UNIVERSITY OF CALIFORNIA
SANTA CRUZ

INVESTIGATION OF ALTERNATIVE mRNA TRANSCRIPT FATE

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

MOLECULAR, CELL, AND DEVELOPMENTAL BIOLOGY

by

Jolene M. Draper

December 2021

The Dissertation of Jolene M. Draper is
approved:

_____
Professor Jeremy Sanford, chair

_____
Distinguished Professor Manuel Ares, Jr.

_____
Professor Alan Zahler

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

Thesis Abstract v

Acknowledgments vii

Chapter 1: Introduction 1

Chapter 2: Isoform-specific translational control is evolutionarily conserved in primates 19

Chapter 3: A method for campus-wide SARS-CoV-2 surveillance at a large public university 42

Appendix: Nanopore sequencing reveals whole mRNA isoform structures engaged with translational machinery 62

Bibliography 71

List of Figures

Chapter 1: Figure 1 13

Chapter 2: Figure 1 24

Chapter 2: Figure 2 26

Chapter 2: Figure 3 28

Chapter 2: Figure 4 30

Chapter 2: Figure 5 31

Chapter 2: Figure 6 33

Chapter 3: Figure 1 47

Chapter 3: Figure 2 49

Chapter 3: Table 1 50

Chapter 3: Figure 3 52

Chapter 3: Figure 4 54

Chapter 3: Figure 5 56

Appendix: Figure 1 63

Appendix: Figure 2 64

Appendix: Figure 3 67

Appendix: Figure 4 68

Appendix: Figure 5 72

Abstract: Investigation of Alternative mRNA Transcript Fate

Jolene Draper

Eukaryotic gene expression from DNA to protein requires multiple RNA maturation steps including 5′ end capping, splicing, and 3′ end cleavage and polyadenylation. In most eukaryotic organisms, genes produce multiple mRNA isoforms by way of alternative splicing, alternative 3′ end maturation, and use of alternative transcriptional start sites. Alternative splicing in particular is a powerful mechanism for expanding gene function through both the protein coding capacity and regulatory potential of individual mRNA molecules. RNA binding proteins (RBPs) recruited to the mRNA during or as a consequence of splicing couple the nuclear history of the message to its cytoplasmic fate. In this thesis I explore how alternative mRNA processing and differential mRNP composition influence gene expression.

Simple measures of steady-state RNA levels by RNA-Seq do not always correlate with ultimate protein output. In some tissues, and for different classes of genes, the relationship appears to be under developmental control, suggesting that mRNA is being repressed at the translational level. Using mass spectrometry to detect relative expression of two protein isoforms from the same gene is extremely difficult since the alternative protein sequence is a small part of the entire protein. To address this problem we assess global mRNA isoform-specific ribosome association that preserves the entirety of the message (Frac-Seq) as a direct and independent signal of the extent to which each isoform engages with translational machinery. We explore alternative isoform ribosome association in two contexts: 1) in a comparative transcriptomics study, where we are able to leverage millions of years of random “experiments” in the form of evolutionary changes in primate genome sequence; and 2) in inchoate human neural differentiation, where alternative splicing is abundant and RNA-protein correlation is particularly low. All

together, these investigations will help us understand how alternative mRNA processing is important for accurate and efficient gene expression.

In addition, in the middle of March 2020, the COVID-19 pandemic suddenly shut down our ability to do research. Uncertainty has been a major theme since then. A leading contributor to personal and public uncertainties was the lack of reliable testing for SARS-CoV-2. A group of UCSC students, staff, and faculty realized we had the technical knowledge to help keep our university community safe. Under the FDA's Emergency Use Authorization (EUA) orders and our Student Health Center's clinical laboratory credentials, we implemented a molecular biology-based SARS-Cov-2 diagnostic test and a campus-wide COVID-19 surveillance program. This initiative has been instrumental in breaking SARS-Cov-2 transmission chains on campus and in the broader community and had an immediate impact on human health and safety. A roadmap for establishing a “pop-up” clinical molecular biology-based diagnostics laboratory is included as a chapter.

Acknowledgements

The text of this dissertation includes the following material accepted for publication: Chang T*, Draper JM*, Van den Bout A*, Kephart E*, Maul-Newby H*, Vasquez Y*, Woodbury J, Randi S, Pedersen M, Nave M, La S, Gallagher N, McCabe MM, Dhillon N, Bjork I, Luttrell M, Dang F, MacMillan JB, Green R, Miller E, Kilpatrick A, Vaske O, Stone MD, and Sanford JR. A method for campus-wide SARS-CoV-2 surveillance at a large public university. PLoS One.

The text of this dissertation includes the following material in preparation for publication: Draper JM*, Philipp J*, Dargyte M, Wallace A, Katzman S, Miga K, Salama S, Haussler D, and Sanford JR. Isoform-specific translational control is evolutionarily conserved in primates.

The co-authors listed in these publications directed and supervised the research which forms the basis for the dissertation.

Chapter 1: Introduction

While the central dogma of molecular biology posits that genetic information flows from DNA to RNA to protein, the *regulation* of that flow of information is determined by complex and combinatorial interactions between these three molecular biopolymers. Gene expression - and ultimately protein production - is heavily influenced by RNA binding proteins (RBPs) and non-coding RNAs (ncRNA) that determine the cellular fate of the message. These *trans*-acting factors recognize and bind to distinct *cis*-elements (structures or sequences) in their messenger RNA (mRNA) targets, forming messenger ribonucleoprotein complexes (mRNPs). mRNP composition is fluid in the lifespan of an mRNA and is determined by both primary mRNA sequence and cellular concentration of mRNAs, RBPs, and ncRNAs.

mRNA biogenesis

Protein coding genes, the basic units of heredity, occupy mostly non-overlapping regions of DNA (Nakayama et al., 2007). Their expression is activated or attenuated by changes in chromatin structure at promoter and enhancer regions, interactions with proteins and RNA, and chemical modifications of those molecules. Messenger RNA (mRNA) synthesis is facilitated by RNA polymerase II in the nucleus, making a single-stranded, antiparallel RNA copy of the DNA.

mRNA maturation

mRNA maturation, including alteration of the ends of the transcript and excision of introns (splicing, discussed below), begins early during transcription while the nascent mRNA is still engaged with chromatin and transcription machinery (Beyer & Osheim, 1988) (Pandya-Jones & Black, 2009) (Herzel & Neugebauer, 2015). The initial nucleotide

is modified with a 7-methylguanylate (m7G) cap, which later facilitates nuclear export and translation (Shatkin, 1985) (Salditt-Georgieff et al., 1980) (Izaurralde et al., 1995). In vertebrates, cleavage and polyadenylation (CPA) of the end of the transcript after an AAUAAA sequence is facilitated by a protein complex called cleavage and polyadenylation specificity factor, CPSF, and requires the presence of the polynucleotide adenylyltransferase, also known as poly(A) polymerase, or PAP (Beaudoing et al., 2000) (Mandel et al., 2008) (Laishram & Anderson, 2010). Downstream GU-rich elements recruit trans-factors and increase CPA efficiency (Wahle & Keller, 1996) (Colgan & Manley, 1997) (Pérez Cañadillas & Varani, 2003). CPA aids in transcription termination and degradation of the trailing RNA downstream of the cleavage site (Dichtl et al., 2002). The addition of a homopolymer string of adenosine nucleotides to the terminus (polyadenylation) and binding of poly(A)-binding protein (PABP) protects the mRNA from premature decay and promotes stability and translation of the molecule (Bernstein et al., 1989) (Ford et al., 1997) (Craig et al., 1998) (Z. Wang et al., 1999).

mRNA splicing

For the majority of eukaryotic genes, nascent mRNAs have long stretches of nucleic acids interspersed within the message (introns) that need to be removed. The remaining sequences (exons) consist of coding information for the final protein product (open reading frame, ORF) and flanking nucleic acid sequences that contain regulatory information and will not be translated (untranslated regions, UTRs). In a well-characterized sequence of two isoenergetic transesterification reactions, the spliceosome cuts out intronic RNA while covalently joining exons together.

The spliceosome, in concert with RBPs known as splicing factors, identifies exon-intron border signals (splice sites) via RNA-RNA and RNA-protein interactions. The spliceosome is a large, dynamic complex composed of uridine-rich small nucleolar RNAs (U snRNAs) and RBPs. Base pairing of a portion of the U1 snRNA to the 5′ splice site marks the upstream exon-intron boundary (Stark et al., 2001) (Nagai et al., 2001). At the

opposite end of the intron, U2 Small Nuclear RNA Auxiliary Factor 1 (U2AF1) binds to the 3′ splice site and U2AF2 binds to a conserved, proximal upstream polypyrimidine tract (Zamore et al., 1992) (K. Zhang et al., 1992). U2AF binding facilitates ATP-dependent, imperfect base pairing of U2 within the U2 snRNP to the sequence around an adenosine ~20-40 nucleotides upstream (Parker et al., 1987) (R. Singh et al., 1995) (Valcárcel et al., 1996) (Gozani et al., 1996) (J. Wu & Manley, 1989). This adenosine, known as the branch point adenosine, is unpaired and bulges out of the double stranded RNA (Yan & Ares, 1996) (Pascolo & Séraphin, 1997). The tri-snRNP, containing basepaired U4/U6, U5, and many proteins, joins the splicing machinery bound to the mRNA, displacing U2AF, U1, and U4 (Johnson & Abelson, 2001) (Sander et al., 2006). The upstream region of U2 base pairs with U6, and the end of a stem loop in U5 interacts with both exons at the future exon-exon boundary (J. A. Wu & Manley, 1991) (Newman, 1997). The ATP-dependent remodeling of this complex alters protein organization and positions the mRNA for the first step of splicing (Raghunathan & Guthrie, 1998) (Laggerbauer et al., 1998) (X. Zhang et al., 2018). Here, the 2′ hydroxyl of the bulged branch point adenosine base attacks the phosphate at the 5′ end of the intron (Konarska et al., 1985) (Steitz & Steitz, 1993) (Chin & Pyle, 1995). This creates a 2′-5′ phosphodiester linkage at the branchpoint, forming an intron lariat and releases the 5′ exon, though it remains bound to the spliceosome (Konarska et al., 1985).

In the second catalytic step, the free 3′ hydroxyl of the detached 5′ exon attacks the phosphate at the 3′ end of the intron, resulting in release of the intron lariat and ligation of exons (Anderson & Moore, 1997). During the second step of splicing, proteins known as the exon junction complex (EJC) are deposited upstream of the newly ligated exons; these proteins remain with the mRNA as it is exported to the cytoplasm and act as a scaffold for protein binding (Z. Zhang & Krainer, 2007) (Hervé Le Hir et al., 2016). Coding sequences and 5′UTRs frequently contain splice sites, whereas 3′UTRs are rarely spliced (Cenik et al., 2010) (Bicknell et al., 2012), in part due to the requirements for nonsense-mediated decay (NMD, see below). The net result of these nuclear processing

steps is a single mature mRNA molecule, coated in RNA binding proteins, with the signals required for nuclear export and varied cytoplasmic processes.

mRNA export and localization

Upon their assembly in the nucleus, mRNPs must travel across the nuclear envelope through the nuclear pore complex in order to be translated by ribosomes. Once they enter the cytoplasm, mRNPs encounter effector proteins of their cytoplasmic fate(s). mRNPs that are subject to spatial or temporal control of expression will bind proteins that direct active or passive transport to subcellular regions. Remodeled mRNPs containing proteins with highly disordered domains can sequester mRNAs within membraneless, phase-separated granules that are translationally- or metabolically-regulated (Kedersha et al., 2000) (di Penta et al., 2009) (Liu-Yesucevitz et al., 2014) (Lin et al., 2015).

The earliest observations of mRNA localization as a means to control gene expression were made in Ascidia, Xenopus, and Drosophila embryos and ova, where developmentally essential mRNAs concentrate in particular regions of the cell, and their mislocalization leads to improper development of the organism . In large cell types such as neurons, long-distance mRNA transport to the axon tip is an efficient way to concentrate factors necessary for cell-cell signalling at the end of the axon via localized protein synthesis (Steward & Levy, 1982) (Garner et al., 1988) (Racca et al., 1997). Disruption of mRNA localization leads to improper development or cell function, highlighting the importance of regulation of mRNP composition.

Protein synthesis

In the cytoplasm, mRNAs may engage with the ribosome, the catalyst of polypeptide synthesis, which translates mRNA into protein For review: (Sonenberg & Hinnebusch, 2009) (Jackson et al., 2010). The ribosome reads the mRNA three nucleotides at a time (codon), elongating the protein chain one amino acid at a time, via

introduction of cognate aminoacylated transfer RNA (aa-tRNA) into the ribosome (Hoagland et al., 1958) (Crick et al., 1961) (Nirenberg et al., 1965). Protein synthesis stops when the ribosome encounters a termination codon (UAA, UAG, or UGA), which is recognized by a release factor (Capecchi, 1967). Though there are important mechanisms for translational control during elongation and termination, regulation of translation is largely governed by translation initiation (Varenne et al., 1984) (Shah et al., 2013) (Arribere & Fire, 2018) (Guimaraes et al., 2020).

Cap-dependent mRNA translation

In eukaryotic, cap-dependent translation, the translation initiation factor eIF4E binds to the m7G cap at the beginning of the message. eIF4G binds both eIF4E and PABP on the poly(A) tail, effectively bringing the mRNA into a loop within the mRNP and activating it for translation (Borman et al., 2000). eIF4G can also enhance the helicase of eIF4A, which unwinds intramolecular base pairing in the 5′UTR (Abramson et al., 1988) (Rogers et al., 1999); (Rogers et al., 2001). An initiator aa-tRNA competent for decoding a start codon bound to eIF2-GTP (together, ternary complex) joins the small (40S) ribosomal subunit and other initiation factors to form the 43S preinitiation complex (43S PIC, so named for its sedimentation coefficient, a factor dependent on shape, density, and mass) (Asano et al., 2000) (Kapp & Lorsch, 2004) (C. R. Singh et al., 2004) (Shin et al., 2011). The 43S PIC recognizes the activated mRNP and scans the mRNA for a start codon (AUG) in the proper nucleotide context (Kozak sequence), the signal for beginning the nascent polypeptide (Kozak, 1991) (Tatyana V. Pestova & Kolupaeva, 2002). Here, hydrolysis of eIF2-GTP to eIF2-GDP leads to displacement of many of the initiation factors, addition of the large (60S) ribosomal subunit, and formation of the final, translationally-competent complex of the 80S ribosome (Unbehaun et al., 2004) (Algire et al., 2005).

An mRNP containing a single elongating ribosome, monosome, is often found during the initial, or pioneer, round of translation (Heyer & Moore, 2016). Many ribosomes

can sequentially engage with an mRNA at the start codon, forming a polysome mRNP, allowing for higher protein output from a single mRNA molecule. Due to the interaction of the end and beginning of the mRNA, bridged by eIFs and PABP, it is possible for the small subunits of the ribosome to immediately reinitiate at the start codon (Pisarev et al., 2007).

Cap-independent translation

Observations that some uncapped viral mRNAs were efficiently translated in cells led to the discovery of cap-independent translation (Pelletier et al., 1988) Jang, 1988; (Elroy-Stein et al., 1989). Here, RNA secondary structures called internal ribosome entry sites (IRESs) promoted ribosome engagement via interaction directly with the 40S ribosomal subunit or via association with IRES trans-acting factors (ITAFs) (T. V. Pestova et al., 1996) (Roberts et al., 1998). For some of these genes, scanning of the 5'UTR is unnecessary and initiator tRNAs can aid assembly of the ribosomal subunits directly at a start codon independently of the ternary complex and 43S PIC (Ji et al., 2004). In some cases, ribosome assembly can occur at a noncanonical start codons with noncanonical initiator tRNAs (Vagner et al., 1995) (Takahashi et al., 2005). Cap-independent translation enhancers (CITEs) within 3'UTRs have also been identified in viruses and in large-scale screens of mRNA sequence elements Yang, 2021 BioRxiv. These structures are also thought to interact with eIFs and ribosomal proteins in a cap-independent manner. The identification of cellular cis-elements and the mechanisms of cap-independent translation initiation are still being investigated and may eventually aid in prediction of mRNA fate.

Translational control

Exquisite translational control is important for maintaining cell health and proper cellular functions. The overall amount of protein that is produced from one mRNA molecule is affected by several factors. Non-optimal codon composition, RNA structure,

and certain nucleotide modifications within the ORF can negatively affect elongation rate of a translating ribosome, which in turn reduces protein output from those messages. AUG (or rarely CUG, UUG, GUG) codons upstream of the correct AUG and Kozak sequence can initiate translation before the main ORF (uORFs) . uORFs often dampen downstream translation, depending on their relative position to the canonical start codon, though uORFs also have the potential to increase downstream translation in specific contexts (Pöyry et al., 2004). Small non-coding microRNAs can act as translational repressors when imperfectly base paired to regulatory elements in 3'UTRs For review: (Behm-Ansmant et al., 2006). Here, the microRNA directs binding of the RNA-induced silencing complex (RISC), which interferes with efficient engagement of translational machinery with the mRNP (van den Berg et al., 2008). Alternative last exons (ALEs) or tandem UTRs remodel the regulatory information in the 3'UTR, thus alternative isoforms may have different sensitivities to translational repression by RISC and microRNAs (discussed below).

Translation can be controlled on a global scale in response to cell state, primarily by modification of eIFs. The kinase mTOR senses environmental stimuli (e.g., nutrients, oxygen, ATP) and phosphorylates eIF4e-binding protein (4E-BP) which releases eIF4e to promote cap binding and mRNA translation activation. The integrated stress response (ISR) detects signals of cellular stress (e.g., nutrient deprivation, abnormal protein homeostasis, viral infection) and in turn phosphorylates eIF2α, preventing assembly of the ternary complex (Harding et al., 2003) (Hinnebusch, 2005). Messages with cap-independent translational machinery are then more likely to engage with ribosomes; indeed, many stress response genes contain IRESs and CITEs for preferential translation during stress conditions.

Protein production from specific messages, or classes of messages, is tuned by binding of RBPs to cis-elements. eIF3 is important for baseline translation, however it also binds directly to RNA structures within 5'UTRs of genes important for cell proliferation and modulates their translation independent of eIF3 global function (Lee et

al., 2015). PABP mRNA contains a conserved homopolymer tract of adenosine monophosphates in its 5'UTR. Binding of PABP to the PABP 5'UTR represses translation, thus acting as a translational autoregulatory negative feedback loop to stop PABP production when its concentration is high (de Melo Neto et al., 1995). hnRNP C bound to the X-linked inhibitor of apoptosis (XIAP) IRES element increases cap-independent translation during cell stress, delaying apoptosis (Holcík et al., 2003). Osmotic-shock induces phosphorylation and consequently relocalization of hnRNP A1 to the cytoplasm where it remodels the XIAP mRNP to repress translation (S. M. Lewis et al., 2007). Ultimately, the cis-elements within the mRNA and the availalbe trans-acting factors determine mRNP composition and mRNA fate.

mRNA decay

Unlike their DNA predecessor, mRNAs are unstable and short-lived in the cell. Steady state mRNA are determined by the counteracting processes of transcription and decay. mRNA turnover is important for controlling the amount and type of transcripts in order to meet the current needs of the cell and respond rapidly to external stimuli.

mRNA decay pathways

mRNAs with long half-lives (days) are enriched for housekeeping and translation function; mRNAs encoding factors required for stress response, the cell cycle, and rapidly regulation like cytokines and transcription factors tend to have a much shorter half-lives (minutes to hours) (Yang et al., 2003); (Schwanhäusser et al., 2011). The length of time an mRNA exists in the cytoplasm is determined by intrinsic mRNA elements and the abundance of trans-acting ncRNA and proteins in decay pathways. For example, AU-rich elements are found in the UTRs of many short-lived mRNA transcripts, however, depending on the composition of RBPs within the cell, the transcript may be stabilized or

rapidly degraded (Spasic et al., 2012). For example, binding of HuR stabilizes the transcript, whereas binding of TTP will recruit mRNA degradation machinery (Fan & Steitz, 1998) (Hau et al., 2007). In animals, a powerful mechanism for post-transcriptional gene silencing is microRNA-induced mRNA degradation. Near perfect base pairing of microRNA and mRNA within the RISC complex promotes recruitment exonucleases that cleave the mRNA, deadenylases that degrade the stabilizing polyA tail, and endonucleases, which degrade the mRNA fragments from the ends For Review, (van den Berg et al., 2008).

Nonsense-mediated decay

Transcription of DNA to mRNA is a one-to-one copying of nucleic acids, but protein translation from mRNA is a three nucleotide-to-one amino acid transfer of information. Thus, the introduction of a non-triplet nucleotide sequence can change the reading frame of the ORF, ultimately changing the protein sequence, removing a stop codon, or introducing a premature termination codon (PTC). When considering the evolution of exons, noisy splicing, and alternative splicing, it is likely that the cell will eventually encounter a non-productive or aberrant mRNA. Nonsense-mediated decay (NMD) is a highly conserved, translation-dependent mRNA surveillance mechanism that identifies these mRNA transcripts and targets them for decay (Peltz et al., 1993) (L. E. Maquat, 1995). During the pioneer round of translation, a ribosome translates mRNA that is still coated with EJCs deposited during pre-mRNA splicing. In the event a ribosome encounters a stop codon upstream of an EJC (and thus upstream of a splice site), the stop codon is interpreted as a premature end to the polypeptide For review: (Lynne E. Maquat et al., 2010). (3′UTRs, or the regulatory nucleic acids after the termination codon, are rarely spliced so are very unlikely to be bound to an EJC.) The EJC acts as a scaffold for NMD proteins (UPF1, UPF2, UPF3a and b) which, when post-translationally modified during NMD, trigger remodeling of the mRNP and recruitment of deadenylases and nucleases (Melero et al., 2012). These transcripts thus engage with ribosomes but,

depending on availability of degradation machinery or presence of inhibitors, are rapidly removed from the cytoplasmic pool of mRNAs.

Alternative transcript generation

Each step of transcript generation is highly regulated, often dynamic, and, like many biological processes, subject to noisy execution. Alternative splicing and alternative end generation are powerful mechanisms for changing protein coding and regulatory potential of a transcript. These three processes can occur alone or in combination with one another, yielding myriad potential alternative isoforms from a static genome.

Alternative mRNA 5′-ends

Cap-dependent 5′-end RNA sequencing of 164 human tissues and cell types revealed that over half of human protein coding genes have more than one possible transcriptional start site (TSS) (Kimura et al., 2006). Alternative transcription initiation leads to an alternative first exon (AFE), changing the 5′UTR sequence and possibly replacing the beginning of the ORF. For the majority of genes with multiple TSSs, sequence conservation at and around the primary TSSs was determined to be high, whereas conservation was generally low around nonprimary TSSs, suggesting that many alternative TSSs may be transcriptional noise (Xu et al., 2019). However, identification of new, nonprimary TSSs increases with diversification of tissue types analysed, and expression from alternative TSSs contributes significantly to tissue-specific isoform expression (Reyes & Huber, 2018). Further, genes with alternative TSSs were found to be enriched for intracellular signalling proteins and post-translational protein-modifying enzymes, which effect expeditious responses to cell cycle changes or external stimuli, suggesting a mechanism for swift, inducible regulation of transcription from alternative promoters. In addition to the ability to increase production of mRNA from an alternative

TSS, many examples of post-transcriptional regulation of alternative 5′UTRs exist in the literature (see below).

Alternative mRNA splicing

Over two thirds of human protein coding genes are multi-exonic; analysis of deep RNA sequencing and microarray data show that over 90% of them are capable of alternative splicing in some context (Pan et al., 2008) (E. T. Wang et al., 2008); (Barbosa-Morais et al., 2012); (Merkin et al., 2012). Alternative inclusion or exclusion of exons (skipped exons, SE), use of alternative splice donor or acceptor sites (A5, A3, respectively), and retention of introns (RI) change the primary sequence of the mRNA.

Serine/arginine-rich (SR) proteins were first identified as a family of non-spliceosomal factors found to influence constitutive and alternative splicing of pre-mRNA (Zahler et al., 1992). SRSF1, for example, has been shown to facilitate the first step of constitutive splicing and influence splice site choice in alternative splicing reporters (Ge & Manley, 1990) (Krainer & Maniatis, 1985) (Krainer et al., 1990). They generally bind cis-elements known as splicing enhancers and promote inclusion of nucleotide sequences. Heterogeneous nuclear RNPs (hnRNPs) typically bind splicing silencers and negatively influence inclusion. Other RBPs block splicing signals or recruit spliceosomal machinery. In the case of MBNL1 and PTBP1, positional binding of the RBP around branch points, splice sites, or within exons encourages exon exclusion, whereas binding downstream of an exon promotes inclusion (Witten & Ule, 2011). Some of these trans-factors remain with the mRNA in the cytoplasm in a splicing-dependent manner and influence cellular fate.

Alternative mRNA 3'-ends

Differences in the end of the transcript arise by two main mechanisms, alternative cleavage and polyadenylation (APA) and alternative splicing (AS). In the event that multiple CPA signal sequences exist within the same 3′UTR, differential cleavage creates

isoforms with distinct transcript termini within the same UTR, called tandem UTRs. It is estimated that ~30% of human mRNAs contain APA sites (Beaudoing et al., 2000). Cell state- or tissue-specific 3′UTR shortening of transcripts has been found in proliferating cells, during differentiation, and in cancer development, suggesting a role for APA in regulating gene expression in response to stress or differentiation (Sandberg et al., 2008) (Mayr & Bartel, 2009) (Lianoglou et al., 2013). In some genes, alternative splicing may incorporate an alternative last exon (ALE), complete with stop codon and 3′UTR, upstream of another last exon. In addition to control by alternative splicing factors, ALE inclusion is influenced by APA and expression levels of CPA factors, with higher concentrations promoting inclusion of the upstream proximal ALE (Cooke et al., 1999) (Movassat et al., 2016). Rarely, alternative splicing within the 3′UTR alters the transcript end after the stop codon. Because there is a high concentration of regulatory information in 3′UTRs, alternative 3′-end processing can have a major impact on the cellular fate of the message.

Alternative isoform fate

It is evident that the fate of each transcript, from biogenesis to decay, is heavily dependent on the composition of its mRNP, which in turn is determined by the mRNA primary sequence. Thus, an important question in the field is how mRNA processing contributes to the cell fate of alternative isoforms from one gene.

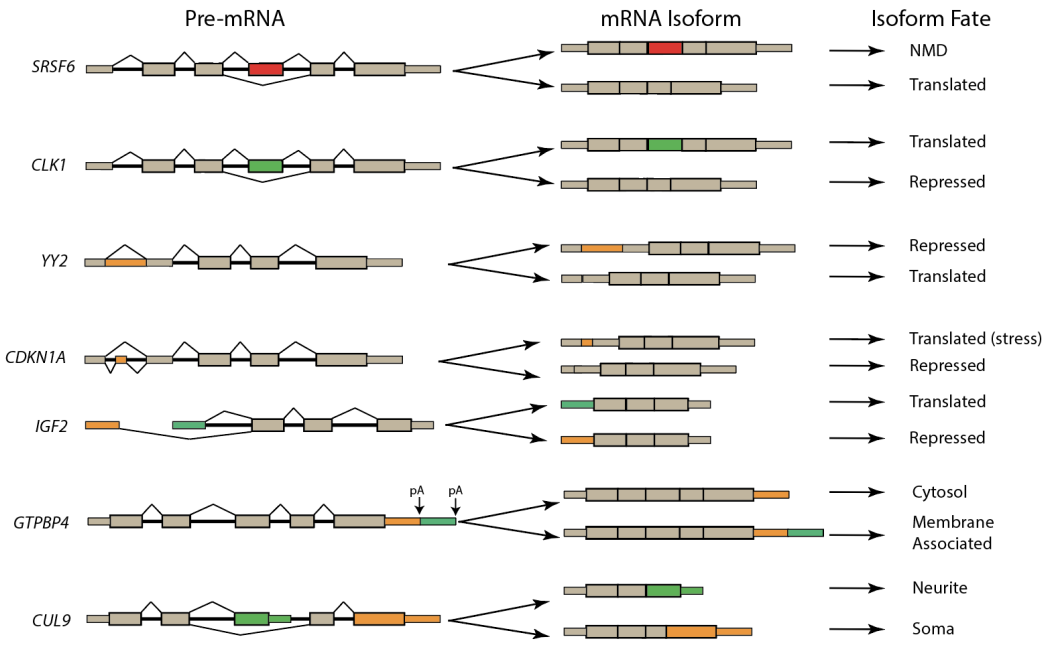

Figure 1: Cartoon diagrams of alternative splicing and the fate of alternative transcripts. See text for references.

Alternative mRNA localization

Alternative mRNA processing can have profound effects on mRNA localization for different isoforms of the same gene. In neurons, messages transported along the cell cytoskeleton to the ends of the neural projections for localized translation were enriched for MBNL binding sites in their 3'UTRs (Taliaferro et al., 2016). When MBNL was depleted, these mRNAs were no longer concentrated within the axon, and neurons exhibited improper cell physiology. Isoforms including their distal alternative last exon were generally enriched for MBNL binding sites and were more often localized to the axon (e.g., CUL9). Conversely, isoforms from the same gene expressing the gene proximal ALE lacked RBP binding sites and did not exhibit active localization within the cell. Similarly, MBNL-binding of long 3'UTR of GTPBP4 in myoblasts is coincident with long isoform localization to the cell membrane (E. T. Wang et al., 2012). APA of GTPBP4

cleaves the 3'UTR upstream of an MBNL binding site; consequently, the short isoform remains diffuse in the cytoplasm.

Alternative splicing coupled to NMD (AS-NMD)

A notable mechanism of gene expression attenuation couples alternative splicing and NMD (AS-NMD) by regulating splicing and inclusion of PTCs within the transcript. AS-NMD in the SR protein family is remarkably prevalent and conserved. For example, the productive isoform for SRSF6 skips exon 3; the ORF remains in frame and is efficiently translated in the cytoplasm. The alternative isoform that includes exon 3 introduces a premature termination codon, which triggers the isoform for NMD, reducing steady state levels of SRSF6 mRNA (Lareau et al., 2007). SRSF2 promotes splicing within its 3'UTR which deposits an EJC after the original termination codon; the ribosome and NMD machinery will interpret the original functional stop codon as a PTC in this context (Sureau et al., 2001). SRSF5 has been shown to autoregulate retention of intron 5, which likewise introduces a PTC into the longer isoform (Lareau & Brenner, 2015). Outside of the SR family, the hnRNP PTBP1, exhibits conserved, tissue-specific skipping of exon 11, which introduces a frameshift in the ORF, creating a downstream PTC (Wollerton et al., 2004). As these splicing factors can influence their own alternative splicing, AS-NMD functions here as an autoregulatory loop. So, while NMD is a mechanism to remove potentially deleterious transcripts, it is also clear that AS-NMD is an important regulatory strategy for certain genes.

Alternative processing coupled to translational control (AS-TC & APA-TC)

The regulatory power of translational control is especially evident in alternative transcripts from a single gene. The insulin-like growth factor II (IGF2) mRNA has several TSSs. When expressed, the leader 1 isoform from the primary TSS is constitutively translated within the cytoplasm. Expression of IRES-containing leader 3 IGF2 mRNA

occurs during development. This isoform is translated in a cap-independent manner, however binding of IGF2 binding protein 3 (IGF2BP3) to the IRES sequence reduces its expression without altering steady state mRNA levels (Pedersen et al., 2002). In murine development, the transcription factor YY2 controls expression of pluripotency and differentiation genes. The 5'UTR of transcription factor YY2 contains a recently inserted alternative intron sequence. Retention of the 117 intronic nucleotides in the alternative isoform increases structure within the 5'UTR (Tahmasebi et al., 2016). During early differentiation, expression of PTBP1 decreases, triggering alternative splicing of the YY2 5'UTR intron. Both the spliced and retained-intron isoform are translated in a cap-dependent manner, however the extra RNA structure of the retained intron renders the RI isoform far more dependent on eIF4-mediated translation activation, thus more sensitive to 4E-BP-mediated translational repression. Thus, protein output from alternative 5'UTR isoforms can be tuned via different mechanisms of translation and the availability of RBPs.

Alternative inclusion of uORFs can also influence translation efficiency of the downstream ORF. In the human gene CDKN1A, inclusion of a skipped exon in the 5'UTR introduces a uORFs with non-canonical start codons (Collier et al., 2018). During UV-induced stress, ISR-mediated phosphorylation of eIF2α prevents ternary complex formation. Subsequently, cap-independent translation from non-canonical start codons in the included isoform enhances expression of the downstream gene, which is necessary for cell cycle regulation and response to UV stress. uORFs often dampen translation from the downstream ORF, as in MDM2 where an alternative leader sequence containing two uORFS leads to underloading of ribosomes on the transcript compared to the alternative leader (Brown et al., 1999).

Changes to the cis-regulatory landscape of an mRNA isoform influences binding of sequence-specific RBPs and non-coding RNAs. Extensive shortening of oncogene 3'UTRs leads to their overexpression in cancer cells (Mayr & Bartel, 2009). Because 3'UTRs are more likely to include negative regulatory information (i.e., microRNA-target

sites), UTR shortening was hypothesized to enhance mRNA stability and translation by removing signals for mRNA decay and translation repression by RISC. Indeed, short isoforms of IGF2BP1, Cyclin D1, and Cyclin D2 showed increased mRNA stability, and short isoforms of IGF2BP1, Cyclin D2, and DICER1 showed alleviation of translational repression compared to long isoforms. Mutation of microRNA binding sites confirmed their role as the cis-elements regulating these observations.

In addition to changes cis-regulatory elements, alternative splicing can change mRNP composition during mRNA processing. For example, alternative association of SRSF1 during SR-mediated AS reveals multifunctional, cell compartment-dependent roles for the splicing factor (Maslon et al., 2014) (Sanford et al., 2004). The activity and localization of many splicing factors, like hnRNPs and SR proteins, depends on the phosphorylation state. Binding of phosphorylated SRSF1 to exon 4 of the SR-protein kinase CLK1 promotes exon 4 inclusion; skipping of this exon results in a truncated, inactive protein isoform (Duncan et al., 1997). Here, hypophosphorylated SRSF1 remains bound to the message through nuclear export and is maintained in the cytoplasmic mRNP. Increased translation of the included isoform was dependent on the level of SRSF1 expression and shuttling ability (Maslon et al., 2014). These data, together with observations that SRSF1 protein is found in polysome fractions and the SRSF1 enhancer cis-elements are sufficient for increased translation of reporter genes (Sanford et al., 2004) (Sanford et al., 2005), support the hypothesis that multifunctional RBPs can couple nuclear processing and translational control of their mRNAs.

Global analysis of translation

Though many individual alternative transcripts are shown to exhibit divergent post-transcriptional cell fates, the consequences of most mRNA isoforms are not yet elucidated. Further, studies evaluating steady state mRNA levels and protein levels

reveal that they are poorly correlated, thus RNA-Seq data are not always a reliable proxy for protein output.

Proteomics

Measuring individual proteins or protein isoforms within the cell is relatively straightforward with benchtop assays if the protein can be tagged or elicits specific antibodies, however the quantification and qualification of the proteome on a global scale by mass spectrometry (MS) has important caveats and technical limitations (Tress et al., 2017) (Fesenko et al., 2017) (Blencowe, 2017). Whole cell MS does not detect whole proteins, rather it relies on digestion of protein by proteases and evaluation of relatively small peptides. Trypsin digests proteins at lysine and arginine amino acid residues and is frequently used for peptide generation. Peptides spanning a constitutive exon and alternative exons would be evidence for alternative protein isoforms. Unfortunately, due to conservation of sequence around splice sites, codons for arginine and lysine residues are overrepresented at exon-exon junctions, thus exon spanning peptides are underrepresented due to the chosen protease (X. Wang et al., 2018). Identification of truncated protein isoforms can also be difficult in the absence of sequence differences or protein-specific, appropriately-sized peptides.

Global analysis of mRNA translation

RNA sequencing is a powerful, straightforward means of measuring steady state levels of RNA though it does not tell the whole story of gene expression. It can, however, be combined with specific biochemical assays to elucidate the story of interest. In this vein, several assays for the sequencing of mRNAs associated with translation machinery are employed as a proxy for translation. The mRNA bound within a ribosome is protected from nucleolytic digestion, thus ribosome footprinting assays that digest unprotected mRNA and sequence ribosome protected RNA fragments (RPFs) elucidate translated

mRNAs Steitz, 1973; (Wolin & Walter, 1988); (Ingolia et al., 2009). The specific size of the protected fragment in ribosome profiling assays indicate the configuration of the ribosome during translation, rate limiting steps, and overall speed of elongation (Ingolia et al., 2011) (Lareau et al., 2014) (Wu et al., 2019). The position of start codons can be empirically determined (with caveats) by treating cells with translation inhibitors that stall initiating ribosomes Ingolia, 2009. RPFs within alternative coding sequences or over alternative exon-exon junctions is evidence supporting alternative-isoform translation (Reixachs-Solé et al., 2020).

Data from ribosome footprinting assays are typically enriched for coding sequences (though RPFs upstream of the expected start codon indicate uORFs or artifacts from drug treatment). What is missing, then, is information about which alternative UTRs, and thus much of the alternative regulatory sequences, are engaged with ribosomes and affect mRNA translation and isoform fate. Polyribosome profiling maintains full-length mRNAs and relies on the inherent ability of larger, denser, heavier mRNPs to sediment faster through a sucrose gradient (Mašek et al., 2011). Here, intact mRNA, replete with RBPs and engaged ribosomes, travel through a sucrose gradient via high-velocity ultracentrifugation and separate depending on the number of ribosomes engaged and the consequently increasing Svedberg unit (S, as in 43S, 80S, etc.). Consecutive fractions of the sucrose gradient are then isolated and sequenced. Because the mRNA is intact, sequencing of polyribosome profiling experiments reveals information about the whole isoform (Sterne-Weiler et al., 2013). Evaluating the relative proportion of the alternative events within bulk cytoplasm, monosome fractions, and separate polyribosome fractions reveals alternative splicing events that may alter ribosome engagement. Differential sedimentation can be interpreted as a proxy for degree of translation (at least ribosome engagement) and can identify alternative events that enrich the message in different fractions.

Chapter 2: Isoform-specific translational control is evolutionarily conserved in primates

Abstract

The process of alternative splicing expands protein-coding capacity and post-transcriptional regulatory mechanisms of many human genes. Here we explore alternative splicing coupled with translational control (AS-TC) across human, chimpanzee, and orangutan induced pluripotent cell lines. Using fractionation-sequencing (Frac-seq), we identified polyribosome-associated mRNA isoforms. We discovered orthologous AS-TC events with either conserved or species-specific translation patterns. Exon sequences associated with conserved sedimentation profiles show strong conservation across vertebrates. Orthologous exons with divergent sedimentation profiles drive species-specific expression of luciferase reporters. These orthologous exons show single nucleotide sequence differences, which significantly alter the strength of RNA binding protein recognition motifs. Together these data show that cis-acting elements regulate AS-TC across primates species.

Introduction

The accurate expression of eukaryotic protein-coding genes requires extensive post-transcriptional processing. These steps include transcript capping, splicing, polyadenylation, nuclear export, localization, translation, and decay. RNA binding proteins (RBPs) and noncoding RNAs play critical roles in maintaining the fidelity and regulation of post-transcriptional gene expression. Together, these factors package protein-coding transcripts into large heterogeneous complexes called messenger ribonucleoprotein complexes (mRNPs) (G. Singh et al., 2015). The RBP composition of these complexes reflect key checkpoints in transcript biogenesis and post-transcriptional regulatory fates (Mitchell & Parker, 2014).

Intervening sequences (introns) are a defining feature of eukaryotic genes. When transcribed into precursor-messenger RNA (pre-mRNA), the process of pre-mRNA splicing assembles the open reading frame by removing introns and joining protein coding or regulatory sequences (exons). This essential step in gene expression is catalyzed by the spliceosome, another dynamic megadalton ribonucleoprotein complex. Splicing also plays important roles in mRNP assembly by recruiting factors required for polyadenylation, nonsense mediated decay (NMD), mRNA export, and translation (Huang & Steitz, 2001; H. Le Hir et al., 2001; Lu & Cullen, 2003; Müller-McNicoll et al., 2016; Nott et al., 2004; Swartz et al., 2007; Wiegand et al., 2003; Woodward et al., 2017). Thus, splicing-dependent mRNP remodeling couples nuclear and cytosolic steps of post-transcriptional gene expression.

Pre-mRNA splicing expands the protein coding capacity of human genes. Alternative splicing (AS) manifests when the spliceosome has the option of ignoring exon or intron sequences. This occurs for a variety of reasons including the interactions between the pre-mRNA and RBPs, epigenetic signals, and RNA polymerase II kinetics (Fu & Ares, 2014). The most obvious impact of alternative splicing is to generate messages from the same gene encoding distinct polypeptides (Gallego-Paez et al.,

2017). However, AS also induces isoform-specific post-transcriptional regulation (Fig 1a). In some cases, AS generates unstable mRNA isoforms through the introduction of premature termination codons (PTCs) that can trigger nonsense-mediated decay (AS-NMD) (Barberan-Soler et al., 2009; Lareau et al., 2007; B. P. Lewis et al., 2003; Ni et al., 2007). In neurons, muscle cells, and drosophila embryos alternative splicing and polyadenylation induce isoform-specific patterns of mRNA localization (Taliaferro et al., 2016; E. T. Wang et al., 2012). Recently, several groups, including our own, discovered that mRNA isoforms can exhibit differential polyribosome association, suggesting that alternative splicing can be coupled to translational control (AS-TC) (Blair et al., 2017; Floor & Doudna, 2016; Reixachs-Solé et al., 2020; Sterne-Weiler et al., 2013; Wong et al., 2016). Although the mechanisms remain to be determined, shuttling pre-mRNA splicing factors, such as SR proteins may play important roles in coupling alternative splicing and translational control (Maslon et al., 2014; Sanford et al., 2004). Alternative 5' and 3' UTRs of mRNA transcripts will likely contain different sequences that influence translational control such as upstream open reading frames (uORFs), internal ribosome entry sites (IRES), and RBP-binding sites (Gebauer et al., 2012; Pfeiffer et al., 2012). Thus, alternative splicing not only expands the protein coding capacity of eukaryotic genes but may also influence the cytoplasmic fate of the resulting mRNA isoforms.

Comparative genomics and transcriptomics are powerful approaches for studying the evolution of gene regulatory mechanisms. For example, comparisons of tissue-specific gene expression and alternative splicing patterns across distantly related vertebrate species revealed that these modes of gene regulation evolve at different rates (Barbosa-Morais et al., 2012; Calarco et al., 2007; Mazin et al., 2018; Merkin et al., 2012). Comparative transcriptomic and proteomic analysis of primate cells also demonstrated that although steady state protein levels are similar for orthologous genes, mRNA levels can vary dramatically. This suggests that overall mRNA levels might evolve under less rigorous evolutionary pressure compared to protein expression levels (Khan et al., 2013).

Isoform-specific post-transcriptional regulation may also be under selective pressure during evolution. (Lareau & Brenner, 2015; B. P. Lewis et al., 2003; Pan et al., 2006; Saltzman et al., 2008). Hundreds of regions in the human genome are ultraconserved with the mouse and rat. Many of these sequences overlap alternative exon sequences associated with AS-NMD, suggesting that regulatory elements and function of these mRNA isoforms are under strong purifying selection. Indeed, ablation or programmed mis-splicing of AS-NMD results in growth defects and loss of tumor suppression in cell culture and mouse embryo models (Leclair et al., 2020; Thomas et al., 2020). Taken together AS-NMD appears to be part of an evolutionarily conserved regulatory mechanism for fine tuning the expression of splicing factors.

By contrast to AS-NMD, the physiological and evolutionary significance of AS-TC is unknown. In this manuscript, we use comparative transcriptomics to test the extent of conservation of AS-TC and to identify functionally important exons that contribute to the coupling. We compared Frac-seq data (subcellular fractionation and high-throughput RNA-sequencing, (Sterne-Weiler et al., 2013) of human, chimpanzee, and orangutan induced pluripotent stem cells (iPSCs). Using these Frac-seq data allows us to identify mRNA isoforms with differential polyribosome association. Our data show that the process of AS-TC is conserved across all three cell lines. We identified alternative splicing events with conserved translational control as well as events with species-specific regulation. We validated the ability of sequence elements associated with isoform-specific polysome association to affect mRNA translation in vivo using luciferase reporters. We further showed that events with conserved translational control show higher sequence conservation specifically in cassette exons, indicating their functional relevance. Taken together our data suggest a conserved mechanism of AS-dependent regulation of mRNA translation.

Results

Frac-Seq analysis of primate iPSCs

We previously developed Frac-seq (Fig 1b) as a method to determine the association of alternative mRNA isoforms with the polyribosome (Sterne-Weiler 2013). In order to test the hypothesis that alternative splicing coupled to translational control is conserved in primates, we applied the Frac-seq methodology to human, chimpanzee, and orangutan iPSCs. We identified and quantified approximately 3000-6000 events spanning a variety of different event types using previously established analysis pipelines (Fig 1c). Within each cell line, we identified over 1000 events with PSI values differing between fractions by more than 10%. We further separated these into events that generate non-productive NMD isoforms (AS-NMD) and events maintaining the integrity of their open reading frame. The latter we consider alternative splicing events that are potentially implicated in translational control (AS-TC events, Fig 1d). Between all three cell lines, we identified orthologous alternative splicing events with differing cellular fate and 362 orthologous alternatively spliced events that are coupled to translational control in all three cell lines (Fig 1e).

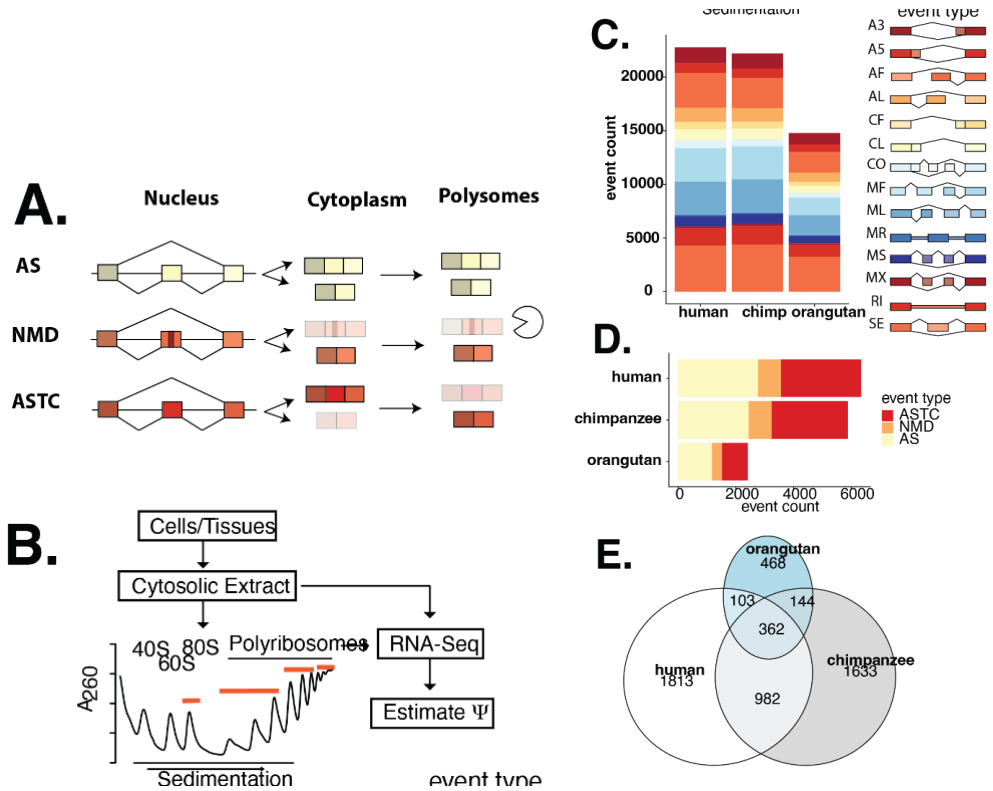

Fig 1. Frac-seq reveals polyribosome associated mRNA isoforms. A) Alternative splicing can influence multiple post-transcriptional regulation pathways. B) Frac-seq (subcellular fractionation and subsequent sequencing of polyA+ selected RNA from fractions) was performed on human, chimpanzee, and orangutan iPSCs. RNA from the total cytosolic lysate, the mono-ribosome (80s), the light (P2-P4), medium (P5-P8), and heavy polyribosome (P9+) was sequenced. B) Identification and quantification of alternatively spliced events was performed using junctionCounts. This pipeline allowed the identification of 14 different event types. C) Alternative splicing events were further classified into AS, AS-TC, and AS-NMD events. The proportions of these three event groups are comparable between the three cell lines. D) Out of the events categorized as either AS, AS-TC, or AS-NMD, about 900 events were identified in all three cell lines. E) Out of the events categorized as AS-TC, over 300 events were identified in all three cell lines.

AS-TC is conserved across primate evolution

To identify patterns of polyribosome association in orthologous AS-TC events, we performed unsupervised hierarchical clustering on the Spearman correlation of the mean PSI for each species, across all fractions. As a control, we performed the same analysis using orthologous splicing events that are not associated with differential polyribosome association. AS events cluster predominantly by species as the fractions within each species are more similar to each other than to other species (Fig 2A). By contrast, orthologous AS-TC events show a stronger correlation for PSI within each fraction across the different species (Fig 2B). Within the AS-TC set of orthologous events, we were interested in identifying potential differences in isoform-specific polyribosome association. To identify orthologous isoforms with the most and least conserved sedimentation profiles we calculated the difference in PSI between pairs of species and summed those differences into a single distance metric. The Manhattan distance for all orthologous AS-TC events results in a right-skewed distribution where the far left and right tails represent the AS-TC events with the most and least conserved sedimentation profile, respectively (Supp Fig 1A). We visualized these subsets of events in correlation heatmaps. The colors indicate the Spearman correlation of PSI values in events with low cumulative distance (conserved sedimentation profiles) (Fig 2C) and in events with high cumulative distance (species-specific sedimentation profiles) (Fig 2D). Interestingly, in events with conserved sedimentation profiles, the same fractions from the different cell lines cluster together neatly, indicating the PSI values in these fractions are more similar to each other than to the other fractions within the same species. Consequently, in events with species-specific sedimentation profiles, the different fractions of each species cluster together, indicating more similarity within the species than the fractions. For example, C12orf29 exhibits a similar sedimentation pattern between species (Fig 2c, left panel), whereas CNN1 alternative splicing generates isoforms with species-specific sedimentation patterns (Fig 2d, right panel). In events with species-specific sedimentation

patterns, the PSI values across the polysomal fractions differ between species, despite a similar nuclear output represented by the PSI value in the cytoplasmic fraction.

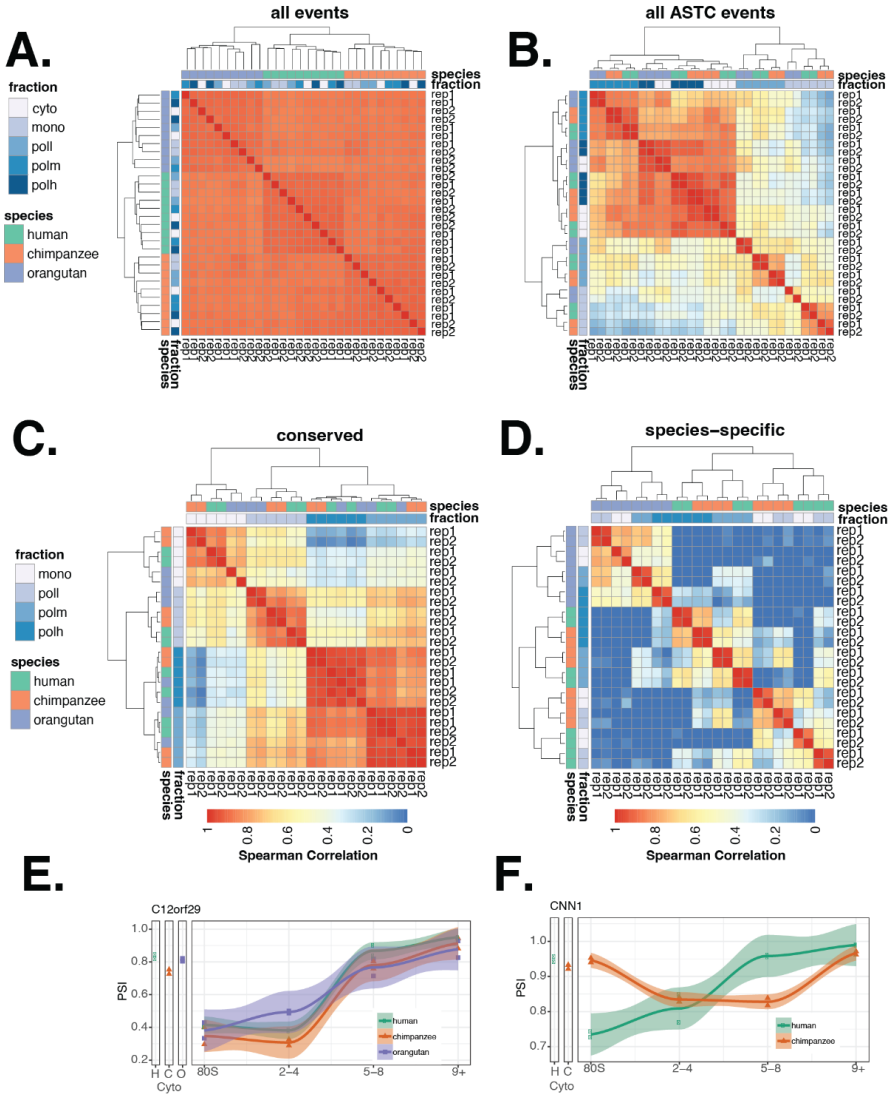

Fig 2. Orthologous AS-TC events exhibit either conserved or species-specific sedimentation profiles.

A) Heatmap of Spearman correlation of PSI values of all identified orthologous events. The columns and rows represent the total cytosolic lysate and the 4 subcellular fractions in each cell line. The colors represent the Spearman correlation of PSI values between pairs of fractions (red = high correlation, blue = low correlation). B) Heatmap of Spearman correlation of PSI values of all orthologous AS-TC events. C,D) Heatmaps of

Spearman correlation of PSI values of orthologous AS-TC events. C) The events in this heatmap exhibit sedimentation profiles consistent across all three cell lines as shown in the example in panel E. D) The events in this heatmap exhibit species-specific sedimentation profiles as shown in the example in panel F. E) The skipped exon event within C12orf29 is an example of an alternative splicing event with conserved sedimentation profiles across all three cell lines. F) The alternative first exon event of CNN1 is an example of an alternative splicing event with species-specific sedimentation profiles.

Exon sequences associated with AS-TC are strongly conserved

To discover determinants associated with AS-TC, we examined the sequence conservation of orthologous mRNA isoforms with similar or species-specific polyribosome profiles. We focused on two fairly simple classes of splicing events (SE), skipped exons and alternative first exons (AFE). We also calculated the phastCons score for AS events that were not associated with differential polyribosome association. Surprisingly, skipped exons that are not associated with AS-TC are much less conserved compared to their flanking exons. By contrast, exons linked to AS-TC exhibit significantly higher phastCons scores and are more similar to their flanking exons. Further, AS-TC exons that have conserved sedimentation profiles between species have elevated phastCons scores relative to the other classes (Fig 3a). Similar effects can be observed for the phastCons scores of alternative first AS-TC and non-AS-TC exons (Fig 3b). We further observe higher sequence conservation in the distal alternative first exons compared to the proximal alternative first exons. The high degree of sequence conservation within AS-TC exons suggests the presence of functional elements within these exons that influence polyribosome association.

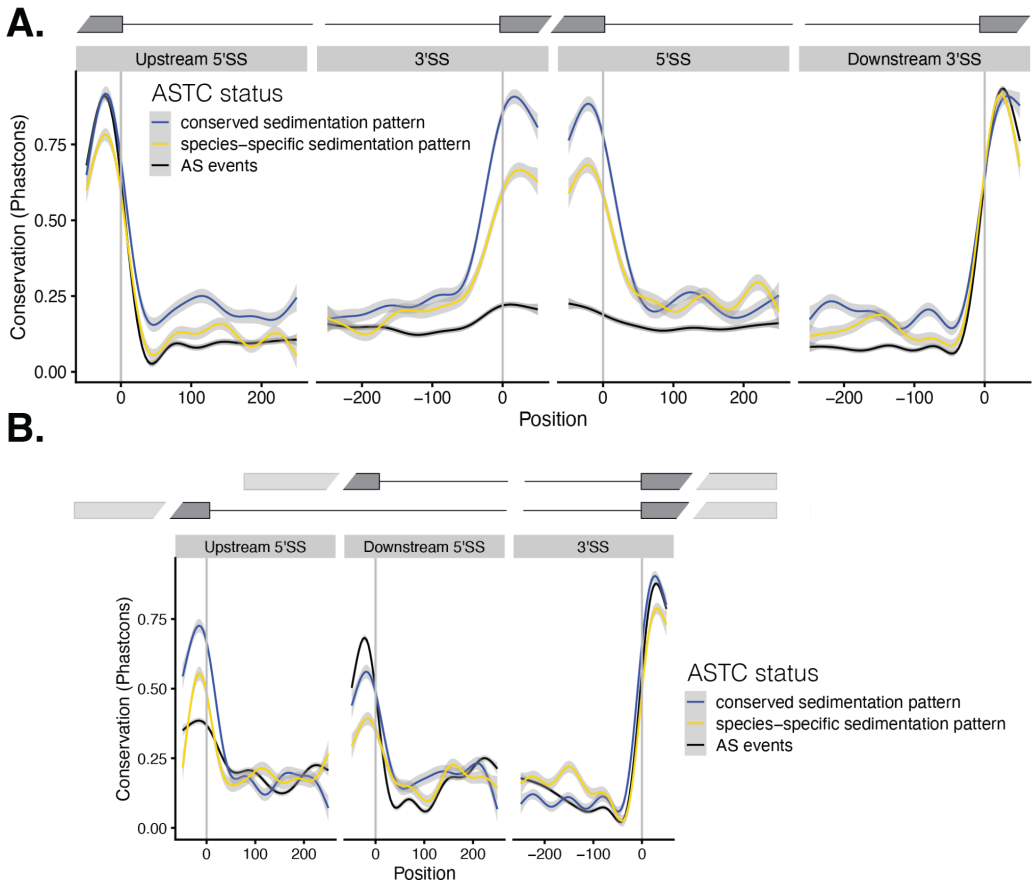

Figure 3

Fig 3. Alternative splicing events with sedimentation profiles consistent across species show higher sequence conservation. A) Sequence conservation of exon/intron boundaries of AS-TC and AS skipped exon events represented by phastCons scores. A lower score indicates less conservation. AS-TC events with conserved sedimentation profiles in blue, AS-TC events with species-specific sedimentation profiles in yellow, and simple AS events in black. B) Sequence conservation of exon/intron boundaries of AS-TC and AS alternative first exon events represented by phastCons scores.

Single nucleotide variants in orthologous AS-TC exons influence translation

In order to test the hypothesis that sequence differences associated with isoform-specific sedimentation patterns regulate isoform-specific polyribosome association, we created luciferase reporters from SE events from exhibiting AS-TC. Fig 4A shows the schematic for the skipped exon luciferase reporters, where either the human (green) or chimpanzee (orange) skipped exon were inserted in-frame, upstream of the firefly luciferase. For each reporter, we measured luciferase activity and steady-state mRNA levels. Frac-seq analysis revealed differential sedimentation profiles for human and chimp isoforms from the GGCX (Fig 4B) and SUMF2 (Fig 4E) genes. The orthologous exon sequences for both of these events differ by only a single nucleotide. Surprisingly, luciferase reporter constructs containing the chimpanzee-derived sequences promoted significantly higher luciferase activity compared to the human sequences (Fig. 4C, F). This effect was likely due to increased translational efficiency, as the steady-state mRNA levels are significantly higher for reporters containing the human SE compared to chimpanzee (Fig. 4D, G). Both pairs of orthologous exon sequences differ at a single position (Fig. 4H, I), which we predict to affect RBP binding sites based on the comparison of surrounding sequences with RNAcompete data.

We also tested the ability of alternative first exon sequences associated with isoform-specific sedimentation to affect mRNA translation. We created pairs of luciferase reporters from three genes exhibiting isoform-specific sedimentation in human and chimpanzee iPSCs corresponding to the proximal (blue) or distal (red) first exon (Fig. 5A). We chose two genes exhibiting species-specific sedimentation profiles (Fig. 45, E) and one gene with conserved sedimentation (Fig 5H). Interestingly, AFE reporters corresponding to the chimpanzee orthologs of CNN1 and UGP2 resulted in stronger luciferase activity compared to their human counterparts. By contrast, orthologous sequences from the MAD2L2 gene resulted in similar expression levels. In all cases, the steady state mRNA levels for each reporter were similar, suggesting that expression

differences were likely due to translation. Pairs of orthologous AF sequences differ in multiple positions between the species and show insertions and deletions, as well.

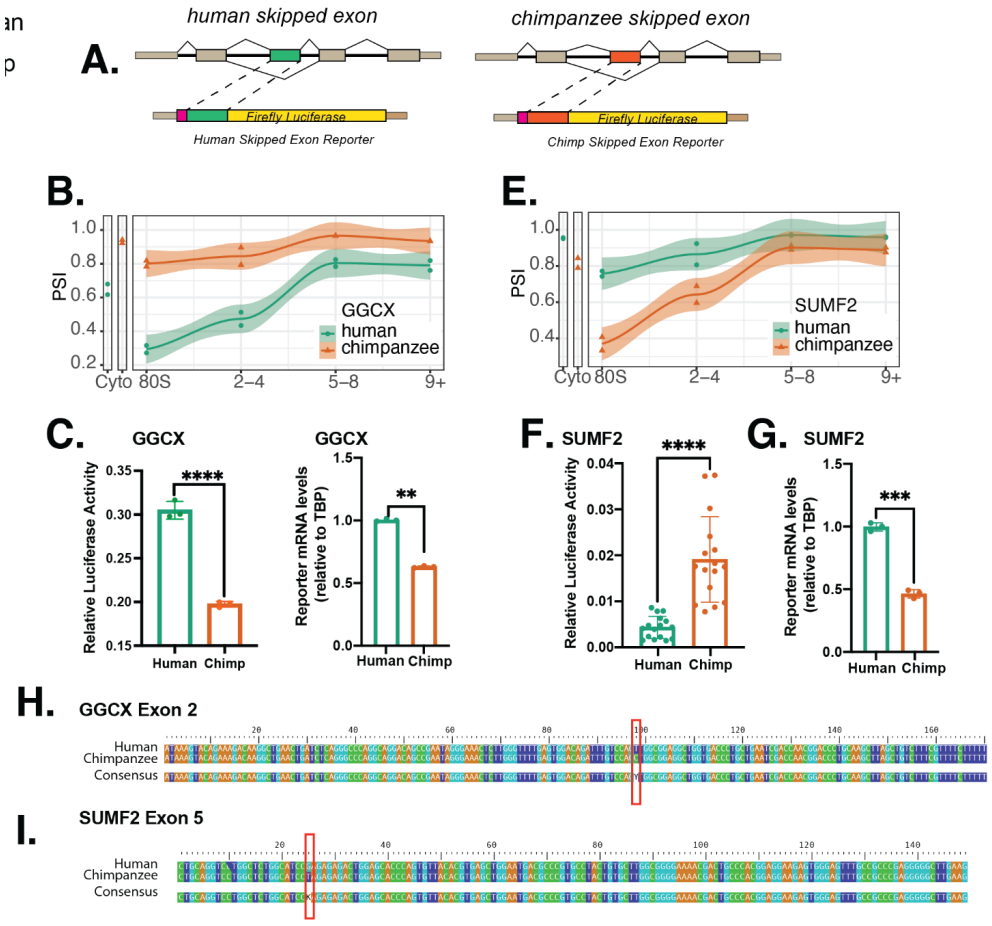

Fig 4. AS-TC cassette exons drive isoform-specific expression. A) Schematic diagram of the pairs of luciferase reporter constructs containing either the human (green) or chimpanzee (orange) cassette exon from different genes exhibiting AS-TC. B,E) Polyribosome sedimentation profiles for isoforms from the GGCX and SUMF2 genes (respectively) in human and chimpanzee iPSCs. C,F) Dual luciferase assays in HEK cells for the two skipped exon events. D,G) qPCR of SE reporter mRNAs. H) Alignment of GGCX exon 2 human and chimpanzee sequences. I) Alignment of SUMF2 exon 5 human and chimpanzee sequences.

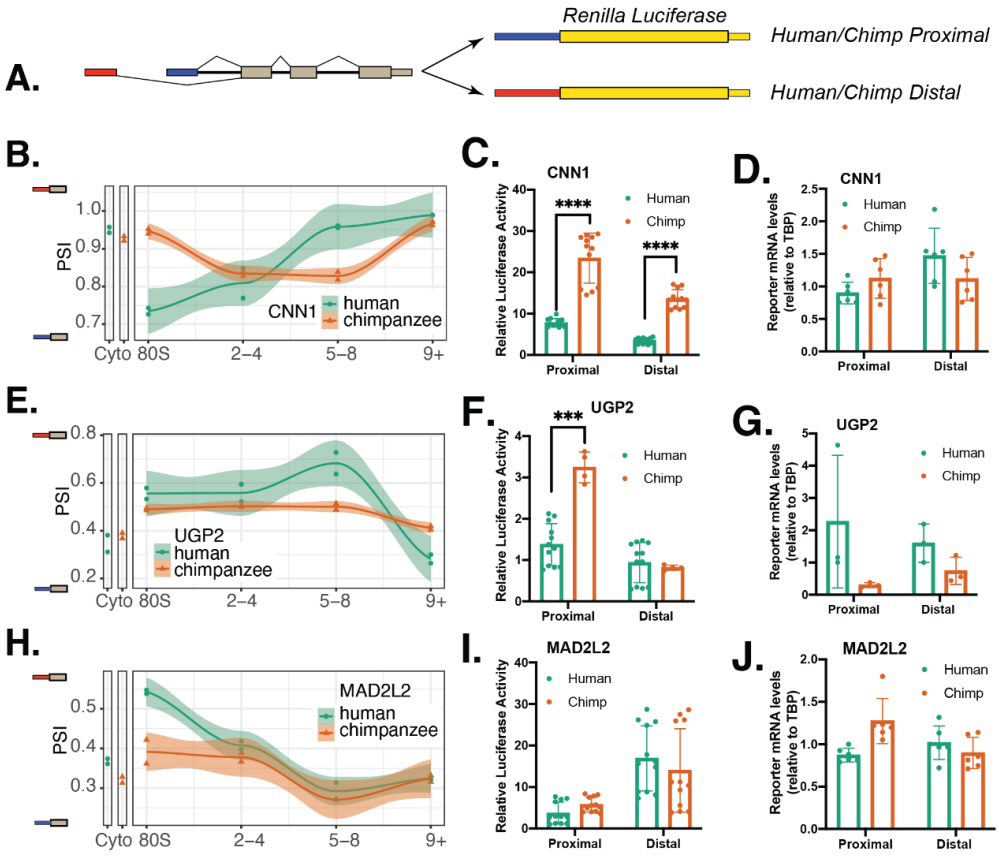

Fig 5. A) Schematic diagram of the pairs of luciferase reporter constructs containing either the distal (red) or proximal (blue) alternative first exons from either human or chimpanzee from different genes exhibiting AS-TC. B,E,H) Polyribosome sedimentation profiles for isoforms from the CNN1, UGP2, and MAD2L2 genes (respectively) in human and chimpanzee iPSCs. C,F,I) Dual luciferase assays in HEK cells for the three alternative first exon events. D,G,J) qPCR of AF reporter mRNAs.

Alternative splicing alters the cis-regulatory landscape of mRNA isoforms

In order to test the effect of alternative splicing on the cis-regulatory landscape of mRNA isoforms, we performed pairwise alignments of alternatively spliced exons (e.g.

skipped exons) between two species at a time (human/chimpanzee and human/orangutan) and identified the sequence differences, predominantly single nucleotide variants (SNV), between the species. We then tested the effect of these sequence differences on predicted RNA binding sites. For that purpose, we scored the similarity of RNAcompete-derived positional weight matrices (PWMs) against the sequences surrounding SNVs in a sliding window, collecting a predicted binding score for each tested RBP (Figure 6A). We performed this analysis on three groups of alternatively spliced events: events that are alternatively spliced but do not exhibit dynamic polysome association in one species but that are AS-TC events in the other species, events that were considered AS-TC in both species, and AS events without dynamic polysome association in both species. To identify changes in the cis-regulatory landscape of mRNA isoforms that might be connected to dynamic, conserved polysome association or species-specific polysome association, we compared the frequency of RBP binding sites that are changed in binding score due to the SNV in the different event groups (Figure 6B, C). The AS events in both species serve as a control. Interestingly, both human/chimpanzee and human/orangutan sequence differences resulted in changes in putative RBP binding sites with frequencies significantly different to the AS/AS control group (Figure 6B,C). Out of the 10 PWMs that showed significant differences in frequency in the human/chimpanzee group and the 21 in the human/orangutan group, 2 were found in both analyses (Figure 6D). The alternatively spliced events analysed for the human/chimpanzee and the human/orangutan comparison are predominantly distinct sets, with an overlap of 205 events (Figure 6E).

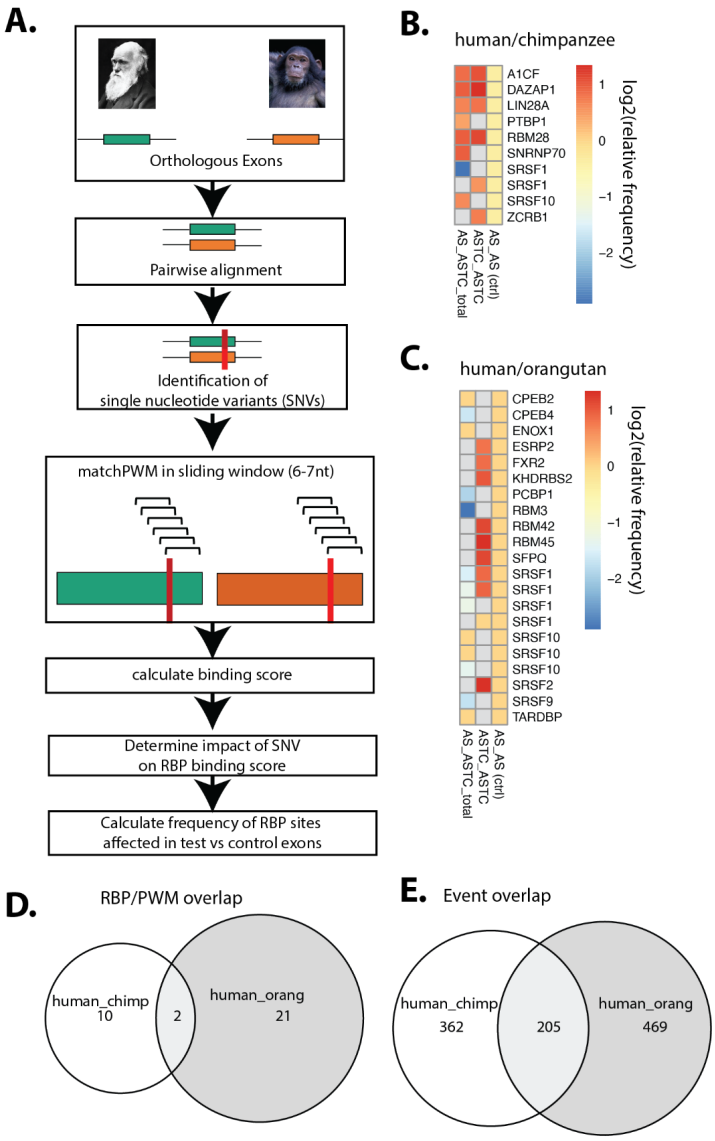

Fig 6. Alternative splicing alters the cis-regulatory landscape of mRNA isoforms. A) Schematic diagram of RBP binding analysis. B) Heatmap of log2 of relative frequencies of predicted RBP binding sites changed through single nucleotide variants between human and chimpanzee skipped exon sequences in AS/AS-TC events, AS-TC/AS-TC events and AS/AS events compared to the AS/AS event control group. C) Heatmap of log2 of relative frequencies of predicted RBP binding sites changed through single nucleotide variants between human and orangutan skipped exon sequences in AS/AS-TC events, AS-TC/AS-TC events and AS/AS events compared to the AS/AS

event control group. D) Euler diagram depicting the overlap of RBPs with significant differences in frequencies compared to the background between the human/chimpanzee and the human/orangutan analysis. E) Euler diagram depicting the overlap of analysed events in the human/chimpanzee and the human/orangutan analysis.

Discussion

One of the most powerful advantages of a comparative transcriptomics study of closely related species is the ability to query myriad sequence differences that have accumulated over millions of years of evolution and remain in the same genic context. Diverging sequence elements that change the cellular fate of orthologous mRNA isoforms may unveil functional elements important for post-transcriptional regulation of gene expression. Employing Frac-Seq, we identified orthologous mRNA isoforms with similar or distinct distributions in sucrose gradients (Figure 1). We then used ENCODE data to investigate the impact of the SNVs on potential protein-RNA interactions. This analysis revealed a distinct set of trans-acting RBPs that recognize sequences associated with AS-TC, suggesting that mRNP composition influences the cytoplasmic fate of mRNA isoforms. Notably, of the PWMs that show variable frequencies in AS-TC-associated skipped exons, the two that are common between human/chimpanzee and human/orangutan are predicted to change binding of SRSF1 and SRSF10 (Figure 5). Both of these SR proteins influence splicing and shuttle between the nucleus and cytoplasm. Whereas SRSF1 has been shown to increase translation efficiency of the message it is bound to (Maslon et al., 2014; Sanford et al., 2004), further studies would be needed to evaluate the cytoplasmic effects of SRSF10 and its protein isoforms.

Our data show that the coupling of alternative splicing to translational control is conserved across multiple primate cell lines, representing approximately 13 million years of primate evolution. We discovered 1,000-3,000 AS-TC events in human, chimp and orangutan iPSC lines. Approximately 30-50% of events found in each cell line were classified as AS-TC in another line and 362 alternative events exhibited AS-TC in all

three species (Figure 1E). By comparing PSI between fractions and species, we found that isoform abundance was strongly correlated by fraction, reflecting similar isoform-specific sedimentation patterns between species (Figure 2B). AS-TC events with shared sedimentation profiles also showed stronger alternative sequence conservation compared to species-specific AS-TC or AS events (Figure 3). The evolutionary conservation of these sequences indicates a potentially important biological function for this mode of gene regulation or the presence of potentially important cis-regulatory elements. Our work does not address the former hypothesis, but the reporter experiments presented in Figure 4 and Figure 5, speak to the latter. For example, reporters based on the fifth exon from the SUMF2 gene differ in only a single nucleotide position (Figure 4I). Using position specific weight matrices developed by the ENCODE consortium to interrogate sequences surrounding polymorphic sites, we identified potential differences in protein-RNA interactions. In the case of SUMF2, the G>T (human>chimp) change at position 41 in exon 5 convert recognition sites for SRSF10 and CNOT4 in the human sequence to an MSI1 binding site in the chimp exon. These potential differences in protein-RNA interactions could impact isoform-specific mRNA stability or translation.

The mechanisms of AS-TC remain unclear, but are certainly varied. Our data suggest that sequence differences between orthologous 5' UTRs and skipped exons result in species-specific polyribosome association of mRNA isoforms as well as influence species-specific expression of luciferase reporters in vivo. For example, orthologous isoforms of SUMF2 containing exon 2 exhibit distinct sedimentation profiles in sucrose gradients. Reporters containing human or chimp exon 2 sequences, which differ by a single nucleotide, have significant differences in luciferase activity (Figure 4E--G, H). We predict that the human sequence binds CNOT4 and SRSF10, whereas the chimp cis-regulatory elements recruits MSI. Because RBPs compete or cooperate with each other, changing the binding potential for one protein may also affect binding of other RBPs, leading to species-specific differences in mRNPs composition.

Several mechanisms were not explored in our analysis and remain viable options for specific transcripts. Changes in mRNA sequence could contribute to isoform-specific differences in RNA secondary structure, which has previously been shown to regulate mRNA translation (Chemla et al., 2020; Mao et al., 2014; Mauger et al., 2019). For example, more stable secondary structures near the start codon attenuates translation initiation (Kudla et al. 2009). Likewise, codon optimality influences translation elongation and RNA stability in a variety of species (Gu et al., 2010; Presnyak et al., 2015; Tuller et al., 2010). In metazoans, codon optimality is important for enhancing the expression of intronless genes but could also influence codon usage within alternatively spliced isoforms, either through inducing a frameshift, modulating GC content, or by varied use of exons with suboptimal codons. Despite these open questions, the work presented here indicates that regulation of gene expression by AS-TC is evolutionarily conserved across the great apes and implicates RNA binding proteins as key regulators of AS-TC in primates. Taken together, these findings suggest that alternative splicing influences mRNP composition by creating complexes with isoform-specific combinations of RNA binding proteins.

Materials & Methods

iPSC generation and culture

Human iPSC line C305 was derived from the GM12878 lymphoblastoid cell line using a proprietary episomal, plasmid-based reprogramming method by Cellular Dynamics. C305 cells were cultured on matrigel in mTeSR. Integration-free chimpanzee (Eip-8919-1A) and orangutan (Jos-C31) iPSCs were generated from primary fibroblasts as previously published by Field et al. (Field et al., 2017) and cultured on vitronectin in Essential 8 media.

Fractionation, polyribosome profiling, RNAseq

Frac-seq experiments were performed as previously published (Sterne-Weiler et al., 2013) using human, chimpanzee, and orangutan iPSC lines, with modification to the number of fractions collected. Cytosolic extracts from cell lines/tissues are fractionated by sucrose gradient centrifugation. We collected the total cytosolic lysate, the monoribosomal fraction (80s), as well as light (P2-4), medium (P5-8), and heavy (P9+) polyribosomal fractions. We then polyA+ selected RNA from each fraction and individually barcoded libraries using the Bioo NEXTflex Directional RNASeq kit. Multiplex-pooled libraries were sequenced on an Illumina HiSeq 4000 using paired-end 2x150bp sequencing, resulting in 75-150M reads per sample with approximately 40-50% junction reads per sample.

Mapping of Illumina short read RNA sequencing

The reads were mapped to the human genome assembly hg38, the chimpanzee (Pan troglodytes) genome assembly panTro6, and the sumatran orangutan (Pongo abelii) genome assembly ponAbe3 using STAR v2.7 (Dobin et al., 2013). Repeat sequences were masked by mapping to repeatMasker sequences (Smit et al. RepeatMasker Open-4.0 at http://repeatmasker.org) using Bowtie2 (Langmead et al., 2009). PCR duplicate removal was performed by collapsing fragments with common start and end positions and CIGAR strings using in house scripts. All data collection and parsing was done with bash and python2.7. Statistical analyses and data visualization were performed using R programming language version 3.5.1.

Identification and quantification of orthologous alternative splicing events

We used Stringtie (Pertea et al., 2015) to identify unannotated transcripts and used CAT transcriptomes (Fiddes et al., 2018) (for human (hg38), chimpanzee (panTro6),

and orangutan (ponAbe3)), together with the stringtie merge command to generate the final transcriptomes for each species. Pairwise alternative splicing events were identified by pairwise comparison of all transcripts with at least one exon-intron junction in common. An alternative event was defined to be a set of exons unique to one transcript that are surrounded by two exon-intron junctions common to both transcripts or by one junction and the transcript terminus.

Alternative splicing events were quantified by counting the number of reads supporting the exon-exon junction and the number of reads supporting the exon-intron junction of each event. PSI (percent spliced in) values were calculated as the ratio of number of reads supporting the included isoform to the number of reads supporting both isoforms.

The event identification and quantification was implemented using in house python scripts. Orthologous alternative splicing events were identified by mapping the event sequences from one species to the genomic sequences of all others. The genes containing the orthologous event coordinates were determined using the CAT transcriptome annotations.

Cross-fraction comparison/ Cross-species comparison / Identification of conserved and species-specific orthologous events

All events identified were filtered to be supported by at least 15 junction reads (per comparison). Alternative splicing events undergoing translational control, or AS-TC events, were defined as events with a change in PSI value (delta PSI) between any two adjacent fractions of at least 0.1. Consequently, alternative splicing events not undergoing translational control were defined as events with a minimum PSI > 0 and delta PSI < 0.1. Alternative splicing events leading to nonsense-mediated decay (AS-NMD) were identified using *in silico* translation of raw transcripts and subsequent identification of premature termination codons (PTCs) (technically CDSinsertion).

For estimating the difference/conservation of the polysome association pattern we calculated the Manhattan distances for each event between each two species. The Manhattan Distance is the sum of differences in mean psi between two species across all fractions. Min/max normalization of the Manhattan distance allowed us to identify events with overall different sedimentation profiles as opposed to events with similar sedimentation profiles at a different y-axis intercept. We ranked all AS-TC and AS-NMD events based on their min/max normalized Manhattan distance and used the top and bottom 10% (= 350 events) for further analysis, considering them the least and most conserved set of events respectively.

Determination of Sequence conservation

To determine the sequence conservation of AS and AS-TC events with conserved or species-specific sedimentation profiles, phastCons (A. Siepel, 2005; Adam Siepel & Haussler, n.d.) scores were obtained from the UCSC genome browser (Kent et al., 2002). For skipped exon events, phastCons scores were obtained for 100nt windows around both splice sites of the cassette exon as well as around the upstream 5'ss and the downstream 3'ss. For alternative first exon events, the scores were obtained for the 300 nucleotides downstream of both transcription start sites as well in 100nt windows around the 5'ss of the two alternative first exons. The scores were visualized with local nonlinear smoothing using a generalized additive model.

RNA purification and RT-qPCR

Total RNA was isolated using the Direct-zol RNA MiniPrep Kit (Zymo Research). 800ng of the RNA were treated with RQ DNase (per protocol). The DNase (Promega) treated RNA was reverse transcribed using the High-Capacity cDNA reverse

transcriptase kit (Applied Biosystems). 1:200 dilutions of the cDNA were made. For the qPCR, we used Luna 2x SYBR premix (total volume 20μl per reaction), 0.25 nM primers and 5μl diluted cDNA. qPCR was performed on QuantStudio 3 Real-Time PCR System (Applied Biosystems, Thermo Fisher) according to MIQE guidelines (Bustin et al., 2009).

Luciferase Reporters

Luciferase activity was assayed 24 hours post transfection using Dual-Glo Luciferase Assay System (Promega). For a 6 well plate, transfections were performed with lipofectamine with either 2μg pLCS plasmid (previously published Sanford et al.) plus 125ng control plasmid (rluc) (for skipped exon events) or 1μg p5UTR (pLightSwith_5UTR, from Switch Gear) 1μg plus 250ng control plasmid (pmir) (for alternative first exon events) per well.

Alignment and identification of SNVs and indels

The sequences of alternative regions of AS-TC events from human and chimpanzee were globally (Needleman-Wunsch alignment) aligned using R Biostrings (Pagès et al., 2020) (function pairwiseAlignment) using default settings. Mismatches, insertions, and deletions were identified in the pairwise alignments using R Biostrings.

RBP binding prediction

RNAcompete (Ray et al., 2009, 2013) data sets available on ENCODE (ENCODE Project Consortium, 2012) were used to predict RBP binding sites affected by single nucleotide sequence differences between human and chimpanzee AS-TC events. The matchPWM (Wasserman & Sandelin, 2004) function from the R Biostrings package (Pagès et al., 2020) was used to score PWMs based on the RNAcompete data in a

sliding window across the identified sequence differences. Matches achieving at least 80% of the maximum score were recorded for both human and chimpanzee datasets. The matches were compared between the species in the form of a deltaPWMscore (e.g. PWMscore(human) - PWMscore(chimp)).

Chapter 3: A method for campus-wide SARS-CoV-2 surveillance at a large public university

Abstract

The systematic screening of asymptomatic and pre-symptomatic individuals is a powerful tool for controlling community transmission of infectious disease on college campuses. Faced with a paucity of testing in the beginning of the COVID-19 pandemic, many universities developed labs focused on SARS-Cov-2 diagnostic testing on campus and in their broader communities. We established the UC Santa Cruz Molecular Diagnostic Lab in early April 2020 and began testing clinical samples just five weeks later. Using a clinically-validated laboratory developed test (LDT) that avoided supply chain constraints, a highly automated sample pooling and processing workflow, and a custom laboratory information management system (LIMS), we expanded testing from a handful of clinical samples per day to thousands per day with the capacity to screen our entire campus population twice per week. In this report we describe the technical, logistical, and regulatory processes that enabled our pop-up lab to scale testing and reporting capacity to thousands of tests per day.

Introduction

The SARS-CoV-2 pandemic radically altered society, the economy, and global health. The rapid spread of SARS-CoV-2 is driven by a combination of asymptomatic carriers and pre-symptomatic transmission (Johansson et al., 2021); (Tindale et al., 2020). Diagnostic testing is one of the critical tools for breaking viral transmission chains (Larremore et al., 2021; Mina et al., 2020). The combination of testing for symptomatic cases, tracing and testing of close contacts, and isolation of infected individuals proved to be a highly effective approach to control community spread of SARS-CoV-2 (Bharti et al., 2020; Hidayat et al., 2020; Smith et al., n.d.). Unfortunately, testing capacity in the USA was constrained by supply chain shortages (Bernard et al., 2020; Esbin et al., 2020; Petersen et al., 2021; Salimnia et al., 2021), the lack of commercial and medical center diagnostic labs capable of rapidly scaling testing, and an uncoordinated response at federal, state, and local levels ((Editors & The Editors, 2020; *[No Title]*, n.d.; Schneider, 2020; Todd, 2020).

Institutions of higher education were dramatically impacted by the pandemic. Instruction shifted to distance learning, research activity was curtailed, student activities and organizations ceased operations, and on campus housing was dramatically reduced (Currie et al., 2021; Penuliar et al., 2020). For college campuses, molecular diagnostic testing for asymptomatic or pre-symptomatic SARS-CoV-2 infections plays an important role in preventing community spread by identifying and isolating cases at the earliest stages of infection (Sina Booeshaghi et al., 2020). Such rapid turnaround typically requires an in-house testing facility. However, many college campuses, especially those without a medical school, lacked their own clinical diagnostic labs at the beginning of the pandemic, hindering the campus testing process and response. To address this challenge, a regulatory mechanism for creating “pop-up” laboratories was developed by which an existing clinical laboratory could extend its license onto non-contiguous research space to allow for the increase in clinical testing capacity in the areas of need.

The primary challenges for campus diagnostic labs are throughput, rapid turnaround time for test results, and a sampling frequency that is less than the viral incubation period (Larremore et al., 2021). Additional considerations such as the biospecimen type and the testing platform play important roles in successful implementation of an institutional-level testing program. The other critical component is a laboratory information management system (LIMS) capable of accessioning, tracing, and reporting in a timely manner to state and local public health agencies on thousands of samples per day through an automated, hands-free process. Finally, given the reality of limited budgets for testing, effective scaling requires that the test be as cost-effective as possible.

In response to the pandemic many university research laboratories pivoted to SARS-CoV-2 molecular diagnostic testing ((“Blueprint for a Pop-up SARS-CoV-2 Testing Lab,” 2020); (McDermott, 2020). This was possible through a relaxed regulatory framework that enabled the creation of temporary COVID-19 testing sites operating under existing campus clinical licenses. These pop-up diagnostic laboratories capitalized on the ingenuity and expertise of faculty, students, and staff to develop and validate laboratory developed tests (LDTs, tests whose application is restricted to the given laboratory; (Center for Devices & Radiological Health, 2019)) to serve the needs of their campus and local communities. In response to the early stages of the COVID-19 pandemic, we established a temporary SARS-CoV-2 testing site at the University of California Santa Cruz, in accordance with the The Clinical Laboratory Improvement Amendments (CLIA, (*Clinical Laboratory Improvement Amendments (CLIA)*, n.d.)) and the California Department of Public Health guidelines (https://covid19.ca.gov/). This clinical laboratory performs diagnostic testing of symptomatic patients for our community’s safety-net healthcare providers and the UC Santa Cruz Student Health Clinic as well as asymptomatic screening on campus.

This paper builds upon the blueprint laid out by our colleagues at UC Berkeley 22, but describes the strategic choices we made to scale testing capacity and overcome the common barriers to SARS-CoV-2 diagnostic testing (Fig 1). In particular we adopted

a distinct liquid handling strategy, using 96 well pipetting heads. By integrating this parallel sample processing strategy with our Laboratory Information Management System, we created a sample collection process that was compatible with 96 well format RNA extraction and facilitated paperless test requisition, accessioning and reporting. The liquid handling strategy also enabled a 'linear sample pooling' approach that allowed for efficient 10:1 pooling and rapid pool deconvolution, if necessary. Taken together, our strategy enabled us to scale cost-effective, rapid turnaround testing capacity to thousands of tests per day, with a lean staff and a modestly equipped laboratory. We believe the approach outlined here can be widely implemented and will be useful for public health efforts during the current pandemic and future outbreaks of infectious disease.

Methods

Assembly of sample collection kits

Sample collection kits consist of racks of barcoded 1.4 mL tubes (Micronics) filled with 0.6 mL transport media (DNA/RNA Shield, Zymo Research). Following automated decapping (Micronics) of the clean tube racks, transport media is added using a MultifloFX liquid dispenser (BioTek) and recapped. The barcodes on each rack of tubes are scanned using a Zianth flatbed scanner and tube barcodes are uploaded into the LIMS. Racks containing collection tubes along with anterior nasal swabs (Typenex Medical) and additional caps are distributed to testing kiosks.

Sample collection at campus kiosks

Participants in the campus asymptomatic testing program check in at a central desk using their student IDs or state issued IDs. After confirming their ID, the receptionist enables the student to select a sample collection tube and scan the 1D barcode. The barcoded tube is linked to an electronic test requisition form (eTRF) for the specific participant. The participant then brings the tube to a supervised self-collection site and

swabs each nostril for 15 seconds, then dunks the swab in the DNA/RNA Shield transport media for 30 seconds, wipes the rim of the tube with a kimwipe and applies a clean cap. Sample tubes are then placed into a collection tube rack for transport to the laboratory.

Automated RNA extraction and RT-qPCR

Each sample tube contains approximately 400 uL of nasal swab in DNA/RNA shield, extraction is initiated by adding 800 uL Viral RNA Buffer (VRB, ZymoResearch) to each sample using a MultifloFX liquid dispenser (BioTek). The addition of the Viral RNA Buffer plus reducing agent at this step has robust mucolytic properties which significantly reduces pipette tip clogging due to sample viscosity. Samples are titurated on the Bravo deck and 540 uL is transferred to a 1.2 mL deep well plate. RNA is extracted using magnetic beads (ZymoResearch) and eluted in 35 uL ddH2O and 5 uL is used as template for RT-qPCR. All automation scripts can be found here on github https://github.com/UCSC-CCDL/Bravo-protocol-files.

LIMS, Sample Accessioning, and Reporting

A custom laboratory information management system (LIMS) was developed in collaboration with Third Wave Analytics (San Francisco) using the Salesforce Lightning Platform and Experience Cloud. This system tracks every sample, well position, plate, RT-qPCR result and applies logic tables to call the presence or absence of SARS-CoV-2. To accession samples, racks are scanned on a Ziath flatbed scanner. The scanner output file is uploaded to the LIMS and barcodes become linked to well positions on the plate. Barcodes are matched against expected barcode IDs from the eTRFs; any unmatched sample tubes are assigned “missing information” status and manually removed from the rack, and the entire rack is rescanned until there is a perfect match to the eTRF database. For pooled samples, a new parent rack containing clean tubes is scanned and every child rack is associated with the parent rack via a barcode scanning step prior to removing an aliquot of the sample. A rack ID from each child plate is also associated with

the parent plate. The LIMS generates a plate definition file containing the barcode IDs for each sample which is loaded into the Design and Analysis software that drives the QuantStudio 6 Pro. Following qPCR results are ingested into LIMS and the logic table is applied to each sample, resulting in a positive, negative, inconclusive, or invalid call. Following review by licensed CLSs, results are reported to medical providers and state and local health agencies via secure HL7 messaging.

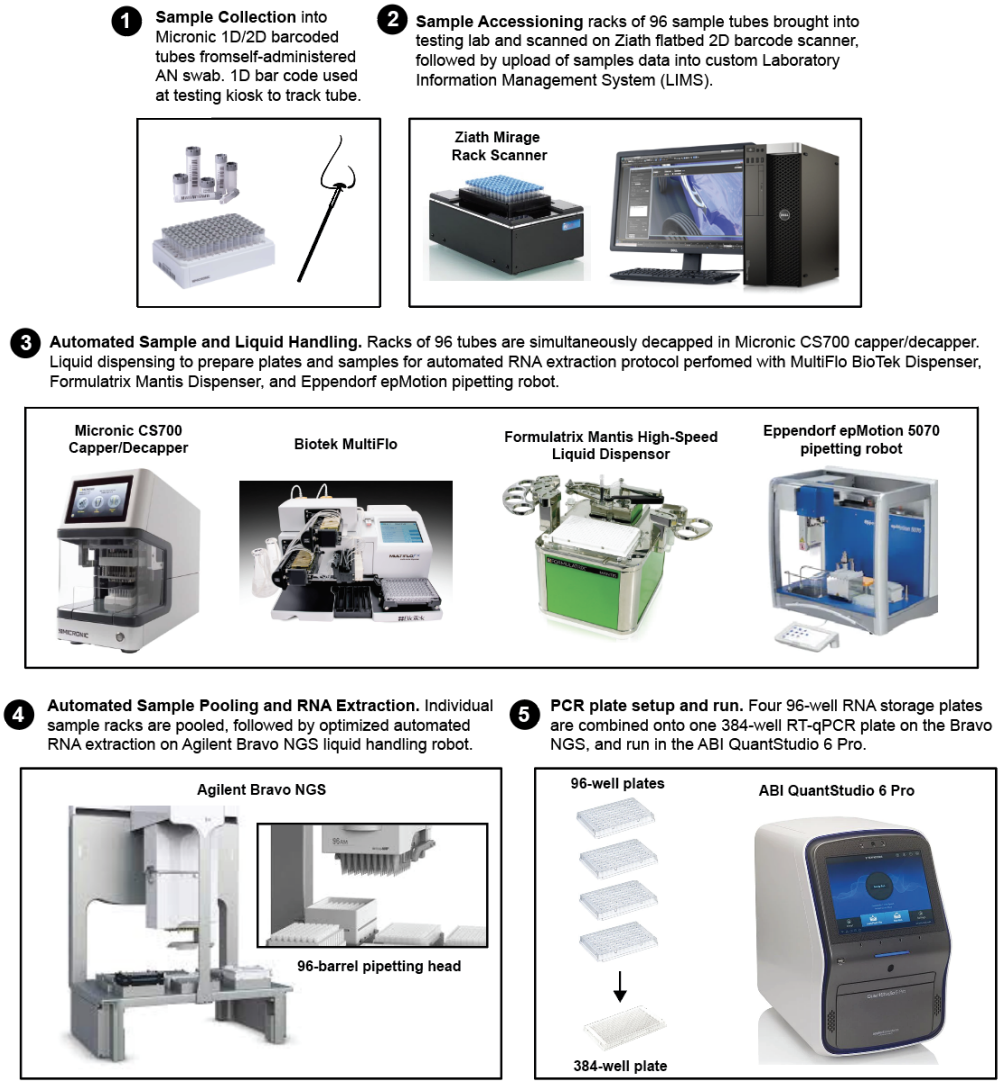

Figure 1. Overview of the UC Santa Cruz Colligan clinical Diagnostic Laboratory workflow.

Multiplex Assay

At low prevalence, sample pooling can greatly accelerate high capacity testing for SARS-CoV-2 (Abid et al., 2020; Ben-Ami et al., 2020; Brynildsrud, 2020; Mallapaty, 2020). We took advantage of the high-sensitivity multiplex assay described above and the 96 channel pipetting capacity off the Bravo NGS-A platform to develop a sample pooling protocol (Fig 2A). In this scheme, individual sample racks (child racks) are combined into a new rack of 96 matrix tubes (parent rack). Each barcoded child rack is associated with a single barcoded parent rack and each child sample tube with its unique rack position (ie. position C4) is associated with a new barcoded parent tube with the identical rack position in the parent rack. Construction of sample pools are managed by the LIMS, which enables accessioning of the parent rack and parent tubes, as well as the association of each child rack and child tube. This process ensures traceability of individual samples. We refer to this strategy as “linear pooling” because the position in the 96 well array determines which samples from each child rack contribute to the pooled sample. This approach is in contrast to reported “matrix” pooling strategies which employ a more complex pool building algorithm in order to deconvolute redundant pools through two PCR reactions (Mallapaty, 2020). In our linear pooling approach, sample deconvolution occurs by re-testing the individual child samples that make up a positive pool. Our laboratory validated 10:1 sample pooling following FDA guidelines (Fig 2B-D). The Ct values for N1 and N2 in the parent (pooled) or children (individual) samples are highly correlated (R=0.96 and 0.95, respectively) with an offset of ~ 3.3 Ct in the pooled samples, as expected for a 10-fold dilution of a positive sample (Fig 2B,C).

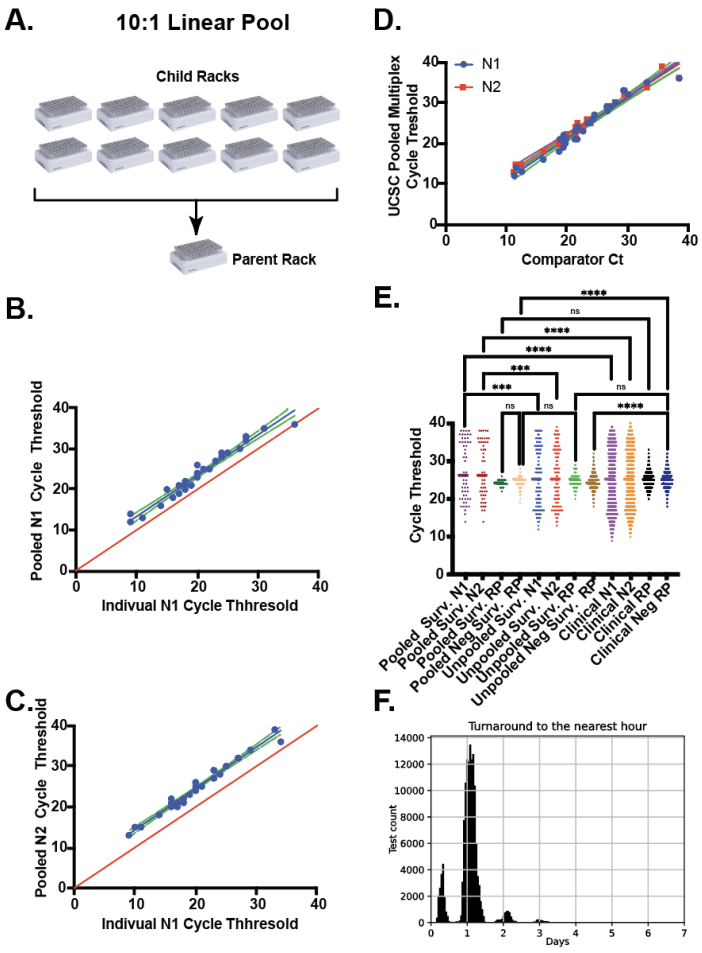

Figure 2. Validation of a pooled sample testing strategy for campus-wide SARS-CoV-2 surveillance. (A) 10:1 linear pooling strategy. Upto 10 individual sample racks (children) are combined into a single pooled sample (parent) rack. (B) UCSC multiplex N1 assay on 30 individual or pooled positive samples. (C) UCSC multiplex N2 assay on 30 individual or pooled positive samples. (D) Analysis of pooled positive samples with the UCSC multiplex or comparator assay. N1 amplicon (blue) and N2 amplicon (red). (E) Overall performance of the UCSC multiplex assay on thousands of clinical, surveillance and pooled surveillance samples. (F) Turnaround time for both clinical and surveillance samples, rounded to the nearest hour.

The sensitivity and specificity of the pooled assay was also validated by the Pangea Labs test (Fig 2D and Table 1). The overall performance of the test on both unpooled and pooled samples has been robust. Fig 2E compares cycle thresholds for N1, N2 and RP, from >10,000 pooled samples, >10,000 unpooled surveillance samples and >30,000 clinical samples. We found no significant difference in the distribution of N1 or N2 cycle thresholds from positive clinical or individual surveillance samples. By contrast we observed significant differences in the cycle thresholds for N1 and N2 from pooled sample tests as compared to unpooled surveillance and clinical samples. This result was expected as positive samples are diluted approximately 10-fold relative to unpooled tests. We observed a slight, but significant difference in Ct values for RP from negative unpooled surveillance samples and negative clinical samples (mean Ct 24.77 and 25.10, respectively). This difference could be due sample collection by a healthcare provider as compared to supervised-self collection of surveillance samples. Finally, we calculated the turnaround time for all COVID-19 tests performed during the fall and winter quarters. Fig 2F shows the distribution of turnaround times (defined as the total time between sample accessioning and reporting) with a mean of ~25 hours and a standard deviation of ~11 hours. A significant number of results were returned in < 12 hours.

Samples Tested 10-Sample Pool	***Comparator Method Result***	
Candidate Test Result	**+**	**-**
Positive	*29*	*1*
Negative	*1*	*29*

Positive Percent Agreement: *29/30 = 97% (95%CI: 84%-100%)*
Negative Percent Agreement: *29/30 = 97% (95%CI: 84%-100%)*

Table 1: Performance of the UCSC Multiplex SARS-CoV-2 Assay on 10-sample pools against the Pangea comparator

Laboratory Information Management System (LIMS)

The blueprint developed by our colleagues at the Integrated Genomics Institute (IGI) at UC Berkeley ("Blueprint for a Pop-up SARS-CoV-2 Testing Lab," 2020) clearly demonstrated that a LIMS is required for sample accessioning, tracing and reporting. Because the liquid handling strategy and sample collection tubes differed from the IGI blueprint, it was necessary to develop a custom LIMS, rather than licensing an established LIMS. Like the IGI, our LIMS consists of three modules (Fig 3). The accessioning module matches sample tube barcodes to electronic test requisitions that are deposited into the LIMS from the sample collection kiosks. The test requisitions are received electronically by the LIMS. When sample racks arrive in the lab they are heat inactivated (30 minutes at 70ºC)(Lista et al., 2020) then the 2D barcodes for all 93 samples are scanned en masse using a flatbed scanner (Ziath). The scanner output file is uploaded to the LIMS and barcodes on the tubes are matched by the LIMS against the barcode associated with the electronic test requisition form for each participant. If a sample barcode cannot be matched against a complete requisition, that sample is flagged by the LIMS and removed from the rack. The accessioning process is then repeated for the entire rack. For pooled sample testing, child racks are scanned and associated with a rack of clean 1.4 mL barcoded tubes in a barcoded parent rack. A fraction of each sample from a child rack is transferred to the parent rack in order to build the pools for testing. Each tube within the parent rack is associated with up to ten individual child samples. The parent plate is scanned and uploaded to the LIMS. Each sample extraction plate, RNA storage plate, and qPCR plate is barcoded and associated with each sample throughout the process, ensuring traceability. The final module links qPCR results to individual samples and applies a logic table to call results as detected, not detected, inconclusive, or invalid. The LIMS enables rapid and facile review by clinical laboratory scientists and timely result reporting to healthcare providers and local and state officials via HL7.

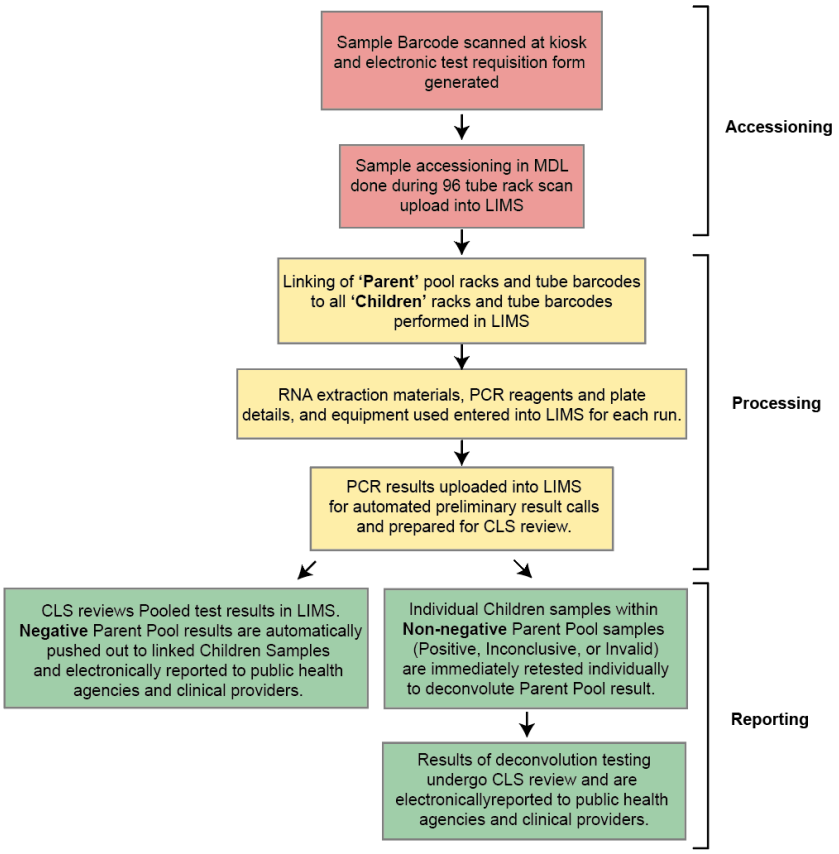

Figure 3. Overview of sample management by the custom laboratory information management system. The workflow is divided into three basic modules: sample accessioning, sample processing and result reporting. To accelerate testing, accessioning and reporting steps are fully automated.

LIMS integration with public health and provider portals allowed scaling

The original implementation of the LIMS had an entirely manual intake of orders and outflow of results. For the kiosks, a flat file was created that could be imported into the LIMS to create new patient and order entries. For reporting results, a pdf result for each patient was generated to send to the provider using a flat file and mail merge, and a flat file was generated for mandatory reporting to the State of California. This minimum-viable-product met requirements but required substantial daily human

intervention. In order to scale up, we implemented an automatic report delivery system integration by contracting with middleware Software-As-A-Service (SAAS) providers (BridgeConnect and Santa Cruz Health Information Exchange). This integration connected the LIMS with the existing systems at the County and State, and with the patient medical portal for approval and communication of results and accessioning of samples from the kiosk sites into the LIMS (Fig 4).

This LIMS has also enabled Accessioning and Resulting to be performed within the existing electronic medical record (EMR) software Point and Click (PNC). Another piece of the integration was creating a staff demographic feed to allow them to be tested within PNC. This iterative approach to integration, with multiple plans, allowed the project to be launched quickly and then later improved so that it could scale. By careful collaboration across various parts of campus we brought together technical experts from genomics, student health, and ITS to provide a seamless experience to patients and students in the community. This information management solution enabled automated test requisitioning, sample accessioning and result reporting and was critical to scaling-up testing capacity.

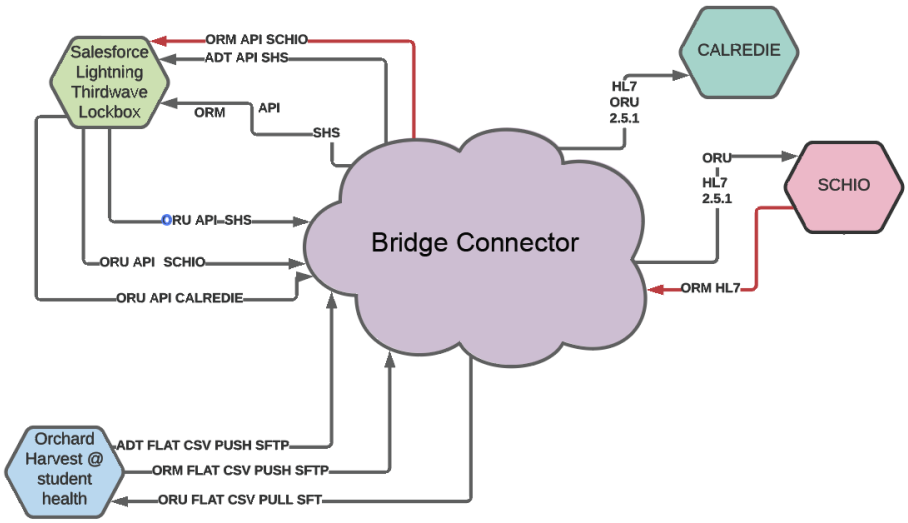

Figure 4. Integration overview between Salesforce Thirdwave LIMS at UCSC CCDL, Orchard Harvest at Student Health, Calredie (California department of public health), and SCHIO for third party integrations in the community. (A)Orchard Harvest, the Laboratory Information System in student health services. Used for orders originating from UCSC affiliates and for communicating results to affiliates. (B) Salesforce Lightning Thirdwave Laboratory Information System. Used by clinical laboratory for processing samples and results. '(C)' Integration platform which converts orders from flat file from Harvest to API Calls in salesforce, and converts results from API in salesforce to flat file for Harvest ingestion. Also exports data to the CDPH in HL7 and to SCHIO for community partner results. (D) Calredie is the CDPH platform for communicating results and orders. (E) Santa Cruz Health Information Exchange - non-profit integration platform provider for community to share data. (F) to be built - order interface from third party community providers directly into the Thirdwave LIMS without having to use the provider portal.

Potential Impact of surveillance on campus-wide SARS-CoV-2 transmission

UC Santa Cruz opened the 2020 academic year under fully remote instruction with low on-campus student density. Approximately 1200 students lived on campus

during Fall quarter; students were tested upon arrival and initially on a twice weekly cadence thereafter. Fig 5A shows the number of daily tests performed during the academic year. The methods described above enabled an efficient scale up of COVID-19 surveillance during the late fall and early winter quarters. During the Fall 2020 quarter we observed an increase in new cases per day during November and December (Fig 5B). Entry testing upon the return of students to campus in the Winter 2021 also discovered a large number of COVID-19 positive individuals. The rapid isolation of positive students coincides with the rapid decline in daily cases and the campus positivity rate (Fig 5C). Between the peak on January 4th and February 1st, the positivity rate declined ~17 fold, to 0.09%. Although there were 227 individuals who tested positive for SARS-CoV-2 between July 14, 2020 and May 1, 2021, case investigations found no evidence of transmission on the UCSC campus. We also compared the positivity rate for samples collected on campus to those that were collected off campus by safety net health care providers and processed by the UC Santa Cruz Colligan Clinical Diagnostics Laboratory (CCDL). Fig 5D shows that compared to symptomatic clinical samples, the positivity rate for samples collected through the asymptomatic testing plan is significantly lower than samples collected from the broader community, as expected. Additionally, recent work from the University of California COVID-19 Task Force compared the COVID-19 incidence of 20-29 year olds on campus and the surrounding community. The COVID-19 incidence was approximately 50% lower on the UCSC campus population than in the surrounding county 20-29 year-olds, suggesting that the multi-layered mitigation measures, including surveillance testing, limited SARS-CoV-2 transmission on campus (Pollock et al., 2021).

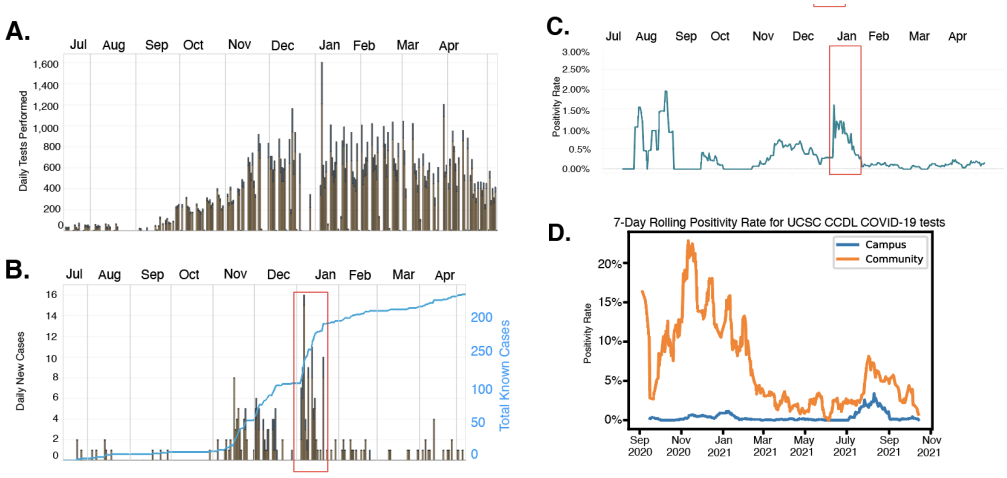

Figure 5. Scaling and impact of asymptomatic testing at UC Santa Cruz during Fall 2020 and Winter 2021 academic quarters. (A) Increase in tests performed per day at the UC Santa Cruz Molecular Diagnostic Lab. Tests for students represented by gold bars, tests for staff in blue. (B) New cases reported on campus per day for students (gold bars) and staff (blue bars) and cumulative case count (blue line). (C) 7 day rolling average positivity rate for UCSC students and staff. Red box signifies a potential spike in positive cases due to imported infections as students repopulated the campus after the winter break. (D) Comparison of positivity rate for clinical samples collected by local health care providers and symptomatic and asymptomatic samples from the UC Santa Cruz campus. All data are available at our campus COVID-19 dashboard https://recovery.ucsc.edu/reporting-covid/covid-tracking/ and are updated in real-time.

Discussion

Even after a full year of the COVID-19 pandemic, scaling up testing remains a major challenge for many institutions (Sina Booeshaghi et al., 2020). Pooled testing is recognized as an important approach for SARS-CoV-2 surveillance (Chan et al., 2021; Mulu et al., 2021; Sawicki et al., 2021; Thanh et al., 2021). We believe our implementation represents a robust process that is applicable on an institutional level.

The approach described here leverages barcoded, SLAS-compatible sample collection tubes to facilitate hands-free accessioning and reporting. To increase testing capacity and conserve resources, we implemented a sample pooling strategy. This approach also accelerates turnaround times and increases throughput as prevalence drops. Finally, a custom LIMS enables sample traceability throughout the testing process as well as deconvolution of pooled samples.

The COVID-19 testing and surveillance efforts at UC Santa Cruz were inspired by the work of our colleagues at other University of California Campuses. However, the relative geographic isolation of Santa Cruz county, combined with limited high complexity diagnostic infrastructure and expertise, drove us to seek solutions that could allow our testing program to scale and remain resilient to supply chain disruptions. The result was a strategy that differed significantly from our colleagues at the IGI Testing Consortium, which provided the blueprint that many labs, including our own, initially followed (“Blueprint for a Pop-up SARS-CoV-2 Testing Lab,” 2020). One of the key initial steps was to create a laboratory developed test using reagents that had not already received Emergency Use Authorization (EUA) from the FDA. This approach allowed us to avoid potential bottlenecks without depleting reagents that were needed by molecular diagnostic labs at medical centers. Instead, we developed an RNA extraction system using research grade reagents from commercial labs. We used a robust one-step RT-PCR reaction mix that was closely related to the gold standard Taq-Path reagent, but was not a component of any test with EUA. We also implemented a multiplex test developed at the Center for Immunity and Infection at Columbia University. This multiplex assay enabled the sensitive detection of SARS-CoV-2 and the endogenous RP transcript, which is an important indicator of sample quality and allowed for pooled sample testing.

Sample collection, accessioning, and result reporting are among the major challenges in scaling up a surveillance program. One important innovation in our process was the use of Society of Laboratory Automation and Screening (SLAS) compatible 1.4 mL barcoded sample collection tubes with 0.6 mL transport medium rather than larger

collection tubes, such as 15 mL falcon tubes (Fig 1). The 1.4 mL sample tubes are critical because they enable hands-free sample accessioning, processing and reporting. Each 1.4 mL tube is labeled with machine-readable 1D and 2D barcodes (on the side and bottom, respectively) that encode an identical numerical identifier. The sample collection tubes are racked into 96 tube arrays, compatible with an automated 96 screw-cap capper/decapper and the Agilent Bravo NGS-A liquid handler 96-well pipetting head. This system eliminated the need for both manual evaluation and recording of sample identifiers and manual transfer of the sample into 96-well extraction plates. The NGS-A's 96 channel pipette head also enables a linear pooling strategy. In this case pooled (parent) sample racks are assembled from up to 10 individual sample racks (children). This process requires a single box of tips per child rack to prevent cross contamination of the individual samples. We found that when assembling parent racks it was important to keep the pipetting head in a fixed position relative to the sample racks in order to avoid carry over contamination during pooling (data not shown). Using this method, a single NGS-A is capable of pooling and extracting 930 samples in approximately 1 hour. This process scales efficiently with additional liquid handlers and staffing. In the event of a positive pool deconvolution occurs by simply retesting individual samples that contributed to the pool.

Like our colleagues at UC Berkeley's IGI, we also developed a custom laboratory information management system (LIMS) using the SalesForce platform. While our initial decisions were informed by the blueprint paper, the key differences described above (1.4 mL matrix tubes, sample pooling and deconvolution) required a customized LIMS that was distinct from the IGI platform. Although development of a custom LIMS was laborious, a fully integrated system for accessioning, tracking and deconvoluting sample pools, as well as reporting results was the single most important element in scaling testing capacity. This system completely eliminated the need for paper test requisition forms and any human readable identifiers.

Conclusions

We describe a process for hands-free accessioning, sample pooling, automated extraction, pool deconvolution and reporting that can be completed in a single day. In theory, a modestly equipped and staffed lab can efficiently process thousands of pooled samples per day. This system enables efficient surveillance for SARS-CoV-2 or pathogens linked to future pandemics on an institutional scale.

Acknowledgments

We would like to thank JoeBen Bevirt (Joby Aviation) and Dr. Nischay Mishra (Columbia University) for technical expertise in automation and SARS-CoV-2 molecular diagnostics, respectively. We thank our colleagues at the UCSF/CZI Biohub Drs. Joe DeRisi and Stephen Miller for testing our clinical and surveillance samples while our lab was briefly shut down during the CZU Lightning Complex Fire evacuations. We thank the staff of the Colligan Clinical Diagnostic Lab (formerly known as the UCSC Molecular Diagnostics Lab) for critical comments on the manuscript. We thank Dr. Susan Weaver and Robert Gosselin for their support on clinical laboratory practice and compliance.

We thank our Santa Cruz County community partners, including Salud Para La Gente, Santa Cruz Community Health, The Santa Cruz County Health Department, Encompass, the Santa Cruz County Jail, Santa Cruz County Probation, and the Santa Cruz County Community Foundation for ongoing collaboration and support throughout the pandemic. Special thanks belong to Santa Cruz Community Health and Salud Para La Gente, both federally qualified health centers, for their partnership in developing and improving our testing lab assays and protocols. Santa Cruz Community Health was our first partner, and assisted in several validation efforts and provided valuable feedback as we designed our testing service protocol and reporting structure. We appreciate the individual efforts of Santa Cruz Community Health CEO Leslie Conner, Medical Director Dr. Casey Kirkhart, COO Philippa Barron, and Operations Director, Stephanie Macwhorter. Salud Para La Gente similarly partnered with us on efforts to continuously improve our COVID-19 testing and was our lead partner in thinking through adoption of AN swabs. This collaboration resulted in a significant positive impact for our community and our campus by improving the swabbing experience and enabling rapid ramp up of testing due to the increased ease of AN swab use. We appreciate the individual efforts of Salud Para La Gente CEO Dori Rose Inda, Medical Director Dr. Amy McEntee, Director of Nursing Danielle Obinger,

Jennifer Wood, Senior Systems Analyst, and Associate Director of Nursing Iriana Hinman.

Appendix: Nanopore sequencing reveals whole mRNA isoform structures engaged with translational machinery

Alternative mRNA processing increases both the protein coding and regulatory capacity of resulting mRNA isoforms. To determine the extent of influence of mRNA processing on differential isoform fate, we use a cellular fractionation and high throughput sequencing approach (Frac-Seq) for quantification of isoform-specific mRNA recruitment to polyribosomes. Here we identify splicing events that alter the polyribosome association of the resulting mRNA isoforms. We use the Rolling Circle to Concatemeric Consensus (R2C2) method developed by the Vollmers lab at UCSC (Volden & Vollmers, 2018) and Oxford Nanopore Sequencing to overcome previous limitations of short-read, PacBio, and full-length cDNA sequencing to collect millions of high-quality, full-length isoform reads. This analysis also revealed that nonsense-mediated decay (NMD) substrates are primarily enriched in the 80S fraction, and less substantially enriched in the light polysome fraction in embryonic stem cells. We also find evidence for attenuation of NMD or stabilization of NMD substrates in the neural cell type. These data finally address the previously impenetrable question of which isoforms are in our dynamically regulated pool.

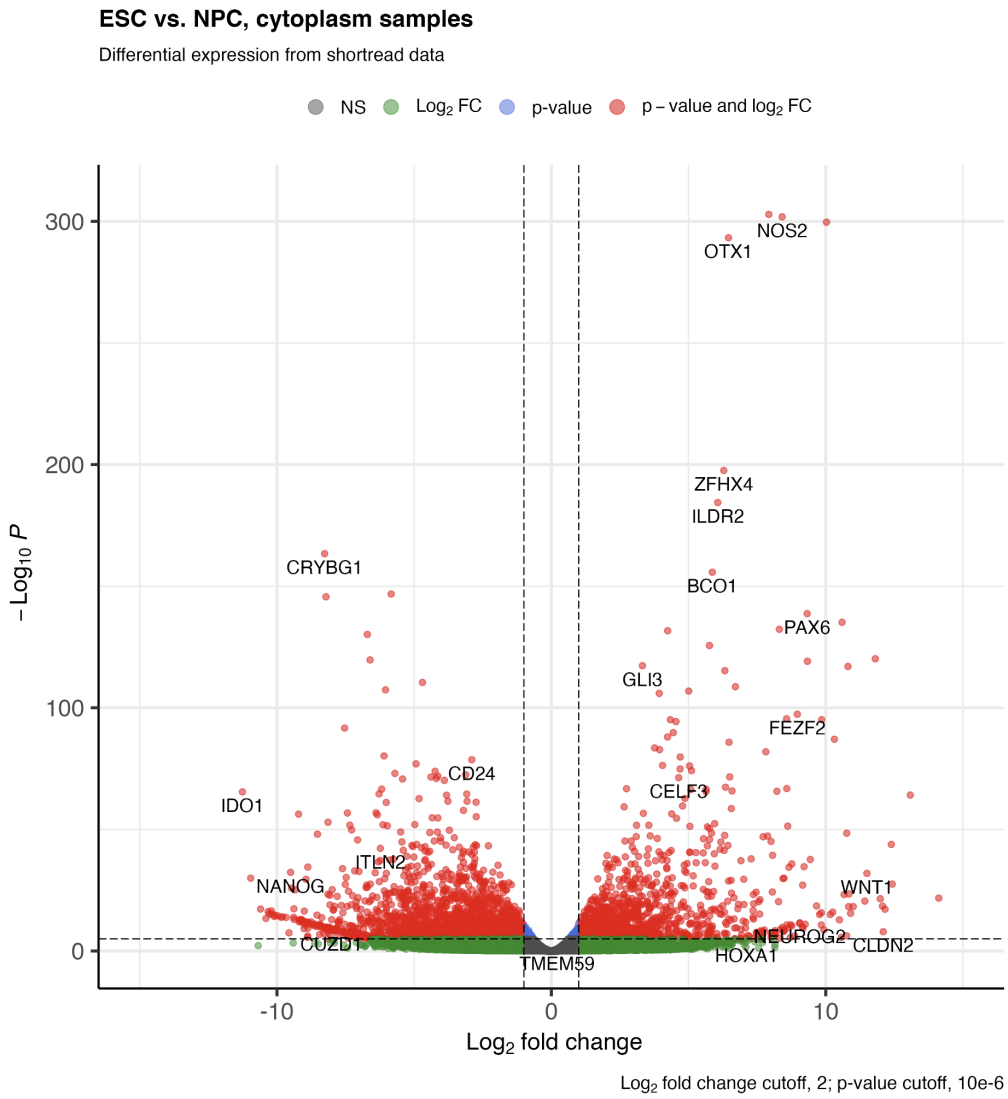

Figure 1. Transcriptome Profiling of ESC and NPC reveals wide-spread changes in gene expression. Volcano plot showing differential expression of cytoplasmic RNAs in H9 and H9-derived NPCs.

Frac-seq short-read analysis of H9 and H9-derived NPCs

To determine how alternatively spliced isoforms are dynamically regulated in stem cells and stem cell derived neural progenitor cells (NPCs), cytoplasmic extracts from monolayer-cultured H9s and H9-derived NPCs were separated on sucrose

gradients. We isolated the monosome fraction (mono), light polysome fraction (poll) and the remainder of the polysomes (polh). RNA was extracted from these fractions and total cytoplasmic RNA, polyA+ selected, and converted to directional RNA Seq libraries (BIOO Scientific qRNA). Biological and technical replicates were sequenced using Hiseq 4000 PE150 (50-100M reads per library). Reads were mapped to the genome with STAR, then assembled into a transcriptome with the aid of a reference annotation (GENCODE) using StringTie. Pairwise alternative splicing events were quantified by counting junction reads (or mean coverage for included form of RI events) using custom scripts (https://github.com/ajw2329/junctionCounts).

Investigating NMD in neural differentiation

Differential splicing analysis revealed a large number of changes occurring during differentiation (3970 events when tested at dPSI > 0, 366 events when tested at dPSI > 0.1). Differential splicing analysis between fractions (e.g. cytosol vs 80S) was performed to identify those events whose isoforms are non-uniformly distributed among the different fractions. This could be an indication of isoform-specific translation regulation, along with decay processes like NMD. In ESCs there are 2479 events (tested at dPSI > 0.1) with non-uniform gradient distributions (i.e. events in which at least one fraction-to-fraction comparison was significant), while in NPCs, there are 1256. This analysis also revealed that NMD substrates are primarily enriched in the 80S fraction, and less substantially enriched in the light polysome fraction.

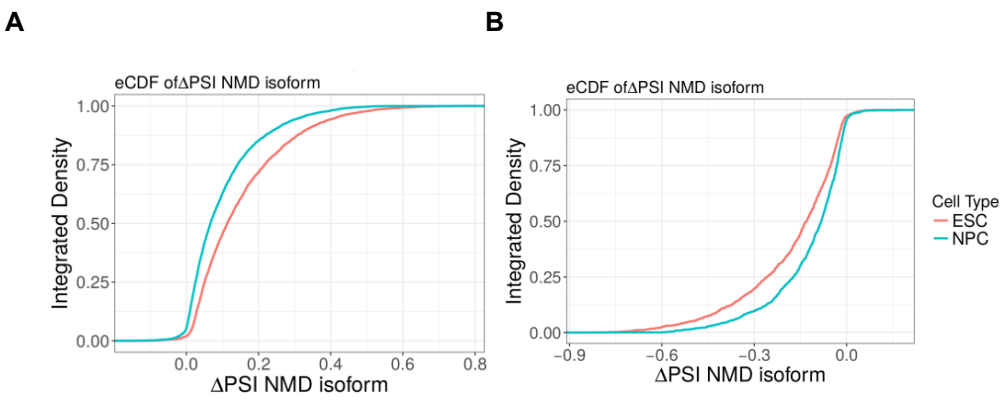

Figure 2: A) eCDF plot of NMD event behavior in input vs monosome B) eCDF plot of NMD event behavior in monsome vs heavy polysome.

The dPSIs for cytosol and monosome were compared for the two cell types. Here, dPSI refers to the change in the fraction of the NMD isoform (i.e. a positive value means an increase in the fraction of the NMD isoform), and the comparisons were conducted exclusively on NMD events. This analysis identified a significant difference in the dPSI distribution between ESC and NPC in the cyto-80S comparison ($p < 2.2e-16$, two-tailed KS-test). Visualization of the eCDFs revealed a leftward shift in NPC relative to ESC, indicating a greater proportion of low dPSI values (Figure 2). Broadly, this can be interpreted to mean a less substantial enrichment of NMD substrates in the 80S fraction of NPCs, which may result from decreased NMD activity, and a more substantial enrichment of NMD substrates in the heavier fractions of NPCs. It seems as though NMD substrates are moving out of the 80S fraction and into the heavier fractions in NPCs, possibly due to a prolonged half-life as a result of downregulated NMD.

Longread sequencing provides highly accurate, full length consensus reads for mRNA transcripts

We prepared full-length cDNAs from the same mRNA used prior for Illumina sequencing, using the R2C2 method developed by the Vollmers lab (Volden, et al., 2018).

Libraries were pooled and sequenced on an ONT PromethION, generating 12.11M reads with read length N50 of 17.6Kb. Reads were basecalled with bonito (https://github.com/nanoporetech/bonito). This combined method gave a median identity of over 99%. We then used FLAIR (Tang et al., 2020) to call mRNA isoforms and quantify their relative abundance within fractions and SQANTI3 for quality control and isoform classification (https://github.com/ConesaLab/SQANTI3).

Average gene body coverage for R2C2 libraries shows slight bias at both ends of the gene; this may be due to the presence of alternative first and last exons and internal priming on genomically encoded As (to be addressed). One ESC monosome library shows appreciable 3'end bias compared to the other libraries (Figure 3). Overall, we identified 29,243 full length isoforms from 12,802 gene loci; 61% of those loci had more than one transcript identified in our libraries. Of those transcripts, 19,797 matched the GENCODE v31 comprehensive human genome annotation, thus between these two cell types and over four cellular fractions, we have identified over 10,000 novel transcripts.

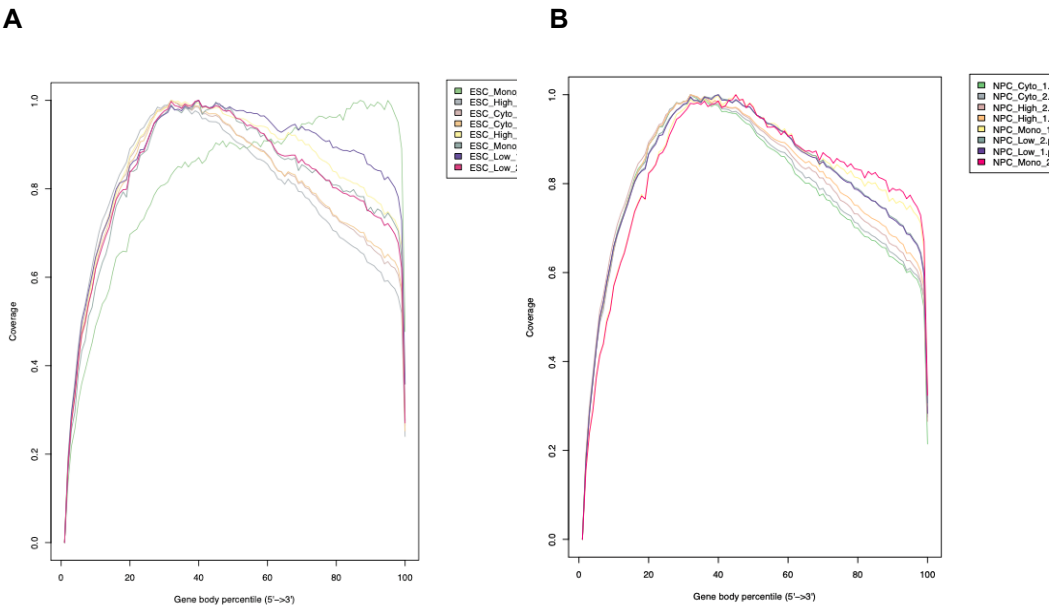

Figure 3 : Gene body coverage for sequencing libraries A) Duplicate cytosol, monsome, low polysome and high polysome libraries from ESCs B) Duplicate cytosol, monsome, low polysome and high polysome libraries from NPCs

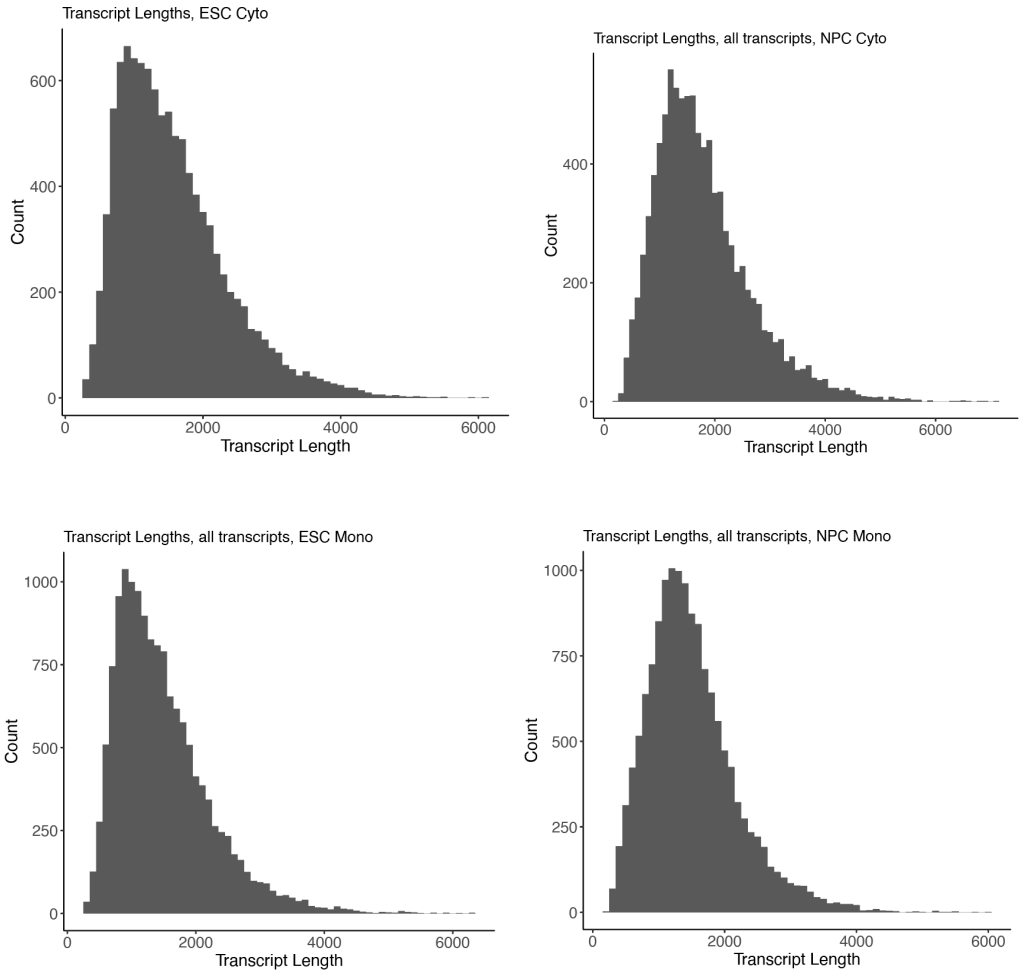

Figure 4: Transcript length histograms for ESC and NPC cytoplasm and monosome fractions

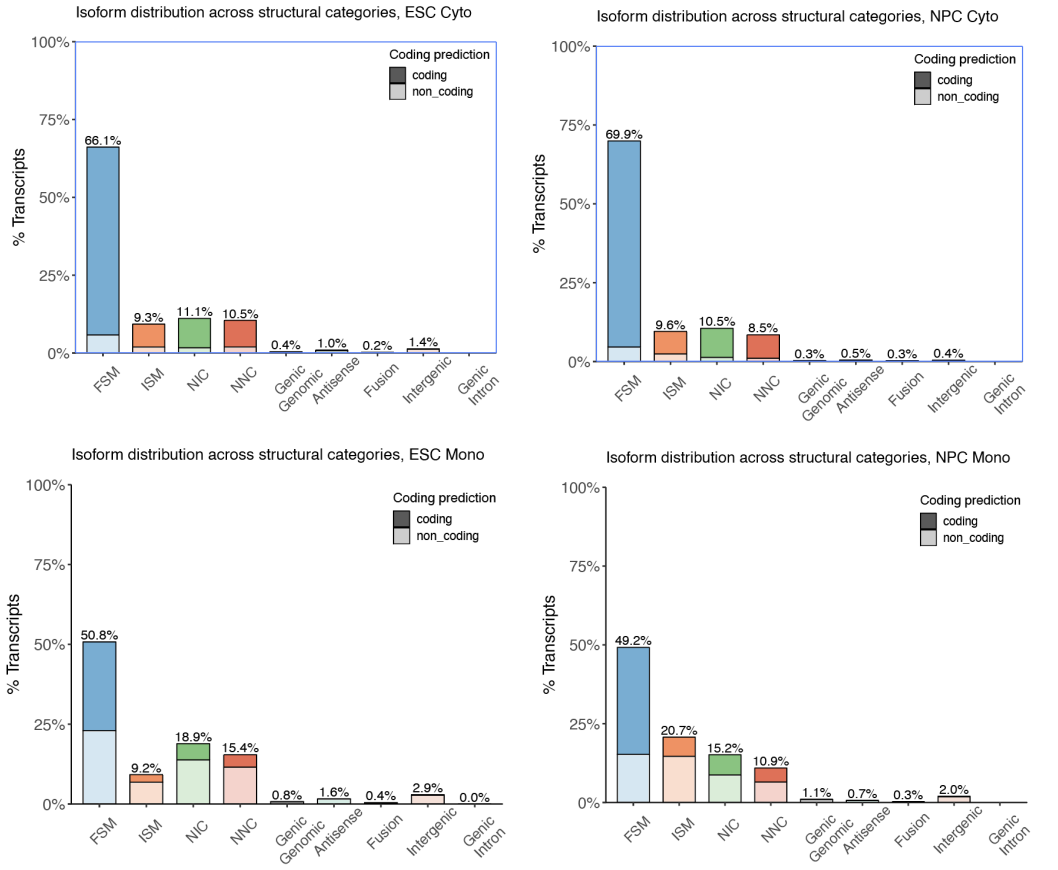

Figure 5: Percent isoform distribution and relative coding potential across transcript structural categories

Transcript length in cytoplasm and monsome fractions were similar, with a slightly larger median transcript length for both fractions in NPCs (Figure 4). When assessing the types of transcripts in these fractions, it is evident that the main class of transcripts are full splice matches (FSM) to transcripts annotated in the GENCODE human genome annotation (v31) for the hg38 assembly (Figure 5). The next largest type of transcript class is ISM, which is an incomplete splice match to the reference annotation. NIC and NNC transcripts are novel isoforms with either a new combination of known and annotated splice sites, or a novel isoform with at least one new, unannotated splice site within the transcript.

Further, we observe that the vast majority of transcripts in the cytoplasm are identified as having a significant ORF, thus are parsed out as coding (darker fraction of the bar). When comparing the coding potential of the transcripts identified in the monosome fraction to the cytoplasm, it is clear that there is a larger fraction of non-coding mRNAs for all transcript classes. This is in line with the idea that the monosome is enriched for targets of translation dependent decay pathways. Further, comparing the relative proportion of coding vs non-coding transcripts between ESC and NPC monosomes, we see a smaller proportion of non-coding transcripts identified in the NPC monosome library.

Bibliography

Abid, S., Ferjani, S., El Moussi, A., Ferjani, A., Nasr, M., Landolsi, I., Saidi, K., Gharbi, H., Letaief, H., Hechaichi, A., Safer, M., Ben Alaya, N. B. E., & Boubaker, I. B.-B. (2020). Assessment of sample pooling for SARS-CoV-2 molecular testing for screening of asymptomatic persons in Tunisia. *Diagnostic Microbiology and Infectious Disease*, *98*(3), 115125.

Abramson, R. D., Dever, T. E., & Merrick, W. C. (1988). Biochemical evidence supporting a mechanism for cap-independent and internal initiation of eukaryotic mRNA. *The Journal of Biological Chemistry*, *263*(13), 6016–6019.

Algire, M. A., Maag, D., & Lorsch, J. R. (2005). Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Molecular Cell*, *20*(2), 251–262.

Anderson, K., & Moore, M. J. (1997). Bimolecular exon ligation by the human spliceosome. *Science*, *276*(5319), 1712–1716.

Arribere, J. A., & Fire, A. Z. (2018). Nonsense mRNA suppression via nonstop decay. *eLife*, *7*. https://doi.org/10.7554/eLife.33292

Asano, K., Clayton, J., Shalev, A., & Hinnebusch, A. G. (2000). A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. *Genes & Development*, *14*(19), 2534–2546.

Barberan-Soler, S., Lambert, N. J., & Zahler, A. M. (2009). Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. *RNA* , *15*(9), 1652–1660.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., Kim, T., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., & Blencowe, B. J. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science*,

338(6114), 1587–1593.

Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., & Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Research*, *10*(7), 1001–1010.

Behm-Ansmant, I., Rehwinkel, J., & Izaurralde, E. (2006). MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. *Cold Spring Harbor Symposia on Quantitative Biology*, *71*, 523–530.

Ben-Ami, R., Klochendler, A., Seidel, M., Sido, T., Gurel-Gurevich, O., Yassour, M., Meshorer, E., Benedek, G., Fogel, I., Oiknine-Djian, E., Gertler, A., Rotstein, Z., Lavi, B., Dor, Y., Wolf, D. G., Salton, M., Drier, Y., & Hebrew University-Hadassah COVID-19 Diagnosis Team. (2020). Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *26*(9), 1248–1253.

Bernard, L., Desoubeaux, G., Bodier-Montagutelli, E., Pardessus, J., Brea, D., Allimonnier, L., Eymieux, S., Raynal, P. I., Vasseur, V., Vecellio, L., Mathé, L., Guillon, A., Lanotte, P., Pourchez, J., Verhoeven, P. O., Esnouf, S., Ferry, M., Eterradossi, N., Blanchard, Y., … Heuzé-Vourc'h, N. (2020). Controlled Heat and Humidity-Based Treatment for the Reuse of Personal Protective Equipment: A Pragmatic Proof-of-Concept to Address the Mass Shortage of Surgical Masks and N95/FFP2 Respirators and to Prevent the SARS-CoV2 Transmission. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.584036

Bernstein, P., Peltz, S. W., & Ross, J. (1989). The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Molecular and Cellular Biology*, *9*(2), 659–670.

Beyer, A. L., & Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Development*, *2*(6), 754–765.

Bharti, N., Exten, C., Fulton, V., & Oliver-Veronesi, R. (2020). Lessons from managing a

campus mumps outbreak using test, trace, and isolate efforts. *American Journal of Infection Control*. https://doi.org/10.1016/j.ajic.2020.11.008

Bicknell, A. A., Cenik, C., Chua, H. N., Roth, F. P., & Moore, M. J. (2012). Introns in UTRs: why we should stop ignoring them. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *34*(12), 1025–1034.

Blair, J. D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., & Floor, S. N. (2017). Widespread Translational Remodeling during Human Neuronal Differentiation. *Cell Reports*, *21*(7), 2005–2016.

Blencowe, B. J. (2017). The Relationship between Alternative Splicing and Proteomic Complexity [Review of *The Relationship between Alternative Splicing and Proteomic Complexity*]. *Trends in Biochemical Sciences*, *42*(6), 407–408.

Blueprint for a pop-up SARS-CoV-2 testing lab. (2020). *Nature Biotechnology*, *38*(7), 791–797.

Borman, A. M., Michel, Y. M., & Kean, K. M. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Research*, *28*(21), 4068–4075.

Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., & Morris, D. R. (1999). Role of two upstream open reading frames in the translational control of oncogene mdm2. *Oncogene*, *18*(41), 5631–5637.

Brynildsrud, O. (2020). COVID-19 prevalence estimation by random sampling in population - optimal sample pooling under varying assumptions about true prevalence. *BMC Medical Research Methodology*, *20*(1), 196.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., & Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, *55*(4), 611–622.

Calarco, J. A., Xing, Y., Cáceres, M., Calarco, J. P., Xiao, X., Pan, Q., Lee, C., Preuss, T.

M., & Blencowe, B. J. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. *Genes & Development*, *21*(22), 2963–2975.

Capecchi, M. R. (1967). A rapid assay for polypeptide chain termination. *Biochemical and Biophysical Research Communications*, *28*(5), 773–778.

Cenik, C., Derti, A., Mellor, J. C., Berriz, G. F., & Roth, F. P. (2010). Genome-wide functional analysis of human 5' untranslated region introns. *Genome Biology*, *11*(3), R29.

Center for Devices, & Radiological Health. (2019, September 2). *Laboratory Developed Tests*. https://www.fda.gov/medical-devices/in-vitro-diagnostics/laboratory-developed-tests

Chan, C. W., Kwon, S., Matushek, S. M., Ciaglia, C., Bethel, C., & Beavis, K. G. (2021). Implementation of a Sample Pooling Strategy for the Direct Detection of SARS-CoV-2 by Real-Time Polymerase Chain Reaction During the COVID-19 Pandemic. *American Journal of Clinical Pathology*. https://doi.org/10.1093/ajcp/aqab035

Chemla, Y., Peeri, M., Heltberg, M. L., Eichler, J., Jensen, M. H., Tuller, T., & Alfonta, L. (2020). A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nature Communications*, *11*(1), 4827.

Chin, K., & Pyle, A. M. (1995). Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA* , *1*(4), 391–406.

Clinical Laboratory Improvement Amendments (CLIA). (n.d.). Retrieved June 21, 2021, from https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA

Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & Development*, *11*(21), 2755–2766.

Collier, A. E., Spandau, D. F., & Wek, R. C. (2018). Translational control of a human CDKN1A mRNA splice variant regulates the fate of UVB-irradiated human keratinocytes. *Molecular Biology of the Cell*, *29*(1), 29–41.

Cooke, C., Hans, H., & Alwine, J. C. (1999). Utilization of splicing elements and

polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Molecular and Cellular Biology*, *19*(7), 4971–4979.

Craig, A. W., Haghighat, A., Yu, A. T., & Sonenberg, N. (1998). Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature*, *392*(6675), 520–523.

Crick, F. H., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, *192*, 1227–1232.

Currie, D. W., Moreno, G. K., Delahoy, M. J., Pray, I. W., Jovaag, A., Braun, K. M., Cole, D., Shechter, T., Fajardo, G. C., Griggs, C., Yandell, B. S., Goldstein, S., Bushman, D., Segaloff, H. E., Patrick Kelly, G., Pitts, C., Lee, C., Grande, K. M., Kita-Yarbro, A., … Killerby, M. E. (2021). Description of a University COVID-19 Outbreak and Interventions to Disrupt Transmission, Wisconsin, August – October 2020. *medRxiv*, 2021.05.07.21256834.

de Melo Neto, O. P., Standart, N., & Martins de Sa, C. (1995). Autoregulation of poly(A)-binding protein synthesis in vitro. *Nucleic Acids Research*, *23*(12), 2198–2205.

Dichtl, B., Blank, D., Sadowski, M., Hübner, W., Weiser, S., & Keller, W. (2002). Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *The EMBO Journal*, *21*(15), 4125–4135.

di Penta, A., Mercaldo, V., Florenzano, F., Munck, S., Ciotti, M. T., Zalfa, F., Mercanti, D., Molinari, M., Bagni, C., & Achsel, T. (2009). Dendritic LSm1/CBP80-mRNPs mark the early steps of transport commitment and translational control. *The Journal of Cell Biology*, *184*(3), 423–435.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , *29*(1), 15–21.

Duncan, P. I., Stojdl, D. F., Marius, R. M., & Bell, J. C. (1997). In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Molecular and Cellular*

Biology, *17*(10), 5996–6001.

Editors, T., & The Editors. (2020). Dying in a Leadership Vacuum. In *New England Journal of Medicine* (Vol. 383, Issue 15, pp. 1479–1480). https://doi.org/10.1056/nejme2029812

Elroy-Stein, O., Fuerst, T. R., & Moss, B. (1989). Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5' sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(16), 6126–6130.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74.

Esbin, M. N., Whitney, O. N., Chong, S., Maurer, A., Darzacq, X., & Tjian, R. (2020). Overcoming the bottleneck to widespread testing: a rapid review of nucleic acid testing approaches for COVID-19 detection. *RNA* , *26*(7). https://doi.org/10.1261/rna.076232.120

Fan, X. C., & Steitz, J. A. (1998). Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *The EMBO Journal*, *17*(12), 3448–3460.

Fesenko, I., Khazigaleeva, R., Kirov, I., Kniazev, A., Glushenko, O., Babalyan, K., Arapidi, G., Shashkova, T., Butenko, I., Zgoda, V., Anufrieva, K., Seredina, A., Filippova, A., & Govorun, V. (2017). Alternative splicing shapes transcriptome but not proteome diversity in Physcomitrella patens. *Scientific Reports*, *7*(1), 2698.

Fiddes, I. T., Armstrong, J., Diekhans, M., Nachtweide, S., Kronenberg, Z. N., Underwood, J. G., Gordon, D., Earl, D., Keane, T., Eichler, E. E., Haussler, D., Stanke, M., & Paten, B. (2018). Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Research*, *28*(7), 1029–1038.

Field, A. R., Jacobs, F. M. J., Fiddes, I. T., Phillips, A. P. R., Reyes-Ortiz, A. M., LaMontagne, E., Whitehead, L., Meng, V., Rosenkrantz, J. L., Haeussler, M., Katzman, S., Salama, S. R., & Haussler, D. (2017). *Structurally conserved primate*

lncRNAs are transiently expressed during human cortical differentiation and influence cell type specific genes. https://doi.org/10.1101/232553

Floor, S. N., & Doudna, J. A. (2016). Tunable protein synthesis by transcript isoforms in human cells. In *eLife* (Vol. 5). https://doi.org/10.7554/elife.10921

Ford, L. P., Bagga, P. S., & Wilusz, J. (1997). The poly(A) tail inhibits the assembly of a 3’-to-5' exonuclease in an in vitro RNA stability system. *Molecular and Cellular Biology*, *17*(1), 398–406.

Fu, X.-D., & Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews. Genetics*, *15*, 689.

Gallego-Paez, L. M., Bordone, M. C., Leote, A. C., Saraiva-Agostinho, N., Ascensão-Ferreira, M., & Barbosa-Morais, N. L. (2017). Alternative splicing: the pledge, the turn, and the prestige : The key role of alternative splicing in human biological systems. *Human Genetics*, *136*(9), 1015–1042.

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

Gebauer, F., Preiss, T., & Hentze, M. W. (2012). From cis-regulatory elements to complex RNPs and back. *Cold Spring Harbor Perspectives in Biology*, *4*(7), a012245.

Ge, H., & Manley, J. L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*, *62*(1), 25–34.

Gozani, O., Feld, R., & Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & Development*, *10*(2), 233–243.

Guimaraes, J. C., Mittal, N., Gnann, A., Jedlinski, D., Riba, A., Buczak, K., Schmidt, A., & Zavolan, M. (2020). A rare codon-based translational program of cell proliferation. *Genome Biology*, *21*(1), 44.

Gu, W., Zhou, T., & Wilke, C. O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Computational Biology*, *6*(2), e1000664.

Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M., & Ron, D. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Molecular Cell*, *11*(3), 619–633.

Hau, H. H., Walsh, R. J., Ogilvie, R. L., Williams, D. A., Reilly, C. S., & Bohjanen, P. R. (2007). Tristetraprolin recruits functional mRNA decay complexes to ARE sequences. *Journal of Cellular Biochemistry*, *100*(6), 1477–1492.

Herzel, L., & Neugebauer, K. M. (2015). Quantification of co-transcriptional splicing from RNA-Seq data. *Methods* , *85*, 36–43.

Heyer, E. E., & Moore, M. J. (2016). Redefining the Translational Status of 80S Monosomes. *Cell*, *164*(4), 757–769.

Hidayat, R., Aini, N., Ilmi, A. F. N., Azzahroh, F., & Giantini, A. (2020). Test, Trace, and Treatment Strategy to Control COVID-19 Infection Among Hospital Staff in a COVID-19 Referral Hospital in Indonesia. *Acta Medica Indonesiana*, *52*(3), 206–213.

Hinnebusch, A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology*, *59*, 407–450.

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., & Zamecnik, P. C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of Biological Chemistry*, *231*(1), 241–257.

Holcík, M., Gordon, B. W., & Korneluk, R. G. (2003). The internal ribosome entry site-mediated translation of antiapoptotic protein XIAP is modulated by the heterogeneous nuclear ribonucleoproteins C1 and C2. *Molecular and Cellular Biology*, *23*(1), 280–288.

Huang, Y., & Steitz, J. A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Molecular Cell*, *7*(4), 899–905.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, *324*(5924), 218–223.

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, *147*(4), 789–802.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C., & Mattaj, I. W. (1995). A cap-binding protein complex mediating U snRNA export. *Nature*, *376*(6542), 709–712.

Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews. Molecular Cell Biology*, *11*(2), 113–127.

Ji, H., Fraser, C. S., Yu, Y., Leary, J., & Doudna, J. A. (2004). Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(49), 16990–16995.

Johansson, M. A., Quandelacy, T. M., Kada, S., Prasad, P. V., Steele, M., Brooks, J. T., Slayton, R. B., Biggerstaff, M., & Butler, J. C. (2021). SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Network Open*, *4*(1), e2035057–e2035057.

Johnson, T. L., & Abelson, J. (2001). Characterization of U4 and U6 interactions with the 5’ splice site using a S. cerevisiae in vitro trans-splicing system. *Genes & Development*, *15*(15), 1957–1970.

Kapp, L. D., & Lorsch, J. R. (2004). The molecular mechanics of eukaryotic translation. *Annual Review of Biochemistry*, *73*, 657–704.

Kedersha, N., Cho, M. R., Li, W., Yacono, P. W., Chen, S., Gilks, N., Golan, D. E., & Anderson, P. (2000). Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *The Journal of Cell Biology*, *151*(6), 1257–1268.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, a. D. (2002). The Human Genome Browser at UCSC. In *Genome*

Research (Vol. 12, Issue 6, pp. 996–1006). https://doi.org/10.1101/gr.229102

Khan, Z., Ford, M. J., Cusanovich, D. A., Mitrano, A., Pritchard, J. K., & Gilad, Y. (2013). Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science*, *342*(6162), 1100–1104.

Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J.-I., Sekine, M., Tsuritani, K., Wakaguri, H., Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N., Yoneyama, T., Otsuka, R., Kanda, K., … Sugano, S. (2006). Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Research*, *16*(1), 55–65.

Konarska, M. M., Grabowski, P. J., Padgett, R. A., & Sharp, P. A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature*, *313*(6003), 552–557.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *The Journal of Biological Chemistry*, *266*(30), 19867–19870.

Krainer, A. R., Conway, G. C., & Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*, *62*(1), 35–42.

Krainer, A. R., & Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell*, *42*(3), 725–736.

Laggerbauer, B., Achsel, T., & Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(8), 4188–4192.

Laishram, R. S., & Anderson, R. A. (2010). The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. *The EMBO Journal*, *29*(24), 4132–4145.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and

memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25.

Lareau, L. F., & Brenner, S. E. (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Molecular Biology and Evolution*, *32*(4), 1072–1079.

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

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926–929.

Larremore, D. B., Wilder, B., Lester, E., Shehata, S., Burke, J. M., Hay, J. A., Tambe, M., Mina, M. J., & Parker, R. (2021). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science Advances*, *7*(1). https://doi.org/10.1126/sciadv.abd5393

Leclair, N. K., Brugiolo, M., Urbanski, L., Lawson, S. C., Thakar, K., Yurieva, M., George, J., Hinson, J. T., Cheng, A., Graveley, B. R., & Anczuków, O. (2020). Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. *Molecular Cell*, *80*(4), 648–665.e9.

Lee, A. S. Y., Kranzusch, P. J., & Cate, J. H. D. (2015). eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, *522*(7554), 111–114.

Le Hir, H., Gatfield, D., Izaurralde, E., & Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO Journal*, *20*(17), 4987–4997.

Le Hir, H., Saulière, J., & Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nature Reviews. Molecular Cell Biology*, *17*(1), 41–54.

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling

of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 189–192.

Lewis, S. M., Veyrier, A., Hosszu Ungureanu, N., Bonnal, S., Vagner, S., & Holcik, M. (2007). Subcellular relocalization of a trans-acting factor regulates XIAP IRES-dependent translation. *Molecular Biology of the Cell*, *18*(4), 1302–1311.

Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & Development*, *27*(21), 2380–2396.

Lin, Y., Protter, D. S. W., Rosen, M. K., & Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell*, *60*(2), 208–219.

Lista, M. J., Page, R., Sertkaya, H., Matos, P. M., Ortiz-Zapater, E., Maguire, T. J. A., Poulton, K., O'Byrne, A. M., Bouton, C., Dickenson, R. E., Ficarelli, M., Howard, M., Betancor, G., Galao, R. P., Pickering, S., Signell, A. W., Wilson, H., Cliff, P., Patel, A., … Martinez-Nunez, R. T. (2020). Resilient SARS-CoV-2 diagnostics workflows including viral heat inactivation. *medRxiv*, 2020.04.22.20074351.

Liu-Yesucevitz, L., Lin, A. Y., Ebata, A., Boon, J. Y., Reid, W., Xu, Y.-F., Kobrin, K., Murphy, G. J., Petrucelli, L., & Wolozin, B. (2014). ALS-linked mutations enlarge TDP-43-enriched neuronal RNA granules in the dendritic arbor. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(12), 4167–4174.

Lu, S., & Cullen, B. R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* , *9*(5), 618–630.

Mallapaty, S. (2020). The mathematical strategy that could transform coronavirus testing. *Nature*, *583*(7817), 504–505.

Mandel, C. R., Bai, Y., & Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences: CMLS*, *65*(7-8), 1099–1122.

Mao, Y., Liu, H., Liu, Y., & Tao, S. (2014). Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic Acids Research*, *42*(8), 4813–4822.

Maquat, L. E. (1995). When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA* , *1*(5), 453–465.

Maquat, L. E., Tarn, W.-Y., & Isken, O. (2010). The pioneer round of translation: features and functions. *Cell*, *142*(3), 368–374.

Mašek, T., Valášek, L., & Pospíšek, M. (2011). Polysome analysis and RNA purification from sucrose gradients. *Methods in Molecular Biology* , *703*, 293–309.

Maslon, M. M., Heras, S. R., Bellora, N., Eyras, E., & Cáceres, J. F. (2014). The translational landscape of the splicing factor SRSF1 and its role in mitosis. *eLife*, e02028.

Mauger, D. M., Cabral, B. J., Presnyak, V., Su, S. V., Reid, D. W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M. J., & McFadyen, I. J. (2019). mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(48), 24075–24083.

Mayr, C., & Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, *138*(4), 673–684.

Mazin, P. V., Jiang, X., Fu, N., Han, D., Guo, M., Gelfand, M. S., & Khaitovich, P. (2018). Conservation, evolution, and regulation of splicing during prefrontal cortex development in humans, chimpanzees, and macaques. *RNA* , *24*(4), 585–596.

McDermott, A. (2020). Inner Workings: Molecular biologists offer "wartime service" in the effort to test for COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9656–9659.

Melero, R., Buchwald, G., Castaño, R., Raabe, M., Gil, D., Lázaro, M., Urlaub, H., Conti, E., & Llorca, O. (2012). The cryo-EM structure of the UPF-EJC complex shows UPF1

poised toward the RNA 3' end. *Nature Structural & Molecular Biology*, *19*(5), 498–505, S1–S2.

Merkin, J., Russell, C., Chen, P., & Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*, *338*(6114), 1593–1599.

Mina, M. J., Parker, R., & Larremore, D. B. (2020). Rethinking Covid-19 Test Sensitivity - A Strategy for Containment. *The New England Journal of Medicine*, *383*(22), e120.

Mitchell, S. F., & Parker, R. (2014). Principles and properties of eukaryotic mRNPs. *Molecular Cell*, *54*(4), 547–558.

Movassat, M., Crabb, T. L., Busch, A., Yao, C., Reynolds, D. J., Shi, Y., & Hertel, K. J. (2016). Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biology*, *13*(7), 646–655.

Müller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & Development*, *30*(5), 553–566.

Mulu, A., Alemayehu, D. H., Alemu, F., Tefera, D. A., Wolde, S., Aseffa, G., Seyoum, T., Habtamu, M., Abdissa, A., Bayih, A. G., & Beyene, G. T. (2021). Evaluation of sample pooling for screening of SARS CoV-2. *PloS One*, *16*(2), e0247767.

Nagai, K., Muto, Y., Pomeranz Krummel, D. A., Kambach, C., Ignjatovic, T., Walke, S., & Kuglstatter, A. (2001). Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. *Biochemical Society Transactions*, *29*(Pt 2), 15–26.

Nakayama, T., Asai, S., Takahashi, Y., Maekawa, O., & Kasama, Y. (2007). Overlapping of genes in the human genome. *International Journal of Biomedical Science: IJBS*, *3*(1), 14–19.

Newman, A. J. (1997). The role of U5 snRNP in pre-mRNA splicing. *The EMBO Journal*, *16*(19), 5797–5800.

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., & Ares, M., Jr. (2007). Ultraconserved elements are associated

with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, *21*(6), 708–718.

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., & O'Neal, C. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America*, *53*(5), 1161–1168.

[No title]. (n.d.). Retrieved June 21, 2021, from https://www.nejm.org/doi/full/10.1056/nejmp2014836

Nott, A., Le Hir, H., & Moore, M. J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes & Development*, *18*(2), 210–222.

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2020). *Biostrings: Efficient manipulation of biological strings. R package version 2.48. 0*.

Pandya-Jones, A., & Black, D. L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA* , *15*(10), 1896–1908.

Pan, Q., Saltzman, A. L., Kim, Y. K., Misquitta, C., Shai, O., Maquat, L. E., Frey, B. J., & Blencowe, B. J. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes & Development*, *20*(2), 153–158.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415.

Parker, R., Siliciano, P. G., & Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, *49*(2), 229–239.

Pascolo, E., & Séraphin, B. (1997). The branchpoint residue is recognized during commitment complex formation before being bulged out of the U2 snRNA-pre-mRNA duplex. *Molecular and Cellular Biology*, *17*(7), 3469–3476.

Pedersen, S. K., Christiansen, J., Hansen, T. v. O., Larsen, M. R., & Nielsen, F. C. (2002). Human insulin-like growth factor II leader 2 mediates internal initiation of translation. *Biochemical Journal*, *363*(Pt 1), 37–44.

Pelletier, J., Kaplan, G., Racaniello, V. R., & Sonenberg, N. (1988). Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Molecular and Cellular Biology*, *8*(3), 1103–1112.

Peltz, S. W., Trotta, C., Feng, H., Brown, A., Donahue, J., Welch, E., & Jacobson, A. (1993). Identification of the cis-acting sequences and trans-acting factors involved in nonsense-mediated mRNA decay. *Protein Synthesis and Targeting in Yeast*, 1–10.

Penuliar, M., Clark, C., Curti, D., Hudson, C., & Philips, B. (2020). Universities and COVID-19 Growth at the Start of the 2020 Academic Year. *medRxiv*, 2020.11.25.20238899.

Pérez Cañadillas, J. M., & Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *The EMBO Journal*, *22*(11), 2821–2830.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295.

Pestova, T. V., Hellen, C. U., & Shatsky, I. N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Molecular and Cellular Biology*, *16*(12), 6859–6869.

Pestova, T. V., & Kolupaeva, V. G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes & Development*, *16*(22), 2906–2922.

Petersen, J., Dalal, S., & Jhala, D. (2021). Criticality of In-House Preparation of Viral Transport Medium in Times of Shortage During COVID-19 Pandemic. *Laboratory Medicine*, *52*(2). https://doi.org/10.1093/labmed/lmaa099

Pfeiffer, B. D., Truman, J. W., & Rubin, G. M. (2012). Using translational enhancers to

increase transgene expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(17), 6626–6631.

Pisarev, A. V., Hellen, C. U. T., & Pestova, T. V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. *Cell*, *131*(2), 286–299.

Pollock, B. H., Kilpatrick, A. M., Eisenman, D. P., Elton, K. L., Rutherford, G. W., Boden-Albala, B. M., Souleles, D. M., Polito, L. E., Martin, N. K., & Byington, C. L. (2021). Safe reopening of college campuses during COVID-19: The University of California experience in Fall 2020. *PloS One*, *16*(11), e0258738.

Pöyry, T. A. A., Kaminski, A., & Jackson, R. J. (2004). What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes & Development*, *18*(1), 62–75.

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124.

Racca, C., Gardiol, A., & Triller, A. (1997). Dendritic and postsynaptic localizations of glycine receptor alpha subunit mRNAs. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(5), 1691–1700.

Raghunathan, P. L., & Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current Biology: CB*, *8*(15), 847–855.

Ray, D., Kazan, H., Chan, E. T., Peña Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B. J., Morris, Q., & Hughes, T. R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotechnology*, *27*(7), 667–670.

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, *499*(7457),

172–177.

Reixachs-Solé, M., Ruiz-Orera, J., Albà, M. M., & Eyras, E. (2020). Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nature Communications*, *11*(1), 1768.

Reyes, A., & Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, *46*(2), 582–592.

Roberts, L. O., Seamons, R. A., & Belsham, G. J. (1998). Recognition of picornavirus internal ribosome entry sites within cells; influence of cellular and viral proteins. *RNA* , *4*(5), 520–529.

Rogers, G. W., Jr, Richter, N. J., Lima, W. F., & Merrick, W. C. (2001). Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *The Journal of Biological Chemistry*, *276*(33), 30914–30922.

Rogers, G. W., Jr, Richter, N. J., & Merrick, W. C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *The Journal of Biological Chemistry*, *274*(18), 12236–12244.

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S., & Darnell, J. E., Jr. (1980). The addition of 5’ cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell*, *19*(1), 69–78.

Salimnia, H., Mitchell, R., Gundel, A., Cambell, A., Gammou, F., Chopra, T., & Fairfax, M. (2021). Pooling samples: a testing option for SARS-CoV-2 during a supply shortage. *Diagnostic Microbiology and Infectious Disease*, *99*(1). https://doi.org/10.1016/j.diagmicrobio.2020.115205

Saltzman, A. L., Kim, Y. K., Pan, Q., Fagnani, M. M., Maquat, L. E., & Blencowe, B. J. (2008). Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Molecular and Cellular Biology*, *28*(13), 4320–4330.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating

cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, *320*(5883), 1643–1647.

Sander, B., Golas, M. M., Makarov, E. M., Brahms, H., Kastner, B., Lührmann, R., & Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Molecular Cell*, *24*(2), 267–278.

Sanford, J. R., Ellis, J. D., Cazalla, D., & Cáceres, J. F. (2005). Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(42), 15042–15047.

Sanford, J. R., Gray, N. K., Beckmann, K., & Cáceres, J. F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes & Development*, *18*(7), 755–768.

Sawicki, R., Korona-Glowniak, I., Boguszewska, A., Stec, A., & Polz-Dacewicz, M. (2021). Sample pooling as a strategy for community monitoring for SARS-CoV-2. *Scientific Reports*, *11*(1), 1–8.

Schneider, E. C. (2020). Failing the Test - The Tragic Data Gap Undermining the U.S. Pandemic Response. *The New England Journal of Medicine*, *383*(4), 299–302.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-limiting steps in yeast protein translation. *Cell*, *153*(7), 1589–1601.

Shatkin, A. J. (1985). mRNA cap binding proteins: essential factors for initiating translation. *Cell*, *40*(2), 223–224.

Shin, B.-S., Kim, J.-R., Walker, S. E., Dong, J., Lorsch, J. R., & Dever, T. E. (2011). Initiation factor eIF2γ promotes eIF2-GTP-Met-tRNAi(Met) ternary complex binding to the 40S ribosome. *Nature Structural & Molecular Biology*, *18*(11), 1227–1234.

Siepel, A. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and

yeast genomes. In *Genome Research* (Vol. 15, Issue 8, pp. 1034–1050). https://doi.org/10.1101/gr.3715005

Siepel, A., & Haussler, D. (n.d.). Phylogenetic Hidden Markov Models. In *Statistical Methods in Molecular Evolution* (pp. 325–351). https://doi.org/10.1007/0-387-27733-1_12

Sina Booeshaghi, A., Tan, F., Renton, B., Berger, Z., & Pachter, L. (2020). Markedly heterogeneous COVID-19 testing plans among US colleges and universities. *medRxiv*, 2020.08.09.20171223.

Singh, C. R., He, H., Ii, M., Yamamoto, Y., & Asano, K. (2004). Efficient incorporation of eukaryotic initiation factor 1 into the multifactor complex is critical for formation of functional ribosomal preinitiation complexes in vivo. *The Journal of Biological Chemistry*, *279*(30), 31910–31920.

Singh, G., Pratt, G., Yeo, G. W., & Moore, M. J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annual Review of Biochemistry*, *84*, 325–354.

Singh, R., Valcárcel, J., & Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, *268*(5214), 1173–1176.

Smith, L. E., Potts, H. W. W., Amlôt, R., Fear, N. T., Michie, S., & James Rubin, G. (n.d.). *Adherence to the test, trace and isolate system: results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study)*. https://doi.org/10.1101/2020.09.15.20191957

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

Spasic, M., Friedel, C. C., Schott, J., Kreth, J., Leppek, K., Hofmann, S., Ozgur, S., & Stoecklin, G. (2012). Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genetics*, *8*(1), e1002433.

Stark, H., Dube, P., Lührmann, R., & Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, *409*(6819), 539–542.

Steitz, T. A., & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(14), 6498–6502.

Sterne-Weiler, T., Martinez-Nunez, R. T., Howard, J. M., Cvitovik, I., Katzman, S., Tariq, M. A., Pourmand, N., & Sanford, J. R. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Research*, *23*(10), 1615–1623.

Steward, O., & Levy, W. B. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *2*(3), 284–291.

Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J., & Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *The EMBO Journal*, *20*(7), 1785–1796.

Swartz, J. E., Bor, Y.-C., Misawa, Y., Rekosh, D., & Hammarskjold, M.-L. (2007). The Shuttling SR Protein 9G8 Plays a Role in Translation of Unspliced mRNA Containing a Constitutive Transport Element. In *Journal of Biological Chemistry* (Vol. 282, Issue 27, pp. 19844–19853). https://doi.org/10.1074/jbc.m701660200

Tahmasebi, S., Jafarnejad, S. M., Tam, I. S., Gonatopoulos-Pournatzis, T., Matta-Camacho, E., Tsukumo, Y., Yanagiya, A., Li, W., Atlasi, Y., Caron, M., Braunschweig, U., Pearl, D., Khoutorsky, A., Gkogkas, C. G., Nadon, R., Bourque, G., Yang, X.-J., Tian, B., Stunnenberg, H. G., … Sonenberg, N. (2016). Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12360–12367.

Takahashi, K., Maruyama, M., Tokuzawa, Y., Murakami, M., Oda, Y., Yoshikane, N., Makabe, K. W., Ichisaka, T., & Yamanaka, S. (2005). Evolutionarily conserved

non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). *Genomics*, *85*(3), 360–371.

Taliaferro, J. M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., Saxena, T., Wang, E. T., Graveley, B. R., Gertler, F. B., Swanson, M. S., & Burge, C. B. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell*, *61*(6), 821–833.

Tang, A. D., Soulette, C. M., van Baren, M. J., Hart, K., Hrabeta-Robinson, E., Wu, C. J., & Brooks, A. N. (2020). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nature Communications*, *11*(1), 1438.

Thanh, T. T., Nhan, N. T. T., Mai, H. K., Trieu, N. B., Huy, L. X., Thuy, H. T. T., Chung, L. T., Anh, N. N., Hong, N. T. T., Thang, B. T., Thu, N. T. H., Chi, L. T. K., Hanh, N. T., Hoang, N. H., Chau, N. V. V., Thwaites, G., Hung, D. T., Van Tan, L., & Yen, N. T. K. (2021). The application of sample pooling for mass screening of SARS-CoV-2 in an outbreak of COVID-19 in Vietnam. *The American Journal of Tropical Medicine and Hygiene*, *104*(4), 1531–1534.

Thomas, J. D., Polaski, J. T., Feng, Q., De Neef, E. J., Hoppe, E. R., McSharry, M. V., Pangallo, J., Gabel, A. M., Belleville, A. E., Watson, J., Nkinsi, N. T., Berger, A. H., & Bradley, R. K. (2020). RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nature Genetics*, *52*(1), 84–94.

Tindale, L. C., Stockdale, J. E., Coombe, M., Garlock, E. S., Lau, W. Y. V., Saraswat, M., Zhang, L., Chen, D., Wallinga, J., & Colijn, C. (2020). Evidence for transmission of COVID-19 prior to symptom onset. *eLife*, *9*. https://doi.org/10.7554/eLife.57149

Todd, B. (2020). The U.S. COVID-19 Testing Failure. In *AJN, American Journal of Nursing* (Vol. 120, Issue 10, pp. 19–20). https://doi.org/10.1097/01.naj.0000718596.51921.f2

Tress, M. L., Abascal, F., & Valencia, A. (2017). Most Alternative Isoforms Are Not

Functionally Important [Review of *Most Alternative Isoforms Are Not Functionally Important*]. *Trends in Biochemical Sciences*, *42*(6), 408–410.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, *141*(2), 344–354.

Unbehaun, A., Borukhov, S. I., Hellen, C. U. T., & Pestova, T. V. (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes & Development*, *18*(24), 3078–3093.

Vagner, S., Waysbort, A., Marenda, M., Gensac, M. C., Amalric, F., & Prats, A. C. (1995). Alternative translation initiation of the Moloney murine leukemia virus mRNA controlled by internal ribosome entry involving the p57/PTB splicing factor. *The Journal of Biological Chemistry*, *270*(35), 20376–20383.

Valcárcel, J., Gaur, R. K., Singh, R., & Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*, *273*(5282), 1706–1709.

van den Berg, A., Mols, J., & Han, J. (2008). RISC-target interaction: cleavage and translational suppression. *Biochimica et Biophysica Acta*, *1779*(11), 668–677.

Varenne, S., Buc, J., Lloubes, R., & Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of Molecular Biology*, *180*(3), 549–576.

Volden, R., & Vollmers, C. (n.d.). *Highly Multiplexed Single-Cell Full-Length cDNA Sequencing of human immune cells with 10X Genomics and R2C2*. https://doi.org/10.1101/2020.01.10.902361

Wahle, E., & Keller, W. (1996). The biochemistry of polyadenylation. *Trends in Biochemical Sciences*, *21*(7), 247–250.

Wang, E. T., Cody, N. A. L., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lécuyer, E., & Burge, C. B. (2012).

Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, *150*(4), 710–724.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476.

Wang, X., Codreanu, S. G., Wen, B., Li, K., Chambers, M. C., Liebler, D. C., & Zhang, B. (2018). Detection of Proteome Diversity Resulted from Alternative Splicing is Limited by Trypsin Cleavage Specificity. *Molecular & Cellular Proteomics: MCP*, *17*(3), 422–430.

Wang, Z., Day, N., Trifillis, P., & Kiledjian, M. (1999). An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. *Molecular and Cellular Biology*, *19*(7), 4552–4560.

Wasserman, W. W., & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. In *Nature Reviews Genetics* (Vol. 5, Issue 4, pp. 276–287). https://doi.org/10.1038/nrg1315

Wiegand, H. L., Lu, S., & Cullen, B. R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11327–11332.

Witten, J. T., & Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. *Trends in Genetics: TIG*, *27*(3), 89–97.

Wolin, S. L., & Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal*, *7*(11), 3559–3569.

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., & Smith, C. W. J. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular Cell*, *13*(1), 91–100.

Wong, Q. W.-L., Vaz, C., Lee, Q. Y., Zhao, T. Y., Luo, R., Archer, S. K., Preiss, T., Tanavde, V., & Vardy, L. A. (2016). Embryonic Stem Cells Exhibit mRNA Isoform Specific Translational Regulation. *PloS One*, *11*(1), e0143235.

Woodward, L. A., Mabin, J. W., Gangras, P., & Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdisciplinary Reviews. RNA*, *8*(3). https://doi.org/10.1002/wrna.1411

Wu, C. C.-C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Molecular Cell*, *73*(5), 959–970.e5.

Wu, J. A., & Manley, J. L. (1991). Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature*, *352*(6338), 818–821.

Wu, J., & Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes & Development*, *3*(10), 1553–1561.

Xu, C., Park, J.-K., & Zhang, J. (2019). Evidence that alternative transcriptional initiation is largely nonadaptive. *PLoS Biology*, *17*(3), e3000197.

Yan, D., & Ares, M., Jr. (1996). Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Molecular and Cellular Biology*, *16*(3), 818–828.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., & Darnell, J. E., Jr. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Research*, *13*(8), 1863–1872.

Zahler, A. M., Lane, W. S., Stolk, J. A., & Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & Development*, *6*(5), 837–847.

Zamore, P. D., Patton, J. G., & Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, *355*(6361), 609–614.

Zhang, K., Saxon, A., & Max, E. E. (1992). Two unusual forms of human immunoglobulin E encoded by alternative RNA splicing of epsilon heavy chain membrane exons. *The Journal of Experimental Medicine*, *176*(1), 233–243.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., & Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Research*, *28*(3), 307–322.

Zhang, Z., & Krainer, A. R. (2007). Splicing remodels messenger ribonucleoprotein

architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(28), 11574–11579.