mitochondrial function and microtubule-based axonal transport. *Experimental neurology*, *307*, 155–163. https://doi.org/10.1016/j.expneurol.2018.06.008

Maltby, C. J., Schofield, J. P. R., Houghton, S. D., O'Kelly, I., Vargas-Caballero, M., Deinhardt, K., & Coldwell, M. J. (2020). A 5′ UTR GGN repeat controls localisation and translation of a potassium leak channel mRNA through G-quadruplex formation. *Nucleic Acids Research*, *48*, 9822–9839. https://doi.org/10.1093/nar/gkaa699

Martin, K. C., Casadio, A., Zhu, H., E, Y., Rose, J. C., Chen, M., Bailey, C. H., & Kandel, E. R. (1997). Synapse-Specific, Long-Term Facilitation of Aplysia Sensory to Motor Synapses: A Function for Local Protein Synthesis in Memory Storage. *Cell*, *91*, 927–938. https://doi.org/10.1016/S0092-8674(00)80484-5

Maselli, R. A., Arredondo, J., Vázquez, J., Chong, J. X., Bamshad, M. J., Nickerson, D. A., Lara, M., Ng, F., Lo, V. L., Pytel, P., & McDonald, C. M. (2018). A presynaptic congenital myasthenic syndrome attributed to a homozygous sequence variant in *LAMA5*. *Annals of the New York Academy of Sciences*, *1413*, 119–125. https://doi.org/10.1111/nyas.13585

Mazzoni, E. O., Mahony, S., Closser, M., Morrison, C. A., Nedelec, S., Williams, D. J., An, D., Gifford, D. K., & Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nature neuroscience*, *16*, 1219.

McCaddon, A., Regland, B., Hudson, P., & Davies, G. (2002). Functional vitamin B12 deficiency and Alzheimer disease. *Neurology*, *58*, 1395–1399. https://doi.org/10.1212/WNL.58.9.1395

Medioni, C., Mowry, K., & Besse, F. (2012). Principles and roles of mRNA localization in animal development. *Development (Cambridge)*. https://doi.org/10.1242/dev.078626

Menon, K. P., Sanyal, S., Habara, Y., Sanchez, R., Wharton, R. P., Ramaswami, M., & Zinn, K. (2004). The Translational Repressor Pumilio Regulates Presynaptic Morphology and Controls Postsynaptic Accumulation of Translation Factor eIF-4E. *Neuron*, *44*, 663–676. https://doi.org/10.1016/j.neuron.2004.10.028

Merianda, T. T., Lin, A. C., Lam, J. S., Vuppalanchi, D., Willis, D. E., Karin, N., Holt, C. E., & Twiss, J. L. (2009). A functional equivalent of endoplasmic reticulum and Golgi in axons for secretion of locally synthesized proteins. *Molecular and Cellular Neuroscience*, *40*, 128–142. https://doi.org/10.1016/j.mcn.2008.09.008

Mertins, P., Tang, L. C., Krug, K., Clark, D. J., Gritsenko, M. A., Chen, L., Clauser, K. R., Clauss, T. R., Shah, P., Gillette, M. A., Petyuk, V. A., Thomas, S. N., Mani, D. R., Mundt, F., Moore, R. J., Hu, Y., Zhao, R., Schnaubelt, M., Keshishian, H., … Carr, S. A. (2018). Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nature Protocols*, *13*, 1632–1661. https://doi.org/10.1038/s41596-018-0006-9

Middleton, S. A., Eberwine, J., & Kim, J. (2019). Comprehensive catalog of dendritically localized mRNA isoforms from sub-cellular sequencing of single mouse neurons. *BMC biology*, *17*, 5. https://doi.org/10.1186/s12915-019-0630-z

Mikl, M., Eletto, D., Lee, M., Lafzi, A., Mhamedi, F., Sain, S. B., Handler, K., & Moor, A. E. (2021). A massively parallel reporter assay reveals focused and broadly encoded RNA localization signals in neurons. *bioRxiv*.

Miller, S., Yasuda, M., Coats, J. K., Jones, Y., Martone, M. E., & Mayford, M. (2002). Disruption of dendritic translation of CaMKIIα impairs stabilization of synaptic plasticity and memory consolidation. *Neuron*, *36*, 507–519. https://doi.org/10.1016/S0896-6273(02)00978-9

Minis, A., Dahary, D., Manor, O., Leshkowitz, D., Pilpel, Y., & Yaron, A. (2014). Subcellular transcriptomics—Dissection of the mRNA composition in the axonal compartment of sensory neurons. *Developmental neurobiology*, *74*, 365–381. https://doi.org/10.1002/dneu.22140

Mitchell, J., Morris, A., & de Belleroche, J. (2009). Thioredoxin reductase 1 haplotypes modify familial amyotrophic lateral sclerosis onset. *Free Radical Biology and Medicine*, *46*, 202–211. https://doi.org/10.1016/j.freeradbiomed.2008.09.041

Moccia, R., Chen, D., Lyles, V., Kapuya, E., Yaping, E., Kalachikov, S., Spahn, C. M., Frank, J., Kandel, E. R., Barad, M., & Martin, K. C. (2003). An Unbiased cDNA Library Prepared from Isolated Aplysia Sensory Neuron Processes Is Enriched for Cytoskeletal and Translational mRNAs. *Journal of Neuroscience*, *23*, 9409–9417. https://doi.org/10.1523/jneurosci.23-28-09409.2003

Mofatteh, M. (2020). mRNA localization and local translation in neurons. *AIMS Neuroscience*, *7*, 299–310. https://doi.org/10.3934/Neuroscience.2020016

Mofatteh, M. (2021). Neurodegeneration and axonal mRNA transportation. *American journal of neurodegenerative disease*, *10*, 1–12.

Mori, Y., Imaizumi, K., Katayama, T., Yoneda, T., & Tohyama, M. (2000). Two cis-acting elements in the 3' untranslated region of α-CaMKII regulate its dendritic targeting. *Nature Neuroscience*, *3*, 1079–1084. https://doi.org/10.1038/80591

Most, D., Ferguson, L., Blednov, Y., Mayfield, R. D., & Harris, R. A. (2015). The synaptoneurosome transcriptome: a model for profiling the emolecular effects of alcohol. *The Pharmacogenomics Journal*, *15*, 177–188. https://doi.org/10.1038/tpj.2014.43

Muddashetty, R. S., Nalavadi, V. C., Gross, C., Yao, X., Xing, L., Laur, O., Warren, S. T., & Bassell, G. J. (2011). Reversible Inhibition of PSD-95 mRNA Translation by miR-125a, FMRP Phosphorylation, and mGluR Signaling. *Molecular Cell*, *42*, 673–688. https://doi.org/10.1016/j.molcel.2011.05.006

Muslimov, I. A., Iacoangeli, A., Brosius, J., & Tiedge, H. (2006). Spatial codes in dendritic BC1 RNA. *Journal of Cell Biology*, *175*, 427–439. https://doi.org/10.1083/jcb.200607008

Muslimov, I. A., Tuzhilin, A., Tang, T. H., Wong, R. K., Bianchi, R., & Tiedge, H. (2014). Interactions of noncanonical motifs with hnRNP A2 promote activity-dependent RNA transport in neurons. *Journal of Cell Biology*, *205*, 493–510. https://doi.org/10.1083/jcb.201310045

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., & Amster-Choder, O. (2011). Translation-Independent Localization of mRNA in *E. coli*. *Science*, *331*, 1081–1084. https://doi.org/10.1126/science.1195691

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*, *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2

Nijssen, J., Aguila, J., Hoogstraaten, R., Kee, N., & Hedlund, E. (2018). Axon-Seq Decodes the Motor Axon Transcriptome and Its Modulation in Response to ALS. *Stem Cell Reports*. https://doi.org/10.1016/j.stemcr.2018.11.005

Nishimune, H., Valdez, G., Jarad, G., Moulson, C. L., Müller, U., Miner, J. H., & Sanes, J. R. (2008). Laminins promote postsynaptic maturation by an autocrine mechanism at the neuromuscular junction. *Journal of Cell Biology*, *182*, 1201–1215. https://doi.org/10.1083/jcb.200805095

Notaras, M., Allen, M., Longo, F., Volk, N., Toth, M., Jeon, N. L., Klann, E., & Colak, D. (2020). UPF2 leads to degradation of dendritically targeted mRNAs to regulate synaptic plasticity and cognitive function. *Molecular Psychiatry*, *25*, 3360–3379. https://doi.org/10.1038/s41380-019-0547-5

Oe, S., & Yoneda, Y. (2010). Cytoplasmic polyadenylation element-like sequences are involved in dendritic targeting of BDNF mRNA in hippocampal neurons. *FEBS Letters*, *584*, 3424–3430. https://doi.org/10.1016/j.febslet.2010.06.040

Ohno, K., Kato, H., Funahashi, S., Hasegawa, T., & Sato, K. (2009). Characterization of CLP36/Elfin/PDLIM1 in the nervous system. *Journal of Neurochemistry*, *111*, 790–800. https://doi.org/10.1111/j.1471-4159.2009.06370.x

Oikonomou, P., Goodarzi, H., & Tavazoie, S. (2014). Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. *Cell Reports*. https://doi.org/10.1016/j.celrep.2014.03.001

Ouwenga, R., Lake, A. M., O'Brien, D., Mogha, A., Dani, A., Dougherty, J. D., O'Brien, D., Mogha, A., Dani, A., & Dougherty, J. D. (2017). Transcriptomic analysis of ribosome-bound mRNA in cortical neurites in vivo. *Journal of Neuroscience*, *37*, 8688–8705. https://doi.org/10.1523/JNEUROSCI.3044-16.2017

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*, 417–419. https://doi.org/10.1038/nmeth.4197

Peixoto, L., Risso, D., Poplawski, S. G., Wimmer, M. E., Speed, T. P., Wood, M. A., & Abel, T. (2015). How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic Acids Research*, *43*, 7664–7674. https://doi.org/10.1093/nar/gkv736

Pimentel, J., & Boccaccio, G. L. (2014). Translation and silencing in RNA granules: a tale of sand grains. *Frontiers in Molecular Neuroscience*, *7*. https://doi.org/10.3389/fnmol.2014.00068

Pisareva, V. P., Skabkin, M. A., Hellen, C. U. T., Pestova, T. V., & Pisarev, A. V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *The EMBO Journal*, *30*, 1804–1817. https://doi.org/10.1038/emboj.2011.93

Poon, M. M., Choi, S.-H. H., Jamieson, C. A. M., Geschwind, D. H., & Martin, K. C. (2006). Identification of process-localized mRNAs from cultured rodent hippocampal neurons. *Journal of Neuroscience*, *26*, 13390–13399. https://doi.org/10.1523/JNEUROSCI.3432-06.2006

Poulopoulos, A., Murphy, A. J., Ozkan, A., Davis, P., Hatch, J., Kirchner, R., & Macklis, J. D. (2019). Subcellular transcriptomes and proteomes of developing axon projections in the cerebral cortex. *Nature*, *565*, 356. https://doi.org/10.1038/s41586-018-0847-y

Prasad, R., Chan, L. F., Hughes, C. R., Kaski, J. P., Kowalczyk, J. C., Savage, M. O., Peters, C. J., Nathwani, N., Clark, A. J. L., Storr, H. L., & Metherell, L. A. (2014). Thioredoxin Reductase 2 (TXNRD2) mutation associated with familial glucocorticoid deficiency (FGD). *The Journal of clinical endocrinology and metabolism*, *99*, E1556–63. https://doi.org/10.1210/jc.2013-3844

Rabani, M., Pieper, L., Chew, G. L., & Schier, A. F. (2017). A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.11.014

Rabin, S. J., Kim, J. M. H., Baughn, M., Libby, R. T., Kim, Y. J., Fan, Y., Libby, R. T., Spada, A. L., Stone, B., & Ravits, J. (2010). Sporadic ALS has compartment-specific aberrant exon splicing and altered cell–matrix adhesion biology. *Human Molecular Genetics*, *19*, 313–328. https://doi.org/10.1093/hmg/ddp498

Raju, C. S., Fukuda, N., López-Iglesias, C., Göritz, C., Visa, N., & Percipalle, P. (2011). In neurons, activity-dependent association of dendritically transported mRNA transcripts with the transacting factor CBF-A is mediated by A2RE/RTS elements. *Molecular Biology of the Cell*, *22*, 1864–1877. https://doi.org/10.1091/mbc.E10-11-0904

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*, 1896–1906. https://doi.org/10.1038/nprot.2007.261

Ratti, A., Fallini, C., Cova, L., Fantozzi, R., Calzarossa, C., Zennaro, E., Pascale, A., Quattrone, A., & Silani, V. (2006). A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of *Msi1* mRNA. *Journal of Cell Science*, *119*, 1442–1452. https://doi.org/10.1242/jcs.02852

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*. https://doi.org/10.1038/nature12311

Reinert, A., Morawski, M., Seeger, J., Arendt, T., & Reinert, T. (2019). Iron concentrations in neurons and glial cells with estimates on ferritin concentrations. *BMC Neuroscience*, *20*, 25. https://doi.org/10.1186/s12868-019-0507-7

Ríos, H., Paganelli, A. R., & Fosser, N. S. (2020). The role of PDLIM1, a PDZ-LIM domain protein, at the ribbon synapses in the chicken retina. *Journal of Comparative Neurology*, *528*, 1820–1832. https://doi.org/10.1002/cne.24855

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*, 896–902. https://doi.org/10.1038/nbt.2931

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*, e47–e47. https://doi.org/10.1093/nar/gkv007

Rodriguez, M. S., Dargemont, C., & Stutz, F. (2004). Nuclear export of RNA. *Biology of the Cell*, *96*, 639–655. https://doi.org/10.1016/j.biolcel.2004.04.014

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L., & Singer, R. H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. *Molecular and cellular biology*, *17*, 2158–2165. https://doi.org/10.1128/mcb.17.4.2158

Rotem, N., Magen, I., Ionescu, A., Gershoni-Emek, N., Altman, T., Costa, C. J., Gradus, T., Pasmanik-Chor, M., Willis, D. E., Ben-Dov, I. Z., Hornstein, E., & Perlson, E. (2017). ALS Along the Axons – Expression of Coding and Noncoding RNA Differs in Axons of ALS models. *Scientific Reports*, *7*, 44500. https://doi.org/10.1038/srep44500

Saal, L., Briese, M., Kneitz, S., Glinka, M., & Sendtner, M. (2014). Subcellular transcriptome alterations in a cell culture model of spinal muscular atrophy point to widespread defects in axonal growth and presynaptic differentiation. *RNA*. https://doi.org/10.1261/rna.047373.114

Salauze, L., van der Velden, C., Lagroye, I., Veyret, B., & Geffard, M. (2005). Circulating antibodies to cysteinyl catecholamines in amyotrophic lateral sclerosis and Parkinson's disease patients. *Amyotrophic Lateral Sclerosis*, *6*, 226–233. https://doi.org/10.1080/14660820510044469

Schrauwen, I., Liaqat, K., Schatteman, I., Bharadwaj, T., Nasir, A., Acharya, A., Ahmad, W., Camp, G. V., & Leal, S. M. (2020). Autosomal Dominantly Inherited GREB1L Variants in Individuals with Profound Sensorineural Hearing Impairment. *Genes*, *11*. https://doi.org/10.3390/genes11060687

Schütz, B. (2005). Imbalanced excitatory to inhibitory synaptic input precedes motor neuron degeneration in an animal model of amyotrophic lateral sclerosis. *Neurobiology of Disease*, *20*, 131–140. https://doi.org/10.1016/j.nbd.2005.02.006

Shan, J., Munro, T. P., Barbarese, E., Carson, J. H., & Smith, R. (2003). A molecular mechanism for mRNA trafficking in neuronal dendrites. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.23-26-08859.2003

Shigeoka, T., Jung, H., Jung, J., Turner-Bridger, B., Ohk, J., Lin, J. Q., Amieux, P. S., & Holt, C. E. (2016). Dynamic axonal translation in developing and mature visual circuits. *Cell*, *166*, 181–192. https://doi.org/10.1016/j.cell.2016.05.029

Shukla, C. J., McCorkindale, A. L., Gerhardinger, C., Korthauer, K. D., Cabili, M. N., Shechner, D. M., Irizarry, R. A., Maass, P. G., & Rinn, J. L. (2018). High-throughput identification of RNA nuclear enrichment sequences. *The EMBO Journal*. https://doi.org/10.15252/embj.201798452

Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome research*, *27*, 491–499. https://doi.org/10.1101/gr.209601.116

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.1

Song, T., Zheng, Y., Wang, Y., Katz, Z., Liu, X., Chen, S., Singer, R. H., & Gu, W. (2015). Specific interaction of KIF11 with ZBP1 regulates the transport of β-actin mRNA and cell motility. *Journal of Cell Science*. https://doi.org/10.1242/jcs.161679

Srivastava, M., Duan, G., Kershaw, N. J., Athanasopoulos, V., Yeo, J. H. C., Ose, T., Hu, D., Brown, S. H. J., Jergic, S., Patel, H. R., Pratama, A., Richards, S., Verma, A., Jones, E. Y., Heissmeyer, V., Preiss, T., Dixon, N. E., Chong, M. M. W., Babon, J. J., & Vinuesa, C. G. (2015). Roquin binds

microRNA-146a and Argonaute2 to regulate microRNA homeostasis. *Nature Communications*, *6*, 6253. https://doi.org/10.1038/ncomms7253

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*, 631–656. https://doi.org/10.1038/s41576-019-0150-2

Stepien, B. K., Oppitz, C., Gerlach, D., Dag, U., Novatchkova, M., Krüttner, S., Stark, A., & Keleman, K. (2016). RNA-binding profiles of {Drosophila} CPEB proteins Orb and Orb2. *Proceedings of the National Academy of Sciences*, *113*. https://doi.org/10.1073/pnas.1603715113

Steward, O., & Levy, W. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. *The Journal of Neuroscience*, *2*, 284–291. https://doi.org/10.1523/JNEUROSCI.02-03-00284.1982

Steward, O. (1997). mRNA Localization in Neurons: A Multipurpose Mechanism? *Neuron*, *18*, 9–12. https://doi.org/10.1016/S0896-6273(01)80041-6

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*. https://doi.org/10.1016/j.cell.2019.05.031

Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J. L., & Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Reports*, *12*, 697–704. https://doi.org/10.1038/embor.2011.76

Sun, Z., Diaz, Z., Fang, X., Hart, M. P., Chesi, A., Shorter, J., & Gitler, A. D. (2011). Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity for the ALS Disease Protein FUS/TLS. *PLoS Biology*, *9*, e1000614. https://doi.org/10.1371/journal.pbio.1000614

Suzuki, N., Akiyama, T., Warita, H., & Aoki, M. (2020). Omics Approach to Axonal Dysfunction of Motor Neurons in Amyotrophic Lateral Sclerosis (ALS). *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00194

Taguchi, Y.-h., Iwadate, M., & Umeyama, H. (2015). Heuristic principal component analysis-based unsupervised feature extraction and its application to gene expression analysis of amyotrophic lateral sclerosis data sets. *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 1–10. https://doi.org/10.1109/CIBCB.2015.7300274

Taliaferro, J. M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., Saxena, T., Wang, E. T., Graveley, B. R., Gertler, F. B., Swanson, M. S., & Burge, C. B. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell*, *61*, 821–833. https://doi.org/10.1016/j.molcel.2016.01.020

Tan, D., Zhou, M., Kiledjian, M., & Tong, L. (2014). The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural & Molecular Biology*, *21*, 679–685. https://doi.org/10.1038/nsmb.2857

Tasdemir-Yilmaz, O. E., & Segal, R. A. (2016). There and back again: Coordinated transcription, translation and transport in axonal survival and regeneration. *Current Opinion in Neurobiology*, *39*, 62–68. https://doi.org/10.1016/j.conb.2016.04.006

Taylor, A. M., Berchtold, N. C., Perreau, V. M., Tu, C. H., Jeon, N. L., & Cotman, C. W. (2009). Axonal mRNA in uninjured and regenerating cortical mammalian axons. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.6130-08.2009

Taylor, J. P., Brown, R. H., & Cleveland, D. W. (2016). Decoding ALS: from genes to mechanism. *Nature*, *539*, 197–206. https://doi.org/10.1038/nature20413

Thelen, M. P., & Kye, M. J. (2020). The Role of RNA Binding Proteins for Local mRNA Translation: Implications in Neurological Disorders. *Frontiers in Molecular Biosciences*, *6*. https://doi.org/10.3389/fmolb.2019.00161

Thomas, M. G., Tosar, L. J. M., Desbats, M. A., Leishman, C. C., & Boccaccio, G. L. (2009). Mammalian Staufen 1 is recruited to stress granules and impairs their assembly. *Journal of Cell Science*, *122*, 563–573. https://doi.org/10.1242/jcs.038208

Tiruchinapalli, D. M., Oleynikov, Y., Kelič, S., Shenoy, S. M., Hartley, A., Stanton, P. K., Singer, R. H., & Bassell, G. J. (2003). Activity-Dependent Trafficking and Dynamic Localization of Zipcode

Binding Protein 1 and β-Actin mRNA in Dendrites and Spines of Hippocampal Neurons. *The Journal of Neuroscience*, *23*, 3251–3261. https://doi.org/10.1523/JNEUROSCI.23-08-03251.2003

Tobias, G., & Koenig, E. (1975). Axonal protein synthesizing activity during the early outgrowth period following neurotomy. *Experimental Neurology*, *49*, 221–234. https://doi.org/10.1016/0014-4886(75)90206-X

Tongiorgi, E., Righi, M., & Cattaneo, A. (1997). Activity-Dependent Dendritic Targeting of BDNF and TrkB mRNAs in Hippocampal Neurons. *The Journal of Neuroscience*, *17*, 9492–9505. https://doi.org/10.1523/JNEUROSCI.17-24-09492.1997

Tóth, E. N., Lohith, A., Mondal, M., Guo, J., Fukamizu, A., & Pourmand, N. (2018). Single-cell nanobiopsy reveals compartmentalization of mRNAs within neuronal cells. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M117.800763

Trcek, T., & Lehmann, R. (2019). Germ granules in *Drosophila*. *Traffic*, *20*, 650–660. https://doi.org/10.1111/tra.12674

Turner-Bridger, B., Caterino, C., & Cioni, J.-M. (2020). Molecular mechanisms behind mRNA localization in axons. *Open Biology*, *10*, 200177. https://doi.org/10.1098/rsob.200177

Tushev, G., Glock, C., Heumüller, M., Biever, A., Jovanovic, M., & Schuman, E. M. (2018). Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron*, *98*, 495–511. https://doi.org/10.1016/j.neuron.2018.03.030

Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*, 2301–2319. https://doi.org/10.1038/nprot.2016.136

Vessey, J. P., Amadei, G., Burns, S. E., Kiebler, M. A., Kaplan, D. R., & Miller, F. D. (2012). An Asymmetrically Localized Staufen2-Dependent RNA Complex Regulates Maintenance of Mammalian Neural Stem Cells. *Cell Stem Cell*, *11*, 517–528. https://doi.org/10.1016/j.stem.2012.06.010

von Kügelgen, N., & Chekulaeva, M. (2020). Conservation of a core neurite transcriptome across neuronal types and species. *Wiley Interdisciplinary Reviews: RNA*. https://doi.org/10.1002/wrna.1590

von Kügelgen, N., Mendonsa, S., Dantsuji, S., Ron, M., Kirchner, M., Zerna, N., Bujanic, L., Mertins, P., Ulitsky, I., & Chekulaeva, M. (2021). Massively parallel identification of zipcodes in primary cortical neurons. *bioRxiv*. https://doi.org/10.1101/2021.10.21.465275

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, *131*, 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wagner, G. P., Kin, K., & Lynch, V. J. (2013). A model based criterion for gene expression calls using RNA-seq data. *Theory in Biosciences*, *132*, 159–164. https://doi.org/10.1007/s12064-013-0178-3

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63. https://doi.org/10.1038/nrg2484

Wells, D. G. (2006). RNA-Binding Proteins: A Lesson in Repression. *Journal of Neuroscience*, *26*, 7135–7138. https://doi.org/10.1523/JNEUROSCI.1795-06.2006

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., & Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*, 1239–1243. https://doi.org/10.1038/nature07002

Williams, T., Ngo, L. H., & Wickramasinghe, V. O. (2018). Nuclear export of RNA: Different sizes, shapes and functions. *Seminars in Cell & Developmental Biology*, *75*, 70–77. https://doi.org/10.1016/j.semcdb.2017.08.054

Wilson, M. P., Plecko, B., Mills, P. B., & Clayton, P. T. (2019). Disorders affecting vitamin B_6 metabolism. *Journal of Inherited Metabolic Disease*, *42*, 629–646. https://doi.org/10.1002/jimd.12060

Wissink, E. M., Fogarty, E. A., & Grimson, A. (2016). High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics*, *17*, 177. https://doi.org/10.1186/s12864-016-2479-7

Wootz, H., Enjin, A., Wallén-Mackenzie, Å., Lindholm, D., & Kullander, K. (2010). Reduced VGLUT2 expression increases motor neuron viability in Sod1G93A mice. *Neurobiology of Disease*, *37*, 58–66. https://doi.org/10.1016/j.nbd.2009.09.006

Wu, D. C., Yao, J., Ho, K. S., Lambowitz, A. M., & Wilke, C. O. (2018). Limitations of alignment-free tools in total RNA-seq quantification. *BMC Genomics*, *19*, 510. https://doi.org/10.1186/s12864-018-4869-5

Wurmus, R., Uyar, B., Osberg, B., Franke, V., Gosdschan, A., Wreczycka, K., Ronen, J., & Akalin, A. (2018). PiGx: reproducible genomics analysis pipelines with GNU Guix. *GigaScience*, *7*, giy123. https://doi.org/10.1093/gigascience/giy123

Yartseva, V., Takacs, C. M., Vejnar, C. E., Lee, M. T., & Giraldez, A. J. (2017). RESA identifies mRNA-regulatory sequences at high resolution. *Nature Methods*. https://doi.org/10.1038/nmeth.4121

Yasuda, K., Zhang, H., Loiselle, D., Haystead, T., Macara, I. G., & Mili, S. (2013). The RNA-binding protein Fus directs translation of localized mRNAs in APC-RNP granules. *Journal of Cell Biology*, *203*, 737–746. https://doi.org/10.1083/jcb.201306058

Yoo, S., Kim, H. H., Kim, P., Donnelly, C. J., Kalinski, A. L., Vuppalanchi, D., Park, M., Lee, S. J., Merianda, T. T., Perrone-Bizzozero, N. I., & Twiss, J. L. (2013). A HuD-ZBP1 ribonucleoprotein complex localizes GAP-43 mRNA into axons through its 3' untranslated region AU-rich regulatory element. *Journal of Neurochemistry*, *126*, 792–804. https://doi.org/10.1111/jnc.12266

Zahr, S. K., Yang, G., Kazan, H., Borrett, M. J., Yuzwa, S. A., Voronova, A., Kaplan, D. R., & Miller, F. D. (2018). A Translational Repression Complex in Developing Mammalian Neural Stem Cells that Regulates Neuronal Specification. *Neuron*, *97*, 520–537.e6. https://doi.org/10.1016/j.neuron.2017.12.045

Zappulo, A., van den Bruck, D., Ciolli Mattioli, C., Franke, V., Imami, K., McShane, E., Moreno-Estelles, M., Calviello, L., Filipchyk, A., Peguero-Sanchez, E., Müller, T., Woehler, A., Birchmeier, C., Merino, E., Rajewsky, N., Ohler, U., Mazzoni, E. O., Selbach, M., Akalin, A., & Chekulaeva, M. (2017). RNA localization is a key determinant of neurite-enriched proteome. *Nature Communications*, *8*, 583. https://doi.org/10.1038/s41467-017-00690-6

Zhang, Q., Fan, L., Hou, F., Dong, A., Wang, Y.-X., & Tong, Y. (2015). New Insights into the RNA-Binding and E3 Ubiquitin Ligase Activities of Roquins. *Scientific Reports*, *5*, 15660. https://doi.org/10.1038/srep15660

Zhang, X., Smits, A. H., van Tilburg, G. B., Ovaa, H., Huber, W., & Vermeulen, M. (2018). Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature Protocols*, *13*, 530–550. https://doi.org/10.1038/nprot.2017.147

Zhang, Y., Gaetano, C. M., Williams, K. R., Bassell, G. J., & Mihailescu, M. R. (2014). FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA. *RNA Biology*, *11*, 1364–1374. https://doi.org/10.1080/15476286.2014.996464

Zhao, W., Pollack, J. L., Blagev, D. P., Zaitlen, N., McManus, M. T., & Erle, D. J. (2014). Massively parallel functional annotation of 3' untranslated regions. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2851

Zhong, J., Zhang, T., & Bloch, L. M. (2006). Dendritic mRNAs encode diversified functionalities in hippocampal pyramidal neurons. *BMC Neuroscience*. https://doi.org/10.1186/1471-2202-7-17

7 Publications

Ludwik, K. A., **von Kuegelgen**, **N.**, & Chekulaeva, M. (2019). Genome-wide analysis of RNA and protein localization and local translation in mESC-derived neurons. *Methods*, *162*, 31–41.

von Kügelgen, **N.**, & Chekulaeva, M. (2020). Conservation of a core neurite transcriptome across neuronal types and species. *Wiley Interdisciplinary Reviews: RNA*, *11*(4), e1590.

von Kügelgen, **N.**, Mendonsa, S., Dantsuji, S., Ron, M., Kirchner, M., Zerna, N., Bujanic, L., Mertins, P., Ulitsky, I., & Chekulaeva, M. (2021). Massively parallel identification of zipcodes in primary cortical neurons. *bioRxiv*.

8 Supplementary Material

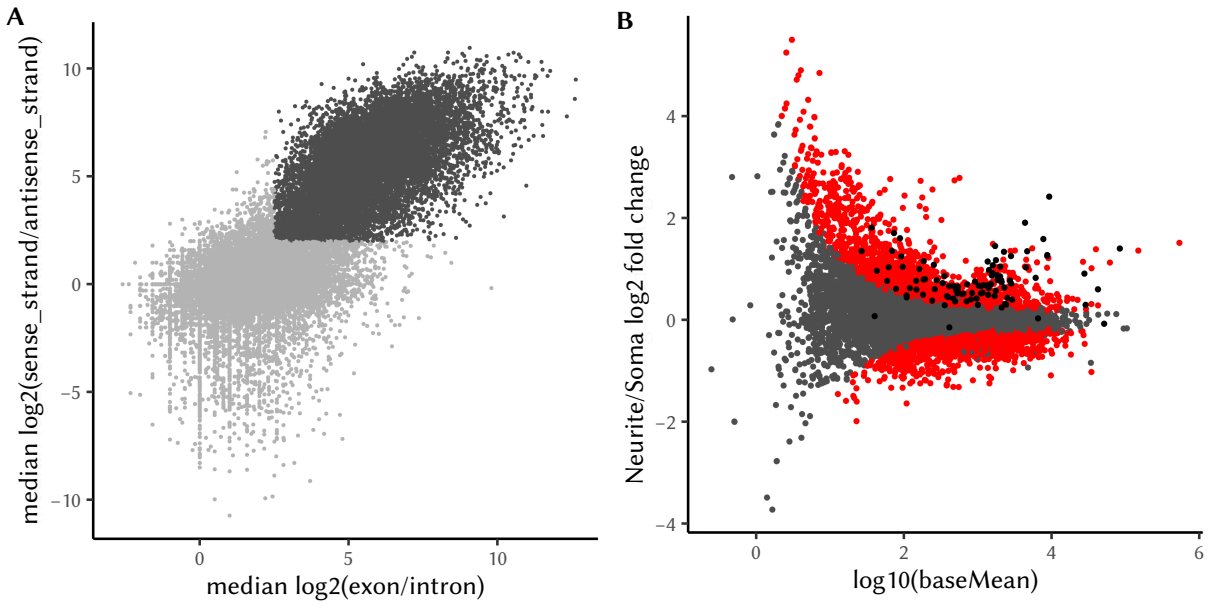


Figure S1: Quality control plots for RNA-Seq data from PCN.
(A) Log2 ratios of reads from exons and introns (x-axis) and sense and anti-sense orientation (y-axis) of all genes. Genes with log2 ratios for exon/intron >2.5 and sense/antisense >2 are marked in dark grey and are the only ones used for any analysis of this data. (B) MA plot of differential expression between neurite and soma compartment (x-axis: log10 mean expression, y-axis log2 fold change). Genes with a significant change (adj. p-value <0.05) are marked red and genes selected for inclusion in the Nzip library are marked black. Data was generated by Sayaka Dantsuji.

Table S1: Primer sequences used for generation of Nzip reporter and sequencing libraries. The 2nd strand primer for generation Nzip sequencing libraries was used as a pool of 6 primers with different lengths of UMI inserts.

Primer	Sequence
Tile amplification (fw)	cggcatggacgagctgtacaagccaTAATTCGATATCCGCATGCTAGC
Tile amplification (rev)	cttcagaaataagtttttgttccaAGACGTGTGCTCTTCCGATC
Nzip library, 1st Strand	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Nzip library, 2nd Strand, with UMI	ACACTCTTTCCCTACACGACGCTCTTCCGATCT- -HNCHNTNNNgcca- *or* -HNCHNTNNNcagcca- *or* -HNCHNTNNNagcgcca- *or* -HNCHNTNNNggctcgcca- *or* -HNCHNTNNNccagcagcca- TAATTCGATATCCGCATGCTAGC
Final library generation, fw (with Illumina adapter and barcode)	AATGATACGGCGACCACCGAGATCTACAC-nnnnnnnn- ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Final library generation, rev (with Illumina adapter and barcode)	CAAGCAGAAGACGGCATACGAGAT-nnnnnnnn- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

Table S2: All genes with significant enrichment in either soma or neurite compartment based on the combined model of all high sensivity datasets. In the last column, the identity of individual datasets, that show sig. enrichment in a given compartment, is abbreviated as follows (datasets identified by first author of the study): Briese - 1, Middleton - 2, Farris - 3, Maciel - 4, Minis - 5, Rotem - 6, Matiolli - 7, Nijssen - 8, Taliaferro (PCN) - 9, Tushev - 10, Taliaferro (CAD) - 11, Taliaferro (N2A) - 12, vonKügelgen - 13, Zappulo - 14

Ensembl gene ID	Gene name	Enrichment (log2fc) in combined model	Adjusted p-value in combined model	Average expression (mean TPM)	No. of datasets with sig. enrichment	Datasets with sig. enrichment in Soma or Neurite
ENSMUSG00000000532	Acvr1b	-1.099	0.0177	28.3	7	Soma: 1,3,5,7,8,13,14
ENSMUSG00000003153	Slc2a3	-1.332	0.0180	75.2	4	Neur: 14; Soma: 3,5,7,8
ENSMUSG00000004508	Gab2	1.315	0.0423	66.6	4	Neur: 3,5,6,7; Soma: 8
ENSMUSG00000004891	Nes	1.348	0.0218	655.7	7	Neur: 1,3,5,6,7,9,11; Soma: 14
ENSMUSG00000005362	Crbn	-1.196	0.0053	27.8	5	Soma: 2,3,5,8,14
ENSMUSG00000007892	Rplp1	0.886	<0.0001	778.2	9	Neur: 1,2,3,5,6,7,8,11,13
ENSMUSG00000011257	Pabpc4	1.296	0.0019	82.5	7	Neur: 3,5,6,7,11,13,14; Soma: 8
ENSMUSG00000013662	Atad1	-0.848	0.0325	54.2	6	Soma: 3,5,7,8,11,14
ENSMUSG00000014602	Kif1a	0.947	0.0119	445.6	8	Neur: 1,2,3,5,6,8,13; Soma: 14
ENSMUSG00000015189	Casd1	-1.428	0.0332	55.0	8	Soma: 1,2,3,5,7,8,11,14
ENSMUSG00000016427	Ndufa1	0.727	0.0117	127.4	8	Neur: 2,4,5,6,7,8,11,13
ENSMUSG00000017778	Cox7c	0.564	0.0307	348.2	9	Neur: 2,3,4,5,6,7,8,11,13
ENSMUSG00000020086	Macroh2a2	-1.235	0.0359	87.0	5	Neur: 5; Soma: 2,7,8,14
ENSMUSG00000020483	Dynll2	0.755	0.0266	261.9	5	Neur: 5,6,9,11; Soma: 7,14
ENSMUSG00000020589	Fam49a	-1.090	0.0126	27.1	5	Neur: 5; Soma: 3,7,8,14
ENSMUSG00000020607	Lratd1	-1.747	0.0297	12.5	5	Neur: 5; Soma: 2,3,7,14
ENSMUSG00000020821	Kif1c	1.485	0.0468	114.7	9	Neur: 3,4,5,6,7,9,13,14; Soma: 8
ENSMUSG00000021546	Hnrnpk	-0.665	0.0117	681.4	2	Soma: 3,8
ENSMUSG00000022661	Cd200	-1.428	0.0126	154.0	7	Soma: 1,2,3,5,7,9,14
ENSMUSG00000024661	Fth1	0.936	0.0106	944.8	10	Neur: 1,2,3,5,6,7,8,9,13,14
ENSMUSG00000024769	Cdc42bpg	1.481	0.0105	14.1	8	Neur: 3,4,5,6,7,11,13,14
ENSMUSG00000025290	Rps24	0.832	<0.0001	803.1	10	Neur: 1,2,3,4,5,6,7,8,11,13
ENSMUSG00000025508	Rplp2	0.705	0.0012	743.9	9	Neur: 1,2,3,5,6,7,8,11,13,14
ENSMUSG00000025967	Eef1b2	0.804	0.0020	613.0	9	Neur: 2,4,5,6,7,8,11,13,14
ENSMUSG00000026664	Phyh	-1.529	0.0056	35.8	4	Neur: 3,13; Soma: 2,7,14
ENSMUSG00000026739	Bmi1	-1.302	0.0350	49.0	6	Soma: 2,3,7,8,11,14
ENSMUSG00000026764	Kif5c	0.917	0.0020	403.0	5	Neur: 4,5,6,7,8,13; Soma: 14
ENSMUSG00000027498	Cstf1	-1.156	0.0225	25.7	2	Soma: 2,3
ENSMUSG00000027935	Rab13	2.322	0.0020	94.9	8	Neur: 3,4,5,6,7,11,13,14
ENSMUSG00000028367	Txn1	0.672	0.0166	238.7	6	Neur: 4,5,7,11,13,14; Soma: 3
ENSMUSG00000028389	Zfp37	-1.076	0.0352	14.4	7	Neur: 7; Soma: 2,3,5,8,11,14
ENSMUSG00000028495	Rps6	0.510	0.0323	1,591.6	7	Neur: 1,3,4,5,6,7,11,13
ENSMUSG00000028524	Sgip1	-1.019	0.0304	30.8	8	Soma: 1,2,3,5,7,8,11,14
ENSMUSG00000028639	Ybx1	0.666	0.0415	1,836.2	9	Neur: 1,2,3,4,5,6,7,8,9,13,14
ENSMUSG00000028648	Ndufs5	0.516	0.0194	221.5	6	Neur: 2,5,6,7,8,11,13
ENSMUSG00000029649	Pomp	0.597	0.0089	84.6	3	Neur: 5,6,11,13; Soma: 3
ENSMUSG00000031066	Usp11	-1.029	0.0327	86.4	5	Soma: 2,3,7,8,13,14
ENSMUSG00000031245	Hmgn5	1.238	0.0059	22.0	6	Neur: 5,6,7,11,13; Soma: 8
ENSMUSG00000031696	Vps35	-0.745	0.0018	119.8	5	Soma: 2,3,5,8,11,14
ENSMUSG00000031703	Itfg1	-0.886	0.0470	66.6	6	Soma: 1,3,5,7,8,11,14
ENSMUSG00000032232	Cgnl1	1.696	0.0001	11.5	7	Neur: 3,4,5,6,7,14; Soma: 8
ENSMUSG00000032262	Elovl4	-0.977	0.0464	24.8	6	Soma: 2,3,5,7,8,14
ENSMUSG00000032518	Rpsa	0.749	0.0046	1,226.8	11	Neur: 1,2,3,5,6,7,8,9,11,13,14
ENSMUSG00000033948	Zswim5	-0.706	0.0275	14.0	5	Soma: 3,5,7,11,14
ENSMUSG00000034586	Hid1	-1.223	0.0310	32.6	7	Soma: 1,2,3,5,7,8,13,14
ENSMUSG00000034648	Lrrn1	-1.938	0.0243	92.6	6	Soma: 3,5,6,7,8,14

Continued on next page

Table S2 – continued from previous page

Ensembl gene ID	Gene name	Enrichment	p.adj	Expr.	No. datasets	Dataset IDs
ENSMUSG00000034892	Rps29	0.867	0.0003	464.5	11	Neur: 1,2,3,4,5,6,7,8,11,13,14
ENSMUSG00000035885	Cox8a	0.594	0.0327	480.4	9	Neur: 1,2,3,5,6,7,8,11,13,14
ENSMUSG00000036052	Dnajb5	-1.466	0.0054	32.1	5	Soma: 2,3,7,8,13,14
ENSMUSG00000036291	Ap5m1	-1.390	0.0196	20.8	4	Soma: 2,5,6,8,14
ENSMUSG00000036473	Tbc1d24	-1.040	0.0179	15.0	6	Neur: 14; Soma: 1,2,3,5,7,8,13
ENSMUSG00000038900	Rpl12	0.774	0.0018	292.9	10	Neur: 1,2,3,5,6,7,8,11,13,14
ENSMUSG00000039001	Rps21	0.676	0.0148	407.7	9	Neur: 2,3,4,5,6,7,8,11,13
ENSMUSG00000039678	Tbc1d13	-1.191	0.0352	52.3	2	Soma: 5,7,8
ENSMUSG00000040952	Rps19	0.602	0.0215	884.7	10	Neur: 1,2,3,4,5,6,7,8,11,13
ENSMUSG00000041203	Trir	0.773	0.0420	66.5	1	Neur: 5,9; Soma: 3
ENSMUSG00000041219	Arhgap11a	1.250	0.0085	65.2	9	Neur: 1,4,5,6,7,9,13,14; Soma: 8
ENSMUSG00000041483	Zfp281	-1.268	0.0175	16.9	5	Soma: 3,5,6,8,11,14
ENSMUSG00000041544	Disp3	-1.252	0.0257	10.6	3	Soma: 6,7,14
ENSMUSG00000041658	Rragb	-0.989	0.0356	31.4	4	Soma: 1,3,5,14
ENSMUSG00000041697	Cox6a1	0.742	0.0066	431.3	11	Neur: 1,2,3,4,5,6,7,8,11,13,14
ENSMUSG00000041775	Mapk1ip1	-1.212	0.0497	33.2	6	Soma: 2,3,5,7,8,14
ENSMUSG00000041841	Rpl37	0.672	0.0093	391.5	9	Neur: 1,2,3,4,5,6,8,11,13
ENSMUSG00000043079	Synpo	1.996	0.0350	36.0	1	Neur: 6
ENSMUSG00000044533	Rps2	0.679	0.0132	1,263.5	12	Neur: 1,2,3,4,5,6,7,8,9,11,13,14
ENSMUSG00000045071	E130308A19Rik	-1.657	0.0217	13.9	7	Neur: 3,13; Soma: 2,7,8,11,14
ENSMUSG00000045435	Tmem60	-1.082	0.0299	22.5	6	Soma: 2,3,5,6,8,14
ENSMUSG00000048970	C1galt1c1	-1.395	0.0196	11.4	6	Soma: 2,3,5,7,8,14
ENSMUSG00000049517	Rps23	0.753	0.0132	386.5	11	Neur: 1,2,3,4,5,6,7,8,9,11,13
ENSMUSG00000050132	Sarm1	-1.086	0.0273	17.7	6	Soma: 1,2,5,7,8,14
ENSMUSG00000050708	Ftl1	0.756	0.0058	1,971.6	10	Neur: 2,3,4,5,6,7,8,9,11,13
ENSMUSG00000053617	Sh3pxd2a	2.138	0.0025	41.7	7	Neur: 1,3,4,5,6,7,9,14
ENSMUSG00000055447	Cd47	-0.989	0.0017	40.5	5	Soma: 2,3,5,7,8,14
ENSMUSG00000056459	Zbtb25	-1.445	0.0299	10.9	3	Soma: 5,6,8,14
ENSMUSG00000057244	Gm6139	0.979	0.0246	87.6	9	Neur: 1,2,3,5,7,8,9,11,14
ENSMUSG00000057322	Rpl38	0.676	0.0010	569.0	10	Neur: 1,2,3,4,5,6,7,8,11,13
ENSMUSG00000057657	Rps18-ps3	0.958	0.0144	12.2	3	Neur: 5,8,11
ENSMUSG00000057716	Tmem178b	-1.240	0.0407	26.6	6	Soma: 1,5,7,8,11,13,14
ENSMUSG00000059291	Rpl11	0.978	0.0002	891.7	8	Neur: 1,3,4,5,6,7,11,13
ENSMUSG00000060143	Gm10076	0.759	0.0013	469.7	9	Neur: 1,2,3,5,7,8,11,13,14
ENSMUSG00000060198	Gm11353	0.836	0.0243	13.8	6	Neur: 3,5,7,8,11,14
ENSMUSG00000061533	Cep128	1.702	0.0087	29.0	8	Neur: 2,3,4,5,6,7,13,14; Soma: 8
ENSMUSG00000061833	Gm6311	1.146	0.0052	101.9	8	Neur: 1,3,5,7,8,9,11,14
ENSMUSG00000061983	Rps12	0.755	0.0233	674.5	10	Neur: 1,2,3,5,6,7,8,9,11,13
ENSMUSG00000062382	Ftl1-ps1	0.645	0.0356	1,421.1	9	Neur: 2,3,4,5,7,8,9,11,13,14
ENSMUSG00000064341	mt-Nd1	1.202	0.0018	2,392.0	9	Neur: 1,2,3,4,5,7,8,13; Soma: 14
ENSMUSG00000064345	mt-Nd2	1.085	0.0154	2,001.0	8	Neur: 1,2,3,4,5,7,13; Soma: 14
ENSMUSG00000064356	mt-Atp8	1.856	<0.0001	55.5	7	Neur: 2,3,4,5,7,8; Soma: 14
ENSMUSG00000064357	mt-Atp6	1.073	0.0249	1,231.9	8	Neur: 1,2,3,4,5,7,8; Soma: 14
ENSMUSG00000064360	mt-Nd3	1.050	0.0154	399.9	7	Neur: 1,2,3,4,5,7,11
ENSMUSG00000064363	mt-Nd4	1.043	0.0126	1,352.3	9	Neur: 1,2,3,4,5,7,8,13; Soma: 14
ENSMUSG00000064367	mt-Nd5	1.332	0.0003	1,647.6	7	Neur: 1,2,3,4,5,7,8
ENSMUSG00000064368	mt-Nd6	1.463	0.0007	259.3	8	Neur: 1,2,3,4,5,7,9; Soma: 14
ENSMUSG00000065430	Mir26a-2	-19.074	0.0225	103.6	0	
ENSMUSG00000065518	Mir423	-19.330	0.0196	164.7	0	
ENSMUSG00000065519	Mir10a	-17.628	0.0467	59.4	0	
ENSMUSG00000065542	Mir224	-17.499	0.0496	99.5	0	

Continued on next page

Table S2 – continued from previous page

Ensembl gene ID	Gene name	Enrichment	p.adj	Expr.	No. datasets	Dataset IDs
ENSMUSG00000065602	Mirlet7f-2	-19.854	0.0145	717.9	1	Soma: 1
ENSMUSG00000065613	Mir92-2	-18.060	0.0370	152.5	0	
ENSMUSG00000067288	Rps28	0.805	0.0007	237.8	9	Neur: 1,2,3,5,6,7,8,11,13
ENSMUSG00000068747	Sort1	-1.187	0.0388	67.3	7	Soma: 2,3,5,6,7,8,13,14
ENSMUSG00000069682	Gm10275	0.808	0.0327	13.4	8	Neur: 2,3,5,7,8,11,13,14
ENSMUSG00000071415	Rpl23	0.573	0.0174	926.9	7	Neur: 2,3,4,5,6,7,13
ENSMUSG00000072789	Gm10420	1.067	0.0323	58.1	7	Neur: 1,3,5,7,8,9,14
ENSMUSG00000073702	Rpl31	0.771	0.0033	622.7	10	Neur: 1,2,3,5,6,7,8,9,11,13
ENSMUSG00000075918	n-R5s2	-30.000	<0.0001	13.7	0	
ENSMUSG00000081926	Gm15536	0.920	<0.0001	71.8	5	Neur: 2,3,5,7,11,14
ENSMUSG00000082762	Gm12366	0.883	0.0289	77.1	8	Neur: 2,3,5,7,8,9,11,14
ENSMUSG00000083863	Gm13341	1.150	0.0237	47.8	5	Neur: 1,2,3,5,8
ENSMUSG00000084289	Gm6977	2.197	0.0009	11.2	6	Neur: 2,3,5,8,11,13
ENSMUSG00000090516	Rps11-ps1	0.606	0.0352	19.8	6	Neur: 2,3,5,11,13,14
ENSMUSG00000093674	Rpl41	0.754	0.0006	2,431.0	10	Neur: 1,2,3,5,6,7,8,9,11,13
ENSMUSG00000095427	Rps2-ps6	1.036	0.0370	100.3	8	Neur: 1,3,5,7,8,9,11,14
ENSMUSG00000096842	Gm10736	1.136	0.0347	1,714.2	9	Neur: 1,2,3,5,7,8,11,13,14
ENSMUSG00000100131	Gm28439	1.554	0.0011	1,593.8	7	Neur: 1,2,3,5,7,8; Soma: 14
ENSMUSG00000100922	Gm8520	1.894	0.0123	11.0	3	Neur: 3,5,9,14
ENSMUSG00000101939	Gm28438	1.367	0.0049	516.1	6	Neur: 1,2,3,5,7,11
ENSMUSG00000106847	Peg13	-1.435	0.0268	27.0	6	Soma: 1,3,5,7,8,11,13
ENSMUSG00000109610	Gm7432	1.001	0.0229	26.0	8	Neur: 1,3,5,7,8,9,11,14
ENSMUSG00000112567	Gm32899	1.779	0.0030	16.8	7	Neur: 2,3,5,7,8,11,14
ENSMUSG00000118164	Gm20570	1.632	0.0033	156.2	4	Neur: 5,7,8; Soma: 14

Table S3: Genes selected for inclusion in the Nzip reporter library.
The table lists all genes selected for the Nzip reporter library, based on the analysis of neurite localisation in primary cortical neurons (von Kügelgen et al., 2021) and publicly available datasets (Briese et al., 2016; Middleton et al., 2019; Minis et al., 2014; Rotem et al., 2017; Taliaferro et al., 2016; Tushev et al., 2018). Listed are the ensembl gene IDs, gene names, number of generated tiles, total length or 3'UTR region covered and the number of datasets in which a gene showed significant enrichment in either compartment.

Ensembl gene ID	Gene name	No. of tiles	3'UTR length	Datasets with sig. enrichment in Soma or Neurite
ENSMUSG00000001289	Pfdn5	18	524	7
ENSMUSG00000004945	Tmem242	29	499	6
ENSMUSG00000006699	Cdc42 (isoform 2)	53	1407	2
ENSMUSG00000006699	Cdc42 (isoform 1)	28	768	2
ENSMUSG00000007950	Abhd8	28	478	6
ENSMUSG00000008668	Rps18	7	160	8
ENSMUSG00000008855	Hdac5	55	1437	5
ENSMUSG00000009073	Nf2	92	2378	4

Continued on next page

Table S3 – continued from previous page

Ensembl gene ID	Gene name	No. of tiles	3'UTR length	No. datasets
ENSMUSG00000013076	Amotl1	230	5826	6
ENSMUSG00000014294	Ndufa2	12	235	7
ENSMUSG00000014313	Cox6c	5	137	8
ENSMUSG00000014602	Kif1a	117	2993	8
ENSMUSG00000015222	Map2	140	3576	8
ENSMUSG00000016252	Atp5e	6	153	8
ENSMUSG00000016427	Ndufa1	5	133	8
ENSMUSG00000017778	Cox7c	4	119	9
ENSMUSG00000018593	Sparc	40	1085	8
ENSMUSG00000018819	Lsp1	23	649	3
ENSMUSG00000019820	Utrn	77	2000	9
ENSMUSG00000020022	Ndufa12	20	582	4
ENSMUSG00000020163	Uqcr11	9	189	6
ENSMUSG00000020431	Adcy1	349	8791	7
ENSMUSG00000020684	Rasl10b	96	2473	10
ENSMUSG00000020821	Kif1c	120	3085	9
ENSMUSG00000020914	Top2a	18	511	5
ENSMUSG00000021215	Net1	79	2052	6
ENSMUSG00000022354	Ndufb9	2	82	7
ENSMUSG00000022594	Lynx1	132	3379	4
ENSMUSG00000022602	Arc	64	1668	-
ENSMUSG00000022623	Shank3	75	1938	5
ENSMUSG00000024346	Pfdn1	26	721	4
ENSMUSG00000024617	Camk2a	132	3372	4
ENSMUSG00000024661	Fth1	7	156	10
ENSMUSG00000025290	Rps24	44	1173	10
ENSMUSG00000025508	Rplp2	15	277	9
ENSMUSG00000025794	Rpl14	13	260	8
ENSMUSG00000025795	Rassf3	101	2586	6
ENSMUSG00000025967	Eef1b2	32	874	9
ENSMUSG00000026031	Cflar (isoform 2)	196	4962	6
ENSMUSG00000026031	Cflar (isoform 1)	32	865	6
ENSMUSG00000026547	Tagln2	24	672	7
ENSMUSG00000026605	Cenpf	77	1993	6
ENSMUSG00000027935	Rab13	29	802	8
ENSMUSG00000028333	Anp32b	17	503	8
ENSMUSG00000028416	Bag1	9	189	7
ENSMUSG00000028639	Ybx1	24	418	9
ENSMUSG00000028998	Tomm7	32	869	7
ENSMUSG00000029580	Actb	24	683	5
ENSMUSG00000029761	Cald1	87	2245	7
ENSMUSG00000031207	Msn	74	1918	6
ENSMUSG00000031245	Hmgn5	20	575	6
ENSMUSG00000031442	Mcf2l	106	2728	8
ENSMUSG00000031818	Cox4i1	3	100	7
ENSMUSG00000032060	Cryab	6	141	6
ENSMUSG00000032330	Cox7a2	9	186	6
ENSMUSG00000032518	Rpsa	32	875	11
ENSMUSG00000034109	Golim4	115	2947	4
ENSMUSG00000034485	Uaca	4	125	6
ENSMUSG00000035203	Epn1	23	398	6
ENSMUSG00000035674	Ndufa3	12	235	7
ENSMUSG00000035885	Cox8a	12	236	9
ENSMUSG00000036751	Cox6b1	13	248	8
ENSMUSG00000037742	Eef1a1	15	285	5
ENSMUSG00000038539	Atf5	21	591	6
ENSMUSG00000038690	Atp5j2	6	141	7
ENSMUSG00000038900	Rpl12	17	500	10
ENSMUSG00000038976	Ppp1r9b	59	1541	4
ENSMUSG00000040842	Szrd1	108	2761	4

Continued on next page

Table S3 – continued from previous page

Ensembl gene ID	Gene name	No. of tiles	3'UTR length	No. datasets
ENSMUSG00000041219	Arhgap11a	112	2849	9
ENSMUSG00000041697	Cox6a1	10	201	11
ENSMUSG00000041841	Rpl37	17	504	9
ENSMUSG00000042066	Tmcc2	45	1191	8
ENSMUSG00000043279	Trim56	270	6812	-
ENSMUSG00000044894	Uqcrq	42	1134	7
ENSMUSG00000045328	Cenpe	21	372	6
ENSMUSG00000046447	Camk2n1	130	3335	5
ENSMUSG00000047721	Bola2	23	404	6
ENSMUSG00000048482	Bdnf	111	2855	-
ENSMUSG00000049517	Rps23	14	267	11
ENSMUSG00000050708	Ftl1	7	163	10
ENSMUSG00000050856	Atp5k	1	68	6
ENSMUSG00000053332	Gas5	162	4118	6
ENSMUSG00000054452	Aes	24	669	8
ENSMUSG00000059534	Uqcr10	10	203	6
ENSMUSG00000060126	Tpt1	24	424	10
ENSMUSG00000061315	Naca	5	130	7
ENSMUSG00000061518	Cox5b	21	372	7
ENSMUSG00000063015	Ccni	39	1044	7
ENSMUSG00000063882	Uqcrh	8	175	7
ENSMUSG00000067288	Rps28	10	207	9
ENSMUSG00000067847	Romo1	2	87	5
ENSMUSG00000068220	Lgals1	18	321	8
ENSMUSG00000071528	Usmg5	3	105	7
ENSMUSG00000071644	Eef1g	21	377	4
ENSMUSG00000079523	Tmsb10	14	271	7
ENSMUSG00000079641	Rpl39	8	181	8
ENSMUSG00000086841	2410006H16Rik	27	462	5
ENSMUSG00000098234	Snhg6	83	2160	5
ENSMUSG00000115783	Bc1	7	168	7

Table S4: Tiles with candidate zipcodes selected for mutational analysis.
The gene names and numbers for all tiles selected for mutational analysis based on the initial Nzip experiment (see figure 3 as well as (upper table) the neurite/soma log2 fold change (log2fc) ratios from the Nzip experiments (including neurite/soma median log2 ratio of the WT control tile from mutation Nzip experiment) and (lower table) the sequences of the selected tiles are listed.

Gene name	Tile No.	Median log2fc (WT tile, Mut. library)	log2fc (Confirmation exp.)	log2fc (initial exp., depolarisation)	log2fc (initial exp.)
Bdnf	56	1.101	1.891	0.829	0.543
Cald1	58	0.908	1.761	1.348	1.336
Camk2n1	12	3.037	2.395	1.174	1.306
Cox5b	6+7	0	2.558; 2.054	1.035; 0.709	1.426; 1.15
Golim4	56	0.717	0.843	1.04	1.272
Kif1c	80	0.603	0.692	-0.638	0.233
Mcf2l	7	1.367	1.879	1.783	2.264
Msn	48	0.559	1.348	1.579	1.254
Ndufa2	11+12	1.255	1.199; -0.225	1.118; 0.557	1.774; 1.477
Rassf3	91	4.163	2.293	0.683	0.727
Rps23	11+12	0.783	1.549; 0.335	1.017; 1.131	1.18; 0.884
Utrn	61	1.434	1.791	1.441	1.321

Gene name	Tile no.	Tile sequence
Bdnf	56	CCTTCTGCTTTAAGTGCCTACATTACCTAACAGTGCTCAAGAGGTTCTCGATTGGAGAA CCACACTCAAATCCATTTATAGCCTCCATCCCATTTCTAAA
Cald1	58	ACTCTAGTGTCTTACAAAAATGTTCTTATCTAGAACATTCTATGCATACAGCGTCATTGT ACAAACTCATACCTCAGCTGACTTAACCAAAATTATCTTA
Camk2n1	12	AAAAACAAAAGTTAAAAAAAAAACACACAAAAAGTAAAAAAATAAAATAAAAAAAATA AAAATCACTATATACACACATATAAAAGAAAAAAGTCTCAGTT
Cox5b	6+7	CATAGCTGGCTAGTCTTATGTCTATAGTCTTTTTTTTTTTAATATATATATATTGAGACC ACCTATCTCTTAATACTAGTTTGGCATCCA
Golim4	56	TTATGTATAAAAGCCCCATATGTAATGTAAAATGTATTATATTGTACATAAATACTCAG AATACAATTATTTTGGTAAATTAGTTTGTATTTTTAATGTC
Kif1c	80	TGATAGAGTCCAAAGGTCTCTGCAGTCAGGGTGAGCTGTGTGGTGCAACGTAGGAGAG TCCAGATGGGGGGTGGCGGGGAATGCCAGTCTCAGCTACCAC
Mcf2l	7	GACATGACCTATGTGTTCCTACCTCAGACATGTCCATCACACACGAATGACCTGTGTGT TCCTACCTCAGACATGACCATGTGTGTTCCTACCTCAGATA
Msn	48	CTGCTGGGTATTCTGGCCTTGTCCCTTCACACCCCCACACCCTTTTCTCCACTTTGGAGC AGCTGCTCCCCCTTTTTTTGTCACATGTGTGCTCTTTTAAT
Ndufa2	11+12	CAACACTGAGGCAAAGTAGTTTTATATAAAAATACTCCTTTATTTCTCCTCAAAAAAAA AAAAAAAAAAAACCCACCAGGTGCCA
Rassf3	91	ATAAAGGAAGCATAAAGCATAAGGGAAGAAAAGGAGCCCCAGAGGCCAAAGAGATAT ATATATATATATTTTAAACCGTTGTTCAGATTCTCCGGTAAAG
Rps23	11+12	ATAACACTGCAATAAGTGGACACTGTCAACATGTGCAGTATTCTAAATGTGCCTTGAAT TTCTATTGTTCCTTGGCAATTGTCAAGTTTA
Utrn	61	AGCATAGATATTTTGAGACGAAGAAAATTGTTTTATATAAGGGGAGAGCCATGATCAC CTTTCTACCTCAGAACCACCTTCCTCCATTGTGTTGGACATA

Table S5: Genes with peaks of neurite enriched tiles across all 3 Nzip experiments.
The table lists the gene names, ID numbers of tiles, mean GC-content of all peaks identified in any of the 3 Nzip experiments, as well as the number of experiments in which a peak was found and the enrichment found in each experiment. Peaks are summarized from all consecutive groups of 2 or more tiles with neurite enrichment of at least log2fc >0.5, from which at least one had log2fc >0.75 and adj. p-value <0.05. Overlapping peaks from different experiments were merged. The last column gives the average and maximum neurite enrichment within a peak as well as the most significant p-value [≈avg (max.), p-value] for each dataset, in which the peak was detected.

Gene name	Peak tiles	mean GC [%]	No. experiments	Enrichment
Abhd8	1 - 4	13.3	1	Initial [0.909 (0.956), p<0.001]
Actb	3 - 6	3.5	1	Confirmation [2.292 (4.498), p=0.039]
Adcy1	28 - 29	7.5	1	Confirmation [1.064 (1.561), p=0.001]
Adcy1	139 - 141	3.7	1	Confirmation [0.619 (0.779), p=0.004]
Amotl1	229 - 230	4.5	2	Initial-depolarized [1.421 (1.495), p<0.001]; Initial [1.227 (1.342), p<0.001]
Amotl1	16 - 17	2.0	1	Confirmation [0.859 (0.862), p=0.001]
Anp32b	3 - 5	6.0	3	Confirmation [0.801 (0.951), p<0.001]; Initial-depolarized [0.803 (1.002), p=0.025]; Initial [0.777 (0.95), p<0.001]
Arc	18 - 20	12.7	1	Confirmation [0.701 (0.787), p=0.008]
Arc	31 - 32	7.5	1	Confirmation [0.708 (0.861), p<0.001]
Bdnf	108 - 110	1.7	1	Initial [1.455 (1.758), p<0.001]
Bola2	18 - 19	5.3	1	Confirmation [0.675 (0.756), p=0.003]
Cald1	58 - 60	3.0	3	Confirmation [1.162 (1.761), p<0.001]; Initial-depolarized [0.959 (1.348), p<0.001]; Initial [0.947 (1.336), p<0.001]
Cald1	74 - 76	6.3	1	Initial [0.633 (0.784), p<0.001]
Cald1	41 - 42	1.5	1	Confirmation [1.082 (1.606), p=0.003]
Camk2a	54 - 56	9.0	1	Initial [0.668 (0.792), p<0.001]
Camk2n1	7 - 8	4.0	1	Initial [0.741 (0.757), p<0.001]
Camk2n1	12 - 13	1.0	1	Initial [0.929 (1.306), p<0.001]
Cflar (iso 1)	13 - 16	0.3	3	Confirmation [1.095 (1.738), p<0.001]; Initial-depolarized [1.31 (1.877), p<0.001]; Initial [1.354 (1.538), p<0.001]
Cflar (iso 2)	51 - 52	4.5	1	Initial [0.8 (0.843), p<0.001]
Cflar (iso 2)	5 - 7	6.7	1	Confirmation [0.723 (0.968), p=0.002]
Cflar (iso 2)	100 - 103	0.5	1	Confirmation [0.893 (1.301), p<0.001]
Cox5b	4 - 8	3.5	2	Confirmation [1.505 (2.558), p<0.001]; Initial [1.085 (1.671), p<0.001]
Gas5	159 - 160	3.5	2	Confirmation [1.621 (2.335), p=0.009]; Initial [1.218 (1.734), p<0.001]
Golim4	4 - 6	1.0	2	Initial-depolarized [1.272 (1.711), p<0.001]; Initial [0.941 (1.053), p<0.001]
Golim4	56 - 57	1.0	1	Initial [0.963 (1.272), p<0.001]
Golim4	77 - 80	2.3	1	Confirmation [0.681 (1.076), p=0.020]
Kif1c	2 - 3	8.0	1	Confirmation [0.914 (1.139), p=0.019]
Map2	4 - 5	1.5	1	Initial [0.762 (0.823), p<0.001]
Mcf2l	56 - 57	1.0	1	Initial [0.734 (0.946), p<0.001]
Mcf2l	3 - 13	0.2	3	Confirmation [1.125 (1.879), p<0.001]; Initial-depolarized [1.238 (1.856), p<0.001]; Initial [1.362 (2.264), p<0.001]
Mcf2l	32 - 33	12.0	1	Confirmation [0.829 (1.031), p=0.020]
Mcf2l	99 - 100	10.0	1	Confirmation [0.704 (0.83), p=0.010]
Msn	48 - 49	5.0	1	Initial [1.198 (1.254), p=0.016]
Ndufa2	11 - 12	1.4	1	Initial [1.626 (1.774), p<0.001]
Rasl10b	16 - 17	3.5	1	Confirmation [0.756 (0.993), p<0.001]
Rasl10b	37 - 39	7.3	1	Confirmation [0.702 (0.801), p<0.001]
Rassf3	4 - 5	7.0	1	Confirmation [0.659 (0.796), p=0.005]
Rassf3	90 - 93	5.5	2	Confirmation [1.448 (2.293), p<0.001]; Initial [0.871 (1.271), p=0.001]

Continued on next page

Table S5 – continued from previous page

Gene name	Peak tiles	mean GC [%]	No. exp.	Enrichment
Rps23	10 - 14	4.0	2	Initial-depolarized [1.038 (1.161), p=0.006]; Initial [0.926 (1.18), p<0.001]
Rps24	20 - 22	4.7	1	Initial [0.688 (0.879), p=0.001]
Shank3	29 - 33	6.6	1	Confirmation [0.623 (0.767), p=0.004]
Shank3	46 - 47	7.5	1	Confirmation [0.757 (0.921), p<0.001]
Shank3	70 - 71	3.5	1	Confirmation [1.259 (1.58), p=0.001]
Snhg6	19 - 20	4.5	1	Initial [0.744 (0.766), p=0.008]
Snhg6	80 - 83	2.7	1	Confirmation [1.99 (3.239), p<0.001]
Szrd1	4 - 5	5.5	1	Confirmation [0.952 (0.953), p<0.001]
Szrd1	52 - 53	9.0	1	Confirmation [0.769 (0.8), p=0.001]
Szrd1	91 - 95	8.4	1	Confirmation [0.737 (0.938), p<0.001]
Tagln2	19 - 20	9.0	1	Confirmation [0.809 (0.901), p<0.001]
Tagln2	4 - 9	4.3	1	Confirmation [0.676 (0.812), p<0.001]
Tmcc2	38 - 39	3.0	1	Initial [2.285 (3.421), p=0.002]
Trim56	267 - 270	4.4	3	Confirmation [0.884 (1.192), p=0.006]; Initial-depolarized [1.041 (1.892), p<0.001]; Initial [1.061 (1.829), p<0.001]
Trim56	142 - 144	3.0	1	Initial [1.047 (1.93), p=0.006]
Utrn	60 - 63	4.3	3	Confirmation [0.948 (1.791), p<0.001]; Initial-depolarized [1.27 (1.573), p<0.001]; Initial [1.231 (1.364), p<0.001]

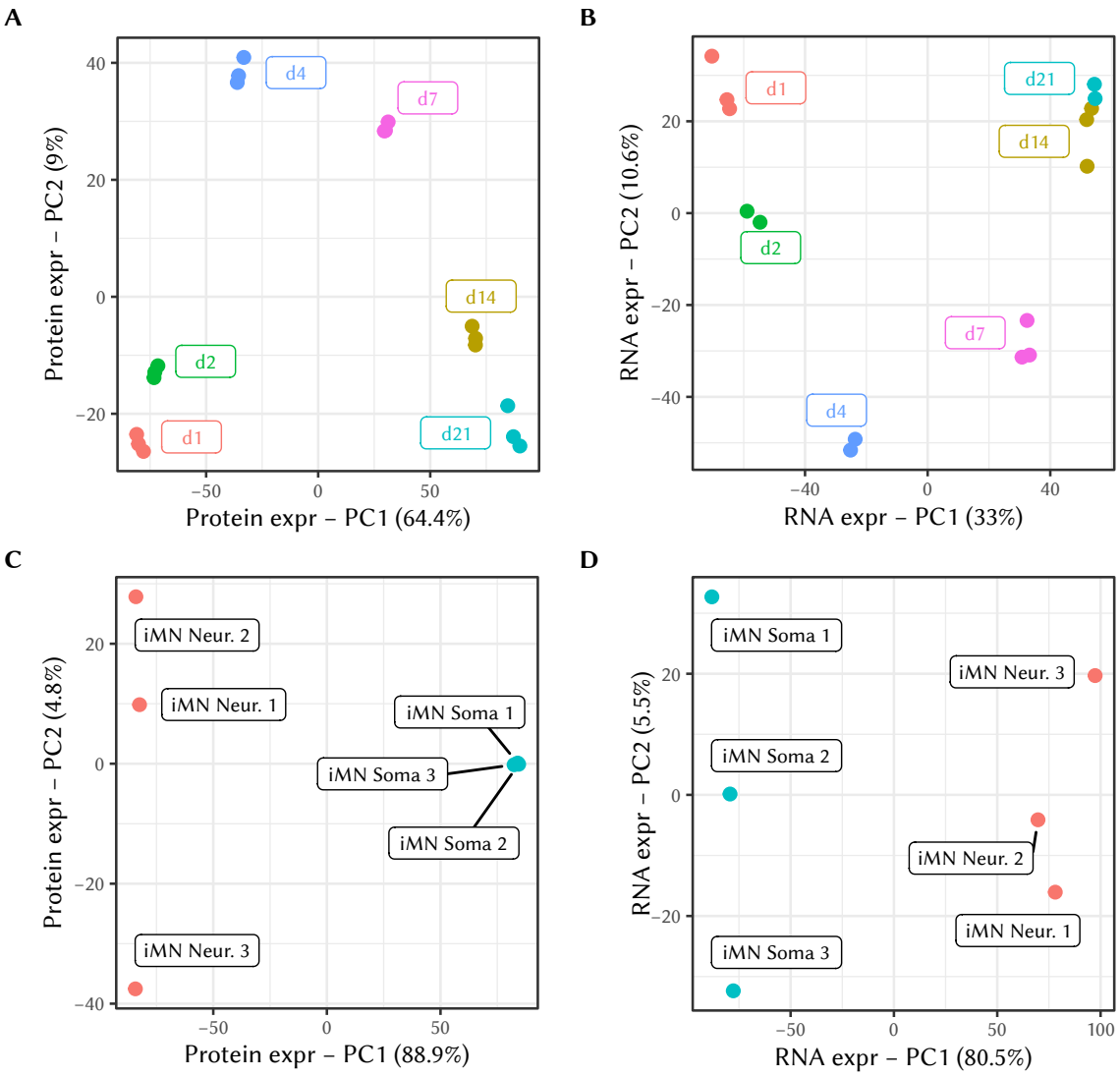


Figure S2: Principle component analysis of wild type iMN.
(A,B) PCA of (A) protein and (B) RNA expression data from different time points throughout the NIL differentiation protocol. (C,D) PCA of (C) protein and (D) RNA expression data from neurite (Neur.) and soma compartment of iMN at day 21. The first two principle components are shown on x- and y-axis, along with their contribution to the total variance. Data was generated by Katarzyna Ludwik.

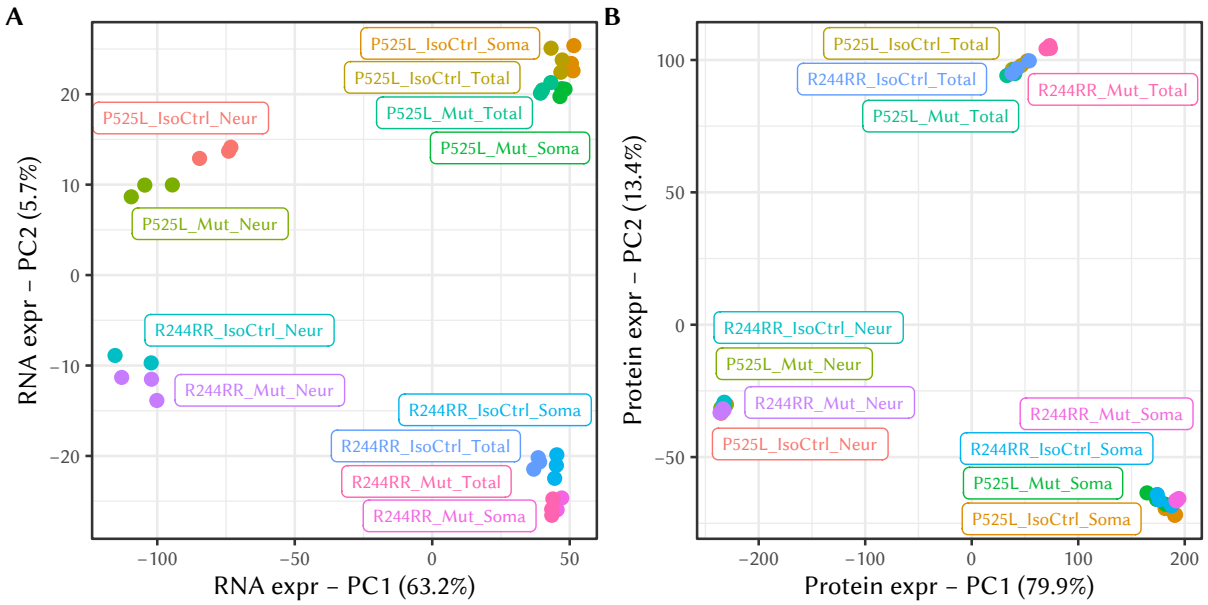


Figure S3: Principle component analysis of FUS mutant and control iMN.
PCA of (A) RNA and (B) protein expression data from neurite or soma compartment, or total cells from both P525L and R244RR FUS mutant or respective isogenic control iMNs. Replicate samples have the colour and respectively coloured labels indicate which mutation (P525L/R244RR), state of FUS (mut/IsoCtrl) and compartment (Neur, Soma, Total). The first two principle components are shown on x- and y-axis, along with their contribution to the total variance. Data was generated by Samantha Mendonsa.

Table S6: All genes with a differential expression score above 2 based on the effect of P525L and R244RR FUS mutations on RNA and protein in neurite, soma and total cells. The table includes Ensemble gene ID; gene name; RNA annotation features (left to right: G-quadruplex motif in any top expressed transcript, any wild type FUS CLIP peak, any mutant FUS CLIP peak (Hoell et al., 2011)); functional annotation of proteins (based signal-p database (Armenteros et al., 2019) or GO cellular compartment terms for plasma membrane components (GO:0005887, GO:0031226, GO:0046658, GO:0019897, GO:0044214) or extracellular proteins (GO:0005576); differential expression score (sum of 0.5 or 1 points for significant (adj. p-value < 0.05) differential expression (log2fc >1 or >1.5) across RNA and protein values in all compartments of both mutations); and for both FUS mutant lines (P525L and R244RR) indication of increased (Up) or reduced (Down) expression in any of the three compartments (N - neurite, S - soma, T - total) for either protein (P) or RNA (R) level or both (P+R).

Ensemble gene ID	Gene name	RNA features	Functional annotation	Score	Diff. expr. P525L	Diff. expr. R244RR
ENSG00000093010	COMT	- ✓ -		9	Up (N: P+R, S: R, T: R)	Up (N: P, S: P+R, T: P+R)
ENSG00000108439	PNPO	- - -		7.5	Down (N: R, T: R)	Up (N: P+R, S: P+R, T: P+R)
ENSG00000175426	PCSK1	✓ - -	Signal-Peptide	6.5	Down (N: P+R, S: R, T: P+R)	Down (N: P, S: P)
ENSG00000102195	GPR50	- - -	PM comp.	6	Down (N: P+R, S: R, T: R)	Down (S: R, T: R)
ENSG00000107438	PDLIM1	- - -		6	Up (N: R, S: R, T: R)	Up (N: P+R, S: P+R, T: R)
ENSG00000141449	GREB1L	- ✓ -		6	Up (N: R, S: R, T: R)	Up (S: P+R, T: P+R)
ENSG00000184470	TXNRD2	- ✓ -		6	Up (S: R, T: R)	Up (S: P+R, T: P+R)
ENSG00000091664	SLC17A6	- - -		5.5	Up (N: R)	Up (N: P+R, S: P+R, T: R)
ENSG00000115457	IGFBP2	- - -	Signal-Peptide; extracellular	5.5	Down (T: R)	Down (N: P+R, S: P+R, T: R)
ENSG00000173698	ADGRG2	- - -	PM comp.	5.5	Up (N: R, S: R, T: R)	Up (N: R, S: R, T: R)
ENSG00000197948	FCHSD1	- - -		5.5	Up (S: R, T: R)	Up (N: P+R, S: R, T: R)
ENSG00000013364	MVP	✓ - -	extracellular	5	Down (N: R, T: P)	Down (N: P, S: P, T: P)
ENSG00000114200	BCHE	- - -	Signal-Peptide; extracellular	5	Up (N: R, S: R, T: R)	Up (N: R, S: R)
ENSG00000140545	MFGE8	- - -	Signal-Peptide; PM comp.; extracellular	5		Up (N: P, S: P+R, T: P+R)
ENSG00000158710	TAGLN2	- - -	extracellular	5	Down (T: R)	Up (N: P+R, S: R, T: P+R)
ENSG00000166426	CRABP1	- - -		5	Down (N: R, S: R, T: R)	Up (N: P+R, T: P)
ENSG00000180543	TSPYL5	✓ - -		5	Up (N: R, S: R, T: R)	Up (N: R, S: P+R, T: R)
ENSG00000184144	CNTN2	- - -	Signal-Peptide	5		Up (N: P, S: P+R, T: P+R)
ENSG00000006128	TAC1	- - -	Signal-Peptide; extracellular	4.5	Down (N: R, S: R, T: R)	Down (S: P+R)
ENSG00000135111	TBX3	- ✓ ✓		4.5	Up (N: R, S: R, T: R)	Up (S: R, T: R)
ENSG00000140945	CDH13	- - -	Signal-Peptide; extracellular	4.5	Down (S: R, T: R)	Down (N: P, S: P, T: P+R)
ENSG00000152192	POU4F1	✓ ✓ -		4.5	Up (N: R, S: R, T: R)	Up (N: R, S: R)
ENSG00000170893	TRH	- - -	Signal-Peptide; extracellular	4.5	Down (S: R, T: R)	Down (N: R, S: P+R, T: P+R)
ENSG00000253953	PCDHGB4	✓ - -	Signal-Peptide; PM comp.	4.5	Up (N: R, S: R, T: R)	Down (N: P+R, S: R, T: R)
ENSG00000049247	UTS2	- - -	Signal-Peptide; extracellular	4		Up (N: R, S: P+R, T: R)
ENSG00000104313	EYA1	- ✓ -		4	Up (N: R, S: R, T: R)	Up (N: R)
ENSG00000109107	ALDOC	- - -	extracellular	4		Down (N: P+R, S: P, T: P+R)
ENSG00000127399	LRRC61	- - -		4	Up (N: R, S: R)	Up (N: R, T: R)
ENSG00000129824	RPS4Y1	- - -		4		Down (S: P+R, T: P+R)

Continued on next page

Table S6 – continued from previous page

Ensemble gene ID	Gene name	RNA feat.	Annotation	Score	Diff. expr. P525L	Diff. expr. R244RR
ENSG00000134115	CNTN6	- - -	Signal-Peptide	4	Up (T: R)	Up (N: P, S: R, T: P)
ENSG00000156466	GDF6	✓ - -	Signal-Peptide	4	Up (N: R)	Up (N: R, S: R, T: R)
ENSG00000196233	LCOR	✓ ✓ ✓		4	Down (N: R)	Up (N: R)
ENSG00000198417	MT1F	- - -		4	Up (N: R)	Down (N: R, S: R, T: R)
ENSG00000203867	RBM20	- ✓ -		4	Up (N: R, T: R)	Up (N: R, T: R)
ENSG00000214076	CPSF1P1	- - -		4	Down (S: R, T: R)	Up (S: R, T: R)
ENSG00000275342	PRAG1	- - -		4	Down (N: R, S: R, T: R)	Down (N: R)
ENSG00000070808	CAMK2A	✓ - -		3.5		Up (S: P+R, T: P+R)
ENSG00000076003	MCM6	- ✓ ✓		3.5	Up (N: R)	Up (N: R, S: P, T: P); Down (N: R, S: P, T: P)
ENSG00000130226	DPP6	✓ - -		3.5	Down (N: R)	Up (N: R, S: R, T: R)
ENSG00000130287	NCAN	- - -	Signal-Peptide; ECM	3.5	Down (N: R)	Up (N: R, S: P, T: P)
ENSG00000139220	PPFIA2	- - -		3.5		Down (N: P+R, S: P+R, T: P)
ENSG00000142798	HSPG2	- - -	Signal-Peptide; extracellular	3.5	Down (T: P)	Up (N: P, S: P, T: P); Down (N: P, S: P, T: P)
ENSG00000146242	TPBG	- - -	Signal-Peptide; PM comp.	3.5	Down (T: P)	Down (N: P, S: R, T: R)
ENSG00000147065	MSN	- - -		3.5	Down (S: R, T: R)	Up (N: R, S: R)
ENSG00000165731	RET	- - -	Signal-Peptide; PM comp.	3.5	Up (N: R, T: R)	Up (N: P, S: P+R, T: P)
ENSG00000170421	KRT8	- - -		3.5	Down (T: R)	Up (N: R, S: R, T: R)
ENSG00000198729	PPP1R14C	- ✓ ✓		3.5	Down (N: R, T: R)	Down (N: R, S: R, T: R)
ENSG00000009694	TENM1	- ✓ -	PM comp.; extracellular	3	Down (T: R)	Down (N: P, T: P)
ENSG00000010292	NCAPD2	- ✓ -		3		Up (N: R, S: P, T: P)
ENSG00000011201	ANOS1	- ✓ -	Signal-Peptide; ECM	3		Up (N: P, S: P, T: P)
ENSG00000066248	NGEF	✓ - -		3	Down (N: R, S: R, T: R)	
ENSG00000072501	SMC1A	- ✓ ✓		3	Down (N: R, S: R, T: R)	Up (N: R, T: P); Down (N: R, T: P)
ENSG00000077092	RARB	- ✓ -		3	Up (N: R, T: R)	Up (N: R)
ENSG00000080345	RIF1	✓ ✓ ✓		3		Up (N: R, S: P, T: P)
ENSG00000087494	PTHLH	- - -	Signal-Peptide; extracellular	3	Down (T: R)	Up (N: R, T: R)
ENSG00000104738	MCM4	- ✓ ✓		3		Up (N: R, S: P, T: P); Down (N: R, S: P, T: P)
ENSG00000106536	POU6F2	- - -		3	Up (N: R, T: R)	Up (S: R)
ENSG00000115461	IGFBP5	✓ - -	Signal-Peptide; extracellular	3	Down (T: R)	Down (S: R, T: R)
ENSG00000118407	FILIP1	✓ - -		3	Up (N: R, S: R, T: R)	
ENSG00000124313	IQSEC2	- - -		3	Down (N: R, S: R, T: R)	Down (N: R)
ENSG00000125630	POLR1B	- - -		3	Up (N: R)	Up (N: R, T: P)
ENSG00000128610	FEZF1	- - -		3	Up (N: R)	Up (S: R, T: R)
ENSG00000131409	LRRC4B	- - -	Signal-Peptide	3	Up (N: P, T: P+R)	
ENSG00000131914	LIN28A	- - -		3	Up (S: R)	Up (N: R, S: P, T: P)
ENSG00000133135	RNF128	- - -	Signal-Peptide	3	Down (T: R)	Up (N: R, T: R)

Continued on next page

Table S6 – continued from previous page

Ensemble gene ID	Gene name	RNA feat.			Annotation	Score	Diff. expr. P525L	Diff. expr. R244RR
ENSG00000133789	SWAP70	✓	-	-		3		Up (N: R, S: P, T: P); Down (N: R, S: P, T: P)
ENSG00000135069	PSAT1	-	✓	✓		3	Down (N: R)	Up (N: R, S: P, T: P)
ENSG00000137766	UNC13C	-	-	-		3		Down (N: R, S: R, T: R)
ENSG00000143429	LSP1P4	-	-	-		3		Up (N: R, S: R, T: R)
ENSG00000146469	VIP	-	-	-	Signal-Peptide; extracellular	3	Down (S: R)	Down (S: R, T: R)
ENSG00000147036	LANCL3	-	-	-		3	Up (T: R)	Up (N: P+R, T: P)
ENSG00000147256	ARHGAP36	-	-	-	Signal-Peptide	3	Down (N: R, S: R, T: R)	
ENSG00000154146	NRGN	✓	-	-		3		Up (N: R, S: R, T: R)
ENSG00000157005	SST	-	-	-	Signal-Peptide; extracellular	3	Up (T: R)	Up (N: R, S: R)
ENSG00000160801	PTH1R	-	-	-	Signal-Peptide; PM comp.	3	Up (N: R, S: R, T: R)	Down (T: R)
ENSG00000161681	SHANK1	-	-	-		3	Up (N: R, S: R, T: R)	
ENSG00000164093	PITX2	-	-	-		3		Down (N: R, S: R, T: R)
ENSG00000164106	SCRG1	-	-	-	Signal-Peptide	3		Down (N: R, S: R, T: R)
ENSG00000164853	UNCX	-	-	-		3		Down (N: R, S: R, T: R)
ENSG00000165675	ENOX2	-	✓	-		3	Down (T: R)	Up (N: R, S: R)
ENSG00000165985	C1QL3	✓	-	-	Signal-Peptide	3	Down (N: R, T: R)	Down (N: P)
ENSG00000166573	GALR1	-	-	-	PM comp.	3		Down (N: R, S: R, T: R)
ENSG00000167244	IGF2	✓	-	-	Signal-Peptide; extracellular	3		Down (N: R, S: R, T: R)
ENSG00000169851	PCDH7	-	✓	✓	Signal-Peptide; PM comp.	3	Up (N: R)	Down (N: P+R, T: P+R)
ENSG00000169860	P2RY1	-	-	-	PM comp.	3	Down (N: R, S: R, T: R)	
ENSG00000170430	MGMT	-	✓	-		3		Up (N: R, S: R, T: R)
ENSG00000172346	CSDC2	✓	-	-		3	Down (N: R, S: R, T: R)	
ENSG00000175183	CSRP2	-	✓	-		3	Down (N: R)	Up (S: P, T: P)
ENSG00000179520	SLC17A8	-	-	-		3	Down (T: R)	Down (S: R, T: R)
ENSG00000179598	PLD6	-	-	-		3		Up (N: R, S: R, T: R)
ENSG00000182836	PLCXD3	-	-	-		3		Up (N: R, S: P+R, T: P)
ENSG00000183337	BCOR	-	✓	✓		3		Down (N: R, S: R, T: R)
ENSG00000183379	SYNDIG1L	-	-	-		3	Up (N: R, S: R, T: R)	
ENSG00000183688	RFLNB	✓	-	✓		3	Down (N: R, S: R, T: R)	
ENSG00000184261	KCNK12	✓	-	-	PM comp.	3		Up (N: R, S: R, T: R)
ENSG00000184445	KNTC1	-	✓	-		3	Up (N: R)	Up (S: P, T: P)
ENSG00000196083	IL1RAP	-	-	-	Signal-Peptide; PM comp.; extracellular	3	Down (N: R, T: R)	Up (N: R)
ENSG00000224597	SVIL-AS1	-	-	-		3	Up (S: R, T: R)	Up (N: R)
ENSG00000243970	PPIEL	-	-	-		3	Down (S: R, T: R)	Up (N: R)
ENSG00000273748	NA	-	-	-		3	Up (N: R, S: R, T: R)	
ENSG00000068784	SRBD1	-	✓	-		2.5	Down (N: R)	Up (N: R, T: P)
ENSG00000078549	ADCYAP1R1	-	-	-	Signal-Peptide; PM comp.	2.5	Down (N: R)	Down (N: R, T: R)
ENSG00000099250	NRP1	-	✓	-	Signal-Peptide	2.5	Up (N: R, T: R); Down (N: R, T: R)	Up (N: R, T: R)

Continued on next page

Table S6 – continued from previous page

Ensemble gene ID	Gene name	RNA feat.	Annotation	Score	Diff. expr. P525L	Diff. expr. R244RR
ENSG00000102038	SMARCA1	- - -		2.5	Down (N: R, T: R)	Up (T: P)
ENSG00000104299	INTS9	- ✓ -		2.5	Down (N: R)	Up (N: R, T: P)
ENSG00000105287	PRKD2	✓ - -		2.5		Up (N: R, S: P, T: P)
ENSG00000105894	PTN	✓ - -	Signal-Peptide; extracellular	2.5	Up (N: R, S: R, T: R)	Up (N: R)
ENSG00000107731	UNC5B	✓ - -	Signal-Peptide	2.5		Up (N: R, S: R, T: R)
ENSG00000112232	KHDRBS2	- - -		2.5		Down (S: P, T: P+R)
ENSG00000114115	RBP1	- - -		2.5	Up (N: R, S: R, T: P+R)	
ENSG00000116962	NID1	- - ✓	Signal-Peptide; extracellular	2.5	Down (T: P)	Down (N: P, T: P)
ENSG00000117154	IGSF21	- - -	Signal-Peptide	2.5	Down (N: R, S: R)	Down (N: R, T: R)
ENSG00000122254	HS3ST2	- - -		2.5	Down (N: R)	Down (S: R, T: R)
ENSG00000122971	ACADS	✓ - -		2.5	Down (N: R)	Up (N: R, S: P); Down (N: R, S: P)
ENSG00000123360	PDE1B	✓ - -		2.5	Down (N: R)	Up (N: R, T: P)
ENSG00000123570	RAB9B	- - -		2.5	Down (N: R, S: R, T: P+R)	
ENSG00000130158	DOCK6	- - -		2.5	Up (N: R)	Up (S: P, T: P)
ENSG00000130702	LAMA5	✓ - -	Signal-Peptide; extracellular	2.5		Up (N: P+R, T: P); Down (N: P+R, T: P)
ENSG00000140836	ZFHX3	✓ ✓ -		2.5	Down (N: R, T: R)	Down (N: R)
ENSG00000141433	ADCYAP1	- - -	Signal-Peptide; extracellular	2.5	Down (N: R, S: R, T: R)	
ENSG00000145868	FBXO38	- ✓ -		2.5	Down (N: R)	Up (N: R, T: P)
ENSG00000151376	ME3	- - -		2.5		Up (N: P, S: P, T: P)
ENSG00000153922	CHD1	- ✓ ✓		2.5	Up (N: R)	Up (N: R, T: P); Down (N: R, T: P)
ENSG00000157483	MYO1E	- ✓ -		2.5	Down (N: R)	Up (S: P, T: P)
ENSG00000159167	STC1	- - -	Signal-Peptide	2.5	Down (N: R, S: R, T: R)	
ENSG00000159713	TPPP3	- - -		2.5	Up (N: R)	Up (N: R, S: R, T: P)
ENSG00000163918	RFC4	- ✓ -		2.5	Up (N: R)	Up (S: P, T: P)
ENSG00000165966	PDZRN4	- - -	Signal-Peptide	2.5	Up (N: R)	Down (S: R, T: R)
ENSG00000166407	LMO1	- - -		2.5	Up (T: R)	Up (N: R, S: R)
ENSG00000173473	SMARCC1	- ✓ ✓		2.5	Up (N: R)	Up (N: R, T: P); Down (N: R, T: P)
ENSG00000182568	SATB1	- ✓ -	Signal-Peptide	2.5	Up (N: R, T: R)	Down (N: R)
ENSG00000185532	PRKG1	✓ ✓ ✓		2.5	Up (T: P)	Up (N: R, S: P, T: P); Down (N: R, S: P, T: P)
ENSG00000186522	SEPTIN10	- - -		2.5	Down (N: R)	Up (S: P, T: P)
ENSG00000196104	SPOCK3	- - -	Signal-Peptide; extracellular	2.5	Up (N: R)	Up (N: R, S: P, T: P); Down (N: R, S: P, T: P)
ENSG00000196482	ESRRG	- - -		2.5	Up (N: R, T: R)	Up (N: R)
ENSG00000198586	TLK1	- ✓ -		2.5	Down (N: R)	Up (N: R, S: P); Down (N: R, S: P)
ENSG00000204131	NHSL2	- - -		2.5	Down (N: R, T: R)	Up (N: R)
ENSG00000043355	ZIC2	✓ - -		2		Down (S: R, T: R)