University of Glasgow

The development of ALICE-tRNA-sequencing and its use in exploring the role of tRNAs in translational control

Thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Bashir Abdirahman Mohamed

Beatson Institute for Cancer Research

University of Glasgow

2021

CANCER RESEARCH UK | BEATSON INSTITUTE

Abstract

Bashir A. Mohamed, Beatson Institute for Cancer Research, University of Glasgow, G61 1BD

Sustaining proliferative signalling and loss of translational control is arguably the most fundamental trait of cancer cells, enabling tumour growth and metastatic dissemination. Transfer RNAs (tRNAs) have long been considered abundant “housekeeping” RNAs, functioning to decipher the universal genetic code. However, exhaustive analyses have implicated tRNA participation in a host of regulatory networks including the cellular stress response and protein synthesis. Recent findings suggest that the expression of tRNAs for synonymous codon usage is dependent on the differentiation/proliferation status of the cell and are coordinated with changes in translation. Although the molecular mechanisms that govern these changes are yet to be elucidated, cellular tRNA composition potentially introduces an additional layer of translational control. tRNAs are the most post-transcriptionally modified RNA species, with well over 50 unique modifications identified in eukaryotes. Consequently, isoacceptor identification and the measuring of the tRNA pool using next generation sequencing has long been an area of interest, with many attempts being made in literature. Using the Escherichia coli dealkylating enzyme AlkB and the novel tRNA high throughput sequencing methodology ALICE-tRNA-seq, we have developed a methodology that can accurately measure relative tRNA pools in vitro and in vivo. We show how other published tRNA sequencing protocols show bias towards tRNA sub populations, with our method showing a more realistic distribution across all tRNAs. We also show relative distribution changes in cellular and genetically modified mouse models of cancer, opening up a high resolution approach to establish the role of tRNAs in translational control and cell fate decisions.

Table of Contents

List of Figures

List of Tables

Abbreviations

3’ UTR	3’ Untranslated region
5’ UTR	5’ Untranslated region
A-site	Acceptor site
ALICE-tRNA-seq	Adapter Ligation Circularization Relinearization tRNA sequencing
APS	ammonium persulfate
ATP	Adenosine triphosphate
cDNA	complementary DNA
CDS	Coding sequence
CPM	Counts per million
DEAD-box	Asp-Glu-Ala-Asp motif
DDX6	DEAD-box helicase 6
DMEM	Dulbecco’s Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal Bovine Serum
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate

GTF	Gene transfer format
HEK293	Human Embryonic Kidney 293 cells
L-glut	L-glutamine
Log2FC	Log2 fold change
m1A	N1-methyladenosine
m1C	N1-methylcytosine
m1G	N1-methylguanosine
m22G	2,2-dimethylguanosine
mRNA	messenger Ribonucleic acid
ns	not significant
NGS	Next generation sequencing
p-body	processing body
PABP	poly(A) binding protein
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PNK	Polynucleotide kinase
PCR	polymerase chain reaction
RNA	Ribonucleic acid
rRNA	Ribosomal RNA

RT	reverse transcription
TAE	Tris-acetate-EDTA
TATA	AT rich promoter (8bp)
TE	Tris-EDTA
TBE	Tris-borate-EDTA
TEMED	Tetramethylethylenediamine
TPM	Transcripts per million
tRNA	Transfer Ribonucleic acid
tRNAi	initiator tRNA

Acknowledgements

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

قُلْ سِيرُوا۟ فِى ٱلْأَرْضِ فَٱنظُرُوا۟ كَيْفَ بَدَأَ ٱلْخَلْقَ ۚ ثُمَّ ٱللَّهُ يُنشِئُ ٱلنَّشْأَةَ ٱلْءَاخِرَةَ ۚ إِنَّ ٱللَّهَ عَلَىٰ كُلِّ شَىْءٍ قَدِيرٌ

Say, [O Prophet,] "Travel throughout the land and see how He originated the creation, then Allah will bring it into being one more time. Surely Allah is the Most Capable of everything.

Ankaboot 29:20

I would like to take this opportunity to thank Martin Bushell and all the members of his lab at the Beatson Institute for Cancer Research, University of Glasgow. All the members of this lab contributed and helped towards this project.

I would like to thank the members of the Jim Norman and Tom Bird labs for providing many of the samples used in this project.

I would like to thank Joseph Waldron, Tobias Schmidt, Ewan Smith and Ania Wilczynska in particular for their help and input into this project.

I would also like to thank my parents for all they have done for me.

Chapter 1: Introduction

1.1 The central dogma of molecular biology

The central dogma, namely, the transcription of DNA to RNA in the nucleus and its translocation to the cytoplasm to be translated into protein has been a central theme in the field of molecular biology for over 50 years. Over the years, the complexity of this theme has given rise to topics such as epigenetics, transcriptional control, post-transcriptional regulation, translational control and post-translational control. The development of the field from its inauguration to now has seen it move from the understanding that gene expression is only controlled at the level of transcription, to the idea that protein output can be regulated at all levels. This complexity is further exemplified by the host of technologically advanced high-throughput sequencing methodologies that exist today. These include but are not limited to; whole-genome sequencing (a method used to analyse the entire genome), ChiP-sequencing (chromatin immunoprecipitation; a method used to identify genome wide protein-DNA binding sites), methylation sequencing (a method used to identify cytosine methylation sites genome-wide), RNA-sequencing (a method used to measure RNA abundance), ribosome profiling (a method used to profile ribosome binding), iCLIP (individual-nucleotide resolution Cross-Linking and immunoprecipitation; a method used to identify RNA-protein interactions). It is clear today that the dynamic and complex nature of transcriptional and translational control to regulate cellular homeostasis and cell fate decisions allows for the fine-tuning of protein synthesis, to meet the differing demands of every type of cell and tissue.

1.2 The need to study translational control

The final step of the gene expression pathway is the translation of mRNAs to proteins and represents the synthesis of the proteome from genomic information. Failure to regulate this step is the hallmark of many diseases, ranging from neurological disorders (Buffington et al., 2014), metabolic disorders (Morita et al., 2014), immunodeficiency (Lucas et al., 2016; Piccirillo et al., 2014) and cancer (Bhat et al., 2015; Truitt and Ruggero 2016; Robichaud et al., 2019).

Uncontrolled and sustained proliferation is a well-established hallmark of cancer (Hanahan and Weinberg, 2011). With 20% of cellular energy being dedicated to a host of cationic pumps, 15% to DNA replication and transcription and 20% to protein synthesis, a considerable amount of resources is dedicated to protein output (Buttgereit and Brand, 1995). Moreover, further cellular energy is dedicated to the translation of mRNAs, since the majority of transcription is targeted towards the synthesis of mRNAs encoding ribosomal proteins and ribosomal RNAs (rRNAs) (Rolfe and Brown, 1997). Therefore, highly malignant cancers further increase protein synthesis mediated energy consumption due to their requirement of increased ribosomal content and translation, in order to maintain continuous and rapid proliferation (Silvera et al., 2010). Since most tumour cells suffer from conditions such as nutritional deprivation and hypoxia (conditions that normally result in the downregulation of mRNA translation), they manage to find a way to bypass traditional homeostasis and uncouple translational regulation during the transformation of the tumour (Robichaud et al., 2019). The proliferation, survival and metastasis of cancer cells, as well as their ability to hijack and change the translational landscape of the cell, makes the study of

translational control an important factor in the identification and development of effective therapeutics.

1.3 An overview of eukaryotic translation

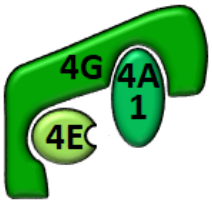
During the process of mRNA transcription, RNA triphosphatase catalyses the removal of the γ-phosphate from the 5' triphosphate, after approximately 30 nucleotides are transcribed (Ramanathan et al., 2016). This results in the transfer of guanosine monophosphate (GMP) on to the transcribing transcript from guanosine triphosphate (GTP). The methylation of the N7 amine of the newly formed cap then results in m7G-capped mRNA (Ramanathan et al., 2016). Once the newly transcribed mRNA successfully undergoes post transcriptional modifications such as splicing and polyadenylation, the mRNA translocates to the cytoplasm, where it can be translated by the ribosome (Bentley, 2014).

Translation of the mRNA is generally categorised into four steps; translation initiation, elongation, termination and ribosome recycling (Bhat et al., 2015). During translation initiation, ribosomal components involved in protein synthesis recognise the start codon and assemble (Sonenberg et al., 2009). The translation initiation step predominantly regulates protein synthesis (Richter and Sonenberg, 2005). Aminoacyl transfer RNAs (charged tRNAs; i.e. tRNAs bound to their designated amino acids), with the correct anticodon decoding the nucleotide sequence of the mRNA, then shuttle the required amino acid to the ribosomal complex (Agris, 2004). The first required aminoacyl-tRNA is initiator methionine (iMet-tRNA), which binds to GTP-bound eukaryotic initiation factor 2 (eIF2), via the methylated adenosine at position 58 of iMet-tRNA. eIF2-bound iMet-tRNA then proceed to assemble with the 40S ribosome subunit (with iMet-tRNA in the P-site), eIF1, eIF1A, eIF3, and eIF5, resulting in the formation of the 43S pre-initiation complex (Agris, 2004).

The aforementioned m7G-cap on the 5’ end of the mRNA proceeds to recruit the eIF4F complex, which consists of the DEAD-box helicase eIF4A, the cap binding protein eIF4E and eIF4G which enables cap-binding (Ramanathan et al., 2016). The eIF4G protein can bind polyA binding protein (PABP), causing the mRNA to circularise (Walsh et al., 2008). The eIF4F complex then recruits the 43S pre-initiation complex, through its interaction with eIF3, and proceeds to scan the 5’ untranslated region (5’ UTR) of the mRNA until is identifies an AUG codon (start codon). Upon recognition of the AUG start codon, eIF2 and other initiation factors are released from the mRNA, resulting in the recruitment of the 60S ribosomal subunit (Hinnebusch et al., 2016). The fully assembled ribosome then proceeds to translate the coding sequence of the mRNA, recruiting tRNAs that can decode the coding sequence, bringing with it the correct amino acid. Individual amino acids are covalently bound sequentially, forming the resultant polypeptide chain (Jackson et al., 2018) (Figure 1.1).

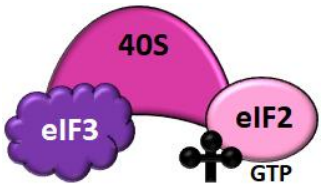
Translation initiation is considered to be the rate limiting step of translation for most mRNAs (Hinnebusch et al., 2016). Therefore, cell fate determination, cancer development and progression are often associated with the altered expression (due to these proteins being downstream of oncogenic signalling) or mutations in the genes of the initiation factors involved in this process (Jackson et al., 2018).

A



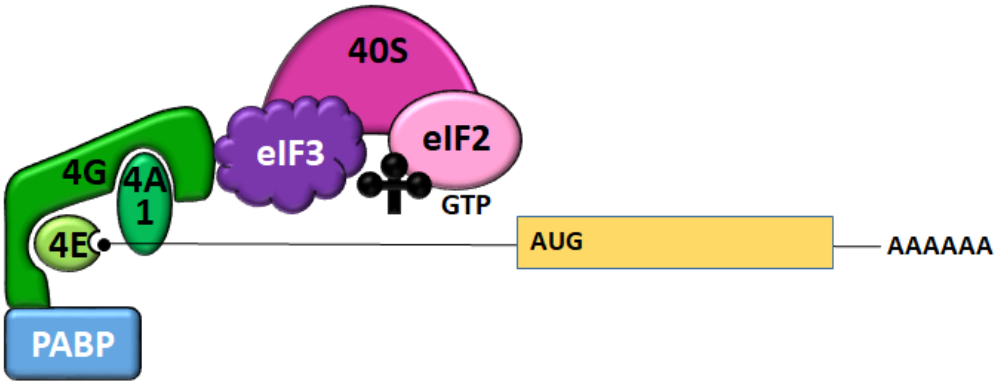
eIF4F complex

B

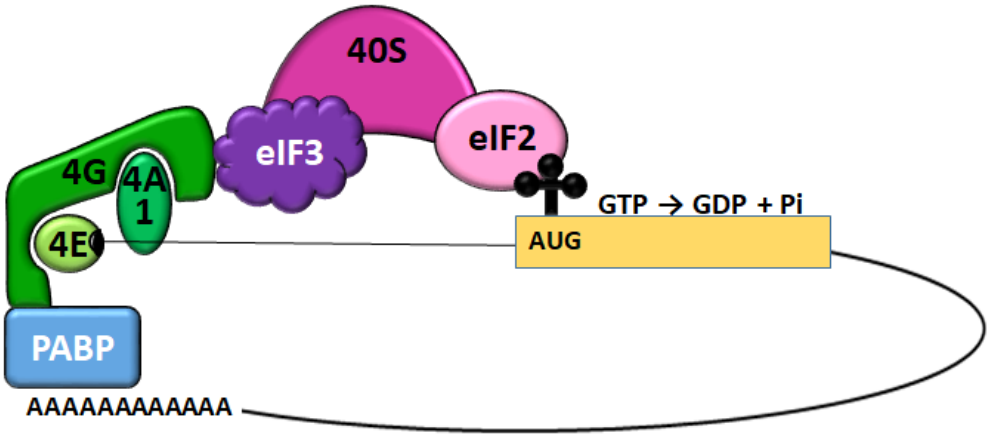


43S pre initiation complex

C



D



E

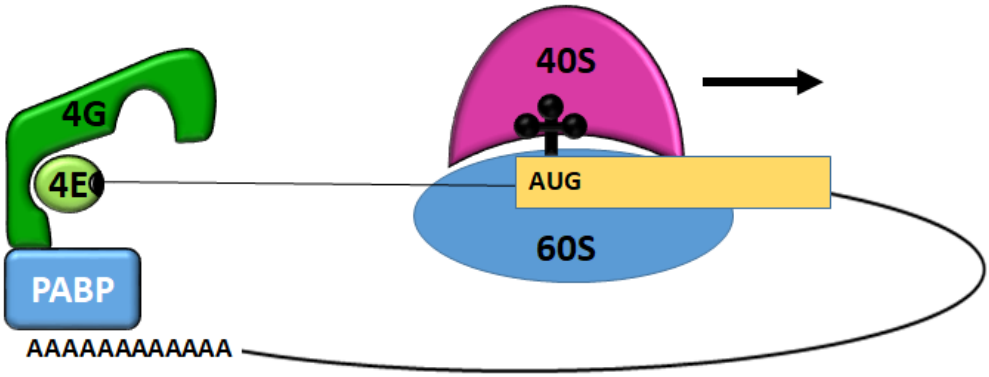


Figure 1.1 Overview of eukaryotic translation. (A) Schematic of the eIF4F complex which consists of the DEAD-box helicase eIF4A1, the cap binding protein eIF4E and eIF4G which enables cap-binding. (B) Schematic of the 43S pre-initiation complex. The first required aminoacyl-tRNA is initiator methionine (iMet-tRNA), which binds to GTP-bound eukaryotic initiation factor 2 (eIF2), via the methylated adenosine at position 58 of iMet-tRNA. eIF2-bound iMet-tRNA then proceed to assemble with eIF3, and eIF5, resulting in the formation of the 43S pre-initiation complex (C) eIF4F recruits the 43S pre-initiation complex (D) Scanning of the mRNA until recognition of the AUG start codon. (E) Recruitment of the 60S subunit and the decoding of the coding sequence.

1.4 tRNA biology and its significance in translational control

Upon the discovery of the universal genetic code, tRNAs were understood to decode their cognate base pairs on the mRNA sequence and deliver the required amino acids to the A-site of the ribosomes, enabling polypeptide synthesis. After comprehensive studies on the structure of tRNAs and their complexes, it was thought that the understanding of tRNA biology was complete (Schimmel, 2017). However, advances in technologies such as high-throughput sequencing (in particular, RNA-seq and proteomics), led to the discovery of the role of tRNAs in adaptive protein synthesis (Li et al., 2011; Javid et al., 2014) and their ability to function as non-coding RNAs in a multitude of regulatory networks (Pekarsky et al., 2016). Furthermore, numerous disease states have been found to coincide with abnormality in aspects of tRNA biology, including tRNA mutations and mutations to the auxiliary proteins that are pertinent to tRNA biogenesis and modifications (Abbott et al., 2014; Blanco and Frye, 2014; Schon et al., 2012; Suzuki et al., 2011). This has resulted in the increased attention into tRNA biology to try and establish the emerging roles of tRNAs in adaptive translation, signalling dynamics and disease.

1.4.1 tRNA biogenesis

The biogenesis of tRNAs consists of many steps including transcription, 5' leader removal, 3' trailer trimming, the splicing of introns, the addition of 3' CCA residues and the covalent modification of a host of nucleoside residues (Phizicky and Hopper, 2010).

In the nucleoplasm, the transcription factor TFIIIC binds to the intragenic A-box and B-box of tRNA genes (which encodes the D-stem and T-stem of the tRNA), resulting in the recruitment of the transcription factor TFIIIB to the upstream region of the tRNA gene (Kirchner and Ignatova, 2015). TFIIIB consists of three subunits; TBP (TATA-binding protein), BDP1 (B-related factor 1) and BRF1 (B double prime 1) and directs the recruitment of RNA polymerase III (pol III), resulting in the transcription of the tRNA (Kirchner and Ignatova, 2015).

Pol III is negatively regulated by the protein Maf1, which was shown to be involved in tRNA-mediated nonsense suppression in yeast (Murawski et al., 1994; Moir et al., 2006). Maf1 has been shown to be conserved throughout eukaryotes, however, Maf1 has been observed to negatively regulate Pol I and Pol II transcription in mammalian cells (Pluta et al., 2001; Reina et al., 2006). Maf1 binds directly to the TFIIIB transcription factor (Desai et al., 2005; Rollins et al., 2007) and Pol III (Gavin et al., 2006; Oficjalska-Pham et al., 2006). Maf1 is regulated via the mTOR and PKA pathways via phosphorylation by the PKA kinase or the mTOR-dependent kinase Sch9 (Huber et al., 2009; Lee et al., 2009; Wei et al., 2009).

Pre-tRNAs that are abnormally transcribed, (i.e. missing 5' leaders and/or 3' trailers), are eliminated during the nuclear surveillance pathway via the degradation of their 3' end by RNase Z in the nucleus (Vogel et al., 2005). Furthermore, tRNAs that lack the required

modifications in the cytosol are degraded via their 5' end by the endonuclease RNase P (Walker and Engelke 2006).

After transcription of the required tRNAs, 5' end processing follows to remove the 5' leader sequence. This is achieved by the endonuclease RNase P, a ribonucleoprotein (RNP) consisting of 10 protein subunits in addition to its RNA component (Walker and Engelke 2006).

tRNA 3' end processing follows, where the 3' trailer sequence is removed from the original transcript in order to facilitate the addition of the CCA sequence to the 3' end. This is achieved by the enzyme RNase Z (also known as tRNase Z), which recognises the N73 discriminator base at the 3' end of tRNAs and cleaves prior to the addition of the CCA bases (Vogel et al., 2005). Interestingly, RNase Z has been shown to have an inability to cleave tRNAs with a mature CCA end, giving the CCA end an anti-determinant characteristic (i.e. an inability to bind its cognate amino acid) (Mohan et al., 1999; Li de la Sierra-Gallay et al. 2006). This is still an area of tRNA biology yet to be fully elucidated. Although the function of RNase Z in tRNA 3' maturation is well documented (as well as its association with γ-tubulin), the link between the RNase Z gene and the increased risk of prostate cancer is still unclear (Tavtigian et al., 2001).

tRNA introns exist on a minority of tRNA genes and are typically found between nucleotides 37 and 38 of the tRNA (Phizicky and Hopper, 2010). Genomic sequencing has revealed that introns exist on at least one family of tRNA genes in all organisms and is a conserved feature (more information on this can be found at the Genomic tRNA Database, http://gtrnadb.ucsc.edu; Chan and Lowe, 2009). Splicing is catalysed by the tRNA splicing endonuclease (TSEN) and the process is inherently simpler than that of spliceosome-

mediated mRNA splicing. This is in part due to the fewer number of proteins involved in the process (Popow et al., 2012). TSEN cleaves intron sites resulting in a 3’ tRNA half with a 5’ hydroxyl group (5’-OH) and a 5’ tRNA half with a 3’ 2’–3’ cyclic phosphate. tRNA ligase then ligates the two ends, and the tRNA is ready for CCA addition (Popow et al., 2012).

tRNA nucleotidyltransferase (CCA-adding enzyme) is responsible for the addition of CCA residues to the 3’ end of tRNAs (Lizano et al., 2007). tRNA nucleotidyltransferase is a unique RNA polymerase due to its ability to synthesize the CCA sequence, without the need of a nucleic acid template (Lizano et al., 2007). The CCA sequence on the 3’ end serves two functions. Firstly, it is the site where aminoacyl tRNA synthetases (aaRSs) covalently bind the cognate amino acid to allow shuttling to the translating ribosome (Ibba and Söll, 2000). Secondly, it is only after charging the tRNA that it is allowed to enter the A-site of the ribosome, therefore fine-tuning its interactions with other structural elements, constraining its structural features (Giegé et al., 2008).

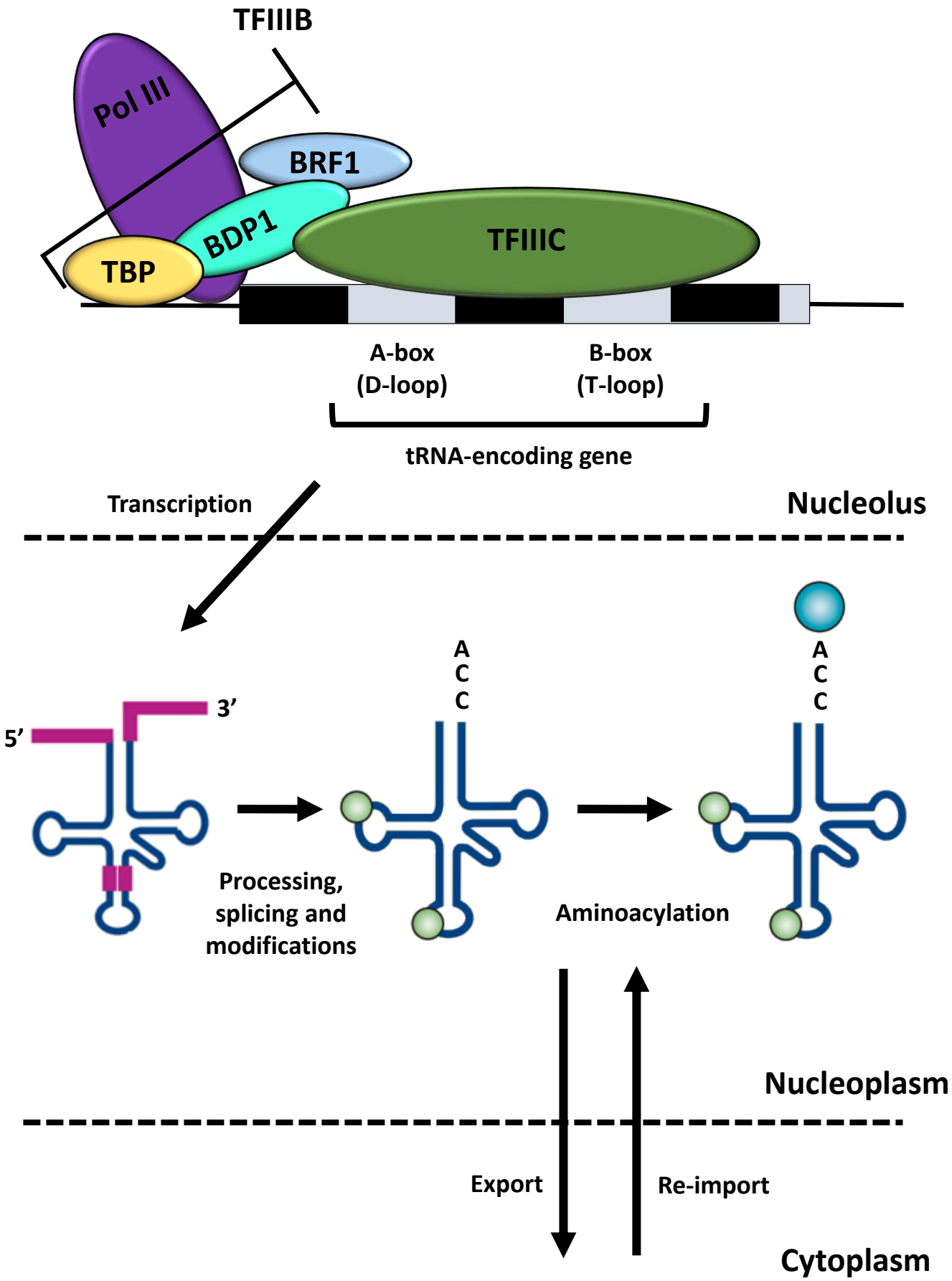

TFIIIB
Pol III
BRF1
BDP1
TBP
TFIIIC
A-box
(D-loop)
B-box
(T-loop)
tRNA-encoding gene
Transcription
Nucleolus
5’
3’
A
C
C
A
C
C
Processing,
splicing and
modifications
Aminoacylation
Nucleoplasm
Export
Re-import
Cytoplasm

Figure 1.2 Overview of tRNA biogenesis. In the nucleolus, the transcription factor TFIIIC binds to the intragenic A-box and B-box of tRNA genes (which encodes the D-stem and T-stem of the tRNA), resulting in the recruitment of the transcription factor TFIIIB to the upstream region of the tRNA gene. TFIIIB consists of three subunits; TBP (TATA-binding protein), BDP1 (B-related factor 1) and BRF1 (B double prime 1) and directs the recruitment of RNA polymerase III (pol III), resulting in the transcription of the tRNA gene. Pre-tRNAs are then processed (addition of the 3' CCA), spliced and modifications are added (green circles). tRNAs then proceed to be aminoacylated (blue circle is amino acid) before being exported into the cytoplasm (Adopted from Kirchner and Ignatova, 2015).

1.4.2 tRNA structure

The canonical tRNA has 76 nucleotides (Rich and Raj-Bhandary, 1976). It is characterized by its classical clover-leaf like secondary structure that can range from 76-93 nucleotides in length, with the 3' end having the newly added CCA sequence. Starting from the tRNAs 3' end, the four arms of the clover-leaf structure are called the acceptor loop, the TΨC loop (Ψ represents pseudouridine – often referred to as the T loop), the anticodon loop and the dihydrouridine loop (D loop) (Holley et al., 1965). When the tRNA folds into its tertiary structure, it forms a 12 base pair TΨC helix, resulting from when the TΨC loop stacks onto the acceptor loop, forming an L-shaped structure (Cramer et al., 1969). Furthermore, the anticodon loop stacks onto the D loop, forming a 10 base pair stem, resulting in the formation of the anticodon/D-loop dumbbell (Crothers et al., 1972).

The newly formed tertiary structure now has two domains, which are joined together via non-canonical pairing of conserved nucleotides from the D-loop and TΨC loop and are joined at right angles (Schimmel, 2017). These conserved nucleotides are essential since they are involved in this intricate tertiary structure formation, resulting in the stabilization of the final L-shaped tertiary structure that is recognised by the ribosome (Schimmel, 2017). tRNAs that have a length greater than 76 are often due to the variable region of the tRNA. However, these excess bases do not interfere with the L-shapped tertiary structure interactions, since they bulge out the back of the tertiary structure (Sigler, 1975). Interestingly, the acceptor stem that attaches the amino acid and the anticodon (which decodes the mRNA codons) are segregated, allowing for the decoding of the mRNA on one end of the tRNA and polypeptide synthesis on the other within the ribosome (Schimmel, 2017).

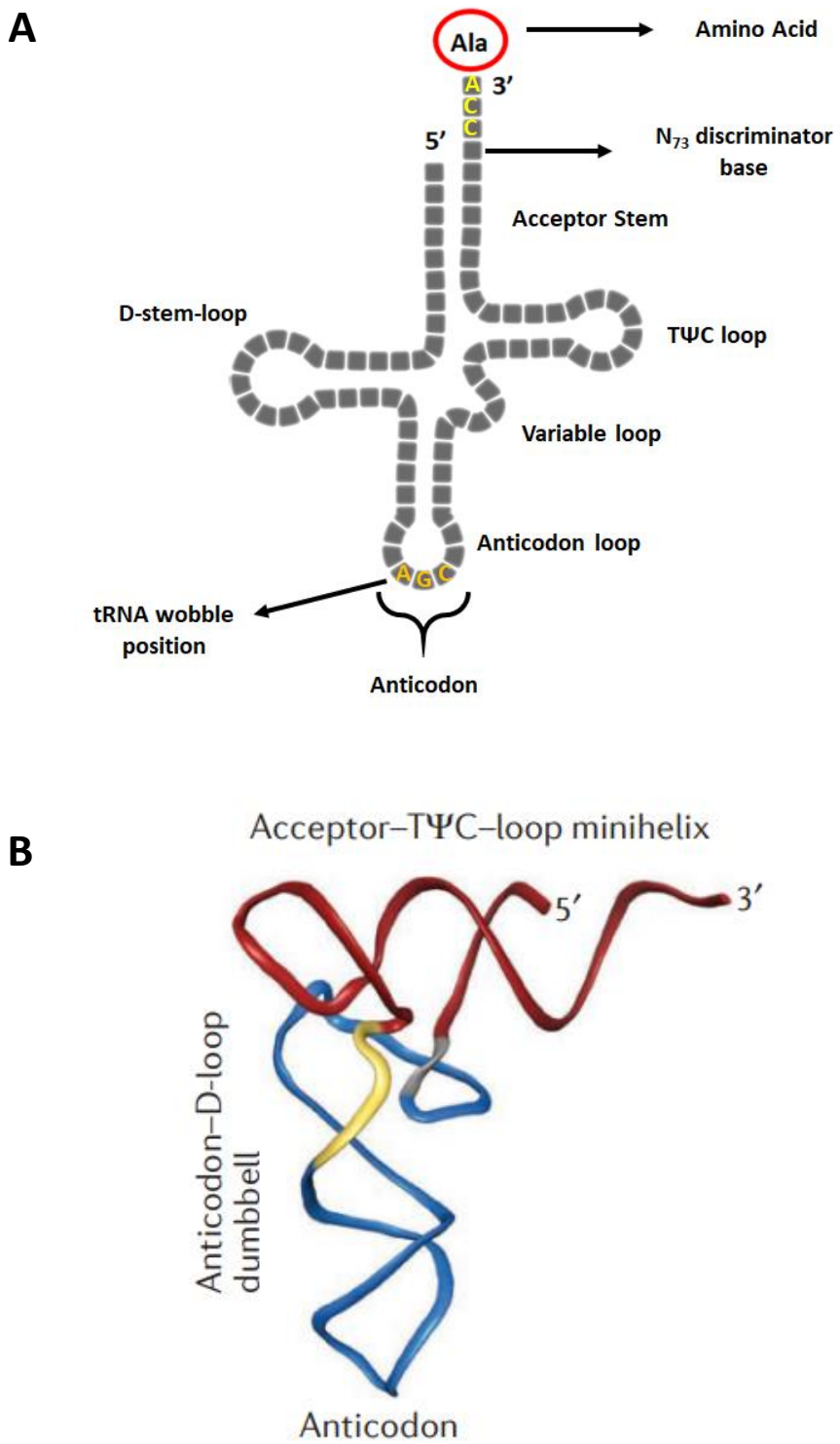


Figure 1.3 tRNA structure. (A) Canonical secondary structure of the tRNA (tRNA-alanine-AGC used as an example). It is characterized by its classical clover-leaf like secondary structure. Starting from the tRNAs 3' end, the four arms of the clover-leaf structure are called the acceptor loop, the TΨC loop, the anticodon loop and the dihydrouridine loop (D-stem loop). 3' terminal CCA is coloured yellow and the anticodon is coloured orange. (B) Tertiary structure of the tRNA. The tertiary structure has two domains, which are joined together via non-canonical pairing of conserved nucleotides from the D-loop and TΨC loop and are joined at right angles. Adopted from Schimmel, 2017

1.4.3 tRNA complexity: isoacceptors, isodecoders, degeneracy and near-cognate binding

Of all RNAs expressed, tRNAs are by far the most abundant (Cantara et al., 2011). Although the reason why individuals vary in the number of tRNA genes is yet to be elucidated (Iben et al., 2014). In the human genome, there are approximately 500 tRNA and tRNA gene-like sequences (Parisien et al., 2013; Abe et al., 2014). With over half of these expressed genes being validated (and more will likely be validated in the future), these expressed genes categorise into 300 different cytoplasmic tRNA sequences and 22 mitochondrial tRNA sequences, which make up the cytoplasmic and mitochondrial tRNA pools respectively (Schimmel, 2017). This therefore results in there being at least 1 tRNA species for every amino acid (see Table 4.1).

tRNAs are generally referred to in two ways; tRNAs that decode the same amino acid but only differ in the anticodon sequence (isoacceptors), and tRNAs that share the same anticodon sequence, but have different body sequences (isodecoders).

During translation, the function of the tRNA is to deliver amino acids to the ribosome so they can be added to the growing polypeptide chain. This is achieved by tRNAs decoding the coding sequence of the mRNA and forming a correct mRNA codon-tRNA anticodon interaction. Due to the degeneracy of the genetic code, more than one codon can code for the same amino acid (with the exception of methionine and tryptophan) (Cantara et al., 2011). Isoacceptors are defined as tRNAs that differ in (but not exclusive to) the anticodon sequence, but deliver the same amino acid (Goodenbour and Pan, 2006). In higher eukaryotes, there is also the presence of isodecoders – tRNAs that have the same anticodon sequence but differ in body sequence (Geslain and Pan, 2010).

The complexity of tRNAs is multiplied further by the existence of the tRNA's own degeneracy rules. The adenosine to inosine modification that exists at position 34 of tRNAs (wobble position) that encode alanine, arginine, isoleucine, leucine, proline, serine, threonine and valine all utilise non-cognate base pairing to decode codons (Maria and Arimbasseri, 2017). This is due to the fact that all of these amino acids can be encode by a codon, whose cognate anticodon containing isoacceptor is not expressed in the genome. For example, the mRNA codon GCC can only be decoded by an A34I edited tRNA-Ala-AGC, as no tRNA with the anticodon GGC is expressed in humans. Inosine has the ability to base pair with adenosines, cytosines and uridines and so an IGC anticodon would be able to decode either GCC, GCU or GCA mRNA codons. The second degeneracy rule that tRNAs employ is G:U base pairing (between the tRNA wobble position and the 3rd position of the mRNA codon) (Stadler and Fire, 2011). For example, CAU codons, which code for histidine, have to be decoded by tRNA-His-GUG as there is no tRNA expressed in the human genome that contains the anticodon AUG . Therefore, CAU codons would be decoded by tRNA-His-GUG by employing G:U base pairing at the wobble position. More examples of these can be found in Table 4.1.

Another form of near-cognate binding comes in the form of tRNA selection at the ribosomes (Tarrant and Von der Haar, 2014). tRNAs are translocated to the ribosome via the phosphorylation of eIF1A, selecting tRNAs carrying the correct amino acids to the A-site of the translating ribosome (Rodnina and Wintermeyer, 2009). However, there are instances when incorrect amino acids are delivered due to near-cognate complementarity between the tRNA anticodon and the mRNA codons (usually when the second nucleotide is not completely complementary), resulting in increased dwell time in relation to the translating ribosome (Tarrant and Von der Haar, 2014). Near-cognate binding at the ribosome A-site

has been shown to decrease translational efficiency (Pape et al., 1999). Furthermore, the ratio between cognate and near-cognate tRNA abundances have been an area of immense interest in the tRNA community, as they could play a significant role in the determination of codon decoding times (Tarrant and Von der Haar, 2014). The relationship between cognate/near-cognate ratios in relation to tRNA abundance and translational efficiency could yield a regulatory function within the cell, and has therefore become a significant area of interest.

1.4.4 tRNA modifications

With the majority of discovered, modified human nucleosides being attributed to tRNAs (over 90), the complexity of their participation in a host of regulatory networks comes as no surprise (Cantara et al., 2011; Saikia et al., 2010). These modifications can range from those that are present in every tRNA, such as dihydrouridine (D) or pseudouridine (Ψ), to others that are restricted to a particular family or group of tRNAs (Schimmel, 2017). Studies conducted on human HEK293T cells and using mathematical models estimated that the average human tRNA consists of approximately 11-13 modified bases (Saikia et al., 2010). Furthermore, others have shown that these modifications are not "all or nothing" with variations in modifications shown to range from 10%-80% depending on the location of the base and the type of modification being observed (Clark et al., 2016). These modifications often determine whether a tRNA molecule is able to participate in translation or any of its additional functions (such as translocation and tRNA regulation) (Schimmel, 2017).

Therefore, a tRNA base being modified is a binary event – the base can either be modified or not. This results in the potential for a host of microspecies existing within a single cell. For example, if the 13 reported modifications that, if absent, result in a certain disease (reported by Torres et al., 2014, Frohlich et al., 2016) were to be considered, the computation of the number of possible tRNA modification combinations is 8192 (213). Given that there are approximately 60 million tRNA species in a single cell (Lodish and Darnell, 1995), then it would seem at first glance that the large collection of microspecies calculated could be accommodated for many times over. However, if the same calculation was to be performed for every isodecoder, then the complexity would build rapidly (Schimmel, 2017).

If the statistical distribution of all possible combinations of modified bases across all tRNAs were to be considered, the number of hypothetical possibilities would be incalculable, and would not be able to be sustained in an entire organism, let alone a single cell. If one was to consider that there were 213 combinations of modified microspecies for a single amino acid-specific tRNA (i.e. considering all possible isodecoders per amino acid), then the total number of combined modifications would easily surpass 10^{100} (2^{13} x 2^{12} x 2^{11} …). This highlights some of the immense complexities that exist in tRNAs and it is no surprise that they are the source of a lot of unanticipated biology (Schimmel, 2017).

The biggest challenge in the tRNA modification field is the inability to accurately measure modifications at a single nucleotide resolution. tRNA modifications have been identified through two dimensional thin layer chromatography (2D-TLC) (Keith, 1995). Here, a combination of P1 nuclease, venom phosphodiesterase and RNase T2 is used to digest the tRNAs into their individual nucleosides. Then, the digested nucleosides are placed on a 2D chromatogram and all nucleosides separate according to their chemical characteristics, depending on the separating solvents used. The limitation to this technique is that you would either have to know what chemical modifications you are intending to find, or would have to use trial and error to elucidate an observed chemical position on the chromatogram. Although you could in theory, calculate the relative modification differences between two biological conditions using input concentration and absorption, you would never be able to work out which tRNA harboured the modification and at what position.

More modern techniques in elucidating tRNA modifications would be the use of high performance liquid chromatography-coupled mass spectrometry (HPLC-MS) (Su et al., 2014). This group used yeast (Saccharomyces cerevisiae) that was stress induced using

hydrogen peroxide to quantitatively measure the modification profile of tRNAs in wild-type and stress conditions. Briefly, total RNA was extracted from yeast cells and the tRNAs were purified using reverse phase HPLC (due to tRNA's unique chromatographic signature). Once tRNAs were purified, they were hydrolysed into their individual nucleosides using a combination of benzonase, phosphodiesterase and alkaline phosphatase before being loaded onto a tandem triple quadruple mass spectrometer (QQQ-MS). Here, nucleosides were ionised and run through a series of quadrupoles to separate the nucleosides according to their chemical characteristics, before being detected and analysed using the high sensitivity dynamic multiple reaction monitoring (DMRM). Relative proportions of modified nucleosides can then be calculated, to give you relative quantitative differentials between biological conditions. Su et al identified 25 tRNA ribonucleoside modifications in *S. cerevisiae*, and claimed a direct link between the translation of stress related proteins and stress induction. Although HPLC-MS is far more advanced and offers high sensitivity quantification in comparison to 2D-TLC, it can only output global modification differentials between biological conditions and does not have the resolution to tell you the type of tRNA that is modified and at what position.

Zheng et al attempted to use a next generation sequencing approach to solve the limitation of tRNA resolution in identifying modified tRNA bases using DM-tRNA-seq (Zheng et al., 2015). For humans, only a select number of modifications and their positions on the tRNA are known (Saikia et al., 2010). tRNAs have traditionally been difficult to sequence due to modifications on the Watson-Crick face of tRNAs (i.e. modifications that exist on the region where complementary base pairing occurs - such as methyl-guanosine, methyl-adenosine and more) (Motorin et al., 2007). This poses a problem because the concept behind next generation sequencing is the production of a cDNA library, which is then amplified using PCR

and subsequently sequenced. Due to modifications present on the Watson-Crick face of tRNAs, traditional reverse transcriptases like SuperScript III would stall when they reach said modifications. To overcome this, Zheng et al utilized a modified library preparation in order to identify the positions of modified bases and to which tRNA group they belonged to. Firstly, they used the Escherichia coli enzyme AlkB to demethylate the methyl groups of modifications that existed on the Watson-Crick face of tRNAs, an enzyme originally known to facilitate in DNA and RNA repair (Trewick et al., 2002; Falnes et al., 2002). They also utilized a mutant form of AlkB (D135S), where an aspartic acid was mutated to serine at position 135 (coinciding with the active site of the protein). The shorter side chain of serine was thought to accommodate bigger modifications like the dimethyl-guanosine, a common modification found in tRNA sequences (Zheng et al., 2015). Then, tRNAs were prepared for sequencing by ligating a 3' adapter to tRNAs, using a DNA-RNA dimer. Once the DNA strand was ligated to the tRNA, it was reverse transcribed using a high fidelity thermostable group II intron reverse transcriptase (TGIRT). TGIRT attaches to the DNA-RNA dimer and uses a process known as template switching, to read the tRNA bases and transcribe them. When TGIRT reaches a modified base, it places a random nucleotide in its place instead of stalling like other traditional reverse transcriptases. When the cDNAs are generated, they are PCR amplified and sequenced. When the tRNAs are mapped, the positions where modifications exist would be identified as mismatches. The advantage of using a methodology like DM-tRNA-seq is the fact that you could identify both the exact position of a modification and the type of tRNA it exists on, while being highly quantitative. However, its limitation is that you would not know what the modification is and the methodology is only as powerful as the known modification profiles that exist (Saikia et al., 2010).

In theory, the best approach to studying tRNA modifications would be the use of the Oxford Nanopore PromethION sequencing technology (Deamer et al., 2016). The Oxford Nanopore sequencing technology works by binding a motor protein to the DNA or RNA samples, which in turn binds to one of the thousands of nanopores in the flow cell. Once bound, the DNA/RNA sequence passes through an electrically resistant membrane via the nanopore which disrupts the current across the membrane when a voltage is applied. This disruption is measured and every base would have its own chemically unique signature, allowing the identification of the traditional bases (A/T/C/G/U) as well as modified bases (Deamer et al., 2016). The unique aspect of the Oxford Nanopore sequencing technology is that it would not require any hydrolysis/digestion reaction to occur on tRNAs prior to sequencing, as it can sequence RNA directly. Therefore, one could determine both a potential modification, as well as knowing the position it occurs at and the type of tRNA it exists on. The current limitation with the Oxford Nanopore sequencing technology however, is the reported error rates in the final sequences outputted (Lu et al., 2016; Johnson et al., 2020). With a median error rate ranging from 5% - 20% per read, in comparison to traditional next generation sequencing (such as Illumina) which has a median error rate of approximately 0.5% (Manley et al., 2016; Johnson et al., 2020), the error rate is far too high. Perhaps as technology advances over the next decade, the Oxford Nanopore sequencing technology could be a real solution in identifying and quantifying differential tRNA modification profiles between biological conditions.

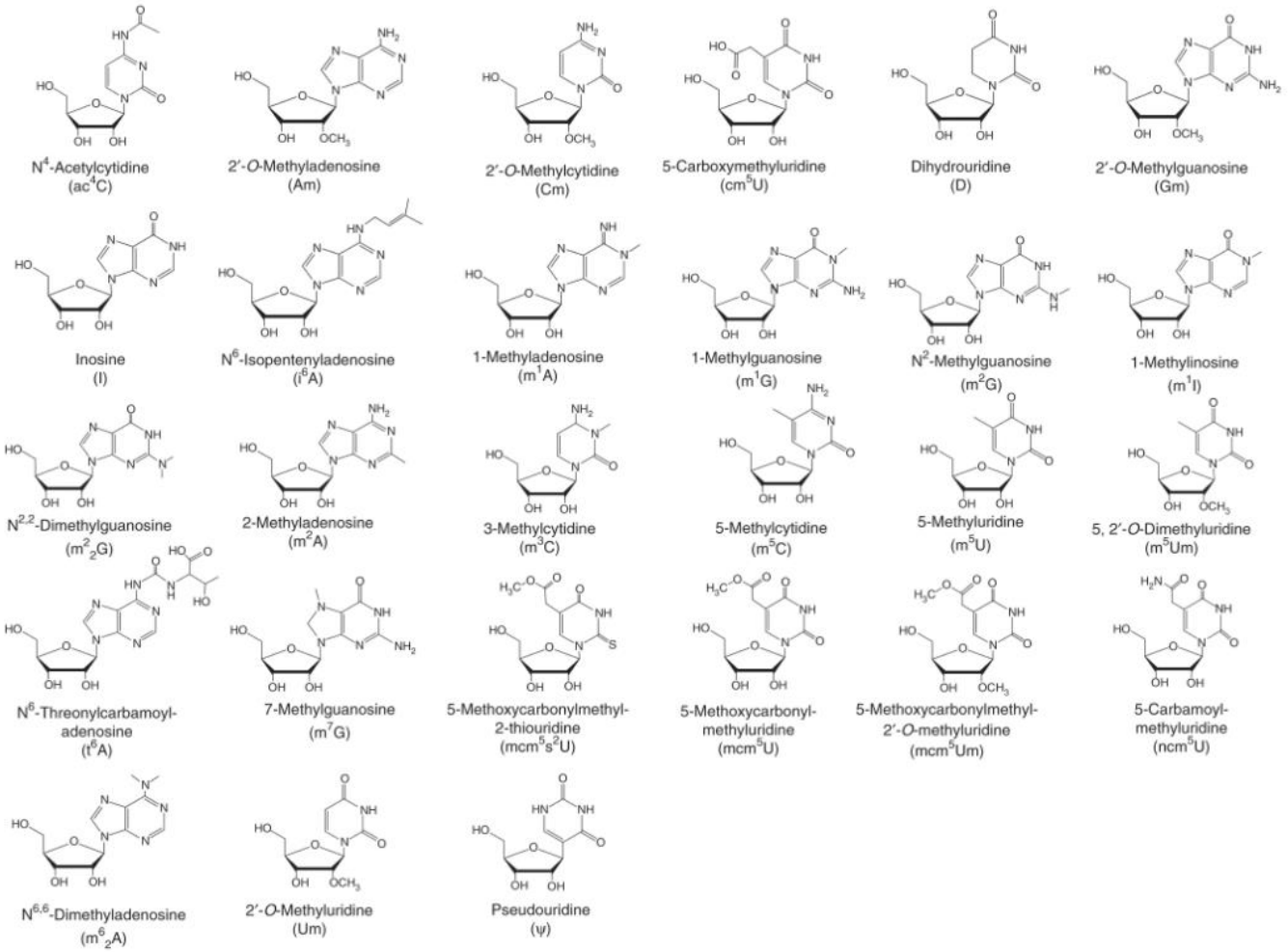


Figure 1.4 tRNA modifications. Some of the tRNA modifications identified by Su et al., 2014. tRNAs were identified using high performance liquid chromatography-coupled mass spectrometry (HPLC-MS).

1.4.5 tRNA supply vs demand

The ready to translate tRNA population has to stay dynamic for several reasons. Firstly, the necessity for tRNAs needs to correspond with codon usage of actively translating mRNAs (Dittmar et al., 2006). Secondly, the gene expression signature of the cell is dependent on physiological and environmental conditions, and varies from tissue to tissue. Therefore, the tRNA pool has to be able to dramatically change in order to meet the demands of the transcriptome (Rak et al., 2018). Furthermore, the codon usage in mRNAs can depend on the gene sets required to be translated, depending on cellular activity (Botzman & Margalit 2011). Gingold et al highlighted this when they showed a difference in codon usage in gene sets associated with vertebrate cell differentiation, in comparison to proliferation associated genes (Gingold et al., 2014).

The supply-to-demand ratio, i.e. the balance between tRNA availability and actively translating codons, seems to depend on translational efficiency and cellular wellbeing (Rak et al., 2018). In unicellular organisms, the choice of codons in the genome (codon demand) correlates highly with tRNA gene copy numbers and measurements in tRNA levels (Dong et al., 1996, Percudani et al., 1997). Similar correlations were observed in tissues with varying cellular states in higher eukaryotes (Gingold et al., 2014; Dittmar et al., 2006). Furthermore, others have shown that the codon usage of highly expressed genes does bias towards codons that match the most abundantly expressed tRNAs (Sharp & Li 1987; Kanaya et al., 1999).

These correlations could be explained in various ways. Firstly, the use of efficiently translated and optimal codons result in certain proteins being highly expressed due to

higher translational speeds (Sorensen et al., 1989; Gardin et al., 2014). Also, codon optimization may be evolutionarily driven, resulting in highly expressed genes exhibiting favourable codons and non-optimal codons being selected against (Frumkin et al., 2018). Because the efficient expression of genes is paramount to the functionality of the cell, the balance between tRNA availability and its coordination with transcriptomic demand also becomes paramount, even though the exact mechanism of how this is achieved is yet to be elucidated (Rak et al., 2018). Depending on the phenotypic diversion of organisms or lifestyle, it is possible that evolution may have optimized gene expression at the level of translational efficiency to match organism demand (Zaborske et al., 2014; Botzman & Margalit 2011; Man & Pilpel 2007; Jiang et al., 2008).

1.4.6 Methodologies to quantitatively measure tRNA abundance and charging

Given the importance of tRNA availability in meeting the transcriptomic demand, a methodology to accurately measure the tRNA-ome has long been an area of interest. However, the reason for trying to measure the tRNA-ome is just as important since the methodologies in literature attempt to answer different questions. In general, tRNAs can be surveyed to investigate either abundance, modifications or charging (Figure 1.5) (Pan, 2018). Methodologies for measuring tRNA modifications were discussed in Section 1.4.4. To date, there is no accurate and unifying methodology that can be used to determine all three of these factors.

Before the emergence of next generation sequencing methodologies, the traditional way of quantitatively measuring tRNA abundance was via the use of hybridization-based microarrays (Dittmar et al., 2006; Gingold et al., 2014). For the microarray approaches, plates with fluorescently labelled sequence probes designed against tRNA sequences (70-80nt in length) are generated. When samples are loaded, tRNAs bind to their complementary sequences. Upon binding, the fluorescently labelled probe emits photons which can be measured and the absorption can be used to quantify relative tRNA levels (Dittmar et al., 2006). The limitations in using hybridization-based approaches are twofold. Firstly, tRNA modifications that exist on the Watson-Crick face of tRNAs can hinder hybridization, therefore biasing quantitation towards tRNAs that are less modified. Secondly, hybridization probes would not be able to distinguish tRNAs that are very similar in sequence. For example, the human genome has 22 genes and 15 sequences for tRNA-alanine-AGC (Pan, 2018). Some of tRNA-alanine-AGC's isodecoders (tRNAs with the same anticodon but different body sequences), only differ by 1 nucleotide. Hybridization probes would therefore not be able to distinguish between many isodecoders and isoacceptors

(Zheng et al., 2015; Cozen et al., 2015; Gogakos et al., 2017; Pinkard et al., 2020). Next generation sequencing methodologies were therefore developed to try and overcome the limitations observed in microarrays.

As mentioned previously, tRNAs have traditionally been difficult to sequence due to modifications on the Watson-Crick face of tRNAs (such as methyl-guanosine, methyl-adenosine and others) (Motorin et al., 2007). This poses a problem because the concept behind next generation sequencing is the production of a cDNA library, which is then amplified using PCR and subsequently sequenced.

Two very similar tRNA-sequencing methodologies reported in literature are ARM-seq (Cozen et al., 2015) and Hydro-seq (Gogakos et al., 2017).

In ARM-seq, Cozen et al utilized the Escherichia coli enzyme AlkB to demethylate the methyl groups of modifications that existed on the Watson-Crick face of tRNAs. Once tRNAs were demethylated, 3' and 5' adapters were ligated to the tRNAs and SuperScript III was used to facilitate reverse transcription. Once the cDNA library was generated, it was PCR amplified and sequenced.

In Hydro-seq, Gogakos et al opted against using AlkB to demethylate modified bases on the Watson-Crick face of tRNAs. Instead, they opted to fragment the tRNAs into 20-30 nucleotide fragments (similar to what is done in traditional RNA-seq). The rationale was that fragmentation would decrease the likelihood of stalling events in the reverse transcription stage and like traditional RNA-seq, the fragments could be used to determine which tRNAs were captured bioinformatically, through multiple iterations of mapping (Gogakos et al., 2017).

Although ARM-seq was developed to sequence tRNA fragments, it is not a suitable methodology to measure global tRNA abundance due to the bias introduced by using 3' and 5' adapter ligation steps (discussed in further detail in Section 3.5.3 and 3.6). Due to the 5' end of the tRNA being involved in the hairpin secondary structure, access to the 5' end of the tRNA becomes challenging. This has also been reported for other small RNAs with secondary structure based problems at their 5' ends due to hairpins affecting library preparation (Liu et al., 2014; Burke et al., 2014; Lama et al., 2019). The Hydro-seq methodology suffers from a similar limitation to that seen in the hybridization-based approaches. Because the tRNAs are being fragmented (and the fragmentation process cannot be controlled since it uses alkaline hydrolysis), isodecoder determination becomes a problem since some isodecoders only differ by 1-2 nucleotides. Furthermore, if the fragmentation results in there being a modification halfway along the fragment, the RT step will still stall and those reads would automatically be discarded. Upon further examination of both the ARM-seq and Hydro-seq sequencing results (discussed in further detail in Section 3.5.3 and 3.6), it was clear that almost 60% of their reads mapped to tRNAs encoding three amino acids.

QuantM-seq attempted to solve this 5' adapter ligation bias by using an annealed adapter where the 5' adapter had an overhang that would be complementary to the CCA of the 3' end of the tRNA, aiding 3' adapter ligation (Pinkard et al., 2020). Once tRNAs were deacylated and demethylated (using a commercial version of wild-type AlkB), the 3' end of the annealed adapter was designed to have both the reverse transcriptase primer binding sites and divergent PCR primer binding sites. Upon ligation, the tRNAs would be reverse transcribed, followed by circularization, and then PCR amplifying the circular cDNA to generate the final library. Truncation events due to the reverse transcriptase stalling would

still result in those fragments being amplified during PCR, since the 3’ adapter already contained the RT and PCR primer binding sites, as opposed to ARM-seq and Hydro-seq, which requires reads to reach the 5’ end of the tRNA in order for the cDNA to be PCR amplified (Pinkard et al., 2020).

The limitation with the QuantM-seq methodology is the lack of use of the D135S AlkB demethylase enzyme. The mutant AlkB is modified so that an aspartic acid is mutated to serine at position 135 (coinciding with the active site of the protein). The shorter side chain of serine was thought to accommodate bigger modifications like the dimethyl-guanosine, a common modification found in tRNA sequences (Zheng et al., 2015). The importance of using the D135S AlkB in combination with the wild-type ALKB is discussed in further detail in Section 3.5.3 and 3.6. Furthermore, the other limitation is that QuantM-seq does not distinguish between mature tRNAs and genuine 3’ halves and 3’ fragments. Furthermore, an extra CCA is added to the 3’ end of tRNAs when they are targeted for degradation (Wellner et al., 2018). Although these reads could easily be discarded bioinformatically, there is no mention of this in their paper. QuantM-seq lacks an ability to distinguish the actual tRNA pool that is participating in translation, as opposed to non-functional tRNAs.

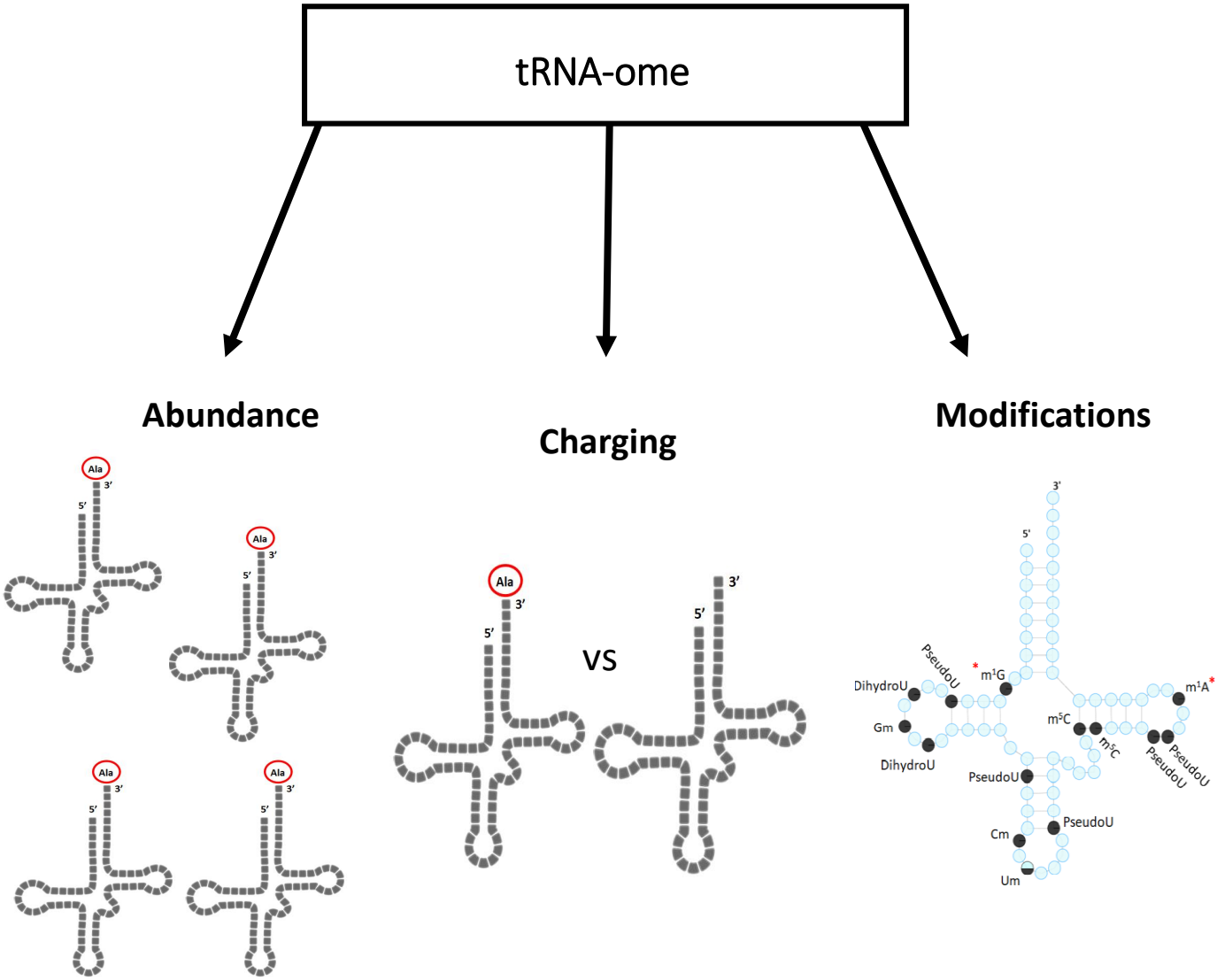
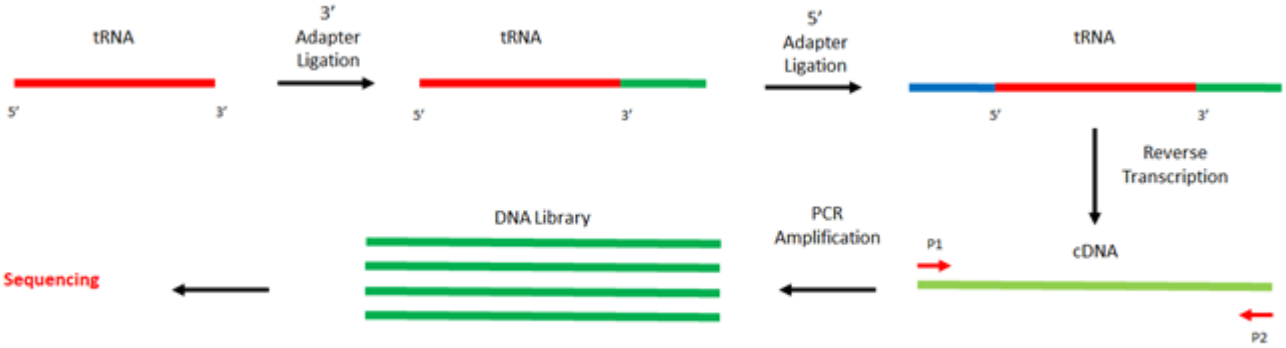


Figure 1.5 A true representation of the tRNA-ome. A true representation of the tRNA-ome would have to take into consideration the abundance of available tRNAs, a measurement of which tRNAs are charged (if one wants to identify ready to translate tRNAs) and the modification profile of said tRNA. In conjunction, the true tRNA landscape can be surveyed.

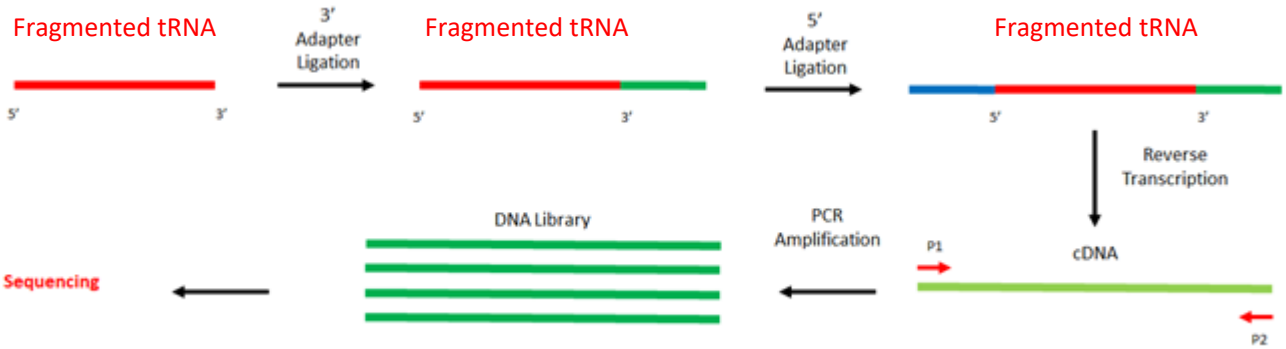
A

Arm-seq



B

Hydro-seq



C

QuantM-seq

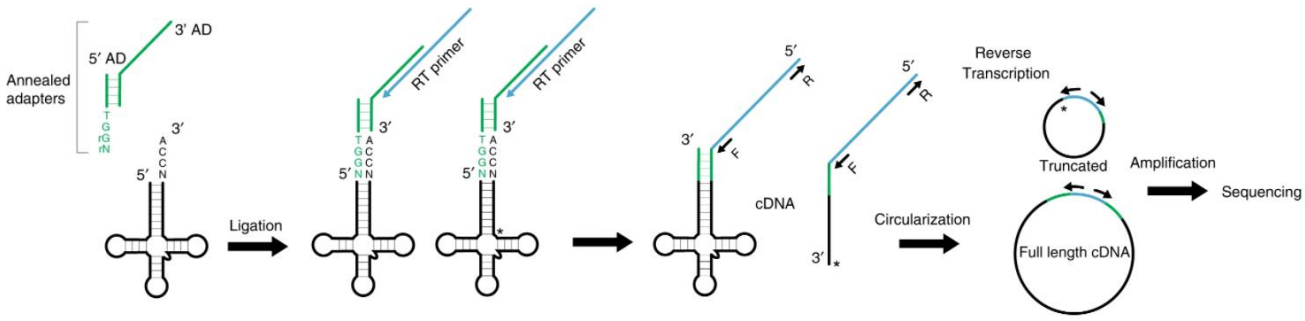


Figure 1.6 Published tRNA-sequencing methodologies. (A) Schematic representing the tRNA library preparation for ARM-seq utilized by Cozen et al., 2015 (B) Schematic representing the tRNA library preparation for Hydro-seq utilized by Gogakos et al., 2017. (C) Schematic representing the tRNA library preparation for QuantM-seq utilized by Pinkard et al., 2019 (also adopted from the same paper). Truncation events (*) due to the reverse transcriptase stalling would not be a problem since the 3' adapter already contained the RT and PCR primer binding sites, as opposed to ARM-seq and Hydro-seq, which requires reads to reach the 5' end of the tRNA in order for the cDNA to be PCR amplified

1.5 Codon optimality

Codon optimality is often used to describe how a cohort of codons on an mRNA sequence contribute to mRNA stability and rates in translation elongation. As discussed in Section 1.4.5, the nuance between tRNA availability, highly expressed genes and the mechanism behind their regulation in humans is an area of immense interest.

mRNA stability may play a vital role in the regulation of codon optimality. Studies into the DEAD-Box helicase DDX6 in mammalian cells showed that mRNAs enriched in A/U ending codons localised to p-bodies where they are kept translationally repressed, which was and dependent on DDX6. Furthermore, through mRNA stability, DDX6 was shown to regulate G/C-rich mRNAs (Courel et al., 2019). Courel et al therefore suggested that codon optimality may be linked with mRNA storage and translational repression, as well as mRNA stability (Courel et al., 2019). Since the optimality of codons is dependent on the translational environment of a cell, the codons that might be directing mRNA turnover has been shown to vary across studies (Presnyak et al., 2015; Courel et al., 2019; Wu et al., 2019; Forrest et al., 2020). The codon composition and distribution of the coding sequence of mRNAs has been shown to influence mRNA stability and translation rates in a host of organisms including Escherichia coli (Frumkin et al., 2018; Boël et al., 2016), yeast (Radhakrishnan et al., 2016; Presnyak et al., 2015), zebrafish (Bazzini et al., 2016; Mishima et al., 2016), mouse (Guimaraes et al., 2020) and human cells (Bornelöv et al., 2019; Hia et al., 2019; Wu et al., 2019; Forrest et al., 2020).

Quantifying the effect of codon optimality and its impact on mRNA stability and translation has been difficult to define. Quantitative metrics include the Codon Adaptive Index (CAI),

the tRNA Adaptive index (tAI) and the Codon Stabilization Coefficient (CSC) (Presnyak et al., 2015). The Codon Adaptive Index uses highly expressed genes to calculate metrics based on the assumption that elevated tRNA expression is proportional to elevated levels in gene expression (Sharp et al., 1986). The tRNA Adaptive index metric modified CAI and assumed that tRNA abundance correlates with tRNA gene copy number, and takes the efficiency of interactions between mRNA codons and the tRNA wobble position into account (dos Reis et al., 2004; Rocha et al., 2004). Since CAI and tAI both represent static codon optimality measures, the Codon Stabilization Coefficient metric improves on these by calculating the Pearson correlation coefficient between global mRNA half-lifes and codon frequencies within a cellular condition (Presnyak et al., 2015; Carneiro et al., 2019). Because the CSCs are calculated for every condition, they are directly dependent on the mRNA half-lives within that cellular condition, which therefore gives a more cell type specific metric of codon optimality. However, in order to obtain a more accurate representation of tRNA demand, mRNA abundance must be taken into consideration, since expression levels can vary across a large range.

1.6 tRNAs and their role in cell fate decisions

There has been a large focus on the role of tRNAs in cell fate decisions, particularly revolved around proliferation and the translation of proliferation-promoting proteins. The tRNA-ome and mRNA-ome varies between tissues (Hernandez-Alias et al., 2020; Dittmar et al., 2006; Pinkard et al., 2020). The codon usage of the mRNA-ome has been shown to differ between tissues and is conserved between humans and mice and reported to often match with the tRNA-ome (Plotkin et al., 2004). Furthermore, translational efficiency for tissue-specific mRNAs have been shown to be impacted by their codon optimality (the codon usage of the most highly expressed mRNAs coincides with the most abundant tRNAs) (Waldman et al., 2010). Non-optimal codons on the other hand, have also been shown to have a greater prevalence in certain biological processes such as the circadian clock (Xu et al., 2013; Zhou et al., 2013) and cell cycle genes, which display distinct codon usage patterns in different cycle phases (Guimaraes et al., 2020; Frenkel-Morgenstern et al., 2012). The working hypothesis is that these distinct codon usage signatures that exist, depending on the tissue and cellular events, could be means for the temporal regulation of gene expression via translational upregulation or repression.

The investigations into proliferation-associated protein expression has shown some correlation with the expressed tRNA population. As mentioned previously, Gingold et al observed that tRNA availability coincided with the upregulation of proliferative mRNAs as well as histone modification changes around tRNA genes and elucidated towards a potential transcriptional programme in the regulation of tRNA and mRNA expression (Gingold et al., 2014). They also observed that proliferative mRNAs were enriched for A/U ending codons in both cancerous and non-cancerous proliferative samples and hence introduced

“proliferative tRNAs” (with proteins associated with pattern specification enriching for G/C ending codons). However, when they clustered the transcriptomes of their samples, it was clear that the samples clustered according to cell/tissue type rather than the proliferative status of the sample. Furthermore, the group utilized hybridization-based tRNA assays to measure tRNA availability (the limitation of which are described in Section 1.4.6).

Guimaraes et al conducted investigations into the effects of serum-deprivation in mouse embryonic fibroblasts and also reported the enrichment of A/U ending mRNA codons but found no differential tRNA expression. This led to them proposing an alternate model; rather than “proliferative tRNAs” being upregulated to meet the transcriptomic demand, there is a global tRNA expression increase to meet the translational boost required in proliferative cells. This could therefore overcome a potential bottleneck in the translation of proliferative mRNAs that are enriched in A/U ending codons (Guimaraes et al., 2020).

Bornelov et al conducted investigations into self-renewing and differentiating stem cells and suggested that the increased translational rates of proliferative mRNAs were down to increases in adenosine to inosine modifications at the wobble position of tRNAs. Since inosine can facilitate in the decoding of non-cognate mRNA codons (potentially compensating for lowly expressed cognate tRNAs), this may be allowing for the increased translation rates of the proliferative mRNAs in self-renewing cells in comparison to differentiating stem cells (Bornelov et al., 2019).

Although the mentioned studies may have revealed potential pathways that result in the translation of non-optimal mRNAs in normal conditions, there is still no distinct mechanism observed in the regulatory dynamic between tRNA availability and mRNA codon usage. This

may in part be due to the lack of a precise methodology to quantitatively measure the tRNA-ome and therefore important information about key tRNAs may be overlooked.

Another gap in the literature is the focus on codon signatures and tRNA availability in the two other major cell states, namely, quiescence and senescence. Cellular quiescence is often defined as cells in the G0 phase of the cell cycle and are characterised by their ability to reverse cell cycle arrest (Luo et al., 2020). Cellular senescence is characterised by permanent proliferative arrest that is often triggered by stimuli such as organelle stress, oncogene activation, telomere dysfunction and DNA damage (Hayflick and Moorhead, 1961; Di Micco et al., 2021). To get a better perspective into how codon optimality and tRNA availability are coordinated, it may be beneficial to identify the tRNA and mRNA signatures in all three major cell states (quiescence, proliferation and senescence). Furthermore, there is a lack of studies using in vivo models. Although in vitro investigations are easier and cheaper to conduct, most codon usage investigations in higher eukaryotes hypothesise that mRNA and tRNA signatures may be tissue specific (Botzman & Margalit 2011). Although Gingold et al looked into the sequencing data of patient tumour samples, there was no real controls into patient gender and age, both of which could output different genetic signatures. Overall, in vivo investigations using mouse models may provide a more global account into the tRNA and mRNA signatures of cells/tissues which may be closer to reality.

1.7 Project aims

As discussed in this chapter, the need to accurately measure the tRNA pool and categorise both the tRNA and mRNA signatures in quiescence, proliferation and senescence (*in vitro* and *in vivo*) would aid in trying to decipher the molecular mechanisms coordinating regulation between tRNA availability and codon usage.

This thesis aims to answer the following questions:

- Can a tRNA-sequencing methodology that avoids the limitations and biases recognised in literature be developed to accurately quantify the tRNA pool?
- Can such a tRNA-sequencing methodology be used successfully both in vitro and in vivo?
- Can such a tRNA-sequencing methodology claim to be more effective then published tRNA-seq methodologies?
- Through tRNA-seq and RNA-seq, can the tRNA and codon signatures of in vitro and in vivo samples be categorised in proliferation, quiescence and senescence
- Can this data be used to identify patterns in tRNA and mRNA signatures in the three major cell states

Chapter 2: Methods and Materials

2.1 Cell culture

2.1.1 HEK293

Passage 14 Human embryonic kidney (HEK293) cells were grown in Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher) and supplemented with 10% Fetal Bovine Serum (FBS, Sigma) and 10% L-glutamine (L-Glut, Sigma). All cells used in the *in vitro* experiments were grown to a confluency of 70%, confirmed by light microscopy, before being processed. When the cells were grown to required need, media is removed and cells treated with 4ml trypsin and incubated at 37°C for 5mins to remove the cells from the plate. 6ml media was added to deactivate trypsin. Cells were spun at 4000rpm for 5min and the pellets washed with PBS. The same process is repeated and the pellets were stored at -80°C until processing.

2.1.2 IMR90s

IMR90 primary human fibroblasts were kindly provided by the Norman Lab, Beatson Institute for Cancer research. Viral supernatant was produced by seeding 70-80% confluent HEK293 cells in 10cm dishes and transducing with either the control vector (pLXSN-control, Clonetech) or the ER:RAS oncogene-induced senescence vector (pLNC-ER:RAS, Addgene). The transfection mixture used consisted of 800µl DMEM, 75µg PEI (polyethylenimine), 2µg VSVg envelope protein plasmid (Clonetech), 8µg Gag.pol packaging plasmid and 20µg of either control or ER:RAS virus. The transfection mixture was briefly vortexed and incubated at room temperature. HEK293s had their media changed and the transfection mix added, incubating for 16 hours. New media was added and incubated further until 24 hours. Then the media was changed again but only 6ml fresh media was added to concentrate the viral

supplement (This is repeated three times in tandem with IMR90 media change). After 48 hours, the 6ml media is collected and filtered through a 0.45µm cellulose acetate membrane. IMR90 cells are seeded 24 hours prior the last media change to coordinate the viral supplement collection and IMR90 media change. The 6ml media collected from the HEK293s are supplemented with 4µg/ml polybrene. When IMR90 media is changed, the 6ml media with viral supplement is added the IMR90 media. This is repeated three times every 3 hours. After the third repeat and IMR90s have been infected, they were grown for 72 hours before being passaged and seeded for antibiotic selection. 400µg/ml neomycin is added to a 1:4 dilution of both IMR90 sets (control and ER:RAS virus), and not allowing the ER:RAS IMR90s to reach confluence. Both sets are then seeded to generate stock cells.

Induction of oncogene induced senescence is achieved by treating both control and ER:RAS cells with 100nM 4-OHT (tamoxifen) and are grown for 10 days, with media changes every two days. After the 10^{th} day, media is removed and cells treated with 4ml trypsin and incubated at 37°C for 5mins to remove the cells from the plate. 6ml media was added to deactivate trypsin. Cells were spun at 4000rpm for 5min and the pellets washed with PBS. The same process is repeated and the pellets were stored at -80°C until processing.

2.2 Genetically modified mouse models

All mouse models used were kindly provided by the Bird Lab, Beatson Institute for Cancer Research. All genetically modified mouse models used were male, between the ages of 8-10 weeks and had similar weights at the time of being celled. All replicates were done together for each experiment to reduce technical variability.

2.2.1 βcatenin/c-MYC model

C57BL/6 mice were generated using wild-type AAV-Cre ($2x10^{11}$GC/mouse) or $Ctnnb1^{ex3}$ heterozygous $R26^{LSL\text{-}myc}$ homozygous with AAV-Cre ($2x10^{11}$GC/mouse). This introduced LoxP sites around exon 3 of βcatenin and the insertion of the humanised $R26^{LSL\text{-}myc}$. All mice were monitored three times per week up until induction. Cre-recombinase activity was induced using the adenovirus AAV8.TBG.PI.Cre.rBG (Addgene, Catalog number: 107787-AAV8; a gift from James M. Wilson, Addgene viral prep # 107787-AAV8; RRID:Addgene_107787), specifically targeting hepatocytes. $2x10^{11}$GC/mouse high dose virus in 100µl PBS was treated to each mouse via tail vein injection. After 4 days, the mice were culled using CO_2 with subsequent cervical dislocation. The caudate lobe of the liver was dissected and snap frozen immediately in liquid nitrogen. The tail sample was taken for retrospective genotyping (which all mice had passed). The caudate lobes were stored at -80°C until processed.

2.2.2 MDM2 model

C57BL/6 $AhCre^{WT}$ mice were crossed with both C57BL/6 $MDM2^{fl/fl}$ and $MDM2^{fl/+}$ to generate $AhCre^{+}$ $MDM2^{fl/fl}$ (introducing LoxP sites around exons 5 and 6 of the MDM2 gene). $AhCre^{WT}$ mice were used as the control and $AhCre^{+}$ $MDM2^{fl/fl}$ for the MDM2 induced senescence experiments. Cre recombinase activity was induced by treatment with the hepatocyte specific AAV8-TBG-Cre adenovirus; $2x10^{11}$GC/mouse high dose virus in 100µl PBS via tail

vein injection. After 4 days, the mice were culled using CO_2 with subsequent cervical dislocation. The caudate lobe of the liver was dissected and snap frozen immediately in liquid nitrogen. The tail sample was taken for retrospective genotyping (which all mice had passed). The caudate lobes were stored at -80°C until processed.

2.3 RNA methods

All RNA extractions for sequencing was performed using the mirVana™ miRNA Isolation Kit (ThermoFisher). Total RNAs were extracted according to manufacturer's instructions.

2.3.1 Harvested cells

Harvested cells were placed on ice and had 600µl of Lysis Binding Buffer (from the mirVana miRNA Isolation Kit) added. Samples were vortexed rigorously until a homogenous lysate was obtained. Organic extraction then starts at Step E of the mirVana miRNA Isolation Kit and both Total RNA and small enriched RNAs (RNAs < 200nt) protocols were performed. All samples were eluted in 100µl water and concentration was measured using a Nanodrop.

2.3.2 Snap frozen tissue homogenisation

Before extraction of RNAs from snap frozen tissues, all work surfaces and equipment were cleaned with 70% ethanol and RNaseZAP (ThermoFisher) to avoid any contamination. The only exception was equipment that was already sterile.

Snap frozen tissues were placed on dry ice when removed from -80°C storage. A Sterallin™ square petri dish (sterilized, Fisher Scientific) was placed onto the dry ice and the temperature was allowed to drop. Sterile forceps were used to transfer the tissue from the Eppendorf to the plate and a sterile disposable scalpel was used to cut off approximately 30-40mg of liver. Dissected liver was placed into a 2ml Precellys CK14 tube with ceramic beads (Bertin Instruments) for homogenisation.

A Precellys Evolution homogeniser (Bertin Instruments) was used to homogenise the tissue. 15-20mins prior to homogenisation, the Cryolys Machine was filled with dry ice and turned on to option 3 (maximum air flow). Compressed air was passed through the Cryolys Machine and into the Precellys Evolution homogeniser lowing its temperature to approximately 4°C.

Just prior to homogenisation, 600µl of Lysis Binding Buffer (from the mirVana miRNA Isolation Kit) was added to the samples. Homogenisation was done at 5500rpm in two cycles, with a 30s pause between cycles. Samples then proceed to organic extraction (Step E) of the mirVana miRNA Isolation Kit and both Total RNA and small enriched RNA (RNAs < 200nt) protocols were performed. All samples were eluted in 100µl water and concentration was measured using a Nanodrop.

2.3.3 tRNA aminoacylation

Once small RNAs were extracted from samples, tRNAs had to be deacylated to remove any bound amino acids to their 3' end prior to size selection. 10X Tris-HCl pH 9.0 was added to small RNA samples and incubated at 37°C for 45mins. RNAs were then column purified using RNA Clean and Concentrator 25 (Zymo) according to manufacturer's instructions.

2.3.4 tRNA demethylation

The CDS for the wild-type AlkB was cloned into the pET-SUMO vector, provided by Tobias Schmidt (Beatson Institute for Cancer Research). The D135S-AlkB variant was produced by site directed mutagenesis (PCR) and expressed in the same system. AlkB was produced in *Escherichia coli* BL21 (DE3). Pre-cultured cells were grown at 37°C overnight in 50µM kanamycin and 40µg/ml chloramphenicol for inoculation. Colonies were chosen for main culture and were grown in the absence of antibiotics. 1mM IPTG was added to induce AlkB overexpression and grown for 4 hours to an OD600 of 1. The cells were collected, pelleted and resuspended in lysis buffer (20mM Tris-HCl pH 7.5, 30mM Imidazol, 1M NaCl, and 50% glycerol), 1mM PMSF, 1mg/ml lysozyme and cOmplete EDTA-free (Roche) to inhibit protease activity. The cells were lysed by sonication and centrifuged at 75000g at 10°C for 30mins. The resultant N-terminal tagged 6x-His-SUMO-AlkB protein was passed through a

Ni-charged HisTrap column using the AKTA fast protein liquid chromatography (GE Healthcare) system. Bound protein was eluted using a linear gradient of up to 500mM Imadazol. Pooled protein was treated with ~500nM Ubiquitin-like Protease 1 (ULP1) at room temperature for 1 hour to cleave the SUMO tag. Protein was further purified by ion-exchange (HiTrap Heparin HP affinity column, GE Healthcare) and gel filtration (Superdex-200). Pooled protein was stored in AlkB buffer (20mM Tris-HCl pH 8.0, 50% glycerol, 0.2M NaCl and 2mM DTT) at -80°C.

AlkB demethylation was carried out in a 100µl reaction; 1µg small RNAs, 120pmol WT AlkB, 240pmol D135S AlkB, 50mM MES pH 5.0, 2mM L-Ascorbic Acid, 300µM α-ketogluterate, 50µM Ferrous Ammonium Sulphate heptahydrate, 50µg/ml Bovine Serum Albumin and nuclease free water to volume. Samples were incubated at room temperature for 2hrs and purified using RNA Clean and Concentrator-5 (Zymo) as per manufacturer's instructions.

2.3.5 Primer extension assay

Designed oligonucleotides are labelled using Adenosine 5'-triphosphate (γ-^{32}P-ATP, 10mM tricine pH 7.6, Perkin Elmer) in a 10µl reaction. The reaction consists of 5 units T4 polynucleotide kinase (PNK, New England Biolabs), 10X polynucleotide kinase buffer, 20pmol primer and nuclease free water to volume. Samples are incubated at 37°C for 30mins. Once labelling is complete, the PNK enzyme is inactivated by incubating at 95°C for 5mins. A further 15µl of nuclease free water is added to take the volume to 25µl. The samples are then purified using G25 microspin columns (Sigma) according to manufacturer's instructions.

100ng tRNAs are used as input into the primer extension assay. Prior to the assay, 100ng RNA, 1µl of the labelled primer and nuclease free water to 6µl and incubated at 95°C for

2mins and then is placed on ice immediately. The reverse transcription mixture is then prepared on ice which consists of 200U SuperScript III, 10X first strand buffer (Thermo Scientific), 0.5mM free nucleotides (dNTPS), 5mM DTT and nuclease free water to 6µl. The two 6µl volumes are combined and incubated at 55°C for 1 hour.

Once the reverse transcriptase reaction is completed, samples are run on a 10% TBE-urea gel (see 2.4.2) for 50mins at 180V (or when the blue formamide dye reaches the bottom of the gel). Once the gel is completed, the gel is fixed in 10% acetic acid/methanol solution for 5mins. The gels are dried for 2 hours at 80°C on a gel dryer dock and is exposed on a phosphor cassette (can vary from 2hr to 16hrs). The phosphor screen is then scanned after exposure using a Typhoon scanning machine.

2.4 RNA and DNA quantification and integrity

2.4.1 Agarose gels

All total RNAs extracted for RNA-seq were first run on a 1% agarose gel to check for RNA integrity. The gels were made by adding 1g agarose (Melford Laboratories Ltd) to 100ml 1x TAE (Tris-acetate-EDTA) and microwaved for 1min and 30sec. When the mixture cools, 10µl of Syber Safe (Thermofisher) was added to visualise the bands before casting. 1x TAE (Tris-acetate-EDTA) was used as the running buffer. 1µg of RNA was added to 2x formamide dye (95% formamide, 20mM Tris HCl pH 7.5, 20mM EDTA and 0.025% bromophenol blue) in a 15µl volume. Samples were loaded onto the gel and ran at 200V for 30mins. HyperLadder™ 1kb (Bioline) was run alongside samples as per manufacturer's instructions. Gels were visualised using a Chemidoc (Bio-Rad).

2.4.2 TBE gels

TBE-urea gels used for small RNA visualization and gel extraction steps were made up from a stock 40% UreaGel system Concentrate (National Diagnostics) and diluted down to the required percentage using the UreaGel system Diluent and 10X TBE (Tris-Boric Acid-EDTA) buffer in a 50ml volume (National Diagnostics). 200µl 25% APS (Ammoniumpersulfate) and 20µl TEMED (Tetramethylethylenediamine) were added to initiate polymerisation. All gels were pre-run at 200V for no less than 1 hour prior to use. All small RNA samples had 2x formamide dye added and heated at 90°C for 2mins prior to loading. A Low Molecular Weight Ladder DNA ladder was run alongside all samples (New England Biolabs).

TBE-gels used for the extraction of DNA libraries were prepared from a 40% 19:1 Acrylamide/Bisacrylamide solution (National Diagnostics) and diluted down to the percentage required. Gels were prepared in a 20ml volume and contained 5X TBE and

nuclease free water. 200µl 10% APS and 12µl TEMED were added to initiate polymerisation. All gels were pre-run at 200V for no less than 30mins prior to use. All samples had 2x formamide dye added and heated at 90°C for 2mins prior to loading. A Low Molecular Weight Ladder DNA ladder was run alongside all samples.

2.4.3 Qubit

The concentration of completed DNA libraries are measured using a Qubit double stranded DNA high sensitivity assay kit (Thermofisher Q32854), as per manufacturer's instructions. 2µl of the final library is used as input.

2.4.4 Library Size Determination

The average size of the final libraries are measured using a D1000 High Sensitivity screen tape (Agilent), as per manufacturer's instructions. 2µl of the final library is used as input. Samples were vortexed for 2000rpm for 1min prior to measurements.

2.5 Next Generation Sequencing

2.5.1 ALICE-tRNA-sequencing

The first step in the ALICE-tRNA-seq methodology is 3’ adapter ligation (3’ adapter = 5’pCAGATCGGAAGAGCACACGTCT-R-NH2). The reaction is prepared in the following way:

	Final Concentration / Volume
tRNAs	10pmol
3' Adapter	30pmol
T4 RNA Ligase 2 Reaction Buffer (NEB)	1X
T4 RNA Ligase 2 (1-249) K227Q (NEB)	200U
RNaseIn (Promega)	40U
PEG8000 (NEB)	20%
DMSO (NEB)	20%
Nuclease Free Water	to 50µl

The reaction is incubated at 16°C for 16 hours and the samples are purified using the RNA Clean and Concentrate 5 as per manufacturer’s instructions and eluted in 13.5µl nuclease free water.

Excess 3’ adapter depletion follows:

	Final Concentration / Volume
3' Adapter Ligated Product	12.5µl
NEB 5' Deadenylase (NEB)	25U
RecJ Buffer (NEB Buffer 2)	1X
RecJf (NEB)	15U
PEG8000 (NEB)	10%
RNaseIn	20U
Nuclease Free Water	to 20µl

The reaction is incubated in a thermocycler at 30°C for 1hr and then 37°C for 30mins and 1100rpm. Samples are purified using the RNA Clean and Concentrate 5 as per manufacturer's instructions and eluted in 12.1µl nuclease free water.

Reverse transcription follows. First, the following is assembled on ice and incubated at 95°C for 2mins and then place on ice immediately for at least 3mins (RT primer = 5'pGATCGTCGGACTGTAGAACTCTGAArCAGACGTGTGCTCTTCCGATCT).

	Final Concentration / Volume
3' Adapter Depleted Product	11.1µl
RT Primer	0.5µM
Nuclease Free Water	to 12µl

Then the following mix is added to the pre-heated samples:

	Final Concentration / Volume
First Strand Buffer	1X
dNTPs	0.5mM
DTT	5mM
RNaseIn	40U
SuperScript III	200U

The reactions are incubated at 55°C for 1hr. RT products are then ran on a 6% TBE-Urea denaturing Gel and are size selected between 75-150nt. Gel slices are crushed and 300µl 1x TE pH 8.0 is added and placed on a thermomixer overnight at 16°C, 550rpm. Spin-X column Centrifuge tubes (Sigma) are used to separate the gel pieces from the eluate and are spun at 14000rpm for 2mins. 50µl AMPure XP magnetic beads, and then immediately 350µl isopropanol are added to the eluate and incubated in a thermomixer at 25°C for 10mins, 550rpm. Then the solution is magnetized for 2mins and the supernatant is discarded. Then, 950µl of freshly prepared 80% ethanol slowly added and incubated for 30sec and the supernatant removed. This step is repeated for a total of 2 ethanol washes, making sure the beads are not incubated in ethanol for an extended period of time. The samples are then dried for 3mins with residual ethanol being removed after 1min. The plate is then removed from the magnetic stand and the bead pellet resuspended in 7.5µl of 1S TE pH 8.0, pipetting

up and down, ensuring the beads are completely resuspended. This is then incubated for 2mins and then magnetized for 3mins until the solution is completely clear. 7µl of supernatant in transferred into a new well and this is the cDNA input for circularization.

The circularization reaction is prepared using the following:

	Final Concentration / Volume
Denatured cDNA	7µl
CircLigase Buffer (Cambio Ltd)	1X
ATP (Cambio Ltd)	200µM
$MnCl_2$ (Cambio Ltd)	2.5mM
Betaine (Thermo)	1M
CircLigase Enzyme (Cambio Ltd)	10U

The reaction is incubated at 60°C for 5hrs. after the reaction is complete, a further 30µl of nuclease free water is added to the samples and purified using a Phenol Chloroform Isoamyl-alcohol (25:24:1 saturated with 10mM Tris pH 8.0, 1mM EDTA, Sigma) and Chloroform (Sigma) clean up and ethanol precipitated in 3X volume of 100% ethanol, 1/10th volume 3M sodium acetate and 1.5µl glycogen (30µg, Promega) and incubate at -20°C overnight. Then, samples are spun at 13000rpm for 30mins at 4°C, supernatant removed and washed with 70% ethanol before being spun again at 13000rpm for 15mins at 4°C and resuspended in 9µl nuclease free water.

The next step is relinearization and is set up:

	Final Concentration / Volume
Circularized product	9µl
RNaseA (Sigma)	10µg

Samples are incubated at 37°C for 1hour and then placed on ice immediately. This is then proceeded by the PCR amplification reaction and is assembled on ice:

	Final Concentration / Volume
Relinearized product	10µl
HF Phsuion Pol Buffer (NEB)	1X
dNTPs	200µM
Fw Primer	0.5µM
Rv Primer	0.5µM
DMSO (NEB)	3%
HF Phsuion DNA Polymerase (NEB)	0.4U
Nuclease Free Water	to 20µl

Fw Primer = 5'AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA

Samples were placed in the thermocycler when at 80°C, under the following conditions:

95°C for 3mins (denaturation), 94°C for 15sec, 62°C for 30sec, 70°C for 30sec (for 8 cycles) and 70°C for 5mins (final extension). DNA loading dye is added to each of the samples and gel extracted in a 5% TBE gel. The exact same steps for the RT gel extractions are repeated for both extraction and clean up except the final resuspension volume is 12µl. Samples had their concentrations measured using the Qubit and average size measured by Tapestation. Samples were then pooled to 1nM and sequenced on a NextSeq500 sequencer as per manufacturer's instructions.

2.5.2 RNA-seq

RNA-seq preparations were done in-house by William Clarke, Beatson Institute for Cancer Research. Samples were checked on an agarose gel prior to library preparation to check for RNA integrity. Libraries were prepared used the TruSeq Stranded mRNA library preparation kit (Illumina) and polyA selected, according to manufacturer's instructions. 1µg RNA was used as input for all samples. Final libraries had their concentrations measured using the Qubit and average size measured by Tapestation. Samples were then pooled to 1nM and sequenced on a NextSeq500 sequencer as per manufacturer's instructions.

2.6 Bioinformatics Analysis

All scripts used in this section can be found at

https://github.com/BashMo1/BashirMohamed_PhD

Since next generation sequencing technically sequences the cDNA generated in library preparations, all uracils will be called as thymines in the upcoming results chapters.

2.6.1 Demultiplexing

Before analysis, samples had to be demultiplexed to separate reads according to their samples. This was done using the Bcl2fastq package using the following code:

```
$ bcl2fastq -p 20 --no-lane-splitting --runfolder-dir . --output-dir fastq_files
```

-p referring the number of threads used

--no-lane-splitting stops the separation of samples into the 4 technical replicates as the sequencing uses 4 lanes

--runfolder-dir . The . allows for the data within the directory to be demultiplexed.

--output-dir specifies the output directory

2.6.2 ALICE-tRNA-seq adapter trimming

Adapter_trimming.sh

Adapters were trimmed using the cutadapt package. The adapter removed was CAGATCGGAAGAGCACACGTCT. A representative example can be found on the github page.

2.6.3 ALICE-tRNA-seq mapping and fragment determination

processsamples.py

The mapping of the tRNAs was adapted from the tRNA Analysis of eXpression (TRAX) programme (Holmes et al., 2020). Prior to mapping, a tRNA database is generated from the original genome in a programme called tRNAscan-SE (Lowe and Chan, 2016). The newly formed mature tRNA sequences are then padded with 20 N bases on both the 5' and 3' ends of the tRNA to allow for tRNAs containing 5' leader and/or 3' trailer sequences to map. The tRNA sequences are then mapped using bowtie2 to both the database of mature tRNA sequences and to the genome with one mismatch (either mapped to hg38 or mm10 genomes, Encode; processsamples.py). This ensures that tRNAs containing introns can be discarded from the final analysis. If reads map to the genome then they are called as pre-tRNAs. If reads map to the mature tRNA database then they are called as tRNA transcripts. If reads map to both the genome and the mature tRNA database, then they are called as tRNA transcripts.

The tRNAs are then classified into four fragment classes. tRNAs are considered to be full length if they are within 5 nucleotides of both the 5' and 3' end. 3' fragments are those sequences that are within 5 nucleotides of the 3' end but not the 5' end. 5' fragments are those within 5 nucleotides of the 5' end but not the 3' end. If sequences are not within 10 nucleotides of either end, then they are categorised as other tRNAs.

Once the tRNAs have been counted, tRNAs of interest can be filtered out and others discarded. The final tRNA counts can then be used as input to carry out quality control, differential expression analysis and statistical analysis.

2.6.4 ALICE-tRNA-seq differential expression

FedvsStarved_isoacceptor_analysis.R; *IMR90_isoacceptor_analysis.R*; *MDM2_isoacceptor_analysis.R*; *bcat_myc_isoacceptor_analysis.R*

After counting and combining full length and 3' tRNA fragments (here refers to tRNAs that had truncated due to modifications on the Watson-Crick face as opposed to genuine 3' tRNA fragments), the counts are checked on a PCA and unfavourable replicates removed. Samples are filtered so every replicate has at least 10 reads, prior to being used as input into DESeq2. Apeglm was used for log fold shrinkage and tRNA isoacceptors were considered statistically significant if the false discovery rate was less than or equal to 0.05. All results after this point were plotted and presented.

2.6.5 A34I mutation analysis

A34I_mutation_analysis.R

Frequency of mutations were calculated at position 34 of tRNAs (output from processsamples.py). If a mismatch occurred at this position (supposed to be read as an A base but instead read as a G base), then its frequency is calculated. This was done for all ALICE-tRNA-seq experiments.

2.6.6 RNA-seq analysis

Transcriptome_editing.py; FedvsStarved_mRNA_analysis.R; *IMR90_mRNA_analysis.R*; *MDM2_mRNA_analysis.R*; *bcat_myc_mRNA_analysis.R*

Human and mouse protein coding sequences were taken from GENCODE (hg38 and mm10 respectively). Prior to mapping, the protein coding sequences fasta file was filtered so that all transcripts had a 3' UTR, 5' UTR, a coding sequence that began with AUG and ended with any of the three stop codons and the coding sequence was divisible by three (transcriptome_editing.py). Reads were then mapped to the filtered transcriptome using salmon. Outputted counts were then ran on a PCA and unfavourable replicates removed. Counts were then used as input into DESeq2 for differential expression using apeglm for log

fold shrinkage. Transcripts were considered to be statistically significant if the false discovery rate was less than or equal to 0.05.

The synonymous codon usage was performed on differential genes that were statistically significant and had a log fold change of ±2. The coding sequences of said transcripts were pulled from the filtered transcriptome fasta file to generate new fasta files for each conditions. Synonymous codon usage was calculated using the SeqinR package and was normalised to the mean TPM for each transcript in each condition to weight for transcript abundance. The newly generated frequencies for each condition were then combined and plotted.

For the relative synonymous codon usage of all transcripts, the same process was followed without the log fold change of ±2 cut off. For the relative synonymous codon usage of the most abundant transcript per gene, the transcript with the highest mean TPM was selected and the rest filtered out.

Amino acid frequencies were calculated by taking the generated fasta files for statistically significant transcripts that had a log fold change of ±2 and the frequency was calculated using the Biostrings programme.

Chapter 3: Development of the ALICE-tRNA-sequencing methodology and its comparison to other tRNA-sequencing methodologies

3.1 Chapter Introduction

The development of a tRNA sequencing methodology to measure the cell's tRNAome has long been a research interest in the field. Many have resorted to using microarrays to investigate the effects on the tRNA pool in differing cell conditions and disease states (Gingold et al., 2014, Grelet et al., 2017), where others have used methods like thin layer chromatography (TLC) and High performance liquid chromatography (HPLC)-coupled mass spectrometry to study tRNA aminoacylation and modifications across various species (Grosjean et al., 2004, Cathopoulis et al., 2007, Su et al., 2014).

The shortfall to using these techniques in an attempt to study the tRNA pool is that they all rely on the successful hybridisation of probes to the tRNAs (utilising oligonucleotides or otherwise). Since all of these methods were born out of studies done on other types of RNAs, such techniques do not translate to tRNAs because of the vast number of modifications they possess. In the case of hybridisation, the most problematic modifications are those that reside on the Watson-Crick face of the tRNAs (Table 3.1). Mass spectrometry methodologies on the other hand, are limited in the sense that observed modifications that are detected cannot be accurately traced back to their isoacceptor/isodecoder, as well as its nucleoside position.

The second limitation is the problem of resolution. tRNAs are generally referred to in two ways; tRNAs that decode the same amino acid but only differ in the anticodon sequence (isoacceptors), and tRNAs that share the same anticodon sequence, but have different body sequences (isodecoders). Hybridisation-based techniques will generally give you information

with regards to isoacceptors, resulting in inferences being made at the amino acid level. Studies utilising such methodologies tend to give you generic information that cannot be mapped back to the genome. This is because modifications on the Watson-Crick face of tRNAs will hinder efforts to hybridise probes, even if they are designed with specificity towards isodecoders. Other methods such as chromatography and mass spectrometry based methodologies, typically used to study tRNA modifications, will accurately display global modification changes, but will typically lack specificity with regards to which tRNAs harbour such modifications and the positions at which they exist. This is due to these methodologies requiring tRNA fractionation in their protocols.

In order to make inferences at the nucleotide/genome level, a high resolution technique, such as high throughput sequencing is required. However, the same limitation of hybridisation still exists; high throughput sequencing of RNAs relies on the production of cDNAs, which is subsequently amplified using PCR before being sequenced. Cozen *et al* (ARM-seq) and Zheng *et al* (DM-seq) both attempted to solve this problem by using the *Escherichia coli* enzyme Alpha-ketoglutarate-dependent dioxygenase (AlkB), an enzyme known for its dealklylating properties in DNA damage protection (Trewick et al., 2002). AlkB has the ability to remove methyl groups from the Watson-crick face of modified tRNA nucleotides, aiding in the production of either full length or close to full length cDNAs post reverse transcription.

In the ARM-seq methodology, Cozen *et al* used the wild-type version of the enzyme (hereafter referred to as WT AlkB) in a demethylation reaction prior to library preparation (Cozen et al., 2015). However, the DM-seq methodology utilised the WT AlkB, in addition to a mutant AlkB, which had an aspartic acid to serine mutation at position 135 (hereafter

referred to as D135S AlkB) (Zheng et al., 2015). The intention behind the aspartic acid to serine mutation was the fact that serine has a shorter amino acid side chain in comparison to aspartic acid, while retaining the characteristics of being both a polar and non-charged amino acid. Also, position 135 in the amino acid sequence of the protein corresponds to the active site of the enzyme. Therefore, it was thought that this mutation would better accommodate larger tRNA modifications like the N^2,N^2-dimethylguanosine ($m^2{}_2g$), a common tRNA modification typically found at position 51 on tRNAs (Zheng et al., 2015, Su et al., 2014).

Another point of discussion is the motivation behind the development of both ARM-seq and DM-seq. Although both of these tRNA-sequencing methodologies were the pioneers in the field, they both attempted to ask different questions. ARM-seq was originally developed to identify tRNA fragments and tRNA halves. Therefore, their methodology required the ligation of 3' and 5' adapters to demethylated tRNAs, before being reverse transcribed and PCR amplified in their library preparations. The advantage of doing this was that only sequences with both adapters bound would ultimately be sequenced. However, DM-seq was a methodology developed to try and identify positions at which modifications were occurring. They did this by using the enzyme TGRIT (Thermostable group II intron reverse transcriptases) in their reverse transcription reactions. TGRIT worked so that whenever it came across a modified base in the tRNA during cDNA production, it would place a random nucleotide at its position as opposed to stalling. Therefore, when the sequencing data is analysed, mismatches that are identified when the sequences are mapped back to tRNA sequences would highlight modified bases. Consequently, neither technique was developed with the intention of measuring or profiling the tRNA pool between two conditions.

To address these points of concern, this chapter will aim to demonstrate how we optimized AlkB demethylation, address the limitations in using ARM-seq and DM-seq to measure the tRNA pool, and how we ultimately generated our own tRNA sequencing methodology, namely, **A**dapter **Li**gation **C**ircilarization R**e**linearization tRNA sequencing (ALICE-tRNA-seq).

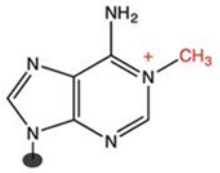
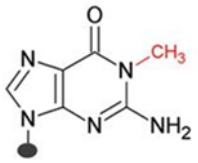
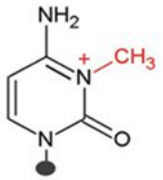
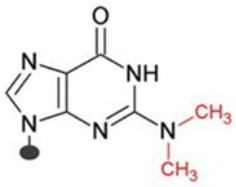
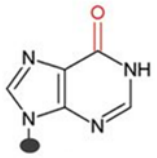
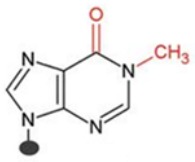
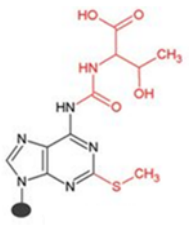
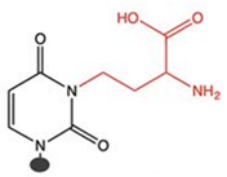
Modification	Symbol	Structure
N^1-methyladenosine*	m^1A	
N^1-methylguanosine*	m^1G	
3-methylcytosine*	m^3C	
N^2, N^2-dimethylguanosine*	m^2_2G	
Inosine	I	
N^1-methyl-inosine	m^1I	
2-methylthio- N^6-threonylcarbamoyl-adenosine	ms^2t^6A	
3-(3-amino-3-carboxypropyl) uridine	acp^3U	

Table 3.1: A table of known modifications that exist on the Watson-Crick face of tRNAs. In the symbol column, modifications with an asterisk (*) can have their modification removed via AlkB demethylation. In the structure column, molecules highlighted in red are the modifications present on the Watson-Crick face. The black dots represent nucleoside binding in the RNA sequence.

3.2 Optimizing AlkB demethylation

The first point of contention to be addressed was the AlkB oxidative demethylation reaction. This was important because different AlkB conditions were utilized in literature. AlkB demethylates modified nucleotides in a two-step, α-ketoglutarate-dependent manner. AlkB first oxidises the methyl group on modified methyl- adenosines, guanosines or cytosines, utilising the conversion of α-ketoglutarate to succinate and then removes the newly formed formaldehyde from the base, resulting in the demethylated nucleotide (see Figure 3.1A for a representative example). To address how successful AlkB demethylation was, primer extension assays were performed. Oligonucleotides were designed to be complementary to the 3' end of selected tRNAs, were subsequently labelled with radioactive ^{32}Phosphorus (^{32}P) and were used as primers in a reverse transcription reaction (Figure 3.1B). Oligonucleotides were designed so they could target particular tRNAs that had unique sequences at their 3' end. Furthermore, all oligonucleotides were purposely designed to be 18 or19 bases long. This was because the vast majority of tRNAs have a m1A modification at position 20 on the tRNA sequence. Therefore, if a tRNA does not undergo an AlkB treatment, then the reverse transcription would automatically stall, serving as a negative control.

In our pilot primer extension assays, HEK293 cells were grown to 70% confluency, and small enriched RNA (RNAs < 200nt) were extracted from the cells. Primers were designed against tRNAs Histidine GTG, Methionine CAT and Tyrosine AUG. Preliminary primer extension assays showed very little cDNA production. The AlkB reaction originally consisted of 1µg RNA, 120pmol WT AlkB, 240pmol D135S AlkB, 50mM MES pH 5.0, 2mM L-Ascorbic Acid, 300µM α-ketogluterate, 50µM Ferrous Ammonium Sulphate heptahydrate, 50µg/ml Bovine Serum Albumin, 300mM NaCl and 2mM $MgCl_2$ in a 100µl reaction (Zheng et al., 2015). It was

clear that the α-ketogluterate, Ferrous Ammonium Sulphate heptahydrate and L-Ascorbic Acid were required for the preliminary oxidation of the methylated bases since they are all co-factors in the reaction. Bovine Serum Albumin increases the general protein concentration by reducing the binding of AlkB to the plastic of the tubes and MES pH 5.0 was used as a buffer to stabilise the pH of the reaction. However, the use of high concentration NaCl and $MgCl_2$ was unclear. As a result, the 300mM NaCl and 2mM $MgCl_2$ were removed from the AlkB reaction buffer since high concentrations of salts may increase the secondary structure of the tRNAs, making it more difficult for the AlkB enzymes to access the modified bases and demethylate them. Enzymatic reactions were set up with +/- salt to test this and resulted in an increased production of full length and close to full length cDNAs in the absence of salts (Figure 3.2).

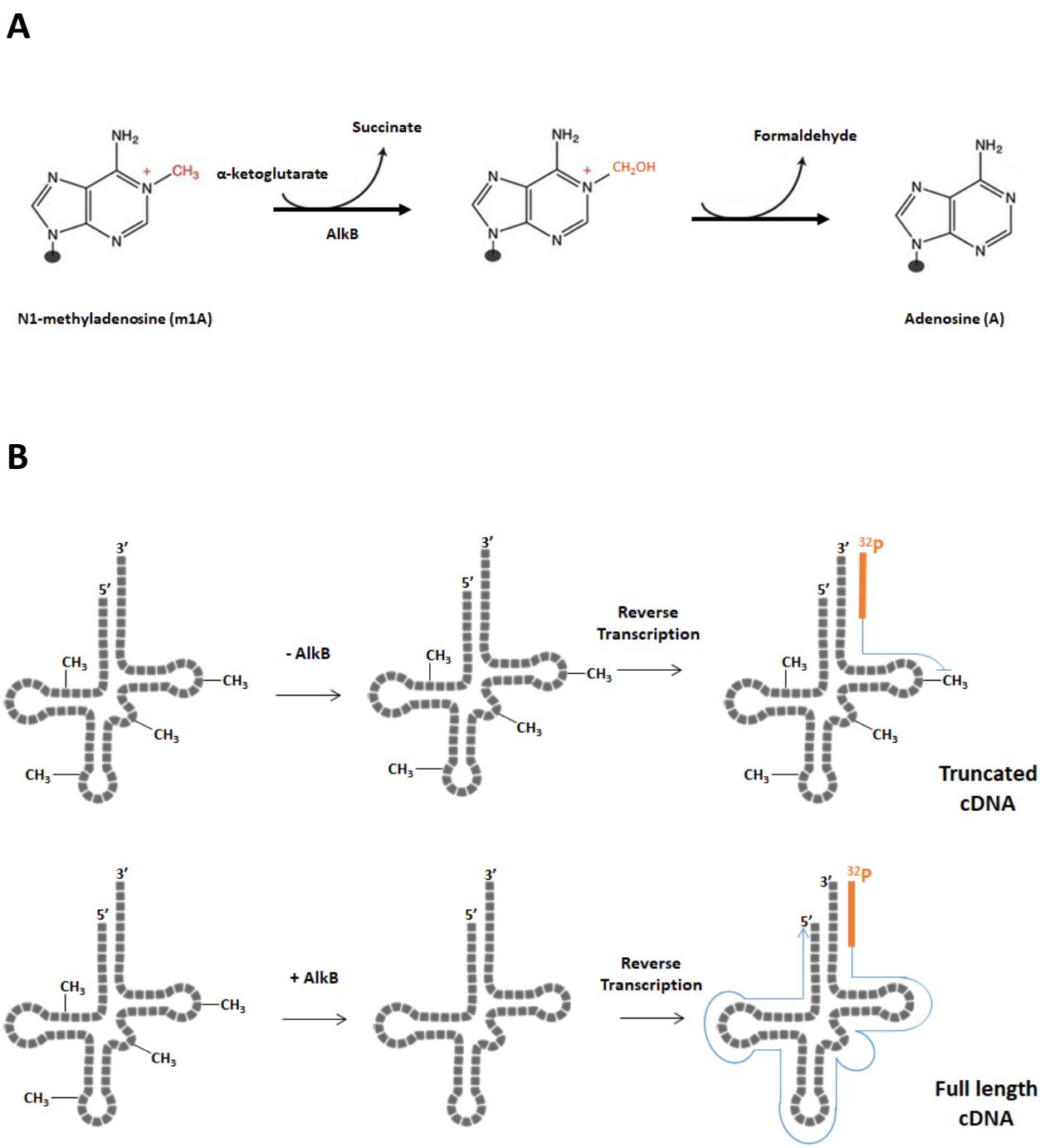


Figure 3.1: A representative schematic of AlkB oxidative demethylation and the primer extension assay. (A) The oxidative demethylation of m1A, resulting in adenosine. (B) In the absence of AlkB, the methylated groups on the Watson-Crick face of the tRNAs block reverse transcription, resulting in truncated cDNAs. Post demethylation, the reverse transcription reaction reaches the 5’ end of the tRNAs, resulting in full length cDNA production.

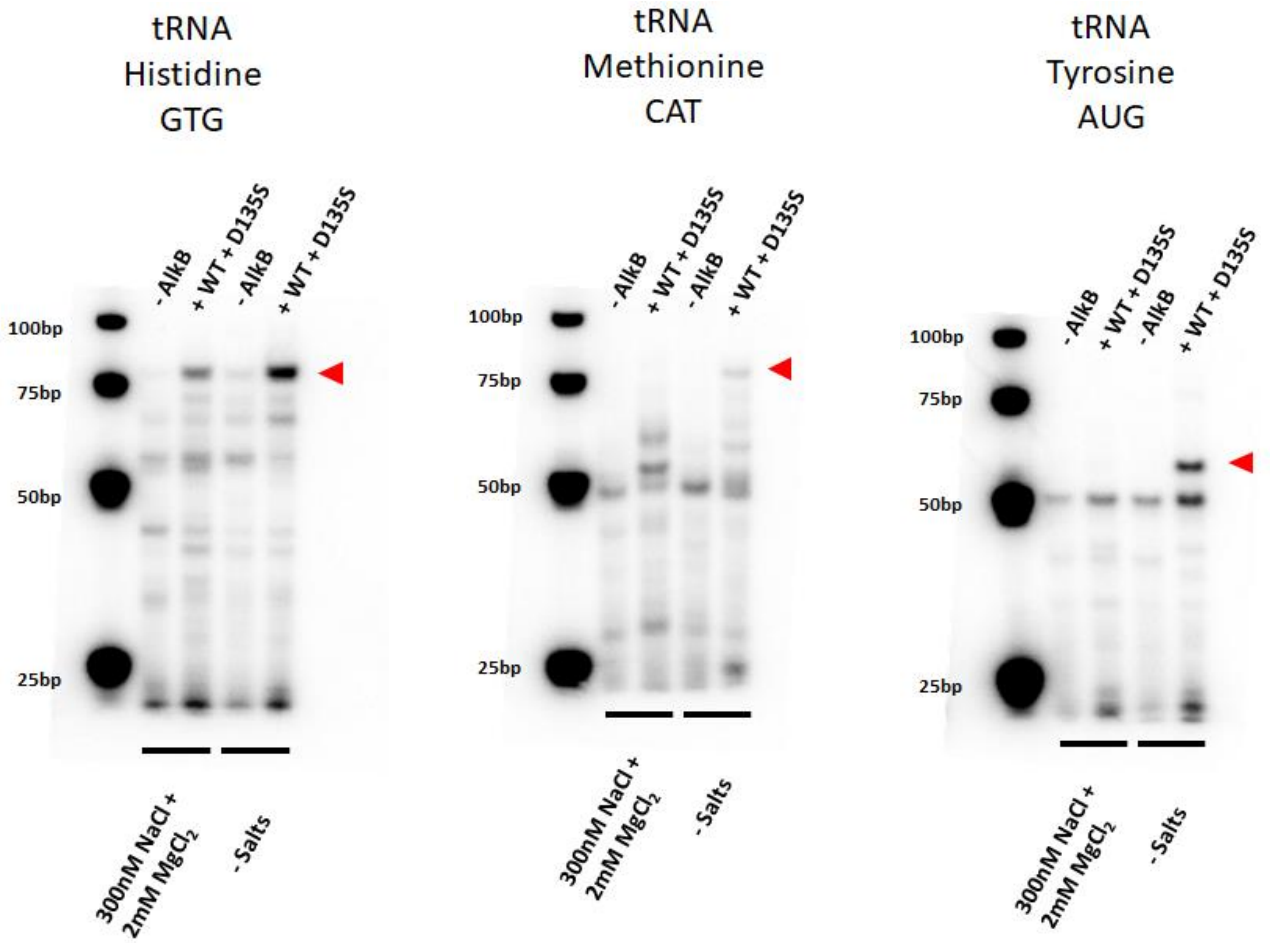


Figure 3.2: Primer Extension Assays of tRNA Histidine GTG, tRNA Methionine CAT and tRNA Tyrosine AUG show increased cDNA production in the absence of salts. The 300mM NaCl and 2mM $MgCl_2$ show to be hindering the production of full length or close to full length cDNAs post reverse transcription. The red arrows indicate the largest produced cDNAs from the assay.

3.3 HEK293 tRNA-sequencing pilot run using the ARM-seq strategy

3.3.1 Library generation and sequencing

After optimizing the AlkB demethylation reaction, a pilot tRNA-seq run was done on HEK293 tRNAs, to determine if the sequencing run was viable to measure and identify the tRNA pool. As this was a pilot, an n=1 was used. Also, a control reaction where no demethylation of the tRNAs was run alongside as a control.

Small RNAs (<200nt) were extracted from HEK293 cells grown to a confluence of 70%. The RNAs were then deacylated, to ensure the removal of bound amino acids to the 3' of the tRNAs. The tRNAs were then size selected on a gel between 50-100nt to ensure the removal of small rRNAs which have the potential to hijack the sequencing run due to their high concentration. The tRNAs were then demethylated using WT and D135S AlkB to remove any modified bases, serving as the input for the library generation (Figure 3.3 A). The tRNAs would then undergo 3' and 5' adapter ligation, before they were reverse transcribed, generating cDNA. The cDNAs were subsequently PCR amplified (12 cycles), generating the library for sequencing (Figure 3.3 B). The libraries were then ran on a 5% TBE gel and libraries between 176-246bp were excised (Figure 3.4 A-B). The average size of the library was calculated using a tapestation and the concentration of the library was determined by Qubit. The concentration and average size was then used to pool the libraries to a 1nM concentration, and was run on a NextSeq5000.

The rationale behind using the ARM-seq methodology for the pilot study was due to the fact that it involves both 3' and 5' tRNA adapter ligation. Both of these adapters would subsequently serve as PCR primer binding sites for the PCR amplification step later in the library protocol. Firstly, this meant that in the analysis of the sequencing data, a clear

distinction could be made with regards to whether a sequence was a mature tRNA, tRNA half or tRNA fragment. This was the main reason for not using DM-seq because a DM-seq library only has a 3' adapter ligation step. This means that it is impossible to distinguish between a tRNA that has been truncated due to the presence of a modification, or whether the sequence was a bona fide tRNA fragment or tRNA half. This lack of clarity would result in the inability to accurately measure the tRNA pool. However, use of the ARM-seq methodology also presented its disadvantage since binding of the 3' and 5' adapter prior to the reverse transcription step would mean that only full length tRNAs would be sequenced. This is because the 5' adapter contains a PCR binding site. Therefore, if the Superscript III enzyme truncates due to a modification that the AlkB enzymes were unable to remove (namely m^1I, ms^2t^6A and acp^3U modifications, see Table 3.1), then these sequences would not be PCR amplified and consequently not detected in the sequencing data.

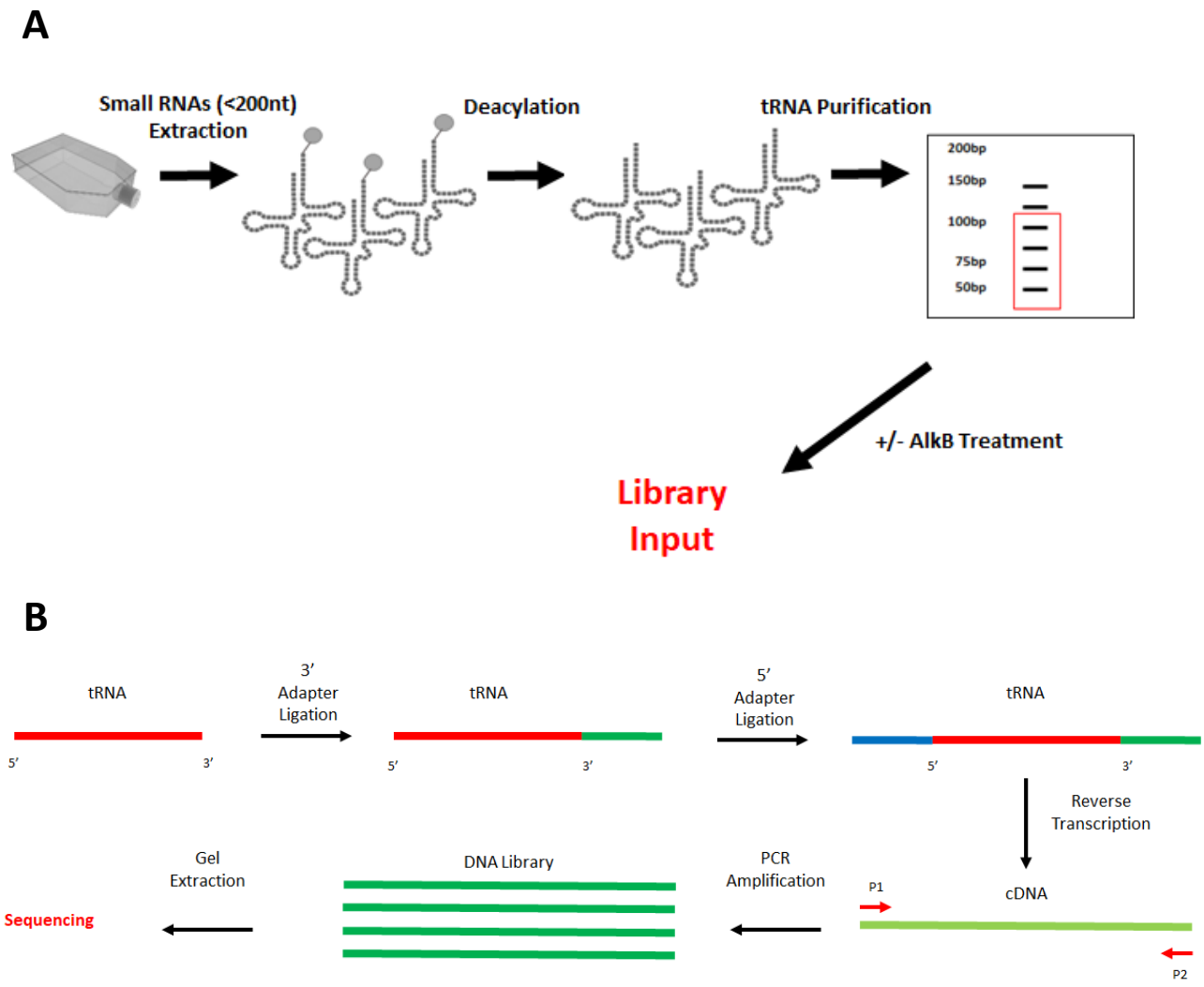
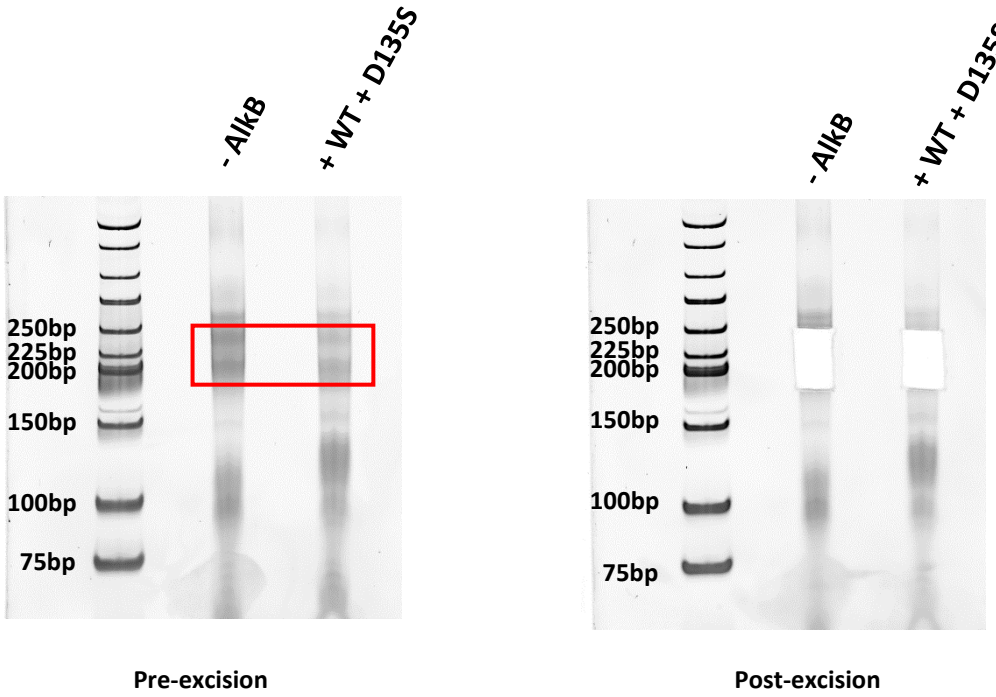


Figure 3.3 Schematic of the pilot tRNA sequencing library preparation and generation. (A) Small enriched RNAs (<200nt) were extracted from HEK293 cells. the RNAs were deacylated to remove any bound amino acids from their 3' end, before being gel extracted and demethylated using WT and D135S AlkB. The resulting tRNAs were then used as the library input. (B) Demethylated tRNAs (red) undergo 3' and 5' adapter ligation (green and blue respectively), followed by reverse transcription (light green) and PCR amplification (dark green). The DNA library is then gel extracted and then sequenced. P1 and P2 represent the PCR primer binding sites for amplification.

A



Pre-excision

Post-excision

B

Extraction Ranges

RNA Size (nt)	Final Library Size (bp)
50	176
120	246

Figure 3.4 Pilot tRNA-seq library gel extraction and extraction ranges. (A) TBE gel representing the small RNA libraries pre- and post-excision. (B) Libraries were cut between 176-246bp, corresponding to the 50-100nt input

3.3.2 Analysis workflow of the pilot tRNA-seq dataset

The pipeline development for the tRNA sequencing data analysis varies significantly to that of common RNA-seq analyses, and it is predicated on some important prerequisites. Firstly, due to the post-transcriptional modifications that occur during tRNA-biogenesis and tRNA maturation, the tRNA sequences cannot be mapped directly onto the genome. This is because tRNAs are transcribed with introns that span the anticodon loop, resulting in splicing events post-transcription (Paushkin et al., 2004). Also, 5' leader and 3' trailer sequences are cleaved during tRNA maturation (Ziehler et al., 2000), as well as 3' CCA sequences being added which is absent from the genomic sequence. Similarly, a single guanine nucleotide is added to the 5' end of histidine tRNAs (Cooley et al., 1982). Therefore, the tRNA sequences would not successfully map back to the genome and the reads would not be identified.

To overcome this, a database of mature tRNAs is formed from the original genome in a programme called tRNAscan-SE (Lowe and Chan, 2016). In brief, the programme works in conjunction with the Infernal software (Nawrocki and Eddy, 2013) and scans the genome for tRNA-like sequences. Once identified, the programme implements profile stochastic grammars called covariance models (i.e it structurally aligns stretches of sequences within the genomic tRNA to estimate the positions where the tRNA would fold to produce the classical hairpin structures). The program then continues to employ this covariance model in iteration, maximising sensitivity and accuracy in isodecoder classification. The tRNAs are then named according to where they are found in the genome. The newly formed mature tRNA sequences are then padded with 20 N bases on both the 5' and 3' ends of the tRNA to allow for tRNAs containing 5' leader and/or 3' trailer sequences to map. The tRNA sequences are then mapped to both the database of mature tRNA sequences and to the

genome. This ensures that tRNAs containing introns can be discarded from the final analysis. If reads map to the genome then they are called as pre-tRNAs. If reads map to the mature tRNA database then they are called as tRNA transcripts. If reads map to both the genome and the mature tRNA database, then they are called as tRNA transcripts.

The tRNAs are then classified into four fragment classes. tRNAs are considered to be full length if they are within 5 nucleotides of both the 5’ and 3’ end. 3’ fragments are those sequences that are within 5 nucleotides of the 3’ end but not the 5’ end. 5’ fragments are those within 5 nucleotides of the 5’ end but not the 3’ end. If sequences are not within 10 nucleotides of either end, then they are categorised as other tRNAs.

Once the tRNAs have been counted, tRNAs of interest can be filtered out and others discarded. The final tRNA counts can then be used as input to carry out quality control, differential expression analysis and statistical analysis.

3.3.3 HEK293 tRNA-sequencing pilot run quality control

To analyse the effectiveness of the ARM-seq methodology in measuring the tRNA pool, it is important to establish the types of tRNAs being detected and their distribution with regards to the amino acids they code for. In the absence of AlkB treatment, 37.33% of the total mapped reads mapped to tRNAs, 6.75% mapped to pre-tRNAs and 55.93% mapped to other RNAs (including snoRNAs, rRNAs and non-coding RNAs). In the condition where RNAs were demethylated with WT and D135S AlkB, 49.43% of the reads mapped to tRNAs, 11.25% mapped to pre-tRNAs and 39.32% mapped to other RNAs (Figure 3.5 A). Although half of the reads seem to be mapping to tRNAs after AlkB demethylation, ~ 50% of reads being lost to other RNAs and pre-tRNAs are too high. However, Cozen *et al* lose ~ 80% of their reads to other RNAs (~60% of them being rRNAs). The stark increase in the percentage of reads mapping to tRNAs is most likely because of the inclusion of a tRNA gel extraction step increasing specificity, which was not done in Cozen *et al*.

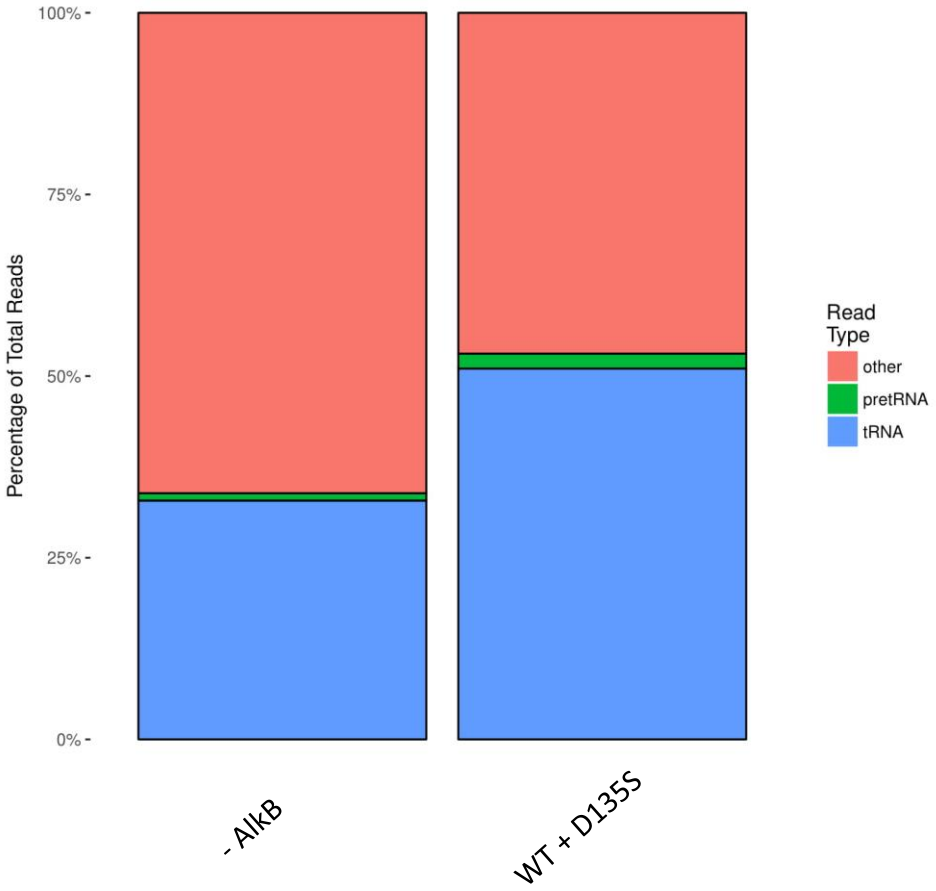
Next, we wished to compare the distribution of isoacceptors +/- demethylation. WT and D135S treatment prior to sequencing dramatically increases the isoacceptor abundance. When the RNAs were not demethylated (- AlkB), 41.06% of the total mapped reads mapped to tRNA-Glycine and 35.07% of the total reads mapped to tRNA-glutamate. In comparison, after WT and D135S treatment, the same isoacceptors took up 10.29% and 14.25% of the total mapped reads (Figure 3.5 B). Furthermore, tRNA-glycine and tRNA-glutamate are the least modified tRNAs across the 20 isoacceptors, which is why the reads bias so heavily towards these two isoacceptors and further highlights the importance of effective demethylation prior to library preparation.

The tRNA reads were then examined and categorised into Full length tRNAs, 3’ tRNAs, 5’tRNAs and other tRNAs (i.e. did not fit into any of the previous categories). In the absence of AlkB treatment, only 2.05% of the tRNA reads were mapped as full length tRNAs. This increased in the demethylated sample which had 5.11% of its tRNA reads being full length tRNAs. The vast majority of the reads were categorised as either 3’ tRNA fragments (36.24% and 60.98% respectively) or 5’ tRNA fragments (56.58% and 29.23% respectively). This is further highlighted in the coverage plot (Figure 3.6 B), where the majority of reads were expected to be highest at the tail end of the tRNAs.

This bias towards 3’ and 5’ tRNAs therefore indicated steps that needed to be considered in future tRNA sequencing experiments. Firstly, because the tRNAs were size selected prior to library preparation, many 3’ and 5’ fragments were identified even though these fragments will typically be approximately 35 nucleotides in size. This suggests that there may be degradation of the RNAs occurring either during the RNA extraction step or during the library preparation. A quality control step that could be added is the running of total RNA on an agarose gel and checking for 28S and 18S integrity before any extractions or library preparations take place. Secondly, the low percentage of full length tRNA reads could be due to the inefficiency of 5’ adapter ligation to the tRNAs. Due to the 5’ end of the tRNA being involved in a hairpin, it is possible that adapter ligation at the 5’ end could be inefficient due to this increased structure and the inability for the ligation enzyme to sufficiently ligate the 5’ adapter. Therefore, it is possible that the 3’ adapter ligated to the full length tRNAs, but because the 5’ adapter didn’t ligate, they would not be detected in the final sequencing data.

Similar problems were reported in Liu *et al* and Lama *et al*, who were investigating the inefficiency of 5’ adapter ligation in the library preparation of pre-microRNAs. Like tRNAs, the 5’ end of pre-microRNAs are involved in a hairpin structure and so found inefficiency when performing 5’ adapter ligations (Liu et al., 2014, Lama et al., 2019).

A


100%
75%
50%
25%
0%
Percentage of Total Reads
Read Type
other
pretRNA
tRNA
- AlkB
WT + D135S

B

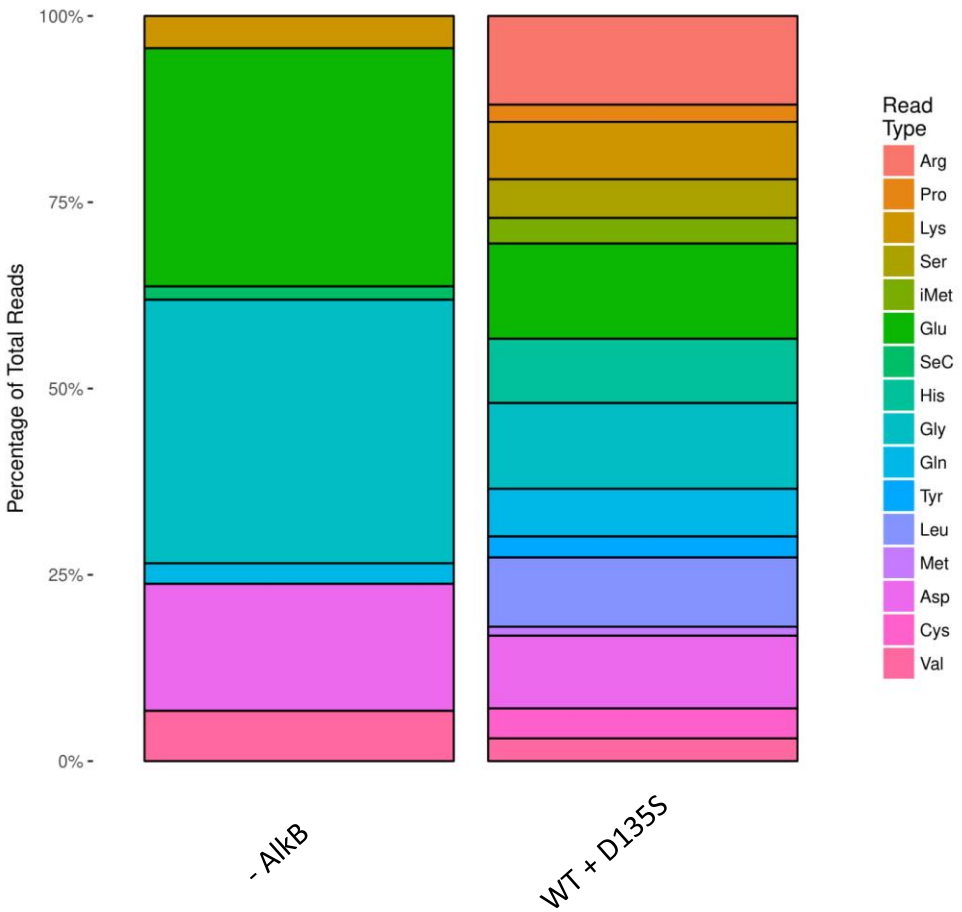

100%
75%
50%
25%
0%
Percentage of Total Reads
Read Type
Arg
Pro
Lys
Ser
iMet
Glu
SeC
His
Gly
Gln
Tyr
Leu
Met
Asp
Cys
Val
- AlkB
WT + D135S

Figure 3.5 AlkB treatment results in higher tRNA reads and a better distribution of reads across the tRNA pool. (A) Total mapped reads categorised into tRNAs (37.33% and 49.43%), pre-tRNAs (6.75% and 11.25%) and other types of RNA (55.93 and 39.32%) for – AlkB and WT + D135S AlkB samples respectively. (B) Isoacceptor distribution across total mapped reads for – AlkB and WT + D135S AlkB samples

A

tRNA Reads		
	- AlkB	**WT + D135S**
Full Length tRNAs	2.05%	5.11%
3' tRNA Fragments	36.24%	60.98%
5' tRNA Fragments	56.58%	29.23%
Other	5.14%	4.68%
Total Reads	6,257,870	13,708,113

B

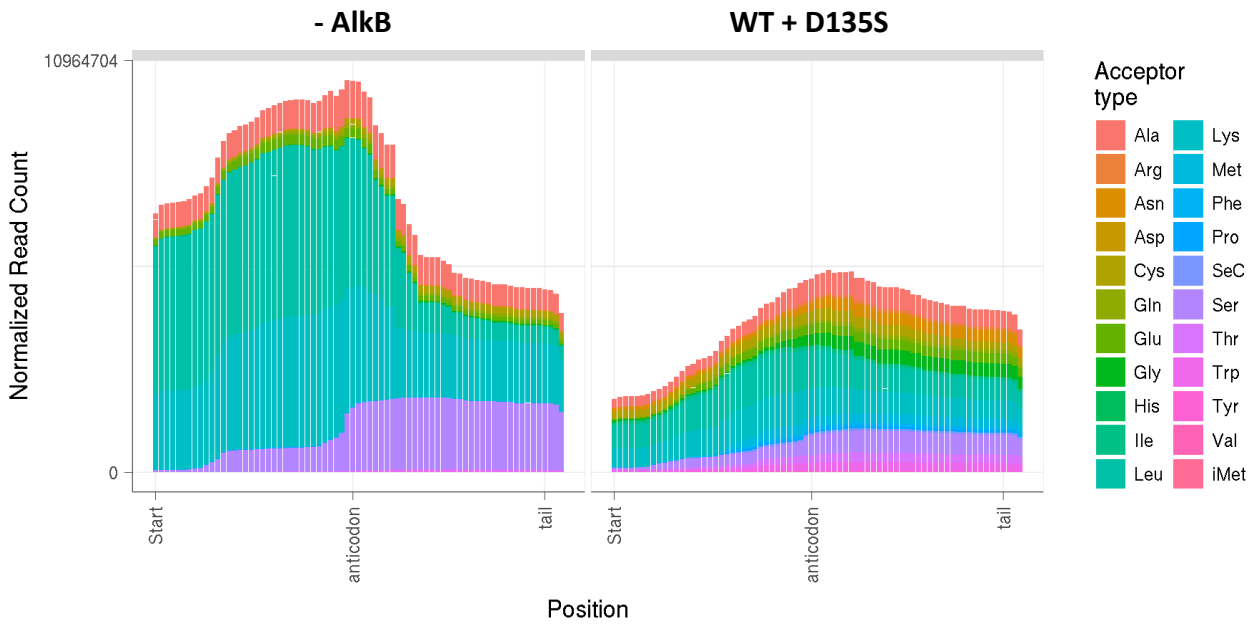


Figure 3.6 ARM-seq tRNA library methodology shows bias towards tRNA halves and tRNA fragments. (A) categorisation of the mapped tRNAs into full length, 3', 5' and other tRNAs. (B) Read coverage of the tRNA fragments, showing the increase of 3' tRNAs after demethylation.

3.4 Optimization of ALICE-tRNA-seq to circumvent the issues found in the pilot study

3.4.1 Experimental design

The limitations in 5' adapter ligation and the heavy bias towards the sequencing of 3' and 5' tRNAs in the previous pilot study resulted in the development of ALICE-tRNA-seq (**A**dapter **Li**gation **C**ircularization R**e**linearization tRNA sequencing). Prior to library preparation, total and small enriched (<200nt) RNAs were extracted from HEK293 cells. The total RNAs were run on a 1% agarose gel in order to determine 28S and 18S integrity as a quality control step to avoid degraded RNAs being sequenced. The total RNAs were also poly(A) selected and prepared for sequencing. The rationale here is that the mRNAs from the same source material is sequenced alongside tRNAs for codon optimality studies and comparison between the two sequencing methodologies. The small enriched RNAs go through the same steps as before, namely deacylation, size selection and demethylation before being used as input in the ALICE-tRNA-seq library generation (Figure 3.7 A).

The ALICE-tRNA-seq library generation begins with 3' adapter ligation, depletion of excess 3' adapter, reverse transcription, circularization, relinearization and PCR amplification (Figure 3.7 B). The rationale for each step and their optimization will be shown.

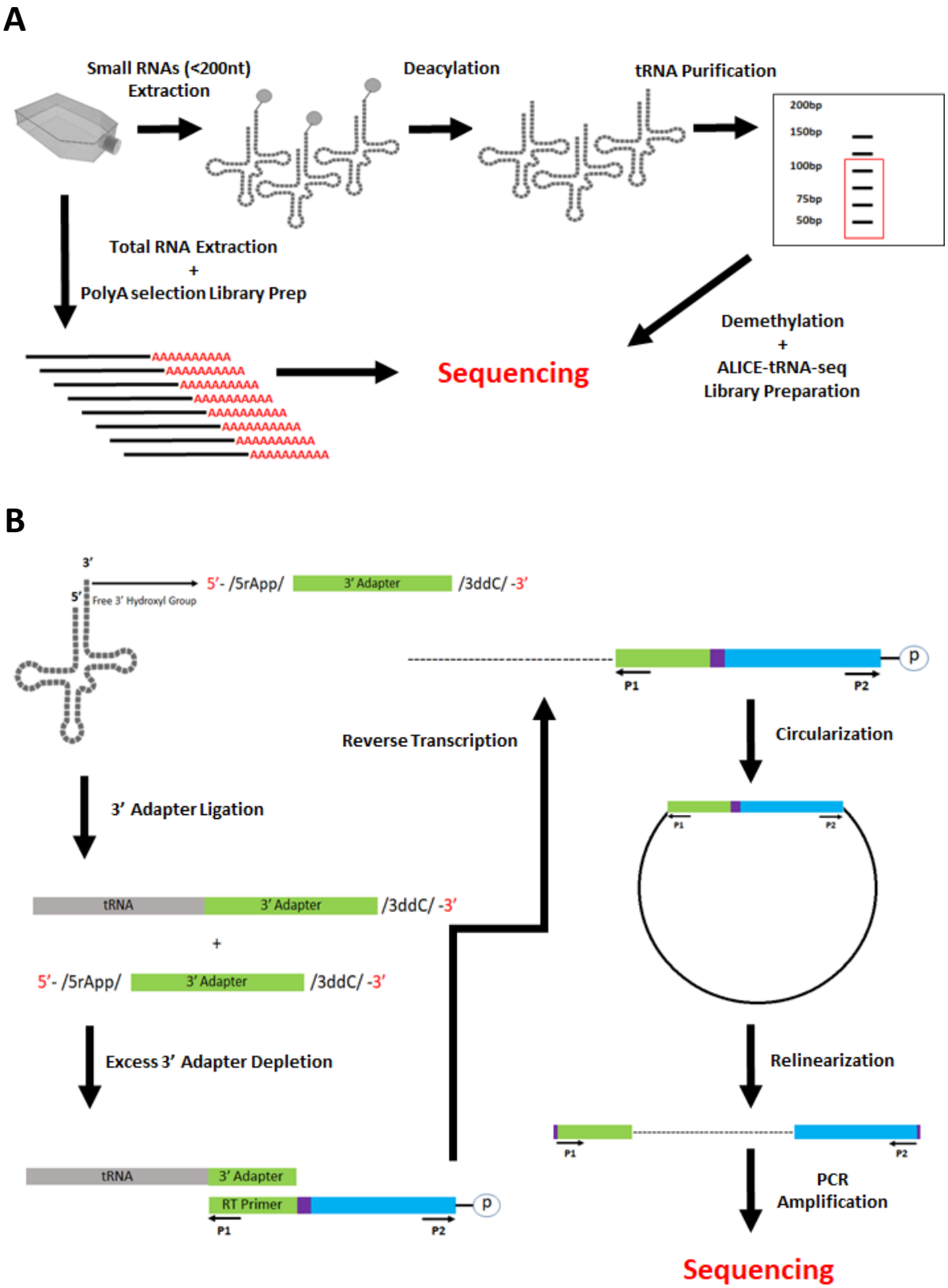


Figure 3.7 Experimental design of ALICE-tRNA-seq. (A) Total and small enriched (<200nt) will be extracted from HEK293 cells for Total RNA and tRNA-seq respectively. The small enriched RNAs will be Deacylated, size selected and demethylated before being used as input into the ALICE-tRNA-seq library preparation. The Total RNAs will be poly(A) selected and sequenced for codon optimality studies. (B) The ALICE-tRNA-seq library preparation will consist of 3' adapter ligation, excess 3' adapter depletion, reverse transcription, circularization, Relinearization and PCR amplification before being sequenced.

3.4.2 RNA extraction quality control and 3’ adapter ligation optimization

Before beginning any library preparation, total RNA is run on a 1% agarose gel to check for 28S and 18S integrity (Figure 3.8 A). Degraded RNA will typically show as a smeared band and cannot be used downstream. Another quality control step is the size selection of tRNAs and the presence of highly saturated mature tRNAs should be seen (Figure 3.8 B). Low quality tRNA will typically present as equal saturation to the rRNA band at approximately 120nt. Furthermore, small RNAs are loaded across multiple lanes to reduce size selection inefficiency since overloaded lanes will result in low quality size separation.

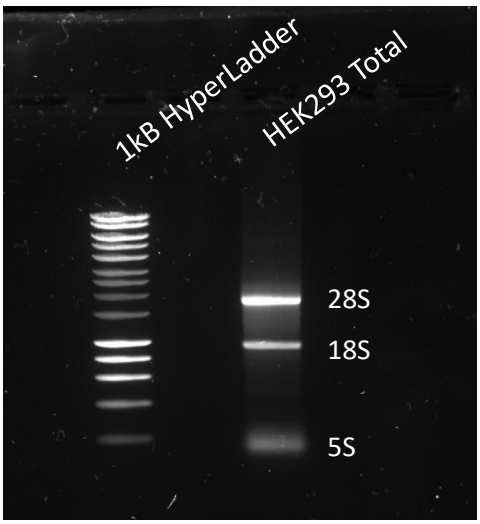
The 3’ adapter ligation step involves the binding of a DNA 3’ adapter with a 5’ pre-adenylated end and a 3’ dideoxy-C end, using T4 RNA Ligase 2 (1-249) K227Q (Figure 3.9 A). The significance of the 5’ pre-adenylated end of the adapter is that it would only ligate with an RNA that has a monophosphate with a free hydroxyl group. This is only found in mature tRNAs, since the hydroxyl group is needed to bind to amino acids (and is also introduced during CCA addition to the acceptor stem in tRNA maturation (Ziehler et al., 2000)). Therefore, specificity is introduced, ruling out the binding of 3’ and 5’ tRNA fragments. The significance of the 3’ dideoxy-C modification is that it will block any potential binding to another adapter since it lacks the ability to bind to another pre-adenylated end, ruling out the possibility of multiple adapters binding to each other.

The percentage PEG8000 was then optimized. PEG8000 acts as a molecular crowding agent, increasing the rate of the enzymatic reaction. 20% PEG8000 was found to be most optimal (Figure 3.9 B). However, it was also observed that using higher concentrations than 20% was detrimental to the production of 3’ adapter ligated tRNAs (Figure 3.9 C).

The percentage DMSO was also optimized. The DMSO acts as a molecular chaperone in the ligation reaction, also increasing the enzymatic rate of the reaction. 20% DMSO was found to be most optimal (Figure 3.9 C).

The time and temperature for which the 3’ adapter ligation reaction took place was also optimized, with 25°C for 16hrs producing the greatest yield (Figure 3.10).

A



B

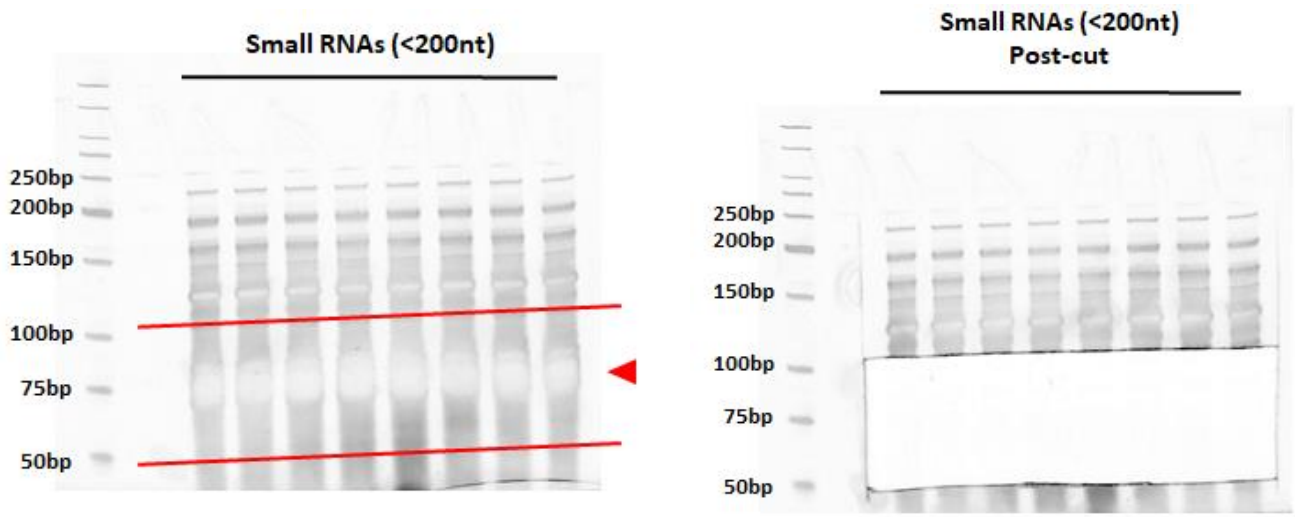
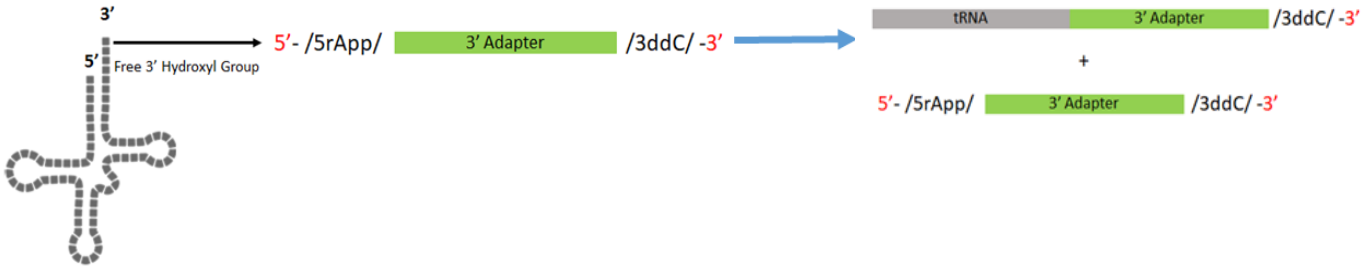
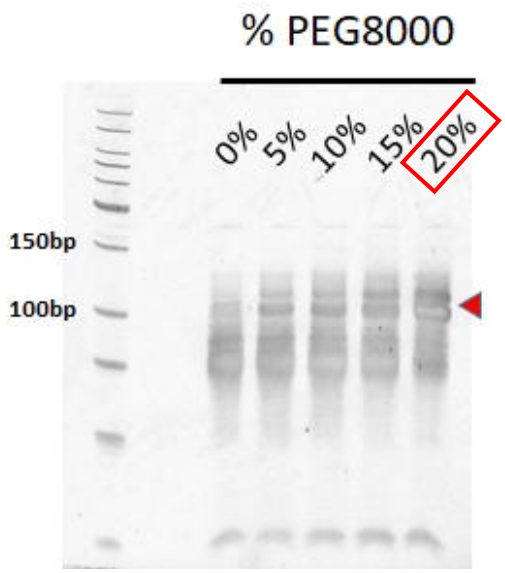


Figure 3.8 RNA extraction quality control. (A) Total RNA run on a 1% agarose gel to check for degradation. This is done by checking the 28S and 18S bands and their ratio. Degraded RNA would generally show smearing and the presence of multiple bands. (B) Size selection of the tRNAs between 50-100nt, with a complementary gel post excision. The red arrow highlights the mature tRNAs. Red lines indicate excised region. Red arrow indicates tRNA population.

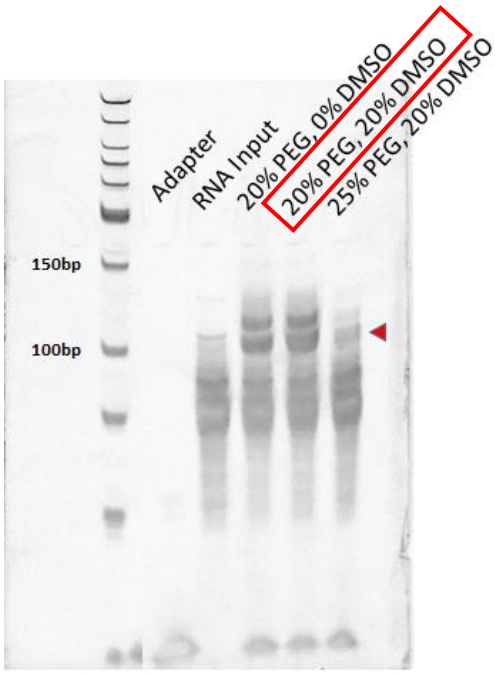
A



B



C



D

Adapter Ligation Ranges

RNA Size (nt)	Post 3' Adapter Ligation (+ 22nt)
50	72
75	97
100	122

Figure 3.9 Optimization of the 3' adapter ligation step. (A) Schematic of the 3' adapter ligation step. (B) Optimization of the percentage PEG8000 used in the 3' adapter ligation reaction. 20% PEG800 results in the highest concentration of bound 3' adapter (C) Optimization of the percentage DMSO used in the 3' adapter ligation reaction. The combination of 20% PEG8000 and 20% DMSO proves most optimal (D) size distribution pre and post 3' adapter ligation. Red arrow indicates tRNA library.

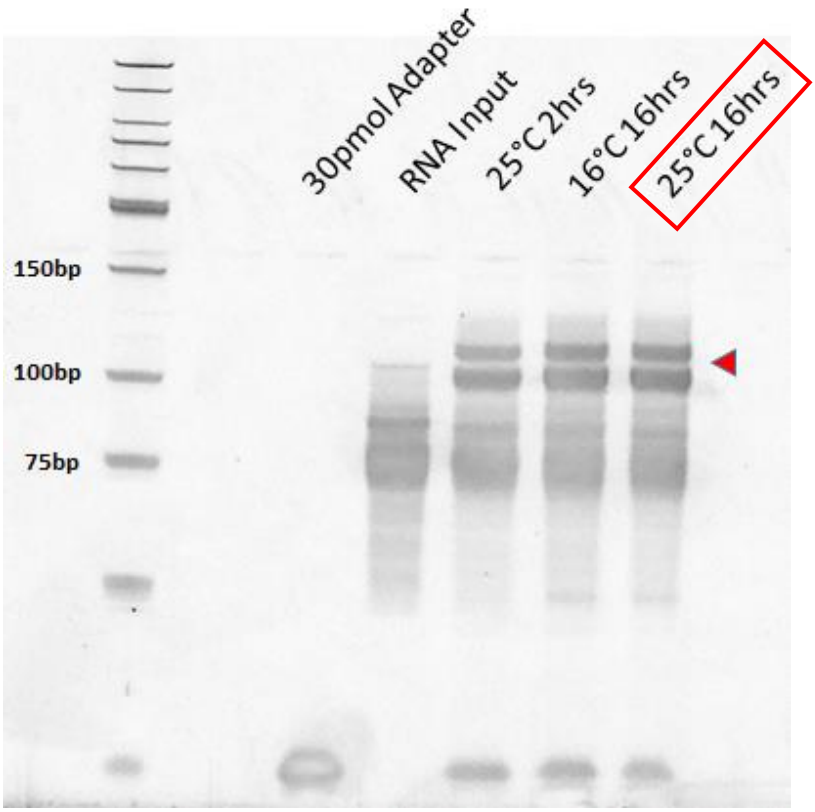


Figure 3.10 Optimization of the 3' adapter ligation reaction temperature and time. 25°C for 16hrs was found to be the most optimal time and temperature for the 3' adapter ligation reaction. Red arrow indicates tRNA library.

3.4.3 Excess 3’ adapter is depleted in a two-step reaction

It is important to remove excess 3’ adapter because it harbours the binding site for the RT primer for reverse transcription in the next step. Failure to remove any excess 3’ adapter would hinder reverse transcription and decrease the yield of cDNA produced. To address this, excess 3’ adapter is depleted in a two-step reaction (Figure 3.11 A). First, deadenylase is used to remove the pre-adenylated group from the 5’ end of the adapter, followed by its digestion by the 5’-3’ exonuclease RecJ. Since RecJ digests the DNA adapter into single nucleotides, they pass through the column when the reaction is cleaned using the RNA-clean and concentrate 5 kit (minimum length of DNA/RNA retained by the column is 17nt). Futhermore, we show that the 3’ adapter ligated tRNAs remain unaffected. This is because the RecJ enzyme only digests DNA. The reaction was carried out at 37°C for 1hr, followed by 30°C for 30mins. Two RecJ endonucleases were used from different suppliers (New England Biolabs and Lucigen) and both successfully depleted excess 3’ adapter (Figure 3.11 B). Deadenylase and RecJ (NEB) were used in all library preparations (due to NEB RecJ being cheaper).

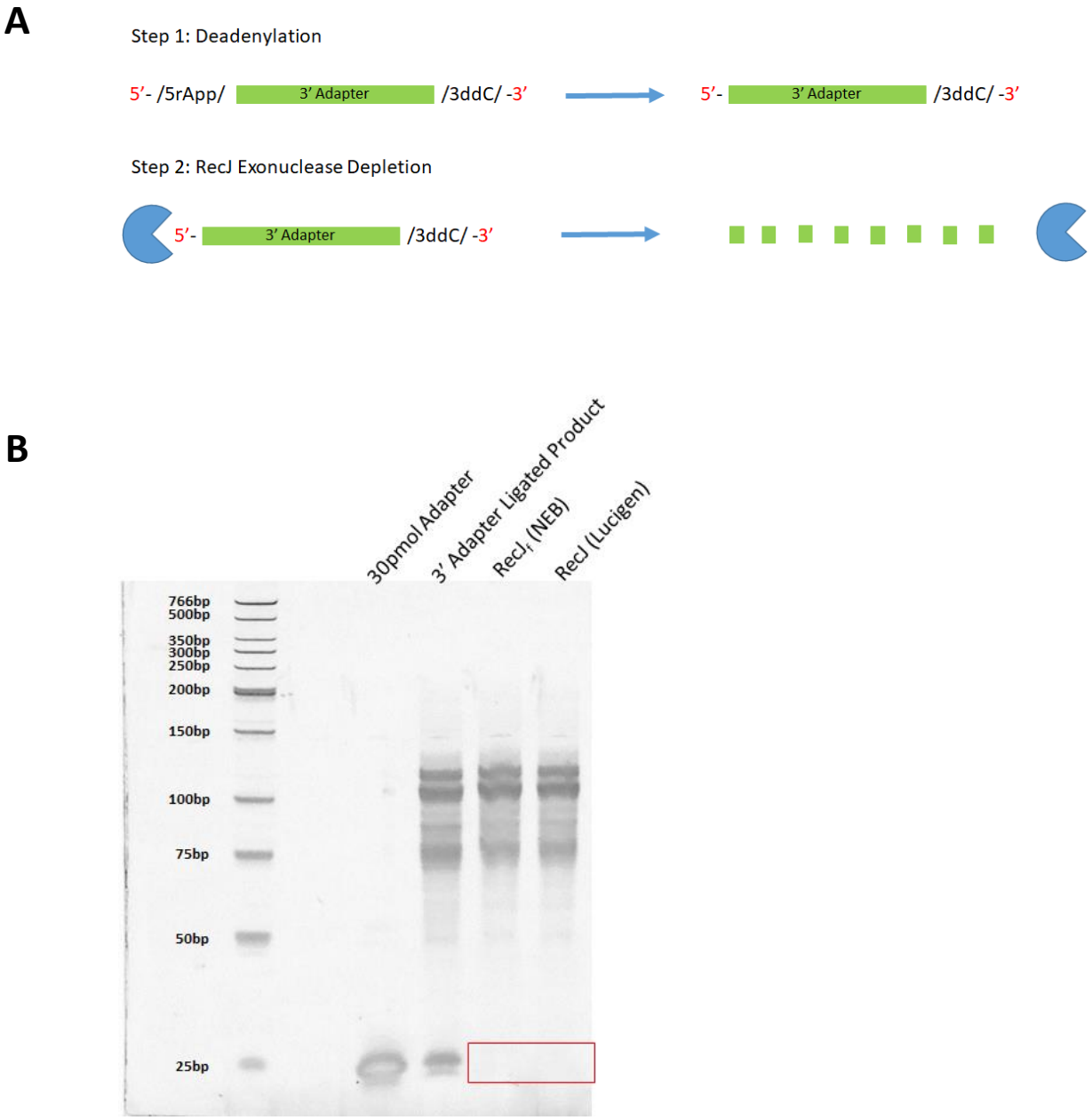


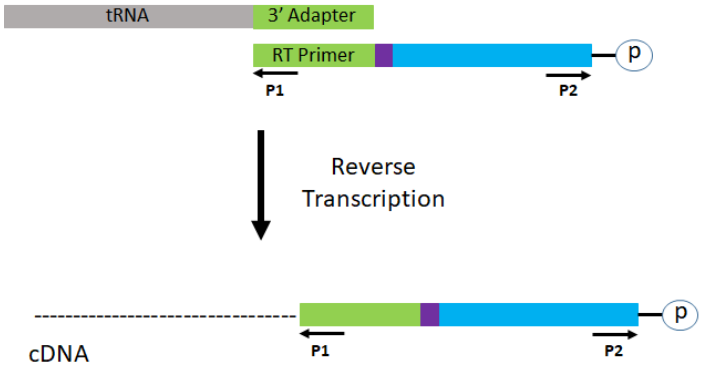
Figure 3.11 Deadenylation and RecJ treatment successfully depletes excess 3' adapter. (A) Schematic of the deadenylation and RecJ reaction to deplete excess 3' adapter. This two step reaction works by removing the pre-adenylated group from the 5' end of the adapter by deadenylase, followed by its digestion by the 5'-3' exonuclease RecJ. (B) 8% TBE gel highlighting the use of two different RecJ enzymes from two suppliers. Both resulted in the successful depletion of excess 3' adapter.

3.4.4 Reverse transcription and the need for a gel extraction step post-RT

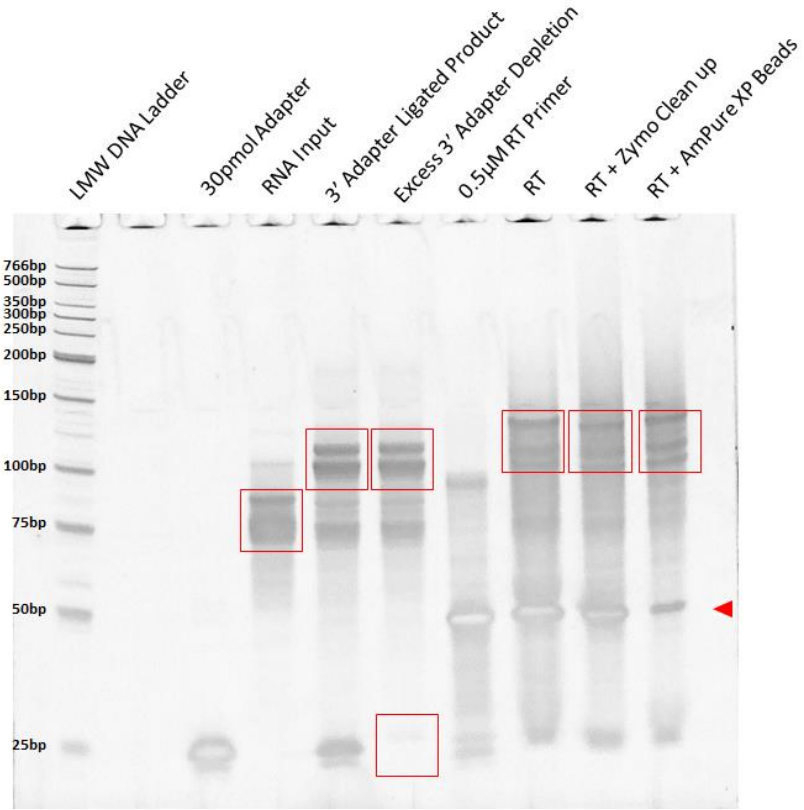
The newly generated 3’ adapter ligated tRNAs now have a RT primer binding site, allowing for cDNA generation without the need for doing a 5’ adapter ligation step (Figure 3.12 A). This is circumvented by the upcoming circularization and relinearization steps. The 3’ adapter ligated tRNAs and RT primer were first heated to 95°C for 5mins and cooled on ice, prior to the addition of the SuperScriptIII (SSIII) reaction mixture, allowing for the complementary binding of the 3’ adapter ligated tRNAs and the RT primer. The RT was then performed at 55°C for 1hr, followed by an RNaseH treatment (37°C for 20mins, as per manufacturer’s instructions), to digest any excess RNAs bound within a RNA/DNA hybrid.

Since the RT primer has a 3’ phosphate group (which acts as a substrate in the circularization reaction), removal of any excess RT primer is significant because the primers also circularize. This would then result in adapter contamination post-PCR, taking up reads whne the samples are sequenced. To address this, an attempt to remove excess RT primer was made using AmPure XP magnetic beads. AmPure XP work by binding to the negatively charged DNA phosphate backbone and by altering the concentration of isopropanol used in the reaction buffer, the size of the DNA purified can be adjusted for. Using 1.16X (of the total volume) isopropanol and using the Zymo Clean and Concentrate-5 columns as a positive control (column can retain sequences >17nt), the AmPure XP magnetic bead clean-up was capable of removing much of the excess RT primer, but not all of it (Figure 3.12 B). Therefore, the decision was made to gel extract the cDNAs (excising sequences between 97-147nt, Figure 3.12 C) and then performing an AmPure XP magnetic bead clean up using 116.7% isopropanol.

A



B



C

Reverse Transcription Ranges

3' adapter ligated tRNA	Post RT (+ 47nt)
72	97
97	122
122	147

Figure 3.12 Gel extraction is required to remove excess RT primer. (A) Schematic of the reverse transcription step. The RT primer is complementary to the 3' adapter ligated tRNA and primes for cDNA production. The RT primer has a 3' free phosphate utilized in the circularization step. The RT primer has divergent PCR primer sites (P1 and P2) (B) 8% TBE gel showing all steps prior to reverse transcription and the need for gel extraction. Both column and bead clean up attempts fail to remove excess RT primer (red arrow). (C) Size distribution expected after reverse transcription. Truncated cDNAs are due to AlkB being unable to remove all Watson-Crick face modifications

3.4.5 Optimizing circularization and relinearization

Once the newly generated cDNAs are purified, they needed to be circularized and subsequently relinearized in order to allow for the divergent PCR primer binding sites to be on opposite ends of the cDNA. The circularization is done by utilizing the 3' free phosphate group and Circ Ligase II, allowing the sequence to be circularized to its 5' end. The RT primer inherently has a RiboC base (a cytosine ribonucleic acid) embedded within its sequence (Figure 3.13 A, purple base), which acts as the digestion site for RNaseA in the relinearization reaction. This results in the aforementioned divergent PCR primer binding sites (Figure 3.13 A, P1 and P2), to be on the 5' and 3' ends of the cDNA and can be successfully PCR amplified. These two steps therefore circumvent the need for a 5' adapter ligation step and similar strategies have been used in the sequencing of other highly structured RNAs in literature (Liu et al., 2014, Burke et al., 2014).

Because small circularized products are difficult to visualise on gels (in this case, sequences range from 97-147nt after reverse transcription) and to avoid the use of radioactive labelled libraries, a simple test based on PCR was used to confirm successful circularization (Figure 3.13 A). If the generated cDNAs are circularized and then relinearized, this would result in the PCR primer binding sites being on opposite ends of the cDNA and hence successful PCR amplification. However, if the circularization step is skipped (positive control), RNaseA treatment on the cDNAs would cleave the RT primer end of the cDNA in two and can therefore not be PCR amplified.

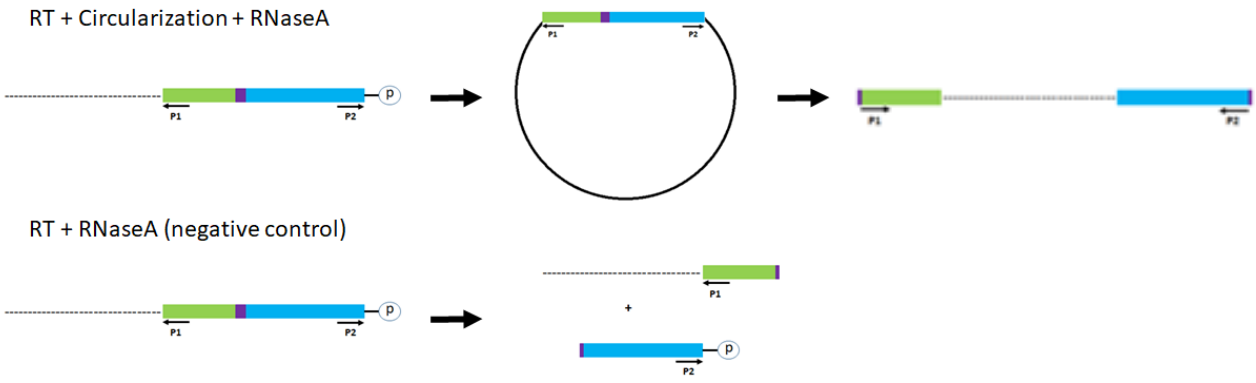
This test was carried out on purified cDNA products and as expected, none of the controls were amplified after PCR, suggesting circularization was successful (Figure 3.13 B). The optimal conditions for circularization were 200µM ATP, 2.5mM $MnCl_2$, 1M betaine and 10U

Circ Ligase II in a 20µl reaction, incubated at 60°C for 5hrs. It was also found that heating the samples (95°C for 5mins) prior to adding the circularization reaction buffer was an effective means to decrease any secondary structure that could hinder circularization. The circularized DNA was then purified using phenol chloroform isoamyl alcohol (25:24:1) and a subsequent ethanol precipitation over night at -20°C. The relinearization reaction was carried out in a 10µl reaction using 10µg RNaseA.

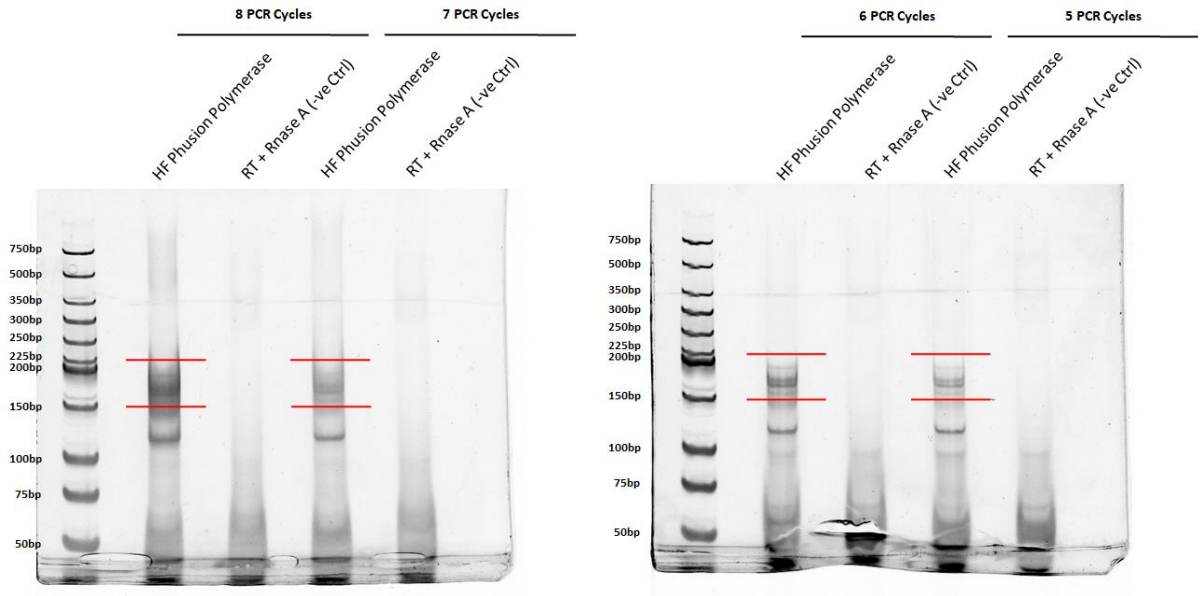
The number of PCR cycles was also optimized (Figure 3.13 B). The significance of having as few PCR cycles as possible is that the PCR reaction can bias against tRNA sequences with a higher concentration of guanine and cytosine bases, resulting in duplication bias. However, this must be balanced with regards to the concentration of library generated post PCR-amplification, since all libraries are size selected (169-219bp are excised, Figure 3.13 C) by gel extraction. Although the number of PCR cycles can be as few as 6, 8 PCR cycles was found to be optimum with regards to the concentration of purified library after gel extraction.

High fusion DNA polymerase was found to be the most effective polymerase for PCR amplification. The reactions were carried out in 20µl (as per manufacturer's instructions). The PCR primers were designed to be applicable with the NextSeq5000 sequencer, with the reverse primers having a unique 6-base barcode index sequence. This results in the ability of demultiplexing pooled libraries from multiple experimental conditions. The library from the 8 PCR-cycles (Figure 3.13 B) was used in an ALICE-tRNA-seq pilot run. The library was gel extracted and purified using AmPure XP magnetic beads. Like section 3.31, the libraries' concentration and size was calculated using a Qubit and Tapestation respectively and the library was diluted to 1nM and run on a NextSeq5000 sequencer.

A



B



C

PCR Amplification Ranges

RNA Size (nt)	Post PCR (+119bp)
50	169
75	194
100	219

Figure 3.13 ALICE-tRNA-seq can go as low as 6 PCR cycles. (A) Schematic depicting how to measure successful circularization. If samples are circularized and then relinearized (utilizing the RiboC base embedded in the sequence, purple block), the PCR primer binding sites (P1 and P2) would be on opposite ends of the cDNA and can therefore be PCR amplified. If the circularization step is skipped, then the cDNA is cleaved and cannot be PCR amplified. (B) 6% TBE gel confirming successful circularization and showing that PCR amplification be reduced to as little as 6 cycles. (C) PCR amplicon range for gel extraction

3.4.6 Pilot ALICE-tRNA-sequencing run on HEK293 tRNAs

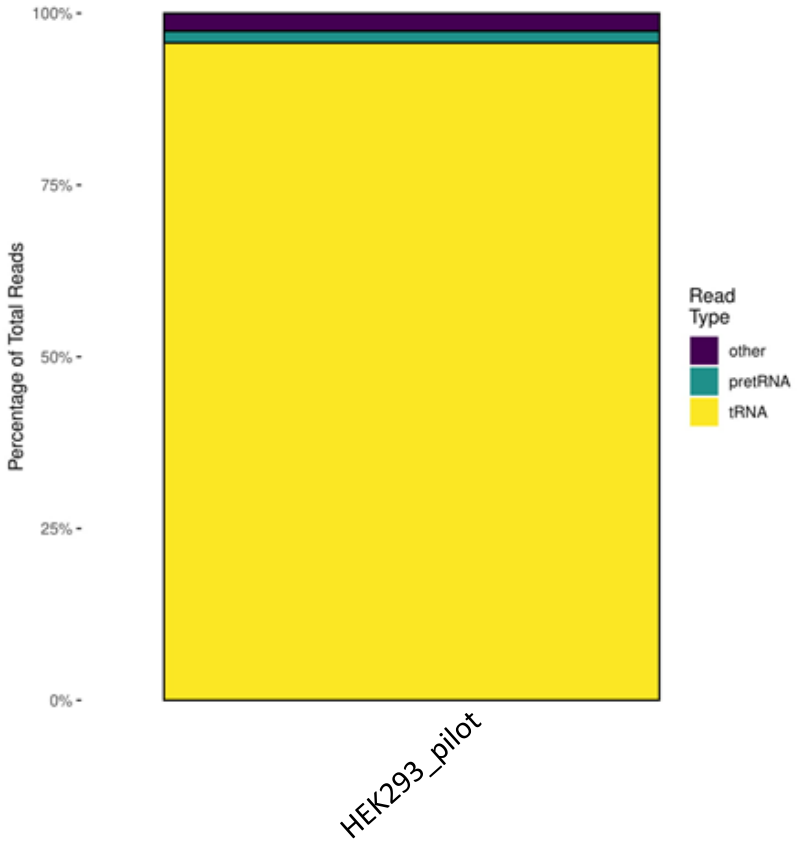
The pilot run of the 8 PCR-cycles library (n=1) vastly improved on the results reported in the ARM-seq methodology pilot run. 93.46% of total mapped reads mapped back to tRNAs (from 25,911,798 reads), with only 1.69% of reads mapping to pre-tRNAs and 2.51% mapping to mitochondria tRNAs (Figure 3.14 A-B). Furthermore, examination of the isoacceptor distribution across the 22 amino acids improved (Figure 3.14 C), with transcripts such as tRNA-Phenylalanine (2.39%), tRNA-threonine (8.45%) and tRNA-isoleucine (2.23%) being detected at higher levels, which were previously very low in abundance (0.58%, 0.56% and 0.2% respectively, Figure 3.5 B) in the ARM-seq pilot run. Also, the bias for tRNA-glycine (41.06%) tRNA-glutamate (35.07%) from the ARM-seq pilot run was reversed, with the isoacceptor proportions in ALICE-tRNA-seq accounting for 5.34% and 1.71% of total reads, respectively (See Table 3.2). This proportionate distribution across all the isoacceptors highlights the significant improvements made using the ALICE-tRNA-seq methodology.

Because the ALICE-tRNA-seq library preparation requires a free hydroxyl group on the 3' end of tRNAs for successful 3' adapter ligation, the methodology naturally corrects the 5' tRNA fragment bias observed in the ARM-seq methodology pilot run. This is further confirmed when only 0.11% of reads (from a total of 25,143,543 reads) mapped back to 5' tRNA fragments after fragment determination in the ALICE-tRNA-seq library preparation (Figure 3.15 A-B) in comparison to the 29.23% detected in the ARM-seq methodology. Furthermore, the read coverage plot (Figure 3.15 A) confirms that all the sequences start from the 3' end and so, truncated tRNAs can be taken into consideration since the removal of a 5' adapter ligation step allows for the identification of tRNAs that may have had cDNA production hampered by modifications that the AlkB enzyme could not remove during demethylation (modifications such as m^1I, ms^2t^6A and acp^3U modifications, see Table 3.1). Furthermore,

the coverage plot shows that more that 60% of the reads pass the anticodon loop, allowing for confidence in their identification. However, most variance in isodecoder sequences occur between the acceptor stem and anticodon loop of the tRNA. The percentage of "other" tRNAs (tRNAs that fail to be within 5 nucleotides of the 3' and 5' end) also decreased from 4.68% (ARM-seq) to 1.84% (ALICE-tRNA-seq).

Improvement in mapping to tRNAs and the more even distribution across all the tRNA isodecoders led to the conclusion of a more robust and technically sound methodology to measure the relative tRNA pool in cells.

A



B

RNA Type	% Reads
tRNA	93.46%
pretRNA	1.69%
havana	0.00%
snRNA	0.63%
mirbase	0.01%
rRNA_pseudogene	0.15%
ribozyme	0.00%
sRNA	0.00%
rRNA	0.90%
misc_RNA	0.26%
snoRNA	0.38%
scaRNA	0.00%
other	2.51%

C

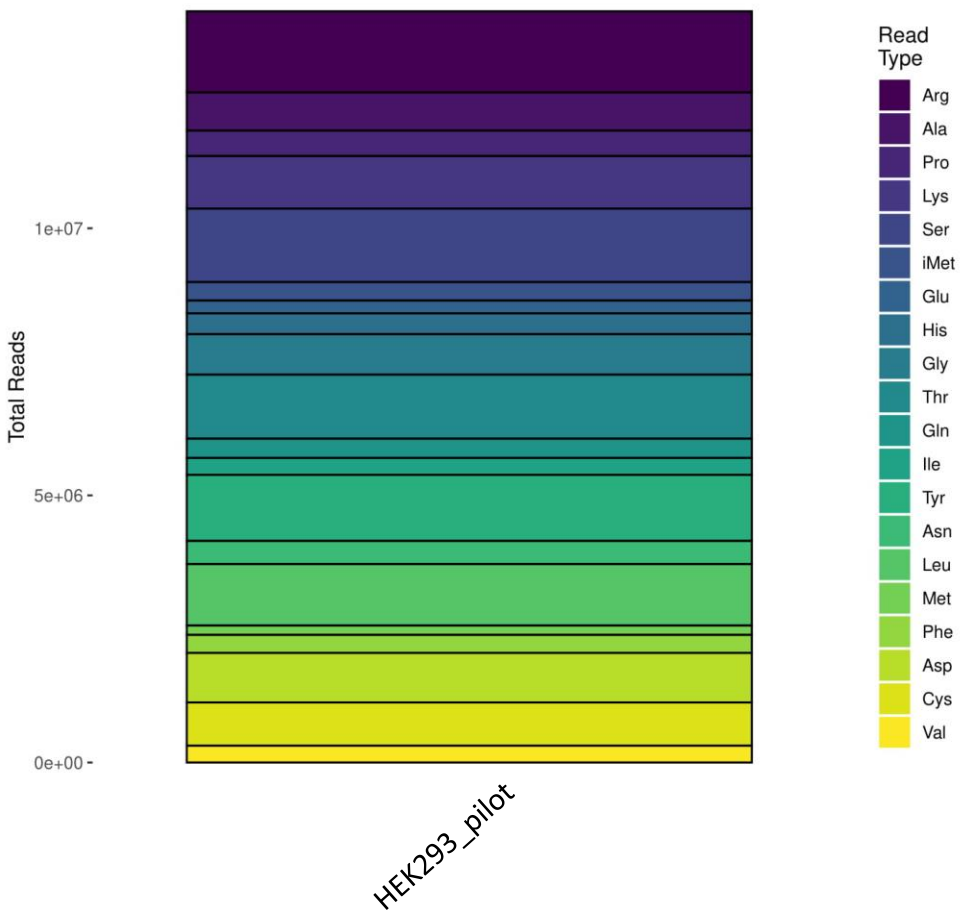
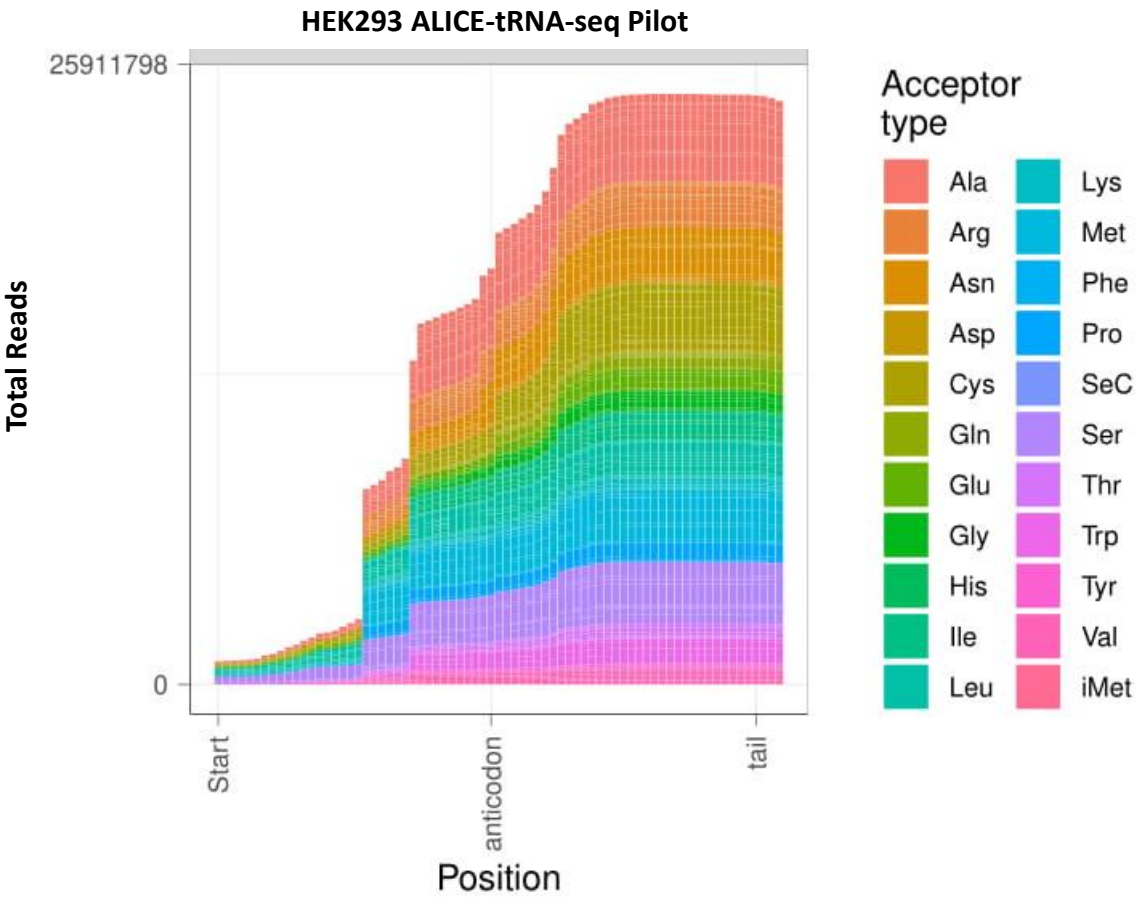


Figure 3.14 ALICE tRNA-seq gives preference to tRNAs over other types of small RNAs. (A-B) Mapping of the sequencing reads to their RNA type. (C) Isodecoder distribution across the 22 amino acids (percentages shown in Table 3.2)

Isodecoder	% of Total Read (ALICE-tRNA-seq)	% of Total Read (ARM-seq, WT+D135S)
Ala	5.03%	0.43%
Arg	10.70%	12.52%
Asn	3.07%	0.34%
Asp	6.54%	6.36%
Cys	5.69%	3.13%
Gln	2.55%	4.98%
Glu	1.71%	14.25%
Gly	5.34%	10.29%
His	2.74%	13.11%
Ile	2.23%	0.20%
iMet	2.43%	3.23%
Leu	8.10%	9.86%
Lys	6.95%	5.60%
Met	1.24%	1.01%
Phe	2.39%	0.58%
Pro	3.36%	1.48%
SeC	0.11%	1.12%
Ser	9.69%	4.27%
Thr	8.45%	0.56%
Trp	0.75%	0.82%
Tyr	8.71%	3.01%
Val	2.24%	2.84%

Table 3.2 Comparison of the isoacceptor distributions in the ALICE-tRNA-seq pilot run vs ARM-seq (WT+D135S) pilot run. A more equal distribution is seen in the pilot run prepared using the ALICE-tRNA-seq methodology in comparison to the ARM-seq methodology. Isodecoders that are biased for in the ARM-seq methodology is reversed in ALICE-tRNA-seq and Isodecoders of very low detection in ARM-seq is improved in ALICE-tRNA-seq.

A



B

	HEK293 ALICE-tRNA-seq Pilot Run
Full Length	6.02%
3' tRNA Fragments	92.03%
5' tRNA Fragments	0.11%
Other	1.84%
Total	25,143,543

Figure 3.15 ALICE-tRNA-seq rectifies bias to 5' tRNA fragments and halves found in the ARM-seq methodology pilot run. (A) Read coverage plot of the HEK293 (n=1) tRNAs prepared using ALICE-tRNA-seq. All reads start from the 3' of the tRNAs with over 60% of reads passing the anticodon loop. (B) Breakdown of fragment determination from the ALICE-tRNA-seq pilot run. 5' tRNA fragments are almost eliminated using the ALICE-tRNA-seq library preparation.

3.5 ALICE-tRNA-sequencing in HEK293 cells and its comparison to published tRNA methodologies

3.5.1 ALICE-tRNA-seq library generation and quality control

As aforementioned, HEK293 cells were grown to 70% confluency and the extracted RNAs went through the same steps as the pilot ALICE-tRNA-seq run (see Figure 3.7 A-B), except 4 replicates were prepared in order to compare to published tRNA-seq methodologies. Figure 3.16 highlights the quality control points from extraction to sequencing. Total RNAs were run on a 1% agarose gel to check for integrity and degradation (Figure 3.16 A). The 28S and 18S bands appeared intact, demonstrating suitability for use in the mRNA-seq and ALICE-tRNA-seq library preparations. The tRNAs were deacylated and size selected on a 6% TEB-Urea gel (Figure 3.16B) with good saturation of the mature tRNAs except for replicate 2. The tRNAs then underwent the ALICE-tRNA-seq library preparation with cDNAs being get extracted after reverse transcription (Figure 3.16 C). As expected, the concentration of replicate 2 was lower than the rest of the replicates. The tRNAs were then circularized, relinearized, PCR amplified and then gel extracted on a 5% TBE gel (Figure 3.16 D). The concentrations and average size were determined by Qubit and Tapestation respectively, before being pooled to 1nM and sequenced on a NextSeq5000. Total RNAs from the same HEK293 cells were poly(A) selected (mRNA selection) and prepared for sequencing using the TruSeq RNA Library Preparation Kit v2.

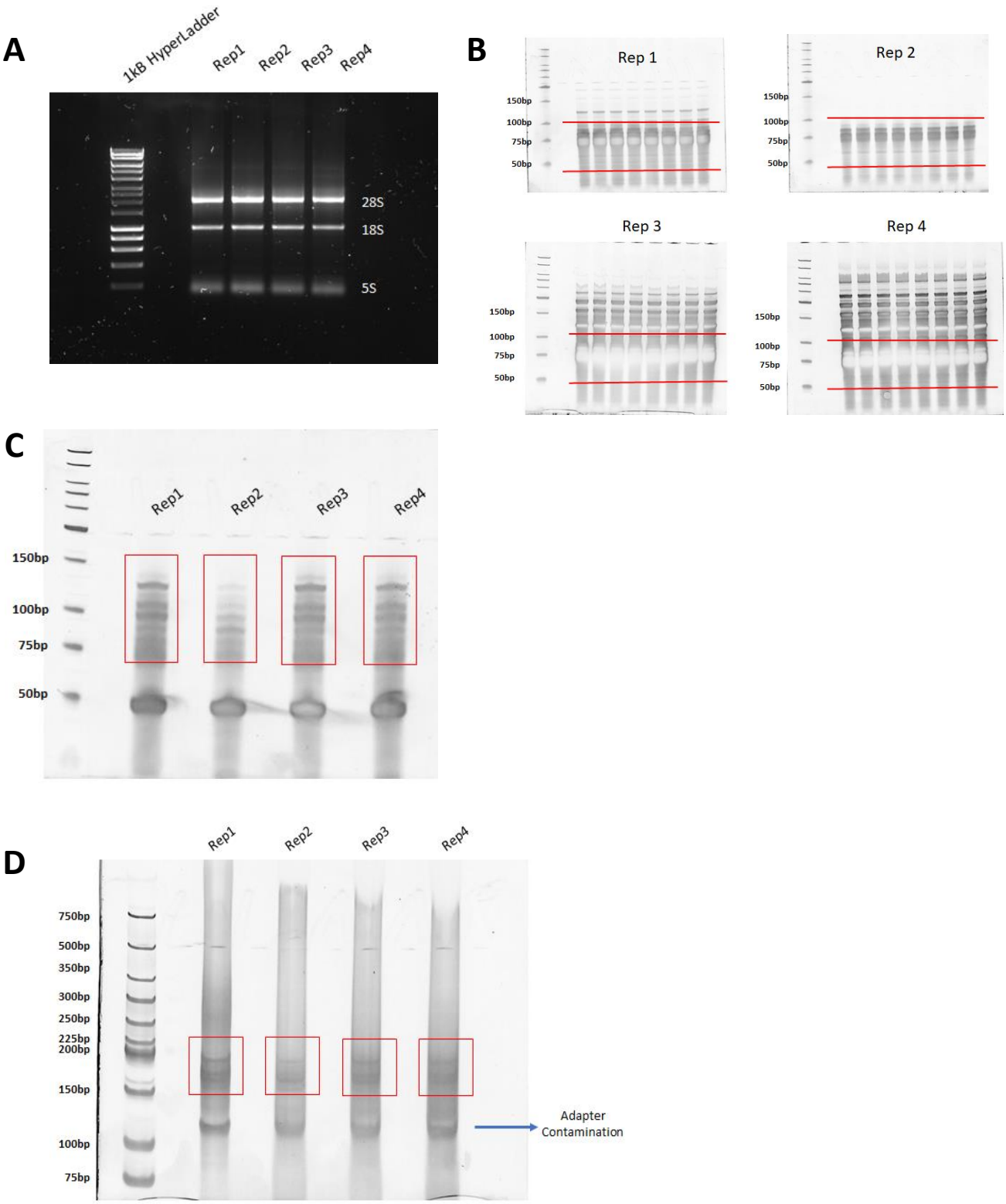


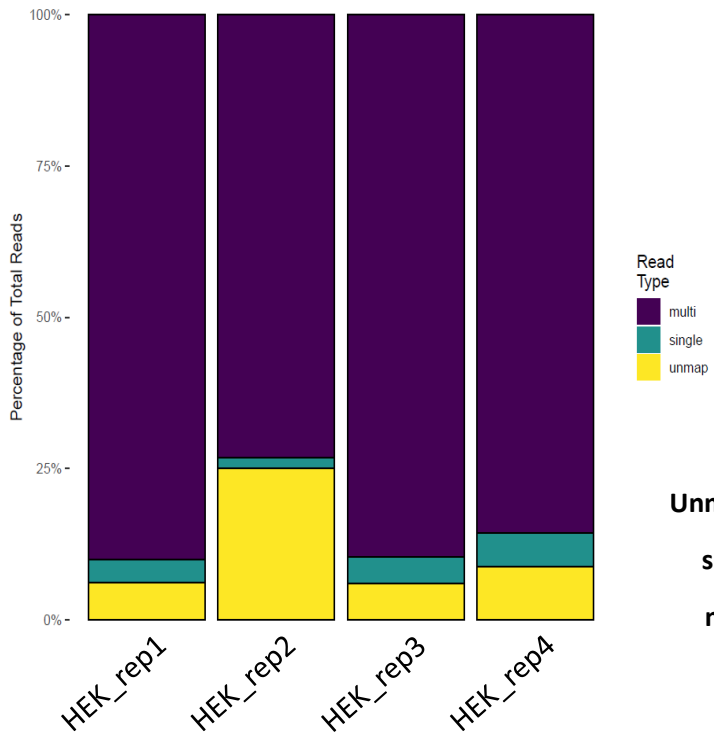
Figure 3.16 HEK293 ALICE tRNA-seq library preparation to be compared to published tRNA-seq methodologies. (A) 1% agarose gel of HEK 293 Total RNA. 28S and 18S bands (and their ratio) confirm no degradation has occurred during extraction. (B) 6% TBE-urea gel for tRNA size selection. Mature tRNAs appear saturated except for rep2. (C) 6%TBE-Urea gel of Post RT gel extraction (red box highlighting excised bands). (D) 5% TBE gel of final libraries to be sequenced. Rep2 was low in concentration and had adapter contamination.

3.5.2 ALICE-tRNA-seq analysis and quality control

The first quality control check to be made on the HEK293 ALICE-tRNA-seq preparation was checking the unmapped percentage and its comparison to single and multimapping events (i.e. tRNAs that map to their associate gene vs unmapped) (Figure 3.17 A). As expected, replicate 2 had to be removed from downstream analyses, with 25.08% of its reads not being mapped to the tRNAs or the genome. This confirmed the adapter contamination seen in the Tapestation when determining the average library size. Replicates 1,3 and 4 all had less than 9% unmapped reads and over 85% multimapping (Figure 3.17 A). The multimapping is a good indication as to the presence of mature tRNAs and therefore a good initial quality control. This is because tRNAs generally have multiple copies across the genome and so many of the tRNAs in the mature tRNA database (see section 3.3.2) generated have multiple copy numbers. The RNA types that the reads map to were also over 90% for tRNAs across the four replicates (Figure 3.17 B-C).

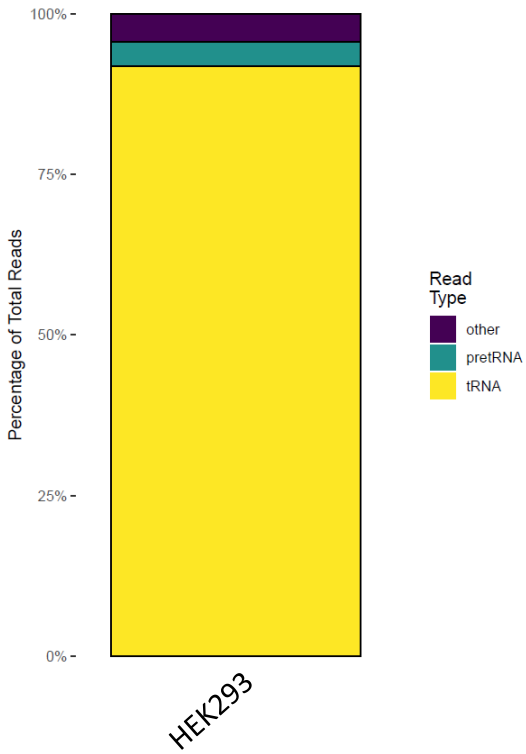
The other quality control step checked before further analysis was the read distribution (Figure 3.18 A) and fragment determination (Figure 3.18 B). As expected, tRNAs starting from the 3' position were plotted, normalized across replicates 1, 3 and 4 (Figure 3.18 A). Over 60% of reads pass the anticodon loop with good distribution across the isodecoders. However, the percentage of full length tRNAs drastically improved in this run, with 26.91%, 31.38% and 20.59% of all mapped reads being full length tRNAs across replicates 1, 3 and 4 (in comparison to 6.02% in the pilot ALICE-tRNA-seq run). From this point onwards, replicate 2 was removed from all downstream analyses.

A



	rep1	rep2	rep3	rep4
Unmapped	6.13%	25.08%	6.04%	8.74%
single	3.79%	1.77%	4.42%	5.56%
multi	90.08%	73.15%	89.54%	85.70%

B

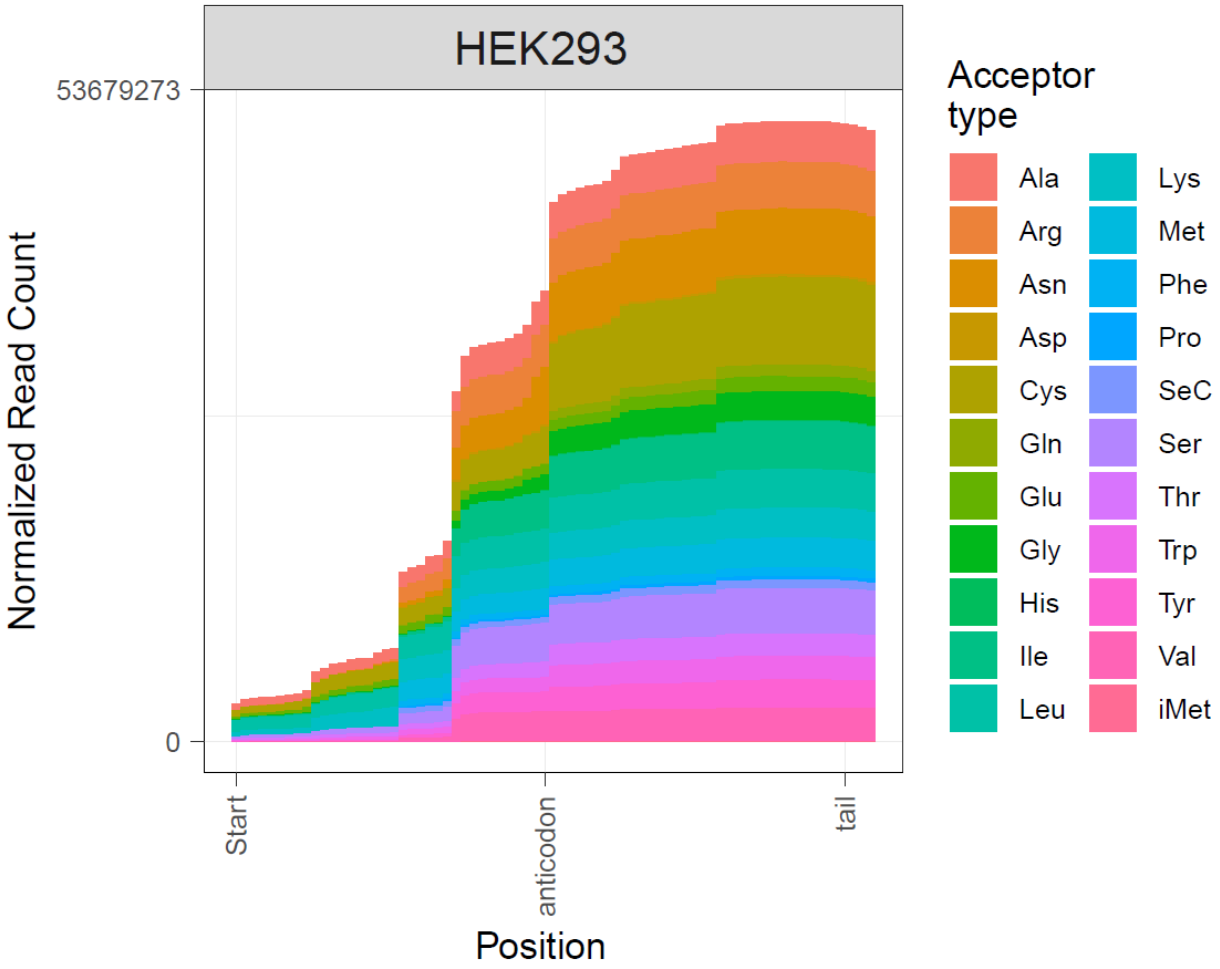


C

	HEK1	HEK2	HEK3	HEK4
tRNA	94.14%	96.57%	92.34%	91.22%
pretRNA	4.07%	2.78%	3.90%	3.76%
havana	0.00%	0.00%	0.00%	0.00%
snRNA	0.00%	0.00%	0.00%	0.00%
mirbase	0.00%	0.00%	0.00%	0.00%
RNA_pseudogene	0.00%	0.00%	0.00%	0.00%
ribozyme	0.00%	0.00%	0.00%	0.00%
sRNA	0.00%	0.00%	0.00%	0.00%
rRNA	0.00%	0.00%	0.00%	0.00%
misc_RNA	0.00%	0.00%	0.00%	0.00%
snoRNA	0.00%	0.00%	0.00%	0.00%
scaRNA	0.00%	0.00%	0.00%	0.00%
other	1.79%	0.66%	3.76%	5.03%

Figure 3.17 Quality control of HEK293 ALICE-tRNA-sequencing results in mapping. (A) Plot showing percentage mapping of reads that underwent multimapping, a single mapping event and unmapped reads. Due to the adapter contamination seen on the replicate 2 Tapestation, the unmapped reads were higher compared to the rest of the replicates, as expected. Replicate 2 was taken out from all downstream analyses. (B) Mapping of reads back to RNA type. As expected, more than 90% of all reads mapped to tRNAs with less than 4.1% of reads mapping to pre-tRNAs. (C) Breakdown of mapping to different RNA types across the 4 replicates.

A



B

	HEK1	HEK2	HEK3	HEK4
Full Length	26.91%	17.28%	31.38%	20.59%
3' tRNA Fragments	66.74%	78.32%	62.34%	72.58%
5' tRNA Fragments	0.00%	0.00%	0.01%	0.01%
Other	6.35%	4.39%	6.27%	6.82%
Total	13,818,791	8,907,264	10,541,343	9,128,165

Figure 3.18 Quality control of HEK293 ALICE-tRNA-sequencing results in fragment determination and coverage. (A) Read coverage plot of tRNA sequences normalized across 3 replicates (replicate 2 removed). As expected, reads start from the 3' with over 60% of all reads passing the anticodon loop. (B) Breakdown of fragment determination across the 4 replicates. Percentage of full length tRNAs vastly improve in comparison to the pilot ALICE-tRNA-seq run. Replicate 2 is removed from all downstream analyses.

3.5.3 Comparison of ALICE-tRNA-seq to published tRNA-seq methodologies

The ALICE-tRNA-seq analysis done for the HEK293 cells are compared to other tRNA-seq methodologies published in literature, conducted on RNA from HEK293 cells. These span a large range of different methods using different approaches to tRNA-seq, summarised below.

The ARM-seq (Cozen et al., 2015) methodology was developed following similarly, to that of standard small RNA-seq methodologies. Namely, adapters are bound to both the 5' and 3' ends of the tRNAs and then undergo reverse transcription and PCR amplification before being sequenced. Hydro-tRNA-seq followed in similar fashion (Gogakos et al., 2017), except they opted to fragment the tRNAs to lengths of approximately 20-30nt (similar to that done in classical RNA-seq except the lengths of the fragments are much shorter). These then have 3' and 5' adapters ligated to the fragmented tRNAs. The difference between these two methodologies is that ARM-seq utilized AlkB to demethylate modified bases prior to library generation, whereas Gogakos et al used fragmentation as a means of avoiding truncation during reverse transcription, by reducing the number of modifications on each fragment. They then used iterative mapping to try and map the sequencing reads to a curated list of mature tRNAs.

The DM-seq methodology (Zheng et al., 2015), like ALICE-tRNA-seq, used WT and D135S AlkB to remove as many methylated bases from the Watson-Crick face of modified tRNAs, prior to library preparation. They avoided 5' adapter ligation by binding a DNA adapter to the 3' of tRNAs and then reverse transcribed their tRNAs using TGRIT (Thermostable group II intron reverse transcriptase). Whenever TGRIT comes across a modified base, it places a random nucleotide in its place as opposed to the stalling of cDNA production.

QuantM-seq (Pinkard et al., 2020) is a circularization based tRNA-seq methodology that uses a commercially available AlkB (which is equivalent to the WT AlkB used in ARM-seq, DM-seq and ALICE-tRNA-seq) to demethylate modified bases. It also utilises a double stranded DNA adapter, where one strand anneals to the 5' of the tRNA and the other strand anneals to the 3' end of the tRNA. Like ALICE-tRNA-seq, the 3' end of the adapter has divergent PCR primer binding sites facing opposite directions. Therefore, after 3'/5' adapter ligation/annealing, the sequences are reverse transcribed and circularized. Since the PCR primer binding sites are facing opposite directions, they PCR directly off of the circularized product, as done in ribosome profiling methodologies (Ingolia et al., 2012), to generate their libraries.

When comparing ALICE-tRNA-seq to the other methodologies, the first aspect that was considered was the tRNA counts distributions across all of the methodologies (Figure 3.19 A). To calculate this, $\log_2$(raw counts + 1) were taken and plotted as a boxplot, with the median counts taken across all of the methodologies. This was because the total number of reads varied from approximately 900,000 reads in ARM-seq to 24,000,000 reads in DM-seq. As expected, there was no normal distribution across the methodologies. However, the distributions seemed to be similar with each methodologies' replicates.

Next, a principal component analysis (PCA) was done to see which methodologies clustered together (Figure 3.19 B). Since a PCA is a measure of variance across multiple components, it was expected that methodologies that cluster together would be similar in either their library preparation or their read counts. As expected, Hydro-seq and ARM-seq, two methodologies that take the classical RNA-seq route ligating 3' and 5' adaptors to their tRNAs, cluster closely together. ALICE-tRNA-seq and QuantM-seq, two circularization based methodologies cluster closely along the first principal component (representing 59.54% of

variance across all the methodologies), along with DM-seq. However, all three methodologies split across the 2nd principal component (representing 19.96% of variance across all the methodologies), with ALICE-tRNA-seq clustering inbetween QuantM-seq and DM-seq.

Next, the correlation coefficients for each replicate from within each of the methodologies was calculated in order to show correlations across the methodologies and to add statistical power to the principal component plots (Figure 3.19 C-D). These were calculated using spearman's rank analysis and the statistical significance for the correlation coefficients were also calculated, with $p<0.05$ set as significant. As expected from the PCA plot, each of the replicates from within each of the experiments correlated with R^2 values greater than 0.99, confirming good reproducibility within all of the methodologies. ARM-seq and Hydro-seq correlate really strongly with each other (with their correlations being statistically significant) with an $R^2>0.8$. ALICE-tRNA-seq and the QuantM-seq were the only other methodologies that strongly correlated, with correlations being statistically significant ($p < 0.05$) and an $R^2>0.83$. This was also expected since the two methodologies are circularization-based with regards to their library preparations. DM-tRNA-seq did not correlate strongly with any of the sequencing methodologies (with it strongest correlation being with ALICE-tRNA-seq with an $R^2=0.38$) but were not statistically significant.

The isodecoder distribution was then compared across the methodologies to elucidate potential bias towards particular isodecoders (Figure 3.20). ARM-seq and Hydro-seq presented the weakest distributions with almost 50% of all reads in ARM-seq mapping to tRNA-Glu (23.02%), tRNA-Gly (13.89%) and tRNA-Lys (14.28%). Similar biases exist in Hydro-seq, with more than 50% of all reads mapping to tRNA-Glu (39.55%), tRNA-Lys (9.36%) and

tRNA-Val (11.15%). The use of just WT AlkB during demethylation prior to library preparation, using high salt concentrations in the demethylation reactions and attempting to ligate 3' and 5' adapters to the tRNAs could be the cause behind these biases. ALICE-tRNA-seq, DM-tRNA-seq and QuantM-seq all provided a much more proportionate distribution across all 22 isodecoders, with each individual technique showing bias towards select isodecoders. This was the case for tRNA-Arg (19.44% and 15.27%) for ALICE-tRNA-seq and QuantM-tRNA-seq respectively and tRNA-Val (18.2%) for DM-tRNA-seq.

Lastly, the isodecoder distributions in all of the tRNA-seq methodologies were compared to the amino acid abundance calculated from HEK293s (Figure 3.21). Briefly, the codons coding each of the amino acids were counted for every transcript in the RNA-seq dataset and normalized to the abundance (transcripts per million, TPM) of each transcript and then were normalized again across its 4 replicates. That ensures that high abundance messages such as house-keeping genes do not skew the percentage read share for each codon. These were then compared to the tRNA-seq isodecoder distributions in the tRNA-seq methodologies. Since the RNA-seq and tRNA-seq came from the same RNA samples for the ALICE-tRNA-seq methodology, their comparison is the fairest. Firstly, what seemed like a bias in the ALICE-tRNA-seq for tRNA-Arg (19.44% read share in tRNAs), is not a bias since the most popular encoded amino acid from the codon read share from the RNA-seq was Arg (10.69%). This will also be true for the QuantM-tRNA-seq with a tRNA-Arg read share of 15.27%. The RNA-seq data also confirms the excess tRNA-Glu in ARM-seq and Hydro-seq (23.02% and 39.55% tRNA read share respectively), with the codons from the RNA-seq having a read share of only 5.34%.

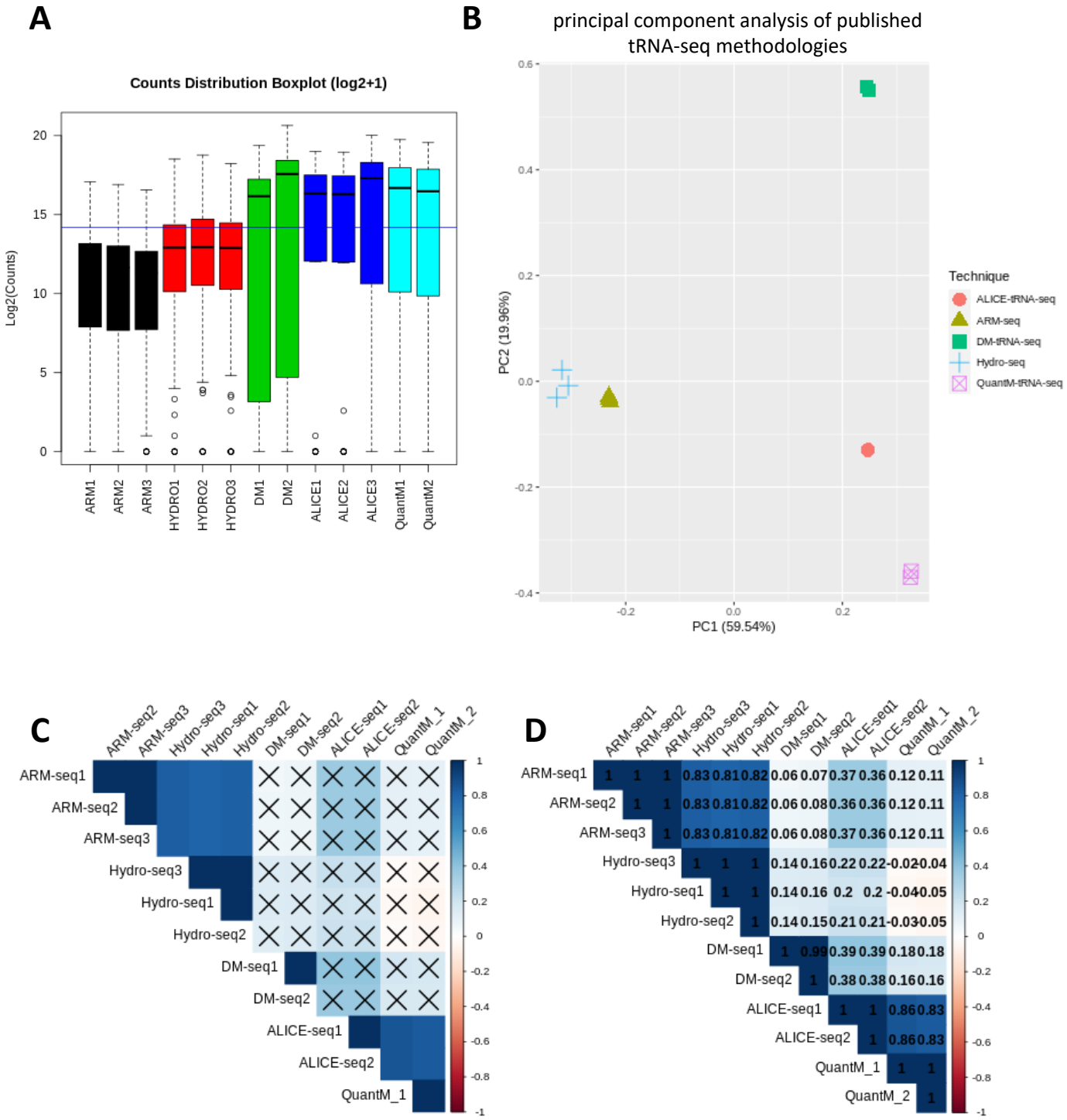


Figure 3.19 ALICE-tRNA-seq isoacceptor distribution comparison to other published tRNA-seq methodologies. (A) Read count distribution of each replicate across the different methodologies. The blue line represents the median counts across of the methodologies. (B) principal component analysis of the five methodologies. All individual replicates cluster together for each methodology with small RNA-seq based methods (Hydro and Arm-seq) clustering and circularization based methods clustering (ALICE and QuantM-seq). PC1 = 59.54% and PC2 = 19.96%. Replicates will overlap if similar in tRNA-counts, like in the case of ALICE-tRNA-seq. (C) Correlation plot with R^2 values represented in the colour map. Statistically significant correlations have no cross ($p < 0.05$), where non-significant correlations are illustrated with a cross (X) within the comparison quadrant. (D) Correlation plot with R^2 values represented in the colour map. Numbers within the comparison quadrants are the R^2 values of each comparison.

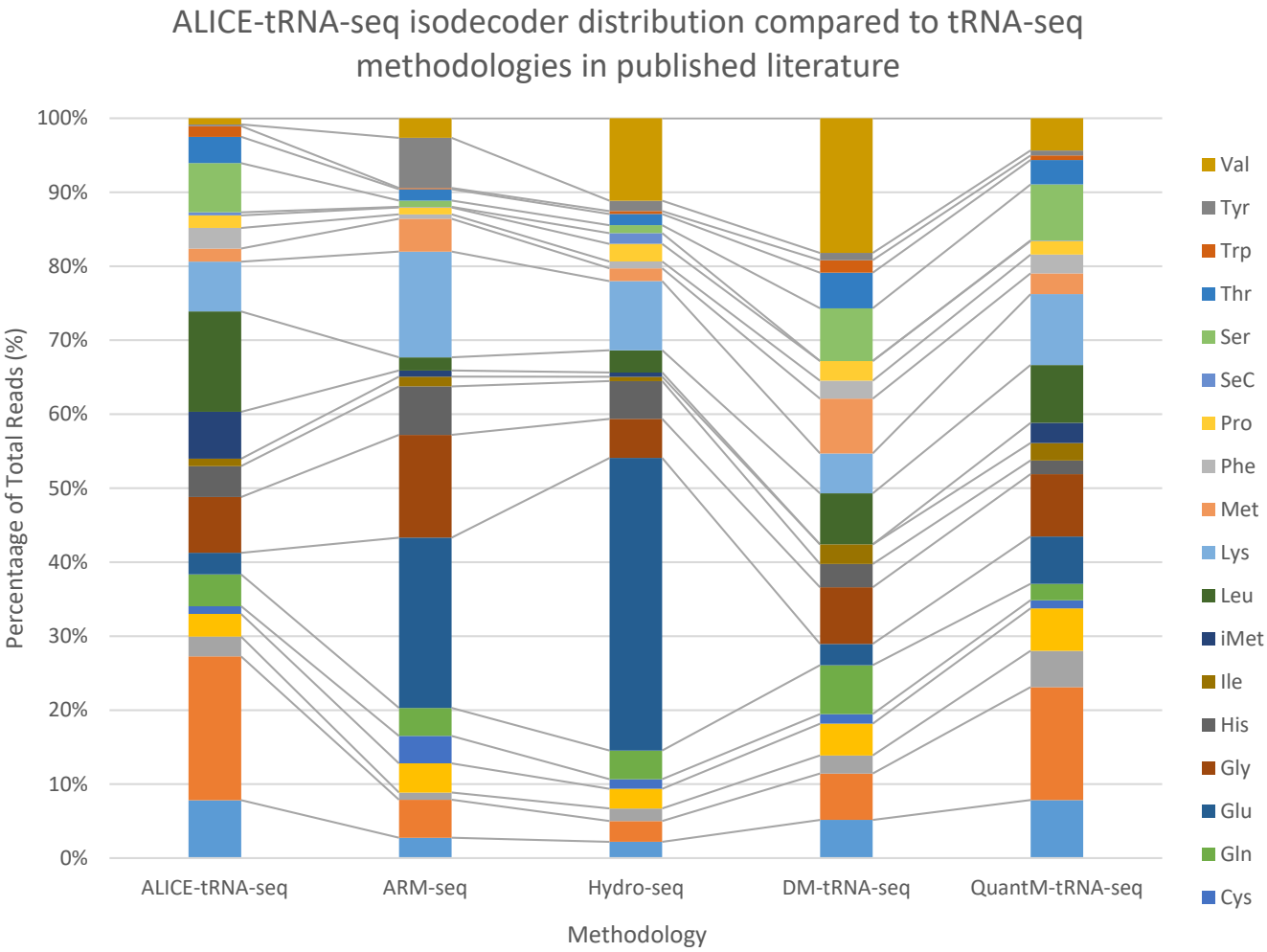


Figure 3.20 ALICE-tRNA-seq isoacceptor distribution comparison to other published tRNA-seq methodologies. Isodecoder counts were normalized within each of the experiments using counts per million (CPM) and the isodecoder read share was subsequently calculated for all 22 amino acids.

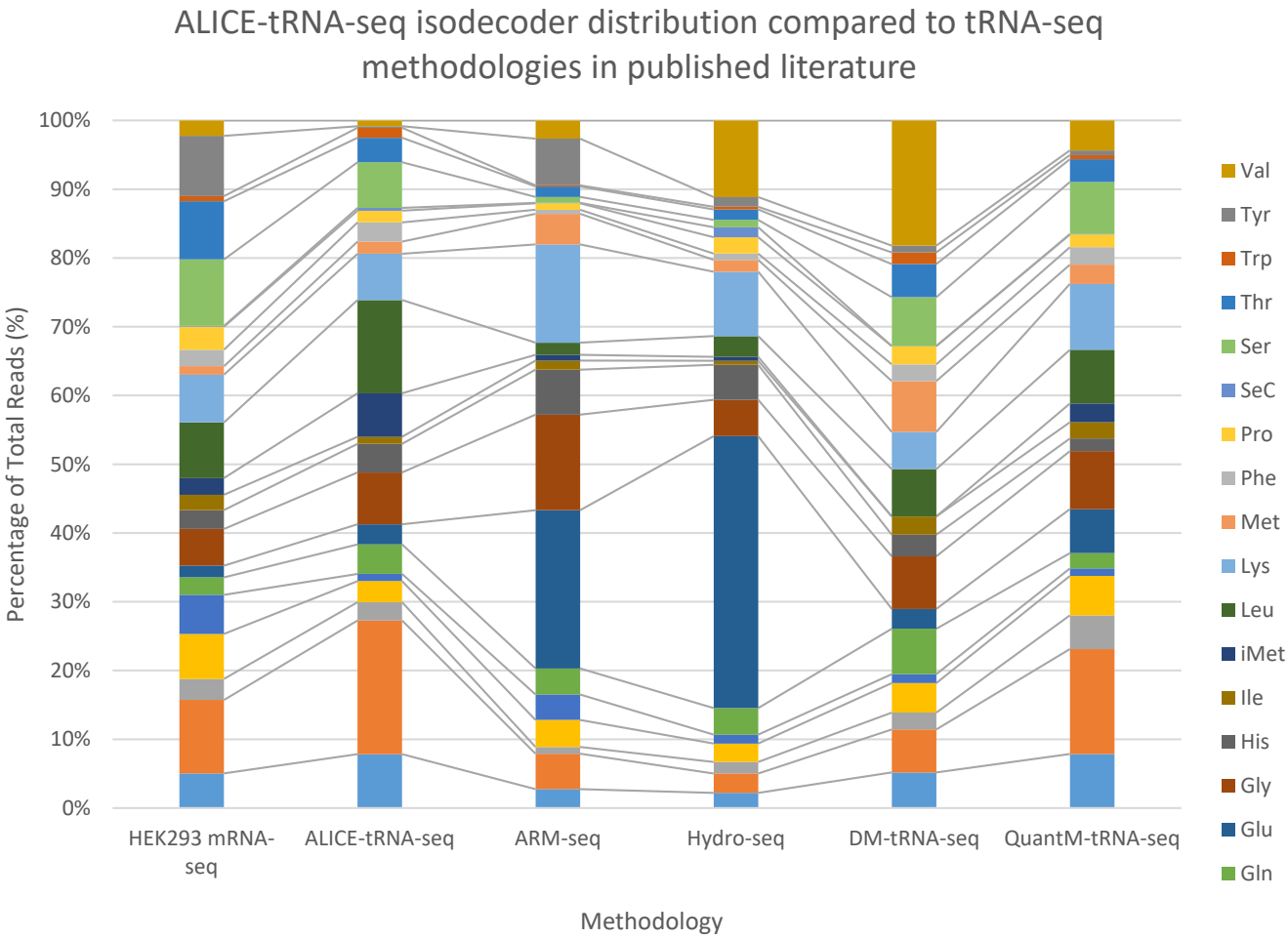


Figure 3.21 Comparison of Isodecoder distribution in ALICE-tRNA-seq and other published tRNA-seq methodologies to HEK293 RNA-seq data. The amino acid distribution from RNA-seq of HEK293's were calculated, normalized to gene abundance and compared to the isodecoder distribution in ALICE-tRNA-seq and other published tRNA-seq methodologies.

3.6 Chapter discussion

In this chapter, we have shown the importance of using both WT and D135S AlkB in an effort to remove as much methylated modifications on the Watson-Crick face of tRNAs prior to library preparations and was shown through primer extension assays and tRNA-sequencing. We also showed the importance of optimizing salt concentrations in the AlkB buffer, something not done by any of the published tRNA-seq methodologies and could enhance the variety of tRNA isodecoders and isoacceptors identified.

We also showed the problems with using an approach that aims to ligate a 5' adapter to tRNA-sequencing libraries and should be avoided. This could be due to the high level of structure at the 5' end of tRNAs and the techniques that circumvent 5' adapter ligation (namely ALICE-tRNA-seq, QuantM-seq and DM-seq), all had proportionate distributions with regards to their isodecoder read shares. We also showed this via sequencing of tRNAs using both ARM-seq and ALICE-tRNA-seq.

With regards to correlation between all the tRNA-seq methodologies, we showed that the ARM-seq and Hydro-seq methodologies (in their current state) were the worst techniques that can be used to measure relative isodecoder abundance. Both techniques presented huge biases towards select tRNAs and would not be suitable for studies relating to codon optimality and their comparison to the tRNAome. However, ARM-seq had always been advertised as a technique especially suitable for the study of 3' and 5' tRNA fragments and tRNA halves. Hydro-seq on the other hand, claimed to be a methodology that can be used to quantify the tRNAome, which we have shown not be accurate.

It was also clear that ALICE-tRNA-seq, QuantM-seq and DM-seq all had favourable isodecoder read share distributions. All three methodologies also circumvent 5' adapter

ligation to their tRNAs in their library preparations. Interestingly, ALICE-tRNA-seq correlates strongest to DM-seq of all the methodologies (even through the correlations were not statistically significant), and the isodecoder distributions for ALICE-tRNA-seq seemed to lie between DM-seq and QuantM-seq which were at opposing extremes (this can also be confirmed in the PCA plot (Figure 3.19 B)).

ALICE-tRNA-seq and QuantM-seq were the two sequencing methodologies that best matched up to the HEK293 RNA-seq dataset. The conclusion to which methodology is stronger to measure the tRNA pool would therefore lie in their library preparations. ALICE-tRNA-seq is the superior methodology because of the use of a 3' pre-adenylated adapter that requires a free monophosphate hydroxyl group (found in mature tRNAs and not found in tRNAs that have not gone through maturation and CCA addition). QuantM-seq on the other hand, utilises a double stranded DNA adapter, where one strand anneals to the 5' end of the tRNA and the other strand anneals to the 3' end of the tRNA (which would have the ability to bind all tRNAs, whether they be mature tRNAs or precursor tRNAs). This targeting in ALICE-tRNA-seq allows us to be confident that the "3' tRNA fragments" we sequence are genuine mature tRNAs that were truncated at the reverse transcription step and not bona fide 3' fragments. The same cannot be said for QuantM-seq, with its biggest limitation being (like DM-seq) that there is no way to distinguish which tRNA reads are genuine, bona fide 3' tRNA fragments and 3' tRNA halves, in comparison to reads that are mature tRNAs but the reverse transcriptase stalled due to a modification in the tRNA sequence.

Chapter 4: Using ALICE-tRNA-seq to investigate changes in the tRNA pool in *in vitro* cell models of quiescence and senescence

4.1 Chapter Introduction

tRNA availability and its link to the translation of the mRNA population has long been an area of interest and has harboured conflicting reports. Links have been made between the decoding accuracy of mRNAs, mRNA stability, translational efficiency and protein folding with tRNA availability (O'Brien et al., 2014; Drummond and Wilke 2008; Man and Pipel 2007). However, the lack of a high resolution technique to accurately measure the tRNA-ome at a single tRNA level to study these links has been a major limitation in the field.

Many diseases, such as cancer, rely heavily upon the alteration of a cell's gene expression signature, with cancers in particular, increasing proliferative capacity by both transcriptional and translational mechanisms (Weinstein et al., 2013; Schwanhüusser et al., 2011). As different tissues vary in both their tRNA-ome and mRNA-ome (Dittmar et al., 2006; Hernandez-Alias et al., 2020), it is important to be able to measure these to try to elucidate the interconnection between them and the molecular mechanisms that govern this.

Guimaraes et al conducted investigations into the effects of serum-deprivation in mouse embryonic fibroblasts and also reported the enrichment of A/T ending mRNA codons but found no differential tRNA expression. This led to them proposing that rather than "proliferative tRNAs" being upregulated to meet the transcriptomic demand, there is a global tRNA expression increase to meet the translational boost required in proliferative cells. This could therefore overcome a potential bottleneck in the translation of proliferative mRNAs that are enriched in A/T ending codons (Guimaraes et al., 2020). However, the tRNA

pools in these investigations were quantified using northern blotting, a technique that relies on oligonucleotides hybridization (the limitations of which are discussed in section 3.1).

In response to stresses and oncogene signalling, healthy cells utilize senescence, a stable cell cycle arrest event as a failsafe mechanism to prevent tumour progression, making escape from senescence a well-established hallmark of cancer (Innes and Gil, 2019; Kuilman et al., 2010; Hanahan and Weinberg, 2000). Although *in vivo* models of stress-induced and oncogene-induced senescence have powerfully established the significance of this protective mechanism (Michaloglou et al., 2005; Kang et al., 2011), the establishment of tumour evasion mechanisms and the evolution of malignant transformation can be better understood by utilizing inducible models of stress-induced and oncogene-induced senescence (Innes and Gil, 2019).

In this chapter, we utilize two *in vitro* models of quiescence and senescence to try and elucidate their tRNA and mRNA profiles to examine how they vary upon changes in cell state, in an attempt to add to previous understandings. The first model discussed is the well-established BJ5TA fibroblasts (from the BJ/hTERT cell line), which are driven from proliferation to quiescence via fetal bovine serum (FBS) starvation. tRNA sequencing confirms the downregulation of tRNAs with A/T wobble anticodons and the upregulation of tRNAS with G/C wobble anticodons (position 34) post starvation. This has previously been established at the mRNA level, where the existence of distinct codon usage signatures in genes encoding for multicellular and cell autonomous processes has been reported (Gingold et al., 2014). mRNA-sequencing showed the preference for A/T ending codons in proliferation, and the preference for G/C ending codons in senescence.

The second model utilised in this chapter uses IMR90 primary human fibroblasts, which undergo oncogene induced senescence via the induction of an estrogen receptor and RAS fusion (Dajee et al., 2002). Briefly, a mutant estrogen receptor (ER) has an H-RAS^{G12V} fused to its ligand binding domain, allowing for the activation of RAS via 4-hydroxy-tamoxifen (4-OHT) treatment, resulting in growth arrest and induction of senescence associated phenotypes (Innes and Gil, 2019). This results in the non-treated fibroblasts remaining in proliferation whereas 4-OHT treated fibroblasts are driven into senescence. In this model, the tRNA sequencing shows that tRNAs with A/T wobble anticodons are upregulated, whereas tRNAs with G/C wobble anticodons are downregulated, following oncogene-induced senescence, which is opposite to that observed in the BJ5TA starved fibroblasts which have entered quiescence. However, we do confirm from the mRNA-sequencing the preference for A/T ending codons in proliferation, and the preference for G/C ending codons in senescence.

4.2 Experimental design

As discussed in Chapter 3, the ALICE-tRNA-seq methodology proves to be an effective means of measuring the tRNA pool *in vitro*. Standard RNA-seq (utilizing polyA selection to capture mRNAs), is used in parallel so that both the tRNA-ome and the mRNA-ome can be examined from the same samples. From the tRNA sequencing, the relative pools between conditions are measured, giving an account to isoacceptors/isodecoders being expressed. mRNA-sequencing measures steady state mRNA levels, meaning that differential expression of genes between conditions can be determined. Weighting codon frequencies by mRNA abundance can be used to allow for more accurate determination of the cellular demand for specific tRNAs and the establishment of codon usage preferences of expressed transcripts.

The BJ5TA fibroblasts were driven from proliferation to quiescence via fetal bovine serum (FBS) starvation (hereafter referred to as Fed vs Starved) (Figure 4.1). Fed cells were the wild-type and have no cell state changes occurring (10% FBS in growth medium), whereas the starved cells had no FBS added to their medium, resulting in growth factor deprivation. Both conditions were grown in tandem for 72h to a confluency of 70%. Both the total RNA and small enriched RNAs (<200nt) are extracted from the same harvested cells allowing for mRNAs and tRNAs to be obtained from the same populations of cells.

The primary human fibroblasts IMR90s were driven from proliferation to senescence via the induction of oncogenic RAS (oncogene induced senescence) (Figure 4.1). To achieve this, IMR90s were infected with retrovirus encoding the mutated estrogen receptor with H-RAS^{G12V} fused to its ligand binding domain (hereafter referred to as Ras). Wildtype IMR90s were infected with an empty retrovirus (hereafter referred to as Empty). Senescence was induced via activation of RAS using 4-hydroxy-tamoxifen (4-OHT) treatment. Empty IMR90s

were treated with an empty retrovirus and are therefore unaffected by 4-OHT and remain in proliferation. Both conditions were grown for 10 days post 4-OHT treatment. The cells were then harvested, lysed and total RNA and small enriched RNAs (<200nt) extracted. 5 replicates were generated for both experimental sets.

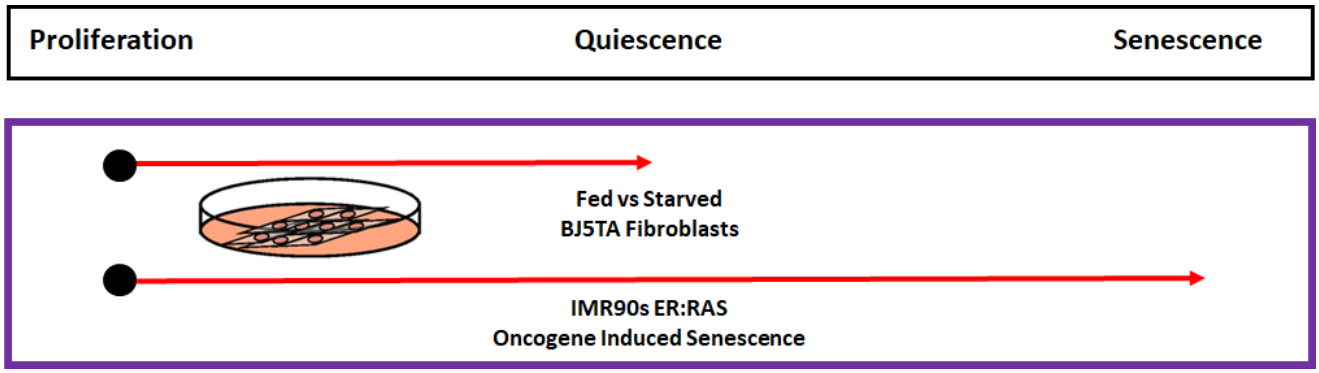


Figure 4.1 Experimental design in trying to establish the tRNA pool *in vitro*. The BJ5TA fibroblasts are driven from proliferation to quiescence via growth factor deprivation (Fed vs Starved). IMR90s (primary human fibroblasts) are driven from proliferation to senescence via oncogene induced senescence. This is achieved by the ectopic expression of the oncogene H-RAS^{G12V} fused to an estrogen receptor (ER:Ras) and induced via 4-hydroxy-tamoxifen (4-OHT) treatment.

4.3 Quality control steps post-extraction and pre-library preparation

4.3.1 Pre-library preparations quality control

Quality control steps were carried out prior to library preparations of the Fed vs Starved samples. Firstly, total RNAs extracted from the replicates were run on a 1% agarose gel to check for RNA integrity (Figure 4.2 A). The gel revealed significant degradation of total RNA from the Fed3 replicate, illustrated by the smearing occurring along the lane. This resulted in both the Fed3 and Starved3 samples being discarded. Starved3 was removed due to the fact that the experimental samples are paired.

New BJ5TA fibroblast replicates were grown, +/- FBS treated and harvested, with extracted total RNAs being run on a 1% agarose gel to check for RNA integrity (Figure 4.2 B). Samples had passed the first quality control steps and were used as input for RNA-seq library preparation.

Small enriched RNAs (<200nt) were deacylated and then size selected for RNAs between 50-100nt on an 8% TBE-Urea gel (Figure 4.2 C). tRNA saturation was clearly visible and rRNAs running at approximately 120nt were avoided (so as to avoid library contamination). tRNAs were then purified and demethylated, prior to being used as input into the ALICE-tRNA-seq library preparation.

The same quality control assessments were carried out for the IMR90 Empty vs Ras experiments. The total RNAs were not degraded (Figure 4.3 A). tRNA saturation was clearly visible for the first 3 replicates for each condition. Empty4, Empty5, Ras4 and Ras5 were size selected on a different day and were oversaturated (hence the darkness observed in the gels). However, inspection of the sample post extraction revealed successful tRNA purification.

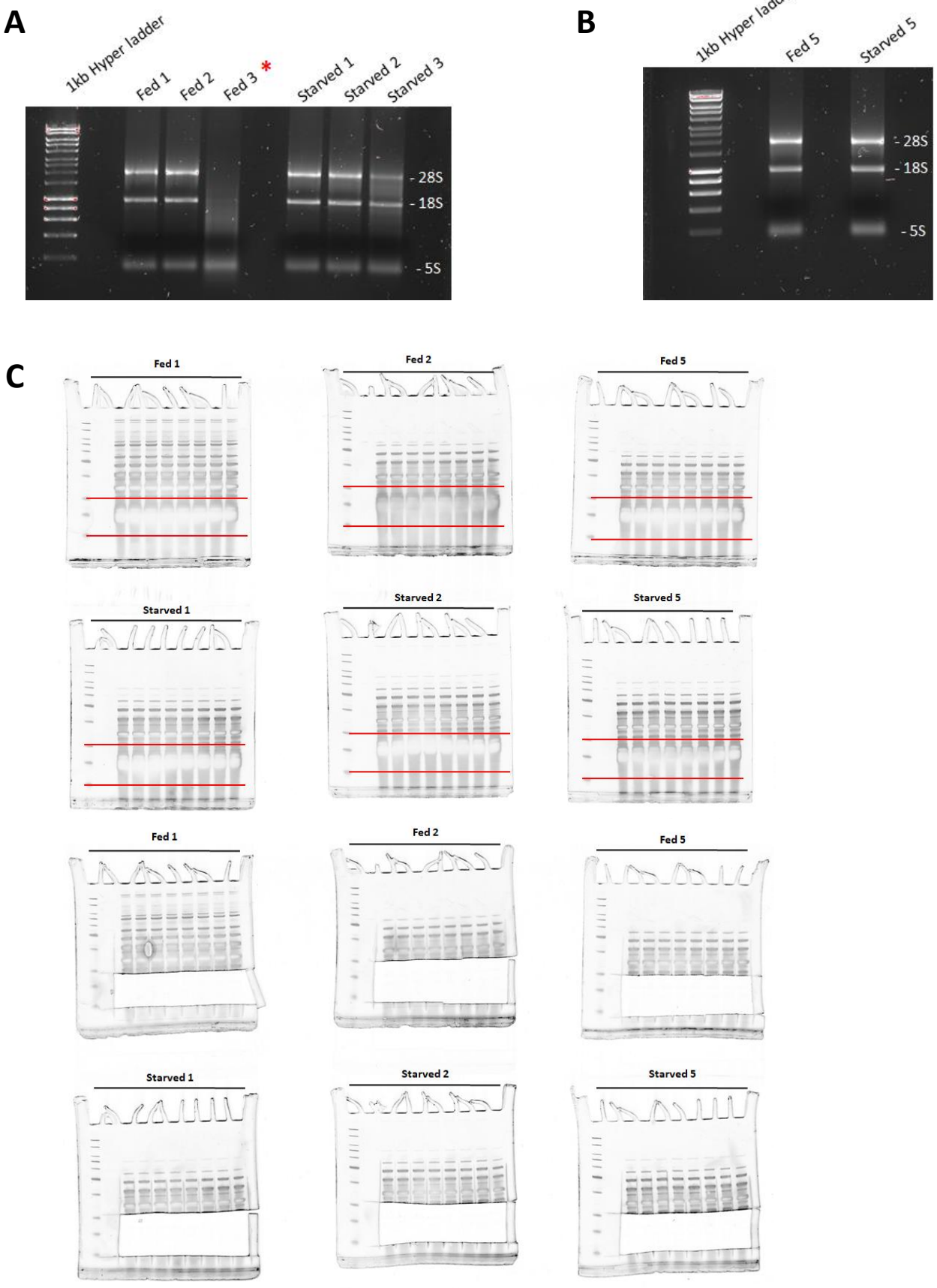


Figure 4.2 Fed vs Starved Total RNA integrity check and tRNA size selection. (A) Total RNAs were run on a 1% agarose gel to check for RNA integrity. Fed3 (red asterisk) was clearly degraded and therefore Fed3 and Starved3 were not used. (B) Fed5 and Starved5 were used as a replacement for the degraded sample in (A). (C) tRNA size selection pre and post excision on a 8% TBE Urea. tRNAs were size selected between 50-100nts

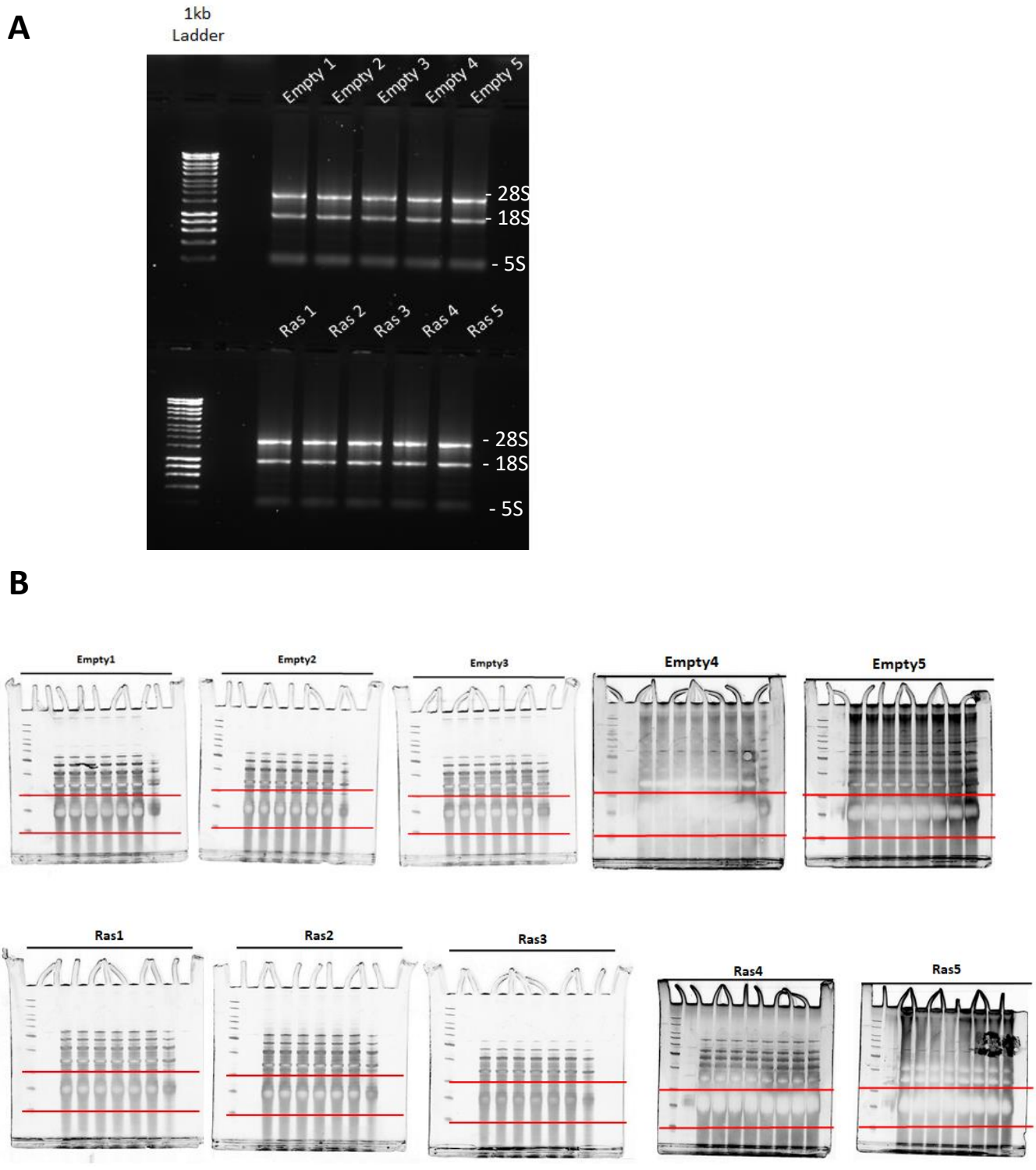


Figure 4.3 IMR90 Total RNA integrity check and tRNA size selection. (A) Total RNAs were run on a 1% agarose gel to check for RNA integrity. (B) tRNA size selection pre-excision on a 8% TBE Urea. The darkness in replicates 4 and five is due to oversaturation. Empty4 does have tRNAs but are too oversaturated. tRNAs were size selected between 50-100nts (red lines)

4.3.2 Library preparations quality control

Quality control steps were carried out post-RT and post-PCR. After reverse transcription, cDNAs between 97-147nt (corresponding to the 50-100nt RNA input length) were gel extracted on a 6% TBE-Urea gel to avoid contamination of excess RT-primer (Figure 4.4 A and Figure 4.6 A) and purified prior to circularization. The gels clearly showed the tRNAs were successfully reverse transcribed, with a good proportion of full length products (~147nt) in addition to the multiple bands, presumably representing truncation events occurring due to modifications on the Watson-Crick face of tRNAs, which the AlkB demethylase enzyme was unable to remove (refer to Table 3.1).

The final libraries were gel extracted after PCR amplification. The final libraries were run on a 5% TBE gel and extracted between 169-219nt (corresponding to the 50-100nt RNA input length), before being purified and pooled for sequencing (Figure 4.4 B and Figure 4.6 B).

In the Fed vs Starved experiment, the concentration of purified library from FBS deprivation replicate2 was too low to pool and therefore, 50% more of the PCR reaction (6µl of PCR reaction loaded instead of 4µl) were loaded onto a 5% TBE-gel to account for the low concentration purified post-excision (Figure 4.5). This was done for both Fed2 and Starved2 since the samples in this experiment were paired.

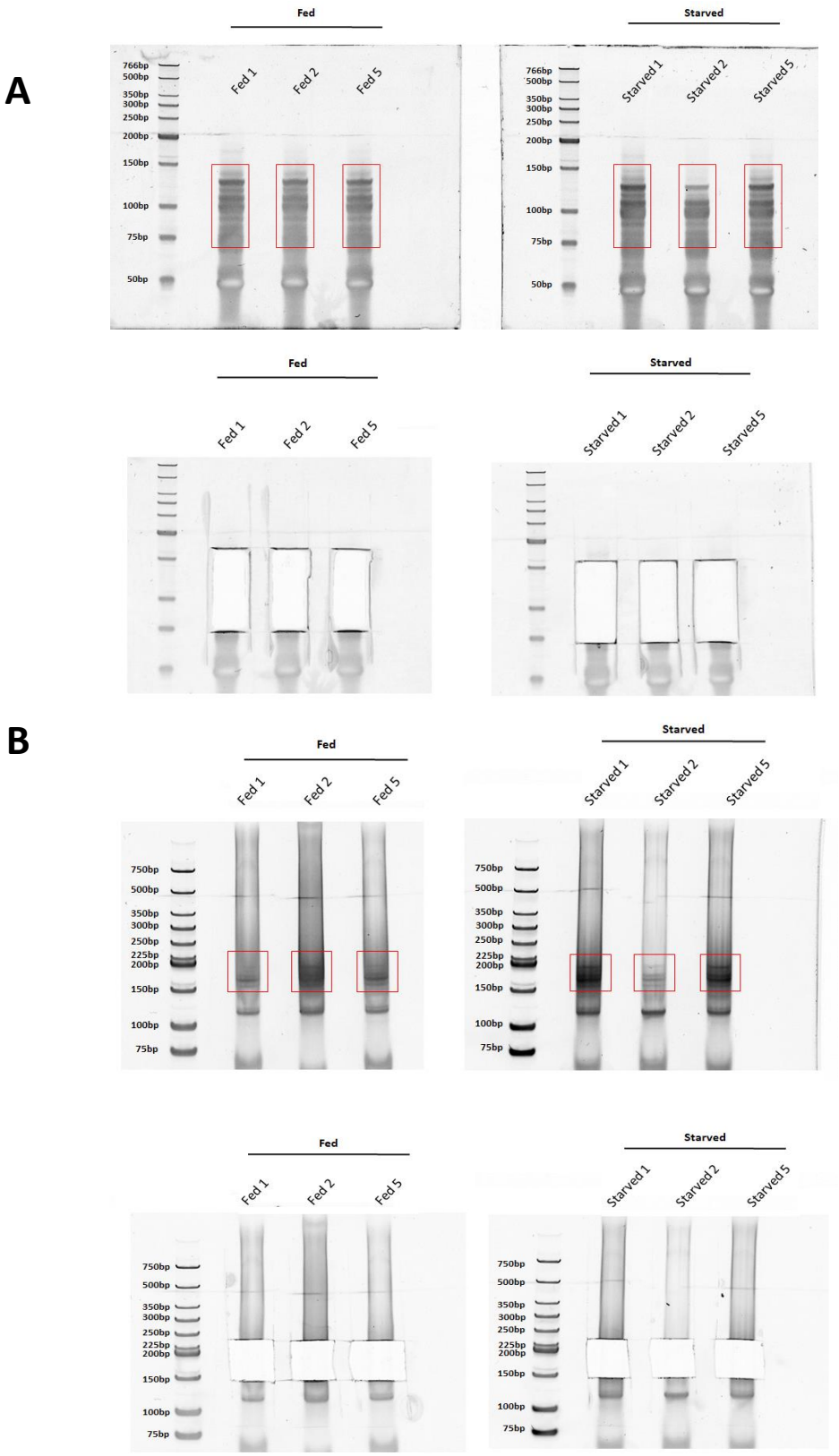


Figure 4.4 Fed vs Starved ALICE-tRNA-seq library preparation post RT and post PCR. (A) tRNA libraries are prepared using ALICE-tRNA-seq with cDNAs being gel extracted between 97-147nt. Gels highlight pre- and post-excision on a 6% TBE-UREA gel. (B) cDNAs were PCR amplified and gel extracted after 8 PCR cycles. DNA libraries were extracted between 169-219nt. Gels highlight pre- and post-excision on a 5% TBE-UREA gel.

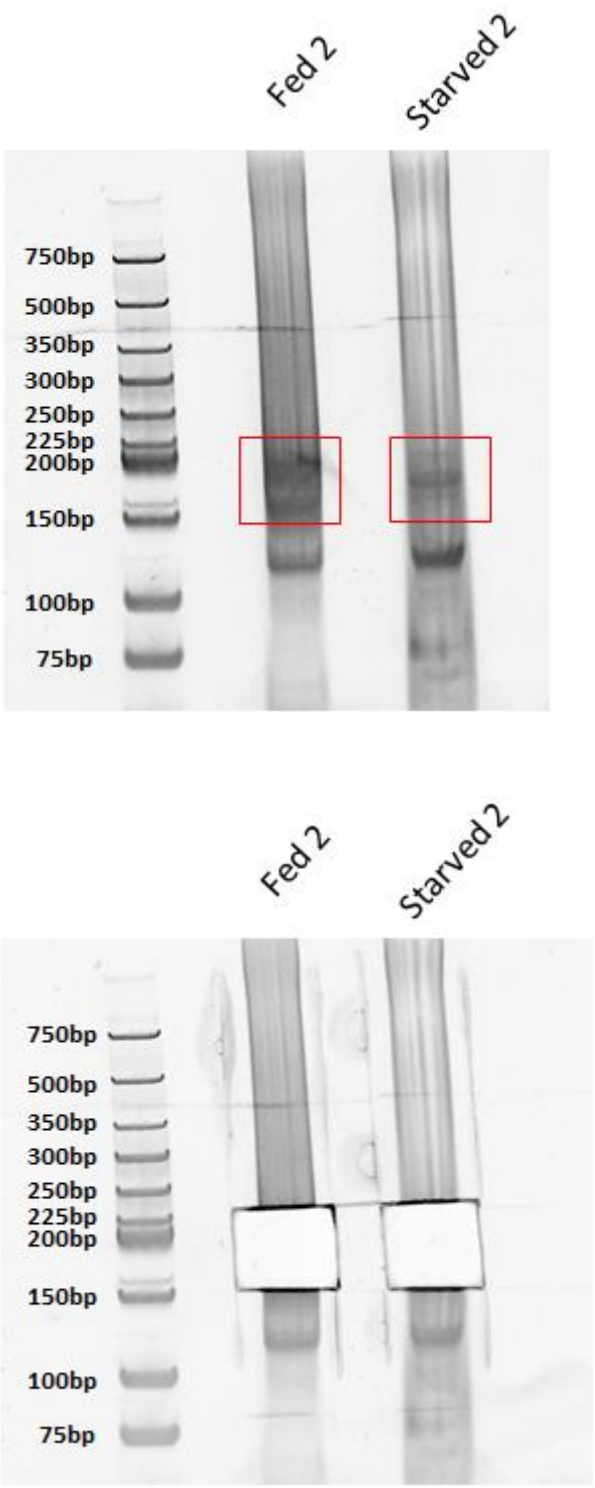
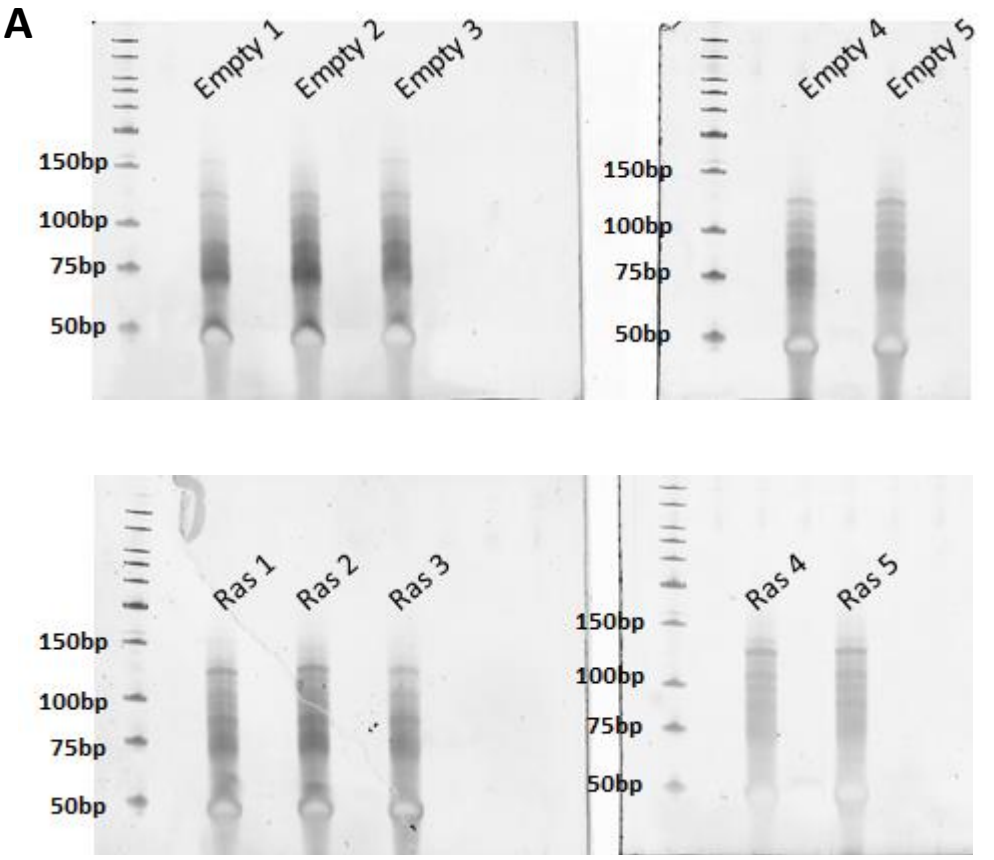


Figure 4.5 50% more Fed2 and Starved2 samples were loaded to increase the concentration of DNA library extracted. Concentration of Fed2 and Starved2 extracted from Figure 4.3 was too low. 50% more of the library was loaded onto a 5% TBE gel to account for the low concentrations initially extracted.


A
Empty 1
Empty 2
Empty 3
Empty 4
Empty 5
150bp
100bp
75bp
50bp
Ras 1
Ras 2
Ras 3
Ras 4
Ras 5

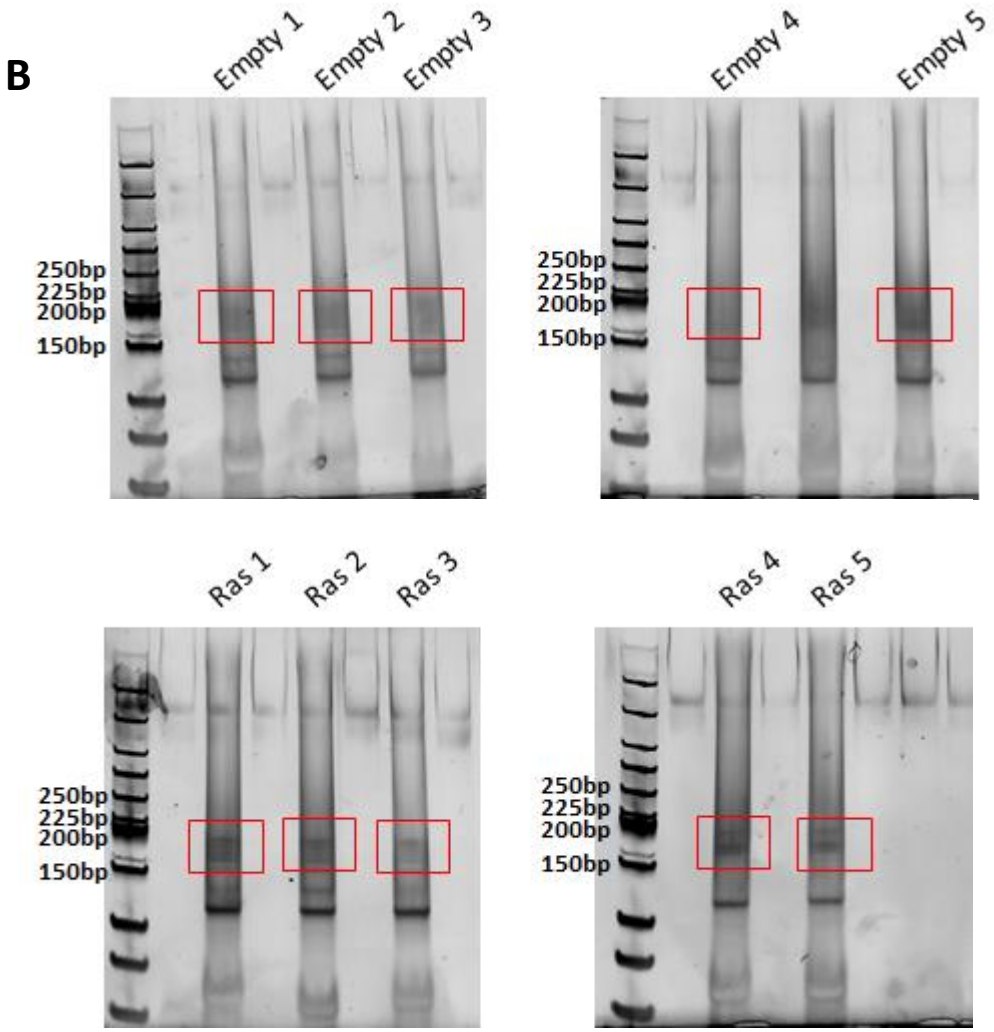

B
Empty 1
Empty 2
Empty 3
Empty 4
Empty 5
250bp
225bp
200bp
150bp
Ras 1
Ras 2
Ras 3
Ras 4
Ras 5

Figure 4.6 IMR90 ALICE-tRNA-seq library preparation post RT and post PCR. (A) tRNA libraries are prepared using ALICE-tRNA-seq with cDNAs being gel extracted between 97-147nt. Gels highlight pre-excision on a 6% TBE-UREA gel. (B) cDNAs were PCR amplified and gel extracted after 8 PCR cycles. DNA libraries were extracted between 169-219nt. Gels highlight pre-excision on a 5% TBE-UREA gel

4.4 Quality control steps post-sequencing

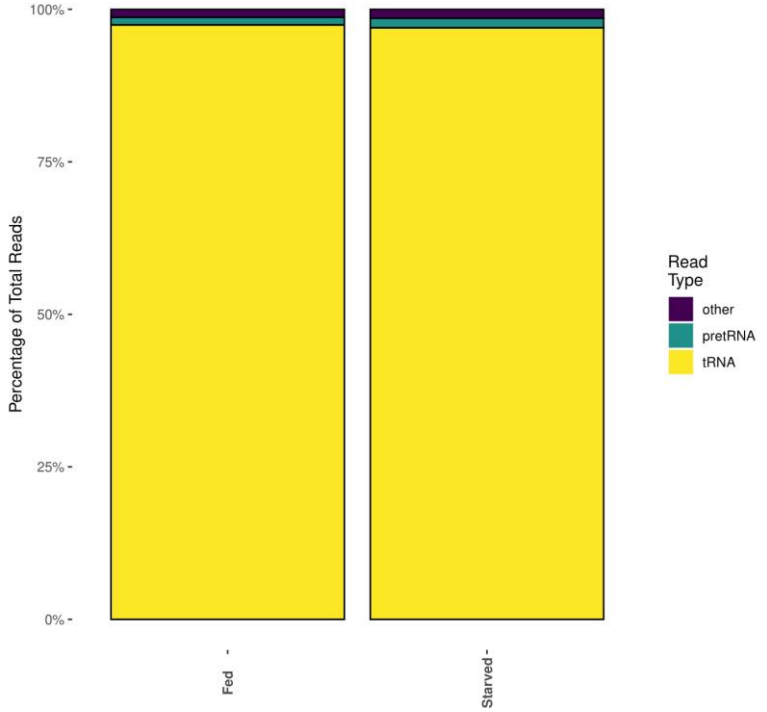
4.4.1 Quality control checks on RNA type mapping and isoacceptor distribution

Sequenced reads had adapters trimmed and were mapped to the genome and the curated database of mature tRNAs. The first quality control step was to elucidate the percentage of reads mapping to tRNAs.

In Fed vs Starved, more than 95% of the mapped reads, mapped successfully to tRNAs (Figure 4.7 A). The distribution of reads to their decoded amino acids were also measured (Figure 4.7 B). There were no clear biases towards any tRNA (as seen in the pilot experiments carried out in Section 3.3).

In the IMR90 Empty vs Ras, more than 79% of the mapped reads, mapped successfully to tRNAs. There were more reads mapping to “other RNAs”, 11.22% and 13.27% for the Empty and Ras conditions respectively (Figure 4.8 A). Upon further examinations, these RNAs mapped to mitochondrial tRNAs (which are removed from the analysis, data not shown). There was also no observed bias in the distribution of reads encoding amino acids (Figure 4.8 B).

A



B

	Fed	Starved
tRNA	96.47%	95.85%
pretRNA	1.24%	1.55%
havana	0.00%	0.00%
snRNA	0.13%	0.20%
mirbase	0.01%	0.01%
rRNA_pseudogene	0.08%	0.07%
ribozyme	0.00%	0.00%
sRNA	0.00%	0.00%
rRNA	0.54%	0.49%
misc_RNA	0.10%	0.20%
snoRNA	0.16%	0.20%
scaRNA	0.00%	0.00%
other	1.27%	1.42%

C

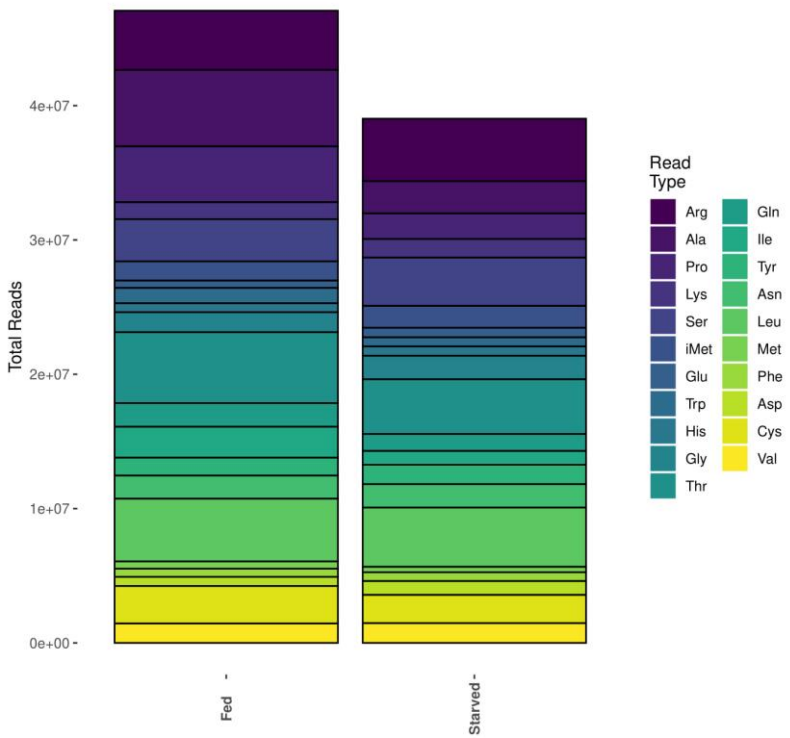
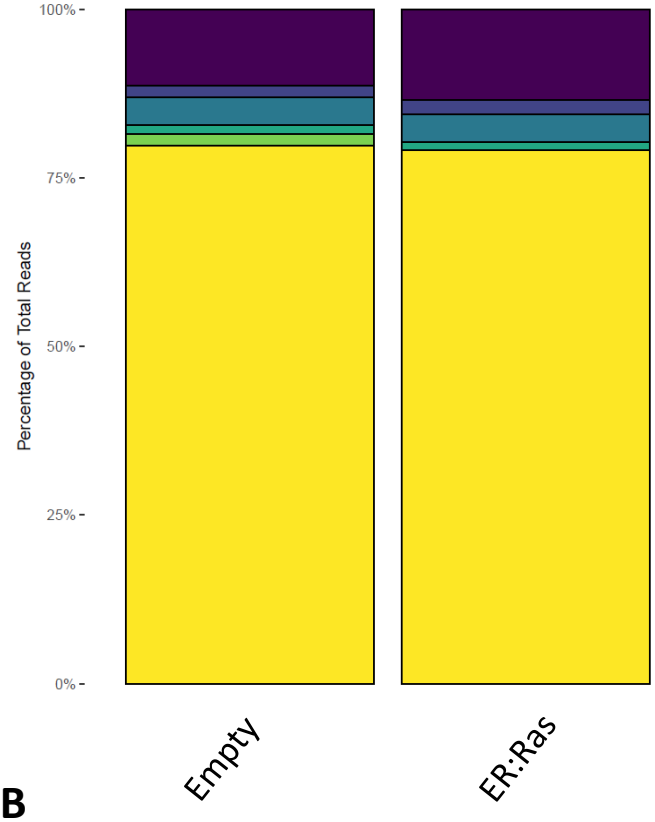


Figure 4.7 Fed vs Starved quality control check on RNA types sequenced and the isoacceptor distributions across the two conditions. (A) Percentage of total reads mapping to their respected RNA types. More than 95% of reads mapped to tRNAs. (B) Breakdown of RNA types mapped to (C) Isoacceptor distribution for both conditions. No clear biases towards any particular isoacceptor were found.

A



	Empty	ER:Ras
tRNA	79.16%	78.05%
pretRNA	1.63%	0.67%
havana	0.00%	0.00%
snRNA	1.41%	1.17%
mirbase	0.01%	0.01%
rRNA_pseudogene	0.43%	0.37%
ribozyme	0.02%	0.03%
sRNA	0.00%	0.00%
rRNA	4.02%	4.03%
misc_RNA	0.38%	0.29%
snoRNA	1.71%	2.11%
scaRNA	0.00%	0.01%
other	11.22%	13.27%

B

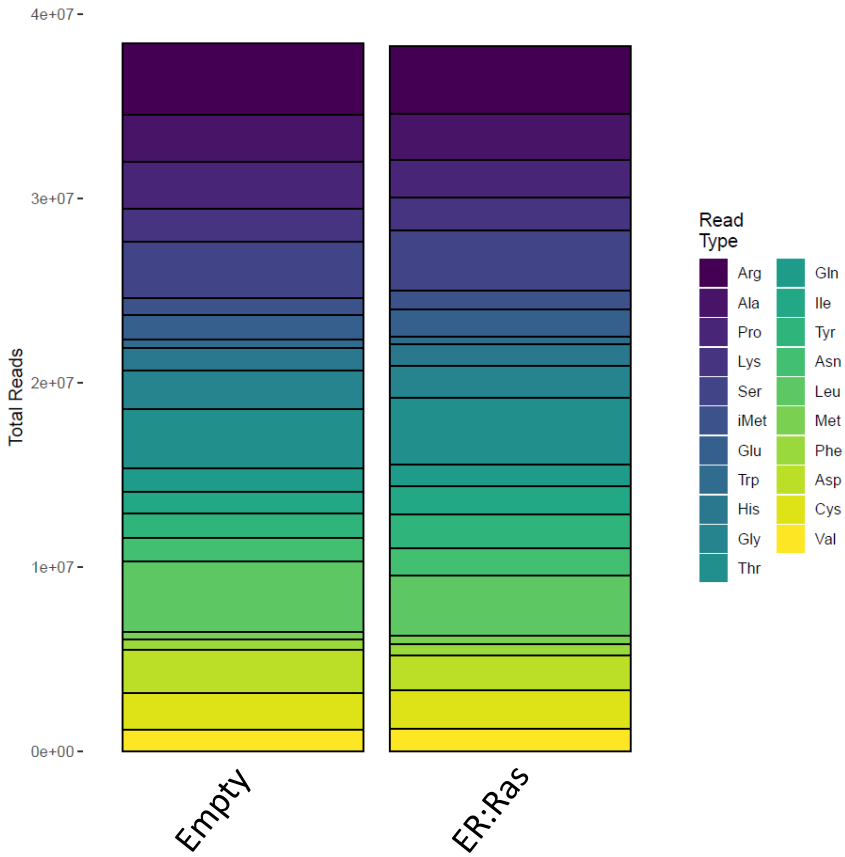


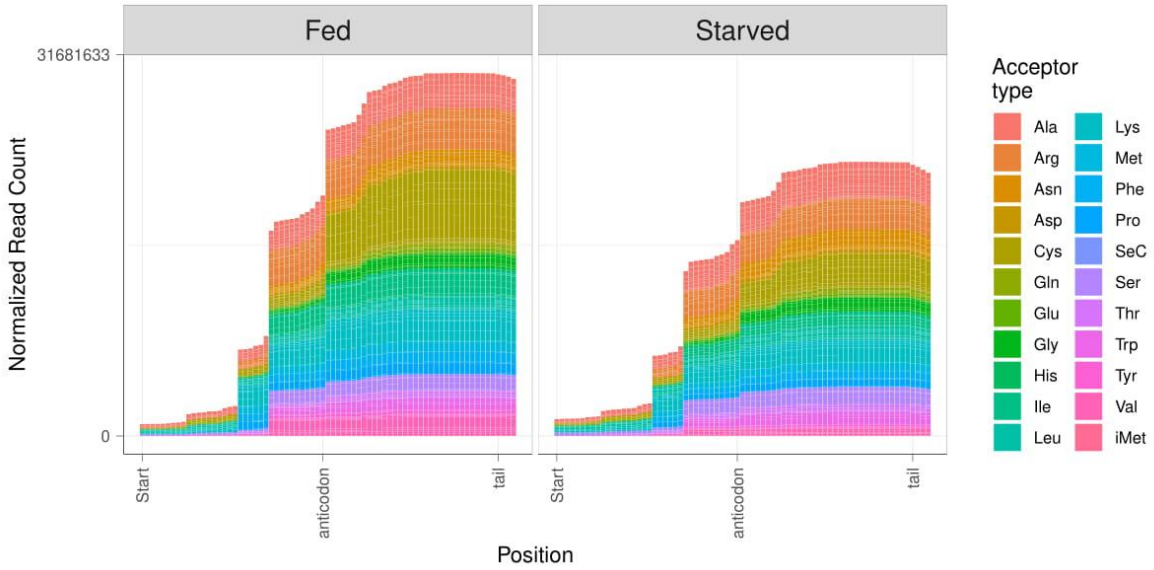
Figure 4.8 IMR90 quality control check on RNA types sequenced and the isoacceptor distributions across the two conditions. (A) Percentage of total reads mapping to their respected RNA types. More than 78% of reads mapped to tRNAs. (B) Breakdown of RNA types mapped to (C) Isoacceptor distribution for both conditions. No clear biases towards any particular isoacceptor were found.

4.4.2 tRNA read coverage and fragment determination

The read coverage for the two experiments was also analysed (Figure 4.9 A and Figure 4.10 A). The reads for each of the replicates were normalized and the tRNA sequences were plotted with respect to where the sequences started. More than 95% of the sequencing reads ended at the 3' end as expected, meaning that the reverse transcription started at this position, which rules out the possibility that these fragments derived from cleaved tRNAs. In Fed vs Starved, more than 60% of the reads overlapped the anticodon loop of the tRNA and more than 55% in Empty vs Ras.

tRNA fragments were also quantified (Figure 4.9 B and Figure 4.10 B). For both experiments, more than 95% of the reads were determined to be either full length tRNAs or sequences that had truncated at the reverse transcription step. No significant 5' fragments were sequenced, as expected.

A

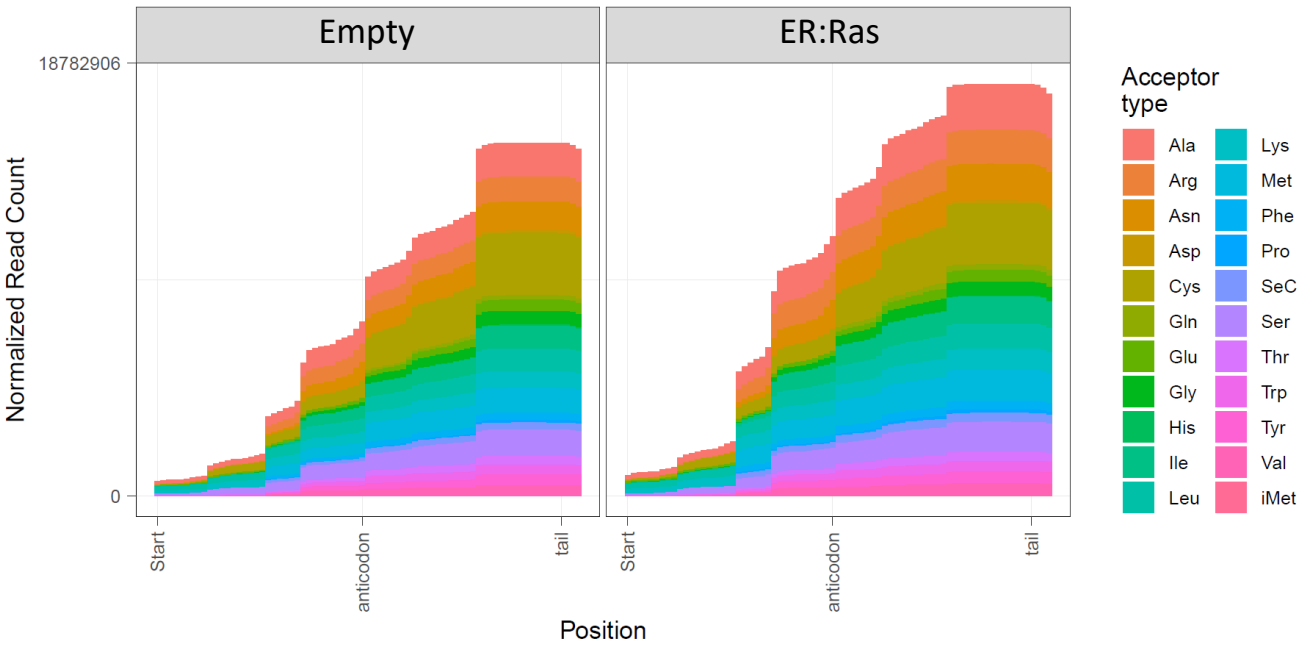


B

	tRNAs					
	Fed1	**Fed2**	**Fed5**	**Starved1**	**Starved2**	**Starved5**
Full Length	13.62%	21.14%	10.11%	20.02%	19.01%	20.94%
3' tRNA Fragments	83.52%	76.76%	87.78%	77.80%	77.51%	76.62%
5' tRNA Fragments	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other	2.85%	2.10%	2.11%	2.17%	3.48%	2.44%
Total Reads	5,830,146	9,547,677	8,201,345	7,697,790	8,441,381	8,014,203

Figure 4.9 Fed vs Starved quality control check on fragment determination and coverage. (A) Read coverage plot of tRNA sequences normalized across 3 replicates. As expected, reads start from the 3’ end with over 60% of all reads passing the anticodon loop. (B) Breakdown of fragment determination across the 3 replicates for each condition. More than 95% of the reads were either sequenced full length tRNAs or sequences that had truncated at the reverse transcription step. No significant 5’ fragments were sequenced as expected.

A



B

	Empty1	Empty2	Empty3	Empty4	Empty5	Ras1	Ras2	Ras3	Ras4
Full Length	33.85%	28.69%	31.35%	10.25%	21.09%	28.40%	32.17%	18.31%	21.43%
3' tRNA Fragments	66.15%	71.31%	68.64%	89.74%	78.90%	71.59%	67.82%	81.68%	78.55%
5' tRNA Fragments	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Other	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%

Figure 4.10 IMR90 quality control check on fragment determination and coverage. (A) Read coverage plot of tRNA sequences normalized across 3 replicates. As expected, reads start from the 3’ end with over 55% of all reads passing the anticodon loop. (B) Breakdown of fragment determination across the 5 and 4 replicates for each condition respectively. More than 98% of the reads were either sequenced full length tRNAs or sequences that had truncated at the reverse transcription step. No significant 5’ fragments were sequenced as expected.

4.4.3 tRNA counts distribution and principal component analysis

The count distribution across the replicates for the Fed and Starved conditions were also calculated (Figure 4.11 A and Figure 4.12 B). Log_2(tRNA counts + 1) were taken to evaluate the distribution of the counts across the replicates and compared to the median across conditions. The median number of counts for each replicate coincided with the condition-wide median, ruling out bias for both experiments.

A principal component analysis was also carried out on the samples (Figure 4.11 B and Figure 4.12A).

In Fed vs Starved, since each replicate for the two conditions were originally harvested from the same cell batch prior to treatment, a batch correction was performed on the samples prior to the PCA being plotted (using Lima batch correct). This allowed for the removal of any technical variance that may have occurred during cell culture, since the experiments were carried out on separate days. The principal component analysis of the tRNA counts showed good separation across the two conditions on PC1 (accounting for 77% of the variance within the experiment).

In Empty vs Ras, Ras1, Ras2, Empty4 and Empty5 were removed from the analysis due to their clustering in the PCA (Figure 4.12 A). The principal component analysis was re-run showing the Empty and RAS replicates separating across PC1, accounting for 92% of the variance.

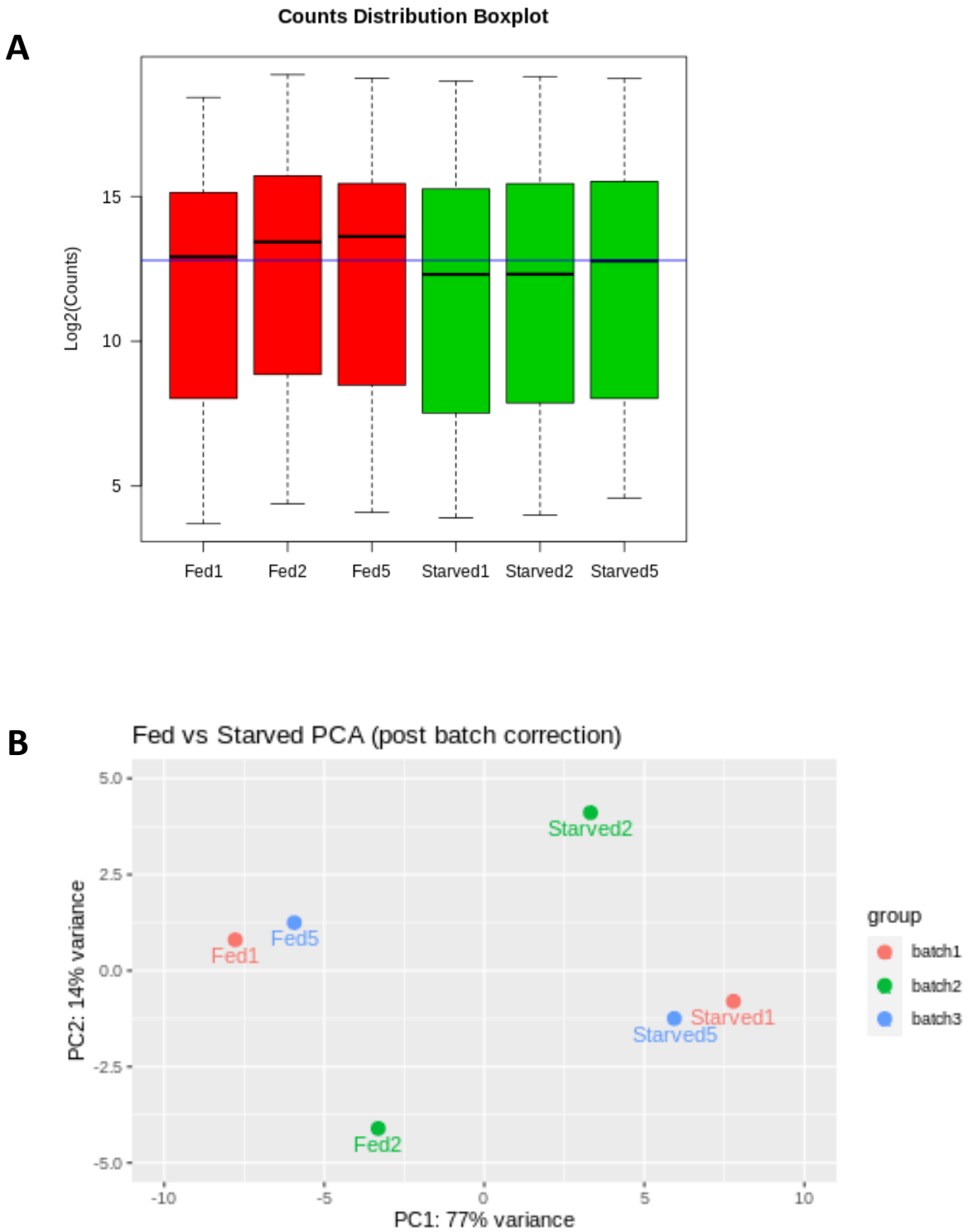


Figure 4.11 Fed vs Starved counts distribution and principle component analysis. (A) Counts distribution of the mapped tRNAs were checked. Log2(counts + 1) were taken across all of the replicates illustrated on a boxplot. All mean counts were close to the median of all counts across the conditions (blue line). (B) Principle component analysis of the tRNA counts to determine where samples clustered. Good separation across the two conditions on PC1 (accounting for 77% of the variance). Lima batch correct was used since the cells were harvested as pairs to remove any technical variation across the replicates.

A

B

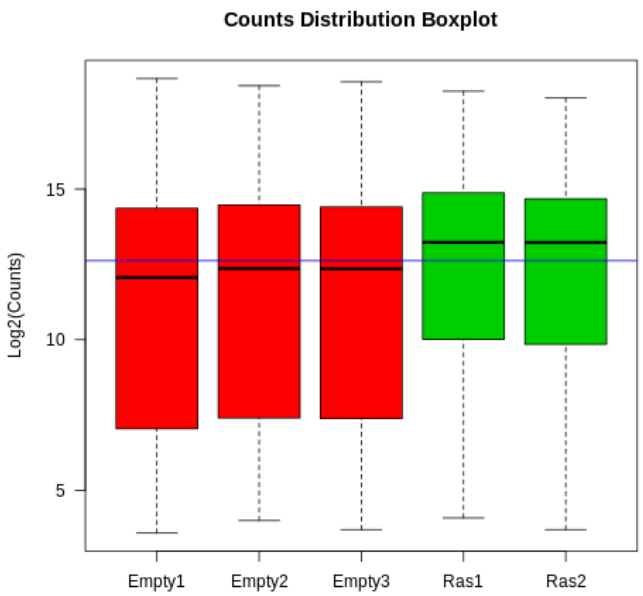


Figure 4.12 IMR90 principle component analysis and counts distribution . (A) Principle component analysis of the tRNA counts to determine where samples clustered. Ras1, Ras2, Empty4 and Empty5 were all removed from the analysis and the PCA rerun. Lima batch correct was used since the cells were harvested as pairs to remove any technical variation across the replicates. (B) Counts distribution of the mapped tRNAs were checked. Log2(counts + 1) were taken across all of the replicates illustrated on a boxplot. All mean counts were close to the median of all counts across the conditions (blue line).

4.4.5 tRNA mismatch frequency at the wobble position to check for adenosine to inosine modification (A34I)

Another quality control step incorporated prior to differential expression was checking for the mismatch frequency at position 34 (tRNA wobble position).

Since there are fewer tRNA anticodons with respect to mRNA codons, tRNAs utilise an adenosine to inosine modification at the tRNA wobble position of select isoacceptors, to accommodate non-cognate base pairing and recognition (Figure 4.13 A). tRNAs also utilize G:T base pairing for select tRNA isoacceptors (Table 4.1).

In next generation sequencing, inosine bases are recognised as guanosines instead of adenosines. Therefore, when reads are mapped back to the mature tRNA database (allowing for only 1 mismatch), reads with the mismatch at position 34 are due to this adenosine to inosine modification (A34I). The tRNA mismatch frequency analysis determined that the inosine modifications only occurred on A bases and not seen in tRNAs with a G/C/T at the wobble position for both experiments (Figure 4.13 B and C).

Amino Acid	mRNA Codon (5' to 3')	tRNA Isoacceptor (5' to 3')	Cognate Base Pairing	A34I Base Pairing	G:T Base Pairing	tRNA Box Type
Arg	CGT	ACG	CGT:ACG	CGT:ICG	-	6-box tRNA
	CGC	Not Expressed	-	CGC:ICG	-	
	CGG	CCG	CGG:CCG	-	-	
	AGG	CCT	AGG:CCT	-	-	
	CGA	TCG	CGA:TCG	CGA:ICG	-	
	AGA	TCT	AGA:TCT	-	-	
Leu	CTT	AAG	CTT:AAG	CTT:IAG	-	6-box tRNA
	CTC	Not Expressed	-	CTC:IAG	-	
	TTG	CAA	TTG:CAA	-	-	
	CTG	CAG	CTG:CAG	-	-	
	TTA	TAA	TTA:TAA	-	-	
	CTA	TAG	CTA:TAG	CTA:IAG	-	
Ser	TCT	AGA	TCT:AGA	TCT:IGA	-	6-box tRNA
	TCC	Not Expressed	-	TCC:IGA	-	
	TCG	CGA	TCG:CGA	-	-	
	AGC	GCT	AGC:GCT	-	-	
	AGT	Not Expressed	-	-	AGT:GCT	
	TCA	TGA	TCA:TGA	TCA:IGA	-	
Ala	GCT	AGC	GCT:AGC	GCT:IGC	-	4-box tRNA
	GCC	Not Expressed	-	GCC:IGC	-	
	GCG	CGC	GCG:CGC	-	-	
	GCA	TGC	GCA:TGC	GCA:IGC	-	
Pro	CCT	AGG	CCT:AGG	CCT:IGG	-	4-box tRNA
	CCC	Not Expressed	-	CCC:IGG	-	
	CCG	CGG	CCG:CGG	-	-	
	CCA	TGG	CCA:TGG	CCA:IGA	-	
Val	GTT	AAC	GTT:AAC	GTT:IAC	-	4-box tRNA
	GTC	Not Expressed	-	GTC:IAC	-	
	GTG	CAC	GTG:CAC	-	-	
	GTA	TAC	GTA:TAC	GTA:IAC	-	
Gly	GGG	CCC	GGG:CCC	-	-	4-box tRNA
	GGC	GCC	GGC:GCC	-	-	
	GGT	Not Expressed	-	-	GGT:GCC	
	GGA	TCC	GGA:TCC	-	-	
Thr	ACT	AGT	ACT:AGT	ACT:IGT	-	4-box tRNA
	ACC	Not Expressed	-	ACC:IGT	-	
	ACG	CGT	ACG:CGT	-	-	
	ACA	TGT	ACA:TGT	ACA:IGT	-	
Glu	GAG	CTC	GAG:CTC	-	-	2-box tRNA
	GAA	TTC	GAA:TTC	-	-	
Gln	CAG	CTG	CAG:CTG	-	-	2-box tRNA
	CAA	TTG	CAA:TTG	-	-	
Lys	AAG	CTT	AAG:CTT	-	-	2-box tRNA
	AAA	TTT	AAA:TTT	-	-	
Ile*	ATT	AAT	ATT:AAT	-	-	2-box tRNA (special case)
	ATC	GAT	ATC:GAT	ATC:IAT	-	
	ATA	TAT	ATA:TAT	ATA:IAT	-	
Asn	AAC	GTT	AAC:GTT	-	-	2-box tRNA (only 1 tRNA expressed)
	AAT	Not Expressed	-	-	AAT:GTT	
His	CAC	GTG	CAC:GTG	-	-	2-box tRNA (only 1 tRNA expressed)
	CAT	Not Expressed	-	-	CAT:GTG	
Phe	TTC	GAA	TTC:GAA	-	-	2-box tRNA (only 1 tRNA expressed)
	TTT	Not Expressed	-	-	TTT:GAA	
Asp	GAC	GTC	GAC:GTC	-	-	2-box tRNA (only 1 tRNA expressed)
	GAT	Not Expressed	-	-	GAT:GTC	
Cys	TGC	GCA	TGC:GCA	-	-	2-box tRNA (only 1 tRNA expressed)
	TGT	Not Expressed	-	-	TGT:GCA	
Tyr	TAC	GTA	TAC:GTA	-	-	2-box tRNA (only 1 tRNA expressed)
	TAT	Not Expressed	-	-	TAT:GTA	
iMet/Met	ATG	CAT	ATG:CAT	-	-	1-box tRNA
Trp	TGG	CCA	TGG:CCA	-	-	1-box tRNA

Table 4.1 A breakdown of the human mRNA:tRNA base pairing that occurs at the ribosome during translation and tRNA degeneracy. All mRNA codons are highlighted in blue (written 5’ to 3’) and tRNA anticodons are highlighted in black (written 3’ to 5’). Cognate binding represents decoding events that take place due to classical Watson-Crick base pairing. Codons that are decoded as a result of the A34I modification at the tRNA wobble position are highlighted in purple. Codons that are decoded as a result of G:T binding are highlighted in green. Not expressed refers to the fact that the tRNA anticodon is not expressed in the human genome and is therefore decoded via A34I modification or G:U non-cognate binding.

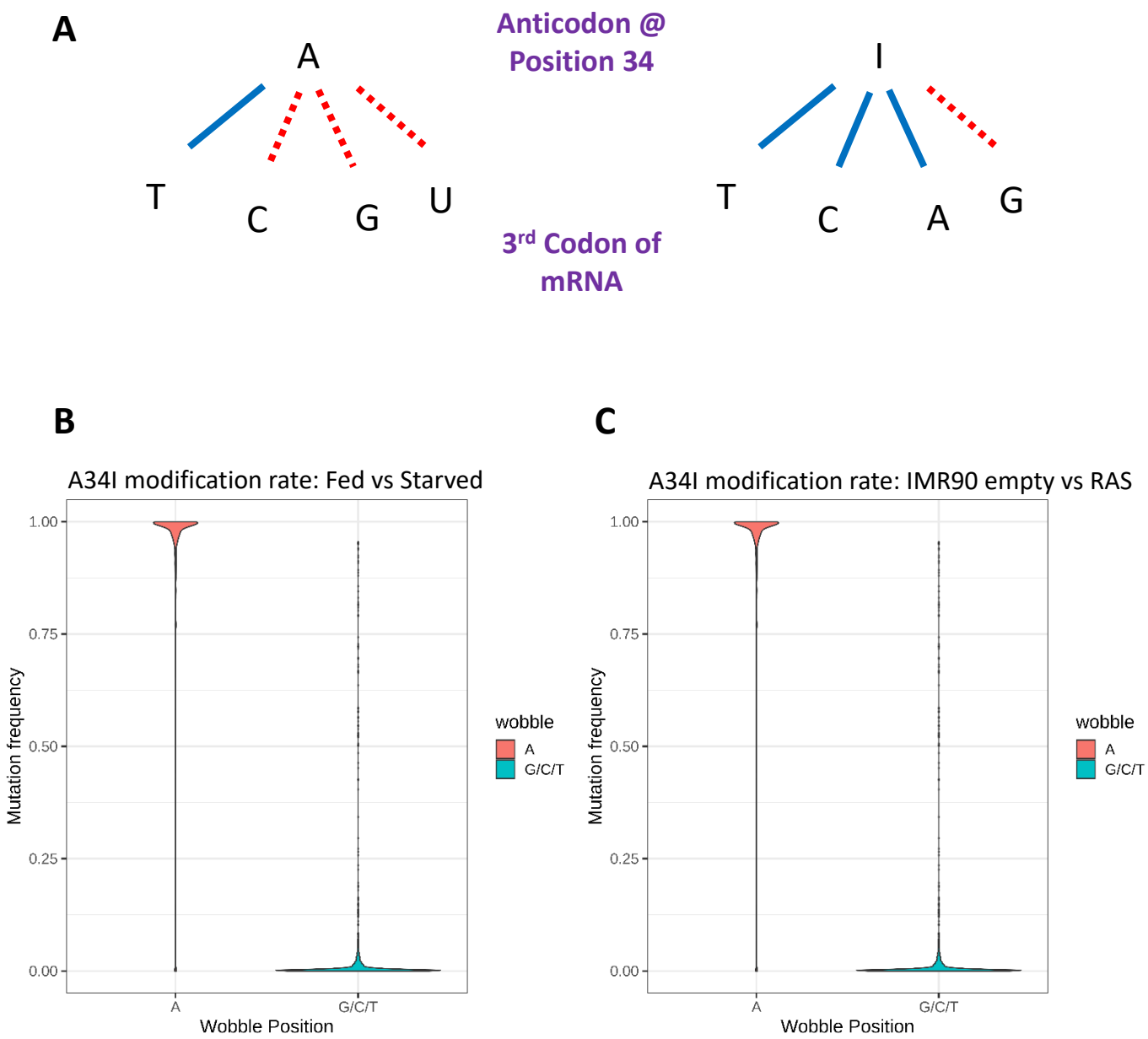


Figure 4.13 tRNA mismatch frequency at position 34 (wobble position) to check for inosine modification (A34I). (A) Schematic of adenosine and inosine base pairing to mRNA codons at the wobble position. Adenosine base pairs with T bases. The inosine modification allows for non-cognate base pairing and recognition of T/C/A bases (bold blue line). Inosine does not pair effectively to G bases (dashed red line). (B - C) Mismatch frequency for tRNAs with A at the wobble position vs G/C/T for both experiments. tRNAs were mapped allowing for 1 mismatch in order to determine the inosine modification at position 34 (wobble position). As expected, inosine modifications should only occur on A bases and not seen in tRNAs with a G/C/T at the wobble position. (B) BJ5TA Fed vs Starved fibroblasts and (C) IMR90 Empty vs Ras.

4.5 Differential tRNA expression changes at the tRNA isoacceptor level

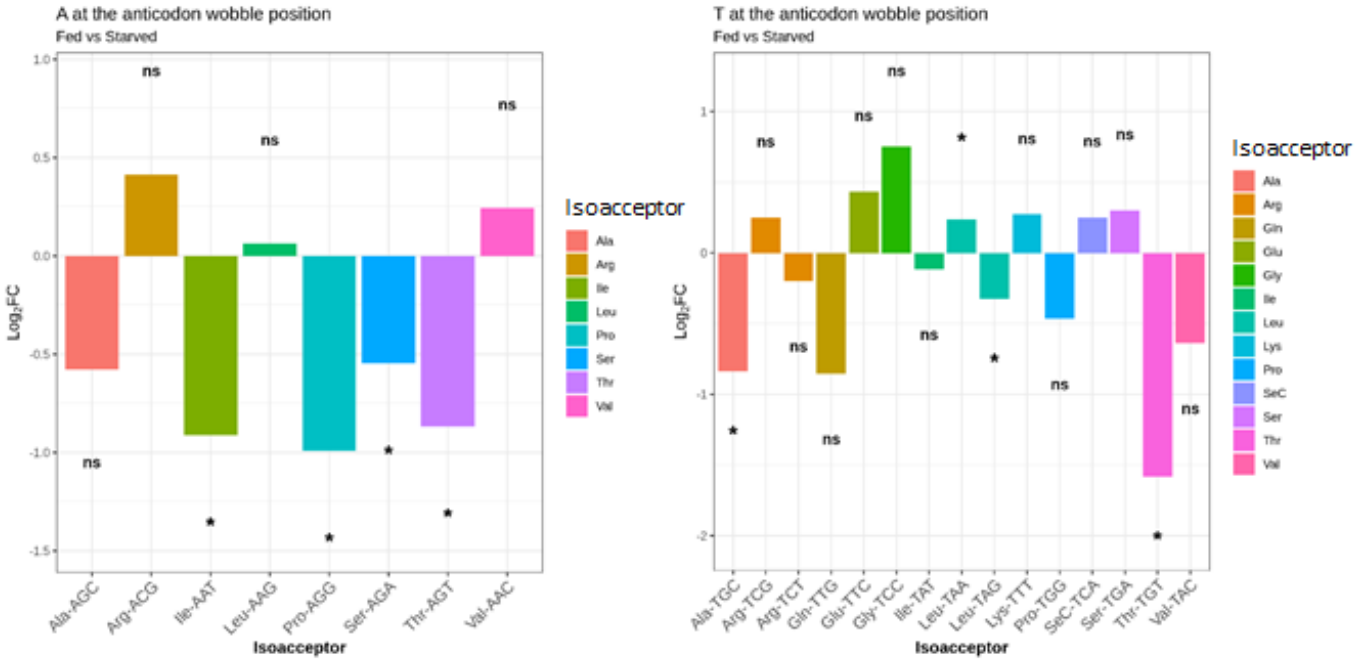
4.5.1 Global tRNA isodecoder changes

Differential expression analysis was performed on filtered tRNA reads (removing tRNAs with counts < 10 in all replicates) using DESeq2 (Love et al., 2014). A minimum count of 10 reads per replicate was used as a cut off point for all isoacceptors. Apeglm log fold change shrinkage (Zhu et al., 2018) was performed in the differential expression analysis. In Fed vs Starved, the Fed condition was set as the wild-type condition, with positive log fold changes referring to upregulation in quiescence and negative log fold changes referring to downregulation in quiescence. In Empty vs Ras, the Empty condition was set as the wild-type condition, with positive log fold changes referring to upregulation in senescence and negative log fold changes referring to downregulation in senescence.

It was previously mentioned in the literature that codons with an A/T at the 3rd position of mRNA codons were preferred for proliferation associated genes and G/C ending codons preferred in senescence associated genes (Gingold et al., 2014). In Fed vs Starved, where cells are being driven from proliferation to quiescence, differential tRNA expression matched these previous findings reported in literature. All statistically significant tRNA isoacceptors with an A or T base at the wobble position were downregulated in quiescence with the exception of tRNA-Leu-TAA (Figure 4.14 A). All statistically significant tRNA isoacceptors with a G or C base at the wobble position were upregulated in senescence, with the exception of tRNA-Ala-CGC (Figure 4.14 B). This suggests that the tRNA availability is regulated to coincide with changes at the mRNA level when cells are driven from proliferation to quiescence. However, this analysis does not make clear what is driving said regulation.

In Empty vs Ras, where cells are driven from proliferation to senescence, global changes at the isoacceptor level were less clear. Only 3 isoacceptors with an A at the wobble position were statistically significant (Figure 4.15 A) with Leu-AAG and Pro-AGG being downregulated as expected and Thr-AGT being upregulated. All statistically significant isoacceptors with a T at the wobble position were all upregulated with the exception of Pro-TGG (going against previously reported findings in literature). Only 3 isoacceptors with a G at the wobble position were statistically significant (Figure 4.15 B) with Asn-GTT and Phe-GAA being upregulated as expected and Gly-CCC being downregulated. All statistically significant isoacceptors with a C at the wobble position were all downregulated with the exception of Ser-CGA. It was clear in this experiment that tRNAs with a purine (A/G) at the wobble position followed the findings reported in literature but pyrimidines (T/C) did not, suggesting that when cells are driven from proliferation to senescence, complementary base pairing of tRNAs and mRNAs was not strictly the limiting factor in meeting the demands of the transcriptome. Furthermore, since isoacceptors that coded for the same amino acids were observed to be moving in both directions in the differential expression analysis, this led to tRNAs being broken down into their box-types and being compared directly to the observed relative synonymous codon usage of the mRNAs. This would allow for the identification of any potential non-cognate binding effects (inosine or G:T base pairing, see Table 4.1), and the role of tRNA degeneracy in decoding the transcriptome.

A



B

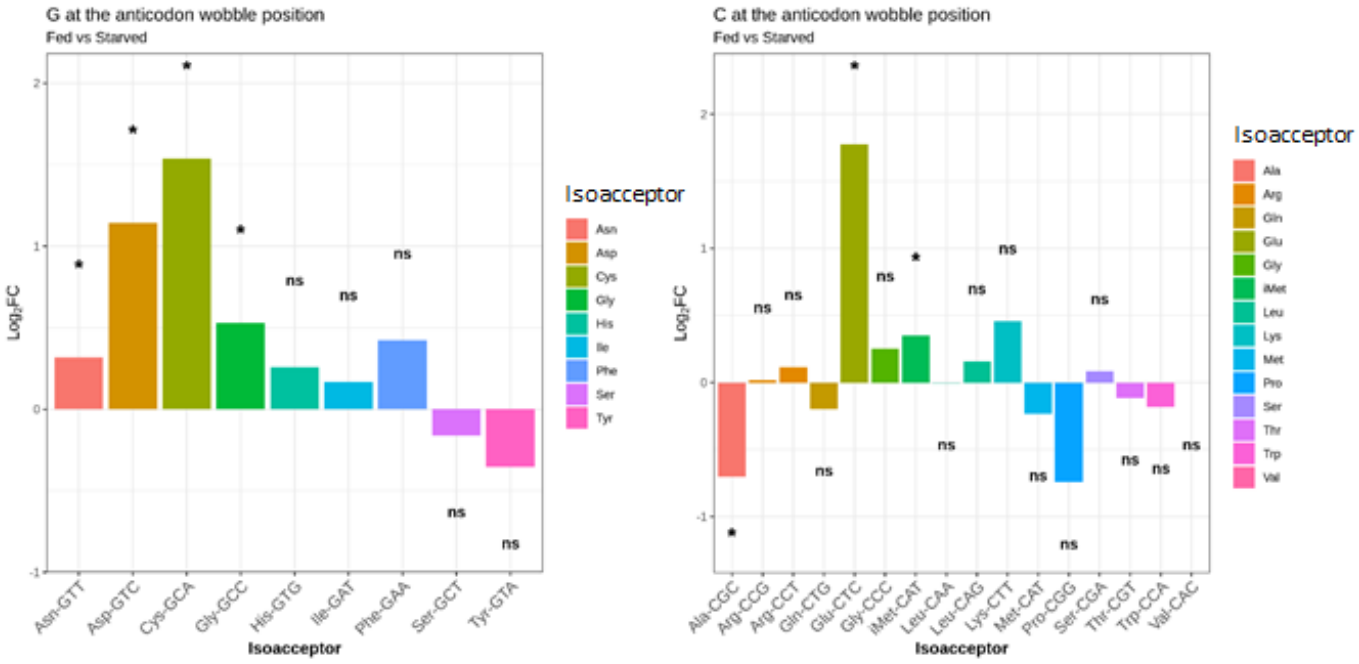
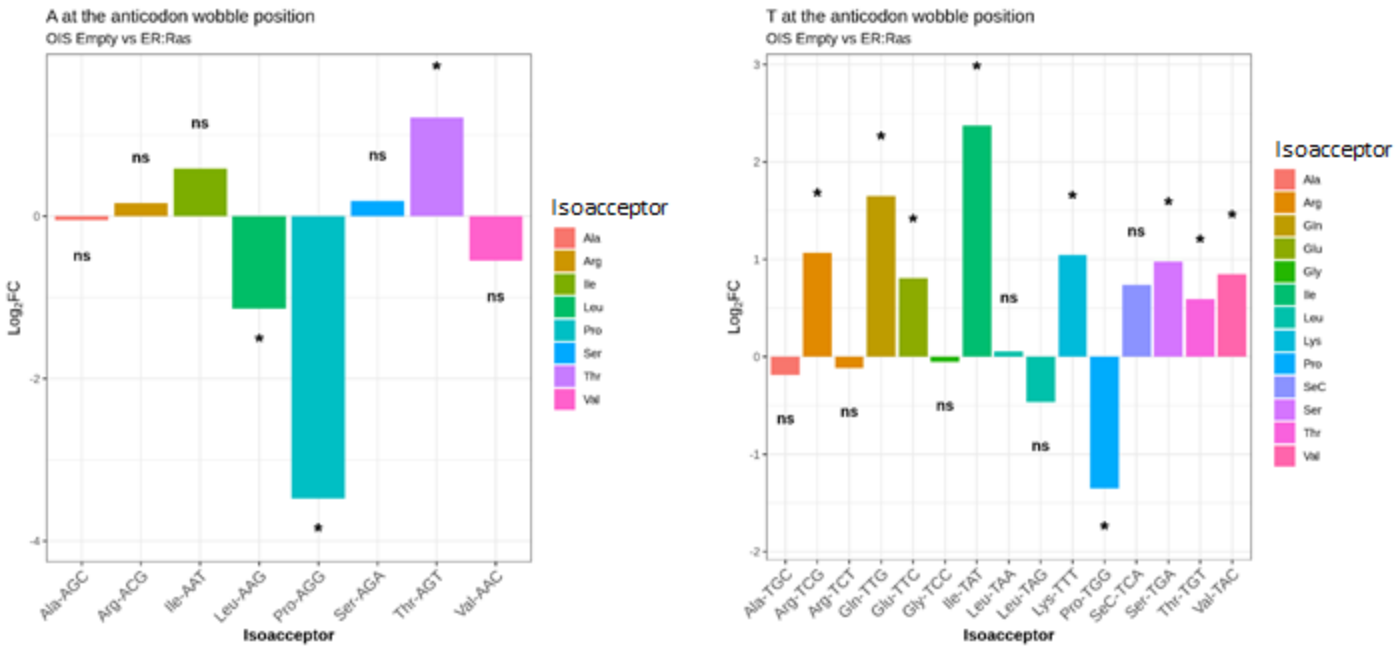


Figure 4.14 Log fold changes at the isoacceptor level match mRNA preferences reported in literature for Fed vs Starved. (A) Log fold changes for tRNA isoacceptors with an A or T at the wobble position. All of the statistically significant isoacceptors are downregulated in quiescence with the exception of tRNA-Leu-TAA. (B) Log fold changes for tRNA isoacceptors with a G or C at the wobble position. All of the statistically significant isoacceptors are upregulated in proliferation with the exception of tRNA-Ala-CGC. * $p < 0.05$. ns = not significant, Log_2(Starved/Fed).

A



B

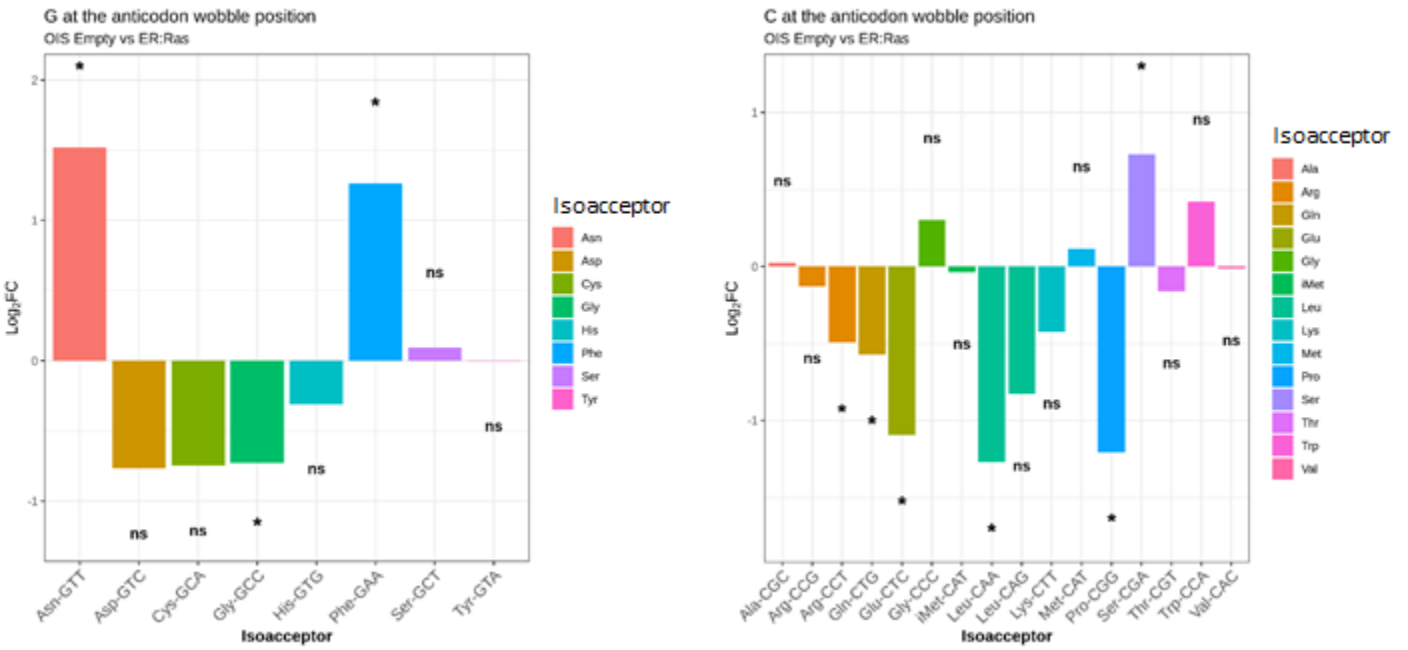


Figure 4.15 Isoacceptors with a purine base at the wobble position confer to mRNA codon preferences observed in literature. (A) Isoacceptors with an A or T at their wobble position. 3 isoacceptors with an A at the wobble position were statistically significant with Leu-AAG and Pro-AGG being downregulated as expected and Thr-AGT being upregulated. All statistically significant isoacceptors with a T at the wobble position were all upregulated with the exception of Pro-TGG. (B) Isoacceptors with an A or T at their wobble position. 3 isoacceptors with a G at the wobble position were statistically significant with Asn-GTT and Phe-GAA being upregulated as expected and Gly-CCC being downregulated. All statistically significant isoacceptors with a C at the wobble position were all downregulated with the exception of Ser-CGA. * $p < 0.05$. ns = not significant, Log_2(Ras/Empty).

4.5.2 Comparison of 6-box tRNAs to the relative synonymous codon usage in mRNA

tRNA box-types is a breakdown of tRNAs according to the number of codons they can decode (see Table 4.1). 6-box tRNAs can decode 6 mRNA codons, although not all complementary anticodons are expressed. Therefore, an inosine modification at position 34 of tRNAs or G:T base pairing is utilised to decode all codons. 4-box tRNAs follows the same reasoning and can decode 4 codons, but only 3 anticodons are expressed. There are two types of 2-box tRNAs, those that express two anticodons and those that express one anticodon. Lastly, 1-box tRNAs are those that decode 1 codon.

The tRNA differential expression analysis was broken down into their box-types and the amino acids they decode. The RNA sequencing that was done in parallel was used to determine the relative synonymous codon usage of the transcriptome. Briefly, RNA reads were mapped to the protein coding transcriptome using salmon allowing for quantification at the transcript level (Patro et al., 2017) and differentially expressed transcripts determined using DESeq2 (Love et al., 2014). Differentially expressed transcripts that were statistically significant were filtered for a ± Log_2 fold change of 2 (4 fold) and the relative synonymous codon usage (RSCU) was calculated, normalizing to transcript abundance. RSCU was plotted for each amino acid to highlight preference according to cell condition.

6-box isoacceptors include tRNA-Arg, tRNA-Leu and tRNA-Ser (Figure 4.16 A-C and Figure 4.17 A-C). All three isoacceptors have an A34I modification (for Arg-ACG, Leu-AAG and Ser-AGA). Also, tRNA-Ser-GCT allows for G:T base pairing with codon AGT.

In Fed vs Starved, tRNA-Leu- TTA (which decodes codon TTA) was the only isoacceptor with an A at the wobble position to be upregulated in quiescence, even though the transcriptomic RSCU shows no preference for it in quiescence compared to proliferation

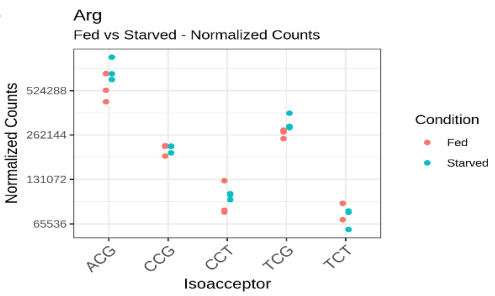
(Figure 4.16 B). tRNA-Leu-TAG (which decodes codon CTA) on the other hand is down regulated in quiescence and again, the transcriptomic RSCU shows no preference for it in quiescence over proliferation.

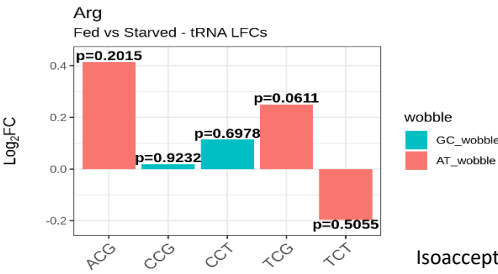
In Empty vs Ras, AGA is the most preferred codon from the RSCU data in both conditions and is preferred in proliferation but tRNA-Arg-TCT shows no statically significant change (Figure 4.17 A). tRNA-Leu-CAA is downregulated in senescence but the codon CTT is slightly preferred in proliferation which would explain why it does not follow G/C anticodons being upregulated (Figure 4.17 B). tRNA-Ser-TGA is upregulated in senescence but the codon TCA is slightly preferred in proliferation. However, the ICT modification would decode this codon in both conditions (Figure 4.17 C).

tRNA degeneracy could explain why certain isoacceptor log fold changes do not match up to preferences at the RSCU level, especially in the Empty vs Ras experiment. Since the A34I modification can decode codons ending with an A/C/T (Table 4.1), this degeneracy in tRNAs could be aiding in the decoding of mRNA codons in both experiments, even though the logfold changes of cognate tRNA isoacceptors are not statistically significant or are changing in the opposite direction to the RSCU preference.

6-box tRNAs – Fed vs Starved

A

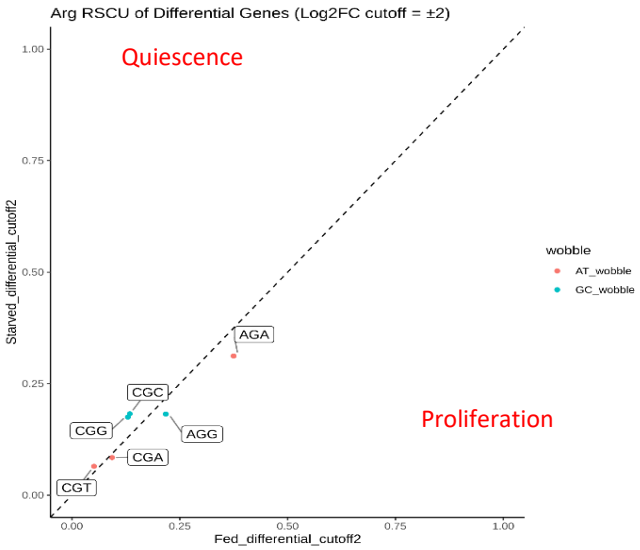




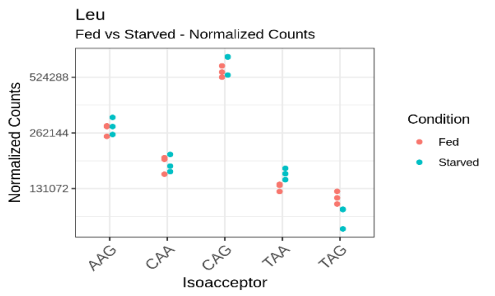
CGT/ CGC CGG AGG CGA AGA

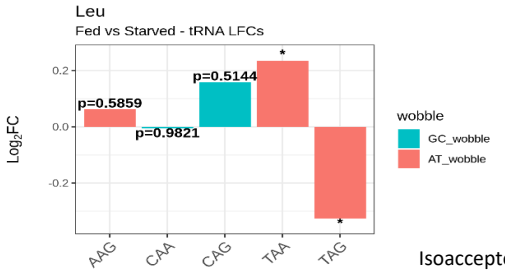
Isoacceptor

mRNA codon



B

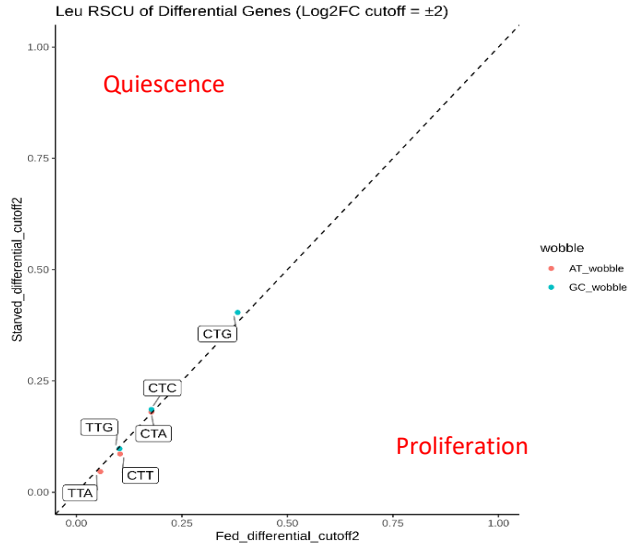




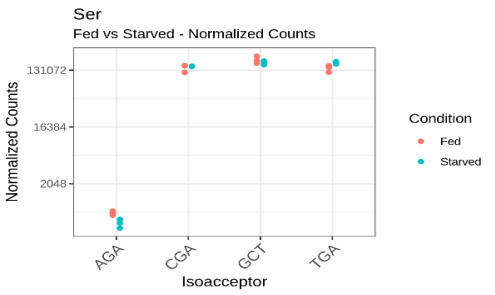
CTT/ CTC TTG CTG TTA CTA

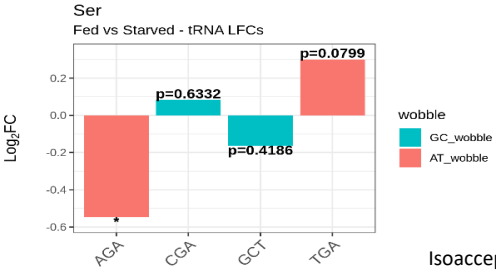
Isoacceptor

mRNA codon



C





TCT/ TCC TCG AGC/ AGT TCA

Isoacceptor

mRNA codon

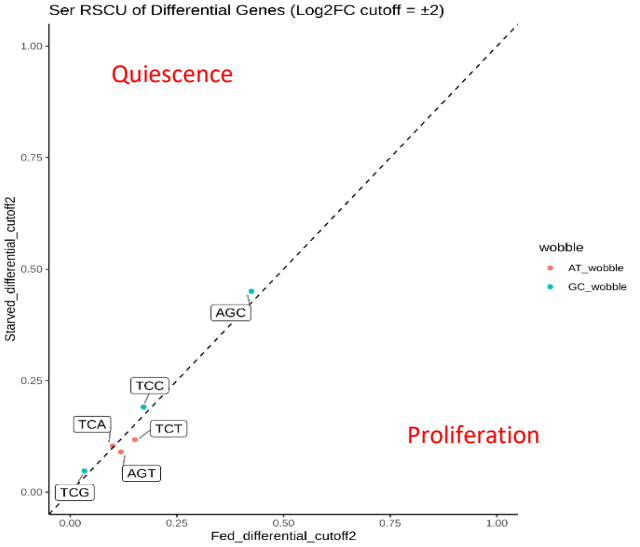


Figure 4.16 6-box tRNA differential expression analysis and relative synonymous codon usage in Fed vs Starved. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Arginine, (B) Leucine and (C) Serine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

6-box tRNAs – Empty vs Ras

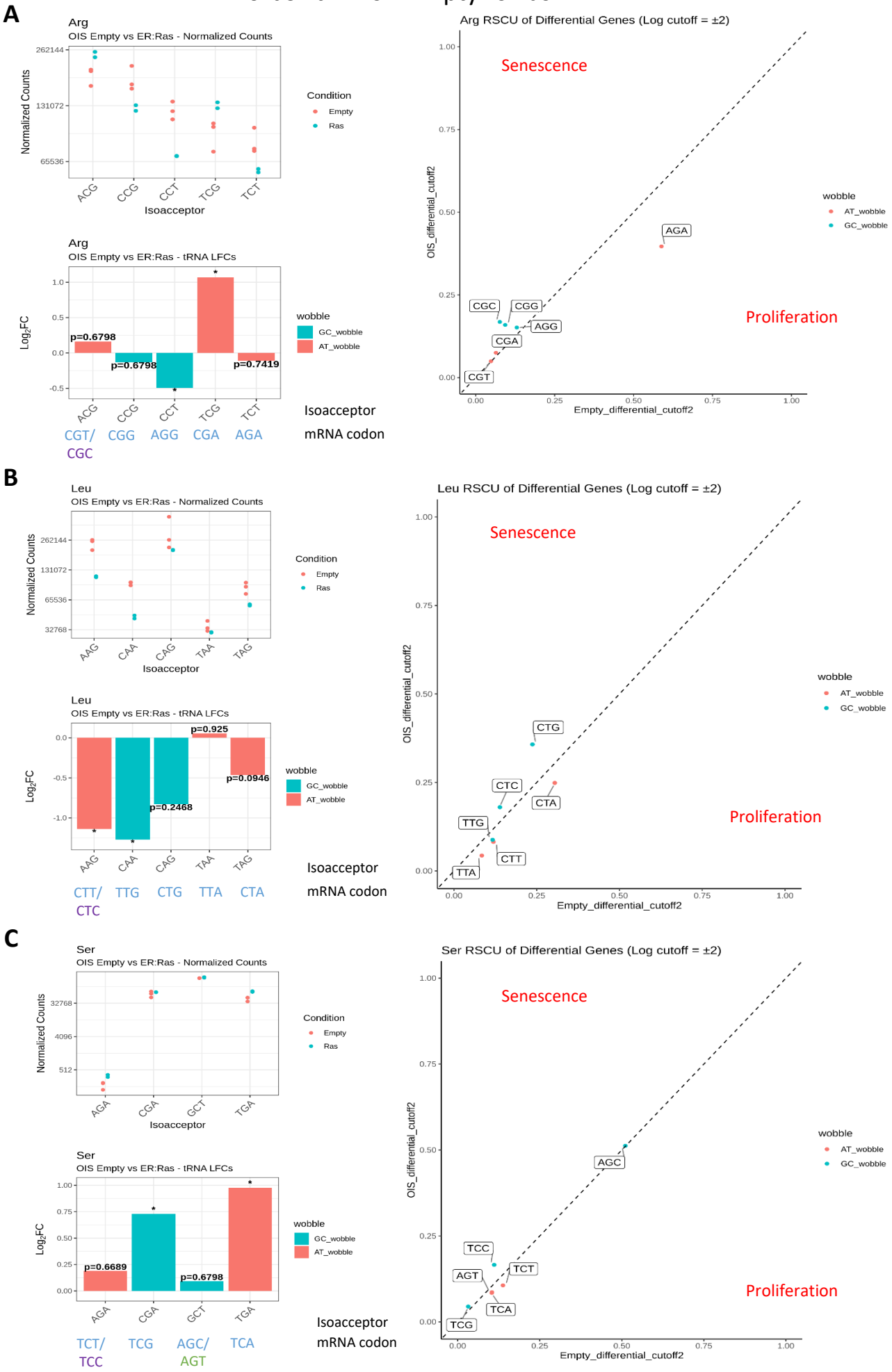


Figure 4.17 6-box tRNA differential expression and relative synonymous codon usage analysis in IMR90 Empty vs Ras. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Arginine, (B) Leucine and (C) Serine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

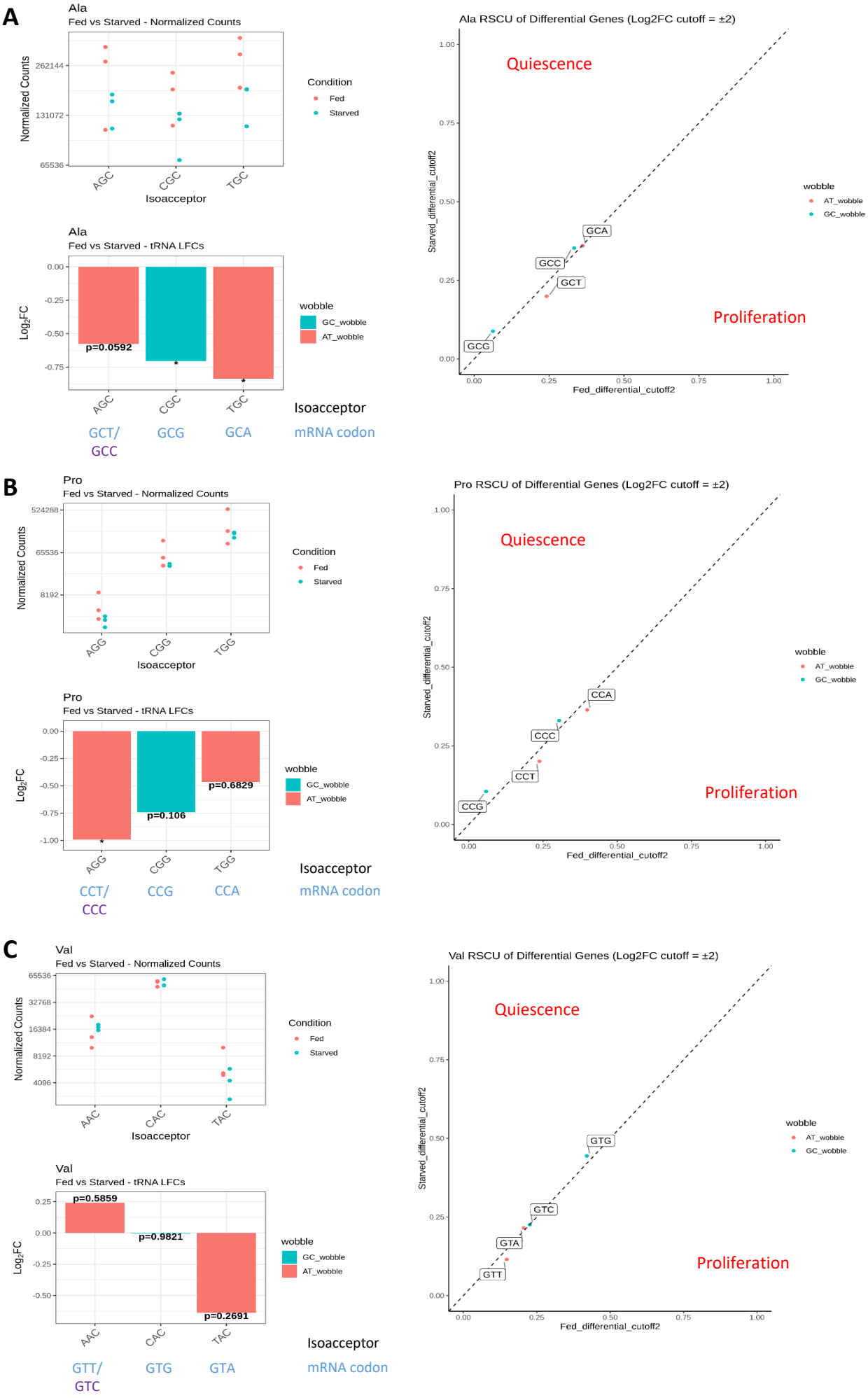
4.5.3 Comparison of 4-box tRNAs to the relative synonymous codon usage in mRNA

The same phenomenon of tRNA degeneracy seen in the 6-box tRNAs is also true for the 4-box tRNAs (Figure 4.18 A-E and 4.19 A-E). 4-box isoacceptors include tRNA-Ala, tRNA-Pro, tRNA-Val, tRNA-Gly and tRNA-Thr.

In Fed vs Starved, tRNA-Ala-CGC (which decodes the GCG codon) was the only isoacceptor to be downregulated in quiescence. The GCG codon is the least preferred synonymous codon and is preferred more in quiescence than in proliferation. However, the GCG codon can also be decoded by IGC (due to A34I) and so could accommodate its decoding in quiescence (Figure 4.18A).

In Empty vs Ras, tRNA-Pro-CGG (which decodes CCG) is downregulated in senescence but the CGC codon is preferred in senescence (Figure 4.19 B). However, the CCG codon is the least preferred synonymous codon. tRNA-Val-TAC (which decodes GTA) is upregulated in senescence even though the GTA codon is preferred in proliferation (Figure 4.19 C). However, the GTA codon can be decoded by IAC, resulting in tRNA degeneracy aiding in meeting transcriptomic demands.

4-box tRNAs – Fed vs Starved



4-box tRNAs – Fed vs Starved

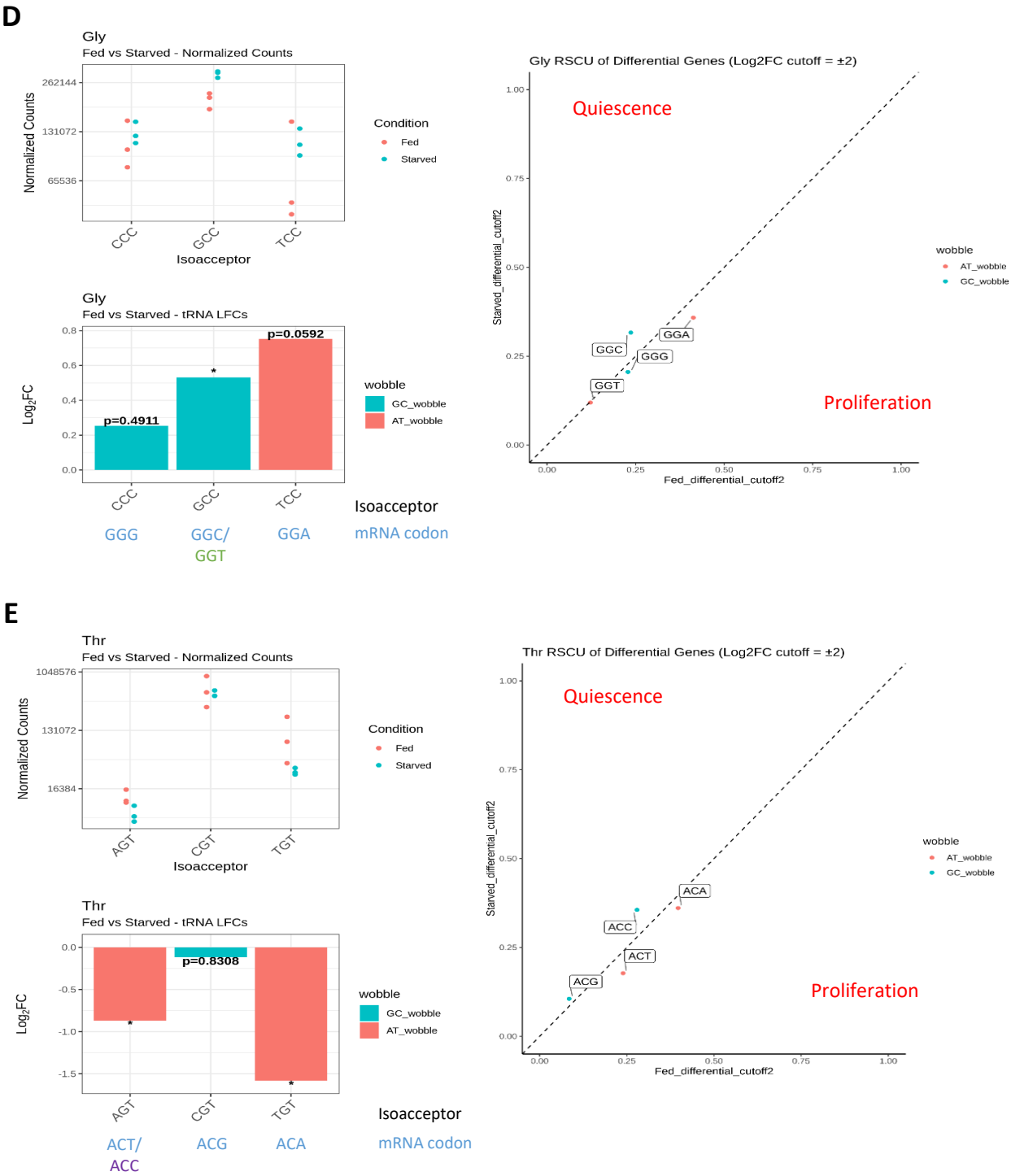
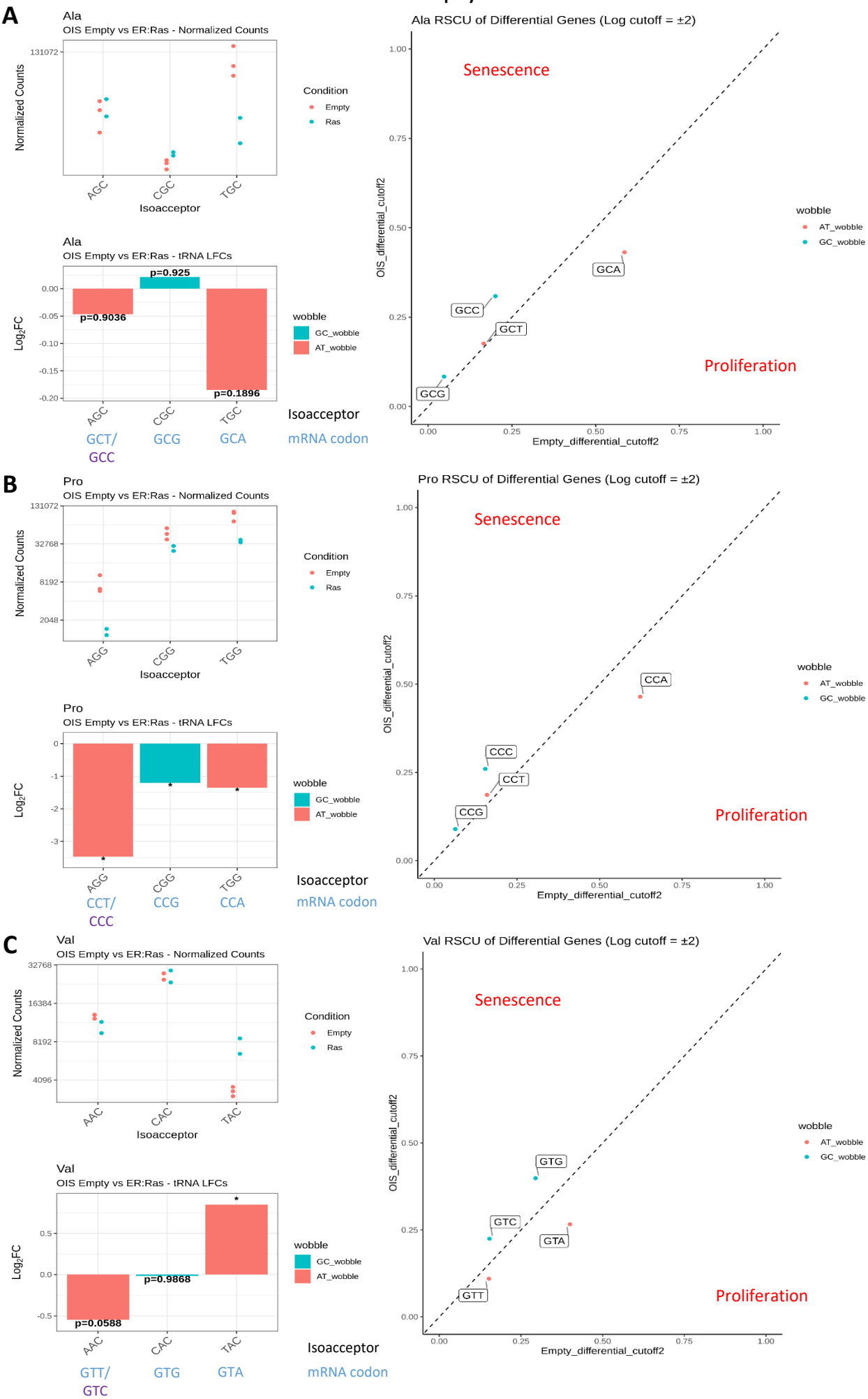


Figure 4.18 4-box tRNA differential expression analysis and relative synonymous codon usage in Fed vs Starved. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Alanine, (B) Proline, (C) Valine, (D) Glycine and (E) Threonine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

4-box tRNAs – Empty vs Ras



4-box tRNAs – Empty vs Ras

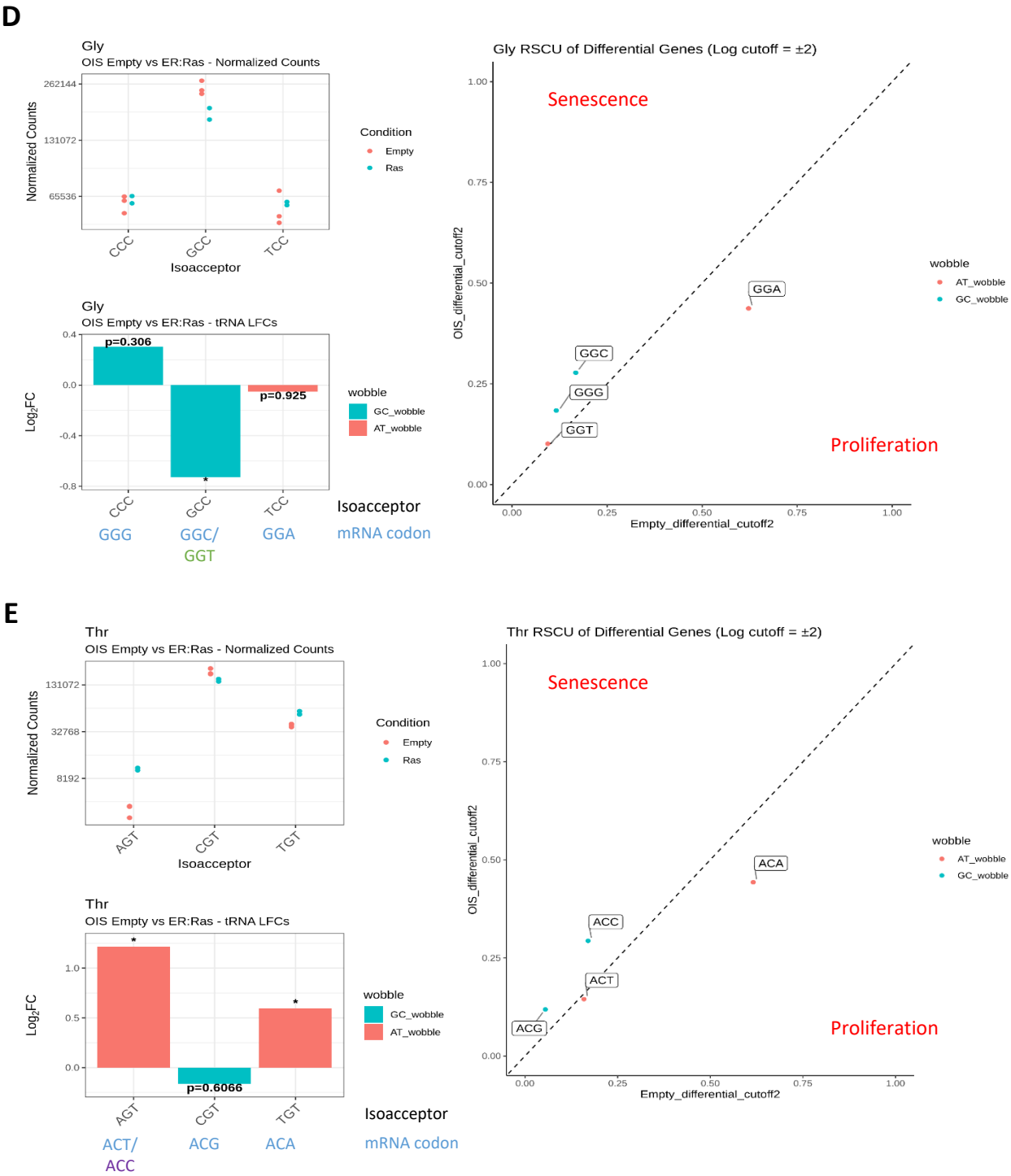


Figure 4.19 4-box tRNA differential expression analysis and relative synonymous codon usage in IMR90 Empty vs Ras. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Alanine, (B) Proline, (C) Valine, (D) Glycine and (E) Threonine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

4.5.4 Comparison of 2-box tRNAs to the relative synonymous codon usage in mRNA

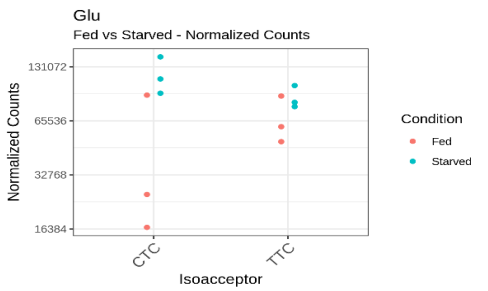
2-box isoacceptors include tRNA-Glu, tRNA-Gln, tRNA-Lys and tRNA-Ile.

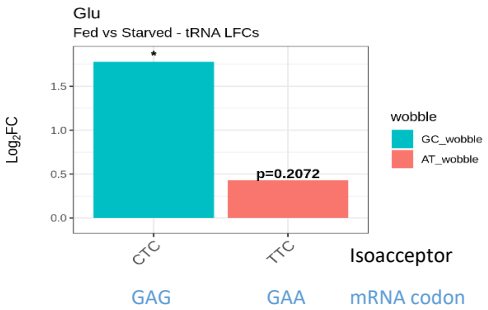
In Fed vs Starved, all the isoacceptors that were differentially expressed and were statistically significant, matched up to the RSCU preferences (Figure 4.20 A-D). Interestingly, tRNA-Ile-GAT showed a very low read count, which could be due to modifications present on the Watson-crick face (as isoleucine is a heavily modified tRNA), preventing it from being sequences as efficiently as the other isoleucine isoacceptors.

In Empty vs Ras, all statistically significant isoacceptors with an A/T at the wobble position were upregulated in senescence even though codons with an A/T at the wobble position were preferred in proliferation. The opposite effect was observed for isoacceptors and codons with a G/C at the wobble position (Figure 4.21 A-C). tRNA-Glu and tRNA-Lys are generally the highest expressed tRNAs from the tRNA-seq data and so its high read count could suffice in meeting the transcriptomic demand, but this would require further investigation. tRNA-Ile-GAT was not identified in the tRNA-sequencing at all (Figure 4.21 D). Due to the low read count observed in Fed vs Starved, it is likely that this isoacceptor was filtered out during analysis.

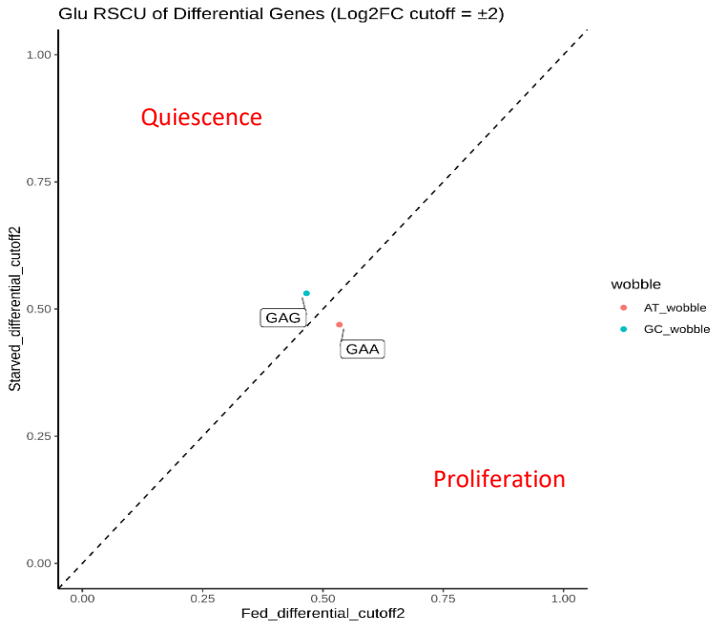
2-box tRNAs – Fed vs Starved

A

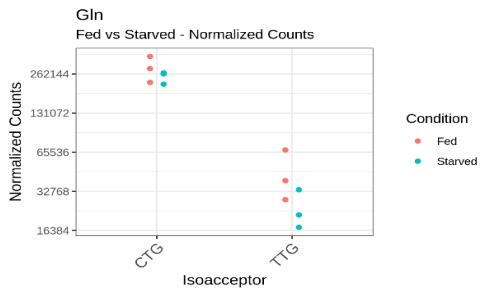


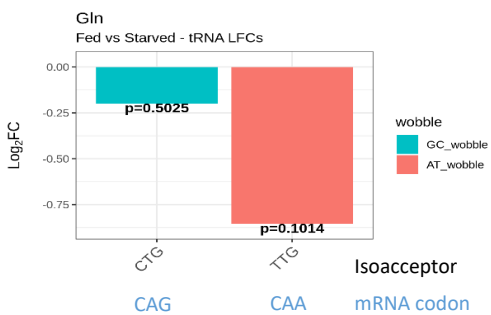


GAG GAA mRNA codon

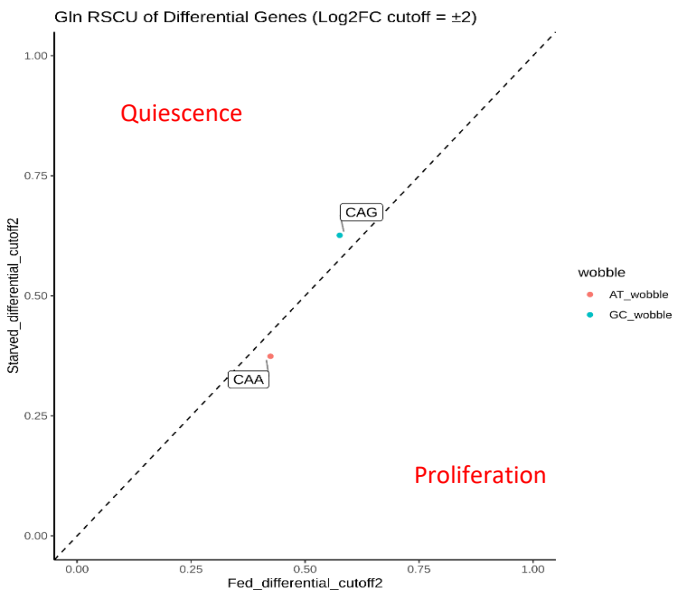


B

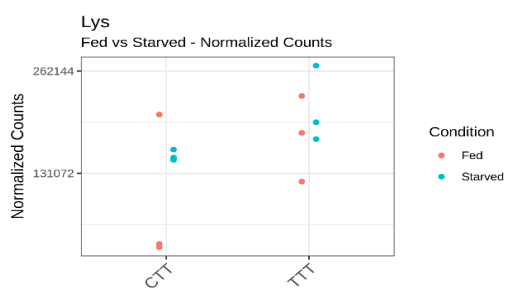


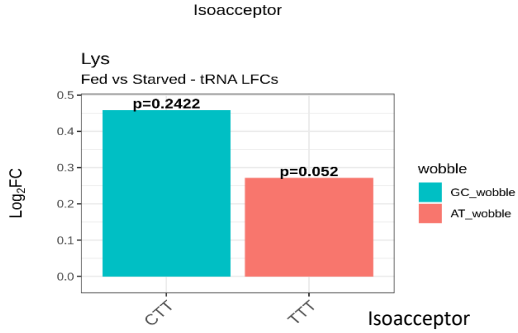


CAG CAA mRNA codon

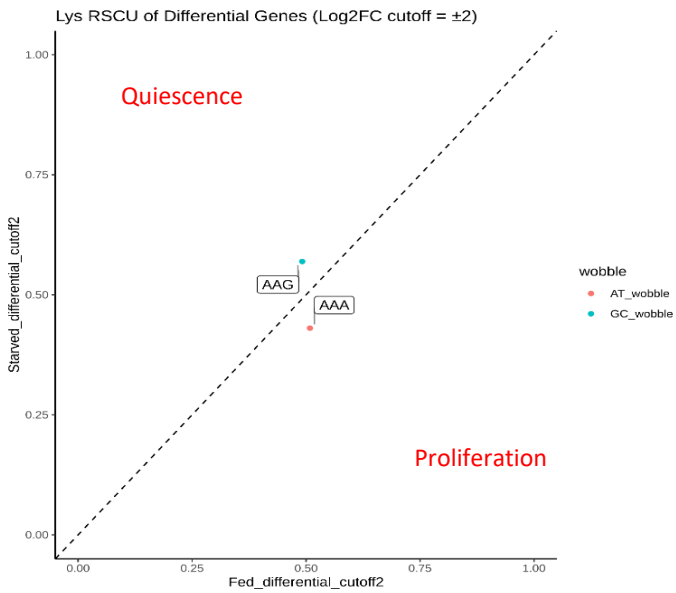


C





AAG AAA mRNA codon



2-box tRNAs – Fed vs Starved

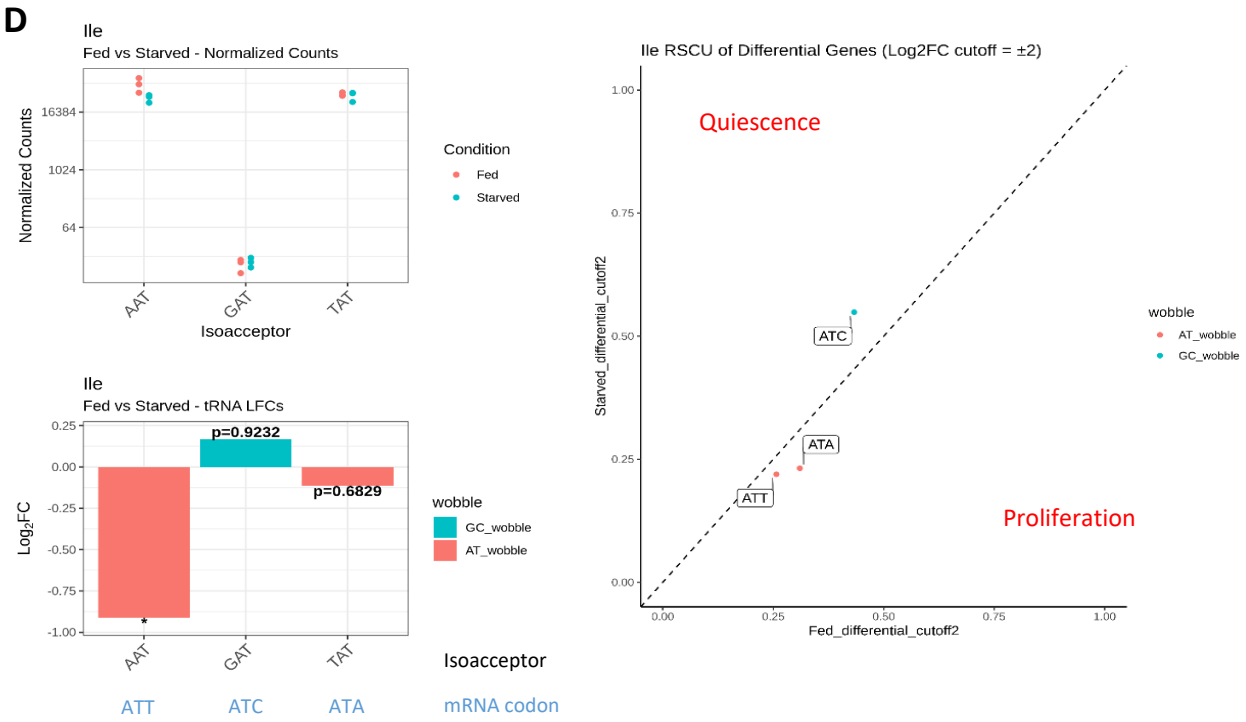
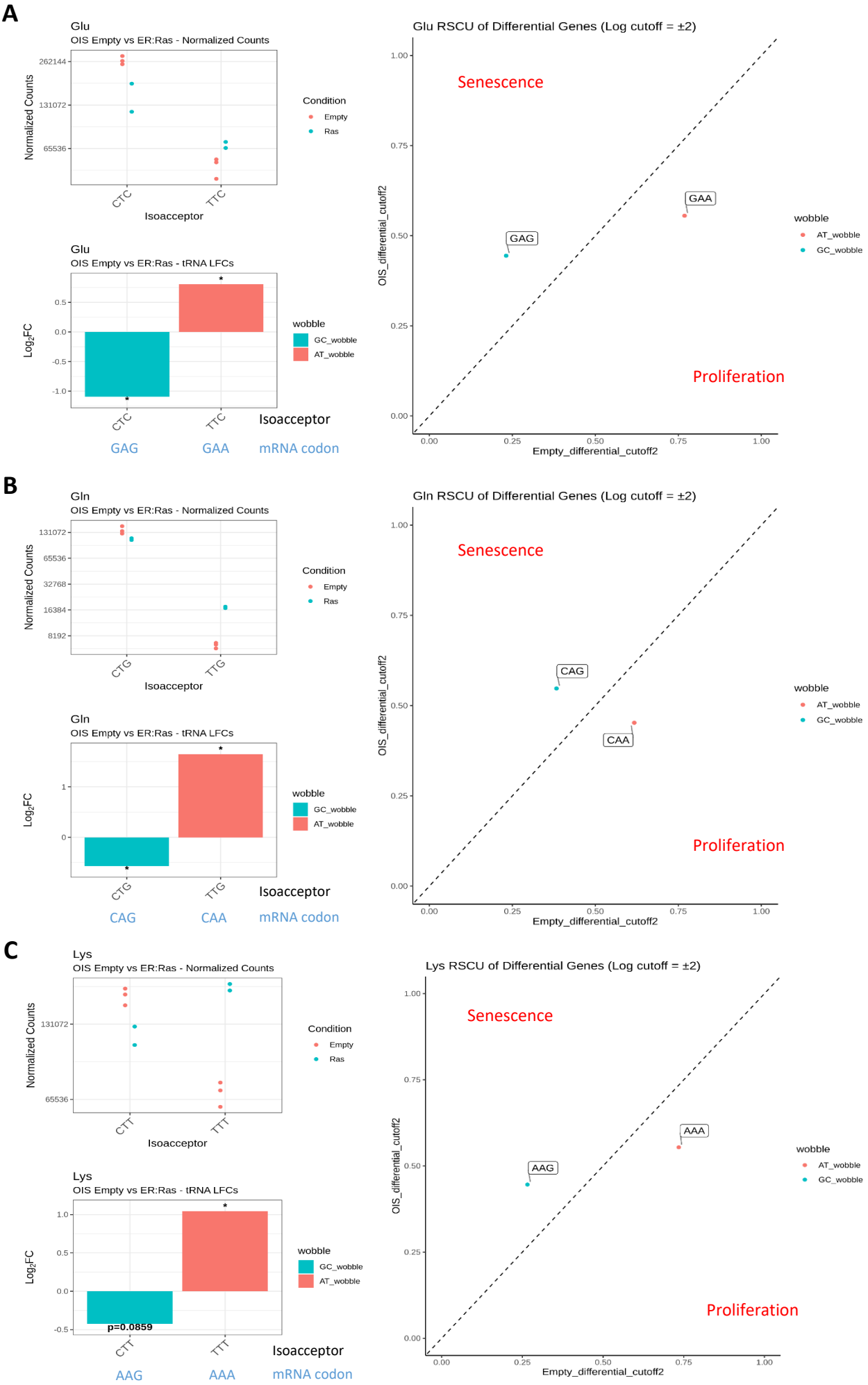


Figure 4.20 2-box tRNA differential expression analysis and relative synonymous codon usage in Fed vs Starved. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Glutamic Acid, (B) Glutamine, (C) Lysine and (D) Isoleucine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

2-box tRNAs – IMR90 Empty vs Ras



2-box tRNAs – IMR90 Empty vs Ras

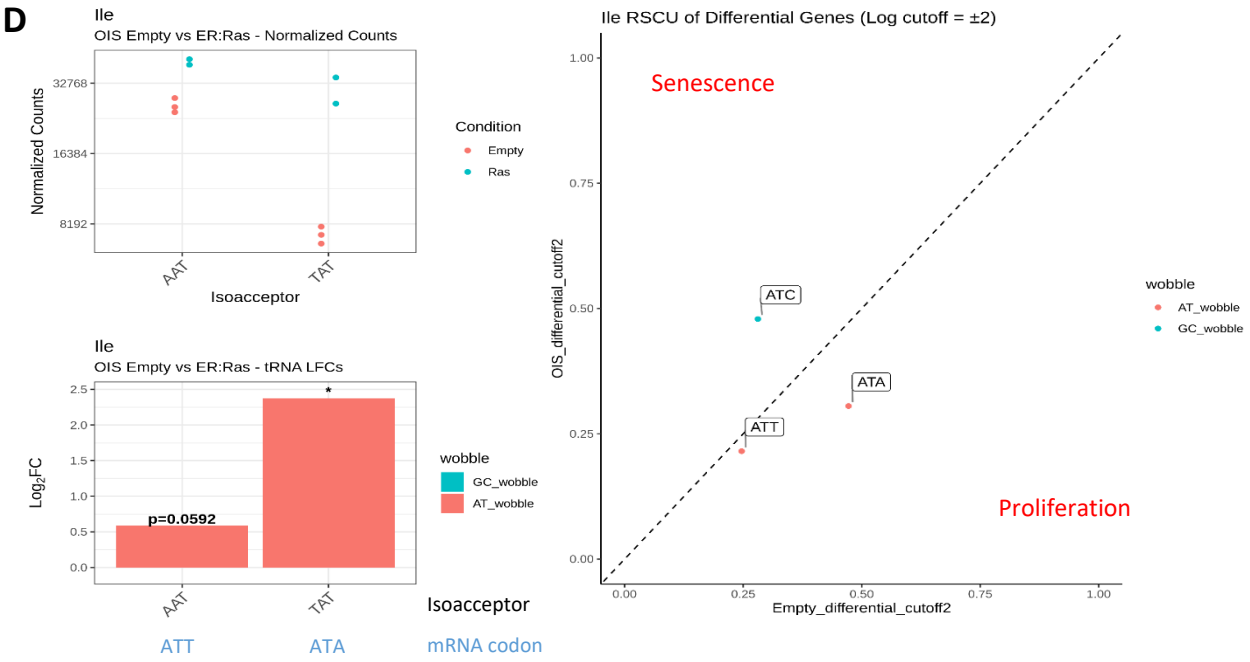


Figure 4.21 2-box tRNA differential expression analysis and relative synonymous codon usage in IMR90 Empty vs Ras. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Glutamic Acid, (B) Glutamine, (C) Lysine and (D) Isoleucine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

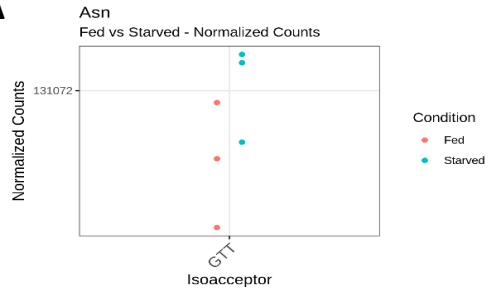
4.5.5 Comparison of 2-box tRNAs (only 1 tRNA expressed) to the relative synonymous codon usage in mRNA

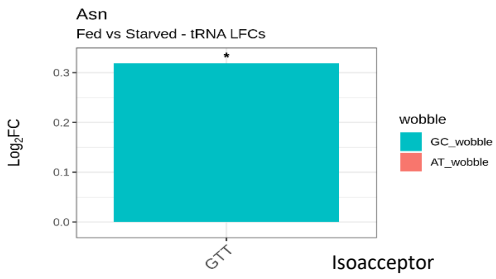
2-box isoacceptors that express only one anticodon are tRNA-Asn, tRNA-His, tRNA-Phe, tRNA-Asp, tRNA-Cys and tRNA-Tyr. All of these isoacceptors utilise G:T base pairing to decode the non-cognate codon.

In Fed vs Starved, all of the 2-box tRNAs that express one anticodon have a G/C at the anticodon are all upregulated in quiescence, where the preferred codon has a G/C codon at its wobble position (Figure 4.22 A-E). tRNA-Tyr-GTA was the only exception but was not statistically significant (Figure 4.22 F). The same was observed in Empty vs Ras (Figure 4.23 A-F).

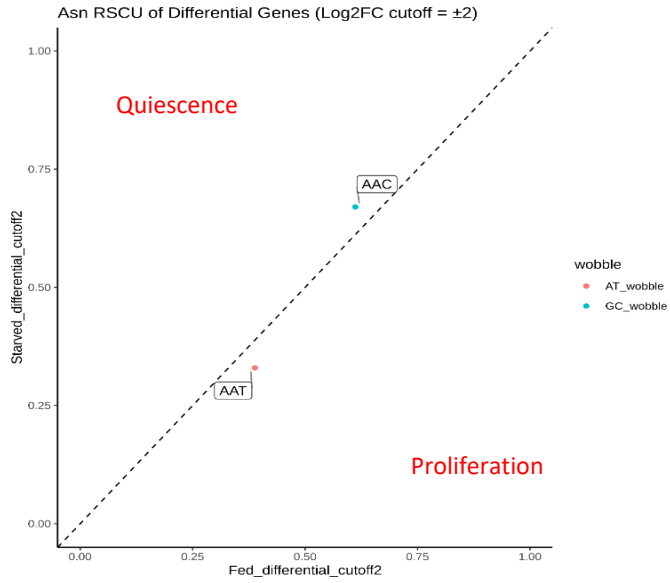
2-box tRNAs (only 1 tRNA expressed) – Fed vs Starved

A

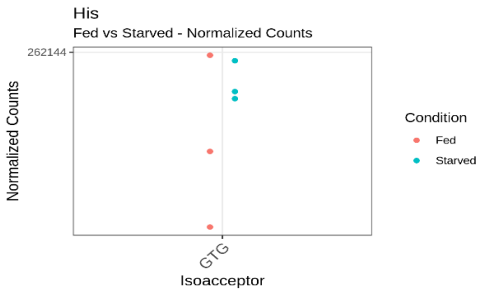


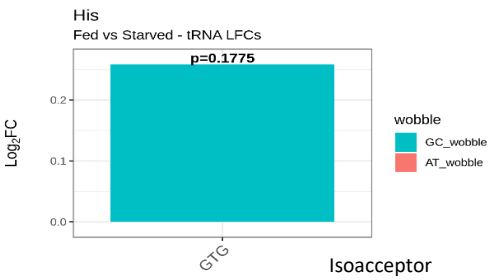


AAC/AAT mRNA codon

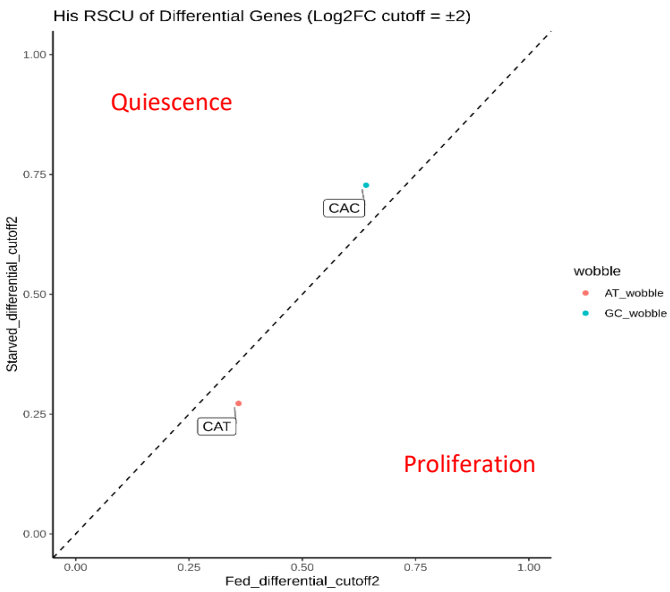


B

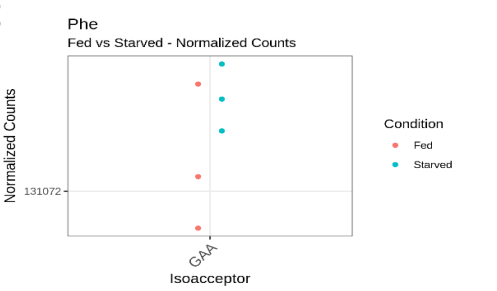


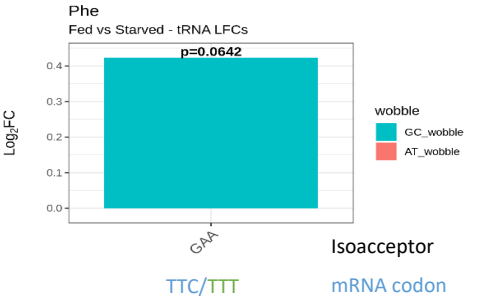


CAC/CAT mRNA codon

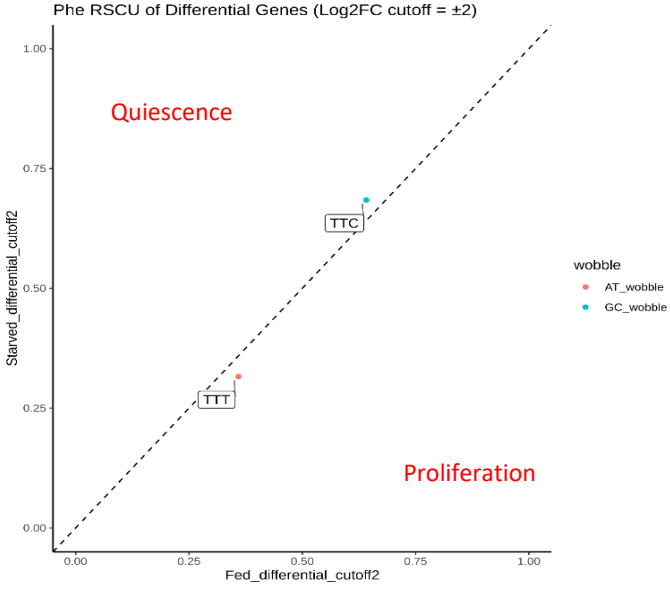


C



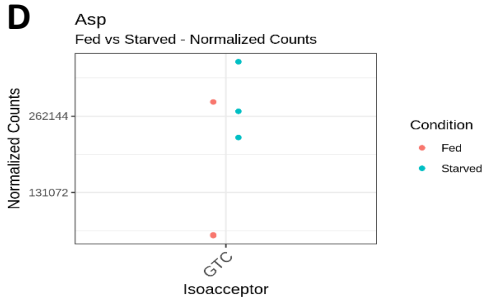


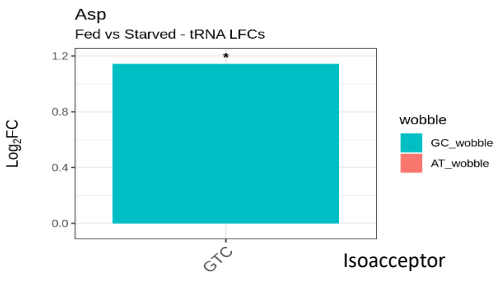
TTC/TTT mRNA codon



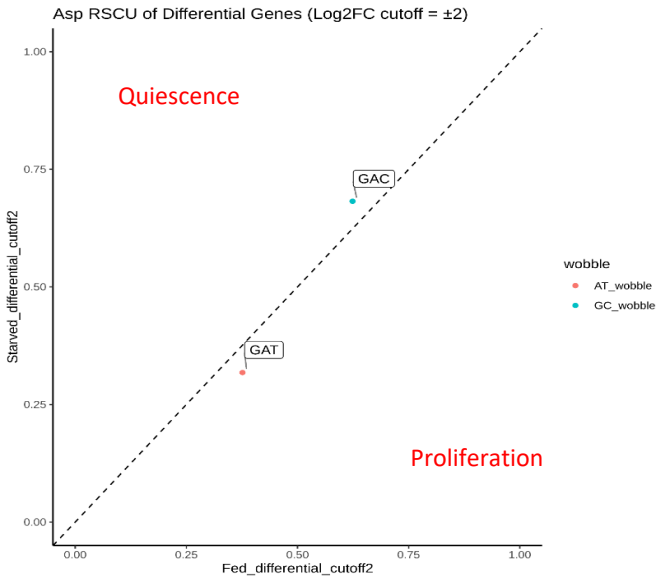
2-box tRNAs (only 1 tRNA expressed) – Fed vs Starved

D

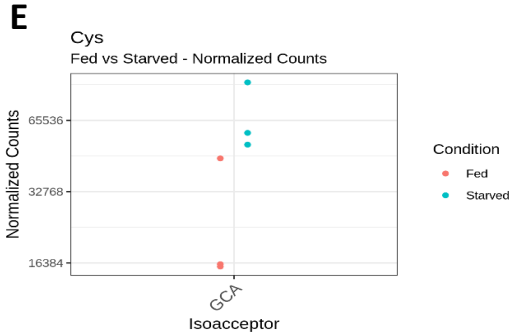


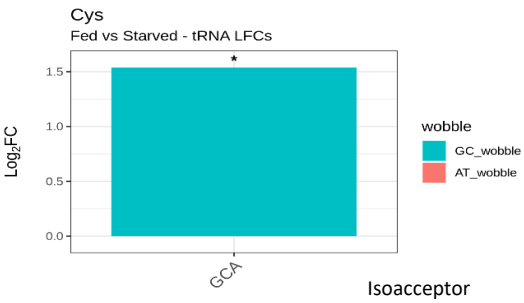


GAC/GAT mRNA codon

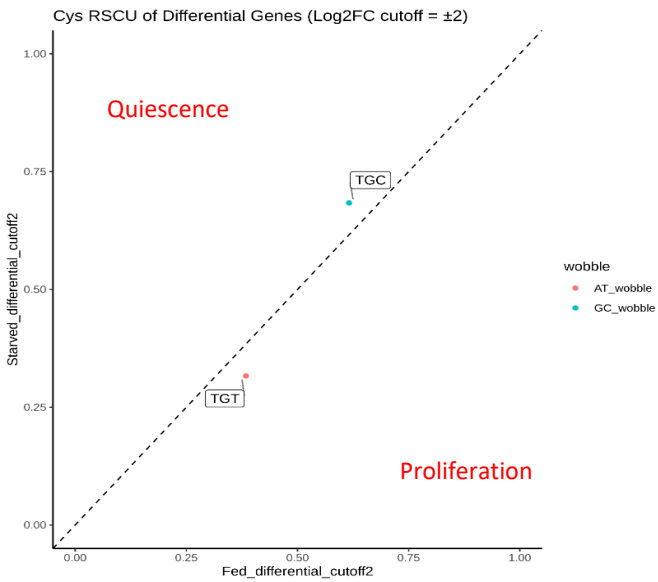


E

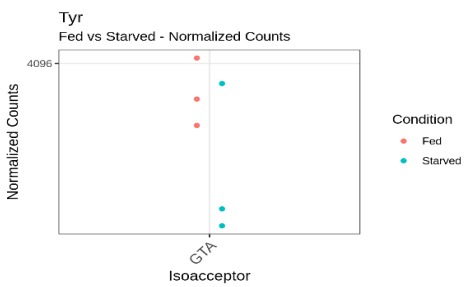


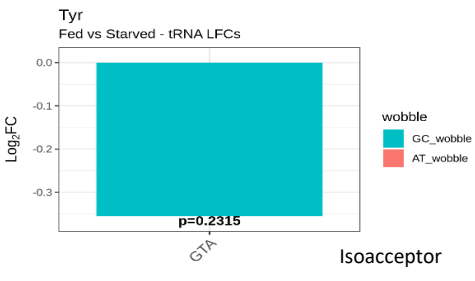


TGC/TGT mRNA codon



F





TAC/TAT mRNA codon

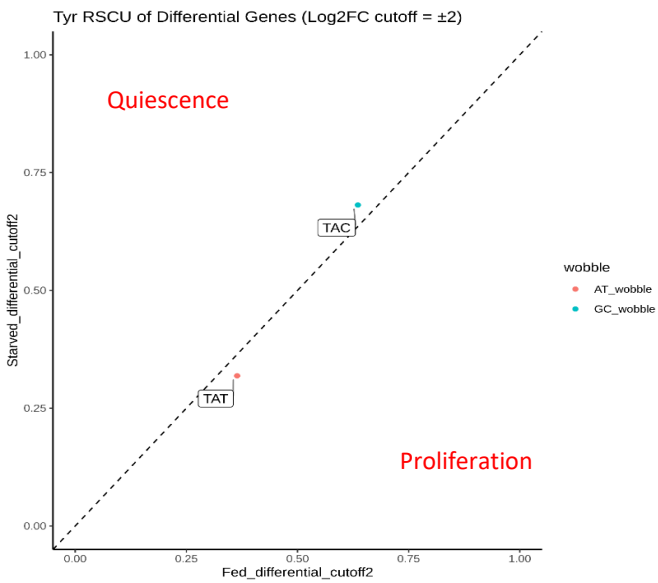
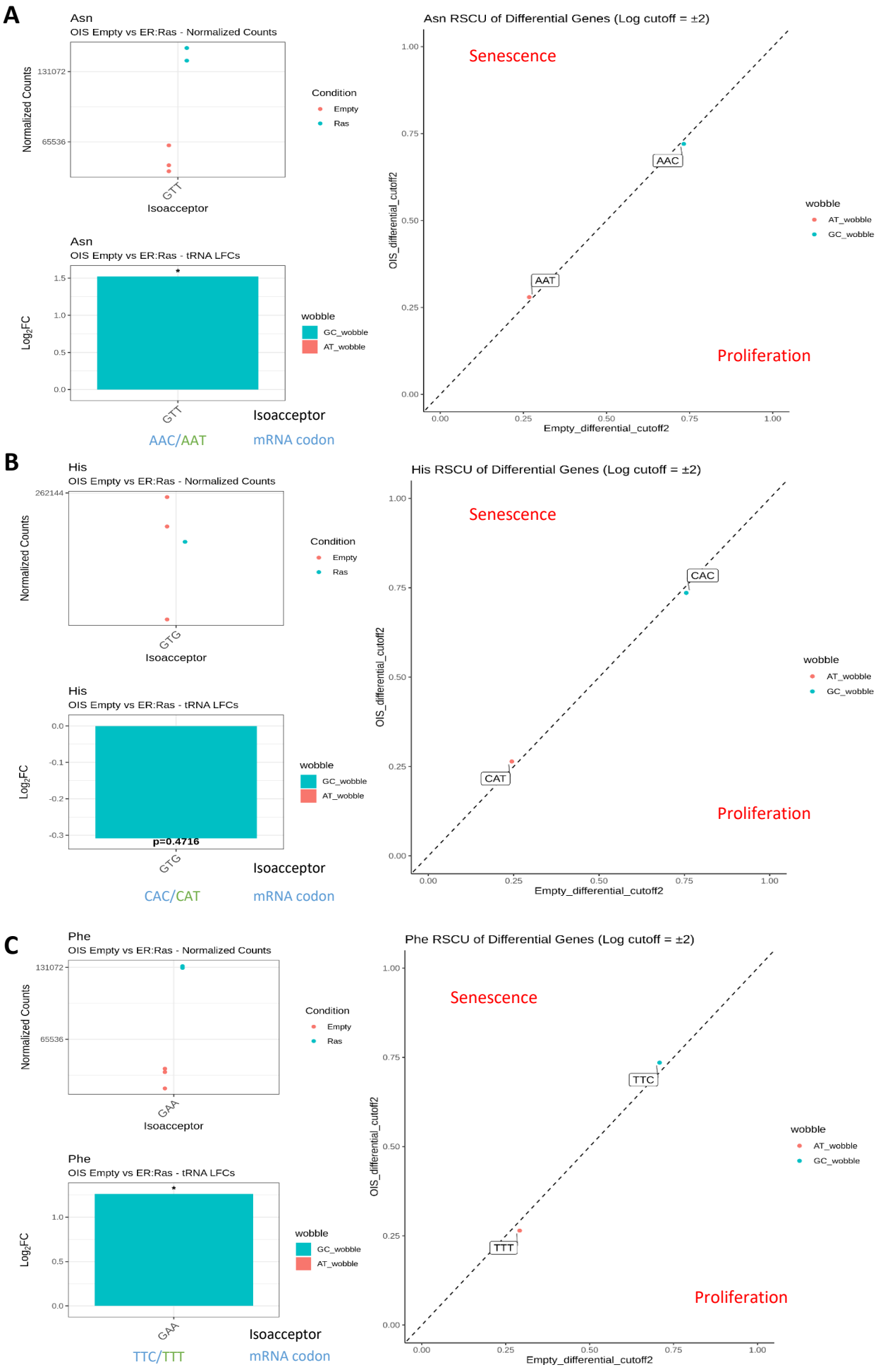


Figure 4.22 2-box tRNA (only 1 tRNA expressed) differential expression analysis in Fed vs Starved. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Asparagine, (B) Histidine, (C) Phenylalanine, (D) Aspartate, (E) Cysteine and (F) Tyrosine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

2-box tRNAs (only 1 tRNA expressed) – IMR90 Empty vs Ras



2-box tRNAs (only 1 tRNA expressed) – IMR90 Empty vs Ras

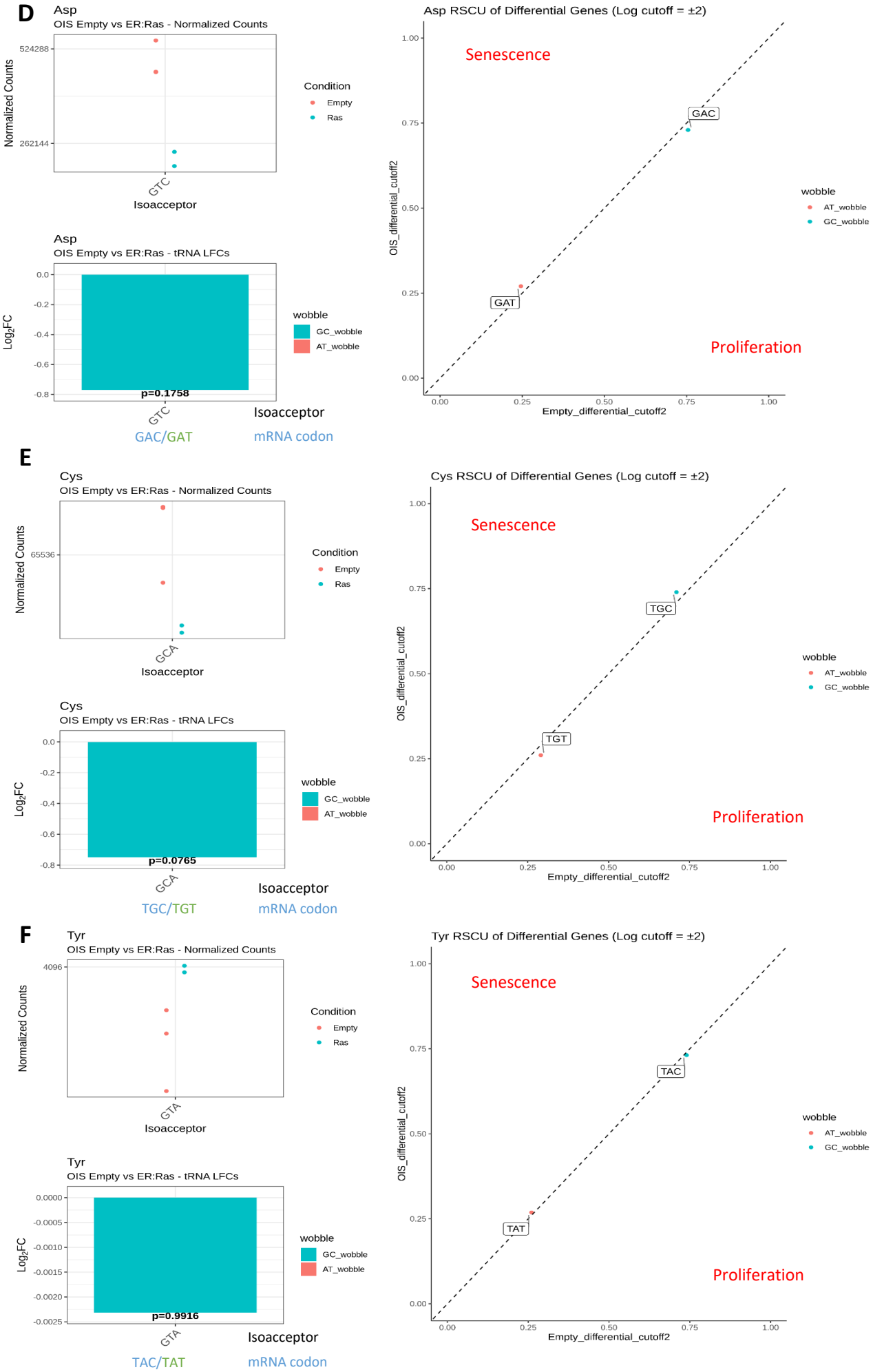


Figure 4.23 2-box tRNA (only 1 tRNA expressed) differential expression analysis in IMR90 Empty vs Ras. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Asparagine, (B) Histidine, (C) Phenylalanine, (D) Aspartate, (E) Cysteine and (F) Tyrosine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

4.6 Global relative synonymous codon usage analysis

The relative synonymous codon usage for all statistically significant ($p < 0.05$), differentially expressed genes with a $\log_2$ fold change of ±2 were calculated (Figure 4.24 A-C). This was done to highlight the codon preference at the wobble position for the transcriptome, depending on the cell state.

In Fed vs Starved, there was a preference for A/T codons at the wobble position of the mRNAs in the Fed condition (proliferation) and G/C codons at the wobble position of the mRNAs in the Starved condition (quiescence) (Figure 4.24 A). This matched up with the tRNA isoacceptor profiles observed in the tRNA sequencing.

In Empty vs Starved, the split between A/T and G/C ending codons was observed (Figure 4.24 B). However, when the codons were split into their single nucleotides at the wobble position, it was clear that A ending codons were the preferred codon in proliferation (empty condition) (Figure 4.24 C). This suggests that it could be the A-ending codons driving synonymous codon usage changes in oncogene-induced senescence. Interestingly, T codons at the wobble position of the mRNAs generally stayed neutral with negligible preferences in both directions. This would suggest the importance of tRNA degeneracy in mRNA decoding since T codons on mRNAs can be decoded either via the A34I modification or G:T base pairing (see Table 4.1). G/C codons at the wobble position of the mRNAs were preferred in senescence, as confirmed in the Fed vs Starved.

Interestingly, when the relative synonymous codon usage was calculated for all transcripts and the most abundant transcript per gene, there was no preference in Fed vs Starved (Figure 4.25 A) and negligible preference in Empty vs Ras (Figure 4.25 B). When these calculations were made, the relative synonymous codon usage for each transcript was

weighted to its abundance to prevent bias. Since these codon preferences are seen in differential genes, this would suggest a regulatory mechanism that drives synonymous codon usage, depending on the cell state. However, it is still unclear whether this preference is driven by the tRNA pool or if the tRNA pool adapts to changes in relative synonymous codon usage.

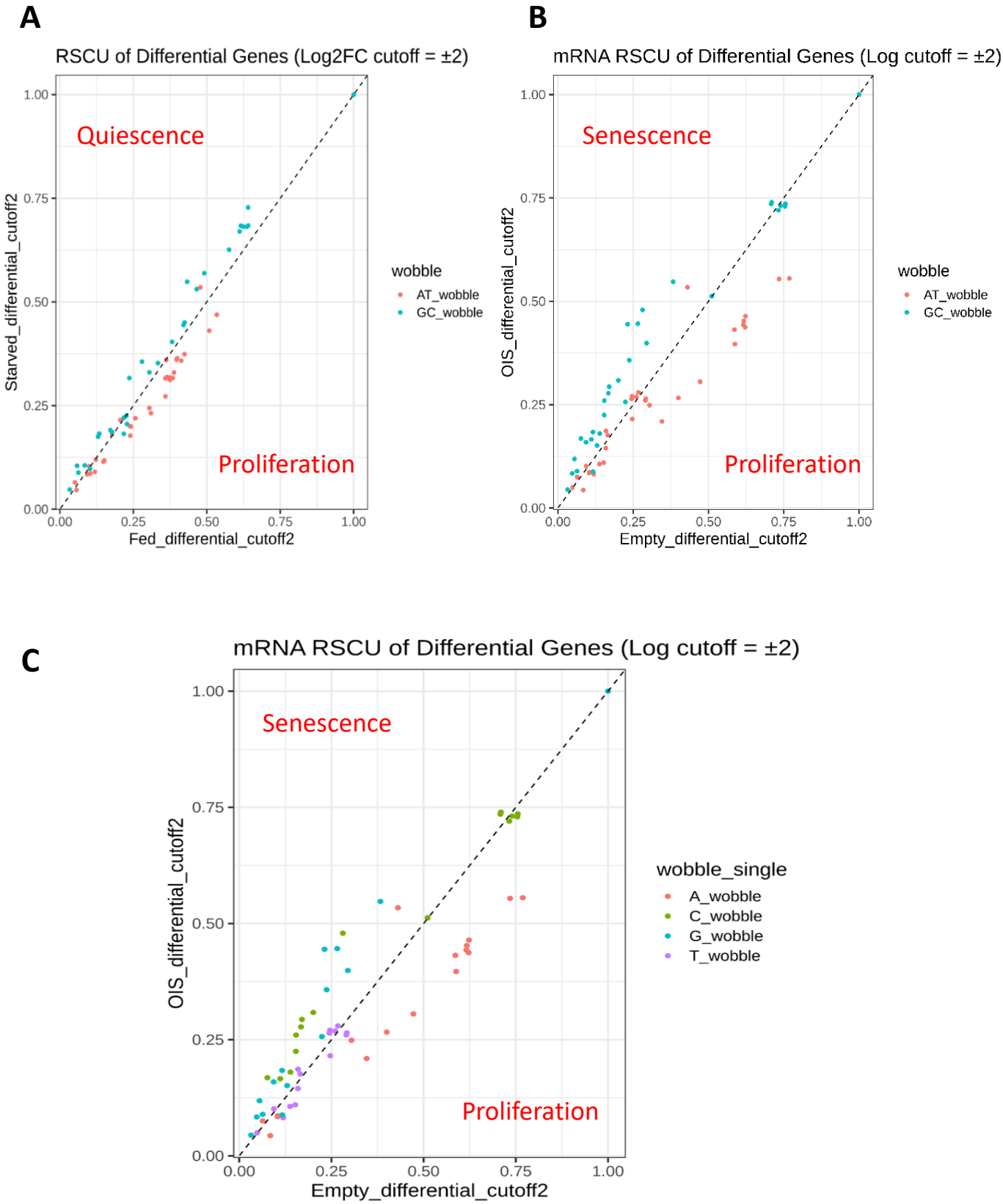


Figure 4.24 Relative synonymous codon usage for statistically significant transcripts with a ±Log_2 fold change of 2. (A) Fed vs Starved shows a preference for A/T in proliferation and G/C in quiescence at the mRNA wobble position. (B) Empty vs Ras oncogene-induced senescence shows a preference for G/C codons in senescence. (C) Empty vs Ras oncogene-induced senescence when split by individual codon at the mRNA wobble position shows a preference for A bases in proliferation but T is split between the two conditions at the mRNA wobble position.

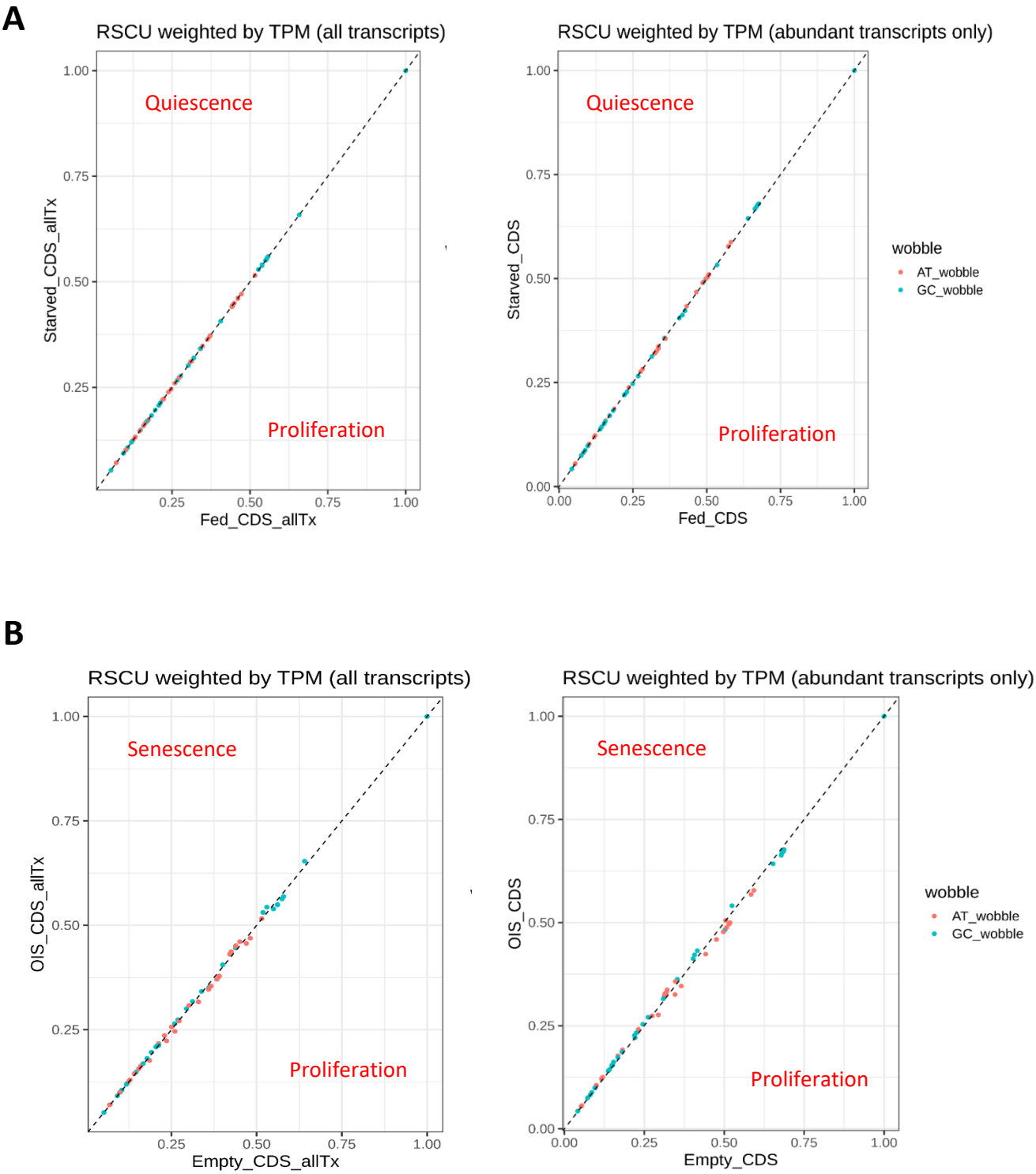


Figure 4.25 Relative synonymous codon usage for all transcripts and the most abundance transcript per gene. (A) Fed vs Starved shows no RSCU preference in any direction when calculated for all transcripts and the most abundant transcript per gene. (B) Empty vs Ras shows negligible RSCU preference in any direction when calculated for all transcripts and the most abundant transcript per gene.

4.7 Investigation into potential A34I modifications at the isodecoder level

The tRNA sequencing results showed the importance of tRNA degeneracy in decoding mRNA sequences that vary in relative synonymous codon usage. The most important modification observed was the adenosine to inosine modification on position 34 of the tRNA wobble position. Since this modification allows for the decoding of A/T/C ending codons, enhancing the flexibility of the isoacceptor pool, it was important to ask how the modification rates changed for the isoacceptors they existed on between conditions.

The mismatch frequency at position 34 was calculated by allowing for 1 mismatch when the tRNAs were mapped back to the mature tRNA database. If the mismatch occurred at position 34 for isoacceptors Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val, and the base was read as a G base, then this would signal an A34I mutation (Section 4.4.5). The mutation frequency was then calculated for all tRNAs in that isoacceptor group and compared between conditions.

For both Fed vs Starved (Figure 4.26 A) and Empty vs Ras (Figure 4.26 B), there were no statistically significant differences within isoacceptors between the two conditions. The mutation rate calculations showed that A34I modifications at these isodecoders were very high in both conditions, suggesting that tRNA degeneracy is not regulated by changes in synonymous codon usage, rather it is a mechanism that appears to be constitutively active.

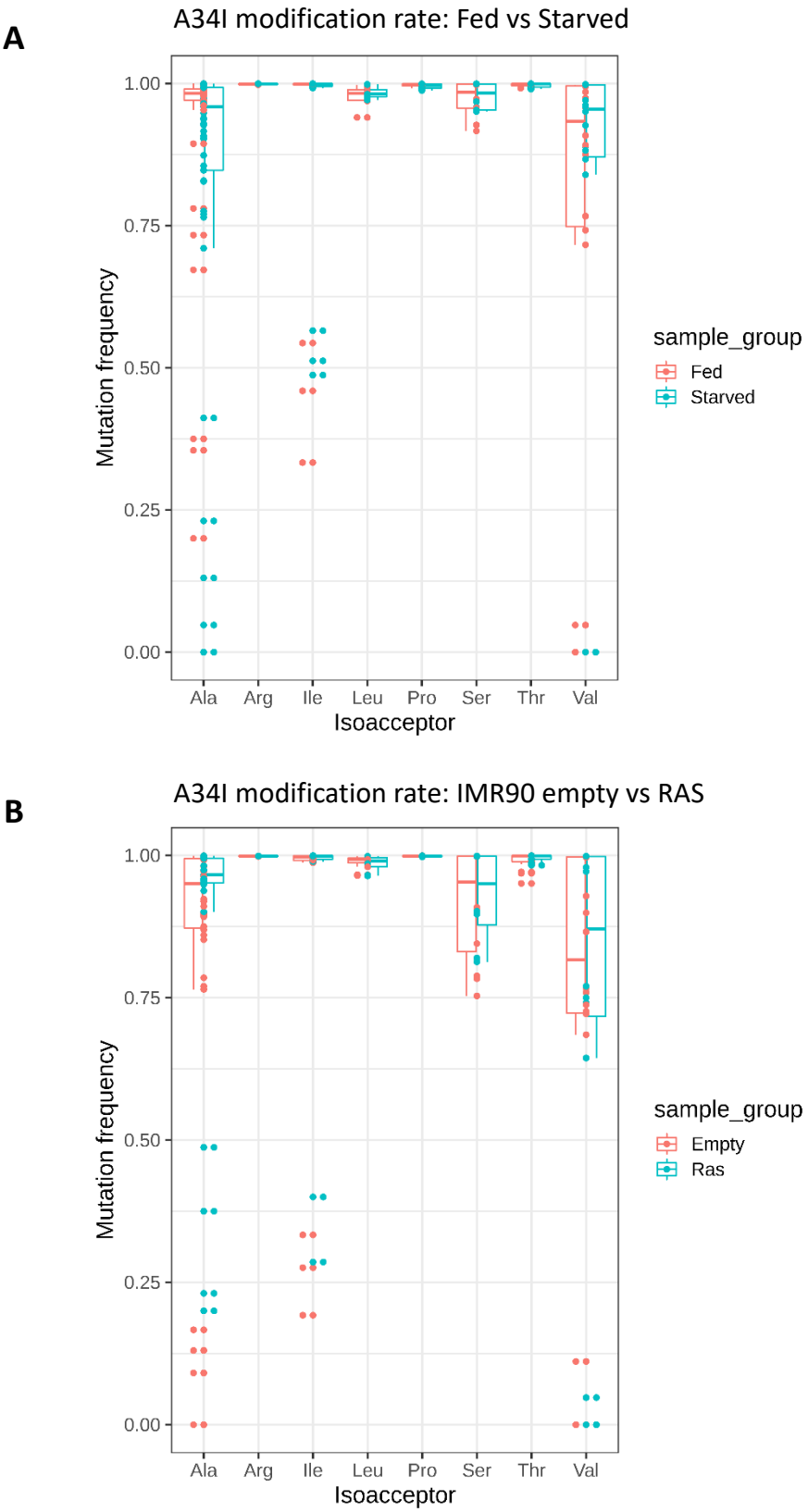


Figure 4.26 A34I modification changes at the isoacceptor level. (A) A34I modification rates were calculated in Fed vs Starved but there was no statistically significant difference between proliferation and quiescence. (B) A34I modification rates were calculated in Empty vs Ras but there was no statistically significant difference between proliferation and senescence.

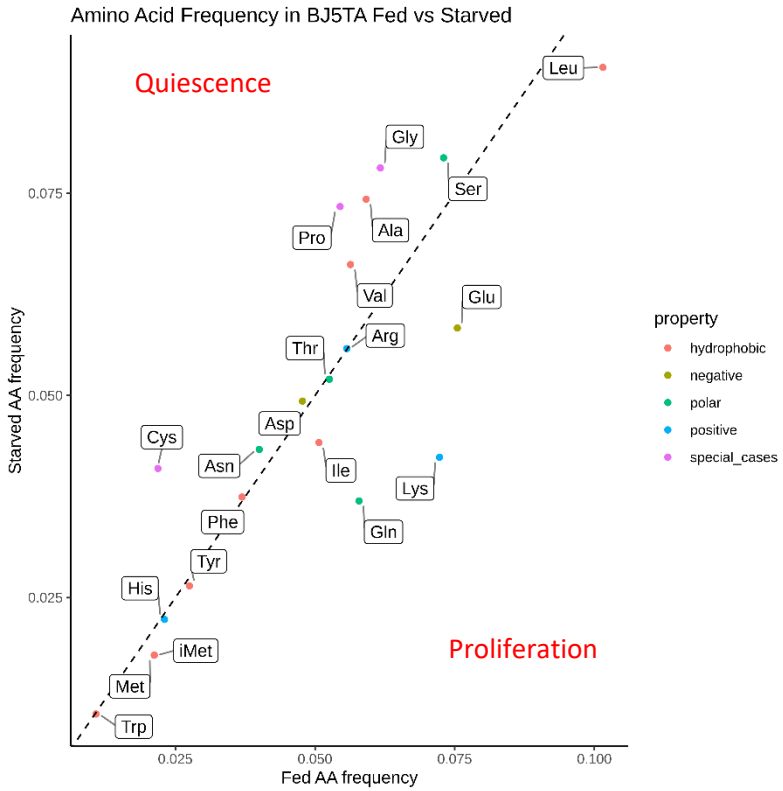
4.8 Investigating if amino acid frequencies change in differentially expressed transcripts

Since some of the 2-box tRNAs all had log fold changes in the same direction in both Fed vs Starved and Empty vs Ras, it was important to check how amino acid frequency varied depending on the cell state. All statistically significant, differentially expressed genes with a $\log_2$ fold change of ±2 were used to calculate the amino acid frequency, normalized against the abundance of each gene.

In Fed vs Starved, where cells are driven from proliferation to quiescence, there were clear differences in amino acid frequency between the two conditions (Figure 4.27 A). Glutamine, isoleucine, lysine, glutamic acid and leucine were all prevalent in the Fed (proliferation) condition. In the starved condition (quiescence), glycine, proline, cysteine and alanine were the prevalent amino acids. This suggests that the amino acid frequencies in the transcripts could be driving some of the tRNA profiles observed, instead of the relative synonymous codon usage.

In Empty vs Ras on the other hand, the difference in amino acid frequencies in the transcripts between proliferation and senescence was even between the two conditions, suggesting the importance of relative synonymous codon usage in the observed tRNA profiles.

A



B

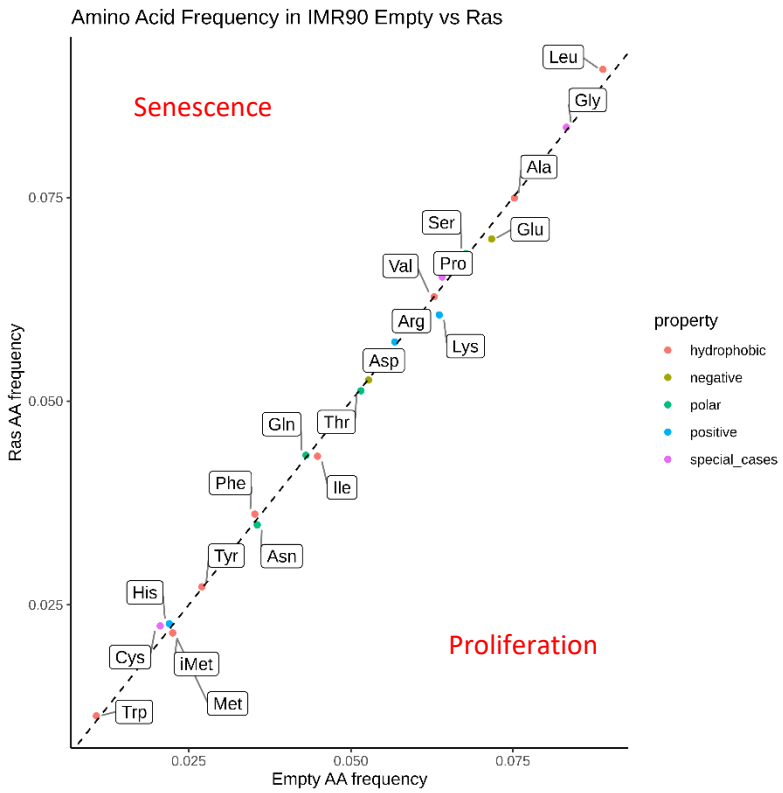


Figure 4.27 Amino acid frequencies in proliferation vs quiescence and proliferation vs senescence. All statically significant, differentially expressed genes with a log2 fold change of ±2 were used to calculate the amino acid frequency, normalized against the abundance of each gene. (A) In Fed vs Starved, quiescence showed prevalence for glutamine, isoleucine, lysine, glutamic acid and leucine and in proliferation, glycine, proline, cysteine and alanine were the prevalent amino acids. (B) Empty vs Ras showed no prevalence for any particular amino acid in their measured transcriptome

4.9 tRNA amino acid frequency vs mRNA amino acid frequency

Lastly, the tRNA amino acid frequencies were correlated with the mRNA amino acid frequencies after being calculated for both experiments (Figure 4.28 and Figure 4.29) to see how tRNA availability matched with the transcriptomic demand. mRNA codons used were calculated from all transcripts per gene and were normalized to their corresponding TPM (transcript per million) to weight for abundance and then summarised by amino acid. The tRNA isoacceptor frequencies were counted after mapping and were normalized to their corresponding CPM (counts per million) and then summarised by amino acid.

Fed tRNA amino acid frequency was plotted against Fed mRNA amino acid frequency, showing moderate correlation (R^2 = 0.689), suggesting tRNA availability matched with the mRNA codons that needed to be translated (Figure 4.28 A). Interestingly, when the same was done for the Starved condition, a very strong correlation was observed (R^2 = 0.999), suggesting tRNA availability matched almost perfectly with the mRNA codons that needed to be translated (Figure 4.28 B). This suggests that when cells are in quiescence, the tRNA pool may suffice in meeting the transcriptomic demands of the cell, but may rely more on tRNA degeneracy in proliferation.

Empty tRNA amino acid frequency was plotted against Empty mRNA amino acid frequency, showing a weak positive correlation (R^2 = 0.4217) (Figure 4.29 A). When the same was done for the Ras condition, a weak positive correlation was also observed (R^2 = 0.3386) (Figure 4.29 B). Again, this may suggest that cells in proliferation (as observed in the starved condition) and cells in senescence may rely heavily on tRNA degeneracy to meet the demands of the transcriptome.

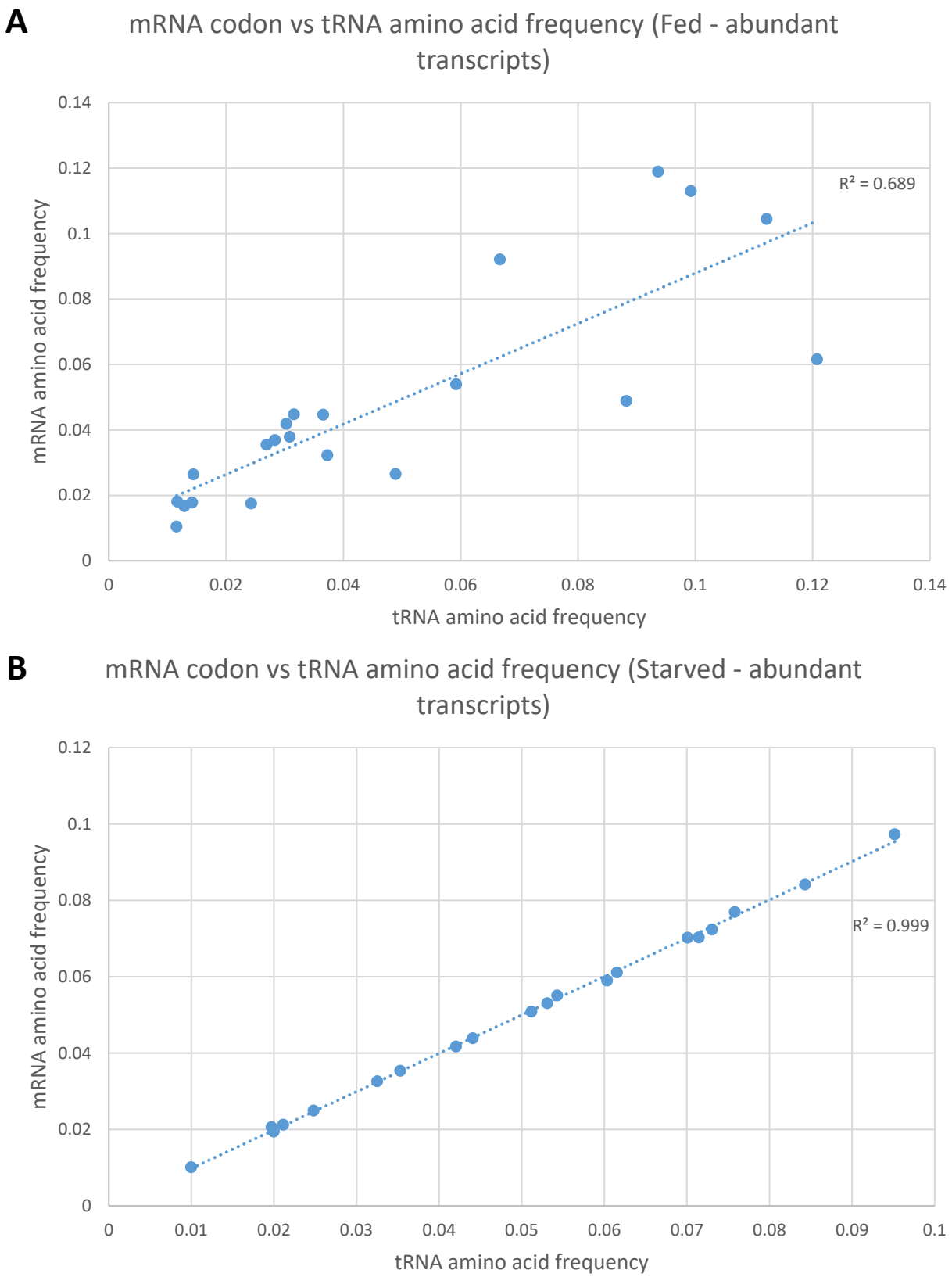


Figure 4.28 Strong correlation observed for tRNA vs mRNA amino acid frequency in Fed vs Starved. (A) Fed mRNA codon frequency of the most abundant transcript for each gene plotted against Fed tRNA isoacceptor frequency. (B) Starved mRNA codon frequency of the most abundant transcript for each gene plotted against Starved tRNA isoacceptor frequency.

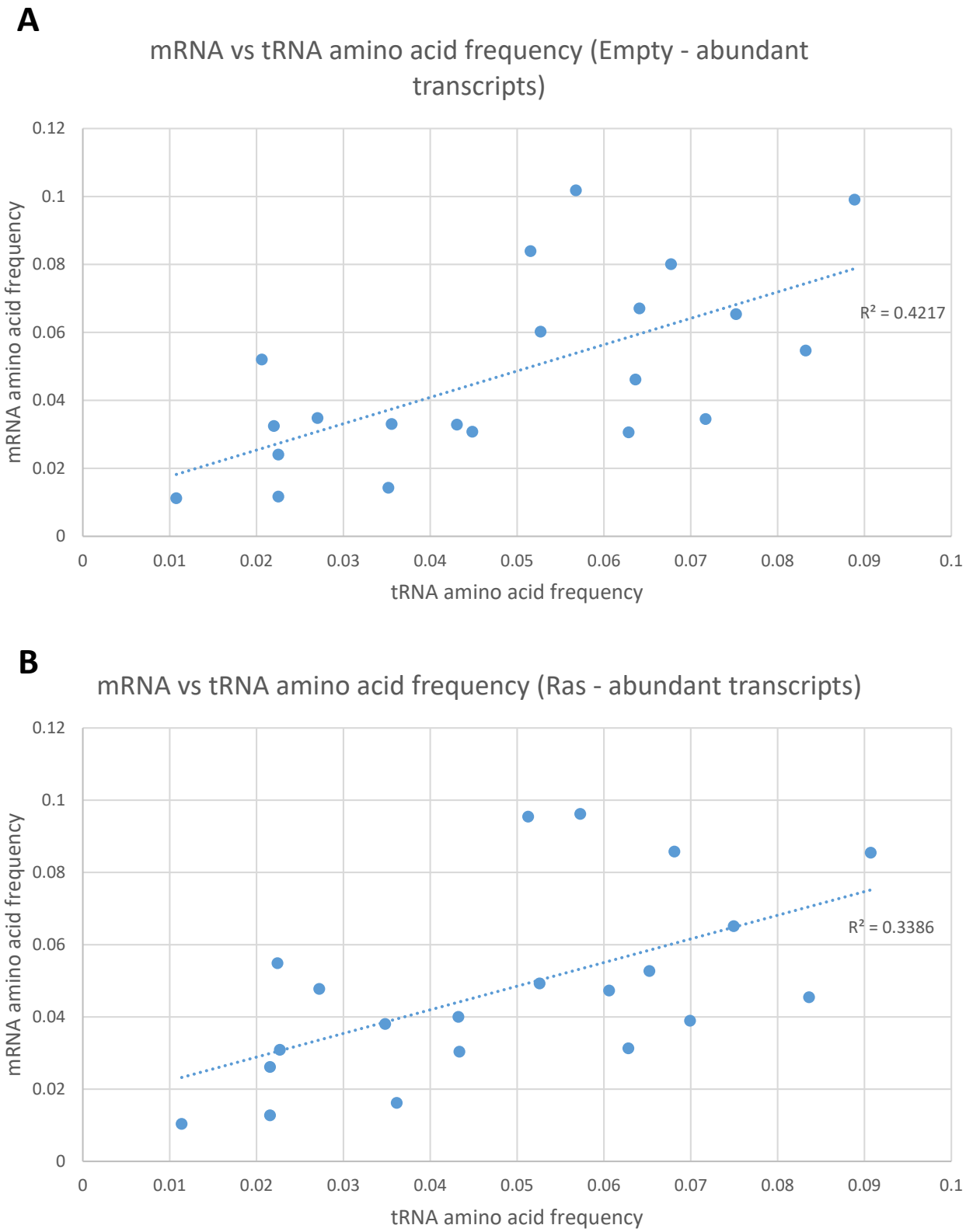


Figure 4.29 Weak positive correlation observed for tRNA vs mRNA amino acid frequency in Empty vs ER:RAS. (A) Empty mRNA codon frequency of the most abundant transcript for each gene plotted against Empty tRNA isoacceptor frequency. (B) ER:RAS mRNA codon frequency of the most abundant transcript for each gene plotted against ER:RAS tRNA isoacceptor frequency.

4.10 Chapter discussion

In this chapter, we utilized two *in vitro* models of quiescence and senescence to try and elucidate their tRNA and mRNA profiles, by sequencing the mRNAs and tRNAs and examine how they vary upon changes in cell state. The Fed vs Starved experiment in BJ5TA fibroblasts was a stress-induced senescence model and had generally followed the trends reported in literature. It was previously mentioned that A/T codons at the 3rd position of mRNA sequences were preferred for proliferation associated genes and G/C codons preferred in senescence associated genes (Gingold et al., 2014). The tRNA sequencing for the Fed vs Starved experiment confirmed this at the isoacceptor level, with almost all tRNA isoacceptors with an A or T base at their wobble position being downregulated in quiescence. When tRNAs did not match the demands highlighted from the relative synonymous codon usage analysis, it was clear that tRNA degeneracy may have been able to meet the demands, either through A34I modifications or by utilising G:T base pairing.

The second model discussed was the oncogene induced senescence model Empty vs Ras in IMR90 primary human fibroblasts. Briefly, a mutant estrogen receptor (ER) has an H-RAS^{G12V} fused to its ligand binding domain, allowing for the activation of RAS via 4-hydroxy-tamoxifen (4-OHT) treatment, resulting in growth arrest and induction of senescence associated phenotypes (Innes and Gil, 2019). This results in the non-treated fibroblasts remaining in proliferation whereas 4-OHT treated fibroblasts are driven into senescence. We were able to establish the tRNA and mRNA profile for this model system. Unlike the Fed vs Starved experiment, the A/T and G/C changes at the tRNA wobble position in proliferation and senescence were not distinctive, but tRNA degeneracy (either through A34I or G:T base pairing) seemed to aid in meeting the transcriptomic demand. Also, at the mRNA level, it was clear that there was a clear preference for A codons at the wobble

position of mRNAs in proliferation and a preference for G/C ending codons in senescence. However T codons are generally balanced across the two conditions, with negligible preferences in both directions.

When the A34I modifications in both experiments were examined, it was clear that the A34I remained high, regardless of the cell state (proliferation/quiescence/senescence), or the type of induction leading to changes in cell state (stress-induced or oncogene-induced). This suggests that the A34I modification is constitutively active and can help in meeting the demand of the transcriptome.

Interestingly, when the amino acid frequencies for both experiments were examined, there was prevalence for particular amino acids in the quiescence (Starved) conditions and the proliferation (Fed) conditions. This suggests that it may not be relative synonymous codon usage alone driving the transcriptomic demand. Although the amino acids were broken down according to their properties, there wasn't a clear property-dependent difference between the two conditions. Interestingly, when proliferation and senescence amino acid frequencies were compared (Empty vs Ras), there was no prevalence for particular amino acids.

Although we managed to build both tRNA and mRNA profiles for the two in vitro models, it is still unclear as to whether the transcriptomic demand drives changes in tRNA expression or vice versa. However, the importance of tRNA degeneracy does give some indication that the tRNAs are able to stay adaptive, regardless of the transcriptomic demand.

Chapter 5: Using ALICE-tRNA-seq to investigate changes in the tRNA pool in *in vivo* models of proliferation and senescence

5.1 Chapter Introduction

The study of cancer *in vivo* allows for the observation of phenotypic endpoints as well as the real time monitoring of human diseases. Measuring the tRNA pool *in vivo* therefore allows the study of the tRNA-ome in primary tissue, which could provide unique information about the translational landscape during disease development.

Hepatocellular carcinomas (HCC), a type of liver cancer that is a leading, worldwide candidate for cancer-related mortality is both an area of interest and a disease that requires improved therapeutic solutions (Zucman-Rossi et al., 2015). Although previous genomic-sequencing investigations have revealed heterogeneity in HCC genomic alterations, the clarification of the molecular mechanisms driving HCC is still an area requiring further elucidation (Schulze et al., 2015).

The WNT/βcatenin signalling pathway, also known as "canonical" WNT signalling, remains one of HCC's most regularly altered pathways (Ally et al., 2017). When WNT ligands bind to their receptors on the cell surface, it results in the stabilization of its transcriptional co-activator βcatenin, which then translocates to the nucleus and activates the transcription of gene targets associated with the TCF (T cell factor/lymphoid enhancer factor family) DNA-binding factors (Nusse and Clevers, 2017). The regulation of βcatenin occurs via kinases and ubiquitin ligases which include (but are not limited to) GSK3α/β, CK1α/δ, the ubiquitin ligase β-TrCP, APC and Axin1, which work to phosphorylate several sites of the βcatenin protein, resulting in its ability to bind β-TrCP, leading to βcatenin being ubiquitinated and targeted for proteasome-mediated degradation (Nusse and Clevers, 2017). In cancer, the activation

of this pathway normally occurs due to mutations in the βcatenin gene CTNNB1 (regularly observed in HCC), the loss of the kinase APC (regularly observed in colorectal cancer) or heightened exposure to WNT ligands (Schulze et al., 2015; Zucman-Rossi et al., 2015; Nusse and Clevers, 2017).

The proto-oncogene MYC is a downstream effector of WNT signalling and βcatenin/TCF mediated transcription in many tissues including the lungs, T-cells and small intestines (De La Coste., 1998; Sansom et al., 2007). However, this interaction between WNT signalling and MYC may not hold true in the liver. MYC expression is not induced in the loss of APC and the subsequent activation of βcatenin (Colnot et al., 2004). Hepatocyte proliferation has been shown to be unaffected by the loss of APC and the deletion MYC (Reed et al., 2008). Furthermore, experimental mouse models of MYC-driven hepatocellular carcinoma often requires the activation of mutations in the CTNNB1 gene (Yim et al., 2018). Although both the WNT/βcatenin and MYC pathways appear to act independently, their functional cross-talk has been shown to result in tumorgenesis in the liver (Bisso et al., 2020).

Another major driver often observed in hepatocellular carcinomas is the dysfunction of the MDM2-p53 axis (Meng et al., 2014). MDM2 is an E3 ubiquitin ligase and a negative regulator of p53. When p53 is in excess, MDM2 binds p53, resulting in its ubiquitination and subsequent targeting for proteasome-mediated degradation (Gannon et al., 2011). Both p53 and MDM2 have been shown to be abnormally expressed in HCC (Jablkowski et al., 2005). When MDM2 regulation fails, the activation of p53 via hypoxia, ribosomal stress, oncogenic activation or genotoxic stress may result in the expression of genes related to cancer initiation and development (Sullivan et al., 2018). However, widespread hepatocyte senescence has been observed in both advanced human liver disease and HCC through p21

and p16 positivity (senescence markers), often driven by p53-mediated senescence or p53-mediated cell death (Gouw et al., 2011, Lu et al., 2015).

In this chapter, we utilize two *in vivo* models; the βcatenin/c-Myc model which drives hepatocytes from quiescence to proliferation and an MDM2 model which drives hepatocytes from quiescence to senescence. The aim of this chapter is to understand if ALICE-tRNA-seq can be used to measure the tRNA-ome *in vivo*. Together, with traditional RNA-seq, we attempt to build both a tRNA profile and a profile for relative synonymous codon usage to try and observe the translational landscape of these hepatocytes, when they transition to either proliferation or senescence.

5.2 Experimental design

After showing that ALICE-tRNA-seq was an effective means of measuring the tRNA pool *in vitro*, the next step was to see if the same could be achieved *in vivo*. RNA-seq and ALICE-tRNA-seq was performed on genetically modified mouse models (*Mus musculus*) to try and capture the tRNA-ome and mRNA-ome. To reduce variability, all sequencing experiments were performed on male mice that were between 8-12 weeks in age and were of similar weight before being culled. All experiments were carried out on the caudate lobe of the liver and liver tissue was collected at the same time. Furthermore, all treatments were carried out at the same time for all replicates.

In the βcatenin/c-Myc model, mice were bred to introduce two loxP sequences around exon3 of βcatenin and introduce a humanised c-MYC (R26-lslMYC) and Cre recombinase (to allow for conditional expression). Upon Cre recombinase activation via AAV8-TBG-PI-Cre-rBG treatment, exon 3 of the βcatenin is cleaved, resulting in the in-frame connection of exons 2 and 4, and the deletion of 76 amino acids in the βcatenin protein, rendering all expressed βcatenin degradation resistant, as well as the expression of the humanised form of c-MYC. Previous studies on mice that had utilized degradation resistant βcatenin alone hypothesised that many of the hepatocytes get stuck in the G_0 phase of the cell cycle (Harada et al., 1999, Barker et al., 2009). Therefore, c-Myc is used in this system and is hypothesised to overcome this repression and drive the tissues into proliferation, via MYC-driven carcinogenesis (Bisso et al., 2020). All sequencing experiments were carried out 4 days post-tamoxifen treatment. Mice that expressed βcatenin/c-Myc had visibly larger livers (due to the hyper proliferation) and βcatenin/c-Myc expression was confirmed histologically via H&E staining (data not shown). All mice in this model were either treated with an empty adenovirus, hereafter referred to as Null (control mice), or the βcatenin/c-Myc/Cre-

recombinase adenovirus, hereafter referred to as Cre (βcatenin/c-Myc induced proliferation).

In the MDM2 model, normal mice were crossed with mice that had exons 5 and 6 floxed (introduction of loxP sequences around said exons) in the MDM2 gene (Grier et al., 2006, Bird et al., 2018). Excision of exons 5 and 6 was induced via treatment with the hepatocyte-specific adenovirus AAV8-TBG-Cre, which activates the TBG (thyroxine binding globulin) promoter, resulting in the expression of Cre recombinase. MDM2 is an E3 ubiquitin ligase and a negative regulator of the tumour suppressor p53 (Gannon et al., 2011). Exons 5 and 6 code for the p53 binding domain of MDM2 and upon their excision, MDM2 is unable to bind p53, resulting in p53 overexpression and therefore the tissues being driven into senescence (Bird et al., 2018). All sequencing experiments were conducted 4 days after adenovirus treatment. Control mice were treated with an empty adenovirus (hereafter referred to as Null) and the floxed MDM2 mice are referred to as Cre (MDM2-induced senescence). Senescence was confirmed via p21 staining (a positive marker of senescence, data not shown) after mice were culled.

In the βcatenin/c-Myc model, hepatocytes were driven from quiescence to proliferation and in the MDM2 model, hepatocytes were driven from quiescence to senescence (Figure 5.1). Both these models were utilised to investigate how the tRNA-ome and mRNA-omes change upon oncogene-induced proliferation and senescence.

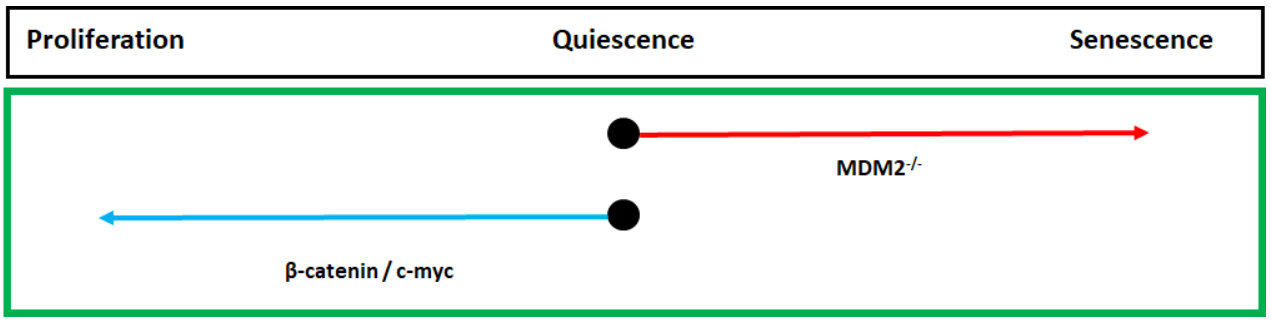


Figure 5.1 Experimental design in trying to establish the tRNA pool *in vivo*. tRNA and RNA-seq was carried out on the MDM2 model, which drives liver tissue from quiescence to senescence and the βcatenin/c-Myc model which drives liver tissue from quiescence to proliferation.

5.3 Confirming AlkB demethylation *in vivo*

Before any tRNA-seq library preparations were conducted, it was essential to check if AlkB (the enzyme used to remove methyl groups from the Watson-Crick face of tRNAs) could demethylate modified nucleotides *in vivo*. To test this, primer extension assays (refer to Figure 3.1) were carried out on tRNA-Gln-CTG, tRNA-iMet-CAT and tRNA-Val-CAC. Primers were designed against the 3' end of these tRNAs. These tRNAs were chosen due to their variation in modifications (tRNA-Gln-CTG being the most modified and tRNA-Val-CAC being the least modified). All primer extension assays were carried out on healthy male and female hepatocytes extracted from primary tissue and the assays were carried out on total RNA and small enriched RNAs (RNAs < 200nt).

Ctrl samples are the RNA samples that had no AlkB treatment and WT + D135S referred to assays that were demethylated using the WT AlkB and the mutant form D135S AlkB. AlkB was able to successfully demethylate modified tRNA bases in both male and female mice hepatocytes for all three chosen tRNAs, as evident from the large increase in full length products following AlkB treatment (Figure 5.2 A-C). For tRNA-Gln-CTG (Figure 5.2 A) and tRNA-iMet-CAT (Figure 5.2 B), almost no full length tRNAs were detected when the RNAs were not treated with AlkB (Ctrl). However, when the same demethylation conditions that were used in the *in vitro* experiments were used, full length tRNA was detected (+ WT + D135S). In the tRNA-Val-CAC assay (Figure 5.2 C), full length tRNAs were observed even in the absence of AlkB but this was expected since tRNA-Val-CAC is one of the least modified mouse tRNAs. However, after demethylation, a clear increase in full length tRNAs were observed.

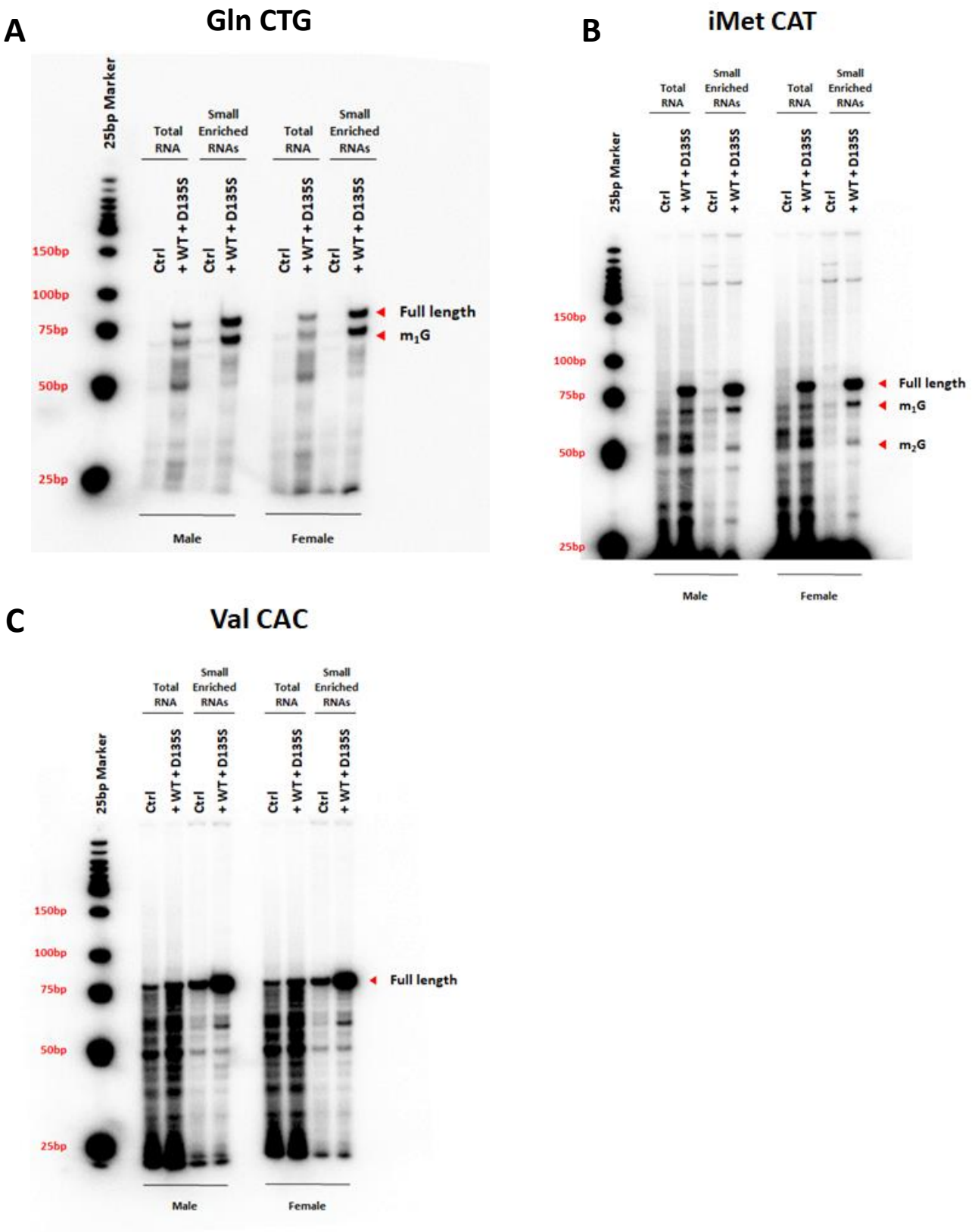


Figure 5.2 AlkB successfully demethylates modified tRNA bases in mouse liver. AlkB demethylation was assessed on the caudate lobe of normal, healthy, male and female mouse liver. The assay was done on both total RNA and small enriched RNAs (RNA<200nt) Primer extension assays were carried out to determine demethylation effectiveness, probing for (A) tRNA-Gln-CUG, (B) tRNA-iMet-CAT and (C) tRNA-Val-CAC. Ctrl in these gels refer to the – AlkB condition and WT + D135S refers to the reaction containing both WT AlkB and the mutant D135S AlkB. Red arrows refer to predicted modifications that may have caused truncation.

5.4 Library and Sequencing quality controls

5.4.1 Library quality controls

The same pre and post library quality control steps done in Chapters 3 and 4 were also conducted on the βcatenin/c-Myc and MDM2 models. All extracted total RNAs from the liver samples were run on a 1% agarose gel to check for RNA integrity. All 5 replicates for both experiments passed the initial RNA integrity check, and were used as input for the RNA-seq library.

tRNAs were extracted from the small enriched RNAs and were size selected between 50-100nt. All 5 replicates for both experiments showed tRNA saturation and were successfully extracted and used as input for the ALICE-tRNA-seq library protocol. All 5 replicates for both experiments underwent successful post-RT and post-PCR gel extraction, before being pooled for sequencing.

5.4.2 Post sequencing quality controls

RNA mapping, amino acid distribution, fragment determination and read coverage was checked for both the βcatenin/c-Myc and MDM2 models prior to analysis. For both models, more than 85% of mapped reads were mapped to tRNAs. tRNA amino acid frequency was also determined and no bias was observed. Furthermore, more than 95% of reads were either full length tRNAs or 3' fragments (tRNAs that were truncated in the reverse transcription step of the ALICE-tRNA-seq due to modifications AlkB could not remove, refer to Table 3.1). More than 60% of reads in the βcatenin/c-Myc model and 55% of reads in the MDM2 model passed the anticodon loop. All tRNA counts were normalized using DESeq2 Size Factors (normalization by CPM).

A principal component analysis was also performed on both models to check how replicates clustered and to determine which replicates would be used in downstream analysis. In the βcatenin/c-Myc model, replicates Null3, Null5 and bcat_myc1 were removed from all downstream analysis (Figure 5.3 A). The principal component analysis was performed on filtered replicates (and renamed to avoid confusion) and good separation was observed along PC1 (accounting for 73% variance) between the quiescence and proliferation samples (Figure 5.3 B).

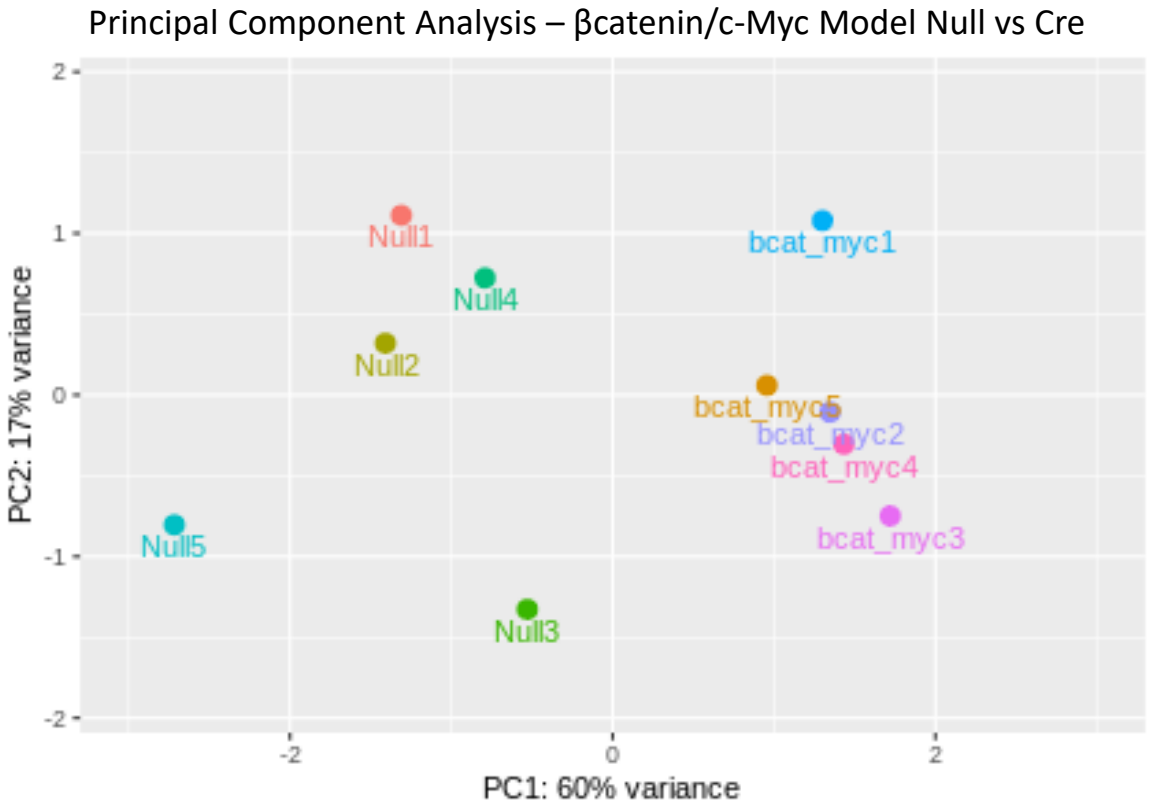
In the MDM2 model, replicates Cre1 and Cre2 were removed from all downstream analysis (Figure 5.4 A). The principal component analysis was performed on filtered replicates (and renamed to avoid confusion) and good separation was observed along PC1 (accounting for 80% variance) between the quiescence and senescence samples (Figure 5.4 B).

For both models, separation was observed along PC2, which accounted for 13% and 10% variance for the βcatenin/c-Myc and MDM2 experiments respectively. However, since

sequencing was performed on 10 different mice for both experiments, some variability was expected and so was ignored. Importantly, separation between quiescence and proliferation and quiescence and senescence was observed for the βcatenin/c-Myc and MDM2 models respectively.

Another quality control step incorporated prior to differential expression was checking for the mismatch frequency at position 34 (tRNA wobble position). Since there are less tRNA anticodons with respect to mRNA codons, tRNAs utilise an adenosine to inosine modification at the tRNA wobble position of select isoacceptors, to accommodate non-cognate base pairing and recognition. As inosines prefer to base-pair with cytosines, inosine bases are recognised as guanosines in next generation sequencing data. As the reference genome contains an adenosine at this position, when reads are mapped back to the mature tRNA database (allowing for only 1 mismatch), reads with the mismatch at position 34 are likely due to this adenosine to inosine modification (A34I). The tRNA mismatch frequency analysis determined that the inosine modifications only occurred on A bases and not seen in tRNAs with a G/C/T at the wobble position for both experiments (Figure 5.5 A-B).

A



B

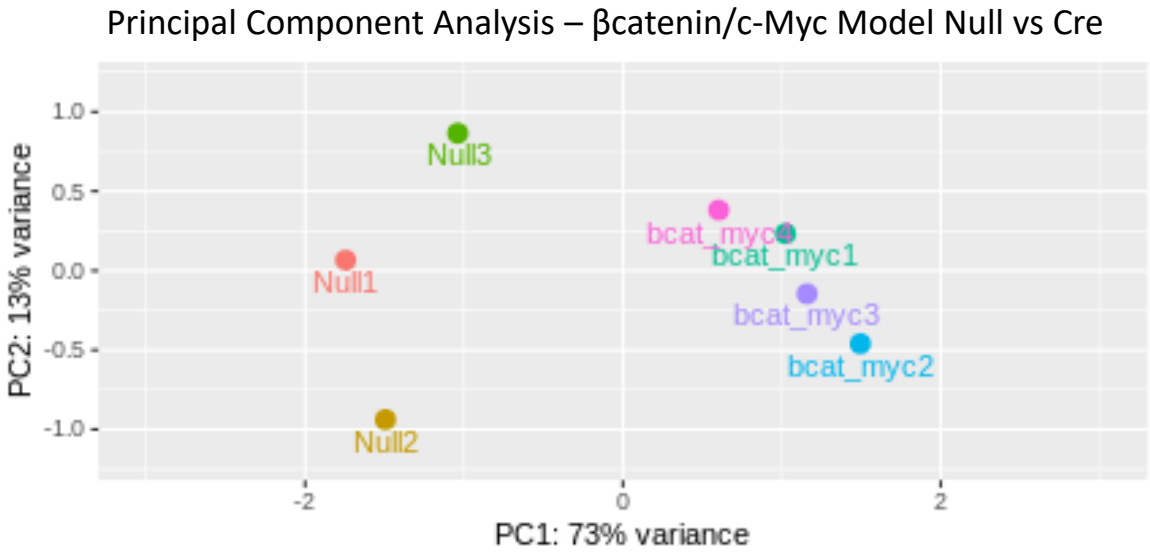


Figure 5.3 Principal component analysis on the βcatenin/c-Myc model. principal component analysis of the tRNA counts to determine where samples clustered. (A) PCA was carried out on all sequenced replicates. Replicates Null3, Null5 and bcat_myc1 were removed from downstream analysis. (B) After the removal of replicates, all remaining replicates where re-named to avoid confusion. Good separation across the two conditions on PC1 (accounting for 73% of the variance).

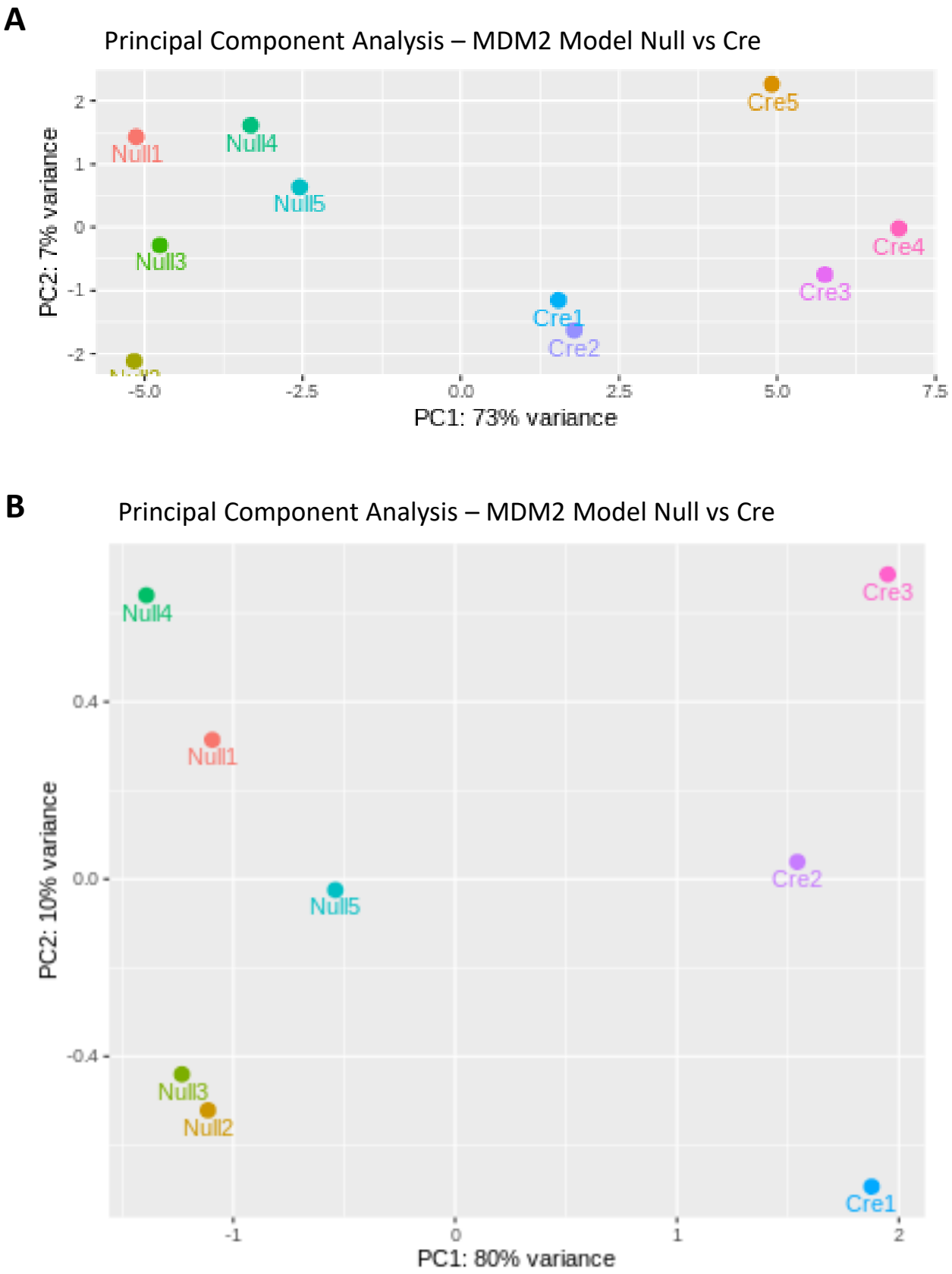
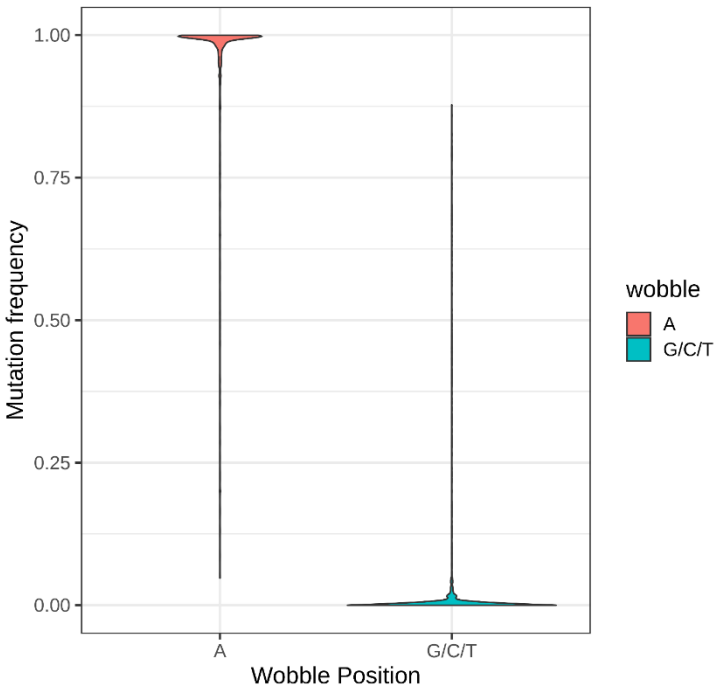


Figure 5.4 Principal component analysis on the MDM2 model. principal component analysis of the tRNA counts to determine where samples clustered. (A) PCA was carried out on all sequenced replicates. Replicates Cre1 and Cre2 were removed from downstream analysis. (B) After the removal of replicates, all remaining replicates where re-named to avoid confusion. Good separation across the two conditions on PC1 (accounting for 73% of the variance).

A



B

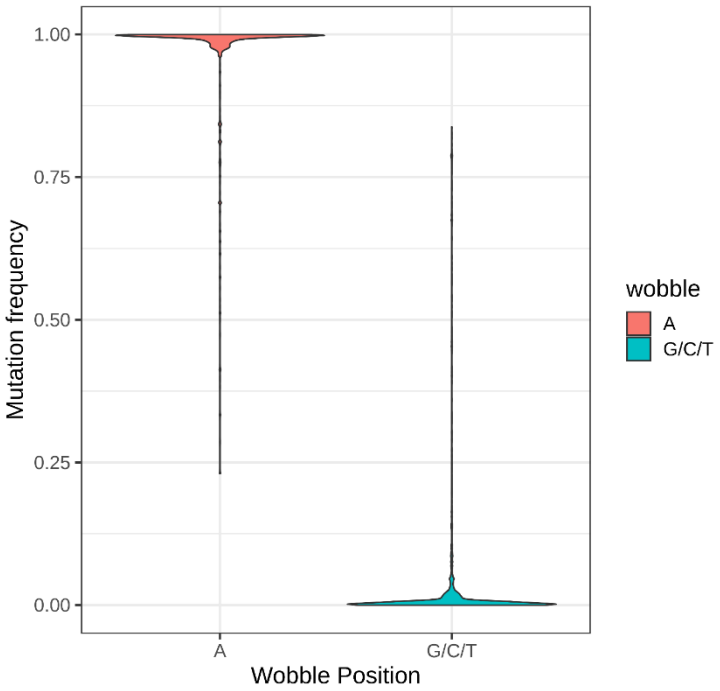


Figure 5.5 tRNA mismatch frequency at position 34 (wobble position) to check for inosine modification (A34I). (A-B) Mismatch frequency for tRNAs with A at the wobble position vs G/C/T for both the βcatenin/c-Myc and MDM2 models respectively. tRNAs were mapped allowing for 1 mismatch in order to determine the inosine modification at position 34 (wobble position). As expected, inosine modifications should only occur on A bases and not seen in tRNAs with a G/C/T at the wobble position. (A) βcatenin/c-Myc model (B) MDM2 model.

5.5 Differential tRNA expression at the isoacceptor level vs mRNA relative synonymous codon usage

tRNA box-types is a breakdown of tRNAs according to the number of codons they can decode (see Table 4.1). 6-box tRNA sets can decode 6 mRNA codons, although not all complementary anticodons are expressed. Therefore, an inosine modification at position 34 of tRNAs or G:T base pairing is utilised to decode all codons. 4-box tRNA sets follow the same reasoning and can decode 4 codons, but only 3 anticodons are expressed. There are two types of 2-box tRNA sets, those that express two anticodons and those that express one anticodon. Lastly, 1-box tRNA sets are those that decode 1 codon.

As explained in Chapter 4, the tRNA differential expression analysis was broken down into their box-types by amino acid. The RNA sequencing that was done in parallel was used to determine the relative synonymous codon usage of the transcriptome.

5.5.1 Comparison of 6-box tRNAs to the relative synonymous codon usage in mRNA

6-box isoacceptors include tRNAs that code for arginine, leucine and serine (Figure 5.6 A-C and Figure 5.7 A-C). Each set contains one isoacceptor that has the potential for A34I modification (Arg-ACG, Leu-AAG and Ser-AGA). Also, tRNA-Ser-GCT allows for G:T base pairing with the codon AGT.

In the βcatenin/c-Myc model, where hepatocytes are being driven from quiescence to proliferation, statistically significant tRNAs with a G/C at the tRNA wobble position are downregulated in proliferation.

tRNA-Arg-CCT is downregulated in proliferation and its cognate codon AGG is slightly preferred in quiescence in comparison to proliferation. Although the mRNA codon AGA is the most preferred codon, changes in its cognate tRNA-Arg-TCT is not statistically significant. Interestingly, tRNA-Arg-ACG (which can undergo A34I modification and can decode its cognate CGT codon as well as codons CGC and CGA, refer to Table 4.1) is the highest expressed tRNA. This may suggest a role for tRNA degeneracy in compensating for lowly expressed tRNAs in the arginine family (Figure 5.6 A).

tRNA-Leu-CAG is downregulated in proliferation and its cognate codon CTG is slightly preferred in quiescence in comparison to proliferation. However, tRNA-Leu-TAG was also down regulated in proliferation despite its cognate codon CTA being preferred in proliferation. However Leu-TAG also decodes CTG codons through wobble base-pairing, which as mentioned above is preferred in quiescence. tRNA-Leu-AAG (which can decode its cognate CTT codon as well as CTC and CTA) was also downregulated in proliferation (Figure 5.6 B).

tRNA-Ser-AGA (which can decode its cognate TCT codon as well as TCC and TCA) was up regulated in proliferation, however, tRNA-Ser-TGA (which encodes TCA) was downregulated in proliferation. Again, tRNA-Ser-AGA through its A34I modification may compensate for losses in tRNA-Ser-TGA expression. mRNA codon AGC was the most preferred codon in both conditions but changes in tRNA-Ser-GCT which decodes it were not statistically significant (Figure 5.6 A).

In the MDM2 model where hepatocytes are driven from quiescence to senescence, there was no correlation between cell state and the base at the tRNA wobble position.

tRNA-Arg-CCG (which decodes codon CGG, which is slightly preferred in senescence), was the only tRNA with a statistically significant log fold change within the arginine family and was upregulated in senescence. tRNA-Arg-ACG (which has the A34I modification) had one of the highest read counts like in the βcatenin/c-Myc model, again suggesting a role for tRNA degeneracy in compensating for lowly expressed tRNAs in the arginine family, but wasn't differentially expressed between the two conditions. The codon AGA was the most preferred codon in both conditions but changes in its cognate tRNA-Arg-TCT were not statistically significant (Figure 5.7 A)

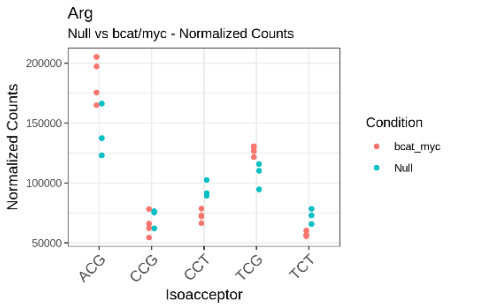
tRNA-Leu-AAG was upregulated in senescence (which decodes its cognate CTT as well as CTC and CTA through the A34I modification). Furthermore, the two most preferred codons were CTA (preferred in quiescence) and CTG (preferred in proliferation), both of which can be decoded by the tRNA-Leu-AAG, suggesting a role for tRNA degeneracy in senescence (Figure 5.7 B).

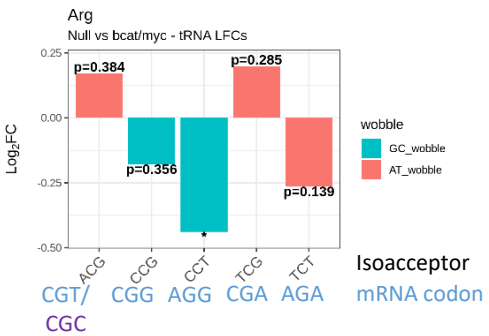
For tRNA-Ser-GGA, the codon AGC was the most preferred codon in both conditions, but there were no statistically significant changes within the tRNA-Ser isoacceptor family.

However, its cognate tRNA-Ser-GCT was the most highly expressed tRNA within the family (figure 5.7 C).

6-box tRNAs – β-catenin/c-Myc Null vs Cre

A

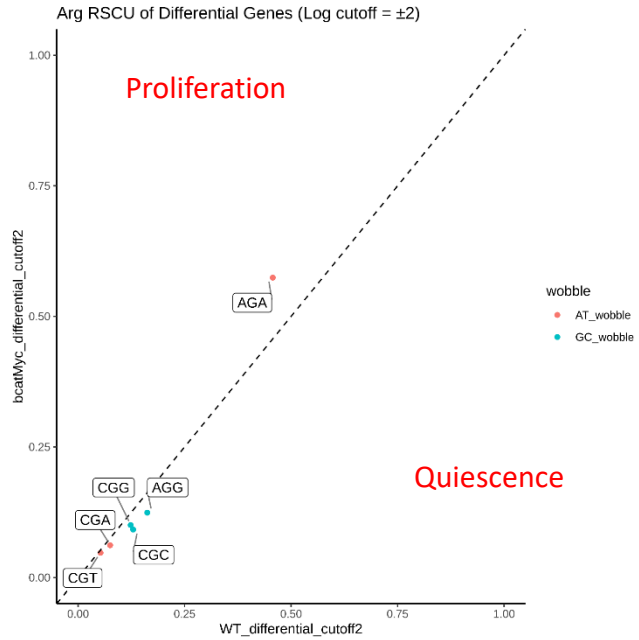




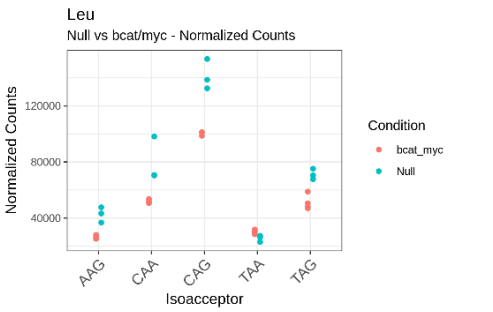
Isoacceptor

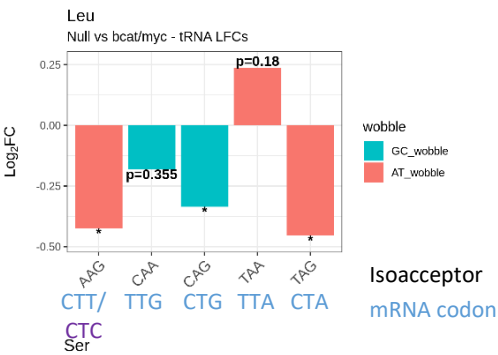
CGT/ CGC | CGG | AGG | CGA | AGA

mRNA codon



B

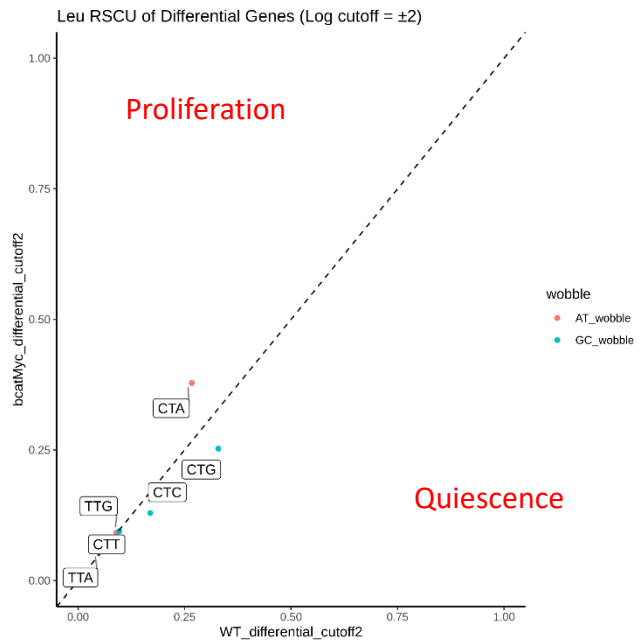




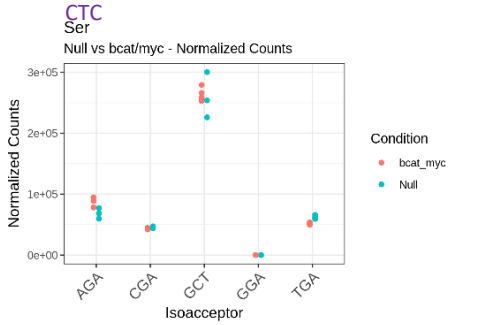
Isoacceptor

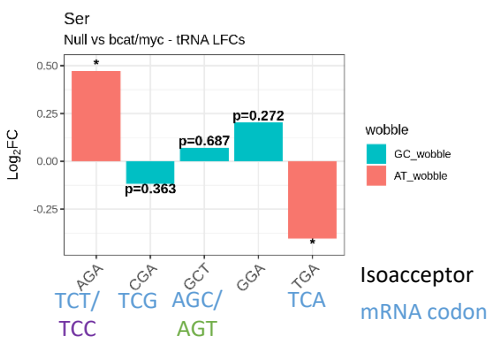
CTT/ CTC | TTG | CTG | TTA | CTA

mRNA codon



C





Isoacceptor

TCT/ TCC | TCG | AGC/ AGT | TCA

mRNA codon

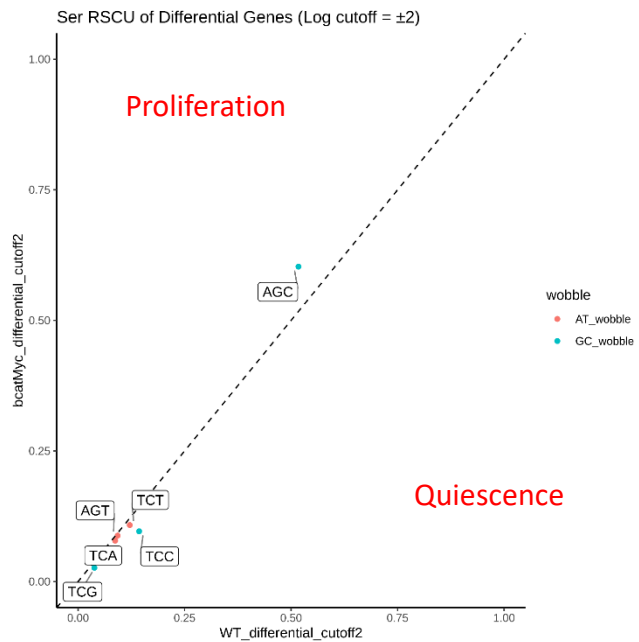
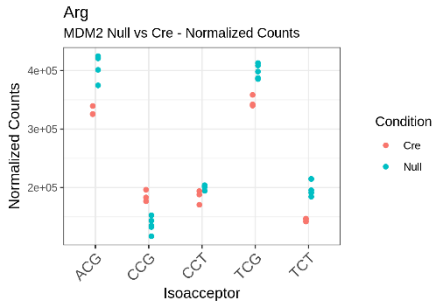
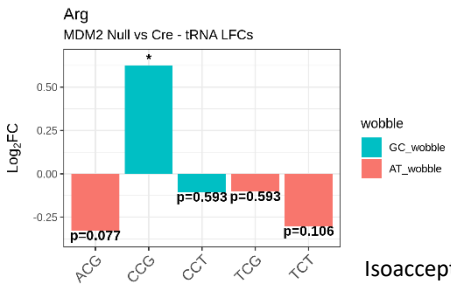


Figure 5.6 6-box tRNA differential expression analysis and relative synonymous codon usage in the βcatenin/c-Myc model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Arginine, (B) Leucine and (C) Serine. All statistically significant log fold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

6-box tRNAs – MDM2 Null vs Cre

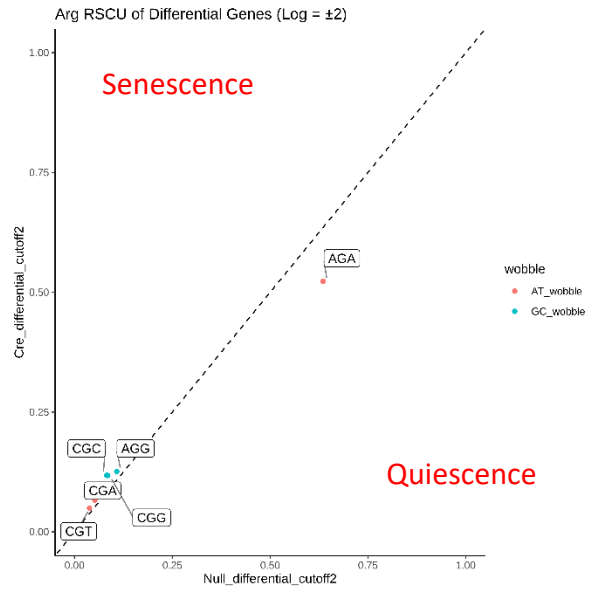
A



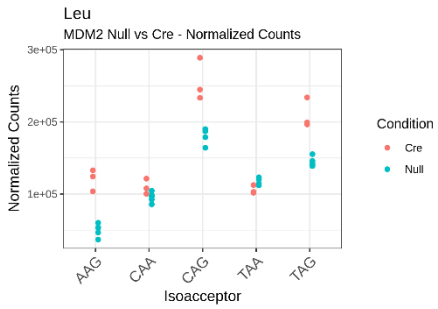


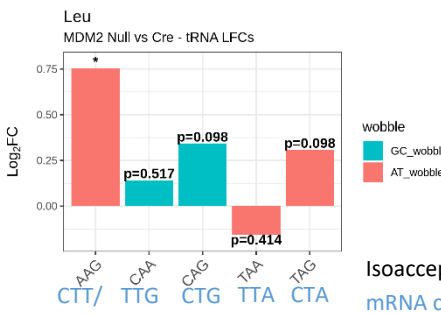
CGT/ CGC | CGG | AGG | CGA | AGA — Isoacceptor

mRNA codon



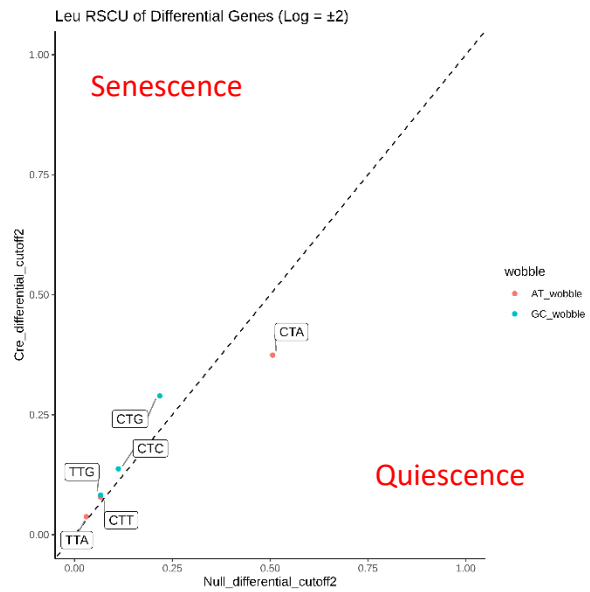
B



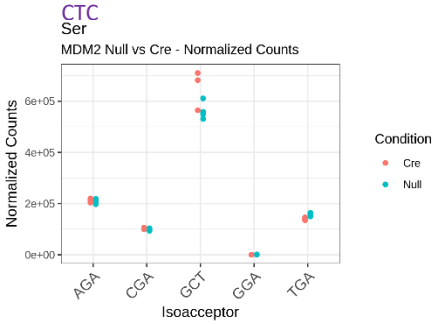


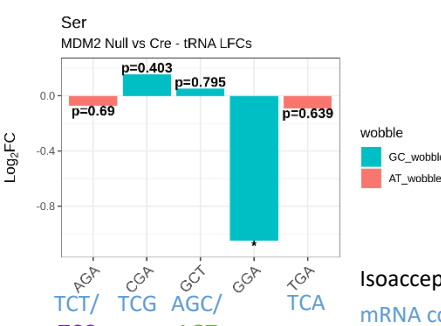
CTT/ CTC | TTG | CTG | TTA | CTA — Isoacceptor

mRNA codon



C





TCT/ TCC | TCG | AGC/ AGT | TCA — Isoacceptor

mRNA codon

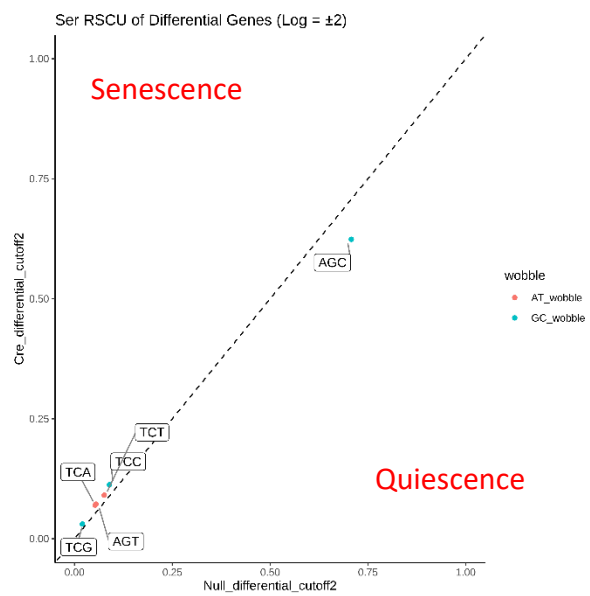


Figure 5.7 6-box tRNA differential expression analysis and relative synonymous codon usage in the MDM2 model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Arginine, (B) Leucine and (C) Serine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

5.5.2 Comparison of 4-box tRNAs to the relative synonymous codon usage in mRNA

4-box isoacceptors include tRNA-Ala, tRNA-Pro, tRNA-Val, tRNA-Gly and tRNA-Thr.

In the βcatenin/c-Myc model, where hepatocytes are being driven from quiescence to proliferation, all statistically significant tRNAs with an A/T at the wobble position were upregulated in proliferation (albeit being only 4 tRNAs). Also, their corresponding mRNA codons with an A/T at its wobble position were preferred in proliferation (Figure 5.8 A-E).

The tRNA-Ala and tRNA-Pro isoacceptor families had no statistically significant tRNA log fold changes. However the most preferred mRNA codon for alanine was GCA, decoded by both its cognate tRNA-Ala-TGC and tRNA-Ala-AGC via its A34I modification which was the most expressed tRNA-Ala isoacceptor (Figure 5.8 A).

For the proline mRNA codon CCA, its cognate tRNA-Pro-TGG was the most expressed tRNA in its isoacceptor family (Figure 5.8 B).

tRNA-Val-AAC was upregulated in in proliferation (which can decode its cognate GTT codon as well as GTC and GTA codons via its A34I modification). The mRNA codon GTA was preferred in proliferation and the GTC codon was preferred in quiescence. Both their cognate tRNAs (tRNA-Val-TAC and tRNA-Val-CAC respectively), were the highest expressed tRNAs in that isoacceptor family (Figure 5.8 C).

tRNA-Thr-AGT was upregulated in proliferation and can decode its cognate codon ACT as well as ACC and ACA via its A34I modification. The most preferred mRNA codon for threonine in both conditions was ACA. Although its cognate tRNA-Thr-TGT was upregulated, due to the low number of reads, it was ignored. However, the ACA codon could still be decoded by tRNA-Thr-AGT (Figure 5.8 E).

In the MDM2 model where hepatocytes were driven from quiescence to senescence, A/T ending mRNA codons were preferred in senescence, showing the opposite of that seen in the βcatenin/c-Myc model (Figure 5.9 A-E).

Both tRNA-Ala and tRNA-Pro had no statistically significant tRNA changes. However, in both isoacceptor families, the most abundant tRNA (tRNA-Ala-AGC and tRNA-Pro-TGG), coincided with the most preferred mRNA codons, GCA (Ala) and CCA (Pro) (Figure 5.9 A and Figure 5.9 B).

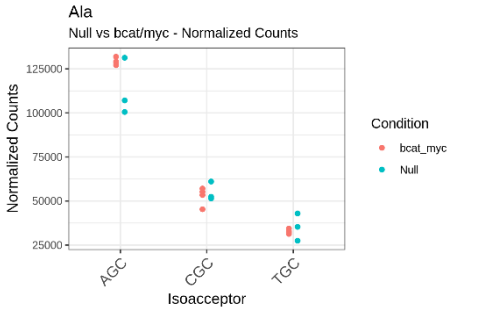
Valine codons GTA (preferred in quiescence, encoded by tRNA-Val-TAC) and GTG (preferred in senescence, encoded by tRNA-Val-CAC) were the most preferred mRNA codons for valine and their cognate tRNAs were the most abundant (Figure 5.9 C).

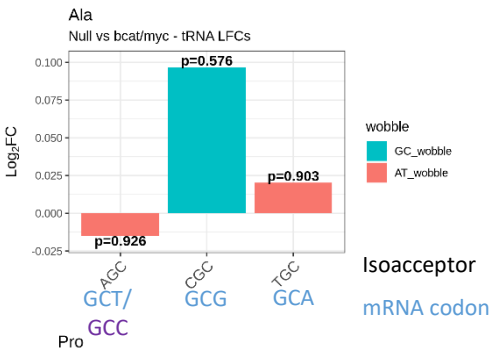
tRNA-Gly-GCC (which encodes its cognate GGC codon and GGT via G:T base pairing) was upregulated in senescence. The most preferred mRNA codon was GGA but changes in its cognate tRNA-Gly-TCC was not statistically significant but was abundant (Figure 5.9 E).

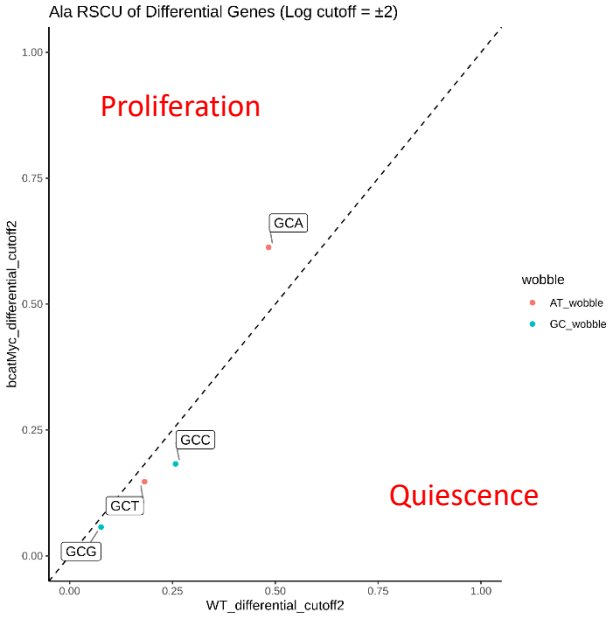
As observed in the βcatenin/c-Myc model, tRNA-Thr-TGT was upregulated (this time in senescence as opposed to proliferation). However, it holds no significance due to its low read counts.

4-box tRNAs – β-catenin/c-Myc Null vs Cre

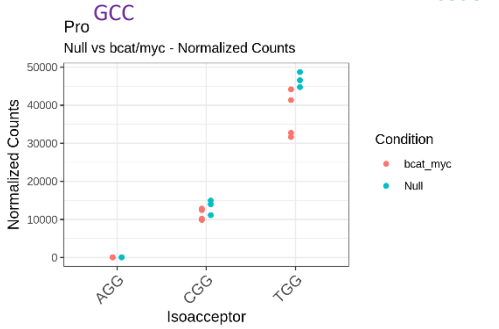
A

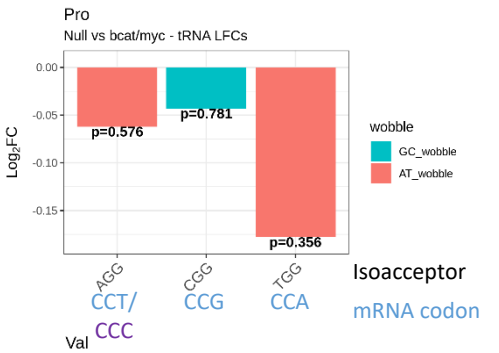


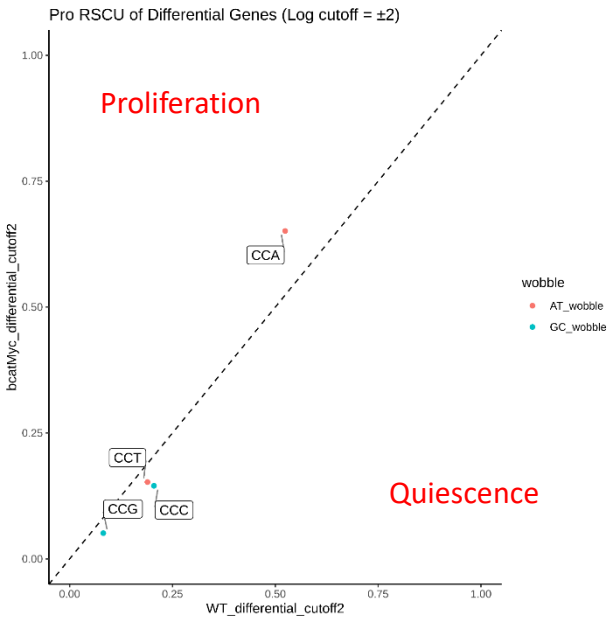




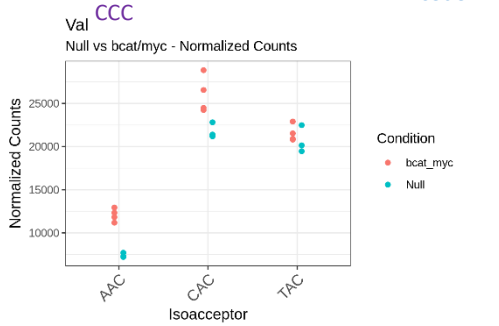
B

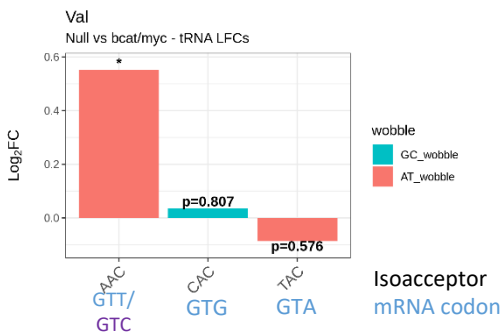


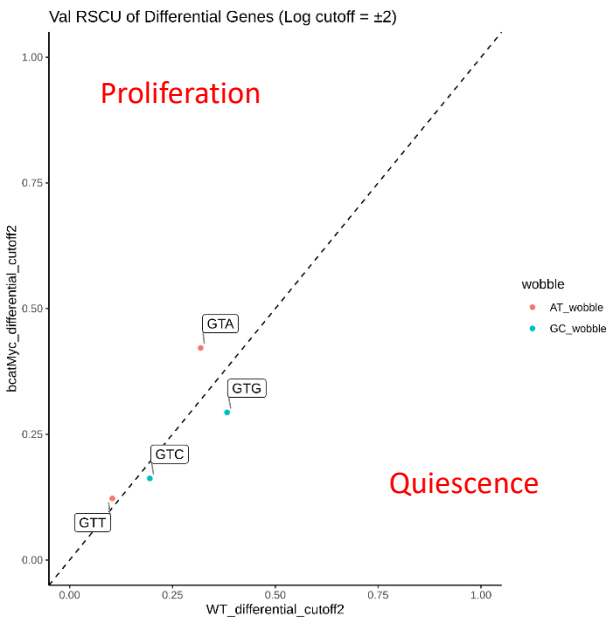




C







4-box tRNAs – β-catenin/c-Myc Null vs Cre

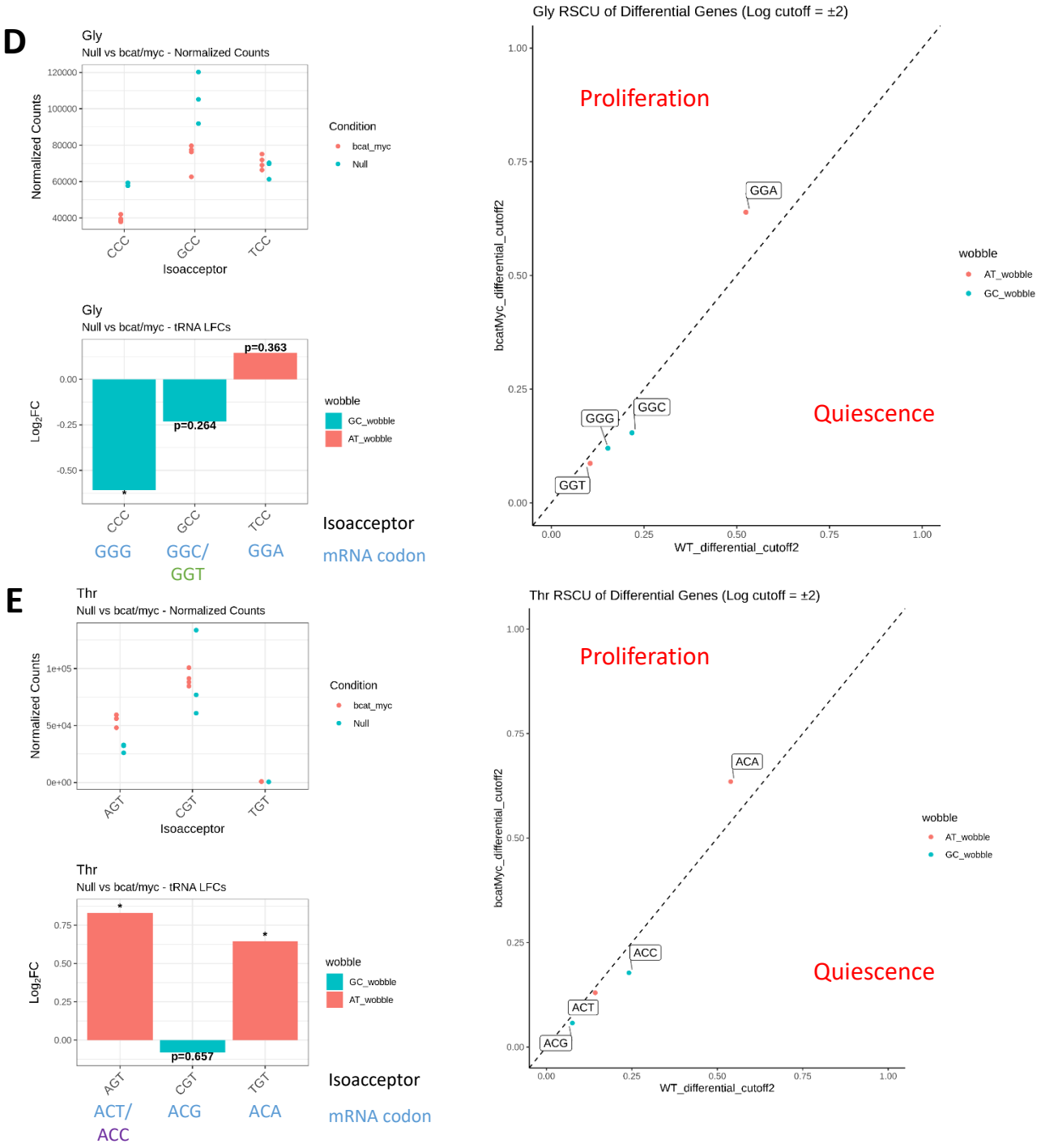
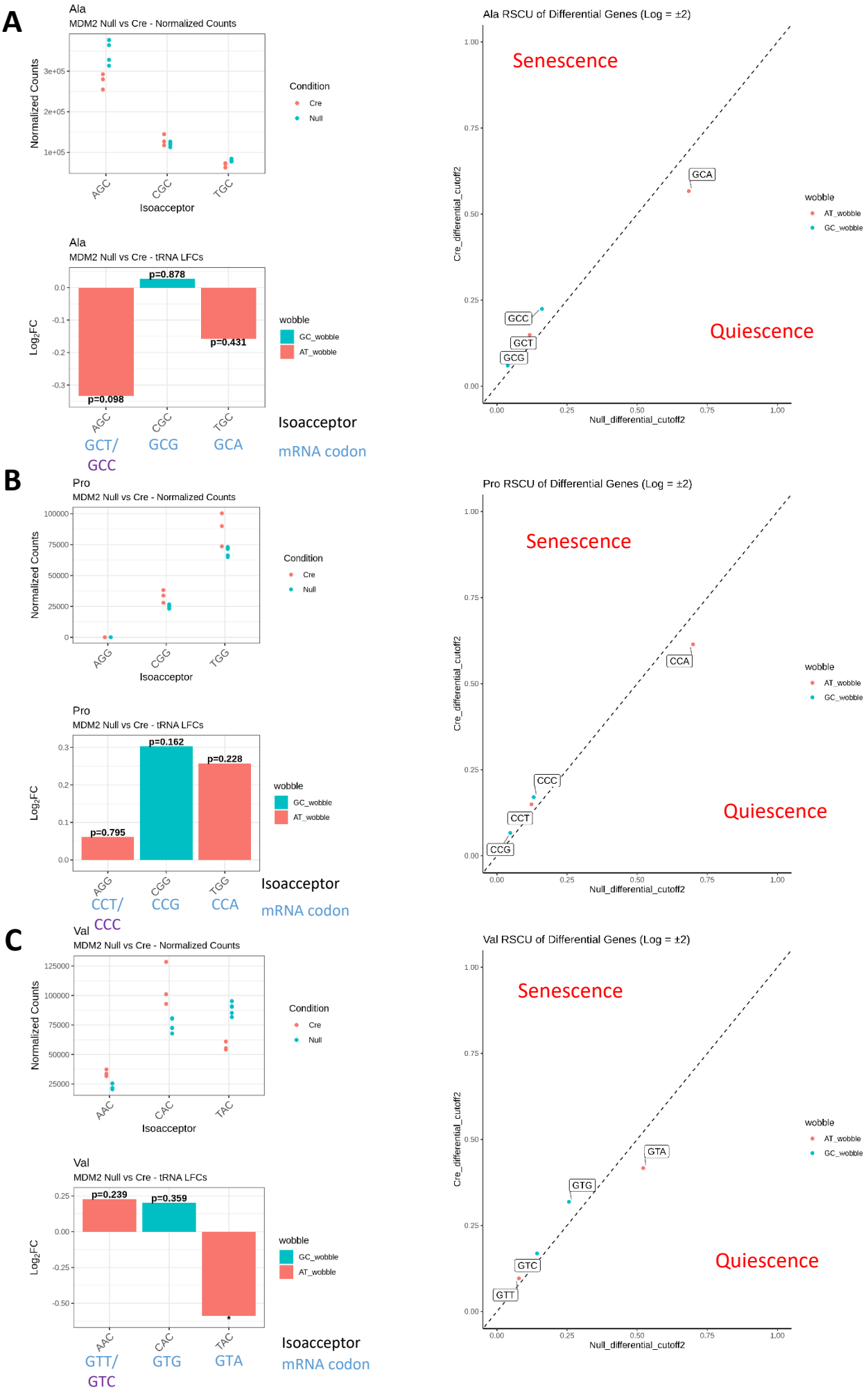


Figure 5.8 4-box tRNA differential expression analysis and relative synonymous codon usage in the βcatenin/c-Myc model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Alanine, (B) Proline, (C) Valine, (D) Glycine and (E) Threonine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

4-box tRNAs – MDM2 Null vs Cre



4-box tRNAs – MDM2 Null vs Cre

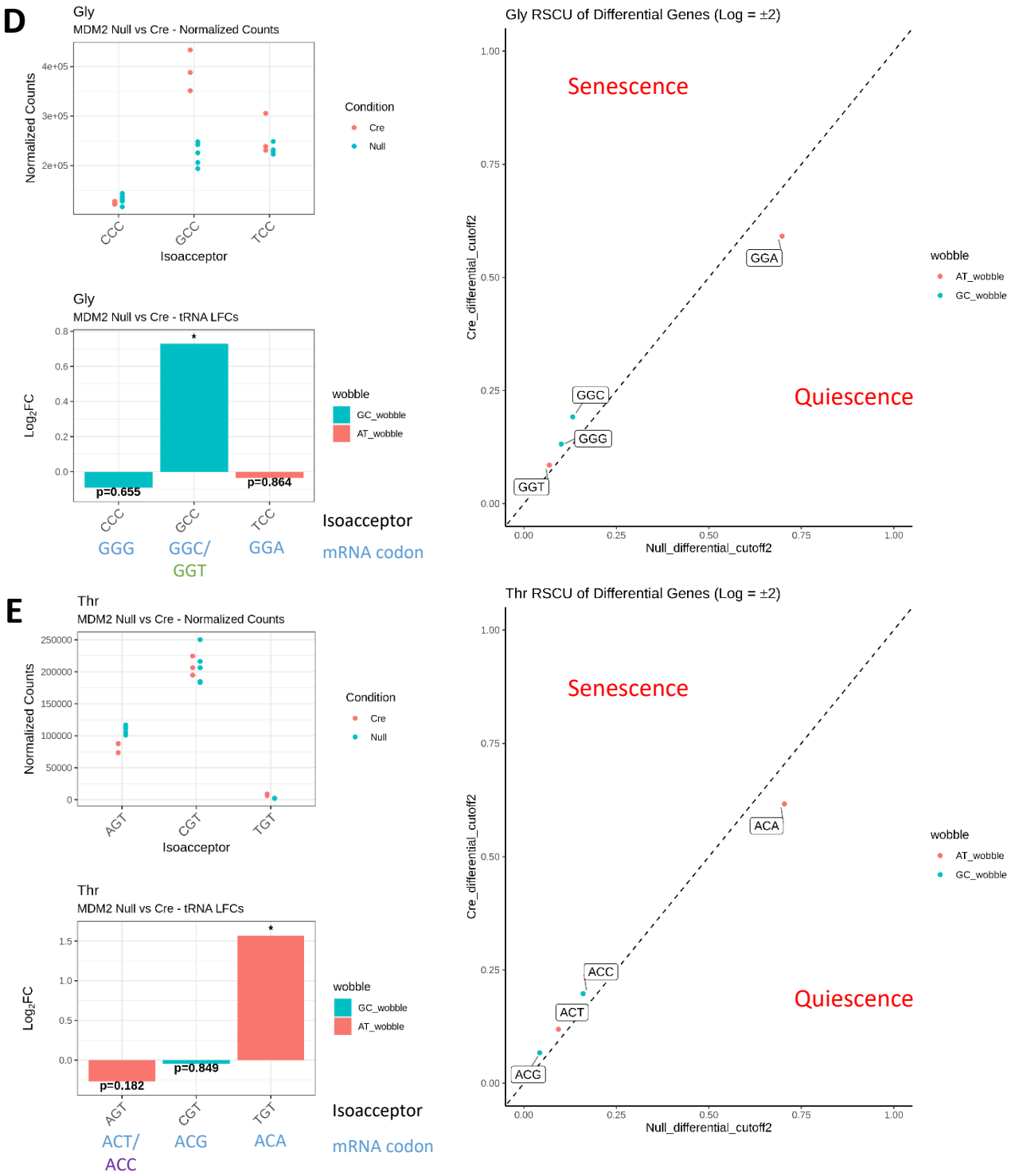


Figure 5.9 4-box tRNA differential expression analysis and relative synonymous codon usage in the MDM2 model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log_2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Alanine, (B) Proline, (C) Valine, (D) Glycine and (E) Threonine. All statistically significant logfold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

5.5.3 Comparison of 2-box tRNAs to the relative synonymous codon usage in mRNA

2-box isoacceptors include tRNA-Glu, tRNA-Gln, tRNA-Lys and tRNA-Ile.

For all the 2-box tRNAs in the βcatenin/c-Myc model, all mRNAs with an A at the wobble position were the most preferred codon (except in the case for tRNA-Ile) (Figure 5.10 A-D).

tRNA-Glu-TTC was down regulated in proliferation even though its cognate mRNA codon GAA was the most preferred glutamic acid codon (Figure 5.10 A).

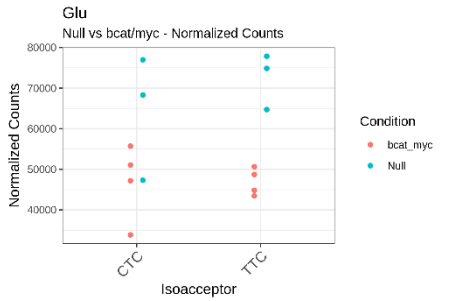
tRNA-Lys-TTT was upregulated in proliferation and its cognate mRNA codon AAA was the most preferred lysine codon (Figure 5.10 C).

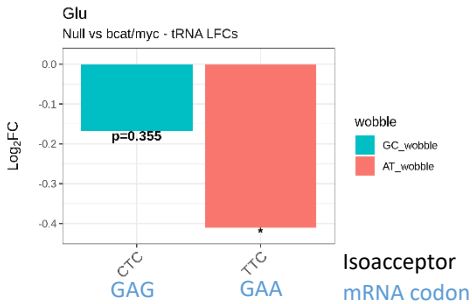
tRNA-Ile-GAT is not expressed in mice. However, tRNA-Ile-AAT via its A34I modification is able to decode mRNA codon ATC, as well as its cognate ATT (which are also the most preferred codons) (Figure 5.10 D).

In the MDM2 model, as in the βcatenin/c-Myc model, all mRNAs with an A at the wobble position were the most preferred codon. However, A ending codons were preferred in quiescence and G/C codons were preferred in senescence (Figure 5.11 A-D).

2-box tRNAs – β-catenin/c-Myc Null vs Cre

A

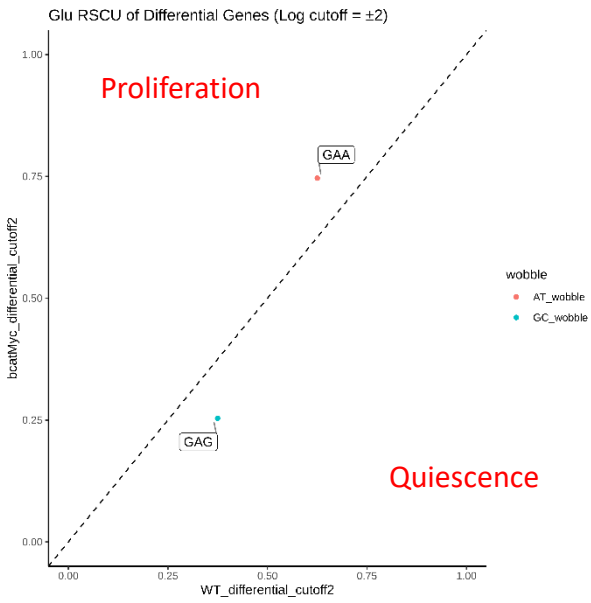




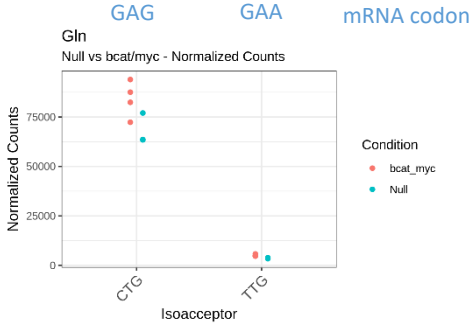
Isoacceptor

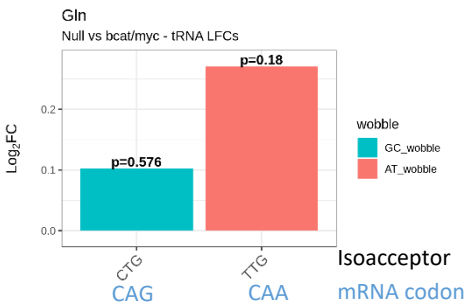
GAG GAA

mRNA codon



B

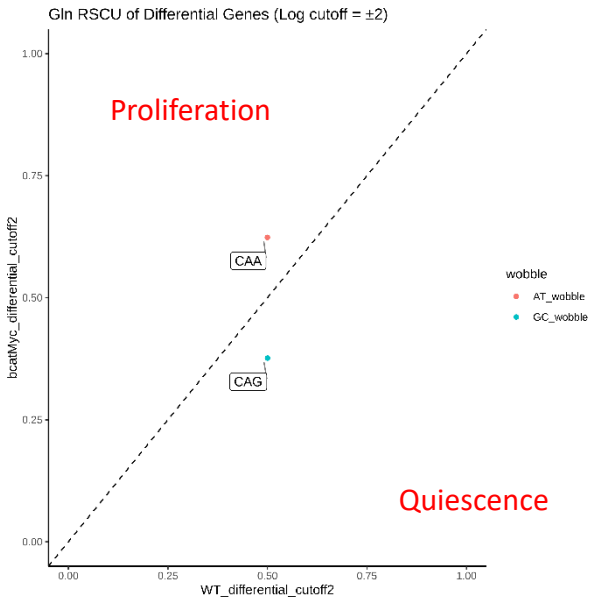




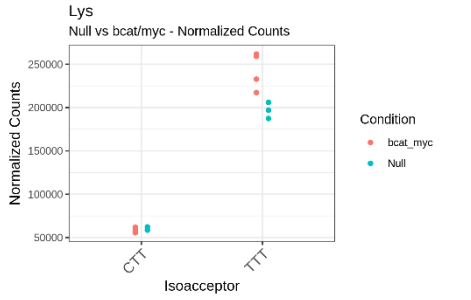
Isoacceptor

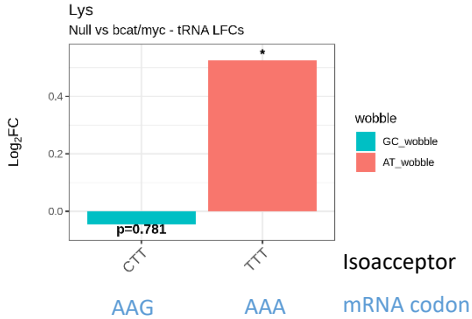
CAG CAA

mRNA codon



C

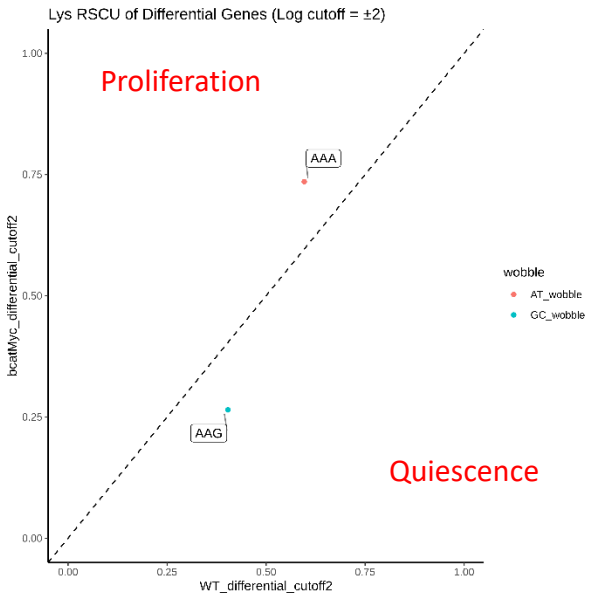




Isoacceptor

AAG AAA

mRNA codon



2-box tRNAs – β-catenin/c-Myc Null vs Cre

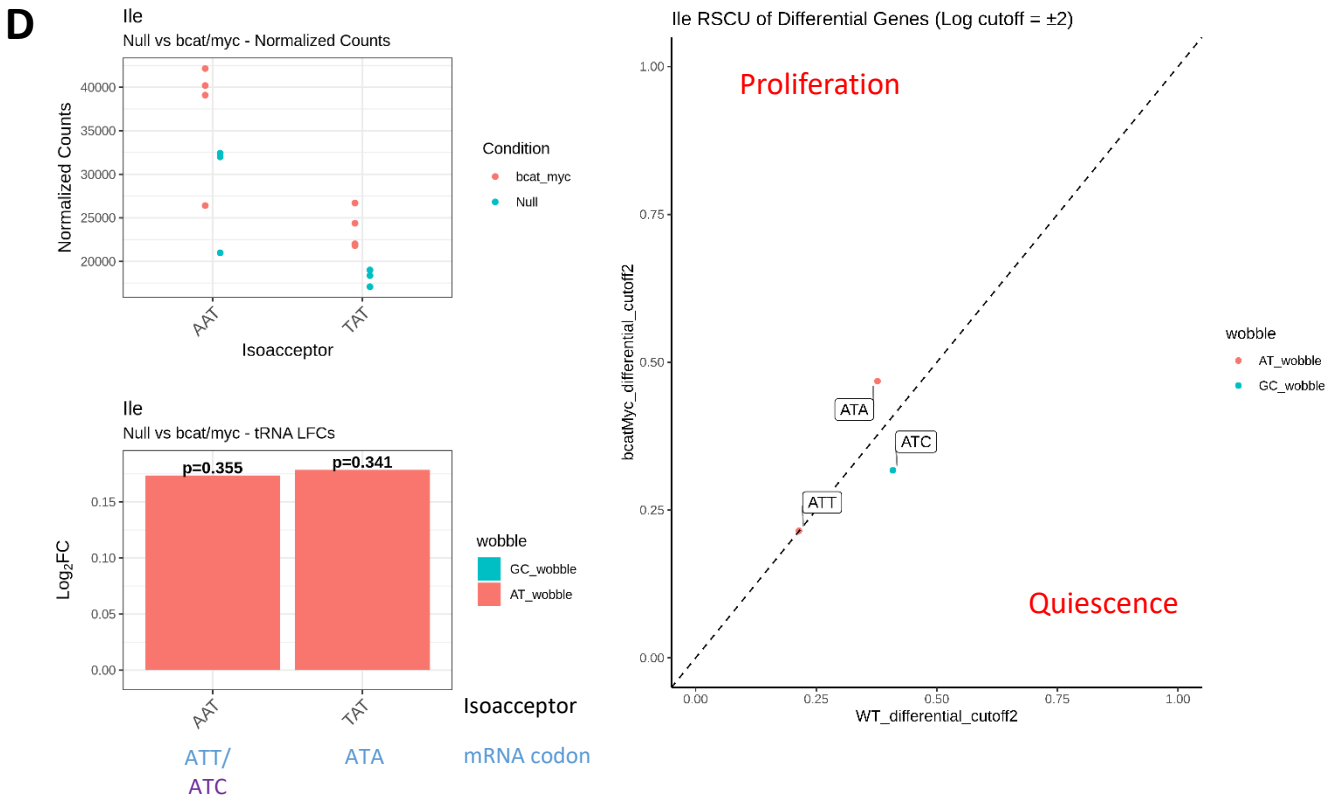
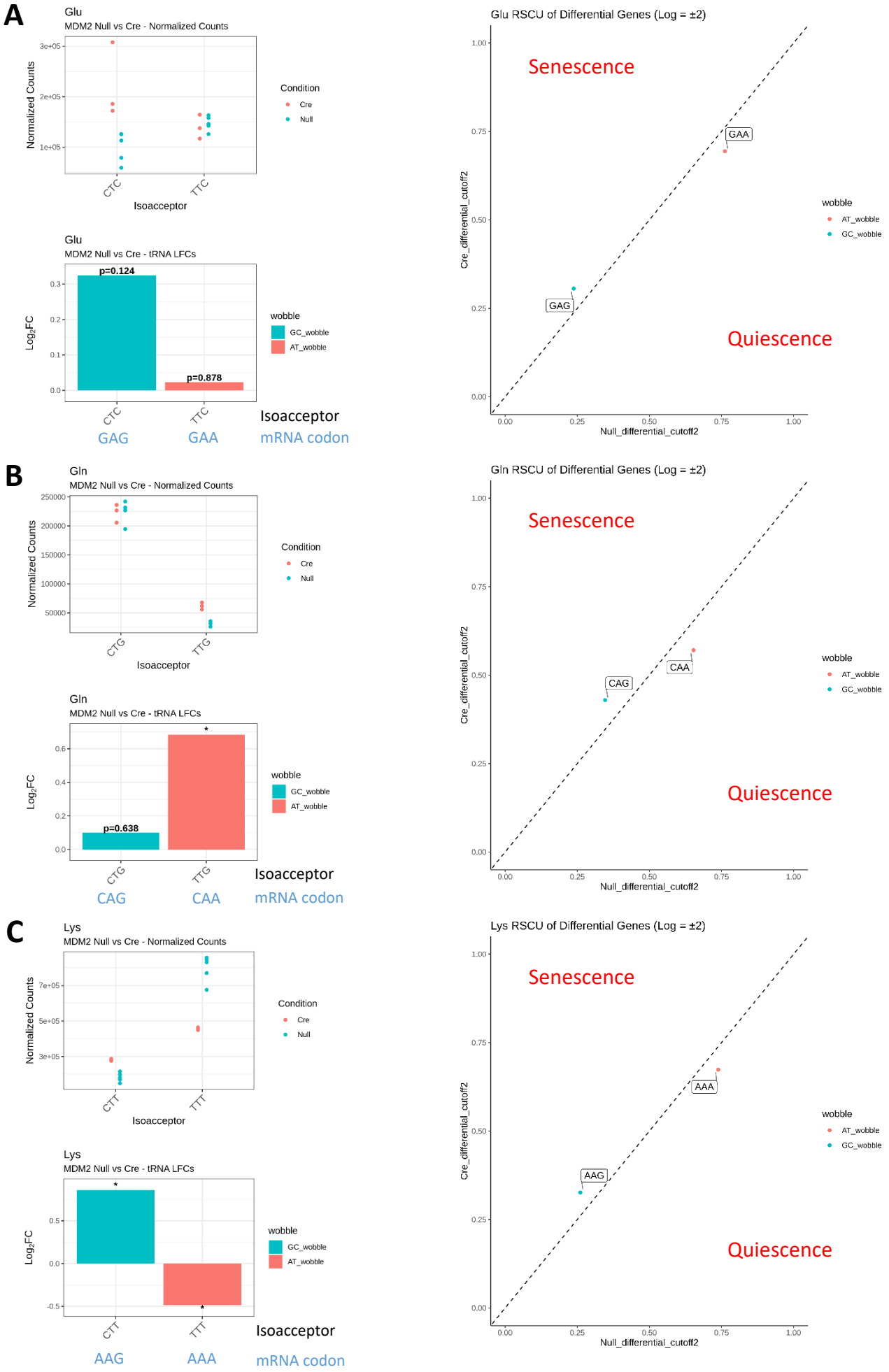


Figure 5.10 2-box tRNA differential expression analysis and relative synonymous codon usage in the βcatenin/c-Myc model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Glutamic Acid, (B) Glutamine, (C) Lysine and (D) Isoleucine. All statistically significant log fold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

2-box tRNAs – MDM2 Null vs Cre



2-box tRNAs – MDM2 Null vs Cre

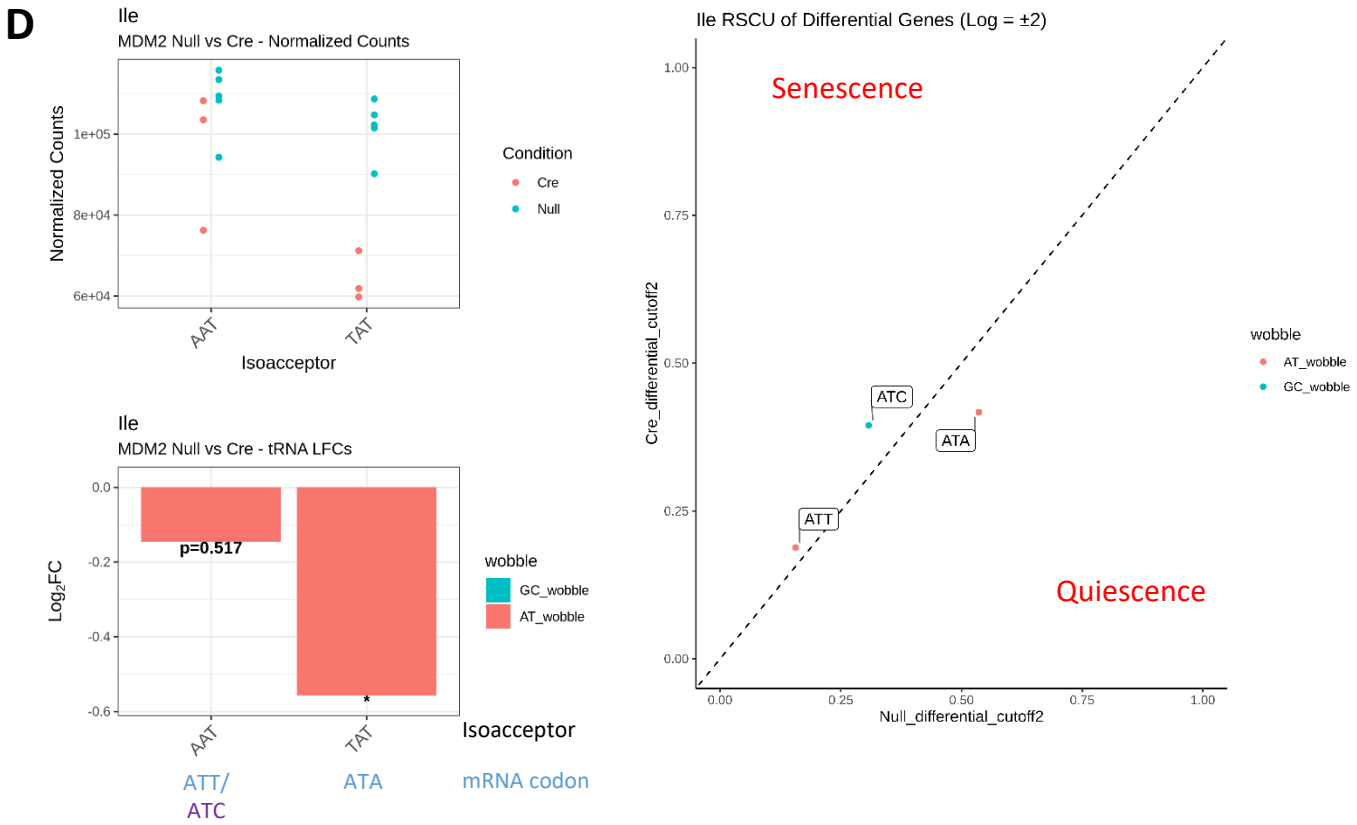


Figure 5.11 2-box tRNA differential expression analysis and relative synonymous codon usage in the MDM2 model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Glutamic Acid, (B) Glutamine, (C) Lysine and (D) Isoleucine. All statistically significant log fold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

5.5.4 Comparison of 2-box tRNAs (only 1 tRNA expressed) to the relative synonymous codon usage in mRNA

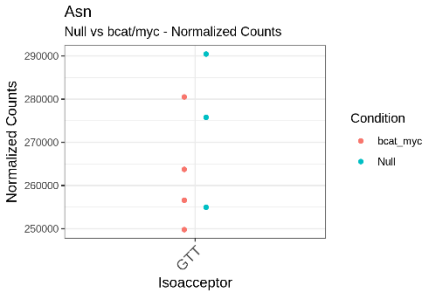
2-box isoacceptors that express only one anticodon are tRNA-Asn, tRNA-His, tRNA-Phe, tRNA-Asp, tRNA-Cys and tRNA-Tyr. All of these isoacceptors utilise G:T base pairing to decode the non-cognate codon.

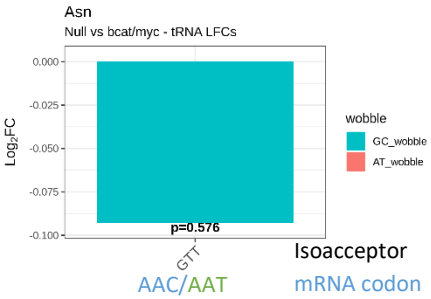
Interestingly, in the βcatenin/c-Myc model, all the mRNA codons decoded by this category of 2-box tRNAs either had a C or T codon at the mRNA wobble position. However, all C ending mRNAs were the preferred codon for each of their amino acids in comparison to the T ending mRNAs. Furthermore, the C ending codons were preferred in proliferation over quiescence. This suggests that although G:T base pairing can be utilised to decode all T ending mRNA codons in this category, the translational machinery prefers cognate base pairing (Figure 5.12 A-F).

In the MDM2 model, the same phenomenon was observed, except all C ending mRNA codons were preferred in quiescence than in senescence. Therefore, where the A34I modification seen in previous sections may be compensating for downregulations in other tRNAs, here the data suggests that G:T base pairing is not the preferred mechanism for decoding mRNA codons.

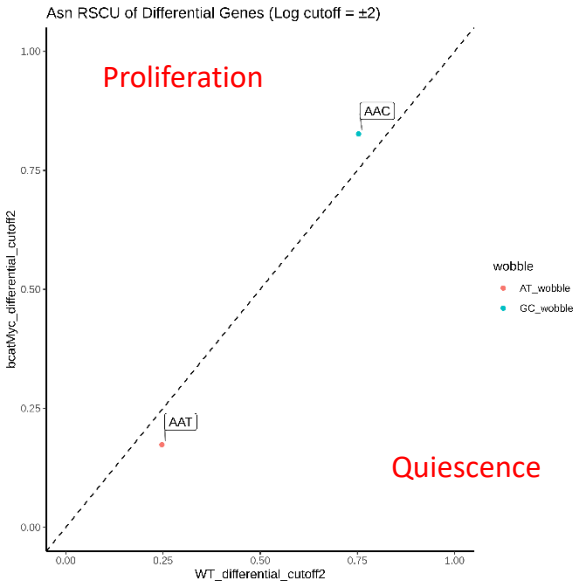
2-box tRNAs (only 1 tRNA expressed) – β-catenin/c-Myc Null vs Cre

A

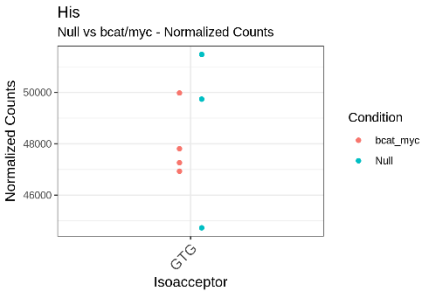


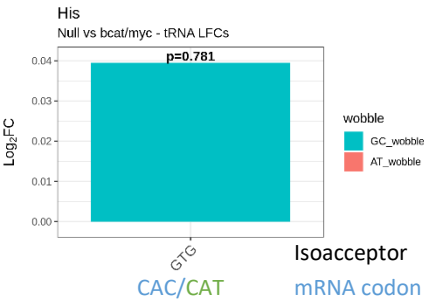


AAC/AAT mRNA codon

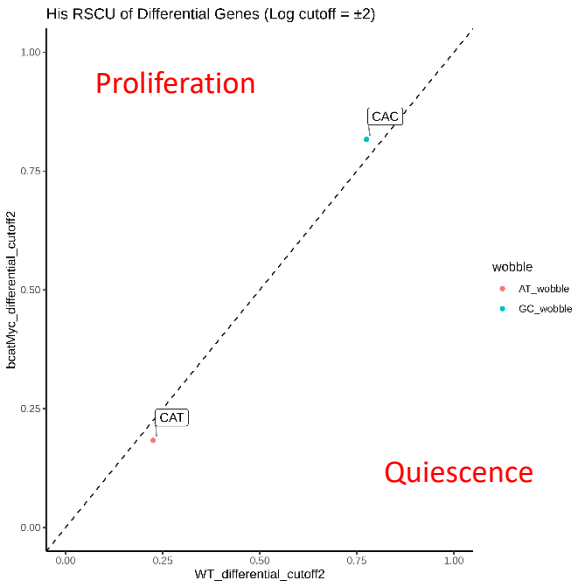


B

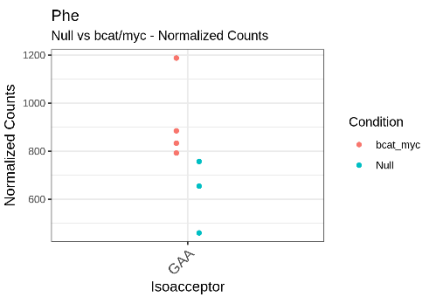


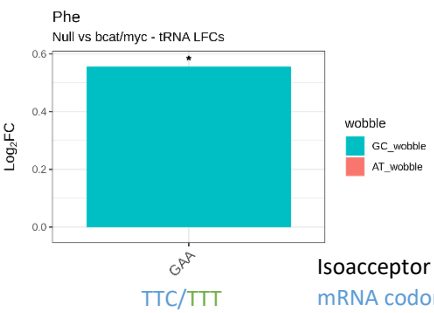


CAC/CAT mRNA codon

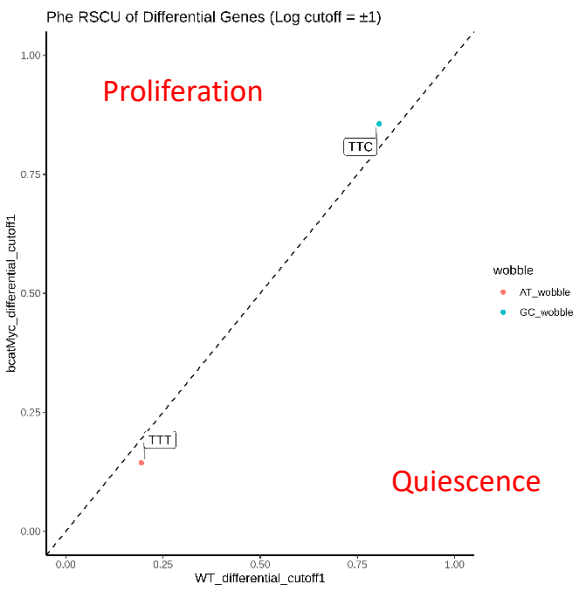


C



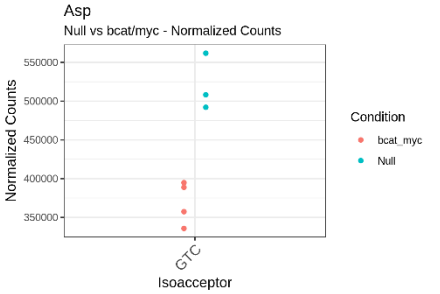


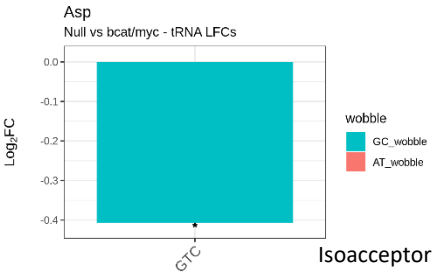
TTC/TTT mRNA codon



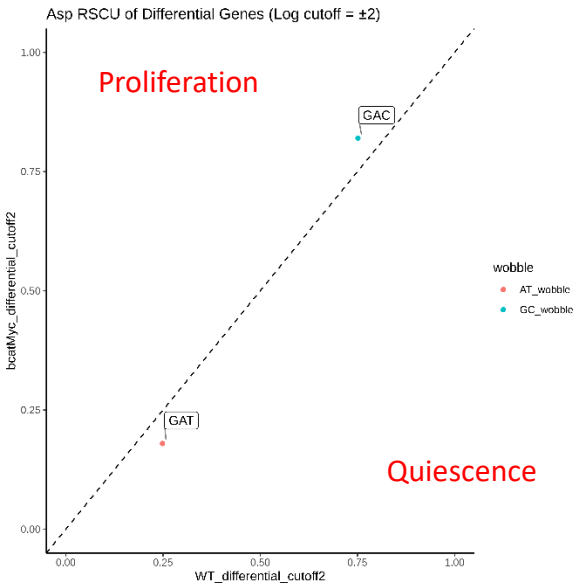
2-box tRNAs (only 1 tRNA expressed) – β-catenin/c-Myc Null vs Cre

D

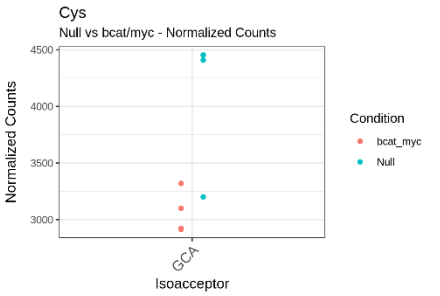


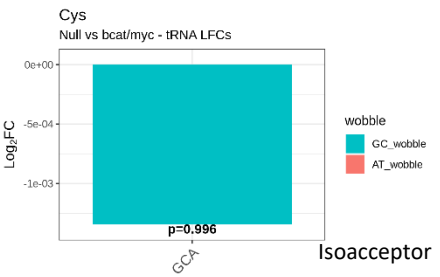


GAC/GAT mRNA codon

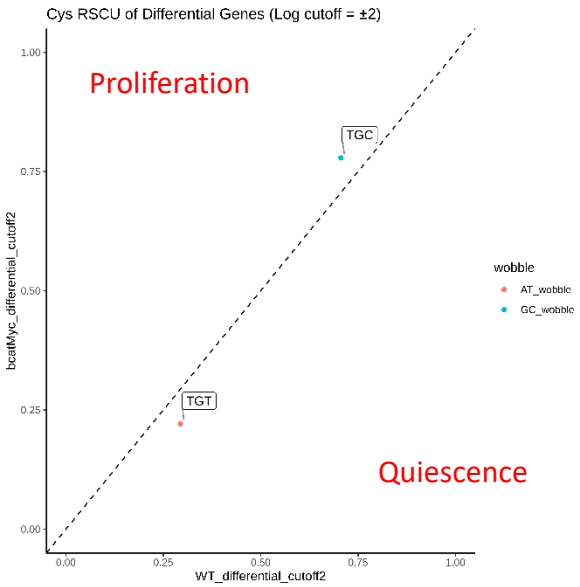


E

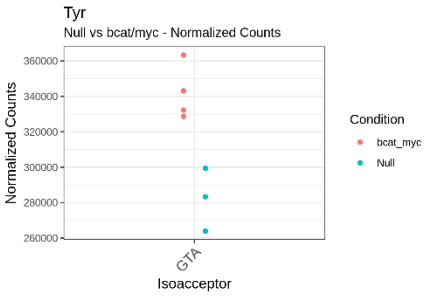


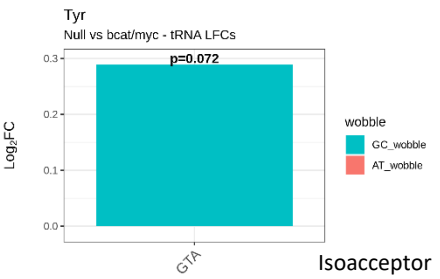


TGC/TGT mRNA codon



F





TAC/TAT mRNA codon

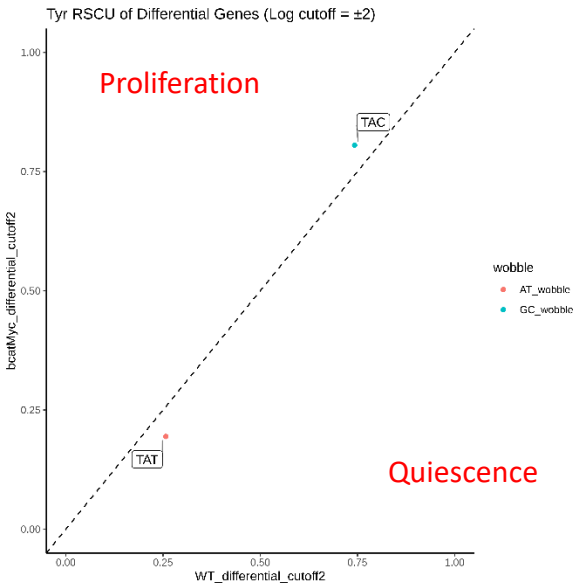
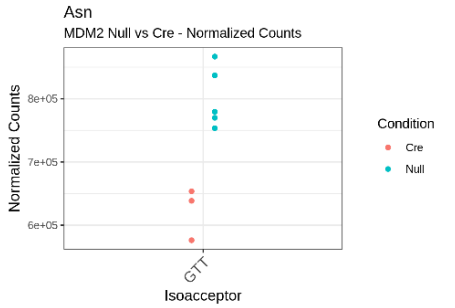
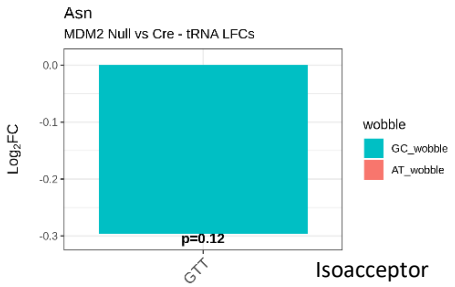


Figure 5.12 2-box tRNA (only 1 tRNA expressed) differential expression analysis in the βcatenin/c-Myc model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Asparagine, (B) Histidine, (C) Phenylalanine, (D) Aspartate, (E) Cysteine and (F) Tyrosine. All statistically significant log fold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

2-box tRNAs (only 1 tRNA expressed) – MDM2 Null vs Cre

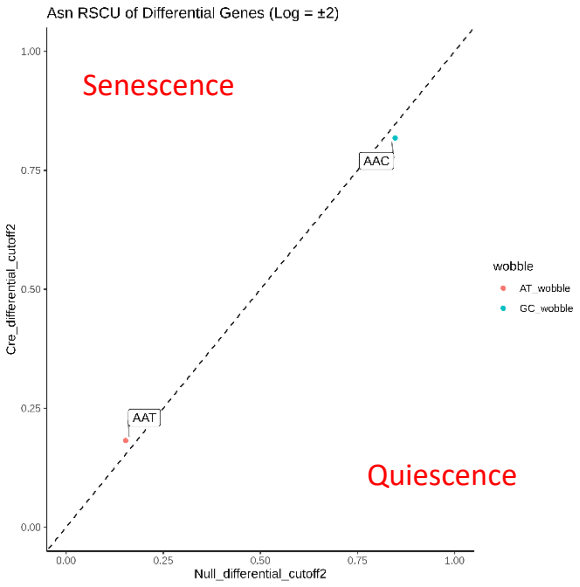
A



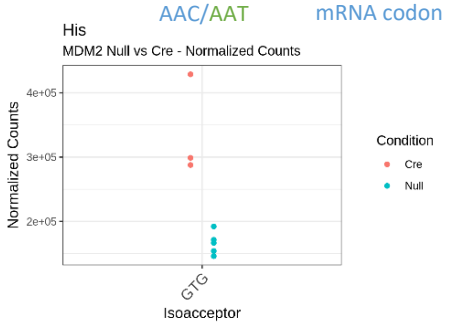


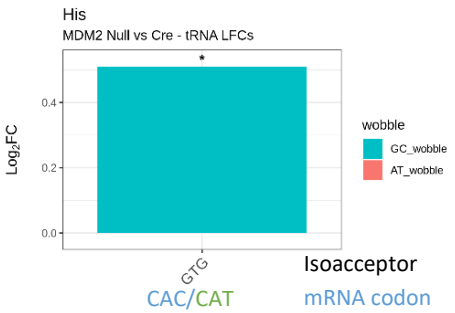
Isoacceptor

AAC/AAT mRNA codon



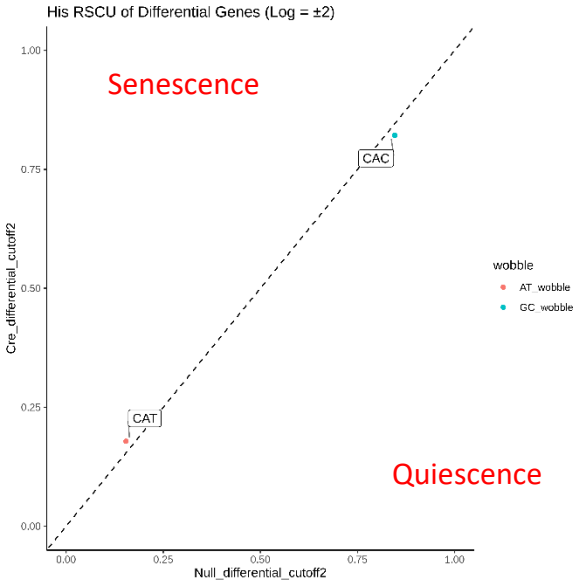
B



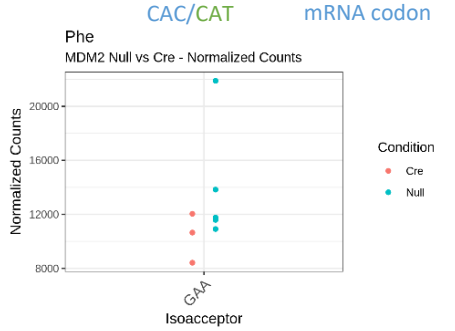


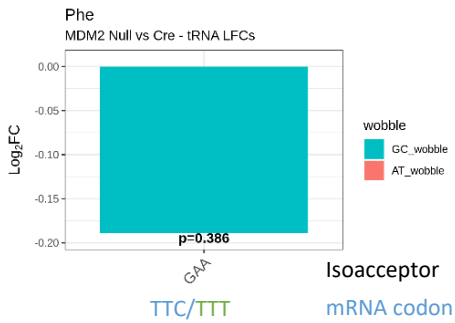
Isoacceptor

CAC/CAT mRNA codon



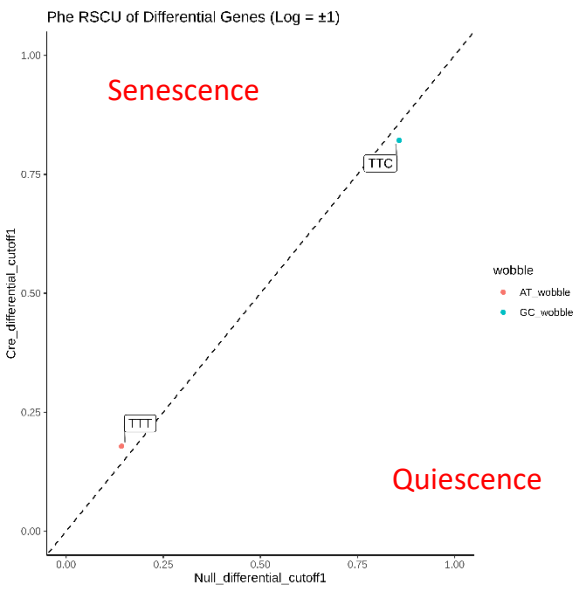
C





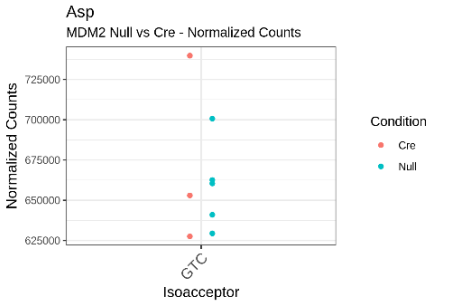
Isoacceptor

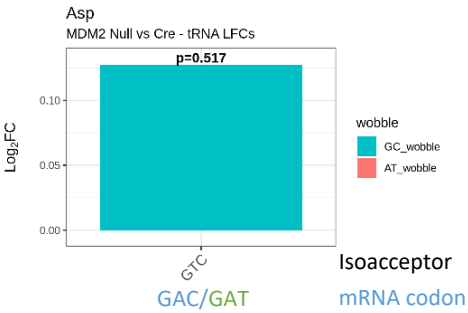
TTC/TTT mRNA codon



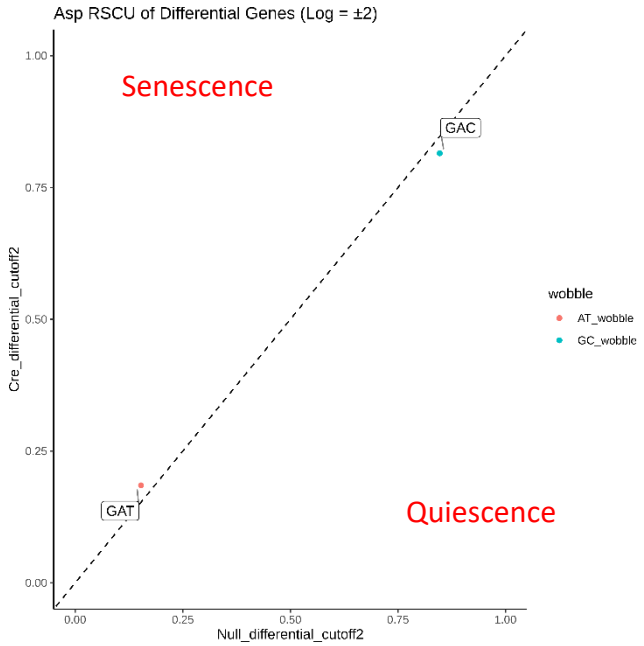
2-box tRNAs (only 1 tRNA expressed) – MDM2 Null vs Cre

D

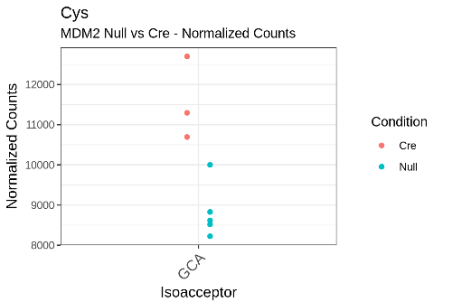


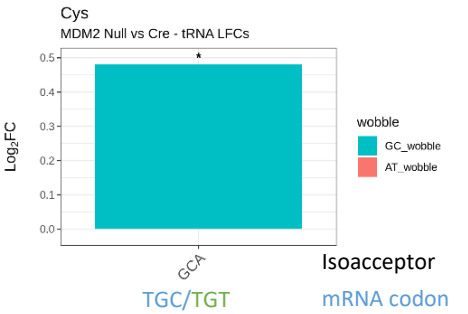


GAC/GAT mRNA codon

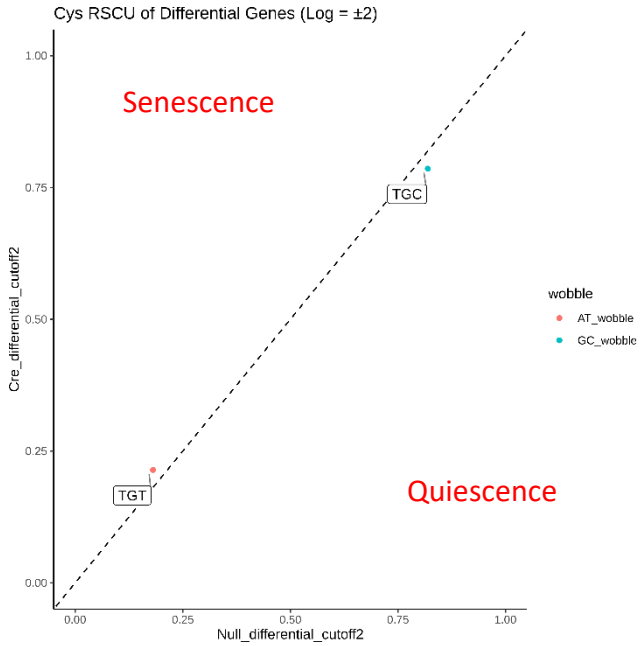


E

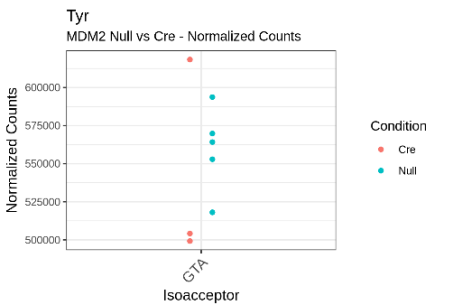


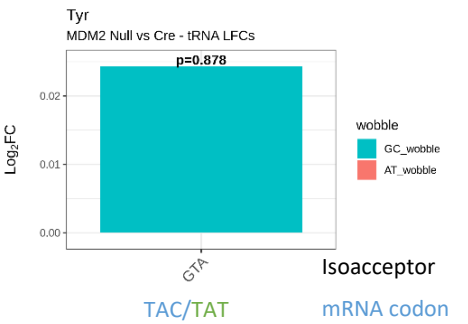


TGC/TGT mRNA codon



F





TAC/TAT mRNA codon

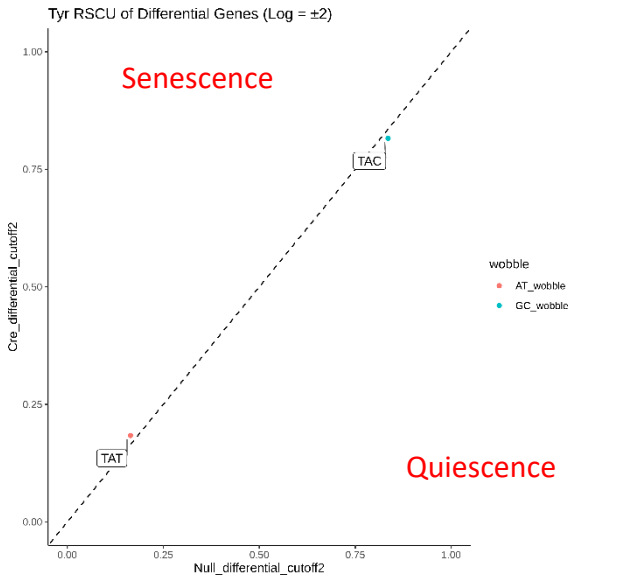


Figure 5.13 2-box tRNA (only 1 tRNA expressed) differential expression analysis in the MDM2 model. tRNA differential expression analysis was broken down into amino acid and box-type. Transcriptome differential expression analysis was filtered for transcripts that were statistically significant and had a Log2 fold change of 2. RSCU was calculated and normalized to transcript abundance. Both differential isoacceptor changes and RSCU was plotted for (A) Asparagine, (B) Histidine, (C) Phenylalanine, (D) Aspartate, (E) Cysteine and (F) Tyrosine. All statistically significant log fold changes are highlighted with an asterisk (*) and all non-significant changes have their adjusted p-vales printed. mRNA codons encoded by the isoacceptor are printed below with complementary base pairing highlighted in blue, non-cognate base pairing due to inosine modification highlighted in purple and G:T base pairing highlighted in green. * $p < 0.05$

5.5.5 Global tRNA isodecoder changes

Overall, for both the βcatenin/c-Myc and MDM2 models, no correlations between the nucleotide at the tRNA wobble position and the cell state was found. However, it was clear in many of the tRNAs in the differential expression analysis that the most preferred mRNA codon generally coincided with the most abundant tRNA.

In the βcatenin/c-Myc model, the majority of mRNA codons with an A at the wobble position were preferred in proliferation. However, tRNA anticodons with a T at the wobble position were mostly downregulated (Figure 5.14 A). However, it was clear from the differential expression analysis that many tRNAs with the A34I modification could be working to compensate these downregulations. All tRNAs with a C at the anticodon wobble position we all downregulated (Figure 5.14 B).

In the MDM2 model, there was also no clear correlation between the nucleotide at the tRNA wobble position and cell state (Figure 5.15 A-B).

A

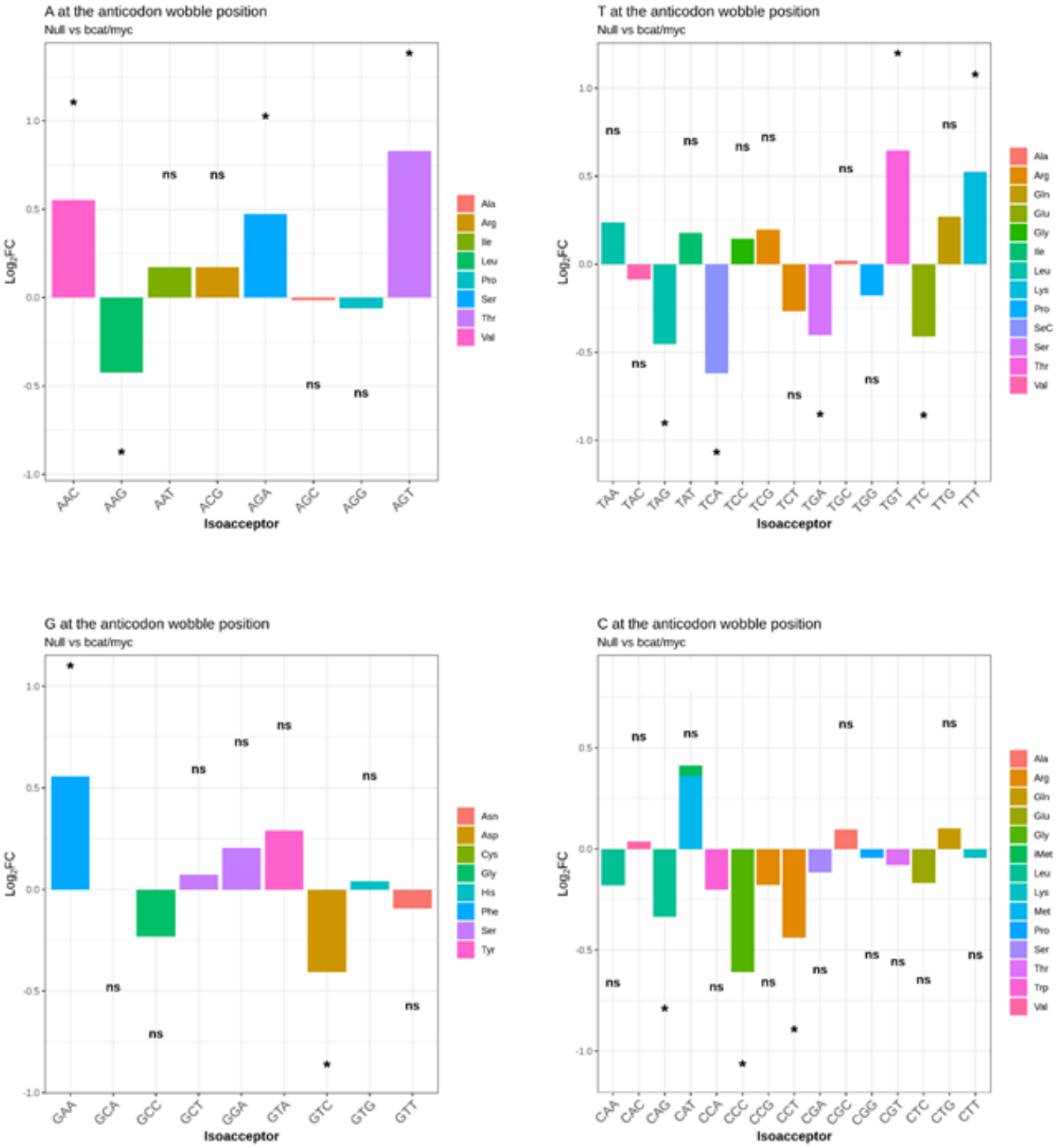


Figure 5.14 Global isoacceptor log fold changes in the βcatenin/c-myc model. (A) Log fold changes for tRNA isoacceptors with an A or T at the wobble position. (B) Log fold changes for tRNA isoacceptors with a G or C at the wobble position. * $p < 0.05$. ns = not significant, Log_2(Cre/Null).

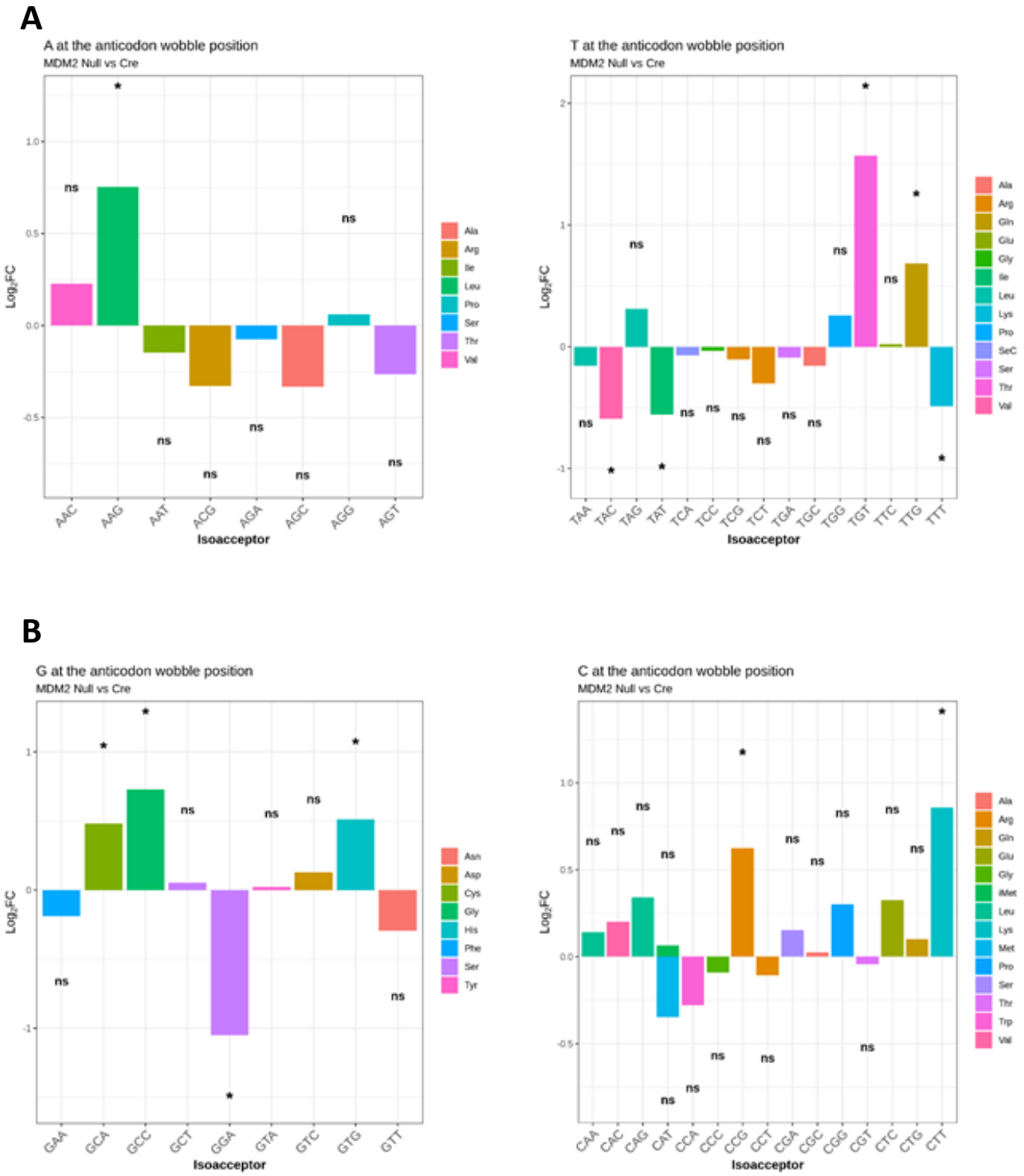


Figure 5.15 Global isoacceptor log fold changes in the MDM2 model. (A) Log fold changes for tRNA isoacceptors with an A or T at the wobble position. (B) Log fold changes for tRNA isoacceptors with a G or C at the wobble position. * $p < 0.05$. ns = not significant, Log_2(Cre/Null).

5.6 Global relative synonymous codon usage analysis

The relative synonymous codon usage for all statistically significant, differentially expressed transcripts with a log2 fold change of ±2 were calculated (all of the individual plots combined) (Figure 5.16 A-B). This was done to highlight the global codon preference at the wobble position for the transcriptome, depending on the cell state.

In the βcatenin/c-Myc model, mRNA codons with either an A or C base at the wobble position were preferred in proliferation and mRNA codons with either a G or T bases were preferred in quiescence (Figure 5.16 A). It was clear from the tRNA differential expression analysis that all of the C ending mRNA codons were decoded by 2-box tRNAs that only express 1 tRNA.

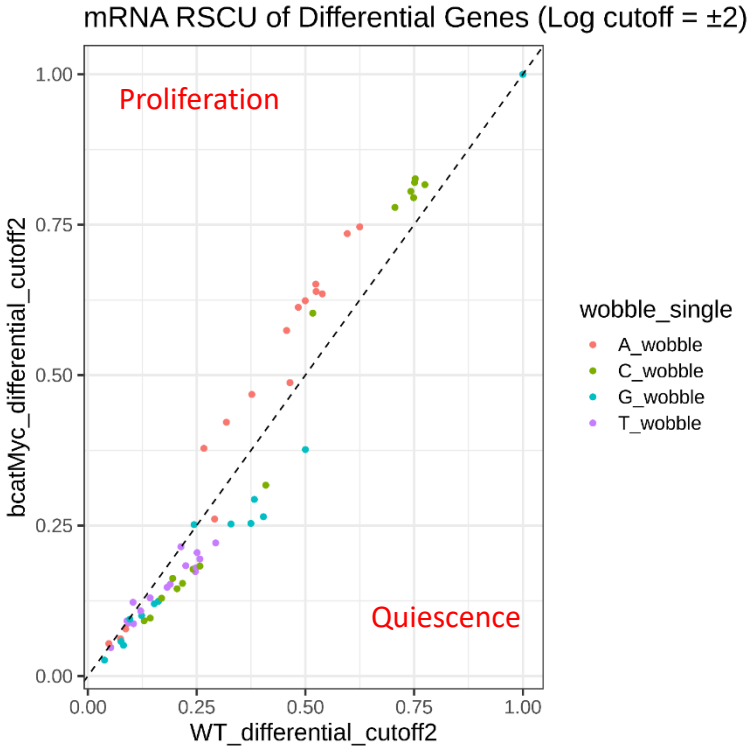
In the MDM2 model, mRNA codons with either an A or C base at the wobble position were preferred in quiescence and mRNA codons with either a G or T base being preferred in senescence (Figure 5.16 B).

When the relative synonymous codon usage was calculated for all transcripts and the most abundant transcript per gene (weighted by transcript abundance, TPM), slight preferences could be observed. In the βcatenin/c-Myc model, there was a slight preference for A/C ending mRNA codons in proliferation and G/T ending mRNA codons in quiescence for all transcripts and the most abundant transcript per gene (Figure 5.17 A).

Interestingly, in the MDM2 model, for all transcripts and the most abundant transcript per gene, there was a distinct preference for A/C ending mRNA codons in senescence and G/T ending mRNA codons in quiescence, opposite of that observed for the differential transcripts (Figure 5.17 B).

This might suggest that in the oncogene induced proliferation model (βcatenin/c-Myc), cells would prefer A ending mRNA codons in order to utilise the A34I modification, and therefore tRNA degeneracy, to try and possibly increase translational efficiency, since the volume of translational events would be much larger in comparison to quiescence. In the oncogene induced senescence model (MDM2), there may be a mechanism regulating the relative synonymous codon usage of differential genes since the difference between them and the transcriptome is so extreme.

A



B

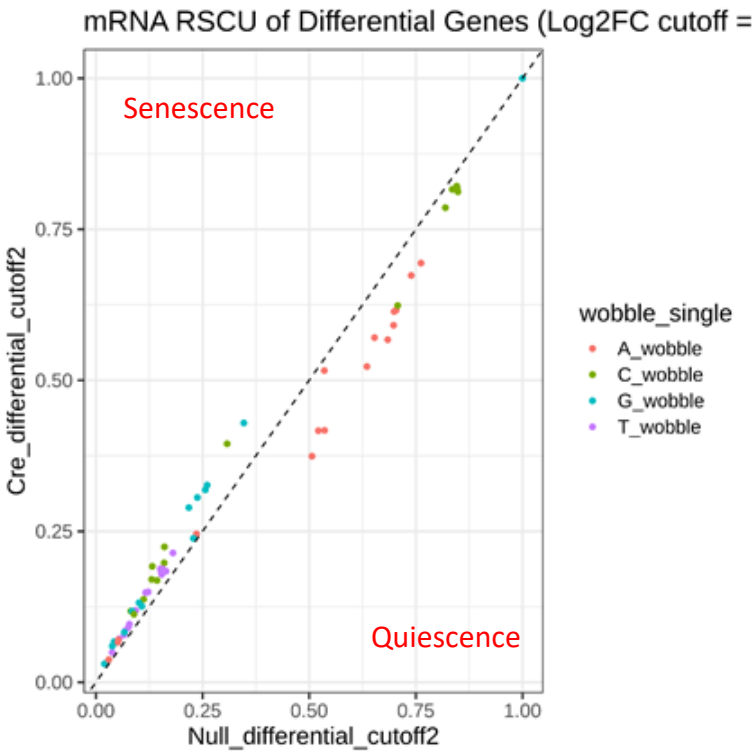


Figure 5.16 Relative synonymous codon usage for statistically significant transcripts with a ±Log$_2$ fold change of 2 in the βcatenin/c-myc and MDM2 models. (A) The βcatenin/c-myc shows a preference for A and C in proliferation and G/C/T in quiescence at the mRNA wobble position. (B) The MDM2 model shows a preference for A and C codons in quiescence and G/C/T in senescence at the mRNA wobble position.

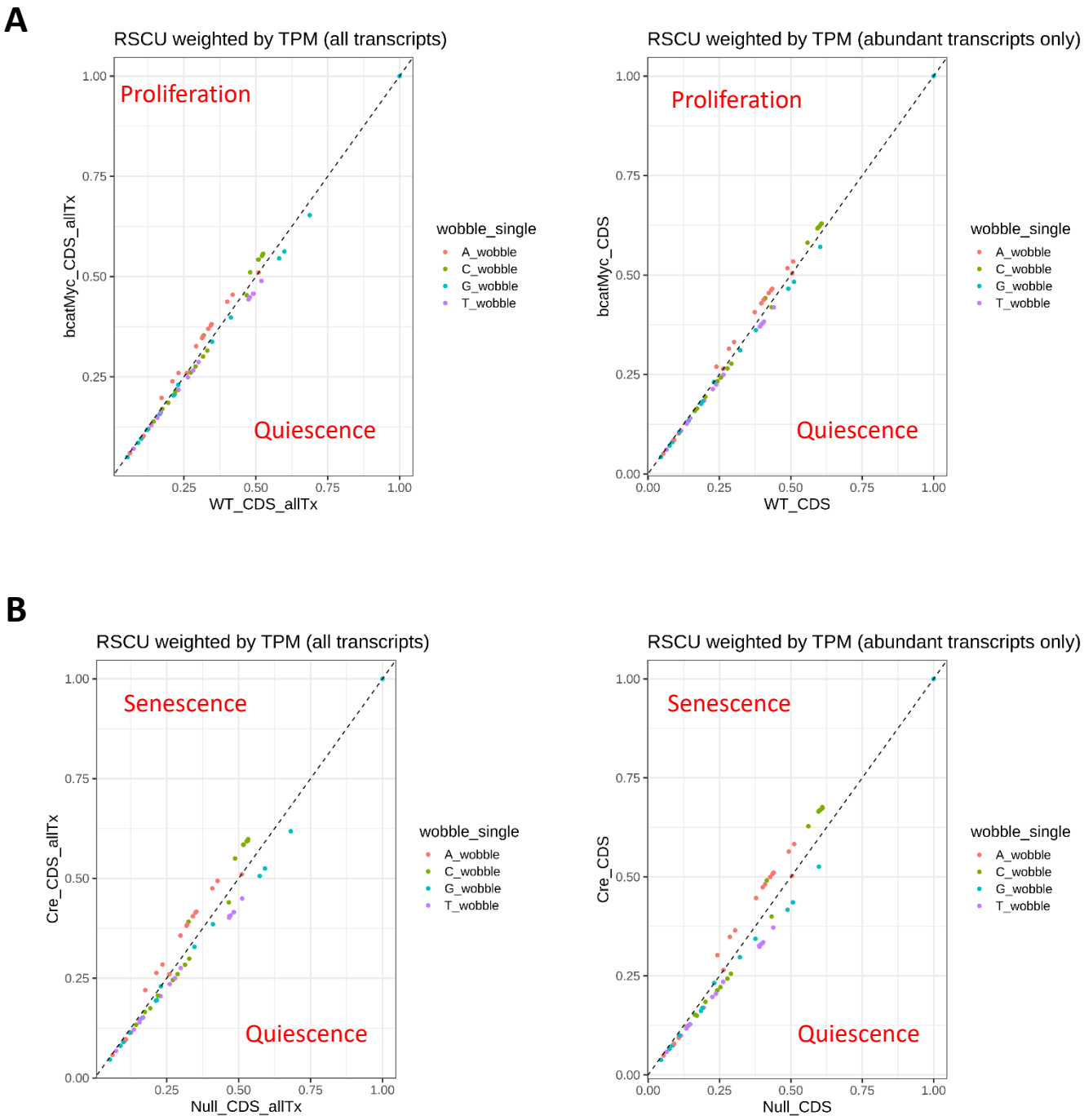


Figure 5.17 Relative synonymous codon usage for all transcripts and the most abundance transcript per gene in the βcatenin/c-myc and MDM2 models. (A) In the βcatenin/c-myc, there is a slight preference for A and C bases at the wobble position proliferation for calculations done for all transcripts and the most abundant transcript per gene. (B) In the MDM2 model, there is a is a preference for A and C bases at the wobble position in senescence for calculations done for all transcripts and the most abundant transcript per gene.

5.7 Investigation into potential A34I modifications at the isodecoder level

The tRNA sequencing results showed the importance of tRNA degeneracy in decoding mRNA sequences that vary in relative synonymous codon usage. The most important modification observed was the adenosine to inosine modification on position 34 of the tRNA wobble position. Since this modification allows for the decoding of A/T/C ending mRNA codons, enhancing the flexibility of the isoacceptor pool, it was important to ask how the modification rates changed for the isoacceptors they existed on between conditions.

The mutation frequency at position 34 was calculated by allowing for 1 mismatch when the tRNAs were mapped back to the mature tRNA database. If the mismatch occurred at position 34 for isoacceptors of Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val, which have an A at the wobble position and the base was read as a G base, then this would signal an A34I mutation. The mutation frequency was then calculated for all tRNAs in that isoacceptor group and compared between conditions.

For both the βcatenin/c-Myc model (Figure 5.18 A) and the MDM2 model (Figure 5.18 B), there were no statistically significant differences within isoacceptors between the two conditions. The mutation rate calculations showed that A34I modifications at these isodecoders were very high in both conditions in both experiments, suggesting that tRNA degeneracy is not regulated by changes in synonymous codon usage, but rather it is a mechanism that appears to be constitutively active.

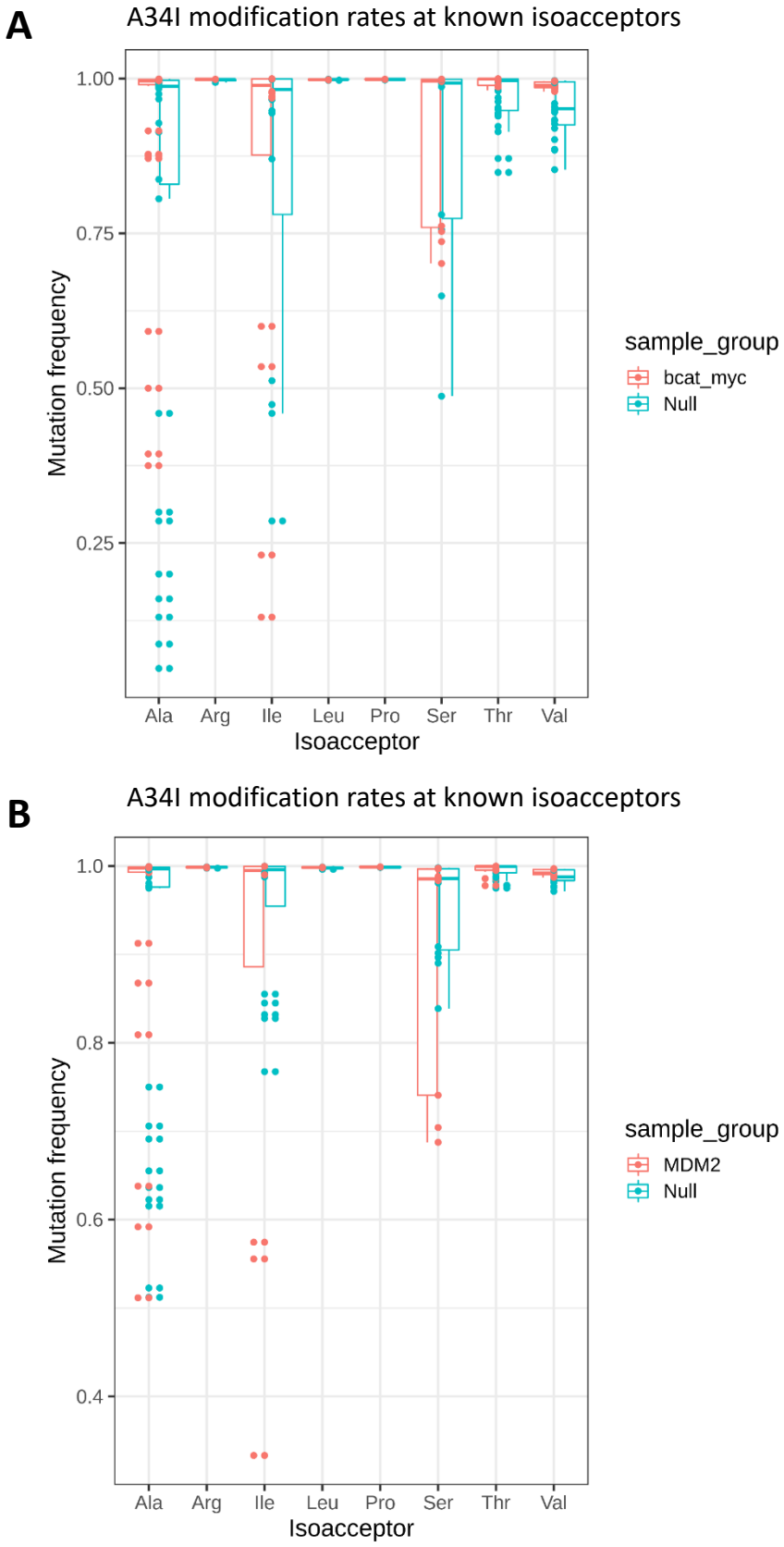


Figure 5.18 A34I modification changes at the isoacceptor level in the βcatenin/c-myc and MDM2 models. (A) A34I modification rates were calculated in the βcatenin/c-myc model but there was no statistically significant difference between quiescence and proliferation. (B) A34I modification rates were calculated in the MDM2 model but there was no statistically significant difference between quiescence and senescence.

5.8 Investigating if amino acid frequencies change in differentially expressed transcripts

It was important to check how amino acid frequency varied depending on the cell state. All statistically significant, differentially expressed genes with a $\log_2$ fold change of ±2 were used to calculate the amino acid frequency, normalized against the abundance of each gene.

In the βcatenin/c-Myc model, amino acids proline and serine were the most enriched amino acids encoded by mRNAs the quiescence condition and arginine, glutamic acid and alanine were the most frequent amino acids in proliferation (Figure 5.19 A).

Greater divergence in amino acid frequency was observed in the senescence condition of the MDM2 model, with arginine, proline and glycine being the most overrepresented amino acids in senescence. Isoleucine, phenylalanine and lysine were also overrepresented in the quiescence condition (Figure 5.19 B).

Although the amino acids were illustrated according to their properties, there was no clear correlation between amino acid property and cell state, suggesting that the changes in amino acid frequency may just be due to the types of transcripts upregulated in each cell condition, and is possibly not driving relative synonymous codon usage.

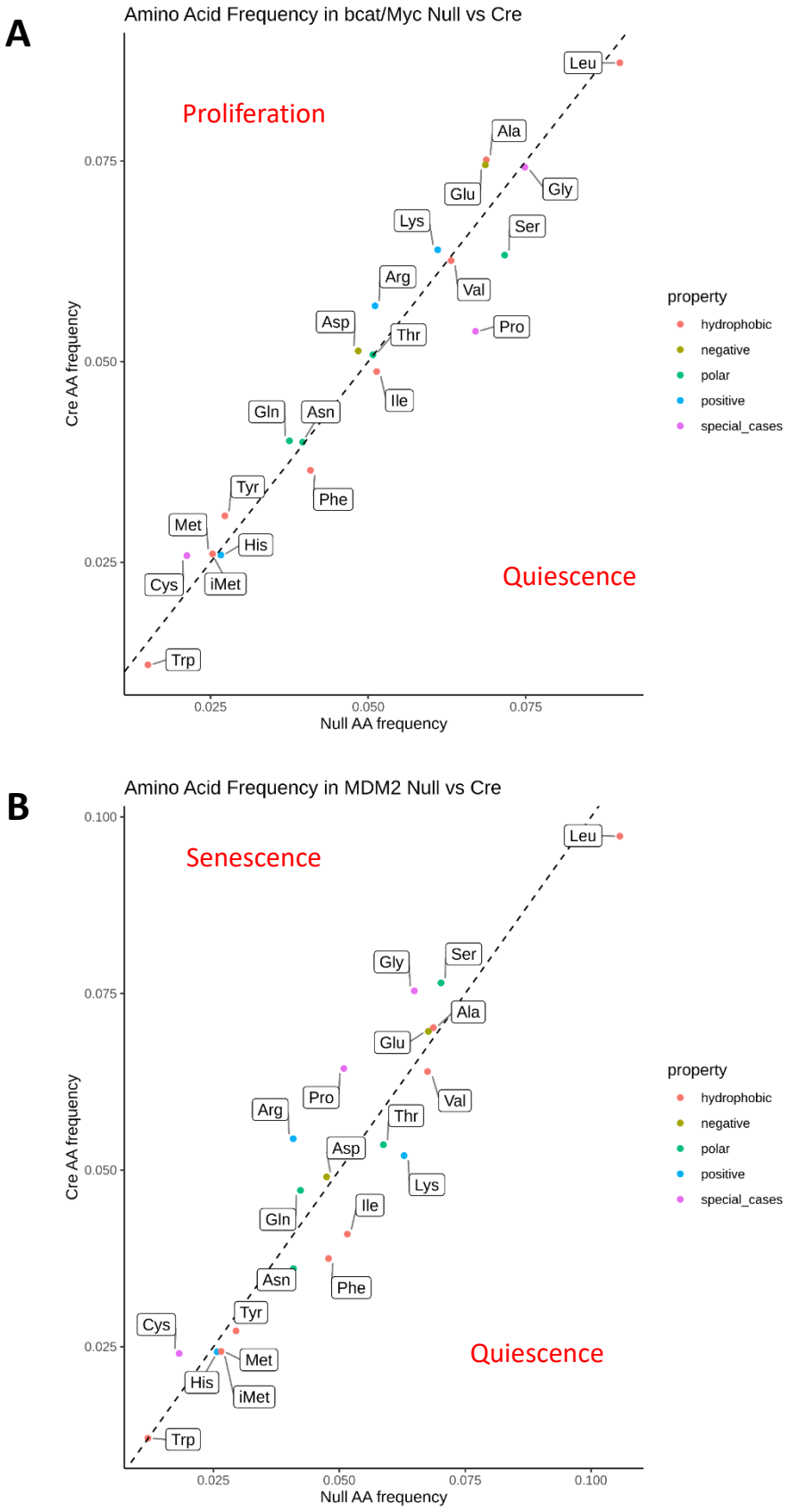


Figure 5.19 Amino acid frequencies in quiescence vs proliferation and quiescence vs senescence. All statically significant, differentially expressed genes with a log2 fold change of ±2 were used to calculate the amino acid frequency, normalized against the abundance of each gene. (A) Amino acid frequencies in the βcatenin/c-Myc model, where cells are driven from quiescence to proliferation. (B) Amino acid frequencies in the MDM2 model, where cells are driven from quiescence to senescence.

5.9 Comparison between the βcatenin/c-Myc and MDM2 models

A principal component analysis was performed on filtered tRNA replicates to see how the βcatenin/c-Myc and MDM2 models clustered. Separation was observed on both the PC1 and PC2 (which accounted for 46% and 33% variance respectively) for the βcatenin/c-Myc induced proliferation tRNAs (green) and the MDM2 induced senescence tRNAs (blue) (Figure 5.20). However, the two control conditions (treated with empty vectors, red and purple) also separated along PC1, but less so on PC2. Since the two control conditions were in quiescence, it was expected that they cluster together. However, given the two experiments utilized two different viral vectors (AAV8-TBG-PI-Cre-rBG for the model and AAV8-TBG-Cre for the MDM2 model), the two induction viruses may be resulting in different downstream effects in the two models.

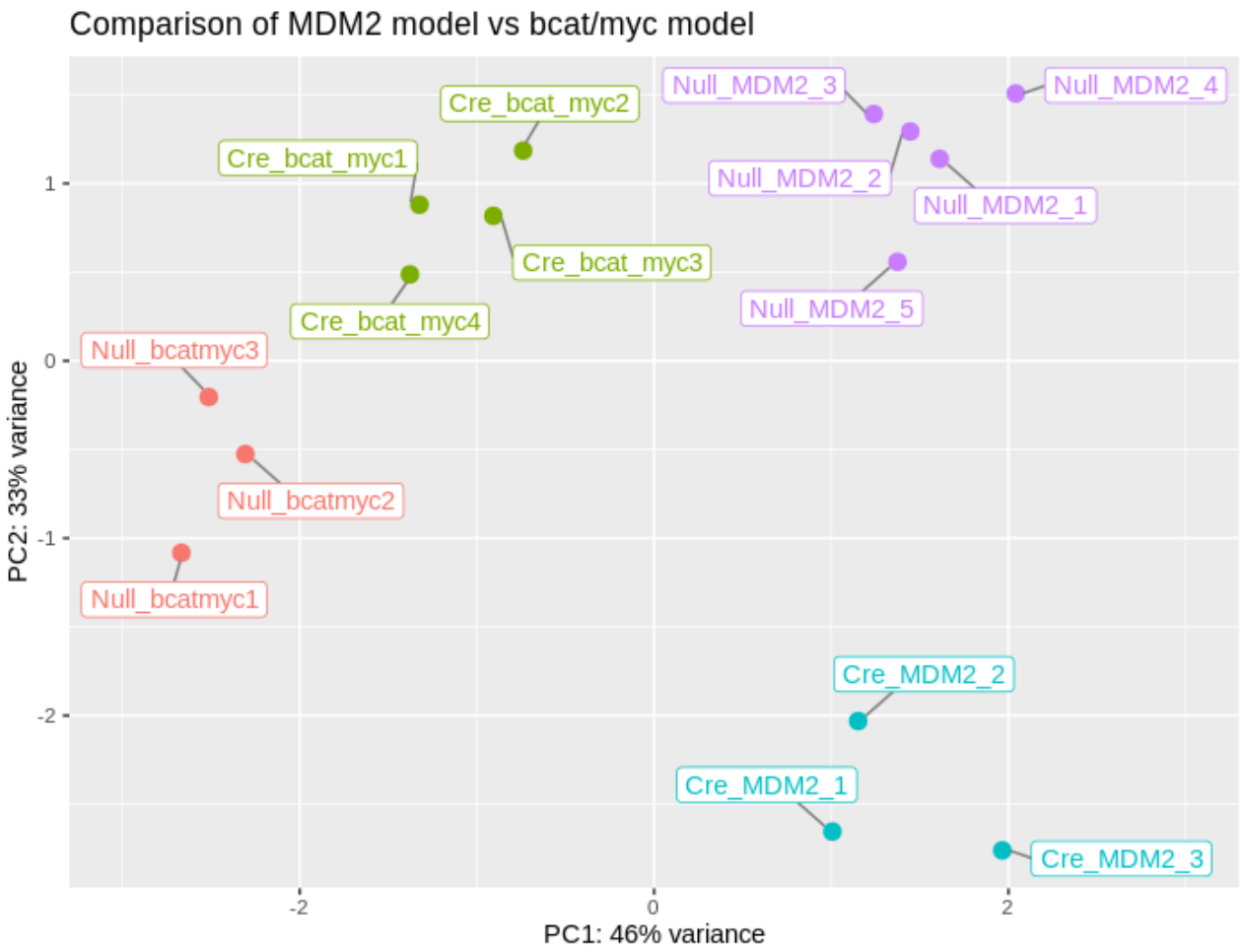


Figure 5.20 principal component analysis on the βcatenin/c-Myc vs MDM2 model. principal component analysis of the tRNA counts to determine where samples clustered. The two cancer models (Cre_MDM2 and Cre_bcat_myc) separated on both axis. The two quiescence conditions (Null_MDM2 and Null_bcat_myc) separated on mainly PC1

5.10 Chapter discussion

In this chapter, we have shown that ALICE-tRNA-seq is a suitable tRNA-sequencing methodology to measure the tRNA pool *in vivo*. After confirming AlkB can demethylate modified tRNA nucleotides in vivo, the tRNA-sequencing showed that we can measure the differential expression of tRNAs in genetically modified mice, without biasing the output of the tRNAs captured. As we did in vitro, we have shown that ALICE-tRNA-seq can identify all 48 isoacceptors, with tRNA mappings accounting for more than 85% of reads.

As shown in chapter 4, we also highlighted the importance of the A34I modification on the tRNA wobble position, as an effective means for potentially compensating for lowly expressed tRNA isoacceptors. This was especially prevalent in the 6-box and 4-box tRNAs, where preferred mRNAs codons coincided with either the most abundantly expressed cognate tRNAs or tRNAs that can bind via the A34I modification.

From the relative synonymous codon usage analysis, we established two major trends. Firstly, in the βcatenin/c-Myc model where hepatocytes are being driven from quiescence to senescence, we find that mRNA codons with an A or C at the wobble position are preferred in proliferation (with G and T ending codons being preferred in quiescence). This is reversed in the MDM2 model, where hepatocytes are being driven from quiescence to senescence, we find that mRNA codons with an A or C at the wobble position are preferred in quiescence (with G and T ending codons being preferred in senescence). We also found that the preferred C ending mRNA codons are decoded by the 2-box tRNAs that only express 1 tRNA (G:T base pairing is used to decode non-cognate bases). We find that mRNAs show the highest preference for cognate base pairing in this class of tRNAs (i.e C ending mRNA codons account for more than 75% of all synonymous codons in Asn, His, Phe, Asp, Cys and Tyr in

comparison to T ending codons) in both proliferation and senescence. This may suggest the cell's hesitancy in using G:T base pairing as a non-cognate means of decoding the transcriptome.

Interestingly in the MDM2 model, when relative synonymous codon usage was calculated for all transcripts and the most abundant transcript per gene, we observed a preference for A and C ending codons in senescence, but was completely reversed for the differential transcripts, where A and C ending codons were preferred in quiescence. This suggests that there may be a mechanism governing relative synonymous codon usage at play for the differentials.

Chapter 6: Discussion

The experiments, analysis and results presented in this thesis aimed to answer the following questions:

- Can a tRNA-sequencing methodology that avoids the limitations and biases recognised in literature be developed to accurately quantify the tRNA pool?
- Can such a tRNA-sequencing methodology claim to be more effective then published tRNA-seq methodologies?
- Can such a tRNA-sequencing methodology be used successfully both *in vitro* and in *vivo*?
- Through tRNA-seq and RNA-seq, can the tRNA and codon signatures of in vitro and in vivo samples be categorised in proliferation, quiescence and senescence
- Can this data be used to identify patterns in tRNA and mRNA signatures in the three major cell states

6.1 ALICE-tRNA-sequencing resolves limitations in current tRNA-sequencing methodologies

Given the importance of tRNA availability in meeting the transcriptomic demand, a methodology to accurately measure the tRNA-ome has long been an area of interest. In general, tRNAs can be surveyed to investigate either abundance, modifications or charging (Pan, 2018). To date, there is no accurate and unifying methodology that can be used to determine all three of these factors. Before the emergence of next generation sequencing methodologies, the traditional way of quantitatively measuring tRNA abundance was via the use of hybridization-based microarrays (Dittmar et al., 2006; Gingold et al., 2014), which have major limitations.

The sequencing of small RNAs requires the following key steps: adapter ligation to the RNA, reverse transcription to generate a cDNA library and the PCR amplification of the cDNA which is ultimately sequenced. Adapter ligation and reverse transcription are both major obstacles in the development of a tRNA sequencing methodology.

The canonical tRNA is characterized by its classical clover-leaf like secondary structure (Rich and Raj-Bhandary, 1976). Traditionally, the sequencing of RNAs begins with the ligation of 5' and 3' adapters, with the latter containing an RT primer binding site for the reverse transcription step. The 5' end of tRNAs are involved in a hairpin secondary structure, making access to the 5' end of the tRNA challenging. This has also been reported for other small RNAs with secondary structure based problems at their 5' ends due to hairpins affecting library preparation (Liu et al., 2014; Burke et al., 2014; Lama et al., 2019). 3' and 5' adapter ligation was a library preparation strategy employed by both ARM-seq (Cozen et al., 2015) and Hydro-seq (Gogakos et al., 2017). The main difference in library preparation between

the two methodologies was that Hydro-seq fragments tRNAs prior to adapter ligation by alkaline hydrolysis (similar to traditional RNA-sequencing).

In our pilot tRNA-seq studies on HEK293 cells, the ARM-seq methodology was used for library preparation (Figure 3.3). The rationale behind this was that in the analysis of the sequencing data, a clear distinction could be made with regards to whether a sequence was a mature tRNA, tRNA half or tRNA fragment. This is because the ligation of both a 3' and 5' adapter means that during the reverse transcription step, the reverse transcriptase would have to reach the 5' end, in order for the cDNA to incorporate both PCR primer binding sites (and subsequently be amplified). However, the sequencing results using this methodology showed that the majority of tRNAs sequenced were 3' tRNA fragments (60.98%) and 5' tRNA fragments (29.23%), with bona fide full length tRNAs only accounting for 5.11% of the mapped reads (Figure 3.6). When the sequencing data for both ARM-seq and Hydro-seq were analysed, the tRNA-coding amino acid distributions showed that almost 50% of all reads in ARM-seq mapped to tRNA-Glu (23.02%), tRNA-Gly (13.89%) and tRNA-Lys (14.28%). Similar biases were also observed in Hydro-seq, with more than 50% of all reads mapping to tRNA-Glu (39.55%), tRNA-Lys (9.36%) and tRNA-Val (11.15%) (Figure 3.20). These biases were not a surprise since tRNA glutamic acid, glycine, lysine and valine happen to be the least modified tRNAs. In the development of ALICE-tRNA-seq (**A**dapter **Li**gation **C**ircularization R**e**linearization tRNA sequencing), we chose to avoid 5' adapter ligation (Figure 3.7).

The other obstacle in the development of a tRNA sequencing methodology is modifications. Unlike other RNAs, over 90 modified human nucleosides have been attributed to tRNAs (Cantara et al., 2011; Saikia et al., 2010). For tRNA sequencing, the modifications that pose

the biggest problem are those that exist on the Watson-Crick face of tRNAs (Table 3.1). This is because traditional reverse transcriptases like SuperScript III would stall when they reach said modifications. This could also explain why ARM-seq and Hydro-seq show huge biases in tRNA-coding amino acid distributions for tRNAs known to be lightly modified. The *Escherichia coli* enzyme Alpha-ketoglutarate-dependent dioxygenase (AlkB), an enzyme known for its dealklylating properties in DNA damage protection (Trewick et al., 2002) has the ability to remove certain methyl groups from modified nucleotides such as methyl-guanosine, methyl-adenosine and more (Table 3.1). There are two versions of AlkB often used in tRNA-sequencing; its wild-type form (WT) and the mutant D135S (coinciding with the active site of the protein). The shorter side chain of serine is thought to accommodate bigger modifications like the dimethyl-guanosine, a common modification found in tRNA sequences (Zheng et al., 2015).

Because Hydro-seq fragments all tRNA sequences prior to library preparations in an attempt to avoid RT stalling due to modifications, they do not use any AlkB. ARM-seq and QuantM-seq (Pinkard et al., 2020) only use the wild-type AlkB in their preparations and DM-seq (Zheng et al., 2015) and ALICE-tRNA-seq utilize both the WT and D135S forms in library preparations. We showed that the use of AlkB was paramount in tRNA-sequencing. When tRNAs were sequenced in the pilot study, tRNA-coding amino acid distributions drastically improved with WT and D135S demethylation prior to library preparation in contrast to no demethylation (Figure 3.5 B). We also showed that WT and D135S was effective in reducing RT stalling in our primer extension assays both *in vitro* (Figure 3.1 B and Figure 3.2) and *in vivo* (Figure 5.2).

DM-seq, QuantM-seq and ALICE-tRNA-seq all incorporate circularization as a means of avoiding 5' adapter ligation and capturing tRNAs that may have truncated cDNAs due to modifications AlkB could not remove. Furthermore, all three methodologies resembled similar anticodon frequencies which correlated well when compared to codon usage of HEK293 cells based on RNA-seq data (Figure 3.21). DM-seq would not be an effective means for measuring the translating tRNA pool, as it does not specifically measure charged tRNAs, but is an excellent methodology for the identification of modified tRNA nucleotides (discussed in Introduction – tRNA modifications). This is because their 3' adapter ligation is non-specific, meaning that it is impossible to distinguish between a tRNA that has been truncated due to the presence of a modification, or whether the sequence was a bona fide tRNA fragment or tRNA half. QuantM-seq tries to overcome this by using an annealed adapter where the 5' adapter had an overhang that would be complementary to the CCA of the 3' end of the tRNA, aiding 3' adapter ligation (Figure 1.6 C). ALICE-tRNA-seq utilizes a 3' adapter with an adenylated 5' end, meaning that the adapter would only bind to adenosines that have a 3' hydroxyl group, which can only be a consequence of deacylation (removal of the amino acid from the 3' end of tRNAs) (Figure 3.7 B). Although QuantM-seq does improve on mature tRNA specificity in comparison to DM-seq, there are limitations in that non-functional tRNAs and 3' tRNA fragments have been shown to still harbour a 3' CCA (Anderson and Ivanov, 2014), and so these would still be sequenced. Therefore, during analysis, a distinction would not be able to be made between mature tRNAs that may have truncated due to a modification and genuine 3' tRNA fragments. Furthermore, tRNAs targeted for degradation have an additional CCA added to their 3' end (Wellner et al., 2018) which would be sequenced in QuantM-seq (although these could be removed bioinformatically but was not mentioned in the paper). Ultimately, the correlation between

QuantM-seq and ALICE-tRNA-seq tRNA sequencing counts had R^2 values greater than or equal to 0.83 and were statistically significant (Figure 3.19 C and D).

However, we believe that ALICE-tRNA-seq provides the best means for measuring the tRNA pool. The 3' adapter ligation provides more specificity to mature tRNAs than the aforementioned methodologies, due to the adapter's adenylated 5' end. Furthermore, the existence of quality control steps such as gel extractions post reverse transcription and post PCR, allows visualisation of the library at each step, allowing the user to be sure that the library preparation is working correctly.

6.2 tRNA availability and synonymous codon usage measured *in vitro* and *in vivo* does not entirely agree with published data

Investigations into the role of tRNAs in the increased rate of translation of proliferative mRNAs has led to three working hypotheses; The expression of proliferative mRNAs results in the upregulation of "proliferative" tRNAs (Gingold et al., 2014), the expression of proliferative mRNAs results in the upregulation of all tRNAs and is not a specific event (Guimaraes et al., 2020); and the expression of proliferative mRNAs results in the upregulation of tRNA modifications, namely A34I (Bornelov et al., 2019).

In vitro, two models were used, in which cells were driven from proliferation to either quiescence or senescence, to try and elucidate their tRNA and mRNA profiles, by sequencing the mRNAs and tRNAs and examining how they vary upon changes in cell state. The first model was serum deprivation in BJ5TA human fibroblasts where cells were driven from proliferation to quiescence. The second model used IMR90 primary human fibroblasts, where cells were driven from proliferation to senescence via the activation of Ras, leading to oncogene induced senescence.

In the serum deprivation experiments (proliferation to quiescence), tRNA sequencing revealed that there was a general downregulation of tRNA isoacceptors with an A/T at the wobble position and an upregulation of tRNA isoacceptors with a G/C (Figure 4.14). This could be to try and accommodate the changes seen at the mRNA level. The relative synonymous codon usage calculated for differential genes from the RNA-sequencing data revealed that mRNA codons with an A/T at the 3rd nucleotide position were preferred in proliferation and codons with a G/C at the 3rd nucleotide position were preferred in quiescence, therefore fitting with tRNA expression changes (Figure 4.24 A). However, when

the synonymous codon usage was calculated for all transcripts and the most abundant transcript per gene, there was no preference for any synonymous codons, suggesting that the differential genes are driving RSCU changes (Figure 4.25 A).

The serum deprivation experiment coincides with Gingold's et al hypothesis that the expression of proliferative mRNAs results in the upregulation of "proliferative" tRNAs. Gingold et al observed that "proliferative" tRNAs coincided with the upregulation of proliferative mRNAs as well as histone modification changes around tRNA genes and elucidated towards a potential transcriptional programme that regulates both tRNA and mRNA expression (Gingold et al., 2014). They also observed that proliferative mRNAs were enriched for A/T ending codons in both cancerous and non-cancerous proliferative samples, while genes associated with pattern specification were enriched for G/C ending codons and hence coined the terms "proliferative tRNAs" and "differentiation tRNAs". Furthermore, Guimaraes et al conducted investigations into the effects of serum-deprivation in mouse embryonic fibroblasts and also reported the enrichment of A/T ending mRNA codons in proliferation but found no differential tRNA expression (Guimaraes et al., 2020). Although our model was a measure of stress-induced quiescence, whereas these reports from literature are comparing proliferation- and pattern specification- associated genes, the data observed in the proliferative state of our BJ5TA human fibroblasts matches. Furthermore, since ALICE-tRNA-seq is a more sensitive means of measuring the tRNA pool (in comparison to northern blots) we can confirm in this model that tRNA availability matches with the transcriptomic demand by upregulating select tRNAs.

However, in the IMR90s, where cells were driven from proliferation to senescence, the tRNA availability did not match with the mRNA demands. From the tRNA-sequencing, the

isoacceptors did not show any pattern with regards to the up/downregulation of A/T vs G/C at the wobble position. In fact, the majority of T ending isoacceptors were upregulated and C ending isoacceptors were downregulated in senescence (Figure 4.15), completely contradicting all three hypotheses. However, when we examined the relative synonymous codon usage of differential genes, we observed A ending mRNA codons being preferred in proliferation and G/C ending mRNA codons being preferred in senescence (which does coincide with published reports) (Figure 4.24 B-C). Interestingly, mRNAs with a T at the 3rd nucleotide position stayed neutral in terms of preference between the two conditions. When the RSCU was calculated for all transcripts and the most abundant transcript per gene, there was almost no preference observed (Figure 4.25 B), suggesting that preference is driven by the differential genes. The fact that tRNA expression is not changing may be to do with downstream effects of RAS signalling.

Similar results to the oncogene induced senescence experiments were also observed in the *in vivo* models. We used two genetically modified mouse models; the βcatenin/c-Myc model which drives hepatocytes from quiescence to proliferation and an MDM2 model which drives hepatocytes from quiescence to senescence. Interestingly, for both experiments, the tRNA sequencing did not match with the calculated relative synonymous codon usage for differential genes. This suggests that there may be a mechanism that cancer models utilize to bypass tRNA expression. An interesting aspect to investigate would be charging rates of the tRNAs. If the tRNAs are being charged with their cognate amino acids more efficiently, the tRNAs would be able to meet transcriptomic demands without having to upregulate tRNA expression. This could also be driven by specific tRNA modifications, but would require further investigation.

In the βcatenin/c-Myc model (quiescence to proliferation), the relative synonymous codon usage of differential genes showed a preference for mRNA codons with an A at the wobble position in proliferation (matching with the IMR90 data). Interestingly, mRNA codons with a C at the wobble position were also preferred in proliferation, contradicting published data (Figure 5.16 A). However, when the C ending mRNAs preferred in proliferation were examined, we found that they were all decoded by the 2-box tRNAs that only express one tRNA (i.e. G:T base pairing is utilized to decode the non-cognate codon, Figure 5.12 A-F). When it came to preferred codon usage for this category, we find that the cognate: non-cognate split was approximately 75%:25% in preference (Figure 5.12 A-F). This led us to conclude that proliferative hepatocytes try to avoid G:T base pairing when decoding proliferative differential genes.

In the MDM2 model (quiescence to senescence), relative synonymous codon usage of differential genes showed a preference for mRNA codons with an A or C at the wobble position in quiescence (Figure 5.16 B). Like observed in the βcatenin/c-Myc model, the C ending mRNAs preferred in proliferation were examined, we found that they were all decoded by the 2-box tRNAs that only express one tRNA (Figure 5.13 A-F), again suggesting that quiescent hepatocytes prefer cognate base pairing over G:T base pairing. Interestingly, in the MDM2 model, when the RSCU was calculated for all transcripts and the most abundant transcript per gene, we observed A-ending and C-ending codons being preferred in senescence (Figure 5.17 B), the mirror effect of what was observed for the RSCU of differential genes. This will require further investigation.

Another form of near-cognate binding comes in the form of tRNA selection at the ribosomes (Tarrant and Von der Haar, 2014). tRNAs are translocated to the ribosome via the

phosphorylation of eIF1A, selecting tRNAs carrying the correct amino acids to the A-site of the translating ribosome (Rodnina and Wintermeyer, 2009). However, there are instances when incorrect amino acids are delivered due to near-cognate complementarity between the tRNA anticodon and the mRNA codons (usually when the second nucleotide is not completely complementary), resulting in increased dwell time in relation to the translating ribosome (Tarrant and Von der Haar, 2014). Near-cognate binding at the ribosome A-site has been shown to decrease translational efficiency (Pape et al., 1999). Furthermore, the ratio between cognate and near-cognate tRNA abundances have been an area of immense interest in the tRNA community, as they could play a significant role in the determination of codon decoding times (Tarrant and Von der Haar, 2014). The relationship between cognate/near-cognate ratios in relation to tRNA abundance and translational efficiency could yield a regulatory function within the cell, and has therefore become a significant area of interest. The redundancies seen (especially in the cancer models) between tRNA availability and synonymous codon usage could be explained by the existence of non-cognate binding occurring at ribosomal A-sites. Although speculation, this could be a significant regulatory process used by the cancer cells to regulate mRNA translational rates, but is an area yet to be elucidated.

The overall observation we found from our studies was that the stress-induced proliferation to quiescence transition matched up with the Gingold's hypothesis. However, all our cancer models showed that the tRNA expression did not match up with the RSCU of differential genes. This suggests that cancers may be able to find a way to meet transcriptomic demand, without the need for instituting global tRNA expression changes (possibly through more efficient tRNA charging).

Interestingly, we found that in all our models, A34I (adenosine to inosine modification at the wobble position of tRNAs) may have been constitutively active in all our models and in some cases, may have been compensating for lowly expressed tRNAs. However, all our data disagrees with the Bornelov model, namely the A34I modifications drives increased translational rates of proliferative mRNAs.

Although the findings reported allows us to classify the mRNAome and tRNAome of popular cancer models, more investigations would have to be conducted to elucidate the mechanism that regulated tRNA availability with translational demands. To potentially answer these questions, it may be useful to couple tRNA-sequencing with ribosome profiling, since the latter would give an indication to the codons being actively translated and indications towards the cell's translation rate. Furthermore, mass spectrometry coupled with DM-tRNA-seq may allow us to categorize global tRNA modification changes with DM-seq allowing us to postulate the tRNAs that have differential modification changes.

In conclusion, we have successfully optimized a tRNA-sequencing methodology that can accurately measure the tRNA pool *in vivo* and *in vitro*, which aids our understanding into how tRNA dynamics changes with transcriptomic demand.

Appendix

Below are the sequenced for ALICE-tRNA-seq library preparations:

	tRNA sequencing adapters (ALICE-tRNA-seq Library Preparation)
3' Adapter	5’pCAGATCGGAAGAGCACACGTCT-R-NH2
RT primer	5’pGATCGTCGGACTGTAGAACTCTGAArCAGACGTGTGCTCTTCCGATCT
PCR Forward Primer	5'AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA

Below are the PCR reverse primers. Each primer has a designed unique 6nt barcode used for demultiplexing pooled samples. Each primer can only be used once for every sequencing run:

PCR Reverse Primer	**Full Sequence**	**Unique barcode**
Rv_primer_1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	CGTGAT
Rv_primer_2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	ACATCG
Rv_primer_3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	GCCTAA
Rv_primer_4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	TGGTCA
Rv_primer_5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	CACTGT

Rv_primer_6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	ATTGGC
Rv_primer_7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	GATCTG
Rv_primer_8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	TCAAGT
Rv_primer_9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	CTGATC
Rv_primer_10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	AAGCTA
Rv_primer_11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	GTAGCC
Rv_primer_12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	TACAAG
Rv_primer_13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	TTGACT
Rv_primer_14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	GGAACT
Rv_primer_15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	TGACAT
Rv_primer_16	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT	GGACGG

Rv_primer_17	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CTCTAC
Rv_primer_18	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GCGGAC
Rv_primer_19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TTTCAC
Rv_primer_20	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GGCCAC
Rv_primer_21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CGAAAC
Rv_primer_22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CGTACG
Rv_primer_23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CCACTC
Rv_primer_24	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GCTACC
Rv_primer_25	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	ATCAGT
Rv_primer_26	CAAGCAGAAGACGGCATACGAGATGCTCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GCTCAT
Rv_primer_27	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	AGGAAT

Rv_primer_28	CAAGCAGAAGACGGCATACGAGATCTTTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CTTTTG
Rv_primer_29	CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TAGTTG
Rv_primer_30	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CCGGTG
Rv_primer_31	CAAGCAGAAGACGGCATACGAGATATCGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	ATCGTG
Rv_primer_32	CAAGCAGAAGACGGCATACGAGATTGAGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TGAGTG
Rv_primer_33	CAAGCAGAAGACGGCATACGAGATCGCCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CGCCTG
Rv_primer_34	CAAGCAGAAGACGGCATACGAGATGCCATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GCCATG
Rv_primer_35	CAAGCAGAAGACGGCATACGAGATAAAATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	AAAATG
Rv_primer_36	CAAGCAGAAGACGGCATACGAGATTGTTGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TGTTGG
Rv_primer_37	CAAGCAGAAGACGGCATACGAGATATTCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	ATTCCG
Rv_primer_38	CAAGCAGAAGACGGCATACGAGATAGCTAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	AGCTAG

Rv_primer_39	CAAGCAGAAGACGGCATACGAGATGTATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GTATAG
Rv_primer_40	CAAGCAGAAGACGGCATACGAGATTCTGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TCTGAG
Rv_primer_41	CAAGCAGAAGACGGCATACGAGATGTCGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GTCGTC
Rv_primer_42	CAAGCAGAAGACGGCATACGAGATCGATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CGATTA
Rv_primer_43	CAAGCAGAAGACGGCATACGAGATGCTGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GCTGTA
Rv_primer_44	CAAGCAGAAGACGGCATACGAGATATTATAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	ATTATA
Rv_primer_45	CAAGCAGAAGACGGCATACGAGATGAATGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	GAATGA
Rv_primer_46	CAAGCAGAAGACGGCATACGAGATTCGGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TCGGGA
Rv_primer_47	CAAGCAGAAGACGGCATACGAGATCTTCGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	CTTCGA
Rv_primer_48	CAAGCAGAAGACGGCATACGAGATTGCCGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TGCCGA

Bibliography

Abbott, J.A., Francklyn, C.S. and Robey-Bond, S.M., 2014. Transfer RNA and human disease. Frontiers in genetics, 5, p.158.

Abe, T., Inokuchi, H., Yamada, Y., Muto, A., Iwasaki, Y. and Ikemura, T., 2014. tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. Frontiers in genetics, 5, p.114.

Agris, P.F., 2004. Decoding the genome: a modified view. Nucleic acids research, 32(1), pp.223-238.

Ally, A., Balasundaram, M., Carlsen, R., Chuah, E., Clarke, A., Dhalla, N., Holt, R.A., Jones, S.J., Lee, D., Ma, Y. and Marra, M.A., 2017. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell, 169(7), pp.1327-1341.

Anderson, P. and Ivanov, P., 2014. tRNA fragments in human health and disease. FEBS letters, 588(23), pp.4297-4304.

Barker, N., Ridgway, R.A., Van Es, J.H., Van De Wetering, M., Begthel, H., Van Den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J. and Clevers, H., 2009. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature, 457(7229), pp.608-611.

Bazzini AA, Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, et al. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition . EMBO J . 2016 Oct 4;35(19):2087–103.

Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. Nature Reviews Genetics, 15(3), pp.163-175.

Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J. and Topisirovic, I., 2015. Targeting the translation machinery in cancer. Nature reviews Drug discovery, 14(4), pp.261-278.

Bird, T.G., Müller, M., Boulter, L., Vincent, D.F., Ridgway, R.A., Lopez-Guadamillas, E., Lu, W.Y., Jamieson, T., Govaere, O., Campbell, A.D. and Ferreira-Gonzalez, S., 2018. TGFβ inhibition restores a regenerative response in acute liver injury by suppressing paracrine senescence. Science translational medicine, 10(454).

Bisso, A., Filipuzzi, M., Gamarra Figueroa, G.P., Brumana, G., Biagioni, F., Doni, M., Ceccotti, G., Tanaskovic, N., Morelli, M.J., Pendino, V. and Chiacchiera, F., 2020. Cooperation between MYC and β-Catenin in liver tumorigenesis requires Yap/Taz. Hepatology, 72(4), pp.1430-1443.

Blanco, S. and Frye, M., 2014. Role of RNA methyltransferases in tissue renewal and pathology. Current opinion in cell biology, 31, pp.1-7.

Boël G, Letso R, Neely H, Price WN, Wong KH, Su M, et al. Codon influence on protein expression in E. coli correlates with mRNA levels. Nature. 2016 Jan 21;529(7586):358–63.

Bornelöv S, Selmi T, Flad S, Dietmann S, Frye M. Codon usage optimization in pluripotent embryonic stem cells. Genome Biol . 2019 Jun 7;20(1).

Botzman, M. and Margalit, H., 2011. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. Genome biology, 12(10), pp.1-11.

Buffington, S.A., Huang, W. and Costa-Mattioli, M., 2014. Translational control in synaptic plasticity and cognitive dysfunction. Annual review of neuroscience, 37, pp.17-38.

Burke, J.M., Bass, C.R., Kincaid, R.P. and Sullivan, C.S., 2014. Identification of tri-phosphatase activity in the biogenesis of retroviral microRNAs and RNAP III-generated shRNAs. Nucleic acids research, 42(22), pp.13949-13962.

Buttgereit, F. and Brand, M.D., 1995. A hierarchy of ATP-consuming processes in mammalian cells. Biochemical Journal, 312(1), pp.163-167.

Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F., 2010. The RNA modification database, RNAMDB: 2011 update. Nucleic acids research, 39(suppl_1), pp.D195-D201.

Carneiro RL, Requião RD, Rossetto S, Domitrovic T, Palhano FL. Codon stabilization coefficient as a metric to gain insights into mRNA stability and codon bias and their relationships with translation. Nucleic Acids Res . 2019 Mar 18;47(5):2216–28.

Chan, P.P. and Lowe, T.M., 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic acids research, 37(suppl_1), pp.D93-D97.

Cathopoulis TJ, Chuawong P, Hendrickson TL. A thin-layer electrophoretic assay for Asp-tRNAAsn/Glu-tRNAGln amidotransferase. Anal Biochem. 2007 Jan 1;360(1):151-3. doi: 10.1016/j.ab.2006.10.019. Epub 2006 Nov 3. PMID: 17113030; PMCID: PMC1800913.

Clark, W.C., Evans, M.E., Dominissini, D., Zheng, G. and Pan, T., 2016. tRNA base methylation identification and quantification via high-throughput sequencing. Rna, 22(11), pp.1771-1784.

Colnot, S., Decaens, T., Niwa-Kawakita, M., Godard, C., Hamard, G., Kahn, A., Giovannini, M. and Perret, C., 2004. Liver-targeted disruption of Apc in mice activates β-catenin signaling

and leads to hepatocellular carcinomas. Proceedings of the National Academy of Sciences, 101(49), pp.17216-17221.

Cooley, L., Appel, B. and Söll, D., 1982. Post-transcriptional nucleotide addition is responsible for the formation of the 5'terminus of histidine tRNA. Proceedings of the National Academy of Sciences, 79(21), pp.6475-6479.

Courel M, Clément Y, Bossevain C, Foretek D, Cruchez OV, Yi Z, et al. Gc content shapes mRNA storage and decay in human cells. Elife. 2019;8:1–32.

Cozen, A.E., Quartley, E., Holmes, A.D., Hrabeta-Robinson, E., Phizicky, E.M. and Lowe, T.M., 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nature methods, 12(9), pp.879-884.

Cramer, F., Erdmann, V.A., Von Der Haar, F. and Schlimme, E., 1969. Structure and reactivity of tRNA. Journal of cellular physiology, 74(S1), pp.163-178.

Crothers, D.M., Seno, T. and Söll, D.G., 1972. Is there a discriminator site in transfer RNA? Proceedings of the National Academy of Sciences, 69(10), pp.3063-3067.

Dajee, M., Tarutani, M., Deng, H., Cai, T. and Khavari, P.A., 2002. Epidermal Ras blockade demonstrates spatially localized Ras promotion of proliferation and inhibition of differentiation. Oncogene, 21(10), pp.1527-1538.

Deamer, D., Akeson, M. and Branton, D., 2016. Three decades of nanopore sequencing. Nature biotechnology, 34(5), pp.518-524.

De La Coste, A., Romagnolo, B., Billuart, P., Renard, C.A., Buendia, M.A., Soubrane, O., Fabre, M., Chelly, J., Beldjord, C., Kahn, A. and Perret, C., 1998. Somatic mutations of the β-catenin

gene are frequent in mouse and human hepatocellular carcinomas. Proceedings of the National Academy of Sciences, 95(15), pp.8847-8851.

Desai N, Lee J, Upadhya R, Chu Y, Moir RD, Willis IM. 2005. Two steps in Maf1-dependent repression of transcription by RNA polymerase III. J Biol Chem 280: 6455–6462.

Di Micco, R., Krizhanovsky, V., Baker, D. and di Fagagna, F.D.A., 2021. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. Nature Reviews Molecular Cell Biology, 22(2), pp.75-95.

Dittmar, K.A., Goodenbour, J.M. and Pan, T., 2006. Tissue-specific differences in human transfer RNA expression. PLoS genetics, 2(12), p.e221.

dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: A test for translational selection. Nucleic Acids Res . 2004;32(17):5036–44.

Dong, H., Nilsson, L. and Kurland, C.G., 1996. Co-variation of trna abundance and codon usage inescherichia coliat different growth rates. Journal of molecular biology, 260(5), pp.649-663.

Drummond, D.A. and Wilke, C.O., 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell, 134(2), pp.341-352.

Falnes, P.Ø., Johansen, R.F. and Seeberg, E., 2002. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature, 419(6903), pp.178-182.

Forrest ME, Pinkard O, Martin S, Sweet TJ, Hanson G, Coller J. Codon and amino acid content are associated with mRNA stability in mammalian cells. Kim YK, editor. PLoS One . 2020 Feb 13;15(2):e0228730.

Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, Hou YM, et al. Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol . 2012

Frohlich et al., 2014 in Jurga, S., Erdmann, V.A. and Barciszewski, J. eds., 2016. Modified nucleic acids in biology and medicine. Springer International Publishing.

Frumkin, I., Lajoie, M.J., Gregg, C.J., Hornung, G., Church, G.M. and Pilpel, Y., 2018. Codon usage of highly expressed genes affects proteome-wide translation efficiency. Proceedings of the National Academy of Sciences, 115(21), pp.E4940-E4949.

Gannon, H.S., Donehower, L.A., Lyle, S. and Jones, S.N., 2011. Mdm2–p53 signaling regulates epidermal stem cell senescence and premature aging phenotypes in mouse skin. Developmental biology, 353(1), pp.1-9.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S. and Futcher, B., 2014. Measurement of average decoding rates of the 61 sense codons in vivo. Elife, 3, p.e03735.

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, et al. 2006. Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631–636.

Geslain, R. and Pan, T., 2010. Functional analysis of human tRNA isodecoders. Journal of molecular biology, 396(3), pp.821-831.

Giegé, R., 2008. Toward a more complete view of tRNA biology. Nature structural & molecular biology, 15(10), pp.1007-1014.

Gingold, H., Dahan, O. and Pilpel, Y., 2012. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. Nucleic acids research, 40(20), pp.10053-10063.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D. and Andersen, L.D., 2014. A dual program for translation regulation in cellular proliferation and differentiation. Cell, 158(6), pp.1281-1292.

Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M. and Tuschl, T., 2017. Characterizing expression and processing of precursor and mature human tRNAs by hydro-tRNAseq and PAR-CLIP. Cell reports, 20(6), pp.1463-1475.

Goodenbour, J.M. and Pan, T., 2006. Diversity of tRNA genes in eukaryotes. Nucleic acids research, 34(21), pp.6137-6146.

Gouw, A.S., Clouston, A.D. and Theise, N.D., 2011. Ductular reactions in human liver: diversity at the interface. Hepatology, 54(5), pp.1853-1863.

Grelet, S., McShane, A., Hok, E., Tomberlin, J., Howe, P. H., & Geslain, R. (2017). SPOt: A novel and streamlined microarray platform for observing cellular tRNA levels. PloS one, 12(5), e0177939. https://doi.org/10.1371/journal.pone.0177939

Grier, J.D., Xiong, S., Elizondo-Fraire, A.C., Parant, J.M. and Lozano, G., 2006. Tissue-specific differences of p53 inhibition by Mdm2 and Mdm4. Molecular and cellular biology, 26(1), pp.192-198.

Grosjean, H., Keith, G. and Droogmans, L., 2004. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. In RNA Interference, Editing, and Modification (pp. 357-391). Humana Press.

Guimaraes JC, Mittal N, Gnann A, Jedlinski D, Riba A, Buczak K, et al. A rare codon-based translational program of cell proliferation. Genome Biol . 2020 Feb 27;21(1):44.

Hanahan, D. and Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. cell, 144(5), pp.646-674.

Harada, N., Tamai, Y., Ishikawa, T.O., Sauer, B., Takaku, K., Oshima, M. and Taketo, M.M., 1999. Intestinal polyposis in mice with a dominant stable mutation of the β-catenin gene. The EMBO journal, 18(21), pp.5931-5942.

Hayflick, L. and Moorhead, P.S., 1961. The serial cultivation of human diploid cell strains. Experimental cell research, 25(3), pp.585-621.

Hernandez-Alias X, Benisty H, Schaefer MH, Serrano L. Translational efficiency across healthy and tumor tissues is proliferation-related. Mol Syst Biol . 2020 Mar;16(3). 5

Hia F, Yang SF, Shichino Y, Yoshinaga M, Murakawa Y, Vandenbon A, et al. Codon bias confers stability to human mRNAs. EMBO Rep . 2019 Nov 5;20(11).

Hinnebusch, A.G., Ivanov, I.P. and Sonenberg, N., 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science, 352(6292), pp.1413-1416.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R. and Zamir, A., 1965. Structure of a ribonucleic acid. Science, pp.1462-1465.

Huber A, Bodenmiller B, Uotila A, Stahl M, Wanka S, Gerrits B, Aebersold R, Loewith R. 2009. Characterization of the rapamycin- sensitive phosphoproteome reveals that Sch9 is a central coordinator of protein synthesis. Genes Dev 23: 1929–1943.

Ibba, M. and Söll, D., 2000. Aminoacyl-tRNA synthesis. Annual review of biochemistry, 69(1), pp.617-650.

Iben, J.R. and Maraia, R.J., 2014. tRNA gene copy number variation in humans. Gene, 536(2), pp.376-384.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature protocols, 7(8), pp.1534-1550.

Innes, A.J. and Gil, J., 2019. IMR90 ER: RAS: A Cell Model of Oncogene-Induced Senescence. In Cellular Senescence (pp. 83-92). Humana Press, New York, NY.

Jablkowski, M., Bocian, A., Bialkowska, J. and Bartkowiak, J., 2005. A comparative study of P53/MDM2 genes alterations and P53/MDM2 proteins immunoreactivity in liver cirrhosis and hepatocellular carcinoma. Journal of experimental & clinical cancer research: CR, 24(1), pp.117-125.

Jackson, R.J., Hellen, C.U. and Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principals of its regulation. Nature reviews Molecular cell biology, 11(2), pp.113-127.

Jacob, D., Thüring, K., Galliot, A., Marchand, V., Galvanin, A., Ciftci, A., Scharmann, K., Stock, M., Roignant, J.Y., Leidel, S.A. and Motorin, Y., 2019. Absolute quantification of noncoding

RNA by microscale thermophoresis. Angewandte Chemie International Edition, 58(28), pp.9565-9569.

Javid, B., Sorrentino, F., Toosky, M., Zheng, W., Pinkham, J.T., Jain, N., Pan, M., Deighan, P. and Rubin, E.J., 2014. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proceedings of the National Academy of Sciences, 111(3), pp.1132-1137.

Jiang, H., Zhang, Y., Sun, J., Wang, W. and Gu, Z., 2008. Differential selection on gene translation efficiency between the filamentous fungus Ashbya gossypii and yeasts. BMC evolutionary biology, 8(1), pp.1-9.

Johnson, L.K., Sahasrabudhe, R., Gill, J.A., Roach, J.L., Froenicke, L., Brown, C.T. and Whitehead, A., 2020. Draft genome assemblies using sequencing reads from Oxford Nanopore Technology and Illumina platforms for four species of North American Fundulus killifish. GigaScience, 9(6), p.giaa067.

Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T., 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene, 238(1), pp.143-155.

Kang, T. W., Yevsa, T., Woller, N., Hoenicke, L., Wuestefeld, T., Dauch, D., Hohmeyer, A., Gereke, M., Rudalska, R., Potapova, A., Iken, M., Vucur, M., Weiss, S., Heikenwalder, M., Khan, S., Gil, J., Bruder, D., Manns, M., Schirmacher, P., Tacke, F., … Zender, L. 2011. Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. Nature, 479(7374), 547–551.

Keith, G., 1995. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. Biochimie, 77(1-2), pp.142-144.

Kirchner, S. and Ignatova, Z., 2015. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nature Reviews Genetics, 16(2), pp.98-112.

Kuilman, T., Michaloglou, C., Mooi, W.J. and Peeper, D.S., 2010. The essence of senescence. 249. Genes Dev, 24(2463-2479), p.250.

Lama, L., Cobo, J., Buenaventura, D. and Ryan, K., 2019. Small RNA-seq: The RNA 5'-end adapter ligation problem and how to circumvent it. Journal of biological methods, 6(1).

Lee SR, Collins K. 2005. Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. J Biol Chem 280: 42744–42749.

Li, L., Boniecki, M.T., Jaffe, J.D., Imai, B.S., Yau, P.M., Luthey-Schulten, Z.A. and Martinis, S.A., 2011. Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites. Proceedings of the National Academy of Sciences, 108(23), pp.9378-9383.

Li de la Sierra-Gallay I, Mathy N, Pellegrini O, Condon C. 2006. Structure of the ubiquitous 39 processing enzyme RNase Z bound to transfer RNA. Nat Struct Mol Biol 13: 376–377.

Liu, X., Zheng, Q., Vrettos, N., Maragkakis, M., Alexiou, P., Gregory, B.D. and Mourelatos, Z., 2014. A MicroRNA precursor surveillance system in quality control of MicroRNA synthesis. Molecular cell, 55(6), pp.868-879.

Lodish, H., Berk, A., Kaiser, C.A., Kaiser, C., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H. and Matsudaira, P., 2008. Molecular cell biology. Macmillan.

Lowe, T.M. and Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic acids research, 44(W1), pp.W54-W57.

Love, M., Anders, S. and Huber, W., 2014. Differential analysis of count data–the DESeq2 package. Genome Biol, 15(550), pp.10-1186.

Lu, H., Giordano, F. and Ning, Z., 2016. Oxford Nanopore MinION sequencing and genome assembly. Genomics, proteomics & bioinformatics, 14(5), pp.265-279.

Lu, W.Y., Bird, T.G., Boulter, L., Tsuchiya, A., Cole, A.M., Hay, T., Guest, R.V., Wojtacha, D., Man, T.Y., Mackinnon, A. and Ridgway, R.A., 2015. Hepatic progenitor cells of biliary origin with liver repopulation capacity. Nature cell biology, 17(8), pp.971-983.

Lucas, C.L., Chandra, A., Nejentsev, S., Condliffe, A.M. and Okkenhaug, K., 2016. PI3Kδ and primary immunodeficiencies. Nature Reviews Immunology, 16(11), pp.702-714.

Man, O. and Pilpel, Y., 2007. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. Nature genetics, 39(3), pp.415-421.

Manley, L.J., Ma, D. and Levine, S.S., 2016. Monitoring error rates in Illumina sequencing. Journal of biomolecular techniques: JBT, 27(4), p.125.

Maraia, R.J. and Arimbasseri, A.G., 2017. Factors that shape eukaryotic tRNAomes: processing, modification and anticodon–codon use. Biomolecules, 7(1), p.26.

Meng, X., Franklin, D.A., Dong, J. and Zhang, Y., 2014. MDM2–p53 pathway in hepatocellular carcinoma. Cancer research, 74(24), pp.7161-7167.

Michaloglou, C., Vredeveld, L.C., Soengas, M.S., Denoyelle, C., Kuilman, T., Van Der Horst, C.M., Majoor, D.M., Shay, J.W., Mooi, W.J. and Peeper, D.S., 2005. BRAF E600-associated senescence-like cell cycle arrest of human naevi. Nature, 436(7051), pp.720-724.

Mishima Y, Tomari Y. Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol Cell. 2016 Mar 17;61(6):874–85.

Moir RD, Lee J, Haeusler RA, Desai N, Engelke DR, Willis IM. 2006. Protein kinase A regulates RNA polymerase III transcription through the nuclear localization of Maf1. Proc Natl Acad Sci 103: 15044–15049.

Mohan A, Whyte S, Wang X, Nashimoto M, Levinger L. 1999. The 39 end CCA of mature tRNA is an antideterminant for eukaryotic 39-tRNase. RNA 5: 245–256.

Morita, M., Gravel, S.P., Chénard, V., Sikström, K., Zheng, L., Alain, T., Gandin, V., Avizonis, D., Arguello, M., Zakaria, C. and McLaughlan, S., 2013. mTORC1 controls mitochondrial activity and biogenesis through 4E-BP-dependent translational regulation. Cell metabolism, 18(5), pp.698-711.

Motorin, Y., Muller, S., Behm-Ansmant, I. and Branlant, C., 2007. Identification of modified residues in RNAs by reverse transcription-based methods. Methods in enzymology, 425, pp.21-53.

Murawski M, Szczesniak B, Zoladek T, Hopper AK, Martin NC, Boguta M. 1994. maf1 mutation alters the subcellular localization of the Mod5 protein in yeast. Acta Biochim Pol 41: 441–448.

Nawrocki, E.P. and Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics, 29(22), pp.2933-2935.

Nusse, R. and Clevers, H., 2017. Wnt/β-catenin signaling, disease, and emerging therapeutic modalities. Cell, 169(6), pp.985-999.

O'Brien, E.P., Ciryam, P., Vendruscolo, M. and Dobson, C.M., 2014. Understanding the influence of codon translation rates on cotranslational protein folding. Accounts of chemical research, 47(5), pp.1536-1544.

Oficjalska-Pham D, Harismendy O, Smagowicz WJ, Gonzalez de Peredo A, Boguta M, Sentenac A, Lefebvre O. 2006. General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. Mol Cell 22: 623–632.

Pan, T., 2018. Modifications and functional genomics of human transfer RNA. Cell research, 28(4), pp.395-404.

Pape, T., Wintermeyer, W. and Rodnina, M., 1999. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. The EMBO journal, 18(13), pp.3800-3807.

Parisien, M., Wang, X. and Pan, T., 2013. Diversity of human tRNA genes from the 1000-genomes project. RNA biology, 10(12), pp.1853-1867.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods, 14(4), pp.417-419.

Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L.Z., Pass, H.I., Kipps, T.J., Liu, C.G. and Croce, C.M., 2016. Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. Proceedings of the National Academy of Sciences, 113(18), pp.5071-5076.

Percudani, R., Pavesi, A. and Ottonello, S., 1997. Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. Journal of molecular biology, 268(2), pp.322-330.

Phizicky, E.M. and Hopper, A.K., 2010. tRNA biology charges to the front. Genes & development, 24(17), pp.1832-1860.

Piccirillo, C.A., Bjur, E., Topisirovic, I., Sonenberg, N. and Larsson, O., 2014. Translational control of immune responses: from transcripts to translatomes. Nature immunology, 15(6), pp.503-511.

Pinkard, O., McFarland, S., Sweet, T. and Coller, J., 2020. Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. Nature communications, 11(1), pp.1-15.

Plotkin JB, Robins H, Levine AJ. Tissue-specific codon usage and the expression of human genes. Proc Natl Acad Sci U S A . 2004 Aug 24;101(34):12588–91.

Pluta K, Lefebvre O, Martin NC, Smagowicz WJ, Stanford DR, Ellis SR, Hopper AK, Sentenac A, Boguta M. 2001. Maf1p, a negative effector of RNA polymerase III in Saccharomyces cerevisiae. Mol Cell Biol 21: 5031–5040.

Popow, J., Schleiffer, A. and Martinez, J., 2012. Diversity and roles of (t) RNA ligases. Cellular and Molecular Life Sciences, 69(16), pp.2657-2670.

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015 Mar 15;160(6):1111–24.

Paushkin, S.V., Patel, M., Furia, B.S., Peltz, S.W. and Trotta, C.R., 2004. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3′ end formation. Cell, 117(3), pp.311-321.

Radhakrishnan A, Chen Y, Martin S, Alhusaini N, Green R, Coller J, et al. The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality Article The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell . 2016;167(1):122-128.e9.

Rak, R., Dahan, O. and Pilpel, Y., 2018. Repertoires of tRNAs: the couplers of genomics and proteomics. Annual review of cell and developmental biology, 34, pp.239-264.

Ramanathan, A., Robb, G.B. and Chan, S.H., 2016. mRNA capping: biological functions and applications. Nucleic acids research, 44(16), pp.7511-7526.

Reed, K.R., Athineos, D., Meniel, V.S., Wilkins, J.A., Ridgway, R.A., Burke, Z.D., Muncan, V., Clarke, A.R. and Sansom, O.J., 2008. B-catenin deficiency, but not Myc deletion, suppresses the immediate phenotypes of APC loss in the liver. Proceedings of the National Academy of Sciences, 105(48), pp.18919-18923.

Reina JH, Azzouz TN, Hernandez N. 2006. Maf1, a new player in the regulation of human RNA polymerase III transcription. PLoS ONE 1: e134. doi: 10.1371/journal.pone.0000134.

Rich, A. and RajBhandary, U.L., 1976. Transfer RNA: molecular structure, sequence, and properties. Annual review of biochemistry, 45(1), pp.805-860.

Richter, J.D. and Sonenberg, N., 2005. Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature, 433(7025), pp.477-480.

Robichaud, N., Sonenberg, N., Ruggero, D. and Schneider, R.J., 2019. Translational control in cancer. Cold Spring Harbor perspectives in biology, 11(7), p.a032896.

Rocha EPC. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. Genome Res . 2004 Nov;14(11):2279–86.

Rodnina, M.V. and Wintermeyer, W., 2009. Recent mechanistic insights into eukaryotic ribosomes. Current opinion in cell biology, 21(3), pp.435-443.

Rolfe, D.F. and Brown, G.C., 1997. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. Physiological reviews, 77(3), pp.731-758.

Rollins J, Veras I, Cabarcas S, Willis I, Schramm L. 2007. Human Maf1 negatively regulates RNA polymerase III transcription via the TFIIB family members Brf1 and Brf2. Int J Biol Sci 3: 292–302.

Saikia, M., Fu, Y., Pavon-Eternod, M., He, C. and Pan, T., 2010. Genome-wide analysis of N1-methyl-adenosine modification in human tRNAs. Rna, 16(7), pp.1317-1327.

Sansom, O.J., Meniel, V.S., Muncan, V., Phesse, T.J., Wilkins, J.A., Reed, K.R., Vass, J.K., Athineos, D., Clevers, H. and Clarke, A.R., 2007. Myc deletion rescues Apc deficiency in the small intestine. Nature, 446(7136), pp.676-679.

Schimmel, P., 2018. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nature reviews Molecular cell biology, 19(1), pp.45-58.

Schon, E.A., DiMauro, S. and Hirano, M., 2012. Human mitochondrial DNA: roles of inherited and somatic mutations. Nature Reviews Genetics, 13(12), pp.878-890.

Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F. and Calatayud, A.L., 2015. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nature genetics, 47(5), pp.505-511.

Schwanhüusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011 May 19;473(7347):337–42.

Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986 Jul 11;14(13):5125–43.

Sharp, P.M. and Li, W.H., 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic acids research, 15(3), pp.1281-1295.

Sigler, P.B., 1975. An analysis of the structure of tRNA. Annual review of biophysics and bioengineering, 4(1), pp.477-527.

Silvera, D., Formenti, S.C. and Schneider, R.J., 2010. Translational control in cancer. Nature Reviews Cancer, 10(4), pp.254-266.

Sonenberg, N. and Hinnebusch, A.G., 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell, 136(4), pp.731-745.

Sørensen, M.A., Kurland, C.G. and Pedersen, S., 1989. Codon usage determines translation rate in Escherichia coli. Journal of molecular biology, 207(2), pp.365-377.

Stadler, M. and Fire, A., 2011. Wobble base-pairing slows in vivo translation elongation in metazoans. Rna, 17(12), pp.2063-2073.

Su, D., Chan, C.T., Gu, C., Lim, K.S., Chionh, Y.H., McBee, M.E., Russell, B.S., Babu, I.R., Begley, T.J. and Dedon, P.C., 2014. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. Nature protocols, 9(4), pp.828-841.

Sullivan, K.D., Galbraith, M.D., Andrysik, Z. and Espinosa, J.M., 2018. Mechanisms of transcriptional regulation by p53. Cell Death & Differentiation, 25(1), pp.133-143.

Suzuki, T., Nagao, A. and Suzuki, T., 2011. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annual review of genetics, 45, pp.299-329.

Tarrant, D. and Von Der Haar, T., 2014. Synonymous codons, ribosome speed, and eukaryotic gene expression regulation. Cellular and Molecular Life Sciences, 71(21), pp.4195-4206.

Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, et al. 2001. A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 27: 172–180.

Torres, A.G., Batlle, E. and de Pouplana, L.R., 2014. Role of tRNA modifications in human diseases. Trends in molecular medicine, 20(6), pp.306-314.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. and Sedgwick, B., 2002. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature, 419(6903), pp.174-178.

Truitt, M.L. and Ruggero, D., 2016. New frontiers in translational control of the cancer genome. Nature Reviews Cancer, 16(5), pp.288-304.

Vogel A, Schilling O, Spath B, Marchfelder A. 2005. The tRNase Z family of proteins: Physiological functions, substrate specificity and structural properties. Biol Chem 386: 1253–1264.

Waldman YY, Tuller T, Shlomi T, Sharan R, Ruppin E. Translation efficiency in humans: Tissue specificity, global optimization and differences between developmental stages. Nucleic Acids Res . 2010 Jan 21;38(9):2964–74.

Walker SC, Engelke DR. 2006. Ribonuclease P: The evolution of an ancient RNA enzyme. Crit Rev Biochem Mol Biol 41: 77–102.

Walsh, D., Arias, C., Perez, C., Halladin, D., Escandon, M., Ueda, T., Watanabe-Fukunaga, R., Fukunaga, R. and Mohr, I., 2008. Eukaryotic translation initiation factor 4F architectural alterations accompany translation initiation factor redistribution in poxvirus-infected cells. Molecular and cellular biology, 28(8), pp.2648-2658.

Wei Y, Tsang CK, Zheng XF. 2009. Mechanisms of regulation of RNA polymerase III-dependent transcription by TORC1. EMBO J 28: 2220–2230.

Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. Vol. 45, Nature Genetics. Nature Publishing Group; 2013. p. 1113–20.

Wellner, K., Betat, H. and Mörl, M., 2018. A tRNA's fate is decided at its 3' end: collaborative actions of CCA-adding enzyme and RNases involved in tRNA processing and degradation. Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1861(4), pp.433-441.

Wu Q, Medina SG, Kushawah G, Devore ML, Castellano LA, Hand JM, et al. Translation affects mRNA stability in a codon-dependent manner in human cells. Elife . 2019 Apr 1;8.

Xu Y, Ma P, Shah P, Rokas A, Liu Y, Johnson CH. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature . 2013;495(7439):116–20.

Yim, S.Y., Shim, J.J., Shin, J.H., Jeong, Y.S., Kang, S.H., Kim, S.B., Eun, Y.G., Lee, D.J., Conner, E.A., Factor, V.M. and Moore, D.D., 2018. Integrated genomic comparison of mouse models reveals their clinical resemblance to human liver cancer. Molecular Cancer Research, 16(11), pp.1713-1723.

Zaborske, J.M., Bauer DuMont, V.L., Wallace, E.W., Pan, T., Aquadro, C.F. and Drummond, D.A., 2014. A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. PLoS biology, 12(12), p.e1002015.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M. and Pan, T., 2015. Efficient and quantitative high-throughput tRNA sequencing. Nature methods, 12(9), pp.835-837.

Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature . 2013 Mar 7;494(7439):111–5.

Zhu A, Ibrahim JG, Love MI (2018). “Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences.” Bioinformatics. doi: 10.1093/bioinformatics/bty895

Ziehler, W.A., Day, J.J., Fierke, C.A. and Engelke, D.R., 2000. Effects of 5 ‘Leader and 3 ‘Trailer Structures on Pre-tRNA Processing by Nuclear RNase P. Biochemistry, 39(32), pp.9909-9916.

Zucman-Rossi, J., Villanueva, A., Nault, J.C. and Llovet, J.M., 2015. Genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology, 149(5), pp.1226-1239.