Defining Protein Synthesis:
New Technologies to Elucidate Translational Control

Nicholas J. Hornstein

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

Abstract

Defining Protein Synthesis: New Technologies to Elucidate Translational Control

Nicholas J. Hornstein

Protein translation has emerged as an important mediator of cellular activity. Here, we discuss efforts to develop and apply new technologies designed to gain insights into translational control. We begin with the application of ribosome profiling to a RiboTag Glioma mouse model which enables translational profiling of transformed cellular populations. This approach demonstrates a number of abnormalities of translation in transformed cells. We go on to report the development of an inexpensive and rapid library preparation methodology which enables high-throughput sequencing of ribosome-protected footprints from small amounts of input material. We apply this technique to a CAMKII RiboTag mouse model to make new insights into cell-type specific translation. Finally, we describe efforts to investigate translation regulatory networks through the development of a technique which couples large-scale perturbation with a genome-wide readout of translation.

Molecular dissection of tissues through the ectopic expression of modified ribosomal proteins commonly relies on tissue-specific genes which act as drivers. In the case of glioma, a gene specific to transformed tissue, but not expressed in normal brain tissue, has not been identified. Chapter 2 focuses on efforts to bypass this through the development of a RiboTag Glioma mouse model which allows for concurrent transformation and the expression of an epitope-tagged ribosomal protein in virally infected cells. This model made possible the isolation of translating mRNA from transformed cellular populations and was used to demonstrate the existence of a number of translational abnormalities in transformed cells.

Conventional ribosome profiling is a powerful tool which allows for the identification of ribosome-protected mRNA footprints. However, it is time-consuming, expensive, and difficult to implement. Based on our experiences with conventional ribosome profiling, we sought to develop a method which could decrease the overall number of enzymatic reactions and purification steps, thereby reducing the time and cost associated with the procedure; these efforts are discussed in Chapter 3. Utilizing a ligation-free library preparation process, which incorporates poly(A)-polymerase, template switching and bead-based purification, we reduced the time, costs and input requirements required to generate a ribosome profiling library while maintaining high library complexity. We applied our ligation-free ribosome profiling technique to a CAMKII RiboTag mouse model which enabled us to identify patterns of cell-type specific translation and the effects of mTOR inhibition in CAMKII-expressing excitatory neurons.

Regulation of protein expression is an essential and highly complex cellular activity. Aberrations of translational control are central to a host of pathologies and have direct clinical relevance. However, our knowledge of the networks which control translation is limited. Chapter 4 details our efforts to develop a highly-scalable technology which enables the identification of gene-specific translational alteration in response to perturbation. Coupled with a large-scale perturbation screen, this technique could lead to the generation of a network for translational control, similar to efforts previously undertaken to understand transcriptional control. By combining the recently developed PLATE-Seq method, which utilizes unique barcode identifiers and pooled library construction, with a technique for the identification and isolation of ribosome associated mRNA, we are able to rapidly and inexpensively determine genome-wide translational states in a highly scalable format.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

Before delving into my acknowledgements I would like to first thank my Thesis and Defense Committees for their valuable advice, comments and guidance. Drs. Donna Farber, Harris Wang, Peter Canoll, Dana Pe'er, and Chaolin Zhang. My growth as a scientist is directly due to the time you have invested in me. Thanks especially to Dr. Peter Canoll who allowed me to rotate in his laboratory and gave me a strong introduction to neuropathology both in his lab and in my medical coursework.

A number of other individuals have been extremely giving of their time, and I have no doubt that I would not be writing this Dissertation today without their help and mentorship. To all of the members of the Sims lab, past and present, thank you providing a wonderful environment in which to perform my PhD research. Specifically, Dr. Jinzhou Yuan for his sagely advice and probing questions; Erin Bush for her positive attitude and wizard-like library preparation skills; Sohani Das Sharma for always lending a helping hand when projects became unmanageable; and Sayantan Bose for introducing me to sequencing. Dr. Christian Gonzalez should be noted especially for, among many lessons, teaching me how to pour my first gradient and introducing me to ribosome profiling while also being a good friend. Dr. Gonzalez contributed the majority of work towards developing ribosome profiling in the Sims lab and this is well-reflected in his being the primary author of the manuscript we submitted which has served as the basis of my second chapter. It should also be noted that Erin Bush and Forest Rey of the Califano lab developed the PLATE-Seq protocol which I discuss in chapter 4. Also thanks to Guomei Tang for generating the CamK-cre-RiboTag mice used in these studies and Daniela Torres for performing the CAMKII-related animal treatments, western blots and immunofluorescence assays used in Chapter 3 experiments.

Additionally, a number of members of the Canoll and Bruce labs contributed to work discussed in my second chapter including Dr. Jennifer Sims, Dr. Angeliki Mela, Franklin Garcia, Dr. Liang Lei, Dr. David A. Gass, and Benjamin Amendolara. Nathalie Bolduc and Andrew Farmer from Clontech Laboratories were extremely helpful in sharing valuable reagents and technical advice which made the work discussed in Chapter Three possible. Finally thanks to the core facilities at Columbia, especially Roxanne Ko and Erin Bush in the Genome Center for consistently going above and beyond in their services.

Because PhD training does not occur in a vacuum, I'd be remiss if I didn't also acknowledge the people in my life who enabled me to focus on research; Subha Perni for always keeping me grounded and my ego in check, and Jennifer Nguyen for taking care of me and showing me great kindness.

While all these individuals helped in some way in my path towards a PhD, the vast majority of credit should rightfully be attributed to my mentor, Dr. Peter Sims. While in the process of joining a lab, I recall receiving strongly worded advice regarding joining the lab of a faculty member who had not yet graduated a student. However, I could not have made a better choice. Dr.Sims invested more time in developing my scientific acumen than I could have any right to expect. Beyond helping with higher-level tasks such as designing and writing a fellowship application, Dr.Sims took an enormous amount of time to teach me more basic skills such as python programming and experimental design. His level of knowledge, ability to teach difficult concepts, and extreme patience are attributes I hope to emulate in the future. Thank you for the time you've spent training me; I am extremely proud to be the first graduate of the Sims Lab.

DEDICATION

This dissertation is dedicated to my parents, Victoria and James.

You've always believed in me, always fought for me, and above all, always loved me.

Thank you for helping to shape me into the man I am today.

CHAPTER 1

Introduction: The Ribosome in Health and Disease, A History

1.1 The Ribosome and Physiology

An Introduction to the Ribosome

Encompassing the lion's share of cellular energy use, protein production is critical for cell function, identity, and survival. The process by which protein production occurs is known as translation. First studied in the early 19th century, research into translation was prodigious through to the early 2000's. However, a lack of technology made it difficult to assess the detailed determinants of translational control on a genome-wide level. A steady series of observations beginning with the discovery of the start codon and existence of reading frames in 1969 by Joan Steitz have evolved to demonstrate that translation is a highly dynamic, selective process able to promote responses essential to cellular survival and integrity (Steitz 1969). While transcription is able to produce an astoundingly complex program of gene expression via control by master regulators and alternative splicing, it does so at the cost of alacrity. Many threats to cellular survival, such as heat, nutrient deprivation, viral infection or hypoxia, are so potently disruptive to cellular function that without a rapid shift of protein production, cellular viability would be compromised. In the case of nutrient deprivation, either by decreased energy or amino acid availability, the primary response –decreasing overall protein production—can reduce cellular ATP utilization by as much as 50%, conserving energy for other critical tasks (Rolfe and Brown 1997). However, there are a number of genes which resist translational inhibition during time of increased cellular stress, through a variety of mechanisms, including upstream open reading

frames or decreased reliance on ribosomal initiating proteins (Holcik et al. 2000, Richter et al. 2010, Spriggs et al. 2010). This ability to selectively enhance the production of proteins with protective functions is critical to translation's role as a rapid responder to both exogenous and endogenous stressors.

The mechanisms by which translational control is achieved are, in part, related to the structural composition of genes. In mammalian systems, the coding region (CDS) is typically flanked upstream and downstream by non-coding leader regions known as the 5' and 3' untranslated regions (UTRs) (Figure 1.1). While the effects of UTR sequence content on CDS translation have been quantified on a gene-by-gene basis, their role in translational control is still poorly understood. Recent work has identified several interesting trends regarding translational control of the CDS due to sequence content in the UTR. While alternative isoforms in rare cases can account for up to two orders of magnitude of translational control, other general trends have emerged such as a relationship between transcript's 3'-UTR length or GC content affecting overall CDS translation rate (Floor et al. 2015).

Much of what we know about translation has built on work by Sydney Brenner, Francis Crick, Leslie Barnett and R.J. Watts-Tobin which demonstrated the three-base periodicity of the genetic code. Published in 1961, the Crick-Brenner experiment, as it came to be known, showed that proflvain-induced insertions or deletions only made the T4 bacteriophage gene rIIB non-functional if one, two, or four bases were removed or inserted (Crick et al. 1961). Furthermore, they found that non-functional mutants could be rescued by further mutations which brought the number of inserted or deleted bases to either zero or a multiple of three. Based on these experiments they inferred that the genetic code was comprised of a three-base DNA codon which corresponded to individual amino acids. Due to the 4-nucleotide genetic sequence and existence

of 20 amino acids, it was inferred that there could be 64 possible codon sequences which allowed for redundancy. Knowledge of codons allowed later researchers the ability to associate specific nucleotide sequences with amino acids, as well as the start and stop codons.

Beyond genetic structure, the most important mediator of translation is the ribosome itself. Consisting of almost 100 proteins, co-factors and rRNA, the assembled ribosome requires an enormous investment of cellular resources to produce and is amongst the most highly evolutionarily conserved sets of proteins in cells (Fox 2010). Due to its large number of constituent proteins, multiple possible phosphorylation states, association with rRNA and highly energetic function, the ribosome has proven challenging to decode and define. Visualization of the ribosome with cryo-electron microscopy (cryo-EM) by Joachim Frank in 1995 revealed important ribosomal structures such as the bifurcating ribosomal tunnel (Frank et al. 1995). A sub 10-Å resolution X-ray crystal structure for the ribosome was obtained by Thomas Steitz as recently as 1998, clarifying the structure of the ribosome and identifying the location of its peptidyl transferase, leading to the 2009 Nobel Prize in Chemistry (Ban et al. 1998). More recently, cryo-EM work, again performed by Joachim Frank, allowed a visualization of the ribosomes structure while actively producing poly-peptides (Valle et al. 2003). These structural studies, combined with previous work by Peter Moore, created a clear picture of the constructed ribosome and its constituent elements (Moore et al. 1968). In the course of understanding the ribosome's function and regulation, it is important to introduce some key proteins which comprise it.

Translation Initiation and Elongation

The two major constituents of the Eukaryotic ribosome, the 40S and 60S subunits, associate with mRNA to form the active 80S complex (Figure 1.1). Of the three main functions of the ribosome during translation—initiation, elongation and termination—ribosomal initiation has been largely accepted, under most circumstances, as the rate limiting step of protein production (Jackson et al. 2010). Cap-dependent ribosomal translation initiation, the dominant form of initiation, begins with the recognition of the m^7Gppp cap by eIF4E as part of the EIF4F complex. Since m^7Gppp modification is one of the earliest steps of post-transcriptional modification, this allows for the efficient translation of mature mRNAs (Furuichi and Shatkin 2000). In mammalian systems, eIF4E complexes two additional proteins, eIF4A, a RNA helicase, and eIF4G, a scaffold protein, to serve as the basis for the eIF4F complex. Due to this complex's involvement in the critical step of 5' cap recognition, it is a key target of endogenous and exogenous translation inhibition. A classic and physiologically significant form of cap-mediated inhibition is caused by competition between the eIF4E binding proteins, 4EBPs, and eIF4G, which shares a structural similarity to the 4EBPs. Under normal conditions, 4EBPs are heavily phosphorylated and have a low affinity for eIF4E, allowing for eIF4E's normal association with the eIF4F complex, and by extension, cap-dependent translation (Pause et al. 1994). However, cellular stress responses lead to the dephosphorylation of 4EBPs, increasing their affinity for eIF4E and greatly diminishing cap-dependent translation as a result.

Beyond eIF4F, ribosomal initiation requires assembly of the ternary complex; an additional member of the pre-initiation complex. Consisting of a methionine charged tRNA capable of recognizing the AUG-start codon, as well as eIF2, an elongation protein involved in ribosomal translocation, and charged with a GTP molecule, the ternary complex can be regulated

via the phosphorylation state of eIF2 in a fashion similar to the regulation of eIF4E. During stress response, phosphorylation of the α-subunit of eIF2 leads to a reduced rate of GDP-GTP exchange, reducing overall formation of the ternary complex and by extension, translation initiation (Clemens 2004). Proper formation of eIF4F and the tertiary complex allows for recruitment of several other proteins, and most importantly, the 40S ribosomal subunit. This assembly of proteins, the pre-initiation complex, is then able to scan the mRNA beginning at its bound m^7Gppp cap until an AUG start-site is located and the 60S subunit is recruited to begin the process of polypeptide elongation.

The process of poly-peptide synthesis, although not the rate-limiting step during normal cap-dependent protein translation, can also exert some effect on specific protein synthesis rates, especially in extreme conditions. Following the introduction of methionine and the formation of the complete 80s ribosome, elongation occurs in a start-stop motion where charged tRNAs are able to introduce necessary amino acids and catalyze peptide bonds (Wolin and Walter 1988). While this occurs rapidly, specific amino acids such as proline and regions of mRNA with complex secondary structures have been shown to reduce elongation rates, leading to ribosome stalling (Peil et al. 2013). Altered elongation rates can also be achieved by modulating the phosphorylation states of ribosomal proteins, such as eEF2. Necessary for GTP-dependent ribosomal translocation, eEF2 is targeted by multiple growth-related regulatory pathways and has reduced ribosomal binding capacity and functionality in its de-phosphorylated state, which is observed in the presence of cellular stressors such as hypoxia and nutrient starvation (Leprivier et al. 2013). Elongation rates can also be controlled by proteins extrinsic to the ribosome. Chaperone proteins, which aid in the proper folding of elongating polypeptide chains, rest at the exit tunnel of the ribosome and can exert a negative effect on protein synthesis in their absence

(Liu and Qian 2014). These processes demonstrate the highly complex state of ribosomal elongation. However, except in states of extreme nutrient deprivation, ribosomal initiation remains the rate-limiting step of protein synthesis (Shah et al. 2013).

Mechanisms of Translational Control

Although multiple regulatory pathways affecting translation have had specific downstream effectors identified, the best understood with an established cis-regulatory element is the mechanistic target of Rapamycin (mTOR). mTOR exists as two structurally distinct serine/threonine kinases. While mTORC1 affects growth, proliferation, protein synthesis, and metabolism, mTORC2 interacts with major regulators of cellular survival such as AKT (Oh and Jacinto 2011, Laplante and Sabatini 2012). Enhanced by extracellular growth factors and nutrients, mTORC1 has a hand in indirectly controlling the activity of many of the essential targets of protein synthesis, including eIF4A helicase, the cap-binding proteins eIF4E and eIF4G. In the case of eIF4A, mTORC1 activates S6 kinase leading to phosphorylation and proteasome degradation of the eIF4A binding protein PDCD4, and increased activity of the eIF4A helicase (Dorrello et al. 2006). Enhanced eIF4E activity is achieved in a similar manner; phosphorylation of eIF4E binding-proteins (4EBPs) by mTORC1 reduces their binding affinity for eIF4E and enhances cap-dependent translation. Although these alterations occur in a cell-wide manner, the effects of mTORC1 inhibition, either by pharmacologic means or nutrient starvation, are observed in only a small subset of genes (Avni et al. 1997). These genes include ribosomal proteins, metabolism-related proteins and translation-related proteins (Thoreen et al. 2012). Surprisingly, the set of genes experimentally identified to have reduced levels of translation due to mTOR inhibition share a common 5' motif. The Terminal Oligo-Pyrimidine (TOP) motif, a

stretch of 4-14 pyrimidines at the start of an mRNA, serves as a CIS-acting factor which allows for these genes to be translationally controlled in a mTOR-dependent manner. Although the list of TOP-motif containing genes is small, their regulation has an outsize effect on cellular energy utilization due to the inclusion of genes involved in ribosome biogenesis and translation.

Beyond the mTOR pathway, a vast network of kinases and phosphatases affects every stage of protein synthesis, although the downstream CIS-acting factors in many cases have not yet been identified (Figure 1.2). A large number of these pathways act on the previously described ternary complex molecule eIF2. eIF2 activity is altered via phosphorylation state during nutrient deprivation, viral infection, iron deficiency, and endoplasmic reticulum stress by the respective protein kinases GCN2, PKR, HRI and PERK. Each of these protein kinases operates independently, and is controlled by a separate regulatory network, making eIF2 a central node for the integration of information regarding these cellular stressors. A key example of the adaptive function of eIF2 is its ability to induce preferential translation of ATF4, a master-regulator of stress-response, during periods of relative global translational repression. (B'chir et al. 2013). eIF2 is not alone in serving as an essential integrator of stress-response signals. eIF4, a molecule highly regulated by mTORC1, is also a target of the protein kinases MNK1 and MNK2. Receiving upstream signals from the MAPK-ERK pathway and the RAS-RAF pathway, MNK1/2 is able to phosphorylate Ser209 on eIF4E, increasing the selective translation of survival-related genes (Wendel et al. 2007).

Translational control can also be affected at the level of transcription. Activity of the transcription factor MYC, normally inhibited by p53 which is in turn inhibited by MDM2, leads to increased mRNA levels of eIF4E, eIF4A, and eIF4G (Pelletier et al. 2015). In turn, both MYC and MDM2 appear to be eIF4E-sensitive mRNAs, indicating a positive-feedforward loop exists

between these regulatory elements (Lin et al. 2008). Due to MYC and MDMs association with oncogenesis and heightened translation, these targets and effectors are crucial to translational control and illustrate the highly complex, specific and interconnected nature of translational regulation.

As evidenced by the various mechanisms described in the previous section, rapid translational control targeting translation initiation and elongation is most commonly achieved via protein kinases and phosphatases. The rapid rate of translational reprogramming during states of cellular stress is in part due to the fact that downstream macromolecules do not require degradation to suppress their effects. In lieu of proteolytic degradation, many of the molecules involved in translational control exhibit altered levels of activity in alternative phosphorylation states. The kinases and phosphatases which exert their effects on the ribosomal proteins involved in translation are downstream effectors of a number of regulatory networks including mTOR, MAPK-ERK, and RAS-RAF. Serving as a point of convergence for these pathways, translation integrates a vast amount of information from a wide-range of intracellular and extracellular conditions. Dysregulation of the networks controlling translation is a common hallmark of pathologies involving aberrant cellular proliferation, growth and protein expression. However, translation's central role, downstream action, and multiple interacting proteins also make it a promising therapeutic target in these diseases.

1.2 The Ribosome and Pathology

The Ribosome in Human Disease

Due to translation's outsize use of cellular resources and role in determining cellular function, it is no surprise that it serves as a point of convergence for several regulatory pathways.

The ability to quickly integrate a vast amount of information from intracellular and extracellular conditions into an altered program of protein expression makes translation a key effector of stress response. However, the importance of protein expression, coupled with the complex nature of the regulatory networks controlling it, and the action of translation itself, make translation into an opportune target for factors aiming to disrupt cellular function and the source of a large number of pathologies.

Proper functioning of ribosomal translation is critical to cellular function; exploring the role of translation as both a cause of, and in response to, pathologies serves to highlight this fact. It took only 90 years from when Friedrich Loeffler first described Klebs-Löffler bacillus in 1884 to advance our understanding of translation to a point where the protein responsible for causing Diptheria and its molecular mechanism could be understood (Loeffler, F. 1884. Mitt. Gesundheitamte 2:421-99). Fragment A, of the Diptheria-causing A-B toxin, acts via ADP-ribosylation of eEF2, rapidly depleting intracellular stores of functional eEF2 and inhibiting cellular protein synthesis (Jorgensen et al. 2005). This minor alteration of a single ribosomal protein destroys the ribosome's ability to catalyze ribosomal elongation, halting cellular protein synthesis, rapidly resulting in cell death. Due to A-B toxins predilection for the upper respiratory tract, this effect is thankfully not wide-spread, however several other organisms have independently developed toxins that act in the same fashion, such as those responsible for Whooping Cough, Cholera, and Pneumonia.

While the ribosome is an effective target for exogenous toxins, it also serves as a key component of the innate immune system's response to viral infection. Relying on the host's ribosomal machinery in order to engage in replication and maintain an infection, viruses are highly dependent on active translation. As mentioned previously, the innate immune system has

the capability to signal through PKR, leading to decreased phosphorylation of EF2 and reduced ternary complex formation. Although this response is variable based on the type and severity of infection, reactive modifications to EF2, along with several other intracellular responses, greatly abrogates ribosomal function in as little as 30 minutes (de Nadal et al. 2011). This rapid and efficient reduction of ribosome function, once again mediated by a single key factor, is critical to inhibiting the propagation of viral infections.

Beyond exogenous and endogenous perturbations, the ribosomal machinery is also susceptible to germline mutations. Although homozygous perturbations to ribosomal proteins commonly results in perinatal lethality, haploinsufficiency results in a number of diseases, summarized in (Bhat et al. 2015, Farley and Baserga 2016) (Table 1.1). Defined as ribosomopathies, these diseases display a wide-range of symptoms but one commonality is clear; aberrations in the expression of ribosomal proteins result in a heightened risk of cancer. While in many cases mutation of ribosomal proteins does not directly induce oncogenesis, attempts to rescue normal expression of ribosomal proteins by upstream pathways controlling translational result in a landscape ripe for malignant transformation. Although germline genetic mutations involving core ribosomal machinery are extremely rare in the general population, alterations in the expression of these same proteins and the upstream mediators of their function are among the most common hallmarks of transformed cells.

First recognized in 1896 when hypertrophic nuclei were observed in malignant cells, dysregulation of translation has a long association with cancer (Pianese 1896). While increased growth and cellular division are common among cancers, the means by which the translational landscape is altered to produce these phenotypes is highly variable but commonly converge on translation. Broadly, inciting events directly related to translation that lead to enhanced anabolic

activity can be separated into three main groups. The first, constitutively activating cell-surface tyrosine kinases normally responsive to growth factors, such as PDGF, can addict cells to increased protein synthesis. Activated growth factor receptors, such as EGFR or PDGFR are commonly associated with several tumor types, such as Wilm's tumor of the kidney or glial tumors of the brain (Ghanem et al. 2010).

Bypassing surface receptors entirely, activity of growth-signaling pathways can additionally be enhanced via removal of downstream tumors suppressors such as the TSC1 and TSC2 proteins. Deletion of either TSC1 or TSC2 hyper-activates mTORC1, decreasing eIF4E binding by 4EBP and leads to development of Tuberous Sclerosis with 100% penetrance. Individuals afflicted with Tuberous Sclerosis experience hundreds of benign tumors in their lifetimes, including harmatomas, angiofibromas, and harmaoblastomas of the heart, eyes, kidney, nervous system and skin (van Slegtenhorst et al. 1997). Tuberous Sclerosis illustrates the extremely deleterious effects of dysregulating a single gene involved in the pathways controlling translational initiation. Mutations of RAS, PTEN and PI3k similarly result in enhanced ribosomal initiation and are associated with their own clinical symptoms and cancers (Mulholland et al. 2012).

Finally, altered expression or activation of many of the key proteins involved in ribosomal initiation can directly promote malignant transformation. Overexpression of eIF4E, for example, has been shown to induce transformation in a variety of primary human cells, and subsequent reduction of eIF4E expression can lead to a reversal of transformation (Lazaris-Karatzas et al. 1990, Ruggero et al. 2004, Larsson et al. 2007). Clinically, eIF4E has been found to be increased 3 to 10 fold in cancers of the blood, brain, lung, prostate, head and neck, bladder, colon and breast and is generally associated with decreased survival (Meric-Bernstam 2008).

The clinical significance of increased activity of core ribosomal proteins is not restricted to eIF4E, but is also observed with eEF2, pEBP1 and pS6, leading to their use as markers of overall survival, progression, and recurrence in multiple forms of cancer. While phosphorylation of some proteins, like pS6, do not seem to directly affect translation, their prognostic value is due to their correlation with increased mTOR network activity and overall cellular translational output.

One particularly interesting neural ribosomopathy is leukoencephalophathy with vanishing white matter syndrome. The mutation or loss of any one of 5 genes encoding the translation initiation factor EIF-2B can lead to profound neurodegeneration resulting in cerebellar ataxia, spasticity, optic atrophy, epilepsy, and eventual death. Although poorly understood, these gross neurologic consequences are due to the improper functioning of a single initiation factor leading to loss of white matter and the appearance of foamy oligodendrocytes with irregular mitochondria and high rates of apoptosis.

Development of Therapeutics Targeting Translation

Existing as a mediator of growth factor signaling and controlling the phosphorylation states of many key ribosomal proteins, the mTOR pathway has become an extremely promising target for anti-cancer therapies. Altering translational control through the actions of protein kinases and phosphatases, the mTOR pathway has a number of targetable candidates. Discovered in 1975 and originally intended as an antifungal agent, Rapamycin led to the discovery of mTOR, originally dubbed "Mammalian Target Of Rapamycin" (Sehgal et al. 1975). As soon as 1981, prior to the discovery of Rapamycin's cellular target, it was established to have anti-cancer activity in cell-line screens (Douros and Suffness 1981). By 2007 however, only a single mTOR inhibitor (Temsirolimus— a water soluble form of rapamycin) had been approved by the FDA as

an anti-cancer agent, while many another therapeutics targeting mTOR had failed to achieve significant results in clinical trials. The failure of these first-generation mTOR inhibitors has been attributed to several reasons. Firstly, the biologic kinetics of rapamycin and early mTOR inhibitors was extremely poor; its ability to inhibit mTOR function is only partial due to its allosteric mechanism of action and mTORC1 specificity (Brachmann et al. 2009). Furthermore, the concentrations required for clinically significant anti-mTOR activity is associated with a poor side-effect profile, including immunosuppression. Finally, and arguably most concerning, due to mTOR's central role in mediating anabolic signaling and complex interaction with multiple convergent pathways, mTOR inhibition could be rescued by upregulating the activity of parallel or upstream pathways, such as AKT or PI3K, or by reducing the function of endogenous mTOR inhibitors such as the TSC complex.

Although results from first generation mTOR inhibitors have been disappointing, they have not lessened interest in leveraging mTOR or translation initiation as therapeutic anti-cancer targets. At current, there are over 200 ongoing clinical trials involving the use of second generation mTOR inhibitors. Whereas the original mTOR inhibitors attenuated mTOR function by allosterically inhibiting mTORC1, second generation compounds target the active site of both mTORC1 and mTORC2—fully inhibiting mTOR function. Furthermore, to reduce resistance via increased AKT or PI3K activity, drugs have been developed which act as dual PI3K-mTOR and AKT-mTOR inhibitors. These drugs are currently in clinic trials and are showing promising results (Bhat et al. 2015). Of the most exciting second-generation inhibitors to be developed are those that target translational initiation directly. These second-generation drugs achieve mTOR inhibition in a variety of ways, including inhibiting the formation of the ternary complex, halting the helicase activity of eIF4A and reducing the interaction between eIF4E and eIF4G (mimicking

cellular methods of translational control). While these drugs have been surprisingly effective in cell-line and animal studies, it remains to be seen how they will fare in clinical trials. In any event, their existence alone heralds a remarkable understanding of the complex network controlling translation, as well as translation's central role in a host of pathologies.

1.3 A History of Translation's Study

Early Insights Propel a Nascent Field

Major advances in our understanding of how translational control is encoded and achieved by cells have regularly followed the development of new technologies which allow for an improved or wholly unique set of observations to be generated. These watershed technologic innovations have in some cases been engineered specifically for the study of translation, while others were cleverly adapted. Regardless, innovations such as the identification of the ribosomal start sequence, the first application of micro-arrays to generate a genome-wide translation measurement, and the development of ribosome profiling have been followed by an explosion in literature advancing our understanding of translational control. A clearer understanding of the technologies which have evolved into those currently used to study translation highlight the importance and necessity of further technology developments required to improve upon those currently available and to promote a deeper understanding of translational control's physiologic function and role in pathologic states.

While the ribosome had been first observed in the mid-1950s by George Emil Parade (winning the Nobel Prize in 1974), the means by which ribosomal initiation occurred remained unclear (Sabatini et al. 1966). Based on observations by Mituru Takanami in 1965 that regions of bacteriophage RNA were protected from enzymatic digestion due to association with the

sterically protective ribosome, Joan Steitz developed a process by which she was able to isolate fragments of RNA which were actively undergoing translation (Takanami et al. 1965). By first digesting ribosome associated-RNA with pancreatic ribonuclease, and then subjecting her sample to centrifugation in a sucrose gradient, Dr. Steitz was able to isolate portions of RNA which were spared from RNAse degradation (Steitz 1969). Combining this with "cutting-edge" cellulose acetate and DEAE chromatography developed by the Sanger group, a set of sequences could be established for the location of polypeptide chain initiation within bacteriophages (Sanger et al. 1965). While knowledge of the AUG initiation codon had been known for several years, the ribosome's ability to initiate protein synthesis internally, and simultaneously, was not. Furthermore, the combination of ribonuclease digestion, sucrose gradation and application of "state-of-the-art" sequencing, is one that has played out multiple times in the course of developing technologies to interrogate translation (Figure 1.3).

The invention of dideoxy chain-termination sequencing, conventionally known as Sanger Sequencing, in 1977 not only lead to Sanger's second Nobel Prize in Chemistry, but also set the stage for refinement of sequencing capabilities and sequencing of the human genome (Sanger et al. 1977). While early RNA sequencing methodologies had been established by Robert Holley and Sanger in the 1960's and 1970's, the human genome project served as a major step forward in the application of genomics and highlighted the importance of genes in health and disease, as well as their highly interconnected nature (Lander et al. 2001). Research groups were spurred to begin decoding the role of transcription in human physiology and disease. The independent discoveries of reverse transcriptase by Howard Termin and David Baltimore in 1970 allowed for the use of sequencing technologies developed for DNA to be utilized with RNA samples, greatly aiding RNA sequencing efforts (Temin and Mizutani 1970).

Genome-Wide Measurements

In order to apply translation-wide measurements on a genome wide scale, developments were first required which would allow for a complete transcriptome to be characterized. The combination of reverse transcription with computerized image processing and hybridization arrays by Floyd Taub attempted to fill this void in 1983 (Taub et al. 1983). Dubbed a DNA-Microarray, this paradigm was able to rapidly generate a measurement of mRNA transcript levels by relying on digital image analysis. While this technique was limited by the requirement for pre-existing capture probes, it enabled researchers to ask questions regarding the means by which translation is controlled on a genome-wide scale. In 1999 David Morris led a group which explored translation's role in overall protein expression. Using a sucrose gradient method similar to Joan Steitz, fibroblast cells were mitogenically stimulated and their ribosome-bound mRNAs were separated into lowly and highly translated groups based on their level of sedimentation in the sucrose gradient. Application of reverse transcription allowed for the use of a 1,200 gene microarray and the discovery that mitogenic stimulation altered the translational expression of only a small fraction (1%) of genes—one of the first instances of translation's tightly controlled nature (Zong et al. 1999).

This method of fractionation, followed by microarray analysis, was repeated to explore the role of translation in several contexts; such as in T-Cell differentiation, glucose starvation, and fMR1 knockout (Mikulits et al. 2000; Brown et al. 2001; Kuhn et al. 2001). Of particular note were experiments performed by Yoav Arava and Patrick Brown, who utilized the previously described paradigm but took great efforts to separate their sucrose gradients into as many as 14 fractions. This work led to insights regarding ribosome density and ORF size, and provided

additional evidence that initiation was the rate limiting step for protein production (Arava et al. 2003). Later experiments by the same group saw the creation of the first ribosome density maps for several ribosome bound mRNAs (Arava et al. 2005). Combining their established fractionation method with oligonucleotide probes complementary to small regions of a specific mRNAs, Arava obtained a clear picture of the location of translating ribosomes on mRNA. This work cemented ribosomal initiation as the rate-limiting step of translation, demonstrated the highly processive nature of translation and began to associate sequence-specific elements of mRNAs to their level of translation control.

Microarrays have also been leveraged in Ribosome Affinity Purification (RAP) experiments which utilize genetically modified ribosomal proteins as targets for immunoprecipitation (Zanetti et al. 2005). Introduction of immunoprecipitable targets to core ribosomal proteins such as EGFP to RPL10a, as in Translating Ribosome Affinity Purification (TRAP), allows for the effective selection of mRNAs undergoing translation from a cellular homogenate (Heiman et al. 2008). A powerful application of this technology is the performance of cell-type specific translational measurements. Cell-type specific promoters induce expression of tagged non-native ribosomal proteins which allows for the purification of translating mRNAs from specific cells of interest in transgenic animals and cell-lines. These immunoprecipitated mRNAs can then be applied to a microarray in order to determine the relative translation of transcripts within specific cell types. However due to the lossy nature of immunoprecipitation, these techniques require large amounts of starting material to be used. Furthermore, due to the genetic construct used in many forms of RAP, including TRAP, the native ribosomal protein continues to be expressed alongside the modified protein, complicating quantification of translation.

Next Generation Sequencing

Microarray analysis of polysome fractionation yielded a number of insights regarding translational control, but suffered from a number of limitations inherent to the technology. While there were some quantification issues arising from an inability to resolve heavy polysome fractions and determine the number of ribosomes bound to mRNA in them, the key drawback of polysome fractionation as a technique was an inability to determine ribosomal position. Knowledge of the translating or scanning ribosomes position is essential to determine if a ribosome is located in the protein coding region, stalled, or contained within an upstream open reading frame (uORF). In 2009, building on the original experimental paradigm designed by Joan Steitz in 1969, Nick Ingolia, in the lab of Jonathan Weissman, developed a library preparation scheme which allowed for the sequencing of ribosome protected footprints and made genome-wide analysis of translation with positional information possible (Ingolia et al. 2009). Dubbed ribosome-profiling, the library preparation scheme began following sucrose-fraction isolation of mRNA fragments protected from RNAse digestion by their active association with ribosomes. Incorporation of adapters necessary for PCR and Illumina sequencing to take place was difficult due to the short nature of these fragments and their lack of a conserved region, such as the Poly-A tail which exists on undigested mRNAs. In the original paper unveiling the ribosome profiling technique, this difficulty was overcome by use of a poly(A) polymerase which generated a 3' tail of adenosine nucleotides, mimicking the function of the poly-A tail in library construction. However, this method was abandoned following its initial demonstration and inter-molecular ligation was used in its place to incorporate the 3' adapter in future ribosome profiling experiments (Ingolia et al. 2012). In both library preparation schemes, reverse

transcription, circularization and a moderately high number of cycles of PCR were then required to generate sequencable libraries. Ribosome profiling allowed for several important advances in comparison to the experiments performed by Yoav Arava. Whereas Arava was able to determine the number of ribosomes associated with mRNA, but not their position, ribosome profiling allows for the genome-wide identification of ribosomal location with nucleotide resolution based on sequence information derived from the ribosome footprint. This positional information can be used to make important observations regarding ribosomal occupancy in the 5' and 3' untranslated regions, allowing discrimination of ribosomes translating protein-coding regions versus upstream open-reading frames and non-canonical start-sites. Furthermore, the effects of mRNA composition and the frequency of translational pausing can additionally be studied with the application of ribosome profiling. Quantitatively, ribosome profiling also bypasses inaccuracies associated with the analysis of heavy polysome fractions (due resolution limitations). The development of a tool able to interrogate translation on a genome-wide scale, with single nucleotide resolution, has allowed for a resurgence of research into translation. Spanning fifty years of research, the developments made in technology that have allowed our understanding of the complex elements by which translation is controlled is staggering. However, several major limitations exist which obfuscate a clear picture of translational control. As will be discussed, new technologies will need to be developed and refined in order to have a more accurate understanding of the elements involved in the normal physiology and pathology of translation.

1.4 Necessity for Improved Tools

New Technologies Required For Further Understanding of Translation

Utilizing a myriad of tools, the importance of translational control to normal physiology and pathologic states has been well established. However, many of the technologies currently utilized in the study of translation suffer from drawbacks which limit their utility, especially when used to explore questions from a systems-biology approach. Due to the high-level of heterogeneity between cell-types, cell-type specific translational measurements are critical in accurately determining cellular translational landscapes. Currently established RAP technologies such as BAC-TRAP and the RiboTag system have been successfully used to perform cell-type specific polysome fractionation with micro-array quantification. However, these techniques do not provide ribosome positional information, or yield a direct quantification of ribosome density. As an additional concern, in the case of BAC-TRAP specifically, the native ribosomal protein continues to be expressed in conjunction with the epitope-modified version, complicating translational analysis. Finally, while these systems excel in labeling ribosomes from cell types of interest that have well characterized cell-type-specific marker genes (which can be used to drive modified ribosome expression), they are difficult to apply to cellular populations which may not have globally unique gene expression.

Ribosome profiling also suffers from several drawbacks. The original ribosome profiling studies were performed using large amounts of material derived from yeast. While this source of material is effectively unlimited, experiments designed to uncover translation's role in diseases which rely on patient's clinical samples or animal-derived samples do not have the same flexibility. Additionally, due to the number of independent enzymatic reactions, gel-purifications and precipitations required, ribosome profiling as described by Ignolia requires about two-weeks of hands on time to perform, several hundred dollars per sample in reagents, and at least 50

nanograms of input material. The development of a ribosome profiling technique which lowered input requirements as well as reduced the time, effort and costs associated with ribosome profiling would greatly aid investigators interested in studying translation.

Finally, the study of translation has been hampered by an inability to perform measurements on a large-scale. The development of a technique amenable to high-throughput measurements would allow for previously impossible systems biology approaches to understanding translation. For example, although it is known that kinases play a major role in cell signaling and translational regulation, and respond to a wide-variety of intra and extracellular perturbations, there has yet to be a systematic effort to characterize their translational effects. This is, in part, due to the staggering effort which would be required for such a project—using conventional ribosome profiling techniques, a study of the 500 human protein kinases in several cellular states would require tens-of-thousands of man-hours. A high-throughput translational measurement technique however would make it possible to determine the effects of each protein kinase in multiple physiologic states for a fraction of the time and cost. Due to their central role in a variety of diseases, such as cancer, as well as their inherent druggability, such a study could yield important clinical insights. An addition application of high-throughput translational screening could be in drug development. The ability to perform screens of compounds targeting translation and rapidly quantitate their effects could be a boon for the development of new therapeutics. Finally, low-cost, high-throughput translational measurements could also aid in the clinical monitoring of patients undergoing chemotherapy targeting translation and serve as an early warning monitor for the occurrence of resistance.

It is the role of this dissertation to describe several works which attempt to address the current limitations of technologies interrogating protein synthesis and apply them to develop new

insights into translational control. In Chapter 2 I discuss efforts to translationally profile tumor-specific cells which required the generation of an animal model which expressed modified ribosomal protein only in transformed cells and the first application of ribosome profiling in tissue. Chapter 3 attempts to address some of the downsides of conventional ribosome profiling through development of a ligation-free approach which utilizes poly-adenylation and template switching to decrease input requirements, cost, and time associated with generating ribosome profiling data. Finally, Chapter 4 focuses on the development of an experimental procedure relying on ribosomal immunoprecipitation and pooled library generation designed to generate translational data in a high-throughput, rapid, and inexpensive manner. I hope to demonstrate that these technologies, as well as the insights we've gained from their application, have helped to advance our understanding of translation and our future ability to interrogate it.

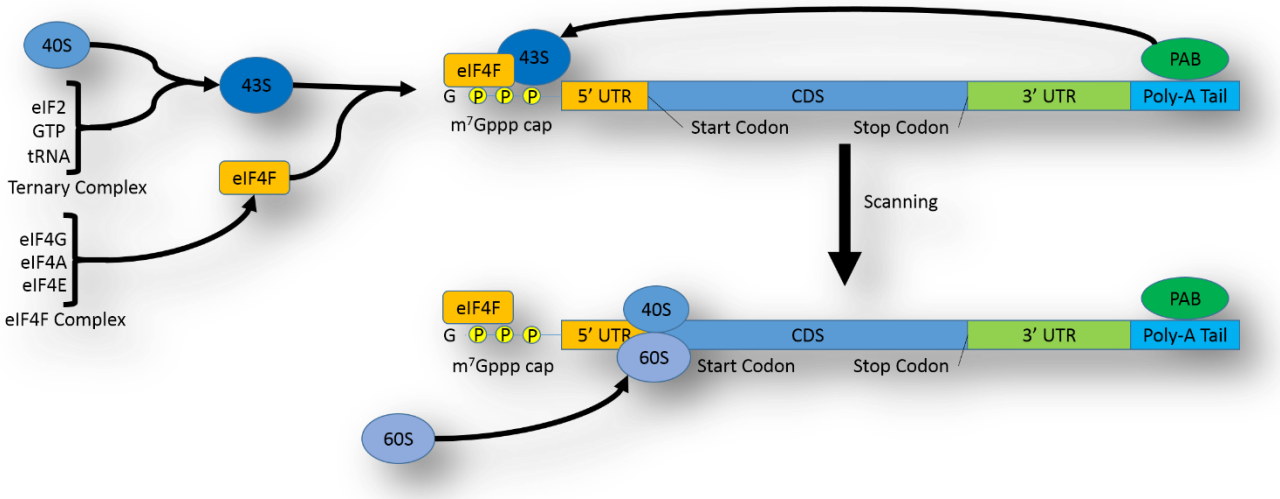

Figure 1.1 Eukaryotic Translation Initiation

Initiation of translation may occur following preassembly of the ternary and eIF4F complexes. The ternary complex, in conjunction with the 40S ribosomal subunit, form the 43S subunit. Working in concert with eIF4F the 43S subunit begins the process of translational scanning in the 5'-untranslated region (UTR). Once a AUG start codon is located, the 60S subunit binds and begins translation of the protein coding region (CDS).

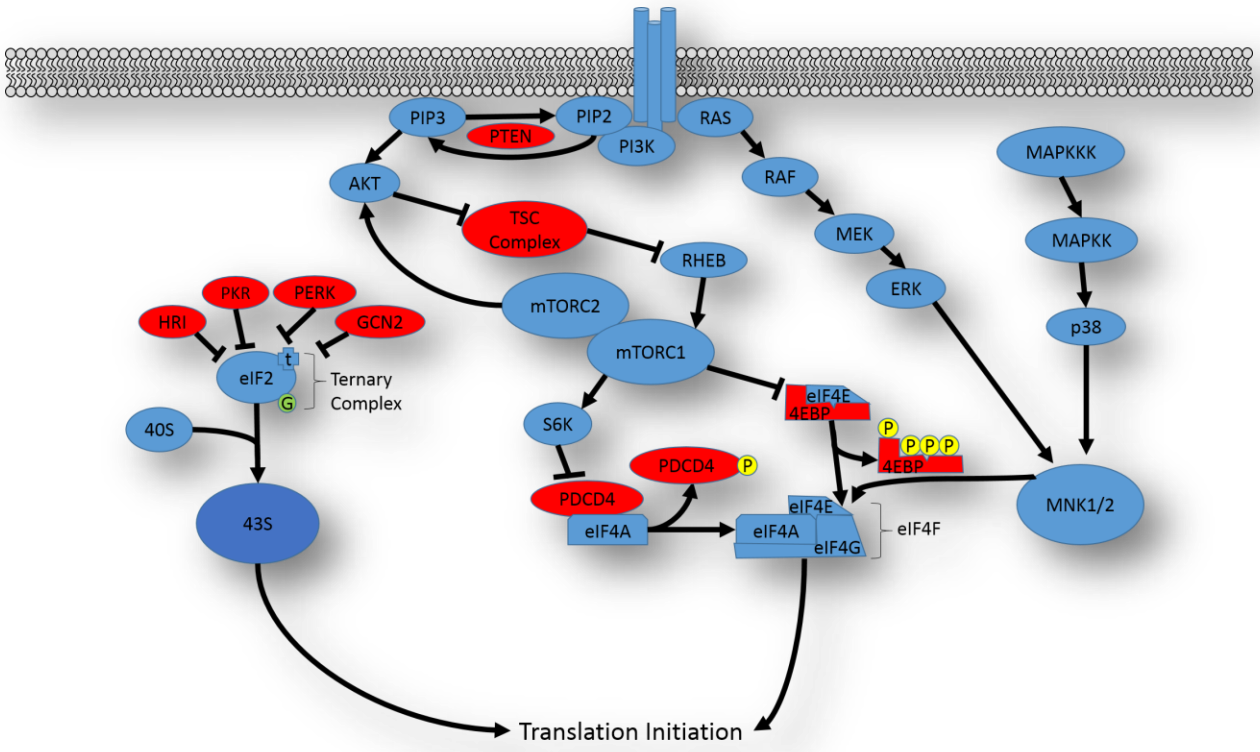

Figure 1.2 Mechanisms Affecting Translation

Multiple pathways converge on translation and affect translation initiation. RAS/RAF, PI3K, mTOR and MNK can affect translation in response to intracellular and extracellular states. Red-colored proteins represent inhibitors of translation initiation.

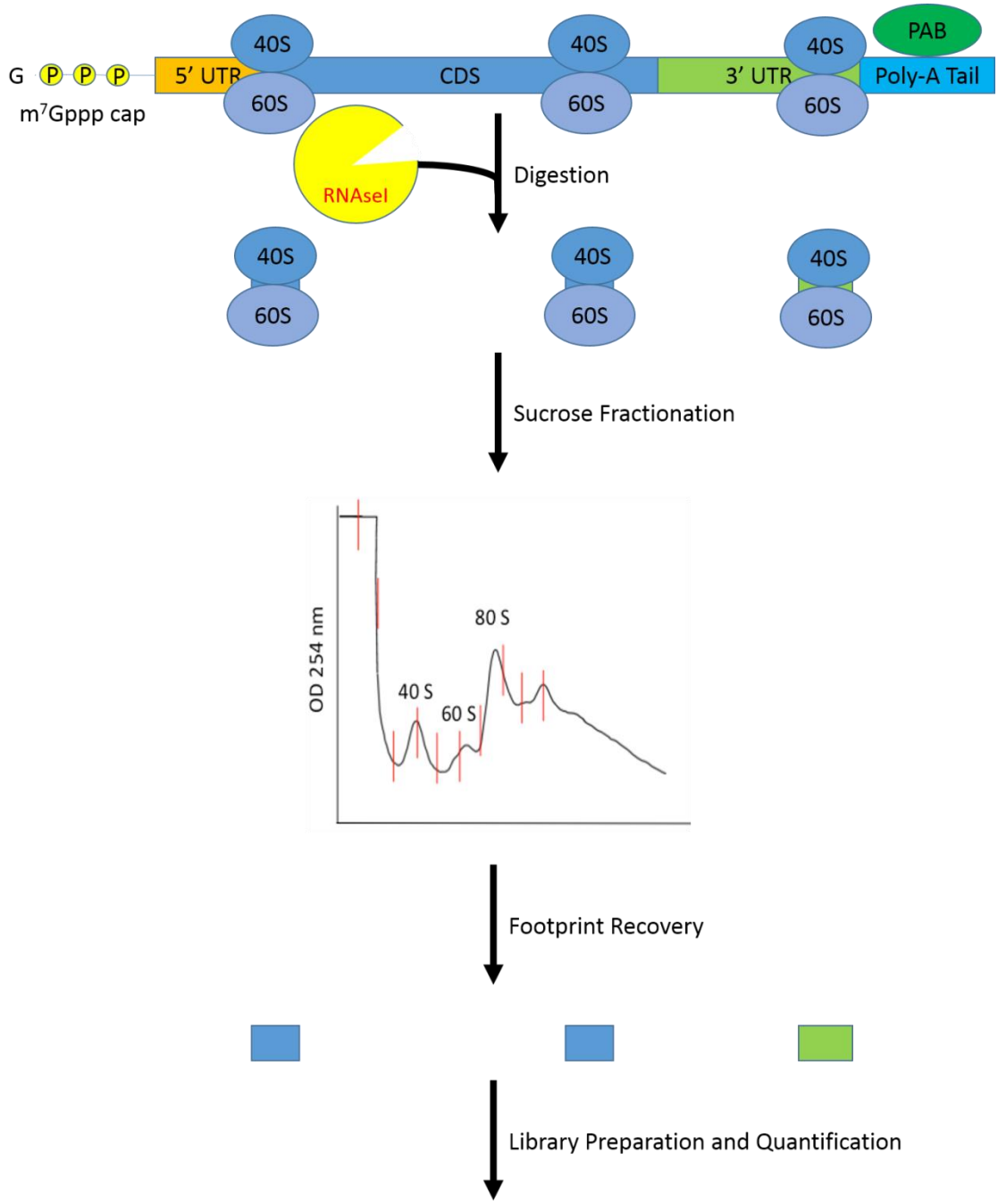

Figure 1.3 Footprint Isolation Experiments

Originally utilized in the 1960's this experimental paradigm has been leveraged successfully for decades in order to yield further insights into mechanisms of translational control. Due to the protective effect of ribosomal binding, fragments of mRNA being actively scanned or translated

by the ribosome can be recovered following digestion of non-bound mRNA with an RNAse. Sucrose fractionation of digested material allows for the recovery of footprints. Sucrose fractionation can also be performed on un-digested material to yield information regarding the number of ribosomes bound to groups of mRNAs. Following fractionation, footprints can be prepared and quantified. Multiple techniques have been used to quantify ribosomal footprints and polysome fractions including sanger sequencing, micro-arrays, and next-generation sequencing.

Table 1.1 Ribasomopathies and Associated Cancers

Disease	**Genetic Defects**	**Associated Cancers**	**Clinical Features**
Diamond-Blackfan Anemia	RPL5, RPL11, RPL36, RPS19, RPS24, RPS17, RPL35A, RPS7, RPS15, RPS27A	Osteosarcoma MDS	Macrocytic Anemia Craniofacial defects Growth retardation Thumb abnormalities
X-linked Dyskeratosis	DKC1	AML Head/Neck Tumors	Oral leukoplakia Cytopenia Hyperpigmentation Nail dystrophy
5q-Syndrome	RPS14	AML	Macrocytic anemia
Schwachman-Diamond Syndrome	SBDS	AML MDS	Neutropenia Pancreatic insufficiency Growth retardation
Cartilage-Hair Hypoplasia	RMRP	Non-Hodgkin Lymphoma Basal Cell Carcinoma	Hypoplastic anemia Growth retardation Hypoplastic hair
Bownen-Conradi Syndrome	EMG1	N/A	Fatal in infancy Growth retardation

Chapter 2

Ribosome Profiling Reveals Cell Type-Specific Translation in Brain Tumors

2.1 Background

Abstract

Transformation of cellular populations is driven by a variety of molecular factors. In cancers generally, and glioma specifically, many of the inciting mutations ultimately affect regulation of protein synthesis. Assessing the translational landscape of gliomas presents a unique scientific challenge as gliomas are heterogeneously composed of multiple cell types recruited by transformed cells into a solid tumor which generates a specialized microenvironment within the brain. In this chapter, I discuss studies undertaken in order to better understand translational regulation in the transformed cells of gliomas. In order to achieve this goal, we developed a tumor-specific ribosome tagging system in the form of a RiboTag glioma mouse model which enabled cell type-specific, genome-wide ribosome profiling of tumors. Furthermore, these studies enabled the application of ribosome profiling to tissue samples. Translational profiling of transformed cells in glioma has been a difficult task due to a lack of a glioma-specific gene which is not naturally expressed in the brain. To bypass this, we developed the RiboTag mouse model which was generated by crossing a mouse containing floxed Trp53 alleles with a mouse designed to express HA-tagged Rpl22 following the action of Cre-recombinase. Transformation of glial progenitor cells and cell-type specific activation of RiboTag expression is achieved with a PDGF-B expressing Cre-recombinant retrovirus which

simultaneously removes native exons of *Trp53* and Rpl22 while inducing HA tagged Rpl22 expression. Surprisingly, while we found that tumor-specific genes have high levels of translation, their overall translation efficiency is low compared to normal brain. Furthermore, we found that the presence of upstream annotated start codons in the 5' leader sequence of genes with 5' ribosomal density was associated with altered ribosomal density in the CDS of genes, leading to differential translation in glioma compared to normal brain.

Introduction

Gliomas are devastating neural malignancies that cause significant morbidity and mortality. Although there will be 20,000 predicted cases of glioma in the United States this year alone, clinical outcomes for the disease have not changed significantly in the past 30 years. Recent efforts by The Cancer Genome Atlas (TCGA) have utilized gene expression profiles of glioblastomas (GBM) to identify four subtypes of gliomas—Proenueral, Neural, Classical and Mesenchymal (Verhaak et al. 2010). Many of the hallmarks of gliomas reside in mutations and deletions of molecules which converge on translation—these include effectors of mTOR and AKT as well as additional mitogenic pathways (Fan et al. 2007; Helmy et al. 2012; Jiang and Liu, 2009; Parsa and Holland, 2004; Rajasekhar et al. 2003; Takeuchi et al. 2005). However, although mutations in these pathways are identified in a high percentage of human tumors, very little is known regarding translational regulation in gliomas.

Existing as both a clinical and scientific challenge, gliomas are diffusively infiltrative into surrounding brain tissue. While techniques exist which allow for physical separation of cells, the dynamic and rapidly responsive process of translation could be perturbed by dissociation. Furthermore, dissociation of transformed cells from cellular populations recruited to

the tumor would be a challenging endeavor. Efforts have been made to characterize tumor-specific translation, but the approaches employed suffer from a number of drawbacks (Doyle et al. 2008; Heiman et al. 2008; Sanz et al. 2009). Attempts to quantify cell-type specific translation have used intact, ribosome-bound RNA to measure translational output. As a result, these studies have lacked important information regarding ribosomal positioning, making it impossible to determine whether increased ribosomal density is relegated to the coding region without additional experiments. Furthermore, the quantification of translation in these studies has been complicated by the presence of the native version of the tagged "cell-specific" ribosomal protein. However, the most concerning feature of previous efforts to quantify translation in glioma is the lack of a glioma-specific marker. While genes like OLIG2 have been previously used to drive expression of modified ribosomal proteins used for the molecular identification of ribosomes derived from cells-of-interest, OLIG2 is normally expressed in glial progenitor cells and mature olgiodendrocytes, obfuscating the translational profile of the intended cellular population. In order to accurately quantify translation in the transformed cells of glioma, a new method which enables the identification and isolation of ribosomes derived from these specific cellular populations is necessary.

Representing a unique opportunity to study a clinical disease progenitors, a recently developed animal model recapitulates many of the genetic and morphologic features seen in Proneural GBMs (Lei et al. 2011; Liu et al. 2011). We have utilized this mouse-model in conjunction with a recently emerged experimental paradigm which allows for molecular dissection of translating ribosomes from heterogeneous tissue through immunoprecipitation of epitope-tagged ribosomes expressed in a cell-type specific manner (Heiman et al. 2008; Sanz et al. 2009). Here, we retrovirally deliver Cre recombinase and PDGF-B to adult mouse brains

which genetically harbor floxed *Rpl22* and *Trp53* alleles. Via the action of cre-recombinase, the fourth exon of *Rpl22* and the seventh exon of *Trp53* are removed, inducing the expression of HA-tagged *Rpl22* and destroying *Trp53* activity. Theoretically, this restricts expression of HA-tagged *Rpl22* to transformed cells descendent from the originally infected cells and creates a transformation-specific epitope to isolate ribosome bound RNA from homogenized tissue. In order to address these concerns related to the accurate quantification of translation, we have applied ribosome profiling, which allows for a genome-wide analysis of protein synthesis and ribosome positioning, to a mouse model which expresses only the non-native version of the tagged protein in cells of interest (Ignolia et al. 2009). Ribosome profiling relies on deep sequencing of ribosome-protected mRNA footprints and has been previously used to investigate a wide variety of scientific questions, including non-canonical translation yeast, and metastasis and invasion in cancer cell lines (Brar et al. 2012; Hsieh et al. 2012).

This chapter describes a strategy for cell type-specific translation measurements through the combination of the tissue specific RiboTag systems with ribosome profiling. We utilized these tools to gain genome-wide ribosomal positioning and translation information, identify genes selectively translated by transformed cells, and demonstrate the role of non-CDS sequences in controlling the translation of genes.

2.2 Results

RiboTag Glioma Model Enables Tumor-Specific Ribosome Profiling

Translational profiling of transformed cellular populations in glioma is a challenging endeavor. Glioma's are known to be diffusively infiltrative, and recruit a number of non-transformed cellular populations into a highly heterogeneous tumor. Although dissociation of

transformed cellular populations is theoretically possible, these manipulations may alter the rapidly responsive translational network. One approach to cell-type specific profiling is the use of molecularly modified ribosomes. However, the specificity of genetic modification relies on the existence of cell type-specific promoters to drive the expression of epitope-tagged ribosomal proteins. In the case of glioma, an exclusive promoter has not been identified, making this a poor approach to characterizing transformed cells in this disease. Previously, OLIG2 was used as a 'glioma specific' promoter with the bacTRAP system (Helmy et al. 2012). While this marker is highly expressed in transformed tissue, several studies have demonstrated that OLIG2 is also highly expressed in untransformed glial progenitors, mature oligodendrocytes and reactive cells recruited to the tumor (Assanah et al. 2006). Due to the lack of a tumor-specific promoter, we developed an alternative methodology which leverages retroviral delivery of Cre recombinase to activate RiboTag expression only in cells exposed to transforming conditions—loss of *Trp53* function and expression of PDGF-B. The ability to both induce transformation and ribosomal labelling in a specific cellular population is achieved by the dual action of Cre recombinase and viral expression of PDGF-B.

The RiboTag mouse is genetically modified such that exon 4 of a core ribosomal protein, *Rpl22,* is flanked by loxP sites followed by an additional copy of *Rpl22's* fourth exon with a HA-tag appended (Figure 2.1A)(Sanz et al. 2009). The RiboTag mouse was crossed with a mouse harboring loxP sites surrounding exon 7 of *Trp53* to generate a hybrid animal homozygous for both alleles (Chen et al. 2005). Cellular transformation is induced in infected cells by stereotactic injection of a replication-incompetent retrovirus expressing platelet-derived growth factor (PDGF-B) and Cre recombinase. Following viral injection, a small population of glial progenitor cells located in subcortical white matter is selectively infected, simultaneously deleting *Trp53*

and inducing expression of HA-tagged ribosomes. Descendants of the originally infected cells retain these genetic alterations and form fatal brain tumors within 30 days.

Similar to gliomas in human patients, the transformed cells in the RiboTag glioma model are highly infiltrative and intermingle with surrounding brain tissue surrounding the injection site. Although multiple cell types, including astrocytes, microglia, neuron, and recruited OLIG2-expressing progenitors exist within the tumor, they lack RiboTag expression. In order to demonstrate the compositional heterogeneity of our RiboTag glioma tumor model, in addition to the specificity of HA tagging, we performed immunofluorescence to stain for canonical marker proteins (Figure 2.1B). HA-staining was highly associated with tumor markers such as OLIG2 and PDGFRA while HA-staining was not observed to co-localize with canonical cellular markers such as GFAP, CD44, RBFOX3 and AIF1.

In order to determine the translational landscape of transformed cells, we first generated tumors in three mice. Brains were isolated, homogenized and material was divided into a portion destined for homogenate RNA-sequencing and the remainder used to generate ribosome profiling libraries. In order to generate ribosome profiling libraries, cellular lysates were digested with RNase I and monosomes were isolated via sucrose gradient fractionation. Following monosome isolation, half the material underwent immunoprecipitation to select for tumor-specific HA-tagged ribosomes (in order to generate a tumor-cell type specific ribosome footprint library), and the remainder was used to generate homogenate ribosome footprint library (Figure 3.1A). We utilized the ribosome's periodic 3-nucleotide codon periodicity in order to demonstrate libraries generated from immunoprecipitated material contain footprints. Calculation of the power spectrum of the 5'-end mapping position for each footprint relative to

the annotated start codon generated a power spectrum with a peak at 0.33/nucleotide, consistent with the anticipated three-base periodicity of ribosomal footprints (Figure 2.1D).

Specificity of the RiboTag model was assessed by calculating enrichment scores of the three tumor samples and examining a variety of tumor and neural lineage markers (Bedard et al. 2007; Cahoy et al. 2008; Lei et al. 2011; Verhaak et al. 2010). Enrichment scores were defined for each gene as the ratio of CDS counts from ribosome profiling libraries, divided by the CDS counts obtained from a sample's homogenate. In agreement with immunofluorescent observations, markers of transformed OPCs and Proneural glioma cells were enriched in libraries generated from immunoprecipitated footprints, while other neural cellular makers were depleted (Figure 2.1E).

Translational Pathways Dysregulated in Tumor Cells

Using ribosome density per gene, we sought to identify biological pathways which were altered in transformed cells. We utilized our three RiboTag immunoprecipitated ribosome profiling libraries and compared them to libraries generated from normal un-infected brains to generate a list of genes with statistically differential translation rates ($P < 0.05$) between the two groups. Differential translation of a gene between these two groups could occur due to compositional differences, transcriptional regulation changes or alterations at the level of translational regulation. This information was utilized by iPAGE, an information theory-based algorithm, in order to determine which gene ontologies were enriched within the tumor-associated libraries (Figure 2.2A)(Goodarzi et al. 2009). Our results, demonstrated in figure 2.2A, associates "Synapse" and "Cation Channel Activity" with increased translation in the normal brain samples, while "DNA Replication" and "Cell Division" ontologies were enriched

in the RiboTag enriched samples. The former ontologies may be related to the large number of neurons in normal cortex, while the latter highly translated ontologies would be anticipated in rapidly dividing transformed cells. Additionally, we identified "Structural Components of the Ribosome" to be highly translated in the RiboTag enriched samples, highlighting global upregulation of protein synthesis within the tumor.

We also sought to determine the translational status of cells in the tumor microenvironment which were not directly descendent from the inciting retro-virally infected cells. Similar to the previous comparison, we performed differential translation rate analysis between the RiboTag enriched ribosome profiling libraries and compared them to ribosome profiling libraries generated from the homogenate tumor-containing brains from which they were derived. From this list, we identified genes which were depleted by RiboTag immunoprecipitation, reasoning that these genes would be upregulated in cells which did not arise from the retrovirus-infected cells. There were roughly 100 genes which had higher translation rates in the tumor homogenate, but were expressed mainly in tumor-associated cells. These genes had a wide range of both positive and negative translation efficiencies (the ratio of ribosome footprints for a given gene divided by homogenate mRNA abundance for that same gene). Of those genes, 14 did not have significantly different levels of RNA abundance, highlighting alterations primarily at the level of translation. Furthermore, 13 of the 14 genes had positive-fold changes in translation efficiency.

Translation Efficiency Reduced in Transformed Cells

In order to explore the role of translational regulation in tumor-associated cells, we calculated mean translation efficiencies from all three mice for genes that were enriched or

depleted by RiboTag (Figure 2.3A). The distribution of translation efficiencies is broad, highlighting the large dynamic range of translational control and its importance in determining protein output—a finding reflected in other mammalian cell types (Ignolia et al. 2011). Comparing the distribution of enriched and depleted genes we were surprised to find that the median translation efficiency of depleted genes was ~25% higher than enriched genes and that there was a significant difference in their translation efficiency distributions ($p = 2x10^{-7}$, two-sample KS-test). In order to determine if this was a tumor-specific phenomenon, we generated ribosome profiling libraries from three normal brains. We found that the translation efficiency distributions of genes enriched or depleted in the RiboTag samples were not significantly different in normal brain (Figure 2.3B). Additionally, comparing the normal brain and tumor homogenate samples demonstrated that >88% of RiboTag-enriched genes had lower translation efficiencies in the tumor homogenate (Figure 2.3C). These results could be explained by global differences in the translational state of transformed cells.

Sequence-Dependent Regulation of 5'-Leader Ribosomal Density

Based on libraries generated from RiboTag immunoprecipitation, we found ribosomal density in the 5'-leader sequences of hundreds of genes. Although annotated as non-coding, ribosomal density in 5'-UTRs has been implicated in several forms of translational control and may have a role in transformed tumor cells (Barbosa et al. 2013; Somers et al. 2013). Previous studies in yeast implied that upstream AUG start codons (uAUGs) in the 5'-UTR can interfere with translation initiation in the downstream CDS by recruiting ribosomes to the 5'-region (Arribere and Gilbert, 2013; Brar et al. 2012). In order to determine if 5'-leader density and sequence content were related to CDS translation, we determined the presence or absence of

non-canonical uAUG in genes with 5'-leader ribosomal density. Figure 2.4A demonstrates these results; CDS translation efficiency distributions are plotted for genes with 5'-leader density in the RiboTag profile and separated into sets of genes which either contain or lack uAUG. Within the RiboTag enriched samples, we found that genes without uAUG were translated ~ two-times more efficiently than genes with uAUG ($p=4.1x10^{-13}$, two-sample KS-test). This trend was also observed in normal brain tissue (Figure 2.4B).

After identifying the significant differences in translation efficiency between genes containing 5'-leader ribosomal occupancy and either containing or lacking uAUG, we sought to determine if 5'-leader efficiency was differentially regulated in our glioma model. We calculated the 5'-leader efficiency for RiboTag and homogenate profiles (see methods). While we previously found that the translation efficiency of the CDS was broadly reduced in tumor compared to normal brain for RiboTag-enriched genes, the 5'-leader efficiency was not as severely affected (~90% reduction compared to ~60%). This difference is illustrated in Figure 2.4C where histograms of the ratio of the 5'-leader efficiency fold-change relative to the CDS translation efficiency fold-change for transcripts containing or lacking uAUGs for RiboTag-enriched genes with 5'-leader density. While uAUG-containing transcripts have relatively similar 5'-leader and CDS translation efficiency fold-changes compared to normal brain, 73% of transcripts lacking uAUG have their ratio of 5'-leader and CDS fold-changes greater than one. The difference in these distributions is statistically significant ($p=0.01$, two-sample KS-test) and can be improved upon by the inclusion of genes with 5'-leader density identified by the three RiboTag mice ($p=4.5x10^{-16}$, two-sample KS-test). A possible interpretation of these results is that translation initiation on non-AUG start codons and canonical initiation are regulated by different factors which are differentially regulated in transformed cells as compared to normal

brain. While examples of non-AUG translation have previously been reported, the mechanism by which this is achieved remains unclear (Chang et al. 2004).

2.3 Discussion

Translational profiling of cell types within the brain represents a unique challenge. Compositionally heterogeneous, and highly interconnected, efforts to dissociate and independently profile neural cell-types are limited by the availability of cell-type specific markers and translation's rapidly responsive and dynamic nature. However, the importance of understanding translational control in the brain, and specifically, in glioma, has led us to develop molecular and analytical tools required to bypass this challenge and others specific to glioma. Here, we have combined a proneural glioma model with the RiboTag system and used ribosome profiling to generate a comprehensive and qualitative picture of translation in glioma.

Recapitulating the human disease of glioma, our model identified several gene ontologies differentially translationally upregulated. These included pathways related to cell adhesion and the extracellular matrix known to play a role in tumor cell invasion. We also identified global alterations in the translation efficiency of tumor-related genes. Globally, the translation efficiency of cells which descended from the retrovirus-infected lineage was lower than genes expressed in other CNS lineages. This result was surprising due to the high total translational output of the tumor as compared to the rest of the brain. Additionally, while ribosomal components seemed to be upregulated in the RiboTag profile, they were almost universally translated more efficiently in normal brain. One possible explanation could be that while translational capacity in glioma is globally increased, the translational machinery may still be saturated by even greater increases in mRNA production. An additional explanation could be that

tight translational control promotes an immature cellular state, as has been previously demonstrated (Signer et al. 2014; Tahmmasebi et al. 2014).

Beyond global translational changes, we also identified widespread 5'-leader ribosomal density in both tumor and normal brain profiles. While 5'-leader ribosomal density has been previously observed, its effect on protein production is poorly understood. The presence of 5'-leader ribosomal density may not represent actively translating ribosomes but could affect the stability of mRNA or translation occurring in the downstream CDS. Alternatively, 5'-leader ribosomes may be actively engaged in the translation of upstream open reading frames (uORFs) generating small peptides, fusion proteins or full-length proteins from unannotated ORFs.

We found that the effects of 5'-leader ribosomal density were strongly dependent on sequence content. Genes containing uAUG in the 5'-leader sequence were found to have lower CDS translation efficiencies as genes lacking a uAUG. Furthermore, in comparing this effect between normal brain and tumor, we found that the fold-change in 5'leader and CDS translation efficiencies greatly diverged for genes lacking uAUG—while CDS translation efficiencies were globally lower in the tumor, 5'-leader translation efficiency was not as greatly reduced. This indicates that the process of inducing ribosomal formation in the 5'-leader must be regulated by alternative pathways which are differentially regulated in tumor as compared to normal brain.

2.4 Conclusion

Cell-type specific ribosome profiling through the use of the RiboTag system has emerged as a powerful tool necessary to study complex and highly heterogeneous tissues such as brain tumors. The altered translational landscape which we've identified in our Proneural glioma model highlights the importance of translational regulation and raises the possibility that

targeting translation machinery may be an effective therapeutic approach for anti-cancer therapies. We hope that the tools described here will be effective in assessing new therapeutic strategies targeting protein translation.

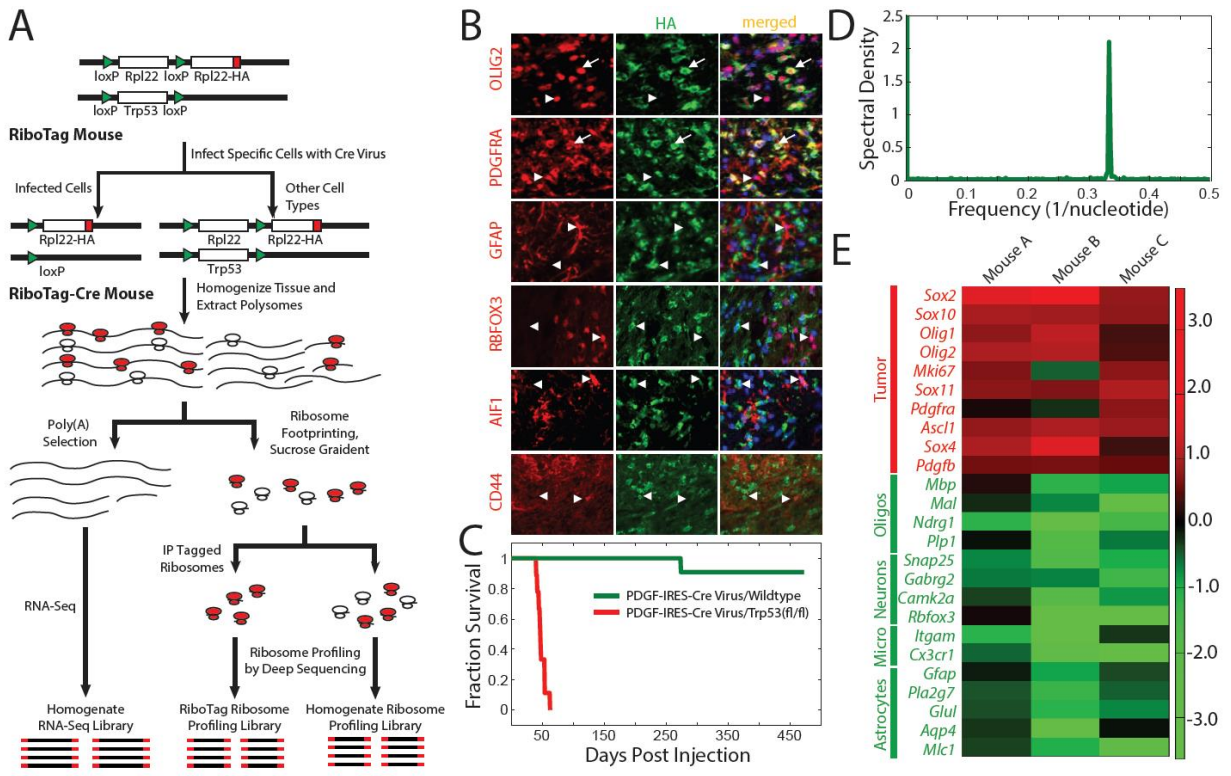

Figure 2.1 RiboTag Mouse Glioma Model and Cell Type-Specific Ribosome Profiling

A, Schematic of the RiboTag glioma mouse model and experimental workflow. Cells infected by a retrovirus that expresses Cre recombinase and PDGF-B express the RiboTag (Rpl22-HA) and harbor a transforming genetic lesion- loss of *Trp53*. Polysomes are extracted from homogenate tumor tissue. Poly(A) RNA is selected from a portion of this for RNA-Seq. The remaining polysomes are digested to monosomes and purified on a sucrose gradient. The purified monosome sample is split in half. One half is converted into a ribosome profiling library. HA-tagged (RiboTag) monosomes originating from the transformed cells are immunoprecipitated from the other half and converted into a ribosome profiling library. Translation rates from the homogenate and RiboTag ribosome profiles are compared to identify genes that are enriched or depleted in the transformed population. **B,** Immunofluorescence staining of tissue sections from

an end-stage RiboTag glioma mouse showing the diversity of cell types present in the tumor. Cells expressing HA (the RiboTag epitope) overlap significantly with OLIG2- and PDGFRA-expressing cells. However, there is essentially no overlap between cells expressing HA and cells expressing GFAP (astrocytes), RBFOX3 (neurons), AIF1 (microglia), or CD44 (reactive astrocytes). **C,** Survival curves for *Trp53*$^{flox/flox}$ and wildtype mice after injection with PDGF-B-IRES-Cre virus indicating a median survival time of 47+/-7 days post injection for our mouse glioma model. **D,** Power spectrum of the 5'-end read positions along CDSs for the first 500 bases of the CDS for all genes with a CDS length of at least 500 bases. This power spectrum was computed from the RiboTag profile of Mouse A, demonstrating that RiboTag immunoprecipitation preserves the expected three-base periodicity arising from codons as indicated by the clear peak at a frequency of ~0.33 nt^{-1}. E) Heat map displaying the translation rate enrichment scores (plotted as score-1 where a score >1 indicates enrichment in the RiboTag profile and a score <1 indicates depletion) for several canonical markers of different cell types across three mice. The enrichment score is calculated by dividing the translation rate in the RiboTag profile by that in the homogenate profile.

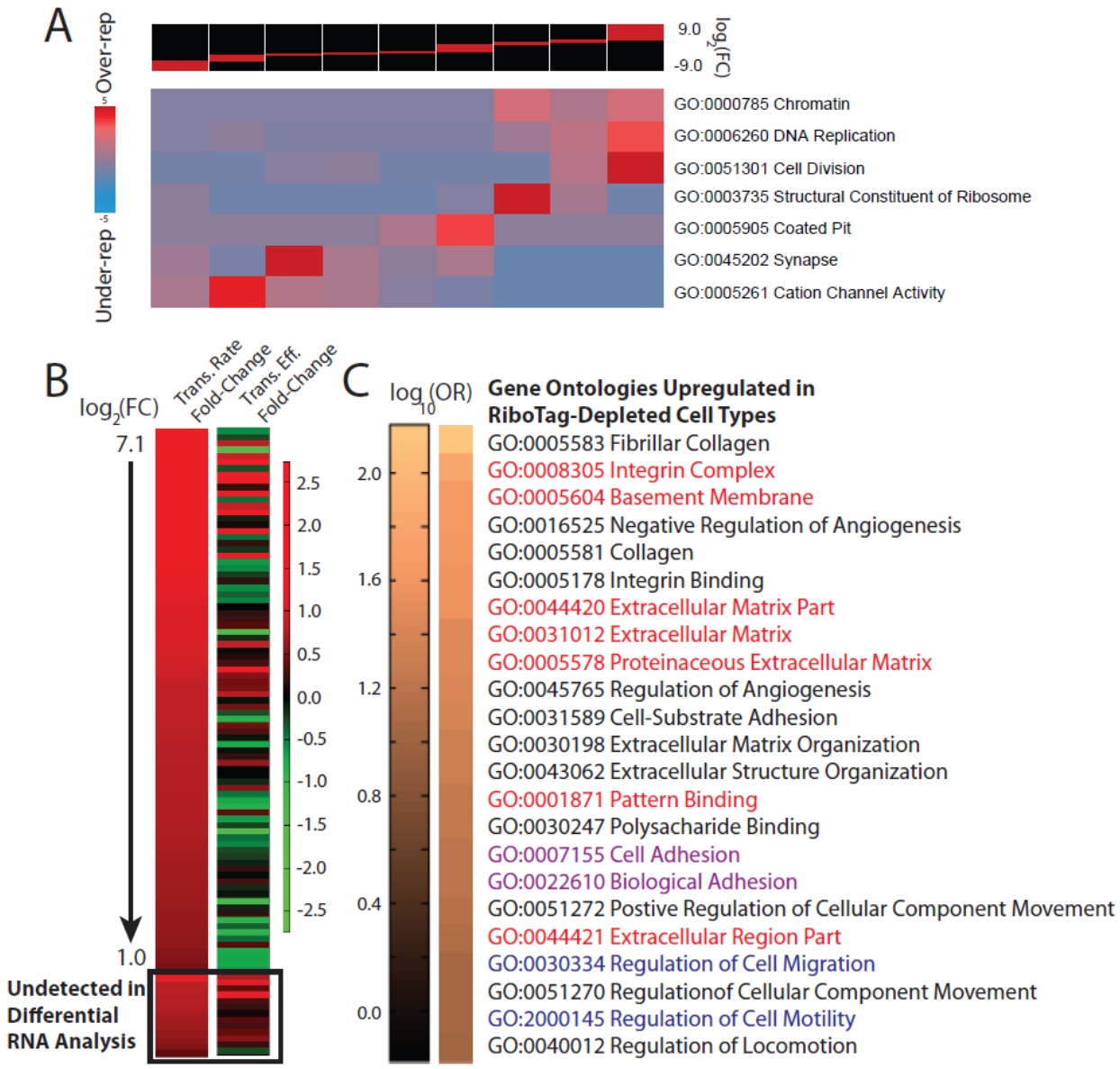

Figure 2.2 Differential Translation Rate Analysis.

A, Information theory-based iPAGE analysis of over- and under-represented gene ontologies in genes with statistically significant ($p<0.05$) high and low translation rate fold-changes indicating high translational output in the RiboTag sample and normal brain, respectively. Chromatin, DNA replication, cell division, and ribosomal pathways are over-represented among genes highly translated in the RiboTag sample, whereas coated pit, synapse, and cation channel activity

pathways are over-represented in the normal brain profile. **B,** Heat map displaying the translation rate fold-change and translation efficiency fold-change from differential translation rate analysis between the tumor homogenate and normal brain ribosome profiles. The genes in this heat map show statistically significant increased translation in the tumor homogenate relative to normal brain but are consistently depleted in the RiboTag profile, indicating expression in tumor associated cells. A subset of these genes, all but one of which exhibited higher translation efficiency in tumor tissue, was not found to have a statistically significant change in RNA abundance. **C,** Gene ontology analysis of upregulated, depleted genes from **B** with heat map of the odds ratio. Pathways in red and blue indicate overlap with Mesenchymal and Classical glioblastoma pathways, respectively. Pathways in purple indicate overlap with both Mesenchymal and Classical pathways.

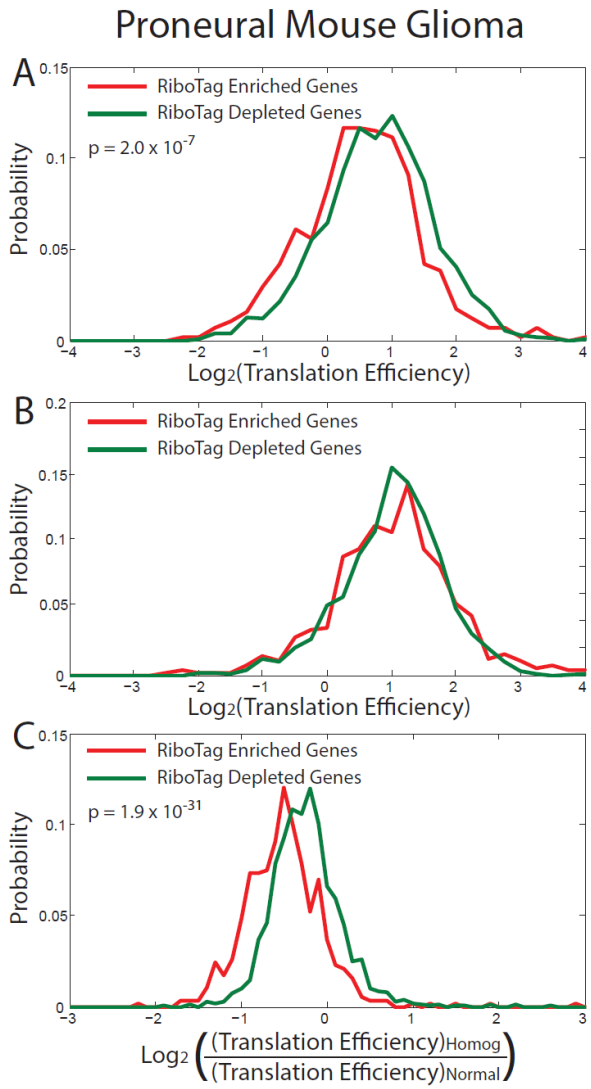

Figure 2.3 Translation Efficiency Analysis.

A, Histograms of the mean translation efficiency for genes that are either enriched (red) or depleted (green) computed from ribosome profiles and RNA-Seq of the homogenate murine tumor samples. The tumor-specific, RiboTag-enriched genes show a statistically significant tendency towards lower translation efficiency. **B,** Histograms of the mean translation efficiency for genes that are either enriched (red) or depleted (green) computed from the ribosome profiles and RNA-Seq of the murine normal brain samples. There is no statistically significant difference between the two gene sets in normal brain. **C,** Histogram of the translation efficiency fold-

change between tumor homogenate and normal brain for the RiboTag-enriched and RiboTag-depleted genes showing that ~90% of RiboTag-enriched genes are translationally downregulated in the murine tumors.

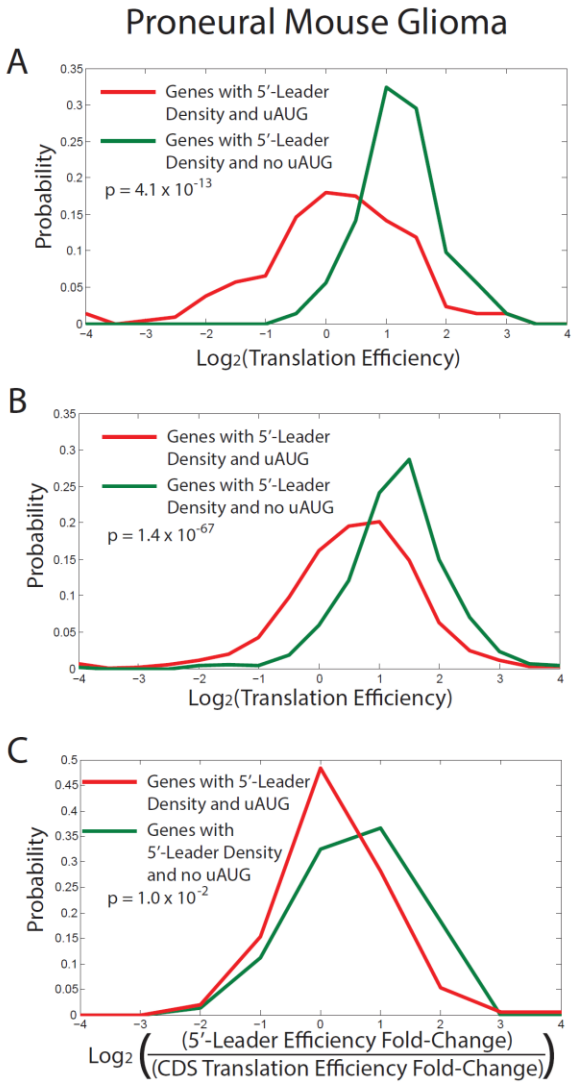

Figure 2.4 Analysis of Non-canonical Translation.

A, Histograms of the translation efficiency of genes with 5'-leader density across all three RiboTag profiles that either contain (red) or lack (green) uAUG. **B,** Histograms of the translation efficiency of genes with 5'-leader density across all three normal brain profiles that either contain (red) or lack (green) uAUG. **C,** Histogram of the ratio of 5'-leader efficiency fold-change to CDS translation efficiency fold-change for genes with 5'-leader density in all three RiboTag mice and either contain (red) or lack (green) uAUG. Fold-change in 5'-leader and CDS translation efficiencies are calculated between the tumor homogenate and normal brain samples.

Table 2.1 RiboTag-Enriched Genes with Higher Translation Rates in Tumor vs. Normal

Adam9	Frmd8	Myo9b	Rpl7a
Aebp1	Gng5	Nasp	Rps12
Afap1l2	Gsx1	Nav2	Rps2
Aldh1a3	H2afy2	Nhsl1	Rps29
Arfgap3	Igf2bp2	Notch1	Rps9
Atp5g2	Igfbp3	Pde8a	Rrm2
Bag3	Incenp	Pdlim4	Sall3
Birc5	Jam3	Pdlim5	Sapcd2
Ccnd1	Kdm6b	Plk1	Scrg1
Cd276	Kif20a	Plxnb3	Socs3
Cdkn2a	Kif2c	Pnlip	Spc24
Ckap2l	Klf3	Ppfibp1	Spry2
Col11a1	Knstrn	Ppp1r18	Spry4
Creb5	Lima1	Prdx4	Sulf2
Cspg4	Lix1l	Qpct	Tagln2
Dcps	Lmnb1	Rbmx	Tk1
Dnmt1	Lnx1	Rcc1	Tmpo
Epn2	Matn4	Rhoc	Trib2
Ezh2	Mcm10	Rpl18	Trim25
Fam101a	Mcm2	Rpl18a	Tubb2b
Fam64a	Metrn	Rpl19	Tubb6
Fam83d	Midn	Rpl21	Vim
Fbl	Morc4	Rpl27a	Ybx1
Fbn2	Moxd1	Rpl28	Zfp488
Fgfrl1	Mpzl1	Rpl35a	

Table 2.2 RiboTag-Depleted Genes with Higher Translation Rates in Tumor vs. Normal

Abca1	Col6a3	Itgam	Postn
Abcc1	Cpne3	Itgb2	Prc1
Abcg2	Ctsc	Kdr	Psmb8
Acan	Ctss	Klhl5	Ptpn12
Aldh1a2	Dcn	Lama4	Ptprb
Anxa2	Dock1	Lamb1	Rrbp1
Aspm	Ecm1	Lgals3bp	S100a6
B2m	Ednrb	Lig1	Scpep1
C1qtnf6	Eltd1	Lmf2	Sec61a1
Cad	Emp1	Lrg1	Serpine1
Cd248	Enpp1	Lum	Serpinf1
Cd74	Ercc6l	Lyn	Slc16a3
Cd93	Esam	Lyz2	Smc4
Cd97	Fancd2	Mmp2	Snx9
Cdca8	Fat1	Mmp9	Spp1
Cdh5	Flna	Mpeg1	Stab1
Chst11	Fn1	Mtap	Syde1
Clic1	Gene	Ncapd3	Thbs1
Col15a1	Gnb2l1	Nid2	Thbs2
Col19a1	Gnb4	Nup188	Tm4sf1
Col3a1	Gpr56	Nup205	Tnc
Col4a1	Grn	Nup93	Top2a
Col4a2	Igfbp7	Pdgfrb	Usp24
Col5a2	Itga1	Pglyrp1	Vwf
Col5a3	Itga7	Plau	Zfp191

CHAPTER 3

Ligation-Free Ribosome Profiling of Cell Type-Specific Translation in the Brain

3.1 Background

Abstract

Ribosome profiling has emerged as a powerful tool for genome-wide measurements of translation, but library construction requires multiple ligation steps and remains cumbersome relative to more conventional deep sequencing experiments. In this chapter, I discuss a new, ligation-free approach to ribosome profiling that does not require ligation. Highlighting this approaches benefits, library construction for ligation-free ribosome profiling can be completed in one day with as little as 1 ng of purified RNA footprints. Ligation-free ribosome profiling was applied to mouse brain tissue to identify new patterns of cell type-specific translation and tested its ability to identify translational targets of mTOR signaling in the brain.

Introduction

Ribosome profiling allows genome-wide measurements of ribosomal occupancy with single nucleotide resolution (Ingolia, 2009). Using deep sequencing as a readout for protein synthesis, the technique has enabled the discovery of previously unannotated open reading frames (ORFs) (Ingolia, 2009;Ingolia, 2011;Brar, 2012;Calviello, 2016) and provided new insights into the mechanisms of translation initiation and elongation (Lareau, 2014), localized translation (Jan, 2014), and the signaling pathways underlying translational control (Thoreen, 2012;Hsieh, 2012). In addition, ribosome profiling has been applied in many cellular contexts

including yeast (Ingolia, 2009), bacteria (Oh, 2011), primary mammalian cells (Ingolia, 2011), and complex tissues (Gonzalez, 2014) to assess the role of translational control in basic physiological processes and its dysregulation in diseases like cancer.

While ribosome profiling is widely used, the library preparation procedure is relatively complex (Ingolia, 2012). Most protocols involve nuclease footprinting of polysomal RNA followed by purification of ribosome-bound mRNA footprints using a sucrose gradient, sucrose cushion, or gel filtration column. After isolation of mRNA footprints by gel electrophoresis, one of multiple library preparation schemes is used to attach universal sequence adapters to the mRNA or cDNA footprints using either single-stranded intermolecular ligation (Guo, 2010;Weinberg, 201682) and/or intramolecular circularization (Ingolia, 2009;Ingolia, 2012) (Figure 3.1A). Because these protocols often involve multiple ligation, gel purification, and nucleic acid precipitation steps, library preparation alone typically takes several days (Ingolia, 2012). Here, we report a new approach to library construction for ribosome profiling that eliminates ligation and requires only one initial gel purification step to isolate RNA footprints (Figure 3.1A). The procedure, which is based on template-switching (Luo, 1990;Zhu, 2001), is highly sensitive and requires only ~1 ng of gel-purified RNA footprints. Following footprint isolation, library construction for ligation-free ribosome profiling can be completed in one day.

In addition to characterizing the performance of ligation-free ribosome profiling, we applied our technique to assess cell type-specific translational regulation in the murine brain. The brain harbors a broad diversity of cell types including astrocytes, oligodendrocytes, microglia, glial progenitors, endothelial cells, and many different types of neurons that likely control translation through different signaling pathways. In addition, many neuron-specific transcripts are translated locally in dendrites, and translational control has been shown to play a key role in

memory (Cho, 2015;Kelleher, 2004;Kandel, 2001;Davis, 1984). We took advantage of a recently reported database of neural cell-specific gene expression (Zhang, 2014) to identify patterns that indicate cell type-specific regulation of translation. As an orthogonal validation for neuron-specific genes, we used the RiboTag system (Sanz, 2009) to purify and identify actively translated transcripts from excitatory neurons in the cortex of Camk2a-Cre/RiboTag mice. Finally, we used our technique to identify the genes controlled by mTOR signaling in the brain by conducting ribosome profiling on the brains of mice treated with AZD-8055, an ATP-competitive inhibitor of mTOR that crosses the blood-brain barrier (Chresta, 2010;Pike, 2013).

3.2 Results

A Ligation-Free Protocol for Ribosome Profiling

Ribosome profiling is more complicated than conventional RNA-Seq because the ribosome-protected mRNA footprints are short (~30 nucleotides) and lack poly(A) tails, which are often used as handles for either isolation or reverse transcription of eukaryotic mRNA. Previously established protocols for ribosome profiling address this problem by single-stranded ligation of a universal adapter to the 3'-end of mRNA footprints to facilitate reverse transcription, which incorporates a longer adapter into the 5'-end of the resulting cDNA (Ingolia et al. 2012). Intramolecular ligation (circularization) of the cDNA effectively attaches a universal adapter to the 3'-end of the cDNA to enable PCR enrichment of the library (Ingolia et al. 2012). Alternatively, a second ligation reaction can be used to attach an adapter to the 3'-end of the cDNA. These ligation reactions are notoriously inefficient and require excess adapter which is typically removed by gel purification and subsequent overnight precipitation of the product (Ingolia et al. 2012). These multi-step procedures and intermediate purification steps require

multiple work days, are intrinsically lossy, and therefore require relatively high input (Ingolia et al. 2012).

To address these issues, we have applied the template switching approach to library construction that has been successfully implemented in other low-input RNA sequencing protocols such as single cell RNA-Seq (Islam et al. 2011;Ramskold et al. 2012;Picelli et al. 2014). Specifically, we have adapted a newly developed version of the SMARTer library construction technology (Clontech) for ribosome profiling (Figure 3.1A). We first polyadenylate dephosphorylated RNA footprints using RNA poly(A) polymerase, similar to the earliest reported protocol for ribosome profiling (Ingolia et al. 2009). We then reverse transcribe the polyadenylated footprints using an enzyme with template-switching activity. In a template switching reaction, the reverse transcriptase (RT) first extends a primer (in this case oligo(dT) linked to a universal sequence on its 5'-end) to produce cDNA. Once the RT reaches the end of the RNA template, the terminal transferase activity intrinsic to the RT adds a low complexity sequence to the 3'-end of the cDNA in a non-template directed fashion. The reaction is carried out in the presence of a second universal sequence adapter that is 3'-terminated with a low-complexity sequence, which hybridizes to the tail added to the cDNA by the RT. Upon hybridization of this second sequence adapter, the RT switches templates and copies the second adapter onto the 3'-end of the cDNA. As a result, both 5' and 3' universal adapters are simultaneously added to the cDNA in a single reaction without single-stranded ligation or intermediate purification steps. We then deplete the resulting product of rRNA using complementary oligonucleotides (Ingolia et al. 2012) and enrich the deep sequencing library by PCR.

Comparison of Ligation-Free and Conventional Ribosome Profiling

We used ligation-free ribosome profiling to measure genome-wide translation in the forebrains of adult mice. Unlike fragments generated in RNA-Seq, ribosome footprints map to the transcriptome with a three-nucleotide periodicity due to the characteristic translocation interval of the ribosome as it translates codons (Ingolia et al. 2009). To verify that the RNA libraries generated using our technique originate from ribosome footprints, we computed the power spectrum of the 5' mapping positions of RNA fragments (Figure 3.1B). As expected, the data are highly periodic with a characteristic frequency of ~0.33 nucleotides^{-1}, similar to what has been observed for conventional ribosome profiling (Ingolia et al. 2009). In addition to three-nucleotide periodicity, ribosome profiling also exhibits a characteristic gene body distribution. The majority of reads are expected to map to the coding sequences (CDS) of transcripts, whereas relatively few should map to the untranslated regions (UTRs)(Ingolia et al. 2009). Many genes have been shown to contain unannotated upstream ORFs (uORFs), and so we also expect that more reads will map to the 5'-UTRs than the 3'-UTRs, as 3'-UTRs are largely depleted of ribosomes. As shown in Figure 3.1C, ligation-free ribosome profiling reads map to the transcriptome with the expected gene body distribution.

To further validate the technique, we compared these results to our previously reported mouse forebrain data that we generated using conventional ribosome profiling (Gonzalez et al. 2014). Figure 3.1D shows that the ribosome footprint counts for each gene across the two data sets are highly correlated. We also compared the gene detection efficiency, saturation properties, and library complexities of the two data sets. We note that in our previously reported experiment with conventional ribosome profiling, we used more input monosomal RNA for library construction than in the current experiment with ligation-free ribosome profiling. In Figures

3.1E and 3.1F, we use downsampling analysis to show that the two data sets are quite similar in terms of both the number of genes detected and number of unique ribosome footprints detected, respectively, at a given sequencing depth. These results imply that the library complexities produced by the two protocols are highly comparable.

In order to determine the sensitivity of both conventional and ligation free ribosome profiling, we generated libraries from a defined 34-base RNA oligonucleotide at five input levels ranging from 0.01 to 100 ng. We constructed Illumina libraries from each dilution using the convention ribosome profiling protocol described by Ingolia *et al* (Ingolia et al. 2012) and the ligation-free protocol described here. We then assessed our yield for each dilution using an Agilent Bioanalyzer (Figure 3.6). We found that the ligation-free method is more sensitive and able to generate detectable libraries from less than 1 ng of input. For both methods we were able to generate quantifiable libraries, however, we were only able to generate libraries at 10 and 100 ng of input when using conventional protocol with nine PCR cycles. In contrast, we were able to generate detectable libraries at all concentrations tested when using the ligation-free protocol with nine PCR cycles. We note that the 10 and 100 ng input libraries made with the ligation-free protocol exhibit over-amplification as evidenced by a broader product length distribution at higher-than-expected molecular weights. To directly compare all of the samples, we kept the number of PCR cycles constant and note that lower cycle numbers could be used to avoid over-amplification of higher input libraries with the ligation-free protocol. In addition, we note higher cycle numbers may result in sufficient library yields for the conventional protocol at lower concentrations, although this could result in increased amplification bias.

Cell Type-Specific Translation in the Brain

One of the key metrics obtainable from ribosome profiling experiments is the translation efficiency (TE), which can be computed for each gene as the ratio of its ribosome footprint density to its expression level measured by RNA-Seq (Ingolia et al. 2009). TE is proportional to the number of ribosomes per transcript averaged over all copies of a given gene.

We used ligation-free ribosome profiling and RNA-Seq to measure TE in the brain of an adult mouse, a complex tissue comprised of many different cell types. Both ribosome footprint densities and expression levels are complicated by cellular composition. This is also true to a large extent for TE, however, because TE is a ratio, the TE measured in homogenized tissue for a cell type-specific gene is accurate for both the tissue and the specific cell type that expresses the gene. Figure 3.2A shows the broad distribution of TEs for genes expressed in the brain of an adult mouse. While this result implies that there is a great deal of translational regulation in the brain, it tells us nothing about the contributions of different cell types.

We validated our TE measurements by performing qPCR on a set of highly translated (Syt1, Snap25) and lowly translated (Trpv6, Tgfb1, Pkd1) genes based on our ribosome profiling data. We first used sucrose gradient fractionation to separate mRNAs based on the number of bound ribosomes and collected fractions. We then used qPCR to assess the relative abundance of each gene in each fraction (Figure 3.7). There are several complications associated with directly comparing qPCR data obtained from polysome profiles and ribosome profiling data. While the majority of transcripts for a highly-translated gene may appear in polysomes with more than five ribosomes per transcript, resolution constraints make it difficult to accurately measure the number of bound ribosomes for each fraction, particularly for heavier polysomes. Furthermore, calculating TE based on log ratios without correcting for cytosolic mRNA levels has been

previously shown to produce an inaccurate estimation of TE (Larsson et al. 2010). While it is difficult to quantitatively compare TE calculated from next-generation sequencing to that obtained from qPCR, we found that the highly translated genes probed are clearly shifted to heavier polysomes compared to the lowly translated genes probed. For example, we found that the maximum abundance of the highly translated genes Syt1 and Snap25 were in the seventh and ninth polysome fractions (greater than five ribosomes per transcript), respectively (Figure 3.7). However, the maximum abundances of Trpv6, Tgfb1, and Pkd1, all of which are lowly translated, were in the fourth and fifth fractions (two or three ribosomes per transcript).

We also compared our ligation-free ribosome profiling and RNA-Seq data with a previously published whole-brain mass-spectrometry data-set obtained from a mouse of similar genetic background and age (Sharma et al. 2015). We found that our ribosome profiling data was better correlated with protein abundance in the brain than our corresponding RNA-Seq measurements (Figure 3.8). Hence, some of the difference in the explained variance may be attributable to the contribution of translation regulation on protein expression. This result is consistent with previously published observations in yeast in which mass spectrometry, RNA-Seq, and ribosome profiling were compared (Ingolia et al. 2009).

A recent study by Zhang and colleagues produced RNA-Seq expression profiles from seven different cell types in the brain by sorting or immune-panning, including astrocytes, neurons, oligodendrocyte progenitor cells (OPCs), newly formed oligodendrocytes, myelinating oligodendrocytes, microglia, and endothelial cells (Zhang et al. 2014). We used this data set to compute cell-type enrichment scores proportional to the specificity with which each gene is expressed in each cell type (see Methods). We then divided the transcriptome into ten gene sets evenly binned by TE and conducted gene set enrichment analysis (GSEA) against rank-ordered

lists of cell-type enrichment scores for each cell type (Subramanian et al. 2005). This analysis allowed us to systematically associate genes with varying degrees of cell type specificity and TE. The normalized enrichment score (NES) for each GSEA is shown in the heatmap in Figure 3.2B (with bin-by-bin and cell type-by-cell type statistical analysis in Figure 3.9), which reveals several interesting patterns. First, we found that microglial genes generally exhibit low TEs. Because we are studying the brains of healthy mice, these microglia are presumably not in an activated state. Previous studies have shown that protein synthesis-associated pathways are upregulated in microglia in certain disease contexts (Chiu et al. 2013), and so these results could be dependent on genotype or other activating conditions such as injury or an inflammatory stimulus. Conversely, neurons, when considered as a broad group, exhibit the highest degree of variation in TE among their cell type-specific genes. As shown in Figure 3.2B, most neuronal genes are either very highly or very lowly translated, suggesting that neuronal genes are under a relatively high degree of translational regulation in comparison to other cell types in the brain.

Translational control is well-known to play an important role in neuronal function and memory formation. Structurally, neurons are highly complex cells that make extensive use of local translation to efficiently modulate protein expression far from the soma (Steward et al. 2001). To validate our observation that neuronal genes are highly translationally regulated, we used the RiboTag system to isolate polysomal mRNAs from a specific neuronal subtype, namely excitatory neurons that express Camk2a. As shown in Figure 3.2C, the RiboTag mouse harbors a modified ribosomal protein L22 (Rpl22) gene with a floxed terminal exon followed by a second copy of the terminal exon with a triple hemagluttinin tag (HA-tag)(Sanz et al. 2009). We crossed the RiboTag mouse with a mouse that expresses Cre recombinase under the control of the Camk2a promoter to produce mice which express HA-tagged ribosomes in Camk2a-expressing

cells. Figure 3.2D shows that, as expected, the HA-tag is expressed exclusively in neurons, marked here by the pan-neuronal marker NeuN (Rbfox3). Hence, we can isolate polysomes from homogenized brain tissue of Camk2a-RiboTag mice and purify mRNA-ribosome complexes that originate from Camk2a-expressing neurons by immunoprecipitation (IP) of the HA-tag (Figure 3.2C). We obtained RNA-Seq expression profiles from both homogenized brain tissue and immunoprecipitated polysomes of two Camk2a-RiboTag mice. We compared the expression levels of each gene in the immunoprecipitated and homogenate profiles and observed that canonical markers of excitatory neurons were enriched by IP, whereas markers of other cell types in the brain, including inhibitory neurons, were depleted by IP (Figure 3.2E). We then repeated the GSEA described above with TE gene sets and genes rank-ordered based on their enrichment by RiboTag IP. This analysis recapitulated the results found for neuronal genes derived from purified neurons in that genes specific to Camk2a-expressing neurons, and not just neurons in general, appear highly translationally regulated (Figure 3.2B). A subset of genes expressed in these neurons exhibit relatively high TE, while the remaining exhibit relatively low TE. Not only do these results provide an orthogonal validation of our GSEA based on pan-neuronal gene expression, they also show that the pattern holds for a specific subtype of excitatory neurons in the cerebral cortex.

Finally, these data reveal a simple developmental trend in the oligodendrocyte lineage. Oligodendrocytes, which are primarily responsible for enwrapping neuronal axons with myelin sheaths, are a unique cell type in that their progenitor cells (OPCs) are widely distributed in the adult brain, where they actively proliferate and differentiate to generate new myelinating oligodendrocytes. Hence, we can detect gene expression and translation from different stages of oligodendrocyte development within homogenized brain tissue. Based on our analysis, OPC-

specific genes are translated more efficiently than those of either newly formed or mature, myelinating oligodendrocytes, which exhibit the lowest TE of the three. As shown in our statistical analysis in Figure 3.9, the comparison between OPCs and myelinating oligodendrocytes is very significant for highly translated genes as is the comparison between newly formed oligodendrocytes and myelinating oligodendrocytes. While one might expect myelinating oligodendrocytes to be less translationally active in comparison to OPCs because they are post-mitotic, their primary role in the brain is to produce large amounts of myelin, which is comprised mainly of proteins and lipids. Nonetheless, we found that most myelin genes have low TE compared to the overall median in the brain ($\log_2$(TE) = -0.02) including Mog (-0.15), Mbp (-0.51), Mobp (-1.42), and Mag (-0.28) with the exception of the transmembrane protein Plp1, which has a TE of 1.02. Hence, despite the importance of protein synthesis to the function of myelinating oligodendrocytes, translation of oligodendrocyte-specific genes is relatively inefficient.

We used gene ontologies (GOs) to further refine these insights into cell type-specific translation. In Figure 3.3, we used GSEA to identify GOs that were strongly associated with cell type-specific genes from each of six cell types in the brain. We then produced heat maps indicating the median TE of each GO. Figure 3.3 contains many of the qualitative patterns found in Figure 3.2B, with neuronal GOs exhibiting a broad range of TEs and microglial and oligodendrocyte GOs exhibiting relatively low TEs. In addition, this analysis reveals some of the gene functions associated with the highly translated and lowly translated neuronal genes. For example, genes associated with synaptic function, particularly those that are released by neurons in a synapse, are generally highly translated. Conversely, sodium, potassium, and most particularly, calcium channels exhibit much lower TEs.

uORFs and 5'-UTRs in the Brain

One of the most intriguing findings of ribosome profiling studies in eukaryotes is the prevalence of unannotated upstream open reading frames (uORFs) which manifest as ribosomal density in the 5'-UTRs of mRNAs(Ingolia et al. 2009;Brar et al. 2012;Calviello et al. 2016;Ingolia et al. 2011). Recent studies have further refined these observations using computational methods to infer which instances of 5'-UTR density actually represent active uORF translation and correlate with direct observations of specific peptides in mass spectrometry (Calviello et al. 2016). Using our mouse brain dataset produced with ligation-free ribosome profiling, we have investigated the 5'-UTR ribosomal density among cell type-specific genes. Figure 3.4A shows that we detect 5'-UTR ribosomal density in a consistent fraction of genes across all cell type-specific gene sets. Previous studies using conventional ribosome profiling have shown that 5'-UTR ribosomal density is associated with different levels of CDS translation depending on sequence context (Brar et al. 2012;Arribere et al. 2013;Gonzalez et al. 2014). Specifically, 5'-UTRs that harbor ribosome density but do not contain AUG sequences are associated with genes with higher TE in the annotated CDS, suggesting a potential regulatory role for upstream ribosomal density. Figure 3.4B shows that this general trend is borne out across all of our cell type-specific gene sets.

We also sought to determine how more general features of the 5'-UTR affect translation efficiency of the corresponding CDS in the brain. Figure 3.4C contains a heat map that simultaneously displays the relationships between CDS TE and both the length and GC-content of the 5'-UTR across the transcriptome. Figures 3.4D and 3.4E display these relationships independently. In general, longer 5'-UTRs are associated with low TE and both high and low

GC-content are associated with low TE. Previous studies have shown that genes with highly structured 5'-UTRs are less abundant at the protein level in yeast (Dvir et al. 2013) which is consistent with the reduced TE associated with long, GC-rich 5'-UTRs observed here.

Translational Targets of mTOR in the Brain

A common application of ribosome profiling is the identification of translational alterations in response to perturbations such as drug treatment or stress. Cells have evolved elegant mechanisms for regulating the translation of specific genes, often through the interaction of signaling molecules with translation factors that control TE through specific cis-regulatory elements in mRNA. We sought to further test the efficacy of our ligation-free ribosome profiling method in the context of this important application by identifying the translational targets of mTOR signaling in the brain.

mTOR plays a crucial role in the translational control of ribosomal proteins and protein factors involved in translation initiation and elongation (Meyuhas et al. 2000). Many of these genes contain a terminal oligopyrimidine or TOP motif in their 5'-UTRs through which translational control is thought to be mediated (Meyuhas et al. 2000). Multiple studies have used ribosome profiling to show that mTOR inhibition causes a coherent decrease in the TEs of the TOP motif-containing genes in cell culture (Hsieh et al. 2012;Thoreen et al. 2012). mTOR is an important drug target in multiple neurological disorders (Wong et al. 2013). For example, rapalog inhibitors of mTOR have been shown to mitigate seizures in certain contexts (Wong et al. 2013). We sought to determine whether mTOR controls the same set of target genes in brain.

We treated mice for one hour with AZD-8055, an ATP-competitive inhibitor of mTOR that has been shown to cross the blood-brain barrier (Chresta et al. 2010;Pike et al. 2013). We

used a competitive inhibitor because previous work has shown that allosteric mTOR inhibitors like rapamycin do not induce the same level of translational alterations as competitive inhibitors (Hsieh et al. 2012). This is, in part, because allosteric compounds do not fully inhibit 4E-BP phosphorylation, which is thought to be the primary mediator of translational control through which mTOR acts (Thoreen et al. 2012). Figure 3.5A shows the effects of AZD-8055 on the phosphorylation of Rps6, which is phosphorylated by the protein kinase Rps6kb1 (i.e. p70S6K) that is activated by mTOR. As expected, Rps6 phosphorylation is clearly detectable in the brain, particularly in neurons, in an untreated mouse, but becomes undetectable in a mouse treated with AZD-8055 based on both immunofluorescence (Figure 3.5, A) and Western blot analysis (Figure 3.10).

We used ligation-free ribosome profiling to compare genome-wide TEs in mice treated with AZD-8055 and vehicle-treated mice. We then conducted a differential TE analysis comparing the treated and untreated conditions to identify genes with significant translational alterations (see Methods). Figure 3.5B shows that overall, the amplitude of the observed alterations in TE are much larger than those found at the level of transcription alone. In addition, Figure 3.5B shows that all of the canonical TOP motif-containing genes exhibit reduced TE in the brains of mice treated with the mTOR inhibitor AZD-8055. Furthermore, most of these TE changes are highly significant based on our differential translation analysis (Figure 3.5, C). Overall, we found 37 genes with significant TE reduction after treatment and fold-change amplitudes greater than two. Of these 37 genes, 25 were in the list of canonical TOP motif-containing genes (Thoreen et al. 2012). Of the remaining 12 genes, all but one are ribosomal proteins and all 12 genes clearly contain TOP motifs. Not only do these results further validate our ligation-free ribosome profiling technique, they also demonstrate rapid and widespread

translational control of the TOP motif-containing genes by mTOR in the brain only one hour following administration of an inhibitor.

3.3 Discussion

We have demonstrated a new approach to library construction for ribosome profiling and used it to show new cell type-specific patterns of protein synthesis in the brain. Through the use of template-switching, we bypassed several inefficient and time-consuming steps associated with conventional ribosome profiling such as ligation, and eliminated almost all gel purification steps. Using ligation-free ribosome profiling, we can construct libraries from as little as 1 ng of purified RNA footprints, and the resulting library complexity and gene detection efficiency is comparable to that of conventional ribosome profiling. Furthermore, due to the elimination of several enzymatic and precipitation steps, the amount of time required to perform library construction with ligation-free ribosome profiling is as little as one day following isolation of RNA footprints.

Although ligation-free ribosome profiling offers the advantages described above, conventional ribosome profiling has some advantages in terms of resolving ribosome footprints. Both the 3'- and 5'-ends of ligation-free ribosome profiling reads are associated with low complexity sequences. Specifically, the 3'-end is poly(dA) and the 5'-end is another low complexity sequence. This complicates precise determination of the ribosome footprint insert sequence- a problem that is resolved by ligation of specific sequence adapters in the conventional library construction protocol. Nonetheless, for the purposes of measuring translation efficiency and other metrics presented here, this shortcoming does not pose a major issue.

Using ligation-free ribosome profiling, we have shown that genes expressed in specific cell types exhibit distinct distributions of translation efficiency in the brain. Interestingly, most

neuron-specific genes have either relatively high or low translation, implying that they are under a high level of translational regulation. We validated these findings in Camk2a-expressing neurons using the RiboTag system, which allows isolation of polysomal mRNA from specific cell types. At the level of gene ontologies, neuron-specific genes involved in synaptic function are efficiently translated as a group compared to, for example, neuron-specific ion channels. We also found that genes associated with three stages of oligodendrocyte differentiation exhibited different translation efficiencies. OPC-specific genes were translated more efficiently than genes specific to newly formed oligodendrocytes, while fully differentiated, myelinating oligodendrocyte-specific genes had the lowest translation efficiency of the three stages. We have also determined the relationship between CDS translation efficiency and the GC-content and length of 5'-UTR sequences in the brain. In general, long, GC-rich 5'-UTRs are associated with low translation efficiency, consistent with the notion that genes containing highly structured 5'-UTRs are lowly translated. Finally, we observed widespread translational repression of genes containing the TOP-motif in response to mTOR inhibition. Our treatment window was just one hour, suggesting that these alterations comprise the earliest effects of competitive mTOR inhibition in the brain.

3.4 Conclusion

Taken together, the above results provide convincing evidence that ligation-free ribosome profiling allows rapid and quantitative translational profiling, even in complex tissues like the mammalian brain. We anticipate that the simplified procedure described here will expand the use of ribosome profiling, and may enable new, low-input or larger-scale applications.

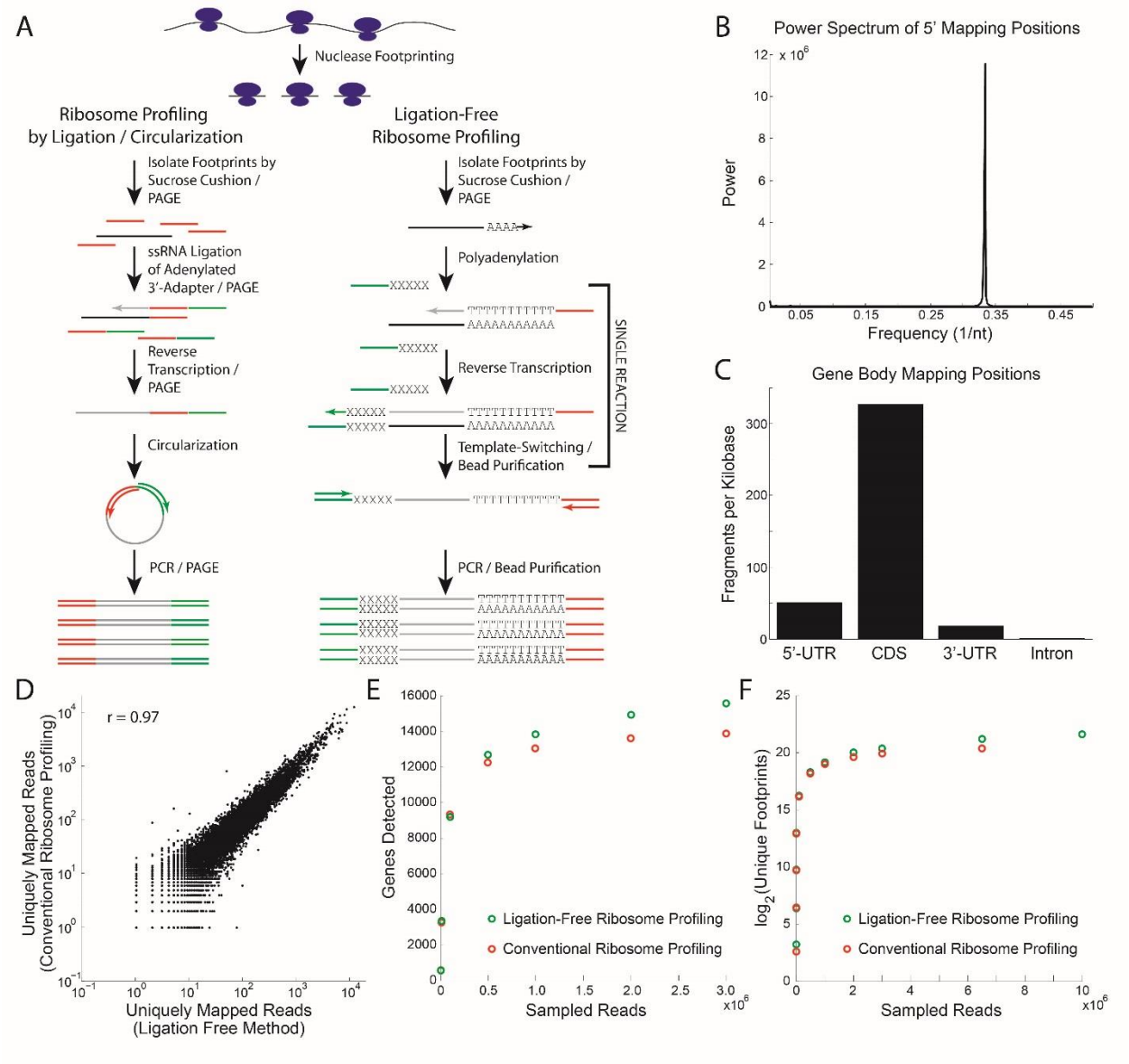

Figure 3.1 Comparison of Ligation-Free Ribosome Profiling to Conventional Methods

A, Schematic of the steps involved in conventional ribosome profiling and ligation-free ribosome profiling. **B,** The power spectrum of 5' mapping positions from CDS reads resulting from ligation-free ribosome profiling method shows clear three-base periodicity that is characteristic of ribosome profiling libraries and reflects the single-codon translocation of the ribosome. **C,** Gene body distribution of mapped reads from ligation-free ribosome profiling show strong

preference for coding-sequence, an additional property inherent to ribosome profiling libraries. **D,** Comparison of the number of uniquely mapped reads per gene in libraries generated with footprints from mouse-forebrains prepared with conventional ribosome profiling strategy and the ligation-free method; the Pearson correlation r = 0.97 indicates a concordance between the two methods. **E,** Saturation analysis showing the number of unique genes detected following downsampling of ligation-free ribosome profiling and conventional ribosome profiling. **F,** Saturation analysis showing the number of unique footprints detected following downsampling of ligation-free ribosome profiling and conventional ribosome profiling.

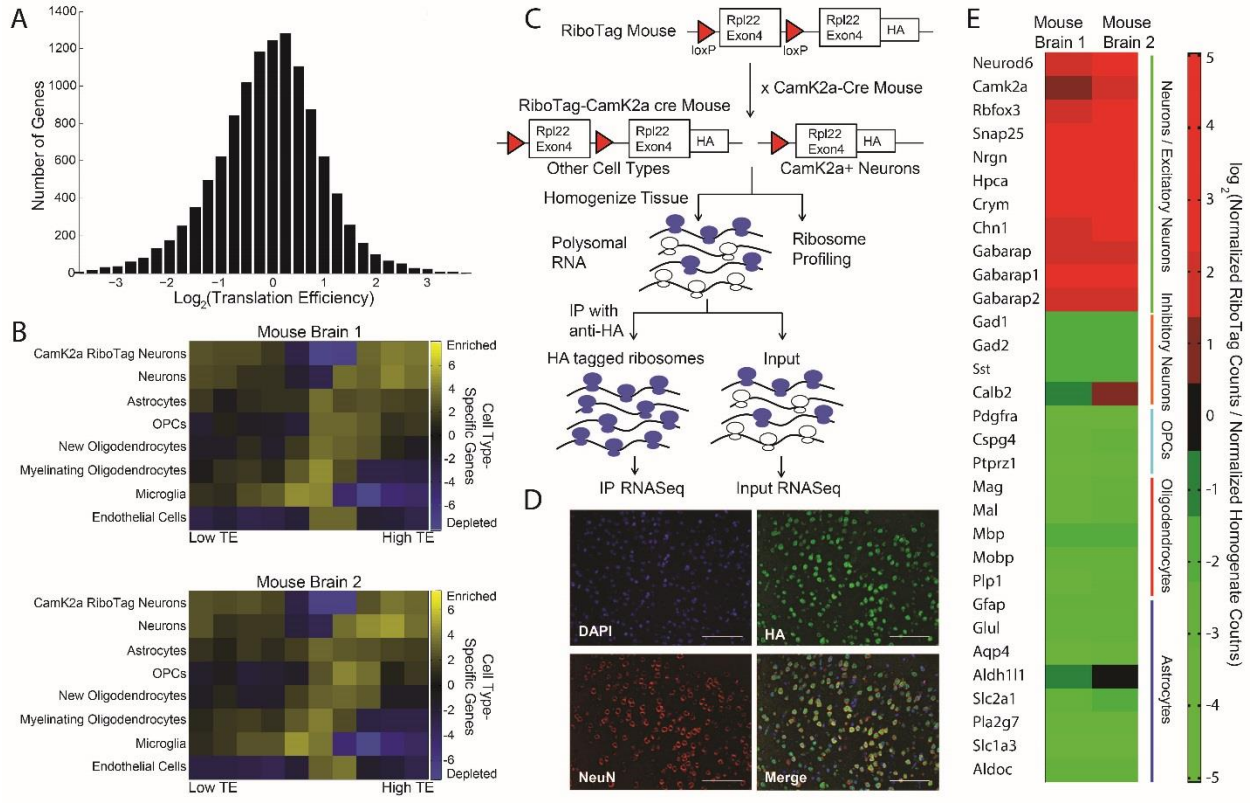

Figure 3.2 Unique Patterns in the Translation Efficiency of Cell Type-Specific Genes in the Brain

A, Histogram showing the broad range of translation efficiencies (TEs) across genes expressed in the mouse brain based on ligation-free ribosome profiling. **B,** TEs measured in two different mouse brains with ligation-free ribosome profiling were combined with cell type-specific RNA-Seq data to systematically associate cell type-specific gene expression and TE. We used gene set enrichment analysis (GSEA) to associate gene sets assembled from genes with similar TEs with a ranked list of all genes order by cell type-specificity for each cell type in the brain. The resulting heatmaps show the enrichment of genes with different TEs in cell type-specific genes for each cell type. Cell type specific genes were identified using either RNA-Seq data from sorted populations or RiboTag RNA-Seq data (for Camk2a-expressing neurons). **C,** A schematic of the RiboTag mouse model shows how the Camk2a-RiboTag mouse was generated. This

provides an orthogonal means of identifying neuron-specific genes that are actively translated. **D,** Fluorescence imaging shows that Rpl22-HA (from the RiboTag allele) expression is specific to Rbfox3+ (NeuN+) cells (a pan-neuronal marker). **E,** Heatmap of the RiboTag enrichment scores following immunoprecipitation of polysomes from Camk2a-RiboTag mouse brains demonstrates strong enrichment of genes specific to excitatory neurons and depletion of genes specific to other cell types in the brain in two different mouse brains.

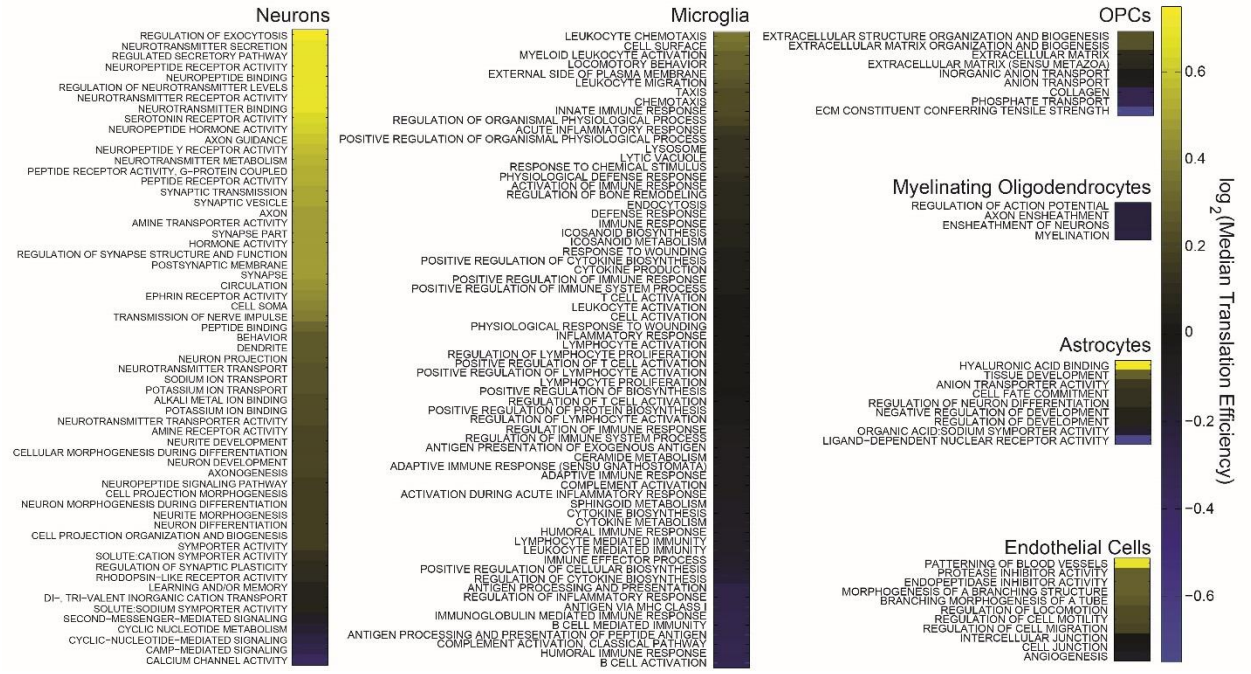

Figure 3.3 Cell-Type Specific Gene Ontologies Recapitulate Global Translation Efficiency Trends

We used GSEA to identify gene ontologies enriched in cell type-specific genes. An enrichment score was calculated for all genes in each cell-type based on RNA-Seq data from sorted neural cell types. This information was placed into six different rank lists, one for each cell type. A gene ontology was defined as being cell-type specific if it had a NES score for a cell-type that was at least three units greater than the next highest NES score. Ligation-free ribosome profiling datasets from two mouse brains were averaged and used to calculate the median translation efficiency for each ontology. Highly enriched ontologies and their median translation efficiencies in descending order are displayed in the heatmaps.

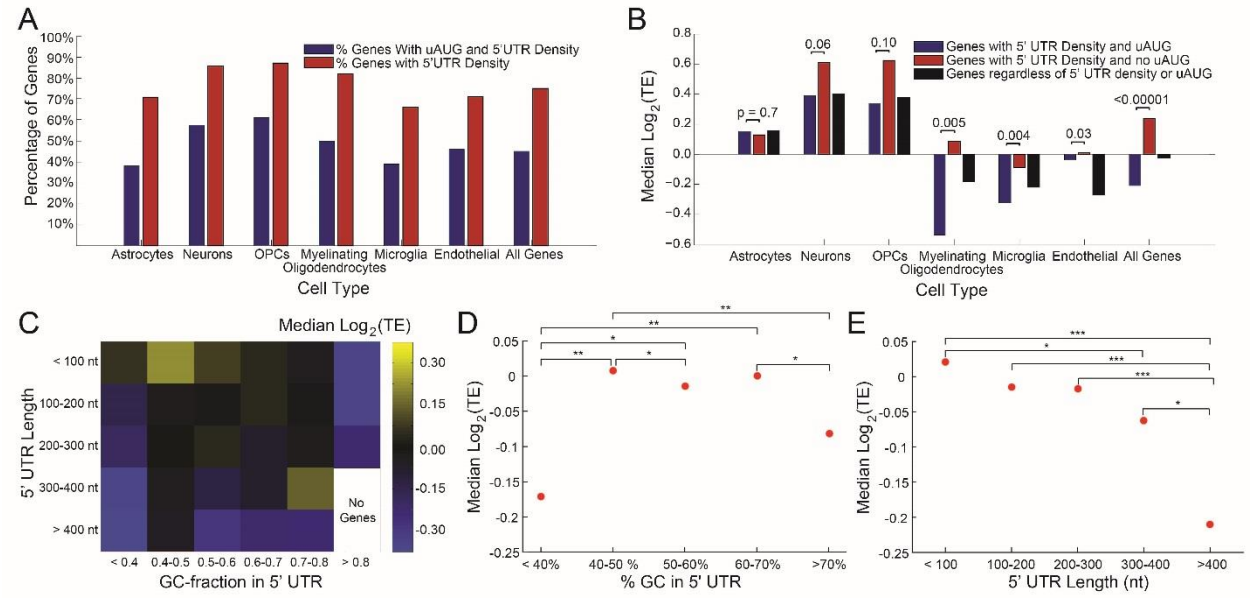

Figure 3.4 Features of 5' UTRs are Associated With CDS Translation

A, The percentage of cell-type specific genes with at least one ribosome footprint mapping to their 5'UTR is plotted alongside with the percentage of cell-type specific genes with 5'UTR ribosomal density and also containing a uAUG sequence. These values are highly consistent across cell types. **B,** Genes containing a uAUG and 5'UTR ribosomal density had lower CDS TE compared to genes without a uAUG. This effect was consistent across multiple cell types and was significant for myelinating, microglial, and endothelial cells. Furthermore, this effect was seen regardless of cell-type specificity. **C,** Heatmap showing the relationship between 5' UTR GC content, 5'UTR length and CDS TE. Very high and very low GC content are associated with lower median TE. As the length of the 5'UTR increases, the median TE of the CDS decreases. * represents at $p \leq .05$, ** represents $p \leq 0.01$ and *** represents $P \leq 0.001$. **D,** The relationships between 5'UTR length and GC content are independently plotted against median TE for each length or GC-content bin. Significance is denoted the same as in Fig. 3.4c.

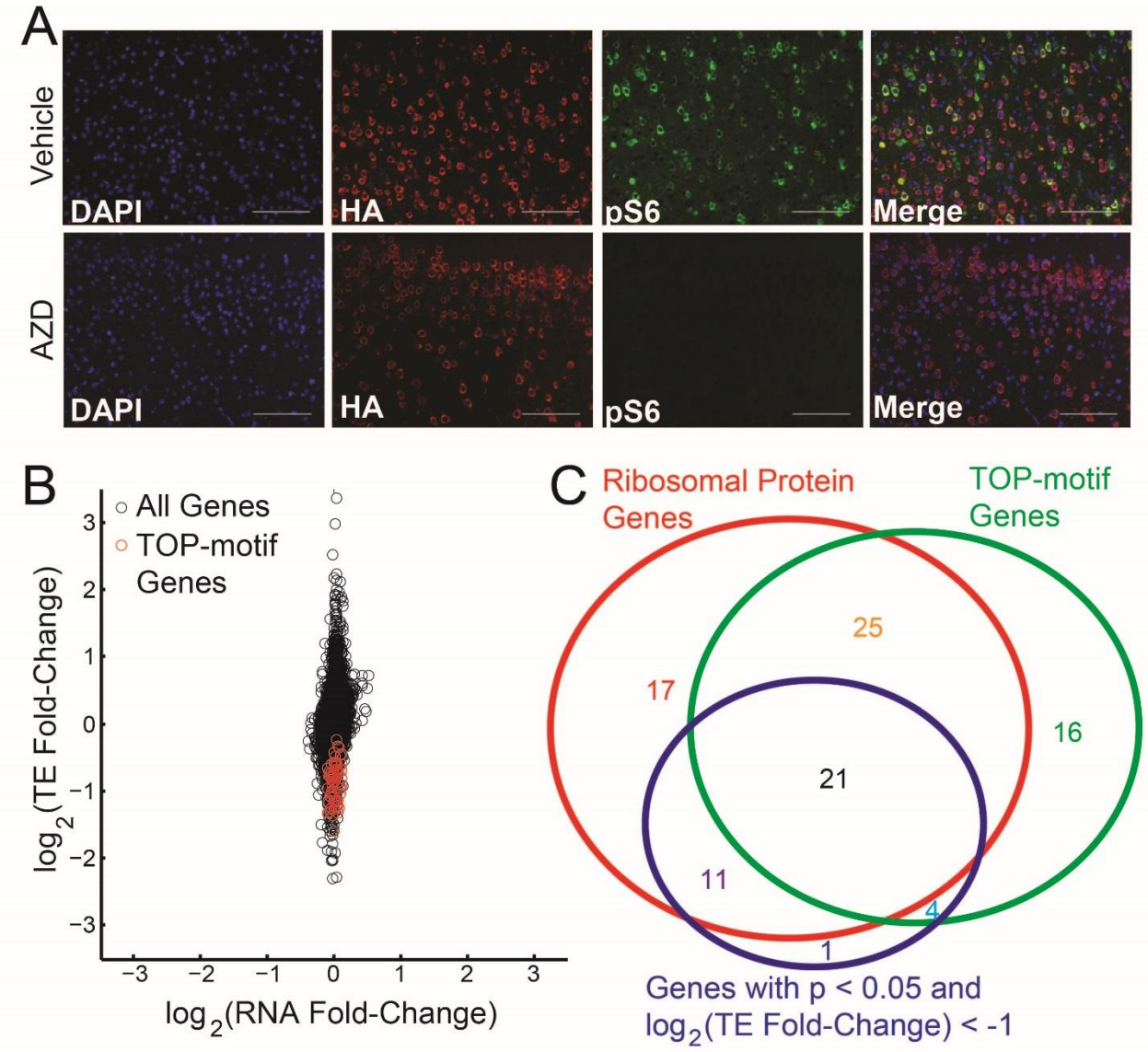

Figure 3.5 mTOR Controls TOP-Motif Containing Genes in the Brain

A, Treatment for one hour with AZD-8055 was sufficient to drastically decrease levels of phosphorylated Rps6 in mouse brains. HA-staining indicates the presence of HA-tagged Rpl22 (RiboTag) in cells expressing Camk2a. **B,** Comparison of RNA and TE fold-changes between AZD-8055-treated and untreated mice. TE exhibits larger amplitude changes that RNA levels in response to mTOR inhibition in the brain. The TE of TOP-motif containing genes are greatly reduced. **C,** Genes with significant differential translation efficiency were identified with

RiboDiff and genes with significant differential RNA expression were identified with DESeq2. The Venn diagram shows the overlap between genes with significant translational reduction after AZD-8055 treatment, ribosomal proteins, and TOP-motif containing genes.

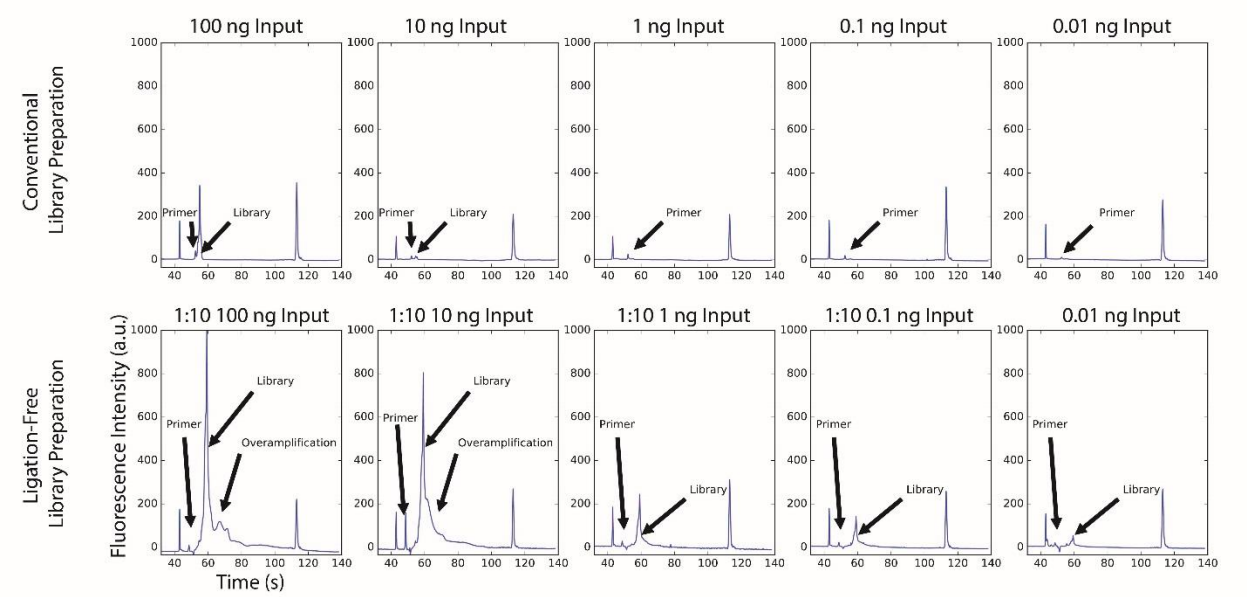

Figure 3.6 Sensitivity of Conventional and Ligation-Free Strategies

Ligation-free and conventional libraries were generated from a serially diluted 34-base RNA oligonucleotide and analyzed via Bioanalyzer following an equal number of PCR cycles for each library. All ligation-free library preparations except for the 0.01 ng sample were loaded onto the Bioanalyzer at a 1:10 dilution to avoid saturating the detector at high concentrations. Detectable libraries were successfully generated for all concentrations using the ligation-free method, but could only be generated using conventional methods for the 100 and 10 ng inputs.

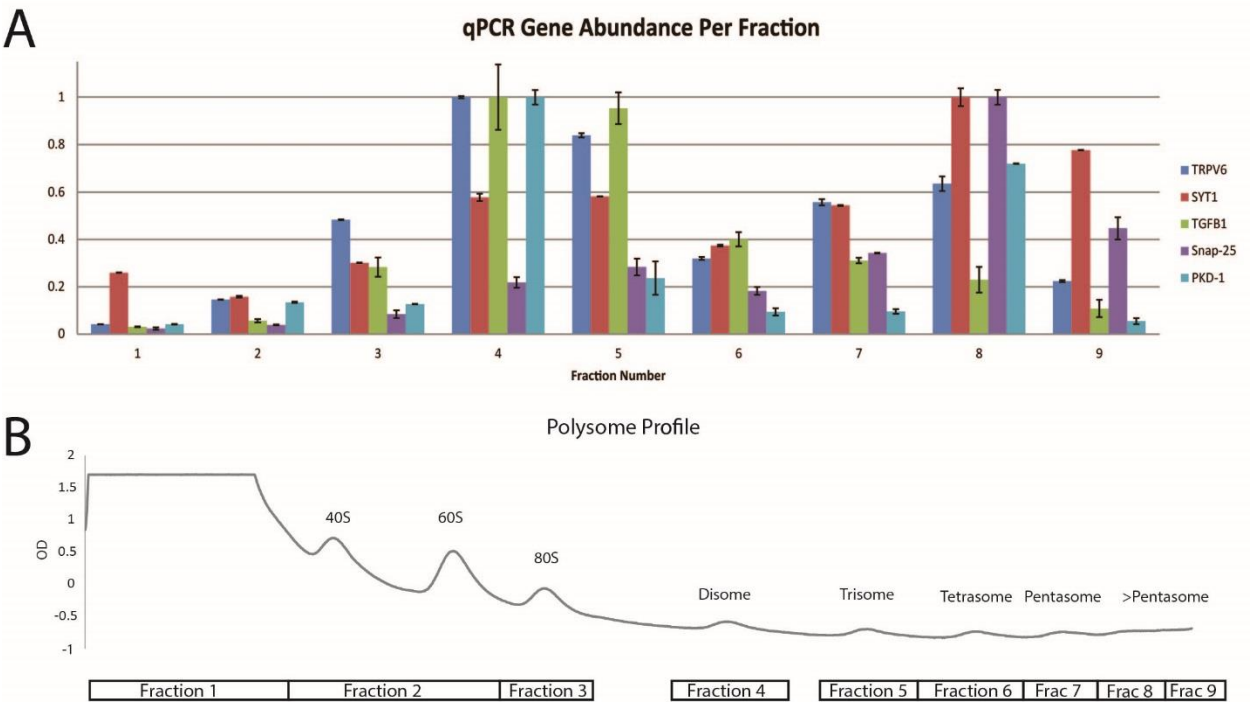

Figure 3.7 Highly Translated Genes are Shifted to Heavier Polysomes

qPCR was performed with 5 probes on fractions isolated from a polysome profile from left frontal lobe brain tissue. **A,** Genes found to be highly translated in ribosome profiling data, Snap-25 and Syt1, were found to be shifted to heavier polysomes; fractions 8 and 9. Genes found to be lowly translated, Tgfb1, Trpv6, and Pkd-1 were found to be most concentrated in lighter polysomes, fractions 4 and 5. **B,** The polysome profile denotes from which portion of the profile fractions were obtained.

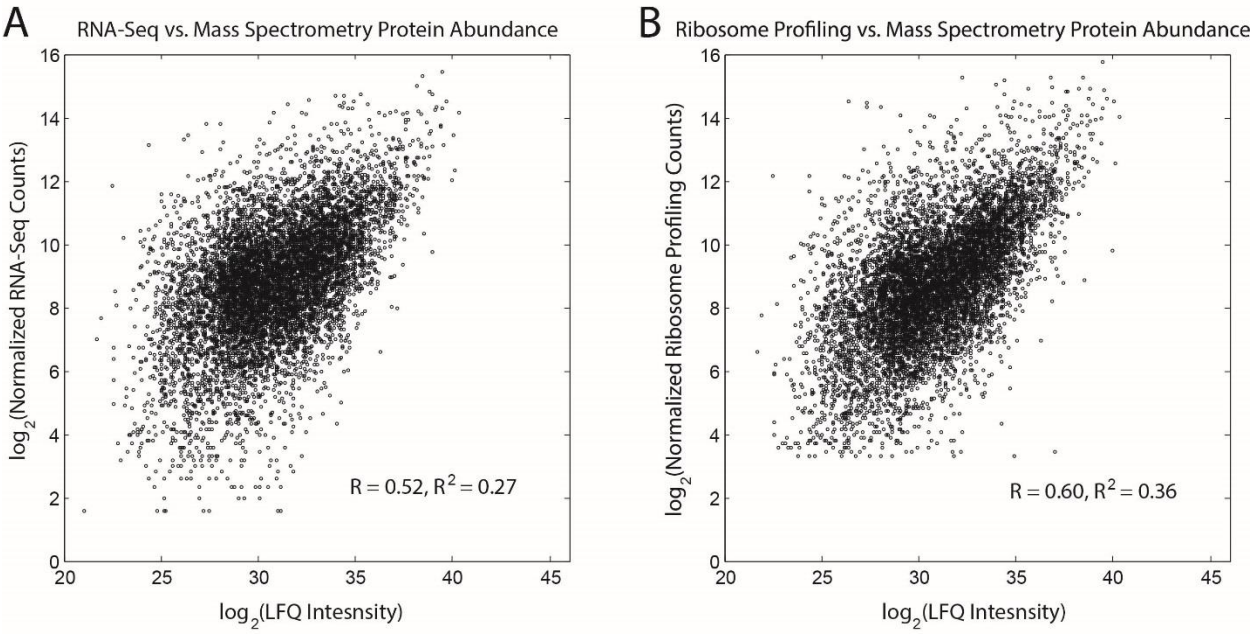

Figure 3.8 Comparison of Ligation-Free Ribosome Profiling and RNA-Seq to Protein Abundances Measured by Mass Spectrometry

RNA-Seq and ligation-free ribosome profiling data from this experiment were plotted against proteomics data from a mouse of the same age and similar background. **A,** RNA-Seq data plotted against whole brain mass spectrometry protein abundance are correlated with r = 0.52 and r^2 = 0.27. **B,** Ligation-free ribosome profiling data plotted against whole brain mass spectrometry protein abundance are better correlated than in a. with r = 0.60 and r^2 = 0.36.

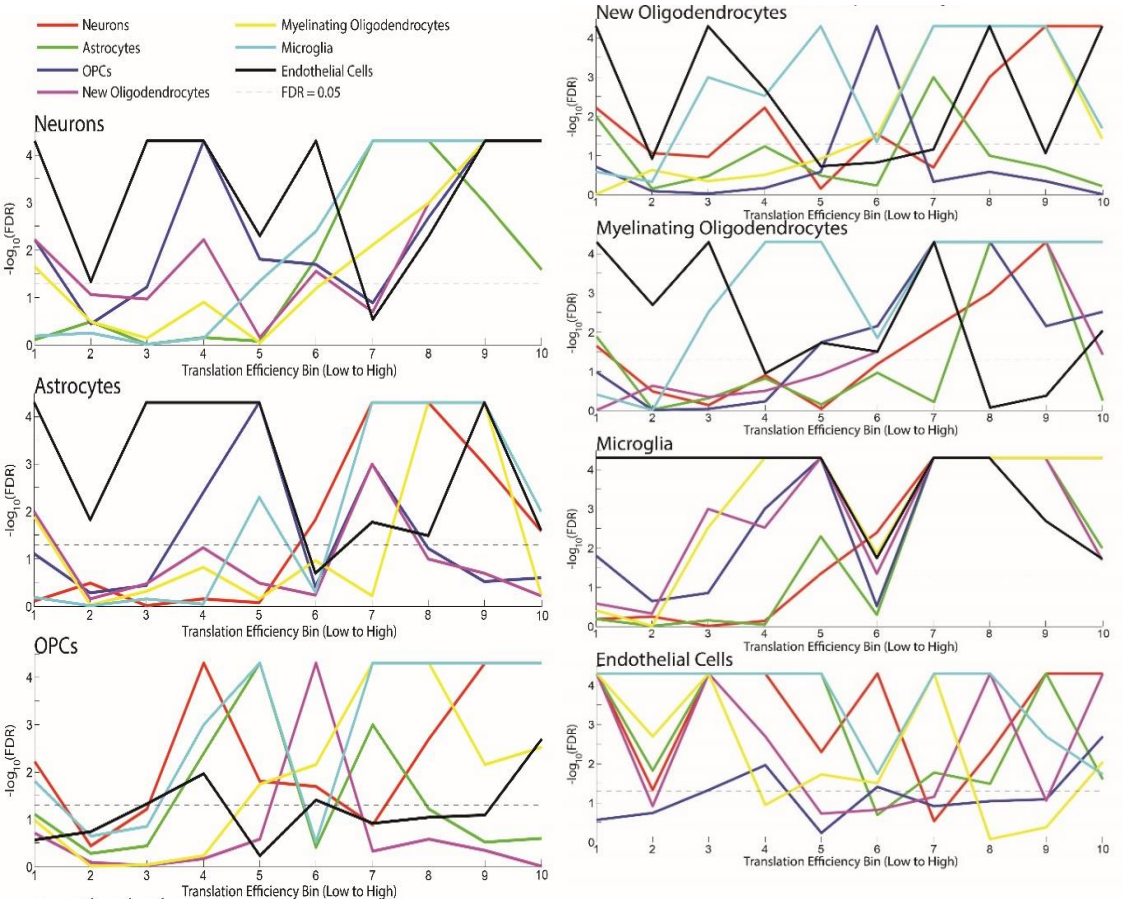

Figure 3.9 Statistics for Cell-Type Specific Translation

FDR-corrected p-values for pairwise comparisons of each cell type at each TE bin for the heatmaps shown in Figure 3.2B computed by GSEA.

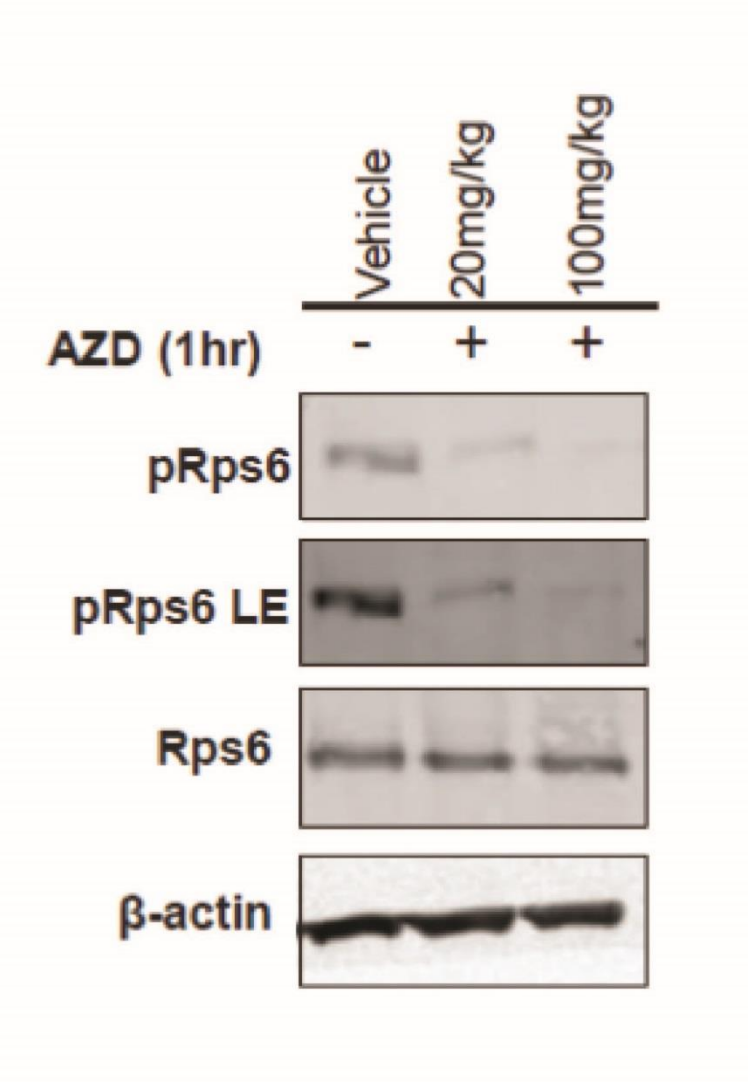

Figure 3.10 Western Blot Analysis of AZD-Treated Mouse Brain

We sacrificed mice one hour after oral administration of AZD-8055 and performed Western blot analysis on homogenized brain tissue. Administration of AZD-8055 in a Camk2a-RiboTag mouse decreases mTOR activity as detected by phosphorylation of Rps6. Phosphorylated Rps6 levels were compared to Rps6 and β-actin levels for vehicle, 20 mg/kg AZD-8055, and 100 mg/kg AZD-8055 treatments (long exposure=LE).

CHAPTER 4

RiboPLATE-Seq: High Throughput Translational Profiling

4.1 Background

Abstract

In cellular systems, protein expression is often dynamically regulated by translation. As the final arbiter of what proteins are expressed, translational control is achieved through the action of a complex network of protein kinases responsive to a host of intra-cellular and extra-cellular conditions. While previous experiments have demonstrated the effects of a handful of protein kinases on gene-specific translational control, to date, there has not been an un-biased and systematic effort to generate a network of translational control. Application of systems biology to similar scientific questions has yielded a number of important insights, such as in the study of transcriptional control where large-scale perturbation studies have allowed for the development of transcription factor networks. In order to assess translation in a similar manner, a technology designed to measure cellular translation states on a genome-wide scale and in the context of large-scale perturbation studies is required. In this chapter I discuss the development of a technique which enables the parallel measurements of translation across a large number of independent samples following compound-induced perturbation. This technology was applied to determine the effects of protein kinase inhibitors in addition to compounds targeting upstream effectors of translation. Our findings demonstrate that translational responses to perturbation can be rapidly assayed in a high-throughput format, facilitating a systems-biology approach to studying translation.

Introduction

In order to maintain their function, cells are often required to respond to physiologic stressors by altering protein expression. Although transcription-level changes can be used to modulate protein output, more rapid alterations can be achieved by adjusting the translation efficiency of specific genes (Andreev et al. 2015). However, in contrast to transcription, the regulatory networks which control translation are poorly understood. It has been demonstrated that translational control can be achieved through the action of protein kinases which can signal to factors that directly interact with cis-regulatory elements of mRNA. However, technologies which could be employed to catalog the genome-wide translational effects of the 500 member family of protein kinases are limited. While there has been recent success in developing genome-wide transcriptional networks based on large numbers of single-gene knockout experiments, these perturbation-based approaches have not been applied towards the development of a gene-specific translational network, in part, due to technological limitations (Kemmeren 2014). Recent advances, such as ribosome profiling, have enabled new insights to be made into translational control but due to high cost and hand-on time, lack the throughput to be utilized in large-scale studies (Ignolia et al. 2009). Although improvements to ribosome profiling, like the recently reported ligation-free approach, have increased the sensitivity and reduced the time and cost necessary to generate translational data, these techniques are still insufficient for applications demanding hundreds of independent experiments (Hornstein et al. 2016). Due to the importance of translation to human physiology and disease, as well as the targetable nature of protein kinases, a high-throughput technology is necessary to understand the role of translational control on a gene-specific level.

High-throughput measurement of translational states is a challenging endeavor for several reasons. Cost, effort and time all scale proportionately with the number of libraries generated in conventional library construction strategies. In applications demanding large numbers of samples, these factors can quickly overwhelm both capacity and budget. Additionally, specific to the case of quantifying translation, a means of identifying and isolating translating mRNA from the total pool of both translated and un-translated mRNA is required.

In order to address difficulties associated with the generation of large numbers of sequencing libraries we employed a soon to be reported highly multiplexed library preparation strategy, Pooled Library Amplification for Transcriptome Expression (PLATE-Seq), to inexpensively and rapidly generate sequencing libraries. PLATE-Seq allows for multiplexed library preparation due to the addition of unique well-specific barcodes during reverse transcription, followed by pooling in later library preparation steps. These methods greatly reduce the cost and effort associated with large-scale library preparation.

In conventional ribosome profiling and previous polysome profiling experiments, ribosome-associated mRNAs were isolated by sucrose fractionation. However, sucrose fractionation is a time-consuming and difficult process which inherently limits that number of samples which can be processed due to rotor capacity. We chose to bypass sucrose fractionation and utilized immunoprecipictation of ribosomal RNA (rRNA) to enrich for ribosome-associated materials. Libraries were generated from samples enriched in ribosome-associated mRNAs, as well as the homogenate they were derived from so that changes at the level of translation could be assessed.

In order to test the ability for this technique (Ribo-PLATE-Seq) to detect differentially translationally regulated genes, we chose to apply a panel of four compounds, two of which are

known to decrease the translation of the well-characterized terminal oligo-pyrimidine (TOP) motif containing family of genes. We performed a drug-perturbation assay on 96 independent samples with 6 replicates for each drug and 72 control replicates. The compounds we used were comprised of two which were known to repress translation of TOP-motif containing gene, one known not to affect translation of TOP-motif containing genes, and a final compound which has not had its translational targets assessed in a genome-wide manner, but has been shown to affect the phosphorylation of eIF4B and S6, two known markers of cap-dependent translation. After generating libraries from these samples and sequencing, a process which is now automated, we found that we were able to detect translational perturbations induced by the two compounds known to repress translation of the TOP-motif containing genes, as well as the third compound which had been shown to affect cap-dependent translation.

4.2 Results

High-Throughput Library Construction

Large-scale application of translational profiling technologies have been hampered by two main limitations. The first is due to an inability to generate multiple ribosome profiling libraries in parallel; library preparation protocols are generally expensive and time-consuming and costs scale linearly with increased sample number. Additionally, ribosome-associated mRNA identification and isolation has conventionally been achieved by use of sucrose fractionation. This method is not amenable to processing more than several samples at a time and severely limits the throughput of approaches designed to study translation. In order to achieve the high-throughput necessary to perform large-scale genomic assays, we have utilized an

enhanced library preparation process amenable to sample pooling and an alternative method to enrich for ribosome-bound mRNAs.

A high-throughput library preparation method for the generation of RNA-Sequencing libraries known as PLATE-Seq was recently developed by members of the Sims and Califano labs. Key to PLATE-Seq is the incorporation of unique sample-specific barcodes during reverse transcription. By adding a unique identifier to libraries generated from individual samples, all steps following reverse transcription can be performed on a pooled library—greatly decreasing the hands-on time and cost associated with large numbers of library preparations. Figure 4.1A illustrates the PLATE-Seq workflow used to measure RNA abundance. Briefly, following cell lysis, samples are transferred to a new multi-well plate with oligo(dT) covalently bound to its surface. Following washing and elution, mRNAs are transferred to a third multi-well plate where reverse transcription is performed with barcoded, adapter-linked oligo(dT) primers. Reverse transcription is followed by exonuclease digestion of single-stranded reverse transcription primers. cDNA products from reverse transcription are then pooled and second-strand synthesis performed with Klenow Large Fragment with adapter-linked random primers. Klenow large fragment's lack of strand-displacement activity yields, at most, a single second-strand synthesis product which contains a sample barcode as well as 5' and 3' PCR adapters (Figure 4.1B). The pooled second-strand synthesis products are enriched in a single PCR reaction and the resultant library sequenced at 2-4 million reads per sample.

The performance of PLATE-Seq was characterized in a fully automated 96-well screen profiling the effects of seven small-molecule perturbations in the BT20 breast cancer cell line. Due to the high-capacity of the system, 12 replicates were generated per condition. PLATE-Seq was compared against the Illumina TruSeq protocol which utilizes a conventional RNA-Seq

library preparation scheme. Due to the input requirements of TruSeq, pooling of material from replicate wells was necessary to generate enough material for library generation. Figures 4.2A shows the distribution of unique genes detected and uniquely mapped reads generated from PLATE-Seq and TruSeq preparation methods. Figures 4.2B show the saturation profile of the aggregate PLATE-Seq data pools as compared to the TruSeq replicate pools. Greater than 75% of the genes detected in the TruSeq replicate pool were detected in the six-well PLATE-Seq data pool, although 6-fold fewer reads were required to achieve these results.

A critical application of perturbation studies is the identification of differentially expressed genes. We sought to determine the similarity between differentially expressed genes identified by PLATE-Seq and TruSeq library preparation methods. DESeq2 was used to identify differentially expressed genes between control replicates generated from PLATE-Seq and compound treated conditions from both library preparation strategies. Figure 4.3 shows the set of differentially expressed genes ($P < 0.05$) identified in both library preparation strategies and projected as a matrix of fold-changes using multidimensional scaling. There was strong overlap of individual drug replicates regardless of the library preparation methodology. Furthermore, drugs with similar mechanisms of action, such as the topoisomerase II inhibitors Idarubicin and Mitoxantrone clustered similarly. These results show that PLATE-Seq and conventional RNA-Seq library preparation methodologies generate comparable gene expression results for the same or similar drugs.

Immunoprecipitation of Ribosome-Bound mRNA

Isolation of ribosome-associated mRNAs is a critical step in obtaining information regarding cellular translational state. In ribosome profiling, this process is performed by sucrose

fractionation following RNase digestion. However, fractionation is a time-consuming process and is inherently limited by the number of samples able to fit in an ultracentrifugation rotor. Furthermore, removal of ribosomal footprints from fractionated material requires several additional steps such as gel-purification which are not amenable to high-throughput applications. Bypassing the complicated step of sucrose fractionation, we enrich for ribosome-associated mRNAs through an immunoprecipitation assay (Figure 4.4). Our experimental paradigm begins following a perturbation screen performed in multi-well plates where samples are lysed in a buffer containing magnesium and cycloheximide designed to preserve the ribosome-mRNA interaction and halt elongation. Lysates are then divided between two plates, with one subjected to PLATE-Seq without modification and the other undergoing pan-ribosomal immunoprecipitation followed by PLATE-Seq. The process of RiboPLATE-Seq begins with pan-ribosomal immunoprecipitation and is performed by exposing lysate in a multi-well plate to biotinylated antibody targeting y10b, an rRNA component of the 60S subunit (Figure 4.5A). Following exposure, streptavidin coated magnetic beads are used to capture the mRNA-ribosome-antibody hybrid. Free mRNA is removed by multiple washes and elution of ribosome-associated mRNA achieved by gentle disruption of the ribosome-mRNA interaction.

In order to test and improve the efficiency of our immunoprecipitation we developed a qPCR-based assay designed to measure the enrichment of ribosome-associated mRNA. In order to determine the depletion of non-ribosome-bound material following immunoprecipitation we added a set of mRNAs unbound to ribosomes (ERCC) to fresh cellular lysate. Following immunoprecipitation, we performed reverse transcription of eluted material using random primers. cDNA was also generated from reserved cellular lysate unexposed to antibody or beads. The gene B2M was chosen to serve as an indicator of enrichment of ribosome-associated

mRNAs due to its known average level of association with ribosomes determined from previous ribosome profiling experiments. qPCR was used to quantify the amounts of B2M and ERCC found in cDNA generated from both the lysate and the eluent following immunoprecipitation (Figure 4.5B). Based on these results, we calculated a depletion of ERCC transcripts of ~ eight-fold.

High-Throughput Identification of Translation Targets

We treated WI-38's, a cell line derived from human lung fibroblasts, with a panel of four drugs to test the ability of Ribo-Plate-Seq to identify differentially translated genes. Several of the drugs chosen, AZD8055, PP242, and MNK-I1 have been previously investigated with ribosome profiling; AZD8055 and PP242 are known to repress translation of Terminal Oligopyrimidine-motif containing transcripts (TOP) while MNK-I1 does not significantly affect their transcription or translation. (Figure 4.6) (Hsieh et al. 2012; Hornstein et al. 2016;Unpublished Sims Lab Data). BKM120 has not been investigated with ribosome profiling, but has been shown to decrease phosphorylation of key-markers of cap-dependent translation including S6 and eIF4B (Serra et al. 2013). Based on these previous observations, we anticipated that we would detect AZD8055, PP242, and BKM120 significantly decreased the translation efficiency of TOP-motif containing genes, while MNK-I1 would not significantly alter the transcription of translation of TOP-motif containing genes. In order to assess the level of depletion of mRNAs not associated with ribosomes in our immunoprecipitated samples, we added ERCC spike-in to half our samples. The experiment was carried out in a 96-well plate with 6 replicates for each drug treatment. Following footprint isolation and library generation, we sequenced the Ribo-PLATE-Seq and PLATE-Seq libraries on the Illumina NextSeq platform

to an average depth of 750,000 and 1,000,000 pass filtered reads, respectively. We quantified the total amount of ERCC detected in all samples and compared the ratio of ERCC detected in Ribo-PLATE-Seq with the amount detected in the PLATE-Seq library generated from the same material (Figure 4.7A). Our depletion of ERCC was ~8-fold, similar to the level of depletion we calculated based on our preliminary qPCR experiments.

In order to determine translationally altered drug targets, we utilized a statistical package, Analysis of Translational Activity (anota), specifically designed to detect differential translation in perturbation studies. Differing from many other statistical tools used to quantify differential translation, anota takes into account the spurious correlation created when comparing mRNA pools to translationally active mRNA on a log scale by performing regression analysis between translationally active mRNA levels and total mRNA levels. Furthermore, by applying linear regression to genes individually, it is better able to statistically handle the drop-out of genes seen between samples in Ribo-PLATE-Seq. Using anota to calculate changes in total mRNA, ribosome-bound mRNA and translation efficiency between control and drug-treated samples, we found that as expected, AZD8055, PP242, and BKM120 treated samples showed significantly reduced ribosome-bound mRNA of TOP-motif containing genes, but did not have a significant change in their total mRNA levels (Figure 4.7B). MNK-I1 did not show significant alteration of either ribosome-bound mRNA or total mRNA for TOP genes. These results were in agreement with previously performed ribosome profiling experiments involving MNK-I1, AZD8055 and PP242, as well as molecular observations for BKM120 which had not been investigated with ribosome profiling.

4.3 Discussion

In this chapter we have demonstrated a new approach to determining cellular translational states in a low-cost, high-throughput and highly scalable platform. This approach relies on a multiplexed barcoding library generation scheme and pan-ribosomal immunoprecipitation of ribosome-associated mRNAs. Using this strategy we generated translational data from 96 independent samples after applying a four-compound panel comprised of drugs known to target key translational mediators and previously investigated with traditional ribosome profiling. We identified differentially translated genes based on anota analysis and found that we were able to detect a significant decrease in translation of the TOP-motif containing family of genes for all compounds which have been previously shown to decrease their translation.

While Ribo-PLATE-Seq offers several advantages to traditional ribosome profiling, it is important to recognize its limitations. Because traditional ribosome profiling libraries are generated from ribosomal footprints, the data generated produce measurements that contain information of the direct genomic position of the ribosome at the time of cellular lysis, as well as a more accurate quantification of the translation efficiency of a gene. However, in the context of perturbation studies, we have demonstrated that Ribo-PLATE-Seq is able to accurately detect previously identified translational alteration for a greatly reduced cost and time investment as compared to traditional ribosome profiling methods.

4.4 Conclusion

Based on the above results, we have demonstrated that Ribo-PLATE-Seq can accurately identify translational perturbations in a genome-wide, inexpensive, and high-throughput manner. We hope that this high-throughput platform will further enable the study of functional genomics and allow for large-scale systems-biology approaches to translational regulation.

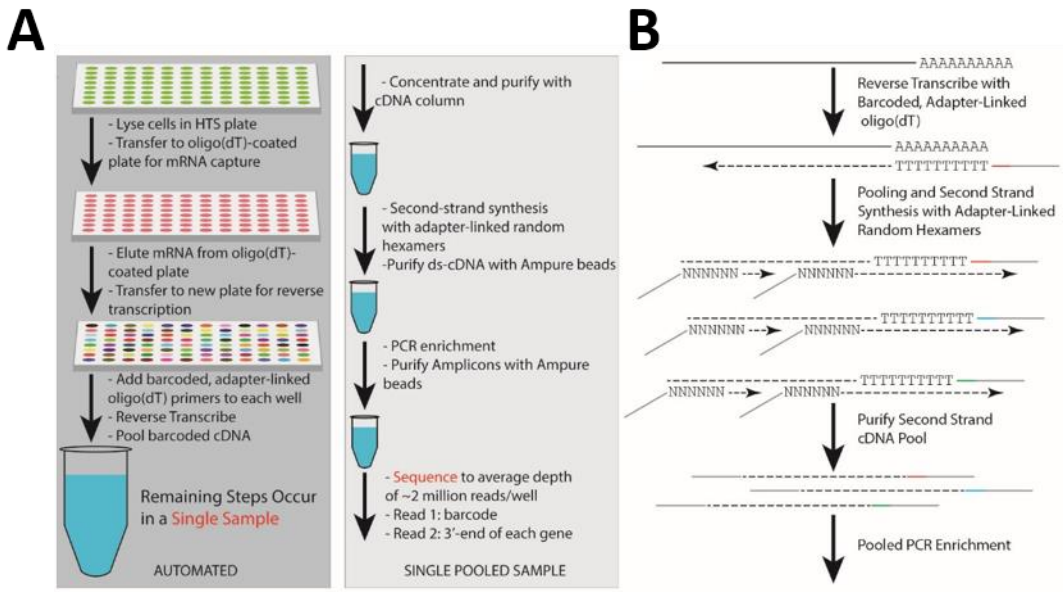

Figure 4.1 Overview PLATE-Seq

A, Following lysis of samples in a multi-well plate, oligo(dT)-coated capture plates are used to purify mRNA from lysate. Following elution from the capture plate, mRNA are transferred to a new plate and reverse transcription performed with barcoded adapter-linked oligo(dT) primers unique to each well. Following reverse transcription, ExoI digestion of free ssDNA primers is necessary to reduce cross-talk between sample wells. Samples are then pooled into a single tube allowing for the remaining second-strand synthesis, PCR enrichment and cleanup steps to occur on the pooled samples. **B,** A schematic showing the molecular steps of library construction.

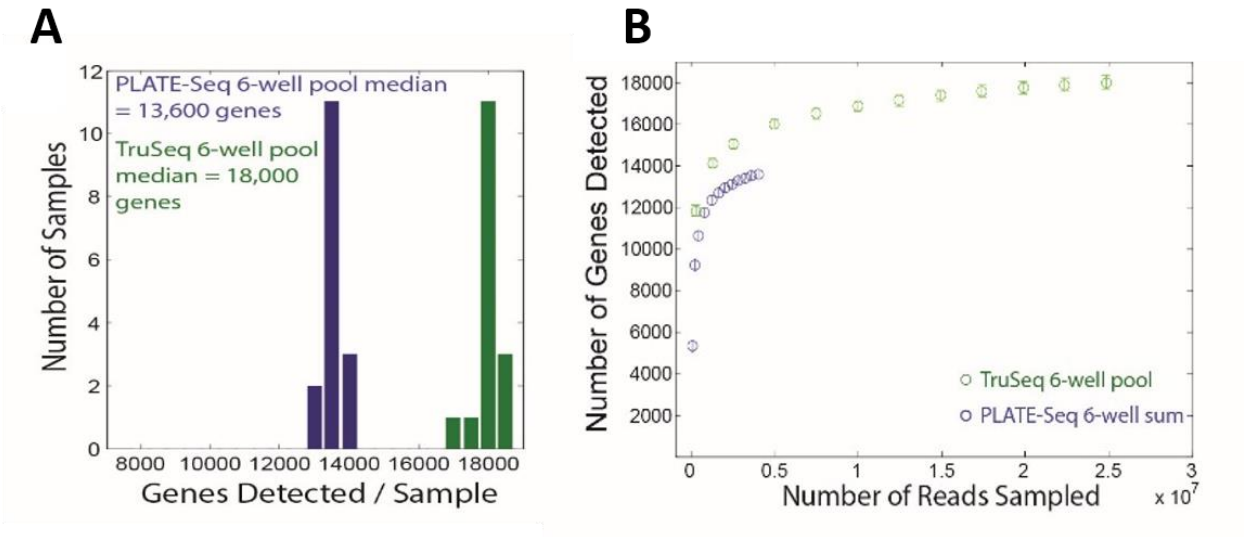

Figure 4.2 Comparison of PLATE-Seq and TruSeq Library Preparations

A, Comparison of the number of unique genes detected when using PLATE-Seq and TRuSeq library preparation methods. **B,** Downsampling analysis demonstrates the number of unique genes detected as a function of reads sampled for TruSeq and PLATE-Seq library preparation methods. Although libraries generated by PLATE-Seq were sequenced to a fraction of the depth of the TruSeq libraries, ~75% of the genes identified by TruSeq were detected by PLATE-Seq.

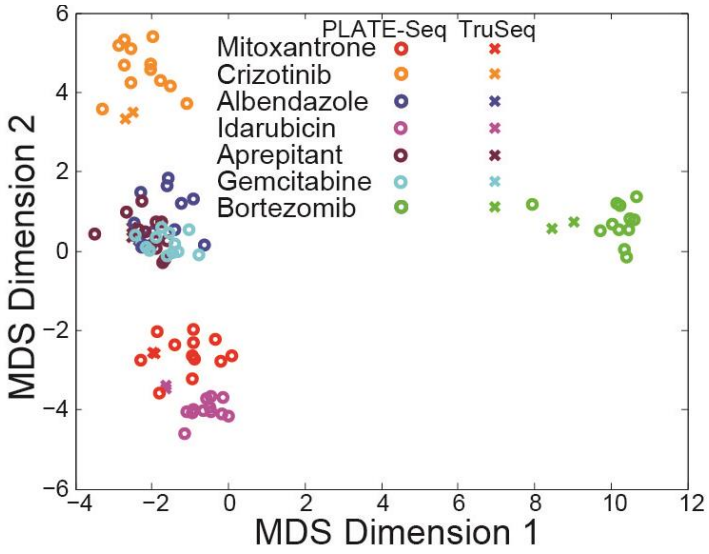

Figure 4.3 Differentially Expressed Genes Detected By PLATE-Seq and TruSeq

MDS clustering of differentially expressed genes identified by PLATE-Seq and TruSeq show that both methods identify similar sets of genes for the same or closely related compounds.

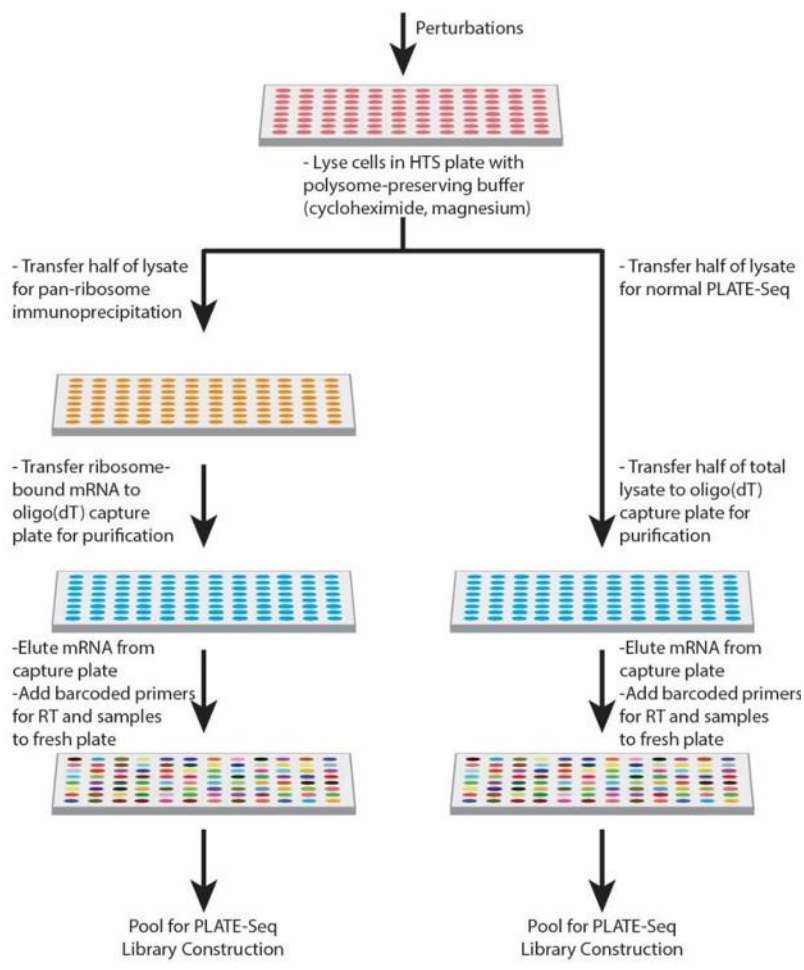

Figure 4.4 Overview Ribo-PLATE-Seq

A workflow overviews the process of generating Ribo-PLATE-Seq and PLATE-Seq sequencing libraries. The main modifications include altering the cellular lysis before to maintain the ribosome-mRNA interaction at the time of lysis, and performing a pan-ribosomal immunoprecipitation on at least half of the lysate.

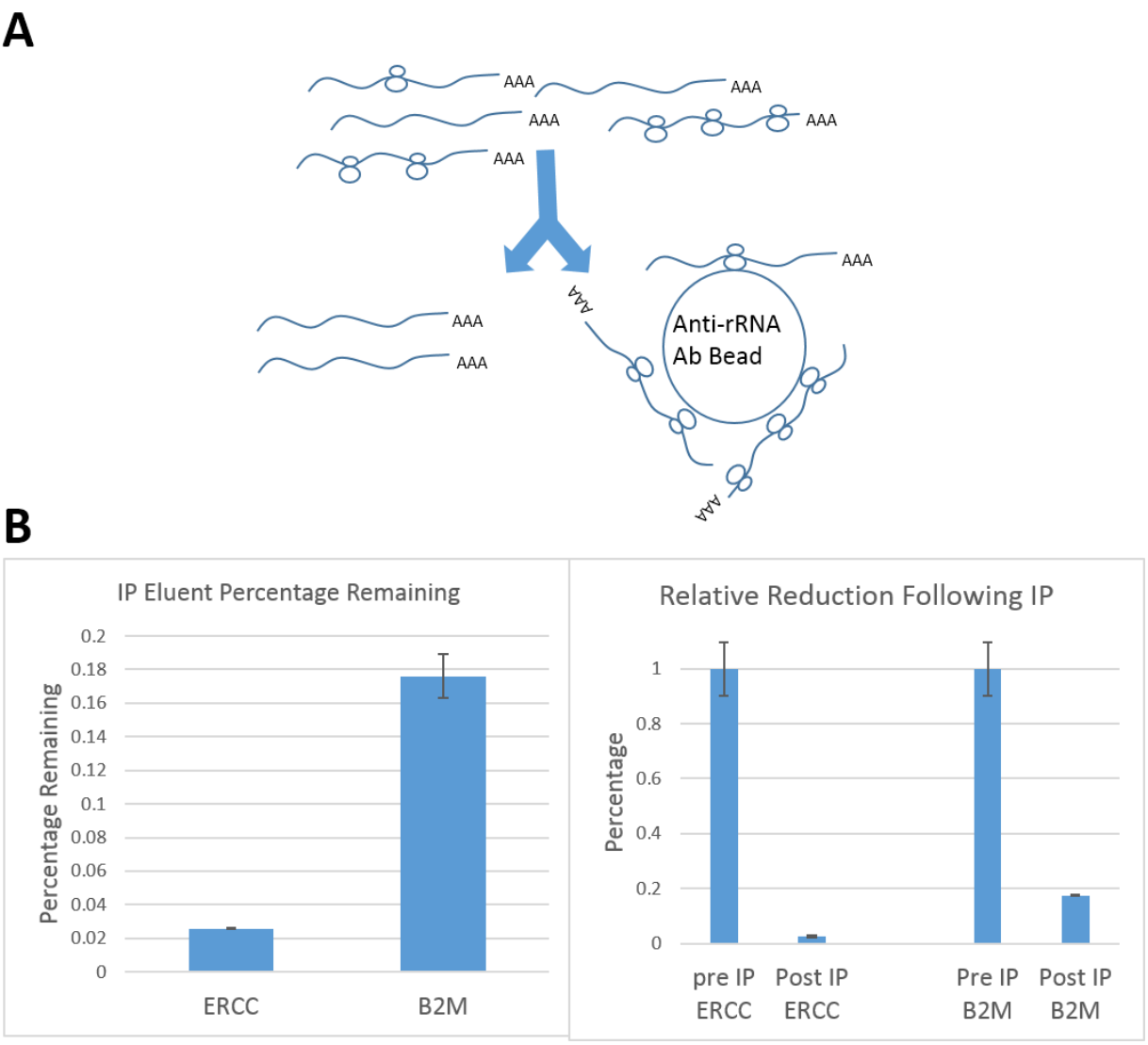

Figure 4.5 Ribo-PLATE-Seq Immunoprecipitation

A, Demonstration of the immunoprecipitation of ribosome-associated mRNAs. Cells are lysed in a buffer containing cycloheximide and magnesium which maintains the mRNA-ribosome interaction. Lysates are then exposed to biotinylated antibody targeting the ribosomal RNA y10b. Following incubation, magnetic streptavidin coated beads bind the mRNA-ribosome-antibody complex through washing steps until the mRNA-ribosome interaction is gently disrupted with the addition of EDTA. **B,** Preliminary qPCR experiments designed to test the performance of

ribosome bound immunoprecipitation. ERCC, a set of known non-ribosome associated mRNAs, was added to cellular lysates and immunoprecipitation carried out. qPCR was performed with primers against ERCC and B2M, a gene previously identified to be translated at an average level. Following IP, ~2% of ERCC remained in the eluent as compared to the original lysate, and ~18% B2M remained. Based on this, we calculated a 8-fold relative enrichment of ribosome-associated transcripts.

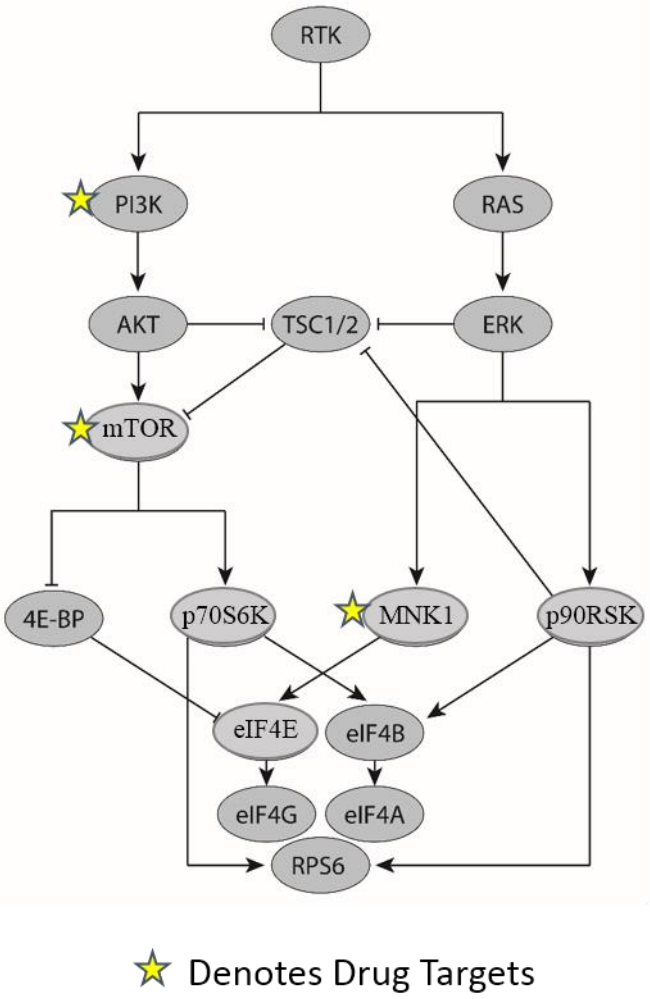

Figure 4.6 Molecular Targets of Compounds Profiled

A, Overview of molecular pathways converging on translational machinery displays the molecular targets of the 4-compound panel used to test for differential translation using Ribo-PLATE-Seq. AZD8055 and PP242 directly target the mTOR complex, reducing translation of TOP-motif containing transcripts. BKM120 is a potent PI3K inhibitor and has similarly been shown at affect the translation of TOP-motif containing transcripts. Although MNK-1 targets eIF4E and by extension eIF4G, MNK-1I has not been shown to affect the translation of TOP-motif containing transcripts.

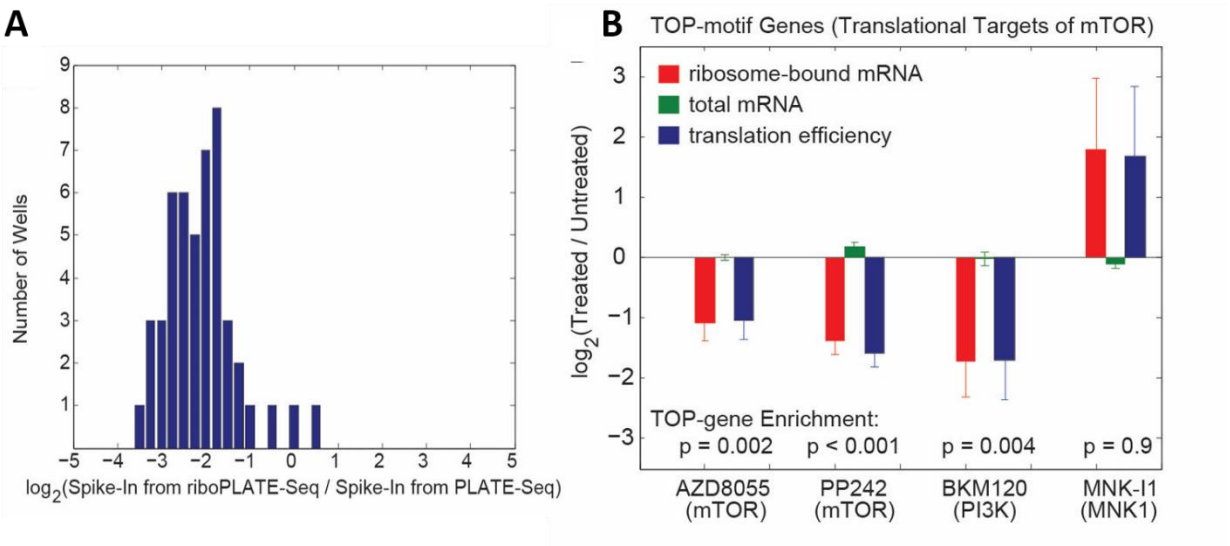

Figure 4.7 Ribo-PLATE-Seq ERCC Depletion and Detection of TOP-Motif Containing Genes

A, Sequencing libraries were generated from samples with and without ribosomal immunoprecipitation. ERCC was included in half of sample wells to serve as a metric for non-ribosome associated mRNA depletion. Comparing libraries generated from samples containing ERCC spike-in, ERCC was depleted in libraries generated from immunoprecipitated material by an average of ~8-fold. **B,** AZD8055, PP242, and BKM120 have been previously shown to decrease the translation of TOP-motif containing genes. Based on differential expression analysis performed with anota between treated and untreated samples, we detected a significant decrease in the translation of TOP-motif containing genes following administration of AZD8055, PP242 and BKM120, but not for MNK-I1 which has not been previously shown to affect the translation of TOP genes.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

In this work we've discussed three recent technology development efforts and applied them to make new insights into translation. Chapter 2 focused on the generation of a mouse model of glioma which allowed for the profiling of transformed cellular populations lacking conventional cellular markers. We used this system to discover translational alterations in glioma by application of conventional ribosome profiling. These efforts demonstrated that the translational landscape of glioma is vastly different as compared to un-transformed cells, as well as the role of the 5' untranslated regions in translation control. Furthermore, we demonstrated that ribosome profiling, after some modifications, could be utilized in mammalian tissue. Stemming from experiences we had in generating translational data for the work discussed in Chapter 2, we sought to develop a technique which would reduce the time, input requirements and cost associated with ribosome profiling. Chapter 3 revolved around these efforts which emerged as ligation-free ribosome profiling, a technique which reduces the generation of ribosome profiling libraries to a single day, with input requirements as low as 1ng and greatly reduced cost as compared to conventional ribosome profiling. This technique was used to investigate cell-type specific translation in the brain and to identify translational targets of mTOR inhibition in excitatory neurons. Finally, Chapter 4 focused on the development of a technique designed to rapidly obtain translational data from perturbation studies in a high-

throughput manner in less than 24 hours for ~ 20 dollars a sample. As a cohesive work, these efforts demonstrate a series of improvements to the technology currently used to assay translation and the generation of new insights into translational control.

Global Translational Perturbations in Glioma

Application of ribosome profiling to the specific population of transformed cells in our glioma mouse model demonstrated a number of interesting findings. Due to the large increase in translation observed in many cancers, we were surprised to find that there was a global reduction in translation efficiency in transformed cells in our glioma model. Agreeing with conventional wisdom, although translation efficiency was reduced, global translation was increased, as evidenced by translational upregulation of key ribosomal proteins. This mismatch between translation efficiency and translation could be explained in several ways. Firstly, while ribosomal machinery and activity may be increased in glioma, greatly enhanced production of mRNA may overwhelm the translational machinery, resulting in a lower than expected translation efficiency. Another possible explanation is that immature cellular states may be aided and maintained by reduced translation efficiency. Regardless of the cause, our results regarding the altered translation state of glioma highlight the striking manner in which translation is altered in the disease and represent a promising therapeutic target. Based on our animal model's ability to accurately recapitulate proneural glioma, we hope that it will be continued to be utilized in efforts designed to test new anti-glioma therapeutics targeting translation such as newly developed non-allosteric mTOR inhibitors.

Ligation Free Ribosome Profiling and Cell-Type Specific Translation

Attempting to improve on the previously reported technique of ribosome profiling, we developed a new ligation-free approach which took advantage of an engineered RNA poly(A) polymerase and M-MulV reverse transcriptase's template switching activity. Incorporating these features into our library preparation scheme, we were able to avoid a number of expensive enzymatic and time-consuming purification steps, reducing the overall preparation time for ribosome profiling from two weeks to as a little as a day. Furthermore, these changes greatly decreased the input requirements of the technique, allowing successful generation of libraries from as little as 1 ng of input. Based on down-sampling analysis of libraries generated with similar amounts of mouse brain material using both conventional and ligation-free methods we also found that the ligation-free method detected a higher number of unique genes and a much higher number of unique footprints, indicative of higher library complexity. Using a CAMKII RiboTag mouse, we obtained translational profiling data from several mouse brains using this method and were able to demonstrate translational trends between different neural cell-types. We found that neuron-specific genes appeared to be split among populations with very high and very low translation efficiency, indicating strict translational control may be important for the function of these cells. Alterations in the translation of maturing oligodendrocytes was also interesting; oligodendrocyte progenitor-specific genes appeared to have high translation efficiency, while new oligodendrocytes specific genes had moderate translation efficiency, and myelinating oligodendrocytes appeared to be associated with low translation efficiency. We also found microglia-specific genes to have low translation efficiency in brain tissue from an animal without infection or disease burden. These results, which were recapitulated in separate animals, indicate that translational control may be an important feature of cellular identity.

Based on results we obtained when previously studying the relationship between CDS ribosomal density and the existence of upstream AUG in genes with and without 5'UTR density, we sought to see if this effect existed in cell-type specific genes as well. We found that while the proportion of genes with and without 5'UTR density did not drastically change between cell-type, genes containing a uAUG and 5'UTR ribosomal density had lower CDS translation efficiency compared to genes without a uAUG. This effect was consistent across multiple cell types and was significant for myelinating, microglial, and endothelial cells, although the effect was seen regardless of cell-type. We also identified a relationship between 5'UTR length, GC-fraction in the 5'UTR and translation efficiency where longer 5'UTRs with very high or very low GC content greatly diminished the translation efficiency of the CDS.

Additionally, we used our ligation free ribosome profiling data obtained from AZD8055 treated and untreated CAMKII RiboTag mice to demonstrate our ability to detect excitatory neuron-specific translational changes following mTOR inhibition in a perturbation assay. We believe that these results, combined with the reduced cost, time and effort associated with ligation-free ribosome profiling will lead to its continued use in the scientific community.

High-Throughput Translational Assaying

Translational control has emerged as an important regulator of cellular activity. However, determining translational output by genomic or proteomic methods have been expensive endeavors due to the lack of throughput of these systems. While recent advances in ribosome profiling library preparation methods, such as ligation-free ribosome profiling, have reduced the effort and input requirements of generating translational measurements, functional genomics applications are still limited. The development of a technique able to perform high-throughput

translational profiling of independent samples would greatly aid in our ability to dissect molecular regulators of translational control through application of large-scale perturbation studies. Such a technique would be crucial, as current technologies are impractical for individually studying the large number of potential translational mediators, such as the 500-member family of protein kinases. In Chapter 4 we explored the development of an experimental technique able to obtain translational data from large number of independent samples while minimizing investments of time and cost. High-throughput translational control assays are hampered by two main issues; the lack of a method for rapid and scalable isolation of ribosome-associated mRNAs, and the directly scaling time and cost associated with the preparation of large numbers of sequencing libraries. We have addressed the latter issue by the application of a recently developed technique, PLATE-Seq. PLATE-Seq introduces sample-specific barcodes during reverse transcription which allows for sample pooling in later steps and can generate large numbers of samples in a single day for a cost under 20 dollars per sample. Bypassing sucrose fractionation, we isolated ribosome associated mRNAs through immunoprecipitation of the ribosome-mRNA complex by targeting a ribosomal RNA found in the 60S subunit followed by gentle dissociation of mRNA from the ribosome. Based on preliminary experiments, we found that the immunoprecipitation method employed had an 8-fold enrichment of ribosome-associated mRNAs. As a test of RiboPLATE-Seq's ability to detect translationally altered genes due to perturbation, we assessed a four-panel compound screen with RiboPLATE-Seq. We were able to recapitulate known translational targets previously identified by ribosome profiling for several compounds, as well as characterize translational targets of a compound not previously interrogated with ribosome profiling which were in agreement with its known molecular effects.

Based on these results we believe that RiboPLATE-Seq represents a promising new technology which will be used to uncover mechanisms of translational control.

5.2 Future Directions

In Chapter 2 we obtained some interesting results regarding the decrease of translation efficiency in the transformed cells of a glioma mouse model. While we are uncertain of the underlying cause of this decrease, it may indicate that the translational machinery in this population of transformed cells is unable to keep up with increases in transcriptional activity. It has been speculated, and in some cases, demonstrated that targeting translational machinery is an effective anti-cancer therapy. The results we've obtained support this hypothesis; testing the effects of known mTOR or PI3K inhibitors in the animal model we've generated would directly demonstrate the effects of translation inhibition in transformed glioma cells. Additionally, based on observations regarding ribosomal density in the 5'-leader region and in the CDS of both transformed and non-transformed cells, we hypothesized that ribosome density in the 5'-leader and the CDS are regulated by alternative pathways which are differentially regulated in tumor as compared to normal brain. In the future, it will be important to determine the role of this alternative pathway in translational control, and how its perturbation in transformed cells affects the overall disease.

While we demonstrated that ligation-free ribosome profiling was able to generate higher-complexity libraries with decreased input requirements, cost and time investment, compared to conventional ribosome profiling, there are still improvements which could be made. Specifically, we've found that the existence of difficult to purify byproduct of our library generation process

complicates efficient sequencing of these libraries. We believe that this contaminant is, in part, due to the large amount of rRNA which is not removed during our ribosomal RNA depletion. In order to address this issue, it will be important to test additional rRNA purification methods, such as removal of rRNA with RNA-based probes earlier in the library generation process. Additionally, we could alter our PCR-protocol to increase the ratio of product to contaminant. Improvements aside, we have transitioned from conventional ribosome profiling in the Sims lab and are currently using ligation-free ribosome profiling to obtain translational information for multiple projects.

The results we've obtained with our application of ligation-free ribosome profiling in Chapter 3 demonstrated that translation is controlled in a cell-type specific manner. However, the means by which cellular populations are able to independently control their translation is unclear. Further studies will be required in order to elucidate both the effects of tight-translational control and the means by which it is achieved. Additionally, we explored the rapid effects of AZD8055, a potent mTOR inhibitor, on CAMKII-expressing excitatory neurons. While we demonstrated that AZD8055 decreases translation of TOP-motif containing genes, a known effect of mTOR inhibitors, we did not see evidence of neuron-specific translational changes in response to mTOR inhibition. Due to mTOR inhibitors clinical use as an anti-seizure medication in patients with neural malignancies, we had hypothesized that excitatory neurons would have specific changes related to this anti-seizure effect. One possible explanation for why we didn't observe these changes may be the minimal treatment time of one hour that was used. In the future, we hope to repeat this experiment with an increased treatment window which may allow us to observe additional excitatory neuron-specific translational changes.

Chapter 4 focused on the development of a new tool able to rapidly and inexpensively obtain translational data from large numbers of samples. Work is ongoing to develop this technique and increase its accuracy and sensitivity. Generating sequencing libraries from immunoprecipitated material derived from small numbers of cells is a challenging endeavor. One issue that complicated our current analysis was the low complexity of RiboPLATE-Seq libraries. This issue lead caused a large amount of drop-out between samples and reduced the detection of unique genes in the RiboPLATE-Seq library as compared to the PLATE-Seq library. We hope to improve by increasing the ratio of lysate which is used to prepare the RiboPLATE-Seq and PLATE-Seq libraries. Additionally, the RiboPLATE-Seq library generation process was technically challenging due to the large number of immunoprecipitations required, leading to variance between samples. To reduce technical variability and decrease the effort associated with this protocol, we designed an automated immunoprecipitation work-flow on the Biomek4000 platform. Using this system we are interested in addressing questions of how translational control is achieved in cellular systems. Small scale studies have previously demonstrated that protein kinases, like mTOR, affect the translation efficiency of specific genes through linear sequence motifs. Due to constraints imposed by traditional ribosome profiling and related techniques designed to measure translation, the vast majority of the 500 known protein kinases have not been investigated. Coupling RiboPLATE-Seq with a kinome-wide siRNA knock-down screen would allow for the identification of translation targets of each kinase and the development of a regulatory network for protein synthesis. Furthermore, this data could be used to identify linear and structural sequence motifs associated with the targets of each protein kinase, leading to elucidation of each protein kinase's cis-regulatory element. Not only could these results shed light on the complicated mechanisms of translational control, but the highly

targetable nature of protein kinases raises the potential for the development of clinically-relevant solutions to diseases where translational aberration is a key feature.

As a final thought, in the preceding chapters we have discussed the development of a number of technologies which allowed us to investigate translational control and cell-type specific translation. As is the case for any new technology, the overall impact of these developments can only fully be assessed based on their acceptance and future implementation. To that end, we have strived to develop tools which have either made unanswerable scientific questions possible, or greatly enhanced the utility of a previously existing experimental paradigm. We hope that, going forward, these tools will continue to be used to develop a deeper understanding of protein synthesis and translational control.

Chapter 6

Materials and Methods

6.1 Molecular Biology

Tissue processing for RNA

Snap frozen tissue samples (5mg) were homogenized at 4°C with a Dounce homogenizer in 1mL of polysome lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$,1mM DTT, 0.5% Triton X-100, 0.024 U/ml TurboDNase, 0.48 U/mL RNasin, and 0.1 mg/ml cycloheximide). Homogenates were centrifuged for 10 min at 4°C, 14,000 x g. The supernatant was removed and used for the isolation of ribosome footprints, total RNA, and polysome immunoprecipitation (IP). SUPERase-In (0.24U/mL) was added to the lysate used for polysome IP to prevent RNA degradation. Total RNA from homogenates was purified using the RNeasy Mini Kit (Qiagen), and RNA integrity was assessed using a Bioanalyzer (Agilent).

Ribosomal Footprint Isolation and Immunoprecipitation (Glioma Studies)

Mouse tissue lysates were digested for 45 minutes at room temperature with RNase I. Monosomes were isolated using sucrose density gradient fractionation. The monosome fraction was split into two samples. RNA from the first sample was extracted with phenol chloroform and used for further purification of ribosomal footprints. HA-tagged monosomes were immunoprecipitated from the second sample to obtain cell type-specific ribosomal footprints (Figure 2.1A).

For ribosomal immunoprecipitation, we coupled 30 uL of mouse monoclonal anti-HA antibody (HA.11, ascites fluid, Covance) to 300 uL of Protein G-coated Dynabeads (30 mg/mL,

Life Technologies) for 1 hour in citrate-phosphate buffer (24 mM citric acid, 52 mM dibasic sodium phosphate, pH 5.0). Beads were washed once in citrate-phosphate buffer and three times in polysome lysis buffer. Beads were added to the lysates and incubated with rotation at 4^{o}C overnight. Beads were then washed four times with 500 uL of polysome lysis buffer. Ribosomes and footprints were released from the beads using EDTA. Beads were incubated with 140 uL of ribosome release buffer (20 mM Tris-HCl pH 7.3, 250 mM NaCl, 0.5% Triton X-100, 50 mM EDTA) for 5 minutes and the supernatant was set aside on ice. This was repeated three more times. The pooled supernatants were then extracted with phenol-chloroform to yield footprints and digested rRNA fragments.

Ribosome footprint isolation of human glioma and non-neoplastic brain tissue was accomplished as described above for the murine specimens without immunoprecipitation.

Polysome Immunoprecipitation (CAMKII Experiments)

100µL of lysate was used as the input, from which RNA was extracted using the RNeasy Mini Kit (Qiagen). The remaining lysate was used for indirect IP of polysomes. We coupled 15 µL of mouse monoclonal anti-HA.11 (ascites, Biolegend) to lysate with rotation at 4°C for 4 hours. We used 150 µL of protein G-coated Dynabeads (30 mg/mL, Life Technologies) and washed them with 600 µL polysome lysis buffer three times. The conjugated lysate was then added to protein G-coated Dynabeads and incubated with rotation at 4°C overnight. Beads were then washed three times with 500 µL of polysome lysis buffer. RNA was extracted from magnetic beads with polysome release buffer (20 mM Tris-HCl pH 7.3, 250 mM NaCl, 0.5% Triton X-100, 50 mM EDTA) four times for 5 min each (140µL x 4). RNA from the pooled

supernatants (560µL) was then extracted with the RNeasy Mini Kit (Qiagen) and RNA integrity was assessed using a Bioanalyzer (Agilent).

Ribosome Profiling and RNA Sequencing Libraries (Glioma Experiments)

Ribosomal footprints were size selected as described previously (Ingolia et al. 2012) and sequencing libraries constructed. We note that, while the exact adapter scheme reported previously was used to construct ribosome profiling libraries for Tumor Mouse A and Normal Mouse A, a different reverse primer was used for amplification of all other ribosome profiling libraries (Ingolia et al. 2012). In particular, we used the following reverse primer that places the barcode sequence for demultiplexing at the 3'-end of the linker sequence appended to each RNA footprint during single-stranded ligation:

5'-CAAGCAGAAGACGGCATACGAGATNNNNNNATTGATGGTGCCTACAG-3'

where NNNNNN represents the barcode sequences used previously(Ingolia et al. 2012). Ribosome profiling libraries were sequenced using the Illumina HiSeq 2000 system at the Columbia Sulzberger Genome Center. Total RNA for each sample was also provided to the Columbia Sulzberger Geneome Center for poly(A)-selection and RNA-Seq using the Illumina TruSeq kit. We note that a total of ten RNA expression profiles were obtained for deconvolution analysis from ten different mice with PDGF-driven, end-stage tumors. Five of these mice were homozygous for both the RiboTag and *Trp53*$^{flox/flox}$. The other five mice were homozygous for *Trp53*$^{flox/flox}$ and did not harbor the RiboTag allele. Ribosome profiling and RNA-Seq libraries for the five human tissue specimens was accomplished using the same procedure.

RNA sequencing libraries (CAMKII Experiments)

RNA samples were provided to the Columbia Sulzberger Genome Center for poly(A)-selection and RNA-Seq using the Illumina TruSeq kit. There were a total of four RNASeq libraries generated for AZD-treated and vehicle control mice. RNASeq libraries were generated from matched samples used in ligation-free ribosome profiling experiments. Four additional libraries were sequenced from non-Ribosome Profiling matched samples; two total input samples and two matched HA-IP samples.

Polysome Profiling and Quantitative PCR Validation

The left frontal lobe, contralateral to the portion used to generate a ligation-free ribosome profiling library, was conserved and used to generate qPCR data from polysome profiles. The tissue sample was lysed with a Dounce homogenizer, as previously described, and fractionated with a 15-50% sucrose gradient at 37,000 RPM for 3.5 hours. Polysome profiles were obtained and RNA was extracted from fractions using an RNA Clean and Concentrator columns (Zymo). cDNA was generated with a high-capacity RNA to cDNA kit (Life Technologies). qPCR was performed on each fraction with five probes representing genes with either high or low TE as found by ribosome profiling; SYT1 (Mm00436858_m1), SNAP25 (Mm01276449_m1), TGFB1 (Mm01178820_m1), PKD1 (Mm00465434_m1), and TRPV6 (Mm00499069_m1) (ThermoFisher). TaqMan Universal Master Mix (Life Technologies) was used to setup qPCR reactions and a Bio-Rad CFX-96 was used to amplify and read plates. All experiments were performed in triplicate. CQ was determined for each sample and an average CQ number was calculated for each set of triplicates. CQ numbers were converted as follows and the highest value for each gene normalized to one.

$$= 2^{1-CQ}$$

These values were then plotted according to the polysome peak from which they they were obtained.

Ribosome Profiling Sensitivity Measurement

A 34-base RNA oligo, 'AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA [Phos]', was purchased from Sigma and used to generate conventional and ligation-free ribosome profiling libraries. Conventional libraries were generated using the protocol described in Ingolia *et al* using the primers described in Gonzalez *et al* (Ignolia et al. 2012; Gonzalez et al. 2014). The template oligo was serially diluted to the following concentrations; 100 ng, 10 ng, 1 ng, 0.1 ng and 0.01 ng. Following dephosphorylation, both conventional and ligation-free construction schemes were used to attempt to generate libraries at each concentration. For the final PCR step for all libraries in both protocols, PCR was restricted to 9 cycles with 90% of the remaining material. Samples were diluted as necessary and assessed with a High-Sensitivity DNA Bioanalyzer Chip (Agilent).

Poly-A Tailing of Size Selected Fragments

Ribosomal footprints were isolated with a sucrose cushion, size-selected and dephosphorylated as previously described (Ingolia et al. 2011;Ingolia et al. 2012). Following dephosphorylation of size-selected footprints, we determined the concentration of input material using a Bioanalyzer (RNA 6000 Pico Chip, Agilent Technologies). We found that quantification with a Bioanalyzer was more accurate than with a RNA Qubit or Nanodrop, due to the presence of Glycoblue (Ambion) as a precipitant. We used a newly developed kit for small RNA library

construction (SMARTer® smRNA-Seq Kit for Illumina®, Clontech Cat. No. 635030) to generate ligation-free ribosome profiling libraries. Between 1ng and 5ng of size-selected material were used as input and diluted with water to a total volume of 7 µL. Ensuring that reagents remained on ice, Polyadenylation mix was prepared by combining 7 µL of RNA input with 2.5 µL of Mix 1 which includes poly(A) polymerase. After adding the Polyadenylation mix, samples were incubated for 5 minutes at 16°C. Following incubation, samples were immediately placed on ice to ensure the Poly-A Tailing reaction did not continue.

Reverse Transcription and Template Switching

Proceeding from the previous step within 5 minutes, samples were allowed to cool for 1 minute in ice. 1 uL of 3' smRNA dT Primer was added to each tube and mixed by pipetting. Samples were incubated for 3 minutes at 72°C and then transferred to ice for 2 minutes. During this incubation step, Reverse Transcription Master Mix was prepared. Reverse Transcription Master Mix consisted of 6.5 µL smRNA Mix 2, 0.5uL RNase Inhibitor, and 2 µL PrimeScript RT. 9 µL Reverse Transcription Master Mix was added to each sample and mixed by pipetting. Samples were placed in a thermocycler pre-heated to 42°C and incubated at 42°C for one hour followed by ten minute incubation at 70°C to heat inactivate the enzyme.

Ribosomal RNA Depletion

Ribosomal RNA (rRNA) was depleted from samples with a Subtraction Oligo Pool as described previously (Ingolia et al. 2012). Briefly, the Subtraction Oligo Pool consists of several dozen short biotinylated oligos complementary to rRNA fragments that commonly contaminate mammalian ribosome profiling libraries. Following hybridization, the oligos are removed with magnetic streptavidin beads. 10 µL of the previous reverse transcription reaction was combined

with 2 µL of the Subtraction Oligo Pool and mixed. The mixture was heated to 100°C for 90 seconds in a thermocycler. Following heating, the mixture was placed into a 100°C heatblock and allowed to cool to 37°C. Upon reaching 37°C, the mixture was removed from the heatblock and incubated for 15 minutes at 37°C in a thermocycler. While the depletion mixture incubated, 37.5 µL myOne Strepavidin C1 DynaBeads (Invitrogen) were prepared for each sample. Streptavidin beads were washed 3 times with an equal volume of 1X Polysome Buffer. Following the final wash, beads were split into 25 µL and 12.5 µL aliquots. After removing the polysome buffer from the 25 µL aliquot of beads, the depletion mixture was added to the beads and the resulting mixture was incubated for 15 minutes at 37°C. The depletion mixture was then recovered from the beads using a magnet and added to the second, 12.5 µL aliquot of beads. The resulting mixture was incubated for 15 minutes at 37°C. Ensuring no beads were carried over, the depleted RT reaction was then recovered using a magnet.

PCR Library Amplification

The SeqAmp DNA Polymerase included in the SMARTer® smRNA-Seq Kit (Clontech) was used to amplify cDNA from the depleted RT reactions. For the experiments reported, we used the Low-Throughput primer set (Clontech Cat. No. 634844), but have also had success using Clontech's High-Throughput primers (included in the SMARTer® smRNA-Seq Kit). PCR reactions were incubated for one minute at 98°C followed by 12 cycles of a two-step protocol of 98°C for ten seconds, and 68°C for ten seconds.

Purification of Libraries

Purification is necessary due to the presence of primers and other contaminants from upstream reactions. Furthermore, it is critical to ensure reduction of a non-product secondary

peak ~ 25 nucleotides smaller than the product peak. The secondary peak increases linearly with PCR cycle number and is inversely related to total input used. Because the secondary peak is similar to the expected peak-size from ribosome profiling and can interfere with sequencing, it is essential to ensure that it is at least less than half the size of the product-peak. We performed two rounds of purification with AMPure XP beads (Beckman Coulter) at a 1.8X and 1.2X ratio (due to differences in product size, the ratio must be altered when used with the High-Throughput primer set).

Validation of Ribosome Profiling Libraries

We used the Qubit dsDNA High-Sensitivity kit (Life Technologies) to quantify libraries prior to pooling. Libraries were evaluated for the presence of primer and secondary peak with the High-Sensitivity Bioanalyzer DNA chip (Agilent Technologies). In order to fully remove primers and to reduce the contribution of the aforementioned no-insert secondary peak, some libraries require an additional round of 1.2X or 1.0X AMPure XP bead cleanup. Sequencing was performed on a NextSeq 500 desktop sequencer with a 75 cycle high-output kit (Illumina). We obtained between 20 and 50 million demultiplexed, pass-filtered, single-end reads for each sample.

qPCR Enrichment Experiments (High-Throughput)

Cell culture media was removed from 10-cm 75% confluent dishes of WI-38 cells and plates were wash with cold PBS supplemented with .1mg/ml cycloheximide. One mL of cold polysome lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 1mM DTT, .5% Triton X-100, .1mg/mL cycloheximide, .5 U/mL SUPERaseIN) was added to the dish and passed through a 23-gauge needle. Following clarification, 2uL of 1:100 ERCC was added to the

cell lysate. 50uL cell lysate was placed in an Eppendorf tube and allowed to rotate at 4C for 4 hours. 1uL biotinylated y10b antibody (Invitrogen) was added to an additional 50uL cell lysate and incubated with rotation for 4 hours at 4C. Following incubation, 5uL washed streptavidin coated dynabeads (Invitrogen) were added to the lysate containing antibody and allowed to incubate for an additional hour. mRNA was eluted from magnetic dynabeads following three washes with polysome wash buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 1mM DTT) and a final elution wash with ribosome release buffer (20 mM Tris-HCl pH 7.3, 250 mM NaCl, 0.5% Triton X-100, 50 mM EDTA). mRNA was purified from the eluent and the additional lysate with the RNAEasy Kit (Qiagen). We generated cDNA via second-strand synthesis with the Applied Biosystems high-capacity RNA to cDNA synthesis kit. We quantified the amounts of ERCC and B2M remaining in both the immunoprecipitated eluent, as well as the undisturbed lysate with TaqMan qPCR probes from ThermoFisher (Hs00984230_m1, Ac03460023_a1). These experiments were repeated in triplicate.

Compound Administration (High-Throughput)

WI-38 human fibroblast cells were seeded on a 96-well plate at a density of 3,000 cells per well in 60uL cell culture media per well (DulBecco's Modified Eagle Medium supplemented with 10% FBS) 36 hours prior to the start of compound administration (doubling time t=24 hours). Stock solutions of AZD8055, PP242, BKM120 and MNK-I1 were prepared in DMSO so that 1 uL of DMSO was added to each experimental well and 1 uL pure DMSO added to control wells. AZD8055, PP242, BKM120 and MNK-I1 were administered to a final concentration of 50nM, .625nM, 100nM and 1uM, respectively. Two hours following compound administration, cell lysis and immunoprecipitation of ribosome bound mRNA began.

Immunoprecipitation of Ribosome Bound mRNA (High-Throughput)

Following treatment, cell culture media was removed and cells were gently washed 1x with cold PBS supplemented with .1mg/ml cycloheximide. All wells received 30 uL of cold polysome lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$,1mM DTT) supplemented with 0.5% Triton X-100, 0.48 U/mL RNasin, and 0.1 mg/ml cycloheximide and pipetted up and down ~5 times. Additionally, every other sample column received 1uL of 1:5000 ERCC spike-in (Life Technologies 4456740). Cells remained at RT for 5 minutes and were then placed on ice. The plate was spun down to remove bubbles at 1400 rpm for 5 minutes. 10 uL from each well was removed and added to a second plate to be used for Plate-Seq. 10 uL of 2X TCL buffer was then added to each of the Plate-Seq wells. .6 uL SUPERASE-in (Life Technologies AM2696) and .6uL y12b-biotinylated antibody (Thermofisher MA516060) was added to each well, the wells sealed, and incubated for 4 hours at 4C while gently shaking. Following 4 hour incubation, 4 uL of washed streptavidin beads (re-suspended in an equal volume of polysome lysis buffer) were added to each well and allowed to incubate for 1 hour at 4C with gentle shaking. Following 1 hour incubation, samples were placed on a bead magnet and washed 3 times with polysome buffer supplemented with .05% Triton X-100 and .1 mg/ml cycloheximide. Following the final wash, 15 uL ribosome release buffer (20 mM Tris-HCl pH 7.3, 250 mM NaCl, 0.5% Triton X-100, 50 mM EDTA) was added to each well and allowed to incubate for 15 minutes. Beads were removed by plate-magnet and the supernatant added to 15 uL of TCL buffer.

6.2 Computational

Bioinformatic Analysis of Ligation-free Ribosome Profiling

Each read contains a G-rich region from terminal transferase activity, followed by a ribosome footprint and a poly-A tail. The first 5 and last 20 bases of each read were removed with fastx_trimmer from the FASTX Toolkit. Because the Poly-A tail can appear at different points in the read, stretches of 'AAAAAAAA' at the 3' end of reads were removed with fastx_clipper; reads shorter than 15 bases after trimming and clipping were discarded. Contaminating rRNA reads were removed by mapping all reads to a rRNA reference library with Bowtie2, allowing for 1 error and outputting reads which did not align (Ingolia et al. 2012). Reads which did not map to the rRNA reference were aligned to the genome and transcriptome with Tophat2 without looking for novel junctions. Following mapping, read counting was performed with HTSeq set in interstrict mode.

Bioinformatic Analysis of Traditional Ribosome Profiling Data

Following Illumina sequencing of our ribosome profiling libraries, we first demultiplexed our raw data using a barcode that was attached to the sequencing template during PCR amplification. The 3'-adapter sequence (CTGTAGGCACCAT) was clipped from each read using fastx_clipper. We discarded all reads that did not contain the adapter sequence or that were shorter than 25 bases after adapter clipping. We then used Bowtie 2 to map the clipped reads to a mouse rRNA reference and discarded any reads that mapped to rRNA(Langmead and Salzberg, 2012). The remaining reads were then mapped to either the mouse (Illumina iGenomes mm10 reference) or human (Illumina iGenomes hg19 reference) transcriptomes using Tophat 2(Kim et al. 2013). We did not attempt to detect novel junctions and obtained ~0.5-11M uniquely mapped reads per mouse sample and ~4-24M uniquely mapped reads per human sample. We used HTSeq

to compute read counts for reads that mapped uniquely to the mouse or human transcriptome in the CDS, 5'-leader (5'-UTR), 3'-UTR, and complete exonic sequence of each gene.

Bioinformatic Analysis of RNA-Seq Data

For RNA-Seq, we received fastq files containing demultiplexed, single-end 100 base reads for each sample from the Columbia Sulzberger Genome Center. We used Tophat 2 to map the reads to the mouse (Illumina iGenomes mm10 reference) or human (Illumina iGenomes hg19 reference) transcriptomes and did not attempt to detect novel junctions, obtaining ~30 million reads per sample. We used HTSeq to compute read counts for reads that uniquely mapped to the transcriptome for each gene.

Calculation of Unique Fragments

The number of unique fragments was calculated for both methods of ribosome profiling with Picard Tools downloaded from the Broad Institute. Picard Tools was used in MarkDuplicates mode and was run files downsampled from the original .bam file output from TopHat that was previously generated for each sample. Downsampling was performed with fastq-sample from the fastq-tools suite. Following sorting and indexing with SamTools, the number of unique fragments was determined with Picard Tools.

Analysis of Translational Activity and RiboTag Enrichment

To analyze differential translation efficiency between the control and AZD-treated samples, we used the recently reported RiboDiff algorithm with the CDS-mapping RNA-Seq and ribosome profiling reads as input (Zong et al. 2015). RiboTag enrichment scores were calculated from two RiboTag IP experiments and two homogenate experiments. RiboTag enrichment scores

were calculated for each gene by first normalizing counts found in RiboTag and homogenate samples by size factors generated from DESeq2. Following normalization, enrichment scores were calculated by dividing normalized RiboTag counts by normalized homogenate counts.

Translation efficiency was also calculated on a per-sample basis by normalizing ribosome profiling and RNA-Seq counts by size factors from DESeq2 and dividing ribosome profiling counts by RNA-Seq counts. We thresholded downstream analyses by removing genes that had less than 37 counts in ribosome profiling and RNASeq data. When the TE of both samples in a group was used, the threshold was increased to 75 counts.

Cell Type-Specific Specific Lists

We used an RNA-Seq database generated from purified representative cell type populations in order to generate rank-lists of cell type-specific genes (Zhang et al. 2014). We created seven cell type-specific enrichment rank lists, one for each of the 7 representative cell types in the database. Enrichment scores for each cell type were calculated for every gene. These scores *E* were calculated for each gene *i* in each cell type *j* were computed from their cell type-specific RNA expression levels $FPKM_{ij}$ using the following equation:

$$E = \frac{FPKM_{ij}}{\sum_k FPKM_{ik}} - \frac{1}{2}$$

This resulted in seven cell type-specific enrichment scores between -0.5 and 0.5 for each gene. This value was later recalculated without including Newly Formed Oligodendrocytes as a cell type (in order to improve enrichment among the remaining cell-types due to significant

overlap between myelinating and newly formed oligodendrocytes). These cell-type enrichment rank-lists were later used in Gene Set Enrichment Analysis (GSEA) and to define which genes were most associated with specific cell-types. Cell-type specific genes were defined as having an enrichment score greater than 0.2.

Gene Set Enrichment Analysis

In order to determine the role of translational regulation in cell-type specific genes, we performed a Gene Set Enrichment Analysis (GSEA) with GSEA software downloaded from the Broad Institute (Subramanian et al. 2005). In all instances of GSEA we performed a "Classic" GSEA analysis in pre-ranked mode. Gene sets were constructed from previously calculated and thresholded TE values for each sample individually and for combined samples as described above. Between 10,201 and 9,904 genes (difference due to previously mentioned thresholding) were ranked based on their TE calculated from untreated RiboTag brains into bins. Equal sized bins spanning 0.75 TE units were constructed around the median and populated with genes based on their TE rank. This was then used as the gene set input for GSEA for each sample.

Cell type-specific enrichment scores, which are described above, were ranked and used to determine if cell type-specific genes were enriched in TE bins. Input to GSEA was a gene-set composed of TE values for a given sample (described above), and a rank-list composed the enrichment scores of a single cell-type. GSEA was then repeated for gene-set with every cell type rank-list. Normalized Enrichment Scores (NES) were generated from the GSEA software and then used to generate figures. The statistical significance of differences in TE between cell-types was calculated using GSEA. The enrichment scores previously calculated for each cell type were used to generate a new comparison score for each gene *i* in each cell type *k* and *j*.

$$Es = \frac{E_{ik}}{E_{ij}}$$

Rank-lists were then generated for each pairwise combination of cell-types composed of calculated comparison scores for each gene. GSEA was run with the same settings as before using the previously generated gene sets based on TE scores.

Gene Ontology Analysis (CAMKII Experiments)

As a secondary means of displaying the cell type-specific translational landscapes we observed, we generated lists of cell type-specific gene ontologies. In order to calculate the enrichment of cell type-specific genes in gene ontologies, a list of 1,400 gene ontologies taken from the iPAGE database was used to create gene-sets where each set was a single ontology (Goodarzi et al. 2009). NES for the enrichment of cell-type specific genes in individual ontologies were produced using this gene-set in conjunction with previously generated rank-lists comprised of enrichment scores (one for each cell-type). A gene ontology was defined as being enriched in an individual cell type if the NES for that cell-type was at least three units higher than the next highest NES for that gene ontology. Median TE was calculated for genes within enriched ontologies and plotted.

Gene Ontology Analysis (Glioma Experiments)

Mutual information-based gene ontology analysis was carried out using the differential translation rate analysis between RiboTag and normal brain profiles with iPAGE(Goodarzi et al. 2009). Gene identification names from the mm10 reference transcriptome annotation were converted to RefSeq identifiers using Babelomics and filtered based on the annotation provided

by iPAGE. We then used log2-transformed fold-change obtained from DESeq for each statistically significant differentially expressed gene ($p < 0.05$) between the RiboTag profiles and the normal brain profiles as input for iPAGE to obtain over- and under-represented gene ontologies across nine bins of translation rate fold-change.

We identified ~100 genes from our differential translation rate analysis between the homogenate and normal brain profiles with both a significantly higher translation rate in the tumor compared to normal brain and on the consensus list of RiboTag-depleted genes. This list was too short to allow iPAGE analysis, and so we identified enriched gene ontologies using Fisher's Exact Test as implemented in FuncAssociate 2.0(Berriz et al. 2009).

5' UTR Analysis

The number of ribosome profiling and RNA-Seq reads mapped to the 5'UTR were counted with HTSeq-counts set in region-interstrict mode for each matched sample. Cell-type specific genes were defined for this analysis as having a previously calculated enrichment value greater than .2. The fraction of cell-type specific genes with 5' UTR ribosomal density was calculated as the percentage of cell-type specific genes with at least 1 ribosomal footprints in the 5' UTR region. Upstream AUG sequences were identified with a custom python script and defined as any AUG sequence found within the 5'UTR region of a gene in genes with 5'UTR density. Median TE was calculated for cell-type specific genes as well as for the subgroups of cell-type specific genes with 5' UTR density and containing uAUG and genes containing 5'UTR density without uAUG. The weighted average of 5'UTR length for each gene was calculated using isoform abundance information from Cufflinks. Cufflinks was quantitated against a reference transcript annotation and otherwise run with default settings. GC content of 5'UTRs

was calculated in the same manner using isoform abundance information from Cufflinks. Genes were sorted into bins defined by GC content and length, and median TE was calculated. The significance of the change in TE due to 5' UTR GC content and 5' UTR length was calculated using the Mann-Whitney U test.

Analysis of Translational Activity, Enrichment, and Differential Translation Rate (Glioma Experiments)

For each gene, we divided the CDS read counts determined above by CDS length and total number of uniquely mapped reads in the sample to determine the CDS ribosome footprint density which we refer to as the translation rate. Similarly, we computed the 5'-leader ribosome footprint density by dividing the 5'-leader read counts determined above by 5'-leader length and total number of uniquely mapped reads in each sample.

To calculate the enrichment scores for each gene, we divided the translation rate measured from each RiboTag ribosome profile by the translation rate measured from the corresponding homogenate ribosome profile. An enrichment score greater than one indicates that the gene is biased towards translation in the transformed cells. We generated a consensus list of enriched genes based on the criteria that the enrichment score is greater than one in all three biological replicates and that the number of uniquely mapped CDS reads in each mouse is greater than ten in the RiboTag ribosome profile. Similarly, we placed genes with enrichment scores less than one and with greater than ten CDS reads in the homogenate ribosome profile in all three biological replicates in the consensus list of depleted genes.

We calculated translation efficiency for a given gene by dividing its translation rate by the RNA read density (number of reads uniquely mapping to the complete exonic sequence for each gene divided by transcript length and total number of uniquely mapped reads from a

sample). Both the differential translation rate and RNA expression analyses were carried out using DESeq(Anders and Huber, 2010) based on the uniquely mapped, CDS ribosome footprint counts and the uniquely mapped exonic RNA counts for each gene computed using HTSeq.

For the human ribosome profiling translation rate analysis depicted in Figure 2.4D, we show genes for which we measured at least at 2-fold increase in translation rate in the tumor tissue compared to normal brain for all possible pairwise comparisons of tumor and normal brain specimens.

6.3 Mouse Models and Tissue Handling

Camk2a-RiboTag Mouse Model

Camk2a-cre mice (JAX ID 005359) have the mouse calcium/calmodulin-dependent protein kinase II alpha (*Camk2a*) promoter driving Cre recombinase expression in the forebrain, specifically in principal excitatory neurons. Camk2a-cre mice were crossed to RiboTag mice (JAX ID 011029) which contain a conditional knock-in allele where exon 4 of the ribosomal protein L22 (Rpl22) is flanked by loxP sites, followed by an identical exon tagged with three repeated hemagglutinin epitope coding sequences (HA-tag). The resulting Camk2a-cre-RiboTag cross expresses the HA-tagged Rpl22 protein in principal excitatory neurons. Camk2a-cre heterozygotes were crossed to homozygous RiboTag mice and genotyped with the following primers for Cre: GCG GTC TGG CAG TAA AAA CTA TC (transgene), GTG AAA CAG CAT TGC TGT CAC TT (transgene), CTA GGC CAC AGA ATT GAA AGA TCT (internal positive control forward), GTA GGT GGA AAT TCT AGC ATC ATC C (internal positive control

reverse), and the following primers for RiboTag: GGG AGG CTT GCT GGA TAT G (forward), TTT CCA GAC ACA GGC TAA GTA CAC (reverse).

Previous reports have shown that recombination with the Camk2a promoter-driven cre begins during the third postnatal week and is completed by the fourth postnatal week, therefore we chose to use mice that were three months old for all experiments.

AZD Drug Delivery and Tissue Collection

AZD-8055 (Selleckchem) was dissolved in Captisol and diluted to a final Captisol concentration of 30% (w/v). A single dose of AZD-8055 was administered by oral gavage (100mg/kg). Vehicle consisted of 30% captisol and was also delivered by oral gavage. Camk2a-cre-RiboTag mice were sacrificed 1 hour after AZD-8055 or vehicle administration, two mice were used per condition. Cervical dislocation was performed and the right frontal lobe of the brain was collected and snap-frozen in liquid nitrogen prior to polysome extraction. The remaining brain lobes were fixed in 4% PFA for 48 hours and embedded in paraffin for histological analysis.

Ribotag Mouse Glioma Model

For experimental induction of murine glioma, transgenic C57BL/6 mice carrying loxP recognition sites at exon 7 of *Trp53* were crossed with RiboTag mice (JAX ID 011029), which carry the HA-affinity tag adjacent to the ribosomal protein *Rpl22*, separated from the natively expressed terminal exon by loxP recognition sites. Hence, the tagged version of *Rpl22* is only expressed following Cre-mediated recombination. These mice were bred to $Trp53^{flox/flox}$ and RiboTag homozygosity. Proneural gliomas were induced *de novo* by stereotactic injection into subcortical white matter of the right frontal lobe of ~$5x10^4$ replication incompetent, retroviral

particles expressing human platelet-derived growth factor (PDGF-B) and Cre recombinase, as described previously(Lei et al. 2011). Two of the three mice in which tumors were induced were 43 days old and the third mouse was 64 days old. Age-matched control mice were injected with an equal volume of serum-free media. Mice were monitored for tumor morbidity by behavior and weight, and sacrificed at 30 days according to Columbia University IACUC protocol #AC-AAAF1710. At this time point, all three mice exhibited symptoms of tumor morbidity and tumors were clearly visible upon removal of the brain. The survival curve in Figure 2.1 indicates a median survival time of 47+/-7 days post injection, and so we sacrificed the animals at 30 days post injection to avoid death due to tumor morbidity at an uncontrolled time so that we could harvest fresh polysomal RNA from the tumor tissue. The right frontal lobe tissue (containing the injection site) and distal tissue from the contralateral hemisphere were snap-frozen in liquid nitrogen immediately following sacrifice. Tissue immediately adjacent to the experimental sample, containing tumor, was fixed in 4% paraformaldehyde (PFA) for 48 hours prior to immunofluorescence. The survival curves depicted in Figure 2.1C were generated by injecting nine *Trp53*$^{flox/flox}$ mice and eleven wildtype C57BL/6 mice (all 6-8 weeks old) with $5x10^4$ viral particles expressing PDGF-B and Cre recombinase(Lei et al. 2011). We note that we have previously reported the use of this specific initiating alteration in conjunction with PDGF-B overexpression (Sonabend et al. 2014).

Human Brain Tumor and Non-Neoplatic Brain Tissue Specimens

The five adult patients included in chapter 2 presented for surgical resection of either malignant glioma or of non-neoplastic brain tissue to relieve epilepsy symptoms. The three

epilepsy patients had no oncological history. Resected tissue specimens were snap-frozen with liquid nitrogen in the operating room to maximize RNA preservation.

6.4 Blotting and Immunofluorescence

Immunofluorescence (CAMKII Experiments)

Fixed brains were embedded in paraffin and tissue sections (5µm) were used for staining. To remove excess paraffin, slides were immersed in xylene then rehydrated by incubation in 100%, 95%, and 75% ethanol. Slides were washed in PBS then water. For antigen retrieval 10mM citrate buffer (pH 6.0) was heated and slides were immersed for 20 minutes, followed by PBS washes. Sections were then permeabilized with 0.5% Triton-X100 in PBS for 15 minutes, blocked in 5% goat serum for 1 hour and incubated with primary antibodies overnight at 4°C. Sections were washed three times in PBS and incubated with AlexaFluor-conjugated secondary antibodies (1:1000, Invitrogen) for one hour at room temperature and counter stained with DAPI. Stained tissue sections were imaged using a Nikon TE2000 epifluorescence microscope.

Immunofluorescence (Glioma Experiments)

At 28 days post injection, mouse brains were fixed in 4% PFA for 48 hours. Brains were then cryoprotected in 30% sucrose for four days and then stored in OCT at -80^{o}C. Cyrosections (10 um) were fixed in 4% PFA for 10 minutes at room temperature, washed in PBS, blocked with 5% horse serum (Sigma) for 30 minutes, and then labeled with primary antibodies overnight at 4^{o}C. Sections were then washed three times with PBS and incubated with AlexaFluor-conjugated secondary antibodies (1:1000, Invitrogen) for one hour at room temperature and counter-stained with DAPI.

Antibodies (CAMKII Experiments)

The following primary antibodies were used for immunofluorescence and western blotting: mouse monoclonal anti-HA.11 ascites (1:500, Biolegend01515), rabbit anti-pS6 S240/244 (1:500, Cell Signaling15), rabbit anti-NeuN (1:500, Cell Signaling943), rabbit anti-pS6 S235/236 (1:1000, Cell Signaling)11), rabbit anti-S6 (1:1000, Cell Signaling17), rabbit anti-β-Actin (1:1000, Cell Signaling970S). The following secondary antibodies were used for immunofluorescence and western blotting: goat anti-rabbit Alexa 488 (1:1000, Invitrogen #A11008), and goat anti-mouse Alexa 568 (1:1000, Invitrogen #A11031).

Antibodies (Glioma Experiments)

The following primary antibodies were used: mouse anti-HA (1:1000, Covance), rabbit anti-PDGFRB (1:500, Cell Signaling), rabbit anti-OLIG2 (1:100, Millipore), rabbit anti-GFAP (1:500, Dako), rabbit anti-RBFOX3 (1:1000, Millipore), rat anti-CD44 (1:150, Calbiochem), and rabbit anti-AIF1 (1:1000, Wako).

Western blot analysis

Tissue was collected one hour after vehicle or AZD-8055 administration (20mg/kg or 100mg/kg AZD-8055). The right frontal brain lobe was lysed from male mice that were 12 weeks old. Tissue was lysed in 1 mL cell extraction buffer (Invitrogen #FNN10011) supplemented with protease (Sigma #P7626) and phosphatase inhibitors (Sigma#P5726, #P0044) with a Dounce homogenizer. Lysate was centrifuged and the supernatant was collected for total protein quantification. 30 ng of total protein was loaded to a NuPAGE 4-12% Bis-Tris gel and subject to gel electrophoresis according to the manufacturer's instructions (Invitrogen

#NP0321BOX). Bands were detected by fluorescent imaging using the Typhoon imaging system.

Microscopy

Stained tissue sections were imaged using a Nikon TE2000 epifluorescence microscope equipped with Metamorph software (Molecular Devices). Micrographs were merged using Metamorph and ImageJ. Immunofluorescence images of the RiboTag cell line were obtained using a Nikon Ti-U epifluorescence microscope equipped with an EM-CCD camera (Andor iXon) and a 532 nm diode-pumped solid state laser.

REFERENCES

1. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).

2. D. E. Andreev *et al.*, Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol* **16**, 90 (2015).

3. Y. Arava, F. E. Boas, P. O. Brown, D. Herschlag, Dissecting eukaryotic translation and its control by ribosome density mapping. *Nucleic Acids Res* **33**, 2421-2432 (2005).

4. Y. Arava *et al.*, Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 3889-3894 (2003).

5. J. A. Arribere, W. V. Gilbert, Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome research* **23**, 977-987 (2013).

6. M. Assanah *et al.*, Glial progenitors in adult white matter are driven to form malignant gliomas by platelet-derived growth factor-expressing retroviruses. *The Journal of neuroscience* **26**, 6781-6790 (2006).

7. D. Avni, Y. Biberman, O. Meyuhas, The 5' terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. *Nucleic Acids Res* **25**, 995-1001 (1997).

8. N. Ban *et al.*, A 9 A resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* **93**, 1105-1115 (1998).

9. C. Barbosa, I. Peixeiro, L. Romao, Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS genetics* **9**, (2013).

10. A. Bedard, P. Tremblay, A. Chernomoretz, L. Vallieres, Identification of genes preferentially expressed by microglia and upregulated during cuprizone-induced inflammation. *Glia* **55**, 777-789 (2007).

11. G. F. Berriz, J. E. Beaver, C. Cenik, M. Tasan, F. P. Roth, Next generation software for functional trend analysis. *Bioinformatics* **25**, 3043-3044 (2009).

12. M. Bhat *et al.*, Targeting the translation machinery in cancer. *Nat Rev Drug Discov* **14**, 261-278 (2015).

13. R. Bonavia, M. M. Inda, W. K. Cavenee, F. B. Furnari, Heterogeneity maintenance in glioblastoma: a social network. *Cancer Res* **71**, 4055-4060 (2011).

14. S. Brachmann, C. Fritsch, S. M. Maira, C. Garcia-Echeverria, PI3K and mTOR inhibitors: a new generation of targeted anticancer agents. *Curr Opin Cell Biol* **21**, 194-198 (2009).

15. G. A. Brar *et al.*, High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**, 552-557 (2012).

16. S. Brenner *et al.*, Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**, 630-634 (2000).

17. V. Brown *et al.*, Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477-487 (2001).

18. J. D. Cahoy *et al.*, A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *The Journal of Neuroscience* **28**, 264-278 (2008).

19. L. Calviello *et al.*, Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods* **13**, 165-170 (2016).

20. Z. B. Chen *et al.*, Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725-730 (2005).

21. I. M. Chiu *et al.*, A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Rep* **4**, 385-401 (2013).

22. J. Cho *et al.*, Multiple repressive mechanisms in the hippocampus during memory formation. *Science* **350**, 82-87 (2015).

23. C. M. Chresta *et al.*, AZD8055 is a potent, selective, and orally bioavailable ATP-competitive mammalian target of rapamycin kinase inhibitor with in vitro and in vivo antitumor activity. *Cancer Res* **70**, 288-298 (2010).

24. M. J. Clemens, Targets and mechanisms for the regulation of translation in malignant transformation. *Oncogene* **23**, 3180-3188 (2004).

25. J. C. Darnell *et al.*, FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-261 (2011).

26. H. P. Davis, L. R. Squire, Protein synthesis and memory: a review. *Psychol Bull* **96**, 518-559 (1984).

27. E. de Nadal, G. Ammerer, F. Posas, Controlling gene expression in response to stress. *Nat Rev Genet* **12**, 833-845 (2011).

28. N. V. Dorrello *et al.*, S6K1- and betaTRCP-mediated degradation of PDCD4 promotes protein translation and cell growth. *Science* **314**, 467-471 (2006).

29. J. Douros, M. Suffness, New antitumor substances of natural origin. *Cancer Treat Rev* **8**, 63-87 (1981).

30. J. P. Doyle *et al.*, Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749-762 (2008).

31. S. Dvir *et al.*, Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc Natl Acad Sci U S A* **110**, E2792-2801 (2013).

32. Q. W. Fan *et al.*, A dual phosphoinositide-3-kinase alpha/mTOR inhibitor cooperates with blockade of epidermal growth factor receptor in PTEN-mutant glioma. *Cancer research* **67**, 7960-7965 (2007).

33. K. I. Farley, S. J. Baserga, Probing the mechanisms underlying human diseases in making ribosomes. *Biochem Soc Trans* **44**, 1035-1044 (2016).

34. S. N. Floor, J. A. Doudna, Tunable protein synthesis by transcript isoforms in human cells. *Elife* **5**, (2016).

35. G. E. Fox, Origin and evolution of the ribosome. *Cold Spring Harb Perspect Biol* **2**, a003483 (2010).

36. J. Frank *et al.*, A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature* **376**, 441-444 (1995).

37. Y. Furuichi, A. J. Shatkin, Viral and cellular mRNA capping: past and prospects. *Adv Virus Res* **55**, 135-184 (2000).

38. V. Gandin *et al.*, Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature* **455**, 684-688 (2008).

39. M. Ghanem, R. Nijman, M. Safan, T. van der Kwast, G. Vansteenbrugge, Expression and prognostic value of platelet-derived growth factor-AA and its receptor alpha in nephroblastoma. *BJU Int* **106**, 1389-1393 (2010).

40. C. Gonzalez *et al.*, Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. *J Neurosci* **34**, 10924-10936 (2014).

41. H. Goodarzi, O. Elemento, S. Tavazoie, Revealing Global Regulatory Perturbations across Human Cancers. *Molecular Cell* **36**, 900-911 (2009).

42. D. A. Guertin, D. M. Sabatini, An expanding role for mTOR in cancer. *Trends Mol Med* **11**, 353-361 (2005).

43. H. Guo, N. T. Ingolia, J. S. Weissman, D. P. Bartel, Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).

44. R. E. Halbeisen, A. P. Gerber, Stress-dependent coordination of transcriptome and translatome in yeast. *PLoS Biol* **7**, e1000105 (2009).

45. M. Heiman *et al.*, A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738-748 (2008).

46. K. Helmy *et al.*, Identification of Global Alteration of Translational Regulation in Glioma In Vivo. *PLoS One* **7**, (2012).

47. J. J. Ho *et al.*, Systemic Reprogramming of Translation Efficiencies on Oxygen Stimulus. *Cell Rep* **14**, 1293-1300 (2016).

48. D. B. Hoelzinger, T. Demuth, M. E. Berens, Autocrine factors that sustain glioma invasion and paracrine biology in the brain microenvironment. *J Natl Cancer Inst* **99**, 1583-1593 (2007).

49. M. Holcik, N. Sonenberg, R. G. Korneluk, Internal ribosome initiation of translation and the control of cell death. *Trends Genet* **16**, 469-473 (2000).

50. N. Hornstein *et al.*, Ligation-free ribosome profiling of cell type-specific translation in the brain. *Genome Biol* **17**, 149 (2016).

51. A. C. Hsieh *et al.*, The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61 (2012).

52. A. C. Hsieh *et al.*, The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-U196 (2012).

53. N. T. Ingolia, G. A. Brar, S. Rouskin, A. M. McGeachy, J. S. Weissman, The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**, 1534-1550 (2012).

54. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).

55. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218 (2009).

56. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789-802 (2011).

57. S. Islam *et al.*, Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* **21**, 1160-1167 (2011).

58. R. J. Jackson, C. U. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127 (2010).

59. C. H. Jan, C. C. Williams, J. S. Weissman, Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014).

60. B. H. Jiang, L. Z. Liu, in *Advances in Cancer Research, Vol 102,* G. F. V. Woude, G. Klein, Eds. (2009), vol. 102, pp. 19-65.

61. G. Johannes, M. S. Carter, M. B. Eisen, P. O. Brown, P. Sarnow, Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc Natl Acad Sci U S A* **96**, 13118-13123 (1999).

62. R. Jorgensen *et al.*, Exotoxin A-eEF2 complex structure indicates ADP ribosylation by ribosome mimicry. *Nature* **436**, 979-984 (2005).

63. E. R. Kandel, The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030-1038 (2001).

64. A. M. Katz *et al.*, Astrocyte-Specific Expression Patterns Associated with the PDGF-Induced Glioma Microenvironment. *PLoS One* **7**, (2012).

65. R. J. Kelleher, 3rd, A. Govindarajan, H. Y. Jung, H. Kang, S. Tonegawa, Translational control by MAPK signaling in long-term synaptic plasticity and memory. *Cell* **116**, 467-479 (2004).

66. P. Kemmeren *et al.*, Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**, 740-752 (2014).

67. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, (2013).

68. A. Kuhn, D. Thu, H. J. Waldvogel, R. L. M. Faull, R. Luthi-Carter, Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *nAture methods* **8**, 945-947 (2011).

69. K. M. Kuhn, J. L. DeRisi, P. O. Brown, P. Sarnow, Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol Cell Biol* **21**, 916-927 (2001).

70. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

71. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *nAture methods* **9**, 357-U354 (2012).

72. M. Laplante, D. M. Sabatini, mTOR signaling in growth control and disease. *Cell* **149**, 274-293 (2012).

73. L. F. Lareau, D. H. Hite, G. J. Hogan, P. O. Brown, Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* **3**, e01257 (2014).

74. O. Larsson *et al.*, Eukaryotic translation initiation factor 4E induced progression of primary human mammary epithelial cells along the cancer pathway is associated with targeted translational deregulation of oncogenic drivers and inhibitors. *Cancer Res* **67**, 6814-6824 (2007).

75. O. Larsson, N. Sonenberg, R. Nadon, Identification of differential translation in genome wide studies. *Proc Natl Acad Sci U S A* **107**, 21487-21492 (2010).

76. A. Lazaris-Karatzas, K. S. Montine, N. Sonenberg, Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* **345**, 544-547 (1990).

77. L. Lei *et al.*, Glioblastoma Models Reveal the Connection between Adult Glial Progenitors and the Proneural Phenotype. *PLoS One* **6**, (2011).

78. G. Leprivier *et al.*, The eEF2 kinase confers resistance to nutrient deprivation by blocking translation elongation. *Cell* **153**, 1064-1079 (2013).

79. S. Levy, D. Avni, N. Hariharan, R. P. Perry, O. Meyuhas, Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc Natl Acad Sci U S A* **88**, 3319-3323 (1991).

80. C. J. Lin, R. Cencic, J. R. Mills, F. Robert, J. Pelletier, c-Myc and eIF4F are components of a feedforward loop that links transcription and translation. *Cancer Res* **68**, 5326-5334 (2008).

81. B. Liu, S. B. Qian, Translational reprogramming in cellular stress response. *Wiley Interdiscip Rev RNA* **5**, 301-315 (2014).

82. C. Liu *et al.*, Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* **146**, 209-221 (2011).

83. G. X. Luo, J. Taylor, Template switching by reverse transcriptase during DNA synthesis. *J Virol* **64**, 4321-4328 (1990).

84. F. Meric-Bernstam, Translation initiation factor 4E (eIF4E): prognostic marker and potential therapeutic target. *Ann Surg Oncol* **15**, 2996-2997 (2008).

85. O. Meyuhas, Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem* **267**, 6321-6330 (2000).

86. W. Mikulits *et al.*, Isolation of translationally controlled mRNAs by differential screening. *FASEB J* **14**, 1641-1652 (2000).

87. D. J. Mulholland *et al.*, Pten loss and RAS/MAPK activation cooperate to promote EMT and metastasis initiated from prostate cancer stem/progenitor cells. *Cancer Res* **72**, 1878-1889 (2012).

88. E. Oh *et al.*, Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell* **147**, 1295-1308 (2011).

89. W. J. Oh, E. Jacinto, mTOR complex 2 signaling and functions. *Cell Cycle* **10**, 2305-2316 (2011).

90. A. T. Parsa, E. C. Holland, Cooperative translational control of gene expression by Ras and Akt in cancer. *Trends in Molecular Medicine* **10**, 607-613 (2004).

91. A. Pause *et al.*, Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* **371**, 762-767 (1994).

92. L. Peil *et al.*, Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc Natl Acad Sci U S A* **110**, 15265-15270 (2013).

93. J. Pelletier, J. Graff, D. Ruggero, N. Sonenberg, Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. *Cancer Res* **75**, 250-263 (2015).

94. G. Pianese, Beitrag zur Histologic und Aetiologies des Karzinoms. *Beitr. Pathol. Anat. Allg. Pathol.* **142**, (1896).

95. S. Picelli *et al.*, Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).

96. K. G. Pike *et al.*, Optimization of potent and selective dual mTORC1 and mTORC2 inhibitors: the discovery of AZD8055 and AZD2014. *Bioorg Med Chem Lett* **23**, 1212-1216 (2013).

97. W. W. Pong, D. H. Gutmann, The ecology of brain tumors: lessons learned from neurofibromatosis-1. *Oncogene* **30**, 1135-1146 (2011).

98. V. K. Rajasekhar *et al.*, Oncogenic Ras and Akt signaling contribute to glioblastoma formation by differential recruitment of existing mRNAs to Polysomes. *Molecular Cell* **12**, 889-901 (2003).

99. D. Ramskold *et al.*, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**, 777-782 (2012).

100. K. Richter, M. Haslbeck, J. Buchner, The heat shock response: life on the verge of death. *Mol Cell* **40**, 253-266 (2010).

101. D. F. Rolfe, G. C. Brown, Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev* **77**, 731-758 (1997).

102. D. Ruggero *et al.*, The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nat Med* **10**, 484-486 (2004).

103. D. D. Sabatini, Y. Tashiro, G. E. Palade, On the attachment of ribosomes to microsomal membranes. *J Mol Biol* **19**, 503-524 (1966).

104. F. Sanger *et al.*, Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-695 (1977).

105. F. Sanger, G. G. Brownlee, B. G. Barrell, A two-dimensional fractionation procedure for radioactive nucleotides. *J Mol Biol* **13**, 373-398 (1965).

106. E. Sanz *et al.*, Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13939-13944 (2009).

107. S. N. Sehgal, H. Baker, C. Vezina, Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J Antibiot (Tokyo)* **28**, 727-732 (1975).

108. P. Shah, Y. Ding, M. Niemczyk, G. Kudla, J. B. Plotkin, Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589-1601 (2013).

109. K. Sharma *et al.*, Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci* **18**, 1819-1831 (2015).

110. M. S. Sheikh, A. J. Fornace, Jr., Regulation of translation initiation following stress. *Oncogene* **18**, 6121-6128 (1999).

111. R. A. Signer, J. A. Magee, A. Salic, S. J. Morrison, Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49-54 (2014).

112. J. Somers, T. Poyry, A. E. Willis, A perspective on mammalian upstream open reading frame function. *International Journal of Biochemistry & Cell Biology* **45**, 1690-1700 (2013).

113. A. M. Sonabend *et al.*, The transcriptional regulatory network of proneural glioma determines the genetic alterations selected during tumor progression. *Cancer Res* **74**, 1440-1451 (2014).

114. A. M. Sonabend *et al.*, Murine cell line model of proneural glioma for evaluation of anti-tumor therapies. *J. Neuro-Oncol.* **112**, 375-382 (2013).

115. K. A. Spriggs, M. Bushell, A. E. Willis, Translational regulation of gene expression during conditions of cell stress. *Mol Cell* **40**, 228-237 (2010).

116. J. A. Steitz, Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* **224**, 957-964 (1969).

117. O. Steward, E. M. Schuman, Protein synthesis at synaptic sites on dendrites. *Annu Rev Neurosci* **24**, 299-325 (2001).

118. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

119. S. Tahmasebi *et al.*, Multifaceted regulation of somatic cell reprogramming by mRNA translational control. *Cell Stem Cell* **14**, 606-616 (2014).

120. M. Takanami, Y. Yan, T. H. Jukes, Studies on the site of ribosomal binding of f2 bacteriophage RNA. *J Mol Biol* **12**, 761-773 (1965).

121. H. Takeuchi *et al.*, Synergistic augmentation of rapamycin-induced autophagy in malignant glioma cells by phosphatidylinositol 3-kinase/protein kinase B inhibitors. *Cancer research* **65**, 3336-3346 (2005).

122. F. E. Taub, J. M. DeLeo, E. B. Thompson, Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. *DNA* **2**, 309-327 (1983).

123. H. M. Temin, S. Mizutani, RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213 (1970).

124. C. C. Thoreen *et al.*, A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109-113 (2012).

125. J. Z. Tsien *et al.*, Subregion- and Cell Type–Restricted Gene Knockout in Mouse Brain. *Cell* **87**, 1317-1326.

126. M. Valle *et al.*, Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* **10**, 899-906 (2003).

127. M. van Slegtenhorst *et al.*, Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. *Science* **277**, 805-808 (1997).

128. R. G. W. Verhaak *et al.*, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* **17**, 98-110 (2010).

129. D. E. Weinberg *et al.*, Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep*, (2016).

130. H. G. Wendel *et al.*, Dissecting eIF4E action in tumorigenesis. *Genes Dev* **21**, 3232-3237 (2007).

131. S. L. Wolin, P. Walter, Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J* **7**, 3559-3569 (1988).

132. M. Wong, Mammalian target of rapamycin (mTOR) pathways in neurological diseases. *Biomed J* **36**, 40-50 (2013).

133. M. Wong, A critical review of mTOR inhibitors and epilepsy: from basic science to clinical trials. *Expert Rev Neurother* **13**, 657-669 (2013).

134. M. E. Zanetti, I. F. Chang, F. Gong, D. W. Galbraith, J. Bailey-Serres, Immunopurification of polyribosomal complexes of Arabidopsis for global analysis of gene expression. *Plant Physiol* **138**, 624-635 (2005).

135. Y. Y. Zaytseva, J. D. Valentino, P. Gulhati, B. M. Evers, mTOR inhibitors in cancer therapy. *Cancer Lett* **319**, 1-7 (2012).

136. Y. Zhang *et al.*, An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-11947 (2014).

137. Y. Y. Zhu, E. M. Machleder, A. Chenchik, R. Li, P. D. Siebert, Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**, 892-897 (2001).

138. Q. Zong, M. Schummer, L. Hood, D. R. Morris, Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc Natl Acad Sci U S A* **96**, 10632-10636 (1999).

139. Y. Zong, Karaletsos, T., Drewe, P., Sreedharan, V.T.T., Singh, K., Hans-Guido, W., Ratsch, G., RiboDiff: Detecting Changes of Translation Efficiency from Ribosome Footprints. *bioRxiv* **10.1101/017111**, (2015).