



RNA DISEASES IN HUMANS – FROM FUNDAMENTAL RESEARCH TO THERAPEUTIC APPLICATIONS

EDITED BY: Naoyuki Kataoka, Akila Mayeda and Kinji Ohno

PUBLISHED IN: Frontiers in Molecular Biosciences and Frontiers in Genetics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88963-097-4

DOI 10.3389/978-2-88963-097-4

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: researchtopics@frontiersin.org

RNA DISEASES IN HUMANS – FROM FUNDAMENTAL RESEARCH TO THERAPEUTIC APPLICATIONS

Topic Editors:

Naoyuki Kataoka, The University of Tokyo, Japan

Akila Mayeda, Fujita Health University, Japan

Kinji Ohno, Nagoya University, Japan

This Research Topic addresses the human diseases caused by a malfunction of the RNA metabolism. We aim at strengthening the link between fundamental research and therapeutic applications.

In eukaryotes, RNA is transcribed from genomic DNA. RNA molecules undergo multiple post-transcriptional processes such as splicing, editing, modification, translation, and degradation. A defect, mis-regulation, or malfunction of these processes often results in diseases in humans, referred to as 'RNA diseases'. There is an increasing number of studies focused on RNA diseases, which are aimed at uncovering the fundamental molecular mechanisms at play in order to develop therapeutic approaches.

Citation: Kataoka, N., Mayeda, A., Ohno, K., eds. (2019). RNA Diseases in Humans – From Fundamental Research to Therapeutic Applications. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-097-4

Table of Contents

05 Editorial: RNA Diseases in Humans—From Fundamental Research to Therapeutic Applications

Naoyuki Kataoka, Akila Mayeda and Kinji Ohno

CHAPTER 1

RNA SPLICING

08 The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer

Ettaib El Marabti and Ihab Younis

19 Alternative Splicing in Neurogenesis and Brain Development

Chun-Hao Su, Dhananjaya D and Woan-Yuh Tarn

28 Alternative Splicing Regulator RBM20 and Cardiomyopathy

Takeshi Watanabe, Akinori Kimura and Hidehito Kuroyanagi

39 Importance of Functional Loss of FUS in FTL/ALS

Shinsuke Ishigaki and Gen Sobue

47 Myelodysplastic Syndrome-Associated SRSF2 Mutations Cause Splicing Changes by Altering Binding Motif Sequences

So Masaki, Shun Ikeda, Asuka Hata, Yusuke Shiozawa, Ayana Kon, Seishi Ogawa, Kenji Suzuki, Fumihiko Hakuno, Shin-Ichiro Takahashi and Naoyuki Kataoka

55 HMGA1a Induces Alternative Splicing of the Estrogen Receptor- α Gene by Trapping U1 snRNP to an Upstream Pseudo-5' Splice Site

Kenji Ohe, Shinsuke Miyajima, Tomoko Tanaka, Yuriko Hamaguchi, Yoshihiro Harada, Yuta Horita, Yuki Beppu, Fumiaki Ito, Takafumi Yamasaki, Hiroki Terai, Masayoshi Mori, Yusuke Murata, Makito Tanabe, Ichiro Abe, Kenji Ashida, Kuniyoshi Kobayashi, Munechika Enjoji, Takashi Nomiya, Toshihiko Yanase, Nobuhiro Harada, Toshiaki Utsumi and Akila Mayeda

CHAPTER 2

TRANSLATION

63 Translation of Hepatitis A Virus IRES is Upregulated by a Hepatic Cell-Specific Factor

Akitoshi Sadahiro, Akira Fukao, Mio Kosaka, Yoshinori Funakami, Naoki Takizawa, Osamu Takeuchi, Kent E. Duncan and Toshinobu Fujiwara

73 Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics

Fouzia Yeasmin, Tetsushi Yada and Nobuyoshi Akimitsu

CHAPTER 3

LONG NON-CODING RNA

83 Identification of Minimal p53 Promoter Region Regulated by MALAT1 in Human Lung Adenocarcinoma Cells

Keiko Tano, Rena Onoguchi-Mizutani, Fouzia Yeasmin, Fumiaki Uchiumi, Yutaka Suzuki, Tetsushi Yada and Nobuyoshi Akimitsu

93 *Wnt/ β -catenin Signaling Pathway Regulates Specific lncRNAs That Impact Dermal Fibroblasts and Skin Fibrosis*

Nathaniel K. Mullin, Nikhil V. Mallipeddi, Emily Hamburg-Shields,
Beatriz Ibarra, Ahmad M. Khalil and Radhika P. Atit

106 *Distinct and Modular Organization of Protein Interacting Sites in Long Non-coding RNAs*

Saakshi Jalali, Shrey Gandhi and Vinod Scaria



Editorial: RNA Diseases in Humans—From Fundamental Research to Therapeutic Applications

Naoyuki Kataoka^{1*}, Akila Mayeda² and Kinji Ohno³

¹ Laboratory of Cell Regulation, Departments of Applied Animal Sciences and Applied Biological Chemistry, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo, Japan, ² Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan, ³ Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

Keywords: RNA, RNA disease, splicing, non-coding RNA (ncRNA), translation, virus

Editorial on the Research Topic

OPEN ACCESS

Edited by:

William Cho,
Queen Elizabeth Hospital (QEH),
Hong Kong

Reviewed by:

Venugopal Thayanithy,
Medical School, University of
Minnesota, United States
Mohammadreza Hajjari,
Shahid Chamran University of
Ahvaz, Iran

*Correspondence:

Naoyuki Kataoka
akataoka@mail.ecc.u-tokyo.ac.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 15 May 2019

Accepted: 26 June 2019

Published: 16 July 2019

Citation:

Kataoka N, Mayeda A and Ohno K
(2019) Editorial: RNA Diseases in
Humans—From Fundamental
Research to Therapeutic Applications.
Front. Mol. Biosci. 6:53.
doi: 10.3389/fmolb.2019.00053

RNA Diseases in Humans—From Fundamental Research to Therapeutic Applications

In higher eukaryotes, many different RNAs are encoded in and transcribed from genomic DNAs. Transcribed RNA molecules undergo multiple post-transcriptional processes such as splicing, editing, modification, translation, and degradation. A defect, mis-regulation, or malfunction of these processes often results in diseases in humans, and recently they have been referred to as “RNA diseases” (Scotti and Swanson, 2016; Kataoka, 2017; Ohno et al., 2018). There is an increasing number of studies focused on RNA diseases, which are aimed at unraveling the fundamental molecular mechanisms and developing therapeutic approaches. The goal of this special issue is to introduce RNA diseases to not only RNA scientists but also clinical scientists who are interested in seeking for cures of the diseases.

Pre-mRNA splicing is one of the major processes for post-transcriptional regulatory steps in eukaryotes. Defects in this step result in many diseases, such as neuromuscular diseases and cancers. Splicing takes place in a large multi-RNA-protein complex, called spliceosome. Over one hundred proteins are involved in this step, and many splicing regulatory proteins have been reported. It has been demonstrated that cancer cells undergo cancer-specific splicing patterns. In this special issue on RNA diseases, El Marabti and Younis describe distinct cases of alternative splicing in cancer. They report several categories of splicing aberrations causing alterations in cancer-related genes, mutations in genes encoding splicing factors or core spliceosomal subunits and the disruptions of the balance of RNA-binding proteins. Myelodysplastic syndrome (MDS) is a heterogeneous group of chronic myeloid neoplasms characterized by ineffective hematopoiesis, peripheral blood cytopenia and a high risk of progression to acute myeloid leukemia (AML). Recently it has been shown that mutations on splicing factors are causes of MDS. SRSF2 is one of the major responsible genes for MDS. Masaki et al. found that MDS associated mutations in SRSF2 alter binding affinity to CCWG sequence in RNA and cause aberrant splicing of a subset of genes including EZH2. Estrogen Receptor (ER) α has a critical role in majority of breast cancer. Ohe et al. analyzed the alternative splicing of ER α by using *in vitro* splicing assays and found that HMGA1a induced exon skipping of a shortened exon 1 of ER α . Regulation of ER α alternative splicing by an HMGA1a gives us a novel insight on 5' splice site regulation by U1 snRNP, as well as a possible target in breast cancer therapy.

Pre-mRNA splicing and associated proteins are also involved in tissue development and diseases. Su et al. describe the importance of alternative splicing during neuronal differentiation. Alternative splicing modulates signaling activity, centriolar dynamics, and metabolic pathways, and it also contributes to neurogenesis and brain development. This review shed light on how splicing defects cause brain disorders and diseases. Among those diseases, one of the most famous disease genes is Fused in sarcoma (FUS). FUS is an RNA binding protein that regulates RNA metabolism including alternative splicing, transcription, and RNA transportation. This protein is genetically and pathologically involved in frontotemporal lobar degeneration (FTLD)/amyotrophic lateral sclerosis (ALS). Ishigaki and Sobue summarize the functions of FUS protein in alternative splicing, transcription, mRNA destabilization, axonal transport, and morphological maintenance of neurons. They conclude that a biological link between loss of FUS function, Tau isoform alteration, aberrant post-synaptic function, and phenotypic expression might lead to the sequential cascade culminating in FTLD. RNA splicing factor is also involved in muscle differentiation and cardiomyopathy. RBM20 is a vertebrate-specific RNA-binding protein, and it has initially been identified as one of dilated cardiomyopathy (DCM)-linked genes. RBM20 is a regulator of heart-specific alternative splicing, and one of the major targets for RBM20 is the titin (*TTN*) gene. As titin is the most important factor for passive tension of cardiomyocytes, extensive heart-specific and developmentally regulated alternative splicing of the *TTN* pre-mRNA by RBM20 plays a critical role in passive stiffness and diastolic function of the heart. Manipulation of the *Ttn* pre-mRNA splicing raises RBM20 as a potential therapeutic target (Watanabe et al.).

Translation of mRNA is another critical step for gene expression. Translation of most mRNAs in cells are mediated by the cap structure at the 5' terminus. However, translation of some cellular mRNAs and viral mRNAs are cap-independent. Internal ribosome entry site (IRES) can recruit eukaryotic initiation factor (eIF) complex and ribosome subunit on mRNA in a cap-independent manner. Hepatitis A virus (HAV) utilizes IRES-dependent translation, but unlike most viral IRESs, HAV IRES-mediated translation requires eIF4E and the 3' end of HAV RNA is polyadenylated. Sadahiro et al. analyzed HAV-IRES-mediated translation in a cell-free system derived from either non-hepatic cells or hepatoma cells and revealed that HAV IRES-mediated translation activity in hepatoma cell extracts is higher as compared to extracts derived from a non-hepatic line. Their results strongly suggest that HAV IRES-mediated translation is upregulated by a hepatic cell-specific activator in a poly(A) tail-independent manner.

Recently, it has been well-accepted that long non-coding RNAs (lncRNAs) also play important roles in gene expression in cells. Some of them can be good markers for cancer progression (Yoshimoto et al., 2016). In the manuscript by Tano et al. demonstrated that MALAT1 has a function in repressing the promoter of p53 (TP53) tumor suppressor gene. The p53

targets were upregulated by MALAT1 knockdown in A549 human lung adenocarcinoma cells, which were mediated by transcriptional activation of p53 through MALAT1 depletion. They also identified a minimal MALAT1-responsive region in the P1 promoter of p53 gene. These results suggest that MALAT1 affects the expression of p53 target genes through repressing p53 promoter activity.

On the other hand, the expression of lncRNAs is also regulated by signal cascade. Mullin et al. identified lncRNAs and protein-coding RNAs that are induced by β -catenin activity in mouse dermal fibroblasts. They identified 111 lncRNAs that are differentially expressed in response to activation of Wnt/ β -catenin signaling. Among them, two novel Wnt signaling-Induced Non-Coding RNA (Wincr) transcripts, Wincr1 and Wincr2, were validated. These two lncRNAs are highly expressed in mouse embryonic skin and perinatal dermal fibroblasts, and Wincr1 expression levels in perinatal dermal fibroblasts affects the expression of key markers of fibrosis. These results suggest that β -catenin signaling-responsive lncRNAs modulate dermal fibroblast behavior and collagen accumulation in dermal fibrosis, providing nodes for therapeutic intervention.

The lncRNAs are thought to exhibit their functions as RNA-Protein complex. Jalali et al. utilized the available genome-scale experimental datasets of RNA binding proteins (RBP) to understand the role of lncRNAs in terms of its interactions with RBPs. Their analysis suggests that density of interaction sites for the RBPs was significantly higher for specific sub-classes of lncRNAs when compared to protein-coding transcripts. The significant enrichment of RBP sites across some lncRNA classes strongly suggests that these interactions might be important in the functional roles of lncRNA such as silencing, mRNA processing, and transport.

Interestingly, lncRNAs also serve as templates of micropeptides. In the review article by Yeasmin et al. they summarize the current progress and view of micropeptides encoded in putative short open reading frames (sORFs) within transcripts previously identified as lncRNAs or transcripts of unknown function (TUFs). There are several lines of evidence for significant divergent roles of micropeptides in many fundamental biological processes and even important relationships with pathogenesis.

In conclusion, RNA and RNA binding proteins are involved in most of gene expression steps in higher eukaryotes. The mutations in them cause RNA diseases in human. We sincerely hope the increasing numbers of scientific findings including papers in this special issue will shed light on the disease mechanisms and lead to the development of novel therapeutic strategies for them.

AUTHOR CONTRIBUTIONS

NK, AM, and KO wrote the paper. NK took the primary responsibility for the final content. NK, AM, and KO read and approved the final manuscript.

REFERENCES

- Kataoka, N. (2017). Modulation of aberrant splicing in human RNA diseases by chemical compounds. *Hum. Genet.* 136, 1237–1245. doi: 10.1007/s00439-017-1789-4
- Ohno, K., Takeda, J. I., and Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip. Rev. RNA.* 9:1. doi: 10.1002/wrna.1451
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3
- Yoshimoto, R., Mayeda, A., Yoshida, M., and Nakagawa, S. (2016). MALAT1 long non-coding RNA in cancer. *Biochim. Biophys. Acta* 1859, 192–199. doi: 10.1016/j.bbagr.2015.09.012

Conflict of Interest Statement: 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.

Copyright © 2019 Kataoka, Mayeda and Ohno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer

Ettaib El Marabti and Ihab Younis*

Biological Sciences Program, Carnegie Mellon University in Qatar, Doha, Qatar

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Daisuke Kaida,
University of Toyama, Japan
Rahul N. Kanadia,
University of Connecticut,
United States
Claudia Ghigna,
Istituto di Genetica Molecolare (IGM),
Italy

*Correspondence:

Ihab Younis
iyounis@andrew.cmu.edu

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 30 April 2018

Accepted: 09 August 2018

Published: 07 September 2018

Citation:

El Marabti E and Younis I (2018) The
Cancer Spliceome: Reprogramming of
Alternative Splicing in Cancer.
Front. Mol. Biosci. 5:80.
doi: 10.3389/fmolb.2018.00080

Alternative splicing allows for the expression of multiple RNA and protein isoforms from one gene, making it a major contributor to transcriptome and proteome diversification in eukaryotes. Advances in next generation sequencing technologies and genome-wide analyses have recently underscored the fact that the vast majority of multi-exon genes under normal physiology engage in alternative splicing in tissue-specific and developmental-specific manner. On the other hand, cancer cells exhibit remarkable transcriptome alterations partly by adopting cancer-specific splicing isoforms. These isoforms and their encoded proteins are not insignificant byproducts of the abnormal physiology of cancer cells, but either drivers of cancer progression or small but significant contributors to specific cancer hallmarks. Thus, it is paramount that the pathways that regulate alternative splicing in cancer, including the splicing factors that bind to pre-mRNAs and modulate spliceosome recruitment. In this review, we present a few distinct cases of alternative splicing in cancer, with an emphasis on their regulation as well as their contribution to cancer cell phenotype. Several categories of splicing aberrations are highlighted, including alterations in cancer-related genes that directly affect their pre-mRNA splicing, mutations in genes encoding splicing factors or core spliceosomal subunits, and the seemingly mutation-free disruptions in the balance of the expression of RNA-binding proteins, including components of both the major (U2-dependent) and minor (U12-dependent) spliceosomes. Given that the latter two classes cause global alterations in splicing that affect a wide range of genes, it remains a challenge to identify the ones that contribute to cancer progression. These challenges necessitate a systematic approach to decipher these aberrations and their impact on cancer. Ultimately, a sufficient understanding of splicing deregulation in cancer is predicted to pave the way for novel and innovative RNA-based therapies.

Keywords: splicing, cancer spliceome, alternative splicing, exons, introns

SPlicing IS AN EVOLUTIONARY CONSERVED AND ESSENTIAL STEP IN GENE EXPRESSION IN EUKARYOTES

Most genes in eukaryotes contain intervening sequences (introns) that disrupt the expressed sequences (exons). Introns in eukaryotes are much longer (median size ~1,000 bp but can be >100,000 bp) compared to exons (median size ~120 bp), making introns the major contributors to the sequence of genes. After transcription, in order for the expressed transcript

(pre-mRNA) to become a suitable message for downstream processes such as translation of the encoded protein, the pre-mRNA from any multi-exon gene has to undergo extensive processing to remove the introns by an extraordinary molecular machine, the spliceosome. Given their large size, introns add a long time, sometimes hours, to the transcription process of genes in eukaryotes. Thus, introns present a conundrum as their transcription, just to be spliced out and degraded, seems to be a wasteful process both in terms of the time it takes to transcribe them and the energy consumed in their transcription as well as their removal and degradation. In addition, the splicing process needs to be extremely efficient as well as be executed with high fidelity. Efficiency is required to make sure all introns are removed from the pre-mRNA on time and in a coordinated manner. Fidelity is of paramount importance because joining exons with any mistake of even one base could have catastrophic effects on the reading frame. Furthermore, the cis-sequences or splice sites at the boundaries of each intron are too simple, sometimes degenerate and highly redundant outside the actual splice sites to serve alone as efficient landmarks for spliceosome assembly. Taken together, these features of introns and splicing in general make the presence of introns in eukaryotes counterintuitive. However, introns are not simply extra sequences that are removed by splicing, but rather have several advantages such as coupling multiple RNA processing events for higher gene expression efficiency as well as regulation and providing a checkpoint for quality control of the mRNA. They also allow any gene that harbor them to have a tremendous capacity for diversification through the process of alternative splicing. Thus, it is likely that the advantages of harboring introns outweigh the disadvantages as their presence in eukaryotic genomes and to some extent their position in the genes are highly conserved (Fedorov et al., 2002; Rogozin et al., 2003), in some cases between humans and the plant *Arabidopsis thaliana*.

DIVERSIFICATION OF TRANSCRIPTOMES BY ALTERNATIVE SPLICING

The advent of high-throughput sequencing has uncovered the fact that most multi-exon genes in eukaryotes undergo at least one event of alternative splicing (Pan et al., 2008), generating two or more distinct mRNAs from the same gene, with the number of alternatively spliced transcripts potentially staggering for some genes. Interestingly, many such transcripts are expressed in a tissue-specific manner, at specific developmental stage, or in a disease-specific manner (Castle et al., 2008; Wang et al., 2008). While the function of some of these alternative transcripts is not always immediately interpretable or even recognized, a plethora of work indicates that alternatively spliced exons are translated and tend to encode important domains in the encoded polypeptide (Kalsotra and Cooper, 2011; Ellis et al., 2012; Weatheritt et al., 2016; Tapial et al., 2017). This suggests an evolutionary conserved molecular design for transcriptome diversification without the need to expand the genome that would require creating genes that are homologous to existing ones that serve similar yet distinct functions (Nilsen and Graveley, 2010).

Alternative splicing is a term used to collectively refer to several splicing events. As shown in **Figure 1**, there are various distinct forms of alternative splicing, including alternative exons (cassette exons: skipped/included whole exons), retained introns, and alternative 5' and 3' splice sites (5' ss and 3' ss). There are also several less obvious alternative splicing events that are tightly coupled to and could be a consequence of transcription regulation such as alternative first and last exons. Nevertheless, all these events are well documented in eukaryotes with remarkable impacts on transcriptome diversification. One class of alternative splicing, intron retention, is often overlooked because it is interpreted as a splicing mistake that lead to an intron not being spliced out. While this might be true in several cases, a lot of evidence points to intron retention being regulated to control the expression of genes post-transcriptionally. In fact, cancer cells of all types are characterized by high levels of retained intron, leading to a higher diversity of their transcriptomes compared to normal cells (Dvinge and Bradley, 2015).

Intron retention and its regulation are obvious in a class of introns referred to as minor or U12 introns, which are conserved in almost all eukaryotes. Unlike the vast majority of introns in cells, which rely on the canonical spliceosome that is composed of U1, U2, U4, U5, and U6 snRNPs for their splicing, minor introns utilize a less abundant and seemingly less efficient spliceosome that is made of U11, U12, U4atac, U5, and U6atac snRNPs. Around 800 minor introns in the human genome are embedded in genes that function in signal transduction and information relay, cell cycle control and DNA damage repair (Turunen et al., 2013). We previously showed that hundreds of U12 introns are extremely conserved and are used as molecular switches that provide rapid control of gene expression that does not depend on transcription of new pre-mRNA especially when the gene product is needed instantaneously such as when cells are under stress (Younis et al., 2013). Given the functions of the genes that host minor introns, it is likely that they are regulated in a similar fashion in cancer.

MECHANISMS OF SPLICING REGULATION

It is of note that some documented alternative splicing events constitute only a small fraction of the processed mRNA that are expressed at any given time. While this suggests that such alternative splicing events represent the expected biological noise of a process that is extremely active in cells, we argue that these events are tightly regulated and serve significant roles in various cell types and tissues. More specifically, the low abundance of these events in one cell type could have evolved to be so because the encoded protein from these specific splicing isoforms have a cell type- or condition-specific function. In addition, some of these events are only expressed at a high level when cells are faced with certain environmental conditions, such as stress, in when the specific splicing isoform becomes absolutely required (Younis et al., 2013). Thus, exhaustive searches are now needed to identify these conditions in which these isoforms become abundant and their function more significant. Finally, some disease tissues show enrichment of these events, suggesting specific functions.

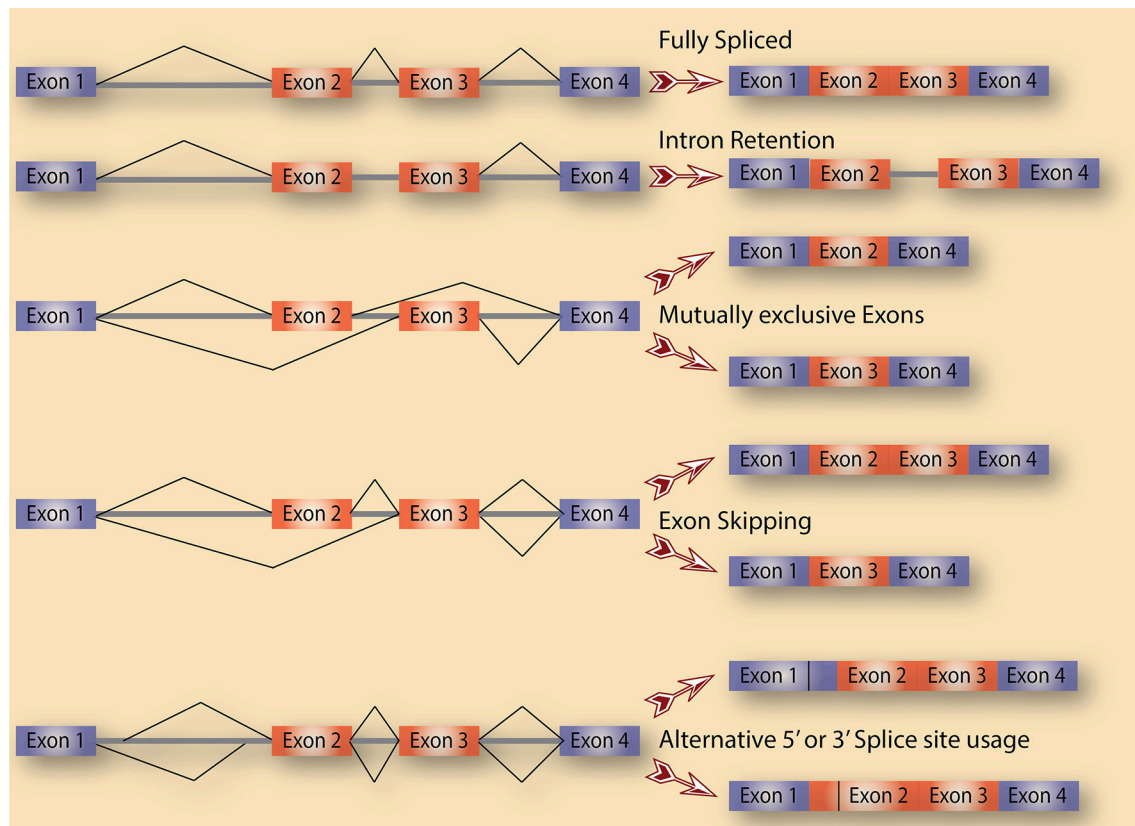


FIGURE 1 | Alternative splicing as a source of diversification. A schematic diagram that depicts the various types of alternative splicing events that potentially exist in cells.

This infers that the abundance of specific alternatively spliced transcripts as well as the choice of the specific alternative splicing events for a given pre-mRNA are under tight regulation. This regulation is dictated by both *cis*-elements in the pre-mRNA itself and *trans*-factors such as RNA binding proteins (RBPs). The fact that the human genome encodes for thousands of RBPs, of which a big fraction functions in RNA splicing and its regulation strongly supports the notion that alternative splicing is not random but rather a highly regulated process and a key step in gene expression regulation.

Splicing factors historically have been classified into hnRNPs, which typically suppress splicing, and SR proteins, which tend to have a positive role in splicing regulation. However, a more thorough analysis of the function of any given hnRNP or SR protein quickly reveals that they do not always conform to these classifications. The ultimate role of an RBP in splicing regulation depends on multiple factors. These include the strength and context of its binding sites on the pre-mRNA in addition to either competitive or cooperative binding of multiple RBPs on or around the regulated exon or intron. This combinatorial regulation makes it very hard to predict the splicing outcome of reduced or increased binding of a single splicing factor in normal or diseased cells. Another complication is that a splicing factor is likely to regulated pre-mRNA splicing of other splicing

factors in endless feedback loops and complex networks. A better approach to understand the regulation of alternative splicing in a given condition requires a systems biology approach in which the expression status and targets of multiple if not all RBPs be assessed to start building these networks of co-regulated pathways.

DEREGULATION OF ALTERNATIVE SPLICING IN CANCER

All data to date indicate that alternative splicing is a well-designed process that is tightly regulated in order to produce a network of alternatively spliced transcripts, which we refer to in this review as the splice-ome (spliceome). Work in the last two decades have moreover showed that the spliceome is significantly altered in disease states, such as cancer (Reviewed in, David and Manley, 2010; Chabot and Shkreta, 2016; Scotti and Swanson, 2016). In fact, every hallmark of cancer can be represented by several examples of proto-oncogenes, tumor suppressor genes, or other genes whose splicing is altered to produce isoforms that are needed for the transformation process (Oltean and Bates, 2014). In this review, we do not aim to provide a comprehensive list of all the cancer-related abnormal alternative splicing events, but rather highlight a few

TABLE 1 | Selected examples of genes with cancer-related alternatively spliced isoforms.

Gene name protein product	Alterations in cancer-related functions due to alternative splicing	Cancer-related alternative splicing	References
<i>TP53</i> /p53	Reduced anti-proliferative response to stress. Reduced tumor suppression	<ul style="list-style-type: none"> Alternative 1st exon Intron 9 retention Intron 9 cryptic exon inclusion Intron 2 retention Cassette Exon 8 Intron 7 alternative 5' ss Intron 2 alternative 5' ss 	Surget et al., 2013
<i>Bcl2l1</i> /BclX	Apoptosis regulator: Anti-apoptotic (BCL-X _L) Pro-apoptotic (BCL-X _S)	<ul style="list-style-type: none"> Cassette Exon 9 	Paronetto et al., 2016
<i>CASP2</i> /Caspase 2	Apoptosis regulator: Anti-apoptotic (Casp2S) Pro-apoptotic (Casp2L)	<ul style="list-style-type: none"> Cassette Exon 2 	Jang et al., 2014
<i>AIMP2</i> /Aimp2	Apoptosis regulation: Anti-apoptotic (AIMP2-DX2) Pro-apoptotic (AIMP2-full length)	<ul style="list-style-type: none"> Cassette Exon 12a Cassette Exon 13 	Anczuków et al., 2012
<i>BIN1</i> / Myc box-dependent-interacting protein 1 (Bin1)	Reduced tumor suppression and apoptosis regulation	<ul style="list-style-type: none"> Intron 5 alternative 3' ss (α deletion) Cassette Exons 7-8 (β deletion) Cassette Exon 11 (γ deletion) Combinations of α, β, and γ deletions 	Liu et al., 2017
<i>TERT</i> /telomerase reverse transcriptase	Reduced replicative senescence Loss of telomerase activity in α , β , γ deletion isoforms	<ul style="list-style-type: none"> Cassette Exon 18 Cassette Exon v6 Cassette Exons v8-10 Cassette Exons v4-5 Cassette Exons v4-7 	Prochazka et al., 2014
<i>CD44</i> /CD44 antigen	Imbalanced regulation of cell division, migration and adhesion	<ul style="list-style-type: none"> Cassette Exon 2 Cassette Exon 3 Cassette Exon 6 Combination of Exons 2,3, and 6 skipping 	Li et al., 2011
<i>IRF3</i> /Interferon regulatory factor 3	Reduced cell growth inhibition Reduced cellular senescence through p53 activation	<ul style="list-style-type: none"> Cassette Exon 3b 	Singh et al., 2004; Radisky et al., 2005
<i>RAC1</i> /Ras-related C3 botulinum toxin substrate 1	Reduced cell growth and cell cycle regulation Reduced cell-cell adhesion formation and contact inhibition	<ul style="list-style-type: none"> Intron 22 alternative 3' ss 	Caldenhoven et al., 1996; Zammarchi et al., 2011
<i>STAT3</i> /Signal transducer and activator of transcription 3	Dominant negative regulation of transcription (STAT3 β)	<ul style="list-style-type: none"> Intron 13 alternative (incomplete) splicing 	Feltes et al., 2002
<i>CDH11</i> /Cadherin11	Enhances invasion when the splice isoform (with unique intracellular domain) is expressed with WT Cadherin 11	<ul style="list-style-type: none"> Mutually exclusive exon 8 (IIIb) or 9 (IIIc) 	Wagner et al., 2003; Zhao et al., 2013; Abou-Fayçal et al., 2017
<i>FGFR2</i> /Fibroblast growth factor receptor 2	Promotes EMT and metastasis	<ul style="list-style-type: none"> Intron 2 alternative 5' ss Cassette Exon 3 	DiFeo et al., 2009
<i>KLF6</i> /Krüppel-like factor 6	Antagonism of tumor suppressor activity	<ul style="list-style-type: none"> Exon 11 skipping with Exon 5 and 6 Inclusion Exon 5 and 6 skipping with Exon 11 Inclusion Exon 5, 6, and 11 skipping Exon 6 skipping and Exon 5 inclusion Partial Exon 5 and 6 splicing 	Zhou et al., 2003; Eckerich et al., 2009; Chakedis et al., 2016
<i>MST1R</i> /Macrophage-stimulating protein receptor or RON (Recepteur d'origine Nantais)	Increased cell motility and invasion		

(Continued)

TABLE 1 | Continued

Gene name protein product	Alterations in cancer-related functions due to alternative splicing	Cancer-related alternative splicing	References
<i>VEGFA</i> /Vascular endothelial growth factor A	Enhanced pro-angiogenic function	<ul style="list-style-type: none"> ○ Cassette Exons 6 and 6b ○ Cassette Exon 7 and 7b ○ Alternative Intron 6 5' ss ○ Alternative Intron 6 3' ss ○ Cassette Exon 8 	Pritchard-Jones et al., 2007; Harper and Bates, 2008
<i>CASP8</i> /Caspase-8	Reduced tumor suppression and pro-apoptotic activity	<ul style="list-style-type: none"> ○ Cassette Exon 8 ○ Alternative Exon 8 and 9 splicing (136 bp insertion between exon 8 and 9) 	Mohr et al., 2005; Olsson and Zhivotovsky, 2011
<i>FAS</i> /Tumor necrosis factor receptor superfamily member 6	Loss of Pro-apoptotic activity	<ul style="list-style-type: none"> ○ Cassette Exon 6 ○ Cassette Exon 8 ○ Intron 5 retention 	van Doorn et al., 2002; Schwerk and Schulze-Osthoff, 2005; Tejedor et al., 2015
<i>MCL1</i> /Induced myeloid leukemia cell differentiation protein (Mcl-1)	Anti-apoptotic	<ul style="list-style-type: none"> ○ Cassette Exon 2 	Shieh et al., 2009
<i>MDM2</i> /E3 ubiquitin-protein ligase (Mdm2)	Reduced p53 binding Enhanced tumor progression.	<ul style="list-style-type: none"> ○ Skipping of Exons 4-9 ○ Skipping of Exons 4-11 ○ Skipping of Exons 5-9 ○ >40 isoforms 	Bartel et al., 2002
<i>GLS</i> /Glutaminase	Deregulation of glutamate metabolism	<ul style="list-style-type: none"> ○ Intron 15 retention leading to cleavage and polyadenylation in Intron 15 and an isoform lacking the canonical last 4 exons. 	van den Heuvel et al., 2012
<i>LDHC</i> /L-lactate dehydrogenase C chain	<ul style="list-style-type: none"> ○ Metabolic rescue in tumor cells ○ Deregulation of cellular energetics 	<ul style="list-style-type: none"> ○ Cassette Exon 3 ○ Skipping of Exon 3-4 ○ Skipping of exons 3, 6, and 7 ○ Cassette Exon7 	Koslowski et al., 2002
<i>MAX</i> / Protein max	Promotion of cell proliferation through enhanced glycolytic metabolism	<ul style="list-style-type: none"> ○ Cassette Exon 2 ○ Cassette Exon 5 	Babic et al., 2013
<i>PKM</i> / Pyruvate kinase PKM	<ul style="list-style-type: none"> ○ Tumor specific metabolism via ○ PKM2 (fetal isoform with Exon10 inclusion) 	<ul style="list-style-type: none"> ○ Mutually exclusive Exons 9 and 10 	Chen et al., 2010; Zhang and Manley, 2013
<i>CCND1</i> / G1/S-specific cyclin-D1	Promotes cell proliferation	<ul style="list-style-type: none"> ○ Intron 4 retention leading to cleavage and polyadenylation in Intron 4 and an isoform lacking the canonical last exon 	Paronetto et al., 2010
<i>EGFR</i> / Epidermal growth factor receptor	Oncogene	<ul style="list-style-type: none"> ○ ALternativer splicing of a combination of Exons9a, 10,16, or 17 ○ Skipping of Exon 2-7 ○ Skipping of Exons 2-22 ○ ... 	Abou-Fayçal et al., 2017
<i>PTEN</i> / Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase (Pten)	Reduced tumor suppression	<ul style="list-style-type: none"> ○ Intron 3 retention ○ Intron 5 retention ○ Inclusion of partial Intron 5 ○ Inclusion of partial Intron 3 ○ Inclusion of 5' end of intron H between Exons 8 and 9 ○ Alternative splicing of the end of intron 5 to splicing acceptor site within intron E 	Agrawal and Eng, 2006; Okumura et al., 2011

that exemplify a deregulated splicing program in cancer that is not a byproduct of the cancer phenotype but a driving force in cancer development and maintenance (Table 1). We thus discuss four main categories of splicing aberrations: (1) Cancer-specific splicing alterations in oncogenes and tumor suppressor genes. (2) Cancer-specific mutations in splicing factors. (3)

Changes in upstream signaling pathways that deregulate splicing factors. And (4) aberrations in spliceosomal components that are linked to cancer. The involvement of alternative splicing in the 10 hallmarks of cancer has been reviewed elsewhere (Sveen et al., 2016). Here we summarize some of these changes to point to the fact that the cancer phenotype in several cancer

types is heavily reliant on altering one or several splicing choices.

CANCER-SPECIFIC SPLICING ALTERATIONS IN ONCOGENES AND TUMOR SUPPRESSOR GENES

Some of the earliest and most studied examples of alternative splicing events that lead to isoforms amiable for cancer are in genes involved in apoptosis such as the members of the Bcl2 family and several caspases. For example, intron 2 of the *Bcl2L1* gene, which encodes the Bcl-X protein, is alternatively spliced. More specifically, the spliceosome has a choice between two 5' splice sites (5' ss) for intron 2. Depending on which 5' ss is chosen, the mRNA produced could be large (Bcl-XL), which encodes a Bcl-X protein with anti-apoptotic function, or small (Bcl-XS), encoding a Bcl-X protein that is missing an essential BH domain and is pro-apoptotic. Another example is Caspase 2 pre-mRNA splicing, whereas the spliceosome faces a choice of including exon 9, generating a caspase 2L mRNA or skipping exon 9, leading to the caspase 2S isoform. The large isoform encodes the pro-apoptotic Casp2L protein, whereas the anti-apoptotic Casp2S protein is encoded by the short isoform. Given that cancer cells are resistant to cell death by apoptosis, they need to ensure the production of Bcl-XS and/or Casp2L. In the absence of mutations in *Bcl2L1* and *Caspase 2* genes that would affect splice sites or other cis-elements leading to Bcl-XS and Casp2L production, cancer cells reprogram the splicing machinery and/or splicing factors that bind to these pre-mRNAs to ensure that the cancer-specific isoforms are enriched.

Several tumor suppressor genes undergo alternative splicing in cancer that leads to either complete or partial loss of function. For example, complex alternative splicing of TP53, which encodes the p53 protein, generates several isoforms with significant impact on the protein function (Surget et al., 2013). Once activated, by DNA damage for example, p53 can induce cell-cycle arrest in either the G1 or G2 phase of the cell cycle. p53 can also activate Growth Arrest and DNA Damage 45 (GADD45), which regulates cell-cycle arrest in the G2/M phases. Thus, the presence of a functional p53 is essential for the multiple cell cycle checkpoints that allow cells to repair DNA damage or commit to apoptosis. Some of the protein products from the TP53 splicing isoforms are dominant negative, and since p53 acts as a tetramer, the production of these dominant negative subunits, even at low level, can have dramatic effects as they act as poison subunits. Four isoforms of these p53 transcripts are depicted in **Figure 2**.

Interestingly, even tumor viruses take advantage of alternative splicing to produce oncoproteins that cause host cell transformation. For example, the production of the two Human Papilloma Virus (HPV) oncoproteins E6 and E7 in patient tissues, which are encoded by one pre-mRNA, depends on alternative splicing. Briefly, unspliced transcripts (that is, the intron is retained) produce the E6 mRNA (and ORF) whereas complete splicing of the pre-mRNA produces the E7 mRNA and

ORF. Other transcripts including E6^ΔE7 or E6*III are generated due to alternative 3'-splice usage (Graham and Faizo, 2017).

These few examples underscore the capacity of alternative splicing to produce two or more proteins from a single gene that could have completely opposite functions with major consequences on cell fate and the transformation process.

CANCER-SPECIFIC MUTATIONS OR ALTERATIONS IN SPLICING FACTORS

Given that cancer cells do reprogram the spliceome, it is not surprising that splicing factors are common targets for deregulation in this disease (Dvinge et al., 2016). These include ESRP1 and ESRP2 (Warzecha et al., 2009), hnRNP A1, hnRNP A2, hnRNP A2/B1, hnRNP H, hnRNP K, and hnRNP M (Moran-Jones et al., 2009; David et al., 2010; Golan-Gerstl et al., 2011; Lefave et al., 2011; Xu et al., 2014; Gallardo et al., 2015), PRPF6 (Adler et al., 2014), PTBP1 (Izaguirre et al., 2012), QKI (Zong et al., 2014), RBFOX2 (Shapiro et al., 2011), RBM4, RBM5, RBM6, and RBM10 (Bonnal et al., 2008; Fushimi et al., 2008; Shapiro et al., 2011; Izaguirre et al., 2012; Bechara et al., 2013; Wang et al., 2014; Hernández et al., 2016), as well as SRSF1, SRSF2, SRSF3, SRSF6, and SRSF10 (Karni et al., 2007; Anczuków et al., 2012; Tang et al., 2013; Jensen et al., 2014; Zhou et al., 2014; Kim et al., 2015).

The SR protein SRSF2, for example, is a splicing factor that is commonly mutated in a collection of neoplastic diseases or cancers of immature blood cells known as Myelodysplastic Syndromes (MDS). Interestingly, mutations in SRSF2 that alter its sequence specificity on its target pre-mRNAs are more likely to be linked to MDS than nonsense mutations, indicating that a gain-of-function (binding to differential pre-mRNA targets) rather than loss-of-function of SRSF2 produces a new set of alternatively spliced mRNAs that are relevant to MDS development (Kim et al., 2015).

Of note, not all changes in splicing factors are due to mutations in their encoding genes as mutation-free disruptions in the repertoire of RNA-binding proteins (splicing factors) due to the imbalance in their expression is emerging as a common feature in many diseases including cancer. For example, frequent upregulation of mutation-free SRSF2 is a driver in the development of Hepatocellular Carcinoma (HCC) (Luo et al., 2017). SRSF1, also a splicing factor, is itself an oncogene whose expression is increased in cancers, including breast cancer (Das and Krainer, 2014; Akerman et al., 2015; Anczuków et al., 2015). These alterations in splicing factors, whether due to mutations or altered expression, tend to have large effects on cell phenotype as these splicing factors bind to and regulate the splicing of hundreds of pre-mRNAs. Thus, cancer cells can alter the splicing of a large number of genes by deregulating a handful of splicing factors. While this might seem to be an overkill, evidence does point to the fact that among the thousands of changes, some have distinct and significant effects on the transformation process. For example, SRSF2 mutants in MDS lead to mis-splicing of hundreds of pre-mRNAs, but one of them, the EZH2 pre-mRNA, encoding a transcriptional regulator that

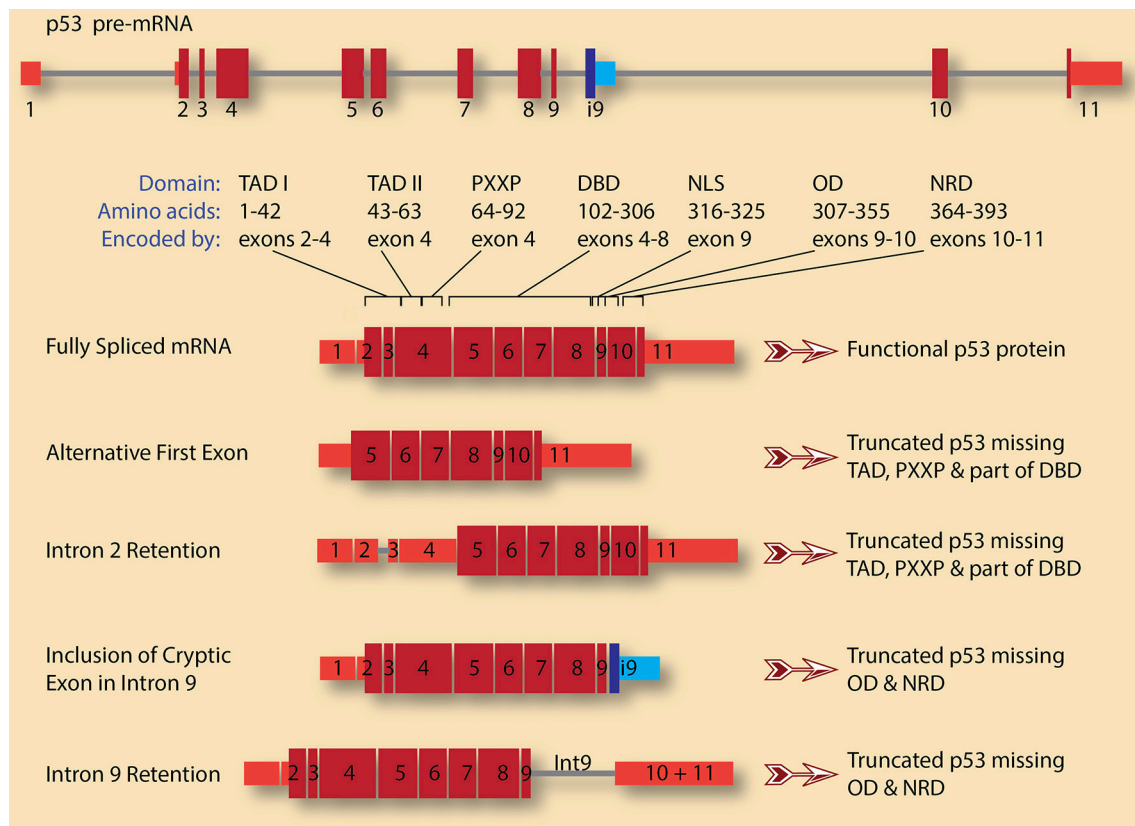


FIGURE 2 | TP53 pre-mRNA splicing has significant consequences on the gene product, p53 protein. A schematic diagram of p53 pre-mRNA shows 11 exons (red boxes for canonical exons and blue boxes for cryptic intronic exon 9, i9) and 10 introns (gray lines). The size and shade of the exon indicate whether it is untranslated region (narrow, light red) or coding (wide, dark red). The various domains and the exons encoding them are also indicated (adopted from Surget et al., 2013). These are: Transactivation domains I and II (TAD I and TAD II), proline-rich region (PXXP), DNA-binding domain (DBD), nuclear localization signal (NLS), oligomerization domain (OD), and negative regulation domain (NRD). A fully spliced mRNA containing the 11 canonical exons encodes for full length and functional p53 protein. Several isoforms with various alternative first exons can be generated for p53 pre-mRNA. In this case, the first exon is what is usually exon 5 in the canonical transcript, leading to the production of p53 protein lacking TADs, PXXP, and part of DBD. This truncated p53 is expected to be dominant negative. A similar protein can be encoded by transcripts in which intron 2 is retained, leading to the usage of start codon in exon 4 rather than the canonical start codon in exon 2. On the other hand, retention of intron 9 and/or inclusion of the cryptic intronic exon 9, i9, change the reading frame causing the loss of the encoded amino acids from exons 10 and 11. The resulting p53 proteins lack OD and NRD. These truncated p53 proteins could compete with wild type p53 for DNA binding but are not functional as they cannot oligomerize.

is required for maintaining the repressed state of many genes during hematopoiesis, stands out. Hematopoietic cells expressing SRSF2 mutants show higher inclusion of a highly conserved “poison” exon in the EZH2 mRNA, leading to degradation of the mRNA by nonsense-mediated decay and loss-of-function of the *EZH2* gene (Kim et al., 2015).

hnRNP proteins also play their share in cancer progression. For example, mis-regulation of a number of hnRNP proteins have been linked to HCC tumor progression, whereas the overexpression of hnRNP A1 in particular has been linked to tumor invasion and metastasis (Zhou Z. J. et al., 2013). The detailed contribution of the RNA splicing-dependent effects of mis-regulation of many hnRNPs in cancer is still under intense investigation by several laboratories and should shed some light on mechanisms as well as potential novel therapeutic targets.

Of note here is that mutations in splicing factors in MDS patients typically cause distinct and sometimes non-overlapping splicing defects, suggesting an alternate underlying mechanism.

Indeed, a recent study has uncovered that mutations on distinct splicing factors in MDS commonly cause elevated R-loops, replication stress, and activation of the ataxia telangiectasia and Rad3-related protein (ATR)-Chk1 pathway (Chen et al., 2018). These effects can lead to deregulated transcription pause release, raising the possibility that the MDS phenotype is related to a transcriptional defect rather than a splicing one.

CHANGES IN UPSTREAM SIGNALING PATHWAYS THAT DEREGULATE SPLICING FACTORS

In order to ensure that several splicing factors and other cellular processes are deregulated, the signaling pathways that relay extracellular signals to splicing factors are often targeted in cancer. The SR protein family is often deregulated as the function of SR proteins tightly depends on their phosphorylation status,

which itself is regulated by upstream kinases. For example, the splicing of the cassette exons in Caspase 9 pre-mRNA is regulated by the splicing factor SRSF1 leading to either caspase 9a or caspase 9b mRNAs. SRSF1 is itself phosphorylated upon activation of multiple signaling pathways, including the PI3K/AKT pathway. Since AKT signaling is often constitutively activated in cancers, such as lung cancer, this leads to constitutive phosphorylation of SRSF1 and deregulated expression of Caspase 9a/9b (Shultz et al., 2010). A similar pathway involving an AKT-hnRNP U axis has also been shown to regulate Caspase 9a/9b ratio (Vu et al., 2013). This deregulated Caspase 9a/9b ratio has marked consequences on apoptosis and contributes to the ability of cancer cells to resist cell death.

Interestingly, several key components of signaling pathways that are typically deregulated in cancer can themselves be alternatively spliced to produce cancer-specific isoforms. For example, the inclusion of exon 6 in the pre-mRNA of the First Apoptosis Signal (Fas) receptor produces an isoform that encodes a membrane bound receptor that plays a key role in relaying extracellular signal that lead to programmed cell death. On the other hand, the Fas isoform with exon 6 being skipped encodes a soluble protein that does not induce apoptosis upon relevant signaling. Epidermal Growth Factor Receptor (EGFR), Insulin Receptor (INSR), Receptor d'Origine Nantais (RON), and Vascular Endothelial Growth Factor Receptor (VEGFR) are among several receptor tyrosine kinases whose splicing is altered in cancer leading to tumor progression or reduced response to therapy (reviewed in, Abou-Fayçal et al., 2017). In the case of VEGFR, one intron retention leads to the production of a shorter and decoy receptor that is dominant negative (Kendall et al., 1996; Vorlová et al., 2011). Similarly, alternative splicing in EGFR pre-mRNA produces several isoforms, some of which are dominant negative whereas others are constitutively active, leading to enhanced tumorigenicity, migration and invasion (Guillaudeau et al., 2012a,b; Piccione et al., 2012; Zhou M. et al., 2013; Zhou Z. J. et al., 2013; Padfield et al., 2015).

ABERRATIONS IN SPLICEOSOMAL COMPONENTS THAT ARE LINKED TO CANCER

It is remarkable that loss-of-function mutations in core components of the spliceosome are not compatible with life, which speaks to the critical role the spliceosome plays in all cells. However, components of the spliceosome can be mutated without complete loss-of-function leading to widespread alterations in splicing and disease.

Patients with MDS, chronic myelomonocytic leukemia (CMML), or chronic lymphocytic leukemia (CLL) acquire mutations in the spliceosomal components SF3B1, SF1, PRPF40B, and U2AF35 besides mutations in the splicing factor SRSF2 and ZRSR2, a component of U11/U12 di-snRNP, (Armstrong et al., 2018). Interestingly, SF3B1 and U2AF35 mutation tend to be missense mutations and mutually exclusive, again suggesting that cells with severe aberrations in spliceosome function are not viable (Armstrong et al., 2018). These mutations

are drivers in cancer and they strongly correlate with prognosis and clinical phenotype.

On the other hand, several genetic diseases are linked to mutations in core components of the spliceosome. These include retinitis pigmentosa, a progressive neurodegeneration of Rod photoreceptors in the retina, which is linked to mutations in PRPF31, PRPF8, BRR2, PRPF4, or PRPF3. Spinal Muscular Atrophy (SMA) is a severe neurodegenerative disease caused by mutations in *SMN1* gene, which encodes a protein that functions in the biogenesis of spliceosomal snRNPs and reduced SMN function in cells has been shown to lead to widespread aberrations in splicing. Mutations in one of the snRNA components of the minor spliceosome, U4atac, have been identified and linked to severe mental retardation and dwarfism, microcephalic osteodysplastic primordial dwarfism type 1 (TALS/MOPD). Despite their low abundance in cells, minor introns are highly conserved and serve as critical molecular switches for the expression of genes that harbor them (Younis et al., 2013). Some of these genes are bona fide oncogenes and tumor suppressor genes, suggesting a role for deregulating minor intron splicing in cancer (Unpublished data).

SYSTEMATIC APPROACHES FOR IDENTIFYING SPLICING ABERRATIONS THAT ARE LINKED TO CANCER

Genetic alterations or mutations in cancer patients that affect the splicing of one gene are relatively easy to study, track, and even propose therapeutic tactics based on fixing the splicing of that one pre-mRNA. However, a major challenge emerges when the alteration is in a splicing factor or core component of the spliceosome as these lead to global (thousands) alterations in splicing affecting a wide range of genes. Still more challenging are cases where the expression of the splicing factors is altered without an obvious underlying genetic mutation. Two of the many challenges are: (1) identifying amongst the thousands of splicing alterations those that significantly contribute to cancer progression, and (2) therapeutically target the splicing factors without having massive side effects that sometimes are worse than the actual disease itself. In order to successfully address these points, it is important to develop a systematic and standardized approach to gain sufficient understanding of splicing deregulation in cancer and their impact on cancer.

Before the advent of next generation sequencing of RNA (referred to here as RNA-seq), transcriptome profiling to identify global splicing changes relied mostly on Gene Expression Exon Microarrays, that contain probes for almost all exons and many introns. While these arrays were a major advance over traditional microarrays with limited number of probes per gene, they tend to be very hard to interpret and generate a lot of false positives if not properly analyzed and replicated. Also, the lack of a large number of probes in introns causes these microarrays to miss a major category of splicing aberration that is intron retention. Nowadays, the gold standard for transcriptome profiling that includes both information on genome-wide expression level

changes as well as splicing disruptions is RNA-seq. While this method is both quantitative and qualitative, it has its own challenges as well. As a start, generating the libraries for RNA-seq from high quality RNA is expensive, tedious, and require well trained personnel. The data generated is massive in size (several gigabytes per sample) and requires powerful computing machines for storage as well as analysis. The analysis itself is a major bottleneck. Several off-the-shelf pipelines exist for mapping raw reads and analysis of differential expression of genes. However, these only scratch the surface and do not fully take advantage of the wealth of data generated by any well-designed RNA-seq experiment. For example, there are several publicly available algorithms that attempt to identify splicing alterations from RNA-seq, but our personal experience with these is that they all fail at capturing the real picture as most of them use statistical models that are not suitable for biological systems, leading to identification of an endless list of statistically significant but small changes that have no or little impact on the phenotype. Thus, many laboratories have opted for their own in-house pipelines that are suited for their own analysis but remain far from suited to apply globally. The major challenge in the analysis of RNA-seq data does not detract from the fact that it has been widely used to generate important databases of splicing alterations in many cancers. The list of splicing aberrations in cancer will grow and our understanding of the molecular basis of these changes as well as their contribution to cancer will improve tremendously in the coming years as our ability to standardize the analysis pipeline improves. In fact, we propose that identifying the right splicing isoforms can be so powerful, they should be used as novel biomarkers for many cancer types and subtypes.

Once enough molecular understanding of the splicing aberrations is gained and their impact on cancer is proven,

innovative RNA-based therapies are required to correct the splicing alterations or induce splicing changes in cancer cells that make them more susceptible to traditional chemotherapy. Only recently, RNA-based therapies, which include a range of mechanisms such as antisense oligonucleotides, RNAi, anti-miRNA, miRNA mimics, aptamers, ribozymes, and others, seemed far-fetched and impractical. However, the recent success of antisense oligonucleotides in correcting the splicing of exon 7 of *SMN* gene in SMA patients and its approval by the FDA speaks to the power of such therapies. In order to start applying such strategies to cancer, it is better to focus first on a few targets with large effects on phenotype. For example, given the large contribution of mis-splicing of genes involved in apoptosis on the ability of cancer cells to resist cell death, these genes are the low hanging fruit. In addition, it is noted that splicing alterations would rarely be the sole driver in cancer progression, we thus do not suggest the use of RNA-based therapies to overcome splicing aberrations as an alternate to traditional therapies, but rather a combination therapy for more effective treatment (we anticipate the effects to be synergistic) with less unfavorable side effects.

AUTHOR CONTRIBUTIONS

EE curated and compiled information toward completion of this review. IY wrote the review.

ACKNOWLEDGMENTS

This publication was made possible by the generous support of the Qatar Foundation through Carnegie Mellon University in Qatar's Seed Research program. The statements made herein are solely the responsibility of the authors.

REFERENCES

- Abou-Fayçal, C., Hatat, A. S., Gazzeri, S., and Eymin, B. (2017). Splice variants of the RTK family: their role in tumour progression and response to targeted therapy. *Int. J. Mol. Sci.* 18:E383. doi: 10.3390/ijms18020383
- Adler, A. S., McClelland, M. L., Yee, S., Yaylaoglu, M., Hussain, S., Cosino, E., et al. (2014). An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev.* 28, 1068–101084. doi: 10.1101/gad.237206.113
- Agrawal, S., and Eng, C. (2006). Differential expression of novel naturally occurring splice variants of PTEN and their functional consequences in Cowden syndrome and sporadic breast cancer. *Hum. Mol. Genet.* 15, 777–787. doi: 10.1093/hmg/ddi492
- Akerman, M., Fregoso, O. I., Das, S., Ruse, C., Jensen, M. A., Pappin, D. J., et al. (2015). Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biol.* 16:119. doi: 10.1186/s13059-015-0682-5
- Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shiroe, N. H., et al. (2015). SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell* 60, 105–117. doi: 10.1016/j.molcel.2015.09.005
- Anczuków, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., et al. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* 19, 220–228. doi: 10.1038/nsmb.2207
- Armstrong, R. N., Steeples, V., Singh, S., Sanchi, A., Boulwood, J., and Pellagatti, A. (2018). Splicing factor mutations in the myelodysplastic syndromes: target genes and therapeutic approaches. *Adv. Biol. Regul.* 67, 13–29. doi: 10.1016/j.jbior.2017.09.008
- Babic, I., Anderson, E. S., Tanaka, K., Guo, D., Masui, K., Li, B., et al. (2013). EGFR mutation-induced alternative splicing of Max contributes to growth of glycolytic tumors in brain cancer. *Cell Metab.* 17, 1000–1008. doi: 10.1016/j.cmet.2013.04.013
- Bartel, F., Taubert, H., and Harris, L. C. (2002). Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell* 2, 9–15. doi: 10.1016/S1535-6108(02)00091-0
- Bechara, E. G., Sebestyén, E., Bernardis, I., Eyra, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* 52, 720–733. doi: 10.1016/j.molcel.2013.11.010
- Bonnal, S., Martínez, C., Förch, P., Bachi, A., Wilm, M., and Valcárcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol. Cell* 32, 81–95. doi: 10.1016/j.molcel.2008.08.008
- Caldenhoven, E., van Dijk, T. B., Solari, R., Armstrong, J., Raaijmakers, J. A., Lammers, J. W., et al. (1996). STAT3beta, a splice variant of transcription factor STAT3, is a dominant negative regulator of transcription. *J. Biol. Chem.* 271, 13221–13227. doi: 10.1074/jbc.271.22.13221
- 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
- Chabot, B., and Shkreta, L. (2016). Defective control of pre-messenger RNA splicing in human disease. *J. Cell Biol.* 212, 13–27. doi: 10.1083/jcb.201510032
- Chakedis, J., French, R., Babicky, M., Jaquish, D., Howard, H., Mose, E., et al. (2016). A novel protein isoform of the RON tyrosine kinase receptor transforms human pancreatic duct epithelial cells. *Oncogene* 35, 3249–3259. doi: 10.1038/onc.2015.384

- Chen, L., Chen, J. Y., Huang, Y. J., Gu, Y., Qiu, J., Qian, H., et al. (2018). The augmented R-Loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell* 69, 412.e6–25. doi: 10.1016/j.molcel.2017.12.029
- Chen, M., Zhang, J., and Manley, J. L. (2010). Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA. *Cancer Res.* 70, 8977–8980. doi: 10.1158/0008-5472.CAN-10-2513
- Choi, J. W., Kim, D. G., Lee, A. E., Kim, H. R., Lee, J. Y., Kwon, N. H., et al. (2011). Cancer-associated splicing variant of tumor suppressor AIMP2/p38: pathological implication in tumorigenesis. *PLoS Genet.* 7:e1001351. doi: 10.1371/journal.pgen.1001351
- Das, S., and Krainer, A. R. (2014). Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol. Cancer Res.* 12, 1195–1204. doi: 10.1158/1541-7786.MCR-14-0131
- David, C. J., Chen, M., Assanah, M., Canoll, P., and Manley, J. L. (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* 463, 364–368. doi: 10.1038/nature08697
- David, C. J., and Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs uncharted. *Genes Dev.* 24, 2343–2364.
- DiFeo, A., Martignetti, J. A., and Narla, G. (2009). The role of KLF6 and its splice variants in cancer therapy. *Drug Resist. Updat.* 12, 1–7. doi: 10.1016/j.drug.2008.11.001
- Dvinge, H., and Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* 7:45. doi: 10.1186/s13073-015-0168-9
- Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* 16, 413–430. doi: 10.1038/nrc.2016.51
- Eckerich, C., Schulte, A., Martens, T., Zapf, S., Westphal, M., and Lamszus, K. (2009). RON receptor tyrosine kinase in human gliomas: expression, function, and identification of a novel soluble splice variant. *J. Neurochem.* 109, 969–980. doi: 10.1111/j.1471-4159.2009.06027.x
- 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
- Fedorov, A., Merican, A. F., and Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16128–16133. doi: 10.1073/pnas.242624899
- Feltes, C. M., Kudo, A., Blaschuk, O., and Byers, S. W. (2002). An alternatively spliced cadherin-11 enhances human breast cancer cell invasion. *Cancer Res.* 62, 6688–6697.
- Fushimi, K., Ray, P., Kar, A., Wang, L., Sutherland, L. C., and Wu, J. Y. (2008). Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15708–15713. doi: 10.1073/pnas.0805569105
- Gallardo, M., Lee, H. J., Zhang, X., Bueso-Ramos, C., Paeon, L. R., McArthur, M., et al. (2015). hnRNP K is a haploinsufficient tumor suppressor that regulates proliferation and differentiation programs in hematologic malignancies. *Cancer Cell* 28, 486–499. doi: 10.1016/j.ccell.2015.09.001
- Golan-Gerstl, R., Cohen, M., Shilo, A., Suh, S. S., Bakács, A., Coppola, L., et al. (2011). Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* 71, 4464–4472. doi: 10.1158/0008-5472.CAN-10-4410
- Graham, S. V., and Faizo, A. A. (2017). Control of human papillomavirus gene expression by alternative splicing. *Virus Res.* 231, 83–95. doi: 10.1016/j.virusres.2016.11.016
- Guillaudeau, A., Durand, K., Bessette, B., Chaunavel, A., Pommepuy, I., Progetti, F., et al. (2012a). EGFR soluble isoforms and their transcripts are expressed in meningiomas. *PLoS ONE* 7:e37204. doi: 10.1371/journal.pone.0037204
- Guillaudeau, A., Durand, K., Rabinovitch-Chable, H., Pommepuy, I., Mestouroux, L., and Robert, S. (2012b). Adult diffuse gliomas produce mRNA transcripts encoding EGFR isoforms lacking a tyrosine kinase domain. *Int. J. Oncol.* 40, 1142–1152. doi: 10.3892/ijo.2011.1287
- Harper, S. J., and Bates, D. O. (2008). VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat. Rev. Cancer* 8, 880–887. doi: 10.1038/nrc2505
- Hernández, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., and Valcárcel, J. (2016). Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA Biol.* 13, 466–472. doi: 10.1080/15476286.2016.1144004
- Izaguirre, D. I., Zhu, W., Hai, T., Cheung, H. C., Krahe, R., and Cote, G. J. (2012). PTBP1-dependent regulation of USP5 alternative RNA splicing plays a role in glioblastoma tumorigenesis. *Mol. Carcinog.* 51, 895–906. doi: 10.1002/mc.20859
- Jang, H. N., Lee, M., Loh, T. J., Choi, S. W., Oh, H. K., Moon, H., et al. (2014). Exon 9 skipping of apoptotic caspase-2 pre-mRNA is promoted by SRSF3 through interaction with exon 8. *Biochim. Biophys. Acta* 1839, 25–32. doi: 10.1016/j.bbagr.2013.11.006
- Jensen, M. A., Wilkinson, J. E., and Krainer, A. R. (2014). Splicing factor SRSF6 promotes hyperplasia of sensitized skin. *Nat. Struct. Mol. Biol.* 21, 189–197. doi: 10.1038/nsmb.2756
- Kalsotra, A., and Cooper, T. A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729. doi: 10.1038/nrg3052
- Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* 14, 185–193. doi: 10.1038/nsmb1209
- Kendall, R. L., Wang, G., and Thomas, K. A. (1996). Identification of a natural soluble form of the vascular endothelial growth factor receptor, FLT-1, and its heterodimerization with KDR. *Biochem. Biophys. Res. Commun.* 226, 324–328. doi: 10.1006/bbrc.1996.1355
- Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., Ramakrishnan, A., et al. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* 27, 617–630. doi: 10.1016/j.ccell.2015.04.006
- Koslowski, M., Türeci, O., Bell, C., Krause, P., Lehr, H. A., Brunner, J., et al. (2002). Multiple splice variants of lactate dehydrogenase C selectively expressed in human cancer. *Cancer Res.* 62, 6750–6755.
- Lefave, C. V., Squatrito, M., Vorlova, S., Rocco, G. L., Brennan, C. W., Holland, E. C., et al. (2011). Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J.* 30, 4084–4097. doi: 10.1038/emboj.2011.259
- Li, Y., Hu, X., Song, Y., Lu, Z., Ning, T., and Cai, H. (2011). Identification of novel alternative splicing variants of interferon regulatory factor 3. *Biochim. Biophys. Acta* 1809, 166–175. doi: 10.1016/j.bbagr.2011.01.006
- Liu, X., Wang, Y., Chang, G., Wang, F., Wang, F., and Geng, X. (2017). Alternative splicing of hTERT Pre-mRNA: a potential strategy for the regulation of telomerase activity. *Int. J. Mol. Sci.* 18:567. doi: 10.3390/ijms18030567
- Luo, C., Cheng, Y., Liu, Y., Chen, L., Liu, L., Wei, N., et al. (2017). SRSF2 regulates alternative splicing to drive hepatocellular carcinoma development. *Cancer Res.* 77, 1168–1178. doi: 10.1158/0008-5472.CAN-16-1919
- Mohr, A., Zwacka, R. M., Jarmy, G., Büneker, C., Schrezenmeier, H., Dohner, K., et al. (2005). Caspase-8L expression protects CD34+ hematopoietic progenitor cells and leukemic cells from CD95-mediated apoptosis. *Oncogene* 24, 2421–2429. doi: 10.1038/sj.onc.1208432
- Moran-Jones, K., Grindlay, J., Jones, M., Smith, R., and Norman, J. C. (2009). hnRNP A2 regulates alternative mRNA splicing of TP53INP2 to control invasive cell migration. *Cancer Res.* 69, 9219–9227. doi: 10.1158/0008-5472.CAN-09-1852
- Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909
- Okumura, N., Yoshida, H., Kitagishi, Y., Nishimura, Y., and Matsuda, S. (2011). Alternative splicing on p53, BRCA1 and PTEN genes involved in breast cancer. *Biochem. Biophys. Res. Commun.* 413, 395–399. doi: 10.1016/j.bbrc.2011.08.098
- Olsson, M., and Zhivotovsky, B. (2011). Caspases and cancer. *Cell Death Differ.* 18, 1441–1449. doi: 10.1038/cdd.2011.30
- Oltean, S., and Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311–5318. doi: 10.1038/onc.2013.533
- Padfield, E., Ellis, H. P., and Kurian, K. M. (2015). Current therapeutic advances targeting EGFR and EGFRvIII in glioblastoma. *Front. Oncol.* 5:5. doi: 10.3389/fonc.2015.00005
- 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
- Paronetto, M. P., Cappellari, M., Busà, R., Pedrotti, S., Vitali, R., Comstock, C., et al. (2010). Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. *Cancer Res.* 70, 229–239. doi: 10.1158/0008-5472.CAN-09-2788

- Paronetto, M. P., Passacantilli, I., and Sette, C. (2016). Alternative splicing and cell survival: from tissue homeostasis to disease. *Cell Death Differ.* 23, 1919–1929. doi: 10.1038/cdd.2016.91
- Piccione, E. C., Lieu, T. J., Gentile, C. F., Williams, T. R., Connolly, A. J., and Godwin, A. K. (2012). A novel epidermal growth factor receptor variant lacking multiple domains directly activates transcription and is overexpressed in tumors. *Oncogene* 31, 2953–2967. doi: 10.1038/nc.2011.465
- Pritchard-Jones, R. O., Dunn, D. B., Qiu, Y., Varey, A. H., Orlando, A., Rigby, H., et al. (2007). Expression of VEGF(XXX)b, the inhibitory isoforms of VEGF, in malignant melanoma. *Br. J. Cancer* 97, 223–230. doi: 10.1038/sj.bjc.6603839
- Prochazka, L., Tesarik, R., and Turanek, J. (2014). Regulation of alternative splicing of CD44 in cancer. *Cell. Signal.* 26, 2234–2239. doi: 10.1016/j.cellsig.2014.07.011
- Radisky, D. C., Levy, D. D., Littlepage, L. E., Liu, H., Nelson, C. M., and Fata, J. E. (2005). Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* 436, 123–127. doi: 10.1038/nature03688
- Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G., and Koonin, E. V. (2003). Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13, 1512–1517. doi: 10.1016/S0960-9822(03)00558-X
- Schwerk, C., and Schulze-Osthoff, K. (2005). Regulation of apoptosis by alternative pre-mRNA splicing. *Mol. Cell* 19, 1–13. doi: 10.1016/j.molcel.2005.05.026
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3
- Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., et al. (2011). An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.* 7:e1002218. doi: 10.1371/journal.pgen.1002218
- Shieh, J. J., Liu, K. T., Huang, S. W., Chen, Y. J., and Hsieh, T. Y. (2009). Modification of alternative splicing of Mcl-1 pre-mRNA using antisense morpholino oligonucleotides induces apoptosis in basal cell carcinoma cells. *J. Invest. Dermatol.* 129, 2497–2506. doi: 10.1038/jid.2009.83
- Shultz, J. C., Goehre, R. W., Wijesinghe, D. S., Murudkar, C., Hawkins, A. J., Shay, J. W., et al. (2010). Alternative splicing of caspase 9 is modulated by the phosphoinositide 3-kinase/Akt pathway via phosphorylation of SRP30a. *Cancer Res.* 70, 9185–9196. doi: 10.1158/0008-5472.CAN-10-1545
- Singh, A., Karnoub, A. E., Palmby, T. R., Lengyel, E., Sondek, J., and Der, C. J. (2004). Rac1b, a tumor associated, constitutively active Rac1 splice variant, promotes cellular transformation. *Oncogene* 23, 9369–9380. doi: 10.1038/sj.onc.1208182
- Surget, S., Khoury, M. P., and Bourdon, J. C. (2013). Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Onco. Targets. Ther.* 7, 57–68. doi: 10.2147/OTT.S53876
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A., and Skotheim, R. I. (2016). Aberrant RNA splicing in cancer: expression changes and driver mutations of splicing factor genes. *Oncogene* 35, 2413–2427. doi: 10.1038/nc.2015.318
- Tang, Y., Horikawa, I., Ajiro, M., Robles, A. I., Fujita, K., Mondal, A. M., et al. (2013). Downregulation of splicing factor SRSF3 induces p53beta, an alternatively spliced isoform of p53 that promotes cellular senescence. *Oncogene* 32, 2792–2798. doi: 10.1038/nc.2012.288
- 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
- Tejedor, J. R., Papasaiakas, P., and Valcarcel, J. (2015). Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol. Cell* 57, 23–38. doi: 10.1016/j.molcel.2014.10.029
- Turunen, J. J., Niemelä, E. H., Verma, B., and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76. doi: 10.1002/wrna.1141
- van den Heuvel, A. P., Jing, J., Wooster, R. F., and Bachman, K. E. (2012). Analysis of glutamine dependency in non-small cell lung cancer: GLS1 splice variant GAC is essential for cancer cell growth. *Cancer Biol. Ther.* 13, 1185–1194. doi: 10.4161/cbt.21348
- van Doorn, R., Dijkman, R., Vermeer, M. H., Starink, T. M., Willemze, R., and Tensen, C. P. (2002). A novel splice variant of the Fas gene in patients with cutaneous T-cell lymphoma. *Cancer Res.* 62, 5389–5392.
- Vorlová, S., Rocco, G., Lefave, C. V., Jodelka, F. M., Hess, K., and Hastings, M. L. (2011). Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* 43, 927–939. doi: 10.1016/j.molcel.2011.08.009
- Vu, N. T., Park, M. A., Shultz, J. C., Goehre, R. W., Hoeflerlin, L. A., Shultz, M. D., et al. (2013). hnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. *J. Biol. Chem.* 288, 8575–8584. doi: 10.1074/jbc.M112.443333
- Wagner, E. J., Curtis, M. L., Robson, N. D., Baraniak, A. P., Eis, P. S., and Garcia-Blanco, M. A. (2003). Quantification of alternatively spliced FGFR2 RNAs using the RNA invasive cleavage assay. *RNA* 9, 1552–1561. doi: 10.1261/rna.5840803
- Wang, E. T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, Y., Chen, D., Qian, H., Tsai, Y. S., Shao, S., Liu, Q., et al. (2014). The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. *Cancer Cell* 26, 374–389. doi: 10.1016/j.ccr.2014.07.010
- Warzecha, C. C., Shen, S., Xing, Y., and Carstens, R. P. (2009). The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol.* 6, 546–562. doi: 10.4161/rna.6.5.9606
- 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
- Xu, Y., Gao, X., Lee, J. H., Huang, H., Tan, H., Ahn, J., et al. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev.* 28, 1191–1203. doi: 10.1101/gad.241968.114
- 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
- Zammarchi, F., de Stanchina, E., Bournazou, E., Supakorndej, T., Martires, K., Riedel, E., et al. (2011). Antitumorigenic potential of STAT3 alternative splicing modulation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17779–17784. doi: 10.1073/pnas.1108482108
- Zhang, J., and Manley, J. L. (2013). Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov.* 3, 1228–1237. doi: 10.1158/2159-8290.CD-13-0253
- Zhao, Q., Caballero, O. L., Davis, I. D., Jonasch, E., Tamboli, P., Yung, W. K., et al. (2013). Tumor-specific isoform switch of the fibroblast growth factor receptor 2 underlies the mesenchymal and malignant phenotypes of clear cell renal cell carcinomas. *Clin. Cancer Res.* 19, 2460–2472. doi: 10.1158/1078-0432.CCR-12-3708
- Zhou, M., Wang, H., Zhou, K., Luo, X., Pan, X., Shi, B., et al. (2013). A novel EGFR isoform confers increased invasiveness to cancer cells. *Cancer Res.* 73, 7056–7067. doi: 10.1158/0008-5472.CAN-13-0194
- Zhou, X., Li, X., Cheng, Y., Wu, W., Xie, Z., Xi, Q., et al. (2014). BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat. Commun.* 5:4581. doi: 10.1038/ncomms5581
- Zhou, Y. Q., He, C., Chen, Y. Q., Wang, D., and Wang, M. H. (2003). Altered expression of the RON receptor tyrosine kinase in primary human colorectal adenocarcinomas: generation of different splicing RON variants and their oncogenic potential. *Oncogene* 22, 186–197. doi: 10.1038/sj.onc.1206075
- Zhou, Z. J., Dai, Z., Zhou, S. L., Fu, X. T., Zhao, Y. M., Shi, Y. H., et al. (2013). Overexpression of HnRNP A1 promotes tumor invasion through regulating CD44v6 and indicates poor prognosis for hepatocellular carcinoma. *Int. J. Cancer* 132, 1080–1089. doi: 10.1002/ijc.27742
- Zong, F. Y., Fu, X., Wei, W. J., Luo, Y. G., Heiner, M., Cao, L. J., et al. (2014). The RNA-binding protein KFI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 10:e1004289. doi: 10.1371/journal.pgen.1004289

Conflict of Interest Statement: 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.

Copyright © 2018 El Marabti and Younis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing in Neurogenesis and Brain Development

Chun-Hao Su¹, Dhananjaya D^{1,2} and Woan-Yuh Tarn^{1,2*}

¹ Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, ² Taiwan International Graduate Program in Molecular Medicine, National Yang-Ming University and Academia Sinica, Taipei, Taiwan

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Hidehito Kuroyanagi,
Tokyo Medical and Dental University,
Japan
Ihab Younis,
Carnegie Mellon University in Qatar,
Qatar

*Correspondence:

Woan-Yuh Tarn
wtarn@ibms.sinica.edu.tw

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 17 November 2017

Accepted: 25 January 2018

Published: 12 February 2018

Citation:

Su C-H, D D and Tarn W-Y (2018)
Alternative Splicing in Neurogenesis
and Brain Development.
Front. Mol. Biosci. 5:12.
doi: 10.3389/fmolb.2018.00012

Alternative splicing of precursor mRNA is an important mechanism that increases transcriptomic and proteomic diversity and also post-transcriptionally regulates mRNA levels. Alternative splicing occurs at high frequency in brain tissues and contributes to every step of nervous system development, including cell-fate decisions, neuronal migration, axon guidance, and synaptogenesis. Genetic manipulation and RNA sequencing have provided insights into the molecular mechanisms underlying the effects of alternative splicing in stem cell self-renewal and neuronal fate specification. Timely expression and perhaps post-translational modification of neuron-specific splicing regulators play important roles in neuronal development. Alternative splicing of many key transcription regulators or epigenetic factors reprograms the transcriptome and hence contributes to stem cell fate determination. During neuronal differentiation, alternative splicing also modulates signaling activity, centriolar dynamics, and metabolic pathways. Moreover, alternative splicing impacts cortical lamination and neuronal development and function. In this review, we focus on recent progress toward understanding the contributions of alternative splicing to neurogenesis and brain development, which has shed light on how splicing defects may cause brain disorders and diseases.

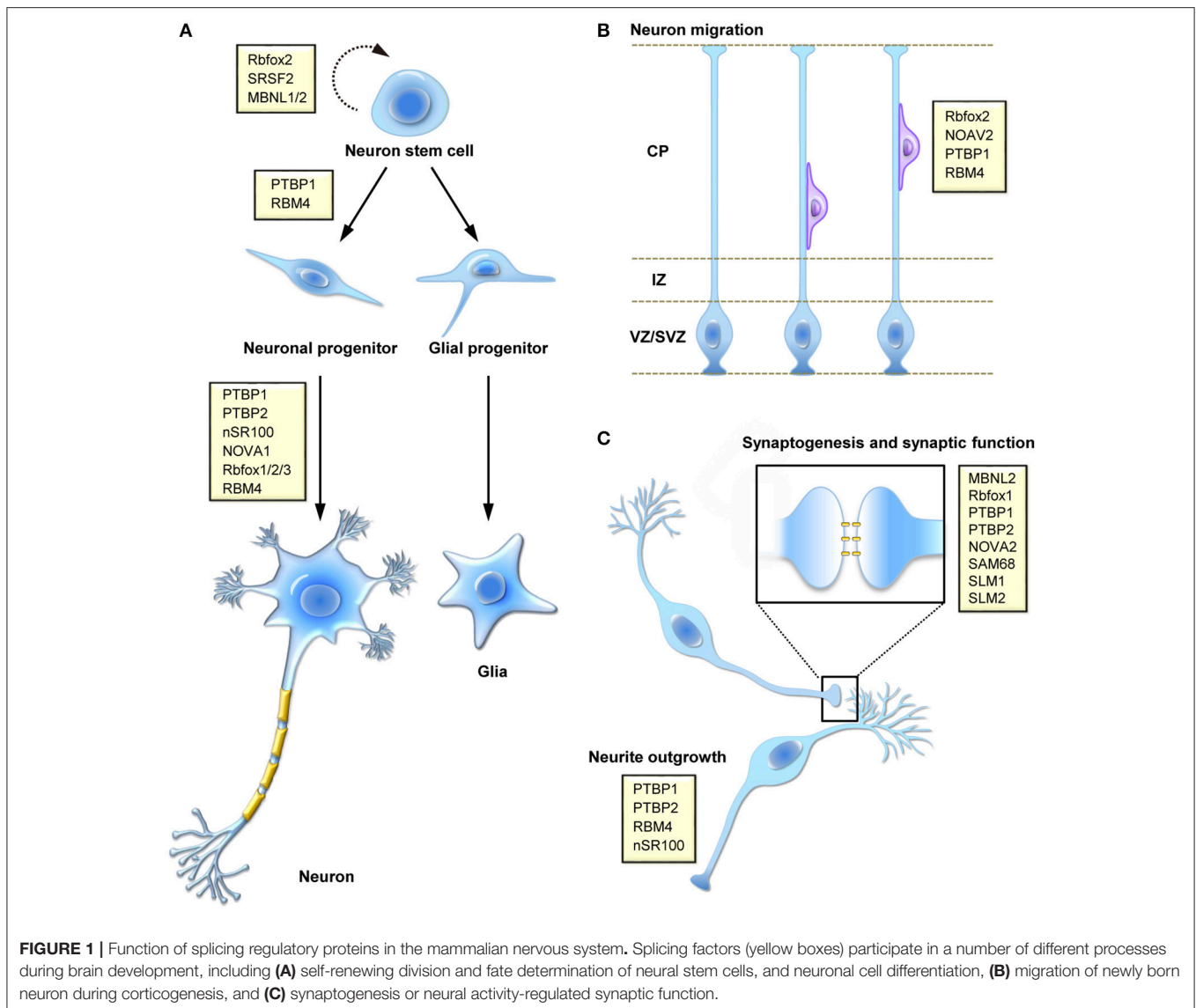
Keywords: alternative splicing, splicing factors, neurogenesis, neuronal differentiation, neuronal migration, neuronal development

INTRODUCTION

Alternative splicing is a crucial step of post-transcriptional gene expression that substantially increases transcriptome diversity and is critical for diverse cellular processes, including cell differentiation and development as well as cell reprogramming and tissue remodeling. Our understanding of the physiological significance and disease implications of alternative splicing has been greatly improved by genetic approaches and RNA deep sequencing. In this review, we focus on alternative splicing in neuronal differentiation from stem/progenitor cells, neuronal migration and functional development of neurons (**Figure 1**).

Alternative Splicing and Its Role in Development

Approximately 95% of human multi-exon genes undergo alternative splicing of precursor mRNAs (pre-mRNAs) (Pan et al., 2008; Wang et al., 2008). In mammals, alternative splicing involves differential use of intron splice sites or the inclusion/exclusion of exons. Alternatively spliced mRNAs may generate protein isoforms with distinct and perhaps antagonistic functions or with altered stability or subcellular localization (Dredge et al., 2001; Matlin et al., 2005). In addition, alternative splicing may introduce premature termination codons into the resulting



mature mRNAs, leading to mRNA downregulation via nonsense-mediated decay (Lareau et al., 2007). Alternative splicing is governed by the interplay between *trans*-acting splicing regulators and *cis*-elements of pre-mRNAs (Matera and Wang, 2014). In general, a splicing activator may enhance splice-site recognition or utilization by the spliceosome, whereas a splicing suppressor may prevent the association of spliceosomal factors with pre-mRNAs or compete off splicing activators. Moreover, alternative splicing is also influenced by transcription rate, histone modifications, and chromatin structure (Kornbliht et al., 2004, 2013; Luco et al., 2010). Alternative splicing may occur in a tissue- or developmental-specific manner or in response to cellular signals and no doubt plays critical roles in many cellular processes (Nilsen and Graveley, 2010).

Alternative splicing provides a means to differentiate gene expression between cell types during development. Tissue-specific regulation of alternative splicing involves the coordinated

actions of splicing factors. Cell type-specific or timely expression of certain splicing regulators is important for precise control of alternative splicing. For example, the RNA-binding protein CUGBP and ETR-3-like factor 1 (CELFI) and muscleblind-like 1 (MBNL1) exhibit switched expression during heart development to regulate splicing of cardiac mRNAs (Kalsotra et al., 2008). Forced expression of embryonic CELFI or ablation of MBNL1 in the adult mouse heart reverts splicing toward embryonic/early postnatal patterns (Kalsotra et al., 2008). Similarly, switching of splicing regulators also occurs in the developing brain (see below). Thus, temporal control of alternative splicing is critical for fetal-to-adult transitions during development. Coordinated splicing networks contribute substantially to the development of various tissues and organs as well as their physiology.

Splicing abnormalities are linked to human genetic diseases, including brain disorder (Raj and Blencowe, 2015; Vuong et al., 2016). For example, familial dysautonomia is caused by

a 5' splice site mutation of the *IKBKAP* gene (Slaugenhaupt et al., 2001). This mutation reduces *IKBKAP* expression via alternative splicing-coupled nonsense-mediated decay, and hence downregulates a set of cell migration-related genes (Anderson et al., 2001; Yoshida et al., 2015). Gene abnormalities in the splicing factor *RBFOX1* gene have been linked to autism spectrum disorder and additional neuromuscular abnormalities (Barnby et al., 2005; Martin et al., 2007; Conboy, 2017). The associations between splicing defects and human disease have been reviewed extensively elsewhere, and will not be emphasized in this review.

Experimental Insights into the Role of Alternative Splicing in Brain Development

Emerging new technologies for RNA studies have greatly enhanced our knowledge of alternative splicing in development. Capture of specific mRNA ribonucleoproteins followed by high-throughput sequencing or splicing microarrays has identified dynamic alternative splicing programs during cell differentiation or development and also revealed the tissue-specific or developmentally regulated RNA-binding landscapes of splicing factors (Rossbach et al., 2014). Use of knockout and transgenic mice has identified the targets and physiological roles of neuronal splicing regulators and revealed how their defects impact brain development and neuronal function (Table 1). Moreover, genetic tagging with a reporter provides a tool for isolating specific cell types for transcriptome comparison (Wang et al., 2011). For example, by using *Tbr2* promoter-driven green fluorescent protein as a tracer, neural progenitor cells (NPCs) can be distinguished from neurons in the developing brain (Zhang et al., 2016). Recently, single-cell profiling techniques enabled the resolution of population heterogeneity and revealed insights into cellular differentiation and development (Darmanis et al., 2015). Computational analysis of deep-sequencing data and annotated databases helped establish the correlation between genetic mutations, splicing variants, and disease (Kircher et al., 2014; Mort et al., 2014). Recently, an unbiased “deep-learning” computational method provided a more powerful link between rare single-nucleotide variations and neurological disorders such as spinal muscular atrophy and autism spectrum disorder (Xiong et al., 2015). Advanced sequencing tools would likely facilitate the detection of cell type- and stimulus-dependent splicing changes and perhaps the identification of previously unrecognized splicing products such as circular RNAs during neuronal development (van Rossum et al., 2016).

Neuronal Differentiation Involves Coordinated Changes in the Expression of Splicing Factors

Genome-wide transcriptome analysis has revealed an exceptionally high level of alternative splicing in the mammalian brain (Yeo et al., 2004). The nervous system adopts alternative splicing for cell differentiation, morphogenesis, the formation of complex neuronal networks, and the establishment/plasticity of delicate synapses (Norris and Calarco, 2012; Zheng and Black, 2013). Splicing regulation may involve some neuron-specific

splicing factors and their interplay with ubiquitous factors (Raj and Blencowe, 2015; Vuong et al., 2016). A switch from predominant expression of PTBP1 to its neuronal paralog PTBP2 (nPTB), which occurs during differentiation of progenitor cells into postmitotic neurons, is important for the stem cell-to-neuron transition (Boutz et al., 2007; Vuong et al., 2016). PTBP1 is downregulated by the neuron-specific microRNA miR-124 (Makeyev et al., 2007). Notably, PTBP1 suppresses the inclusion of exon 10 of *PTBP2*, producing an exon 10-skipped mRNA that is susceptible to nonsense-mediated decay (Figure 2). Thus, PTBP1 restricts the level of PTBP2 in non-neuronal cells or NPCs. RBM4 is a ubiquitous RNA-binding protein, but its level is elevated during neuronal differentiation of mouse embryonal carcinoma P19 cells (Tarn et al., 2016). Interestingly, RBM4 acts in the same manner as PTBP1 to suppress exon 11/10 of *PTBP1/PTBP2* in myoblast cells, and it downregulates PTBP1/PTBP2 levels (Lin and Tarn, 2011; Figure 2). However, during neuronal differentiation of mesenchymal stem cells, RBM4 induces the skipping of mammalian-specific exon 9 of *PTBP1*, which produces a functional PTBP1 isoform with compromised splicing activity compared with full-length PTBP1 (Su et al., 2017). Therefore, RBM4 attenuates the activity of PTBP1 in splicing regulation (Su et al., 2017; Figure 2). Notably, *PTBP2* does not contain an exon equivalent of exon 9 of *PTBP1*, so PTBP2 is likely resistant to regulation by RBM4 during stem cell differentiation. On the other hand, the neural-specific SR-related protein of 100 kDa (nSR100/SRRM4) promotes exon 10 inclusion of *PTBP2* and thus maintains PTBP2 level in neurons (Calarco et al., 2009).

PTBP1 and PTBP2 regulate overlapping but distinct repertoires of splicing events. PTBP1 suppresses the splicing of a subset of neural targets to inhibit neuronal differentiation. PTBP2 expression is elevated in differentiating neuronal cells and activates certain neural targets that promote differentiation (Boutz et al., 2007). Nevertheless, PTBP2 is downregulated as cells mature and undergo synaptogenesis. This sequential downregulation of PTBP1 and PTBP2 is important for two transitions of splicing regulation throughout neuronal differentiation and maturation and for functional expression of postsynaptic density protein-95 (PSD-95) via splicing control (Zheng et al., 2012). Both RBM4 and PTBP1 have preference for CU-rich *cis*-elements and hence antagonize each other during splicing regulation; thus, in general, they function oppositely in cell differentiation.

Besides the above, the neuron-specific splicing regulator Nova-1 can negatively autoregulate its own expression by suppressing exon 4 inclusion (Dredge et al., 2005). A study revealed that RBM4 promotes *Nova-1* exon 4 inclusion during differentiation and maturation of brown adipocytes (Lin J. C. et al., 2016), but whether this regulation occurs in neurons is unclear. Moreover, all three *Rbfox* family members exploit a conserved mechanism of splicing autoregulation to produce a splice isoform with a truncated RNA-recognition motif; this isoform has dominant-negative activity in splicing (Damianov and Black, 2010). The splicing switch of *RBFOX3* from the truncated isoform to the full-length protein occurs in a development-dependent manner, and the

TABLE 1 | Examples of the function of neuronal splicing regulators in neuronal differentiation and brain development.

Splicing regulators	Targets	Knockout/downregulation of splicing regulators		References
		Target exon	Phenotypes	
nSR100/SRRM4	<i>Protrudin (Zfyve27)</i>	suppressed inclusion of exon L between exon 8 and 9	impaired neurite outgrowth	Ohnishi et al., 2017
Ptbp1 & 2	<i>PSD-95 (Dlg4)</i>	exon 18 inclusion	impaired development of glutamatergic neurons	Zheng et al., 2012
Ptbp1	<i>Flna</i>	included the poison exon	brain specific malformation	Zhang et al., 2016
Ptbp2	<i>Dnm1</i>	altered mutually exclusive selection of exons 9a/9b	impaired synaptic function, and caused seizures and behavioral deficits	Li et al., 2014
Nova2	<i>Dab1</i>	exon 7bc (9bc) inclusion	impaired radial migration and Purkinje neuron migration	Yano et al., 2010
Rbfox3	<i>Numb</i>	repressed Numb exon 12 inclusion	impaired neuronal differentiation	Kim et al., 2013
Rbfox1	<i>Snap25</i>	altered mutually exclusive selection of exons 5a/5b	caused seizure	Gehman et al., 2011
SRSF1	<i>ApoER2</i>	promote exon 19 inclusion	impaired synapse formation and function	Hinrich et al., 2016
hnRNP H1/H2	<i>TRF2</i>	exon 7 (TRF2-S) inclusion	impaired neuronal differentiation.	Grammatikakis et al., 2016
RBM4	<i>Numb</i>	increased exon 3 skipping and exon 9 inclusion	impaired neuronal differentiation and neuronal outgrowth	Tarn et al., 2016

PSD-95: postsynaptic density protein 95.

Dnm1: dynamin1.

Flna: filamin A.

Dab1: disabled homolog-1.

hnRNP: heterogeneous nuclear ribonucleoprotein.

Snap25: synaptosomal-associated protein 25.

ApoER2: apolipoprotein E receptor 2.

TRF2: telomeric repeat-binding factor 2.

latter is necessary for late neuronal differentiation (Kim et al., 2013).

Together, precise timing and level control of splicing regulators is critical for dynamic alternative splicing regulation during cell differentiation and development.

Alternative Splicing in Self-renewal and Differentiation of Stem Cells

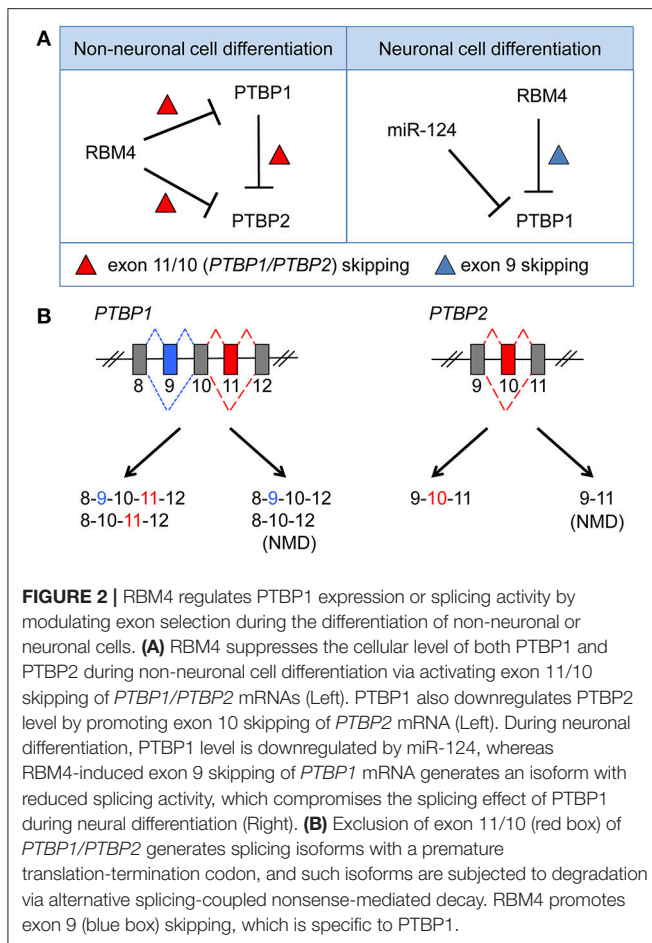
Alternative splicing also plays a critical role in self-renewal of pluripotent cells as well as in cell-fate determination and reprogramming (Graveley et al., 2011; Ye and Belloch, 2014). Genome-wide RNA sequencing (RNA-seq) studies have revealed that stem cells and differentiated cells exhibit different splicing profiles (Pritsker et al., 2005). Fine-tuning the expression of several stemness-related transcription factors such as Oct4, Nanog, Sox2, and Tcf3 is important for pluripotency maintenance (Chen et al., 2008; Kim et al., 2008). In particular, different isoforms of *Tcf3* and *Oct4* influence self-renewal of stem cells (Atlasi et al., 2008; Salomonis et al., 2010). The forkhead box transcription factor FoxP1 plays a hierarchical role in the transcription network of pluripotency; the switching of its mutually exclusive exons controls pluripotency and reprogramming of embryonic stem cells (Gabut et al., 2011). Several splicing factors modulate alternative splicing in embryonic stem cells and contribute positively (such as Rbfox2 and SRSF2) or negatively (such as MBNL1/2) to maintaining the stem cell splicing program (Ye and Belloch, 2014). Thus, alternative splicing plays a

critical role in the decision between stem cell self-renewal and differentiation.

Alternative splicing modulates the activity of certain histone modification enzymes in neuronal cells and hence influences the epigenetic status (Fiszbein and Kornblihtt, 2016). The histone methyltransferase G9a is a suppressor of pluripotency-related genes (Kellner and Kikyo, 2010). During neuronal differentiation of neuroblastoma neuro-2a cells, alternative exon inclusion of G9a promotes its nuclear localization and hence increases the dimethylation of histone 3 lysine 9 (H3K9me2). Thus, the regulation of G9a alternative splicing is necessary for efficient neuronal differentiation (Fiszbein et al., 2016). More intriguingly, alternative splicing also modulates the activity of the demethylase LSD1 (Laurent et al., 2015). Therefore, the balanced methylation of H3K9 is likely important for regulating gene expression profiles during neuronal differentiation.

Alternative Splicing in Differentiation of Neuronal Stem/Progenitor Cells

Transcriptome profiling demonstrated the dynamic nature of alternative splicing events in different cell types, brain regions, and developmental stages (Johnson et al., 2009; Zhang et al., 2014; Yan et al., 2015). RNA-seq analysis of purified NPCs and differentiating neurons in the mouse cortex revealed an alternative splicing switch for a set of neuron-specific exons during differentiation (Zhang et al., 2016). Analysis of human cerebral organoids and fetal neocortex also revealed different splicing patterns in intermediate progenitor cells, radial



glial cells, immature neurons, and neurons during cortical development (Camp et al., 2015; Zhang et al., 2016). Therefore, splicing regulation establishes cell type- and stage-specific gene expression profiles during neurogenesis and brain development, which rely on proper expression and function of splicing regulators (Raj and Blencowe, 2015; Vuong et al., 2016; Baralle and Giudice, 2017).

Among neuronal splicing regulators, PTBP1 is exclusively expressed in embryonic stem cells and NPCs, whereas PTBP2 and Rbfox proteins are mainly expressed in neurons. A recent report showed that PTBP1 and Rbfox antagonistically modulate neuronal fate via their roles in regulating alternative exon selection (Zhang et al., 2016). Rbfox switches the centrosomal isoform of *Ninein* to the non-centrosomal form as a result of alternative splicing and hence influences centriolar dynamics and promotes NPC differentiation. On the other hand, PTBP1 suppresses a premature stop codon-containing exon of *filamin A* (*Flna*) in NPCs and hence maintains apical progenitors. Genetic mutations that generate aberrant *Flna* splice isoforms in NPCs are linked to periventricular nodular heterotopia, a neuronal migration disorder. Thus, a better understanding of the mechanisms of neuronal alternative splicing may provide plausible treatment strategies for neuronal disorders.

The Notch receptors play a critical role in fate decisions of various stem/progenitor cells, and Numb is a critical effector of Notch signaling. Alternative splicing of exons 3 and 9 of *Numb* generate four different isoforms, which differentially modulate Notch activity. The detail of how alternative splicing of *Numb* modulates cell differentiation is not completely known. Rbfox3 can regulate alternative splicing of *Numb*, and Rbfox3 depletion impairs neurogenesis in the hippocampal dentate gyrus (Kim et al., 2013; Lin Y. S. et al., 2016). Our recent study showed that RBM4 determines the selection of two alternative exons, and its overexpression preferentially produces a *Numb* isoform with the highest potential to promote *Mash1* expression and subsequent differentiation of neuronal progenitor cells. Moreover, additional splicing regulators of *Numb* have been implicated in either cancer progression or tumor suppression (Bechara et al., 2013; Zong et al., 2014). Thus, it is conceivable that fine-tuning the expression of *Numb* isoforms during fate decision of neuronal progenitor cells may constitute a combinatorial effect of multiple splicing regulators.

Different Alternative Splicing Patterns in Neurons and Glia

Brain tissues comprise a variety of cell types including neural precursor cells, neurons, and various subtypes of neuroglia. Tantalizing issues remain as to whether and how alternative splicing influences neural fate determination and which splicing regulators are involved (Raj and Blencowe, 2015). Expression of specific alternatively spliced isoforms in distinct neurons has been reported in *Caenorhabditis elegans* and *Drosophila* (Lah et al., 2014; Norris et al., 2014). For example, UNC75 and EXC7 (respective homologs of mammalian CELF and Hu/ELAV) differentially modulate alternative splicing of *unc-16* in GABAergic motor neurons and cholinergic motor neurons (Norris et al., 2014). The energy requirement of different types of brain cells varies; the oxidative and glycolytic pathways predominate in neurons and astrocytes, respectively (Magistretti and Allaman, 2015). Transcriptome profiling has revealed distinct *pyruvate kinase M* (*PKM*) splice isoforms, i.e., *PKM1* and *PKM2* in neurons and glial cells, respectively (Zhang et al., 2014). The *PKM1* and *PKM2* isoforms result from mutually exclusive exon selection. Selective expression of *PKM* isoforms is also critical for regulating glucose metabolism in muscle and cancer (Christofk et al., 2008). Gradual switching of embryonic *PKM2* to adult *PKM1* occurs during mouse brain development and during neuronal differentiation of human mesenchymal stem cells (Su et al., 2017). RBM4 antagonizes PTBP1 activity and hence promotes the *PKM2*-to-*PKM1* switch. Overexpression of RBM4 or *PKM1* increases oxygen consumption and accordingly facilitates neuronal differentiation. These results support the high energy demand of neurons. Because neuroenergetics is dynamic and changes in response to neuronal activity such as glutamatergic stimulation and hypoxia (Bélanger et al., 2011), whether the expression of the splice isoforms of certain synthetic enzymes, including *PKMs*, is coordinately changed remains to be investigated. *PKM* is involved not only in cell metabolism but also in the

modulation of gene expression. PKM2 acts coordinately with β -catenin during gene activation underlying the epithelial-to-mesenchymal transition and thus promotes cell proliferation and tumorigenesis (Yang et al., 2011). A recent report demonstrated that the RNA binding protein Quaking maintains neural stem cell functions during early brain development by preventing the PKM2 switch to PKM1 (Hayakawa-Yano et al., 2017).

Alternative Splicing in Neuronal Migration and Brain Development

The mammalian cerebral cortex has a highly organized six-layered structure consisting of a variety of neuron subtypes (Molyneaux et al., 2007). Positioning of newborn neurons that originate from the ventricular zone and subventricular zone in the embryonic cortical plate occurs in a birth date-dependent “inside-out” manner (Cooper, 2008; Gao and Godbout, 2013). Several signaling cascades regulate neuronal migration in the cortical plate, including the Reelin-Disabled homolog 1 (*Dab1*) pathway (Franco et al., 2011; Gao and Godbout, 2013). Upon binding to the very low density lipoprotein receptor (VLDLR) or apolipoprotein E receptor 2 (ApoER2), Reelin induces differential phosphorylation of the cytosolic adaptor protein *Dab1* and elicits subsequent downstream events that link *Dab1* to the control of neuronal migration. *Reeler* mutant mice and mice with spontaneous or targeted mutations of *Dab1* or either of the receptors exhibit similar phenotypes characterized by ataxia, tremors, and a reeling gait (D’Arcangelo et al., 1995; Howell et al., 1997; Sheldon et al., 1997; Trommsdorff et al., 1999). Differential exon selection of *Dab1* occurs during brain development, resulting in multiple splice isoforms (Gao et al., 2012). *Nova2* suppresses the inclusion of mouse *Dab1* exon 9b/c (Yano et al., 2010). *Nova2* knockout causes neuronal migration defects in both the cerebral cortex and cerebellum due to increasing aberrant exon 9 b/c-containing *Dab1*. Differential selection of exons 7 and 8 of *Dab1* is also intriguing because these two exons encode a domain containing critical tyrosines that are targets of Reelin-mediated phosphorylation. Moreover, ApoER2 also undergoes alternative splicing. The exon 19-containing domain of ApoER2 is important for synapse formation and function via its interaction with PSD-95 (Beffert et al., 2005; Hinrich et al., 2016). Exon 19 inclusion is reduced in the brain of Alzheimer’s patients. It has been shown that SRSF1 inhibits exon 19 inclusion of *ApoER2* and that blocking SRSF1-binding sites using an antisense oligonucleotide has therapeutic potential (Hinrich et al., 2016). Reelin signaling also plays a role in dendritic spine formation and modulates synaptic plasticity in the developing and adult brain (D’Arcangelo, 2014). Therefore, imbalance of splicing factors likely affects neuronal migration and cortical lamination.

Alternative Splicing in Neurologic Functions

Alternative splicing also regulates neurologic functions such as axon guidance and synaptogenesis. A number of neuronal

mRNAs undergo alternative exon selection to generate isoforms in response to neuronal stimulation. Synaptic activity promotes exon 19 inclusion of ApoER2, which then binds Reelin and enhances long-term potentiation (Beffert et al., 2005). Moreover, alternative splicing of the synaptic cell-adhesion molecules neuroligins and neuroligins generates multiple isoforms, and interactions between the various isoforms modify their activity toward glutamatergic and GABA-mediated synaptogenesis. Therefore, alternative splicing can shape the strength and functions of synapses. PTBP2 and Sam68 are involved in splicing regulation of neuroligins (Resnick et al., 2008; Iijima et al., 2011). Notably, Sam68 activity is regulated by depolarization-induced calcium/calmodulin-dependent kinase IV, indicating that neuronal activity controls the diversity of neuroligins via splicing regulation and hence influences synaptic functions (Iijima et al., 2011). Moreover, alternative splicing also regulates the dynamics of neuronal transcriptomes. In pilocarpine-stimulated neurons, exclusion of a cryptic “poison” exon of the sodium channel *Scn9a* mRNA increases the SCN9A level (Eom et al., 2013). A more recent report revealed that neurons can rapidly regulate the expression of several dendritic mRNAs by removing introns that are retained in existing transcripts stored in the nucleus (Mauger et al., 2016). Thus, rapid and signal-responsive splicing regulation is critical for neurological functions.

Perspectives

The combination of various genetic tools and RNA-seq has advanced our knowledge of the impact of alternative splicing on neural development and function. Recently, the use of cell-surface or genetically engineered fluorescent protein markers and fluorescence-activated cell sorting has enabled the isolation of stem/progenitor cells and specific neuronal types (Zhang et al., 2016). Using Cre recombinase-expressing mouse lines, one can manipulate the temporal expression of a splicing regulator or wild-type or disease-related mutant in specific types of neurons and investigate changes in the transcriptome or splicing patterns or isolate target mRNA ribonucleoproteins (Möröy and Heyd, 2007). Single-cell RNA-seq has begun to clarify cell-to-cell transcriptome variability. Since mammalian brains comprise complex and diverse neuronal cell types, to decipher alternative splicing patterns at the single-neuron level still remains challenged. More recently, a single-cell topological data analysis revealed time-series gene expression changes of individual cells throughout murine embryonic stem cell differentiation into motor neurons (Rizvi et al., 2017). With the aid of new technologies, future investigations will paint a more comprehensive picture and define the dynamic scope of how splicing programming determines stem/progenitor cell fate determination and differentiation into the various brain cell types as well as neural circuit development. Emerging in situ sequencing and single-cell fluorescence in situ hybridization strategies (Liu and Trapnell, 2016) may allow revealing topological changes of alternative splicing in a brain network and perhaps unveiling pathological mechanisms at the single-cell level.

AUTHOR CONTRIBUTIONS

C-HS, DD, and W-YT: Jointly wrote this review; W-YT: Defined the scope of the review and edited the draft. All authors read and approved the final manuscript.

REFERENCES

- Anderson, S. L., Coli, R., Daly, I. W., Kichula, E. A., Rock, M. J., Volpi, S. A., et al. (2001). Familial dysautonomia is caused by mutations of the IKAP gene. *Am. J. Hum. Genet.* 68, 753–758. doi: 10.1086/318808
- Atlasi, Y., Mowla, S. J., Ziaee, S. A., Gokhale, P. J., and Andrews, P. W. (2008). OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells* 26, 3068–3074. doi: 10.1634/stemcells.2008-0530
- Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi: 10.1038/nrm.2017.27
- Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D. E., Mott, R., et al. (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *Am. J. Hum. Genet.* 76, 950–966. doi: 10.1086/430454
- Bechara, E. G., Sebestyén, E., Bernardis, I., Eyra, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell.* 52, 720–733. doi: 10.1016/j.molcel.2013.11.010
- Beffert, U., Weeber, E. J., Durudas, A., Qiu, S., Masiulis, I., Sweatt, J. D., et al. (2005). Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. *Neuron* 47, 567–579. doi: 10.1016/j.neuron.2005.07.007
- Bélanger, M., Casey, M., Cormier, M., Filion, A. L., Martin, G., Aubut, S., et al. (2011). Maintenance and decline of physical activity during adolescence: insights from a qualitative study. *Int. J. Behav. Nutr. Phys. Act.* 8, 117. doi: 10.1186/1479-5868-8-117
- 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
- Calarco, J. A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., et al. (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* 138, 898–910. doi: 10.1016/j.cell.2009.06.012
- Camp, J. G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuninger, M., et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15672–15677. doi: 10.1073/pnas.1520760112
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117. doi: 10.1016/j.cell.2008.04.043
- Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., et al. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452, 230–233. doi: 10.1038/nature06734
- Conboy, J. G. (2017). Developmental regulation of RNA processing by Rbfox proteins. *Wiley Interdiscip. Rev. RNA* 8:e1398. doi: 10.1002/wrna.1398
- Cooper, J. A. (2008). A mechanism for inside-out lamination in the neocortex. *Trends Neurosci.* 31, 113–119. doi: 10.1016/j.tins.2007.12.003
- D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I., and Curran, T. (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* 374, 719–723.
- Damianov, A., and Black, D. L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA* 16, 405–416. doi: 10.1261/rna.1838210
- D'Arcangelo, G. (2014). Reelin in the years: controlling neuronal migration and maturation in the mammalian brain. *Adv. Neurosci.* 2014:597395. doi: 10.1155/2014/597395

ACKNOWLEDGMENTS

The work from our laboratory that is reported in this review was supported by the Ministry of Science and Technology grant 106-2311-B-001-015.

- Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., et al. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7285–7290. doi: 10.1073/pnas.1507125112
- Dredge, B. K., Polydorides, A. D., and Darnell, R. B. (2001). The splice of life: alternative splicing and neurological disease. *Nat. Rev. Neurosci.* 2, 43–50. doi: 10.1038/35049061
- Dredge, B. K., Stefani, G., Engelhard, C. C., and Darnell, R. B. (2005). Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* 24, 1608–1620. doi: 10.1038/sj.emboj.7600630
- Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J. L., et al. (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *Elife* 2013, 2. doi: 10.7554/eLife.00178
- Fiszbein, A., Giono, L. E., Quaglini, 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
- Fiszbein, A., and Kornblihtt, A. R. (2016). Histone methylation, alternative splicing and neuronal differentiation. *Neurogenesis* 3:e1204844. doi: 10.1080/23262133.2016.1204844
- Franco, S. J., Martinex-Garay, I., Gil-Sanz, C., Harkins-Perry, S. R., and Müller, U. (2011). Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex. *Neuron* 69, 482–497. doi: 10.1016/j.neuron.2011.01.003
- Gabut, M., Samavarchi-Teheani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H. K., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 147, 132–146. doi: 10.1016/j.cell.2011.08.023
- Gao, Z., and Godbout, R. (2013). Reelin-Disabled-1 signaling in neuronal migration: splicing takes the stage. *Cell. Mol. Life Sci.* 70, 2319–2329. doi: 10.1007/s00018-012-1171-6
- Gao, Z., Poon, H. Y., Li, X., Palmesino, E., Glubrecht, D. D., Colwill, K., et al. (2012). Splice-mediated motif switching regulates disabled-1 phosphorylation and SH2 domain interactions. *Mol. Cell. Biol.* 32, 2794–2808. doi: 10.1128/MCB.00570-12
- Gelman, L. T., Stoilov, P., Maguire, J., Damianov, A., Lin, C. H., Shinu, L., et al. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* 7, 706–711. doi: 10.1038/ng.841
- Grammatikakis, I., Zhang, P., Panda, A. C., Kim, J., Maudsley, S., Adelmohsen, K., et al. (2016). Alternative splicing of neuronal differentiation factor TRF2 regulated by HNRNPH1/H2. *Cell Rep.* 5, 926–934. doi: 10.1016/j.celrep.2016.03.080
- 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
- Hayakawa-Yano, Y., Suyama, S., Nogami, M., Yugami, M., Koya, I., Furukawa, T., et al. (2017). An RNA-binding protein, Qki5, regulates embryonic neural stem cells through pre-mRNA processing in cell adhesion signaling. *Genes Dev.* 31, 1910–1925. doi: 10.1101/gad.300822.117
- Hinrich, A. J., Jodelka, F. M., Chang, J. L., Brutman, D., Bruno, A. M., Briggs, C. A., et al. (2016). Therapeutic correction of ApoER2 splicing in Alzheimer's disease mice using antisense oligonucleotides. *EMBO Mol. Med.* 8, 328–345. doi: 10.15252/emmm.201505846
- Howell, B. W., Hawkes, R., Soriano, P., and Cooper, J. A. (1997). Neuronal position in the developing brain is regulated by mouse disabled-1. *Nature* 389, 733–737. doi: 10.1038/39607
- Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatzer, T., Richard, S., et al. (2011). SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. *Cell* 147, 1601–1614. doi: 10.1016/j.cell.2011.11.028

- Johnson, M. B., Kawasawa, Y. I., Mason, C. E., Krsnik, Z., Coppola, G., Bogdanović, D., et al. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* 62, 494–509. doi: 10.1016/j.neuron.2009.03.027
- Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B., et al. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20333–20338. doi: 10.1073/pnas.0809045105
- Kellner, S., and Kikyo, N. (2010). Transcriptional regulation of the *Oct4* gene, a master gene for pluripotency. *Histol. Histopathol.* 25, 405–412. doi: 10.14670/HH-25.405
- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049–1061. doi: 10.1016/j.cell.2008.02.039
- Kim, K. K., Nam, J., Mukoyama, Y. S., and Kawamoto, S. (2013). Rbfox3-regulated alternative splicing of numb promotes neuronal differentiation during development. *J. Cell Biol.* 200, 443–458. doi: 10.1083/jcb.2012.06146
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892
- Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA* 10, 1489–1498. doi: 10.1261/rna.7100104
- Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165. doi: 10.1038/nrm3525
- Lah, G. J., Li, J. S., and Millard, S. S. (2014). Cell-specific alternative splicing of *Drosophila Dscam2* is crucial for proper neuronal wiring. *Neuron* 83, 1376–1388. doi: 10.1016/j.neuron.2014.08.002
- Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929. doi: 10.1038/nature05676
- Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., et al. (2015). A specific LSD/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell.* 57, 957–970. doi: 10.1016/j.molcel.2015.01.010
- Li, Q., Zheng, S., Han, A., Lin, C. H., Stoilov, P., Fu, X. D., et al. (2014). The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. *Elife* 3:e01201. doi: 10.7554/eLife.01201
- Lin, J. C., Lu, Y. H., Liu, Y. R., and Lin, Y. J. (2016). RBM4a-regulated splicing cascade modulates the differentiation and metabolic activities of brown adipocytes. *Sci. Rep.* 6:20665. doi: 10.1038/srep20665
- Lin, J. C., and Tarn, W. Y. (2011). RBM4 down-regulates PTB and antagonizes its activity in muscle cell-specific alternative splicing. *J. Cell Biol.* 193, 509–520. doi: 10.1083/jcb.201007131
- Lin, Y. S., Wang, H. Y., Huang, D. F., Hsieh, P. F., Lin, M. Y., Chou, C. H., et al. (2016). Neuronal splicing regulator Rbfox3 (NeuN) regulates adult hippocampal neurogenesis and synaptogenesis. *PLoS ONE* 11:e0164164. doi: 10.1371/journal.pone.0164164
- Liu, S., and Trapnell, C. (2016). Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Res.* 5:F1000 Faculty Rev-182. doi: 10.12688/f1000research.7223.1
- Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000. doi: 10.1126/science.1184208
- Magistretti, P. J., and Allaman, I. (2015). A cellular perspective on brain energy metabolism and function imaging. *Neuron* 86, 883–901. doi: 10.1016/j.neuron.2015.03.035
- Makeyev, E. V., Zhang, J., Carrasco, M. A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell.* 27, 435–448. doi: 10.1016/j.molcel.2007.07.015
- Martin, C. L., Duvall, J. A., Ilkin, Y., Simon, J. S., Arreaza, M. G., Wilkes, K., et al. (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 144B, 869–876. doi: 10.1002/ajmg.b.30530
- 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
- Matlin, A. J., Clark, F., and Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398. doi: 10.1038/nrm1645
- Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* 92, 1266–1278. doi: 10.1016/j.neuron.2016.11.032
- Molyneaux, B. J., Arlotta, P., Menezes, J. R., and Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* 8, 427–437. doi: 10.1038/nrn2151
- Möröy, T., and Heyd, F. (2007). The impact of alternative splicing *in vivo*: mouse models show the way. *RNA* 13, 1155–1171. doi: 10.1261/rna.554607
- Mort, M., Sterne-Weiler, T., Li, B., Ball, E. V., Cooper, D. N., Radivojac, P., et al. (2014). MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol.* 15:R19. doi: 10.1186/gb-2014-15-1-r19
- Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909
- Norris, A. D., and Calarco, J. A. (2012). Emerging roles of alternative pre-mRNA splicing regulation in neuronal development and function. *Front. Neurosci.* 6:122. doi: 10.3389/fnins.2012.00122
- Norris, A. D., Gao, S., Norris, M. L., Ray, D., Ramani, A. K., Fraser, A. G., et al. (2014). A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol. Cell.* 54, 946–959. doi: 10.1016/j.molcel.2014.05.004
- Ohnishi, T., Shirane, M., and Nakayama, K. L. (2017). SRRM4-dependent neuron-specific alternative splicing of protruding transcripts regulates neurite outgrowth. *Sci. Rep.* 7:41130. doi: 10.1038/srep41130
- 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
- Pritsker, M., Doniger, T. T., Kramer, L. C., Westcot, S. E., and Lemischka, I. R. (2005). Diversification of stem cell molecular repertoire by alternative splicing. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14290–14295. doi: 10.1073/pnas.0502132102
- Raj, B., and Blencowe, B. J. (2015). Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* 87, 14–27. doi: 10.1016/j.neuron.2015.05.004
- Resnick, M., Segall, A., G. G. R., Lupowitz, Z., and Zisapel, N. (2008). Alternative splicing of neuexins: a role for neuronal polypyrimidine tract binding protein. *Neurosci. Lett.* 439, 235–240. doi: 10.1016/j.neulet.2008.05.034
- Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., et al. (2017). Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* 35, 551–560. doi: 10.1038/nbt.3854
- Rosbach, O., Hung, L. H., Khrameeva, E., Schreiner, S., König, J., Curk, T., et al. (2014). Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* 11, 146–155. doi: 10.4161/rna.27991
- Salomonis, N., Schlieve, C. R., Pereira, L., Wahlquist, C., Colas, A., Zamboni, A. C., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10514–10519. doi: 10.1073/pnas.0912260107
- Sheldon, M., Rice, D. S., D’Arcangelo, G., Yoneshima, H., Nakajima, K., Mikoshiba, K., et al. (1997). Scrambler and yotari disrupt the disabled gene and produce a reeler-like phenotype in mice. *Nature* 389, 730–733.
- Slaugenhaupt, S. A., Blumenfeld, A., Gill, S. P., Leyne, M., Mull, J., Cuajungco, M. P., et al. (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* 68, 598–605. doi: 10.1086/318810
- Su, C. H., Hung, K. Y., Hung, S. C., and Tarn, W. Y. (2017). RBM4 regulates neuronal differentiation of mesenchymal stem cells by modulating alternative splicing of pyruvate kinase M. *Mol. Cell Biol.* 37:e00466-16. doi: 10.1128/MCB.00466-16
- Tarn, W. Y., Kuo, H. C., Yu, H. I., Liu, S. W., Tseng, C. T., Dhananjaya, D., et al. (2016). RBM4 promotes neuronal differentiation and neurite outgrowth by modulating numb isoforms expression. *Mol. Biol. Cell.* 27, 1676–1683. doi: 10.1091/mbc.E15-11-0798
- Trommsdorff, M., Gotthardt, M., Hiesberger, T., Shelton, J., Stockinger, W., Nimpf, J., et al. (1999). Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* 97, 689–701. doi: 10.1016/S0092-8674(00)80782-5

- van Rossum, D., Verheijen, B. M., and Pasterkamp, R. J. (2016). Circular RNAs: novel regulators of neuronal development. *Front. Mol. Neurosci.* 9:74. doi: 10.3389/fnmol.2016.00074
- Vuong, C. K., Black, D. L., and Zheng, S. (2016). The neurogenetics of alternative splicing. *Nat. Rev.* 17, 265–281. doi: 10.1038/nrn.2016.27
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, 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., Zhang, H., Young, A. G., Qiu, R., Argalian, S., Li, X., et al. (2011). Transcriptome analysis of neural progenitor cells by a genetic dual reporter strategy. *Stem Cells* 29, 1589–1600. doi: 10.1002/stem.699
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., et al. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347:1254806. doi: 10.1126/science.1254806
- Yan, Q., Weyn-Vanhenhenryck, S. M., Wu, J., Slona, S. A., Zhang, Y., Chen, K., et al. (2015). Systemic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3445–3450. doi: 10.1073/pnas.1502849112
- Yang, W., Xia, Y., Ji, H., Zheng, Y., Liang, J., Huang, W., et al. (2011). Nuclear PKM2 regulates β -catenin transactivation upon EGFR activation. *Nature* 480, 118–122. doi: 10.1038/nature10598
- Yano, M., Hatajawa-Yano, Y., Mele, A., and Darnell, R. B. (2010). Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron* 66, 848–858. doi: 10.1016/j.neuron.2010.05.007
- Ye, J., and Blelloch, R. (2014). Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* 15, 271–280. doi: 10.1016/j.stem.2014.08.010
- Yeo, G., Holste, D., Kreiman, G., and Burge, C. B. (2004). Variation in alternative splicing across human tissues. *Genome Biol.* 5:R74. doi: 10.1186/gb-2004-5-10-r74
- Yoshida, M., Kataoka, N., Miyauchi, K., Ohe, K., Iida, K., Yoshida, S., et al. (2015). Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc. Natl. Acad. Sci. U.S.A.* 122, 2764–2769. doi: 10.1073/pnas.1415525112
- Zhang, X., Chen, M. H., Wu, X., Kodani, A., Fan, J., Doan, R., et al. (2016). Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* 166, 1147–1162. doi: 10.1016/j.cell.2016.07.025
- Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keefe, S., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947. doi: 10.1523/JNEUROSCI.1860-14.2014
- Zheng, S., and Black, D. L. (2013). Alternative pre-mRNA splicing in neurons: growing up and extending its reach. *Trends Genet.* 29, 442–448. doi: 10.1016/j.tig.2013.04.003
- Zheng, S., Gray, E. E., Chawla, G., Porse, B. T., O'Dell, T. J., and Black, D. L. (2012). PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. *Nat. Neurosci.* 15, 381–388. doi: 10.1038/nn.3026
- Zong, F. Y., Fu, X., Wei, W. J., Luo, Y. G., Heiner, M., Cao, L. J., et al. (2014). The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 10:e1004289. doi: 10.1371/journal.pgen.1004289

Conflict of Interest Statement: 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.

Copyright © 2018 Su, D and Tarn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing Regulator RBM20 and Cardiomyopathy

Takeshi Watanabe^{1,2}, Akinori Kimura^{3,4} and Hidehito Kuroyanagi^{1,4,5*}

¹ Laboratory of Gene Expression, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan,

² Department of Psychosomatic Dentistry, Graduate School of Medical and Dental Science, Tokyo Medical and Dental University (TMDU), Tokyo, Japan, ³ Division of Pathology, Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan, ⁴ Laboratory for Integrated Research Projects on Intractable Diseases Advanced Technology Laboratories, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan, ⁵ Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA, United States

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Ihab Younis,
Carnegie Mellon University in Qatar,
Qatar
Claudia Ghigna,
Istituto di genetica molecolare (IGM),
Italy

*Correspondence:

Hidehito Kuroyanagi
kuroyana.end@tmd.ac.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 02 October 2018

Accepted: 09 November 2018

Published: 28 November 2018

Citation:

Watanabe T, Kimura A and
Kuroyanagi H (2018) Alternative
Splicing Regulator RBM20 and
Cardiomyopathy.
Front. Mol. Biosci. 5:105.
doi: 10.3389/fmolb.2018.00105

RBM20 is a vertebrate-specific RNA-binding protein with two zinc finger (ZnF) domains, one RNA-recognition motif (RRM)-type RNA-binding domain and an arginine/serine (RS)-rich region. *RBM20* has initially been identified as one of dilated cardiomyopathy (DCM)-linked genes. RBM20 is a regulator of heart-specific alternative splicing and *Rbm20*^{ΔRRM} mice lacking the RRM domain are defective in the splicing regulation. The *Rbm20*^{ΔRRM} mice, however, do not exhibit a characteristic DCM-like phenotype such as dilatation of left ventricles or systolic dysfunction. Considering that most of the *RBM20* mutations identified in familial DCM cases were heterozygous missense mutations in an arginine-serine-arginine-serine-proline (RSRSP) stretch whose phosphorylation is crucial for nuclear localization of RBM20, characterization of a knock-in animal model is awaited. One of the major targets for RBM20 is the *TTN* gene, which is comprised of the largest number of exons in mammals. Alternative splicing of the *TTN* gene is exceptionally complicated and RBM20 represses >160 of its consecutive exons, yet detailed mechanisms for such extraordinary regulation are to be elucidated. The *TTN* gene encodes the largest known protein titin, a multi-functional sarcomeric structural protein specific to striated muscles. As titin is the most important factor for passive tension of cardiomyocytes, extensive heart-specific and developmentally regulated alternative splicing of the *TTN* pre-mRNA by RBM20 plays a critical role in passive stiffness and diastolic function of the heart. In disease models with diastolic dysfunctions, the phenotypes were rescued by increasing titin compliance through manipulation of the *Ttn* pre-mRNA splicing, raising RBM20 as a potential therapeutic target.

Keywords: RBM20, dilated cardiomyopathy (DCM), alternative splicing, isoform switching, mutation, arginine/serine (RS)-rich region, titin, nuclear localization

INTRODUCTION

Cardiomyopathy is a myocardial disease with cardiac dysfunction. Cardiomyopathy is roughly classified as genetic cardiomyopathy including hypertrophic cardiomyopathy (HCM), and mixed (genetic and acquired) cardiomyopathy such as dilated cardiomyopathy (DCM; Dadson et al., 2017). HCM is a disease in which hypertrophy of the ventricle occurs despite the absence of high blood pressure or valvular disease that cause ventricular hypertrophy (Elliott, 2014). More than half

of the HCM patients carry mutations in one of eight sarcomere genes (Sabater-Molina et al., 2018). DCM is another common form of cardiomyopathy, affecting ~1 in 250–500 in general population (McKenna et al., 2017) and characterized by left ventricular dilatation and systolic dysfunction in the absence of abnormal loading conditions or coronary artery disease (Rampersaud et al., 2011; McCartan et al., 2012). Mortality rate of DCM is high as a result of heart failure (Kirk et al., 2009). Among the idiopathic DCM cases, 20–35% are familial, with autosomal dominant inheritance in most cases (Kimura, 2016). A next generation sequencing method has recently identified more than 400 potentially causative mutations in 60 genes both in familial and sporadic DCM cases (Pérez-Serra et al., 2016). These DCM-associated genes can be classified into various functional groups such as muscle contraction, Ca^{2+} handling, and nuclear function. Such molecular genetic complexity makes it difficult to elucidate the mechanisms bringing about the common phenotypes of DCM (Hershberger et al., 2013). Among the DCM-linked mutations, 25% were mapped to the *TTN* gene (Herman et al., 2012; Hershberger et al., 2013; Fatkin and Huttner, 2017). The human *TTN* gene has 364 exons, the largest number of exons in a single gene in mammals, 363 of which are coding. The *TTN* gene encodes the largest known protein titin, a multi-functional sarcomeric structural protein specific to striated muscles (Gigli et al., 2016). Titin plays a major role in passive tension of cardiomyocytes (Hidalgo and Granzier, 2013). The *TTN* pre-mRNA undergoes extensive alternative splicing, leading to tissue-specific and developmentally regulated titin isoforms.

RBM20, encoding RNA Binding Motif Protein-20 (RBM20) has initially been identified as one of the DCM-linked genes (Brauch et al., 2009). Genetic abnormalities in *RBM20* have been identified in about 2–3% of familial and sporadic DCM cases (Li et al., 2010; Refaat et al., 2012; Kayvanpour et al., 2017). Recently, *RBM20* has been identified as a crucial RNA-binding protein that controls the splicing of *TTN* (Guo et al., 2012). However, roles of *RBM20* in the pathophysiology of DCM is still unclear. Only a few studies have addressed molecular mechanisms of splicing regulation by *RBM20*, and are controversial. In this review, we summarize the literature on *RBM20* and discuss effects of mutations found in the DCM patients. We also summarize recent attempts to manipulate the *RBM20* functions in various disease models. Finally, we will discuss open questions about the functions of *RBM20* and its relevance to DCM.

STRUCTURE OF RBM20

RBM20 is a vertebrate-specific RNA-binding protein (Zerbino et al., 2018). The human *RBM20* gene resides on chromosome 10 and is composed of 14 exons. Human *RBM20* protein consists of 1,227 amino acid residues and is relatively large for a splicing regulator, yet it has only three conserved recognizable functional domains: two zinc finger (ZnF) domains and one RNA-Recognition Motif (RRM)-type RNA-binding domain (Figure 1). Sequence alignment of *RBM20* proteins from various vertebrate species revealed three other conserved regions (Guo et al., 2012;

Murayama et al., 2018; Zahr and Jaalouk, 2018): a leucine (L)-rich region at the N-terminus, an arginine/serine (RS)-rich region just downstream from the RRM domain and a glutamate (E)-rich region between the RS-rich region and the ZnF2 domain (Figure 1).

Vertebrates have two proteins homologous to *RBM20*; *matrin3*, and *ZNF638* have two RRM domains sandwiched by two ZnF domains and these domains are most related to those of *RBM20* (Coelho et al., 2016). *RBM20* is highly expressed in the heart and the skeletal muscle (Filippello et al., 2013), whereas *matrin3* and *ZNF638* are widely expressed across different cell types (Coelho et al., 2016).

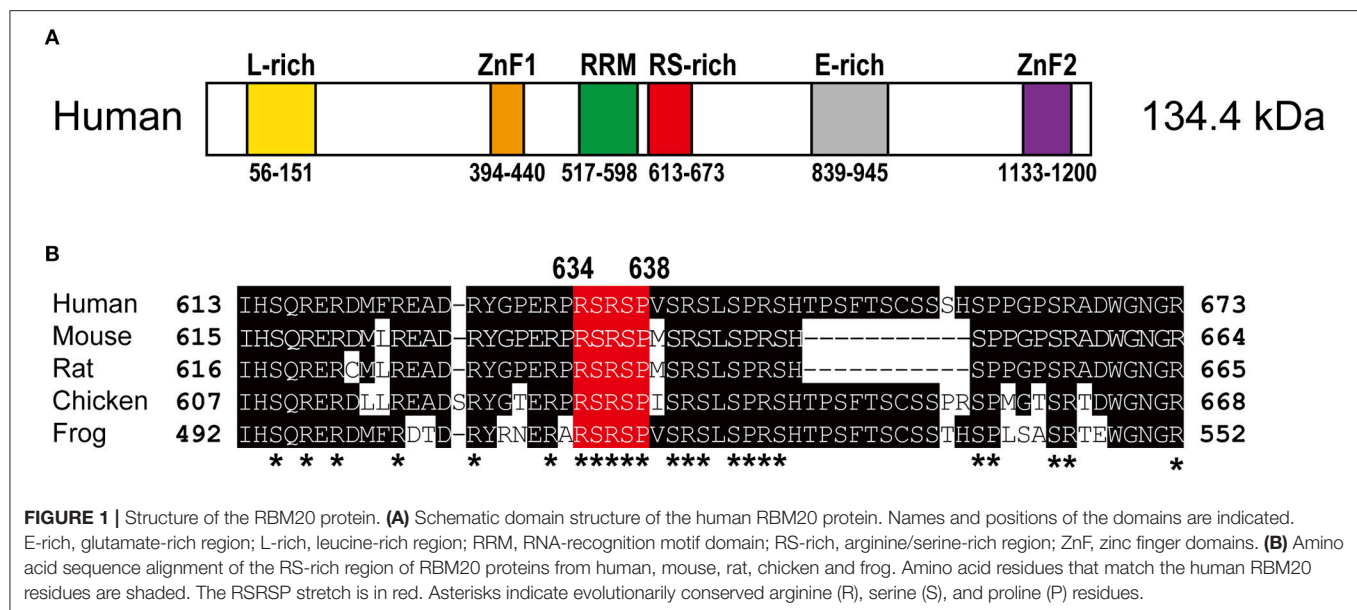
RBM20 MUTATIONS IN DCM PATIENTS

The *RBM20* mutations identified so far in familial as well as sporadic DCM cases are listed in Table 1. The list clearly revealed that almost all of the *RBM20* mutations are heterozygous missense mutations and are enriched in a hot spot composed of an arginine-serine-arginine-serine-proline (RSRSP) stretch at aa 634–638 in the RS-rich region (Table 1; Figure 1B). This situation is unusual considering that most of the missense mutations were mapped to the RRM domains in our previous genetic screening for loss- or reduction-of-function mutants for splicing factors (Kuroyanagi et al., 2006, 2007, 2013). We will discuss later how these mutations would affect the function of *RBM20*.

ALTERNATIVE SPLICING OF THE TTN GENE

Single titin protein spans half of sarcomere, with its N- and C-termini in the Z-disk and the M-band, respectively (Figure 2). It is composed of four structural and functional regions located in Z-disk, I-band, A-band, and M-band (Figure 2A; Wang et al., 1979). Titin is attached to the Z-disk and the thick filament via its Z-disk and A-band segments, respectively (Wang et al., 1979). The I-band region is not attached to any of the solid structures and therefore functions as a molecular spring that generates passive tension when sarcomeres are stretched during diastole (Horowitz et al., 1986). In the elastic I-band region of titin, there are six domains from the N-terminus as follows: proximal immunoglobulin (Ig) repeat domain, N2B-unique element, middle Ig repeat domain, N2A-unique element, proline-glutamate-valine-lysine (PEVK) domain and distal Ig repeat domain (Figure 2A; Labeit et al., 1990; Bang et al., 2001; Lange et al., 2005).

The *TTN* gene can potentially produce an mRNA of more than 100 kb. Deduced from its sequence, the predicted full-length mRNA would produce a protein composed of ~39,000 amino acid residues whose molecular weight (MW) is 4.2 MDa (Bang et al., 2001; Guo and Sun, 2017). However, when SDS-PAGE was performed on rat myocardium samples, six titin isoforms of roughly 3.0 to 3.9 MDa were mainly identified (Zhu and Guo, 2017). While almost all exons encoding the Z-disk, A-band and M-band regions are constitutively included,



many exons encoding the elastic I-band region are alternatively spliced in tissue-specific and developmentally regulated manners (**Figure 2B**). Exons encoding the proximal and distal Ig repeat domains are constitutively included in all isoforms. Exon 49, encoding the N2B-unique element, is a long exon (2,646 bp in human) and is specifically included in the heart but excluded in the skeletal muscle. Exon 50 is a constitutive exon and encodes an Ig domain. Exons 51–101 are alternatively spliced and encode the middle Ig repeat domain. Exons 102–108 and exons 109–224 encode the N2A-unique element and PEVK domain, respectively (Lewinter et al., 2007). Among these exons, alternative splicing of exons encoding the middle Ig repeat domain (exons 51–100) and the PEVK domain (exons 116–218) are exceptionally complicated (Guo et al., 2010). In the myocardium, two major titin isoforms are expressed; the shorter N2B isoform contains only the N2B-unique element among the variable domains; the longer N2BA isoforms including fetal types called fetal cardiac titin (FCT) contain the N2B- and N2A-unique elements and a variable length of the middle Ig repeat domain (**Figure 1**). In the skeletal muscle, the N2A isoform containing all of the variable domains except for the N2B-unique element is expressed (Lewinter et al., 2007; Guo and Sun, 2017; **Figure 2B**).

RBM20 is the major regulator of the heart-specific *TTN* pre-mRNA splicing. In the spontaneous *Rbm20* mutant rat strain lacking a 95-kb region spanning from exons 2–14, titin N2B is no longer expressed while N2BA is predominantly expressed in heterozygotes, and an extraordinarily large isoform N2BA-G is exclusively expressed in homozygotes (Guo et al., 2012); **Figure 2B**. The N2BA isoform is also predominantly expressed in the heart of a DCM patient carrying a heterozygous missense mutation S635A in the *RBM20* gene (Guo et al., 2012). In a human heart-failure cohort, low expression of endogenous RBM20 was correlated with the splicing pattern of the *TTN* gene (Maatz et al., 2014). These reports indicated that splicing control of many of the *TTN* exons is extremely sensitive to the amount of functional RBM20 protein.

RBM20 REGULATES HEART-SPECIFIC ALTERNATIVE SPLICING

High-throughput RNA sequencing (RNA-seq) of cardiac transcriptomes from the *Rbm20*-null rats and human DCM patients with and without mutations in *RBM20* revealed 31 genes whose alternative splicing is RBM20-dependent in both rats and humans (Guo et al., 2012). Crosslinking and immunoprecipitation coupled with RNA-seq (CLIP-seq) experiments of endogenous RBM20 in rat cardiomyocytes identified 80 direct target exons in 18 genes (Maatz et al., 2014). RBM20 predominantly represses cassette exons including those in the *Ttn* and *Ryr2* genes by binding to upstream and/or downstream intron(s) of the target exons (Li et al., 2013; Maatz et al., 2014). Mutually exclusive exons are also enriched among the RBM20 target exons (Guo et al., 2012; Maatz et al., 2014). For instance, RBM20 represses exons 15 and 16 and promotes inclusion of exon 14 in the *Camk2d* gene encoding Ca^{2+} /calmodulin-dependent protein kinase II- δ (CaMKII- δ); RBM20 represses exons 5–7 and promotes inclusion of exon 4 in the *Ldb3* gene encoding Lim domain binding protein 3 (LDB3).

A UCUU core element has been identified as a precise RNA recognition element (RRE) for RBM20 by photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) experiments with epitope-tagged human RBM20 in human embryonic kidney 293 (HEK293) cells and by the CLIP-seq experiments with rat cardiomyocytes (Maatz et al., 2014). In the *Ttn* pre-mRNA, RBM20-binding sites were identified in many of the introns between exon 50 and exon 219 but were almost excluded from the constitutively spliced regions (Maatz et al., 2014). The introns in the alternatively spliced regions were retained in the wild-type rat heart (Li et al., 2013), suggesting that RBM20 represses exons 51–218 by inhibiting excision of most if not all of the introns in this region.

TABLE 1 | *RBM20* mutations identified in DCM patients and their symptoms other than ventricular dilatation.

References	Domain	Mutation	Origin	Type	Effect on RBM20	Splicing regulation ^a	AF	VA	HF or SD
Brauch et al., 2009	RS-rich	R634Q	Familial	Hetero	Unknown	Defective	0/10	1/10	0/10
		R636S	Familial	Hetero	Unknown	Defective	1/13	0/13	2/13
		R636H	Familial	Hetero	Unknown	Unknown	0/2	0/2	0/2
		S637G	Familial	Hetero	Unknown	Defective	0/4	1/4	1/4
		P638L	Familial	Hetero	Unknown	Defective	2/16	6/16	5/16
Li et al., 2010	RRM	V535I	Sporadic ^b	Hetero	Unknown	Unaffected	1/1	0/1	1/1
	RS-rich	R634Q	Sporadic	Hetero	Unknown	Defective	0/1	0/1	1/1
		R634W	Familial	Hetero	Exclusion from nucleus ^c	Defective	0/1	0/1	0/1
		R636C	Familial	Hetero	Unknown	Unknown	1/2	1/2	1/2
		R636H	Familial	Hetero	Unknown	Unknown	0/3	1/3	1/3
	Others	R716Q	Familial ^d	Hetero	Unknown	Defective	0/18	0/18	3/18
Millat et al., 2011	RS-rich	S637G	Familial	Hetero	Unknown	Unknown	Unknown	Unknown	Unknown
Refaat et al., 2012	L-rich	L83I	Sporadic	Hetero	Unknown	Unknown	0/1	0/1	Unknown
	RS-rich	P638L	Sporadic	Hetero	Unknown	Defective	1/1	0/1	Unknown
	E-rich	D888N	Sporadic	Hetero	Unaffected ^c	Unknown	1/1	0/1	Unknown
	Others	S455L	Sporadic	Hetero	Unknown	Unknown	0/1	0/1	Unknown
		R703S	Sporadic	Hetero	Unknown	Unknown	1/1	0/1	Unknown
		G1031X	Sporadic	Hetero	Unaffected ^c	Effective	0/1	0/1	Unknown
		P1081R	Sporadic	Hetero	Unaffected ^c	Unknown	0/1	0/1	Unknown
		E1206K	Sporadic	Hetero	Unaffected ^c	Unknown	0/1	0/1	Unknown
Guo et al., 2012	RS-rich	S635A	Sporadic	Hetero	Exclusion from nucleus ^c	Defective	Unknown	Unknown	Unknown
Wells et al., 2013	RS-rich	R636H	Familial	Hetero	Unknown	Unknown	Unknown	Unknown	Unknown
Chami et al., 2014	RS-rich	R636H	Familial	Hetero	Unknown	Unknown	Unknown	Unknown	Unknown
Zhao et al., 2015	Others	R1182H	Unknown	Hetero	Unknown	Unknown	Unknown	Unknown	Unknown
Beqqali et al., 2016	E-rich	E913K	Familial	Hetero	Protein destabilization	Defective	Unknown	Unknown	4/9
Murayama et al., 2018	RS-rich	R634W	Familial	Hetero	Exclusion from nucleus ^c	Defective	0/2	0/2	1/2
	Others	G1031X	Sporadic ^e	Homo	Unaffected ^c	Effective	1/2	1/2	1/2

Note that residues within the RSRSP stretch in the RS-rich region are in red.

^aEffect of *RBM20* mutations on alternative splicing regulation of the *TTN* gene was assessed by using *TTN* splicing reporters (Guo et al., 2012; Murayama et al., 2018).

^bThe patient also has an A698T mutation in *LDB3*.

^cEffect of *RBM20* mutations on *RBM20* nuclear localization and phosphorylation of the RSRSP stretch was assessed by using mouse cDNAs with equivalent mutations (Murayama et al., 2018).

^dIncomplete penetrance.

^eThe patient is homozygous for the *RBM20* mutation due to uniparental disomy, whereas his mother is heterozygous and asymptomatic.

AF, atrial fibrillation; HF, heart failure; SD, sudden death; VA, ventricular arrhythmias.

THE RSRSP STRETCH IN THE RS-RICH REGION IS CRUCIAL FOR NUCLEAR LOCALIZATION OF RBM20

Recently, it has been reported that both of the two serine residues in the RSRSP stretch are constitutively phosphorylated in cells and that single amino acid substitutions in the stretch disrupted nuclear localization of full-length *RBM20* protein (Murayama et al., 2018; **Figure 3**). Moreover, *Rbm20*^{S637A} knock-in mouse mimicking the S635A mutation showed a remarkable effect on

the titin isoform expression like in the *Rbm20*-null rat strain (Murayama et al., 2018). These findings indicated that the RSRSP stretch is a crucial part of the *RBM20* nuclear localization signal (NLS) and that the DCM-linked mutation in the RSRSP stretch fully disrupted the alternative splicing control by *RBM20*.

Many of splicing factors such as SR-protein family members are known to have RS-rich regions or RS domains consisting of multiple serine-arginine (SR) and arginine-serine (RS) dipeptides (Zahler et al., 1992). The RS domains of the SR proteins are extensively phosphorylated on the serine

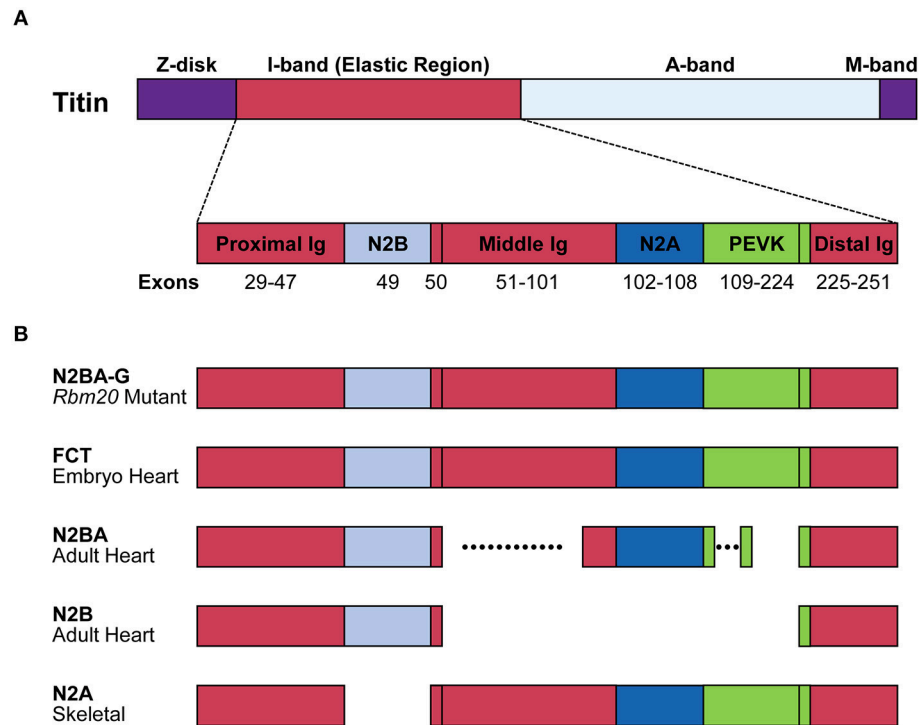


FIGURE 2 | Structure of the titin protein isoforms. **(A)** Schematic domain structure of the titin protein. Names and positions of the domains are indicated. Corresponding exons are indicated below each domain. Distal Ig, distal Ig repeat domain; Middle Ig, middle Ig repeat domain; N2A, N2A-unique element; N2B, N2B-unique element; Proximal Ig, proximal Ig repeat domain. **(B)** Schematic structures of the titin isoforms. Names of the isoforms and the tissues that mainly express the isoforms are indicated on the left. Dotted lines indicate highly variable alternatively spliced regions. FCT, fetal cardiac titin.

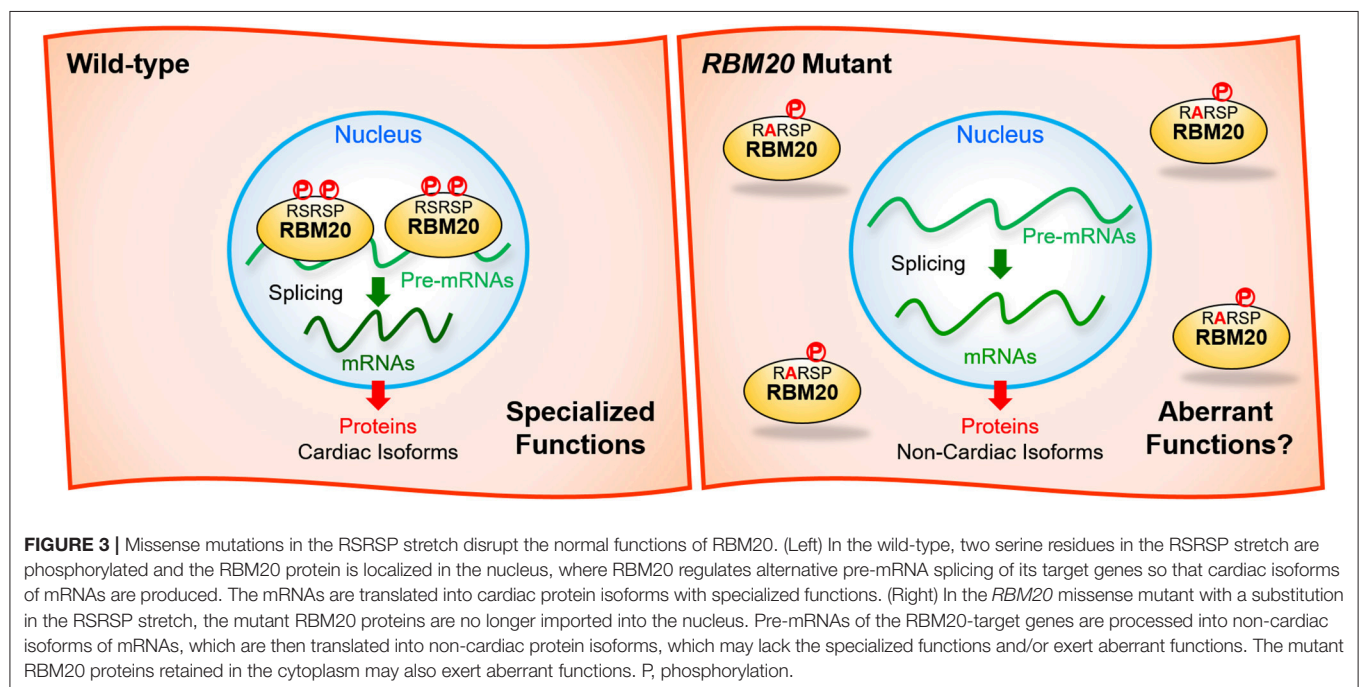


FIGURE 3 | Missense mutations in the RSRSP stretch disrupt the normal functions of RBM20. (Left) In the wild-type, two serine residues in the RSRSP stretch are phosphorylated and the RBM20 protein is localized in the nucleus, where RBM20 regulates alternative pre-mRNA splicing of its target genes so that cardiac isoforms of mRNAs are produced. The mRNAs are translated into cardiac protein isoforms with specialized functions. (Right) In the *RBM20* missense mutant with a substitution in the RSRSP stretch, the mutant RBM20 proteins are no longer imported into the nucleus. Pre-mRNAs of the RBM20-target genes are processed into non-cardiac isoforms of mRNAs, which are then translated into non-cardiac protein isoforms, which may lack the specialized functions and/or exert aberrant functions. The mutant RBM20 proteins retained in the cytoplasm may also exert aberrant functions. P, phosphorylation.

residues and this phosphorylation plays an important role in regulating subcellular localization, protein-protein interaction and splicing regulation activities of the SR

proteins (Xiao and Manley, 1997; Yeakley et al., 1999). It is therefore reasonable to suggest that the RBM20-mediated heart-specific alternative splicing is dynamically regulated

during development and under pathological conditions via dynamic phosphorylation/dephosphorylation of the RSRSP stretch.

Many RBM20-interacting proteins have been identified in HEK293 cells by quantitative stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics experiments and some of these interactions were affected by the S635A mutation (Maatz et al., 2014). Totally distinct subcellular localization of the wild-type and mutant RBM20 proteins (Murayama et al., 2018) might be the major cause of this distinct interactomes. Even with the information about the RBM20 interactome, it is still unclear how the interaction with these proteins leads to repression or switching of its target exons.

THE RRM DOMAIN OF RBM20 IS CRUCIAL FOR SPLICING REGULATION *IN VIVO*

Molecular mechanisms of RBM20-mediated alternative splicing have been analyzed mostly by utilizing *TTN* reporter minigenes expressed in non-cardiac cells. However, the results from such studies utilizing distinct reporter minigenes are controversial as to which of the conserved domains are crucial for the regulation. Mutations in the RSRSP stretch but not in the RRM domain affected repression of the 5' PEVK exons (Guo et al., 2012). The RRM domain and the E-rich region were crucial for repressing exon 242 (Liss et al., 2018). The RSRSP stretch and the E-rich region but not putative RNA-binding domains RRM, ZnF1, or ZnF2 were crucial for repression of a chimeric exon 51/218 (Murayama et al., 2018). Full-length RBM20 (Maatz et al., 2014) or the RRM domain alone (Dauksaite and Gotthardt, 2018) does not necessarily bind to any RNA molecules containing UCUU element(s) in electrophoretic mobility shift assays (EMSAs). Therefore, other RNA element(s) and/or other RNA-binding protein(s) might be involved in the recognition of the authentic target pre-mRNAs by RBM20.

Functions of the RRM domain *in vivo* has been assessed by deleting exons 6 and 7 of the *Rbm20* gene in the mouse. N2BA and N2BA-G isoforms of titin proteins predominated in the left ventricles (LVs) of the *Rbm20*^{ΔRRM} heterozygotes and homozygotes, respectively, and alternative splicing of the *Camk2d* and *Ldb3* genes was evidently affected as in the *Rbm20*-deficient rats (Methawasin et al., 2014), indicating that the RRM domain is crucial for the splicing regulation of these genes *in vivo*. However, the heterozygous or homozygous *Rbm20*^{ΔRRM} mice did not show any significant differences in cardiac chamber geometry and dimensions compared to wild-type controls (Methawasin et al., 2014), suggesting that switching of the titin isoforms to N2BA-G together with the splicing change in *Camk2d* and *Ldb3* *per se* does not cause DCM-like phenotypes such as LV chamber dilatation and systolic dysfunction. This is consistent with that there is not a reported familial case where a missense mutation is mapped to the RRM domain (Table 1). Interestingly, the phenotypes of the *Rbm20*^{ΔRRM} mice are distinct from those of the *Rbm20*-deficient rats (Guo et al., 2012) and *Rbm20* knockout (KO) mice in which exons 4 and 5 are deleted for a frame-shift (van den

Hoogenhof et al., 2018): the heterozygous and homozygous *Rbm20* null rats and mice showed LV dilatation in addition to drastic splicing changes in the *Ttn*, *Camk2d* and *Ldb3* genes. These observations suggest that RBM20(ΔRRM) protein retains some regulatory functions. Elucidation of transcripts differentially affected between the *Rbm20*^{ΔRRM} and *Rbm20* KO mice would lead to identification of genes crucial for the progression of the DCM-like phenotypes in the rodent models.

The only *RBM20* missense mutation outside of the RSRSP stretch in familial DCM cases with complete penetrance was mapped to a highly conserved glutamate (E) residue in the E-rich region (Beqqali et al., 2016; van den Hoogenhof et al., 2018; Table 1). The E913K mutation has been shown to decrease the amount of total RBM20 protein and to affect the *TTN* splicing in the heart of a patient heterozygous for the mutation (Beqqali et al., 2016), suggesting that the E-rich region is crucial for the stability of RBM20. Other missense mutations outside of the RSRSP stretch were identified only in sporadic cases (Table 1) and no experimental evidence of altered splicing has been demonstrated. Therefore, it is unclear whether these mutations affected RBM20 functions and hence caused the DCM phenotypes.

The only *RBM20* non-sense mutation in the DCM patients reported so far is G1031X in sporadic cases (Table 1). This mutation is in exon 11 and likely causes non-sense-mediated mRNA decay (NMD) of mature mRNAs (Schweingruber et al., 2013), leading to haploinsufficiency of *RBM20*. Notably, one of the patients is homozygous for the G1031X mutation due to uniparental disomy, whereas his mother is asymptomatic even with the heterozygous G1031X mutation (Murayama et al., 2018). It is therefore under debate whether heterozygous non-sense mutations in *RBM20* leading to haploinsufficiency would cause the DCM phenotypes.

ANALYSIS OF THE RBM20 FUNCTIONS WITH PLURIPOTENT STEM CELLS

Expression profiling throughout *in vitro* cardiogenesis in embryoid bodies (EBs) derived from mouse embryonic stem cells (mESCs) revealed that *Rbm20* became expressed as early as *Nkx2-5*, a marker for cardiac progenitors, consistent with *Rbm20* induction during *in vivo* cardiogenesis between E7.5 and E8.5 (Beraldi et al., 2014). Even though *Rbm20* is maximally expressed at day 9 of *in vitro* differentiation, the transition of the titin isoforms was apparent at day 24, which was suppressed by *Rbm20* knockdown (Beraldi et al., 2014), suggesting that RBM20-mediated splicing regulation is reproduced in the *in vitro* cardiogenesis.

In vitro differentiation of human induced-pluripotent stem cell (hiPSC)-derived cardiomyocytes (hiPSC-CMs) from familial DCM patients carrying different *RBM20* mutations have been utilized for gene expression profiling during *in vitro* cardiogenesis. Cytological analysis of the hiPSC-CMs revealed that the *RBM20* mutations disorganized sarcomere structures (Wyles et al., 2016b; Streckfuss-Bömeke et al., 2017).

The *RBM20* hiPSC-CMs were defective in Ca^{2+} handling machinery with prolonged Ca^{2+} levels in the cytoplasm and higher Ca^{2+} spike amplitude (Wyles et al., 2016b; Streckfuss-Bömeke et al., 2017), consistent with an increased risk of malignant ventricular arrhythmias in DCM patients with *RBM20* mutations than those with *TTN* mutations (van den Hoogenhof et al., 2018). The *RBM20* hiPSC-CMs have also been utilized for demonstrating their increased susceptibility to β -adrenergic stress and therapeutic rescue by a β -blocker carvedilol and a Ca^{2+} channel blocker verapamil (Wyles et al., 2016a).

REGULATION OF TITIN COMPLIANCE BY RBM20

Titin-based passive tension in the cardiomyocytes occupies a large proportion of the passive stiffness of the whole myocardium (Rivas-Pardo et al., 2016). It has a negative correlation with molecular weight or amino acid sequence length of titin's spring region. For instance, N2B has a shorter elastic region compared to N2BA, thus giving higher passive stiffness to the cardiomyocytes. It is therefore believed that the ratio of the titin isoforms as well as the total amount of titin protein influence the myocardial passive stiffness (Lahmers et al., 2004). The ratio of the N2B isoform to the N2BA isoform varies from species to species (Neagoe et al., 2003) and from ventricles to atria (Fukuda et al., 2003). It is also dynamically regulated during development (Lahmers et al., 2004; Opitz et al., 2004; Warren et al., 2004; Opitz and Linke, 2005; **Figure 2B**) and under pathological conditions. Amount of the N2BA isoform was increased in heart failure with DCM, heart failure with reduced ejection fraction (HFrEF) and chronic ischemic cardiomyopathy (Makarenko et al., 2004; Nagueh et al., 2004; Borbély et al., 2008), and a reduction of the N2BA isoform was observed in diastolic dysfunction resulting from hypertensive heart disease (Warren et al., 2003).

In the *Rbm20* ^{Δ RRM} heterozygote mice, the compliance of titin proteins is increased and diastolic stiffness of a LV chamber is reduced without significant effect on the chamber geometry or dimensions; beneficial effects on diastolic function dominated under conditions of exercise over an unfavorable effect on end-systolic elastance (Methawasin et al., 2014). Splicing changes in the *Ldb3* and *Camk2d* genes in the *Rbm20* ^{Δ RRM} heterozygotes caused minimal effects on the LDB3 isoforms and no apparent effects on phosphorylation of known CaMKII- δ targets (Methawasin et al., 2014). The *Rbm20* ^{Δ RRM} mice have therefore been utilized to prove a concept that increasing the titin compliance is beneficial to disease models where the mice suffer from lowered diastolic stiffness of the hearts (see below).

RBM20 AS A POTENTIAL THERAPEUTIC TARGET

Heart failure with preserved ejection fraction (HFpEF) is a complex syndrome that includes diastolic dysfunction,

exercise intolerance and concentric hypertrophic remodeling. Deletion of *Ttn* exons 251–269, corresponding to the I-band–A-band junction (IAjxn) of titin, increases strain on the spring region and causes an HFpEF-like syndrome in mice (Granzier et al., 2014). Upon constitutive or inducible, heart-specific heterozygous deletion of the *RBM20* RRM domain in the *Ttn* ^{Δ IAjxn} mice, compliant titin isoforms were expressed, diastolic function was normalized, exercise performance was improved and pathological hypertrophy was attenuated (Bull et al., 2016). HFpEF model mice can also be prepared by performing transverse aortic constriction (TAC) surgery with deoxycorticosterone acetate (DOCA) pellet implantation; inducible, heart-specific heterozygous deletion of the RRM domain in this model mice has also been shown to ameliorate diastolic dysfunction and to recover exercise intolerance (Methawasin et al., 2016).

Deletion of *Ttn* exon 49, corresponding to the N2B-unique element, results in small hearts with reduced sarcomere length and increased passive tension leading to diastolic dysfunction in mice (Radke et al., 2007). Heterozygous deletion of the *RBM20* RRM domain from the *Ttn* N2B KO mice restored the cardiac dimension and improved the diastolic function (Hinze et al., 2016).

Deletion of *Ttn* constitutive exons 30–38, corresponding to nine proximal Ig domains in the spring region, increased diastolic stiffness leading to diastolic dysfunction (Chung et al., 2013) and caused mild kyphosis, a phenotype associated with skeletal muscle myopathy (Buck et al., 2014) in mice. *RBM20* was upregulated at the protein level in the *Ttn* IG KO soleus muscle leading to further shortening of titin and heterozygous deletion of the *RBM20* RRM domain from the *Ttn* IG KO mice restored the length of titin in the soleus (Buck et al., 2014).

These examples demonstrated that diastolic dysfunctions could be rescued by increasing titin compliance through manipulation of *Ttn* pre-mRNA splicing in the animal models, and raised the *Ttn* splicing regulator *RBM20* as a potential therapeutic target. Recently, high-throughput screening of small compounds were performed with *RBM20*-sensitive *TTN* splicing reporters and cardenolides have been shown to inhibit *RBM20*-mediated repression of *TTN* exons at least in part by reducing *RBM20* protein level in cultured cells (Liss et al., 2018).

Extracellular and intracellular signals affecting the titin isoform ratios via *RBM20* have also been investigated. Thyroid hormone T3 can promote developmental titin isoform transitions in primary rat cardiomyocytes via the phosphatidylinositol-3-kinase (PI3K)/AKT pathway (Krüger et al., 2008) and the effect of T3 on titin is dependent on *RBM20* (Zhu et al., 2015). Western blot analysis of the cardiomyocytes with anti-*RBM20* antibody detected two bands with apparent molecular weight of 135 kDa and 145 kDa and the amounts of the two isoforms were differentially affected by T3 and/or a PI3K inhibitor, implying that the activity of *RBM20* is regulated by post-translational modification(s) (Zhu et al., 2015). Insulin can also promote the developmental transition of the titin isoforms in primary rat cardiomyocytes by increasing the amount of *RBM20* proteins through the PI3K-Akt-mTOR kinase axis (Zhu et al., 2017).

REGULATION OF CIRCULAR RNA PRODUCTION FROM THE *TTN* GENE BY RBM20

Circular RNAs (circRNAs) can be generated by “back-splicing” of pre-mRNAs (Li et al., 2018) and some of the circRNAs have physiological functions as miRNA sponges *in vivo* (Hansen et al., 2013; Memczak et al., 2013). The *TTN* gene has been shown to produce a variety of circRNAs mostly from the alternatively spliced exons (Li et al., 2013; Khan et al., 2016). Recent RNA-seq experiments identified thousands of circRNAs in the human heart (Khan et al., 2016) and >1,000 circRNAs in the mouse heart (Aufiero et al., 2018). However, only a small subset of circRNAs expressed in the heart are evolutionarily conserved (Aufiero et al., 2018), implying that most of the cardiac circRNAs are non-functional. Forty-three of the human cardiac circRNAs including those from the *TTN* and *CAMK2D* genes were differentially expressed in heart samples from DCM patients compared with those from controls (Khan et al., 2016). Thirty-eight of the mouse cardiac circRNAs were differentially expressed in the *Rbm20* KO mice, 11 of which were generated from the *Ttn* gene in an RBM20-dependent manner (Aufiero et al., 2018). One may therefore say that RBM20 switches from N2BA-G production to circRNA production from the *TTN* gene, although physiological and pathological functions of the circRNAs from the *TTN* and other genes remain to be elucidated.

OPEN QUESTIONS AND FUTURE RESEARCH DIRECTIONS

Although it has been genetically shown that the *RBM20* mutations cause DCM, subsequent cardiac transcriptome analyses and animal models have not yet specified RBM20-regulated genes whose aberrant splicing are critically linked to each of the DCM symptoms such as systolic dysfunction, left ventricle dilatation and a risk of ventricular arrhythmia. RBM20 mutant protein with a missense mutation in the RSRSP stretch may exert aberrant functions in the cytoplasm (Figure 3). Identification of such critical splice variants or aberrant effects may lead to development of new therapeutics for DCM symptoms not restricted to those caused by the *RBM20* mutations. To validate the candidate events, it is necessary to genetically restore cardiac isoforms and/or reduce aberrant isoforms in an appropriate animal model that show DCM-like phenotypes as in the *RBM20*-linked DCM patients. Ventricular

arrhythmia has not been reported for the ΔRRM or KO mouse models, and therefore another animal model that phenocopies the *RBM20*-linked DCM is awaited. A recent large-scale genome-wide association study (GWAS) of >1 million people including 60,620 atrial fibrillation cases have identified *RBM20* as one of genes near risk variants (Nielsen et al., 2018), suggesting its implication in atrial cardiomyopathy.

TTN exons 51–124 and alternative splicing events in *CAMK2D* and *LDB3* are hyper-sensitive to reduction in the amount of functional RBM20 protein, whereas *TTN* exons 125–218 are much less affected by the heterozygous mutations (Guo et al., 2012; Beqqali et al., 2016; Murayama et al., 2018). Biochemical and biophysical analysis of the splicing control of *TTN* by RBM20 should address the following points: (1) how tens of consecutive cassette exons can be synchronously repressed depending on the amount of a single factor RBM20, and (2) how other tens of consecutive cassette exons can be almost completely repressed by less amount of RBM20. Application of recent technical progress in direct sequencing of full-length cDNA/mRNA to the extremely long *TTN* transcripts would elucidate precise splicing patterns of the RBM20-dependent isoforms, which will help understanding the regulation mechanisms.

RBM20 is expressed at early cardiogenesis, yet a compliant titin FCT isoform is expressed in the embryonic heart. RBM20 is also expressed in the skeletal muscle, yet the splicing patterns of the *TTN* mRNAs are totally different between these tissues. So far, it is unknown how the activity of RBM20 is regulated during development and in different tissues. Elucidating such mechanisms will lead to further understanding of heart-specific alternative splicing regulation as well as to future possible therapeutics that manipulate the activity of RBM20.

AUTHOR CONTRIBUTIONS

TW, AK, and HK drafted the manuscript, and revised critically the manuscript for important intellectual content.

FUNDING

The study leading to this article was supported by Grants-in-Aid for Scientific Research (KAKENHI, Grant Numbers JP15H01350, JP17H05596, JP17H03633, and JP15KK0252) from Japan Society for the Promotion of Science (JSPS) (to HK) and a grant from Takeda Science Foundation (to HK).

REFERENCES

- 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
- Bang, M. L., Centner, T., Fornoff, F., Geach, A. J., Gotthardt, M., McNabb, M., et al. (2001). The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ. Res.* 89, 1065–1072. doi: 10.1161/hh2301.100981
- Beqqali, A., Bollen, I. A. E., Rasmussen, T. B., van den Hoogenhof, M. M., van Deutekom, H. W. M., Schafer, S., et al. (2016). A mutation in the glutamate-rich region of RNA-binding motif protein 20 causes dilated cardiomyopathy through missplicing of titin and impaired Frank-Starling mechanism. *Cardiovasc. Res.* 112, 452–463. doi: 10.1093/cvr/cvw192

- Beraldi, R., Li, X., Martinez Fernandez, A., Reyes, S., Secreto, F., Terzic, A., et al. (2014). Rbm20-deficient cardiogenesis reveals early disruption of RNA processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. *Hum. Mol. Genet.* 23, 3779–3791. doi: 10.1093/hmg/ddu091
- Borbély, A., van Heerebeek, L., and Paulus, W. J. (2008). Transcriptional and Posttranslational Modifications of Titin. *Circ. Res.* 104, 12–14. doi: 10.1161/CIRCRESAHA.108.191130
- Brauch, K. M., Karst, M. L., Herron, K. J., de Andrade, M., Pellikka, P. A., Rodeheffer, R. J., et al. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* 54, 930–941. doi: 10.1016/j.jacc.2009.05.038
- Buck, D., Smith, J. E., Chung, C. S., Ono, Y., Sorimachi, H., Labeit, S., et al. (2014). Removal of immunoglobulin-like domains from titin's spring segment alters titin splicing in mouse skeletal muscle and causes myopathy. *J. Gen. Physiol.* 143, 215–230. doi: 10.1085/jgp.201311129
- Bull, M., Methawasin, M., Strom, J., Nair, P., Hutchinson, K., and Granzier, H. (2016). Alternative splicing of titin restores diastolic function in an HFpEF-like genetic murine model (*Ttn*^{ΔIAJxn}) novelty and significance. *Circ. Res.* 119, 764–772. doi: 10.1161/CIRCRESAHA.116.308904
- Chami, N., Tadros, R., Lemarbre, F., Lo, K. S., Beaudoin, M., Robb, L., et al. (2014). Nonsense mutations in BAG3 are associated with early-onset dilated cardiomyopathy in French Canadians. *Can. J. Cardiol.* 30, 1655–1661. doi: 10.1016/j.cjca.2014.09.030
- Chung, C. S., Hutchinson, K. R., Methawasin, M., Saripalli, C., Smith, J. E., Hidalgo, C. G., et al. (2013). Shortening of the elastic tandem immunoglobulin segment of titin leads to diastolic dysfunction. *Circulation* 128, 19–28. doi: 10.1161/CIRCULATIONAHA.112.001268
- Coelho, M. B., Attig, J., Ule, J., and Smith, C. W. J. (2016). Matrin3: connecting gene expression with the nuclear matrix. *Wiley Interdiscip. Rev. RNA* 7, 303–315. doi: 10.1002/wrna.1336
- Dadson, K., Hauck, L., and Billia, F. (2017). Molecular mechanisms in cardiomyopathy. *Clin. Sci.* 131, 1375–1392. doi: 10.1042/CS20160170
- Dauksaite, V., and Gotthardt, M. (2018). Molecular basis of titin exon exclusion by RBM20 and the novel titin splice regulator PTB4. *Nucleic Acids Res.* 46, 5227–5238. doi: 10.1093/nar/gky165
- Elliott, P. M. (2014). 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy. *Eur. Heart J.* 35, 2733–2779. doi: 10.1093/eurheartj/ehu284
- Fatkin, D., and Huttner, I. G. (2017). Titin-truncating mutations in dilated cardiomyopathy. *Curr. Opin. Cardiol.* 32, 232–238. doi: 10.1097/HCO.0000000000000382
- Filippello, A., Lorenzi, P., Bergamo, E., and Romanelli, M. G. (2013). Identification of nuclear retention domains in the RBM20 protein. *FEBS Lett.* 587, 2989–2995. doi: 10.1016/j.febslet.2013.07.018
- Fukuda, N., Wu, Y., Farman, G., Irving, T. C., and Granzier, H. (2003). Titin isoform variance and length dependence of activation in skinned bovine cardiac muscle. *J. Physiol.* 553, 147–154. doi: 10.1113/jphysiol.2003.049759
- Gigli, M., Begay, R. L., Morea, G., Graw, S. L., Sinagra, G., Taylor, M. R. G., et al. (2016). A review of the giant protein titin in clinical molecular diagnostics of cardiomyopathies. *Front. Cardiovasc. Med.* 3:21. doi: 10.3389/fcvm.2016.00021
- Granzier, H. L., Hutchinson, K. R., Tonino, P., Methawasin, M., Li, F. W., Slater, R. E., et al. (2014). Deleting titin's I-band/A-band junction reveals critical roles for titin in biomechanical sensing and cardiac function. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14589–14594. doi: 10.1073/pnas.1411493111
- Guo, W., Bharmal, S. J., Esbona, K., and Greaser, M. L. (2010). Titin diversity–alternative splicing gone wild. *J. Biomed. Biotechnol.* 2010:753675. doi: 10.1155/2010/753675
- Guo, W., Schafer, S., Greaser, M. L., Radke, M. H., Liss, M., Govindarajan, T., et al. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* 18, 766–773. doi: 10.1038/nm.2693
- Guo, W., and Sun, M. (2017). RBM20, a potential target for treatment of cardiomyopathy via titin isoform switching. *Biophys. Rev.* 10, 15–25. doi: 10.1007/s12551-017-0267-5
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993
- Herman, D. S., Lam, L., Taylor, M. R. G., Wang, L., Teekakirikul, P., Christodoulou, D., et al. (2012). Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* 366, 619–628. doi: 10.1056/NEJMoa1110186
- Hershberger, R. E., Hedges, D. J., and Morales, A. (2013). Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* 10, 531–547. doi: 10.1038/nrcardio.2013.105
- Hidalgo, C., and Granzier, H. (2013). Tuning the molecular giant titin through phosphorylation: role in health and disease. *Trends Cardiovasc. Med.* 23, 165–171. doi: 10.1016/j.tcm.2012.10.005
- Hinze, F., Dieterich, C., Radke, M. H., Granzier, H., and Gotthardt, M. (2016). Reducing RBM20 activity improves diastolic dysfunction and cardiac atrophy. *J. Mol. Med.* 94, 1349–1358. doi: 10.1007/s00109-016-1483-3
- Horowitz, R., Kempner, E. S., Bisher, M. E., and Podolsky, R. J. (1986). A physiological role for titin and nebulin in skeletal muscle. *Nature* 323, 160–164. doi: 10.1038/323160a0
- Kayvanpour, E., Sedaghat-Hamedani, F., Amr, A., Lai, A., Haas, J., Holzer, D. B., et al. (2017). Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals. *Clin. Res. Cardiol.* 106, 127–139. doi: 10.1007/s00392-016-1033-6
- Khan, M. A., Reckman, Y. J., Aufiero, S., van den Hoogenhof, M. M., van der Made, I., Beqqali, A., et al. (2016). RBM20 regulates circular RNA production from the titin gene. *Circ. Res.* 119, 996–1003. doi: 10.1161/CIRCRESAHA.116.309568
- Kimura, A. (2016). Molecular genetics and pathogenesis of cardiomyopathy. *J. Hum. Genet.* 61, 41–50. doi: 10.1038/jhg.2015.83
- Kirk, R., Naftel, D., Hoffman, T. M., Almond, C., Boyle, G., Caldwell, R. L., et al. (2009). Outcome of pediatric patients with dilated cardiomyopathy listed for transplant: a multi-institutional study. *J. Heart Lung Transplant.* 28, 1322–1328. doi: 10.1016/j.healun.2009.05.027
- Krüger, M., Sachse, C., Zimmermann, W. H., Eschenhagen, T., Klede, S., and Linke, W. A. (2008). Thyroid hormone regulates developmental titin isoform transitions via the phosphatidylinositol-3-kinase/ AKT pathway. *Circ. Res.* 102, 439–447. doi: 10.1161/CIRCRESAHA.107.162719
- Kuroyanagi, H., Kobayashi, T., Mitani, S., and Hagiwara, M. (2006). Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms *in vivo*. *Nat. Methods* 3, 909–915. doi: 10.1038/nmeth944
- Kuroyanagi, H., Ohno, G., Mitani, S., and Hagiwara, M. (2007). The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing *in vivo*. *Mol. Cell. Biol.* 27, 8612–8621. doi: 10.1128/MCB.01508-07
- Kuroyanagi, H., Watanabe, Y., and Hagiwara, M. (2013). CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the unc-32 gene in neuron-specific manners in *Caenorhabditis elegans*. *PLoS Genet.* 9:e1003337. doi: 10.1371/journal.pgen.1003337
- Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.-L., et al. (1990). A regular pattern of two types of 100-residue motif in the sequence of titin. *Nature* 345, 273–276. doi: 10.1038/345273a0
- Lahmers, S., Wu, Y., Call, D. R., Labeit, S., and Granzier, H. (2004). Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium. *Circ. Res.* 94, 505–513. doi: 10.1161/01.RES.0000115522.52554.86
- Lange, S., Xiang, F., Yakovenko, A., Vihola, A., Hackman, P., Rostkova, E., et al. (2005). The kinase domain of titin controls muscle gene expression and protein turnover. *Science* 308, 1599–1603. doi: 10.1126/science.1110463
- Lewinter, M. M., Wu, Y., Labeit, S., and Granzier, H. (2007). Cardiac titin: Structure, functions and role in disease. *Clin. Chim. Acta* 375, 1–9. doi: 10.1016/j.cca.2006.06.035
- Li, D., Morales, A., Gonzalez-Quintana, J., Norton, N., Siegfried, J. D., Hofmeyer, M., et al. (2010). Identification of novel mutations in RBM20 in patients with dilated cardiomyopathy. *Clin. Transl. Sci.* 3, 90–97. doi: 10.1111/j.1752-8062.2010.00198.x
- Li, S., Guo, W., Dewey, C. N., and Greaser, M. L. (2013). Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res.* 41, 2659–2672. doi: 10.1093/nar/gks1362

- Li, X., Yang, L., and Chen, L.-L. (2018). The biogenesis, functions, and challenges of circular RNAs. *Mol. Cell* 71, 428–442. doi: 10.1016/j.molcel.2018.06.034
- Liss, M., Radke, M. H., Eckhard, J., Neuenschwander, M., Dauksaite, V., von Kries, J.-P., et al. (2018). Drug discovery with an RBM20 dependent titin splice reporter identifies cardenolides as lead structures to improve cardiac filling. *PLoS ONE* 13:e0198492. doi: 10.1371/journal.pone.0198492
- Maatz, H., Jens, M., Schafer, S., Heinig, M., Kirchner, M., et al. (2014). RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J. Clin. Invest.* 124, 3419–3430. doi: 10.1172/JCI74523
- Makarenko, I., Opitz, C. A., Leake, M. C., Neagoe, C., Kulke, M., Gwathmey, J. K., et al. (2004). Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. *Circ. Res.* 95, 708–716. doi: 10.1161/01.RES.0000143901.37063.2f
- McCartan, C., Mason, R., Jayasinghe, S. R., and Griffiths, L. R. (2012). Cardiomyopathy classification: ongoing debate in the genomics era. *Biochem. Res. Int.* 2012, 1–10. doi: 10.1155/2012/796926
- McKenna, W. J., Maron, B. J., and Thiene, G. (2017). Classification, epidemiology, and global burden of cardiomyopathies. *Circ. Res.* 121, 722–730. doi: 10.1161/CIRCRESAHA.117.309711
- Memczak, S., Jens, M., Elefantioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928
- Methawasin, M., Hutchinson, K. R., Lee, E.-J., Smith, J. E., Saripalli, C., Hidalgo, C. G., et al. (2014). Experimentally increasing titin compliance in a novel mouse model attenuates the frank-starling mechanism but has a beneficial effect on diastole. *Circulation* 129, 1924–1936. doi: 10.1161/CIRCULATIONAHA.113.005610
- Methawasin, M., Strom, J. G., Slater, R. E., Fernandez, V., Saripalli, C., and Granzier, H. (2016). Experimentally increasing the compliance of titin through RNA binding Motif-20 (RBM20) inhibition improves diastolic function in a mouse model of heart failure with preserved ejection fraction clinical perspective. *Circulation* 134, 1085–1099. doi: 10.1161/CIRCULATIONAHA.116.023003
- Millat, G., Bouvagnet, P., Chevalier, P., Sebbag, L., Dulac, A., Dauphin, C., et al. (2011). Clinical and mutational spectrum in a cohort of 105 unrelated patients with dilated cardiomyopathy. *Eur. J. Med. Genet.* 54, e570–e575. doi: 10.1016/j.ejmg.2011.07.005
- Murayama, R., Kimura, M., Togo-ohno, M., and Yamasaki, Y. (2018). Phosphorylation of the RSRP stretch is critical for splicing regulation by RNA-Binding Motif Protein 20 (RBM20) through nuclear localization. *Sci. Rep.* 8:8970. doi: 10.1038/s41598-018-26624-w
- Nagueh, S. F., Shah, G., Wu, Y., Torre-Amione, G., King, N. M. P., Lahmers, S., et al. (2004). Altered titin expression, myocardial stiffness, and left ventricular function in patients with dilated cardiomyopathy. *Circulation* 110, 155–162. doi: 10.1161/01.CIR.0000135591.37759.AF
- Neagoe, C., Opitz, C. A., Makarenko, I., and Linke, W. A. (2003). Gigantic variety: expression patterns of titin isoforms in striated muscles and consequences for myofibrillar passive stiffness. *J. Muscle Res. Cell Motil.* 24, 175–189. doi: 10.1023/A:1026053530766
- Nielsen, J. B., Thorolfsson RB, Fritsche LG, Zhou W., Skov MW., Graham SE., et al. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* 50, 1234–1239. doi: 10.1038/s41588-018-0171-3
- Opitz, C. A., Leake, M. C., Makarenko, I., Benes, V., and Linke, W. A. (2004). Developmentally regulated switching of titin size alters myofibrillar stiffness in the perinatal heart. *Circ. Res.* 94, 967–975. doi: 10.1161/01.RES.0000124301.48193.E1
- Opitz, C. A., and Linke, W. A. (2005). Plasticity of cardiac titin/connectin in heart development. *J. Muscle Res. Cell Motil.* 26, 333–342. doi: 10.1007/s10974-005-9040-7
- Pérez-Serra, A., Toro, R., Sarquella-Brugada, G., De Gonzalo-Calvo, D., Cesar, S., Carro, E., et al. (2016). Genetic basis of dilated cardiomyopathy. *Int. J. Cardiol.* 224, 461–472. doi: 10.1016/j.ijcard.2016.09.068
- Radke, M. H., Peng, J., Wu, Y., McNabb, M., Nelson, O. L., Granzier, H., et al. (2007). Targeted deletion of titin N2B region leads to diastolic dysfunction and cardiac atrophy. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3444–3449. doi: 10.1073/pnas.0608543104
- Rampersaud, E., Siegfried, J. D., Norton, N., Li, D., Martin, E., and Hershberger, R. E. (2011). Rare variant mutations identified in pediatric patients with dilated cardiomyopathy. *Prog. Pediatr. Cardiol.* 31, 39–47. doi: 10.1016/j.pppedcard.2010.11.008
- Refaat, M. M., Lubitz, S. A., Makino, S., Islam, Z., Michael Frangiskakis, J., Mehdi, H., et al. (2012). Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *HRTHM* 9, 390–396. doi: 10.1016/j.hrthm.2011.10.016
- Rivas-Pardo, J. A., Eckels, E. C., Popa, I., Kosuri, P., Linke, W. A., and Fernández, J. M. (2016). Work done by titin protein folding assists muscle contraction. *Cell Rep.* 14, 1339–1347. doi: 10.1016/j.celrep.2016.01.025
- Sabater-Molina, M., Pérez-Sánchez, I., Hernández del Rincón, J. P., and Gimeno, J. R. (2018). Genetics of hypertrophic cardiomyopathy: a review of current state. *Clin. Genet.* 93, 3–14. doi: 10.1111/cge.13027
- Schweingruber, C., Rufener, S. C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay — mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim. Biophys. Acta Gene Regul. Mech.* 1829, 612–623. doi: 10.1016/j.bbagr.2013.02.005
- Streckfuss-Bömeke, K., Tiburcy, M., Fomin, A., Luo, X., Li, W., Fischer, C., et al. (2017). Severe DCM phenotype of patient harboring RBM20 mutation S635A can be modeled by patient-specific induced pluripotent stem cell-derived cardiomyocytes. *J. Mol. Cell. Cardiol.* 113, 9–21. doi: 10.1016/j.jmcc.2017.09.008
- van den Hoogenhof, M. M. G., Beqqali, A., Amin, A. S., van der Made, I., Aufiero, S., Khan, M. A. F., et al. (2018). RBM20 mutations induce an arrhythmogenic dilated cardiomyopathy related to disturbed calcium handling. *Circulation* 138, 1330–1342. doi: 10.1161/CIRCULATIONAHA.117.031947
- Wang, K., McClure, J., and Tu, A. (1979). Titin: major myofibrillar components of striated muscle. *Proc. Natl. Acad. Sci. U.S.A.* 76, 3698–3702. doi: 10.1073/pnas.76.8.3698
- Warren, C. M., Jordan, M. C., Roos, K. P., Krzesinski, P. R., and Greaser, M. L. (2003). Titin isoform expression in normal and hypertensive myocardium. *Cardiovasc. Res.* 59, 86–94. doi: 10.1016/S0008-6363(03)00328-6
- Warren, C. M., Krzesinski, P. R., Campbell, K. S., Moss, R. L., and Greaser, M. L. (2004). Titin isoform changes in rat myocardium during development. *Mech. Dev.* 121, 1301–1312. doi: 10.1016/j.mod.2004.07.003
- Wells, Q. S., Becker, J. R., Su, Y. R., Mosley, J. D., Weeke, P., D'Aoust, L., et al. (2013). Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* 6, 317–326. doi: 10.1161/CIRCGENETICS.113.000011
- Wyles, S., Hrstka, S., Reyes, S., Terzic, A., Olson, T., and Nelson, T. (2016a). Pharmacological modulation of calcium homeostasis in familial dilated cardiomyopathy: an *in vitro* analysis from an RBM20 patient-derived iPSC model. *Clin. Transl. Sci.* 9, 158–167. doi: 10.1111/cts.12393
- Wyles, S. P., Li, X., Hrstka, S. C., Reyes, S., Oommen, S., Beraldi, R., et al. (2016b). Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Hum. Mol. Genet.* 25, 254–265. doi: 10.1093/hmg/ddv468
- Xiao, S. H., and Manley, J. L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.* 11, 334–344. doi: 10.1101/gad.11.3.334
- Yeakley, J. M., Tronchère, H., Olesen, J., Dyck, J. A., Wang, H. Y., and Fu, X. D. (1999). Phosphorylation regulates *in vivo* interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J. Cell Biol.* 145, 447–455. doi: 10.1083/jcb.145.3.447
- Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* 6, 837–847. doi: 10.1101/gad.6.5.837
- Zahr, H. C., and Jaalouk, D. E. (2018). Exploring the crosstalk between LMNA and splicing machinery gene mutations in dilated cardiomyopathy. *Front. Genet.* 9:231. doi: 10.3389/fgene.2018.00231
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46, D754–D761. doi: 10.1093/nar/gkx1098
- Zhao, Y., Feng, Y., Zhang, Y.-M., Ding, X.-X., Song, Y.-Z., Zhang, A.-M., et al. (2015). Targeted next-generation sequencing of candidate genes reveals novel mutations in patients with dilated cardiomyopathy. *Int. J. Mol. Med.* 36, 1479–1486. doi: 10.3892/ijmm.2015.2361

- Zhu, C., and Guo, W. (2017). Detection and quantification of the giant protein titin by SDS-agarose gel electrophoresis. *MethodsX* 4, 320–327. doi: 10.1016/j.mex.2017.09.007
- Zhu, C., Yin, Z., Ren, J., McCormick, R. J., Ford, S. P., and Guo, W. (2015). RBM20 is an essential factor for thyroid hormone-regulated titin isoform transition. *J. Mol. Cell Biol.* 7, 88–90. doi: 10.1093/jmcb/mjv002
- Zhu, C., Yin, Z., Tan, B., and Guo, W. (2017). Insulin regulates titin pre-mRNA splicing through the PI3K-Akt-mTOR kinase axis in a RBM20-dependent manner. *Biochim. Biophys. Acta Mol. Basis Dis.* 1863, 2363–2371. doi: 10.1016/j.bbdis.2017.06.023

Conflict of Interest Statement: 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.

Copyright © 2018 Watanabe, Kimura and Kuroyanagi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Importance of Functional Loss of FUS in FTL/ALS

Shinsuke Ishigaki^{1,2*} and Gen Sobue^{3,4*}

¹ Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan, ² Department of Therapeutics for Intractable Neurological Disorders, Nagoya University Graduate School of Medicine, Nagoya, Japan, ³ Brain and Mind Research Center, Nagoya University, Nagoya, Japan, ⁴ Research Division of Dementia and Neurodegenerative Disease, Nagoya University Graduate School of Medicine, Nagoya, Japan

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Hitomi Tsuiji,
Nagoya City University, Japan
Woan-Yuh Tarn,
Academia Sinica, Taiwan

*Correspondence:

Shinsuke Ishigaki
ishigaki-ns@umin.net
Gen Sobue
sobueg@med.nagoya-u.ac.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 28 February 2018

Accepted: 17 April 2018

Published: 03 May 2018

Citation:

Ishigaki S and Sobue G (2018)
Importance of Functional Loss of FUS
in FTL/ALS. *Front. Mol. Biosci.* 5:44.
doi: 10.3389/fmolb.2018.00044

Fused in sarcoma (FUS) is an RNA binding protein that regulates RNA metabolism including alternative splicing, transcription, and RNA transportation. FUS is genetically and pathologically involved in frontotemporal lobar degeneration (FTLD)/amyotrophic lateral sclerosis (ALS). Multiple lines of evidence across diverse models suggest that functional loss of FUS can lead to neuronal dysfunction and/or neuronal cell death. Loss of FUS in the nucleus can impair alternative splicing and/or transcription, whereas dysfunction of FUS in the cytoplasm, especially in the dendritic spines of neurons, can cause mRNA destabilization. Alternative splicing of the *MAPT* gene at exon 10, which generates 4-repeat Tau (4R-Tau) and 3-repeat Tau (3R-Tau), is one of the most impactful targets regulated by FUS. Additionally, loss of FUS function can affect dendritic spine maturation by destabilizing mRNAs such as Glutamate receptor 1 (GluA1), a major AMPA receptor, and Synaptic Ras GTPase-activating protein 1 (SynGAP1). Moreover, FUS is involved in axonal transport and morphological maintenance of neurons. These findings indicate that a biological link between loss of FUS function, Tau isoform alteration, aberrant post-synaptic function, and phenotypic expression might lead to the sequential cascade culminating in FTL. Thus, to facilitate development of early disease markers and/or therapeutic targets of FTL/ALS it is critical that the functions of FUS and its downstream pathways are unraveled.

Keywords: FUS, FTL/ALS, tau, GluA1, SynGAP

INTRODUCTION

Amyotrophic lateral sclerosis (ALS), characterized by selective motor neuronal loss in the central nervous system, and frontotemporal lobar degeneration (FTLD), which is distinguished by changes in character, abnormal behaviors, language impairments, and progressive dementia, have recently been recognized as two ends of the spectrum of one disease (Robberecht and Philips, 2013). This notion is supported by the genetic determinants underlying familial FTL/ALS (Renton et al., 2014) and lines of evidence showing a pathological continuity between ALS and FTL (Riku et al., 2014). RNA binding proteins (RBPs) such as transactive response (TAR) DNA-binding protein 43 (TDP-43), and fused in sarcoma (FUS) genetically and pathologically link the two neurodegenerative diseases to a single disease state (Van Langenhove et al., 2012). These genes are causative for familial ALS and FTL, and are pathological hallmarks of both familial and

sporadic FTL/ALS in which TDP-43 or FUS-positive inclusions are observed (Kwiatkowski et al., 2009; Lagier-Tourenne and Cleveland, 2009; Vance et al., 2009; Mackenzie et al., 2011; Strong and Volkening, 2011). Additionally, FTL has also been classified as a tauopathy characterized by an accumulation of phosphorylated microtubule-associated protein tau (Tau) in affected neurons (Seelaar et al., 2011).

FUS was originally identified as a fusion protein that resulted from a chromosomal translocation in human myxoid liposarcomas, in which the N-terminal portion of FUS was translocated and fused to the transcription factor CHOP (Croizat et al., 1993). FUS functions as a regulator of multiple aspects of RNA metabolism, including transcription, alternative splicing, and mRNA transport, as well as DNA damage regulation (Bertolotti et al., 1996; Wang et al., 2008; Schwartz et al., 2012; Tan et al., 2012). Whole-body knockout of FUS in highly homogenous inbred C57B6 strain mice resulted in early neonatal death due to immune system defects (Hicks et al., 2000), whereas FUS KO in outbred mice had no developmental impairments (Kuroda et al., 2000). These suggest that the inbred background lacks genes required to compensate for the FUS KO effects. Similar to TDP-43 pathology, FUS-related FTL/ALS pathology is characterized by mislocalization of FUS to the cytoplasm and a concomitant reduction in nuclear expression in affected neurons (Neumann et al., 2009; Deng et al., 2010; Mackenzie et al., 2011). Redistribution of FUS from the nucleus to the cytoplasm implies that the loss of nuclear FUS is causal for FUS-associated ALS/FTL. Indeed, loss of FUS leads to neuronal cell death in *Drosophila* and zebrafish (Kabashi et al., 2011; Wang et al., 2011). On the other hand, accumulation of FUS in the cytoplasm is strongly associated with stress granules, which are non-membranous, cytoplasmic ribonucleoprotein (RNP) granules composed of mRNAs, translation initiation factors, ribosomes, and other RBPs. These granules are induced by various cellular stresses, such as oxidative stress, glucose starvation, mitochondrial dysfunction, and viral infection that inhibit translation initiation. The stress granule associated gain-of-toxicity hypothesis of FUS has been well reviewed elsewhere (Gao et al., 2017).

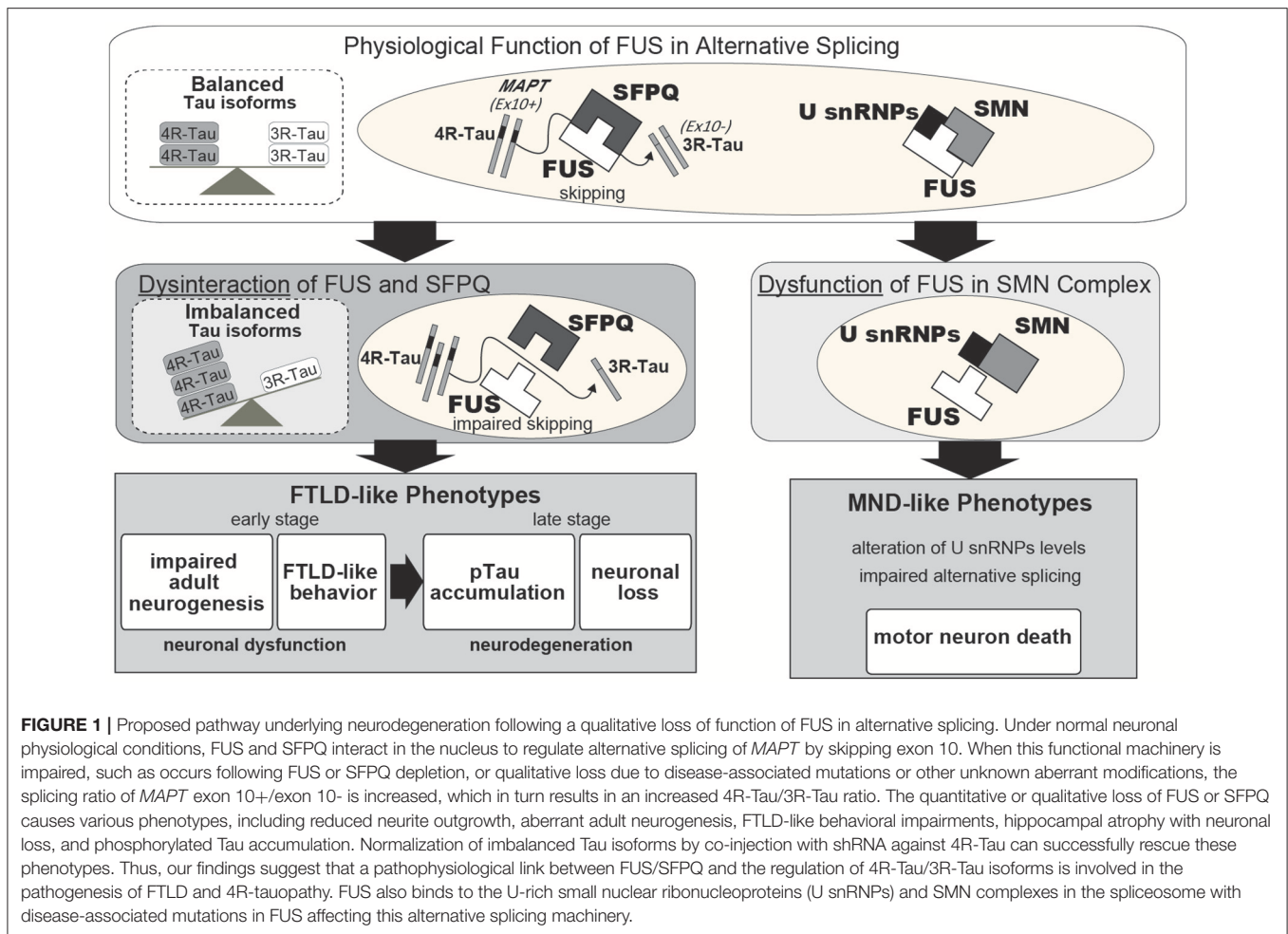
This review provides an overview of recent findings that reveal the effects of functional loss of FUS on the pathogenesis of FTL/ALS. First, loss of FUS in the nucleus leads to imbalanced Tau isoforms due to insufficient skipping of exon 10 in the *MAPT* gene. Second, loss of FUS in the cytoplasm causes decreased stability in GluA1 and SynGAP α 2 mRNA resulting in aberrant maturation of dendritic spines. In addition, we summarize the roles of FUS in neurite maintenance and axonal transport, and provide a briefly overview of the FUS liquid-phase-transition, which may alter its various physiological functions and contribute to the development of toxic cellular effects under pathological conditions. Thus, the functional properties of FUS may influence multiple cellular processes of neurons and/or glial cells whose dysfunction could be the most plausible explanation for neuronal toxicity mediated by loss of FUS.

QUANTITATIVE AND QUALITATIVE LOSS OF FUNCTION OF FUS

Although recent reports have suggested that loss-of-FUS-function in motor neurons might not contribute to motor neuron degeneration in ALS (Scekic-Zahirovic et al., 2016; Sharma et al., 2016), lines of evidence suggest that loss-of-FUS-function in cerebral neurons can contribute to neuronal dysfunction and neurodegeneration in FTL. FUS-deficient mice generated either via silencing or FUS knock-out exhibit behavioral impairments (Kino et al., 2015; Udagawa et al., 2015). However, recovery of wild-type FUS in the FUS-silenced mice rescued the behavioral phenotypes, whereas a disease-associated mutant did not (Ishigaki et al., 2017).

Although FUS pathology is detected in both ALS and FTL cases, the majority of disease-causing mutations within FUS are associated with ALS cases. Nevertheless, a subset of familial and sporadic ALS cases involving FUS gene mutations have been shown to have cognitive dysfunction or mental retardation (Bäumer et al., 2010; Huang et al., 2010; Yan et al., 2010; Belzil et al., 2012; Yamashita et al., 2012). Moreover, a spectrum of cognitive impairments have been observed in a considerable subpopulation of ALS patients (Swinnen and Robberecht, 2014). Taken together, the clinical data and FUS-silenced mice model findings support the hypothesis that FUS dysfunction results in early cognitive impairments.

In familial and sporadic FTL/ALS cases, which are, respectively, characterized by mutations in the FUS coding sequence or the presence of a basophilic inclusion body (BIBD), the affected motor neurons exhibit dislocation of FUS with the protein accumulating in the cytoplasm rather than the nucleus. Cytoplasmic mislocalization of FUS is presumably the first step in the disease cascade; therefore, quantitative loss-of-FUS is thought to be causal for FTL/ALS. However, disease-associated mutations do not trigger complete mislocalization of FUS to the cytoplasm as a moderate amount of the protein remains localized in the nucleus (Kino et al., 2011). This implies that the FUS mutants are non-functional, and that this, in conjunction with the quantitative reduction in protein, culminates in neuronal dysfunction and FTL/ALS pathophysiology. It has been reported that FUS binds U-rich small nuclear ribonucleoproteins (snRNPs) and the SMN complex, which is the machinery for snRNP biogenesis, and hence compromises precursor mRNA splicing, leading to FUS-associated FTL/ALS (Tsuiji et al., 2013; Sun et al., 2015). In our recent study, the presence of disease-associated mutations in FUS disrupted formation of a high molecular weight FUS complex by impeding interactions with a second protein, Splicing factor, proline- and glutamine-rich (SFPQ). The impaired FUS functionality suggests that the pathophysiological features of FTL/ALS also arise from qualitative losses in FUS and SFPQ (Ishigaki et al., 2017; **Figure 1**). Another group recently reported on the presence of possible SFPQ mutations in familial ALS cases (Thomas-Jinu et al., 2017). These findings suggest that aberrant interactions between FUS and its spliceosome binding partners in the nucleus of neurons might lead to neuronal dysfunction and



subsequent neurodegeneration. However, future pathological studies examining the FUS/SFPQ nuclear interaction in both FTLD/ALS and tauopathies are necessary.

The cell selectivity of neurodegenerative diseases such as FTLD/ALS has remained a mystery. For instance, the pathology of FTLD/ALS in both motor neurons and cortical neurons involves major selective neuronal vulnerability. Glial cells, such as astrocytes and microglial cells, have also been linked with FTLD/ALS as modifiers of the non-cell-autonomous mechanism of disease pathogenesis, while cerebellar neurons are typically spared in FTLD/ALS (Boill  e et al., 2006; Yamanaka et al., 2008a,b). The cell and region specific-selectivity of FTLD/ALS, however, cannot be explained by the expression pattern of FUS since it is expressed ubiquitously throughout the CNS (Kwiatkowski et al., 2009). We previously found that the profiles of FUS-mediated gene expression and alternative splicing in motor neurons are similar to those of cortical neurons, but differ from those in cerebellar neurons despite the similarity of their innate transcriptome signature. The gene expression profiles in glial cells were similar to those in motor and cortical neurons. Given that motor and cortical neurons are the major affected cell-types in FTLD/ALS, whereas glial cells are modifiers and cerebellar neurons are spared, it is possible that

the FUS-regulated transcriptome profiles in each cell-type may determine the cellular fate in association with FTLD/ALS and that neuron-glia interactions may be involved in the pathogenesis (Fujioka et al., 2013). Indeed, FUS silencing caused glial cell proliferation in the brain of non-human primates (Endo et al., 2017).

Taken together, these findings indicate that both quantitative and qualitative losses of FUS function are likely involved in the pathogenesis of FTLD/ALS, and should provide clues for therapeutics that clarify the functional properties of FUS.

FUS FUNCTION IN THE NUCLEUS: REGULATION OF ALTERNATIVE SPLICING AND TRANSCRIPTION

Since FUS plays a role in multiple aspects of RNA metabolism, transcriptome deterioration could be the most plausible explanation for neuronal toxicity mediated by loss-of-FUS. In support of this, numerous neuronal function-associated molecules in FUS-regulated transcriptome profiles have been identified (Ishigaki et al., 2012; Lagier-Tourenne et al., 2012; Rogelj et al., 2012; Fujioka et al., 2013; Honda et al., 2013;

Nakaya et al., 2013). The alteration of gene expression and/or alternative splicing of these genes may have a large impact on neuronal function which contributes to the neurodegeneration observed in FTL/ALS. We speculate that disruptions to FUS functionality could result in a partial effect rather than fatal damage by altering isoforms or expression levels of these genes. Thus, it is possible that neurodegeneration only results after the transcriptional disruption triggered by loss of FUS functionality reaches a critical threshold such that the expression of individual genes and alternative splicing events are not critical by themselves.

To gain a better understanding of this mechanism it is necessary to narrow down the list of FUS-regulated genes to those most likely to be disease-associated. Among the genes, alternative splicing of *MAPT* exon 10 has been shown to be relevant to FTL/ALS pathogenesis (Orozco et al., 2012; Ishigaki et al., 2017). *MAPT* encodes the Tau protein, a microtubule-binding protein in which aberrant accumulation of the phosphorylated form in affected neurons causes tauopathies, such as Alzheimer's disease and FTL. It has been reported that the ratio of 4-repeat Tau (4R-Tau)/3-repeat Tau (3R-Tau) is high in tauopathies, including FTL, progressive supranuclear palsy (PSP), and corticobasal degeneration (CBD) (Hong et al., 1998; Yoshida, 2006; Umeda et al., 2013). Our previous study demonstrated that the intranuclear FUS/SFPQ complex regulates alternative splicing of *MAPT* exon 10, which generates two Tau isoforms harboring either three or four microtubule-binding repeats (3R-Tau and 4R-Tau), respectively (Ishigaki et al., 2017). FUS- or SFPQ-silenced mice exhibit an increase in 4R-Tau leading to FTL-like behavior, reduced adult neurogenesis, phosphorylated Tau accumulation, and neuronal loss (Figure 1; Ishigaki et al., 2017). These findings suggest that the impaired Tau isoform ratio generated in response to dysregulation of alternative splicing by the aberrant FUS-SFPQ complex could be an early pathogenic factor for FTL/ALS and tauopathies. A report of familial FTL characterized by a Q140H substitution in FUS with accompanying abnormal Tau isoform ratios supports the idea (Ferrer et al., 2015).

Additional targets of FUS-mediated exon skipping could likewise contribute to FTL/ALS pathogenesis. Among these genes is FUS itself in which FUS-mediated splicing at exon 7 contributes to autoregulation of expression with the exon 7 skipped variant undergoing nonsense-mediated decay (NMD). The auto regulatory function is deficient in ALS-associated FUS mutants (Zhou et al., 2013).

Other FUS-regulated genes, such as *NTNG1* or *BRAF*, which could be important for neuronal cell survival, have been identified in multiple reports (Orozco and Edbauer, 2013). Further study is necessary to evaluate their significance in FTL/ALS pathogenesis.

FUNCTION OF FUS IN THE DENDRITIC SPINE: MRNA STABILIZATION

While FUS is enriched in the nucleus, a percentage of the protein is localized to the soma and neuronal processes (Fujii and

Takumi, 2005; Aoki et al., 2012; Yasuda et al., 2013). Moreover, in dendrites many RNA binding proteins, including FUS, are involved in the local translation machinery to regulate synaptic function and morphology (Fujii and Takumi, 2005; Qiu et al., 2014; Sephton et al., 2014). Binding of FUS to the 3'UTR of target mRNAs is an important determinant of translational efficiency and mRNA stability (Colombrita et al., 2012; Lagier-Tourenne et al., 2012; Rogelj et al., 2012). Thus, these findings suggest that the cytoplasmic function of FUS may be involved in regulating mRNA stability, translation, and transport.

Masuda et al. reported that FUS participates in the alternative polyadenylation machinery with FUS binding nascent RNAs and interacting with the CPSF and CSTF complexes (Masuda et al., 2015). In addition, we have shown that FUS regulates GluA1 mRNA stability in cooperation with CPSF6, PAN2, and PABP, while it also controls the mRNA stability of SynGAP α 2, an isoform of SynGAP1, with ELAVL proteins in a 3'UTR length-dependent manner. FUS-silencing reduced the number of mature dendritic spines both *in vitro* and *in vivo*. Recovering expression of either GluA1 or the SynGAP α 2 isoform in FUS-deficient mice partially ameliorated abnormal behaviors and the impaired dendritic spine maturation caused by FUS-depletion, suggesting that FUS-mediated GluA1 mRNA stability and control of SynGAP α 2 isoform-specific expression is critical for these phenotypes (Udagawa et al., 2015; Yokoi et al., 2017).

These results, taken together, suggest that the loss of regulatory control of synaptic molecule mRNA stability in response to impaired FUS functionality causes synaptic dysfunction and could lead to post-synapse impairments in FTL/ALS.

MAINTENANCE OF NEURONAL MORPHOLOGY BY FUS

It is known that post-synapse impairments in neurodegenerative disorders including FTL/ALS might be an early pathological change (Sephton and Yu, 2015; Herms and Dorostkar, 2016). For instance, misrouting of Tau protein into the somatodendritic compartment is recognized as an early pathological event in Alzheimer disease (AD) and other tauopathies (Ballatore et al., 2007; Hoover et al., 2010). Similarly, FUS^{R521G} transgenic mice exhibited a reduction of dendritic arbors and mature spines (Sephton et al., 2014), and overexpression of FUS^{R521C} exhibited dendritic and synaptic defects accompanied with damaged splicing of *Bdnf* (Qiu et al., 2014).

It was demonstrated that neurite outlength is reduced in FUS-silenced primary cortical neurons but can be recovered by overexpressing wild-type FUS, whereas disease-associated mutants had no effect (Ishigaki et al., 2017). Similarly, iPSC-derived neurons in familial ALS patients harboring mutations in FUS exhibited shorter neurites compared to controls (Ichihayashi et al., 2016). Moreover, rescue by co-silencing 4R-Tau ameliorated the toxic effects of FUS-silencing on neurite outlength (Ishigaki et al., 2017). Thus, FUS dysfunction induces

abnormal neuronal morphology, which may be attributable to alterations in tau isoforms. Indeed, 4R-Tau functions in suppression of microtubule dynamics by stabilizing microtubule interactions and 4R-Tau overexpression affected neurite outlength in a dose-dependent manner (Panda et al., 2003; Ishigaki et al., 2017). Thus, the morphological abnormalities in neurites might be one of the earliest biomarkers and could thus be used in therapeutic screens or as a diagnostic tool.

REGULATION OF AXONAL FUNCTION BY FUS

Some studies have implicated FUS in the regulation of neuronal pre-synaptic function with disease-associated FUS mutants impairing its regulatory role (Sasayama et al., 2012; Armstrong and Drapeau, 2013; Schoen et al., 2015). Errichelli et al. reported that circular RNA expression, which is involved in axon guidance, was affected in motor neurons of FUS KO mice (Errichelli et al., 2017). Axonal transport defects have been reported for ALS/FTLD-associated mutations of FUS (Baldwin et al., 2016; Chen et al., 2016). Moreover, Guo et al. found that axonal transport was affected by disease-associated mutations of FUS in human iPSC-derived motor neurons (Guo et al., 2017). Since axonal transport defects appear in mice carrying mutations in SOD1 that cause ALS and in *Drosophila* carrying mutations in TDP-43 and C9orf72 (Williamson and Cleveland, 1999; Baldwin et al., 2016), further investigation to clarify the common downstream pathomechanism is necessary.

LIQUID-PHASE TRANSITION OF FUS AND ITS PATHOLOGICAL AND PHYSIOLOGICAL FUNCTIONS

Recent studies have unveiled a novel protein property of FUS, liquid-liquid phase transitions that lead to the formation of various proteinaceous membrane-less organelles. It has been demonstrated that FUS undergoes a liquid-liquid phase separation before converting into the insoluble form of the protein, a process that is promoted by mutations, phosphorylation, or the presence of RNA (Murakami et al., 2015; Patel et al., 2015; Chong and Forman-Kay, 2016; Monahan et al., 2017). Similar to hnRNP A2, the low complexity domain (LCD) in the C-terminal region of FUS is responsible for the liquid-liquid phase separation (Xiang et al., 2015; Murray et al., 2017). Other RNA-binding proteins such as TDP-43, TIA1, TAF15, and dipeptide repeat proteins synthesized from mutant C9orf72 also contain LCDs (Boeynaems et al., 2017; Harrison and Shorter, 2017). Moreover, it has been reported that Tau also undergoes a liquid-liquid phase separation in solution with 4R-Tau more prone to form liquid droplets than 3R-Tau (Ambadipudi et al., 2017).

These findings strongly suggest a biochemical link between RNA-binding proteins and other amyloid-formable proteins

including Tau and its association with RNA processing in neurodegenerative diseases. Since those findings were based on *in vitro* experiments, further investigation is necessary to clarify whether/how liquid-liquid phase transitions are associated with biological function, and whether transitions that occur in the cytoplasm of dendritic spines and/or the nucleus utilize the same or a different molecular process.

CONCLUSIONS

Accumulating *in vitro* and *in vivo* evidences indicate that FUS dysfunction might be involved in the pathomechanism of FTLD/ALS and other neurodegenerative diseases including tauopathies. FUS directly impacts RNA metabolism via alternative splicing, transcription, and mRNA stabilization,

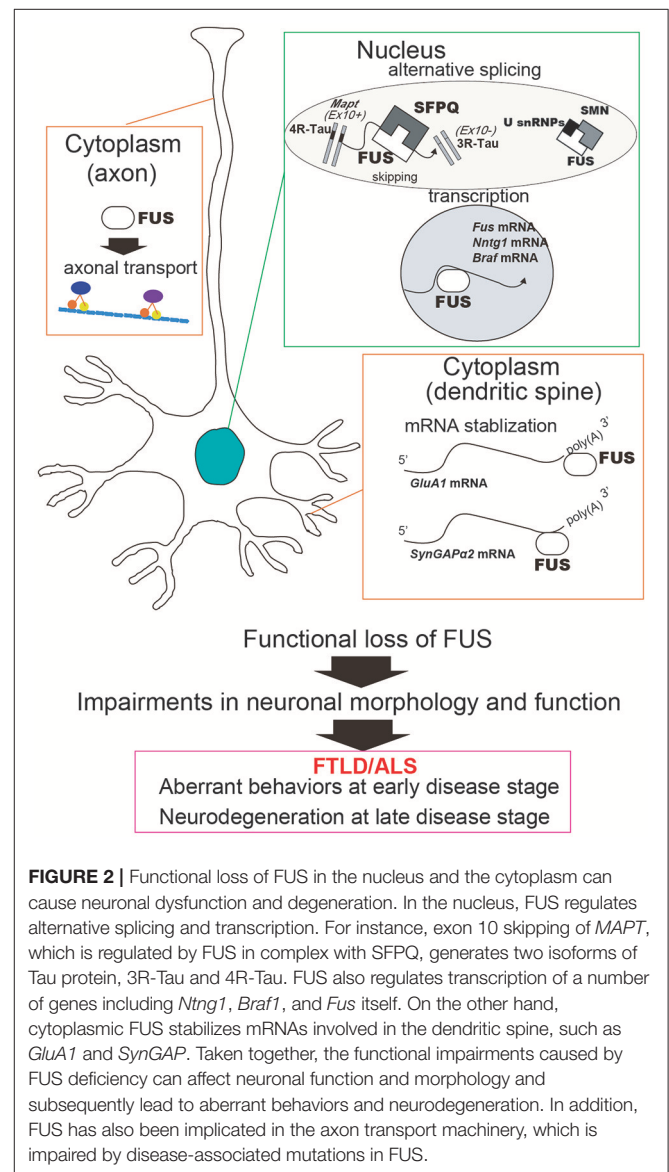


FIGURE 2 | Functional loss of FUS in the nucleus and the cytoplasm can cause neuronal dysfunction and degeneration. In the nucleus, FUS regulates alternative splicing and transcription. For instance, exon 10 skipping of *MAPT*, which is regulated by FUS in complex with SFPQ, generates two isoforms of Tau protein, 3R-Tau and 4R-Tau. FUS also regulates transcription of a number of genes including *Ntng1*, *Braf1*, and *Fus* itself. On the other hand, cytoplasmic FUS stabilizes mRNAs involved in the dendritic spine, such as *GluA1* and *SynGAP*. Taken together, the functional impairments caused by FUS deficiency can affect neuronal function and morphology and subsequently lead to aberrant behaviors and neurodegeneration. In addition, FUS has also been implicated in the axon transport machinery, which is impaired by disease-associated mutations in FUS.

all of which can subsequently influence neuronal/synaptic functions and lead to impaired behaviors during the early disease stage and neurodegeneration at the late disease stage (Figure 2). Although this review focused on the loss of FUS function, FUS toxicity could affect RNA metabolism as well; for instance, overexpression of mutant FUS has been shown to disrupt target gene expression (Coady and Manley, 2015). It indicates that loss-of-function and/or gain-of-toxicity of FUS might influence RNA metabolism pathways and subsequent cellular phenomenon. Indeed, in neurons with simultaneous depletion of FUS and TAF15 the gene expression profiles were similar to that in ALS patient-derived neurons bearing the ALS mutation FUS^{R521G} (Kapeli et al., 2016). This is further supported by similar transcriptome profiles in TDP-43 Drosophila models following both loss and gain of FUS function (Vanden Broeck et al., 2013). Consequently, to determine the utility of FUS and its downstream pathways as early disease markers and/or therapeutic targets of FTLD/ALS, it is crucial that their functional properties be more precisely clarified.

REFERENCES

- Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* 8:275. doi: 10.1038/s41467-017-00480-0
- Aoki, N., Higashi, S., Kawakami, I., Kobayashi, Z., Hosokawa, M., Katsuse, O., et al. (2012). Localization of fused in sarcoma (FUS) protein to the post-synaptic density in the brain. *Acta Neuropathol.* 124, 383–394. doi: 10.1007/s00401-012-0984-6
- Armstrong, G. A., and Drapeau, P. (2013). Loss and gain of FUS function impair neuromuscular synaptic transmission in a genetic model of ALS. *Hum. Mol. Genet.* 22, 4282–4292. doi: 10.1093/hmg/ddt278
- Baldwin, K. R., Godena, V. K., Hewitt, V. L., and Whitworth, A. J. (2016). Axonal transport defects are a common phenotype in Drosophila models of ALS. *Hum. Mol. Genet.* 25, 2378–2392. doi: 10.1093/hmg/ddw105
- Ballatore, C., Lee, V. M., and Trojanowski, J. Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* 8, 663–672. doi: 10.1038/nrn2194
- Bäumer, D., Hilton, D., Paine, S. M., Turner, M. R., Lowe, J., Talbot, K., et al. (2010). Juvenile ALS with basophilic inclusions is a FUS proteinopathy with FUS mutations. *Neurology* 75, 611–618. doi: 10.1212/WNL.0b013e3181e d9cde
- Belzil, V. V., Langlais, J. S., Daoud, H., Dion, P. A., Brais, B., and Rouleau, G. A. (2012). Novel FUS deletion in a patient with juvenile amyotrophic lateral sclerosis. *Arch. Neurol.* 69, 653–656. doi: 10.1001/archneurol.2011.2499
- Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P., and Tora, L. (1996). hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* 15, 5022–5031.
- Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., et al. (2017). Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* 65, 1044.e5–1055.e5. doi: 10.1016/j.molcel.2017.02.013
- Boillée, S., Yamanaka, K., Lobsiger, C. S., Copeland, N. G., Jenkins, N. A., Kassiotis, G., et al. (2006). Onset and progression in inherited ALS determined by motor neurons and microglia. *Science* 312, 1389–1392. doi: 10.1126/science.1123511
- Chen, Y., Deng, J., Wang, P., Yang, M., Chen, X., Zhu, L., et al. (2016). PINK1 and parkin are genetic modifiers for FUS-induced neurodegeneration. *Hum. Mol. Genet.* 25, 5059–5068. doi: 10.1093/hmg/ddw310
- Chong, P. A., and Forman-Kay, J. D. (2016). Liquid-liquid phase separation in cellular signaling systems. *Curr. Opin. Struct. Biol.* 41, 180–186. doi: 10.1016/j.sbi.2016.08.001
- Coady, T. H., and Manley, J. L. (2015). ALS mutations in TLS/FUS disrupt target gene expression. *Genes Dev.* 29, 1696–1706. doi: 10.1101/gad.267286.115
- Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F. E., Buratti, E., et al. (2012). TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *J. Biol. Chem.* 287, 15635–15647. doi: 10.1074/jbc.M111.333450
- Crozat, A., Aman, P., Mandahl, N., and Ron, D. (1993). Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* 363, 640–644. doi: 10.1038/363640a0
- Deng, H. X., Zhai, H., Bigio, E. H., Yan, J., Fecto, F., Ajroud, K., et al. (2010). FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. *Ann. Neurol.* 67, 739–748. doi: 10.1002/ana.22051
- Endo, K., Ishigaki, S., Masamizu, Y., Fujioka, Y., Watakabe, A., Yamamori, T., et al. (2017). Silencing of FUS in the common marmoset (*Callithrix jacchus*) brain via stereotaxic injection of an adeno-associated virus encoding shRNA. *Neurosci. Res.* doi: 10.1016/j.neures.2017.08.006. [Epub ahead of print].
- Errichelli, L., Dini Modigliani, S., Laneve, P., Colantoni, A., Legnini, I., Caputo, D., et al. (2017). FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* 8:14741. doi: 10.1038/ncomms14741
- Ferrer, I., Legati, A., García-Monco, J. C., Gomez-Beldarrain, M., Carmona, M., Blanco, R., et al. (2015). Familial behavioral variant frontotemporal dementia associated with astrocyte-predominant tauopathy. *J. Neuropathol. Exp. Neurol.* 74, 370–379. doi: 10.1097/NEN.0000000000000180
- Fujii, R., and Takumi, T. (2005). TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J. Cell Sci.* 118(Pt 24), 5755–5765. doi: 10.1242/jcs.02692
- Fujioka, Y., Ishigaki, S., Masuda, A., Iguchi, Y., Udagawa, T., Watanabe, H., et al. (2013). FUS-regulated region- and cell-type-specific transcriptome is associated with cell selectivity in ALS/FTLD. *Sci. Rep.* 3:2388. doi: 10.1038/srep02388
- Gao, F. B., Almeida, S., and Lopez-Gonzalez, R. (2017). Dysregulated molecular pathways in amyotrophic lateral sclerosis-frontotemporal dementia spectrum disorder. *EMBO J.* 36, 2931–2950. doi: 10.15252/embj.201797568

AUTHOR CONTRIBUTIONS

SI: Conception and design, manuscript writing, editing, and figure design. GS: Conception and design, manuscript writing, and editing.

ACKNOWLEDGMENTS

This work was supported by Mext Grant-in-aid project, Scientific Research on Innovation Area (Brain Protein Aging and Dementia control), by Mext Grant-in-Aid for Scientific Research on Innovative Areas (Comprehensive Brain Science Network), by Mext KAKENHI grant number 15K09310, and by CREST from JST.

This work was also supported by the Integrated Research on Neuropsychiatric Disorders and Integrated Research on Depression, Dementia and Development Disorders projects carried out under the Strategic Research Program for Brain Sciences and Brain/MINDS of the Japan Agency for Medical Research and Development.

- Guo, W., Naujock, M., Fumagalli, L., Vandoorne, T., Baatsen, P., Boon, R., et al. (2017). HDAC6 inhibition reverses axonal transport defects in motor neurons derived from FUS-ALS patients. *Nat. Commun.* 8:861. doi: 10.1038/s41467-017-00911-y
- Harrison, A. F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438. doi: 10.1042/BCJ20160499
- Hermes, J., and Dorostkar, M. M. (2016). Dendritic spine pathology in Neurodegenerative diseases. *Annu. Rev. Pathol.* 11, 221–250. doi: 10.1146/annurev-pathol-012615-044216
- Hicks, G. G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., et al. (2000). FUS deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nat. Genet.* 24, 175–179. doi: 10.1038/72842
- Honda, D., Ishigaki, S., Iguchi, Y., Fujioka, Y., Udagawa, T., Masuda, A., et al. (2013). The ALS/FTLD-related RNA-binding proteins TDP-43 and FUS have common downstream RNA targets in cortical neurons. *FEBS Open Biol.* 4, 1–10. doi: 10.1016/j.fob.2013.11.001
- Hong, M., Zhukareva, V., Vogelsberg-Ragaglia, V., Wszolek, Z., Reed, L., Miller, B. I., et al. (1998). Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17. *Science* 282, 1914–1917. doi: 10.1126/science.282.5395.1914
- Hoover, B. R., Reed, M. N., Su, J., Penrod, R. D., Kotilinek, L. A., Grant, M. K., et al. (2010). Tau mislocalization to dendritic spines mediates synaptic dysfunction independently of neurodegeneration. *Neuron* 68, 1067–1081. doi: 10.1016/j.neuron.2010.11.030
- Huang, E. J., Zhang, J., Geser, F., Trojanowski, J. Q., Strober, J. B., Dickson, D. W., et al. (2010). Extensive FUS-immunoreactive pathology in juvenile amyotrophic lateral sclerosis with basophilic inclusions. *Brain Pathol.* 20, 1069–1076. doi: 10.1111/j.1750-3639.2010.00413.x
- Ichiyanagi, N., Fujimori, K., Yano, M., Ishihara-Fujisaki, C., Sone, T., Akiyama, T., et al. (2016). Establishment of *in vitro* FUS-associated familial amyotrophic lateral sclerosis model using human induced pluripotent stem cells. *Stem Cell Rep.* 6, 496–510. doi: 10.1016/j.stemcr.2016.02.011
- Ishigaki, S., Fujioka, Y., Okada, Y., Riku, Y., Udagawa, T., Honda, D., et al. (2017). Altered tau isoform ratio caused by loss of FUS and SFPQ function leads to FTL-like phenotypes. *Cell Rep.* 18, 1118–1131. doi: 10.1016/j.celrep.2017.01.013
- Ishigaki, S., Masuda, A., Fujioka, Y., Iguchi, Y., Katsuno, M., Shibata, A., et al. (2012). Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Sci. Rep.* 2:529. doi: 10.1038/srep00529
- Kabashi, E., Bercier, V., Lissouba, A., Liao, M., Brustein, E., Rouleau, G. A., et al. (2011). FUS and TARDBP but not SOD1 interact in genetic models of amyotrophic lateral sclerosis. *PLoS Genet.* 7:e1002214. doi: 10.1371/journal.pgen.1002214
- Kapeli, K., Pratt, G. A., Vu, A. Q., Hutt, K. R., Martinez, F. J., Sundararaman, B., et al. (2016). Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses. *Nat. Commun.* 7:12143. doi: 10.1038/ncomms12143
- Kino, Y., Washizu, C., Aquilanti, E., Okuno, M., Kurosawa, M., Yamada, M., et al. (2011). Intracellular localization and splicing regulation of FUS/TLS are variably affected by amyotrophic lateral sclerosis-linked mutations. *Nucleic Acids Res.* 39, 2781–2798. doi: 10.1093/nar/gkq1162
- Kino, Y., Washizu, C., Kurosawa, M., Yamada, M., Miyazaki, H., Akagi, T., et al. (2015). FUS/TLS deficiency causes behavioral and pathological abnormalities distinct from amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* 3:24. doi: 10.1186/s40478-015-0202-6
- Kuroda, M., Sok, J., Webb, L., Baechtold, H., Urano, F., Yin, Y., et al. (2000). Male sterility and enhanced radiation sensitivity in TLS^(-/-) mice. *EMBO J.* 19, 453–462. doi: 10.1093/emboj/19.3.453
- Kwiatkowski, T. J., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* 323, 1205–1208. doi: 10.1126/science.1166066
- Lagier-Tourenne, C., and Cleveland, D. W. (2009). Rethinking ALS: the FUS about TDP-43. *Cell* 136, 1001–1004. doi: 10.1016/j.cell.2009.03.006
- Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* 15, 1488–1497. doi: 10.1038/nn.3230
- Mackenzie, I. R., Ansorge, O., Strong, M., Bilbao, J., Zinman, L., Ang, L. C., et al. (2011). Pathological heterogeneity in amyotrophic lateral sclerosis with FUS mutations: two distinct patterns correlating with disease severity and mutation. *Acta Neuropathol.* 122, 87–98. doi: 10.1007/s00401-011-0838-7
- Masuda, A., Takeda, J., Okuno, T., Okamoto, T., Ohkawara, B., Ito, M., et al. (2015). Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev.* 29, 1045–1057. doi: 10.1101/gad.255737.114
- Monahan, Z., Ryan, V. H., Janke, A. M., Burke, K. A., Rhoads, S. N., Zerze, G. H., et al. (2017). Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* 36, 2951–2967. doi: 10.15252/embj.201696394
- Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S., Rees, E., Miyashita, A., et al. (2015). ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* 88, 678–690. doi: 10.1016/j.neuron.2015.10.030
- Murray, D. T., Kato, M., Lin, Y., Thurber, K. R., Hung, I., McKnight, S. L., et al. (2017). Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* 171, 615.e16–627. e6. doi: 10.1016/j.cell.2017.08.048
- Nakaya, T., Alexiou, P., Maragkakis, M., Chang, A., and Mourelatos, Z. (2013). FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* 19, 498–509. doi: 10.1261/rna.037804.112
- Neumann, M., Rademakers, R., Roeber, S., Baker, M., Kretschmar, H. A., and Mackenzie, I. R. (2009). A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain* 132(Pt 11), 2922–2931. doi: 10.1093/brain/awp214
- Orozco, D., and Edbauer, D. (2013). FUS-mediated alternative splicing in the nervous system: consequences for ALS and FTL. *J. Mol. Med.* 91, 1343–1354. doi: 10.1007/s00109-013-1077-2
- Orozco, D., Tahirovic, S., Rentzsch, K., Schwenk, B. M., Haass, C., and Edbauer, D. (2012). Loss of fused in sarcoma (FUS) promotes pathological Tau splicing. *EMBO Rep.* 13, 759–764. doi: 10.1038/embor.2012.90
- Panda, D., Samuel, J. C., Massie, M., Feinstein, S. C., and Wilson, L. (2003). Differential regulation of microtubule dynamics by three- and four-repeat tau: implications for the onset of neurodegenerative disease. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9548–9553. doi: 10.1073/pnas.1633508100
- Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnke, M., Hein, M. Y., et al. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* 162, 1066–1077. doi: 10.1016/j.cell.2015.07.047
- Qiu, H., Lee, S., Shang, Y., Wang, W. Y., Au, K. F., Kamiya, S., et al. (2014). ALS-associated mutation FUS-R521C causes DNA damage and RNA splicing defects. *J. Clin. Invest.* 124, 981–999. doi: 10.1172/JCI72723
- Renton, A. E., Chiò, A., and Traynor, B. J. (2014). State of play in amyotrophic lateral sclerosis genetics. *Nat. Neurosci.* 17, 17–23. doi: 10.1038/nn.3584
- Riku, Y., Watanabe, H., Yoshida, M., Tatsumi, S., Mimuro, M., Iwasaki, Y., et al. (2014). Lower motor neuron involvement in TAR DNA-binding protein of 43 kDa-related frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *JAMA Neurol.* 71, 172–179. doi: 10.1001/jamaneurol.2013.5489
- Robberecht, W., and Philips, T. (2013). The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* 14, 248–264. doi: 10.1038/nrn3430
- Rogelj, B., Easton, L. E., Bogu, G. K., Stanton, L. W., Rot, G., Curk, T., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* 2:603. doi: 10.1038/srep00603
- Sasayama, H., Shimamura, M., Tokuda, T., Azuma, Y., Yoshida, T., Mizuno, T., et al. (2012). Knockdown of the Drosophila fused in sarcoma (FUS) homologue causes deficient locomotive behavior and shortening of motoneuron terminal branches. *PLoS ONE* 7:e39483. doi: 10.1371/journal.pone.0039483
- Seckic-Zahirovic, J., Sendscheid, O., El Oussini, H., Jambeau, M., Sun, Y., Mersmann, S., et al. (2016). Toxic gain of function from mutant FUS protein is crucial to trigger cell autonomous motor neuron loss. *EMBO J.* 35, 1077–1097. doi: 10.15252/embj.201592559
- Schoen, M., Reichel, J. M., Demestre, M., Putz, S., Deshpande, D., Proepper, C., et al. (2015). Super-resolution microscopy reveals presynaptic localization of the ALS/FTD related protein FUS in hippocampal neurons. *Front. Cell. Neurosci.* 9:496. doi: 10.3389/fncel.2015.00496
- Schwartz, J. C., Ebmeier, C. C., Podell, E. R., Heimiller, J., Taatjes, D. J., and Cech, T. R. (2012). FUS binds the CTD of RNA polymerase

- II and regulates its phosphorylation at Ser2. *Genes Dev.* 26, 2690–2695. doi: 10.1101/gad.204602.112
- Seelaar, H., Rohrer, J. D., Pijnenburg, Y. A., Fox, N. C., and van Swieten, J. C. (2011). Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *J. Neurol. Neurosurg. Psychiatry* 82, 476–486. doi: 10.1136/jnnp.2010.212225
- Sephton, C. F., Tang, A. A., Kulkarni, A., West, J., Brooks, M., Stubblefield, J. J., et al. (2014). Activity-dependent FUS dysregulation disrupts synaptic homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4769–E4778. doi: 10.1073/pnas.1406162111
- Sephton, C. F., and Yu, G. (2015). The function of RNA-binding proteins at the synapse: implications for neurodegeneration. *Cell. Mol. Life Sci.* 72, 3621–3635. doi: 10.1007/s00018-015-1943-x
- Sharma, A., Lyashchenko, A. K., Lu, L., Nasrabad, S. E., Elmaleh, M., Mendelsohn, M., et al. (2016). ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function. *Nat. Commun.* 7:10465. doi: 10.1038/ncomms10465
- Strong, M. J., and Volkening, K. (2011). TDP-43 and FUS/TLS: sending a complex message about messenger RNA in amyotrophic lateral sclerosis? *FEBS J.* 278, 3569–3577. doi: 10.1111/j.1742-4658.2011.08277.x
- Sun, S., Ling, S. C., Qiu, J., Albuquerque, C. P., Zhou, Y., Tokunaga, S., et al. (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* 6:6171. doi: 10.1038/ncomms7171
- Swinnen, B., and Robberecht, W. (2014). The phenotypic variability of amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* 10, 661–670. doi: 10.1038/nrneurol.2014.184
- Tan, A. Y., Riley, T. R., Coady, T., Bussemaker, H. J., and Manley, J. L. (2012). TLS/FUS (translocated in liposarcoma/fused in sarcoma) regulates target gene transcription via single-stranded DNA response elements. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6030–6035. doi: 10.1073/pnas.1203028109
- Thomas-Jinu, S., Gordon, P. M., Fielding, T., Taylor, R., Smith, B. N., Snowden, V., et al. (2017). Non-nuclear pool of splicing factor SFPQ regulates axonal transcripts required for normal motor development. *Neuron* 94:931. doi: 10.1016/j.neuron.2017.04.036
- Tsujii, H., Iguchi, Y., Furuya, A., Kataoka, A., Hatsuta, H., Atsuta, N., et al. (2013). Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO Mol. Med.* 5, 221–234. doi: 10.1002/emmm.201202303
- Udagawa, T., Fujioka, Y., Tanaka, M., Honda, D., Yokoi, S., Riku, Y., et al. (2015). FUS regulates AMPA receptor function and FTL/ALS-associated behaviour via GluA1 mRNA stabilization. *Nat. Commun.* 6:7098. doi: 10.1038/ncomms8098
- Umeda, T., Yamashita, T., Kimura, T., Ohnishi, K., Takuma, H., Ozeki, T., et al. (2013). Neurodegenerative disorder FTDP-17-related tau intron 10 +16C → T mutation increases tau exon 10 splicing and causes tauopathy in transgenic mice. *Am. J. Pathol.* 183, 211–225. doi: 10.1016/j.ajpath.2013.03.015
- Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 323, 1208–1211. doi: 10.1126/science.1165942
- Vanden Broeck, L., Naval-Sánchez, M., Adachi, Y., Diaper, D., Dourlen, P., Chapuis, J., et al. (2013). TDP-43 loss-of-function causes neuronal loss due to defective steroid receptor-mediated gene program switching in *Drosophila*. *Cell Rep.* 3, 160–172. doi: 10.1016/j.celrep.2012.12.014
- Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2012). The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum. *Ann. Med.* 44, 817–828. doi: 10.3109/07853890.2012.665471
- Wang, J. W., Brent, J. R., Tomlinson, A., Shneider, N. A., and McCabe, B. D. (2011). The ALS-associated proteins FUS and TDP-43 function together to affect *Drosophila* locomotion and life span. *J. Clin. Invest.* 121, 4118–4126. doi: 10.1172/JCI57883
- Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., et al. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454, 126–130. doi: 10.1038/nature06992
- Williamson, T. L., and Cleveland, D. W. (1999). Slowing of axonal transport is a very early event in the toxicity of ALS-linked SOD1 mutants to motor neurons. *Nat. Neurosci.* 2, 50–56. doi: 10.1038/4553
- Xiang, S., Kato, M., Wu, L. C., Lin, Y., Ding, M., Zhang, Y., et al. (2015). The LC Domain of hnRNP A2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* 163, 829–839. doi: 10.1016/j.cell.2015.10.040
- Yamanaka, K., Boillee, S., Roberts, E. A., Garcia, M. L., McAlonis-Downes, M., Mikse, O. R., et al. (2008a). Mutant SOD1 in cell types other than motor neurons and oligodendrocytes accelerates onset of disease in ALS mice. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7594–7599. doi: 10.1073/pnas.0802556105
- Yamanaka, K., Chun, S. J., Boillee, S., Fujimori-Tonou, N., Yamashita, H., Gutmann, D. H., et al. (2008b). Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nat. Neurosci.* 11, 251–253. doi: 10.1038/nn2047
- Yamashita, S., Mori, A., Sakaguchi, H., Suga, T., Ishihara, D., Ueda, A., et al. (2012). Sporadic juvenile amyotrophic lateral sclerosis caused by mutant FUS/TLS: possible association of mental retardation with this mutation. *J. Neurol.* 259, 1039–1044. doi: 10.1007/s00415-011-6292-6
- Yan, J., Deng, H. X., Siddique, N., Fecto, F., Chen, W., Yang, Y., et al. (2010). Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. *Neurology* 75, 807–814. doi: 10.1212/WNL.0b013e3181f07e0c
- Yasuda, K., Zhang, H., Loisele, D., Haystead, T., Macara, I. G., and Mili, S. (2013). The RNA-binding protein FUS directs translation of localized mRNAs in APC-RNP granules. *J. Cell Biol.* 203, 737–746. doi: 10.1083/jcb.201306058
- Yokoi, S., Udagawa, T., Fujioka, Y., Honda, D., Okado, H., Watanabe, H., et al. (2017). 3'UTR length-dependent control of SynGAP isoform $\alpha 2$ mRNA by FUS and ELAV-like proteins promotes dendritic spine maturation and cognitive function. *Cell Rep.* 20, 3071–3084. doi: 10.1016/j.celrep.2017.08.100
- Yoshida, M. (2006). Cellular tau pathology and immunohistochemical study of tau isoforms in sporadic tauopathies. *Neuropathology* 26, 457–470. doi: 10.1111/j.1440-1789.2006.00743.x
- Zhou, Y., Liu, S., Liu, G., Oztürk, A., and Hicks, G. G. (2013). ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. *PLoS Genet.* 9:e1003895. doi: 10.1371/journal.pgen.1003895

Conflict of Interest Statement: 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.

Copyright © 2018 Ishigaki and Sobue. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Myelodysplastic Syndrome-Associated SRSF2 Mutations Cause Splicing Changes by Altering Binding Motif Sequences

So Masaki^{1,2*}, Shun Ikeda¹, Asuka Hata¹, Yusuke Shiozawa³, Ayana Kon³, Seishi Ogawa³, Kenji Suzuki², Fumihiko Hakuno⁴, Shin-Ichiro Takahashi⁴ and Naoyuki Kataoka^{1,4*}

¹ Laboratory for Malignancy Control Research, Medical Innovation Center, Kyoto University Graduate School of Medicine, Kyoto, Japan, ² Laboratory of Molecular Medicinal Science, Department of Pharmaceutical Sciences, Ritsumeikan University, Shiga, Japan, ³ Department of Pathology and Tumor Biology, Kyoto University Graduate School of Medicine, Kyoto, Japan, ⁴ Laboratory of Cell Regulation, Departments of Applied Animal Sciences and Applied Biological Chemistry, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo, Japan

OPEN ACCESS

Edited by:

Philipp Kapranov,
Huaqiao University, China

Reviewed by:

Zhixiang Lu,
Harvard Medical School,
United States
Woan-Yuh Tarn,
Academia Sinica, Taiwan

*Correspondence:

So Masaki
smasaki@fc.ritsumei.ac.jp
Naoyuki Kataoka
akataoka@mail.ecc.u-tokyo.ac.jp

Specialty section:

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

Received: 15 November 2018

Accepted: 29 March 2019

Published: 16 April 2019

Citation:

Masaki S, Ikeda S, Hata A,
Shiozawa Y, Kon A, Ogawa S,
Suzuki K, Hakuno F, Takahashi S-I and
Kataoka N (2019) Myelodysplastic
Syndrome-Associated SRSF2
Mutations Cause Splicing Changes by
Altering Binding Motif Sequences.
Front. Genet. 10:338.
doi: 10.3389/fgene.2019.00338

Serine/arginine-rich splicing factor 2 (SRSF2) is a member of the SR protein family that is involved in both constitutive and alternative mRNA splicing. Mutations in SRSF2 gene are frequently reported in myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). It is imperative to understand how these mutations affect SRSF2-mediated splicing and cause MDS. In this study, we characterized MDS-associated SRSF2 mutants (P95H, P95L, and P95R). We found that those mutants and wild-type SRSF2 proteins showed nuclear localization in HeLa cells. *In vitro* splicing reaction also revealed that mutant proteins associated with both precursor and spliced mRNAs, suggesting that the mutants directly participate in splicing. We established the human myeloid leukemia K562 cell lines that stably expressed myc-tagged wild-type or mutant SRSF2 proteins, and then performed RNA-sequence to analyze the splicing pattern of each cell line. The results revealed that both wild-type and mutants affected splicing of approximately 3,000 genes. Although splice site sequences adjacent to the affected exons showed no significant difference compared to the total exons, exonic motif analyses with both inclusion- and exclusion-enhanced exons demonstrated that wild-type and mutants have different binding sequences in exons. These results indicate that mutations of SRSF2 in MDS change binding properties of SRSF2 to exonic motifs and this causes aberrant splicing.

Keywords: myelodysplastic syndrome, splicing, SRSF2, exonic splicing enhancer, aberrant splicing, EZH2 (enhancer of zeste homolog 2)

INTRODUCTION

Serine/Arginine rich (SR) proteins are essential splicing factors that also confer regulatory activity of alternative splicing (Manley and Krainer, 2010; Howard and Sanford, 2015; Kataoka, 2017). SR protein family consists of 11 proteins in human. SR proteins contain one or two RNA binding domain (RBD) at amino-terminus and multiple repeats of

Serine-Arginine dipeptides at carboxy-terminus. SRSF2, which was originally called SC35, is a member of the SR protein family (Fu and Maniatis, 1990; Fu and Ares, 2014; Kataoka, 2017). During splicing, SRSF2 promotes exon recognition by binding to exonic splicing enhancer (ESE) motifs in precursor of mRNA (pre-mRNA) through its RBD. This promotes both the binding of U2AF heterodimer and U1 snRNP to the upstream 3' splice site and to the downstream 5' splice site, respectively (Chen and Manley, 2009; Fu and Ares, 2014; Kataoka, 2017).

Recently SRSF2 was found to be one of the major responsible genes of myelodysplastic syndrome (MDS). MDS is a heterogeneous group of chronic myeloid neoplasms characterized by many symptoms, such as ineffective hematopoiesis, peripheral blood cytopenia and a high risk of progression to acute myeloid leukemia (Fu and Maniatis, 1990; Cazzola et al., 2013). Mutations in splicing factors represent a novel class of driver mutations in human cancers and affect about 50% of patients with myelodysplasia (Graubert et al., 2011; Papaemmanuil et al., 2011, 2013; Yoshida et al., 2011; Walter et al., 2013; Haferlach et al., 2014). Somatic mutations are frequently found in genes SF3B1, SRSF2, U2AF1 and ZRSR2 (Graubert et al., 2011; Papaemmanuil et al., 2011; Yoshida et al., 2011). Interestingly, the common feature of those gene products in pre-mRNA splicing is a 3' splice site recognition, implicating that MDS pathogenesis is caused most likely by abnormal pre-mRNA splicing. The mutations are heterozygous and clustered in specific amino acid residues for SF3B1, SRSF2 and U2AF1 (Yoshida et al., 2011; Darman et al., 2015; Ilagan et al., 2015; Kim et al., 2015). These 'hotspot' mutations are predicted to be a 'gain-of-function' mutations by affecting protein structure (Yoshida et al., 2011; Darman et al., 2015; Ilagan et al., 2015; Kim et al., 2015). Among them, the hotspot mutations in SRSF2 were found in proline residue residing slightly outside of its RNA binding domain. Changes in alternative splicing patterns with SRSF2 mutations were reported in culture cells, mouse models and primary human samples (Przychodzen et al., 2013; Brooks et al., 2014; Shao et al., 2014; Ilagan et al., 2015; Kim et al., 2015; Komeno et al., 2015; Shirai et al., 2015; Zhang et al., 2015; Obeng et al., 2016; Mupo et al., 2017). In spite of these important findings, the precise mechanism for aberrant splicing in the cells carrying MDS mutations still remains largely unclear.

To address the mechanism of aberrant splicing in MDS, we have prepared expression plasmids for both wild-type and MDS-causing mutants of SRSF2, P95H, P95L and P95R. Transcriptome analyses using total RNAs recovered from those cell lines revealed that both wild-type and mutants affected splicing of approximately 3,000 genes. Motif analyses with both inclusion- and exclusion-enhanced exons demonstrated that the mutants have different binding sequences in exons compared to wild-type protein. These results strongly suggest that MDS mutations in SRSF2 alter binding properties of SRSF2 to exonic motifs and this results in aberrant splicing.

RESULTS

Localization of Wild-Type and Mutant SRSF2 Proteins in HeLa Cells and Association With Pre-mRNA and mRNA *in vitro*

In order to gain insights on how mutations in SRSF2, one of the splicing factors mutated in MDS patients, affect splicing, we have prepared the SRSF2 mutant cDNAs carrying three kinds of mutations at 95th position of Proline residue (P95H, L and R) found in MDS patients. We first determined their subcellular localization in HeLa cells. Both wild-type and mutants of SRSF2 cDNAs were transfected into HeLa cells and those proteins were expressed as fusions with a myc-tag. When the myc-SRSF2 wild-type protein was expressed, it was localized in both speckles and nucleoplasm, likely due to overexpression (Figure 1A, panel a). The subcellular localization of mutant SRSF2 proteins was similar to that of the wild-type, although mutant proteins exhibited slightly less numbers of nuclear speckles (Figure 1A).

We have also carried out *in vitro* splicing reaction with both Flag-tagged wild-type and mutant SRSF2 proteins, followed by immunoprecipitation of RNA from the mixture in order to test whether mutant proteins were able to support splicing or not. For this assay, we used the immunoglobulin μ chain (IgM) pre-mRNA, whose splicing is SRSF2-dependent (Mayeda et al., 1999). HeLa cell extract was mixed with cell lysates from HEK293T cells that express either wild-type or each mutant SRSF2 protein. The Flag-vector transfected cell lysate was used as a negative control. The results of splicing reaction are shown in Figure 1B. With a negative control, splicing did not take place (lane 1), indicating the dependency of IgM pre-mRNA to SRSF2. As shown in Figure 1B (lanes 2–5), both wild-type and each mutant supported IgM splicing *in vitro*, indicating mutant SRSF2 proteins have a splicing supporting activity even stronger than wild-type protein. Immunoprecipitation was carried out with the reaction mixtures using anti-Flag tag antibody. Flag-SRSF2 wild-type precipitated both pre-mRNA and mRNA (Figure 1B, lane 7). Mutant SRSF2 proteins also precipitated pre-mRNA and mRNA more efficiently than wild-type protein (Figure 1B, lane 8–10). Immunoprecipitated SRSF2 proteins were also detected by western blotting with anti-Flag M2 antibody (Figure 1C), and it turned out the comparable amounts of both wild-type and mutant SRSF2 proteins were precipitated (Figure 1C, lanes 7–10). Take all the results together, we concluded that mutant SRSF2 proteins are able to cause change of splicing pattern in cells, and the mutants likely have higher affinity to SRSF2-mediated ESE sequences.

Detection of Splicing Pattern Changes in K562 Cells Expressing Mutant SRSF2 Proteins

We generated to generate cell lines that stably express either wild-type or SRSF2 mutants with K562 cells, a myelogenous leukemia cell line. After establishment, we checked the protein expression level by western blotting by using anti-myc tag antibody. As

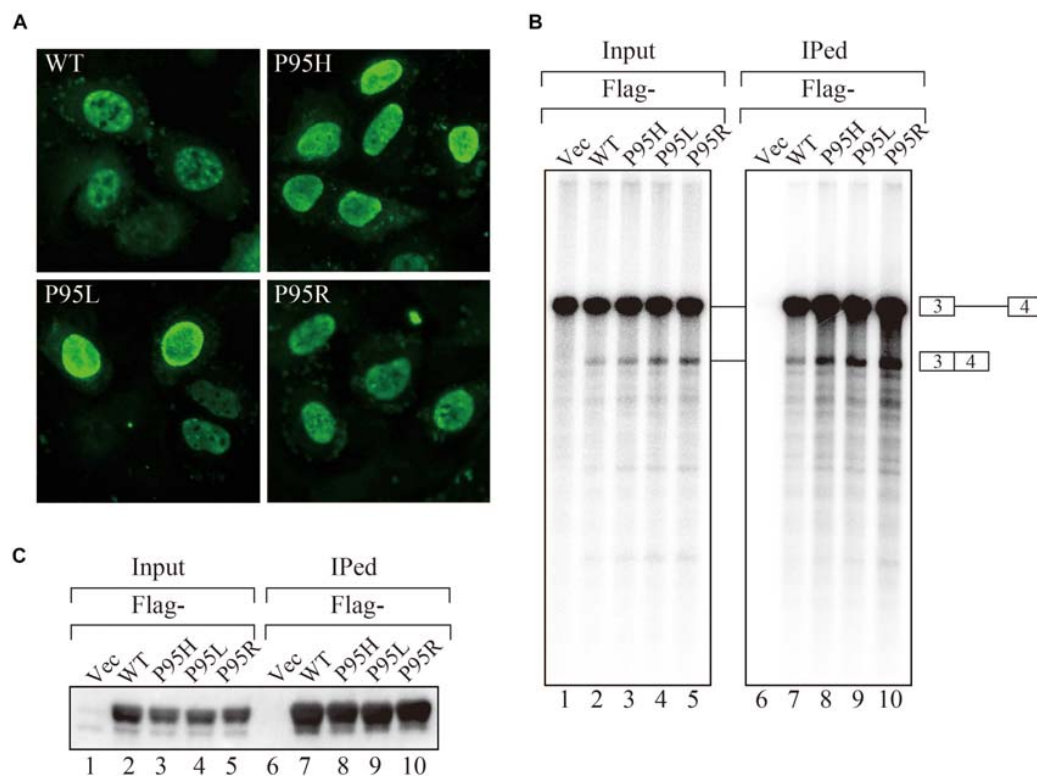


FIGURE 1 | MDS-causing SRSF2 mutant proteins are localized in the nuclei and associate with pre-mRNA and mRNA *in vitro*. **(A)** HeLa cells were transfected with expression vectors encoding either wild-type or mutant SRSF2 proteins (P95H, P95L, and P95R) fused with myc-tag. **(B)** *In vitro* splicing reaction in the presence of either Flag-tagged wild-type SRSF2 protein or mutant proteins followed by immunoprecipitation of RNAs by anti-Flag M2 antibody. **(C)** Immunoprecipitation of SRSF2 proteins by anti-Flag antibody from *in vitro* splicing reaction. Five percent of total reaction mixture was used as an input. The anti-Flag M2 antibody was used for detection of Flag-tagged SRSF2 proteins.

shown in **Figure 2A**, all cells expressed myc-tagged SRSF2 proteins except myc-vector transfected cells. During selection of the stable cell lines, cells expressing relatively large amount of mutant SRSF2 proteins tend to die, likely due to toxicity of mutant proteins and cells expressing high amount of mutant proteins may have been eliminated during selection of the clones. Total RNAs derived from those cell lines were applied to RNA sequence analyses. The reads were mapped to human genome with more than 90% efficiency (**Figure 2B**). We determined splicing changes by comparing wild-type and mutants splicing patterns with vector transfected one. It turned out that more than 5000 splicing events in more than 3000 genes were detected with both wild-type and mutants (**Figure 2C**). Drawing Venn diagram demonstrated that overlapping genes for all of them are 1406 genes, and there are also many genes either specific to each protein (**Figure 2D**).

ESE-Like Motifs Were Enriched in Exons Skipped by SRSF2 Mutant Proteins

Since we identified many genes whose splicing patterns were affected by the expression of both wild-type and mutant SRSF2 proteins, we investigated the motifs of exons enriched in both wild-type and mutants regulated genes. Specific sequence features

in the exon were searched by multiple expectation-maximization for motif elicitation (MEME) (Bailey et al., 2006) algorithm to find the enrichment or depletion of novel sequence motifs. Surprisingly, enriched motifs for all mutants are purine-rich sequences (**Figure 3A**). The motif for wild-type SRSF2-excluded exons is AGGTRAG (R indicates purine residue), in which the purine stretch is separated by T residue (**Figure 3A**). It has been shown that SRSF2 proteins are able to bind purine-rich ESEs (Cavaloc et al., 1999). These results strongly suggested that MDS-causing mutations in proline residue of SRSF2 cause reduction of the affinity to purine-rich ESEs.

Mutant SRSF2 Proteins Prefer CCWG Motif for Exon Inclusion

We also carried out motif analysis for included exons, and we found that a purine-rich motif appeared as the most frequently appeared motif with a wild-type protein (**Figure 3B**), consistent with the previous findings that SRSF2 binds to purine-rich ESE to promote exon inclusion (Cavaloc et al., 1999). With SRSF2 P95L, a similar A/G rich motif was also found (**Figure 3B**), suggesting this mutation has slight effect on recognizing sequence of SRSF2 protein. In contrast, with two other mutant proteins, SRSF2 P95H and P95R, CCWG (W: weak as T or A) containing

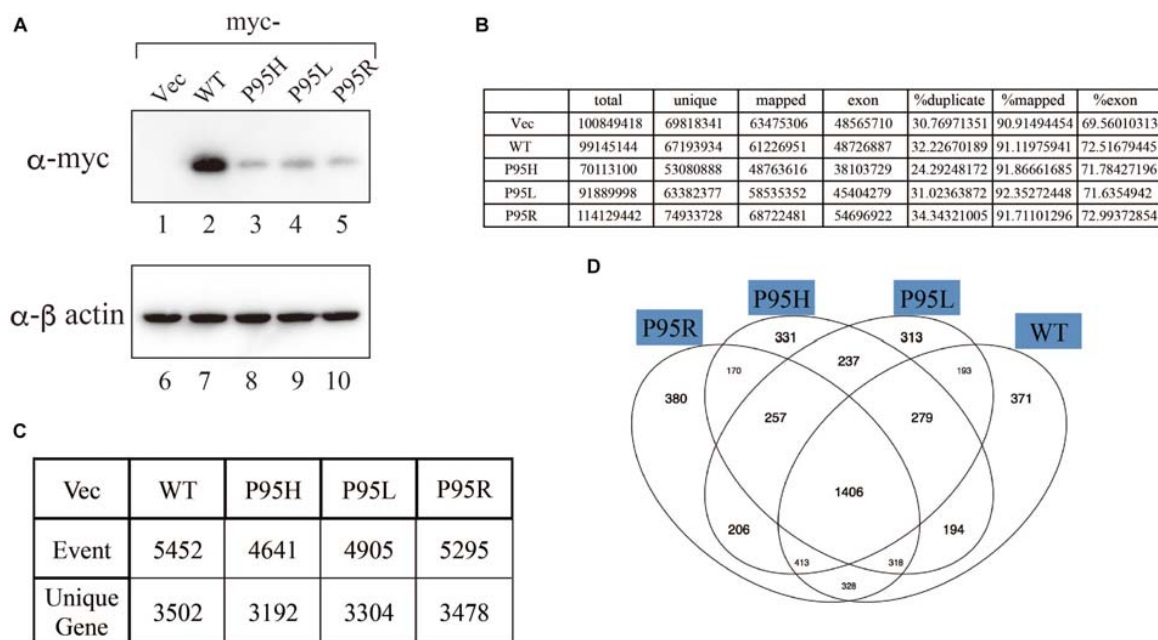


FIGURE 2 | Analyses of splicing changes in K562 cells stably expressing either wild-type or mutant SRSF2 protein. **(A)** Immunoblotting analyses of K562 cell lines stably expressing myc-tagged wild-type or mutant SRSF2 proteins. A myc-vector was used as a negative control. **(B)** Numbers of reads from RNA sequence analyses and mapping percentage of sequence tags. **(C)** Numbers of altered splicing events and genes identified by comparing with myc-vector transfected cells. **(D)** Venn diagram comparing the differentially spliced genes among myc-vector, WT and three SRSF2 mutants expressing cells.

motif was identified in inclusion-promoted exon at the top (**Figure 3B**). These results strongly suggested that purine-rich motif and CCWG-containing motif function as ESE for wild-type and mutants, respectively, and MDS-causing mutations alter the high affinity of SRSF2 from purine-rich motif to CCWG motif.

We have confirmed the splicing change with wild-type and mutant proteins by choosing several genes to determine splicing changes by RT-PCR. Among them, the splicing change of EZH2 gene is shown in **Figure 3D**. EZH2 is also known as one of the responsible genes for causing MDS (Ernst et al., 2010). With RNA sequencing analysis, the number of reads for exon 9.5, which has two CCWG motifs and premature termination codons (**Figure 3C**), was reduced with SRSF2 wild-type expression. In contrast, mutant protein expression increased the numbers of exon 9.5 reads, suggesting that mutant proteins promote exon 9.5 inclusion whereas wild-type enhances skipping of this exon. To test this possibility, we carried out RT-PCR analysis by amplifying Exon9–Exon10 region of EZH2 mRNA. The results indicate that wild-type protein expression reduced the ratio of exon 9.5 included mRNA (**Figure 3D**, lane 2). In contrast, all of the mutant proteins increased exon 9.5-included mRNA ratio (**Figure 3D**, lanes 3–5). These results indicate that SRSF2 wild-type and mutants have an opposite effect on CCWG-containing exon splicing.

DISCUSSION

In this manuscript we have analyzed the splicing changes in K562 cells stably expressing either wild-type or mutant SRSF2

proteins. As expected, both wild-type and mutants affected many splicing events of various genes, and CCWG motif was found in inclusion promoted exons with mutant proteins. CCNG or GGNG motif was previously reported as the binding sequences of SRSF2 protein by SELEX (Cavaloc et al., 1999). Indeed, three mutant proteins promote CCWG-motif containing pseudo exon of EZH2, which is one of the responsible genes for MDS (**Figure 3D**) (Ernst et al., 2010). This mis-splicing may affect EZH2 protein level in cells and the reduction of EZH2 protein level would result in mis-regulation of epigenetics (Kim et al., 2015; Shirahata-Adachi et al., 2017; Kon et al., 2018; Shiozawa et al., 2018). All three SRSF2 mutants were able to support SRSF2-dependent substrate (IgM) splicing *in vitro* more efficiently and associate with IgM pre-mRNA and mRNA more strongly than wild-type protein (**Figure 1B**). There are several CCWG motifs in 3' exon of IgM pre-mRNA (Mayeda et al., 1999). It is highly likely mutant proteins bind to these motifs more strongly than wild-type (**Figure 4**). On the other hand, the AGGTRAG motif was identified in exons excluded by wild-type SRSF2 (**Figure 3A**). In this motif, purine-stretch is separated by T (U in RNA). It was demonstrated that U residue splitting purine-stretch in ESE abolishes splicing-promoting activity *in vitro* (Watakabe et al., 1989). This motif may bind to SRSF2 with low affinity. Alternatively, it was demonstrated that SR protein overexpression can cause exon skipping which depends on their prevalent actions on a flanking constitutive exon, and it requires collaboration of more than one SR protein (Han et al., 2011). The AGGTRAG sequence might be a binding motif of 'weak' SR protein, not a direct binding motif for SRSF2 protein, to support exon inclusion. When overexpressed, wild-type SRSF2 protein can

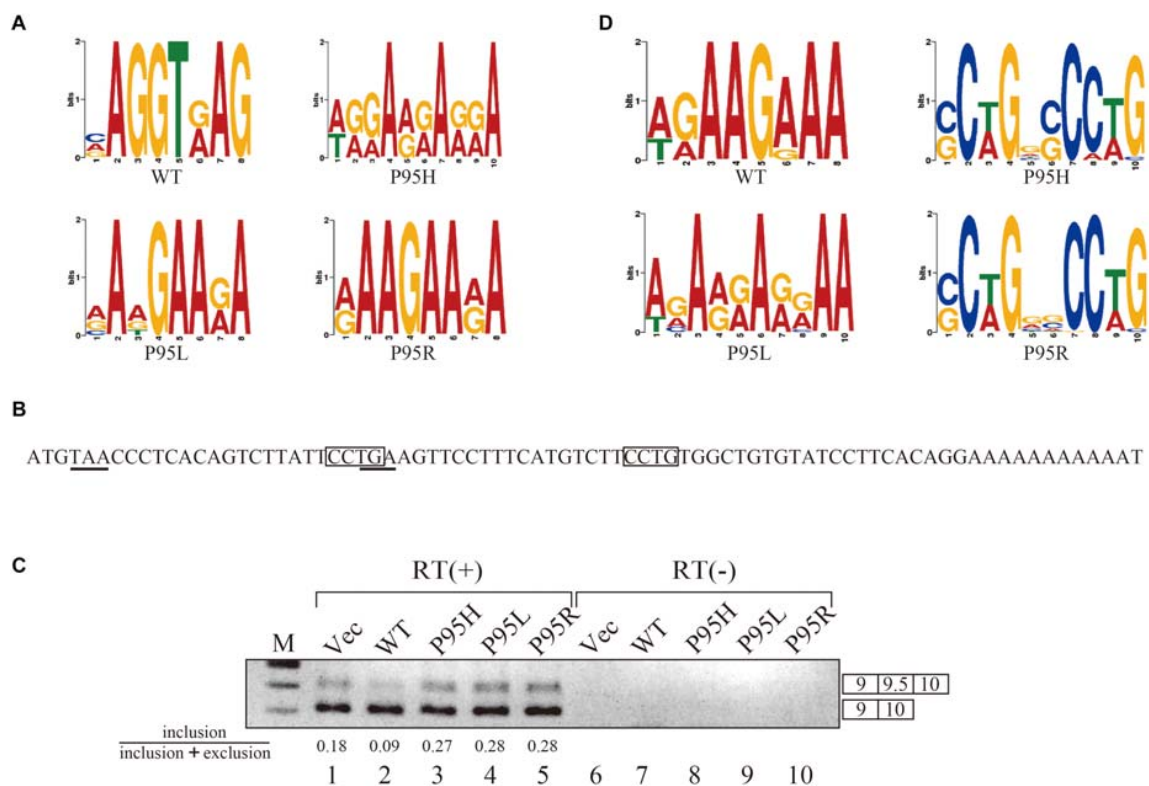


FIGURE 3 | SRSF2 mutants promote exon 9.5 of EZH2 gene through binding to CCWG motif. **(A)** Sequence logos for the motifs enriched in inclusion-type exons with WT, P95H, P95L, and P95R are presented. **(B)** Sequence logos for the motifs enriched in inclusion-type exons with WT, P95H, P95L, and P95R are shown. **(C)** Sequence of exon 9.5 (pseudo exon) in human EZH2 gene. In frame stop codons are underlined. CCWG motif, found with SRSF2 P95H and P95R proteins in **(B)**, are indicated with open boxes. **(D)** Splicing pattern of EZH2 pre-mRNA in myc-vector, SRSF2 WT, P95H, P95L, and P95R are indicated. The ratio of exon9.5 included product in total splicing products is shown beneath the panel. The schematic representations of PCR products are shown at the right side of the panel. RT, Reverse Transcriptase. M, DNA size marker.

bind to the flanking exon as a 'strong' SR protein more efficiently. The purine-rich ESE like sequence motifs were identified in exclusion promoted exons with mutant proteins (**Figure 3A**). Mutant proteins may have lower affinity to purine-rich sequences than wild-type (**Figure 4**). It is assumed that MDS mutations in Proline residue affect the conformation of RNA binding domain of SRSF2, although this proline residue is outside of RNA binding domain consensus. Indeed, 3D clustering analysis with protein structure predicted that this Proline residue is able to contact with RNA (Kamburov et al., 2015). Therefore, this residue is a part of RNA binding domain. By changing Proline residue to Histidine, Arginine or Leucine, the affinity with specific RNA sequence can be altered. Comparing crystal structures of these mutants-RNA complexes with that of wild-type would reveal the mechanism for the recognition of different RNA sequences. We believe this also uncover why mutant SRSF2 proteins have lower affinity to purine-rich ESEs.

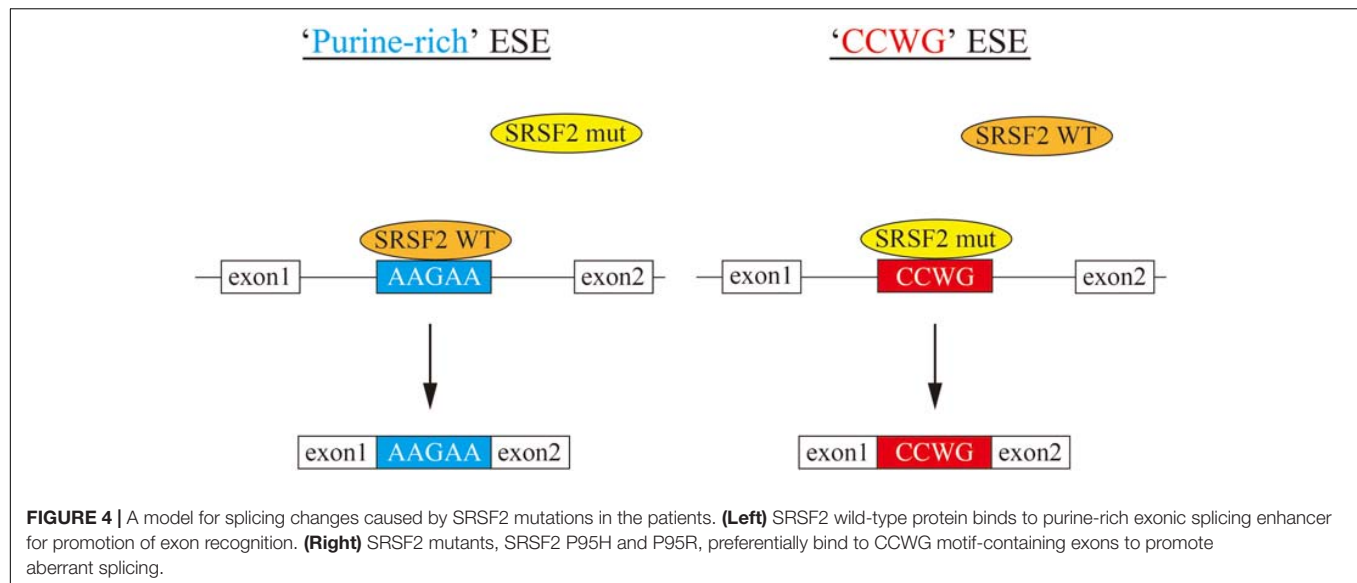
Despite of the common features described above, each mutant contains peculiar subset of genes whose splicing patterns were specifically changed (**Figure 2D**). These splicing changes in certain genes may confer the pathological difference among the MDS patients. Further analyses of specific splicing changes in each mutant are also required.

Most recently several groups demonstrated that MDS responsible mutations in SRSF2 and U2AF1 cause augment of R loop (Chen et al., 2018; Nguyen et al., 2018). Enhanced R loop formation activate the ataxia telangiectasia and Rad3-related protein (ATR)-Chk1 pathway, which likely contributes to MDS phenotype (Nguyen et al., 2018). Efficient formation of R loop can occur by slowing down rearrangement of mRNA-protein complexes during/after splicing. The different binding affinity of SRSF2 mutants to RNA may be involved in this step. It would be of a great interest in which portions of genes form R loops and whether enrichment of CCWG motif can be observed in those regions or not.

MATERIALS AND METHODS

Plasmid Construction

The cDNA of human SRSF2/SC35 was amplified by Reverse Transcription and Polymerase Chain Reaction (RT-PCR). The cDNA was cloned between BamHI and XhoI sites of either myc- or Flag-pCDNA3 vector. The resultant plasmid was used as a template in order to prepare mutant cDNAs that harbor MDS mutations, such as P95H, P95L and P95R. Point Mutations were



introduced by QuikChangeTM Site-Directed Mutagenesis Kit (STRATAGENE) with myc-SRSF2 plasmid in accordance with the manufacturer's recommendations.

Cell Culture and Establishment of Stable Cell Lines

K562 cells were cultured at 37°C with 5% CO₂ in RPMI1640 supplemented with 10% (v/v) fetal bovine serum (Sigma-Aldrich) and 1% (v/v) penicillin/streptomycin antibiotics (standard medium). K562 cells were transfected with pcDNA3-myc-SRSF2(WT), pcDNA3-myc-SRSF2(P95H), pcDNA3-myc-SRSF2(P95L), pcDNA3-myc-SRSF2(P95R) or empty vector plasmid, using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. Transfected K562 cells were selected with 100 mg/mL G418, and then resistant cells were isolated with limiting dilution methods in 96-well plates. Obtained each stable clone-cells were maintained in 50 mg/mL G418 RPMI medium.

For HEK293T cells, DMEM supplemented with 10% (v/v) fetal bovine serum (Sigma-Aldrich) and 1% (v/v) Antibiotics Antimycotics (SIGMA) was used for cell culture and Lipofectamine 2000 was also used for transfection. Total cell lysates from transfected HEK293T cells were prepared as previously described (Kataoka and Dreyfuss, 2004, 2008; Kataoka, 2016).

In vitro Transcription, Splicing and Immunoprecipitation of RNAs

For *in vitro* transcription, pC3-C4 plasmid linearized by HindIII was used as a template and performed as described previously (Watakabe et al., 1989, 1993). *In vitro* splicing reaction with HeLa cell nuclear extracts complemented with Flag-SRSF2 protein expressing HEK293T total cell lysates was carried out as previously described (Kataoka et al., 2001, 2011; Kim et al., 2001; Kataoka and Dreyfuss, 2004; Kawano et al., 2004; Kataoka, 2016).

Immunoprecipitation of RNAs from splicing reaction was accomplished by using anti-Flag M2 agarose beads (SIGMA) by the protocol described previously (Kataoka et al., 2001, 2011; Kataoka and Dreyfuss, 2004; Kataoka, 2016).

Antibodies, Western Blotting and Immunostaining of HeLa Cells

The antibodies used for immunoblotting and immunostaining are as follows: anti-myc (MC045, Nacalai Tesque, Japan), anti-Flag M2 (Sigma-Aldrich), fluorescein isothiocyanate-conjugated goat anti-mouse F(ab')₂ (Cappel Laboratories, Durham, NC, United States), peroxidase-conjugated goat anti-mouse IgG antibodies (Jackson Immuno Research Laboratories, West Grove, PA, United States). For western blotting, the cells were lysed in CelLytic M Cell Lysis Reagent (Sigma-Aldrich) containing a protease inhibitor cocktail (Roche). The lysates were boiled with SDS-sample buffer at 95°C for 3 min. The samples were subjected to SDS-PAGE, transferred to PVDF membranes by iBlot system (Invitrogen), and incubated with primary antibodies. The membranes were washed and incubated with horseradish peroxidase-conjugated secondary antibody. Finally, chemiluminescence was detected using Chemi-Lumi One Super kit (Nacalai Tesque), and luminescence images were analyzed by LAS 4000 (GE Lifesciences). Immunostaining of HeLa cells with anti-myc antibody was performed as described previously (Kataoka et al., 2011).

RNA Recovery, RNA Sequence and Alternative Splicing Analysis

Total RNAs from K562 stable cells were performed by using RNeasy Mini Kit (QIAGEN). The synthesis and amplification of complementary DNA were performed using SMARTer Ultra Low Input RNA Kit for Sequencing, version 3 (Clontech). Each sample applied illumina GA. The reads were trimmed to 99 bases and were mapped on hg19 genomes and gencode v7 protein

coding transcripts by tophat (v2.0.9). To compute junctions in each sample, we processed the gencode GTF file by eval package (v2.2.8) and applied the mapped reads to juncBASE (v0.6) packages by following the options: ‘-min overhang = 6 -l 99 -c 3.’ For comparing each sample, we calculated “Percent Splicing index” (PSI) and corrected *p*-value and Benjamini & Hedgehog multiple test using pairwise fishers test getASEvents w reference.py by following options: ‘-jcn seq len = 186 -method = BH’ comparing with vector and wild-type.

Motif Analysis

To search RNA binding motif of mutated SRSF2, we collected 100 base exon side sequences of exclusion junctions separating upper or lower than vector’s PSI in each sample and the upper and lower sequences was applied MEME v4.10.0 as following that options: ‘-minw 4-maxw 10 -maxsize 1000000.’

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RT-PCR reaction was accomplished as described previously (Wang et al., 2017). Briefly, 1 µg of total RNA was used for reverse transcription with prime Script Reverse Transcriptase (TAKARA, Japan). The produced cDNA was used for PCR reaction by using the following primers and PCR cycles. Cycle conditions were as follows: 94°C for 2 min; followed by 33 cycles of 94°C denaturation for 10 s, 58°C annealing for 15 s, and 72°C elongation for 30 s; with a final incubation at 72°C for 2 min in a PCR Thermal Cycler (BIOMETRA). PCR products were separated by electrophoresis and stained with ethidium bromide. The primers for PCR are as follows; EZH2 exon9 F, 5’-AAGCGGAAGAACACAGAAAC-3’, EZH2 exon10 R, 5’-CAGAGGAGCTCGAAGTTTCA-3’. For quantitation analysis

of alternative splicing products, the signals were measured by ImageJ software [U.S. National Institutes of Health, Bethesda (Schneider et al., 2012)].

AUTHOR CONTRIBUTIONS

SM and NK started this project, conceived and designed the experiments. SM, AH, YS, AK, and NK performed the experiments including RNA sequencing. SI and NK analyzed RNA sequencing results. SM, SI, AH, YS, AK, SO, FH, S-IT, and NK analyzed the data. SM, SI, AH, YS, AK, KS, FH, S-IT, and NK contributed reagents, materials, and analysis tools. SM and NK wrote the manuscript. NK took the primary responsibility for the final content. SM, SI, AH, YS, AK, SO, KS, FH, S-IT, and NK read and approved the final manuscript.

FUNDING

This work was supported by Grants-in-Aid for Scientific Research (23112706 to SM, 18K06012 to NK). This work was also supported in part by research funding from Dainippon Sumitomo Pharma Co., Ltd.

ACKNOWLEDGMENTS

We would like to thank the members of Takahashi lab and Dr. Aiko Sugiyama for helpful discussion and support for experiments. We are grateful to Dr. Gideon Dreyfuss (University of Pennsylvania) for kind gifts of myc- and Flag-pCDNA3 vectors.

REFERENCES

- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Brooks, A. N., Choi, P. S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., et al. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS One* 9:e87361. doi: 10.1371/journal.pone.0087361
- Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* 5, 468–483. doi: 10.1017/s1355838299981967
- Cazzola, M., Della Porta, M. G., and Malcovati, L. (2013). The genetic basis of myelodysplasia and its clinical relevance. *Blood* 122, 4021–4034. doi: 10.1182/blood-2013-09-381665
- Chen, L., Chen, J. Y., Huang, Y. J., Gu, Y., Qiu, J., Qian, H., et al. (2018). The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell.* 69:e6. doi: 10.1016/j.molcel.2017.12.029
- 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
- Darman, R. B., Seiler, M., Agrawal, A. A., Lim, K. H., Peng, S., Aird, D., et al. (2015). Cancer-associated SF3B1 hotspot mutations induce cryptic 3’ splice site selection through use of a different branch point. *Cell Rep.* 13, 1033–1045. doi: 10.1016/j.celrep.2015.09.053
- Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* 42, 722–726. doi: 10.1038/ng.621
- Fu, X. D., and Ares, M. Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. doi: 10.1038/nrg3778
- Fu, X. D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* 343, 437–441. doi: 10.1038/343437a0
- Graubert, T. A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C. L., Shao, J., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* 44, 53–57. doi: 10.1038/ng.1031
- Haeflrich, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., and Nagae, G. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28, 241–247. doi: 10.1038/leu.2013.336
- Han, J., Ding, J. H., Byeon, C. W., Kim, J. H., Hertel, K. J., Jeong, S., et al. (2011). SR proteins induce alternative exon skipping through their activities on the flanking constitutive exons. *Mol. Cell. Biol.* 4, 793–802. doi: 10.1128/MCB.01117-10
- Howard, J. M., and Sanford, J. R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* 6, 93–110. doi: 10.1002/wrna.1260
- Ilagan, J. O., Ramakrishnan, A., Hayes, B., Murphy, M. E., Zebari, A. S., Bradley, P., et al. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res.* 25, 14–26. doi: 10.1101/gr.181016.114
- Kamburov, A., Lawrence, M. S., Polak, P., Leshchiner, I., Lage, K., Golub, T. R., et al. (2015). Comprehensive assessment of cancer missense mutation clustering in

- protein structures. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5486–E5495. doi: 10.1073/pnas.1516373112
- Kataoka, N. (2016). Purification of RNA-protein splicing complexes using a tagged protein from in vitro splicing reaction mixture. *Methods Mol. Biol.* 1421, 45–52. doi: 10.1007/978-1-4939-3591-8_5
- Kataoka, N. (2017). Modulation of aberrant splicing in human RNA diseases by chemical compounds. *Hum. Genet.* 136, 1237–1245. doi: 10.1007/s00439-017-1789-4
- Kataoka, N., Diem, M. D., Kim, V. N., Yong, J., and Dreyfuss, G. (2001). Magoh, a human homolog of *Drosophila mago nashi* protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J.* 20, 6424–6433. doi: 10.1093/emboj/20.22.6424
- Kataoka, N., Diem, M. D., Yoshida, M., Hatai, C., Dobashi, I., and Dreyfuss, G. (2011). Specific Y14 domains mediate its nucleo-cytoplasmic shuttling and association with spliced mRNA. *Sci. Rep.* 1:92. doi: 10.1038/srep00092
- Kataoka, N., and Dreyfuss, G. (2004). A simple whole cell lysate system for in vitro splicing reveals a stepwise assembly of the exon-exon junction complex. *J. Biol. Chem.* 279, 7009–7013. doi: 10.1074/jbc.M307692200
- Kataoka, N., and Dreyfuss, G. (2008). Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. *Methods Mol. Biol.* 488, 357–365. doi: 10.1007/978-1-60327-475-3_23
- Kawano, T., Kataoka, N., Dreyfuss, G., and Sakamoto, H. (2004). Ce-Y14 and MAG-1, components of the exon-exon junction complex, are required for embryogenesis and germline sexual switching in *Caenorhabditis elegans*. *Mech. Dev.* 121, 27–35. doi: 10.1016/j.mod.2003.11.003
- Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., and Ramakrishnan, A. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* 27, 617–630. doi: 10.1016/j.ccell.2015.04.006
- Kim, V. N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* 293, 1832–1836. doi: 10.1126/science.1062829
- Komono, Y., Huang, Y. J., Qiu, J., Lin, L., Xu, Y., and Zhou, Y. (2015). SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. *Mol. Cell. Biol.* 35, 3071–3082. doi: 10.1128/MCB.00202-15
- Kon, A., Yamazaki, S., Nannya, Y., Kataoka, K., Ota, Y., and Nakagawa, M. M. (2018). Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood* 131, 621–635. doi: 10.1182/blood-2017-01-762393
- Manley, J. L., and Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 24, 1073–1074. doi: 10.1101/gad.1934910
- Mayeda, A., Screaton, G. R., Chandler, S. D., Fu, X. D., and Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.* 19, 1853–1863. doi: 10.1128/mcb.19.3.1853
- Mupo, A., Seiler, M., Sathiaselan, V., Pance, A., Yang, Y., and Agrawal, A. A. (2017). Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia* 31, 720–727. doi: 10.1038/leu.2016.251
- Nguyen, H. D., Leong, W. Y., Li, W., Reddy, P. N. G., Sullivan, J. D., Walter, M. J., et al. (2018). Spliceosome mutations induce r loop-associated sensitivity to ATR Inhibition in myelodysplastic syndromes. *Cancer Res.* 78, 5363–5374. doi: 10.1158/0008-5472.CAN-17-3970
- Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., et al. (2016). Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell* 30, 404–417. doi: 10.1016/j.ccell.2016.08.006
- Papaemmanuil, E., Cazzola, M., Boultonwood, J., Malcovati, L., Vyas, P., Bowen, D., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* 365, 1384–1395. doi: 10.1056/NEJMoa1103283
- Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616–3627. doi: 10.1182/blood-2013-08-518886
- Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M. A., Padgett, R., Maciejewski, J. P., et al. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* 122, 999–1006. doi: 10.1182/blood-2013-01-480970
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., et al. (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* 21, 997–1005. doi: 10.1038/nsmb.2906
- Shiozawa, Y., Malcovati, L., Galli, A., Sato-Otsubo, A., Kataoka, K., Sato, Y., et al. (2018). Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat. Commun.* 9:3649. doi: 10.1038/s41467-018-06063-x
- Shirahata-Adachi, M., Iriyama, C., Tomita, A., Suzuki, Y., Shimada, K., and Kiyoi, H. (2017). Altered EZH2 splicing and expression is associated with impaired histone H3 lysine 27 tri-Methylation in myelodysplastic syndrome. *Leuk. Res.* 63, 90–97. doi: 10.1016/j.leukres.2017.10.015
- Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., et al. (2015). Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. *Cancer Cell* 27, 631–643. doi: 10.1016/j.ccell.2015.04.008
- Walter, M. J., Shen, D., Shao, J., Ding, L., White, B. S., Kandath, C., et al. (2013). Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* 27, 1275–1282. doi: 10.1038/leu.2013.58
- Wang, D. O., Ninomiya, K., Mori, C., Koyama, A., Haan, M., Kitabatake, M., et al. (2017). Transport granules bound with nuclear cap binding protein and exon junction complex are associated with microtubules and spatially separated from eIF4E granules and p bodies in human neuronal processes. *Front. Mol. Biosci.* 4:93. doi: 10.3389/fmolb.2017.00093
- Watakabe, A., Inoue, K., Sakamoto, H., and Shimura, Y. (1989). A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin mu chain pre-mRNAs. *Nucleic Acids Res.* 17, 8159–8169. doi: 10.1093/nar/17.20.8159
- Watakabe, A., Tanaka, K., and Shimura, Y. (1993). The role of exon sequences in splice site selection. *Genes Dev.* 7, 407–418. doi: 10.1101/gad.7.3.407
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64–69. doi: 10.1038/nature10496
- Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., et al. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4726–E4734. doi: 10.1073/pnas.1514105112

Conflict of Interest Statement: 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.

Copyright © 2019 Masaki, Ikeda, Hata, Shiozawa, Kon, Ogawa, Suzuki, Hakuno, Takahashi and Kataoka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



HMGA1a Induces Alternative Splicing of the *Estrogen Receptor- α* Gene by Trapping U1 snRNP to an Upstream Pseudo-5' Splice Site

Kenji Ohe^{1*}, Shinsuke Miyajima², Tomoko Tanaka³, Yuri Hamaguchi³, Yoshihiro Harada¹, Yuta Horita¹, Yuki Beppu¹, Fumiaki Ito¹, Takafumi Yamasaki¹, Hiroki Terai¹, Masayoshi Mori¹, Yusuke Murata¹, Makito Tanabe³, Ichiro Abe⁴, Kenji Ashida⁵, Kunihisa Kobayashi⁴, Munechika Enjoji¹, Takashi Nomiyama³, Toshihiko Yanase³, Nobuhiro Harada⁶, Toshiaki Utsumi² and Akila Mayeda^{7*}

OPEN ACCESS

Edited by:

Emanuele Buratti,
International Centre for Genetic
Engineering and Biotechnology, Italy

Reviewed by:

Claudia Ghigna,
Istituto di Genetica Molecolare (IGM),
Italy
Rosanna Asselta,
Humanitas Università, Italy

*Correspondence:

Kenji Ohe
ohekenji@fukuoka-u.ac.jp
Akila Mayeda
mayeda@fujita-hu.ac.jp

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 30 March 2018

Accepted: 22 May 2018

Published: 08 June 2018

Citation:

Ohe K, Miyajima S, Tanaka T, Hamaguchi Y, Harada Y, Horita Y, Beppu Y, Ito F, Yamasaki T, Terai H, Mori M, Murata Y, Tanabe M, Abe I, Ashida K, Kobayashi K, Enjoji M, Nomiyama T, Yanase T, Harada N, Utsumi T and Mayeda A (2018) HMGA1a Induces Alternative Splicing of the Estrogen Receptor- α Gene by Trapping U1 snRNP to an Upstream Pseudo-5' Splice Site. *Front. Mol. Biosci.* 5:52. doi: 10.3389/fmolb.2018.00052

¹ Department of Pharmacotherapeutics, Faculty of Pharmaceutical Sciences, Fukuoka University, Fukuoka, Japan,

² Department of Breast Surgery, Fujita Health University, Toyoake, Japan, ³ Department of Endocrinology and Diabetes Mellitus, Faculty of Medicine, Fukuoka University, Fukuoka, Japan, ⁴ Department of Endocrinology and Diabetes Mellitus, Fukuoka University Chikushi Hospital, Chikushino, Japan, ⁵ Department of Medicine and Bioregulatory Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, ⁶ Department of Biochemistry, Fujita Health University, Toyoake, Japan, ⁷ Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan

Objectives: The high-mobility group A protein 1a (HMGA1a) protein is known as a transcription factor that binds to DNA, but recent studies have shown it exerts novel functions through RNA-binding. We were prompted to decipher the mechanism of HMGA1a-induced alternative splicing of the estrogen receptor alpha (ER α) that we recently reported would alter tamoxifen sensitivity in MCF-7 TAMR1 cells.

Methods: Endogenous expression of full length ER α 66 and its isoform ER α 46 were evaluated in MCF-7 breast cancer cells by transient expression of HMGA1a and an RNA decoy (2'-O-methylated RNA of the HMGA1a RNA-binding site) that binds to HMGA1a. RNA-binding of HMGA1a was checked by RNA-EMSA. *In vitro* splicing assay was performed to check the direct involvement of HMGA1a in splicing regulation. RNA-EMSA assay in the presence of purified U1 snRNP was performed with psoralen UV crosslinking to check complex formation of HMGA1a-U1 snRNP at the upstream pseudo-5' splice site of exon 1.

Results: HMGA1a induced exon skipping of a shortened exon 1 of ER α in *in vitro* splicing assays that was blocked by the HMGA1a RNA decoy and sequence-specific RNA-binding was confirmed by RNA-EMSA. RNA-EMSA combined with psoralen UV crosslinking showed that HMGA1a trapped purified U1 snRNP at the upstream pseudo-5' splice site.

Conclusions: Regulation of ER α alternative splicing by an HMGA1a-trapped U1 snRNP complex at the upstream 5' splice site of exon 1 offers novel insight on 5' splice site regulation by U1 snRNP as well as a promising target in breast cancer therapy where alternative splicing of ER α is involved.

Keywords: estrogen receptor alpha, HMGA1a, alternative splicing, U1 snRNP, breast cancer

INTRODUCTION

Cancer-associated alternative splicing has been extensively studied in various steps from tumor initiation to progression and metastasis (Oltean and Bates, 2014; Chen and Weiss, 2015). These cancer-associated alternative splicing events are aberrantly regulated by multifunctional bona fide splicing factors [serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs)] and tissue-specific RNA-binding proteins (David and Manley, 2010) possessing oncogenic potential *per se* (Karni et al., 2007). In addition, several recent reports have evoked attention on the importance of alternative splicing of the clinically important biomarkers in breast cancer (Inoue and Fry, 2015), and the deregulation and involvement of splicing factors in breast cancer-associated alternative splicing (Silipo et al., 2015).

Estrogen receptor alpha (ER α) is an important biomarker as well as the key factor in estrogen-dependent growth of breast cancer. We have recently found that high-mobility group A protein 1a (HMGA1a) is involved in alternative splicing of ER α (Ohe et al., 2018). The resulting transcript, called ER α 46 (Flouriou et al., 2000), is known for its function in partially inhibiting mitogenic activity of full length ER α (Penot et al., 2005).

HMGA1a is originally known as HMGI (Lund et al., 1983), a nonhistone DNA-binding protein of the UBF/HMG family, and an oncoprotein which induces cancerous transformation (Reeves, 2010). In addition to its DNA-binding properties exerted by its three AT hooks, we have shown HMGA1a protein binds to a specific RNA sequence of 5'-GC(U)GCUACAAG-3', adjacently upstream the authentic 5' splice site of exon 5 in the *Presenilin-2* (PS2) gene, interacts with U1-70K (Manabe et al., 2003), and traps U1 snRNP to this 5' splice site to inhibit normal

dissociation of U1 snRNP from the spliceosome and induces aberrant splicing of PS2 exon 5 (Ohe and Mayeda, 2010). This effect was also found in HIV-1 splice site regulation (Tsuruno et al., 2011). A report of an aptamer search for the second AT hook motif of HMGA1 proteins shows binding to a G-rich motif of 5'-GGGGNGNGGNUGGGGNGG-3' (Maasch et al., 2010). Another study shows HMGA1a binds the second loop of 7SK RNA (5'-UGCGC-3') (Eilebrecht et al., 2011b). Recently, it has been reported that AT hook proteins such as HMGA1a bind RNA with one order higher affinity than DNA (Filarsky et al., 2015). Proteomics studies have reported that HMGA1a interacts with mRNA processing proteins by GST pull-down and farwestern experiments (Sgarra et al., 2005, 2008; Pierantoni et al., 2007). Thus, the recent description of the role of HMGA1a in RNA metabolism in addition to its role as a DNA-binding transcription factor make it more multi-functional protein than previously believed.

MATERIALS AND METHODS

Plasmids, Antisense Oligonucleotides, 2'-O-Methyl RNAs, Recombinant Protein, Purified U1 snRNP

Plasmids were constructed as indicated below.

Plasmid for CDC-ER α pre-mRNA was constructed by nearby exon and intron sequences of the splice sites of ER α exon 1 attached to the 5' ends of the primers which were used to amplify CDC14-15 (chicken delta crystallin exon 14-15) (Sawa et al., 1988; Kataoka et al., 2000) by PCR. The primers used for this PCR reaction were δ ESR α 1_FW and δ ESR α 1_RV. The PCR fragment was digested with Hind III and self-ligated resulting in pSP64-CDC-ER α with shortened ER α exon 1 (58 bp) and

TABLE 1 | Oligonucleotides used in this study.

δ ESR α 1_FW	5'- CGCAAGCTT GCGGCTACACGGTGC CGAGCGCGGCCCGCCGGCATTCTACAGTACCCGCGtccccagggagcagcaaa-3'
δ ESR α 1_RV	5'- CGCAAGCTT GCGGGCCAC CTGGAAAAAGAGCACAGCCCGAGGTTAGAGGCTgtcctccaaccacattaataatgagaa-3'
FW:forward, RV:reverse	Shortened ER α exon 1 (bold) and flanking intron sequence in upper case (Hind III is italicized) and CDC14-15 intron sequence is in lower case
ER α Ex1-5SS_wt_FW	5'- ATACAAGCTTCCCAGCGGCTACACG GT GCGCGAGGCGCGCCCGCCGGCATTCTACAGGtaccgcgGAATTCGTAATCATGGTCA TAGCTG-3'
ER α Ex1-5SS_HMGBSmut_FW	5'- ATACAAGCTTCCCAGCG AGAT ACG GT GCGCGAGGCGCGCCCGCCGGCATTCTACAGGtaccgcgGAATTCGTAATCATGGTCA TAGCTG-3'
ER α Ex1-5SS_A5SSmut_FW	5'- ATACAAGCTTCCCAGCGGCTACACG GT GCGCGAGGCGCGCCCGCCGGCATTCTACT Cct accgcgGAATTCGTAATCATGGTCA TAGCTG-3'
ER α Ex1-5SS_P5SSmut_FW	5'- ATACAAGCTTCCCAGCGGCTACAC TCC GCGCGAGGCGCGCCCGCCGGCATTCTACAGGtaccgcgGAATTCGTAATCATGGTCA TAGCTG-3'
ER α Ex1-5SS_RV	5'-CCCAGCTTGTATTCTATAGTGTACCTAAATCGTATGT-3'
FW:forward, RV:reverse	Hind III site is italicized, HMGA1a binding sequence is underlined, intron sequence in lower case, GU of pseudo 5' splice site is in bold, mutated nucleotides are in red.
α P5'SS	5'-CTCGCGCACCGT-3'
α A5'SS	5'-CGCGGGTACCTG-3'
α U1	5'-CTGCCAGGTAAGTAT-3'
α U2	5'-AGGCCGAGAAGCGAT-3'
cont	5'-AGGGAGTATGTGAATGCC-3'

flanking intron sequence placed in the intron of CDC14-15. pSP64-CDC-ER α plasmid was linearized using SmaI for RNA synthesis.

The riboprobe template plasmids for ER α Ex1-5SS_wt, ER α Ex1-5SS_HMGBSmut, ER α Ex1-5SS_A5SSmut, ER α Ex1-5SS_P5SSmut were constructed by PCR using pSP64 as template and digested by HindIII. The forward primers were ER α Ex1-5SS_wt_FW, ER α Ex1-5SS_HMGBSmut_FW, α Ex1-5SS_A5SSmut_FW, ER α Ex1-5SS_P5SSmut_FW, and the reverse primer is ER α Ex1-5SS_RV (Table 1). The HindIII-digested PCR products were self-ligated at the Hind III site to obtain the pSP64-ER α Ex1-5SS_wt, pSP64-ER α Ex1-5SS_HMGBSmut, pSP64-ER α Ex1-5SS_A5SSmut, pSP64-ER α Ex1-5SS_P5SSmut plasmids. These plasmids were linearized using EcoRI.

All of the linearized plasmids were transcribed with SP6 RNA polymerase in the presence of [α - 32 P]UTP. The sequences for antisense oligonucleotides used for RNase H cleavage of the upstream pseudo-5' splice site (α P5'SS) and authentic-5' splice site (α A5'SS) and the ones used for digesting 5' ends of U1 (α U1) and U2 (α U2) snRNA, as well as the control oligonucleotide (cont), are shown in Table 1. The sequences for 2'-O-methyl RNA_HMGA1a_mut and 2'-O-methyl RNA_HMGA1a_wt that were purchased from Fasmac Co., Ltd. are also shown in Table 2. Recombinant HMGA1a was a kind gift from K. Kanameki and Y. Muto (Musashino, Tokyo, Japan) purified from *E. coli* and microdialyzed in buffer D (Mayeda and Krainer, 1999b) for *in vitro* splicing analyses. Purified U1 snRNP was kindly provided by the laboratory of R. Lührmann (Göttingen, Germany).

Electrophoretic Mobility Shift Assays (EMSA)

EMSA was performed as previously described (Ohe and Mayeda, 2010) with the following minor modifications. Each reaction mixture was incubated at 30°C for 15 min and RNA-protein complexes were analyzed by 5% polyacrylamide gel electrophoresis (PAGE) (acrylamide: bisacrylamide ratio 30:1 [wt/wt]) at 4°C.

In Vitro Splicing, Psoralen UV Crosslinking Assay

In vitro splicing was performed as described (Ohe and Mayeda, 2010), with minor modifications. 32 P-labeled pre-mRNA (~20 fmol) was incubated at 30°C for 2 h in a 12.5 μ l reaction mixture containing 3 mM ATP, 20 mM creatine phosphate, 20 mM HEPES-NaOH (pH 7.3), 3.5 mM MgCl₂, 2% (wt/vol) low-molecular-weight polyvinyl alcohol (Sigma), and 3.5 μ l of HeLa cell nuclear extract (CilBiotech). Recombinant protein was added first to the probe prior adding nuclear extract. Where indicated 20 pmol of SRSF1 was added. Psoralen-mediated UV crosslinking assay was performed as previously described fixing the incubation time at 5 min (Ohe and Mayeda, 2010).

Immunoblot Assays

MCF-7 cells were washed twice with PBS and resuspended in 1 ml TRIzol reagent (Invitrogen). After passing through a 25 G needle five times, protein was purified from the interphase and organic phase by precipitating with 6 V of a solution for precipitating

TABLE 2 | 2'-O-methyl RNA used in this study.

2'-O-methyl RNA_HMGA1a_mut	5'-CGCCGAGAUCAAG-3'
2'-O-methyl RNA_HMGA1a_wt	5'-CGCUGCUACAAG-3'

protein (50% Ethanol, 24.5% Acetone, 24.5% Methanol, 1% distilled water). Twenty micrograms of protein was boiled in sample buffer and separated by 10% sodium dodecyl sulfate-PAGE (SDS-PAGE), transferred to nitrocellulose membrane and analyzed by an ER α antibody (HC-20) (Santa-Cruz, sc-543) which recognizes the C-terminus of the protein or HMGA1a antibody (FL95; Santa Cruz). Anti-rabbit immunoglobulin G conjugated to alkaline phosphatase (Promega) was used as secondary antibody and detected by BCIP (5-bromo-4-chloro-3-indolylphosphate) (Promega).

Statistical Analysis

Data are presented as mean \pm SD. Group differences were analyzed by Students *t*-test using Microsoft Excel.

RESULTS

HMGA1a Binds a Sequence in ER α Exon 1

In our previous reports, we showed that HMGA1a binds to RNA in sequence-specific manner (Manabe et al., 2003, 2007; Ohe and Mayeda, 2010). Here, we searched for other binding sites in other diseases. Since HMGA1a has been intensely studied in breast cancer as a transcription factor and that its expression correlates with its malignant potential, we were motivated to seek whether its RNA-binding characteristics have any involvement in the development of the disease. Accordingly, we found a candidate HMGA1a RNA-binding site in the *estrogen receptor alpha* (ER α) gene. It functioned in MCF-7 cells, induced estrogen-dependent growth in these cells as well as nude mice, and tamoxifen-resistant MCF-7 TAMR1 cells were sensitized to tamoxifen (Ohe et al., 2018). Here we show the *in vitro* analyses of HMGA1a-trapped U1 snRNP at the upstream pseudo 5' splice site adjacent the HMGA1a RNA-binding site in ER α exon1, which is located downstream of several non-coding exons.

The candidate HMGA1a RNA-binding site was located 33 nucleotides upstream the authentic 5' splice site (Figure 1A: HMGA1aBS-wt) of ER α exon 1. The 5' splice site score of the adjacent pseudo-5' splice site (MaxEnt: 8.67; Yeo and Burge, 2004) is comparable to the authentic 5' splice site (MaxEnt: 8.63) of this exon. The HMGA1a RNA-binding candidate sequence is 5'-GCGGCUACACG-3', a two-base mismatch of the original one we found previously, 5'-GCGGCUACAAG-3' (Manabe et al., 2003, 2007; Ohe and Mayeda, 2010) (mismatch underlined). RNA electrophoretic mobility shift assay (EMSA) (Figure 1B).

HMGA1a Induces Exon Skipping of a Shortened ER α Exon 1 in Vitro

ER α 46 has been reported to be expressed through differential regulation of translation (Barraille et al., 1999), thus we needed to exclude this possibility by testing HMGA1a in an *in vitro*

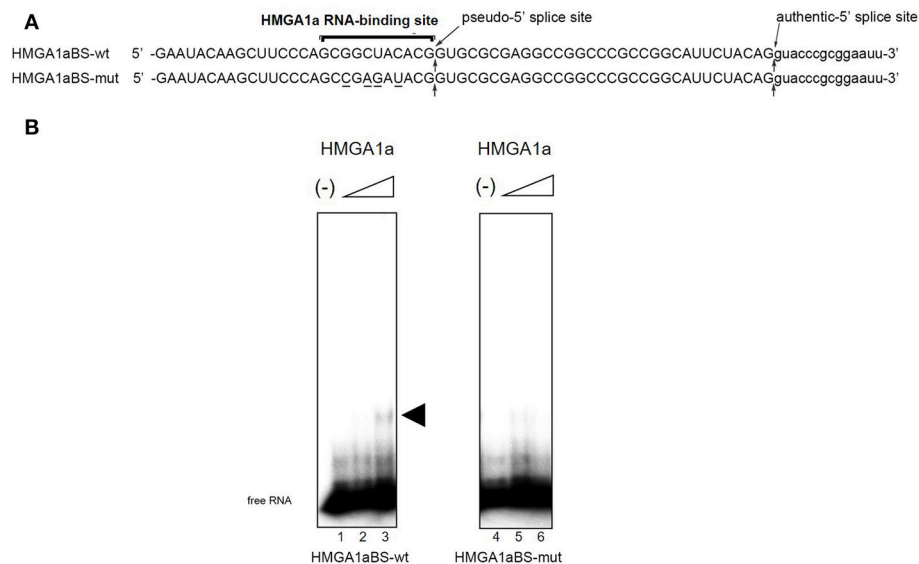


FIGURE 1 | HMGA1a regulates alternative splicing of the *ER α* gene in MCF-7 cells. **(A)** Sequences of the HMGA1a RNA-binding sequence in *ER α* exon 1a is indicated (HMGA1a_BSwt). Arrows show the pseudo-5' splice site and authentic 5' splice site. Mutated bases are underlined (HMGA1a_BSmut). **(B)** HMGA1a binding to 32 P-labeled RNAs were analyzed by RNA-EMSA. Increasing amounts of recombinant HMGA1a (0, 191, 383 ng in 15 μ l) are indicated by minus signs and triangles. The shifted HMGA1a bound 32 P-labeled RNA is indicated by arrowhead.

splicing assay where translational regulation is not observed. A heterologous pre-RNA transcript with the splice sites and flanking sequences of *ER α* exon 1 were inserted in the intron of a conventional splicing substrate, CDC14-15 (Sawa et al., 1988; Kataoka et al., 2000), designated CDC-ER α . Exon 1 was shortened for the limitations of length in splicing RNA *in vitro* (Mayeda and Krainer, 1999a). The RNA transcribed from CDC-ER α includes the HMGA1a RNA-binding sequence in a shortened *ER α* exon1 with flanking intron sequences inserted in intron between CDC exon 14 and 15 (Figure 2A). The splicing recapitulated that of cultured MCF-7 cells (Ohe et al., 2018), except that CDC-ER α showed exon skipping using HeLa nuclear extract in our *in vitro* splicing assay (Figure 2B, lane 2), thus we added SRSF1 and exon inclusion was observed (Figure 2B, lane 3). When HMGA1a was added to this reaction, an increase of exon exclusion was observed (Figure 2B, lane 4). Since HMGA1a is known to be highly expressed in HeLa cells, it is possible that the endogenous HMGA1a in HeLa nuclear extract induced exon skipping in our *in vitro* splicing assays before SRSF1 was added (Figure 2B, lane 2). Next, in order to observe the decoy effect of PS2 HMGA1a RNA-binding sequence, we extended the splicing reaction to 3 h without adding SRSF1. In this condition, no exon inclusion of CDC-ER α was observed (Figure 2B, lane 6). To confirm whether this exon skipping event was due to RNA-binding of HMGA1a, 2'-O-methyl RNA of the PS2 HMGA1a RNA-binding sequence (Manabe et al., 2007) was added to the reaction. While 2'-O-methyl RNA of mutant HMGA1a RNA-binding sequence showed no exon inclusion at 8 and 16 μ M (Figure 2B, lane 7,8), 2'-O-methyl RNA of wild-type sequence showed clear exon inclusion at the same concentration (Figure 2B, lane 9,10). We believe that

this induction of exon inclusion is significant in such limited exon inclusion conditions of this pre-mRNA substrate *in vitro*. The decoy effect of PS2 HMGA1a RNA-binding sequence was observed to inhibit alternative splicing induced by endogenous HMGA1a protein of MCF-7 cells (Ohe et al., 2018). Here we focus on HMGA1a-induced exon skipping of CDC-ER α and conducted further experiments to decipher the mechanism.

HMGA1a Anchors U1 snRNP to the Upstream Pseudo-5' Splice Site of *ER α* Exon 1

In our previous report, we showed HMGA1a prevents normal dissociation of U1 snRNP from the 5' splice site only when the 5' splice site is adjacently downstream the HMGA1a RNA-binding sequence (Ohe and Mayeda, 2010). In the case of *ER α* exon 1, HMGA1a binds adjacently upstream a pseudo-5' splice site located 33 nucleotides upstream of the authentic splice site (Figure 1). We tested whether HMGA1a could trap U1 snRNP to the pseudo-5' splice site and block U1 snRNP-binding to the authentic 5' splice site of *ER α* exon 1 by psoralen-mediated UV crosslinking assay (Figure 3B). We used the same RNA (HMGA1aBS-wt) in the RNA-EMSA experiments (Figure 1A), which contains the pseudo-5' splice site and authentic 5' splice site of *ER α* exon 1. Mutants of each 5' splice site (Figure 3, A5'SSmut: mutant of authentic 5' splice site, P5'SSmut: mutant of pseudo-5' splice site) were used to define each U1 snRNP/5' splice site crosslink. When purified U1 snRNP was added (5-min incubation), two crosslinks with fast and slow mobility were detected using HMGA1aBS-wt RNA (Figure 3B, lane 3). In the presence of HMGA1a, the fast mobility band increased

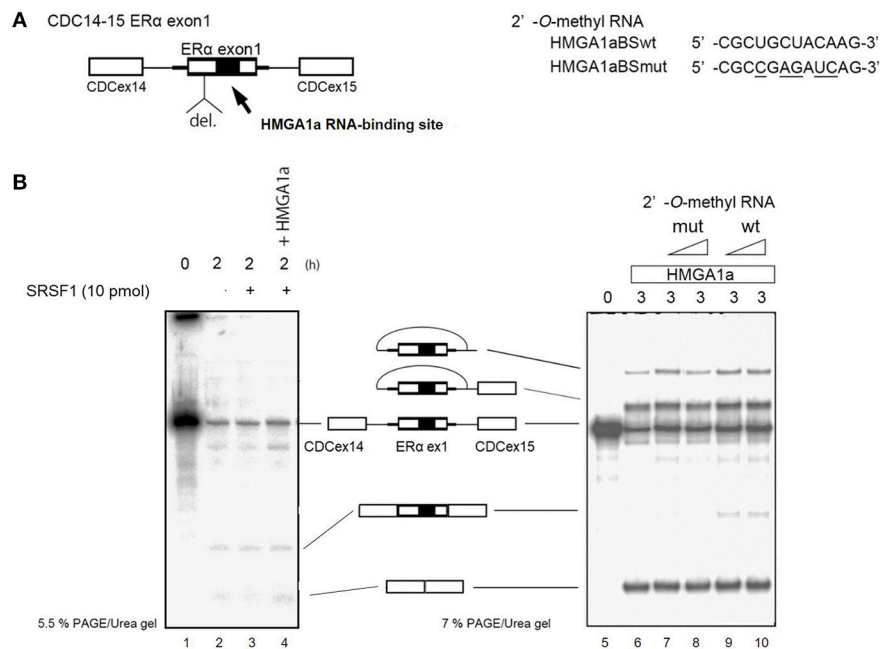


FIGURE 2 | HMGA1a regulates alternative splicing of ER α exon 1 *in vitro*. **(A)** Scheme shows CDC-ER α pre-mRNA: shortened [indicated as del. (deletion of nucleotides)] ER α exon 1 and flanking intron sequence (bold line) inserted in the intron (thin line) of CDC14-15 pre-mRNA. The 2'-O-methyl RNA used in **(B)** are shown on right side. **(B)** (left panel) HMGA1a induces exon skipping of CDC-ER α (lane 1). SRSF1 was added as indicated (lanes 3,4). Two hundred and eighty-seven nanograms in 12.5 μ l reaction of recombinant HMGA1a was added in lane 4. **(B)** (right panel) 2'-O-methyl RNA of HMGA1a RNA-binding sequence increases exon inclusion. Fifty and one hundred picomoles of 2'-O-methyl RNA are indicated as triangles. White arrowhead shows increase of exon inclusion by HMGA1a RNA-binding sequence 2'-O-methyl RNA.

intensity, while the slow mobility band was almost completely abolished (**Figure 3B** lane 4). The slow mobility crosslink found in HMGA1aBS-wt could not be detected using A5'SSmut, thus can be designated as the U1 snRNA/authentic 5' splice site-crosslink. The fast mobility crosslink is difficult to judge because of its similar mobility with the internal crosslinks (**Figure 3B**, lane 2). This is also the case for A5'SS mut (**Figure 3B**, compare lanes 8, 9 with lane 7) and P5'SS mut (**Figure 3B**, compare lanes 13, 14 with lane 12). When using P5'SS mut, the intensity of the slow mobility crosslink (U1 snRNA/authentic 5' splice site) did not decrease, but rather increased in the presence of HMGA1a.

Taken together, an aberrant complex of HMGA1a-mediated trapping of U1 snRNP to the upstream 5' splice site inhibited binding of U1 snRNP to the authentic-5' splice site (**Figure 4**).

DISCUSSION

HMGA1a is originally known as a DNA-binding transcription factor but we found it exerts abnormal exon skipping of the *presenilin-2* gene in sporadic Alzheimer's disease through sequence-specific RNA binding to a sequence, 5'-GCUGCUACAAG-3' (Manabe et al., 2003, 2007; Ohe and Mayeda, 2010). It has also been recently reported of other RNA targets for HMGA1a: regulating Vpr mRNA expression of the HIV gene (Tsuruno et al., 2011); and binding to 7SK snRNA through its DNA-binding domain and thereby affecting its own

transcriptional regulation activity (Eilebrecht et al., 2011a,b,c). HMGA1a is known to be multifunctional (Reeves, 2001), with various functions in normal as well as pathophysiological contexts. Indeed, it has been reported as an important component of senescence-associated heterochromatic foci (Narita et al., 2006). This study expands the number of RNA targets of HMGA1a with important pathophysiological function as well as suggesting a novel mechanism in 5' splice site choice.

Regulation of 5' splice site function by upstream 5' splice sites has been analyzed in previous studies and uncovered silencing sequences (Yu et al., 2008) as well as the discovery of upstream 5' splice sites functioning as enhancers of the authentic 5' splice site (Hicks et al., 2010). How the intron-proximal 5' splice site is favored when two comparable 5' splice sites exist, is not known (Roca et al., 2013). Here we showed an example of such two competing 5' splice sites in exon 1 of the *estrogen receptor alpha* (*ERα*) gene. The 5' splice site scores of the two 5' splice sites are comparable both with a MAXENT score of 8.6. In normal expression of the *ERα* gene the intron-proximal 5' splice site is utilized. Regulation of U1 snRNP binding to alternative 5' splice sites has been reported to occur during A complex formation, with no difference of U1 snRNP binding to upstream and downstream 5' splice site in E complex (Hodson et al., 2012). U1 snRNP has been reported to protect a region that is 23 nucleotides upstream into the exon and 12 nucleotides downstream into the intron in PTB-independent conditions (Sharma et al., 2011).

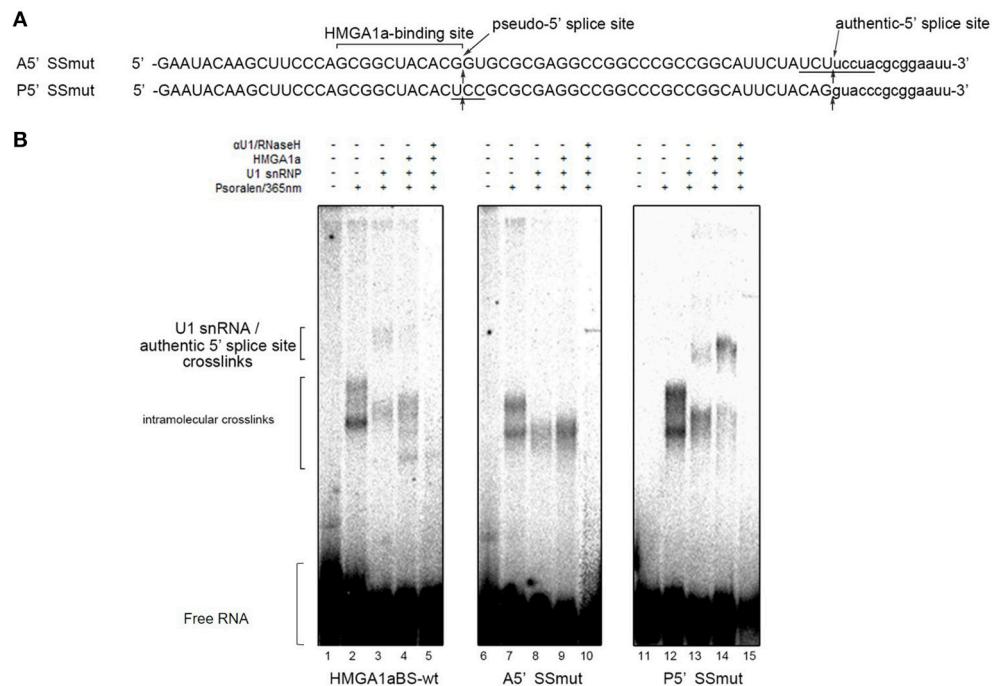


FIGURE 3 | U1 snRNP is anchored to the upstream pseudo-5' splice site in the presence of HMGA1a. **(A)** Aligned sequences of the short RNAs containing mutated authentic 5' splice site (Amut) and mutated pseudo-5' splice site (Pmut). Mutated bases are underlined. Arrows show the pseudo-5' splice site and authentic-5' splice site. **(B)** Radiolabeled RNA of **(A)** was crosslinked to purified U1 snRNP (234 ng in 12.5 μ l) in the presence (287 ng in 12.5 μ l) or absence of HMGA1a by psoralen and 365 nm UV light.

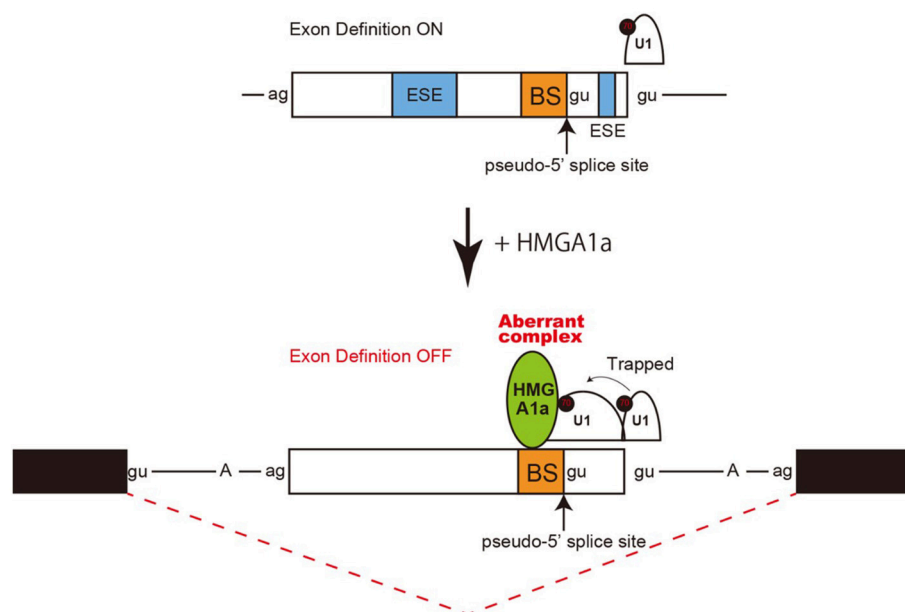


FIGURE 4 | Model and mechanism of HMGA1a-induced ER α alternative splicing. HMGA1a traps U1 snRNP to an upstream pseudo 5' splice site and disrupts the function of the downstream authentic 5' splice site.

Since the two 5' splice sites of ER α exon 1 are 33 nucleotides apart, the length in-between would allow both 5' splice sites to bind U1 snRNP but would be in close proximity. We believe the two 5' splice sites found in ER α exon 1 would be a good natural model in studying how the intron-proximal 5' splice site is favored when the upstream 5' splice site is in the closest range of simultaneous binding of U1 snRNP. From RNA-EMSA and psoralen crosslinking assays in this study, HMGA1a binded to an upstream sequence adjacent a pseudo 5' splice site and inhibited U1 snRNP binding to the authentic 5' splice site only in the presence of the upstream pseudo 5' splice site. We believe U1 snRNP simultaneously binded to both 5' splice sites of ER α exon 1 in the absence of HMGA1a, but when HMGA1a is added, U1 snRNP is trapped to form an aberrant complex which inhibits U1 snRNP binding to the authentic 5' splice site (Figure 4).

However, there is still various limitations in this study. First, if U1 snRNP protection at the upstream 5' splice site is extended more than U1 snRNP binding in PTB independent conditions (Sharma et al., 2011), U1 snRNP binding to authentic-5' splice site would be inhibited. This may be the case because we observed an increase of U1 snRNP binding to the authentic 5' splice site when the upstream 5' splice site was mutated (Figure 3B, lane 14). An RNase H protection assay with an antisense oligonucleotide directed to check the occupancy of the authentic 5' splice site would be able to answer this point. Second, there is still a possibility of HMGA1a trapping U1 snRNP to the authentic 5' splice site. Though we previously found that HMGA1a traps U1 snRNP when the HMGA1a RNA-binding site is adjacent or at least within ten nucleotides of the 5' splice site (Ohe and Mayeda, 2010), the sequence between the two 5' splice sites of ER α exon 1 is extremely GC-rich (24 out of 32 nucleotides; 75%) and there is a strong possibility of secondary structure leading to close enough range of the HMGA1a RNA-binding site and 5' splice site for HMGA1a-induced U1 snRNP trapping. This also well explains the increased crosslink of U1 snRNP and the authentic 5' splice site when the upstream 5' splice site was mutated (Figure 3B, lane 14).

Besides the precise mechanism of HMGA1a-induced silencing of the downstream authentic 5' splice site of ER α exon 1, this event consequently induced exon skipping and expression of the isoform ER α 46. Using decoy RNA that binded to HMGA1a and inhibited HMGA1a-induced exon skipping of ER α , we observed

enhanced estrogen-dependent tumor growth and sensitization of tamoxifen-resistant tumor cells to tamoxifen due to increased expression of full length ER α by correction of alternative splicing (Ohe et al., 2018). We hope this HMGA1a-targeted therapy, along with its RNA-binding site, will enlighten a novel strategy in overcoming tamoxifen-resistant breast cancer.

AUTHOR CONTRIBUTIONS

KO conceived the original idea, conducted experiments (Figures 1B, 2, 3), and wrote the paper. SM, TT, YH, YHa, YHo, YB, FI, TaY, HT, MM, YM, MT, IA, KA, KK, ME, TN, ToY, NH, TU, AM helped conducting the experiments and provided materials as well as advice on the paper. All the authors contributed substantially to the conception or design, data, interpretation of the results, they all drafted and revised the content, and finally approved for publishment. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was supported by a Grant-in-Aid for Scientific Research (C) to KO (grant 24591920 and 15K10050) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and a Grant-in-Aid from Fujita Health University to AM and TU.

ACKNOWLEDGMENTS

We thank Professor Reinhard Lührmann (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany) for kindly providing purified U1 snRNP. We thank Dr. Kanako Kanameki and Professor Yutaka Muto (Faculty of Pharmacy and Research Institute of Pharmaceutical Sciences, Musashino University, Musashino, Tokyo, Japan) for kindly providing purified HMGA1a.

We thank the Radioisotope Research Center of Fujita Health University and Fukuoka University for instrumental support.

REFERENCES

- Barraille, P., Chinestra, P., Bayard, F., and Faye, J. C. (1999). Alternative initiation of translation accounts for a 67/45 kDa dimorphism of the human estrogen receptor ER α . *Biochem. Biophys. Res. Commun.* 257, 84–88. doi: 10.1006/bbrc.1999.0334
- Chen, J., and Weiss, W. A. (2015). Alternative splicing in cancer: implications for biology and therapy. *Oncogene* 34, 1–14. doi: 10.1038/ncr.2013.570
- David, C. J., and Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 24, 2343–2364. doi: 10.1101/gad.1973010
- Eilebrecht, S., Becavin, C., Leger, H., Benecke, B. J., and Benecke, A. (2011a). HMGA1-dependent and independent 7SK RNA gene regulatory activity. *RNA Biol.* 8, 143–157. doi: 10.4161/rna.8.1.14261
- Eilebrecht, S., Benecke, B. J., and Benecke, A. (2011b). 7SK snRNA-mediated, gene-specific cooperativity of HMGA1 and P-TEFb. *RNA Biol.* 8, 1084–1093. doi: 10.4161/rna.8.6.17015
- Eilebrecht, S., Brysbaert, G., Wegert, T., Urlaub, H., Benecke, B. J., and Benecke, A. (2011c). 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. *Nucleic Acids Res.* 39, 2057–2072. doi: 10.1093/nar/gkq1153
- Filarsky, M., Zillner, K., Araya, I., Villar-Garea, A., Merkl, R., Langst, G., et al. (2015). The extended AT-hook is a novel RNA binding motif. *RNA Biol.* 12, 864–876. doi: 10.1080/15476286.2015.1060394
- Flouriot, G., Brand, H., Denger, S., Metivier, R., Kos, M., Reid, G., et al. (2000). Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is

- able to repress hER-alpha activation function 1. *EMBO J.* 19, 4688–4700. doi: 10.1093/emboj/19.17.4688
- Hicks, M. J., Mueller, W. F., Shepard, P. J., and Hertel, K. J. (2010). Competing upstream 5' splice sites enhance the rate of proximal splicing. *Mol. Cell Biol.* 30, 1878–1886. doi: 10.1128/MCB.01071-09
- Hodson, M. J., Hudson, A. J., Cherny, D., and Eperon, I. C. (2012). The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites. *Nucleic Acids Res.* 40, 6850–6862. doi: 10.1093/nar/gks322
- Inoue, K., and Fry, E. A. (2015). Aberrant splicing of estrogen receptor, HER2, and CD44 genes in breast cancer. *Genet. Epigenet.* 7, 19–32. doi: 10.4137/GEG.S35500
- Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* 14, 185–193. doi: 10.1038/nsmb1209
- Kataoka, N., Yong, J., Kim, V., N., Velazquez, F., Perkinson, R., A., Wang, F., et al. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* 6, 673–682. doi: 10.1016/S1097-2765(00)00065-4
- Lund, T., Holtlund, J., Fredriksen, M., and Laland, S. G. (1983). On the presence of two new high mobility group-like proteins in HeLa S3 cells. *FEBS Lett.* 152, 163–167. doi: 10.1016/0014-5793(83)80370-6
- Maasch, C., Vater, A., Buchner, K., Purschke, W. G., Eulberg, D., Vonhoff, S., et al. (2010). Polyethylenimine-polyplexes of Spiegelmer NOX-A50 directed against intracellular high mobility group protein A1 (HMGA1) reduce tumor growth *in vivo*. *J. Biol. Chem.* 285, 40012–40018. doi: 10.1074/jbc.M110.178533
- Manabe, T., Katayama, T., Sato, N., Gomi, F., Hitomi, J., Yanagita, T., et al. (2003). Induced HMGA1a expression causes aberrant splicing of Presenilin-2 pre-mRNA in sporadic Alzheimer's disease. *Cell Death Differ.* 10, 698–708. doi: 10.1038/sj.cdd.4401221
- Manabe, T., Ohe, K., Katayama, T., Matsuzaki, S., Yanagita, T., Okuda, H., et al. (2007). HMGA1a: sequence-specific RNA-binding factor causing sporadic Alzheimer's disease-linked exon skipping of presenilin-2 pre-mRN. *Genes Cells* 12, 1179–1191. doi: 10.1111/j.1365-2443.2007.01123.x
- Mayeda, A., and Krainer, A. R. (1999a). Mammalian *in vitro* splicing assays. *Methods Mol. Biol.* 118, 315–321.
- Mayeda, A., and Krainer, A. R. (1999b). Preparation of HeLa cell nuclear and cytosolic S100 extracts for *in vitro* splicing. *Methods Mol. Biol.* 118, 309–314.
- Narita, M., Krizhanovsky, V., Nunez, S., Chicas, A., Hearn, S. A., Myers, M. P., et al. (2006). A novel role for high-mobility group A proteins in cellular senescence and heterochromatin formation. *Cell* 126, 503–514. doi: 10.1016/j.cell.2006.05.052
- Ohe, K., and Mayeda, A. (2010). HMGA1a trapping of U1 snRNP at an authentic 5' splice site induces aberrant exon skipping in sporadic Alzheimer's disease. *Mol. Cell Biol.* 30, 2220–2228. doi: 10.1128/MCB.00114-10
- Ohe, K., Miyajima, S., Abe, I., Tanaka, T., Hamaguchi, Y., Harada, Y., et al. (2018). HMGA1a induces alternative splicing of estrogen receptor alpha in MCF-7 human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* doi: 10.1016/j.jsbmb.2018.04.007. [Epub ahead of print].
- Oltean, S., and Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311–5318. doi: 10.1038/onc.2013.533
- Penot, G., Le Peron, C., Merot, Y., Grimaud-Fanouillere, E., Ferriere, F., Boujrad, N., et al. (2005). The human estrogen receptor-alpha isoform hERalpha46 antagonizes the proliferative influence of hERalpha66 in MCF7 breast cancer cells. *Endocrinology* 146, 5474–5484. doi: 10.1210/en.2005-0866
- Pierantoni, G. M., Esposito, F., Giraud, S., Bienvenut, W. V., Diaz, J. J., and Fusco, A. (2007). Identification of new high mobility group A1 associated proteins. *Proteomics* 7, 3735–3742. doi: 10.1002/pmic.200700148
- Reeves, R. (2001). Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* 277, 63–81. doi: 10.1016/S0378-1119(01)00689-8
- Reeves, R. (2010). Nuclear functions of the HMG proteins. *Biochim. Biophys. Acta* 1799, 3–14. doi: 10.1016/j.bbagr.2009.09.001
- Roca, X., Krainer, A. R., and Eperon, I. C. (2013). Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev.* 27, 129–144. doi: 10.1101/gad.209759.112
- Sawa, H., Ohno, M., Sakamoto, H., and Shimura, Y. (1988). Requirement of ATP in the second step of the pre-mRNA splicing reaction. *Nucleic Acids Res.* 16, 3157–3164. doi: 10.1093/nar/16.8.3157
- Sgarra, R., Furlan, C., Zammitti, S., Lo Sardo, A., Maurizio, E., Di Bernardo, J., et al. (2008). Interaction proteomics of the HMGA chromatin architectural factors. *Proteomics* 8, 4721–4732. doi: 10.1002/pmic.200800193
- Sgarra, R., Tessari, M. A., Di Bernardo, J., Rustighi, A., Zago, P., Liberatori, S., et al. (2005). Discovering high mobility group A molecular partners in tumour cells. *Proteomics* 5, 1494–1506. doi: 10.1002/pmic.200401028
- Sharma, S., Maris, C., Allain, F. H., and Black, D. L. (2011). U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol. Cell* 41, 579–588. doi: 10.1016/j.molcel.2011.02.012
- Silipo, M., Gautrey, H., and Tyson-Capper, A. (2015). Deregulation of splicing factors and breast cancer development. *J. Mol. Cell Biol.* 7, 388–401. doi: 10.1093/jmcb/mjv027
- Tsuruno, C., Ohe, K., Kuramitsu, M., Kohma, T., Takahama, Y., Hamaguchi, Y., et al. (2011). HMGA1a is involved in specific splice site regulation of human immunodeficiency virus type 1. *Biochem. Biophys. Res. Commun.* 406, 512–517. doi: 10.1016/j.bbrc.2011.02.059
- Yeo, G., and Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394. doi: 10.1089/1066527041410418
- Yu, Y., Maroney, P. A., Denker, J. A., Zhang, X. H., Dybkov, O., Luhrmann, R., et al. (2008). Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* 135, 1224–1236. doi: 10.1016/j.cell.2008.10.046

Conflict of Interest Statement: 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.

Copyright © 2018 Ohe, Miyajima, Tanaka, Hamaguchi, Harada, Horita, Beppu, Ito, Yamasaki, Terai, Mori, Murata, Tanabe, Abe, Ashida, Kobayashi, Enjoji, Nomiya, Yanase, Harada, Utsumi and Mayeda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Translation of Hepatitis A Virus IRES Is Upregulated by a Hepatic Cell-Specific Factor

Akitoshi Sadahiro¹, Akira Fukao², Mio Kosaka², Yoshinori Funakami², Naoki Takizawa³, Osamu Takeuchi¹, Kent E. Duncan⁴ and Toshinobu Fujiwara^{2*}

¹ Laboratory of Infection and Prevention, Department of Virus Research, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan, ² Laboratory of Biochemistry, Graduate School of Pharmaceutical Sciences, Kindai University, Osaka, Japan, ³ Laboratory of Virology, Institute of Microbial Chemistry (BIKAKEN), Tokyo, Japan, ⁴ Center for Molecular Neurobiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Alain Krol,
UPR9002 Architecture et Réactivité
de l'ARN, France
Marcelo Lopez-Lastra,
Pontificia Universidad Católica
de Chile, Chile
Marc Fabian,
McGill University, Canada

*Correspondence:

Toshinobu Fujiwara
tosinobu@phar.kindai.ac.jp

Specialty section:

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

Received: 28 March 2018

Accepted: 19 July 2018

Published: 10 August 2018

Citation:

Sadahiro A, Fukao A, Kosaka M,
Funakami Y, Takizawa N, Takeuchi O,
Duncan KE and Fujiwara T (2018)
Translation of Hepatitis A Virus IRES Is
Upregulated by a Hepatic
Cell-Specific Factor.
Front. Genet. 9:307.
doi: 10.3389/fgene.2018.00307

Many viruses strongly prefer to infect certain cell types, a phenomenon known as “tropism.” Understanding tropism’s molecular basis is important for the design of vaccines and antiviral therapy. A common mechanism involves viral protein interactions with cell-specific surface receptors, but intracellular mechanisms involving translation have also been described. In this report, we focus on Hepatitis A Virus (HAV) tissue tropism from the standpoint of the translational machinery. HAV genomic RNA, like other positive stranded RNA viruses, is devoid of a cap structure and its translation is driven by highly structured RNA sequences termed internal ribosome entry site (IRES) in the 5′ untranslated region (UTR). Unlike most viral IRESs, HAV IRES-mediated translation requires eIF4E and the 3′ end of HAV RNA is polyadenylated. However, the molecular mechanism of HAV IRES-mediated translation initiation remains poorly understood. We analyzed HAV-IRES-mediated translation in a cell-free system derived from either non-hepatic cells (HeLa) or hepatoma cells (Huh-7) that enables investigation of the contribution of the cap and the poly(A) tail. This revealed that HAV IRES-mediated translation activity in hepatoma cell extracts is higher as compared to extracts derived from a non-hepatic line. Our data suggest that HAV IRES-mediated translation is upregulated by a hepatic cell-specific activator in a poly(A) tail-independent manner.

Keywords: Hepatitis A Virus (HAV), internal ribosome entry site (IRES), translation regulation, tropism, translation initiation

INTRODUCTION

In eukaryotes, the vast majority of cellular mRNAs are capped at the 5′ end and polyadenylated at the 3′ end. Translation initiation on these mRNAs follows a well-defined pathway (Green et al., 2016) involving multiple stages governed by a large number of eIFs interacting with the mRNA and the ribosome (Jackson et al., 2010). The first step is recognition of the m⁷G cap structure by the eukaryotic initiation factor (eIF) 4F complex, which consists of an ATP-dependent RNA helicase eIF4A, a large scaffold protein eIF4G, and a cap binding protein eIF4E. eIF4G interacts with the poly(A) binding protein PABP and the interaction between these three proteins (eIF4E, eIF4G, and PABP) leads to circularization of the capped mRNA (Tarun and Sachs, 1995; Le et al., 1997; Imataka et al., 1998). The 43S pre-initiation complex (PIC), which consists of the ternary complex

(eIF2-GTP-Met-tRNAi), several eIFs including eIF1, 1A, 3, and 5, and the 40S small ribosomal subunit, is recruited to the mRNA via the eIF4F complex. After binding to mRNA, the 43S PIC scans the mRNA 5' UTR in a 5' to 3' direction until the Met-tRNAi recognizes the start AUG codon. The recognition of the start AUG codon leads to conformational changes that produce a stable 48S initiation complex. Subsequently some eIFs are released, and then the 60S large ribosome subunit is recruited to form an 80S initiation complex ready to synthesize encoded peptide.

Capping and polyadenylation of cellular mRNA occurs in the nucleus, co-transcriptionally (Moteki and Price, 2002; Richard and Manley, 2009). However, many RNA viruses replicate in the cytoplasm and viral proteins are synthesized from their mRNA without cap structure. Picornaviruses that are positive-stranded RNA viruses replicate in the cytoplasm, but their mRNAs have a poly(A) tail synthesized by a viral RNA polymerase. In contrast, picornavirus mRNA lacks a cap structure. Instead of cap structure, a viral protein (VPg) is covalently linked to the 5' end. Picornavirus mRNA is translated via an internal ribosome entry site (IRES) in its 5' UTR. Viral IRES are highly structured RNA sequences that recruit the ribosome onto mRNA in a cap independent manner. IRES elements were initially reported in poliovirus (PV) and encephalomyocarditis virus (EMCV) genomes (Jang et al., 1988; Pelletier and Sonenberg, 1988). It has been demonstrated that the viral mRNAs of all members of the *Picornaviridae* family are translated in an IRES dependent manner (Martinez-Salas et al., 2015). Picornavirus IRESs can be classified into three groups by eIFs/cellular proteins requirement and scanning mechanism. The picornavirus type I IRESs (e.g., Enterovirus, Rhinovirus) and type II IRESs (e.g., Cardiovirus and Aphthovirus) require almost all of initiation factors, except for eIF4E, to stimulate the IRES activity. Moreover, some cellular proteins, known as IRES trans-acting factors (ITAFs) stimulate the type I and II IRESs activity. ITAFs have typically been identified using biochemical approaches. They are usually RNA-binding proteins (e.g., polypyrimidine tract binding protein; PTB) and can regulate IRES-dependent translation both positively and negatively (Graff et al., 1998; Hunt and Jackson, 1999; Gosert et al., 2000; Yi et al., 2000; Cordes et al., 2008). The picornavirus type I and type II IRES are classified by scanning mechanism. The 3' border of type I and II IRES harbors the Yn-Xm-AUG motif in which Yn (pyrimidine-rich tract; $n = 8-10$ nt) is separated by a spacer (Xm; $m = 18-20$ nt spacer) from an AUG triplet. In the case of type I IRES, recruited ribosome to Yn-Xm-AUG motif scans the downstream spacer region (at a distance of ~ 160 nt) for translation initiation on the authentic AUG (Hellen et al., 1994). On the other hand, in the case of type II IRES, ribosome is recruited to AUG codon of Yn-Xm-AUG motif directly without scanning (Kaminski et al., 1994). The picornavirus IRES type III (e.g., HAV) requires all of initiation factors including eIF4E and some ITAFs to stimulate the IRES activity.

The host and tissue tropisms of viruses are determined by multiple host and viral factors. In general it is well accepted that there are two major types of viral tropism, receptor-dependent

and -independent tropism. The translation machinery is one of the important factors for determining receptor-independent tropism, since the host translation machinery is used to express viral proteins, and thus, to regulate viral propagation (Niepmann, 2009). Tissue-specific expression of ITAFs could help to explain viral tissue tropism on the translational level. For example, ITAF₄₅, also known as erbB-3-binding protein 1 (Ebp1), binds to foot-and-mouth disease virus (FMDV) IRES and stimulates IRES activity. ITAF₄₅ expresses in proliferating cells during the S phase but not during cell cycle arrest (Radomski and Jost, 1995). Therefore, FMDV IRES is activated only in proliferating tissue, characterizing the virus as endotheliotropic (Pilipenko et al., 2000). miR-122 that is expressed preferentially in liver cells is one of the *trans*-acting factors for regulating HCV translation (Lagos-Quintana et al., 2002; Sempere et al., 2004; Chang et al., 2004; Fu et al., 2005). miR-122 contributes to stimulation of HCV translation by recognition of a *cis*-acting element on the 5' UTR of its mRNA, and thus, miR-122 is one factor determining HCV liver tropism on the level of translation (Henke et al., 2008).

Hepatitis A Virus (HAV), a member of the *Picornaviridae* family, is one of the major causative agents of acute hepatitis. HAV infection does not cause chronic liver disease, but superinfection of HAV with hepatitis B virus or HCV may affect the natural history of HBV and HCV related to liver cirrhosis and cancer (Lemon et al., 2017). HAV mRNA contains an IRES element in its 5' UTR, a single open reading frame, and is polyadenylated (Brown et al., 1991). The translational mechanism driven from Picornaviral IRES is determined by characteristic of the IRES element. IRES elements of picornaviruses are categorized into three groups and the HAV IRES belongs to type III. In the case of HAV IRES-mediated translation, the first event of translation initiation is that eIF4F complex recognizes the IRES element (Borman et al., 2001). Interestingly, it has been demonstrated that eIF4E, the cap binding protein, is required for activation of HAV IRES-mediated translation although HAV mRNA does not have cap structure (Ali et al., 2001). Subsequently the 43S PIC and the 60S large ribosome subunit are recruited onto HAV mRNA, followed by synthesis of the viral polyprotein. HAV IRES activity was found to be stimulated by two ITAFs, PTB and poly(rC) binding protein 2 (PCBP2) (Graff et al., 1998; Gosert et al., 2000). Conversely, two negative ITAFs that repress HAV IRES activity were identified: GAPDH and La autoantigen (Yi et al., 2000; Cordes et al., 2008). It was suggested that HAV IRES-mediated translation could explain hepatotropism (Borman et al., 1997). However, the reported positive and negative ITAFs for HAV translation are broadly expressed in many tissues. Thus, the hepatocyte specific translational mechanism of HAV mRNA remains unclear.

In this study, we focus on hepatic cell-specific translation mediated by the HAV IRES. We previously established a cell-free translation system using human cervical carcinoma cell (HeLa) extracts (Fukao et al., 2009). Thus, to analyze hepatic cell-specific translation mechanisms, we adapted this system to the human hepatoma cell (Huh-7) and established a cell-free translation system using Huh-7. By measuring HAV IRES-mediated translation activity in Huh-7 and HeLa cell extracts,

we show that an HAV IRES reporter mRNA is translated more efficiently in Huh-7 cell extracts than in HeLa cells. Translation extract mixing experiments demonstrate that this effect is due to a positive-acting factor in the Huh-7 cell extracts. Moreover, translational enhancement mediated by the HAV IRES in Huh-7 cell extracts was also observed with a reporter mRNA lacking a poly(A) tail. Our results support the idea that HAV is highly hepatotropic on a translational level and suggest that liver cell-specific components can stimulate translation of this IRES through a mechanism that does not involve the poly(A) tail.

MATERIALS AND METHODS

Plasmids

Reporter plasmid (pBSII-Nluc-A114) encoding NanoLuc (Nluc) luciferase was constructed previously (Fukao et al., 2014). To obtain the plasmid encoding HAV-IRES-Nluc-A114 and EMCV IRES-Nluc-A114 (pBSII-HAV-IRES-Nluc-A114 and pBSII-EMCV-IRES-Nluc-A114), the total synthesized HAV-IRES sequence (Borman et al., 1995) (invitrogen) and the total synthesized EMCV-IRES were inserted upstream of Nluc gene of pBSII-Nluc-A114.

In vitro Transcription

Plasmids were linearized downstream of A114 or the Nluc gene with HindIII or XbaI, respectively. Linearized plasmids with HindIII or XbaI were used as templates to synthesize mRNA with or without a poly(A) tail. mRNAs were synthesized in the presence of either 7mGpppG (cap), ApppG (Acap) or in the absence of cap analog (Nocap) with previously described protocol (Duncan et al., 2009). The synthesized RNAs were purified by RNeasy Mini Kit (QIAGEN).

Cell Culture and in vitro Translation

Huh-7 and HeLa cells were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 5% fetal bovine serum. Cells were detached for preparation of cell extracts by 2.5 g/l-Trypsin Solution (Nacalai) when cell outgrowth had reached 90% confluence. Detached cells were collected by centrifugation at $700 \times g$ for 2 min and washed two times in phosphate buffered saline (PBS) at 4°C. Pelleted cells were suspended in 1 volume of ice cold lysis buffer (Thoma et al., 2004). After 5 min on ice, cells were lysed with 12 strokes with needle. Following centrifugation of the homogenate at $10,000 \times g$ for 10 min at 4°C, supernatant was collected as cell extracts. These cell extracts were dialyzed with lysis buffer, and dialyzed cell extracts were used for translation reaction. A total of 8 µl of nuclease-untreated HeLa or Huh-7 cell extracts, 10 µl of reaction buffer (30 mM Hepes-KOH buffer [pH 7.4], 8 mM creatine phosphate, 0.5 mM spermidine, 1 mM ATP, 0.2 mM GTP, 20 µM amino acids, 1.5 mM magnesium acetate, 80 mM potassium acetate, 40 µg/ml creatine kinase), and 2 µl of 10 ng/µl reporter mRNA were incubated at 37°C for 30 min. The reaction was stopped by liquid N₂, and the luciferase reporter assay was performed according to the manufacturer's protocol (NanoGlo Luciferase Assay Systems, Promega).

Northern Blot Analysis

Total RNA was extracted from *in vitro* translation reaction mixtures after luciferase re- porter assays using ISOGEN II (Nippon Gene). Samples were separated in a 1.0% formaldehyde-containing agarose gel and transferred onto a nylon membrane (Pall). After blotting RNAs onto nylon membrane, the membrane was dyed by methylene blue solution (0.04% methylene blue, 0.5M NaOAc [pH 5.3]). The stained membrane was subjected to analysis under bright field with the LAS-4000 image analyzer (Fuji). Reporter mRNAs were detected with digoxigenin (DIG)-labeled RNA probe complementary to the Nluc gene in hybridization buffer (50% formamide, 750 mM NaCl, 75 mM Trisodium Citrate, 2% Blocking Reagent, 0.1% SDS, 0.1% N-Lauroylsarcosine, 200 ng/µl yeast tRNA) at 60°C for 12 h. The hybridization signal was detected with CDP Star (Roche) as the reaction substrate according to the manufacturer's instructions. The membrane was subjected to analysis with an LAS-4000 image analyzer.

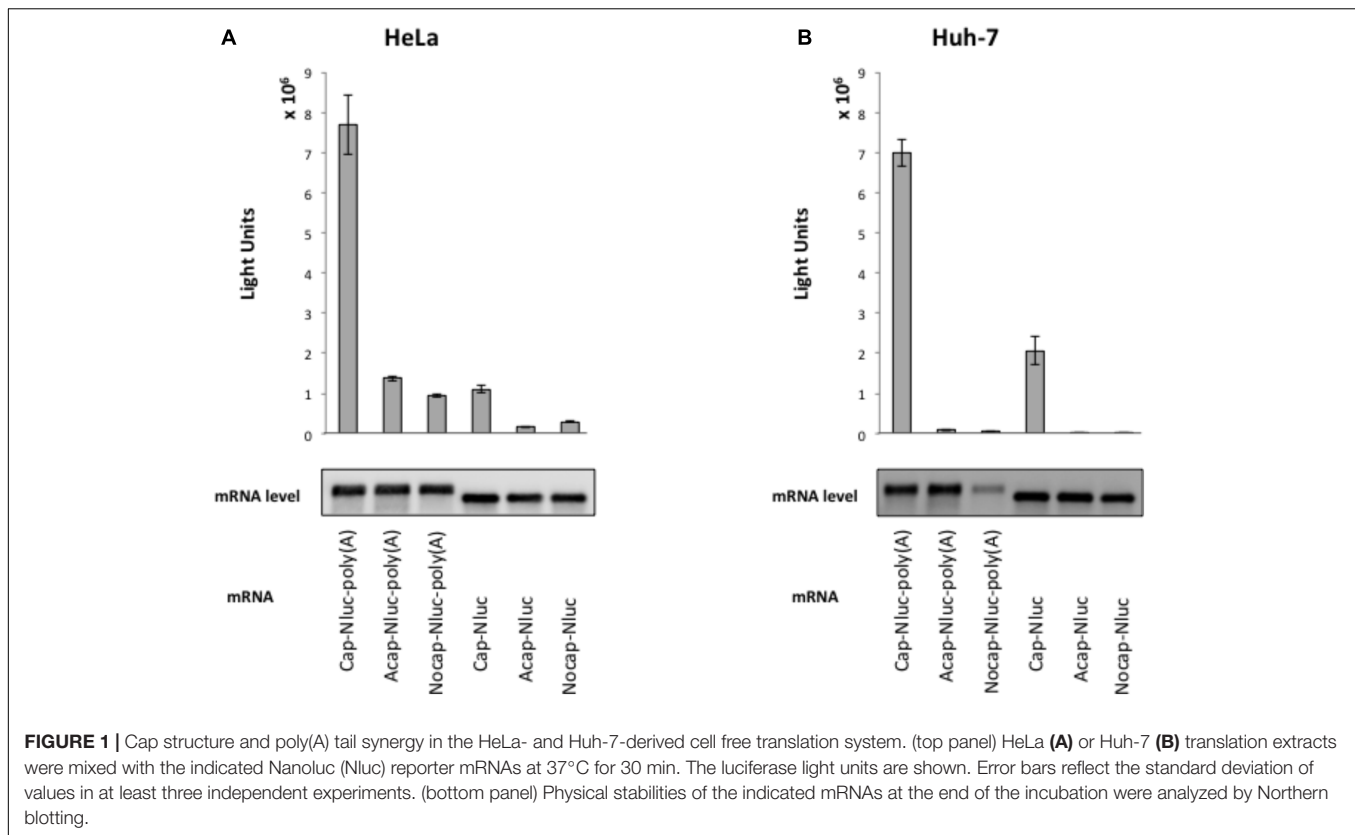
Sucrose Density Gradient Assay

In vitro translation was scaled up to 40 µl and performed at 37°C for 10 min. After 10 min incubation, 0.5 µl of 100 mg/ml cycloheximide was added to stop the translation reaction. The reaction mixture was loaded on the top of 11 ml of a linear 5–25% sucrose gradient (5–25% sucrose in 20 mM HEPES-KOH [pH 7.6], 150 mM potassium acetate, 5 mM MgCl₂). After centrifugation at 38,000 rpm for 2.5 h at 4°C in a HITACHI P40ST rotor, 11 fractions (each 1 ml) were collected from the top of the gradient using a piston gradient fractionator (BIOCOMP). Total RNA was extracted from each fraction, and Nluc mRNA was detected by Northern blotting.

RESULTS

Cap- and Poly(A)-Dependent Translation Huh-7 Cell Extracts

To elucidate HAV IRES-mediated translation preference for cell extracts, we established a cell-free translation system from HeLa cell and human liver cell (hepatoma) Huh-7 extracts. To assess our *in vitro* translation system, we measured Nluc activity from reporter mRNAs with an m⁷GpppG cap, analogous to the physiological cap, or a non-physiological ApppG cap analog (Acap) that is not recognized by eIF4E and with or without poly (A) tail. Reporter mRNAs were incubated with HeLa and Huh-7 cell extracts at 37°C for 30 min. In HeLa cell extracts, translation activity of cap-Nluc-poly(A) mRNA was approximately 7 times higher than that of cap-Nluc mRNA, and 5.5 times higher than that of Acap-Nluc-poly(A) mRNA (Figure 1A). In Huh-7 cell extracts, translation activity of cap-Nluc-poly(A) mRNA was approximately 3.5 times higher than that of cap-Nluc mRNA, and 75 times higher than that of Acap-Nluc-poly(A) mRNA (Figure 1B). Moreover, translation activity of the other reporter mRNAs [Nocap-Nluc-poly(A), Acap-Nluc, Nocap-Nluc] were 8–500 times lower than cap-Nluc-poly(A) mRNA. In principle, differences in mRNA stability could contribute to the observed



differences in luciferase activity. To address this issue, we measured reporter mRNA levels at the end of the different translation reactions. We detected no significant differences in the amount of reporter mRNAs in HeLa (Figure 1A). On the other hand, in Huh-7 cell extracts, Nocap-Nluc-poly(A) mRNA was degraded after translation reaction, but the other reporter mRNAs were stable after translation reaction (Figure 1B). These data indicates that the observed differences of translation activity are due to different translational efficiencies, except for Nocap-Nluc-poly(A) in Huh-7. These results show that our *in vitro* translational system with HeLa and Huh-7 extracts can analyze cap- and poly(A)-dependent translation.

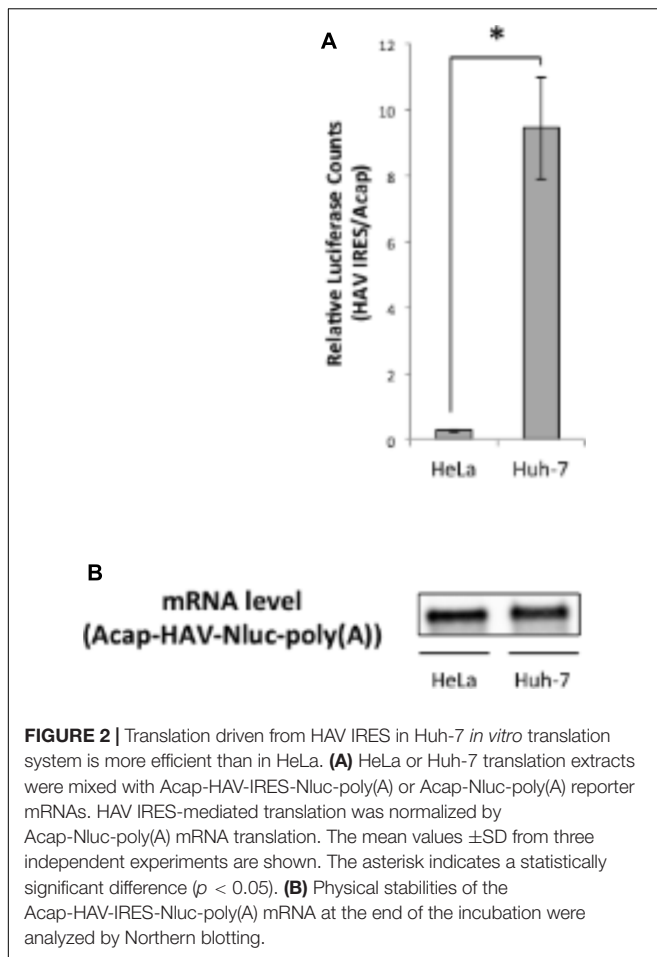
HAV IRES-Mediated Translation Is Enhanced in Huh-7 Cell Extracts

Next, we analyzed whether translation mediated by the HAV IRES is enhanced in liver-derived hepatoma cells by comparing *in vitro* translation activities of HAV IRES reporters in HeLa or Huh-7 extracts. We used a monocistronic Nluc reporter mRNA containing a HAV IRES element in the 5' UTR, as this configuration has been argued to be superior to dicistronic mRNAs for evaluating IRES-mediated translation (Song et al., 2006; Jünemann et al., 2007). The HAV IRES reporter mRNA not attached cap analog was unstable after 30 min incubation in our Huh-7 *in vitro* translation system, compared with a non-canonical Acap attached reporter mRNA (Supplementary Figure S1). Therefore, to analyze HAV IRES

mediated translation and to inhibit reporter mRNA degradation, an Acap is attached to the 5' end of the reporter mRNA which conjugated HAV IRES element. The translation activity of Acap-HAV-IRES-Nluc-poly(A) mRNA in Huh-7 cell extracts was approximately 35 times higher than in that in HeLa cell extracts (Figure 2A). Moreover, the translation activity driven by HAV IRES was analyzed in another liver-derived hepatoma cells, HepG2. Compared with HeLa cell, HAV IRES-mediated translation was enhanced in HepG2 cell extracts (Supplementary Figure S2). To determine whether this was due to effects on mRNA stability or translation, we again examined the amount of reporter mRNA after translation reactions by Northern blotting. The indicated reporter mRNA [Acap-HAV-Nluc-poly(A)] level in Huh-7 cell extracts was slightly lower than in HeLa cell extracts (Figure 2B). On the other hand, the reporter mRNA level in HepG2 cell is similar to those in HeLa cell extracts (Supplementary Figure S3). These results imply that HAV IRES-mediated translation is more efficient in extracts from the liver-derived Huh-7 or HepG2 cells than in HeLa, which have a cervical origin, not because of mRNA stability but because of translation regulation.

80S Ribosome Assembly on a HAV IRES Reporter mRNA Is More Efficient in Huh-7 Extracts

We next performed sucrose density gradient assays to analyze which step of HAV IRES-mediated translation might be more



efficient in Huh-7 cell extracts vs. HeLa cell extracts. 28S rRNA and 18S rRNA, the components of 60S and 40S ribosomal subunits respectively, were mainly collected at fraction number 8-9 (bottom panels of **Figures 3A,B**). Thus, we defined fraction number 8-9 as the 80S fraction. Cap-Nluc-poly(A) mRNAs were mainly present in the ribosome-free fractions (fraction number 2-4) in both HeLa and Huh-7 cell extracts (**Figure 3A**). However, approximately 10% of cap-Nluc-poly(A) mRNAs were present with 80S fractions (fraction number 8-9) in HeLa and Huh-7 cell extract, indicating that the 80S initiation complex formed on the added mRNA suggesting that the mRNAs were translating. When Acap-HAV-IRES-Nluc-poly(A) mRNAs were used as reporters, only a small amount of Acap-HAV-IRES-Nluc-poly(A) mRNAs were present in 80S fractions in HeLa cell extracts, while approximately 20% of the reporter mRNAs were present in 80S fractions in Huh-7 cell extracts (**Figure 3B**). These results suggest that the 80S initiation complex is formed much more efficiently on the Acap-HAV-IRES-Nluc-poly(A) mRNA in Huh-7 extracts than in HeLa extracts. Thus, the difference in translational efficiencies between these two extracts is mediated mostly or exclusively at the level of translation initiation.

A HAV IRES Reporter mRNA Lacking a Poly(A) Tail Is Still Translated More Efficiently in Huh-7 Cell Extracts Than in HeLa Cell Extracts

The HAV mRNA is polyadenylated and the poly(A) tail is a known translational enhancer element. Thus, we next asked whether the enhanced HAV IRES-mediated translation in Huh-7 cell extracts depended on a poly(A) tail. We compared the translational activity of HAV IRES reporter mRNAs with or without poly(A) tails in HeLa and Huh-7 cell extracts. In both types of extracts, the relative translational activity of Acap-HAV-IRES-Nluc-poly(A) was 3-4 times lower than that of Acap-HAV-IRES-Nluc-poly(A) (**Figures 4A,B**). This was due to differences in translation efficiency, not mRNA stability, since levels of the reporter mRNAs at the end of the reactions were similar (**Figures 4A,B**). These results show that the poly(A) tail promotes translation of HAV IRES-containing mRNAs in both HeLa and Huh-7 cell extracts.

Next, we compared relative translational activity of the HAV IRES reporter mRNA lacking a poly(A) tail in HeLa cell extracts and Huh-7 cell extracts. The enhancement of translation mediated by HAV IRES in Huh-7 cell extracts still occurred even when the poly(A) tail was deleted from the reporter mRNAs (**Figures 4C,D**). This result clearly demonstrates that the difference in translational efficiencies of the HAV IRES in Huh-7 cell extracts vs. HeLa cell extracts is independent of the poly(A) tail.

Huh-7 Cell Extracts Contain an Activator of HAV IRES-Mediated Translation

In principle, there are two different mechanistic explanations for the difference in HAV IRES translational efficiency in Huh-7 and HeLa extracts: (i) HAV IRES-mediated translation is repressed in HeLa cell extracts by a repressor or (ii) HAV IRES-mediated translation is stimulated in Huh-7 cell extracts by an activator. To distinguish between these possibilities, we performed *in vitro* translation assays where we mixed a constant amount of Huh-7 cell extracts with increasing amounts of HeLa cell extract and *vice versa*.

First, we analyzed whether addition of HeLa cell extracts repress relative Nluc activity from Acap-HAV-IRES-Nluc-poly(A) mRNA in the Huh-7 *in vitro* translation system. Addition of increasing amounts of HeLa cell extracts to Huh-7 extracts did not affect HAV IRES-mediated translation (**Figure 5A**). As a control, we also analyzed the effect of adding Huh-7 cell extracts in parallel (**Figure 5B**). This also did not have any additional effect, as expected. Thus, if HeLa cells do not contain a repressor, it is not active under these assay conditions. Next, we analyzed whether addition of Huh-7 cell extracts to HeLa cell extracts increased relative luciferase activity of Acap-HAV-IRES-Nluc-poly(A) mRNA. The relative luciferase activity from Acap-HAV-IRES-Nluc-poly(A) reporter mRNA was enhanced by addition of increasing amounts of Huh-7 cell extracts to HeLa cell extracts (**Figure 5D**). This was a specific effect of

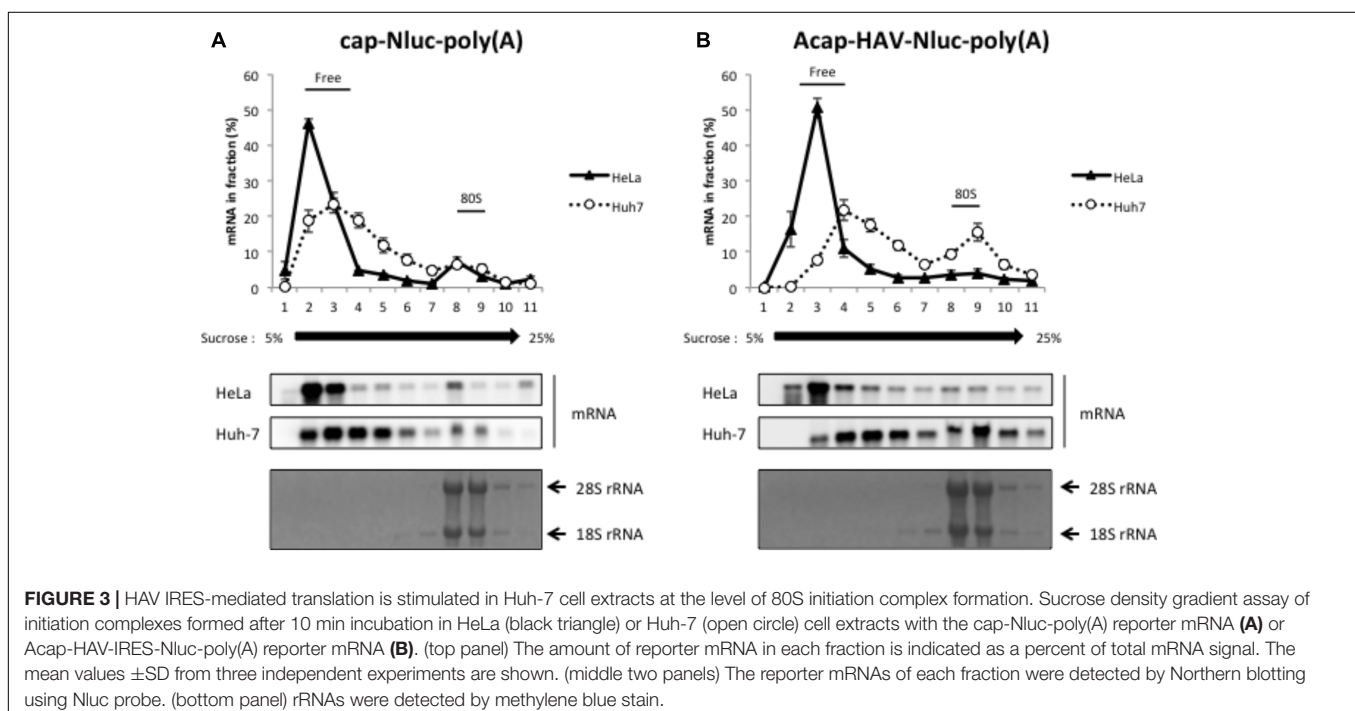
adding Huh-7 extracts, since addition of HeLa cell extracts to HeLa cell extracts did not affect the relative luciferase activity (**Figure 5C**). Taken together, these results strongly suggest that a component of Huh-7 cell extracts stimulates HAV IRES-mediated translation.

DISCUSSION

A fundamental aspect of viral biology is that different viruses infect different classes of cells, a phenomenon known as “tropism.” There is strong evidence that this can involve cell type-specific extracellular viral receptors, but there is also evidence that intracellular factors are involved (Henke et al., 2008; Schieck et al., 2013; Yan et al., 2013). Given the central importance of translation for viral replication, factors that regulate viral translation in a cell-selective manner could affect viral cell tropism. However, direct evidence for this idea has only been generated in a limited number of cases (Malnou et al., 2002; Guest et al., 2004). A limitation has been that most mechanistic studies using *in vitro* translation systems did not use extracts from the relevant cell types. In this study, we used cell-free translation systems to provide evidence for HAV hepatotropism at the translational level. By comparing HAV-IRES translational activity in extracts derived from target (liver) and non-target (cervical) cell types, we provide evidence for the existence of a factor in liver-derived cell types that functions in a poly(A) tail-independent manner to enhance translation initiation driven by the HAV IRES.

Hepatitis A Virus, a member of the picornavirus family, exhibits hepatotropism, meaning that it selectively infects liver cells, and thereby causes hepatitis A. Why does HAV infect

liver cells selectively? Dotzauer et al. (2000) reported that liver-specific expression of the asialoglycoprotein receptor (ASGPR) contributes to hepatotropism of HAV infection through direct interaction with HAV-specific immunoglobulin A. However, for other viruses such as PV and HCV, IRES-mediated translation occurs in a tissue-specific manner driven by cell-specifically expressed *trans*-acting factors (Guest et al., 2004; Henke et al., 2008). Therefore, we hypothesized that HAV IRES-mediated translation might also be a mechanism underlying HAV hepatotropism. To investigate this hypothesis, we established an *in vitro* cell-free translation system from liver hepatoma Huh-7 cell extracts. We directly compared translational activity mediated by the HAV IRES in these extracts to parallel extracts derived from HeLa cells, a non-liver cell line. Previously, a cell-free translation system with rabbit reticulocyte lysates (RRL) mixed with HeLa cell extracts has been utilized to analyze the HAV IRES-mediated translation mechanism (Ali et al., 2001; Borman et al., 2001; Michel et al., 2001). We thought that a cell-free translational system from liver derived cells was more likely to capture natural regulation of the HAV IRES and thus would be useful to study hepatotropism at the translational level. Thus, we adapted a cell-free translation system involving only human cell extracts (Fukao et al., 2009) to Huh-7 cells. Both cap- and poly(A)-dependent translation occurred in our *in vitro* translation system (**Figure 1**). In this study, cell extracts were not treated with micrococcal nuclease because the HAV IRES-mediated translation activity was decreased by nuclease treatment (data not shown). The reason why nuclease treatment decreases HAV IRES-mediated translation in Huh-7 cell extracts is not known, but endogenous RNA, including mRNA and miRNA, may be required for HAV IRES-mediated translation.



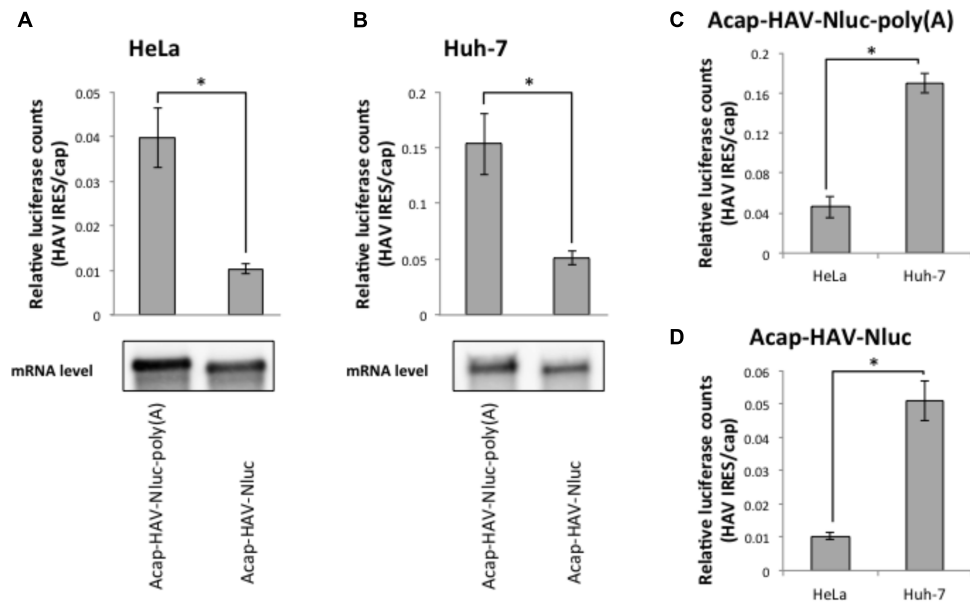


FIGURE 4 | Enhanced activity of the HAV IRES in Huh-7 vs. HeLa extracts is independent of the poly(A) tail. (top panel) HeLa (A) or Huh-7 (B) *in vitro* translation were performed with Acap-HAV-IRES-Nluc-poly(A) or Acap-HAV-IRES-Nluc reporter mRNAs. HAV IRES-mediated translation was normalized by cap-dependent translation. The mean values \pm SD from three independent experiments are shown. (bottom panel) Physical stabilities of the Acap-HAV-IRES-Nluc-poly(A) and Acap-HAV-IRES-Nluc mRNAs at the end of the incubation were analyzed by Northern blotting. (C) Comparison of Acap-HAV-IRES-Nluc-poly(A) relative light units between in HeLa and in Huh-7 cell extracts. (D) Comparison of Acap-HAV-IRES-Nluc relative light units between in HeLa and in Huh-7 cell extracts. The asterisks indicate in (A–D) statistically significant differences ($p < 0.05$).

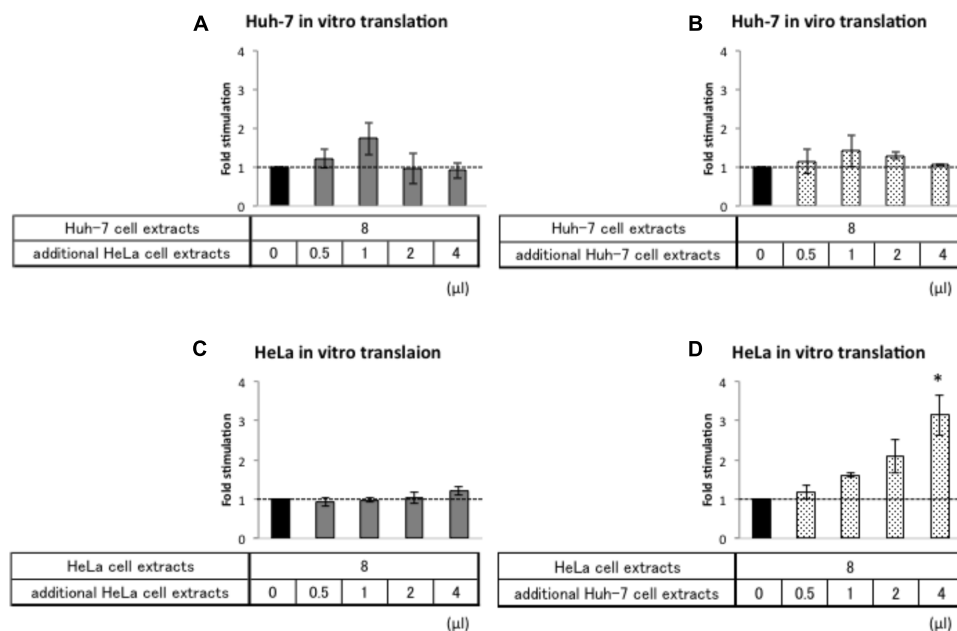


FIGURE 5 | Huh-7 cell extracts contain a factor that promotes HAV IRES-mediated translation. Huh-7-based (A,B) or HeLa-based (C,D) *in vitro* translation reaction were performed with Acap-HAV-IRES-Nluc-poly(A) reporter mRNA. *In vitro* translation reactions were carried out in the presence of additional cell extracts (gray bars: HeLa, dotted bars: Huh-7) or absence of them (black bar). Each bottom panel indicates the volume of additional cell extracts added. The total volume of reaction mixture was not changed. HAV IRES-mediated translation was normalized by cap-dependent translation. Fold stimulation by additional cell extracts was calculated by dividing relative light units of Acap-HAV-IRES-Nluc-poly(A) obtained in translational effector-added translation reactions by those in lysis buffer-added translation reactions, which are set as 1 (symbolized as dotted line). The mean values \pm SD from three independent experiments are shown. The asterisk indicates statistically significant differences in additional HeLa cell extracts (C) or Huh-7 cell extracts (D) in HeLa-based *in vitro* translation system ($p < 0.05$).

It has been reported that translation mediated by HAV IRES is efficient in Huh-7 cells by evaluation of the translational level of HAV IRES-conjugated mRNA derived from transfected plasmids (Mackiewicz et al., 2010). Our data about HAV IRES-mediated translation provide direct evidence for this proposal. Moreover, the results of sucrose density gradient analysis emphasize that HAV IRES-mediated translation undergoes translation initiation more efficiently in Huh-7 cell extracts than in HeLa cell extracts. These data provide direct evidence that the HAV mRNA translational mechanism is more efficient in extracts from targeted cell types. On the other hand, EMCV mRNA translation was not enhanced in Huh-7 cell extracts, compared with HAV mRNA (**Supplementary Figure S3**). These data means that HAV IRES-mediated translation is enhanced specifically in targeted cell extracts. Translation enhancement is one of the contributing factors to viral tropism, expected to be partially determined HAV hepatotropism on the translational level.

All picornavirus mRNAs are polyadenylated, and the poly(A) tail of picornaviruses is necessary for infectivity and minus-strand RNA synthesis (Spector and Baltimore, 1974; Hruby and Roberts, 1976; Herold and Andino, 2001). In this study, we analyzed the poly(A) tail function for HAV IRES-mediated translation. Our data suggested that the poly(A) tail stimulates HAV IRES-mediated translation in HeLa cell extracts, consistent with a previous report (Bergamini et al., 2000). This result also indicates that our *in vitro* translation system can analyze the poly(A)-dependency of cap-independent translation. In Huh-7 cell extracts, we confirmed that the poly(A) tail promoted HAV IRES-mediated translation. These results suggest that HAV IRES-mediated translation can be stimulated by a poly(A) tail in Huh-7 cell as well as in HeLa cell extracts. The poly(A) tail interacts with PABP, which mediates efficient translation by circularization of HAV IRES mRNA via an IRES-eIF4G-PABP-poly(A) interaction (Michel et al., 2001). Our results show that HAV IRES mRNA without a poly(A) tail is nevertheless translated more efficiently in Huh-7 cell extracts than in HeLa cell extracts. Thus, efficient translation of HAV IRES mRNA in Huh-7 cell extracts does not depend on mRNA circularization via the poly(A) tail. Rather our data are most consistent with recognition of the HAV IRES element by a *trans*-acting factor in Huh-7 cells that stimulates translation in a poly(A) tail-independent manner.

Known IRES *trans*-acting factors (ITAFs) can either promote IRES-mediated translation or repress it (Graff et al., 1998; Gosert

et al., 2000; Yi et al., 2000; Cordes et al., 2008). Moreover, for HCV, a specific miRNA is known to stimulate IRES-mediated translation (Henke et al., 2008). Our data clearly show that HAV IRES-mediated translation was promoted specifically by a liver cell-specific activator present in our Huh-7 cell translation extracts. This factor could be a liver-specific ITAF, miRNA, or a completely new class of translational regulatory factor. Identifying the factor will be crucial to fully elucidating the mechanism of HAV IRES-mediated translational enhancement in liver cells. We expect that our Huh-7 cell *in vitro* translation system will be instrumental for studying HAV mRNA translation. However, we also expect it to be useful for mechanistic analysis of liver-specific translational regulation of other mRNAs.

AUTHOR CONTRIBUTIONS

AS performed the measurements and he was involved in planning the work. AF performed the measurements and he was involved in planning and supervised the work. MK, NT, and YF aided in interpreting the results and worked on the manuscript. OT and KD processed the experimental data, performed the analysis, drafted the manuscript, and designed the figures. TF contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. All authors discussed the results and commented on the manuscript.

FUNDING

This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (“Neo-taxonomy of non-coding RNAs”) and a Grant-in-Aid for Scientific Research (B) from Japan Ministry of Education, Culture, Sports, Science and Technology. Part of this study was supported by the MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2014–2018 (S1411037).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00307/full#supplementary-material>

REFERENCES

- Ali, I. K., Mckendrick, L., Morley, S. J., and Jackson, R. J. (2001). Activity of the hepatitis A virus IRES requires association between the cap-binding translation initiation factor (eIF4E) and eIF4G. *J. Virol.* 75, 7854–7863. doi: 10.1128/Jvi.75.17.7854-7863.2001
- Bergamini, G., Preiss, T., and Hentze, M. W. (2000). Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* 6, 1781–1790. doi: 10.1017/S1355838200001679
- Borman, A. M., Bailly, J. L., Girard, M., and Kean, K. M. (1995). Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation *in vitro*. *Nucleic Acids Res.* 23, 3656–3663. doi: 10.1093/nar/23.18.3656
- Borman, A. M., Le Mercier, P., Girard, M., and Kean, K. M. (1997). Comparison of picornaviral IRES-driven internal initiation of translation in cultured cells of different origins. *Nucleic Acids Res.* 25, 925–932. doi: 10.1093/nar/25.5.925
- Borman, A. M., Michel, Y. M., and Kean, K. M. (2001). Detailed analysis of the requirements of hepatitis A virus internal ribosome entry segment for the eukaryotic initiation factor complex eIF4F. *J. Virol.* 75, 7864–7871. doi: 10.1128/jvi.75.17.7864-7871.2001
- Brown, E. A., Day, S. P., Jansen, R. W., and Lemon, S. M. (1991). The 5' nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation *in vitro*. *J. Virol.* 65, 5828–5838.
- Chang, J., Nicolas, E., Marks, D., Sander, C., Lerro, A., Buendia, M. A., et al. (2004). miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA

- and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol.* 1, 106–113. doi: 10.4161/rna.1.2.1066
- Cordes, S., Kusov, Y., Heise, T., and Gauss-Muller, V. (2008). La autoantigen suppresses IRES-dependent translation of the hepatitis A virus. *Biochem. Biophys. Res. Commun.* 368, 1014–1019. doi: 10.1016/j.bbrc.2008.01.163
- Dotzauer, A., Gebhardt, U., Bieback, K., Göttke, U., Kracke, A., Mages, J., et al. (2000). Hepatitis A virus-specific immunoglobulin A mediates infection of hepatocytes with hepatitis A virus via the asialoglycoprotein receptor. *J. Virol.* 74, 10950–10957. doi: 10.1128/JVI.74.23.10950-10957.2000
- Duncan, K. E., Strein, C., and Hentze, M. W. (2009). The SXL-UNR corepressor complex uses a PABP-mediated mechanism to inhibit ribosome recruitment to msl-2 mRNA. *Mol. Cell* 36, 571–582. doi: 10.1016/j.molcel.2009.09.042
- Fu, H., Tie, Y., Xu, C., Zhang, Z., Zhu, J., Shi, Y., et al. (2005). Identification of human fetal liver miRNAs by a novel method. *FEBS Lett.* 579, 3849–3854. doi: 10.1016/j.febslet.2005.05.064
- Fukao, A., Mishima, Y., Takizawa, N., Oka, S., Imataka, H., Pelletier, J., et al. (2014). MicroRNAs trigger dissociation of eIF4AI and eIF4AII from target mRNAs in humans. *Mol. Cell* 56, 79–89. doi: 10.1016/j.molcel.2014.09.005
- Fukao, A., Sasano, Y., Imataka, H., Inoue, K., Sakamoto, H., Sonenberg, N., et al. (2009). The ELAV protein HuD stimulates cap-dependent translation in a Poly(A)- and eIF4A-dependent manner. *Mol. Cell* 36, 1007–1017. doi: 10.1016/j.molcel.2009.11.013
- Gosert, R., Chang, K. H., Rijnbrand, R., Yi, M., Sangar, D. V., and Lemon, S. M. (2000). Transient expression of cellular polypyrimidine-tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites in vivo. *Mol. Cell. Biol.* 20, 1583–1595. doi: 10.1128/MCB.20.5.1583-1595.2000
- Graff, J., Cha, J., Blyn, L. B., and Ehrenfeld, E. (1998). Interaction of poly(rC) binding protein 2 with the 5' noncoding region of hepatitis A virus RNA and its effects on translation. *J. Virol.* 72, 9668–9675.
- Green, K. M., Linsalata, A. E., and Todd, P. K. (2016). RAN translation-What makes it run? *Brain Res.* 1647, 30–42. doi: 10.1016/j.brainres.2016.04.003
- Guest, S., Pilipenko, E., Sharma, K., Chumakov, K., and Roos, R. P. (2004). Molecular mechanisms of attenuation of the Sabin strain of poliovirus type 3. *J. Virol.* 78, 11097–11107. doi: 10.1128/JVI.78.20.11097-11107.2004
- Hellen, C. U., Pestova, T. V., and Wimmer, E. (1994). Effect of mutations downstream of the internal ribosome entry site on initiation of poliovirus protein synthesis. *J. Virol.* 68, 6312–6322.
- Henke, J. I., Goergen, D., Zheng, J., Song, Y., Schüttler, C. G., Fehr, C., et al. (2008). microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J.* 27, 3300–3310. doi: 10.1038/emboj.2008.244
- Herold, J., and Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell* 7, 581–591. doi: 10.1016/S1097-2765(01)00205-2
- Hruby, D. E., and Roberts, W. K. (1976). Encephalomyocarditis virus RNA: variations in polyadenylic acid content and biological activity. *J. Virol.* 19, 325–330.
- Hunt, S. L., and Jackson, R. J. (1999). Polypyrimidine-tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. *RNA* 5, 344–359. doi: 10.1017/S1355838299981414
- Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.* 17, 7480–7489. doi: 10.1093/emboj/17.24.7480
- Jackson, R. J., Hellen, C. U., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127. doi: 10.1038/nrm2838
- Jang, S. K., Kräusslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, A. C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* 62, 2636–2643.
- Jünemann, C., Song, Y., Bassili, G., Goergen, D., Henke, J., and Niepmann, M. (2007). Picornavirus internal ribosome entry site elements can stimulate translation of upstream genes. *J. Biol. Chem.* 282, 132–141. doi: 10.1074/jbc.M608750200
- Kaminski, A., Belsham, G. J., and Jackson, R. J. (1994). Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. *EMBO J.* 13, 1673–1681.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* 12, 735–739. doi: 10.1016/S0960-9822(02)00809-6
- Le, H., Tanguay, R. L., Balasta, M. L., Wei, C. C., Browning, K. S., Metz, A. M., et al. (1997). Translation initiation factors eIF-iso4G and eIF-4B interact with the poly(A)-binding protein and increase its RNA binding activity. *J. Biol. Chem.* 272, 16247–16255. doi: 10.1074/jbc.272.26.16247
- Lemon, S. M., Ott, J. J., Van Damme, P., and Shouval, D. (2017). Type A viral hepatitis: a summary and update on the molecular virology, epidemiology, pathogenesis and prevention. *J. Hepatol.* doi: 10.1016/j.jhep.2017.08.034 [Epub ahead of print].
- Mackiewicz, V., Cammas, A., Desbois, D., Marchadier, E., Pierredon, S., Beaulieux, F., et al. (2010). Nucleotide variability and translation efficiency of the 5' untranslated region of hepatitis A virus: update from clinical isolates associated with mild and severe hepatitis. *J. Virol.* 84, 10139–10147. doi: 10.1128/JVI.02598-09
- Malnou, C. E., Poyry, T. A. A., Jackson, R. J., and Kean, K. M. (2002). Poliovirus internal ribosome entry segment structure alterations that specifically affect function in neuronal cells: molecular genetic analysis. *J. Virol.* 76, 10617–10626. doi: 10.1128/jvi.76.21.10617-10626.2002
- Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G., and Diaz-Toledano, R. (2015). Picornavirus IRES elements: RNA structure and host protein interactions. *Virus Res.* 206, 62–73. doi: 10.1016/j.virusres.2015.01.012
- Michel, Y. M., Borman, A. M., Paulous, S., and Kean, K. M. (2001). Eukaryotic initiation factor 4G-poly(A) binding protein interaction is required for poly(A) tail-mediated stimulation of picornavirus internal ribosome entry segment-driven translation but not for X-mediated stimulation of hepatitis C virus translation. *Mol. Cell. Biol.* 21, 4097–4109. doi: 10.1128/MCB.21.13.4097-4109.2001
- Moteki, S., and Price, D. (2002). Functional coupling of capping and transcription of mRNA. *Mol. Cell.* 10, 599–609. doi: 10.1016/S1097-2765(02)00660-3
- Niepmann, M. (2009). Internal translation initiation of picornaviruses and hepatitis C virus. *Biochim. Biophys. Acta* 1789, 529–541. doi: 10.1016/j.bbagr.2009.05.002
- Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–325. doi: 10.1038/334320a0
- Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I., et al. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev.* 14, 2028–2045. doi: 10.1101/gad.14.16.2028
- Radomski, N., and Jost, E. (1995). Molecular cloning of a murine cDNA encoding a novel protein, p38-2G4, which varies with the cell cycle. *Exp. Cell Res.* 220, 434–445. doi: 10.1006/excr.1995.1335
- Richard, P., and Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23, 1247–1269. doi: 10.1101/gad.1792809
- Schieck, A., Schulze, A., Gahler, C., Muller, T., Haberkorn, U., Alexandrov, A., et al. (2013). Hepatitis B virus hepatotropism is mediated by specific receptor recognition in the liver and not restricted to susceptible hosts. *Hepatology* 58, 43–53. doi: 10.1002/hep.26211
- Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004). Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* 5:R13. doi: 10.1186/gb-2004-5-3-r13
- Song, Y., Friebe, P., Tzima, E., Junemann, C., Bartenschlager, R., and Niepmann, M. (2006). The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. *J. Virol.* 80, 11579–11588. doi: 10.1128/JVI.00675-06
- Spector, D. H., and Baltimore, D. (1974). Requirement of 3'-terminal poly(adenylic acid) for the infectivity of poliovirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 71, 2983–2987. doi: 10.1073/pnas.71.8.2983
- Tarun, S. Z. J., and Sachs, A. B. (1995). A common function for mRNA 5' and 3' ends in translation initiation in yeast. *Genes Dev.* 9, 2997–3007. doi: 10.1101/gad.9.23.2997
- Thoma, C., Bergamini, G., Galy, B., Hundsdoerfer, P., and Hentze, M. W. (2004). Enhancement of IRES-mediated translation of the c-myc and BiP mRNAs by

- the poly(A) tail is independent of intact eIF4G and PABP. *Mol. Cell* 15, 925–935. doi: 10.1016/j.molcel.2004.08.021
- Yan, H., Peng, B., He, W., Zhong, G., Qi, Y., Ren, B., et al. (2013). Molecular determinants of hepatitis B and D virus entry restriction in mouse sodium taurocholate cotransporting polypeptide. *J. Virol.* 87, 7977–7991. doi: 10.1128/JVI.03540-12
- Yi, M., Schultz, D. E., and Lemon, S. M. (2000). Functional significance of the interaction of hepatitis A virus RNA with glyceraldehyde 3-phosphate dehydrogenase (GAPDH): opposing effects of GAPDH and polypyrimidine tract binding protein on internal ribosome entry site function. *J. Virol.* 74, 6459–6468. doi: 10.1128/JVI.74.14.6459-6468.2000
- Conflict of Interest Statement:** 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.

Copyright © 2018 Sadahiro, Fukao, Kosaka, Funakami, Takizawa, Takeuchi, Duncan and Fujiwara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics

Fouzia Yeasmin¹, Tetsushi Yada² and Nobuyoshi Akimitsu^{1*}

¹ Isotope Science Centre, The University of Tokyo, Tokyo, Japan, ² Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Fukuoka, Japan

OPEN ACCESS

Edited by:

Kinji Ohno,
Nagoya University, Japan

Reviewed by:

Malgorzata Kloc,
Houston Methodist Research Institute,
United States
Jonathan Perreault,
Institut National de la Recherche
Scientifique (INRS), Canada

*Correspondence:

Nobuyoshi Akimitsu
akimitsu@ric.u-tokyo.ac.jp

Specialty section:

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

Received: 23 January 2018

Accepted: 09 April 2018

Published: 25 April 2018

Citation:

Yeasmin F, Yada T and Akimitsu N
(2018) Micropeptides Encoded in
Transcripts Previously Identified as
Long Noncoding RNAs: A New
Chapter in Transcriptomics and
Proteomics. *Front. Genet.* 9:144.
doi: 10.3389/fgene.2018.00144

Integrative analysis using omics-based technologies results in the identification of a large number of putative short open reading frames (sORFs) with protein-coding capacity within transcripts previously identified as long noncoding RNAs (lncRNAs) or transcripts of unknown function (TUFs). sORFs were previously overlooked because of their diminutive size and the difficulty of identification by bioinformatics analyses. There is now growing evidence of the existence of potentially functional micropeptides produced from sORFs within cells of diverse species. Recent characterization of a few of these revealed their significant divergent roles in many fundamental biological processes, where some also show important relationships with pathogenesis. Recent works therefore provide new insights for exploring the wealth of information that may lie within sORF-encoded short proteins. Here, we summarize the current progress and view of micropeptides encoded in sORFs of protein-coding genes.

Keywords: lncRNAs, TUFs, sORFs, micropeptides, translation

INTRODUCTION

Identification of a large number of RNA transcripts by genome-wide analysis suggests a complex network of transcripts that includes tens of thousands of long noncoding RNAs (lncRNAs) and transcripts of unknown function (TUFs) (Carninci et al., 2005; Willingham et al., 2006; Birney et al., 2007; Kapranov et al., 2007). Recent studies have suggested that lncRNAs and TUFs in the human genome represent the greatest source for short open reading frames (sORFs), which were previously overlooked because of their small size and the lack of evidence for “codingness” (Frith et al., 2006; Cohen, 2014; Pauli et al., 2015). As a result, sORFs embedded in lncRNAs and TUFs have not been adequately studied.

sORF-encoded micropeptides first attracted the attention of a group of scientists during their study of lncRNA (Rohrig et al., 2002). From that point, many studies have been carried out to identify potential sORF candidates, and whether there are any more of them that can encode functional micropeptides. Recent advancements in bioinformatics, proteomics and transcriptomics have revealed that traditional computational algorithms used in searches for many potent ORFs may have included oversights as many studies have now identified hundreds of non-annotated

sORFs that have coding potential for micropeptides (Ingolia et al., 2011; Slavoff et al., 2013; Bazzini et al., 2014) from yeast (Smith et al., 2014) to plants (Hanada et al., 2013; Laressergues et al., 2015) and humans (Ingolia et al., 2014; Ma et al., 2014). sORF-encoded proteins have emerged as a new, functional class because of their role in many biological activities (Crappé et al., 2014). The diverse biological functions of this new group of short proteins have attracted the attention of the scientific community and increased interest in studying them in more detail (Saghatelian and Couso, 2015; Makarewich and Olson, 2017).

Here, we give a brief overview of the various approaches recently used to identify sORF- encoded micropeptides and their biological function. Based on the results of previous studies, we also try to identify the potential ideas and strategies that can be implemented to characterize other micropeptides' functionalities. Finally, we review the diverse biological function of micropeptides that have been found up until recently, from plants to animals. These suggesting that many biologically significant micropeptides may be concealed in the hidden world of proteomes.

**MORE DEVELOPED TECHNIQUES
IDENTIFY MORE POTENT
sORF-ENCODED MICROPEPTIDES**

Traditional computational prediction of protein-coding ORFs relies on a number of stringent criteria to remove meaningless ORFs, such as size cutoff of 300 nucleotides, AUG start

codon usage, and sequence conservation (Gish and States, 1993; Kochetov, 2005), rendering them inappropriate for sORF detection. Hunting for these tiny treasures has therefore posed a great challenge.

However, with the advancement of technology, the challenge has begun to be addressed effectively. Both computational and experimental approaches have made it easier to explore the complexity of the small proteome. Several approaches have been taken to systematically annotate sORFs with coding potential. Along with other conventional strategies, such as cross-species comparison, examination of codon content and coding features used to identify ORFs, various metrics and methods have been developed and are playing prominent roles in identifying putative sORFs (Table 1).

Ribosome profiling has emerged as a technique for comprehensively and quantitatively measuring translation (Ingolia et al., 2014; Smith et al., 2014). Based on modification of ribosome foot printing, it is mainly premised on deep sequencing of ribosome-protected mRNA fragments to obtain a global snapshot of translation. Application of ribosome profiling has provided several key findings, including prodigious use of non-ATG initiation codons, as well as identification of polycistronic genes, upstream ORFs and overlapping ORFs. Hundreds of putative non-annotated protein-coding sORFs have recently been identified in eukaryotic genomes by using this technique (Ingolia et al., 2011; Bazzini et al., 2014).

However, ribosome occupancy does not always mean true translation, as indicated by the identification of many well-characterized nuclear lncRNAs in a ribosome profiling assay (Brannan et al., 1990; Guttman et al., 2013). Many ORFs

TABLE 1 | Computational and experimental approaches to protein-coding sORFs.

Metrics and methods to identify sORF (including both computational and experimental)		Description	References
Computing-based method	sORFinder, HAltORF, uPEPPERoni	Web based tools to locate sORF having coding potential	Hanada et al., 2010; Vanderperre et al., 2012; Skarshewski et al., 2014
	PhyloCSF	A computational method examines evolutionary conservation of a sORF across species	Lin et al., 2011
Transcriptomic-based method	Ribosome profiling	A deep sequencing- based tool of ribosome protected mRNA fragments to obtain a global snapshot of translation	Ingolia et al., 2011
	Poly-ribo seq	A combination of ribosome profiling and polysome to enrich more potent protein coding ORFs	Aspden et al., 2014
	Ribosome releasing scores (RRS)	These three metrics are developed and combined with ribosome profiling to assist in identification of true protein coding ORFs	Guttman et al., 2013
	Fragment length organization similarity score (FLOSS)		Ingolia et al., 2014
	ORF regression algorithm for translation evaluation RPFS (ribosome-protected mRNA fragments) (ORF-RATER)		Fields et al., 2015
Proteomics-based	Proteo genomics	A combined approach of proteomics and genomics	Slavoff et al., 2013

are associated with ribosomes to regulate the translation of downstream ORFs. This suggests ribosome profiling is not sufficient evidence of protein synthesis. To differentiate more effective protein-coding transcripts from noncoding RNAs, several algorithms and metrics have been developed based on their ribosome-profiling characteristics, including RRS (Guttman et al., 2013), FLOSS (Ingolia et al., 2014), ORF-RATER (Fields et al., 2015), and Ribo taper (Calviello et al., 2016).

Poly-Ribo-Seq, a modification of a ribosome-profiling method, enriches polysomes that are more likely to be actively translating mRNA into proteins. Poly-Ribo-Seq was successfully used to identify several sORFs in the *Drosophila* genome (Galindo et al., 2007; Aspden et al., 2014).

Mass spectrometry (MS) peptidomics and proteomics experiments have recently been applied to identify sORF-encoded micropeptides. MS is advantageous compared with ribosome profiling, as it directly detects the peptide generated from ORFs and therefore validates the production of peptides. However, the bias of MS toward more abundant proteins means it only detects the peptides abundant in cells. Analysis of tandem mass spectrometry (MS/MS) data that mapped expressed peptides to their encoding genomic loci and transcriptome data generated by ENCODE has identified 85 unique peptides that match with 69 lncRNAs (Bánfai, 2012). Slavoff et al. developed a modified proteomic strategy, known as proteogenomics to identify and validate more potent sORFs, wherein they compiled a custom mRNA-seq derived polypeptide database to identify MS fragmentation spectra. In this approach, the proteome is enriched to isolate small polypeptides before proteomic analysis. Through this strategy, 86 uncharacterized sORFs (sORF-encoded polypeptides) of 90 were identified in K562 cells (Slavoff et al., 2013). There are also still some difficulties to consider. The average tissue content of micropeptides is very low, and they are often subjected to degradation or loss during sample preparation, which further impedes their identification. As a result, many micropeptides produced in cells may be absent in MS analysis. New and alternative extraction methods may prove more effective in extracting and identifying micropeptides. For example, Schwaid et al. described an affinity-based approach that could enrich and identify cysteine-containing human sORF-encoded polypeptides (ccSEPs) in cells. They were able to identify 16 novel sSEPs from previously uncharacterized sORFs (Schwaid et al., 2013). MS-based methods have thus, to date, identified a limited number of micro-proteins.

sORF-ENCODED MICROPEPTIDES: INSIGHTS INTO THEIR FUNCTION

Small peptides have high recognition because of their important roles in diverse biological processes (Fricker, 2005; Boonen et al., 2009; Cabrera-Quio et al., 2016). The largest and most extensively studied class of small peptides are classical bioactive peptides, which are derived from larger precursor proteins and contain N-terminal signal sequences. Hormones and neuropeptides are considered the best examples of bioactive molecules (Hashimoto et al., 2001; Cunha et al., 2008). Most of these peptides

act as ligands of membrane receptors (Boonen et al., 2009). Micropeptides differ from these bioactive small peptides in that they are not processed from large peptides but rather are translated from sORFs previously identified as lncRNAs and TUFs. Four initial studies (Rohrig et al., 2002; Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007) were pioneering in opening up new avenues for sORF research. Their studies showed how a sORF can be involved in different developmental contexts with apparently different biological roles during morphogenesis.

As described above, advancements in technologies over the past few years have led to the discovery of several hundred of putative coding sORFs in various species. However, it is still unknown how many of these newly discovered sORF-encoded peptides are functional. Existence of a peptide does not always imply it has a function. Experimental demonstration is important in revealing their biological effects. Several approaches can be used to validate candidate-translated sORFs (Housman and Ulitsky, 2016). Recently some micropeptides have been characterized and found to play important roles in fundamental biological processes such as RNA decapping (D'Lima et al., 2017), DNA repair (Slavoff et al., 2014), stress signaling (Matsumoto et al., 2017), apoptosis (Guo et al., 2003), muscle formation (Bi et al., 2017), metabolic homeostasis (Lee et al., 2015), and calcium homeostasis (Magny et al., 2013; Anderson et al., 2015, 2016; Nelson et al., 2016; **Figure 1**). The following section briefly explains commonly used strategies for deciphering the functions of short proteins that are necessary for their characterization (**Figure 2**).

IN SILICO (OR COMPUTATIONAL) CHARACTERIZATION

Evolutionary conservation is an important sign that a gene is functional. One hallmark of the sORFs studied thus far is evolutionary conservation of micropeptides. An evolutionary conserved micropeptide called *polished rice* (*pri*) or *tarsal-less* (*tal*) was identified in *Drosophila*, while the *Tribolium* orthologue is known as *mille-pattes* (*mlpt*) (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007). These micropeptides were characterized based on their conservation. Homology-based searching among species for unannotated micropeptides may be performed to predict any conserved biological function (**Figure 2**). The best example of homology-based characterization is the identification of a group of micropeptides, namely, myoregulin (MLN), phospholamban (PLN), and sarcolipin (SLN). They share conserved peptide sequences from flies to vertebrates involved in Ca^{2+} homeostasis through inhibiting SERCA activity (Magny et al., 2013) in muscle. There is a sequence and structural similarity among these peptides. Later, another two micropeptides, endoregulin (ELN), and another-regulin (ALN), were also characterized based on their shared amino acids, and found to show similar functions to MLN/PLN/SLN, but in nonmuscle cell types (Anderson et al., 2016).

Thus, identification and characterization based on sequence features is a reasonable approach for deciphering the biological

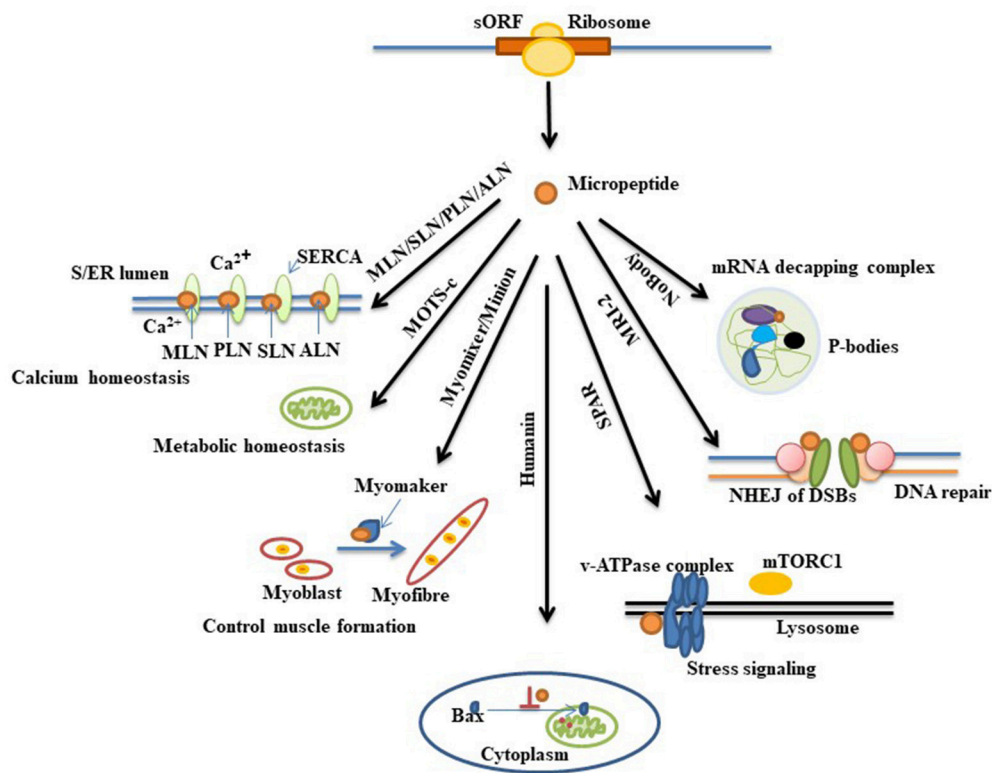


FIGURE 1 | Diverse biological function of recently annotated micropeptides. Micropeptides are found to be involved in many biological processes. Myoregulin (MLN), phospholamban (PLN), sarcoplipin (SLN), and another regulin (ALN) are a group of peptides that interact with the protein SERCA (a Ca^{2+} Pump) in sarcoplasmic and endoplasmic reticulum (S/ER) and maintain Ca^{2+} homeostasis in the cell. MOTS-c and humanin are mitochondrial sORF-encoded micropeptides that display important roles in metabolic homeostasis and apoptosis, respectively. Humanin suppresses apoptosis by preventing the translocation of an apoptosis inducing protein, Bax (Bcl2-associated X protein), from cytoplasm to mitochondria. Another micropeptide named MRI-2 is found to enhance non-homologous end joining (NHEJ) of double-strand DNA breaks (DSBs) by associating with other DNA end-binding proteins (Ku proteins). Myomixer, minion, SPAR, and NoBody, four other micropeptides that have been recently discovered, have distinct biological roles wherein myomixer and minion stimulate the fusion of myoblast to form myofiber during muscle formation by participating with another protein, myomaker. The micropeptide SPAR is localized into lysosome where it interacts with the lysosomal v-ATPase complex and regulates mTORC1 protein activation during stress signaling. NoBody, a p-body (processing-body, which is involved in mRNA turnover) dissociating micropeptide, shows its function by interacting with the mRNA decapping complex.

function of new unannotated micropeptides. Computational predictions of functional sORFs use several key features to identify potential sORFs. Canonical protein-coding ORFs show striking sequence features as measured by the ratio of Ka and Ks (Ka/ Ks < 1, the ratio of synonymous versus nonsynonymous codon substitution), suggesting that canonical protein coding genes are under selective pressure during evolution. Compared with canonical protein coding genes, it is difficult to score statistically significant values for very short sequences because the number of possible changes is low (Ladoukakis et al., 2011). Mackowiak and his group brought a new computational approach to identify conserved sORFs using comparative genomics (Mackowiak et al., 2015). Three qualitative features of coding sequence conservation specific to known micropeptides and canonical proteins were analyzed in their study. The first is the conservation of amino acid sequences by phylogenetic codon substitution frequencies (PhyloCSF). Second is the conservation of the reading frame, which is the conservation of in-frame start and stop codons in related species. The third is a drop

in nucleotide sequence conservation around the start and stop codons using PhastCons (Siepel et al., 2005). The combination of these three features has identified about 2,000 sORFs in five systems: human, mouse, zebrafish, fruit fly, and the nematode *Caenorhabditis elegans*. Translation and protein expression of some of these predicted sORFs have also been confirmed by experimental evidence.

Although functional characterization of sORFs based on sequence conservation is useful, it is not applicable for all. Some non-conserved sORFs may evolve as newly coding ORFs that can also be present and be involved with regulatory functions.

FUNCTIONAL PROTEOMICS

Although some sORFs are found to be highly conserved across species, most show relatively low sequence conservation compared with known protein-coding genes (Carvunis et al., 2012; Slavoff et al., 2013). Therefore, although homology-based functional characterization is reasonable, as mentioned above, it

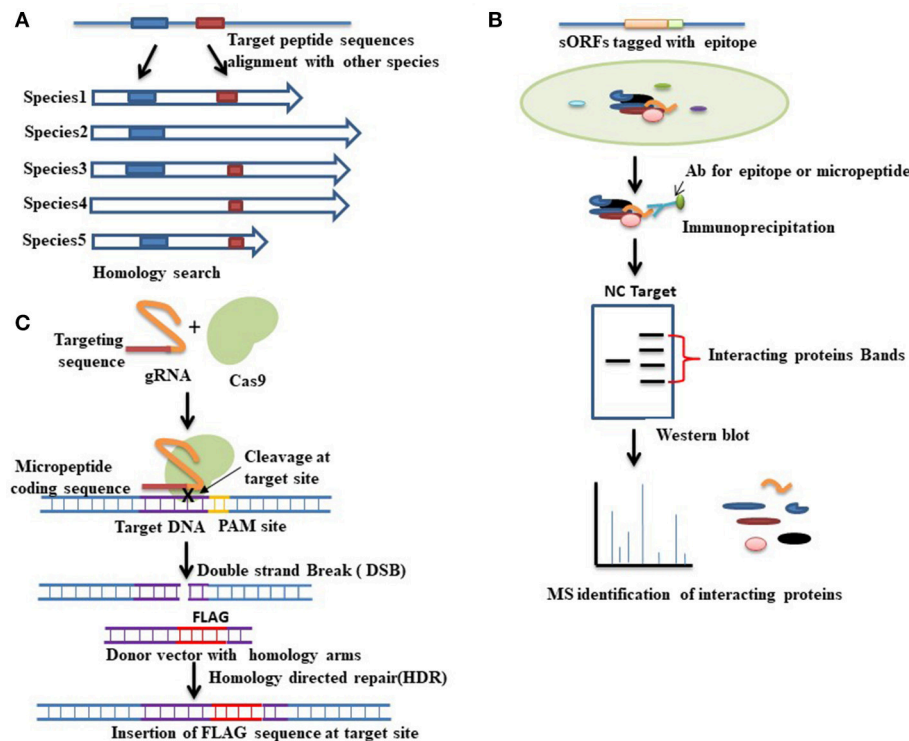


FIGURE 2 | Various approaches for functional characterization of micropeptides. **(A)** Evolutionary conservation of a peptide sequence is suggestive of functionality. Homology- based searching among species thus can be performed to identify whether the target peptide sequence shares any functional similarity with other proteins. Here the blue and red boxes indicate the conserved sequences among species. **(B)** Functional proteomics is a commonly used approach for identifying the interacting proteins of a target protein. In this method, first, immunoprecipitation is conducted by using an antibody (Ab) that is designed either against the epitope tagged with a target micropeptide or directly against the micropeptide. Western blot is then performed followed by mass spectrometry analysis to separate and identify the interacting proteins. Red brackets indicate the bands of interacting proteins that are separated by western blot analysis. A negative control (NC) denotes an empty vector that also runs for comparison. The nature of the interacting protein will thus provide clues about the function of the target micropeptide. **(C)** CRISPR-cas9 mediated gene editing approaches can also be used to check the coding potential of sORFs. To verify the coding potential, an epitope tag (FLAG) can be inserted at the downstream of the sORF into the endogenous locus. CRISPR-cas9 mediated gene editing is started by the recognition of the target site, which is mediated by a guide RNA (gRNA). Guide RNA guides the cas9 endonuclease to a specific location in the genome sequence, which is immediately adjacent to a protospacer adjacent motif (PAM). Upon recognition, the cas9 creates a double strand break (DSB) at the target site. This DSB can then be repaired either by non-homologous end joining (NHEJ) or by homology directed repair (HDR). HDR is used to insert an epitope tag at the target site where a donor vector with homology to the targeted locus must be provided. The donor vector must contain the epitope tag that has to be knocked-in at the target site. Expression of the engineered fusion protein can then be verified by western blot analysis.

has difficulty finding species-specific functional peptides. Several of the micropeptides characterized thus far exert their functions by interacting with other proteins. Several studies have applied functional proteomics successfully to identify the interacting partners. For example, Matsumoto and colleagues employed functional proteomics to study a LINC00961-encoded short protein. This micropeptide interacts with the lysosomal v-ATPase complex to regulate mTORC1 (a rapamycin protein complex) activation (**Figure 1**) and muscle regeneration. This interaction with the v-ATPase complex and regulation of mTORC1 is specific to the amino acid response. It is therefore known as a small regulatory polypeptide of the amino acid response, or SPAR (Matsumoto et al., 2017).

By employing functional proteomics, another group also characterized and identified the biological significance of another unreported micropeptide, named NoBody (D'Lima et al., 2017). By performing immunoprecipitation and MS analysis, the

researchers found NoBody to be a component of the mRNA decapping protein complex that cross-links to EDC4 (enhancer of mRNA decapping 4). The mRNA decapping complex removes the 5' cap from mRNAs to promote 5'-3' decay. Molecular components of this pathway localize to p-bodies. Manipulation of NoBody expression is anticorrelated with the P-body number. NoBody regulates the P-body number in cells by interacting with decapping proteins. This micropeptide is therefore called the non-annotated P-body dissociating polypeptide (NoBody).

However, traditional immunoprecipitation methods very often result in the enrichment of many nonspecific interactions of micropeptides. For example, functional proteomics analysis of a micropeptide named modulator of retroviral infection (MRI) has revealed that it is associated with ku70 and ku80, two essential proteins that are involved in the nonhomologous end joining DNA repairing mechanism (Slavoff et al., 2014). Association of MRI with ku70/ku80 suggests that it is involved in the cellular

DNA repairing mechanism. Although the immunoprecipitation of MRI also enriched for heat shock protein 70 family members protein, imaging studies ruled out cytosolic heat shock proteins as bona fide interactors that might be formed after the cells are lysed during the immunoprecipitation (Slavoff et al., 2014; Grundy et al., 2016). Such a problem thus demands a better approach for identifying micropeptide associated proteins and protein complexes. Recently Chu and colleagues applied an in-situ proximity tagging method to elucidate micropotein-protein interactions (MPIs) for an uncharacterized micropotein called c11orf98 (Chu et al., 2017). This method relies on an engineered ascorbate peroxidase (APEX) (Rhee et al., 2013). When APEX fusion protein is expressed in the cells and treated with hydrogen peroxide (H_2O_2) in the presence of biotin-phenol, the proteins proximal to the APEX fusion protein are labeled with biotin. The proteins, that are biotinylated, can then be enriched and analyzed by MS. Thus, the analysis of biotinylated proteins provides valuable information about the protein environment of fusion protein. Since the interactions take place in the context of a living cell, the enrichment of nonspecific interactors is reduced. By applying this approach, it was revealed that c11orf98 interacts with nucleolar proteins nucleoplasm and nucleolin (Chu et al., 2017), which suggests that the application of APEX tagging is useful to characterize uncharacterized micropeptides.

These studies suggest that functional proteomics may be implemented to understand the function and biological nature of an unannotated short protein through identifying direct binding partners or components (Figure 2).

GENE EDITING APPROACHES

Recently developed Clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (cas9) mediated gene editing technology has become a powerful approach among scientists to study a gene's function. CRISPR-cas9 mediated gene editing strategies can also be used for identifying and verifying coding potential of sORF encoded peptides. An epitope tag can be knocked-in into the endogenous locus of a micropeptide in-frame with the predicted sORF to produce a fusion protein using CRISPR/cas9-mediated homologous recombination (Figure 2). Detection of the engineered fusion protein by western blot analysis provides the evidence that the mRNA is translated into a stable peptide. This powerful knock-in technique also simplifies many downstream applications that are important for functional characterizing of a gene. For example, immunoprecipitation to identify binding partners of the target proteins. Immunocytochemistry can also be performed in epitope-tagged samples to check the subcellular localization of the fusion protein, which may provide important information about its involvement in biological processes. Recently some research groups have implemented this new technology to verify sORF-encoded peptides (Galindo et al., 2007; Slavoff et al., 2014; Anderson et al., 2015). By using CRISPR-cas9 homologous recombination, an epitope tag was inserted at the downstream of the sORF to confirm whether the sORF containing gene was actively transcribed from its

native chromosomal context and translated into a stable peptide. Identification and validation of some sORF-encoded peptides by CRISPR-cas9 mediated gene editing technologies thus indicate the possible successful application of them in identifying and verifying other sORF-encoded peptides.

DIVERSE BIOLOGICAL FUNCTIONS OF MICROPEPTIDES

In Plants

The first eukaryotic micropeptide was identified in plants by a group of researchers studying legumes. A gene called early nodulin 40 (Enod40), previously annotated as lncRNA, was found to encode two short peptides of 12 and 24 amino acids (AAs) in plants, where they interact with a sucrose-synthesizing enzyme during root nodule organogenesis (Rohrig et al., 2002). Since the discovery of the first micropeptide in plants, others have also been functionally characterized. The 36 AAs peptide, which is encoded by the POLARIS (PLS) gene in *Arabidopsis*, has been shown to affect root growth and leaf vascular patterning (Casson et al., 2002; Chilley et al., 2006). Another two micropeptides, 76 AAs Brk1 (Brk) and 53 AAs ROTUNDIFOLIA (ROT4), were also found to be involved with leaf morphogenesis. In maize, the recessive mutation of Brk1 results in several morphological defects of leaf epithelia (Frank and Smith, 2002). However, ROT4 regulates polar cell proliferation in lateral organs and leaf morphogenesis in *Arabidopsis* (Narita et al., 2004). In *Arabidopsis*, two other best-characterized micropeptides were reported: a 51 AAs ROT18/DLV1 and a 25 AAs kiss of death (KOD), which are involved in plant organogenesis (Wen et al., 2004; Valdivia et al., 2012; Guo et al., 2015) and programmed cell death regulation (Blanvillain et al., 2011), respectively. Recently two newer micropeptides have also been identified in maize, Zm401p10 and Zm908p11 with 89 and 97 AAs, respectively, which are involved in pollen development (Ma et al., 2008; Wang et al., 2009; Dong et al., 2013). Characterizations of these micropeptides indicate their functional diversity ranging from plant development to growth, nodulation, organogenesis, pollen development, and cell death.

In Animals

The first identification of micropeptides in animals came from the study of lncRNAs in *Drosophila*. The sORFs of the long noncoding RNA, namely, polished rice or tarsal-less (tal), encode four micropeptides from 11 to 32 AAs are required during the embryonic development of flies (Galindo et al., 2007; Kondo et al., 2007, 2010). By triggering proteasome-mediated protein processing, the *pri* micropeptide converts a transcription factor, shavenbaby (Svb), from a repressor into an activator (Zanet et al., 2015). Since then, a handful of micropeptides have been functionally characterized (Table 2). To identify the characterizing signal molecules from the nonannotated translated sORFs, the Pauli group identified a micropeptide, Toddler, which acts as a motogen, a signal that promotes cell migration. Toddler activates G-protein-coupled APJ (apelin) signaling for this function (Pauli et al.,

TABLE 2 | Micropeptides and their diverse biological functions.

Origin	Micro-peptides	Conservation	Method of identification/characterization	Function	Size (AAs)	References
Plant	Early nodulin 40 (Enod 40)	Plants	<i>In vitro</i> translation	Nodule organogenesis	12.24	Rohrig et al., 2002
	POLARIS (PLS)		Gene expression analysis by promoter trapping; Mutation analysis	Leaf morphogenesis	36	Casson et al., 2002; Chilley et al., 2006
	Brick1 (Brk)	Plants and animals	Mutation analysis	Leaf morphogenesis	76	Frank and Smith, 2002
	ROTUNDIFOLIA (ROT4)	Plants	Screening of a mutant in <i>Arabidopsis thaliana</i>	Leaf morphogenesis	53	Narita et al., 2004
	ROT18/DLV1	Plants	Gain of function screening of genes responsible for fruit growth and development in <i>Arabidopsis</i>	Plant organogenesis	51	Wen et al., 2004; Guo et al., 2015
	Kiss of death (KOD)		Gene expression analysis by promoter trapping	Programmed cell death regulation	25	Blarvillain et al., 2011
	Zm401p10, Zm908p11	Poaceae	Bioinformatics analysis	Pollen development	89.97	Ma et al., 2008; Wang et al., 2009; Dong et al., 2013
Animal	Polished rice (Pri)	Insects	Mutation analysis	Fly embryogenesis	11–32	Galindo et al., 2007; Kondo et al., 2007, 2010
	Toddler	Vertebrates	Ribo-seq-based search for novel signaling peptides	Promotes cell migration	58	Pauli et al., 2014
	AGD3	Mammals	Sequencing analysis.	Involve in stem cell differentiation	63	Kikuchi et al., 2009
	Myoregulin (MLN)	Mammals	Bioinformatics approaches; Homology-based characterization	Calcium homeostasis	46	Magny et al., 2013; Anderson et al., 2015
	DWOLF	Lamprey	PhyloCSF search; Gain and loss of function	Enhance muscle performance	34	Nelson et al., 2016
	Myomixer	Vertebrates	CRISPR-cas9 mediated loss of function screening of genes required for myoblast fusion	Functionally involve in controlling muscle performance	84	Bi et al., 2017
	MRI-2	Mammals	HPLC-MS/MS screening combining with RNA seq; characterized by functional proteomics	DNA repairing process	69	Slavoff et al., 2014
	NoBody	Mammals	HPLC-MS/MS screening combining with RNA seq; characterized by functional Proteomics	mRNA recycling	68	D'Lima et al., 2017
	SPAR	Human and mouse	Proteomics strategy	Regulate muscle regeneration	90	Matsumoto et al., 2017
	Humanin	Different species	Functional expression screening	Involve in program cell death	24	Hashimoto et al., 2001; Guo et al., 2003
	MOTS-c	14 species	<i>In silico</i> search for potential sORFs in human 12srRNA.	Metabolic Homeostasis	16	Lee et al., 2015
	Minion	Mammalian species	RNA seq analysis of uninjured and regenerating muscle	Muscle formation	84	Zhang et al., 2017
	HOXB-AS3	Primates	Ribosome profiling	Suppresses colon cancer growth	53	Huang et al., 2017

2014). AGD3, previously classified as a TUF, encodes a small protein of 63 AAs and has been found to show involvement in human stem cell differentiation (Kikuchi et al., 2009). Recently a group of micropeptides was found to show a prominent role in calcium homeostasis, both in skeletal and nonskeletal muscle cells, through the binding and inhibiting of a well-known Ca^{2+} ATP-ase pump, SERCA, thereby influencing regular muscle contraction (Magny et al., 2013; Anderson et al., 2015). Nelson et al. described the opposite activity of another lncRNA-derived micropeptide in mammalian muscle, called DWOLF (dwarf open reading frame). This

micropeptide enhances SERCA activity by displacing those inhibitory proteins and boosts muscle performance. DWOLF is abundantly expressed in the mouse heart, and is suppressed in ischemic human heart tissue, suggesting a possible link with heart failure (Nelson et al., 2016). Myomixer, a micropeptide of 84 AAs also has a function in the muscle but is unlike DWOLF or other micropeptides in this group. Myomixer plays a role in controlling muscle formation by associating with a fusogenic membrane protein, myomaker, and favors formation of multinucleated myofibers in mice (Bi et al., 2017). Recently, another peptide known as minion (microprotein inducer of

fusion), which is specific for skeletal muscle, has been identified. Functional characterization of this microprotein revealed that like myomixer, minion also controls cell fusion, and muscle formation by associating with myomaker (Zhang et al., 2017). The functionality of micropeptides has also been found in the DNA repairing process. For example, a 69 AAs small peptide, MRI-2, has been identified as a novel factor of the non-homologous end join factor (NHEJ). MRI-2 stimulates NHEJ by interacting with Ku protein, a DNA end-binding protein (Slavoff et al., 2014). As more micropeptides are characterized, more hidden functions are unfolded, as exemplified by another micropeptide that is encoded by a putative lncRNA HOXB-AS3. This conserved 53 AAs peptide, HOXB-AS3, inhibits tumorigenesis by the regulation of PKM alternative splicing and metabolic reprogramming of colon cancer cells (Huang et al., 2017). NoBody and SPAR are two additional examples of functional micropeptides, which as we described above, have been characterized recently by their distinct biological significance.

According to Weissman, some micropeptides might also be immunogenic without a clear functional role. For example, micropeptides derived from human-infecting cytomegalovirus (HCMV) lncRNA β 2.7, were found to robustly stimulate T cell memory responses only in humans with a history of HCMV infection (Fields et al., 2015). Very recently, another group of scientists identified some micropeptides that exhibited differential regulation upon viral infection (Razooky et al., 2017). These indicate that there may be more sORFs that are involved with certain diseases. Thus, translation of some ORFs that have been previously overlooked may contribute in important ways to cell biology.

Biologically significant micropeptides are not only found to be encoded by nuclear-encoded transcripts. Mitochondrial genomes also contribute in the proteome by producing biologically important micropeptides. Humanin, a signaling peptide encoded by mitochondrial sORFs, is functionally involved with programmed cell death. It inhibits translocation of an apoptosis-inducing protein, Bax (Bcl2-associated x-protein), from cytoplasm to mitochondria, and thereby regulates apoptosis (Guo et al., 2003). Humanin also shows neuroprotective effects and is known as a peptide against neurotoxicity related diseases (Matsuoka et al., 2006). Another micropeptide of 16 AAs was also found to be encoded by mitochondrial 12sRNA, named MOTS-c. MOTS-c shows endocrine-like effects on muscle metabolism, insulin sensitivity and weight regulation (Lee et al., 2015). Identification of the mitochondrial-encoded peptides humanin and MOTS-c suggests the possible existence of more potent sORFs in mitochondria along with their role as regulators of biological processes.

The diverse biological functions of these micropeptides serve as an indication that we are at the very beginning of exploring the mystery of micropeptides.

CONCLUSIONS

Technological advances have uncovered the existence of several hundred putative sORF-encoded micropeptides throughout the genomes. Recent identification and characterization of a small number of sORF-encoded micropeptides and their biological role indicate that there is a hidden world of active peptides waiting to be explored. A great deal of effort is still needed to validate whether each of these peptides is biologically important or if they are just transcriptional/translational noise. Some widely used approaches, such as homology-based functionality search, functional proteomics, gene editing technologies, and massive sequencing-based approach, can be implemented on uncharacterized micropeptides to reveal their biological relevance. Tiny size, low abundance, rapid degradation and loss during sample preparation often make it difficult to work with micropeptides, demanding more sensitive and sophisticated methods. Thus, there are many technical challenges in facilitating the study of micropeptides.

Functional studies of micropeptides in a wide range of species demonstrate that they have important biological functions, including involvement in human pathogenesis. HOXB-AS3, DWORF and humanin are some examples of this group, which show involvement in cancer, heart diseases, and neurotoxicity related diseases, respectively. In addition to these, involvement of a group of newly identified micropeptides against viral infection mediated pathogenesis also suggest that there are more micropeptides that may be involved with certain diseases in humans. These findings indicate that micropeptides may represent new opportunities for drug therapies.

Although some of the micropeptides are functionally characterized, the exact mechanism of their mode of action is unclear. Complete understanding of their action may play an important role in therapeutic purposes, where a drug may be designed by modulating or mimicking their function to regulate any biological pathway they may be involved in.

These recent findings provide new insights into sORF-encoded micropeptides as a new and important class of biological molecules and offer new avenues of research in the proteomics world.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- 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
- Anderson, D. M., Makarewich, C. A., Anderson, K. M., Shelton, J. M., Bezprozvannaya, S., Bassel-Duby, R., et al. (2016). Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* 9:ra119. doi: 10.1126/scisignal.aaj1460
- Aspden, J. L., Eyre-Walker, Y. C., Philips, R. J., Amin, U., Mumtaz, M. A. S., Brocard, M., et al. (2014). Extensive translation of small open

- reading frames revealed by Poly-Ribo-Seq. *Elife* 3:e03528. doi: 10.7554/eLife.03528
- Bánfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W. E., et al. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 22, 1646–1657. doi: 10.1101/gr.134767.111
- 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
- Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J. R., Shelton, J. M., et al. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. *Science* 356, 323–327. doi: 10.1126/science.aam9361
- Birney, E., Stamatioupanopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874
- Blanvillain, R., Young, B., Cai, Y. M., Hecht, V., Varoquaux, F., Delorme, V., et al. (2011). The Arabidopsis peptide kiss of death is an inducer of programmed cell death. *EMBO J.* 30, 1173–1183. doi: 10.1038/emboj.2011.14
- Boonen, K., Creemers, J. W., and Schoofs, L. (2009). Bioactive peptides, networks and systems biology. *BioEssays* 31, 300–314. doi: 10.1002/bies.200800055
- Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* 10, 28–36. doi: 10.1128/MCB.10.1.28
- Cabrera-Quio, L. E., Herberg, S., and Pauli, A. (2016). Decoding sORF translation - from small proteins to gene regulation. *RNA Biol.* 13, 1051–1059. doi: 10.1080/15476286.2016.1218589
- Calviello, L., Mukherjee, N., Wyler, E., Zaubner, H., Hirsekorn, A., Selbach, M., et al. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* 13, 165–170. doi: 10.1038/nmeth.3688
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563. doi: 10.1126/science.1112014
- Carvunis, A. R., Rolland, T., Wapinski, I., Calderwood, M. A., Yildirim, M. A., Simonis, N., et al. (2012). Proto-genes and *de novo* gene birth. *Nature* 487, 370–374. doi: 10.1038/nature11184
- Casson, S. A., Chilley, P. M., Topping, J. F., Evans, I. M., Souter, M. A., and Lindsey, K. (2002). The POLARIS gene of Arabidopsis encodes a predicted peptide required for correct root growth and leaf vascular patterning. *Plant Cell* 14, 1705–1721. doi: 10.1105/tpc.002618
- Chilley, P. M., Casson, S. A., Tarkowski, P., Hawkins, N., Wang, K. L. C., Hussey, P. J., et al. (2006). The POLARIS peptide of Arabidopsis regulates auxin transport and root growth via effects on ethylene signaling. *Plant Cell* 18, 3058–3072. doi: 10.1105/tpc.106.040790
- Chu, Q., Rathore, A., Diedrich, J. K., Donaldson, C. J., Yates, J. R. III., and Saghatelian, A. (2017). Identification of microprotein-protein interactions via apex tagging. *Biochemistry* 56, 3299–3306. doi: 10.1021/acs.biochem.7b00265
- Cohen, S. M. (2014). Everything old is new again: (linc) RNAs make proteins! *EMBO J.* 33, 937–938. doi: 10.1002/embj.201488303
- Crappé, J., Van Criekinge, W., and Menschaert, G. (2014). Little things make big things happen: a summary of micropeptide encoding genes. *EuPA Open Proteomics* 3, 128–137. doi: 10.1016/j.euprot.2014.02.006
- Cunha, F. M., Berti, D. A., Ferreira, Z. S., Klitzke, C. F., Markus, R. P., and Ferro, E. S. (2008). Intracellular peptides as natural regulators of cell signaling. *J. Biol. Chem.* 283, 24448–24445. doi: 10.1074/jbc.M801252200
- D’Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., et al. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* 13, 174–180. doi: 10.1038/nchembio.2249
- Dong, X., Wang, D., Liu, P., Li, C., Zhao, Q., Zhu, D., et al. (2013). Zm908p11, encoded by a short open reading frame (sORF) gene, functions in pollen tube growth as a profilin ligand in maize. *J. Exp. Bot.* 64, 2359–2372. doi: 10.1093/jxb/ert093
- Fields, A. P., Rodriguez, E. H., Jovanovic, M., Stern-Ginossar, N., Haas, B. J., Mertins, P., et al. (2015). A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Mol. Cell* 60, 816–827. doi: 10.1016/j.molcel.2015.11.013
- Frank, M. J., and Smith, L. G. (2002). A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells. *Curr. Biol.* 12, 849–853. doi: 10.1016/S0960-9822(02)00819-9
- Fricker, L. D. (2005). Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J.* 7, E449–E455. doi: 10.1208/aapsj070244
- Frith, M. C., Forrest, A. R., Nourbakhsh, E., Pang, K. C., Kai, C., Kawai, J., et al. (2006). The abundance of short proteins in the mammalian proteome. *PLoS Genet.* 2:e52. doi: 10.1371/journal.pgen.0020052
- Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A., and Couso, J. P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5:e106. doi: 10.1371/journal.pbio.0050106
- Gish, W., and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* 3, 266–272. doi: 10.1038/ng0393-266
- Grundly, G. J., Rulten, S. L., Arribas-Bosacoma, R., Davidson, K., Kozik, Z., Oliver, A. W., et al. (2016). The Ku-binding motif is a conserved module for recruitment and stimulation of non-homologous end-joining proteins. *Nat. Commun.* 7:11242. doi: 10.1038/ncomms11242
- Guo, B., Zhai, D., Cabezas, E., Welsh, K., Nouraini, S., Satterthwait, A. C., et al. (2003). Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature* 423, 456–461. doi: 10.1038/nature01627
- Guo, P., Yoshimura, A., Ishikawa, N., Yamaguchi, T., Guo, Y., and Tsukaya, H. (2015). Comparative analysis of the RTFL peptide family on the control of plant organogenesis. *J. Plant Res.* 128, 497–510. doi: 10.1007/s10265-015-0703-1
- Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251. doi: 10.1016/j.cell.2013.06.009
- Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K., and Shiu, S. H. (2010). sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics* 26, 399–400. doi: 10.1093/bioinformatics/btp688
- Hanada, K., Higuchi-Takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., et al. (2013). Small open reading frames associated with morphogenesis are hidden in plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2395–2400. doi: 10.1073/pnas.1213958110
- Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., et al. (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer’s disease genes and Abeta. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6336–6341. doi: 10.1073/pnas.101133498
- Housman, G., and Ulitsky, I. (2016). Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochim. Biophys. Acta* 1859, 31–40. doi: 10.1016/j.bbagr.2015.07.017
- 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
- Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., et al. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379. doi: 10.1016/j.celrep.2014.07.045
- Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi: 10.1016/j.cell.2011.10.002
- Kapranov, P., Willingham, A. T., and Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 8, 413–423. doi: 10.1038/nrg2083
- Kikuchi, K., Fukuda, M., Ito, T., Inoue, M., Yokoi, T., Chiku, S., et al. (2009). Transcripts of unknown function in multiple-signaling pathways involved in human stem cell differentiation. *Nucleic Acids Res.* 37, 4987–5000. doi: 10.1093/nar/gkp426
- Kochetov, A. V. (2005). AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. *Bioinformatics* 21, 837–840. doi: 10.1093/bioinformatics/bti136
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* 9, 660–665. doi: 10.1038/ncb1595
- Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., et al. (2010). Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila embryogenesis*. *Science* 329, 336–339. doi: 10.1126/science.1188158

- Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A., and Couso, J. P. (2011). Hundreds of putatively functional small open reading frames in *Drosophila*. *Genome Biol.* 12:R118. doi: 10.1186/gb-2011-12-11-r118
- Lauressergues, D., Couzigou, J. M., San Clemente, H., Martinez, Y., Dunand, C., Bécard, G., et al. (2015). Primary transcripts of microRNAs encode regulatory peptides. *Nature* 520, 90–93. doi: 10.1038/nature14346
- Lee, C., Zeng, J., Drew, B. G., Sallam, T., Martin-Montalvo, A., Wan, J., et al. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab.* 21, 443–454. doi: 10.1016/j.cmet.2015.02.009
- Lin, M. F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27, i275–i282. doi: 10.1093/bioinformatics/btr209
- Ma, J., Ward, C. C., Jungreis, I., Slavoff, S. A., Schwaid, A. G., Neveu, J., et al. (2014). Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J. Proteome Res.* 13, 1757–1765. doi: 10.1021/pr401280w
- Ma, J., Yan, B., Qu, Y., Qin, F., Yang, Y., Hao, X., et al. (2008). Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J. Cell. Biochem.* 105, 136–146. doi: 10.1002/jcb.21807
- Mackowiak, S. D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., et al. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.* 16:179. doi: 10.1186/s13059-015-0742-x
- Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., et al. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 341, 1116–1120. doi: 10.1126/science.1238802
- Makarewicz, C. A., and Olson, E. N. (2017). Mining for micropeptides. *Trends Cell Biol.* 27, 685–696. doi: 10.1016/j.tcb.2017.04.006
- Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., et al. (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* 541, 228–232. doi: 10.1038/nature21034
- Matsuoka, M., Hashimoto, Y., Aiso, S., and Nishimoto, I. (2006). Humanin and colivelin: neuronal-death-suppressing peptides for Alzheimer's disease and amyotrophic lateral sclerosis. *CNS Drug Rev.* 12, 113–122. doi: 10.1111/j.1527-3458.2006.00113.x
- Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., et al. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana*. *Plant J.* 38, 699–713. doi: 10.1111/j.1365-3113X.2004.02078.x
- Nelson, B. R., Makarewicz, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351, 271–275. doi: 10.1126/science.aad4076
- Pauli, A., Norris, M. L., Valen, E., Chew, G. L., Gagnon, J. A., Zimmerman, S., et al. (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* 343:1248636. doi: 10.1126/science.1248636
- Pauli, A., Valen, E., and Schier, A. F. (2015). Identifying (non-)coding RNAs and small peptides: challenges and opportunities. *Bioessays* 37, 103–112. doi: 10.1002/bies.201400103
- Razooky, B. S., Obermayer, B., O'May, J. B., and Tarakhovsky, A. (2017). Viral infection identifies micropeptides differentially regulated in smORF-containing lncRNAs. *Genes* 8:206. doi: 10.3390/genes8080206
- Rhee, H. W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., et al. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* 339, 1328–1331. doi: 10.1126/science.1230593
- Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1915–1920. doi: 10.1073/pnas.022664799
- Saghatelian, A., and Couso, J. P. (2015). Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* 11:909. doi: 10.1038/nchembio.1964
- Savard, J., Marques-Souza, H., Aranda, M., and Tautz, D. (2006). A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* 126, 559–569. doi: 10.1016/j.cell.2006.05.053
- Schwaid, A. G., Shannon, D. A., Ma, J., Slavoff, S. A., Levin, J. Z., Weerapana, E., et al. (2013). Chemoproteomic discovery of cysteine-containing human short open reading frames. *J. Am. Chem. Soc.* 135, 16750–16753. doi: 10.1021/ja406606j
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050. doi: 10.1101/gr.3715005
- Skarszewski, A., Stanton-Cook, M., Huber, T., Al Mansoori, S., Smith, R., Beatson, S. A., et al. (2014). uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation. *BMC Bioinformatics* 15:36. doi: 10.1186/1471-2105-15-36
- Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., and Saghatelian, A. (2014). A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *J. Biol. Chem.* 289, 10950–10957. doi: 10.1074/jbc.C113.533968
- 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
- Smith, J. E., Alvarez-Dominguez, J. R., Kline, N., Huynh, N. J., Geisler, S., Hu, W., et al. (2014). Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Rep.* 7, 1858–1866. doi: 10.1016/j.celrep.2014.05.023
- Valdivia, E. R., Chevalier, D., Sampedro, J., Taylor, I., Niederhuth, C. E., and Walker, J. C. (2012). DVL genes play a role in the coordination of socket cell recruitment and differentiation. *J. Exp. Bot.* 63, 1405–1412. doi: 10.1093/jxb/err378
- Vanderperre, B., Lucier, J. F., and Roucou, X. (2012). HALtORF: a database of predicted out-of-frame alternative open reading frames in human. *Database* 2012:bas025. doi: 10.1093/database/bas025
- Wang, D., Li, C., Zhao, Q., Zhao, L., Wang, M., Zhu, D., et al. (2009). Zm401p10, encoded by an anther-specific gene with short open reading frames, is essential for tapetum degeneration and anther development in maize. *Funct. Plant Biol.* 36, 73–85. doi: 10.1071/FP08154
- Wen, J., Lease, K. A., and Walker, J. C. (2004). DVL, a novel class of small polypeptides: overexpression alters *Arabidopsis* development. *Plant J.* 37, 668–677. doi: 10.1111/j.1365-3113X.2003.01994.x
- 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
- Zanet, J., Benrabah, E., Li, T., Pelissier-Monier, A., Chanut-Delalande, H., Ronsin, B., et al. (2015). Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* 349, 1356–1358. doi: 10.1126/science.aac5677
- Zhang, Q., Vashisht, A. A., O'Rourke, J., Corbel, S. Y., Moran, R., Romero, A., et al. (2017). The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* 8:15664. doi: 10.1038/ncomms15664

Conflict of Interest Statement: 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.

Copyright © 2018 Yeasmin, Yada and Akimitsu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of Minimal *p53* Promoter Region Regulated by MALAT1 in Human Lung Adenocarcinoma Cells

Keiko Tano¹, Rena Onoguchi-Mizutani¹, Fouzia Yeasmin¹, Fumiaki Uchiumi², Yutaka Suzuki³, Tetsushi Yada⁴ and Nobuyoshi Akimitsu^{1*}

¹ Isotope Science Center, The University of Tokyo, Tokyo, Japan, ² Department of Gene Regulation, Faculty of Pharmaceutical Sciences, Tokyo University of Science, Noda-shi, Chiba-ken, Japan, ³ Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, ⁴ Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Kitakyushu, Japan

OPEN ACCESS

Edited by:

Kinji Ohno,
Nagoya University, Japan

Reviewed by:

Tohru Yoshihisa,
University of Hyogo, Japan
Akio Masuda,
Nagoya University Graduate School of
Medicine, Japan

*Correspondence:

Nobuyoshi Akimitsu
akimitsu@ric.u-tokyo.ac.jp

Specialty section:

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

Received: 03 August 2017

Accepted: 27 November 2017

Published: 26 March 2018

Citation:

Tano K, Onoguchi-Mizutani R, Yeasmin F, Uchiumi F, Suzuki Y, Yada T and Akimitsu N (2018) Identification of Minimal *p53* Promoter Region Regulated by MALAT1 in Human Lung Adenocarcinoma Cells. *Front. Genet.* 8:208. doi: 10.3389/fgene.2017.00208

The MALAT1 long noncoding RNA is strongly linked to cancer progression. Here we report a MALAT1 function in repressing the promoter of *p53* (*TP53*) tumor suppressor gene. *p21* and *FAS*, well-known *p53* targets, were upregulated by MALAT1 knockdown in A549 human lung adenocarcinoma cells. We found that these upregulations were mediated by transcriptional activation of *p53* through MALAT1 depletion. In addition, we identified a minimal MALAT1-responsive region in the P1 promoter of *p53* gene. Flow cytometry analysis revealed that MALAT1-depleted cells exhibited G1 cell cycle arrest. These results suggest that MALAT1 affects the expression of *p53* target genes through repressing *p53* promoter activity, leading to influence the cell cycle progression.

Keywords: noncoding RNA, TP53, transcription, genetic, adenocarcinoma, promoter regions, genetic

INTRODUCTION

Recent large-scale transcriptome analyses have revealed that transcription is spread throughout the mammalian genome (>90%), including in noncoding regions. This yields large numbers of noncoding RNAs (ncRNAs) (Birney et al., 2007), including small ncRNAs (20–200 nt), and long ncRNAs (lncRNA) (>200 nt). Increasing numbers of studies have revealed the unique features and functions of lncRNAs in broad biological processes as well as development of diseases (Mercer et al., 2009). Considering their critical roles in cellular and molecular mechanisms, we cannot exclude the involvement of lncRNAs from molecular analyses of biological processes and disease pathogenesis.

Recent studies have reported several lncRNAs exhibiting cancer-specific expression patterns (Tano and Akimitsu, 2012). Metastasis associated lung adenocarcinoma transcript 1 (MALAT1) is a well-known cancer-related lncRNA over 8,000 nt in length. MALAT1 was originally identified in early-stage nonsmall cell lung cancer (NSCLC) with a high propensity for metastasis (Ji et al., 2003). Although Malat1 is not essential in living mice maintained under normal breeding conditions (Nakagawa et al., 2012; Zhang et al., 2012), MALAT1 has an important role in the study of human cancer: MALAT1 is overexpressed in many solid tumors (Lin et al., 2007; Gutschner et al., 2013a). In addition, high expression levels of MALAT1 correlated with poor prognosis in patients with NSCLC (Ji et al., 2003). MALAT1 contributes in metastasis phenotype of human lung cancer cells (Tano et al., 2010; Gutschner et al., 2013b). These results highlight that revealing the function of MALAT1 is important to understand the cancer biology.

MALAT1 functions in regulating gene expression in several capacities. Intracellularly, MALAT1 is stably retained in the nucleus (Miyagawa et al., 2010; Tani et al., 2012). MALAT1 specifically localizes to nuclear speckles, which are subnuclear structures enriched for pre-mRNA splicing factors (Hutchinson et al., 2007), and regulates alternative splicing by modulating the distribution and levels of active pre-mRNA splicing factors (SR proteins) in nuclear speckles (Tripathi et al., 2010). Another report demonstrated MALAT1 involvement in regulated transcriptional programs (Yang et al., 2011). MALAT1 was shown to regulate the relocation of growth control genes from the repressive environment of polycomb bodies (PcGs) to the gene activation milieu of interchromatin granules (ICGs) in response to growth signals by interacting with unmethylated Pc2. This leads to the promotion of E2F1 sumoylation and activation of transcription of genes associated with growth control. In addition, identification of genomic region interacted with MALAT1 revealed that MALAT1 associates with transcriptionally active genes (Engreitz et al., 2014; West et al., 2014). These results highlight the distinct role of MALAT1 in the regulation of gene expression program.

We previously demonstrated a link between characteristics of cancer metastasis and gene regulation by MALAT1 (Tano et al., 2010). We showed that MALAT1 promotes cell migration, which is one of the most important features for metastasis, through regulation of several migration-related genes at the transcriptional and/or post-transcriptional level (Tano et al., 2010). We also performed microarray analysis of MALAT1-knockdown cells and found that MALAT1 was involved in the repression of several genes associated with tumor suppression.

In this study, we re-examined the previous microarray data and found that several MALAT1-regulated genes were *p53* (*TP53*) target genes, such as *p21* (*CDKN1A*) and *FAS*, suggesting the possibility that MALAT1 is involved in the repression of these genes through *p53*. We showed that upregulation of both *p21* and *FAS* in MALAT1-depleted A549 lung adenocarcinoma cells was repressed by knockdown of *p53* and inhibition of *p53* activity by PFT- α . Further, we found that depletion of MALAT1 leads to upregulation of *p53* through activation of *p53* promoter. We identified -153 to -111 of the P1 *p53* promoter as a MALAT1-responsive region. This is the first report showing that MALAT1 affects expression of *p53* target genes through negative regulation of specific elements in the *p53* promoter. Finally, we showed that depletion of MALAT1 resulted in cell cycle arrest in G1. Together our results indicate that MALAT1 may have additional functions in repressing tumor suppression to promote cancer progression.

MATERIALS AND METHODS

Cell Culture and RNA Interference

A549 and H1299 (kindly supplied by Dr. Hideki Matsumoto, Fukui University, Japan) cells were grown at 37°C with 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM) or RPMI 1640 medium, respectively, supplemented with 10% fetal bovine serum and penicillin/streptomycin.

RNA interference was performed using Lipofectamine RNAiMAX (Invitrogen, Tokyo, Japan), according to the

manufacturer's instructions. The siRNA sequences were as follows (sense/antisense): *p53* (siRNA 1), 5'-GTGAGCGCTTCGAGATGTTCC-3'/5'-AACATCTCGAAGCGCTCACGC-3'; and *p53* (siRNA 2), 5'-gacTccagTggTaaTcTcTT-3'/5'-gTagaTTaccacTggagTcTT-3'. The MALAT1 siRNA and the negative control siRNA sequences were described previously (Tano et al., 2010). Efficient reduction of each gene by siRNA was confirmed by quantitative real-time PCR analysis.

Plasmid Constructs

The human *p53* promoter pGL2 (Basic) luciferase plasmids containing *p53* promoter fragments (356, 200, and 100 bp) were purchased from Addgene (MA, USA). For the construction of 5'-end deletion mutant reporter plasmids, each fragment was amplified by PCR and cloned into the pGL2 basic reporter vector. The primers used for PCR cloning were as follows: pGL2-177bp-F-SacI, 5'-cagaccGAGCTCctctcccaactccatttc-3'; pGL2-165bp-F-SacI, 5'-cagaccGAGCTCctcatttcttcttctc-3'; pGL2-148bp-F-SacI, 5'-cagaccGAGCTCctccggcaggcgattac-3'; pGL2-140bp-F-SacI, 5'-cagaccGAGCTCggcgattacttgccctt-3'; pGL2-130bp-F-SacI, 5'-cagaccGAGCTCttgccttactgtcatggcg-3'; pGL2-122bp-F-SacI, 5'-cagaccGAGCTCactgtcatggcgactgtcc-3'; pGL2-110bp-F-SacI, 5'-cagaccGAGCTCgactgtccagcttctgtgcag-3'; and pGL2-Re-HindIII, 5'-aatcccAAGCTTctagacttttgagaa gctcaaaacttttag-3'. The pGL2-200bp *p53* promoter plasmid was used as a template.

For the construction of deletion mutant plasmids, in which a part of the MALAT1 response element was deleted, we performed site-directed mutagenesis using primers as follows: *p53pro-del1*-F, 5'-GCTTCCTCCGGCAGGCGG-3'; *p53pro-del1*-Re, 5'-AAA TGGAGTTGGGGAGGAGGGTGC-3'; *p53pro-del2*-F, 5'-GGA TTACTTGCCCTTACTTGTCATG-3'; *p53pro-del2*-Re, 5'-CCG GAGGAAGCAAAGGAAATG-3'; *p53pro-del3*-F, 5'-CCTTACT TGTCATGGCGACTG-3'; and *p53pro-del3*-Re, 5'-TAATCCGC CTGCCGGAGG-3'.

Reverse Transcription and Quantitative Real-Time PCR Analysis

Total RNA was prepared using the RNAiso Plus kit (Takara Bio, Shiga, Japan), and 500 ng of RNA was reverse transcribed to produce cDNA with the PrimeScript RT Master Mix (Takara Bio). Real-time PCR was carried out with the Thermal Cycler Dice using the SYBR Premix Ex Taq II (Takara Bio, Shiga, Japan). The sequences of primer sets used in this analysis were shown in Table 1.

RT-PCR

Total RNA was extracted using RNAiso Plus (Takara Bio), and 500 ng of RNA was reverse transcribed to produce cDNA with the PrimeScript RT Master Mix (Takara Bio). PCR was performed on 1/10 (2 μ l) of the cDNA using Ex Taq HS polymerase (Takara Bio). PCR conditions were 30 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 1 min. PCR products were separated on 2% agarose gels.

Western Blot Analysis

Cells were lysed with RIPA buffer containing 125 mM Tris-HCl (pH 8.0), 375 mM sodium chloride, 2.5 mM EDTA, 0.25% sodium dodecyl sulfate, 2.5% Triton X-100, 0.25% sodium deoxycholate, 1 mM PMSF, 5 µg/ml Leupeptin, and 1 µl Aprotinin Solution

(Wako, Osaka, Japan), and centrifuged at 15,000 rpm to remove debris. Protein extracts (10 µg) were resolved by sodium dodecyl sulfate 10% polyacrylamide gel electrophoresis and transferred to a PVDF membrane. After blocking with 3% BSA in TBST, the blot was probed with the DO-1 monoclonal anti-human p53 antibody (Medical and Biological Laboratories, Aichi, Japan), followed by peroxidase labeled anti-mouse antibody (no. NIF 825, GE Healthcare). For α-tubulin detection, rabbit anti-α-tubulin antibody (Medical and Biological Laboratories) and peroxidase labeled anti-rabbit antibody (no. NA934VS; GE Healthcare) were used. The horseradish peroxidase-labeled antibodies were detected by Immobilon Western Chemiluminescent HRP Substrate (Millipore) using a Lumino image analyzer, LAS4000 (Fujifilm).

TABLE 1 | Primers for p53 target genes.

Gene	Forward	Reverse
p21	tcactgtctgtaccctgtgc	ggcgcttgagggtgtagaaa
FAS	gtggaccgcgtcagtaag	tctagcaacagacgtagaacca
DNAJC15,	agattcagatagagtccagcat	tggatgaaccaagatttgt
p53	cgtgtggagatttggatgac	ttgtagtggatgggtgacagtc
pre-mRNA of p53	gtggaaggaaattgcgtgt	accacccttaaccctcct
BTG2	gcgagcagaggcttaaggt	gggaaaccagtgggtttgta
SULF2	aagcacggctccgactact	gtcggaagaagctcacg
HMOX1	gggtgataagaggccaaga	agctcctgcaactcctcaaa
RPS27L	tctatcccggaagttgatgc	caagctcagcctaccagac
ACTA2	ccctgaagtaccgataagaaca	ggcaacacgaagctcattg
FDXR	gtgacacagccgtgattctg	cttctgtgatgccgtctctc
GAPDH	gcaccgtcaaggctgagaac	tggtgaagacgccagtga

Massive Transcriptional Start Site (TSS) Analysis

TSS analysis was performed as described in Tani et al. (2012). Briefly, cells transfected with control or MALAT1 siRNA were harvested and RNA was extracted using RNAiso Plus. Thirty micrograms of the obtained total RNA was subjected to oligo-capping by treatment of BAP, TAP, and RNA oligo ligation.

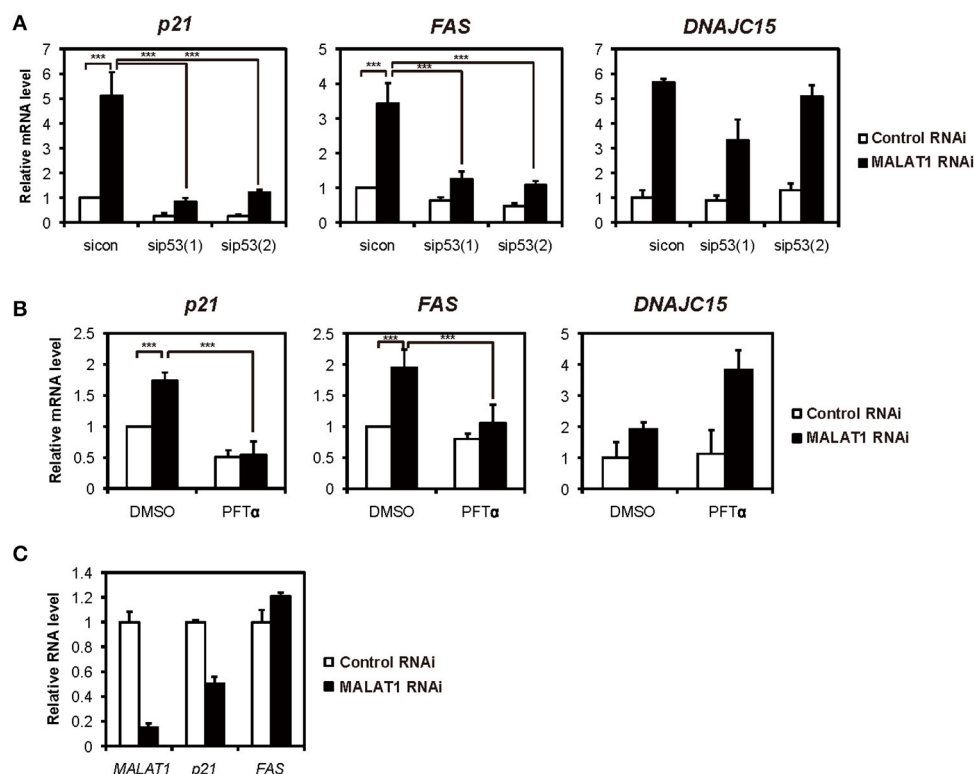


FIGURE 1 | Increased expression levels of both *p21* and *FAS* mediated by p53 in MALAT1-knockdown cells. **(A,B)** Real-time PCR analyses were performed to assess the effects of knockdown of p53 by p53 siRNA 1 (sip53(1)) and p53 siRNA 2 (sip53(2)) **(A)** or inhibition of p53 by PFT-α in the presence of doxorubicin **(B)** on the expression of *p21*, *FAS* and *DNAJC15* genes in MALAT1-knockdown cells. Relative mRNA expression levels are presented as ratios to the level of that in control cells. Data are presented as means ± standard deviation (SD) of three independent experiments (****P* < 0.001, two-way ANOVA followed by Bonferroni *t*-test as post-hoc test), except for *DNAJC15*, which shows the average of duplicate experiments. **(C)** p53-null H1299 cells were transfected with MALAT1 siRNA, and *p21* and *FAS* expression levels were analyzed by real-time PCR. Values represent the means ± SD of duplicate measurements.

After DNase I treatment, polyA-containing RNA was selected by oligo-dT powder. First strand cDNA was synthesized from random hexamers and amplified with 15 cycles of PCR using Gene Amp PCR kits (Perkin Elmer). The PCR fragments were size fractionated by 12% polyacrylamide gel electrophoresis and the 150–250 bp fraction was recovered. The quality and quantity of the obtained single-stranded first strand cDNAs were assessed by BioAnalyzer (Agilent).

One nanogram of the size-fractionated cDNA was used for the massively paralleled sequencing by an Illumina GA Sequencer. Clusters (15,000–20,000) were generated per tile and 36 cycles of the sequencing reactions were performed according to the manufacturer's instructions. The 36-base long tags corresponding to the 5'-ends of transcripts were generated by the sequencer. The obtained sequences were mapped onto human genomic sequences (hg18 of UCSC

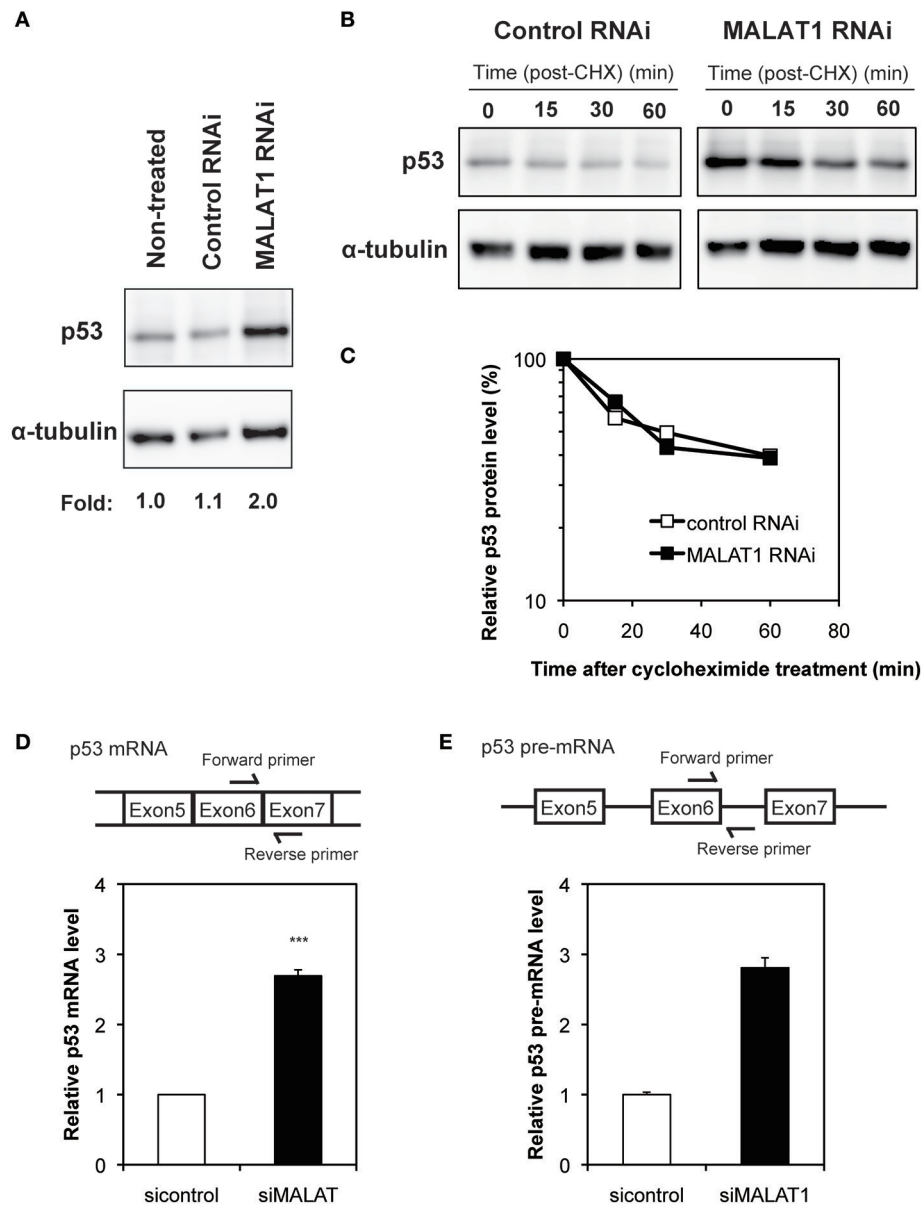


FIGURE 2 | Increased expression level of p53 in MALAT1-knockdown cells. **(A)** Expression levels of p53 and α -tubulin were determined by Western blotting. **(B)** Determination of p53 protein stability in MALAT1-knockdown cells. Cells were treated with cycloheximide (CHX) (100 μ M) to inhibit further protein synthesis and harvested at the indicated times. Remained Proteins of p53 and α -tubulin were measured by western blot analysis. The signal intensities of p53 bands were normalized by those of α -tubulin bands. **(C)** Quantitative analysis of p53 protein levels in **(B)** using a lumino image analyzer. **(D)** Expression levels of p53 mRNA were analyzed by real-time RT-PCR. Data are presented as means \pm SD of three independent experiments (*** P < 0.001, Student's t -test). A schematic diagram to detect matured p53 mRNAs is shown at upper area. **(E)** Pre-mRNA level of p53 was examined by real-time RT-PCR. Values represent the means \pm SD of duplicate measurements. A schematic diagram to detect pre-matured p53 mRNAs is shown at upper area.

Genome Browser) using the sequence alignment program Eland.

Luciferase Assays

Luciferase assays were performed using the dual-luciferase reporter assay system (Promega). Cells were co-transfected with the *p53* promoter reporter plasmids containing firefly luciferase and internal control reporter plasmids containing Renilla luciferase using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions. At 24 h after transfection, cells were harvested and luciferase activity was measured following the manufacturer's protocol.

RESULTS AND DISCUSSION

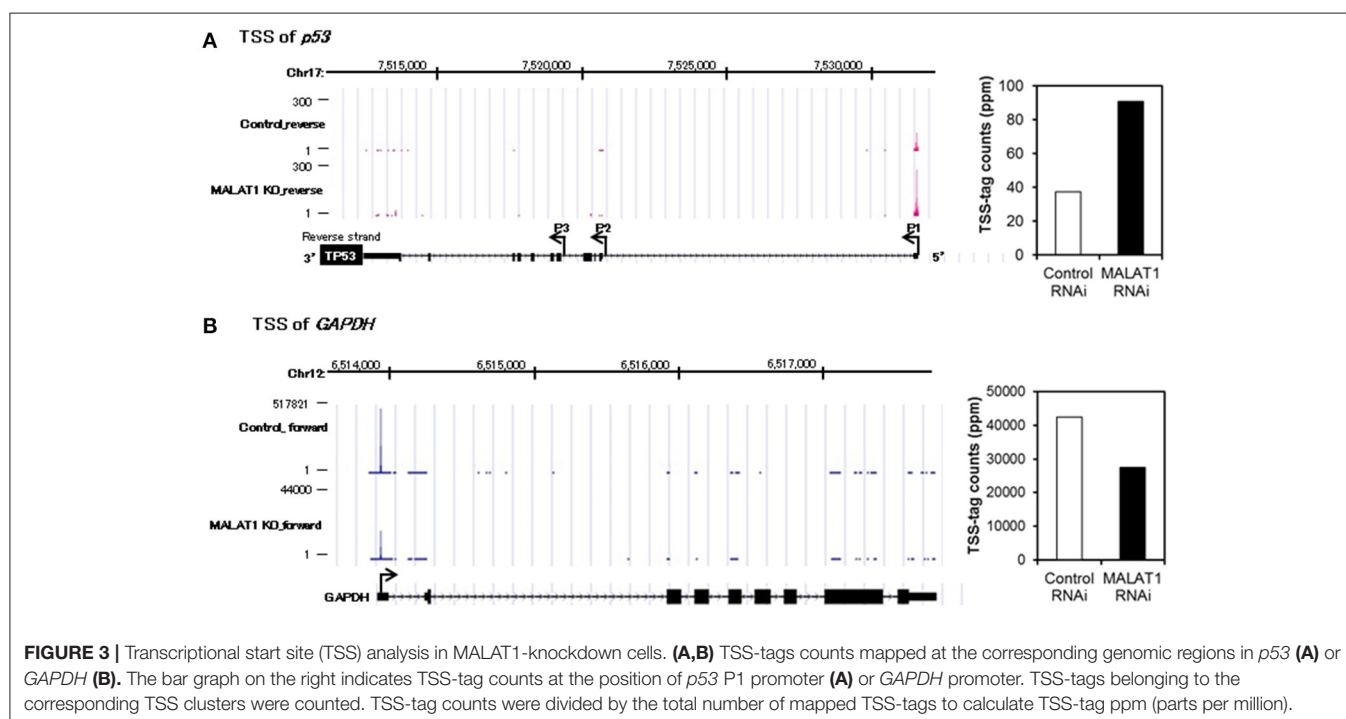
Upregulation of Both *p21* and *FAS* in MALAT1-Knockdown A549 Cells Was Mediated by *p53*

Previously, we showed that several *p53* target genes, including *p21* and *FAS*, were upregulated by knockdown of MALAT1 in A549 cells, in which *p53* is intact (Tano et al., 2010). This prompted the hypothesis that MALAT1 represses the expression of *p21* and *FAS* genes through *p53* activity. To test this hypothesis, we examined whether the *p53* target genes were upregulated through *p53* activity in MALAT1-knockdown cells. First, we confirmed the upregulation of *p53* target genes upon MALAT1 knockdown (Figure 1A and Figure S3). We then found that upregulation of *p21* and *FAS* in MALAT1-knockdown cells was repressed by siRNA-mediated *p53* depletion (Figure 1A). In contrast, upregulation of *DNAJC15*, which is not a *p53*

target gene, was not repressed by *p53* depletion in MALAT1-knockdown cells. Generally, knockdown efficiency is not 100%; therefore, we still detected some upregulation of *p21* and *FAS* mRNAs upon MALAT1 knockdown even in the *p53*-knockdown cells. To further investigate whether *p53* is involved in the upregulation of *p21* and *FAS* mRNAs, we examined Pifithrin- α (PFT α), a specific inhibitor of *p53*, on the upregulation of *p21* and *FAS* in MALAT1-knockdown cells. The increased expression levels of *p21* and *FAS* in MALAT1-knockdown cells were inhibited by PFT- α , and this inhibitory effect was not observed with *DNAJC15* (Figure 1B). Furthermore, we found that MALAT1-knockdown-mediated upregulation of both *p21* and *FAS* was not observed in *p53*-null H1299 cells (Figure 1C). These results suggest that upregulation of *p21* and *FAS* in MALAT1-knockdown cells was mediated by *p53* activity.

To investigate the mechanism by which MALAT1 affects *p53* activity, we next examined the expression levels of *p53* in MALAT1-knockdown cells. Western blot analysis revealed that the *p53* protein levels in MALAT1-knockdown cells were increased (Figure 2A). Since changes in *p53* protein levels have been attributed to increases in *p53* protein stability, we investigated the half-life of *p53* protein in MALAT1-knockdown cells. The half-life of *p53* in MALAT1-knockdown cells was not changed compared with control cells (Figures 2B,C), suggesting that the increase in *p53* protein in MALAT1-knockdown cells was not due to increased protein stability.

We then determined whether changes in *p53* mRNA levels were the basis of increased *p53* protein in MALAT1-knockdown cells. qRT-PCR revealed that the level of *p53* mRNA was increased by approximately three-fold in MALAT1-knockdown cells (Figure 2D). Pre-mRNA level of *p53* gene was also



upregulated in MALAT1-knockdown cells (**Figure 2E**). We also observed upregulation of mature and pre-mature *p53* mRNAs upon MALAT1 depletion in HCT116 cell, a human colorectal carcinoma cell line expressing normal *p53* (**Figure S4**). These results suggest that knockdown of MALAT1 resulted in increased *p53* expression by increasing *p53* mRNA levels.

MALAT1 Negatively Regulates *p53* Promoter Activity

To further investigate the mechanism for increased expression of *p53* mRNA in MALAT1-knockdown cells, we analyzed *p53* promoter activity. Since the *p53* gene (*TP53*) has several alternative promoters (*p53* P1 promoter, which is located within a 356 bp region upstream of the major transcription start site of the *p53* gene, *p53* P2 promoter and *p53* P3 promoter) (Tuck and Crawford, 1989; Hollstein and Hainaut, 2010), we first performed massive transcriptional start site (TSS) analysis (Suzuki et al., 1997; Tsuchihara et al., 2009), which can easily monitor the genome-wide positions of TSSs. The number of TSS-tags corresponds to the number of transcripts that initiate from the site, therefore TSS analysis enables the determination

of regions that can upregulate transcription of the *p53* gene in MALAT1-knockdown cells. TSS analysis revealed that the number of TSS-tags at the position of the P1 promoter of *p53* gene was upregulated by more than two-fold in MALAT1-knockdown cells (**Figure 3**). In contrast, the TSS-tag counts of *GAPDH* were not increased in MALAT1-knockdown cells. This result suggests that transcription from the *p53* P1 promoter was upregulated in MALAT1-knockdown cells. The *p53* P1 promoter is located within the noncoding exon 1, and this promoter region can drive the transcript that encodes the active form of *p53* protein (Wang and El-Deiry, 2006).

To examine whether the P1 promoter of *p53* gene is activated in MALAT1-knockdown cells, we performed luciferase reporter assays. We transfected MALAT1-knockdown cells with a reporter plasmid containing the P1 promoter, a 356-bp region located −344 to +12 relative to the major TSS (Tuck and Crawford, 1989; Hollstein and Hainaut, 2010). The results showed that promoter activity of the 356 bp region was increased in MALAT1-knockdown cells by more than four-fold compared with control cells, suggesting increased P1 promoter activity in MALAT1-knockdown cells (**Figure 4A**). The promoter activity

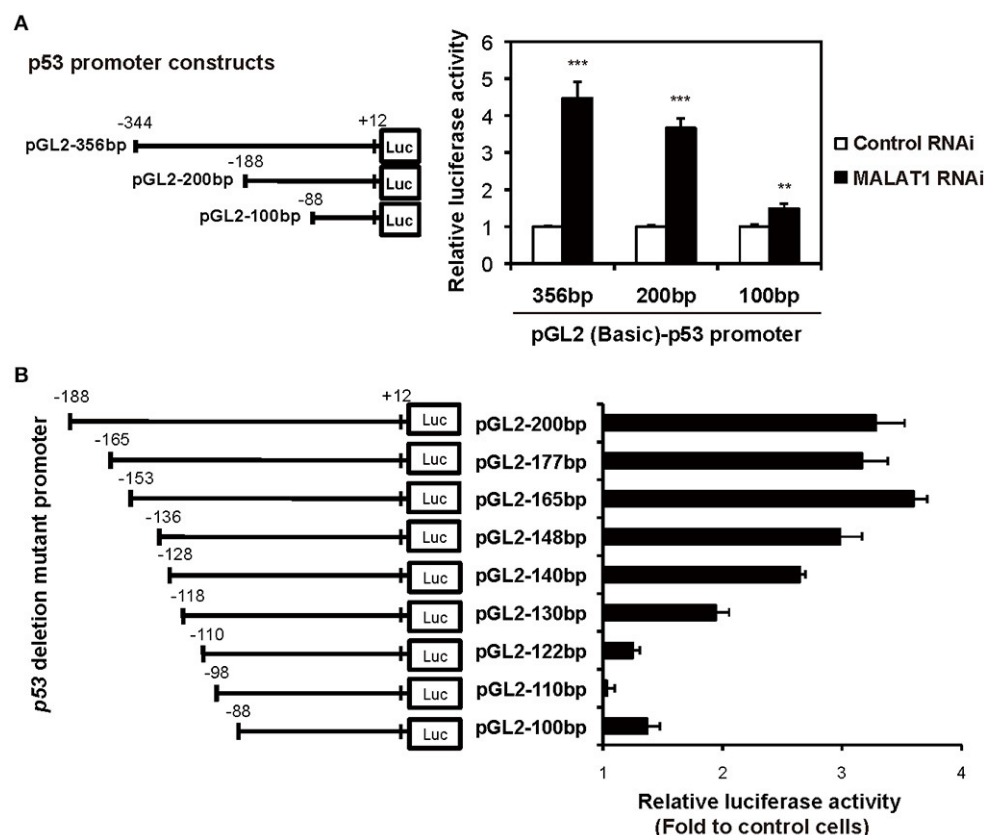


FIGURE 4 | Determination of the MALAT1-responsive region in the human *p53* promoter. **(A)** The activities of the human *p53* promoter regions (pGL2-356, −200, and −100 bp) in MALAT1-knockdown cells were examined using luciferase assays. The data are presented as fold induction compared with the activity in control cells. The *p53* promoter-reporter firefly luciferase activity is indicated relative to the activity of the Renilla luciferase control. All transfections were normalized with equal amounts of pGL2-Control vector. Data are means \pm SD ($n = 3$ each, $**P < 0.01$, $***P < 0.001$, Student's *t*-test). **(B)** The relative luciferase activities of the 5' deletion mutant promoters in MALAT1-knockdown cells compared with control cells were determined. Data is shown as fold induction compared with the activity in the control cells (activity in the control cells was set at 1.0).

of the 200-bp region (−188 to +12) was also elevated in MALAT1-knockdown cells. However, further deletion to 100 bp (−88 to +12) resulted in a two- to three-fold reduction in promoter activity. These results suggest that the P1 promoter region from −188 to −88 contain *cis* elements that may be regulated by MALAT1.

To further elucidate the MALAT1-responsive region in the P1 promoter, we constructed a series of seven 5′ deletion mutant reporter plasmids (pGL2-177, −165, −148, −140, −130, −122, and −110 bp), in which the 5′ ends of the 200-bp region of the *p53* promoter was deleted. The deletion mutant plasmids were transfected into MALAT1-knockdown cells and assayed for

luciferase activity. The reporter activity of the *p53* promoter in MALAT1-knockdown cells was gradually reduced by deletion from 165 bp (−153 to +12) to 122 bp (−110 to +12) of the *p53* promoter (**Figure 4B**). This finding suggests a possible MALAT1-responsive region in the *p53* promoter. We speculated that MALAT1 could regulate the activity of the transcription factors that bind to this region (between −153 to −110) in the *p53* promoter.

We next evaluated possible binding sites for transcription factors in the MALAT1-responsive region in the *p53* promoter. As summarized in **Figure 5A**, 15 binding sites for transcription factors were predicted using TRANSFAC database analysis of

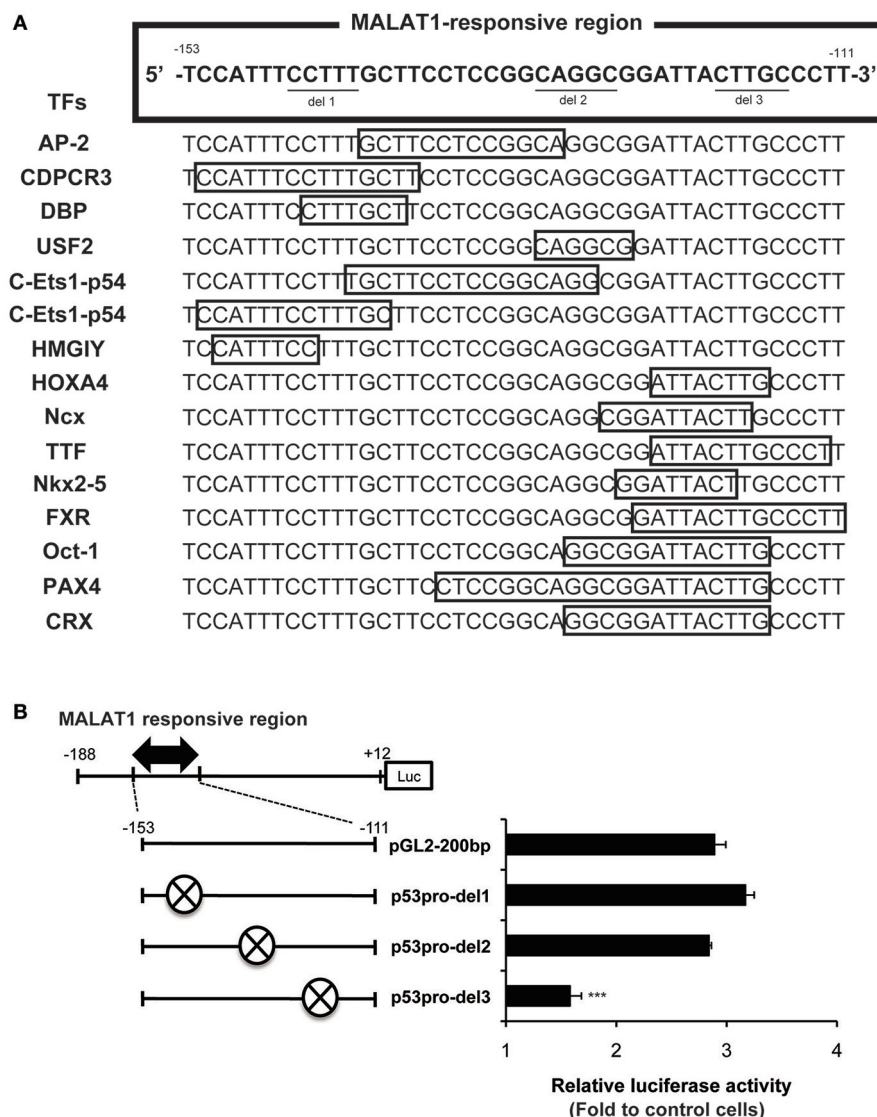


FIGURE 5 | Analysis of the MALAT1-responsive region in the human *p53* promoter. **(A)** TRANSFAC database analysis of the MALAT1-response region (−153 to −111) in the *p53* promoter. Transcription factors (TFs, shown on the left) and their binding sites (open boxes) predicted by TRANSFAC are shown. **(B)** The relative luciferase activities of the deletion mutant promoters (p53pro-del1, del2, and del3), in which parts of MALAT1-response elements were deleted, were determined. The deleted sequences (crosshatch, ×) of each mutant promoter are CCTTT (p53pro-del1), CAGGC (p53pro-del2), and CTTGC (p53pro-del3). Data are means ± SD ($n = 3$ each, *** $P < 0.001$ compared with the luciferase activity of pGL2-200 bp).

the MALAT1-response region (−153 to −111). Interestingly, approximately half of these predicted binding sites were located in the 3′-downstream portion of the MALAT1-responsive region. To evaluate the contribution of these binding sites in the 3′-downstream portion of the MALAT1-responsive region, we constructed mutant reporter plasmids with deletions in this region and examined luciferase activity in MALAT1-knockdown cells. A five-nucleotide deletion in the 3′-downstream region (p53pro-del3) resulted in decreased luciferase activity compared with the original sequence in MALAT1-knockdown cells. Other mutants with deletions in other sites, but not in the 3′-downstream element (p53pro-del1 and p53pro-del2), did not show reduced luciferase activity (Figure 5B). These results indicate that binding sites in the 3′-downstream portion of the MALAT1-responsive region is responsible for transcriptional regulation of *p53* by MALAT1. Therefore, we speculated that MALAT1 modulates *p53* promoter activity through regulation of transcription factors predicted to bind to this region.

A previous report suggested that MALAT1 interacts with SR splicing factors, such as SRSF1, which is also known as SF2/ASF, and localizes to nuclear speckles to regulate alternative splicing of pre-mRNA (Tripathi et al., 2010). To rule out the possibility that upregulation of *p53* expression through activation of *p53* promoter is arbitrarily attributed to abnormal alternative splicing caused by knockdown of MALAT1, we examined the effect of depletion of SRSF1 on *p53* expression. Knockdown of *SRSF1* increased the inclusion of exon 19 of *MGEA6*, which is regulated by SRSF1 as well as MALAT1 (Tripathi et al., 2010), indicating changes in alternative splicing in *SRSF1* knockdown cells (Figure S1A). However, knockdown of *SRSF1* did not affect expression of *p53* mRNA (Figure S1B). This result suggests that increased expression of *p53* mRNA in MALAT1-knockdown cells is independent of MALAT1 regulation of alternative splicing. This finding strongly supports our idea that MALAT1 affects *p53* expression through regulating *p53* promoter activity.

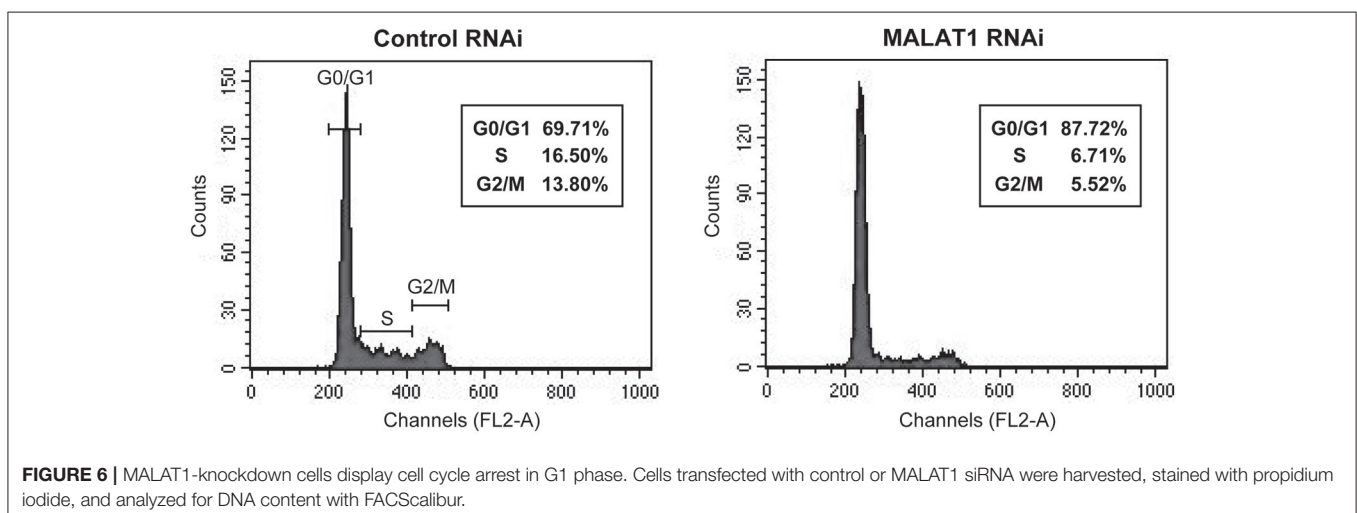
MALAT1 is known to localize to nuclear speckles. To determine whether nuclear speckle localization of MALAT1 is necessary for regulation of *p53* expression, we examined *p53*

expression levels when localization of MALAT1 was disrupted by knockdown of RNPS1 or SRm160, which are nuclear speckle proteins that contribute to MALAT1 nuclear speckle localization (Miyagawa et al., 2010). Expression level of *p53* was not increased by disrupting the localization of MALAT1 to nuclear speckles using knockdown of RNPS1 and SRm160 genes instead of MALAT1 knockdown (Figure S2). We confirmed that knockdown of RNPS1 and SRm160 had little effect on MALAT1 expression (less than two-fold). This result indicates that localization of MALAT1 to nuclear speckles is not necessary for controlling expression of *p53*. This finding indicates the possibility that MALAT1 functions in mechanisms at other nuclear structures in addition to nuclear speckles.

Depletion of MALAT1 Induces G1 Cell Cycle Arrest

p53 mRNA levels are tightly regulated during the cell cycle, with its transcription induced before DNA synthesis and a peak production at the G1-S cell cycle transition (Wang and El-Deiry, 2006). In addition, p21, which is upregulated by *p53* in MALAT1-knockdown cells, is a major mediator of G1 cell cycle arrest (Chen et al., 1996). To explore the biological significance of MALAT1-mediated *p53* expression, we examined whether MALAT1 knockdown leads to cell cycle arrest in G1. Flow cytometry analysis revealed that depletion of MALAT1 in A549 cells exhibited a higher proportion of cells in G0/G1 (approximately 90%) and lower proportion in S (7%) and G2/M (6%), compared with control cells (Figure 6), indicating that depletion of MALAT1 leads to cell cycle arrest in G1 phase. Therefore, the function of MALAT1 in regulating *p53* promoter activity would contribute to the regulation of cell cycle progression, especially in G1 phase.

In the present study, we demonstrated that upregulation of *p21* and *FAS* in MALAT1-depleted cells was mediated by *p53*. We also showed that depletion of MALAT1 increased the expression of *p53* at both the protein and mRNA levels. In addition, we found that depletion of MALAT1 led to increased activity of the *p53* promoter through specifically affecting 3′-downstream



elements in the MALAT1-responsive region in *p53*. Together results suggest that MALAT1 reduces the expression of *p21* and *FAS* through repression of *p53* promoter activity.

We also demonstrated that depletion of MALAT1 in A549 cells resulted in G1 cell cycle arrest. Tripathi et al. showed that depletion of MALAT1 in human diploid fibroblasts leads to defects in cell cycle progression in G1/S and activation of *p53* (Tripathi et al., 2013). However, the mechanisms by which MALAT1 regulates *p53* expression were not determined. Herein, we propose a novel function for MALAT1 in regulating *p53* promoter activity and controlling cell cycle progression.

Our data identified eight transcription factors that were predicted to bind to 3'-downstream portion of the MALAT1-responsive region, including HOXA4, Ncx, TTF, Nkx2-5, FXR, Oct-1, PAX4, and CRX. Although *p53* is a well-known tumor suppressor subject to multiple regulations at the transcriptional level (Tripathi et al., 2013), none of these candidate transcription factors were previously identified as *p53* regulatory factors. Of these, Nkx2-5 and FXR are most the likely candidates of *p53* regulatory factors because TSS analysis revealed that the number of TSS-tags of their target genes, *ECE-1* and *SOCS3*, respectively, were increased in MALAT1-depleted cells. Therefore, these factors may be responsible for the MALAT1-mediated regulation of the *p53* promoter. Further studies are required to elucidate the precise molecular mechanism of *p53* promoter regulation by MALAT1 with responding transcription factors.

Together our findings help elucidate the novel regulation of the *p53* promoter by the long noncoding RNA, MALAT1, and further our understanding of MALAT1 function in cancer biology.

AUTHOR CONTRIBUTIONS

KT: carried out most experiments; RO-M, YS, and TY: carried out part of FACS, NGS, and bioinformatics, respectively; FY carried

out part of RT-qPCR; KT, FU, and NA: designed this study; KT, RO-M, YS, TY, FU, and NA: wrote the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Tomoaki Tanaka (Chiba University) for providing helpful discussion and advice. This work was financially supported by the Suzuken Memorial Foundation, the Naito Foundation, MEXT KAKENHI noncoding RNA and MEXT KAKENHI Grant Number 221S0002.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2017.00208/full#supplementary-material>

Figure S1 | Increased expression of *p53* mRNA in MALAT1-knockdown cells is independent of alternative splicing regulated by MALAT1. **(A)** Changes in alternative splicing in *SRSF1* knockdown cells are shown by increased inclusion of exon 19 of *MGEA6*. RT-PCR analysis was performed using primers specific for *SRSF1*-regulated alternative exons in *MGEA6*. Alternative exon-included (upper band) and exon-excluded bands (lower band) are shown. **(B)** Quantitative real-time PCR analysis of *SRSF1* expression levels (left) and *p53* mRNA levels (right) in *SRSF1* knockdown cells compared with control cells. Values represent the means \pm SD of duplicate measurements.

Figure S2 | Localization of MALAT1 to nuclear speckles is not necessary for regulation of *p53* expression. **(A,B)** Quantitative real-time PCR analysis of *p53* mRNA levels in RNPS1- **(A)** or SRm160- **(B)** knockdown cells. Values represent the means \pm SD of duplicate measurements. **(C,D)** Quantitative real-time PCR analysis of MALAT1 expression levels in RNPS1- **(C)** or SRm160- **(D)** knockdown cells. Values represent the means \pm SD of duplicate measurements.

Figure S3 | Increased expression levels of indicated *p53* target mRNAs upon MALAT1 knockdown. Real-time PCR analyses determined the expression levels of indicated RNAs those are normalized by GAPDH mRNA. Data are presented as means \pm errors of two independent experiments.

Figure S4 | Increased expression levels of pre-matured and matured *p53* mRNAs in MALAT1-knockdown cells. Real-time PCR analyses were performed to assess the indicated RNAs those are normalized by GAPDH mRNA. Data are presented as means \pm standard deviation (SD) of three independent experiments (* $P < 0.05$, ** $P < 0.01$, Student's *t*-test).

REFERENCES

- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874
- Chen, X., Ko, L. J., Jayaraman, L., and Prives, C. (1996). *p53* levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev.* 10, 2438–2451. doi: 10.1101/gad.10.19.2438
- Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., et al. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* 159, 188–199. doi: 10.1016/j.cell.2014.08.018
- Gutschner, T., Hammerle, M., and Diederichs, S. (2013a). MALAT1 - a paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* 91, 791–801. doi: 10.1007/s00109-013-1028-y
- Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013b). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850
- Hollstein, M., and Hainaut, P. (2010). Massively regulated genes: the example of TP53. *J. Pathol.* 220, 164–173. doi: 10.1002/path.2637
- Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8:39. doi: 10.1186/1471-2164-8-39
- Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., et al. (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041. doi: 10.1038/sj.onc.1206928
- Lin, R., Maeda, S., Liu, C., Karin, M., and Edgington, T. S. (2007). A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* 26, 851–858. doi: 10.1038/sj.onc.1209846
- 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
- Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., et al. (2010). Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* 18, 738–751. doi: 10.1261/rna.028639

- Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., et al. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* 18, 1487–1499. doi: 10.1261/rna.033217.112
- Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A., and Sugano, S. (1997). Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* 200, 149–156.
- Tani, H., Mizutani, R., Salam, K. A., Tano, K., Ijiri, K., Wakamatsu, A., et al. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22, 947–956. doi: 10.1101/gr.130559.111
- Tano, K., and Akimitsu, N. (2012). Long non-coding RNAs in cancer progression. *Front. Genet.* 3:219. doi: 10.3389/fgene.2012.00219
- Tano, K., Mizuno, R., Okada, T., Rakwal, R., Shibato, J., Masuo, Y., et al. (2010). MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett.* 584, 4575–4580. doi: 10.1016/j.febslet.2010.10.008
- Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011
- Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S. M., Wu, X., et al. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* 9:e1003368. doi: 10.1371/journal.pgen.1003368
- Tsuchihara, K., Suzuki, Y., Wakaguri, H., Irie, T., Tanimoto, K., Hashimoto, S., et al. (2009). Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Res.* 37, 2249–2263. doi: 10.1093/nar/gkp066
- Tuck, S. P., and Crawford, L. (1989). Characterization of the human p53 gene promoter. *Mol. Cell. Biol.* 9, 2163–2172.
- Wang, S., and El-Deiry, W. S. (2006). p73 or p53 directly regulates human p53 transcription to maintain cell cycle checkpoints. *Cancer Res* 66, 6982–6989. doi: 10.1158/0008-5472.CAN-06-0511
- West, J. A., Davis, C. P., Sunwoo, H., Simon, M. D., Sadreyev, R. I., Wang, P. I., et al. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802. doi: 10.1016/j.molcel.2014.07.012
- Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., et al. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147, 773–788. doi: 10.1016/j.cell.2011.08.054
- Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., et al. (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* 2, 111–123. doi: 10.1016/j.celrep.2012.06.003

Conflict of Interest Statement: 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 reviewer AM and handling Editor declared their shared affiliation.

Copyright © 2018 Tano, Onoguchi-Mizutani, Yeasmin, Uchiumi, Suzuki, Yada and Akimitsu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Wnt/ β -catenin Signaling Pathway Regulates Specific lncRNAs That Impact Dermal Fibroblasts and Skin Fibrosis

Nathaniel K. Mullin^{1†}, Nikhil V. Mallipeddi^{1†}, Emily Hamburg-Shields¹, Beatriz Ibarra¹, Ahmad M. Khalil^{2*‡} and Radhika P. Atit^{1,2,3*‡}

¹ Department of Biology, Case Western Reserve University, Cleveland, OH, United States, ² Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, United States, ³ Department of Dermatology, Case Western Reserve University, Cleveland, OH, United States

OPEN ACCESS

Edited by:

Kinji Ohno,
Nagoya University, Japan

Reviewed by:

Kaushendra Tripathi,
University of Alabama at Birmingham,
United States
Félix Recillas-Targa,
Universidad Nacional Autónoma
de México, Mexico

*Correspondence:

Radhika P. Atit
rpa5@case.edu
Ahmad M. Khalil
dr.ahmad.khalil@gmail.com

[†]Co-first authors

[‡]Co-senior authors

Specialty section:

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

Received: 19 September 2017

Accepted: 06 November 2017

Published: 21 November 2017

Citation:

Mullin NK, Mallipeddi NV,
Hamburg-Shields E, Ibarra B,
Khalil AM and Atit RP (2017)
Wnt/ β -catenin Signaling Pathway
Regulates Specific lncRNAs That
Impact Dermal Fibroblasts and Skin
Fibrosis. *Front. Genet.* 8:183.
doi: 10.3389/fgene.2017.00183

Wnt/ β -catenin signaling is required for embryonic dermal fibroblast cell fate, and dysregulation of this pathway is sufficient to promote fibrosis in adult tissue. The downstream modulators of Wnt/ β -catenin signaling required for controlling cell fate and dermal fibrosis remain poorly understood. The discovery of regulatory long non-coding RNAs (lncRNAs) and their pivotal roles as key modulators of gene expression downstream of signaling cascades in various contexts prompted us to investigate their roles in Wnt/ β -catenin signaling. Here, we have identified lncRNAs and protein-coding RNAs that are induced by β -catenin activity in mouse dermal fibroblasts using next generation RNA-sequencing. The differentially expressed protein-coding mRNAs are enriched for extracellular matrix proteins, glycoproteins, and cell adhesion, and many are also dysregulated in human fibrotic tissues. We identified 111 lncRNAs that are differentially expressed in response to activation of Wnt/ β -catenin signaling. To further characterize the role of mouse lncRNAs in this pathway, we validated two novel Wnt signaling- Induced Non-Coding RNA (Wincr) transcripts referred to as *Wincr1* and *Wincr2*. These two lncRNAs are highly expressed in mouse embryonic skin and perinatal dermal fibroblasts. Furthermore, we found that *Wincr1* expression levels in perinatal dermal fibroblasts affects the expression of key markers of fibrosis (e.g., *Col1a1* and *Mmp10*), enhances collagen contraction, and attenuates collective cell migration. Our results show that β -catenin signaling-responsive lncRNAs may modulate dermal fibroblast behavior and collagen accumulation in dermal fibrosis, providing new mechanistic insights and nodes for therapeutic intervention.

Keywords: gene expression, *Mmp10*, *Wincr1*, *Wincr2*, dermis, development

INTRODUCTION

Wnt/ β -catenin signaling has a diverse role in both embryonic mouse skin development and in human skin diseases such as pilomatricomas, dermal hypoplasia, and fibrosis (Wang et al., 2007; Lam and Gottardi, 2011; Hamburg and Atit, 2012; Lim and Nusse, 2012). β -catenin is a key transducer of the Wnt/ β -catenin signaling pathway and a regulator of transcription

(van Amerongen and Nusse, 2009; Schuijers et al., 2014; McCrea and Gottardi, 2016). Cell type and context-specific target gene expression provide specificity for the diverse functions of the Wnt signaling pathway (Nakamura et al., 2009; Nusse and Clevers, 2017). Elucidating the downstream modulators of the Wnt signaling pathway is critical to our understanding of how this pathway influences distinct cell types in development and disease.

Dermal fibroblasts are key contributors to hair follicle development, regional identity, skin patterning, wound healing, and skin fibrosis (Chang et al., 2002; Eames and Schneider, 2005; Rendl et al., 2005; Enzo et al., 2015). We have previously demonstrated that dermal Wnt/ β -catenin activity is required for mouse embryonic dermal fibroblast identity and hair follicle initiation (Atit et al., 2006; Ohtola et al., 2008; Tran et al., 2010; Chen et al., 2012). Activation of the Wnt signaling pathway in humans is also a common feature in fibrosis of varying organs such as lung, liver, kidney, and skin (Enzo et al., 2015). We found that sustained Wnt/ β -catenin activation in dermal fibroblasts is sufficient to cause dermal fibrosis in the adult mouse (Akhmetshina et al., 2012; Hamburg and Atit, 2012; Mastrogiannaki et al., 2016). Transcriptome analysis of mouse fibrotic dermis showed an increase in mRNA levels of regulatory genes, such as *Col7a1*, *Ccn3/Nov*, *Biglycan*, and *Matrix Metalloproteinase 16 (Mmp16)*, a subset of which are also up-regulated in human skin fibrosis and tumor stroma (Hamburg-Shields et al., 2015). Thus, studying the various mechanisms of β -catenin-mediated gene regulation will provide new insights into how context-specific transcriptional targets are activated and repressed in skin development and disease.

In recent years, lncRNAs have emerged as crucial intermediate facilitators for a variety of cellular pathways and gene expression networks (Rinn and Chang, 2012; Wan and Wang, 2014). Regulatory lncRNAs are defined as >200 nt-long RNA molecules readily present within both the cytosol and nucleus but lacking protein-coding capacity. Various studies have implicated lncRNAs in directly facilitating gene expression, both in *cis* and in *trans* (Khalil et al., 2009; Moran et al., 2012). Furthermore, studies have documented the existence of differentially expressed lncRNAs within specific fibrotic conditions (Huang et al., 2015; Micheletti et al., 2017; Piccoli et al., 2017; Qu et al., 2017). Recently, TGF β -responsive lncRNAs have been shown to directly control the expression of fibrotic genes, demonstrating a role for lncRNAs in a pathway with a similar function as Wnt/ β -catenin signaling (Fu et al., 2016; Wang et al., 2016). An increasing number of studies are investigating the direct role that lncRNAs play in influencing Wnt/ β -catenin signaling (Fan et al., 2014; Vassallo et al., 2015; Wang et al., 2015; Ma et al., 2016).

Wnt/ β -catenin signaling-induced lncRNAs that function as downstream effectors of Wnt signaling to modulate gene expression have yet to be fully characterized. Here, we expressed a stabilized version of β -catenin protein from the endogenous locus in neonatal mouse primary dermal fibroblasts to identify Wnt/ β -catenin signaling-responsive mRNAs and lncRNAs. We functionally characterized one of the most differentially expressed lncRNAs, *GM12603*, referred to as *Wincrl* in the context of

fibrotic gene expression and fibroblast behavior. Our results show that *Wincrl* has gene-regulatory and functional roles in key behaviors of dermal fibroblasts such as collective cell migration and collagen contraction.

MATERIALS AND METHODS

Animals and Ethics

Ctnnb1 ^{Δ ex3/+} (Harada et al., 1999), *Engrailed1Cre (En1Cre)* (Kimmel et al., 2000), *Gt(ROSA)26Sor^{tm1(rtTA,EGFP)}Nagy* (*R26rtTA*) Jax labs stock: 005670 (Belteki et al., 2005), *Teto-deltaN89 β -catenin* (Mukherjee et al., 2010) were maintained on mixed genetic background (CD1, C57Bl6) and genotyped as previously described. Triple transgenic *Engrailed1Cre/+; Rosa26rtTA/+; Teto-deltaN89 β -catenin/+* experimental mice were generated. For each experiment, a minimum of three mutants with litter-matched controls were studied except where otherwise noted. Animals of both sexes were randomly assigned to all studies. Case Western Reserve Institutional Animal Care and Use Committee approved all animal procedures in accordance with AVMA guidelines (Protocol 2013-0156, approved 21 November 2014, Animal Welfare Assurance No. A3145-01).

Collection and Culture of Primary Dermal Fibroblasts

Whole ventral skin from *Ctnnb1* ^{Δ ex3/+} mice was dissected from the trunk of postnatal days 4–7 (P4–7) mice and minced (Harada et al., 1999). Skin was incubated in Dulbecco's Modified Eagle's Medium: Nutrient Mixture F-12 (DMEM/F12) (Thermo Fisher Cat. No. 11320) and 50 mg/mL Liberase DL (Roche Cat. No. 5401160001) at 37°C under constant rotation for 1 h. Dermis was dissociated by vigorous pipetting and passing through an 18 gauge needle (BD Cat. No. 305196). Cells from each animal were cultured individually in complete growth media (DMEM (Thermo Fisher Cat. No. 11995065) + 10% Fetal Bovine Serum (FBS) (Thermo Fisher Cat. No. 10082147) + 1% Penicillin/Streptomycin (Invitrogen Cat. No. 15140122) + 1% Antibiotic-Antimycotic (Invitrogen Cat. No. 14240062) in 5% CO₂ except where otherwise noted. After 90 min, media was removed and replaced with fresh culture media. After two passages, Adenovirus was administered in basal DMEM (no FBS). Dermal fibroblasts were infected with Adenovirus-Cre (366–500 MOI) or Adenovirus-GFP (366 MOI) (Adenovirus purchased from University of Iowa). After 48 h, recombination of *Ctnnb1* Exon 3 was confirmed with site specific PCR under standard conditions. Primers *Neo PMR* (5' AGACTGCCTTGGGAAAAGCG 3') and *Cat-AS5* (5' ACGTGTGGCAAGTTCGCGTCATCC 3') were used to identify the targeted allele with a ~500 bp amplicon and *GF2* (5'GGTAGGTGAAGCTCAGCGCAGAGC 3') and *Cat-AS5* identified the recombined allele with an expected amplicon of ~700 bp (Harada et al., 1999).

For *in vivo* induction of stabilized β -catenin expression in *Engrailed1Cre/+; Rosa26rtTA/+; Teto-deltaN89 β -catenin/+* mice, pregnant dams were given 80 μ g of doxycycline (Sigma

Cat. No. D9891) per gram of body weight by intraperitoneal injection at embryonic day 12.5 (E12.5) and embryonic cranial and dorsal dermal skin were harvested at E13.5. *In vitro* stabilized β -catenin expression was induced by treating *Engrailed1*Cre/+; *Rosa26rtTA*/+; *Teto-deltaN89 β -catenin*/+ P4 ventral dermal fibroblasts cells with 2 μ g/mL doxycycline in complete growth media for 4 days. For induction-reversal experiments, duplicate cultures were then switched to complete growth media without doxycycline for 48 h prior to RNA isolation.

Whole-Genome RNA Sequencing

Total RNA was extracted from embryonic tissues *in vivo* and dermal fibroblasts *in vitro* using TRIzol reagent (Thermo Fisher Cat. No. 15596026). RNA was isolated using the RNeasy MinElute kit (Qiagen Cat. No. 74204) with DNaseI (Qiagen Cat. No. 79254) treatment, following the manufacturer's protocol. RNA concentration and quality were measured using the NanoDrop 8000 UV-Vis Spectrophotometer.

Libraries were prepared by the CWRU Genomics Sequencing Core, using the True-Seq Stranded Kit (Illumina). Paired-end sequencing was carried out on the Illumina HiSeq 2500 platform. Resulting 100 bp reads were mapped to the mm10 mouse genome release using TopHat and Cufflinks. Mapped raw reads were counted, normalized to total mapped reads, and used for differential gene expression (Cuffdiff and Seqmonk standard settings). Differential expression was defined as an absolute fold change greater than 2, and Benjamini–Hochberg adjusted *P*-value less than 0.05. Normalized mapped reads are available on Gene Expression Omnibus (GSE103870) using the private token from the editor.

Quantitative PCR and Primers

Total RNA was extracted from P4 ventral dermal fibroblasts between passage 3–6 as mentioned above. cDNA was generated using the Invitrogen High Capacity RNA-to-cDNA Reverse Transcription Kit (Thermo Fisher Cat. No. 4374966). Relative mRNA quantities of select genes were determined using the Applied Biosystems StepOnePlus Real-Time PCR System (Life Technologies Cat. No. 4376600) and the Δ Ct or $\Delta\Delta$ Ct method where applicable (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008). In all plots, sample and control RQs were normalized to the mean RQ of the control group. *Axin2* quantity was measured relative to the reference gene *ActB* using Taqman probes from Thermo Fisher (Mm00443610_m1 and Mm02619580_g1, respectively) or *HPRT* (mm1545399_m1). *Winc1* isoform1 (F: TGATCCCACTGAAAATGCTG, R:GGTG ATTTGACCTGCCATCT) and *Winc2* (F:GGCCTGGATAG AGGTCTCC, R:TAGTTCTCTCCATCGGTTTCC) quantities were measured relative to reference gene *RPL32* (F:TTAAGC GAACTGGCGGAAAC, R:TTGTTGCTCCATAACCGATG) using custom-designed primers (Invitrogen) and SyBr Green reagents (Invitrogen Cat. No. 4367659). *Mmp10* and *Colla1* quantities (Mm01168399_m1 and Mm00801666_g1, respectively) were measured relative to reference gene *ActB* expression, using TaqMan Gene Expression Master Mix (Thermo Fisher Cat. No. 4369016). Non-coding gene primers

were designed using Primer3Plus¹. Primer sequences for *RPL32* and *Mmp10* were acquired through the MGH Primer Bank Site². Primer sequences for *Colla1* were identified previously (He et al., 2005).

Statistics were performed using GraphPad Prism 7. For experiments in which cells from the same animal were used in control and experimental conditions, a paired *t*-test was used. In all other instances, an unpaired *t*-test was used. Significance for all purposes was defined as **P*-value ≤ 0.05 , ***P*-value ≤ 0.01 , ****P*-value ≤ 0.001 , *****P*-value ≤ 0.0001 . Paired samples are shown as dots connected by a line where applicable. Expression across different tissues are shown as mRNA levels relative to reference gene ($2^{-\Delta Ct}$). In all other plots, individual relative quantities ($2^{-\Delta\Delta Ct}$) are shown along with mean and standard error of the mean.

Bioinformatic Analysis

Heatmaps were generated using all mRNA or lncRNA genes considered significantly differentially expressed [abs(fold change) > 2 , adjusted *P*-value < 0.05] between GOF and control samples. Color was assigned to each sample on a basis of Z-score of fragments per kilobase mapped (FPKM) compared to other samples' FPKM of the same gene (row Z-score). Heatmaps were clustered hierarchically based on the aggregate differential gene expression profiles using the DESeq2 package in R.

DAVID Functional Annotation Clustering³ was performed on all differentially expressed protein-coding genes [abs(fold change) > 2 , adjusted *P*-value < 0.05]. Gene lists were entered as Gene Symbols. All settings were default. The top five ranked clusters are shown, with clusters enriched among up-regulated genes shown as positive and clusters enriched among down-regulated genes shown as negative.

Overrepresentation of predicted transcription factor binding sites was determined using oPOSSUM 3.0 Single Site Analysis (SSA)⁴ (Ho Sui et al., 2007; Kwon et al., 2012). Gene lists were loaded into the browser-based software by gene symbol, and putative transcription factor binding sites (TFBS) were scored based on their overrepresentation in the areas surrounding the transcription start site (TSS) of such genes, as compared to prevalence in the rest of the genome. All 29,347 genes in the oPOSSUM database were used as background. For analysis of TFBS enrichment around lncRNAs, custom FASTA files were generated for the promoter sequence of each gene. Proximity to TSS was defined as ± 5 kb for all analysis. All JASPAR PBM profiles were queried. Each TFBS was assigned a Z-score and a Fisher score for overrepresentation. TFBS were then plotted according to these scores using GraphPad Prism 7. Thresholds were determined as follows: Z-score – Mean + (2 \times Standard Deviation), Fisher score – Mean + Standard Deviation as previously described (Kwon et al., 2012).

Gene Set Enrichment Analysis (GSEA) was used to identify known gene expression signatures similar to the differential

¹<http://primer3plus.com/>

²<https://pga.mgh.harvard.edu/primerbank/>

³<https://david.ncifcrf.gov/>

⁴<http://opossum.cisreg.ca/oPOSSUM3/>

expression of genes in GOF dermal fibroblasts (Mootha et al., 2003; Subramanian et al., 2005). GSEA was run as a Java Applet. MSigDB C2 Curated Gene Sets were queried. Gene sets smaller than 15 genes and larger than 500 genes were excluded from the analysis and 1000 *gene_set* permutations were used. All other settings remained default.

The human matrisome gene list was accessed through the MIT Matrisome Project at MatrisomeDB⁵. The intersection of this list with others was performed in Microsoft Excel.

Coexpression networks were constructed using Cytoscape version 3.4.0 (Shannon et al., 2003). Candidate genes were selected on the basis of significant differential upregulation (adjusted *P*-value < 0.05 and fold change > 2) and inclusion in the Matrisome gene list (see above). An FPKM table was constructed in Microsoft Excel consisting of these candidate genes as well as *Axin2*, *Wincrl*, and *Wincrl2*. This table was loaded into Cytoscape as an unassigned table. The plugin “Expression Correlation” was used (with a Low Cutoff of -1 and a High Cutoff of 0.6) to generate a correlation network. The built-in tool “Network Analyzer” was used to further analyze the network for node degree. A custom style was created to visualize the network, in which edge color correlates with strength of correlation (red:high CC). Node size correlated with number of first neighbors (degree).

Human Fibrotic Disease Study Expression Analysis

Microarray data from nine human studies of various fibrotic conditions were analyzed using Gene Expression Omnibus (GEO). Data were analyzed using the GEO2R tool. Genes with average fold change > 1.5 and adjusted *P*-value < 0.05 were considered significantly differentially expressed.

In Vitro Lentivirus Overexpression and Gapmer Knockdown of *Wincrl*

Full-length isoform 1 of *Wincrl* was assembled from synthetic oligonucleotides and PCR products and subcloned into pMA-T (Life Technologies, Ref. No. 1725930). *Wincrl* lentivirus was generated by subcloning *Wincrl* from pMA-T into the NcoI-EcoRV region of pENTR4 (Thermo Fisher Cat. No. A10465) vector and then swapping it into the pGK destination vector (Addgene Cat. No. 19068). Lentivirus was generated in 293T (f-variant) cells (ATCC Cat. No. 3216) by simultaneously transfecting pGK Destination Vector containing *Wincrl* (4.5 μ g), PMD2G helper plasmid (1.6 μ g) (Addgene Cat. No. 12259), and pCMV-dR874 plasmid (3.2 μ g), using Lipofectamine 3000 reagent (18 μ L) (Invitrogen Cat. No. L30000015) in 1.5 mL Opti-MEM (Thermo Fisher Cat. No. 31985062) media in a 6 cm dish. Cells were selected in Puromycin (2 μ g/mL) (Sigma Cat. No. P8833). Virus was collected from 293T conditioned cell media at 24 and 52 h after transfection, clarified with 45 μ m pore filter (PES membrane, Thermo Fisher Cat. No. 7252545) and frozen. Lentivirus conditioned media was titrated on 293T cells at 3.25–50% dilutions.

After 3 days, infected cells were selected by treatment with Puromycin (2 μ g/mL) for 3 days. Stable expressing P4 mouse ventral dermal fibroblast cells lines were propagated for further experiments. *Wincrl* overexpression was confirmed by qRT-PCR as described.

Wincrl knockdown was achieved with custom LNA-GapmeRs designed and synthesized by Exiqon⁶. Sequences targeting *Wincrl* (GACTAGGATGATAGAT) and a negative (scrambled) control (AACACGTCTATACGC) were acquired. LNA GapmeRs were reconstituted to 50 μ M in tissue culture-grade water as per manufacturer's instruction. Transfection was carried out using Lipofectamine 3000 Reagent in Opti-MEM. Dermal fibroblasts were seeded in a 12-well cell culture treated plate and transfected in 500 μ L of Opti-MEM containing 2 μ L Lipofectamine 3000 and a final LNA GapmeR concentration of 50 nM. Active transfection was carried out for 6 h, at which time the Lipofectamine 3000 containing media was replaced with complete growth media. Additional LNA GapmeRs were added for a final concentration of 100 nM. After 48 h of unassisted transfection, cells were harvested and RNA was extracted and analyzed for gene expression changes as described above. The knockdown of *Wincrl* expression was confirmed by qRT-PCR.

Proliferation, Migration, and Contraction Functional Assays

To measure cell proliferation, a standard growth curve assay was performed on P4 ventral dermal fibroblasts between passage 4–6 as described above. 30,000 cells were plated in duplicate into a 12 well plate, with cell number being assessed using Trypan blue (Invitrogen/Gibco Cat. No. 1520061) exclusion on the Cell Countess (Invitrogen Cat. No. C10281). Collective cell migration was assessed using a qualitative scratch assay (Liang et al., 2007). Briefly, a 200 μ L pipet tip was used to scratch the monolayer and create a 400 μ m gap. Images of cells were taken at time (T) 0, 15, and 22 h. Images were taken on Leica S6D microscope with MC120 HD camera with Leica software. Cell contraction was assessed using the Cell Contraction Assay Kit (Cell Biolabs Inc., CBA-201), following manufacturer instruction. Images of cells were taken with Olympus IX71 microscope with Olympus BX60 camera using Olympus DP controller software. All images were analyzed in Image J software (Schneider et al., 2012).

RESULTS

Global RNA Expression Is Altered after β -catenin Stabilization in Dermal Fibroblasts

To identify coding and non-coding RNAs downstream of Wnt/ β -catenin signaling, we infected neonatal dermal fibroblasts carrying *Ctnnb1* ^{Δ ex3/+} with Adenovirus Cre (Ad-Cre) (Figure 1A). Recombination of Exon 3 of *Ctnnb1* produces

⁵<http://matrisomeproject.mit.edu/other-resources/>

⁶<http://www.exiqon.com/gapmers>

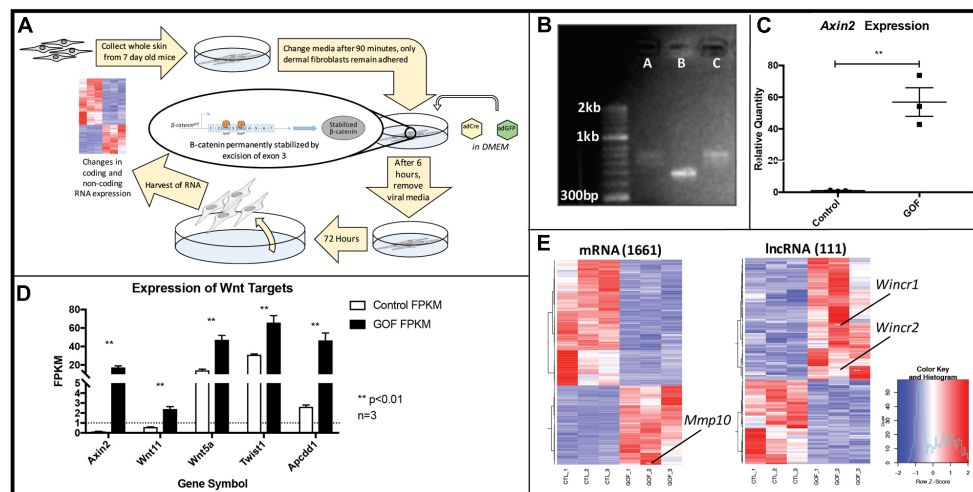


FIGURE 1 | *In vitro* genetic recombination of β -catenin (*Ctnnb1* $^{\Delta ex3}$) in dermal fibroblasts results in a strong Wnt signaling expression signature as well as global expression changes in coding and non-coding genes. **(A)** Schematic of the workflow **(B)** *Ctnnb1* $^{\Delta ex3}$ recombination by Adenovirus-Cre (AdCre) infection was confirmed by *Ctnnb1* exon3 site-specific PCR in lanes A and C, compared to the control lane B, infected with Adenovirus-GFP. **(C)** Relative mRNA quantity of *Axin2*, a direct β -catenin target is significantly higher in β -catenin stabilized gain of function dermal fibroblasts (GOF) (P -value = 0.0035, $n = 3$). **(D)** Known β -catenin targets are higher in GOF dermal fibroblasts by RNAseq ($n = 3$). **(E)** Heat map showing hierarchical clustering (based on entities and samples) of all differentially expressed mRNAs (1661) and lncRNAs (111) (P -value < 0.05, fold change > 2) in response to β -catenin stabilization.

a stabilized form of β -catenin protein, referred to as Gain of Function (GOF), which lacks the phosphorylation site for degradation and constitutively activates the canonical Wnt signaling pathway. Using site-specific primers, we confirm the consistent and effective Ad-Cre-mediated excision of *Ctnnb1* Exon 3 by 48 h post-infection, compared to fibroblasts carrying the same transgene and transduced with Adenovirus GFP (**Figure 1B**) (Harada et al., 1999). *Axin2*, a well-established direct target of β -catenin (Jho et al., 2002), was up-regulated by 40 to 70-fold in cells at 72 h post-infection across all GOF samples (**Figure 1C**) further demonstrating successful excision of *Ctnnb1* Exon 3 and activation of the pathway.

Next, we isolated total RNA and performed expression profiling on control and GOF neonatal dermal fibroblasts ($n = 3$) by whole-genome RNA-sequencing at 72 h post Ad-Cre infection (deposited in GEO, GSE103870). We quantified gene expression using FPKM (see Materials and Methods) and further verified GOF status by examining FPKM values of several well-known Wnt/ β -catenin signaling targets in dermal fibroblasts *in vivo*, such as *Wnt5a*, *Wnt11*, *Apcdd1*, and *Twist2* (Budnick et al., 2016). All target gene expression levels were significantly higher across all three GOF samples, serving as quality control for the sample preparation and sequencing (**Figure 1D**). Within expressed mRNAs and lncRNAs (FPKM > 1), GOF samples independently clustered together via Pearson correlation against the controls (**Supplementary Figure S1**). In total, we identified 1,661 mRNAs and 111 lncRNAs that were differentially expressed across all GOF samples (fold change > 2, adjusted P -value < 0.05) (**Figure 1E**). These findings demonstrate that Wnt signaling differentially regulates lncRNAs as well as mRNAs in dermal fibroblasts.

Stabilization of β -catenin Leads to Dysregulation of Matrisome Genes Relevant to Human Fibrosis

We performed Gene Ontology (GO) Analysis for all differentially expressed mRNAs. These analyses identified key functional groups affected by Wnt/ β -catenin signaling in dermal fibroblasts. Through DAVID, the differentially expressed mRNAs were grouped by functional clusters, with positive enrichment scores indicating functional clusters comprised of up-regulated genes and negative scores indicating clusters of down-regulated genes (**Figure 2A**). The highest scoring functional clusters enriched in up-regulated genes included Glycoprotein, Membrane, Cell junction, and Wnt signaling (fold change > 2, adjusted P -value < 0.05). Functions including Extracellular Matrix and Secreted/Glycoprotein were also highly enriched among those genes that were down-regulated in GOF dermal fibroblasts (fold change, <0.5-adjusted P -value < 0.05).

Since DAVID functional clustering analysis highlighted a strong overrepresentation of ECM genes, we used an existing gene set consisting of matrix encoding and related genes or “matrisome” (Naba et al., 2012), to further annotate the categories of ECM-associated genes. We identified 127 significantly up-regulated and 131 significantly down-regulated matrisome genes in GOF samples compared to controls. Of all significantly dysregulated genes in GOF samples, 9.1% were included in the matrisome gene list. Both up-regulated and down-regulated gene sets revealed expression changes in ECM Glycoproteins, Secreted factors, and ECM-regulators (**Figure 2B**). Of the 258 differentially expressed matrisome genes, 59 were annotated as ECM glycoproteins, 16 are Collagen proteins, 10 are Proteoglycans and 34 are ECM affiliated genes, while the

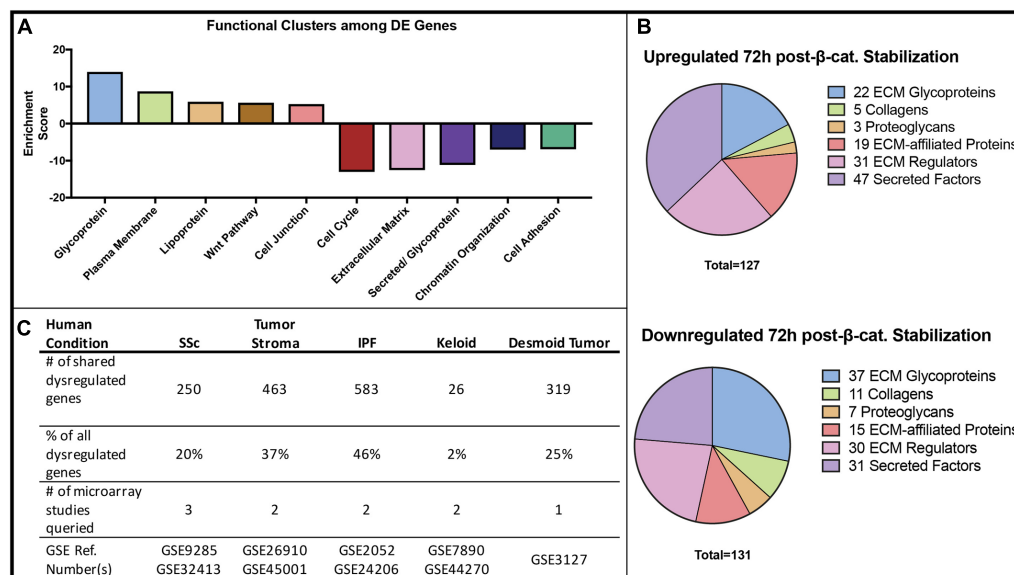


FIGURE 2 | Gene Ontology and disease comparison analysis of all β -catenin responsive genes. **(A)** The functional annotation (DAVID) of differentially expressed (DE) genes in β -catenin stabilized GOF. **(B)** A defined matrisome gene dataset (Naba et al., 2012) overlapped with differentially expressed genes in GOF yielded 127 up-regulated and 131 down-regulated matrisome genes consisting of ECM glycoproteins, secreted factors and ECM-regulators. **(C)** Differentially expressed genes in GOF dataset are also dysregulated in a variety of human fibrotic conditions.

rest serve ECM regulatory functions. Within the various functions of differentially expressed genes, the ECM-encoding and matrisome-related genes are of note due to their function in fibrogenesis.

We utilized Gene Set Enrichment Analysis (GSEA) to identify biological signatures enriched in GOF samples (Subramanian et al., 2005). We input 3114 mouse genes (adjusted P -value < 0.05) expressed in our control and GOF fibroblasts, 2371 of which corresponded to human genes with microarray identifiers in the GSEA database (**Supplementary Figure S3A**). The top four gene sets enriched in GOF samples were targets of *Suz12*, *EED*, and exhibited histone modification of H3K4Me2/H3K27Me3. They are all related to the function of the Polycomb Repressive Complex 2 (PRC2), a well-characterized epigenetic mechanism of gene repression (**Supplementary Figures S3B–F**). A recent study found functional links between lncRNAs, Wnt/ β -catenin signaling, and components of PRC2 in liver cancer stem cells (Zhu et al., 2016). The ontology analysis suggests that epigenetic mechanisms may serve as intermediate modulators of Wnt/ β -catenin signaling.

Finally, we compared the list of differentially expressed genes in our GOF mouse dermal fibroblast dataset to genes differentially expressed in human biopsied fibrotic tissue such as systemic sclerosis (SSc), desmoid tumors, idiopathic lung fibrosis, and tumor stroma (Hamburg-Shields et al., 2015). Of the genes differentially expressed in GOF mouse dermal fibroblasts compared to controls, we found varying percentages to also be dysregulated in human fibrotic conditions (**Figure 2C**). Our analysis shows that the expression changes in GOF dermal fibroblasts *in vitro* after 48 h of stabilization of β -catenin is consistent with findings from profiling *in vivo*

fibrotic mouse dermis in that Wnt/ β -catenin signaling may regulate the expression of matrisome genes and contribute to fibrotic conditions (Hamburg-Shields et al., 2015). Therefore, we demonstrate that the Wnt/ β -catenin signaling-responsive genes in dermal fibroblasts are also dysregulated in at least one type of human fibrotic tissue.

lncRNAs *Winc1* and *Winc2* Positively Respond to β -catenin Activity in a Tissue-Specific Manner

We next focused on the 111 differentially expressed lncRNAs in the GOF samples as compared to controls. Two top novel candidate Wnt Induced Non-Coding RNAs (*Winc1* and *Winc2*) were highly differentially expressed (**Figure 3A**). *Winc1* isoform 1 (ENSMUST00000146678.1) (*GM12603*) was expressed at 37 FPKM in control samples and at 551 FPKM in the GOF samples (average 14x fold-change). *Winc2* (*GM12606*) was expressed at 3.24 FPKM in control samples and 28.27 FPKM in GOF samples, marking over an 8x fold change between conditions. Therefore, we validated increase of *Winc1* with isoform 1 specific PCR primers and *Winc2* expression by qRT-PCR in four independent biological replicates of both GOF and controls in P4 ventral dermal fibroblasts *in vitro* (**Figures 3B,C**).

To investigate the tissue-specific expression profile of *Winc1* *in vivo*, we used embryonic skin in which we have previously shown a strong instructive role for Wnt/ β -catenin signaling in dermal fibroblast identity and hair follicle initiation (Atit et al., 2006; Ohtola et al., 2008; Fu et al., 2009; Tran et al., 2010; Chen et al., 2012). We assayed for *Winc1* *in vivo*

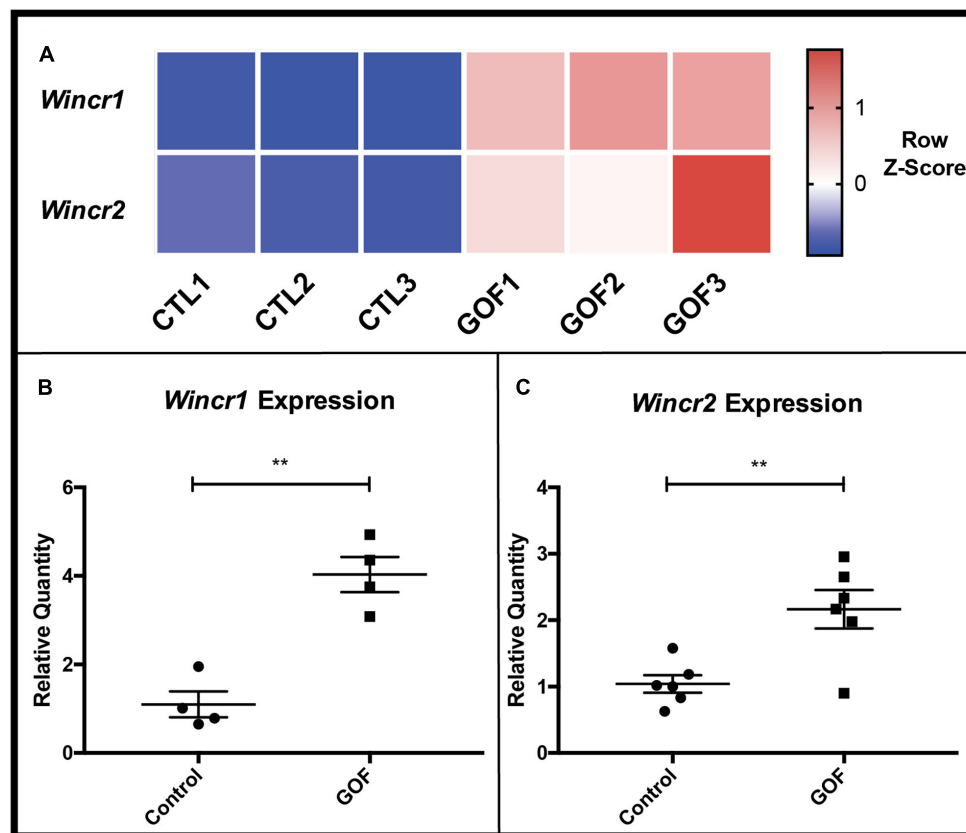


FIGURE 3 | β -catenin stabilized GOF have increased expression of *Winc1* and *Winc2*. **(A)** Heatmap representative of RNA-Seq FPKM values of two candidate lncRNAs, *Winc1* and *Winc2* from control and GOF dermal fibroblasts ($n = 3$). **(B,C)** Differential expression of lncRNA targets analyzed by qRT-PCR. lncRNA expression levels are significantly up-regulated within *in vitro* β -catenin stabilized GOF fibroblasts vs. the control (*Winc1*: $n = 4$, *Winc2*: $n = 6$). ** P -value ≤ 0.01 .

in embryonic cranial and dorsal dermis during hair follicle initiation at E13.5 by qRT-PCR. *Winc1* displayed a clear tissue-specific expression pattern during normal development in E13.5 mouse embryos. We found *Winc1* transcript to be abundant in the cranial (head) and dorsal skin, as compared to expression in the embryonic liver, heart, and gut ($n = 2$) (Figure 4A). *Winc1*'s enhanced presence in embryonic cranial and dorsal skin vs. other tissues suggests that our lncRNA candidate could play a role in fibroblast biology. *Winc2* followed a similar tissue-specific trend in embryonic tissues (Supplementary Figure S4A).

Winc1 Dynamically Responds to Wnt/ β -catenin Signaling Activity

We further tested β -catenin-responsive expression of *Winc1* with doxycycline-inducible stabilized β -catenin (β -cat^{istab}) in E13.5 *Engrailed1Cre/+*; *Rosa26rtTA/+*; *Teto-deltaN89 β -catenin/+* cranial and dorsal dermal fibroblasts *in vitro* (Mukherjee et al., 2010). Relative *Axin2* mRNA expression level was measured by qRT-PCR to confirm GOF status after doxycycline administration (Figure 4B). In response to β -cat^{istab}, *Winc1* was significantly higher in embryonic dorsal and cranial dermal fibroblasts (Figure 4C). In the same *in vitro* GOF

samples assayed for *Winc1* expression, an increase in *Winc2* expression was observed but did not reach statistical significance (Supplementary Figures S4B,C).

We tested for dynamic β -catenin-responsive expression of *Winc1* in doxycycline inducible-reversible levels of β -cat^{istab} *in vitro* in P4 ventral dermal fibroblasts. Addition of doxycycline and subsequent withdrawal in culture media allowed us to induce and reverse expression of stabilized β -catenin as shown by *Axin2* mRNA levels in *Engrailed1Cre/+*; *Rosa26rtTA/+*; *Teto-deltaN89 β -catenin/+* P4 dermal fibroblasts (Figure 4D). Similarly to the previous result, *Winc1* responded dynamically to the induction and reversal of β -cat^{istab} levels (Figure 4D), further establishing *Winc1* as a Wnt/ β -catenin signaling-responsive lncRNA in embryonic and perinatal dermal fibroblasts obtained from cranial and trunk skin.

Winc1 Influences Expression of Some Matrisome Genes

We next utilized co-expression network construction to identify and separate putative regulatory targets of both β -catenin and *Winc1*. It has been shown that correlation of expression between two genes across samples can imply a common regulatory factor (Allocco et al., 2004). Using FPKM values from control and

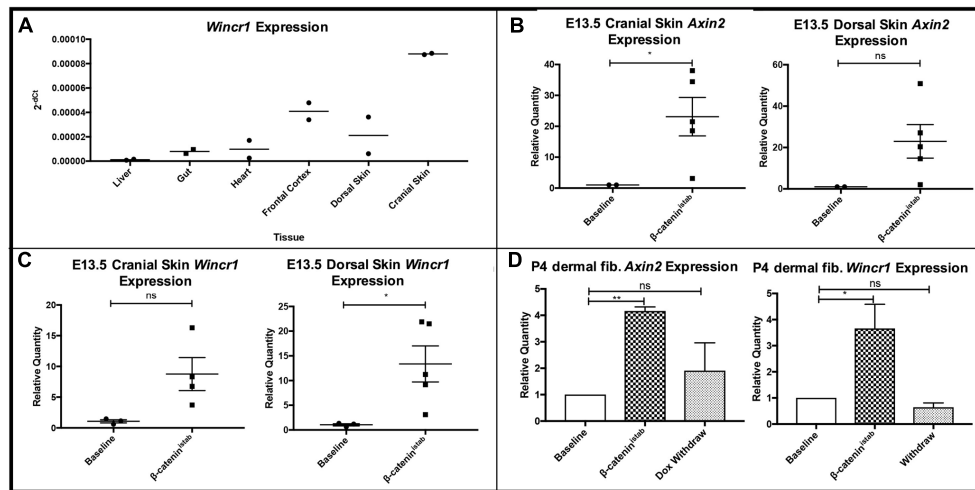


FIGURE 4 | *Wincrl* is differentially expressed in embryonic and perinatal mesenchyme and dynamically responds to changes in β -catenin activity levels. **(A)** Survey of *Wincrl* mRNA levels in embryonic (E13.5) tissues reveals highest expression in embryonic cranial and dorsal skin ($n = 2$). **(B)** *Axin2* expression levels were used to validate β -catenin activity following doxycycline induction for 4 days in *Engrailed1Cre/+; Rosa26rtTA/+; Teto- $\Delta N89$ (β -cat^{istab})* E13.5 primary fibroblasts *in vitro*. (P -value = 0.0237, $n = 2, 5$). **(C)** Relative quantity of *Axin2* and *Wincrl* mRNA in β -cat^{istab} E13.5 embryonic fibroblasts. **(D)** Relative quantity of *Axin2* and *Wincrl* mRNA in β -cat^{istab} P4 fibroblasts following doxycycline induction for 4 days and subsequent withdrawal for 2 days *in vitro*. * P -value ≤ 0.05 , ** P -value ≤ 0.01 .

GOF samples, we constructed a network of up-regulated mouse matrisome genes connected on the basis of correlation coefficient ($CC > 0.6$). Genes were grouped into two clusters: those that correlate with *Axin2* (97 genes), and those that do not correlate with *Axin2* (6 genes). By excluding genes whose expression correlated with that of *Axin2*, a known direct target of β -catenin, we identified a small number of matrisome genes that might be indirect targets of β -catenin regulation. Relevant fibrosis-related marker genes were selected as possessing some or all of the following traits:

- Up-regulated in β -cat^{istab} dermal fibroblasts by comparison of FPKM (Figure 1E)
- Inclusion in the Matrisome gene set (Figure 2B)
- Overexpressed in human fibrotic tissue microarray (Figure 2C)
- Exclusion from the *Axin2* cluster of the correlation network (Supplementary Figure S5)

Of the genes that passed all four criteria, *Matrix Metalloproteinase-10* (*Mmp10*) was selected for functional investigation. Furthermore, there was no significant enrichment of predicted *Tcf/Lef* binding motifs within 5 kb of the transcriptional start site of *Mmp10* and other differentially expressed mRNAs (Supplementary Figure S2).

To elucidate *Wincrl* function in gene regulation and dermal fibroblast biology, we either overexpressed or knocked down *Wincrl* in P4 ventral dermal fibroblasts *in vitro*. For overexpression, we generated a lentivirus construct to overexpress *Wincrl* (LV-*Wincrl*). For knockdowns of *Wincrl*, we utilized LNA GapmeRs (*Wincrl* GapmeRs). We confirmed increase in *Wincrl* RNA expression levels after infection of LV-*Wincrl* (Figure 5A) and reduction after transient transfection

with *Wincrl* GapmeR (Figure 5D). Dermal fibroblasts with either over- or knocked down expression of *Wincrl* and the parent control cells were then used to investigate changes in the expression of dermal identity and fibrosis-related gene candidates.

We found *Mmp10* mRNA level was highly correlated with levels of *Wincrl* in dermal fibroblasts (Figures 5A,B,E). After infection with LV-*Wincrl*, expression of *Mmp10* increased dramatically compared to the control (Figure 5B). Conversely, knockdown of *Wincrl* via GapmeRs resulted in a decrease in *Mmp10* expression by $\sim 30\%$ of basal levels (Figure 5E). Also, the expression of *Colla1*, a key fibrotic marker, correlated negatively with overexpression of *Wincrl* (Figure 5C). However, we did not observe a significant difference in *Colla1* mRNA levels upon knockdown of *Wincrl* (Figure 5F). The mRNA levels of markers for fibroblasts identity *Platelet derived growth factor receptor alpha* (*Pdgfra*) and β -catenin responsive genes and fibrotic markers such as *Loxl4*, *Col5a1*, which were up-regulated in GOF, did not correlate with *Wincrl* mRNA level, indicating that *Wincrl* does not regulate their expression in our system (Supplementary Figure S6). Co-expression analysis also identified *Has2* and *Methylthioadenosine phosphorylase* (*Mtap*), the latter of which shares a locus with *Wincrl*. However, the relative mRNA levels of *Has2* and *Mtap* did not correlate with changes in *Wincrl* level (Supplementary Figure S4). Thus, our findings demonstrate that *Mmp10* is a key target of *Wincrl*.

To test if *Wincrl* and β -catenin have synergistic effects on gene expression, we infected *Engrailed1Cre/+; R26rtTA/+; Teto- β -catenin* (β -cat^{istab}) P4 ventral dermal fibroblast with LV-*Wincrl*. We validated induction of *Axin2* in β -cat^{istab} condition and confirmed comparable overexpression of *Wincrl* in lentivirus infected β -cat^{istab} condition (Supplementary Figure S7). *Mmp10* and *Colla1* mRNA levels in LV-*Wincrl*+ β -cat^{istab} were not

significantly altered from LV-*Wincrl* only (Supplementary Figure S5). These results suggest that Wnt/ β -catenin signaling and *Wincrl* do not synergize to regulate *Mmp10* and *Col1a1* mRNA and likely function in a linear pathway.

***Wincrl* Mitigates Migration and Enhances Collagen Gel Contraction of Dermal Fibroblasts**

Given the gene regulatory effect of *Wincrl* on *Mmp10* expression, we sought to determine functional effects of *Wincrl* on cellular behavior of dermal fibroblasts. In order to prevent discrepancies, we used the same stably infected LV-*Wincrl* lines in all functional assays (characterized in Figure 5A). Two different dermal fibroblast lines with over-expression of *Wincrl* had comparable rates of proliferation to controls over 7 days (Supplementary Figure S8). Next, we used an *in vitro* scratch assay to study the rate of collective cell migration of dermal fibroblasts in a wound healing setting. This multistep process is relevant in many biological processes such as embryonic development and wound healing (Liang et al., 2007). Compared to control, LV-*Wincrl* had attenuated collective cell migration into the cleared area in both serum containing (Supplementary Figure S8) and serum-free conditions within 14–22 h (Figures 6A,B).

Collagen type1 gel embedded with fibroblasts is considered an *in vitro* model for fibrotic process and matrix turnover (Fang et al., 2004). We determined whether *Wincrl* affects the ability of dermal fibroblasts for collagen contraction, a key fibroblast function. We embedded control and LV-*Wincrl* cells in type I rat tail collagen for 48 h to increase the mechanical load and then released the gels. We found LV-*Wincrl* dermal fibroblasts had much higher capacity for processing collagen than controls (Figures 6C,D). The difference in collagen gel shrinkage by 14 h was significant and consistent across the two LV-*Wincrl* cell lines. Thus, gain of function *Wincrl* positively correlated with collagen gel contraction, thereby linking *Wincrl* to a key fibroblast functionality.

In summary, we have identified key mRNAs and lncRNAs that are responsive to Wnt/ β -catenin signaling levels and we have demonstrated that a lncRNA responsive to Wnt/ β -catenin activity, *Wincrl*, is a potentially new regulator of dermal fibroblast behavior and ECM-related gene expression.

DISCUSSION

Long non-coding RNAs have emerged as key regulators of many cellular processes, and their dysregulation has been observed in many human diseases (Wan and Wang, 2014). Studies of novel lncRNAs have led to the discovery of new mechanisms of gene regulation paving the way toward novel therapeutic strategies. In our current study, through the genetic stabilization of β -catenin in primary dermal fibroblasts, we identified downstream lncRNAs and mRNAs responsive to Wnt/ β -catenin signaling. Our genetically targeted culture systems enabled us to observe the effects of β -catenin's activity levels on gene expression in dermal fibroblasts without the paracrine influences

of other cell types that would be present in whole skin expression analysis. Utilizing this system, we have identified Wnt/ β -catenin signaling-induced lncRNAs such as *Wincrl*, that act as putative regulators of key genes such as *Mmp10* and *Col1a1* in ECM biology. *Wincrl* also affects complex cellular behaviors such as collective cell migration and collagen processing and contraction.

Our expression profiling study allowed us to identify genes that were under the influence of Wnt/ β -catenin signaling in dermal fibroblasts. Consistent with our previous gene expression analysis from mouse fibrotic dermis after 21 days of sustained Wnt/ β -catenin signaling (Hamburg-Shields et al., 2015), gene ontology analysis through functional clustering shows a strong enrichment of ECM and ECM-regulatory gene expression within 72 h of Wnt/ β -catenin signaling in dermal fibroblasts. Our analysis also implies that downstream targets of Wnt/ β -catenin signaling may alter matrix deposition, remodeling, and fibroblast behavior. Comparison of the Wnt/ β -catenin signaling activation gene signature in dermal fibroblasts with those dysregulated in various contexts of human fibrotic diseases confirms that the gene expression signature of β -catenin stabilization in our model is indeed relevant to human fibrotic disease. Thus, β -catenin stabilization in dermal fibroblasts is a relevant signaling pathway in matrix construction and remodeling and a relevant node for therapeutic intervention.

Analysis of the gene expression signature induced by β -catenin stabilization reveals enrichment for PRC2 targets, signifying epigenetic regulation. Unbiased gene set enrichment analysis of genes expressed in dermal fibroblasts shows an enrichment of PRC2 targets. This epigenetic regulatory repressive complex has been shown to interact with lncRNAs, suggesting that a portion of the β -catenin-dependent expression signature controlling ECM and fibrosis-related genes may be regulated by PRC2 or other mechanisms (Rinn and Chang, 2012; Wan and Wang, 2014). While further studies are needed to elucidate the mechanism of β -catenin's influence on fibrotic genes, these analyses of the coding gene changes following β -catenin activation guided our identification of novel lncRNAs as regulators of gene expression and fibroblast behavior.

Using sustained and inducible-reversible systems of Wnt/ β -catenin signaling activation, *Wincrl* was identified as a Wnt/ β -catenin signaling induced lncRNA candidate due to its consistent response to β -catenin activity levels in multiple contexts in mouse dermal fibroblasts. Further context-specific functional roles of *Wincrl* will be elucidated in future *in vivo* studies with fibroblast-restricted mutants. Our *in vitro* modulations of *Wincrl* expression levels in dermal fibroblasts demonstrated that this lncRNA has a robust gene regulatory effect on distinct genes. Identification of other *Wincrl* targets will require extensive expression profiling in dermal fibroblasts and other cell types. In addition, our studies reveal that β -catenin stabilization and overexpression of *Wincrl* do not synergize to increase *Mmp10* levels, suggesting that β -catenin and *Wincrl* may function linearly in an uncharacterized genetic or epigenetic pathway to modulate key genes in matrix production and turnover. Our current data show that while *Wincrl* levels

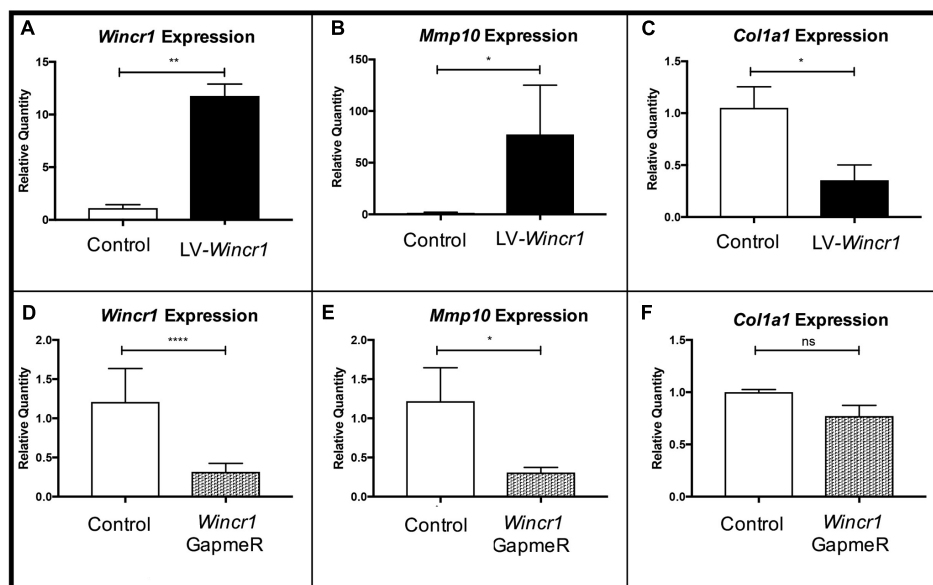


FIGURE 5 | Gene regulatory function of *Winc1*. **(A,D)** Relative quantity of *Winc1* RNA is significantly altered as a result of lentivirus or GapmeR infection. **(B,E)** Relative quantity of *Mmp10* mRNA levels correlate with changes in *Winc1* expression levels. **(C,F)** In LV-*Winc1*, relative quantity of *Col1a1* is significantly reduced and was comparable between control and *Winc1* GapmeR ($n = 5$ biological replicates). * P -value ≤ 0.05 , ** P -value ≤ 0.01 , **** P -value ≤ 0.0001 .

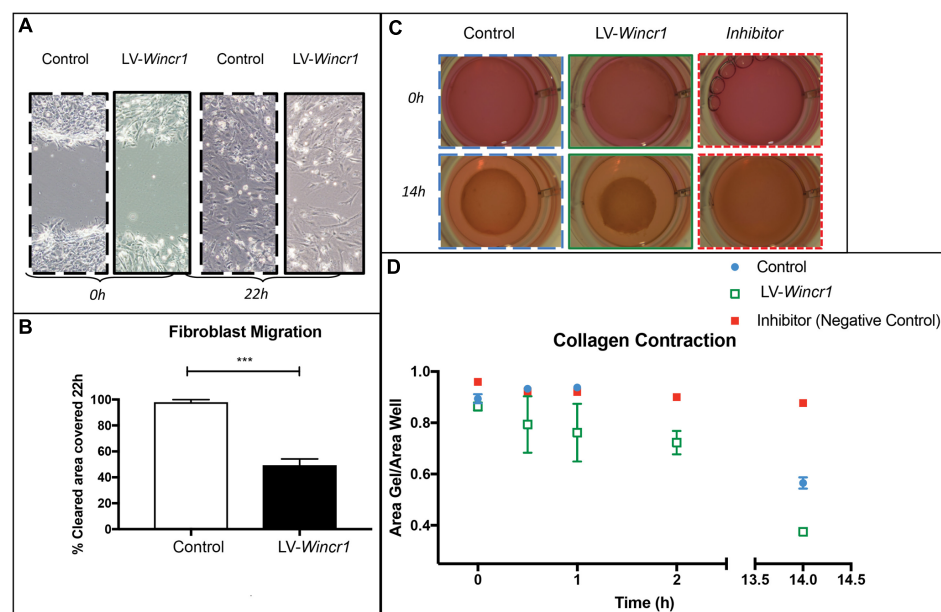


FIGURE 6 | *Winc1* function in collective cell migration and fibrotic response. **(A)** Dermal fibroblasts with LV-*Winc1* exhibit decreased migration compared to uninfected controls. **(B)** Migration is significantly diminished after 22 h following a scratch in serum-free media ($n = 3$) and representative of three separate experiments. **(C,D)** Type I collagen gel contraction assay demonstrates LV-*Winc1* fibroblasts have a higher ability to process collagen. The difference in collagen shrinkage is visible and consistent across three technical replicates from two independent cell lines by 14 h. Points represent the quantitation of collagen gel area normalized to plate area after 14 h using ImageJ software ($n = 2$ between Control and LV-*Winc1* at 14 h). *** P -value ≤ 0.001 .

are increased after stabilizing β -catenin, *Winc1* can also independently regulate expression of protein coding genes. Identifying other *Winc1* targets will allow us to further refine our understanding of how *Winc1* intersects with Wnt/ β -catenin

signaling for gene regulation and fibroblast cell behavior. In our current study, we elaborate on the role of *Winc1* in fibroblast gene regulation and cellular behavior, given that this is a context in which Wnt/ β -catenin signaling is known to be important.

Our query of several fibrotic genes suggests that *Wincrl* has a restricted gene regulatory role. Specifically, *Mmp10* expression was consistently responsive to *Wincrl* levels, suggesting it is a putative regulatory target of this novel lncRNA. *Wincrl*'s gene regulatory role does not appear to be pan-ECM, but instead can regulate a defined set of genes. Further mechanistic-based experiments are required to gain insight into how *Wincrl* regulates *Mmp10*, and such studies would be helpful in predicting other targets through binding motifs or other means. *Mmp10/Stromelysin-2* has emerged as a key player in fibrosis and “degradomics” (Overall and Kleinfeld, 2006; Schlage et al., 2015; Sokai et al., 2015). MMPs can promote collagen contraction by increasing turnover (Daniels et al., 2003; Bildt et al., 2009), and we found increased contractile behavior in LV-*Wincrl* dermal fibroblasts. *Mmp10* is not a collagenase, but it has the ability to regulate expression of other MMPs, such as *Mmp13*, that have collagenolytic function and promote the resolution of matrix in wound healing. It is not clear if *Mmp10* has a similar role in mitigating ECM accumulation in chronic fibrosis settings that are the result of excessive ECM production and inadequate resolution (McKleroy et al., 2013; Rohani et al., 2015). It is tempting to speculate that *Wincrl* participates as a negative feedback to the pro-fibrotic Wnt/ β -catenin pathway in dermal fibroblasts. Future functional studies with *Wincrl* in fibrosis models will be needed to demonstrate if it has a fibroprotective function.

Wnt/ β -catenin signaling has diverse roles in skin development, tissue homeostasis, and disease. Thus, identifying new genetic and epigenetic regulatory mechanisms will provide novel insight in our understanding of this important pathway in dermal fibroblasts. Discovery of this lncRNA and its preliminary regulatory links to β -catenin signaling, skin development, and known fibrotic protein-coding gene expression indicate that lncRNAs may play an important regulatory role in directing Wnt/ β -catenin signaling. This is of particular interest in the context of fibrosis, a condition that can be caused by activation of this pathway, but with gene regulatory intermediates that will require further characterization in order to be of clinical value. The robust responsiveness of *Wincrl* to Wnt/ β -catenin signaling indicates the possibility of this and other lncRNAs as targets for therapeutic intervention to treat fibrosis, but also as circulating diagnostic biomarkers (Tang et al., 2015; Vencken et al., 2015; Zhou et al., 2015). Further studies are needed to elucidate the gene regulatory mechanism and *in vivo* functions of *Wincrl* and other β -catenin responsive lncRNAs. Also, mechanistic understanding into *Wincrl* may lead to novel insights into the Wnt signaling pathway and how it regulates key genetic networks throughout embryonic development and adult diseases.

AUTHOR CONTRIBUTIONS

NKM, NVM, RA, and AK contributed to experimental design, experiment collection, analysis, writing, and figure preparation. EH-S contributed to experimental

design and writing of the manuscript. BI contributed to analysis.

FUNDING

This research was partially supported by the Case Western Reserve University Skin Diseases Research Center Pilot and Feasibility Program (NIH P30 AR039750). This work was supported by the following grants: NIH-NIDCR-R01DE01870 (RA), Global Fibrosis Foundation (RA), NIBIB:1P41EB021911-01 (AK), NIH-NIA F30 AG045009 (EH-S), GAANN Fellowship (BI), Case Western Reserve University ENGAGE (NVM) and SOURCE Programs (NKM).

ACKNOWLEDGMENTS

We thank to past and current members of the Atit laboratory for excellent discussion and advice. Thanks to Gregg DiNuoscio and Anna Jussila for technical support. Thanks to the Case Genomics Core and Bioinformatics Core Services (Dr. Ricky Chan).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2017.00183/full#supplementary-material>

FIGURE S1 | Heatmaps of all lncRNAs and mRNAs expressed above 1FPKM. Biological replicates cluster together on the basis of global expression for both (A) lncRNA and (B) mRNAs by Pearson correlation (based on entities and samples).

FIGURE S2 | Enrichment of predicted transcription factor binding site (TFBS) analysis on promoters of mRNA. Enrichment of TFBS within 5 Kb of the transcriptional start site (TSS) of up-regulated (A) and differentially expressed (P -value = 0.05, fold change > 2) (B) mRNAs after activating Wnt/ β -catenin signaling. Statistically enriched TFBS (upper right quadrant) are not in the TCF/LEF family of transcription factors associated with Wnt signaling pathway (red color in A).

FIGURE S3 | Gene set enrichment analysis (GSEA) shows strong enrichment in Polycomb Repressive Complex 2 (PRC2) targets among genes up-regulated after β -catenin stabilization. (A–E) Top enriched gene sets from C2 database are Suz12, EED, and PRC2 targets, as well as H3 methylation sites. Gene sets enriched in the β -catenin stabilized GOF condition were plotted based on FWER P -value and Normalized Enrichment Score. (F) Leading edge analysis of the top four gene sets show common genes driving the enrichment of these signatures in GOF dermal fibroblasts.

FIGURE S4 | LncRNA *Wincrl* is differentially expressed in response to β -catenin activity at E13.5. (A) Endogenous expression of *Wincrl* within distinct tissues from an E13.5 wild type embryo via qRT-PCR. Steady state mRNA expression level was verified across two independent embryonic litters. (B) Relative quantity of *Wincrl* in E13.5 β -cat^{stab} cultured cranial and dorsal dermal fibroblasts after 4 days of β -catenin stabilization. (C) *Wincrl* expression is not significantly altered after inducing β -catenin stabilization.

FIGURE S5 | Co-expression network of genes up-regulated in β -catenin GOF and included in mouse Matrisome. Network includes all genes up-regulated (P -value < 0.05, fold change > 2) in GOF samples and included in the mouse Matrisome. Network edges indicate expression correlation between genes, with

correlation coefficient above 0.6 shown. Network is separated into clusters based on correlation with *Axin2* expression, or absence of such correlation (to the right).

FIGURE S6 | Manipulation of *Winc1* expression does not affect all pro-fibrotic genes. **(A–G)** Expression of additional matrisome genes as a result of *Winc1* overexpression. We identified these mRNA targets from the analysis of our dataset, co-expression network generation, known fibroblast identity markers, matrisome gene targets, and literature screens.

FIGURE S7 | Lack of synergistic expression of *Mmp10* in cells with LV-*Winc1* and LV-*Winc1* + β cat^{istab}. **(A)** *Axin2* mRNA levels, a measure of β -catenin activation, is significantly higher in β cat^{istab} condition (P -value = 0.0152, n = 2). **(B)** *Winc1*

level is significantly higher in LV-*Winc1* and LV-*Winc1* + β cat^{istab} samples (n = 2). **(C)** There is no significant difference in relative quantity of *Mmp10* and *Col1a1* mRNA between LV-*Winc1* and LV-*Winc1* + β cat^{istab} samples, (n = 2). **(C,D)** Relative quantity of *Col1a1* mRNA is significantly reduced in the presence of LV-*Winc1* (P -value = 0.0023, n = 2). Representative of three different experiments. * P -value \leq 0.05, ** P -value \leq 0.01.

FIGURE S8 | Overexpression of *Winc1* does not affect fibroblast proliferation.

(A) Proliferation curve of control and LV-*Winc1* infected primary dermal fibroblasts showing comparable rate of proliferation (n = 3) and representative of two different experiments. **(B)** Migration is significantly diminished after 15 h following a scratch in the monolayer of LV-*Winc1* cells in serum-free media (n = 3) and representative of two separate experiments.

REFERENCES

- Akhmetshina, A., Palumbo, K., Dees, C., Bergmann, C., Venalis, P., Zerr, P., et al. (2012). Activation of canonical Wnt signalling is required for TGF- β -mediated fibrosis. *Nat. Commun.* 3:735. doi: 10.1038/ncomms1734
- Allocco, D. J., Kohane, I. S., and Butte, A. J. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* 5:18. doi: 10.1186/1471-2105-5-18
- Atit, R., Sgaier, S. K., Mohamed, O. A., Taketo, M. M., Dufort, D., Joyner, A. L., et al. (2006). β -catenin activation is necessary and sufficient to specify the dorsal dermal fate in the mouse. *Dev. Biol.* 296, 164–176. doi: 10.1016/j.ydbio.2006.04.449
- Belteki, G., Haigh, J., Kabacs, N., Haigh, K., Sison, K., Costantini, F., et al. (2005). Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res.* 33, 1–10. doi: 10.1093/nar/gni051
- Bildt, M. M., Bloemen, M., Kuijpers-Jagtman, A. M., and Von den Hoff, J. W. (2009). Matrix metalloproteinase inhibitors reduce collagen gel contraction and alpha-smooth muscle actin expression by periodontal ligament cells. *J. Periodontol. Res.* 44, 266–274. doi: 10.1111/j.1600-0765.2008.01127.x
- Budnick, I., Hamburg-Shields, E., Chen, D., Torre, E., Jarrell, A., Akhtar-Zaidi, B., et al. (2016). Defining the identity of mouse embryonic dermal fibroblasts. *Genesis* 54, 415–430. doi: 10.1002/dvg.22952
- Chang, H. Y., Chi, J.-T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., et al. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12877–12882. doi: 10.1073/pnas.162488599
- Chen, D., Jarrell, A., Guo, C., Lang, R., and Atit, R. (2012). Dermal β -catenin activity in response to epidermal Wnt ligands is required for fibroblast proliferation and hair follicle initiation. *Development* 139, 1522–1533. doi: 10.1242/dev.076463
- Daniels, J. T., Cambrey, A. D., Ocleston, N. L., Garrett, Q., Tarnuzzer, R. W., Schultz, G. S., et al. (2003). Matrix metalloproteinase inhibition modulates fibroblast-mediated matrix contraction and collagen production in vitro. *Invest. Ophthalmol. Vis. Sci.* 44, 1104–1110. doi: 10.1167/iops.02-0412
- Eames, B. F., and Schneider, R. A. (2005). Quail-duck chimeras reveal spatiotemporal plasticity in molecular and histogenic programs of cranial feather development. *Development* 132, 1499–1509. doi: 10.1242/dev.01719
- Enzo, M. V., Rastrelli, M., Rossi, C. R., Hladnik, U., and Segat, D. (2015). The Wnt/ β -catenin pathway in human fibrotic-like diseases and its eligibility as a therapeutic target. *Mol. Cell. Ther.* 3:1. doi: 10.1186/s40591-015-0038-2
- Fan, Y., Shen, B., Tan, M., Mu, X., Qin, Y., Zhang, F., et al. (2014). Long non-coding RNA UCA1 increases chemoresistance of bladder cancer cells by regulating Wnt signaling. *FEBS J.* 281, 1750–1758. doi: 10.1111/febs.12737
- Fang, Q., Liu, X., Abe, S., Kobayashi, T., Wang, X. Q., Kohyama, T., et al. (2004). Thrombin induces collagen gel contraction partially through PAR1 activation and PKC- ϵ . *Eur. Respir. J.* 24, 918–924. doi: 10.1183/09031936.04.00005704
- Fu, J., Jiang, M., Mirando, A. J., Yu, H. M., and Hsu, W. (2009). Reciprocal regulation of Wnt and Gpr177/mouse Wntless is required for embryonic axis formation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18598–18603. doi: 10.1073/pnas.0904894106
- Fu, X.-L., Liu, D.-J., Yan, T.-T., Yang, J.-Y., Yang, M.-W., Li, J., et al. (2016). Analysis of long non-coding RNA expression profiles in pancreatic ductal adenocarcinoma. *Sci. Rep.* 6:33535. doi: 10.1038/srep33535
- Hamburg, E. J., and Atit, R. P. (2012). Sustained Beta-Catenin activity in dermal fibroblasts is sufficient for skin fibrosis. *J. Invest. Dermatol.* 132, 2469–2472. doi: 10.1038/jid.2012.155
- Hamburg-Shields, E., DiNuoscio, G. J., Mullin, N. K., Lafayatis, R., and Atit, R. P. (2015). Sustained β -catenin activity in dermal fibroblasts promotes fibrosis by up-regulating expression of extracellular matrix protein-coding genes. *J. Pathol.* 235, 686–697. doi: 10.1002/path.4481
- Harada, N., Tamai, Y., Ishikawa, T.-O., Sauer, B., Takaku, K., Oshima, M., et al. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J.* 18, 5931–5942. doi: 10.1093/emboj/18.21.5931
- He, Z., Feng, L., Zhang, X., Geng, Y., Parodi, D. A., Suarez-Quian, C., et al. (2005). Expression of Colla1, Colla2 and procollagen I in germ cells of immature and adult mouse testis. *Reproduction* 130, 333–341. doi: 10.1530/rep.1.00694
- Ho Sui, S. J., Fulton, D. L., Arenillas, D. J., Kwon, A. T., and Wasserman, W. W. (2007). OPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res.* 35(Suppl. 2), W245–W252. doi: 10.1093/nar/gkm427
- Huang, C., Yang, Y., and Liu, L. (2015). Interaction of long noncoding RNAs and microRNAs in the pathogenesis of idiopathic pulmonary fibrosis. *Physiol. Genomics* 47, 463–469. doi: 10.1152/physiolgenomics.00064.2015
- Jho, E.-H., Zhang, T., Domon, C., Joo, C.-K., Freund, J.-N., and Costantini, F. (2002). Wnt/ β -catenin/Tcf signaling induces the transcription of *Axin2*, a negative regulator of the signaling pathway. *Mol. Cell. Biol.* 22, 1172–1183. doi: 10.1128/MCB.22.4.1172
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., 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
- Kimmel, R. A., Turnbull, D. H., Blanquet, V., Wurst, W., Loomis, C. A., and Joyner, A. L. (2000). Two lineage boundaries coordinate vertebrate apical ectodermal ridge formation. *Genes Dev.* 14, 1377–1389. doi: 10.1101/gad.14.11.1377
- Kwon, A. T., Arenillas, D. J., Worsley Hunt, R., and Wasserman, W. W. (2012). oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda)* 2, 987–1002. doi: 10.1534/g3.112.003202
- Lam, A. P., and Gottardi, C. J. (2011). β -catenin signaling. *Curr. Opin. Rheumatol.* 23, 562–567. doi: 10.1097/BOR.0b013e32834b3309
- Liang, C.-C., Park, A. Y., and Guan, J.-L. (2007). In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* 2, 329–333. doi: 10.1038/nprot.2007.30
- Lim, X., and Nusse, R. (2012). Wnt signaling in skin development, homeostasis, and disease. *Cold Spring Harb. Perspect. Biol.* 5:a008029. doi: 10.1101/cshperspect.a008029
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Ma, Y., Yang, Y., Wang, F., Moyer, M.-P., Wei, Q., Zhang, P., et al. (2016). Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/ β -catenin signalling pathway via suppression of activator protein 2 α . *Gut* 65, 1494–1504. doi: 10.1136/gutjnl-2014-308392

- Mastrogiannaki, M., Lichtenberger, B. M., Reimer, A., Collins, C. A., Driskell, R. R., and Watt, F. M. (2016). β -Catenin stabilization in skin fibroblasts causes fibrotic lesions by preventing adipocyte differentiation of the reticular dermis. *J. Invest. Dermatol.* 136, 1130–1142. doi: 10.1016/j.jid.2016.01.036
- McCrea, P. D., and Gottardi, C. J. (2016). Beyond β -catenin: prospects for a larger catenin network in the nucleus. *Nat. Rev. Mol. Cell Biol.* 17, 55–64. doi: 10.1038/nrm.2015.3
- McKleroy, W., Lee, T.-H., and Atabai, K. (2013). Always cleave up your mess: targeting collagen degradation to treat tissue fibrosis. *Am. J. Physiol. Lung. Cell Mol. Physiol.* 304, L709–L721. doi: 10.1152/ajplung.00418.2012
- Micheletti, R., Plaisance, I., Abraham, B. J., Sarre, A., Ting, C.-C., Alexanian, M., et al. (2017). The long noncoding RNA Wisper controls cardiac fibrosis and remodeling. *Sci. Transl. Med.* 9:eaa19118. doi: 10.1126/scitranslmed.aai9118
- Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately down-regulated in human diabetes. *Nat. Genet.* 34, 267–273. doi: 10.1038/ng1180
- Moran, V. A., Niland, C. N., and Khalil, A. M. (2012). Co-Immunoprecipitation of long noncoding RNAs. *Methods Mol. Biol.* 925, 219–228. doi: 10.1007/978-1-62703-011-3_15
- Mukherjee, A., Soyak, S. M., Li, J., Ying, Y., Szwarc, M. M., He, B., et al. (2010). A mouse transgenic approach to induce β -catenin signaling in a temporally controlled manner. *Transgenic Res.* 20, 827–840. doi: 10.1007/s11248-010-9466-6
- Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A., and Hynes, R. O. (2012). The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell Proteomics* 11:M111.014647. doi: 10.1074/mcp.M111.014647
- Nakamura, Y., de Paiva Alves, E., Veenstra, G. J., and Hoppler, S. (2009). Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β -catenin recruitment. *Bioinformatics* 143, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Nusse, R., and Clevers, H. (2017). Wnt/ β -Catenin signaling, disease, and emerging therapeutic modalities. *Cell* 169, 985–999. doi: 10.1016/j.cell.2017.05.016
- Ohtola, J., Myers, J., Akhtar-Zaidi, B., Zuzindlak, D., Sandesara, P., Yeh, K., et al. (2008). β -Catenin has sequential roles in the survival and specification of ventral dermis. *Development* 135, 2321–2329. doi: 10.1242/dev.021170
- Overall, C. M., and Kleinfeld, O. (2006). Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* 6, 227–239. doi: 10.1038/nrc1821
- Piccoli, M.-T., Gupta, S. K., Viereck, J., Foinquinos, A., Samolovac, S., Kramer, F. L., et al. (2017). Inhibition of the cardiac fibroblast-enriched lncRNA Meg3 prevents cardiac fibrosis and diastolic dysfunction novelty and significance. *Circ. Res.* 121, 575–583. doi: 10.1161/CIRCRESAHA.117.310624
- Qu, X., Du, Y., Shu, Y., Gao, M., Sun, F., Luo, S., et al. (2017). MIAT Is a Pro-fibrotic long non-coding RNA governing cardiac fibrosis in post-infarct myocardium. *Sci. Rep.* 7:42657. doi: 10.1038/srep42657
- Rendl, M., Lewis, L., and Fuchs, E. (2005). Molecular dissection of mesenchymal-epithelial interactions in the hair follicle. *PLOS Biol.* 3:e331. doi: 10.1371/journal.pbio.0030331
- Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *arXiv* 81, 145–166. doi: 10.1146/annurev-biochem-051410-092902
- Rohani, M. G., McMahan, R. S., Razumova, M. V., Hertz, A. L., Cieslewicz, M., Pun, S. H., et al. (2015). MMP-10 regulates collagenolytic activity of alternatively activated resident macrophages. *J. Invest. Dermatol.* 135, 2377–2384. doi: 10.1038/jid.2015.167
- Schlage, P., Kockmann, T., Sabino, F., Kizhakkedathu, J. N., and Auf dem Keller, U. (2015). Matrix metalloproteinase 10 degradomics in keratinocytes and epidermal tissue identifies bioactive substrates with pleiotropic functions. *Mol. Cell. Proteomics* 14, 3234–3246. doi: 10.1074/mcp.M115.053520
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Schuijers, J., Mokry, M., Hatzis, P., Cuppen, E., and Clevers, H. (2014). Wnt-induced transcriptional activation is exclusively mediated by TCF/LEF. *EMBO J.* 33, 146–156. doi: 10.1002/embj.201385358
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sokai, A., Handa, T., Tanizawa, K., Oga, T., Uno, K., Tsuruyama, T., et al. (2015). Matrix metalloproteinase-10: a novel biomarker for idiopathic pulmonary fibrosis. *Respir. Res.* 16, 120. doi: 10.1186/s12931-015-0280-9
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Tang, J., Jiang, R., Deng, L., Zhang, X., Wang, K., and Sun, B. (2015). Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. *Oncotarget* 6, 4505–4515. doi: 10.18632/oncotarget.2934
- Tran, T. H., Jarrell, A., Zentner, G. E., Welsh, A., Brownell, I., Scacheri, P. C., et al. (2010). Role of canonical Wnt signaling/ β -catenin via Dermo1 in cranial dermal cell development. *Development* 137, 3973–3984. doi: 10.1242/dev.056473
- van Amerongen, R. E. E., and Nusse, R. (2009). Towards an integrated view of Wnt signaling in development. *Development* 136, 3205–3214. doi: 10.1242/dev.033910
- Vassallo, I., Zinn, P., Lai, M., Rajakannu, P., Hamou, M.-F., and Hegi, M. E. (2015). WIF1 re-expression in glioblastoma inhibits migration through attenuation of non-canonical WNT signaling by downregulating the lncRNA MALAT1. *Oncogene* 35, 12–21. doi: 10.1038/nc.2015.61
- Vencken, S. F., Greene, C. M., and Mckiernan, P. J. (2015). Non-coding RNA as lung disease biomarkers. *Thorax* 70, 501–503. doi: 10.1136/thoraxjnl-2014-206193
- Wan, D. C., and Wang, K. C. (2014). Long Noncoding RNA: significance and potential in skin biology. *Cold Spring Harb. Perspect. Med.* 4:a015404. doi: 10.1101/cshperspect.a015404
- Wang, X., Reid Sutton, V., Omar Peraza-Llanes, J., Yu, Z., Rosetta, R., Kou, Y. C., et al. (2007). Mutations in X-linked PORCN, a putative regulator of Wnt signaling, cause focal dermal hypoplasia. *Nat. Genet.* 39, 836–838. doi: 10.1038/ng2057
- Wang, Y., Chen, R., and Correspondence, Z. F. (2015). The Long Noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. *Stem Cell* 16, 413–425. doi: 10.1016/j.stem.2015.03.003
- Wang, Z., Jinnin, M., Nakamura, K., Harada, M., Kudo, H., Nakayama, W., et al. (2016). Long non-coding RNA TSIX is up-regulated in scleroderma dermal fibroblasts and controls collagen mRNA stabilization. *Exp. Dermatol.* 25, 131–136. doi: 10.1111/exd.12900
- Zhou, X., Yin, C., Dang, Y., Ye, F., and Zhang, G. (2015). Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci. Rep.* 5:11516. doi: 10.1038/srep11516
- Zhu, P., Wang, Y., Huang, G., Ye, B., Liu, B., Wu, J., et al. (2016). lnc- β -Catm elicits EZH2-dependent β -catenin stabilization and sustains liver CSC self-renewal. *Nat. Struct. Mol. Biol.* 23, 631–639. doi: 10.1038/nsmb.3235

Conflict of Interest Statement: 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.

Copyright © 2017 Mullin, Mallipeddi, Hamburg-Shields, Ibarra, Khalil and Atit. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Distinct and Modular Organization of Protein Interacting Sites in Long Non-coding RNAs

Saakshi Jalali^{1,2*}, Shrey Gandhi¹ and Vinod Scaria^{1,2*}

¹ GN Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology, New Delhi, India, ² CSIR Institute of Genomics and Integrative Biology, Academy of Scientific and Innovative Research, New Delhi, India

Background: Long non-coding RNAs (lncRNAs), are being reported to be extensively involved in diverse regulatory roles and have exhibited numerous disease associations. LncRNAs modulate their function through interaction with other biomolecules in the cell including DNA, RNA, and proteins. The availability of genome-scale experimental datasets of RNA binding proteins (RBP) motivated us to understand the role of lncRNAs in terms of its interactions with these proteins. In the current report, we demonstrate a comprehensive study of interactions between RBP and lncRNAs at a transcriptome scale through extensive analysis of the crosslinking and immunoprecipitation (CLIP) experimental datasets available for 70 RNA binding proteins.

Results: Our analysis suggests that density of interaction sites for these proteins was significantly higher for specific sub-classes of lncRNAs when compared to protein-coding transcripts. We also observe a positional preference of these RBPs across lncRNA and protein coding transcripts in addition to a significant co-occurrence of RBPs having similar functions, suggesting a modular organization of these elements across lncRNAs.

Conclusion: The significant enrichment of RBP sites across some lncRNA classes is suggestive that these interactions might be important in understanding the functional role of lncRNA. We observed a significant enrichment of RBPs which are involved in functional roles such as silencing, splicing, mRNA processing, and transport, indicating the potential participation of lncRNAs in such processes.

Keywords: long non-coding RNAs, RNA binding proteins, protein-lncRNA interactions, Argonaute (ago), MALAT1

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Keith W. Vance,
University of Bath, United Kingdom
Nobuyoshi Akimitsu,
The University of Tokyo, Japan

*Correspondence:

Saakshi Jalali
saakshi.jalali@gmail.com
Vinod Scaria
vinods@igib.in

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 27 September 2017

Accepted: 14 March 2018

Published: 04 April 2018

Citation:

Jalali S, Gandhi S and Scaria V (2018)
Distinct and Modular Organization of
Protein Interacting Sites in Long
Non-coding RNAs.
Front. Mol. Biosci. 5:27.
doi: 10.3389/fmolb.2018.00027

BACKGROUND

The recent years have seen the discovery of a large number of novel transcripts which belong to the long non-coding RNA (lncRNA) class in humans and other model organisms (Pauli et al., 2012). This has been largely contributed by the availability of high-throughput methodologies for transcriptome annotation, including tiling microarrays (Hafner et al., 2010a; Furey, 2012) and deep sequencing (Roberts et al., 2011). The recent genome-wide analyzes of lncRNA genes in Humans have annotated over 83,215 transcripts from 32,446 lncRNAs genes (Derrien et al., 2012; Harrow et al., 2012). The lncRNA superset presently includes a number of sub-classes which include 3 prime overlapping ncRNA, antisense, bidirectional promoter lncRNA, lincRNA, macro lncRNA, miscRNA, non-coding, processed transcripts, pseudogene, retained intron, sense intronic, sense overlapping, and TEC. By definition lncRNAs encompass all transcripts > 200 nucleotides in length and no ORF coding for more than 30 amino acids (Mercer et al., 2009). The biogenesis

and regulation of lncRNAs have not been studied in great detail, though it is believed that they are transcribed majorly by Polymerase II and are capped and polyadenylated (Goodrich and Kugel, 2006; Gibb et al., 2011). One particular class of lncRNAs, the large intergenic non-coding RNA has been primarily discovered through their association with epigenetic marks in the genome (Cabili et al., 2011; Cao, 2014). We have recently shown extensive similarities and specific dissimilarities in epigenetic regulation of lncRNAs in comparison to protein-coding genes (Sati et al., 2012). The precise biological function of many of the lncRNAs are not known, though a handful of the candidates have been recently shown to be mechanistically involved in gene regulation and associated with diseases (Wapinski and Chang, 2011). Recent reports from our group also suggest processing of a subset of lncRNAs to smaller RNAs (Jalali et al., 2012), and that a subset of lncRNAs could be potentially targeted by microRNAs (Jalali et al., 2013), thus constituting an intricate and yet poorly understood network of non-coding RNA mediated regulation.

Mechanistically, the characterization of lncRNA could be generalized as a function of its interactions with other biomolecules in the cell: DNA, RNA, protein, and small-molecules (Bhartiya et al., 2012). Current studies have showed that molecular and computational biology techniques can act as catalyst in discovering lncRNA-mediated regulation via understanding their interactions with different biomolecules (Jalali et al., 2015). Recent reports have also suggested the possibility of protein-lncRNA interactions and regulatory interactions mediated through them (Kung et al., 2013). The present understanding of protein-lncRNA interactions are limited to a handful of candidates associated with proteins involved in epigenetic modifications as in the cases of HOTAIR (Gupta et al., 2010), Anril (Kogo et al., 2011), and Xist (Arthold et al., 2011); splicing as in the case of MALAT1 (or NEAT2) (Tripathi et al., 2010) conserved nuclear ncRNA; transcriptional regulation through interaction with transcription factors as in the case of Gas5 (Kino et al., 2010) and few other candidates like Meg3 (Zhao et al., 2006), DHFR (Blume et al., 2003), and Gomafu (Sheik Mohamed et al., 2010). It has been recommended that computational methods for predicting protein-RNA interactions, though less accurate, could be potentially used to guide in experiments (Puton et al., 2012). Recently experimental methodologies to understand protein-RNA interactions on a genomic-scale, including CLIP-seq (Darnell, 2012) and variants thereof (Hafner et al., 2010a; Jain et al., 2011; Konig et al., 2011) has provided insights into the target-sites of a number of RNA binding proteins with much higher resolution (Popov and Gil,

2010). The availability of genome-scale maps of RNA binding proteins provide a novel opportunity toward understanding patterns of RNA binding proteins interaction sites in different transcript classes and derive clues on the interaction networks, regulation and functional consequences of these interactions.

Recently, Li and coworkers showed the interaction between protein and lncRNAs, in addition to their association with disease causing SNPs. They have deposited all the interaction data in form of bed files in starBase 2.0 database, the same datasets are also included in our current study (Li et al., 2014). Tartaglia and coworkers have also employed a novel algorithm catRAPID to evaluate the binding tendency of protein with RNAs (Livi et al., 2015). A similar study by Park et al. has also attempted to explore the possible functions of lncRNAs by focusing at the RBP-lncRNA interactions. lncRNAtor functionally annotates lncRNA molecules based on their expression profiles and co-expression with mRNAs. It also encompass lncRNAs interaction data with 57 RBPs for 5 organisms (Park et al., 2014).

The functional interactions of lncRNAs could be potentially summarized as the sum total of the interactions between other biomolecules independently or in context of one another. The interaction of lncRNAs with genomic DNA and its involvement in chromatin organization (Lee and Bartolomei, 2013) and with other RNA species (Salmena et al., 2011; Bhartiya et al., 2012; Jalali et al., 2015) including microRNAs (Jalali et al., 2013) has been explored at length. Though there have been a number of reports characterizing functional roles of lncRNAs through their association with proteins (Wilusz et al., 2009), no systematic analysis reports has been published on mapping or on characterizing the functional domains of lncRNAs for protein-binding sites. Our study focuses on providing a platform to explore these interactions at a larger scale using computational approaches to functionally indict the lncRNA molecules.

In the present report, we have performed a comprehensive analysis of 70 experimental RNA binding protein datasets available in the public domain. We have derived the peak information (or the most probable site of interaction between protein and RNA) for these RNA binding protein sites at a genome-scale from doRiNA (Blin et al., 2015), starBase (Yang et al., 2011; Li et al., 2014), and CLIPdb (Yang et al., 2015) and analyzed their binding sites in lncRNAs and protein coding transcripts. Our analysis suggests 6 lncRNA subtypes (viz; antisense, lincRNA, miscRNA, processed transcripts, retained intron, and sense intronic) to be largely enriched for protein-binding sites compared to other subclasses hence potentially contribute to a novel layer of regulatory interactions mediated through protein-RNA interactions in ncRNA transcripts. Our analysis shows the distribution of RBP binding sites on the lncRNA loci as opposed to only protein coding transcripts. In our study, we also reveal an interesting pattern of positional clustering of RBP target sites in lncRNAs suggesting a modular organization of regulatory sites in lncRNAs. We also propose how the functionally similar proteins co-occur in both protein coding and lncRNA transcripts. To our knowledge, this is the most comprehensive study on the comparison of lncRNA-RBP interactions as opposed to protein coding loci.

Abbreviations: lncRNA, long non-coding RNA; DNA, Deoxyribonucleic acid; CLIP, cross-linking immunoprecipitation; HITS-CLIP, UV cross-linking and immunoprecipitation with high-throughput sequencing; PAR-CLIP, Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation; iCLIP, individual-nucleotide resolution Cross-Linking and Immuno Precipitation; ncRNA, Non-coding RNA; lincRNA, long intergenic RNA; TEC, To be Experimentally Confirmed; RIP-seq, RNA Immunoprecipitation sequencing; CLASH, cross-linking ligation and sequencing of hybrids; CIMS, Crosslinking induced mutation site; CITS, crosslinking induced truncation analysis; RISC, RNA-induced silencing complex; RNAi, RNA interference; miRNA, microRNA; RBP, RNA binding protein; UTR, untranslated region; CDS, coding sequence; UCSC, University of California, Santa Cruz.

METHODS

Long Non-coding RNA Datasets

We used the comprehensive compendium of lncRNAs available from GENCODE Version 24 (August 2015 freeze, GRCh38, Ensembl 83, 84) (<http://www.encodegenes.org/>) (Harrow et al., 2012). The lncRNA dataset had a total of 32,446 genes encompassing 83,215 transcripts having 3,14,672 exons comprising of both Ensembl and Havana annotations. lncRNA transcripts were assigned into 13 biotypes, viz, 3prime overlapping ncRNA, antisense, bidirectional promoter lncRNA, lincRNA, macro lncRNA, miscRNA, non-coding, processed transcripts, pseudogene, retained intron, sense intronic, sense overlapping, and TEC. We also extracted the 19,655 protein coding genes with 79,930 transcripts and their 7,11,466 exons.

Genome Scale Datasets for Protein-RNA Interactions

We have compiled and analyzed the protein-RNA interaction datasets from public domain for 70 unique proteins derived from 51 publications across 3 databases (detailed in **Table 2**). The RBP binding sites were downloaded from 3 databases namely: starBase v2.0 (Yang et al., 2011; Li et al., 2014), doRiNA 2.0 (Blin et al., 2015), and ClipDB v1.0 (last updated: April, 2015) (Yang et al., 2015). The ClipDB database consisted of datasets analyzed using 4 different softwares PARalyzer (Corcoran et al., 2011), CIMS (Crosslinking induced mutation site) (Moore et al., 2014), CITS (Weyn-Vanhentenryck et al., 2014), and Piranha (Uren et al., 2012).

These datasets comprise of positions of interaction of RNA binding protein and RNA target sites derived after PAR-CLIP (Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation), HITS-CLIP-seq (High Throughput sequencing of RNA isolated by crosslinking immunoprecipitation), RIP-seq (RNA immunoprecipitation), iCLIP (individual nucleotide resolution crosslinking and immunoprecipitation), PAR-iCLIP (Photoactivatable Ribonucleoside Enhanced individual nucleotide resolution crosslinking and immunoprecipitation) and CLASH (cross-linking ligation and sequencing of hybrids) followed by sequencing of the pull-down fraction of RNA. The sequenced RNA is further used to identify exact or probable binding site using various bioinformatic approaches. In case of ClipDB, the peak calling and identification were done using PARalyzer, CIMS, Piranha, and CITS software tools. Hence, we stored each

of files derived from all databases in form of peaks as separate files for downstream analysis. Details of all the techniques and methodologies used to process the data used in our analysis is given in **Table 1**.

All these RNA binding sites were leftover to hg38/GRCh38 assembly using the CrossMap-0.2.2 tool (Zhao et al., 2014). The peak information was available for proteins as shown in **Supplementary Tables 1A,B**. In total, we considered 7 datasets for our study, namely: (1) starBase; (2) doRiNA; (3) Clipdb-PARalyzer; (4) CLIPdb-CIMS; (5) CLIPdb-CITS; (6) CLIPdb-Piranha-stranded; and (7) CLIPdb-Piranha-non-stranded.

Mapping of RNA Binding Protein Interaction Sites

The peaks of the RNA binding protein interaction sites were mapped to the lncRNA exons using bespoke perl script and BEDtools (v2.17.0) (Quinlan and Hall, 2010). The most probable site of interaction (or the peaks) between protein and RNA were derived from datasets taken from doRiNA, starBase, and CLIPdb databases which were processed through standard computational pipelines (as listed in **Table 2**), offering an easy comparability at the analysis point of view. Further, we tried to analyze the binding sites in each of the individual lncRNA subclasses as defined by GENCODE annotations (i.e., 3 prime overlapping ncRNA, antisense, bidirectional promoter lncRNA, lincRNA, macro lncRNA, miscRNA, non-coding, processed transcripts, pseudogene, retained intron, sense intronic, sense overlapping, and TEC). Similarly, we also plotted the distribution of the binding sites across the protein-coding exons derived from the GENCODE v24 annotation file.

We further tested the significance of binding frequency for each of the lncRNA biotype when compared to the protein coding transcripts. The normalized frequency of binding was calculated by dividing the unique number of RBP peaks mapped from each dataset by unique number of bases of lncRNA/protein coding/random transcripts per kb. Statistical unpaired *t*-test was applied using R (version 3.1.3) (R Core Team, 2015) script, to test if any of the lncRNA biotypes had significantly higher RBP binding frequency as compared to the protein coding transcripts.

Combinatorial Patterns for RNA Binding Protein Interaction Sites in lncRNAs

We explored the possibility of positional clustering of RNA binding protein interaction sites across the lncRNA and

TABLE 1 | The summary of total datasets examined in our study for analysis.

Database	Technique used	Software to predict binding sites	Total datasets	No of unique RBP
starBase	CLASH, HITS-CLIP, iCLIP, PAR-CLIP	Ago PAR-CLIP raw data were reanalyzed using PARalyzer v1.1, other CLIP-identified binding sites clusters/peaks were used directly from their respective publications	86	38
CLIPdb	HITS-CLIP, iCLIP, PAR-CLIP, PAR-iCLIP	PARalyzer, CIMS, Piranha, CITS	430	59
doRiNA	iCLIP, HITS-CLIP, CLIP-Seq, PAR-CLIP	Binding sites directly adopted from their respective publications	79	37

TABLE 2 | List of the 70 RNA binding proteins derived from the respective databases.

S. no.	Name	Full name	doRiNA	starBase	Clipdb	Publications
1	AGO1	Argonaute RISC catalytic component 1	✓	✓	✓	Hafner et al., 2010b; Helwak et al., 2013; Memczak et al., 2013
2	AGO2	Argonaute RISC catalytic component 2	✓	✓	✓	Hafner et al., 2010b; Gottwein et al., 2011; Kishore et al., 2011; Lipchina et al., 2011; Haecker et al., 2012; Riley et al., 2012; Skalsky et al., 2012; Karginov and Hannon, 2013; Memczak et al., 2013; Xue et al., 2013
3	AGO3	Argonaute RISC catalytic component 3	✓	✓	✓	Hafner et al., 2010b
4	AGO4	Argonaute RISC catalytic component 4	✓	✓	✓	Hafner et al., 2010b
5	ALKBH5	Alpha-Ketoglutarate-Dependent Dioxygenase AlkB Homolog 5	✓	✓	✓	Baltz et al., 2012
6	ATXN2	Ataxin 2	✓		✓	Yokoshi et al., 2014
7	C17ORF85/NCBP3	Chromosome 17 Open Reading Frame 85/Nuclear Cap Binding Subunit 3	✓	✓	✓	Baltz et al., 2012
8	C22ORF28/RTCB	Chromosome 22 Open Reading Frame 28 /RNA 2,3-Cyclic Phosphate And 5-OH Ligase	✓	✓		Baltz et al., 2012
9	CAPRIN1	Cytoplasmic Activation- And Proliferation-Associated Protein 1	✓	✓	✓	Baltz et al., 2012
10	CPSF1	Cleavage And Polyadenylation Specific Factor 1			✓	Martin et al., 2012
11	CPSF2	Cleavage And Polyadenylation Specific Factor 2			✓	Martin et al., 2012
12	CPSF3	Cleavage And Polyadenylation Specific Factor 3			✓	Martin et al., 2012
13	CPSF4	Cleavage And Polyadenylation Specific Factor 4			✓	Martin et al., 2012
14	CPSF6	Cleavage And Polyadenylation Specific Factor 6			✓	Martin et al., 2012
15	CPSF7	Cleavage And Polyadenylation Specific Factor 7			✓	Martin et al., 2012
16	CSTF2	Cleavage Stimulation Factor Subunit 2			✓	Martin et al., 2012; Yao et al., 2012
17	CSTF2T	Cleavage Stimulation Factor Subunit 2, Tau Variant			✓	Martin et al., 2012
18	DGCR8	DiGeorge Syndrome Critical Region 8	✓	✓	✓	Macias et al., 2012
19	EIF4A3	Eukaryotic Translation Initiation Factor 4A3	✓	✓	✓	Saulière et al., 2012
20	ELAVL1/HUR	Embryonic Lethal Abnormal Vision Drosophila)-Like binding protein 1/Human Antigen R	✓	✓	✓	Lebedeva et al., 2011; Mukherjee, 2011; Friedersdorf and Keene, 2014
21	EWSR1	Ewing Sarcoma Breakpoint Region 1	✓	✓	✓	Hoell et al., 2011; Paronetto et al., 2014
22	EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit			✓	Kaneko et al., 2013
23	FBL	Fibrillarin			✓	Kishore et al., 2013
24	FIP1L1	Factor Interacting With PAPOLA And CPSF1			✓	Martin et al., 2012
25	FMR1	Fragile X Mental Retardation 1	✓	✓	✓	Ascano et al., 2012
26	FOX2	RNA binding protein fox-1 homolog 2	✓			Yeo et al., 2009
27	FUS	Fused in Sarcoma	✓	✓	✓	Hoell et al., 2011; Lagier-Tourenne et al., 2012; Nakaya et al., 2013; Yokoshi et al., 2014
28	FXR1	Fragile X Mental Retardation, Autosomal Homolog 1	✓	✓	✓	Ascano et al., 2012
29	FXR2	Fragile X Mental Retardation Autosomal Homolog 2	✓	✓	✓	Ascano et al., 2012
30	HNRNPA1	Heterogeneous Nuclear Ribonucleoprotein A1			✓	Huelga et al., 2012
31	HNRNPA2B1	Heterogeneous Nuclear Ribonucleoprotein A2/B1			✓	Huelga et al., 2012
32	HNRNPC	Heterogeneous Nuclear Ribonucleoprotein C (C1/C2)	✓	✓	✓	Zarnack et al., 2010, 2013
33	HNRNPD	Heterogeneous Nuclear Ribonucleoprotein D			✓	Yoon et al., 2014
34	HNRNPF	Heterogeneous Nuclear Ribonucleoprotein F			✓	Huelga et al., 2012
35	HNRNPH	Heterogeneous Nuclear Ribonucleoprotein H1			✓	Katz et al., 2010
36	HNRNPL	Heterogeneous Nuclear Ribonucleoprotein L	✓		✓	Shankarling et al., 2014
37	HNRNPM	Heterogeneous Nuclear Ribonucleoprotein M			✓	Huelga et al., 2012

(Continued)

TABLE 2 | Continued

S. no.	Name	Full name	doRiNA	starBase	Clipdb	Publications
38	HNRNPU	Heterogeneous Nuclear Ribonucleoprotein U (Scaffold Attachment Factor A)			✓	Huelga et al., 2012; Xiao et al., 2012
39	IGF2BP1	Insulin Like Growth Factor 2 mRNA Binding Protein 1	✓	✓	✓	Hafner et al., 2010b
40	IGF2BP2	Insulin Like Growth Factor 2 mRNA Binding Protein 2	✓	✓	✓	Hafner et al., 2010b
41	IGF2BP3	Insulin Like Growth Factor 2 mRNA Binding Protein 3	✓	✓	✓	Hafner et al., 2010b
42	LIN28A	Lin-28 Homolog A	✓	✓	✓	Wilbert et al., 2012; Hafner et al., 2013
43	LIN28B	Lin-28 Homolog B	✓	✓	✓	Graf et al., 2013; Hafner et al., 2013
44	METTL3	Methyltransferase-Like Protein 3	✓	✓	✓	Ping et al., 2014
45	MOV10	Moloney Leukemia Virus 10 RISC Complex RNA Helicase	✓	✓	✓	Sievers et al., 2012
46	NOP56	Nucleolar Protein 5A			✓	Kishore et al., 2013
47	NOP58	Nucleolar Protein 5			✓	Kishore et al., 2013
48	NUDT21	Nudix (Nucleoside Diphosphate Linked Moiety X)-Type Motif 21			✓	Martin et al., 2012
49	PTBP1	Polypyrimidine Tract Binding Protein 1		✓	✓	Xue et al., 2013; Raj et al., 2014
50	PTBP1 and PTBP2	Polypyrimidine Tract Binding Protein 1 and Polypyrimidine Tract Binding Protein 2			✓	Xue et al., 2009
51	PUM2	Pumilio RNA Binding Family Member 2	✓	✓	✓	Hafner et al., 2010b
52	QKI	Quaking homolog KH domain RNA binding	✓	✓	✓	Hafner et al., 2010b
53	RBM10	RNA Binding Motif Protein 10	✓			Wang et al., 2013
54	RBPM5	RNA Binding Protein With Multiple Splicing	✓			Farazi et al., 2014
55	RTCB	RNA 2',3'-Cyclic Phosphate And 5'-OH Ligase			✓	Baltz et al., 2012
56	SFRS1	Splicing factor arginine/serine-rich 1	✓	✓		Sanford et al., 2009
57	SRRM4	Serine/Arginine Repetitive Matrix 4			✓	Raj et al., 2014
58	TAF15	TATA-Box Binding Protein Associated Factor 15	✓	✓	✓	Hoell et al., 2011; Ibrahim et al., 2013
59	TARDBP/TDP-43	Transactive response DNA binding protein 43 kDa	✓	✓	✓	Tollervey et al., 2011; Yokoshi et al., 2014
60	TIA1	T-Cell-Restricted Intracellular Antigen-1	✓	✓	✓	Ule et al., 2010
61	TIAL1	TIA-1-Related Protein	✓	✓		Ule et al., 2010
62	TNRC6A	Trinucleotide repeat-containing gene 6A		✓	✓	Hafner et al., 2010b
63	TNRC6B	Trinucleotide repeat-containing gene 6B		✓	✓	Hafner et al., 2010b
64	TNRC6C	Trinucleotide repeat-containing gene 6C		✓	✓	Hafner et al., 2010b
65	U2AF65	U2 Small Nuclear Ribonucleoprotein Auxiliary Factor (65kD)		✓		Zarnack et al., 2013
66	UPF1	Up-Frameshift Suppressor 1 Homolog		✓		Zünd et al., 2013
67	WDR33	WD Repeat Domain 33			✓	Schönemann et al., 2014
68	WTAP	Wilms Tumor 1 Associated Protein	✓			Ping et al., 2014
69	YTHDF2	YTH N6-Methyladenosine RNA Binding Protein 2			✓	Wang et al., 2014
70	ZC3H7B	Zinc Finger CCCH-Type Containing 7B	✓	✓	✓	Baltz et al., 2012

protein coding transcripts. For this, we calculated the co-occurrence binding frequencies for each of the 70 RBPs from the six datasets for each of the lncRNAs and protein coding transcripts in the annotation list. For this analysis we did not consider the CLIPdb-Piranha-non-stranded) dataset due to lack of strand orientation information. Bespoke shell scripts were used to identify RBP sites which co-occurred with each other and were therefore clubbed together.

The coordinates for each RBP peak dataset were intersected separately with both the lncRNA and protein coding exons

using BEDtools. These intersecting coordinates were then used to calculate the number of bases which were shared between each of the protein datasets to examine their co-occurrence. The values were further normalized by dividing it with the total number of unique bases of individual RBP datasets which were intersecting with lncRNA and protein coding exons. The mapping percentage in protein coding transcripts provided the baseline for co-occurrence frequency of the binding sites. These co-occurrence frequencies were calculated independently for all the RBP across six datasets.

Positional Preference of RNA Binding Protein Interaction Sites in lncRNAs

We also examined the positional preference of the RNA binding protein interaction sites across the length of lncRNA transcript. As the length of the transcripts varied considerably in our analysis therefore, we briefly define the length of the transcripts as divided into three equal parts. The length of long non-coding transcripts were normalized to 100 nucleotides and arbitrary divided into three equal parts viz., 5 prime end, the middle region, and 3 prime end for comparisons. The notation 5 prime, middle region, and 3 prime denote the positions of the three equal fragments and have no bearing with 5 prime and 3 prime UTRs. Except for datasets analyzed using Piranha, which did not have strand information of the called RBP peaks, all other datasets were used to check for their positional preference. The unique number of bases intersecting with each of the three lncRNA segments was calculated for each dataset. These were further normalized by dividing these values with the unique number of bases in the respective lncRNA segment. Percentage preference was calculated for each segment and the positional location of RNA protein-binding sites were enumerated and plotted as heatmaps.

Additionally, we also plotted the counts of the RNA binding protein interactions sites in protein coding transcripts derived from GENCODE annotation file and the mappings were divided into 3 regions: 5 prime UTR, coding exons, and 3 prime UTR of the coding genes. The CLIPdb-Piranha-non-stranded dataset were not used for the analysis due to the lack of strand information of the peaks.

RESULTS

Analysis of Mapping of RNA Binding Proteins Datasets

We analyzed publicly available datasets for 70 RNA binding proteins derived from seven datasets encompassing five technologies viz. PAR-CLIP, HITS-CLIP, iCLIP, RIP-seq, and CLASH. The experimental datasets were downloaded for RNA binding proteins from three databases (details in Table 1). The experiments briefly included high-throughput genome-scale analysis of RNA protein interactions through pull down and sequencing. The derived data in form of interaction sites (or peaks) which were pre-processed using different computational pipelines including PARalyzer, CIMS, Piranha, and CITS for each of the proteins and were mapped onto the hg38 build of the Human reference genome. The total number of peaks mapping to the genome for respective datasets corresponding to each RNA binding protein has been detailed in Supplementary Tables 1A,B. Each of the dataset was kept as a separate file even if the name of the RNA binding protein was same. This was followed to maintain the identity of each dataset as there were differences in number of peaks for same proteins across different databases which could be attributed to the different experimental protocols used for processing including difference in cell lines, conditions or end points, or downstream computational

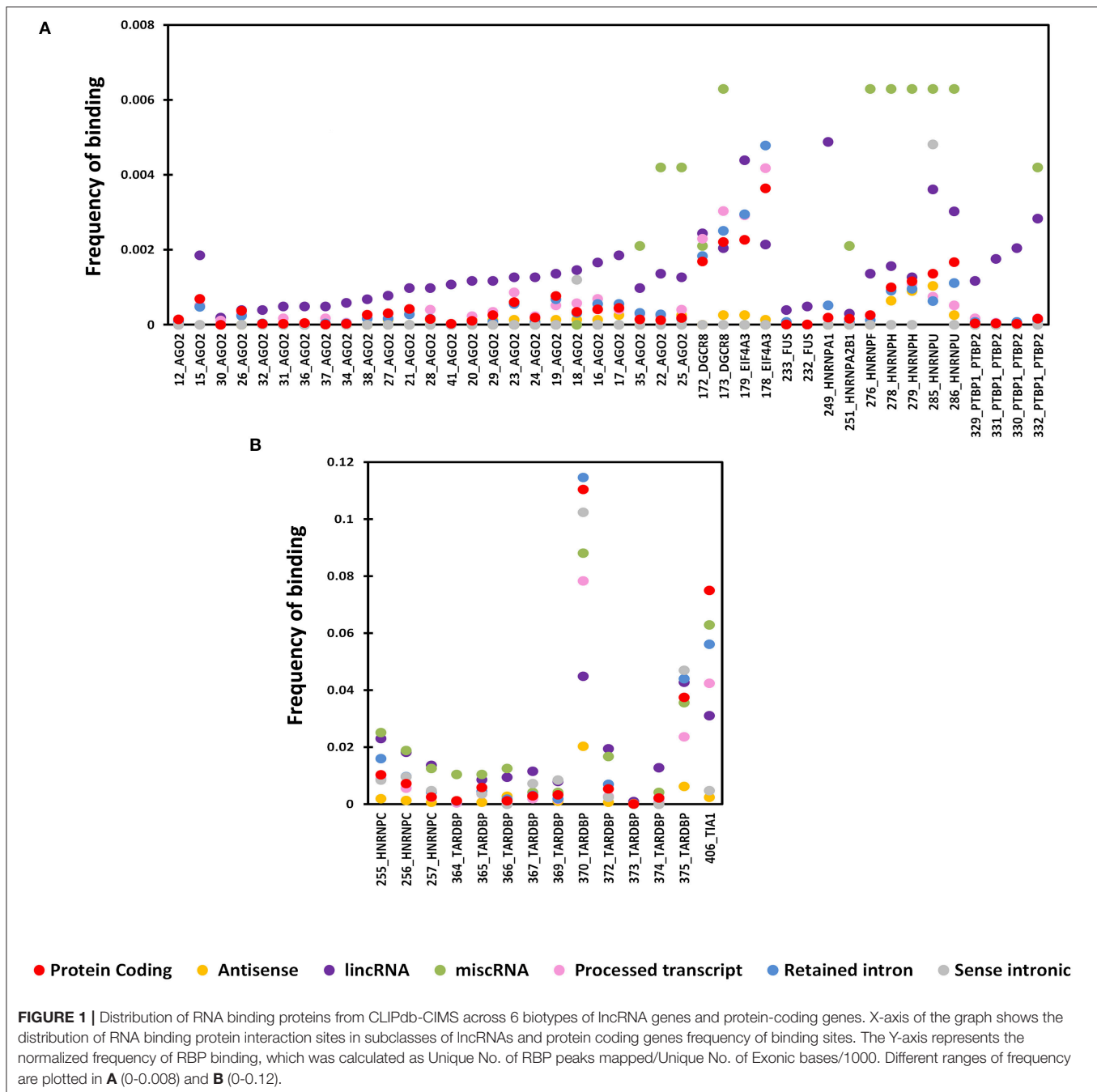
processing. As same protein was present in more than one dataset, we did not group them as one because different databases had differences in the number and position of peaks owing to the differences in the peak calling softwares and computational pipelines adopted by the users. Nevertheless, the differences in the global frequencies have not been influenced by these.

Comparison of RNA Binding Protein Interaction Sites Within lncRNAs and Protein Coding Genes

We compared the interaction sites for each of the RNA binding proteins in lncRNAs as well as protein-coding transcripts. Toward this end, we used the transcript annotations as provided by GENCODE V24 (Harrow et al., 2012) for protein-coding transcripts and lncRNAs. In total the dataset comprised of 79,930 protein-coding transcripts from 19,655 genes and 83,215 lncRNA transcripts arising out of 32,446 genic loci. We analyzed the distribution of RNA binding protein interaction sites across lncRNAs and protein coding transcripts.

All proteins showed distinct frequency distribution across both protein-coding and long non-coding transcripts. In general, RBP binding was higher in protein coding transcripts when compared to long non-coding transcripts. But when we looked closely, few of RBPs showed higher enrichment for lncRNA subclass when compared to protein coding transcripts. We tested the significance of the enrichment of RBP sites across lncRNA subtypes as opposed to protein coding transcripts using paired *t*-test. We observed that six of the biotypes including antisense, lincRNA, miscRNA, processed transcripts, retained intron, and sense intronic were more enriched (p -value ≤ 0.05) for RBP sites as opposed to protein coding transcripts in some or the other RBP dataset.

We plotted the binding frequencies of RBPs in lncRNAs and protein coding transcripts for each of the seven datasets as separate graphs. Those datasets and biotypes which had a significantly higher binding for RBPs have been plotted (Figure 1, Supplementary Figures 1, 2). The RBP binding frequency for CLIPdb-CIMS dataset was significantly higher in lincRNA class when compared to protein coding transcripts for all proteins, while HNRNP (F, H, and U) protein had consistent enrichment for miscRNA class (Figure 1). HNRNP complexes help in processing of pre-mRNAs into functional, translatable mRNAs in the cytoplasm. AGO group from CLIPdb-Piranha-non-stranded dataset were mostly enriched for miscRNA, sense intronic, and lincRNA class compared to protein coding transcript while most of proteins showed enrichment for miscRNA and lincRNA classes (Supplementary Figure 1). In Supplementary Figure 1B, we observed miscRNA and lincRNA class to be mostly enriched for most of proteins including AGO proteins, CSTF2 in sense intronic and DGCR8 in retained intron class. AGO2 protein is an important part of RNA-induced silencing complex (RISC) and is required for RNA-mediated gene silencing (RNAi). CSTF2 plays role in polyadenylation and 3'-end



cleavage of mammalian pre-mRNAs. DGCR8 is a component of the microprocessor complex that acts as a RNA- and heme-binding protein that is involved in the initial step of microRNA (miRNA) biogenesis. For the starBase, CLIPdb-CITS, doRiNA, Clipdb-PARalyzer datasets RBPs showed higher frequency distribution for lncRNAs (miscRNA, retained intron processed transcript) compared to protein coding transcripts (Supplementary Figures 2A–D), ATXN2 protein from Supplementary Figure 2D had a comparable binding frequency in miscRNA class to protein coding transcripts. This protein is involved in EGFR trafficking, acting as negative

regulator of endocytic EGFR internalization at the plasma membrane. Proteins from CLIPdb-Piranha-stranded had enrichment for miscRNA class when compared to protein coding transcripts (Supplementary Figure 2E).

We additionally chose a random set of 1,000,000 (1 million) genomic loci as a control set with an average length of 240 bases and mapped the RBP sites across this control set. The frequencies of protein binding sites across these random genomic loci, lncRNA, and protein coding transcripts of randomly chosen RBPs from each of the six datasets have been depicted in the Supplementary Figure 3, to illustrate that the frequency of

protein binding sites in lncRNAs is not an arbitrary event. The observed RBP frequency was significantly lower for these random positions when compared to protein coding transcripts and lncRNAs. This clearly substantiates the fact that the observed RBP distribution frequencies are not just due to randomness but are inherently due to the class of RNA they bind.

Combinatorial Patterns for Protein-Binding Sites in lncRNAs Show Similar Proteins Have Overlapping Binding Sites

The seven datasets considered in this study were observed to map onto lncRNA transcripts as well as protein-coding transcripts. To understand whether they map to common subset of loci in the respective transcripts, we evaluated the positional overlaps

of the binding sites for each protein from these seven datasets individually. The counts of overlaps were measured as proportion of the total number of independent occurrences of binding sites for each protein. The overlaps were counted separately for all positions in the protein coding transcripts and in lncRNAs. The mapping in protein coding transcripts served as the control set which provided a fair idea of the general overlap in the genomic scale.

Four proteins from the CLIPdb-CITS dataset CSTF2, HNRNPC, TARDBP, and TIA1 showed maximum co-occurrence with their respective set of proteins both in protein coding and lncRNAs transcripts while CSTF2, HNRNPC, and TIA1 co-occurred with each other as well. Our analysis revealed that similar functioning proteins have significantly higher overlapping binding sites with each other, as expected, while EZH2 was an exception in this dataset (Figure 2).

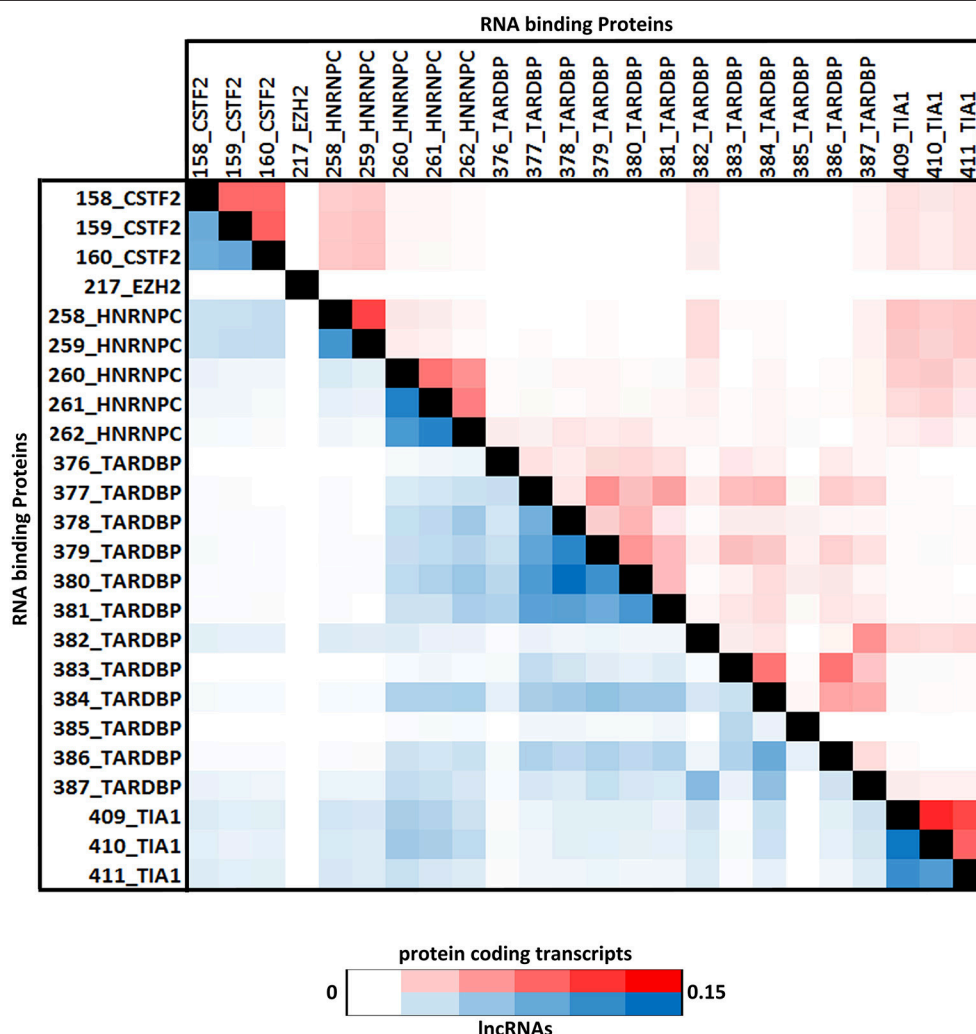


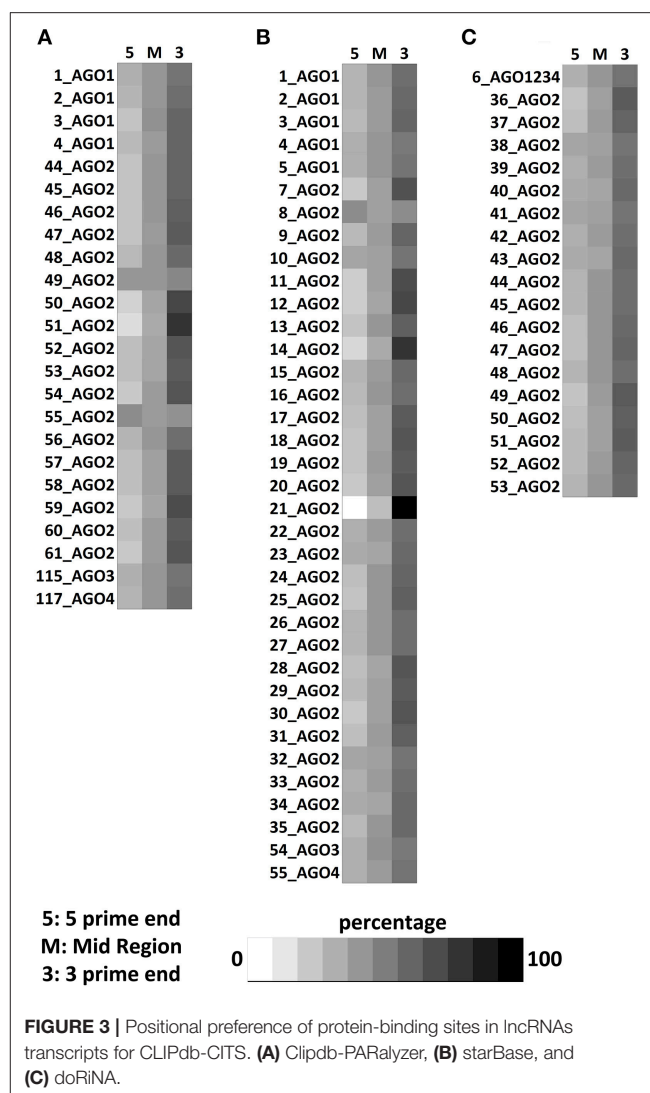
FIGURE 2 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for CLIPdb-CITS RBPs. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins. The diagonal blocks represent a value of 1 corresponding to the exact overlap between the individual protein datasets.

Similarly, RBPs from other five datasets also showed same behavior of co-occurrence between the same set of proteins as shown in **Supplementary Figures 4–8** as heatmap. ELAVL1 co-occurred with HUR proteins from doRiNA dataset with high co-occurrence binding frequency as both being the alternate name of same protein. HNRNPF co-occurred with HNRNPU; both are part of the same HNRNP complex, infact all the HNRNP proteins are related to each other.

While protein having similar function such as AGO and DGCR8 proteins were co-occurring in both the doRiNA and CLIPdb-CIMS datasets. Similarly, TNRC6 (A-C) proteins co-occurred with AGO proteins from CLIPdb-Piranha-stranded, Clipdb-PARalyzer, and starBase datasets, from previous observations it has been seen that functionally related proteins co-occur as in case of TNRC6 with Argonautes, as they have shown to be to play important roles in microRNA mediated regulation of transcripts (Baillat and Shiekhhattar, 2009; Chen et al., 2009). ATXN2 and TARDB from Clipdb-PARalyzer are known to associate in one complex depending on RNA where they bind, we observed them to co-occur in our analysis (Elden et al., 2010). From Clipdb-PARalyzer dataset CSTF2 co-occurred with CPSF proteins. Argonaute protein was observed to co-occur with FUS, HNRNP, PTBP1, and PTBP2 from CLIPdb-CIMS datasets and from literature it has been reported that all these proteins interact with each except AGO, hence we believe if other proteins co-occur then AGO should also functionally correlate with these proteins. From starBase dataset, we also observed TAF15 and FUS co-occurred. In addition, we also observed that FUS and TARDB proteins co-occurred from Clipdb-PARalyzer dataset and AGO group of proteins from CLIPdb-CIMS dataset co-occurred with HNRNP2B1, HNRNPF, HNRNPM, and HNRNPU proteins. There were other proteins also which co-occurred but with low co-occurrence binding frequency. There was no stark difference in the overlaps of the binding sites between protein coding transcripts and lncRNA sites for each of the proteins considered in our analysis.

Positional Clustering of the Protein-Binding Sites

Positional preferences of the RNA binding protein interaction sites were examined across the entire length lncRNAs. The entire length of transcript was calculated by summing up the lengths of individual exons falling in a transcript and then calculating the position of the mapped RNA binding protein interaction site across this calculated length. As the length of the transcript varied therefore, the entire length was arbitrarily divided into three equal parts viz. 5 prime end, middle region, and 3 prime end. Our analysis revealed that the number of RNA binding protein interaction sites for most of the proteins were in majorly mapping to the 3 prime end and the mid segment of the transcripts as shown in **Figure 3** and **Supplementary Figure 9**. To observe the frequencies of binding sites in protein coding transcripts, we mapped and analyzed the RNA binding protein interaction sites in the protein coding transcripts. The binding frequencies for RBPs were evaluated in protein coding transcripts which were divided as 5 prime UTR, CDS, and 3 prime UTR. The data for



the same was derived from GENCODE annotation file in form of bed files. We observed that RNA binding protein interaction sites were distributed in 3 prime UTR, 5 prime UTR, and coding exons and frequencies varied for each protein. The HUR/ELAV1 protein showed a positional preference toward the 3 prime end across the lncRNA transcript and the same has been reported recently by Wang and group (Wang et al., 2015) (**Figure 4**, **Supplementary Figures 10, 11**).

We further observed that AGO proteins across the three datasets, namely; Clipdb-PARalyzer, starBase, and doRiNA showed to have a positional preference in protein coding and lncRNA transcripts (**Figures 3, 4**). When we examined the mapping for the three datasets in protein coding transcripts, we observed that AGO protein showed preference toward the 3 prime UTR. Previous reports have shown AGO proteins bound to miRNAs to target toward 3 prime end of mRNA thereby affecting its translation (Pillai et al., 2004). Such positional preference for AGO proteins is an established fact when targeting the 3' end of mRNAs leading to post-transcriptional silencing.

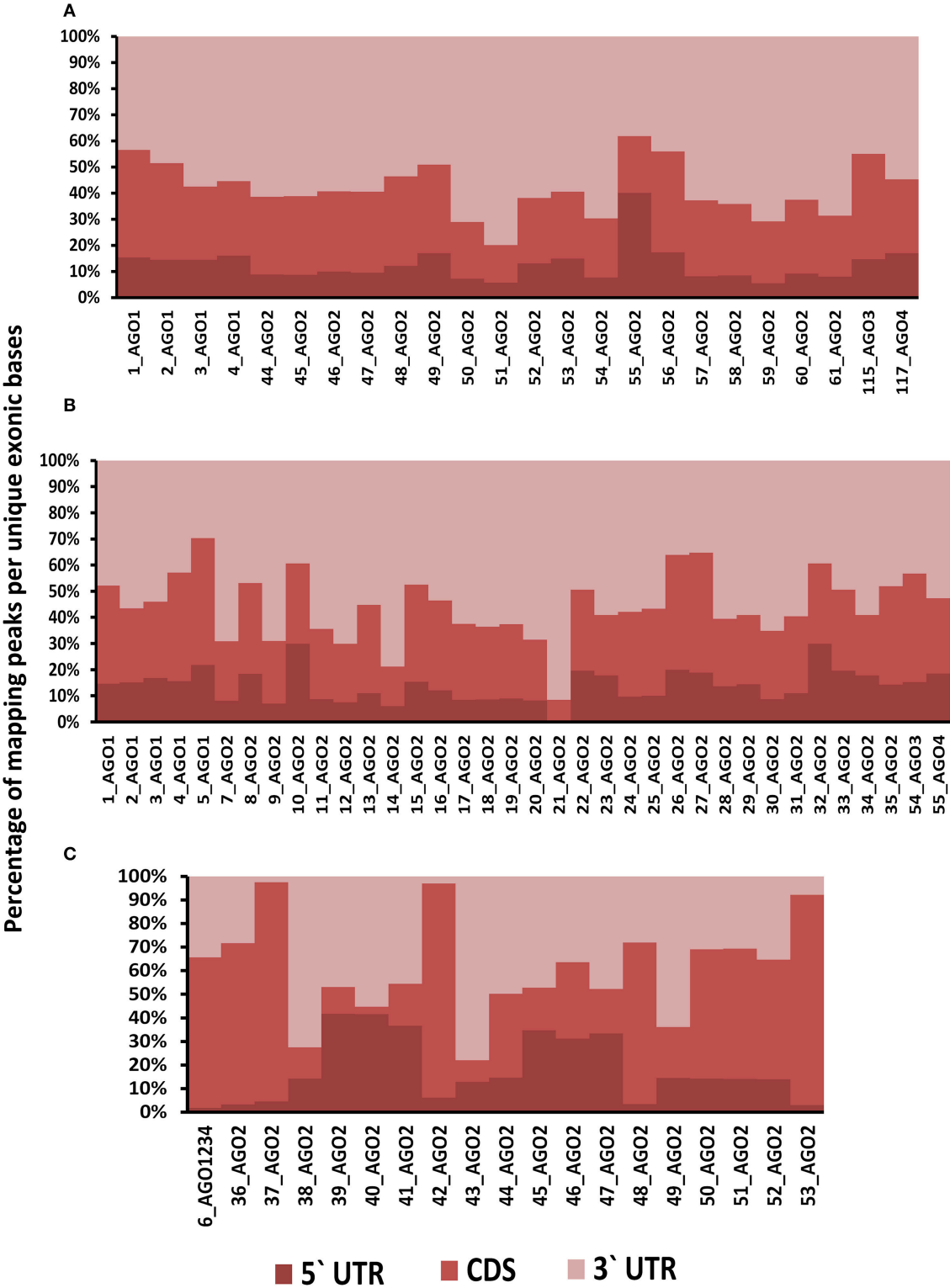


FIGURE 4 | (A) Distribution of RNA binding proteins sites from (A) Clipdb-PARalyzer, (B) starBase, and (C) doRiNA datasets across protein coding genes. X-axis of the graph depicts the RBP sites in protein coding genes and Y-axis is the percentage of mapping of these binding sites across the 5'UTR, CDS, and 3' UTR.

We observed similar positional preference for AGO protein in lncRNAs, thereby suggesting certain regulatory roles.

High Frequencies of RNA Binding Protein Interaction Sites in a Subset of Transcripts

We also observed that many well-known lncRNAs including XIST, NEAT1, OIP5-AS1, and MALAT1 had large number of RNA binding protein sites across their length. A subset of well-annotated lncRNA genes had consistently large number of binding sites for majority of the proteins considered. MALAT1 (metastasis associated lung adenocarcinoma transcript 1), a well-studied lncRNA with intricate roles in the pathophysiology of cancer Metastases is one of such candidate (Gutschner et al., 2013). MALAT1 is highly conserved amongst mammals and is known to be localized in nucleus. We plotted the binding sites for all RBPs to the full-length of MALAT1 transcripts and the same is shown in **Figure 5** for ClipDB-CIMS, CLIPdb-CITS, and CLIPdb-Piranha-stranded datasets. We combined all the datasets for each protein within a database and divided them into three classes (Cytoplasmic, Nuclear, or Both) based on their cellular localization. The distribution profiles for all the RBPs across the MALAT1 gene was derived using UCSC Genome Browser (Meyer et al., 2013).

We observed that the RBPs known to be localizing in nucleus were shown to have higher binding sites across MALAT1 when compared to other RBPs. The functional interaction of MALAT1 with a number of RNA binding proteins have been previously studied (Tripathi et al., 2010), suggesting extensive functional link to the interactions and thereby providing interesting insights

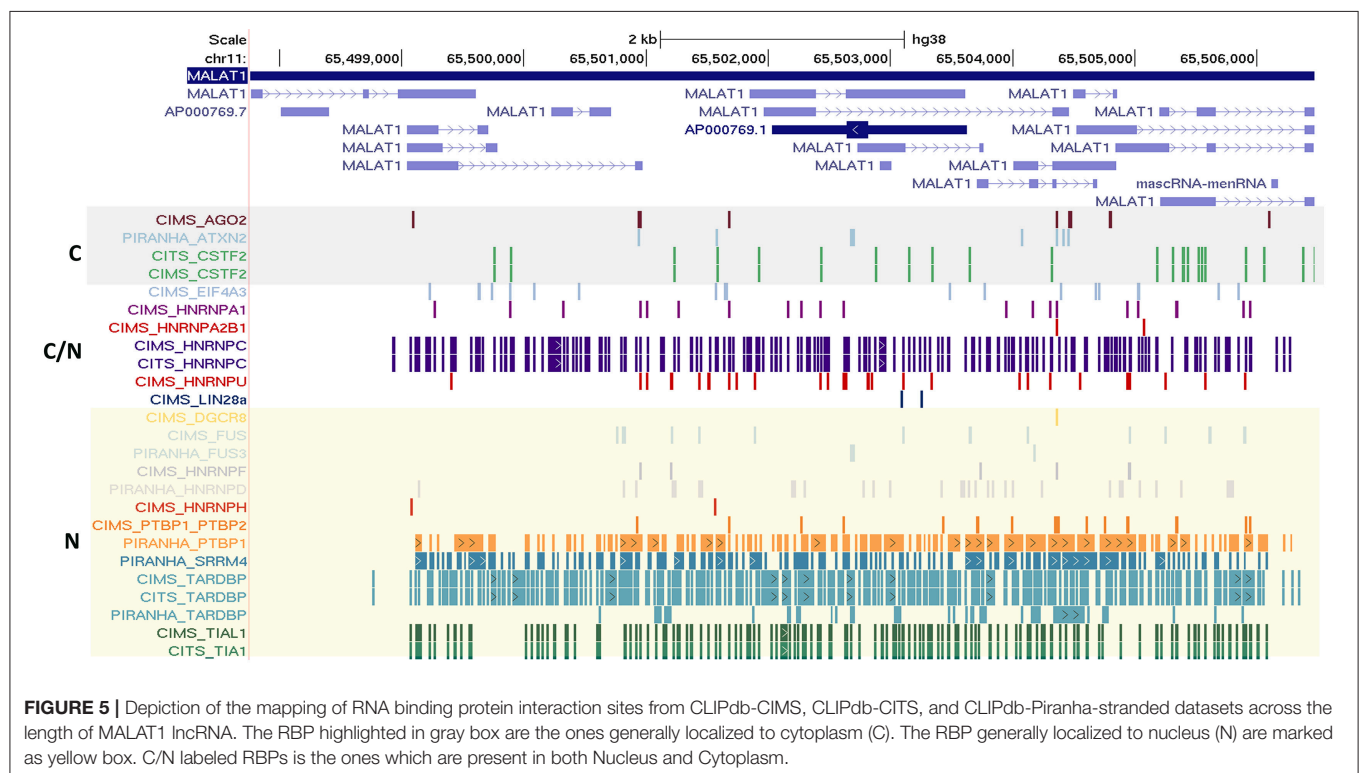
for lncRNA functions and biological regulatory networks they take part in. The mapping for all other datasets across the MALAT1 lncRNA is shown in **Supplementary Figures 12, 13**.

DISCUSSION

lncRNAs have lately emerged as one of the major transcript forms encoded by the human genome, the numbers growing as much as the number of protein-coding transcripts over the years. GENCODE v24 has 83,215 lncRNA loci compared to 79,930 protein-coding gene loci. The functional role of many candidate lncRNAs have been extensively studied in the recent past, nevertheless the general lack of conservation of lncRNAs, even between closely related organisms, barring a handful of candidate lncRNAs has restricted the possibility to model functionalities of lncRNAs in model systems.

The availability of genome-scale assays for evaluating protein-binding sites in RNA (Kishore et al., 2011), has offered new opportunities to address this issue at much higher confidence and resolution than which were provided by computational approaches (Bellucci et al., 2011; Puton et al., 2012). To date, seven datasets for genome-scale protein-RNA interactions are available in public domain (i.e., doRiNA, Clipdb, starBase) and the present analysis makes use of all these available datasets. We show such approaches involving repurposing of datasets could provide immense insights into the biological functions with potential regulation of lncRNAs.

In the present study, we have used the peak information (or the most probable site of interaction between protein and



RNA) from seven datasets processed through standardized computational pipeline for accurate assessment of protein-RNA interaction sites (doRiNA, Clipdb, starBase). This allowed us to compare the frequencies of the protein binding sites in systematic fashion. It has not escaped our attention that the datasets encompass a diverse set of experiments; cell line, and experimental protocols, nevertheless; our findings hold true despite these differences available in public domain as part of this analysis encompassing six experimental databases of RNA binding proteins. For instance, one of the most studied RBP, the Argonaute datasets showed similar trends regardless of the diverse experimental protocols (HITS-CLIP, iCLIP, PAR-CLIP) and analysis methodologies employed.

The RBPs considered in our study are known to be involved in varied types of functional roles including silencing, splicing, stability, mRNA processing, and transport. In the current study, we observed RBPs enriched for specific lncRNA biotypes are involved in diverse functions, suggesting their probable functional mechanism of action. RBPs such as AGO, DGCR8, EWSR1, TNRC6A/B/C, and FUS, involved in maintenance of the stability of RNA, were having significant enrichment for the lncRNA, miscRNA, retained intron subclasses suggesting they might be acting as either transporters or as sponges for these RBPs. Another set of RBPs such as CPSF complex, FBL, TAF15, and HNRNP family, playing a role in mRNA processing were shown to be enriched in lncRNA subclasses, signifying that lncRNAs in turn might be acting as guides. These proteins might be also involved in mechanism of lncRNA biogenesis. Enrichment was also observed for proteins such as ATXN2, C17ORF85, and HNRNPs which predominantly are involved in the export and transporting of RNA moieties, in addition to proteins such as EIF4A3, FOX2, PTBP1, QKI, SFRS1, SRRM4 among others which are predominantly involved in splicing. Hence our analysis suggests that interaction of lncRNA with such types of RBPs surely provide hints about the possible functional roles lncRNAs might be playing which can be validated by experimental approaches.

We also highlight the localization of lncRNAs and RBPs within a cell. We classified the RBPs based on their known localization within the cells and overlapped it with MALAT1, which is an established nuclear enriched lncRNA. The results indicated that the intensity of nuclear localized RBPs were higher for MALAT1 across all the seven datasets. This further strengthened the fact that these bindings were not an arbitrary event and are indeed interacting with the co-localized lncRNAs.

The present analysis reveals a set of interesting characteristics of protein-RNA interaction in the context of lncRNAs: (1) high frequency of RNA-protein interaction sites in lncRNAs subclasses; (2) co-occurrence of RNA binding protein interaction sites; and (3) positional preference for the binding sites across the transcript length. This analysis, to our best of knowledge is the most comprehensive analysis of RNA binding protein interaction sites in lncRNAs, and provides the basis for further analysis on the functional consequences of these patterns. It has also not escaped our attention that targeting protein-interaction sites and thus the functionalities could be in the

future therapeutically explored. Recent reports from other laboratories have explored the possibility of targeting RNA structures using small molecules (Jamal et al., 2012; Bose et al., 2013). Further availability of genome-scale protein-RNA interaction datasets and availability of tools to query RNA secondary structures at genome scale (Hofacker, 2003) would provide us with immense opportunities toward understanding the entire repertoire of functional RNA interactions and phenotypic correlates at a genome-scale level. This would also form the much-needed resource of knowledge to potentially query and understand consequences of genomic variations at these loci.

CONCLUSION

The interactions between proteins and RNA molecules can provide the essential insights into the functioning of the lncRNAs. In this study, we highlight the enrichment of RBP sites across some of the lncRNA transcript classes in comparison with protein coding transcripts. We have systematically demonstrated that proteins having similar functional roles showed a higher co-occurrence across both lncRNA and protein coding transcripts. Also, the positional preference of most of RBPs agreed with their possible functional roles. Our study gives a compendium of lncRNA and RBP interactions suggesting a large number of functional roles which they can play including silencing, splicing, mRNA processing, export or transport.

AUTHOR CONTRIBUTIONS

VS conceptualized the analysis. Data analysis was performed by SJ and SG. SJ prepared the data summaries and visualization. SJ and SG wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This work was funded by the Council of Scientific and Industrial Research (CSIR), India through Grant GENCODE-C (BSC0123).

ACKNOWLEDGMENTS

Authors would like to acknowledge Dr. S. Ramachandran and Dr. Sheetal Gandotra for their valuable discussions which helped in compiling the analysis of this study and writing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2018.00027/full#supplementary-material>

Supplementary Figure 1 | (A,B) Distribution of RNA binding proteins from CLIPdb-Piranha-stranded across six biotypes of lncRNA genes and protein-coding genes. X-axis of the graph shows the distribution of RNA binding protein interaction sites in subclasses of lncRNAs and protein coding genes frequency of binding sites.

Supplementary Figure 2 | Distribution of RNA binding proteins from (A) starBase, (B) CLIPdb-CITS, (C) doRiNA, (D) Clipdb-PARalyzer, and (E) CLIPdb-Piranha-non-stranded across 6 biotypes of lncRNA genes and protein-coding genes. X-axis of the graph shows the distribution of RNA binding protein interaction sites in subclasses of lncRNAs and protein coding genes frequency of binding sites.

Supplementary Figure 3 | Distribution of RNA binding proteins CLIPdb-CIMS (CSTF2), starBase (AGO1), CLIPdb-CITS (HNRNPC), doRiNA (AGO2), Clipdb-PARalyzer (AGO2), and CLIPdb-Piranha-non-stranded (AGO2) across lncRNA, Protein Coding Transcript and Random Genomic Loci. X-axis of the graph represents random RBPs selected from each dataset and Y-axis depicts the normalized frequency of RNA binding protein interaction sites. The frequency is calculated as the number of unique RBP peaks per unique number of exonic bases per kilobase mapped.

Supplementary Figure 4 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for doRiNA dataset RBPs. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins.

Supplementary Figure 5 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for CLIPdb-CIMS dataset RBPs. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins.

Supplementary Figure 6 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for CLIPdb-Piranha-non-stranded dataset RBP. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins.

Supplementary Figure 7 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for starBase dataset RBPs. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins.

Supplementary Figure 8 | The Heatmap depicts the combinatorial patterns of clustered protein-binding sites across lncRNAs (blue in color) and protein coding transcripts (red in color) for Clipdb-PARalyzer dataset RBPs. The scale here signifies the number of overlapping binding sites per total number of occurrences for the independent proteins.

Supplementary Figure 9 | Positional preference of protein-binding sites in lncRNAs transcripts for (A) Clipdb-PARalyzer, (B) CLIPdb-CIMS, (C) starBase, (D) doRiNA, (E) CLIPdb-CITS, and (F) CLIPdb-Piranha-stranded.

Supplementary Figure 10 | (A) Distribution of RNA binding proteins sites from (A) Clipdb-PARalyzer, (B) CLIPdb-CIMS, and (C) starBase datasets across Refseq genes. X-axis of the graph depicts the distribution of RNA binding protein interaction sites in refseq genes and Y-axis is the frequency of binding sites.

Supplementary Figure 11 | (B) Distribution of RNA binding proteins sites from (A) doRiNA, (B) CLIPdb-CITS and (C) CLIPdb-Piranha-stranded datasets across Refseq genes. X-axis of the graph depicts the distribution of RNA binding protein interaction sites in refseq genes and Y-axis is the frequency of binding sites.

Supplementary Figure 12 | Depiction of the mapping of RNA binding protein interaction sites from Clipdb-PARalyzer datasets across the length of MALAT1 lncRNA. The RBP highlighted in gray box are the ones generally localized to cytoplasm (C). The RBP generally localized to nucleus (N) are marked as yellow box. C/N labeled RBPs is the ones which are present in both Nucleus and Cytoplasm.

Supplementary Figure 13 | Depiction of the mapping of RNA binding protein interaction sites from ClipDB (doRiNA and starBase datasets) across the length of MALAT1 lncRNA. The RBP highlighted in gray box are the ones generally localized to cytoplasm (C). The RBP generally localized to nucleus (N) are marked as yellow box. C/N labeled RBPs is the ones which are present in both Nucleus and Cytoplasm.

Supplementary Tables 1 | (A) Detailed list of publically available datasets derived from Starbase and Dorina databases. (B) Detailed list of publically available datasets derived from CLIPdb database.

REFERENCES

- Arthold, S., Kurowski, A., and Wutz, A. (2011). Mechanistic insights into chromosome-wide silencing in X inactivation. *Hum. Genet.* 130, 295–305. doi: 10.1007/s00439-011-1002-0
- Ascano, M. Jr., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386. doi: 10.1038/nature11737
- Baillat, D., and Shiekhata, R. (2009). Functional dissection of the human TNRC6 (GW182-related) family of proteins. *Mol. Cell. Biol.* 29, 4144–4155. doi: 10.1128/MCB.00380-09
- Baltz, A. G., Munschauer, M., Schwanhäusser, 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
- Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods* 8, 444–445. doi: 10.1038/nmeth.1611
- Bhartiya, D., Kapoor, S., Jalali, S., Sati, S., Kaushik, K., Sachidanandan, C., et al. (2012). Conceptual approaches for lncRNA drug discovery and future strategies. *Expert Opin. Drug Discov.* 7, 503–513. doi: 10.1517/17460441.2012.682055
- Blin, K., Dieterich, C., Wurmus, R., Rajewsky, N., Landthaler, M., and Akalin, A. (2015). DoRiNA 2.0-upgrading the dorina database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* 43, D160–D167. doi: 10.1093/nar/gku1180
- Blume, S. W., Meng, Z., Shrestha, K., Snyder, R. C., and Emanuel, P. D. (2003). The 5'-untranslated RNA of the human dhfr minor transcript alters transcription pre-initiation complex assembly at the major (core) promoter. *J. Cell. Biochem.* 88, 165–180. doi: 10.1002/jcb.10326
- Bose, D., Jayaraj, G. G., Kumar, S., and Maiti, S. (2013). A molecular-beacon-based screen for small molecule inhibitors of miRNA maturation. *ACS Chem. Biol.* 8, 930–938. doi: 10.1021/cb300650y
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611
- Cao, J. (2014). The functional role of long non-coding RNAs and epigenetics. *Biol. Proced. Online* 16:11. doi: 10.1186/1480-9222-16-11
- Chen, C.-Y. A., Zheng, D., Xia, Z., and Shyu, A.-B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat. Struct. Mol. Biol.* 16, 1160–1166. doi: 10.1038/nsmb.1709
- Corcoran, D. L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R. L., Keene, J. D., et al. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* 12:R79. doi: 10.1186/gb-2011-12-8-r79
- Darnell, R. (2012). CLIP (Cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. *Cold Spring Harb. Protoc.* 7, 1146–1160. doi: 10.1101/pdb.prot072132
- 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
- Elden, A. C., Kim, H. -J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466, 1069–1075. doi: 10.1038/nature09320

- Farazi, T. A., Leonhardt, C. S., Mukherjee, N., Mihailovic, A., Li, S., Max, K. E., et al. (2014). Identification of the RNA recognition element of the RBPMS family of RNA-binding proteins and their transcriptome-wide mRNA targets. *RNA* 20, 1090–1102. doi: 10.1261/rna.045005.114
- Friedersdorf, M. B., and Keene, J. D. (2014). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol.* 15:R2. doi: 10.1186/gb-2014-15-1-r2
- Furey, T. S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* 13, 840–852. doi: 10.1038/nrg3306
- Gibb, E. A., Brown, C. J., and Lam, W. L. (2011). The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer* 10:38. doi: 10.1186/1476-4598-10-38
- Goodrich, J. A., and Kugel, J. F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. *Nat. Rev. Mol. Cell Biol.* 7, 612–616. doi: 10.1038/nrm1946
- Gottwein, E., Corcoran, D. L., Mukherjee, N., Skalsky, R. L., Hafner, M., Nusbaum, J. D., et al. (2011). Viral microRNA targetome of KSHV-Infected primary effusion lymphoma cell lines. *Cell Host Microbe* 10, 515–526. doi: 10.1016/j.chom.2011.09.012
- Graf, R., Munschauer, M., Mastrobuoni, G., Mayr, F., Heinemann, U., Kempa, S., et al. (2013). Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. *RNA Biol.* 10, 1146–1159. doi: 10.4161/rna.25194
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975
- Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850
- Haecker, I., Gay, L. A., Yang, Y., Hu, J., Morse, A. M., McIntyre, L. M., et al. (2012). Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion Lymphomas. *PLoS Pathog.* 8:e1002884. doi: 10.1371/journal.ppat.1002884
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010a). PAR-CLIP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* 2034. doi: 10.3791/2034
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010b). Transcriptome-wide identification of RNA-binding protein and MicroRNA target sites by PAR-CLIP. *Cell* 141, 129–141. doi: 10.1016/j.cell.2010.03.009
- Hafner, M., Max, K. E. A., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., et al. (2013). Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *RNA* 19, 613–626. doi: 10.1261/rna.036491.112
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111
- Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665. doi: 10.1016/j.cell.2013.03.043
- Hoell, J. I., Larsson, E., Runge, S., Nusbaum, J. D., Duggimpudi, S., Farazi, T. A., et al. (2011). RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* 18, 1428–1431. doi: 10.1038/nsmb.2163
- Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res.* 31, 3429–3431. doi: 10.1093/nar/gkg599
- Huelga, S. C., Vu, A. Q., Arnold, J. D., Liang, T. D., Liu, P. P., Yan, B. Y., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* 1, 167–178. doi: 10.1016/j.celrep.2012.02.001
- Ibrahim, F., Maragkakis, M., Alexiou, P., Maronski, M. A., Dichter, M. A., and Mourelatos, Z. (2013). Identification of *in vivo* Conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep.* 3, 301–308. doi: 10.1016/j.celrep.2013.01.021
- Jain, R., Devine, T., George, A. D., Chittur, S. V., Baroni, T. E., Penaval, L. O., et al. (2011). RIP-chip analysis: RNA-binding protein immunoprecipitation-microarray (chip) profiling. *Methods Mol. Biol.* 703, 247–263. doi: 10.1007/978-1-59745-248-9_17
- Jalali, S., Bhartiya, D., Lalwani, M. K., Sivasubbu, S., and Scaria, V. (2013). Systematic transcriptome wide analysis of lncRNA-miRNA interactions. *PLoS ONE* 8:e53823. doi: 10.1371/journal.pone.0053823
- Jalali, S., Jayaraj, G., Scaria, V., Kapranov, P., Cheng, J., Dike, S., et al. (2012). Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biol. Direct* 7:25. doi: 10.1186/1745-6150-7-25
- Jalali, S., Kapoor, S., Sivasadas, A., Bhartiya, D., and Scaria, V. (2015). Computational approaches towards understanding human long non-coding RNA biology. *Bioinformatics* 31, 2241–2251. doi: 10.1093/bioinformatics/btv148
- Jamal, S., Periwal, V., and Scaria, V. (2012). Computational analysis and predictive modeling of small molecule modulators of microRNA. *J. Cheminform.* 4:16. doi: 10.1186/1758-2946-4-16
- Kaneko, S., Son, J., Shen, S. S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1258–1264. doi: 10.1038/nsmb.2700
- Karginov, F. V., and Hannon, G. J. (2013). Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. *Genes Dev.* 27, 1624–1632. doi: 10.1101/gad.215939.113
- Katz, Y., Wang, E. T., Airolidi, E. M., and Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015. doi: 10.1038/nmeth.1528
- Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3:ra8. doi: 10.1126/scisignal.2000568
- Kishore, S., Gruber, A. R., Jedlinski, D. J., Syed, A. P., Jorjani, H., and Zavolan, M. (2013). Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol.* 14:R45. doi: 10.1186/gb-2013-14-5-r45
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* 8, 559–564. doi: 10.1038/nmeth.1608
- Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., et al. (2011). Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71, 6320–6326. doi: 10.1158/0008-5472.CAN-11-1021
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., et al. (2011). iCLIP—transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* 2638. doi: 10.3791/2638
- Kung, J. T. Y., Colognori, D., and Lee, J. T. (2013). Long noncoding RNAs: past, present, and future. *Genetics* 193, 651–669. doi: 10.1534/genetics.112.146704
- Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* 15, 1488–1497. doi: 10.1038/nn.3230
- Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., et al. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* 43, 340–352. doi: 10.1016/j.molcel.2011.06.008
- Lee, J. T., and Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323. doi: 10.1016/j.cell.2013.02.016
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). StarBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248
- Lipchina, I., Elkabetz, Y., Hafner, M., Sheridan, R., Mihailovic, A., Tuschl, T., et al. (2011). Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev.* 25, 2173–2186. doi: 10.1101/gad.17221311
- Livi, C. M., Klus, P., Delli Ponti, R., and Tartaglia, G. G. (2015). CatRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics* 32, 773–775. doi: 10.1093/bioinformatics/btv629
- Macias, S., Plass, M., Stajda, A., Michlewski, G., Eyra, E., Cáceres, J. F., et al. (2012). DGCR8 HITS-CLIP reveals novel functions for the microprocessor. *Nat. Struct. Mol. Biol.* 19, 760–766. doi: 10.1038/nsmb.2344

- Martin, G., Gruber, A. R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep.* 1, 753–763. doi: 10.1016/j.celrep.2012.05.003
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928
- 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
- Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., et al. (2013). The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.* 41, D64–D69. doi: 10.1093/nar/gks1048
- Moore, M. J., Zhang, C., Gantman, E. C., Mele, A., Darnell, J. C., and Darnell, R. B. (2014). Mapping argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* 9, 263–293. doi: 10.1038/nprot.2014.012
- Mukherjee, N. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* 43, 327–339. doi: 10.1016/j.molcel.2011.06.007
- Nakaya, T., Alexiou, P., Maragkakis, M., Chang, A., and Mourelatos, Z. (2013). FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* 19, 498–509. doi: 10.1261/rna.037804.112
- Park, C., Yu, N., Choi, I., Kim, W., and Lee, S. (2014). LncRNAator: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 30, 2480–2485. doi: 10.1093/bioinformatics/btu325
- Paronetto, M. P., Bernardis, I., Volpe, E., Bechara, E., Sebestyén, E., Eyra, E., et al. (2014). Regulation of FAS exon definition and apoptosis by the ewing sarcoma protein. *Cell Rep.* 7, 1211–1226. doi: 10.1016/j.celrep.2014.03.077
- Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhout, N. L., Levin, J. Z., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. doi: 10.1101/gr.133009.111
- Pillai, R. S., Artus, C. G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA* 10, 1518–1525. doi: 10.1261/rna.7131604
- Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 24, 177–189. doi: 10.1038/cr.2014.3
- Popov, N., and Gil, J. (2010). Epigenetic regulation of the INK4B-ARF-INK4a locus: in sickness and in health. *Epigenetics* 5, 685–690. doi: 10.4161/epi.5.8.12996
- Puton, T., Kozłowski, L., Tuszyńska, I., Rother, K., and Bujnicki, J. M. (2012). Computational methods for prediction of protein-RNA interactions. *J. Struct. Biol.* 179, 261–268. doi: 10.1016/j.jsb.2011.10.001
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z. Y., et al. (2014). A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* 56, 90–103. doi: 10.1016/j.molcel.2014.08.011
- R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Found. Stat. Comput. Available online at: <http://www.r-project.org>
- Riley, K. J., Rabinowitz, G. S., Yario, T. A., Luna, J. M., Darnell, R. B., and Steitz, J. A. (2012). EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J.* 31, 2207–2221. doi: 10.1038/emboj.2012.63
- Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics* 27, 2325–2329. doi: 10.1093/bioinformatics/btr355
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014
- Sanford, J. R., Wang, X., Mort, M., Vanduy, N., Cooper, D. N., Mooney, S. D., et al. (2009). Splicing factor SRSF1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 19, 381–394. doi: 10.1101/gr.082503.108
- Sati, S., Ghosh, S., Jain, V., Scaria, V., and Sengupta, S. (2012). Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* 40, 10018–10031. doi: 10.1093/nar/gks776
- Saulière, J., Murigneux, V., Wang, Z., Marquet, E., Barbosa, I., Le Tonquèze, O., et al. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *TL - 19. Nat. Struct. Mol. Biol.* 19, 1124–1131. doi: 10.1038/nsmb.2420
- Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A. R., Keller, W., et al. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.* 28, 2381–2393. doi: 10.1101/gad.250985.114
- Shankarling, G., Cole, B. S., Mallory, M. J., and Lynch, K. W. (2014). Transcriptome-wide RNA interaction profiling reveals physical and functional targets of hnRNP L in human T cells. *Mol. Cell Biol.* 34, 71–83. doi: 10.1128/MCB.00740-13
- Sheik Mohamed, J., Gaughwin, P. M., Lim, B., Robson, P., and Lipovich, L. (2010). Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* 16, 324–337. doi: 10.1261/rna.1441510
- Sievers, C., Schlumpf, T., Sawarkar, R., Comoglio, F., and Paro, R. (2012). Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic Acids Res.* 40:e160. doi: 10.1093/nar/gks697
- Skalsky, R. L., Corcoran, D. L., Gottwein, E., Frank, C. L., Kang, D., Hafner, M., et al. (2012). The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.* 8:e1002484. doi: 10.1371/journal.ppat.1002484
- Tollervey, J. R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., et al. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* 14, 452–458. doi: 10.1038/nn.2778
- Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011
- Ule, J., Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N. M., et al. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.* 8:e1000530. doi: 10.1371/journal.pbio.1000530
- Uren, P. J., Bahrami-Samani, E., Burns, S. C., Qiao, M., Karginov, F. V., Hodges, E., et al. (2012). Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* 28, 3013–3020. doi: 10.1093/bioinformatics/bts569
- Wang, Z., Bhattacharya, A., and Ivanov, D. N. (2015). Identification of Small-Molecule Inhibitors of the HuR/RNA Interaction using a fluorescence polarization screening assay followed by NMR validation. *PLoS One.* 10:e0138780. doi: 10.1371/journal.pone.0138780
- Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120. doi: 10.1038/nature12730
- Wang, Y., Gogol-Döring, A., Hu, H., Fröhler, S., Ma, Y., Jens, M., et al. (2013). Integrative analysis revealed the molecular mechanism underlying RBM10-mediated splicing regulation. *EMBO Mol. Med.* 5, 1431–1442. doi: 10.1002/emmm.201302663
- Wapinski, O., and Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001
- Weyn-Vanhenenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152. doi: 10.1016/j.celrep.2014.02.005
- Wilbert, M. L., Huelga, S. C., Kapeli, K., Stark, T. J., Liang, T. Y., Chen, S. X., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* 48, 195–206. doi: 10.1016/j.molcel.2012.08.004
- Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23, 1494–1504. doi: 10.1101/gad.1800909
- Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., et al. (2012). Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol. Cell* 45, 656–668. doi: 10.1016/j.molcel.2012.01.009
- Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., et al. (2013). Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* 152, 82–96. doi: 10.1016/j.cell.2012.11.045

- Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y. S., et al. (2009). Genome-wide analysis of PTB-RNA Interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or Skipping. *Mol. Cell* 36, 996–1006. doi: 10.1016/j.molcel.2009.12.003
- Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., and Qu, L. H. (2011). StarBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and degradome-Seq data. *Nucleic Acids Res.* 39, D202–D209. doi: 10.1093/nar/gkq1056
- Yang, Y.-C. T., Di, C., Hu, B., Zhou, M., Liu, Y., Song, N., et al. (2015). CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics* 16:51. doi: 10.1186/s12864-015-1273-2
- Yao, C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X., et al. (2012). Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18773–18778. doi: 10.1073/pnas.1211101109
- Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X. -D., and Gage, F. H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 16, 130–137. doi: 10.1038/nsmb.1545
- Yokoshi, M., Li, Q., Yamamoto, M., Okada, H., Suzuki, Y., and Kawahara, Y. (2014). Direct binding of Ataxin-2 to distinct elements in 3' UTRs promotes mRNA stability and protein expression. *Mol. Cell* 55, 186–198. doi: 10.1016/j.molcel.2014.05.022
- Yoon, J. -H., De, S., Srikantan, S., Abdelmohsen, K., Grammatikakis, I., Kim, J., et al. (2014). PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat. Commun.* 5, 1–15. doi: 10.1038/ncomms6248
- Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., et al. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of alu elements. *Cell* 152, 453–466. doi: 10.1016/j.cell.2012.12.023
- Zarnack, K., Rot, G., Kayikci, M., Turner, D. J., Luscombe, N. M., and Ule, J. (2010). Technical reports. *Nat. Struct. Mol. Biol.* 17, 909–916. doi: 10.1038/nsmb.1838
- Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J. P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30, 1006–1007. doi: 10.1093/bioinformatics/btt730
- Zhao, J., Zhang, X., Zhou, Y., Ansell, P. J., and Klibanski, A. (2006). Cyclic AMP stimulates MEG3 gene expression in cells through a cAMP-response element (CRE) in the MEG3 proximal promoter region. *Int. J. Biochem. Cell Biol.* 38, 1808–1820. doi: 10.1016/j.biocel.2006.05.004
- Zünd, D., Gruber, A. R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat. Struct. Mol. Biol.* 20, 936–943. doi: 10.1038/nsmb.2635

Conflict of Interest Statement: 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 reviewer NA and handling Editor declared their shared affiliation.

Copyright © 2018 Jalali, Gandhi and Scaria. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership