frontiers | Research Topics

RNA machines

Edited by
Nikolay Shirokikh, Kirk Blomquist Jensen,
Jane E. A. Reid and Nehal Thakor

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-3688-9
DOI 10.3389/978-2-8325-3688-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

RNA machines

Topic editors

Nikolay Shirokikh — Australian National University, Australia

Kirk Blomquist Jensen — South Australian Health and Medical Research Institute (SAHMRI), Australia

Jane E. A. Reid — Australian National University, Australia

Nehal Thakor — University of Lethbridge, Canada

Citation

Shirokikh, N., Jensen, K. B., Reid, J. E. A., Thakor, N., eds. (2023). *RNA machines*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3688-9

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 27 September 2023
DOI 10.3389/fgene.2023.1290420

OPEN ACCESS

EDITED AND REVIEWED BY
William C. Cho,
QEH, Hong Kong SAR, China

*CORRESPONDENCE
Nikolay E. Shirokikh,
nikolay.shirokikh@anu.edu.au

RECEIVED 07 September 2023
ACCEPTED 18 September 2023
PUBLISHED 27 September 2023

CITATION
Shirokikh NE, Jensen KB and Thakor N (2023), Editorial: RNA machines. *Front. Genet.* 14:1290420. doi: 10.3389/fgene.2023.1290420

Editorial: RNA machines

Nikolay E. Shirokikh[1]*, Kirk Blomquist Jensen[2] and Nehal Thakor[3]

[1]The John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia, [2]School of Biological Sciences, Faculty of Sciences, University of Adelaide, Adelaide, SA, Australia, [3]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

KEYWORDS

RNA, splicing, non-coding RNA, short non-coding RNA, RNA translational control, RNA viruses, RNA synthetic biology, ribozymes

Editorial on the Research Topic
RNA machines

The omnipotent RNA machines

RNA has been escaping the limelight for decades while it was used to decipher the genetic code (Gardner et al., 1962; Nirenberg et al., 1965; Nirenberg, 2004), establish the rules of complementary nucleotide pairing (Magasanik et al., 1950), develop sequencing techniques (Jou et al., 1972), showcase foundational biomolecule interactions (Fox et al., 2018; Tauber et al., 2020) and assist in protein production (Spirin et al., 1988). The central dogma of molecular biology, an information flow typical to all living things, has been conceived to be centred around RNA (Boivin and Vendrely, 1947). Yet even now, we often employ a term "encoded in the DNA", whereas the actual code is written in the RNA and its decoding is also performed by the ancient RNA molecular machinery of the cells. Notably, newer molecule types, classes and functions consistently emerge and re-emerge in the RNA world. The expanding diversity of RNA modifications and the new varieties of "nonocoding" but information-rich RNA remain as a vast and constantly replenisheable reservoir of biologically active molecules (Mercer et al., 2009; Roundtree et al., 2017). RNA now has been intertwined in more and more intricate cellular and viral processes and activities, tying it into the majority–perhaps, all–of the biological pathways.

Refinement of splicing by and for RNA

A fundamentally distinctive feature of the RNA is its versatility. Consequently, we find the involvement of RNA function across all stages of gene expression. In eukaryotic and some prokaryotic cells, RNA-determined splicing is one of the cornerstone processes of gene expression. Splicing is driven by introns which likely originate from ancient retroelements (RNA retroviruses derivatives) but now are tightly controlled in the cells (Xiong and Eickbush, 1990; Flavell, 1995; Boeke and Stoye, 1997; Koonin, 2006; Hoskins and Moore, 2012; Zimmerly and Semper, 2015). Regulation of intron activity during splicing, such as in alternative splicing, creates an extra layer of diversity which is evolutionarily malleable and is thought to be extremely important in cell differentiation of complex organisms such as vertebrates, and in complex organs such as the brain (Santoni et al., 1989; Hoskins and Moore, 2012; Braunschweig et al., 2013; Irimia et al., 2014; Weatheritt et al., 2016; Ha et al.,

2021). Interrupted genes themselves and splicing have been discovered relatively early in eukaryotic viruses, and then in the genomes of eukaryotic and some prokaryotic cells (Berget et al., 1977; Chow et al., 1977; Darnell, 1978; Keller and Noon, 1984; Apirion and Miczak, 1993; Belfort et al., 1995; Herbert and Rich, 1999; Benler and Koonin, 2022; Vosseberg et al., 2022). The constituents and structures of the major (type I) eukaryotic splicing machinery are well-established, and include the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6, and a number of auxiliary protein factors (Wahl et al., 2009; Will and Lührmann, 2011; Zhao and Pyle, 2017; Kastner et al., 2019; van der Feltz and Hoskins, 2019). The action of snRNA-organised spliceosomal complexes on the introns of nuclear transcripts involves first complex E, where U1 binds to the GU sequence at the 5′splice site donor, U2 Auxiliary Factor 1 binds to the AG sequence at the 3′splice site acceptor, U2 Auxiliary Factor 2 binds polypyrimidine tract between the branch point and 3′splice site, and Splicing Factor 1 binds to the branch point. A series of rearrangements then occurs towards Complex A with U2 binding the branch point, Complex B with U4, U5 and U6 joining and releasing U1, and Complex C with U4 release, lariat formation catalysis by U2 and U6 and then exon ligation (Rogers and Wall, 1980; Maniatis, 1991; Stamm et al., 2005; Wahl et al., 2009; Wang et al., 2015; Kastner et al., 2019). Manuel et al. summarise the impact of alternative splicing on the proteome diversity in a comprehensive review also describing splicing contribution to the generation of noncoding and circular RNAs. They address the important challenges in the prevalence of functional alternative splicing, the artefact-prone difficulties it creates in genome annotations, and the linking between observed RNA splice-types and peptide datasets, including advances and limitations of the current technologies.

Splicing picture was not complete until the alternative “minor” (type II) splicing apparatus has been identified (Tarn and Steitz, 1996; Sharp and Burge, 1997). Initially considered as indeed a “minor” variation that rarely occurs with efficiency and operates with a near-complete own set of snRNPs (U11, U12, U4atac, and U6atac, while U5 remains the same), AU-AC or GU-AG acceptor and donor definitions and the branch sequence, U12-type splicing has been found much more prolific (Park et al., 2016; Moyer et al., 2020; Fast, 2021). While not universally encountered, type II spliceosone and its introns are remarkably conserved (human genome harbours about 700 U12-type introns, but it can be much more (Larue et al., 2021)) and were recently highlighted interacting with the type I splicing machinery in diverse ways. In an insightful review, Akinyi and Frilander enlist cooperation between the U2 and U12 types using the example of *SNRNP48* and *RNPC3* genes (encoding U11 and U12 proteins) where feedback-stabilised U11 and U12 interactions lead to the recognition of 3′splice site 1 and synthesis of a non-productive, nuclear-retained transcript–as opposed to the use of 3′splice site 2 that results in cytoplasmic export and translation. They further showcase examples of direct competition between U2 and U12 types (for *HNRNPLL*, *ZNF207*, *C1orf112*, *NCBP2*, *PRMT1*, *dZRSR2/Urp*, *CTNNBL1*, *CUL4A*, *SPAG16*, *Prospero*, *SRSF10*, *MAPK8/9*, *TMEM87a/b*, *CENATAC* and other genes), or cryptic activation of U2 splicing in a deficit of U12 machinery (for *SNRPE*, *RCD8/EDC4*, *SLC9A8*, *MAPK12*, *LKB1* and other genes), and discuss implications in the context of spliceosomal diseases including Peutz-Jager Syndrome, spondyloepiphyseal dysplasia tarda and Cerebral palsy.

Remarkably, splicing has always been linked to RNA polymerase II transcriptional dynamics (Barrass et al., 2015; Naftelberg et al., 2015; Herzel et al., 2017; Milligan et al., 2017; Ragan et al., 2019). In a similarity to the multitude of transcriptional regulators, splicing is modulated by an array of non-constitutive protein factors (Lin and Fu, 2007; Sapra et al., 2009; Änkö et al., 2012; Vuong et al., 2016). Previously, many of such regulators have been implicated in the production of unusual RNA types such as micro and circular RNA (Melamed et al., 2013; Conn et al., 2015; Salzman, 2016; Eger et al., 2018; Pillman et al., 2018; Ratnadiwakara et al., 2018), cases of alternative processing and splicing of micro-exons (Ustianenko et al., 2017; Torres-Méndez et al., 2019; Head et al., 2021), and are known contributors to the physiologically significant development, differentiation and pathogenesis processes (Irimia et al., 2014; Mochizuki et al., 2021). In an original phylogenetic research article, Huang et al. feature the conservation and diversity of SYF2, an important splicing factor that interacts with cyclin D-type binding-protein 1, a cell cycle regulator at the G1/S transition. SYF2 has been characterised as essential or stimulatory in a variety of proliferative situations, including cancer (Guo et al., 2014; Yan et al., 2015; Zhang et al., 2015). Huang et al. *d*emonstrate conservation of the phylogenetic and splicing patterns of SYF2 in animals, while its mRNA abundance patterns were substantially different across the different tissues of mammals. They demonstrate SYF2 is associated with the occurrence of cancer in breast, lung, spleen and reproductive organs, making SYF2 and its RNA interactors valuable therapeutic targets.

RNA regulators conducting without a code

To act, RNA not necessarily needs to be decoded. Micro RNA and various noncoding RNA functions have been prominent in the transition space from transcriptional to post-transcriptional control and translation (Mehler and Mattick, 2007; Mercer et al., 2009; Änkö and Neugebauer, 2010; Schonrock et al., 2012; Salmanidis et al., 2014; Statello et al., 2021). With the advent of various high-throughput sequencing technologies, we began to substantially broaden the horizons of our understanding towards the “rare” RNA transcript type diversity (Kapranov et al., 2007; Ma et al., 2013; Gil and Ulitsky, 2020). Noncoding RNA of various types have emerged as functionally active in determining cell differentiation and development (Fatica and Bozzoni, 2014). Ni et al. dive into the classification of Terminus-Associated Non-coding RNAs (TANRs; as well as mRNA 5′-end associated noncoding RNAs) in an insightful review of these emerging RNA types. They highlight Terminus-Associated Short Non-coding RNAs (TASRs) and their antisense (aTASRs) varieties, Transcription Termination Site Associated RNAs (TTSa-RNAs), Transcription Boundary-Associated RNAs (TBARs), Terminus-Associated Long RNAs (TALRs) and 3′UTR-associated RNAs (uaRNAs), and explore evidence for their biogenesis, including the same and independent promoter models. Ni et al. further enlist organisation and discovery technology of several prominent TANRs and summarise the demonstrated TANR functions,

including transcriptional interference, promoter and terminator juxtapositioning, transcription termination assistance, micro RNA sponging and sequence-directed RNA cleavage and modification. They note that functionally-relevant TANRs can originate also from long noncoding RNA genes, with *MALAT1* prominently exemplified by its translation-activating MALAT1-associated small cytoplasmic RNA (*MascRNA*) (Wilusz et al., 2012).

While individual micro RNAs may appear as less multifaceted regulators and interactors compared to the longer non-coding RNAs, a developing view is that micro RNAs function in networks, collectively targeting the entire pathways of the cells (Gao, 2008; Ryan et al., 2010; Bracken et al., 2016; Dragomir et al., 2018). Such networks offer a high degree of versatility, sophistication and accuracy of control. An interesting expansion of this idea is presented in a review of Budrass et al. where they thoroughly describe a new micro RNA regulation network and intersect it with the protein control network of a matching complexity, as found for the chaperones Heat Shock Protein 40s (Hsp40s; often referred to as J-proteins by their encoding *DNAJ* genes) (Laufen et al., 1999; Han and Christen, 2003). J-proteins are extremely conserved (from bacteria to human) and function as a "tailoring kit" for situationally activating Hsp70 proteins that have far less client discrimination. J-proteins are devise (over 40 in humans) and possess specific client binding, localisation and additional enzymatic activities (Cyr et al., 1994; Jiang et al., 2019). Budrass et al. review micro RNA target site predictions in J-protein mRNAs, demonstrate their conservation across mammals as well as vertebrates, and intriguingly showcase co-targeting of certain J-protein mRNAs (including *DNAJ A1, A2, B1, B4, B5, B6b, B9, C13, C21,* and *C23*) by micro RNAs of identical and different families, at one or multiple sites, opening new area of complex combinatorial regulatory opportunities.

Translation of RNA and damage control

Decoding of the messenger(m)RNA into the proteins is the most crucial function performed by the RNA (Topisirovic and Sonenberg, 2011; Hershey et al., 2012; Shirokikh, 2022). Translational control is involved in nutrient, stress condition and damage sensing (Holcik and Sonenberg, 2005; Bramham et al., 2016; Ross et al., 2018; 2019; Janapala et al., 2019; Xie et al., 2019). Translation employs the most complete repertoire of RNA activities, including direct basepairing interactions, complex structure formation, functionally modified nucleotides, energised intermediates such as the aminoacyl-tRNA, precise macromolecular interactions as happens in the ribosomal subunit binding and elongation cycle dynamics, and ribozyme catalysis in the ribosomal peptidyl transferase centre. Translation initiation is the most responsive protein biosynthesis regulator in eukaryotes (Kozak, 1992; Pisarev et al., 2005; Sonenberg and Hinnebusch, 2009; Archer et al., 2016; Shirokikh and Preiss, 2018), and within it the accuracy of start codon recognition is especially important, whereby a small mistake can lead to misfolded proteins and adverse cell effects including malignancy (Fekete et al., 2005; Lomakin et al., 2006; Cheung et al., 2007; Asano, 2014; Thakur et al., 2020; Gleason et al., 2022). Accurate start codon recognition involves protein initiator tRNA "carriers"—of which a GTPase eukaryotic translation initiation factor 5B (eIF5B) is the most conserved, having its bacterial and archaeal counterparts (Ross et al., 2018; Shirokikh and Preiss, 2018). In an immersive mini-review, Chukka et al. discuss eIF5B at the crossroads the ribosomal, transfer and messenger RNA interactions. They highlight canonical and most conserved eIF5B functions in initiator tRNA stabilisation on the ribosomal small subunit and subunit joining. Chukka et al. also provide an outlook into the less obvious eIF5B activities in checkpointing eukaryotic small subunit maturation, conveying initiation with the alternative initiator tRNA carrier eIF2A active in certain stresses and interacting with viral (*e.g.*, hepatitis C virus and classical swine fever virus) and cellular (*e.g.*, X-linked inhibitor of apoptosis) internal ribosome entry sites (IRESes). eIF5B action in upstream Open Reading Frame (uORF)-regulated genes is reviewed and its overall cell survival-promoting and thus, malignancy-maintaining role is brought into the focus as an attractive drug target.

Another interesting activity of RNA tightly linked to translation is that of a cellular protection and damage sensing. Oxygenation and oxidative environments present a substantial challenge to the nucleic acids-based life, and especially so to the RNA which can become oxidised in a diverse ways. It has been known that oxidised RNA induces translational errors and may be neurodegenerative or carcinogenic (Tanaka et al., 2007; Fimognari, 2015; Guo et al., 2020). Seixas et al. thoroughly present types of oxidative RNA damage in a mini review covering the most ubiquitous oxidating agents, RNA injuries including 8-oxo-7,8-dihydroguanosine, 8-oxo-7,8-dihydroadenosine, 5-hydroxycytosine and 5-hydroxyuridine, and their effects on mRNA, tRNA, ribosomal and micro RNA function. Seixas et al. emphasise the known passive (scavenging) and active (repair) RNA injury protection systems in prokaryotes noting open questions in comprehensive identification of these components across all life.

Multifunctional RNA in viruses and synthetic biology

RNA is often multifunctional in viruses where there can be certain restrictions on genome size, and in synthetic biology designs where vector and delivery limitations apply to the RNA length, together with the considerations of economy and cost (Afonin et al., 2014; Rossetto and Pari, 2014; Dao et al., 2015; Richert-Pöggeler et al., 2021). From its discovery, Human Immunodeficiency Virus 1 (HIV-1) has been intriguing the researchers with the multitude of functions of its highly-structured RNA modules, containing proteins in all three open reading frames and often with an overlap (splicing-, scanning-, frameshifting- and shunting-controlled), which can be synthesised from the 5′cap or IRES (Ohlmann et al., 2014; Guerrero et al., 2015; Reitz and Gallo, 2015; De Breyne and Ohlmann, 2019). In the DNA-integrated form the HIV-1 provirus can produce partially-spliced and fully-spliced transcripts, among the latter the *tat* mRNA, encoding the essential Tat transcriptional regulator of the virus. All viral transcripts share the initial 289 nt and thus the 59 nt of the highly structured *trans*-activation response (TAR) RNA element. TAR controls host translation via activation and suppression of Protein Kinase RNA-activated (PKR), and further activates viral transcription upon Tat binding (Guerrero et al., 2015). Inspired by their recent discovery that *tat* can have an IRES element

active in latent infection (Khoury et al., 2020), Khoury et al. in an original research article embarked on a *tour de force* to explore cellular proteins possibly interacting with *tat* and modulating it. Discovering 243 significantly interacting proteins by *tat*-3×MS2-stem-loop-directed pull-down and mass spectrometry in latent and productive T-cells, they used knockdown of several top hits to identify Signal Recognition Particle 14 (SRP14) and High-mobility group box 3 (HMGB3) proteins affecting HIV infection the most. Using RNA modification protection, Khoury et al. located the SRP14 and HMGB3 binding sites nearby the *tat* start codon. Most intriguingly, SRP14 and HMGB3 negatively regulated latent and productive infection, while stimulating and repressing Tat synthesis, respectively. Khoury et al. propose SRP14 and HMGB3 alter the efficiency of the *tat* IRES, opening new depths in the lentiviral host interactions and additional pathways to manipulate HIV reactivation.

Interestingly, HIV-1 RNA may contain other RNA regulators, riboswitches (Ooms et al., 2004; Boeras et al., 2017). Riboswitches are compact structural modules of RNA conditionally obstructing (or promoting) a certain process (Garst et al., 2011; Breaker, 2012). Riboswitches can be sensitive to an interaction with another macromolecule or a small compound, or physical conditions such as temperature, pH, salinity, *etc.*, (Mironov et al., 2002; Nahvi et al., 2002; Serganov and Nudler, 2013). Riboswitches can be placed ahead of an "amplifying" stage of a synthetic construct expression, such as transcription or translation, and thus are among the most interesting synthetic biology tools (Breaker, 2018; Kavita and Breaker, 2022). In a brief research report, Korniakova et al. present a new fluoride-sensitive vector design incorporating a fluoride riboswitch (Ren et al., 2012) in the reporter 5′UTR, downstream of the testable promoter region. The plasmid allows to decouple cloning of powerful and potentially cell-damaging promoters from their functional testing, while maintaining same arrangement of the vector for the ease of cloning and comparisons.

Perspectives for RNA as a molecular machine of design

It may not be an overstatement to name the RNA an ultimate molecular machine of life. RNA often performs in relatively straightforward ways built on direct molecular recognition through tertiary structure and basepairing, as happens during micro RNA target binding, and in distinct enzymatic reactions, as occurs during the intron lariat formation and excision. In the other cases, RNA performs as the structural and enzymatic core of conveying molecular machines such as the ribosome, where it uses chemical energy to process, transform and realise biological information. It is quite remarkable that the RNA can "work" with all types of biological macromolecules, be an enzyme and a substrate, carry and decode the genetic information, signal, receive and operate with chemical potentials, making it a "complete", self-sufficient molecule. This self-sufficiency contains a value for novel synthetic biology designs, that is, being recognised in RNA vaccines and gene replacement therapeutics of more sophisticated construction, such as self-amplifying RNA (Rodríguez-Gascón et al., 2014; Brito et al., 2015; Pardi et al., 2018; Dolgin, 2021). It also contains a substantial combinatorial challenge of finding an optimal function in a sea of interactions and activities. We can hope to keep learning from extant (and extinct) life to identify new elements of RNA control, and employ approaches based on artificial intelligence to devise new RNA modules and their applications (Lv et al., 2021; Mohanty and Mohanty, 2021). Thankfully and as exemplified in this Research Topic, we cannot stop to continuously discover new, sometimes unexpected, nuances of the RNA-based processes.

Author contributions

NS: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing. KJ: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing. NT: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing.

Funding

This work was supported by the National Health and Medical Research Council (NHMRC) Investigator grant (GNT1175388) and The Bootes Foundation grant (to NS); New Frontiers in Research Fund Exploration grant (NFRFE-2019-01047), the Natural Sciences and Engineering Research Council of Canada-Discovery Grant Program (RGPIN-2017-05463) (to NT).

Acknowledgments

The authors acknowledge the support of their group members in preparing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Afonin, K. A., Viard, M., Koyfman, A. Y., Martins, A. N., Kasprzak, W. K., Panigaj, M., et al. (2014). Multifunctional RNA nanoparticles. *Nano Lett.* 14, 5662–5671. doi:10.1021/nl502385k

Änkö, M.-L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., et al. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* 13, R17. doi:10.1186/gb-2012-13-3-r17

Änkö, M.-L., and Neugebauer, K. M. (2010). Long noncoding RNAs add another layer to pre-mRNA splicing regulation. *Mol. Cell* 39, 833–834. doi:10.1016/j.molcel.2010.09.003

Apirion, D., and Miczak, A. (1993). RNA processing in prokaryotic cells. *BioEssays* 15, 113–120. doi:10.1002/bies.950150207

Archer, S. K., Shirokikh, N. E., Beilharz, T. H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* 535, 570–574. doi:10.1038/nature18647

Asano, K. (2014). Why is start codon selection so precise in eukaryotes? *Translation* 2, e28387. doi:10.4161/trla.28387

Barrass, J. D., Reid, J. E. A., Huang, Y., Hector, R. D., Sanguinetti, G., Beggs, J. D., et al. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biol.* 16, 282. doi:10.1186/s13059-015-0848-1

Belfort, M., Reaban, M. E., Coetzee, T., and Dalgaard, J. Z. (1995). Prokaryotic introns and inteins: A panoply of form and function. *J. Bacteriol.* 177, 3897–3903. doi:10.1128/jb.177.14.3897-3903.1995

Benler, S., and Koonin, E. V. (2022). Recruitment of mobile genetic elements for diverse cellular functions in prokaryotes. *Front. Mol. Biosci.* 9, 821197. doi:10.3389/fmolb.2022.821197

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci.* 74, 3171–3175. doi:10.1073/pnas.74.8.3171

Boeke, J., and Stoye, J. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelement. Available at: https://www.ncbi.nlm.nih.gov/books/NBK19468 (Accessed January 7, 2023).

Boeras, I., Seufzer, B., Brady, S., Rendahl, A., Heng, X., and Boris-Lawrie, K. (2017). The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Sci. Rep.* 7, 6902. doi:10.1038/s41598-017-06883-9

Boivin, A., and Vendrely, R. (1947). On the possible role of the two nuclear acids in the living cell. *Experientia* 3, 32–34. doi:10.1007/BF02155119

Bracken, C. P., Scott, H. S., and Goodall, G. J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. *Nat. Rev. Genet.* 17, 719–732. doi:10.1038/nrg.2016.134

Bramham, C. R., Jensen, K. B., and Proud, C. G. (2016). Tuning specific translation in cancer metastasis and synaptic memory: control at the MNK–eIF4E Axis. *Trends Biochem. Sci.* 41, 847–858. doi:10.1016/j.tibs.2016.07.008

Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152, 1252–1269. doi:10.1016/j.cell.2013.02.034

Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* 4, a003566. doi:10.1101/cshperspect.a003566

Breaker, R. R. (2018). Riboswitches and translation control. *Cold Spring Harb. Perspect. Biol.* 10, a032797. doi:10.1101/cshperspect.a032797

Brito, L. A., Kommareddy, S., Maione, D., Uematsu, Y., Giovani, C., Berlanda Scorza, F., et al. (2015). "Chapter seven - self-amplifying mRNA vaccines," in Advances in genetics *nonviral vectors for gene therapy*. Editors L. Huang, D. Liu, and E. Wagner (Cambridge, MA: Academic Press), 179–233. doi:10.1016/bs.adgen.2014.10.005

Cheung, Y.-N., Maag, D., Mitchell, S. F., Fekete, C. A., Algire, M. A., Takacs, J. E., et al. (2007). Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection *in vivo*. *Genes Dev.* 21, 1217–1230. doi:10.1101/gad.1528307

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8. doi:10.1016/0092-8674(77)90180-5

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., et al. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134. doi:10.1016/j.cell.2015.02.014

Cyr, D. M., Langer, T., and Douglas, M. G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* 19, 176–181. doi:10.1016/0968-0004(94)90281-X

Dao, B. N., Viard, M., Martins, A. N., Kasprzak, W. K., Shapiro, B. A., and Afonin, K. A. (2015). Triggering RNAi with multifunctional RNA nanoparticles and their delivery. *DNA RNA Nanotechnol.* 2, 1–12. doi:10.1515/rnan-2015-0001

Darnell, J. E. (1978). Implications of RNA·RNA splicing in evolution of eukaryotic cells. *Science* 202, 1257–1260. doi:10.1126/science.364651

De Breyne, S., and Ohlmann, T. (2019). Focus on translation initiation of the HIV-1 mRNAs. *Int. J. Mol. Sci.* 20, 101. doi:10.3390/ijms20010101

Dolgin, E. (2021). The tangled history of mRNA vaccines. *Nature* 597, 318–324. doi:10.1038/d41586-021-02483-w

Dragomir, M., Mafra, A. C. P., Dias, S. M. G., Vasilescu, C., and Calin, G. A. (2018). Using microRNA networks to understand cancer. *Int. J. Mol. Sci.* 19, 1871. doi:10.3390/ijms19071871

Eger, N., Schoppe, L., Schuster, S., Laufs, U., and Boeckel, J.-N. (2018). "Circular RNA splicing," in Circular RNAs: Biogenesis and functions *advances in experimental medicine and biology*. Editor J. Xiao (Singapore: Springer), 41–52. doi:10.1007/978-981-13-1426-1_4

Fast, N. M. (2021). Intron splicing: U12 spliceosomal introns not so 'minor' after all. *Curr. Biol.* 31, R912–R914. doi:10.1016/j.cub.2021.06.008

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi:10.1038/nrg3606

Fekete, C. A., Applefield, D. J., Blakely, S. A., Shirokikh, N., Pestova, T., Lorsch, J. R., et al. (2005). The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection *in vivo*. *EMBO J.* 24, 3588–3601. doi:10.1038/sj.emboj.7600821

Fimognari, C. (2015). Role of oxidative RNA damage in chronic-degenerative diseases. *Oxidative Med. Cell. Longev.* 2015, e358713. doi:10.1155/2015/358713

Flavell, A. J. (1995). Retroelements, reverse transcriptase and evolution. *Comp. Biochem. Physiology Part B Biochem. Mol. Biol.* 110, 3–15. doi:10.1016/0305-0491(94)00122-B

Fox, A. H., Nakagawa, S., Hirose, T., and Bond, C. S. (2018). Paraspeckles: where long noncoding RNA meets phase separation. *Trends Biochem. Sci.* 43, 124–135. doi:10.1016/j.tibs.2017.12.001

Gao, F.-B. (2008). Posttranscriptional control of neuronal development by microRNA networks. *Trends Neurosci.* 31, 20–26. doi:10.1016/j.tins.2007.10.004

Gardner, R. S., Wahba, A. J., Basilio, C., Miller, R. S., Lengyel, P., and Speyer, J. F. (1962). Synthetic polynucleotides and the amino acid code, vii. *Proc. Natl. Acad. Sci.* 48, 2087–2094. doi:10.1073/pnas.48.12.2087

Garst, A. D., Edwards, A. L., and Batey, R. T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* 3, a003533. doi:10.1101/cshperspect.a003533

Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21, 102–117. doi:10.1038/s41576-019-0184-5

Gleason, A. C., Ghadge, G., Sonobe, Y., and Roos, R. P. (2022). Kozak similarity score algorithm identifies alternative translation initiation codons implicated in cancers. *Int. J. Mol. Sci.* 23, 10564. doi:10.3390/ijms231810564

Guerrero, S., Batisse, J., Libre, C., Bernacchi, S., Marquet, R., and Paillart, J.-C. (2015). HIV-1 replication and the cellular eukaryotic translation apparatus. *Viruses* 7, 199–218. doi:10.3390/v7010199

Guo, C., Chen, Q., Chen, J., Yu, J., Hu, Y., Zhang, S., et al. (2020). 8-Hydroxyguanosine as a possible RNA oxidative modification marker in urine from colorectal cancer patients: evaluation by ultra performance liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B* 1136, 121931. doi:10.1016/j.jchromb.2019.121931

Guo, J., Yang, L., Huang, J., Liu, X., Qiu, X., Tao, T., et al. (2014). Knocking down the expression of SYF2 inhibits the proliferation of glioma cells. *Med. Oncol.* 31, 101. doi:10.1007/s12032-014-0101-x

Ha, K. C. H., Sterne-Weiler, T., Morris, Q., Weatheritt, R. J., and Blencowe, B. J. (2021). Differential contribution of transcriptomic regulatory layers in the definition of neuronal identity. *Nat. Commun.* 12, 335. doi:10.1038/s41467-020-20483-8

Han, W., and Christen, P. (2003). Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system. *J. Biol. Chem.* 278, 19038–19043. doi:10.1074/jbc.M300756200

Head, S. A., Hernandez-Alias, X., Yang, J.-S., Ciampi, L., Beltran-Sastre, V., Torres-Méndez, A., et al. (2021). Silencing of SRRM4 suppresses microexon inclusion and promotes tumor growth across cancers. *PLOS Biol.* 19, e3001138. doi:10.1371/journal.pbio.3001138

Herbert, A., and Rich, A. (1999). RNA processing and the evolution of eukaryotes. *Nat. Genet.* 21, 265–269. doi:10.1038/6780

Hershey, J. W. B., Sonenberg, N., and Mathews, M. B. (2012). Principles of translational control: an overview. *Cold Spring Harb. Perspect. Biol.* 4, a011528. doi:10.1101/cshperspect.a011528

Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* 18, 637–650. doi:10.1038/nrm.2017.63

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol.* 6, 318–327. doi:10.1038/nrm1618

Hoskins, A. A., and Moore, M. J. (2012). The spliceosome: A flexible, reversible macromolecular machine. *Trends Biochem. Sci.* 37, 179–188. doi:10.1016/j.tibs.2012.02.009

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved Program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi:10.1016/j.cell.2014.11.035

Janapala, Y., Preiss, T., and Shirokikh, N. E. (2019). Control of translation at the initiation phase during glucose starvation in yeast. *Int. J. Mol. Sci.* 20, 4043. doi:10.3390/ijms20164043

Jiang, Y., Rossi, P., and Kalodimos, C. G. (2019). Structural basis for client recognition and activity of Hsp40 chaperones. *Science* 365, 1313–1319. doi:10.1126/science.aax1280

Jou, W. M., Haegeman, G., Ysebaert, M., and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237, 82–88. doi:10.1038/237082a0

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi:10.1126/science.1138341

Kastner, B., Will, C. L., Stark, H., and Lührmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* 11, a032417. doi:10.1101/cshperspect.a032417

Kavita, K., and Breaker, R. R. (2022). Discovering riboswitches: the past and the future. *Trends Biochem. Sci.* 48, 119–141. doi:10.1016/j.tibs.2022.08.009

Keller, E. B., and Noon, W. A. (1984). Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci.* 81, 7417–7420. doi:10.1073/pnas.81.23.7417

Khoury, G., Mackenzie, C., Ayadi, L., Lewin, S. R., Branlant, C., and Purcell, D. F. J. (2020). Tat IRES modulator of tat mRNA (TIM-TAM): A conserved RNA structure that controls tat expression and acts as a switch for HIV productive and latent infection. *Nucleic Acids Res.* 48, 2643–2660. doi:10.1093/nar/gkz1181

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biol. Direct* 1, 22. doi:10.1186/1745-6150-1-22

Kozak, M. (1992). Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* 8, 197–225. doi:10.1146/annurev.cb.08.110192.001213

Larue, G. E., Eliáš, M., and Roy, S. W. (2021). Expansion and transformation of the minor spliceosomal system in the slime mold Physarum polycephalum. *Curr. Biol.* 31, 3125–3131.e4. doi:10.1016/j.cub.2021.04.050

Laufen, T., Mayer, M. P., Beisel, C., Klostermeier, D., Mogk, A., Reinstein, J., et al. (1999). Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc. Natl. Acad. Sci.* 96, 5452–5457. doi:10.1073/pnas.96.10.5452

Lin, S., and Fu, X.-D. (2007). SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.* 623, 107–122. doi:10.1007/978-0-387-77374-2_7

Lomakin, I. B., Shirokikh, N. E., Yusupov, M. M., Hellen, C. U., and Pestova, T. V. (2006). The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* 25, 196–210. doi:10.1038/sj.emboj.7600904

Lv, H., Shi, L., Berkenpas, J. W., Dao, F.-Y., Zulfiqar, H., Ding, H., et al. (2021). Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. *Briefings Bioinforma.* 22, bbab320. doi:10.1093/bib/bbab320

Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 925–933. doi:10.4161/rna.24604

Magasanik, B., Vischer, E., Doniger, R., Elson, D., and Chargaff, E. (1950). The separation and estimation of ribonucleotides in minute quantities. *J. Biol. Chem.* 186, 37–50. doi:10.1016/s0021-9258(18)56284-0

Maniatis, T. (1991). Mechanisms of alternative pre-mRNA splicing. *Science* 251, 33–34. doi:10.1126/science.1824726

Mehler, M. F., and Mattick, J. S. (2007). Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* 87, 799–823. doi:10.1152/physrev.00036.2006

Melamed, Z., Levy, A., Ashwal-Fluss, R., Lev-Maor, G., Mekahel, K., Atias, N., et al. (2013). Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. *Mol. Cell* 50, 869–881. doi:10.1016/j.molcel.2013.05.007

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi:10.1038/nrg2521

Milligan, L., Sayou, C., Tuck, A., Auchynnikava, T., Reid, J. E., Alexander, R., et al. (2017). RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. *eLife* 6, e27082. doi:10.7554/eLife.27082

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., et al. (2002). Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* 111, 747–756. doi:10.1016/S0092-8674(02)01134-0

Mochizuki, Y., Funayama, R., Shirota, M., Kikukawa, Y., Ohira, M., Karasawa, H., et al. (2021). Alternative microexon splicing by RBFOX2 and PTBP1 is associated with metastasis in colorectal cancer. *Int. J. Cancer* 149, 1787–1800. doi:10.1002/ijc.33758

Mohanty, E., and Mohanty, A. (2021). Role of artificial intelligence in peptide vaccine design against RNA viruses. *Inf. Med. Unlocked* 26, 100768. doi:10.1016/j.imu.2021.100768

Moyer, D. C., Larue, G. E., Hershberger, C. E., Roy, S. W., and Padgett, R. A. (2020). Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.* 48, 7066–7078. doi:10.1093/nar/gkaa464

Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198. doi:10.1146/annurev-biochem-060614-034242

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., and Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. *Chem. Biol.* 9, 1043–1049. doi:10.1016/S1074-5521(02)00224-7

Nirenberg, M. (2004). Historical review: deciphering the genetic code – a personal account. *Trends Biochem. Sci.* 29, 46–54. doi:10.1016/j.tibs.2003.11.009

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., et al. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. U. S. A.* 53, 1161–1168. doi:10.1073/pnas.53.5.1161

Ohlmann, T., Mengardi, C., and López-Lastra, M. (2014). Translation initiation of the HIV-1 mRNA. *Transl. (Austin)* 2, e960242. doi:10.4161/2169074X.2014.960242

Ooms, M., Huthoff, H., Russell, R., Liang, C., and Berkhout, B. (2004). A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* 78, 10814–10819. doi:10.1128/JVI.78.19.10814-10819.2004

Pardi, N., Hogan, M. J., Porter, F. W., and Weissman, D. (2018). mRNA vaccines — A new era in vaccinology. *Nat. Rev. Drug Discov.* 17, 261–279. doi:10.1038/nrd.2017.243

Park, S. J., Jung, H. J., Nguyen Dinh, S., and Kang, H. (2016). Structural features important for the U12 snRNA binding and minor spliceosome assembly of Arabidopsis U11/U12-small nuclear ribonucleoproteins. *RNA Biol.* 13, 670–679. doi:10.1080/15476286.2016.1191736

Pillman, K. A., Phillips, C. A., Roslan, S., Toubia, J., Dredge, B. K., Bert, A. G., et al. (2018). miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking. *EMBO J.* 37, e99016. doi:10.15252/embj.201899016

Pisarev, A. V., Shirokikh, N. E., and Hellen, C. U. T. (2005). Translation initiation by factor-independent binding of eukaryotic ribosomes to internal ribosomal entry sites. *Comptes Rendus Biol.* 328, 589–605. doi:10.1016/j.crvi.2005.02.004

Ragan, C., Goodall, G. J., Shirokikh, N. E., and Preiss, T. (2019). Insights into the biogenesis and potential functions of exonic circular RNA. *Sci. Rep.* 9, 2048. doi:10.1038/s41598-018-37037-0

Ratnadiwakara, M., Mohenska, M., and Änkö, M.-L. (2018). Splicing factors as regulators of miRNA biogenesis – links to human disease. *Seminars Cell & Dev. Biol.* 79, 113–122. doi:10.1016/j.semcdb.2017.10.008

Reitz, M. S., and Gallo, R. C. (2015). "171 - human immunodeficiency viruses," in *Mandell, douglas, and bennett's principles and practice of infectious diseases*. Editors J. E. Bennett, R. Dolin, and M. J. Blaser (Philadelphia: W.B. Saunders), 2054–2065. doi:10.1016/B978-1-4557-4801-3.00171-5

Ren, A., Rajashankar, K. R., and Patel, D. J. (2012). Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. *Nature* 486, 85–89. doi:10.1038/nature11152

Richert-Pöggeler, K. R., Vijverberg, K., Alisawi, O., Chofong, G. N., Heslop-Harrison, J. S., Pat), and Schwarzacher, T. (2021). Participation of multifunctional RNA in replication, recombination and regulation of endogenous plant pararetroviruses (EPRVs). *Front. Plant Sci.* 12, 689307. doi:10.3389/fpls.2021.689307

Rodríguez-Gascón, A., del Pozo-Rodríguez, A., and Solinís, M. Á. (2014). Development of nucleic acid vaccines: use of self-amplifying RNA in lipid nanoparticles. *Int. J. Nanomedicine* 9, 1833–1843. doi:10.2147/IJN.S39810

Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. *Proc. Natl. Acad. Sci.* 77, 1877–1879. doi:10.1073/pnas.77.4.1877

Ross, J. A., Bressler, K. R., and Thakor, N. (2018). Eukaryotic initiation factor 5B (eIF5B) cooperates with eIF1A and eIF5 to facilitate uORF2-mediated repression of ATF4 translation. *Int. J. Mol. Sci.* 19, 4032. doi:10.3390/ijms19124032

Ross, J. A., Dungen, K. V., Bressler, K. R., Fredriksen, M., Khandige Sharma, D., Balasingam, N., et al. (2019). Eukaryotic initiation factor 5B (eIF5B) provides a critical cell survival switch to glioblastoma cells via regulation of apoptosis. *Cell Death Dis.* 10, 57–15. doi:10.1038/s41419-018-1283-5

Rossetto, C. C., and Pari, G. S. (2014). PAN's labyrinth: molecular biology of kaposi's sarcoma-associated herpesvirus (KSHV) PAN RNA, a multifunctional long noncoding RNA. *Viruses* 6, 4212–4226. doi:10.3390/v6114212

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi:10.1016/j.cell.2017.05.045

Ryan, B. M., Robles, A. I., and Harris, C. C. (2010). Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer* 10, 389–402. doi:10.1038/nrc2867

Salmanidis, M., Pillman, K., Goodall, G., and Bracken, C. (2014). Direct transcriptional regulation by nuclear microRNAs. *Int. J. Biochem. Cell Biol.* 54, 304–311. doi:10.1016/j.biocel.2014.03.010

Salzman, J. (2016). Circular RNA expression: its potential regulation and function. *Trends Genet.* 32, 309–316. doi:10.1016/j.tig.2016.03.002

Santoni, M. J., Barthels, D., Vopper, G., Boned, A., Goridis, C., and Wille, W. (1989). Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. *EMBO J.* 8, 385–392. doi:10.1002/j.1460-2075.1989.tb03389.x

Sapra, A. K., Änkö, M.-L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., et al. (2009). SR protein family members display diverse activities in the formation of nascent and mature mRNPs *in vivo*. *Mol. Cell* 34, 179–190. doi:10.1016/j.molcel.2009.02.031

Schonrock, N., Harvey, R. P., and Mattick, J. S. (2012). Long noncoding RNAs in cardiac development and pathophysiology. *Circulation Res.* 111, 1349–1362. doi:10.1161/CIRCRESAHA.112.268953

Serganov, A., and Nudler, E. (2013). A decade of riboswitches. *Cell* 152, 17–24. doi:10.1016/j.cell.2012.12.024

Sharp, P. A., and Burge, C. B. (1997). Classification of introns: U2-Type or U12-type. *Cell* 91, 875–879. doi:10.1016/S0092-8674(00)80479-1

Shirokikh, N. E., and Preiss, T. (2018). Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *WIREs RNA* 9, e1473. doi:10.1002/wrna.1473

Shirokikh, N. E. (2022). Translation complex stabilization on messenger RNA and footprint profiling to study the RNA responses and dynamics of protein biosynthesis in the cells. *Crit. Rev. Biochem. Mol. Biol.* 57, 261–304. doi:10.1080/10409238.2021.2006599

Sonenberg, N., and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745. doi:10.1016/j.cell.2009.01.042

Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S., and Alakhov, Y. B. (1988). A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 242, 1162–1164. doi:10.1126/science.3055301

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., et al. (2005). Function of alternative splicing. *Gene* 344, 1–20. doi:10.1016/j.gene.2004.10.022

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9

Tanaka, M., Chock, P. B., and Stadtman, E. R. (2007). Oxidized messenger RNA induces translation errors. *Proc. Natl. Acad. Sci.* 104, 66–71. doi:10.1073/pnas.0609737104

Tarn, W.-Y., and Steitz, J. A. (1996). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT–AC) intron *in vitro*. *Cell* 84, 801–811. doi:10.1016/S0092-8674(00)81057-0

Tauber, D., Tauber, G., Khong, A., Van Treeck, B., Pelletier, J., and Parker, R. (2020). Modulation of RNA condensation by the DEAD-box protein eIF4A. *Cell* 180, 411–426. doi:10.1016/j.cell.2019.12.031

Thakur, A., Gaikwad, S., Vijjamarri, A. K., and Hinnebusch, A. G. (2020). eIF2α interactions with mRNA control accurate start codon selection by the translation preinitiation complex. *Nucleic Acids Res.* 48, 10280–10296. doi:10.1093/nar/gkaa761

Topisirovic, I., and Sonenberg, N. (2011). Translational control by the eukaryotic ribosome. *Cell* 145, 333–334. doi:10.1016/j.cell.2011.04.006

Torres-Méndez, A., Bonnal, S., Marquez, Y., Roth, J., Iglesias, M., Permanyer, J., et al. (2019). A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat. Ecol. Evol.* 3, 691–701. doi:10.1038/s41559-019-0813-6

Ustianenko, D., Weyn-Vanhentenryck, S. M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. *WIREs RNA* 8, e1418. doi:10.1002/wrna.1418

van der Feltz, C., and Hoskins, A. A. (2019). Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Crit. Rev. Biochem. Mol. Biol.* 54, 443–465. doi:10.1080/10409238.2019.1691497

Vosseberg, J., Schinkel, M., Gremmen, S., and Snel, B. (2022). The spread of the first introns in proto-eukaryotic paralogs. *Commun. Biol.* 5, 476–479. doi:10.1038/s42003-022-03426-5

Vuong, C. K., Black, D. L., and Zheng, S. (2016). The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* 17, 265–281. doi:10.1038/nrn.2016.27

Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi:10.1016/j.cell.2009.02.009

Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., et al. (2015). Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158. doi:10.3892/br.2014.407

Weatheritt, R. J., Sterne-Weiler, T., and Blencowe, B. J. (2016). The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123. doi:10.1038/nsmb.3317

Will, C. L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707. doi:10.1101/cshperspect.a003707

Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L., and Sharp, P. A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* 26, 2392–2407. doi:10.1101/gad.204438.112

Xie, J., de Souza Alves, V., von der Haar, T., O'Keefe, L., Lenchine, R. V., Jensen, K. B., et al. (2019). Regulation of the elongation phase of protein synthesis enhances translation accuracy and modulates lifespan. *Curr. Biol.* 29, 737–749. doi:10.1016/j.cub.2019.01.029

Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362. doi:10.1002/j.1460-2075.1990.tb07536.x

Yan, S., Deng, Y., Qiang, Y., Xi, Q., Liu, R., Yang, S., et al. (2015). SYF2 is upregulated in human epithelial ovarian cancer and promotes cell proliferation. *Tumour Biol.* 36, 4633–4642. doi:10.1007/s13277-015-3111-1

Zhang, S., Shi, W., Chen, Y., Xu, Z., Zhu, J., Zhang, T., et al. (2015). Overexpression of SYF2 correlates with enhanced cell growth and poor prognosis in human hepatocellular carcinoma. *Mol. Cell Biochem.* 410, 1–9. doi:10.1007/s11010-015-2533-9

Zhao, C., and Pyle, A. M. (2017). Structural insights into the mechanism of group II intron splicing. *Trends Biochem. Sci.* 42, 470–482. doi:10.1016/j.tibs.2017.03.007

Zimmerly, S., and Semper, C. (2015). Evolution of group II introns. *Mob. DNA* 6, 7. doi:10.1186/s13100-015-0037-5

frontiers in Genetics

REVIEW
published: 15 September 2020
doi: 10.3389/fgene.2020.552444

Terminus-Associated Non-coding RNAs: Trash or Treasure?

Wen-Juan Ni, Fuhua Xie and Xiao-Min Leng**

School of Basic Medicine, Gannan Medical University, Ganzhou, China

3′ untranslated regions (3′ UTRs) of protein-coding genes are well known for their important roles in determining the fate of mRNAs in diverse processes, including trafficking, stabilization, translation, and RNA–protein interactions. However, non-coding RNAs (ncRNAs) scattered around 3′ termini of the protein-coding genes, here referred to as terminus-associated non-coding RNAs (TANRs), have not attracted wide attention in RNA research. Indeed, whether TANRs are transcriptional noise, degraded mRNA products, alternative 3′ UTRs, or functional molecules has remained unclear for a long time. As a new category of ncRNAs, TANRs are widespread, abundant, and conserved in diverse eukaryotes. The biogenesis of TANRs mainly follows the same promoter model, the RNA-dependent RNA polymerase activity-dependent model, or the independent promoter model. Functional studies of TANRs suggested that they are significantly involved in the versatile regulation of gene expression. For instance, at the transcriptional level, they can lead to transcriptional interference, induce the formation of gene loops, and participate in transcriptional termination. Furthermore, at the posttranscriptional level, they can act as microRNA sponges, and guide cleavage or modification of target RNAs. Here, we review current knowledge of the potential role of TANRs in the modulation of gene expression. In this review, we comprehensively summarize the current state of knowledge about TANRs, and discuss TANR nomenclature, relation to ncRNAs, cross-talk biogenesis pathways and potential functions. We further outline directions of future studies of TANRs, to promote investigations of this emerging and enigmatic category of RNA.

Keywords: 3′ termini, 3′ UTR, ncRNA, biogenesis, function

OPEN ACCESS

Edited by:
Nikolay Shirokikh, Australian National University, Australia

Reviewed by:
Vladimir P. Zhdanov, Boreskov Institute of Catalysis (RAS), Russia
Lucy Barrett, University of Western Australia, Australia
Takuya Imamura, Hiroshima University, Japan

****Correspondence:***
Fuhua Xie
fuhuax2004@163.com
Xiao-Min Leng
lengxm@mail2.sysu.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *16 April 2020*
Accepted: *25 August 2020*
Published: *15 September 2020*

Citation:
Ni W-J, Xie F and Leng X-M (2020) Terminus-Associated Non-coding RNAs: Trash or Treasure? Front. Genet. 11:552444. doi: 10.3389/fgene.2020.552444

INTRODUCTION

The encyclopedia of DNA elements (ENCODE) project aims to reveal functional elements of the human genome, thereby providing new insights into gene and genome functions (Ding et al., 2014; Moraes and Goes, 2016). For instance, RNA sequencing revealed that eukaryotic genomes are pervasively transcribed, using different regions to generate abundant and versatile non-coding RNAs (ncRNAs) (Kapranov et al., 2007; Forrest and Carninci, 2009; Jacquier, 2009; Clark et al., 2011; Jensen et al., 2013; Lu and Lin, 2019). Well-characterized ncRNAs, such as long non-coding RNAs (lncRNAs), small nucleolar RNAs (snoRNAs), and microRNAs (miRNAs), have been found to be variably produced. LncRNAs mainly derive from intergenic regions, introns, and antisense strands (Ayupe et al., 2015; Liu et al., 2015). SnoRNAs mainly arise from introns and

intergenic regions. The possible origins of miRNAs resemble those of snoRNAs. Consistently, similar percentages of intronic snoRNAs and intronic miRNAs have been reported in different eukaryotes (Mattick, 2003; Brown et al., 2008; Scott and Ono, 2011). Additionally, many new ncRNAs located at the 3′ and 5′ termini of genes have also been detected (Kapranov et al., 2007; Jacquier, 2009; Djebali et al., 2012; Ma et al., 2017; Laudadio et al., 2018). Owing to the absence of specific patterns in most 3′ end-associated ncRNAs and the limitations of the RNA sequencing technologies, these ncRNAs have usually been ignored for the past decade.

Investigation of the full landscape of 3′ untranslated regions (3′ UTRs) across species and cell types has contributed substantially to our understanding of their biogenesis and functions. Studies on the functions of 3′ UTRs focused primarily on their role in the regulation of gene expression, including mRNA trafficking, translational control, metabolism, and mRNA-protein structures (Wickens et al., 1997; Andreassi and Riccio, 2009; Denti et al., 2013; Jia et al., 2013; Pánek et al., 2016; Mayr, 2017). However, ncRNAs found around 3′ termini are usually not identified as biologically important. Indeed, the presence of terminus-associated small RNAs (TASRs) in both human and mouse genomes was firstly reported in 2007. These RNAs are usually scattered at both strands of protein-coding genes and do not exhibit unique lengths, specific base compositions, or typical secondary structures (Kapranov et al., 2007). Other small RNAs have also been detected at the 3′ ends of genes in both human and chicken genomes (Taft et al., 2009; Wei et al., 2011). Interestingly, these 3′ end-associated small RNAs are significantly different from the characteristic transcription initiation RNAs (tiRNAs) (Taft et al., 2009). In addition, transient transcriptome sequencing (TT-seq) has detected short-lived RNAs downstream of the polyadenylation [poly(A)] sites in human K562 cells. However, these RNAs are difficult to detect as they are usually cleaved from these sites, resulting in unprotected 5′ ends (Schwalb et al., 2016). Thus, these terminus-associated non-coding RNAs (TANRs) did not attract attention due to the lacking of unique length ranges and typical secondary structures.

From a technical perspective, transcriptome sequencing and microarrays show limitations for the discovery of TANRs. Indeed, transcriptome sequencing requires the construction of cDNA libraries and TANRs are often discarded during the rRNA removal step of this process or mixed with annotated transcript fragments afterward. In fact, as mixed fragments, they can partially or completely overlap with the annotated transcripts. Overlapping RNAs can be mapped as part of the annotated transcripts, alternative UTR regions, or even discarded. Furthermore, if some TANRs do not overlap with annotated transcripts, these would be filtered out as erroneous transcripts during bioinformatic analyses. In similarity, transcriptome microarrays are based on available information on annotated transcripts, usually excluding TANRs. Hence, TANRs have been mostly ignored in gene expression studies, given their lack of specific patterns, the uncertainty of their transcriptional origin, and other methodological difficulties (Jacquier, 2009; Yu et al., 2018).

Although TANRs usually are not identified as high-value targets, increasing evidence implied that they are important molecules for several cellular activities. For instance, the detection of diverse TANRs in eukaryotes suggested that they are widespread, abundant, and conserved. Moreover, studies of their biogenesis and functions pointed at TANRs as versatile molecules regulating gene expression. Since the biogenesis and functions of many TANRs are still unclear, and an increasing number of TANRs have been reported, elucidating their biological functions and mechanisms of action has become a new frontier in the field of RNA research.

DISCOVERY OF TANRs

Applications and breakthroughs of next-generation sequencing (NGS) and gene array in transcriptomics have revealed eukaryotic genomes can generate a multitude of diverse RNA species (Willingham et al., 2006; Kapranov et al., 2007; Jacquier, 2009). Owing to the presence of bidirectional promoters, one more lncRNA and many small ncRNAs have been found around the corresponding mRNA transcription start sites (TSSs) (Seila et al., 2008; Neil et al., 2009; Xu et al., 2009) (**Figure 1**). These RNAs can be generally termed as promoter-associated RNAs (PARs), including promoter-associated non-coding RNAs (pancRNAs) (Yamamoto et al., 2016; Uesaka et al., 2017), promoter upstream transcripts (PROMPTs) (Preker et al., 2011), upstream antisense RNAs (UaRNAs) (Flynn et al., 2011), stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs) (Neil et al., 2009; Xu et al., 2009), promoter-associated long RNAs (PALRs) (Kapranov et al., 2007), tiRNAs (Taft et al., 2009), and other PARs (Jiang et al., 2007). When the attention was shifted to the 3′ terminus, diverse ncRNAs were also discovered. These were divided into different subclasses: TASRs (Kapranov et al., 2007), antisense TASRs (aTASRs) (Kapranov et al., 2010), terminus-associated small nucleolar RNAs (TASNRs) (Leng et al., 2014), transcription termination site associated RNAs (TTSa-RNAs) (Valen et al., 2011; Laudadio et al., 2018), transcription boundary-associated RNAs (TBARs) (Yu et al., 2018), terminus-associated long RNAs (TALRs) (Yue et al., 2010), and 3′ UTR-associated RNAs (uaRNAs) (Mercer et al., 2011) (**Figure 2**). For a clearer distinction, we highlight that the abbreviation "UaRNAs" stands for "upstream antisense RNAs," while "uaRNAs" indicates "3′ UTR-associated RNAs." The different methods used for TANR identification together with the main characteristics of TANRs are summarized in **Table 1**.

Terminus-associated small RNAs were firstly reported to cluster at the 3′ termini of mRNAs (Kapranov et al., 2007). In mammals, there are about 200 TASR copies per cell (total numbers for all protein-coding genes), constituting approximately 3% of the small RNA library (Kapranov et al., 2007; Djebali et al., 2012). In *Arabidopsis thaliana*, TASR peaks were identified on 287 protein-coding genes, demonstrating that TASRs mainly accumulated in leaves and young seedlings (Ma et al., 2017). Altogether, the relevance of TASRs in both mammals and plants has been underestimated and the study of these small

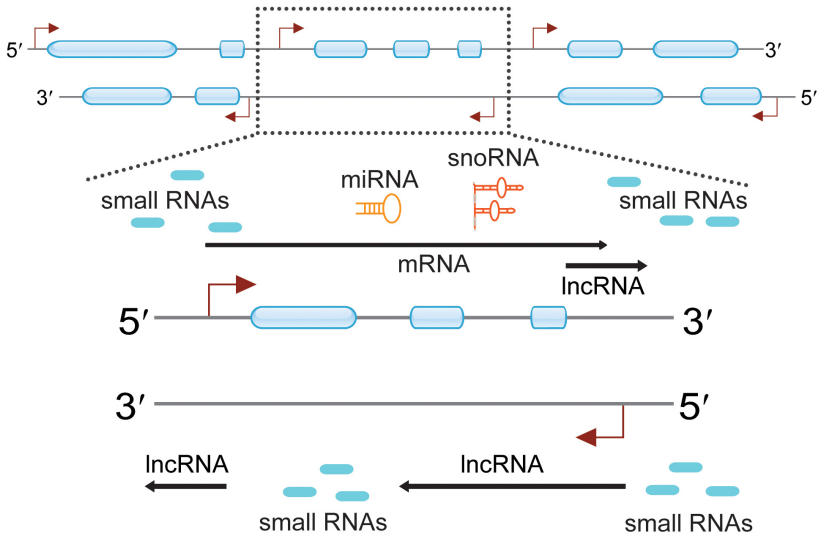


FIGURE 1 | Pervasive transcription across eukaryotic genomes can generate a multitude of diverse RNA species. Genomic regions are indicated with two thin lines marked with direction (5′ to 3′). Exons are presented as blue boxes, and transcription start sites (TTSs) are indicated with red angled arrows. Certain regions of the genome were highlighted and indicated with dotted frame lines. The zoomed in regions enlist many ncRNAs are observed around the protein-coding genes. These ncRNAs include small RNAs, lncRNAs, miRNAs, and snoRNAs.

RNAs did not receive priority in recent years for their lack of specific patterns. Considering their specific location within 3′ UTRs and the presence of poly(A) tails at their 3′ ends, TASRs have been regarded as degraded mRNA products or alternative 3′ UTRs in many studies.

Interestingly, a novel type of TASRs containing polyU tails at their 5′ end have been identified and renamed aTASRs, because they are antisense to 3′ UTRs (Kapranov et al., 2010). There are about 600 aTASR copies per cell, corresponding to 702 RefSeq-annotated protein-coding genes. Meanwhile, 1258 transcripts with non-genomically encoded 5′ poly(U) stretches closely associated with the 3′ termini of known RNAs can also be found in the UCSC Genome Browser database (Kapranov et al., 2010). Since aTASRs display a stretch of U residues at their 5′ ends but no poly(A) at their 3′ ends, they would be discarded in a conventional transcriptome analysis or library construction. Thus, direct RNA sequencing without prior conversion of RNA to cDNA would facilitate the discovery of novel ncRNAs (Furlan et al., 2020).

Argonaute (AGO) proteins are highly specialized binding small RNAs and can regulate gene expression at both transcriptional and posttranscriptional level by interacting with other proteins (Meister, 2013). By sequencing AGO1/2 immunoprecipitated libraries, several TTSa-RNAs were identified in *Homo sapiens*, particularly clustered close to the 3′ termination sites of mRNAs (Valen et al., 2011). Such TTSa-RNAs were found to be originated from 2822 protein-coding genes on average. Additionally, TTSa-RNAs are rich in G residues at their 5′ end and have a peculiar oligo(A) tail at their 3′ end (Laudadio et al., 2018). Compared to TASRs and aTASRs, TTSa-RNAs display shorter lengths (22 to 24 nt) and a specific cellular localization (enriched in nucleus). Beyond linear TASRs, aTASRs, and TTSa-RNAs, hairpin TASNRs (for some given genes) have been found in the yeast species related to *Schizosaccharomyces pombe* Lindner (Leng et al., 2014). In addition to small RNAs, lncRNAs, such as TALRs (for a given gene), uaRNAs (3′ UTR-associated RNAs) (about 1000 copies per cell on average in human), antisense CUTs, and SUTs (about 1000 copies per cell on average), have also been reported (Neil et al., 2009; Xu et al., 2009; Yue et al., 2010; Mercer et al., 2011). Given the evolutionary pressure toward the conservation of 3′ UTR regions, TANRs are usually conserved among different species.

Regarding the genomic location of these ncRNAs, TASRs, aTASRs, TTSa-RNAs, and uaRNAs are located within 3′ UTRs. In particular, TANRs and aTASRs start from poly(A) signal sites, while TTSa-RNAs end at the cleavage sites. Furthermore, TALRs and a small subset of antisense CUTs/SUTs usually overlap with 3′ UTRs. On the other hand, TASNRs are located downstream of 3′ UTRs. As indicated by their name, aTASRs and antisense CUTs/SUTs are located on the antisense strand, while other ncRNAs are located on the sense strand (**Figure 2**). According to their length (more or less than 200 nt), TALRs, uaRNAs, and antisense CUTs/SUTs are classified as lncRNAs, whereas others are considered small RNAs. Overall, TANRs vary considerably in their genomic location, strand, and length (**Table 1**).

BIOGENESIS OF TANRs

Studies of the biogenesis of ncRNAs are required to elucidate their functions and potential roles in the regulation of gene expression (Kim et al., 2009; Li et al., 2010). MiRNAs are currently the best-described small regulatory ncRNAs that follow a specific biogenesis pathway, requiring DROSHA/DGCR8, DICER1, and

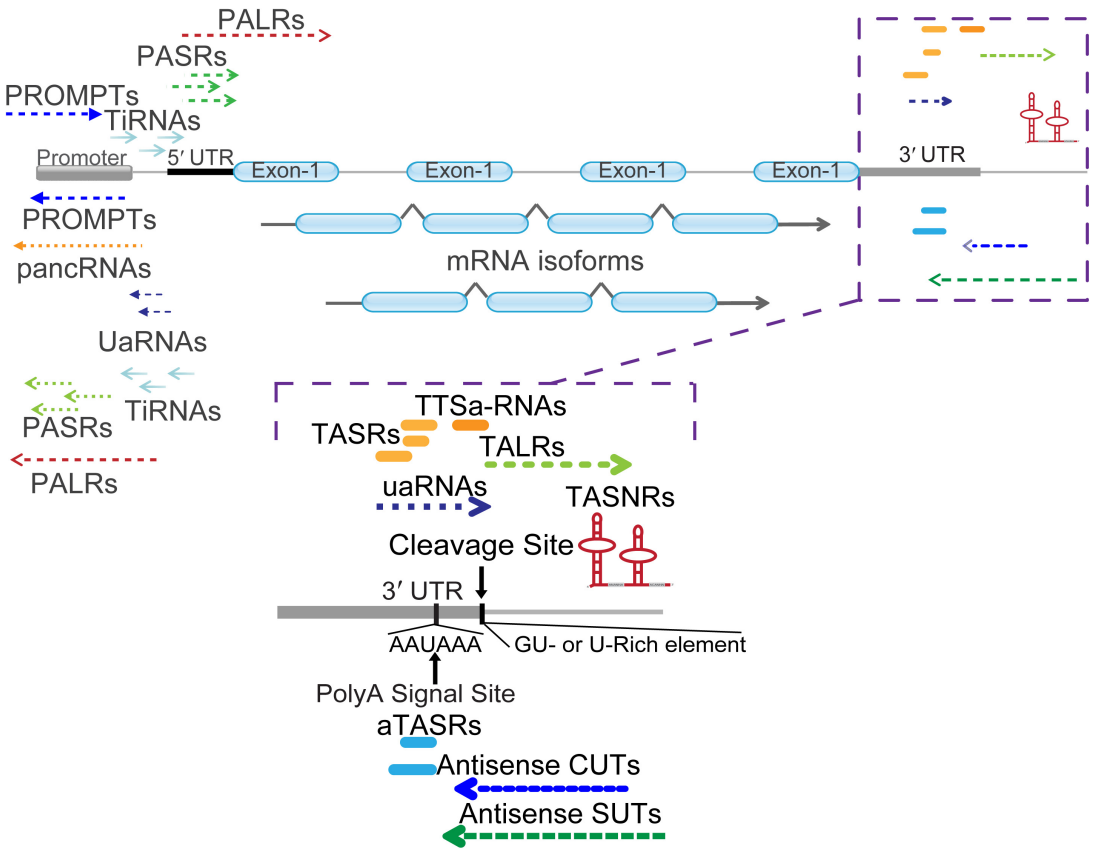


FIGURE 2 | Classification of mRNA 5′ and 3′ end associated ncRNAs. Different types of ncRNAs are indicated with detailed names. PALRs, promoter-associated long RNAs; PROMPTs, promoter upstream transcripts; UaRNAs, upstream antisense RNAs; pancRNAs, promoter-associated non-coding RNAs; PASRs, promoter-associated small RNAs; TiRNAs, transcription-initiation RNAs. The zoomed in regions enlist numerous ncRNAs located at the 3′ termini of mRNA. Cleavage site (AAUAA) and poly(A) signal site (GU- or U-rich element) were indicated with vertical arrow lines and letters. TASRs, terminus-associated small RNAs; aTASRs, antisense TASRs; TASNRs, terminus-associated small nucleolar RNAs; TTSa-RNAs, transcription termination site associated RNAs; TALRs, terminus-associated long RNAs; uaRNAs, 3′ untranslated region (UTR)-associated RNAs; antisense CUTs, antisense cryptic unstable transcripts; antisense SUTs, antisense stable unannotated transcripts.

AGO proteins (Daugaard and Hansen, 2017; Saeed et al., 2020). As TANRs are a novel class of ncRNAs, most but not all proteins associated to their biogenesis are unknown. According to their maturation process, the biogenesis of TANRs can generally occur by one of three models: the same promoter model, the RdRP activity-dependent model, and the independent promoter model.

The Same Promoter Model

In the same promoter model of TANR biogenesis, firstly, the transcription of TANR precursors is coupled to that of the upstream mRNAs using the same promoter. Then, maturation of TANRs occurs by posttranscriptional cleavage. Considering that the maturation processes of TASRs, TASNRs, TTSa-RNAs, TALRs, and uaRNAs share many characteristics, we summarize them altogether.

Terminus-associated small RNAs are located within the 3′ UTR of genes where no histone modifications marking active promoters or enrichment for RNA polymerase II (RNAPII) occupancy are found (Mercer et al., 2011). Hence, it is reasonable to infer that for their maturation TASRs undergo posttranscriptional cleavage. Studies on the biogenesis of TALRs and TTSa-RNAs also suggested that their maturation mainly depends on posttranscriptional cleavage from the corresponding mRNAs (Yue et al., 2010; Laudadio et al., 2018). However, this maturation process significantly differs from that of miRNAs. Firstly, evidence of the formation of secondary structures and of the corresponding passenger strands, characteristic of miRNA maturation, has not been found for these ncRNAs (Valen et al., 2011). Furthermore, genome-wide studies of TTSa-RNAs also determined that the regions flanking TTSa-RNAs do not tend to form hairpin structures more than randomly picked genomic regions (Laudadio et al., 2018). Secondly, altered expression of DICER and AGO2, required for miRNA biogenesis, had no effects on TTSa-RNA biogenesis (Valen et al., 2011; Laudadio et al., 2018). Importantly, defined sites within the polyA tail and approximately 75% of mRNA 3′ ends carry at least one TTSa-RNA read, suggesting that mRNA 3′ end processing is involved in their biogenesis (Valen et al., 2011). However, TTSa-RNAs are not by-products of mRNA degradation, since they display upstream poly(A) signals and are specifically loaded on AGO

TABLE 1 | List of TANRs described in this article.

TANRs	Technology	5′ Cap	3′ polyA	Species	Length	Strand	Structure	References
TASRs	Tiling array	No	Yes	*H. sapiens*, *M. musculus* and *A. thaliana*	22–200 nt	Sense	Linear	Kapranov et al. (2007); Djebali et al. (2012), Ma et al. (2017)
aTASRs	Helicos single-molecule sequencing	No	No	*H. sapiens* and *M. musculus*	<200 nt	Antisense	Linear	Kapranov et al. (2010)
TASNRs	Northern blot, RT-PCR, RACE	No	No	*Schizosaccharomyces group*	<200 nt	Sense	Hairpin	Leng et al. (2014)
TTSa-RNAs	RNA sequencing	No	No	*H. sapiens*	22–24 nt	Sense	Linear	Valen et al. (2011); Laudadio et al. (2018)
TALRs	RACE, RT-PCR	Yes	Yes	*H. sapiens*	>200 nt	Sense	Linear	Yue et al. (2010)
uaRNAs	CAGE, SAGE, Microarray	Yes	Yes	*H. sapiens*, *M. musculus* and *D. melanogaster*	>200 nt	Sense	Linear	Mercer et al. (2011)
Antisense CUTs/SUTs	Tiling array, RNA sequencing	Yes	Yes	*S. cerevisiae*	200–500 nt	Antisense	Linear	Neil et al. (2009); Xu et al. (2009)

proteins. Moreover, TTSa-RNAs tend to carry a G residue in the first position at the 5′ end and an oligo(A) tail (four or more As) at the 3′ end, supporting the hypothesis that TTSa-RNAs undergo posttranscriptional cleavage from the corresponding mRNAs (Laudadio et al., 2018).

Notably, detailed studies on the biogenesis of TASNRs and uaRNAs strongly indicated the same promoter model as the typical one for the biogenesis of most TANRs. In particular, two TASNR precursors (*rpl26-snR49* and *rpl29-snR93*) highly overlapped with upstream mRNAs; no promoters were detected between mature TASNRs and their precursors; and promoter deletion analysis confirmed that the precursor of TASNR *snR49* and the corresponding upstream mRNA used the same promoter for the regulation of their transcription. Thus, TASNRs undergo processing from precursors during maturation (Leng et al., 2014). As for uaRNAs, no active promoters or enrichment for RNAPII occupancy have been found within the 3′ UTR; however, exon-intron junctions have been detected (Mercer et al., 2011). Moreover, a detailed study on the biogenesis of the uaRNA *FLJ11812* in human cells confirmed that the maturation of this ncRNA depends on posttranscriptional cleavage, and that the TIA1 protein is responsible for this process (Ge et al., 2014).

Although TASNRs and uaRNAs exploit the same promoters of their respective upstream protein-coding genes for transcription, their precursors originate differently. Indeed, uaRNAs may derive from their corresponding mRNAs through maturation by cleavage similarly to TTSa-RNAs. Conversely, TASNR precursors are different transcripts from their corresponding mRNAs, although highly overlapping. As for TASRs and TALRs, it is still unknown whether they are cleaved from their corresponding mRNAs.

The RdRP Activity-Dependent Model

The RNA-dependent RNA polymerase (RdRP) plays a key role in RNA silencing in fungi, plants, and worms by generating double-stranded RNAs (dsRNAs) from RNA templates (Duempelmann et al., 2020). In the RdRP activity-dependent model of TANR biogenesis, RdRP can *de novo* synthesize antisense TANRs at the 3′ termini of mRNAs by using the sense mRNAs as templates. For instance, it has been reported that the human telomerase reverse transcriptase (TERT) RdRP can perform *de novo* synthesis of short interfering RNAs (siRNAs) that are complementary to template RNAs (Maida et al., 2016). Thus, *de novo* RNA synthesis by RdRP suggests the existence of a novel RNA copying mechanism. Recent studies strongly indicated that the biogenesis of aTASRs depends on RdRP. Indeed, aTASRs contain non-genomically encoded poly(U) stretches at their 5′ ends that are complementary to the 3′ poly(A) tails of mRNAs (Kapranov et al., 2010). These double-stranded and complementary RNAs have been detected in both human cells and plants (Kapranov et al., 2010; Ma et al., 2017). In *A. thaliana*, aTASR fragments were preferentially incorporated into AGO4 and aTASR accumulation was significantly decreased in *rdr2* (RNA-dependent RNA polymerase 2), *nrpd1a* (RNA polymerase IVa), and *nrpd1b* (RNA polymerase IVb) mutants. Thus, RdRPs and RNA polymerase IV are responsible for the biogenesis of some aTASRs, even though the detailed mechanisms remain unknown (Ma et al., 2017). However, the endogenous biochemical pathway that mediates copying of aTASRs in human cells still requires further investigation (Kapranov et al., 2010).

The Independent Promoter Model

In the independent promoter model, TANRs on the antisense strand have their own promoters. As independent transcripts, their biogenesis is usually regulated by their upstream promoter regions. Although TANRs include only a small number of antisense CUTs/SUTs, several studies indicated that independent promoters are primarily responsible for their biogenesis in *Saccharomyces cerevisiae*. This conclusion derived from the fact that the transcriptional initiation sites of antisense CUTs or SUTs are located in nucleosome-free regions (NFRs), corresponding to promoter regions. Thus, independent transcription is the main biogenesis mechanism of antisense CUTs or SUTs (Neil et al., 2009; Xu et al., 2009).

FUNCTIONS OF TANRs

The existence of different pathways of TANR biogenesis suggests that they are important for some cellular activities. NcRNAs typically function by forming various ribonucleoproteins (RNPs) together with several proteins. Well-known functional RNP

particles include snoRNA ribonucleoproteins (snoRNPs) and miRNA-AGO ribonucleoproteins (miRNPs). These RNPs contain the respective RNAs and a small set of associated proteins (Bachellerie et al., 2002; Bartel, 2004). Within miRNPs, miRNAs usually cause degradation and translational repression of target mRNAs through the formation of miRNA-mRNA duplexes. However, miRNA–mRNA interactions are dynamically regulated by different physiological or pathological conditions (Ni and Leng, 2015). As a group of widely studied functional proteins, AGO proteins associate with a diverse variety of ncRNAs, thereby providing functional and regulatory support for ncRNA-mediated modulation of gene expression (Joshua-Tor and Hannon, 2011; Daugaard and Hansen, 2017). It was reported that TANRs enriched in different subcellular compartments (cytoplasm and nucleus) can interact with different AGO proteins in eukaryotes. Hence, TANRs may regulate gene expression at both transcriptional and posttranscriptional levels (**Figure 3**).

Transcriptional Regulation by TANRs

Members of eukaryotic AGO protein family are key players of gene expression (Meister, 2013). Interestingly, a previous study showed that synthetic small RNAs fully complementary to a TALR located beyond the 3′ terminus of progesterone receptor (*PR*) mRNA could modulate *PR* transcription (Younger and Corey, 2011). This provides new insights into the function of TANRs with high nuclear localization. Firstly, the TALR is loaded onto AGO2 upon addition of exogenous miRNA mimics. Then, the complex formed of miRNA mimics, TALR, and AGO2 is recruited to the promoter region of an upstream gene. Finally, a gene loop juxtaposing the promoter and terminator is formed, resulting in altered regulation of transcription (**Figure 3**) (Yue et al., 2010). Notably, the formation of gene loops is thought to mediate long-distance transcriptional regulation in different eukaryotes (Bratkowski et al., 2018). However, functional studies of AGO1 and AGO2-associated TTSa-RNAs strongly argued against their specific recruitment on chromatin given their nucleoplasm/chromatin abundance, although Gene Ontology (GO) analysis suggested that genes giving rise to TTSa-RNAs are significantly enriched in the regulation of cell cycle progression and DNA integrity checkpoints (Laudadio et al., 2018). On the other hand, evidence of transcriptional stalling via RNAPII backtracking triggering nucleolytic degradation of the nascent RNA indicates that TTSa-RNAs may be implied in

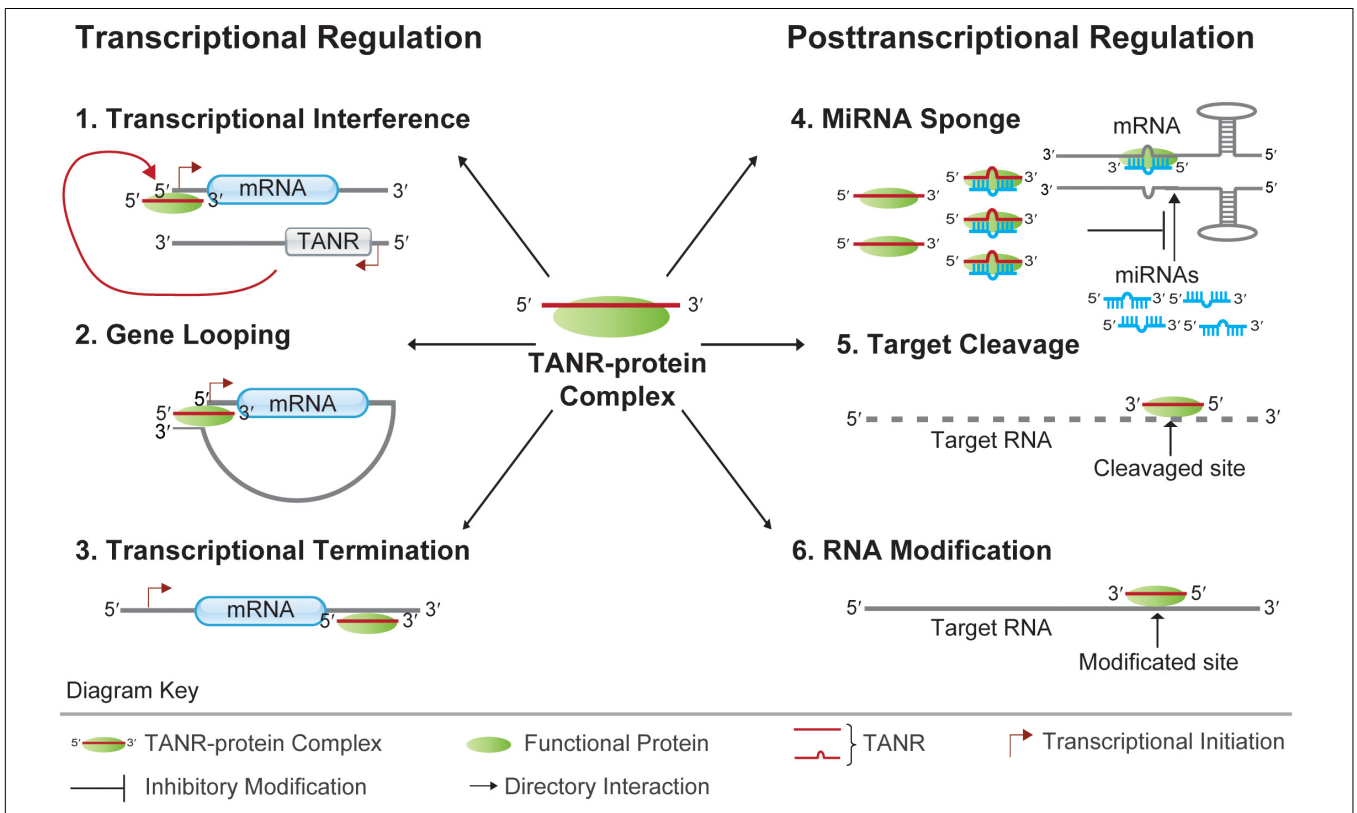


FIGURE 3 | Functions of TANRs. TANRs function at the transcriptional level: 1. Transcriptional Interference: TNAR together with its associated functional protein regulates the transcription of its sense protein-coding gene via binding to the sense promoter region; 2. Gene Looping: TNAR together with its associated functional protein regulates the transcription of its upstream gene by juxtaposing the promoter and terminator together; 3. Transcriptional Termination: TNAR together with its associated functional protein regulates the transcription termination of its upstream gene by binding to the protein-coding gene 3′ UTR. TANRs function at the posttranscriptional level: 4. MiRNA Sponge: TNAR together with its associated functional protein regulates mRNA translation by sponging miRNA from its target mRNA; 5. Target Cleavage: TNAR together with its associated functional protein can direct cleavage of target mRNA at specific site. 6. RNA Modification: TNAR together with its associated functional protein can guide RNA modification at specific site.

the termination of mRNA transcription (Valen et al., 2011). For instance, a recent study in *A. thaliana* indicated that promoter-proximal RNAPII stalling can regulate plant gene transcription (Thomas et al., 2020). Thus, it is reasonable to infer that mammalian TTSa-RNAs might participate in the regulation of gene transcription through the modulation of transcriptional termination (**Figure 3**).

In plants, AGO1 represses target RNAs in the cytoplasm, while AGO4 usually directs *de novo* DNA methylation in the nucleus (Baulcombe, 2004; Vaucheret, 2008; Carbonell, 2017). Site-specific DNA methylation signals were observed on several genomic loci corresponding to the peaks of many TASRs associated with AGO4 in *A. thaliana* (Ma et al., 2017). Furthermore, some aTASRs are preferentially incorporated into AGO4. Thus, a subset of the TASRs and aTASRs reported in *A. thaliana* may be involved in site-specific DNA methylation (Ma et al., 2017). However, it is not clear if TANR-mediated gene looping is required to guide DNA methylation.

In *S. cerevisiae*, antisense CUTs/SUTs usually couple the transcriptional regulation of neighboring genes. As overlapping and divergent transcripts, they may act as local regulatory signals for transcriptional interference (**Figure 3**) (Neil et al., 2009; Xu et al., 2009). In addition, transcriptional interference mediated by *cis*-acting antisense CUTs/SUTs involves several chromatin modifiers (such as Set2p, Set1p, Rcoi1p, and Eaf3p) (Nevers et al., 2018). In a recent related report, the transcription of approximately 20% of *S. cerevisiae* genes was found to be repressed by antisense ncRNAs via a chromatin-based transcription interference mechanism. Hence, using near-base-pair-resolution techniques in antisense CUTs/SUTs-inducible strains would reveal the relationship between antisense transcription and repression of sense gene expression, nucleosome occupancy, and transcription-associated histone modifications (Gill et al., 2020).

Posttranscriptional Regulation by TANRs

Previous reports suggested that 3′ UTRs can function *in trans* to regulate cell proliferation and differentiation in the absence of corresponding protein-coding transcripts (Rastinejad and Blau, 1993; Amack et al., 1999; Jenny et al., 2006). For example, expression of oskar 3′ UTR in *Drosophila* could rescue the egg-less defect of oskar null-mutants in the absence of the Oskar protein. Indeed, the oskar 3′ UTR functions as a scaffold for trafficking and accumulation of Staufen during oogenesis (Jenny et al., 2006). Moreover, in *A. thaliana*, the overexpressed ncRNA *IPS1* can act as a competing endogenous RNA (ceRNA) that positively regulates the expression of *PHO2 by* sequestering *miR-399* from its target site (Franco-Zorrilla et al., 2007). Also, in human embryonic stem cells (hESCs) and human umbilical vein endothelial cells (HUVECs), the uaRNA *FLJ11812* derived from the 3′ UTR of *TGFB2* can be targeted by *miR-4459*. Conversely, uaRNA *FLJ11812* can upregulate the levels of the proteins CDC20B and ATG13, whose coding genes can also be targeted by *miR-4459*. Thus, this uaRNA acts as a ceRNA by sponging *miR-4459* from its target mRNAs (Lu et al., 2015). Therefore, uaRNAs can act as decoys to sponge miRNAs from their target mRNAs (**Figure 3**). Alternatively, they may act as scaffolds to form regulatory RNA-protein complexes that are functional even in the absence of their corresponding proteins (Mercer et al., 2011).

Structural and functional analyses of ncRNAs in fission yeast suggested that some TANRs act as guide snoRNAs. By forming specific snoRNPs, these snoRNAs can direct methylation or pseudouridylation of target RNAs. Notably, most of such site-specific modifications can affect cell growth *in vivo*. For example, TANR *snR49* was predicted to mediate pseudouridylation of 18S rRNA at the U121 and U305 sites. Upon deletion of TANR *snR49*, the corresponding modifications on rRNA disappeared with consequent delay of cell growth. Furthermore, posttranscriptional modifications of target RNAs by TASNRs are conserved in yeasts (Leng et al., 2014). Thus, TASNRs can act as guide RNAs for targeted RNA modifications (**Figure 3**).

Since the production of human aTASRs is positively correlated with that of their associated mRNAs, functional studies of aTASRs were based on the corresponding transcripts. These transcripts corresponded to functionally annotated proteins and were further analyzed. Functional enrichment analysis suggested that they are related to translation. Indeed, the GO categories of "structural constituent of ribosome," "translation," and "RNA binding" were all significantly overrepresented. Due to the bias of enrichment analysis toward highly synthesized transcripts, all human genes were used as background for a second estimation of enrichment. Nevertheless, similar results were obtained, with the GO biological function category "translation" scoring as the top hit (Kapranov et al., 2010). In *A. thaliana*, aTASRs associated with cytoplasmic AGO1 are proposed to mediate target RNA cleavage (**Figure 3**) (Ma et al., 2017). Perhaps, synthetic aTASR mimics would help to reveal their mechanism of translational regulation.

PERSPECTIVES AND DISCUSSION

Terminus-associated non-coding RNAs were identified years ago, however, their definition is somewhat confused for researchers. Regarding the nomenclature, the abbreviation "UaRNAs" has been used to indicate "upstream antisense RNAs" and sometimes "3′ UTR-associated RNAs." In terms of timing, upstream antisense RNAs (UaRNAs) were reported before 3′ UTR-associated RNAs (uaRNAs) (Chiu et al., 2006). uaRNAs were then defined according to their specific genomic location within mRNA 3′ UTRs (Mercer et al., 2011). However, UaRNAs were also later discovered and studied (Flynn et al., 2011). Therefore, the abbreviation "uaRNAs" has been given different meanings in separate studies, possibly causing confusion (Chiu et al., 2006; Flynn et al., 2011; Mercer et al., 2011; Lu et al., 2015; Ogami et al., 2017; Yu et al., 2018). To some degree, TTSa-RNAs and TASNRs broadly belong to the same class of TASRs. Indeed, when TASRs were first and systemically described, no identifiable patterns, such as genomic locations, lengths, and subcellular localizations, were unraveled (Kapranov et al., 2007). In contrast, TTSa-RNAs enriched in small RNA libraries of AGO1/2 immunoprecipitates are located before the cleavage sites of mRNAs with restricted lengths (approximately 23 nt) and exhibit nuclear localization (Valen et al., 2011;

Laudadio et al., 2018). Moreover, unlike other TANRs, TASNRs are a well-known group of snoRNAs (Leng et al., 2014). Regarding CUTs and SUTs, although no clear partition between CUTs and SUTs exists, some ncRNAs defined as CUTs have been redefined as SUTs (Jacquier, 2009; Neil et al., 2009; Xu et al., 2009). Recently, a uniform annotation system for transcript boundaries has been proposed. This annotation is based on their genomic positions and sequence lengths, and provides suggestions for additional classifications of TANRs, for instance according to their biogenesis pathways, modes of action, and biological outputs (Yu et al., 2018). However, as more and diverse TANRs are found in other eukaryotes, a new, more elaborate nomenclature for TANR classification should be proposed, including detailed information on their genomic location, originating strand, biogenesis pathway, and functions.

The discovery of novel transcripts around annotated transcripts also challenges the concept of gene (Gerstein et al., 2007; Gingeras, 2007). Indeed, not only mRNAs but also lncRNAs can generate functional TANRs. A well-known example is *MALAT1*-associated small cytoplasmic RNA (mascRNA), generated from the nascent lncRNA metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*). MascRNA is located at the 3′ end of mature *MALAT1*, and its maturation is dependent on RNase P (Wilusz et al., 2008). Functional studies of mascRNA found that this ncRNA is involved in cardiovascular innate immunity (Gast et al., 2016). Surprisingly, mascRNA could function as a translational enhancer when placed downstream of *cGFP in vivo* (Wilusz et al., 2012). Studies of the function and biogenesis of mascRNA suggested that TANRs originating from lncRNAs also play an important role in regulating gene expression. Furthermore, several studies have found that ends of both some mRNAs and certain lncRNAs contained conserved secondary structures that might generate TANRs (Kertesz et al., 2010; Nguyen et al., 2016; Yu et al., 2017). Hence, the possible presence of TANRs should not be ignored in either protein-coding or non-protein-coding loci.

Although TANRs can derive from different pathways, their biogenesis might involve the cross-talk of several regulatory mechanisms. For instance, transcription and posttranscriptional processing are important steps of the maturation of TANRs. Furthermore, the carboxy-terminal domain (CTD) of RNAPII is important for coupling mRNA transcription and processing (McCracken et al., 1997; Proudfoot et al., 2002; Ahn et al., 2004; Bentley, 2005). Indeed, by interacting with splicing and 3′ cleavage factors, RNAPII couples transcription, splicing, and cleavage of mRNA precursors (Ahn et al., 2004). Meanwhile, terminal sites are associated with pausing of RNA polymerase (Schwalb et al., 2016). Thus, whether TANRs maturate during a coupled process of transcription and posttranscriptional cleavage or they are derived from RNAPII backtracking remains unknown. Thus, new methods for detecting nascent RNAs or the use of mutants in mRNA 3′ end maturation pathways may shed some light on TANR biogenesis (Wissink et al., 2019; Furlan et al., 2020).

Given the heterogeneity of TANRs, unraveling their functions has become one of the most basic and pressing issues. 3′ UTRs usually harbor critical elements for gene expression, such as miRNA response elements (MREs). Therefore, TANRs that contain MREs may act as miRNA sponges, thus protecting the corresponding mRNAs from translational repression or degradation. For instance, uaRNA *FLJ11812* functions as a ceRNA by sponging *miR-4459* from its target mRNAs, thereby providing a novel direction for functional studies (Ge et al., 2014; Lu et al., 2015). Furthermore, the formation of gene loops juxtaposing the promoter and terminator has been reported in several organisms, and gene looping is thought to mediate long-distance transcriptional regulation (Bratkowski et al., 2018). However, it is unclear whether TANR-mediated gene looping is required for guiding DNA methylation, mRNA processing, or other processes. Nevertheless, the occurrence of miRNA sponging and gene looping provides novel directions for functional studies of sense TANRs. As for antisense ncRNAs, the discovery of chromatin-based transcription interference also suggested a new mechanism of TANR function (Gill et al., 2020).

Once a TANR is discovered, it is challenging to know how to study its function. Basic information, such as the abundance of related mRNAs, the secondary structure, and the subcellular localization of TANRs, aids in understanding their possible functions. For detailed functional studies, induced upregulation and downregulation of TANRs represent an appropriate strategy for primary functional studies. To achieve upregulation, overexpression or synthesis of certain TANRs represents available methods. However, for most TANRs overlapping with certain 3′ UTRs that harbor important regulatory elements, some technical issues need to be overcome to eliminate the potential impact of induced downregulation on the upstream transcripts. Currently, siRNA screens and the application of CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 to delete certain DNA regions provide useful tools for functional annotation of TANRs in a native context (Zhao et al., 2017).

In summary, the discovery of TANRs in different eukaryotes suggested that they are abundant and conserved. Moreover, studies of the biogenesis and functions of TANRs indicated that they can play important roles in different cellular activities. However, since TANRs represent a novel group of ncRNAs, their biogenesis and functions still require further research. As more information about different TANRs is being reported, their involvement in the regulation of gene expression is due to be unfolded in full and presents one more intriguing observation of the versatility of RNA function.

AUTHOR CONTRIBUTIONS

FX and X-ML contributed to the conception of the study. W-JN and X-ML wrote the manuscript. W-JN, FX, and X-ML discussed and improved the revised manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (81760506).

REFERENCES

Ahn, S. H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* 13, 67–76. doi: 10.1016/s1097-2765(03)00492-1

Amack, J. D., Paguio, A. P., and Mahadevan, M. S. (1999). Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum. Mol. Genet.* 8, 1975–1984. doi: 10.1093/hmg/8.11.1975

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

Ayupe, A. C., Tahira, A. C., Camargo, L., Beckedorff, F. C., Verjovski-Almeida, S., and Reis, E. M. (2015). Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol.* 12, 877–892. doi: 10.1080/15476286.2015.1062960

Bachellerie, J. P., Cavaille, J., and Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* 84, 775–790. doi: 10.1016/s0300-9084(02)01402-5

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.

Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431, 356–363.

Bentley, D. L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* 17, 251–256. doi: 10.1016/j.ceb.2005.04.006

Bratkowski, M., Unarta, I. C., Zhu, L., Shubbar, M., Huang, X., and Liu, X. (2018). Structural dissection of an interaction between transcription initiation and termination factors implicated in promoter-terminator cross-talk. *J. Biol. Chem.* 293, 1651–1665. doi: 10.1074/jbc.m117.811521

Brown, J. W., Marshall, D. F., and Echeverria, M. (2008). Intronic noncoding RNAs and splicing. *Trends Plant Sci.* 13, 335–342. doi: 10.1016/j.tplants.2008.04.010

Carbonell, A. (2017). Plant ARGONAUTEs: features, functions, and unknowns. *Methods Mol. Biol.* 1640, 1–21. doi: 10.1007/978-1-4939-7165-7_1

Chiu, A., Wu, X., Subtelny, A., and Sharp, P. (2006). Characterizing polyadenylated uaRNAs suggests a potential role for Pabpn1. *Environ. Prot. Chem. Ind.* 29:S1. doi: 10.1176/appi.pn.2014.3a2

Clark, M. B., Amaral, P. P., Schlesinger, F. J., Dinger, M. E., Taft, R. J., Rinn, J. L., et al. (2011). The reality of pervasive transcription. *PLoS Biol.* 9:e1000625. doi: 10.1371/journal.pbio.1000625

Daugaard, I., and Hansen, T. B. (2017). Biogenesis and function of ago-associated RNAs. *Trends Genet.* 33, 208–219. doi: 10.1016/j.tig.2017.01.003

Denti, M. A., Viero, G., Provenzani, A., Quattrone, A., and Macchi, P. (2013). mRNA fate: life and death of the mRNA in the cytoplasm. *RNA Biol.* 10, 360–366. doi: 10.4161/rna.23770

Ding, N., Qu, H., and Fang, X. (2014). [The ENCODE project and functional genomics studies]. *Yi Chuan* 36, 237–247.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108.

Duempelmann, L., Skribbe, M., and Buhler, M. (2020). Small RNAs in the transgenerational inheritance of epigenetic information. *Trends Genet.* 36, 203–214. doi: 10.1016/j.tig.2019.12.001

Flynn, R. A., Almada, A. E., Zamudio, J. R., and Sharp, P. A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10460–10465. doi: 10.1073/pnas.1106630108

Forrest, A. R., and Carninci, P. (2009). Whole genome transcriptome analysis. *RNA Biol.* 6, 107–112. doi: 10.4161/rna.6.2.7931

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079

Furlan, M., Tanaka, I., Leonardi, T., De Pretis, S., and Pelizzola, M. (2020). Direct RNA sequencing for the study of synthesis, processing, and degradation of modified transcripts. *Front. Genet.* 11:394. doi: 10.3389/fgene.2020.00394

Gast, M., Schroen, B., Voigt, A., Haas, J., Kuehl, U., Lassner, D., et al. (2016). Long noncoding RNA MALAT1-derived mascRNA is involved in cardiovascular innate immunity. *J. Mol. Cell Biol.* 8, 178–181. doi: 10.1093/jmcb/mjw003

Ge, D., Han, L., Huang, S., Peng, N., Wang, P., Jiang, Z., et al. (2014). Identification of a novel MTOR activator and discovery of a competing endogenous RNA regulating autophagy in vascular endothelial cells. *Autophagy* 10, 957–971. doi: 10.4161/auto.28363

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681. doi: 10.1101/gr.6339607

Gill, J. K., Maffioletti, A., Garcia-Molinero, V., Stutz, F., and Soudet, J. (2020). Fine chromatin-driven mechanism of transcription interference by antisense noncoding transcription. *Cell Rep.* 31:107612. doi: 10.1016/j.celrep.2020.107612

Gingeras, T. R. (2007). Origin of phenotypes: genes and transcripts. *Genome Res.* 17, 682–690. doi: 10.1101/gr.6525007

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* 10, 833–844. doi: 10.1038/nrg2683

Jenny, A., Hachet, O., Závorszky, P., Cyrklaff, A., and Weston, M. D. (2006). A translation-independent role of oskar RNA in early *Drosophila* oogenesis. *Development* 133, 2827–2833. doi: 10.1242/dev.02456

Jensen, T. H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. *Mol. Cell* 52, 473–484. doi: 10.1016/j.molcel.2013.10.032

Jia, J., Yao, P., Arif, A., and Fox, P. L. (2013). Regulation and dysregulation of 3'UTR-mediated translational control. *Curr. Opin. Genet. Dev.* 23, 29–34. doi: 10.1016/j.gde.2012.12.004

Jiang, H., Kim, D., and Morris, K. V. (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12422–12427. doi: 10.1073/pnas.0701635104

Joshua-Tor, L., and Hannon, G. J. (2011). Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb. Perspect. Biol.* 3:a003772. doi: 10.1101/cshperspect.a003772

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341

Kapranov, P., Ozsolak, F., Kim, S. W., Foissac, S., Lipson, D., Hart, C., et al. (2010). New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. *Nature* 466, 642–646. doi: 10.1038/nature09190

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., et al. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107. doi: 10.1038/nature09322

Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139. doi: 10.1038/nrm2632

Laudadio, I., Formichetti, S., Gioiosa, S., Klironomos, F., Rajewsky, N., Macino, G., et al. (2018). Characterization of transcription termination-associated RNAs: new insights into their biogenesis, tailing, and expression in primary tumors. *Int. J. Genomics* 2018:1243858.

Leng, X. M., Diao, L. T., Li, B., Bi, Y. Z., Chen, C. J., Zhou, H., et al. (2014). The ribosomal protein rpl26 promoter is required for its 3' sense terminus ncRNA transcription in *Schizosaccharomyces pombe*, implicating a new transcriptional mechanism for ncRNAs. *Biochem. Biophys. Res. Commun.* 444, 86–91. doi: 10.1016/j.bbrc.2014.01.018

Li, Y., Liu, X., Huang, L., Guo, H., and Wang, X. J. (2010). Potential coexistence of both bacterial and eukaryotic small RNA biogenesis and functional related protein homologs in Archaea. *J. Genet. Genomics* 37, 493–503. doi: 10.1016/s1673-8527(09)60069-2

Liu, J., Wang, H., and Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328. doi: 10.1111/pbi.12336

Lu, W., Han, L., Su, L., Zhao, J., Zhang, Y., Zhang, S., et al. (2015). A 3'UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. *Stem Cells Dev.* 24, 1133–1140. doi: 10.1089/scd.2014.0353

Lu, Z., and Lin, Z. (2019). Pervasive and dynamic transcription initiation in *Saccharomyces cerevisiae*. *Genome Res.* 29, 1198–1210. doi: 10.1101/gr.245456.118

Ma, X., Han, N., Shao, C., and Meng, Y. (2017). Transcriptome-wide discovery of PASRs (promoter-associated small RNAs) and TASRs (terminus-associated small RNAs) in *Arabidopsis thaliana*. *PLoS One* 12:e0169212. doi: 10.1371/journal.pone.0169212

Maida, Y., Yasukawa, M., and Masutomi, K. (2016). De novo RNA synthesis by RNA-dependent RNA polymerase activity of telomerase reverse transcriptase. *Mol. Cell. Biol.* 36, 1248–1259. doi: 10.1128/mcb.01021-15

Mattick, J. S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25, 930–939. doi: 10.1002/bies.10332

Mayr, C. (2017). Regulation by 3'-untranslated regions. *Annu. Rev. Genet.* 51, 171–194. doi: 10.1146/annurev-genet-120116-024704

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361. doi: 10.1038/385357a0

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* 14, 447–459. doi: 10.1038/nrg3462

Mercer, T. R., Wilhelm, D., Dinger, M. E., Solda, G., Korbie, D. J., Glazov, E. A., et al. (2011). Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res.* 39, 2393–2403. doi: 10.1093/nar/gkq1158

Moraes, F., and Goes, A. (2016). A decade of human genome project conclusion: scientific diffusion about our genome knowledge. *Biochem. Mol. Biol. Educ.* 44, 215–223. doi: 10.1002/bmb.20952

Neil, H., Malabat, C., D'aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042. doi: 10.1038/nature07747

Nevers, A., Doyen, A., Malabat, C., Neron, B., Kergrohen, T., Jacquier, A., et al. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Res.* 46, 6009–6025. doi: 10.1093/nar/gky342

Nguyen, T. C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F. H., et al. (2016). Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.* 7:12023.

Ni, W.-J., and Leng, X.-M. (2015). Dynamic miRNA-mRNA paradigms: new faces of miRNAs. *Biochem. Biophys. Rep.* 4, 337–341. doi: 10.1016/j.bbrep.2015.10.011

Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J. J., et al. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* 31, 1257–1271. doi: 10.1101/gad.302604.117

Pánek, J., Koláø, M., Herrmannová, A., and Valášek, L. S. (2016). A systematic computational analysis of the rRNA-3' UTR sequence complementarity suggests a regulatory mechanism influencing post-termination events in metazoan translation. *RNA* 22, 957–967. doi: 10.1261/rna.056119.116

Preker, P., Almvig, K., Christensen, M. S., Valen, E., Mapendano, C. K., Sandelin, A., et al. (2011). PROMoter uPstream transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res.* 39, 7179–7193. doi: 10.1093/nar/gkr370

Proudfoot, N. J., Furger, A., and Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* 108, 501–512. doi: 10.1016/s0092-8674(02)00617-7

Rastinejad, F., and Blau, H. M. (1993). Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell* 72, 903–917. doi: 10.1016/0092-8674(93)90579-f

Saeed, S., Zahra, V. A., Sharif, M., Ali, A., Javad, D. S., and Sadegh, A. J. (2020). Small regulatory noncoding RNAs in *Drosophila melanogaster*: biogenesis and biological functions. *Brief. Funct. Genomics* 19, 309–323. doi: 10.1093/bfgp/elaa005

Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., et al. (2016). TT-seq maps the human transient transcriptome. *Science* 352, 1225–1228. doi: 10.1126/science.aad9841

Scott, M. S., and Ono, M. (2011). From snoRNA to miRNA: dual function regulatory non-coding RNAs. *Biochimie* 93, 1987–1992. doi: 10.1016/j.biochi.2011.05.026

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., et al. (2008). Divergent transcription from active promoters. *Science* 322, 1849–1851. doi: 10.1126/science.1162253

Taft, R. J., Glazov, E. A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G. J., et al. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* 41, 572–578.

Thomas, Q. A., Ard, R., Liu, J., Li, B., Wang, J., Pelechano, V., et al. (2020). Transcript isoform sequencing reveals widespread promoter-proximal transcriptional termination in *Arabidopsis*. *Nat. Commun.* 11:2589.

Uesaka, M., Agata, K., Oishi, T., Nakashima, K., and Imamura, T. (2017). Evolutionary acquisition of promoter-associated non-coding RNA (pancRNA) repertoires diversifies species-dependent gene activation mechanisms in mammals. *BMC Genomics* 18:285. doi: 10.1186/s12864-017-3662-1

Valen, E., Preker, P., Andersen, P. R., Zhao, X., Chen, Y., Ender, C., et al. (2011). Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat. Struct. Mol. Biol.* 18, 1075–1082. doi: 10.1038/nsmb.2091

Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci.* 13, 350–358. doi: 10.1016/j.tplants.2008.04.007

Wei, W., Pelechano, V., Jarvelin, A. I., and Steinmetz, L. M. (2011). Functional consequences of bidirectional promoters. *Trends Genet.* 27, 267–276. doi: 10.1016/j.tig.2011.04.002

Wickens, M., Anderson, P., and Jackson, R. J. (1997). Life and death in the cytoplasm: messages from the 3' end. *Curr. Opin. Genet. Dev.* 7, 220–232. doi: 10.1016/s0959-437x(97)80132-3

Willingham, A. T., Dike, S., Cheng, J., Manak, J. R., Bell, I., Cheung, E., et al. (2006). Transcriptional landscape of the human and fly genomes: nonlinear and multifunctional modular model of transcriptomes. *Cold Spring Harb. Symp. Quant. Biol.* 71, 101–110. doi: 10.1101/sqb.2006.71.068

Wilusz, J. E., Freier, S. M., and Spector, D. L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135, 919–932. doi: 10.1016/j.cell.2008.10.012

Wilusz, J. E., Jnbaptiste, C. K., Lu, L. Y., Kuhn, C. D., Joshua-Tor, L., and Sharp, P. A. (2012). A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* 26, 2392–2407. doi: 10.1101/gad.204438.112

Wissink, E. M., Vihervaara, A., Tippens, N. D., and Lis, J. T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.* 20, 705–723. doi: 10.1038/s41576-019-0159-6

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037. doi: 10.1038/nature07728

Yamamoto, N., Agata, K., Nakashima, K., and Imamura, T. (2016). Bidirectional promoters link cAMP signaling with irreversible differentiation through promoter-associated non-coding RNA (pancRNA) expression in PC12 cells. *Nucleic Acids Res.* 44, 5105–5122. doi: 10.1093/nar/gkw113

Younger, S. T., and Corey, D. R. (2011). Transcriptional regulation by miRNA mimics that target sequences downstream of gene termini. *Mol. Biosyst.* 7, 2383–2388. doi: 10.1039/c1mb05090g

Yu, D., Ma, X., Zuo, Z., Shao, W., Wang, H., and Meng, Y. (2017). Bioinformatics resources for deciphering the biogenesis and action pathways of plant small RNAs. *Rice* 10:38.

Yu, D., Ma, X., Zuo, Z., Wang, H., and Meng, Y. (2018). Classification of transcription boundary-associated RNAs (TBARs) in animals and plants. *Front. Genet.* 9:168. doi: 10.3389/fgene.2018.00168

Yue, X., Schwartz, J. C., Chu, Y., Younger, S. T., Gagnon, K. T., Elbashir, S., et al. (2010). Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nat. Chem. Biol.* 6, 621–629. doi: 10.1038/nchembio.400

Zhao, W., Siegel, D., Biton, A., Tonqueze, O. L., Zaitlen, N., Ahituv, N., et al. (2017). CRISPR-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Res.* 45, 10800–10810. doi: 10.1093/nar/gkx675

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fgene.2021.680725

The RNA-Binding Proteins SRP14 and HMGB3 Control HIV-1 Tat mRNA Processing and Translation During HIV-1 Latency

*Georges Khoury[1†], Michelle Y. Lee[1†], Sri H. Ramarathinam[2], James McMahon[3], Anthony W. Purcell[2], Secondo Sonza[1], Sharon R. Lewin[3,4,5] and Damian F. J. Purcell[1]**

[1] Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia, [2] Infection and Immunity Program, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia, [3] Victorian Infectious Diseases Service, The Royal Melbourne Hospital at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia, [4] Department of Infectious Diseases, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia, [5] Department of Infectious Diseases, Alfred Hospital and Monash University, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Jane E. A. Reid,
Australian National University, Australia

Reviewed by:
Fatah Kashanchi,
George Mason University, United States
Elton J. R. Vasconcelos,
University of Leeds, United Kingdom

***Correspondence:**
Damian F. J. Purcell
dfjp@unimelb.edu.au

†Present address:
Georges Khoury and Michelle Y. Lee,
Division of Microbiology and Immunology, Yerkes National Primate Research Center, Emory University, Atlanta, GA, United States

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 15 March 2021
Accepted: 17 May 2021
Published: 14 June 2021

Citation:
Khoury G, Lee MY, Ramarathinam SH, McMahon J, Purcell AW, Sonza S, Lewin SR and Purcell DFJ (2021) The RNA-Binding Proteins SRP14 and HMGB3 Control HIV-1 Tat mRNA Processing and Translation During HIV-1 Latency. Front. Genet. 12:680725. doi: 10.3389/fgene.2021.680725

HIV-1 Tat protein is essential for virus production. RNA-binding proteins that facilitate Tat production may be absent or downregulated in resting CD4$^+$ T-cells, the main reservoir of latent HIV in people with HIV (PWH) on antiretroviral therapy (ART). In this study, we examined the role of Tat RNA-binding proteins on the expression of Tat and control of latent and productive infection. Affinity purification coupled with mass spectrometry analysis was used to detect binding partners of MS2-tagged *tat* mRNA in a T cell-line model of HIV latency. The effect of knockdown and overexpression of the proteins of interest on Tat transactivation and translation was assessed by luciferase-based reporter assays and infections with a dual color HIV reporter virus. Out of the 243 interactions identified, knockdown of SRP14 (Signal Recognition Particle 14) negatively affected *tat* mRNA processing and translation as well as Tat-mediated transactivation, which led to an increase in latent infection. On the other hand, knockdown of HMGB3 (High Mobility Group Box 3) resulted in an increase in Tat transactivation and translation as well as an increase in productive infection. Footprinting experiments revealed that SRP14 and HMGB3 proteins bind to TIM-TAM, a conserved RNA sequence-structure in *tat* mRNA that functions as a Tat IRES modulator of *tat* mRNA. Overexpression of SRP14 in resting CD4$^+$ T-cells from patients on ART was sufficient to reverse HIV-1 latency and induce virus production. The role of SRP14 and HMGB3 proteins in controlling HIV Tat expression during latency will be further assessed as potential drug targets.

Keywords: HIV-1, latency, *tat* mRNA, SRP14, HMGB3, mRNA processing, translation

INTRODUCTION

The persistence of a reservoir of resting CD4$^+$ T-cells harboring silent provirus is the major impediment toward attaining a cure for HIV-1 (Deeks et al., 2016). Despite two decades of intensive investigation, the mechanisms contributing toward establishment and maintenance of latent infection *in vivo* remain incompletely understood (Khoury et al., 2018a). To date, blocks

to the initiation of transcription have been the most widely studied and targeted for reversal of latent infection in clinical trials (Ait-Ammar et al., 2020). In these trials, reactivation of virus in the form of cell-associated HIV-1 RNA is detected but decay of the reservoir is almost never induced (Zerbato et al., 2019). Successful production of HIV-1 protein would be required to prime the immune response after reactivation of virus, and results of these trials suggest that additional blocks are present in the latently infected cell that prevent complete processing of RNA or translation to occur. The production of multiply spliced (MS) mRNA is required for successful production of HIV-1 virions, and these transcripts are not always detected after treatment with latency-reversing agents (Pasternak and Berkhout, 2018). We recently found that production of MS RNA is a better indication of virus reactivation *ex vivo* (Zerbato et al., 2021). In addition, a recent study showed that latent infection in blood $CD4^+$ T cells from HIV-1 infected individuals on suppressive therapy was due to blocks in the post-initiation stages of transcription, including elongation, RNA-capping and splicing (Yukl et al., 2018).

The HIV-1 regulatory protein, Tat, is essential for successful transcription of the HIV-1 genome and virus production in natural infection (Ott et al., 2011). Whilst transcription initiation is a Tat-independent process, Tat is required for efficient elongation of transcription as the association of the negative factors DSIF (DRB Sensitivity-Inducing Factor) and NELF (Negative ELongation Factor) with the RNA polymerase II causes promoter-proximal pausing (Ping and Rana, 2001). Tat liberates its co-factor, the positive transcription elongation factor-b (P-TEFb), from its sequestration in the inactive 7SK snRNP complex (Krueger et al., 2010; Muniz et al., 2010), to associate with and phosphorylate DSIF and the carboxy-terminus of the stalled RNA polymerase II resulting in release of the transcriptional complex (Ping and Rana, 2001). Notably, Tat is also involved in the other post-transcriptional processes of RNA polyadenylation and splicing (Chiu et al., 2001, 2002; Schapira et al., 2003; Jablonski et al., 2010) where additional blocks to HIV-1 transcription in latent infection have been suggested. There are multiple pieces of evidence supporting the importance of Tat in diverting the integrated provirus away from latent infection. Fluctuations in the levels of Tat protein are a strong indicator of whether a cell will enter latency (Weinberger et al., 2005; Razooky et al., 2015). Nuclear retention of the multiply spliced RNAs that encode Tat or the presence of low levels of Tat protein contribute to maintaining the cells in a latent state (Lassen et al., 2006). Exogenous Tat delivered into latently infected cells inhibited proviral entry into latency, whilst established latency can be reversed by Tat (Donahue et al., 2012). Over time on suppressive therapy, mutations detrimental to Tat function accumulate, contributing to the persistence of latent provirus (Yukl et al., 2009).

We recently characterized the presence of an RNA regulatory element underlying Tat-encoding sequence, which we termed TIM-TAM (Tat IRES modulator of *tat* mRNA) (Khoury et al., 2020). We showed that TIM-TAM is involved in regulating latent infection, as viruses carrying a silent mutation disrupting the secondary structure of TIM-TAM resulted in a restriction in establishment of latency in primary $CD4^+$ T-cells. Furthermore, reactivation of latent HIV-1 from the CCL19 primary cell model (Saleh et al., 2011; Spina et al., 2013) infected with virus carrying mutated TIM-TAM was affected after treatment with PMA/PHA (Phorbol Myristate Acetate/PHytohAemagglutinin). The TIM-TAM was also shown to exhibit the properties of an internal ribosome entry site (IRES), RNA structures that facilitate translation initiation in a cap-independent manner.

Proteins required for processing and translation of *tat* mRNAs may be differentially expressed in cells carrying latent provirus and in cells undergoing productive infection. The recent study by Moron-Lopez et al. (2020) showed that human splice factors were differentially expressed between unstimulated and activated cells from ART-suppressed individuals. Activation of the splice acceptor 3 (SA3) site in the HIV-1 genome is required for production of *tat* mRNA, the use of which is tightly regulated by splicing silencers and enhancers (Saliou et al., 2009). In addition to translation initiation factors, auxiliary proteins unique to a particular IRES element may affect IRES function positively or negatively (Lozano and Martínez-Salas, 2015). Several proteins, including HNRNPA1, DDX3, hRIP and HuR (Monette et al., 2009; Rivas-Aravena et al., 2009; Liu et al., 2011) have been reported to impact the function of the HIV-1 5′UTR IRES element.

In this study, we used affinity purification coupled mass spectrometry analysis to identify cellular RNA-binding proteins that interact with *tat* mRNA during productive and latent infection. Using a principal component analysis-based scoring system, a short-list of thirteen proteins were chosen for follow-up investigation. Knockdown (KD) and overexpression assays were used to investigate the roles of these proteins in latent and productive infection of HIV-1 with a dual-fluorescent reporter virus. The effect of KD of these proteins on the various stages of Tat expression—Tat mRNA splicing, Tat translation and Tat transactivation—were explored using luminescence-based reporter systems. The affinity purification approach and downstream validation assay systems allowed identification of Tat-RNA binding proteins that are differentially expressed in resting and activated $CD4^+$ T-cells isolated from people living with HIV on ART and are potential druggable targets for modulation of HIV-1 latency.

MATERIALS AND METHODS

RNP Purification by MS2 Selection Affinity

HIV 5′UTRtat1-Tat (NL4-3, nt 455–743; 5777-6044; 8369–8414) and Tat (nt 5,830–6,044; 8369–8414) fragments were amplified and cloned into NheI and EcoRI restriction sites of pcDNA3.1(-)::MS2 plasmid, upstream of 3× MS2 stem-loops (BamHI-KpnI). RNAs were generated by run-off transcription with T7 MEGAscript kit (Promega) using KpnI linearized plasmids. DNA templates were digested with RQ1 RNase-Free DNase (Promega) and RNA were recovered by lithium chloride precipitation then dissolved in MilliQ water. J-Lat cells [clone 6.3, from NIH AIDS Reagent Program, (Jordan et al., 2003)] were expanded in RPMI supplemented with 10% FBS to 10^8 cells per sample, left untreated or stimulated with TNF-α (20 ng/mL) for 24 h. Activation of virus

expression was validated by flow cytometry (GFP^+ expression) and $p24^{CA}$-ELISA as previously described (Khoury et al., 2020). Cells were then collected and lyzed with ice-cold lysis buffer (50 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl, 0.1% IgePal) supplemented with protease inhibitors (Roche) for 30 min at 4°C. Protein cell extracts were dialyzed before use against buffer D 1× [Hepes KOH pH 7.9 20 mM, KCl 100 mM, glycerol 20%, EDTA 0.2 mM, DTT 0.5 mM) + MgCl2 3 mM + protease inhibitor tablet (Roche)] for 2 h at 4°C, followed by centrifugation for 10 min at 1,700 × g at 4°C. Five hundred pmol of MS2-tagged RNAs into 100 µl buffer D 1× were denatured by 10 min heating at 65°C, followed by slow cooling at room temperature with addition of 7.75 µl of 62.5 mM $MgCl_2$ to a final concentration of 4.5 mM $MgCl_2$. After 10 min incubation at room temperature, RNAs were incubated with a fivefold molar excess of purified MBP-MS2 fusion protein at 4°C for 20 min. The RNA-MS2:MS2-MBP complexes formed were incubated with amylose beads (200 µl, GE Healthcare) equilibrated in buffer D for 2 h at 4°C. After three washes with 500 µl of buffer D, 1 mg of protein extract supplemented with 5 µM of yeast tRNAs (Sigma-Aldrich) was added. After 20 min of incubation at 4°C with constant agitation, three successive washes were performed in Buffer D and RNP complexes eluted twice by incubation for 20 min at room temperature with 200 µl of Buffer D containing 10 mM maltose. Half of the eluted RNP complexes were processed in solution for mass spectrometry analyses. For western-blot analysis, 10% of the eluted material was used. Experiments were repeated three independent times using different batches of RNA and protein lysate.

Mass Spectrometry Analysis

TCEP (10 mM, Thermo Fisher Scientific) was added to the eluted sample to reduce the cysteine bonds in proteins, and heated for 20 min at 60°C followed by alkylation of cysteines using 25 mM Iodoacetamide (Sigma-Aldrich) for 20 min at RT in the dark. Samples were subjected to proteolytic cleavage by addition of 1 µg of trypsin protease from porcine pancreas per 100 µg of protein and incubated overnight at 37°C under agitation. Reactions were stopped by addition of formic acid and the resultant peptides were concentrated using C18-packed tips (BondElut, Agilent/Varian). The peptides were subject to online trapping using a PepMap100 trap column at 15 µL/min followed by separation on PepMap 100 C18 nanocolumn (50 cm × 75µm) using a 30-min gradient of Buffer B (80% ACN 0.1% Formic acid) over Buffer A (0.1% Formic acid). The online separated peptides were analyzed using a Q-Exactive plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The survey scans were acquired at 70,000 resolution from 375 to 1,800 m/z, the ion accumulation target was set to 3e6 with maximum injection time of 120 ms. A total of 12 most intense ions (with charge more than 2) were sequentially isolated and fragmented by higher-energy collisional dissociation (HCD) set to 27%, at a resolution of 17,500, target of 1e5 ions and maximum injection time of 120 ms. To identify protein groups, data acquired was converted into mgf and searched against Swissprot human database (version 2016_12) using *ProteinPilot software* (v4.0, SCIEX) with the following search parameters: Iodoacetamide alkylation, trypsin enzyme digestion, instrument-specific settings for TripleTOF 5600 + (MS tolerance 0.05 Da, MS/MS tolerance 0.1 Da, charge state + 2– + 5), biological modification probabilistic features on, thorough ID algorithm, and detected protein threshold 0.05. Mass Spectrometry Interaction Statistics (MiST) (Jäger et al., 2011) analysis was conducted to sort proteins based on specificity, reproducibility and abundance over the three replicates. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD025782.

Generation of Knockdown RFP^+ Jurkat Cell Lines

Three or four Sherwood UltramiR shRNA viral particles (10^6–10^7 TU/mL) against each of the 15 protein targets of interest were obtained from TransOmic Technologies. The pZIP (SFFV) shRNA-mir lentivectors constitutively express the short hairpin RNA (shRNA-mir), puromycin selection marker and red fluorescent protein (RFP) driven by the Spleen Focus Forming Virus (SFFV) promoter. Jurkat cells (2.1×10^5) were transduced with 100 µL of pooled pseudoviruses for each protein target separately at a MOI of 2.5. Spinoculation was conducted at 1,200 × g for 2 h at 23°C in a flat-bottom 96-well plate. Cells were incubated at 37°C for 72 h then sorted on RFP^+ expression on an Astrios cell sorter (Beckman Coulter). Bulk populations were then kept in culture under 0.7 µg/mL puromycin selection. Single clones expressing high levels of RFP^+ cells were selected from the bulk population on the BD FACSAria III. Bulk and single clones were maintained under constant puromycin selection then surviving clones were expanded. Cells were collected for RT-qPCR analysis as well as immunoblotting.

Cloning of Tat Expression Reporter Constructs

For the Tat/GH1 constructs, HIV-1 sequences were derived from pNL4-3 (M. Martin, NIH, Bethesda, MD, United States), GH1 sequences were amplified from pØhGH (Nichols Institute Diagnostics), LucF from pGL4.13[luc2/SV40] (Promega), and LucR from pGL4.73[hRluc/SV40] (Promega). These segments were assembled by SOE-PCR and cloned into the pcDNA3.1- (Invitrogen) backbone via use of the Xho1 and XbaI sites. The NT5C3 splicing constructs were cloned into the pcDNA3.1- backbone through cleavage of NhoI and NheI where the NT5C3 cellular sequences were amplified from HeLa cell genomic DNA and the other components derived as for the Tat/GH1 constructs.

DNA Transfection of KD RFP + Cells and Luciferase Analysis

Bulk populations or single clones of the KD RFP^+ Jurkat cell lines (6×10^4 cells) were plated into a round-bottom 96-well plate. Media was completely removed and cells were transfected with 1.6 µg reporter constructs using DMRIE-C reagent (1:2.5, Invitrogen) in 62.5 µL Opti-MEM (Gibco). For transactivation assays, 200 ng of Tat WT and 1 µg of Tat hGH reporter construct with 300 ng of both LTR-lucFirefly and lucRenilla. For translation

assays, 300 ng of Tat-Cap or IRES LucF reporter constructs were used with 300 ng of lucRenilla for normalization. For splicing assays, 300 ng of Tat^+ or Tat- reporter constructs with 300 ng of LTR-lucFirefly. Three hours post-transfection, DNA:DMRIE-C mix was removed and media replenished. Twenty-four hours post-transfection, cells were collected and luciferase activity was measured following lysis in 40 μL of 1X passive lysis buffer using a FLUOstar plate reader with the dual-luciferase reporter assay (Promega).

Virus Production and Transduction

R7GEmTB dual color reporter virus was produced by replacing mCherry into R7GEmC [obtained from NIH AIDS Reagent Program, (Calvanese et al., 2013)] with mtagBFP2 fluorescent protein. SRP14 and PTB cDNA were cloned into pInducer10 lentivector (Meerbrey et al., 2011) by replacing tRFP-shRNA cassette with protein ORF-T2A-mtagBFP2. Viral stocks were generated by transfecting the proviral constructs into HEK 293T cells with Lipofectamine 2000 (Invitrogen) in serum free media (Opti-MEM, Gibco). Supernatants were collected after 72 h, clarified by centrifugation and 0.45 μm filtration to clear cell debris. Particles were concentrated using microcon centrifugal filter device (30K, Merck Millipore) or pelleted by overlaying supernatant on a cushion of 20% (w/v) sucrose in TNE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) in Ultra Clear Thinwall tubes and centrifugation at 24,200 rpm for 2 h at 4°C (Beckman SW28/SW41Ti rotor). Viral particle pellets were resuspended into Opti-MEM and stored at −80°C. Virus titres were quantified by measuring $p24^{CA}$ levels by capture ELISA as previously described (Khoury et al., 2020) and titration into TZM-bl cells (Sarzotti-Kelsoe et al., 2014). KD RFP^+ Jurkat cells and $CD4^+$ T-cells (10^6 cells) were infected with 40 and 100 ng R7GEmTB virus (+Env 92HT593.1), respectively. Cells were spinoculated at 1,200 × g for 2 h at 23°C in a 96 well flat-bottom plate and incubated at 37°C for 3 days. Cells were collected, washed and resuspended in 100 μL of 1X PBS then loaded onto cover slips that had been pre-coated with poly-L-Lysine 0.01% (Sigma-Aldrich). Cells were incubated for 1 h on the coverslip then rinsed with PBS 1X and fixed with 2% PFA. Cells were treated with glycine 0.2 M for 10 min at room temperature then rinsed with PBS 1X and H_2O. Coverslips were inverted on a microscope slide containing ProLong Gold Antifade Mountant (Thermo Fisher Scientific). Images were acquired on LSM710 confocal microscope using *Zeiss Zen software*. J-Lat 10.6, 8.4 and A2 (obtained from the NIH AIDS reagent program) were infected via spinoculation at 1,500 × g for 2 h at 23°C followed by 1 h incubation at 37°C before replenishing the culture with fresh media without or with 5 μg/mL doxycycline. Two days post-infection, cells were washed and stained with Near-IR Live/Dead fixable dead cell staining (Invitrogen). Finally, cells were washed, fixed with 1% formaldehyde and acquired using a Fortessa flow cytometry instrument (BD Bioscience). Analysis was performed using *FlowJo Software, version 10.4.2*.

Ethics

The studies involving the use of blood samples from HIV negative donors were reviewed and approved by the Human Research and Ethics Committees from the University of Melbourne (15-09VIC-03 and 17-08VIC-01). All HIV-1 seronegative donors were recruited by the Red Cross Blood Bank (Melbourne, Australia) and provided written informed consent for the use of their blood products for the research. The use of blood samples from people living with HIV was approved by the Alfred Hospital (HREC214/15) for the study entitled Large volume peripheral blood mononuclear cells (PBMCs) collection by leukapheresis to define HIV persistence in HIV-infected adults. All participants provided informed consent and the protocol was approved by the local Institutional Review Board.

Participant Details

PBMCs from people living with HIV on ART with a viral load < 20 copies/mL for ≥3 years were collected by leukapheresis (Alfred Hospital, Melbourne, Australia; **Supplementary Table 1**) and stored in liquid nitrogen. Resting $CD4^+$ T-cells (purity > 95%) were isolated from PBMCs by negative selection using $CD4^+$ T-cell isolation kit (Miltenyi Biotec) supplemented with anti-CD69 (clone L78, BD) and anti-HLA-DR (clone 2-O6). $CD4^+$ T-cells were activated with anti-CD3/CD28 (coated α-CD3 clone OKT3, BD, 1 μg/mL and soluble α-CD28 clone L293, BD, 0.5 μg/mL) for 48 h at 37°C.

Western-Blot

Resting and activated $CD4^+$ T-cells isolated from patients under ART (30×10^6 cells) were lysed for 30 min on ice using ice-cold IGEPAL cell lysis buffer (Tris-HCl pH 8.8 50 mM, NaCl 150 mM, IGEPAL 1%, EDTA 1 mM) supplemented with protease inhibitor cocktail (Complete^{TM}, Mini, EDTA-free Protease Inhibitor Cocktail, Roche). Cell lysate was cleared and quantified using the Bio-Rad Protein Assay Reagent (Bio-Rad). Equal amounts of each sample (20 μg) were loaded on 15% SDS-PAGE, transferred to nitrocellulose membrane. Blots were probed with anti-SRP14 (B-3, sc-377012, Santa Cruz) at 1:10, anti-HMGB3 (clone 546519, R&D) 1:1,000, anti-PTB (clone 7, #325000, Invitrogen) at 1:500 and anti-GAPDH (14C10, Cell Signaling) at 1:1,000. Antibodies were detected using either goat anti-rabbit (Invitrogen, #656120) or goat anti-mouse IgG (H + L) HRP (Invitrogen, #626520) at 1/5,000 and developed with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientifc). Images were visualized using an *MF-ChemiBis 3.2 imaging system (DNR)*.

Quantitative Real-Time PCR Analysis

RNA from resting and activated $CD4^+$ T-cells isolated from patients under ART (15×10^6 cells) and bulk and individual clone of interest for the KD RFP + Jurkat cells as well as untransduced Jurkats were extracted using TRIzol following the manufacturer's protocol. RNA (500 ng) was DNase-treated with RQ1 DNase (Promega) and reverse transcribed using 4 U OmniScript (Qiagen) with dNTP (0.5 mM), random hexamers (2 μM), oligo$(dT)_{15}$ (1 μM), and RNasin (10 U) in 30 μL reactions for 1 h at 37°C. Real-time PCR was performed on CFX Connect real-time PCR detection system (Bio-Rad) using Fast SYBR Green Master Mix (Applied Biosystems), forward/reverse primers (**Supplementary Table 2**) at 500 nM final concentration,

2 μL of undiluted cDNA in reactions with a final volume of 20 μL. Each sample was assayed in duplicate and normalized on GAPDH content and untransduced Jurkat cells for KD RFP + Jurkats or 18S content and activated $CD4^+$ T-cells for primary $CD4^+$ T-cells. Cycling conditions used were 95°C for 10 min for activation of the DNA polymerase followed by 45 cycles of denaturation at 95°C for 3 s and annealing/extension at 60°C for 20 s. Melt curve analysis was performed each time with a starting temperature of 60°C, increasing to 95°C with 0.5°C increments every 0.05 s. Data were analyzed on *CFX Manager 3.0* software employing $2^{-\Delta\Delta Ct}$ to determine fold changes.

T-Cell Electroporation

Electroporation of resting $CD4^+$ T-cells was performed using an Amaxa human T-cell Nucleofector kit (Lonza). Purified resting $CD4^+$ T-cells (5.3×10^6) were re-suspended in 100 μl nucleofector solution (4.5:1 ratio of human T-cell nucleofector solution:supplement) and transfected with 1.2 μg DNA per 10^6 cell using Amaxa nucleofector program U-014 for high viability. Cells were then transferred into a 12-well plate containing pre-equilibrated media, followed by half-media change 6 h post-nucleofection. Cells were maintained for 48 h at 37°C in media supplemented with 1% FBS and 1 U/mL IL-2, with or without 5 μg/mL doxycycline. After 2 days, cells were collected and stained for CD25 (PE-Cy7, clone 2A3, BD) and HLA-DR (V605, clone L243 BioLegend) activation markers in comparison to PHA stimulated (10 μg/mL) cells. Culture supernatant was also harvested and viral RNA was isolated using QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's protocol. vRNA was DNase treated (2 U/μg RNA, Promega) and reversed transcribed using Omniscript as described above. HIV RNA levels were assessed using droplet digital PCR (ddPCR) using *pol* primers/probe (**Supplementary Table 2**). Thermal cycling was conducted as follows: 95°C for 10 min, 40 cycles of 94°C for 30 s and 60°C for 60 s, followed by 98°C for 10 min (ramp rate 2°C/s for each step) on a C1000 Touch Thermal cycler (Bio-Rad). The droplets were subsequently read on a QX200 droplet reader (Bio-Rad) and the data were analyzed with *Quanta-Soft 1.7.4 software*. The limit of detection of our assay was of 58 copies/mL.

Recombinant Proteins

Production and purification of MBP-MS2 protein was done as previously described (Deckert et al., 2006). cDNA encoding SRP14, HMGB3 and PTB were cloned into NheI and HindIII restriction sites of pET28a + bacterial vector (Novagen) allowing the expression of N-terminally 6×-His tagged proteins. All recombinant proteins were expressed in E. coli BL21 (DE3) Codon + bacteria grown in LB/Kanamycin and induced by addition of 0.1 mM IPTG overnight at 37°C (for PTB and SRP14) or 30°C (for HMGB3). Cells were sonicated in lysis buffer containing 20 mM potassium phosphate pH 7.4 (pH to 6 for HMGB3), 1M KCl, 1 mM DTT, 20 mM imidazole and protease inhibitor cocktail (Roche), then treated with 1.5 mg/mL lysozyme (from chicken egg white, Sigma-Aldrich) for 20 min on ice. Soluble proteins were purified by polyethyleneimine (0.4% final) followed by ammonium sulfate precipitation (30% for SRP14 and PTB; 40% then 60% precipitation for HMGB3). Proteins were purified by affinity purification using Ni-NTA affinity resin (Qiagen) coupled to an Äkta pure FPLC (GE Healthcare). The bound proteins were eluted with a linear imidazole gradient (20 mM to 500 mM). Proteins were buffer exchanged into buffer D (20 mM HEPES–KOH pH 7.9, 100 mM KCl, 0.2 mM EDTA, 3 mM $MgCl_2$, 0.5 mM DTT, 20% glycerol) using PD-10 desalting columns (GE Healthcare) followed by concentration using 3, 10, and 30 k cutoff amicon columns (Merck) for SRP14, HMGB3 and PTB, respectively. Proteins concentration was determined using UV spectroscopy at 280 nm. Purity was confirmed by 12.5% SDS-PAGE and western-blot analysis using protein specific antibodies and anti-His HRP conjugated antibody (clone J099B12, BioLegend) at 1:500.

Footprinting Assays

Footprinting analysis was performed by SHAPE as previously described (Mortimer and Weeks, 2007). Briefly, *tat2* RNA was generated by run-off transcription using Sp6 MEGAscript kit (Promega) and pSP65::tat2 construct linearized with ClaI. *In vitro* transcribed *tat2* RNA (1 pmol) was probed in 3× folding buffer (333 mM HEPES–KOH pH 8, 333 mM NaCl, 33.3 mM $MgCl_2$) in the presence or absence of SRP14, HMGB3 or PTB recombinant protein at 5, 10 and 20 protein/RNA molar ratio (0.41–1.6 μM). RNP complexes were formed by incubation for 20 min at room temperature then probed with 1-methyl-7-nitroisatoic anhydride (1M7, 65 mM in DMSO) or DMSO for 4 min at 37°C, then recovered by ethanol precipitation. Primer extension was conducted as described previously with 0.4 μM fluorescently labeled odp3102 primer (6-FAM or HEX, Sigma-Aldrich, **Supplementary Table 2**). The dideoxy sequencing reactions were generated using unmodified RNA, labeled primers (PET or NED, Applied Biosystems) and 0.5 mM ddGTP. cDNAs were recovered by ethanol precipitation and separated by capillary electrophoresis with LIZ500 size standard (ABI 3130, AGRF). Data was processed using the *QuShape software* (Karabiber et al., 2013). Protections induced by SRP14, HMGB3 or PTB binding were indicated by a reduction in the normalized SHAPE reactivities.

RESULTS

Detection of 243 Putative Tat-RNA and Cellular Protein Interactions

To explore MS RNA binding partners during latency, we initiated a proteomic approach based on affinity chromatography purification of RNA-protein complexes (Maenner et al., 2010; Bar et al., 2011) formed upon incubation of *in vitro* transcribed *tat* RNA with protein lysate, followed by protein identification by mass spectrometry. An overview of the processes is shown in **Figure 1A**. Protein lysates were prepared from the HIV-1 latently infected T cell line, J-Lat6.3 where the cells were either left untreated (latent infection) or activated with TNF-α (productive infection). The RNA of interest, 5′UTRtat1-Tat and Tat with three binding sites for the MS2 coat protein fused at their 3′-ends were used as baits. In parallel, a control RNA that only contained the three MS2 binding sites was used. RNA

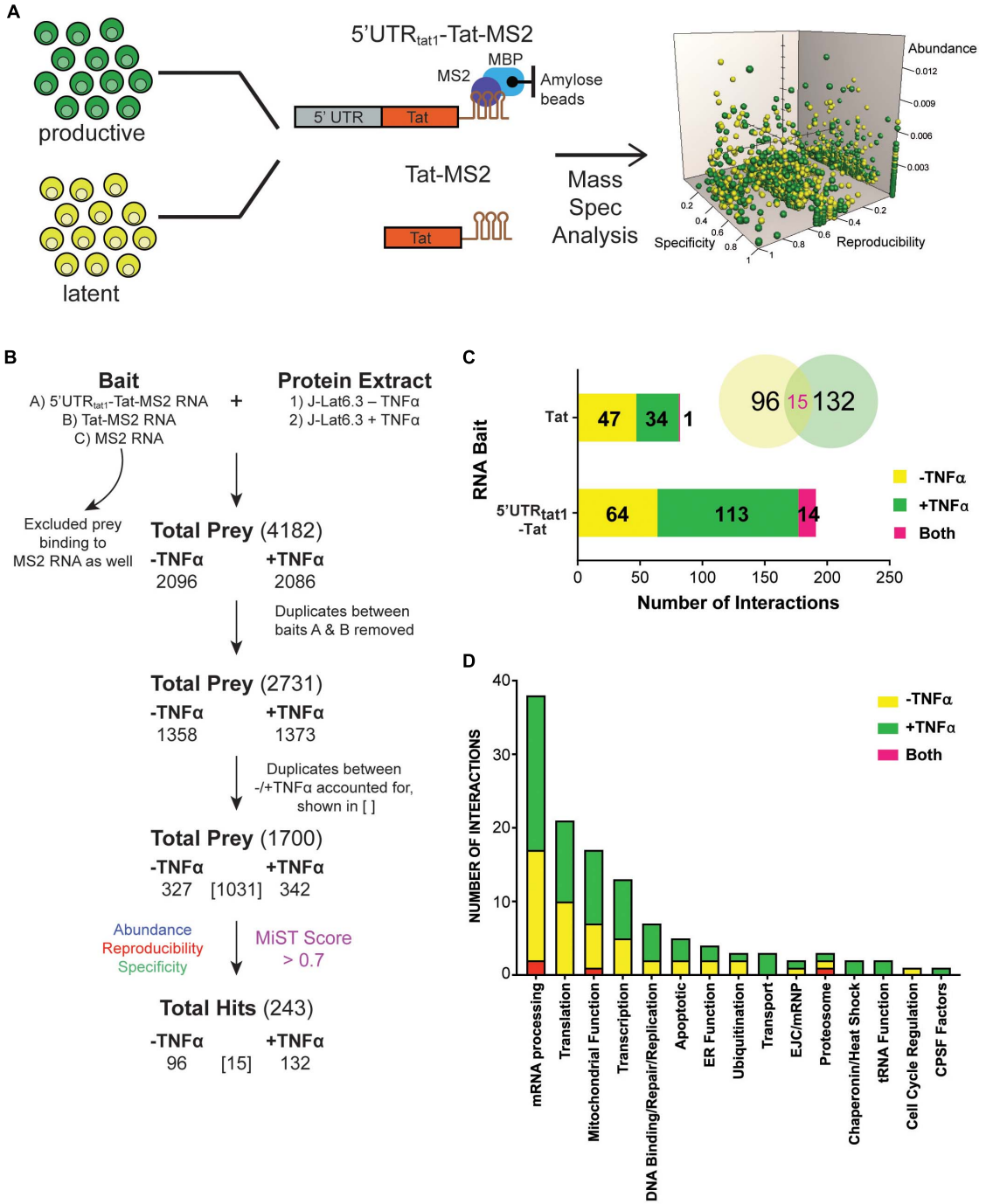


FIGURE 1 | Enrichment in mRNA processing and translation factors assembling on *tat* mRNA. **(A)** Representation of the affinity purification strategy used for isolation of cellular RNA binding proteins that interact with *tat* mRNA. Whole cell lysates were prepared from the latently infected T-cells, J-Lat clone 6.3 where untreated cells were used as the latent sample and TNF-α activated cells were used as the productively infected sample. Cellular proteins that interact with *tat* mRNA were pulled down through MBP-MS2 affinity purification using two RNA baits that contained Tat exon 2 with or without the 5′UTR from tat1 mRNA and fused to three MS2 sequences. Eluted RNA-protein complexes were analyzed by mass spectrometry. MiST analysis was used to determine the most biologically relevant interactions after combining data from three replicates. **(B)** The filtering steps applied to refine the mass spectrometry data. Prey binding to the control bait, MS2 RNA were excluded, followed by removal of duplicate prey captured by baits A (5′UTRtat-Tat) and B (Tat RNA). Prey were then defined as derived from uniquely –TNFa or +TNFa lysates or from both. Lastly, MiST analysis was applied to score interactions and a confidence threshold of 0.7 was used to refine the list of candidates for relevant Tat mRNA: cellular protein interactions. **(C)** Breakdown of the 243 proteins selected by MiST analysis by bait and lysate. **(D)** Annotation of the 243 proteins based on GO biological process.

retention to amylose beads was mediated by the MS2-MBP fusion protein containing the MS2 coat protein RNA binding domain and the *E. coli* maltose binding protein (MBP). RNP complexes were eluted by maltose and the protein composition of the purified RNPs were analyzed by mass spectrometry. The proteins identified by MS were quantitatively scored using the mass spectrometry interaction statistics (MiST) platform devised by Jäger et al. (2011).

Several filtering steps described in **Figure 1B** were applied to refine the protein preys obtained by MBP-MS2 pull-down for generation of a list of 1,700 unique proteins where duplicates between the different conditions were accounted for. These 1,700 proteins were quantitatively scored based on their abundance, reproducibility and specificity and 243 putative Tat RNA:protein interactions were identified using a confidence threshold of 0.7 for biological relevance (**Figure 1B**).

As expected, when assigned to their respective bait (**Figure 1C**, bar graph) or protein lysate (**Figure 1C**, Venn Diagram), we observed a larger number of proteins interacting with the longer RNA bait (191 proteins interacting with the 5′UTRtat1-Tat-MS2 vs. 82 with Tat-MS2). Moreover, a higher number of interactions were detected with the lysates prepared from activated T-cells (147 proteins for +TFNα vs. 111 proteins for −TNFα). Analysis of the 243 proteins by broad gene ontology terms (Spliceosome database, Cvitkovic and Jurica, 2013) showed that the top two overrepresented annotated biological processes were mRNA processing and translation (**Figure 1D**). Analysis of the GO molecular function also showed a predominance of RNA-binding proteins (**Supplementary Figure 1**).

Knockdown of Tat RNA Binding Proteins Affect Latent and Productive Infection of HIV-1

From the 243 proteins identified through MS2 chromatography affinity purification, thirteen proteins were selected to follow through with validation studies (**Figure 2A** and **Supplementary Table 3**). The main criteria for selection were the MiST score, the role of the protein in cellular pathways and the novelty of the protein in the context of regulation of HIV-1 latent infection. Two additional proteins, PTB1 and HSP90A were selected as controls as both proteins have previously been shown to be involved in regulation of HIV-1 latency (Lassen et al., 2006; Anderson et al., 2014).

For further refinement of the list of candidates for in-depth investigation, we assessed the effect of knockdown of the proteins on HIV-1 latent and productive infection. Jurkat T-cells were transduced with shRNA expressing lentivectors targeting each of the fifteen protein targets for knockdown (KD). MCM5 KD induced high cell death hence MCM5 was excluded from further investigation. Successfully transduced cells were sorted into a bulk population by RFP^+ expression (**Supplementary Figure 2A**) where the degree of knockdown was heterogeneous between individual cells. Success of the knockdown of the gene targets in the bulk Jurkat cell populations were confirmed by western blot and densitometry analysis (**Supplementary Figure 2B**). To assess the outcome of protein knockdown on HIV-1 infection, we used a single-round dual-fluorescent reporter virus (DuoFluo, R7GEmTB) based on the R7GEmC backbone described in Calvanese et al. (2013). In our DuoFluo virus, HIV-1 5′-LTR controls eGFP expression indicative of productive infection and mTagBFP2, controlled by the EF1α promotor, is the marker of latent infection (**Figure 2B**). The fluorescent phenotype of cells and corresponding profile of HIV-1 infection are shown in the grid (**Figure 2B**). We confirmed the ability of the eGFP/mTagBFP2 expressing dual-fluorescent reporter virus to identify the presence of latently (blue, BFP^+) and productively (green, GFP^+ or cyan, GFP^+ BFP^+) infected Jurkat and primary $CD4^+$ T-cells by fluorescence microscopy after infection with R7GEmTB (**Supplementary Figure 3**).

All bulk populations of the KD RFP^+ Jurkat cells were infected with the DuoFluo virus and collected for flow cytometry analysis after 72 h. Levels of eGFP and mTagBFP2 expression were assessed by gating on the RFP^+ population and compared against DuoFluo infected untransduced RFP- parental Jurkat cells. An increase in productive infection was detected after knockdown of FLNA, HMGB3, PTBP1, HSP90AA1, and KIF2C (**Figure 2C**, green bars). On the other hand, a dramatic increase in latent infection was seen after knockdown of TOP2A, SRP14, HNRNPH1, DDX1, and HNRNPL (**Figure 2C**, blue bars). The increase in latent infection after knockdown of TOP2A and HNRNPH1 was coupled with a decrease in productive infection.

PTB, known to facilitate the export of multiply spliced (MS) mRNAs to the cytoplasm (Lassen et al., 2006), appears to play a role in controlling latent and productive infection, as increases in both forms of infection were observed after knockdown of the protein (productive: 70.9%, latent: 7.61%). However, the effect on latent infection was smaller than expected. Interestingly, two genes, SRP14 and HMGB3 have not been previously reported to play a role in the regulation of HIV-1 replication and their knockdown here had very marked effects on latent and productive infection, respectively. Knockdown of SRP14 had no effect on productive infection but increased dramatically the percentage of cells entering into latency (17.0% vs. Jurkat 1.71%, **Figure 2D**). In contrast, knockdown of HMGB3 had no effect on latent infection, but increased the percentage of productively infected cells (88.8% vs. Jurkat 65.2%, **Figure 2D**). These data suggest that SRP14 is a negative regulator of latent infection, whilst HMGB3 is a negative regulator of productive infection.

Knockdown of SRP14 and HMGB3 Strongly Modulates Splicing at SA3

As a strong block to multiply splicing of HIV-1 mRNA was recently characterized in $CD4^+$ T-cells isolated from patients under ART (Yukl et al., 2018), we examined the role of knockdown of the RNA binding proteins on *tat* mRNA splicing using an HIV-1 splicing reporter, which harbors SD1 5′ss and SA3, SA4a,b,c and SA5 3′ss (Ropers et al., 2004). In this splicing reporter construct, Tat-exon1,2 (nt 1-839/5590-6044 NL4-3) was placed in the context of a human gene to recapitulate HIV integration in latently infected cells. The context of HIV-1 integration in the latent cell line, ACH2 cells (Clouse et al., 1989, Folks et al., 1989) was used as the basis of the design hence NT5C3

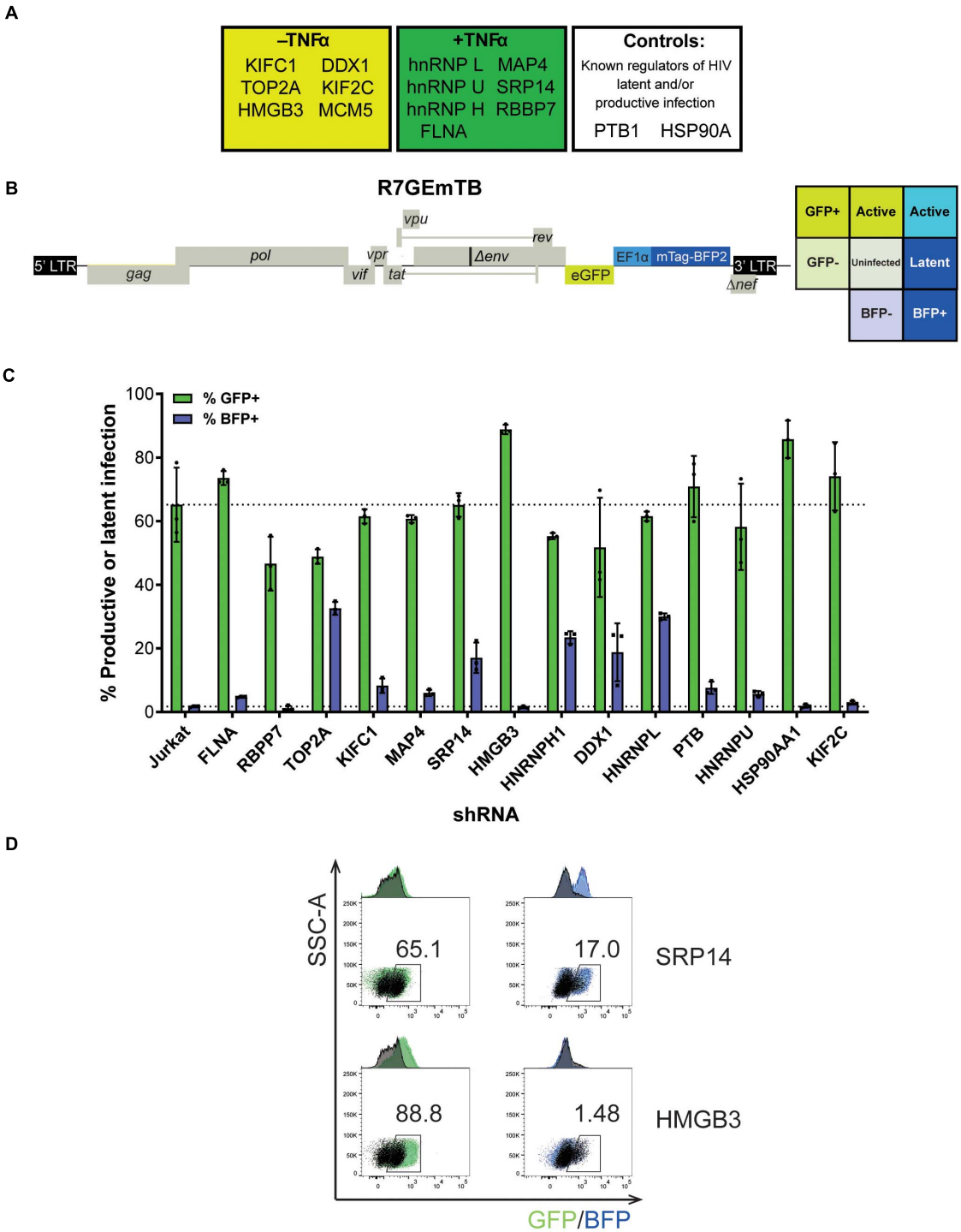


FIGURE 2 | SRP14 and HMGB3 knockdown impact HIV latent and productive infection. **(A)** Short-list of 13 proteins chosen for follow-up validation and two control proteins, PTB and HSP90A that are known regulators of HIV-1 replication. **(B)** Genome of the single round DuoFluo virus [R7/E-/GFP-EF1α-mTagBFP2 (R7GEmTB)] used in this study. The fluorescence profiles that correspond to latent or productive infection are shown in the grid. **(C)** Bulk populations of RFP^+ shRNA knockdown (KD) Jurkats were infected with the single round DuoFluo virus pseudotyped with the dual-tropic 92HT593.1 envelope, and collected 48 h later for flow cytometry analysis. Percentage of productive or latently infected cells after knockdown of the specific protein is shown, with productive infection in green and latent infection in blue. Data shown are means of three independent experiments ± SEM. **(D)** Scatter plots highlighting the shift in GFP^+ or BFP^+ populations after infection of RFP^+ SRP14 (top) or HMGB3 (bottom) KD Jurkats with DuoFluo virus. Black population represent infected, untransduced Jurkat cells, while green/blue populations represent infected SRP14/HMGB3 KD RFP^+ Jurkats. Values represent the percentage of GFP^+ or BFP^+ cells in KD RFP^+ Jurkat cells.

exon 5 and intron 5 were incorporated upstream of the HIV-1 5′-LTR (**Figure 3A**). Renilla luciferase (LucR) was introduced at the 3′ end of Tat exon 2, thus an increase in LucR would indicate splicing at SA3, SA4a,b,c or SA5. Co-transfection of the splicing reporter with an LTR-LucF reporter cassette allows Firefly luciferase to be used as a specific readout of splicing at SA3 and Tat production. The effect of cellular proteins on use of SA3 in the Tat$^+$ splicing reporter were compared to a matched Tat- splicing reporter lacking both the HIV introns and Tat exon 2 (**Figure 3A**). An increase in LucR expression in the Tat- context shows the promotion of splicing between the cellular splice site, SDc site at the 3′ end of NT5C3 exon 5 and the SAc upstream of the 5′-LTR.

Bulk populations of the KD RFP + Jurkat cells were co-transfected with the Tat$^+$ or Tat- splicing reporter and the LTR-LucF cassette and harvested for luciferase analysis 24 h later. The LucF/LucR ratio was calculated for both the Tat$^+$ and Tat- contexts in response to the knockdown of each protein target. Fold changes in Tat splicing over transfected RFP- untransduced Jurkat cells were reported in **Figure 3B**. Knockdown of SRP14 decreased splicing at SA3 by sixfold compared to untransduced Jurkats, whereas knockdown of HMGB3 and PTB increased the use of SA3 by 8.9- and 8.8-fold, respectively (**Figure 3B**). None of the other ten proteins of interest affected splicing at SA3. The effect of SRP14 and HMGB3 KD on use of SA3 is consistent with the effects of these proteins on latent and productive infection, suggesting an important role of these proteins in Tat mRNA processing and regulation of HIV-1 infection.

Knockdown of SRP14 and HMGB3 Impacts Tat Expression and Function

Major blocks of HIV transcription and translation have been reported during latency (Khoury et al., 2018a). Due to the central role of Tat protein in promoting HIV transcription and post-transcriptional events, this warranted a deeper investigation of the role of SRP14, HMGB3 and PTB on Tat expression. One caveat of using bulk populations of the KD RFP$^+$ Jurkat cells is the large clonal variation. To circumvent this obstacle, single clones were sorted following 12 days of puromycin selection and assessed through expression of RFP by flow cytometry (**Supplementary Figure 2C**). To examine KD efficiency in the various clones, changes in mRNA levels compared to untreated Jurkat T-cells was determined by RT-qPCR. Importantly, we observed across all single clones tested a significant reduction in the levels of *SRP14*, *HMGB3* and *PTB* mRNA compared to untransduced Jurkats (FC vs. Jurkat ≥50%, **Supplementary Figure 2D**).

Next, bulk and single SRP14 (B5, C10, and G4), HMGB3 (C2, D4, E7, and G9) or PTB (C10, C11, D5, and D8) shRNA

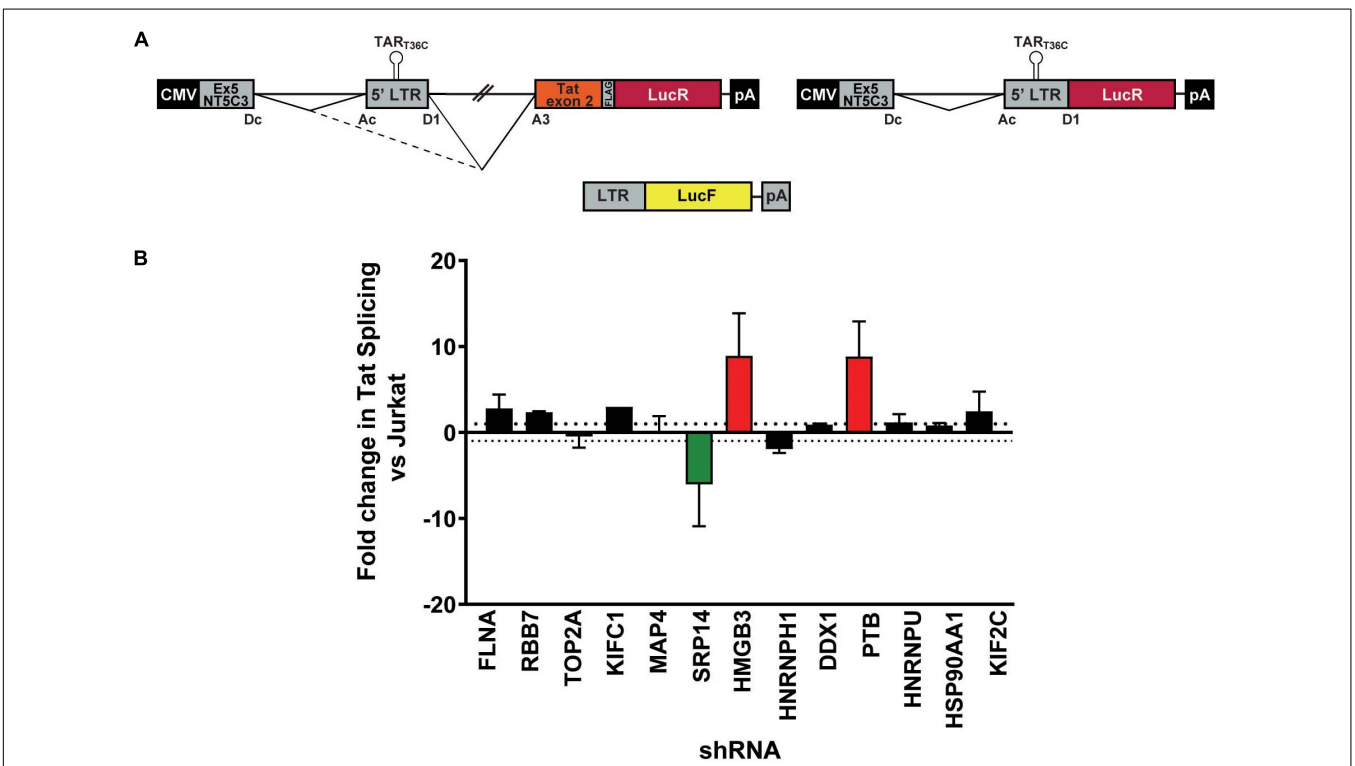


FIGURE 3 | SRP14 and HMGB3 knockdown influence *tat* mRNA splicing. **(A)** Schematic representation of the NT5C3-Tat splicing reporter construct (Tat+), the corresponding Tat- control and the transactivation reporter cassette (LTR-LucF). The splicing patterns that lead to SA3 activation and Tat expression are indicated. RFP$^+$ shRNA KD Jurkats were co-transfected with either the Tat$^+$ or Tat– splicing reporter vectors and LTR-LucF, and 24 h later harvested for luciferase readout. **(B)** The fold changes in Tat splicing (Luciferase Firefly/Renilla ratio for Tat$^+$ vs. Tat– cells) over untransduced Jurkat cells were determined. Data represent mean of three independent experiments ± SEM. D, donor; A, acceptor splice site.

KD clones were transfected with Tat expression constructs to assess the effect on Tat transactivation and translation. We previously characterized a highly conserved element underlying the Tat open reading frame, named TIM-TAM (for Tat IRES modulator of *tat* mRNA) and characterized its role in controlling Tat translation through cap- and IRES-dependent mechanisms (Khoury et al., 2020). Moreover, we developed a model system to assess Tat cap and IRES translation (Nguyen et al., 2019). In our IRES-dependent Tat translation expression cassette, the Tat encoding exons have been incorporated into the human growth hormone gene (GH1), where readthrough transcription and alternative splicing would allow low level expression of GH1-Tat protein through an IRES-dependent mechanism (**Figure 4A**). In the control construct, Tat was placed under the control of a cytomegalovirus (CMV) promoter allowing Tat expression in a cap-dependent manner. In both contexts, Tat was cloned in phase

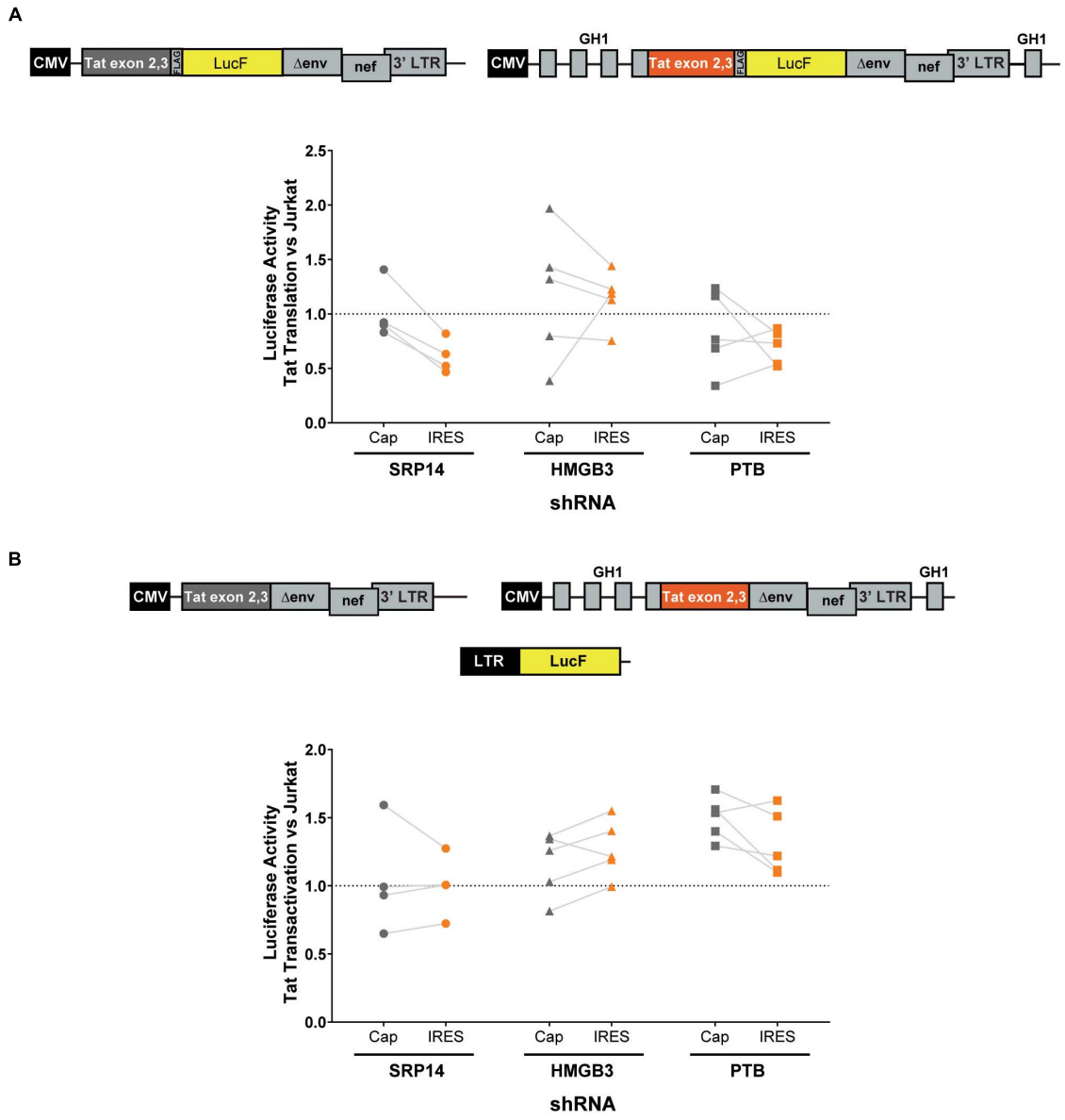


FIGURE 4 | SRP14 and HMGB3 knockdown control Tat -cap and -IRES dependent translation. Diagrams depicting Tat reporter constructs used to study Tat cap- (left) or IRES- (right) dependent translation **(A)** and transactivation **(B)**. Luciferase Firefly expression, which is produced in fusion with Tat (in A) or under the control of the 5′-LTR (in B), is used as a readout for Tat translation and transactivation, respectively. Bulk populations or single clones selected for knockdown of SRP14, HMGB3 or PTBP1 were transfected with Tat cap or IRES expression constructs to assess the effect on Tat translation **(A)** or transactivation **(B)**. Cells were harvested for luciferase analysis 24 h later. Firefly luciferase activity (LucF) was normalized on Renilla luciferase activity (LucR) and shown as a fold change over untransduced Jurkats. Data shows mean of two independent experiments with single clones and bulk populations. CMV, cytomegalovirus promoter; GH1, human growth hormone gene.

with LucF hence lucF expression is a marker of Tat translation. Bulk population and single clones of the KD RFP$^+$ Jurkat cells were transfected with cap or IRES Tat-LucF expression constructs and harvested for luciferase analysis 24 h later. Fold changes in luciferase activity were then calculated in comparison to RFP- untransduced Jurkats. In the SRP14 KD RFP + Jurkats, Tat translation was reduced from the cap-dependent context in the bulk population and 2 of 3 single clones and from the IRES-dependent context in the bulk population and all three single clones (**Figure 4A**). In contrast, HMGB3 knockdown induced an increase in Tat translation from the cap-dependent context for the bulk population and 2 of 4 single clones, and in the bulk population and 3 of 4 single clones for the IRES-dependent context (**Figure 4A**). Lastly, knockdown of PTB resulted in a decrease in Tat translation in 2 of 4 single clones transfected with Tat-lucF construct, as well as all 4 single clones and the bulk population transfected with Tat IRES construct (**Figure 4A**).

We next investigated the effect of SRP14 and HMGB3 KD on Tat transactivation by using a modified expression cassette system, where Tat is translated through an IRES- or cap-dependent pathway and the luminescence readout is induced by Tat transactivation of an HIV-1 LTR-LucF reporter cassette. The effect on Tat transactivation after knockdown of SRP14 was variable across the single clones for both cap- and IRES-Tat translation, however, there was a clear increase in Tat transactivation for 3 out of 4 Tat-cap transfected single clones and 4 out of 4 Tat-IRES transfected single clones after knockdown of HMGB3 compared to untransduced Jurkats (**Figure 4B**). Knockdown of PTBP1 resulted in an increase in Tat transactivation for all single clones transfected with Tat-Cap and Tat-IRES (**Figure 4B**). These data demonstrate a role of SRp14 and HMGB3 in controlling HIV-1 latent and productive infection in Tat-dependent manner.

SRP14 and HMGB3 Binds in the Vicinity of Tat Start Codon

Whilst SRP14 and HMGB3 were detected in the RNP complexes (**Supplementary Figure 4**), further investigation was required to assess direct interaction of these proteins with *tat* mRNA. To delineate SRP14, HMGB3, and PTB binding sites on multiply spliced RNA, we performed footprinting assays coupled to SHAPE (selective 2′ hydroxyl acylation analyzed by primer extension) analysis. We have recently determined *tat1* and *tat2* mRNA folding using enzymatic and chemical probing and identified a highly conserved sequence-structure within MS RNA, TIM-TAM (Khoury et al., 2020). TIM-TAM forms the apical part of an irregular stem-loop structure $SLS3_{A3}$ that harbors the Tat start codon. TIM-TAM controls the timing and level of Tat translation during the early and late phases of infection, while promoting latent infection and virus reactivation. Footprinting assays were performed on RNP complexes formed by *tat2* transcript (**Figure 5A**) and recombinant proteins at three different [RNA]/[protein] ratios 5, 10, and 20. Normalized shape reactivities and probing data were used to determine the binding sites of SRP14, HMGB3, and PTB. At the lowest [RNA]/[protein] ratio, protections were mainly detected on $SLS3_{A3}$ including TIM-TAM and its bordering sequences. Strong protections were also observed in the 5′ untranslated region, more specifically on the TAR, PBS, and DIS elements (**Figure 5B**). Upon increasing SRP14, HMGB3, and PTB concentration, protections of TIM-TAM were reinforced and new ones were detected on the Tat start codon and in the vicinity of SA3. Altogether, these data are consistent with direct binding of SRP14, HMGB3 as well as PTB to MS RNA highlighting a potential role of these RNA binding proteins in controlling Tat expression during latent infection.

SRP14 Reactivates HIV-1 Latently Infected Cells and Virus Production

To assess the role of SRP14 in controlling latent infection, we tested the effect of its overexpression on virus reactivation in a T-cell line model of latent infection, using the J-Lat 10.6, 8.4 and A2 clones. J-Lats are Jurkat derived cells containing one stably integrated, but transcriptionally silenced full-length HIV-1 genome with GFP in place of the nef gene (Jordan et al., 2003). J-Lat cells were transduced with pInducer-SRP14-T2A-mtagBFP2 lentiviral vectors and cultured in the absence or presence of doxycycline (+Dox) for 2 days. Representative scatter plot highlighting mtagBFP2 expression following treatment of transduced cells with 5 μg/mL doxycycline is shown in **Figure 6A**. SRP14 expression was validated by western-blot analysis (data not shown). The different J-Lat clones exhibit variable levels of basal GFP expression. Upon doxycycline treatment, we observed an increase in GFP expression in all T-cell lines. Indeed, a significant increase in mean fluorescence intensity (MFI) was detected for BFP$^+$GFP$^+$ cells vs. GFP$^+$ cells alone (**Figure 6B**).

To determine whether differences in RNA binding protein levels in resting and activated CD4$^+$ T-cells might be involved in the observed effect on HIV expression, we measured SRP14, HMGB3 and PTB protein and RNA levels in resting and α-CD3/CD28 stimulated CD4$^+$ T-cells from patients living with HIV on ART. Western-blot analysis indicated low expression levels of SRP14 and HMGB3 in resting cells, and upon stimulation a 2.5- and 18.8-fold increase in SRP14 and HMGB3 protein level was detected, respectively (**Figure 7A**). Similar results were observed with RT-qPCR analysis of *HMGB3* mRNA levels between resting and activated CD4$^+$ T-cells as we detected a 28.5-fold increase in *HMGB3* mRNA expression following CD4$^+$ T-cell activation (**Figure 7B**). However, no changes in *SRP14* mRNA levels were seen in response to T-cell activation. Interestingly, while a 4.4-fold increase in *PTB* mRNA expression was observed, no significant change in PTB protein expression was detected following stimulation of CD4$^+$ T-cells.

PTB was identified as an HIV RNA binding protein that induces virus reactivation and release of replication competent virus in resting CD4$^+$ T-cells from patient on ART (Lassen et al., 2006). To examine whether SRP14 might also act as a positive factor for HIV-1 gene expression, resting CD4$^+$ T-cells isolated from people living with HIV on ART were electroporated with SRP14 or PTB Dox-inducible expression constructs alone or in combination using an Amaxa nucleofector. After 48 h, virion release into the culture supernatant was assessed using

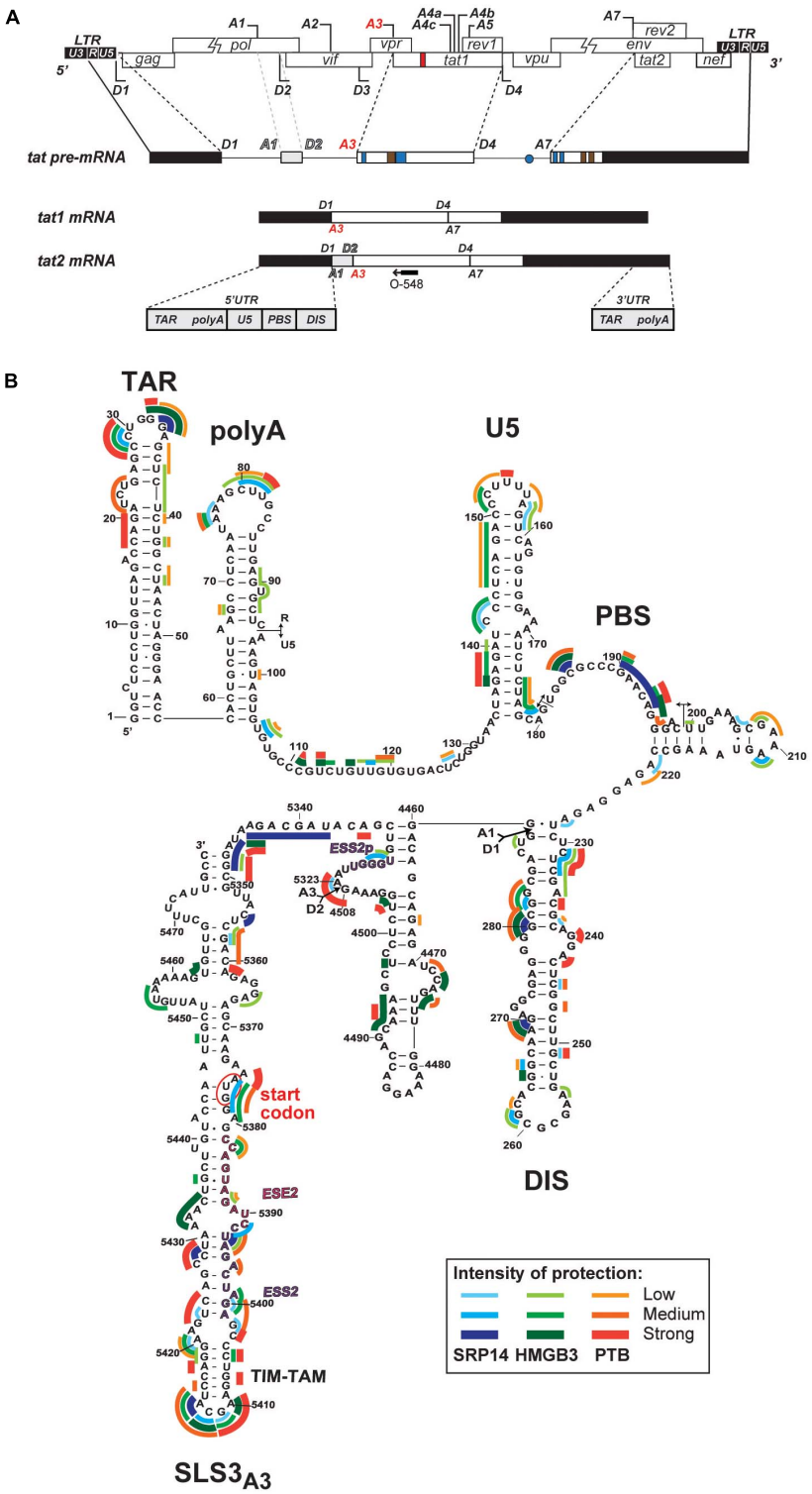


FIGURE 5 | Interaction of SRP14, HMGB3 and PTB with *tat* mRNA. **(A)** Schematic representation of HIV-1 proviral genome, *tat* pre-mRNA and multiple splice events leading to *tat1* and *tat2* mRNAs production during the early phases of HIV-1 infection. D represent donor splice sites, A represent acceptor splice sites, 5′UTR/3′UTR indicate 5′/3′ untranslated regions. The hybridisation site used for primer extension with O-548 (also known as Odp3102) is indicated by black arrow. **(B)** Probing of SRP14, HMGB3, and PTB binding sites on *tat2* mRNA. *Tat2* transcript (1 pmol) was modified with 1M7 (65 mM) in the absence or presence of increasing concentration (0.41—0.83—1.6 μM) of recombinant proteins. Conditions of modification are given in Materials and Methods. Protections generated by SRP14, HMBG3 and PTB recombinant proteins are indicated on the secondary structure model of *tat2* mRNA by blue, green and red lines, respectively. Pale, medium, and dark colors indicate the intensity of protections (low, medium and strong protections). Numbering of nucleotides and positions of the *cis* regulatory elements are given in reference to HIV-1 BRU (K02013). The start codon of Tat protein is circled.

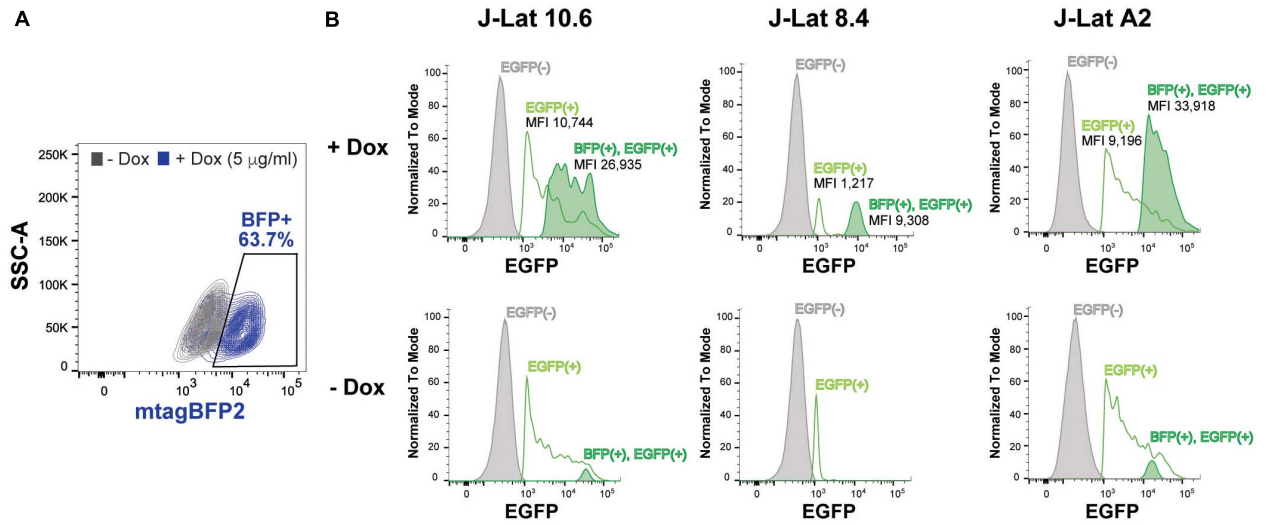


FIGURE 6 | SRP14 expression reactivates latently infected T-cell lines. **(A)** J-Lat cells were transduced with SRP14-T2A-mtagBFP2 lentivector and cultured with (+Dox) or without (–Dox) doxycycline at 5 μg/mL. Cells were analyzed for mtagBFP2 expression by flow cytometry 72 h post-infection. **(B)** Histogram depicting $EGFP^{+}$ expression in latently infected T-cells following transduction with SRP14-T2-mtagBFP2. Virus reactivation in J-Lat 10.6, 8.4 and A2 clones is shown by an increase in mean fluorescence intensity (MFI) of $BFP^{(+)}EGFP^{(+)}$ cells in comparison to $EGFP^{(-)}$ and $EGFP^{(+)}$ cells.

an RT-ddPCR assay. One out of the four patient $CD4^{+}$ T-cells electroporated with SRP14 and $SRP14^{+}$ PTB presented an upregulation of virus production upon doxycycline treatment (**Figure 7C**). Virus production following SRP14 overexpression was not coupled with an increase in cell activation as only 0.9% and 1.3% of SRP14 transfected cells expressed middle (CD25) and late (HLA-DR) activation markers, respectively (**Figure 7D**).

DISCUSSION

We have characterized *tat* RNA:cellular protein interactions differentially expressed between productive and latent infection. Out of the 243 proteins identified by mass spectrometry, multiple cellular factors were investigated for their putative roles in the control of Tat expression and viral replication. A consistent effect on Tat expression and HIV-1 replication was exerted by both SRP14 and HMGB3, where SRP14 acts as a positive regulator of Tat expression and negative regulator of latent infection while HMGB3 acts as a negative regulator of Tat expression and negative regulator of productive infection (**Figure 8**). However, the exact mechanisms exerted by SRP14 and HMGB3 on the pathways of Tat expression have not been determined. It should be noted that knockdown of SRP14 and HMGB3 affected to a larger degree Tat expression in the IRES context. This suggests that SRP14 and HMGB3 proteins are acting through the Tat IRES pathway, by directly interacting with TIM-TAM or by acting as scaffolds for other RNA-binding proteins.

In a recent study, Yukl's group identified HIV multiple splicing as a common block in three primary cell models of latent infection and in peripheral $CD4^{+}$ T-cells isolated from HIV infected ART suppressed individuals (Moron-Lopez et al., 2020), confirming previous findings from the same group that showed a series of blocks to HIV proximal elongation, distal transcription/polyadenylation and splicing preventing HIV expression in $CD4^{+}$ T-cells from blood of HIV infected patients on ART (Yukl et al., 2018). Prior studies have shown that MS RNA encoding Tat protein inhibits the establishment of HIV latency (Donahue et al., 2012). When present, Tat activates virus replication at a higher rate than any of the known LRAs (Razooky et al., 2015; Khoury et al., 2018b) by potently inducing HIV transcription and splicing (Khoury et al., 2018b). Moreover, by controlling its own production at the splicing (Jablonski et al., 2010) and translational levels (Charnay et al., 2009, Khoury et al., 2020), Tat acts as a switch for productive and latent HIV infection (Khoury et al., 2020).

siRNA knockdown of TOP2A in Sup T1 cells (Balakrishna et al., 2013), MAP4 in TZM-bl, HEK 293T and HeLa P4.2 cells (Brass et al., 2008; König et al., 2008; Gallo and Hope, 2012), HNRNPH1 in 293T cells (König et al., 2008), DDX1 in HeLa cells (Edgcomb et al., 2012) and HNRNPU in HeLa P4/R5 cells (Zhou et al., 2008) inhibit HIV-1 replication, corroborating our findings following shRNA KD of these proteins in Jurkat cells. Furthermore, an enrichment in mRNA processing proteins was observed in two previous HIV pull-down assays that used HIV-1 5′ leader sequence and unspliced RNA (Vallejos et al., 2011; Knoener et al., 2017). Although the pull-down methods and cell types used in both these studies were distinct, common proteins with our *tat* RNA pull-down assay were identified such as HNRNPL, HNRNPU, SRP14, and TERA that were isolated from HeLa cells arrested at G2/M with the HIV-1 5′UTRgag (Vallejos et al., 2011), as well as PPIA, GSTP1, STIP1, PHB, NUDC, FLNB, FUBP1, DEK, MAP4, CLIC1, and CD47 (**Supplementary Table 4**) identified from Jurkat cells infected with NL4-3 and unspliced HIV-1 RNA-cellular protein complexes (Knoener et al., 2017).

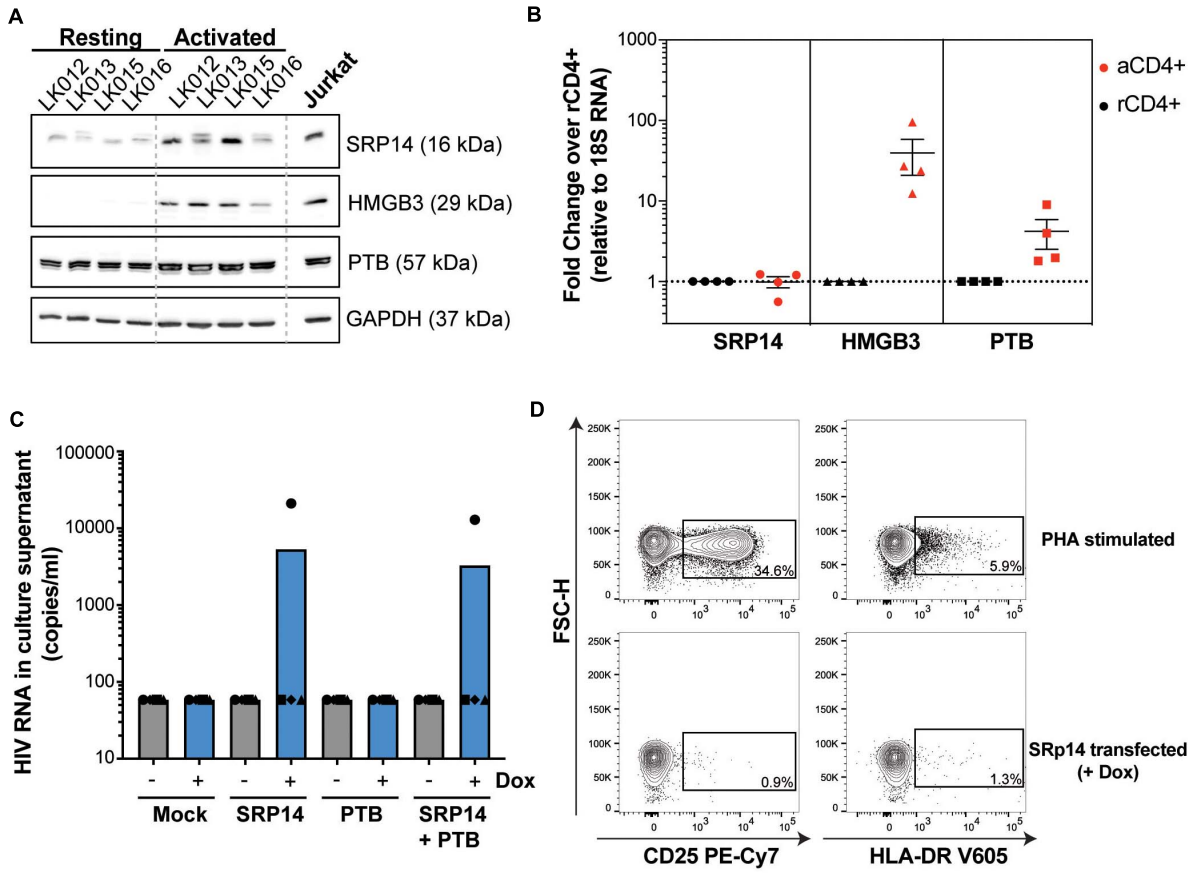


FIGURE 7 | SRP14 and PTB expression induce HIV expression in primary CD4$^+$ T-cells from patients on ART. **(A)** SRP14, HMGB3 and PTB protein expression in resting CD4$^+$ T-cells isolated from patient under antiretroviral therapy (LK, leukophoresis) increase in response to T-cell stimulation by α-CD3/CD28 (1 μg/mL; 0.5 μg/mL). Expression of SRP14, HMGB3, and PTB was measured by western-blot in resting and activated CD4$^+$ T-cells 72 h post-stimulation. GAPDH was used as a loading control and Jurkat protein lysate as a positive control for antibody detection of the various proteins. **(B)** SRP14, HMGB3, and PTB RNA levels were assessed by RT-qPCR and shown as fold change in activated (aCD4$^+$) vs. resting CD4$^+$ (rCD4$^+$) after normalization on 18S RNA. **(C)** Effect of SRP14, HMGB3 and PTB overexpression following doxycycline (+Dox) treatment on HIV-1 RNA levels in the culture supernatant 48 h post-stimulation. **(D)** Plots depicting the expression of middle (CD25) and late (HLA-DR) activation markers on SRP14 transfected cells and PHA stimulated cells.

It should be noted that the large isoform of PTB, PTBP1, can overcome the nuclear retention of MS RNA in latently infected cells when overexpressed (Lassen et al., 2006). However, the levels detected in resting CD4$^+$ T-cells from patients on ART were no different to activated CD4$^+$ T-cells. Hence the specific isoform involved in Tat expression may be distinct to the isoform responsible for reversing aberrant accumulation of MS RNA. Further studies are required to determine the contribution of PTB to latency.

SRP14 (Signal Recognition Particle 14) is part of the signal recognition particle RNP complex, which functions by arresting the ribosome during translation so that secretory proteins can be correctly targeted to the endoplasmic reticulum (Weichenrieder et al., 2000; Lakkaraju et al., 2008). SRP14 together with SRP9 inhibits both cap- and IRES-dependent initiation. By binding to 40S ribosomal subunits, SRP9/14 prevents 48S complex formation hence interfering with the recruitment of mRNA to 40S subunits (Ivanova et al., 2015). HMGB3 (High Mobility Group Box 3) is thought to be a DNA-binding protein that can remodel chromatin structures and can also act as a nucleic acid sensor—these are putative functions inferred from its shared homology with other proteins in the HMG-Box family (Nemeth et al., 2003). An analysis of the entire mRNA bound proteome in HEK 293 cells, however, identified HMGB3 as a potential RNA-binding protein (Baltz et al., 2012) although to date, no specific RNA binding associated function has been reported for HMGB3. In the present study, SRP14 and HMGB3 appear to have a positive and inhibitory role on Tat IRES translation, respectively. To our knowledge, these findings are the first to link SRP14 and HMGB3 to HIV expression. The exact mechanisms of which require investigation.

Our findings highlight SRP14 and HMGB3 as potential targets of pharmaceutical intervention. This still does not present an ideal situation as, akin to LRAs used in the past, targeting of these cellular proteins do not confer specificity against latent HIV-1. These proteins, however, are unlikely to have an involvement in gene expression pathways as extensive as that of the epigenetic modifiers and proteins involved in the NFκB and

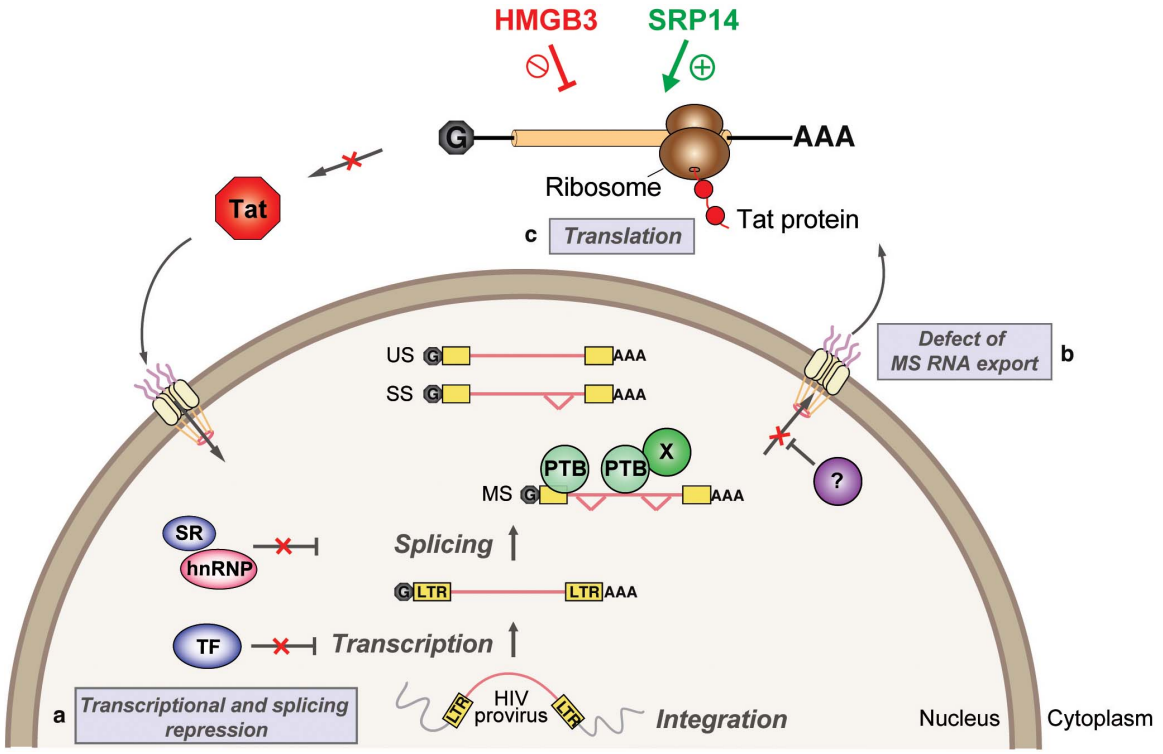


FIGURE 8 | Model depicting SRP14 and HMGB3 roles in post-transcriptional regulation of HIV-1 gene expression in latently infected cells. Multiple factors have been implicated in the establishment and maintenance of latency including **(a)** transcriptional and splicing repression by sequestration of essential transcription factors (TF) and/or splicing factors; **(b)** defect in multiply spliced (MS) RNA export due to insufficient levels of Rev and/or RNA binding factors such as PTB (for polypyrimidine tract binding protein); and **(c)** aberrant localization of MS RNA in the nucleus imposes a block on translation. Due to its central role in transcriptional transactivation and production of full-length viral mRNAs, small stochastic changes in Tat expression, the absence or impairment of Tat function would favor latent infection. By targeting *tat* mRNA, SRP14 activates Tat translation and hence stimulates HIV replication, while HMGB3 protein inhibits Tat translation and consequently impedes HIV expression in resting CD4$^+$ T cells.

NFAT pathways, and hence may only impact upon the expression of a limited array of cellular genes.

There are several limitations in this study that should be acknowledged. Given the large number of cells required to complete the pull-down assay, we chose J-Lat cells to use for the screening of Tat RNA binding proteins during latency given their transformed nature and homogeneity of the cell population. While latently infected T-cell lines are useful for screening, a common limitation of using T-cell lines is that the intrinsic cellular factors and environment in T-cell lines are very different to that of primary T-lymphocytes hence the cellular factors that influence HIV latency may differ between T-cell lines and primary CD4$^+$ T-cells. Another caveat is the pull-down assay was performed on a mixed population of productively infected cells, as 100% reactivation of J-Lat cells is never achieved even after treatment with strong mitogens, including TNF-α. Virus expression from HIV latently infected T-cells could be sorted into two sub-populations; inducible (GFP$^+$) and non-inducible proviruses (GFP−), following treatment with TNF-α. The reporter systems used in this study assess different stages of the Tat expression pathway. However, these assays do not allow evaluation of these stages independently of the others. Further validation studies are required to determine the exact mechanisms exerted by SRP14 and HMGB3 on the pathways of Tat expression. Finally, in this study we did not address whether SRP14 expression is sufficient to circumvent MS RNA retention in the nucleus of resting CD4$^+$ T-cells, and whether it can induce production of replication competent virus. Further experiments are also required to test the effect of HMGB3 and SRP14 on *tat* mRNA export and/or stability.

Reactivation of HIV-1 can result from fluctuations in the levels of Tat protein (Weinberger et al., 2005; Khoury et al., 2020) as well as cellular factors involved in reinforcement of silent infection (Burnett et al., 2009). Hence, targeting Tat expression may serve as the basis for development of a more biologically relevant "shock and kill" strategy, a process that could lead to the discovery of an effective and durable functional cure for HIV-1.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: PRIDE, PXD025782.

ETHICS STATEMENT

The studies involving the use of blood samples from HIV negative donors were reviewed and approved by the Human Research and Ethics Committees from the University of Melbourne (15-09VIC-03 and 17-08VIC-01). All HIV-1 seronegative donors were recruited by the Red Cross Blood Bank (Melbourne, Australia) and provided written informed consent for the use of their blood products for the research. The use of blood samples from people living with HIV was approved by the Alfred Hospital (HREC214/15) for the study entitled Large volume peripheral blood mononuclear cells (PBMCs) collection by leukapheresis to define HIV persistence in HIV-infected adults. All participants provided informed consent and the protocol was approved by the local Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GK, ML, and DP conceived and designed the study. GK, ML, and SR acquired the data. GK, ML, SS, and DP analyzed and interpreted the data. JM, AP, and SL provided resources. GK and ML drafted the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This study was supported by the National Health and Medical Research Council of Australia (NHMRC) program and project grants (#1052979 and #1129320 to DP), NHMRC Principal Research Fellowship (#1137739 to AP) and Australian Centre for HIV and Hepatitis Virology Research (ACH^2) (2015-21 to DP).

ACKNOWLEDGMENTS

We would like to thank the Department of Infectious Diseases at the Alfred Hospital and the Lewin/Cameron Clinical Research Group at the Peter Doherty Institute for providing the leukapheresis samples. We thank L. Wakim for providing the pInducer10 lentivector and her insight into T-cell infections. We thank M. Olshansky for his assistance in the bioinformatics analysis, N. Peirce, D. Yurick, J. Edwards, and A. Uldrich for their technical help. We thank the DMI Flow Facility staff for their advice and generous assistance in maintaining the flow cytometers and sorters used in this study. We also acknowledge instrumentation, training and technical support by the Monash Biomedical Proteomics Facility. Computational resources were supported by the R@CMon/Monash Node of the Nectar Research Cloud, an initiative of the Australian Government's Super Science Scheme and the Education Investment Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.680725/full#supplementary-material

Supplementary Figure 1 | GO Biological Process analysis of the full set of proteins detected by mass spectrometry. List of proteins that were uniquely detected in the pull-down assay using untreated J-Lat 6.3 **(A)** or TNF-α activated **(B)** lysates were analyzed through *Enrichr* and overrepresented GO terms are depicted in the bar graphs. Examples of proteins belonging to the top classes are listed.

Supplementary Figure 2 | Protein and mRNA profiles of KD RFP^+ Jurkat cells. **(A)** Representative histograms of RFP^+ shRNA transduced Jurkats (bulk cells) where SRP14, HMGB3 or PTB were targeted for knockdown. **(B)** Representative western blots showing knockdown in protein levels of the three genes of interest, SRP14, HMGB3, PTB in their respective bulk shRNA-generated cell lines. Numbers below the blots show relative amounts of proteins in comparison to untransduced Jurkats. **(C)** Expression of RFP in bulk or single clone Jurkat cells targeted for knockdown of SRP14, HMGB3 or PTB. **(D)** Fold change in SRP14, HMGB3 and PTB mRNA levels in their respective bulk shRNA-generated cell lines or single clones compared to untreated Jurkat cells as determined by RT-qPCR.

Supplementary Figure 3 | R7GEmTB reporter virus label latently infected cells. Jurkat and $CD4^+$ T-cells were infected with R7GEmTB dual color reporter virus and subjected to confocal microscopy to assess the infection phenotype. GFP^+ and GFP^+ BFP^+ cells, representative of productive infection are shown in green and cyan, while BFP^+ cells representative of latent infection are shown in blue.

Supplementary Figure 4 | Western-blot analysis of RNP complexes. Analysis of the protein content of the RNP complexes formed on *tat* mRNA by western-blot using antibodies directed against SRP14 and HMGB3. MBP-MS2 was used as a loading control (ponceau stain). JL− and JL+ represent total lysates prepared from J-Lat 6.3 left untreated or treated with TNF-α, respectively.

Supplementary Table 1 | Participant demographics.

Supplementary Table 2 | Primers and probes used in this study. The number, sequence, and usage of each primer are given. Restriction sites are underlined, start and stop codons are in boldfaces.

Supplementary Table 3 | List of protein hits selected for follow-up analysis. Summary of the gene ID, uniprot, gene symbol, protein name, function, mass (in kDa), and protein length (in amino acid, aa) are indicated.

Supplementary Table 4 | List of proteins identified by affinity purification coupled to mass spectrometry in the RNP complexes formed with 5′UTRtat1-Tat-MS2 and Tat-MS2 RNAs. The number of peptides identified for each bait (5′UTRtat-Tat-MS2 and Tat-MS2) and prey (protein) as well as their fold-enrichment in resting and activated cells are indicated.

REFERENCES

Ait-Ammar, A., Kula, A., Darcis, G., Verdikt, R., De Wit, S., Gautier, V., et al. (2020). Current status of latency reversing agents facing the heterogeneity of HIV-1 cellular and tissue reservoirs. *Front. Microbiol.* 10:3060. doi: 10.3389/fmicb.2019.03060

Anderson, I., Low, J. S., Weston, S., Weinberger, M., Zhyvoloup, A., Labokha, A. A., et al. (2014). Heat shock protein 90 controls HIV-1 reactivation from latency. *Proc. Natl. Acad. Sci. U S A.* 111, E1528–E1537.

Balakrishna, S. L., Satyanarayana, N., and Kondapi, A. K. (2013). Involvement of human topoisomerase II isoforms in HIV-1 reverse transcription. *Arch. Biochem. Biophys.* 532, 91–102. doi: 10.1016/j.abb.2013.01.010

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell.* 46, 674–690. doi: 10.1016/j.molcel.2012.05.021

Bar, A., Marchand, V., Khoury, G., Dreumont, N., Mougin, A., Robas, N., et al. (2011). Structural and functional analysis of the Rous Sarcoma virus negative regulator of splicing and demonstration of its activation by the 9G8 SR protein. *Nucleic Acids Res.* 39, 3388–3403. doi: 10.1093/nar/gkq1114

Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A., Xavier, R. J., et al. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319, 921–926. doi: 10.1126/science.1152725

Burnett, J. C., Miller-Jensen, K., Shah, P. S., Arkin, A. P., and Schaffer, D. V. (2009). Control of stochastic gene expression by host factors at the HIV promoter. *PLoS Pathog.* 5:e1000260. doi: 10.1371/journal.ppat.1000260

Calvanese, V., Chavez, L., Laurent, T., Ding, S., and Verdin, E. (2013). Dual-color HIV reporters trace a population of latently infected cells and enable their purification. *Virology* 446, 283–292. doi: 10.1016/j.virol.2013.07.037

Charnay, N., Ivanyi-Nagy, R., Soto-Rifo, R., Ohlmann, T., Lopez-Lastra, M., and Darlix, J. L. (2009). Mechanism of HIV-1 Tat RNA translation and its activation by the Tat protein. *Retrovirology* 6:74. doi: 10.1186/1742-4690-6-74

Chiu, Y. L., Coronel, E., Ho, C. K., Shuman, S., and Rana, T. M. (2001). HIV-1 Tat protein interacts with mammalian capping enzyme and stimulates capping of TAR RNA. *J. Biol. Chem.* 276, 12959–12966. doi: 10.1074/jbc.m007901200

Chiu, Y.-L., Ho, C. K., Saha, N., Schwer, B., Shuman, S., and Rana, T. M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. *Mol. Cell.* 10, 585–597. doi: 10.1016/s1097-2765(02)00630-5

Clouse, K. A., Powell, D., Washington, I., Poli, G., Strebel, K., Farrar, W., et al. (1989). Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. *J. Immunol.* 142, 431–438.

Cvitkovic, I., and Jurica, M. S. (2013). Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.* 41, 132–141.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., et al. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell Biol.* 26, 5528–5543. doi: 10.1128/mcb.00582-06

Deeks, S. G., Lewin, S. R., Ross, A. L., Ananworanich, J., Benkirane, M., Cannon, P., et al. (2016). International AIDS Society global scientific strategy: Towards an HIV cure 2016. *Nat. Med.* 22, 839–850.

Donahue, D. A., Kuhl, B. D., Sloan, R. D., and Wainberg, M. A. (2012). The viral protein Tat can inhibit the establishment of HIV-1 latency. *J. Virol.* 86, 3253–3263. doi: 10.1128/jvi.06648-11

Edgcomb, S. P., Carmel, A. B., Naji, S., Ambrus-Aikelin, G., Reyes, J. R., Saphire, A. C. S., et al. (2012). DDX1 is an RNA-dependent ATPase involved in HIV-1 Rev function and virus replication. *J. Mol. Biol.* 415, 61–74. doi: 10.1016/j.jmb.2011.10.032

Folks, T. M., Clouse, K. A., Justement, J., Rabson, A., Duh, E., Kehrl, J. H., et al. (1989). Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. *Proc. Natl. Acad. Sci. U S A* 86, 2365–2368. doi: 10.1073/pnas.86.7.2365

Gallo, D. E., and Hope, T. J. (2012). Knockdown of MAP4 and DNAL1 produces a post-fusion and pre-nuclear translocation impairment in HIV-1 replication. *Virology.* 422, 13–21. doi: 10.1016/j.virol.2011.09.015

Ivanova, E., Berger, A., Scherrer, A., Alkalaeva, E., and Strub, K. (2015). Alu RNA regulates the cellular pool of active ribosomes by targeted delivery of SRP9/14 to 40S subunits. *Nucleic Acids Res.* 43, 2874–2887. doi: 10.1093/nar/gkv048

Jablonski, J. A., Amelio, A. L., Giacca, M., and Caputi, M. (2010). The transcriptional transactivator Tat selectively regulates viral splicing. *Nucleic Acids Res.* 38, 1249–1260. doi: 10.1093/nar/gkp1105

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., et al. (2011). Global landscape of HIV-human protein complexes. *Nature* 481, 365–370.

Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* 22, 1868–1877. doi: 10.1093/emboj/cdg188

Karabiber, F., McGinnis, J. L., Favorov, O. V., and Weeks, K. M. (2013). QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 19, 63–73. doi: 10.1261/rna.036327.112

Khoury, G., Darcis, G., Lee, M. Y., Bouchat, S., Van Driessche, B., Purcell, D. F., et al. (2018a). *The Molecular Biology of HIV Latency. In: HIV vaccine and cure - The Path Towards Finding an Effective Cure and Vaccine*. Cham: Springer Nature.

Khoury, G., Mota, T. M., Li, S., Tumpach, C., Lee, M. Y., Jacobson, J., et al. (2018b). HIV latency reversing agents act through Tat post translational modifications. *Retrovirology* 15:36.

Khoury, G., Mackenzie, C., Ayadi, L., Lewin, S. R., Branlant, C., and Purcell, D. F. J. (2020). Tat IRES modulator of tat mRNA (TIM-TAM): a conserved RNA structure that controls Tat expression and acts as a switch for HIV productive and latent infection. *Nucleic Acids Res.* 48, 2643–2660. doi: 10.1093/nar/gkz1181

Knoener, R. A., Becker, J. T., Scalf, M., Sherer, N. M., and Smith, L. M. (2017). Elucidating the in vivo interactome of HIV-1 RNA by hybridization capture and mass spectrometry. *Sci. Rep.* 7:16965.

König, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M. C., Irelan, J. T., et al. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 Replication. *Cell* 135, 49–60. doi: 10.1016/j.cell.2008.07.032

Krueger, B. J., Varzavand, K., Cooper, J. J., and Price, D. H. (2010). The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. *PLoS One.* 5:e12335. doi: 10.1371/journal.pone.0012335

Lakkaraju, A. K., Mary, C., Scherrer, A., Johnson, A. E., and Strub, K. (2008). SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites. *Cell* 133, 440–451. doi: 10.1016/j.cell.2008.02.049

Lassen, K. G., Ramyar, K. X., Bailey, J. R., Zhou, Y., and Siliciano, R. F. (2006). Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. *PLoS Pathog.* 2:e68. doi: 10.1371/journal.ppat.0020068

Liu, J., Henao-Mejia, J., Liu, H., Zhao, Y., and He, J. J. (2011). Translational regulation of HIV-1 replication by HIV-1 rev cellular cofactors Sam68, eIF5A, hRIP, and DDX3. *J. Neuroimmun. Pharmacol.* 6, 308–321. doi: 10.1007/s11481-011-9265-8

Lozano, G., and Martínez-Salas, E. (2015). Structural insights into viral IRES-dependent translation mechanisms. *Curr. Opin. Virol.* 12, 113–120. doi: 10.1016/j.coviro.2015.04.008

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., et al. (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* 8:e1000276. doi: 10.1371/journal.pbio.1000276

Meerbrey, K. L., Hu, G., Kessler, J. D., Roarty, K., Li, M. Z., Fang, J. E., et al. (2011). The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc. Natl. Acad. Sci. U S A.* 108, 3665–3670. doi: 10.1073/pnas.1019736108

Monette, A., Ajamian, L., Lopez-Lastra, M., and Mouland, A. J. (2009). Human immunodeficiency virus type 1 (HIV-1) induces the cytoplasmic retention of heterogeneous nuclear ribonucleoprotein A1 by disrupting nuclear import: implications for HIV-1 gene expression. *J. Biol. Chem.* 284, 31350–31362.

Moron-Lopez, S., Telwatte, S., Sarabia, I., Battivelli, E., Montano, M., Macedo, A. B., et al. (2020). Human splice factors contribute to latent HIV infection in primary cell models and blood CD4+ T cells from ART-treated individuals. *PLoS Pathog.* 16:e1009060. doi: 10.1371/journal.ppat.1009060

Mortimer, S. A., and Weeks, K. M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129, 4144–4145. doi: 10.1021/ja0704028

Muniz, L., Egloff, S., Ughy, B., Jády, B. E., and Kiss, T. (2010). Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator tat. *PLoS Pathog.* 6:e1001152. doi: 10.1371/journal.ppat.1001152

Nemeth, M. J., Curtis, D. J., Kirby, M. R., Garrett-Beal, L. J., Seidel, N. E., Cline, A. P., et al. (2003). Hmgb3: an HMG-box family member expressed in primitive hematopoietic cells that inhibits myeloid and B-cell differentiation. *Blood* 102, 1298–1306. doi: 10.1182/blood-2002-11-3541

Nguyen, W., Jacobson, J., Jarman, K. E., Jousset Sabroux, H., Harty, L., McMahon, J., et al. (2019). Identification of 5-substituted 2-acylaminothiazoles that activate Tat mediated transcription in HIV-1 latency models. *J. Med. Chem.* 62, 5148–5175. doi: 10.1021/acs.jmedchem.9b00462

Ott, M., Geyer, M., and Zhou, Q. (2011). The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host. Microb.* 10, 426–435. doi: 10.1016/j.chom.2011.11.002

Pasternak, A. O., and Berkhout, B. (2018). What do we measure when we measure cell-associated HIV RNA. *Retrovirology* 15:13.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450.

Ping, Y.-H., and Rana, T. M. (2001). DSIF and NELF Interact with RNA Polymerase II Elongation Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA Polymerase II and DSIF during Transcription Elongation *. *J. Biol. Chem.* 276, 12951–12958. doi: 10.1074/jbc.m006130200

Razooky, B. S., Pai, A., Aull, K., Rouzine, I. M., and Weinberger, L. S. (2015). A hardwired HIV latency program. *Cell* 160, 990–1001. doi: 10.1016/j.cell.2015.02.009

Rivas-Aravena, A., Ramdohr, P., Vallejos, M., Valiente-Echeverría, F., Dormoy-Raclet, V., Rodríguez, F., et al. (2009). The Elav-like protein HuR exerts translational control of viral internal ribosome entry sites. *Virology.* 392, 178–185. doi: 10.1016/j.virol.2009.06.050

Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C., et al. (2004). Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. *J. Biol. Chem.* 279, 29963–29973. doi: 10.1074/jbc.m404452200

Saleh, S., Wightman, F., Ramanayake, S., Alexander, M., Kumar, N., Khoury, G., et al. (2011). Expression and reactivation of HIV in a chemokine induced model of HIV latency in primary resting CD4+ T cells. *Retrovirology* 8:80. doi: 10.1186/1742-4690-8-80

Saliou, J. M., Bourgeois, C. F., Ayadi-Ben Mena, L., Ropers, D., Jacquenet, S., Marchand, V., et al. (2009). Role of RNA structure and protein factors in the control of HIV-1 splicing. *Front. Biosci.* 14, 2714–2729. doi: 10.2741/3408

Sarzotti-Kelsoe, M., Bailer, R. T., Turk, E., Lin, C. L., Bilska, M., Greene, K. M., et al. (2014). Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods.* 409, 131–146. doi: 10.1016/j.jim.2013.11.022

Schapira, M., Raaka, B. M., Das, S., Fan, L., Totrov, M., Zhou, Z., et al. (2003). Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proc. Natl. Acad. Sci. U S A.* 100, 7354–7359. doi: 10.1073/pnas.1131854100

Spina, C. A., Anderson, J., Archin, N. M., Bosque, A., Chan, J., Famiglietti, M., et al. (2013). An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. *PLoS Pathog.* 9:e1003834. doi: 10.1371/journal.ppat.1003834

Vallejos, M., Deforges, J., Plank, T. D., Letelier, A., Ramdohr, P., Abraham, C. G., et al. (2011). Activity of the human immunodeficiency virus type 1 cell cycle-dependent internal ribosomal entry site is modulated by IRES trans-acting factors. *Nucleic Acids Res.* 39, 6186–6200. doi: 10.1093/nar/gkr189

Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. (2000). Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* 408, 167–173. doi: 10.1038/35041507

Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P., and Schaffer, D. V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* 122, 169–182. doi: 10.1016/j.cell.2005.06.006

Yukl, S., Pillai, S., Li, P., Chang, K., Pasutti, W., Ahlgren, C., et al. (2009). Latently-infected CD4+ T cells are enriched for HIV-1 Tat variants with impaired transactivation activity. *Virology* 387, 98–108. doi: 10.1016/j.virol.2009.01.013

Yukl, S. A., Kaiser, P., Kim, P., Telwatte, S., Joshi, S. K., Vu, M., et al. (2018). HIV latency in isolated patient CD4+T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci. Transl. Med.* 10:eaa9927.

Zerbato, J. M., Khoury, G., Zhao, W., Gartner, M. J., Pascoe, R. D., Rhodes, A., et al. (2021). Multiply spliced HIV RNA is a predictive measure of virus production ex vivo and in vivo following reversal of HIV latency. *EBioMedicine* 65:103241. doi: 10.1016/j.ebiom.2021.103241

Zerbato, J. M., Purves, H. V., Lewin, S. R., and Rasmussen, T. A. (2019). Between a shock and a hard place: challenges and developments in HIV latency reversal. *Curr. Opin. Virol.* 38, 1–9. doi: 10.1016/j.coviro.2019.03.004

Zhou, H., Xu, M., Huang, Q., Gates, A. T., Zhang, X. D., Castle, J. C., et al. (2008). Genome-scale RNAi screen for host factors required for HIV replication. *Cell. Host. Microb.* 4, 495–504. doi: 10.1016/j.chom.2008.10.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 21 June 2021
doi: 10.3389/fgene.2021.689922

Deciphering Network Crosstalk: The Current Status and Potential of miRNA Regulatory Networks on the HSP40 Molecular Chaperone Network

Lion Budrass[1], *Richard P. Fahlman*[2,3] *and Sue-Ann Mok*[1,2]*

[1] Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada, [2] Department of Biochemistry, University of Alberta, Edmonton, AB, Canada, [3] Department of Oncology, University of Alberta, Edmonton, AB, Canada

Molecular chaperone networks fulfill complex roles in protein homeostasis and are essential for maintaining cell health. Hsp40s (commonly referred to as J-proteins) have critical roles in development and are associated with a variety of human diseases, yet little is known regarding the J-proteins with respect to the post-transcriptional mechanisms that regulate their expression. With relatively small alterations in their abundance and stoichiometry altering their activity, post-transcriptional regulation potentially has significant impact on the functions of J-proteins. MicroRNAs (miRNAs) are a large group of non-coding RNAs that form a complex regulatory network impacting gene expression. Here we review and investigate the current knowledge and potential intersection of miRNA regulatory networks with the J-Protein chaperone network. Analysis of datasets from the current version of TargetScan revealed a great number of predicted microRNAs targeting J-proteins compared to the limited reports of interactions to date. There are likely unstudied regulatory interactions that influence chaperone biology contained within our analysis. We go on to present some criteria for prioritizing candidate interactions including potential cooperative targeting of J-Proteins by multiple miRNAs. In summary, we offer a view on the scope of regulation of J-Proteins through miRNAs with the aim of guiding future investigations by identifying key regulatory nodes within these two complex cellular networks.

Keywords: J-proteins, Hsp40, microRNAs, chaperone, target prediction

OPEN ACCESS

Edited by:
Nehal Thakor,
University of Lethbridge, Canada

Reviewed by:
Elif Pala,
Sanko University, Turkey
Piyush Khandelia,
Birla Institute of Technology and Science, India

***Correspondence:**
Sue-Ann Mok
sueann@ualberta.ca

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *01 April 2021*
Accepted: *19 May 2021*
Published: *21 June 2021*

Citation:
Budrass L, Fahlman RP and Mok SA (2021) Deciphering Network Crosstalk: The Current Status and Potential of miRNA Regulatory Networks on the HSP40 Molecular Chaperone Network. Front. Genet. 12:689922. doi: 10.3389/fgene.2021.689922

INTRODUCTION

MicroRNA (miRNA) networks of gene regulation and molecular chaperone networks both consist of complex webs of interactions with broad implications in shaping the proteome. In both networks, the activity of individual molecules impacts many target or client molecules, resulting in broad regulation of protein homeostasis (proteostasis) (Hipp et al., 2014). Here, we discuss the intersection of these two major cellular networks. Specifically, we review the current reported miRNA regulatory interactions on the HSP40 family of chaperones, commonly referred to as J-proteins. Furthermore, we will consider the potential expanded network of interactions predicted

by bioinformatic analysis. For this study, we will focus on J-proteins which represent the largest and most diverse group of molecular chaperones (Kampinga and Craig, 2010).

MOLECULAR CHAPERONES

The cellular proteostasis network coordinates protein synthesis, degradation, and stress responses to ensure the correct folding, concentration, and localization of proteins to effectively carry out their cellular functions (Hipp et al., 2014). Molecular chaperones are an integral component of all proteostasis processes. They are a large group of ∼300 proteins (Brehme et al., 2014) that operate to recognize and deal with protein misfolding issues arising throughout the proteome. A multitude of functions is attributed to individual chaperone members with respect to facilitating protein folding, protein disaggregation, sequestration of aggregates, and directing misfolded proteins toward cellular degradation pathways (Kim et al., 2013; Kaushik and Cuervo, 2018; Nillegoda et al., 2018). Individual chaperones can often be grouped into distinct protein families e.g., J-proteins, Hsp60, Hsp70, Hsp90, sHsp. The most well-studied mechanistic aspect of chaperone activities is their ability to reversibly bind, and release unfolded and misfolded substrates (often termed clients) to promote their proper folding and prevent aggregation. Members of the major Heat shock protein 70 (Hsp70) and Heat shock protein 90 (Hsp90) families interact with hundreds of client proteins (Kerner et al., 2005; Taipale et al., 2012). In contrast, some chaperones members of the J-protein family show evidence of more discrete client binding profiles (Fotin et al., 2004; Gong et al., 2009; Kakkar et al., 2016a,b; Craig and Marszalek, 2017). However, chaperone interactions are not limited to client binding. Chaperones are known to work cooperatively with other chaperones and components of the proteostasis network, functioning as larger protein complexes (Taipale et al., 2014; Rizzolo et al., 2017; Freilich et al., 2018; Karunanayake and Page, 2021). For example, HSP70s have ATPase activities that allow for cycling between client binding and release—J-protein and Bag chaperone families strongly stimulate this cycling by promoting ATP hydrolysis and release, respectively (Freilich et al., 2018). Another factor contributing to the complex nature of chaperone activities is the different subcellular localization and expression levels of individual chaperones that directly influence client binding and stoichiometry of formed chaperone complexes (Craig and Marszalek, 2017).

The J-Protein Family of Chaperones. Potential for Regulation to Shape Cellular Proteostasis

The modular nature of chaperone complexes is proposed to contribute to the fine-tuning of chaperone recruitment and processing of specific clients in the cell. One large family of chaperones that facilitate modularity is the J-protein family. There are over 40 identified J-protein family members in humans (Kampinga and Craig, 2010), which are listed in **Table 1**. All members share a characteristic J-domain that facilitates modes of Hsp70 binding and ATP hydrolysis (Karzai and McMacken, 1996; Jiang et al., 2007; Kityk et al., 2018; Faust et al., 2020) and thus, are commonly referred to as J-proteins. However, beyond this shared domain, there is incredible diversity between J-protein members with respect to their structural and functional domains: individual J-proteins possess unique combinations of different client binding domains, localization signals, and enzymatic activities (Kampinga and Craig, 2010). J-proteins are roughly classified into three groups by structure. The A-class J-proteins share an overall domain structure similar to the *E.coli* J-proteins whereas the B-class J-proteins have only partially retained these domains. The remaining J-proteins are simply categorized as C-class (Kampinga and Craig, 2010).

Multiple J-proteins can compete for interactions with the same Hsp70. Therefore, varying the expression levels of individual J-proteins could, in turn, fine-tune the proteostasis of specific client subsets in a cell. Furthermore, there is evidence that changes in the balance of chaperone concentrations can have significant effects on chaperone complex formation and function (Kanelakis et al., 1999; Kundrat and Regan, 2010; Cabrera et al., 2019). For example, increasing the cellular concentrations of the mitochondrial J-protein, DNAJA3, interferes with the ability of the mitochondrial Hsp70 to bind substrates resulting in protein aggregation and mitochondria fragmentation (Lee B. et al., 2015). More broadly, J-proteins may be used to modulate chaperone network function to deal with the specific proteomes of different tissues (Uhlén et al., 2015). Indeed, individual J-proteins do show variations in tissue-specific expression (Hageman and Kampinga, 2009) and mutations in J-proteins are associated with highly tissue-specific diseases (Koutras and Braun, 2014; Sarparanta et al., 2020) such as early-childhood-onset recessive dilated cardiomyopathy and ataxia (Davey et al., 2006; Sparkes et al., 2007) and recessive distal hereditary motor neuropathy (Blumen et al., 2012).

Little is currently known regarding the mechanisms of how cells discretely modulate the expression of J-proteins in a tissue-specific manner or in response to stimuli or stress. Classically studied mechanisms of chaperone regulation are transcription factor activation (e.g., Heat Shock Factor 1) of a broad subset of chaperone gene targets during stress conditions such as heat shock (Zou et al., 1998; Anckar and Sistonen, 2011; Zheng et al., 2016) or ER-stress (Lee et al., 2003; Acosta-Alvear et al., 2007). Nonetheless, while some J-proteins exhibit stress-induced expression, most of the members of the family are constitutively expressed to cell or tissue specific levels (Zhao et al., 2008; Kakkar et al., 2012). In contrast to the stress response-activated transcription factors, even less is known about the post-transcriptional regulation of chaperone protein expression by other cellular factors such as microRNAs.

Considering the increasing understanding of J-proteins in protein folding-related diseases, such as the reported reduction of several J-proteins in Parkinson's Disease (Hasegawa et al., 2018), a more thorough understanding of the regulation of these proteins is warranted.

TABLE 1 | J-proteins and their validated miRNA targeting.

Hsp40	Targeting miRNA	References	Hsp40	Targeting miRNA	References
DNAJA1	-	-	DNAJC6	-	-
DNAJA2	-	-	DNAJC7	-	-
DNAJA3	-	-	DNAJC8	-	-
DNAJA4	-	-	DNAJC9	-	-
DNAJB1	miR-370, miR-543	Evert et al., 2018	DNAJC10	-	-
DNAJB2a	-	-	DNAJC11	-	-
DNAJB2b	-	-	DNAJC12	-	-
DNAJB4	-	-	DNAJC13	-	-
DNAJB5	-	-	DNAJC14	-	-
DNAJB6a	-	-	DNAJC15	-	-
DNAJB6b	-	-	DNAJC16	-	-
DNAJB7	-	-	DNAJC17	-	-
DNAJB8	-	-	DNAJC18	-	-
DNAJB9	miR-25-32-92-363-367 family	Wang et al., 2020	DNAJC19	-	-
DNAJB11	miR-29b	Beitzinger et al., 2007	DNAJC20	-	-
DNAJB12	miR-148-152 family	Ma et al., 2020	DNAJC21	-	-
DNAJB13	-	-	DNAJC22	-	-
DNAJB14	-	-	DNAJC23	-	-
DNAJC1	-	-	DNAJC24	-	-
DNAJC2	-	-	DNAJC25	-	-
DNAJC3	miR-200 family	Belgardt et al., 2015	DNAJC26	-	-
DNAJC4	-	-	DNAJC27	-	-
DNAJC5	-	-	DNAJC28	-	-
DNAJC5B	-	-	DNAJC29	-	-
DNAJC5G	-	-	DNAJC30	-	-

MicroRNA-MEDIATED SILENCING OF mRNA TRANSCRIPTS

Since the initial discovery of short non-coding RNAs regulating mRNA translation (Lee et al., 1993; Wightman et al., 1993; Reinhart et al., 2000), it has become apparent that microRNAs (miRNAs) function in the regulation of a large portion of the cellular transcriptome. It is estimated that each miRNA family targets on average more than 400 human mRNAs, and over half of human mRNAs have canonical conserved target sequences in their 3' untranslated regions (UTRs) (Friedman et al., 2009).

Endogenous miRNAs arise from long primary transcripts. A series of cellular processing events, depending on the transcript origin of the miRNA, produce the final mature miRNA in the Ago protein-containing silencing complexes (Bartel, 2018). The mature ∼22 nucleotides (nt) miRNA, guide the Ago protein-containing complexes to their target mRNAs via base pairing. In canonical targeting of mRNAs, this involves contiguous base pairing of the 5' seed region of the miRNA (nts 2-7) (Bartel, 2018). Base pairing with additional 3' nucleotides in the miRNA can occur but has been reported to have minimal effects on silencing efficacy (Grimson et al., 2007; Wee et al., 2012; Salomon et al., 2015). While several mechanisms have been reported regarding the silencing of mRNAs by miRNAs, the repression mechanism dependent on the TNRC6 adaptor protein family is the dominant mechanism in humans, as recently reviewed and discussed (Jonas and Izaurralde, 2015; Bartel, 2018). In this mechanism, TNRC6 family proteins bind several miRNA-ago complexes and therefore enhance the silencing of several miRNAs to one mRNA.

CURRENT STATE OF REPORTED miRNA REGULATORS OF J-PROTEINS

While there have been numerous reports in the literature describing correlations of J-Protein expression with miRNA expression, there are relatively few examples where the target sequence in the 3' UTR of the J-Protein mRNAs has been experimentally validated (see **Table 1**). Reports only describing anti-correlations in J-Protein expression with miRNA expression were omitted as indirect regulator interaction networks cannot be ruled out without further investigation. Some of these excluded reports include examples with compelling data where miRNA-dependent regulation is through the 3' UTR of a target mRNA, such as the down regulation of DNAJC6 upon miR-146b-5p expression (Kirchmeyer et al., 2018). Cases where the target sequence for a miRNA was not verified were also excluded (Mitra et al., 2012; Yang et al., 2014; Mycko et al., 2015). Increasing complexity of regulation of long non-coding RNAs (lncRNA) (Goodall and Wickramasinghe, 2021) and their interactions with miRNA regulatory networks (López-Urrutia et al., 2019)

could interfere in the miRNA-mRNA interactions in these cases. We will now briefly summarize verified microRNA targeting of J-proteins.

DNAJB1 is mostly known for the chimeric transcript it forms with PRKACA, which codes for the catalytic domain of protein kinase A in fibrolamellar hepatocellular carcinoma (Honeyman et al., 2014). It furthermore has been argued to be involved in p53-mediated apoptosis through degradation of PDCD5 (Cui et al., 2014). An investigation on a model for Spinocerebellar Ataxia Type 3 (SCA3) revealed a functional role for DNAJB1 in the clearance of mutant polyglutamine (polyQ) protein ataxin-3 aggregates. miR-370 and miR-543, which were both upregulated in SCA3 were shown to specifically target DNAJB1 mRNA. This study highlights possible disease implications miRNAs could have through their interactions with chaperones (Evert et al., 2018).

DNAJC3 is an ER-localized J-protein and co-chaperone to HSPA5. A loss-of-function mutation leads to diabetes mellitus and multisystemic neurodegeneration (Synofzik et al., 2015) and in mice, DNAJC3 knockout mice had a phenotype of partial loss of pancreatic beta-cells (Ladiges et al., 2005). In mouse models of beta cell stress and obesity, miRNA-200 family was found to have a role in promoting the apoptosis of pancreatic beta cells (Belgardt et al., 2015). Transcriptome analysis of miRNA-200 targets in mice revealed DNAJC3 which was then validated as a direct target.

DNAJB9 has recently been shown to inhibit p53-induced apoptosis (Lee H. J. et al., 2015). In models of chemotherapy resistance in acute myeloid leukemia, Wang et al. (2020) identified a regulatory network that involves the direct downregulation of DNAJB9 by miR-32. While miR-32 inhibited DNAJB9, it was in turn modulated by the lncRNA, small nucleolar RNA host gene 5 (SNHG5), creating an axis of control between DNAJB9-miR-32-SNHG5, possibly causing chemotherapy resistance. An analogous regulatory network has also been reported for DNAJB12, an ER-related J-protein (Ma et al., 2020), where the direct targeting of anti-apoptotic DNAJB12 by miR-152-3p is negatively modulated by the lncRNA HCG18 in gastric cancer models. Both studies highlight the complexity of gene regulation through the miRNA network, including the involvement of factors such as lncRNAs.

Additional evidence of miRNA targeting of J-Proteins can be taken where miRNA-Argonaute protein complexes have been identified to associate with the 3' UTRs. As an example, immunoprecipitation experiments revealed DNAJB11 as a component of miR-29-Ago complexes (Beitzinger et al., 2007). While this is strongly suggestive of a regulatory interaction, the authors indicate that not all the interactions identified lead to biological downregulation upon validation.

MicroRNA TARGET PREDICTION OF J-PROTEINS

While this list of validated regulatory interactions of miRNAs with J-proteins is quite limited, the correlative data in the literature suggests there are significantly more interactions awaiting validation.

To obtain a more global perspective on the potential miRNA network of interactions on the J-Proteins we performed an *in silico* analysis of miRNA target predictions. To this end, we utilized the most recent version of Targetscan (version 7.2) which identifies predicted canonical mRNA target sequences with 7–8 nt stretches of complementarity to the miRNA seed sequence (Agarwal et al., 2015).

Within the context of miRNA target prediction, both evolutionary conservation aspects of the miRNAs themselves and particular putative miRNA target sequences within a mRNA exist. With respect to our analysis, there are miRNAs and miRNA families that are broadly conserved among vertebrates and miRNAs conserved among mammals (Bartel, 2018). While there may be more recently evolved miRNAs, it has been proposed that many of these have too few targets under selective pressure to enable target predictions with any confidence (Friedman et al., 2009). A major caveat of any predicted miRNA target analyses is that not all mRNAs with 7–8 nt complimentary sites to the miRNA seed sequence exhibit regulation by that miRNA (Baek et al., 2008; Selbach et al., 2008).

An initial analysis for putative miRNA targets of the J-protein family with cumulative context scores of ≤ -0.1 (Agarwal et al., 2015) as a first-pass threshold, results in 1,337 potential miRNA target sequences for 212 different miRNAs or miRNA families. This minimal criterion yields an unwieldy number of potential sites for experimental validation and most likely consists of a high proportion of false identifications. As mRNAs with evolutionary conserved potential miRNA target sequences in their 3' UTRs exhibit a higher probability of responding to the activity of a miRNA (Baek et al., 2008), strategies to identify more likely miRNA target sequences include choosing sites that exhibit conservation (Bartel, 2009; Friedman et al., 2009). As a result, we applied the criteria for evolutionarily conserved mRNA target sequences for both broadly conserved miRNAs among vertebrates and miRNAs conserved among mammals. This additional parameter reduces the likelihood of false positive identifications in the dataset, while including many possible cross-species interactions. Thus, although our additional analysis criteria decreases our false negative rate of prediction it will also miss some potentially biologically relevant chaperone-miRNA interactions that are not conserved among species. It should be noted that this new criteria also leads to the exclusion of target predictions for some highly probable targets. One example being the targeting of DNAJB5 by miR-21 (Lampis et al., 2018), where miR-21 expression was demonstrated to lead to 3' UTR dependent regulation of DNAJB5. This miR-21:DNAJB5 interaction was excluded from our presentation as a valid target as the predicted target sequence was not experimentally verified. With a selection criterion for conserved putative mRNA target sites for conserved miRNAs, Targetscan identifies 164 and 72 predicted targets respectively for either broadly conserved miRNAs or miRNAs conserved among mammals. This level of analysis reveals significant variations between the members of the J-protein family. As seen in **Figure 1**, nine mRNAs have no predicted canonical miRNA target sequences, such as DNAJA3,

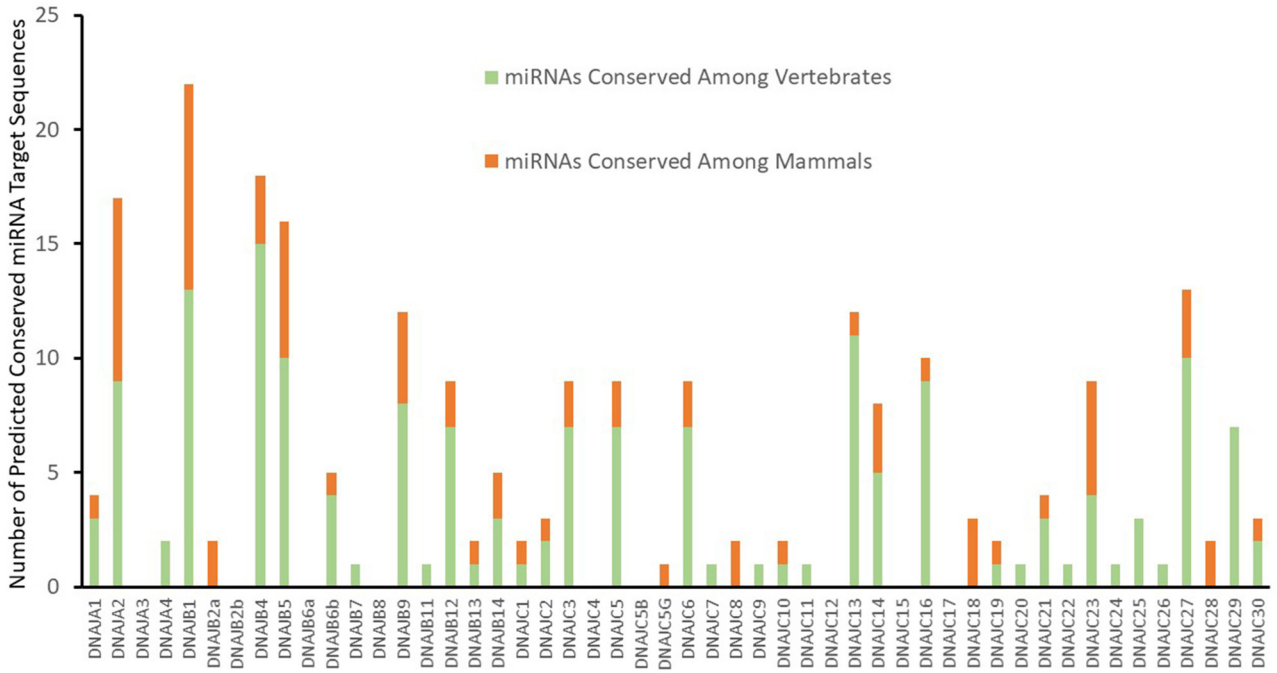


FIGURE 1 | Predicted targeting by conserved miRNA Families. Predicted conserved J-Protein target sites of miRNAs from TargetScan (Agarwal et al., 2015) with a context score of <−0.1. The number of predicted conserved miRNA targets for each of the J-proteins are indicated. The proportion of predicted targets by miRNAs and miRNA families broadly conserved throughout vertebrates or only within mammals are also indicated.

DNAJB8, and DNAJC4. On the opposite side of the spectrum, several J-Protein mRNAs contain a high number of predicted conserved miRNA targets. The mRNAs for DNAJA2, DNAJB1, DNAJB4, and DNAJB5 each have >15 predicted conserved miRNA target sequences within their 3' UTRs. Among these, only two of the predicted target sequences for DNAJB1 have been experimentally verified (Evert et al., 2018).

The specific predicted conserved target interactions for all the J-Protein members by broadly conserved miRNAs or those conserved among mammals are shown in **Figures 2**, **3** respectively. With the reported known interactions listed in **Table 1** highlighted in these figures, it is apparent how few of these predicted 234 interactions have been investigated. This therefore emphasizes an area of research in chaperone biology that is primed for further investigation.

Predicting Strong miRNA Candidates of J-Protein Regulation

Despite the large number of predicted potential interactions, many predicted conserved miRNA target sequences are known to not significantly regulate expression levels of their targets in cells (Baek et al., 2008). Proteomic analysis of the same miRNAs in different cellular backgrounds reveals cell-specific differences (Ludwig et al., 2016; Piragasam et al., 2020) and changes in protein abundances that may counter interactions shown by miRNA target predictions (Piragasam et al., 2020). These proteomic analyses are holistic in that they reveal both direct and indirect impacts on protein abundance through miRNAs. Results that counter predictions can thus not be used to rule out potential interactions. Nonetheless, these types of analyses may provide some insight and guidance for predicting interactions that might lead to stronger regulation of a given target which can then be prioritized for investigation. In the next few paragraphs, we examine two potential miRNA mechanisms that may enhance their effects on target transcripts and highlight potential examples in our J-protein-miRNA dataset.

Multiple miRNA Target Sites

The degree of regulation of mRNAs by miRNAs depends on the identities and abundance of the particular miRNAs in the cell, the number of target mRNA sites in a cell, as well as the specific binding efficacy for a given miRNA target site. For each given miRNA:mRNA interaction, the regulation is typically modest with repression being <50% (Baek et al., 2008; Selbach et al., 2008). However, enhanced repression is often observed when multiple miRNA target sites are within the same 3' UTR as these effects are typically additive. The predictions presented in **Figures 2**, **3** reveal several individual miRNAs that are predicted to have multiple mRNA target sequences within a given 3' UTR. These include the three predicted target sites for miR-23-3p within the 3' UTR of DNAJC6 and an additional 12 miRNA:mRNA interactions with two predicted miRNA target sites. Of the 12 predictions with two predicted target sites in an mRNA for the same miRNA, the targeting of DNAJB1 by miR-370 (Evert et al., 2018) and DNAJC3 by the miRNA-200 family (Belgardt et al., 2015) have been validated.

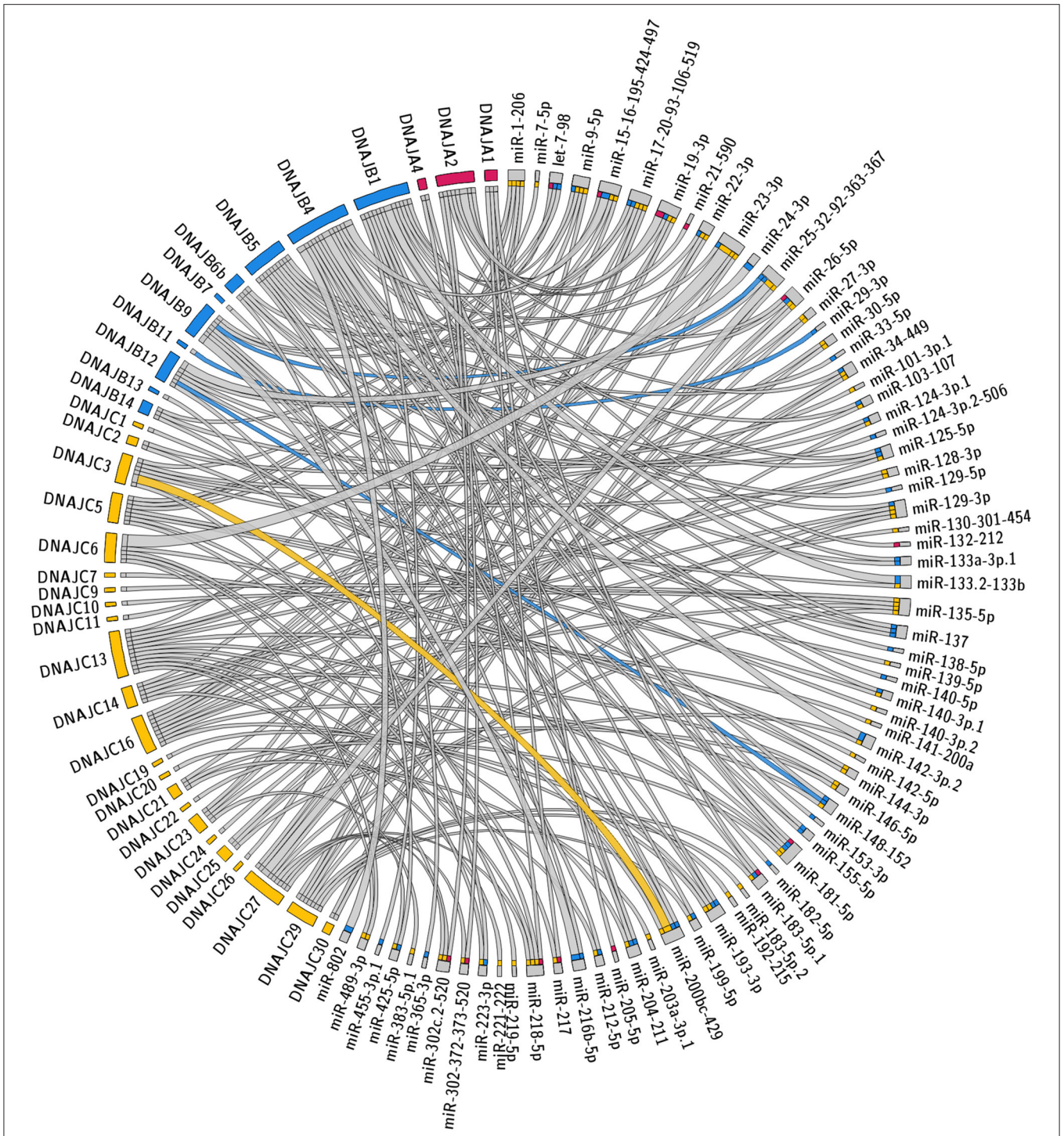


FIGURE 2 | Predicted miRNA targets for miRNAs broadly conserved in vertebrates. Interaction Circos plot (Krzywinski et al., 2009) for TargetScan depicting targeting of J-Proteins by miRNA families that are broadly conserved throughout vertebrates. For each J-Protein, miRNAs with a context score of <−0.1 were extracted from TargerScan (Agarwal et al., 2015). Unconserved target sequences for miRNAs that were not conserved were excluded from the analysis. A colorized line indicates miRNA targeting that has been experimentally verified and the width of the bridging lines is proportional to the number of predicted target sequences.

Cooperative miRNA Targeting

Complex patterns of miRNA expression exist in different cell and tissue types (Landgraf et al., 2007; Chaulk et al., 2016). The intersection of both miRNA and mRNA expression patterns can yield increasingly complex combinatorial regulatory networks of regulation. This regulation becomes even more complex when considering that multiple miRNAs can simultaneously target the same 3' UTR of an mRNA and lead to differential outcomes

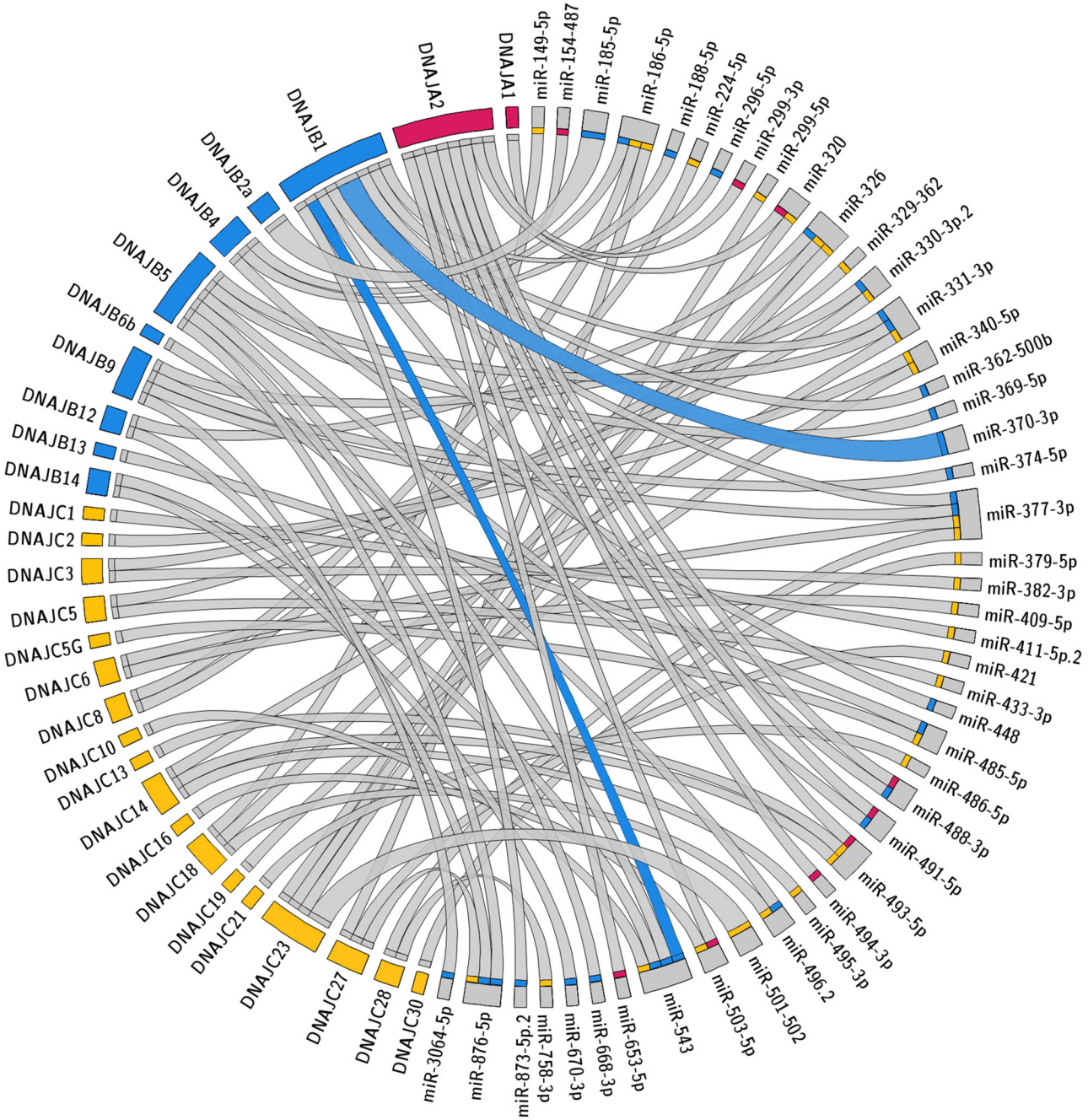


FIGURE 3 | Predicted miRNA targets for miRNAs conserved within mammals. Interaction Circos plot (Krzywinski et al., 2009) for TargetScan depicting miRNA targeting of the J-Protein by miRNA families that are conserved throughout mammals. For each J-Protein, miRNAs with a context score of <−0.1 were extracted from TargerScan (Agarwal et al., 2015). Unconserved target sequences for miRNAs that were not conserved were excluded from the analysis. A colorized line indicates miRNA targeting that has been experimentally verified and the width of the bridging lines is proportional to the number of predicted target sequences.

for the same transcript depending on a cell's given miRNA signature pattern.

The potential for multiple different miRNAs targeting the same J-Protein is summarized in **Figures 2**, **3**. Here it should be noted that the co-targeting of DNAJB1 by miR-543 and miR-370 has been documented (Evert et al., 2018). There is another level of complexity regarding co-targeting that is not revealed in the presentation of the figures, that is, their spatial proximity. In some cases, miRNAs are reported to act cooperatively on the same mRNA, specifically those with target sequences within 8–40 nt of each other (Grimson et al., 2007; Sætrom et al., 2007). This is a result of the TNRC6 proteins being able to associate with

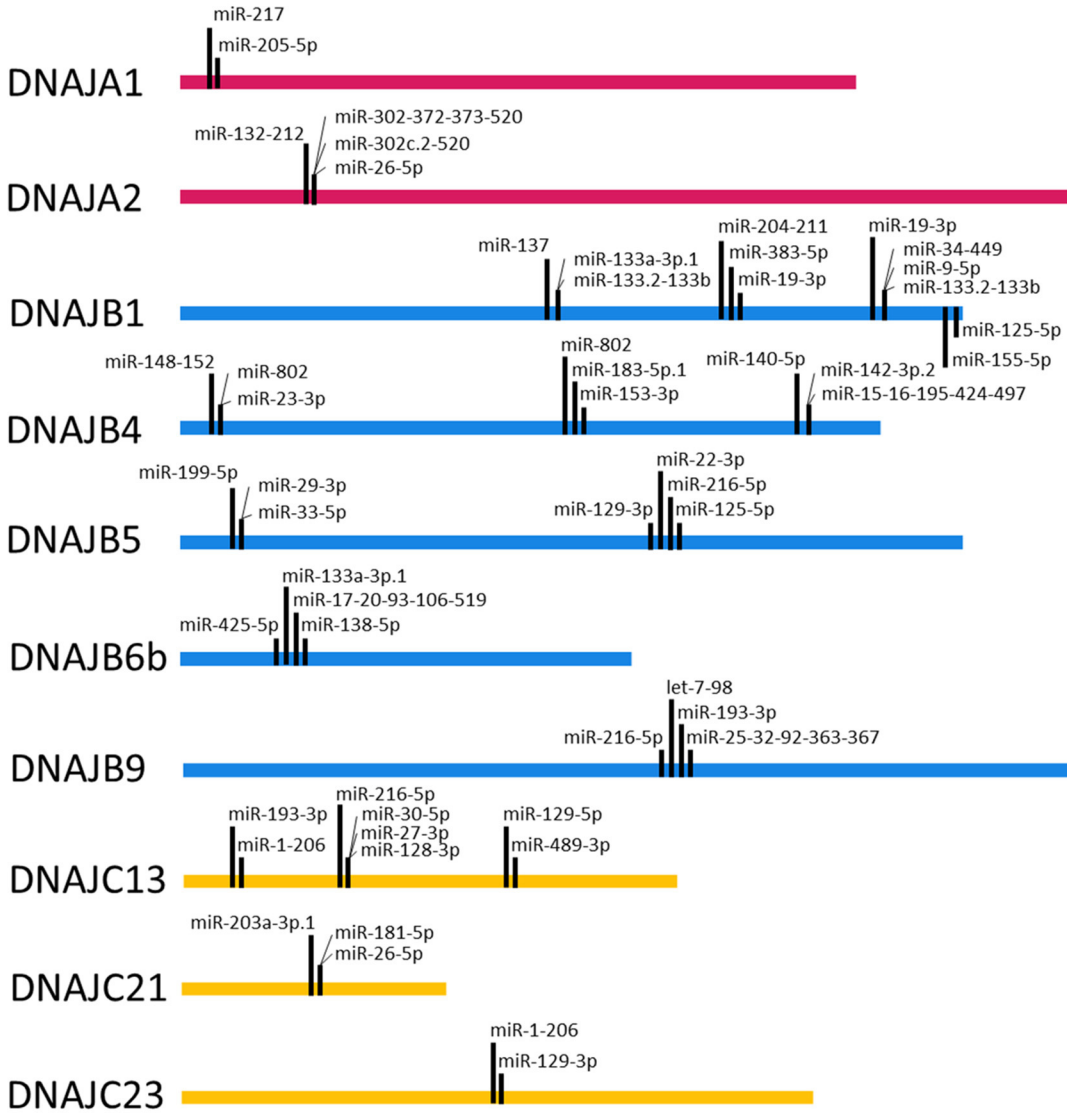


FIGURE 4 | Target Site Proximity. For each J-Protein, target sites were identified on TargetScan (Agarwal et al., 2015) with a context score of <-0.1 and then manually scanned for the occurrence of miRNA seed sequences <40 nucleotides apart. The locations of identified close proximity target sites are visualized on the ribbon corresponding to the nucleotide length of the 3' UTR for each given J-Protein. Only miRNA families that are broadly conserved through vertebrates were included in the analysis.

multiple Ago protein complexes simultaneously (Briskin et al., 2020).

To query the potential for cooperative interactions between the predicted miRNA target sequences in the 3' UTRs of the J-protein mRNAs, target sequences within 8–40 nts of each other were identified. **Figure 4** depicts the miRNA target sequences that meet this criterion for the miRNA families broadly conserved among vertebrates. Intriguingly, the analysis reveals a potential for another level of complexity with regard to miRNA regulation. While the analysis identified multiple examples of potential miRNA target sequences within 8–40 nts, such as the predicted miR-217 and miR-205-5p target sequences in the 3' UTR of the DNAJA1 mRNA, there are also potential combinations of mutual exclusivity. For example, as shown in **Figure 4**, the DNAJB4 mRNA has predicted miRNA target sequences for the miR-148-152 family within 40 nts of the predicted target sequences for miR-802 and miR-23-3p. As the predicted target sequences for miR-802 and miR-23-3p partially overlap, then if both sequences exhibit bonafide mRNA targeting in cells, they would have to be mutually exclusive in their targeting by these miRNAs. This leads to the prediction that miR-148-152 could act cooperatively with either miR-802 and miR-23-3p but that these two miRNAs could not bind to the same DNAJB4 mRNA to regulate its expression. While this form of potential regulation leads to numerous instances of a Boolean logic type of regulation behavior, it is currently unclear whether this behavior would be

recapitulated in cells and would be highly dependent on miRNA complex concentrations.

CONCLUSION

Perhaps as a result of the enormous inherent complexities of both the chaperone networks and miRNA regulatory networks, there has been relatively little reported work validating their intersection of regulation. Our analysis offers starting points for the exploration of miRNA and J-protein interactions. With the reported linkages and interest in both miRNAs and J-proteins in human diseases such as neurodegeneration and cancer, there are ample possibilities that we have outlined for future investigations into the interplay of these systems.

AUTHOR CONTRIBUTIONS

LB: drafted the manuscript and collected literature and data. RF: data analysis and edited the manuscript. SAM: designed the content of the article and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by NSERC Discovery Grant program funding to SAM. LB was supported by fellowships from the Neuroscience and Mental Health Institute (University of Alberta) and SynAD (University of Alberta).

REFERENCES

Acosta-Alvear, D., Zhou, Y., Blais, A., Tsikitis, M., Lents, N. H., Arias, C., et al. (2007). XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Mol. Cell* 27, 53–66. doi: 10.1016/j.molcel.2007.06.011

Agarwal, V., Bell, G. W., Nam, J., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005. doi: 10.7554/eLife.05005

Anckar, J., and Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* 80, 1089–1115. doi: 10.1146/annurev-biochem-060809-095203

Baek, D., Shin, C., Gygi, S. P., Camargo, F. D., Villien, J., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71. doi: 10.1038/nature07242

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Bartel, D. P. (2018). Metazoan microRNAs. *Cell* 173, 20–51. doi: 10.1016/j.cell.2018.03.006

Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E., and Meister, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* 4, 76–84. doi: 10.4161/rna.4.2.4640

Belgardt, B., Ahmed, K., Spranger, M., Latreille, M., Denzler, R., Kondratiuk, N., et al. (2015). The microRNA-200 family regulates pancreatic beta cell survival in type 2 diabetes. *Nat. Med.* 21, 619–627. doi: 10.1038/nm.3862

Blumen, S. C., Astord, S., Robin, V., Vignaud, L., Toumi, N., Cieslik, A., et al. (2012). A rare recessive distal hereditary motor neuropathy with HSJ1 chaperone mutation. *Ann. Neurol.* 71, 509–519. doi: 10.1002/ana.22684

Brehme, M., Voisine, C., Rolland, T., Wachi, S., Soper, J., Zhu, Y., et al. (2014). A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* 9, 1135–1150. doi: 10.1016/j.celrep.2014.09.042

Briskin, D., Wang, P. Y., and Bartel, D. P. (2020). The biochemical basis for the cooperative action of microRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 117, 17764–17774. doi: 10.1073/pnas.1920404117

Cabrera, Y., Dublang, L., Fernández-Higuero, J. A., Albesa-Jové, D., Lucas, M., Viguera, A. R., et al. (2019). Regulation of human Hsc70 ATPase and chaperone activities by Apg2: role of the acidic subdomain. *J. Mol. Biol.* 431, 444–461. doi: 10.1016/j.jmb.2018.11.026

Chaulk, S. G., Ebhardt, H. A., and Fahlman, R. P. (2016). Correlations of microRNA:microRNA expression patterns reveal insights into microRNA clusters and global microRNA expression patterns. *Mol. BioSyst.* 12, 110–119. doi: 10.1039/C5MB00415B

Craig, E. A., and Marszalek, J. (2017). How do J-proteins get Hsp70 to do so many different things? *Trends Biochem. Sci.* 42, 355–368. doi: 10.1016/j.tibs.2017.02.007

Cui, X., Choi, H., Choi, Y., Park, S., Sung, G., Lee, Y., et al. (2014). DNAJB1 destabilizes PDCD5 to suppress p53-mediated apoptosis. *Cancer Lett.* 357, 307–315. doi: 10.1016/j.canlet.2014.11.041

Davey, K. M., Parboosingh, J. S., McLeod, D. R., Chan, A., Casey, R., Ferreira, P., et al. (2006). Mutation of DNAJC19, a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition. *J. Med. Genet.* 43, 385–393. doi: 10.1136/jmg.2005.036657

Evert, B. O., Nalavade, R., Jungverdorben, J., Matthes, F., Weber, S., Rajput, A., et al. (2018). Upregulation of miR-370 and miR-543 is associated with reduced expression of heat shock protein 40 in spinocerebellar ataxia type 3. *PLoS ONE* 13:e0201794. doi: 10.1371/journal.pone.0201794

Faust, O., Abayev-Avraham, M., Wentink, A. S., Maurer, M., Nillegoda, N. B., London, N., et al. (2020). HSP40 proteins use class-specific regulation to drive HSP70 functional diversity. *Nature* 587, 489–494. doi: 10.1038/s41586-020-2906-4

Fotin, A., Cheng, Y., Grigorieff, N., Walz, T., Harrison, S. C., and Kirchhausen, T. (2004). Structure of an auxilin-bound clathrin coat and its implications for the mechanism of uncoating. *Nature* 432, 649–653. doi: 10.1038/nature03078

Freilich, R., Arhar, T., Abrams, J. L., and Gestwicki, J. E. (2018). Protein–protein interactions in the molecular chaperone network. *Acc. Chem. Res.* 51, 940–949. doi: 10.1021/acs.accounts.8b00036

Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105. doi: 10.1101/gr.082701.108

Gong, Y., Zhang, Z., Kakihara, Y., Emili, A., Krogan, N., Houry, W. A., et al. (2009). An atlas of chaperone-protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. *Mol. Syst. Biol.* 5:275. doi: 10.1038/msb.2009.26

Goodall, G. J., and Wickramasinghe, V. O. (2021). RNA in cancer. *Nat. Rev. Cancer* 21, 22–36. doi: 10.1038/s41568-020-00306-0

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105. doi: 10.1016/j.molcel.2007.06.017

Hageman, J., and Kampinga, H. H. (2009). Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. *Cell Stress Chaperones* 14, 1–21. doi: 10.1007/s12192-008-0060-2

Hasegawa, T., Yoshida, S., Sugeno, N., Kobayashi, J., and Aoki, M. (2018). DnaJ/Hsp40 family and parkinson's disease. *Front. Neurosci.* 11:743. doi: 10.3389/fnins.2017.00743

Hipp, M. S., Park, S., and Hartl, F. U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* 24, 506–514. doi: 10.1016/j.tcb.2014.05.003

Honeyman, J. N., Simon, E. P., Robine, N., Chiaroni-Clarke, R., Darcy, D. G., Lim, I. I. P., et al. (2014). Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science* 343, 1010–1014. doi: 10.1126/science.1249484

Jiang, J., Maes, E. G., Taylor, A. B., Wang, L., Hinck, A. P., Lafer, E. M., et al. (2007). Structural basis of J cochaperone binding and regulation of Hsp70. *Mol. Cell* 28, 422–433. doi: 10.1016/j.molcel.2007.08.022

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433. doi: 10.1038/nrg3965

Kakkar, V., Kuiper, E. F. E., Pandey, A., Braakman, I., and Kampinga, H. H. (2016a). Versatile members of the DNAJ family show Hsp70 dependent anti-aggregation activity on RING1 mutant parkin C289G. *Sci. Rep.* 6:34830. doi: 10.1038/srep34830

Kakkar, V., Månsson, C., de Mattos, E., Bergink, S., van der Zwaag, M., van Waarde, M. W. H., et al. (2016b). The S/T-rich motif in the DNAJB6 chaperone delays polyglutamine aggregation and the onset of disease in a mouse model. *Mol. Cell* 62, 272–283. doi: 10.1016/j.molcel.2016.03.017

Kakkar, V., Prins, L. C. B., and Kampinga, H. H. (2012). DNAJ Proteins and protein aggregation diseases. *Curr. Top. Med. Chem.* 12, 2479–2490. doi: 10.2174/1568026611212220004

Kampinga, H. H., and Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* 11, 579–592. doi: 10.1038/nrm2941

Kanelakis, K. C., Morishima, Y., Dittmar, K. D., Galigniana, M. D., Takayama, S., Reed, J. C., et al. (1999). Differential effects of the hsp70-binding protein BAG-1 on glucocorticoid receptor folding by the hsp90-based chaperone machinery. *J. Biol. Chem.* 274, 34134–34140. doi: 10.1074/jbc.274.48.34134

Karunanayake, C., and Page, R. C. (2021). Cytosolic protein quality control machinery: Interactions of Hsp70 with a network of co-chaperones and substrates. *Exp. Biol. Med.* doi: 10.1177/1535370221999812. [Epub ahead of print].

Karzai, A. W., and McMacken, R. (1996). A bipartite signaling mechanism involved in dnaj-mediated activation of the *Escherichia coli* DnaK protein. *J. Biol. Chem.* 271, 11236–11246. doi: 10.1074/jbc.271.19.11236

Kaushik, S., and Cuervo, A. M. (2018). The coming of age of chaperone-mediated autophagy. *Nat. Rev. Mol. Cell Biol.* 19, 365–381. doi: 10.1038/s41580-018-0001-6

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H., Stines, A. P., et al. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122, 209–220. doi: 10.1016/j.cell.2005.05.028

Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., and Hartl, F. U. (2013). Molecular chaperone functions in protein folding and proteostasis. *Ann. Rev. Biochem.* 82, 323–355. doi: 10.1146/annurev-biochem-060208-092442

Kirchmeyer, M., Servais, F. A., Hamdorf, M., Nazarov, P. V., Ginolhac, A., Halder, R., et al. (2018). Cytokine-mediated modulation of the hepatic miRNome: miR-146b-5p is an IL-6-inducible miRNA with multiple targets. *J. Leukoc. Biol.* 104, 987–1002. doi: 10.1002/JLB.MA1217-499RR

Kityk, R., Kopp, J., and Mayer, M. P. (2018). Molecular mechanism of J-domain-triggered ATP hydrolysis by Hsp70 chaperones. *Mol. Cell* 69, 227–237.e4. doi: 10.1016/j.molcel.2017.12.003

Koutras, C., and Braun, J. E. A. (2014). J protein mutations and resulting proteostasis collapse. *Front. Cell. Neurosci.* 8:191. doi: 10.3389/fncel.2014.00191

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kundrat, L., and Regan, L. (2010). Balance between folding and degradation for Hsp90-dependent client proteins: a key role for CHIP. *Biochemistry* 49, 7428–7438. doi: 10.1021/bi100386w

Ladiges, W. C., Knoblaugh, S. E., Morton, J. F., Korth, M. J., Sopher, B. L., Baskin, C. R., et al. (2005). Pancreatic β-cell failure and diabetes in mice with a deletion mutation of the endoplasmic reticulum molecular chaperone gene P58IPK. *Diabetes* 54, 1074–1081. doi: 10.2337/diabetes.54.4.1074

Lampis, A., Carotenuto, P., Vlachogiannis, G., Cascione, L., Hedayat, S., Burke, R., et al. (2018). MIR21 drives resistance to heat shock protein 90 inhibition in cholangiocarcinoma. *Gastroenterology* 154, 1066–1079.e5. doi: 10.1053/j.gastro.2017.10.043

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129, 1401–1414. doi: 10.1016/j.cell.2007.04.040

Lee, A., Iwakoshi, N. N., and Glimcher, L. H. (2003). XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol. Cell. Biol.* 23, 7448–7459. doi: 10.1128/MCB.23.21.7448-7459.2003

Lee, B., Ahn, Y., Kang, S., Park, Y., Jeon, Y., Rho, J. M., et al. (2015). Stoichiometric expression of mtHsp40 and mtHsp70 modulates mitochondrial morphology and cristae structure via Opa1L cleavage. *Mol. Biol. Cell* 26, 2156–2167. doi: 10.1091/mbc.E14-02-0762

Lee, H. J., Kim, J. M., Kim, K. H., Heo, J. I., Kwak, S. J., and Han, J. A. (2015). Genotoxic stress/p53-induced DNAJB9 inhibits the pro-apoptotic function of p53. *Cell Death Differ.* 22, 86–95. doi: 10.1038/cdd.2014.116

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

López-Urrutia, E., Bustamante Montes, L. P., Ladrón de Guevara Cervantes, D., Pérez-Plasencia, C., and Campos-Parra, A. D. (2019). Crosstalk between long non-coding RNAs, micro-RNAs and mRNAs: deciphering molecular mechanisms of master regulators in cancer. *Front. Oncol.* 9:669. doi: 10.3389/fonc.2019.00669

Ludwig, K. R., Dahl, R., and Hummon, A. B. (2016). Evaluation of the mirn23a cluster through an iTRAQ-based quantitative proteomic approach. *J. Proteome Res.* 15, 1497–1505. doi: 10.1021/acs.jproteome.5b01101

Ma, P., Li, L., Liu, F., and Zhao, Q. (2020). HNF1A-induced lncRNA HCG18 facilitates gastric cancer progression by upregulating DNAJB12 via miR-152-3p. *OncoTargets Ther.* 13, 7641–7652. doi: 10.2147/OTT.S253391

Mitra, A., Rostas, J. W., Dyess, D. L., Shevde, L. A., and Samant, R. S. (2012). Micro-RNA-632 downregulates DNAJB6 in breast cancer. *Lab. Invest.* 92, 1310–1317. doi: 10.1038/labinvest.2012.87

Mycko, M. P., Cichalewska, M., Cwiklinska, H., and Selmaj, K. W. (2015). miR-155-3p drives the development of autoimmune demyelination by regulation of heat shock protein 40. *J. Neurosci.* 35, 16504–16515. doi: 10.1523/JNEUROSCI.2830-15.2015

Nillegoda, N. B., Wentink, A. S., and Bukau, B. (2018). Protein disaggregation in multicellular organisms. *Trends Biochem. Sci.* 43, 285–300. doi: 10.1016/j.tibs.2018.02.003

Piragasam, R. S., Hussain, S. F., Chaulk, S. G., Siddiqi, Z. A., and Fahlman, R. P. (2020). Label-free proteomic analysis reveals large dynamic changes to the cellular proteome upon expression of the miRNA-23a-27a-24-2 microRNA cluster. *Biochem. Cell Biol.* 98, 61–69. doi: 10.1139/bcb-2019-0014

Reinhart, B. J., Pasquinelli, A. E., Bettinger, J. C., Horvitz, H. R., Basson, M., Ruvkun, G., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906. doi: 10.1038/35002607

Rizzolo, K., Huen, J., Kumar, A., Phanse, S., Vlasblom, J., Kakihara, Y., et al. (2017). Features of the chaperone cellular network revealed through systematic interaction mapping. *Cell Rep.* 20, 2735–2748. doi: 10.1016/j.celrep.2017.08.074

Sætrom, P., Heale, B. S. E., Snøve, O., Aagaard, L., Alluin, J., and Rossi, J. J. (2007). Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* 35, 2333–2342. doi: 10.1093/nar/gkm133

Salomon, W., Jolly, S., Moore, M., Zamore, P., and Serebrov, V. (2015). Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides. *Cell* 162, 84–95. doi: 10.1016/j.cell.2015.06.029

Sarparanta, J., Jonson, P. H., Kawan, S., and Udd, B. (2020). Neuromuscular diseases due to chaperone mutations: a review and some new results. *Int. J. Mol. Sci.* 21:1409. doi: 10.3390/ijms21041409

Selbach, M., Fang, Z., Schwanhäusser, B., Khanin, R., Thierfelder, N., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63. doi: 10.1038/nature07228

Sparkes, R., Patton, D., and Bernier, F. (2007). Cardiac features of a novel autosomal recessive dilated cardiomyopathic syndrome due to defective importation of mitochondrial protein. *Cardiol. Young* 17, 215–217. doi: 10.1017/S1047951107000042

Synofzik, M., Haack, T., Kopajtich, R., Gorza, M., Rapaport, D., Greiner, M., et al. (2015). Absence of BiP co-chaperone DNAJC3 causes diabetes mellitus and multisystemic neurodegeneration. *Am. J. Hum. Genet.* 96:514. doi: 10.1016/j.ajhg.2015.02.003

Taipale, M., Krykbaeva, I., Koeva, M., Kayatekin, C., Westover, K., Karras, G., et al. (2012). Quantitative analysis of Hsp90-client interactions reveals principles of substrate recognition. *Cell* 150, 987–1001. doi: 10.1016/j.cell.2012.06.047

Taipale, M., Tucker, G., Peng, J., Krykbaeva, I., Lin, Z., Larsen, B., et al. (2014). A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* 158, 434–448. doi: 10.1016/j.cell.2014.05.039

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419

Wang, D., Zeng, T., Lin, Z., Yan, L., Wang, F., Tang, L., et al. (2020). Long non-coding RNA SNHG5 regulates chemotherapy resistance through the miR-32/DNAJB9 axis in acute myeloid leukemia. *Biomed. Pharmacother.* 123:109802. doi: 10.1016/j.biopha.2019.109802

Wee, L., Flores-Jasso, C., Salomon, W., and Zamore, P. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055–1067. doi: 10.1016/j.cell.2012.10.036

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862. doi: 10.1016/0092-8674(93)90530-4

Yang, Z., He, M., Wang, K., Sun, G., Tang, L., and Xu, Z. (2014). Tumor suppressive microRNA-193b promotes breast cancer progression via targeting DNAJC13 and RAB22A. *Int. J. Clin. Exp. Pathol.* 7, 7563–7570.

Zhao, X., Braun, A., and Braun, J. (2008). Biological roles of neural J proteins. *Cell Mol. Life Sci.* 65, 2385–2396. doi: 10.1007/s00018-008-8089-z

Zheng, X., Krakowiak, J., Patel, N., Beyzavi, A., Ezike, J., Khalil, A. S., et al. (2016). Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. *eLife* 5:e18638. doi: 10.7554/eLife.18638

Zou, J., Guo, Y., Guettouche, T., Smith, D. F., and Voellmy, R. (1998). Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* 94, 471–480. doi: 10.1016/S0092-8674(00)81588-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 20 July 2021
doi: 10.3389/fgene.2021.700744

At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression

Maureen V. Akinyi and Mikko J. Frilander*

Institute of Biotechnology/Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland

Many eukaryotic species contain two separate molecular machineries for removing non-coding intron sequences from pre-mRNA molecules. The majority of introns (more than 99.5% in humans) are recognized and excised by the major spliceosome, which utilizes relatively poorly conserved sequence elements at the 5′ and 3′ ends of the intron that are used for intron recognition and in subsequent catalysis. In contrast, the minor spliceosome targets a rare group of introns (approximately 0.5% in humans) with highly conserved sequences at the 5′ and 3′ ends of the intron. Minor introns coexist in the same genes with major introns and while the two intron types are spliced by separate spliceosomes, the two splicing machineries can interact with one another to shape mRNA processing events in genes containing minor introns. Here, we review known cooperative and competitive interactions between the two spliceosomes and discuss the mechanistic basis of the spliceosome crosstalk, its regulatory significance, and impact on spliceosome diseases.

Keywords: RNA processing, mRNA splicing, minor spliceosome, major spliceosome, exon definition, cryptic splicing, minor spliceosome disease

OPEN ACCESS

Edited by:
Jane E. A. Reid,
Australian National University, Australia

Reviewed by:
Raymond O'Keefe,
The University of Manchester, United Kingdom
Ihab Younis,
Carnegie Mellon University in Qatar, Qatar

****Correspondence:***
Maureen V. Akinyi
maureen.akinyi@helsinki.fi

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *26 April 2021*
Accepted: *28 June 2021*
Published: *20 July 2021*

Citation:
Akinyi MV and Frilander MJ (2021) At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression. Front. Genet. 12:700744. doi: 10.3389/fgene.2021.700744

INTRODUCTION

The removal on non-coding intervening sequences (introns) and ligation of coding sequences (exons) in pre-cursor messenger RNA (pre-mRNA), is carried out by a dynamic and complex machinery known as the spliceosome. The majority of metazoan organisms contain two parallel but analogous spliceosomes: the major or U2-dependent spliceosome which excises approximately 99.5% of introns, depending on the organism, and the minor or U12-dependent spliceosome which excises about 0.5% of introns. The respective intron types are similarly referred to as either major or U2-type introns, and minor or U12-type introns (Turunen et al., 2013a). The number of U12-type introns varies between species: for instance, in humans approximately 700 genes contain U12-type introns, while only 19 are found in *Drosophila*. More recently, an investigation of the genome of slime mold *Physarum polycephalum* revealed an exceptional case of >20,000 minor introns in a single genome (Larue et al., 2021). The typical architecture of minor intron containing genes (MIGs) includes a single U12-type intron per gene, flanked by multiple U2-type introns. However, a small subset of MIGs contain two or even three U12-type introns (Burge et al., 1998; Levine and Durbin, 2001; Moyer et al., 2020). The origin of the two parallel machineries and the disproportionate distribution of the two intron types in present day genomes has been the subject

of ongoing debate (Burge et al., 1998; Lynch and Richardson, 2002; Roy and Gilbert, 2006; Lin et al., 2010; Baumgartner et al., 2019). Nonetheless, it is safe to conclude that both machineries are ancient, related to one another, and originate from group II self-splicing introns (Burge et al., 1998; Shi, 2017). Interestingly, U12-type introns have reportedly been lost in multiple phylogenetic lineages, suggesting that they can be dispensable (Bartschat and Samuelsson, 2010). On the other hand, recognition sequences and locations of U12-type introns in individual genes are highly conserved in organismal lineages that have retained them (Moyer et al., 2020). These properties suggest that U12-type introns may serve an indispensable regulatory function in present-day organisms. Intriguingly, some introns harbor either tandem or overlapping splice sites that enable intron recognition by both major and minor spliceosomes, which in some cases, has been shown to have regulatory significance (Scamborova et al., 2004; Janice et al., 2013; Hafez and Hausner, 2015).

A key distinguishing feature between U12-type and U2-type introns is the conservation of intron recognition sequences i.e., the 5′ splice site (5′ss), 3′ splice site (3′ss) and the branch point sequence (BPS). Additionally, U2-type introns distinctively contain a polypyrimidine tract (PPT) upstream of the 3′ss. Splice site sequences are relatively weakly conserved in U2-type introns, which leads to more flexible splice site choices that fuel alternative splicing processes. In contrast, splice site sequences within U12-type introns are significantly more conserved (**Figure 1A**), which translates to less flexibility in splice site choice and consequently, reduced levels of alternative splicing in minor introns. Despite these differences, the overall splicing chemistry and spliceosome assembly is similar between the two intron types and has been covered in depth in several reviews (Patel and Steitz, 2003; Singh and Cooper, 2012; Turunen et al., 2013a; Matera and Wang, 2014; Jutzi et al., 2018; Wilkinson et al., 2020). Briefly, the 5′ss and BPS are initially recognized either by separate U1 and U2 snRNPs (major spliceosome) or a U11/U12 di-snRNP complex (minor spliceosome). Additionally, the PPT and 3′ss in U2-type introns are recognized by the U2AF1/U2AF2 protein heterodimer, whereas the U12-type intron 3′ss is recognized by the ZRSR2 protein. Following this initial recognition, the entry of U4/U6.U5 tri-snRNP (major spliceosome) or U4atac/U6atac.U5 tri-snRNP (minor spliceosome) leads to the formation of catalytic structures and intron excision.

Efforts aimed at understanding the functional significance of U12-type introns have suggested that the splicing of minor introns is slower or less efficient than that of major introns and could be used as a rate-limiting step to control or fine-tune mRNA levels of MIGs (Patel et al., 2002; Younis et al., 2013; Niemelä et al., 2014). Accordingly, several studies have reported elevated levels of retained U12-type introns under physiological conditions and in human diseases caused by mutations in minor spliceosome components. The retention of a single U12-type intron can disrupt the reading frame which would likely lead to downregulation at the protein level, either through nuclear retention and degradation of mRNAs containing unspliced U12-type introns (Niemelä et al., 2014; Ogami et al., 2018; Palazzo and Lee, 2018) or nonsense-mediated decay (NMD) due to introduction of premature termination codons (PTC) (Kurosaki et al., 2019). Consequently, studies examining the regulatory significance of U12-type introns or minor spliceosome diseases have mostly focused on intron retention, which is typically the predominant outcome of regulated or defective U12-type intron splicing. In this review, we instead focus on the mechanisms that involve interplay between adjacent minor and major spliceosomes during nuclear mRNA processing. We review mechanisms of minor and major spliceosome interactions, their potential regulatory significance under physiological conditions and their impact on minor spliceosome diseases.

COOPERATION AND COMPETITION IN RECOGNITION OF ADJACENT SPLICE SITES AS A MEANS OF REGULATION

Interaction between the minor and major spliceosomes during pre-mRNA processing can lead to two opposing outcomes: cooperation or competition. Cooperative interactions can facilitate mutual activation of adjacent spliceosomes during splicing through exon and intron definition mechanisms. Alternatively, the two spliceosomes can compete with one another for access to introns that harbor splice site recognition sequences for both machineries. Under normal physiological conditions, competitive interactions can be identified by the alternate use of either U12-type or U2-type splice sites, which result in different mRNA isoforms. In contrast, cooperative interactions between the two spliceosomes are more challenging to identify, as they typically do not lead to changes in alternative splice site usage under physiological conditions, except in the few cases where such interactions have been exploited for regulatory purposes. Additionally, diseases affecting minor spliceosome functions lead to a variety of alternative splicing choices in both interaction types.

Cooperation Between the Minor and Major Spliceosomes

At the mechanistic level, mutual interactions between the major and minor spliceosomes on the same pre-mRNA are predominantly mediated by exon definition interactions, in which the initial recognition of introns takes place by pairing splice sites across exons instead of introns (Robberson et al., 1990; Berget, 1995). Subsequently, the juxtaposition of exon-definition complexes enables the cross-intron pairing of splice sites through protein-protein interactions. Exon definition mechanisms are particularly useful in describing intron recognition mechanisms in vertebrates, which have the characteristic pre-mRNA architecture of relatively short exons separated by long introns. In contrast, the recognition of short introns occurs through intron-definition mechanisms whereby initial splice site pairing takes place across introns (Berget, 1995).

Both exon and intron definition mechanisms rely on protein-protein interactions to connect spliceosomal complexes

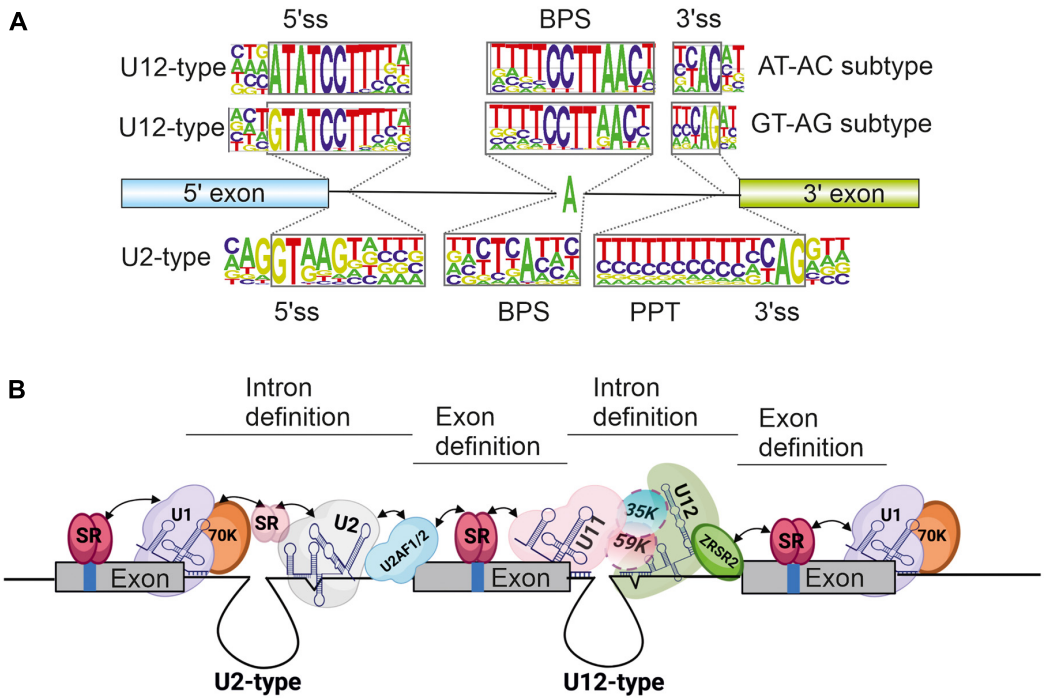


FIGURE 1 | (A) Consensus splice site sequences of minor and major introns. Minor or U12-type introns can additionally be grouped into AT-AC and GT-AG subtypes based on the first and last di-nucleotides. **(B)** Schematic of exon definition interactions taking place in a typical minor intron containing gene flanked by U2-type introns. Selected snRNA and protein components are indicated. Exons are depicted as solid boxes and introns as lines. Regulatory elements within exons are depicted as blue bars. Figure in panel **(B)** was modified from Turunen et al. (2013a). Created by BioRender.com.

assembled on the 5′ss or PPT/3′ss, to enable cross-exon and cross-intron communication. Proteins containing arginine and serine rich domains (RS domains) are the main facilitators of these interactions (**Figure 1B**). These include the SR protein family of splicing regulators and several integral spliceosome components present in both the major and minor spliceosomes (Long and Caceres, 2009). The latter group includes the U1-70K protein that is part of the U1 snRNP in the major spliceosome (Theissen et al., 1986; Spritz et al., 1987; Cho et al., 2011), its paralog U11-35K in the minor spliceosome which is also a component of the U11 snRNP (Will et al., 2004; Niemelä et al., 2015), the U2AF1/2 heterodimer involved in initial recognition of the PPT and 3′ss of U2-type introns, and the ZRSR2 protein that functions in recognition of the U12-type 3′ss (Tronchère et al., 1997; Shen et al., 2010). Additionally, recent work examining exon-bridging disruptions between major and minor spliceosomes provided evidence for a role of the U11-59K protein in exon-definition interactions (Olthof et al., 2021). Further regulation of both exon and intron definition mechanisms is provided by embedded exonic and intronic sequence elements, that are bound by RNA binding proteins (RBPs) including SR-proteins and heterogenous ribonucleoproteins (hnRNPs). Both SR and hnRNP proteins predominantly facilitate and regulate recognition of major introns (Berget, 1995; Reed, 1996; De Conti et al., 2013), but have also been shown to similarly interact with the minor spliceosome (Hastings and Krainer, 2001; Shen and Green, 2007; Turunen et al., 2013b). Consistently, exon-definition interactions between the major and minor spliceosomes have been demonstrated both *in vitro*, between U1 snRNP and the minor spliceosome (Wu and Krainer, 1996), and *in vivo*, between U11/U12 di-snRNP and the upstream U2AF1/2 bound to the major spliceosome PPT and 3′ss (**Figure 2**; Niemelä et al., 2015; Verbeeren et al., 2010, 2017; Olthof et al., 2021).

Currently, the only known cooperative interactions between the spliceosomes with demonstrated regulatory function have been reported for the *SNRNP48* and *RNPC3* genes. These encode the U11-48K and U11/U12-65K proteins, respectively, that are components of the U11/U12 di-snRNP. Both genes contain a conserved sequence element in non-coding regions which include a tandem repeat of the U12-type 5′ss consensus sequence. The tandem 5′ss sequences are recognized by the U11 snRNP, but are not used as splicing donors by the minor spliceosome. Instead, the binding of the U11/U12 di-snRNP activates an upstream U2-type 3′ss (**Figure 2**). Thus, the element has been aptly named a U11 snRNP-binding splicing enhancer, or USSE (Verbeeren et al., 2010). Interestingly, not only is the whole sequence stretch between the upstream 3′ss and the USSE element highly conserved, but the distance between the two sites also appears to be under evolutionary pressure to maintain optimal exon definition interactions between the two spliceosomes (Niemelä et al., 2015). These regulatory circuits function as autoregulatory or cross regulatory feedback mechanisms for both genes, promoting the formation of unproductive mRNA isoforms that are either degraded by NMD

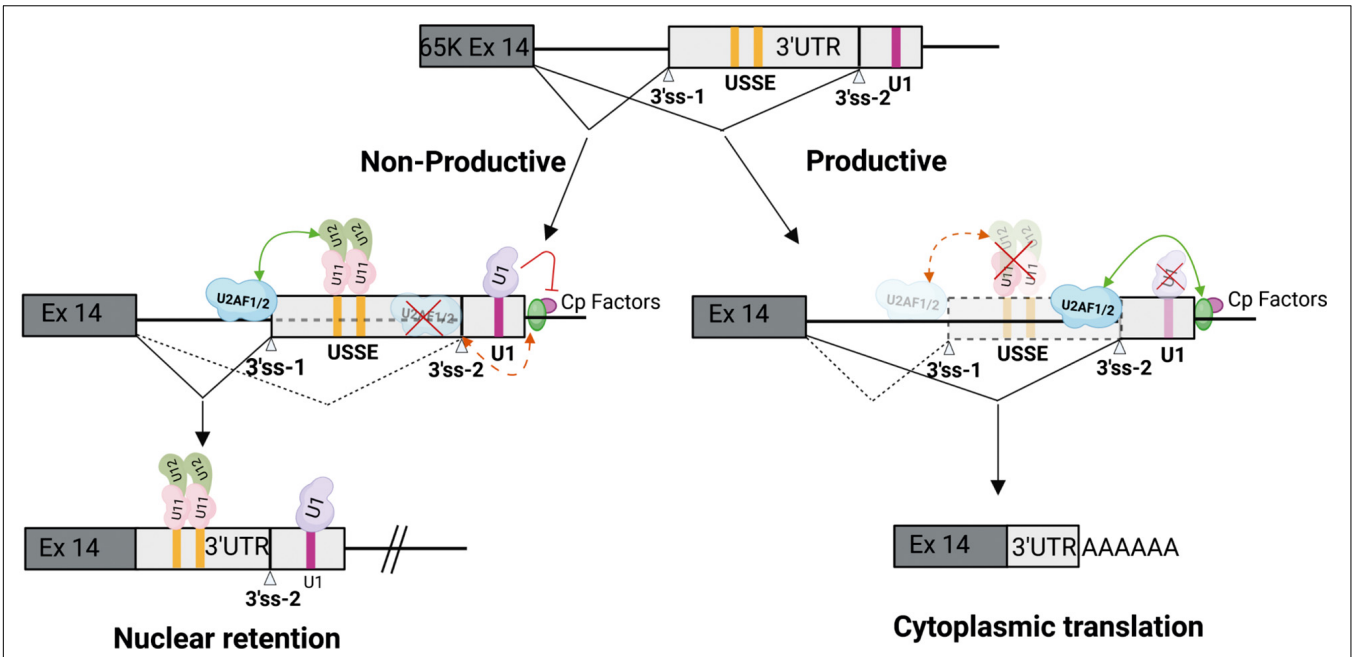


FIGURE 2 | Cooperation between minor and major spliceosomes. Feedback regulation in the minor *RNPC3*(65K) gene is mediated by protein-protein interactions between the U11/U12 di-snRNP bound to the USSE and upstream U2AF1/2 to enhance recognition of an upstream 3′ss. Cp factors denote cleavage and polyadenylation factors. This leads to formation of either productive mRNA translated in the cytoplasm, or unproductive mRNA that is retained in the nucleus. Modified from Verbeeren et al. (2017). Created by BioRender.com.

machinery (*SNRNP48*) due to inclusion of a PTC, or are retained in the nucleus due to an export-incompetent mRNA isoform (*RNPC3*) (**Figure 2**; Verbeeren et al., 2010, 2017; Niemelä et al., 2015). Notably, in *RNPC3* the same autoregulatory mechanism is also dynamically regulated during neuronal differentiation (Verbeeren et al., 2017), and is thus reminiscent of the post-transcriptional regulatory programs involving microexons and other RBPs (Ustianenko et al., 2017; Müller-mcnicoll et al., 2019).

Impact of Splicing Diseases on Spliceosome Cooperation

Further evidence for cooperative interactions between the two spliceosomes has emerged from global transcriptome analyses of diseases which partially compromise the function of the minor spliceosome. A typical outcome in these diseases is intron retention resulting from splicing defects. However, several studies have also demonstrated that the splicing defects are not limited to U12-type introns but can also spread to the flanking U2-type introns (Argente et al., 2014; Madan et al., 2015; Cologne et al., 2019; de Wolf et al., 2021; Olthof et al., 2021) in a subset of mRNAs. The most plausible explanation for these observations is that splicing of the affected U2-type introns is dependent on stabilizing exon-definition interactions with the neighboring U12-type introns (**Figure 1B**). Currently, evidence supporting this outcome is still somewhat limited and systematic surveys or more detailed mechanistic studies are needed to confirm these possibilities. Furthermore, there is currently no evidence of the regulatory significance of such interactions and it is possible that these interactions rather serve to reinforce constitutive splicing patterns.

Competition Between Minor and Major Spliceosomes

Competition between the two spliceosomes represents a special subclass of alternative splicing where an individual intron can be spliced by either the minor or the major spliceosome. In these instances, competing splice sites are typically in close proximity to one another on the pre-mRNA and the resulting mRNA isoforms are also often annotated in public databases. Depending on the positioning of the splice sites, such introns have been referred to as twintrons or nested introns, which both refer to instances where either minor or major intron is located within the other intron type (Levine and Durbin, 2001; Scamborova et al., 2004; Chang et al., 2007; Hafez and Hausner, 2015). Nested U2-type and U12-type introns can either have separate 5′ss and 3′ss sequences or they can share one splice site, but not both. Another possibility is that the two introns are interlocked and have a partially overlapping configuration (**Figure 3**). The exact frequency and functional significance of these juxtaposed U2-type and U12-type splice sites in the genomic context has not been systematically characterized. Furthermore, it is likely that at least a subset of such events may have been annotated as standard alternative U2-type splice sites utilized by the major spliceosome (Levine and Durbin, 2001; Chang et al., 2007), particularly with the GT-AG subclass of U12-type introns, that

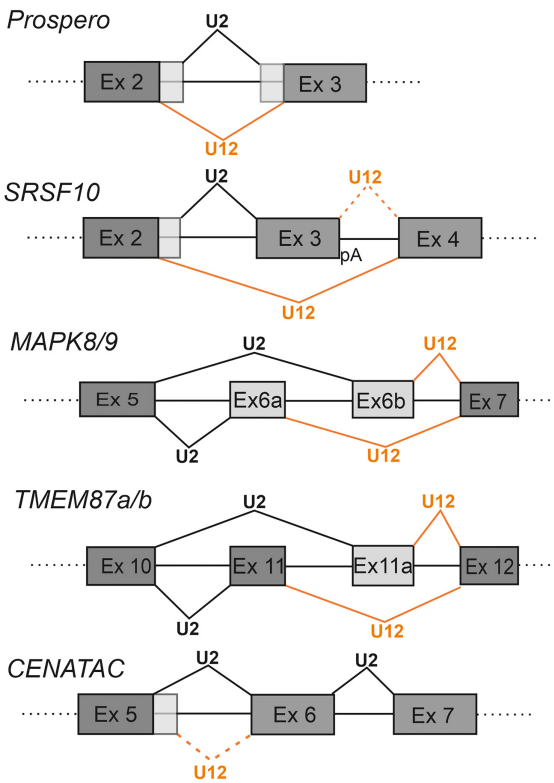


FIGURE 3 | Competition between minor and major spliceosomes. Examples of known genes utilizing adjacent U12-and U2-type introns in regulating gene expression. The U2-type splicing is indicated with black lines and U12-type splicing with orange lines. Alternative exon sequences are indicated with light gray shading. With *SRSF10* an alternative poly-A site (pA) downstream of exon 3 is indicated. In the same panel, the rare U12-type splice site usage downstream of the exon 3 is indicated by dotted orange line.

can be misannotated as major introns. A small subset of these competition events involving adjacent U2-type and U12-type splice sites appear to have regulatory significance as suggested by Janice et al. (2013), who identified 18 twintron arrangements in the human genome that were evolutionarily conserved in vertebrates. An alternative hypothesis is that the nested introns may represent evolutionary intermediates in the process of minor to major intron conversion that has been suggested as an explanation for the low numbers of minor introns in present-day genomes (Burge et al., 1998; Mount et al., 2007; Lin et al., 2010; Janice et al., 2013; Moyer et al., 2020).

The first known and best characterized case of nested U12-type and U2-type introns has been described for the *prospero* (*pro*) gene in *Drosophila* (Otake et al., 2002; Scamborova et al., 2004). In this case the U2-type intron is located inside of an AT-AC subtype U12-type intron, resulting in either shorter (*pro-S;* minor spliceosome) or longer (*pro-L;* major spliceosome) mRNA isoforms. The resulting *pro* protein isoforms differ in the homeodomain region, which may affect DNA binding specificity (**Figure 3**). The balance between the two isoforms is developmentally regulated *via* a purine-rich enhancer element that binds the *Drosophila* hnRNPA1 homologs, Hrp36/Hrp38 (Scamborova et al., 2004; Borah et al., 2009).

A recent study on the *SRSF10* gene, a member of SR family splicing regulators, illustrates the use of competing nested U12-type and U2-type introns in regulating not only the levels of the SRSF10 protein, but also other members of the same family (Meinke et al., 2020). The *SRSF10* regulatory module resembles that of the *prospero* circuit as it contains a U2-type intron embedded within a U12-type intron with AT-AC termini (**Figure 3**). Splicing through the minor pathway leads to skipping of exon 3 and formation of the full-length *SRSF10* mRNA. In contrast, use of the major pathway leads to inclusion of exon 3 and formation of a truncated mRNA that utilizes a polyadenylation signal in exon 3. Competitive recognition by either the major or minor spliceosome, thus determines inclusion or exclusion of exon 3, which also harbors an exonic splicing enhancer (ESE) that is specific for the SRSFR10 protein and is involved in SRSF10 autoregulation. Levels of the functional *SRSF10* isoform spliced by the minor spliceosome correlate not only with the activity of the minor spliceosome, but also with overall levels of other SR proteins, in a tissue-specific and developmental manner. This further suggests that regulation of the SR-protein family as a group, may be linked to the activity of the minor spliceosome (Meinke et al., 2020).

Other known cases of nested introns with an external U12-type intron and internal U2-type intron have been described for the *HNRNPLL*, *ZNF207,* and *C1orf112* genes (Janice et al., 2013) but a clear regulatory role (if any) for these splicing events has not yet been determined. Similarly, instances of the reverse arrangement of nested introns in which the U2-type intron is located externally and the U12-type intron internally, have been reported in *NCBP2, PRMT1, dZRSR2/Urp, CTNNBL1, CUL4A,* and *SPAG16* genes (Lin et al., 2010; Janice et al., 2013). Of these, *NCBP2,* a subunit of the nuclear cap binding complex, represents a well-characterized regulatory switch where use of the major splicing pathway results in a truncated protein that lacks a large part of the RNA Recognition Motif (RRM) (Pabis et al., 2010). The truncated NCBP2 form does not support heterodimer formation with the other subunit (NCBP1) or cap binding, but instead has independent roles in transcription and RNA processing (Pabis et al., 2010).

Examples of more complex arrangements in which U12-type and U2-type introns are found in an interlocked and partially overlapping arrangement, have been described for the c-Jun N-terminal Kinase (JNK) family genes (*MAPK8-9*) and for *TMEM87a* and *TMEM87b* genes. Here, competition between minor and major pathways leads to mutually exclusive incorporation of alternatively spliced exons into the mRNA. In the *MAPK8/9* genes, U12-type and U2-type splice sites have an interlocked configuration where mutual competition of two U2-type 3′ss and two U12-type 5′ss leads to inclusion of either the alternative exon 6a or exon 6b (exons 7a and 7b in later genome assemblies, respectively) (Chang et al., 2007). In this case competition involves use of either the U2-type 3′ss upstream of exon 6b or the U12-type 5′ss downstream of exon 6a (**Figure 3**). The different JNK isoforms exhibit tissue-specific expression, such that the exon 6a isoform is

predominantly expressed in neurons owing to the activity of neuronal splicing regulators, such as Nova (Relógio et al., 2005), whereas the isoform containing exon 6b is expressed ubiquitously (Chang et al., 2007). A recent study by Olthof et al. (2019) identified comparable examples in other MIGs, including the *TMEM87a* and *TMEM87b* genes. Similar to the *MAPK* circuit, both *TMEM87a and TMEM87b* contain an upstream U2-type intron with a 3′ss embedded in the downstream U12-type intron (**Figure 3**). Competition between the U2-type 3′ss upstream of exon 11a and the U12-type 5′ss downstream of exon 11 results in inclusion of either alternatively spliced exon 11 or exon 11a. Akin to the *MAPK* family genes, the different isoforms of *TMEM87a* and *TMEM87b* exhibit tissue-specific expression (Olthof et al., 2019), indicating that additional yet-to-be characterized regulatory factors play a role in splicing pattern selection in different tissues. Furthermore, the configuration of mutually exclusive alternative exons suggests that for both *MAPK8/9* and *TMEM 87a/b* genes, the regulation is linked to exon definition interactions between the minor and major spliceosomes across exons 6a/6b and 11/11a, respectively.

Typically, competing U12-type and U2-type introns harbor distinct splice sites for either splicing pathway. However, in rare cases, one of the splice sites, usually the 3′ss, can be shared between the two splicesosomes. One such case has been described in the transcriptomic analysis of patients suffering from Microcephalic Osteodysplastic Primordial Dwarfism type I/Taybi-Linder Syndrome (MOPD 1/TALS) by Cologne et al. (2019). The study described an alternative splicing switch in patient cells from U12- to U2-type 5′ss usage within intron 5 of the *CCDC84* gene (later renamed *CENATAC*; de Wolf et al., 2021), while the 3′ss was shared between the two intron types (**Figure 3**). Splicing by the major spliceosome is expected to increase levels of the CENATAC protein since use of the minor pathway leads to incorporation of a PTC, and possibly a decay of the target mRNA by the NMD pathway. Interestingly, the CENATAC protein has recently been identified as a novel component of the minor spliceosome and particularly necessary for splicing of the AT-AN subtype of U12-type introns (de Wolf et al., 2021). Both the U12-type and U2-type 5′ss sequences are phylogenetically highly conserved, suggesting that the competing 5′ss elements are part of an autoregulatory feedback mechanism regulating the cellular levels of the CENATAC protein. A similar case has been observed for the *MAPK12* gene in a cell line carrying a U12 snRNA mutation linked to cerebellar ataxia. In that case minor spliceosome dysfunction induces an exon skipping event where the U2-type 5′ss of the upstream intron is used together with the 3′ss of the downstream U12-type intron (Norppa and Frilander, 2021; **Figure 4**).

In both the *MAPK12* and *CENATAC* cases with a shared 3′ss, there is another, albeit undetectable functional change upstream of the 3′ss. In the major pathway this region constitutes a pyrimidine-rich PPT recognized by the U2AF2 protein present in the U2AF1/2 heterodimer. In contrast, the minor spliceosome does not use the PPT or the U2AF1/2 heterodimer for intron recognition, but rather relies on BPS recognition by the U12 snRNA. A plausible explanation for the intron 3′ end compatibility between the two spliceosomes is that the minor intron BPS is also pyrimidine-rich and can, in a suitable context, also serve as a PPT for major introns, as suggested earlier (Burge et al., 1998).

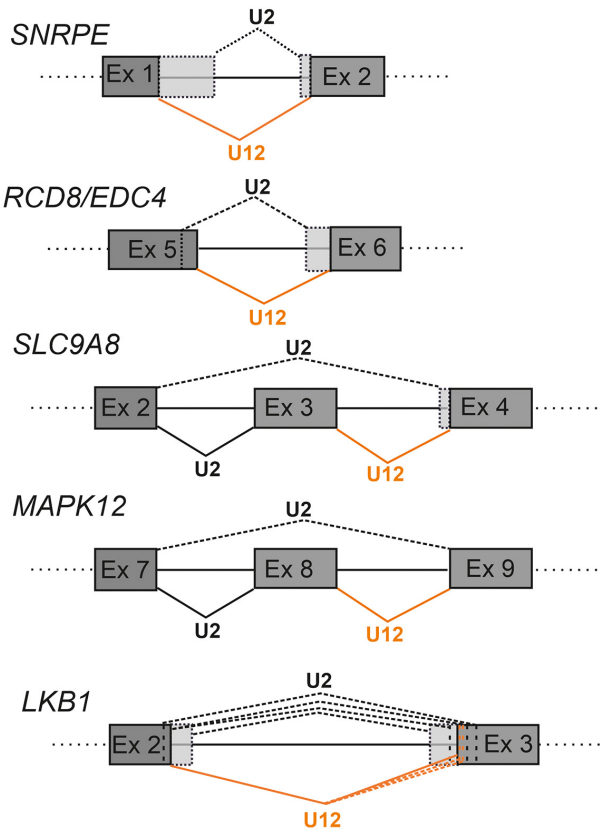


FIGURE 4 | Schematics showing examples of cryptic splice site activation in defective minor intron splicing. The U2-type splicing is indicated with black lines and U12-type splicing with orange lines. Constitutive splicing is indicated with solid lines and cryptic splicing with dotted lines. Alternative exon sequences are indicated with light gray shading. With LKB1 the additional cryptic splice sites are indicated by dotted vertical lines.

A unique case of competition has been described for the non-productive use of components of both splicing pathways to regulate the ratio between unspliced genomic RNA and subgenomic spliced mRNA species in the Rous Sarcoma Virus (RSV). The genomic RNA of this retrovirus contains an inhibitory splicing element designated the Negative Regulator of Splicing (NRS), the function of which is to suppress all splicing of the viral RNA to ensure production of unprocessed genomic RNA, that is subsequently packaged into virions (Gontarek et al., 1993). Interestingly, the NRS contains overlapping binding sites for both U1 and U11 snRNPs (Gontarek et al., 1993; Hibbert et al., 1999). Investigations into the NRS function revealed that the binding of U1 snRNP is essential for splicing inhibition through non-productive interactions between U1 snRNP and with the factors bound to the downstream 3′ss (Cook and McNally, 1999; Hibbert et al., 1999). In this process the U11 snRNP is a competitive antagonist for U1 binding, thereby fine-tuning the NRS activity (McNally et al., 2004, 2006).

Impact of Splicing Diseases on Spliceosome Competition

Human diseases that compromise functions of either spliceosome can have two differential outcomes in competitive contexts between the two spliceosomes. First, in cases of competitive intron recognition that leads to a balanced expression of spliceosome-specific mRNA isoforms, such as those described above, the most likely outcome is a shift in the balance between the isoforms. This effect may vary depending on the specific mutation/spliceosomal defect, and may be combined with other outcomes, such as alterations in exon definition mechanisms or increased levels of intron retention.

An alternative possibility is the activation of cryptic splice sites near introns that do not display any obvious competition between the two spliceosomes. "Cryptic" in this context refers to splice sites that are not used under physiological conditions and are predictably also not annotated or documented in public databases. Thus, cryptic splice sites can be thought of as pseudo splices sites that are weaker in strength compared to authentic sites, and as such, are not efficiently recognized by the spliceosome in a context specific manner. Such splice sites tend to be activated as a consequence of mutations in either authentic splice sites or alternatively, in spliceosome components or regulators. Both lead to defects in splice site recognition either at the level of single introns, or more broadly (Kapustin et al., 2011). While cryptic splice sites may often lead to context-specific alternative splicing, mutations can also inadvertently generate new splice site sequences that closely match consensus splice site sequences and thus result in disease-specific alternative splicing of transcripts (Buratti et al., 2007). In essence, under physiological conditions the level of competition between cryptic splice sites and authentic sites is low, and the use of cryptic sites only becomes visible in disease contexts (Kapustin et al., 2011; Jaganathan et al., 2019). In more benign settings, cryptic splice sites are thought to give rise to tissue-specific alternative splicing, creating splice isoforms with diverse functions in different tissues (Jaganathan et al., 2019).

The consequences of competitive intron recognition by either minor and major spliceosome on cryptic splicing have thus far been studied in cases where the splicing of U12-type introns has been compromised (Turunen et al., 2008; Cologne et al., 2019; de Wolf et al., 2021; Norppa and Frilander, 2021; Olthof et al., 2021). In such cases, the outcome is typically increased minor intron retention combined with the activation of nearby cryptic U2-type splice sites. In more rare cases, splicing of the U12-type intron in question appears to be unperturbed and the splicing defect is observed only due to activation of U2-type cryptic splice sites such as those described for the *SNRNPE*, *RCD8/EDC4* and *SLC9A8* genes (**Figure 4**; Turunen et al., 2008; de Wolf et al., 2021; Norppa and Frilander, 2021).

Analysis of simple U12-type splice site mutations have been reported in a small number of detailed studies including the *LKB1* gene (also known as *STK11*) implicated in the Peutz-Jager Syndrome (Hastings et al., 2005) and the *WDFY1* gene (Chang et al., 2007). The *LKB1* case is particularly illuminating as the A > G mutation of the first nucleotide of the intron only changes the U12-type 5′ss subtype, with the U12-type 5′ss still matching the consensus sequence (**Figure 4**). Nevertheless, this leads to a complex pattern of cryptic splice site activation by both minor and major spliceosomes, suggesting that even small changes in splice site strength can tip the balance in the competition between authentic and cryptic splice sites (**Figure 4**).

Other genes in which splice site mutations have been shown to lead to activation of a U2-type cryptic splice sites include *SEDL,* which has been linked to spondyloepiphyseal dysplasia tarda (SEDT) (Shaw et al., 2003) and *AP4M1* that has been linked to Cerebral palsy (Verkerk et al., 2009) both of which exhibit activation of cryptic splice sites as a consequence of the mutation. More recently, an analysis of single nucleotide polymorphisms (SNPs) in U12-type introns and MIGs revealed that such variants have a much wider connection to disease than previously thought (Olthof et al., 2020). In contrast to these splice site point mutations, mutations in components of the minor spliceosomal machinery are characterized by high levels of intron retention and additionally, cryptic splice activation in a larger number of mRNAs as evidenced by several transcriptomic studies (Madan et al., 2015; Merico et al., 2015; Cologne et al., 2019; de Wolf et al., 2021). The studies described above provide evidence for complex cryptic splicing events that not only disturb the splicing of immediate surroundings of the affected U12-type introns, but also extend further and influence the splicing of more distal U2-type introns, possibly as a consequence of disrupting the exon definition networks.

DISCUSSION

Until very recently, the outlook on the regulatory significance of the minor spliceosome functioning in parallel with the major spliceosome has been static. The main focus in the field has been on the inefficient splicing of U12-type introns under physiological conditions and the increased intron retention events observed in minor spliceosome diseases. However, rapidly accumulating transcriptomic data from an increasing number of minor spliceosome diseases is now challenging this narrow view by providing robust evidence of widespread crosstalk mechanisms between the minor and major spliceosomes, which were previously only reported in single-gene investigations. The fact that the locations of U12-type introns are known and highly conserved, presents a unique opportunity or lens through which the parallel functioning of both spliceosomes, particularly in exon and intron definition contexts, can be examined. Importantly, under physiological conditions, crosstalk between the two spliceosomes appears mostly to function as a mechanism for reinforcing constitutive splicing patterns through exon definition interactions, and to a lesser extent, as a mechanism for regulating balanced expression of mRNA isoforms that are dependent on either spliceosome. The less studied role of this crosstalk in regulating gene expression is particularly intriguing, as the few genes described here that utilize adjacent U12- and U2-type splice sites to regulate their expression, highlight an overlooked yet significant regulatory mechanism for some MIGs. Consequently, examining such crosstalk mechanisms has the potential to contribute to current understanding of the

evolutionary significance of both spliceosomes functioning in parallel. Interestingly, for genes currently known to harbor these adjacent splice sites both cooperative and competitive outcomes have been observed, with the latter being more common. Competition between splice sites is a common occurrence for U2-type introns, due to the more degenerate splice sites in these introns and has been discussed extensively within the context of alternative splicing (Chen and Manley, 2009; Wang et al., 2015; Dvinge, 2018; Ule and Blencowe, 2019). In contrast, competition between U12- and U2-type splices sites is relatively understudied and much remains to be understood regarding such competitive events, including the effects of enhancers and silencers. It would thus be interesting to determine if weaker U12- and U2-type splice sites are a common feature of nested introns and whether specific U12-type subtypes are preferred for these competitive events.

In disease contexts the primary splicing defect can lead to additional defects in both competitive and collaborative interactions between the two spliceosomes. The occurrence of additional mRNA isoforms as a consequence of the loss of crosstalk between the two spliceosomes also influences the interpretation of transcriptomic data derived from minor spliceosome disease patient cells. A typical transcriptome-level workflow aims to identify the affected minor introns and MIGs, and optionally, use the intron retention levels to estimate the downregulation of MIGs for downstream analyses. Thus, the presence of novel transcripts arising from cryptic splice site usage and the loss of exon definition interactions can either exacerbate the effect on expression levels or lead to formation of mutant proteins, which may contribute to the pathology of the given disease. Additionally, the *MAPK8/9* (Chang et al., 2007) and *TMEM87a/b* (Olthof et al., 2019) examples, as well as the MOPD1/TALS analyses (Cologne et al., 2019) have demonstrated that differential U12- or U2-type splice site usage is not static, and can change in a tissue-specific manner, necessitating extended analyses of multiple cells and tissue types. Such analyses may provide insight into the specificity that is observed in minor spliceosome diseases, in which developmental processes and neuronal tissues are particularly affected (Jutzi et al., 2018; Olthof et al., 2020).

At a more technical level, the activation and detection of cryptic splice sites poses an additional challenge for data analysis. Most software used in alternative splicing analyses rely on existing annotations when detecting alternative splicing events (Jiang and Chen, 2021). As cryptic splice sites are not normally annotated in public databases, they tend to be ignored by most alternative splicing analysis tools. This can be mitigated by using software that allows for the identification of *de novo* events, such as KisSplice used in the Cologne et al. (2019) study or MAJIQ which examines local splicing variation complexity (Vaquero-Garcia et al., 2016). In our recent work on Mosaic Variegated Aneuploidy (MVA) caused by mutations in *CENATAC* (de Wolf et al., 2021), we documented complex patterns of cryptic splice site activation as a consequence of minor intron splicing defects using Whippet (Sterne-Weiler et al., 2018), which is particularly suited for deciphering complex alternative splicing events. It is, however likely that in future, analyses of complex splicing events in transcripts may be resolved by the use of long-read sequencing methods that are less sensitive to annotation biases.

In summary, recent transcriptome-wide investigations have uncovered a substantial number of crosstalk interactions between the major and minor spliceosomes. An outstanding question related to these splicing events concerns the identification of true regulatory events from cases representing opportunistic cryptic splice site activations. An additional unanswered question is whether either spliceosome can be regulated individually in a manner that would influence the crosstalk. There is some evidence of specific splicing factors such as hnRNP H/F and SRSF10 being linked to regulation of the minor spliceosome (McNally et al., 2006; Turunen et al., 2013b; Meinke et al., 2020), but their generality remains to be confirmed. The currently known instances of cooperative and competitive interactions between the two spliceosomes highlighted in this review emphasize their functional and regulatory potential and set the stage for future investigations of their significance.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to the article and approved the submitted version.

FUNDING

This work has been supported by Sigrid Jusélius Foundation and Jane and Aatos Erkko Foundation awarded to MF.

ACKNOWLEDGMENTS

The authors would like to thank the members of Frilander laboratory for stimulating discussions and feedback during this work.

REFERENCES

Argente, J., Flores, R., Gutiérrez−Arumí, A., Verma, B., Martos−Moreno, G. Á, Cuscó, I., et al. (2014). Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. *EMBO Mol. Med.* 6, 299–306. doi: 10.1002/emmm.201303573

Bartschat, S., and Samuelsson, T. (2010). U12 type introns were lost at multiple occasions during evolution. *BMC Genomics* 11:106. doi: 10.1186/1471-2164-11-106

Baumgartner, M., Drake, K., and Kanadia, R. N. (2019). An Integrated Model of Minor Intron Emergence and Conservation. *Front. Genet.* 10:1113. doi: 10.3389/fgene.2019.01113

Berget, S. M. (1995). Exon Recognition in Vertebrate Splicing. *J. Biol. Chem.* 270, 2411–2414. doi: 10.1074/jbc.270.6.2411

Borah, S., Wong, A. C., and Steitz, J. A. (2009). Drosophila hnRNP A1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the prospero twintron. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2577–2582. doi: 10.1073/pnas.0812826106

Buratti, E., Chivers, M., Královičová, J., Romano, M., Baralle, M., Krainer, A. R., et al. (2007). Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* 35, 4250–4263. doi: 10.1093/nar/gkm402

Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). Evolutionary Fates and Origins of U12-Type Introns. *Mol. Cell* 2, 773–785. doi: 10.1016/s1097-2765(00)80292-0

Chang, W. C., Chen, Y. C., Lee, K. M., and Tarn, W. Y. (2007). Alternative splicing and bioinformatic analysis of human U12-type introns. *Nucleic Acids Res.* 35, 1833–1841. doi: 10.1093/nar/gkm026

Chen, M., and Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754. doi: 10.1038/nrm2777

Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R., et al. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8233–8238. doi: 10.1073/pnas.1017700108

Cologne, A., Benoit-Pilven, C., Besson, A., Putoux, A., Campan-Fournier, A., Bober, M. B., et al. (2019). New insights into minor splicing—a transcriptomic analysis of cells derived from TALS patients. *RNA* 25, 1130–1149. doi: 10.1261/rna.071423.119

Cook, C. R., and McNally, M. T. (1999). Interaction between the Negative Regulator of Splicing Element and a 3 splice site: requirement for U1 small nuclear ribonucleoprotein and the 3 splice site branch Point/Pyrimidine tract. *J. Virol.* 73, 2394–2400. doi: 10.1128/jvi.73.3.2394-2400.1999

De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4, 49–60. doi: 10.1002/wrna.1140

de Wolf, B., Oghabian, A., Akinyi, M. V., Hanks, S., Tromer, E. C., van Hooff, J., et al. (2021). Chromosomal instability by mutations in a novel specificity factor of the minor spliceosome. *EMBO J.* doi: 10.15252/embj.2020106536

Dvinge, H. (2018). Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett.* 592, 2987–3006. doi: 10.1002/1873-3468.13119

Gontarek, R. R., McNally, M. T., and Beemon, K. (1993). Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. *Genes Dev.* 7, 1926–1936. doi: 10.1101/gad.7.10.1926

Hafez, M., and Hausner, G. (2015). Convergent evolution of twintron-like configurations: one is never enough. *RNA Biol.* 12, 1275–1288. doi: 10.1080/15476286.2015.1103427

Hastings, M. L., and Krainer, A. R. (2001). Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* 7, 471–482. doi: 10.1017/S1355838201002552

Hastings, M. L., Resta, N., Traum, D., Stella, A., Guanti, G., and Krainer, A. R. (2005). An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at noncanonical cryptic splice sites. *Nat. Struct. Mol. Biol.* 12, 54–59. doi: 10.1038/nsmb873

Hibbert, C. S., Gontarek, R. R., and Beemon, K. L. (1999). The role of overlapping U1 and U11 5' splice site sequences in a negative regulator of splicing. *RNA* 5, 333–343. doi: 10.1017/S1355838299981347

Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.e24. doi: 10.1016/j.cell.2018.12.015

Janice, J., Jakalski, M., and Makałowski, W. (2013). Surprisingly high number of Twintrons in vertebrates. *Biol. Direct* 8:4. doi: 10.1186/1745-6150-8-4

Jiang, W., and Chen, L. (2021). Alternative splicing: human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.* 19, 183–195. doi: 10.1016/j.csbj.2020.12.009

Jutzi, D., Akinyi, M. V., Mechtersheimer, J., Frilander, M. J., and Ruepp, M.-D. (2018). The emerging role of minor intron splicing in neurological disorders. *Cell Stress* 2, 40–54. doi: 10.15698/cst2018.03.126

Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., et al. (2011). Cryptic splice sites and split genes. *Nucleic Acids Res.* 39, 5837–5844. doi: 10.1093/nar/gkr203

Kurosaki, T., Popp, M. W., and Maquat, L. E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* 20, 406–420. doi: 10.1038/s41580-019-0126-2

Larue, G. E., Eliáš, M., and Roy, S. W. (2021). Expansion and transformation of the minor spliceosomal system in the slime mold Physarum polycephalum. *Curr. Biol.* doi: 10.1016/j.cub.2021.04.050

Levine, A., and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.* 29, 4006–4013. doi: 10.1093/nar/29.19.4006

Lin, C. F., Mount, S. M., Jarmoowski, A., and Makaowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10:47. doi: 10.1186/1471-2148-10-47

Long, J. C., and Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* 417, 15–27. doi: 10.1042/BJ20081501

Lynch, M., and Richardson, A. O. (2002). The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* 12, 701–710. doi: 10.1016/S0959-437X(02)00360-X

Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., et al. (2015). Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* 6, 1–14. doi: 10.1038/ncomms7042

Matera, A. G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121. doi: 10.1038/nrm3742

McNally, L. M., Yee, L., and McNally, M. T. (2004). Two regions promote U11 small nuclear ribonucleoprotein particle binding to a retroviral splicing inhibitor element (negative regulator of splicing). *J. Biol. Chem.* 279, 38201–38208. doi: 10.1074/jbc.M407073200

McNally, L. M., Yee, L., and McNally, M. T. (2006). Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. *J. Biol. Chem.* 281, 2478–2488. doi: 10.1074/jbc.M511215200

Meinke, S., Goldammer, G., Weber, A. I., Tarabykin, V., Neumann, A., Preussner, M., et al. (2020). Srsf10 and the minor spliceosome control tissue-specific and dynamic sr protein expression. *Elife* 9:e56075. doi: 10.7554/eLife.56075

Merico, D., Roifman, M., Braunschweig, U., Yuen, R. K. C., Alexandrova, R., Bates, A., et al. (2015). Compound heterozygous mutations in the noncoding RNU4ATAC cause Roifman Syndrome by disrupting minor intron splicing. *Nat. Commun.* 6:8718. doi: 10.1038/ncomms9718

Mount, S. M., Gotea, V., Lin, C. F., Hernandez, K., and Makałowski, W. (2007). Spliceosomal small nuclear RNA genes in 11 insect genomes. *RNA* 13, 5–14. doi: 10.1261/rna.259207

Moyer, D. C., Larue, G. E., Hershberger, C. E., Roy, S. W., and Padgett, R. A. (2020). Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.* 48, 7066–7078. doi: 10.1093/nar/gkaa464

Müller-mcnicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *J. Mol. Cell Biol.* 11, 930–939. doi: 10.1093/jmcb/mjz043

Niemelä, E. H., Oghabian, A., Staals, R. H. J., Greco, D., Pruijn, G. J. M., and Frilander, M. J. (2014). Global analysis of the nuclear processing of transcripts with unspliced U12-type introns by the exosome. *Nucleic Acids Res.* 42, 7358–7369. doi: 10.1093/nar/gku391

Niemelä, E. H., Verbeeren, J., Singha, P., Nurmi, V., Frilander, M. J., and Niemelä, E. H. (2015). Evolutionarily conserved exon definition interactions with U11 snRNP mediate alternative splicing regulation on U11-48K and U11/U12-65K genes. *RNA Biol.* 12, 1256–1264. doi: 10.1080/15476286.2015.1096489

Norppa, A. J., and Frilander, M. J. (2021). The integrity of the U12 snRNA 3' stem–loop is necessary for its overall stability. *Nucleic Acids Res.* 49, 2835–2847. doi: 10.1093/nar/gkab048

Ogami, K., Chen, Y., and Manley, J. L. (2018). RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA* 4:8. doi: 10.3390/ncrna4010008

Olthof, A. M., Hyatt, K. C., and Kanadia, R. N. (2019). Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics* 20:686. doi: 10.1186/s12864-019-6046-x

Olthof, A. M., Rasmussen, J. S., Campeau, P. M., and Kanadia, R. N. (2020). Disrupted minor intron splicing is prevalent in Mendelian disorders. *Mol. Genet. Genomic Med.* 8:e1374. doi: 10.1002/mgg3.1374

Olthof, A. M., White, A. K., Mieruszynski, S., Doggett, K., Lee, M. F., Chakroun, A., et al. (2021). Disruption of exon-bridging interactions between minor and

major spliceosomes reult in alternative splicing around minor introns. *Nucleic Acids Res.* 49, 3524–3545. doi: 10.1093/nar/gkab118

Otake, L. R., Scamborova, P., Hashimoto, C., and Steitz, J. A. (2002). The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in Drosophila. *Mol. Cell* 9, 439–446. doi: 10.1016/S1097-2765(02)00441-0

Pabis, M., Neufeld, N., Shav-Tal, Y., and Neugebauer, K. M. (2010). Binding properties and dynamic localization of an alternative isoform of the cap-binding complex subunit CBP20. *Nucleus* 1, 412–421. doi: 10.4161/nucl.1.5.12839

Palazzo, A. F., and Lee, E. S. (2018). Sequence determinants for nuclear retention and cytoplasmic export of mRNAs and lncRNAs. *Front. Genet.* 9:440. doi: 10.3389/fgene.2018.00440

Patel, A. A., McCarthy, M., and Steitz, J. A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* 21, 3804–3815. doi: 10.1093/emboj/cdf297

Patel, A. A., and Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* 4, 960–970. doi: 10.1038/nrm1259

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* 6, 215–220. doi: 10.1016/S0959-437X(96)80053-0

Relógio, A., Ben-Dov, C., Baum, M., Ruggiu, M., Gemund, C., Benes, V., et al. (2005). Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.* 280, 4779–4784. doi: 10.1074/jbc.M411976200

Robberson, B. L., Cote, G. J., Berget, S. M., and Mcclean, M. (1990). Exon Definition May Facilitate Splice Site Selection in RNAs with Multiple Exons Downloaded from. *Mol. Cell. Biol.* 10, 84–94. doi: 10.1128/mcb.10.1.84-94.1990

Roy, S. W., and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* 7, 211–221. doi: 10.1038/nrg1807

Scamborova, P., Wong, A., and Steitz, J. A. (2004). An Intronic Enhancer Regulates Splicing of the Twintron of Drosophila melanogaster prospero Pre-mRNA by Two Different Spliceosomes. *Mol. Cell. Biol.* 24, 1855–1869. doi: 10.1128/mcb.24.5.1855-1869.2004

Shaw, M., Brunetti-Pierri, N., Kádasi, L., Kovácová, V., Van Maldergem, L., De Brasi, D., et al. (2003). Identification of three novel *SEDL* mutations, including mutation in the rare, non-canonical splice site of exon 4. *Clin. Genet.* 64, 235–242. doi: 10.1034/j.1399-0004.2003.00132.x

Shen, H., and Green, M. R. (2007). RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat. Struct. Mol. Biol.* 14, 597–603. doi: 10.1038/nsmb1263

Shen, H., Zheng, X., Luecke, S., and Green, M. R. (2010). The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* 24, 2389–2394. doi: 10.1101/gad.1974810

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. doi: 10.1038/nrm.2017.86

Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006

Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E., and Francke, U. (1987). The human u1-70k snrnp protein: CDNA cloning, chromosomal localization, expression,alternative splicing and RNA-binding. *Nucleic Acids Res.* 15, 10373–10391. doi: 10.1093/nar/15.24.10373

Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H., and Blencowe, B. J. (2018). Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop. *Mol. Cell* 72, 187–200.e6. doi: 10.1016/j.molcel.2018.08.018

Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., et al. (1986). Cloning of the human cDNA for the U1 RNA-associated 70K protein. *EMBO J.* 5, 3209–3217. doi: 10.1002/j.1460-2075.1986.tb04631.x

Tronchère, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* 388, 397–400. doi: 10.1038/41137

Turunen, J. J., Niemelä, E. H., Verma, B., and Frilander, M. J. (2013a). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76. doi: 10.1002/wrna.1141

Turunen, J. J., Verma, B., Nyman, T. A., and Frilander, M. J. (2013b). HnRNPH1/H2, U1 snRNP, and U11 snRNP cooperate to regulate the stability of the U11-48K pre-mRNA. *RNA* 19, 380–389. doi: 10.1261/rna.036715.112

Turunen, J. J., Will, C. L., Grote, M., Lührmann, R., and Frilander, M. J. (2008). The U11-48K Protein Contacts the 5' Splice Site of U12-Type Introns and the U11-59K Protein. *Mol. Cell. Biol.* 28, 3548–3560. doi: 10.1128/mcb.01928-07

Ule, J., and Blencowe, B. J. (2019). Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* 76, 329–345. doi: 10.1016/j.molcel.2019.09.017

Ustianenko, D., Weyn-Vanhentenryck, S. M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. *Wiley Interdiscip. Rev. RNA* 8:e1418. doi: 10.1002/wrna.1418

Vaquero-Garcia, J., Barrera, A., Gazzara, M. R., Gonzalez-Vallinas, J., Lahens, N. F., Hogenesch, J. B., et al. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* 5:e11752. doi: 10.7554/eLife.11752

Verbeeren, J., Niemelä, E. H., Turunen, J. J., Will, C. L., Ravantti, J. J., Lührmann, R., et al. (2010). An Ancient Mechanism for Splicing Control: U11 snRNP as an Activator of Alternative Splicing. *Mol. Cell* 37, 821–833. doi: 10.1016/j.molcel.2010.02.014

Verbeeren, J., Verma, B., Niemelä, E. H., Yap, K., Makeyev, E. V., and Frilander, M. J. (2017). Alternative exon definition events control the choice between nuclear retention and cytoplasmic export of U11/U12-65K mRNA. *PLoS Genet.* 13:e1006824. doi: 10.1371/journal.pgen.1006824

Verkerk, A. J. M. H., Schot, R., Dumee, B., Schellekens, K., Swagemakers, S., Bertoli-Avella, A. M., et al. (2009). Mutation in the AP4M1 Gene Provides a Model for Neuroaxonal Injury in Cerebral Palsy. *Am. J. Hum. Genet.* 85, 40–52. doi: 10.1016/j.ajhg.2009.06.004

Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., et al. (2015). Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158. doi: 10.3892/br.2014.407

Wilkinson, M. E., Charenton, C., and Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* 89, 359–388.

Will, C. L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., et al. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* 10, 929–941. doi: 10.1261/rna.7320604

Wu, Q., and Krainer, A. R. (1996). U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* 274, 1005–1008. doi: 10.1126/science.274.5289.1005

Younis, I., Dittmar, K., Wang, W., Foley, S. W., Berg, M. G., Hu, K. Y., et al. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *Elife* 2:e00780. doi: 10.7554/eLife.00780

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

MINI REVIEW
published: 27 August 2021
doi: 10.3389/fgene.2021.737433

Established and Emerging Regulatory Roles of Eukaryotic Translation Initiation Factor 5B (eIF5B)

Prakash Amruth Raj Chukka[1,2,3,4], *Stacey D. Wetmore*[1,2,3,4]* *and Nehal Thakor*[1,2,5,6,7]*

[1] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada, [2] Southern Alberta Genome Sciences Centre (SAGSC), University of Lethbridge, Lethbridge, AB, Canada, [3] Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada, [4] Canadian Centre of Research in Advanced Fluorine Technologies (C-CRAFT), University of Lethbridge, Lethbridge, AB, Canada, [5] Department of Biological Sciences, University of Lethbridge, Lethbridge, AB, Canada, [6] Department of Neuroscience, Canadian Centre for Behavioral Neuroscience (CCBN), University of Lethbridge, Lethbridge, AB, Canada, [7] Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Translational control (TC) is one the crucial steps that dictate gene expression and alter the outcome of physiological process like programmed cell death, metabolism, and proliferation in a eukaryotic cell. TC occurs mainly at the translation initiation stage. The initiation factor eIF5B tightly regulates global translation initiation and facilitates the expression of a subset of proteins involved in proliferation, inhibition of apoptosis, and immunosuppression under stress conditions. eIF5B enhances the expression of these survival proteins to allow cancer cells to metastasize and resist chemotherapy. Using eIF5B as a biomarker or drug target could help with diagnosis and improved prognosis, respectively. To achieve these goals, it is crucial to understand the role of eIF5B in translational regulation. This review recapitulates eIF5B's regulatory roles in the translation initiation of viral mRNA as well as the cellular mRNAs in cancer and stressed eukaryotic cells.

Keywords: eukaryotic initiation factor 5B (eIF5B), mRNA translation, Non-canonical Translation Ianitiation, IRES, uORF

OPEN ACCESS

Edited by:
Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:
Toshinobu Fujiwara,
Kindai University, Japan
Hiroaki Imataka,
University of Hyogo, Japan

***Correspondence:**
Stacey D. Wetmore
stacey.wetmore@uleth.ca
Nehal Thakor
nthakor@uleth.ca

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *06 July 2021*
Accepted: *10 August 2021*
Published: *27 August 2021*

Citation:
Chukka PAR, Wetmore SD and Thakor N (2021) Established and Emerging Regulatory Roles of Eukaryotic Translation Initiation Factor 5B (eIF5B). Front. Genet. 12:737433. doi: 10.3389/fgene.2021.737433

INTRODUCTION

Cap-dependent or canonical translation initiation is an intricate process and highly regulated in eukaryotes. It involves multiple initiation factors ranging from small proteins to complex multidomain proteins. The process begins when 5′ cap is recognized by eukaryotic initiation factor 4F (eIF4F) complex and mRNA is recruited onto the 43S preinitiation complex (43S-PIC) (**Supplementary Figure 1**). Subsequently, this leads to the formation of 48S preinitiation complex (48S-PIC), which scans the 5′ untranslated region (UTR) of the mRNA and recognizes the start codon (AUG). In the final step, eukaryotic initiation factor 5B (eIF5B) promotes the association of small (40S) and large (60S) ribosomal subunits to form elongation competent 80S initiation complex (80S IC) (**Supplementary Figure 1**).

eIF5B, also known as IF-M2A/hIF2, is a "Arrokoth" shaped GTPase discovered in 1975. It is universally conserved among all eukaryotes and encoded by the *EIF5B* gene (Pestova et al., 2000; Roll-Mecak et al., 2000). eIF5B mediates the association of the 40S and 60S ribosomal subunits during eukaryotic translation initiation (Merrick et al., 1975; Lee et al., 2002; Shin et al., 2002).

Although eIF5B is highly conserved, its depletion does not have a substantial effect on cell viability under normal conditions (Lee et al., 2014; Ho et al., 2018; Ross et al., 2019, 2020). Conversely, reduced levels of eIF5B under stress conditions significantly affects cell viability (Lee et al., 2014; Ho et al., 2018).

eIF5B consists of a highly conserved functional C-terminal region (human 587–1220, yeast 397–1,002) and a less conserved N-terminal region (human 1–586, yeast 1–396) (Choi et al., 1998; Lee et al., 1999). Deleting the N-terminal region does not affect cell viability and many *in vitro* studies have shown that N-terminally truncated eIF5B is active (human 587–1,220, yeast 397–1,002) (Pestova et al., 2000; Lee et al., 2002; Shin et al., 2002; Fringer et al., 2007; Pisareva and Pisarev, 2014). On the other hand, the functional C-terminal consists of four domains: G domain (human 629–850, yeast 401–625), domain II (human 856–948, yeast 630–745), domain III (human 951–1,082, 755–855), and domain IV (human 1,076–1,220, yeast 859–1,002) (Pestova et al., 2001; Nag et al., 2016; Huang and Fernandez, 2020). Additionally, domains III and IV are connected by a helix h12 whose deletion in yeast yields non-functional eIF5B similar to the defects observed in ΔeIF5B and Δdomain IV cells (Fringer et al., 2007; Shin et al., 2011).

Studies on eIF5B demonstrated a newer function that parallels the role of eIF2, a Met-tRNA$_i^{Met}$ delivering eukaryotic initiation factor, during stress conditions (Thakor and Holcik, 2012; Holcik, 2015; Sharma et al., 2016). The mechanism is activated when α-subunit of eIF2 is phosphorylated and sequestered by eIF2B. Under these conditions, eIF5B promotes translation of specific proteins by delivering Met-tRNA$_i^{Met}$ to the eukaryotic ribosomes, due to its homology to IF-2 which delivers met-tRNAf^{Met} to the bacterial ribosomes (Ross et al., 2019). Additionally, recent studies have unveiled a mechanistic role of eIF5B during the canonical and non-canonical eukaryotic translation initiation (Ross et al., 2019). We and others have also clearly implicated eIF5B in oncogenesis (Wang et al., 2016; Ross et al., 2019; Suresh et al., 2020). Accordingly, this review provides insights into the coordination of eIF5B with other eukaryotic initiation factors to carry out diverse functions during translation initiation in yeast and human cells. This review covers the established roles of eIF5B in 40S ribosome maturation, formation of 48S PIC, stabilization of Met-tRNA$_i^{Met}$, 60S ribosomal recruitment and 80S complex formation. We also highlight the emerging roles of eIF5B during Met-tRNA$_i^{Met}$ delivery by coordinating with eIF2A, uORF-mediated translation initiation, and IRES-mediated translation initiation. We further discuss how eIF5B acts as a nexus between non-canonical translation and the survival of cancer cells.

Role of eIF5B in Pre-40S Ribosome Subunit Maturation

During ribosome biogenesis, the large and small subunits undergo a translation-like cycle where eIF5B mediates the association of the pre-40S and 60S ribosomal subunits, which acts as a quality control step (Lebaron et al., 2012). The resulting complex is not a true 80S initiation-complex (80S IC) as it lacks initiator tRNA and mRNA. As a result, the 60S ribosomal subunit is displaced by the termination factor Rli-1 during 40S subunit maturation (Strunk et al., 2012; Woolford and Baserga, 2013). The 80S-like complex ensures the proper functioning of pre-ribosomes before translation (Strunk et al., 2012). A study on the YKK392 yeast strain devoid of eIF5B has shown a negative effect on the ribosomal subunit association, resulting in the accumulation of pre-40S subunits (Strunk et al., 2012). In general, this accumulation is not detrimental, but delays the formation of the 80S-like ribosomal complex and slows down cell growth (Strunk et al., 2012). Deleting the eIF5B coding gene *FUN12* in yeast also results in the accumulation of pre-18S rRNA, and decreased levels of 27S pre-rRNA and 40S ribosomes (Lebaron et al., 2012; Strunk et al., 2012). Thus, eIF5B is essential for catalyzing the ribosome maturation process in yeast.

eIF5B Interacts With eIF5 to Stimulate the Formation of 48S Initiation Complex

Human eIF5B and eIF5 synergistically mediate the efficient formation of the 48S initiation complex (48S IC) (Pisareva and Pisarev, 2014). The interaction between domain IV of eIF5B and the C-terminus of eIF5 could be crucial for efficient 48S IC formation (Lin et al., 2018). Affinity studies inferred a higher affinity of eIF5B for eIF5 compared with eIF1A (Lin et al., 2018). It has been hypothesized that human eIF5 and eIF5B together stimulate 43S preinitiation complex (PIC) rearrangement to increase the yield of functional 48S IC (Pisareva and Pisarev, 2014). Additionally, eIF5B deletion results in the destabilization of 48S IC, which was suggested to be induced by eIF5 (Pisareva and Pisarev, 2014). This phenomenon was observed in both optimal and non-optimal AUG context, suggesting the role of eIF5B in 48S IC stabilization (Pisareva and Pisarev, 2014). eIF5 has additional roles as a GTPase-activating protein (GAP) and a GDP-dissociation inhibitor (GDI) (Paulin et al., 2001). While eIF5 induces eIF2-GTP hydrolysis, eIF5B promotes the release of eIF2 from 48S IC (Unbehaun et al., 2004; Pisarev et al., 2006). Furthermore, eIF5B assists in establishing 48S IC on a bona fide AUG and prevents leaky scanning along with eIF5 (Pisareva and Pisarev, 2014; Lin et al., 2018). Since codon scanning and selection during translation initiation are essential for generating functional proteins, these studies collectively provide compelling evidence that eIF5B coordinates with eIF5 to aid the establishment of an efficient 48S IC on a bona fide start codon.

Met-tRNA$_i^{Met}$ Stabilization by eIF5B

After eIF2-GTP delivers Met-tRNA$_i$ to the P-site of the 40S ribosomal subunit, GAP eIF5 induces eIF2-GTP hydrolysis (Paulin et al., 2001; Majumdar and Maitra, 2005). eIF2B disrupts the eIF5/eIF2-GDP interaction and facilitates eIF2-GDP release (Jennings and Pavitt, 2010; Jennings et al., 2016). In the absence of eIF2, domains III and IV of eIF5B extend into the inter-subunit space and stabilize Met-tRNA$_i^{Met}$ (Huang and Fernandez, 2020). Conformational changes in domains III and IV facilitate interactions between basic amino acids in domain IV and $_{73}$ACCA$_{76}$-Met of Met-tRNA$_i^{Met}$ (Fernandez et al., 2013). This creates a kink in the Met-tRNA$_i^{Met}$ stem structure that

is not seen in elongation tRNA, which helps Met-tRNA$_i^{Met}$ simultaneously interact with both mRNA and eIF5B (Wang et al., 2020). Additionally, methionine of tRNA$_i$ positions itself in the hydrophobic pocket formed by eIF5B and the uL16 loop of the 60S subunit, and single-molecule experiments suggested that the hydrophobic pocket acts as a residue selectivity filter (Wang et al., 2020). Ultimately, the eIF5B-initiator tRNA complex places the tRNA$_i$ aminoacyl end out of the peptidyl transfer center, awaiting GTP hydrolysis. Thus, these eIF5B interactions are vital for stabilization and correct positioning Met-tRNA$_i^{Met}$ in the initiation complex after eIF2 is displaced from the ribosome, suggesting the absence of eIF5B could delay the transition into elongation.

Interaction of eIF5B With eIF1A for Ribosome Recruitment

eIF1A is one of the earliest discovered interacting partners of eIF5B. Similar to eIF5B, eIF1A is considered to be a universal translational factor (Sorensen et al., 2001). The interaction between these two initiation factors is thought to be important for eIF5B recruitment and mediating the joining of the large and small ribosomal subunits (Fringer et al., 2007). In contrast, another study proposed that eIF5 recruits eIF5B and eIF1A disrupts their interaction after eIF2-GTP hydrolysis (Lin et al., 2018). In the absence of the ribosome, interactions between eIF5B and eIF1A are disrupted due to intramolecular interactions within each initiation factor (Nag et al., 2016). Interactions between eIF1A and eIF5B are only established after codon recognition when C-terminal tail (CTT) of eIF1A is displaced from the P-site to interact with the adjacent eIF5B (Yu et al., 2009). Domains III and IV of eIF5B interact with the oligonucleotide/oligosaccharide-binding (OB) domain and CTT of eIF1A (Nag et al., 2016). These interactions are critical for the recruitment of the 60S subunit and formation of the 80S complex (Olsen et al., 2003; Acker et al., 2006). Thus, these studies suggest that interactions between the two universally conserved eIF5B and eIF1A initiation factors are necessary to mediate the association of the 40 and 60S subunits by eIF5B, which forms a viable 80S IC.

Ribosomal Subunit Association Is Expedited by eIF5B

One primary role of eIF5B is to promote the ribosomal association of the 40S and 60S subunits (**Figure 1A**; Lee et al., 2002; Shin et al., 2002). The long retention time (30–60 s) of eIF5B on ribosomes indirectly prevents their collision on mRNA before 80S IC transitions into the elongation step (Wang et al., 2019). Apart from ribosomal subunit joining, *in vivo* studies have shown the stabilization of the halfmer polysome (43S PIC + 80S on an mRNA) by eIF5B (Lee et al., 2002). To participate in ribosomal subunit association, eIF5B must be in an active form, which has been proposed to be achieved through a domain release mechanism (Nag et al., 2016). The process begins when the inherently rigid domains III and IV of eIF5B-GDP become flexible upon GTP binding (Kuhle and Ficner, 2014).

Each domain of eIF5B has a specific function during ribosome association: (1) the G-domain interacts with the 60S subunit and is involved in GTP hydrolysis, (2) domain II is anchored to the 40S subunit, (3) domain III also anchors to the 40S subunit, and promotes GTP hydrolysis when Met-tRNA$_i^{Met}$ is delivered, and (4) domain IV interacts with t-RNA, eIF1A, and eIF5 (Fernandez et al., 2013; Nag et al., 2016; Lin et al., 2018; Huang and Fernandez, 2020). GTP hydrolysis is not required for ribosomal subunit association, but is required for the release of eIF5B and 80S IC transition into the elongation step (Shin et al., 2002; Huang and Fernandez, 2020). If domain III does not recognize a proper Met-tRNA$_i^{Met}$ delivery or ribosomal association, eIF5B could be trapped in the P/A site, which hampers the recruitment of a new aminoacyl-tRNA, delays the transition to elongation, and obstructs new ribosome recruitment. Although ribosome recruitment occurs even in the absence of eIF5B, 80S IC formation is inefficient and the transition time into the elongation step is longer, resulting in a slow growth phenotype (Fringer et al., 2007; Jiang et al., 2016). Overall, once eIF5B induces the 60S and 40S subunit association, eIF5B-GDP is released, making the 80S initiation complex elongation competent (Fringer et al., 2007).

eIF5B Interaction With eIF2A in Non-canonical Translation Initiation

When availability of ternary complex (eIF2-GTP-Met-tRNA$_i^{Met}$) is low under stress conditions, initiation factor eIF2A has been shown to deliver the Met-tRNA$_i^{Met}$ to the 40S ribosome by coordinating with eIF5B (Kim et al., 2018). eIF5B alone plays a major role in Met-tRNA$_i^{Met}$ delivery during translation initiation on certain virus mRNAs (Pestova et al., 2008; Yamamoto et al., 2014). *In vitro* studies involving pull down assays suggest that domain IV of eIF5B modestly interacts with the M domain of eIF2A (462–502 aa), but all eIF5B domains are required for a high affinity interaction (Kim et al., 2018). A predicted model suggested that eIF5B domain IV might also be responsible for interacting and delivering Met-tRNA$_i^{Met}$ (Kim et al., 2018). Unlike eIF2, eIF2A does not depend on GTP to deliver Met-tRNA$_i^{Met}$, but may rely on GTP hydrolysis by eIF5B for its release from the ribosome (Adams et al., 1975; Zoll et al., 2002; Kim et al., 2018). Studies in *S. cerevisiae* and *Caenorhabditis elegans* showed slow growth phenotype when eIF5B/iffb-1 was knocked down, and the growth deteriorated even more when both eIF5B and eIF2A were knocked down (Zoll et al., 2002; Kim et al., 2018). However, under conditions like hypoxia, the initiation process could be solely eIF5B dependent as depletion of eIF2A has no effect on protein synthesis (Ho et al., 2018). These studies clearly imply that eIF2A augments eIF5B function during Met-tRNA$_i^{Met}$ delivery, and domain IV of eIF5B that helps stabilize Met-tRNA$_i^{Met}$ during normal conditions could be indispensable for Met-tRNA$_i^{Met}$ delivery under stress conditions.

Role of eIF5B in IRES-Mediated Translation

Internal ribosome entry site (IRES) is a secondary structure present on the mRNA of both viral and cellular origins

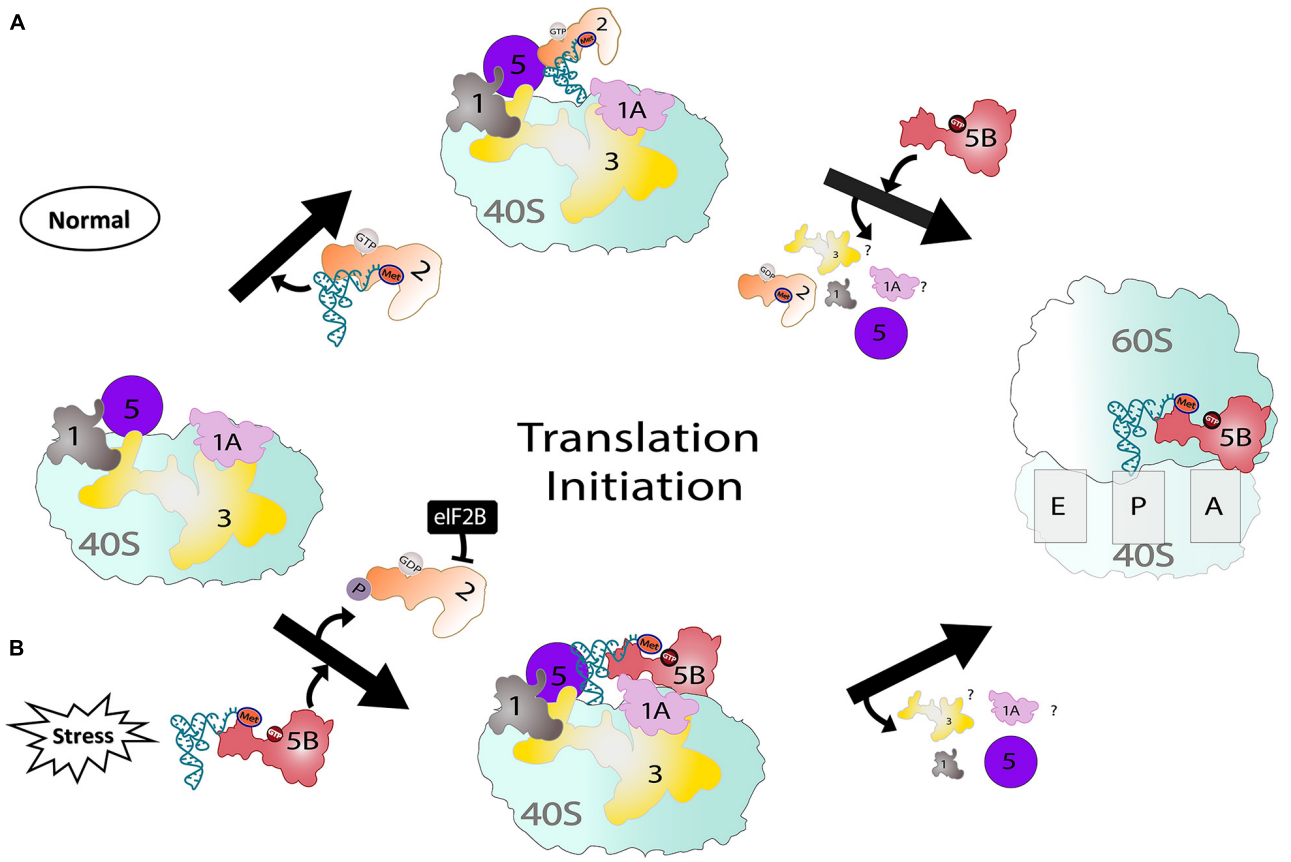


FIGURE 1 | Two pathways showing eIF2-dependent (normal) and eIF5B-dependent translation (stress). **(A)** Under normal conditions, eIF2 delivers Met-tRNA$_i^{Met}$ to the ribosome and, upon release of eIF2, eIF5B stabilizes Met-tRNA$_i^{Met}$. Once 48S PIC forms, eIF5B assists in large and small ribosomal subunit joining. **(B)** Under stress conditions, eIF2 is phosphorylated and sequestered by eIF2B. However, eIF5B parallels the role of eIF2 and delivers Met-tRNA$_i^{Met}$ to the P-site of the 40S ribosome in addition to its usual function during translation.

(Pestova et al., 2008; Thakor and Holcik, 2012; Sharma et al., 2016). These IRES elements can recruit ribosomes directly without a requirement for the 5′-m7G cap during certain stress conditions (Holcik et al., 2000; Thakor and Holcik, 2012; Sharma et al., 2016). Several viral mRNAs contain IRES elements, including but not limited to hepatitis C virus (HCV), classical swine fever virus (CSFV), poliovirus (PV), and coxsackie B virus (CBV) (Pestova et al., 2008; Shatsky et al., 2008; Terenin et al., 2008; Yamamoto et al., 2014). CSFV and HCV IRES in particular depend on eIF5B for translation in the absence of eIF2 (**Figure 1B**; Pestova et al., 2008; Shatsky et al., 2008; Yamamoto et al., 2014). X-linked inhibitor of apoptosis (XIAP, a caspase inhibitor in eukaryotes) mRNA also contains IRES and its translation is eIF5B dependent upon eIF2 sequestration (Yoon et al., 2006; Thakor and Holcik, 2012; Thakor et al., 2017). In general, XIAP is an important anti-apoptotic protein that plays a substantial role in preventing programmed cell death (Perrelet et al., 2000; LaCasse et al., 2008). Nevertheless, IRES elements of viruses such as cricket paralysis virus (CrPV) do not depend on eIF5B for translation (Deniz et al., 2009; Kerr et al., 2016). Thus, although there is strong evidence that eIF5B is involved in IRES mediated translation of several viral and anti-apoptotic mRNA, not all IRES-containing mRNAs require eIF5B for translation initiation.

Regulation of uORF Mediated Translation by eIF5B

Upstream open reading frame (uORF) elements are present in the 5′ untranslated regions (UTR) of mRNA across different species (Chew et al., 2016). These elements inhibit mRNA translation under normal conditions and promote their translation during stress conditions that induce eIF2 phosphorylation (**Figures 2A,B**; Dever et al., 1992; Barbosa et al., 2013; Pakos-Zebrucka et al., 2016; Ross et al., 2018; Chen and Tarn, 2019). For example, *activating transcriptional factor 4* (*ATF4*) has two uORF regions that keep protein levels low until phosphorylation of α-subunit of eIF2 (**Figure 2**; Holcik, 2015; Sharma et al., 2016). During *ATF4* mRNA translation under normal conditions, eIF5B prevents leaky mRNA scanning and ribosomes are recruited onto uORFs, which represses the translation of *ATF4* ORF (Ross et al., 2018). In contrast, under stress or eIF5B depleted conditions, ribosomal recruitment

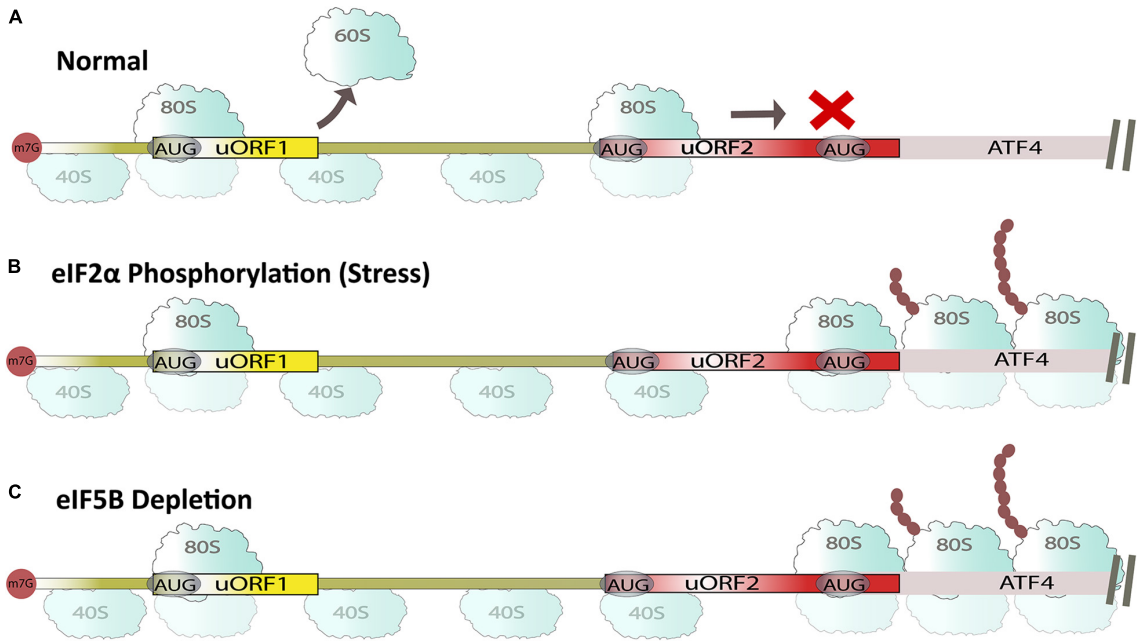


FIGURE 2 | Summary of uORF-mediated ATF4 translation. **(A)** Under non-aberrant normal conditions, ribosomes are recruited on uORF regions, repressing ATF4 expression in the presence of eIF5B. **(B)** Under aberrant stress conditions, ribosomes scan past uORF2, resulting in the synthesis of ATF4 in the presence of eIF5B. **(C)** eIF5B depletion in normal cells also promotes ATF4 synthesis.

occurs on the main ORF AUG sequence, causing de-repression of *ATF4* mRNA translation (**Figure 2C**; Ross et al., 2018). This in turn leads to the upregulation of other proteins like C/EBP homologous protein (CHOP) and growth arrest and DNA damage-inducible protein (GADD34) (Pakos-Zebrucka et al., 2016). This was further corroborated by our recent study on human embryonic kidney 293T (HEK-293T) cells, where depleting eIF5B induced ER stress upregulating mRNA and protein levels of CHOP and GADD34 (Bressler et al., 2021). eIF5B requires cooperativity with initiation factors eIF5 and eIF1A to repress ATF4 expression, and the mechanism of repression is predominantly dependent on uORF2 (Ross et al., 2018). Apart from ATF4, general control non-depressible 4 (GCN4), a transcription factor containing 4 uORFs, is repressed both in starved and non-starved yeast cells in the absence of eIF5B (Shin et al., 2002; Murakami et al., 2018). Similarly, knocking down eIF5B decreased the levels of programmed death-ligand 1 (PD-L1) in heme starved non-small cell lung cancer (NSLC) and lewis lung carcinoma (LLC) cells (Suresh et al., 2020). In contrast, eIF5B negatively regulates p21 and p27 in serum starved THP1 cells, which have been linked to cell cycle progression, and anti-apoptosis (Lee et al., 2014). Numerous uORF dependent proteins have been identified whose expression is regulated by eIF5B, and many of these proteins have implications in cancer resistance and cell cycle. Thus, regulating eIF5B could impact such pathways necessary for cancer survival.

eIF5B Promotes Cancer Cell Survival

Previous studies suggest eIF5B mediates the IRES containing subset of mRNA translation to resist apoptosis in cancer cells (Ross et al., 2019). Translation of IRES-containing mRNAs, which encode anti-apoptotic proteins such as XIAP, Bcl-xL, and Cellular Inhibitor of Apoptosis Protein 1 (cIAP1) as well as Nuclear factor erythroid 2-related factor 2 (Nrf2), are regulated by eIF5B in glioblastoma multiforme (GBM) cells (Ross et al., 2019, 2020). There is also growing evidence of high eIF5B expression levels in various malignancies like GBM, lung adenocarcinoma (LUAD) and hepatocellular carcinoma (HCC), signifying the importance of eIF5B as the stress-related tumorigenic eIF (Wang et al., 2016; Ross et al., 2019; Suresh et al., 2020). In fact, high eIF5B levels were associated with poor prognosis for HCC patients, while low eIF5B levels resulted in smaller tumor sizes, lower vascular invasions, and better survival rates (Wang et al., 2016). Studies on GBM cells show eIF5B depletion leads to reduced cell growth due to inhibition of the NF-κB pathway and sensitization to temozolomide (TMZ)-mediated apoptosis (Ross et al., 2019, 2020). Further, eIF5B aids the survival of GBM cells under hypoxic conditions by acting as one of the essential translational factors for the synthesis of hypoxia-response proteins and regulates carbon metabolism (Ho et al., 2018). In HCC, eIF5B indirectly promotes metastasis and proliferation by upregulating ArfGAP with SH3 Domain, Ankyrin Repeat and PH Domain 1 (ASAP1) expression both *in vivo* and *in cellulo* (Wang et al., 2016). Depletion of eIF5B in maraba virus infected and uninfected U2OS cells resulted in reduced Bcl-xL expression at both the transcriptional and translational levels (Hassanzadeh et al., 2019). This suggests eIF5B plays a role in regulating apoptosis during oncolytic virus treatment (Hassanzadeh et al., 2019). Depletion of eIF5B reduced tumor mass and propagation in lewis lung carcinoma (LLC)

and non-small cell lung cancer (NSLC) cell lines, respectively, demonstrating the dependence of cancer cells on eIF5B for growth and proliferation (Suresh et al., 2020). In line with these findings, high levels of eIF5B under heme depletion induce translation of integrated stress response (ISR) dependent PD-L1, which inhibits T-cell activity. This clearly indicates that eIF5B promotes the survival of LLC and NSLC malignancies (Suresh et al., 2020). Additionally, under serum deprivation in the THP1 cell line, an acute increase in eIF5B levels is observed (Lee et al., 2014). When eIF5B is depleted, global translation is reserved and early G0 phase prohibition occurs, indicating the regulatory role of eIF5B in the cell cycle (Lee et al., 2014). eIF5B also regulates developmental pathways, in particular the mammalian target of rapamycin (mTOR) and mitogen-activated protein kinase (MAPK) pathways, which are activated by epidermal growth factor receptor (EGFR) (Jiang et al., 2016). Thus, these studies highlight that supressing the activity of eIF5B disrupts many pathways, alluding to its importance in oncogenesis.

Considering its pivotal role in cancer cells, eIF5B is emerging as a therapeutic target for cancer treatment. Numerous small molecules and proteins have been identified that could inhibit the activity of eIF5B. For example, a small molecule denoted LWW31 was suggested to inhibit eIF5B activity and lower the viability of cancer cells (Wu et al., 2016). Similarly, ribavirin triphosphate, a guanosine triphosphate analog, has been hypothesized to inhibit the activity of eIF5B (Galmozzi et al., 2012). Additionally, proteins like Puf6p and HIV-1 matrix have been identified, which have shown to repress the function of eIF5B in yeast and human, respectively (Wilson et al., 1999; Deng et al., 2008). These promising works highlight that further research is required to screen and identify clinically relevant small molecules and peptides for targeting eIF5B in "hard-to-treat" and "high-fatality" cancers such as GBM.

CONCLUDING REMARKS

The research works summarized in this review clearly illustrate that eIF5B is a crucial factor for canonical translation initiation. Its role in uORF- and IRES-mediated translation initiation has also been unequivocally established. To this end, the ability of eIF5B to interact with eIF5, eIF2A, and eIF1A as well as to bind and deliver initiator tRNA is critical for mRNA translation. eIF5B is overexpressed in several malignancies and its aberrant expression has been linked to glioblastoma, lung carcinoma, and hepatocellular carcinoma. Due to egregious levels of eIF5B and its conspicuous role in non-canonical translation, eIF5B is an important initiation factor for oncogenesis. It would be ideal to target eIF5B with the goal to regulate non-canonical translation using new small molecules. To achieve this goal, multi-omics (transcriptome, metabolome, as well as translatome and proteome) studies in cancer models are required to establish eIF5B as a biomarker for certain types of cancer. Additionally, fundamental biomedical and pre-clinical studies are necessary to establish eIF5B as a therapeutic target for cancer treatments. Finally, drug discovery research that includes the integration of computer-aided drug design (CADD) with machine learning and is complimented by traditional wet bench experiments must be done. Cancers with upregulated eIF5B have been challenging to treat, and this could change if druggability of eIF5B is further explored as it is a viable therapeutic target.

AUTHOR CONTRIBUTIONS

NT proposed the idea for mini-review. PC wrote the manuscript. NT and SW edited the manuscript and assisted with the figure concepts. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the New Frontiers in Research Fund-Exploration grant (NFRFE-2019-01047), the Natural Sciences and Engineering Research Council of Canada-Discovery Grant Program (RGPIN-2017-05463), the Canada Foundation for Innovation-John R. Evans Leaders Fund (35017 and 2016-04568), Canada Foundation of Innovation (22770), and the Board of Governors Research Chair Program of the University of Lethbridge (UofL). PC thanks RNA innovation for a student fellowship, which was partially supported by the NSERC-CREATE grant (510937-2018).

ACKNOWLEDGMENTS

We thank Dr. Mehdi Jafarnejad, Katrina Tylor, and Kush Patel for critically reading this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.737433/full#supplementary-material

REFERENCES

Acker, M. G., Shin, B.-S., Dever, T. E., and Lorsch, J. R. (2006). Interaction between eukaryotic initiation factors 1A and 5B is required for efficient ribosomal subunit joining. *J. Biol. Chem.* 281, 8469–8475. doi: 10.1074/jbc.m600210200

Adams, S. L., Safer, B., Anderson, W. F., and Merrick, W. C. (1975). Eukaryotic initiation complex formation. Evidence for two distinct pathways. *J. Biol. Chem.* 250, 9083–9089. doi: 10.1016/s0021-9258(19)40696-0

Barbosa, C., Peixeiro, I., and Romao, L. (2013). Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9:e1003529. doi: 10.1371/journal.pgen.1003529

Bressler, K. R., Ross, J. A., Ilnytskyy, S., Vanden Dungen, K., Taylor, K., Patel, K., et al. (2021). Depletion of eukaryotic initiation factor 5B (eIF5B) reprograms the cellular transcriptome and leads to activation of endoplasmic reticulum (ER) stress and c-Jun N-terminal kinase (JNK). *Cell Stress Chaperones* 26, 253–264. doi: 10.1007/s12192-020-01174-1

Chen, H. H., and Tarn, W. Y. (2019). uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol.* 16, 1327–1338. doi: 10.1080/15476286.2019.1632634

Chew, G. L., Pauli, A., and Schier, A. F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* 7:11663.

Choi, S. K., Lee, J. H., Zoll, W. L., Merrick, W. C., and Dever, T. E. (1998). Promotion of met-tRNAiMet binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. *Science* 280, 1757–1760. doi: 10.1126/science.280.5370.1757

Deng, Y., Singer, R. H., and Gu, W. (2008). Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. *Genes Dev.* 22, 1037–1050. doi: 10.1101/gad.1611308

Deniz, N., Lenarcic, E. M., Landry, D. M., and Thompson, S. R. (2009). Translation initiation factors are not required for Dicistroviridae IRES function in vivo. *RNA* 15, 932–946. doi: 10.1261/rna.1315109

Dever, T. E., Feng, L., Wek, R. C., Cigan, A. M., Donahue, T. F., and Hinnebusch, A. G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* 68, 585–596. doi: 10.1016/0092-8674(92)90193-g

Fernandez, I. S., Bai, X. C., Hussain, T., Kelley, A. C., Lorsch, J. R., Ramakrishnan, V., et al. (2013). Molecular architecture of a eukaryotic translational initiation complex. *Science* 342:1240585. doi: 10.1126/science.1240585

Fringer, J. M., Acker, M. G., Fekete, C. A., Lorsch, J. R., and Dever, T. E. (2007). Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining. *Mol. Cell Biol.* 27, 2384–2397. doi: 10.1128/mcb.02254-06

Galmozzi, E., Aghemo, A., and Colombo, M. (2012). Eukaryotic initiation factor 5B: a new player for the anti-hepatitis C virus effect of ribavirin?. *Med. Hypotheses* 79, 471–473. doi: 10.1016/j.mehy.2012.06.026

Hassanzadeh, G., Naing, T., Graber, T., Jafarnejad, S. M., Stojdl, D. F., Alain, T., et al. (2019). Characterizing Cellular Responses During Oncolytic Maraba Virus Infection. *Int. J. Mol. Sci.* 20:580. doi: 10.3390/ijms20030580

Ho, J. J. D., Balukoff, N. C., Cervantes, G., Malcolm, P. D., Krieger, J. R., and Lee, S. (2018). Oxygen-Sensitive Remodeling of Central Carbon Metabolism by Archaic eIF5B. *Cell Rep.* 22, 17–26. doi: 10.1016/j.celrep.2017.12.031

Holcik, M. (2015). Could the eIF2alpha-Independent Translation Be the Achilles Heel of Cancer? *Front. Oncol.* 5:264. doi: 10.3389/fonc.2015.00264

Holcik, M., Sonenberg, N., and Korneluk, R. G. (2000). Internal ribosome initiation of translation and the control of cell death. *Trends Genet.* 16, 469–473. doi: 10.1016/s0168-9525(00)02106-5

Huang, B. Y., and Fernandez, I. S. (2020). Long-range interdomain communications in eIF5B regulate GTP hydrolysis and translation initiation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 1429–1437. doi: 10.1073/pnas.1916436117

Jennings, M. D., Kershaw, C. J., White, C., Hoyle, D., Richardson, J. P., Costello, J. L., et al. (2016). eIF2beta is critical for eIF5-mediated GDP-dissociation inhibitor activity and translational control. *Nucleic Acids Res.* 44, 9698–9709.

Jennings, M. D., and Pavitt, G. D. (2010). eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. *Nature* 465, 378–381. doi: 10.1038/nature09003

Jiang, X., Jiang, X., Feng, Y., Xu, R., Wang, Q., and Deng, H. (2016). Proteomic Analysis of eIF5B Silencing-Modulated Proteostasis. *PLoS One* 11:e0168387. doi: 10.1371/journal.pone.0168387

Kerr, C. H., Ma, Z. W., Jang, C. J., Thompson, S. R., and Jan, E. (2016). Molecular analysis of the factorless internal ribosome entry site in Cricket Paralysis virus infection. *Sci. Rep.* 6:37319.

Kim, E., Kim, J. H., Seo, K., Hong, K. Y., An, S. W. A., Kwon, J., et al. (2018). eIF2A, an initiator tRNA carrier refractory to eIF2α kinases, functions synergistically with eIF5B. *Cell. Mol. Life Sci.* 75, 4287–4300. doi: 10.1007/s00018-018-2870-4

Kuhle, B., and Ficner, R. (2014). eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *EMBO J.* 33, 1177–1191. doi: 10.1002/embj.201387344

LaCasse, E. C., Mahoney, D. J., Cheung, H. H., Plenchette, S., Baird, S., and Korneluk, R. G. (2008). IAP-targeted therapies for cancer. *Oncogene* 27, 6252–6275. doi: 10.1038/onc.2008.302

Lebaron, S., Schneider, C., Van Nues, R. W., Swiatkowska, A., Walsh, D., Bottcher, B., et al. (2012). Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nat. Struct. Mol. Biol.* 19, 744–753. doi: 10.1038/nsmb.2308

Lee, J. H., Choi, S. K., Roll-Mecak, A., Burley, S. K., and Dever, T. E. (1999). Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4342–4347. doi: 10.1073/pnas.96.8.4342

Lee, J. H., Pestova, T. V., Shin, B. S., Cao, C., Choi, S. K., and Dever, T. E. (2002). Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proc. Natl. Acad. Sci. U. S. A* .99, 16689–16694. doi: 10.1073/pnas.262569399

Lee, S., Truesdell, S. S., Bukhari, S. I., Lee, J. H., Letonqueze, O., and Vasudevan, S. (2014). Upregulation of eIF5B controls cell-cycle arrest and specific developmental stages. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4315–E4322.

Lin, K. Y., Nag, N., Pestova, T. V., and Marintchev, A. (2018). Human eIF5 and eIF1A Compete for Binding to eIF5B. *Biochemistry* 57, 5910–5920. doi: 10.1021/acs.biochem.8b00839

Majumdar, R., and Maitra, U. (2005). Regulation of GTP hydrolysis prior to ribosomal AUG selection during eukaryotic translation initiation. *EMBO J.* 24, 3737–3746. doi: 10.1038/sj.emboj.7600844

Merrick, W. C., Kemper, W. M., and Anderson, W. F. (1975). Purification and characterization of homogeneous initiation factor M2A from rabbit reticulocytes. *J. Biol. Chem.* 250, 5556–5562. doi: 10.1016/s0021-9258(19)41216-7

Murakami, R., Singh, C. R., Morris, J., Tang, L., Harmon, I., Takasu, A., et al. (2018). The Interaction between the Ribosomal Stalk Proteins and Translation Initiation Factor 5B Promotes Translation Initiation. *Mol. Cell Biol.* 38, e00067–18.

Nag, N., Lin, K. Y., Edmonds, K. A., Yu, J., Nadkarni, D., Marintcheva, B., et al. (2016). eIF1A/eIF5B interaction network and its functions in translation initiation complex assembly and remodeling. *Nucleic Acids Res.* 44, 7441–7456.

Olsen, D. S., Savner, E. M., Mathew, A., Zhang, F., Krishnamoorthy, T., Phan, L., et al. (2003). Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. *EMBO J.* 22, 193–204. doi: 10.1093/emboj/cdg030

Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A., and Gorman, A. M. (2016). The integrated stress response. *EMBO Rep.* 17, 1374–1395.

Paulin, F. E. M., Campbell, L. E., O'brien, K., Loughlin, J., and Proud, C. G. (2001). Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr. Biol.* 11, 55–59. doi: 10.1016/s0960-9822(00)00025-7

Perrelet, D., Ferri, A., Mackenzie, A. E., Smith, G. M., Korneluk, R. G., Liston, P., et al. (2000). IAP family proteins delay motoneuron cell death in vivo. *Eur. J. Neurosci.* 12, 2059–2067. doi: 10.1046/j.1460-9568.2000.00098.x

Pestova, T. V., De Breyne, S., Pisarev, A. V., Abaeva, I. S., and Hellen, C. U. (2008). eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J.* 27, 1060–1072. doi: 10.1038/emboj.2008.49

Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., et al. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7029–7036.

Pestova, T. V., Lomakin, I. B., Lee, J. H., Choi, S. K., Dever, T. E., and Hellen, C. U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403, 332–335. doi: 10.1038/35002118

Pisarev, A. V., Kolupaeva, V. G., Pisareva, V. P., Merrick, W. C., Hellen, C. U., and Pestova, T. V. (2006). Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* 20, 624–636. doi: 10.1101/gad.1397906

Pisareva, V. P., and Pisarev, A. V. (2014). eIF5 and eIF5B together stimulate 48S initiation complex formation during ribosomal scanning. *Nucleic Acids Res.* 42, 12052–12069. doi: 10.1093/nar/gku877

Roll-Mecak, A., Cao, C., Dever, T. E., and Burley, S. K. (2000). X-Ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. *Cell* 103, 781–792. doi: 10.1016/s0092-8674(00)00181-1

Ross, J. A., Ahn, B. Y., King, J., Bressler, K. R., Senger, D. L., and Thakor, N. (2020). Eukaryotic initiation factor 5B (eIF5B) regulates temozolomide-mediated apoptosis in brain tumour stem cells (BTSCs). *Biochem. Cell Biol.* 98, 647–652. doi: 10.1139/bcb-2019-0329

Ross, J. A., Bressler, K. R., and Thakor, N. (2018). Eukaryotic Initiation Factor 5B (eIF5B) Cooperates with eIF1A and eIF5 to Facilitate uORF2-Mediated Repression of ATF4 Translation. *Int. J. Mol. Sci.* 19:4032. doi: 10.3390/ijms19124032

Ross, J. A., Dungen, K. V., Bressler, K. R., Fredriksen, M., Khandige Sharma, D., Balasingam, N., et al. (2019). Eukaryotic initiation factor 5B (eIF5B) provides a critical cell survival switch to glioblastoma cells via regulation of apoptosis. *Cell Death Dis.* 10:57. doi: 10.1038/s41419-018-1283-5

Sharma, D. K., Bressler, K., Patel, H., Balasingam, N., and Thakor, N. (2016). Role of Eukaryotic Initiation Factors during Cellular Stress and Cancer Progression. *J. Nucleic Acids* 2016:8235121. doi: 10.1155/2016/8235121

Shatsky, I., Terenin, I., Dmitriev, S., and Andreev, D. (2008). Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.* 15, 836–841. doi: 10.1038/nsmb.1445

Shin, B.-S., Acker, M. G., Kim, J.-R., Maher, K. N., Arefin, S. M., Lorsch, J. R., et al. (2011). Structural integrity of {alpha}-helix H12 in translation initiation factor eIF5B is critical for 80S complex stability. *RNA* 17, 687–696. doi: 10.1261/rna.2412511

Shin, B. S., Maag, D., Roll-Mecak, A., Arefin, M. S., Burley, S. K., Lorsch, J. R., et al. (2002). Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. *Cell* 111, 1015–1025. doi: 10.1016/S0092-8674(02)01171-6

Sorensen, H. P., Hedegaard, J., Sperling-Petersen, H. U., and Mortensen, K. K. (2001). Remarkable conservation of translation initiation factors: IF1/eIF1A and IF2/eIF5B are universally distributed phylogenetic markers. *IUBMB Life* 51, 321–327. doi: 10.1080/152165401317190842

Strunk, B. S., Novak, M. N., Young, C. L., and Karbstein, K. (2012). A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* 150, 111–121. doi: 10.1016/j.cell.2012.04.044

Suresh, S., Chen, B., Zhu, J., Golden, R. J., Lu, C., Evers, B. M., et al. (2020). eIF5B drives integrated stress response-dependent translation of PD-L1 in lung cancer. *Nat. Cancer* 1, 533–545. doi: 10.1038/s43018-020-0056-0

Terenin, I. M., Dmitriev, S. E., Andreev, D. E., and Shatsky, I. N. (2008). Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.* 15, 836–841. doi: 10.1038/nsmb.1445

Thakor, N., and Holcik, M. (2012). IRES-mediated translation of cellular messenger RNA operates in eIF2alpha- independent manner during stress. *Nucleic Acids Res.* 40, 541–552. doi: 10.1093/nar/gkr701

Thakor, N., Smith, M. D., Roberts, L., Faye, M. D., Patel, H., Wieden, H. J., et al. (2017). Cellular mRNA recruits the ribosome via eIF3-PABP bridge to initiate internal translation. *RNA Biol.* 14, 553–567. doi: 10.1080/15476286.2015.1137419

Unbehaun, A., Borukhov, S. I., Hellen, C. U., and 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 Dev.* 18, 3078–3093. doi: 10.1101/gad.1255704

Wang, J., Johnson, A. G., Lapointe, C. P., Choi, J., Prabhakar, A., Chen, D. H., et al. (2019). eIF5B gates the transition from translation initiation to elongation. *Nature* 573, 605–608. doi: 10.1038/s41586-019-1561-0

Wang, J., Wang, J., Shin, B. S., Kim, J. R., Dever, T. E., Puglisi, J. D., et al. (2020). Structural basis for the transition from translation initiation to elongation by an 80S-eIF5B complex. *Nat. Commun.* 11:5003.

Wang, Z. G., Zheng, H., Gao, W., Han, J., Cao, J. Z., Yang, Y., et al. (2016). eIF5B increases ASAP1 expression to promote HCC proliferation and invasion. *Oncotarget* 7, 62327–62339. doi: 10.18632/oncotarget.11469

Wilson, S. A., Sieiro-Vazquez, C., Edwards, N. J., Iourin, O., Byles, E. D., Kotsopoulou, E., et al. (1999). Cloning and characterization of hIF2, a human homologue of bacterial translation initiation factor 2, and its interaction with HIV-1 matrix. *Biochem. J.* 342, 97–103. doi: 10.1042/bj3420097

Woolford, J. L. Jr., and Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics* 195, 643–681. doi: 10.1534/genetics.113.153197

Wu, C. Y., Wang, D. H., Wang, X., Dixon, S. M., Meng, L., Ahadi, S., et al. (2016). Rapid Discovery of Functional Small Molecule Ligands against Proteomic Targets through Library-Against-Library Screening. *ACS Comb. Sci.* 18, 320–329. doi: 10.1021/acscombsci.5b00194

Yamamoto, H., Unbehaun, A., Loerke, J., Behrmann, E., Collier, M., Burger, J., et al. (2014). Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nat. Struct. Mol. Biol.* 21, 721–727. doi: 10.1038/nsmb.2859

Yoon, A., Peng, G., Brandenburger, Y., Zollo, O., Xu, W., Rego, E., et al. (2006). Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* 312, 902–906. doi: 10.1126/science.1123835

Yu, Y., Marintchev, A., Kolupaeva, V. G., Unbehaun, A., Veryasova, T., Lai, S. C., et al. (2009). Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res.* 37, 5167–5182. doi: 10.1093/nar/gkp519

Zoll, W. L., Horton, L. E., Komar, A. A., Hensold, J. O., and Merrick, W. C. (2002). Characterization of mammalian eIF2A and identification of the yeast homolog. *J. Biol. Chem.* 277, 37079–37087. doi: 10.1074/jbc.M207109200

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

MINI REVIEW
published: 10 January 2022
doi: 10.3389/fgene.2021.821535

Bacterial Response to Oxidative Stress and RNA Oxidation

*André F. Seixas, Ana P. Quendera, João P. Sousa, Alda F. Q. Silva, Cecília M. Arraiano and José M. Andrade**

Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Bacteria have to cope with oxidative stress caused by distinct Reactive Oxygen Species (ROS), derived not only from normal aerobic metabolism but also from oxidants present in their environments. The major ROS include superoxide O_2^-, hydrogen peroxide H_2O_2 and radical hydroxide $HO^\bullet$. To protect cells under oxidative stress, bacteria induce the expression of several genes, namely the SoxRS, OxyR and PerR regulons. Cells are able to tolerate a certain number of free radicals, but high levels of ROS result in the oxidation of several biomolecules. Strikingly, RNA is particularly susceptible to this common chemical damage. Oxidation of RNA causes the formation of strand breaks, elimination of bases or insertion of mutagenic lesions in the nucleobases. The most common modification is 8-hydroxyguanosine (8-oxo-G), an oxidized form of guanosine. The structure and function of virtually all RNA species (mRNA, rRNA, tRNA, sRNA) can be affected by RNA oxidation, leading to translational defects with harmful consequences for cell survival. However, bacteria have evolved RNA quality control pathways to eliminate oxidized RNA, involving RNA-binding proteins like the members of the MutT/Nudix family and the ribonuclease PNPase. Here we summarize the current knowledge on the bacterial stress response to RNA oxidation, namely we present the different ROS responsible for this chemical damage and describe the main strategies employed by bacteria to fight oxidative stress and control RNA damage.

Keywords: nucleotide modification, oxidative stress, 8-oxo-G, RNA oxidation, ROS, quality control of damaged RNA

OPEN ACCESS

Edited by:
Jane E. A. Reid, Australian National University, Australia

Reviewed by:
Indra Mani Sharma, National Cancer Institute at Frederick, United States

****Correspondence:***
José M. Andrade
andrade@itqb.unl.pt

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *24 November 2021*
Accepted: *21 December 2021*
Published: *10 January 2022*

Citation:
Seixas AF, Quendera AP, Sousa JP, Silva AFQ, Arraiano CM and Andrade JM (2022) Bacterial Response to Oxidative Stress and RNA Oxidation. Front. Genet. 12:821535. doi: 10.3389/fgene.2021.821535

INTRODUCTION

Oxidative Stress and Reactive Oxygen Species

Oxygen is an abundant component of the earth's atmosphere and a crucial player in many forms of life. However, oxygen comes with a risk that is the production of highly reactive oxygen species (ROS), which chemically modify by oxidation a variety of macromolecules (RNA, DNA, proteins, and lipids) (Markkanen 2017). Consequently, the structure and therefore the function of these macromolecules can be affected, usually resulting in cell toxicity. In bacteria and other organisms, ROS can be either generated intracellularly, as result of aerobic metabolism or exogenously from the outside environment, as consequence of local exposure to increased levels of oxidative agents (Dubbs and Mongkolsuk 2012; Fasnacht and Polacek 2021). Cells possess mechanisms to counteract oxidation and can tolerate low levels of ROS. Indeed, at low levels, ROS can act as signaling molecules controlling different cellular processes, such as quorum sensing, biofilm formation or bacterial self-destruction (Zhao and Drlica 2014; Dryden et al., 2017). However, when there is an imbalance between the amount of ROS and the ability to eliminate them, cells suffer oxidative stress

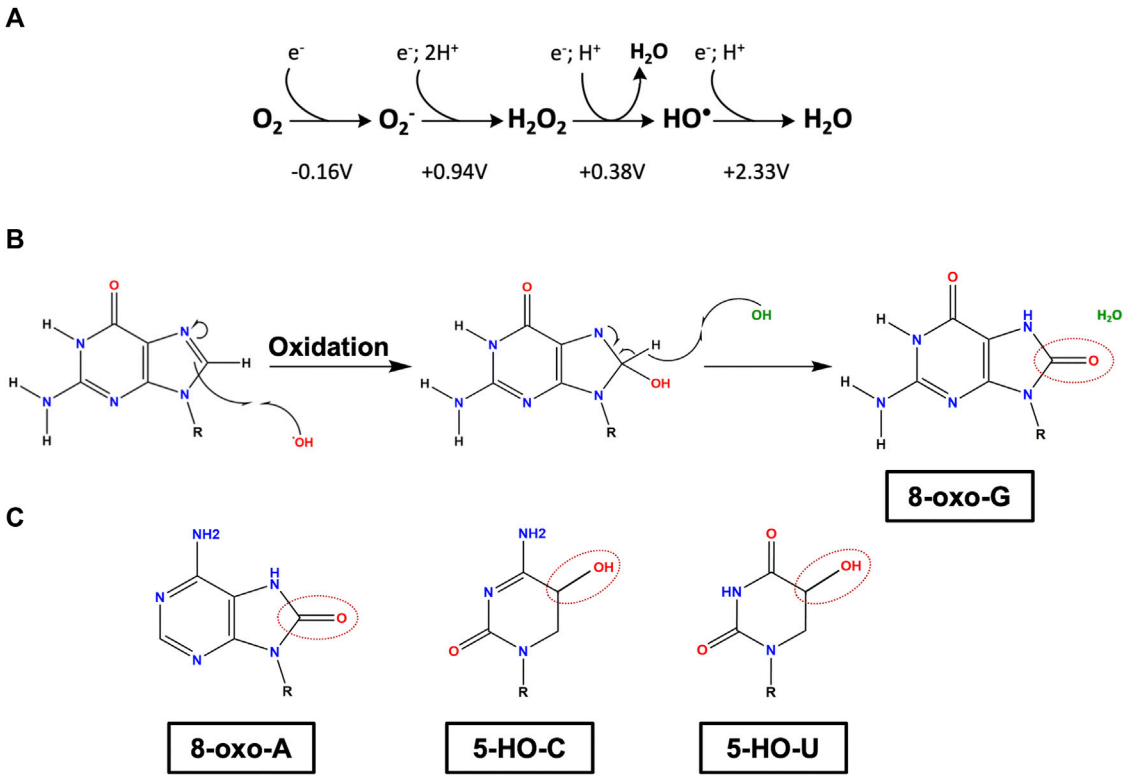


FIGURE 1 | Main oxidative lesions found in RNA. **(A)** Redox state of molecular oxygen. From left to right: molecular oxygen (O_2), superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radical ($HO^•$) and water (H_2O). The reduction potentials are shown regarding the standard concentration of O_2 to be 1M at pH 7 (adapted from Imlay 2009). **(B)** Reaction of a guanosine with a hydroxyl radical, yielding 8-oxo-7,8-dihydroguanosine (8-oxo-G). **(C)** Three modified RNA nucleobases: oxidized adenosine 8-oxo-7,8-dihydroadenosine (8-oxo-A), oxidized cytosine 5-hydroxycytosine (5-HO-C) and oxidized uridine 5-hydroxyuridine (5-HO-U). Oxidation sites are marked as dashed circles. R represents the ribose group.

(Berghoff and Klug 2012). Three naturally occurring ROS species, superoxide anion (O_2–), hydrogen peroxide (H_2O_2), and hydroxyl radical ($HO^•$), display different reactivities and are the ones with major relevance in aerobic environments and oxidative stress (**Figure 1A**).

Superoxide Anion O_2^-

O_2^- is a non-diffusible free radical formed when O_2 removes one electron from a donor. It can be produced both intracellularly and exogenously (Kehrer 2000; Imlay 2009). Inside bacterial cells, the major source of O_2^- is the autoxidation of enzymes like dehydrogenases, glutathione reductase and cytochromes P450 (Farr and Kogoma 1991; Gort and Imlay 1998). In eukaryotic cells, the most important source of ROS is the mitochondria *via* the electron transport chain system (Malla and Kwakye 2021). In phagocytes, O_2^- is largely produced by NADPH oxidase and is used for destruction of pathogens (Shatwell and Segal 1996; Kong and Lin 2010). Exogenously O_2^- can come from quinone-like compounds such as plumbagin and menadione or herbicidal paraquat (Farr et al., 1985; Suntres 2002; Criddle et al., 2006). Superoxide anion O_2^- has a half-life in the range of microseconds and it oxidizes several targets like ascorbate and thiols (Kehrer 2000). Additionally, it can be dismutated to H_2O_2 and O_2 through the action of superoxide dismutase (SOD) (Fridovich 1989) and act as a reducing agent of transition metals like Fe^{3+} (Farr and Kogoma 1991).

Hydrogen Peroxide H_2O_2

Hydrogen peroxide is the simplest member of the class of peroxides. The main sources of internal H_2O_2 are the autoxidation of flavoenzymes and the dismutation of O_2^- and contrary to superoxide anion, hydrogen peroxide can easily cross the cellular membrane (Imlay 2009). There are specific environments with residual amounts of oxygen that can be hubs of H_2O_2 accumulation, such as the anoxic interfaces in soil, hydrothermal vents (Tapley et al., 1999; Ogino et al., 2018) and the intestine epithelium (Espey 2013). One possible theory to explain the presence of H_2O_2 in the intestine is its production by lactic acid bacteria (LAB) (Imlay 2009). There is evidence that LAB metabolism can use oxygen and release H_2O_2 to their environment (Seki et al., 2004). Released H_2O_2 can also cause growth arrest of other bacterial species (Pericone et al., 2000; Tong et al., 2007; Bao et al., 2017). H_2O_2 is more stable when compared to O_2^- with a half-life in the range of minutes (Kehrer 2000), and its decomposition is mediated by peroxidases like catalase (Winterbourn 2013). The most relevant reaction in which H_2O_2 is involved is undoubtedly the Fenton reaction (Fenton 1894). In this reaction, H_2O_2 interacts with metal

cofactors such as iron (Fe^{2+}) producing the highly reactive hydroxyl radicals $HO^{\bullet}$ (Shcherbik and Pestov 2019).

Hydroxyl Radicals $HO^{\bullet}$

$HO^{\bullet}$ is one of the most dangerous ROS due to a half-life in the range of nanoseconds and its extreme reactivity given the high electron potential of +2.33 V (**Figure 1A**). It can attack most of the organic molecules in the vicinity of its formation, reacting either with proteins, lipids, or nucleic acids, especially with thiamine and guanosine (Farr and Kogoma 1991). $HO^{\bullet}$ can arise from absorption of a photon by tryptophan residues and from photochemical reactions in natural waters (Vaughan and Blough 1998). However, the most relevant source of $HO^{\bullet}$ in biological systems is the Fenton reaction since iron is such an important micronutrient (Liochev et al., 1999). Note that the ROS levels of the different species are interconnected. As mentioned above, O_2^- reduces Fe^{3+} and its dismutation consequently produces H_2O_2; hence when the intracellular concentration of O_2^- increases, the concentration of H_2O_2 and subsequently $HO^{\bullet}$ also increases. Contrarily to O_2^-, which can be eliminated by SOD, $HO^{\bullet}$ cannot be eliminated by an enzymatic reaction (Reiter et al., 1995).

Reactive Oxygen Species and Cell Damage

An increase in ROS species is perilous to the cell as they cause an extensive damage to several targets, which include all biomolecules like proteins, lipids, and nucleic acids (DNA and RNA). As result of oxidation, the structure and function of these macromolecules are affected, with potentially toxic consequences that can ultimately result in cell death. Most protein oxidative modifications occur in the side chains of amino acids like oxidation of thiols and the formation of carbonyl groups affecting protein structure, conformation, and function (Dalle-Donne et al., 2006; Morzel et al., 2006; Zhang et al., 2013). Cysteine and methionine are the most susceptible amino acids to oxidation because they contain reactive sulfur atoms (Shacter 2000). In humans, most oxidized proteins are eliminated but some can accumulate and contribute to the damage associated with several diseases (Stadtman 2001). When ROS attack lipids, the products of this oxidation are lipid peroxides that mostly affect polyunsaturated fatty acids, which are critical components of cellular membranes (Niki 2009; Yin et al., 2011). Consequently, the accumulation of lipid peroxides results in changes in membrane permeability and fluidity (Dobretsov et al., 1977) and can affect ion channels, inactivate membrane transport proteins, disrupt homeostasis, and affect signaling pathways (Welsch 1987; Blair 2008; Su et al., 2019).

Nucleic acids are very susceptible to chemical damage given the reactivity of the nitrogen and oxygen atoms of the nucleobases (Simms and Zaher 2016). Among the four DNA nucleobases, guanine is the most prone to oxidation given its lower reduction potential (Steenken and Jovanovic 1997) and its oxidation can interfere with DNA metabolism like transcription and replication (Lee and Kang 2019). Contrarily to other macromolecules, damaged DNA cannot be replaced so it needs to be repaired to remain viable (Lee and Kang 2019). The main repair mechanism to correct non-bulky DNA lesions caused by oxidative stress is the base excision repair (Ignatov et al., 2017). Unlike DNA, no repair mechanisms have yet been identified for oxidized RNA, which imposes a challenge to the cell in order to keep homeostasis. This is even more important as RNA is more susceptible to oxidation than DNA (see below).

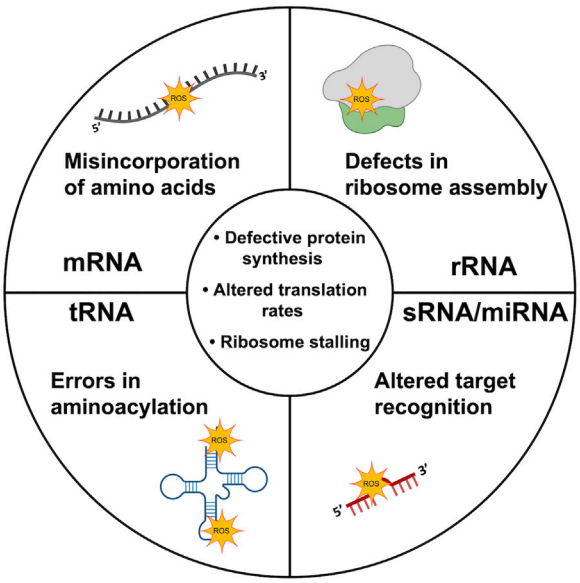


FIGURE 2 | Oxidation of different RNA species. Virtually all RNA species are vulnerable to RNA oxidation. The scheme depicts major consequences of oxidation damage of rRNA, mRNA, tRNA and sRNA/miRNA molecules. In the middle of the scheme are represented defects common to oxidation of all RNA molecules. Molecule templates were retrieved from BioRender database.

RNA OXIDATION

RNA accounts for 80–90% of the total nucleic acids (Hangauer et al., 2013; Li et al., 2020), and it affects the post-transcriptional regulation of gene expression (dos Santos et al., 2018; Quendera et al., 2020). ROS can attack RNA in different ways, whether it is through formation of abasic site, strand breaks or modifications of base and sugar moieties (Liu et al., 2012; Küpfer and Leumann 2014; Tanaka and Chock 2021). When compared to DNA, RNA molecules are more prone to oxidation for several reasons: 1) Being mostly single-stranded, the nucleobases of the RNA molecules are more exposed to ROS (Hofer et al., 2005); 2) RNA is less associated with proteins compared to DNA, making it more vulnerable (Cobley et al., 2018); 3) In physiological conditions the concentration of ribonucleotides is much higher than deoxyribonucleotides; therefore, the modifications are more likely to be incorporated into RNA than in DNA (Yan and Zaher 2019) and 4) RNA lacks repair mechanisms in contrast to what happens with DNA (Küpfer and Leumann 2014).

RNA molecules can be oxidized by $HO^{\bullet}$ derived from the Fenton reaction by detaching the hydrogen from the C-H

bonds, resulting in several oxidation adducts, namely: 8-oxo-7,8- dihydroguanosine (8-oxo-G), 8-oxo-7,8-dihydroadenosine (8-oxo-A), 5-hydroxyuridine (5-HO-U) and 5-hydrocytosine (5-HO-C) (Simms and Zaher 2016; Yan and Zaher 2019) (**Figures 1B,C**). Although all four nucleobases are affected by ROS, the most common oxidation lesion found in cells is the 8-oxo-G given that guanosine is the base with the lowest reduction potential (Radak and Boldogh 2010; Shcherbik and Pestov 2019; Estevez et al., 2021). Under physiological conditions, the frequency of 8-oxo-G is approximately of 1 per 100.000 of unmodified guanosines (Shen et al., 2000). This modification has a high propensity to cause mutations in the open reading frame leading to changes in gene expression and the development of aberrant proteins (Ishii and Sekiguchi 2019; Yan and Zaher 2019). Furthermore, 8-oxo-G can dramatically reduce the rate of peptide-bond formation, strongly impairing protein synthesis (Thomas et al., 2019).

When compared to DNA, the study of RNA damage by oxidation has been neglected, and this is mainly due to the assumption that RNAs tend to have a short half-life due to a rapid turnover before causing deleterious effects (Malla and Kwakye 2021; Tanaka and Chock 2021). This neglection resulted in an underestimation of the toxic effect of RNA oxidation on cells (Li et al., 2020). Next, we will provide examples on how RNA oxidation affects different RNA species within a cell (**Figure 2**).

Ribosomal RNA

Ribosomal RNA composes the vast majority of total RNA molecules within a cell and is responsible for the structural and functional core of the ribosome (Cech 2000; Andrade et al., 2018; Dos Santos et al., 2020). Despite rRNAs being highly structured molecules, it is well established that rRNAs are a target for oxidation, not only in bacteria but also in higher organisms, and this has been associated with diseases (Ahmad et al., 2017; Willi et al., 2018). As we could expect, oxidation-induced modifications in the rRNA will affect important functional sites of the ribosome. In *Escherichia coli* it was shown that both the 16S rRNA (component of the small ribosomal subunit) and the 23S rRNA (component of the large ribosomal subunit) are subject to oxidation by H_2O_2 *in vivo*, although the 23S rRNA seems more prone to nucleobase oxidation (Willi et al., 2018). Specifically, it was shown that modification of three inner residues of the 23S rRNA (A2451, C2063, U2585) located in the catalytic center of the ribosome affects multiple steps of translation, such as peptide bond formation or tRNA binding to the ribosome. Oxidation of rRNA at specific positions within the ribosome can lead to defects in ribosome assembly (Simms and Zaher 2016). In addition, it was recently shown that the absence of methylations in the 23S rRNA at specific positions could make it more susceptible to oxidative stress, suggesting that oxidative damage to RNA can be regulated by the presence of post-transcriptional nucleoside modifications like methylation (Estevez et al., 2021).

Transfer RNA

Transfer RNAs are structured molecules with great abundance and stability in the cell. These molecules decode the different codons on mRNA and allow the correct amino acid to be added to the nascent polypeptide. Certain post-transcriptional modifications normally found in tRNAs like the 5-methylaminomethyl-2-thiouridine (mnm5s2U) were also found to reduce oxidation in tRNAs (Estevez et al., 2021). Oxidation of the tRNA nucleobases is likely to induce a profound effect on the overall structure of the molecule, thus affecting protein synthesis. An *in vitro* study where liquid chromatography tandem mass spectrometry was applied to *E. coli* tRNA showed that upon oxidation induced by UV radiation, several tRNA nucleosides were modified with the 8-oxo-G adduct (Sun et al., 2018). Specifically, oxidative damage was mostly found to affect the anticodon and variable loop regions sequences of tRNA (Sun et al., 2020), which correlates with translational errors in the decoding process and aminoacylation (Yan and Zaher 2019). Zhong et al. (2015) showed that under oxidative stress caused by H_2O_2, the vast majority of tRNAs in *E. coli* were decreased *in vivo*. This was shown to occur due to nonspecific extended tRNA degradation, in contrast to eukaryotic cells where tRNAs are cut in halves (Thompson et al., 2008). However, another report using a different *E. coli* strain revealed the reduction of only $tRNA^{Gly}$ upon oxidation also with H_2O_2, with the global pool of tRNA remaining unchanged (Leiva et al., 2020). Therefore, the extent of tRNAs affected by oxidation is still debatable. The reduction of tRNA levels significantly slows down the translation elongation speed (Zhu and Dai 2019). This could be beneficial for bacteria as oxidative stress increases the probability of protein misfolding, which can be toxic to the cell. Concomitantly, overexpression of tRNAs was found to protect cells against oxidative stress, presumably by decreasing ribosome jamming found upon oxidation treatment, which causes the premature dissociation of RNA from ribosomes (Zhong et al., 2015).

Messenger RNA

Messenger RNA is responsible for carrying the information encoded in DNA to synthesize proteins. The 8-oxo-G modification can affect the translation process in *E. coli* by interfering with the A site of mRNA codons leading to the incorporation of a different amino acid (Simms et al., 2014). Recent studies expanded the mechanism by which 8-oxo-G affects the decoding process. This adduct was found to induce a *syn* conformation of the mRNA molecule, causing it to pair preferentially with a tRNA carrying adenosine instead of cytidine, therefore, disrupting the translation rate (Thomas et al., 2019; Zhu and Dai 2019). From these studies it was also possible to conclude that the mRNA alterations caused by 8-oxo-G seem to affect mRNA-tRNA interactions in the ribosome decoding center. From the three main steps of the translation process, it is the elongation step that seems to be most affected by oxidative stress, with increasing levels of stalled ribosomes found on the mRNA molecule upon oxidation treatment, as observed in *E. coli* and cyanobacteria (Nishiyama et al., 2006; Zhong et al., 2015).

Small Non-coding RNA

Small non-coding RNAs play an important role in the regulation of gene expression (Kavita et al., 2018; Quendera et al., 2020). In the literature there is no current evidence of direct oxidation of sRNAs in prokaryotes, although it seems likely that they are also targets for ROS, as it is the case of microRNAs (miRNAs) in eukaryotes. Namely, the microRNA miR-184 is highly oxidized and enriched in 8-oxo-G nucleotides upon H_2O_2 treatment (Wang et al., 2015). This oxidative modification influences the function of miR-184, which caused a substantial downregulation of Bcl-xL and Bcl-w mRNAs, promoting apoptosis to a higher extent. Recently, it was also reported that the miRNA miR-1 is preferentially oxidized in cardiac hypertrophy cases (Seok et al., 2020). Therefore, it is reasonable to think that like in miRNAs, bacterial sRNAs can also be oxidized under specific conditions. Further research is necessary to clarify whether this assumption proves to be true.

However, several studies demonstrated the importance of bacterial and even archaeal sRNAs in response to oxidative stress, namely by regulating the expression of important enzymes that protect the cell from ROS attacks. For instance, several sRNAs are known to regulate catalase expression in different bacteria. In *Deinococcus radiodurans*, the sRNA OsiA binds directly to the catalase *kat*A mRNA and stabilizes the transcript (Chen et al., 2019). In this bacterium it was also shown that in conditions of oxidative stress, another sRNA called OsiR is highly expressed and base-pairs with the mRNA of KatE2, resulting in increased transcript levels and increased translation of KatE2, a second catalase of this bacterium (Gao et al., 2020). In *Pseudomonas stutzeri* A1501, the NfiS sRNA was also found to bind to the mRNA of the catalase coding gene *kat*B (Zhang et al., 2019). Other sRNAs can regulate the expression of superoxide dismutase, as it the case of *Staphylococcus aureus* RscA which represses the translation of *sodA* mRNA and promotes the oxidative stress response via SodM (Lalaouna et al., 2019). Nevertheless, the role of sRNA is not restricted to regulation of ROS scavenging enzymes. sRNAs have been shown to act on multiple targets (Andrade et al., 2013) and may contribute for bacterial resistance to oxidative stress, namely through regulation of the expression of alternative transcription factors (Fröhlich and Gottesman 2018).

QUALITY CONTROL PATHWAYS TO ELIMINATE OXIDATED RNA

Oxidized nucleobases incorporated into either DNA or RNA are potentially mutagenic and constitute a major challenge for cell survival. In bacteria, oxidized DNA bases (which block replication or originate mutations) can be removed by base excision repair enzymes (namely MutM, endonucleases IV and VIII) (for more information see Imlay 2013). When the oxidative damage is too profuse, SOS response is triggered, allowing RecA protein to repair DNA oxidation-induced lesions. RecA itself can also be oxidized, but this can be reversed by Msr proteins to maintain a level of reduced functional RecA necessary to carry out both efficient recombination and SOS system repair (Henry et al., 2021). In a striking contrast, no similar mechanisms of repair have yet been reported for oxidized RNA. Such pathways may exist but simply have not been recognized yet. In favor of this view, a mechanism of repair for alkylated RNA by oxidative demethylases has already been reported, proving that damages in RNA bases are repairable (Aas et al., 2003). However, two RNA-binding proteins, MutT and PNPase, have been identified to participate in surveillance mechanisms that recognize and degrade oxidized RNA.

MutT/Nudix Family Members

Incorporation of oxidized ribonucleotides during RNA synthesis can lead to mutations (Li et al., 2014). *E. coli* RNA polymerase is ten-times less efficient in using 8-oxo-GTP than GTP (Taddei et al., 1997) and the human RNA polymerase II presents the same discriminatory activity (Hayakawa et al., 1999). MutT protein (a member of the Nudix family) was found to hydrolyze both 8-oxo-dGTP and 8-oxo-GTP, preventing their integration into DNA and RNA, respectively (Maki and Sekiguchi 1992; Taddei et al., 1997). In *Mycobacterium tuberculosis*, MutT1 and ADP-ribose pyrophosphatase (ADPRase) were shown to degrade 8-oxo-dGTP and 8-oxo-GTP/GDP (Patil et al., 2013). Such enzymatic activity is well conserved and human cells have four proteins responsible for degradation of oxidized nucleotides that are MutT homologues, namely MTH1, MTH2, MTH3 and NUDT5 (reviewed in Rudd et al., 2016).

Polynucleotide Phosphorylase

Bacterial PNPase is widely conserved and a major 3′-5′ exoribonuclease involved in RNA degradation (dos Santos et al., 2018). In *E. coli*, PNPase was found to have high affinity for 8-oxo-G containing RNA when compared to undamaged RNA (Hayakawa et al., 2001). In agreement with this result, PNPase-deficient *E. coli* cells are more sensitive to hydrogen peroxide and other oxidants than the wild-type, and the levels of 8-oxo-G were found to strongly accumulate in the PNPase mutant strain, as result of deficient elimination of these defective RNA molecules. Complementation of the mutant corrects both the growth defect and reduces the levels of 8-oxo-G, suggesting a direct role of PNPase in controlling the oxidative damage (Wu et al., 2009). The importance of PNPase to fight oxidative stress has been extended to other bacterial species, such as *D. radiodurans* (Chen et al., 2007) and *Yersinia pestis* (Henry et al., 2012), where PNPase was required to protect cells under oxidative stress. Moreover, human PNPase (hPNPase) is located mainly in the mitochondria where oxidative damage is elevated, which presupposes a role for hPNPase in controlling the level of oxidized RNA (Chen et al., 2006). PNPase-knocked down HeLa cells under H_2O_2 treatment presented reduced growth and higher 8-oxo-G content than the control (Wu and Li 2008). Overall, PNPase is a major enzyme acting in the quality control of RNA and can discriminate among oxidized and non-oxidized forms of RNA, showing higher affinity to binding and degradation of oxidation modified RNAs in different organisms.

Another example of a protein that binds oxidized RNA is the mammalian Y box-binding protein 1 (YB-1), which specifically binds with high affinity to 8-oxo-G-containing RNA

oligonucleotides (Hayakawa et al., 2002). The YB-1 domain responsible for RNA-binding is the cold shock domain and it has more than 40% homology to the major cold shock protein in *E. coli*, CspA (Matsumoto and Wolffe 1998). It is a multifunctional protein that can bind to both RNA and DNA and is important for transcriptional and translational regulation. When full length YB-1 was introduced and overexpressed in *E. coli*, bacteria became more resistant to paraquat oxidative stress (Hayakawa et al., 2002). However, when the RNA-binding domain was removed, cells were more susceptible to the oxidant. YB-1 could work as an RNA chaperone and be involved in the sequestration of oxidized RNA, preventing the translation of 8-oxo-G-containing mRNAs (Li et al., 2006). On the other hand, YB-1 could bind to oxidized ribonucleotides and recruit other factors for subsequent degradation. It is possible that both PNPase and YB-1 could have a cooperative function that protects cells against oxidized RNA, however the detailed mechanisms are still elusive (Ishii and Sekiguchi 2019).

SCAVENGER ENZYMES AGAINST FREE RADICALS IN BACTERIA

Bacteria (and other organisms) are only able to tolerate a certain amount of non-harmful levels of ROS. To mitigate this threat, bacteria have evolved several mechanisms to fight oxidative stress and avoid the imbalance of ROS levels that lead to cell toxicity. From these, the presence of scavenging enzymes that consume ROS (superoxide dismutases, catalases and peroxidases) are critical in self-defense mechanisms against oxidative stress in bacteria.

Superoxide Dismutase for Degradation of Superoxide

Superoxide dismutase is a family of enzymes widespread among bacteria, which act as a metalloenzyme that converts superoxide anion to hydrogen peroxide and molecular oxygen (Steinman 1988). *E. coli* has two cytoplasmic enzymes (Fe-SOD and Mn-SOD) and a periplasmic enzyme (Cu-Zn-SOD) (Imlay 2013). High titers of cytoplasmic SODs maintain steady-state superoxide levels at a sub-nanomolar concentration (Imlay and Fridovich 1991) in order to avoid toxicity by O_2^-. Being structurally and kinetically similar, the role of Mn-SOD and Fe-SOD is to ensure that SOD activity is present under a wide range of metal bioavailability. The two isozymes are coordinately regulated in response to iron levels: when iron is abundant, Fur (Ferric uptake regulator) inhibits Mn-SOD transcription while Fe-SOD is synthesized and activated by iron; when iron levels are reduced, Fur is deactivated and there is transcription of both Mn-SOD and the sRNA RyhB, which will repress Fe-SOD mRNA (Tardat and Touati 1991; Massé and Gottesman 2002). In anaerobic conditions, only Fe-SOD synthesis persists to prepare *E. coli* for subsequent aeration and consequent superoxide formation. Fe-SOD enzyme is favored by evolution in anaerobic habitats due to the biological availability of iron. Regarding periplasmic SODs, they protect bacteria from superoxide that leaks from respiratory chain components to the outer part of the cytoplasmic membrane. These SODs are under the control of the RNA polymerase sigma factor RpoS (also called σ^{38}, Sigma S or KatF), and are strongly induced when aerobic cells enter stationary phase (Imlay 2009).

Catalases and Peroxidases for Degradation of Hydrogen Peroxide

Catalases are common in most bacteria, with exceptions found in Gram-positive bacteria like streptococci, enterococci and leuconostocs (Mishra and Imlay 2012). *E. coli* has three main enzymes that can degrade hydrogen peroxide: alkyl hydroperoxide reductase (Ahp), catalase G (KatG) and catalase E (KatE) (Seaver and Imlay, 2001). Ahp is a two-component (AhpC-AhpF) thiol-based peroxidase that transfers electrons from NADH to H_2O_2, reducing it to water. KatG belongs to the catalase-peroxidase family and is only weakly expressed in exponential cells. However, the transcriptional regulator OxyR strongly induces both *ahp*CF and *kat*G when cells are stressed by exogenous H_2O_2 (see below). KatE is strongly expressed in stationary phase cells only, since it is induced by the RpoS system (Schellhorn and Hassan, 1988). These functionally redundant enzymes are used at different conditions: peroxidases are the primary scavenger at low H_2O_2 concentrations, whereas catalases are used at higher levels of H_2O_2 or when cells are starved (Imlay, 2013). Additionally, in the *E. coli* periplasm, cytochrome c peroxidase (Ccp) receives electrons from the respiratory chain and transfers them directly to H_2O_2 (Partridge et al., 2007). Ccp instead of eliminating H_2O_2 allows the cell to use it as a terminal oxidant to support respiration (Khademian and Imlay, 2017). This function is beneficial only when molecular oxygen is unavailable.

PERCEIVING REDOX SIGNALS FROM REACTIVE OXYGEN SPECIES

To protect cells under oxidative stress, bacteria induce the expression of several genes, namely the SoxRS, OxyR and PerR systems. The activation of these multigene transcriptional pathways is critical to the bacterial response to oxidative stress and will be next presented.

SoxRS System

Bacteria have evolved transcriptional regulators capable of perceiving redox signals from ROS, and consequently regulate enzymatic antioxidant systems (Lushchak and Storey 2021). The SoxRS regulon is a response mechanism to O_2^- and to redox-cycling compounds, but not to H_2O_2 (Gu and Imlay 2011). This system is controlled by two proteins, which activate consecutively two stages of transcription (González-Flecha and Demple 2000). SoxR is a homodimer, each with an iron-sulfur cluster [2Fe-2S] center (Hidalgo et al., 1997) and whose activity depends on its oxidation state. Only the oxidized form ($[2Fe\text{-}2S]^{2+}$) is active and able to induce

transcription of *sox*S (Gaudu and Weiss 1996). SoxS is a second transcriptional activator that induces the transcription of more than 100 genes with functions in antioxidant defense, damage repair and maintenance of central metabolisms that allow the cell to survive under oxidative stress (Nunoshiba et al., 1992; Wu and Weiss 1992). Some of the genes regulated by SoxS are *sod*A, *nfo*, *zwf*, *fum*C, *acn*A, *fur* and *ntr*A (González-Flecha and Demple 2000; Touati 2000; Pomposiello et al., 2003; Blanchard et al., 2007). After oxidative stress, SoxR is again reduced and inactivated through reducing systems encoded by *rse*C and *rsx*ABCDGE operon. Therefore, these systems maintain SoxR in its reduced state, preventing activation of the SoxRS regulon (Koo et al., 2003). Also, when oxidative stress decreases, SoxS is rapidly degraded by proteolysis (Griffith et al., 2004). The transcription factor SoxR is widely distributed in Actinobacteria and Proteobacteria. However, while in enteric bacteria SoxR activates the expression of SoxS which in turn controls the expression of a diverse class of genes involved in the stress response, in nonenteric bacteria (like *Pseudomonas aeruginosa* or *Streptomyces coelicolor*) SoxS is absent, and SoxR directly activates a small sets of genes in response to redox-cycling agents (Chiang and Schellhorn 2012; Kobayashi et al., 2014; Mettert and Kiley 2015).

OxyR System

The OxyR system is primarily responsible for recognizing and maintaining hydrogen peroxide levels in the cell (Åslund et al., 1999). OxyR is a LysR-type transcriptional regulator, widespread in most Gram-negative bacteria but it has also been reported in some Gram-positive bacteria, such as *S. aureus* and *S. coelicolor* (Morikawa et al., 2006; Oh et al., 2007). The transcriptional activity of OxyR activity depends on its redox state, as it can either act as activator if in its oxidized form or as repressor when in reducing conditions (Christman et al., 1989; Heo et al., 2010). OxyR acts as a tetramer and specifically binds to the 5' promoter regions of target genes at a conserved sequence motif (Toledano et al., 1994). In *E. coli*, the OxyR regulon spans more than 20 different genes, involved in several molecular mechanisms of adaptive response to redox stress, such as H_2O_2 detoxification (*kat*E and *ahp*CF), heme biosynthesis (*hem*H), reductant supply (*grx*A, *gor*A, *trx*C), repression of iron import (*fur*, encoding Fur regulator of ferric ion uptake) and others (Zheng et al., 1999; Zheng et al., 2001a; Zheng et al., 2001b). Additionally, OxyR upregulates the expression of OxyS, a small regulatory RNA that directs peroxide stress response (González-Flecha and Demple 1999; Fröhlich and Gottesman 2018). Although OxyS does not directly affect the combat against excessive H_2O_2, this sRNA seems to play a role in protection from H_2O_2-induced mutagenesis (Altuvia et al., 1997). Additionally, depletion of OxyS has shown to result in considerably higher levels of H_2O_2 in E. *coli* (González-Flecha and Demple 1999).

Regulation by OxyR in *E. coli* begins by sensing the levels of H_2O_2 at a specific cysteine residue in the protein (C199). Under regular conditions, OxyR is present in its reduced form (Zheng et al., 1998). Increased levels of H_2O_2 result in the rapid oxidation of OxyR: the peroxide molecule reacts with the thiol group of C199, leading to interaction with the C208-SH to form an intramolecular disulphide bond, inducing conformational changes which alter the DNA binding properties of OxyR, allowing effective interaction with RNA polymerase (Lee et al., 2004). This molecular mechanism is then reversed through feedback regulation, since oxidized OxyR induces the expression of the *grx*A and *gor* genes, encoding glutaredoxin 1 and glutathione reductase, which act to reduce OxyR (Zheng et al., 1998). Although the OxyR system has been mostly studied in *E. coli*, different bacteria have evolved in the direction to adapt their particular OxyR regulon to better suit their environmental niches: they may display differences in the molecular mechanism of H_2O_2 regulation, the number of OxyR homologs or the type of genes present in their regulon (further reviewed in Sen and Imlay 2021).

PerR System

Despite OxyR importance, the most prevalent system for preventing peroxide oxidative stress in Gram-positive bacteria is the PerR system. PerR belongs to the ferric uptake regulator (Fur) superfamily of metalloregulatory transcription factors, firstly identified in *Bacillus subtilis* where it is found to be the key regulator of the H_2O_2 response (Bsat et al., 1998). PerR has also been identified in some Gram-negative bacteria (van Vliet et al., 1999; Li et al., 2004), found primarily in association with OxyR (Wu et al., 2006). PerR is a transcriptional factor that acts as a dimer and in response to metal ions inside the cell through metal-catalyzed oxidation (Mongkolsuk and Helmann 2002). Each subunit contains a structural site that binds irreversibly to zinc (Zn^{2+}) and a regulatory metal binding site (Jacquamet et al., 2009). In *B. subtilis*, PerR binds either to manganese (Mn^{2+}) or to ferrous iron (Fe^{2+}), which is preferred in most conditions (Sen and Imlay 2021). The metal-bound conformation of PerR binds to DNA and this binding occurs at a specific Per box (Fuangthong and Helmann 2003) located in the promoter region or downstream from it, acting then as a repressor (Dubbs and Mongkolsuk 2012).

Regulation by PerR is based on the oxidation of ferrous iron (Fe^{2+}) into ferric iron (Fe^{3+}) at the regulatory site. Under non-stress conditions, binding of the regulatory metal stabilizes the conformation of the PerR dimer to better interact with DNA, which leads to the repression of the PerR regulon (Jacquamet et al., 2009). When intracellular H_2O_2 level increase, excess peroxide causes the oxidation of the iron ions in the regulatory sites of PerR through the Fenton reaction (Lee and Helmann 2006). PerR then undergoes conformational changes that render it unable to interact with DNA (Imlay 2015), leading to the derepression of PerR regulated genes. Most of these genes are involved in peroxide stress metabolism and protection (*kat*, *ahp*CF operon, *mrg*A), but some participate in metal homeostasis (*hem*AXCDBL operon) and surfactant production (*srf*A) (Chen et al., 1995; Bsat et al., 1996, 1998; Hayashi et al., 2005). Similarly to OxyR, different bacteria may have evolved to use PerR in different ways to better adapt to their environmental niche but PerR homologs tend to normally regulate similar groups of genes.

CONCLUSION

Aerobic organisms have developed over time the ability to maintain the levels of intracellular ROS within specific limits, to prevent ROS insults that can damage distinct classes of biomolecules. However, when self-defense mechanisms fail to protect the cell, ROS levels increase and as result of this imbalance, the cell faces oxidative stress. Due to its chemistry, RNA is highly susceptible to oxidation, which leads to the modification of nucleobases (namely, the formation of 8-oxo-G) causing RNA injury with potential mutagenic consequences. Since RNA is at the center of cellular regulatory pathways, the chemical damage of RNA triggered by ROS has a wide impact in gene expression. The importance of this impact is even stressed as RNA is highly expressed, comprising most nucleic acids within a cell, with the existence of many different classes of RNAs (mRNA, rRNA, tRNA and sRNA). Virtually, all these different RNA species are vulnerable to oxidation-induced damage, affecting translation at various levels. Not only the rate of protein synthesis is slowed but also the formation of aberrant proteins is potentiated, what may affect cell fitness and even cause cell death. The consequences of RNA damage by oxidation are widespread through all domains of life. Interestingly, in humans, the oxidation of RNA is implicated in several diseases, like cancer (Guo et al., 2020), diabetes (Cejvanovic et al., 2018), and neurodegenerative disorders (Fimognari 2015). Bacteria developed quality control mechanisms to specifically degrade oxidized RNA, in order to guarantee RNA fidelity. These surveillance mechanisms also rely on RNA-binding proteins that show higher affinity to bind oxidized RNA rather than non-damaged RNA. However, the number of proteins which were identified to participate in these mechanisms is rather small, with the best-known examples relying on PNPase and MutT. Nevertheless, it seems likely that other RNA-binding proteins participate in the surveillance of oxidized RNAs. Overall, the study of the bacterial response to oxidative stress and oxidation of RNA is an area that will certainly attract more attention in the years to come.

AUTHOR CONTRIBUTIONS

JA outlined the manuscript. AS wrote the first draft with the help of AQ. AS, AQ, JS, AS and JA wrote the final manuscript. AS designed the figures. JA supervised the work. All authors edited and approved the final version of this article.

FUNDING

Research in our lab was supported by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER through COMPETE2020–Programa Operacional Competitividade e Internacionalização (POCI) and by FCT–Fundação para a Ciência e a Tecnologia (Portugal) through Grant PTDC/BIA-MIC/32525/2017 to JMA and Doctoral Fellowships to AFS (PD/BD/146136/2019) and APQ (PD/BD/135487/2018).

ACKNOWLEDGMENTS

We would like to thank Guillem Hernandez Gras for helping in the preparation of **Figure 1**.

REFERENCES

Aas, P. A., Otterlei, M., Falnes, P. Ø., Vågbø, C. B., Skorpen, F., Akbari, M., et al. (2003). Human and Bacterial Oxidative Demethylases Repair Alkylation Damage in Both RNA and DNA. *Nature* 421, 859–863. doi:10.1038/nature01363

Ahmad, W., Ijaz, B., Shabbiri, K., Ahmed, F., and Rehman, S. (2017). Oxidative Toxicity in Diabetes and Alzheimer's Disease: Mechanisms behind ROS/RNS Generation. *J. Biomed. Sci.* 24, 1–10. doi:10.1186/s12929-017-0379-z

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., and Storz, G. (1997). A Small, Stable RNA Induced by Oxidative Stress: Role as a Pleiotropic Regulator and Antimutator. *Cell* 90, 43–53. doi:10.1016/S0092-8674(00)80312-8

Andrade, J. M., Pobre, V., and Arraiano, C. M. (2013). Small RNA Modules Confer Different Stabilities and Interact Differently with Multiple Targets. *PLoS One* 8, e52866. doi:10.1371/journal.pone.0052866

Andrade, J. M., Santos, R. F., Chelysheva, I., Ignatova, Z., and Arraiano, C. M. (2018). The RNA -binding Protein Hfq Is Important for Ribosome Biogenesis and Affects Translation Fidelity. *EMBO J.* 37. doi:10.15252/embj.201797631

Aslund, F., Zheng, M., Beckwith, J., and Storz, G. (1999). Regulation of the OxyR Transcription Factor by Hydrogen Peroxide and the Cellular Thiol-Ddisulfide Status. *Proc. Natl. Acad. Sci.* 96, 6161–6165. doi:10.1073/PNAS.96.11.6161

Bao, X., Yang, J., De Soet, J. J., Liu, H., Gao, X., Van Loveren, C., et al. (2017). Factors Influencing the Competition between *Streptococcus Oligofermentans* and *Streptococcus Mutans* in Dual-Species Biofilms. *Caries Res.* 51, 507–514. doi:10.1159/000479044

Berghoff, B., and Klug, G. (2012). Small RNAs with a Role in the Oxidative Stress Response of Bacteria," in *Regulatory RNAs In Prokaryotes*. Vienna: Springer Vienna, 1–14. doi:10.1007/978-3-7091-0218-3_1

Blair, I. A. (2008). DNA Adducts with Lipid Peroxidation Products. *J. Biol. Chem.* 283, 15545–15549. doi:10.1074/jbc.R700051200

Blanchard, J. L., Wholey, W.-Y., Conlon, E. M., and Pomposiello, P. J. (2007). Rapid Changes in Gene Expression Dynamics in Response to Superoxide Reveal SoxRS-dependent and Independent Transcriptional Networks. *PLoS One* 2, e1186. doi:10.1371/journal.pone.0001186

Bsat, N., Chen, L., and Helmann, J. D. (1996). Mutation of the *Bacillus Subtilis* Alkyl Hydroperoxide Reductase (ahpCF) Operon Reveals Compensatory Interactions Among Hydrogen Peroxide Stress Genes. *J. Bacteriol.* 178, 6579–6586. doi:10.1128/jb.178.22.6579-6586.1996

Bsat, N., Herbig, A., Casillas-Martinez, L., Setlow, P., and Helmann, J. D. (1998). Bacillus Subtiliscontains Multiple Fur Homologues: Identification of the Iron Uptake (Fur) and Peroxide Regulon (PerR) Repressors. *Mol. Microbiol.* 29, 189–198. doi:10.1046/j.1365-2958.1998.00921.x

Cech, T. R. (2000). The Ribosome Is a Ribozyme. *Science* 289, 878–879. doi:10.1126/science.289.5481.878

Cejvanovic, V., Kjær, L. K., Mørup Bergholdt, H. K., Henriksen, T., Weimann, A., Ellervik, C., et al. (2018). RNA Oxidation and Iron Levels in Patients with Diabetes. *Free Radic. Biol. Med.* 129, 532–536. doi:10.1016/j.freeradbiomed.2018.10.420

Chen, H.-W., Rainey, R. N., Balatoni, C. E., Dawson, D. W., Troke, J. J., Wasiak, S., et al. (2006). Mammalian Polynucleotide Phosphorylase Is an Intermembrane Space RNase that Maintains Mitochondrial Homeostasis. *Mol. Cel. Biol.* 26, 8475–8487. doi:10.1128/mcb.01002-06

Chen, L., Keramati, L., and Helmann, J. D. (1995). Coordinate Regulation of *Bacillus Subtilis* Peroxide Stress Genes by Hydrogen Peroxide and Metal Ions. *Proc. Natl. Acad. Sci.* 92, 8190–8194. doi:10.1073/pnas.92.18.8190

Chen, X., Wurtmann, E. J., Van Batavia, J., Zybailov, B., Washburn, M. P., and Wolin, S. L. (2007). An Ortholog of the Ro Autoantigen Functions in 23S rRNA

Maturation in *D. Radiodurans*. *Genes Dev.* 21, 1328–1339. doi:10.1101/gad.1548207

Chen, Y., Xue, D., Sun, W., Han, J., Li, J., Gao, R., et al. (2019). sRNA OsiA Stabilizes Catalase mRNA during Oxidative Stress Response of *Deincoccus Radiodurans* R1. *Microorganisms* 7, 422. doi:10.3390/microorganisms7100422

Chiang, S. M., and Schellhorn, H. E. (2012). Regulators of Oxidative Stress Response Genes in *Escherichia coli* and Their Functional Conservation in Bacteria. *Arch. Biochem. Biophys.* 525, 161–169. doi:10.1016/j.abb.2012.02.007

Christman, M. F., Storz, G., and Ames, B. N. (1989). OxyR, a Positive Regulator of Hydrogen Peroxide-Inducible Genes in *Escherichia coli* and *Salmonella typhimurium*, Is Homologous to a Family of Bacterial Regulatory Proteins. *Proc. Natl. Acad. Sci.* 86, 3484–3488. doi:10.1073/PNAS.86.10.3484

Cobley, J. N., Fiorello, M. L., and Bailey, D. M. (2018). 13 Reasons Why the Brain Is Susceptible to Oxidative Stress. *Redox Biol.* 15, 490–503. doi:10.1016/j.redox.2018.01.008

Criddle, D. N., Gillies, S., Baumgartner-Wilson, H. K., Jaffar, M., Chinje, E. C., Passmore, S., et al. (2006). Menadione-induced Reactive Oxygen Species Generation via Redox Cycling Promotes Apoptosis of Murine Pancreatic Acinar Cells. *J. Biol. Chem.* 281, 40485–40492. doi:10.1074/jbc.M607704200

Dalle-Donne, I., Aldini, G., Carini, M., Colombo, R., Rossi, R., and Milzani, A. (2006). Protein Carbonylation, Cellular Dysfunction, and Disease Progression. *J. Cel. Mol. Med.* 10, 389–406. doi:10.1111/j.1582-4934.2006.tb00407.x

Dobretsov, G. E., Borschevskaya, T. A., Petrov, V. A., and Vladimirov, Y. A. (1977). The Increase of Phospholipid Bilayer Rigidity after Lipid Peroxidation. *FEBS Lett.* 84, 125–128. doi:10.1016/0014-5793(77)81071-5

Dos Santos, R. F., Andrade, J. M., Pissarra, J., Deutscher, M. P., and Arraiano, C. M. (2020). Hfq and RNase R Mediate rRNA Processing and Degradation in a Novel RNA Quality Control Process. *MBio* 11. doi:10.1128/mBio.02398-20

dos Santos, R. F., Quendera, A. P., Boavida, S., Seixas, A. F., Arraiano, C. M., and Andrade, J. M. (2018). Major 3′-5′ Exoribonucleases in the Metabolism of Coding and Non-coding RNA. *Prog. Mol. Biol. Transl. Sci.* 159, 101–155. doi:10.1016/bs.pmbts.2018.07.005

Dryden, M. S., Cooke, J., Salib, R. J., Holding, R. E., Biggs, T., Salamat, A. A., et al. (2017). Reactive Oxygen: A Novel Antimicrobial Mechanism for Targeting Biofilm-Associated Infection. *J. Glob. Antimicrob. Resist.* 8, 186–191. doi:10.1016/j.jgar.2016.12.006

Dubbs, J. M., and Mongkolsuk, S. (2012). Peroxide-sensing Transcriptional Regulators in Bacteria. *J. Bacteriol.* 194, 5495–5503. doi:10.1128/JB.00304-12

Espey, M. G. (2013). Role of Oxygen Gradients in Shaping Redox Relationships between the Human Intestine and its Microbiota. *Free Radic. Biol. Med.* 55, 130–140. doi:10.1016/j.freeradbiomed.2012.10.554

Estevez, M., Valesyan, S., Jora, M., Limbach, P. A., and Addepalli, B. (2021). Oxidative Damage to RNA Is Altered by the Presence of Interacting Proteins or Modified Nucleosides. *Front. Mol. Biosci.* 8, 1–11. doi:10.3389/fmolb.2021.697149

Farr, S. B., and Kogoma, T. (1991). Oxidative Stress Responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* 55, 561–585. doi:10.1128/mmbr.55.4.561-585.199110.1128/mr.55.4.561-585.1991

Farr, S. B., Natvig, D. O., and Kogoma, T. (1985). Toxicity and Mutagenicity of Plumbagin and the Induction of a Possible New DNA Repair Pathway in *Escherichia coli*. *J. Bacteriol.* 164, 1309–1316. doi:10.1128/jb.164.3.1309-1316.1985

Fasnacht, M., and Polacek, N. (2021). Oxidative Stress in Bacteria and the Central Dogma of Molecular Biology. *Front. Mol. Biosci.* 8, 1–13. doi:10.3389/fmolb.2021.671037

Fenton, H. J. H. (1894). LXXIII.-Oxidation of Tartaric Acid in Presence of Iron. *J. Chem. Soc. Trans.* 65, 899–910. doi:10.1039/CT8946500899

Fimognari, C. (20152015). Role of Oxidative RNA Damage in Chronic-Degenerative Diseases. *Oxidative Med. Cell Longevity* 2015, 1–8. doi:10.1155/2015/358713

Fridovich, I. (1989). Superoxide Dismutases. *J. Biol. Chem.* 264, 7761–7764. doi:10.1016/S0021-9258(18)83102-7

Fröhlich, K. S., and Gottesman, S. (2018). Small Regulatory RNAs in the Enterobacterial Response to Envelope Damage and Oxidative Stress. *Microbiol. Spectr.* 6, 1–16. doi:10.1128/microbiolspec.rwr-0022-2018

Fuangthong, M., and Helmann, J. D. (2003). Recognition of DNA by Three Ferric Uptake Regulator (Fur) Homologs in *Bacillus Subtilis*. *J. Bacteriol.* 185, 6348–6357. doi:10.1128/JB.185.21.6348-6357.2003

Gao, L., Chen, X., Tian, Y., Yan, Y., Zhan, Y., Zhou, Z., et al. (2020). The Novel ncRNA OsiR Positively Regulates Expression of *katE2* and Is Required for Oxidative Stress Tolerance in *Deinococcus Radiodurans*. *Ijms* 21, 3200. doi:10.3390/ijms21093200

Gaudu, P., and Weiss, B. (1996). SoxR, a [2Fe-2S] Transcription Factor, Is Active Only in its Oxidized Form. *Proc. Natl. Acad. Sci.* 93, 10094–10098. doi:10.1073/pnas.93.19.10094

González-flecha, B., and Demple, B. (2000). Genetic Responses to Free Radicals: Homeostasis and Gene Control. *Ann. N. Y. Acad. Sci.* 899, 69–87. doi:10.1111/j.1749-6632.2000.tb06177.x

Gonza´lez-Flecha, B., and Demple, B. (1999). Role for the *oxyS* Gene in Regulation of Intracellular Hydrogen Peroxide in *Escherichia coli*. *J. Bacteriol.* 181, 3833–3836. doi:10.1128/JB.181.12.3833-3836.1999

Gort, A. S., and Imlay, J. A. (1998). Balance between Endogenous Superoxide Stress and Antioxidant Defenses. *J. Bacteriol.* 180, 1402–1410. doi:10.1128/jb.180.6.1402-1410.1998

Griffith, K. L., Shah, I. M., and E. Wolf, R., Jr (2004). Proteolytic Degradation of *Escherichia coli* Transcription Activators SoxS and MarA as the Mechanism for Reversing the Induction of the Superoxide (SoxRS) and Multiple Antibiotic Resistance (Mar) Regulons. *Mol. Microbiol.* 51, 1801–1816. doi:10.1046/j.1365-2958.2003.03952.x

Gu, M., and Imlay, J. A. (2011). The SoxRS Response of *Escherichia coli* Is Directly Activated by Redox-Cycling Drugs rather Than by Superoxide. *Mol. Microbiol.* 79, 1136–1150. doi:10.1111/j.1365-2958.2010.07520.x

Guo, C., Chen, Q., Chen, J., Yu, J., Hu, Y., Zhang, S., et al. (2020). 8-Hydroxyguanosine as a Possible RNA Oxidative Modification Marker in Urine from Colorectal Cancer Patients: Evaluation by Ultra Performance Liquid Chromatography-Tandem Mass Spectrometry. *J. Chromatogr. B* 1136, 121931. doi:10.1016/j.jchromb.2019.121931

Hangauer, M. J., Vaughn, I. W., and McManus, M. T. (2013). Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *Plos Genet.* 9, e1003569. doi:10.1371/journal.pgen.1003569

Hayakawa, H., Hofer, A., Thelander, L., Kitajima, S., Cai, Y., Oshiro, S., et al. (1999). Metabolic Fate of Oxidized Guanine Ribonucleotides in Mammalian Cells. *Biochemistry* 38, 3610–3614. doi:10.1021/bi9823611

Hayakawa, H., Kuwano, M., and Sekiguchi, M. (2001). Specific Binding of 8-Oxoguanine-Containing RNA to Polynucleotide Phosphorylase Protein. *Biochemistry* 40, 9977–9982. doi:10.1021/bi010595q

Hayakawa, H., Uchiumi, T., Fukuda, T., Ashizuka, M., Kohno, K., Kuwano, M., et al. (2002). Binding Capacity of Human YB-1 Protein for RNA Containing 8-oxoguanine. *Biochemistry* 41, 12739–12744. doi:10.1021/bi0201872

Hayashi, K., Ohsawa, T., Kobayashi, K., Ogasawara, N., and Ogura, M. (2005). The H 2 O 2 Stress-Responsive Regulator PerR Positively Regulates srfA Expression in Bacillus Subtilis. *J. Bacteriol.* 187, 6659–6667. doi:10.1128/JB.187.19.6659-6667.2005

Henry, A., Shanks, J., van Hoof, A., and Rosenzweig, J. A. (2012). TheYersinia Pseudotuberculosisdegradosome Is Required for Oxidative Stress, while its PNPase Subunit Plays a Degradosome-independent Role in Cold Growth. *FEMS Microbiol. Lett.* 336, 139–147. doi:10.1111/j.1574-6968.12000.x

Henry, C., Loiseau, L., Vergnes, A., Vertommen, D., Mérida-Floriano, A., Chitteni-Pattu, S., et al. (2021). Redox Controls RecA Protein Activity via Reversible Oxidation of its Methionine Residues. *Elife* 10. doi:10.7554/eLife.63747

Heo, Y.-J., Chung, I.-Y., Cho, W.-J., Lee, B.-Y., Kim, J.-H., Choi, K.-H., et al. (2010). The Major Catalase Gene (katA) of *Pseudomonas aeruginosa* PA14 Is under Both Positive and Negative Control of the Global Transactivator OxyR in Response to Hydrogen Peroxide. *J. Bacteriol.* 192, 381–390. doi:10.1128/JB.00980-09

Hidalgo, E., Ding, H., and Demple, B. (1997). Redox Signal Transduction: Mutations Shifting [2Fe-2S] Centers of the SoxR Sensor-Regulator to the Oxidized Form. *Cell* 88, 121–129. doi:10.1016/S0092-8674(00)81864-4

Hofer, T., Badouard, C., Bajak, E., Ravanat, J.-L., Mattsson, Å., and Cotgreave, I. A. (2005). Hydrogen Peroxide Causes Greater Oxidation in Cellular RNA Than in DNA. *Biol. Chem.* 386, 333–337. doi:10.1515/BC.2005.040

Ignatov, A. V., Bondarenko, K. A., and Makarova, A. V. (2017). Non-bulky Lesions in Human DNA: The Ways of Formation, Repair, and Replication. *Acta Naturae* 9, 12–26. doi:10.32607/20758251-2017-9-3-12-26

Imlay, J. A., and Fridovich, I. (1991). Assay of Metabolic Superoxide Production in *Escherichia coli*. *J. Biol. Chem.* 266, 6957–6965. doi:10.1016/s0021-9258(20)89596-9

Imlay, J. A. (2009). Oxidative Stress. *EcoSal Plus* 3, 231–237. doi:10.1128/ecosalplus.5.4.4

Imlay, J. A. (2013). The Molecular Mechanisms and Physiological Consequences of Oxidative Stress: Lessons from a Model Bacterium. *Nat. Rev. Microbiol.* 11, 443–454. doi:10.1038/nrmicro3032

Imlay, J. A. (2015). Transcription Factors that Defend Bacteria against Reactive Oxygen Species. *Annu. Rev. Microbiol.* 69, 93–108. doi:10.1146/ANNUREV-MICRO-091014-104322

Ishii, T., and Sekiguchi, M. (2019). Two Ways of Escaping from Oxidative RNA Damage: Selective Degradation and Cell Death. *DNA Repair* 81, 102666. doi:10.1016/j.dnarep.2019.102666

Jacquamet, L., Traoré, D. A. K., Ferrer, J.-L., Proux, O., Testemale, D., Hazemann, J.-L., et al. (2009). Structural Characterization of the Active Form of PerR: Insights into the Metal-Induced Activation of PerR and Fur Proteins for DNA Binding. *Mol. Microbiol.* 73, 20–31. doi:10.1111/j.1365-2958.2009.06753.x

Kavita, K., de Mets, F., and Gottesman, S. (2018). New Aspects of RNA-Based Regulation by Hfq and its Partner sRNAs. *Curr. Opin. Microbiol.* 42, 53–61. doi:10.1016/j.mib.2017.10.014

Kehrer, J. P. (2000). The Haber-Weiss Reaction and Mechanisms of Toxicity. *Toxicology* 149, 43–50. doi:10.1016/S0300-483X(00)00231-6

Khademian, M., and Imlay, J. A. (2017). *Escherichia coli* Cytochrome C Peroxidase Is a Respiratory Oxidase that Enables the Use of Hydrogen Peroxide as a Terminal Electron Acceptor. *Proc. Natl. Acad. Sci. USA* 114, E6922–E6931. doi:10.1073/pnas.1701587114

Kobayashi, K., Fujikawa, M., and Kozawa, T. (2014). Oxidative Stress Sensing by the Iron-Sulfur Cluster in the Transcription Factor, SoxR. *J. Inorg. Biochem.* 133, 87–91. doi:10.1016/j.jinorgbio.2013.11.008

Kong, Q., and Lin, C.-l. G. (2010). Oxidative Damage to RNA: Mechanisms, Consequences, and Diseases. *Cell. Mol. Life Sci.* 67, 1817–1829. doi:10.1007/s00018-010-0277-y

Koo, M.-S., Lee, J.-H., Rah, S.-Y., Yeo, W.-S., Lee, J.-W., Lee, K.-L., et al. (2003). A Reducing System of the Superoxide Sensor SoxR in *Escherichia coli*. *EMBO J.* 22, 2614–2622. doi:10.1093/emboj/cdg252

Küpfer, P. A., and Leumann, C. J. (2014). "Oxidative Damage on RNA Nucleobases"*in* Chemical Biology Of Nucleic Acids. Berlin, Heidelberg: Springer Berlin Heidelberg, 75–94. doi:10.1007/978-3-642-54452-1_5

Lalaouna, D., Baude, J., Wu, Z., Tomasini, A., Chicher, J., Marzi, S., et al. (2019). RsaC sRNA Modulates the Oxidative Stress Response of *Staphylococcus aureus* during Manganese Starvation. *Nucleic Acids Res.* 47, 9871–9887. doi:10.1093/nar/gkz728

Lee, C., Lee, S. M., Mukhopadhyay, P., Kim, S. J., Lee, S. C., Ahn, W.-S., et al. (2004). Redox Regulation of OxyR Requires Specific Disulfide Bond Formation Involving a Rapid Kinetic Reaction Path. *Nat. Struct. Mol. Biol.* 11, 1179–1185. doi:10.1038/nsmb856

Lee, J.-W., and Helmann, J. D. (20062006 44070). The PerR Transcription Factor Senses H2O2 by Metal-Catalysed Histidine Oxidation. *Nature* 440, 363–367. doi:10.1038/nature04537

Lee, T.-H., and Kang, T.-H. (2019). DNA Oxidation and Excision Repair Pathways. *Ijms* 20, 6092. doi:10.3390/ijms20236092

Leiva, L. E., Pincheira, A., Elgamal, S., Kienast, S. D., Bravo, V., Leufken, J., et al. (2020). Modulation of *Escherichia coli* Translation by the Specific Inactivation of tRNAGly under Oxidative Stress. *Front. Genet.* 11. doi:10.3389/fgene.2020.00856

Li, H., Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2004). Differential Gene Expression in Response to Hydrogen Peroxide and the Putative PerR Regulon of *Synechocystis* Sp. Strain PCC 6803. *J. Bacteriol.* 186, 3331–3345. doi:10.1128/JB.186.11.3331-3345.2004

Li, Z., Chen, X., Liu, Z., Ye, W., Li, L., Qian, L., et al. (2020). Recent Advances: Molecular Mechanism of RNA Oxidation and its Role in Various Diseases. *Front. Mol. Biosci.* 7, 1–7. doi:10.3389/fmolb.2020.00184

Li, Z., Malla, S., Shin, B., and Li, J. M. (2014). Battle against RNA Oxidation: Molecular Mechanisms for Reducing Oxidized RNA to Protect Cells. *WIREs RNA* 5, 335–346. doi:10.1002/wrna.1214

Li, Z., Wu, J., and DeLeo, C. (2006). RNA Damage and Surveillance under Oxidative Stress. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* 58, 581–588. doi:10.1080/15216540600946456

Liochev, S. I., Benov, L., Touati, D., and Fridovich, I. (1999). Induction of the soxRS Regulon of *Escherichia coli* by Superoxide. *J. Biol. Chem.* 274, 9479–9481. doi:10.1074/jbc.274.14.9479

Liu, M., Gong, X., Alluri, R. K., Wu, J., Sablo, T., and Li, Z. (2012). Characterization of RNA Damage under Oxidative Stress in *Escherichia coli*. *Biol. Chem.* 393, 123–132. doi:10.1515/hsz-2011-0247

Lushchak, V. I., and Storey, K. B. (2021). Oxidative Stress Concept Updated: Definitions, Classifications, and Regulatory Pathways Implicated. *EXCLI J.* 20, 956–967. doi:10.17179/excli2021-3596

Maki, H., and Sekiguchi, M. (1992). MutT Protein Specifically Hydrolyses a Potent Mutagenic Substrate for DNA Synthesis. *Nature* 355, 273–275. doi:10.1038/355273a0

Malla, S., and Kwakye, A. (2021). RNA Oxidation: Role of Polynucleotide Phosphorylase in the Quality Control of Oxidized RNA. *J. Life Sci.* 3, 43–60. doi:10.36069/jols/20210603

Markkanen, E. (2017). Not breathing Is Not an Option: How to deal with Oxidative DNA Damage. *DNA Repair* 59, 82–105. doi:10.1016/j.dnarep.2017.09.007

Massé, E., and Gottesman, S. (2002). A Small RNA Regulates the Expression of Genes Involved in Iron Metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 99, 4620–4625. doi:10.1073/pnas.032066599

Matsumoto, K., and Wolffe, A. P. (1998). Gene Regulation by Y-Box Proteins: Coupling Control of Transcription and Translation. *Trends Cel Biol.* 8, 318–323. doi:10.1016/S0962-8924(98)01300-2

Mettert, E. L., and Kiley, P. J. (2015). Fe-S Proteins that Regulate Gene Expression. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1853, 1284–1293. doi:10.1016/j.bbamcr.2014.11.018

Mishra, S., and Imlay, J. (2012). Why Do Bacteria Use So many Enzymes to Scavenge Hydrogen Peroxide? *Arch. Biochem. Biophys.* 525, 145–160. doi:10.1016/j.abb.2012.04.014

Mongkolsuk, S., and Helmann, J. D. (2002). Regulation of Inducible Peroxide Stress Responses. *Mol. Microbiol.* 45, 9–15. doi:10.1046/j.1365-2958.2002.03015.x

Morikawa, K., Ohniwa, R. L., Kim, J., Maruyama, A., Ohta, T., and Takeyasu, K. (2006). Bacterial Nucleoid Dynamics: Oxidative Stress Response in *Staphylococcus aureus*. *Genes Cells* 11, 409–423. doi:10.1111/j.1365-2443.2006.00949.x

Morzel, M., Gatellier, P., Sayd, T., Renerre, M., and Laville, E. (2006). Chemical Oxidation Decreases Proteolytic Susceptibility of Skeletal Muscle Myofibrillar Proteins. *Meat Sci.* 73, 536–543. doi:10.1016/j.meatsci.2006.02.005

Niki, E. (2009). Lipid Peroxidation: Physiological Levels and Dual Biological Effects. *Free Radic. Biol. Med.* 47, 469–484. doi:10.1016/j.freeradbiomed.2009.05.032

Nishiyama, Y., Allakhverdiev, S. I., and Murata, N. (2006). A New Paradigm for the Action of Reactive Oxygen Species in the Photoinhibition of Photosystem II. *Biochim. Biophys. Acta (Bba) - Bioenerg.* 1757, 742–749. doi:10.1016/j.bbabio.2006.05.013

Nunoshiba, T., Hidalgo, E., Amábile Cuevas, C. F., and Demple, B. (1992). Two-stage Control of an Oxidative Stress Regulon: the *Escherichia coli* SoxR Protein Triggers Redox-Inducible Expression of the *soxS* Regulatory Gene. *J. Bacteriol.* 174, 6054–6060. doi:10.1128/jb.174.19.6054-6060.1992

Ogino, T., Maegawa, S., Shigeno, S., Fujikura, K., and Toyohara, H. (2018). Highly Sensitive Avoidance Plays a Key Role in Sensory Adaptation to Deep-Sea Hydrothermal Vent Environments. *PLoS One* 13, e0189902–10. doi:10.1371/journal.pone.0189902

Oh, S.-Y., Shin, J.-H., and Roe, J.-H. (2007). Dual Role of OhrR as a Repressor and an Activator in Response to Organic Hydroperoxides in *Streptomyces Coelicolor*. *J. Bacteriol.* 189, 6284–6292. doi:10.1128/JB.00632-07

Partridge, J. D., Poole, R. K., and Green, J. (2007). The *Escherichia coli yhjA* Gene, Encoding a Predicted Cytochrome C Peroxidase, Is Regulated by FNR and OxyR. *Microbiology* 153, 1499–1509. doi:10.1099/mic.0.2006/004838-0

Patil, A. G. G., Sang, P. B., Govindan, A., and Varshney, U. (2013). *Mycobacterium tuberculosis* MutT1 (Rv2985) and ADPRase (Rv1700) Proteins Constitute a Two-Stage Mechanism of 8-Oxo-dGTP and 8-Oxo-GTP Detoxification and Adenosine to Cytidine Mutation Avoidance. *J. Biol. Chem.* 288, 11252–11262. doi:10.1074/jbc.M112.442566

Pericone, C. D., Overweg, K., Hermans, P. W. M., and Weiser, J. N. (2000). Inhibitory and Bactericidal Effects of Hydrogen Peroxide Production by *Streptococcus Pneumoniae* on Other Inhabitants of the Upper Respiratory Tract. *Infect. Immun.* 68, 3990–3997. doi:10.1128/IAI.68.7.3990-3997.2000

Pomposiello, P. J., Koutsolioutsou, A., Carrasco, D., and Demple, B. (2003). SoxRS-regulated Expression and Genetic Analysis of the *yggX* Gene of *Escherichia coli*. *J. Bacteriol.* 185, 6624–6632. doi:10.1128/JB.185.22.6624-6632.2003

Quendera, A. P., Seixas, A. F., Dos Santos, R. F., Santos, I., Silva, J. P. N., Arraiano, C. M., et al. (2020). RNA-binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria. *Front. Mol. Biosci.* 7, 78. doi:10.3389/fmolb.2020.00078

Radak, Z., and Boldogh, I. (2010). 8-Oxo-7,8-dihydroguanine: Links to Gene Expression, Aging, and Defense against Oxidative Stress. *Free Radic. Biol. Med.* 49, 587–596. doi:10.1016/j.freeradbiomed.2010.05.008

Reiter, R. J., Melchiorri, D., Sewerynek, E., Poeggeler, B., Barlow-Walden, L., Chuang, J., et al. (1995). A Review of the Evidence Supporting Melatonin's Role as an Antioxidant. *J. Pineal Res.* 18, 1–11. doi:10.1111/j.1600-079X.1995.tb00133.x

Rudd, S. G., Valerie, N. C. K., and Helleday, T. (2016). Pathways Controlling dNTP Pools to Maintain Genome Stability. *DNA Repair* 44, 193–204. doi:10.1016/j.dnarep.2016.05.032

Schellhorn, H. E., and Hassan, H. M. (1988). Transcriptional Regulation of *katE* in *Escherichia coli* K-12. *J. Bacteriol.* 170, 4286–4292. doi:10.1128/jb.170.9.4286-4292.1988

Seaver, L. C., and Imlay, J. A. (2001). Hydrogen Peroxide Fluxes and Compartmentalization inside Growing *Escherichia coli*. *J. Bacteriol.* 183, 7182–7189. doi:10.1128/JB.183.24.7182-7189.2001

Seki, M., Iida, K.-i., Saito, M., Nakayama, H., and Yoshida, S.-i. (2004). Hydrogen Peroxide Production in Streptococcus Pyogenes : Involvement of Lactate Oxidase and Coupling with Aerobic Utilization of Lactate. *J. Bacteriol.* 186, 2046–2051. doi:10.1128/JB.186.7.2046-2051.2004

Sen, A., and Imlay, J. A. (2021). How Microbes Defend Themselves from Incoming Hydrogen Peroxide. *Front. Immunol.* 12, 1104. doi:10.3389/FIMMU.2021.667343

Seok, H., Lee, H., Lee, S., Ahn, S. H., Lee, H.-S., Kim, G.-W. D., et al. (2020). Position-specific Oxidation of miR-1 Encodes Cardiac Hypertrophy. *Nature* 584, 279–285. doi:10.1038/s41586-020-2586-0

Shacter, E. (2000). Quantification and Significance of Protein Oxidation in Biological Samples*. *Drug Metab. Rev.* 32, 307–326. doi:10.1081/DMR-100102336

Shatwell, K. P., and Segal, A. W. (1996). NADPH Oxidase. *Int. J. Biochem. Cel Biol.* 28, 1191–1195. doi:10.1016/S1357-2725(96)00084-2

Shcherbik, N., and Pestov, D. G. (2019). The Impact of Oxidative Stress on Ribosomes: From Injury to Regulation. *Cells* 8, 1379. doi:10.3390/cells8111379

Shen, Z., Wu, W., and Hazen, S. L. (2000). Activated Leukocytes Oxidatively Damage DA, RNA, and the Nucleotide Pool through Halide-dependent Formation of Hydroxyl Radical. *Biochemistry* 39, 5474–5482. doi:10.1021/bi992809y

Simms, C. L., Hudson, B. H., Mosior, J. W., Rangwala, A. S., and Zaher, H. S. (2014). An Active Role for the Ribosome in Determining the Fate of Oxidized mRNA. *Cel Rep.* 9, 1256–1264. doi:10.1016/j.celrep.2014.10.042

Simms, C. L., and Zaher, H. S. (2016). Quality Control of Chemically Damaged RNA. *Cel. Mol. Life Sci.* 73, 3639–3653. doi:10.1007/s00018-016-2261-7

Stadtman, E. R. (2001). Protein Oxidation in Aging and Age-Related Diseases. *Ann. N. Y. Acad. Sci.* 928, 22–38. doi:10.1111/j.1749-6632.2001.tb05632.x

Steenken, S., and Jovanovic, S. V. (1997). How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* 119, 617–618. doi:10.1021/ja962255b

Steinman, H. M. (1988). "Bacterial Superoxide Dismutases" in *Oxygen Radicals In Biology And Medicine* (Boston, MA: Springer US, 641–646. doi:10.1007/978-1-4684-5568-7_101

Su, L.-J., Zhang, J.-H., Gomez, H., Murugan, R., Hong, X., Xu, D., et al. (20192019). Reactive Oxygen Species-Induced Lipid Peroxidation in Apoptosis, Autophagy, and Ferroptosis. *Oxidative Med. Cell Longevity* 2019, 1–13. doi:10.1155/2019/5080843

Sun, C., Jora, M., Solivio, B., Limbach, P. A., and Addepalli, B. (2018). The Effects of Ultraviolet Radiation on Nucleoside Modifications in RNA. *ACS Chem. Biol.* 13, 567–572. doi:10.1021/acschembio.7b00898

Sun, C., Limbach, P. A., and Addepalli, B. (2020). Characterization of UVA-Induced Alterations to Transfer RNA Sequences. *Biomolecules* 10, 1527. doi:10.3390/biom10111527

Suntres, Z. (2002). Role of Antioxidants in Paraquat Toxicity. *Toxicology* 180, 65–77. doi:10.1016/S0300-483X(02)00382-7

Taddei, F., Hayakawa, H., Bouton, M.-F., Cirinesi, A.-M., Matic, I., Sekiguchi, M., et al. (1997). Counteraction by MutT Protein of Transcriptional Errors Caused by Oxidative Damage. *Science* 278, 128–130. doi:10.1126/science.278.5335.128

Tanaka, M., and Chock, P. B. (2021). Oxidative Modifications of RNA and its Potential Roles in Biosystem. *Front. Mol. Biosci.* 8, 1–13. doi:10.3389/fmolb.2021.685331

Tapley, D. W., Buettner, G. R., and Shick, J. M. (1999). Free Radicals and Chemiluminescence as Products of the Spontaneous Oxidation of Sulfide in Seawater, and Their Biological Implications. *Biol. Bull.* 196, 52–56. doi:10.2307/1543166

Tardat, B., and Touati, D. (1991). Two Global Regulators Repress the Anaerobic Expression of MnSOD in *Escherichia coli*: Fur (Ferric Uptake Regulation) and Arc (Aerobic Respiration Control). *Mol. Microbiol.* 5, 455–465. doi:10.1111/j.1365-2958.1991.tb02129.x

Thomas, E. N., Simms, C. L., Keedy, H. E., and Zaher, H. S. (2019). Insights into the Base-Pairing Preferences of 8-oxoguanosine on the Ribosome. *Nucleic Acids Res.* 47, 9857–9870. doi:10.1093/nar/gkz701

Thompson, D. M., Lu, C., Green, P. J., and Parker, R. (2008). tRNA Cleavage Is a Conserved Response to Oxidative Stress in Eukaryotes. *Rna* 14, 2095–2103. doi:10.1261/rna.1232808

Toledano, M. B., Kullik, I., Trinh, F., Baird, P. T., Schneider, T. D., and Storz, G. (1994). Redox-dependent Shift of OxyR-DNA Contacts along an Extended DNA-Binding Site: A Mechanism for Differential Promoter Selection. *Cell* 78, 897–909. doi:10.1016/S0092-8674(94)90702-1

Tong, H., Chen, W., Merritt, J., Qi, F., Shi, W., and Dong, X. (2007). Streptococcus Oligofermentans Inhibits Streptococcus Mutans through Conversion of Lactic Acid into Inhibitory H2O2: a Possible Counteroffensive Strategy for Interspecies Competition. *Mol. Microbiol.* 63, 872–880. doi:10.1111/j.1365-2958.2006.05546.x

Touati, D. (2000). Sensing and Protecting against Superoxide Stress inEscherichia Coli- How many Ways Are There to triggersoxRSresponse? *Redox Rep.* 5, 287–293. doi:10.1179/135100000101535825

van Vliet, A. H. M., Baillon, M.-L. A., Penn, C. W., and Ketley, J. M. (1999). *Campylobacter Jejuni* Contains Two Fur Homologs: Characterization of Iron-Responsive Regulation of Peroxide Stress Defense Genes by the PerR Repressor. *J. Bacteriol.* 181, 6371–6376. doi:10.1128/JB.181.20.6371-6376.1999

Vaughan, P. P., and Blough, N. V. (1998). Photochemical Formation of Hydroxyl Radical by Constituents of Natural Waters. *Environ. Sci. Technol.* 32, 2947–2953. doi:10.1021/es9710417

Wang, J.-X., Gao, J., Ding, S.-L., Wang, K., Jiao, J.-Q., Wang, Y., et al. (2015). Oxidative Modification of miR-184 Enables it to Target Bcl-xL and Bcl-W. *Mol. Cel* 59, 50–61. doi:10.1016/j.molcel.2015.05.003

Welsch, C. W. (1987). Enhancement of Mammary Tumorigenesis by Dietary Fat: Review of Potential Mechanisms. *Am. J. Clin. Nutr.* 45, 192–202. doi:10.1093/ajcn/45.1.192

Willi, J., Küpfer, P., Evéquoz, D., Fernandez, G., Katz, A., Leumann, C., et al. (2018). Oxidative Stress Damages rRNA inside the Ribosome and Differentially Affects the Catalytic center. *Nucleic Acids Res.* 46, 1945–1957. doi:10.1093/nar/gkx1308

Winterbourn, C. C. (2013). *The Biological Chemistry of Hydrogen Peroxide*. 1st ed. Elsevier, 3–25. doi:10.1016/B978-0-12-405881-1.00001-XThe Biological Chemistry of Hydrogen Peroxide

Wu, H.-J., Seib, K. L., Srikhanta, Y. N., Kidd, S. P., Edwards, J. L., Maguire, T. L., et al. (2006). PerR Controls Mn-dependent Resistance to Oxidative Stress in *Neisseria Gonorrhoeae*. *Mol. Microbiol.* 60, 401–416. doi:10.1111/J.1365-2958.2006.05079.X

Wu, J., Jiang, Z., Liu, M., Gong, X., Wu, S., Burns, C. M., et al. (2009). Polynucleotide Phosphorylase Protects *Escherichia coli* against Oxidative Stress. *Biochemistry* 48, 2012–2020. doi:10.1021/bi801752p

Wu, J., and Li, Z. (2008). Human Polynucleotide Phosphorylase Reduces Oxidative RNA Damage and Protects HeLa Cell against Oxidative Stress. *Biochem. Biophysical Res. Commun.* 372, 288–292. doi:10.1016/j.bbrc.2008.05.058

Wu, J., and Weiss, B. (1992). Two-stage Induction of the soxRS (Superoxide Response) Regulon of *Escherichia coli*. *J. Bacteriol.* 174, 3915–3920. doi:10.1128/jb.174.12.3915-3920.1992

Yan, L. L., and Zaher, H. S. (2019). How Do Cells Cope with RNA Damage and its Consequences? *J. Biol. Chem.* 294, 15158–15171. doi:10.1074/jbc.REV119.006513

Yin, H., Xu, L., and Porter, N. A. (2011). Free Radical Lipid Peroxidation: Mechanisms and Analysis. *Chem. Rev.* 111, 5944–5972. doi:10.1021/cr200084z

Zhang, H., Zhan, Y., Yan, Y., Liu, Y., Hu, G., Wang, S., et al. (2019). The *Pseudomonas Stutzeri*-specific Regulatory Noncoding RNA NfiS Targets *katB* mRNA Encoding a Catalase Essential for Optimal Oxidative Resistance and Nitrogenase Activity. *J. Bacteriol.* 201. doi:10.1128/JB.00334-19

Zhang, W., Xiao, S., and Ahn, D. U. (2013). Protein Oxidation: Basic Principles and Implications for Meat Quality. *Crit. Rev. Food Sci. Nutr.* 53, 1191–1201. doi:10.1080/10408398.2011.577540

Zhao, X., and Drlica, K. (2014). Reactive Oxygen Species and the Bacterial Response to Lethal Stress. *Curr. Opin. Microbiol.* 21, 1–6. doi:10.1016/j.mib.2014.06.008

Zheng, M., Doan, B., Schneider, T. D., and Storz, G. (1999). OxyR and SoxRS Regulation of *Fur*. *J. Bacteriol.* 181, 4639–4643. doi:10.1128/JB.181.15.4639-4643.1999

Zheng, M., Wang, X., Doan, B., Lewis, K. A., Schneider, T. D., and Storz, G. (2001a). Computation-directed Identification of OxyR DNA Binding Sites in *Escherichia coli*. *J. Bacteriol.* 183, 4571–4579. doi:10.1128/JB.183.15.4571-4579.2001

Zheng, M., Wang, X., Templeton, L. J., Smulski, D. R., Larossa, R. A., and Storz, G. (2001b). DNA Microarray-Mediated Transcriptional Profiling of the *Escherichia coli* Response to Hydrogen Peroxide. *J. Bacteriol.* 183, 4562–4570. doi:10.1128/JB.183.15.4562-4570.2001

Zheng, M., Åslund, F., and Storz, G. (1998). Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation. *Science* 279, 1718–1722. doi:10.1126/SCIENCE.279.5357.1718

Zhong, J., Xiao, C., Gu, W., Du, G., Sun, X., He, Q.-Y., et al. (2015). Transfer RNAs Mediate the Rapid Adaptation of *Escherichia coli* to Oxidative Stress. *Plos Genet.* 11, e1005302–24. doi:10.1371/journal.pgen.1005302

Zhu, M., and Dai, X. (2019). Maintenance of Translational Elongation Rate Underlies the Survival of *Escherichia coli* during Oxidative Stress. *Nucleic Acids Res.* 47, 7592–7604. doi:10.1093/nar/gkz467

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 21 January 2022
doi: 10.3389/fgene.2021.591543

Fluoride-Controlled Riboswitch-Based Dampening of Gene Expression for Cloning Potent Promoters

*Vesta Korniakova, Aurélie Devinck, Marie-Christine Groleau, Eric Déziel and Jonathan Perreault**

INRS - Centre Armand-Frappier Santé Biotechnologie, Boulevard des Prairies, Laval, QC, Canada

Bioreporter systems based on detectable enzyme activity, such as that of beta-galactosidase or luciferase, are key in novel bacterial promoter discovery and study. While these systems permit quantification of gene expression, their use is limited by the toxicity of the expressed reporter enzymes in a given host. Indeed, the most potent promoters may be overlooked if their activity causes a lethal overproduction of the reporter genes when screening for transcriptional activity of potential promoter sequences with the *luxCDABE* cassette. To overcome this limitation, a variation of the mini-CTX-*lux* plasmid has been designed which allows reduction of promoter activity *via* the addition of an adjacent fluoride riboswitch. The riboswitch adds a layer of regulation between the promoter and the reporter gene, allowing cloning of stronger promoters by weakening expression, while giving the potential to induce with fluoride to provide a good signal for weaker promoters, thus circumventing limitations associated with reporter toxicity. We noticed the riboswitch potential portability issues between species, suggesting caution when using riboswitches non-native to the species where it is being used. This study introduces a new molecular biology tool which will allow for the identification of previously unverifiable or uncharacterized potent promoters and also provides a cloning vector for translational fusion with luciferase in a plasmid compatible with many species such as from the genera *Burkholderia* and *Pseudomonas*.

Keywords: plasmid, promoter, regulatory region, luciferase, translational fusion, fluoride riboswitch, *Burkholderia*, *Pseudomonas*

OPEN ACCESS

Edited by:
Kirk Blomquist Jensen, South Australian Health and Medical Research Institute, Australia

Reviewed by:
Dominique Belin, Université de Genève, Switzerland
Nikolay Shirokikh, Australian National University, Australia

****Correspondence:***
Jonathan Perreault
jonathan.perreault@inrs.ca

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *04 August 2020*
Accepted: *15 December 2021*
Published: *21 January 2022*

Citation:
Korniakova V, Devinck A, Groleau M-C, Déziel E and Perreault J (2022) Fluoride-Controlled Riboswitch-Based Dampening of Gene Expression for Cloning Potent Promoters. Front. Genet. 12:591543. doi: 10.3389/fgene.2021.591543

INTRODUCTION

Reporter genes encoding for proteins which are easily detectable through sensitive and simple means (colorimetry, fluorescence, luminescence) are key elements to numerous gene expression studies and critical to decipher regulatory elements, including the discovery of new promoters and their characterization in terms of strength and dynamics. Common reporter proteins include β-galactosidase, Green Fluorescent Protein (GFP) and luciferase; detected either by spectrophotometry, fluorimetry or luminometry, respectively. As a rule of thumb, when gene regulatory elements are cloned upstream of a reporter gene, a high reporter protein signal indicates a strong promoter, while a low signal is attributed to a weak promoter. Strategies have been developed to allow for weak promoter detection and characterization *via* reporter gene assays (Guo et al., 2019),

however to our knowledge, no strategy for the detection and study of circumstantially lethal potent promoters, have been suggested. Classical gene reporter assays may have biases against the most potent promoters. The toxicity caused by overexpression of reporter proteins could inhibit the growth of potential clones causing an important gap in new promoter discovery.

Many DNA cloning experiments are not successful and are deemed to be technical mysteries. This failure may appear initially as a cloning gap in full genome screens or as an absence of transformed colonies for a given construct in a species of interest other than the shuttle species. A possible reason for these failures may be that an overexpression of the detectable protein in the designed construct has caused a lethal metabolic burden for the cell and thus an absence of viable construct-validated clones. Previously, it has been shown through genome sequencing of clone-based assemblies that many occurring cloning gaps were not technical failures but rather a consequence of the sequences coding for toxic products (Kimelman et al., 2012). As cloning and transformation experiments often involve propagating the construct across different species, constructs must be compatible with the cloning hosts being manipulated in order for an experiment to be successful. While toxicity level thresholds of different reporter genes, their substrates, or byproducts vary depending on the host organism, overexpression of any protein can potentially be toxic (Bolognesi and Lehner, 2018). In fact, toxicity has been previously reported for luciferase substrate N-decyl aldehyde in *Saccharomyces cerevisae* and *Caenorhabditis elegans* (Hollis et al., 2001); for constitutive expression of *Gaussia princeps* luciferase (Gluc) in *Escherichia coli* (Liu et al., 2014); for β-galactosidase expression in *E. coli* under osmotic stress (Malakar et al., 2014); and for GFP in *S. cerevisae* (Kintaka et al., 2016). One study concluded that for a number of proteins, the overexpression burden limit in *S. cerevisae* is achieved for normally non-harmful proteins once it constitutes up to 15% of the total cellular proteins (Eguchi et al., 2018).

The goal of the cloning strategy described in this paper was to decouple the cloning and transformation process from the evaluation of promoter activity in a host organism. To the best of our knowledge, no such strategy has previously been described. To this end we believe riboswitches, ligand-specific RNA *cis*-acting gene regulatory elements, may be important tools for dampening the strength of potent constitutive or potent uncharacterized inducible promoters. By sandwiching an appropriate riboswitch between the potent promoter and the reporter gene, expression levels may be controlled and adjusted down to viable levels during promoter screening or characterization assays (**Figures 1A,B**). Many expression vectors are designed for protein induction *via* inducible promoters to avoid toxicity of the protein to be purified during cloning and growth, however to our knowledge no vector exists for controlled promoter read-out *via* inducible 5' UnTranslated Regions (UTRs). Different approaches may be used to reduce expression levels, such as copy number or forced chromosomal integration, this was reviewed in (Camps, 2010) and (Marschall et al., 2016). Others (Xu et al., 2013) have evaluated copy number of plasmids to optimize fatty acid production in *E. coli* and found, expectedly, that high copy number plasmids incurred higher expression which was deleterious to growth in certain conditions. In principle, copy number could be controlled either by mutating the *ori* or, at least in the case of *ColE1* plasmids, mutating and changing the ratios of RNAII and/or RNAI, which control plasmid replication (Standley et al., 2019). However, in addition to limiting us to *ColE1* plasmids, this has been done such that copy number can be achieved with different mutants, but we are not aware of a system that allows control of copy number in a manner similar to induction systems.

When choosing the right promoter-dampening riboswitch for an experiment it is important to consider its compatibility with the shuttle and final host species. Criteria to consider include regulation range dynamics of the riboswitch (including fold induction and regulation mechanism), and whether or not the trigger ligand is endogenous to the species and what systems exist, if any, to control ligand concentrations inside the cell. For this study, we identified the fluoride riboswitch (F RS) as a potential candidate for mitigating the potency of two promoters to be used in a reporter gene system: the constitutive S7 ribosomal protein gene promoter (P_{S7}) from *Paraburkholderia xenovorans* strain LB400 and the P1 integron promoter originally from R388, a trimethoprim-resistance broad-host-range plasmid (Zolg and Hanggi, 1981; DeShazer and Woods, 1996; Massey et al., 2011). These promoters were meant to be cloned upstream of the *lux* cassette in a strategically redesigned version of mini-CTX-*lux* (Becher and Schweizer, 2000), a high copy plasmid in *E. coli* or a single copy chromosomal integration plasmid for *Burkholderia thailandensis* and *Pseudomonas aeruginosa*. The fluoride riboswitch has the advantage of controlling gene expression according to concentrations of fluoride, a non-cellular metabolite. This conserved RNA structure is widespread across bacteria and archaea and is known to upregulate, in the presence of fluoride ions, the expression of proteins which manage its exportation, such as the CrcB proteins and the CLC proteins, fluoride-specific channels which act as fluoride/proton antiporters (Weinberg et al., 2010; Baker et al., 2012; Stockbridge et al., 2012). The atomic resolution structure of this riboswitch, was shown to have a four base pairs pseudoknot and two small pseudoknots of a single base pair, with the ligand, fluoride ions, coordinated to 3 Mg^{2+} ions, themselves coordinated by water and the ribose-phosphate backbone (Ren et al., 2012). This widespread riboswitch regulates numerous genes and uses different expression platforms, sometimes with obvious Rho-independent transcription terminators and sometimes presumably through translation regulation (Weinberg et al., 2010; Baker et al., 2012). Typically, F RS have a $K_D \sim 50$ μM, according to *in vitro* assays performed with instances of this riboswitch from four different species, but the concentration added in media that will trigger the riboswitch is much higher (mM range) due to active export of fluoride ions (Baker et al., 2012). Additionally, the fluoride riboswitch from a thermophilic archaeon has been previously used as an alternative strategy to inducible promoters for regulating gene expression in hyperthermophiles (Speed et al., 2018).

Our redesigned plasmid, which we named pVK-f-*lux,* features optimal cloning features for allowing to easily swap promoters and 5′UTRs as needed in order to find the right combination for a particular experiment. Additionally, it is optimized to allow for

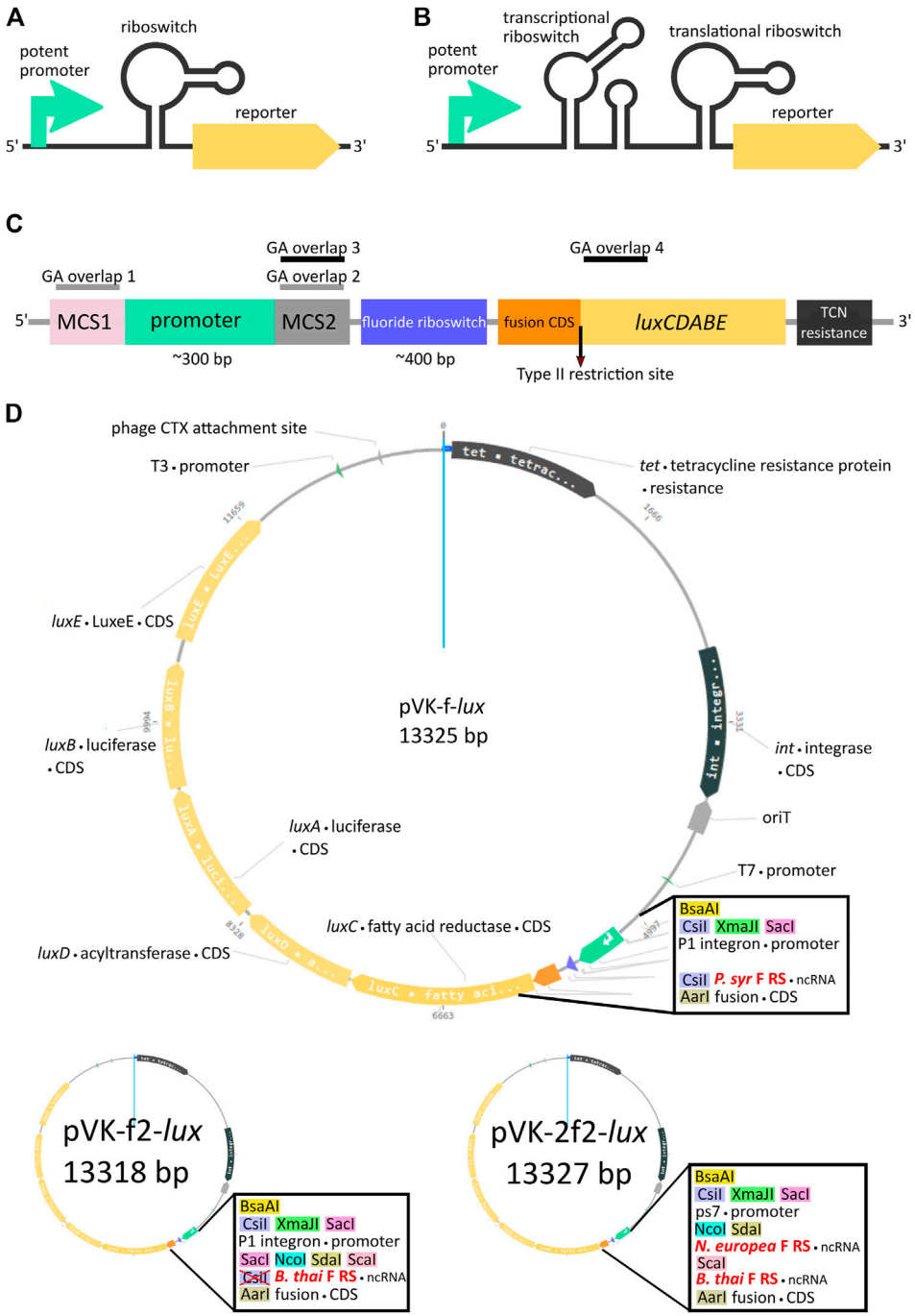


FIGURE 1 | Riboswitch-mediated reporter expression under the control of a potent promoter. Schematic diagram of a potent promoter dampening strategy with one **(A)** or two **(B)** riboswitches. **(C)** Key features of pVK-f-*lux*, a parts-swappable mini-CTX-*lux* derived backbone for convenient cloning of promoters and RNA regulatory parts. MCS1 contains CsiI-XmaJI-SacI restriction sites and MCS2 contains NcoI-SdaI-ScaI restriction sites. A type IIS restriction site is located at the beginning of *luxC* allowing digestion into the second codon for scar-free translational fusion. GA overlaps for double digestions with SacI and NcoI or ScaI and AarI have been designed for allowing interchangeability of parts as described (**Supplementary Material**: Quick User Manual for pVK-f-*lux*). These double digestions are buffer compatible and yield fragments visible on an agarose gel (approximately 300 and 400 bp in length, see **Supplementary Figure S3** for more details). **(D)** Plasmid map of pVK-f-*lux* and pVK-f2-*lux*.

scar-free translational fusion cloning, a feature not often present in reporter vectors, but essential for studying many *cis*-regulatory RNA elements. Our plasmid is designed for Gibson assembly (GA) cloning but may also be used with a restriction enzyme digestion and ligation approach.

METHODS

DNA Amplification and Reporter Plasmid Construct Assembly

Oligonucleotides from Integrated DNA Technologies (25 nmoles DNA oligonucleotides and 500 ng of gBlocks® Gene Fragments) were used. DNA parts for GA cloning were amplified using the Q5® High-Fidelity DNA Polymerase (New England Biolabs) using a touchdown-gradient PCR protocol as previously described (Korbie and Mattick, 2008) and appropriate primers and template as specified in **Supplementary Table S1**. Touchdown annealing cycles (−1.2°C/cycle for 10 cycles) were performed from 71°C down to 60.2°C and were followed by 20 cycles of constant annealing temperatures (with five tubes in a gradient from 55°C to 72°C). Reactions with the most specificity were chosen for further cloning steps. Backbone vectors, as per specific cloning attempts (**Supplementary Table S1** and **Supplementary Table S2**) were linearized using restriction enzymes (Thermo Fisher Scientific) as described. GA cloning was carried out using diluted PCR products, unpurified restriction digestion products and the NEBuilder® HiFi DNA Assembly Master Mix kit (New England Biolabs) according to the manufacturer protocol.

Bacterial Strains and Construction of Reporter Strains

All strains and clones used in this study are enumerated in **Supplementary Table S3**. All strains were grown at 37°C on a rotary agitator in liquid Luria Broth (Alpha Biosciences) or on Petri dishes of Luria agar (Alpha Biosciences). GA-cloned plasmid constructs were transformed into either *E. coli* strain DH5*α* or strain SM10λpir as follows. Either 2 μL of the GA reaction or 25 ng of the plasmid of interest was added to 100 μL of thawed chemically competent cells on ice and incubated for 20 min. A thermal shock was performed for 40 s at 42°C followed by a 3-min incubation on ice. 300 μL of Luria broth was added to the mix and cells were incubated at 37°C for 1 h with agitation at 250 rpm. 150 μL of cells were spread on a prewarmed selection plate and incubated overnight at 37°C.

The constructs were integrated into the chromosome of *B. thailandensis* E264 by bi-parental conjugation with *E. coli* SM10λpir as follows. Pellets from 1.5 ml of overnight cultures diluted to 0.5 OD_{600} for both *E. coli* SM10λpir donor strains and for *B. thailandensis* E264 *WT* strain were obtained by centrifugation at 7,000 *g*. Each pellet was resuspended in 25 μL of LB and pooled into a single drop on an antibiotic-free Luria agar dish for overnight incubation at 37°C. The resulting growth was resuspended in 1 ml of liquid Luria Broth using a sterile Q-tip and 100 μL was plated on Luria agar selection plates for *B. thailandensis* E264.

Liquid and solid selection media were supplemented with 15 μg/ml tetracycline for *E. coli* strains; and with 25 μg/ml tetracycline, 50 μg/ml gentamycine and 15 μg/ml polymyxin for *B. thailandensis* E264. When required, FH_4KO_2 was added to Luria agar selection plates or to liquid media in concentrations ranging from 0 to 31 mM. Transformed reporter strains were verified for luminescence signal using a microplate reader (Cytation 3; BioTek Instruments, Inc.). Plasmids were extracted from transformed *E. coli* strains using the Presto™ Mini Plasmid Kit (Geneaid) and sequences were confirmed by Sanger sequencing carried out at Genome Quebec (Montreal, Canada).

Testing Gibson Assembly-Based Cloning in the Plasmid

The designed overhangs of the 5′UTR DNA sequence which excludes the promoter region, were tested for compatibility for GA with a ScaI and AaRI digested backbone, by carrying out GAs with inserts containing the NcoI-SdaI-ScaI left overlap sequence of 17 nts and the AarI/luxC right overlap sequence of 21–24 nts (more details in the cloning flow chart in **Supplementary Material**).

Lux Reporter Assay

To assess time-course riboswitch regulation dynamics in bacteria, *E. coli* DH5*α*, *E. coli* SM10 and *B. thailandensis* strains transformed with constructs of interest containing the P1 promoter and a 5′UTR translational fusion with the bacterial operon *luxCDABE* were first cultured overnight in LB supplemented with the same antibiotic composition as during transformation. Next, cells were pelleted by centrifugation at 15,000 x *g* for 3 min and washed twice with M9 Minimal Media (M9-MM). Cells were then suspended in fresh M9-MM. M9-MM was prepared by combining 200 ml of sterile M9 salts (64 g/L Na_2HPO_4-$7H_2O$, 15 g/L KH_2PO_4, 2.5 g/L NaCl, 5.0 g/L NH_4Cl with 2 ml of sterile 1 M $MgSO_4$ or $MgCl_2$, 20 ml of 20% glucose and 100 μL of sterile 1 M $CaCl_2$ in a total volume of 1,000 ml of sterile deionised water. Assays were carried out in 96-well microplates from Greiner Bio-One (Microplate, 96 well, PS F-bottom [chimney well], white, med. binding Ref: 655095). Each well contained a total culture volume of 200 μL of antibiotic supplemented 0.5X M9-MM. Cultures were adjusted to an initial optical density at 600 nm (OD_{600}) of approximately 0.06. To seal the plate, an optical film was used before reading. Luminescence and OD_{600} readings were recorded at 20 min intervals with a multi-mode microplate reader (Cytation 3; BioTek Instruments, Inc.) for assay total run times between 30 and 60 h.

Luciferase Reporter Data Analysis

Average blank values for each time-point were subtracted from each corresponding well sample reading. All readings were cropped to start as soon as the OD_{600} reading hit 0.1 for an individual well. For each assay the number of time points used was adjusted to be the same for all samples unless otherwise stated. For determining the plateau OD_{600} value, the average value of the last 58 OD_{600} readings of a time-course luciferase assay was calculated.

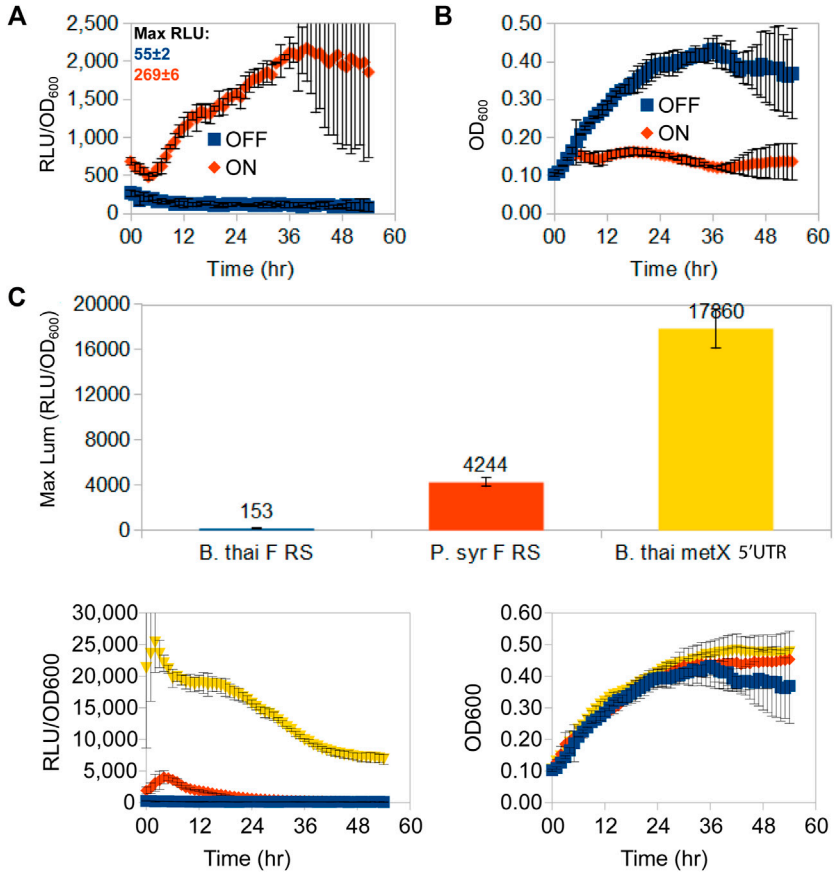


FIGURE 2 | Time-course luminescence induction curves of *B. thailandensis* E264 clones containing the *B. thailandensis* fluoride riboswitch and repression effect. A 54 h time-course *lux* expression assay comparing *lux*/OD **(A)** and growth curves **(B)** of *B. thailandensis* E264 containing the P1-*B. thai* F RS-*lux* (pVK-f2-*lux*) constructs (See **Supplementary Table S3**: *B. thailandensis E264*/P1 + *B. thai* F) for either an ON-induced state of the fluoride riboswitch in the presence of 31 mM F^-, or an OFF-repressed state fluoride riboswitch in the absence of F^-. Cultures were grown and measured on the same 96-well microplate assay run and maximum peak levels of luciferase expression are indicated above curves in Relative Luminescence units (RLU) for un-normalised-to-OD signal strength comparison. The data points represent the means and standard deviations of triplicate values. **(C)** *In vivo* repression capability of three OFF-repressed riboswitch-containing constructs in *B. thailandensis* E264: *B. thailandensis* fluoride riboswitch construct (pVK-f2-*lux*—*B. thai* F RS), the *P. syringae* fluoride riboswitch (pVK-f-*lux*—*P. syr* F RS), and the *B. thailandensis metX* 5′UTR (*B. thai metX* 5′UTR). OFF-repression was achieved with 0 mM F- for fluoride riboswitch constructs and with 0.05 mM methionine for the *metX* 5′UTR construct which is suspected to have a regulatory element (Leyn et al., 2014; and unpublished data). RLU per OD_{600} and growth curves are shown as well (bottom left and right respectively). The data points represent the means of triplicate values.

Fold Induction of Total Luciferase Activity

$$FI\ of\ \text{lumtotal} = \frac{f^{-ligand}(n)}{f^{+ligand}(n)}$$

where:

$$f^{-ligand}(n) = \frac{\sum_{i=m}^{n}\left(\frac{lum}{OD}\right)^{-ligand}(i)}{}; f^{+ligand}(n) = \frac{\sum_{i=m}^{n}\left(\frac{lum}{OD}\right)^{+ligand}(i)}{}$$

Where *ligand* is fluoride ions (F-);

$(lum/OD)^{-ligand}(n)$ is the luminescence reading for a culture containing a P1+ F RS-*lux* construct of interest in absence of supplementary fluoride for the time point *n* normalized to its OD at 600 nm.

$(lum/OD)^{+ligand}(n)$ is the luminescence reading for a culture containing a P1+ F RS-*lux* construct of interest in presence of supplementary fluoride for the time point *n* normalized to its OD at 600 nm.

For the calculation of error on fold induction, standard deviation of both triplicates were combined with the following formula. $\frac{\sigma FI}{FI} = \sqrt{\left(\frac{\sigma F0}{av.F0}\right)^2 + \left(\frac{\sigma Fx}{av.Fx}\right)^2}$

Where FI is fold induction; σFI is standard deviation of fold induction; σF0 is standard deviation of triplicate at 0 mM F^-; av. F0 is average of triplicate at 0 mM F^-; and similarly for Fx (representing the triplicates of each fluoride concentration tested).

Peak lum/ OD_{600}

The maximum peak luminescence value, in Relative Luminescence Units (RLU), was determined for a time-course

TABLE 1 | Cloning success of the promoters upstream of the *lux* operon depends on the choice of 5′UTR.

Cloning attempt		Cloning outcome
Promoter	**5′ UTR**	
P_{S7}	5′-AGGAGC-3′ RBS	failed
P_{S7}	*B. thailandensis* fluoride riboswitch	failed
P_{S7}	*P. syringae* fluoride riboswitch	failed
P_{S7}	*B.cereus* and *B.thailandensis* fluoride riboswitch	failed
P_{S7}	*N.europea* and *B.thailandensis* fluoride riboswitch	successful
P1 integron	5′-AGGAGC-3′ RBS	failed
P1 integron	5′-AGGAGU-3′ RBS	failed
P1 integron	*E. coli thiM* TPP riboswitch	failed
P1 integron	*B. thailandensis thiC* riboswitch	failed
P1 integron	*B. thailandensis* mini-*ykkC* riboswitch	failed
P1 integron	*B. thailandensis metK* 5′UTR [a,b]	failed
P1 integron	*B. thailandensis* fluoride riboswitch	successful
P1 integron	*P. syringae* fluoride riboswitch	successful
B. thailandensis metK promoter	5′-AGGAGC-3′ RBS	successful
P1 integron	*B. thailandensis metX* 5′UTR[b]	successful
P1 integron	*B. thailandensis metZ* 5′UTR [b]	successful
P1 integron	*B. cereus* fluoride riboswitch + *B. thailandensis* fluoride riboswitch	successful (but not inducible)
P1 integron	*P. aeruginosa* PA14 *yybP-ykoY* riboswitch	successful

[a]*Three different construct designs of varying lengths were attempted (not shown).*
[b]*The metK, metX and metZ 5′UTR, from B. thailandensis were suspected to have regulatory elements (Leyn et al., 2014), which we confirmed (unpublished data).*

reporter assay. This value was divided by the corresponding time-point OD_{600}.

Average Lum/OD_{600}

The average RLU value for all time points of each technical triplicate (i.e., for each clone, a single pre-culture divided in three wells for cultures with measurements over ~ 2 days, or as described in figures and text) was divided by the average OD_{600} value for all time points of each triplicate, respectively. Additional replicate experiments were performed to ensure reproducibility, but were not included in statistics. In cases where reproducibility could not be assessed, it is mentioned in the text.

For the double fluoride riboswitch construct, because the luminescence was close to background, for each assay a restricted window of time was used to calculate FI. This window was selected when the average luminescence of triplicates over 1 hour (i.e., for nine data points) was greater than standard deviation for at least 2 hours in a row (in other words, when luminescence was above background). Such luminescence levels were observed only at concentrations of 31 and 62 mM fluoride and, to use the same time window for both concentrations, we limited ourselves from 22h00–29h20 for *E. coli* DH5α and from 7h00–12h00 for *B. thailandensis.*

RESULTS AND DISCUSSION

The pVK-F-*Lux* Plasmid Allows for Rapid Mixing and Matching Promoters and 5′UTRs

To study *cis*-regulatory RNA elements, we wanted to devise a luciferase reporter with a potent promoter to provide a strong signal. After multiple cloning attempts, sequencing of the only two clones with inserts of the correct size revealed mutations which would explain the lack of luminescence in these clones (**Supplementary Figure S1**). In that context, the most likely reason for these failed cloning experiments appeared to be a selective pressure against strong expression of this reporter. To us, this highlighted challenges related to the study of strong promoters and how many strong promoters might have been overlooked in past screening attempts. We thus used a fluoride riboswitch as a way to dampen expression independently of the promoter cloned upstream to design a new reporter tool with unique features.

For our design we chose to include three different single cutter restriction enzyme sites as GA overlap flanking sequences for the promoter region, both for the 5′ end, designated as the Multiple Cloning Site 1 (MCS1), and for the 3′end, designated as the Multiple Cloning Site 2 (MCS2) (**Figures 1C,D**). Two versions of the plasmid were constructed, containing either the *P. syringae* fluoride riboswitch (*P. syr* F RS), termed pVK-f-*lux*, or the *B. thailandensis* E264 fluoride riboswitch (*B. thai F RS*) (**Supplementary Figure S2**), termed pVK-f2-*lux* (**Figure 1D**). For the 3′ end of the 5′ UTR part, AarI, a type IIS restriction site was incorporated. If the inner restriction sites are used to cut the backbone to insert a promoter sequence, then the remainder of the MCS1 and MCS2 sequences on the linearized backbone are sufficient as GA overhangs (GA overlap 1 and GA overlap 2 in **Figure 1C**) and can be added to the insert of interest. Similarly, by digesting the backbone for a 5′ UTR part insert using the innermost (in relation to the insert position) MCS2 site, and the AarI site, the backbone remainder of the MCS2 site may be used for the 5′ GA overhang (GA overlap 3 in **Figure 1C**) and the beginning 15–20 nucleotides of *luxC* may be used as the 3′ end GA

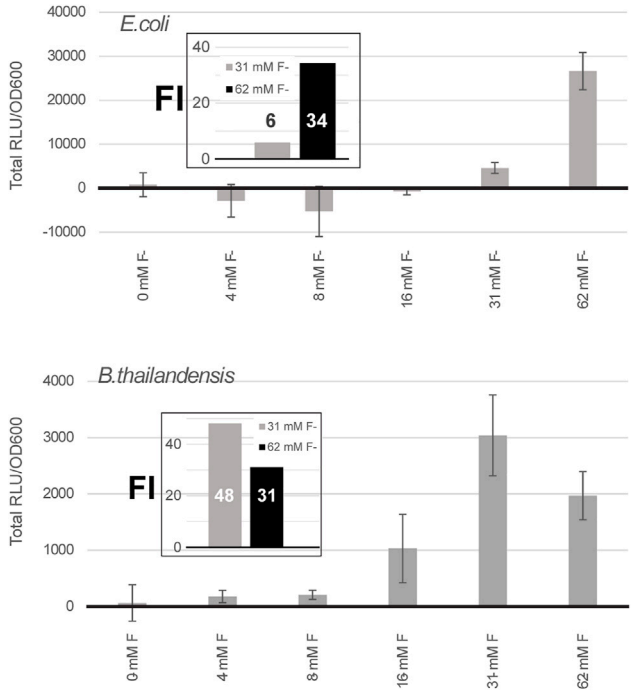


FIGURE 3 | A double fluoride riboswitch construct with P_{S7} is strongly repressed and inducible. Total luciferase activity in cultures was calculated over an incubation period of a few hours, ~ 7 h for *E. coli* DH5α and 5 h for *B. thailandensis,* these time points were selected because they were above background, as expression remained low even when induced (**Supplementary Figure S6**). Both were transformed with the *B. thailandensis* E264 fluoride riboswitch construct which has an additional fluoride riboswitch from *Nitrosomonas europea* upstream (pVK-2f2-*lux*). The plasmid is replicative in *E. coli* strains and integrative in *B. thailandensis* E264. Both pre-culture media (LB) and expression media (0.5X M9-MM) were supplemented with appropriate antibiotics depending on the species (see Methods). Fluoride induction concentrations are as indicated. Concentrations were chosen according to the maximum induction effect. The values for each sample represent the means and standard deviations of triplicates on the same microplate.

overhang (GA overlap 4 in **Figure 1C**). Short inserts (18 and 21 bp) could not be cloned, because a small insert size is already known to be detrimental for Gibson assembly (Roth et al., 2014).

The AarI RE site is positioned to cut the backbone directly after the 2nd nucleotide of the 2nd codon of *luxC*, which allows achieving a scar-free translational fusion (i.e., with no MCS sequence between a regulatory element under study and the start codon) with a choice of the desired start codon. This could be useful given that some known examples of non-AUG start codons are important for translational regulation (Hecht et al., 2017). If scar-free translational fusion is not a priority it is recommended to re-incorporate the AarI site and all depleted restriction sites into the constructs to allow for the flexibility of further cloning by using new plasmid constructs as backbones for new experiments rather than needing to start over from the original backbone, as required for many plasmid assembly methods. If at any point the innermost restriction enzyme recognition site of an MCS, in relation to the insert, is not ideal for a particular cloning strategy, when using a newly assembled backbone, there are still two other restriction sites which may be used at each MCS, thus avoiding, in most cases, the necessity to domesticate any inserts (**Supplementary Figure S3**). Additional information for using this plasmid is available (**Supplementary Material**: Quick User Manual for pVK-f-*lux*).

The Fluoride Riboswitch Dampens Reporter Gene Expression and Facilitates Potent Promoter Cloning

Having chosen the fluoride riboswitch as the candidate regulatory RNA for independent promoter dampening triggered by a non-cellular metabolite, we first needed to evaluate its usefulness to repress elevated expression levels with the option of re-activating expression. For this, we measured luminescence for a fluoride riboswitch reporter construct (pVK-f2-*lux*) in its original host organism (*B. thailandensis* E264) (**Figure 2**). Depending on the conditions and time, fluoride supplementation causes a ~ 20X induction of luminescence (RLU/OD_{600}) (**Figure 2A**). The addition of fluoride, or general ion content of media, had no apparent impact on osmotic pressure, since equivalent amounts of chloride (NaCl) made no difference (**Supplementary Figure S4**). In *B. thailandensis* E264, for OFF conformations (absence of fluoride), the repression capacity at the maximum peak expression of a 54 h time-course luciferase expression curve of the *B. thai* F RS was 28X greater than that of the *P. syr* F RS, and 117X greater than that of the *B. thailandensis* E264 *metX* UTR used as a control (*B. thai metX* 5′UTR; **Figure 2C**).

Cloning attempts of P_{S7} were unsuccessful with the presence of one fluoride riboswitch, while cloning attempts of the P1 integron promoter were only successful in presence of a riboswitch—either the fluoride riboswitch sequence from *B. thailandensis* E264, the fluoride riboswitch sequence from *P. syringae,* three different 5′UTRs involved in methionine metabolism from *B. thailandensis* E264 or the *yybP-ykoY* riboswitch from *P. aeruginosa* PA14. This suggests that a 5′UTR dampening tool is imperative to repress promoter potency and reduce its toxicity. Cloning the AGGAGC RBS by itself downstream of the P1 promoter was unsuccessful using many GA design strategies (**Supplementary Table S1**). However, cloning this RBS was successful when integrated within one of the above-mentioned 5′UTR or when in tandem with a weak promoter such as that of the *metK* promoter from *B. thailandensis* E264. Moreover, we have successfully cloned the P_{S7} promoter using a construct comprising two fluoride riboswitches, one from *Nitrosomonas europea* ATCC 19718 and a second from *B. thailandensis* E264, both Betaproteobacteria.

Not all riboswitches enabled successful cloning of the strong P1 promoter, and only a combination of two riboswitches enabled cloning of the P_{S7} promoter, suggesting that the tested riboswitches alone did not sufficiently repress *lux* expression in their OFF conformations. As a reference, one study reported that mRNA coding for the S7 protein was among the top 3% in terms of total RNA quantity, highlighting how strong this promoter is, while *metK* mRNA was in the top 10%

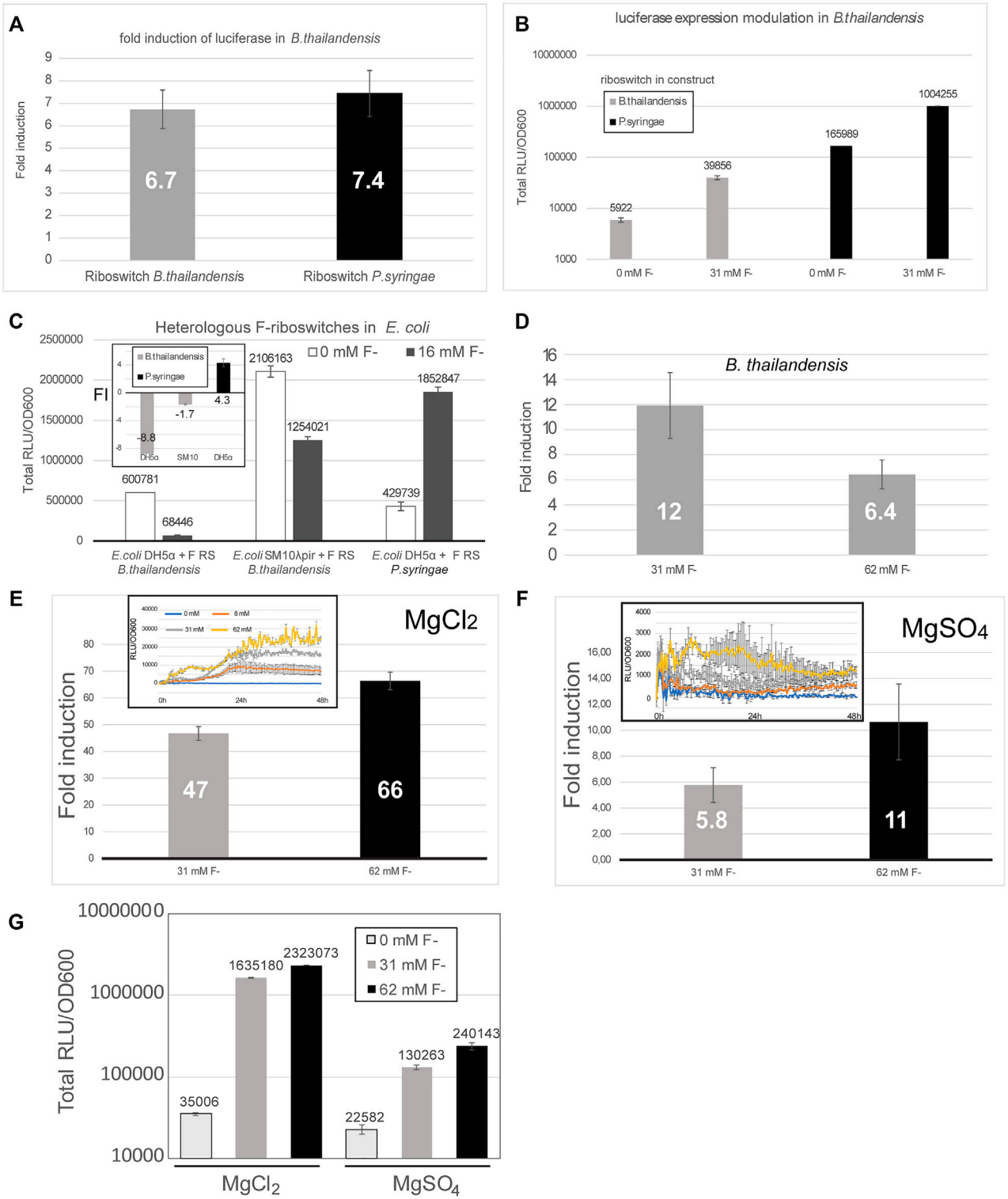


FIGURE 4 | The fluoride riboswitch modulation is affected by media and host strain. **(A)** Fold induction (FI) of total luciferase activity in cultures was calculated over an incubation period of ~ 40 h (*B. thailandensis* transformed with the *B. thailandensis* E264 fluoride riboswitch construct—pVK-f2-*lux*) and ~55 h (*B. thailandensis* E264, transformed with *P. syringae* fluoride riboswitch construct—pVK-f-*lux*). Both plasmids are replicative in *E. coli* strains and integrative in *B. thailandensis* E264. Both pre-culture media (LB) and expression media (0.5X M9-MM) were supplemented with appropriate antibiotics depending on the species (see Methods). **(B)** The same results are represented with the direct luminescence (sum) data. **(C)** Fluoride induction concentrations were 16 mM for *E. coli* DH5α and *E. coli* SM10λpir transformants and 31 mM for *B. thailandensis* E264. Concentrations were chosen according to the maximum induction effect. The FI values for each sample represent the means and standard deviations of triplicates on the same microplate. **(D)** Fold induction (FI) of *B. thailandensis* E264 transformed with the *B. thailandensis* E264 fluoride riboswitch for 31 mM and 62 mM of fluoride. **(E–G)** Different culturing conditions ($MgCl_2$ vs $MgSO_4$) were also evaluated. A much stronger induction by fluoride can be noticed with $MgCl_2$.

(Gorochowski et al., 2019). A full list of successful and unsuccessful cloning experiments enumerated in **Table 1** highlights the relation between expression levels and successful cloning of the promoter.

The fluoride riboswitch may be a useful tool for screening for a wide range of promoters and not only those which are potent. For example, a library of unknown sequences containing possible promoters may be cloned into pVK-f-*lux* with an up-regulating fluoride riboswitch in the target host such that the same library of clones can be screened for reporter gene activity in the presence or absence of fluoride. In this way, a screen in the absence of fluoride (with maximum repression), would yield those clones containing the most potent promoters. In parallel, another screen with added fluoride to re-activate riboswitch-mediated repression would allow to detect weaker promoters. Ideally, fluoride threshold tolerance of the target species during transformation and reporter assays as well as the timing of the expression pattern should be determined prior to screening.

To test the possibility of using the fluoride riboswitch as a screening tool to mediate promoter potency during transformation of strong promoters (such as P_{S7} or P1 integron promoters), the viability of *E. coli* DH5α and *E. coli* SM10λpir transformant cells was assessed in the presence of 10 and 15 mM fluoride on selection plates. *E. coli* DH5α showed reduced viability at both concentrations as illustrated by the reduced number of visible colonies (**Supplementary Figure S5**). Colonies which grew on fluoride supplemented plates were also visibly smaller for all tested constructs. Sequenced plasmid extractions of overnight inoculations of the P_{S7} promoter + *B. thai* F RS-*lux* in liquid media with corresponding fluoride concentrations revealed non-functional mutants for all cases even those for which *lux* expression should have been repressed in fluoride supplemented media. We suspect that the cloning failure of P_{S7} in this experiment was due to insufficient repression of *lux* expression rather than fluoride concentration levels in the transformation media as sequencing results were similar to previous Gibson assembly attempts in absence of fluoride and using different GA designs for the same construct (**Supplementary Material**). Overall 33% of sequenced clones of the P_{S7} promoter + *B. thai* F RS-*lux* attempts from **Supplementary Figure S5** contained a 56 nt addition and a point mutation of C285T, in reference to the P_{S7} sequence, 37% had a gap, and 30% had an unrelated sequence included as the insert (**Supplementary Table S4**). The effect of fluoride on growth in liquid expression media was also tested and *B. thailandensis* E264 clones with chromosomally integrated P1+riboswitch-*lux* constructs for the *B. thailandensis* or the *P. syringae* fluoride riboswitches. All were shown to grow equally well in 31.25 and 62.5 mM fluoride supplemented liquid media (as seen in **Figure 2**), demonstrating their tolerance to fluoride presence. The P_{S7} *N. europea* + *B. thailandensis* fluoride riboswitches construction was tested using different fluoride concentrations in *E. coli* DH5α and in *B. thailandensis* E264. In *E. coli* DH5α we saw an increase in luciferase expression at 62 mM of fluoride with a fold change of 25 compared to the condition without fluoride (**Figure 3**). This data suggests that with a double fluoride riboswitch construct we are able to clone the strong P_{S7} promoter and analyze its action in relation to luciferase expression. However, in spite of an apparently strong induction, this double-riboswitch construct is limited by a strong repression. Indeed, the ~ 30 fold induction from **Figure 3** hides a very low expression even when driven by P_{S7} (**Supplementary Figure S6**).

Luciferase expression was tested in *E. coli* DH5α, *E. coli* SM10λpir and *B. thailandensis* E264, each transformed with a construct containing the fluoride riboswitch originating either from *B. thailandensis* or from *P. syringae* (**Figures 4A–C**). While riboswitch modulation varied between conditions, for *B. thailandensis* it was coherent with the expected induction mechanism of F RS (**Figures 4A,B,D–G**), *B. thailandensis* E264 demonstrated an up-regulation, with a 5.1 fold change, up to 25 fold (**Figure 2A**) with the addition of fluoride, and even ~ 65 fold in media with $MgCl_2$ vs $MgSO_4$ (**Figure 4E**) which we noticed fortuitously. In contrast, *E. coli* DH5α and *E. coli* SM10λpir yielded a very small up-regulation (if any) and we even observed a down-regulation in some assays (**Figure 4C** and **Supplementary Figure S7A**). In other words, we did not obtain reliable results with regards to fluoride-mediated induction for the shuttle vector *E. coli*, for which the riboswitches are not native. Nonetheless, repression apparently still occurred since our ability to clone constructs in *E. coli* closely paralleled the strength of promoters, most likely because of viability issues of constructs that strongly expressed luciferase (**Table 1**). It should also be noted that expression, and FI, varied considerably depending on media used (0.5X M9 with or without sulfur supplementation, i.e., $MgCl_2$ or $MgSO_4$) (**Figures 4E–G**). Moreover, fluoride ions being toxic, a pleiotropic effect is expected, such as reduced growth, especially for *E. coli* in presence of 62 mM fluoride, but also with regards to expression. Indeed, a slight fluoride-dependent repression was observed in *B. thailandensis* transformed with the *B. thai metZ* 5′UTR plasmid, even if this UTR harbors no F RS (**Supplementary Figure S8**). Other riboswitches were tested as well (**Table 1**), but either did not dampen expression enough to allow cloning or did not provide as good of a modulation (less than two fold). Regulation sensitivity thresholds were also tested for the P1 + *B. thai* F RS (pVK-f2-*lux*, **Supplementary Table S2**) constructs transformed into *B. thailandensis* E264 and visible regulation effect was achieved at 3.9 mM for *B. thailandensis* E264/P1 + *B. thai* F (**Supplementary Figure S7**). Characterization of presented constructs across different strains is important for understanding the limitations of a fluoride riboswitch-mediated *lux* reporter system. To this end, plasmid replication levels and luciferase expression levels were determined in *E. coli* SM10λpir and *E. coli* DH5α. *E. coli* SM10λpir expresses far more luciferase than *E. coli* DH5α at similar fitness levels based on the plateau OD_{600} value, however counterintuitively *E. coli* DH5α produces 1.4X more of pVK-f2-*lux* than the former (**Supplementary Table S5**).

When running a time-course expression assays, the duration should be optimized to ensure that an expression peak is attained for the given strain and media conditions. Even if addition of 8 mM or 16 mM fluoride induced luciferase expression in most relevant assays, some discrepancies were observed between some clones, both with regards to exact expression quantitation and

growth curves. Additionally, peaks were reached at different times in different fluoride concentrations highlighting the importance of a sufficiently long assay run (**Supplementary Figure S9**). Oscillating expression (with ups and downs) may also be observed when using the fluoride riboswitch in reporter assays as fluoride concentrations inside bacteria will vary according to the activity of the fluoride export pumps (**Supplementary Figure S9**).

Characterization is not only important for understanding how different conditions affect riboswitch dynamics but also how they may affect the *lux* cassette enzymes, as the system is composed of five different enzymes (*luxCDABE*). We did test the system's sensitivity to unrelated inducers/repressors. Effect of chloride (up to 15.6 mM) was tested on *E. coli* clones carrying pVK-f2-*lux,* however no regulatory effect was observed (**Supplementary Figure S4**). Methionine addition (up to 125 mM) to *B. thailandensis* E264 clones carrying pVK-f2-*lux* also did not have a regulatory effect.

CONCLUSION

In this study we designed a plasmid which allows for straightforward swapping of promoters and 5′ UTR translationally fused sequences directly from PCR amplified inserts using Gibson assembly-based homologous cloning. We also included the fluoride riboswitch as a tool for modulating reporter gene expression under the control of strong constitutive promoters, such as the P1 integron promoter, in order to circumvent possible reporter overexpression toxicity in both shuttle and final host species, even if it still has limitations, as exemplified by the cloning of the particularly strong P_{S7} promoter which required the combined repression of two riboswitches. We also illustrate that riboswitches used as cloning tools need to be characterized across shuttle species as well as the target species to ensure optimal use. Indeed, we discovered that in its native species the *B. thailandensis* fluoride riboswitch upregulates expression when supplemented with fluoride, yet this gene induction does not translate well to *E. coli*. This is not the first time such a phenomenon has been observed, there are several accounts of riboswitches discovered in metagenomes, or in bacteria difficult to transform, that do not modulate gene expression in model organisms like *E. coli* (personal communication, Ronald Breaker). Several reasons may explain this phenomenon: the difference in riboswitch expression platform folding kinetics due to difference in RNA polymerase activity across species or the wide gap in the GC% of their respective genomes (67 *vs* 50% for *B. thailandensis* and *E. coli*, respectively). Nevertheless, the repression (even if not necessarily relieved by fluoride) permitted cloning both in *E. coli* and *B. thailandendis*. This work may also serve as an example of riboswitch use to improve current cloning tools. Other riboswitches whose ligands are independent of the host organism's metabolism and have less pleiotropic effects than fluoride, such as the theophylline synthetic riboswitch (Topp et al., 2010), may provide alternatives to apply the same approach, potentially circumventing some of the project-specific limitations that can be encountered the same way different resistance markers can be more or less appropriate for a given cloning experiment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

VK designed vector. VK and AD performed experiments and wrote manuscript. M-CG and ED provided strains and some reagents as well as critical advice and revised manuscript. JP helped to conceive vector, design experiments and revised the manuscript.

FUNDING

JP thanks support from Natural Sciences and Engineering Council of Canada (NSERC) (418240-2012-RGPIN and RGPIN-2019-06403 to JP). JP is a junior 2 FRQS research scholar. AD was supported by a scholarship from the Armand-Frappier foundation. This study was also supported by Canadian Institutes of Health Research operating grant MOP-142466 to ED.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.591543/full#supplementary-material

REFERENCES

Baker, J. L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R. B., and Breaker, R. R. (2012). Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* 335, 233–235. doi:10.1126/science.1215063

Becher, A., and Schweizer, H. P. (2000). Integration-proficient *Pseudomonas aeruginosa* Vectors for Isolation of Single-Copy Chromosomal lacZ and Lux Gene Fusions. *Biotechniques* 29, 948–952. doi:10.2144/00295bm04

Bolognesi, B., and Lehner, B. (2018). Reaching the Limit. *Elife* 7, e39804. doi:10.7554/eLife.39804

Camps, M. (2010). Modulation of ColE1-like Plasmid Replication for Recombinant Gene Expression. *Dnag* 4, 58–73. doi:10.2174/187221510790410822

Deshazer, D., and Woods, D. E. (1996). Broad-host-range Cloning and Cassette Vectors Based on the R388 Trimethoprim Resistance Gene. *Biotechniques* 20, 762–764. doi:10.2144/96205bm05

Eguchi, Y., Makanae, K., Hasunuma, T., Ishibashi, Y., Kito, K., and Moriya, H. (2018). Estimating the Protein burden Limit of Yeast Cells by Measuring the Expression Limits of Glycolytic Proteins. *Elife* 7, e34595. doi:10.7554/eLife.34595

Guo, Y., Hui, C.-Y., Liu, L., Zheng, H.-Q., and Wu, H.-M. (2019). Improved Monitoring of Low-Level Transcription in *Escherichia coli* by a β-Galactosidase

α-Complementation System. *Front. Microbiol.* 10, 1454. doi:10.3389/fmicb.2019.01454

Hecht, A., Glasgow, J., Jaschke, P. R., Bawazer, L. A., Munson, M. S., Cochran, J. R., et al. (2017). Measurements of Translation Initiation From all 64 Codons in *E. coli*. *Nucleic Acids Res.* 45 (7), 3615–3626.

Hollis, R. P., Lagido, C., Pettitt, J., Porter, A. J. R., Killham, K., Paton, G. I., et al. (2001). Toxicity of the Bacterial Luciferase Substrate,n-Decyl Aldehyde, toSaccharomyces cerevisiaeandCaenorhabditis Elegans. *FEBS Lett.* 506, 140–142. doi:10.1016/s0014-5793(01)02905-2

Kimelman, A., Levy, A., Sberro, H., Kidron, S., Leavitt, A., Amitai, G., et al. (2012). A Vast Collection of Microbial Genes that Are Toxic to Bacteria. *Genome Res.* 22, 802–809. doi:10.1101/gr.133850.111

Kintaka, R., Makanae, K., and Moriya, H. (2016). Cellular Growth Defects Triggered by an Overload of Protein Localization Processes. *Sci. Rep.* 6, 31774. doi:10.1038/srep31774

Liu, M., Blinn, C., Mcleod, S. M., Wiseman, J. W., Newman, J. V., Fisher, S. L., et al. (2014). Secreted Gaussia Princeps Luciferase as a Reporter of *Escherichia coli* Replication in a Mouse Tissue Cage Model of Infection. *PLoS One* 9, e90382. doi:10.1371/journal.pone.0090382

Malakar, P., Singh, V. K., Karmakar, R., and Venkatesh, K. V. (2014). Effect on β-galactosidase Synthesis and burden on Growth of Osmotic Stress in *Escherichia coli*. *Springerplus* 3, 748. doi:10.1186/2193-1801-3-748

Marschall, L., Sagmeister, P., and Herwig, C. (2016). Tunable Recombinant Protein Expression in *E. coli*: Enabler for Continuous Processing? *Appl. Microbiol. Biotechnol.* 100, 5719–5728. doi:10.1007/s00253-016-7550-4

Massey, S., Johnston, K., Mott, T. M., Judy, B. M., Kvitko, B. H., Schweizer, H. P., et al. (2011). *In Vivo* bioluminescence Imaging of *Burkholderia Mallei* Respiratory Infection and Treatment in the Mouse Model. *Front. Microbiol.* 2, 174.doi:10.3389/fmicb.2011.00174

Ren, A., Rajashankar, K. R., and Patel, D. J. (2012). Fluoride Ion Encapsulation by Mg2+ Ions and Phosphates in a Fluoride Riboswitch. *Nature* 486, 85–89. doi:10.1038/nature11152

Roth, T. L., Milenkovic, L., and Scott, M. P. (2014). A Rapid and Simple Method for DNA Engineering Using Cycled Ligation Assembly. *PLoS One* 9, e107329. doi:10.1371/journal.pone.0107329

Speed, M. C., Burkhart, B. W., Picking, J. W., and Santangelo, T. J. (2018). An Archaeal Fluoride-Responsive Riboswitch Provides an Inducible Expression System for Hyperthermophiles. *Appl. Environ. Microbiol.* 84, e02306. doi:10.1128/AEM.02306-17

Standley, M. S., Million-Weaver, S., Alexander, D. L., Hu, S., and Camps, M. (2019). Genetic Control of ColE1 Plasmid Stability that Is Independent of Plasmid Copy Number Regulation. *Curr. Genet.* 65, 179–192. doi:10.1007/s00294-018-0858-0

Stockbridge, R. B., Lim, H.-H., Otten, R., Williams, C., Shane, T., Weinberg, Z., et al. (2012). Fluoride Resistance and Transport by Riboswitch-Controlled CLC Antiporters. *Proc. Natl. Acad. Sci. USA* 109, 15289–15294. doi:10.1073/pnas.1210896109

Topp, S., Reynoso, C. M. K., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., et al. (2010). Synthetic Riboswitches that Induce Gene Expression in Diverse Bacterial Species. *Appl. Environ. Microbiol.* 76 (23), 7881–7884. doi:10.1128/aem.01537-10

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., et al. (2010). Comparative Genomics Reveals 104 Candidate Structured RNAs from Bacteria, Archaea, and Their Metagenomes. *Genome Biol.* 11, R31. doi:10.1186/gb-2010-11-3-r31

Xu, P., Gu, Q., Wang, W., Wong, L., Bower, A. G., Collins, C. H., et al. (2013). Modular Optimization of Multi-Gene Pathways for Fatty Acids Production in *E. coli*. *Nat. Commun.* 4 (1), 1 8.

Zolg, J. W., and Hänggi, U. J. (1981). Characterization of a R Plasmid-Associated, Trimethoprim-Resistant Dihydrofolate Reductase and Determination of the Nucleotide Sequence of the Reductase Gene. *Nucl. Acids Res.* 9, 697–710. doi:10.1093/nar/9.3.697

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fgene.2022.873869

OPEN ACCESS

EDITED BY
Nikolay Shirokikh,
Australian National University, Australia

REVIEWED BY
Moaz Ahmad,
National Institutes of Health,
United States
Shixiang Yao,
Southwest University, China
Abdulmojeed Yakubu,
Nasarawa State University, Nigeria

*CORRESPONDENCE
Jing-Fang Yang,
yangjf@mail.ccnu.edu.cn
Yun-Sheng Chen,
chenyunshenglw@163.com

†These authors contributed equally to this work

SPECIALTY SECTION
This article was submitted to RNA, a section of the journal Frontiers in Genetics

RECEIVED 11 February 2022
ACCEPTED 18 July 2022
PUBLISHED 02 September 2022

CITATION
Huang B-X, Jia Z-C, Yang X, Cheng C-L, Liu X-R, Zhang J, Chen M-X, Yang J-F and Chen Y-S (2022), Genome-wide comparison and in silico analysis of splicing factor SYF2/NTC31/p29 in eukaryotes: Special focus on vertebrates.
Front. Genet. 13:873869.
doi: 10.3389/fgene.2022.873869

Genome-wide comparison and *in silico* analysis of splicing factor SYF2/NTC31/p29 in eukaryotes: Special focus on vertebrates

Bao-Xing Huang[1†], Zi-Chang Jia[2,3†], Xue Yang[3], Chao-Lin Cheng[4], Xiao-Rong Liu[1], Jianhua Zhang[4], Mo-Xian Chen[2], Jing-Fang Yang[5*] and Yun-Sheng Chen[1*]

[1]Clinical Laboratory, Shenzhen Children's Hospital, Shenzhen, China, [2]Co-Innovation Center for Sustainable Forestry in Southern China and Key Laboratory of National Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, College of Biology and the Environment, Nanjing Forestry University, Nanjing, China, [3]State Key Laboratory of Crop Biology, College of Life Science, Shandong Agricultural University, Taian, China, [4]Department of Biology, Hong Kong Baptist University, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China, [5]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

The gene *SYF2*—an RNA splicing factor—can interact with Cyclin D-type binding protein 1 (GICP) in many biological processes, including splicing regulation, cell cycle regulation, and DNA damage repair. In our previous study we performed genome-wide identification and functional analysis of *SYF2* in plant species. The phylogenetic relationships and expression profiles of *SYF2* have not been systematically studied in animals, however. To this end, the gene structure, genes, and protein conserved motifs of 102 *SYF2* homologous genes from 91 different animal species were systematically analyzed, along with conserved splicing sites in 45 representative vertebrate species. A differential comparative analysis of expression patterns in humans and mice was made. Molecular bioinformatics analysis of *SYF2* showed the gene was conserved and functional in different animal species. In addition, expression pattern analysis found that *SYF2* was highly expressed in hematopoietic stem cells, T cells, and lymphoid progenitor cells; in ovary, lung, and spleen; and in other cells and organs. This suggests that changes in *SYF2* expression may be associated with disease development in these cells, tissues, or organs. In conclusion, our study analyzes the *SYF2* disease resistance genes of different animal species through bioinformatics, reveals the relationship between the *SYF2* genotype and the occurrence of certain diseases, and provides a theoretical basis for follow-up study of the relationship between the *SYF2* gene and animal diseases.

KEYWORDS

alternative splicing, SYF2, expression profile, gene family, proteogenomics

Introduction

In eukaryotic cells, gene expression can be roughly subdivided into three steps: transcription, splicing, and translation, and is performed by RNA polymerases, spliceosomes, and ribosomes. In 1977, scientists first discovered that adenovirus mRNA and its corresponding DNA transcription template did not form a continuous hybrid double strand, but was instead an extended circular single strand DNA at different locations. This suggests that genetic information is transferred from DNA to mRNA, not only by transcription, but also by RNA splicing (Berget et al., 1977). The pre-mRNA introns and exons produced during transcription are arranged alternately. Only after intron excision and exon splicing is complete can mature mRNA be generated and enter the translation process with coherent information (Chen et al., 2012). This is known as the splicing process. RNA splicing is an important process for regulating cell differentiation, proliferation, and survival, and is equally important in gene regulation. Splicing factors participate in the splicing process of RNA precursors, and their presence causes the final protein products to show different functional and structural characteristics, thereby increasing genetic diversity. In recent years, sequencing technology and transcriptome analyses have revealed that alternative splicing is ubiquitous in various species (Chen et al., 2021; Song et al., 2021), and can lead to profound changes in gene expression patterns during development (Chen M.-X. et al., 2020). The molecular mechanism of RNA splicing consists of a two-step transesterification reaction (Toro et al., 2007; Wan et al., 2016). This deceptively simple chemical reaction is difficult to perform in cells on its own, and it requires a spliceosome to complete it. During RNA splicing, a large and highly dynamic molecular machine in the cell nucleus is pieced together from many different components. It is a ribosomal protein complex that recognizes the splicing site of the RNA precursor, and catalyzes the splicing reaction. Its size is 60 S, and it is mainly dynamically composed of a variety of non-SNRNPs, assembled small nuclear ribonucleoproteins, and RNA (Madhani & Guthrie, 1994; Will & Lührmann, 1997). It is formed at various stages of splicing with the addition of snRNA. In addition, spliceosomes are classified into major and minor spliceosomes due to differences in the proportion of intron excisions they monitor. Approximately 99.5% of intron excision reactions, which are the primary spliceosomes during splicing, are monitored. The major spliceosomes can monitor approximately 99.5% of the intron resection response, while secondary spliceosomes are less efficient in monitoring intron excision reactions, accounting for only 0.5% (Lorkovic et al., 2005; Chen & Moore, 2015). When splicing occurs, there are two steps to the assembly of the spliceosome. First, the identification of the 3′ and 5′ splice sites is completed in a base-complementary manner, and the U2 snRNP is guided to bind to the branch site to form a splice precursor. The resulting product then combines with U4, U5, and U6 snRNP trimers to form a spliceosome (Will & Luhrmann, 2011). The average human body contains approximately 100,000 spliceosomes per cell. Spliceosomes can also be classified into two types, type I and type II. The first spliceosome contains five major snRNP subcomplexes: U1, U2, and U4 to U6. Five snRNPs, U5, U11, U12, U4atac, and U6atac, constitute the type II spliceosome (Chen & Moore, 2015).

SYF2, also called p29, or CBPIN, or NTC31, encodes a nuclear protein. *SYF2* interacts with Cyclin D-type binding-protein 1 (GICP), is involved in cell cycle regulation and pre-mRNA splicing, and plays an important role in cancer progression.

As a cell cycle regulator, *SYF2* induces the transition of the G1-to-S phase to promote cell proliferation by interacting with cyclin-D-type binding protein 1 (Witzel et al., 2010). Cyclin D1 induces the cell cycle transition of G1-to-S phase by interacting with cyclin-dependent kinase 4/6 (CDK4/6), thereby promoting cell proliferation (Tao et al., 2020).

SYF2 participates in the progress of diverse tumor entities, such as breast cancer (Shi et al., 2017), gastric cancer (GC) (Liu et al., 2019; Tao et al., 2020), human epithelial ovarian cancer (EOC) (Yan et al., 2015), hepatocellular carcinoma (Zhang et al., 2015), esophageal squamous cell carcinoma (ESCC) (Zhu et al., 2014), and glioma (Guo et al., 2014), in a cell cycle-dependent pathway. There is a positive correlation between *SYF2* expression and proliferation of cancer, with *SYF2* a potential novel tumor marker and an oncogene. *SYF2* might potentially be the molecular target for the treatment of cancer, i.e., knocking down *SYF2* would lead to cell cycle G1/S phase arrest, and hence to inhibition of cancer cell proliferation.

SYF2 may promote the replication checkpoint and S-phase arrest (slowdown) through both splicing-dependent and independent mechanisms. *SYF2* regulates DNA replication and cell cycle progression through AS regulation of ECT2-Ex5 (Tanaka et al., 2020). ECT2 is a protooncogene with an important role in the cytokinesis phase of the cell cycle. The ECT2-Ex5+ isoform promotes S-phase accumulation. The p29 gene is involved in intervertebral disc (IVD) degeneration (Cherif et al., 2022). *SYF2* interacts with PRP17, which is involved in the splicing and cell cycle (Ben-Yehuda et al., 2000). *SYF2* was hypomethylated in all superovulated oocytes (Huo et al., 2020).

SYF2 is involved in many biological processes, such as splicing regulation, cell cycle regulation, and DNA damage repair. Our previous study performed genome-wide identification and functional analysis of *SYF2* in plant species (Tian et al., 2019). However, the phylogenetic relationships and expression profiles of *SYF2* in animals have not been systematically studied. In this study, we used a variety of bioinformatics methods to systematically analyze the gene structures, gene and protein motifs, and splicing conservation of the animal *SYF2* gene family. The differential expression patterns of *SYF2* in different diseases, different organs and

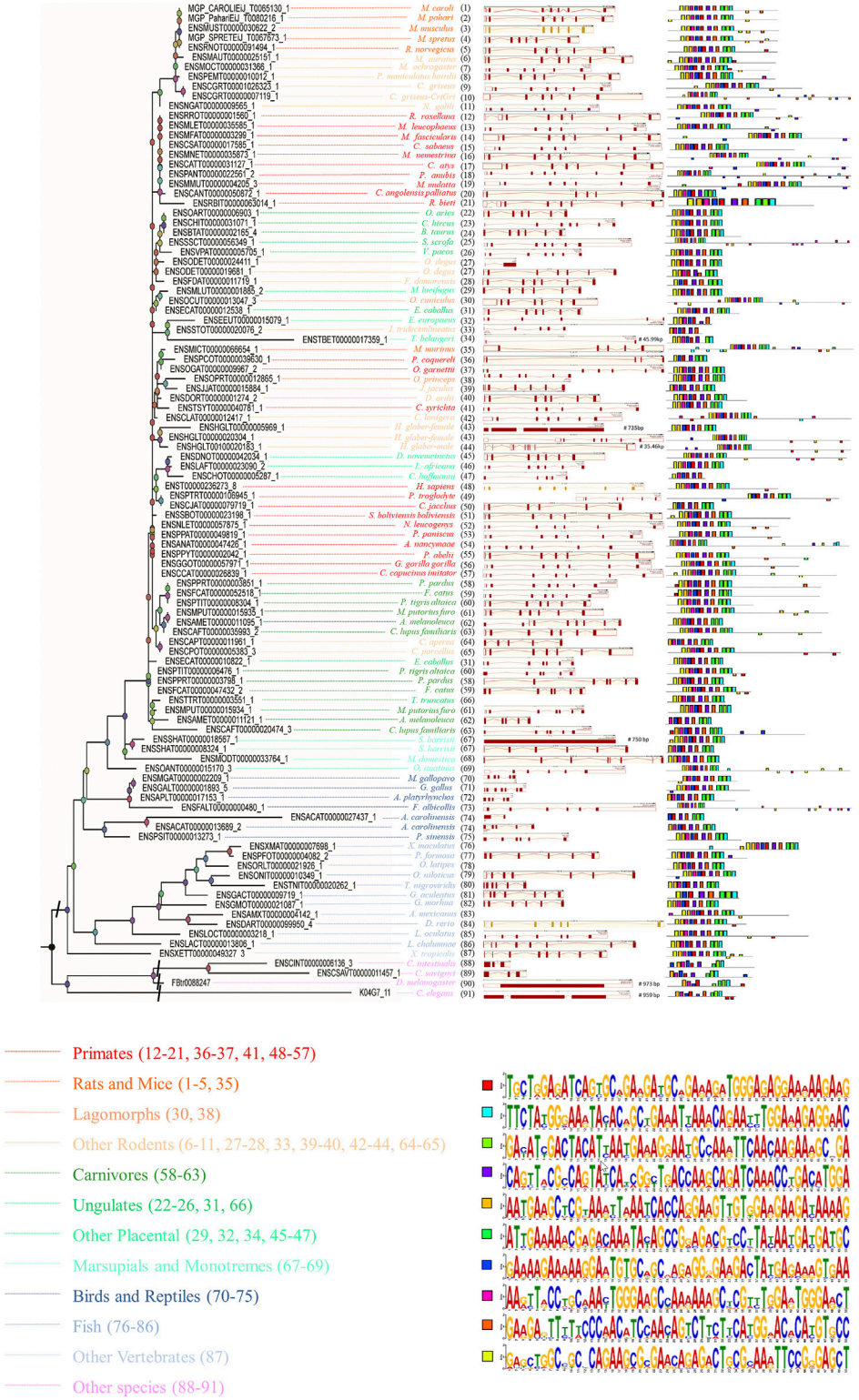


FIGURE 1
The phylogenetic relationship analysis of animal *SYF2* genes. The phylogenetic relationship is listed in the left panel. The gene structure is listed in the middle panel, and conserved motifs of cDNA sequences are listed in the right panel. The sequence of each conserved motif is listed below the phylogenetic tree.

tissues, different cells, different developmental stages, and different sampling time points in humans and mice are also discussed. This suggests that *SYF2* may be involved in the development of disease, and may be a molecular target for the treatment of cancer. This study aims to reveal the relationship between *SYF2* genotypes and biological disease processes from the perspective of bioinformatics, and to provide some basic theories for subsequent research into *SYF2* as a novel tumor marker.

Results

Phylogenetic tree construction of animal *SYF2* genes

In order to gain a deeper understanding of the function of *SYF2*, the possible *SYF2* genes in different animal species were determined according to the amino acid sequence of human *SYF2* (*Homo sapiens*, *ENST00000236273_8*). Ultimately, we found by alignment a total of 102 homologous sequences from 91 animal species, including 23 primates, 6 rats and mice, 18 other rodents, 12 carnivores, 11 fish, 8 ungulates, 7 birds and reptiles, 6 other placentals, 4 marsupials and monotremes, 2 lagomorphs, 1 other vertebrate and 4 other species (outgroup) (Supplementary Table S1). For construction of the phylogenetic tree of the *SYF2* gene, using the 102 amino acid sequence similarity of 91 animal species, see Figure 1, left. The phylogenetic tree has five main branches: primates, vertebrates, mammals, rodents and lagomorphs, and other species. These five main branches are then subdivided into twelve smaller branches. Among them, rodents and lagomorphs include rats, mice, lagomorphs, and other rodents. Other mammals include other placentals, marsupials, monotremes, carnivores, and ungulates. Other vertebrates include birds, reptiles, and fish.

Conserved motif analysis of *SYF2*

To further study the conserved nature of *SYF2*s in animals, the gene structure, cDNA, and conserved peptide motif of *SYF2*s were analyzed. The genetic structure of *SYF2* of each animal species is shown in Figure 1 and Supplementary Table S2, which shows the number of introns and exons per sequence, and the presence of untranslated regions other than CDS. In general, the gene structure of this family is diverse, with each gene containing from 1 to 9 exons. We found different exon-intron organization in different subgroups within the same general class (Figure 1). For instance, among all *SYF2* members, there are two *degus*, whose sequences exist in a subgroup, where *ENSODET00000019681_1* has seven exons and *ENSODET00000024411_1* has only one long exon (Figure 1). Similarly, two *harrisii* exist in the same subgroup, and a similar situation exists: *ENSSHAT00000008324_1* has 6 exons, while *ENSSHAT00000018567_1* has only one long exon (Figure 1). Overall, the exon-intron distribution pattern of the *SYF2* gene varied across all animal species involved in this study and also within the same genus. It is indicated that the changes of gene structure are of great significance to the development and evolution of their gene families. Furthermore, the conserved motifs of each *SYF2* were compared and analyzed using MEME software, and it was found that 76 of the 102 sequences shared the same 10 motifs and had similar organizational structures. The remaining 26 sequences had some changes in the number or structure of conserved motifs. In addition, some sequences had less conserved motifs, thus indicating their functional diversity. The sequence of *C. angolensis palliatus* (*ENSCANT00000050872_1*), *O. degus* (*ENSODET00000019681_1*), *M. murinus* (*ENSMICT00000066654_1*), *J. jaculus* (*ENSJJAT00000015884_1*), *H. glaber-female* (*ENSHGLT00000005969_1*), *C. aperea* (*ENSCAPT00000011961_1*), *C. lupus familiaris* (*ENSCAFT00000020474_3*), *S. harrisii* (*ENSSHAT00000008324_1*), *M. gallopavo* (*ENSMGAT00000002209_1*), *A. platyrhynchos* (*ENSAPLT00000017153_1*), *A. carolinensis* (*ENSACAT00000013689_2*), *C. intestinalis* (*ENSCINT00000006136_3*), and *C. elegans* (*K04G7_11*) had nine motifs; *F. damarensis* (*ENSFDAT00000011719_1*), *T. belangeri* (*ENSTBET00000017359_1*), *P. troglodyte* (*ENSPTRT00000106945_1*), *C. jacchus* (*ENSCJAT00000079719_1*), *A. melanoleuca* (*ENSAMET00000011121_1*), *G. aculeatus* (*ENSGACT00000009719_1*), *C. savignyi* (*ENSCSAVT00000011457_1*), and *D. melanogaster* (FBtr0088247) had eight motifs; *E. europaeus* (*ENSEEUT00000015079_1*), *P. coquereli* (*ENSPCOT00000039630_1*), and *A. carolinensis* (*ENSACAT00000027437_1*) had seven motifs; *I. tridecemlineatus* (*ENSSTOT00000020076_2*) and *C. hoffmanni* (*ENSCHOT00000005287_1*) had six motifs, suggesting that the *SYF2* gene sequence is diverse in different species. Furthermore, there was no correlation between conserved motifs and gene structure after comparative analysis. For instance, *S. harrisii* (*ENSSHAT00000018567_1*) had only one long exon, but contained all 10 motifs, while *C. hoffmanni* (*ENSCHOT00000005287_1*) had six exons and five introns. These sequences had only six motifs.

In addition, peptide level analysis was performed. A total of 102 sequences in different species were annotated with the *SYF2* domain (Figure 2), and MEME analysis was used to predict unknown protein motifs. Among the 102 animals belonging to different species, the animals with 8 conserved motifs accounted for the majority. Motifs shown in bright red (NERNAKFNKKAERFYGKYTAEIKQNLERGTA), pale green (DYAAAQLRQYHRLTKQIKPDMETYERL) blue (HGEEFFPTSNSLLHGTHVPSTEEIDRMV), green (EKRDKYSRRRPYNDDADIDYI), purple (KKECAARGEDYEKVKLLEISAEDAERWERKKKRKNPDLGFS), ginger (RNEARKLNHQEVVEEDKRLKLPANWEAKKARLEWELQEEEK), and dark blue (AEELAAQKREQRLRKFRELHL) occur in nearly every one of

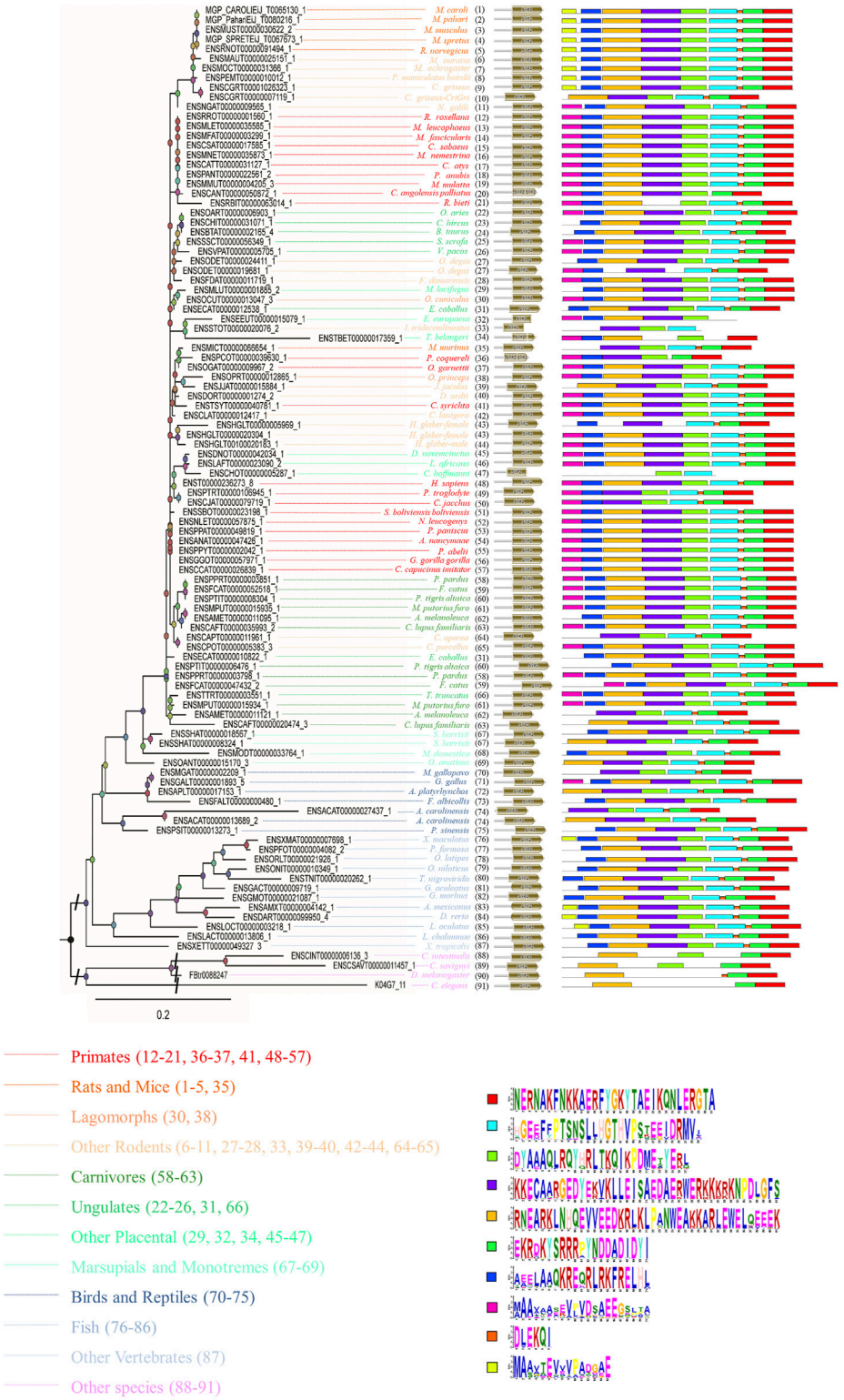


FIGURE 2
The phylogenetic relationship analysis of animal *SYF2* proteins. The phylogenetic relationship is listed in the left panel, the peptide structure is listed in the middle panel, and conserved motifs of peptide sequences are listed in the right panel. The sequence of each conserved motif is listed below the phylogenetic tree.

them, accounting for 70% of all motifs analyzed in this paper. Furthermore, the rose red motifs (MAAXAASE VPVDSAEEGSLTA) were concentrated in primate sequences, and only sparsely distributed in sequences of other animal species. The yellow motif (MAAXTEVVVPADGAE) only existed in rat, mouse, fish, and other rodent sequences, but was not detected in other animal sequences. In summary, through the analysis of conserved motifs at the RNA/cDNA and protein levels, it can be found that there is little difference in codon usage among *SYF2* orthologs, and the number and position of motifs are obviously similar. This indicates that *SYF2* is highly conserved among different proteins and cDNAs in different animal species.

Construction of the *SYF2* protein interaction network

We further explored how *SYF2* plays a role in biological regulatory processes. Next, we employed the tool STRING to construct a protein interaction network of *SYF2* in different species (Figure 3). We used protein intercrops to interact with *SYF2* in organisms, including two animal species (*Homo sapiens*, *Mus musculus*), yeast, and two plant species (*Arabidopsis* and *Oryza sativa*).

Human XAB2 protein is a protein interactor of *SYF2*. It consists of 15 repeated tetrapeptides, and was identified by interaction with xeroderma pigmentosa histone A (XPA) (Nakatsu et al., 2000; Kuraoka et al., 2008). It is a novel component involved in transcriptional coupling repair and transcription, and plays a role in mitotic cell cycle regulation (Hou et al., 2016). Furthermore, there is also a clear interaction between CDC40 and *SYF2*. The CDC40 (PRP17) gene in *S. cerevisiae*, whose mutation results in sensitivity to temperature changes, was originally identified in CDC40-1 (Orna & Martin, 2002). It plays a variety of roles in cell cycle progression. Its mutation causes the cell cycle to stall (Kaplan & Kupiec, 2007). In addition, Cdc40p, Slu7p, Prp22p, Prp18p, Prp16p, and Prp8p acted as pre-mRNA splicing factors during the second splicing reaction stage (Vijayraghavan et al., 1989; Schwer & Gross, 1998).

In the mouse protein interaction map, we also found proteins with high interaction with *SYF2*. Prpf19 is a functionally diverse protein (Yin et al., 2012), is highly conserved, and participates in splicing as a splicing factor (He et al., 2021).

Analysis of transcript isoforms and conserved splice sites

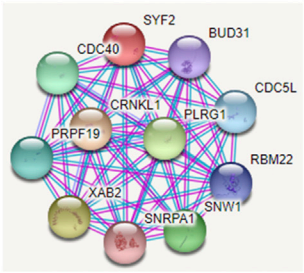
To further understand the splicing patterns and conserved splicing sites of the *SYF2* gene, we extracted some available animal *SYF2* gene transcription subtypes from the Ensembl database and then selected 45 representative animals for alternative splicing analysis (Figure 4). A total of 106 splice isomers were obtained from 45 *SYF2* genes, with an average of 2–3 transcripts per gene. Among them, the human and *Oryctolagus cuniculus SYF2* genes annotated 4 subtypes, the largest number of annotated subtypes in these animals. When comparing the conserved motifs of the transcription subtype with the genome structure (Figure 4, right), it was found that the original transcripts of most genes contained the most motifs, while the remaining replacement transcripts usually contained fewer motifs. Tthe variable splicing event was then analyzed. First, in *Mandrillus leucophaeus*, *Macaca fascicularis*, *Macaca nemestrina*, *Cercocebus atys*, *Colobus angolensis palliatus*, *Carlito syrichta*, *Callithrix jacchus*, *Saimiri boliviensis boliviensis*, *Pan paniscus*, *Aotus nancymaae*, *Gorilla gorilla*, *Cebus capucinus imitator*, *Panthera pardus*, *Felis catus*, *Panthera Tigris altaica*, etc., there was a large amount of exon skipping. Second, an in-depth study found that the number of alternative splicing events of the first exon and the last exon (AFE and ALE), accounted for the bulk of the total alternative splicing events in *SYF2*, which indirectly led to the generation of many truncated transcript subtypes, such as in *Mus caroli*, *Mus pahari*, *Rhinopithecus roxellana*, *Macaca fascicularis*, *Cercocebus atys*, *Rhinopithecus bieti*, *Microcebus murinus*, *Gorilla gorilla gorilla*, *Oryctolagus cuniculus,* and so on. Moreover, alternative transcription initiation and alternative polyadenylation of several transcripts have been found, such as in *Capra hircus*, *Microcebus Murinus*, and *Danio rerio*.

Analysis of the *SYF2* expression profile in animals

To further investigate the association between the animal *SYF2* gene and certain cell, tissue, and organ diseases, we analyzed the expression patterns of the *SYF2* gene in different animal species. Mice have genome sequences that are highly similar to humans, with gene homology as high as 78.5%. In addition, mice are close to humans in terms of biological evolution, their tissue and organ structure and cell functions are similar to those of humans, while their placenta formation and early embryonic development are also similar to humans. Therefore, we performed a comparative analysis of *SYF2* gene expression patterns in humans and mice. Through the BAR HeatMapper Plus tool, we reconstructed the expression profile to include three aspects: human disease (Figure 5), human and mouse tissues and organs (Figures 6 and 7), and human and mouse cell types and development stages (Supplementary Figures S1–S3).

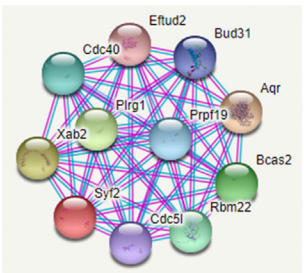
First, accumulation of *SYF2* was found in breast tumor lumen, triple-negative breast cancer, and HER2-positive breast cancer (Figure 5). Second, based on multiple datasets, *SYF2*s were highly expressed in spleen and lung tissue in both humans and mice (Figures 6 and 7). However, there are also differences in the

Homo sapiens



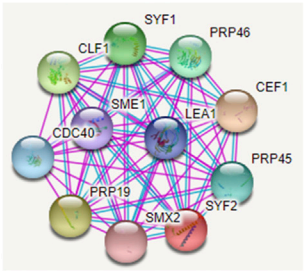
Gene ID	Gene name
ENSP00000236273	SYF2
ENSP00000366557	CRNKL1
ENSP00000351137	XAB2
ENSP00000424417	PLRG1
ENSP00000261531	SNW1
ENSP00000357928	CDC40
ENSP00000227524	PRPF19
ENSP00000360532	CDC5L
ENSP00000199814	RBM22
ENSP00000386023	BUD31
ENSP00000254193	SNRPA1

Mus musculus



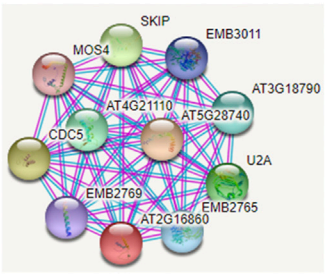
Gene ID	Gene name
ENSMUSP00000030622	Syf2
ENSMUSP00000047157	Aqr
ENSMUSP00000019614	Xab2
ENSMUSP00000114968	Plrg1
ENSMUSP00000005830	Bcas2
ENSMUSP00000025506	Rbm22
ENSMUSP00000044305	Cdc40
ENSMUSP00000136858	Prpf19
ENSMUSP00000125590	Bud31
ENSMUSP00000024727	Cdc5l
ENSMUSP00000021306	Eftud2

Saccharomyces cerevisiae



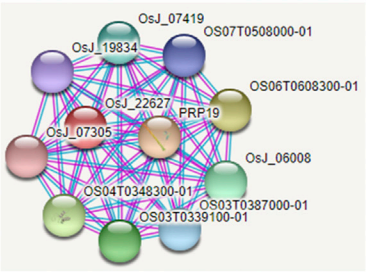
Gene ID	Gene name
YGR129W	SYF2
YMR213W	CEF1
YLL036C	PRP19
YLR117C	CLF1
YDR416W	SYF1
YPL151C	PRP46
YAL032C	PRP45
YDR364C	CDC40
YPL213W	LEA1
YOR159C	SME1
YFL017W-A	SMX2

Arabidopsis thaliana



Gene ID	Gene name
AT2G16860.1	AT2G16860
AT5G28740.1	AT5G28740
AT1G09770.1	CDC5
AT1G77180.2	SKIP
AT1G09760.1	U2A
AT4G21110.1	AT4G21110
AT3G18790.1	AT3G18790
AT2G38770.1	EMB2765
AT5G13010.1	EMB3011
AT3G13200.1	EMB2769
AT3G18165.1	MOS4

Oryza sativa



Gene ID	Gene name
OS06T0711600-01	OsJ_22627
OS10T0466300-01	PRP19
OS06T0608300-01	OS06T0608300-01
OS04T0348300-01	OS04T0348300-01
OS03T0339100-01	OS03T0339100-01
OS02T0234800-01	OsJ_06008
OS02T0602500-02	OsJ_07419
OS03T0387000-01	OS03T0387000-01
OS07T0508000-01	OS07T0508000-01
OS06T0106100-01	OsJ_19834
OS02T0586500-02	OsJ_07305

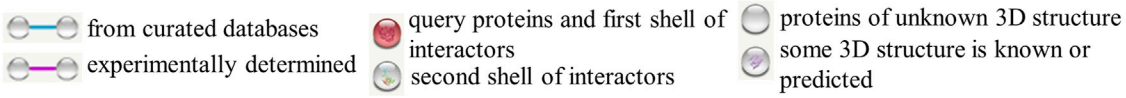


FIGURE 3
Protein interaction network diagrams of representative animal and plant species. Known interactions, either determined by experiments (pink line), or from curated databases (blue line), are presented in protein–protein interaction networks. *SYF2s* of *Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Oryza sativa* were used as query proteins for analysis by the STRING database. Highly scored interactors are presented in the form of a network diagram. Empty notes are proteins with unknown 3D structures, while filled notes are proteins with known or predicted 3D structures in the current database.

locations and levels of *SYF2* expression between humans and mice. Human *SYF2* accumulates specifically in the reproductive organs, bladder, thyroid, and colon (Figure 6), while mouse *SYF2* is abundant in tissues such as the cerebellum and thymus (Figure 7). Third, analysis of cell-type expression profiles showed that in humans, *SYF2* is high in common lymphoid progenitors and hematopoietic stem cells. By contrast, mouse *SYF2* accumulates in both native thymus-derived CD4-positive αβ T cells and induced T regulatory cells (Supplementary Figures S1 and S3). Moreover, *SYF2* was expressed at high levels in mice of different strains, and at different sampling times, and different developmental and somite stages (Supplementary Figure S2). The

FIGURE 4
Summary of splice isoforms of the animal *SYF2* gene. Transcript isoforms from 45 animal SYF2 genes are summarized (left and middle panel). Conserved protein motifs and sequences of potential protein products from splicing isoforms are illustrated (right panel and bottom of the figure, respectively), with additional annotation to define exon–exon boundaries (blue lines between boxes).

developmental map showed a high abundance of human *SYF2* in both the fetal and juvenile stages (Supplementary Figure S1). Across the developmental stages in mice, the accumulation of *SYF2* was higher in two stages—the embryonic stage and a few days after parturition (Supplementary Figure S2). In summary, the expression patterns of *SYF2* in humans and mice are not consistent, indicating that different species have different expression patterns due to the existence of different transcription and translation patterns. However, the study of *SYF2* in different species will help to reveal more possible regulatory roles of *SYF2*, which is conducive to the further analysis of *SYF2* function. Comparative analysis of expression patterns in humans and mice can help provide a theoretical basis for research into, and the treatment of some diseases. The comparison of human and mouse *SYF2* gene expression patterns across tissues, cell types, and developmental stages is summarized in Table 1.

Discussion

Phylogenetic and splicing pattern analysis indicates *SYF2* conserved among animals

Numerous existing reports suggest that 15–35% of human disease is caused by mis-splicing or mis-assembly of spliceosome complex proteins (Shi, 2017). However, underlying evidence for how mis-splicing causes disease is lacking. Therefore, understanding the underlying mechanisms of splicing regulation will not only contribute to the decoding of the eukaryotic splicing machinery, but may also provide new targets for clinical drug discovery. We performed phylogenetic and splicing pattern analyses of *SYF2* in this work to reveal its structural conservation and potential regulatory mechanisms across different animal species.

Phylogenetic topology shows that *SYF2* proteins can be divided into 12 groups: primates, rats and mice, lagomorphs, other rodents, carnivores, ungulates, other placentas, marsupials and monotremes, birds and reptiles, fish, other vertebrates, and other species. In this, all vertebrate species are aggregated into one large group, showing distant relationships with other species, such as *Ciona intestinalis*, *Ciona savignyi*, *Caenorhabditis elegans*, and *Drosophila melanogaster* (Figure 1). Furthermore, animal *SYF2* were subjected to conserved splicing pattern analysis (Figure 4). Similar to *SYF2* previously reported in plants (Tian et al., 2019), truncated transcripts exist for the animal *SYF2* gene, resulting in the creation of a conserved protein form with N-terminal truncation (Figure 4). Splice site analysis revealed that AFE and ALE were the most prominent AS events in the numerous animal species involved. In terms of the number of transcript isoforms, the *SYF2* gene in different animal species generally has more than one transcript isoform, but it is worth noting that each transcript isoform has a similar structure, suggesting that they have similar functions in the regulation of gene expression. Moreover, most protein isoforms corresponding to each transcriptional isoform were considered functional. Therefore, the relevant biological functions of *SYF2* protein isoforms in animal species need further study.

Differential expression patterns of animal *SYF2*s reveal functional diversity

SYF2 induces a transition from the G1 to S phase to promote cell proliferation, and does so by interacting with the cyclin-D-type binding protein 1 (Witzel et al., 2010). Embryonic and juvenile stages are important periods of cell growth and development in the developmental cycle of animals. Human and mouse expression profiling data indicate significant enrichment of *SYF2* both in human fetal and juvenile stages, and in mouse embryonic and postpartum periods. Previous studies have also shown that disruption of *SYF2* in mice leads to embryonic lethality (Chen et al., 2012). In *Arabidopsis*, high enrichment of *SYF2* was clearly detected in the shoot apex during flowering transformation, but decreased enrichment and repressed expression of *SYF2* were found in more mature pollen (Tian et al., 2019). This suggests that *SYF2* can participate in embryonic developmental regulation by mediating cell cycle regulation.

The regulation of the cell cycle by *SYF2* is also associated with the occurrence of many cancers. In the detection of disease expression profiles, *SYF2* was significantly expressed in breast tumor lumen, triple-negative breast cancer, and HER2-positive breast cancer (Figure 5). In addition, in the organ-tissue-related expression profiles, *SYF2* was enriched in the human ovary, testis, spleen, lung, bladder, thyroid, and colon (Figure 6), as well as in the mouse cerebellum, thymus, spleen and lung (Figure 7). Among these, many organ diseases caused by abnormal expression of *SYF2* have been confirmed. For example, overexpression of *SYF2* affects the cell cycle or cell proliferation leading to the occurrence and progress of breast cancer (Shi et al., 2017), non-small cell lung cancer (Liu et al., 2015; Chen et al., 2019), and ovarian cancer (Yan et al., 2015).

Interestingly, the comparative analysis of human and mouse expression data revealed that *SYF2* expression patterns were different in different cells, tissues, organs, and developmental stages of humans and mice. These results indicate that the regulatory patterns of transcription and translation vary by species, although this is not absolute. Some similarities have been detected in expression in certain organs and tissues. For instance, high expression of *SYF2* was detected in organs such as the lung and spleen of both humans and mice (Figures 6 and 7). These results provide new entry points for the treatment of certain organ diseases.

In this article, we compared and analyzed the expression patterns of humans and mice, and summarized the experimental

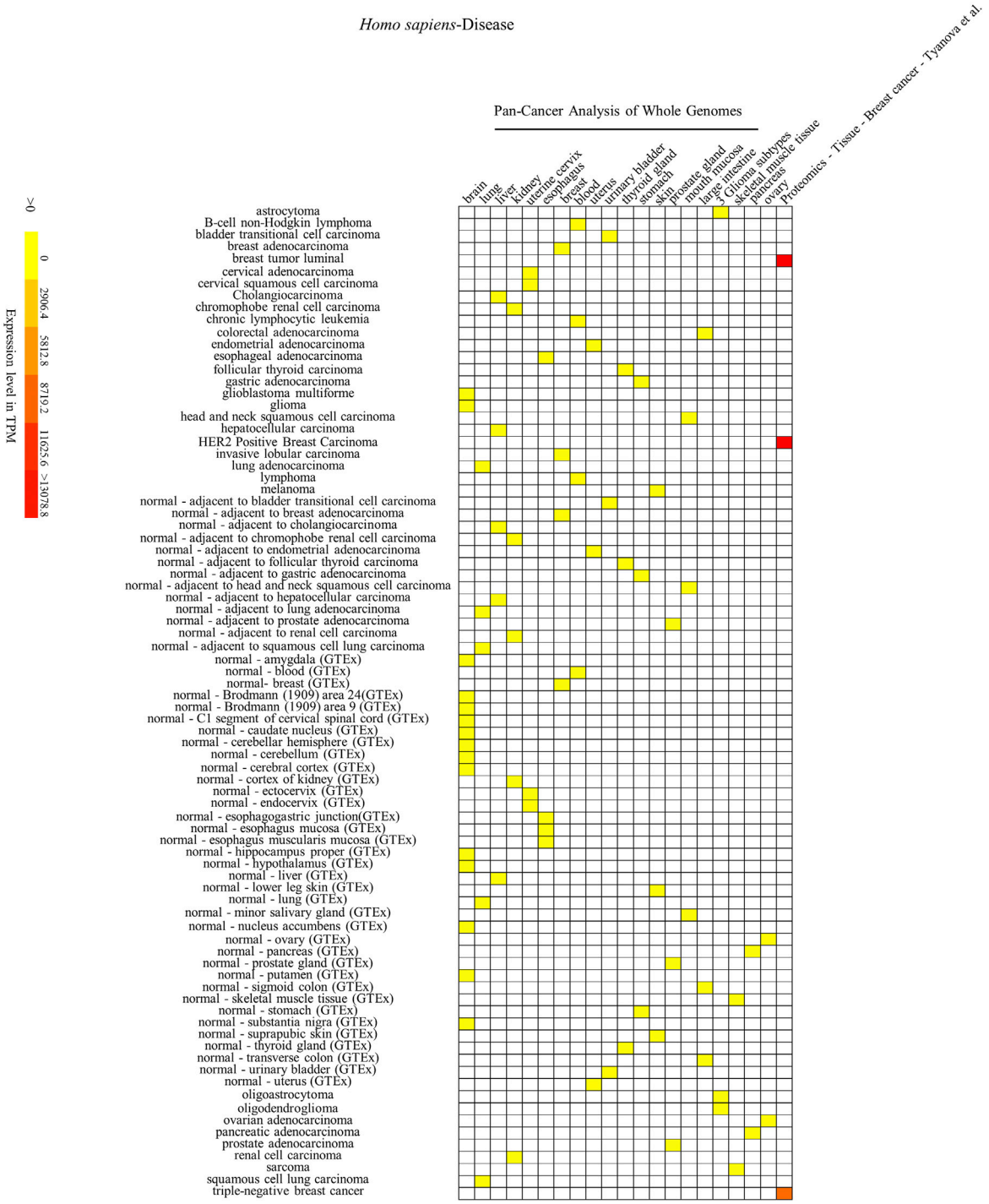


FIGURE 5
Heatmap of *Homo sapiens*' SYF2 gene expression in different diseases. Disease database is from the International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG). GTEx means Genotype-Tissue Expression (GTEx) Project. No. 1–21 represent 21 data source projects, which are Pan-Cancer Analysis of Whole Genomes – (brain, lung, liver, kidney, uterine cervix, esophagus, breast, blood, uterus, urinary bladder, thyroid gland, stomach, skin, prostate gland, mouth mucosa, large intestine) (1–16), 3 Glioma subtypes (17), Pan-Cancer Analysis of Whole Genomes – (skeletal muscle tissue, pancreas, ovary) (18–20), Proteomics – Tissue – Breast Cancer – Tyanova (21), respectively. Baseline expression levels are in TPM (transcripts per million). The raw data were reorganized and presented as heatmaps using online BAR HeatMapper Plus software (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi).

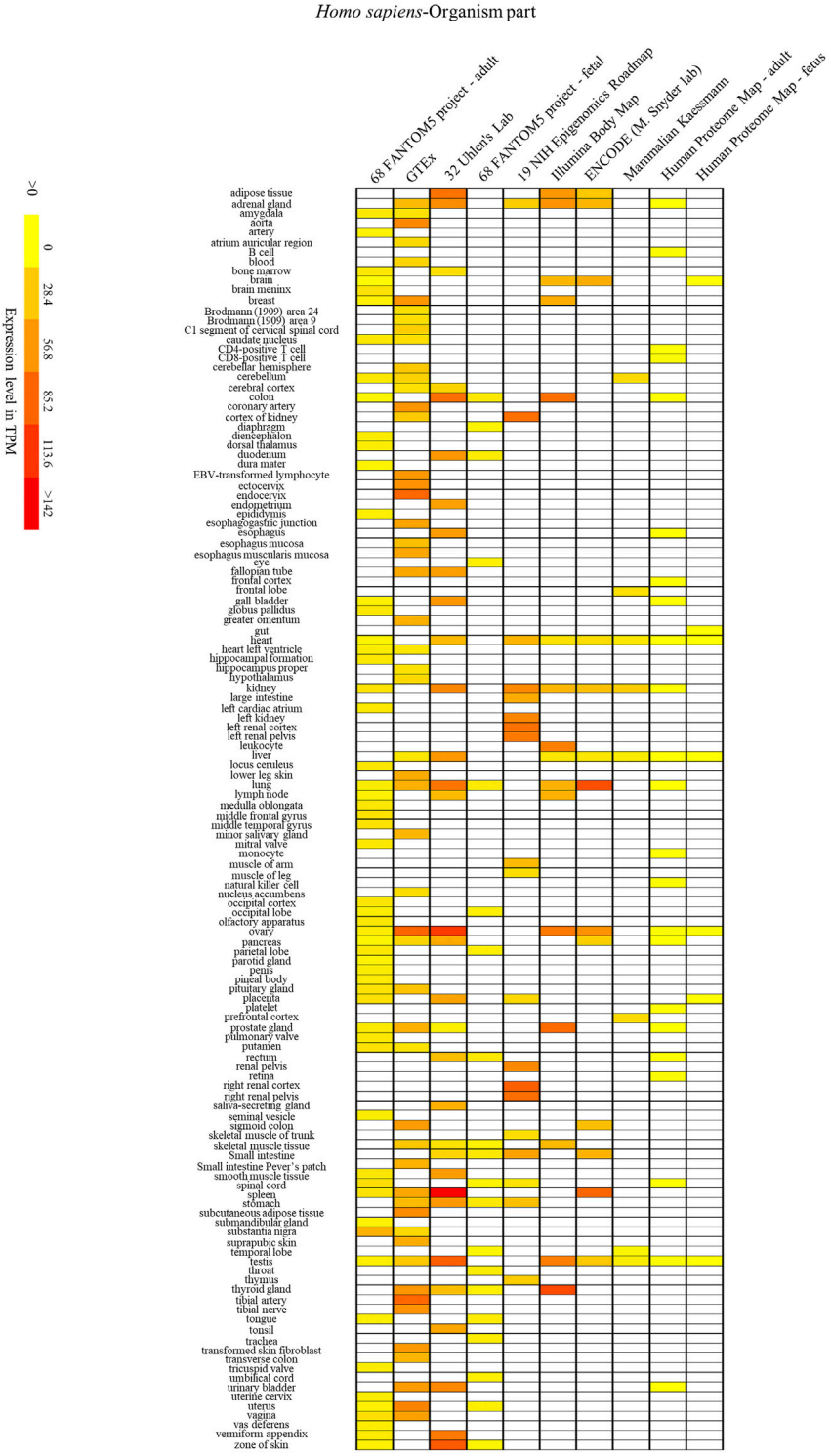


FIGURE 6
The human SYF2 gene is specifically expressed in different organs. Nos. 1–10 represent 10 data source projects, which are 68 FANTOM5 project – adult (1), GTEx (2), 32 Uhlen's Lab (3), 68 FANTOM5 project – fetal (4), 19 NIH Epigenomics Roadmap (5), Illumina Body Map (6), ENCODE (M. Snyder lab) (7), Mammalian Kaessmann (8), Human Proteome Map – adult (9), Human Proteome Map – fetus (10), respectively. Baseline expression levels are in TPM (transcripts per million). The raw data were reorganized and presented as heatmaps using online BAR HeatMapper Plus software (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi).

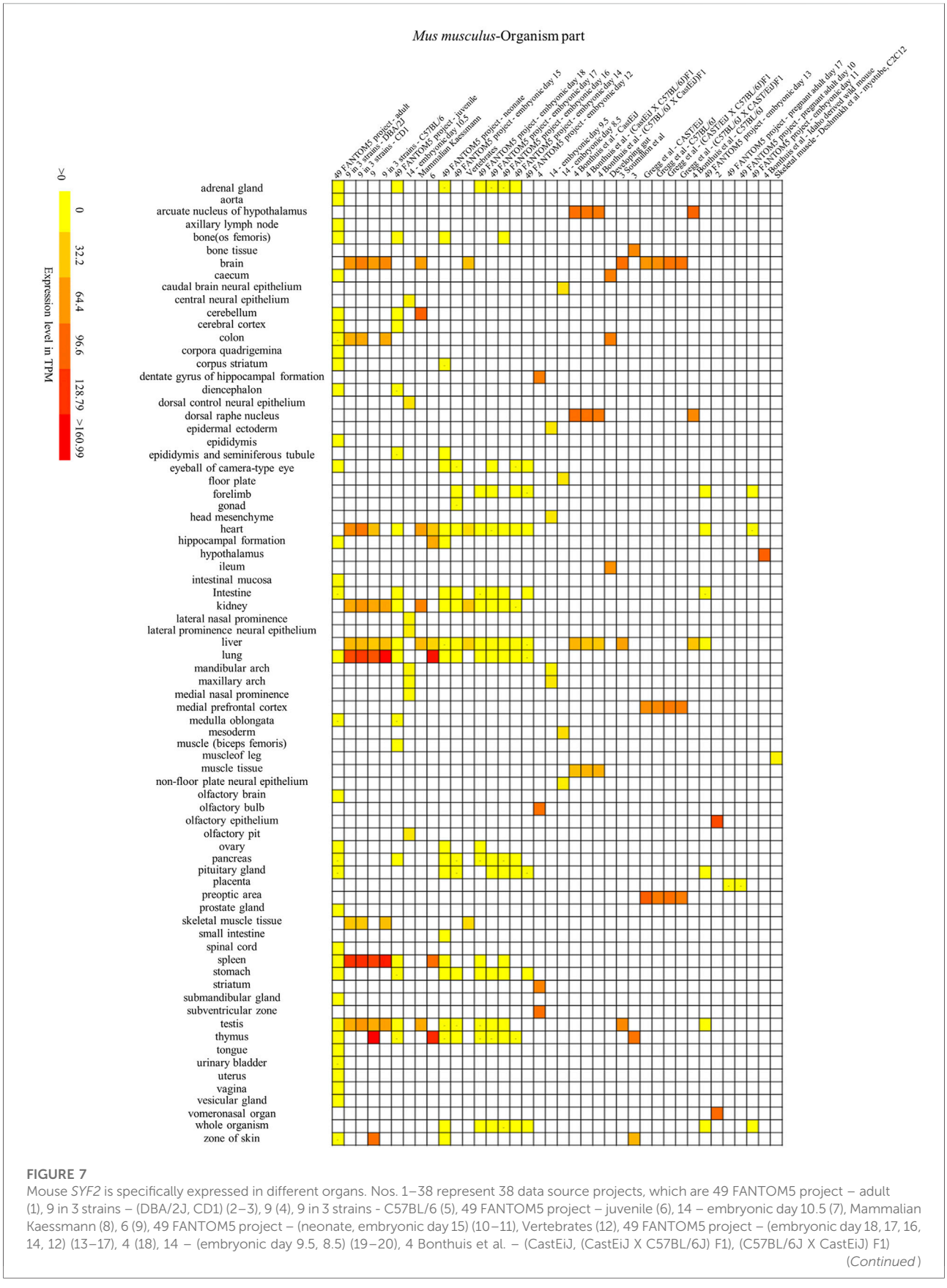


FIGURE 7
Mouse *SYF2* is specifically expressed in different organs. Nos. 1–38 represent 38 data source projects, which are 49 FANTOM5 project – adult (1), 9 in 3 strains – (DBA/2J, CD1) (2–3), 9 (4), 9 in 3 strains - C57BL/6 (5), 49 FANTOM5 project – juvenile (6), 14 – embryonic day 10.5 (7), Mammalian Kaessmann (8), 6 (9), 49 FANTOM5 project – (neonate, embryonic day 15) (10–11), Vertebrates (12), 49 FANTOM5 project – (embryonic day 18, 17, 16, 14, 12) (13–17), 4 (18), 14 – (embryonic day 9.5, 8.5) (19–20), 4 Bonthuis et al. – (CastEiJ, (CastEiJ X C57BL/6J) F1), (C57BL/6J X CastEiJ) F1)
(*Continued*)

FIGURE 7
(21–23), Developing gut (24), 3 Soumillon et al. (25), 3 (26), Gregg et al. – (CAST/EiJ, C57BL/6J, (CAST/EiJ X C57BL/6J) F1, (C57BL/6J X CAST/EiJ) F1) (27–30), 4 Bonthuis et al. – C57BL/6J (31), 49 FANTOM5 project – embryonic day 13 (32), 2 (33), 49 FANTOM5 project – (pregnant adult day 17, 10, embryonic day 11) (34–36), 4 Bonthuis et al. – Idaho derived wild mouse (37), Skeletal muscle – Deshmukh et al. – myotube, C2C12 (38), respectively. Baseline expression levels are in TPM (transcripts per million). The raw data were reorganized and presented as heatmaps using online BAR HeatMapper Plus software (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi).

TABLE 1 Comparison of human and mouse *SYF2* gene expression patterns in tissues, cell types and developmental stages.

	Human	Mouse
Tissues	Reproductive organs	The cerebellum
	Bladder	Thymus
	Thyroid and colon	
Cell-type expression profiles	Lymphoid progenitors	Native thymus-derived CD4-positive αβ T cells
	Hematopoietic stem cells	Induced T regulatory cells
Developmental stages	Fetal stages	The embryonic stage
	Juvenile stages	A few days after parturition

data in each project by means of bioinformatics. We offered preliminarily speculation on the possible function of *SYF2*, which is expected to provide a direction and a theoretical basis for research into clinically relevant diseases. Analysis of results may be affected by different experimental and sampling conditions between projects. However, modern SWATH-MS proteomics technology (Chen M. X. et al., 2020; Shen et al., 2021) could be used to study the potential function of *SYF2* further, and to verify the existing analysis results. It would be helpful to explore expression of the potential function of *SYF2*, and in so doing create more possibilities for the treatment of diseases caused by its abnormal expression.

Comparison of *SYF2* in animals, yeast, and plants

The *SYF2* in animals, yeast, and plants (*Arabidopsis*, *Oryza sativa*) was compared. First, by analyzing the interaction network, it was found that there were only 1–2 common interacting proteins among the three species, which indicates that *SYF2* has specific regulatory networks in animals, plants, and yeast (Figure 3). Second, transcriptional isoforms of *SYF2* averaged 2–3 in animal species (Figure 4), with only one copy in most yeast and plants (Tian et al., 2019). This suggests that *SYF2* may play a more important role in animal species. The functional roles of *SYF2* in these three species are conserved to some extent.

Conclusion

Throughout the study, we screened 102 *SYF2* genes in 91 animal species and analyzed their phylogeny, gene structure, gene and protein motifs, conservation of splicing patterns, and expression patterns. Analysis of related structures, motifs, and splicing patterns showed that *SYF2* is highly conserved in many animal species. In addition, the analysis of expression patterns showed that *SYF2* is associated with the occurrence of cancer in breast, lung, spleen, and reproductive organs, as well as other diseases. These results are intended to help reveal the possible relationship between the *SYF2* genotype and the occurrence of certain diseases, which can provide information about subsequent *SYF2* expression in studies where animals provide the basis.

Materials and methods

Identification and screening of *SYF2* protein sequences in animals

In the Ensembl database (http://asia.ensembl.org/), protein BLAST was performed based on the *Homo sapiens SYF2* protein sequence (ENST00000236273_8) as a template. All available gene sequences were found in animal genomes, and further screening was performed through HMMER 3.2.1 (Johnson et al., 2010).

Construction of the *SYF2* gene phylogenetic tree in animals

A phylogenetic tree was constructed using the protein sequences of 102 *SYF2* genes obtained from the Ensembl database. Where a gene had more than one transcript, the longest coding sequence was selected. Selected sequences were subjected to comparative analysis by Muscle V3.8 (Edgar, 2004), after which a root phylogenetic tree was constructed using maximum likelihood implemented in PhyML V3.03730

(Gascuel, 2010). Finally, FigTree V1.4.3.3831 was used to edit and present the phylogenetic tree (Morariu et al., 2008). The reliability of the phylogenetic tree was tested by bootstrapping repeated sampling. Nucleotide sites were randomly selected from the original sequence to form a new set of gene sequences, and the same method was used to construct another phylogenetic tree. The topology of this phylogenetic tree was repeatedly compared with the structure of the original tree. Internal branches of the original phylogenetic tree with the same sequence separation as the bootstrap value were assigned a value of 1, while other internal branches were assigned a value of 0. We calculated the percentage of eigenvalue 1 obtained for each internal branch of the original phylogenetic tree to verify the reliability of the phylogenetic tree (Katsura et al., 2017; Fan et al., 2021).

Analysis of gene structures, protein domains and MEME motifs

All necessary *SYF2*-related gene and protein sequence information, as well as intron and exon structure information, was downloaded from Ensembl. Subsequently, the Gene Structure Display Server 2.0 (http://gsds.gao-lab.org/) was used to reconstruct gene structure (Hu et al., 2014). The HMMER website (https://www.ebi.ac.uk/Tools/hmmer/) was used to predict the protein structure domain (Potter et al., 2018). The cDNA and amino acid sequences of all the screened genes were entered into the MEME (https://meme-suite.org/meme/tools/meme), and the 10 most conserved motifs corresponding to the sequences were systematically predicted and analyzed.

Construction of protein interaction networks

The interacting proteins of *Homo sapiens* (ENST00000236273_8), *Mus musculus* (ENSMUST00000030622_2) and *Saccharomyces cerevisiae* (YGR129W) were analyzed through the STRING online database (https://string-db.org/), and the proteins with high interaction rankings were presented through the protein–protein interaction network. Finally, predicted functional partners (confidence cutoff of 0.900) of *SYF2* proteins were presented in the form of an interaction network drawn by Cytoscape 3.8 software.

Analysis and identification of conserved splicing profiles and splice sites

Useful splice isomer sequences for the *SYF2* gene were collected from Ensembl. The cDNA sequence information corresponding to the gene was entered into the MEME to obtain the corresponding motif information for each transcript.

Analysis of *SYF2* expression by online microarray datasets

The required *SYF2* expression data were downloaded through the Expression Atlas (https://www.ebi.ac.uk/gxa/home). The online BAR HeatMapper Plus software (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi) (Chen M. X. et al., 2020) was then used to rearrange the obtained original data as required, before finally presenting it in the form of a heatmap.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization, Y-SC, J-FY, and M-XC; writing original draft preparation, XY, Z-CJ, and B-XH; writing review and editing, JZ, X-RL, C-LC, B-XH, Y-SC, and Z-CJ; funding, J-FY and Y-SC. The final version of the manuscript was agreed by all authors.

Funding

This work was supported by the Program for Science Technology and Innovation Committee of Shenzhen (2021N062-JCYJ20210324115408023), the Natural Science Foundation of Jiangsu Province (SBK2020042924), the Scientific Research Innovation Team of Young Scholars in Colleges and Universities of Shandong Province (2019KJE011), the National Natural Science Foundation of China (NSFC32001932), and the Hong Kong Research Grant Council (AoE/M-05/12, AoE/M-403/16, GRF12100318, 12103219, 12103220).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.873869/full#supplementary-material

References

Ben-Yehuda, S., Dix, I., Russell, C. S., McGarvey, M., Beggs, J. D., and Kupiec, M. (2000). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*. *Genetics* 156, 1503–1517. doi:10.1093/genetics/156.4.1503

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–3175. doi:10.1073/pnas.74.8.3171

Chen, W., and Moore, M. J. (2015). Spliceosomes. *Curr. Biol.* 25, R181–R183. doi:10.1016/j.cub.2014.11.059

Chen, C. H., Chu, P. C., Lee, L., Lien, H. W., Lin, T. L., Fan, C. C., et al. (2012). Disruption of murine mp29/Syf2/Ntc31 gene results in embryonic lethality with aberrant checkpoint response. *PLoS One* 7, e33538. doi:10.1371/journal.pone.0033538

Chen, Z. Y., Liu, H. Y., Jiang, N., and Yuan, J. M. (2019). LncRNA HOST2 enhances gefitinib-resistance in non-small cell lung cancer by down-regulating miRNA-621. *Eur. Rev. Med. Pharmacol. Sci.* 23, 9939–9946. doi:10.26355/eurrev_201911_19560

Chen, M.-X., Zhang, K.-L., Zhang, M., Das, D., Fang, Y.-M., Dai, L., et al. (2020a). Alternative splicing and its regulatory role in woody plants. *Tree Physiol.* 40, 1475–1486. doi:10.1093/treephys/tpaa076

Chen, M. X., Zhang, K. L., Gao, B., Yang, J. F., Tian, Y., Das, D., et al. (2020b). Phylogenetic comparison of 5' splice site determination in central spliceosomal proteins of the U1-70K gene family, in response to developmental cues and stress conditions. *Plant J.* 103, 357–378. doi:10.1111/tpj.14735

Chen, M. X., Mei, L. C., Wang, F., Boyagane Dewayalage, I. K. W., Yang, J. F., Dai, L., et al. (2021). PlantSPEAD: a web resource towards comparatively analysing stress-responsive expression of splicing-related proteins in plant. *Plant Biotechnol. J.* 19, 227–229. doi:10.1111/pbi.13486

Cherif, H., Mannarino, M., Pacis, A. S., Ragoussis, J., Rabau, O., Ouellet, J. A., et al. (2022). Single-cell RNA-seq analysis of cells from degenerating and non-degenerating intervertebral discs from the same individual reveals new biomarkers for intervertebral disc degeneration. *Int. J. Mol. Sci.* 23, 3993. doi:10.3390/ijms23073993

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340

Fan, T., Zhao, Y. Z., Yang, J. F., Liu, Q. L., Tian, Y., Debatosh, D., et al. (2021). Phylogenetic comparison and splice site conservation of eukaryotic U1 snRNP-specific U1-70K gene family. *Sci. Rep.* 11, 12760. doi:10.1038/s41598-021-91693-3

Gascuel, O., Dufayard, J. F., Lefort, V., Anisimova, M., and Hordijk, W. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi:10.1093/sysbio/syq010

Guo, J., Yang, L., Huang, J., Liu, X., Qiu, X., Tao, T., et al. (2014). Knocking down the expression of SYF2 inhibits the proliferation of glioma cells. *Med. Oncol.* 31, 101. doi:10.1007/s12032-014-0101-x

He, Y., Huang, C., Cai, K., Liu, P., Chen, X., Xu, Y., et al. (2021). PRPF19 promotes tongue cancer growth and chemoradiotherapy resistance. *Acta Biochim. Biophys. Sin.* 53, 893–902. doi:10.1093/abbs/gmab059

Hou, S., Li, N., Zhang, Q., Li, H., Wei, X., Hao, T., et al. (2016). XAB2 functions in mitotic cell cycle progression via transcriptional regulation of CENPE. *Cell Death Dis.* 7, e2409. doi:10.1038/cddis.2016.313

Hu, B., Jin, J., Guo, A. Y., He, Z., Ge, G., and Gao, G. (2014). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi:10.1093/bioinformatics/btu817

Huo, Y., Yan, Z. Q., Yuan, P., Qin, M., Kuo, Y., Li, R., et al. (2020). Single-cell DNA methylation sequencing reveals epigenetic alterations in mouse oocytes superovulated with different dosages of gonadotropins. *Clin. Epigenet.* 12, 75. doi:10.1186/s13148-020-00866-w

Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC BioinformaticsBMC Bioinforma.* 1111, 431431. doi:10.1186/1471-2105-11-431

Kaplan, Y., and Kupiec, M. (2007). A role for the yeast cell cycle/splicing factor Cdc40 in the G(1)/S transition. *Curr. Genet.* 51, 123–140. doi:10.1007/s00294-006-0113-y

Katsura, Y., Stanley, C. E., Jr., Kumar, S., and Nei, M. (2017). The reliability and stability of an inferred phylogenetic tree from empirical data. *Mol. Biol. Evol.* 34, 718–723. doi:10.1093/molbev/msw272

Kuraoka, I., Ito, S., Wada, T., Hayashida, M., Tanaka, K., Saijo, M., et al. (2008). Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* 283, 940–950. doi:10.1074/jbc.M706647200

Liu, Y., Ni, T., Xue, Q., Lv, L., Chen, B., Cui, X., et al. (2015). Involvement of p29/SYF2/fSAP29/NTC31 in the progression of NSCLC via modulating cell proliferation. *Pathol. Res. Pract.* 211, 36–42. doi:10.1016/j.prp.2014.07.013

Liu, B., Li, G., Zhang, Z., and Wu, H. (2019). Influence of miR-376c-3p/SYF2 Axis on the progression of gastric cancer. *Technol. Cancer Res. Treat.* 18, 1533033819874808. doi:10.1177/1533033819874808

Lorkovic, Z. J., Lehner, R., Forstner, C., and Barta, A. (2005). Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA* 11, 1095–1107. doi:10.1261/rna.2440305

Madhani, H. D., and Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.* 28, 1–26. doi:10.1146/annurev.ge.28.120194.000245

Morariu, V. I., Srinivasan, B. V., Raykar, V. C., Duraiswami, R., and Davis, L. S. (2008). "Automatic online tuning for fast Gaussian summation," in Conference on Neural Information Processing Systems.

Nakatsu, Y., Asahina, H., Citterio, E., Rademakers, S., Vermeulen, W., Kamiuchi, S., et al. (2000). XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J. Biol. Chem.* 275, 34931–34937. doi:10.1074/jbc.M004936200

Orna, D., and Martin, K. (2002). Mutations in genes of *Saccharomyces cerevisiae* encoding pre-mRNA splicing factors cause cell cycle arrest through activation of the spindle checkpoint. *Nucleic Acids Res.* 30, 4361–4370. doi:10.1093/nar/gkf563

Potter, S. C., Aurélien, L., Eddy, S. R., Youngmi, P., Rodrigo, L., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi:10.1093/nar/gky448

Schwer, B., and Gross, C. H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *Embo J.* 17, 2086–2094. doi:10.1093/emboj/17.7.2086

Shen, C. C., Chen, M. X., Xiao, T., Zhang, C., Zhu, F. Y., Zhang, K. L., et al. (2021). Global proteome response to Pb(II) toxicity in poplar using SWATH-MS-based quantitative proteomics investigation. *Ecotoxicol. Environ. Saf.* 220, 112410. doi:10.1016/j.ecoenv.2021.112410

Shi, F., Cai, F. F., Cai, L., Lin, X. Y., Zhang, W., Wang, Q. Q., et al. (2017). Overexpression of SYF2 promotes cell proliferation and correlates with poor prognosis in human breast cancer. *Oncotarget* 8, 88453–88463. doi:10.18632/oncotarget.18188

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. doi:10.1038/nrm.2017.86

Song, T., Das, D., Ye, N.-H., Wang, G.-Q., Zhu, F.-Y., Chen, M.-X., et al. (2021). Comparative transcriptome analysis of coleorhiza development in japonica and Indica rice. *BMC Plant Biol.* 21, 514. doi:10.1186/s12870-021-03276-z

Tanaka, I., Chakraborty, A., Saulnier, O., Benoit-Pilven, C., Vacher, S., Labiod, D., et al. (2020). ZRANB2 and SYF2-mediated splicing programs converging on ECT2 are involved in breast cancer cell resistance to doxorubicin. *Nucleic Acids Res.* 48, 2676–2693. doi:10.1093/nar/gkz1213

Tao, Y., Zhao, Y., Peng, Y., Ma, X., Sun, C., and Xu, K. (2020). MicroRNA-621 inhibits the growth of gastric cancer cells by targeting SYF2. *Arch. Biochem. Biophys.* 688, 108406. doi:10.1016/j.abb.2020.108406

Tian, Y., Chen, M. X., Yang, J. F., Achala, H. H. K., Gao, B., Hao, G. F., et al. (2019). Genome-wide identification and functional analysis of the splicing component SYF2/NTC31/p29 across different plant species. *Planta* 249, 583–600. doi:10.1007/s00425-018-3026-3

Toro, N., Jiménez-Zurdo, J. I., and García-Rodríguez, F. M. (2007). Bacterial group II introns: Not just splicing. *FEMS Microbiol. Rev.* 31, 342–358. doi:10.1111/j.1574-6976.2007.00068.x

Vijayraghavan, U., Company, M., and Abelson, J. (1989). Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev.* 3, 1206–1216. doi:10.1101/gad.3.8.1206

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* 353, 895–904. doi:10.1126/science.aag2235

Will, C. L., and Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 9, 320–328. doi:10.1016/s0955-0674(97)80003-8

Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707–330. doi:10.1101/cshperspect.a003707

Witzel, I.-I., Koh, Li F., and Perkins, Neil D. (2010). Regulation of cyclin D1 gene expression. *Biochem. Soc. Trans.* 38, 217–222. doi:10.1042/BST0380217

Yan, S., Deng, Y., Qiang, Y., Xi, Q., Liu, R., Yang, S., et al. (2015). SYF2 is upregulated in human epithelial ovarian cancer and promotes cell proliferation. *Tumour Biol.* 36, 4633–4642. doi:10.1007/s13277-015-3111-1

Yin, J., Zhu, J. M., and Shen, X. Z. (2012). New insights into pre-mRNA processing factor 19: A multi-faceted protein in humans. *Biol. Cell* 104, 695–705. doi:10.1111/boc.201200011

Zhang, S., Shi, W., Chen, Y., Xu, Z., Zhu, J., Zhang, T., et al. (2015). Overexpression of SYF2 correlates with enhanced cell growth and poor prognosis in human hepatocellular carcinoma. *Mol. Cell. Biochem.* 410, 1–9. doi:10.1007/s11010-015-2533-9

Zhu, J., Ji, L., Zhang, J., Yang, L., Guan, C., Wang, Y., et al. (2014). Upregulation of SYF2 in esophageal squamous cell carcinoma promotes tumor cell proliferation and predicts poor prognosis. *Tumour Biol.* 35, 10275–10285. doi:10.1007/s13277-014-2305-2

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 09 February 2023
DOI 10.3389/fgene.2023.1089053

OPEN ACCESS

EDITED BY
Nikolay Shirokikh,
Australian National University, Australia

REVIEWED BY
Alexander F. Palazzo,
University of Toronto, Canada
Loredana Le Pera,
National Research Council (CNR), Italy

*CORRESPONDENCE
Benoit Laurent,
benoit.laurent@usherbrooke.ca

SPECIALTY SECTION
This article was submitted to RNA, a section of the journal Frontiers in Genetics

RECEIVED 03 November 2022
ACCEPTED 23 January 2023
PUBLISHED 09 February 2023

CITATION
Manuel JM, Guilloy N, Khatir I, Roucou X and Laurent B (2023), Re-evaluating the impact of alternative RNA splicing on proteomic diversity.
Front. Genet. 14:1089053.
doi: 10.3389/fgene.2023.1089053

Re-evaluating the impact of alternative RNA splicing on proteomic diversity

Jeru Manoj Manuel[1,2], Noé Guilloy[2], Inès Khatir[1,2], Xavier Roucou[2,3,4] and Benoit Laurent[1,2]*

[1]Research Center on Aging, Centre Intégré Universitaire de Santé et Services Sociaux de l'Estrie-Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, QC, Canada, [2]Department of Biochemistry and Functional Genomics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada, [3]Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS), Sherbrooke, QC, Canada, [4]Quebec Network for Research on Protein Function Structure and Engineering, PROTEO, Québec, QC, Canada

Alternative splicing (AS) constitutes a mechanism by which protein-coding genes and long non-coding RNA (lncRNA) genes produce more than a single mature transcript. From plants to humans, AS is a powerful process that increases transcriptome complexity. Importantly, splice variants produced from AS can potentially encode for distinct protein isoforms which can lose or gain specific domains and, hence, differ in their functional properties. Advances in proteomics have shown that the proteome is indeed diverse due to the presence of numerous protein isoforms. For the past decades, with the help of advanced high-throughput technologies, numerous alternatively spliced transcripts have been identified. However, the low detection rate of protein isoforms in proteomic studies raised debatable questions on whether AS contributes to proteomic diversity and on how many AS events are really functional. We propose here to assess and discuss the impact of AS on proteomic complexity in the light of the technological progress, updated genome annotation, and current scientific knowledge.

KEYWORDS

alternative splicing, RNA, isoform proteins, alternative proteins, ghost proteome

Introduction

Alternative splicing (AS) is a key process by which genes produce more than a single mRNA, hence contributing to the transcriptome complexity. In this process, specific exons of a gene can be included or excluded in the final RNA. Protein-coding genes and lncRNA genes can generate multiple splice variants from one gene through AS (Mercer et al., 2011; Khan, Wellinger, and Laurent 2021). From plants to humans, AS is a powerful mechanism that increases transcriptome plasticity and can control the expression level of certain genes (Castle et al., 2008; Gueroussov et al., 2015; Muhammad et al., 2022). Indeed, RNA splice variants arising from AS can exhibit different mRNA stabilities and structures. In humans, it is estimated that 95% of genes undergo AS, which underscores its importance (Castle et al., 2008; Pan et al., 2008; Nilsen and Graveley 2010). Three transcripts are produced in average from each protein-coding gene (Khan, Wellinger, and Laurent 2021). Importantly, splice variants produced from protein-coding genes can potentially encode for distinct protein isoforms. For a given gene, the most expressed transcript is usually defined as coding for the canonical protein. This canonical status is determined based on the transcript expression across different tissues of an organism, the conservation of its exon combination with other species, and/or the existence of a functional role for the protein (Osmanli et al., 2022). Compared to their canonical proteins, isoform

proteins can lose or gain certain domains and, therefore, can differ in their functional properties by the alteration of localization signals, sequences for post-translational modifications, or interaction with other proteins (Kriventseva et al., 2003; Stamm et al., 2005; Leoni et al., 2011; Light and Elofsson 2013). Advances in proteomics have shown that the proteome is indeed diverse due to the presence of numerous protein isoforms. For the past decades, with the help of advanced high-throughput technologies such as RNA sequencing (RNA-seq), a full catalog of alternatively spliced transcripts has been established, but the functional significance of most AS events remains still largely unknown. Hence, the identification of numerous alternatively spliced transcripts raises important and debatable questions: how many AS events are real and not mere artefacts of splicing machinery? How many AS products are functional? Does AS really expand proteomic diversity? We propose here to re-evaluate and discuss the impact of AS on proteomic diversity in the light of the technological progress, updated genome annotation, and current scientific knowledge.

Alternative splicing and proteomic diversity: Two different visions

Whether AS is a major source of proteome complexity has always been a contentious issue in the field. For example, on this debatable question, Benjamin J. Blencowe and Michael L. Tress et al. have mutually expressed their contrasting opinions few years ago (Tress et al., 2017a; Blencowe 2017; Tress et al., 2017b).

Michael L. Tress and colleagues claimed that AS might not be the key to proteome complexity. They argued that most genes only expressed one main transcript across multiple cell lines (Gonzalez-Porta et al., 2013), and hence, one single main protein isoform can be detected by high-resolution mass spectroscopy (Abascal et al., 2015; Ezkurdia et al., 2015). The abundance of alternatively spliced variants identified from more than 100 different tissues at various developmental stages was, therefore, in contrast with the low number of multiple protein isoforms per gene. They found that only 2% of genes had multiple isoform proteins (246 genes with splice event-specific peptide evidence over 12,716 human genes for which at least two peptides have been detected) (Abascal et al., 2015). As few genes provided reliable evidence for more than one isoform, the authors stated that alternative variants were not abundant at the protein level (Tress, Abascal, and Valencia 2017a). One possible reason could be the misidentification of a good peptide spectrum with multiple assigned peaks. However, the discrepancies between transcriptomics and proteomics experiments are difficult to explain solely on a technical issue. They described that alternatively spliced exons were not under selective pressure and are evolving neutrally (Tress, Abascal, and Valencia 2017a). This observation suggested in their opinion that AS events were not evolutionary innovations and that most alternatively spliced variants were not functionally important if translated.

In response to Tress et al. (2017a) and Tress et al. (2017b), Benjamin J Blencowe agreed that AS events were mostly specific to species and, hence, are under relaxed selection pressure (Blencowe 2017). However, he pointed out that even though alternatively spliced transcripts were expressed at lower levels than their corresponding main protein isoforms, it did not mean that these splice variants were not translated or did not have a relevant function in a given cell or tissue type. Blencowe argued that protein abundance was predominantly related to transcript abundance (Liu, Beyer, and Aebersold 2016) and that many splice variants identified by transcriptomics have been detected in polysome fractions and were likely translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). Finally, Blencowe stated that the low detection rate of protein isoforms by LC-MS/MS cannot be interpreted since their identification is limited by the coverage and sensitivity of the technology. Indeed, the peptide number largely exceeds the number of sequencing cycles provided using a mass spectrometer, thereby limiting the detection of splice variants compared to a constitutively expressed sequence (Blencowe 2017).

Different perspectives: Right and wrong at the same time?

These two visions highlight the AS potential role in proteomic diversity on two different ends of the spectrum. The limitations to the available technology and the scientific knowledge at the time the studies were conducted have potentially skewed the interpretation to opposite ends. In this section, we discuss critical points that should be considered to assess the impact of AS on proteomic diversity.

Alternative splicing: Real or artefact of splicing machinery?

The widespread presence of alternatively spliced transcripts has raised the question of whether they are artefacts of splicing machinery or have a biological purpose (Graveley 2001). With the high complexity of eukaryotic genes and the level of splice-site conservation, numerous AS events are expected to happen along the processing of pre-mRNAs, regardless of their functional relevance (Modrek and Lee 2002). However, having a reduced fidelity of the spliceosome to promote proteome diversification could be problematic for a cell since basic molecular mechanisms cannot afford to jeopardize levels of essential proteins (Hsu and Hertel 2009). Consequently, high degrees of specificity and fidelity are required for pre-mRNA splicing to ensure the correct expression of critical functional mRNAs. Indeed, even though the frequency of aberrant spliced transcripts varies widely among loci, tissues, and species, the minimum splicing error rate in vertebrates is around 0.1% aberrant transcripts per intron (Skandalis 2016). The spliceosome is extremely accurate in selecting splice junctions with error frequencies as low as one per 10^5 splicing events (Fox-Walsh and Hertel 2009). This estimation was only performed on specific transcripts (i.e., UBA52, RPL23, HPRT, POLB, and TRPV1), so the extent to which the spliceosome is error-prone remains to be globally assessed. Although the spliceosome is prone to errors, mis-spliced mRNAs can be degraded from cells through nonsense-mediated RNA decay (NMD) or other RNA quality control steps (Saudemont et al., 2017; Garcia-Moreno and Romao 2020). Therefore, the spliceosome is unlikely responsible for generating artefactual splice variants.

An evolutionary perspective of alternative splicing

The importance and functionality of AS events are often associated on whether these events are conserved during evolution. Generally,

95% of human multiexon genes undergo AS (Pan et al., 2008), but this ratio is 60.7% in the fruit fly (*Drosophila melanogaster*) (Graveley et al., 2011), 25% in the nematode (*Caenorhabditis elegans*) (Ramani et al., 2011), and only 2.9% in the green alga (*Volvox carteri*) (Kianianmomeni et al., 2014). Organisms with more complexity tend to have a higher ratio of AS events. There is a strong positive correlation between the number of unique cell types—referred as organism complexity—and the number of AS events (Chen et al., 2014). The study of the evolutionary landscape of AS over ~350 million years of evolution in vertebrates showed significant differences in AS complexity among vertebrate species, with primates harboring the highest complexity (Barbosa-Morais et al., 2012; Merkin et al., 2012). These studies demonstrated that the variation in gene expression was conserved at the tissue-specific level, while AS was conserved at the species-specific level, suggesting that AS diverged faster than gene expression. Moreover, AS event types varied in their frequency among different organisms. In animals, exon skipping is the most common AS event, which represents around 50% of all AS events (Pan et al., 2008), while in plants, intron retention is the most abundant AS event type (Reddy et al., 2013). Most AS events have variable tissue specificities and appear to be evolving neutrally (Wang et al., 2008). However, a subset of AS events is conserved between species and displays tissue specificity. For example, around 20% of alternative exons are conserved between humans and mice (Modrek and Lee 2002; Abascal et al., 2015). These conserved events are significantly enriched in genes that function in common biological processes and pathways. Alternative exons in these splicing "networks" allow the tissue-specific rewiring of protein–protein interaction networks (Buljan et al., 2012; Ellis et al., 2012; Irimia et al., 2014; Tapial et al., 2017). Investigating these networks in different tissues and organs has revealed that these conserved isoforms play a prominent role in the regulation of neuronal development (Boutz et al., 2007; Jiao et al., 2008; Laurent et al., 2015; Fiszbein et al., 2016), immunity (Zikherman and Weiss 2008), and muscle differentiation (Nakka et al., 2018). However, this evolutionary conservation does not mean that alternative exons, which are not evolutionarily conserved, are not significant and do not participate in proteomic diversity. These isoforms could be expressed in a lineage-specific manner, or they might have just recently evolved. For instance, the exonization of intronic sequences such as repetitive elements is now widely documented in many genomes. In primate and human genomes, Alu elements are the most abundant transposable elements that can generate new exons (i.e., Alu exons) and lead to novel spliced transcripts (Krull, Brosius, and Schmitz 2005). Ribosome profiling and proteomics data from human tissues and cell lines showed that some Alu-derived exons can be translated and present in human proteins (Lin et al., 2016), suggesting that some Alu exons can contribute to proteomic diversity. However, in primates and humans, the high number and complexity of AS events might not reflect the functional expansion of the transcriptome but could be explained by the nearly neutral theory (Ohta 1992). Weak selection results in an excess of neutral or slightly deleterious mutations, including those affecting AS regulation. A reduction of intron splicing accuracy, mutations introducing cryptic splicing signals, and transposable element insertion events can generate novel AS events that produce non-functional spliced transcripts (Pickrell et al., 2010). Since these mutations are not removed by purging selection, they can persist and some of them can selectively give novel functional entities, for example, AS events that become functional.

Correlation between transcription and translation

One common argument supporting AS contribution to proteomic complexity is that protein abundance is predominantly related to transcript abundance (Liu, Beyer, and Aebersold 2016). Therefore, even low levels of alternatively spliced transcripts have a chance to be translated into functional proteins. However, there are many regulatory mechanisms that can balance the level of protein expression: the translation rate, the degradation rate, the protein synthesis rate, and transport (Vogel and Marcotte 2012; McManus, Cheng, and Vogel 2015). Different subsets of genes exhibit different types of regulation. At a steady state, mRNA levels correlate with protein levels even during dynamic processes such as proliferation or differentiation (Hsieh et al., 2012; Vogel and Marcotte 2012; Kristensen, Gsponer, and Foster 2013; Li et al., 2014). However, the mRNA levels of some genes are proxies for the corresponding protein levels because of post-transcriptional and translational mechanisms (Liu and Aebersold 2016; Liu, Beyer, and Aebersold 2016). For short-term adaptation such as stress response, the regulation of the transcript level of specific genes is unadapted to the cellular response and post-transcriptional mechanisms (e.g., increase of translation or increase of protein degradation) are thereby more efficient. For instance, changes in the translation rate could positively or negatively affect the mRNA–protein ratio (Lackner et al., 2012; Cheng et al., 2016) and, hence, foster a significant contribution of alternatively spliced transcripts to proteomic diversity.

Another argument supporting AS contribution to proteomic complexity is that many splice variants identified by transcriptomics have been detected in polysome fractions and, hence, are likely to be translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). However, there may be significant levels of alternatively spliced transcripts that do not pass co-translational quality control mechanisms and are degraded. Aberrant polypeptides and mRNAs can be detected and eliminated by mRNA quality control systems while engaging the ribosome (Inada 2017). Because the ribosome has a central role in quality control processes, alternatively spliced transcripts associated with the ribosome are not necessarily translated into proteins.

What is new on proteomic diversity?

Re-evaluating the impact of AS on proteomic diversity necessitates examining the newest developments in this field of investigation, more specifically the technological progress, the update of genome annotation, and the latest advances in scientific knowledge.

Technological and technical advances

As highlighted by Blencowe, LC-MS/MS has some limitations in identifying all potential protein isoforms in a complex sample. The number of peptides exceeds the number of sequencing cycles provided using a mass spectrometer, and hence, the detection of alternative

splice isoforms present in low quantities is limited and could potentially explain why so few alternative isoforms can be detected in proteomics experiments (Blencowe 2017). To address this issue, the integration of RNA-seq with a data-independent acquisition method acquiring all theoretical spectra has been implemented to reduce peptide mapping uncertainty, improve quantitative accuracy, and detect novel peptides (Liu et al., 2017; Jeong, Kim, and Paik 2018; Agosto et al., 2019). This proteogenomic approach yielded high reproducibility between technical and biological replicates and enabled the quantification of a large fraction of the proteome with quantitative accuracy (Poulos et al., 2020). Another limitation to the detection of alternative splice isoforms is also attributed to enzymes used to digest protein samples. The standard protease used in shotgun proteomics is the trypsin that digests at K or R residues, hereby producing short peptides (around six amino acids) and limiting the proteome coverage and detection of isoform proteins (Wang et al., 2018). Other proteases (e.g., chymotrypsin, LysC, LysN, AspN, GluC, and ArgC) have been used to cover complementary fractions of the proteome and improve the detection of specific peptides (Giansanti et al., 2016). A combination of several enzymes could be the best approach to reach comprehensive peptide identification.

Another challenge is to improve the identification of potentially functional transcripts. The development of long-read sequencing technologies has transformed the field since we can now obtain the entire RNA sequence in a single read (Marx 2023). The full-length transcript recovery and quantification helped advance transcript-level analyses of AS processes, distinguish novel isoform changes, and improve the ability to identify functional isoforms (Uapinyoying et al., 2020; De Paoli-Iseppi, Gleeson, and Clark 2021; Hu et al., 2021; Troskie et al., 2021; Wright et al., 2022). For instance, alternative isoforms and tumor-specific isoforms arising from aberrant splicing during liver tumorigenesis were recently identified by single-molecule real-time long-read RNA sequencing (Chen et al., 2019). Another study combined long-read sequencing with polysome profiling and ribosome foot printing data to predict isoform-specific translational status in the rat hippocampus (Wang X et al., 2019). Indeed, single-molecule sequencing also provides the opportunity to improve ribosome profiling quantification by adapting existing methods for translation studies. For example, quantification of the translation of individual transcript isoforms using ribosome-protected mRNA fragments revealed evolutionary conserved impacts of differential splicing on the proteome (Reixachs-Sole et al., 2020). Finally, the single-cell revolution could also help address more accurately the impact of AS on proteomic diversity. Single-cell differential splicing analyses revealed novel differentially expressed splicing junctions (Liu et al., 2021). Single-cell proteomics is now taking the center stage. Novel quantitative single-cell proteomics approaches are capable of consistently quantifying thousands of proteins per cell across thousands of individual cells using limited instrument time and display ultra-high sensitivity to detect changes in a single-cell proteome (Schoof et al., 2021; Brunner et al., 2022). The technology could be applied for detecting specific protein isoforms in a particular cell type and, hence, could give unprecedented insights into the isoform proteome in health and disease. Interestingly, there are now integrated strategies that can profile single-cell proteome and transcriptome in a single reaction, highlighting the promising potential of highly multiplexed single-cell analyses (Genshaft et al., 2016; Specht et al., 2021).

Finally, an additional challenge is that most proteomic data were focused on the identification of proteins derived from alternatively spliced transcripts in steady-state conditions (Blakeley et al., 2010; Ezkurdia et al., 2012; Alfaro et al., 2017). However, most RNA splicing changes have been associated with changes in physiological conditions (e.g., stress response and hypoxia) or between normal and disease states (Ly et al., 2014). Some studies have also addressed the issue of whether targeted perturbations in RNA splicing patterns manifest as changes in the proteomic composition. For example, by depleting a spliceosome component (i.e., PRPF8) and using quantitative proteomics, it was established that significant changes in RNA relative abundance showed consistent changes in protein production (Liu et al., 2017). Using a similar approach, it would be interesting to determine more broadly how changes in AS for a subset of transcripts reflect in differential protein expression and assess the contribution of AS to proteomic complexity.

Genome annotation

Historically, mRNAs were defined as monocistronic and expected to encode a single protein. In addition, open reading frames (ORFs) shorter than 100 codons were automatically discarded from genome annotations as proteins of this length were deemed too short to be functional (Cheng et al., 2011). However, the annotation rules have considerably limited the exploration of the proteome. Based on the potential polycistronic nature of genes, a deeper ORF annotation from an exhaustive transcriptome has predicted all possible alternative ORFs (altORFs), which are defined as potential protein-coding ORFs located either in UTRs of transcripts, in alternative reading frames within the coding sequence of mRNAs, or in non-coding RNAs (Samandi et al., 2017; Brunet et al., 2018; Brunet et al., 2019). Numerous altORFs were identified to be both in-frame and out-of-frame of annotated ORFs. Many annotated altORFs are conserved in eukaryotes, suggesting that alternative proteins encoded from these alternative start codons might have a function across species. The community used ribosome profiling to capture all translation events across the genome and confirmed the translation of many altORFs (Bazzini et al., 2014; Ji et al., 2015; Samandi et al., 2017; Weaver et al., 2019). Combined with large-scale proteomics, these studies have led to the identification and functional relevance of alternative proteins translated from many altORFs located within mature transcripts (Saghatelian and Couso 2015; Na et al., 2018; Rothnagel and Menschaert 2018; Orr et al., 2020). Many functional studies showed that alternative proteins play central functions in the maintenance of cellular homeostasis (Delcourt et al., 2018; Cardon et al., 2020; Vergara et al., 2020; Brunet et al., 2021a; Cao et al., 2021; Ichihara, Nakayama, and Matsumoto 2022). In humans, mutations creating or deleting altORFs have been associated with physiopathological conditions such as amyotrophic lateral sclerosis (ALS) (Brunet et al., 2021b), craniofrontonasal syndrome (Tavares et al., 2019), and thrombocythemia (Wiestner et al., 1998). Interestingly, mutations found in cancers that are silent for reference proteins can impact the expression of alternative proteins resulting from the mutated mRNA, suggesting that alternative proteins could be new biomarkers of pathologies (Child, Miller, and Geballe 1999; Liu et al., 1999; Barbosa, Peixeiro, and Romao 2013; Sendoel et al., 2017; Schulz et al., 2018).

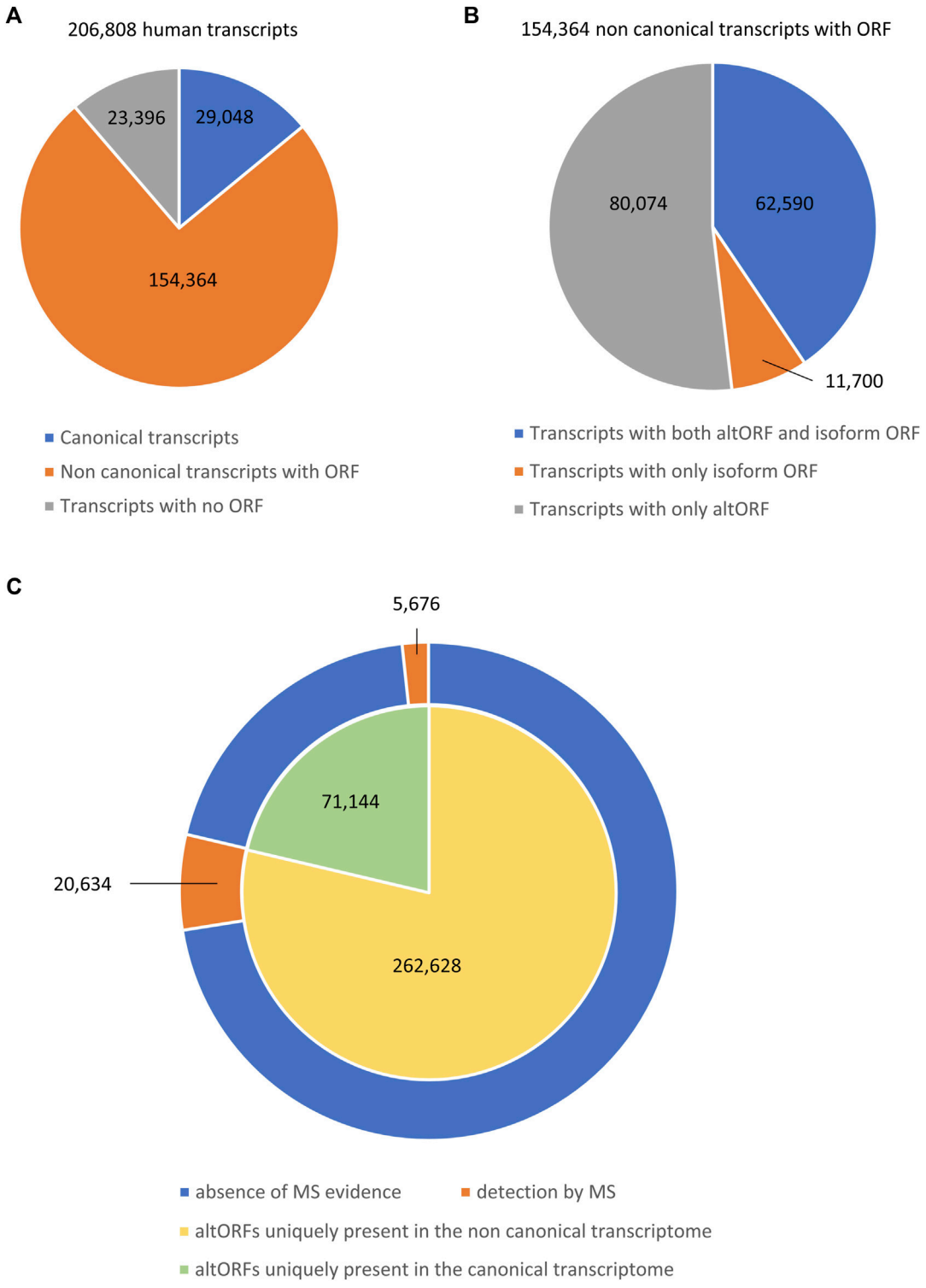


FIGURE 1
Composition of the human transcriptome. **(A)** Pie chart showing the number of different transcripts from the human reference genome (GRCh38 v95). Three types of transcripts are represented: canonical transcripts encoding a reference protein (blue), non-canonical transcripts generated through alternative splicing that contain an ORF (orange), and transcripts that do not have an annotated ORF (gray). **(B)** Pie chart showing the proportion of different sub-types of non-canonical transcripts containing an ORF. Three sub-types of transcripts are represented: non-canonical transcripts with both an alternative ORF (altORF) and an isoform ORF (blue), non-canonical transcripts with only an isoform ORF (orange), and non-canonical transcripts with only an altORF (gray). **(C)** Double pie chart representing the distribution of altORFs uniquely present in the canonical transcriptome (green) or the non-canonical transcriptome (yellow). Using the OpenProt database (Brunet et al., 2019), the evidence obtained by mass spectrometry (MS) of altORF-related proteins is represented in orange in the ring, while the absence of evidence is represented in blue.

A major problem is that alternative proteins expressed from these altORFs are usually not represented in the conventional protein databases (Brunet, Leblanc, and Roucou 2020; Cardon, Fournier, and Salzet 2021). Therefore, these alternative proteins represent a "ghost proteome" that was not considered until recently. Data-driven tools such as the sORF repository (Olexiouk, Van Criekinge, and Menschaert 2018) or the OpenProt database (Brunet et al., 2021a) have now been developed to offer a broader view of proteomes. The existence of thousands of altORFs hidden within known coding sequence of mRNAs raises the question of whether AS could also contribute to proteomic diversity through these small alternative proteins. To address this question, we performed a computational analysis using Ensembl human genome annotation (GRCh38 v95) and the OpenProt database (version 1.6) to determine the impact of AS on this hidden proteome. We identified a total of 206,808 transcripts including 29,048 transcripts defined as canonical as they encode reference proteins (Figure 1A). These transcripts might contain altORFs coding for alternative proteins. We also identified 154,364 transcripts (74.6%) that we categorized as non-canonical since they derive from AS but are not referenced to encode for reference proteins (Figure 1A). However, these transcripts may encode isoforms of reference proteins and/or contain an altORF. Finally, we identified 23,396 transcripts (11.3%) with no ORF according to the OpenProt database (Figure 1A). We next analyzed the non-canonical coding transcriptome landscape. Among these 154,364 transcripts, we identified 62,590 transcripts (40.5%) that contain both an ORF coding for an isoform of a reference protein and an altORF (Figure 1B). We found 80,074 transcripts (51.9%) only containing altORFs and 11,700 transcripts (7.6%) only containing an ORF coding for an isoform of a reference protein (Figure 1B). Our analysis highlights that AS generates numerous transcripts that do not encode for an isoform of a reference protein, supporting the claim by Tress and colleagues that AS might not be the key to proteomic complexity (Tress, Abascal, and Valencia 2017a). However, these transcripts contain altORFs that can potentially code for alternative proteins. These altORFs might also be commonly present in the related canonical transcripts as they could be located in the exons that are not directly affected by AS. We analyzed the distribution of these altORFs and identified 71,144 altORFs that were uniquely present in the canonical transcriptome (29,048 transcripts), while 262,628 altORFs were uniquely present in the non-canonical transcriptome (154, 364 transcripts) (Figure 1C). It represents an average of 2.4 unique altORFs per canonical transcript and 1.7 unique altORFs per non-canonical coding transcript. Using the OpenProt database that encompasses 87 ribosome profiling and 114 mass spectrometry studies from several species, tissues, and cell lines (Brunet et al., 2019), we looked for mass spectrometry evidence for all these altORFs. We found that 5,676 unique altORFs (7.98%) in canonical transcripts had evidence in mass spectrometry, while 20,634 unique altORFs (7.85%) in non-canonical transcripts produced alternative proteins detected by mass spectrometry (Figure 1C). This result clearly indicates that AS can indeed contribute to the human proteomic diversity through the translation of altORFs within mature RNAs.

Contribution of long non-coding RNAs and circular RNAs

Long non-coding RNAs (lncRNAs) represent an important part of the transcriptome (Liu et al., 2005; Derrien et al., 2012). LncRNAs are transcripts of 200 nucleotides or more that should not harbor protein-encoding ORFs (Dinger et al., 2008; Khalil et al., 2009; Derrien et al., 2012). Genome-wide translation profiling has recently revealed that small ORFs identified in lncRNA genes can code for micropeptides, polypeptides with a length of less than 100 amino acids essential for cellular growth (Chen et al., 2020). Other small peptides produced from lncRNAs have also been reported in functional studies (Odermatt et al., 1997; MacLennan and Kranias 2003; Slavoff et al., 2013; Ruiz-Orera et al., 2014; Pang, Mao, and Liu 2018; Wang J et al., 2019; Hartford Corrine and Lal, 2020; Nita et al., 2021; Mise et al., 2022). Eukaryotic lncRNA genes are usually composed of multiple exons with an average of 2.49 exons per human lncRNA gene (Khan, Wellinger, and Laurent 2021). LncRNA transcripts are efficiently spliced with a very similar distribution of AS-type events to that of protein-coding transcripts (Khan, Wellinger, and Laurent 2021). Hence, lncRNAs also generate multiple splice variants whose functional relevance can be associated with RNA-based differential functions (Khan, Wellinger, and Laurent 2021). Although the majority of alternatively spliced lncRNAs are likely non-functional, some of them can produce micropeptides. Indeed, specific splice variants of lncRNAs have the unique capability to produce functional micropeptides that are not encoded by the lncRNA of reference, that is, HOXB-AS3 lncRNA (Huang et al., 2017), LINC00948 lncRNA (Anderson et al., 2015), and LINC00665 lncRNA (Guo et al., 2020). Therefore, the proteomic diversity also depends on AS of lncRNAs. With a total of 354,855 lncRNA genes identified in 17 different species, the exact contribution of lncRNA splice variants to the proteomic complexity remains to be precisely determined and will be a major challenge in the field.

Circular RNAs (circRNAs) are produced from the back-splicing of linear RNAs where upstream splice-acceptor sites are covalently linked to downstream splice-donor sites to form an RNA loop structure (Kristensen et al., 2019). CircRNAs can be conserved during evolution and exhibit a tissue- or cell-specific expression (Kristensen et al., 2019; Santer, Bär, and Thum 2019). CircRNAs are functionally important as they act as microRNA decoys or scaffolds that sequester specific proteins (Chen et al., 2020). Due to their circular shape, circRNAs were not predicted to be translated, but there is growing evidence that circRNAs containing small ORFs can produce micropeptides that have a functional relevance (Legnini et al., 2017; Pamudurti et al., 2017; Liang et al., 2019; Lei et al., 2020; Sinha et al., 2022). It has been hypothesized that AS, particularly exon skipping, drives the formation of circRNAs. However, *in silico* analyses of AS and circRNA production in the human heart revealed that only 10% of circRNAs are produced from alternatively spliced exons, while 90% of circRNAs come from constitutive exons (Aufiero et al., 2018). Therefore, it is possible that AS can also impact the proteomic composition *via* circRNAs containing small ORFs, even though this contribution probably remains limited since circRNAs are described to largely be non-functional products of splicing errors (Xu and Zhang 2021). Future studies on circRNA translation will help uncover the circRNA-driven hidden proteome and enlighten on the functional importance of these novel proteins.

Perspectives

Although MS combined with long-read sequencing and ribosome profiling data has significantly improved the identification of new isoform proteins, many MS fragment spectra still remain unidentified and could potentially result from alternative proteins, micropeptides translated from lncRNAs, circRNAs, or other RNAs (Makarewich and

Olson 2017). Moreover, identifying isoform proteins or small proteins using "bottom–up" MS is challenging. An alternative form of a protein must have a tryptic peptide with more than eight amino acids in the region that differs from the canonical protein to be identified correctly. In addition, this peptide must be suitable for ionization and fragmentation. For small proteins with less than 100 amino acids, the chance to have unique detected peptides is strongly reduced compared to large proteins. Size selection, enrichment of small-size proteins, and careful selection of proteases may improve detection of low abundant proteins and micropeptides. Furthermore, matching MS spectra with custom databases will also help successfully identify novel isoform proteins or small-size micropeptides. "Top–down" proteomics, which characterizes intact proteins in complex mixtures without prior digestion, could be a good alternative approach. However, this method requires long ion accumulation, activation, and detection times and has not been achieved on a large scale due to lack of methods integrated with tandem MS. Despite significant advances, identifying new isoform proteins in the proteome complexity remains a challenge, and further improvements (e.g., methodology, filtering criteria, and database) will be required to substantially improve this situation in the future.

Determining which alternatively spliced transcripts produce proteins with important biological functions (i.e., isoform proteins, alternative proteins, and micropeptides) is the key to confirm the real impact of AS on proteomic complexity. To date, relatively few isoform and alternative proteins have been studied at the functional level, and the biological significance of AS-derived proteome remains obscure. For some AS events, functional consequences can be easily inferred based on changes in the protein sequence. Some alternatively spliced transcripts can encode protein isoforms, which lose or gain specific domains. Interestingly, 50% of AS events in the human transcriptome preserve the ORF and 65% of these frame-preserving splice variants are detected in polysome fractions and, hence, are likely translated (Weatheritt, Sterne-Weiler, and Blencowe 2016). This observation indicates that alternatively spliced transcripts with no frame preservation are potentially eliminated by quality control processes such as NMD. Indeed, some AS events can lead to the inclusion of highly conserved "poison" exons, which contain a premature truncation codon (Leclair et al., 2020). Although these exons do not contribute to the protein-coding capacity, their AS coupled to NMD plays an autoregulatory role in gene expression and protein abundance. Hence, the functional consequences of AS are not always obvious, and many studies failed to detect any differences in the activity of isoform proteins. However, the absence of functional relevance does not mean that there are no functional differences. Therefore, determining the biological function of a single AS event or an AS-derived product will be a major challenge of the proteomic era in the upcoming years.

AS also has a strong clinical relevance since dysregulations of AS have been associated with many chronic diseases including cancer (Ouyang et al., 2021; Zhang et al., 2021). It is, therefore, critical to advance the functional characterization of the AS-derived proteome, but the identification of AS events without regard to their contribution to proteomic diversity is also essential. Indeed, it is key to further study any potential AS alterations in diseases or pathological conditions as they could be valuable prognostic and diagnostic biomarkers. Such investigations could also provide tools for the development of therapeutics. Two splicing-based therapeutic agents are currently tested in clinical trials: small-molecule splicing modulators and antisense oligonucleotides (ASOs). Small-molecule drugs modulate the splicing activity by directly targeting the spliceosome and splicing factors. Surprisingly, these compounds do not induce global splicing inhibition but rather selective changes in AS for genes related to cell proliferation and apoptosis (Folco, Coil, and Reed 2011; Vigevani et al., 2017). However, potential problems of off-target effects require that AS mechanisms are fully understood before further clinical use. In contrast, ASOs are emerging as more secure therapeutic agents to modulate splicing. ASOs can specifically neutralize splice sites, inhibit the recruitment of specific RNA-binding proteins or inhibit the expression of specific splice variants (Rinaldi and Wood 2018). For instance, clinical applicability of ASO-based strategies has been successful in the treatment of patients with spinal muscular atrophy (Hua et al., 2008). ASOs could be used to specifically target specific disease-related splice variants, but advancing knowledge on the functional roles of isoform proteins is, hence, critical for efficient clinical interventions. Regardless of its contribution to proteomic diversity, targeting AS is now recognized an important area for clinical intervention.

Conclusion

On the contentious question "Does alternative splicing really expand proteomic diversity?," we can hereby affirm that AS indeed participates to proteomic complexity in many ways, that is, isoform proteins, alternative proteins, and micropeptides. In the light of this re-evaluation, the AS-related ghost proteome fills a gap and enlarges our vision of the current proteome. Importantly, the remaining limitations on the original question should be taken in consideration in future research endeavors. To continue assessing AS contribution to proteomic complexity, deeper ORF annotation and improvement of technologies and methodologies will be key to functional proteomic discoveries. With a repertoire of alternatively spliced transcripts now significantly expanded, more extensive functional studies on AS and its related proteome are necessary to unravel their unexpected implications in a variety of biological processes.

Author contributions

JMM and IK wrote the manuscript, and NG performed the bioinformatics analysis. XR supervised the bioinformatics analysis and revised the manuscript. BL designed, supervised the experiments, wrote, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by a grant from the Canadian Institutes of Health Research to BL (PJT-166109). JMM was supported by a fellowship from the RNA Innovation NSERC CREATE program. IK was supported by a fellowship from the Faculty of Medicine and Health Sciences at Université de Sherbrooke.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abascal, F., Ezkurdia, I., Rodriguez-Rivas, J., Rodriguez, J. M., del Pozo, A., Vazquez, J., et al. (2015). 'Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11, e1004325. doi:10.1371/journal.pcbi.1004325

Agosto, L. M., Gazzara, M. R., Radens, C. M., Sidoli, S., Baeza, J., Garcia, B. A., et al. (2019). Deep profiling and custom databases improve detection of proteoforms generated by alternative splicing. *Genome Res.* 29, 2046–2055. doi:10.1101/gr.248435.119

Alfaro, J. A., Ignatchenko, A., Ignatchenko, V., Sinha, A., Boutros, P. C., and Kislinger, T. (2017). Detecting protein variants by mass spectrometry: A comprehensive study in cancer cell-lines. *Genome Med.* 9, 62. doi:10.1186/s13073-017-0454-9

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi:10.1016/j.cell.2015.01.009

Aufiero, S., van den Hoogenhof, M. M. G., Reckman, Y. J., Beqqali, A., van der Made, I., Kluin, J., et al. (2018). Cardiac circRNAs arise mainly from constitutive exons rather than alternatively spliced exons. *RNA* 24, 815–827. doi:10.1261/rna.064394.117

Barbosa, C., Peixeiro, I., and Romao, L. (2013). 'Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9, e1003529. doi:10.1371/journal.pgen.1003529

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi:10.1126/science.1230612

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., et al. (2014). 'Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993. doi:10.1002/embj.201488411

Blakeley, P., Siepen, J. A., Lawless, C., and Hubbard, S. J. (2010). Investigating protein isoforms via proteomics: A feasibility study. *Proteomics* 10, 1127–1140. doi:10.1002/pmic.200900445

Blencowe, B. J. (2017). 'The relationship between alternative splicing and proteomic complexity. *Trends Biochem. Sci.* 42, 407–408. doi:10.1016/j.tibs.2017.04.001

Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., et al. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 21, 1636–1652. doi:10.1101/gad.1558107

Brunet, M. A., Brunelle, M., Lucier, J. F., Delcourt, V., Levesque, M., Grenier, F., et al. (2019). 'OpenProt: A more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res.* 47, D403–D10. doi:10.1093/nar/gky936

Brunet, M. A., Jacques, J. F., Nassari, S., Tyzack, G. E., McGoldrick, P., Zinman, L., et al. (2021a). 'The FUS gene is dual-coding with both proteins contributing to FUS-mediated toxicity. *EMBO Rep.* 22, e50640. doi:10.15252/embr.202050640

Brunet, M. A., Leblanc, S., and Roucou, X. (2020). 'Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp. Cell Res.* 393, 112057. doi:10.1016/j.yexcr.2020.112057

Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A., and Roucou, X. (2018). 'Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* 28, 609–624. doi:10.1101/gr.230938.117

Brunet, M. A., Lucier, J. F., Levesque, M., Leblanc, S., Jacques, J. F., Al-Saedi, H. R. H., et al. (2021b). 'OpenProt 2021: Deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res.* 49, D380–D388. doi:10.1093/nar/gkaa1036

Brunner, A. D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., et al. (2022). 'Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* 18, e10798. doi:10.15252/msb.202110798

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G. P., Fuxreiter, M., Bateman, A., et al. (2012). 'Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46, 871–883. doi:10.1016/j.molcel.2012.05.039

Cao, X., Khitun, A., Luo, Y., Na, Z., Phoodokmai, T., Sappakhaw, K., et al. (2021). Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nat. Commun.* 12, 508. doi:10.1038/s41467-020-20841-6

Cardon, T., Fournier, I., and Salzet, M. (2021). 'Shedding light on the ghost proteome. *Trends Biochem. Sci.* 46, 239–250. doi:10.1016/j.tibs.2020.10.003

Cardon, T., Franck, J., Coyaud, E., Laurent, E. M. N., Damato, M., Maffia, M., et al. (2020). 'Alternative proteins are functional regulators in cell reprogramming by PKA activation. *Nucleic Acids Res.* 48, 7864–7882. doi:10.1093/nar/gkaa277

Castle, J. C., Zhang, C., Shah, J. K., Kulkarni, A. V., Kalsotra, A., Cooper, T. A., et al. (2008). 'Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40, 1416–1425. doi:10.1038/ng.264

Chen, H., Gao, F., He, M., Ding, X. F., Wong, A. M., Sze, S. C., et al. (2019). 'Long-Read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. *Hepatology* 70, 1011–1025. doi:10.1002/hep.30500

Chen, J., Zachery Cogan, J., Nuñez, J. K., Fields, A. P., Britt Adamson, D. N., Matthias Mann, L., et al. (2020). 'Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146.

Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. (2014). 'Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* 31, 1402–1413. doi:10.1093/molbev/msu083

Cheng, H., Chan, W. S., Li, Z., Wang, D., Liu, S., and Zhou, Y. (2011). 'Small open reading frames: Current prediction techniques and future prospect. *Curr. Protein Pept. Sci.* 12, 503–507. doi:10.2174/138920311796957667

Cheng, Z., Teo, G., Krueger, S., Rock, T. M., Koh, H. W., Choi, H., et al. (2016). Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.* 12, 855. doi:10.15252/msb.20156423

Child, S. J., Miller, M. K., and Geballe, A. P. (1999). Translational control by an upstream open reading frame in the HER-2/neu transcript. *J. Biol. Chem.* 274, 24335–24341. doi:10.1074/jbc.274.34.24335

De Paoli-Iseppi, R., Gleeson, J., and Clark, M. B. (2021). 'Isoform age - splice isoform profiling using long-read technologies. *Front. Mol. Biosci.* 8, 711733. doi:10.3389/fmolb.2021.711733

Delcourt, V., Brunelle, M., Roy, A. V., Jacques, J. F., Salzet, M., Fournier, I., et al. (2018). The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1. *Mol. Cell Proteomics* 17, 2402–2411. doi:10.1074/mcp.RA118.000593

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi:10.1101/gr.132159.111

Dinger, M. E., Paulo AmaralTim, R. M., Marjan, E., and Askarian-Amiri, P. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation'Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18, 1433–1445. doi:10.1101/gr.078378.108

Ellis, J. D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J. A., et al. (2012). 'Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892. doi:10.1016/j.molcel.2012.05.037

Ezkurdia, I., del Pozo, A., Frankish, A., Rodriguez, J. M., Harrow, J., Ashman, K., et al. (2012). 'Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol. Biol. Evol.* 29, 2265–2283. doi:10.1093/molbev/mss100

Ezkurdia, I., Rodriguez, J. M., Carrillo-de Santa Pau, E., Vazquez, J., Valencia, A., and Tress, M. L. (2015). 'Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.* 14, 1880–1887. doi:10.1021/pr501286b

Fiszbein, A., Giono, L. E., Quaglino, A., Berardino, B. G., Sigaut, L., von Bilderling, C., et al. (2016). Alternative splicing of G9a regulates neuronal differentiation. *Cell Rep.* 14, 2797–2808. doi:10.1016/j.celrep.2016.02.063

Folco, E. G., Coil, K. E., and Reed, R. (2011). 'The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes & Dev.* 25, 440–444. doi:10.1101/gad.2009411

Fox-Walsh, K. L., and Hertel, K. J. (2009). 'Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1766–1771. doi:10.1073/pnas.0813128106

Garcia-Moreno, J. F., and Romao, L. (2020). 'Perspective in alternative splicing coupled to nonsense-mediated mRNA decay. *Int. J. Mol. Sci.* 21, 9424. doi:10.3390/ijms21249424

Genshaft, A. S., Li, S., Gallant, C. J., Darmanis, S., Prakadan, S. M., Ziegler, C. G., et al. (2016). 'Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* 17, 188. doi:10.1186/s13059-016-1045-6

Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. (2016). 'Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* 11, 993–1006. doi:10.1038/nprot.2016.057

Gonzalez-Porta, M., Frankish, A., Rung, J., Harrow, J., and Brazma, A. (2013). 'Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* 14, R70. doi:10.1186/gb-2013-14-7-r70

Graveley, B. R. (2001). Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107. doi:10.1016/s0168-9525(00)02176-4

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473–479. doi:10.1038/nature09715

Gueroussov, S., Gonatopoulos-Pournatzis, T., Irimia, M., Raj, B., Lin, Z. Y., Gingras, A. C., et al. (2015). An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* 349, 868–873. doi:10.1126/science.aaa8381

Guo, B., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., et al. (2020). 'Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 39, e102190. doi:10.15252/embj.2019102190

Hartford Corrina, R., and Lal., A. (2020). 'When long noncoding becomes protein coding. *Mol. Cell. Biol.* 40, e00528-19. doi:10.1128/MCB.00528-19

Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., et al. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 485, 55–61. doi:10.1038/nature10912

Hsu, S. N., and Hertel, K. J. (2009). 'Spliceosomes walk the line: Splicing errors and their impact on cellular function. *RNA Biol.* 6, 526–530. doi:10.4161/rna.6.5.9860

Hu, Y., Fang, L., Chen, X., Zhong, J. F., Li, M., and Wang, K. (2021). 'LIQA: Long-read isoform quantification and analysis. *Genome Biol.* 22, 182. doi:10.1186/s13059-021-02399-8

Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F., and Krainer, A. R. (2008). 'Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet.* 82, 834–848. doi:10.1016/j.ajhg.2008.01.014

Huang, J-Z., Chen, M., Chen, D., Gao, X- C., Zhu, S., Huang, H., et al. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68, 171–184. doi:10.1016/j.molcel.2017.09.015

Ichihara, K., Nakayama, K. I., and Matsumoto, A. (2022). 'Identification of unannotated coding sequences and their physiological functions. *J. Biochem.*, mvac064. doi:10.1093/jb/mvac064

Inada, T. (2017). 'The ribosome as a platform for mRNA and nascent polypeptide quality control. *Trends Biochem. Sci.* 42, 5–15. doi:10.1016/j.tibs.2016.09.005

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi:10.1016/j.cell.2014.11.035

Jeong, S. K., Kim, C. Y., and Paik, Y. K. (2018). 'ASV-ID, a proteogenomic workflow to predict candidate protein isoforms on the basis of transcript evidence. *J. Proteome Res.* 17, 4235–4242. doi:10.1021/acs.jproteome.8b00548

Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). 'Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* 4, e08890. doi:10.7554/eLife.08890

Jiao, Y., Robison, A. J., Bass, M. A., and Colbran, R. J. (2008). 'Developmentally regulated alternative splicing of densin modulates protein-protein interaction and subcellular localization. *J. Neurochem.* 105, 1746–1760. doi:10.1111/j.1471-4159.2008.05280.x

Khalil, Ahmad M., Guttman, Mitchell, Huarte, Maite, Garber, Manuel, Raj, Arjun, Rivea Morales, Dianali, et al. (2009). 'Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11667–11672. doi:10.1073/pnas.0904715106

Khan, M. R., Wellinger, R. J., and Laurent, B. (2021). 'Exploring the alternative splicing of long noncoding RNAs. *Trends Genet.* 37, 695–698. doi:10.1016/j.tig.2021.03.010

Kianianmomeni, A., Ong, C. S., Ratsch, G., and Hallmann, A. (2014). 'Genome-wide analysis of alternative splicing in Volvox carteri. *BMC Genomics* 15, 1117. doi:10.1186/1471-2164-15-1117

Kristensen, A. R., Gsponer, J., and Foster, L. J. (2013). 'Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* 9, 689. doi:10.1038/msb.2013.47

Kristensen, Lasse S., MariaAndersenStagstedHansen, S. Lotte V. W. Karoline K. Ebbesen, Thomas B., Kjems, Jørgen, Hansen, T. B., and Kjems, J. (2019). 'The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691. doi:10.1038/s41576-019-0158-7

Kriventseva, E. V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M. S., et al. (2003). 'Increase of functional diversity by alternative splicing. *Trends Genet.* 19, 124–128. doi:10.1016/S0168-9525(03)00023-4

Krull, M., Brosius, J., and Schmitz, J. (2005). 'Alu-SINE exonization: En route to protein-coding function. *Mol. Biol. Evol.* 22, 1702–1711. doi:10.1093/molbev/msi164

Lackner, D. H., Schmidt, M. W., Wu, S., Wolf, D. A., and Bahler, J. (2012). 'Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biol.* 13, R25. doi:10.1186/gb-2012-13-4-r25

Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., et al. (2015). A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell* 57, 957–970. doi:10.1016/j.molcel.2015.01.010

Leclair, N. K., Brugiolo, M., Urbanski, L., Lawson, S. C., Thakar, K., Yurieva, M., et al. (2020). 'Poison exon splicing regulates a coordinated network of SR protein expression during differentiation and tumorigenesis. *Mol. Cell* 80, 648–665. doi:10.1016/j.molcel.2020.10.019

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). 'Circ-ZNF609 is a circular RNA that can Be translated and functions in myogenesis. *Mol. Cell* 66, 22–37. doi:10.1016/j.molcel.2017.02.017

Lei, M., Zheng, G., Ning, Q., Zheng, J., and Dong, D. (2020). 'Translation and functional roles of circular RNAs in human cancer. *Mol. Cancer* 19, 30. doi:10.1186/s12943-020-1135-7

Leoni, G., Le Pera, L., Ferre, F., Raimondo, D., and Tramontano, A. (2011). 'Coding potential of the products of alternative splicing in human. *Genome Biol.* 12, R9. doi:10.1186/gb-2011-12-1-r9

Li, G. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635. doi:10.1016/j.cell.2014.02.033

Liang, W. C., Wong, C. W., Liang, P. P., Shi, M., Cao, Y., Rao, S. T., et al. (2019). Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol.* 20, 84. doi:10.1186/s13059-019-1685-4

Light, S., and Elofsson, A. (2013). 'The impact of splicing on protein domain architecture. *Curr. Opin. Struct. Biol.* 23, 451–458. doi:10.1016/j.sbi.2013.02.013

Lin, L., Jiang, P., Park, J. W., Wang, J., Lu, Z. X., Lam, M. P., et al. (2016). 'The contribution of Alu exons to the human proteome. *Genome Biol.* 17, 15. doi:10.1186/s13059-016-0876-5

Liu, C., Bai, B., Skogerbø, G., Cai, L., Deng, W., Zhang, Y., et al. (2005). 'NONCODE: An integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* 33, D112–D115. doi:10.1093/nar/gki041

Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., et al. (1999). 'Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.* 21, 128–132. doi:10.1038/5082

Liu, S., Zhou, B., Wu, L., Sun, Y., Chen, J., and Liu, S. (2021). Single-cell differential splicing analysis reveals high heterogeneity of liver tumor-infiltrating T cells. *Sci. Rep.* 11, 5325. doi:10.1038/s41598-021-84693-w

Liu, Y., and Aebersold, R. (2016). 'The interdependence of transcript and protein abundance: New data--new complexities. *Mol. Syst. Biol.* 12, 856. doi:10.15252/msb.20156720

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. *Cell* 165, 535–550. doi:10.1016/j.cell.2016.03.014

Liu, Y., Gonzalez-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., et al. (2017). Impact of alternative splicing on the human proteome. *Cell Rep.* 20, 1229–1241. doi:10.1016/j.celrep.2017.07.025

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., et al. (2014). 'A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* 3, e01630. doi:10.7554/eLife.01630

MacLennan, D. H., and Kranias, E. G. (2003). Phospholamban: A crucial regulator of cardiac contractility. *Nat. Rev. Mol. Cell Biol.* 4, 566–577. doi:10.1038/nrm1151

Makarewich, C. A., and Olson, E. N. (2017). 'Mining for micropeptides. *Trends Cell Biol.* 27, 685–696. doi:10.1016/j.tcb.2017.04.006

Marx, V. (2023). 'Method of the year: Long-read sequencing. *Nat. Methods* 20, 6–11. doi:10.1038/s41592-022-01730-w

McManus, J., Cheng, Z., and Vogel, C. (2015). 'Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation. *Mol. Biosyst.* 11, 2680–2689. doi:10.1039/c5mb00310e

Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., et al. (2011). 'Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30, 99–104. doi:10.1038/nbt.2024

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). 'Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593–1599. doi:10.1126/science.1228186

Mise, S., Matsumoto, A., Shimada, K., Hosaka, T., Takahashi, M., Ichihara, K., et al. (2022). 'Kastor and Polluks polypeptides encoded by a single gene locus cooperatively regulate VDAC and spermatogenesis. *Nat. Commun.* 13, 1071. doi:10.1038/s41467-022-28677-y

Modrek, B., and Lee, C. (2002). 'A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19. doi:10.1038/ng0102-13

Muhammad, S., Xu, X., Zhou, W., and Wu, L. (2022e1758). *Alternative splicing: An efficient regulatory approach towards plant developmental plasticity*. Wiley Interdiscip Rev RNA.

Na, C. H., Barbhuiya, M. A., Kim, M. S., Verbruggen, S., Eacker, S. M., Pletnikova, O., et al. (2018). 'Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res.* 28, 25–36. doi:10.1101/gr.226050.117

Nakka, K., Ghigna, C., Gabellini, D., and Dilworth, F. J. (2018). 'Diversification of the muscle proteome through alternative splicing. *Skelet. Muscle* 8, 8. doi:10.1186/s13395-018-0152-3

Nilsen, T. W., and Graveley, B. R. (2010). 'Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi:10.1038/nature08909

Nita, A., Matsumoto, A., Tang, R., Shiraishi, C., Ichihara, K., Saito, D., et al. (2021). 'A ubiquitin-like protein encoded by the "noncoding" RNA TINCR promotes keratinocyte proliferation and wound healing. *PLoS Genet.* 17, e1009686. doi:10.1371/journal.pgen.1009686

Odermatt, A., Taschner, P. E., Scherer, S. W., Beatty, B., Khanna, V. K., Cornblath, D. R., et al. (1997). 'Characterization of the gene encoding human sarcolipin (SLN), a proteolipid associated with SERCA1: Absence of structural mutations in five patients with brody disease. *Genomics* 45, 541–553. doi:10.1006/geno.1997.4967

Ohta, T. (1992). 'Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. *Genetics* 130, 917–923. doi:10.1093/genetics/130.4.917

Olexiouk, V., Van Criekinge, W., and Menschaert, G. (2018). 'An update on sORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 46, D497–D502. doi:10.1093/nar/gkx1130

Orr, M. W., Mao, Y., Storz, G., and Qian, S. B. (2020). 'Alternative ORFs and small ORFs: Shedding light on the dark proteome. *Nucleic Acids Res.* 48, 1029–1042. doi:10.1093/nar/gkz734

Osmanli, Z., Falgarone, T., Samadova, T., Aldrian, G., Leclercq, J., Shahmuradov, I., et al. (2022). *The Difference in Structural States between Canonical Proteins and Their Isoforms Established by Proteome-Wide Bioinformatics Analysis*, 12.*Biomolecules*

Ouyang, J., Zhang, Y., Xiong, F., Zhang, S., Gong, Z., Yan, Q., et al. (2021). 'The role of alternative splicing in human cancer progression. *Am. J. Cancer Res.* 11, 4642–4667.

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). 'Translation of CircRNAs. *Mol. Cell* 66, 9–21. doi:10.1016/j.molcel.2017.02.021

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi:10.1038/ng.259

Pang, Y., Mao, C., and Liu, S. (2018). 'Encoding activities of non-coding RNAs. *Theranostics* 8, 2496–2507. doi:10.7150/thno.24677

Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* 6, e1001236. doi:10.1371/journal.pgen.1001236

Poulos, R. C., Hains, P. G., Shah, R., Lucas, N., Xavier, D., Manda, S. S., et al. (2020). 'Strategies to enable large-scale proteomics for reproducible research. *Nat. Commun.* 11, 3793. doi:10.1038/s41467-020-17641-3

Ramani, A. K., Calarco, J. A., Pan, Q., Mavandadi, S., Wang, Y., Nelson, A. C., et al. (2011). Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* 21, 342–348. doi:10.1101/gr.114645.110

Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). 'Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi:10.1105/tpc.113.117523

Reixachs-Sole, M., Ruiz-Orera, J., Alba, M. M., and Eyras, E. (2020). 'Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nat. Commun.* 11, 1768. doi:10.1038/s41467-020-15634-w

Rinaldi, C., and Wood, M. J. A. (2018). 'Antisense oligonucleotides: The next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* 14, 9–21. doi:10.1038/nrneurol.2017.148

Rothnagel, J., and Menschaert, G. (2018). Short open reading frames and their encoded peptides. *Proteomics* 18, e1700035.

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014). 'Long non-coding RNAs as a source of new peptides. *Elife* 3, e03523. doi:10.7554/eLife.03523

Saghatelian, A., and Couso, J. P. (2015). 'Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* 11, 909–916. doi:10.1038/nchembio.1964

Samandi, S., Roy, A. V., Delcourt, V., Lucier, J. F., Gagnon, J., Beaudoin, M. C., et al. (2017). 'Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. *Elife* 6.

Santer, L., Bär, C., and Thomas, T. (2019). 'Circular RNAs: A novel class of functional RNA molecules with a therapeutic perspective. *Mol. Ther.* 27, 1350–1363. doi:10.1016/j.ymthe.2019.07.001

Saudemont, B., Popa, A., Parmley, J. L., Rocher, V., Blugeon, C., Necsulea, A., et al. (2017). 'The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome Biol.* 18, 208. doi:10.1186/s13059-017-1344-6

Schoof, E. M., Furtwangler, B., Uresin, N., Rapin, N., Savickas, S., Gentil, C., et al. (2021). 'Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* 12, 3341. doi:10.1038/s41467-021-23667-y

Schulz, J., Mah, N., Neuenschwander, M., Kischka, T., Ratei, R., Schlag, P. M., et al. (2018). 'Loss-of-function uORF mutations in human malignancies. *Sci. Rep.* 8, 2395. doi:10.1038/s41598-018-19201-8

Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., et al. (2017). 'Translation from unconventional 5' start sites drives tumour initiation. *Nature* 541, 494–499. doi:10.1038/nature21036

Sinha, T., Panigrahi, C., Das, D., and Chandra Panda, A. (2022). Circular RNA translation, a path to hidden proteome. *Wiley Interdiscip. Rev. RNA* 13, e1685.

Skandalis, A. (2016). 'Estimation of the minimum mRNA splicing error rate in vertebrates. *Mutat. Res.* 784-785, 34–38. doi:10.1016/j.mrfmmm.2016.01.002

Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., et al. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* 9, 59–64. doi:10.1038/nchembio.1120

Specht, H., Emmott, E., Petelski, A. A., Huffman, R. G., Perlman, D. H., Serra, M., et al. (2021). 'Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* 22, 50. doi:10.1186/s13059-021-02267-5

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., et al. (2005). Function of alternative splicing. *Gene* 344, 1–20. doi:10.1016/j.gene.2004.10.022

Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., et al. (2017). 'An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 27, 1759–1768. doi:10.1101/gr.220962.117

Tavares, R., Kague, E., Musso, C. M., Alegria, T. G. P., Freitas, R. S., Bertola, D. R., et al. (2019). 'Craniofrontonasal syndrome caused by introduction of a novel uATG in the 5'UTR of EFNB1. *Mol. Syndromol.* 10, 40–47. doi:10.1159/000490635

Tress, M. L., Abascal, F., and Valencia, A. (2017a). 'Alternative splicing may not Be the key to proteome complexity. *Trends Biochem. Sci.* 42, 98–110. doi:10.1016/j.tibs.2016.08.008

Tress, M. L., Abascal, F., and Valencia, A. (2017b). 'Most alternative isoforms are not functionally important. *Trends Biochem. Sci.* 42, 408–410. doi:10.1016/j.tibs.2017.04.002

Troskie, R. L., Jafrani, Y., Mercer, T. R., Ewing, A. D., Faulkner, G. J., and Cheetham, S. W. (2021). 'Long-read cDNA sequencing identifies functional pseudogenes in the human transcriptome. *Genome Biol.* 22, 146. doi:10.1186/s13059-021-02369-0

Uapinyoying, P., Goecks, J., Knoblach, S. M., Panchapakesan, K., Bonnemann, C. G., Partridge, T. A., et al. (2020). 'A long-read RNA-seq approach to identify novel transcripts of very large genes. *Genome Res.* 30, 885–897. doi:10.1101/gr.259903.119

Vergara, D., Verri, T., Damato, M., Trerotola, M., Simeone, P., Franck, J., et al. (2020). 'A hidden human proteome signature characterizes the epithelial mesenchymal transition program. *Curr. Pharm. Des.* 26, 372–375. doi:10.2174/1381612826666200129091610

Vigevani, L., Gohr, A., Webb, T., Irimia, M., and Valcarcel, J. (2017). 'Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun.* 8, 2100. doi:10.1038/s41467-017-02007-z

Vogel, C., and Marcotte, E. M. (2012). 'Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi:10.1038/nrg3185

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509

Wang, J., Zhu, S., Meng, N., He, Y., Lu, R., and Yan, G. R. (2019). 'ncRNA-Encoded peptides or proteins and cancer. *Mol. Ther.* 27, 1718–1725. doi:10.1016/j.ymthe.2019.09.001

Wang, X., Codreanu, S. G., Wen, B., Li, K., Chambers, M. C., Liebler, D. C., et al. (2018). Detection of proteome diversity resulted from alternative splicing is limited by trypsin cleavage specificity. *Mol. Cell Proteomics* 17, 422–430. doi:10.1074/mcp.RA117.000155

Wang, X., You, X., Langer, J. D., Hou, J., Rupprecht, F., Vlatkovic, I., et al. (2019). 'Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. *Nat. Commun.* 10, 5009. doi:10.1038/s41467-019-13037-0

Weatheritt, R. J., Sterne-Weiler, T., and Blencowe, B. J. (2016). 'The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123. doi:10.1038/nsmb.3317

Weaver, J., Mohammad, F., Buskirk, A. R., and Storz, G. (2019). *Identifying small proteins by ribosome profiling with stalled initiation complexes*, 10. mBio.

Wiestner, A., Schlemper, R. J., van der Maas, A. P., and Skoda, R. C. (1998). 'An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat. Genet.* 18, 49–52. doi:10.1038/ng0198-49

Wright, D. J., Hall, N. A. L., Irish, N., Man, A. L., Glynn, W., Mould, A., et al. (2022). 'Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. *BMC Genomics* 23, 42. doi:10.1186/s12864-021-08261-2

Xu, C., and Zhang, J. (2021). 'Mammalian circular RNAs result largely from splicing errors. *Cell Rep.* 36, 109439. doi:10.1016/j.celrep.2021.109439

Zhang, Y., Qian, J., Gu, C., and Yang, Y. (2021). 'Alternative splicing and cancer: A systematic review. *Signal Transduct. Target Ther.* 6, 78. doi:10.1038/s41392-021-00486-7

Zikherman, J., and Weiss, A. (2008). Alternative splicing of CD45: The tip of the iceberg. *Immunity* 29, 839–841. doi:10.1016/j.immuni.2008.12.005

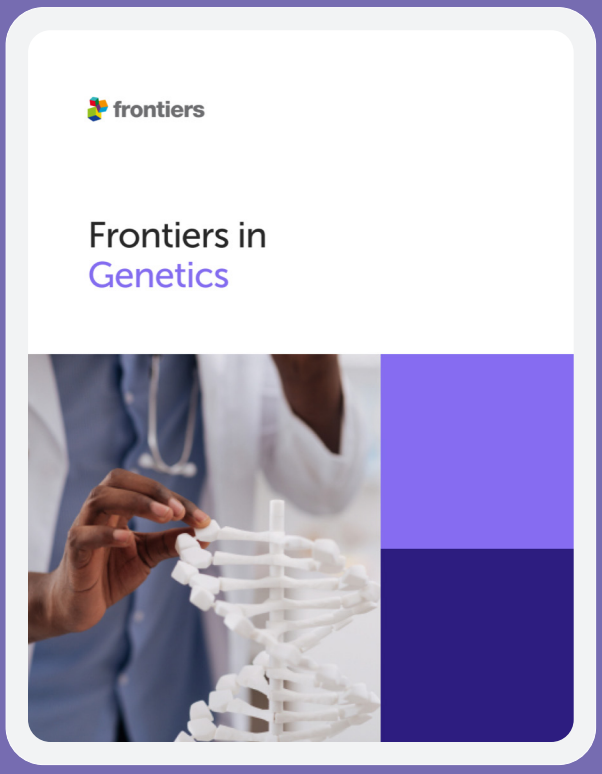
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact