2019-07-31

Intron and Small RNA Localization in Mammalian Neurons

Harleen Saini
University of Massachusetts Medical School

INTRON AND SMALL RNA LOCALIZATION IN MAMMALIAN NEURONS

A Dissertation Presented

By

Harleen Saini

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial filfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

JULY 31st, 2019

NEUROSCIENCE

INTRON AND SMALL RNA LOCALIZATION IN MAMMALIAN NEURONS

A Dissertation Presented
By
Harleen Saini

This work was undertaken in the Graduate School of Biomedical Sciences
Program in Neuroscience

Under the mentorship of

Melissa J. Moore, Ph.D., Co-Thesis Advisor

Sean R. Eddy, Ph.D., Co-Thesis Advisor

Phillip D. Zamore, Ph.D., Member of Committee

Jeanne Lawrence, Ph.D., Member of Committee

Sean Ryder, Ph.D., Member of Committee

Athma Pai, Ph.D., Member of Committee

Brenton R. Graveley, Ph.D., External Member of Committee

William Theurkauf, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

July 31st 2019

ACKNOWLEDGEMENTS

I am deeply grateful to everyone who has been a part of this journey. First and foremost, my mentors and role models, Melissa Moore and Sean Eddy. Melissa, for teaching me how to look at things upside down, inside out, and every which way with an open mind, and for making me believe that I could do anything given the will. Sean, for taking me in, keeping me excited about my research even when things got hard, teaching me how to express myself, and getting me excited about all kinds of questions that lay ahead. Between the two of them, brilliant ideas abound, bringing a constant sense of excitement to every conversation.

I also found invaluable camaraderie in both laboratories: Alicia Bicknell and Tom Jones have been instrumental in my personal and professional development. At UMass, I had the pleasure of discussing science and life with Guramrit Singh, Emiliano Ricci, Joerg Braun, Mihir Metkar, Chris Roy, Bo Han, Zhao Zhang, Blandine Mercier, Akiko Noma, and Hakan Ozadam. At Harvard, there were no pipettes, no hum of the centrifuge, or the smell of 200 proof ethanol to stand by but there were many lively discussions about probability theory, Markov models, neural networks, the genetic code, the origin of life, and occasionally, music and science fiction, thanks to Kate Shulgina, Grey Wilburn, Peter Koo, Cara Weisman, Tim Dunn, Nick Carter, Jenny Chen, and Mary Richardson. Thanks also to Elena Rivas and William Gao for teaching me about RNA structure prediction and for providing feedback on my research.

I am grateful to Phil Zamore for always steering me in the right direction and my advisory committee - Marc Freeman, Bill Theurkauf, Jeanne Lawrence, and Sean Ryder - for their guidance and support. I would also like to express my appreciation to Athma Pai and Brent Graveley for agreeing to serve on my dissertation examination committee.

I am thankful to HHMI for providing me with funds that paid a part of my stipend but more importantly, gave me the confidence and independence to pursue an open-ended exploratory project. I am also thankful to Harvard MCB for hosting me and treating me like one of their own.

Special mention goes to Laureen Murtha, Rina Paladino, Hongyan Wang, Tina Nesbeda, Michelle Merry, and Veronica Chudik for making everyday operations at the Moore lab, the RTI, and the Eddy lab run smoothly. I could not have dreamt of such support for pursuing my scientific passions.

And I could have done none of this without the support of my family, as well as friends, and especially my partner, Mike Do, who has always inspired me and naturally brought the best out of me.

To all my teachers, both in front of and behind the scenes.

". . . that there is such a lot to find in it,
but such a lot of it to find it in."

Bill Bryson, In a Sunburned Country

INTRON AND SMALL RNA LOCALIZATION IN MAMMALIAN NEURONS

by Harleen Saini

ABSTRACT

RNA molecules are diverse in form and function. They include messenger RNAs (mRNAs) that are templates for proteins, splice products such as introns that can generate functional noncoding RNAs, and a slew of smaller RNAs such as transfer RNAs (tRNAs) that help decode mRNAs into proteins. RNAs can show distinct patterns of subcellular localization that play an important role in protein localization. However, RNA distribution in cells is incompletely understood, with prior studies focusing primarily on RNAs that are long (>200 nucleotides), fully processed, and polyadenylated. We examined the distribution of RNAs in neurons. Neuronal compartments can be separated by long distances and play distinct roles, raising the possibility that RNA localization is especially overt and functionally meaningful in these cells. In our exploration, we physically dissected projections from cell bodies of neurons from the rat brain and sequenced total RNA. We describe two main findings. First, we identified excised introns that are enriched in neuronal projections and confirmed their localization by single-molecule fluorescence in situ hybridization. These are a previously unknown set of circular RNAs in neuronal projections: tailless lariats that possess a non-canonical C branchpoint. Second, we observed a highly abundant population of small (20-150 nucleotide) RNAs in neuronal projections, most of which are tRNAs. For both circular introns and tRNAs, we did not observe known RNA localization signals. Thus, many types of RNA, if sufficiently stable, appear free to diffuse to distant locations, their localization perhaps aided by the movement of large organelles in the confines of neuronal projections. Our survey of RNA molecules across subcellular compartments provides a foundation for investigating the function of these molecules and the mechanisms that localize them.

Contents

List of Figures

A	Adenine
T	Thymine
C	Cytosine
G	Guanine
U	Uracil
I	Inosine
bp	A base pair of DNA
DNA	DeoxyRibonucleic Acid
EI	Exon-Intron
IE	Intron-Exon
kb	kilobase
miRNA	micro RNA
mRNA	messenger RNA
NMD	Nonsense Mediated Decay
nt	mucleotide
PASseq	PolyA-site sequencing
PolyA+	Polyadenylated
Poly-	Nonadenylated
pre-mRNA	precursor to mRNA
PTC	Premature Termination Codon
RiboSeq	Ribosome profiling
RIN	RNA integrity number
RNA	RiboNucleic Acid
RNAseq	RNA sequencing
RP mRNAs	Ribosomal Protein-encoding mRNAs
rRNA	ribosomal RNA
s	second (time)
smFISH	single-molecule Fluorescence In Situ Hybridization
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
pre-tRNA	precursor to tRNA
TPM	transcripts per million
tRNA	transfer RNA
5′	The 5 prime end of a DNA or RNA molecule
3′	The 3 prime end of a DNA or RNA molecule
-OH	hydroxyl
µ	micro ($1x10^{-6}$ standard units)

PREFACE

The work presented in **Chapter II** is under review and a preprint version is avialable at doi: https://doi.org/10.1101/623447

Saini H., Bicknell A. A., Eddy S. R., Moore M. J. (2019) Free circular introns with an unusual branchpoint in neuronal projections. *Under review.*

CHAPTER I

Introduction

This dissertation presents an exploration of the various types of RNA distributed across subcellular compartments of cells that have extended morphologies. Based on the findings of this exploration, two types of RNA are of principal concern. The first are the parts of the messenger RNA (mRNA) that, typically, do not code for protein sequences, the introns. The second are a class of RNAs that decode mRNA to protein, the transfer RNAs (tRNAs). The subcellular distributions of these RNAs are examined in neurons, a cell type that tends to be extremely polarized, with one end customized to receive information and the other to transmit it. Neuronal compartments can be separated by a substantial distance, providing the opportunity for RNA localization to play a functional role that is particularly prominent.

In this introduction, I review topics that are important to the exploration of RNA localization in neurons. First, are there *a priori* reasons for localizing RNA within different subcellular compartments? Second, what has been reported about intron localization within neurons? Third, what is known about the subcellular distribution of small noncoding RNAs (<200 nt, which includes tRNA) in neurons? Finally, I will discuss the challenges that hinder a deeper understanding of these topics, and how research described in this dissertation takes steps toward meeting those challenges.

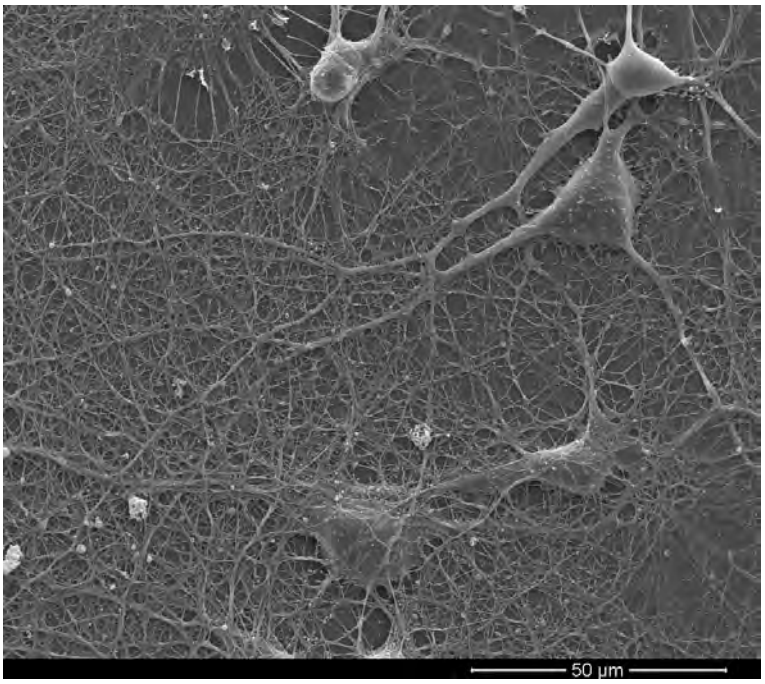


FIGURE I.1: Extended and asymmetric neuronal morphology.

A scanning electron micrograph of primary hippocampal neurons from rat embryos, maintained in culture for 14 days (Image courtesy of Alicia A. Bicknell). The bounds of this image do not encompass the entire denritic network of a single cell, nor do they capture the full length of an axon.

Protein and RNA localization in asymmetric cells

Many cell types are polarized, and the operation of particular subcellular compartments depends to a large degree on the localization of molecular machinery, especially proteins. In principle, localization could occur by synthesizing proteins as soon as the blueprints (the mRNA) are released from the nucleus, and then shipping the proteins to where they are needed. Alternatively, mRNAs could be shipped to those locations and translated locally. Are there reasons for favoring one over the other? These questions become especially acute for cell types that

have extended morphologies (Figure I.1).

Neurons provide an informative point of discussion. Broadly speaking, the textbook neuron has three compartments: the soma, which includes the nucleus and cytoplasm; the dendrites, which receive and integrate information; and the axon, which transmits this information to where it needs to go (and can be over a meter long in some mammals). Importantly, neuronal projections (dendrites and axons) can be very thin and long, perhaps providing a bottleneck that makes the advantages of transporting mRNA versus proteins clear.

The minimum diameter of neuronal projections is thought to be constrained by the need to fit a molecular transport system and allow bidirectional movement of cargo [Sterling and Laughlin, 2015]. The transport machinery typically consists of microtubule tracks (~30 nm diameter) and motor proteins (length of kinesin is ~10 nm and dynein is ~15 nm). The cellular membrane and its underlying cytoskeleton add ~30 nm to the diameter. With room for cargo (proteins, organelles, RNA, etc.) and for ions to flow with relatively little resistance to support the propagation of electrical signals, the minimum necessary thickness of neuronal projections approaches the diameter of the finest axons observed, which is ~100 nm [Faisal et al., 2005]. Given these spatial constraints, is it favorable to transport protein or mRNA?

One could argue that an mRNA is massive compared to the protein it encodes

and thus proteins might be the preferred cargo for transport. On average, each nucleotide has a molecular mass of ~330 Da whereas each amino acid has a mass of ~110 Da, and three nucleotides are needed to encode one amino acid. That already shows the molecular mass of an RNA to be at least 9 times that of its encoded protein. In addition to the sequence that encodes protein (open reading frame; ORF), an mRNA often comprises untranslated regions (UTRs), a stretch of adenosines at its 3′ end (polyA tail), and is studded with proteins [Dreyfuss et al., 2002], putting the total mass of an mRNA at least an order of magnitude greater than the mass of the protein it encodes. Additionally, transport of the mRNA alone is also not sufficient. For translation to occur, the entire translation machinery (ribosomes, tRNAs, translation factors, etc.) must also be localized. Given these challenges, why ship RNA at all?

A need for mRNA delivery is apparent for myelin binding protein (MBP), one of the earliest studied cases of mRNA localization [Trapp et al., 1987]. In vertebrates, some axons are insulated by a fatty deposit called myelin, to enhance signal transmission. Myelin is a proteolipid, an extension of the plasma membrane of glial cells that wraps around axons. MBP is a lipophilic protein that provides structural support to myelin. Due to its lipophilicity, MBP can rapidly adhere to any membrane that it encounters. Glial cells express MBP under tight subcellular spatial control [Barbarese et al., 1988] possibly because ectopic expression of MBP can result in its assembly at the wrong place and can interfere with its

delivery to the right place [Müller et al., 2013]. A convenient way to segregate proteins would be to deliver their mRNA to the appropriate subcellular location and translate locally. Indeed, MBP expression is regulated by transporting its mRNA to glial projections. MBP mRNA localization is mediated by a protein binding to its 3′UTR that pulls the mRNA along microtubule tracks [Munro et al., 1999].

The case of MBP illustrates that mRNA localization and local translation can ensure protein expression in a specific subcellular compartment. An additional reason for localizing RNA could be to control the timing and amount of protein expression on site. A single mRNA can produce a few to thousands of protein copies depending on translation efficiency and mRNA stability. Signals within the mRNA itself can determine these parameters and can be present in the noncoding parts (5′ or 3′ UTRs) of the mRNA such that upon translation, the protein sequence is unaffected [Roy and Jacobson, 2013]. In case of MBP, the localization signal is embedded in its 3′ UTR. Another dendritically-localized mRNA Arc/Arg3.1 contains a stop codon that is recognized as premature by the nonsense-mediated decay (NMD) machinery. It is therefore subject to degradation via NMD immediately following its translation [Giorgi et al., 2007] producing only a few copies of protein per mRNA. Thus, mRNA localization and local translation offers many opportunities for spatial and temporal control of protein expression.

Hundreds of RNAs have been observed to localize to distal cellular compartments, many of which engage with the translation machinery for local protein synthesis [Briese et al., 2016, Cajigas et al., 2012, Gumy et al., 2011, Minis et al., 2014, Miyashiro et al., 1994, Poon et al., 2006, Poulopoulos et al., 2019, Taliaferro et al., 2016, Taylor et al., 2009, Zivraj et al., 2010]. While some RNAs localize specifically to dendrites (e.g. Map2; Garner et al. [1988]) or axons (e.g. Tau; Litman et al. [1993]), others are depleted from cellular projections (e.g. Srsf5; this work) and some are found in all cellular compartments (e.g. ribosomal protein-encoding mRNAs or RP mRNA; this work and Shigeoka et al. [2018]). It remains to be determined if various RNAs show asymmetric distribution because of active localization, depletion or aggregation. Furthermore, some RNAs may be homogeneous in concentration but appear enriched relative to other RNAs. Nevertheless, RNA localization is both common and controlled, and serves important functions at least for localized protein production.

Study of these questions may also have practical relevance. The importance of intracellular transport in neurons is emphasized by the debilitating diseases that result from impaired transport. Interestingly, there are examples of neurodegeneration arising from deficiency in protein localization (amyotrophic lateral sclerosis; Taylor et al. [2016]) and of mRNA splicing and transport (spinal muscular atrophy; Zhang et al. [2008]).

Subcellular localization and functions of intron sequences

In the 1970s, an attempt to map cytoplasmic mRNA to nuclear DNA led to a surprising observation: chunks of sequence present in DNA were missing from the cytoplasmic mRNA [Berget et al., 1977, Chow et al., 1977]. Although an intact copy of DNA is transcribed, parts of RNA ("intervening sequences" or introns) can be cut out and other parts ("expressed sequences" or exons) stitched together to form a mature RNA by a process called splicing. For most eukaryotic mRNAs, recognition of exon/intron boundaries, intron removal, and exon ligation is conducted by a dynamic collaboration between multiple proteins and RNA molecules, collectively forming a massive ribonucleoprotein complex called the spliceosome [Jurica and Moore, 2003].

From this perspective, introns seemed fairly easy to define as parts of an RNA that must be removed to form the mature mRNA. However, due to alternative splicing, not all introns are constitutively removed from an mRNA, nor are all exons constitutively retained in mature mRNA. Even exon-intron and intron-exon boundaries can vary [Nilsen and Graveley, 2010]. What, then, is an intron? In this dissertation, a spliceosomal intron is defined as an mRNA sequence that is at least sometimes removed by the process of splicing and contains consensus motifs that are recognized by the spliceosome — a 5′ splice site consensus at its 5′ end, a 3′ splice site consensus at its 3′ end, and a branch site consensus.

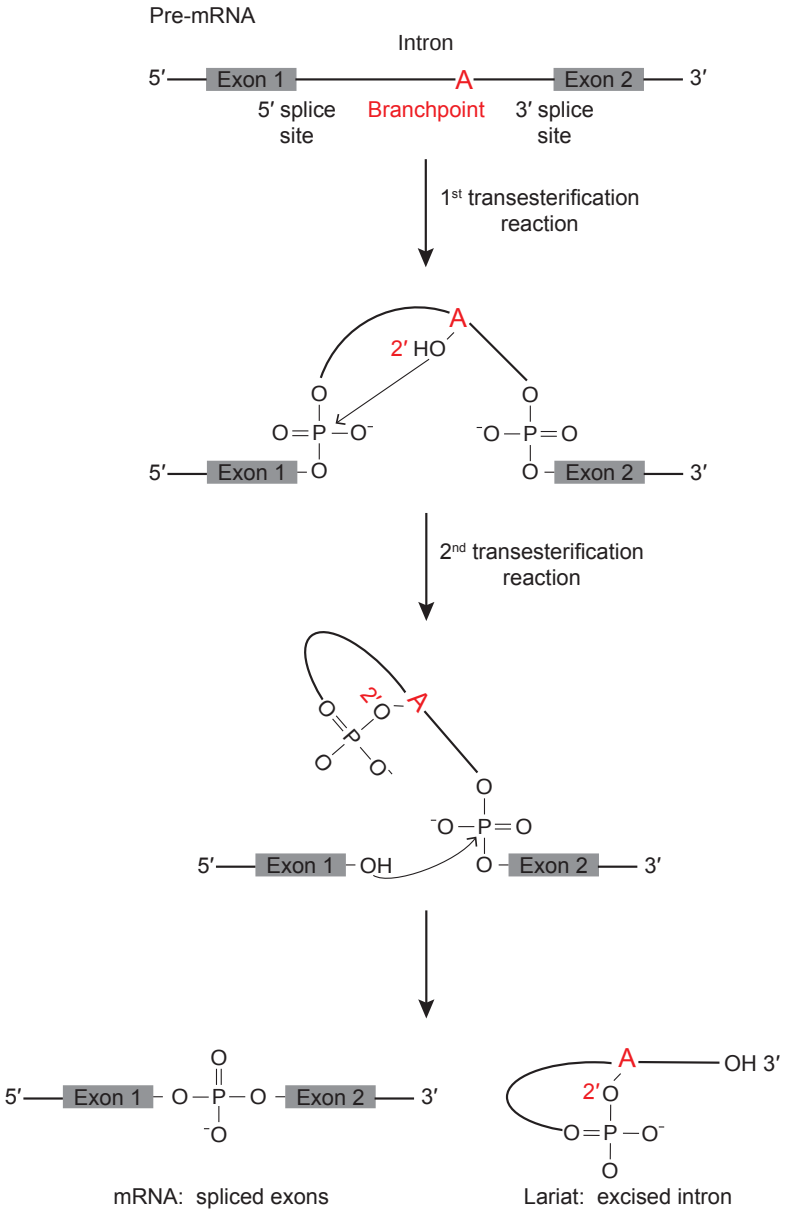


FIGURE I.2: Pre-mRNA splicing releases ligated exons (mature mRNA) and an excised intron lariat. Splicing involves two transesterification reactions that result in intron excision and exon ligation.

Intron removal and exon ligation, or splicing, involves two biochemical reactions. First, the 2′-OH of an adenosine at the branch site attacks the phosphodiester at the 5′ splice site, releasing a 3′-OH on the upstream exon and a lariat intermediate with a 2′-5′ phosphodiester bond. Then the 3′-OH of the upstream exon attacks the 3′ splice site, releasing an intron lariat and ligating the exons (Figure I.2). Once an intron is removed from the RNA, it is usually degraded by the combined actions of the debranching enzyme, which hydrolyzes the 2′-5′ phosphodiester bond, and exonucelolytic digestion [Lodish et al., 2000].

While constitutively spliced introns can be removed fairly quickly from the mRNA (on a time scale of 1–10 minutes in human cells; Singh and Padgett [2009]) and concomitantly with RNA transcription, alternatively spliced introns or introns flanking alternatively spliced exons are often removed relatively slowly or less efficiently [Pai et al., 2017, Pandya-Jones et al., 2013]. Thus, not only can intron boundaries vary, so can the the rate of splicing across introns. Some introns undergo splicing after an extended delay, in response to a signal, and have been called “detained introns” (these can persist in the nucleus for over an hour after transcription, Boutz et al. [2015]). At steady state, transcripts with detained introns show a higher concentration in the nucleus relative to the cytoplasm and appear to be withheld in the nucleus (similar examples in neurons are presented in Yap et al. [2012] and Mauger et al. [2016]; Figure I.3 (a)). Since steady state RNA abundance is determined by its rate of synthesis and decay, it is possible

that there are no barriers to the nuclear export of transcripts with detained introns but rather these transcripts are less stable in the cytoplasm. If detained introns are subject to degradation in the cytoplasm, then they will appear relatively enriched in the nucleus. The efficiency of nuclear export and the stability of transcripts with detained introns in the cytoplasm has not been directly tested.

What do we know about the fate of introns after they are spliced? The default assumption seems to be that introns are degraded rapidly in the nucleus after splicing but few studies have directly measured their half-lives [Clement et al., 1999, Moore, 2002]. However, multiple studies have provided evidence for stable excised introns in a variety of cell types [Gardner et al., 2012, Hesselberth, 2013, Morgan et al., 2019, Parenteau et al., 2019, Talhouarne and Gall, 2014, 2018, Zhang et al., 2013]. In the nucleus, some excised introns can function as precursors to small nucleolar RNAs (snoRNAs; an important class of noncoding RNAs broadly involved in rRNA and snRNA maturation) and microRNAs (miRNAs; noncoding RNAs involved in translation regulation and RNA degradation) [Hubé and Francastel, 2015]. More recently, two independent groups showed that environmental stress or nutrient deprivation correlated with an accumulation of a subset of excised linear introns in yeast cells [Morgan et al., 2019, Parenteau et al., 2019]. The authors posited that high intron abundance might sequester splicing factors thereby enhancing overall cell survival by globally inhibiting splicing and restricting cell growth. If the goal is to sequester splicing factors, then

the sequence or identity of accumulated introns might be irrelevant as long as enough introns accumulate. Although the two studies observed similar effects of intron accumulation, the identities of excised introns were indeed different.

The subcellular location of these excised linear introns is unknown but in a different study, specific lariat introns were detected in the cytoplasm of a variety of cell types from different vertebrate organisms [Talhouarne and Gall, 2018]. These introns also lack sequence conservation and the set of intron lariats varies between different cell types and organisms. A common feature of cytoplasmic lariat introns was that the 2′-5′ branch comprised a cytosine rather than the canonical adenosine. The presence of these lariat introns in both the nuclear and cytoplasmic compartments suggests that they might originate in the nucleus and localize to the cytoplasm after splicing. Whether these lariat introns have a functional impact in the cytoplasm or if they are localized to particular cytoplasmic domains remains to be determined.

A better known class of introns found in the cytoplasm includes retained introns, i.e., introns that are not spliced out of the mRNA. These are different from detained introns which are expected to be ultimately spliced, albeit in response to a signal and are enriched in the nucleus. Intron retention is a form of alternative splicing and transcripts with retained introns can exit the nucleus [Baralle and Giudice, 2017, Hilleren and Parker, 2003]. Intriguingly, in a large scale analysis

of polyadenylated RNAs from a variety of cell types, neuronal and immune cells showed a greater incidence of intron retention compared to cells from other organs (e.g., muscle, heart, liver, colon, etc.) [Braunschweig et al., 2014]. Since splicing does not always precede polyadenylation, it is unclear whether these retained introns are merely splicing intermediates that are restricted to the nucleus or legitimate cytoplasm-localized alternative splicing products. Examining the subcellular location of mRNAs with retained introns can provide insight to their function.

In some cases, cytoplasmic transcripts with retained introns are nonfunctional and are degraded by the cell's quality control mechanisms [Hilleren and Parker, 2003]. However, there are cases in which cytoplasmic intron retention serves a function. In neurons, the importance of intron retention is demonstrated most clearly in the case of axon guidance, in which neurons extend to connect with their downstream target. To reach these targets, axons usually follow conserved paths during development by adaptable attraction or repulsion to external cues. ROBO3, a protein that senses and responds to the extracellular protein, SLIT, plays an important role in axon development [Chen et al., 2008]. Robo3 expresses both a fully spliced mRNA isoform (Robo3.1) and a second isoform with intron 26 retained (Robo3.2). Proteins translated from its two isoforms differ in the C-terminus and respond antagonistically to SLIT: Robo3.2 expression results in repulsion to SLIT whereas Robo3.1 expression mutes this

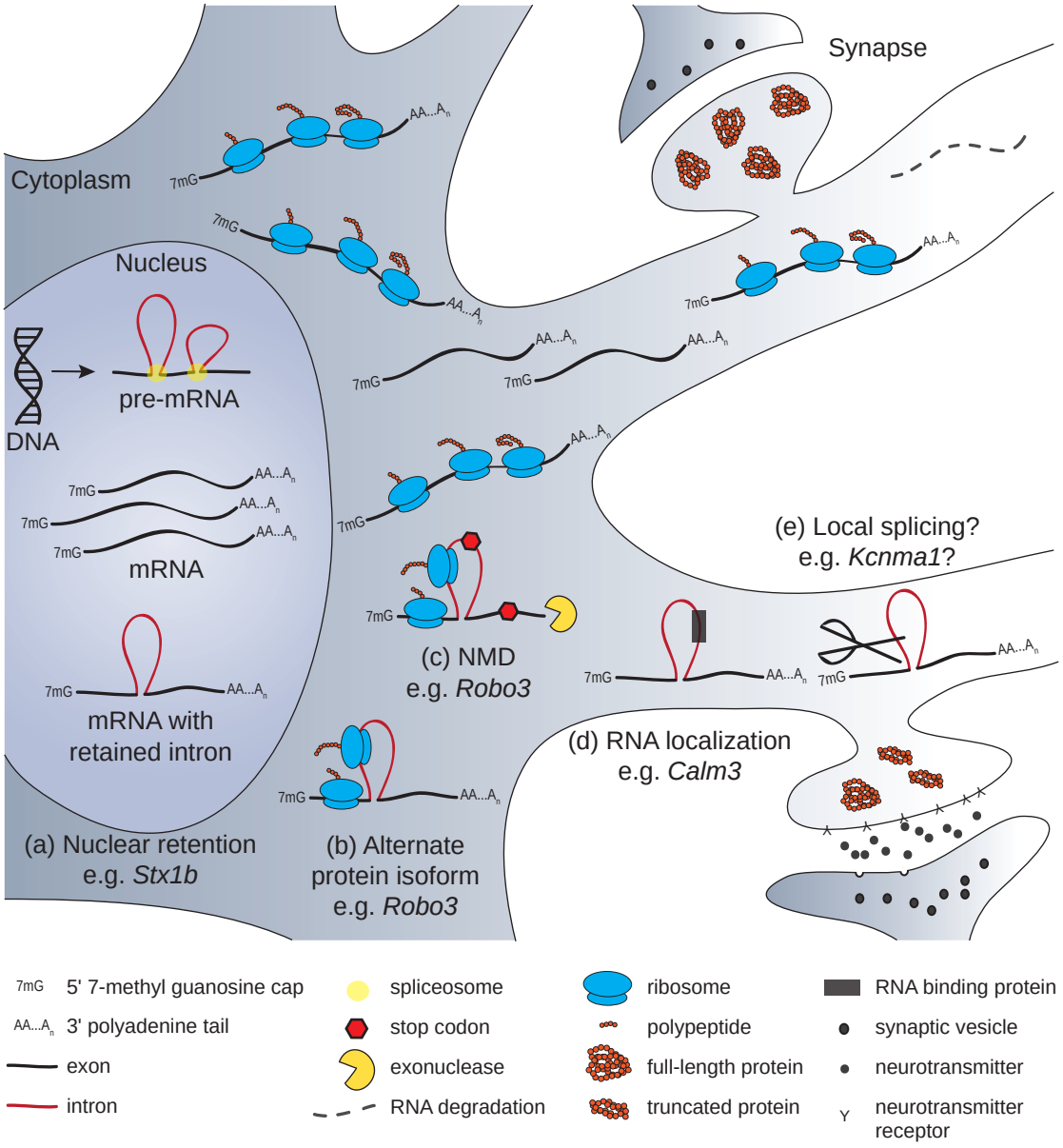


FIGURE I.3: Functional effects of retained introns. Intron retention is a form of alternative splicing. Some transcripts with retained introns are found enriched in the nucleus, for e.g. (a) *Stx1b* [Yap et al., 2012]. In this case, by retaining an intron, steady state abundance of the mature mRNA decreases in the cytoplasm and is thought to be a regulatory mechanism to control protein production. Translation from transcripts with retained introns can alter the protein sequence, for e.g. (b) *Robo3* [Chen et al., 2008]. Intron retention can introduce premature stop codons and render the mRNA sensitive to nonsense-mediated decay, thus reducing mRNA half-life and consequently inhibiting protein expression, for e.g. (c) *Robo3* [Colak et al., 2013]. Some retained introns contain RNA localization signals, for e.g. (d) *Calm3* [Sharangdhar et al., 2017]. A more controversial function for intron retention is localized splicing to allow the production of splice variants on site, for e.g. (e) *Kcnma1* [Bell et al., 2010].

repulsion (Figure I.3 (b)). Moreover, intron retention in Robo3 contributes to the temporal control of protein expression from the appropriate isoform, the mechanism for which is described below. Alternative splicing of Robo3 and its consequences are evolutionarily conserved, having been observed in mouse and chick [Chen et al., 2008, Colak et al., 2013].

As mentioned above for Arc, retained introns often contain in-frame or premature termination codons (PTC) that can target the transcript for NMD [Giorgi et al., 2007]. In addition to its well-known role in degrading erroneous transcripts, the NMD pathway plays a regulatory role in normal protein expression [Bicknell et al., 2012]. Upon translation termination, the NMD machinery degrades mRNAs containing a premature termination codon, thus limiting the number of proteins that can be generated from each transcript. In case of Robo3, protein expression from Robo3.2, the intron-retaining isoform, is controlled spatially and temporally by altering the activity of the NMD pathway in the growing axon [Colak et al., 2013] (Figure I.3 (c)).

Thus, retained introns in cytoplasmic RNAs can generate alternative proteins and alter transcript stability. Additionally, some retained introns contain protein-binding sites that target them to specific subcellular locations (for e.g. Calm3; Figure I.3 (d)). Under special circumstances, an intron may even be spliced in the cytoplasmic compartment. Although our current understanding is that all splicing

occurs in the nucleus, in platelets that lack a nucleus, spliceosomal components, pre-mRNAs, and splicing have been observed [Denis et al., 2005]. In addition, it is known that the stress response mRNA XBP1 contains an intron that is excised in the cytoplasm by a spliceosome-independent mechanism [Back et al., 2006, Hayashi et al., 2007]. Are similar mechanisms of cytoplasmic splicing at work in neurons? Some provocative results [Bell et al., 2010, Glanzer et al., 2005] (Figure I.3 (e)) have suggested that cytoplasmic splicing might be more common in the nervous system where cells send dendrites and axons to locations that are distant from the nucleus and rely on local translation. Although controversial [Steitz et al., 2008], this possibility is tantalizing as it would add another layer of flexibility to the local regulation of gene expression at individual synapses or in growing axons and dendrites. If cytoplasmic splicing does occur in neurons, this would fundamentally transform our current understanding of the physical separation of pre-mRNA processing and mRNA translation in eukaryotic cells.

However, it is challenging to test for splicing activity in neuronal projections without knowing potential targets of splicing. Although previous studies have tried to identify mRNAs with retained introns in dendrites, those experiments involved many rounds of RNA and DNA amplification, which amplifies the signal but can also amplify noise. Also missing from these experiments was evidence for the excised intron. Therefore, we decided to sequence RNAs from neuronal projections to hunt for retained and excised introns. This is a first step toward a

comprehensive catalog of intron sequences that localize to projections. Such a catalog would be a resource for investigating the full breadth of functions served by introns.

Distribution of tRNAs and other short RNAs in neurons

As mentioned above, localized translation requires the presence of translation machinery, which includes ribosomes, tRNAs, and various translation factors. Ribosomes can be actively transported and are clearly visible in electron microscopy images of axons and dendrites [Steward and Falk, 1985, Zelená, 1970]. Not much is known about the distribution of tRNAs; whether they are localized, whether all tRNAs or a subset of tRNAs are localized, and how such localization may occur, are open questions.

Consideration of tRNAs raises an important point. These molecules are distinct from the more commonly studied RNAs in that they are short, they lack a 5′ cap or 3′ polyadenine tail, and are transcribed by RNA polymerase III. For technical reasons, most of RNA localization studies have been focused on long, polyadenylated RNAs (translated by RNA polymerase II). As a consequence, little is known about the localization of nonadenylated RNA, even though these species are diverse and often abundant. Thus, there may be an entire ecosystem of RNA in distal cellular locations awaiting discovery.

mRNAs are distributed in cells by a variety of localization mechanisms. Some mRNAs contain clearly defined localization elements, for e.g., A2RE in MBP mRNA [Munro et al., 1999] and ZBP-1 in β-actin [Hüttelmaier et al., 2005], that enable their interaction with motor proteins for microtubule-mediated localization. Other mRNAs can contain self-destruct signals that are activated or inactivated in a spatially and temporally controlled manner, for e.g. by intron retention in Robo3 mRNA (described above). When the length of an RNA is not constrained, there can be ample room to include sequences that can play a part in RNA localization. However, RNAs with structural and/or functional constraints, such as tRNAs, snRNAs, or miRNAs, might not be amenable to incorporation of extra sequence. A variety of functional, mostly noncoding, RNAs comprise the pool of RNAs of limited size. Most of these have been classified based on their function: miRNA, piRNA, tRFs (tRNA fragments), tRNAs, snRNAs, 5S rRNA, 5.8S rRNA. How are these small RNAs distributed in large cells? Is the subcellular localization of small RNAs dependent on their association with larger protein complexes, or can they be distributed independently, perhaps by diffusion or propulsion by the movement of large organelles and macromolecular complexes?

Small RNAs also show asymmetric distribution in cells of all shapes and sizes. For example, snRNAs are enriched in the nucleus. Although snRNAs shuttle to the cytoplasm for post-transcriptional modifications that are essential to their function, they are imported back into the nucleus by a multi-protein complex

that assembles on processed snRNAs in the cytoplasm and disassembles in the nucleus [Will and Lührmann, 2001]. Once in the nucleus, mature snRNAs associate with spliceosomal proteins and pre-mRNAs. As a result, snRNAs are expected to be relatively depleted from the cytoplasm and even more so from cellular regions far away from the nucleus, such as neuronal projections. On the other hand, 5S and 5.8S rRNAs are deeply embedded in the ribosome and are therefore expected to be enriched at sites of translation in the cytoplasm.

Do tRNAs, their synthetases, and amino acids follow ribosomes too or are they uniformly distributed throughout the cell? How do tRNAs move around in cells? In a recent study (not yet peer reviewed; [Koltun et al., 2019]), as part of a larger experiment to visualize regions of translation activity, the authors transfected primary cortical neurons from mice with fluorescently labeled tRNAs and recorded tRNA mobility. tRNAs could be observed in all parts of the neuron and a subset colocalized with ribosomes. The authors calculated the median velocity of tRNAs to be 0.022 µm/s in the soma and 0.015 µm/s in projections, which they claim is greater than what they would expect by diffusion. In comparison, kinesin-1 family motor proteins that transport organelles, proteins, and RNA, have velocities ranging between ~0.5–1 µm/s in axons [Maday et al., 2014]; live cell imaging of Camk2a mRNA in dendrites of primary rat hippocampal neurons showed oscillatory and directional movement with a transport velocity of 0.04 ± 0.01 µm/s [Rook et al., 2000].

Interestingly, in Koltun et al. [2019], translation inhibition by puromycin, which can interfere with tRNA association with ribosomes, made the tRNA signal less punctate or more diffuse. Perhaps tRNA localization depends on a combination of diffusion and capture by ribosomes or tRNA synthetases.

One goal of this dissertation is a comprehensive exploration of projection-localized RNAs, an exploration that is not limited to polyadenylated species and does not discriminate against smaller RNAs. The hope is that casting this wide net will lead to the identification of new classes of localized RNAs. These data can used to ask if a logic exists for the localization of RNA types, and to provide clues for mechanisms of localization.

Challenges to studying cytoplasmic introns and other RNAs

The preceding sections have reviewed several topics that converge on the need to survey, systematically and comprehensively, the distribution of RNA molecules in distinct subcellular compartments. With regard to introns, some are retained in cytoplasmic transcripts and are probably more prevalent in neurons [Braunschweig et al., 2014, Buckley et al., 2011]. A wide range of functions, from protein expression regulation to RNA localization, can potentially be attributed to retained introns [Chen et al., 2008, Colak et al., 2013, Sharangdhar et al., 2017]. Excised introns in the cytoplasm could be a source of a functional noncoding RNA

[Talhouarne and Gall, 2018]. These possibilities expand when looking beyond introns to other small RNAs. However, studying the cytoplasmic localization of RNA sequences *en masse* can be challenging for several reasons, some of which are described below.

Subcellular fractionation to separate nuclei from cytoplasm or cellular projections

Almost all types of RNAs can now be subjected to high-throughput sequencing. To study the full range of RNAs in a specific cellular compartment, RNA must be isolated specifically from that compartment. This can be accomplished in several ways. If the goal is to separate nuclei from the cytoplasm, the classical approach involves biochemical fractionation of cells or tissues. First, the samples are lysed with a hypotonic solution and mild detergent. The lysis conditions are optimized to be strong enough to burst the cell membrane but gentle enough to keep the nuclear membrane intact. The lysate is subjected to centrifugation which causes the nuclei to pellet. The supernatant or “cleared” cytoplasmic lysate can then be collected for further experimentation.

While this method has historically proven quite useful, especially when experiments require large quantities of nuclear and/or cytoplasmic lysate, the purity of resulting cellular fractions can vary depending on the lysis conditions, handling of the samples, and the morphological complexity of the cells. Fractionation of

neuronal cells (for example, separating projections from soma) by this method is unsuitable because of the loss of projections during cell dissociation and lysis. Mild lysis conditions can result in cellular projections pelleting with the nuclei, harsh lysis could cause the nuclear membrane to rupture and blur the separation of nucleo-cytoplasmic fractions. This becomes especially problematic when identifying intron sequences in the cytoplasm because even minimal contamination from the nucleus could yield false positives.

Particularly for neurons or cells with extended morphologies, other methods to isolate projections include microdissection by laser capture [Zivraj et al., 2010] or microsuction of cellular contents with a pipette [Buckley et al., 2011]. When handled carefully, these methods can provide highly pure lysate, but the amount of lysate and RNA yield is often low (a few picograms of RNA per cellular fraction). Preparing cDNA libraries from picograms of RNA might necessitate many rounds of RNA and/or DNA amplification, which can amplify signal but also introduce noise or technical artifacts.

Another approach to separate neuronal projections from soma that has proven reasonably successful is by culturing neurons on platforms that provide a physical separation between cellular projections and soma. Microfluidic chambers with channels that are up to 3 µm in diameter and can be over a 100 µm long have been used to isolate dendrites/axons of a variety of neurons [Taylor et al., 2005].

A similar principle underlies semipermeable membranes, again with up to 3 µm holes in them [Poon et al., 2006] that provide physical separation between cellular compartments. A major advantage of this approach is that it can be scaled up relatively easily to acquire greater quantities of lysate that can yield higher quantities (several micrograms) of RNA. For experiments in Chapters II and III, we used semipermeable membranes with 1 µm holes to separate neuronal projections from somata.

Capturing a fair representation of localized RNA

For technical reasons, many of the currently available datasets examining projection-localized RNAs are enriched for polyadenylated RNAs that are ~200 nt or longer. If the goal is to study mRNA localization, then there are clear advantages to enriching for long polyadenylated RNAs: one, selecting polyadenylated RNAs excludes the very abundant rRNA from cDNA libraries, and two, selecting RNAs greater than ~200 nt excludes other abundant classes of RNA, such as tRNAs and snRNAs. Although such datasets are enriched for mRNAs, they would omit nonadenylated RNAs (for example, excised introns), circular RNAs, and the whole set of smaller RNAs.

To capture a more complete representation of all localized RNAs without overwhelming the data with rRNA reads, RNAseq libraries can be prepared after

rRNA depletion (commercially available kits for commonly used model organisms make this task easy). An initial step to prepare libraries for RNAseq involves converting the RNA to cDNA by reverse transcription. Several techniques are available to prime the reverse transcription reaction. The more commonly used approach is to prime with an oligo-dT primer, but that will only capture polyadenylated RNAs or transcripts with internal string of adenines. An alternative is to use random hexamers for priming the reverse transcription reaction [Zhang et al., 2012]. Another approach involves ligating an adaptor to the 3′ ends of RNAs that can serve as a hybridization target for the primer. The adaptor ligation approach works well both long and short RNAs [Heyer et al., 2015]. Alternatively, an RNA polyA polymerase could be used to add a polyA tail to RNA fragments that can then be primed for reverse transcription with an oligo-dT primer [Zhu et al., 2001]. This method works especially well in sequencing intact, shorter RNAs.

To capture long RNAs, RNAseq libraries in Chapter II were prepared after rRNA depletion using the random hexamer approach [Zhang et al., 2012], whereas small RNAseq libraries in Chapter III were prepared using the polyadenylation method [Zhu et al., 2001].

Genome annotation of introns, exons, and repetitive RNAs

Genome annotation is essential to characterize the types of RNAs represented in RNAseq data. When studying alternative splicing events, such as intron retention, the gene structure must be annotated too. The exon/intron structure of genes is not easy to recognize by sequence alone and becomes even more challenging in the case of alternative splicing. Although spliceosomal introns contain consensus sequences at their 5′ and 3′ ends, these are short (~6 nt) and still poorly characterized for mammalian genomes. If the splice site consensus were a conserved hexamer, we could expect to find it every 4096 nt (4^6) in a random genomic sequence that has uniform nucleotide composition. Thus in a genome with ~3 billion base pairs, a specific hexanucleotide sequence would be expected to appear ~1.4 million times by chance. Of course, the genome is more complex than that and we already know a lot about protein sequences and RNA structure. Thankfully, the genomes of many commonly used organisms have been annotated, curated, and are publicly available [O'Leary et al., 2016, Zerbino et al., 2018]. However, these annotations are not perfect or complete. When investigating retained introns from RNAseq data, false positives could easily creep in where exon/intron boundaries are poorly defined or alternative splicing is incompletely understood.

The annotation problem also applies to classes of RNAs that are pseudogenized

or repeated in the genome. This can be confounded further for noncoding RNAs, where there is no open reading frame to help define gene boundaries. In addition, just because something is annotated and assigned a name, does not mean it is functional. Many pseudogenes are expressed and annotated, although many of them might be inert.

Dissertation overview

Chapter II focuses on a systematic investigation of intron sequences in neuronal projections.

Chapter III provides evidence that small RNAs are abundant in neuronal projections, and largely comprise tRNAs and tRNA fragments.

Chapter IV discusses the impact of these results, the limitations of the experiments, and new questions that have emerged from the research presented in this dissertation.

CHAPTER II

Free circular introns with an unusual branchpoint in neuronal projections

Harleen Saini[1,2], Alicia A. Bicknell[1†], Sean R. Eddy[2,3], Melissa J. Moore[1†]

[1]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, United States

[2]Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, Cambridge, United States

[3]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, United States

[†] Present address: Moderna Therapeutics, Cambridge, United States

Contribution summary

H.S. and M.J.M. conceived the project. A.A.B. performed ribosome profiling and polyA+ RNAseq; all other experiments were done by H.S. H.S. analyzed the data under S.R.E. and M.J.M.s supervision. H.S., S.R.E, and M.J.M. wrote the final document.

Abstract

The polarized structure of axons and dendrites in neuronal cells depends in part on RNA localization. Previous studies have looked at which polyadenylated RNAs are enriched in neuronal projections or at synapses, but less is known about the distribution of non-adenylated RNAs. By physically dissecting projections from cell bodies of primary rat hippocampal neurons and sequencing total RNA, we found an unexpected set of free circular introns with a non-canonical branchpoint enriched in neuronal projections. These introns appear to be tailless lariats that escape debranching. They lack ribosome occupancy, sequence conservation, and known localization signals, and their function, if any, is not known. Nonetheless, their enrichment in projections has important implications for our understanding of the mechanisms by which RNAs reach distal compartments of asymmetric cells.

Introduction

In polarized cells, such as neurons and oocytes, RNA localization to distinct subcellular compartments is important for spatial control of protein expression [Holt and Bullock, 2009]. Known mechanisms for asymmetric distribution of RNA include active transport [e.g., *Actb*, Ross et al., 1997], spatially restricted capture by an anchor [e.g., *Nanos*, Forrest and Gavis, 2003], and control of RNA degradation [e.g., *Hsp83*, Bashirullah et al., 2001]. In some cases, RNA localization depends on splicing. For example, a detained intron (i.e., an intron with regulated post-transcriptional splicing, as opposed to constitutive co-transcriptional splicing) restricts *Srsf5* mRNA export from the nucleus [Boutz et al., 2015], and a retained intron (i.e., an alternative unspliced isoform) promotes dendritic localization of *Calm3* [Sharangdhar et al., 2017]. The *Robo3* gene, which is important for commissural axon development in mice, expresses both a fully spliced mRNA and another retaining intron 26, and these isoforms encode different proteins that have opposing functions in axon guidance [Chen et al., 2008]. Spatial and temporal control of protein expression from the intron-retaining *Robo3* isoform depends on its susceptibility to nonsense-mediated decay due to the presence of a premature termination codon in the retained intron [Colak et al., 2013].

Provocatively, some retained introns have been proposed to undergo splicing in

dendrites [Glanzer et al., 2005]. For example, an intron in the calcium-activated potassium channel *Kcnma1* was reported to undergo splicing in dendrites of rat hippocampal neurons [Bell et al., 2010], and this was suggested to be a mechanism for locally tailoring calcium-activated potassium currents. Because pre-mRNA splicing by the spliceosome is generally thought to be restricted to the nucleus [Steitz et al., 2008], this proposal has been controversial, and it has not yet been independently confirmed.

The interplay between intron retention and neuronal RNA localization has been studied in several individual cases [Bell et al., 2010, Buckley et al., 2011, Chen et al., 2008, Khaladkar et al., 2013, Ortiz et al., 2017, Sharangdhar et al., 2017]. In this work, our aim was to systematically identify localized RNAs in primary rat hippocampal neurons by sequencing total RNA (rRNA depleted) as opposed to polyadenylated (polyA+) RNA, with a particular focus on the repertoire of projection-localized introns (both retained and excised). Our analyses identify hundreds of transcripts with retained introns. Unexpectedly, we also found a set of free circular introns localized to distal neuronal projections.

Results

Experimental Design and Validation

To physically separate cellular projections from cell bodies, we cultured dissociated primary rat hippocampal cells on membranes with 1 µm diameter pores [Poon et al., 2006]. These cultures are a mixture of neuronal and glial cells; we add a DNA replication inhibitor to block cell division and prevent dividing glia from overgrowing post-mitotic neurons. We refer to the projections as "neuro-glial" projections because both neuronal (Map2-immunopositive) and non-neuronal (Gfap/Vimentin-immunopositive) projections extend through the pores and continue growing on the underside of the membrane, whereas cell bodies and nuclei are restricted to the top surface (Figures II.1 and II.2). Lysates prepared by scraping the underside are highly enriched for projections ("projection" samples), while lysates prepared from the top surface comprise whole cells with nuclei and projections ("whole cell" samples).

To capture both adenylated (polyA+) and non-adenylated (polyA-) long RNAs in our lysates, we prepared rRNA-depleted total RNAseq libraries (mean insert size ~200 nt) from five biological replicates (10 samples total). The RNAseq libraries were subjected to paired-end sequencing (100-125 nt reads) on the Illumina platform to obtain 30-80 million mate pairs per sample.

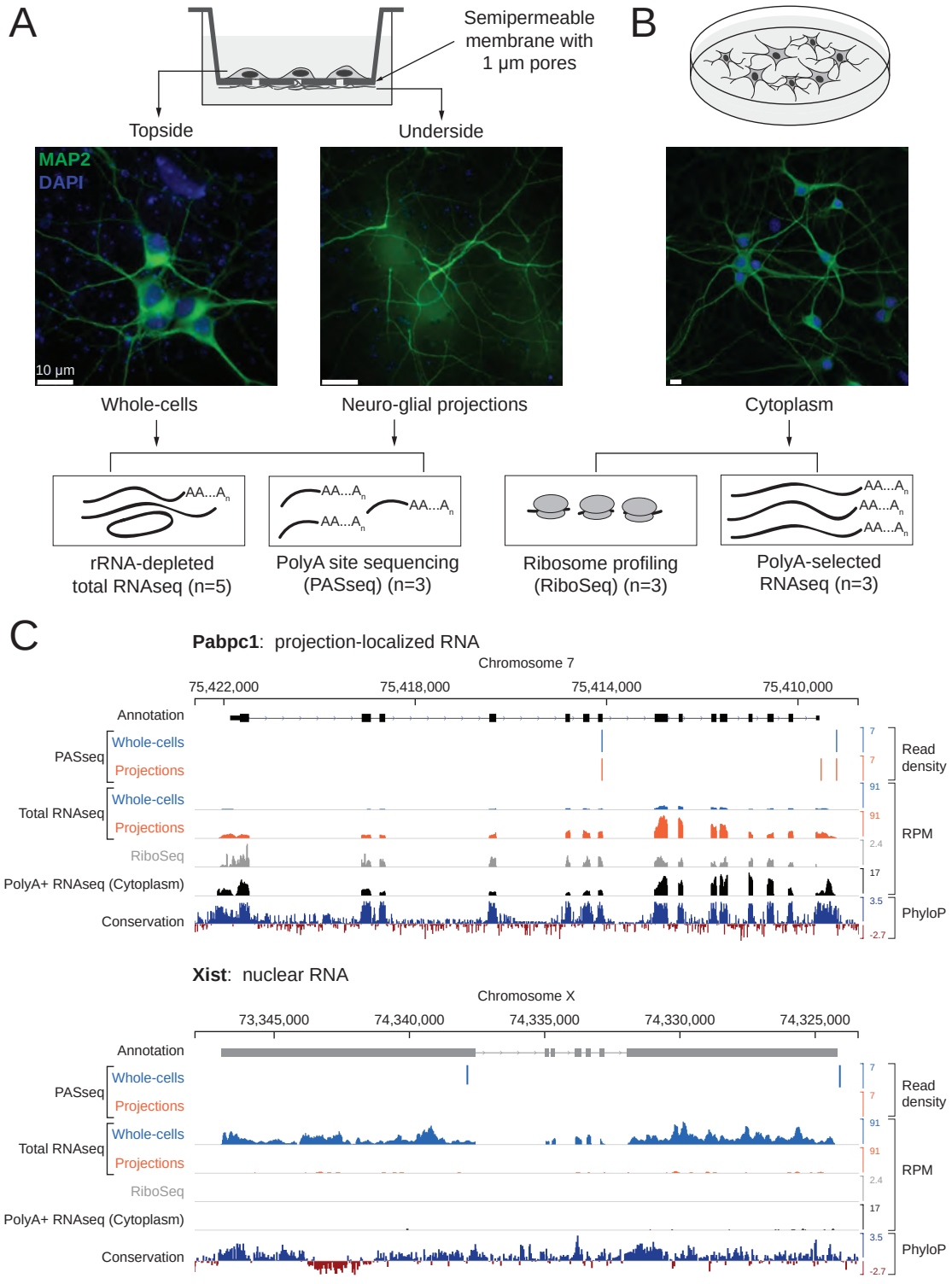


FIGURE II.1: Experimental design and data validation (legend continues on the next page).

FIGURE II.1: (A) Imaging of MAP2 protein immunostaining (neuronal marker, green) and DAPI fluorescence (nuclear marker, blue) confirms that the bottom surface of neuronal cultures on a semipermeable membrane (cartoon, top) consists only of neuro-glial projections. Total RNAseq and polyA site RNAseq (PASseq) datasets were generated from the top surface ("whole cell") and bottom surface ("projection") lysates. (B) Standard plate cultures, fractionated to remove nuclei, were used to prepare ribosome profiling and cytoplasmic polyA+ RNAseq datasets. (C) Genome browser plots of read densities (sum of three replicates) and sequence conservation [PhyloP scores on 20 aligned vertebrate genomes, Pollard et al., 2010] for a projection-localized mRNA and a nuclear noncoding RNA (RPM= Reads per million mapped).

We generated additional datasets to help interpret projection and whole cell RNAseq data. To distinguish retained introns from intronic polyadenylation (polyA) sites, we prepared polyA-site sequencing (PASseq) libraries [Ashar-Patel et al., 2017] from a subset of samples (three biological replicates each, 6 samples total). To identify coding exons, we generated ribosome profiling [Ricci et al., 2014] and polyA+ selected RNAseq libraries [Heyer et al., 2015] from the purified cytoplasmic (nuclei-depleted) fraction of primary rat hippocampal neurons cultured on plates (three biological replicates) (Figure II.1*A, B*).

For preliminary characterization of these data, we aligned RNAseq reads to the rat genome (Ensembl release 81, Rnor_6.0) [Zerbino et al., 2018] (Figure II.16). Ribosomal RNAs and small (<150 nt) noncoding RNAs (e.g., snRNAs, tRNAs) were depleted effectively. Larger abundant nonpolyadenylated noncoding RNAs were observed as expected, such as SRP (or 7SL, the RNA component of the signal recognition particle) and RMRP (the RNA component of mitochondrial RNase P); both were more prevalent in projection than in whole cell libraries. Reads mapping to the mitochondrial genome were also highly abundant and

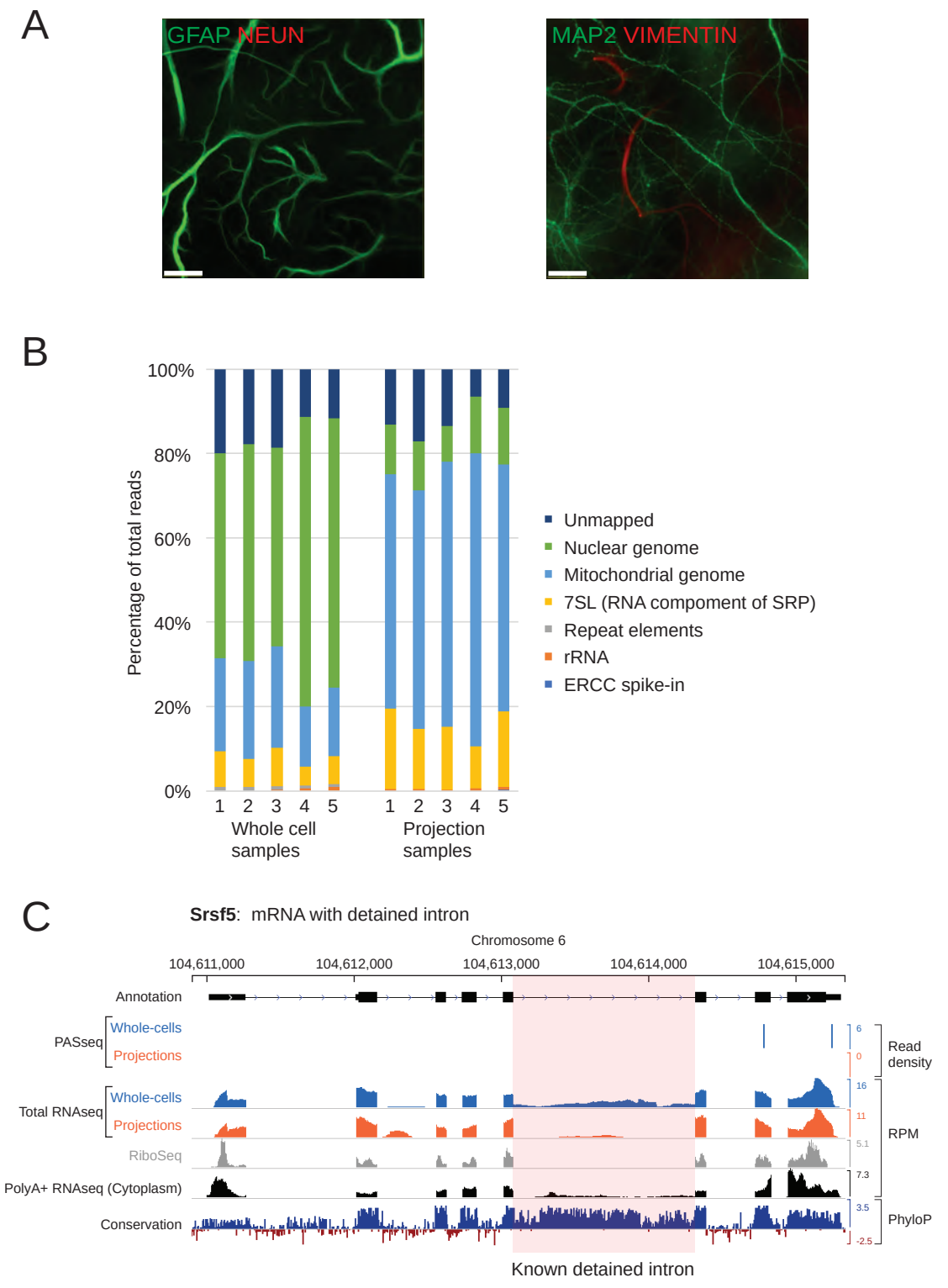


FIGURE II.2: Separation of neuro-glial projections from cell bodies (legend continues on the next page).

FIGURE II.2: (A) GFAP and Vimentin (VIM) immunopositive projections pass through the semipermeable membrane and are visible on the bottom surface by immunofluorescence imaging. (B) Percentage of total RNAseq reads from whole-cells and projections aligning to various genomic regions. A majority of reads from projection RNAseq data align to the mitochondrial genome. (C) Genome browser plot of read densities and sequence conservation for *Srsf5*, an mRNA with a known detained intron (highlighted in pink).

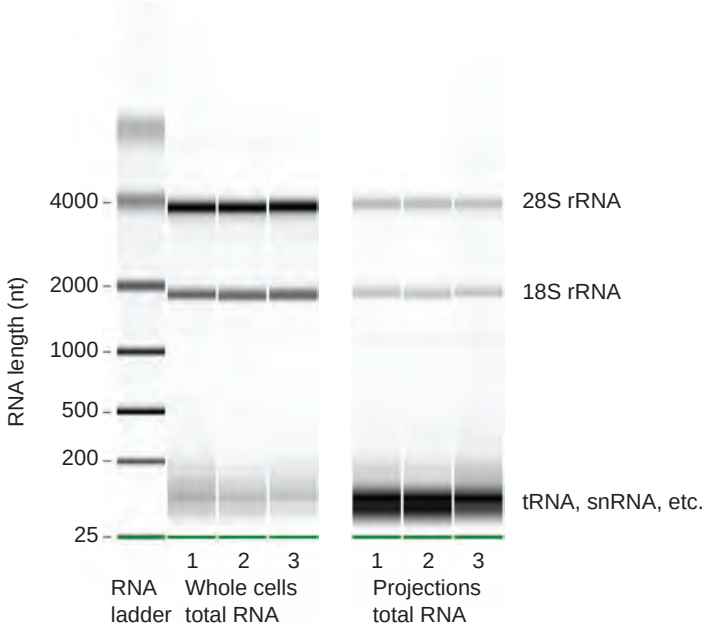


FIGURE II.3: Microcapillary electrophoresis of total RNA (Bioanalyzer RNA pico assay) from whole cells and projections of three biological replicates. Only RNAs > ~150 nt were sequenced for this study. In projections, the band at ~100 nt was striking. By sequencing RNAs within 20-150 nt size range, we found that they were mostly full length and fragmented tRNAs (Chapter III).

enriched in projection libraries, consistent with the abundance of mitochondria in neurons and neuronal projections [Palay, 1956] (Figure II.2). To minimize spurious genome alignments due to expected abundant species, we filtered out reads mapping entirely to rRNA, SRP, mitochondrial genome, and repeat elements cataloged by RepeatMasker [Jurka, 2000] before moving on to other analyses described below.

To qualitatively assess our success in separating projections from whole cells, we aligned filtered reads to the rat genome with TopHat2 [Kim et al., 2013] and visualized read density tracks of known nuclear and projection localized RNAs on a rat genome browser [Integrative Genomics Viewer, Robinson et al., 2011]. Known nuclear-localized RNAs such as the noncoding RNA *Xist* [Brown et al., 1992] and *Srsf5* mRNA (which contains a detained intron) [Boutz et al., 2015] were depleted from projections relative to whole cells (Figure II.1*C* and II.2). Conversely, the known projection-localized *Pabpc1* mRNA [Poon et al., 2006] was enriched in projection data.

RNAs enriched in projections

To comprehensively and quantitatively evaluate how well our datasets distinguish known localized RNAs, we employed Kallisto and Sleuth [Bray et al., 2016, Pimentel et al., 2017] for differential expression analysis of annotated RNA transcripts in projections versus whole cells. RNA abundances (in TPM, Transcripts Per Million) of biological replicates were well correlated (Spearman's correlation coefficient $\geq$ 0.83 in projection samples, $\geq$ 0.88 in whole cell) (Figure II.5), but comparisons between projection and whole cell datasets showed substantial differences (Figure II.4*A*). As expected, known nuclear noncoding RNAs, including *Xist*, *Malat1*, *Meg3*, snoRNAs, and scaRNAs, were among the 1,486 genes significantly depleted (> 1.5 fold and q-value < 0.01) from projections.

In contrast, 1,440 transcripts were significantly enriched in projections, including known projection-localized mRNAs such as *Pabpc1*, *Map2*, *Dlg4* (neuronal), and *Gfap* (glial) [Cajigas et al., 2012, Garner et al., 1988, Poon et al., 2006, Sarthy et al., 1989]. Gene ontology analysis showed that the set of projection-enriched mRNAs were significantly enriched for genes involved in mitochondrial functions (cytochrome-c oxidase activity) as well as nearly the entire set of ribosomal protein encoding mRNAs (RP mRNAs) (Figure II.6). Indeed, seventy annotated RP mRNA isoforms were enriched more than two-fold in projections (black dots in Figure II.4*A*). RP mRNA enrichment in projections is puzzling because ribosomal proteins are imported into the nucleus for ribosome assembly but their enrichment in distal cellular locations has been observed consistently in other RNA localization studies (for e.g., most recently, Shigeoka et al. [2018] studied RP mRNA enrichment in *Xenopus* retinal cell axons), and we discuss it later.

To test the accuracy of our RNAseq abundance measurements, we performed single-molecule fluorescence in situ hybridization (smFISH) on example RNAs. Using exon-hybridizing RNAScope probe sets, we probed for six different mRNAs spanning a range of TPM values and projection:whole cell ratios (*Polr2a*, 2:4; *Ppib*, 43:50; *Ubc*, 162:146; *Pabpc1*, 40:12; *Srsf5*, 18:37; *Rpl4*, 297:96) (Figure II.4*B-G*). Consistent with their RNAseq projection:whole cell TPM ratios, *Polr2a*, *Ppib*, and *Ubc* mRNAs exhibited predominantly cytoplasmic localization, with a strong gradient in spot density highest in cell bodies tapering off into the

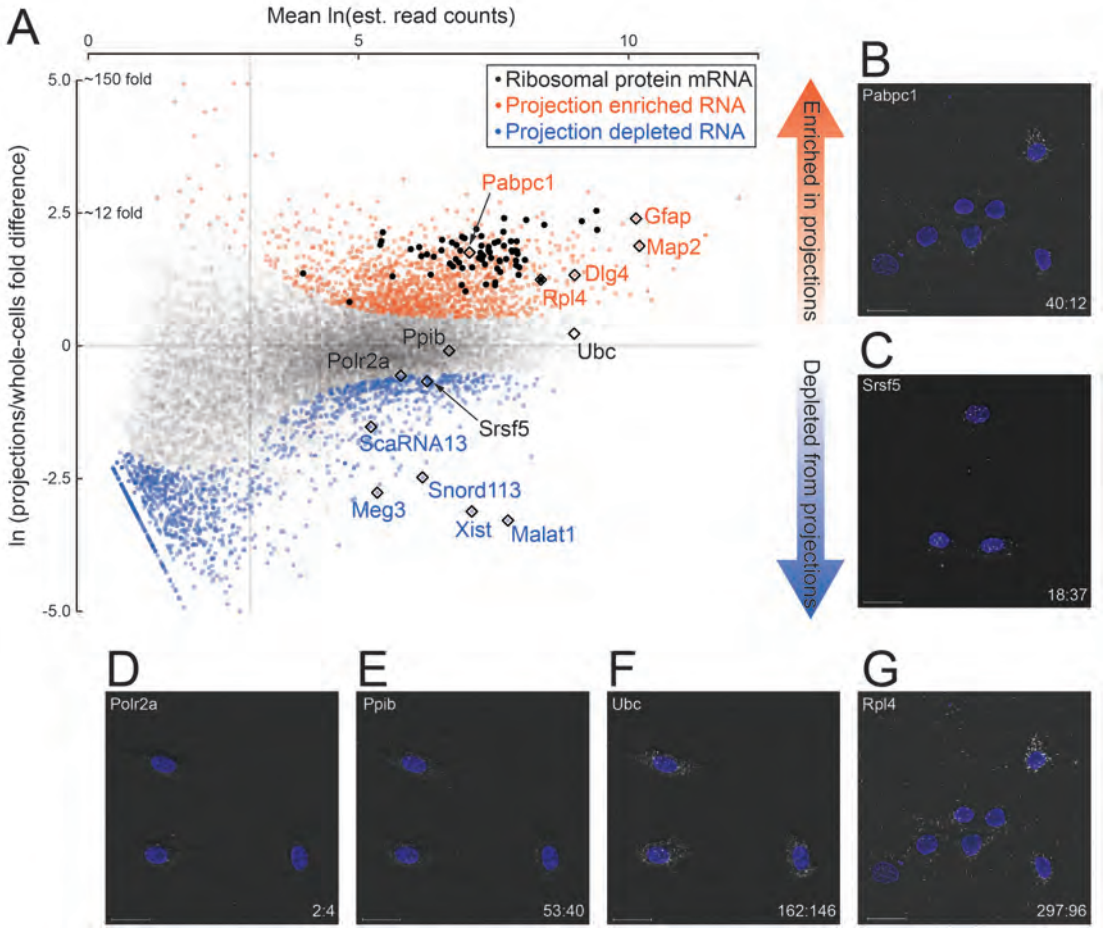


FIGURE II.4: Quantitative analysis of RNA localization. (A) Scatter plot comparing log mean read counts and log fold difference in projections versus whole cells, for 19,815 RNA transcripts with non-zero read counts. 1,440 (orange dots) are significantly enriched (q-value < 0.01 and fold-change > 1.5) in projections, 1,486 (blue dots) are enriched in whole cells, and 16,899 (gray dots) show no significant enrichment in either sample. Ribosomal protein encoding RNAs are shown as black dots; rhombi enclose labeled RNAs. (B-G) smFISH validation of examples of projection-localized (B) versus projection-depleted (C), and low (D), medium (E), high (F), and higher (G) abundance mRNAs. Upper left corner shows gene name, lower right corner shows projection:whole cell TPM. Scale bars = 20µm.

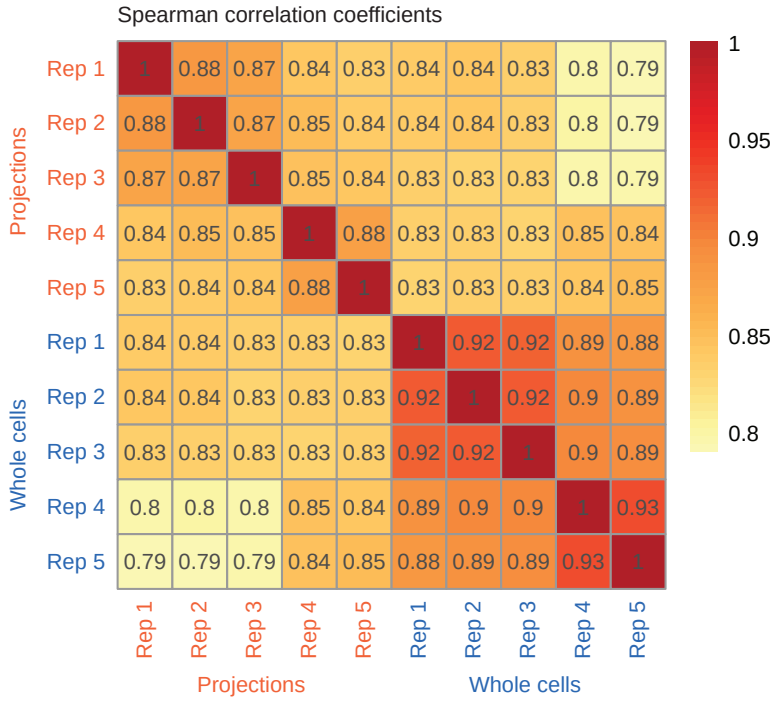


FIGURE II.5: Correlation between biological replicates (TPM values) of whole cell and projection RNAseq datasets. The color gradient corresponds to the value of Spearman's correlation coefficient shown in each cell of the heatmap.

projections. *Srsf5* mRNA was almost entirely confined to the cell bodies, and consistent with the presence of a detained intron [Boutz et al., 2015], *Srsf5* mRNA spots were also visible in the nucleus. In contrast, *Pabpc1* and *Rpl4* mRNA spots exhibited a nearly uniform distribution throughout cell bodies and projections, with relatively few discernible spots in nuclei.

We concluded from these and other analyses that our RNAseq datasets reliably detect and quantify projection-localized RNAs in primary rat hippocampal cultures. We then turned to our main interest in localization of intron sequences.

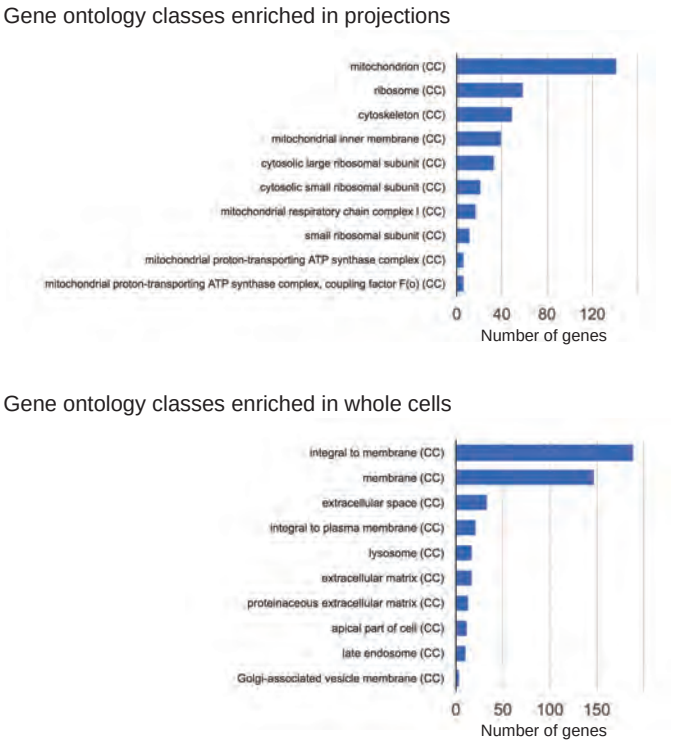


FIGURE II.6: Gene ontology classes significantly represented by genes enriched in projections or whole cells. CC = Cellular Component sub-ontology classification.

Intron regions enriched in projections

Because of alternative splicing, transcribed genomic regions cannot be easily separated into introns and exons. To facilitate a comprehensive analysis, we define an "intron region" as a genomic interval that is annotated as intronic (and not exonic) in all annotated transcript isoforms that span it (Ensembl release 81, Rnor_6.0, annotation downloaded on July 24, 2015) (Figures II.7*A* and II.17). Out of a total of 190,180 such intron regions, we considered 57,432 to have reliable coverage (at least one read in each of the five biological replicates, with mean read density $>$ 0.005 mapped reads/intron region length) in the whole cell

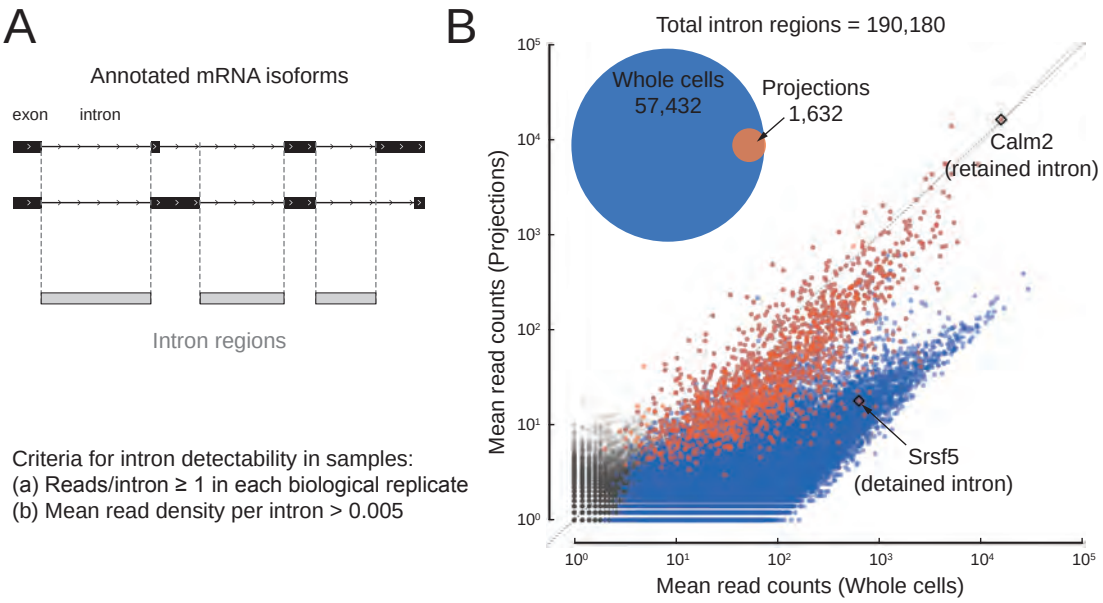


FIGURE II.7: A subset of introns localize to projections. (A) Cartoon illustrating how we define intron regions as the intersection of all annotated introns. For two mock alternatively spliced isoforms, black rectangles are exons, lines are introns; vertical dotted lines mark the intron region boundaries. (B) Scatter plot comparing mean mapped reads per intron region (+1 pseudocount from five biological replicates of projections versus whole cell RNAseq data. Rhombi enclose labeled RNAs. Inset: Venn diagram showing the number of intron regions that we consider to show reliable read coverage (at least one read in each replicate and mean read density $>$ 0.005 reads/nucleotide).

libraries, but only 1,632 met these criteria in projections (Figure II.7*B* - inset). For the 33 intron regions that we considered reliably covered in projections but not in whole cells, individual examination showed that all had coverage in the whole cell libraries but had just missed the cut.

Introns are expected to be spliced and degraded in the nucleus and thus strongly depleted from projections, but a scatter plot of intron region abundance in projections versus whole cells (Figure II.7*B*) shows a bimodal distribution. Intron regions that show coverage in projections define a subpopulation that

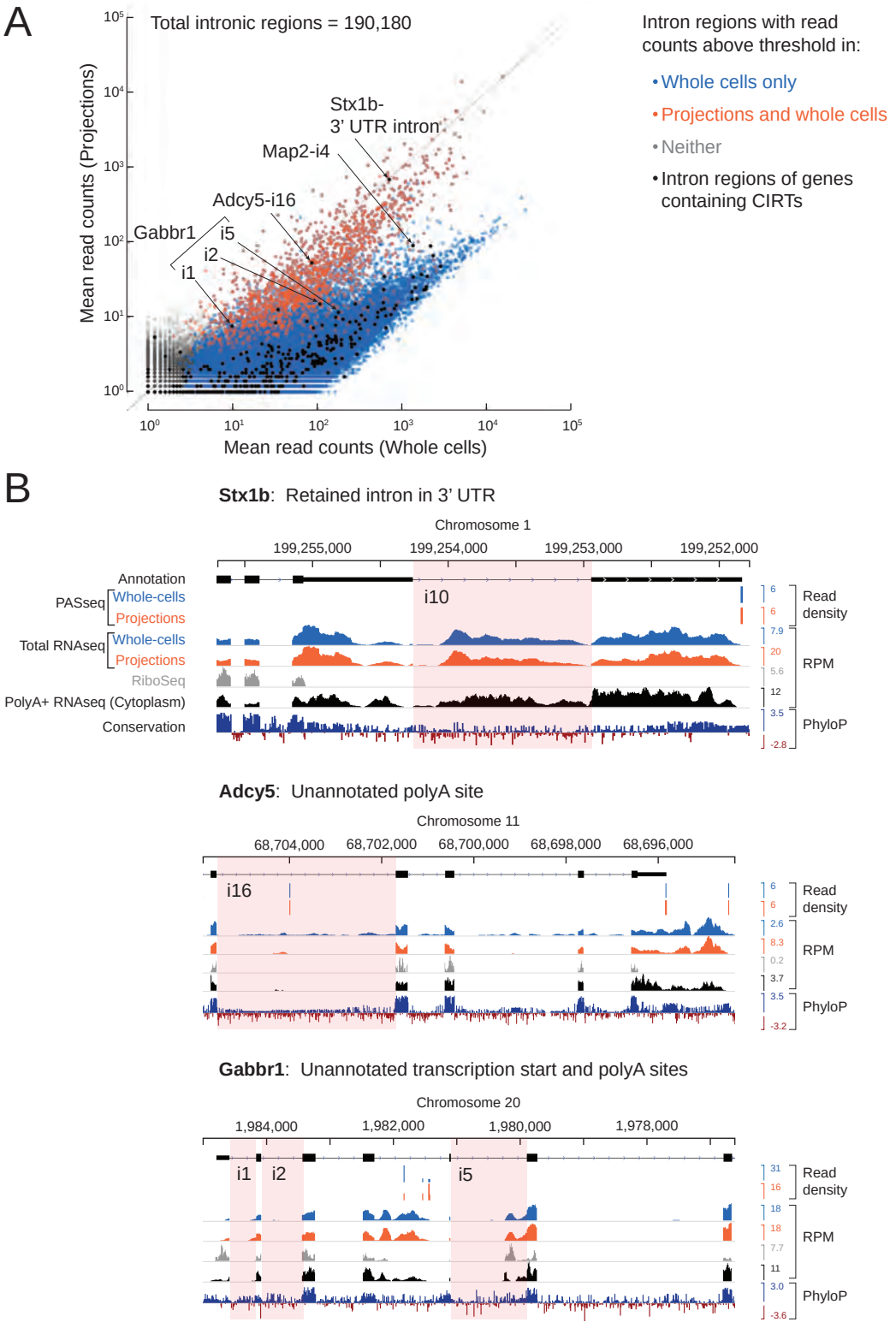


FIGURE II.8: Comparison with previously reported retained introns in dendrites (legend continues on the next page).

FIGURE II.8: Comparison with previously reported retained introns in dendrites. (A) In Buckley et al. [2011], 27 genes with cytoplasmic intron retaining sequences (CIRTs) were reported. All introns (n=436, black dots) of genes containing CIRTs (n=25) are overlaid on the scatter plot from Figure II.1. Six intron regions pass the detection threshold in our projection RNAseq data and are labeled. (B) Genome browser plots for introns labeled in (A) with the relevant intron highlighted in pink. A retained intron in the 3′UTR of *Stx1b* was also observed by Yap et al. [2012]. *Gabbr1* has multiple annotated isoforms, of which only one is shown. RNAseq reads on *Map2-i4* (not shown here) come from an unannotated exon.

has similar read coverage in projections and whole cells. This population is interesting because of the restrictive way that we define "intron regions": no annotated transcript isoform of a given gene shows these regions as exonic, but their abundance in projections suggests that they are either unspliced, excised but stable, or independent transcripts. They include, for example, known (but unannotated) neuron-specific retained introns, such as in *Calm2* [homolog of mouse *Calm3*, Sharangdhar et al., 2017].

As we looked at randomly selected examples of the 1,632 intron regions in the rat genome browser, we found, unsurprisingly, that many cases simply represented an unannotated alternative splicing event (i.e., alternative 5′ or 3′ splice site) or an unannotated transcription start site (TSS) or polyadenylation site (PAS) within an annotated intron. We found it useful to distinguish them into classes depending on the presence or absence of reads spanning the unspliced 5′ or 3′ splice sites (5′ exon-intron and intron-3′ exon reads, EI and IE) and spliced exon-exon junctions (EE) (Figures II.9*A, B* and II.10).

Intron regions with high EI reads but few IE reads (Figure II.9*B-(i) upper left*;

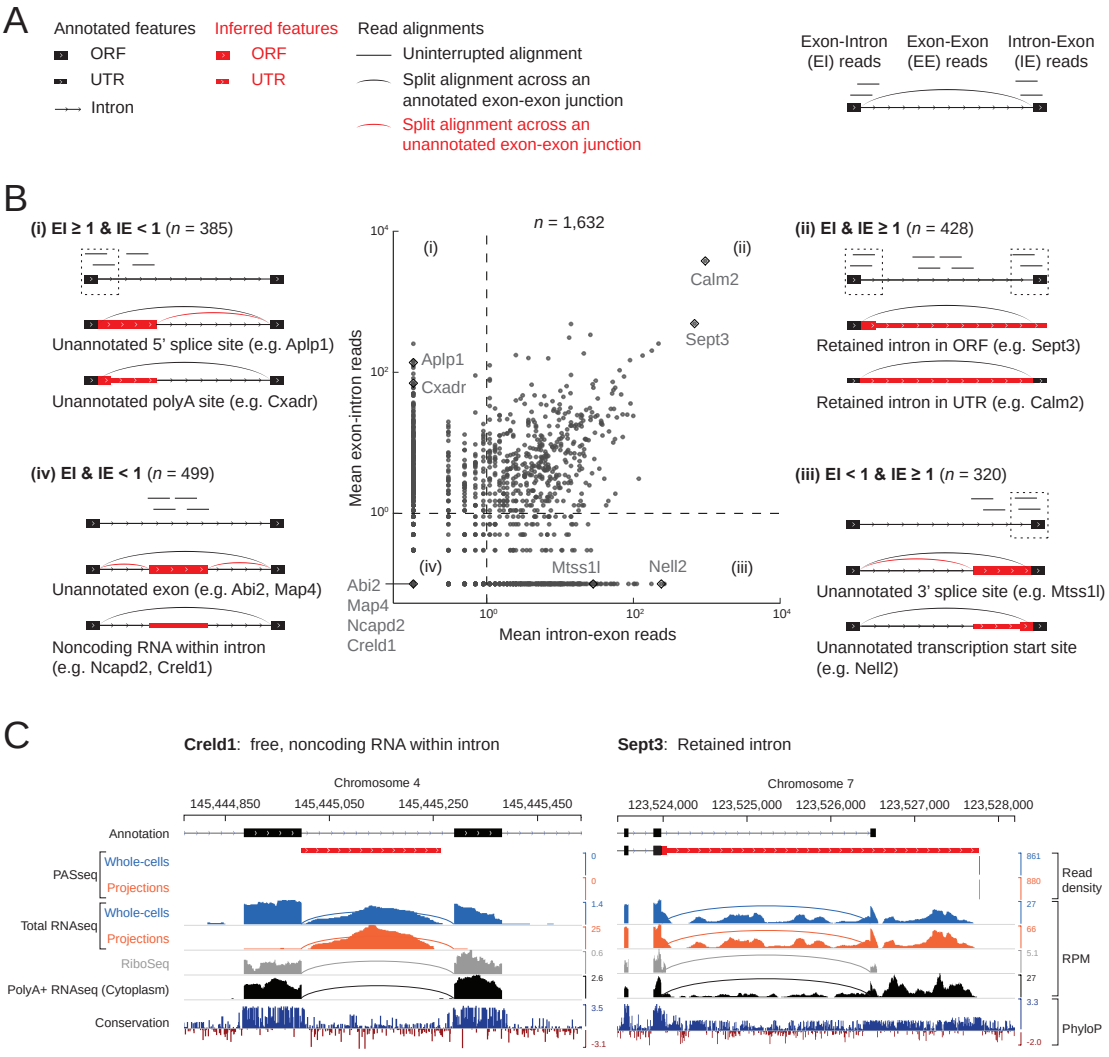


FIGURE II.9: Classification of 1,632 projection-localized intron regions. (A) Description of symbols used to show annotated gene structure and types of read alignments. Rectangles represent annotated (black or inferred (or unannotated, red) coding (ORF, Open Reading Frame) and noncoding (UTR, UnTranslated Region) exons. Lines represent introns and arrows point in the 5′ to 3′ direction of the transcript. Uninterrupted read alignments are shown as lines whereas arcs depict split reads connecting spliced exon-exon junctions. (B) Scatter plot of mean read coverage (+0.1 pseudocount) of 50 nt exon-intron (EI) versus intron-exon (IE) boundary regions in projection RNAseq samples. Dashed lines indicate thresholds set to EI=1 and IE=1 that divide introns into four quadrants, i-iv. Representative sketches depicting the situations within each quadrant are shown. *n* shows the number of intron regions in each quadrant. (C) Genome browser views of read coverage (sum of three biological replicates) and phyloP conservation for a free (Creld1-intron 5) and retained intron (Sept3 - 3′ terminal intron). RPM = reads per million mapped.

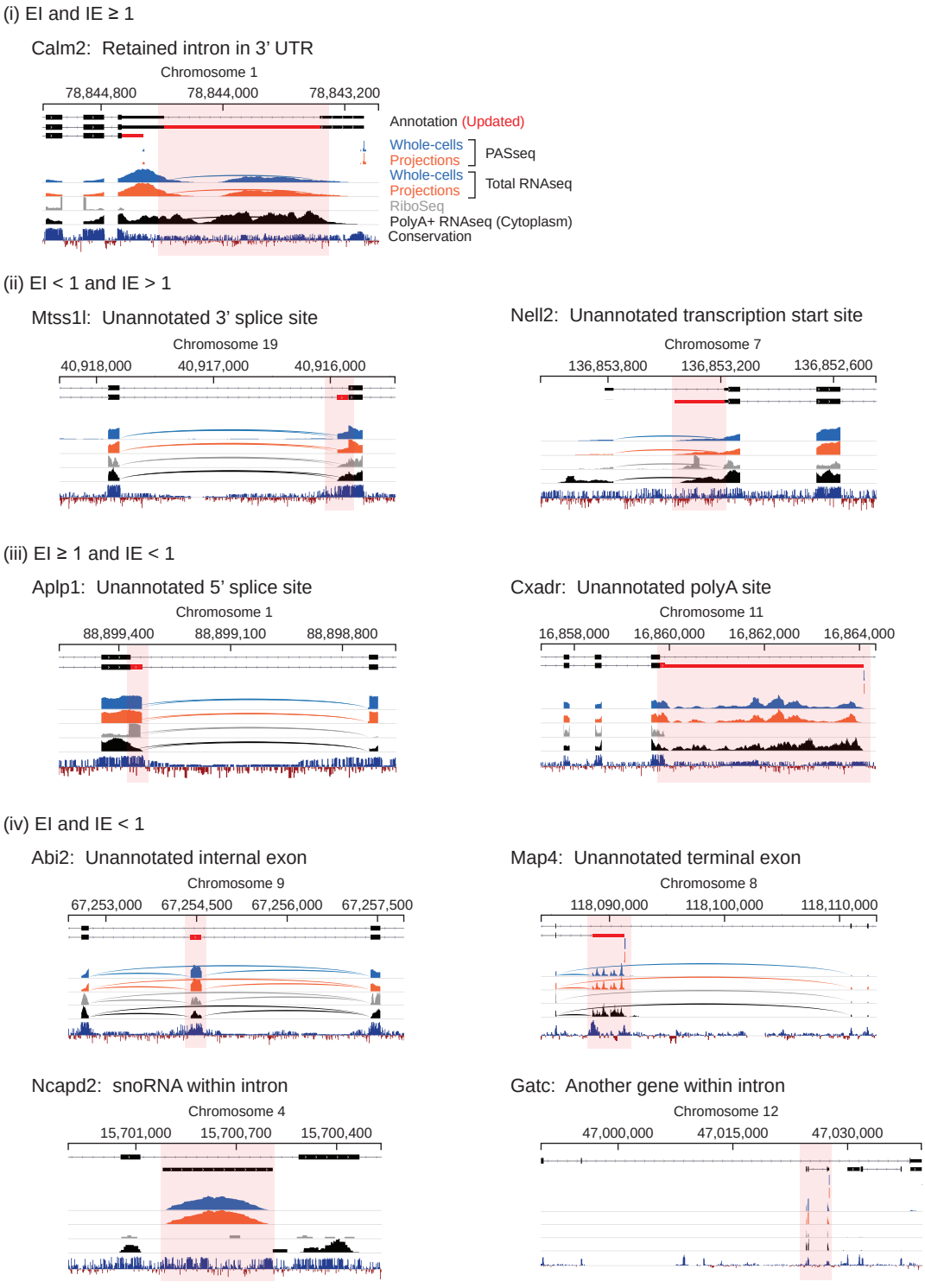


FIGURE II.10: Genome browser views of introns labeled in Figure II.4.

n=385 regions) correspond to unannotated alternative 5′ splice sites (e.g., *Aplp1*) or unannotated alternative polyadenylation sites (e.g., *Cxadr*). Conversely, regions with high IE but few EI reads (Figure II.9*B-(iii) lower right*; *n*=320 regions) correspond to unannotated alternative 3′splice sites (e.g., *Mtss1l*) or unannotated alternative transcriptional start sites (e.g., *Nell2*).

High EI and IE reads correspond predominantly to retained introns (Figure II.9*B-(ii) upper right*; *n*=428 regions), the most abundant of which are a 3′ UTR intron in *Calm2* and the last intron in *Sept3* (intron 10). Previous work in mouse neurons showed that the dsRNA binding protein STAU2 interacts with the *Calm3* intron to promote dendritic localization of the intron-retaining mRNA isoform [Sharangdhar et al., 2017]. In our data, the intron-retaining isoform of *Calm2* predominated in both projection and whole cell libraries, with no selective enrichment in either cell compartment (Figure II.9). Retention of *Sept3* intron 10 results in a protein isoform with a different C-terminal sequence than the canonical isoform. Conservation of this alternative coding sequence suggests it is likely functional (Figure II.9*C*).

We looked specifically at *Kcnma1* intron 23, which has previously been reported to be retained, localized to rat primary hippocampal neuron projections [Bell et al., 2008], and spliced locally in dendrites upon neuronal activation [Bell et al., 2010]. In our data, no *Kcnma1* introns appear to be retained or localized to

projections (Figure II.14). We detect only spliced *Kcnma1* transcript isoforms in projections.

A fourth class of intron regions in projections had both low EI and IE reads (Figure II.9*B-(iv) lower left*; n = 499 regions). Some of these regions proved to harbor a gene transcribed from the same strand and contained within the intron of a different gene, such as *Cox6a1* within the last intron of *Gatc* (Figure II.14). Others corresponded to unannotated alternative terminal exons, as in *Map4* (Figure II.14). We used the presence of intronic polyadenylation sites from our PASseq data to identify these two cases (n = 96 combined). Another subset corresponded to unannotated alternative cassette exons, as in *Abi2* (Figure II.14), which we identified using evidence of ribosome occupancy (mean ribosome profiling reads $\geq$ 5 per replicate; n = 221). The remaining 278 regions showed no evidence of ribosome occupancy in our ribosome profiling data, nor of polyadenylation sites in our PASseq data, and thus they appeared to correspond to nonpolyadenylated noncoding RNAs – possibly free introns or other genes within an intron – and we looked into them further.

Circular introns with noncanonical branchpoints in projections

One explanation for detecting “free” intron regions in projection samples is that they could correspond to intron-encoded small RNAs such as small nucleolar

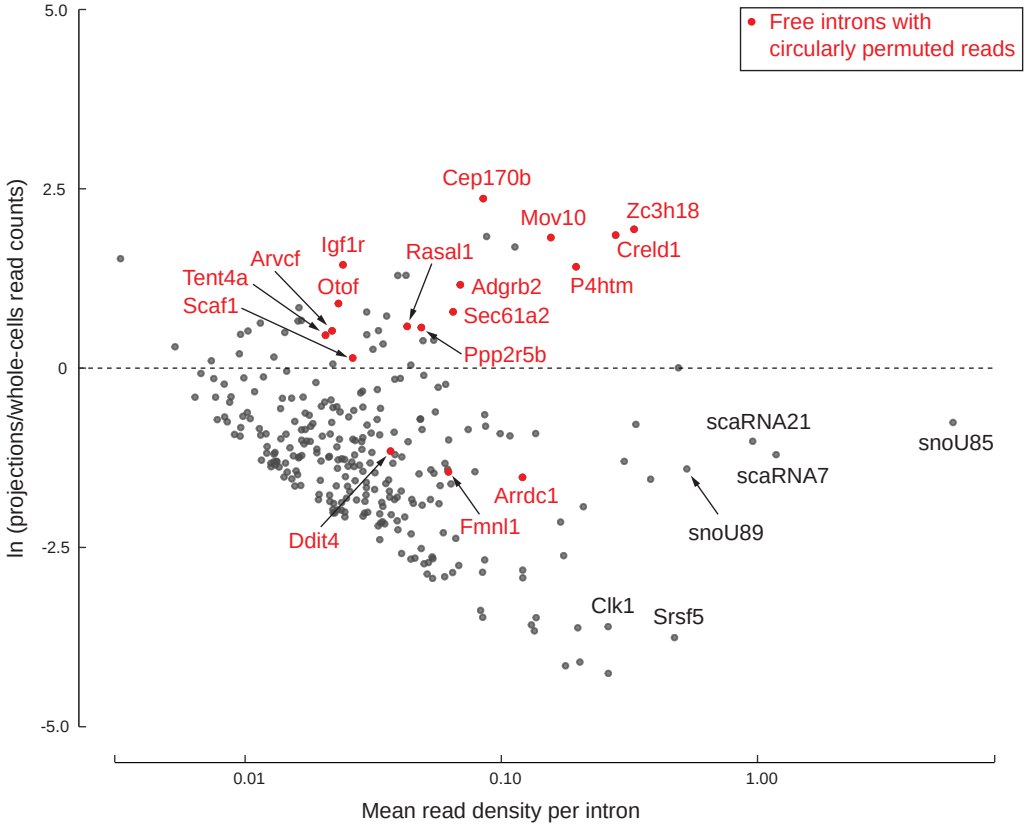


FIGURE II.11: 278 free introns detected in projections. Scatter plot of log ratio of mean read density per intron region in projections versus whole cells, and read density averaged across all replicates of projections and whole cells. Examples of previously known detained introns and intronic snoRNAs/scaRNAs are labeled in black. Introns with circularly permuted reads and without known functional annotation are labeled in red.

RNAs (snoRNAs) or small Cajal body RNAs (scaRNAs) that are nuclear-localized yet abundant enough that we detect reads in both projection and whole cell data. We plotted relative enrichment (log ratio in projections/whole cells) versus average abundance of each of the 278 intron regions (Figure II.11). This showed that most regions are indeed depleted in projection data, and we found that many do correspond to known snoRNAs and scaRNAs. Projection-depleted regions also included known detained (unspliced and nuclear-localized) introns of highly

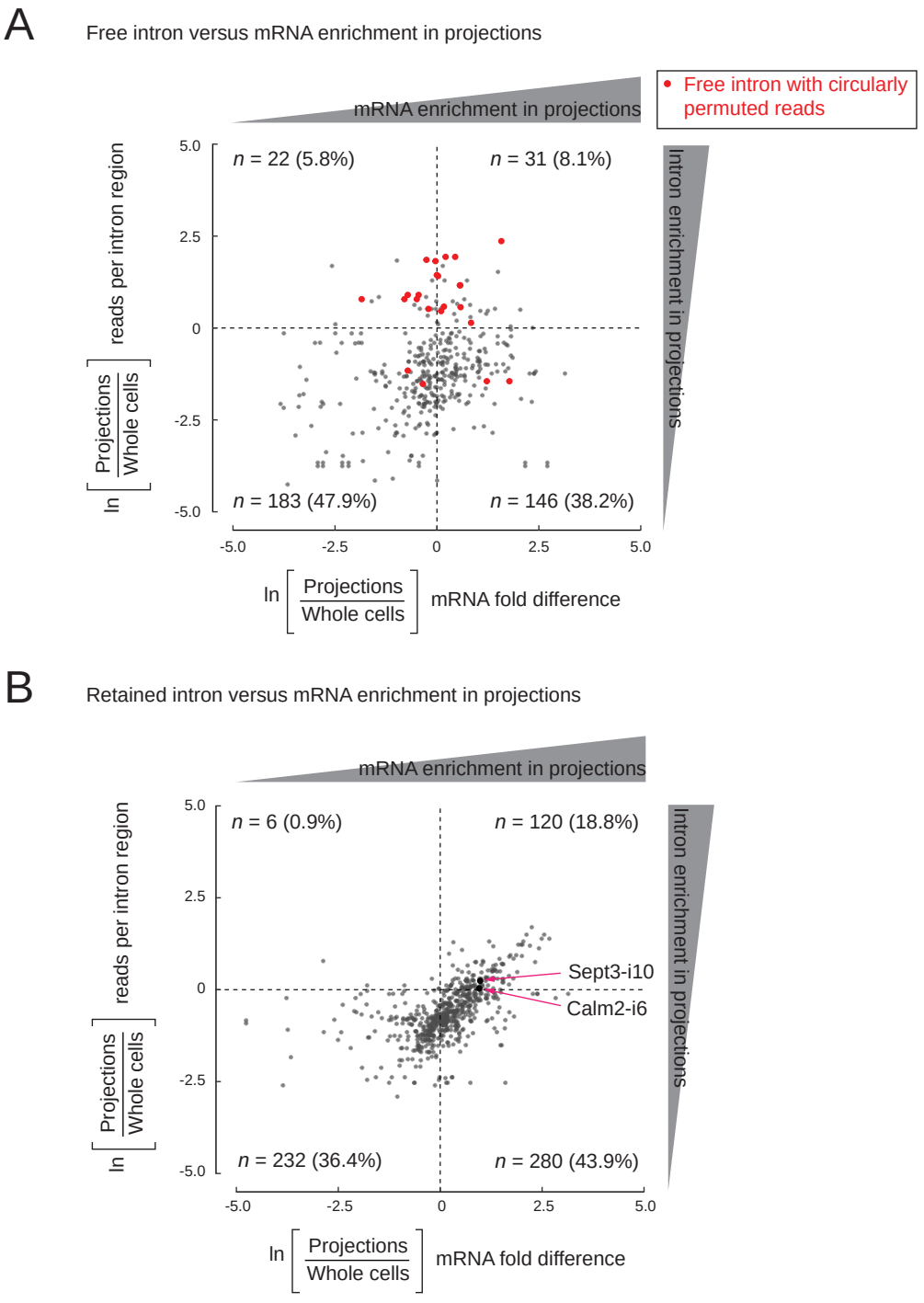


FIGURE II.12: Comparison of intron enrichment versus mRNA enrichment in projections.

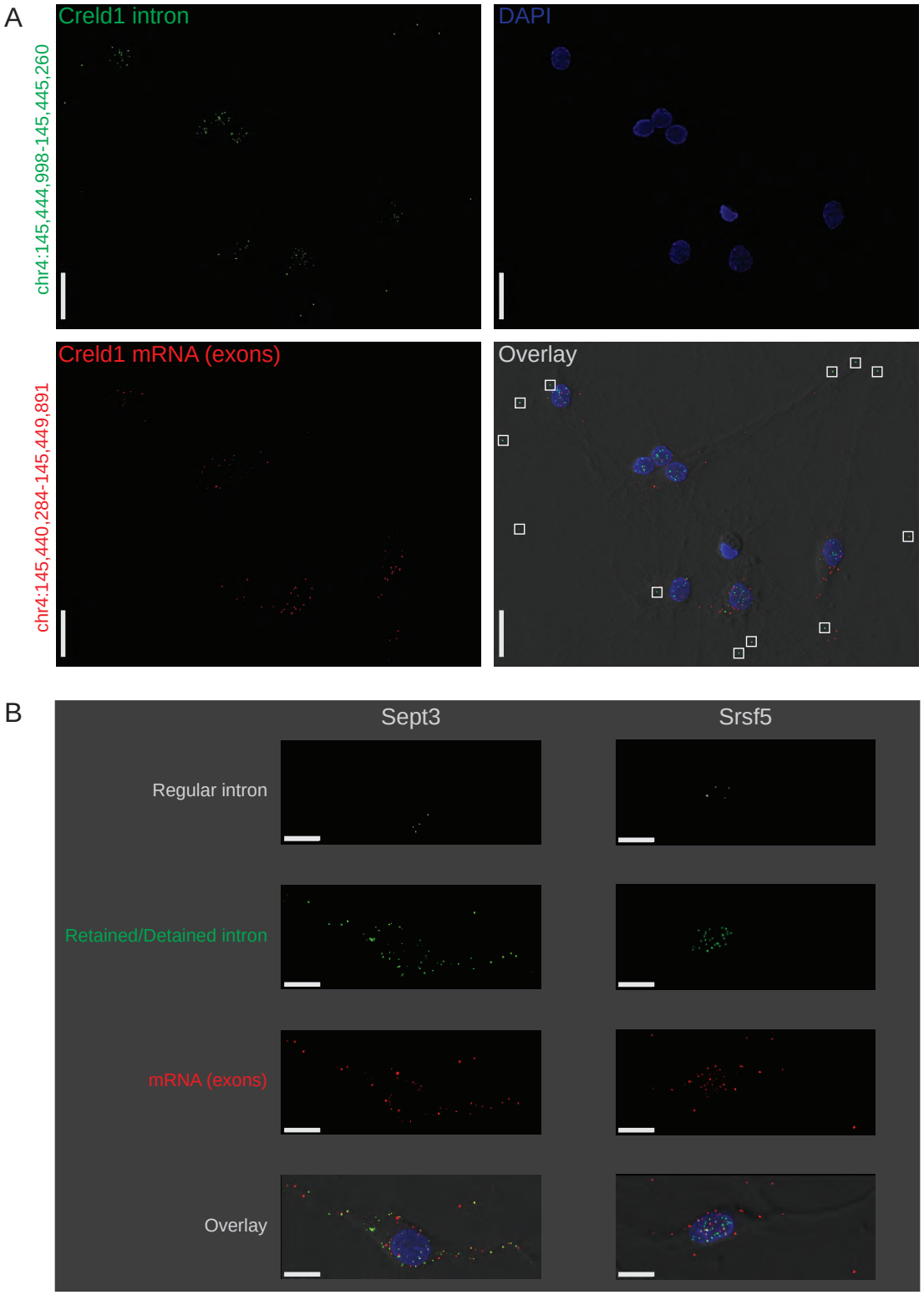


FIGURE II.13: Subcellular location of regular, retained, detained, and free introns validated by smFISH.

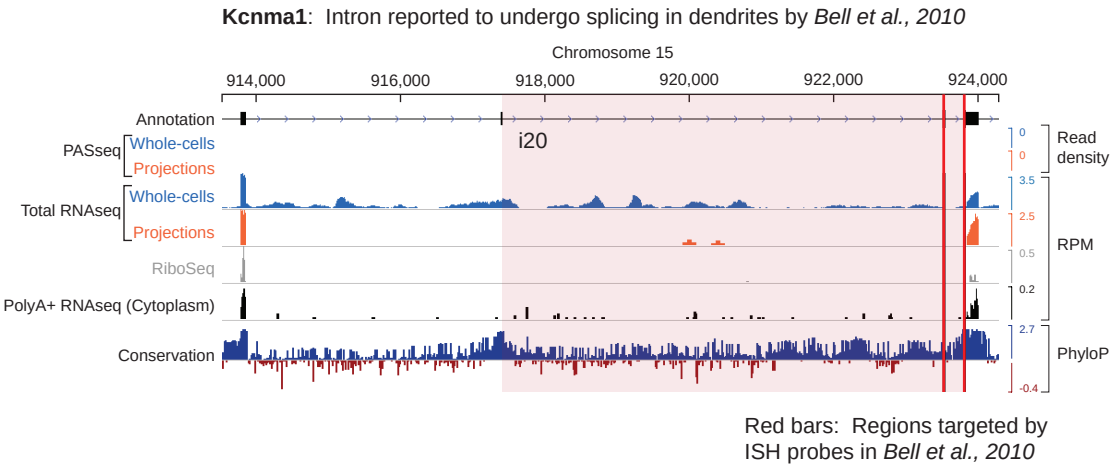


FIGURE II.14: Genome browser view of *Kcnma1* shows only the spliced isoform in projections. The intron reported to undergo splicing in dendrites is highlighted in pink, and red bars indicate the genomic locus targeted by ISH probes in the study by Bell et al. [2010].

expressed genes, including *Srsf5* and *Clk1*. We attribute their low EI and IE read counts in projection data to statistical fluctuation around our thresholds; in whole cell data, we see high EI and IE read counts for these abundant nuclear-localized introns.

However, many other intron regions were both abundant and enriched in projections (Figure II.11). These have features indicative of a previously described class of RNAs called “stable intronic sequence RNAs” [sisRNAs; Gardner et al., 2012]. SisRNAs are circular lariat products of splicing (i.e., free introns) that are inefficiently debranched in the nucleus and exported to the cytoplasm via an NXF1/NXT1-dependent mechanism [Talhouarne and Gall, 2018]. The most prominent examples in our projection libraries, such as *Creld1* (Figures II.9*C* and II.11), *Zc3h18*, and *Mov10*, were devoid of exon-intron and

intron-exon reads, with higher read density on the intron than on the flanking exons. Further, these species were not detected in polyA+ RNAseq and ribosome profiling data. We also observed a lack of read coverage over a 20-30 nt region at the 3' end of the intron.

Spliceosome-mediated intron excision from pre-mRNA releases a lariat molecule in which the branchpoint nucleotide, predominantly an adenosine [Taggart et al., 2017], is linked 2′-5′ to the 5′ end of the intron. Reverse transcriptase can occasionally traverse a 2′-5′ linkage, so lariat branchpoints result in circularly permuted reads in RNAseq data. To test whether projection-enriched free intron species had characteristics of an intron lariat, we searched for circularly permuted reads within them using find_circ.py [Memczak et al., 2013]. For fourteen projection-enriched intron regions, we detected numerous circularly permuted reads (Figure II.15*A*), enabling us to identify the branchpoints. In all 14 cases, the branchpoint was a cytosine, C (n = 12), or guanine, G (n = 2), instead of the canonical adenosine, A. The spliceosome can use C or G as a branchpoint nucleotide, but the lariat debranching enzyme is inefficient at hydrolyzing the 2′-5′ linkage at these residues [Jacquier and Rosbash, 1986]. Thus lariats with C or G branchpoints might be expected to be more stable than other introns. We attribute the lack of read coverage at the 3′ end of these introns to exonucleolytic degradation of the lariat tail, leaving stable circular RNAs with a 2′-5′ linkage.

We asked if these 14 introns had any other unusual sequence features. We derived sequence logos centered at their aligned 5′ splice sites, branchpoints, and 3′ splice sites. Their 5′ and 3′ splice sites conform to the standard consensus sequences, but their branchpoint follows a CC consensus 17-49 nucleotides upstream of 3′ splice site (Figure II.15*B*). Mismatches in read alignments at the branchpoint and the relative scarcity of circularly permuted reads versus linear reads were consistent with the existence of a 2′-5′ linkage that the reverse transcriptase traverses with lower efficiency and accuracy. Other than conservation of 5′ and 3′ splice site sequences, none of the 14 free introns showed notable phylogenetic sequence conservation (Figure II.15*C*). Although we hypothesize that the C branch site is an essential feature in stabilizing these projection-localized free introns, C branchpoints were not notably conserved across multi-species alignments of these introns, indicating little evolutionary pressure to conserve this feature.

Discussion

In this study, we conducted a comprehensive analysis of transcripts localizing to neuro-glial projections of primary rat hippocampal cells. Our data add to the growing compendium of localized RNAs identified using high-throughput methods in diverse rat/mouse neuronal cell types (e.g., motor neurons, Briese

et al. [2016]; dorsal root ganglia, Gumy et al. [2011], Minis et al. [2014]; retinal ganglion cells, Zivraj et al. [2010]; neuropil, Cajigas et al. [2012]; cortical cells, Taliaferro et al. [2016], Taylor et al. [2009], Poulopoulos et al. [2019]; primary hippocampal cells, Poon et al. [2006], Miyashiro et al. [1994]). Because our main intent was to identify and characterize intron sequences in projections, we created rRNA-depleted total RNAseq libraries from rat hippocampal cells grown on membranes that provided physical separation between projections and cell bodies (Figure II.1). As expected, 97% of the annotated intron regions that we could detect across all samples met our expression cutoffs in whole cells only (Figure II.7). Of the 3% of intron regions that are also detectable in projections, the majority turned out to be attributable to incomplete annotation of alternative mRNA isoforms or to intron-encoded snoRNAs/scaRNAs (Figure II.9). Integration of layers of information – polyA+ RNAseq, ribosome profiling, and PASseq libraries – from parallel hippocampal cultures proved essential for us to characterize these regions.

Full-length introns present at high abundance in projections fell into two classes, one expected (retained introns) and one unexpected (free introns). As a class, retained introns displayed no strong tendency toward preferential projection localization (Figure II.12). Thus, while some transcripts may contain localization elements in an alternatively retained intron [Sharangdhar et al., 2017], this does not appear to be a general function of intron retention in neurons. Some

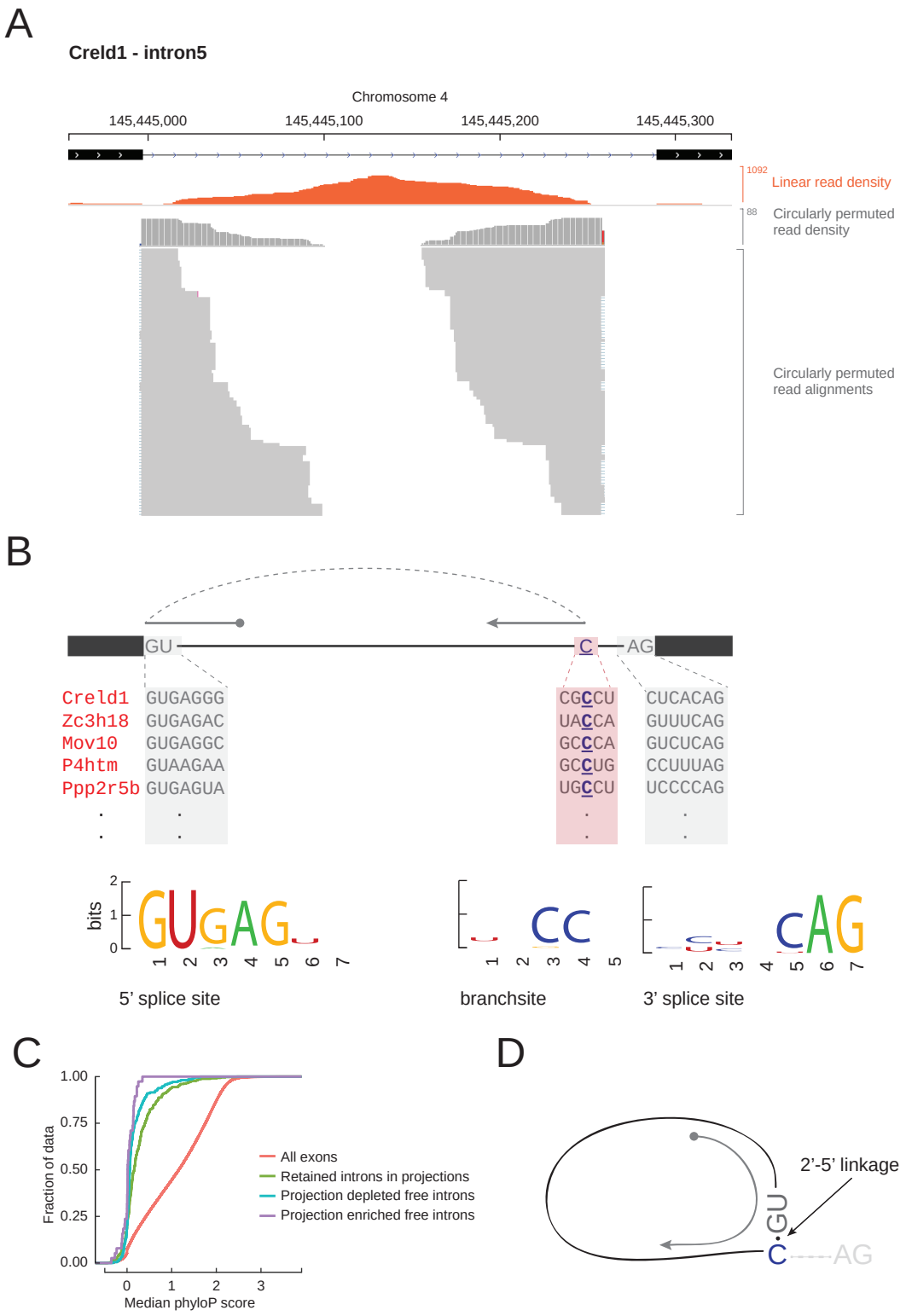


FIGURE II.15: Free circular introns with a noncanonical branchpoint (legend continues on the next page)

FIGURE II.15: Free circular introns with a noncanonical branchpoint. (A) Projection RNAseq read density for linearly mapped (orange) and circularly permuted (gray) reads across Creld1 - intron5. Alignment mismatches at the junction of circularly permuted reads are colored. Individual circularly permuted read alignments are shown as horizontal gray bars, with each bar representing one read. (B) Schematic of circularly permuted read alignment and sequence composition at 5′ splice sites, branch sites (the branchpoint nucleotide is underlined), and 3′ splice sites of five example introns. (C) Cumulative distribution function plots for phyloP sequence conservation showing lack of conservation in projection-enriched free introns relative to exons, or even to retained introns. *x*-axis represents median phyloP score over a 50 nt sliding window with 10 nt step size. (D) Inferred molecular structure of projection-localized free circular introns.

retained introns clearly express alternative protein isoforms, as we observed for *Sept3* intron 10 (Figure II.9). Notably, *Kcnma1* intron 23, previously proposed to undergo local splicing in primary rat hippocampal dendrites [Bell et al., 2010], had almost no coverage in our projection libraries in the same cell types (Figure II.14). Further, the retained introns that we identified in neuro-glial projections exhibit almost no overlap with the set of "cytoplasmic intron sequence-retaining transcripts" (CIRTs) previously reported to localize to rat primary hippocampal dendrites (Figure II.8) [Buckley et al., 2011]. CIRTs and *Kcnma1* intron 23 were initially identified by sequencing RNA from 15-300 individually dissected dendrites. While the method was state-of-the-art at the time, the low amount of input RNA necessitates multiple rounds of RNA and cDNA amplification prior to sequencing. In contrast, we started with high RNA input, restricted cDNA amplification to $\leq$ 15 cycles, and sequenced deeply to capture even low abundance RNA species.

Our most surprising finding was the detection of free introns in projections

(Figure II.11), which we infer to be lariat species containing a cytosine or guanine branchpoint instead of the canonical adenosine (Figure II.15). The most likely explanation for the persistence of such species is the inability of debranching enzyme to cleave the 2′-5′ bond at a C or G branch. That stable circular introns (or stable intron sequences, sisRNAs) can escape the nucleus and accumulate in the cytoplasm has already been reported by Talhouarne and Gall [2018]. In that study, the authors observed sisRNAs across diverse samples, including cell lines from multiple vertebrate species and mouse fibroblasts, red blood cells, liver, and brain. Despite this pervasiveness, these introns' general lack of sequence conservation and absence of interspecies overlap with sisRNAs strongly disfavors any evolutionarily conserved function. Rather, it seems more likely that sisRNAs are gene expression byproducts, perhaps with little positive or negative influence on cellular function; they may simply be noise in the system. We hypothesize that substituting any intron branchpoint from a canonical A to C/G has little effect on splicing while making the excised intron resistant to lariat debranching, leading to a stable 2′-5′ circular RNA without much phenotypic consequence, and that such substitutions are tolerated at low frequency in any organism with spliceosomal introns in its genes.

Our study extends previous work on sisRNAs by the Gall lab by finding that sisRNAs are enriched in neuro-glial projections. If these free circular introns lack evolutionary signatures of function, why would they appear to be enriched in

projections? We observe that because RNAseq experiments measure relative rather than absolute abundances, relative enrichment does not necessarily imply active localization. We hypothesize that mRNAs are typically occupied by polyribosomes in the cell body, whereas an untethered stable noncoding RNA may more freely diffuse throughout the entire volume of the cell, including distant projections. Unlike most mRNAs, free circular introns showed no evidence of ribosome occupancy. It has previously been observed for *Actb* mRNA in neurons that a lack of ribosomal engagement leads to faster mRNA diffusion kinetics than is observed for actively translating mRNAs [Katz et al., 2016]. The intracellular environment of neurons and neuronal projections is full of highly motile proteins and organelles that are themselves actively transported [Stępkowski et al., 2017], as indeed we see reflected by the prevalence of mitochondrial RNA in our projection libraries ($>$ half of all reads in our projection libraries were of mitochondrial origin; Figure II.2). We imagine that the movement of large objects within the confines of narrow cellular projections mixes the surrounding cytoplasm and could well force unanchored molecules to move in both anterograde and retrograde directions. Thus, free stable circular introns may appear to be *relatively* enriched in projections simply because they are long-lived and more freely diffusing than other RNAs that are depleted in projections because they are less stable or tethered in the cell body (to polyribosomes, for example).

This hypothesis has implications for another long-standing puzzle in neuronal RNA localization. All transcriptome-wide studies of the mRNA content of dendrites, axons, or synapses, including ours, have observed an enrichment of the complete set of ribosomal protein (RP) mRNAs [Ainsley et al., 2014, Briese et al., 2016, Cajigas et al., 2012, Moccia et al., 2003, Nakayama et al., 2017, Poon et al., 2006, Puthanveettil et al., 2013, Shigeoka et al., 2018, Taliaferro et al., 2016, Zivraj et al., 2010, to cite a few]. Ribosomes are assembled in the nucleus, so localization of RP mRNAs to distant projections is puzzling. Recently, ribosome protein synthesis and incorporation of some individual ribosomal proteins into assembled ribosomes in axons of Xenopus retinal cells has been shown to occur and to be important for axon development [Shigeoka et al., 2018], however, only a subset of RP mRNAs in axons are translated. It has been previously shown that in quiescent or growth-arrested cells (such as neurons), 30-40% of RP mRNA molecules sediment in polysome profiles as free mRNPs, compared to only ~10% for non-RP mRNAs [Meyuhas et al., 1987]. Thus, it appears that in neurons, a substantial fraction of RP mRNAs are not engaged with ribosomes, and like our free circular introns, unengaged RP mRNAs may be more free to diffuse about the cell. We propose that the relative enrichment of RP mRNAs in projections could simply be due to their low ribosome engagement, not to an active transport mechanism.

Methods and Materials

Primary hippocampal neuron culture

Animals were handled in accordance with protocols approved by the Institutional Animal Care and Use Committee at University of Massachusetts Medical School and Harvard University. All experiments were performed on primary hippocampal neurons of embryonic day 19 rat fetuses. Pregnant Sprague Dawley rats purchased from Charles River Laboratories at 19 days of gestation were euthanized by carbon dioxide asphyxiation immediately followed by diaphragm puncture to ensure death of the animal prior to surgical removal of fetuses. Fetuses were transferred to pre-cooled dishes and placed on ice. Fetal brains were gently extracted under sterile conditions and submerged in ice-cold sterile HBSS (Hank's Balanced Salt Solution, Gibco™ 14185052) for transport to the tissue-culture laboratory. The brains were transferred to ice-cold Hibernate E (BrainBits) for microdissection of the hippocampi in a sterile tissue culture hood. Isolated hippocampi were transferred to freshly prepared, pre-warmed pre-dissociation solution (comprising Hibernate E, EBSS, Papain and DNase from Worthington LK003176) and incubated at 37℃ for 30 minutes. Pre-dissociation solution was replaced by 2 ml MEM complete media (containing 50 ml 10x MEM (Invitrogen 11430-030), 15 ml 20% glucose, 15 ml 7.5% sodium bicarbonate, 2 ml 1N HCl, 400 ml water, 5 ml 200 mM glutamax (Gibco™ 35050061), 50 ml

heat inactivated horse serum, 5 ml penicillin-streptomycin (Gibco™ 15140122)) and the hippocampi were dissociated by gentle trituration, first with a regular and then a fire-polished (with reduced tip diameter) glass pipette. Additional MEM complete media was added to the dissociated cells. Cells were counted using a hemocytometer and plated at optimized densities on poly-D-lysine hydrobromide (Sigma #P0899) coated plates/surfaces. Two hours after plating, MEM complete media was replaced with Neurobasal media (500 ml Neurobasal (Invitrogen 21102-049), 1.3 mL 200 mM glutamax, 1X antibiotic/antimycotic (Gibco™ 15240062), 1x B27 (Gibco™ 17504001)). We obtained ~13 embryos per rat and ~1 million cells per embryo.

Mature neurons are post-mitotic but glial cells divide. To prevent glial cells from crowding out the neurons, we treated the culture with a DNA replication inhibitor, Cytosine β-D-arabinofuranoside (Sigma-Aldrich #C1768), on the third day in vitro (DIV). Half of the neurobasal media was replaced with fresh media on DIV = 6, and then on DIV = 12. Cells were harvested on DIV = 14.

Physical separation of neuro-glial projections from cell bodies

To separate neuro-glial projections from cell bodies, primary neurons were cultured on Falcon® Permeable Support with Polyethylene Terephthalate (PET) membranes (Corning Life Sciences #C353102) that have 1µm pores such that

the cell bodies remain on top of the membrane, but cellular projections (axons, dendrites, and glial projections) can grow through the pores to the underside of the membrane [Poon et al., 2006]. Lysate collected from the top surface of the membranes comprises whole cells (cell bodies and projections) whereas lysate from the underside of the membranes comprises neuro-glial projections and some cytoplasmic blebs.

Primary hippocampal neurons were cultured at a density of 0.2 x 10^6 per well on Falcon® Permeable Support designed to fit 6-well tissue culture plates (Corning Life Sciences #C353102). The plating density was optimized to ensure healthy neuronal cultures for harvesting on DIV = 14. Before setting up the neuronal culture, membranes were immersed in poly-D-lysine hydrobromide solution (0.1 mg/ml in 0.05 M sodium borate pH 8.5) overnight at 37℃. Before plating neurons, membranes were rinsed three times with sterile water and incubated at 37℃ for at least 2 hours while immersed in MEM complete media. Neuronal health was assessed for every biological replicate by immunofluorescence imaging using a dendritic marker, MAP2, to visually inspect cell morphology.

RNA isolation from neurons cultured on semipermeable membranes

To extract RNA from neuro-glial projections, we removed media from the Falcon® Permeable Support, turned it upside-down with the projections-side of the

membrane facing upward, applied 200 µl of TRIzol® reagent (ThermoFisher #15596018) to the membrane, quickly scraped the surface with one stroke using a cell-scraper, tilted the membrane and gently collected the lysate from the edge of the membrane with a pipette. Lysate was similarly collected from the cell body side of the membrane by applying 500 µl of TRIzol® reagent.

RNA was extracted following the steps recommended by TRIzol® reagent manual with minor modifications. Briefly, after application of TRIzol® reagent, the lysate was transferred to 15 ml Falcon tubes and vortexed for 30 seconds, followed by a 5 minute incubation at room temperature to dissociate nucleoprotein complexes. 200 µl of chloroform per 1 ml of TRIzol® was then added for phase separation. The tubes were vigorously shaken by hand and centrifuged at 12,000 x g for 15 min at 4 ℃. The aqueous phase was transferred to fresh tubes and the chloroform wash was repeated two more times. RNA was precipitated by mixing with 100% isopropanol, incubating at room temperature for 15 min, and centrifugation at 12,000 x g at 4℃ for 15 min. The supernatant was removed without disturbing the precipitated RNA. The RNA precipitate was rinsed two times with 75% ethanol, air dried, and dissolved in RNase-free water by incubating for 10 min at 55-60℃. RNA was stored at -80℃ until needed.

Typical RNA yield from neuronal cultures on 8 x 6-well plates with the semipermeable membrane inserts was ~5 µg from the projections lysate for

every ~50 µg from the whole cell lysate. The quality and quantity of RNA was assessed by Nanodrop™ UV spectrophotometer (A260/A280 measurements) and Bioanalyzer Pico RNA microcapillary electrophoresis (Figure II.3).

rRNA-depleted total RNAseq library preparation

RNAseq libraries were prepared following the protocol published in [Zhang et al., 2013]. Briefly, 5 µg of RNA from each sample was treated with TURBO™ DNase (ThermoFisher AM2238) followed by clean-up and enrichment of RNA $>$ ~150 nt using RNA clean and concentrator (Zymo Research #R1013). DNase activity was tested beforehand to ensure DNA digestion. Total RNA was depleted of rRNA using the Ribo-Zero rRNA removal kit (Illumina #MRZH11124), following manufacturer's protocol. RNA was hydrolyzed using 5X first strand buffer (provided with Superscript III reverse transcriptase, ThermoFisher #18080044) at 94℃ for 4 minutes and 50 seconds and immediately moved to ice. The fragmented RNA was reverse transcribed using random hexamers (ThermoFisher #N8080127) and Superscript III to make single-stranded cDNA. To make strand specific libraries, the second strand complementary to the cDNA was transcribed with DNA polymerase I (New England Biolabs #M0209S) using dUTP instead of dTTP. Illumina sequencing adapters were ligated to the double stranded cDNA. The dUTP-containing strand was degraded using Uracil-DNA Glycosylase (New England Biolabs #M0280S). The resulting single-stranded

cDNA was amplified with 13 or 15 PCR cycles followed by size-selection using Pippin Prep (Sage Science). Before subjecting to Illumina sequencing, the RNAseq library quality was assessed by running the samples on the Bioanalyzer and Sanger sequencing a subset of TOPO-TA cloned products. Only those libraries with sufficient final concentration ($>$ 12 nM), a product of appropriate size range (mode 300 nt size), and comprising expected RNA sequences (for instance, exon regions of abundant mRNAs) were selected for Illumina sequencing.

A total of five biological replicates were sequenced over a span of 2 years, replicates 1-3 (paired-end 100) in 2014 and replicates 4,5 (paired-end 125) in 2016.

PolyA site sequencing

PolyA site sequencing (PASseq) libraries were prepared following the protocol in Ashar-Patel et al. [2017] and Heyer et al. [2015]. 2-5 µg of total RNA from each sample (3 biological replicates of projection and whole-cell lysates) was treated with DNase and fragmented as described above. The RNA was reverse transcribed with Superscript III using an anchored oligo-dT primer containing Illumina sequencing adapters and a unique barcode for each sample. Single-stranded RNA was degraded with RNaseI. The cDNA was denatured (65℃ for 5 min) and resolved by electrophoresis on a 10% polyacrylamide gel to select

160-210 nt sized fragments (for a 50-100 nt expected insert size sans the adapter sequences). To extract cDNA from the gel, a piece of the gel containing the cDNA was cut at the appropriate location, crushed, and nutated overnight in a solution of 300 mM sodium chloride and 10 mM EDTA. The solution was recovered from gel pieces by centrifugation in Corning™ Costar™ Spin-X™ columns (#07200386) at 10,000 x g for 3 minutes. cDNA was precipitated using isopropanol, followed by washes in 70% ethanol. The cDNA was then circularized using CircLigase (EpiCentre BioTechnologies #CL4115K) and amplified with 12-14 PCR cycles. The amplified DNA library was further enriched for a product of size 180-280 nt to exclude insert-less product (150 nt) using Pippin Prep.

Ribosome profiling and polyA+ RNAseq from the cytoplasmic fraction

Ribosome profiling and corresponding RNAseq libraries were prepared from fractionated cytoplasmic lysate of primary hippocampal neurons following the protocol in Ricci et al. [2014] and Heyer et al. [2015].

Rat primary hippocampal neurons were cultured (as described above) on poly-D-lysine coated 6-cm plates (three plates per sample) at a density of 1 x 10^6 cells per plate. After 14 days in vitro, cyclohexamide was added to the media at 100 µg/µl final concentration for 10 min to stall translation. The plates were placed on ice where the media was removed and the cells were washed two

times with 2 ml ice-cold PBS containing 100 µg/ml cyclohexamide. The cells were lysed in 200 µl lysis buffer (10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 1% Triton™ X-100, 2 mM DTT, 100 µg/ml cycloheximide, protease inhibitor (Complete, EDTA-free, Roche)). The lysate was collected by scraping the plates, transferred to a clean microcentrifuge tube and incubated on ice for 5 min, followed by centrifugation at 1300 x g for 10 min to pellet the nuclei. The supernatant was recovered, flash frozen in liquid nitrogen, and stored at -80℃ until needed. Half of the lysate was used for ribosome profiling and the other half for polyA+ RNAseq library preparation.

To purify ribosome occupied RNA sequences for ribosome profiling, the RNA was digested with 300 units of RNase T1 (Fermentas) and 500 ng of RNase A (Ambion) for 30 minutes at room temperature to break down polysomes into monosomes. The monosomes were purified by density gradient ultracentrifugation. Lysates were fractionated by centrifugation through a 10-50% (weight/volume) linear sucrose gradient (20 mM HEPES-KOH, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT, 100 µg/ml cyclohexamide) at 35,000 r.p.m. for 2 hours and 40 minutes at 4℃. A gradient fractionator (Brandel) was used to identify and collect the monosome enriched fraction by measuring absorbance at 254 nm. RNA was extracted from the monosome fraction and resolved by electrophoresis on a denaturing polyacrylamide gel to select RNA fragments ranging from 26 - 32 nt in size.

To extract RNA for ribosome profiling and polyA+ RNAseq, SDS was added to 1% final volume, and proteinase K (Invitrogen) was added to a final concentration of 200 µg/ml. The samples were incubated at 42℃ for 45 minutes. One volume of acid phenol/chlorofom (Ambion AM9720, pH 4.5) was added and the samples were vortexed for 30 seconds followed by centrifugation at 12,000 x g for 15 minutes. The supernatant was transferred to a clean microcentrifuge tube and 0.1 volume of sodium acetate (3 M, pH 5.2) and 10 mM final concentration of $MgCl_2$ were added. To precipitate RNA, 1 volume of 100% isopropanol was added to the solution and centrifuged at 12,000 x g for 35 minutes. The precipitated RNA was rinsed with 70% ethanol, air dried, and reconstituted in 5 µl water.

For polyA+ RNAseq library preparation, the RNA was partially hydrolyzed using Fragmentation Reagent (Ambion) prior to cDNA library preparation.

To prepare cDNA libraries from ribosome occupied RNA fragments and from RNA fragments from the cytoplasmic lysate, the 3′ ends of RNA fragments were dephosphorylated with T4 polynucleotide kinase (New England BioLabs #M0201S). A preadenylated DNA adaptor sequence was ligated to the 3′-hydroxyl ends of the RNA fragments using T4 RNA Ligase (T4 RNA Ligase 2, truncated K227Q, NEB #M0351S). The ligated RNA product was reverse transcribed using Superscript III and a barcoded primer with sequence complementarity to the adaptor. The reverse transcription primer also contained

adaptors required by Illumina sequencers. The resulting cDNAs were enriched for desired product size, circularized, and amplified following the steps described for PASseq library preparation.

Ribosome profiling libraries were amplified with 8 PCR cycles whereas polyA+ RNAseq libraries were amplified with 13 PCR cycles and sequenced on Illumina HiSeq sequencers for 50 nt single reads.

Immunofluorescence staining

Cells were cultured either on semipermeable membrane inserts or Thermo Scientific™ Lab-Tek™ II chamber slides as described in previous sections. At DIV = 14, media was removed, the cells were rinsed two times with phosphate buffered saline (PBS, pH = 7.4) and subsequently treated with fixative (4% paraformaldehyde) for 10 min at room temperature. The fixative was removed and the cells were rinsed three times with PBS followed by permeabilization with 0.1% Triton™ X-100 for 5 minutes. Cells were then rinsed three times with PBS and quenched with 50 mM ammonium chloride (in double-distilled water) for 10-15 minutes, followed by three more rinses with PBS. The fixed and permeabilized cells were then incubated with 10% normal goat serum in PBS (blocking solution) for 30 minutes at room temperature. For cells grown on membranes, the membranes were cut out of their plastic support system using a

sharp blade and transferred to small chambers for the next steps. The samples were kept moist at all times during the protocol.

The membranes or slides were incubated overnight at 4℃ with primary antibody diluted in blocking solution then rinsed three times with PBS. Fluorescently-labeled secondary antibody was applied for 1 hour in the dark at room temperature and rinsed by three washes with PBS. The membranes were cut in half and placed on glass slides with either the whole cell or the projection side on top. ProLong™ Gold antifade media with DAPI was applied to the membranes before covering them with a 0.16-0.19 mm thick cover glass. The edges of cover glass were sealed with transparent nail-polish and allowed to set overnight in the dark at room temperature. The samples were imaged on DeltaVision or Zeiss Cell Discoverer microscopes.

smFISH

smFISH was performed following ACDBio protocol. Briefly, cells were cultured on Ibidi poly-D-lysine coated chambered coverslips. At DIV=14, media was removed, cells were rinsed two times with PBS, fixed for 30 minutes at room temperature, then rinsed three times with PBS. The cells were dehydrated by incubating them in sequentially higher concentrations of ethanol (50%, 70%, and 100%, respectively) for 5 minutes each and a final immersion in 100% ethanol

for at least 10 min at room temperature. The dehydration was then reversed by incubation in 70% ethanol for 2 minutes, 50% ethanol for 10 minutes, and rinsing with PBS at room temperature. They were then treated with Protease III (ACD Bio) for 10 minutes at room temperature. smFISH probes were hybridized following manufacturer's instructions. In the end, the cells were counterstained with DAPI and mounted with ProLong™ Glass Antifade Mountant (P36980). Probes were multiplexed to image up to three different targets in the same sample. The samples with imaged on Zeiss Cell Discoverer at the Harvard Center for Biological Imaging.

COMPUTATIONAL ANALYSIS

An outline for the data analysis workflow is shown in Figures II.16 and II.17. Descriptive placeholder input and output filenames (enclosed in <>) are used in the commands shown. If reusing these commands, please replace placeholder filenames with appropriate ones and omit enclosing "<" and ">".

rRNA-depleted, total RNAseq genome alignment

Paired-end reads from five biological replicates (10 samples total) of projection and whole cell RNAseq are provided in the fastq format at GSE129924. Reads from replicates 1-3 are 2 $\times$ 100 nt, whereas replicates 4 and 5 are 2 $\times$ 125 nt long.

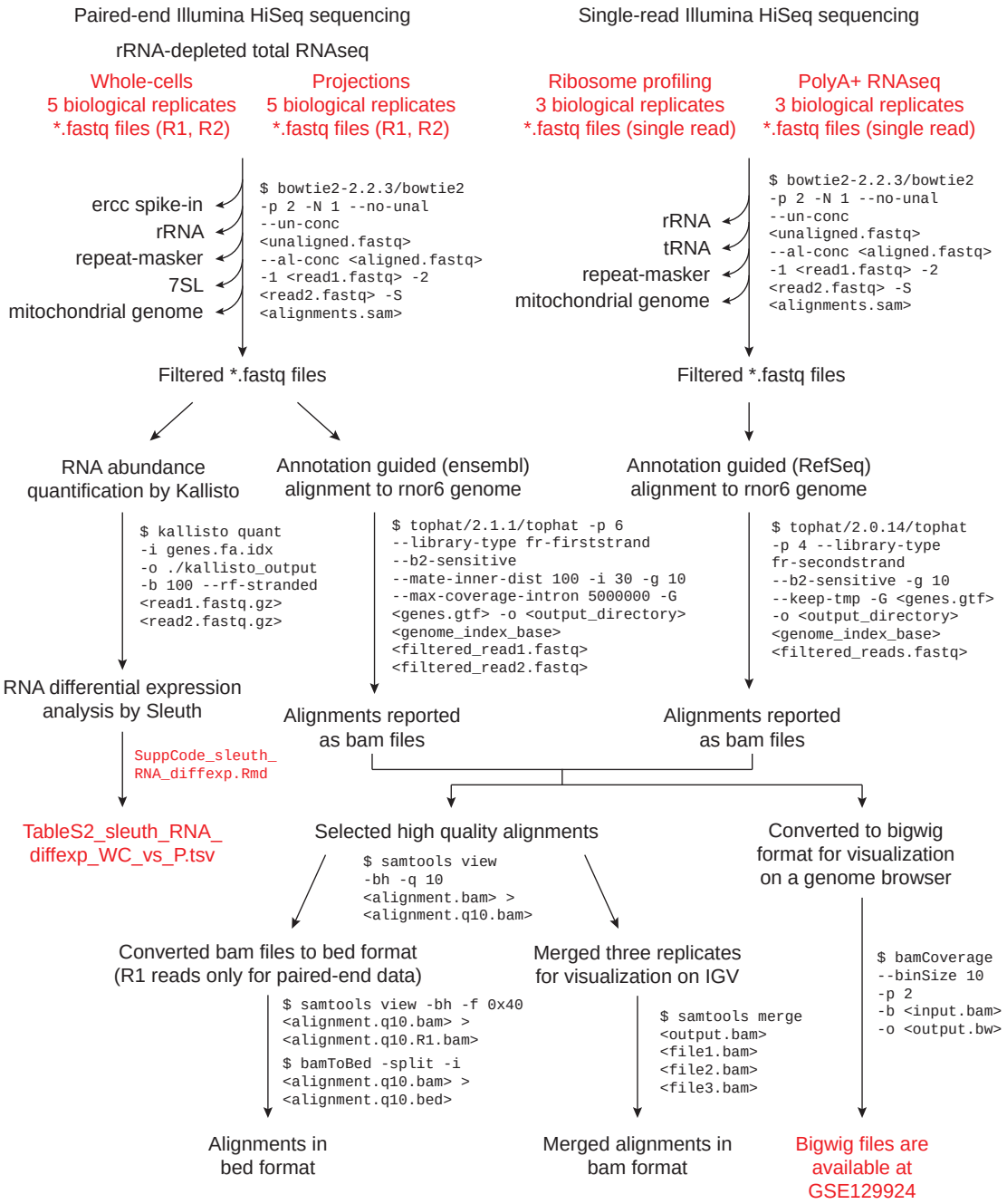
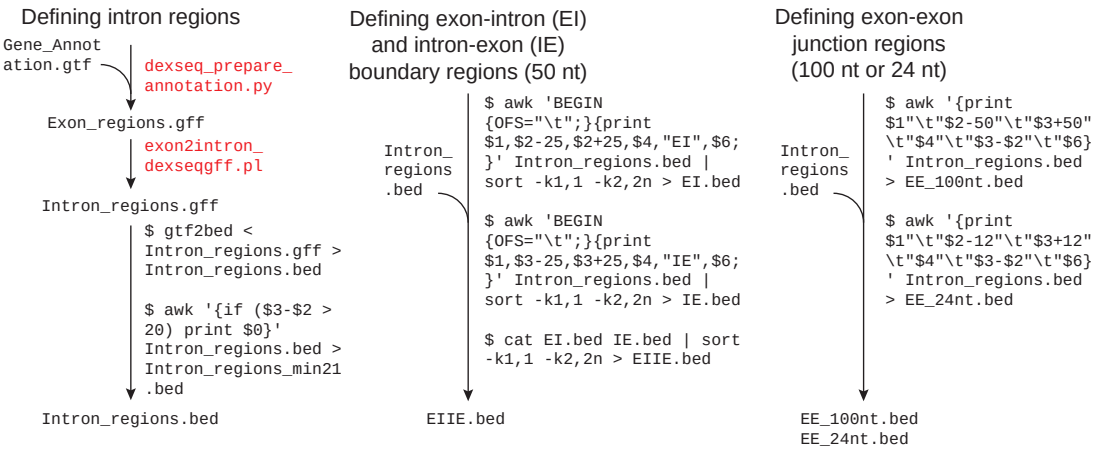


FIGURE II.16: Data analysis workflow for sequence alignments and RNA abundance quantitation. Files provided as supplementary data or available at GSE129924 are in red text.



Counting the number of reads on regions of interest

Library type	Intron regions	EI and IE boundary regions	EE junction regions
Total RNA-seq	`$ coverageBed -sorted -S -nobuf -a Intron_regions.bed -b alignment.q10.bed`	`$ coverageBed -sorted -S -f 0.60 -nobuf -a EIIE.bed -b alignment.q10.bed`	`$ coverageBed -S -f 1.0 -counts -sorted -a EE_100nt.bed -b alignment.q10.bam`
PolyA+ RNA-seq	`$ coverageBed -sorted -s -nobuf -a Intron_regions.bed -b alignment.q10.bed`	`$ coverageBed -sorted -s -f 0.60 -nobuf -a EIIE.bed -b alignment.q10.bed`	`$ coverageBed -s -f 1.0 -counts -sorted -a EE_100nt.bed -b alignment.q10.bam`
Ribosome profiling	`$ coverageBed -sorted -s -nobuf -a Intron_regions.bed -b alignment.q10.bed`	`$ coverageBed -sorted -s -f 0.60 -nobuf -a EIIE.bed -b alignment.q10.bed`	`$ coverageBed -s -f 1.0 -counts -sorted -a EE_24nt.bed -b alignment.q10.bam`
PAS-seq	`$ coverageBed -sorted -s -nobuf -a Intron_regions.bed -b alignment.q10.bed`		

Output provided in: TableS3_intron_master.tsv

FIGURE II.17: Data analysis workflow to count reads on intron regions. Files provided as supplementary data or available at GSE129924 are in red text.

Replicates 4 and 5 also contain ERCC RNA spike-in. Reads from all 10 samples were aligned to ERCC, rRNA, repeat elements cataloged by RepeatMasker [Jurka, 2000], 7SL or SRP (the RNA component of signal recognition particle), and the mitochondrial genome, serially in that order, using bowtie2 version 2.2.3 [Langmead and Salzberg, 2012]. Unaligned reads after each step were passed on for the next alignment. The following parameters were used for each alignment:

```
$ bowtie2-2.2.3/bowtie2 -p 2 -N 1 --no-unal \\
--un-conc <unaligned.fastq> --al-conc <aligned.fastq> \\
-1 <read1.fastq> -2 <read2.fastq> -S <alignments.sam>
```

Unaligned or filtered reads were then mapped to the rat genome (Ensembl release 81, Rnor_6.0, annotation downloaded on July 24, 2015) [Zerbino et al., 2018] with TopHat version 2.1.1 [Kim et al., 2013] using the following parameters:

```
$ tophat/2.1.1/tophat -p 6 --library-type fr-firststrand \\
--b2-sensitive --mate-inner-dist 100 -i 30 -g 10 \\
--max-coverage-intron 5000000 -G <genes.gtf> -o <output_directory> \\
<genome_index_base> <filtered_read1.fastq> <filtered_read2.fastq>
```

High quality read alignments were selected (using SAMtools version 1.4.1) for visualization on the rat genome browser.

```
$ samtools view -bh -q 10 <alignment.bam> > <alignment.q10.bam>
```

PolyA site identification from PASseq data

Six samples (three from whole cell and three from projection lysates) were barcoded and sequenced in one lane on NextSeq 500 with 150 cycles. The barcoded reads were parsed using Illumina's bcl2fastq version 1.8.4 conversion software.

```
$ bcl2fastq --barcode-mismatches 1 -R <run_directory> \\
-o <output_directory> --use-bases-mask I5y*n 2> <output.log>
```

Reads were trimmed using cutadapt version 1.7.1 [Martin, 2011] to remove stretches of A's from the 3′ end and barcode sequence from the 5′ end, and then selected for minimum 25 nt resulting read length. We used the fastx toolkit version 0.0.14 (`http://hannonlab.cshl.edu/fastx_toolkit/`) to further select reads with sequencing quality greater than 35. Parsed, trimmed, and filtered fastq files are provided on GSE129924.

```
$ cutadapt -a AAAAAAAA -o <A_trimmed.fastq> <input.fastq> \\
2> <output.log>
$ cutadapt -u 7 -m 25 -o <A_barcode_trimmed.fastq> \\
<A_trimmed.fastq> 2> <output.log>
$ fastq_quality_filter -v -q 35 -p 50 -i <A_barcode_trimmed.fastq> \\
-o <A_barcode_trimmed_hiqual.fastq> 2> <output.log>
```

For PAS site identification, we used cleanUpdTSeq v.1.0.2 [Sheppard et al., 2013], which calculates the probability of a genomic locus to be a true polyadenylation site. The reads were first aligned to the rat genome using TopHat version 2.1.1 [Langmead and Salzberg, 2012], then reads with mapping quality greater than 10 were selected using SAMtools version 1.4.1 [Li et al., 2009].

```
$ tophat/2.1.1/tophat -p 4 --library-type fr-secondstrand \\
--b2-very-sensitive --no-novel-juncs -i 30 -g 10 \\
-G <genes.gtf> -o <output_directory>\\
<genome_index_base> <PASseq_reads.fastq>


$ samtools view -bh -q 10 <alignment.bam> > <alignment.q10.bam>
```

Strand-specific 3′ end read alignment coordinates were extracted using BEDTools version 2.26.0 [Quinlan and Hall, 2010].

```
$ bedtools genomecov -3 -d -strand - -ibam <alignment.q10.bam> \\
-g <rnor6.ChromInfo.txt> > <output.q10.3cov.minus.txt>
$ bedtools genomecov -3 -d -strand + -ibam <alignment.q10.bam> \\
-g <rnor6.ChromInfo.txt> > <output.q10.3cov.plus.txt>
```

Only those genomic loci with more than 5 read alignments were considered for further analyses.

```
$ awk '{if ($3 > 5) print $0}' <*.3cov.minus.txt> \\
> <*.3cov.minus.non0.txt>
$ awk '{if ($3 > 5) print $0}' <*.3cov.plus.txt> \\
> <*.3cov.plus.non0.txt>
```

To prepare data for use in CleanUpdTseq, files were converted to bed format and a unique identifier for each site was added. Each line of the bed file contained

single nucleotide genomic coordinates and the number of reads whose 3′ ends aligned at that locus.

```
$ awk '{OFS="\t"}; {print $1, $2, $2, "SampleID_"NR, $3, "+"}' \\
> <3cov.plus.bed>
$ awk '{OFS="\t"}; {print $1, $2, $2, "SampleID_"NR, $3, "-"}' \\
> <3cov.minus.bed>
```

Bed files from the plus and minus strands were combined and sorted, and then they were split by chromosome.

```
$ cat <3cov.plus.bed> <3cov.minus.bed> | sort -k 1,1 -k 2,2n \\
> output.bed


$ awk '{if ($1=="chr1") print $0}' <*.bed > > <*.chr1.bed>
$ awk '{if ($1=="chr2") print $0}' <*.bed > > <*.chr2.bed>
...
$ awk '{if ($1=="chrY") print $0}' <*.bed > > <*.chrY.bed>
```

A custom R script was used to run cleanUpdTSeq to identify polyA sites. Genomic loci with a high probability of being a true polyadenylation site (p-value < 0.001) were selected for further analysis or visualization on the rat genome browser.

```
$ Rscript ./cleanUpdtseq.R $FILE $NAME.nbp rat
```

Cytoplasmic polyA+ RNAseq and ribosome profiling genome alignment

The adapter was trimmed from fastq files, and only reads longer than 24 nt were kept. Reads were filtered for rRNA, tRNA, repeat elements, and mitochondrial genome following the steps described in the PASseq methods section. Reads were aligned to the genome with TopHat using RefSeq annotation (RGSC 6.0/rn6, Jul. 2014, downloaded on July 7, 2016) as a reference. The alignment is reported in bam file format.

```
$ tophat/2.0.14/tophat -p 4 --library-type fr-secondstrand \\
--b2-sensitive -g 10 --keep-tmp -G <genes.gtf> \\
-o <output_directory> <genome_index_base> <filtered_reads.fastq>
```

Differential expression analysis

To quantify annotated transcript abundance, we used Kallisto version 0.44.0 [Bray et al., 2016] with the following parameters:

```
$ kallisto quant -i genes.fa.idx -o ./kallisto_output \\
-b 100 --rf-stranded <read1.fastq.gz> <read2.fastq.gz>
```

Reference sequences (fasta format) of protein coding and noncoding RNAs were downloaded from ensembl.org [Zerbino et al., 2018] in January, 2018. Transcripts from the mitochondrial genome were omitted from this analysis. The sequence

of nuclear noncoding RNA, *Xist*, was imported from RefSeq (NR_132635.1) [O'Leary et al., 2016] because it was not annotated in the rat genome reference data downloaded on July 24, 2015 (Ensembl release 81, Rnor_6.0).

To identify RNAs enriched in projections, we compared transcript abundances in projections to whole cells using Sleuth version 0.30.0 [Pimentel et al., 2017]. The code used to run sleuth is provided as SuppCode_sleuth_RNA_diffexp.Rmd and the output is provided as Table S2 (TableS2_sleuth_RNA_diffexp_WC_vs_P.tsv).

Gene ontology analysis

To identify gene families enriched in projections or whole cells, we used GeneCodis `http://genecodis.cnb.csic.es/` (accessed on November 5, 2018) [Carmona-Saez et al., 2007] with default parameters, focusing on the cellular component for gene ontology classification. RNAs enriched in projections (n = 1,440) or whole cells (n = 1,297) were selected using a q-value cutoff of $<$ 0.01 and $\log_e$ mean read counts across all libraries $>$ 1 and compared against a background comprising all RNAs that had $\log_e$ mean read counts $>$ 1 (n = 19,461).

Intron quantification

To define intron regions, we first extracted the genomic coordinates of all annotated exons in any isoform of a gene as a .gff file using dexseq_prepare_annotation.py (provided as supplementary code) from DEXSeq [Anders et al., 2012]. Intron regions were defined as the non-exon regions of each gene. Intron region coordinates were derived from exons using a program written by Alejandro Reyes (ORCID: 0000-0001-8717-66) copied from `http://seqanswers.com/forums/showpost.php?p=137918&postcount=4`.

To count the number of reads aligning to intron regions, exon-intron (EI) boundaries, intron-exon (IE) boundaries, and exon-exon (EE) junctions, we used BEDTools version 2.26.0 [Quinlan and Hall, 2010]. Reads were aligned to the genome using TopHat, a splice-sensitive alignment algorithm, as described above. Only reads with high quality alignments were counted on regions of interest.

The intron reference file was first converted from .gff to .bed file format using BEDOPS [Neph et al., 2012]. EI and IE regions span 50 nt, 25 nt from the exon and 25 nt from the intron. The EE region coordinates include the intron region and 50 nt of each flanking exon. To count reads crossing the EE junction, only those read alignments that started and ended at the EE coordinates were considered. Since EE regions are longer than 100 nt because they include the

intron, the only way a 100 nt long read would start and end at the EE coordinates is if the alignment splits. In case of ribosome profiling and polyA+ RNAseq data, where the minimum read length is 24 nt, the EE regions included the intron region and 12 nt of each flanking exon. The commands used to align the reads to the rat genome, define regions of interest, and count reads on them are shown in Figures II.16 and II.17.

Circularly permuted read alignments

To align reads allowing circular permutation, we used find_circ.py version 1.2 [Memczak et al., 2013] following instructions provided on the github repository `https://github.com/marvin-jens/find_circ`. Reads that failed to align to the rat genome using TopHat (see above) were used to search for circularly permuted alignments. The following commands were executed:

```
$ unmapped2anchors.py <unmappedreads.bam> > <unmapped_anchors.qfa>


$ bowtie2 -p 8 --score-min=C,-15,0 --reorder --mm -q \\
-x <Bowtie2Index/genome> -U <unmapped_anchors.qfa> | \\
find_circ.py -G <genome.fa> -B <anchors.bam> --noncanonical \\
-R <spliced_reads.txt> -s <stats.txt> > <circs.txt>
```

Sequence conservation

To assess sequence conservation of regions of interest (free intron regions enriched in and depleted from projections, retained introns in projections, and all annotated exons), we used the PhyloP sequence conservation scores from 20 aligned vertebrate genomes [Pollard et al., 2010] downloaded from the UCSC database [Kent et al., 2002]. Each nucleotide has a PhyloP conservation score. For every region of interest, we calculated the average PhyloP score across 50 nt windows with a 10 nt interval (using bigWigAverageOverBed version 2 from Kent et al. [2010]) and took the median score across all windows in the region.

```
$ bedtools makewindows -w 50 -s 10 -b <region.bed> \\
-i srcwinnum > <windows.bed>

$ bigWigAverageOverBed -bedOut=<avg_phyloscores.bed> \\
<rn6.phyloP20way.bw> <windows.bed> <avg_phyloscores.tab>
```

Data and code availability

All raw data (fastq format) and corresponding coverage files (bigwig format) are available at NCBI GEO under accession number GSE129924, and a tarball of supplementary tables and code can be found on `http://eddylab.org/publications.html#Saini19`

Acknowledgments

We thank Guiping Wang and Xiaowei Zhuang (Harvard University) for providing reagents and technical assistance with smFISH experiments; the deep sequencing core facility at University of Massachusetts Medical School (UMass) for HiSeq sequencing; Massachusetts Green High Performance Computing Center (UMass) for computational resources; Harvard Center for Biological Imaging for access to microscopes and image processing resources; Harvard MCB graphics for feedback on data presentation; Phillip Zamore (UMass) and members of the Eddy lab (Harvard University) for discussions and critical reading of the manuscript. This work was supported by funding from HHMI (S.R.E.), NIH (R01-GM53007 to M.J.M.), and an HHMI International Student Research Fellowship (H.S.). M.J.M. was an HHMI Investigator at the time this study was conducted.

CHAPTER III

Full-length and fragmented tRNAs abound in neuronal projections

Harleen Saini

RNA Therapeutics Institute, University of Massachusetts Medical School,

Worcester, United States

Department of Molecular and Cellular Biology, Howard Hughes Medical Institute,

Harvard University, Cambridge, United States

Contribution summary

H.S. performed all the experiments and analyzed the data with feedback from Alicia A. Bicknell, Tom A. Jones (Harvard University), Sean R. Eddy and Melissa J. Moore. This chapter was written by H.S.

Introduction

Sometimes experiments yield unexpected results. In Chapter II, we extracted RNA from neuro-glial projections and observed a high intensity signal between 25-100 nt in a total RNA microcapillary electrophoresis profile (Figure II.3). In a typical cell, rRNA comprises ~80% or more of total RNA by mass [Lodish et al., 2000]. Indeed, when RNA from intact cells is size-separated by denaturing gel electrophoresis and stained with a dye, 28S and 18S rRNAs appear as two prominent bands of ~4 kb and ~2 kb length. As expected, the total RNA gel electrophoresis profile from whole cell samples showed two sharp and distinct peaks corresponding to rRNAs. However, the most prominent signal in projection samples was from RNA species of 25–150 nt in size. The RNA integrity number (RIN) [Schroeder et al., 2006] for RNA from projections was reproducibly low, likely due to the large shift in the abundance from 28S and 18S rRNA to 25–150 nt RNA species. Here, I describe my investigation into the meaning of RIN and the small RNA composition of neuronal projections.

Results

The most abundant RNA species in neuro-glial projections are <200 nt long

To characterize RNAs that localize to neuronal projections, we had cultured primary rat hippocampal cells on membranes with 1 µm holes and allowed them to grow in vitro for 14 days (Figure II.1). The membrane served as a physical barrier between cellular projections and somata. Although the projections can pass through the 1 µm holes and continue growing on the underside of the membrane, the cell bodies and nuclei do not and are only detectable on top of the membrane. Thus, by scraping the underside of these membranes, we obtained lysate enriched in neuro-glial projections ("projection" samples) whereas lysate from the top surface of the membrane comprised cell bodies, including nuclei and projections ("whole cell" samples). I extracted RNA from both whole cell and projection samples and subjected it to microcapillary gel electrophoresis (Bioanalyzer Pico RNA kit) to check the size-separated composition, concentration, and integrity of purified RNA.

Predictably, the most intense signal in the size-separated total RNA trace from whole cells was at ~4 kb (28S rRNA; 4802 nt, Hadjiolov et al. [1984]) and ~2 kb (18S rRNA; 1874 nt, Chan et al. [1984]) regions (Figure III.1 A). The highest intensity signal in the projections total RNA trace, however, was between 25 and 150 nt (Figure III.2 A). Although our initial goal was to sequence longer RNAs

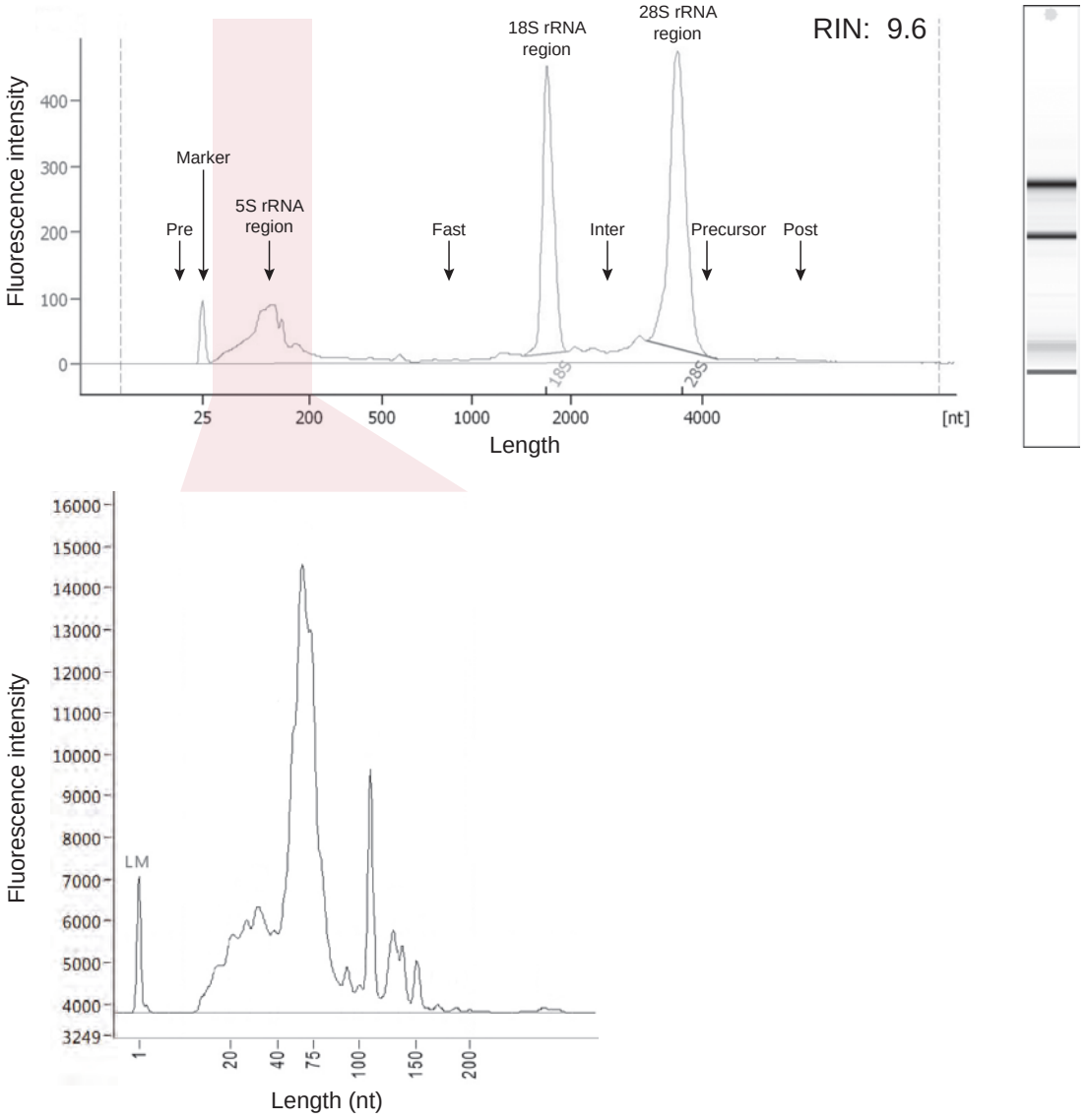


FIGURE III.1: Electrophoresis profile of total RNA (A) and small RNA (B) from whole cell lysate. (A) Total RNA from whole cell samples was subjected to capillary electrophoresis on an Agilent Bioanalyzer (pico RNA kit). The RNA integrity number (RIN) is shown in the upper right corner of the electropherogram. Regions taken into consideration for calculating RIN are labeled. A transformed view of the electropherogram is presented on the right. (B) Small RNAs of <200 nt length were purified and resolved by capillary electrophoresis on an Agilent Fragment Analyzer (small RNA kit). The lower molecular weight marker is labeled as LM. The numbers above the peaks show the approximate sizes of RNA molecules.

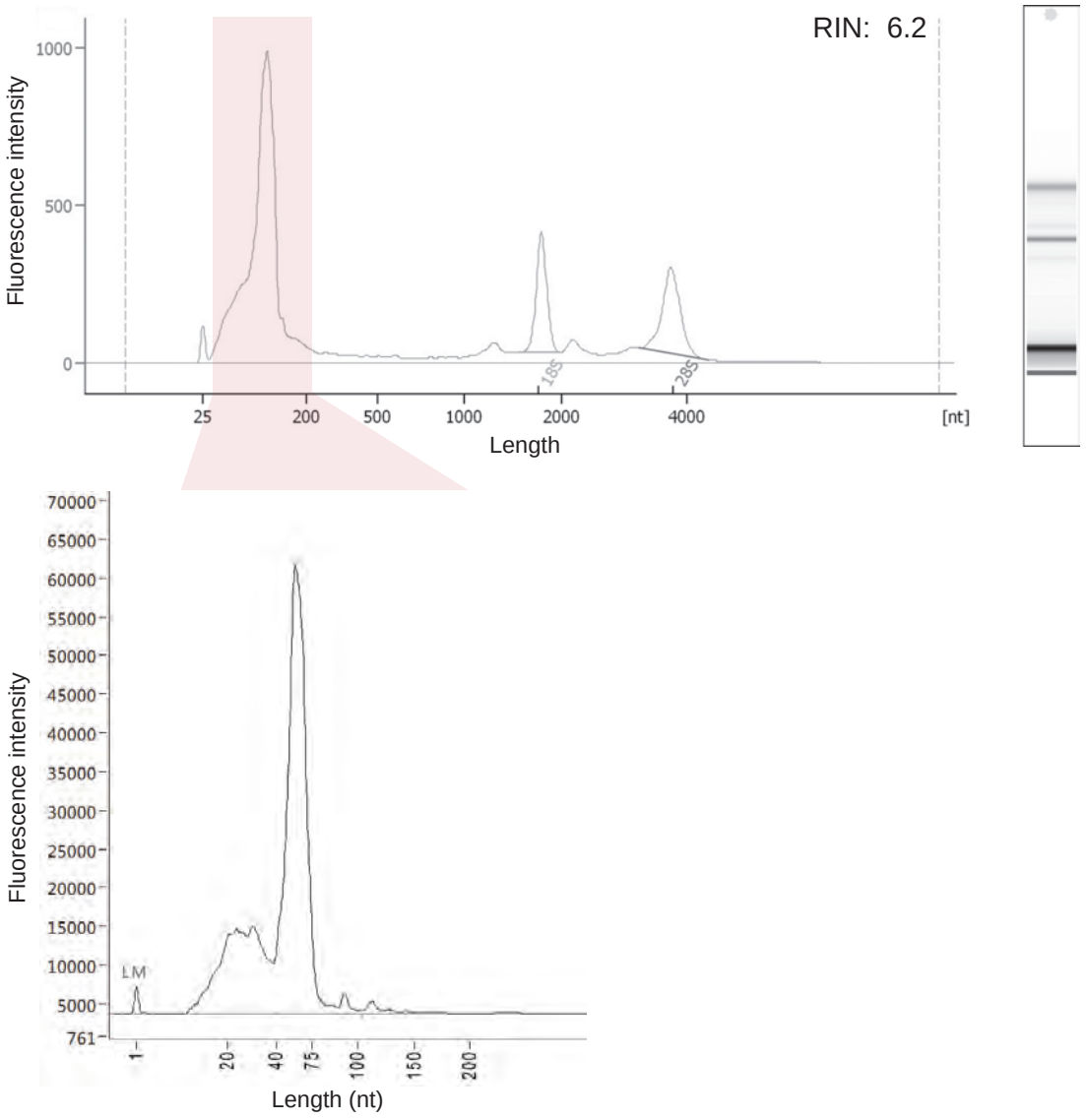


FIGURE III.2: Electrophoresis profile of total RNA (A) and small RNA (B) from projection lysate. Samples were processed the same way as shown in Figure III.1

(>200 nt) from these samples, we could not ignore what appeared to be the most abundant RNA species in projections and decided to investigate the small RNA (<200 nt) composition of neuro-glial projections.

To better resolve small RNAs, I first isolated them using a column (Zymo RNA clean and concentrator) that selectively binds RNAs longer than ~200 nt. The small RNAs can be recovered from the flowthrough by ethanol precipitation. I subjected purified small RNAs to microcapillary electrophoresis (Agilent Fragment Analyzer, small RNA kit) customized to resolve lower molecular weight RNAs (10–200 nt). The projections samples showed a distinct and prominent peak at ~60 nt (Figure III.2 B), whereas whole cell samples showed peaks at ~65 nt and ~110 nt (Figure III.1 B). Note that a margin of error in RNA length estimation is expected since it is inferred by comparing to a ladder containing marker RNAs of known sizes, and not by sequencing. We suspected that the peak at ~60 nt might comprise tRNAs (75–95 nt; Goodenbour and Pan [2006]) and at ~110 nt might contain 5.8S rRNA (156 nt), 5S rRNA (120 nt) or snRNAs (100–300 nt) and decided to check the small RNA composition by sequencing.

Low RIN value does not necessarily imply RNA degradation

The atypical total RNA profile from neuro-glial projections, with a high intensity signal from lower molecular weight RNA species resulted in a low RNA integrity

number (Figure III.2 A). A commonly used signature of intact RNA isolated from cells/tissues is the predominance, 'sharpness' (not smeared), and relative signal intensity of 28S and 18S rRNAs in a total RNA profile from denaturing gel electrophoresis. More recently, the RNA integrity number or RIN has gained popularity as a semiquantitative measure for RNA degradation [Schroeder et al., 2006]. RIN is especially customized to evaluate RNA integrity from widely used automated capillary based electrophoresis devices (for e.g. Bioanalyzer by Agilent Technologies).

The RIN algorithm was parametrized based on total RNA profiles of a variety of human, mouse, and rat tissues. RIN values range from 0 to 10, with lower numbers suggesting greater RNA degradation. The calculation takes into consideration signal intensity, area under the peak, and relative signals at regions labeled in Figure III.1, namely, 5S, 18S, 28S rRNA regions and pre, fast, inter, precursor, and post regions. A degraded RNA sample usually presents decreased signal in the 18S and 28S regions, with a corresponding increase in the inter, fast, and 5S regions due to accumulation of smaller partially hydrolyzed RNA molecules.

In our samples, RIN values for total RNA from whole cells were almost always >9, suggesting minimal RNA degradation, but RIN values for total RNA from projections were consistently low at ~5 (Figure III.3). Although one interpretation

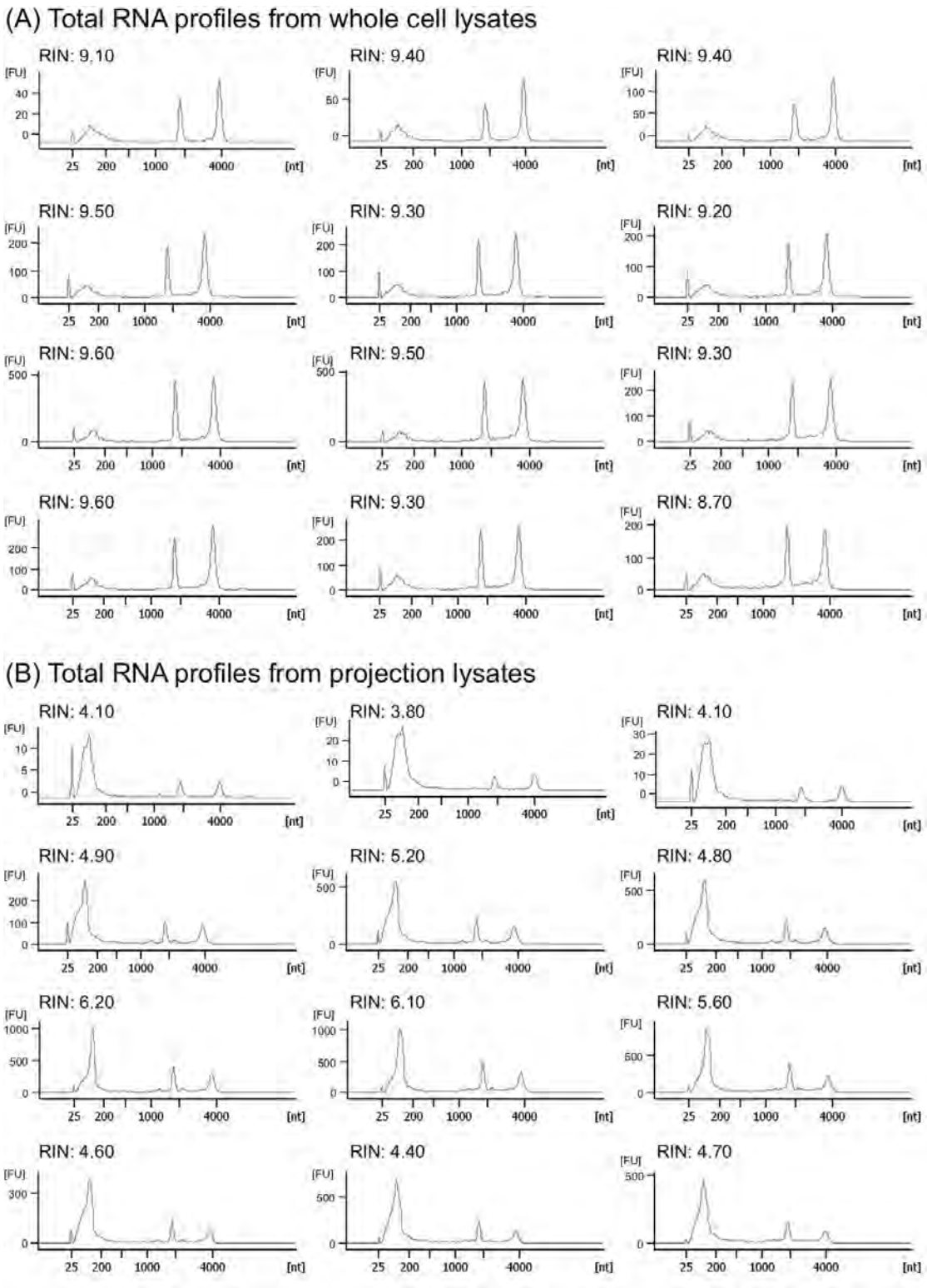


FIGURE III.3: Reproducibility of total RNA electrophoresis profiles from whole cell and projection lysates. Total RNA electrophoresis profiles from 12 biological replicates of whole cell (A) and projection (B) samples are shown. RIN for each profile is shown on the upper left corner of the electropherogram. (FU = Fluorescence Units).

of a low RIN is RNA degradation, it is important to remember the samples (RNA from intact cells/tissues) used to parameterize RIN and the features used for calculating RIN.

We had isolated RNA from a subcellular fraction and the RNA profile showed a strong signal in the 5S region and a relatively weak signal in the 18S and 28S regions. The RIN algorithm expects the highest signal from 28S and 18S regions and a lower signal in the 5S region, thus our projection samples naturally incur a penalty. However, visual examination of the total RNA trace from projections showed several features that are inconsistent with degradation. The 28S and 18S rRNA peaks have sharp boundaries as opposed to being blurry or smeared, indicating rRNAs are likely intact. Partially degraded samples also tend to present a higher intensity signal in the inter and fast regions due to accumulation of low molecular weight RNA fragments. RNA from projection samples did not show an inflated signal in the inter and fast regions.

Additional reasons that argued against RNA degradation were based on how the samples were collected. First, we had extracted RNA from projections and whole cells simultaneously using the exact same protocol and reagents. RNA from whole cell samples appeared reproducibly intact suggesting that inadvertent exposure to RNases or our handling of the samples was unlikely to cause degradation. Second, the total RNA profile from whole cells and projections

was consistent across several biological replicates processed over a span of at least 2 years and remained unchanged despite minor variations in the RNA extraction protocol (Figure III.3). Either RNA degradation in projections samples was highly reproducible or the profile is representative of intact RNAs. Third, the distribution of small RNAs in cellular projections of rat hippocampal cells maintained in culture for 14 days is unknown. Perhaps, in projections, 5S or 5.8S rRNA or tRNAs are highly abundant, or maybe there is an unknown set of RNAs enriched in neuro-glial projections.

Thus, it seemed likely that the low RIN values for projection samples were caused by a relative shift in abundance from 28S and 18S rRNAs to smaller RNAs rather than RNA degradation and it seemed worthwhile to identify the small RNAs composition in neuro-glial projections by sequencing them.

tRNAs abound in neuro-glial projections

To capture a fair representation of small RNAs, which would include tRNAs based on the size distribution, we used Takara's SMARTer smRNAseq protocol [Zhu et al., 2001] (Figure III.4). The libraries were prepared followed manufacturer's instructions. Briefly, target RNAs were polyadenylated followed by reverse transcription using an oligo dT primer with an Illumina sequencing compatible adaptor overhang. The reverse transcriptase used in this protocol has terminal

transferase activity such that it adds non-templated nucleotides to the cDNA upon releasing the RNA template. An Illumina compatible sequencing adaptor is incorporated to the 3′ end of the cDNA by template switching. The cDNA library was amplified using 8 PCR cycles and sequenced on Illumina's HiSeq 4000 machine to get 100 nt single-end reads.

To test whether the sequenced libraries included a representative fraction of input RNA, we checked the read length distribution (Figure III.5), which satisfactorily mimicked the RNA length distribution in Figures III.1 B and III.2 B. We then proceeded to align the reads sequentially to annotated classed of RNA: rRNAs, tRNAs, snRNAs, repeat elements (cataloged by RepeatMasker [Jurka, 2000]), and the nuclear genome (Figure III.6).

On average, ~36% of the reads from projections samples aligned to tRNAs compared to ~16% from whole cell samples. In contrast, as a class, the least abundant small RNAs in projections were snRNAs, with ~3% of the reads aligning to them as compared to ~16% from whole cells. snRNAs are integral components of the spliceosome. Although some snRNAs undergo post-transcriptional processing in the cytoplasm, their depletion from projections is expected because the spliceosome assembles in the nucleus [Will and Lührmann, 2001].

In sequencing long RNAs from projections, we had found that >50% of our reads aligned to the mitochondrial genome, which is consistent with the high abundance

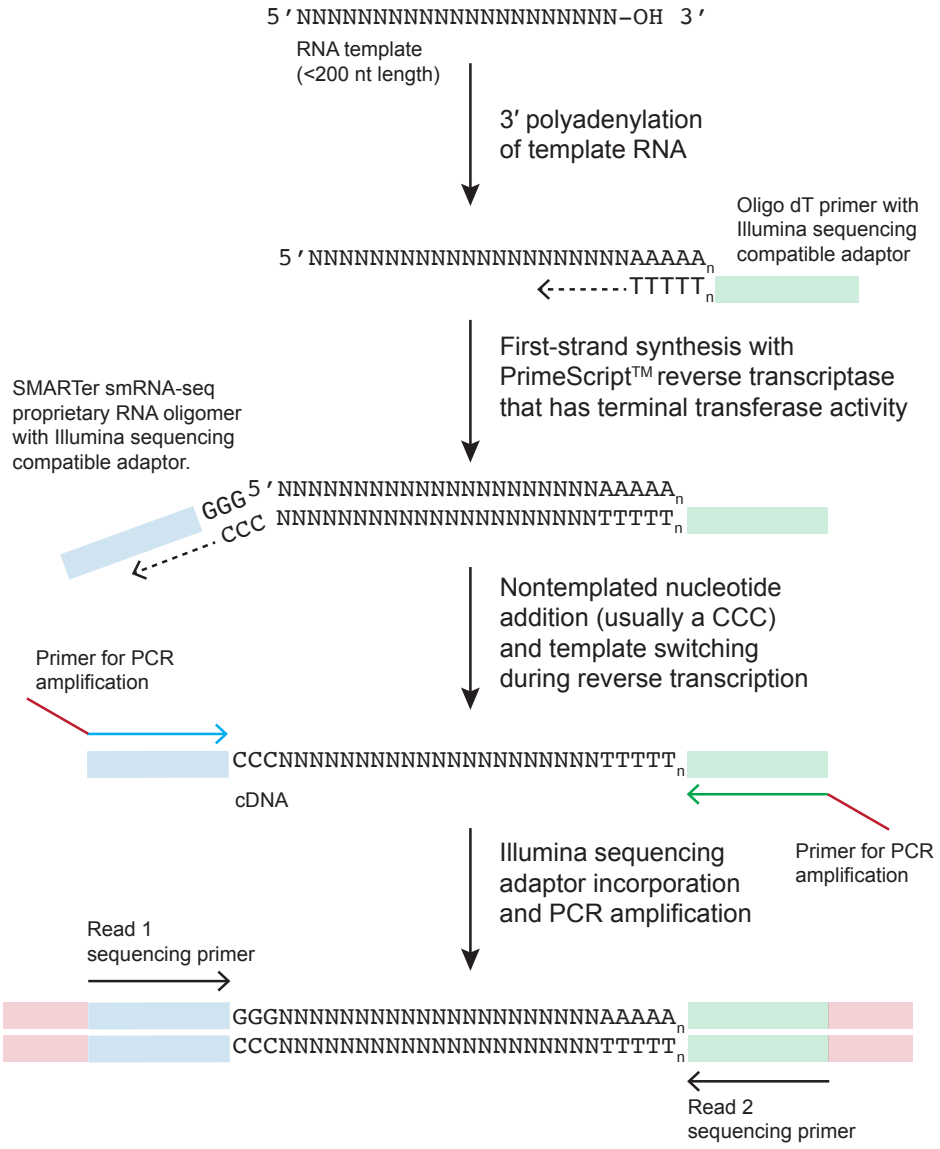


FIGURE III.4: Library preparation schematic for small RNAseq using SMARTer smRNA-Seq Kit (Takara Bio). This figure is an adaptation of Figure 1 in the SMARTer smRNA-Seq user manual.

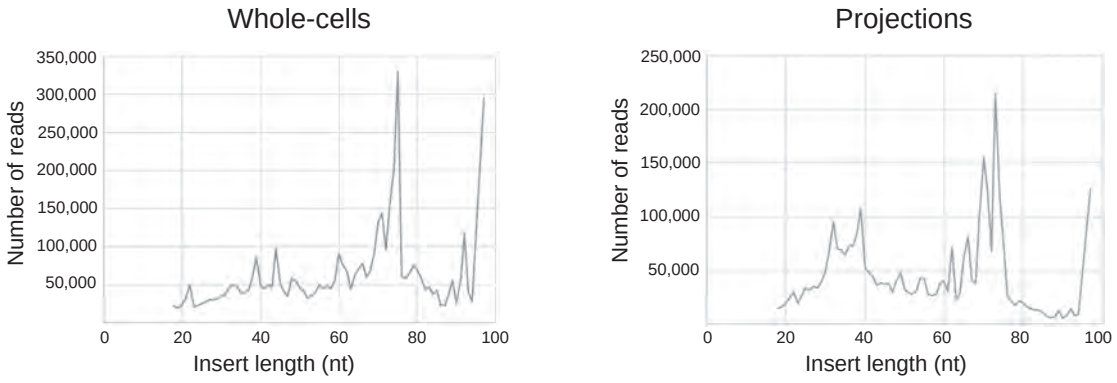


FIGURE III.5: Read length distribution of sequenced RNAs from whole cells and projections. RNAseq libraries were sequenced to obtain single-end 100 nt long reads. The insert length distribution is shown here after removing 3 nts from the 5′ end (presumably incorporated due to reverse transcriptase's terminal transferase activity) and the sequence including and following a stretch of 6 or more adenosines from the 3′ end of reads. The spike at ~100 nt represents RNAs equal to or longer than ~100 nt.

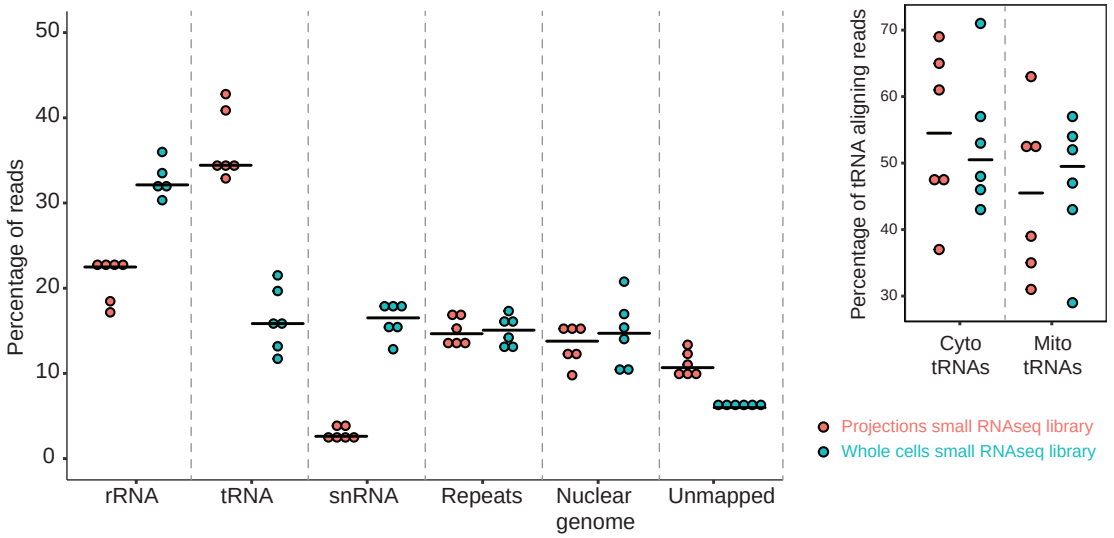


FIGURE III.6: Read alignment classification. Percentage of reads aligned to rRNA (45S), tRNAs, snRNAs, repeat elements (cataloged by RepeatMasker [Jurka, 2000]), and the nuclear genome are shown. Each filled circle represents a small RNAseq library from biological replicates of projection (red) or whole cell (cyan) samples. The plot on the upper right shows the percentage of tRNA aligning reads that aligned to cytoplasmic (cyto) or mitochondrial (mito) tRNAs. The horizontal black bars mark the median percentage.

of mitochondria in neuronal projections (Chapter II; [Palay, 1956]). We wondered whether projection-localized tRNAs were predominantly mitochondrial and found that, of all the reads aligning to tRNAs, on average only ~44% from projection samples and ~46% from whole cell samples aligned to mitochondrial tRNAs (Figure III.6). For translation to occur in projections, in addition to mRNA localization, the translation machinery including ribosomes and tRNAs must be localized. To ask whether all essential tRNAs were localized to projections, we decided to examine the cytoplasmic tRNA composition in more detail.

Neuro-glial projections comprise all essential tRNA isodecoders

To infer tRNA expression from RNAseq data, read alignment only to the genome is inadequate. tRNAs are short and many have multiple copies in the genome, which makes aligning reads uniquely to a single genomic locus challenging, if not impossible. Moreover, some tRNAs contain introns that are absent from mature tRNAs, necessitating proper intron annotation and use of a splice-sensitive aligner. Mature tRNAs also contain a non-genomically-encoded sequence at their 3′ end, for example, a CCA (or a CCACCA for those targeted for degradation), which can hinder read alignment to the genome. To complicate things further, most tRNAs contain modified nucleotides that are often misread by the reverse transcriptase. So, on the one hand, to overcome reads aligning to multiple tRNA loci, high alignment stringency is preferred but on the other hand, alignment

stringency must be relaxed to allow alignments with mismatches due to tRNA modifications.

An effective strategy for tRNA quantification from short RNAseq that overcomes many of these challenges was recently developed by Cozen et al. [2015] (ARM-Seq pipeline). Instead of aligning reads only to the genome, the ARM-seq pipeline uses genomic tRNA annotation to prepare a reference database of mature tRNA sequences and aligns reads to both mature tRNAs and the genome (Figure III.7). ARM-Seq output includes convenient visual representations of the alignments as well as quantified estimates for tRNA expression and a classification of reads aligning to pre-tRNAs, mature intact tRNAs and tRNA fragments.

ARM-Seq requires tRNA annotation by tRNAScan-SE [Lowe and Eddy, 1997] as input. We therefore annotated tRNAs in the rat genome (Ensembl release 81, Rnor_6.0) and identified 1,747 tRNA loci. To put the number of rat tRNA loci in perspective, 504 tRNA loci were identified in the 2001 draft of the human genome using the same program [Lander et al., 2001] (Figure III.8, an adaptation of Figure 34 in Lander et al. [2001]). The inflated tRNA locus counts in the rat genome are largely due to an expansion in ProCCU and AlaGCU tRNAs, which had 407 and 894 loci in rat versus 11 and 25 in human. Whether all of these tRNA loci are transcribed or only a subset, is unknown. The annotation alone

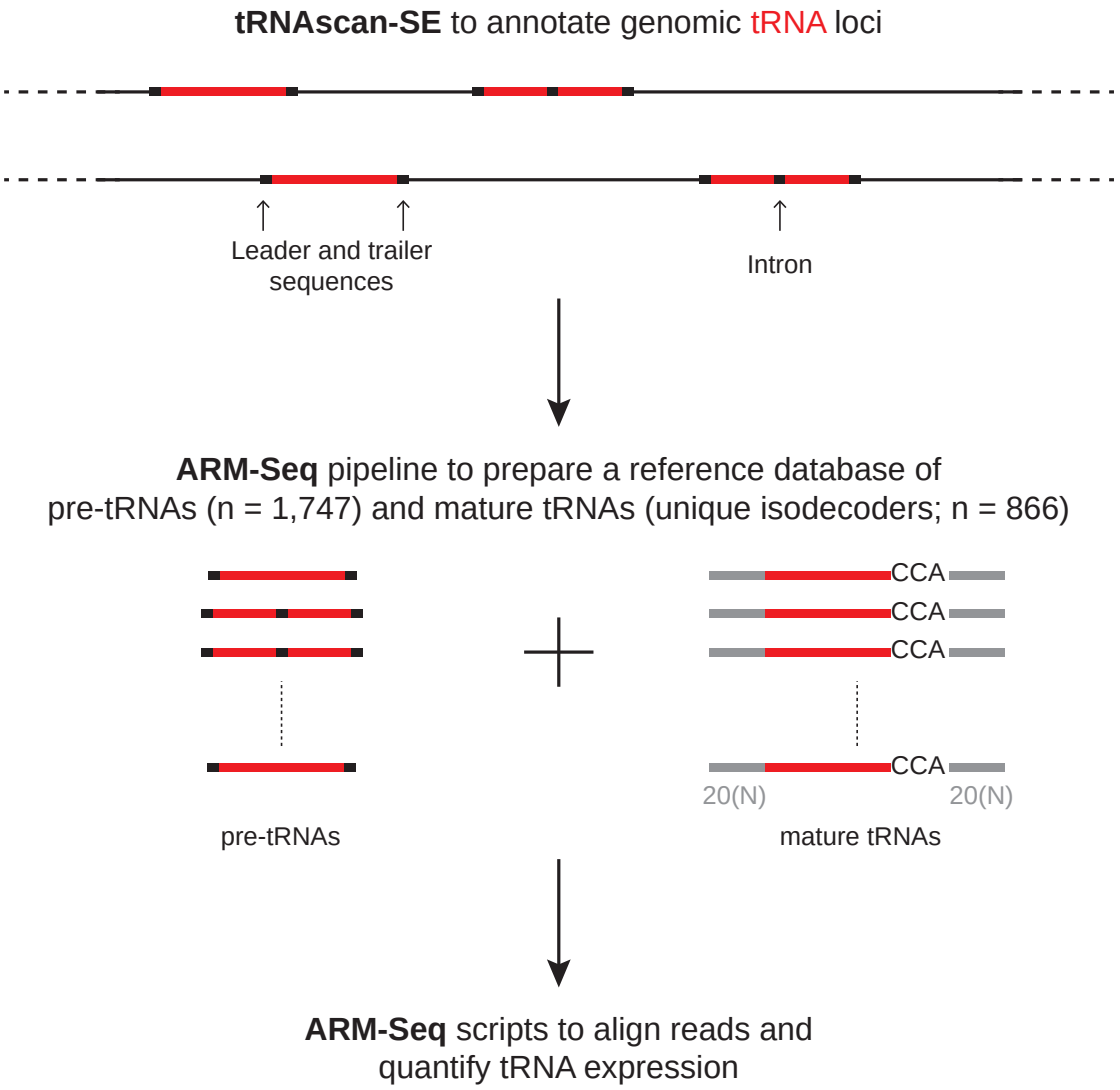


FIGURE III.7: Schematized ARM-Seq tRNA alignment pipeline. Genomic tRNA loci were identified using tRNAscan-SE. ARM-Seq extracts the tRNA loci from tRNAscan output to generate a reference comprising pre-tRNAs and mature tRNA (devoid of introns, leader/trailer sequences, with CCA at the 3′ ends, and 20 N nucleotides at both ends of the tRNAs). Reads are aligned to this reference and the alignments are classified as pre-tRNAs, whole tRNAs, or tRNA fragments.

Amino acid	Codon	Anticodon	No. of human tRNA loci	No. of rat tRNA loci	No. of unique tRNAs	No. of expressed tRNAs
Phe	UUU	AAA	0	0	0	0
Phe	UUC	GAA	14	8	4	4
Leu	UUA	UAA	8	2	2	2
Leu	UUG	CAA	6	5	3	2
Leu	CUU	AAG	13	10	7	2
Leu	CUC	GAG	0	0	0	0
Leu	CUA	UAG	2	3	3	3
Leu	CUG	CAG	6	9	4	4
Ile	AUU	AAU	13	10	5	2
Ile	AUC	GAU	1	1	1	0
Ile	AUA	UAU	5	3	2	2
Met	AUG	CAU	17	14	7	5
Val	GUU	AAC	20	35	21	2
Val	GUC	GAC	0	0	0	0
Val	GUA	UAC	5	3	2	1
Val	GUG	CAC	19	7	5	3
Ser	UCU	AGA	10	30	22	4
Ser	UCC	GGA	0	0	0	0
Ser	UCA	UGA	5	4	4	4
Ser	UCG	CGA	4	3	3	3
Pro	CCU	AGG	11	407	158	2
Pro	CCC	GGG	0	2	1	0
Pro	CCA	UGG	10	7	3	2
Pro	CCG	CGG	4	3	1	1
Thr	ACU	AGU	8	21	17	5
Thr	ACC	GGU	0	0	0	0
Thr	ACA	UGU	10	6	5	4
Thr	ACG	CGU	7	4	4	4
Ala	GCU	AGC	25	894	441	7
Ala	GCC	GGC	0	4	4	0
Ala	GCA	UGC	10	12	9	4
Ala	GCG	CGC	5	7	6	3
Tyr	UAU	AUA	1	0	0	0
Tyr	UAC	GUA	11	3	2	2
*	UAA	UUA	0	0	0	0
*	UAG	CUA	0	0	0	0
His	CAU	AUG	0	0	0	0
His	CAC	GUG	12	11	2	1
Gln	CAA	UUG	11	6	5	4
Gln	CAG	CUG	21	10	6	6
Asn	AAU	AUU	1	0	0	0
Asn	AAC	GUU	33	15	8	5
Lys	AAA	UUU	16	6	3	3
Lys	AAG	CUU	22	17	8	6
Asp	GAU	AUC	0	2	2	0
Asp	GAC	GUC	10	14	3	3
Glu	GAA	UUC	14	10	5	4
Glu	GAG	CUC	8	10	4	2
Cys	UGU	ACA	0	1	1	0
Cys	UGC	GCA	30	38	14	5
*	UGA	UCA	0	0	0	0
Trp	UGG	CCA	7	7	6	6
Arg	CGU	ACG	9	6	3	4
Arg	CGC	GCG	0	0	0	0
Arg	CGA	UCG	7	5	4	4
Arg	CGG	CCG	5	3	3	3
Ser	AGU	ACU	0	1	1	0
Ser	AGC	GCU	7	16	11	3
Arg	AGA	UCU	5	7	7	6
Arg	AGG	CCU	4	7	5	5
Gly	GGU	ACC	0	11	8	0
Gly	GGC	GCC	11	12	4	3
Gly	GGA	UCC	5	9	2	2
Gly	GGG	CCC	8	5	4	4

FIGURE III.8: tRNA annotation and expression in rat hippocampal cells. Each amino acid is encoded by one or more three-letter codons. All 64 codons are shown in this figure, with corresponding tRNA anticodons and the amino acids they encode. Stop codons are represented by an asterisk. tRNAs were annotated using tRNAScan-SE. A total of 1,747 tRNA loci were identified in the rat genome, which comprised 894 unique tRNA isodecoders (No. of unique tRNAs), a subset of which showed coverage in RNAseq data (No. of expressed tRNAs; red text; Appendix A). The black lines point to wobble base pair usage. Black rectangles enclose expressed isodecoders from tRNAs expected to decode by wobble rules. This figure was inspired by Figure 34 from Lander et al. [2001] and the number of human tRNA loci are copied from the same reference.

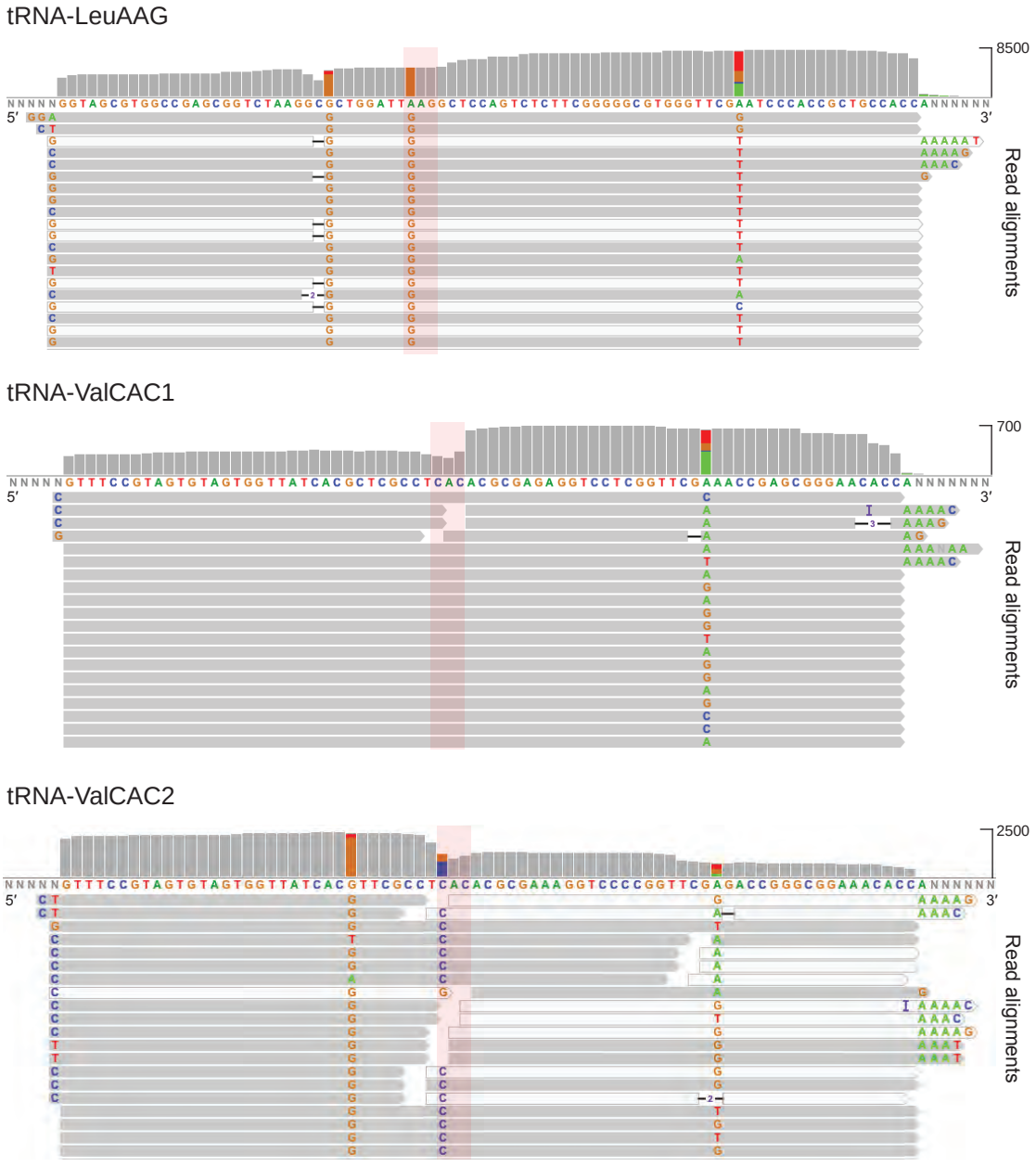


FIGURE III.9: Full length and fragmented tRNAs with mismatches, deletions, or truncations at modified nucleotides. Read alignments to three representative tRNAs are shown. For each tRNA, the top panel shows a histogram of read coverage at each nucleotide, with the maximum number of reads shown in the upper right corner. The tRNA sequence is shown below the histogram and selected individual read alignments are shown as horizontal bars. The shading of each bar is representative of the alignment quality, darker color representing a better alignment. Mismatched nucleotides are labeled in the alignment bars and deletions are shown as black dashes. Mismatches are also shown in the histogram as colored bars with each nucleotide assigned a specific color (A = green, C = blue, G = orange, T = red). The anticodon is highlighted in pink.

does not provide information about transcriptional activity but RNAseq data can. To determine the number and types of isodecoders (same anticodon but different tRNA sequence) expressed in our samples, we proceeded to align RNAseq reads to the tRNA reference following the ARM-seq pipeline.

Satisfyingly, in agreement with wobble rules, we found that reads from both whole cell and projections samples covered all 45 of the 61 tRNA isodecoders whereas the few wobble base pairing anticodon tRNA loci identified by tRNAScan-SE (for example, IleAUC, ProCCC, etc.) showed no coverage [Lander et al., 2001] (Figure III.8). Additionally, for tRNAs ProCCU and AlaGCU, 158 and 441 unique isodecoders were identified, of which only 2 (from 8 identical loci) and 7 (from 12 identical loci) isodecoders were expressed.

Of the tRNAs represented in RNAseq data, many had individual reads aligning across the entire length of the tRNA with signatures of modified nucleotides appearing as mismatches or deletions in read alignment (Figure III.9). For example, an adenosine (A) at the wobble position in the anticodon (shaded in pink) is often enzymatically deaminated to form inosine (I), which is read as guanine by most commercially available reverse transcriptases [Limbach and Paulines, 2017]. In tRNA-LeuAAG, which is expected to undergo A to I modification according to the wobble rules (Figure III.9), 100% of the reads at the wobble adenosine show guanine. Many reads aligning to LeuAAG also

show 1 or 2 nt deletions at the ~27th nucleotide position from the 5′ end. This is another common tRNA modification site where the guanine is modified to a N^2,N^2-dimethyl-G [Pan, 2018] in human cells. Detecting potential tRNA modifications as alignment mismatches was reassuring because it suggested that we had indeed captured mature tRNAs in our dataset.

Despite the slew of tRNA modifications encountered by the reverse transcriptase, many reads spanned the entire length of tRNA molecules. However, we also observed reads that only partially covered some of the tRNAs (Appendix A). The partial coverage was in general skewed toward the 3′ ends of tRNAs (e.g. tRNA-ValCAC1 in Figure III.9) but there were clear cases where the 5′ ends showed a greater number of reads aligning to them (e.g. tRNA-ValCAC2 in Figure III.9). The 3′ skew in coverage could be due to the reverse transcriptase occasionally disengaging from the template at modified or highly structured sites, or it could arise from tRNA fragments. However, for reads aligning to the 5′ ends of tRNAs, the most likely explanation is the existence of tRNA fragments with a 3′-OH that were available for polyadenylation, reverse transcription, and amplification for inclusion into the sequenced libraries. The sequencing method used here requires a 3′-OH, therefore tRNA fragments with a 3′ phosphate are expected to be underrepresented in these libraries.

Thus, these data provide evidence for the presence of all essential tRNA

isodecoders and 5′ tRNA fragments in neuro-glial projections. Although we cannot resolve the source of apparent 3′ fragments, further analyses of the data could provide clues to whether the 3′ reads originated from fragments or due to reverse transcriptase's premature disengagement from the template.

Discussion

In this chapter, we investigated the composition of small RNAs (<200 nt) that appeared strikingly abundant in neuronal projections (Figure III.1 and III.2). This study emerged from a curious observation in Chapter II, where we had extracted RNA from neuro-glial projections to study localized intron sequences: in analyzing RNA integrity by microcapillary electrophoresis, RNA extracted from cellular projections showed an exaggerated signal between 25 and 150 nt (or 5S region). High signal in the 5S region did not appear to be due to RNA degradation and we decided to identify the small RNA composition by sequencing them.

By using a sequencing method (Figure III.4) that could capture a fair representation of small RNAs from whole cells and projections, we found that ~36% of reads from projection samples aligned to both cytoplasmic and mitochondrial tRNAs (Figure III.6). Since translation does occur in neuro-glial projections, it was unsurprising to find tRNAs, an essential component of the translation machinery, in these samples. Deeper examination of the data

revealed that all essential tRNAs were detectable in projection samples (Figure III.8). Interestingly, we also found evidence for fragmented tRNAs in whole cells as well as neuro-glial projections (Figure III.9).

Although we answered the question about the composition of the prominent band by RNAseq, further analyses of these data can provide valuable insight to the types of tRNA modifications and the nature of tRNA fragments found in primary rat hippocampal cells and their projections. In mammalian cells, recent work has implicated tRNA fragments translation regulation [Yamasaki et al., 2009], cell proliferation [Blanco et al., 2016, Goodarzi et al., 2015], and epigenetic inheritance [Sharma et al., 2016]. Just by looking at read coverage across all tRNAs (Appendix A), it appears that many tRNAs fragment in the anticodon loop. A comprehensive examination of read ends could provide clues about the sites and mechanism of tRNA fragmentation.

Acknowledgments

I am grateful to the deep sequencing core facility at University of Massachusetts Medical School (UMass) for HiSeq sequencing (and resequencing with no charge!); Massachusetts Green High Performance Computing Center (UMass) for computational resources; Alicia Bicknell, Phillip Zamore, Kate Shulgina and members of the Eddy lab (Harvard University) for constructive feedback on

this research. This work was supported by funding from HHMI (S.R.E.), NIH (R01-GM53007 to M.J.M.), and HHMI International Student Research Fellowship (H.S.). M.J.M. was an HHMI Investigator at the time this study was conducted.

CHAPTER IV

Discussion

Research presented in this dissertation was aimed at characterizing intron sequences that localize to neuronal projections (Chapter II) and investigating the composition of a particularly abundant set of RNAs that we observed during our exploration of total RNA isolated from neuro-glial projections (Chapter III). Here, I discuss the implications of our findings, the limitations of our experiments, new questions that originate from this work and possible future research directions.

Two main results emerge from this dissertation research:

1. Although intron lariats are thought to be degraded fairly quickly after splicing [Ooi et al., 2001, Ruskin and Green, 1985], several free circular introns with characteristics of a lariat can localize to neuronal projections. Features common to these projection-localized circular introns are the usage of a cytosine branchpoint, lack of ribosome occupancy, lack of sequence conservation, and lack of any recognizable RNA localization signals.

2. The most abundant class of RNAs in projections of primary rat hippocampal cells included short RNAs of less than ~150 nt length. Cytoplasmic and mitochondrial tRNAs predominated in the pool of projection-localized short RNAs. tRNAs with all essential anticodons (in agreement with the wobble hypothesis; Crick [1966]) were represented in neuronal projections.

Introns in neuro-glial projections

Chapter II of this dissertation described our exploration of intron sequences that localize to neural projections (axons, dendrites, and some glial projections) of rat hippocampal cells in primary culture. The search for projection-localized intron sequences was motivated in part by the observation that intron retention might be particularly prevalent in neurons [Braunschweig et al., 2014] and a provocative suggestion that splicing can occur in dendrites [Glanzer et al., 2005]. Regardless of the question of splicing, retained introns can impact mRNA localization [Sharangdhar et al., 2017], stability [Colak et al., 2013], or the translation products [Chen et al., 2008], whereas excised introns can serve noncoding functions [Curtis et al., 2012, Morgan et al., 2019, Ooi et al., 1998, Parenteau et al., 2019]. An exhaustive account of cytoplasmic intron sequences from neurons could yield insight to general mechanisms of RNA splicing and localization. Therefore, as a first step toward understanding the origin, function, and fate of cytoplasmic introns, I sought to determine a comprehensive set of intron sequences that localize to neuronal projections.

Culturing primary cells on membranes with small pores (1 µm diameter) allowed for the physical separation of cellular projections from the somata; while the projections could pass through the pores, the somata could not. Using a conventional RNAseq method, we sequenced rRNA-depleted total RNA [Zhang

et al., 2012]. We captured ~200 nt or longer, adenylated and nonadenylated, linear and circular RNAs. We also sequenced the remaining shorter (<200 nt) RNAs [Zhu et al., 2001] because of their high abundance in projections (Chapter III). Comparative analysis of total RNA extracted from the projections versus whole-cells allowed us to identify projection-enriched RNAs. For a deeper understanding of localized RNAs, we complemented RNAseq datasets with polyA+ RNAseq, ribosome profiling, and polyA-site sequencing libraries from the same cells.

Thus, by integrating information from a variety of datasets, we provide evidence that neuronal projections contain transcripts with retained introns, as well as a diverse set of free introns. The presence of free introns outside of the nucleus was unexpected and raised questions about their origin, stability, nuclear export, localization, and function. These introns showed no evidence of ribosome occupancy and were not represented in polyA+ RNAseq, indicating that they are noncoding and nonadenylated RNAs. Circularly permuted read alignment on these introns suggested that they were tailless lariats with a noncanonical cytosine branchpoint. Similar circular introns from a different set of genes were recently observed by Talhouarne and Gall [2018] in a variety of vertebrate cell types. Although pervasive, the intron sequences we found lack conservation. Furthermore, their functions remain enigmatic.

In considering the possibility of splicing in dendrites, the location of origin of these circular introns becomes an important question: were they excised in the nucleus and then transported to projections independent of their mRNA or were they spliced locally in cellular projections? Given that these introns are detectable in the nucleus, a parsimonious explanation is that they move to projections after splicing in the nucleus, instead of the other way around. The apparent lack of an unspliced template (no reads across exon-intron or intron-exon boundaries) in projections further supports this idea. An alternative is that splicing in projections occurs so quickly that the unspliced template is undetectable. However, if splicing were so fast, there would appear to be no regulatory advantage to localizing an unspliced transcript. Furthermore, there is no evidence yet for the presence of the spliceosome in neural projections. Noncanonical mechanisms of splicing might be hypothesized to mediate intron removal from the mRNA in projections. In that case, would a new mechanism of splicing use the same splice sites as the spliceosome? The existence of a alternate splicing pathway that recognizes spliceosomal splice sites seems unparsimonious.

Instead, the possibility that these circular introns originated in the nucleus seems plausible and can be tested. We propose that the circular form of these introns might inhibit their degradation and contribute to their distribution to cellular extremities, perhaps simply by diffusion. Although the spliceosome can utilize a cytosine branchpoint [Taggart et al., 2017], the debranching enzyme is unable

to hydrolyze a 2′-5′ bond at a cytosine [Jacquier and Rosbash, 1986], which could result in a circular RNA that is resistant to exonucleolytic degradation. If that were the case, then mutating the branchpoint C to the canonical A might linearize these introns and render them susceptible to exonucleolytic degradation. Consequently, their localization to distal cellular projections would be expected to be impaired. Instead, if these introns were localized by some unknown active transport, then changing the branchpoint might alter the stability but not necessarily their localization.

The question then arises whether all introns with a C branchpoint are stable and localize to cellular projections. If so, then the expectation would be that all lariats with a C branchpoint present in the nucleus would be detectable in projections. Exceptions might include lariats with a short half-life. Comprehensive analysis of branchpoint usage in nuclear and cytoplasmic lariats could provide clues about the role of C branchpoint usage in intron localization. However, identifying lariat branchpoints using standard sequencing methods is challenging because most reverse transcriptases (to prepare cDNA, usually an early step for any sequencing protocol) introduce mutations, deletions, or fall off the RNA when they encounter a 2′-5′ bond. Enriching for circular RNA species prior to cDNA preparation and using a permissive reverse transcriptase has been used previously to map lariat branchpoints in mammalian cells [Taggart et al., 2017], and in principle, could be used here as well.

Thus, my results and those of others [Talhouarne and Gall, 2014, 2018] indicate that circular introns are more widespread than previously anticipated. Tampering with the localization of these circular introns, one way or another as proposed here, could provide useful information about the origin of these introns and perhaps even provide clues about their functionality.

tRNAs in neuro-glial projections

While exploring the total RNA landscape of neuronal projections, a population of RNAs of <150 nt length stood out. Were these RNAs degradation intermediates or an unknown class of intact RNAs enriched in projections? Chapter III provides evidence that argues against a degraded RNA sample. Sequencing the small RNAs from neuronal projections showed that on average, ~36% of the reads aligned to tRNAs, as compared to ~16% from whole cell samples. In fact, all essential anticodon tRNAs are represented in RNAseq data from neuronal projections. This makes sense because for translation to occur in distal cellular locations, tRNAs are required just as much as mRNAs, ribosomes, and translation factors.

Systematic analysis of tRNA expression required using unconventional methods for read alignment and abundance estimation. tRNAs are diverse and each one can have multiple copies in the genome, therefore aligning reads directly to the

genome alone is ineffective because reads might not align uniquely to one locus. tRNAs also undergo extensive post-transcriptional processing such that reads from a mature tRNA might fail to align to the genomic sequence. Fortunately, Cozen et al. [2015] recently developed a computational pipeline to assess tRNA expression from sequencing data. Following guidelines presented in Cozen et al. [2015] allowed me to estimate tRNA expression and abundance in our samples. I found many reads that aligned end-to-end to mature tRNAs as well as those that aligned to parts of a tRNA in a pattern that indicated tRNA fragmentation.

Lately, tRNA fragments have received much attention for their roles in transcription [Sharma et al., 2016] or translation regulation [Thompson and Parker, 2009, Yamasaki et al., 2009], and cell differentiation [Blanco et al., 2016, Goodarzi et al., 2015]. Non uniform read coverage on some of the tRNAs from neuronal projections indicate the presence on tRNA fragments (Appendix A). Identifying the types of tRNA fragments could provide clues regarding the fragmentation mechanism or their potential functionality. For instance, stress induced cleavage of tRNAs can have a global impact on translation [Yamasaki et al., 2009]. On the other hand, cleavage of a subset of tRNAs with a specific anticodon could, in theory, affect translation efficiency of mRNAs containing corresponding codons. Thus, examining the cleavage sites and the ratio of tRNA fragments to intact tRNAs could be insightful for our understanding of translation regulation, especially in neuronal projections.

Many open questions remain that can be answered by further analyses of these data. Regarding tRNAs in the unique geometries of neurons, characterizing tRNA fragments and modifications in projection versus whole cell samples could provide novel insight regarding tRNA localization, processing, and their impact on local translation. More broadly, these datasets can be mined to study the expression and distribution of other classes of projection-localized small RNAs.

General limitations and closing remarks

A caveat of the experiments presented in this dissertation is that we do not know the volume of the lysate from which RNA was extracted, therefore, we can only measure relative abundances of RNAs in each sample. An actively localized RNA would be expected to appear "enriched" relative to other RNAs, but so can a homogeneously distributed RNA if a majority of RNAs are depleted from projections. It is important to note that RNAs can diffuse to cellular extremities (for e.g., Nanos mRNA localization in Drosophila embryo depends on its diffusion; Forrest and Gavis [2003]), especially if the RNA is stable and not sequestered to subcellular locations either due to transport by motor proteins or by association with non-uniformly distributed molecular complexes.

Another limitation of these experiments arises from the heterogeneous composition of our primary hippocampal cell cultures. Although we observe both

neuronal and glial projections passing through the holes in the membranes on which they are cultured, it is possible that they represent only a subset of all the cells in culture. Thus the relative cellular composition on the projections side might differ from that on the whole cell side.

RNA processing and localization could vary in different cell types too. RNAs that are localized to projections by all cells might show an exaggerated enrichment in projections than those that are localized by only a subset of cells. Perhaps the extreme enrichment of ribosomal protein encoding RNA (RP mRNA) observed by us and others is a result of RP mRNAs being localized to projections of all cells as opposed to an RNA that is expressed by a specific cell type.

Not knowing whether a specific cell-type or all cells expressed the RNA molecules of interest can also limit our ability to understand the functional relevance of localized RNAs. For example, we are unable to determine whether all circular introns are expressed by one cell type or if different cells express different introns. Although we were able to confirm the localization of a few candidate mRNAs and introns by smFISH, a more comprehensive atlas of localized RNAs would be a better indicator of RNA enrichment in individual cellular projections.

Similar limitations apply to our analyses of tRNA localization in these cells. Collectively, we find all tRNAs represented in projections but it is possible that different cells express different subsets. If such differences exist, it would be

important to know because they could impact translation efficiency and might even serve as a mechanism to regulate translation differently in different cells.

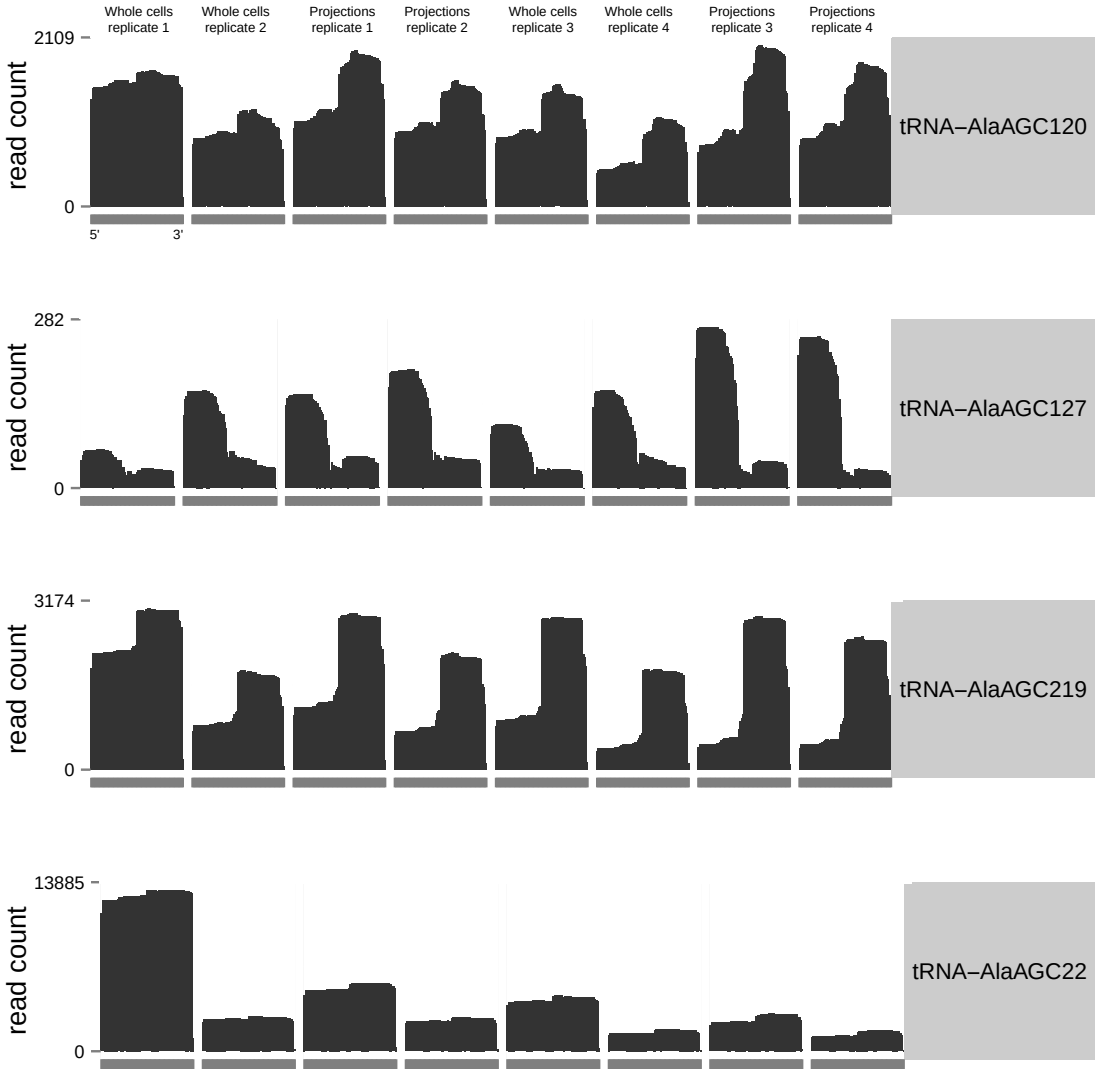
The aforementioned caveats are important considerations in interpreting our results. A recurring theme is the heterogeneity present in our cultures, among cell types and cellular compartments within a type. We could have used more homogeneous cultures. However, the drawback of that approach is that homogeneity is highly artificial, and could influence the kinds and distributions of RNAs in a way that is non-physiological due to a lack of interactions among different cell types. Moving forward, it would be important to examine RNA types and localizations in the context of normal tissues at subcellular resolution. This thesis provides a set of tools, first-principle-validations, and likely RNA candidates to help guide such an examination.

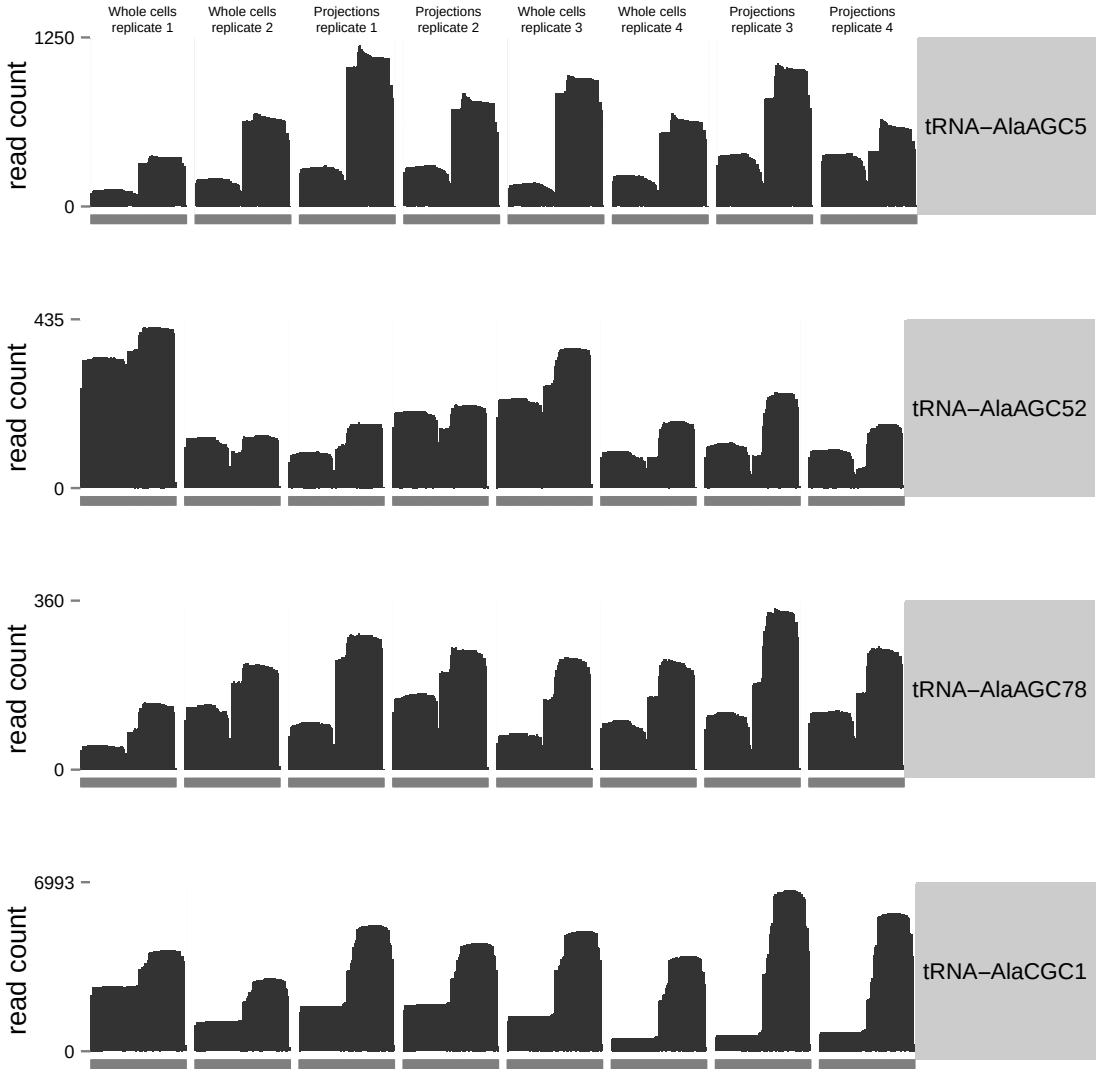
Appendix A

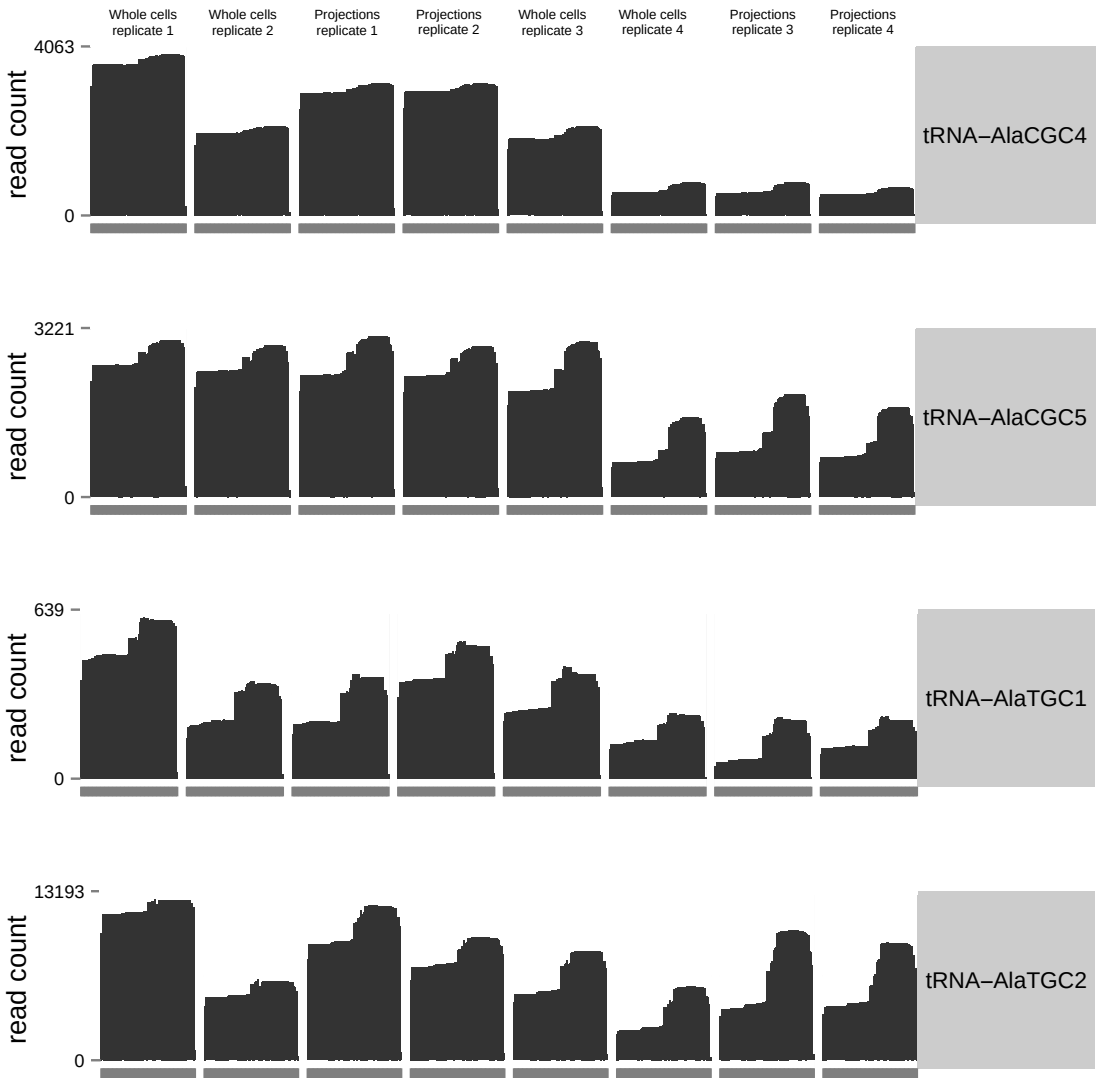
Appendix A: Read coverage on tRNA isodecoders

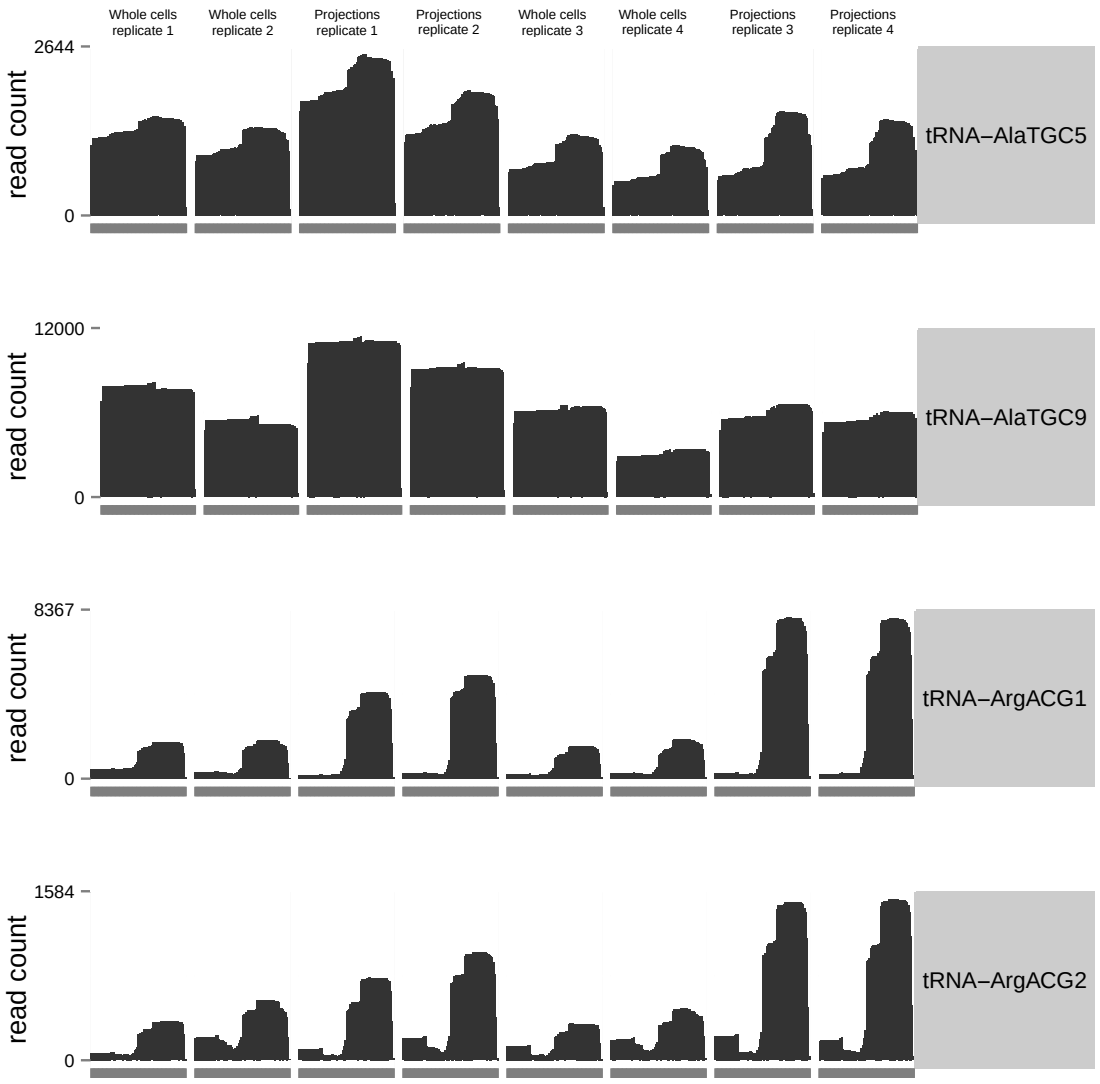
A histogram of read coverage at each nucleotide for all tRNA isodecoders that are expressed in our samples are shown for four biological replicates of whole cell and projection samples. Horizontal gray bars under the histogram depict 5′ to 3′ orientation of the tRNA labeled on the right in each row.

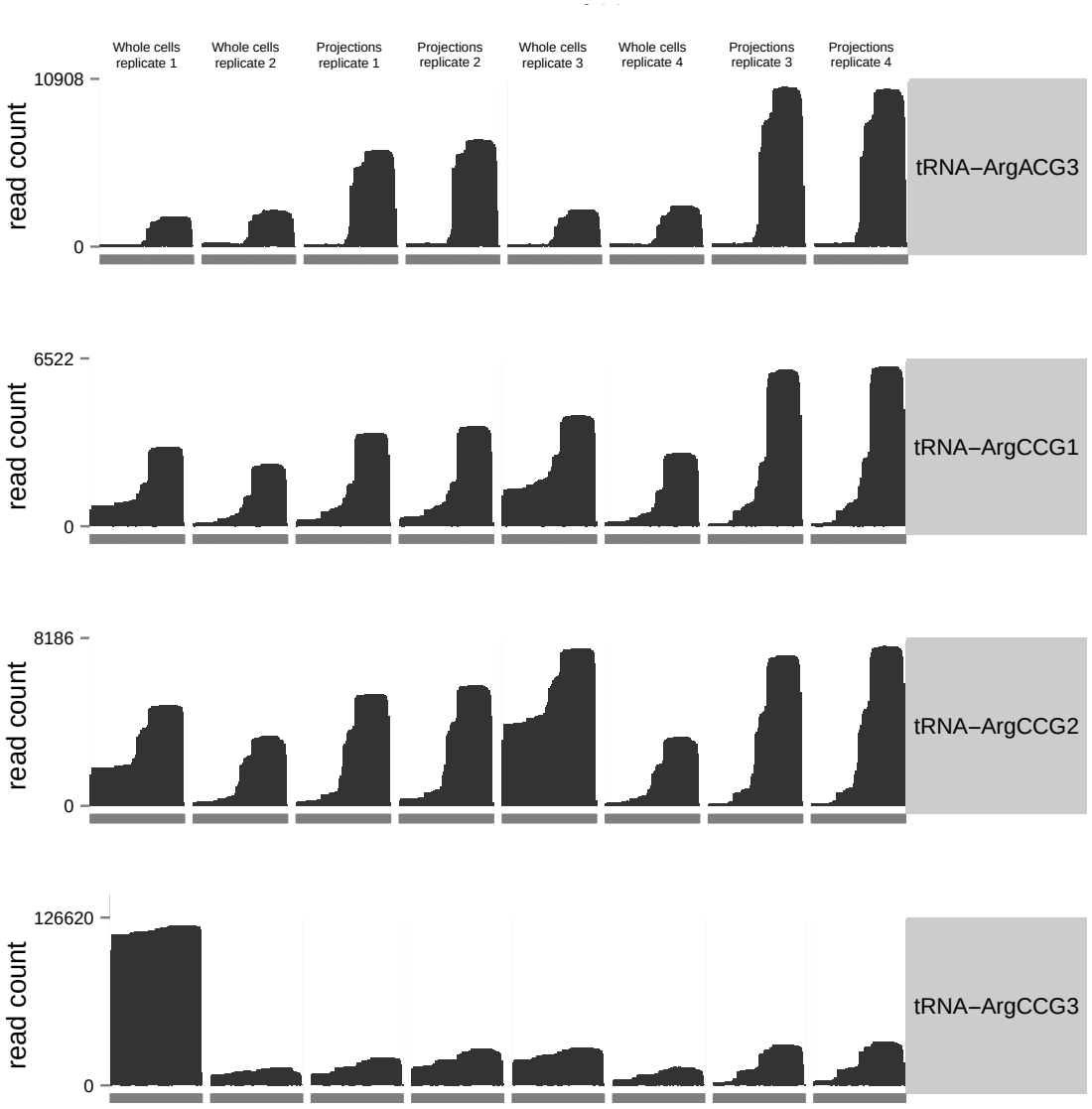
These plots are a part of the output from the ARM-Seq pipeline [Cozen et al., 2015] described in Chapter III.

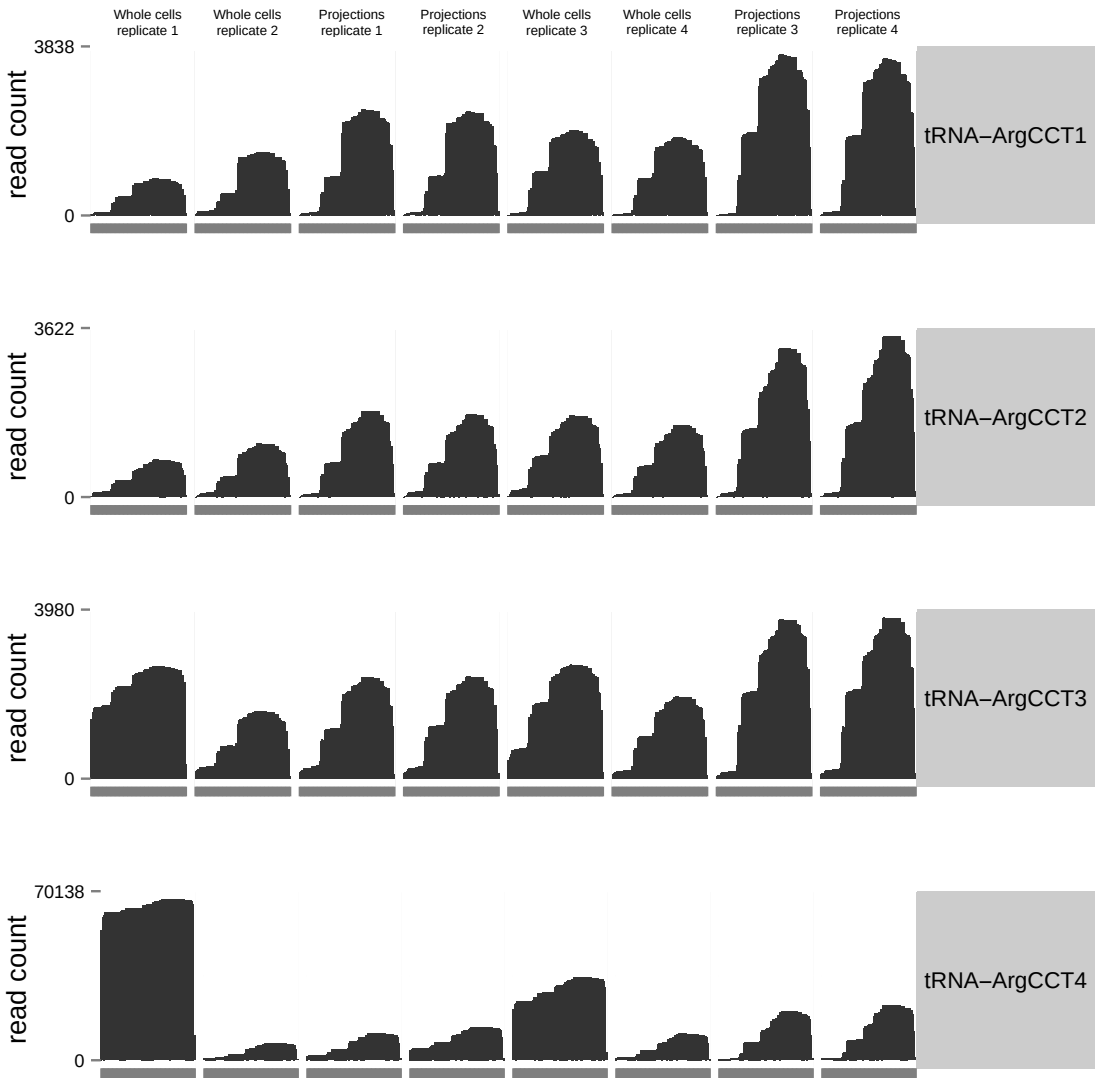

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
2109
read count
0
5'
3'
tRNA−AlaAGC120
282
read count
0
tRNA−AlaAGC127
3174
read count
0
tRNA−AlaAGC219
13885
read count
0
tRNA−AlaAGC22

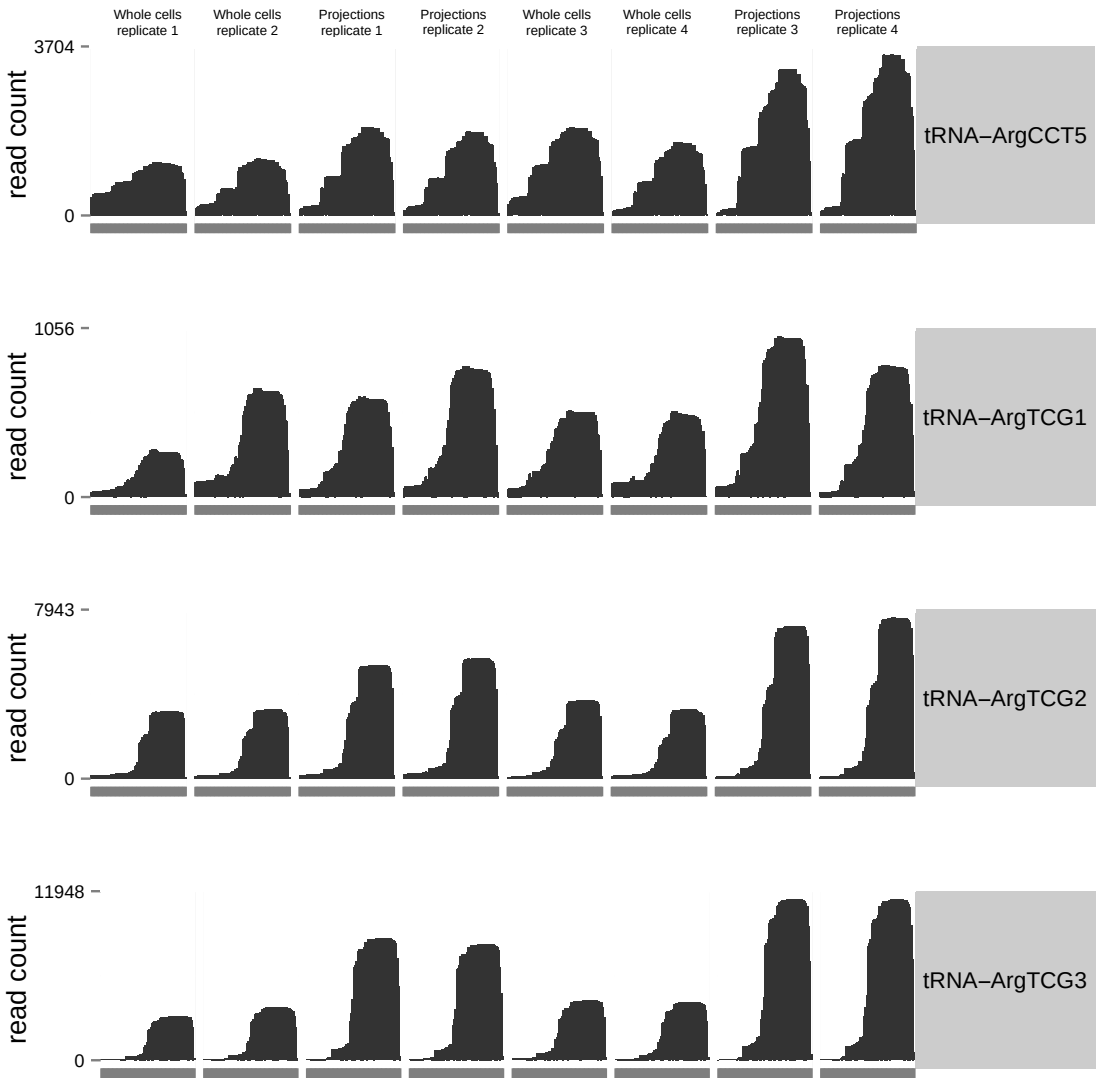

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
1250
read count
0
tRNA−AlaAGC5
435
read count
0
tRNA−AlaAGC52
360
read count
0
tRNA−AlaAGC78
6993
read count
0
tRNA−AlaCGC1

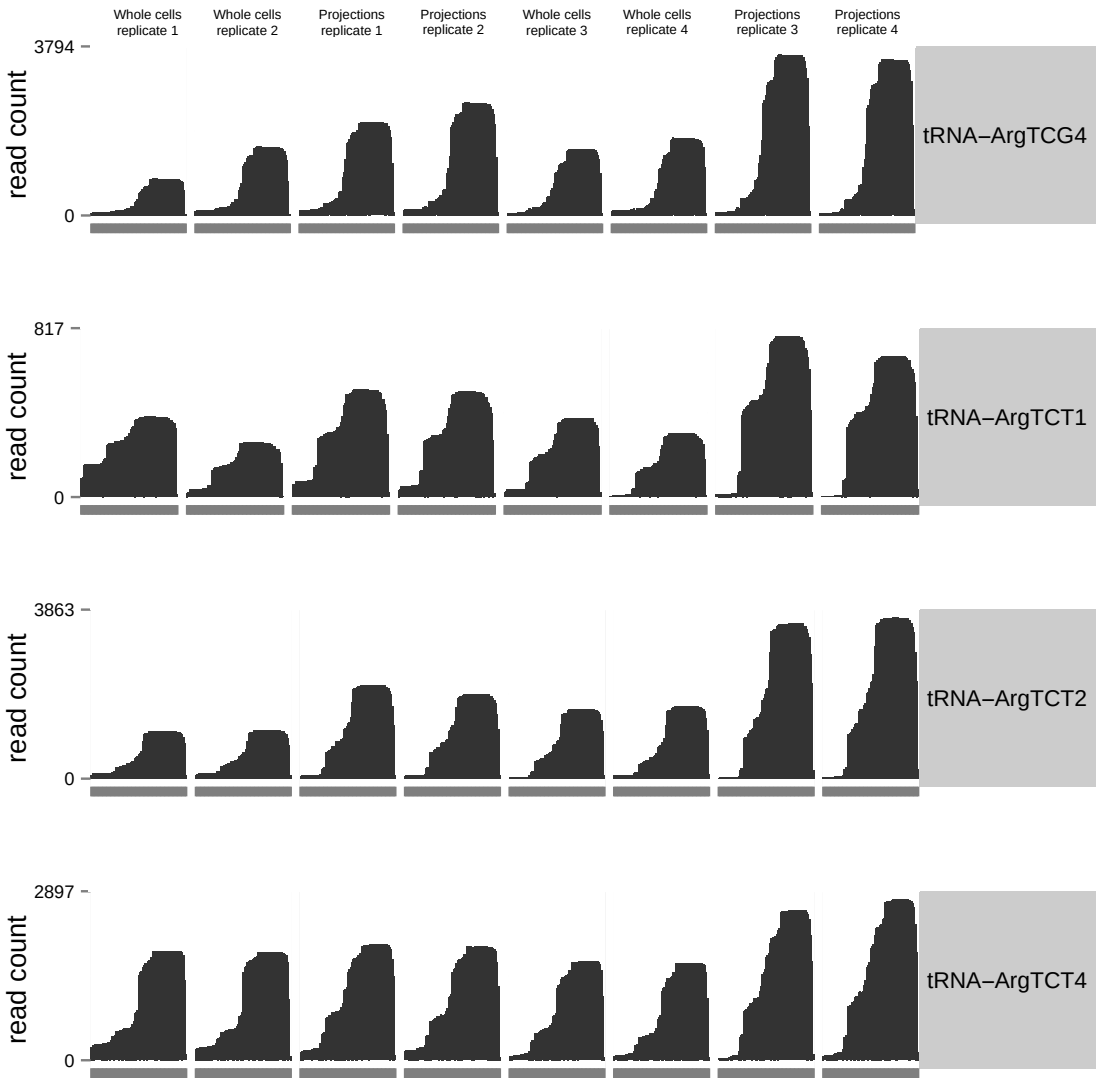

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
4063
read count
0
tRNA–AlaCGC4
3221
read count
0
tRNA–AlaCGC5
639
read count
0
tRNA–AlaTGC1
13193
read count
0
tRNA–AlaTGC2

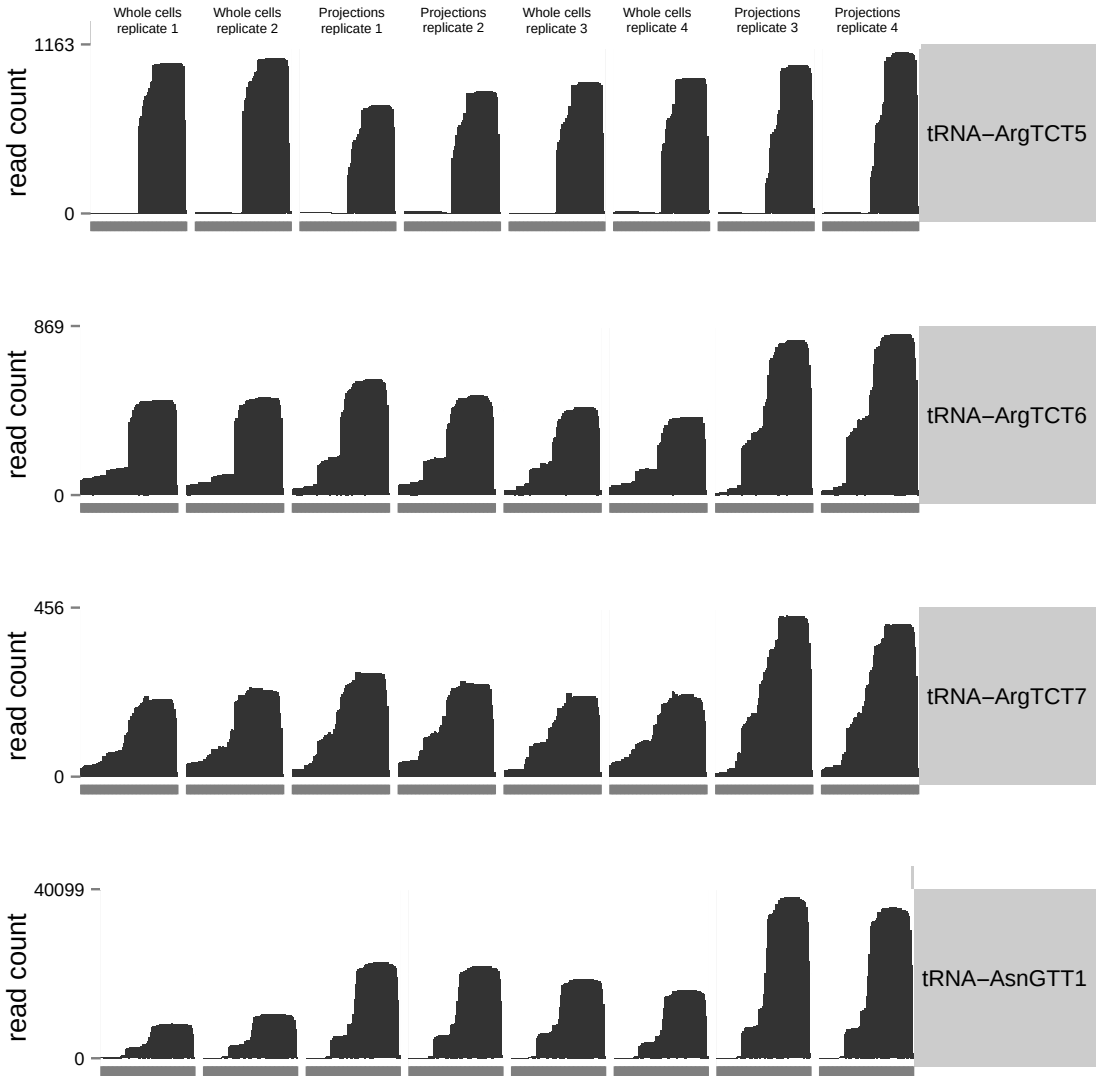

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
2644
read count
0
tRNA–AlaTGC5
12000
read count
0
tRNA–AlaTGC9
8367
read count
0
tRNA–ArgACG1
1584
read count
0
tRNA–ArgACG2

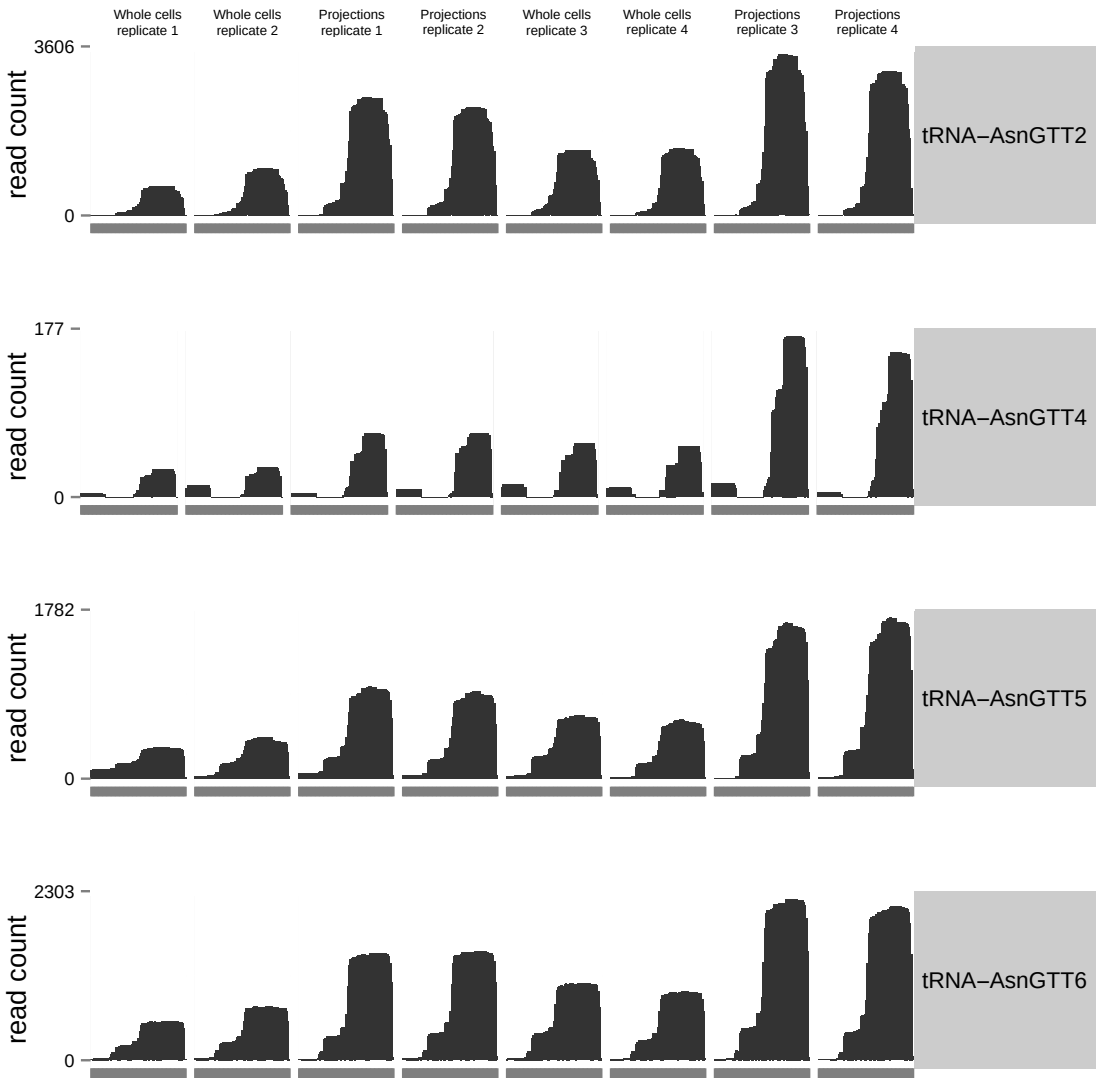

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
10908
read count
0
tRNA–ArgACG3
6522
read count
0
tRNA–ArgCCG1
8186
read count
0
tRNA–ArgCCG2
126620
read count
0
tRNA–ArgCCG3

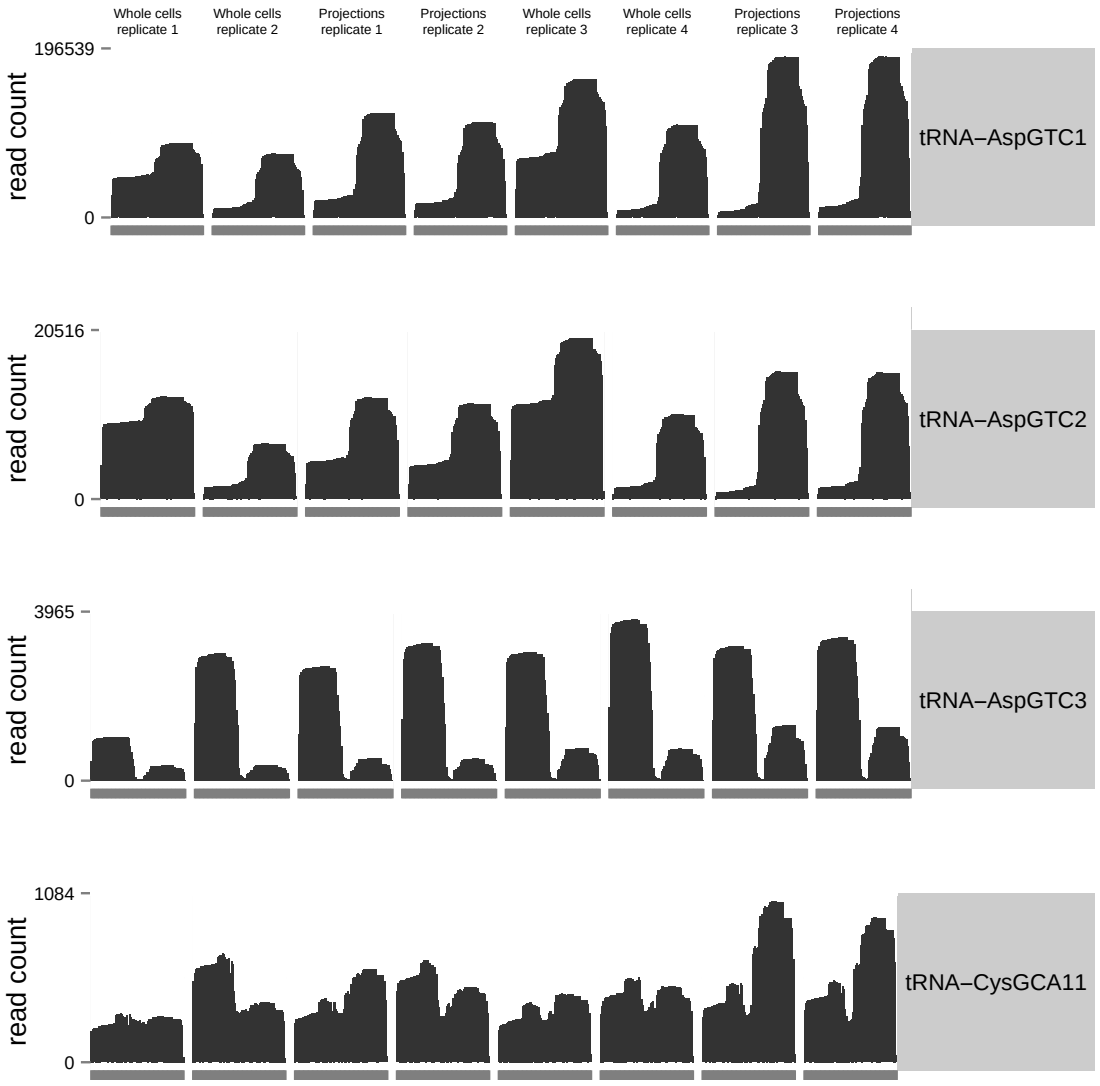

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
3838
0
tRNA–ArgCCT1
3622
0
tRNA–ArgCCT2
3980
0
tRNA–ArgCCT3
70138
0
tRNA–ArgCCT4

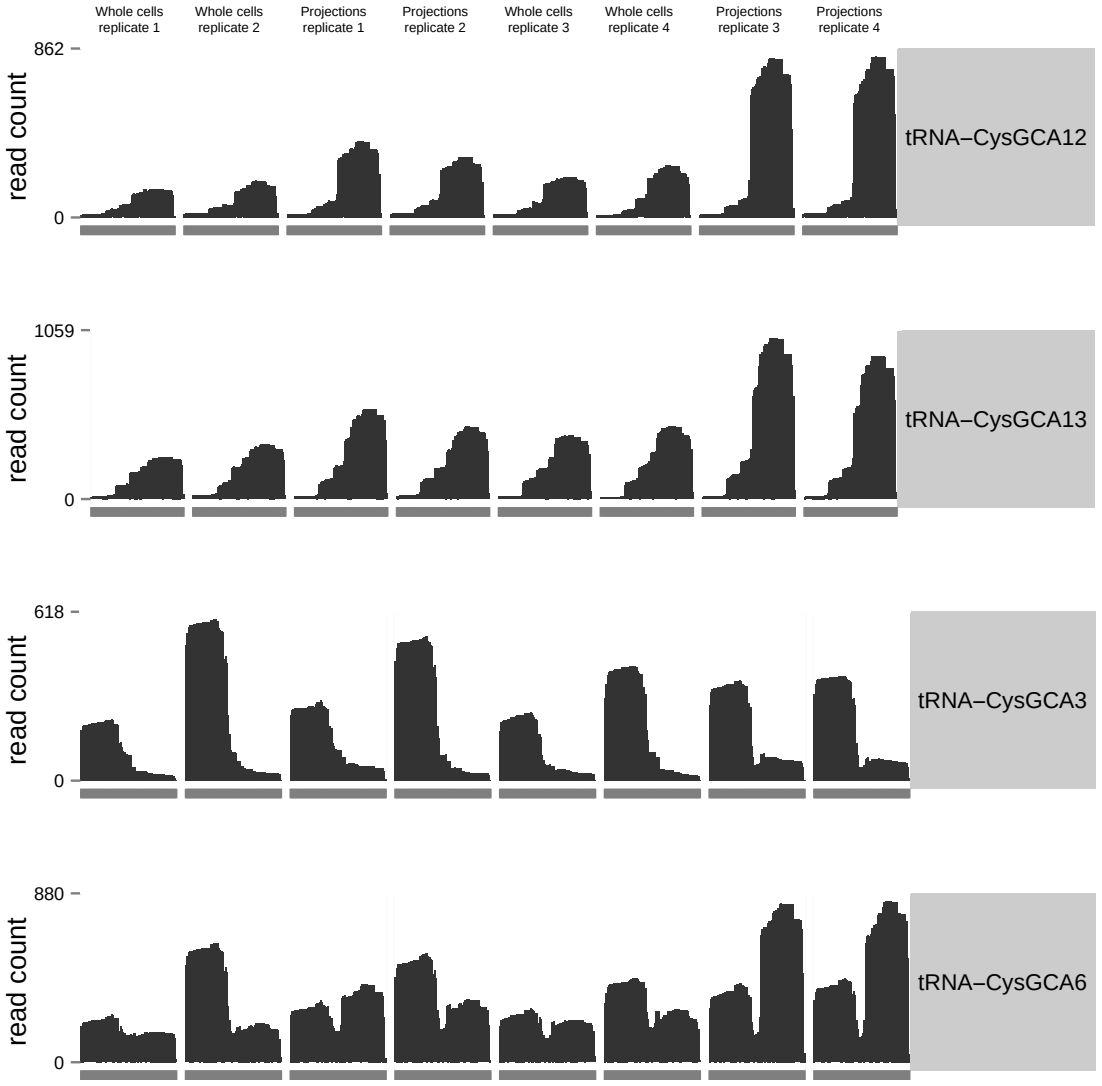

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
3704
read count
0
tRNA–ArgCCT5
1056
read count
0
tRNA–ArgTCG1
7943
read count
0
tRNA–ArgTCG2
11948
read count
0
tRNA–ArgTCG3

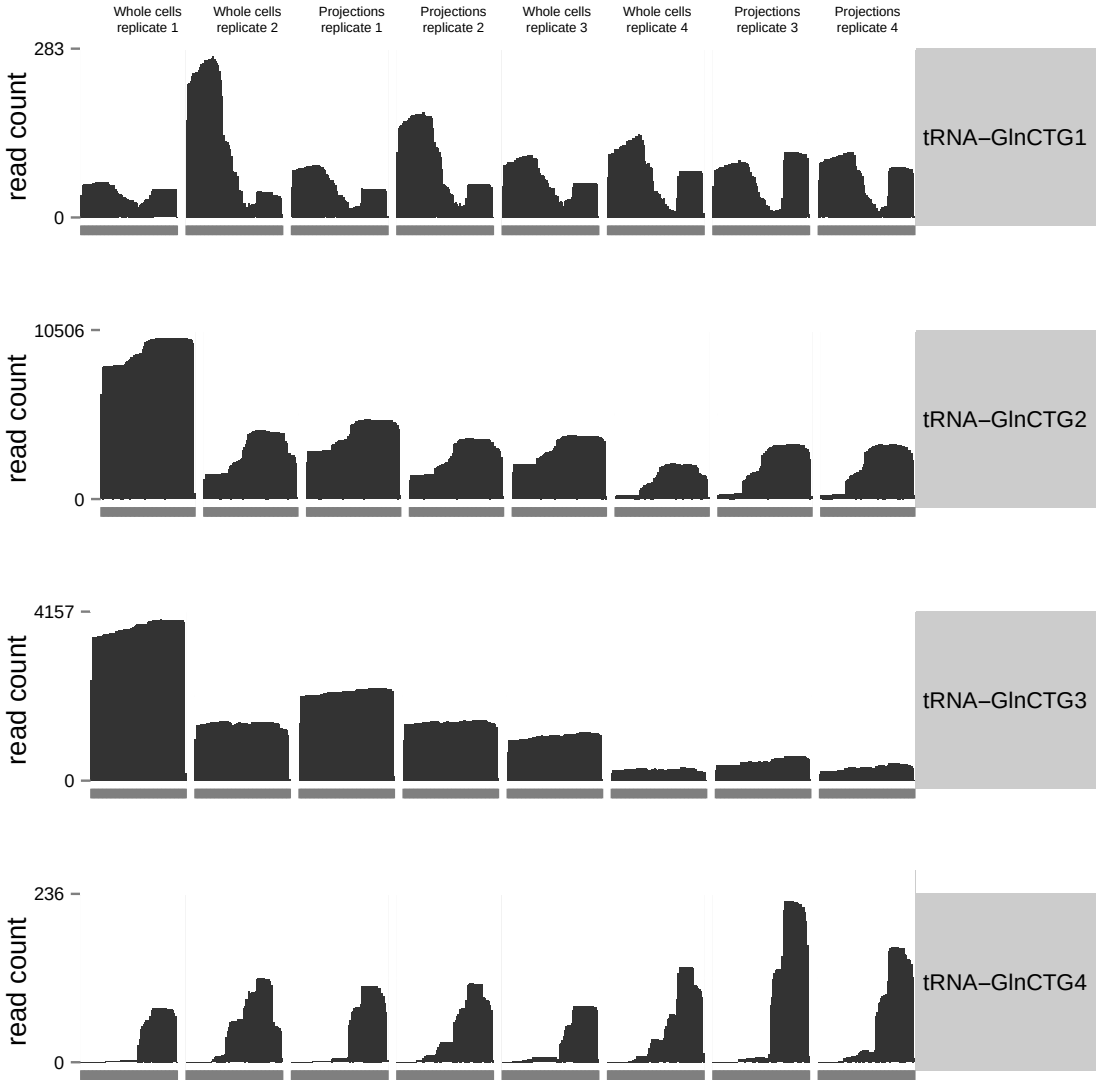

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
3794
read count
0
tRNA–ArgTCG4
817
read count
0
tRNA–ArgTCT1
3863
read count
0
tRNA–ArgTCT2
2897
read count
0
tRNA–ArgTCT4

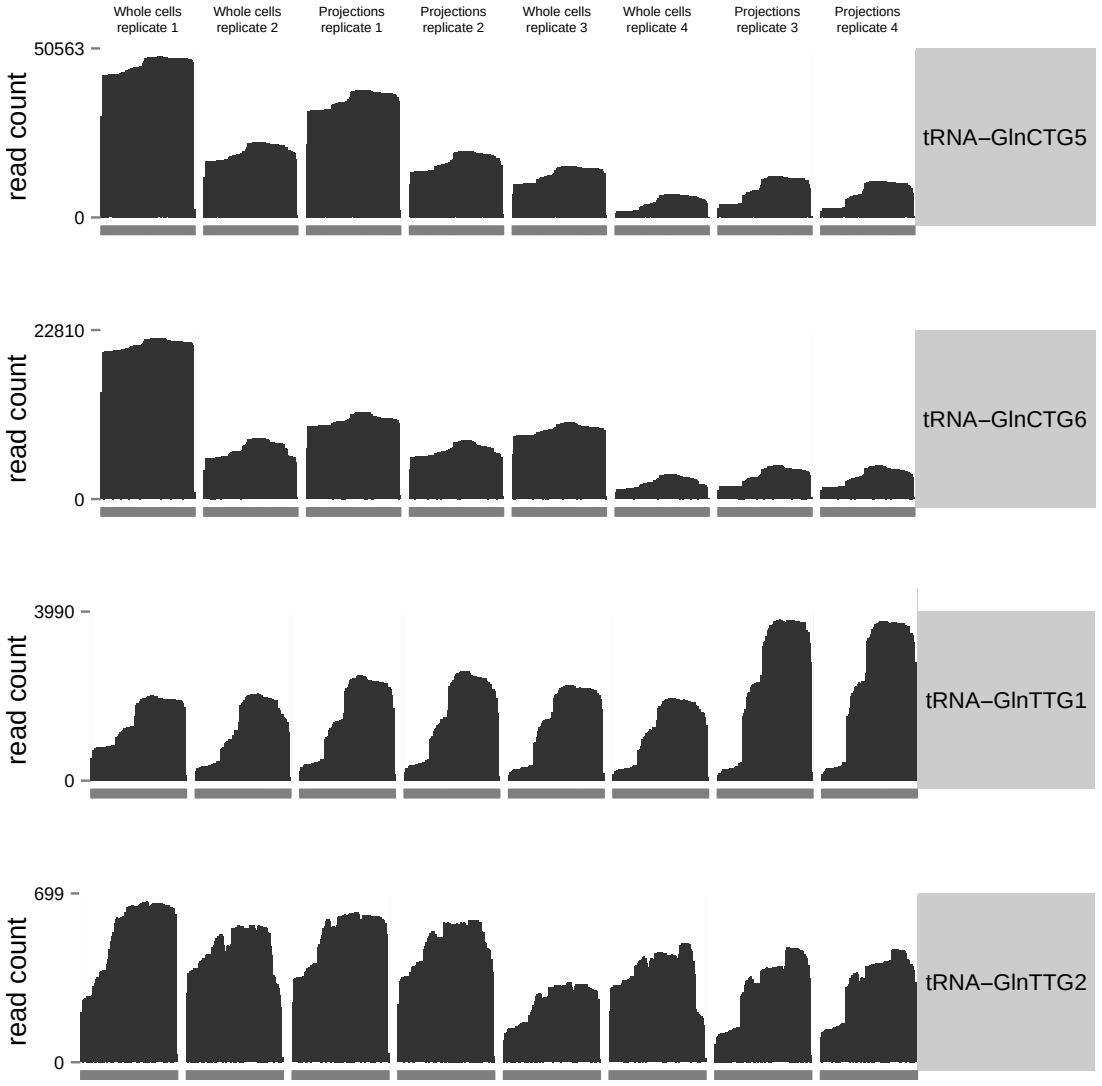

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
1163
read count
0
tRNA–ArgTCT5
869
read count
0
tRNA–ArgTCT6
456
read count
0
tRNA–ArgTCT7
40099
read count
0
tRNA–AsnGTT1

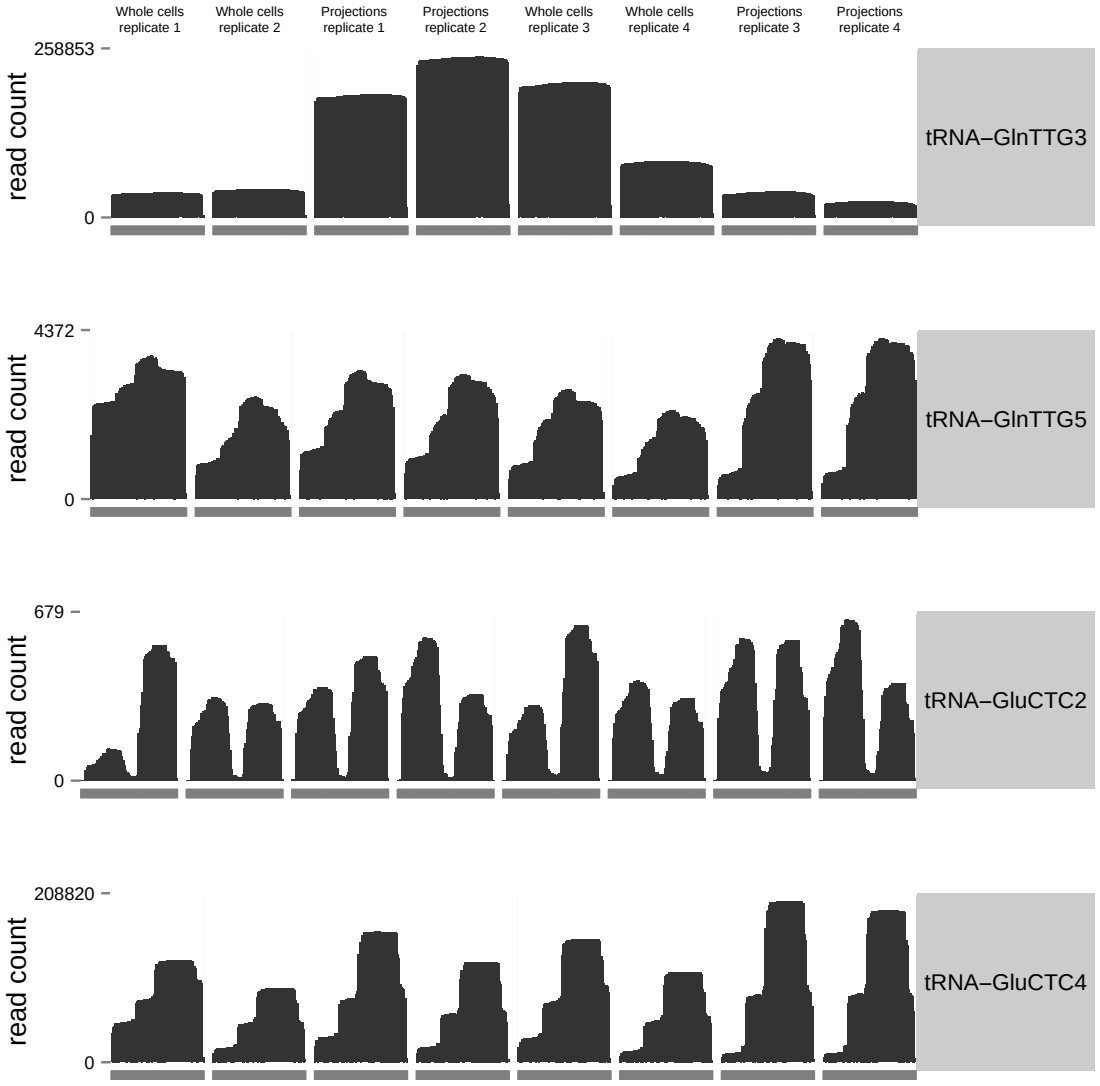

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
3606
0
tRNA–AsnGTT2
177
0
tRNA–AsnGTT4
1782
0
tRNA–AsnGTT5
2303
0
tRNA–AsnGTT6

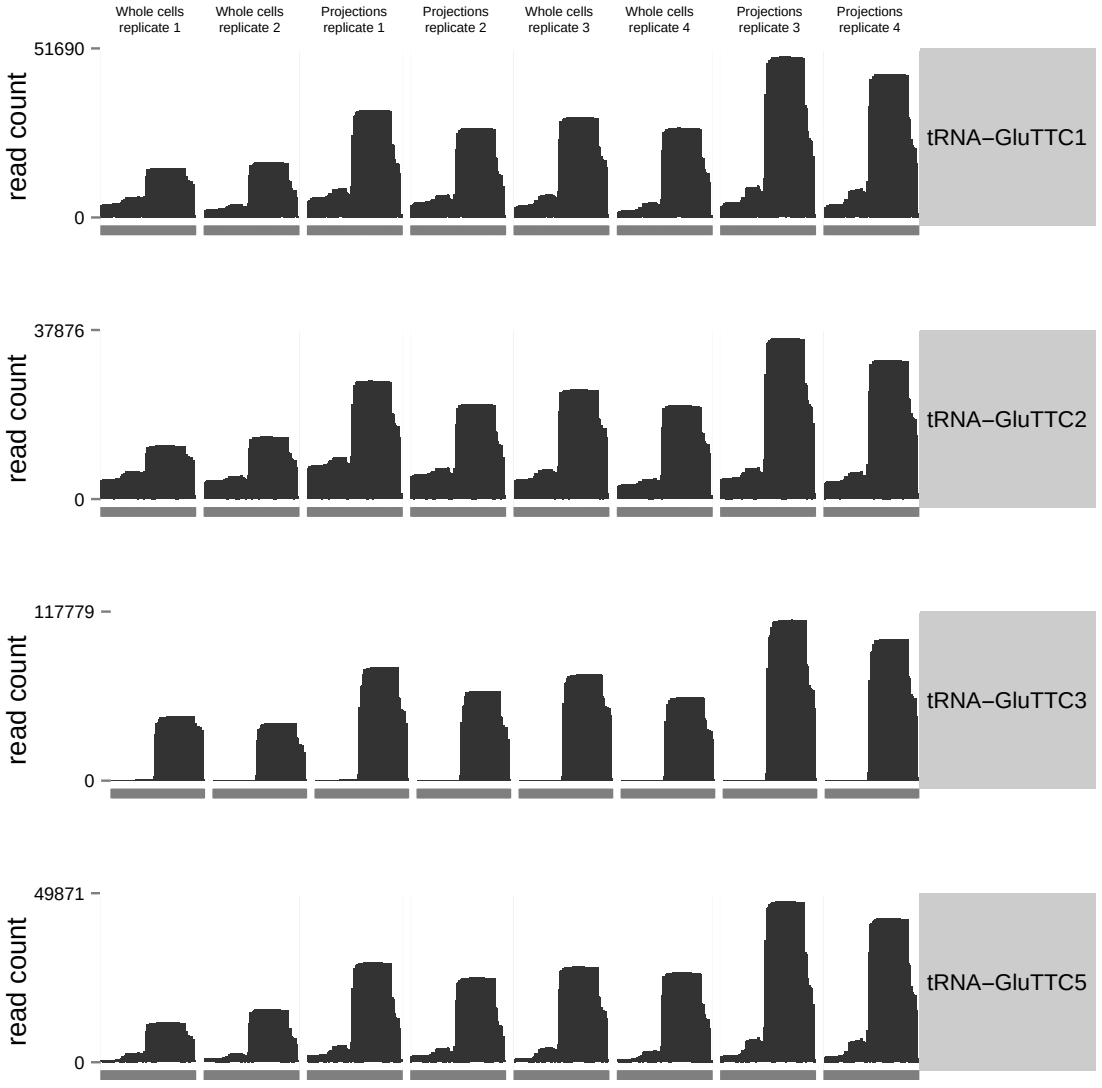

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
196539
read count
0
tRNA–AspGTC1
20516
read count
0
tRNA–AspGTC2
3965
read count
0
tRNA–AspGTC3
1084
read count
0
tRNA–CysGCA11

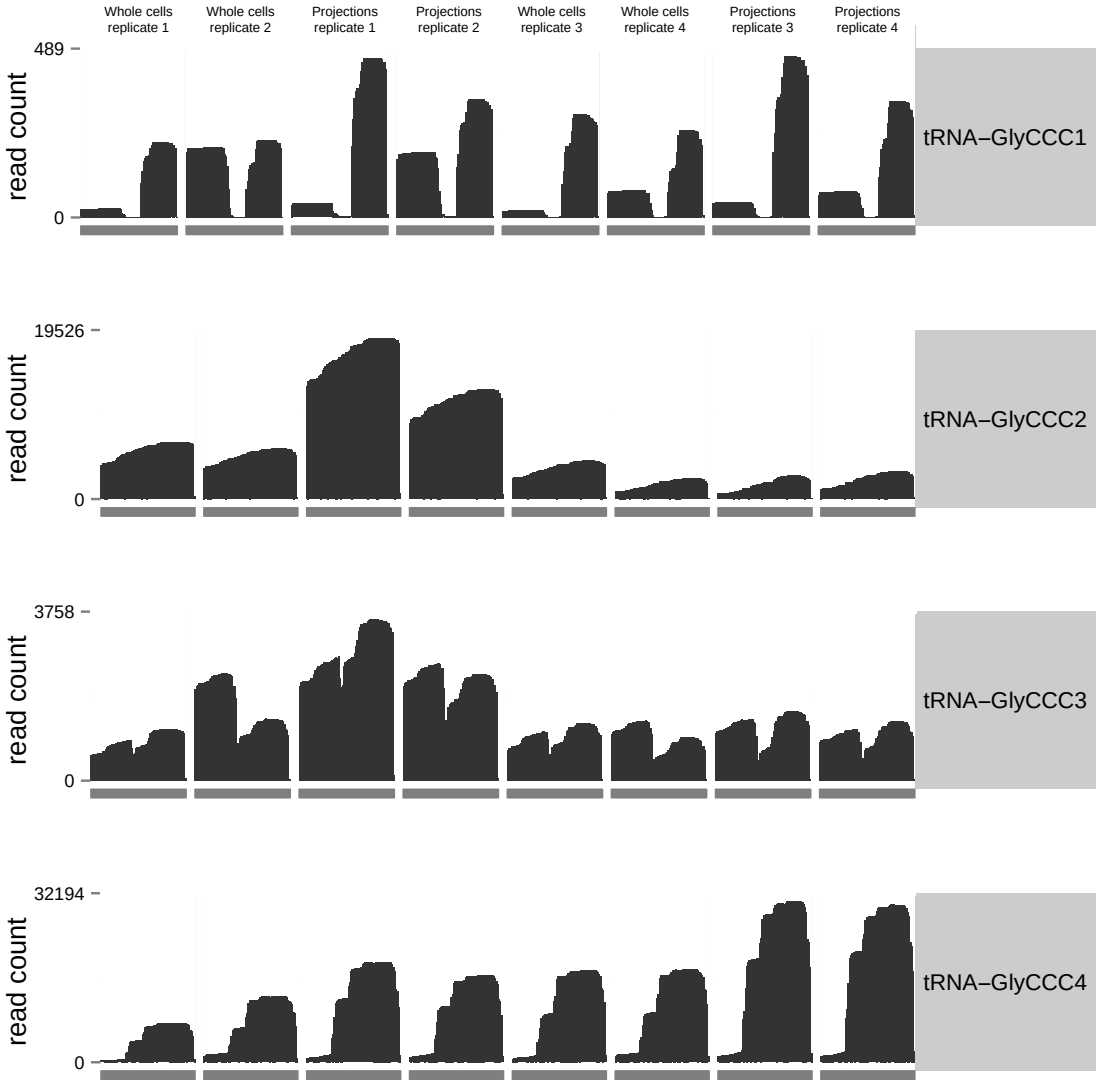

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
862
0
tRNA–CysGCA12
1059
0
tRNA–CysGCA13
618
0
tRNA–CysGCA3
880
0
tRNA–CysGCA6

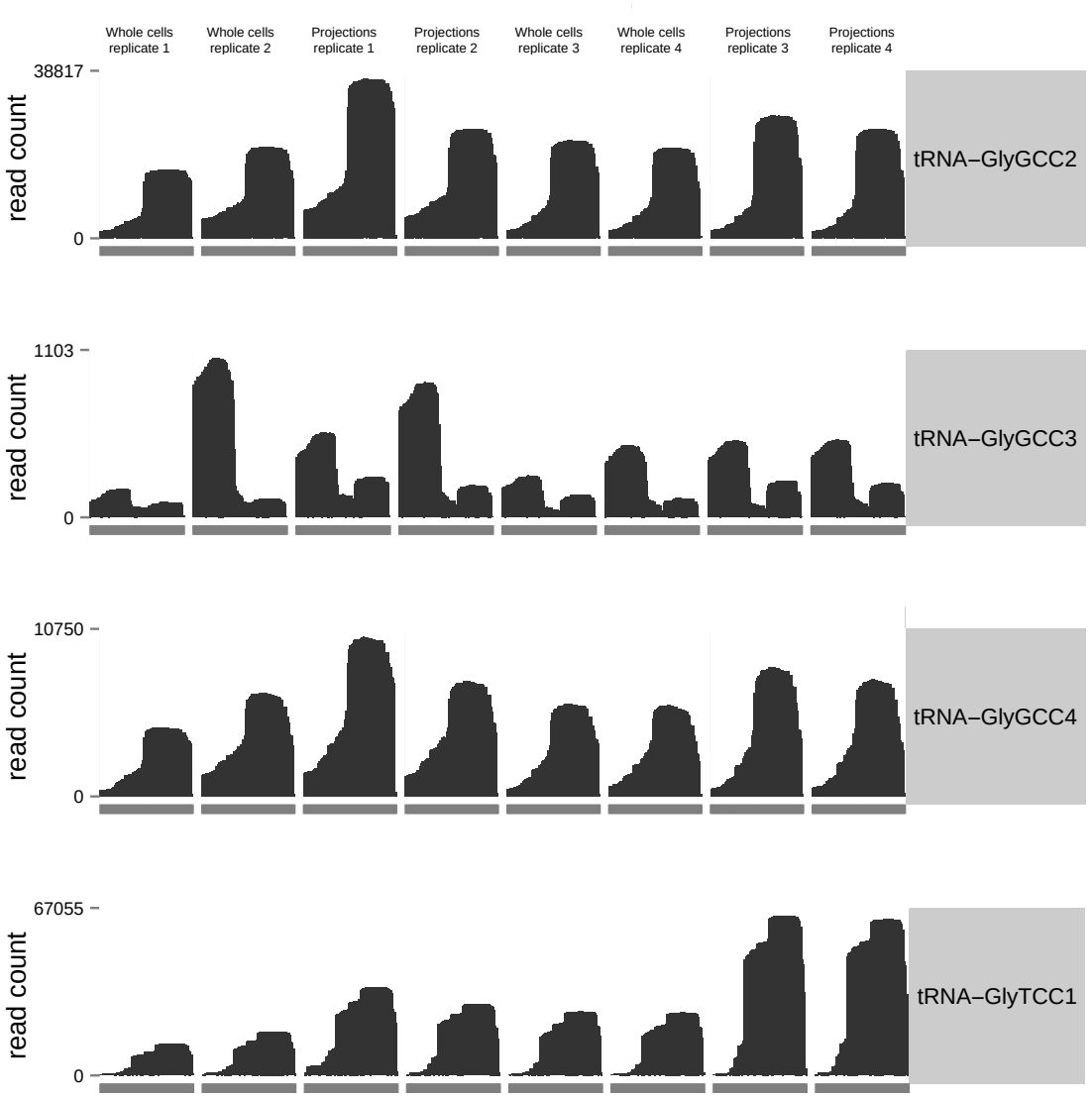

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
283
0
tRNA–GlnCTG1
10506
0
tRNA–GlnCTG2
4157
0
tRNA–GlnCTG3
236
0
tRNA–GlnCTG4

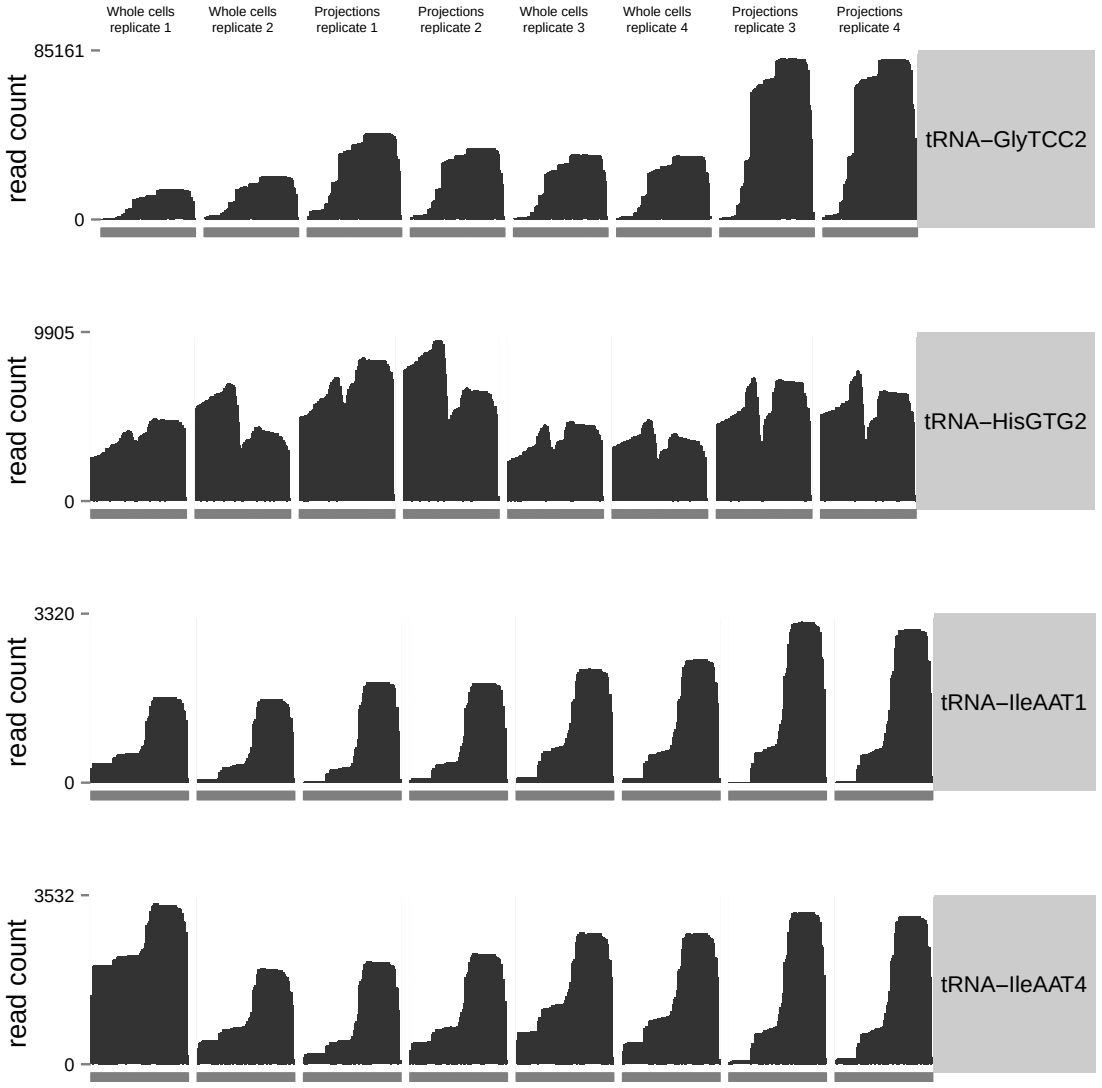

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
50563
read count
0
tRNA−GlnCTG5
22810
read count
0
tRNA−GlnCTG6
3990
read count
0
tRNA−GlnTTG1
699
read count
0
tRNA−GlnTTG2

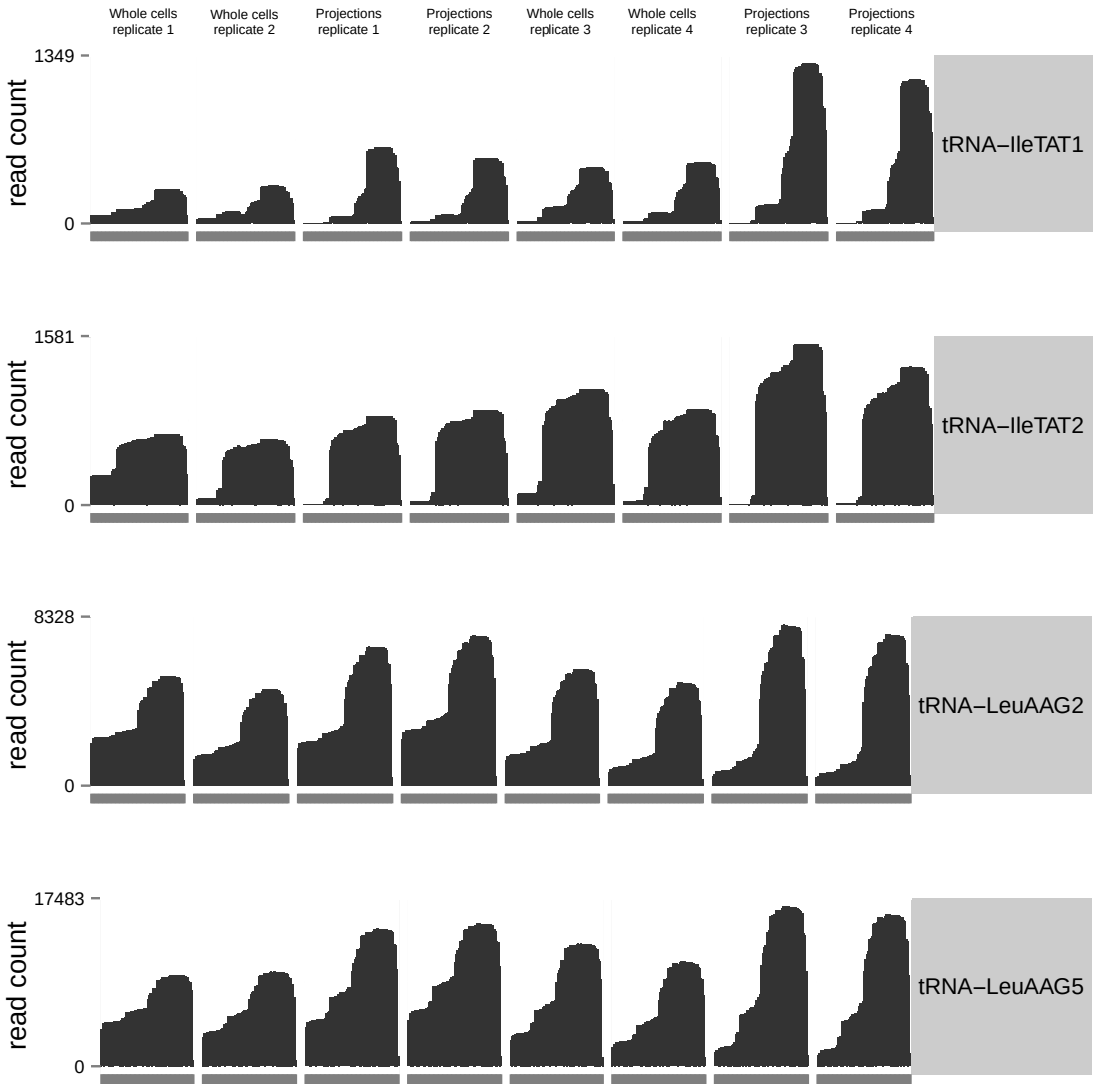

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
258853
read count
0
tRNA–GlnTTG3
4372
read count
0
tRNA–GlnTTG5
679
read count
0
tRNA–GluCTC2
208820
read count
0
tRNA–GluCTC4

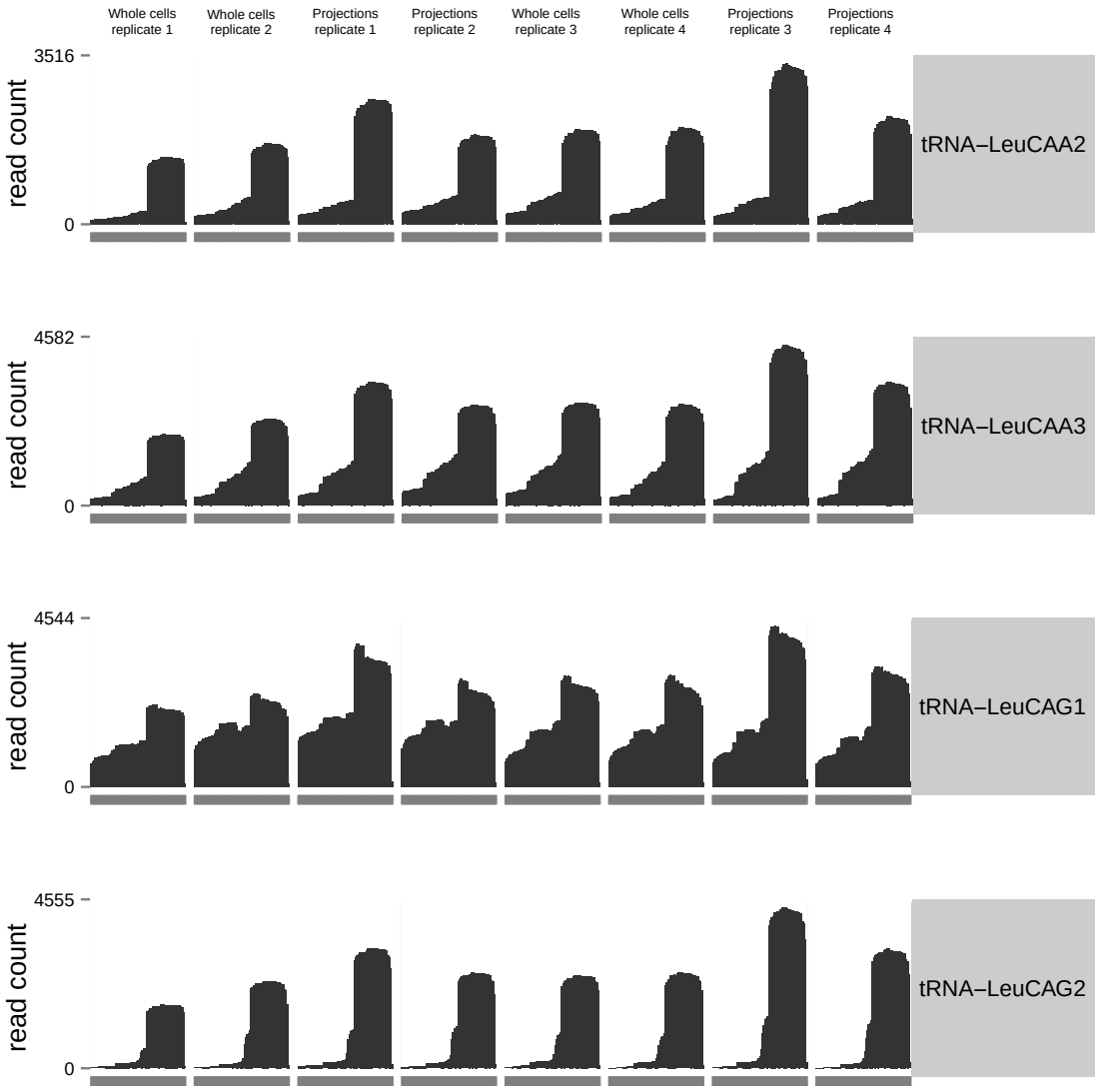

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
51690
0
tRNA−GluTTC1
37876
0
tRNA−GluTTC2
117779
0
tRNA−GluTTC3
49871
0
tRNA−GluTTC5

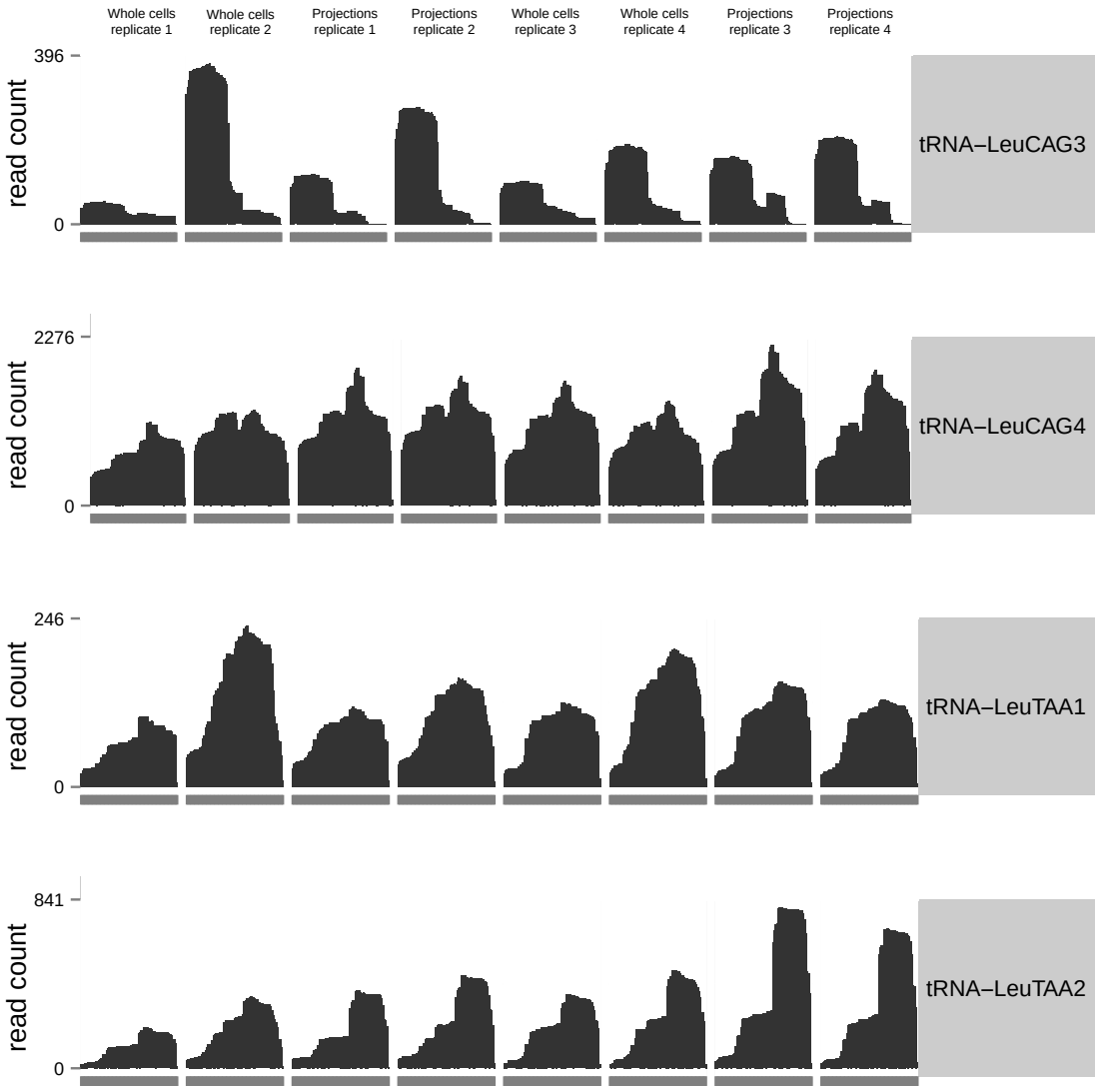

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
489
0
tRNA−GlyCCC1
19526
0
tRNA−GlyCCC2
3758
0
tRNA−GlyCCC3
32194
0
tRNA−GlyCCC4

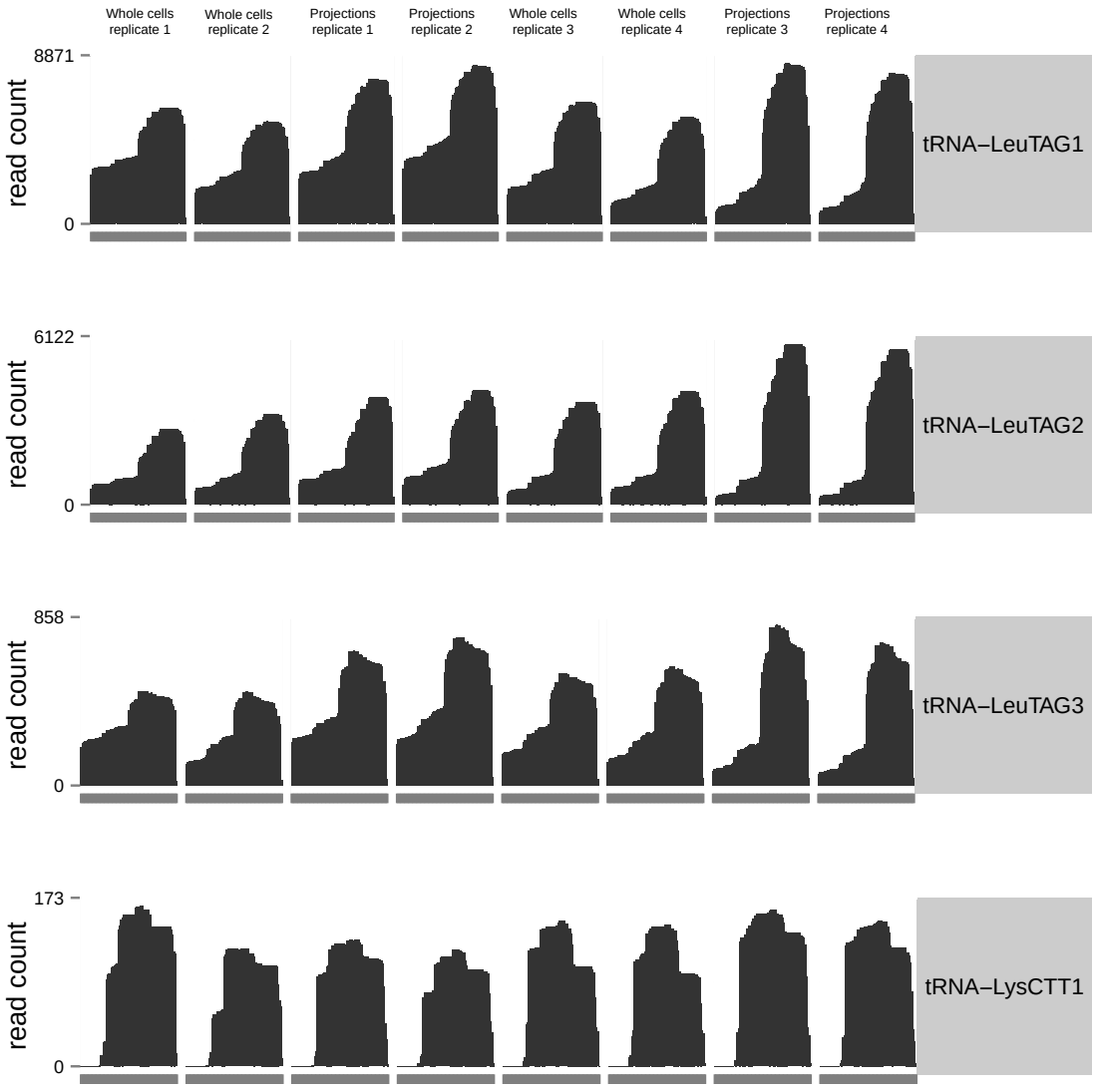

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
38817
read count
0
tRNA−GlyGCC2
1103
read count
0
tRNA−GlyGCC3
10750
read count
0
tRNA−GlyGCC4
67055
read count
0
tRNA−GlyTCC1

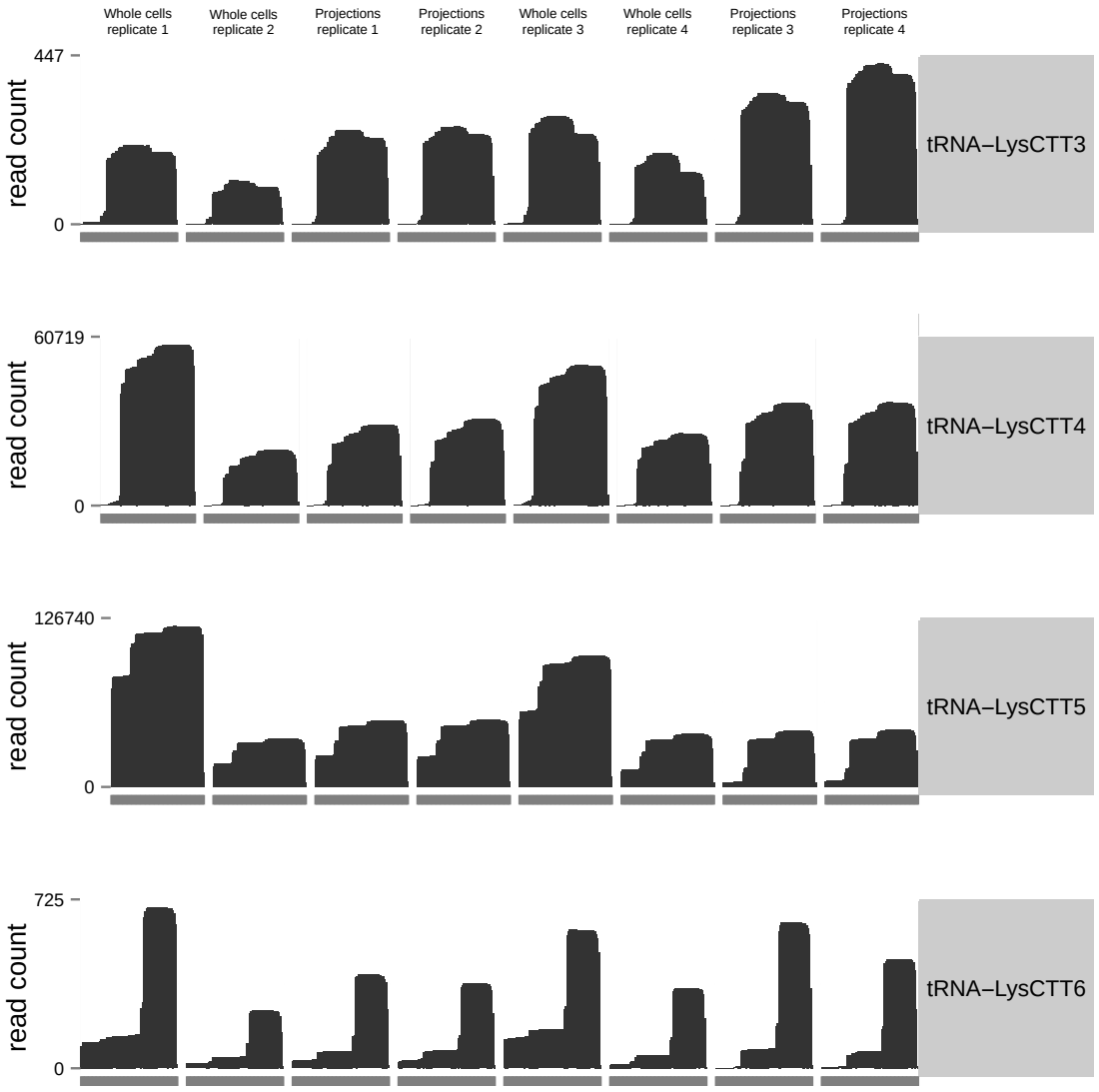

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
85161
read count
0
tRNA–GlyTCC2
9905
read count
0
tRNA–HisGTG2
3320
read count
0
tRNA–IleAAT1
3532
read count
0
tRNA–IleAAT4

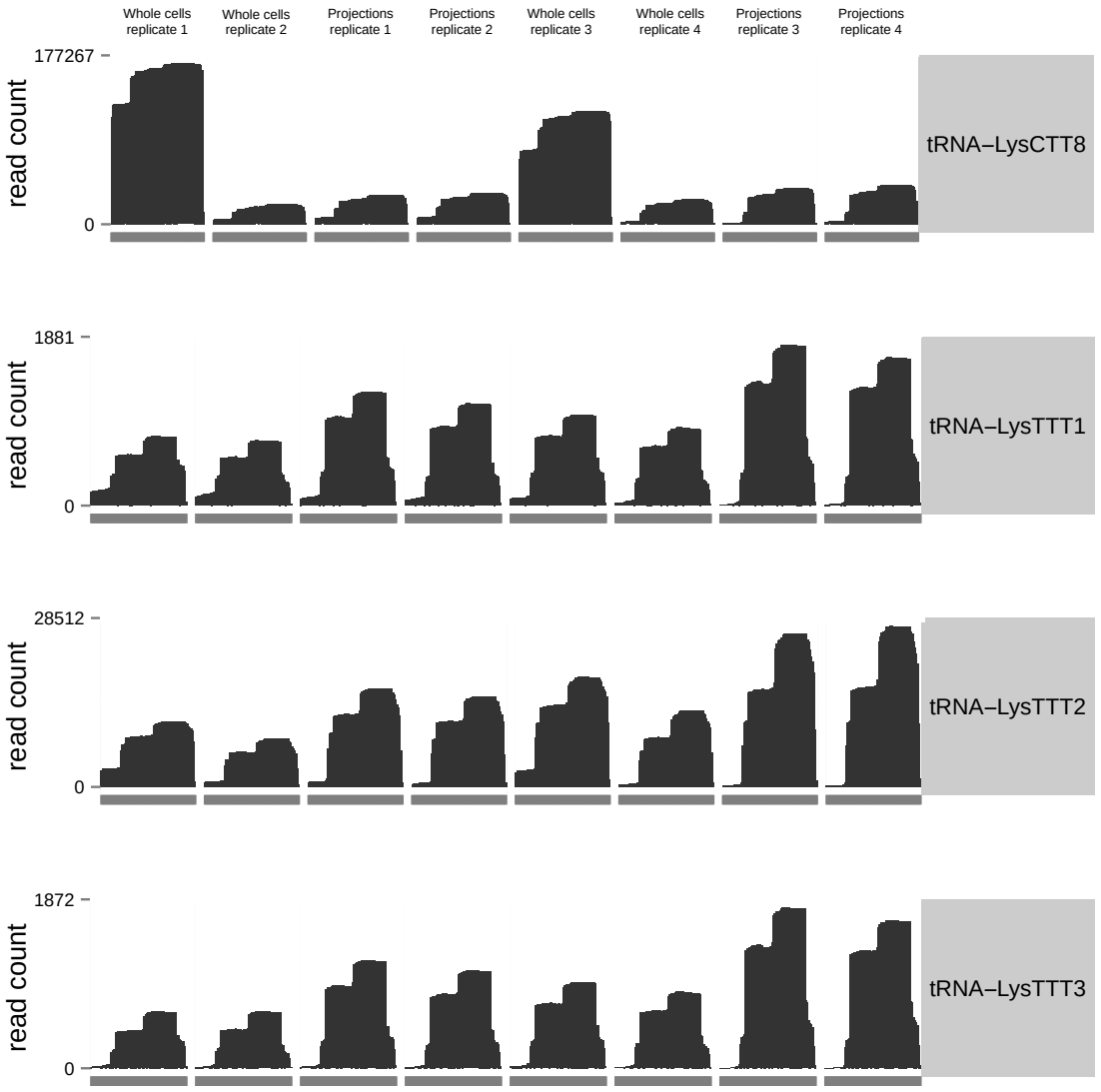

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
1349
0
tRNA–IleTAT1
1581
0
tRNA–IleTAT2
8328
0
tRNA–LeuAAG2
17483
0
tRNA–LeuAAG5

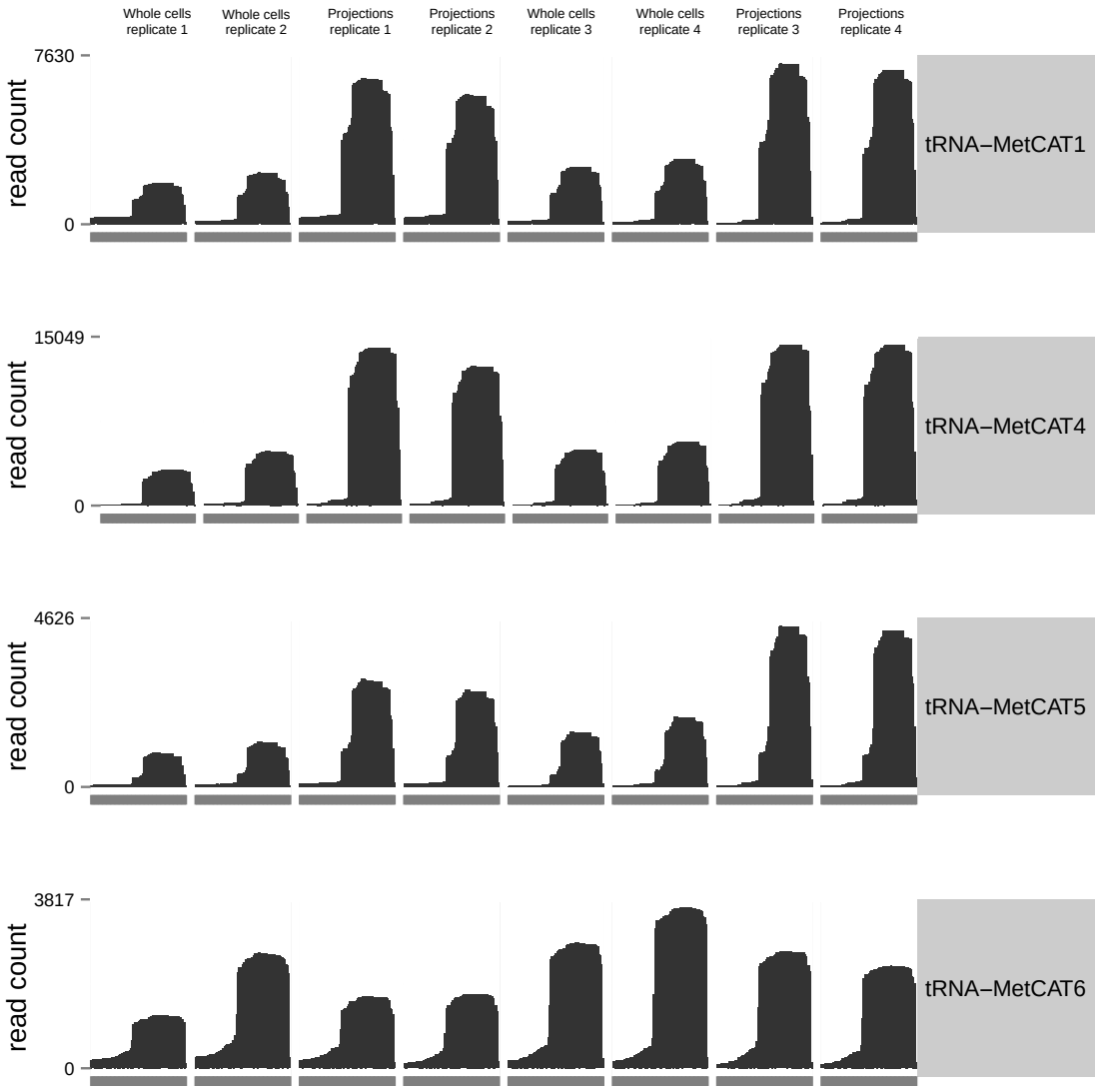

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
3516
0
tRNA–LeuCAA2
4582
0
tRNA–LeuCAA3
4544
0
tRNA–LeuCAG1
4555
0
tRNA–LeuCAG2

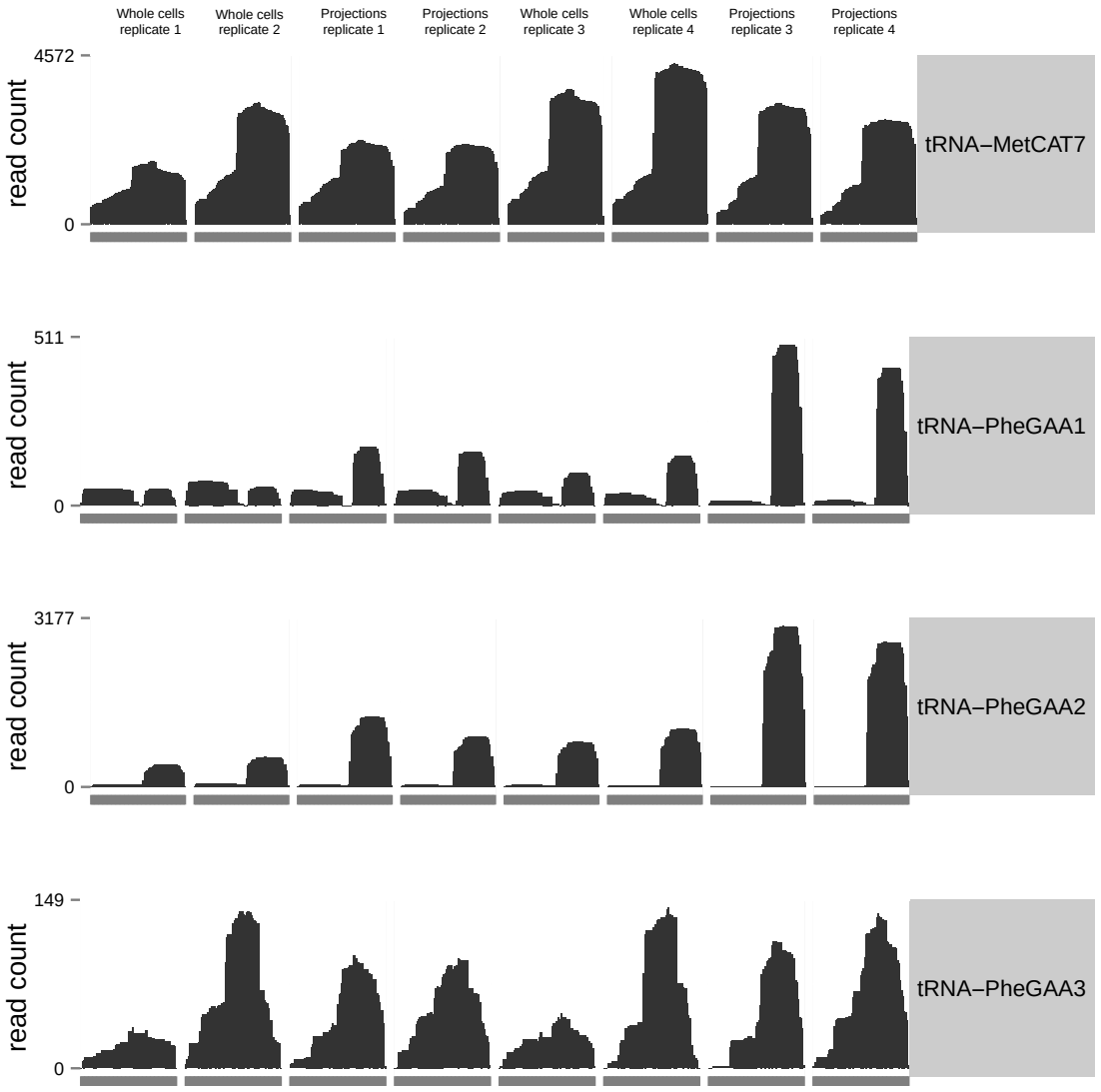

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
396
0
tRNA–LeuCAG3
2276
0
tRNA–LeuCAG4
246
0
tRNA–LeuTAA1
841
0
tRNA–LeuTAA2

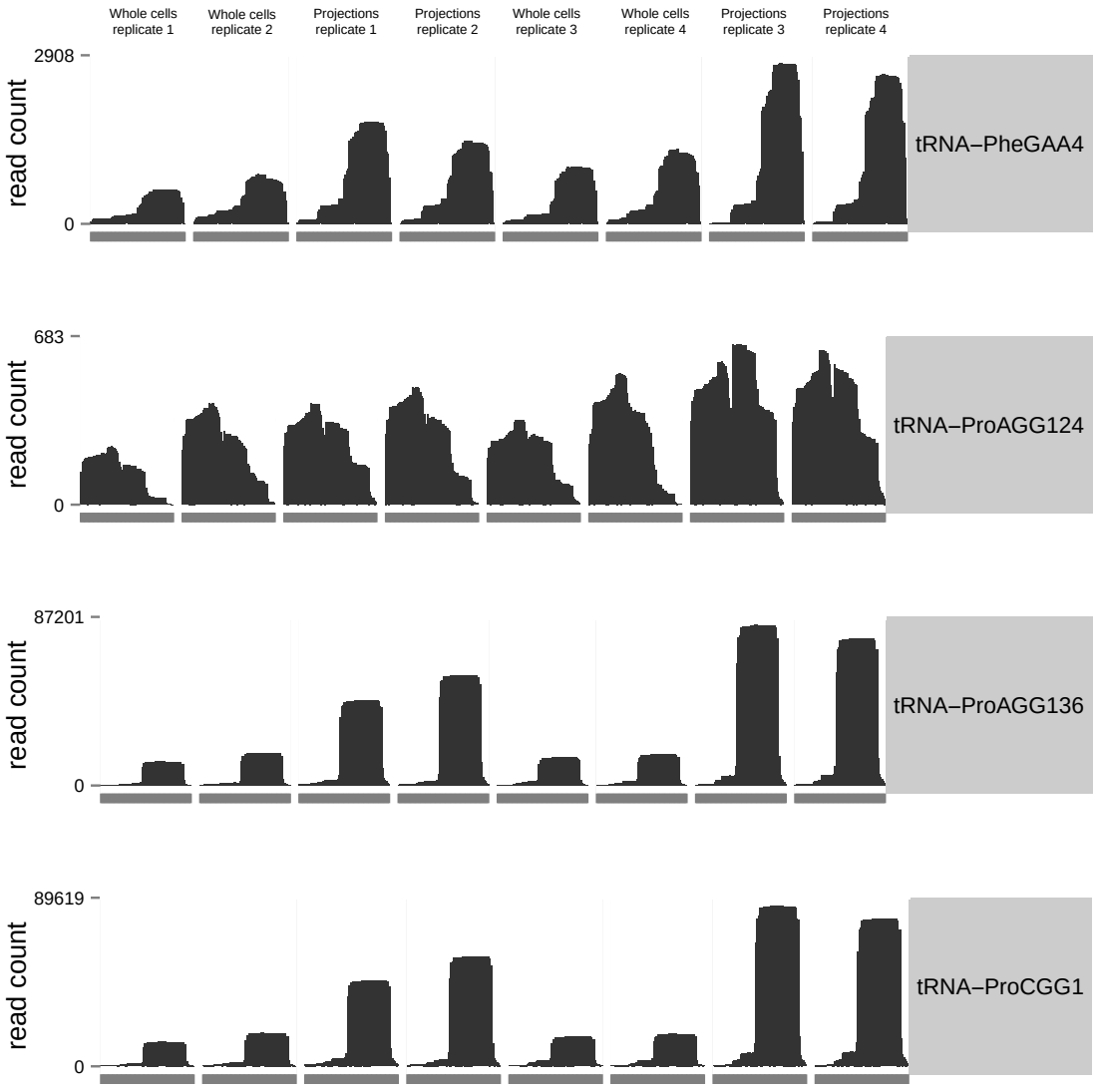

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
8871
read count
0
tRNA−LeuTAG1
6122
read count
0
tRNA−LeuTAG2
858
read count
0
tRNA−LeuTAG3
173
read count
0
tRNA−LysCTT1

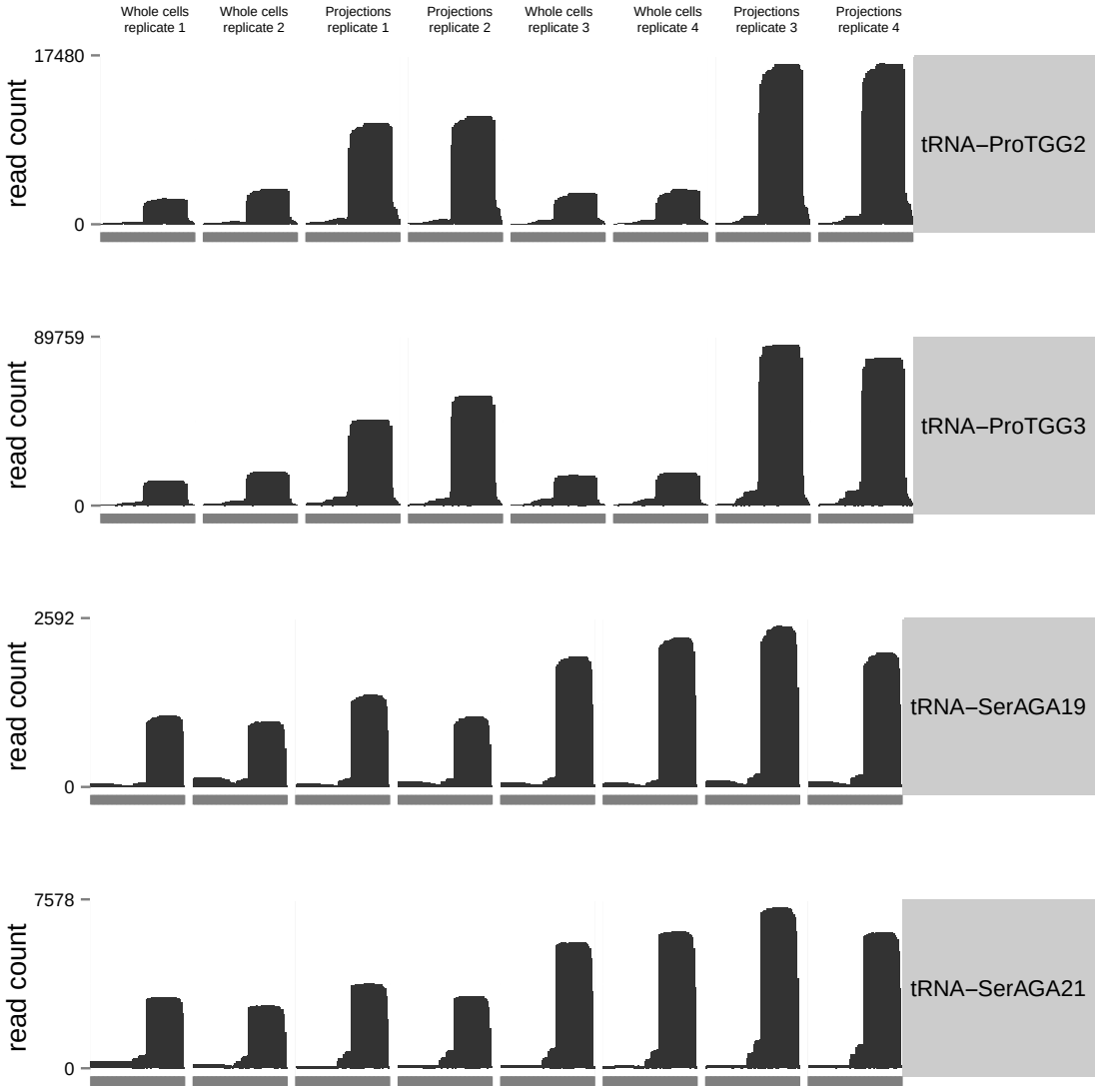

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
447
0
tRNA−LysCTT3
60719
tRNA−LysCTT4
126740
tRNA−LysCTT5
725
tRNA−LysCTT6

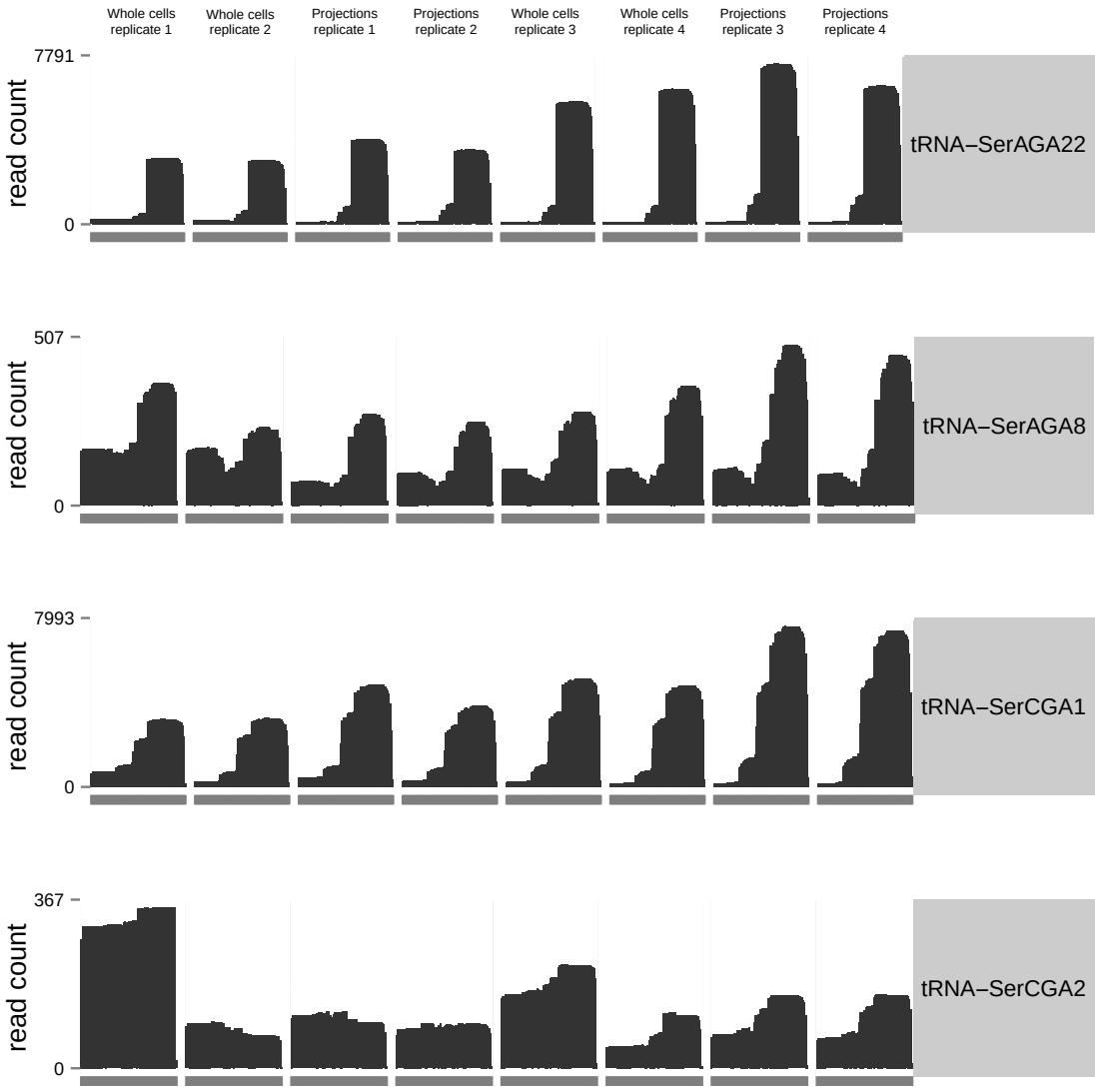

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
177267
0
tRNA−LysCTT8
1881
0
tRNA−LysTTT1
28512
0
tRNA−LysTTT2
1872
0
tRNA−LysTTT3

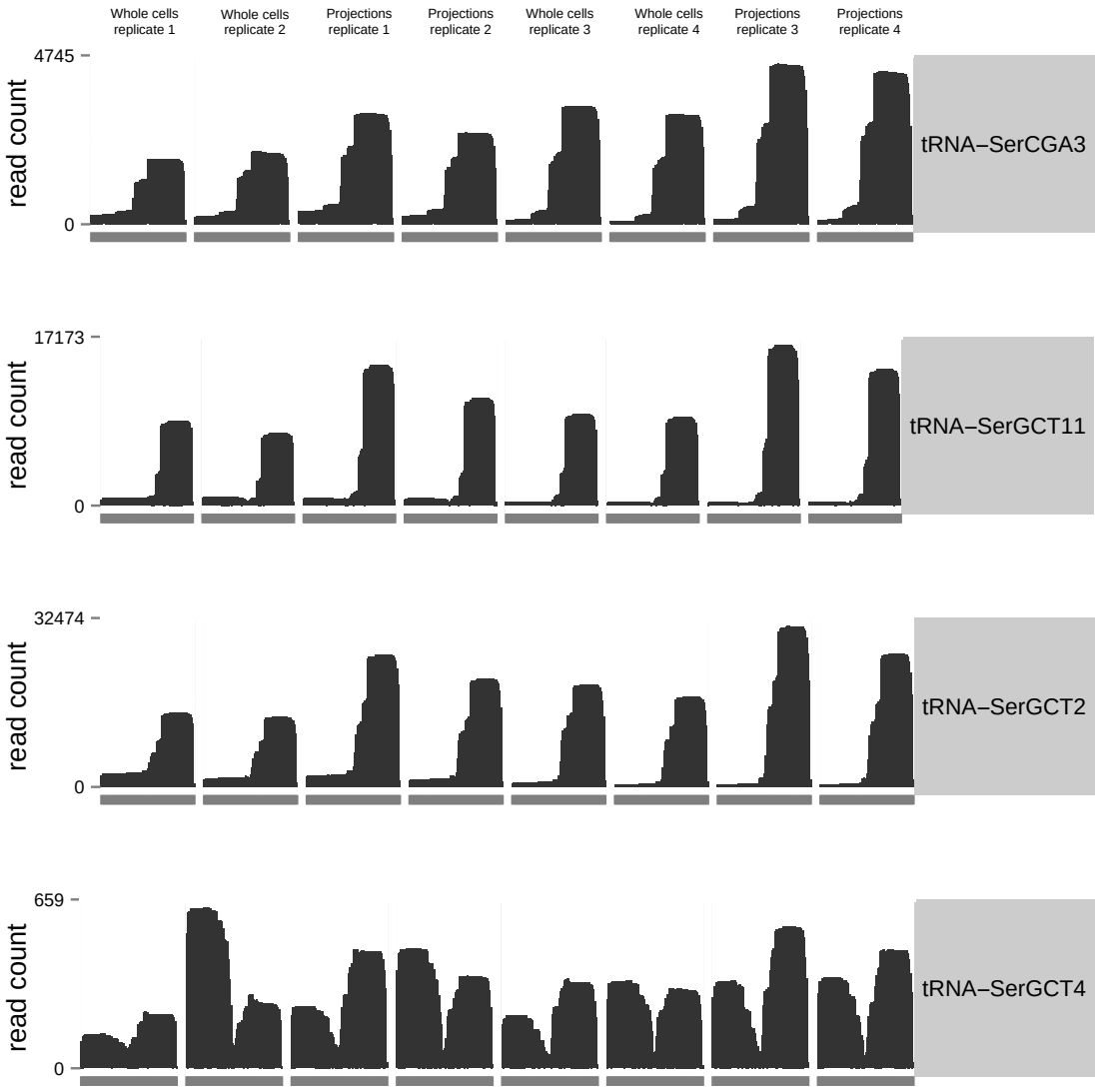

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
7630
read count
0
tRNA−MetCAT1
15049
read count
0
tRNA−MetCAT4
4626
read count
0
tRNA−MetCAT5
3817
read count
0
tRNA−MetCAT6

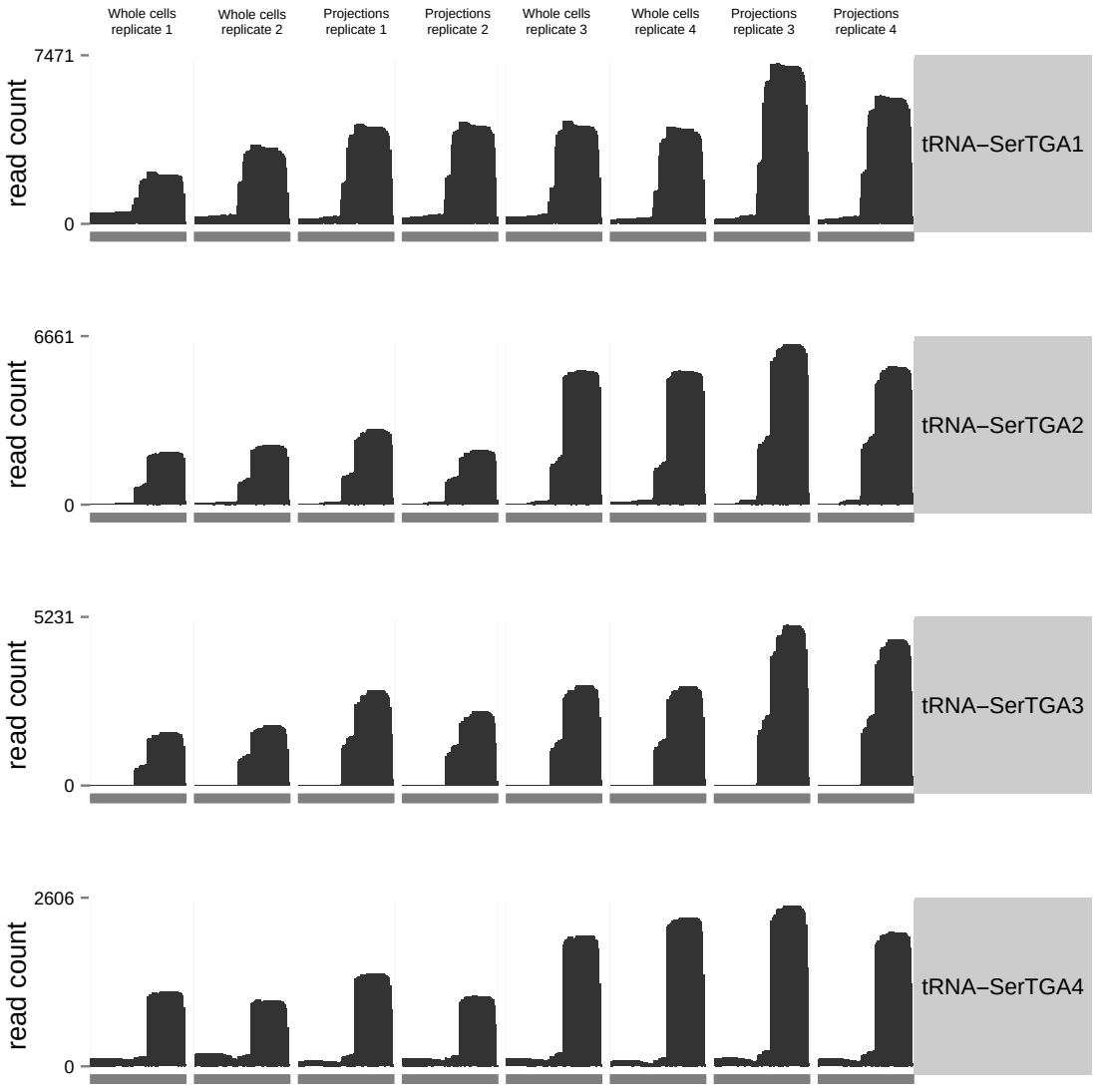

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
4572
read count
0
tRNA–MetCAT7
511
read count
0
tRNA–PheGAA1
3177
read count
0
tRNA–PheGAA2
149
read count
0
tRNA–PheGAA3

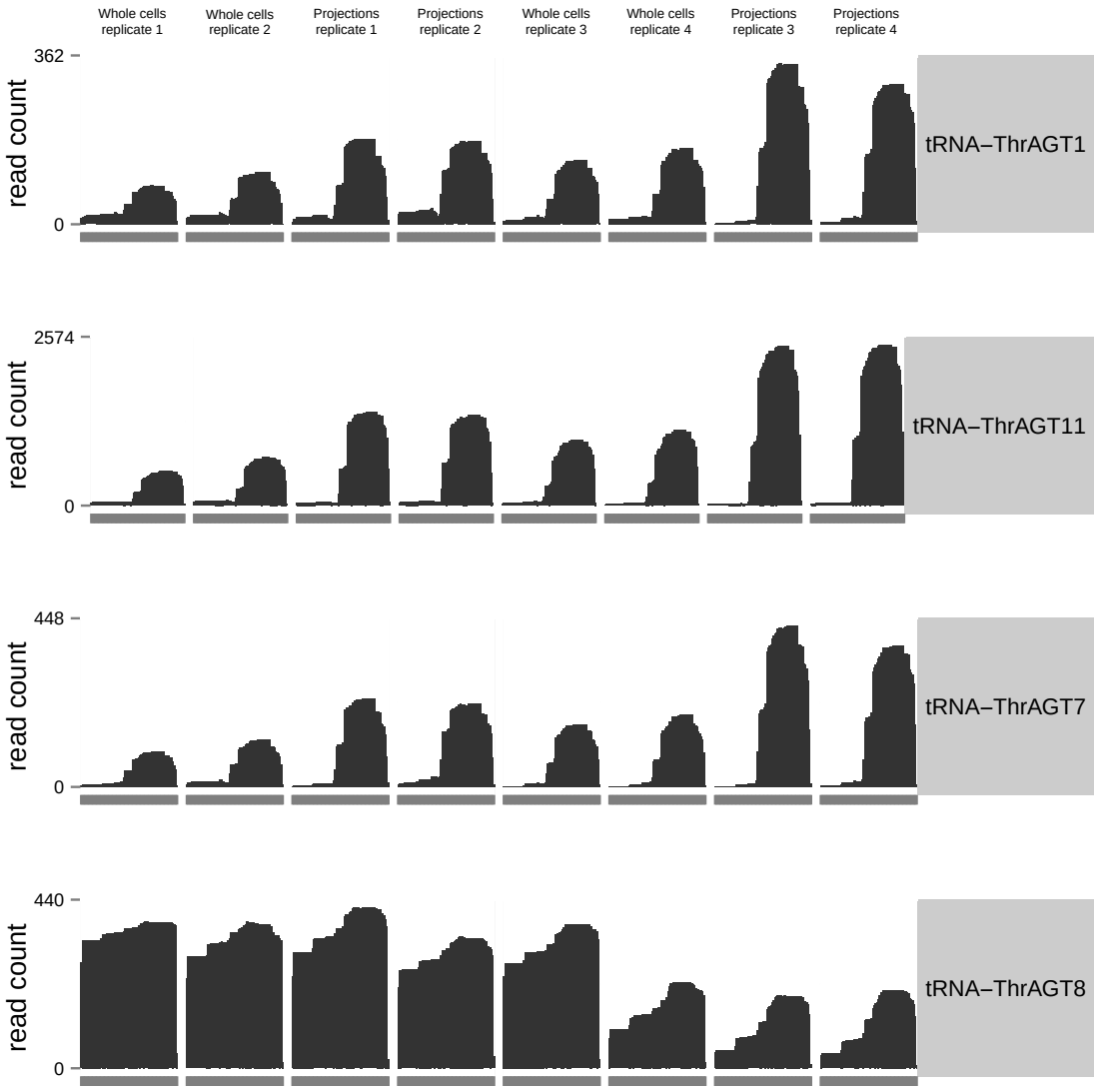

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
2908
0
tRNA–PheGAA4
683
0
tRNA–ProAGG124
87201
0
tRNA–ProAGG136
89619
0
tRNA–ProCGG1

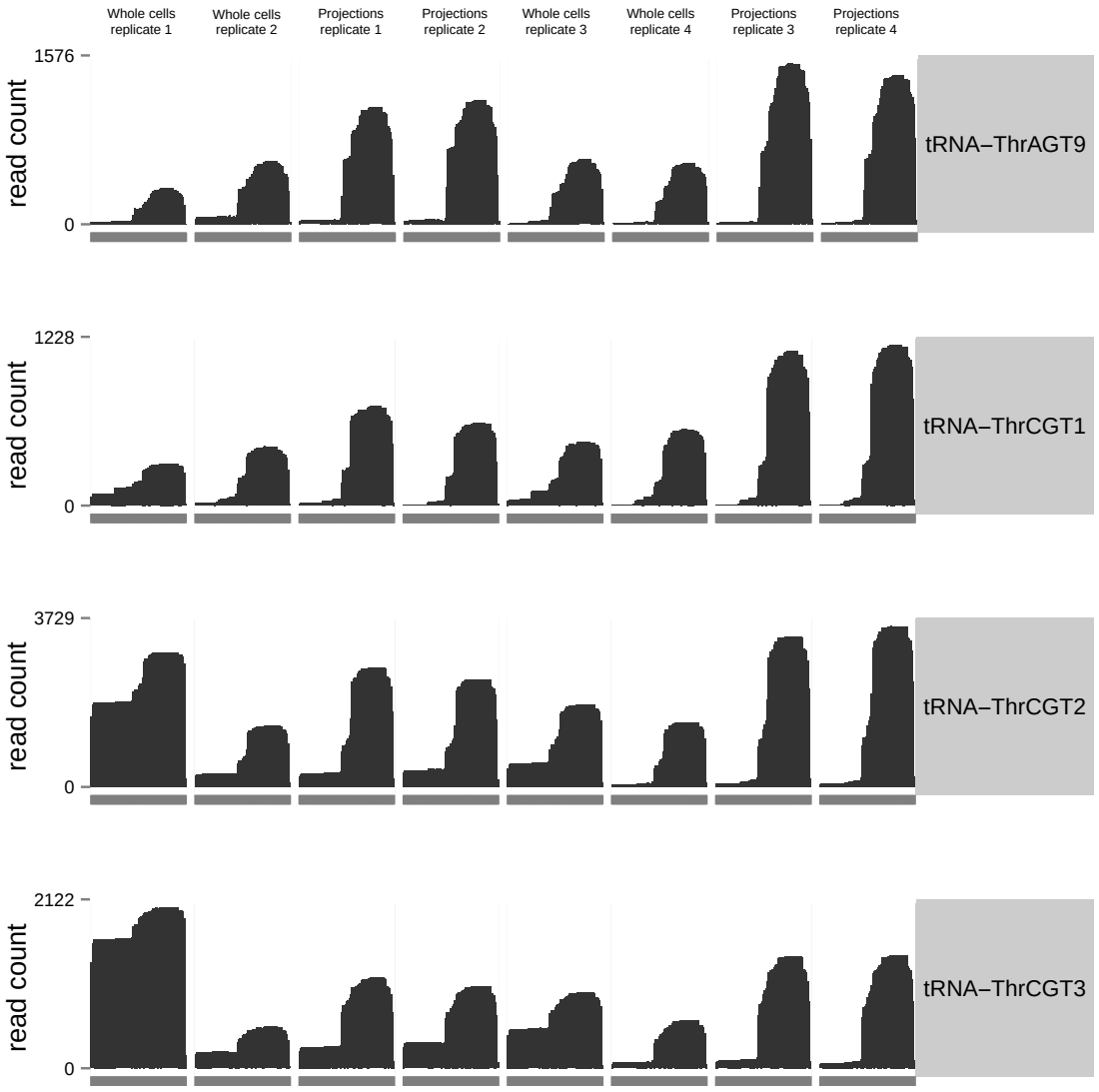

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
17480
0
tRNA−ProTGG2
89759
tRNA−ProTGG3
2592
tRNA−SerAGA19
7578
tRNA−SerAGA21

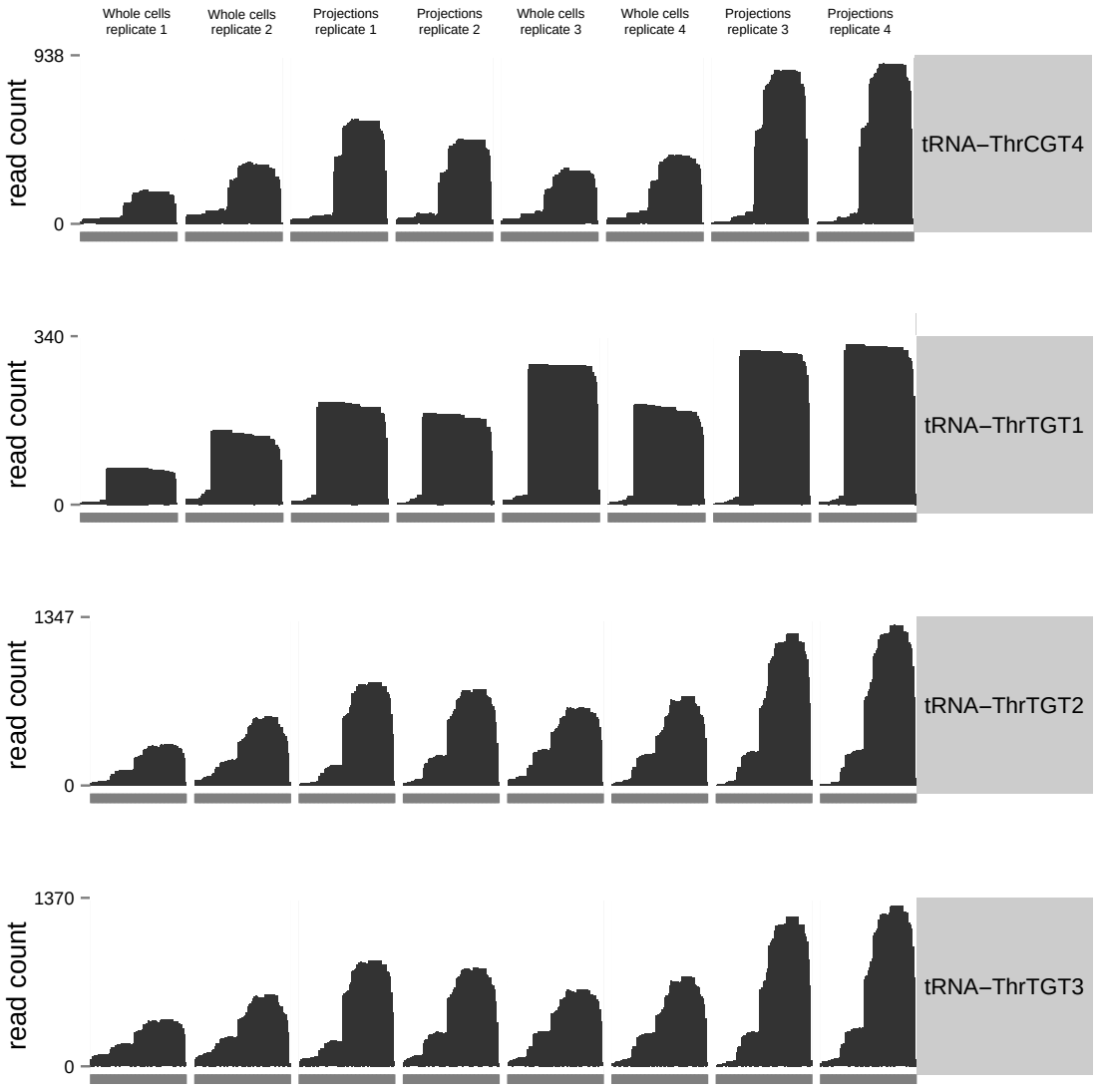

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
7791
read count
0
tRNA−SerAGA22
507
read count
0
tRNA−SerAGA8
7993
read count
0
tRNA−SerCGA1
367
read count
0
tRNA−SerCGA2

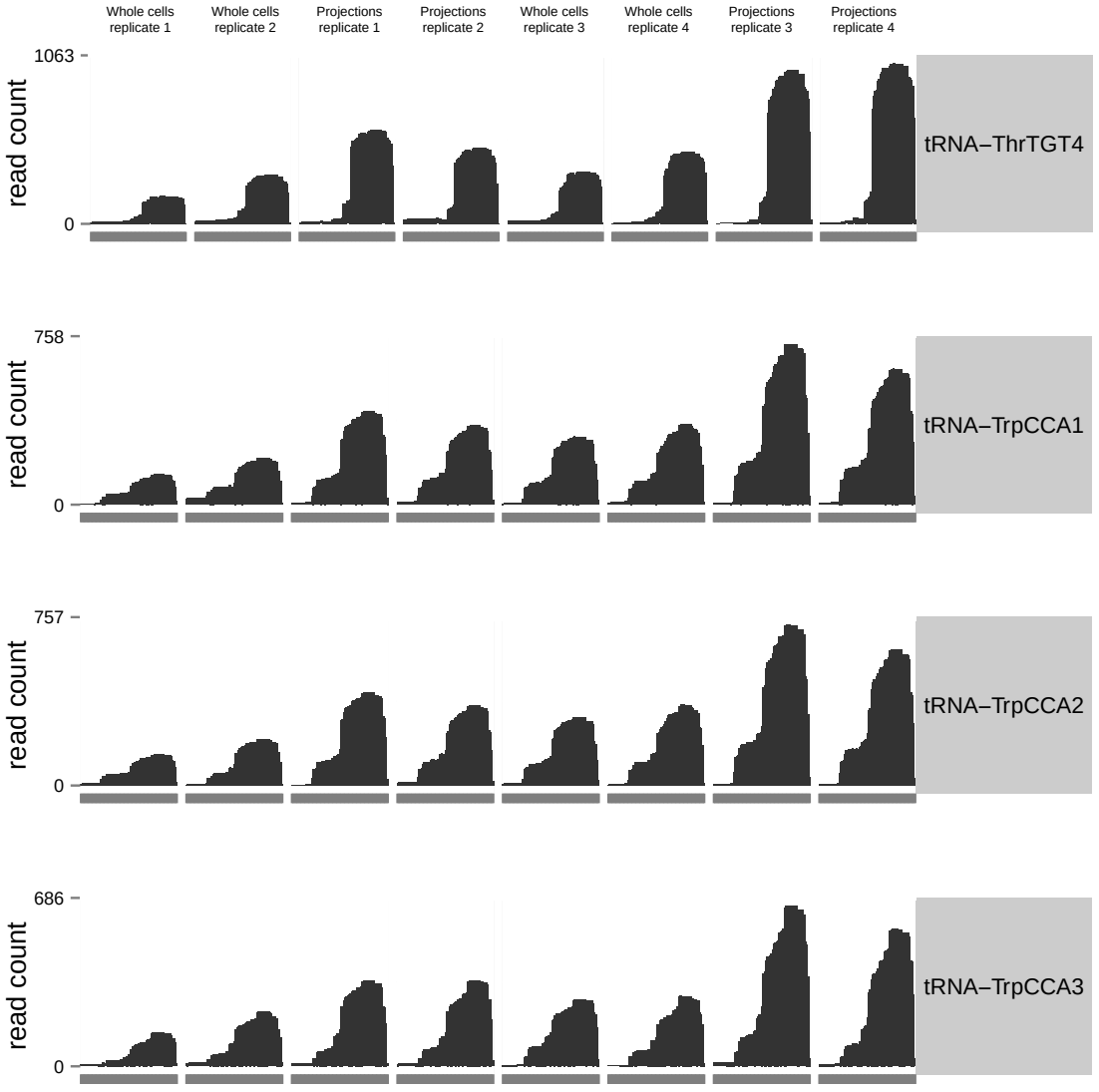

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
4745
0
tRNA−SerCGA3
17173
0
tRNA−SerGCT11
32474
0
tRNA−SerGCT2
659
0
tRNA−SerGCT4

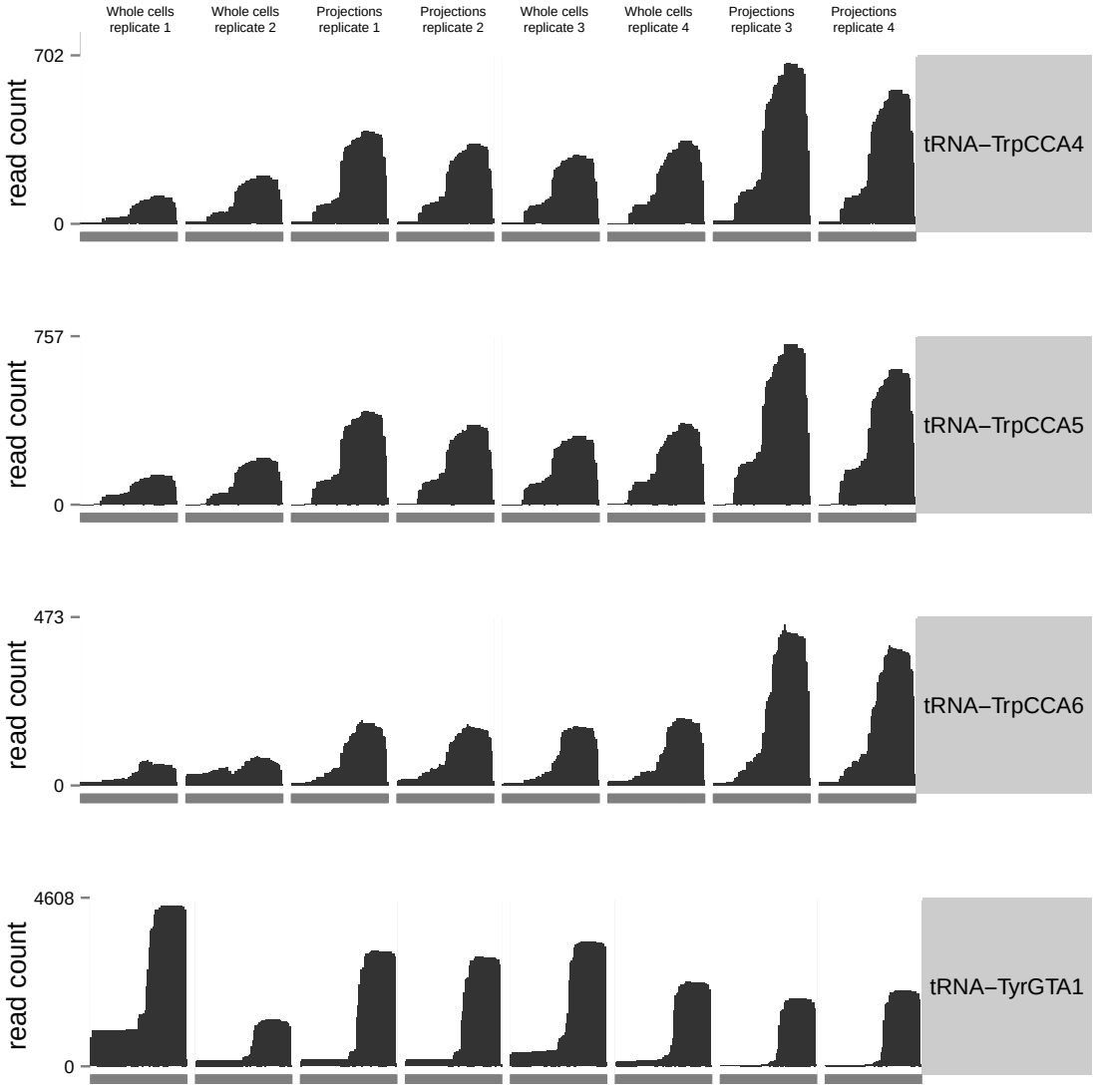

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
7471
read count
0
tRNA–SerTGA1
6661
read count
0
tRNA–SerTGA2
5231
read count
0
tRNA–SerTGA3
2606
read count
0
tRNA–SerTGA4

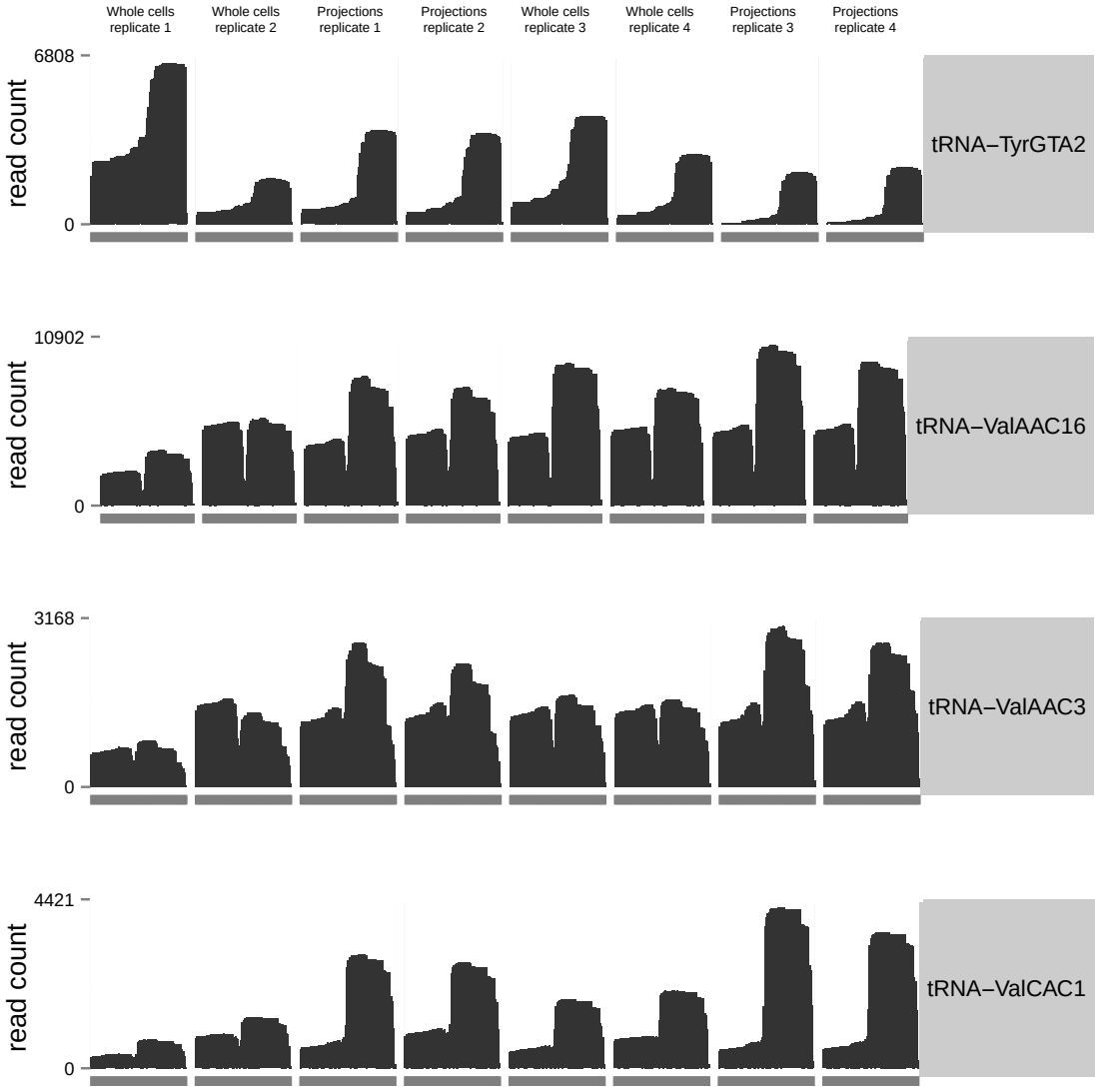

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
362
0
tRNA−ThrAGT1
2574
0
tRNA−ThrAGT11
448
0
tRNA−ThrAGT7
440
0
tRNA−ThrAGT8

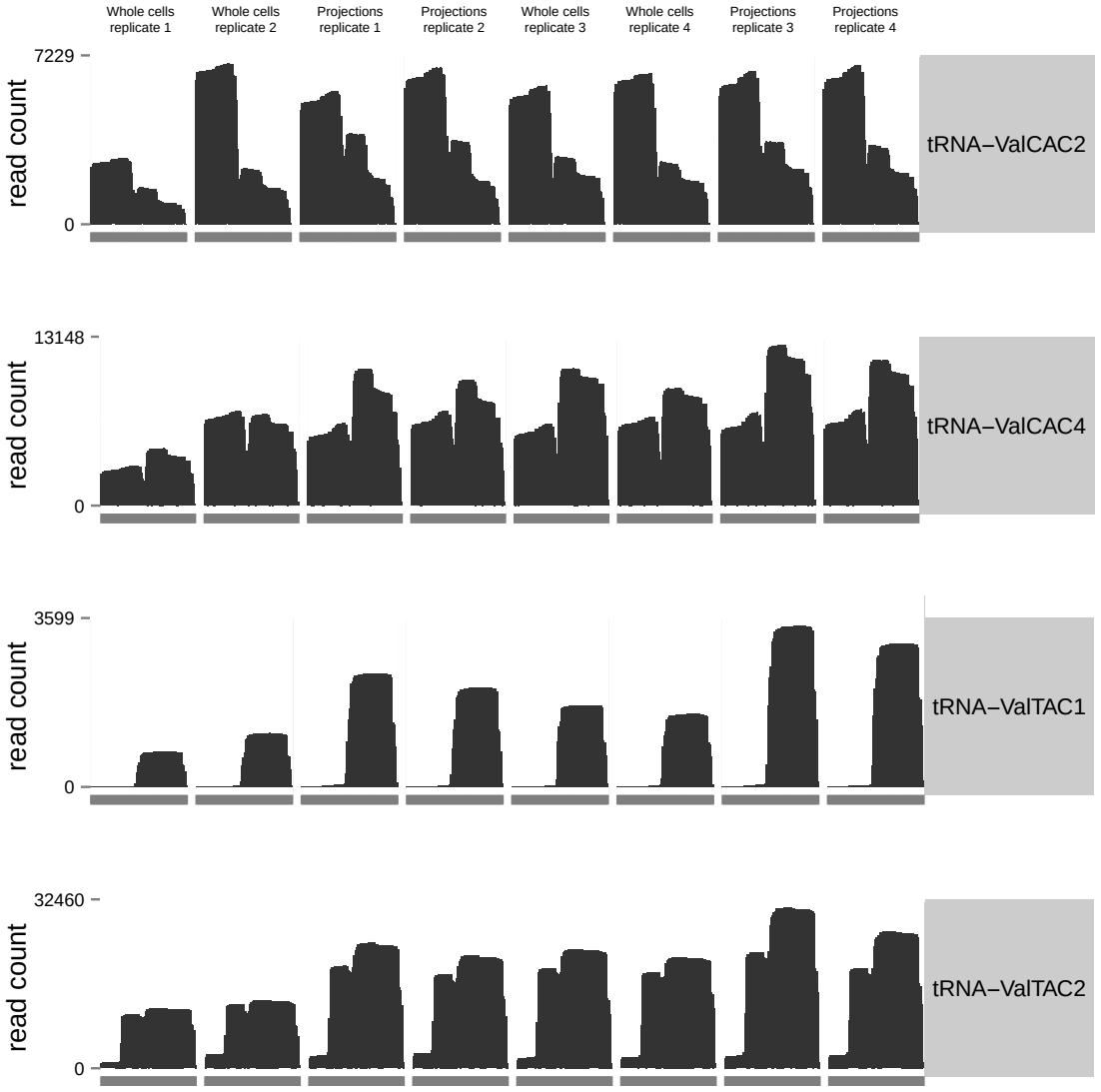

Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
1576
0
tRNA−ThrAGT9
1228
tRNA−ThrCGT1
3729
tRNA−ThrCGT2
2122
tRNA−ThrCGT3


Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
938
0
tRNA−ThrCGT4
340
0
tRNA−ThrTGT1
1347
0
tRNA−ThrTGT2
1370
0
tRNA−ThrTGT3


Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
1063
0
tRNA−ThrTGT4
758
0
tRNA−TrpCCA1
757
0
tRNA−TrpCCA2
686
0
tRNA−TrpCCA3


Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
702
0
tRNA−TrpCCA4
757
0
tRNA−TrpCCA5
473
0
tRNA−TrpCCA6
4608
0
tRNA−TyrGTA1


Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
read count
6808
0
tRNA–TyrGTA2
10902
0
tRNA–ValAAC16
3168
0
tRNA–ValAAC3
4421
0
tRNA–ValCAC1


Whole cells replicate 1
Whole cells replicate 2
Projections replicate 1
Projections replicate 2
Whole cells replicate 3
Whole cells replicate 4
Projections replicate 3
Projections replicate 4
7229
read count
0
tRNA–ValCAC2
13148
read count
0
tRNA–ValCAC4
3599
read count
0
tRNA–ValTAC1
32460
read count
0
tRNA–ValTAC2

Bibliography

Ainsley, J. A., Drane, L., Jacobs, J., Kittelberger, K. A., and Reijmers, L. G. (2014). Functionally diverse dendritic mRNAs rapidly associate with ribosomes following a novel experience. *Nature Communications*, 5(1):4510.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, 22(10):2008–2017.

Ashar-Patel, A., Kaymaz, Y., Rajakumar, A., Bailey, J. A., Karumanchi, S. A., and Moore, M. J. (2017). FLT1 and transcriptome-wide polyadenylation site (PAS) analysis in preeclampsia. *Scientific Reports*, 7(1):12139.

Back, S. H., Lee, K., Vink, E., and Kaufman, R. J. (2006). Cytoplasmic IRE1alpha-mediated XBP1 mRNA splicing in the absence of nuclear processing and endoplasmic reticulum stress. *The Journal of Biological Chemistry*, 281(27):18691–706.

Baralle, F. E. and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, 18(7):437–451.

Barbarese, E., Barry, C., Chou, C.-H. J., Goldstein, D. J., Nakos, G. A., Hyde-DeRuyscher, R., Scheld, K., and Carson, J. H. (1988). Expression and Localization of Myelin Basic Protein in Oligodendrocytes and Transfected Fibroblasts. *Journal of Neurochemistry*, 51(6):1737–1745.

Bashirullah, A., Cooperstock, R. L., and Lipshitz, H. D. (2001). Spatial and temporal control of RNA stability. *Proceedings of the National Academy of Sciences*, 98(13):7025–7028.

Bell, T. J., Miyashiro, K. Y., Sul, J.-Y., Buckley, P. T., Lee, M. T., et al. (2010). Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. *Proceedings of the National Academy of Sciences*, 107(49):21152–7.

Bell, T. J., Miyashiro, K. Y., Sul, J.-Y., McCullough, R., Buckley, P. T., et al. (2008). Cytoplasmic BKCa channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. *Proceedings of the National Academy of Sciences*, 105(6):1901–1906.

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, 74(8):3171–3175.

Bicknell, A. A., Cenik, C., Chua, H. N., Roth, F. P., and Moore, M. J. (2012). Introns in UTRs: why we should stop ignoring them. *BioEssays*, 34(12):1025–34.

Blanco, S., Bandiera, R., Popis, M., Hussain, S., Lombard, P., et al. (2016). Stem cell function and stress response are controlled by protein synthesis. *Nature*, 534(7607):335–340.

Boutz, P. L., Bhutkar, A., and Sharp, P. A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes & Development*, 29(1):63–80.

Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, 24(11):1774–1786.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527.

Briese, M., Saal, L., Appenzeller, S., Moradi, M., Baluapuri, A., and Sendtner, M. (2016). Whole transcriptome profiling reveals the RNA content of motor axons. *Nucleic Acids Research*, 44(4):e33–e33.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, 71(3):527–42.

Buckley, P. T., Lee, M. T., Sul, J.-Y., Miyashiro, K. Y., Bell, T. J., Fisher, S. A., Kim, J., and Eberwine, J. (2011). Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron*, 69(5):877–84.

Cajigas, I. J., Tushev, G., Will, T. J., tom Dieck, S., Fuerst, N., and Schuman, E. M. (2012). The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. *Neuron*, 74(3):453–466.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. M., and Pascual-Montano, A. (2007). GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology*, 8(1):R3.

Chan, Y. L., Gutell, R., Noller, H. F., and Wool, I. G. (1984). The nucleotide sequence of a rat 18 S ribosomal ribonucleic acid gene and a proposal for the secondary structure of 18 S ribosomal ribonucleic acid. *The Journal of Biological Chemistry*, 259(1):224–30.

Chen, Z., Gore, B. B., Long, H., Ma, L., and Tessier-Lavigne, M. (2008). Alternative Splicing of the Robo3 Axon Guidance Receptor Governs the Midline

Switch from Attraction to Repulsion. *Neuron*, 58(3):325–332.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1):1–8.

Clement, J. Q., Qian, L., Kaplinsky, N., and Wilkinson, M. F. (1999). The stability and fate of a spliced intron from vertebrate cells. *RNA*, 5(2):206–20.

Colak, D., Ji, S.-J., Porse, B. T. T., and Jaffrey, S. R. R. (2013). Regulation of Axon Guidance by Compartmentalized Nonsense-Mediated mRNA Decay. *Cell*, 153(6):1252–1265.

Cozen, A. E., Quartley, E., Holmes, A. D., Hrabeta-Robinson, E., Phizicky, E. M., and Lowe, T. M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature Methods*, 12(9):879–84.

Crick, F. H. (1966). Codon–anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology*, 19(2):548–55.

Curtis, H. J., Sibley, C. R., and Wood, M. J. A. (2012). Mirtrons, an emerging class of atypical miRNA. *Wiley Interdisciplinary Reviews: RNA*, 3(5):617–632.

Denis, M. M., Tolley, N. D., Bunting, M., Schwertz, H., Jiang, H., et al. (2005). Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell*, 122(3):379–91.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular cell biology*, 3(3):195–205.

Faisal, A. A., White, J. A., and Laughlin, S. B. (2005). Ion-Channel Noise Places Limits on the Miniaturization of the Brain's Wiring. *Current Biology*, 15(12):1143–1149.

Forrest, K. M. and Gavis, E. R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. *Current Biology*, 13(14):1159–68.

Gardner, E. J., Nizami, Z. F., Talbot, C. C., and Gall, J. G. (2012). Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. *Genes & Development*, 26(22):2550–9.

Garner, C. C., Tucker, R. P., and Matus, A. (1988). Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites. *Nature*, 336(6200):674–677.

Giorgi, C., Yeo, G. W., Stone, M. E., Katz, D. B., Burge, C., Turrigiano, G., and Moore, M. J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell*, 130(1):179–91.

Glanzer, J., Miyashiro, K. Y., Sul, J.-Y., Barrett, L., Belt, B., Haydon, P., and Eberwine, J. (2005). RNA splicing capability of live neuronal dendrites. *Proceedings of the National Academy of Sciences*, 102(46):16859–64.

Goodarzi, H., Liu, X., Nguyen, H., Zhang, S., Fish, L., and Tavazoie, S. (2015). Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell*, 161(4):790–802.

Goodenbour, J. M. and Pan, T. (2006). Diversity of tRNA genes in eukaryotes. *Nucleic Acids Research*, 34(21):6137–6146.

Gumy, L. F., Yeo, G. S. H., Tung, Y.-C. L., Zivraj, K. H., Willis, D., Coppola, G., Lam, B. Y. H., Twiss, J. L., Holt, C. E., and Fawcett, J. W. (2011). Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. *RNA*, 17(1):85–98.

Hadjiolov, A., Georgiev, O., Nosikov, V., and Yavachev, L. (1984). Primary and secondary structure of rat 28 S ribosomal RNA. *Nucleic Acids Research*, 12(8):3677–3693.

Hayashi, A., Kasahara, T., Iwamoto, K., Ishiwata, M., Kametani, M., Kakiuchi, C., Furuichi, T., and Kato, T. (2007). The Role of Brain-derived Neurotrophic Factor (BDNF)-induced XBP1 Splicing during Brain Development. *Journal of Biological Chemistry*, 282(47):34525–34534.

Hesselberth, J. R. (2013). Lives that introns lead after splicing. *Wiley Interdisciplinary Reviews: RNA*, 4(6):677–691.

Heyer, E. E., Ozadam, H., Ricci, E. P., Cenik, C., and Moore, M. J. (2015). An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Research*, 43(1):e2–e2.

Hilleren, P. J. and Parker, R. (2003). Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. *Molecular Cell*, 12(6):1453–65.

Holt, C. E. and Bullock, S. L. (2009). Subcellular mRNA Localization in Animal Cells and Why It Matters. *Science*, 326(5957):1212–1216.

Hubé, F. and Francastel, C. (2015). Mammalian introns: when the junk generates molecular diversity. *International Journal of Molecular Sciences*, 16(3):4429–52.

Hüttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G. J., Condeelis, J., and Singer, R. H. (2005). Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature*, 438(7067):512–5.

Jacquier, A. and Rosbash, M. (1986). RNA splicing and intron turnover are greatly diminished by a mutant yeast branch point. *Proceedings of the National Academy of Sciences*, 83(16):5835–9.

Jurica, M. S. and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molecular Cell*, 12(1):5–14.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics*, 16(9):418–20.

Katz, Z. B., English, B. P., Lionnet, T., Yoon, Y. J., Monnier, N., Ovryn, B., Bathe, M., and Singer, R. H. (2016). Mapping translation 'hot-spots' in live cells by

tracking single molecules of mRNA and ribosomes. *eLife*, 5(JANUARY2016):1–16.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6):996–1006.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, 26(17):2204–2207.

Khaladkar, M., Buckley, P. T., Lee, M. T., Francis, C., Eghbal, M. M., Chuong, T., Suresh, S., Kuhn, B., Eberwine, J., and Kim, J. (2013). Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons. *PloS One*, 8(10):e76194.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36.

Koltun, B., Ironi, S., Gershoni-Emek, N., Barrera, I., Hleihil, M., Nanguneri, S., Sasmal, R., Agasti, S. S., Nair, D., and Rosenblum, K. (2019). Measuring mRNA translation in neuronal processes and somata by tRNA-FRET. *bioRxiv*, page DOI: 10.1101/646216.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–9.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Limbach, P. A. and Paulines, M. J. (2017). Going global: the new era of mapping modifications in RNA. *Wiley Interdisciplinary Reviews. RNA*, 8(1):e1367.

Litman, P., Barg, J., Rindzoonski, L., and Ginzburg, I. (1993). Subcellular localization of tau mRNA in differentiating neuronal cell culture: implications for neuronal polarity. *Neuron*, 10(4):627–38.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). *Molecular cell biology*. W.H. Freeman.

Lowe, T. M. and Eddy, S. R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, 25(5):955–964.

Maday, S., Twelvetrees, A. E. E., Moughamian, A. J. J., and Holzbaur, E. L. L. (2014). Axonal Transport: Cargo-Specific Mechanisms of Motility and Regulation. *Neuron*, 84(2):292–309.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10.

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. *Neuron*, 92(6):1266–1278.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–338.

Meyuhas, O., Thompson, E. A., and Perry, R. P. (1987). Glucocorticoids selectively inhibit translation of ribosomal protein mRNAs in P1798 lymphosarcoma cells. *Molecular and Cellular Biology*, 7(8):2691–9.

Minis, A., Dahary, D., Manor, O., Leshkowitz, D., Pilpel, Y., and Yaron, A. (2014). Subcellular transcriptomics-dissection of the mRNA composition in the axonal compartment of sensory neurons. *Developmental Neurobiology*, 74(3):365–81.

Miyashiro, K., Dichter, M., and Eberwine, J. (1994). On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. *Proceedings of the National Academy of Sciences*, 91(23):10800–4.

Moccia, R., Chen, D., Lyles, V., Kapuya, E., E, Y., et al. (2003). An unbiased cDNA library prepared from isolated Aplysia sensory neuron processes is enriched for cytoskeletal and translational mRNAs. *The Journal of Neuroscience*, 23(28):9409–17.

Moore, M. J. (2002). Nuclear RNA turnover. *Cell*, 108(4):431–4.

Morgan, J. T., Fink, G. R., and Bartel, D. P. (2019). Excised linear introns regulate growth in yeast. *Nature*, 565(7741):606–611.

Müller, C., Bauer, N. M., Schäfer, I., and White, R. (2013). Making myelin basic protein -from mRNA transport to localized translation. *Frontiers in Cellular Neuroscience*, 7:169.

Munro, T. P., Magee, R. J., Kidd, G. J., Carson, J. H., Barbarese, E., Smith, L. M., and Smith, R. (1999). Mutational Analysis of a Heterogeneous Nuclear Ribonucleoprotein A2 Response Element for RNA Trafficking. *Journal of Biological Chemistry*, 274(48):34389–34395.

Nakayama, K., Ohashi, R., Shinoda, Y., Yamazaki, M., Abe, M., et al. (2017). RNG105/caprin1, an RNA granule protein for dendritic mRNA localization, is essential for long-term memory formation. *eLife*, 6:DOI: 10.7554/eLife.29677.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics*, 28(14):1919–1920.

Nilsen, T. W. and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280):457–463.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745.

Ooi, S. L., Dann, C., Nam, K., Leahy, D. J., Damha, M. J., and Boeke, J. D. (2001). RNA lariat debranching enzyme. *Methods in Enzymology*, 342:233–48.

Ooi, S. L., Samarsky, D. A., Fournier, M. J., and Boeke, J. D. (1998). Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: Intron length effects and activity of a precursor snoRNA. *RNA*, 4(9):1096–110.

Ortiz, R., Georgieva, M. V., Gutiérrez, S., Pedraza, N., Fernández-Moya, S. M., and Gallego, C. (2017). Recruitment of Staufen2 Enhances Dendritic Localization of an Intron-Containing CaMKIIα mRNA. *Cell Reports*, 20(1):13–20.

Pai, A. A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., and Burge, C. B. (2017). The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *eLife*, 6:DOI: 10.7554/eLife.32537.

Palay, S. L. (1956). Synapses in the central nervous system. *The Journal of Biophysical and Biochemical Cytology*, 2(4 Suppl):193–202.

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Research*, 28(4):395–404.

Pandya-Jones, A., Bhatt, D. M., Lin, C.-H., Tong, A.-J., Smale, S. T., and Black, D. L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA*, 19(6):811–827.

Parenteau, J., Maignon, L., Berthoumieux, M., Catala, M., Gagnon, V., and Abou Elela, S. (2019). Introns are mediators of cell response to starvation. *Nature*, 565(7741):612–617.

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, 14(7):687–690.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20(1):110–21.

Poon, M. M., Choi, S.-H., Jamieson, C. A. M., Geschwind, D. H., and Martin, K. C. (2006). Identification of process-localized mRNAs from cultured rodent hippocampal neurons. *The Journal of Neuroscience*, 26(51):13390–9.

Poulopoulos, A., Murphy, A. J., Ozkan, A., Davis, P., Hatch, J., Kirchner, R., and Macklis, J. D. (2019). Subcellular transcriptomes and proteomes of developing axon projections in the cerebral cortex. *Nature*, 565(7739):356–360.

Puthanveettil, S. V., Antonov, I., Kalachikov, S., Rajasethupathy, P., Choi, Y.-B.,

et al. (2013). A strategy to capture and characterize the synaptic transcriptome. *Proceedings of the National Academy of Sciences*, 110(18):7464–7469.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2.

Ricci, E. P., Kucukural, A., Cenik, C., Mercier, B. C., Singh, G., Heyer, E. E., Ashar-Patel, A., Peng, L., and Moore, M. J. (2014). Staufen1 senses overall transcript secondary structure to regulate translation. *Nature Structural and Molecular Biology*, 21(1):26–35.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1):24–26.

Rook, M. S., Lu, M., and Kosik, K. S. (2000). CaMKIIalpha 3' untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage. *The Journal of Neuroscience*, 20(17):6385–93.

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L., and Singer, R. H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. *Molecular and Cellular Biology*, 17(4):2158–65.

Roy, B. and Jacobson, A. (2013). The intimate relationships of mRNA decay and translation. *Trends in Genetics*, 29(12):691–9.

Ruskin, B. and Green, M. (1985). An RNA processing activity that debranches RNA lariats. *Science*, 229(4709):135–140.

Sarthy, P. V., Fu, M., and Huang, J. (1989). Subcellular localization of an intermediate filament protein and its mRNA in glial cells. *Molecular and Cellular Biology*, 9(10):4556–9.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7(1):3.

Sharangdhar, T., Sugimoto, Y., Heraud-Farlow, J., Fernández-Moya, S. M., Ehses, J., Ruiz de Los Mozos, I., Ule, J., and Kiebler, M. A. (2017). A retained intron in the 3'-UTR of Calm3 mRNA mediates its Staufen2- and activity-dependent localization to neuronal dendrites. *EMBO Reports*, 18(10):1762–1774.

Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, 351(6271):391–396.

Sheppard, S., Lawson, N. D., and Zhu, L. J. (2013). Accurate identification of polyadenylation sites from 3' end deep sequencing using a naive Bayes classifier. *Bioinformatics*, 29(20):2564–71.

Shigeoka, T., Koppers, M., Wong, H. H.-W., Lin, J. Q., Dwivedy, A., et al. (2018). On-site ribosome remodeling by locally synthesized ribosomal proteins in axons. *bioRxiv*, page DOI: 10.1101/500033.

Singh, J. and Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, 16(11):1128–1133.

Steitz, J. A., Dreyfuss, G., Krainer, A. R., Lamond, A. I., Matera, A. G., and Padgett, R. A. (2008). Where in the cell is the minor spliceosome? *Proceedings of the National Academy of Sciences*, 105(25):8485–8486.

Stępkowski, T. M., Męczyńska-Wielgosz, S., and Kruszewski, M. (2017). mitoLUHMES: An Engineered Neuronal Cell Line for the Analysis of the Motility of Mitochondria. *Cellular and Molecular Neurobiology*, 37(6):1055–1066.

Sterling, P. and Laughlin, S. (2015). *Principles of Neural Design*. The MIT Press.

Steward, O. and Falk, P. M. (1985). Polyribosomes under developing spine synapses: growth specializations of dendrites at sites of synaptogenesis. *Journal of Neuroscience Research*, 13(1-2):75–88.

Taggart, A. J., Lin, C.-L., Shrestha, B., Heintzelman, C., Kim, S., and Fairbrother, W. G. (2017). Large-scale analysis of branchpoint usage across species and cell lines. *Genome Research*, 27(4):639–649.

Talhouarne, G. J. and Gall, J. G. (2014). Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes. *RNA*, 20(9):1476–1487.

Talhouarne, G. J. S. and Gall, J. G. (2018). Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proceedings of the National Academy of Sciences*, 115(34):E7970–E7977.

Taliaferro, J. M. M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., et al. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell*, 61(6):821–833.

Taylor, A. M., Berchtold, N. C., Perreau, V. M., Tu, C. H., Li Jeon, N., and Cotman, C. W. (2009). Axonal mRNA in uninjured and regenerating cortical mammalian axons. *The Journal of Neuroscience*, 29(15):4697–707.

Taylor, A. M., Blurton-Jones, M., Rhee, S. W., Cribbs, D. H., Cotman, C. W., and Jeon, N. L. (2005). A microfluidic culture platform for CNS axonal injury, regeneration and transport. *Nature Methods*, 2(8):599–605.

Taylor, J. P., Brown, R. H., and Cleveland, D. W. (2016). Decoding ALS: from genes to mechanism. *Nature*, 539(7628):197–206.

Thompson, D. M. and Parker, R. (2009). Stressing out over tRNA cleavage. *Cell*, 138(2):215–9.

Trapp, B. D., Moench, T., Pulley, M., Barbosa, E., Tennekoon, G., and Griffin, J. (1987). Spatial segregation of mRNA encoding myelin-specific proteins. *Proceedings of the National Academy of Sciences*, 84(21):7773–7777.

Will, C. L. and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology*, 13(3):290–301.

Yamasaki, S., Ivanov, P., Hu, G.-f., and Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. *The Journal of Cell Biology*, 185(1):35–42.

Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., and Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes & Development*, 26(11):1209–23.

Zelená, J. (1970). Ribosome-like particles in myelinated axons of the rat. *Brain Research*, 24(2):359–363.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., et al. (2018). Ensembl 2018. *Nucleic Acids Research*, 46(D1):D754–D761.

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., and Chen, L.-L. (2013). Circular intronic long noncoding RNAs. *Molecular Cell*, 51(6):792–806.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell*, 133(4):585–600.

Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence*, 3(1):9.

Zhu, Y., Machleder, E., Chenchik, A., Li, R., and Siebert, P. (2001). Reverse Transcriptase Template Switching: A SMART™ Approach for Full-Length cDNA Library Construction. *BioTechniques*, 30(4):892–897.

Zivraj, K. H., Tung, Y. C. L., Piper, M., Gumy, L., Fawcett, J. W., Yeo, G. S. H., and Holt, C. E. (2010). Subcellular Profiling Reveals Distinct and Developmentally Regulated Repertoire of Growth Cone mRNAs. *Journal of Neuroscience*, 30(46):15464–15478.