



RNA SPLICING AND BACKSPlicing: DISEASE AND THERAPY

EDITED BY: Rosanna Asselta, Stefano Duga, Emanuele Buratti and
Eladio Andrés Velasco
PUBLISHED IN: Frontiers in Genetics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-235-7

DOI 10.3389/978-2-88966-235-7

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 SPLICING AND BACKSPLICING: DISEASE AND THERAPY

Topic Editors:

Rosanna Asselta, Humanitas University, Italy

Stefano Duga, Humanitas University, Italy

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

Eladio Andrés Velasco, Instituto de Biología y Genética Molecular, Spain

Citation: Asselta, R., Duga, S., Buratti, E., Velasco, E. A., eds. (2020). RNA Splicing and Backsplicing: Disease and Therapy. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-235-7

Table of Contents

- 05 Editorial: RNA Splicing and Backsplicing: Disease and Therapy**
Rosanna Asselta, Stefano Duga, Eladio Andrés Velasco and Emanuele Buratti
- 08 Control of mRNA Splicing by Intragenic RNA Activators of Stress Signaling: Potential Implications for Human Disease**
Raymond Kaempfer, Lena Ilan, Smadar Cohen-Chalamish, Orli Turgeman, Lise Sarah Namer and Farhat Osman
- 14 Minigene Splicing Assays Identify 12 Spliceogenic Variants of BRCA2 Exons 14 and 15**
Eugenia Fraile-Bethencourt, Alberto Valenzuela-Palomo, Beatriz Díez-Gómez, María José Caloca, Susana Gómez-Barrero and Eladio A. Velasco
- 28 Global Increase in Circular RNA Levels in Myotonic Dystrophy**
Karol Czubak, Katarzyna Taylor, Agnieszka Piasecka, Krzysztof Sobczak, Katarzyna Kozłowska, Anna Philips, Saam Sedehizadeh, J. David Brook, Marzena Wojciechowska and Piotr Kozłowski
- 45 Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing**
Megan Stevens and Sebastian Oltean
- 54 Identification of hsa_circ_0001821 as a Novel Diagnostic Biomarker in Gastric Cancer via Comprehensive Circular RNA Profiling**
Shan Kong, Qian Yang, Chenxue Tang, Tianyi Wang, Xianjuan Shen and Shaoqing Ju
- 66 An Altered Splicing Registry Explains the Differential ExSpeU1-Mediated Rescue of Splicing Mutations Causing Haemophilia A**
Dario Balestra, Iva Maestri, Alessio Branchini, Mattia Ferrarese, Francesco Bernardi and Mirko Pinotti
- 75 To Be or Not to Be: Circular RNAs or mRNAs From Circular DNAs?**
Leire Iparraguirre, Iñigo Prada-Luengo, Birgitte Regenbergh and David Otaegui
- 81 Validation and Classification of Atypical Splicing Variants Associated With Osteogenesis Imperfecta**
Lulu Li, Yixuan Cao, Feiyue Zhao, Bin Mao, Xiuzhi Ren, Yanzhou Wang, Yun Guan, Yi You, Shan Li, Tao Yang and Xiuli Zhao
- 93 RNA-Seq Perspectives to Improve Clinical Diagnosis**
Guillermo Marco-Puche, Sergio Lois, Javier Benítez and Juan Carlos Trivino
- 100 Comprehensive Assessment of BARD1 Messenger Ribonucleic Acid Splicing With Implications for Variant Classification**
Logan C. Walker, Vanessa Lilian Lattimore, Anders Kvist, Petra Kleiblova, Petra Zemankova, Lucy de Jong, George A. R. Wiggins, Christopher Hakkaart, Simone L. Cree, Raquel Behar, Claude Houdayer, kConFab Investigators, Michael T. Parsons, Martin A. Kennedy, Amanda B. Spurdle and Miguel de la Hoya
- 109 Exploring the RNA Gap for Improving Diagnostic Yield in Primary Immunodeficiencies**
Jed J. Lye, Anthony Williams and Diana Baralle

- 119** *Consequences of Making the Inactive Active Through Changes in Antisense Oligonucleotide Chemistries*
Khine Zaw, Kane Greer, May Thandar Aung-Htut, Chalermchai Mitrpant, Rakesh N. Veedu, Sue Fletcher and Steve D. Wilton
- 125** *Circular RNAs: Potential Regulators of Treatment Resistance in Human Cancers*
Shivapriya Jeyaraman, Ezanee Azlina Mohamad Hanif, Nurul Syakima Ab Mutalib, Rahman Jamal and Nadiyah Abu
- 134** *Expression Changes Confirm Genomic Variants Predicted to Result in Allele-Specific, Alternative mRNA Splicing*
Eliseos J. Mucaki, Ben C. Shirley and Peter K. Rogan
- 150** *Dynamic Supraspliceosomes are Assembled on Different Transcripts Regardless of Their Intron Number and Splicing State*
Naama Sebbag-Sznajder, Yehuda Brody, Hodaya Hochberg-Laufer, Yaron Shav-Tal, Joseph Sperling and Ruth Sperling
- 164** *Innovative Therapeutic and Delivery Approaches Using Nanotechnology to Correct Splicing Defects Underlying Disease*
Marc Suñé-Pou, María J. Limeres, Cristina Moreno-Castro, Cristina Hernández-Munain, Josep M. Suñé-Negre, María L. Cuestas and Carlos Suñé
- 186** *Contribution of mRNA Splicing to Mismatch Repair Gene Sequence Variant Interpretation*
Bryony A. Thompson, Rhiannon Walters, Michael T. Parsons, Troy Dumenil, Mark Drost, Yvonne Tiersma, Noralane M. Lindor, Sean V. Tavtigian, Niels de Wind, Amanda B. Spurdle and the InSiGHT Variant Interpretation Committee



Editorial: RNA Splicing and Backsplicing: Disease and Therapy

Rosanna Asselta^{1,2*}, Stefano Duga^{1,2*}, Eladio Andrés Velasco^{3,4} and Emanuele Buratti⁵

¹ Department of Biomedical Sciences, Humanitas University, Milan, Italy, ² Humanitas Clinical and Research Center, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy, ³ Splicing and Genetic Susceptibility to Cancer, Instituto de Biología y Genética Molecular, Valladolid, Spain, ⁴ Consejo Superior de Investigaciones Científicas (CSIC-UVa), Valladolid, Spain, ⁵ International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

Keywords: splicing, backsplicing, RNA processing, circRNA, therapy

Editorial on the Research Topic

RNA Splicing and Backsplicing: Disease and Therapy

Splicing has been extensively studied in recent years both under physiological and pathological conditions. In particular, high-throughput RNA sequencing has allowed a much deeper knowledge on the breadth of alternative splicing in gene expression regulation. Besides the multiplicity of transcripts originating from “conventional” linear splicing, an additional layer of complexity is provided by backsplicing, mostly occurring at annotated exon boundaries, which produces circular RNAs (circRNAs). These are covalently closed RNA rings particularly stable compared to their linear counterparts because they are resistant to exonucleolytic decay; besides being potential therapeutic agents and targets, they also represent attractive biomarkers. In parallel, the increasing data from human genomes are providing, for the first time, population-wide landscapes of genetic variants potentially impacting on splicing, both in monogenic and complex diseases.

Splicing alterations have been particularly studied in autoimmune diseases, where they can be responsible for the generation of autoantigens, and in cancer, where the neo-antigens originating from splicing derangement can impact on tumor immunogenicity and have important consequences on the efficacy of immunotherapy. Moreover, the generation of alternatively spliced isoforms was shown to be implicated in drug resistance to chemotherapy. As a result, a number of approaches were developed to modulate splicing both to correct splicing mutations and to promote/silence specific splicing events as potential therapies, including antisense oligonucleotides, modified U1 snRNAs, small molecules acting on splicing, and trans-splicing. These post-transcriptional interventions have substantial advantages over traditional gene therapy, including no need to deliver large DNA constructs and no concerns regarding the tissue specificity and expression level of the transgene.

In view of these many implications in human diseases and the tremendous efforts aimed at translating the molecular understanding of splicing into the clinic, the purpose of this Research Topic was to provide an overview on new data on RNA splicing and backsplicing in disease and therapy. The articles in this topic range from understanding the mechanisms of splicing regulation, to evaluating the impact of splicing isoforms and circRNA detection on disease diagnosis and prognosis, up to possible therapeutic applications of splicing correction.

SPlicing REGULATION AND ITS IMPACT ON GENE EXPRESSION

Despite decades of molecular studies, the splicing machinery is still only partially understood in its fine mechanistic details, and the work by Sebbag-Sznajder et al. points to the role of

OPEN ACCESS

Edited by:

William C. Cho,
QEH, Hong Kong

Reviewed by:

Dario Balestra,
University of Ferrara, Italy
Maria Paola Paronetto,
Foro Italico University of Rome, Italy

*Correspondence:

Rosanna Asselta
rosanna.asselta@hunimed.eu
Stefano Duga
stefano.duga@hunimed.eu

Specialty section:

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

Received: 06 November 2020

Accepted: 20 November 2020

Published: 08 December 2020

Citation:

Asselta R, Duga S, Velasco EA and
Buratti E (2020) Editorial: RNA Splicing
and Backsplicing: Disease and
Therapy. *Front. Genet.* 11:626835.
doi: 10.3389/fgene.2020.626835

dynamic supraspliceosomes in the splicing regulation of all pre-mRNAs. These gigantic structures (21 MDa) package RNAPII transcribed pre-mRNAs into complexes composed of four native spliceosomes connected by the transcript. Supraspliceosomes also contain pre-mRNA processing factors (e.g., cap-binding proteins, 3'-end processing components, and the ADAR1 and ADAR2 editing enzymes). The supraspliceosome thus emerges as a principal regulator of multiple pre-mRNA processing steps (Sebbag-Sznajder et al.).

Concerning specifically the molecular mechanisms involved in splicing regulation, it has been recently demonstrated that intragenic elements within pre-mRNA can promote RNA structures that potentially activate the RNA-dependent protein kinase PKR, which in turn induces nuclear eIF2 α phosphorylation and thereby strongly enhances mRNA splicing efficiency. This mechanism has been extensively studied for the tumor necrosis factor- α gene and for fetal/adult globin genes. Knowledge of these regulatory mechanisms is crucial to interpret the functional consequences of genetic variations mapping in these intragenic regulatory elements. This may eventually help explaining the impact on splicing of numerous human β -thalassemia mutations (Kaempfer et al.).

Splicing regulation can also have dramatic effects on the downstream events, as exemplified by the alternative splicing event involving exon 2 of Bcl-x, which can result in the production of the anti-apoptotic Bcl-xL or the pro-apoptotic Bcl-xS isoforms. The control of the Bcl-xL/Bcl-xS splicing ratio, reviewed by Stevens and Oltean, is provided by serine/arginine-rich (SR) proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs), transcription factors, and cytokines, and represents a clear example of the importance of splicing modulation (Stevens and Oltean).

ANNOTATION AND FUNCTIONAL INTERPRETATION OF VARIANTS IMPACTING ON SPLICING

A precise knowledge on the mechanism of splicing is also essential for the functional annotation of variants in order to correctly interpret genetic variation and its relationship with human diseases. Regarding this issue, two reviews in our Topic address the importance to incorporate an analysis on the consequences of genetic variations on RNA metabolism to improve clinical diagnosis (Marco-Puche et al.; Lye et al.). Indeed, it is estimated that RNA sequencing would increase the diagnostic rate by up to 10–35%, even though the dynamic nature of the transcriptome, which changes according to tissue type, cellular conditions, and environmental factors, makes this approach more complicated than classic DNA sequencing. Significant advances are being made in bioinformatics to define, homogenize, and monitor the transcriptomic information in order to endure reproducibility and repeatability, which are mandatory for clinical utility of these data. Lye et al. propose the introduction of RNAseq-based analysis for patients who have a clinical presentation of primary immunodeficiency, but despite having undergone whole-exome/whole-genome

sequencing, remain undiagnosed (Lye et al.). Thompson et al. investigated the contribution of splicing-assay data to the classification of mismatch-repair (MMR) gene sequence variants (Thompson et al.).

An important aspect when considering the impact of genetic variation on RNA metabolism is how we validate and classify variants. In this specific area, three manuscripts of our Research Topic address this problem in relation to osteogenesis imperfecta and to the risk of breast cancer, providing useful information on the consequences of atypical splicing variants and improving the clinical interpretation of variants of *COL1A1/COL1A2* and *BRCA2* genes through systematic functional assays in splicing reporter minigenes (Li et al.; Fraile-Bethencourt et al.). Likewise, Walker et al. showed that the comprehensive assessment of alternative splicing events of the breast/ovarian cancer gene *BARD1* is a valuable strategy for the classification of spliceogenic and truncating variants (Walker et al.).

Mucaki et al. evaluated the quality of *in-silico* predictions by information-theory-based variant analysis on the effect of nucleotide substitutions detected in different cancers on splicing. In this study, RT-PCR and RNAseq data confirmed in most cases predictions even though not all events were detected by both techniques stressing the importance of multiple approaches in functional validation.

CircRNAs AS BIOMARKERS IN HUMAN DISEASES

In recent years, there has been a growing interest in circRNAs as potential players in human diseases, especially cancer, and as useful clinical diagnostic markers. In this frame, a comprehensive circRNA profiling, described in the work of Kong et al., led to the discovery of cancer-specific circRNAs in gastric cancer.

Even though in most cases circRNAs are considered the product of backsplicing, at least a fraction of them were shown to originate from the transcription of circular DNAs of chromosomal origin, as discussed in the review by Iparraquirre et al.

A role for circRNAs in different types of treatment resistance has also emerged, and the mini-review from Jeyaraman et al. focuses on the possible role of circRNAs as regulators of treatment resistance in human cancers based on their regulatory role on specific cancer-related networks (Jeyaraman et al.).

Finally, an extensive dysregulation of circRNA biogenesis, leading a global increase in circRNA levels, was found in patients with myotonic dystrophy and circRNA levels were found to be associated with muscle weakness and alternative-splicing changes (Czubak et al.).

THERAPEUTIC APPROACHES EXPLOITING SPLICING CORRECTION

The precise knowledge on the mechanisms of splicing in physiology and disease has been instrumental to design novel therapeutic strategies for splicing-derived pathologies, which are nicely reviewed by Suñé-Pou et al. In their review, they focus on

nanotechnology-based gene delivery strategies to overcome the challenges and barriers facing nucleic acid-based therapeutics, which still represent a major obstacle to the clinical translation of splicing-correction approaches (Suñé-Pou et al.).

Among the different strategies explored to correct splicing, the use of modified spliceosomal U1snRNAs has demonstrated to successfully correct splicing mutations in several cellular and mouse models of human disease. Concerning coagulation factor VIII, a number of splicing defects leading to hemophilia A were analyzed by a mini-gene expression approach to experimentally determine the best modified U1snRNA variant able to correct multiple hemophilia A-causing mutations by re-directing the use of the proper 5' splice site (Balestra et al.).

Interfering with splicing for therapeutic purposes does not only involve restoring aberrant splicing but, in some cases, inducing aberrant splicing (exon skipping) can be exploited to exclude a mutated in-frame exon from the mature transcript. This is the case of the *DMD* gene that, when mutated, is responsible for Duchenne (DMD) and Becker (BMD) muscular dystrophies. The observation that most mutations producing internally truncated but partially functional protein are associated with the milder BMD phenotype suggested the idea to induce targeted exon skipping as a treatment strategy for severe DMD. This has been obtained using antisense oligonucleotides (AO) designed to anneal to splicing sites. Different strategies to improve AO potency have been explored, in order to improve cellular uptake or increase stability and specificity of AOs, such as the use of 2'-O-methyl (2'-OMe) or locked nucleic acid (LNA), which increase binding affinity and resistance against nuclease degradation. In their work, Zaw et al., by testing a set of AOs targeting exons 16, 23, and 51 of human *DMD*, showed that the incorporation of LNAs into 2'-OMe antisense sequences increases their potency as steric blockers of

splicing. However, this increased potency came at the price of the activation of alternative cryptic splice sites, suggesting the need to carefully check for possible off-target effects when using this exon-skipping inducing molecules (Zaw et al.).

In summary, the articles presented in this Research Topic illuminate several aspects of RNA splicing and backspacing processes, ranging from their role in the regulation of gene expression, up to their involvement in diseases and related therapeutics/biomarkers. The next few years promise to shed further light on these aspects. In particular, we expect to have a comprehensive catalog of alternative-splicing/backspacing genetic variants in different human populations, along with expression levels of the associated relevant genes (i.e., a complete catalog of splicing quantitative trait loci, sQTLs, as well as circQTLs in different tissues). This will be instrumental for improving both disease diagnosis and patient treatments, giving even more added value for advocating a precision medicine approach.

AUTHOR CONTRIBUTIONS

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

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

Copyright © 2020 Asselta, Duga, Velasco and Buratti. 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.



Control of mRNA Splicing by Intragenic RNA Activators of Stress Signaling: Potential Implications for Human Disease

Raymond Kaempfer*, Lena Ilan, Smadar Cohen-Chalamish, Orli Turgeman, Lise Sarah Namer and Farhat Osman

Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

OPEN ACCESS

Edited by:

Rosanna Asselta,
Humanitas University, Italy

Reviewed by:

Tohru Yoshihisa,
University of Hyogo, Japan
Rahul N. Kanadia,
University of Connecticut,
United States

*Correspondence:

Raymond Kaempfer
kaempfer@hebrew.edu

Specialty section:

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

Received: 28 February 2019

Accepted: 30 April 2019

Published: 14 May 2019

Citation:

Kaempfer R, Ilan L,
Cohen-Chalamish S, Turgeman O,
Namer LS and Osman F (2019)
Control of mRNA Splicing by
Intragenic RNA Activators of Stress
Signaling: Potential Implications
for Human Disease.
Front. Genet. 10:464.
doi: 10.3389/fgene.2019.00464

A critical step in the cellular stress response is transient activation of the RNA-dependent protein kinase PKR by double-helical RNA, resulting in down-regulation of protein synthesis through phosphorylation of the α chain of translation initiation factor eIF2, a major PKR substrate. However, intragenic elements of 100–200 nucleotides in length within primary transcripts of cellular genes, exemplified by the *tumor necrosis factor* (*TNF*)- α gene and fetal and adult *globin* genes, are capable of forming RNA structures that potently activate PKR and thereby strongly enhance mRNA splicing efficiency. By inducing nuclear eIF2 α phosphorylation, these PKR activator elements enable highly efficient early spliceosome assembly yet do not impair translation of the mature spliced mRNA. The *TNF*- α RNA activator of PKR folds into a compact pseudoknot that is highly conserved within the phylogeny. Upon excision of β -*globin* first intron, the RNA activator of PKR, located in exon 1, is silenced through strand displacement by a short sequence within exon 2, restricting thereby the ability to activate PKR to the splicing process without impeding subsequent synthesis of β -globin essential for survival. This activator/silencer mechanism likewise controls splicing of α -*globin* pre-mRNA, but the exonic locations of PKR activator and silencer sequences are reversed, demonstrating evolutionary flexibility. Impaired splicing efficiency may underlie numerous human β -thalassemia mutations that map to the β -*globin* RNA activator of PKR or its silencer. Even where such mutations change the encoded amino acid sequence during subsequent translation, they carry the potential of first impairing PKR-dependent mRNA splicing or shutoff of PKR activation needed for optimal translation.

Keywords: mRNA splicing control, intragenic RNA activators of PKR, activation of PKR, eIF2 α phosphorylation, PKR silencer elements, *TNF*- α gene, β -*globin* gene, human β -thalassemia mutations

Abbreviations: 2-APRE, 2-aminopurine response element; PKR, protein kinase RNA-activated; eIF2 α , eukaryotic initiation factor 2 α -chain; TNF, tumor necrosis factor; IFN, interferon; nt, nucleotide; PBMC, peripheral blood mononuclear cells; UTR, untranslated region.

INTRODUCTION

Phosphorylation of the α -chain of eukaryotic translation initiation factor 2 (eIF2 α) is critical for mounting the integrated cellular stress response (Harding et al., 2003; Muaddi et al., 2010). Transient phosphorylation of eIF2 α blocks GDP/GTP exchange needed for recycling of eIF2 between rounds of protein synthesis, inducing translational repression (Sonenberg and Hinnebusch, 2009). The RNA-dependent protein kinase PKR is a prominent eIF2 α kinase having a major role in the IFN-mediated antiviral response. IFNs, including IFN- γ , induce high levels of *PKR* gene transcription in the cell (Stark et al., 1998). To become activated, PKR must undergo ATP-dependent *trans*-autophosphorylation upon engaging, through its tandem RNA binding motifs, double-stranded RNA generated during virus replication (Meurs et al., 1990). Highly ordered double-stranded RNA structures rather than specific sequences are needed to activate PKR (Bevilacqua and Cech, 1996). Once activated by double-stranded RNA, PKR will phosphorylate eIF2 α , blocking translation and virus spread from infected cells (Stark et al., 1998).

We review here the discovery and mode of action of a novel class of regulatory RNA elements inside cellular genes that activate PKR to control thereby not only their translation but in particular, enhance their mRNA splicing. Once transcribed into single-stranded RNA, these short non-coding elements fold into structures that act in *cis* to potentially activate PKR, rendering splicing highly efficient (Osman et al., 1999; Ilan et al., 2017; Namer et al., 2017) or repressing translation of the encoded mRNA (Ben-Asouli et al., 2002; Cohen-Chalamish et al., 2009), in each case by inducing eIF2 α phosphorylation. We address potential implications of these RNA elements for human disease.

REGULATION OF GENE EXPRESSION BY INTRAGENIC ELEMENTS THAT ACTIVATE PKR

Linear double-stranded RNA, generated in the course of virus infection, was considered to be the classical activator of PKR. That notion was shattered by the discovery of short elements within cellular genes that once transcribed, fold into RNA structures capable of activating PKR even more effectively and use this property to control gene expression. Thus, human *IFN- γ* mRNA contains a 5'-terminal 203-nt element that folds into a pseudoknot that potently activates PKR, inducing thereby eIF2 α phosphorylation and attenuating its own translation by an order of magnitude (Ben-Asouli et al., 2002; Cohen-Chalamish et al., 2009). This negative feedback loop prevents induction of pathological hyper-inflammation by limiting production of IFN- γ , a prominent inflammatory cytokine (Ben-Asouli et al., 2002). This intragenic element also couples *IFN- γ* mRNA translation to the level of PKR in the cell (Ben-Asouli et al., 2002). Extensive mutational analysis combined with structure probing showed that the RNA activator of PKR is denatured by ribosome passage and undergoes dynamic refolding to allow

PKR activation in the course of translation (Cohen-Chalamish et al., 2009). Because both activation of PKR and phosphorylation of eIF2 α substrate are transient events, followed promptly by dephosphorylation that inactivates PKR while restoring eIF2 α activity, intragenic RNA activators of PKR function locally as *cis*-acting control elements (Osman et al., 1999; Ben-Asouli et al., 2002; Namer et al., 2017).

TNF- α mRNA SPLICING DEPENDS ON ACTIVATION OF PKR AND PHOSPHORYLATION OF ITS eIF2 α SUBSTRATE

The inflammatory cytokine *TNF- α* is not only critical for protective immunity and the anti-tumor response but also a major mediator of inflammatory diseases. *TNF- α* is expressed promptly during the immune response, *TNF- α* mRNA levels becoming maximal within 3 h in stimulated human PBMC (Jarrous et al., 1996). To achieve such efficient expression, splicing of *TNF- α* mRNA uses activation of PKR. The adenine analog 2-aminopurine, a competitive inhibitor of ATP in binding kinases, especially PKR, blocks splicing of all three *TNF- α* introns (Jarrous et al., 1996). Splicing of *TNF- α* pre-mRNA is controlled by the 104-nt 2-APRE located within the 3'-UTR (Figure 1A; Osman et al., 1999). This *cis*-acting RNA element activates PKR more potently than does double-stranded RNA and enhances *TNF- α* mRNA splicing by over an order of magnitude when PKR expression is increased (Osman et al., 1999). Mutational analysis, including compensatory mutations that restore base pairing and secondary structure of RNA, showed that the 2-APRE renders nuclear splicing of *TNF- α* pre-mRNA not only strictly dependent on PKR activation but also highly efficient, yet does not cause translational repression (Osman et al., 1999; Namer et al., 2017). In contrast to *TNF- α* , the closely related *TNF- β* (lymphotoxin) gene does not harbor an intragenic activator of PKR and its mRNA is spliced sluggishly; yet, upon transposition of the *TNF- α* element into the *TNF- β* 3'-UTR, splicing of *TNF- β* pre-mRNA became as efficient as that of *TNF- α* , showing that the 2-APRE functions as an autonomous splicing control element (Osman et al., 1999; Namer et al., 2017).

Protein kinase RNA-activated activation requires its homodimerization on the activating RNA to permit *trans*-autophosphorylation leading to kinase activation (Zhang et al., 2001; Dey et al., 2005). Given that activation of PKR requires at least 33 and optimally 80 base pairs of double-helical RNA (Manche et al., 1992; Bevilacqua and Cech, 1996), how could the 2-APRE, having only 104 nt, activate PKR so potently? Extensive genetic analysis, validated by gain-of-function mutations, revealed that the *TNF- α* RNA activator of PKR folds into a compact pseudoknot that constrains the RNA into two double-helical stacks with parallel axes (Figure 1A), each long enough to bind a PKR monomer, promoting efficient kinase dimerization enabling activation (Namer et al., 2017). The *TNF- α* pseudoknot is highly conserved in the phylogeny over 400 million years, from teleost fish to humans. Indeed,

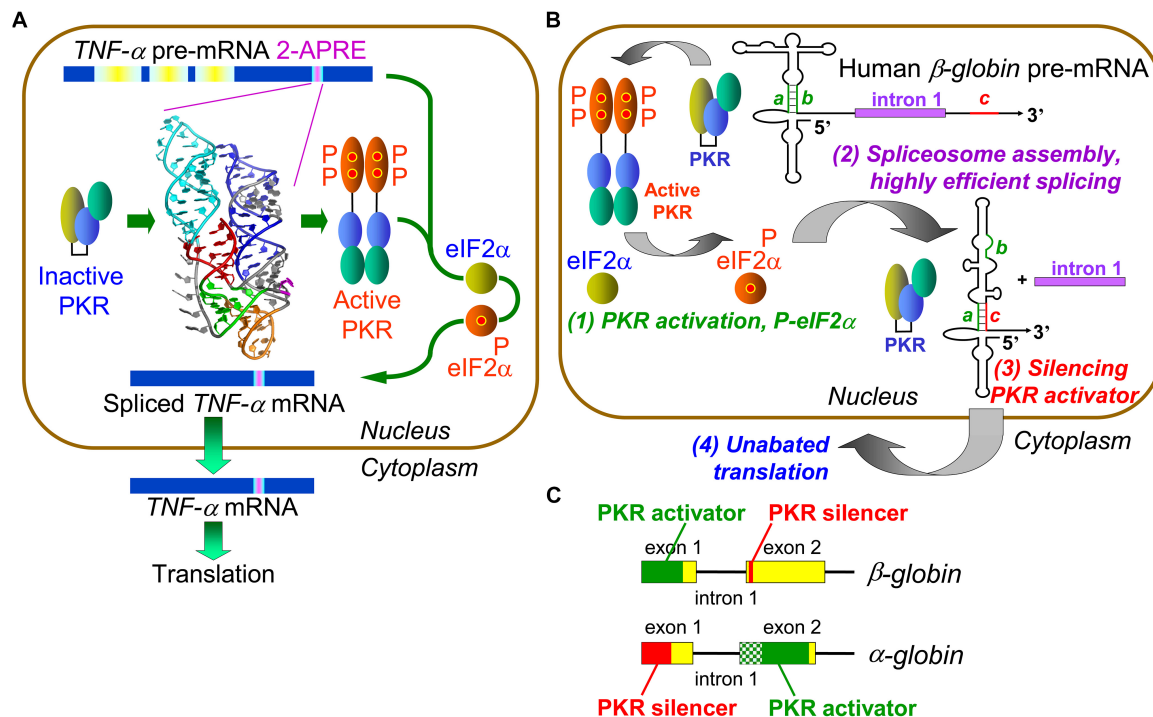


FIGURE 1 | Control of mRNA splicing by intragenic RNA activators of PKR and eIF2 α phosphorylation. **(A)** An RNA element in *TNF- α* pre-mRNA controls splicing by activating PKR and thereby inducing eIF2 α phosphorylation. The 104-nt *TNF- α* activator of PKR (2-APRE), located within the 3'-UTR, potentially activates PKR through its pseudoknot structure that constrains the RNA into two parallel helices, each capable of binding a PKR monomer, facilitating PKR dimerization on RNA needed for kinase activation. Once PKR is activated, it must induce eIF-2 α phosphorylation to enable efficient splicing of *TNF- α* mRNA. **(B)** Regulation of *β -globin* pre-mRNA splicing by an intragenic RNA activator of PKR and silencing of the PKR activator upon intron excision. (1) Within the RNA activator of PKR, consisting of 5'-terminal 124 nt within exon 1, 5 bp helix *a-b* (green) is critical for activity. Once PKR is activated, it must induce eIF-2 α phosphorylation to enable efficient spliceosome assembly and mRNA splicing (2). Upon excision of intron 1, sequence *c* located near the start of exon 2 (red) displaces strand *b* from strand *a*, a rearrangement that silences the ability of mature *β -globin* mRNA to activate PKR (3). This renders activation of PKR transient, serving solely to promote splicing yet allowing for unimpeded synthesis of *β -globin* protein (4) (after Kaempfer et al., 2018). **(C)** Elements that activate PKR or silence PKR activation map into opposite exons in α - and β -*globin* pre-mRNA. The core of the α -*globin* RNA activator of PKR is shown in solid green within exon 2; maximal PKR activation also requires upstream RNA sequence shown in shaded green.

turbot 2-APRE RNA activates human PKR and enhances human *TNF- β* mRNA splicing as effectively as does the human element (Namer et al., 2017).

Local activation of PKR not only enhances *TNF- α* mRNA splicing but also increases protein yield correspondingly, without repressing translation (Namer et al., 2017). Surprisingly, we discovered that PKR activation promotes efficient *TNF- α* mRNA splicing by inducing eIF2 α phosphorylation (Figure 1A). Expression of non-phosphorylatable mutant eIF2 α abrogated PKR-dependent splicing. Phosphorylation of eIF2 α is not only strictly needed but also sufficient to achieve highly efficient splicing. Blocking rapid dephosphorylation of eIF2 α with salubrinal, which increases phospho-eIF2 α globally in the cell (Boyce et al., 2005), sufficed to raise splicing the efficiency of *TNF- β* pre-mRNA to that of *TNF- α* pre-mRNA (Namer et al., 2017). eIF2 α phosphorylation upregulates *TNF- α* mRNA splicing in human PBMC, demonstrating its physiological relevance. Therefore, stress-induced PKR-mediated eIF2 α phosphorylation has not only a major role in down-regulating translation but also plays a key positive role in rendering splicing highly efficient.

INTRAGENIC RNA ACTIVATORS OF PKR CONTROL GLOBIN GENE EXPRESSION AT mRNA SPLICING

To analyze the molecular mechanism underlying highly efficient splicing of *TNF- α* mRNA induced by its intragenic RNA activator of PKR and mediated by eIF2 α phosphorylation, we offered *in vitro* transcribed *TNF- α* precursor RNA as substrate for splicing in HeLa cell nuclear extract. That attempt failed, owing to prompt and complete degradation of *TNF- α* pre-mRNA. However, it led to our discovery that splicing of *β -globin* exon1-intron1-exon2 template, serving as positive control for splicing (Krainer et al., 1984), also depends strictly on the activation of PKR (Ilan et al., 2017). This came as a surprise, given that *globin* gene expression has long served as a paradigm for eukaryotic gene regulation. Indeed, splicing of human α -*globin*, β -*globin* as well as fetal γ -*globin* pre-mRNAs depends heavily on PKR activation induced by intragenic RNA activator elements (Ilan et al., 2017). Hence, PKR activation is used more broadly within the human genome beyond inflammatory cytokine genes, to control mRNA

splicing. Excision of β -globin intron 1, the first splicing event, was blocked by anti-PKR antibodies as well as by PKR depletion, where it could be restored with recombinant PKR.

The β -globin RNA activator of PKR maps into the first exon (Figure 1B, step 1); mutation of short helix a - b in the β -globin activator severely impairs both PKR activation and mRNA splicing. Efficient splicing of each of α -, β - and γ -globin pre-mRNA species depends strictly on activation of PKR and nuclear eIF2 α phosphorylation and is inhibited by non-phosphorylatable mutant eIF2 α or anti-phospho-eIF2 α antibodies (Ilan et al., 2017). Activation of PKR and eIF2 α phosphorylation are required at an early step in β -globin spliceosome assembly, formation of Complex A (Figure 1B, step 2). As shown for β -globin pre-mRNA, activation of PKR and phosphorylation of eIF2 α mediate splicing not only *in vitro* but also in intact cells (Ilan et al., 2017).

INTRAGENIC RNA-MEDIATED SILENCING OF PKR ACTIVATORS UPON SPLICING

The RNA activator of PKR is contained within β -globin exon 1 and thus maintained in spliced mRNA, where it could strongly down-regulate translation as shown for *IFN- γ* mRNA (Ben-Asouli et al., 2002; Cohen-Chalamish et al., 2009), creating a paradox. Yet, during erythroid development, globin mRNA is translated massively, reaching 95% of total protein in reticulocytes as compared to <0.1% in proerythroblasts (Nienhuis and Benz, 1977). How is maximal translation achieved? Excision of β -globin first intron juxtaposes short 5-nucleotide sequence c , located near the start of exon 2, to exon 1, inducing strand displacement within exon 1 that destroys helix a - b at the core of the PKR activator, resulting in silencing of the activator once β -globin mRNA is spliced (Figure 1B, step 3 and Figure 1C). Mutation of either strand a or c abrogated silencing whereas compensatory mutation restored it (Ilan et al., 2017). Splicing of α -globin pre-mRNA is regulated similarly except that locations of PKR activator and silencer are reversed between exons 2 and 1, demonstrating evolutionary flexibility in control of PKR activation during and upon splicing (Figure 1C). The silencing mechanism allows for highly efficient PKR-dependent splicing, followed promptly by shutoff of PKR activation, to permit undisturbed, maximal translation of the spliced mRNA product (Ilan et al., 2017). This mechanism assures that the ability to activate PKR remains transient, serving only to enable efficient splicing, without hindering globin synthesis (Figure 1B, step 4).

INTRAGENIC RNA ELEMENTS THAT ACTIVATE PKR OR SILENCE PKR ACTIVATORS ARE POTENTIAL SOURCES OF HUMAN DISEASE

Protein kinase RNA-activated activator and silencer RNA structures were defined for the human β -globin gene (HBB) by truncation, mutational analysis, and in-line probing of the RNA

(Figure 1B; Ilan et al., 2017). Figure 2 depicts the resulting RNA secondary structure. The activator of PKR is comprised of nucleotides 1–124 from the 5' end; truncation of only a few nucleotides from either side sufficed to abrogate PKR activation. The AUG start codon, located at positions 51–53, forms part of strand a that generates the helix at the core of the PKR activator. Replacement of CACCA in strand a , including A of the start codon, by complementary nucleotides had a pronounced negative effect on splicing efficiency. Replacement of CGUGG in opposite strand b , which includes the sequence that engages the AUG codon, by complementary nucleotides largely abrogated splicing efficiency both in cells and *in vitro*, whereas efficient splicing was restored by ab double mutation (Ilan et al., 2017).

Inspection of the human gene mutation data base (HBB¹) shows that numerous human β -thalassemia mutations map to the β -globin RNA activator of PKR or to its silencer

¹<http://www.hgmd.org>

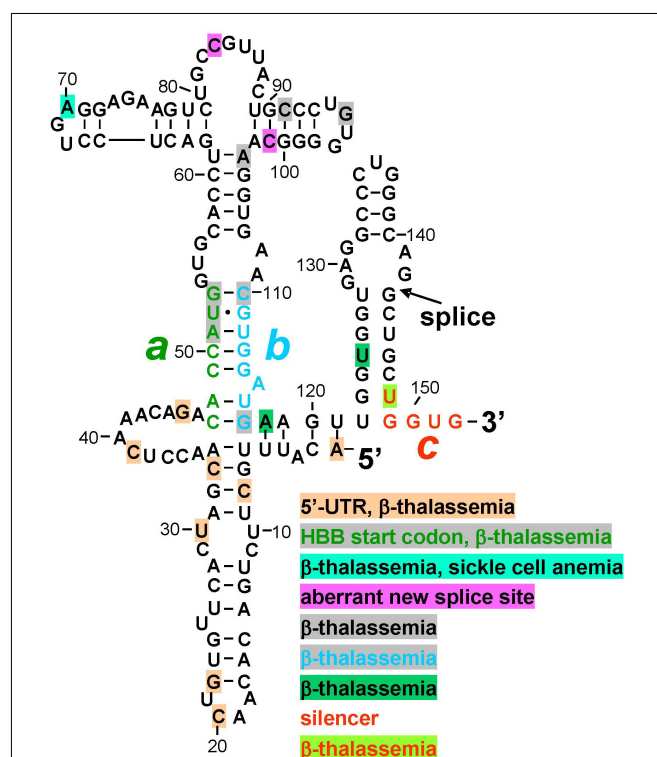


FIGURE 2 | Mutations in human β -globin RNA activator of PKR and silencer of PKR activation are associated with β -thalassemia. Structure of the RNA activator of PKR (nucleotides 1–124), determined by in-line probing and mutagenesis, with a key role for helix strands a (green) and b (cyan). Strand a includes the AUG start codon. Position of first splice junction is shown, as is start of exon 2 containing PKR silencer c (red), upon excision of intron 1 but before displacement of strand b by sequence c validated by in-line probing and mutagenesis (see Figure 1B). Nucleotide mutations associated with β -thalassemia in the human gene mutation database (<http://www.hgmd.org>), are marked by shading in various colors, see text. HBB, human β -globin gene.

(**Figure 2**). Thus, regulatory mutations were reported within the 5'-UTR, many without mechanism. However, C33G mutation reduced the β -globin transcript level (Ho et al., 1996) while A1G mutation led, in mouse β -major globin, to over twofold decreased mRNA expression (Myers et al., 1986); splicing was not analyzed. Other β -thalassemia mutations map to the AUG initiation codon within strand *a*, to the bifurcation loop structure, to nucleotides in strand *b* that form G-C base pairs with strand *a* to generate the helix indispensable for PKR activation, as well as to nucleotides adjoining strand *b* and the 3' end of the PKR activator structure (**Figure 2**). Within silencer sequence *c*, U148A mutation destabilizes base pairing with strand *a*. Because splicing precedes translation, the effect of these mutations may be manifested at the level of PKR activation needed for high splicing efficiency, or at shutoff of PKR activation directly upon splicing, even before they can affect the sequence of the translated β -globin protein product that is more readily diagnosed in patients and traditionally reported on the data base. Thus, although mutation of the AUG start codon can severely impact translation, this codon also has a dual function in controlling splicing, being located at the heart of the RNA activator of PKR and base pairing with the silencer.

Splicing-defective mutations reported within the β -globin PKR activator domain (**Figure 2**) create aberrant splice donor sites that alter protein sequence; aberrant splice donor site mutations are lacking for downstream β -globin exons 2 and 3.

Minimal sequences encoding the α -globin RNA activator of PKR and silencer (**Figure 1C**) were defined thus far only by truncation analysis (Ilán et al., 2017); their RNA structure remains to be determined. Therefore, it is too early to perform a similar analysis for α -globin. Nonetheless, numerous mutations characterized as leading to α -thalassemia, hemolytic anemia, or variant α -globin proteins (HBA1, HBA2; see footnote 1) map to the PKR activator and silencer domains as delineated at present.

Following the pattern for adult β -globin, the RNA activator of PKR of γ -globin, the fetal form of β -globin, is located within the first exon and γ -globin mRNA splicing is strictly dependent on PKR activation and eIF2 α phosphorylation (Ilán et al., 2017), but the structure of the γ -globin PKR activator element was not yet analyzed.

Mutational analysis of the *TNF- α* RNA activator of PKR (2-APRE, **Figure 1A**) demonstrated its exquisite sensitivity to mutations, even to a single nucleotide change or base pair inversion, in activating PKR and rendering splicing highly efficient (Namer et al., 2017). Length and nature of the *TNF- α* 3'-UTR impart great lability to the primary transcript, rendering detection and analysis of splicing defects in human patients difficult. Assay of *TNF- α* splicing efficiency necessarily must be done in primary PBMC or transfected cells, using real-time polymerase chain reaction or ribonuclease protection analysis (Namer et al., 2017). It is thus no surprise that among mutations reported as yielding a *TNF- α* phenotype, the human gene mutation

data base lacks thus far mutations mapping into the 2-APRE, remote from the coding region. Focus has been on control of *TNF- α* mRNA translation by microRNAs (TNFA; see footnote 1).

FUTURE PERSPECTIVES

The discovery of intragenic elements that once transcribed, control splicing by activating PKR in the nucleus or by silencing the ability to activate PKR, adds a new dimension to the analysis and interpretation of human gene mutations. As shown for the RNA activators of PKR within *IFN- γ* mRNA and *TNF- α* pre-mRNA, even single-nucleotide substitutions or the inversion of a single base pair can lead to loss of the ability of the RNA element to activate PKR (Cohen-Chalamish et al., 2009; Namer et al., 2017). This demonstrates the exquisite sensitivity of the PKR protein molecule to the RNA structure that it must interact with in order to achieve kinase activation, resulting in eIF2 α phosphorylation that in turn enhances mRNA splicing efficiency. As shown for β -globin pre-mRNA, on the other hand, silencing of the ability of its RNA activator of PKR to activate the kinase, through effective base pairing with RNA encoded by a separate silencer element that induces conformational changes within the RNA activator structure, in this case through strand displacement, is likewise sensitive to mutation (Ilán et al., 2017).

Thus, short intragenic RNA elements that activate PKR or that silence PKR activators are not only essential for controlling efficient mRNA splicing but also create potential etiology for human disease. In a broader sense, this novel perspective may account for and/or contribute to the phenotype of gene mutations analyzed hitherto primarily for their effect on protein sequence. Even where such mutations change the encoded amino acid sequence during subsequent translation in the cytoplasm, they also carry the potential of first impairing PKR-dependent mRNA splicing in the nucleus or the shutoff of PKR activation needed for optimal translation. That concept extends to silent mutations and to mutations that alter amino acid sequence without having a major effect on protein function.

AUTHOR CONTRIBUTIONS

OT searched the human gene mutation database. OT and RK analyzed the human mutation data. LI, SC-C, LN, FO, and RK designed and performed the experiments and analyzed the results. RK wrote the manuscript. All authors read and approved the final version of the manuscript for submission.

FUNDING

This work was supported by United States Congressionally Directed Medical Research Programs award (W81XWH-17-1-0647).

REFERENCES

- Ben-Asouli, Y., Banai, Y., Pel-Or, Y., Shir, A., and Kaempfer, R. (2002). Human interferon- γ mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. *Cell* 108, 221–232. doi: 10.1016/s0092-8674(02)00616-5
- Bevilacqua, P. C., and Cech, T. R. (1996). Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry* 35, 9983–9994. doi: 10.1021/bi9607259
- Boyce, M., Bryant, K. F., Jousse, C., Long, K., Harding, H. P., Scheuner, D., et al. (2005). A selective inhibitor of eIF2 α dephosphorylation protects cells from ER stress. *Science* 307, 935–939. doi: 10.1126/science.1101902
- Cohen-Chalamish, S., Hasson, A., Weinberg, D., Namer, L. S., Banai, Y., Osman, F., et al. (2009). Dynamic refolding of IFN- γ mRNA enables it to function as PKR activator and translation template. *Nat. Chem. Biol.* 5, 896–903. doi: 10.1038/nchembio.234
- Dey, M., Cao, C., Dar, A. C., Tamura, T., Ozato, K., Sicheri, F., et al. (2005). Mechanistic link between PKR dimerization, autophosphorylation, and eIF2 α substrate recognition. *Cell* 122, 901–913. doi: 10.1016/j.cell.2005.06.041
- Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calton, M., et al. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol. Cell* 11, 619–633. doi: 10.1016/s1097-2765(03)00105-9
- Ho, P. J., Rochette, J., Fisher, C. A., Wonke, B., Jarvis, M. K., Yardumian, A., et al. (1996). Moderate reduction of beta-globin gene transcript by a novel mutation in the 5' untranslated region: a study of its interaction with other genotypes in two families. *Blood* 87, 1170–1178.
- Ilan, L., Osman, F., Namer, L. S., Eliahu, E., Cohen-Chalamish, S., Ben-Asouli, Y., et al. (2017). PKR activation and eIF2 α phosphorylation mediate human globin mRNA splicing at spliceosome assembly. *Cell Res.* 27, 688–704. doi: 10.1038/cr.2017.39
- Jarrous, N., Osman, F., and Kaempfer, R. (1996). 2-Aminopurine selectively inhibits splicing of tumor necrosis factor alpha mRNA. *Mol. Cell. Biol.* 16, 2814–2822. doi: 10.1128/mcb.16.6.2814
- Kaempfer, R., Namer, L. S., Osman, F., and Ilan, L. (2018). Control of mRNA splicing by noncoding intragenic RNA elements that evoke a cellular stress response. *Int. J. Biochem. Cell. Biol.* 105, 20–23. doi: 10.1016/j.biocel.2018.09.021
- Krainer, A. R., Maniatis, T., Ruskin, B., and Green, M. R. (1984). Normal and mutant human β -globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36, 993–1005. doi: 10.1016/0092-8674(84)90049-7
- Manche, L., Green, S. R., Schmedt, C., and Mathews, M. B. (1992). Interactions between double-stranded RNA regulators and the protein kinase DAI. *Mol. Cell. Biol.* 12, 5238–5248. doi: 10.1128/mcb.12.11.5238
- Meurs, E., Chong, K., Galabru, J., Thomas, N. S., Kerr, I. M., Williams, B. R., et al. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell* 62, 379–390. doi: 10.1016/0092-8674(90)90374-n
- Muaddi, H., Majumder, M., Peidis, P., Papadakis, A. I., Holcik, M., Scheuner, D., et al. (2010). Phosphorylation of eIF2 α at serine 51 is an important determinant of cell survival and adaptation to glucose deficiency. *Mol. Biol. Cell* 21, 3220–3231. doi: 10.1091/mbc.E10-01-0023
- Myers, R. M., Tilly, K., and Maniatis, T. (1986). Fine structure genetic analysis of a beta-globin promoter. *Science* 232, 613–618. doi: 10.1126/science.3457470
- Namer, L. S., Osman, F., Banai, Y., Masquida, B., Jung, R., and Kaempfer, R. (2017). An ancient pseudoknot in TNF- α pre-mRNA activates PKR, inducing eIF2 α phosphorylation that potentially enhances splicing. *Cell Rep.* 20, 188–200. doi: 10.1016/j.celrep.2017.06.035
- Nienhuis, A. W., and Benz, E. J. (1977). Regulation of hemoglobin synthesis during the development of the red cell. *N. Engl. J. Med.* 297, 1318–1328. doi: 10.1056/NEJM197712152972404
- Osman, F., Jarrous, N., Ben-Asouli, Y., and Kaempfer, R. (1999). A cis-acting element in the 3'-untranslated region of human TNF- α mRNA renders splicing dependent on the activation of protein kinase PKR. *Genes Dev.* 13, 3280–3293. doi: 10.1101/gad.13.24.3280
- Sonenberg, N., and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745. doi: 10.1016/j.cell.2009.01.042
- Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, R. D. (1998). How cells respond to interferons. *Annu. Rev. Biochem.* 67, 227–264. doi: 10.1146/annurev.biochem.67.1.227
- Zhang, F., Romano, P. R., Nagamura-Inoue, T., Tian, B., Dever, T. E., Mathews, M. B., et al. (2001). Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop. *J. Biol. Chem.* 276, 24946–24958. doi: 10.1074/jbc.M102108200

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 Kaempfer, Ilan, Cohen-Chalamish, Turgeman, Namer and Osman. 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.



Minigene Splicing Assays Identify 12 Spliceogenic Variants of *BRCA2* Exons 14 and 15

Eugenia Fraile-Bethencourt¹, Alberto Valenzuela-Palomo¹, Beatriz Díez-Gómez¹, María José Caloca², Susana Gómez-Barrero³ and Eladio A. Velasco^{1*}

¹ Splicing and Genetic Susceptibility to Cancer, Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain,

² Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain, ³ VISAVET-Universidad Complutense de Madrid, Madrid, Spain

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Logan Walker,
University of Otago, New Zealand
Katarzyna Gaweda-Walerych,
Mossakowski Medical Research
Centre (PAN), Poland
Mads Thomassen,
Odense University Hospital, Denmark

*Correspondence:

Eladio A. Velasco
eavelsam@ibgm.uva.es

Specialty section:

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

Received: 05 February 2019

Accepted: 07 May 2019

Published: 28 May 2019

Citation:

Fraile-Bethencourt E,
Valenzuela-Palomo A, Díez-Gómez B,
Caloca MJ, Gómez-Barrero S and
Velasco EA (2019) Minigene Splicing
Assays Identify 12 Spliceogenic
Variants of *BRCA2* Exons 14 and 15.
Front. Genet. 10:503.
doi: 10.3389/fgene.2019.00503

A relevant fraction of *BRCA2* variants is associated with splicing alterations and with an increased risk of hereditary breast and ovarian cancer (HBOC). In this work, we have carried out a thorough study of variants from *BRCA2* exons 14 and 15 reported at mutation databases. A total of 294 variants from exons 14 and 15 and flanking intronic sequences were analyzed with the online splicing tools NNSplice and Human Splicing Finder. Fifty-three out of these 294 variants were selected as candidate splicing variants. All variants but one, were introduced into the minigene MGBR2_ex14-20 (with exons 14–20) by site-directed mutagenesis and assayed in MCF-7 cells. Twelve of the remaining 52 variants (23.1%) impaired splicing at different degrees, yielding from 5 to 100% of aberrant transcripts. Nine variants affected the natural acceptor or donor sites of both exons and three affected putative enhancers or silencers. Fluorescent capillary electrophoresis revealed at least 10 different anomalous transcripts: ▼(E14q5), Δ(E14p10), Δ(E14p246), Δ(E14q256), Δ(E14), Δ(E15p12), Δ(E15p13), Δ(E15p83), Δ(E15) and a 942-nt fragment of unknown structure. All transcripts, except for Δ(E14q256) and Δ(E15p12), are expected to truncate the *BRCA2* protein. Nine variants induced severe splicing aberrations with more than 90% of abnormal transcripts. Thus, according to the guidelines of the American College of Medical Genetics and Genomics, eight variants should be classified as pathogenic (c.7008-2A > T, c.7008-1G > A, c.7435+1G > C, c.7436-2A > T, c.7436-2A > G, c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G), one as likely pathogenic (c.7008-3C > G) and three remain as variants of uncertain clinical significance or VUS (c.7177A > G, c.7447A > G and c.7501C > T). In conclusion, functional assays by minigenes constitute a valuable strategy to primarily check the splicing impact of DNA variants and their clinical interpretation. While bioinformatics predictions of splice site variants were accurate, those of enhancer or silencer variants were poor (only 3/23 spliceogenic variants) which showed weak impacts on splicing (~5–16% of aberrant isoforms). So, the Exonic Splicing Enhancer and Silencer (ESE and ESS, respectively) prediction algorithms require further improvement.

Keywords: breast cancer, *BRCA2*, DNA variants, splicing, hybrid minigenes

INTRODUCTION

Since the discovery of the breast cancer genes *BRCA1* (OMIM #113705) and *BRCA2* (OMIM #600185) (Miki et al., 1994; Wooster et al., 1995), nearly 17,000 different variants of both genes have been recorded at the ClinVar database¹ (date last accessed; November 2018). Germline inactivating variants in *BRCA1* and *BRCA2* confer high lifetime risks of breast and ovarian cancers (Mavaddat et al., 2013). Also, other cancer types, such as prostate, pancreatic and melanoma, are associated with pathogenic variants in these genes (Petrucelli et al., 2013). Despite the high penetrance of *BRCA* pathogenic variants, they are responsible for only ~15–20% of hereditary breast and ovarian cancer (HBOC) (Stratton and Rahman, 2008). In fact, HBOC is a highly genetically heterogeneous disease with about 25 known or proposed susceptibility genes (Nielsen et al., 2016). Apart from the *BRCA* genes, *PALB2* (OMIM #610355), *ATM* (OMIM #607585), and *CHEK2* (OMIM #604373) have a prominent contribution since, in a recent study, more than 30% of pathogenic variants were found in these genes (Buys et al., 2017).

Commonly, the variants are classified attending to their predicted effect on the protein so that truncating variants (frameshift and nonsense) are directly classified as pathogenic, while intronic, missense and synonymous variants are usually considered to be variants of uncertain clinical significance (VUS). In fact, VUS are identified by a relevant proportion of *BRCA* genetic tests (~20%), which hamper genetic counseling and subsequent preventive or therapeutic actions, since risk assessment is solely based on family history (Radice et al., 2011; Eccles et al., 2015; Ricks et al., 2015).

Furthermore, other upstream gene-expression processes, such as transcription or splicing, can be impaired if regulatory motifs are targeted by nucleotide variations (Wang and Cooper, 2007). Splicing is the process by which introns are removed from a pre-mRNA and exons are consecutively joined. This mechanism is performed in the nucleus by the spliceosome, a macrocomplex constituted by 5 small nuclear ribonucleoproteins (snRNPs) and many other associated proteins (De Conti et al., 2012). The spliceosome recognizes in the pre-mRNA specific sequences which define the exons/introns boundaries and other elements needed to carry out the process. These sequences are: the acceptor or 3' splice site (3'ss), the donor or 5' splice site (5'ss), the branch point, the polypyrimidine tract and the auxiliary *cis* sequences known as splicing regulatory elements (SREs) where enhancer or silencer *trans* factors can bind. Therefore, any change in the sequence may disrupt splicing (Cartegni et al., 2002). Splicing variants usually break the 3'ss or 5'ss leading to abnormal splicing events such as exon skipping, alternative site usage or intron retention. However, they may also create new splicing sites or strengthen cryptic ones that would then be recognized. Other mechanism that may alter splicing is the disruption of exonic/intronic splicing enhancers (ESEs/ISEs) or the creation of exonic/intronic splicing silencers (ESSs/ISSs) (Abramowicz and Gos, 2018). Nevertheless, it is extremely difficult to identify active

SREs and predict the impact of the DNA variants on splicing given the low accuracy of SRE-detection softwares. Therefore, splicing variants can induce abnormal transcripts that either introduce premature termination codons (PTC), in-frame loss of essential protein domains or even inclusion of new translated sequences. Consequently, variants with impact on splicing (or *spliceogenic* variants) may be associated with an increased risk of a given disease. This ethiopathogenic mechanism has been so far underestimated, even though some authors have suggested that spliceogenic variants may represent more than 60% of disease-causing mutations (López-Bigas et al., 2005).

Previous studies have shown that a significant number of splicing variants have been detected in *BRCA2* (Spurdle et al., 2008; Rebbeck et al., 2018). In fact, previous results from our group showed that more than 50% of tested variants of *BRCA2* exons 16–27 impaired splicing (Acedo et al., 2012, 2015; Fraile-Bethencourt et al., 2017, 2018). Likewise, at least 24 different *BRCA2* alternative transcripts have been identified. They are helpful to interpret the splicing outcomes of genetic variations (Fackenthal et al., 2016) and suggest a fine regulation of *BRCA2* exon processing. This feature is supported by the fact that several ESE-rich regions have been functionally mapped by exonic deletions throughout most *BRCA2* exons. Thus, these motifs would be involved in precise exon recognition and alternative splicing events (Acedo et al., 2015; Fraile-Bethencourt et al., 2017). Moreover, we showed that functional mapping is an optimal approach that improves ESE-software predictions and facilitates the identification of spliceogenic mutations of this sort of *cis*-elements.

In this work, we have extended our analysis to *BRCA2* exons 14 and 15 by carrying out an in-depth study of candidate spliceogenic variants. We have explored the presence of splicing enhancers in exons 14 and 15 and have undertaken RNA assays of 52 selected variants from both exons.

MATERIALS AND METHODS

Ethical approval for this study was obtained from the Ethics Review Committee of the Hospital Universitario Río Hortega de Valladolid (6/11/2014).

Bioinformatics: Databases and *in silico* Analysis

We collected *BRCA2* variants from the main databases: ClinVar², the BRCA Share Database (UMD³) (Beroud et al., 2016) and the Breast Cancer Information Core (BIC⁴) (Supplementary Table S1). Variants and transcripts were annotated according to the Human Genome Variation Society (HGVS) guidelines on basis of the *BRCA2* GenBank sequence NM000059.1. In order to simplify, we identified transcripts with a shortened code that combines the following symbols (Lopez-Perolio et al., 2019): Δ (skipping of reference exonic sequences), ▼ (inclusion of

¹<https://www.ncbi.nlm.nih.gov/clinvar>

²<https://www.ncbi.nlm.nih.gov/clinvar/>

³<http://www.umd.be/BRCA2/>

⁴<https://research.nhgri.nih.gov/projects/bic/index.shtml>

reference intronic sequences), E (exon), p (acceptor shift), q (donor shift). When necessary, the exact number of skipped or retained nucleotides is indicated. For example, transcript Δ (E14p10) indicates the use of an alternative acceptor site 10-nt downstream that causes a 10-nt deletion.

In silico analysis was made with the online softwares: NNSplice⁵ (Reese et al., 1997) and Human Splicing Finder version 3.1 (HSF⁶) that contain several prediction algorithms of different splicing motifs (Desmet et al., 2009). The following matrices were used: MaxEntScan (MES) (Yeo and Burge, 2004), the HSF branch point detection tool, ESE-finder (Cartegni et al., 2003), the HSF matrices for 9G8 and Tra2 β and the HSF matrix for hnRNPA1. All the analyses were carried out with the default threshold values of NNSplice and HSF (NNSplice, 0.4; MES, 3.0; Branch point – no cut-off; SRE (0–100 scale): SF2/ASF, 72.98; SF2/ASF (IgM – BRCA1), 70.51; SC35, 75.05; SRp40, 78.08; SRp55 73.86; 9G8, 59.245; Tra2 β , 75.964 and hnRNPA1, 65.476.

Minigene Construction and Mutations

The minigene MGBR2_14-20 was built as previously reported (Fraile-Bethencourt et al., 2017). A total of 52 variants and 8 microdeletions were introduced into the wild type (wt) minigene by site-directed mutagenesis with the QuikChange Lightning Kit (Agilent, Santa Clara, CA, United States), following the manufacturer's instructions (**Supplementary Table S2**). All mutant clones were confirmed by sequencing (MacroGen, Madrid, Spain).

MCF-7 Transfections

Approximately 2×10^5 MCF-7 cells (human breast adenocarcinoma cell line) were plated in four-well plates (Nunc, Roskilde, Denmark). They were grown to 90% confluency in 0.5 mL of medium (MEME, 10% fetal bovine serum, 2 mM glutamine, 1% non-essential amino acids and 1% penicillin/streptomycin). Then, 1 μ g of minigene was transfected into MCF-7 cells using low toxicity Lipofectamine (Life Technologies, Carlsbad, CA, United States) in GibcoTM Opti-MemTM medium (Thermo Fisher Scientific, Waltham, MA, United States). Cells were incubated during 48h and then treated with cycloheximide 300 μ g/ml (Sigma-Aldrich, St. Louis, MO, United States) for 4 h to inhibit the nonsense-mediated mRNA decay (NMD). The RNA was purified with the Genematrix Universal RNA Purification Kit (EURx, Gdańsk, Poland) with on-column DNase I digestion.

siRNA Assays

SR proteins were silenced in MCF7 cells by small interfering RNAs (siRNA) against the main SR proteins: SRSF1 (SF2), SRSF2 (SC35), SRSF3 (SRp20), SRSF5 (SRp40), SRSF7 (9G8), SRSF9 (SRp30c), and Tra2 β (**Supplementary Table S3**), using anti-Luciferase siRNA as negative control. Approximately 1.5×10^5 cells were subjected to a two-hit transfection in Optimem medium (Gibco – Life Technologies, Carlsbad, CA, United States) with 3 μ l of Oligofectamine (Thermo Fisher

Scientific) and the specific siRNA at a final concentration of 0.08 μ M on day 2. Then, 2 μ g of the wt minigene were transfected with low toxicity Lipofectamine (Thermo Fisher Scientific) on day 4, and RNA was extracted on day 5. Silencing was confirmed by qPCR using 10 ng of cDNA in 25 μ l reaction (**Supplementary Table S3**). Amplification was made with SG qPCR Master Mix (Eurx, Gdańsk, Poland). Each siRNA/minigene transfection as well as all the qPCR experiments were carried out in duplicate.

RT-PCR and Transcripts Amplification

Retrotranscription was carried out with 400 ng of RNA and the RevertAid First Strand cDNA Synthesis Kit (Life Technologies), using the specific minigene primer RTPSPL3-RV (5'-TGAGGAGTGAATTGGTCGAA-3'). Samples were incubated at 42°C for 1 h, followed by 5 min at 70°C. Transcripts were amplified with Platinum Taq DNA polymerase (Life Technologies) using 40 ng of cDNA and the primers pMAD_607FW (Patent P201231427, CSIC) and RTBR2_ex17RV2 (5'-GGCTTAGGCATCTATTAGCA-3'). PCR consisted of: denaturation step at 94°C for 2 min, followed by 35 cycles 94°C-30 s, 60°C-30 s and 72°C-1 min/kb, and a final extension step at 72°C for 5 min. Transcripts were sequenced at the MacroGen Spain facility.

In order to relatively quantify all transcripts, semi-quantitative fluorescent RT-PCRs were undertaken in triplicate with the primers pMAD_607FW (FAM-labeled) and RTBR2_ex17RV2 and Platinum Taq DNA polymerase (Life Technologies) under standard conditions except that 26 cycles were herein applied (Acedo et al., 2015). FAM-labeled products were run with LIZ-1200 Size Standard at the MacroGen facility and analyzed with the Peak Scanner software V1.0 (Life Technologies). Only peak heights ≥ 50 RFU (Relative Fluorescence Units) were considered.

RESULTS

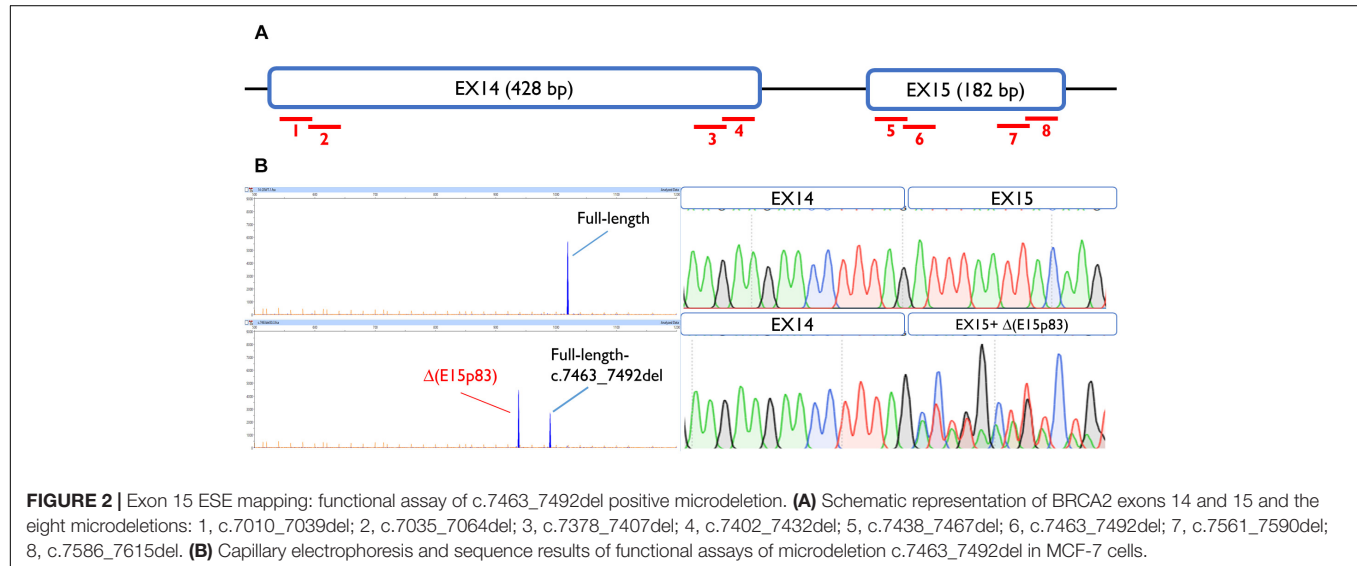
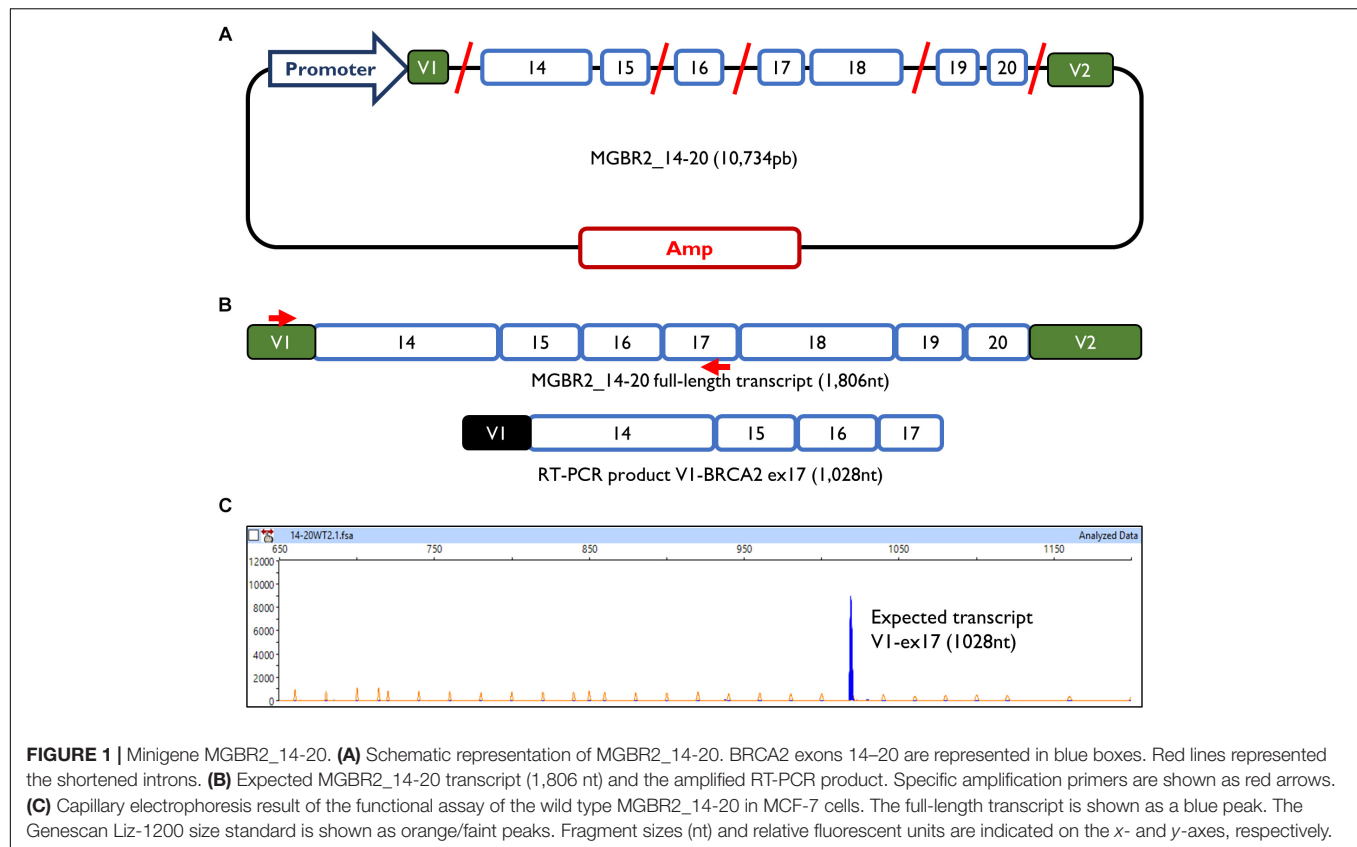
Minigene Construction and ESE Mapping

The minigene MGBR2_14-20 had previously been used to study variants of exons 16, 17, and 18, proving that it is a reliable and robust tool to functionally assay splicing variants (Fraile-Bethencourt et al., 2017, 2018; Montalban et al., 2018). The MGBR2_14-20 is a 10.7 Kb construct which, after transfection in MCF-7 cells, produces a transcript with the following structure: V1-BRCA2_exons from 14 to 20-V2 (1,806 nt) (**Figure 1**). To study exons 14 and 15, cDNA was amplified with a forward primer located in V1 (pMAD_607FW) and a reverse primer located in exon 17 (RTBR2_Ex17RV2), with an expected transcript size of 1028 nt (**Figures 1B,C**).

To map ESEs, 30-nt overlapping microdeletions were performed along the first and the last 55-nt of exons 14 and 15 (Acedo et al., 2015), always preserving the splice site conserved positions (the first 2 nt and the last 3 nt of the exon). None of the deletions but one altered splicing, suggesting the absence of *cis*-regulatory motifs within these segments. Only exon 15 deletion c.7463_7492

⁵http://www.fruitfly.org/seq_tools/splice.html

⁶<http://www.umd.be/HSF3/>



impaired splicing generating a major aberrant transcript (62%) with a deletion of 83 nt [$\Delta(E15p83)$]. This transcript was caused by use of a cryptic 3'ss 83-nt downstream, which is stronger than the canonical one, according to Max Ent Scan (MES; 6.18 vs. 5.16) (Figure 2). To determine which ESEs were implied in exon 15 recognition, siRNA experiments against the main splicing factors (SFSR1, SFSR2, SFSR3, SFSR5, SRSF7, SFSR9, and Tra2 β) were accomplished

(Supplementary Figure S1). However, none of them affected the recognition of exons 14 and 15.

Variant Collection and Bioinformatics Analysis

A total of 294 different variants, spread throughout exons 14 and 15 and flanking introns, were collected from the main databases

(ClinVar, UMD and BIC) (**Supplementary Table S1**). They were analyzed *in silico* with MES and NNSplice for splice site prediction, and with ESE/ESS estimation algorithms integrated in Human Splicing Finder (HSF). Potential splicing variants were selected following these criteria: creation or disruption of splice sites (according to MES or NNSplice); disruption of the branch point; disruption of the polypyrimidine tract; elimination of enhancers or creation of silencers. Some of the selected variants had a combined effect, for example, they were predicted to simultaneously create an ESS and removed an ESE. A total of 53 candidate variants (~18%) that included 19 intronic, 18 missense, 5 nonsense, 8 synonymous, and 3 frameshift variants were selected (**Table 1**). According to their previous clinical classification, the selection contained: 8 benign or likely benign variants, 30 VUS and 15 pathogenic or likely pathogenic variants.

Bioinformatics indicated that 13 variants disrupted the natural splice sites, three decreased their scores (one disrupted the polypyrimidine tract), 11 created new splice sites, one decreased the branch point score (HSF: 79.34→55.5), three altered intronic splicing elements (ISEs and ISSs) and 22 altered exonic splicing elements (ESEs and ESSs) (**Table 1**). Exceptionally, variants c.7008-5T > C (ivs13-5T > C) and c.7435+5T > C (ivs14+5T > C) were also selected because, even though the bioinformatics did not show significant score changes (MES: 10.37→8.48 and 5.64→5.56, respectively), these positions are relevant for exon recognition. Thus, c.7008-5T > C is the closest position of the polypyrimidine tract to the canonical acceptor site, and c.7435+5T > C is part of the consensus 5'ss sequence and +5 changes were previously associated with disease (Sharma et al., 2014; Montalban et al., 2018).

Functional Splicing Assays Using the Minigene MGBR2_14-20

A total of 52 variants were genetically engineered in the wt MGBR2_14-20 by using specific primers (**Supplementary Table S2**). Despite 53 variants were initially selected, mutagenesis experiments did not work for c.7598C > G variant. The 52 mutant minigenes were checked by Sanger sequencing and assayed in MCF-7 cells. Results showed that 12 of them (23%) altered splicing (**Table 2** and **Figure 3**), seven of which had previously been classified as pathogenic and five as VUS (**Table 3**). Among these 12 variants, there were 9 intronic, 2 missense and 1 nonsense changes. Functionally, the 9 intronic variants (c.7008-3C > G, c.7008-2A > T, c.7008-1G > A, c.7435+1G > C, c.7436-2A > T, c.7436-1G > A, c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G) disrupted the natural splice sites, the two missense changes (c.7177A > G and c.7447A > G) triggered the use of *de novo* splice sites and originated other transcripts, and the nonsense one (c.7501C > T) probably altered SREs despite it was primarily selected because of the creation of a new 5'ss (**Table 1**). According to Vallée et al. (2016), nine variants induced severe splicing disruptions as they produced more than 60% of aberrant transcripts, ranging from 92.8 to 100% (**Table 3**).

Acceptor Site Variants

Exon 14 3'ss was affected by c.7008-3C > G, c.7008-2A > T, and c.7008-1G > A, whereas c.7008-5T > C only produced the canonical transcript (**Table 2**). Curiously, while the main outcome of c.7008-3C > G was exon skipping (Δ E14), the variants c.7008-2A > T and c.7008-1G > A produced mostly the aberrant transcript Δ (E14p10), in which a cryptic 3'ss was recognized by the spliceosome 10-nt downstream. This cryptic 3'ss was not detected by NNSplice or MES tools. The loss of 10-nt at the beginning of exon 14 would generate a PTC 27 codons downstream (p.Thr2337Asnfs*27). Our results also revealed the use of another cryptic 3'ss (MES = 4.44) within exon 14, located 246-nt downstream the natural one, originating the transcript Δ (E14p246) (13%) in the c.7008-3C > G assay (**Table 2**). The transcript Δ (E14p246) led to an in-frame deletion of 82 amino acids from position p.2337 to p.2418 (p.Thr2337_Arg2418del).

The branch point (c.7436-22C > T), polypyrimidine tract (c.7436-14T > G) and -4 (c.7436-4A > G, c.7436-4A > T) variants of exon 15 did not impair splicing. Other exon 15 acceptor variants, such as c.7436-2A > T and c.7436-1G > A, mainly caused isoform Δ (E15p13) through use of a cryptic acceptor 13-nt downstream (**Table 2**). The use of this cryptic acceptor would provoke a frameshift deletion, leading to a PTC in the protein (p.Asp2479Valfs*41). Like exon 14 cryptic acceptor, this exon 15 cryptic 3'ss was not detected by NNSplice or MES. Variant c.7436-1G > A also produced the minor transcript Δ E15p83 (3.7%) a 83-nt deletion that introduced a frameshift (p.Asp2479Alafs*32) and a PTC 32 codons downstream. This transcript was generated by a cryptic acceptor site 83-nt downstream (MES = 6.28). In summary, we found 5 variants (c.7008-3C > G, c.7008-2A > T, c.7008-1G > A, c.7436-2A > T, and c.7436-1G > A) that altered 3'ss recognition of exons 14 and 15, leading to aberrant splicing (**Table 2**). Remarkably, all of them showed the total absence of canonical transcript, except for c.7008-3C > G that produced 7% of the full-length transcript. Moreover, our results unveiled exon 14 and 15 cryptic splice sites that are only recognized when natural acceptors are disrupted.

Donor Site Variants

Seven variants were predicted to disrupt donor sites: c.7435+1G > C, c.7435+3A > G, c.7435+5T > C, and c.7435+6G > A (exon 14) and c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G (exon 15; **Table 1**). Among the exon 14 variants, only c.7435+1G > C impaired splicing (**Table 2**) producing a single transcript with a 5-nt insertion [∇ (E14q5)], due to the recognition of a cryptic 5'ss in ivs14. The ∇ (E14q5) is an aberrant splicing isoform which leads to PTC (p.Asp2479Glyfs*4). Surprisingly, this cryptic donor was not detected by NNSplice software as the canonical one was. Regarding exon 15 donor variants, our results showed that all of them (c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G) produced Δ (E15) as unique transcript (**Table 2**), which generates a PTC eight codons downstream (p.Asp2479Alafs*8).

Splicing Regulatory Element-Variants

A total of 26 SRE variants were selected according to the criteria above described and assayed in MCF-7 cells (**Table 1**). Only

TABLE 1 | Bioinformatics analysis of *BRCA2* exons 14 and 15 selected variants.

HGVS variants	Splice site		ESEs/ISEs ¹	ESS	Summary of predictions ²
	<i>NNSplice</i>	<i>MaxEnt</i>	<i>ESE Finder + other motifs</i>	<i>hnRNPA1 by HSF</i>	
<i>Exon 14 (c.7008_7435)</i>	3'SS: 0.56 5'SS: < 0.4	3'SS: 10.37 5'SS: 5.64			
c.7008-5T > C	3'SS, 0.56→0.53	3'SS, 10.37→8.48			↓ 3'SS
c.7008-3C > G	3'SS, 0.56→ < 0.4	3'SS, 10.37→1.64			[−] 3'SS
c.7008-2A > T	3'SS, 0.56→ < 0.4	3'SS, 10.37→2.0			[−] 3'SS
c.7008-1G > A	3'SS, 0.56→ < 0.4	3'SS, 10.37→1.62			[−] 3'SS
c.7009A > G			[+] 3 ESEs; [−] 1 ESE	[+] 67.62	[−]ESEs; [+] ESS
c.7010C > T			[+]1 ESE; [−]1 ESE		[−]ESEs
c.7024C > T			[−] 2 ESEs	[+] 71.67	[−]ESEs; [+] ESS
c.7030A > G	5'SS, <0.4→0.71	5'SS, −1.55→6.62	[+]2 ESEs	[+] 76.90	[+] 5'SS; [−]ESEs; [+] ESS
c.7037A > G			[−] 2 ESEs	[+] 72.38	[−]ESEs; [+] ESS
c.7157C > A			[−] 2 ESEs	[+] 69.76	[−]ESEs; [+] ESS
c.7170T > G			[+]2 ESEs	[+]2 (67.62;76.43)	[+]ESEs; [+] ESS
c.7177A > G			[−] 1 ESE	[+] (69.52)	[−]ESEs; [+] ESS
c.7180A > T	3'SS, <0.4→0.49		[+]1 ESE	[−] (67.14)	[+] 3'SS
c.7182A > G	3'SS, <0.4→0.45				[+] 3'SS
c.7203A > G	3'SS, <0.4→0.46		[−] 2 ESEs	[+] (72.86)	[+] 3'SS; [−]ESEs; [+] ESS
c.7261C > G			[−] 1 ESE	[+] (72.38)	[−]ESEs; [+] ESS
c.7266T > A	5'SS, <0.4→0.76	5'SS, 2.7→8.34			[+] 5'SS
c.7294A > G			[−] 2 ESEs	[+] (65.48)	[−]ESEs; [+] ESS
c.7296A > G			[−] 4 ESEs	[+] (74.76)	[−]ESEs; [+] ESS
c.7330G > T			[−] 2 ESEs	[+] (68.33)	[−]ESEs; [+] ESS
c.7339A > G				[+] (71.43)	[+] ESS
c.7397C > T			[−] 4 ESEs	[+] (70.24)	[−]ESEs; [+] ESS
c.7418G > A	5'SS, <0.4→0.58		[+]2 ESEs		[+] 5'SS
c.7428A > G				[+] (71.43)	[+] ESS
c.7435+1G > C		5'SS, 5.64→−2.62			[−] 5'SS
c.7435+3A > G		5'SS, 5.64→−1.06			[−] 5'SS
c.7435+5T > C		5'SS, 5.64→5.56			[−] 5'SS
c.7435+6G > A			[−] 1 ISE		[−]ISE; +6 conserved nt;
c.7435+7T > G			[−] 1 ISE	[−] (70.48)	[−]ISEs; [−] ESS
c.7435+10G > A			[+]1 ISE; [−]1 ESE	[−] (70.48)	[−]ISEs; [−] ESS
<i>Exon 15 (c.7436_7617)</i>	3'SS: 0.9 5'SS: 0.99	3'SS: 5.16 5'SS: 9.8			
HGVS variants	Splice site		ESEs	ESS	
	<i>NNSplice</i>	<i>MaxEnt</i>	<i>ESE Finder + other motifs</i>	<i>hnRNPA1</i>	
c.7436-22C > T			<i>Branch point: 79.34→55.5</i>	[+](73.81)	↓ Branch point
c.7436-14T > G				[+] (69.05)	↓ 3'SS
c.7436-4A > G	3'SS, 0.9→ < 0.4	3'SS, 5.16→4.54			[−] 3'SS
c.7436-4A > T	3'SS, 0.9→ < 0.4	3'SS, 5.16→4.73			[−] 3'SS
c.7436-2A > T	3'SS, 0.9→ < 0.4	3'SS, 5.16→−3.2			[−] 3'SS
c.7436-1G > A	3'SS, 0.9→ < 0.4	3'SS, 5.16→ −3.58			[−] 3'SS
c.7447A > G	3'SS, <0.4→0.73		[+] 2 ESEs; [−]1 ESE	[+]2 (66.43; 71.19)	[+] 5'SS
c.7466A > G	5'SS, <0.4→0.54	5'SS, 1.53→6.64	[+] 4 ESEs	[+] (76.90)	[+] 5'SS
c.7467T > C			[+] 1 ESE; [−] 2 ESEs	[+] (67.14)	[−]ESEs; [+] ESS
c.7471C > T			[+] 1 ESE; [−] 3 ESEs		[−]ESEs
c.7471delC			[−] 3 ESEs		[−]ESEs
c.7472A > G			[−] 3 ESEs	[−] (75.48)	[−]ESEs; [−] ESS
c.7474_7475delGA			[−] 4 ESEs	[−] (75.48)	[−]ESEs; [−] ESS
c.7492A > G			[+] 1 ESE	[+] 2 (71.43; 76.43)	[+]ESEs; [+] ESS
c.7501C > T	5'SS, <0.4→0.96	5'SS, 2.44→10.19	[−] 1 ESE	[+] (65.72)	[+] 5'SS; [−]ESEs; [+] ESS

(Continued)

TABLE 1 | Continued

HGVS variants	Splice site		ESEs	ESS
	NNSplice	MaxEnt	ESE Finder + other motifs	hnRNPA1
c.7544C > T			[−] 1 ESE	[+] 5'SS; [−]ESEs
c.7598C > G		5'SS, −4.0→4.27	[−] 3 ESEs	[−]ESEs
c.7601C > T			[−] 2 ESEs	[−]ESEs
c.7611_7615delTAAAC	3'SS, <0.4→0.86		[+] 2 ESEs; [−]1 ESE	[+] (66.67)
c.7617G > A	5'SS, 0.99→0.8		[−] 2 ESEs	[−] (73.10)
c.7617+1G > A	5'SS, 0.99→ < 0.4	5'SS, 9.8→1.62	[−] 2 ESEs	[↓] 5'SS
c.7617+1G > T	5'SS, 0.99→ < 0.4	5'SS, 9.8→1.62	[−] 2 ESEs	[−] 5'SS
c.7617+2T > G	5'SS, 0.99→ < 0.4	5'SS, 9.8→2.15	[+] 1 ESE	[−] (66.19)

¹ESE, Exonic Splicing Enhancer; ISE, Intronic Splicing Enhancer; ESS, Exonic Splicing Silencer; ISS, Intronic Splicing Silencer. Full ESE predictions are available at <https://figshare.com/s/246eed89fce86af1e0a6>. ²Summary of predictions: [−], disruption; [+], creation; ↑, score increase; ↓, score decrease.

TABLE 2 | Quantification of the transcripts found by capillary electrophoresis after functional assays of BRCA2 exons 14 and 15 variants.

Variants	Transcripts						
	Full-length	▼(E14q5) r.7435_7436ins 7435+1_7435+5	Δ(E14p10) r.7008_7017del	Δ(E14p246) r.7008_7253del	Δ(E14q256) r.7008_7263del	Δ(E14) r.7008_7435del	Other transcripts
Exon 14 (c.7008_7435)							
c.7008-3C>G	7.2%±0.1%			13.4%±0.1%		79.3%±0.2%	
c.7008-2A>T			88.1%±0.4%			11.9%±0.4%	
c.7008-1G>A			86.5%±0.7%			13.5%±0.7%	
c.7177A>G	95.2%±0.1%				1.5%±0.02%		3.3%±0.1% (942nt)
c.7435+1G>C		100%					
Exon 15 (c.7436_7617)							
	Full-length	Δ(E15p12) r.7436_7447del	Δ(E15p13) r.7436_7448del	Δ(E15p83) r.7436_7518del	Δ(E15) r.7436_7617del		
c.7436-2A>T			100%				
c.7436-1G>A			96.3%±0.1%	3.7%±0.1%			
c.7447A>G	89.9%±0.3%	10.1%±0.3%					
c.7501C>T	84.0%±0.1%			16.0%±0.1%			
c.7617+1G>A					100%		
c.7617+1G>T					100%		
c.7617+2T>G					100%		

Short descriptions of transcripts were annotated according to the ENIGMA consortium: Δ: skipping; E: exon; ▼: insertion; p: alternative 3'ss; q: alternative 5'ss; and the number of nucleotides inserted or deleted. HGVS annotations of transcripts are shown below ENIGMA description. **PTC-transcripts:** ▼(E14q5), Δ(E14p10), Δ(E14q256), Δ(E14), Δ(E15p13), Δ(E15p83), and Δ(E15).

c.7177A > G altered weakly splicing resulting in about 5% of aberrant transcripts (Table 2). This matches with the creation of a new donor site that was not detected by the splicing prediction software. Conversely, none of the exon 15 SRE-variants impaired splicing even though microdeletion tests had revealed a presumed ESE interval (c.7463_7492).

New Splice Site Variants

We have assayed 10 variants of this type, six in exon 14 (c.7030A > G, c.7180A > T, c.7182A > G, c.7203A > G, c.7266T > A, and c.7418G > A) and four in exon 15 (c.7447A > G, c.7466A > G, c.7501C > T, and c.7611_7615delTAAAC) (Table 1). Results showed that two exon 15 variants (c.7447A > G and c.7501C > T) slightly disrupted splicing (Table 2). The variant c.7447A > G generated

a new acceptor but most of its outcome was a full-length transcript (Table 2). The variant c.7501C > T was predicted to create a new strong donor (NNSplice: < 0.4→0.96 and MES: 2.44→10.19) (Table 1). However, only 16% of the transcripts made use of this cryptic donor 83-nt upstream of the natural one (Table 2).

Analysis of Transcripts

The so called full-length or canonical transcript (expected size: 1028-nt) was amplified with primers placed on vector exon V1 and BRCA2 exon 17. Apart from the canonical transcript, we have detected other ten different ones (Figure 3). Aberrant exon 14 splicing produced six different isoforms, but only five of them were characterized: Δ(E14), Δ(E14p10), Δ(E14p246), Δ(E14q256), and ▼(E14q5). A 942-nt transcript of unknown

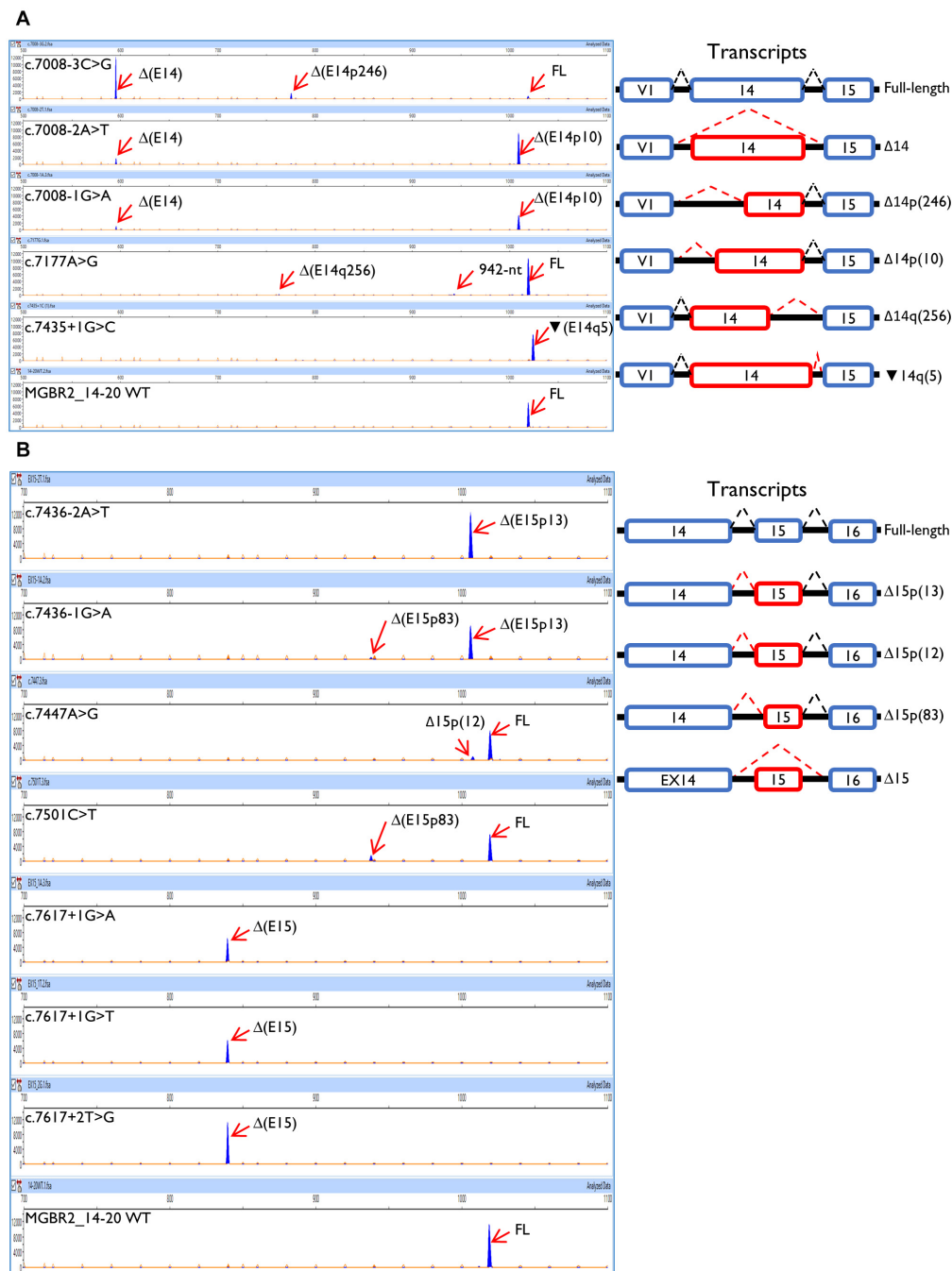


FIGURE 3 | Functional assays of the BRCA2 exons 14 and 15 variants using the MGBR2_14-minigene. **(A)** Exon 14 variants. **(B)** Exon 15 variants. On the left, capillary electropherograms are shown. Transcripts were amplified using pMAD_607FW-FAM and RTBR2_Ex17RV2. Labeled transcripts are shown as blue peaks, LIZ1200 was used as size standard (orange peaks). The expected size of the full length transcript is 1028 nt (1018–1019 nt according to Peak Scanner). On the right, splicing patterns are represented. While blue boxes are natural exons, red boxes represent aberrant exons; dashed black and red lines show canonical and aberrant splicing events, respectively.

structure could also be detected by capillary electrophoresis. All exon 14 isoforms, except for $\Delta(E14p246)$, introduced PTCs into the mRNA. The isoform $\Delta(E14p246)$ was seen in up to 13% of the c.7008-3C > G transcripts (Table 2). Exon 15 variants produced 4 different isoforms besides the canonical

one: $\Delta(E15p12)$, $\Delta(E15p13)$, $\Delta(E15q83)$, and $\Delta(E15)$. The $\Delta(E15p13)$, $\Delta(E15q83)$ and $\Delta(E15)$ isoforms created PTCs while $\Delta(E15p12)$ (new acceptor 12-nt downstream) kept the reading frame, although this isoform only accounted for 10% of the transcripts (Table 2).

TABLE 3 | Classification of spliceogenic variants.

HGVS	Type	Main RNA outcome	Protein effect	Previous classification ¹	ACMG classification ²	ENIGMA ³
Exon 14 (c.7008_7435)						
c.7008-3C > G	Intronic	r.7008_7435del	Thr2337Phefs*17	VUS	PS3, PM2, PP3: L. Pathogenic	Class-4
c.7008-2A > T	Intronic	r.7008_7017del	Thr2337Asnfs*27	Pathogenic	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7008-1G > A	Intronic	r.7008_7017del	Thr2337Asnfs*27	VUS	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7177A > G	Missense (p.Met2393Val)	Full-length	–	VUS	PM2, PP3: VUS	Class-3
c.7435+1G > C	Intronic	r.7435_7436ins 7435+1_7435+5	Asp2479Glyfs*4	Pathogenic	PVS1, PS3, PM2: Pathogenic	Class-4
Exon 15 (c.7436_7617)						
c.7436-2A > T	Intronic	r.7436_7448del	Asp2479Valfs*41	Pathogenic	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7436-1G > A	Intronic	r.7436_7448del	Asp2479Valfs*41	VUS	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7447A > G	Missense (p.Ser2483Gly)	Full-length	–	VUS	PM2, PP3: VUS	Class-3
c.7501C > T	Nonsense (p.Gln2501Ter)	Full-length	–	Pathogenic	PVS1, PM2, PP3: Pathogenic (nonsense)	Class-5 (nonsense)
c.7617+1G > A	Intronic	r.7436_7617del	Asp2479Alafs*8	Pathogenic	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7617+1G > T	Intronic	r.7436_7617del	Asp2479Alafs*8	Pathogenic	PVS1, PS3, PM2, PP3: Pathogenic	Class-4
c.7617+2T > G	Intronic	r.7436_7617del	Asp2479Alafs*8	Pathogenic	PVS1, PS3, PM2, PP3: Pathogenic	Class-4

¹VUS, Variants of Uncertain Clinical Significance. ²ACMG criteria: PVS1, null variant (nonsense, frameshift, canonical ± 1 or ± 2 ss, etc.) in a gene where LOF is a known mechanism of disease; PS3, well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product (PS3 was only used for severe splicing alterations according to Vallée et al., 2016); PM2, absent from controls in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium; PP3, multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.).

³ENIGMA criteria: Class 4: Variant considered extremely likely to alter splicing based on position, namely IVS ± 1 or IVS ± 2 , or G > non-G at last base of exon if first 6 bases of the intron are not GTRGT and/or variants are predicted bioinformatically to alter the use of the native donor/acceptor site. Class 3: "In the absence of clinical evidence to assign an alternative classification, variant allele tested for mRNA aberrations is found to produce mRNA transcript(s) predicted to encode intact full-length protein..."

DISCUSSION

Due to the implementation of Next Generation Sequencing in the clinical setting (Slavin et al., 2015), a large number of variants have been detected in disease-responsible genes. HBOC and the breast cancer susceptibility genes are not exceptions, where thousands of different variants have been reported although many of them are considered as VUS (Spurdle et al., 2012; Slavin et al., 2017). In this context, the functional and clinical classifications pose a challenge for Medical Genetics. We have herein functionally assayed 52 *BRCA2* variants using the minigene MGBR2_14-20, whose reliability had been previously proven (Fraile-Bethencourt et al., 2017, 2018; Montalbán et al., 2018). We found 12 variants that altered splicing, nine of which would severely alter the protein. This study forms part of a comprehensive study of our group concerning potential splicing *BRCA2* variants, where 22 exons and up to 335 variants have been assayed using three minigenes: MGBR2_2-9 (Fraile-Bethencourt et al., 2019), MGBR2_14-20 (Fraile-Bethencourt et al., 2017, 2018) and MGBR2_19-27 (Acedo et al., 2012, 2015). The following advantages of the minigene technology should be underlined: (i) analysis of a single allele outcome without the interference of the wt counterpart of a patient sample; (ii) simple and fast quantification of generated transcripts by fluorescent capillary electrophoresis with minimum hands-on time versus other proposed methods (Farber-Katz et al., 2018); (iii) versatility, one single multi-exon minigene allows to assay variants from different exons; (iv) capability of analysis

in many cell types to check effects derived from tissue-specific alternative splicing; (v) high reproducibility of splicing physiological/pathological patterns. In fact, we have previously provided many examples of the minigene reproducibility. In the case of *BRCA2* exons 14 and 15, variants c.7008-2A > T, c.7617+1G > A, and c.7617+2T > G displayed similar patterns in patient RNA and minigene assays (Vreeswijk et al., 2009; Colombo et al., 2013; de Garibay et al., 2014). Moreover, another 31 variants of this and other constructs replicated previously reported patient splicing outcomes (Acedo et al., 2015; Fraile-Bethencourt et al., 2017, 2018, 2019), confirming that splicing reporters are robust and valuable tools to test the impact of variants on splicing, especially when patient RNA is not available.

Here, we have shown that nine variants drastically disrupted the splicing pattern. We found five 3'ss disrupting variants, one of which (c.7008-3C > G) provoked exon 14 skipping, whereas the rest of them induced the use of cryptic sites (c.7008-2A > T, c.7008-1G > A, c.7436-2A > T, and c.7436-1G > A). Moreover, variant c.7435+1G > C, which disrupt the exon 14 donor site, provoked the use of a cryptic donor. Curiously, neither c.7435+3A > G, c.7435+5T > C nor c.7435+6G > A, which are also part of the consensus 5'ss, affected splicing. This may be due to the low frequency of +5T and +6G at these positions, so that any nucleotide change only equals or improves splice site strength. Conversely, +3A is the main nucleotide at this position (71%) but +3G is also relatively frequent (24%) (Zhang, 1998). Thus, a substitution A to G might have a reduced or no splicing

impact, as it is the case of the c.7435+3A > G variant. As a matter of fact, variant c.9501+3A > T produced 87% of the canonical transcript (Acedo et al., 2015). On the other hand, the three variants of exon 15 donor site c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G resulted in exon skipping. In this regard, it was recently recommended the use of a combination of the computational tools HSF plus Splice Site Finder-like to select candidate splice site variants with high sensitivity and specificity (Moles-Fernández et al., 2018). According to HSF, variants c.7435+3A > G, c.7435+5T > C, and c.7435+6G > A showed only minimal changes of the splice site scores ($\pm 1\%$) so that they should have been excluded from subsequent functional tests.

Concerning other splicing motifs, minigenes also allow the identification of regulatory sequences (splicing enhancers and silencers) and splicing factors involved in the specific regulation (Baralle and Buratti, 2017). Indeed, the SRE mapping constitutes an interesting experimental approach since it identifies critical regions for exon recognition. In this context, our group previously found exonic variants that disrupted splicing through elimination of ESEs or creation of *de novo* silencers, such as BRCA2 variants c.7985C > G (predicted missense p.Thr2662Arg) or c.8009C > A (predicted nonsense p.Ser2670*) (Fraile-Bethencourt et al., 2017). The splicing assays showed how both variants elicited complete splicing aberrations (mainly exon 18 skipping). However, they were *a priori* classified as missense and nonsense variants, respectively, due to their predicted protein effect. After microdeletion

mapping, we have identified a putative ESE-region in exon 15 (c.7463_7492). These ESEs might be involved in exon 15 3'ss recognition since their loss produced the use of a cryptic acceptor 83-nt downstream (Figure 2). Curiously, we did not find any ESE-variants that affected pre-mRNA processing. Only, c.7501C > T, which lays near to this presumed ESE interval, provoked a similar outcome to that of r.7463_7492 deletion (Table 2). In summary, we have tested a total of 117 different variants in minigene MGBR2_14-20, from exons 14–18 (Table 4), 51 of which (43.6%) induced abnormal splicing patterns: 31 disrupted the natural splice sites (16 3'ss and 15 5'ss), 11 affected SREs, six created *de novo* splice sites and three altered the polypyrimidine tract.

Transcript Interpretation

The BRCA2 exons 14 (c.7008_7435) and 15 (c.7436_7617) encode for amino acids 2336 to p.2539, which are part of the DNA Binding Domain (DBD; p.2459_p.3190). The DBD is the largest conserved region of BRCA2 and is composed of a helical domain (HD), three oligonucleotide binding sites (OB) and a tower domain (TD) (Guidugli et al., 2014). Specifically, exons 14 and 15 are part of the HD (p.2481_2667) that binds to the protein DSS1 (deleted in split-hand/split-foot syndrome) in the region comprised by the residues 2472–2957 (Marston et al., 1999). Among these residues, a total of 125 are strictly conserved from human to sea urchin. DSS1 is important for BRCA2 stability, since its loss leads to a

TABLE 4 | Summary of spliceogenic variants tested in minigene MGBR2_14-20.

DNA variant	Splicing motif ¹	Splicing outcome ²	Clinical interpretation
Exon 14 (this work)			
c.7008-3C > G	[–] 3'SS	Δ(E14) 79.3%; Δ(E14p246) 13.4%; CT 7.2%	Likely Pathogenic
c.7008-2A > T	[–] 3'SS	Δ(E14p10) 88.1%; Δ(E14) 11.9%	Pathogenic
c.7008-1G > A	[–] 3'SS	Δ (E14p10) 86.5%; Δ (E14) 13.5%	Pathogenic
c.7177A > G	[–]ESE/[+]ESS	Δ (E14q256) 1.5%; 942-nt, 3.3%	VUS
c.7435+1G > C	[–] 5'SS	▼(E14q5) 100%	Pathogenic
Exon 15 (this work)			
c.7436-2A > T	[–] 3'SS	Δ (E15p13) 100%	Pathogenic
c.7436-1G > A	[–] 3'SS	Δ (E15p13) 96.3%; Δ (E15p83) 3.7%	Pathogenic
c.7447A > G	[+] 3'SS	Δ (E15p12) 10.1%; CT 89.9%	VUS
c.7501C > T	[–]ESE/[+]ESS	Δ (E15p83) 16%; CT 84%	Pathogenic (Nonsense)
c.7617+1G > A	[–] 5'SS	Δ (E15) 100%	Pathogenic
c.7617+1G > T	[–] 5'SS	Δ (E15) 100%	Pathogenic
c.7617+2T > G	[–] 5'SS	Δ (E15) 100%	Pathogenic
Exon 16 (Fraile-Bethencourt et al., 2018)			
c.7618-2A > T	[–] 3'SS	Δ (E16p44) 96.9%; Δ (E16p55) 1.8%; Δ (E16) 1.3%	Pathogenic
c.7618-2A > G	[–] 3'SS	Δ (E16p44) 97.2%; Other transcripts 2.8%	Pathogenic
c.7618-1G > A	[–] 3'SS	Δ (E16p44) 91.5%; Δ (E16p55) 4.7%; Others 2.4%	Pathogenic
c.7618-1G > C	[–] 3'SS	Δ (E16p44) 92.6%; Δ (E16) 2.8%; Δ (E16p55) 1.9%	Pathogenic
c.7802A > G	[+] 5'SS	Δ (E16q4) (45.7%); CT 54.3%	Pathogenic
c.7805G > C	[–] 5'SS	Δ (E16) 77.6%; Δ (E16q100) 14.4%; ▼(E16q20) 6.5%	Pathogenic
c.7805+1G > A	[–] 5'SS	Δ (E16) 88%; Δ (E16q100) 10.1%	Pathogenic
c.7805+3A > C	[–] 5'SS	Δ (E16) 75.3%; Δ (E16q100) 13.3% ▼(E16q20) 3.8%	Pathogenic

(Continued)

TABLE 4 | Continued

DNA variant	Splicing motif ¹	Splicing outcome ²	Clinical interpretation
Exon 17 (Fraile-Bethencourt et al., 2017; Montalban et al., 2018)			
c.7806-9T > G	Pyr	Δ (E17) 41.5%; ▼(E17p8) 36.3%; Δ (E17p69) 22.2%	Likely Pathogenic
c.7806-2A > G	[−] 3'SS	Δ (E17p20) 51.8%; Δ (E17p69) 28.1%; Δ (E17) 20.1%	Pathogenic
c.7806-1G > A	[−] 3'SS	Δ (E17p1) 100%	Pathogenic
c.7806-1G > T	[−] 3'SS	Δ (E17p20) 100%	Pathogenic
c.7806-1_7806-2dup	[+] ³ 'SS	Δ (E17p2) 92.6%; Δ (E17) 5.1%; Δ (E17p69) 2.3%	Pathogenic
c.7975A > G	[−] 5'SS	Δ (E17) 26.2%; CT 73.8%	VUS
c.7976G > C	[−] 5'SS	Δ (E17) 100%	Likely Pathogenic
c.7976G > A	[−] 5'SS	Δ (E17) 100%	Likely Pathogenic
c.7976+1G > A	[−] 5'SS	Δ (E17) 100%	Pathogenic
c.7976+5G > T	[−] 5'SS	Δ (E17) 100%	Likely Pathogenic
Exon 18 (Fraile-Bethencourt et al., 2017)			
c.7977-7C > G	[+] ³ 'SS/Pyr	▼(E18p6) 78.4%; Δ (E18) 21.6%	Likely Pathogenic
c.7977-6T > G	Pyr	CT 66.7%; Δ (E18) 31%; Δ (E18p191) 2.3%	VUS
c.7977-3_7978del	[−] 3'SS	Δ (E18) 90%; Δ (E18p191) 10%	Pathogenic
c.7977-2A > T	[−] 3'SS	Δ (E18) 93.3%; Δ (E18p191) 6.7%	Pathogenic
c.7977-1G > T	[−] 3'SS	Δ (E18) 91.5%; Δ (E18p191) 7%; Δ (E18p236) 1.5%	Pathogenic
c.7977-1G > C	[−] 3'SS	Δ (E18) 89.8%; Δ (E18p191) 10.2%	Pathogenic
c.7985C > G	[−]ESE/[+]ESS	Δ (E18) 90.2%; Δ (E18p191) 5%; others 4.8%	Likely Pathogenic
c.7988A > T	[+] ⁵ 'SS [−]ESE	CT 84.2%; Δ (E18) 8.6%; others 7.2%	VUS
c.7992T > A	[−]ESE/[+]ESS	CT 68.6%; Δ (E18) 31.4 %	VUS
c.8007A > G	[−]ESE/[+]ESS	CT 84.8%; Δ (E18) 15.2%	VUS
c.8009C > A	[−]ESE/[+]ESS	Δ (E18) 91.2%; Δ (E18p191) 4.8%; CT 4%	Pathogenic
c.8009C > T	[−]ESE/[+]ESS	CT 76.6%; Δ (E18) 23.4%	VUS
c.8009C > G	[−]ESE/[+]ESS	CT 79.9%; Δ (E18) 20.1%	VUS
c.8023A > G	[+] 5'SS	Δ (E18q309) 93%; other aberrant transcripts 7%	Likely Pathogenic
c.8035G > T	[+] 5'SS	Δ (E18q298) 93.6%; 878-nt transcript 4%; CT: 2.4%	Likely Pathogenic
c.8072C > T	[−]ESE/[±]ESS	CT 94.9%; Δ (E18) 5.1%	VUS
c.8168A > G	[+] ⁵ 'SS	CT 69.6%; Δ (E18q164) 25.9%; Δ (E18) 4.5%	VUS
c.8249_8250del	[−]ESE/[−]ESS	CT 93.0%; Δ (E18) 7.0%	VUS
c.8331G > A	[−] 5'SS	Δ (E18) 52%; CT 40.7%; aberrant transcripts 7.3%	Likely Pathogenic
c.8331+1G > T	[−] 5'SS	Δ (E18) 81%; Ex18-del157 6.4%; Δ (E17q151, E18) 6.1%; ▼(E17q58)+ Δ (E18) 3.7%; others 2.8%	Pathogenic
c.8331+2T > C	[−] 5'SS	Δ (E18) 87.1%; Δ (E17q151,E18) 12.9%	Pathogenic

¹[−], disruption; [+], creation; [±], simultaneous creation and disruption; 5'SS, 5' splice site; 3'SS, 3' splice site; ESE, Exonic Splicing Enhancer; ESS, Exonic Splicing Silencer; Pyr, polypyrimidine tract. ²Transcript annotation according to Lopez-Perolio et al. (2019); CT, canonical transcript.

reduction of BRCA2 levels in human cells (Li et al., 2006). Moreover, exons 14 and 15 coding region is also recognized by FANCD2 (Fanconi anemia group D2) protein, which binds to the BRCA2 protein between codons 2350 and 2545 (Hussain et al., 2004). FANCD2, like BRCA2, is one of the 16 proteins that form the Fanconi Anemia complex, aimed to repair DNA interstrand crosslinks. However, it was shown that BRCA2-FANCD2 association has an independent function in the Fanconi Anemia pathway. The BRCA2-FANCD2 complex is involved in the restart of the replication fork, by protecting the nascent DNA strands from degradation (Raghunandan et al., 2015). Taken together, exons 14 and 15 contain crucial sequences of BRCA2, owing to its function in homologous recombination and other relevant biological processes. Moreover, the biological relevance of exons 14 and 15 is supported by the presence of numerous pathological nonsense and frameshift variants at the mutation databases. Hence, the exon 14–15 spliceogenic

variants that induce PTC-transcripts may be associated with an increased risk of HBOC.

Clinical Interpretation of Spliceogenic Variants

Twelve variants altered splicing with different patterns. While some variants caused the total absence of canonical transcript, others originated just ~5% of aberrant transcripts (Table 2). Variants c.7008-2A > T, c.7008-1G > A, c.7435+1G > C, c.7436-2A > T, c.7436-1G > A, c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G did not produce the canonical transcript. Moreover, all the transcripts generated by these variants were frameshift transcripts. Thus, following the criteria of the American College of Medical Genetics and Genomics (ACMG), these eight variants should be classified as pathogenic variants (Table 3). Variant c.7008-3C > G produced Δ(E14) as the major transcript, but the full-length (~7%) and the in-frame transcript

Δ (E14p246) (~13%) were also identified. The Δ (E14p246) isoform contains a deletion of 82 non-conserved amino acids (p.Thr2337_Arg2418del) that form part of the FANCD2 binding site (p.Thr2350_Val2545). At the UMD database, c.7008-3C > G is classified as a “causal” variant because of the skipping of exon 14, but ClinVar shows it as VUS. According to the ACMG criteria, this variant should be cataloged as likely pathogenic (Table 3). On the other hand, according to the guidelines of the ENIGMA consortium⁷ eight variants should be classified as Class 4 (Likely pathogenic), three as VUS and one as pathogenic, though this is due to its predicted nonsense effect (Table 3).

On the other hand, missense variants c.7177A > G (p.Met2393Val) and c.7447A > G (p.Ser2483Gly) produced the canonical transcript as the main outcome and only 5 and 10% of aberrant isoforms, respectively (Table 2). The canonical transcript generated by these two variants carried missense changes, but *in silico* predictions with the PolyPhen tool⁸ suggested that p.Met2393Val and p.Ser2483Gly were both benign changes. So, following the ACMG criteria, c.7177A > G and c.7447A > G remain as VUS (Table 3). Finally, variant c.7501C > T generated mainly the canonical transcript (84%) that includes a nonsense change (pathogenic according to ClinVar) so this change should be classified as pathogenic under a combined splicing-protein viewpoint. Altogether, we have re-classified three variants (c.7008-3C > G, c.7008-1G > A, and c.7436-1G > A) from VUS to pathogenic or likely pathogenic, and we have provided further support for the classification of six spliceogenic variants (c.7008-2A > T, c.7435+1G > C, c.7436-2A > T, c.7617+1G > A, c.7617+1G > T, and c.7617+2T > G). Interestingly, c.7617G > A is classified as “causal” in the UMD database⁹ and indicated that causes exon 15 skipping but no functional proofs were provided. However, the functional assay of MGBR2_14-20-c.7617G > A only showed the canonical transcript (Supplementary Figure S2). In fact, NNSplice, HSF and MES estimated just a small decrease (−19, −11.6, and −24%, respectively) of the donor site score. Therefore, c.7617G > A behaves as a neutral variant from the splicing perspective. Finally, the minigene MGBR2_14-20 results of 5 exons suggest that a total of 39 spliceogenic variants should be classified as pathogenic or likely pathogenic (Table 4), lending further support to this strategy for the clinical interpretation of variants.

In summary, splicing is a finely regulated mechanism which can be altered by any change in the sequence. The disruption of this process might cause serious effects on the protein, from the loss of important domains to the generation of a PTC. Thus, splicing alteration is a common mechanism of gene inactivation, which is often involved in human disease. Here, we have revealed 12 spliceogenic variants of BRCA2 exons 14 and 15. The minigene based assays offer a relevant information about effects of splicing variants, since they allow to functionally assay almost any DNA change, to quantify all generated transcripts, including very rare

ones, and to initially study the splicing regulation. Moreover, we have detected an ESE region that seems to be regulating exon 15 splicing, and therefore constitutes a hypothetical hotspot for putative ESE-mutations. Indeed, pSAD-based minigenes have constituted an invaluable technology to functionally test variants of other disease genes such as GRN (Frontotemporal Dementia), SERPINA1 (Severe alpha-1 antitrypsin deficiency) and CHD7 (Charge Syndrome) genes (Lara et al., 2014; Villate et al., 2018). Altogether, these results highlight, once more, the importance of RNA assays to know the splicing effects of DNA variants to give support to their clinical interpretation and consequently to activate preventive and/or therapeutic interventions.

AUTHOR CONTRIBUTIONS

EF-B contributed to the bioinformatics analysis, minigene construction, manuscript writing, and performed most of the splicing functional assays. BD-G and AV-P participated in minigene construction, mutagenesis experiments, and functional assays. SG-B and MC carried out the data collection of variants and their computer analysis, as well as manuscript editing. EV conceived the study and the experimental design, supervised all the experiments, and wrote the manuscript. All authors contributed to data interpretation, revisions of the manuscript, and approved the final version of the manuscript.

FUNDING

EV's lab was supported by grants from the Spanish Ministry of Science, Innovation and Universities, Plan Nacional de I+D+I 2013-2016, Instituto de Salud Carlos III (PI13/01749 and PI17/00227) co-funded by FEDER from Regional Development European Funds (European Union), and grants CSI090U14 (ORDEN EDU/122/2014) and CSI242P18 (actuación cofinanciada P.O. FEDER 2014-2020 de Castilla y León) from the Consejería de Educación, Junta de Castilla y León. EF-B was supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015–2019). AV-P was supported by a predoctoral fellowship from Consejería de Educación, Junta de Castilla y León (2018–2022).

ACKNOWLEDGMENTS

We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00503/full#supplementary-material>

⁷ <https://enigmaconsortium.org/library/general-documents/enigma-classification-criteria/>

⁸ <http://genetics.bwh.harvard.edu/pph2/>

⁹ <https://goo.gl/bG2V4n>

REFERENCES

- Abramowicz, A., and Gos, M. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268. doi: 10.1007/s13353-018-0444-447
- Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B., and Velasco, E. A. (2015). Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum. Mutat.* 36, 210–221. doi: 10.1002/humu.22725
- Acedo, A., Sanz, D. J., Durán, M., Infante, M., Pérez-Cabornero, L., Miner, C., et al. (2012). Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res.* 14:R87. doi: 10.1186/bcr3202
- Baralle, D., and Buratti, E. (2017). RNA splicing in human disease and in the clinic. *Clin. Sci.* 131, 356–368. doi: 10.1042/CS20160211
- Beroud, C., Letovsky, S. I., Braastad, C. D., Caputo, S. M., Beaudoux, O., Bignon, Y. J., et al. (2016). BRCA share: a collection of clinical BRCA gene variants. *Hum. Mutat.* 37, 1318–1328. doi: 10.1002/humu.23113
- Buy, S. S., Sandbach, J. F., Gammon, A., Patel, G., Kidd, J., Brown, K. L., et al. (2017). A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer* 123, 1721–1730. doi: 10.1002/cncr.30498
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298. doi: 10.1038/nrg775
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31, 3568–3571.
- Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., et al. (2013). Comparative in vitro and in silico analyses of variants in splicing regions of brca1 and brca2 genes and characterization of novel pathogenic mutations. *PLoS One* 8:e57173. doi: 10.1371/journal.pone.0057173
- De Conti, L., Baralle, M., and Buratti, E. (2012). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4, 49–60. doi: 10.1002/wrna.1140
- de Garibay, G. R., Acedo, A., García-Casado, Z., Gutiérrez-Enríquez, S., Tosar, A., Romero, A., et al. (2014). Capillary electrophoresis analysis of conventional splicing assays: IARC analytical and clinical classification of 31 BRCA2 genetic variants. *Hum. Mutat.* 35, 53–57. doi: 10.1002/humu.22456
- Desmet, F. O., Hamroun, D., Lalonde, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37:e67. doi: 10.1093/nar/gkp215
- Eccles, D. M., Mitchell, G., Monteiro, A. N. A., Schmutzler, R., Couch, F. J., Spurdle, A. B., et al. (2015). BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. *Ann. Oncol.* 26, 2057–2065. doi: 10.1093/annonc/mdv278
- Fackenthal, J. D., Yoshimatsu, T., Zhang, B., de Garibay, G. R., Colombo, M., De Vecchi, G., et al. (2016). Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. *J. Med. Genet.* 53, 548–558. doi: 10.1136/jmedgenet-2015-103570
- Farber-Katz, S., Hsuan, V., Wu, S., Landrith, T., Vuong, H., Xu, D., et al. (2018). Quantitative Analysis of BRCA1 and BRCA2 Germline Splicing Variants Using a Novel RNA-Massively Parallel Sequencing Assay. *Front. Oncol.* 8:286. doi: 10.3389/fonc.2018.00286
- Fraile-Bethencourt, E., Díez-Gómez, B., Velásquez-Zapata, V., Acedo, A., Sanz, D. J., and Velasco, E. A. (2017). Functional classification of DNA variants by hybrid minigenes: identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet.* 13:e1006691. doi: 10.1371/journal.pgen.1006691
- Fraile-Bethencourt, E., Valenzuela-Palomo, A., Díez-Gómez, B., Acedo, A., and Velasco, E. A. (2018). Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays. *Front. Genet.* 9:188. doi: 10.3389/fgene.2018.00188
- Fraile-Bethencourt, E., Valenzuela-Palomo, A., Díez-Gómez, B., Goia, E., Acedo, A., Buratti, E., et al. (2019). Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J. Pathol.* doi: 10.1002/path.5268 [Epub ahead of print].
- Guidugli, L., Carreira, A., Caputo, S. M., Ehlen, A., Galli, A., Monteiro, A. N., et al. (2014). Functional assays for analysis of variants of uncertain significance in BRCA2. *Hum. Mutat.* 35, 151–164. doi: 10.1002/humu.22478
- Hussain, S., Wilson, J. B., Medhurst, A. L., Hejna, J., Witt, E., Ananth, S., et al. (2004). Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum. Mol. Genet.* 13, 1241–1248. doi: 10.1093/hmg/ddh135
- Lara, B., Martínez, M. T., Blanco, I., Hernández-Moro, C., Velasco, E. A., Ferrarotti, I., et al. (2014). Severe alpha-1 antitrypsin deficiency in composite heterozygotes inheriting a new splicing mutation QOMadrid. *Respir. Res.* 15, 125. doi: 10.1186/s12931-014-0125-y
- Li, J., Zou, C., Bai, Y., Wazer, D. E., Band, V., and Gao, Q. (2006). DSS1 is required for the stability of BRCA2. *Oncogene* 25, 1186–1194. doi: 10.1038/sj.onc.1209153
- López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903. doi: 10.1016/j.febslet.2005.02.047
- Lopez-Perolio, I., Leman, R., Behar, R., Lattimore, V., Pearson, J. F., Castéra, L., et al. (2019). Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. *J. Med. Genet.* 2019, 1–9. doi: 10.1136/jmedgenet-2018-105834
- Marston, N. J., Richards, W. J., Hughes, D., Bertwistle, D., Marshall, C. J., and Ashworth, A. (1999). Interaction between the Product of the Breast Cancer Susceptibility Gene BRCA2 and DSS1, a Protein Functionally Conserved from Yeast to Mammals. *Mol. Cell. Biol.* 19, 4633–4642. doi: 10.1128/MCB.19.7.4633
- Mavaddat, N., Peock, S., Frost, D., Ellis, S., Platte, R., Fineberg, E., et al. (2013). Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J. Natl. Cancer Inst.* 105, 812–822. doi: 10.1093/jnci/djt095
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66–71.
- Moles-Fernández, A., Duran-Lozano, L., Montalbán, G., Bonache, S., López-Perolio, I., Menéndez, M., et al. (2018). Computational tools for splicing defect prediction in breast/ovarian cancer genes: how efficient are they at predicting rna alterations? *Front. Genet.* 9:366. doi: 10.3389/fgene.2018.00366
- Montalbán, G., Fraile-Bethencourt, E., López-Perolio, I., Pérez-Segura, P., Infante, M., Durán, M., et al. (2018). Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: BRCA2 c.7976+5G > T as a case study. *Hum. Mutat.* 39, 1155–1160. doi: 10.1002/humu.23583
- Nielsen, F. C., van Overeem Hansen, T., and Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat. Rev. Cancer* 16, 599–612. doi: 10.1038/nrc.2016.72
- Petrucelli, N., Daly, M. B., and Feldman, G. L. (2013). “BRCA1 and BRCA2 hereditary breast and ovarian cancer”. *GeneReviews® [Internet]* eds MP., Adam, H. H., Ardinger Seattle, WA: University of Washington, Seattle 1–11.
- Radice, P., De, S. S., Caleca, L., and Tommasi, S. (2011). Unclassified variants in BRCA genes: guidelines for interpretation. *Ann. Oncol.* 22(Suppl. 1), i18–i23. doi: 10.1093/annonc/mdq661
- Raghuveer, M., Chaudhury, I., Kelich, S. L., Hanenberg, H., and Soback, A. (2015). FANCD2, FANCD1 and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi Anemia core complex. *Cell Cycle* 14, 342–353. doi: 10.4161/15384101.2014.987614
- Rebeck, T. R., Friebe, T. M., Friedman, E., Hamann, U., Huo, D., Kwong, A., et al. (2018). Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum. Mutat.* 39, 593–620. doi: 10.1002/humu.23406
- Reese, M. G., Eeckman, F. H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323. doi: 10.1089/cmb.1997.4.311
- Ricks, T. K., Chiu, H.-J., Ison, G., Kim, G., McKee, A. E., Kluetz, P., et al. (2015). Successes and challenges of PARP inhibitors in cancer therapy. *Front. Oncol.* 5:222. doi: 10.3389/fonc.2015.00222
- Sharma, N., Sosnay, P. R., Ramalho, A. S., Douville, C., Franca, A., Gottschalk, L. B., et al. (2014). Experimental assessment of splicing variants using expression minigenes and comparison with in silico predictions. *Hum. Mutat.* 35, 1249–1259. doi: 10.1002/humu.22624
- Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., et al. (2017). The contribution of pathogenic variants in breast cancer susceptibility

- genes to familial breast cancer risk. *NPJ Breast Cancer* 3, 22. doi: 10.1038/s41523-017-0024-8
 - Slavin, T. P., Niell-Swiler, M., Solomon, I., Nehoray, B., Rybak, C., Blazer, K. R., et al. (2015). Clinical application of multigene panels: challenges of next-generation counseling and cancer risk management. *Front. Oncol.* 5:208. doi: 10.3389/fonc.2015.00208
 - Spurdle, A. B., Couch, F. J., Hogervorst, F. B. L., Radice, P., and Sinilnikova, O. M. (2008). Prediction and assessment of splicing alterations: implications for clinical testing. *Hum. Mutat.* 29, 1304–1313. doi: 10.1002/humu.20901
 - Spurdle, A. B., Healey, S., Devereau, A., Hogervorst, F. B. L., Monteiro, A. N. A., Nathanson, K. L., et al. (2012). ENIGMA-evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum. Mutat.* 33, 2–7. doi: 10.1002/humu.21628
 - Stratton, M. R., and Rahman, N. (2008). The emerging landscape of breast cancer susceptibility. *Nat. Genet.* 40, 17–22.
 - Vallée, M. P., Di Sera, T. L., Nix, D. A., Paquette, A. M., Parsons, M. T., Bell, R., et al. (2016). Adding in silico assessment of potential splice aberration to the integrated evaluation of brca gene unclassified variants. *Hum. Mutat.* 37, 627–639. doi: 10.1002/humu.22973
 - Villate, O., Ibarluzea, N., Fraile-Bethencourt, E., Valenzuela, A., Velasco, E. A., Grozeva, D., et al. (2018). Functional analyses of a novel splice variant in the CHD7 gene. found by next generation sequencing, confirm its pathogenicity in a spanish patient and diagnose him with CHARGE syndrome. *Front. Genet.* 9, 26–31. doi: 10.3389/fgene.2018.00007
 - Vreeswijk, M. P. G., Kraan, J. N., van der Klift, H. M., Vink, G. R., Cornelisse, C. J., Wijnen, J. T., et al. (2009). Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum. Mutat.* 30, 107–114. doi: 10.1002/humu.20811
 - Wang, G.-S., and Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761. doi: 10.1038/nrg2164
 - Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789–792. doi: 10.1038/378789a0
 - 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
 - Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* 7, 919–932. doi: 10.1093/hmg/7.5.919
- 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 Fraile-Bethencourt, Valenzuela-Palomo, Díez-Gómez, Caloca, Gómez-Barrero and Velasco. 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.



Global Increase in Circular RNA Levels in Myotonic Dystrophy

Karol Czubak¹, Katarzyna Taylor², Agnieszka Piasecka², Krzysztof Sobczak², Katarzyna Kozłowska³, Anna Philips³, Saam Sedehizadeh⁴, J. David Brook⁴, Marzena Wojciechowska^{1,4} and Piotr Kozłowski^{1*}

¹ Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland,

² Laboratory of Gene Therapy, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland, ³ European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, ⁴ Queen's Medical Centre, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

OPEN ACCESS

Edited by:

Rosanna Asselta,
Humanitas University, Italy

Reviewed by:

Alessandra Perfetti,
Policlinico San Donato (IRCCS),
Italy

Andy Berglund,
University at Albany,
United States

*Correspondence:

Piotr Kozłowski
kozlowp@ibch.poznan.pl

Specialty section:

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

Received: 21 February 2019

Accepted: 19 June 2019

Published: 18 July 2019

Citation:

Czubak K, Taylor K, Piasecka A,
Sobczak K, Kozłowska K, Philips A,
Sedehizadeh S, Brook JD,
Wojciechowska M and Kozłowski P
(2019) Global Increase in Circular
RNA Levels in Myotonic Dystrophy.
Front. Genet. 10:649.
doi: 10.3389/fgene.2019.00649

Splicing aberrations induced as a consequence of the sequestration of muscleblind-like splicing factors on the dystrophin myotonia protein kinase transcript, which contains expanded CUG repeats, present a major pathomechanism of myotonic dystrophy type 1 (DM1). As muscleblind-like factors may also be important factors involved in the biogenesis of circular RNAs (circRNAs), we hypothesized that the level of circRNAs would be decreased in DM1. To test this hypothesis, we selected 20 well-validated circRNAs and analyzed their levels in several experimental systems (e.g., cell lines, DM muscle tissues, and a mouse model of DM1) using droplet digital PCR assays. We also explored the global level of circRNAs using two RNA-Seq datasets of DM1 muscle samples. Contrary to our original hypothesis, our results consistently showed a global increase in circRNA levels in DM1, and we identified numerous circRNAs that were increased in DM1. We also identified many genes (including muscle-specific genes) giving rise to numerous (>10) circRNAs. Thus, this study is the first to show an increase in global circRNA levels in DM1. We also provided preliminary results showing the association of circRNA level with muscle weakness and alternative splicing changes that are biomarkers of DM1 severity.

Keywords: circular RNAs (circRNAs), myotonic dystrophy type 1 (DM1), circRNA biogenesis, CIRI2, droplet digital PCR (ddPCR), DM1 biomarkers

INTRODUCTION

Myotonic dystrophy type 1 (*dystrophin myotonia 1*, DM1, OMIM: 160900) is the most common form of adult-onset muscular dystrophy, affecting approximately 1 in 8,000 people worldwide. DM1 is an autosomal dominant disorder caused by an expansion of CTG repeats in the 3' untranslated region of the *dystrophin myotonia protein kinase* (DMPK) gene (Brook et al., 1992; Fu et al., 1992; Mahadevan et al., 1992). Unaffected individuals have between 5 and ~34 repeats, whereas in DM1 patients, the triplet repeat is expanded, often to hundreds or even thousands of copies (Brook et al., 1992). The pathogenesis of DM1 is strongly linked to the expression of mutation-containing transcripts and is manifested through the nuclear accumulation of mutant transcripts in characteristic foci (Taneja et al., 1995). The presence of these mutant transcripts causes the sequestration of muscleblind-like (MBNL) proteins [including MBNL1, the main MBNL family protein in muscles (Fardaei et al., 2002; Konieczny et al., 2014), MBNL2, and MBNL3], which normally regulate alternative splicing of pre-messenger RNAs (pre-mRNAs) encoding proteins

critical for skeletal, cardiac, and nervous system function (Miller et al., 2000; Pascual et al., 2006). Thus, their sequestration and functional insufficiency result in aberrant alternative splicing of many target genes. For example, mis-splicing of the *CLCN1* exon 7, the *INSR* exon 11, and the *BIN1* exon 11 were shown to be associated with reduced chloride conductance, lower insulin responsiveness, and muscle weakness, respectively (Philips et al., 1998; Savkur et al., 2001; Mankodi et al., 2002; Ho et al., 2004; Fugier et al., 2011). A pathomechanism similar to that observed in DM1 was also proposed for myotonic dystrophy type 2 (*dystrophia myotonica* 2, DM2, OMIM: 602668), a disease caused by an expansion of CCTG repeats in the first intron of the *CCHC-type zinc finger nucleic acid binding protein* gene (Liquori et al., 2001). However, in this study, we mainly focused on DM1.

The results of a recent study suggest that in addition to a function in alternative splicing, MBNLs may play an important role in the biogenesis of a recently recognized class of RNA molecules called circular RNAs (circRNAs) (Ashwal-Fluss et al., 2014). Unlike other types of RNA, circRNAs are very stable molecules. Due to the low expression level of the initially identified circRNAs, they were considered byproducts of aberrant RNA splicing. However, with the dissemination of RNA-Seq technology, research has revealed that circRNAs are abundant among a variety of transcriptomes (Memczak et al., 2013; Salzman et al., 2013). Although the levels of most circRNAs are low, there are examples of circRNAs with levels comparable with or higher than those of their linear counterparts (Jeck et al., 2013). Most circRNAs are encoded by protein-coding genes and derived from their exons, which may indicate that transcription of circRNAs is directed by RNA polymerase II and that their biogenesis is mediated by the spliceosome. In the majority of cases, head-to-tail junctions of circular transcripts are flanked by canonical splice sites (Ashwal-Fluss et al., 2014; Starke et al., 2015). Reportedly, the formation of circRNAs may occur both post-transcriptionally and cotranscriptionally (Wilusz and Sharp, 2013; Ashwal-Fluss et al., 2014; Kramer et al., 2015), and their biogenesis competes with the formation of linear transcripts (mRNA). The mechanisms of this competition are tissue specific and conserved from flies to humans (Ashwal-Fluss et al., 2014; Rybak-Wolf et al., 2015). To date, no function has been assigned for the vast majority of circRNAs, with exceptions such as circCDR1as, *Sry* circRNA, or circHIPK3 (*hsa_circ_0000284*), which can act as microRNA sponges (Hansen et al., 2011; Memczak et al., 2013; Zheng et al., 2016). Other functions, such as involvement in protein and/or RNA transport (Memczak et al., 2013), regulating synaptic functions in neural tissue (You et al., 2015), or acting as templates for translation of functional peptides [e.g., (Li and Lytton, 1999)], have also been proposed for circRNAs.

The precise mechanism of circRNA generation remains unknown. However, several mechanisms of circRNA biogenesis have been proposed (Salzman et al., 2012; Jeck et al., 2013; Salzman et al., 2013). All of these proposed mechanisms assume the generation of circRNAs by head-to-tail splicing (back-splicing). One of the proposed mechanisms suggests that RNA-binding proteins (RBPs), which bind to specific motifs in introns flanking

circRNA-coding exons, play an important role in circRNA biogenesis (Ashwal-Fluss et al., 2014; Conn et al., 2015). Back-splicing is facilitated by the interaction between RBPs, which bring the introns closer together. The *Drosophila* Mbl protein (ortholog of human MBNLs) may be a circRNA-biogenesis RBP (Ashwal-Fluss et al., 2014). Interestingly, one of Mbl-regulated circRNAs is circMBNL1/circMbl, a circRNA generated from the second exon of the *MBNL1/Mbl* gene. The introns flanking this circRNA contain highly conserved MBNL/Mbl-binding motifs. Furthermore, the exogenous expression of Mbl stimulates circRNA production from endogenous MBNL1/Mbl transcripts in both humans and flies. Mbl-binding sequences in both introns are necessary, suggesting that Mbl induces circularization by bridging the two flanking introns. Importantly, downregulation of Mbl in both fly cell culture and fly neural tissue leads to a significant decrease in circMbl level, whereas the elevated level of Mbl increases the level of circMbl as well as other circRNAs, suggesting a general role for MBNLs/Mbl in circRNA biogenesis (Ashwal-Fluss et al., 2014).

In this work, we aimed to test the level of circRNAs in DM1. Since MBNL proteins may be involved in circRNA biogenesis (Ashwal-Fluss et al., 2014), we hypothesized that the generation of at least some circRNAs (e.g., circRNAs characterized by multiple MBNL-binding sites in their flanking introns) would be downregulated by the diminished functional levels of MBNLs, which are sequestered in mutant RNA foci (Miller et al., 2000). To test this hypothesis, we selected 20 well-validated circRNAs and analyzed their expression levels in several experimental systems, including cultured human myoblasts and skeletal muscle biopsy samples from patients and healthy individuals. In addition, we used muscles from the *HSA*^{1R} transgenic mouse model of DM1 (Mankodi et al., 2000; Sobczak et al., 2013). The analysis of circRNA expression levels was performed with in-house-designed droplet digital PCR (ddPCR) (Hindson et al., 2011; Miotke et al., 2014) assays. We also expanded this analysis and explored global levels of circRNAs using RNA-Seq data from an “exploratory cohort” of DM1 muscle samples of quadriceps femoris (QF) and tibialis anterior (TA) (<http://www.dmseq.org/>).

In summary, we found no downregulation of the analyzed circRNAs in DM (both DM1 and DM2) samples compared with those in non-DM samples. Therefore, these results question the role of MBNL proteins in circRNA biogenesis in muscles. Interestingly, in our experimental systems that are characterized by a lower level of functional MBNLs, we discovered a consistent increase in circRNA levels. As a result, we identified a subset of circRNAs that were upregulated in DM1 samples and could be used as novel biomarkers. Although the obtained data do not confirm our hypothesis regarding the link between MBNL sequestration and disrupted circRNA biogenesis in DM1 (and DM2), we do not exclude the possibility of the existence of individual circRNAs that are regulated by MBNLs. Additionally, we demonstrated that elevated circRNA levels associate with molecular (alternative splicing) and clinical (muscle weakness) symptoms of DM severity. However, the role of individual circRNAs altered in DM1 and their global function in DM1 pathogenesis remain to be determined.

MATERIALS AND METHODS

Complementary DNA Samples

Four complementary DNA (cDNA) sample sets (Table 1 and described later) were used in this study. These sets included samples from myoblast cell lines (CL) derived from human skeletal muscles, muscle biopsy (BP) samples from DM1 and DM2 patients and corresponding healthy controls, and samples from the *HSA^{LR}* transgenic mouse model of DM1 (MM). For the purpose of cDNA generation, total RNA was extracted using the standard protocol, as previously described (Chomczynski and Sacchi, 1987). Reverse transcription was performed according to the manufacturers' recommendations. All reverse transcription reactions were performed with the use of random hexamers. The particular reverse transcriptases (RTs) used in the analyzed sample sets are indicated later. The DM1-specific splicing aberrations in the muscle sample sets used in this study were evaluated before (Wojciechowska et al., 2018) and are shown (BP_DM2) in Figure S1. The splicing aberrations in DM1 samples deposited in the DMseq database and analyzed in this study [see section Analysis of Next-Generation Sequencing Data] were also recently demonstrated (Wang et al., 2018).

The sample sets: i) CL_DM1 (generated with SuperScript III RT, Invitrogen, Carlsbad, CA, USA) consisted of three DM1 samples extracted from DM1 myoblast CL (9886, >200 CTG repeats; 10010, >200 CTGs; and 10011, >350 CTGs) and three sex- and age-matched control samples extracted from non-DM myoblast CL (9648, 10104, 10701) as described in Wojciechowska et al. (2014); ii) BP_DM1 (iScript RT, Bio-Rad) consisted of five DM1 and six control QF muscle samples; iii) BP_DM2 (GoScript RT, Promega) consisted of nine DM2 and four control samples, derived from QF or biceps branchii muscles; iv) MM_DM1 (SuperScript III RT, Invitrogen) consisted of 10 DM1-model and 10 control samples of the *HSA^{LR}* transgenic mouse model of DM1 and control background *FVB* mice, respectively. RNA was extracted from gastrocnemius muscle (Mankodi et al., 2000).

The samples, experimental protocols, and methods reported in this study were carried out in accordance with the approval of the local ethics committees: NRESCommittee.EastMidlands-Nottingham2 and the University of Rochester Research Subjects Review Board. Informed consent was obtained from all subjects.

TABLE 1 | Characteristics of sample sets used in the study.

	Sample set ID	Subset	# of samples
CL	CL_DM1	control	3
		DM1	3
	BP_DM1	control	6
BP	BP_DM1	DM1	5
		control	4
	BP_DM2	DM2	9
MM	MM_DM1	control, <i>FVB</i>	10
		DM1 model, <i>HSA^{LR}</i>	10

CL, human cell lines; BP, human muscle biopsies; MM, mouse muscles.

Selection of Circular RNAs for Experimental Analyses

Twenty circRNAs (Table 2) whose levels were experimentally evaluated in our study were selected from previously detected (Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2013; Rybak-Wolf et al., 2015; Zhang et al., 2016) circRNAs deposited in circBase (December 2016) (Glazar et al., 2014; <http://www.circbase.org/>). We considered only circRNAs validated by at least 20 next-generation sequencing (NGS) reads in at least two of the previously mentioned studies. Fourteen circRNAs were selected based on the relatively high level in different types of cells/tissues and a relatively high ($\geq 10\%$ in Jeck et al., 2013) proportion compared with that of their linear counterparts (mRNA). Four circRNAs were selected based on a high number ($n \geq 10$) of potential MBNL-binding sites (YGCY motifs; Goers et al., 2010) in adjacent (300 nt upstream and 300 nt downstream) sequences of their flanking introns. Two additional circRNAs selected for analysis were circCDR1as and circMBNL1. Additionally, eight circRNAs were experimentally analyzed for the purpose of validation of the most differentiated circRNAs identified based on RNA-Seq data analysis of control and DM1 QF samples (see later). In the mouse sample set, we analyzed seven circRNAs. Five of them, i.e., circCamsap1, circHippk3, circNfatc3, circZkscan1, and circCdr1as, were selected based on orthology to the human circRNAs analyzed in our study (see Table 2). Two circRNAs, i.e., circZfp609 and circBnc2, were selected based on their recently reported role in the skeletal muscle (Legnini et al., 2017; Wang et al., 2019).

PCR Assays Design and Validation

For the experimental analysis of selected circRNAs, we designed PCR assays that allowed the amplification and parallel analysis of circRNAs and their linear counterparts. Each assay consisted of one primer common to the circular and linear transcript and two primers specific for either circular or linear transcript. The only exceptions were assays designed for circCDR1as (circRNA generated from a single-exon transcript) and circMBNL1, which consisted of four primers (two for the circular transcript and two for the linear transcript). Primer sequences are shown in Table S1.

The PCR products of the designed assays were validated by analysis in agarose gel electrophoresis (the length of each product was as expected). Briefly, PCR was performed in a 10- μ l reaction composed of 0.3 μ l of a 10- μ M dilution of forward and reverse primers (0.6 μ l in total; primers were synthesized by Sigma-Aldrich, Saint Louis, MO, USA), 0.125- μ l deoxynucleotide triphosphate mix (concentration of each nucleotide was 10 mM) (Promega), 0.05- μ l GoTaq DNA Polymerase (concentration 5 u/ μ l) (Promega), 2- μ l 5 \times colorless GoTaq reaction buffer (containing 7.5 mM MgCl₂) (Promega), 6.225- μ l deionized water, and 1- μ l cDNA template. The following cycling conditions were used: 2 min at 95°C, followed by 35 cycles at 95°C for 20 s, 58–60°C (different for individual assays) for 20 s, and 72°C for 20 s, followed by 5 min at 72°C. The obtained PCR products were visualized on a standard 1.5% agarose gel. Additionally, the specificity of each product was confirmed by Sanger sequencing

TABLE 2 | CircRNAs selected for analysis.

circRNA	circBase ID	Genome localization (hg 19)	Homing gene	circRNA:mRNA ratio (Jeck et al., 2013)	Number of potential MBNL-binding motifs
circRNAs selected based on high level in different cells/tissues					
circASXL1	hsa_circ_0001136	20:30954186 30956926	ASXL1	286%	6
circCASPAP1	hsa_circ_0001900	9:138773478 138774924	CASPAP1	253%	10
circFAM13B	hsa_circ_0001535	5:137320945 137324004	FAM13B	34%	5
circHIPK3	hsa_circ_0000284	11:33307958 33309057	HIPK3	721%	2
circMBOAT2	hsa_circ_0000972	2:9048750 9098771	MBOAT2	19%	5
circMIB1	hsa_circ_0000835	18:19345732 19359646	MIB1	26%	1
circNFATC3	hsa_circ_0000711	16:68155889 68160513	NFATC3	52%	3
circPHC3	hsa_circ_0001359	3:169854206 169867032	PHC3	22%	5
circPIP5K1C	hsa_circ_0000871	19:3660963 3661999	PIP5K1C	11%	2
circSCMH1	hsa_circ_0000061	1:41536266 41541123	SCMH1	23%	3
circSHKBP1	hsa_circ_0000936	19:41089303 41089623	SHKBP1	14%	9
circUBAP2_e7-8	hsa_circ_0001851	9:33971648 33973235	UBAP2	82%	4
circUBAP2_e9-12	hsa_circ_0001847	9:33953282 33963789	UBAP2	63%	3
circZKSCAN1	hsa_circ_0001727	7:99621041 99621930	ZKSCAN1	99%	14
circRNAs selected based on high number of potential MBNL-binding motifs					
circCCDC134	hsa_circ_0001238	22:42204878 42206295	CCDC134	98%	10
circFOXK2	hsa_circ_0000816	17:80521229 80526077	FOXK2	26%	12
circPDCD11	hsa_circ_0000258	10:105197771 105198565	PDCD11	129%	11
circPROSC	hsa_circ_0001788	8:37623043 37623873	PROSC	15%	11
circRNAs additionally included in the analysis					
circCDR1as	hsa_circ_0001946	X:139865339 139866824	CDR1as	–	5
circMBNL1	hsa_circ_0001348	3:152017193 152018156	MBNL1	–	7

performed on an ABI Prism 3130 genetic analyzer (Applied Biosystems, Carlsbad, CA, USA) according to the manufacturer's general recommendations.

Droplet Digital PCR

The level of circRNAs was analyzed with the use of the ddPCR technique (Hindson et al., 2011; Miotke et al., 2014) developed by Bio-Rad. ddPCR involves partitioning the analyzed sample into many low-volume droplet reactions, and only a fraction of these reactions contains one (in most cases) or more template molecules (positive droplets). The final concentration of the analyzed templates was determined by Poisson statistical analysis of the number of positive and negative droplets. ddPCR analyses were performed according to the manufacturer's general recommendations. Briefly, reactions were carried out in a total volume of 20 µl, containing 10-µl 2× EvaGreen Supermix (Bio-Rad), 1 µl 4 µM forward primer, 1 µl 4 µM reverse primer, and different amounts of cDNA template, determined on the basis of optimization reactions performed for each analyzed gene/transcript. A QX200 ddPCR droplet generator (Bio-Rad) was used to divide the reaction mixture into up to 20,000 droplets. The initial dilution of the cDNA samples ensured that most of the generated droplets contained zero or one template molecule. The thermal parameters of the PCR were as follows: 5 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at annealing temperature (optimized for each gene) and 45 s at 72°C, followed by 2 min at 72°C, 5 min at 4°C, enzyme inactivation at 90°C for 5 min and holding at 12°C. The amplified products were analyzed using a QX200 droplet reader (Bio-Rad). The exact number of cDNA particles (representing particular transcripts) was calculated

based on the number of positive (containing template cDNA molecules) and negative (without template cDNA molecules) droplets using QuantaSoft (Bio-Rad) version 1.7.4.019 software, which utilizes Poisson distribution statistics.

In the analyses, we took the factor of the aforementioned cDNA dilution into account. Importantly, in our analysis, we used the following exclusion criteria: i) from the analysis of the level of a particular circRNA, we excluded samples with less than 10 positive droplets corresponding to the linear counterpart of this circRNA; ii) in the individual sample set, we did not consider the analysis of a particular circRNA if more than half of the samples (including DM and control samples) were excluded from the analysis in step i. Additionally, due to the limited amount of RNA samples, not all of the originally selected circRNAs were tested in the BP_DM2 sample set.

For each analyzed circRNA, their levels in particular samples were calculated as a fraction of circular particles (FCP) constituted by the amount of circRNA particles (C) in a total number of particles [circRNAs (C) and their linear counterparts (L)] generated from a particular gene:

$$FCP = C / (C + L) \quad (1)$$

The only exception was circCDR1as for which both linear and circular transcripts are generated from the same single exon (PCR primers designed for analysis of linear transcripts are also specific to cDNA generated from circular transcripts). Thus, the equation in this case is as follows:

$$FCP = C / L \quad (2)$$

Additionally, the levels of circRNAs and their linear counterparts were normalized against the levels of housekeeping genes (i.e., *ACTB* and *GAPDH*).

Analysis of Next-Generation Sequencing Data

For the purpose of global analysis of circRNA expression, we used the RNA-Seq data [Gene Expression Omnibus (GSE86356)] deposited in the DMseq database (Wang et al., 2018) (<http://www.dmseq.org/>). From the data sets of 126 samples derived from different muscle tissues, we chose the data sets of muscles represented by the highest number of samples, i.e., QF and TA. To avoid potential technical variations for analysis, we selected only samples for which sequencing data were generated with uniform procedures. For each sample, paired-end sequencing libraries were prepared from rRNA-depleted total RNA. Reverse transcription was performed using random primers, followed by second strand cDNA synthesis, end repair, adenylation, and ligation of adapters. Sequencing was performed using an Illumina HiSeq 2000 system (Illumina, San Diego, CA, USA), followed by processing with standard HiSeq 2000 software. Reads were mapped to the human genome (GRCh37/hg19) using Hisat2 (Kim et al., 2015). For the analysis, we selected data sets for 23 QF samples (11 control samples and 12 DM1 samples) and 27 TA samples (six control samples and 21 DM1 samples). The GSM accession numbers of selected samples are shown in **Table S2**. The average number of mappable reads in selected samples was ~29 million (ranging from ~18 to ~97 million reads; median ~26 million reads) and constituted 92% of the total library size on average. The length of reads was 60 nt. The detection and quantification of circRNAs and their linear mRNA counterparts in the selected samples was performed with CIRI2 (Gao et al., 2018), which uses maximum likelihood estimation based on multiple seed matching. This tool enables the identification of back-spliced junction reads and the filtration of false positives derived from repetitive sequences and mapping errors. The normalized level of circRNAs was calculated either as a number of circRNA-specific reads per million mappable reads (RPM) or as a fraction of circRNA-specific reads in a total number of circRNA-specific and corresponding linear reads (FCR). Note that FCR corresponds to FCP calculated based on the number of circular and linear RNA particles. The level of circRNAs was also normalized against the number of reads specific to individual housekeeping genes (e.g., *ACTB* or *GAPDH*).

Statistical Analyses

All statistical analyses were performed using Statistica (StatSoft, Tulsa, OK, USA) or Prism v. 5.0 (GraphPad, San Diego, CA, USA). All *p*-values were provided for two-sided tests. If necessary, the false discovery ratio (FDR) was calculated according to the Benjamini–Hochberg procedure (<http://www.biostathandbook.com/multiplecomparisons.html>). To compare the observed and expected (equal occurrence of increases and decreases) proportion of increased and decreased circRNAs in DM samples in particular sample sets or RNA-Seq data sets, we used the χ^2 test. All human genome positions indicated in

this report refer to the February 2009 (GRCh37/hg19) human reference sequence. The functional association analysis of the genes corresponding to circRNAs was performed with the use of DAVID Bioinformatics Resources (Huang et al., 2009a; Huang et al., 2009b). The computational prediction of exons in MBNL1 was performed with the GENSCAN online tool (<http://genes.mit.edu/GENSCAN.html>), using the default filters (i.e., organism: vertebrate; suboptimal exon cutoff: 1.00; print options: predicted peptides only). The exon prediction was performed in the sequence of the second exon of *MBNL1* flanked by directly adjacent 1-kb fragments of downstream and upstream introns (coordinates of analyzed sequence: chr3:152016193–152019155). Correlations of circRNA levels with DM1 severity were performed for TA samples with the use of phenotypic (ankle dorsiflexion force) and splicing alteration data deposited in the DMseq database.

RESULTS

Selection of Circular RNA Species for Expression Analysis in DM1

To check whether the level of individual circRNAs is affected in DM1, we selected 20 circRNAs reported in previous studies (Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2013; Zhang et al., 2014; Rybak-Wolf et al., 2015) and deposited in circBase (Glazar et al., 2014; <http://www.circbase.org/>). To avoid falsely identified circRNAs, we considered only circRNAs validated by at least 20 NGS reads in at least two previous studies. Fourteen circRNAs (**Table 2**) were selected based on their relatively high levels (compared with other circRNAs) in different types of cells/tissues and relatively high ($\geq 10\%$ in Jeck et al., 2013) expression levels compared with that of their linear counterparts (mRNAs). Four circRNAs (**Table 2**) were selected based on a high number ($n \geq 10$) of potential MBNL-binding sites (YGCT motifs; Goers et al., 2010) in adjacent (300 nt upstream and 300 nt downstream) sequences of their flanking introns. Additionally, we selected circCDR1as (hsa_circ_0001946) (Hansen et al., 2011; Memczak et al., 2013), the well-studied circRNA generated from the antisense transcript of the *CDR1* gene (*CDR1as*), and circMBNL1 (hsa_circ_0001348) (**Table 2**), which derives from the second exon of *MBNL1*, which is reportedly involved in the self-regulation of *MBNL1* expression (Konieczny et al., 2017) and linked to circRNA biogenesis (Ashwal-Fluss et al., 2014).

Design of Assays to Analyze Circular RNA Expression

For each selected circRNA, we designed PCR assays allowing amplification and parallel analysis of a given circRNA and its linear mRNA counterpart. Each assay consisted of three primers as follows: one primer common to both the circular and linear transcripts and two primers specific for either the circular or linear transcript (**Figure 1A**, **Table S1**). The size of circRNA-specific amplicons was analyzed by agarose gel electrophoresis (**Figure 1B**), and the predicted back-splice sites were subsequently confirmed by Sanger sequencing (**Figure 1C** and **Figure S2**).

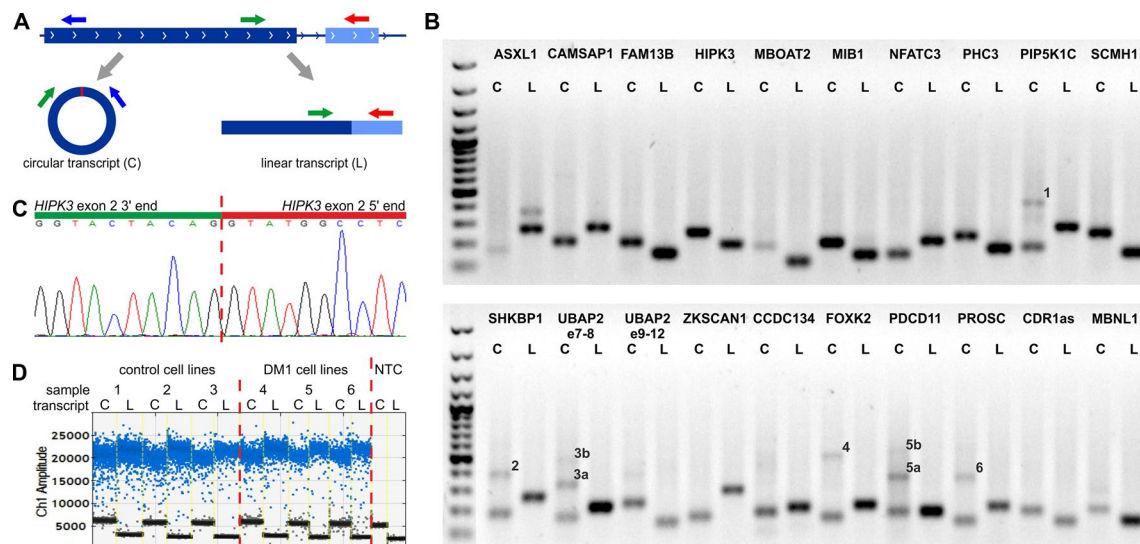


FIGURE 1 | Design and validation of the assays used for analysis of circRNA levels. **(A)** Each assay consisted of three primers as follows: one primer (green arrow) common to the circular (C) and linear (L) transcripts (circRNA and mRNA, respectively) and two primers specific for either circular (blue arrow) or linear (red arrow) transcripts. The primers specific to linear transcripts were located in either the downstream or upstream exon, outside of circRNA-coding exons. **(B)** Gel electrophoresis confirming the size of circRNA-specific (C) and linear, mRNA-specific (L) amplicons. Additional bands in some tracks corresponding to circRNAs indicate the occurrence of circRNA-related concatemers (see You et al., 2015), labeled by numbers on the gel. The expected lengths of the indicated concatemers are: 1–427 bp, 2–383 bp, 3a–314 bp, 3b–473 bp, 4–498 bp, 5a–357 bp, 5b–537 bp, and 6–360 bp. An additional band in ASXL1 linear transcript track corresponds to the alternative transcript containing alternatively included (97-nt long) exon 5. The first track is the GeneRuler 1-kb DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA). PCR reaction was performed for a control CL sample from CL_DM1 sample set. **(C)** Exemplary result of Sanger sequencing of the predicted back-splice site of circHIPK3. Results of sequencing of back-splice sites of other circRNAs are shown in **Figure S2**. **(D)** Exemplary result of the ddPCR analysis of circHIPK3 in the myoblast CL (CL_DM1) sample set. Sample number and type are indicated above the graph. NTC—no template control. Ch1 Amplitude—relative fluorescence signal in channel 1. Each blue dot represents one copy of either circular or linear transcript (positive droplets), while the black dots represent negative (empty) droplets. For each sample, the number of positive and negative droplets was used to calculate the concentration of the analyzed transcript.

The specific assays were employed for quantification of cDNA copies corresponding to circRNA and linear mRNA transcripts using ddPCR that enables absolute quantification of nucleic acid templates (**Figure 1D**; for details, see Materials and Methods).

Additionally, gel electrophoresis of the PCR product specific for circMBNL1 revealed an additional longer band. Analysis of this additional band led to the identification and characterization of a new circRNA (circMBNL1') consisting of the second exon of *MBNL1* and a 93-nt fragment of the large (~114-kb long) downstream intron 2 (**Figure S3**). The analysis of the surrounding sequence with the GENESCAN online tool identified (with high confidence) the incorporated fragment of intron as an exon, with canonical 5' and 3' splice sites.

Analysis of Expression Levels of the Selected Circular RNAs in DM Samples

Human myoblast cell lines (CL), as well as skeletal muscle biopsy (BP) tissues from DM1, DM2, and non-DM controls, were used to compare expression levels of circRNAs in DM and unaffected samples in three different sample sets (defined in Materials and Methods and **Table 1**). As shown in **Figure 2**, four circRNAs (i.e., circCAMSAP1, circHIPK3, circNFATC3, and circZKSCAN1) that were consequently analyzed in all sample sets, in all

but two cases (circNFATC3 in the CL_DM1 sample set and circZKSCAN1 in the BP_DM1 sample set) showed an increase in DM samples. Also, the other selected circRNAs tend to be rather increased than decreased in DM samples (**Figure S4**) (for details, see **Table 3** and **Table S3**). The marginally significant differences of the individual circRNA levels are indicated by asterisks on the graphs. A similar effect was observed when the circRNAs levels were normalized against the levels of housekeeping genes (*GAPDH* and *ACTB*; data not shown).

The disadvantage of analysis of human biopsy samples is that they may not always be of homogenous quality (e.g., different sample sources or divergent tissue and/or RNA treatment protocols may result in differences in RNA integrity). Moreover, the limited access to this type of samples and consequently small sample sets does not always allow the detection (with appropriate statistical support) of smaller changes in the levels of analyzed transcripts. Therefore, in the next step, we used cDNA samples from muscles of the commonly used and well-characterized mouse model of DM1 (*HSA*^{LR}, Mankodi et al., 2000) and compared them with samples from control background (*FVB*) mice. For analysis, we selected five mouse circRNAs (circCamsap1, circHipk3, circNfatc3, circZkscan1, and circCdr1as) that are orthologs of the human circRNAs analyzed in this study. Additionally, we analyzed two circRNAs reportedly

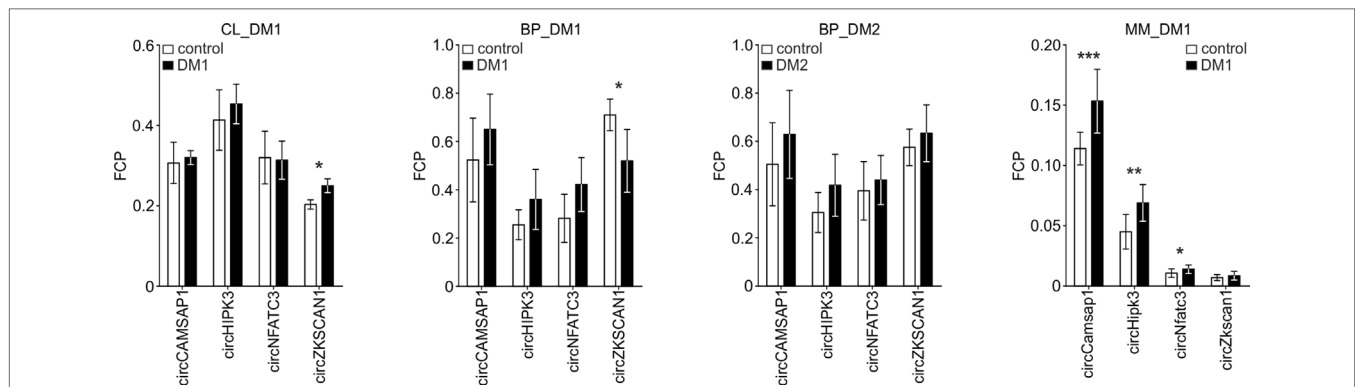


FIGURE 2 | Comparison of circCAMSAP1, circHIPK3, circNFATC3, and circZKSCAN1 levels in DM and control samples. The tested sample sets are indicated above the graphs. The values indicated by white or black bars are the averaged fraction of circular particles (FCPs) calculated for either control or DM samples, respectively. The whiskers represent the standard deviation (SD) values. The differences in circRNA levels were compared with *t*-tests, and the asterisks represent the following significance levels: **p* < 0.05; ***p* < 0.01; ****p* < 0.001. The exact *p*-values, SD values, and numbers of samples analyzed are indicated in **Table S3**.

TABLE 3 | Results of experimental analyses of circRNA expression levels.

	DM1/DM2 status			
	CL_DM1	BP_DM1	BP_DM2	MM_DM1
circCAMSAP1	↑ (<i>p</i> = 0.69)	↑ (<i>p</i> = 0.27)	↑ (<i>p</i> = 0.27)	↑ (<i>p</i> = 0.0009)
circHIPK3	↑ (<i>p</i> = 0.48)	↑ (<i>p</i> = 0.12)	↑ (<i>p</i> = 0.14)	↑ (<i>p</i> = 0.002)
circNFATC3	↓ (<i>p</i> = 0.89)	↑ (<i>p</i> = 0.06)	↑ (<i>p</i> = 0.49)	↑ (<i>p</i> = 0.049)
circZKSCAN1	↑ (<i>p</i> = 0.02)	↓ (<i>p</i> = 0.03)	↑ (<i>p</i> = 0.40)	↑ (<i>p</i> = 0.2)
circASXL1	↑ (<i>p</i> = 0.91)	ex	↓ (<i>p</i> = 0.15)	—
circFAM13B	↑ (<i>p</i> = 0.31)	↓ (<i>p</i> = 0.12)	—	—
circMBOAT2	↓ (<i>p</i> = 0.51)	ex	—	—
circMIB1	↓ (<i>p</i> = 1.00)	↓ (<i>p</i> = 0.91)	—	—
circPHC3	↓ (<i>p</i> = 0.55)	↑ (<i>p</i> = 0.16)	↑ (<i>p</i> = 0.26)	—
circPIP5K1C	↓ (<i>p</i> = 0.87)	↑ (<i>p</i> = 0.08)	—	—
circSCMH1	↑ (<i>p</i> = 0.42)	↑ (<i>p</i> = 0.42)	—	—
circSHKBP1	↓ (<i>p</i> = 1.00)	ex	—	—
circUBAP2_e9-12	↑ (<i>p</i> = 0.09)	ex	—	—
circUBAP2_e7-8	↑ (<i>p</i> = 0.66)	↓ (<i>p</i> = 0.81)	—	—
circCCDC134	↑ (<i>p</i> = 0.55)	ex	—	—
circFOXK2	↑ (<i>p</i> = 0.27)	↑ (<i>p</i> = 0.08)	—	—
circPDCD11	↑ (<i>p</i> = 0.55)	ex	—	—
circPROSC	↑ (<i>p</i> = 0.79)	↑ (<i>p</i> = 0.27)	—	—
circCDR1as	↓ (<i>p</i> = 0.70)	↓ (<i>p</i> = 0.15)	—	↓ (<i>p</i> = 0.07)
circMBNL1	↑ (<i>p</i> = 0.82)	ex	—	—
circBnc2	—	—	—	↑ (<i>p</i> = 0.25)
circZfp609	—	—	—	↑ (<i>p</i> = 0.0005)
chi², <i>p</i>-value	0.17	0.4	0.1	0.06

↑, circRNA level increased; ↓, circRNA level decreased; ex, excluded from analysis due to low number of positive droplets (see Materials and Methods); —, not analyzed; **bolded**, circRNAs analyzed in all sample sets.

involved in muscle development, i.e., circZfp609 (ortholog of human circZNF609) and circBnc2 (ortholog of human circBNC2) (Legnini et al., 2017; Wang et al., 2019). As shown in **Table 3** and **Figure S4**, the levels of four out of seven circRNAs tested in mice (i.e., circCamsap1, circHipk3, circNfatc3, and circZfp609) were significantly increased in *HSA*^{LR} mice.

In conclusion, our experimental analyses show a trend toward some increase of circRNA level in DM (especially supported in the DM1 mouse model).

Analysis of Circular RNA Levels in DM1 With RNA-Seq Data Sets

CircRNAs selected for the experiments described previously may not be representative, and global circRNA level changes may be too small to be detected with a few circRNAs. Therefore, in the next step, to better evaluate the global circRNA level, we used the RNA-Seq data deposited in the DMseq database (Wang et al., 2018). For the analysis, we selected data sets of muscle samples most frequently represented in the database, QF muscle

(11 control samples and 12 DM1 samples) and TA muscle (six control samples and 21 DM1 samples). To avoid potential technical variations in analysis, we selected only samples with sequencing data generated with uniform procedures (for details, see Materials and Methods).

In total, in QF samples, we detected 22,816 distinct circRNAs (“all”; a substantial fraction were confirmed by just a few reads), 4,168 (18%) of which were classified as “validated” (confirmed by at least five reads in at least two samples), and 152 (0.7%) were classified as “common” (present in all or all but one sample of either control or DM1 samples). In the case of TA samples, the “all” group contained 38,403 circRNAs, and the “validated” and “common” groups contained 7,537 (20% of “all”) and 403 (1% of “all”) circRNAs, respectively. As expected, the fraction of known (deposited in circBase and in Maass et al., 2017) circRNAs increased with the level of validation in both QF and TA (Table 4, Table S4, Table S5).

To compare the global level of circRNA in control and DM1 samples, in each sample, we summarized the number of reads [normalized as reads per million mappable reads (RPMs)] mapping to back-splice sequences (circRNA level) and mapping to the corresponding linear-splice sequences (linear mRNA level). As shown in Figure 3, the average global level of “common” circRNAs was significantly increased in DM1 samples ($p = 0.004$ in QF and $p < 0.0001$ in TA). Importantly, no difference was detected compared with corresponding linear transcripts ($p = 0.6$ and $p = 0.1$ in QF and TA, respectively). The increased level of circRNA in DM1 samples was also visible for “all” and “validated” circRNAs (Figure S5). Similar results were obtained when the level of transcripts (number of reads) was normalized against the level of individual housekeeping genes, e.g., *ACTB* or *GAPDH* (data not shown).

The previously mentioned changes in circRNA levels may be a reflection of an increase or decrease of expression from a particular gene or genome region. To control for this effect, we also normalized the levels of circRNAs against the levels of their linear counterparts, calculating the level of circRNAs as fraction of circRNA-specific reads in a total number of circRNA-specific and corresponding linear reads (FCR). Again, the cumulative

value or averaged FCRs were higher in DM1 samples than in control samples (right graphs in Figure 3 and Figure S5). Additionally, in this analysis, circular transcripts of “common” circRNAs accounted for ~5–10% of their linear counterparts.

Differential Expression of Individual Circular RNAs

Although it was not the main purpose of the study, by using the generated data, we also analyzed the differential expression of individual circRNAs. This analysis was limited to only the sets of “common” circRNAs ($n = 152$ in QF and $n = 403$ in TA) with expression levels detectable in the vast majority of analyzed samples. The difference in circRNA levels was calculated for the level of circRNAs normalized as RPMs and FCRs of individual circRNAs and expressed as log2 of fold change in DM1 samples vs. control samples. In both QF and TA, the changes in circRNA levels calculated with two normalization methods were highly correlated (Figure S6), indicating that circRNA changes do not depend on the expression of genes (level of their primary transcripts) from which they are generated. The results of the analyses are shown in Table S6 and Table S7 and graphically summarized in the form of volcano plots (Figure 4). The list of circRNAs differentially expressed (RPM value difference at the p -value level < 0.05) in both QF and TA are shown in Table 5; note that four circRNAs are significantly differentiated after adjustment for multiple comparisons in both tissues. Most of the top differentially expressed circRNAs are deposited in circBase, and the majority of them are encoded by exons of known genes (Tables S4, S5, S6, and S7). As shown in Figure 4, log2 fold change values are substantially shifted toward positive values, indicating an excess of circRNAs with increased levels in DM1 samples. This effect is in line with the global increase in circRNA levels in DM1 (in both QF and TA) described previously. For example, assuming that results fulfilling the following thresholds are significant (p -value < 0.05 and log2 fold change ≤ -1 or ≥ 1), we obtained 38 and 120 differentially expressed circRNAs in QF and TA, respectively. Among these circRNAs, circRNAs with increased expression levels in DM1 (Figure 4) were substantially

TABLE 4 | Number of circRNAs in quadriceps femoris (QF) and tibialis anterior (TA) tissues in different validation groups.

		“All”			“Validated”			“Common”		
		Total	Known	New	Total	Known	New	Total	Known	New
QF	control	11,960	5,085 (42.5%)	6,875 (57.5%)	3,566	2,425 (68.0%)	1,141 (32.0%)	152	135 (88.8%)	17 (11.2%)
	DM1	16,131	6,503 (40.3%)	9,628 (59.7%)	4,078	2,715 (66.6%)	1,363 (33.4%)	152	135 (88.8%)	17 (11.2%)
	control + DM1	22,816	8,319 (36.5%)	14,497 (63.5%)	4,168	2,765 (66.3%)	1,403 (33.7%)	152	135 (88.8%)	17 (11.2%)
TA	control	10,022	4,593 (45.8%)	5,429 (54.2%)	4,515	3,007 (66.6%)	1,508 (33.4%)	403	336 (83.4%)	67 (16.6%)
	DM1	34,720	11,014 (31.7%)	23,706 (68.3%)	7,536	4,614 (61.2%)	2,922 (38.8%)	403	336 (83.4%)	67 (16.6%)
	control + DM1	38,403	11,816 (30.8%)	26,587 (69.2%)	7,537	4,615 (61.2%)	2,922 (38.8%)	403	336 (83.4%)	67 (16.6%)

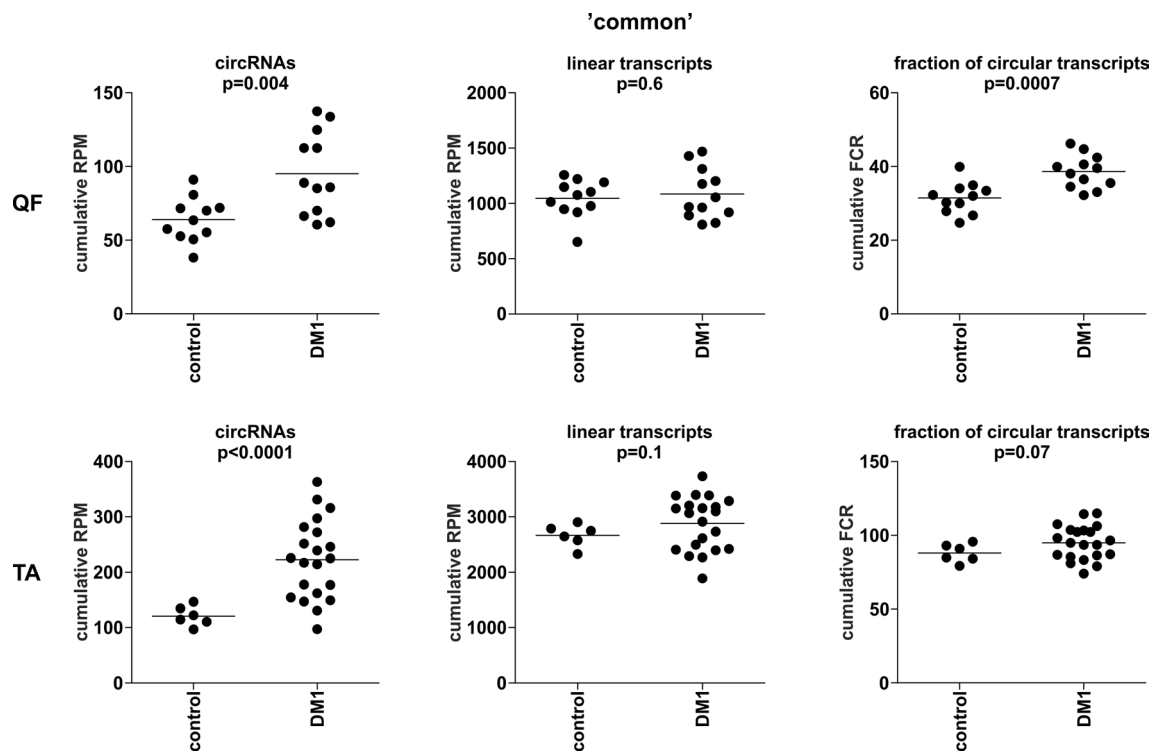


FIGURE 3 | Comparison of cumulative levels of circRNAs and linear RNAs in control and myotonic dystrophy type 1 (DM1) muscle samples. Dot plots depicting the cumulative level of “common” circRNAs and linear transcripts in quadriceps femoris (QF) (upper panel) and tibialis anterior (TA) (lower panel). From the left (in each panel): cumulative reads per million mappable reads (RPM) of circRNAs, cumulative RPM of linear transcripts, and cumulative fraction of circRNA-specific reads in a total number of circRNA-specific and corresponding linear reads (FCR). The false discovery ratio (FDR)-corrected p -value (t -test with correction for not-equal variance) of the differences between control and DM1 samples is shown above each dot plot.

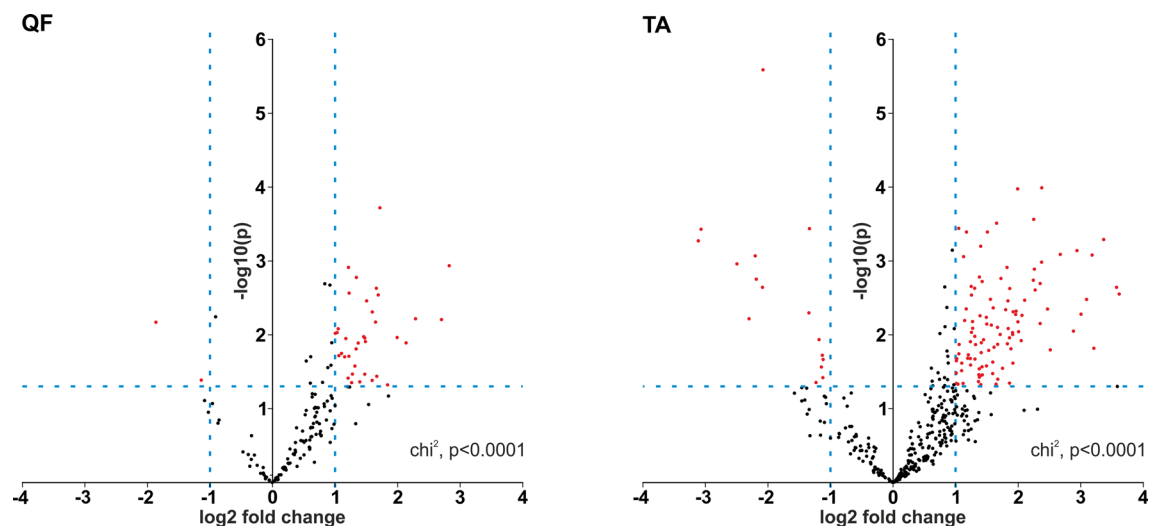


FIGURE 4 | Prevalence of circRNAs with increased levels in DM1. Volcano plots depicting differences in the levels of “common” circRNAs (dots) in DM1 and control samples in QF (left-hand side) and TA (right-hand side). Positive and negative values of \log_2 fold change indicate increased and decreased circRNAs in DM1. Each red dot represents circRNA fulfilling the following criteria of expression change: p -value < 0.05 and \log_2 fold change ≤ -1 or ≥ 1 (the thresholds indicated by dotted lines).

TABLE 5 | CircRNAs differentially expressed in both QF and TA muscles.

circRNA genome localization	Homolog gene	circBase ID	QF			TA		
			log2 fold change	p-value	FDR-corrected p-value	log2 fold change	p-value	FDR-corrected p-value
1:200816768 200822623	CAMSAP2	hsa_circ_0141534	0.835	0.002	0.049	0.946	0.001	0.019
1:35824526 35827390	ZMYM4	hsa_circ_0011536	1.005	0.010	0.073	1.050	0.0004	0.016
1:59805630 59844509	FGGY	hsa_circ_0006633	0.611	0.020	0.091	1.253	0.009	0.053
10:126628943 126631876	n/a	hsa_circ_0006545	1.264	0.044	0.146	1.426	0.035	0.118
10:34558585 34573173	PARD3	hsa_circ_0018168	1.370	0.013	0.073	1.671	0.008	0.050
11:33307959 33309057	HIPK3	hsa_circ_0000284	0.917	0.002	0.049	0.831	0.008	0.049
2:110919180 110920712	NPHP1	hsa_circ_0056019	1.718	0.0002	0.029	1.052	0.018	0.078
2:215632206 215646233	BARD1	hsa_circ_0001098	1.596	0.005	0.064	2.942	0.001	0.019
2:240929491 240946787	NDUFA10	hsa_circ_0001118	2.287	0.006	0.064	1.240	0.026	0.096
21:30693542 30702014	BACH1	hsa_circ_0001181	1.277	0.034	0.133	1.602	0.004	0.039
3:170906491 170912424	TNIK	hsa_circ_0002387	1.225	0.003	0.049	1.319	0.014	0.064
3:196118684 196129890	UBXN7	hsa_circ_0001380	1.031	0.009	0.073	1.175	0.0004	0.016
6:158703295 158735300	n/a	hsa_circ_0142312	1.103	0.018	0.091	1.673	0.013	0.064
6:158733083 158735300	n/a	hsa_circ_0142313	1.215	0.001	0.049	1.921	0.005	0.040
6:170846322 170858201	PSMB1	hsa_circ_0078784	1.661	0.002	0.049	1.960	0.005	0.041
7:18705836 18706099	HDAC9	hsa_circ_0007904	1.667	0.036	0.138	1.013	0.022	0.089
7:80418622 80440017	SEMA3C	hsa_circ_0004365	1.994	0.011	0.073	2.048	0.012	0.061
7:99621042 99621930	ZKSCAN1	hsa_circ_0001727	0.606	0.045	0.146	0.636	0.049	0.143
8:52773405 52773806	PCMTD1	hsa_circ_0001801	1.051	0.008	0.073	1.129	0.001	0.019
9:37424842 37426651	GRHPR	hsa_circ_0001861	1.647	0.007	0.064	1.373	0.035	0.118

bolded, circRNAs differentially expressed in QF and TA at the level of FDR-corrected p -value < 0.05 .

overrepresented [i.e., 36 (95%) in QF (χ^2 , $p < 0.0001$) and 104 (87%) in TA (χ^2 , $p < 0.0001$)]. Similar bias toward circRNAs increased in DM1 may also be seen with other methods of normalization (e.g., such as FCR or normalization against the level of housekeeping genes; data not shown) as well as with other cutoff thresholds. Among circRNAs for which both RPM and FCR values were decreased in DM1, we studied whether MBNL may contribute to their biogenesis. We conducted the analysis of introns (300 nt upstream and 300 nt downstream from circRNA-generating exons) flanking these circRNAs. However, we did not show enrichment of potential MBNL-binding motifs (n ranging from 1 to 9, in most cases $n \leq 5$) that would justify the role of MBNLs in their biogenesis. The interesting exception was circGSE1 (having as many as 30 potential MBNL-binding sites), with decreased RPM and RCF values in DM1 in TA [log2 fold change = -2.1 ; false discovery rate (FDR)-corrected p -value = 0.0001 and log2 fold change = -0.9 ; FDR-corrected p -value = 0.1, respectively].

The functional association analysis of genes corresponding to circRNAs either increased or decreased in DM1 in TA (67 distinct genes at $p < 0.01$ for differences in RPM, Table S7) showed the strongest association (enrichment) with the following UniProt keywords: “phosphoprotein” [number of involved genes (n) = 46, fold enrichment (FE) = 1.7, Benjamini corrected p -value (p_{BC}) = 0.0005] and “alternative splicing” (n = 52, FE = 1.5, p_{BC} = 0.001; Figure S7). The genes were also associated with the Gene Ontology cellular component (CC) term “nucleoplasm” (n = 24, FE = 2.5, p_{BC} = 0.004; Figure S7). A similar analysis performed for QF (18 distinct genes) also showed an enrichment of genes associated with alternative splicing and nucleus localization

keywords/terms among the top results (Figure S7), but the associations were nonsignificant due to the much smaller number of analyzed genes.

Identification of Multi-circRNA Genes

During the analysis, we noticed that a substantial number of circRNAs were generated from multi-circRNA genes (MCGs), which give rise to more than one circRNA. Furthermore, 14 MCGs in QF and 59 MCGs in TA (top-MCGs) generated more than 10 distinct circRNAs. As shown in Figure 5A (empty bars), cumulatively 69 and 78% of circRNAs were generated from MCGs, and 7 and 13% were generated from top-MCGs in QF and TA, respectively. The top-MCGs from which the highest numbers of circRNAs were generated were *titin* (*TTN*: 44 circRNAs in QF and 86 circRNAs in TA; cumulatively 96 distinct circRNA species), *nebulin* (*NEB*: 41 and 59; cumulatively 66), and *triadin* (*TRDN*: 24 and 37; cumulatively 39). All three genes are strongly related to biological functions and highly expressed in skeletal muscles. Other top-MCGs strongly related to the function of skeletal muscles are *dystrophin* (*DMD*), *myopalladin*, *myomesin 1*, and *myosin IXA*. Notably, the previously mentioned muscle-related multiexon MCGs were strongly enriched in new (not present in circBase) circRNAs (~ 95 vs 34%/39% in all “validated” circRNAs in QF/TA samples). This finding may have been observed because skeletal muscle tissues were not comprehensively studied (reported in the circBase) in the context of circRNA discovery.

The maps of genomic regions giving rise to circRNAs generated from top-MCGs common to QF and TA are shown

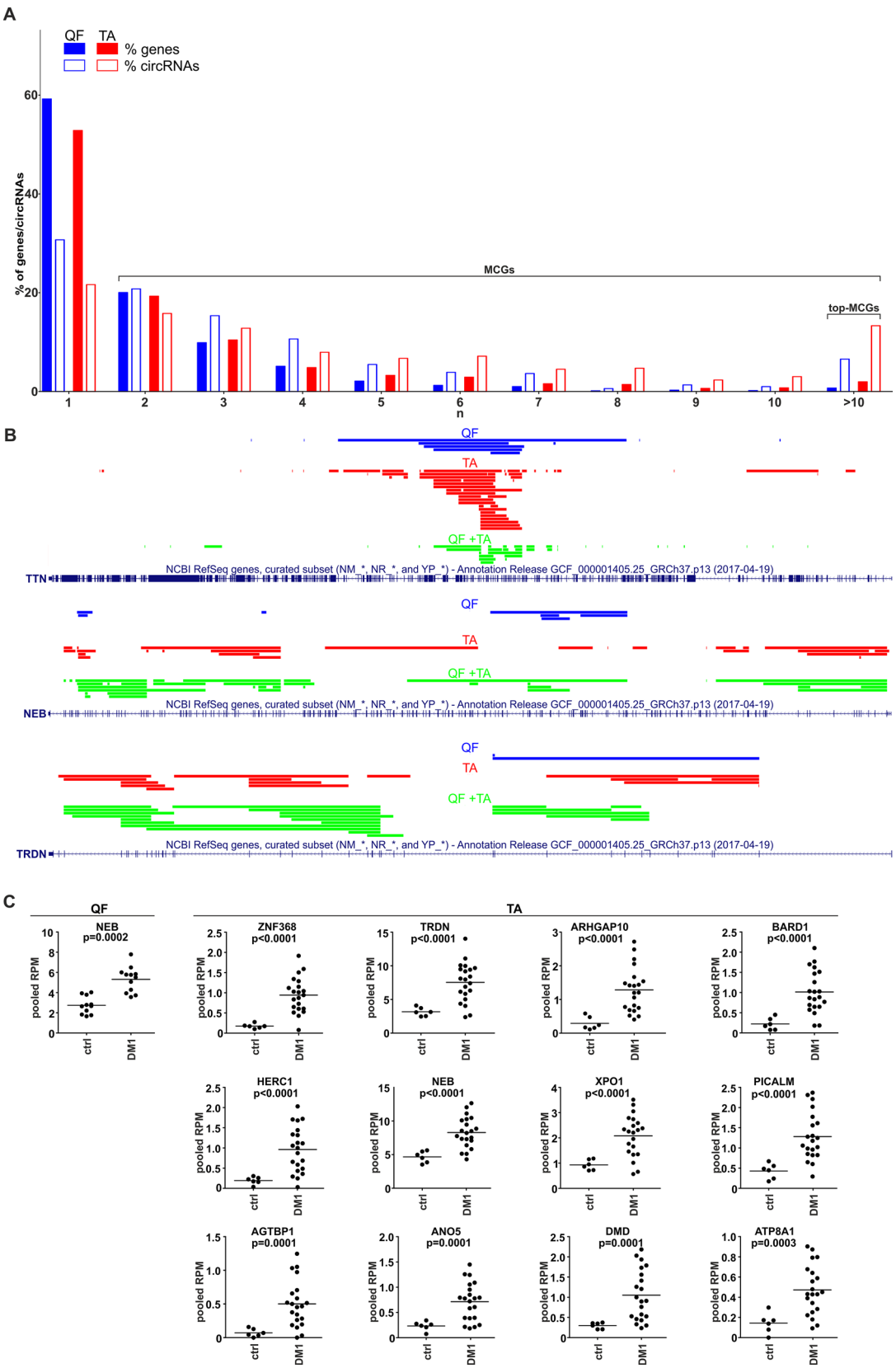


FIGURE 5 | Continued

FIGURE 5 | CircRNAs generated from multi-circRNA genes (MCGs). **(A)** Bar graph showing the percentage of genes that generate a particular number (n) of distinct circRNA species (solid bars) and the percentage of circRNAs generated from these genes (empty bars). Blue and red bars represent QF and TA, respectively. For example, in QF, the genes generating more than 10 circRNAs constitute ~1% of all circRNA-generating genes but generate ~7% of all circRNAs. **(B)** The maps of *TTN*, *NEB*, and *TRDN* (RefSeq tracks) with schematic representation of regions (color lines) overlapping exons giving rise to circRNAs (presented with the use of University of California—Santa Cruz Genome Browser). Blue, red, and green lines represent circRNAs specific to QF, specific to TA, common to QF and TA, respectively. **(C)** Dot plots depicting levels (pooled RPMs) of top-MCG-specific circRNA pools most profoundly differentiated between control (ctrl) and DM1 samples in QF and TA. The FDR-corrected p -value is shown above each graph. In each graph, each dot represents pooled circRNA-specific RPM values in the individual sample.

in **Figure 5B** and **Figure S8**. As shown in the figures, the back-splice sites of almost all circRNAs overlapped with the splice sites of canonical exons; therefore, almost all circRNAs may derive from the sequences of canonical exons. Moreover, a substantial fraction of circRNAs were common to QF and TA (green lines, QF + TA), and tissue-specific circRNAs mostly resulted from the higher number of circRNAs detected in TA. Interestingly, in most cases, circRNA-annotated sequences were not randomly distributed and clustered in the center of the gene. The effect was especially visible for circRNAs common to QF and TA. The most profound example of this distribution was *TTN*. The opposite example was *NEB* in which circRNA-annotated sequences were more or less randomly distributed over the entire gene. The observed distributions do not indicate that circRNAs are preferentially generated from exons flanked by long introns (Jeck et al., 2013).

The Level of Circular RNA Pools Generated From Particular Multi-circRNA Genes Increases in DM1

Considering circRNAs as competing regulators of linear transcripts, any circRNA generated from a particular gene may affect its linear-transcript-dependent expression. Therefore, in the next step, we compared the cumulative level of circRNAs generated from particular top-MCGs (circRNA pools) in control and DM1 samples. As shown in **Table S8** and **Table S9**, the cumulative RPM value of circRNA pools increased in DM1 samples in 11 out of 14 and 59 out of 59 top-MCGs in QF and TA, respectively. Similar results were also obtained for pooled FCRs (**Table S8** and **Table S9**), as well as for circRNA pools obtained with the other methods of circRNA level normalization (e.g., against the level of housekeeping genes; data not shown). In eight cases (i.e., *GBE1*, *SMARCC1*, *BIRC6*, *SENP6*, *CHD2*, *MYBPC1*, *MAP4K3*, and *RALGAP2*), the circRNA pools were increased, although none of the individual circRNAs constituting these pools were significantly differentiated. The levels of the most profoundly differentiated circRNA pools (FDR-corrected p -value < 0.0005) in QF and TA are shown in **Figure 5C**.

Circular RNA Levels Are Associated With DM Severity

The comparison of the global circRNA level in TA with a phenotypic biomarker of muscle strength (ankle dorsiflexion force) associated with DM1 severity showed a substantial correlation [correlation coefficient (R) = -0.85; p < 0.001]. Significant negative correlation with muscle strength (p < 0.05;

R < -0.434) showed also 117 (out of 403) individual “common” circRNAs and 42 (out of 59) top-MCGs-specific circRNA pools (**Figure 6**, **Table S10**).

In the next step, we compared the circRNA level with the level of early-, medium-, and late-responding alternatively spliced exons, being molecular biomarkers of DM1 severity. As shown in **Figure 6** and **Table S10**, the global circRNA level was significantly correlated with the percent spliced-in (PSI) values of all analyzed exons. The strongest correlation showed exon 7 of *MBNL1* (R = 0.88; p < 0.001), exon 8 of *CAPZB* (R = -0.85; p < 0.001), exon 29 of *CACNA1S* (R = -0.83; p < 0.001), and exon 22 of *ATP2A1* (R = -0.82; p < 0.001). Negative correlations were obtained for exons alternatively excluded in DM1. In contrast, exon 7 of *MBNL1* and exon 7 of *NFIX*, both alternatively included in DM1, showed positive correlations. Similar correlations were obtained for a substantial fraction of individual circRNAs, as well as for the top-MCG-specific circRNA pools (**Table S10**).

DISCUSSION

Splicing aberrations induced by functional inactivation of MBNL-splicing factors constitute a main pathomechanism of DM1. Previous research suggested that in addition to a function in alternative splicing, MBNL proteins participate in the biogenesis of circRNA, bringing circRNA-flanking introns closer together and facilitating back-splicing (circularization) (Ashwal-Fluss et al., 2014). Thus, downregulation of circRNAs would be expected in DM1 (and in DM2) cells in which expanded CUG (CCUG in DM2) repeats attract MBNLs, leading to their sequestration.

To test whether circRNA levels are changed in DM1 and to verify the role of MBNLs in the biogenesis of circRNA, we analyzed the expression level of up to 20 circRNAs in myoblast CL and skeletal muscle samples derived from patients with DM1 and healthy controls. Among the selected circRNAs were those with a relatively high number ($n \geq 10$) of potential MBNL-binding motifs in flanking introns, as well as circMBNL1, which is regulated by MBNL1 (Konieczny et al., 2017). Additionally, circCDR1as and circHIPK3, the highly expressed and most extensively studied circRNAs, were among the selected circRNAs (Hansen et al., 2011; Memczak et al., 2013; Zheng et al., 2016). None of the circRNAs tested in our analysis showed a consistent decrease of level in DM1. There was also no decrease in the levels of circRNAs in muscles from patients with DM2 or in muscles from the transgenic mouse model of DM1. All of the previous results question the role of MBNLs as important factors in circRNA biogenesis in muscles. The discrepancy between our study and earlier reports may be because previous analyses were

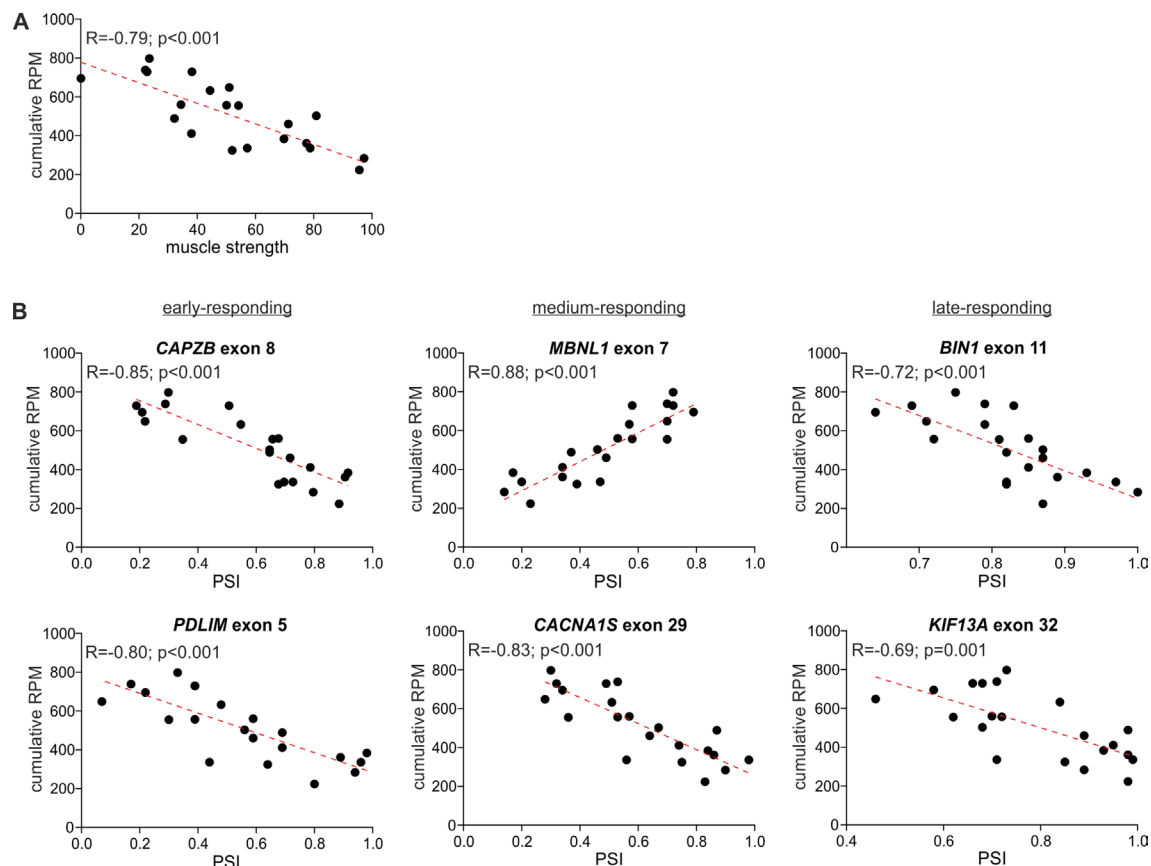


FIGURE 6 | Correlation of global circRNA level with disease severity. For each plot, the R -value, p -value, and the trendline (red-dotted line) are shown. **(A)** A scatter plot showing the correlation of global circRNA levels normalized as RPMs (Y-axis) and muscle strength (X-axis). **(B)** Scatter plots showing correlations of global circRNA levels (Y-axis) and PSI values of early-, medium-, and late-responding exons alternatively spliced in DM1 (X-axis). Each dot represents an individual TA sample.

performed in artificial models (artificially generated circRNA genes) in which some of the tested processes (e.g., interaction of MBNLs/Mbl with artificial, usually shorter introns) may take place differently, and the stoichiometry of interacting proteins and RNA particles may be different from those in natural mammalian tissues. Additionally, the previous experiments were mostly performed with the fly Mbl splicing factor. Potentially, human orthologs may not have the exact same circRNA-generation activity, and we cannot exclude the possibility that decreased levels of MBNLs, although they induce aberrations in alternative splicing, are still sufficient for circRNA processing. Furthermore, it is possible that MBNLs play a role in the biogenesis of specific individual circRNAs, which were not tested experimentally in our study. CircGSE1, flanked by multiple MBNL-binding motifs and decreased in DM1, may be an example of such a circRNA. MBNL1-dependent biogenesis of circGSE1 may be additionally supported by the fact that, contrasted with other circRNAs, its increased level is associated with lower DM1 severity (Table S10). Another example of circRNA decreased in DM1 and associated with lower DM1 severity is circFGFR1 (Table S10). In contrast to our original hypothesis, the previously mentioned experiments showed a trend toward a global increase

in circRNA levels in DM1 samples. Although changes in levels of individual circRNAs are small and nonsignificant in most cases, circRNAs with increased levels in DM samples were prevalent in most of our experiments. Additionally, analysis of a higher number of samples from mouse skeletal muscles that provided a better statistical power to detect smaller changes in circRNA levels showed that four out of the seven tested circRNAs were significantly increased in the mouse model of DM1, including circHipp3 regulating cell growth and differentiation (Zheng et al., 2016) and protein-coding circZfp609 playing role in myogenesis (Legnini et al., 2017; Wang et al., 2019).

To check whether the global circRNA level is indeed increased in DM1, we used publicly available RNA-Seq data sets deposited in the DMseq database (<http://www.dmseq.org/>). The advantage of such data is that they are generated by an independent experimenter blind to the hypotheses tested in particular studies (also in ours). The increased global level of circRNA in DM1 was confirmed in two independent sets of samples, consisting of samples from two different skeletal muscles, QF and TA.

CircRNAs, generated either cotranscriptionally or posttranscriptionally (Wilusz and Sharp, 2013; Ashwal-Fluss et al., 2014; Kramer et al., 2015), compete with their linear

counterparts (mRNAs) for their shared linear precursor (pre-mRNA). However, notably, some circRNAs are the main or exclusive products generated from their precursors (e.g., circCDR1as). The generation of circRNA may be a mechanism of mRNA downregulation (Ashwal-Fluss et al., 2014). Alternatively, disturbances and delays in mRNA maturation may increase the duration of the immature transcript and shift the balance of pre-mRNA processing in favor of circRNA biogenesis (Ashwal-Fluss et al., 2014; Zhang et al., 2016; Liang et al., 2017). In DM1, such disturbances in transcript maturation may be caused by the sequestration of MBNLs and aberrations in splicing. The increased global level of circRNA in DM1 may simply be a side effect of splicing aberrations or secondary effect of the chronic pathological state of DM1, not dependent on MBNL1 or splicing alterations.

Furthermore, as the levels of circRNAs are altered in such disorders as Duchenne muscular dystrophy (DMD) or dilated cardiomyopathy (Khan et al., 2016; Legnini et al., 2017), it may suggest that deregulation of circRNAs is generally associated with a muscle pathological state. It may be supported by the results of recent studies demonstrating changes of circRNA levels in different muscle diseases and physiological conditions. For example, it was shown that several circRNAs [e.g., circZNF609 (Legnini et al., 2017; Wang et al., 2019), circQKI (Legnini et al., 2017), circBNC2 (Legnini et al., 2017), circFGFR4 (Li et al., 2018a), and circFUT10 (Li et al., 2018b)] are differentiated in different muscle conditions and may be involved in the regulation of myoblast proliferation and muscle cell development (well reviewed in Greco et al., 2018).

Also, there are some facts that may link elevated global levels of circRNA or increased levels of an individual circRNA with DM1 pathogenesis. First, the global circRNA level, as well as the levels of substantial fractions of MCG-specific circRNA pools and individual circRNAs were negatively correlated with molecular and clinical biomarkers of DM1 severity. Second, gene ontology analysis of the circRNA genes that were increased in DM1 showed enrichment of the aberrant splicing, phosphoprotein, and nucleoplasm terms. It seems particularly interesting considering that aberrant alternative splicing is one of the most prominent molecular markers of DM1 (Philips et al., 1998; Wang et al., 2007) and it is linked with hyperphosphorylation of CUGBP1 protein (Kuyumcu-Martinez et al., 2007; Wang et al., 2009). Additionally, as recently shown (Ketley et al., 2014; Wojciechowska et al., 2014), utilization of kinase inhibitors alleviated some of the molecular symptoms of DM1, among others, diminishing the nuclear fraction of mutant DMPK transcripts (Ketley A et al., IDMC-11, San Francisco 2017). Third, DM pathogenesis may be also linked with the increase of circZfp609, which was observed in our study. It was recently shown that the level of circZfp609, as well as the level of its human ortholog (circZNF609), is increased in proliferating myoblasts and is downregulated during myogenesis (decreased in more differentiated muscle cells). Functional tests demonstrated that circZfp609/circZNF609 plays a role in promoting myoblast proliferation (possibly by sponging miR-194-5p) (Legnini et al., 2017; Wang et al., 2019). This suggests that an increased level of circZfp609/circZNF609 may delay muscle differentiation and maturation. An increase

of circZNF609 was also observed in DMD cells (Legnini et al., 2017) that suggests a link between both dystrophies, i.e., DM and DMD. Finally, recent results by Voellenkle et al. showed that the levels of four-out-of-nine-tested circRNAs were significantly increased in DM1 patients and correlated with muscle weakness (Voellenkle et al., 2019).

By using generated circRNA data sets, we also performed analyses of individual circRNAs and MCG-specific circRNA pools. The analyses led to the identification of many circRNAs and circRNA pools that were significantly differentiated between DM1 and control samples. In both types of analyses and in both analyzed tissues, there was a substantial excess of circRNAs or circRNA pools in DM1. This finding is consistent with the observation of the global increase in circRNA levels in DM1 samples. Although many of the changes in circRNA and circRNA pools reached statistical significance ($p < 0.05$, even after FDR correction), whether the differentiated circRNAs/circRNA pools are specific and biologically relevant to DM1 or result from a global increase in circRNA levels in DM1 cannot be established. One hint as to the role of circRNAs in DM1 may be found in the functional association analysis, which showed that terms related to alternative splicing and nuclear localization were among the strongest associations of genes giving rise to differentiated circRNAs. Other links between aberrations in circRNA levels and DM1 pathogenesis come from the observed associations between circRNA levels and muscle weakness, as well as between circRNA levels and abnormalities of alternative splicing of well-known DM biomarkers. Additionally, the transcripts of at least 10 (DMD, *KIF1B*, *MYBPC1*, *NEB*, *NCOR2*, *PICALM*, *RERE*, *SMARCC1*, *UBAP2*, and *USP25*) out of 63 identified top-MCGs were previously shown to be aberrantly spliced in DM1 (Du et al., 2010; Nakamori et al., 2013). Nonetheless, the changes in individual circRNAs require further experimental validation. Moreover, notably, the power of this analysis is limited due to the depth of coverage (adjusted for mRNA analysis) that does not allow reliable estimation of low-level circRNAs.

Interestingly, among the top-MCGs, there are genes highly expressed and strongly associated with the biological function of skeletal muscles [e.g., *TTN* (total number of circRNAs generated in both QF and TA, $n = 96$), *NEB* ($n = 66$), *TRDN* ($n = 39$), *DMD* ($n = 33$), *myopalladin* ($n = 22$), *myomesin 1* ($n = 18$), or *myosin IXA* ($n = 14$)]. All of these genes are large multiexon genes, including *DMD* (2.1 Mbp, up to 81 exons), the largest human gene, and *TTN* (0.3 Mbp, up to 362 exons), which has the highest number of exons (Figure 5B and Figure S8). A large number of exons increase the number of potential splicing donor/acceptor pairs, which may facilitate the generation of different circRNAs. Alternatively, the higher number of circRNAs generated from multiexon genes may also result from higher chances/numbers of aberrations occurring during processing of their transcripts.

In conclusion, our results indicate that MBNL deficiency does not cause the expected decrease in circRNA levels in DM1 cells and tissues. In contrast, the global level of circRNAs is elevated in DM1. However, the role of the increased level of circRNAs in the pathogenesis of DM1 is unknown and requires further investigation.

CONTRIBUTION TO THE FIELD STATEMENT

Recently, a great deal of interest has been focused on a new class of RNA molecules called circular RNAs (circRNAs). To date, thousands of circRNAs have been found in different human cells/tissues. Although the function of circRNAs remains mostly unknown, circRNAs have emerged as an important component of the RNA–RNA and RNA–protein interactome. Thus, intensive efforts are being made to fully understand the biology and function of circRNAs, especially their role in human diseases. As an important role in the biogenesis of circRNA may be played by MBNL-splicing factors, in this study, we used DM1 (to a lesser extent, DM2) as a natural model in which the level of MBNLs is decreased. In contrast to the expected effect, our results consistently showed a global increase in circRNA levels in DM1. As a consequence, whole genome transcriptome analysis revealed dozens of circRNAs with significantly altered (mostly increased) levels in DM1. Furthermore, we observed that the circRNA levels were in many cases strongly associated with DM1 severity.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.dmseq.org/>.

ETHICS STATEMENT

The samples, experimental protocols, and methods reported in this study were carried out in accordance with the approval of the local ethics committees: NRESCommittee.EastMidlands-Nottingham2 and the University of Rochester Research Subjects Review Board. Informed consent was obtained from all subjects.

AUTHOR CONTRIBUTIONS

KC—designed assays for circRNA identifications and testing, performed all experimental analyses, interpreted the results, performed statistical analyses, prepared figures, tables, and supplementary materials, and wrote the manuscript;

KT—participated in data interpretation, performed alternative splicing analysis, participated in manuscript preparation, and prepared cDNA samples; APi—participated in data interpretation, performed alternative splicing analysis, and participated in manuscript preparation; KS—provided CL and patient tissue samples and participated in data interpretation and the manuscript preparation; KK—performed computational circRNA identification and analysis; APh—supervised computational analyses and participated in the manuscript preparation; SS—provided patient tissue samples and participated in the manuscript preparation; JDB—provided patient and mouse tissue samples and participated in the manuscript preparation; MW—provided CL and patient tissue samples and participated in data interpretation and the manuscript preparation; PK—conceived and coordinated the study, supervised the experiments and statistical analyses, interpreted the results, and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This manuscript has been released as a preprint at bioRxiv (Czubak et al., 2019).

FUNDING

This work was supported by the National Science Center PL [2016/21/N/NZ5/00508 (to KC), 2014/15/B/NZ2/02453 (to KS), 2014/15/D/NZ2/02305 (to AP), 2017/24/C/NZ1/00112 (to KT), and 2014/13/B/NZ5/03214 (to MW)]; Muscular Dystrophy UK [17GRO-PG12-0149 (to JDB)]; Wellcome Trust [Seeding Drug Discovery grant number 107562/Z/15/Z; 2015 (to JDB)]; and Myotonic Dystrophy Support Group (to JDB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00649/full#supplementary-material>

REFERENCES

- Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66. doi: 10.1016/j.molcel.2014.08.019
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68, 799–808. doi: 10.1016/0092-8674(92)90154-5
- Chomczynski, P., and Sacchi, N. (1987). Single-step method of Rna isolation by acid guanidinium thiocyanate phenol chloroform extraction. *Anal. Biochem.* 162, 156–159. doi: 10.1016/0003-2697(87)90021-2
- Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., et al. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134. doi: 10.1016/j.cell.2015.02.014
- Czubak, K., Taylor, K., Piasecka, A., Sobczak, K., Kozłowska, K., Philips, A. et al., (2019). Global increase in circRNA levels in myotonic dystrophy. bioRxiv (Preprint) doi: 10.1101/489070
- Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., et al. (2010). Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.* 17, 187–193. doi: 10.1038/nsmb.1720
- Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshire, M. G., Harper, P. S., et al. (2002). Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* 11, 805–814. doi: 10.1093/hmg/11.7.805
- Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255, 1256–1258. doi: 10.1126/science.1546326
- Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., et al. (2011). Misregulated alternative splicing of BIN1 is associated with T

- tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med.* 17, 720–725. doi: 10.1038/nm.2374
- Gao, Y., Zhang, J., and Zhao, F. (2018). Circular RNA identification based on multiple seed matching. *Brief. Bioinformatics* 19, 803–810. doi: 10.1093/bib/bbx014
- Glazar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA* 20, 1666–1670. doi: 10.1261/rna.043687.113
- Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010). MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res.* 38, 2467–2484. doi: 10.1093/nar/gkp1209
- Greco, S., Cardinali, B., Falcone, G., and Martelli, F. (2018). Circular RNAs in muscle function and disease. *Int. J. Mol. Sci.* 19, 3454. doi: 10.3390/ijms19113454
- Hansen, T. B., Wiklund, E. D., Bramsen, J. B., Villadsen, S. B., Statham, A. L., Clark, S. J., et al. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* 30, 4414–4422. doi: 10.1038/emboj.2011.359
- Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83, 8604–8610. doi: 10.1021/ac202028g
- Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004). Muscleblind proteins regulate alternative splicing. *EMBO J.* 23, 3103–3112. doi: 10.1038/sj.emboj.7600300
- Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157. doi: 10.1261/rna.035667.112
- Ketley, A., Chen, C. Z., Li, X., Arya, S., Robinson, T. E., Granados-Riveron, J., et al. (2014). High-content screening identifies small molecules that remove nuclear foci, affect MBNL distribution and CELF1 protein levels via a PKC-independent pathway in myotonic dystrophy cell lines. *Hum. Mol. Genet.* 23, 1551–1562. doi: 10.1093/hmg/ddt542
- Khan, M. A. F., Reckman, Y. J., Auferio, S., van den Hoogenhof, M. M. G., 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
- Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Koniczny, P., Stepniak-Koniczna, E., and Sobczak, K. (2014). MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* 42, 10873–10887. doi: 10.1093/nar/gku767
- Koniczny, P., Stepniak-Koniczna, E., Taylor, K., Sznajder, L. J., and Sobczak, K. (2017). Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1 transcript. *Nucleic Acids Res.* 45, 1760–1775. doi: 10.1093/nar/gkw1158
- Kramer, M. C., Liang, D., Tatomer, D. C., Gold, B., March, Z. M., Cherry, S., et al. (2015). Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* 29, 2168–2182. doi: 10.1101/gad.270421.115
- Kuyumcu-Martinez, N. M., Wang, G. S., and Cooper, T. A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* 28, 68–78. doi: 10.1016/j.molcel.2007.07.027
- Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Standler, O., et al. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 6622–6637, e29. doi: 10.1016/j.molcel.2017.02.017
- Li, H., Wei, X., Yang, J., Dong, D., Hao, D., Huang, Y., et al. (2018a). circFGFR4 promotes differentiation of myoblasts via binding miR-107 to relieve its inhibition of Wnt3a. *Mol. Ther. Nucleic Acids* 11, 272–283. doi: 10.1016/j.omtn.2018.02.012
- Li, H., Yang, J., Wei, X., Song, C., Dong, D., Huang, Y., et al. (2018b). CircFUT10 reduces proliferation and facilitates differentiation of myoblasts by sponging miR-133a. *J. Cell. Physiol.* 233, 4643–4651. doi: 10.1002/jcp.26230
- Li, X. F., and Lytton, J. (1999). A circularized sodium-calcium exchanger exon 2 transcript. *J. Biol. Chem.* 274, 8153–8160. doi: 10.1074/jbc.274.12.8153
- Liang, D., Tatomer, D. C., Luo, Z., Wu, H., Yang, L., Chen, L. L., et al. (2017). The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting. *Mol. Cell* 68,940–954 e943. doi: 10.1016/j.molcel.2017.10.034
- Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., et al. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293, 864–867. doi: 10.1126/science.1062125
- Maass, P. G., Glazar, P., Memczak, S., Dittmar, G., Hollfinger, I., Schreyer, L., et al. (2017). A map of human circular RNAs in clinically relevant tissues. *J. Mol. Med. (Berl)* 95, 1179–1189. doi: 10.1007/s00109-017-1582-9
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255, 1253–1255. doi: 10.1126/science.1546325
- Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., et al. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* 289, 1769–1773. doi: 10.1126/science.289.5485.1769
- Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., et al. (2002). Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell* 10, 35–44. doi: 10.1016/S1097-2765(02)00563-4
- 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
- Miller, J. W., Urbinati, C. R., Teng-Ummuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., et al. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* 19, 4439–4448. doi: 10.1093/emboj/19.17.4439
- Miotke, L., Lau, B. T., Rumma, R. T., and Ji, H. P. (2014). High sensitivity detection and quantitation of DNA copy number and single nucleotide variants with single color droplet digital PCR. *Anal. Chem.* 86, 2618–2624. doi: 10.1021/ac403843j
- Nakamori, M., Sobczak, K., Puwanant, A., Welle, S., Eichinger, K., Pandya, S., et al. (2013). Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol.* 74, 862–872. doi: 10.1002/ana.23992
- Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006). The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* 74, 65–80. doi: 10.1111/j.1432-0436.2006.00060.x
- Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* 280, 737–741. doi: 10.1126/science.280.5364.737
- Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885. doi: 10.1016/j.molcel.2015.03.027
- Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., and Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genet.* 9, e1003777. doi: 10.1371/journal.pgen.1003777
- Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7, e30733. doi: 10.1371/journal.pone.0030733
- Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* 29, 40–47. doi: 10.1038/ng704
- Sobczak, K., Wheeler, T. M., Wang, W., and Thornton, C. A. (2013). RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy. *Mol. Ther.* 21, 380–387. doi: 10.1038/mt.2012.222
- Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. H., et al. (2015). Exon circularization requires canonical splice signals. *Cell Rep.* 10, 103–111. doi: 10.1016/j.celrep.2014.12.002
- Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* 128, 995–1002. doi: 10.1083/jcb.128.6.995
- Voellenkle, C., Perfetti, A., Carrara, M., Fuschi, P., Renna, L. V., Longo, M., et al. (2019). Dysregulation of circular RNAs in Myotonic Dystrophy Type 1. *Int. J. Mol. Sci.* 20, 1938. doi: 10.3390/ijms20081938
- Wang, E. T., Treacy, D., Eichinger, K., Struck, A., Estabrook, J., Olafson, H., et al. (2018). Transcriptome alterations in myotonic dystrophy skeletal muscle and heart. *Hum. Mol. Genet.* 28, 1312–1321. doi: 10.1093/hmg/ddy432

- Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G., and Cooper, T. A. (2007). Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J. Clin. Invest.* 117, 2802–2811. doi: 10.1172/JCI32308
- Wang, G. S., Kuyumcu-Martinez, M. N., Sarma, S., Mathur, N., Wehrens, X. H., and Cooper, T. A. (2009). PKC inhibition ameliorates the cardiac phenotype in a mouse model of myotonic dystrophy type 1. *J. Clin. Invest.* 119, 3797–3806. doi: 10.1172/JCI37976
- Wang, Y. H., Li, M. L., Wang, Y. H., Liu, J., Zhang, M. L., Fang, X. T., et al. (2019). A Zfp609 circular RNA regulates myoblast differentiation by sponging miR-194-5p. *Int. J. Biol. Macromol.* 121, 1308–1313. doi: 10.1016/j.ijbiomac.2018.09.039
- Wilusz, J. E., and Sharp, P. A. (2013). Molecular biology. A circuitous route to noncoding RNA. *Science* 340, 440–441. doi: 10.1126/science.1238522
- Wojciechowska, M., Sobczak, K., Kozłowski, P., Sedehizadeh, S., Wojtkowiak-Szlachcic, A., Czubak, K., et al. (2018). Quantitative methods to monitor RNA biomarkers in Myotonic Dystrophy. *Sci. Rep.* 8, 5885. doi: 10.1038/s41598-018-24156-x
- Wojciechowska, M., Taylor, K., Sobczak, K., Napierala, M., and Krzyzosiak, W. J. (2014). Small molecule kinase inhibitors alleviate different molecular features of myotonic dystrophy type 1. *RNA Biol.* 11, 742–754. doi: 10.4161/rna.28799
- You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., et al. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* 18, 603–610. doi: 10.1038/nn.3975
- Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X. H., Chen, L. L., and Yang, L. (2014). Complementary sequence-mediated Exon circularization. *Cell* 159, 134–147. doi: 10.1016/j.cell.2014.09.001
- Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J. L., et al. (2016). The Biogenesis of nascent circular RNAs. *Cell Rep.* 15, 611–624. doi: 10.1016/j.celrep.2016.03.058
- Zheng, Q. P., Bao, C. Y., Guo, W. J., Li, S. Y., Chen, J., Chen, B., et al. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 7, 11215. doi: 10.1038/ncomms11215

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 Czubak, Taylor, Piasecka, Sobczak, Kozłowska, Philips, Sedehizadeh, Brook, Wojciechowska and Kozłowski. 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.



Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing

Megan Stevens* and Sebastian Oltean*

Institute of Biomedical and Clinical Science, Medical School, College of Medicine and Health, University of Exeter, Exeter, United Kingdom

OPEN ACCESS

Edited by:

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

Reviewed by:

Claudio Sette,
University of Rome Tor Vergata,
Italy

Benoit Chabot,

Université de Sherbrooke,
Canada

Lucie Grodecká,
Center of Cardiovascular and
Transplant Surgery,
Czechia

*Correspondence:

Megan Stevens
m.stevens2@exeter.ac.uk
Sebastian Oltean
s.oltean@exeter.ac.uk

Specialty section:

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

Received: 17 June 2019

Accepted: 31 July 2019

Published: 06 September 2019

Citation:

Stevens M and Oltean S (2019)
Modulation of the Apoptosis Gene
Bcl-x Function Through
Alternative Splicing.
Front. Genet. 10:804.
doi: 10.3389/fgene.2019.00804

Apoptosis plays a vital role in cell homeostasis during development and disease. Bcl-x, a member of the Bcl-2 family of proteins, is a mitochondrial transmembrane protein that functions to regulate the intrinsic apoptosis pathway. An alternative splicing (AS) event in exon 2 of Bcl-x results in two isoforms of Bcl-x with antagonistic effects on cell survival: Bcl-xL (long isoform), which is anti-apoptotic, and Bcl-xS (short isoform), which is pro-apoptotic. Bcl-xL is the most abundant Bcl-x protein and functions to inhibit apoptosis by a number of different mechanisms including inhibition of Bax. In contrast, Bcl-xS can directly bind to and inhibit the anti-apoptotic Bcl-xL and Bcl-2 proteins, resulting in the release of the pro-apoptotic Bak. There are multiple splice factors and signaling pathways that influence the Bcl-xL/Bcl-xS splicing ratio, including serine/arginine-rich (SR) proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs), transcription factors, and cytokines. Dysregulation of the AS of Bcl-x has been implicated in cancer and diabetes. In cancer, the upregulation of Bcl-xL expression in tumor cells can result in resistance to chemotherapeutic agents. On the other hand, dysregulation of Bcl-x AS to promote Bcl-xS expression has been shown to be detrimental to pancreatic β -cells in diabetes, resulting in β -cell apoptosis. Therefore, manipulation of the splice factor, transcription factor, and signaling pathways that modulate this splicing event is fast emerging as a therapeutic avenue in the treatment of cancer and diabetes.

Keywords: alternative splicing, apoptosis, Bcl-x, RNA-binding proteins, isoform

INTRODUCTION

Programmed cell death, known as apoptosis, plays a role in cell homeostasis in development and disease. The complex mechanism of apoptosis involves distinct regulatory pathways: the death receptor-mediated (extrinsic) pathway and the mitochondrial (intrinsic) pathway. The extrinsic pathway initiates apoptosis through a death ligand binding to a death receptor, such as tumor necrosis factor (TNF)- α binding to TNF receptor 1 (TNFR1). This results in the recruitment of several death domains, leading to activation of the apoptosis proteins caspase-8 and caspase-10. It is the intrinsic pathway that is regulated by the Bcl-2 family of proteins. The intrinsic pathway is activated by internal stimuli such as DNA damage, oxidative stress, or hypoxia, resulting in a loss of mitochondrial outer membrane (MOM) integrity and release of cytochrome c into the cytoplasm, which forms a complex with Apaf-1 and caspase-9 to form the

apoptosome. The apoptosome goes on to activate caspase-3, which then activates cytoplasmic endonucleases (CAD/ICAD) and proteases, leading to degradation of chromosomal DNA, chromatin condensation, cytoskeletal reorganization, and cell disintegration (Elmore, 2007).

Bcl proteins can be either pro- or anti-apoptotic depending on the Bcl-2 homology (BH) domains present (Danial, 2007). Bcl-x is a member of the Bcl-2 family of proteins with multiple BH3 domains. It is a transmembrane protein that lies within the mitochondria and regulates mitochondrial outer membrane permeabilization (MOMP) and release of cytochrome c into the cytoplasm in response to different stimuli (Shamas-Din et al., 2013).

An alternative splicing (AS) event in exon 2 of Bcl-x results in two isoforms of Bcl-x with antagonistic effects on cell survival: Bcl-xL (long isoform), which is anti-apoptotic, and Bcl-xS (short isoform), which is pro-apoptotic (**Figure 1**). This review will focus on how the Bcl-x AS event is regulated in health and disease, as well as discussing how manipulation of Bcl-x splicing could be a potential therapeutic avenue in disease.

BCL-X SPLICE ISOFORMS

Mechanism of AS

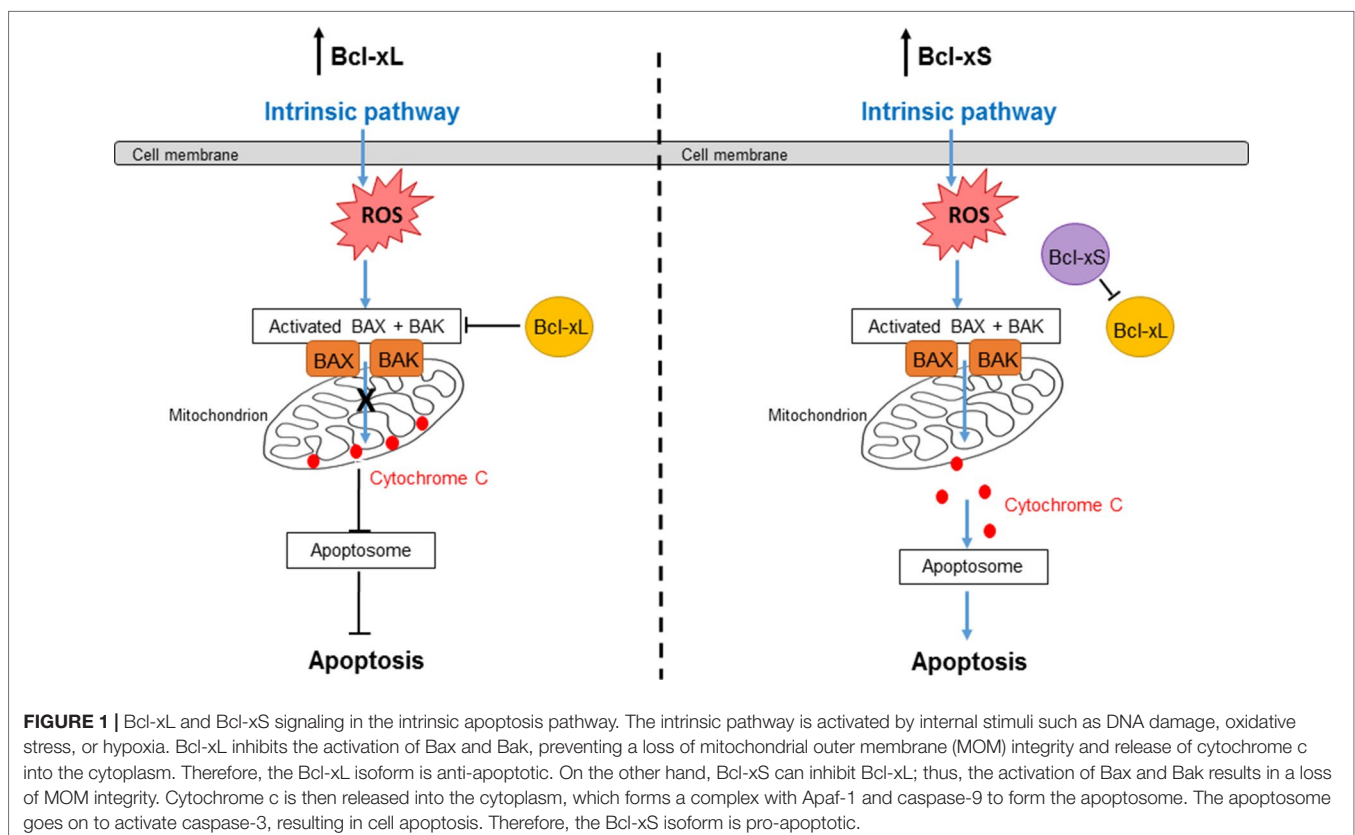
AS is a key process in genetic diversity through which a single pre-mRNA transcript can give rise to multiple protein isoforms; therefore, AS increases the coding capacity of a gene (Matlin

et al., 2005). AS is a highly regulated process; dysregulation can result in cellular dysfunction and disease, including cancer (Oltean and Bates, 2014), diabetes (Stevens and Oltean, 2016; Juan-Mateu et al., 2016), and cardiomyopathy (Guo et al., 2012).

The splicing reaction generally involves the removal of introns from the pre-mRNA and the joining of exons, which is carried out by a macromolecular complex of small nuclear ribonucleoproteins and accessory proteins, known as the spliceosome (Will and Luhrmann, 2011). The spliceosome assembles on splice sites within the pre-mRNA transcript, and the splicing reaction occurs through two transesterification reactions: generation of the branch point and splicing at the 3' or 5' splice site (Shi, 2017).

AS is a process where whole exons or parts of exons/introns are included/excluded in the final mRNA transcript. There are four main types of AS event: 1) cassette exon (whole-exon skipping or retention), 2) intron retention (intron remains), 3) alternative 3' or 5' splice site (different splice sites within an exon), and 4) mutually exclusive exons (two exons alternate inclusion/exclusion). In some instances, AS can alter the protein that is encoded, which can have an effect on function.

AS is regulated by cis- and trans-acting elements. Cis-acting elements can be divided into four subgroups: exonic and intronic splicing enhancers and exonic and intronic splicing silencers. Cis-acting elements recruit trans-acting splice factors to the splice site to either facilitate or suppress the splicing reaction (Matlin et al., 2005). Splice factors are RNA-binding proteins and include



serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Fu and Ares, 2014). Similar to transcription factors, splice factors are also integral parts of cellular signaling pathways. Modification of splice factor activity and availability through intracellular and extracellular signals results in changes in AS, thus changes in the protein repertoire and cell function.

Transcription factors, termed trans-acting factors, are sequence-specific DNA-binding proteins that bind to response elements. Transcription factors can regulate pre-mRNA splicing through three key mechanisms: 1) influencing transcription elongation rates, 2) binding to pre-mRNA to recruit splice factors, and 3) blocking the association of splice factors with the pre-mRNA (reviewed in Rambout et al., 2018).

AS of Bcl-x

Bcl-x represents an example of an apoptotic protein whose function is tightly regulated by AS. The Bcl-x gene consists of three exons. Within exon 2, alternative usage of two 5' splice sites yields two splice variants of Bcl-x, which have antagonistic effects on cell survival. If the proximal 5' splice site is selected in exon 2, the long isoform (Bcl-xL) is expressed, which has an anti-apoptotic function. On the other hand, if the distal 5' splice site is selected, the short isoform (Bcl-xS) is expressed, which promotes cell death (Boise et al., 1993) (**Figure 1**).

Regarding the protein, Bcl-xL is a 233-amino acid protein containing four BH domains, a loop between BH3 and BH4, and a transmembrane region. On the other hand, Bcl-xS lacks an internal 63-amino acid segment that contains the conserved BH1 and BH2 domains; therefore, it only contains the BH3 and BH4 domains. The BH1 and BH2 domains are essential for the interaction of Bcl-xL with death agonists, thus suppressing their activity.

Bcl-xL is the most abundant Bcl-x protein and functions to inhibit apoptosis by a number of different mechanisms. It

can directly inhibit Bax through binding to it and preventing it from binding to the MOM due to the presence of the BH1 and BH2 domains, induce the translocation of MOM bound Bax to the cytoplasm, and sequester tBid, which is an activator of Bax (Billen et al., 2008; Edlich et al., 2011). As a consequence, Bcl-xL prevents apoptosis through inhibition of MOM (Shamas-Din et al., 2013). Overexpression of Bcl-xL has been reported to be correlated with increased cell and tissue survival (Yip and Reed, 2008), including pancreatic islet β -cells (Federici et al., 2001; Carrington et al., 2009). In addition, increased expression of Bcl-xL promotes the progression of breast and urothelial cancer (Espana et al., 2004; Yoshimine et al., 2013) and plays a role in chemotherapy resistance (Amundson et al., 2000).

Regulation of the 5' splice site selected in Bcl-x exon 2 is a critical factor in determining whether a cell is susceptible or resistant to apoptosis. Bcl-xS is the pro-apoptotic isoform of Bcl-x. Bcl-xS can directly bind to and inhibit the anti-apoptotic Bcl-xL and Bcl-2 proteins by forming heterodimers, resulting in the release of the pro-apoptotic Bak (Lindenboim et al., 2001; Plotz et al., 2012). Cell culture studies have shown that increasing the Bcl-xS isoform relative to Bcl-xL can induce apoptosis in cancer cells and pancreatic β -cells (Mercatante et al., 2001; Barbour et al., 2015).

REGULATION OF BCL-X SPLICE SITE SELECTION

There are multiple splice factors and signaling pathways that influence the Bcl-xL/Bcl-xS splicing ratio (**Figure 2**). SR proteins reported to be implicated in the homeostatic regulation of Bcl-x splicing include SRSF1 (Paronetto et al., 2007; Cloutier et al., 2008), SRSF2 (Merdzhanova et al., 2008), SRSF3 (Bielli et al., 2014a), SRSF7 (Bielli et al., 2014a), SRSF9

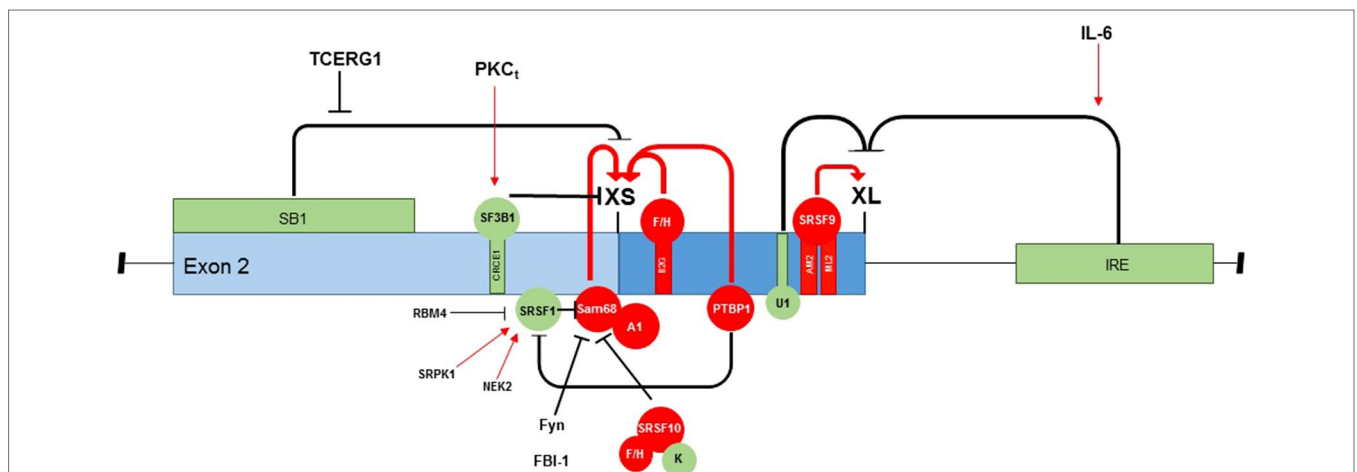


FIGURE 2 | Regulation of Bcl-x exon 2 splice site selection. There are multiple splice factors and signaling pathways that influence 5' splice site selection in exon 2 of the Bcl-x pre-mRNA, which function to alter the Bcl-xL/Bcl-xS splicing ratio. This schematic summarizes the RNA-binding proteins that can both promote splice site selection (red) and inhibit splice site selection (green), pre-mRNA-binding motifs within the exon, and known signaling pathways involved in splice site selection.

(Cloutier et al., 2008), and SRSF10 (Shkreta et al., 2016), as well as the following hnRNPs: A1 (Paronetto et al., 2007), PTBP1 (Bielli et al., 2014a), K (Revil et al., 2009), and F/H (Garneau et al., 2005; Dominguez et al., 2010). RNA-binding proteins include Sam68 (Paronetto et al., 2007), SF3B1 (Massiello et al., 2006), RBM4 (Wang et al., 2014), RBM11 (Pedrotti et al., 2012), RBM25 (Zhou et al., 2008), RBM10 (Inoue et al., 2014), and TRA2 β (Bielli et al., 2014a), in addition to the transcription factors TCERG1 and FBI-1 (Montes et al., 2012; Bielli et al., 2014b).

RNA-Binding Proteins Implicated in Bcl-x AS

The RNA-binding protein Sam68 complexed with hnRNP A1 can bind to the Bcl-x pre-mRNA to promote selection of the distal 5' splice site and the production of Bcl-xS (Paronetto et al., 2007). This interaction is modulated by the Fyn kinase, which is normally activated through protein kinase C (PKC) signaling (Hsu et al., 2009). Although this collaboration between Sam68 and hnRNP A1 is not thought to make a substantial contribution to Bcl-x splicing under normal growth conditions, where very little Bcl-xS is produced, they have been reported to play a critical role when cells are subjected to DNA damage by treatment with oxaliplatin (Cloutier et al., 2018). Oxaliplatin is proposed to promote tyrosine dephosphorylation on Sam68 by counteracting Fyn kinase activity; therefore, dephosphorylated Sam68 can more readily interact with hnRNP A1 to activate the distal 5' splice site in exon 2 of Bcl-x (Paronetto et al., 2007). Furthermore, depletion of hnRNP A1 or mutations that impair its interaction with Sam68 attenuated Bcl-xS splice isoform production (Paronetto et al., 2007). In addition to Bcl-x, Sam68 has been also proposed to regulate splice site selection in another apoptosis gene, *BIRC5*, modulating the expression of the anti-apoptotic DEX3 isoform (Gayvan-Cervantes et al., 2017).

Studies have also shown that the interaction of Sam68 with the Bcl-x pre-mRNA is also modulated by SRSF1 and SRSF10. hnRNP A1 is known to compete with SRSF1 to cause switches in 5' splice sites (Eperon et al., 2000). SRSF1 is suggested to decrease the use of the distal 5' splice site; in cells treated with SRSF1, the proximal splice site was found to be used exclusively, resulting in the expression of only Bcl-xL (Paronetto et al., 2007). Furthermore, SRSF1 activity is modulated *via* phosphorylation by the kinases NEK2 and SRPK1, both of which have been reported to contribute to apoptosis resistance through the expression of Bcl-xL (Naro et al., 2014). On the other hand, SRSF1 itself is antagonized by the splice factors PTBP1 and RBM4. PTBP1 is reported to bind to a polypyrimidine tract located between the two 5' splice sites in exon 2; upon binding of PTBP1 to this site, the distal splice site is favored, and Bcl-xS is transcribed (Bielli et al., 2014a). Mechanistically, PTBP1 was reported to displace the binding of SRSF1 to the proximal splice site, therefore repressing the expression of Bcl-xL (Bielli et al., 2014a). RBM4 competes with SRSF1 to bind to the same regulatory element in the Bcl-x pre-mRNA, promoting the expression of Bcl-xS (Wang et al., 2014). SRSF1 is implicated in splice site regulation in many

genes, including several other genes in the apoptosis pathway. An example is that of *BIM* and *BIN1*; SRSF1 overexpression has been shown to promote the AS of anti-apoptotic isoforms, thus promoting cell survival (Anczukow et al., 2012).

SRSF10 has been reported to collaborate with hnRNP A1/A2 and Sam68 to drive the DNA damage-induced increase in Bcl-xS (Cloutier et al., 2018). SRSF10 interacts with the repressor hnRNP K and stimulatory hnRNP F/H proteins in normally growing cells, resulting in repression of the Bcl-xS splice site; however, upon DNA damage, SRSF10 becomes dephosphorylated, and its interaction with hnRNP F/H is decreased, thus allowing the stimulatory hnRNP F/H to bind to G-rich regulatory elements located downstream of the Bcl-xS splice site, activating its expression (Shkreta et al., 2016). SRSF2, which is upregulated by the transcription factor E2F1, has also been reported to bind to the same G-rich regulatory elements located downstream of the Bcl-xS splice site, resulting in increased cell apoptosis (Merdzhanova et al., 2008). hnRNP K has been found to bind to CX-rich sequences in a silencer element located upstream of the Bcl-xS 5' splice site, which results in the repression of Bcl-xS (Revil et al., 2009).

SRSF9 is a splice factor reported to be involved in upregulating the anti-apoptotic splice variant, Bcl-xL. SRSF9 binds to two elements (ML2 and AM2) within the B3 region of the Bcl-x pre-mRNA located immediately upstream of the Bcl-xL donor site, resulting in a shift in splicing to the Bcl-xL 5' splice site (Michelle et al., 2012). In addition, the B3 region also contains an element that represses Bcl-xL splice site selection, which is bound by U1 snRNP; however, SRSF9 appears to counteract the repressive activity of upstream U1 snRNP-binding sites (Michelle et al., 2012).

In addition to RNA-binding proteins, it has been previously shown that two G-quadruplexes (G4s) form in the Bcl-x pre-mRNA, each of which is close to the two alternative 5' splice sites (Weldon et al., 2017). Furthermore, G4 ligands have been shown to affect Bcl-x splicing, which act independently at the two splice sites depending on their structure (Weldon et al., 2017).

Components of the exon junction complex (EJC), which is deposited on the mRNA concomitantly with splicing to coordinate mRNA export and surveillance, including eIF4A3, Y14, RNPS1, SAP18, and Acinus, have been reported to regulate Bcl-x splicing, with their knock-down shown to encourage production of a Bcl-xS variant (Michelle et al., 2012). Indeed, Bcl-x was the first mammalian gene for which a role of EJC components in splicing was demonstrated. In addition, depletion of these components of the EJC was also shown to effect the splicing of other apoptosis genes, including *Bim* and *Mc11*, inducing the synthesis of pro-apoptotic splice variants (Michelle et al., 2012).

Transcription Factors Implicated in Bcl-x AS

TCERG1 is a human nuclear factor implicated in transcriptional elongation and pre-mRNA splicing. TCERG1 has been reported to regulate the splicing of Bcl-x in a promoter-dependent manner; it promotes the splicing of Bcl-xS through the SB1

regulatory element within the first part of exon 2 (Montes et al., 2012). The proposed mechanism for this regulation is that TCERG1 modulates the elongation rate of RNA polymerase II to relieve pausing of the putative polymerase pause site, thus activating the pro-apoptotic Bcl-xS splice site (Montes et al., 2012). In concordance, TCERG1 has been proposed to sensitize cells to apoptosis through changes in mitochondrial membrane permeabilization (Montes et al., 2015). Interestingly, TCERG1 has also been reported to regulate splicing of the apoptosis gene *Fas/CD95*, promoting the expression of pro-apoptotic Fas (Montes et al., 2015).

FBI-1 is a BTB/POZ-domain Krüppel-like zinc-finger transcription factor. It has recently been reported to play a direct role in the regulation of AS through its interaction with Sam68, reducing its binding to the Bcl-x pre-mRNA (Bielli et al., 2014b). Like Sam68, FBI-1 is overexpressed in human cancers (Aggarwal et al., 2010; Bielli et al., 2011). Through its interaction with Sam68, FBI-1 promotes splicing of the anti-apoptotic Bcl-xL isoform, thus increasing cell survival (Bielli et al., 2014b).

Signaling Pathways Implicated in Bcl-x AS

The phosphoinositide 3-kinase (PI3K) pathway has been proposed as a key survival pathway regulating the alternative 5' splice site selection of the Bcl-x pre-mRNA, increasing Bcl-xL expression in non-small cell lung cancer (NSCLC) cells (Shultz et al., 2012). Protein kinase C_α (PKC_α), atypical PKC, is downstream of PI3K and has been implicated in regulating this AS mechanism and the expression of the splice factor SF3B1, which is an RNA trans-factor that interacts with CREC1 to regulate the 5' splice site selection of the Bcl-x pre-mRNA (Massiello et al., 2006; Shultz et al., 2012). On the other hand, the classical PKC pathway has been implicated in Bcl-x AS in non-transformed cells (HEK293 cells), where PKC inhibitors were shown to increase the expression of Bcl-xS; such changes in the Bcl-x splicing ratio were not observed when cancer cells were treated with PKC inhibitors (Revil et al., 2009).

Interleukin 6 (IL-6) acts as both a pro-inflammatory cytokine and an anti-inflammatory myokine. Treatment of K562 leukemia cells with IL-6 resulted in a reduction in the Bcl-xL/Bcl-xS ratio; nucleotides 1–176 of the downstream intron were found to be required for the IL-6 effect (Li et al., 2004). It is likely that IL-6 has specific downstream targets that directly regulate Bcl-x splicing; however, the exact mechanism is yet to be elucidated.

Ceramide is an important regulator of cell stress responses and growth mechanisms. A family of ceramide-regulated enzymes known as ceramide-activated protein phosphatases includes the serine/threonine-specific protein phosphatase PP1; endogenous ceramide has been reported to modulate the activity of SR proteins in a PP1-dependent manner (Chalfant et al., 2001). Regarding Bcl-x AS, ceramide has been shown to modulate 5' splice site selection, increasing the mRNA expression of Bcl-xS, which correlated to an increased sensitization to chemotherapy (Chalfant et al., 2002). More recently, two ceramide-responsive cis-elements within exon 2 of the Bcl-x pre-mRNA have been identified that function

to regulate 5' splice site selection in response to ceramide (Massiello et al., 2004).

ROLE OF BCL-X SPLICING IN DISEASE

Cancer

Cancer cells often avoid apoptosis through a change in the expression of genes that control apoptosis, including Bcl-x (Fernald and Kurokawa, 2013). Experimentally increased Bcl-xL expression has been observed in several cancer types, and high Bcl-xL expression is correlated with reduced cellular sensitivity to chemotherapeutic agents (Olopade et al., 1997; Takehara et al., 2001; Mercatante et al., 2002). Altered control of the expression of splice factors, resulting in a change in the balance of pro- and anti-apoptotic splice variants, has also been implicated in cancers. For example, SRSF1 has been shown to be increased in breast cancer (Karni et al., 2007); SRSF1 is associated with an increase in the Bcl-xL/Bcl-xS ratio (Paronetto et al., 2007). On the other hand, SRSF2 is reported to be upregulated in lung cancer (Gout et al., 2012), which results in a decrease in the Bcl-xL/Bcl-xS ratio (Merdzhanova et al., 2008). The reasons for this difference are not yet clear. Furthermore, both FBI-1 and Sam68 are overexpressed in human cancers (Aggarwal et al., 2010; Bielli et al., 2011), which results in an upregulation of Bcl-xL and cell survival (Bielli et al., 2014b).

The splice factor hnRNP K represses the expression of Bcl-xS, suggesting an anti-apoptotic mechanism in cancer cells (Revil et al., 2009). Increases in the expression and changes in the cellular distribution of hnRNP K have been demonstrated in many cancer types, indicating it to be a prognostic marker of cancer (Pino et al., 2003; Carpenter et al., 2006; Chen et al., 2008). It has been proposed that hnRNP K interacts with the phosphatase 2A (PP2A) inhibitor SET to promote tumorigenesis through a reduction in Bcl-xS levels (Almeida et al., 2014).

BC200 is a long non-coding RNA (lncRNA) that has been shown to be upregulated in breast cancer. Interestingly, a knockout (KO) of BC200 suppressed tumor cell growth both *in vitro* and *in vivo* through increased expression of the Bcl-xS isoform (Singh et al., 2016). Therefore, BC200 is proposed to play an oncogenic role in breast cancer through binding to the Bcl-x pre-mRNA and recruiting hnRNP A1/B2 (Singh et al., 2016).

The splicing suppressor RBM4 has recently been implicated in tumorigenesis; its expression is significantly decreased in cancer patients, and its level is positively correlated with improved survival (Wang et al., 2014). Mechanistically, RBM4 antagonizes SRSF1 and upregulates the expression of the pro-apoptotic Bcl-xS isoform, thus acting as a tumor suppressor (Wang et al., 2014).

Therefore, in general, an upregulation of the pro-survival Bcl-xL is often observed in cancer cells, which may be, at least in part, due to the dysregulation of certain splice factors involved in Bcl-x pre-mRNA splicing.

Diabetes

Although the evidence thus far for the role of Bcl-x AS in diabetes is limited, it is clear that β -cell apoptosis plays a major

role in the pathogenesis of diabetes, which correlates with the increased expression of the pro-apoptotic Bcl-xS splice isoform. Further research is needed to elucidate the mechanisms by which Bcl-x AS is regulated in diabetes to determine the splice factors and signaling pathway involved.

MANIPULATION OF BCL-X SPLICING AS A POTENTIAL THERAPEUTIC AVENUE

Manipulating the expression of the Bcl-x isoform ratio is emerging as a potential therapeutic avenue in certain disease types. This includes some form of cancers where tumor cells are resistant to chemotherapeutic agents due to the increased expression of anti-apoptotic Bcl-xL and diabetes where β -cells undergo apoptosis correlating to a shift in the splicing ratio to promote the pro-apoptotic Bcl-xS isoform relative to Bcl-xL.

In cancer, one of the key causes of chemoresistance is the resistance of cancer cells to apoptosis (Fulda, 2009). The manipulation of splice factor expression has been shown to sensitize cancer cells to therapeutic treatments, acting either as pro-survival factors that diminish drug-induced apoptosis or as pro-apoptotic factors to potentiate the effects of chemotherapeutics. An example is the anti-apoptotic splice factor SRSF1. Downregulation of SRSF1 in cervical cancer cells with the AURKA kinase inhibitor VX-680 altered the splicing of Bcl-x to increase Bcl-xS, sensitizing the cells to VX-680-induced apoptosis (Moore et al., 2010). Furthermore, silencing SRSF1 in cancer cell lines has been shown to facilitate apoptosis induced by gemcitabine (Adesso et al., 2013). Similarly, hnRNP K has also been reported to interfere with the tumor response to chemotherapeutics. In acute myeloid leukemia cells, a reduction in hnRNP K is required in order for NSC606985, a camptothecin analogue, to trigger cell apoptosis (Go et al., 2009).

Splice-switching oligonucleotides (SSOs) are anti-sense oligonucleotides that hybridize to pre-mRNA sequences, blocking the binding of splice factors, thus redirecting the splicing machinery to an alternative pathway and modifying splicing of the gene. In cancer, a commonly reported therapeutic effect of SSOs is of targeting the Bcl-x pre-mRNA to redirect splicing from Bcl-xL to Bcl-xS, resulting in pro-apoptotic and chemosensitizing effects in various cancer cell lines (Mercatante et al., 2002; Bauman et al., 2009; Bauman et al., 2010; Li et al., 2015). However, the effects of the Bcl-x SSOs appear to vary depending on the expression profile of the target cells, which was suggested to be attributed to the endogenous levels of the Bcl-xL variant; tumor cells with higher endogenous levels of Bcl-xL were reported to be more susceptible to the effects of Bcl-x SSOs, which is likely to be due to the SSO being able to produce enough Bcl-xS to promote apoptosis due to the higher transcription levels of Bcl-x (Mercatante et al., 2001; Mercatante et al., 2002).

Certain chemical classes of G4 ligands, including ellipticine and quindoline derivatives, have been reported to have diverse effects on 5' splice site usage in Bcl-x; for example, the ellipticine GQC-05 antagonizes the Bcl-xL 5' splice site and activates the Bcl-xS 5' splice site, thus inducing cell apoptosis

(Weldon et al., 2017). Such ligands may have the potential to switch Bcl-x splicing in a therapeutic manner.

Islet transplantation is fast becoming a realistic alternative treatment option for patients with a brittle form of type I diabetes (Ricordi and Strom, 2004). However, retaining islet viability is a problem. One study reported that delivering the Bcl-xL variant to the islets *via* protein transduction resulted in an improvement in islet viability, thus preserving islets for transplantation (Klein et al., 2004). In addition, insulin-like growth factor 1 (IGF-1) was shown to be protective against type I diabetes in non-obese diabetic mice as shown by reduced β -cell apoptosis resulting from increased expression of the anti-apoptotic Bcl-xL and Bcl-2 (Chen et al., 2004).

A major problem that comes with the manipulation of AS events as a potential therapeutic option is how to target the splicing event within a particular cell type or tissue. It is of major concern that although reducing the Bcl-xS/Bcl-xL ratio in diabetes would have therapeutic effects regarding β -cell apoptosis, a decrease in the Bcl-xS/Bcl-xL ratio at a whole-organism level may promote tumor cell survival. One potential avenue that could be used to address this issue is using exosomes to deliver SSOs or splicing modulating drugs to target tissues. Exosomes are naturally occurring nanovesicular structures that are secreted by most cell types and are suggested to be the “next-generation” carrier for gene therapy. Although there are few publications to date, Alvarez-Erviti et al. (2011) have successfully delivered siRNA to the mouse brain *via* targeted exosomes.

CONCLUSION

AS of Bcl-x is a tightly regulated event that determines the apoptotic potential of the cell. In general, most cell types predominantly express the anti-apoptotic Bcl-xL isoform. However, a further upregulation of Bcl-xL expression in tumor cells can result in resistance to chemotherapeutic agents. On the other hand, dysregulation of Bcl-x AS to promote Bcl-xS expression has been shown to be detrimental to pancreatic β -cells in diabetes, resulting in β -cell apoptosis. Therefore, manipulation of the splice factor, transcription factor, and signaling pathways that modulate this splicing event is fast emerging as a therapeutic avenue in the treatment of cancer and diabetes; however, further research is required to investigate whether Bcl-x splicing can be modulated in a cell/tissue-specific manner.

AUTHOR CONTRIBUTIONS

MS wrote the manuscript. SO helped with revisions and approved the final version.

FUNDING

Funding for this study was supported by grants to SO: British Heart Foundation (PG/15/53/31371), Diabetes UK (17/0005668)

REFERENCES

- Adesso, L., Calabretta, S., Barbagallo, F., Capurso, G., Pillozzi, E., Geremia, R., et al. (2013). Gemcitabine triggers a pro-survival response in pancreatic cancer cells through activation of the MNK2/eIF4E pathway. *Oncogene*. 32, 2848–2857. doi: 10.1038/onc.2012.306
- Aggarwal, A., Hunter, W. J., 3rd, Aggarwal, H., Silva, E. D., Davey, M. S., Murphy, R. F., et al. (2010). Expression of leukemia/lymphoma-related factor (LRF/POKEMON) in human breast carcinoma and other cancers. *Exp. Mol. Pathol.* 89, 140–148. doi: 10.1016/j.yexmp.2010.05.002
- Almeida, L. O., Garcia, C. B., Matos-Silva, F. A., Curti, C., and Leopoldino, A. M. (2014). Accumulated SET protein up-regulates and interacts with hnRNPK, increasing its binding to nucleic acids, the Bcl-xS repression, and cellular proliferation. *Biochem. Biophys. Res. Commun.* 445, 196–202. doi: 10.1016/j.bbrc.2014.01.175
- Alvarez-Erviti, L., Seow, Y., Yin, H., Betts, C., Lakhal, S., and Wood, M. J. (2011). Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. *Nat. Biotech.* 29, 341–345. doi: 10.1038/nbt.1807
- Amundson, S. A., Myers, T. G., Scudiero, D., Kitada, S., Reed, J. C., and Fornace, A. J., Jr. (2000). An informatics approach identifying markers of chemosensitivity in human cancer cells. *Cancer Res.* 60, 6101–6110.
- Anczukow, 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
- Barbour, S. E., Nguyen, P. T., Park, M., Emani, B., Lei, X., Kambalapalli, M., et al. (2015). Group VIA phospholipase A2 (iPLA2 β) modulates Bcl-x 5'-splice site selection and suppresses anti-apoptotic Bcl-x(L) in β -cells. *J. Biol. Chem.* 290, 11021–11031. doi: 10.1074/jbc.M115.648956
- Bauman, J. A., Jearawiriyapaisarn, N., and Kole, R. (2009). Therapeutic potential of splice-switching oligonucleotides. *Oligonucleotides*. 19, 1–13. doi: 10.1089/oli.2008.0161
- Bauman, J. A., Li, S. D., Yang, A., Huang, L., and Kole, R. (2010). Anti-tumor activity of splice-switching oligonucleotides. *Nucleic Acids Res.* 38, 8348–8356. doi: 10.1093/nar/gkq731
- Bielli, P., Busa, R., Paronetto, M. P., and Sette, C. (2011). The RNA-binding protein Sam68 is a multifunctional player in human cancer. *Endocr. Relat. Cancer* 18, R91–R102. doi: 10.1530/ERC-11-0041
- Bielli, P., Bordini, M., Di Biasio, V., and Sette, C. (2014a). Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5' splice site selection. *Nucleic Acids Res.* 42, 12070–12081. doi: 10.1093/nar/gku922
- Bielli, P., Busa, R., Di Stasi, S. M., Munoz, M. J., Botti, F., Kornblihtt, A. R., et al. (2014b). The transcription factor FBI-1 inhibits SAM68-mediated BCL-X alternative splicing and apoptosis. *EMBO Rep.* 15, 419–427. doi: 10.1002/embr.201338241
- Billen, L. P., Kokoski, C. L., Lovell, J. F., Leber, B., and Andrews, D. W. (2008). BCL-XL inhibits membrane permeabilization by competing with Bax. *PLoS Biol.* 6, e147. doi: 10.1371/journal.pbio.0060147
- Boise, L. H., Gonzalez-Garcia, M., Postema, C. E., Ding, L., Lindsten, T., Turka, L. A., et al. (1993). Bcl-x, a Bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell*. 74, 597–608. doi: 10.1016/0092-8674(93)90508-N
- Butler, A. E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R. A., and Butler, P. C. (2003). Beta-cell deficit and increased beta-cell apoptosis in human with type 2 diabetes. *Diabetes*. 52, 102–110. doi: 10.2337/diabetes.52.1.102
- Carpenter, B., McKay, M., Dundas, S. R., Lawrie, L. C., Telfer, C., and Murray, G. I. (2006). Heterogeneous nuclear ribonucleoprotein K is overexpressed, aberrantly localized and is associated with poor prognosis in colorectal cancer. *Br. J. Cancer*. 95, 921–927. doi: 10.1038/sj.bjc.6603349
- Carrington, E. M., McKenzie, M. D., Jansen, E., Myers, M., Fynch, S., Kos, C., et al. (2009). Islet beta-cells deficient in Bcl-xL develop but are abnormally sensitive to apoptotic stimuli. *Diabetes*. 58, 2316–2323. doi: 10.2337/db08-1602
- Chalfant, C. E., Ogretmen, B., Galadari, S., Kroesen, B. J., Pettus, B. J., and Hannun, Y. A. (2001). FAS activation induces dephosphorylation of SR proteins; dependence on the *de novo* generation of ceramide and activation of protein phosphatase 1. *J. Biol. Chem.* 276, 44848–44855. doi: 10.1074/jbc.M106291200
- Chalfant, C. E., Rathman, K., Pinkerman, R. L., Wood, R. E., Obeid, L. M., Ogretmen, B., et al. (2002). *De novo* ceramide regulates the alternative splicing of caspase 9 and Bcl-x in A549 lung adenocarcinoma cell. Dependence on protein phosphatase-1. *J. Biol. Chem.* 277, 12587–12595. doi: 10.1074/jbc.M112010200
- Chen, W., Salojin, K. V., Mi, Q. S., Grattan, M., Meagher, T. C., Zucker, P., et al. (2004). Insulin-like growth factor (IGF)-I/IGF-binding protein-3 complex: therapeutic efficacy and mechanism of protection against type 1 diabetes. *Endocrinology*. 145, 627–638. doi: 10.1210/en.2003-1274
- Chen, Y., Li, W., and Zhang, S. (2008). hnRNP K expression and its clinical significance in human lung cancer tissues. *Zhongguo Fei. Ai. Za. Zhi*. 11, 241–245. doi: 10.3779/j.issn.1009-3419.2008.02.017
- Cloutier, A., Shkreta, L., Toutant, J., Durand, M., Thibault, P., and Chabot, B. (2018). hnRNP A1/A2 and Sam68 collaborate with SRSF10 to control the alternative splicing response to oxaliplatin-mediated DNA damage. *Sci. Rep.* 8, 2206. doi: 10.1038/s41598-018-20360-x
- Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T., and Chabot, B. (2008). Antagonistic effects of the SRP30c protein and cryptic 5' splice sites on the alternative splicing of the apoptotic regulator Bcl-x. *J. Biol. Chem.* 283, 21315–21324. doi: 10.1074/jbc.M800353200
- Danial, N. N. (2007). BCL-2 family proteins: critical checkpoints of apoptotic cell death. *Clin. Cancer Res.* 13, 7254–7263. doi: 10.1158/1078-0432.CCR-07-1598
- Dominguez, C., Fiset, J. F., Chabot, B., and Allain, F. H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* 17, 853–861. doi: 10.1038/nsmb.1814
- Edlich, F., Banerjee, S., Suzuki, M., Cleland, M. M., Arnoult, D., Wang, C., et al. (2011). Bcl-x(L) retrotranslocates Bax from the mitochondria into the cytosol. *Cell*. 145, 104–116. doi: 10.1016/j.cell.2011.02.034
- Elmore, S. (2007). Apoptosis: a review of programmed cell death. *Toxicol. Pathol.* 35, 495–516. doi: 10.1080/01926230701320337
- Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G., et al. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell Biol.* 20, 8303–8318. doi: 10.1128/MCB.20.22.8303-8318.2000
- Espana, L., Fernandez, Y., Ruvio, N., Torregrosa, A., Blanco, J., and Sierra, A. (2004). Overexpression of Bcl-xL in human breast cancer cells enhances organ-selective lymph node metastasis. *Breast Cancer Res. Treat.* 87, 33–44. doi: 10.1023/B:BREA.0000041579.51902.89
- Federici, M., Hribal, M. L., Ranalli, M., Marselli, L., Porzio, O., Lauro, D., et al. (2001). The common Arg972 polymorphism in insulin receptor substrate-1 causes apoptosis of human pancreatic islets. *FASEB J.* 15, 22–24. doi: 10.1096/fj.00-0414fje
- Fernald, K., and Kurokawa, M. (2013). Evading apoptosis in cancer. *Trends. Cell Biol.* 23, 620–633. doi: 10.1016/j.tcb.2013.07.006
- 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
- Fulda, S. (2009). Tumor resistance to apoptosis. *Int. J. Cancer*. 124, 511–515. doi: 10.1002/ijc.24064
- Go, F. H., Wu, Y. L., Zhao, M., Liu, C. X., Wang, L. S., and Chen, G. Q. (2009). Protein kinase C-delta mediates down-regulation of heterogeneous nuclear ribonucleoprotein K protein: involvement in apoptosis induction. *Exp. Cell Res.* 315, 3250–3258. doi: 10.1016/j.yexcr.2009.09.005
- Garneau, D., Revil, T., Fiset, J. F., and Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* 280, 22641–22650. doi: 10.1074/jbc.M501070200
- Gayvan-Cervantes, J., Gonzalez-Torres, C., Maldonado, V., Zampieri, C., Ceballos-Cancino, G., and Melendez-Zajgla, J. (2017). Protein Sam68 regulates the alternative splicing of surviving DEX3. *J. Biol. Chem.* 291, 13745–13757. doi: 10.1074/jbc.M117.800318
- Gout, S., Brambilla, E., Boudria, A., Drissi, R., Lantuejoul, S., Gazzeri, S., et al. (2012). Abnormal expression of the pre-mRNA splicing regulators SRSF1, SRSF2, SRPK1 and SRPK2 in non small cell lung carcinoma. *PLoS One*. 7, e46539. doi: 10.1371/journal.pone.0046539
- Guo, W., Schafer, S., Greaser, M. L., Radke, M. H., Liss, M., Givindarajan, T., et al. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* 18, 766–773. doi: 10.1038/nm.2693

- Hsu, K. L., Fan, H. J., Chen, Y. C., Huang, Y. S., Chen, C. H., Wu, J. C., et al. (2009). Protein kinase C- γ kinase cascade mediates the oleic acid-induced disassembly of neonatal rat cardiomyocyte adherens junctions. *Int. J. Biochem. Cell Biol.* 41, 1536–1546. doi: 10.1016/j.biocel.2008.12.016
- Inoue, A., Yamamoto, N., Kimura, M., Nishio, K., Yamane, H., and Nakajima, K. (2014). RBM10 regulates alternative splicing. *FEBS Lett.* 588, 942–947. doi: 10.1016/j.febslet.2014.01.052
- Juan-Mateu, J., Villate, O., and Eizirik, D. L. (2016). Mechanisms in endocrinology. Alternative splicing: the new frontier in diabetes research. *Eur. J. Endocrinol.* 174, R225–R238. doi: 10.1530/EJE-15-0916
- 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
- Klein, D., Ribeiro, M. M., Mendoza, V., Jayaraman, S., Kenyon, N. S., Pileggi, A., et al. (2004). Delivery of Bcl-XL or its BH4 domain by protein transduction inhibits apoptosis in human islets. *Biochem. Biophys. Res. Commun.* 323, 473–478. doi: 10.1016/j.bbrc.2004.08.116
- Li, C. Y., Chu, J. Y., Yu, J. K., Huang, X. Q., Liu, X. J., Shi, L., et al. (2004). Regulation of alternative splicing of Bcl-x by IL-6, GM-CSF and TPA. *Cell Res* 14, 473–479. doi: 10.1038/sj.cr.7290250
- Li, Z., Li, Q., Han, L., Tian, N., Liang, Q., Li, Y., et al. (2015). Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines. *Oncol. Rep.* 35(1013), 1019. doi: 10.3892/or.2015.4465
- Lindenboim, L., Borner, C., and Stein, R. (2001). Bcl-x(S) can form homodimers and heterodimers and its apoptotic activity requires localization of Bcl-x(S) to the mitochondria and its BH3 and loop domains. *Cell Death Differ.* 8, 933–942. doi: 10.1038/sj.cdd.4400888
- Luciani, D. S., White, S. A., Windenmaier, S. B., Saran, V. V., Taghizadeh, F., Hu, X., et al. (2013). Bcl-2 and Bcl-xL suppress glucose signaling in pancreatic β -cells. *Diabetes* 62, 170–182. doi: 10.2337/db11-1464
- Massiello, A., Salas, A., Pinkerman, R. L., Roddy, P., Roesser, J. R., (2004). Identification of two RNA cis-elements that function to regulate the 5' splice site selection of Bcl-x pre-mRNA in response to ceramide. *J. Biol. Chem.* 279, 15799–15804. doi: 10.1074/jbc.M313950200
- Massiello, A., Roesser, J. R., and Chalfant, C. E. (2006). SAP155 binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5' splice site selection of Bcl-x pre-mRNA. *FASEB J.* 20, 1680–1682. doi: 10.1096/fj.05-5021fje
- 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
- Mercatante, D. R., Bortner, C. D., Cidlowski, J. A., and Kole, R. (2001). Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. Analysis of apoptosis and cell death. *J. Biol. Chem.* 276, 16411–16417. doi: 10.1074/jbc.M009256200
- Mercatante, D. R., Mohler, J. L., and Kole, R. (2002). Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J. Biol. Chem.* 277, 49374–49382. doi: 10.1074/jbc.M209236200
- Merdzhanova, G., Edmond, V., De Seranno, S., Van den Broeck, A., Corcos, L., Brambilla, C., et al. (2008). E2F1 controls alternative splicing pattern of genes involved in apoptosis through upregulation of the splicing factor SC35. *Cell Death Differ.* 15, 1815–1823. doi: 10.1038/cdd.2008.135
- Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., et al. (2012). Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol. Cell Biol.* 32, 954–967. doi: 10.1128/MCB.06130-11
- Montes, M., Cloutier, A., Sanchez-Hernandez, N., Michelle, L., Lemieux, B., Blanchette, M., et al. (2012). TCERG1 regulates alternative splicing of the Bcl-x gene by modulating the rate of RNA polymerase II transcription. *Mol. Cell Biol.* 32, 751–762. doi: 10.1128/MCB.06255-11
- Montes, M., Coiras, M., Becerra, S., Morneo-Castro, C., Mateos, E., Majuelos, J., et al. (2015). Functional consequences for apoptosis by transcription elongation regulator 1 (TCERG1)-mediated Bcl-x and Fas/CD95 alternative splicing. *PLoS One* 10, e0139812. doi: 10.1371/journal.pone.0139812
- Moore, M. J., Wang, Q., Kennedy, C. J., and Silver, P. A. (2010). An alternative splicing network links cell-cycle control to apoptosis. *Cell* 142, 625–636. doi: 10.1016/j.cell.2010.07.019
- Naro, C., Barbagallo, F., Chieffi, P., Bourgeois, C. F., Paronetto, M. P., and Sette, C. (2014). The centrosomal kinase NEK2 is a novel splicing factor kinase involved in cell survival. *Nucleic Acids Res.* 42, 3218–3227. doi: 10.1093/nar/gkt1307
- Olopade, O. I., Adeyanju, M. O., Safa, A. R., Hagos, F., Mick, R., Thompson, C. B., et al. (1997). Overexpression of BCL-X protein in primary breast cancer is associated with high tumor grade and nodal metastases. *Cancer J. Sco. Am.* 3, 230–237.
- Oltean, S., and Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311–5318. doi: 10.1038/onc.2013.533
- Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E., and Sette, C. (2007). The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* 176, 929–939. doi: 10.1083/jcb.200701005
- Pedrotti, S., Busa, R., Compagnucci, C., and Sette, C. (2012). The RNA recognition motif protein RBM11 is a novel tissue-specific splicing regulator. *Nucleic Acids Res.* 40, 1021–1032. doi: 10.1093/nar/gkr819
- Pino, I., Pio, R., Toledo, G., Zabalegui, N., Vicent, S., Rey, N., et al. (2003). Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer* 41, 131–143. doi: 10.1016/S0169-5002(03)00193-4
- Plotz, M., Hossini, A. M., Gillissen, B., Daniel, P. T., Stockfleth, E., and Eberle, J. (2012). Mutual recognition of Bcl-2 proteins independent of the BH3 domain as shown by the BH3-domain lacking protein Bcl-x(AK). *PLoS One* 7, e34549. doi: 10.1371/journal.pone.0034549
- Rambout, X., Dequiedt, F., and Maquat, L. E. (2018). Beyond transcription: roles of transcription factors in pre-mRNA splicing. *Chem. Rev.* 118, 4339–4364. doi: 10.1021/acs.chemrev.7b00470
- Revil, T., Pelletier, J., Toutant, J., Cloutier, A., and Chabot, B. (2009). Heterogeneous nuclear ribonucleoprotein K represses the production of pro-apoptotic Bcl-xS splice isoform. *J. Biol. Chem.* 284, 21458–21467. doi: 10.1074/jbc.M109.019711
- Ricordi, C., and Strom, T. B. (2004). Clinical islet transplantation: advances and immunological challenges. *Nat. Rev. Immunol.* 4, 259–268. doi: 10.1038/nri1332
- Shamas-Din, A., Kale, J., Leber, B., and Andrews, D. W. (2013). Mechanisms of action of Bcl-2 family proteins. *Cold Spring Harb. Perspect. Biol.* 5, a008714. doi: 10.1101/cshperspect.a008714
- Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. doi: 10.1038/nrm.2017.86
- Shkreta, L., Toutant, J., Durand, M., Manley, J. L., and Chabot, B. (2016). SRSF10 connects DNA damage to the alternative splicing of transcripts encoding apoptosis, cell-cycle control, and DNA repair factors. *Cell Rep.* 17, 1990–2003. doi: 10.1016/j.celrep.2016.10.071
- Shultz, J. C., Vu, N., Schultz, M. D., Mba, M. U., Shapiro, B. A., and Chalfant, C. E. (2012). The proto-oncogene PKC δ regulates the alternative splicing of Bcl-x pre-mRNA. *Mol. Cancer Res.* 10, 660–669. doi: 10.1158/1541-7786.MCR-11-0363
- Singh, R., Gupta, S. C., Peng, W. X., Zhou, N., Pochampally, R., Atfi, A., et al. (2016). Regulation of alternative splicing of Bcl-x by BC200 contributes to breast cancer pathogenesis. *Cell Death Res.* 7, e2262. doi: 10.1038/cddis.2016.168
- Socha, L., Silva, D., Lesage, S., Goodnow, C., and Petrovsky, N. (2003). The role of endoplasmic reticulum stress in nonimmune diabetes: NOD.k iHEL, a novel model of beta cell death. *Ann. N. Y. Acad. Sci.* 1005, 178–183. doi: 10.1196/annals.1288.022
- Stevens, M., and Oltean, S. (2016). Alternative splicing in CKD. *J. Am. Soc. Nephrol.* 27, 1596–1603. doi: 10.1681/ASN.2015080908
- Takehara, T., Liu, X., Fujimoto, J., Friedman, S. L., and Takahashi, H. (2001). Expression and role of Bcl-xL in human hepatocellular carcinomas. *Hepatology* 34, 55–61. doi: 10.1053/jhep.2001.25387
- 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
- Weldon, C., Dacanay, J. G., Gokhale, V., Boddupally, P. V. L., Behm-Ansmant, I., Burley, G.-A., et al. (2017). Specific G-quadruplex ligands modulate the alternative splicing of Bcl-x. *Nucleic Acids Res.* 46, 886–896. doi: 10.1093/nar/gkx1122

- Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707. doi: 10.1101/cshperspect.a003707
- Yip, K. W., and Reed, J. C. (2008). Bcl-2 family proteins and cancer. *Oncogene*. 27, 6398–6406. doi: 10.1038/onc.2008.307
- Yoshimine, S., Kikuchi, E., Kosaka, T., Mikami, S., Miyajima, A., Okada, Y., et al. (2013). Prognostic significance of Bcl-xL expression and efficacy of Bcl-xL targeting therapy in urothelial carcinoma. *Br. J. Cancer*. 108, 2132–2120. doi: 10.1038/bjc.2013.216
- Zhou, A., Ou, A. C., Cho, A., Benz, E. J., Jr., and Huang, S. C. (2008). Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection. *Mol. Cell Biol.* 28, 5924–5936 doi: 10.1128/MCB.00560-08

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 Stevens and Oltean. 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.



Identification of hsa_circ_0001821 as a Novel Diagnostic Biomarker in Gastric Cancer via Comprehensive Circular RNA Profiling

Shan Kong^{1†}, Qian Yang^{1†}, Chenxue Tang¹, Tianyi Wang¹, Xianjuan Shen^{2*} and Shaoqing Ju^{1*}

¹ Department of Laboratory Medicine, Affiliated Hospital of Nantong University, Nantong, China, ² Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China

OPEN ACCESS

Edited by:

Stefano Duga,
Humanitas University, Italy

Reviewed by:

Marco Manfrini,
Maria Cecilia Hospital, Italy
Konrad Huppl,
National Cancer Institute (NCI),
United States

*Correspondence:

Xianjuan Shen
juanxia819@163.com
Shaoqing Ju
jsq814@hotmail.com

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 29 April 2019

Accepted: 21 August 2019

Published: 20 September 2019

Citation:

Kong S, Yang Q, Tang C, Wang T,
Shen X and Ju S (2019) Identification
of hsa_circ_0001821 as a Novel
Diagnostic Biomarker in Gastric
Cancer via Comprehensive
Circular RNA Profiling.
Front. Genet. 10:878.
doi: 10.3389/fgene.2019.00878

Background: The morbidity and mortality of gastric cancer (GC) remain high worldwide. With the advent of the Human Genome Sequencing Project, circular RNAs (circRNAs) have attracted widespread attention in cancer research due to their stable ring structure. Our aim was to identify differentially expressed circRNAs in GC and explore their potential roles in GC diagnosis, treatment, and prognostic prediction.

Methods: Large-scale gene screening was performed in three pairs of GC tissues and adjacent noncancerous tissues using high-throughput sequencing. The expression of hsa_circ_0001821 was detected in 80 pairs of tissue samples by quantitative real-time PCR (qRT-PCR). Stability of the ring structure of hsa_circ_0001821 RNA was verified by exonuclease digestion assay, and its diagnostic value was evaluated by receiver operating characteristic (ROC) analysis. In addition, the location of hsa_circ_0001821 in GC cells was detected by nucleoplasm separation assay.

Results: A total of 25,303 circRNAs were identified, among which 2,007 circRNAs were differentially expressed (fold change > 2.0, $P < 0.05$). Further validation disclosed that hsa_circ_0001821 was significantly downregulated in the 80 pairs of GC tissues and 30 whole-blood specimens obtained from the GC patients. The specificity of hsa_circ_0001821 in GC was higher than that in other solid tumors. In addition, hsa_circ_0001821 was relatively stable after RNA exonuclease digestion. Clinicopathological parameter analysis showed that hsa_circ_0001821 was negatively correlated with tumor depth ($r = -0.255$, $P = 0.022$) and lymph node metastasis ($r = -0.235$, $P = 0.036$). Area under the curve (AUC) analysis showed that the diagnostic efficiency of circulating hsa_circ_0001821 in distinguishing GC patients was higher than that in GC tissues (0.872, 95%CI: 0.767–0.977 vs. 0.792, 95%CI: 0.723–0.861). Combined use of circulating hsa_circ_0001821 with the existing tumor markers yielded the largest AUC of 0.933. Finally, hsa_circ_0001821 was demonstrated to mainly locate in the cytoplasm, implying that it played a potential regulatory role in GC at the posttranscriptional level.

Conclusion: Hsa_circ_0001821 may prove to be a new and promising potential biomarker for GC diagnosis.

Keywords: circular RNA, gastric cancer, high-throughput sequencing, biomarker, diagnosis

INTRODUCTION

Gastric cancer (GC) remains one of the most common malignant tumors worldwide. According to the latest statistics released by the World Health Organization (WHO) Cancer Control Program, over seven million people die of cancer worldwide each year, with about 700,000 of them suffering from GC (Siegel et al., 2019). Meanwhile, approximately 934,000 new cases of GC are diagnosed every year, among which about 43% (400,000) occur in China with morbidity and mortality rates about twofold higher than the world average (Cheng et al., 2016; Miller et al., 2016). Usually, patients with advanced GC may have a 50–70% chance of recurrence after surgery, and their 5-year survival rate is often less than 30% (Sun and Yan, 2016). Currently, the early detection rate of GC is less than 10%, and the disease is usually diagnosed in the advanced stage or when metastasis has already occurred. Therefore, it is of particular significance to screen out specific and sensitive biomarkers and strengthen the research on GC pathogenesis for the sake of improving the diagnosis and treatment of GC.

Using the Human Genome Sequencing Project, scientists have found that the proportion of protein-coding genes in the transcriptome is much lower than that in noncoding RNAs (ncRNAs), and about 80% transcription products are ncRNAs (Prasanth and Spector, 2007). Initially, most ncRNAs were considered to be the “noise” of genome transcription and therefore largely ignored. However, with the reduction of sequencing cost and the emerging of new-generation sequencing technology and in-depth sequencing of complementary DNA (cDNA) pools or libraries, ncRNAs have been identified to act as regulatory factors to control gene expression at multiple cell levels and maintain telomere elongation. Meanwhile, they are viewed as guides of molecular repair with important biological functions in life activities and disease occurrence (Taft et al., 2010). Presently, three specific ncRNAs are widely reported in cancer research, including microRNAs (miRNAs), long noncoding RNAs (lncRNAs) > 200 nt, and newly discovered circular RNAs (circRNAs) (Beermann et al., 2016).

circRNAs are a group of endogenous ncRNA molecules that widely exist in human cells. Current studies have demonstrated that circRNA is produced by special variable shear, and its 3′ and 5′ ends are joined together by covalent bonding to form a closed circular structure. Compared with other types of ncRNAs, circRNA is well tolerable by RNA exonuclease, relatively stable, and not easily degradable, making it a highly variable competitive endogenous RNA (ceRNA) (Lasda and Parker, 2014). A single circRNA molecule contains a large number of miRNA response elements that can bind or release a large number of miRNAs instantaneously, so circRNA acts very efficiently and stably as a ceRNA (Qu et al., 2015). Evidence shows that circRNAs can regulate gene expression at the transcriptional level *via* binding miRNAs as a molecule sponge. On the other hand, circRNAs might bind to RNA-binding proteins or other RNA translation proteins through complementary base pairs, interfering with the normal function of genes at the posttranscriptional level. These findings provide a new direction for the exploration of circRNAs as targets for disease diagnosis and prognostic prediction.

To find differentially expressed circRNAs, we detected circRNA expression in three pairs of GC tissues by high-throughput

sequencing in the present study and identified 2,007 significantly differentially expressed circRNAs *via* circRNA sequencing. Subsequently, we chose hsa_circ_0001821 as our study object to further our investigation in 80 pairs of GC tissues and 30 whole-blood samples from GC patients and evaluate the clinical utility of hsa_circ_0001821 in GC diagnosis by receiver operating characteristic (ROC) analysis in an attempt to provide a novel biomarker for GC research.

RESULTS

Identification of Deregulated circRNAs in GC Tissues

To investigate the expression profiles of circRNAs in GC tissues, we conducted high-throughput sequencing in three GC tissues *vs.* three matched noncancerous tissues and identified a total of 25,303 circRNA targets, including 20,036 known circRNAs and 5,267 undefined circRNAs. The heatmap was depicted as a direct approach to visualize the distributions of the dataset for circRNA profiles (Figure 1A). Volcano plots depicted 2,007 differentially expressed circRNAs in the GC tissues (fold change > 2.0, $P < 0.5$) (Figure 1B), from which 16 significantly different circRNAs were selected ($2.0 < \text{fold change} < 6.0$, $P < 0.05$). The details regarding these circRNAs are presented in Table 1. Knowing that the specific parental gene is a lncRNA closely associated with GC evolution and progression (Xu et al., 2017; Zhao et al., 2018), we finally chose hsa_circ_0001821 as our research target. We first detected hsa_circ_0001821 expression in 20 pairs of GC tissues by quantitative real-time PCR (qRT-PCR) and found that it was significantly downregulated in the GC tissues (Figure 1C).

Methodological Evaluation of hsa_circ_0001821 in GC Cells

According to the human reference genome (GRCh37/hg19) from the Ensembl genome database (<http://www.ensembl.org>), hsa_circ_0001821 is located at chr8_128902834_128903244+, and the length of its mature transcript is 410 bp (Figure 2A). To verify the specificity and accuracy of the amplification procedure, the PCR amplification products were subjected to 2.5% agarose gel electrophoresis. The single electrophoresis bands were consistent with the size of the primer amplification product (Figure 2B). To verify the ring structure of hsa_circ_0001821, we designed polymerized primers and reverse primers for their cyclization sites (Figure 2C). qRT-PCR was then performed using genomic DNA (gDNA) and cDNA as templates and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as the negative control. Agarose electrophoresis assay showed that hsa_circ_0001821 could be amplified from the PCR products using cDNA as the template, while a negative result was observed in the control group using gDNA as the template (Figure 2D). Besides, the reverse shear site of hsa_circ_0001821 was confirmed by Sanger sequencing (Figure 2E). Knowing that circRNA is relatively stable compared with linear RNA and not easily degraded by RNA exonuclease, we performed the RNA exonuclease digestion assay. RNA exonuclease was added to total RNA isolated from SGC-7901

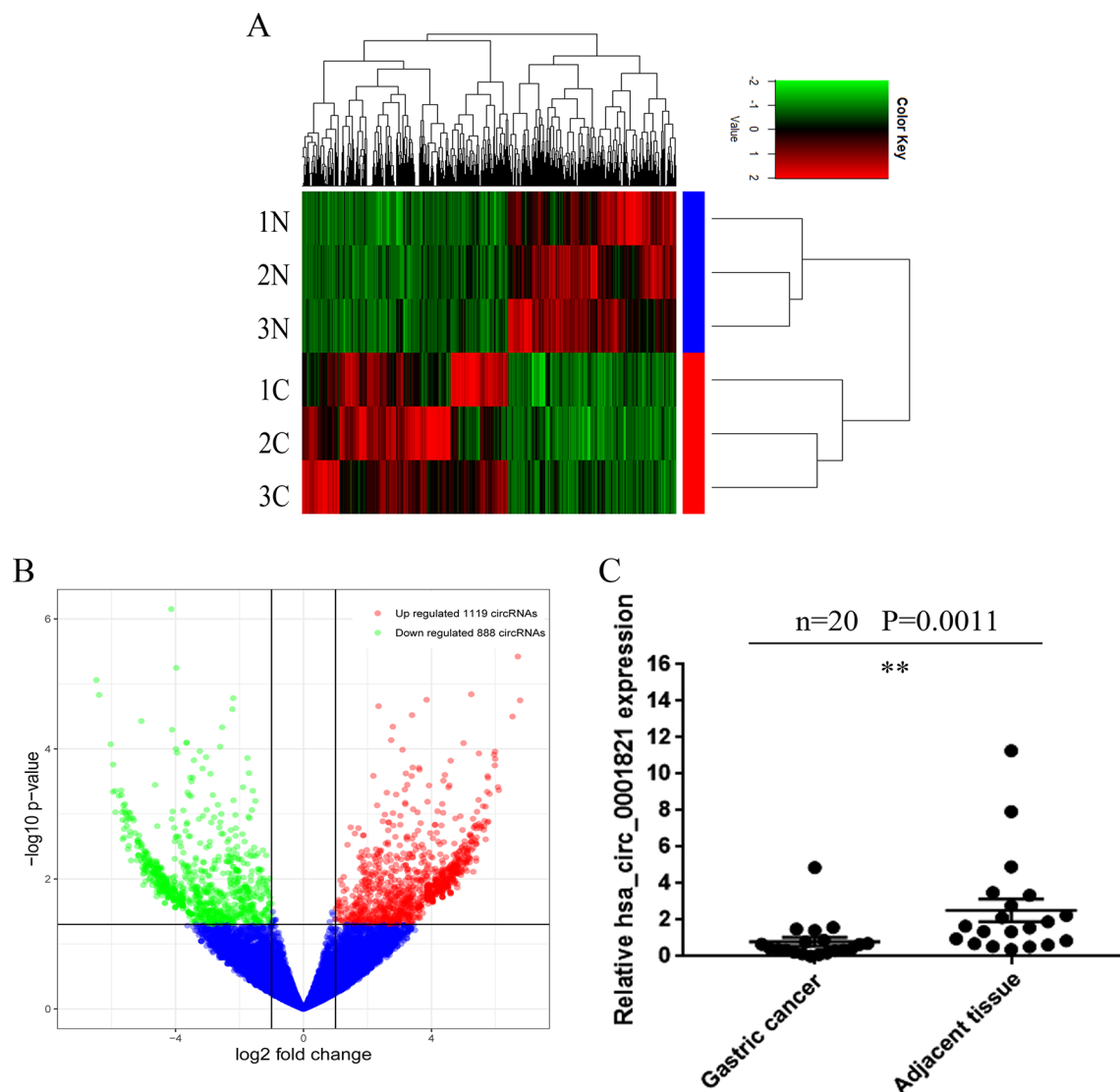


FIGURE 1 | Identification of deregulated circular RNAs (circRNAs) in gastric cancer (GC) tissues. **(A)** Clustered heatmap. Each row represents a tissue sample, and each column represents a circRNA. The color scale reflects the \log_2 signal strength from green (low intensity) to black (medium intensity) to red (strong intensity). **(B)** Volcano plots. The red points in plot indicate the differentially upregulated expression of circRNAs with statistical significance while the green points indicate the downregulated circRNAs. **(C)** Initial verification of hsa_circ_0001821 expression in 20 pairs of GC tissues by quantitative real-time PCR (qRT-PCR). ** $P < 0.01$ were considered significant.

and BGC-823 cells, and the expression of hsa_circ_0001821 and linear PVT1 was detected by qRT-PCR. Compared with that of linear PVT1, hsa_circ_0001821 expression was not significantly reduced after RNA exonuclease treatment, indicating that hsa_circ_0001821 had a relatively stable structure (Figure 2F).

Correlation Analysis of hsa_circ_0001821 Expression and the Clinicopathological Parameters in GC Patients

As shown in Table 2, the expression of hsa_circ_0001821 in GC tissues was significantly correlated with tumor depth ($P = 0.0030$) and lymph node metastasis ($P = 0.0072$). However, we did not find any association between the hsa_circ_0001821 expression

and other clinicopathological parameters, such as gender ($P = 0.8285$), age ($P = 0.1887$), histological differentiation ($P = 0.0696$), tumor size ($P = 0.8900$), CEA ($P = 0.0977$), CA199 ($P = 0.0864$), and CA125 ($P = 0.7259$). Furthermore, the Spearman correlation analysis also indicated that hsa_circ_0001821 expression was negatively correlated with tumor depth ($r = -0.255$, $P = 0.022$) and lymph node metastasis ($r = -0.235$, $P = 0.036$) (Table 3).

Validation of hsa_circ_0001821 Expression in Different Tumor Tissues

For large-sample verification, 60 pairs of GC tissues were collected. The result of qRT-PCR showed that the expression of hsa_circ_0001821 in GC tissues was significantly lower than that

TABLE 1 | A total of 16 significantly differentially expressed circRNAs identified via circRNA sequencing.

circRNA	Gene	Fold change	P value	Chr	Type
hsa_circ_0001364	GNB4	2.251337211	0.033633989	3	Exons
hsa_circ_0001821	PVT1	3.287035969	0.001654397	8	Exons
hsa_circ_0041638	ENO3	4.89323537	0.008760362	17	Exons
hsa_circ_0012397	FAF1	3.060338579	0.046433022	1	Exons
hsa_circ_0000665	UBE2I	5.659752071	0.042535215	16	Exons
hsa_circ_0009109	DCAF6	2.800439031	0.039049502	1	Exons
hsa_circ_0004119	RAB23	4.324163375	0.008493806	6	Exons
hsa_circ_0008192	PTBP3	3.677016974	0.002126932	9	Exons
hsa_circ_0001998	FUT8	-3.222198465	0.018780302	14	Exons
hsa_circ_0069338	SEPSECS	-2.876947158	0.018265975	4	Exons
hsa_circ_0015262	SUCO	-2.506859008	0.020001902	1	Exons
hsa_circ_0006559	TXNDC11	-5.320530101	0.005367026	16	Exons
hsa_circ_0001818	UBR5	-3.361273221	0.020285073	8	Exons
hsa_circ_0094976	ALG9	-4.805587534	0.041057838	11	Exons
hsa_circ_0063809	CELSR1	-4.065924801	0.007520253	22	Exons
hsa_circ_0001766	PDIA4	-2.841537714	0.000632629	7	Exons

in adjacent normal tissues ($P < 0.0001$) (Figure 3A). To verify the organ specificity of hsa_circ_0001821, the relative expression of hsa_circ_0001821 was calculated in 20 pairs of breast cancer tissues (Figure 3B), 22 pairs of lung cancer tissues (Figure 3C), and 20 pairs of colorectal cancer (CRC) tissues (Figure 3D). The results showed that the hsa_circ_0001821 expression in breast cancer and lung cancer was not statistically significant while it was increased in CRC tissues. With the clinicopathological parameters of these patients taken into account, we may conclude that hsa_circ_0001821 was organ specific in GC.

Evaluation of the Diagnostic Value of hsa_circ_0001821 in GC Patients

To see whether hsa_circ_0001821 could be utilized as a potential GC diagnostic marker, we depicted the ROC curve and calculated the area under the curve (AUC) based on the data obtained from the 80 pairs of GC tissues. The AUC of hsa_circ_0001821 in differentiating GC tissues from noncancerous ones was 0.792 (95%CI: 0.723–0.861, $P < 0.001$) (Figure 4A). In view of the noninvasiveness of liquid biopsy, we also detected the expression of hsa_circ_0001821 in peripheral blood samples of 30 GC patients and collected 30 fresh normal whole-blood samples as the healthy control. Consistent with the finding in the tissue samples, the hsa_circ_0001821 expression was also downregulated in the peripheral blood samples of GC patients (Figure 4B). Then we performed ROC analysis to verify the clinical utility of circulating hsa_circ_0001821 in GC diagnosis. The data showed that the AUC of circulating hsa_circ_0001821 in distinguishing GC patients from the healthy donors was 0.872 (95%CI: 0.767–0.977), which is higher than that of CEA (0.839, 95%CI: 0.740–0.937), CA199 (0.771, 95%CI: 0.649–0.893), and CA125 (0.742, 95%CI: 0.613–0.871) (Figure 4C). More importantly, the combined use of circulating hsa_circ_0001821 and the existing tumor markers CEA, CA199, and CA125 yielded the largest AUC of 0.933 (Figure 4C). Statistical analysis also showed that the combination of circulating hsa_circ_0001821 and CA199 significantly provided a sensitivity of 93.33% (Table 4).

Exploration of the Downstream Regulatory Network of hsa_circ_0001821 in GC Cells

To investigate the functional mechanism of hsa_circ_0001821 in GC cells, the expression levels of hsa_circ_0001821 in five GC cell lines (SGC-7901, HGC-27, BGC-823, AGS, and MKN-1) were detected, using the normal gastric mucosal epithelial GES-1 cells as the control. Similarly, hsa_circ_0001821 showed a significantly lower expression level in the five GC cell lines ($P < 0.01$) (Figure 5A). We further extracted RNA from SGC-7901 cells by nucleoplasm separation and found that hsa_circ_0001821 accounted for a higher proportion in the cytoplasm, suggesting that it might participate in GC progression mainly through posttranscriptional regulation (Figure 5B). Next, the potential circRNA–miRNA–mRNA regulatory axis in GC was predicted by using high-throughput sequencing and bioinformatics analysis. As shown in Figure 5C, seven miRNAs (miR-1208, miR-1825, miR-197, miR-203, miR-339-3p, miR-526b, and miR-1827) and their corresponding target mRNAs were depicted, which may provide a new direction in exploring the regulatory network of hsa_circ_0001821 in GC in the future.

DISCUSSION

circRNAs are a subclass of ncRNAs widely expressed in mammalian cells. Ample evidence has shown that circRNAs are mainly produced by precursor messenger RNAs (pre-mRNAs) via variable splicing (Ashwal-Fluss et al., 2014). circRNAs were first detected in RNA viruses in the 1970s (Kos et al., 1986). Then in 1979, researchers discovered for the first time under the electron microscope that circRNAs were also present in the cytoplasm of eukaryotic cells (Hsu and Coca-Prados, 1979). It was also reported that some exon-derived circRNAs existed in the mitochondria of yeast and human cells (Arnberg et al., 1980). However, due to the immature technology at that time, circRNAs were only regarded as a kind of low-abundance RNA molecule formed by the incorrect splicing of exon transcripts and were not further studied. With the development of high-throughput sequencing technology and bioinformatics, circRNAs have been

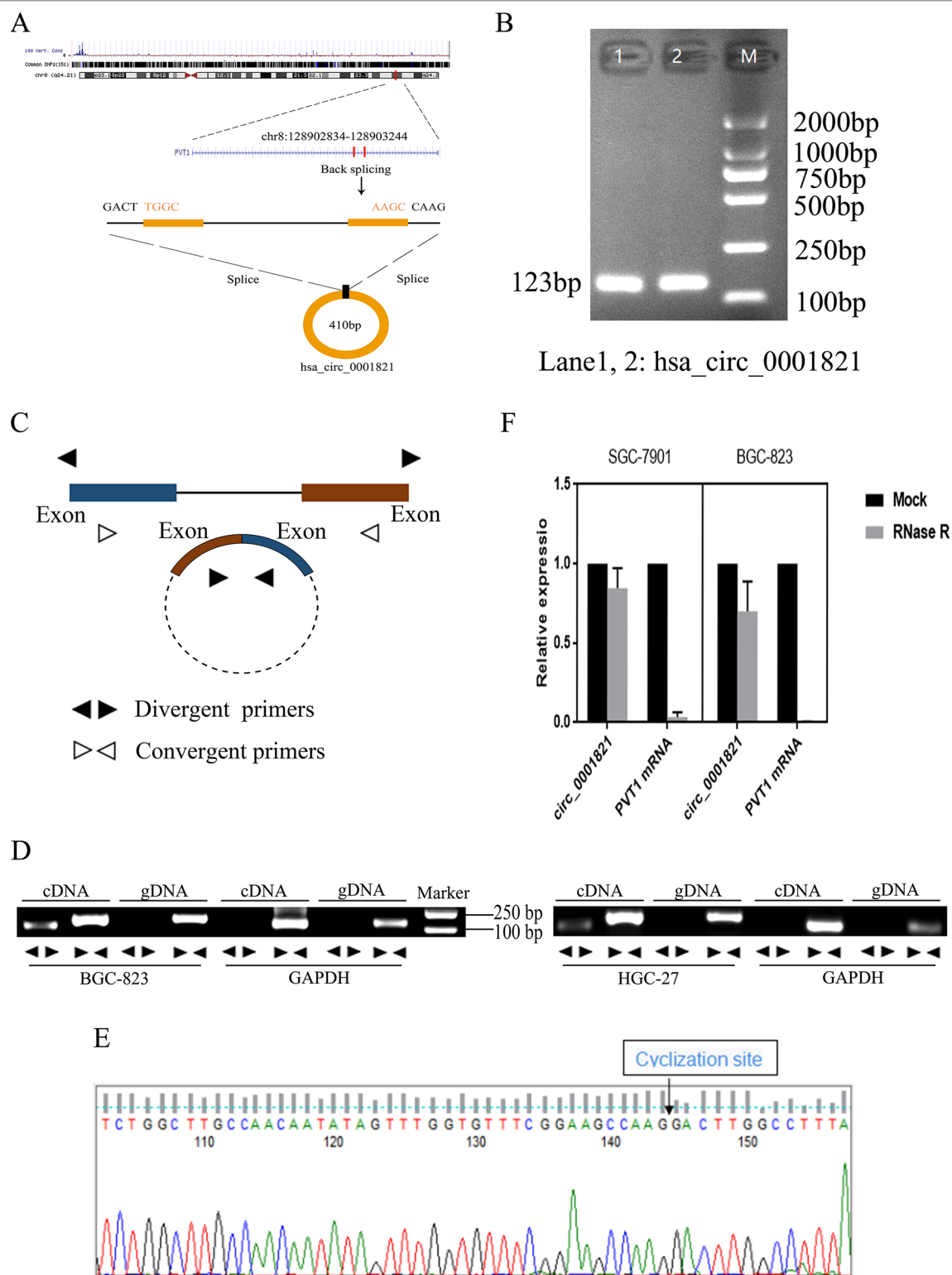


FIGURE 2 | Methodological evaluation of hsa_circ_0001821 in gastric cancer (GC) cells. **(A)** The origin of hsa_circ_0001821 searched through the University of California Santa Cruz (UCSC) genome browser. **(B)** Verification of the size of the primer amplification product (123 bp) by agarose gel electrophoresis. **(C)** A schematic diagram for the design of polymerized primers and reverse primers. **(D)** Verification of the ring structure of hsa_circ_0001821. **(E)** Detection of the cyclization site by Sanger sequencing. **(F)** Hsa_circ_0001821 was tolerable to the degradation of RNA exonuclease in SGC-7901 and BGC-823 cells.

TABLE 2 | The association between hsa_circ_0001821 expression and the clinicopathological parameters in GC patients.

Characteristics	n	High expression (n=18)	Low expression (n=62)	P-Value
Gender				
male	55	12	43	0.8285
female	25	6	19	
Age(years)				
≥60	62	16	46	0.1887
<60	18	2	16	
Histological differentiation				
Poorly	46	7	39	0.0696
Moderately	34	11	23	
Tumor depth				
T1-T2	19	9	10	0.0030**
T3-T4	61	9	52	
Lymph node metastasis				
Yes	56	8	48	0.0072**
No	24	10	14	
Tumor size(cm)				
≥5	30	7	23	0.8900
<5	50	11	39	
CEA(ng/ml)				
>5.0	27	9	18	0.0977
≤5.0	53	9	44	
CA199(U/ml)				
>37.0	19	7	12	0.0864
≤37.0	61	11	50	
CA125(U/ml)				
>35.0	24	6	18	0.7259
≤35.0	56	12	44	

Statistical analyses were carried out using Pearson χ^2 test.

** $P < 0.01$ was considered significant.

TABLE 3 | Spearman correlation analysis of hsa_circ_0001821 expression and the clinicopathological parameters in GC patients.

Variables	hsa_circ_0001821 expression level	
	Spearman correlation	P-Value
Gender	-0.046	0.684
Age(years)	0.022	0.846
Histological differentiation	0.010	0.927
Tumor depth	-0.255	0.022*
Lymph node metastasis	-0.235	0.036*
Tumor size(cm)	0.004	0.969
CEA(ng/ml)	-0.177	0.116
CA199(U/ml)	-0.193	0.087
CA125(U/ml)	-0.045	0.689

* $P < 0.05$ was considered significant.

widely found in eukaryotic cells, and their expression levels are specific to species, tissues, and time (Memczak et al., 2013).

More studies have reported that circRNAs are involved in the development of malignant tumors. A study revealed that a group of circRNAs specifically participated in the invasive growth of pancreatic ductal adenocarcinoma cells (Li et al., 2018b). Besides, a series of circRNAs has been reported to be aberrantly expressed in hepatocellular carcinoma (HCC), making early detection of liver cancer possible (Zhang et al., 2018). In our study, we

used three pairs of fresh GC tissues and their corresponding benign adjacent tissues to identify a number of circRNAs with significant expression differences through circRNA-seq. From the 2,007 differentially expressed circRNAs, we selected 16 circRNAs for initial verification, among which we finally selected hsa_circ_0001821 in view of the relationship between its parental genes and GC. Our present study showed that hsa_circ_0001821 was significantly downregulated in both GC tissues and whole-blood specimens, implying the potential role of hsa_circ_0001821 in GC evolution. However, one contradiction was that the expression trend of hsa_circ_0001821 in GC tissues *via* high-throughput sequencing was contrary to that detected by qRT-PCR. A similar situation also appeared in Li's article (Li et al., 2018a). After a simple analysis, we suspected that the small sample size and individual differences might be the main factors accounting for the contradictory results. High-throughput sequencing consisted of only three pairs of GC tissues, which may not represent the total number of GC patients, and there might be individual differences between each GC patient.

An ideal tumor marker should be organ specific. Our study discovered that hsa_circ_0001821 was not significantly expressed in breast cancer and lung cancer tissues but upregulated in CRC tissues, which is opposite to the finding in GC tissues. Analysis on the clinicopathological parameters also showed that hsa_circ_0001821 was significantly correlated with tumor depth and lymph node metastasis of GC patients. Spearman correlation analysis also indicated that decreased hsa_circ_0001821 expression was negatively correlated with tumor depth and lymph node metastasis. However, no correlation was observed between hsa_circ_0001821 and other tumors, supporting its organ specificity in GC. Knowing that circRNAs are mainly present in white blood cells (WBCs) but are excessively depleted in the serum samples, we chose whole-blood samples to isolate circulating circRNAs. ROC analysis proved that the AUC of circulating hsa_circ_0001821 in distinguishing GC patients from the healthy donors was 0.872, which is higher than that in GC tissues and other laboratory markers of CEA, CA199, and CA125. More importantly, combining circulating hsa_circ_0001821 with other existing tumor markers yielded a maximum AUC of 0.933. These results suggest that hsa_circ_0001821 could be utilized as a biomarker with favorable sensitivity and specificity in GC.

As circRNAs are connected at the 3' and 5' ends by exon or intron cyclization forming a complete ring structure, they are not easily degraded by exonuclease and therefore more stable than linear RNAs (Lasda and Parker, 2014). It was found in our study that hsa_circ_0001821 was not significantly degraded after RNA exonuclease treatment as compared with linear PVT1, indicating that hsa_circ_0001821 is relatively stable. Evidence has shown that increased lncRNA PVT1 expression is closely correlated with GC progression. For example, PVT1 was reported to participate in angiogenesis *via* activating the STAT3/VEGFA axis in GC (Zhao et al., 2018). Besides, it was highly responsible for cisplatin resistance and multidrug resistance in GC cells (Zhang et al., 2015; Zhang et al., 2017). Other reports showed that PVT1 might serve as a promising biomarker for early detection and prognostic prediction of GC (Kong et al., 2015; Yuan et al., 2016). It was found in our study that hsa_circ_0001821 originating from

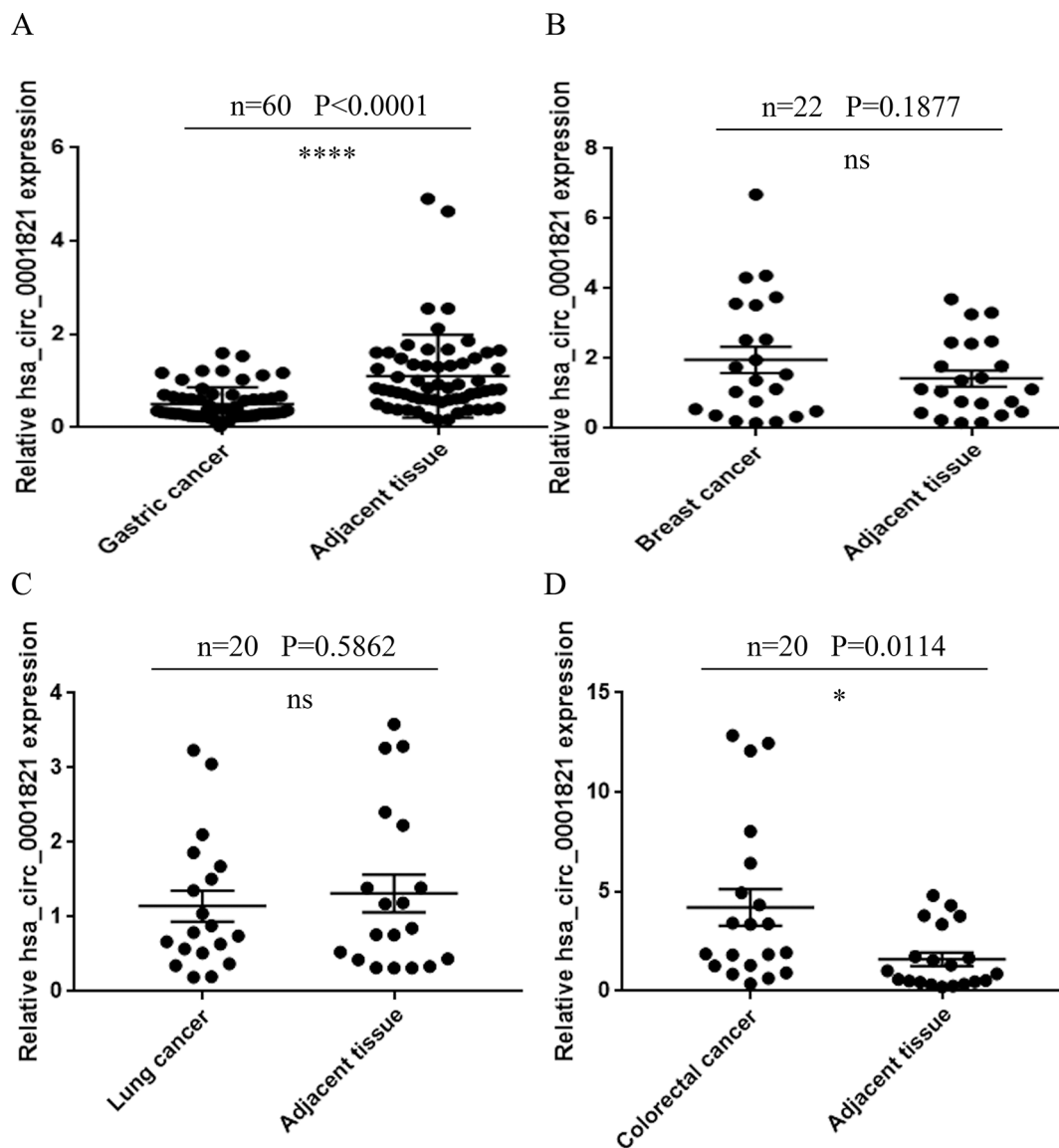


FIGURE 3 | Validation of hsa_circ_0001821 expression in different tumor tissues. **(A)** Extended verification of hsa_circ_0001821 expression in 60 pairs of gastric cancer (GC) tissues by quantitative real-time PCR (qRT-PCR). Detection of hsa_circ_0001821 expression in breast cancer **(B, n = 22)**, lung cancer **(C, n = 20)**, and colorectal cancer tissues **(D, n = 20)**. * $P < 0.05$ and **** $P < 0.0001$ were considered significant.

its parent gene PVT1 (which was upregulated in GC) had a more stable ring structure and was significantly downregulated in GC. This different expression trend between hsa_circ_0001821 and its parental gene PVT1 will inspire us to explore its regulatory axis in our future research.

In view of the characteristics and their increasing importance in tumor development, we believed that circRNAs have advantages in acting as clinical diagnostic markers, and we hope that further study on the circRNA-associated mechanism in GC development would shed new light on GC treatment. It was found in our study that hsa_circ_0001821 was significantly downregulated in the five GC cell lines. Our nucleoplasm separation assay indicated that hsa_circ_0001821 accounted for a higher proportion in the

cytoplasm, suggesting that it may play a regulatory role in GC progression at the posttranscriptional level. Additionally, the circRNA-miRNA-mRNA regulatory axis in GC was predicted. The bioinformatics analysis illustrated that hsa_circ_0001821 could potentially interact with miR-1208, miR-1825, miR-197, miR-203, miR-339-3p, miR-526b, and miR-1827. Among these miRNAs, miR-197 was found to exert an inhibitory effect on human gastric carcinogenesis and progression by regulating the MTDH/PTEN/AKT signaling pathway (Liao et al., 2018). Besides, miR-203 was able to inhibit the malignant phenotype of GC cells and served as a noninvasive biomarker for predicting prognosis and metastasis in GC patients (Imaoka et al., 2016; Zhou et al., 2016; Gao et al., 2017; Li et al., 2019). The rs8506G >

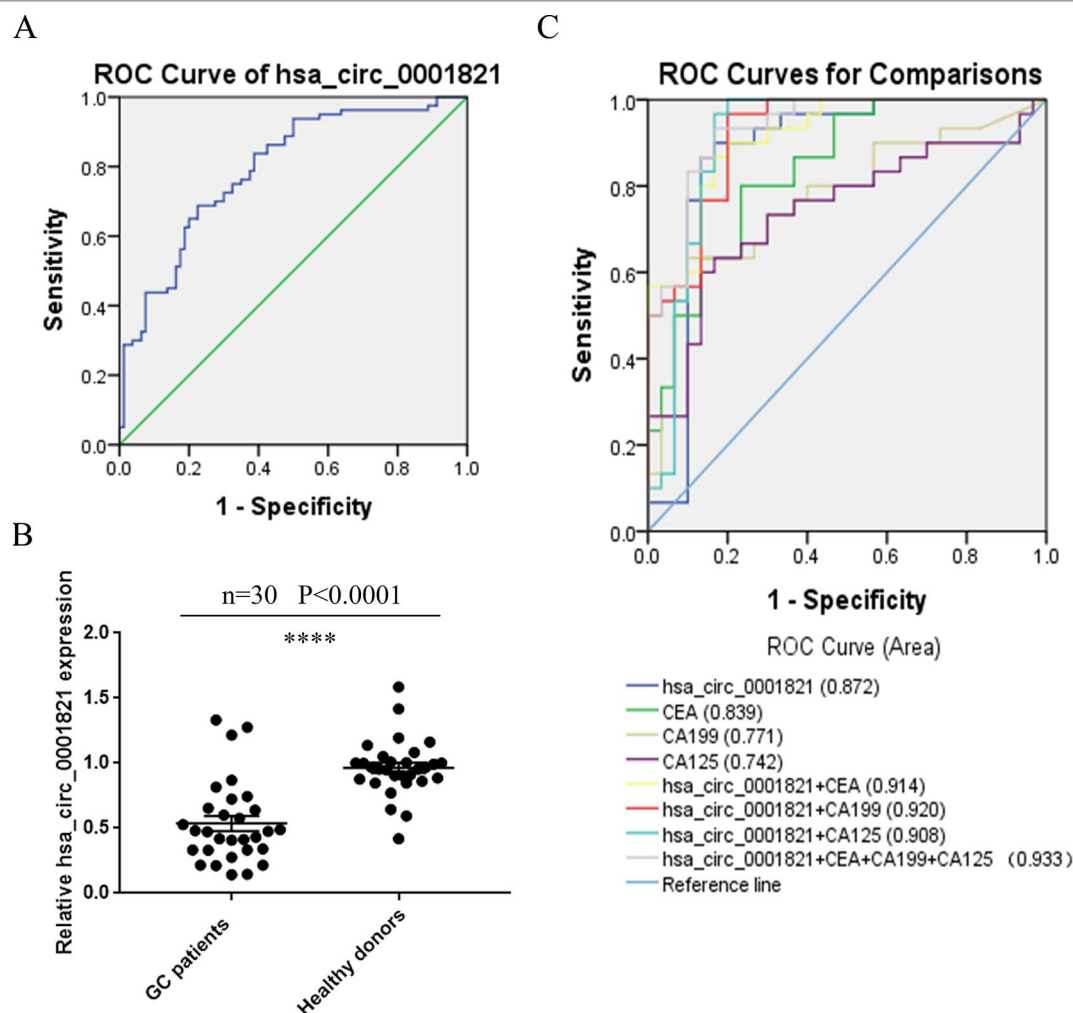
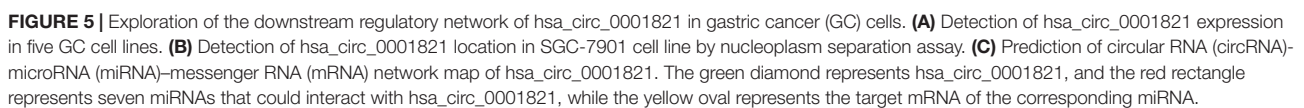


FIGURE 4 | Evaluation of the diagnostic value of hsa_circ_0001821 in gastric cancer (GC) patients. **(A)** Receiver operating characteristic (ROC) analysis of hsa_circ_0001821 in differentiating GC tissues from noncancerous tissues ($n = 80$). **(B)** Detection of circulating hsa_circ_0001821 expression in the whole-blood samples from GC patients ($n = 30$) and healthy donors ($n = 30$). **(C)** The construction of the joint diagnostic model containing circulating hsa_circ_0001821 and existing laboratory indicators. **** $P < 0.0001$ was considered significant.

TABLE 4 | Evaluation of the diagnostic values of combination of hsa_circ_0001821, CEA, CA199 and CA125.

	SEN, %	SPE, %	ACCU, %	PPV, %	NPV, %
hsa_circ_0001821	86.67(26/30)	86.67(26/30)	86.67(52/60)	86.67(26/30)	86.67(26/30)
CEA	63.33(19/30)	86.67(26/30)	75.00(45/60)	82.61(19/23)	70.27(26/37)
CA199	60.00(18/30)	90.00(27/30)	75.00(45/60)	85.71(18/21)	69.23(27/39)
CA125	60.00(18/30)	86.67(26/30)	73.33(44/60)	81.82(18/22)	68.42(26/38)
hsa_circ_0001821+CEA	90.00(27/30)	80.00(24/30)	85.00(51/60)	81.82(27/33)	88.89(24/27)
hsa_circ_0001821+CA199	93.33(28/30)	80.00(24/30)	86.67(52/60)	82.35(28/34)	92.31(24/26)
hsa_circ_0001821+CA125	90.00(27/30)	73.33(22/30)	81.67(49/60)	77.14(27/35)	88.00(22/25)
hsa_circ_0001821+CEA+CA199+CA125	93.33(28/30)	63.33(19/30)	78.33(47/60)	71.79(28/39)	90.48(19/21)

SEN, sensitivity; SPE, specificity; ACCU, overall accuracy; PPV, positive predictive value; NPV, negative predictive value.



In summary, we identified approximately 2,007 circRNAs that were significantly differentially expressed in GC through high-throughput sequencing. Among these circRNAs, hsa_circ_0001821 was significantly downregulated in both GC tissues and whole-blood specimens. These data suggest that

September 2019 | Volume 10 | Article 878

MATERIALS AND METHODS

Specimen Collection

From September 2016 to December 2018, 80 pairs of GC tissues were collected in the Affiliated Hospital of Nantong University (Nantong, China). The tissue samples were added to an RNA fixative agent (Biotek, Beijing, China) immediately after excision and stored at -80°C . In addition, a total of 60 peripheral blood samples (stored in EDTA tubes), including 30 GC patients and 30 healthy controls, were also included in this study. All the included patients were diagnosed by professional pathologists and clinicians and did not receive preoperative chemotherapy or radiotherapy. All the samples described above were collected in accordance with the Code of Ethics of the World Medical Association, and informed consent was obtained for experimentation with human subjects. The study was approved by the ethics committee of the local hospital (ethical review report number: 2018-L055).

Cell Culture

Human GC cell lines (SGC-7901, HGC-27, BGC-823, AGS, and MKN-1) were purchased from the Stem Cell Bank of the Chinese Academy of Sciences (Shanghai, China). Human normal gastric epithelial GES-1 cells were used as the normal control. All cell lines were cultured in RPMI 1640 medium (Corning, Manassas, VA) supplemented with 10% fetal bovine serum (FBS, Gibco, Grand Island, NY), 1% penicillin and streptomycin in a humidified incubator (37°C , 5% CO_2).

Nucleoplasm Separation Assay

The nuclear/cytoplasmic RNA was isolated from SGC-7901 cells using a PARIS™ Kit (Thermo Fisher Scientific) following the protocol and subjected to qRT-PCR analysis. Up to 10^7 fresh cultured cells were collected for the experiment. After one wash with phosphate-buffered saline (PBS), cells were resuspended in 300- μl ice-cold cell fractionation buffer, incubated on ice for 5–10 min, and centrifuged at 4°C , $500 \times g$, for 3 min. Then the cytoplasmic fraction was carefully aspirated away from the nuclear pellet. Subsequently, approximately 400- μl ice-cold cell disruption buffer and an equivalent volume of $2\times$ lysis/binding solution were added to the nuclear pellet. After mixing upside down, 400- μl 100% ethanol was added to the mixture. Then the sample mixture was drawn through a filter cartridge. Following orderly washing, centrifugation, and filtration, RNA was eluted twice with elution solution at 95°C . Finally, the isolated nuclear/cytoplasmic RNA was stored at -80°C for later use.

RNA Exonuclease Digestion Assay

Ribonuclease R (RNase R) was purchased from Genesee Biotech Co., Ltd (Guangzhou, China). About 3–4 U/ μg of RNase R was added to 10- μg total RNA extracted from SGC-7901 and BGC-823 cells. Subsequently, we configured a total of 50- μl digestion reaction system containing 5- μl $10\times$ reaction buffer and then added RNase-free water to make up the total volume. Next, the reaction mixture was incubated at 37°C for 30 min and

kept at 70°C for 10 min to inactivate the enzyme before reverse-transcription reaction was performed.

Total RNA Extraction and qRT-PCR

Total cell and tissue RNA were extracted using TRIzol reagent (Invitrogen, Karlsruhe, Germany), while the peripheral blood samples were pretreated with erythrocyte lysate (Beyotime, Shanghai, China), and then RNA was extracted with TRIzol reagent. Total RNA in each sample was quantified as indicated by NanoDrop™ One (Thermo Fisher Scientific, USA). RNA integrity and gDNA contamination were verified by standard denaturalized agarose gel electrophoresis, and purity was determined by spectrophotometry at 260–280 nm. cDNA was synthesized using reverse-transcription reagent (Thermo Fisher Scientific). The relative expression of hsa_circ_0001821 was normalized by the housekeeping gene GAPDH. All primers used in this study were synthesized by RiboBio Corporation (Suzhou, China). The sequences of the target gene are as follows: hsa_circ_0001821: 5'-tggaatgtaagaccccgact-3' (forward) and 5'-ccatcttgaggggcatctt-3' (reverse); PVT1: 5'-gcatggagcttcgtcaagt-3' (forward) and 5'-gccacagcctccctaaac-3' (reverse); GAPDH: 5'-gaacgggaagctcactgg-3' (forward) and 5'-gcctgctccaccactct-3' (reverse). All qRT-PCR assays were performed on the LightCycler 480 system for a total of 20 μl . The $2^{-\Delta\Delta\text{CT}}$ method was used to calculate the relative expression level, and the $\Delta\Delta\text{Ct}$ value was presented as the difference between the experimental group ($\text{Ct}_{\text{target}} - \text{Ct}_{\text{reference}}$) and the calibrator group ($\text{Ct}_{\text{target}} - \text{Ct}_{\text{reference}}$). All experiments were performed independently three times.

High-Throughput Sequencing

Total RNA was isolated from the tissues using HiPure Total RNA Mini Kit (Magen, Germany). The RNA concentration was determined using the Qubit 3.0 fluorometer (Invitrogen, Carlsbad, CA), and RNA integrity assays were performed using the Agilent 2100 Bioanalyzer (Applied Biosystems, Carlsbad, CA). A RIN value over 7.0 was considered eligible. RNA-seq library was prepared with approximately 2- μg total RNA using KAPA RNA HyperPrep Kit with RiboErase (HMR) for Illumina® (Kapa Biosystems, Inc., Woburn, MA). Briefly, total RNA was incubated at 37°C for 30 min with 10 units RNase R (Epicentre Technologies, Madison, WI) after removal of ribosomal RNA. Next, the RiboMinus RNase R (+) RNA was fragmented, and then first-strand and directional second-strand syntheses were performed. Subsequently, a tailing/adaptor ligation approach was performed with the purified cDNA. Finally, the purified, adapter-ligated DNA was amplified. Each library was diluted to 10 nM and pooled equimolar prior to clustering. Paired-end (PE150) sequencing was performed on all samples.

Identification of Differentially Expressed circRNAs via circRNA-Seq

As for the screening of differentially expressed circRNAs, the reads were first mapped to the latest UCSC transcript set using Bowtie 2 version 2.1.0 (Langmead and Salzberg, 2012) and

the gene expression level was estimated using RSEM v1.2.15 (Li and Dewey, 2011). Trimmed mean of *M*-value (TMM) was used to normalize the gene expression. Differentially expressed genes were identified using the edgeR program (Robinson et al., 2010). Genes showing altered expression with $P < 0.05$ and more than twofold changes were considered differentially expressed. Uncharacterized circRNAs were regarded as new circRNAs or less studied circRNAs. We firstly used DCC software to identify circRNAs in RNA-seq. The specific procedure is that the DCC software combined with STAR software to compare the sequencing reads to the reference genome, and then the DCC software filtered out the linear sequences aligned to the reference genome. The circRNAs containing the junction site were then identified from the unpaired sequences. After the DCC recognized the circRNA, it was compared to see whether the circRNA_ID (chromosomal coordinates) was in the circBase/circBank database, and if so, the corresponding ID was given; if not, it was represented by NA.

Construction of the circRNA-miRNA-mRNA Regulatory Network via Bioinformatics Software

Based on the circRNA-seq data, we firstly searched for hsa_circ_0001821-targeted miRNAs in the CircInteractome database (<https://circinteractome.nia.nih.gov>) and found that the context + score percentile of seven miRNAs was greater than 85. Secondly, we searched the miRDB database (<http://mirdb.org/miRDB/index.html>) for the downstream target genes of the above seven miRNAs and selected the top 10 genes for network mapping.

Statistical Analysis

The statistical analysis was conducted by GraphPad Prism 7.0 (GraphPad Software, La Jolla, CA) and SPSS 20.0 (SPSS, Inc., Chicago, USA). The clustered heatmap and volcano plots were generated

via R version 3.5.1 (*R: A Language and Environment for Statistical Computing*, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2018, <https://www.R-project.org>). Student's *t* test was performed on data of two groups, and paired *t* test was used for comparison of cancerous tissues and adjacent noncancerous tissues. When there were more than two groups of data to compare, we used one-way ANOVA. The ROC curve was established to evaluate the diagnostic value. Youden index (also known as the correct index, Youden index = specificity + sensitivity – 1) was calculated to assess the authenticity of the screening test. The correlation between hsa_circ_0001821 and the clinicopathological parameters was evaluated by chi-square test and Spearman correlation test. A *P* value of less than 0.05 was considered statistically significant.

DATA AVAILABILITY

The datasets generated for this study can be found in GEO database, GSE131414.

ETHICS STATEMENT

The study was approved by the ethics committee of the Affiliated Hospital of Nantong University.

AUTHOR CONTRIBUTIONS

SK wrote the manuscript and performed the experiences; QY helped write the manuscript and perform the experiences; CT helped collect the data; TW interpreted the results; XS and SJ conceived and designed the project, gave vital suggestions and approved the final version.

FUNDING

This project was supported by grants from the National Natural Science Foundation of China (81871720).

REFERENCES

- Arnberg, A. C., Van Ommen, G. J., Grivell, L. A., Van Bruggen, E. F., and Borst, P. (1980). Some yeast mitochondrial RNAs are circular. *Cell* 19 (2), 313–319. doi: 10.1016/0092-8674(80)90505-X
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56 (1), 55–66. doi: 10.1016/j.molcel.2014.08.019
- Beermann, J., Piccoli, M. T., Viereck, J., and Thum, T. (2016). Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiol. Rev.* 96 (4), 1297–1325. doi: 10.1152/physrev.00041.2015
- Cheng, X. J., Lin, J. C., and Tu, S. P. (2016). Etiology and prevention of gastric cancer. *Gastrointest. Tumors* 3 (1), 25–36. doi: 10.1159/000443995
- Fan, Q. H., Yu, R., Huang, W. X., Cui, X. X., Luo, B. H., and Zhang, L. Y. (2014). The has-miR-526b binding-site rs8506G > a polymorphism in the lincRNA-NR_024015 exon identified by GWASs predispose to non-cardia gastric cancer risk. *PLoS One* 9 (3), e90008. doi: 10.1371/journal.pone.0090008
- Gao, P., Wang, S., Jing, F., Zhan, J., and Wang, Y. (2017). microRNA-203 suppresses invasion of gastric cancer cells by targeting ERK1/2/Slug/E-cadherin signaling. *Cancer Biomark.* 19 (1), 11–20. doi: 10.3233/CBM-160167
- Hsu, M. T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 280 (5720), 339–340. doi: 10.1038/280339a0
- Imaoka, H., Toiyama, Y., Okigami, M., Yasuda, H., Saigusa, S., Ohi, M., et al. (2016). Circulating microRNA-203 predicts metastases, early recurrence, and poor prognosis in human gastric cancer. *Gastric Cancer* 19 (3), 744–753. doi: 10.1007/s10120-015-0521-0
- Kong, R., Zhang, E. B., Yin, D. D., You, L. H., Xu, T. P., Chen, W. M., et al. (2015). Long noncoding RNA PVT1 indicates a poor prognosis of gastric cancer and promotes cell proliferation through epigenetically regulating p15 and p16. *Mol. Cancer* 14, 82. doi: 10.1186/s12943-015-0355-8
- Kos, A., Dijkema, R., Arnberg, A. C., van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 323 (6088), 558–560. doi: 10.1038/323558a0
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi: 10.1038/nmeth.1923

- Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA* 20 (12), 1829–1842. doi: 10.1261/rna.047126.114
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi: 10.1186/1471-2105-12-323
- Li, B., Wang, F., Li, X., Sun, S., Shen, Y., and Yang, H. (2018a). Hsa_circ_0008309 may be a potential biomarker for oral squamous cell carcinoma. *Dis. Markers* 2018, eCollection. doi: 10.1155/2018/7496890
- Li, J., Zhang, B., Cui, J., Liang, Z., and Liu, K. (2019). miR-203 inhibits the invasion and EMT of gastric cancer cells by directly targeting annexin A4. *Oncol. Res.* doi: 10.3727/096504018X15444387696532
- Li, Z., Yanfang, W., Li, J., Jiang, P., Peng, T., Chen, K., et al. (2018b). Tumor-released exosomal circular RNA PDE8A promotes invasive growth via the miR-338/MACC1/MET pathway in pancreatic cancer. *Cancer Lett.* 432, 237–250. doi: 10.1016/j.canlet.2018.04.035
- Liao, Z., Li, Y., Zhou, Y., Huang, Q., and Dong, J. (2018). MicroRNA-197 inhibits gastric cancer progression by directly targeting metadherin. *Mol. Med. Rep.* 17 (1), 602–611. doi: 10.3892/mmr.2017.7908
- 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 (7441), 333–338. doi: 10.1038/nature11928
- Miller, K. D., Siegel, R. L., Lin, C. C., Mariotto, A. B., Kramer, J. L., Rowland, J. H., et al. (2016). Cancer treatment and survivorship statistics. *CA Cancer J. Clin.* 66 (4), 271–289. doi: 10.3322/caac.21349
- Prasanth, K. V., and Spector, D. L. (2007). Eukaryotic regulatory RNAs: an answer to the ‘genome complexity’ conundrum. *Genes Dev.* 21 (1), 11–42. doi: 10.1101/gad.1484207
- Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., et al. (2015). Circular RNA: a new star of noncoding RNAs. *Cancer Lett.* 365 (2), 141–148. doi: 10.1016/j.canlet.2015.06.003
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69 (1), 7–34. doi: 10.3322/caac.21551
- Sun, W., and Yan, L. (2016). Gastric cancer: current and evolving treatment landscape. *Chin. J. Cancer* 35 (1), 83. doi: 10.1186/s40880-016-0147-6
- Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.* 220 (2), 126–139. doi: 10.1002/path.2638
- Xu, M. D., Wang, Y., Weng, W., Wei, P., Qi, P., Zhang, Q., et al. (2017). A positive feedback loop of lncRNA-PVT1 and FOXM1 facilitates gastric cancer growth and invasion. *Clin. Cancer Res.* 23 (8), 2071–2080. doi: 10.1158/1078-0432.CCR-16-0742
- Yuan, C. L., Li, H., Zhu, L., Liu, Z., Zhou, J., and Shu, Y. (2016). Aberrant expression of long noncoding RNA PVT1 and its diagnostic and prognostic significance in patients with gastric cancer. *Neoplasia* 63 (3), 442–449. doi: 10.4149/314_150825N45
- Zhang, C., Zhang, C., Lin, J., and Wang, H. (2018). Circular RNA hsa_circ_0091579 serves as a diagnostic and prognostic marker for hepatocellular carcinoma. *Cell. Physiol. Biochem.* 51 (1), 290–300. doi: 10.1159/000495230
- Zhang, X. W., Bu, P., Liu, L., Zhang, X. Z., and Li, J. (2015). Overexpression of long non-coding RNA PVT1 in gastric cancer cells promotes the development of multidrug resistance. *Biochem. Biophys. Res. Commun.* 462 (3), 227–232. doi: 10.1016/j.bbrc.2015.04.121
- Zhang, X. W., Liu, L., Zhang, X. Z., and Bo, P. (2017). Kanglaite inhibits the expression of drug resistance genes through suppressing PVT1 in cisplatin-resistant gastric cancer cells. *Exp. Ther. Med.* 14 (2), 1789–1794. doi: 10.3892/etm.2017.4650
- Zhao, J., Du, P., Cui, P., Qin, Y., Hu, C., Wu, J., et al. (2018). LncRNA PVT1 promotes angiogenesis via activating the STAT3/VEGFA axis in gastric cancer. *Oncogene* 37 (30), 4094–4109. doi: 10.1038/s41388-018-0250-z
- Zhou, P., Jiang, N., Zhang, G. X., and Sun, Q. (2016). MiR-203 inhibits tumor invasion and metastasis in gastric cancer by ATM. *Acta Biochim. Biophys. Sin. (Shanghai)* 48 (8), 696–703. doi: 10.1093/abbs/gmw063

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 Kong, Yang, Tang, Wang, Shen and Ju. 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.



An Altered Splicing Registry Explains the Differential ExSpeU1-Mediated Rescue of Splicing Mutations Causing Haemophilia A

Dario Balestra^{1*}, Iva Maestri², Alessio Branchini¹, Mattia Ferrarese¹, Francesco Bernardi¹ and Mirko Pinotti¹

¹ Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy, ² Department of Experimental and Diagnostic Medicine, University of Ferrara, Ferrara, Italy

OPEN ACCESS

Edited by:

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

Reviewed by:

Brage Storstein Andresen,
University of Southern
Denmark, Denmark
Xavier Roca,
Nanyang Technological
University, Singapore

*Correspondence:

Dario Balestra
blsdra@unife.it

Specialty section:

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

Received: 01 July 2019

Accepted: 12 September 2019

Published: 10 October 2019

Citation:

Balestra D, Maestri I, Branchini A,
Ferrarese M, Bernardi F and Pinotti M
(2019) An Altered Splicing Registry
Explains the Differential ExSpeU1-
Mediated Rescue of Splicing
Mutations Causing Haemophilia A.
Front. Genet. 10:974.
doi: 10.3389/fgene.2019.00974

The exon recognition and removal of introns (splicing) from pre-mRNA is a crucial step in the gene expression flow. The process is very complex and therefore susceptible to derangements. Not surprisingly, a significant and still underestimated proportion of disease-causing mutations affects splicing, with those occurring at the 5' splice site (5'ss) being the most severe ones. This led to the development of a correction approach based on variants of the spliceosomal U1snRNA, which has been proven on splicing mutations in several cellular and mouse models of human disease. Since the alternative splicing mechanisms are strictly related to the sequence context of the exon, we challenged the U1snRNA-mediated strategy in the singular model of the exon 5 of coagulation factor (F)VIII gene (*F8*) in which the authentic 5'ss is surrounded by various cryptic 5'ss. This scenario is further complicated in the presence of nucleotide changes associated with FVIII deficiency (Haemophilia A), which weaken the authentic 5'ss and create/strengthen cryptic 5'ss. We focused on the splicing mutations (c.602-32A > G, c.602-10T > G, c.602G > A, c.655G > A, c.667G > A, c.669A > G, c.669A > T, c.670G > T, c.670+1G > T, c.670+1G > A, c.670+2T > G, c.670+5G > A, and c.670+6T > C) found in patients with severe to mild Haemophilia A. Minigenes expression studies demonstrated that all mutations occurring within the 5'ss, both intronic or exonic, lead to aberrant transcripts arising from the usage of two cryptic intronic 5'ss at positions c.670+64 and c.670+176. For most of them, the observed proportion of correct transcripts is in accordance with the coagulation phenotype of patients. In co-transfection experiments, we identified a U1snRNA variant targeting an intronic region downstream of the defective exon (Exon Specific U1snRNA, U1sh7) capable to re-direct usage of the proper 5'ss (~80%) for several mutations. However, deep investigation of rescued transcripts from +1 and +2 variants revealed only the usage of adjacent cryptic 5'ss, leading to frameshifted transcript forms. These data demonstrate that a single ExSpeU1 can efficiently rescue different mutations in the *F8* exon 5, and provide the first evidence of the applicability of the U1snRNA-based approach to Haemophilia A.

Keywords: RNA splicing, splicing mutations, human disease, ExSpeU1, Haemophilia A

INTRODUCTION

In higher eukaryotes, the information necessary for protein synthesis is scattered across the gene, where the coding segments (exons) represent a minor proportion. Therefore, the exon recognition and the removal of the non-coding sequences (introns) from pre-mRNA are essential for proper gene expression, and this process (splicing) is carried out by a huge macromolecular complex named spliceosome. The first step of the spliceosome assembly involves binding of the small nuclear ribonucleoprotein U1 (U1snRNP) to the 5' splice site (5'ss) by complementarity with the 5' tail of its RNA component, U1snRNA (Roca et al., 2005). Not surprisingly, nucleotide changes occurring at the 5'ss, by interfering with its recognition and eventually leading to aberrant splicing events, are commonly associated with severe clinical phenotypes and are widely (9%) reported in human inherited diseases (<http://www.hgmd.org/>). This information led us to develop a correction strategy based on U1snRNAs variants designed to restore the complementarity with the mutated 5'ss (compensatory U1snRNA) (Pinotti et al., 2009). Unexpectedly, we also demonstrated the correction potential of engineered U1snRNAs targeting intronic sequences downstream of the defective exon (Exon Specific U1snRNAs; ExSpeU1), which are active on mutations occurring at the 5'ss, 3'ss as well as within exon (Alanis et al., 2012). The efficacy has been proven both in several cellular (Glaus et al., 2011; Schmid et al., 2011; Balestra et al., 2015; van der Woerd et al., 2015; Dal Mas et al., 2015a; Dal Mas et al., 2015b; Rogalska et al., 2016; Tajnik et al., 2016; Scalet et al., 2017; Scalet et al., 2018; Balestra et al., 2019; Balestra and Branchini, 2019; Scalet et al., 2019) and animal (Balestra et al., 2014; Balestra et al., 2016; Rogalska et al., 2016; Donadon et al., 2018b; Donadon et al., 2019; Lin et al., 2019) models of human disease.

However, the exon definition is very complex and, besides the splice sites, involves a series of splicing regulatory elements, which lead to the choice of the correct splice junctions and disfavor usage of the several cryptic splice sites (De Conti et al., 2013). Therefore, depending on the context, nucleotide changes can trigger different aberrant splicing mechanisms with which correction approaches, such as the U1snRNA-mediated one, must cope.

Here, we challenged the ExSpeU1s in the *F8* exon 5 as a model of context in which the authentic 5'ss is surrounded by various cryptic 5'ss (**Figure 1A**), a scenario further complicated by the occurrence of nucleotide changes at the 5'ss that are associated with coagulation factor VIII (FVIII) deficiency (Haemophilia A, HA) (Bolton-Maggs and Pasi, 2003). Minigene expression studies indicated that these changes alter the delicate interplay among 5'ss, which can be apparently re-balanced by a ExSpeU1 targeting a downstream intronic sequence. However, the deep investigation of the rescued transcripts from +1 and +2 variants revealed that the ExSpeU1 re-directed the usage of the newly-created cryptic 5'ss, thus vanishing the correction attempt. These data demonstrate the applicability of the ExSpeU1 to HA-causing mutations and strengthen the importance of the sequence context in dictating the splicing outcome.

MATERIALS AND METHODS

Creation of Expression Vectors

To create the pF8wt plasmid, the genomic region of the human *F8* gene (NC_000023.11) spanning from c.602-464 to c.670+773 was amplified from genomic DNA of a normal subject with primers i4F-i5R using high-fidelity Pfu DNA-polymerase (Transgenomic, Glasgow, UK). The *F8* amplicon was sequentially cloned into the pTB expression vector (gift of prof. F. Pagani, ICGEB, Trieste, Italy) by exploiting the *NdeI* restriction site.

To create the pF8-32A > G, pF8-10T > G, pF8.602G > A, pF8.655G > A, pF8.667G > A, pF8.669A > G, pF8.669A > T, pF8.670G > T, pF8+1G > T, pF8+1G > A, pF8+2T > G, pF8+5G > A, and pF8+6T > C plasmids, the nucleotide changes were introduced into the pF8wt minigene by site-directed mutagenesis (QuickChange II Site-Directed Mutagenesis Kit, Stratagene, La Jolla, CA, USA).

The pU1F8d, pU1F8s7, pU1F8s16, pU1F8s25 expression vectors for the modified U1snRNAs were created by replacing the sequence between the *BglII* and *XbaI* restriction sites with a PCR generated with a U1-specific forward primer (containing the modified 5' tail of the U1snRNA) and a reverse primer base-pairing downstream the *XbaI* cloning site.

The pU7a,b,c,d expression vectors for the modified U7snRNAs were created as previously reported (Balestra et al., 2015). Briefly, a PCR containing the modified binding site of the engineered U7snRNA has been generated by using the primers indicated in **Supplementary Table 1** and cloned into the pSP64 plasmid (gift from Franco Pagani, ICGEB, ITA) after digestion with *StuI* and *XbaI* restriction sites.

All vectors have been validated by sequencing.

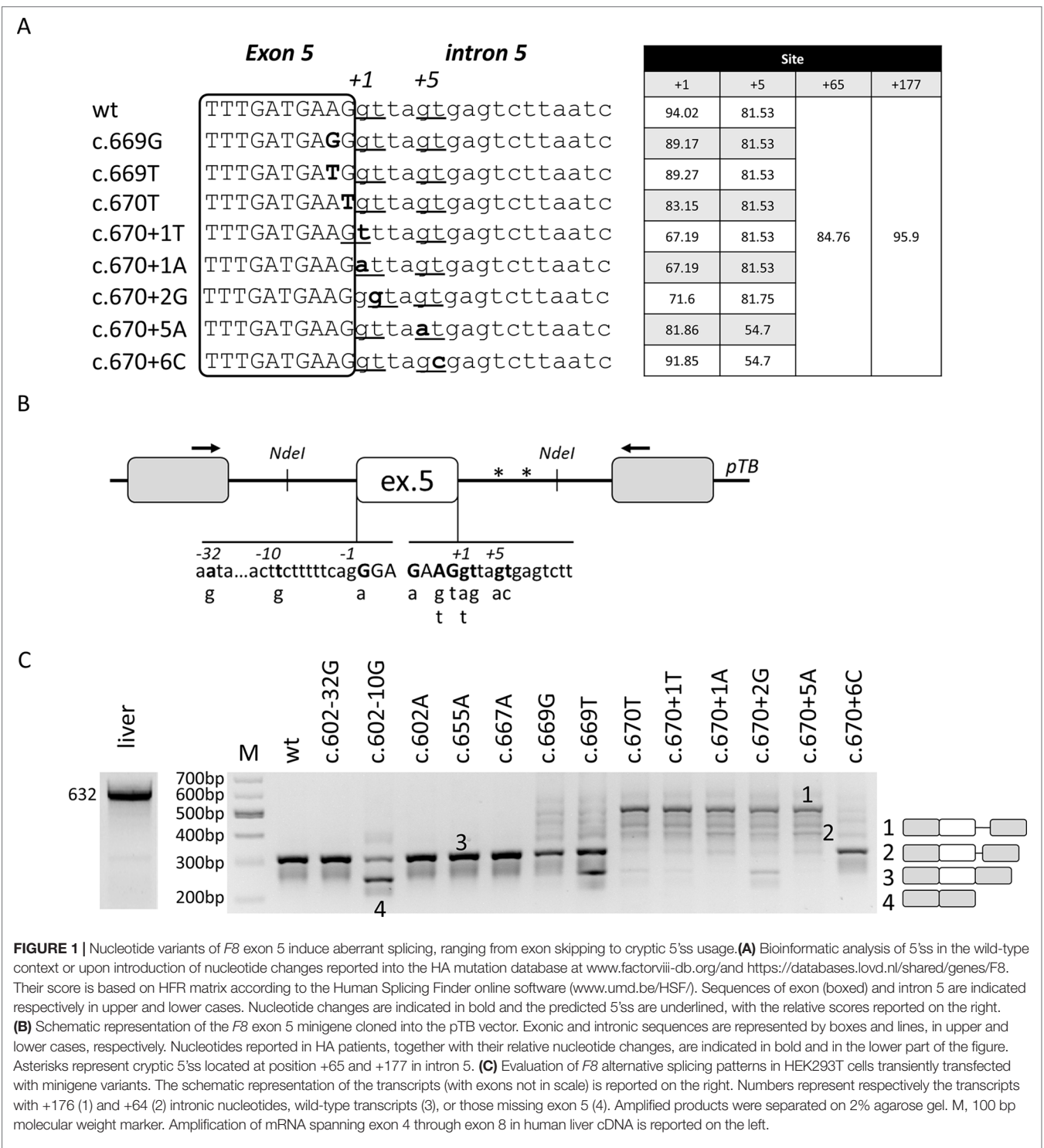
Sequences of oligonucleotides are provided in **Supplementary Table 1**.

Expression in Mammalian Cells and mRNA Studies

Human Embryonic Kidney 293T (HEK293T) cells were cultured as previously described (Farrarese et al., 2018). Cells were seeded on twelve-well plates and transfected with the Lipofectamine 2000 reagent (Life Technologies, Carlsbad, CA, USA), according to the manufacturer's protocol.

Five hundred nanograms of pF8 minigene variants were transfected alone or with a molar excess (1.5X) of the pU1/pU7 plasmids. Total RNA was isolated 24 h post-transfection with Trizol (Life Technologies), reverse-transcribed with random primers and amplified using the Pfu DNA-polymerase (Transgenomic, Glasgow, UK) with primers Alfa and Bra. The same DNA polymerase and primers 4F and 8R were used to evaluate the *F8* splicing patterns in human liver. Densitometric analysis for the quantification of correct and aberrant transcripts was performed using the ImageJ software (<https://imagej.net>).

For denaturing capillary electrophoresis analysis, the amplified fragments were labeled by using primers Alfa and the fluorescently-labeled (FAM dye) Bra and run on an ABI-3100 instrument (Waltham, MA, USA).



Three independent experiment were conducted for each variant and condition.

Computational Analysis

Computational prediction of splice sites and of splicing regulatory elements was conducted by using the Human Splicing Finder (www.umd.be/HSF/) online software.

RESULTS

The Computational Analysis Predicts Several Competing 5'ss

The bioinformatic analysis (www.umd.be/HSF3/index.html) predicts that *F8* exon 5 is well defined, as demonstrated by the high scores of the 5'ss and 3'ss (94,02 and 94,62, respectively)

(Figure 1A). Moreover, it predicts three cryptic 5'ss in intron 5, two of them located at nucleotide positions +65 and +177 bp, and one in the proximity (+5) of the authentic one. All of them have a score (81, 85, and 96 for +5, +65, and +177 cryptic 5'ss, respectively) close to that of the authentic 5'ss and, based on the HSF matrix, above the threshold of 80.

In this model, we chose to analyze a panel of nucleotide variants occurring at the 5'ss (c.667G > A, c.669A > G, c.669A > T, c.670G > T, c.670+1G > T, c.670+1G > A, c.670+2T > G, c.670+5G > A, c.670+6T) or 3'ss (c.602-32A > G, c.602-10T > G, c.602G > A), and associated with different degree of HA severity (Table 1). The introduction of nucleotide changes at the 5'ss is predicted to weaken the authentic 5'ss, with changes at the highly conserved positions +1 and +2 being the most detrimental ones (Figure 1A). Interestingly, due to the genomic context of *F8* exon 5, the introduction of the c.670+1G > T and c.670+2T > G variants is predicted to insert new and shifted alternative 5'ss that, if used, would lead to aberrant transcripts differing for only one nucleotide in size, as compared to the correct one (-1 for c.670+1G > T and +1 for c.670+2T > G).

This provided us with an informative model to assess 5'ss competition and aberrant splicing mechanisms as well as the suitability of the ExSpeU1 as a correction approach.

The *F8* Exon 5 Nucleotide Changes Lead to Aberrant Splicing, Ranging From Exon Skipping to Usage of Cryptic Intronic 5'ss

To investigate splicing mechanisms in the *F8* exon 5 context, we created a *F8* minigene including the *F8* exon 5 and the surrounding introns (Figure 1B). Expression of the wild type (wt) minigene in HEK293T cells indicated that exon 5 is well defined, and this pattern recapitulates the complete inclusion observed in human liver mRNA (Figure 1C), thus validating our experimental approach.

Since even exonic nucleotide changes such as missense mutations might affect the splicing code, we screened for the presence of exonic splicing enhancers (ESEs), which were predicted by computational analysis. To this purpose we exploited antisense U7snRNA variants designed to target and mask the

candidate ESEs (U7a,b,c) or partially the authentic 5'ss (U7d). Co-expression of these U7snRNA variants with the wt minigene revealed that only the control U7d affected splicing, and partially induced exon 5 skipping (Supplementary Figure 1). Since these data did not provide elements for a selection among the many exonic nucleotide variations annotated in the HA databases (www.factorviii-db.org/and https://databases.lovd.nl/shared/genes/F8), we only investigated the c.655G > A change, being the most frequent missense change in *F8* exon 5.

Expression studies with the minigene variants and the analysis of splicing patterns by RT-PCR and conventional electrophoresis demonstrated that some changes (c.602-32A > G, c.602G > A, c.655G > A, and c.667G > A) were ineffective on splicing (Figure 1C). Differently, the c.602-10T > G, c.669A > T, c.670G > T, c.670+1G > T, and c.670+2T > G variants led, to various extent, to exon 5 skipping. Moreover, all mutations within the 5'ss (c.669A > T, c.669A > G, c.670G > T, c.670+1G > T, c.670+1G > A, c.670+1G > T, c.670+2T > G, c.670+5G > A, and c.670+6T) led to alternative transcripts originating from the usage of cryptic 5'ss located 65 and 177 bp downstream of the authentic 5'ss (Figure 1C). All aberrant transcripts, confirmed by sequencing (Supplementary Figure 2), account for a deleted and frame-shifted mRNA, with two premature stop codons at intronic positions c.673-678.

Taken together these data identified mutations causing aberrant splicing and provided candidates to explore splicing correction by ExSpeU1s.

A Unique ExSpeU1 Is Able to Rescue Multiple Mutations but Not Variants at +1 and +2 Positions Due to an Altered Splicing Registry

In the attempt to restore proper exon definition, we designed a compensatory U1snRNA and three ExSpeU1 with perfect complementarity to the wild-type 5'ss or the adjacent intronic sequences, respectively (Figure 2A). The efficacy of these U1snRNA variants has been initially evaluated on the c.669A > T change since mutations at -2 position of the 5'ss have been previously shown to be rescuable by the modified U1snRNA-based approach (Alanis et al.,

TABLE 1 | Features of the nucleotide changes reported in the *F8* mutation databases (www.factorviii-db.org/and https://databases.lovd.nl/shared/genes/F8) investigated in the study. The HA coagulation phenotype is defined on the basis of FVIII cofactor activity levels (severe <1%; moderate, 1-5%; mild 5-40%).

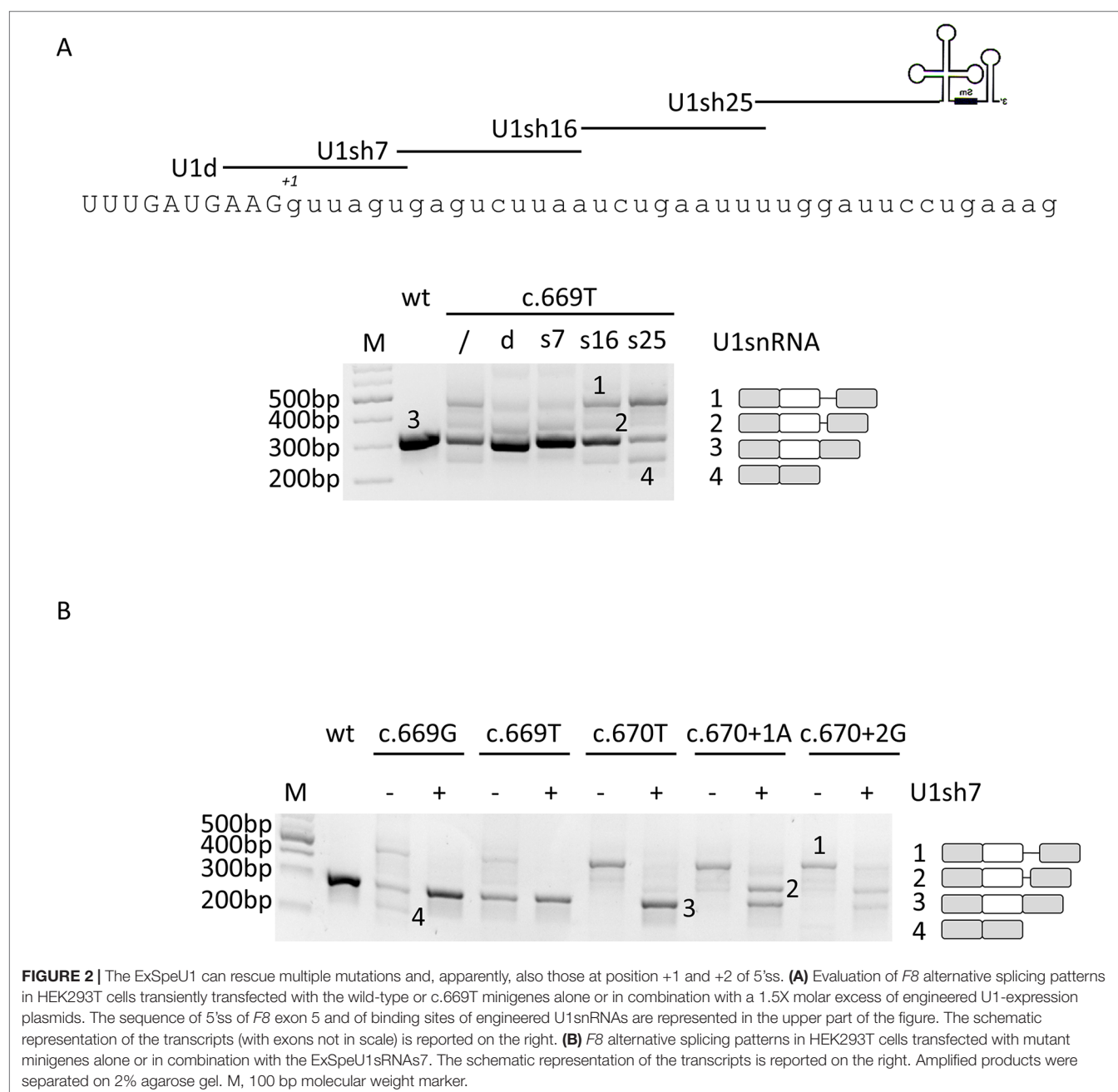
Location in gene	No of Patients reported	Mutation	Amino acid change	Coagulation phenotype
Intron 4	1	c.602-32A > G		Severe
Intron 4	1	c.602-10T > G		Mild
Exon 5	1	c.602G > A	p.Gly201Glu	Severe
Exon 5	15	c.655G > A	p.Ala219Thr	Mild
Exon 5	1	c.667G > A	p.Glu223Lys	Severe
Exon 5	1	c.669A > G	p.Glu223Glu	Mild
Exon 5	2	c.669A > T	p.Glu223Asp	Severe
Exon 5	2	c.670G > T	p.Gly224Trp	Moderate
Intron 5	3	c.670+1G > T		Severe
Intron 5	1	c.670+1G > A		Severe
Intron 5	1	c.670+2T > G		Severe
Intron 5	3	c.670+5G > A		Severe
Intron 5	4	c.670+6T > C		Mild/Moderate

2012). Co-transfection experiments led us to select the compensatory U1sRNA (U1d) and one ExSpeU1 (U1sh7) that appreciably rescued splicing. In particular, the densitometric analysis of bands revealed that co-expression of the U1d and U1sh7 was associated with an increase of correctly spliced transcripts (from $52 \pm 3\%$ to $71 \pm 3\%$ or $75 \pm 4\%$ for U1d and U1sh7, respectively) (**Figure 2A**).

Based on these results and on the fact that the ExSpeU1, by binding to a less conserved intronic region, potentially ensures higher exon specificity (Rogalska et al., 2016; Donadon et al., 2019), the U1sh7 was selected for further investigation on an enlarged panel of variants. Co-transfection experiments showed that the U1sh7 remarkably rescued the c.669A > G, c.669A > T, and c.670G >

T variants (> 80% of correct transcripts), and also appeared to have a correction effect (from 0% to ~40%) on those at the conserved +1 (c.670+1A) and +2 (c.670+2G) positions (**Figure 2B**).

The unexpected appearance of transcripts with a size compatible with correct splicing even for mutants at positions +1 and +2 prompted us to further analyze the splicing outcome by fluorescent labeling of amplicons followed by denaturing capillary electrophoresis (**Figure 3** and **Supplementary Figure 3**). This approach, in cells expressing the +1 and +2 variants alone, revealed the presence of trace levels of aberrant transcripts differing for only a few nucleotides (-1, +1 and +4 bp). In particular, the -1 and +1 aberrant transcripts were identified



only in the c.670+1T and c.670+2G context ($7.1 \pm 2.4\%$ and $1.3 \pm 0.7\%$, respectively), while the +4 transcript was detected for the c.670+1T ($1.1 \pm 0.8\%$), c.670+2G ($1.4 \pm 0.8\%$), and c.669T mutant ($1 \pm 0.6\%$).

Through this approach, we then evaluated the effect of the U1sh7, which re-directed the spliceosome on the exon-intron junction, as indicated by the generally decreased transcripts arising from the usage of the distal cryptic 5'ss. However, due to the mutations at the crucial positions, the U1sh7 forced the usage of the adjacent cryptic 5'ss, created/strengthened by mutations, and remarkably increased the proportion of +4 forms for the

c.670+1T (from 1.1 ± 0.8 to $72.6 \pm 2.2\%$) and c.670+1A (from 0 to $55.3 \pm 2.1\%$) and +1 forms for the c.670+2G (from ~1% to ~34%) variants.

Concerning the other mutants, the denaturing capillary electrophoresis permitted us to complete the proper evaluation of the U1sh7-mediated rescue on the selected panel of mutations. As shown in **Figure 3B**, the co-expression of the U1sh7 led to an appreciably increased proportion of correct transcripts for the c.602-10T > G variant at the 3'ss and the c.669A > T, c.669A > G, c.670G > T, c.670+5G > A, and c.670+6T variants at the 5'ss.

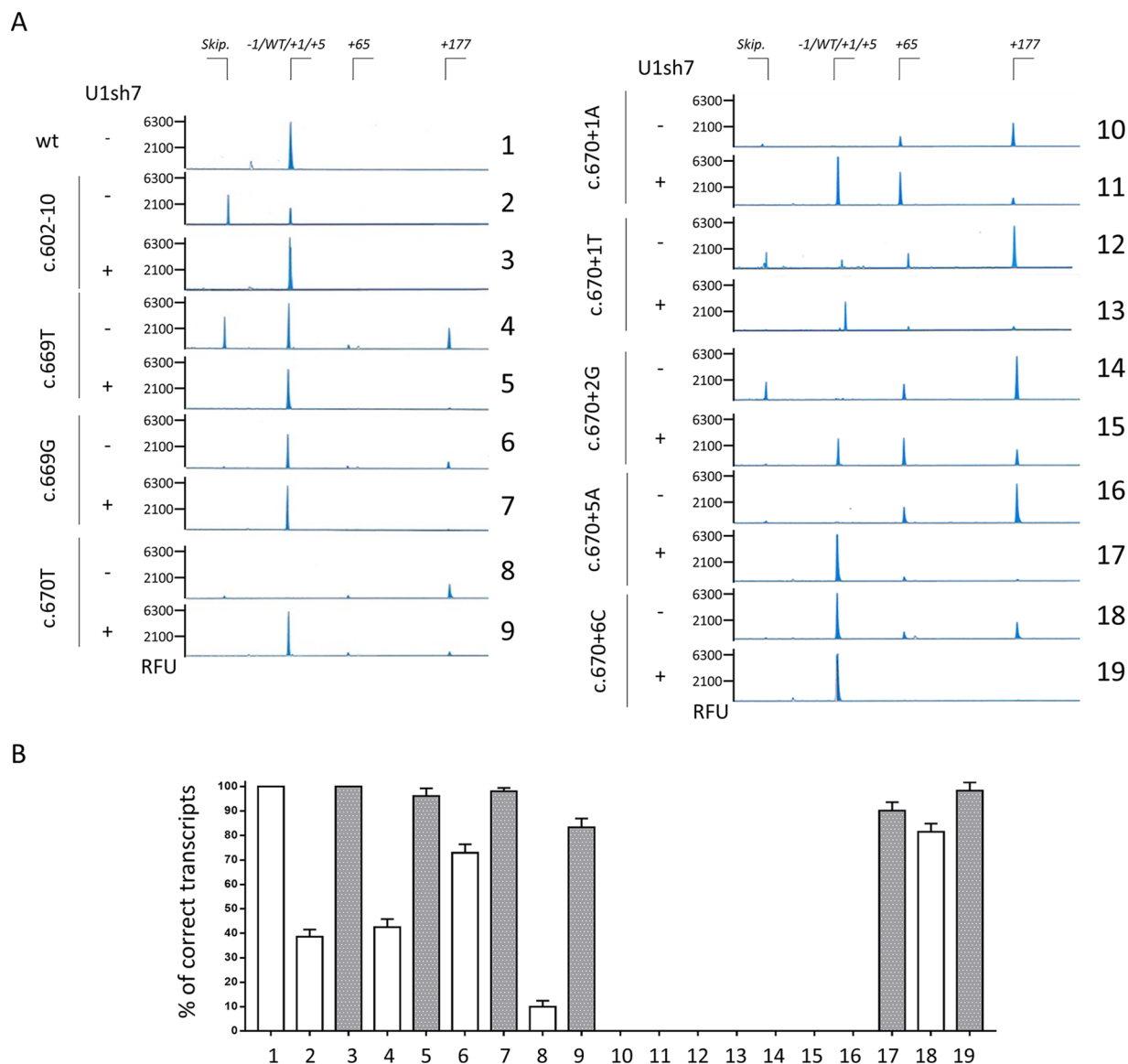


FIGURE 3 | Capillary electrophoresis-mediated analysis of splicing reveals the usage of an altered splicing registry. **(A)** Analysis of splicing patterns in HEK293T cells expressing mutant minigenes alone or in combination with ExSpeU1sRNAs7 by denaturing capillary electrophoresis of fluorescently-labelled PCR products. The amount of transcripts is represented by the area under each peak. The scheme of transcripts is reported on top. RFU: Relative Fluorescence Units. **(B)** Relative amount of correctly spliced transcripts in HEK293T cells transfected as in panel A and analyzed by denaturing capillary electrophoresis. The white and grey histograms report the percentage of correct transcripts expressed as mean \pm SD from three independent experiments before or after treatment with U1sh7.

Taken together, our data dissected further the aberrant splicing patterns associated with HA-causing mutations and identified a unique ExSpeU1 able to rescue multiple mutations, except for +1/+2 variants suffering from an unfavorable context.

DISCUSSION

The advent of next-generation sequencing has enormously expanded the number of gene variations associated with human diseases, thus posing the problem of identifying the causative ones. This is particularly difficult for nucleotide changes that, being at the exon-intron boundaries or within introns, are candidate to affect splicing since their precise effect is hardly predictable by computational tools. The scenario is further complicated by the overlapped splicing and amino acid codes within exons, which might lead to missense changes exerting their pathogenic roles by altering the splicing process rather than the protein biology (Tajnik et al., 2016; Donadon et al., 2018a).

In this context, the experimental evaluation of the impact of nucleotide changes on splicing is mandatory to help diagnosis and counseling. Here, we addressed this issue in a singular gene context, namely the *F8* exon 5, where various exonic changes and multiple cryptic 5'ss are respectively located within or in the proximity of the authentic 5'ss, thus complicating the selection of the right one.

The analysis of the splicing pattern of the c.602-10A > G nucleotide change, with appreciable levels (~38%) of correctly spliced transcripts, is consistent with the mild coagulation phenotype reported in the patient. Conversely, the c.602G > A (p.G201E), c.655G > A (p.A219T), and c.667G > A (p.E223G) missense variants are not associated with significant splicing alterations, indicating that FVIII deficiency is mainly caused by the underlying amino acid substitutions impairing protein biosynthesis/function. This observation is also strengthened by the data with antisense U7snRNAs that do not support the presence of important regulatory exonic elements in the *F8* exon 5, which might have been altered by exonic changes. However, in the proximity of the 5'ss, the splicing code overlaps with the amino acid registry, with the splicing one being the first used in the gene expression flow. Concerning mutations occurring within the 5'ss, it is worth noting that the exonic c.669A > G (p.E223E), c.669A > T (p.E223D), and c.670G > T (p.G224W) variants clearly alter splicing. Whereas the c.669A > T is mainly associated with exon skipping and loss of exon definition, the c.669A > G and particularly the c.670G > T variants lead to partial intron retention, with the usage of a strong intronic 5'ss at position +177. Noticeably, splicing analysis of variants at different positions of the same triplet (c.667G > A, c.669A > G, and c.669A > T) coding for glutamic acid at position 233 in FVIII resulted in different splicing outcomes, ranging from exon skipping to the usage of cryptic splice 5'ss or null. This finding highlights the need for careful evaluation of the effects of exonic changes on splicing, due to the limited help of bioinformatic analysis in predicting the pathogenicity of nucleotide changes. It is worth noting that levels of correct transcripts for the c.669A > T (p.E223D) (~40%) and c.670G > T (~10%) variants suggest that the associated FVIII deficiency (moderate/severe) would arise

from a combination of splicing and protein impairment. On the other hand, the mutations at the intronic positions +1/+2/+5 were not compatible with correct splicing. Differently, the +6 variant led to remarkable levels of correctly spliced transcripts, in accordance with the severe or mild coagulation phenotypes reported in HA patients, respectively.

The knowledge of aberrant splicing patterns lays the foundation for the exploration of correction approaches for therapeutic purposes, as we did in several other human disease models. Intervention at the mRNA level has the advantage of maintaining the physiological gene regulation and is based on delivery of small coding cassettes, thus allowing the exploitation of any viral vector strategy. Among the different strategies, engineered U1snRNAs demonstrated the ability of rescuing multiple mutation types, including changes at 5'ss, 3'ss as well as within exons, in cellular and animal models of human disease (Glaus et al., 2011; Schmid et al., 2011; Balestra et al., 2014; Balestra et al., 2015; van der Woerd et al., 2015; Dal Mas et al., 2015a; Dal Mas et al., 2015b; Balestra et al., 2016; Rogalska et al., 2016; Tajnik et al., 2016; Scalet et al., 2017; Scalet et al., 2018; Donadon et al., 2018b; Donadon et al., 2019; Scalet et al., 2019). Noticeably, modified U1snRNAs were shown to preserve their correction effect even when targeting intronic regions downstream of the defective exon (ExSpeU1) through a mechanism that, unlike antisense oligonucleotides blocking an intronic element, involves U1snRNP assembly, spliceosome activation, and recruitment of splicing factors (Martínez-Pizarro et al., 2018; Rogalska et al., 2016). Moreover, this second generation of modified U1snRNAs potentially ensures higher exon and gene specificity since their base-pairing ability with intronic, and thus less conserved, sequences, as supported by recent studies (Donadon et al., 2019; Rogalska et al., 2016). Notwithstanding, the off-target effect of each ExSpeU1 has to be carefully assessed when approaching clinics.

In the *F8* exon 5 context, we identified an ExSpeU1 able to restore proper exon definition in the presence of multiple mutations, located at both 3'ss or 5'ss. Interestingly, we observed that variants at +1 and +2 position of 5'ss, not thought to be rescuable by modified U1snRNA due to their high degree of conservation, resulted in transcripts with size comparable with that of correctly spliced ones. Recently, the T > C transition at position +2 of 5'ss has been demonstrated to be rescuable by modified U1snRNA (Scalet et al., 2019), a finding compatible with the observation that a small fraction of introns removed by U2-type spliceosome has cytidine at position +2 (Burset et al., 2000). There are also examples of mutations at +1 position that are compatible with correct processing (Hartmann et al., 2010), which potentially open the possibility to rescue them. Here, to dissect the elusive nature of transcripts, we exploited the denaturing capillary electrophoresis, a strategy able to distinguish amplicons differing by only one nucleotide. The analysis of the splicing patterns revealed the usage of cryptic 5'ss created by mutations (c.670+1G > T and c.670+2T > G) leading to transcripts respectively shorter or longer of one single nucleotide. Unfortunately, due to the altered registry of the 5'ss, the ExSpeU1 further promoted the usage of the 5'ss other than the authentic one, thus vanishing the correction effect.

In conclusion, through molecular characterization of various *F8* exon 5 variants occurring at the 3'ss or 5'ss, or within the exon, we demonstrated for the first time the ability of a unique ExSpeU1 to rescue multiple HA-causing *F8* mutations. Moreover, our findings highlight the need to investigate the effect on splicing of nucleotide changes, particularly of those occurring in exonic sequences, and suggest a careful inspection of the sequence context and evaluation of transcripts to avoid over-interpretations, with implications for diagnosis and counseling.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/Supplementary Files.

ETHICS STATEMENT

Ethics approval for this study was not required as per the local legislation. Notwithstanding, the DNA sample from control patient was used after obtaining the informed consent.

REFERENCES

- Alanis, E. F., Pinotti, M., Mas, A. D., Balestra, D., Cavallari, N., Rogalska, M. E., et al. (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* 21, 2389–2398. doi: 10.1093/hmg/dds045
- Balestra, D., Faella, A., Margaritis, P., Cavallari, N., Pagani, F., Bernardi, F., et al. (2014). An engineered U1 small nuclear RNA rescues splicing-defective coagulation F7 gene expression in mice. *J. Thromb. Haemost.* 12, 177–185. doi: 10.1111/jth.12471
- Balestra, D., Barbon, E., Scalet, D., Cavallari, N., Perrone, D., Zanibellato, S., et al. (2015). Regulation of a strong F9 cryptic 5'ss by intrinsic elements and by combination of tailored U1snRNAs with antisense oligonucleotides. *Hum. Mol. Genet.* 24, 4809–4816. doi: 10.1093/hmg/ddv205
- Balestra, D., Scalet, D., Pagani, F., Rogalska, M. E., Mari, R., Bernardi, F., et al. (2016). An exon-specific U1snRNA induces a robust factor IX activity in mice expressing multiple human FIX splicing mutants. *Mol. Ther. - Nucleic Acids* 5, e370. doi: 10.1038/mtna.2016.77
- Balestra, D., and Branchini, A. (2019). Molecular mechanisms and determinants of innovative correction approaches in coagulation factor deficiencies. *Int. J. Mol. Sci.* 20, 3036. doi: 10.3390/ijms20123036
- Balestra, D., Giorgio, D., Bizzotto, M., Fazzari, M., Ben Zeev, B., Pinotti, M., et al. (2019). Splicing mutations impairing CDKL5 expression and activity can be efficiently rescued by U1snRNA-based therapy. *Int. J. Mol. Sci.* 20, 4130. doi: 10.3390/ijms20174130
- Bolton-Maggs, P. H. B., and Pasi, K. J. (2003). Haemophilias A and B. *Lancet* 361, 1801–1809. doi: 10.1016/S0140-6736(03)13405-8
- Burset, M., Seledtsov, I. A., and Solovyev, V. V. (2000). Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* 28, 4364–4375. doi: 10.1093/nar/28.21.4364
- Dal Mas, A., Fortugno, P., Donadon, I., Levati, L., Castiglia, D., and Pagani, F. (2015a). Exon-specific U1s correct SPINK5 exon 11 skipping caused by a synonymous substitution that affects a bifunctional splicing regulatory element. *Hum. Mutat.* 36, 504–512. doi: 10.1002/humu.22762
- Dal Mas, A., Rogalska, M. E., Bussani, E., and Pagani, F. (2015b). Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. *Am. J. Hum. Genet.* 96, 93–103. doi: 10.1016/j.ajhg.2014.12.009
- De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4, 49–60. doi: 10.1002/wrna.1140

AUTHOR CONTRIBUTIONS

All authors contributed significantly to the manuscript. The manuscript was conceived and prepared by DB, MP, and FB. IM performed the capillary electrophoresis analysis and AB, MF, and DB performed the experiments. Overall, manuscript clarity was reviewed by all authors, and all approved its content.

FUNDING

Authors would like to acknowledge the support provided by the Early Career Bayer Haemophilia Awards Program (BHAP 2017, DB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00974/full#supplementary-material>

- Donadon, I., McVey, J. H., Garagiola, I., Branchini, A., Mortarino, M., Peyvandi, F., et al. (2018a). Clustered *F8* missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. *Haematologica* 103, 344–350. doi: 10.3324/haematol.2017.178327
- Donadon, I., Pinotti, M., Rajkowska, K., Pianigiani, G., Barbon, E., Morini, E., et al. (2018b). Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. *Hum. Mol. Genet.* 27, 2466–2476. doi: 10.1093/hmg/ddy151
- Donadon, I., Bussani, E., Riccardi, F., Licastro, D., Romano, G., Pianigiani, G., et al. (2019). Rescue of spinal muscular atrophy mouse models with AAV9-Exon-specific U1 snRNA. *Nucleic Acids Res.* 47, 7618–7632. doi: 10.1093/nar/gkz469
- Ferrarese, M., Testa, M. F., Balestra, D., Bernardi, F., Pinotti, M., and Branchini, A. (2018). Secretion of wild-type factor IX upon readthrough over F9 pre-peptide nonsense mutations causing hemophilia B. *Hum. Mutat.* 39, 702–708. doi:10.1002/humu.23404
- Glaus, E., Schmid, F., Da Costa, R., Berger, W., and Neidhardt, J. (2011). Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR Splice defect in patient-derived cells. *Mol. Ther.* 19, 936–941. doi: 10.1038/mt.2011.7
- Hartmann, L., Neveling, K., Borkens, S., Schneider, H., Freund, M., Grassman, E., et al. (2010). Correct mRNA processing at a mutant TT splice donor in FANCC ameliorates the clinical phenotype in patients and is enhanced by delivery of suppressor U1 snRNAs. *Am. J. Hum. Genet.* 87, 480–493. doi: 10.1016/j.ajhg.2010.08.016
- Lin, Y. H., Chu, C. C., Fan, H. F., Wang, P. Y., Cox, M. M., and Li, H. W. (2019). A 5'-to-3' strand exchange polarity is intrinsic to RecA nucleoprotein filaments in the absence of ATP hydrolysis. *Nucleic Acids Res.* 47, 5126–5140. doi: 10.1093/nar/gkz189
- Martínez-Pizarro, A., Dembic, M., Pérez, B., Andresen, B. S., and Desviat, L. R. (2018). Intronic PAH gene mutations cause a splicing defect by a novel mechanism involving U1snRNP binding downstream of the 5' splice site. *PLoS Genet.* 14, 1–17. doi: 10.1371/journal.pgen.1007360
- Pinotti, M., Balestra, D., Rizzotto, L., Maestri, I., Pagani, F., and Bernardi, F. (2009). Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood* 113, 6461–6464. doi: 10.1182/blood-2009-03-207613
- Roca, X., Sachidanandam, R., and Krainer, A. R. (2005). Determinants of the inherent strength of human 5' splice sites. *RNA* 11, 683–698. doi: 10.1261/rna.2040605

- Rogalska, M. E., Tajnik, M., Licastro, D., Bussani, E., Camparini, L., Mattioli, C., et al. (2016). Therapeutic activity of modified U1 core spliceosomal particles. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms11168
- Scalet, D., Balestra, D., Rohban, S., Bovolenta, M., Perrone, D., Bernardi, F., et al. (2017). Exploring splicing-switching molecules for seckel syndrome therapy. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1863, 15–20. doi: 10.1016/j.bbdis.2016.09.011
- Scalet, D., Sacchetto, C., Bernardi, F., Pinotti, M., Van De Graaf, S. F. J., and Balestra, D. (2018). The somatic FAH C.1061C > A change counteracts the frequent FAH c.1062+5G > A mutation and permits U1snRNA-based splicing correction. *J. Hum. Genet.* 63, 683–686. doi: 10.1038/s10038-018-0427-x
- Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-causing variants of the conserved +2T of 5' splice sites can be rescued by engineered U1snRNAs. *Hum. Mutat.* 40, 48–52. doi: 10.1002/humu.23680
- Schmid, F., Glaus, E., Barthelmes, D., Fliegau, M., Gaspar, H., Nürnberg, G., et al. (2011). U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. *Hum. Mutat.* 32, 815–824. doi: 10.1002/humu.21509
- Tajnik, M., Rogalska, M. E., Bussani, E., Barbon, E., Balestra, D., Pinotti, M., et al. (2016). Molecular basis and therapeutic strategies to rescue factor IX variants that affect splicing and protein function. *PLoS Genet.* 12, 1–16. doi: 10.1371/journal.pgen.1006082
- van der Woerd, W. L., Mulder, J., Pagani, F., Beuers, U., Houwen, R. H. J., and van de Graaf, S. F. J. (2015). Analysis of aberrant pre-messenger RNA splicing resulting from mutations in ATP8B1 and efficient in vitro rescue by adapted U1 small nuclear RNA. *Hepatology* 61, 1382–1391. doi:10.1002/hep.27620

Conflict of Interest: MP is inventor of a patent (PCT/IB2011/054573) on modified U1snRNAs.

The remaining 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 Balestra, Maestri, Branchini, Ferrarese, Bernardi and Pinotti. 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.



To Be or Not to Be: Circular RNAs or mRNAs From Circular DNAs?

Leire Iparraguirre^{1†}, Iñigo Prada-Luengo^{2†}, Birgitte Regenberg² and David Otaegui^{1*}

¹ Neurosciences Area, Biodonostia Health Research Institute, San Sebastián, Spain, ² Department of Biology, University of Copenhagen, Copenhagen, Denmark

OPEN ACCESS

Edited by:

Rosanna Asselta,
Humanitas University,
Italy

Reviewed by:

Argyris Papanonis,
University Medical Center
Göttingen, Germany
Tobias Jakobi,
Heidelberg University, Germany

*Correspondence:

David Otaegui
david.otaegui@biodonostia.org

[†]These authors share first authorship

Specialty section:

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

Received: 16 May 2019

Accepted: 05 September 2019

Published: 11 October 2019

Citation:

Iparraguirre L, Prada-Luengo I,
Regenberg B and Otaegui D (2019)
To Be or Not to Be: Circular RNAs or
mRNAs From Circular DNAs?
Front. Genet. 10:940.
doi: 10.3389/fgene.2019.00940

In recent years, there has been a growing interest in circular RNAs (circRNAs) since they are involved in a wide spectrum of cellular functions that might have a large impact on phenotype and disease. CircRNAs are mainly recorded by RNA-Seq and computational methods focused on the detection of back-splicing junction sequences considered the diagnostic feature of circRNAs. While some protocols remove linear RNA prior to sequencing, many have characterized circRNAs by sorting through total RNA sequencing data without excluding the possibility that some linear RNA can provide the same signal as a circRNA. Recent studies have revealed that circular DNAs of chromosomal origin are common in eukaryotic genomes and that they can be transcribed. Transcription events across the junction of circular DNAs would result in a transcript with a junction similar to those present in circRNAs. Therefore, in this report, we want to draw attention to transcripts from such circular DNAs both as an interesting new player in the transcriptome and also as a confounding factor that must be taken into account when studying circRNAs.

Keywords: circular RNA, mRNAs, NGS, back-splicing junctions, Circular DNAs

SIGNIFICANCE

There is a growing interest in circRNAs due to their implication in many biological processes and diseases in addition to their biomarker potential. They are mainly detected by the presence of reads mapping their backsplicing junction. Nevertheless, circRNAs are no longer the only transcripts containing such a junction since recent studies have revealed that circular DNAs are common and can be transcribed resulting in transcripts that would mimic a circRNA signal. Therefore, this new type of chimeric transcript can change the way in which circRNA analysis is being done and impact some of the results already reported.

ARE CIRCULAR RNAs THE ONLY CHIMERIC TRANSCRIPTS?

Circular RNA (circRNAs) were rediscovered a few years ago as non-canonically spliced RNA forms present in different organisms including humans (Salzman et al., 2012; Jeck et al., 2013). They are covalently closed transcripts formed through an RNA back-splicing event, where a splice donor of a downstream exon joins to an upstream splice acceptor leading to covalently closed transcripts that are characterized by the presence of a back-splicing junction that makes circRNAs distinguishable from their linear counterparts (**Figure 1A**) (Zhang et al., 2016; Wilusz, 2018).

Since their rediscovery, the scientific community has drawn its attention to circRNAs and has investigated their involvement in several cellular processes in health and disease (Haque

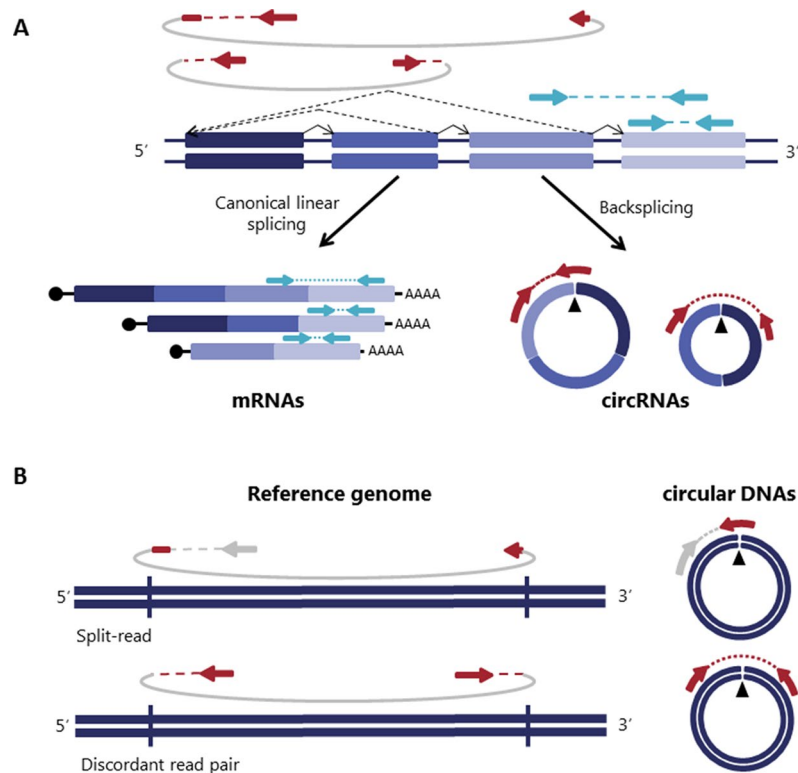


FIGURE 1 | Detection of both circular RNA and DNAs is based on reads spanning the junction. **(A)** Backspliced joining between exons and canonical exon joining are depicted with non-continuous and continuous lines respectively giving rise to circRNAs and mRNAs. Paired-end reads spanning the backspliced junctions are shown in red and paired-end reads consistent with a linear transcript are shown in blue. **(B)** Similar to circRNAs, circular DNAs (eccDNAs) are detected based on structural-read variants consistent with a circularization event depicted in red [adapted from (Møller et al., 2018)].

and Harries, 2017), their potential role as biomarkers (Abu and Jamal, 2016), and their regulatory functions (Floris et al., 2016). CircRNAs are now known to be abundant and stable in the cytosol and the nucleus (Salzman et al., 2012; Jeck et al., 2013; Li et al., 2015) and have also been found free in biofluids (Bahn et al., 2015; Memczak et al., 2015; Chen et al., 2018) and in extracellular vesicles (Kyoung Mi et al., 2017). The biomarker potential of circRNAs has been intensely studied, in fact, there have been published many case-control studies seeking for differentially expressed circRNAs that could be biomarkers of different diseases. To date, circRNAs have been implicated in several diseases including cancer (Kristensen et al., 2017; Arnaiz et al., 2018), neurological disorders (Akhter, 2018), cardiovascular diseases (Aufiero et al., 2019) and immune-related diseases (Iparraguirre et al., 2017; Liu et al., 2019). At the same time, getting to fully understand their biogenesis, characteristics, functions, and implications in human biology remain as open questions for researchers in the field.

Although the function of most of the circRNAs remains unknown, it has been shown that some circRNAs can act as microRNA sponges, regulating the microRNA levels and their activity (Hansen et al., 2013; Memczak et al., 2013; Zheng et al., 2016). They are involved in gene expression regulation by regulating the transcription of their parental

genes, competing with the linear splicing or sponging proteins (Ashwal-Fluss et al., 2014; Li et al., 2018). Interestingly, ribosome profiling studies have recently shown that circRNAs can also be translated both *in vitro* and *in vivo* (Legnini et al., 2017; Yang et al., 2017).

The main feature of circRNAs, and responsible for most of their special properties, is their circularity. Therefore, besides detecting their characteristic back-spliced junction, testing the circularity of these molecules, is one of the key points in every circRNA study. Nevertheless, many studies have based their discovery of circRNAs on total RNA and might thereby have interpreted some linear chimeric transcripts as circRNAs, resulting in false positive circRNA detections. To circumvent this problem most studies have confirmed the circularity of the transcripts found by total RNA-seq using RNase R, Northern blot or electrophoretic methods (Jeck and Sharpless, 2014). However, these circularity validations have also sometimes revealed transcripts that seem to be linear, rather than circular confirming that the detection of circRNAs starting from total RNA can lead to some false positives. These false positives have been attributed to technical artifacts or transcripts derived from uncommon events such as exon duplications or transplicing events (Jeck and Sharpless, 2014; Szabo and Salzman, 2016). That said, the option of having found true, biologically active and functional linear transcripts that contain a sequence

equivalent to a backsplicing junction (from now on called chimeric linear transcripts), has been somewhat overlooked because a source of such linear RNA has not been known for healthy cells.

CIRCULAR DNAs AS A SOURCE OF CHIMERIC LINEAR TRANSCRIPTS

Most of the human genome is organized in linear chromosomes, however, some exceptions have long been accepted such as mitochondrial DNA, and chromosomal aberrations such as DNA circles carrying oncogenes (e.g. double minutes) (Benner et al., 1991; Nathanson et al., 2014; Turner et al., 2017) and ring chromosomes (Tümer et al., 2004). It was not until recently that different circular DNAs such as microDNAs (Shibata et al., 2012) or extrachromosomal circular DNAs (eccDNAs) were found to also arise from large parts of different eukaryotic genomes including human and yeast (Møller et al., 2015; Kumar et al., 2017; Møller et al., 2018).

Circular DNAs are formed when two ends of a linear DNA are joint together resulting in a junction similar to the backspliced junction on circRNAs commonly called breakpoint junction that is detected based on structural-read variants consistent with a circularization event (Gresham et al., 2010; Møller et al., 2018; Prada-Luengo et al., 2019) (Figure 1B). They usually range from a hundred bases to megabase circles and can contain full exons and genes (Shibata et al., 2012; Møller et al., 2015; Kumar et al., 2017; Turner et al., 2017; Møller et al., 2018) and while some regions of the genome are more commonly found on circular DNA (Sinclair and Guarente, 1997; ; Møller et al., 2016; Turner et al., 2017; Møller et al., 2018), most circular DNA appear to

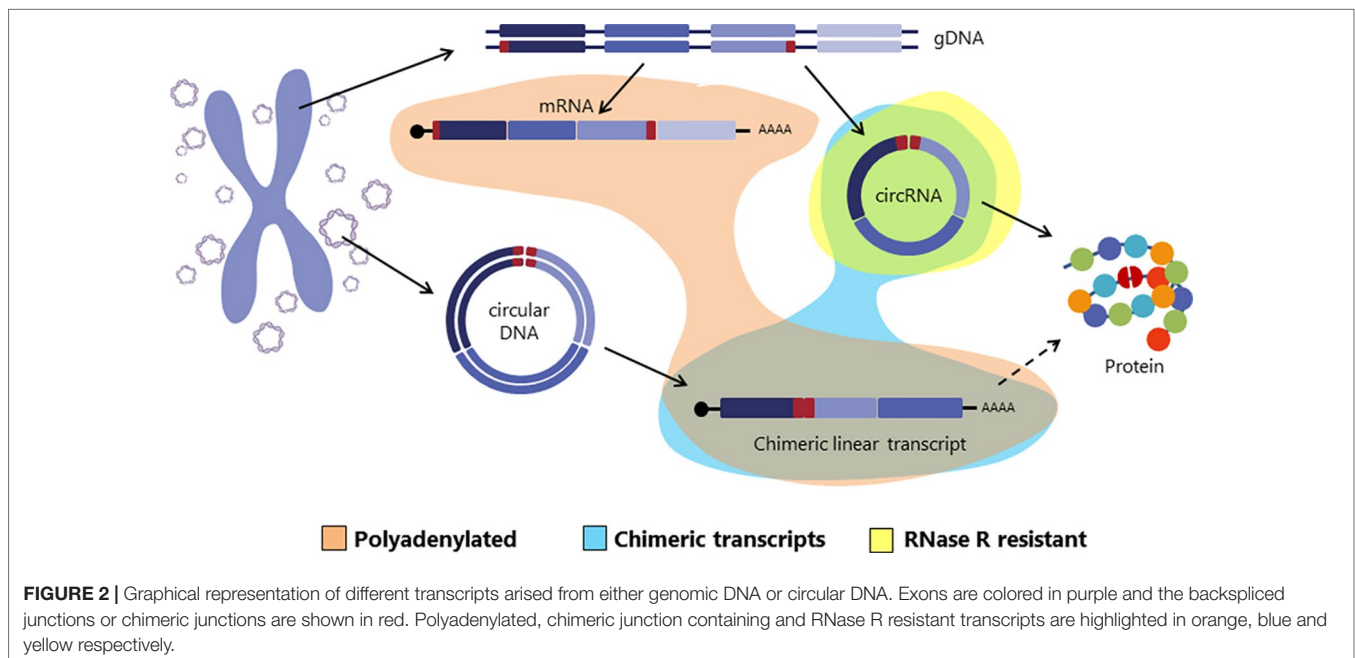
occur at random (Shibata et al., 2012; Møller et al., 2015; Kumar et al., 2017; Møller et al., 2018).

Interestingly, in a recent paper Møller et al. identified thousands of eccDNAs in leucocytes and muscle cells in healthy controls. With the idea of investigating whether eccDNAs could be transcribed, an mRNA library was also sequenced from muscle tissue and analyzed for transcription events across the breakpoint junction of the detected eccDNA finding several matches (Møller et al., 2018). This finding suggests that circular DNA in healthy tissue is transcribed, giving rise to linear and polyadenylated transcripts that will carry a sequence equivalent to the backsplicing sequence of circRNAs (Møller et al., 2018) (Figure 2).

The transcriptional evidence of circular DNAs, together with their abundance, lead us to suggest that circular DNAs could be a natural source of a substantial amount of linear RNAs carrying chimeric junctions. In many cases, these chimeric junctions might be indistinguishable from the backsplicing junctions of circRNAs, and therefore they might be confounding factors in circRNA studies. In the following paragraphs, we will explain the data supporting this proposal.

circRNA DETECTION: ALL THAT GLITTERS IS NOT GOLD

As previously introduced, circRNAs are formed through a non-canonical splicing event called backsplicing. Transcripts resulting from this backsplicing event have a covalently closed loop structure with neither 5'–3' polarity, nor a polyadenylated tail and more importantly, they are characterized by the presence of a scrambled exon order relative to the linear transcript (Zhang et al., 2016; Wilusz, 2018). This scrambled exon order becomes



evident in the backspliced junction that connects a 5' downstream sequence with an upstream 3' sequence. Thus, all the circRNA detection algorithms exploit the presence of the back-spliced junctions as a diagnostic feature for the identification of circRNA (Figure 1A).

Different methods have been adapted for the detection of these back-spliced junctions. Commercial arrays containing probes targeting these backspliced regions have been widely used in biomarker screening studies (Iparraguirre et al., 2017; Liu et al., 2017; Sui et al., 2017; Li et al., 2018). The subsequent validation is often also based on the amplification of the backsplicing junctions using divergent primers (Panda and Gorospe, 2018). Many other papers have conducted a high throughput sequencing analysis that overcomes one of the main limitations of the arrays allowing to detect not only the annotated circRNAs but also *de novo* RNA circularization events from genomic regions where no circRNA were annotated by previous studies. Several bioinformatic pipelines have been developed for the detection of circRNAs in RNA-Seq datasets, but all of them are based on the presence of reads crossing over the back-splicing junctions and finding the most reliable one is still a challenge for bioinformaticians (Hansen et al., 2016; Hansen, 2018; Prada-Luengo et al., 2019).

Two main approaches can be followed for the detection of circRNAs in RNA-Seq data. Firstly, many circRNA RNA-seq studies are based on RNase R treated samples in order to deplete all the linear RNAs before sequencing. Although this approach is specially designed for the circRNA detection it is worth noting that RNase R degradation is variable, that there are rare cases of RNase R resistant linear RNAs and RNase R sensitive circRNAs (Szabo and Salzman, 2016) and that this treatment requires a high RNA input which could be limiting for some tissues. Other circRNA studies choose to sequence either total, ribosomal-depleted (ribo-), or non-polyadenylated (polyA-) RNA, where both linear and circular RNAs can be found (Salzman et al., 2012; Memczak et al., 2013; Broadbent et al., 2015; Lu et al., 2015; Memczak et al., 2015). This approach avoids the use of RNase R, which reduces the RNA amount needed for the sequencing and allows studying the expression of other types of RNAs from the same dataset. It has been demonstrated that with a good sequencing depth and quality and a carefully data-analysis, true circRNAs can be detected from total RNA sequencing (Wang et al., 2017), however, in this second approach, a later circularity confirmation for circRNAs is needed.

With the discovery of linear chimeric RNAs transcribed from circular DNAs, circRNAs are no longer the only transcripts with chimeric junctions. Therefore, it is of utmost importance to note that whereas the first approach will significantly enrich the RNA sample in circRNAs so that most of the detected chimeric junctions will correspond to true circRNAs, the second one might overestimate the number of circRNA transcripts by attributing to circRNAs the signal coming from both circRNAs and the linear chimeric transcripts transcribed from circular DNAs. Consequently, taking into account the coexistence of circRNAs and linear chimeric transcripts, the need of circularity tests and functionality assays gains importance and special care should be taken regarding not

only experimental but also computational methods to avoid mistaking chimeric transcripts from circular DNAs with circRNAs formed by backsplicing.

DISCUSSION

The circRNA field is still at an early stage, however, circRNAs have already shown to be astonishing molecules, implicated in many processes, with a great biomarker potential and that can also change the way we understand the transcription and translation processes. For these reasons, they are gaining attention and the circRNA field is at the moment one of the most active RNA research fields. However, there are still many conflicts, controversies and open questions (Li, 2019) that have to be discussed.

In this report, and in light of the recent advances in the circular DNA field, we want to point out the transcription from extrachromosomal circular DNA as one of the main natural sources of linear transcripts with back-spliced signals that could be interfering with circRNA data (Møller et al., 2018). From now on, apart from the technical artifacts, duplications and transplicing events that could lead to false positives in the circRNA detection, we should take also into account the existence of this new type of chimeric transcripts. Therefore circularization tests and functional assays are more important than ever.

In any case, these chimeric linear transcripts should not only be considered as a mere confounding factor for circRNA studies. Despite the technical implications for the circRNA characterization, the existence of these circRNA-like chimeric linear RNA molecules coming from eccDNAs adds a new type of molecule to the ever-growing list of RNAs and expands our vision about the complexity of the transcriptome and its regulation. Moreover, these linear RNA molecules coming from eccDNA could also present functions similar to the circRNA, including the regulatory functions or the potential to be translated. Gene products from eccDNA transcripts could potentially contribute to the phenotype of somatic cells and tissue as reported in yeast (Gresham et al., 2010; Demeke et al., 2015). However in this nascent field, more data and research is needed to start scratching the surface of the iceberg.

AUTHOR CONTRIBUTIONS

LI, IP-L, BR, and DO wrote the paper.

FUNDING

This study has been funded by Instituto de Salud Carlos III through the project "PI17/00189" (Co-funded by European Regional Development Fund/European Social Fund) "Investing in your future"). IP-L and BR were supported by the Danish Council for Independent Research, 6108-00171B and LI was supported by the Department of Education of the Basque Government [grant number PRE_2018_2_0081].

REFERENCES

- Abu, N., and Jamal, R. (2016). Circular RNAs as promising biomarkers: a mini-review. *Front. Physiol.* 7 (AUG), 1–6. doi: 10.3389/fphys.2016.00355
- Akhter, R. (2018). Circular RNA and Alzheimer's disease. *Adv. Exp. Med. Biol.* 1087, 239–243. doi: 10.1007/978-981-13-1426-1_19
- Arnaiz, E., Sole, C., Manterola, L., Iparraguirre, L., Otaegui, D., and Lawrie, C. (2018). CircRNAs and cancer: 170 Biomarkers and master regulators. *Semin. Cancer Biol.* 58, 90–99. doi: 10.1016/j.semcancer.2018.12.002
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). CircRNA biogenesis competes with Pre-mRNA splicing. *Mol. Cell* 56 (1), 55–66. doi: 10.1016/j.molcel.2014.08.019
- Aufiero, S., Reckman, Y. J., Pinto, Y. M., and Creemers, E. E. (2019). Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* 16 (8), 503–514. doi: 10.1038/s41569-019-0185-2
- Bahn, J. H., Zhang, Q., Li, F., Chan, T.-M., Lin, X., Yong, K., et al. (2015). The landscape of MicroRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.* 61 (1), 221–230. doi: 10.1373/clinchem.2014.230433
- Benner, S. E., Wahl, G. M., and Von Hoff, D. D. (1991). Double minute chromosomes and homogeneously staining regions in tumors taken directly from patients versus in human tumor cell lines. *Anticancer Drugs* 2 (1), 11–25. doi: 10.1097/00001813-199102000-00002
- Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., and Sabeti, P. C. (2015). Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-183 coding RNA and circular RNA. *BMC Genomics*, 1–22. doi: 10.1186/s12864-015-1603-4
- Chen, D., Zhang, L., Tan, K., and Jing, Q. (2018). Application of droplet digital PCR in quantitative detection of the cell-free circulating circRNAs. doi: 10.1080/13102818.2017.1398596
- Demeke, M. M., Foulquié-Moreno, M. R., Dumortier, F., and Thevelein, J. M. (2015). Rapid evolution of recombinant *Saccharomyces cerevisiae* for xylose fermentation through formation of extra-chromosomal circular DNA. *PLoS Genet.* 11 (3), e1005010. doi: 10.1371/journal.pgen.1005010
- Floris, G., Zhang, L., Follasa, P., and Sun, T. (2016). Regulatory Role of Circular RNAs and Neurological Disorders. *Mol. Neurobiol.* 54, 5156–5165. doi: 10.1007/s12035-016-0055-4
- Gresham, D., Usaite, R., Germann, S. M., Lisby, M., Botstein, D., and Regenberg, B. (2010). Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. *PNAS* 107 (43), 18551–18556. doi: 10.1073/pnas.1014023107
- Hansen, T. B. (2018). Improved circRNA identification by combining prediction algorithms. *Front. Cell Dev. Biol.* 6 (March), 1–9. doi: 10.3389/fcell.2018.00020
- 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 (7441), 384–388. doi: 10.1038/nature11993
- Hansen, T. B., Venø, M. T., Damgaard, C. K., and Kjems, J. (2016). Comparison of circular RNA prediction tools. *Nucleic Acids Res.* 44 (6). doi: 10.1093/nar/gkv1458
- Haque, S., and Harries, L. W. (2017). Circular RNAs (circRNAs) in health and disease. *Genes (Basel)* 8 (353), 1–17. doi: 10.3390/genes8120353
- Iparraguirre, L., Muñoz-Culla, M., Prada-Luengo, I., Castillo-Triviño, T., Olascoaga, J., and Otaegui, D. (2017). Circular RNA profiling reveals that circular RNAs from ANXA2 can be used as new biomarkers for multiple sclerosis. *Hum. Mol. Genet.* 26 (18), 3564–3572. doi: 10.1093/hmg/ddx243
- Jeck, W. R., and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32 (5), 453–461. doi: 10.1038/nbt.2890
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19 (2), 141–157. doi: 10.1261/rna.035667.112
- Kristensen, L. S., Hansen, T. B., Venø, M. T., and Kjems, J. (2017). Circular RNAs in cancer : opportunities and challenges in the field. *Nat. Publ. Gr.* 37 (5), 555–565. doi: 10.1038/onc.2017.361
- Kumar, P., Dillon, L. W., Shibata, Y., Jazaeri, A. A., Jones, D. R., and Dutta, A. (2017). Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol. Cancer Res.* 15 (9), 1197–1205. doi: 10.1158/1541-7786.MCR-17-0095
- Kyoung Mi, K., Abdelmohsen, K., Mustapic, M., Kapogiannis, D., and Gorospe, M. (2017). RNA in extracellular vesicles. *WIREs RNA* 8 (4), 1–14. doi: 10.1002/wrna.1413
- Legnini, I., Timoteo, G. Di, Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66 (1), 22–37.e9. doi: 10.1016/j.molcel.2017.02.017
- Li, H. (2019). Intriguing circles : conflicts and controversies in circular RNA. *WIREs RNA* 10 (e1538), 1–19. doi: 10.1002/wrna.1538
- Li, S., Teng, S., Xu, J., Su, G., Zhang, Y., Zhao, J., et al. (2018a). Microarray is an efficient tool for circRNA profiling. *Brief. Bioinform.* 1–12. doi: 10.1093/bib/bby006
- Li, X., Yang, L., and Chen, L. (2018). The biogenesis , functions , and challenges of circular RNAs. *Mol. Cell* 71 (3), 428–442. doi: 10.1016/j.molcel.2018.06.034
- Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., et al. (2015). Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 25 (8), 981–984. doi: 10.1038/cr.2015.82
- Liu, C., Li, X., Nan, F., Shen, N., Yang, L., Chen, L., et al. (2019). Structure and degradation of circular RNAs regulate PKR activation in innate immunity article structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* 177, 1–16. doi: 10.1016/j.cell.2019.03.046
- Liu, W., Zhang, J., Zou, C., Xie, X., Wang, Y., Wang, B., et al. (2017). Microarray expression profile and functional analysis of circular RNAs in osteosarcoma. *Cell Physiol. Biochem.* 43 (3), 969–985. doi: 10.1159/000481650
- Lu, T., Cui, L., Zhou, Y. A. N., Zhu, C., Fan, D., Gong, H. A. O., et al. (2015). Transcriptome-wide investigation of circular RNAs in rice. *RNA* 21 (12), 2076–2087. doi: 10.1261/rna.052282.115
- 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 (7441), 333–338. doi: 10.1038/nature11928
- Memczak, S., Papavasileiou, P., Peters, O., and Rajewsky, N. (2015). Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* 10 (10), 1–13. doi: 10.1371/journal.pone.0141214
- Møller, H. D., Bojsen, R. K., Tachibana, C., Parsons, L., Botstein, D., and Regenberg, B. (2016). Genome-wide Purification of Extrachromosomal Circular DNA from Eukaryotic Cells. *J. Vis. Exp.* e54239, 240 doi: 10.3791/54239
- Møller, H. D., Mohiyuddin, M., Prada-luengo, I., Sailani, M. R., Halling, J. F., Plomgaard, P., et al. (2018). Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* 9, 1069, 1–12. doi: 10.1038/s41467-018-03369-8
- Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D., and Regenberg, B. (2015). Extrachromosomal circular DNA is common in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 112 (24), E3114–E3122. doi: 10.1073/pnas.1508825112
- Nathanson, D. A., Gini, B., Mottahedeh, J., Visnyei, K., Koga, T., Gomez, G., et al. (2014). Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Sci. Rep.* 343, 72–76. doi: 10.1126/science.1241328
- Panda, A. C., and Gorospe, M. (2018). Detection and Analysis of Circular RNAs by RT-PCR. 8. doi: 10.21769/BioProtoc.2775.Detection
- Prada-Luengo, I., Krogh, A., Maretty, L., and Regenberg, B. (2019). Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *bioRxiv*, 654194. doi: 10.1101/654194
- Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7 (2). doi: 10.1371/journal.pone.0030733
- Shibata, Y., Kumar, P., Layer, R., Willcox, S., Gagan, J. R., Griffith, J. D., et al. (2012). Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* (80-) 336 (6077), 82–86. doi: 10.1126/science.1213307
- Sinclair, D. A., and Guarente, L. (1997). Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* 91 (7), 1033–1042. doi: 10.1016/S0092-8674(00)80493-6
- Sui, W., Shi, Z., Xue, W., Ou, M., Zhu, Y., Chen, J., et al. (2017). Circular RNA and gene expression profiles in gastric cancer based on microarray chip technology. *Oncol. Rep.* 37 (3), 1804–1814. doi: 10.3892/or.2017.5415

- Szabo, L., and Salzman, J. (2016). Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* 17 (11), 679–692. doi: 10.1038/nrg.2016.114
- Tümer, Z., Harboe, T. L., Blennow, E., Kalscheuer, V. M., Tommerup, N., and Brøndum-Nielsen, K. (2004). Molecular cytogenetic characterization of ring chromosome 15 in three unrelated patients. *Am. J. Med. Genet.* 344, 340–344. doi: 10.1002/ajmg.a.30035
- Turner, K. M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nat. Publ. Gr.* 543 (7643), 122–125. doi: 10.1038/nature21356
- Wang, J., Liu, K., Liu, Y., Lv, Q., Zhang, F., and Wang, H. (2017). Evaluating the bias of circRNA predictions from total RNA-Seq data. *Oncotarget* 8 (67), 110914–110921. doi: 10.18632/oncotarget.22972
- Wilusz, J. E. (2018). A 360 view of circular RNAs : from biogenesis to functions. *WIREs RNA*, 1–17. doi: 10.1002/wrna.1478
- Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by. *Cell Res.* 27 (5), 626–641. doi: 10.1038/cr.2017.31
- Zhang, Y., Shao, C., Li, C., Zhao, Z., and Xie, J. (2016). [Research progress of circRNA and its significance in forensic science]. *Fa Yi Xue Za Zhi* 32 (2), 131–133.
- Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., et al. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 7 (6), 1–13. doi: 10.1038/ncomms11215

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

Copyright © 2019 Iparraguirre, Prada-Luengo, Regenber and Otaegui. 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.



Validation and Classification of Atypical Splicing Variants Associated With Osteogenesis Imperfecta

Lulu Li^{1†}, Yixuan Cao^{1†}, Feiyue Zhao¹, Bin Mao¹, Xiuzhi Ren², Yanzhou Wang³, Yun Guan⁴, Yi You¹, Shan Li¹, Tao Yang¹ and Xiuli Zhao^{1*}

¹ Department of Medical Genetics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & School of Basic Medicine, Peking Union Medical College, Beijing, China, ² Department of Orthopaedics, The People's Hospital of Wuqing District, Tianjin, China, ³ Department of Pediatric Orthopaedics, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China, ⁴ Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University, School of Medicine, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Eladio Andrés Velasco,
Institute of Biology and Molecular
Genetics (IBGM), Spain

Reviewed by:

Eugenia Fraile-Bethencourt,
Oregon Health & Science University,
United States

Andrés Fernando Muro,
International Centre for Genetic
Engineering and Biotechnology, Italy

*Correspondence:

Xiuli Zhao
xiulizhao@ibms.pumc.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 06 June 2019

Accepted: 13 September 2019

Published: 18 October 2019

Citation:

Li L, Cao Y, Zhao F, Mao B, Ren X,
Wang Y, Guan Y, You Y, Li S, Yang T
and Zhao X (2019) Validation and
Classification of Atypical Splicing
Variants Associated With
Osteogenesis Imperfecta.
Front. Genet. 10:979.
doi: 10.3389/fgene.2019.00979

Osteogenesis Imperfecta (OI) is a rare inherited bone dysplasia, which is mainly caused by mutations in genes encoding type I collagen including *COL1A1* and *COL1A2*. It has been well established to identify the classical variants as well as consensus splicing-site-variants in these genes in our previous studies. However, how atypical variants affect splicing in OI patients remains unclear. From a cohort of 867 OI patients, we collected blood samples from 34 probands which contain 29 variants that are located close to splice donor/acceptor sites in either *COL1A1* or *COL1A2*. By conducting minigene assay and sequencing analysis, we found that 17 out of 29 variants led to aberrant splicing effects, while no remarkable aberrant splicing effect was observed in the remaining 12 variants. Among the 17 variants that affect splicing, 14 variants led to single splicing influence: 9 led to exon skipping, 2 resulted in truncated exon, and 3 caused intron retention. There were three complicated cases showing more than one mutant transcript caused by recognition of several different splice sites. This functional study expands our knowledge of atypical splicing variants, and emphasizes the importance of clarifying the splicing effect for variants near exon/intron boundaries in OI.

Keywords: osteogenesis imperfecta, *COL1A1*, *COL1A2*, minigene splicing assay, atypical splicing variants

INTRODUCTION

Osteogenesis imperfecta (OI), also known as brittle bone disease, is an inherited skeletal dysplasia characterized by frequent fractures, blue sclerae, bone deformity, and relaxation of skin and ligament. OI is considered as a rare bone disease and its prevalence is reported to be 1 in 15,000 live births (Stoll et al., 1989). Based on phenotypes, patients with OI can be categorized into 4 types according to Sillence et al. (1979): patients with the mildest phenotype and with blue sclerae (type I); lethal (type II); the severe form with progressively skeletal deformity (type III); moderate OI with variable bone deformity (type IV). Recently types V–XIX OI were grouped according to genetic and clinical characteristics (Rauch and Glorieux, 2004; Forlino and Marini, 2016; Lindert et al., 2016; Marini et al., 2017).

OI is mainly caused by abnormal structure and quantity of type I collagen, functioning as the main matrix in bone tissue. Type I collagen is encoded by *COL1A1* (MIM# 120150) and *COL1A2* (MIM#

120160) (Marini et al., 2007). Mutations in other collagen-related genes have been reported to contribute to OI development as well, including *IFITM5*, *SERPINF1*, *CRTAP*, *P3H1*, *PIIB*, *SERPINH1*, *FKBP10*, *PLOD2*, *BMP1*, *SP7*, *TMEM38B*, *WNT1*, *CREB3L1*, *SPARC*, and *MBTPS2* (Byers and Pyott, 2012; Rohrbach and Giunta, 2012; Lindert et al., 2016; Gagliardi et al., 2017). Nevertheless, around 90% of OI are autosomal dominant inheritance with a familial history and are caused by mutations in *COL1A1* and *COL1A2*.

Typical mutation spectrum of OI includes missense, nonsense, frameshift, and splice site mutations. Despite of these classical mutations, it was shown that a large portion of DNA variants disrupted splicing in cancer-related diseases (Sanz et al., 2010). However, it was rarely reported whether similar DNA variants have an impact on aberrant splicing in OI patients. RNA splicing is essential for transcription processing and for the correct protein synthesis. Human genes undergo alternative splicing therefore different transcripts can be generated (Johnson et al., 2003). The process of splicing initiates from recognition of core splicing signal, including splicing donor (gt), splicing acceptor (ag) and a branch point (Wang and Burge, 2008). The splicing process is catalyzed by the spliceosome, which contains five uridine rich ribonucleoproteins (U1, U2, U4, U5, and U6) and more than 200 associated proteins (Zhou et al., 2002). During the splicing process, a cryptic splice site may be activated due to the variants and generate aberrant splicing products (Sun and Chasin, 2000). Therefore, studying the splicing effects caused by the variants is important for understanding the pathogenesis and molecular mechanisms of OI.

Because of the very low expression levels of *COL1A1*/*COL1A2* in peripheral blood, RNAs from the tissue of OI patients would be ideal for examining whether the variants can affect RNA splicing. However, the availability of the tissue of OI patients is limited. Therefore, a minigene assay, which is based on patients' genomic DNA, represents a valid and powerful approach to study the splicing pattern (Cooper, 2005; Ahlborn et al., 2015; Fraile-Bethencourt et al., 2019).

It has been reported that variants at splicing sites can drive to splicing effects in some OI patients (Schleit et al., 2015; Schwarze et al., 1999). However, most of these variants were typical splicing variants which were located at splicing donor/acceptor sites in introns. A recent study reported splicing effects in 40 OI patients harboring the variants in introns (Schleit et al., 2015). Although the pathogenicity of variants at splicing sites has been well studied, atypical splicing sites beyond the splicing sites (GT-AG) were rarely reported. To determine whether a variant has an impact on splicing efficiency, we selected 34 OI probands carrying 29 different variants which were located close to the splicing sites in introns or exons of *COL1A1* or *COL1A2*. Based on minigene assays and sequence analysis, 17 variants showed aberrant splicing effects while 12 variants presented no splicing consequences. The aberrant splicing was further classified into 3 patterns: exon skipping, truncated exon/intron retention resulted from recognition of alternative splice sites and compound aberrant splicing. Current findings enriched the splicing patterns, and

suggested that atypical splicing variants may represent a large group of pathogenic mutations of OI.

METHODS AND MATERIALS

Variant Nomenclature

The variants of *COL1A1* and *COL1A2* were named according to variant nomenclature provided by Human Genome Variation Society (<http://www.hgvs.org/munomen>). The genomic DNA and cDNA sequences of *COL1A1* (NC_000017.11) and *COL1A2* (NC_000007.14) were obtained from National Center for Biotechnology Information (NCBI) reference sequence and University of California, Santa Cruz (UCSC) Genome browser database (<http://genome.ucsc.edu/>). The altered proteins were named based on the sequencing of mutant transcripts.

Subjects

A total number of 867 patients (from 489 families) diagnosed as OI were recruited for this study from 2014 to 2018. Information of their phenotypes, including number of fractures, blue sclerae, affected skeletal location, and bone deformity were recorded after obtaining patients' informed consent. Tissue samples, including peripheral blood and/or skin, were collected to detect the variants. After sequence analysis, 34 probands from different families carrying *COL1A1* or *COL1A2* variants close to the exon/intron boundaries were enrolled for minigene splicing assay. All variants identified in this study have been submitted to the Osteogenesis Imperfecta Variant Database (<http://oi.gene.le.ac.uk/>).

In Silico Analysis

Online software ESE Finder 3.0 and Human Splicing Finder (version 3.1) were used to predict the splicing effect of each of the variants. Analysis of ESE Finder was performed to detect exonic splicing enhancers for SR proteins as well as alterations in splice sites. SRProteins matrix library was used to analyze the variants located in exons and SpliceSites matrix library was used for variants in introns. All analyses were performed with default threshold values.

Whole Exome Sequencing (WES)

Genomic DNA was extracted from the peripheral blood, and 1–3 µg genomic DNA was used for WES as described previously (Li et al., 2019). Sequencing was carried out on HiSeq 4000 System (Illumina) as 150 bp paired-end runs after DNA fragmentation, end pair ligation, purification and size distribution assessment. Sequencing analysis was performed using the Pipeline (version 1.3.4; Illumina).

Sanger Sequencing

Sanger sequencing was employed to verify the variants in *COL1A1* and *COL1A2* after WES, and to verify the splicing variants after minigene assay. The process was described previously (You et al.,

2018). Briefly, genomic DNA was isolated using a proteinase K and phenol–chloroform method. Primers were designed by Primer3 (<http://primer3.ut.ee/>). Sequencing was conducted in Applied Biosystems 3730xl DNA Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). Result of Sanger sequencing was analyzed using CodonCode Aligner (version 6.0.2.6; CodonCode, Centerville, MA, USA). The sequence results were aligned to reference sequences *COL1A1* (NC_000017.11) and *COL1A2* (NC_000007.14) and DNA alignment was conducted using DNAMAN (version 6.0, LynnonBiosoft, USA).

Minigene Assay

Twenty-nine variants close to intron–exon boundary in *COL1A1* and *COL1A2* from 34 probands were selected for the minigene splicing assay (Figure 1A). The fragments of interests varying from 808 bp to 2,510 bp (Table S1) which contain the putative splicing variant along with flanking exons were amplified by high fidelity PCR. The PCR was carried out using HS DNA polymerase (TaKaRa, Shiga, Japan) and forward and reverse primers with restriction sites for BamHI or MluI (New England Biolabs, Ipswich, MA, USA). Primers were designed for each target fragment using Primer3 (<http://primer3.ut.ee/>) (Table S1). The amplified target fragments were cloned into the pCAS2 vector

(Figure 1B) using restriction endonucleases BamHI, MluI, and T4 DNA ligase (New England Biolabs). The constructed vector was further transformed into *E. coli* DH5α Competent Cells (TaKaRa, Shiga, Japan), followed by sequencing verification. Both the purified constructs of wild type and mutant type were transferred into HEK293T cells using Invitrogen Lipofectamine 3000 Transfection Kit (Thermo Fisher Scientific). HEK293T cell line was selected to eliminate endogenous interference for its low expression of type I collagen. After 24 h incubation, RNA was isolated using Trizol reagent (Invitrogen). One microgram total RNA was used for RT-PCR using PrimeScript RT reagent kit with gDNA Eraser (TaKaRa). PCR products were separated on 1% agarose gel containing ethidium bromide. The target DNA bands were purified using GeneJET Gel Extraction Kit (Thermoscientific, Lithuania), followed by DNA sequencing with ABI3730xl (Thermo Fisher Scientific, Waltham, MA, USA). The procedure was summarized in the schematic map (Figure 1C).

Fibroblasts Assay

Skin samples were collected from probands PUMC-253, 371, 98, 401, and 216 following the skin biopsy process or surgical operation. Cleaned dermal tissues were cut into small pieces of 1 mm² and washed with PBS. After transferring the dermal pieces

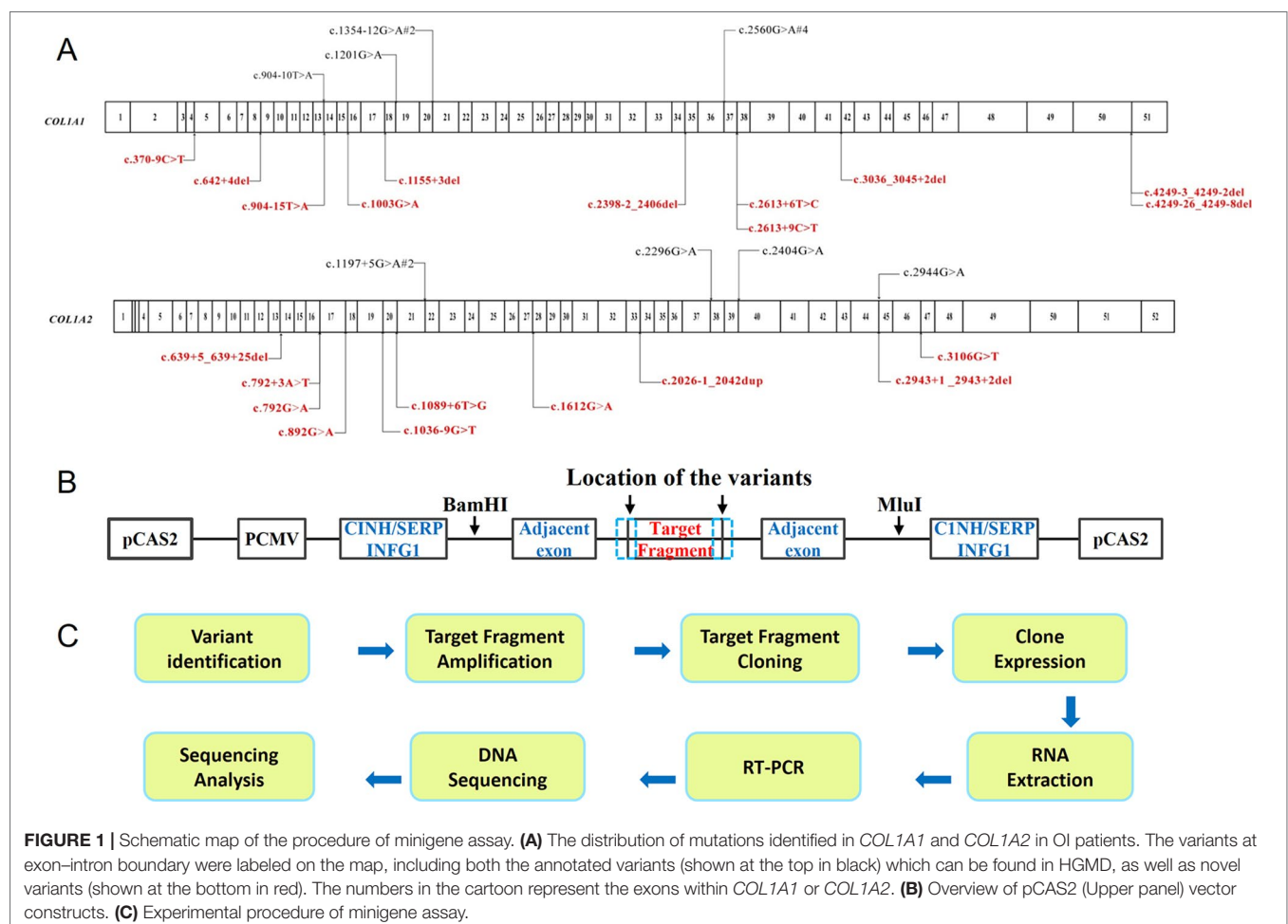


FIGURE 1 | Schematic map of the procedure of minigene assay. **(A)** The distribution of mutations identified in *COL1A1* and *COL1A2* in OI patients. The variants at exon–intron boundary were labeled on the map, including both the annotated variants (shown at the top in black) which can be found in HGMD, as well as novel variants (shown at the bottom in red). The numbers in the cartoon represent the exons within *COL1A1* or *COL1A2*. **(B)** Overview of pCAS2 (Upper panel) vector constructs. **(C)** Experimental procedure of minigene assay.

into a cell culture flask, skin tissue was attached on the flask in humid environment overnight and fibroblasts were cultured in fibroblast culture medium [F12 (Gibco, NY, USA) containing 15% FCS (Gibco, Australia) and 1% antibiotics (Sigma)]. RNA was isolated using Trizol reagent (Invitrogen) when dermal fibroblasts were cultured for 3 passages. After reverse transcription, PCR products were separated on 1% agarose gel followed by sequencing confirmation.

RESULTS

We enrolled a cohort of 867 OI patients and 72 OI patients (from 26 families) carried 22 different classical splicing mutations (with gt/ag mutations) in *COL1A1* and *COL1A2* (Table S2). This research focused on the atypical splicing variants that are located close to intron–exon boundaries, in order to determine whether such variants affect splicing. Details of variants found by whole exome sequencing or Sanger sequencing, expected variant type, actual variant type by minigene analysis, alteration of nucleotide, amino acid change, and the classification of the OI type were shown in Table 1. All 34 probands were germline heterozygotes with variation of *COL1A1* or *COL1A2*, and each cell contained a normal allele and a mutant allele. Minigene assay showed that the normal alleles only formed wild type transcripts. So in the following results the transcripts from the mutant alleles will be mainly clarified.

SPLICING EFFECT ANALYZED BY MINIGENE ASSAY

Among the 34 probands, there were 29 different variants and 17 variants displayed aberrant splicing based on findings in minigene assay and 12 did not show any splicing consequence (Table 1). RT-PCR of RNA extracted from fibroblasts was also conducted for 5 variants (c.642+4delA in *COL1A1*, c.1089+6T > G in *COL1A2*, c.1197+5G > A in *COL1A2*, c.2026-1_2042dup in *COL1A2*, c.792G > A in *COL1A2*) (Table 1), and results from fibroblasts were in line with findings of minigene assay. In general, two main types of single-splicing-effects were categorized: exon skipping (Figure 2A), and alternative splice sites activation (Figures 2B, C). The latter one can be further separated into two subtypes: partial exon deletion resulted from the alternative splice sites in exons (Figure 2B), and intron retention caused by alternative splice sites in introns (Figure 2C). The results from minigene assay were then compared with the predictions made by *in silico* tools: Human Splicing Finder (version 3.1) and ESE Finder 3.0 (Table S3). Both tools only correctly predicted a portion of aberrant splicing, and hence a minigene assay is a solid method to verify the splicing pattern.

Variants Only Led to Exon Skipping in OI Patients

Nine variants in this study were observed with only exon skipping, as indicated by minigene assay (Table 1). These variants include c.1155+3delA, c.2398-2_2406del, c.2613+6T>C in *COL1A1*, and c.639+5_639+25del c.792+3A>T, c.1089+6T>G, c.1197+5G>A,

c.2943+1_2943+2delgt, c.792G>A in *COL1A2*. None of these variants drove to frameshift alterations or premature stop codons.

Eight of these variants with exon skipping effects are located in introns. Notably, the variant, c.792G > A in *COL1A2* (PUMC-371) in the exon 16 displayed the exon skipping effect as well (Figure 3). Generally c.792G > A (p.Lys264Lys) was regarded as a synonymous mutation, but this variant was found at the last nucleotide in exon 16 of *COL1A2*, so we suspect it may affect splicing. Minigene analysis confirmed our conjecture and showed a wild type (Figure 3A lower panel) and a mutant transcript (Figure 3A upper panel) with exon 16 skipping. The schematic splicing map was shown in Figure 3B. To validate the results obtained from the minigene assay, RNA was isolated from skin fibroblasts of the patient, followed by sequencing of RT-PCR products (Figure 3C). The endogenous expression was in agreement with findings from minigene assay.

Partial Exon Deletion Caused by Cryptic Splice Site Activation in Exon

Recognition of Alternative Donor Site in Exon

Variant c.3036_3045+2del in *COL1A1* (PUMC-480) led to the activation of cryptic donor site in the exon (Figure 4). Two different transcripts were found by minigene analysis: a wild type transcript from the normal allele and a mutant transcript with disrupted signal after exon 40 (Figure 4A). After further T clone sequencing, the mutant transcripts were divided into two segments: only exon 41 skipping in transcript 1 (33%), and a partial skipping of exon 41 in transcript 2 (67%). An alternative donor splice site in exon 41 c.3029_3030 GT was recognized, which led to a truncated exon 41 (Figure 4Ab). Variant c.642+4delA in *COL1A1* (PUMC-401) also resulted in the utilization of an alternative donor site (c.617_618GT) and generated truncated exon 8 (Figure 5Ad).

Recognition of Alternative Acceptor Site in Exon

Three variants were found with alternative splicing acceptor site-induced aberrant splicing. Variant c.642+4delA in *COL1A1* (PUMC-401) was observed that an AG site (c.660_661AG) in exon 9 in *COL1A1* was utilized as the splicing acceptor (Figures 5Ab, B). Consequently, a truncated exon 9 was generated. There were two variants c.4249-26_4249-8del in *COL1A1* (PUMC-276) and c.4249-3_4249-2del in *COL1A1* (PUMC-290) which showed the same splicing effects (Figure S1). The minigene results of both variants showed an alternative AG site (c.4395+1147_4395+1148AG) in the UTR sequence, which was used as the 3' splice site, resulted in the deletion of exon 51 and partial of 3' UTR (Figure S1B).

Intron Retention Caused by Alternative Splice Site in Intron

Recognition of Alternative Donor Site in Intron

In proband PUMC-401 (c.642+4delA in *COL1A1*), one mutant transcript with alternative donor site in intron 7 was recognized (Figure 5Ac). The alternative splice site c.589-62_589-61gt, which is located in intron 7, was selected preferentially as donor site during splicing. As a result, part of intron 7 (96bp) was inserted in the mutant transcript.

TABLE 1 | Splicing analysis of the atypical *COL1A1* variants and atypical *COL1A2* variants.

Proband number	Gene	Location	Mutation	Expect effect	Result of Minigene assay				
					Effect observed in Minigene assay	Nucleotide change	Predicted amino acid change	Fibroblasts assay	OI type
Variants in introns									
PUMC-229	COL1A1	Intron 4	c.370-9C > T	Splicing	No aberration	No aberration	No aberration	NA	III
PUMC-401	COL1A1	Intron 8	c.642+4delA	Splicing	Skipping of exon 8/ Skipping of exon 8 and partial exon 9/Retention of intron 7 and skipping of exon 8/26 bp del in exon 8	r.589_642del/r.589_661del/ r.587_588ins588+1_588+96/ r.617-642del	(p.Gly197_Ser214del)/ (p.Gly197Valfs*44)/ (p.Pro196_Gly197ins32)/ (p.Glu207Serfs*17)	Skipping of exon 8/ Retention of intron 7 and skipping of exon 8/26 bp del in exon 8/36 bp del in exon 8/ Retention of intron 8	IV
PUMC-15	COL1A1	Intron 13	c.904-10T > A	Splicing	Retention of 8bp	r.369_370ins370-8_370-1	(p.Gly302Valfs*242)	NA	I
PUMC-105	COL1A1	Intron 13	c.904-15T > A	Splicing	Retention of 13bp	r.903_904ins904-13_904-1	(p.Gly302Valfs*12)	NA	I
PUMC-186	COL1A1	Intron 17	c.1155+3delA	Splicing	Skipping of exon 17	r.1057_1155del	(p.Gly353_Ala385del)	NA	I
PUMC-369	COL1A1	Intron 20	c.1354-12G > A	Splicing	Retention of 10bp	r.903_904ins904-10_904-1	(p.Gly452Profs*26)	NA	I
PUMC-189	COL1A1	Intron 20	c.1354-12G > A	Splicing	Retention of 10bp	r.903_904ins904-10_904-1	(p.Gly452Profs*26)	NA	I
PUMC-111	COL1A1	Intron 34-exon 35	c.2398-2_2406del	Splicing	Skipping of exon 35	r.2398_2451del	(p.Gly800_Pro817del)	NA	IV
PUMC-109	COL1A1	Intron 37	c.2613+6T > C	Splicing	Skipping of exon 37	r.2560_2613del	(p.Gly854_Pro871del)	NA	IV
PUMC-469	COL1A1	Intron 37	c.2613+9C > T	Splicing	No aberration	No aberration	No aberration	NA	III
PUMC-480	COL1A1	Exon 41-intron 41	c.3036_3045+2del	Splicing	Skipping of exon 41/ partial exon 41 del	r.2938_3045del/r.3029_3045del	(p.Gly980_Glu1015del)/ (p.Glu1011Glyfs*19)	NA	IV
PUMC-276	COL1A1	Intron 50	c.4249-26_4249-8del	Splicing	Exon 51_3'UTRdel	r.4249_4395+1148del	(p.Thr1417Ala*64)	NA	IV
PUMC-290	COL1A1	Intron 50	c.4249-3_4249-2del	Splicing	Exon 51_3'UTRdel	r.4249_4395+1148del	(p.Thr1417Ala*64)	NA	IV
PUMC-391	COL1A2	Intron 13	c.639+5_639+25del	Splicing	Skipping of exon 13	r.595_639del	(p.Gly199_Thr213del)	NA	I
PUMC-90	COL1A2	Intron 16	c.792+3A > T	Splicing	Skipping of exon 16	r.739_792del	(p.Gly247_Lys264del)	NA	III
PUMC-224	COL1A2	Intron 19	c.1036 -9G > T	Splicing	No aberration	No aberration	No aberration	NA	IV
PUMC-98	COL1A2	Intron 20	c.1089+6T > G	Splicing	Skipping of exon 20	r.1036_1089del	(p.Gly346_Pro363del)	Skipping of exon 20	I
PUMC-216	COL1A2	Intron 21	c.1197+5G > A	Splicing	Skipping of exon 21	r.1090_1197del	(p.Gly364_Arg399del)	Skipping of exon 21	IV
PUMC-37	COL1A2	Intron 21	c.1197+5G > A	Splicing	Skipping of exon 21	r.1090_1197del	(p.Gly364_Arg399del)	NA	IV
PUMC-253	COL1A2	Intron 33-exon 34	c.2026-1_2042dup	Splicing	No aberration	r.2042_2043insertGGGTGCTC CTGGTGCTGT	(p.Val681_Gly682insGlyAlaProGlyAlaVal)	No aberration	IV
PUMC-312	COL1A2	Intron 44	c.2943+1-2943+2delgt	Splicing	Skipping of exon 44	r.2836_2943del	(p. Gly946_Thr981del)	NA	III
Variants in exons									
PUMC-234	COL1A1	Exon 16	c.1003G > A	Missense	No aberration	c.1003G > A	No aberration	NA	IV
PUMC-2	COL1A1	Exon 19	c.1201G > A	Missense	No aberration	c.1201G > A	No aberration	NA	III
PUMC-351	COL1A1	Exon 37	c.2560G > A	Missense	No aberration	c.2560G > A	No aberration	NA	IV
PUMC-23	COL1A1	Exon 37	c.2560G > A	Missense	No aberration	c.2560G > A	No aberration	NA	IV
PUMC-41	COL1A1	Exon 37	c.2560G > A	Missense	No aberration	c.2560G > A	No aberration	NA	IV
PUMC-339	COL1A1	Exon 37	c.2560G > A	Missense	No aberration	c.2560G > A	No aberration	NA	IV
PUMC-371	COL1A2	Exon 16	c.792G > A	Synonymous	Skipping of exon 16	c.739_792del	(p.247Gly_264Lysdel)	Skipping of exon 16	III
PUMC-479	COL1A2	Exon 18	c.892G > A	Missense	No aberration	c.892G > A	No aberration	NA	I
PUMC-448	COL1A2	Exon 28	c.1612G > A	Missense	No aberration	c.1612G > A	No aberration	NA	III
PUMC-441	COL1A2	Exon 38	c.2296G > A	Missense	No aberration	c.2296G > A	No aberration	NA	IV
PUMC-296	COL1A2	Exon 40	c.2404G > A	Missense	Retention of 49bp/No aberration	c.2403_2404ins2404-49_2404-1/c.2404G > A	(p.Gly802Alafs*5)/No aberration	NA	I
PUMC-430	COL1A2	Exon 45	c.2944G > A	Missense	No aberration	c.2944G > A	No aberration	NA	III
PUMC-115	COL1A2	Exon 48	c.3106G > T	Missense	No aberration	c.3106G > T	No aberration	NA	IV

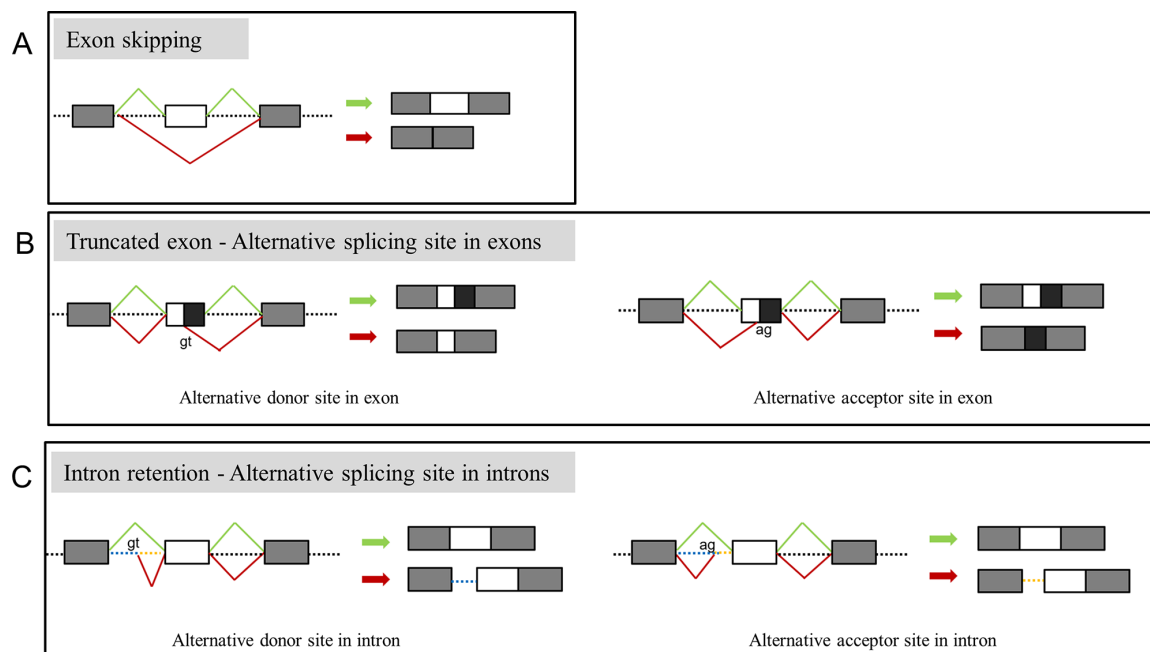


FIGURE 2 | Representation of main splicing effects in OI patients. **(A)** Variants resulting in exon skipping. **(B)** Variants resulting in truncated exon caused by recognition of alternative donor (left) or acceptor (right) in exons. **(C)** Variants resulting in intron retention caused by recognition of alternative donor (left) or acceptor (right) in introns. Splicing products in green indicate the wild type transcript, products in red indicate the aberrant splicing.

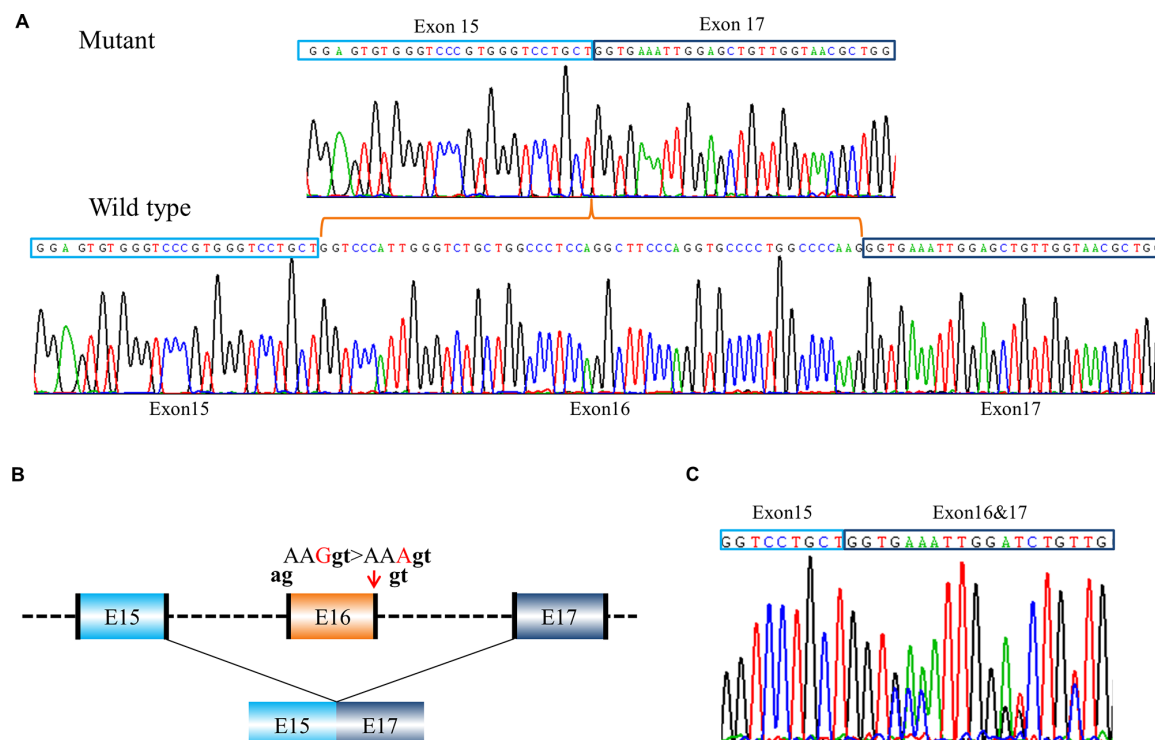
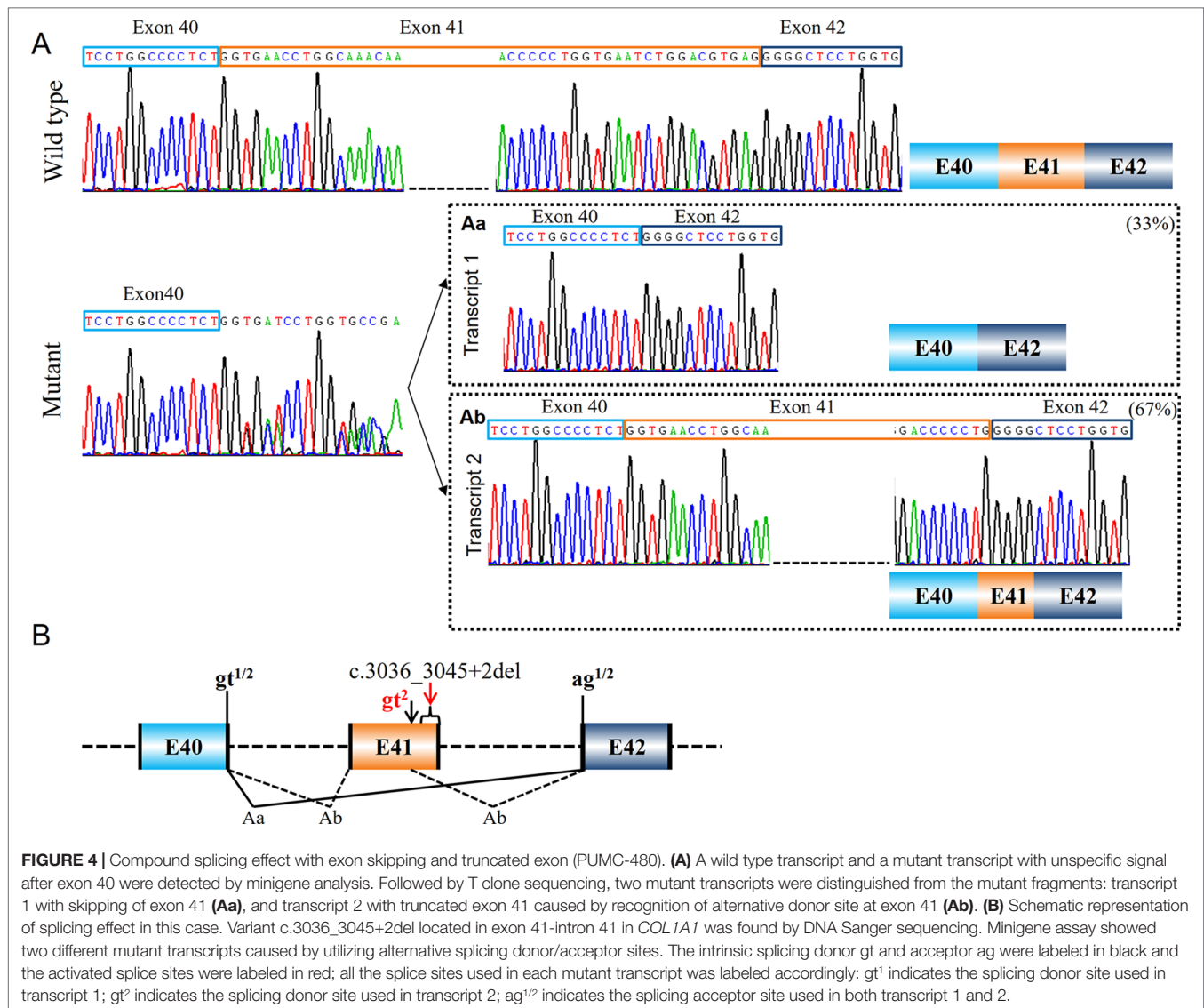


FIGURE 3 | A case of exon skipping resulted from a synonymous mutation (PUMC-371). **(A)** Sequencing analysis by minigene assay indicated a wild type transcript and a mutant transcript. Compared with the wild type transcript, the mutant transcript showed exon 16 skipping. **(B)** Schematic representation of the splicing effect. A synonymous mutation c.792G > A (p.Lys264Lys) in *COL1A2* at the last nucleotide in exon 16 was found by DNA Sanger sequencing. Splicing assay indicated the skipping of exon 16, c.739_792del (p.Gly247_Lys264del). The dinucleotide in black indicated intrinsic splicing donor or acceptor. **(C)** Sequencing analysis of RT-PCR products from patient's fibroblasts confirmed exon 16 skipping.



Recognition of Alternative Acceptor Site in Intron

Five probands (PUMC-15, PUMC-105, PUMC-369, PUMC-189, and PUMC-296) were found with intron retention caused by alternative acceptor site in intron in this study (Table 1). In particular, PUMC-296 (Figure 6) carried a missense mutation c.2404G > A in *COL1A2* indicated by Sanger sequencing. Such change took place in the first nucleotide in exon 40, therefore agGG altered to agAG. An alternative 3' splice site in intron 39, c.2404-51_2404-50ag, was recognized during splicing in one of the mutant transcripts (Figure 6Aa). This led to an insertion of 49bp (retention of partial intron 39) in the mRNA.

Compound Splicing Effects Resulted From Numbers of Aberrant Splicing Transcripts

During splicing, more than one transcript can be generated because of the existence of alternative splicing. This makes some aberrant splicing cases even more complicated. In this study, there were three variants generating more than one mutant

transcript showed by minigene assay: c.642+4delA in *COL1A1* (PUMC-401), c.3036_3045+2del in *COL1A1* (PUMC-480), and c.2404G > A in *COL1A2* (PUMC-296) (Table 1).

Patient PUMC-401 with a variant c.642+4delA in *COL1A1* formed four mutant splicing isoforms (Figures 5Aa–Ad). Four pairs of alternative splice sites utilized in this patient were labelled in the schematic map (Figure 5B): Splicing of *gt*¹ (c.588+1_588+2*gt*) and *ag*¹ (c.643-2_643-1*ag*) generated transcript 1 with skipping of exon 8 (Figure 5Aa); Splicing of *gt*² (c.588+1_588+2*gt*) and *ag*² (c.660-661*AG*) formed transcript 2 with deletion of exon 8 and partial exon 9 (Figure 5Ab); Splicing of *gt*³ (c.589-62_589-61*gt*) and *ag*³ (c.643-2_643-1*ag*) generated transcript 3 with deletion of exon 8 and insertion of partial intron 7 (Figure 5Ac); Transcript 4 was generated by splicing of *gt*⁴ (c.588+1_588+2*gt*) and *ag*⁴ (c.589-2_589-1*ag*), together with *gt*⁴ (c.617-618*GT*) and *ag*⁴ (c.643-2_643-1*ag*) with the effect of truncated exon 8 (Figure 5Ad). Exon skipping (transcript 1, 55%) and intron retention (transcript 3, 27%) were the most dominant isoforms (Figure 5). Sanger

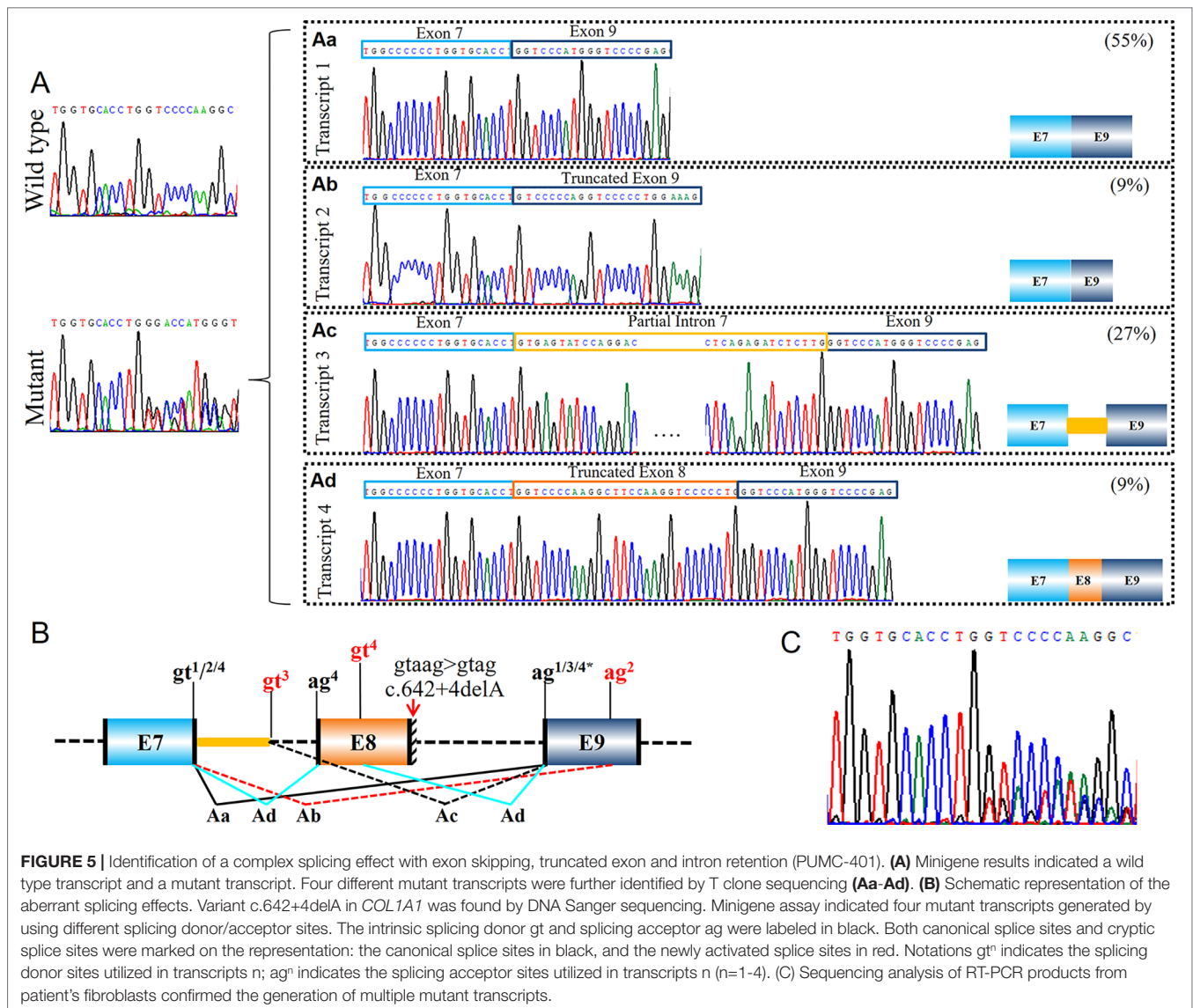


FIGURE 5 | Identification of a complex splicing effect with exon skipping, truncated exon and intron retention (PUMC-401). **(A)** Minigene results indicated a wild type transcript and a mutant transcript. Four different mutant transcripts were further identified by T clone sequencing (**Aa–Ad**). **(B)** Schematic representation of the aberrant splicing effects. Variant c.642+4delA in COL1A1 was found by DNA Sanger sequencing. Minigene assay indicated four mutant transcripts generated by using different splicing donor/acceptor sites. The intrinsic splicing donor gt and splicing acceptor ag were labeled in black. Both canonical splice sites and cryptic splice sites were marked on the representation: the canonical splice sites in black, and the newly activated splice sites in red. Notations gtⁿ indicates the splicing donor sites utilized in transcripts n; agⁿ indicates the splicing acceptor sites utilized in transcripts n (n=1–4). **(C)** Sequencing analysis of RT-PCR products from patient's fibroblasts confirmed the generation of multiple mutant transcripts.

sequencing results of dermal fibroblasts confirmed that multiple transcripts were generated (**Figure 5C**), and that alternative donor/acceptor sites were utilized *in vivo* (**Figure S2**). Similarly, we found exon skipping (transcript 1, 17%) and intron retention (transcript 2, 25%) were the most prevalent mutant isoforms (**Figure S2**). However, 36 bp deletion in exon 8 (transcript 4, 4%) and retention of intron 8 (transcript 5, 4%) were only found in fibroblast assay (**Figure S2**) and skipping of exon 8 and partial exon 9 (transcript 2, 9%) were only found in minigene assay (**Figure 5**).

Similarly, patient PUMC-480 formed two different mutant transcripts (**Figure 4A**). Two pairs of alternative splice sites were used: Utilizing gt¹ (c.2937+1_2937+2gt) and ag¹ (c.3046-2_3046-1ag) generated transcript 1 with exon 41 skipping (33%); Utilizing gt² (c.3029-3030 GT) and ag² (c.3046-2_3046-1ag) generated transcript 2 with deletion of partial exon 41 (67%). Another patient, PUMC-296 (c.2404G > A in COL1A2) showed a missense variant at the first nucleotide in exon 40. Two transcripts were found by minigene analysis: one with a missense

variant (c.2404G > A, 60%) and the other with an insertion of 49 bp (40%) by recognition of alternative 3' splice site (c.2404-51_2404-50ag) in intron 39 (**Figure 6A**).

No Remarkable Aberrant Splicing Effect

Among the 29 variants, 12 variants did not show any splicing consequence indicated by minigene assay (**Table 1**). Most of them were missense variants at the first nucleotide in the exons. While some of them (c.370-9C > T in COL1A1, c.2613+9C > T in COL1A1, c.1036-9G > T in COL1A2, c.2026-1_2042dup in COL1A2) carried the variants in introns without aberrant splicing, and they were excluded from the pathogenic variants. In particular, variant c.2026-1_2042dup in COL1A2 (PUMC-253) should be highlighted. This variant may cause aberrant splicing because the duplication covered the 3' boundary of intron 33 to the 5' partial exon 34 (**Figure S3**). After verification using minigene assay, two transcripts

were observed: a wild type transcript from the normal allele and a mutant transcript from the mutant allele (**Figure S3A**). Because the mutant transcript showed the same pattern c.2026-1_2042dup as the sequencing results, no splicing effect was found. RT-PCR of RNA extracted from fibroblasts of this patient confirmed that no aberration was observed (**Figure S3C**).

RELATIONSHIP BETWEEN GENOTYPES AND PHENOTYPES

According to the clinical features of OI including fracture frequency, presence of blue sclerae and bone deformity, the 29 variants (34 OI patients) were classified into different phenotypical groups (**Table 1**): 8 variants were grouped as type I, 8 were type III, and the remaining 13 were type IV OI. Most of the variants with aberrant splicing corresponded to a mild phenotype (e.g. type I or type IV OI). For example, PUMC-296 who was identified the variant c.2404G > A in *COL1A2* leading to multiple mutant transcripts, presented a mild phenotype: 0.3 fracture times per year without other skeletal problems.

Those exhibited severe phenotype (type III OI), the minigene analysis showed no aberrant splicing (confirmed as no aberration for intronic variants or missense mutation for exonic variants) or exon skipping effect. For instance, PUMC-371 (c.792G > A in *COL1A2* consequent to skipping of exon 16) displayed rather severe phenotypes: with more than 30 times total fracture times (2.9 times yearly of fracture frequency), short stature (Z score = -6.32), presence of dentinogenesis imperfecta and disability of walking.

Moreover, the patients with aberrant splicing effects caused by intronic variants (n = 17) often expressed relatively milder phenotypes: only 11.76% (2 in 17) of them were OI type III, and 88.24% (15 in 17) were OI type I or type IV. Regarding the exonic variants, a large proportion led to a severe type III phenotype, being 30.77% (4 in 13).

DISCUSSION

The splicing effects of 29 suspected atypical splicing variants associate with OI were examined in current study. Among 29 variants, 17 were identified with aberrant splicing, and 12 were not observed any abnormal splicing effect. The splicing effects can be classified as (i) exon skipping or (ii) alternative splice site induced intron retention or partial exon deletion. We further conducted skin fibroblast RT-PCR sequencing and confirmed the findings in the minigene assay, suggesting it is a reliable approach to assess the splicing consequences.

THE MECHANISM OF ABERRANT SPLICING GENERATION

Pre-mRNA splicing occurs when exons and introns are precisely recognized. Two theories were proposed about the splicing

initiation: the intron definition and exon definition (Keren et al., 2010). In intron definition, 5' splice site (GT) and 3' splice site (AG) as well as branch site (YNYURAY) are recognized and mRNA splicing mechanism places across the introns. Variants locate at any of these sites will impair the transcription (Vijayraghavan et al., 1986). While in the exon definition, exons are identified by their naturally high GC proportion. Though exon definition was believed to be the main mechanism of the evolution of alternative splicing (Ram and Ast, 2007), the core intronic splicing signal was still widely studied and believed to be crucial for aberrant splicing. In this study, 89% (17/19) aberrant splicing was caused by the intronic variants (**Table 1**), supporting this notion. To explore the mechanisms underlying aberrant splicing, we further analyzed our results and found the following three main causative reasons for aberrant splicing.

Canonical 5' Splice Site Cannot Be Recognized

This can be resulted from the alteration of an adjacent nucleotide, for example in patient PUMC-371 (c.792G > A in *COL1A2*), such variant changed the consensus sequence AAGgt to AAAgt (**Figure 3**). It was known that the conservation of last nucleotide at 3' exon is G > A/T (Roca et al., 2012; Roca et al., 2013). The alteration from G to A changed the conservation, and disrupted the base-pairing between U1 small nuclear RNA (snRNA) and the donor site (Roca et al., 2013). In addition, unrecognition of authentic donor can be also caused by the inexistence of 5' splice site resulted from a deletion (**Figure 4**). PUMC-480 (c.3036_3045+2del in *COL1A1*) belongs to this instance, and the disappearance of the canonical donor site induced exon skipping or the activation of a cryptic donor site.

Both Canonical 5' and 3' Splice Sites Are Deactivated

The deactivation of both splice sites can lead to rather complicated case, for instance, PUMC-401 (c.642+4delA in *COL1A1*). The deletion changed 5' intronic consensus sequence gtaag to gtag (**Figure 5**). The conservation of +4 site in intron is A > T/G (Roca et al., 2012), so the variant led to deactivation of canonical donor site and the selection of alternative donor/acceptor sites for all mutant transcripts both from minigene results (**Figures 5Aa–Ad**) and from cultured fibroblasts (**Figure S2**). The alteration near 5' intronic site caused the deactivation of acceptor site in adjacent intron (**Figure 5Ab**), but the reasons remain to be elucidated. Similar effects were reported by Schwarze et al. (1999) that variant c.642+1G > A in *COL1A1* led to multiple mutant transcripts caused by employing alternative donor sites. As both variants are located in intron 8 of *COL1A1*, and it was showed that introns 5, 6, and 9 were removed before introns 7 and 8 (Schwarze et al. 1999). This could be one of the reasons that both studies found the compound transcripts when variants are located in intron 8 of *COL1A1*.

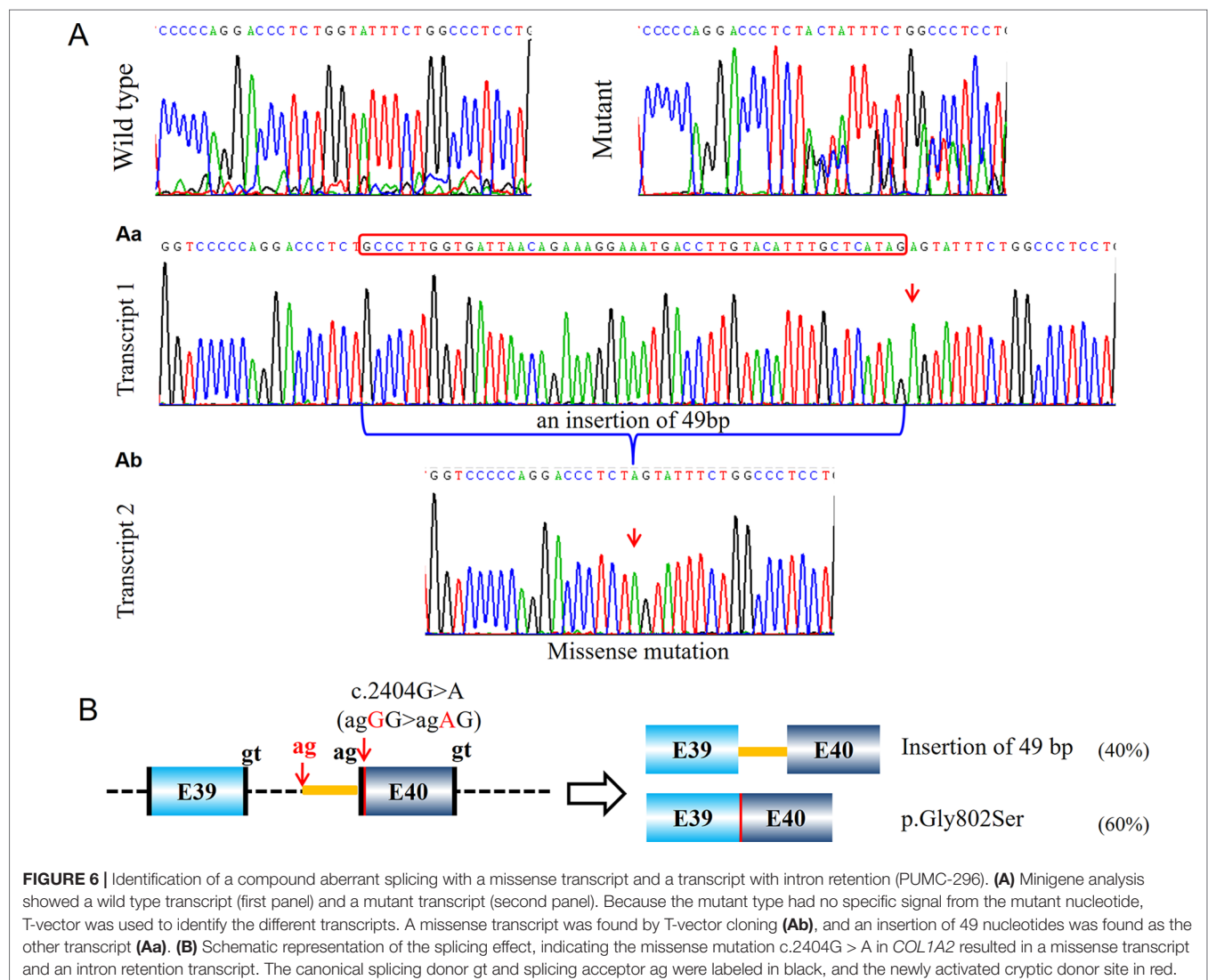
Canonical 3' Splice Site Cannot Be Recognized

A 3' splice site includes a branch point, a polypyrimidine tract and a splicing acceptor site (Wahl et al., 2009). One possible reason leading to the unrecognition of 3' splice site is the changing of nucleotide adjacent to the splicing acceptor as happened in PUMC-296 (c.2404G > A in *COL1A2*) (Figure 6). The acceptor site is recognized through non-Watson-Crick interaction by pairing with donor site and branch point (Wilkinson et al., 2017). Wilkinson et al. (2017) reported that the first 10 nucleotides of 5' exon are always well ordered to facilitate the mRNA processing. The boundary of an 3' splice site and 5' exon is always consensus as Y₁₀NCAG/G, where Y stands for pyrimidine and N equals A/G/C/T (Sun and Chasin, 2000). Therefore, the alteration at the first nucleotide in PUMC-296 resulted in deactivation of canonical acceptor, and instead a cryptic acceptor site was selected. Another reason is that the variants may be located at polypyrimidine tract (PPT) region. Variants in probands PUMC-276, PUMC-290, PUMC-15, PUMC-105, PUMC-369, and PUMC-189 belong to this case. It was known

that by binding to different locations of sequences, polypyrimidine tract-binding protein1 (PTBP1) can induce either exon skipping or inclusion (Hamid and Makeyev, 2017). Sanz et al. (2010) reported that variants affecting PPT region resulted in the exon skipping. Consistently, in PUMC-276 and PUMC-290, both of the two variants caused skipping of exon 51 and deletion of partial 5'UTR (Figure S1). The remaining four variants mentioned above led to insertion of part of PPT region (Table 1).

RELATIONSHIP BETWEEN ABERRANT SPLICING AND PHENOTYPE

Most of the aberrant splicing found in this study corresponds to mild phenotypes (Type I or type IV OI) (Table 1). Type I collagen is a protein of triple helix structure comprised of two alpha1 chains and one alpha2 chain (Marini et al., 2017). Its synthesis involves the correct post-translational modifications, folding and secretion (Ishikawa and Bächinger, 2013). Variants within its



encoding genes, *COL1A1* and *COL1A2*, have two main types of collagen defects: quantitative defect and structure defect (Marini et al., 2007). The structure alterations generally cause more severe phenotypes due to excessive post translational modification (Ishikawa and Bächinger, 2013). The collagen defect mechanism can be classified into two types: (I) Synthesizing of single *COL1A1* allele consequences in haploinsufficiency. This involves nonsense-mediated mRNA decay, or frameshift/splicing mutation-induced pre-termination codon, and most of them being mild OI type (Rauch et al., 2010); (II) The helical mutations of *COL1A1* or *COL1A2* induced structural change of type I collagen. Missense mutations in triple-helical domain can result in dominant negative effect, thus impair the collagen folding and synthesis. The helical mutations are mostly glycine substitutions and the severity varies from mild to severe levels (Rauch et al., 2010; Lindahl et al., 2015).

Among the aberrant splicing in this research, we noticed that all OI patients with more than one mutant transcripts (e.g. PUMC-401, PUMC-480, and PUMC-296) have mild phenotypes, being either type I or type IV OI (Table 1). Haploinsufficiency could be the main reason, as one wild type allele may fulfill the normal functions. Regarding the mutant allele, although there were many different mutant transcripts, some of them led to premature termination codon (e.g. PUMC-296, Figure 6), and induced the degradation of those transcripts (Kervestin and Jacobson, 2012). Therefore, in principle, only a small proportion of defective transcripts affect the collagen function.

Similarly, variants locate at the polypyrimidine tract (PPT) region (e.g. PUMC-15, PUMC-105, PUMC-369, PUMC-189, PUMC-276 and PUMC-290) have mild phenotypes as well. What need to be noted here is PUMC-15, 105, 369, and 189, among whom all the variants resulted in an insertion of part of PPT sequence, generated the premature termination codon and therefore resulted in the degradation of the defective transcript (Kervestin and Jacobson, 2012). Their phenotypes (type I OI) are in agreement with the protein alteration (Table 1).

The most dominant splicing effect is exon skipping. However, we did not observe a strong correlation between exon skipping and phenotype (Table 1). Most of patients with exon skipping expressed milder clinical manifestations (type I or type IV OI) than those with missense mutations. There are only three (PUMC-90, 312, and 371) patients have severer phenotype, with two exon 16, and one exon 44 skipping, respectively. Depending on the location of skipped exons, the severity of OI can vary from mild to severe level (Thomas and DiMeglio, 2016). Even if the skipping did not change the Gly-X-Y triplet pattern, the chain alignment may still have causative effect on collagen folding (Marini et al., 2017). If the variant occurs at the C-terminal region of propeptide, this may be associated with protein folding delay, thus further affect the correct assembly of collagen (Symoens et al., 2014). The locations of both alteration (Marini et al., 2007) and modifier genes (Riordan and Nadeau, 2017) contribute to different phenotypes, and details remain to be elucidated.

Although a large proportion of structural defects of collagen was due to the classical splicing mutations (Marini et al., 2007), atypical variants in the introns or exons that are close to the splice sites are also important and hence should be highlighted in

future sequencing analysis. Among the recruited 867 OI patients, we found 17 atypical splicing variants and 22 typical splicing variants. Thus the atypical splicing variants represent a high proportion (44%, 17/39). For the first time, our study examined and classified the atypical (exon/intron border exclusive) splicing variants associated with OI, which helps to identify the causative mutation and establish the correlation between splicing effect and OI phenotypes.

DATA AVAILABILITY STATEMENT

The Datasets Generated for This Study Can Be Found in the Osteogenesis Imperfecta Variant Database ([Http://Oi.Gene.Le.Ac.Uk/](http://Oi.Gene.Le.Ac.Uk/)).

ETHICS STATEMENT

All procedures performed in this study involving human participants were approved by Institutional Review Board (IRB) of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China (015-2015). Informed consent was obtained from all adult participants/legal guardians of children under age 18.

AUTHOR CONTRIBUTIONS

LL and YC performed the minigene assay, sequencing analysis, and wrote the manuscript. FZ, BM, and YY carried out plasmid construction. SL and TY conducted data collection as well as data analysis. XR and YW helped with recruiting patients and YG helped to discuss the data and helped writing the final manuscript. XZ conceived the study and supervised this research. All authors performed critical reading and approved the final version of manuscript.

FUNDING

This study was supported by grants from National Key Research and Development Program of China (2016YFE0128400, 2016YFC0905100), CAMS Innovation Fund for Medical Sciences (CIFMS, 2016-I2M-3-003) and National Natural Science Foundation of China (81472053).

ACKNOWLEDGMENTS

The authors would like to thank all OI patients and their families for their participation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00979/full#supplementary-material>

REFERENCES

- Ahlborn, L. B., Dandanel, M., Steffensen, A. Y., Jonson, L., Nielsen, F. C., and Hansen, T. V. (2015). Splicing analysis of 14 BRCA1 missense variants classifies nine variants as pathogenic. *Breast Cancer Res. Treat.* 150 (2), 289–298. doi: 10.1007/s10549-015-3313-7
- Byers, P. H., and Pyott, S. M. (2012). Recessively inherited forms of osteogenesis imperfecta. *Annu. Rev. Genet.* 46, 475–497. doi: 10.1146/annurev-genet-110711-155608
- Cooper, T. A. (2005). Use of minigene systems to dissect alternative splicing elements. *Methods* 37 (4), 331–340. doi: 10.1016/j.ymeth.2005.07.015
- Forlino, A., and Marini, J. C. (2016). Osteogenesis imperfecta. *Lancet* 387 (10028), 1657–1671. doi: 10.1016/S0140-6736(15)00728-X
- Fraile-Bethencourt, E., Valenzuela-Palomo, A., Díez-Gómez, B., Caloca, M. J., Gómez-Barrero, S., and Velasco, E. A. (2019). Minigene splicing assays identify 12 spliceogenic variants of BRCA2 exons 14 and 15. *Front. Genet.* 10, 503. doi: 10.3389/fgene.2019.00503
- Gagliardi, A., Besio, R., Carnemolla, C., Landi, C., Armini, A., Aglan, M., et al. (2017). Cytoskeleton and nuclear lamina affection in recessive osteogenesis imperfecta: A functional proteomics perspective. *J. Proteomics* 167, 46–59. doi: 10.1016/j.jprot.2017.08.007
- Hamid, F. M., and Makeyev, E. V. (2017). A mechanism underlying position-specific regulation of alternative splicing. *Nucleic Acids Res.* 45 (21), 12455–12468. doi: 10.1093/nar/gkx901
- Ishikawa, Y., and Bächinger, H. P. (2013). A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta (BBA) — Mole. Cell Res.* 1833 (11), 2479–2491. doi: 10.1016/j.bbamcr.2013.04.008
- Johnson, J. M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P. M., Armour, C. D., et al. (2003). Genome wide survey of human alternative pre-mRNA splicing with exon junction microarray. *Science* 302, 2141–2144. doi: 10.1126/science.1090100
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11 (5), 345–355. doi: 10.1038/nrg2776
- Kervest, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell. Biol.* 13 (11), 700–712. doi: 10.1038/nrm3454
- Li, L., Bin, M., Li, S., Xiao, J., Wang, H., Zhang, J., et al. (2019). Genotypic and phenotypic characterization of Chinese patients with osteogenesis imperfecta. *Hum. Mutat.* 40 (5), 588–600. doi: 10.1002/humu.23718
- Lindahl, K., Astrom, E., Rubin, C. J., Grigelioniene, G., Malmgren, B., Ljunggren, O., et al. (2015). Genetic epidemiology, prevalence, and genotype-phenotype correlations in the Swedish population with osteogenesis imperfecta. *Eur. J. Hum. Genet.* 23 (8), 1042–1050. doi: 10.1038/ejhg.2015.81
- Lindert, U., Cabral, W. A., Ausavarat, S., Tongkobpetch, S., Ludin, K., Barnes, A. M., et al. (2016). MBTPS2 mutations cause defective regulated intramembrane proteolysis in X-linked osteogenesis imperfecta. *Nat. Commun.* 7, 11920. doi: 10.1038/ncomms11920
- Marini, J. C., Forlino, A., Bachinger, H. P., Bishop, N. J., Byers, P. H., Paepe, A., et al. (2017). Osteogenesis imperfecta. *Nat. Rev. Dis. Primers* 3, 17052. doi: 10.1038/nrdp.2017.52
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., et al. (2007). Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* 28 (3), 209–221. doi: 10.1002/humu.20429
- Ram, O., and Ast, G. (2007). SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet.* 23 (1), 5–7. doi: 10.1016/j.tig.2006.10.002
- Rauch, F., and Glorieux, F. H. (2004). Osteogenesis imperfecta. *Lancet* 363 (9418), 1377–1385. doi: 10.1016/S0140-6736(04)16051-0
- Rauch, F., Lalic, L., Roughley, P., and Glorieux, F. H. (2010). Relationship between genotype and skeletal phenotype in children and adolescents with osteogenesis imperfecta. *J. Bone Miner. Res.* 25 (6), 1367–1374. doi: 10.1359/jbmr.091109
- Riordan, J. D., and Nadeau, J. H. (2017). From peas to disease: modifier genes, network resilience, and the genetics of health. *Am. J. Hum. Genet.* 101 (2), 177–191. doi: 10.1016/j.ajhg.2017.06.004
- Roca, X., Akerman, M., Gaus, H., Berdeja, A., Bennett, C. F., and Krainer, A. R. (2012). Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.* 26 (10), 1098–1109. doi: 10.1101/gad.190173.112
- 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 (2), 129–144. doi: 10.1101/gad.209759.112
- Rohrbach, M., and Giunta, C. (2012). Recessive osteogenesis imperfecta: clinical, radiological, and molecular findings. *Am. J. Med. Genet. Part C: Semin. Med. Genet.* 160C (3), 175–189. doi: 10.1002/ajmg.c.31334
- Sanz, D. J., Acedo, A., Infante, M., Duran, M., Perez-Cabornero, L., Esteban-Cardena, E., et al. (2010). A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin. Cancer Res.* 16 (6), 1957–1967. doi: 10.1158/1078-0432.CCR-09-2564
- Schleif, J., Bailey, S. S., Tran, T., Chen, D., Stowers, S., Schwarze, U., et al. (2015). Molecular outcome, prediction, and clinical consequences of splice variants in COL1A1, which encodes the proalpha1(I) chains of type I procollagen. *Hum. Mutat.* 36 (7), 728–739. doi: 10.1002/humu.22812
- Schwarze, U., Starman, B. J., and Byers, P. H. (1999). Redefinition of exon7 in the COL1A1 gene of type I collagen by an intron8 splice donor site mutation in a form of osteogenesis imperfecta: influence of intron splice order on outcome of splice site mutation. *Am. J. Hum. Genet.* 65, 336–344. doi: 10.1086/302512
- Sillence, D. O., Senn, A., and Darks, D. M. (1979). Genetic heterogeneity in osteogenesis imperfecta. *J. Med. Genet.* 16, 101–116. doi: 10.1136/jmg.16.2.101
- Stoll, C., Dott, B., Roth, M., and Alembik, Y. (1989). Birth prevalence rates of skeletal dysplasias. *Clin. Genet.* 35, 88–92. doi: 10.1111/j.1399-0004.1989.tb02912.x
- Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* 20 (17), 6414–6425. doi: 10.1128/MCB.20.17.6414-6425.2000
- Symoens, S., Hulmes, D. J., Bourhis, J. M., Coucke, P. J., De Paepe, A., and Malfait, F. (2014). Type I procollagen C-propeptide defects: study of genotype-phenotype correlation and predictive role of crystal structure. *Hum. Mutat.* 35 (11), 1330–1341. doi: 10.1002/humu.22677
- Thomas, I. H., and DiMeglio, L. A. (2016). Advances in the classification and treatment of osteogenesis imperfecta. *Curr. Osteoporos. Rep.* 14 (1), 1–9. doi: 10.1007/s11914-016-0299-y
- Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J., et al. (1986). Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway particularly assembly of the spliceosome. *EMBO J.* 5 (7), 1683–1695. doi: 10.1002/j.1460-2075.1986.tb04412.x
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136 (4), 701–718. doi: 10.1016/j.cell.2009.02.009
- Wang, Z., and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14 (5), 802–813. doi: 10.1261/rna.876308
- Wilkinson, M. E., Fica, S. M., Galej, W. P., Norman, C. M., Newman, A. J., and Nagai, K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science* 358 (6368), 1283–1288. doi: 10.1126/science.aar3729
- You, Y., Wang, X., Li, S., Zhao, X., and Zhang, X. (2018). Exome sequencing reveals a novel MFN2 missense mutation in a Chinese family with Charcot-Marie-Tooth type 2A. *Exp. Ther. Med.* 16 (3), 2281–2286. doi: 10.3892/etm.2018.6513
- Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185. doi: 10.1038/nature01031

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

Copyright © 2019 Li, Cao, Zhao, Mao, Ren, Wang, Guan, You, Li, Yang and Zhao. 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.



RNA-Seq Perspectives to Improve Clinical Diagnosis

Guillermo Marco-Puche¹, Sergio Lois¹, Javier Benítez^{2*} and Juan Carlos Trivino^{1*}

¹ Bioinformatics Group, Sistemas Genómicos, Paterna, Spain, ² Human Genetics Group, Spanish National Cancer Research Center, Madrid, Spain

OPEN ACCESS

Edited by:

Eladio Andrés Velasco,
Institute of Biology and Molecular
Genetics (IBGM), Spain

Reviewed by:

Rahul N Kanadia,
University of Connecticut,
Mansfield,
United States

Elton J. R. Vasconcelos,
University of Leeds,
United Kingdom

*Correspondence:

Javier Benítez
jbenitez@cniio.es
Juan Carlos Trivino
jc.trivino@sistemasgenomicos.com

Specialty section:

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

Received: 20 August 2019

Accepted: 22 October 2019

Published: 12 November 2019

Citation:

Marco-Puche G, Lois S, Benítez J
and Trivino JC (2019) RNA-
Seq Perspectives to Improve
Clinical Diagnosis.
Front. Genet. 10:1152.
doi: 10.3389/fgene.2019.01152

In recent years, high-throughput next-generation sequencing technology has allowed a rapid increase in diagnostic capacity and precision through different bioinformatics processing algorithms, tools, and pipelines. The identification, annotation, and classification of sequence variants within different target regions are now considered a gold standard in clinical genetic diagnosis. However, this procedure lacks the ability to link regulatory events such as differential splicing to diseases. RNA-seq is necessary in clinical routine in order to interpret and detect among others splicing events and splicing variants, as it would increase the diagnostic rate by up to 10–35%. The transcriptome has a very dynamic nature, varying according to tissue type, cellular conditions, and environmental factors that may affect regulatory events such as splicing and the expression of genes or their isoforms. RNA-seq offers a robust technical analysis of this complexity, but it requires a profound knowledge of computational/statistical tools that may need to be adjusted depending on the disease under study. In this article we will cover RNA-seq analyses best practices applied to clinical routine, bioinformatics procedures, and present challenges of this approach.

Keywords: RNA-Seq - RNA sequencing, transcriptomics, bioinformatics, clinical routine, tissue-specific expression, variants of uncertain significance (VUS), alternative splicing (AS), DEG (differentially expressed genes)

INTRODUCTION

In recent years, the use of next-generation sequencing (NGS) for the diagnosis of Mendelian or rare genetic disorders has entered routine clinical practice. The increasing ability to sequence entire genomes in a cost-effective manner has allowed the identification of approximately 260 novel rare genetic diseases per year (Boycott et al., 2017). Focusing on the ~1.5% of the human genome represented by coding sequences, diagnostic rates of whole-exome sequencing (WES) vary widely by inherited condition, and they range from 28 to 55% (Retterer et al., 2016). By extending the focus to deep intronic and regulatory variants in non-coding regions, including structural and non-exonic variants not detectable by WES, whole-genome sequencing (WGS) increased the diagnostic rate by more than 17% (Lionel et al., 2018). The high rate of undiagnosed cases is related to at least two important limitations: (i) the catalog of Mendelian phenotypes is as yet far from complete (~300 new Mendelian phenotypes are added to the OMIM database each year (Chong et al., 2015)); and (ii) although the interpretation of protein-coding regions of the genome is reliable, our understanding of non-coding variation and its functional interpretation is still limited.

Recently, different studies reported on how the application of RNA sequencing (RNA-seq) can help to shed light on the possible pathogenicity of variants of unknown significance (VUS) identified through DNA sequencing studies such as WES and WGS, as it provides direct insight into the transcriptional

alterations caused by VUS and thus improves diagnostic rates (Cummings et al., 2017; Kremer et al., 2017). Alternative splicing (AS) is considered to be a key cellular process in ensuring functional complexity in higher eukaryotes (Chen et al., 2012). Remarkably, this process is estimated to affect more than 88% of human protein-coding genes (Kampa et al., 2004). The major effector of the RNA splicing reaction is the spliceosome, a complex of hundreds of interacting proteins, and small nuclear RNAs (snRNAs) including the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5, and U6 (Tazi et al., 2009). Each intron of the pre-mRNA is flanked by a 5'-exon and a 3'-exon and contains different conserved splicing signals recognized by the spliceosome: the 5'-splice site, the branch point sequence, the 3'-splice site, and the polypyrimidine tract located 5-40 bp upstream of the 3' end of the intron (Cartegni et al., 2002) (**Supplementary Figure 1**). Since these splicing signals are not sufficient for splicing regulation, the fidelity of pre-mRNA splicing depends on interactions between *trans*-acting factors (proteins and ribonucleoproteins) and *cis*-acting elements (pre-mRNA sequences), including exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), and intronic splicing silencer (ISS) elements (Blencowe, 2006), that exert their effects by facilitating the binding of splicing factors, which in turn positively or negatively regulate inclusion of a particular exon.

Due to its underlying complexity, AS can lead to disease in different ways. The most common alterations of the splicing process are in *cis*-acting regulatory elements that are located either in core consensus sequences (5' splice site, 3' splice site, and branch point) or in regulatory elements that modulate spliceosome recruitment (Singh and Cooper, 2012). Some authors estimate that up to 62% of all disease-causing single nucleotide variants (SNVs) may affect RNA splicing (Lopez-Bigas et al., 2005). In terms of evolutionary conservation, about 50% of the synonymous positions in codons of conserved alternatively spliced mRNAs are under selection pressure, suggesting that conserved alternative exons and their flanking introns are strongly enriched in splicing regulatory elements (Blencowe, 2006). In this regard, it has been estimated that up to 25% of synonymous substitutions can disrupt normal splicing in the same way as non-synonymous variants or premature termination codons (Pagani et al., 2005), suggesting that those regions should also be routinely examined. Different examples of Mendelian disorders have already been associated with transcriptional perturbations introduced by both synonymous and non-synonymous variants (Slaugenhaupt et al., 2001; Cassini et al., 2019) (**Supplementary Table 1**). Since RNA-seq is not a part of current diagnostic genetic testing routine, these estimates seem to reflect a significant proportion of potentially diagnosable cases that remain unresolved at present. Some authors demonstrate the utility of RNA-seq to diagnose 10% of patients with mitochondrial diseases and identify candidate genes for the remaining 90% (Kremer et al., 2017).

SECTION 1: TOWARDS CLINICAL APPLICATION OF RNA SEQUENCING

During the past years, the importance of RNA-seq as a clinical diagnostic tool has increased. The possibility to analyze new types

of potential pathological variants in clinical routine has led to an increase in the diagnostic rate without an excessive increment in cost or time. However, some issues of RNA-seq analysis must be resolved to ensure the diagnostic quality of the study.

RNA-seq can complement the limitations of purely genetic information by probing variations in RNA with different additional studies (Kremer et al., 2017). First, the expression level of a gene or transcript outside of its physiological range can be measured. Second, cases with allele-specific expression (ASE), and therefore their association with disease predisposition, can be identified (Byron et al., 2016). Third, aberrant splicing can be recognized, which is known to be a major cause of Mendelian disorders (Tazi et al., 2009; Singh and Cooper, 2012; Scotti and Swanson, 2016).

Different studies suggest that 9 to 30% (Stenson et al., 2017) of disease-causing variants have an impact on RNA expression. The measurement of gene expression is thus expected to represent an improvement of the clinical routine; for example, some authors correlate the under-expression of certain genes with loss of function (LOF). This strategy has already been used in the identification of under-expression of *RARS2* in blood, which is associated with global developmental delay, seizures, microcephaly, hypotonia, and progressive scoliosis (Fresard et al., 2018).

Variable expressivity and incomplete penetrance are recurrent genetic issues in variant interpretation and may result from a combination of allelic variation, modifier genes, and/or environmental factors. A genetic condition with a reduced penetrance or high variability of symptoms may be a challenge for diagnosis. Allele-specific expression refers to the differential abundance of the allele copies and is thought to be relevant for as much as 50% of all human genes (Cooper et al., 2013). This differential allele expression can favor either the mutant or the wild-type allele and hence may influence clinical penetrance in different directions (Cartegni et al., 2002). Assuming a recessive condition, ASE-based analysis can help to reveal mono-allelic expression (MAE). For example, variants located in conserved splice sites of exon 12 of the *SPAST* gene lead to exon skipping and cause hereditary spastic paraplegia (HSP). Degradation of aberrant transcripts by a nonsense-mediated decay (NMD) mechanism results in ASE of the *SPAST* wild-type allele (Lopez-Bigas et al., 2005). In contrast, asymptomatic carriers of autosomal dominant retinitis pigmentosa (adRP) are protected from the disease by ASE of the wild-type *PRPF31* allele (Byron et al., 2016). In this context, ASE-based analyses may complement DNA resequencing studies such as WES or WGS for the identification of causative and low-frequency regulatory variants (Lappalainen et al., 2013) or disease-associated predisposition variants (Valle et al., 2008; De La Chapelle, 2009).

SECTION 2: RNA-SEQ, BIOINFORMATICS APPROACH AND NEW PERSPECTIVES FOR KNOWLEDGE OF GENETIC VARIATION

RNA-seq data processing after NGS sequencing is mandatory for an appropriate analysis. As noted in Conesa et al. (2016) there is no optimal pipeline for all the different applications and scenarios in RNA-seq. However, data processing steps must be

included in clinical routine in order to guarantee the quality and reproducibility of the study.

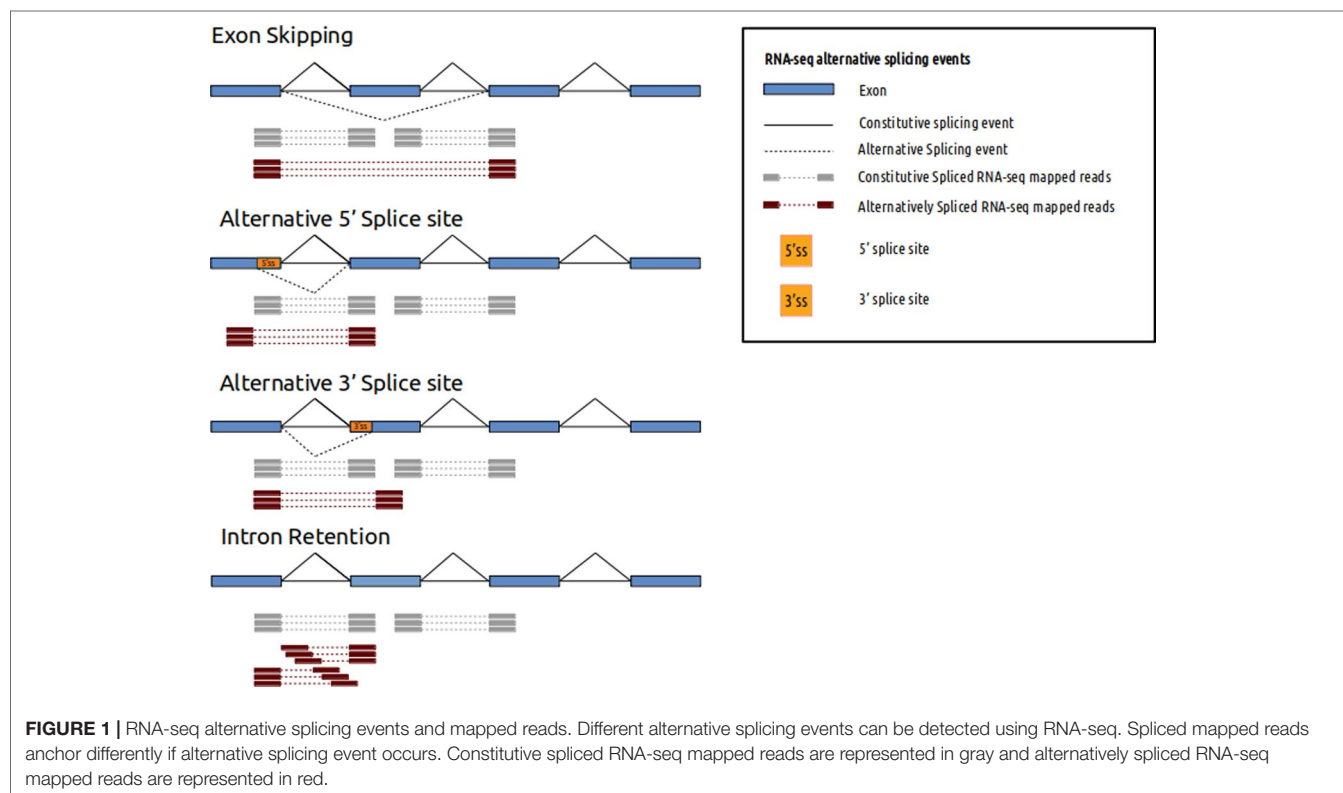
Usually RNA-seq data analysis must start with raw-data quality control. This allows obtaining a general idea of the quality of the sequencing and deciding if the quality requirements for the clinical routine are met. For this purpose different bioinformatics tools such as FastQC (Andrews, 2010) allow to control the most important and general parameters for global evaluation, such as Phred quality score, read length distribution, GC content, k-mer over-representation, adapter content, and duplicated reads. In case of adapter removal, specific bioinformatics tools may be necessary; some of the most referenced tools are CutAdapt (Compeau et al., 2013), FASTX-Toolkit (Gordon, 2010), and Trimmomatic (Bolger et al., 2014). For example, adapter presence or reduced read quality could lead to read misalignment or altered gene expression estimation and splicing event detection.

In the next step, raw-data reads are mapped against a human reference genome using a splice-aware alignment algorithm, such as STAR (Dobin et al., 2013), TopHat2 (Kim, 2013) or HiSAT2 (Kim et al., 2015). Splice-aware aligners allow reads to partially align into splice junctions between exons (Figure 1). In this step, there are important variables that must be evaluated and adjusted according to the type of study and phenotype. For example, the reference version of the genome (Guo et al., 2017) has an impact on the sensitivity and the specificity of variants identified. On the other hand, reference genome annotation files (such as bed or gtf) have a positive impact on mapping performance, quantification, and detection

of differential expression and alternative splicing (Wu et al., 2013). To enrich reference genome annotation, some helpful databases that can be incorporated are SpliceDisease (Finotello et al., 2014) and ASpedia (Wang et al., 2016). SpliceDisease links experimentally supported and manually curated splicing-mutation disease entries with genes and diseases. ASpedia provides genomic annotations extracted from DNA, RNA and proteins, transcription, and regulatory elements obtained from NGS datasets, and isoform-specific functions collected from published datasets.

After mapping the reads to the genome, there are some technical and biological biases that can affect the sensitivity threshold. The 3' end bias of the mapped transcripts could either indicate a technical issue of reduced performance of the number of priming positions from which reverse transcriptase can start cDNA synthesis (Finotello et al., 2014) or a biological issue of RNA degradation by 5' exonuclease (Wang et al., 2016). Assessment of this type of bias is mandatory for the acceptance or rejection of clinical routine samples, and this can be done with quality control tools such as RSeQC (Li et al., 2015).

Prior to assessing differential expression of genes and their isoforms, mapped reads must be quantified. Tools like HTSeq (Anders et al., 2015), FeatureCounts (Liao et al., 2014), and GenomicAlignments (Lawrence et al., 2013) allow quantification of the number of mapped reads within a specific gene feature. Several biases like gene length (Gao et al., 2011) or GC content (Risso et al., 2011) may affect the quantification process and have a negative impact on the differential expression analysis (DEA). To reduce these biases, several methods have been



described. Some methods normalize the read counts based on gene length and library size (total number of reads per replicate). As described in Conesa et al. (2016) the most employed methods involve the use of RPKM units (reads per kilobase of transcript and per million mapped reads) for single-end reads (Mortazavi et al., 2008), FPKM units for paired-end reads (fragments per kilobase of transcript per million mapped reads), and TPM units (transcripts per million). Other more complex normalization strategies are based on a theoretical initial distribution or on housekeeping genes (Evans et al., 2018).

At the isoform level, other quantification methods such as Cufflinks (Evans et al., 2018) and RSEM (Li and Dewey, 2011) are employed. Before testing differential expression between patients, it is mandatory to control technical batch effects and possible biological bias related to biopsy site, gender, or age. Principal component analysis (PCA) or Multi-Dimensional Scaling (MDS) are useful tools for monitoring these effects. After obtaining counts for the gene or transcript level, the count data is processed with different statistical methods such as R/Bioconductor package DESeq2 (Love et al., 2014), (Anders and Huber, 2010), edgeR (Robinson et al., 2009), or SVA (Leek et al., 2012). These tools use batch effect adjustment or modeling to reduce this technical bias. A whole functional RNA-seq pipeline provided by ENCODE can be found in: <https://github.com/ENCODE-DCC/rna-seq-pipeline>

Allele-specific expression can be identified by correlating allele counts obtained from RNA-seq and DNA resequencing. This comparison can be processed using pileLettersAt from the R/Bioconductor package GenomicAlignments (Lawrence et al., 2013). Some authors indicate that the sensitivity of ASE estimation depends on different technical variables such as variant coverage, allele frequency, or the number of alternative alleles (Kremer et al., 2017).

As stated in the American College of Medical Genetics guidelines (Richards et al., 2015), splice site prediction tools such as GeneSplicer (Pertea et al., 2001), Human Splicing Finder (Desmet et al., 2009), and MaxEntScan (Yeo and Burge, 2004) have a higher sensitivity (~90–100%) relative to the specificity (~60–80%) in predicting site abnormalities. It is recommended to use different algorithms to build a single piece of evidence regarding splice site variations. Other algorithms like LeafCutter (Li et al., 2018) rely on RNA-seq data and are able to identify variable splicing events such as: exon skipping, exon truncation, exon elongation, new exon, and complex splicing (or any other splicing event or combinations of the ones mentioned) using short-read RNA-seq data and focusing on excised introns (not relying on predefined models like other tools such as Cufflinks (Roberts et al., 2011)).

SECTION 3: ISSUES TO BE ADDRESSED IN THE TRANSCRIPTOMIC APPROACH

Due to the dynamic nature of the transcriptome, RNA-seq studies present an important technical complexity. Even if RNA-seq studies can be introduced into clinical routine, some conceptual problems should be solved in the coming years.

Different authors point out that one of the major difficulties in transcriptomic analysis and its application to clinical routine is tissue-specific expression (Cummings et al., 2017), where genes and especially their isoforms can present a wide spectrum of splicing events and expression patterns depending on the tissue or cell type. This point is essential for a correct clinical interpretation of the variants (Melé et al., 2015), (Wang et al., 2008), but presents a problem in the initial selection of material for clinical routine. It is mandatory to assess invasiveness when obtaining the material related to the studied disease. Regarding this issue, it is documented that "noninvasive" material such as fibroblasts and blood present 68 and 70.6% of detectable expression of OMIM genes (Cummings et al., 2017; Fresard et al., 2018). This data indicates that using these tissues could help solve a broad spectrum of clinical studies using RNA-seq technology. For example, in neurological diseases, blood tissue presents a detectable expression of 76% of the genes associated with their phenotypes (Fresard et al., 2018).

However, tissue-specific expression may confound RNA-seq analyses and manifests the necessity to select the optimal tissue, whose basal gene expression profile allows monitoring all genes associated with the studied phenotype. For the efficient inclusion of RNA-seq analysis into clinical routine, new biological knowledge is required and additional bioinformatics tools need to be developed. In this context, new databases based on large-scale studies have been collecting and integrating information focused on the relationships between genes, isoforms, and tissues. The database established by the GTEX consortium is one of the most important and widely referenced databases (Melé et al., 2015). As noted in Cummings et al. (2017), the GTEX database is used for tissue selection depending on the clinical case. This information can become the mainstay of new algorithms for the *in silico* selection of optimal tissue depending on the specific disease or phenotype studied for clinical RNA-seq analysis. Some tools using such algorithms have already been described, such as for example PAGE (Nelakuditi).

Additionally, this type of database homogenizes the transcriptomic information from large-scale analyses and could be a valuable source of control samples for statistical contrast and the identification of relatively high frequency variants or splicing events. For this initiative to succeed, and to overcome the inter-analysis barriers, the homogenization of sequencing protocols, starting materials, coverage of analysis, patient description, and bioinformatics pipelines is essential (Cummings et al., 2017). In addition, it is necessary to define the laboratory and bioinformatics parameters and tools that allow monitoring and controlling this process. For example, from a laboratory point of view, assessment of the quality and quantity of extracted RNA, or the library preparation strategy and its possible relationship with technical bias for the NGS process are some of the most important parameters to consider (Wai et al., 2019). To control this bias, different mathematical methods, such as principal component analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (tSNE) based on expression have been proposed (Dey et al., 2017). Another important consideration is the definition of RNA spike-in control mixtures (Devonshire et al., 2010). These elements allow the evaluation of the technical and biological

variability, and are essential for the identification of confounding effects, normalization processes, and quality control.

Regarding technical sensitivity and specificity of RNA-seq applied to clinical routine, the dynamic nature of transcriptomics and the complexity of some alterations, for example, splicing events or ASE deviation, multiplies the number of technical and biological variables to be considered during bioinformatics analysis (Costa et al., 2014). This complexity is reflected in the need to design mathematical methods capable of absorbing if not all, at least part of the variability present in this type of study. In this respect, there are different obstacles for bioinformatics analysis of RNA-seq data. Among them are the mapping process and the possible effect of different factors on the identification of variants, such as the presence of neighboring SNPs and small indels in the unbiased identification of ASE (Wood et al., 2015; Byron et al., 2016), junction events (Williams et al., 2014), or the isoform assembly process, where the length of reads, library preparation strategy, the initial coverage, and GC content of the transcripts could affect the accuracy of the transcript identification process (Mantere et al., 2019; Wai et al., 2019).

FINAL REMARKS

The RNA-seq approach holds the promise to become an interesting clinical routine tool to increase the genetic diagnostic rate. This methodology may increase our knowledge about genetic alterations and their association to genetic diseases with the inclusion of other types of variants, such as splicing events or aberrant gene expression. This type of alterations is usually not detected by DNA resequencing analyses and may be one of the main reasons of the moderate diagnostic rate of this methodology in some diseases.

REFERENCES

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11 (10), R106. doi: 10.1186/gb-2010-11-10-r106
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinf.* 31 (2), 166–169. doi: 10.1093/bioinformatics/btu638
- Andrews, S., Andrews, S. FASTQC A Quality Control tool for High Throughput Sequence Data. 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell* 126 (1), 37–47. doi: 10.1016/j.cell.2006.06.023
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinf.* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., and Baynam, G. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* 100 (5), 695–705. doi: 10.1016/j.ajhg.2017.04.003
- Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., and Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17 (5), 257–271. doi: 10.1038/nrg.2016.10
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3 (4), 285–298. doi: 10.1038/nrg775
- Cassini, T. A., Duncan, L., Rives, L. C., Newman, J. H., Phillips, J. A., and Koziura, M. E. (2019). Whole genome sequencing reveals novel *IGHMBP2* variant leading to

However, due to the dynamic nature of the transcriptome, RNA-seq analysis presents a high complexity, with the concomitant need to consider different technical and biological variables. The control and the effect of these possible fluctuations are currently under investigation. In this context, a deeper and more specific knowledge of the technical and bioinformatics area that varies with the analyzed disease seems necessary to guarantee a meaningful clinical outcome. In this sense, great advances are being made in bioinformatics to define, homogenize, and monitor the transcriptomic information in order to break the inter-analysis barrier, which is mandatory for clinical reproducibility. However certain issues remain outstanding that should be further defined and resolved in the coming years.

AUTHOR CONTRIBUTIONS

All authors contributed to manuscript writing, revision, read and approved the submitted version.

FUNDING

JB's lab is partially funded by grant PI16/00440 from Instituto de Salud Carlos III (ISCIII), cofunded by European Regional Development Fund (ERDF).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01152/full#supplementary-material>

- unique cryptic splice-site and Charcot-Marie-Tooth phenotype with early onset symptoms. *Mol. Genet. Genomic Med.* 7 (6), e00676. doi: 10.1002/mgg3.676
- Chen, L., Tovar-Corona, J. M., and Urrutia, A. O. (2012). Alternative splicing: a potential source of functional innovation in the eukaryotic genome. *Int. J. Evol. Biol.* 2012, 1–10. doi: 10.1155/2012/596274
- Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., and Smith, J. D. (2015). The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. *Am. J. Hum. Genet.* 97 (2), 199–215. doi: 10.1016/j.ajhg.2015.06.009
- Compeau, P. E. C., Pevzner, P. A., Tesler, G., Papoutsoglou, G., Roscito, J. G., and Dahl, A. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17 (1), 10–12. doi: 10.14806/ej.17.1.200
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., and McPherson, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17 (1), 13. doi: 10.1186/s13059-016-0881-8
- Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* 132, 1077–1130. doi: 10.1007/s00439-013-1331-2
- Costa, V., Angelini, C., De Feis, I., and Ciccociola, A. (2014). Uncovering the complexity of transcriptomes with RNA-SEQ. *Role Bioinforma Agric.* 2010, 207–247. doi: 10.1201/b16568-11
- Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Foley, A. R., and Bolduc, V. (2017). HHS Public Access. *Sci. Transl. Med.* 9 (386), 1–25. doi: 10.1126/scitranslmed.aal5209

- De La Chapelle, A. (2009). Genetic predisposition to human disease: allele-specific expression and low-penetrance regulatory loci. *Oncogene* 28 (38), 3345–3348. doi: 10.1038/onc.2009.194
- Desmet, F. O., Hamroun, D., Lalande, M., Collod-Bèroud, G., Claustres, M., and Bèroud, C. (2009). Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37 (9), 1–14. doi: 10.1093/nar/gkp215
- Devonshire, A. S., Elawarapu, R., and Foy, C. A. (2010). Evaluation of external RNA controls for the standardisation of gene expression biomarker measurements. *BMC Genomics* 11 (1), 662. doi: 10.1186/1471-2164-11-662
- Dey, K. K., Hsiao, C. J., and Stephens, M. (2017). Visualizing the structure of RNA-seq expression data using grade of membership models. *PloS Genet.* 13 (3), 1–23. doi: 10.1371/journal.pgen.1006599
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., and Jha, S. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinf.* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635
- Evans, C., Hardin, J., and Stoebe, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief Bioinform.* 19 (5), 776–792. doi: 10.1093/bib/bbx008
- Finotello, F., Lavezzo, E., Bianco, L., Barzon, L., Mazzon, P., and Fontana, P. (2014). Reducing bias in RNA sequencing data: a novel approach to compute counts. *BMC Bioinf.* 15 (Suppl 1), 1–12. doi: 10.1186/1471-2105-15-S1-S7
- Fresard, L., Smail, C., Smith, K. S., Ferraro, N. M., Teran, N. A., and Kernohan, K. D. (2018). Identification of rare-disease genes in diverse undiagnosed cases using whole blood transcriptome sequencing and large control cohorts. *bioRxiv* 25 (6), 408492. doi: 10.1101/408492
- Gao, L., Fang, Z., Zhang, K., Zhi, D., and Cui, X. (2011). Length bias correction for RNA-seq data in gene set analyses. *Bioinf.* 27 (5), 662–669. doi: 10.1093/bioinformatics/btr005
- Gordon, A. FASTX-Toolkit. 2010. https://github.com/agordon/fastx_toolkit.
- Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D. C., and Shyr, Y. (2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 109 (2), 83–90. doi: 10.1016/j.ygeno.2017.01.005
- Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., and Cawley, S. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 14 (3), 331–342. doi: 10.1101/gr.2094104
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2 accurate alignment of transcriptomes in the presence of insertions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36
- Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., and Iuso, A. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8, 1–11. doi: 10.1038/ncomms15824
- Lappalainen, T., Sammeth, M., Friedländer, M. R., T Hoen, P. A. C., Monlong, J., and Rivas, M. A. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nat.* 501 (7468), 506–511. doi: 10.1038/nature12531
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., and Gentleman, R. (2013). Software for computing and annotating genomic ranges. *PloS Comput. Biol.* 9 (8), 1–10. doi: 10.1371/journal.pcbi.1003118
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf.* 28 (6), 882–883. doi: 10.1093/bioinformatics/bts034
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12 (1), 323. doi: 10.1186/1471-2105-12-323
- Li, X., Nair, A., Wang, S., and Wang, L. (2015). Quality control of RNA-seq experiments. *RNA Bioinf.* 28 (16), 137–146. doi: 10.1007/978-1-4939-2291-8_8
- Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., and Im, H. K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50 (1), 151–158. doi: 10.1038/s41588-017-0004-9
- Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinf.* 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656
- Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., and Hosseini, S. M. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* 20 (4), 435–443. doi: 10.1038/gim.2017.119
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease?. *FEBS Lett.* 579 (9), 1900–1903. doi: 10.1016/j.febslet.2005.02.047
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8
- Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read sequencing emerging in medical genetics. *Front. Genet.* 10, 1–14. doi: 10.3389/fgene.2019.00426
- Melè, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., et al. The human transcriptome across tissues and individuals. 2015.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5 (7), 621–628. doi: 10.1038/nmeth.1226
- Nelakuditi, V. PAGE: Panel Analysis of Gene Expression. <https://page.ccm.sickkids.ca/About>.
- Pagani, F., Raponi, M., and Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci.* 102 (18), 6368–6372. doi: 10.1073/pnas.0502288102
- Pertea, M., Lin, X., and Salzberg, S. L. (2011). 2001 GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* 29 (5), 1185–1190. doi: 10.1093/nar/29.5.1185
- Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., and Gibellini, F. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* 18 (7), 696–704. doi: 10.1038/gim.2015.148
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., and Gastier-Foster, J. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi: 10.1038/gim.2015.30
- Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinf.* 12 (1), 480. doi: 10.1186/1471-2105-12-480
- Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinf.* 27 (17), 2325–2329. doi: 10.1093/bioinformatics/btr355
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinf.* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17 (1), 19–32. doi: 10.1038/nrg.2015.3
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18 (8), 472–482. doi: 10.1016/j.molmed.2012.06.006
- Slaugenaupt, S. A., Blumenfeld, A., Gill, S. P., Leyne, M., Mull, J., and Cuajungco, M. P. (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* 68 (3), 598–605. doi: 10.1086/318810
- Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., and Heywood, S. (2017). The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* 136 (6), 665–677. doi: 10.1007/s00439-017-1779-6
- Tazi, J., Bakkour, N., and Stamm, S. (2009). Alternative splicing and disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1792 (1), 14–26. doi: 10.1016/j.bbdis.2008.09.017
- Valle, L., Serena-Acedo, T., Liyanarachchi, S., Hampel, H., Comeras, I., and Li, Z. (2008). Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Sci.* 321 (5894), 1361–1365. doi: 10.1126/science.1159397

- Wai, H., Douglas, A. G. L., and Baralle, D. (2019). RNA splicing analysis in genomic medicine. *Int. J. Biochem. Cell Biol.* 108, 61–71. doi: 10.1016/j.biocel.2018.12.009
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., and Mayr, C. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nat.* 456 (7221), 470–476. doi: 10.1038/nature07509
- Wang, L., Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J. E., and Dasari, S. (2016). Measure transcript integrity using RNA-seq data. *BMC Bioinf.* 17 (1), 1–16. doi: 10.1186/s12859-016-0922-z
- Williams, A. G., Thomas, S., Wyman, S. K., and Holloway, A. K. (2014). RNA-seq data: challenges in and recommendations for experimental design and analysis. *Curr. Protoc. Hum. Genet.* 83, 11.13.1–11.13.20. doi: 10.1002/0471142905.hg1113s83
- Wood, D. L. A., Nones, K., Steptoe, A., Christ, A., Harliwong, I., and Newell, F. (2015). Recommendations for accurate resolution of Gene and isoform allele-specific expression in RNA-seq data. *PLoS One* 10 (5), 1–27. doi: 10.1371/journal.pone.0126911
- Wu, P. Y., Phan, J. H., and Wang, M. D. (2013). Assessing the impact of human genome annotation choice on RNA-seq expression estimates. *BMC Bioinf.* 14 Suppl 1 (Suppl 1), 1–13. doi: 10.1186/1471-2105-14-S11-S8
- Yeo, G., and Burge, C. B. MaxEntScan. 2004; <https://github.com/esebesty/maxentscan>.

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

Copyright © 2019 Marco-Puche, Lois, Benítez and Trivino. 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.



Comprehensive Assessment of *BARD1* Messenger Ribonucleic Acid Splicing With Implications for Variant Classification

Logan C. Walker^{1*}, Vanessa Lilian Lattimore¹, Anders Kvist², Petra Kleiblova^{3,4}, Petra Zemankova⁴, Lucy de Jong¹, George A. R. Wiggins¹, Christopher Hakkaart¹, Simone L. Cree¹, Raquel Behar⁵, Claude Houdayer⁶, kConFab Investigators^{7,8}, Michael T. Parsons⁹, Martin A. Kennedy¹, Amanda B. Spurdle⁹ and Miguel de la Hoya⁵

OPEN ACCESS

Edited by:

Eladio Andrés Velasco,
Institute of Biology and Molecular
Genetics (IBGM), Spain

Reviewed by:

Lucie Grodecká,
Center of Cardiovascular and
Transplant Surgery (Czechia),
Czechia
Elena Bueno Martínez,
Spanish National Research Council
(CSIC), Spain

*Correspondence:

Logan C. Walker
logan.walker@otago.ac.nz

Specialty section:

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

Received: 12 August 2019

Accepted: 21 October 2019

Published: 19 November 2019

Citation:

Walker LC, Lattimore VL, Kvist A,
Kleiblova P, Zemankova P,
de Jong L, Wiggins GAR, Hakkaart C,
Cree SL, Behar R, Houdayer C,
Parsons MT, Kennedy MA,
Spurdle AB and de la Hoya M (2019)
Comprehensive Assessment of
BARD1 Messenger Ribonucleic
Acid Splicing With Implications
for Variant Classification.
Front. Genet. 10:1139.
doi: 10.3389/fgene.2019.01139

¹ Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand, ² Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden, ³ Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czechia, ⁴ Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czechia, ⁵ Molecular Oncology Laboratory, CIBERONC, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain, ⁶ Department of Genetics, F76000 and Normandy University, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen University Hospital, Rouen, France, ⁷ Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia, ⁸ Research Department, Peter MacCallum Cancer Center, Melbourne, VIC, Australia, ⁹ Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

Introduction: Case-control analyses have shown *BARD1* variants to be associated with up to >2-fold increase in risk of breast cancer, and potentially greater risk of triple negative breast cancer. *BARD1* is included in several gene sequencing panels currently marketed for the prediction of risk of cancer, however there are no gene-specific guidelines for the classification of *BARD1* variants. We present the most comprehensive assessment of *BARD1* messenger RNA splicing, and demonstrate the application of these data for the classification of truncating and splice site variants according to American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines.

Methods: Nanopore sequencing, short-read RNA-seq (whole transcriptome and targeted), and capillary electrophoresis analysis were performed by four laboratories to investigate alternative *BARD1* splicing in blood, breast, and fimbriae/ovary related specimens from non-cancer affected tissues. Splicing data were also collated from published studies of nine different tissues. The impact of the findings for PVS1 annotation was assessed for truncating and splice site variants.

Results: We identified 62 naturally occurring alternative spliced *BARD1* splicing events, including 19 novel events found by next generation sequencing and/or reverse transcription PCR analysis performed for this study. Quantitative analysis showed that naturally occurring splicing events causing loss of clinically relevant domains or nonsense mediated decay can constitute up to 11.9% of overlapping natural junctions, suggesting that aberrant splicing can be tolerated up to this level. Nanopore sequencing of whole

BARD1 transcripts characterized 16 alternative isoforms from healthy controls, revealing that the most complex transcripts combined only two alternative splicing events. Bioinformatic analysis of ClinVar submitted variants at or near *BARD1* splice sites suggest that all consensus splice site variants in *BARD1* should be considered likely pathogenic, with the possible exception of variants at the donor site of exon 5.

Conclusions: No *BARD1* candidate rescue transcripts were identified in this study, indicating that all premature translation-termination codons variants can be annotated as PVS1. Furthermore, our analysis suggests that all donor and acceptor (IVS+/-1,2) variants can be considered PVS1 or PVS1_strong, with the exception of variants targeting the exon 5 donor site, that we recommend considering as PVS1_moderate.

Keywords: breast cancer, mRNA splicing, nanopore sequencing, RNAseq analysis, variant classification, ACMG

INTRODUCTION

The *BARD1* gene (MIM# 601593) was identified in 1996 as the result of a yeast two-hybrid screen for proteins that interact with the breast and ovarian cancer associated BRCA1 protein (Wu et al., 1996). The *BARD1* reference transcript contains 11 exons and produces a full length 777 amino acid protein which is structurally related to BRCA1 as both contain N-terminal RING finger domains and two carboxy-terminal (BRCT) domains (Miki et al., 1994; Wu et al., 1996). The interaction of BARD1 to BRCA1 is mediated by their respective RING domains leading to the proposal that *BARD1* is a candidate breast and ovarian cancer predisposing gene. Various lines of evidence suggest BARD1 may act as a potent tumor suppressor, including the ability to induce TP53-dependent apoptosis (Irminger-Finger et al., 2001), and the observation that homozygous loss of *BARD1* in mice is embryonically lethal, mimicking the properties of *BRCA1* (McCarthy et al., 2003). Furthermore, numerous studies of individuals who have a family history of breast cancer have found rare and functionally deleterious variants in *BARD1* (Ishitobi et al., 2003; Karppinen et al., 2004; De Brakeleer et al., 2010; Ratajska et al., 2012). Case-control analyses have shown *BARD1* loss of function variants to be associated with a low (< 2-fold) to moderate (> 2-fold) increase in risk of breast cancer (Couch et al., 2017; Kurian et al., 2017; Slavin et al., 2017) and up to five-fold increase in risk of triple negative breast cancer (Shimelis et al., 2018). However, the utility of *BARD1* sequencing to identify actionable pathogenic variants in a clinical setting remains undefined and requires a thorough investigation of all possible ways a variant might lead to loss of function. Sequence variants play an important role in the regulation of pre-messenger RNA (mRNA) splicing (Scotti and Swanson, 2016), and there is an established link between aberrant splicing of cancer predisposition genes and breast cancer risk (Walker et al., 2010; Whiley et al., 2010; Whiley et al., 2011). Thus, investigating the role of *BARD1* variants in the production of aberrant mRNA transcripts can be used to assess the likelihood of sequence variants causing functional changes that confer pathogenicity (Walker et al., 2013).

Determining the effect of sequence variants on the expression of mRNA splice isoforms and interpreting which spliceogenic

variants are potentially deleterious is a major challenge. Reverse transcriptase-polymerase chain reaction (RT-PCR) has been the major technology used to assess mRNA splicing in a variety of cancer susceptibility genes, including *BARD1*. However, incorrect positioning of PCR primers can result in key splicing events not being detected and lead to a misinterpretation of splicing events. For example, *BRCA1* c.594-2A > C was originally classed as pathogenic and associated with an aberrant mRNA profile that included exon 10 skipping (out-of-frame) but no consideration was given to natural alternative splicing (Tesoriero et al., 2005). More recently, we showed that *BRCA1* c.594-2A > C occurs in *cis* with *BRCA1* c.641A > G and should not be considered as a high-risk pathogenic variant because the out-of-frame splicing alteration did not affect the predominant alternative spliced event, Δ(E9_E10), which retains tumor suppressor activity (de la Hoya et al., 2016).

Massively parallel complementary DNA sequencing (RNA-seq) has further advanced our ability to characterize and quantify gene transcripts, and will therefore become a key technology for measuring gene expression changes in clinical diagnostics. Recent studies have begun to demonstrate the utility of RNA-seq for identifying mRNA splicing events in breast cancer susceptibility genes, including *BRCA1* (Davy et al., 2017; de Jong et al., 2017; Hojny et al., 2017), *BRCA2* (Davy et al., 2017), *PALB2* (Lopez-Perolio et al., 2019), and *BARD1* (Davy et al., 2017). These studies revealed that key advantages of using RNA-seq over RT-PCR is the ability to quantitatively assess multiple splicing events across the whole transcript in one sequencing assay. Furthermore, long-read nanopore sequencing is capable of generating sequences of full-length transcripts and thus can resolve complex exon structures of full-length mRNAs from genes expressing a large number of isoforms (de Jong et al., 2017). Several reports have profiled *BARD1* transcripts to characterize “naturally occurring” mRNA splice isoforms across multiple tissue types (Li et al., 2007; Lombardi et al., 2007; Sporn et al., 2011; Bosse et al., 2012; Zhang et al., 2012; Pilyugin and Irminger-Finger, 2014; Davy et al., 2017). However, despite previous published reports of *BARD1* splicing, current catalogues of alternatively spliced events (e.g., Ensembl—ENSG00000138376) only account for a fraction of transcripts associated with this gene.

We present the most comprehensive assessment of *BARD1* mRNA splicing generated by both RT-PCR and RNA-seq (long-read and short-read) platforms across multiple tissue types. Furthermore, we also utilize American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (Richards et al., 2015), bioinformatic splicing, and population frequency data to evaluate potential pathogenicity of *BARD1* variants located at canonical splice sites. Results from our study provide an important basis to standardize the clinical classification and reporting of *BARD1* genetic variants.

MATERIALS AND METHODS

Ribonucleic Acid Samples

RNA samples assessed in this study were isolated from different tissue types, including 47 human lymphoblastoid cell lines (LCLs) derived from female healthy controls, an epithelial enriched area of nine healthy breast samples from women with breast tumors (SCAN-B study, ClinicalTrials.gov identifier: NCT02306096), two normal fimbria tissues obtained from prophylactic oophorectomies performed in post-menopausal women without cancer, commercially available RNA from one non-malignant breast tissue (Clontech 636576), and one pool of three non-malignant ovarian tissues (Clontech 636555) (**Supplementary Figure S1**).

Nanopore-Sequencing—MinION Platform

Laboratory 1

The Oxford Nanopore MinION Genomic DNA sequencing of LCL RNA was carried out as previously described (de Jong et al., 2017). Briefly, PCR products were prepared for sequencing using the Nanopore Sequencing Kit SQK-NSK007 (R9 Version). Primer sequences for *BARD1* exons 1 and 11 are as follows: 5'-CTCGACCGCCTGGAGAAG-3' and 5'-CTGGCTTGGGCTTTCTACTG-3'. The raw electrical signal was uploaded to Metrichor (version 1.107), using the 2D Basecalling RNN for SQK-NSK007. Full-length alternative isoform analysis of RNA (FLAIR; <https://github.com/BrooksLabUCSC/flair>) was used to identify novel and known isoforms of *BARD1*. Sequence reads in FASTA format were aligned to the GRCh38 using the align module, which implements minimap2 with the splice option. Aligned reads were then corrected and collapsed using the respective modules of FLAIR with default settings. Annotation for known isoforms were provided by GENCODE (v29).

Targeted Ribonucleic Acid Sequencing—Illumina Platform

Laboratory 2

RNA-sequencing of a 36 LCLs from kConFab [18 sample pairs with/without nonsense-mediated mRNA decay (NMD) inhibition] was carried out using Kapa RNA HyperPrep Kit (Roche) according to manufacturer. Briefly, 250 ng of total RNA were chemically fragmented (mean fragment length 200 bp).

PCR amplifications were run for 8 and 12 cycles for pre- and post-hybridization PCR, respectively. Plexes of six barcoded samples (166 ng of each) were hybridized with custom-designed SeqCap EZ Choice CZECA NCA v1.2, Roche (Soukupova et al., 2018). Libraries were paired-end sequenced on NextSeq 500 with NextSeq 500/550 Mid Output Kit v2.5 (150 cycles). Splice junctions were included if they were identified in at least three LCLs with an average of more than four reads per LCL.

Whole Ribonucleic Acid sequencing—Illumina Platform

Laboratory 1

RNA-sequencing of a single LCL from a female healthy control was carried out as described previously (Lattimore et al., 2018). Briefly, libraries were prepared from total RNA using poly(A) enrichment of the mRNA (mRNA-Seq) to remove ribosomal RNA (rRNA). The calculation of the percentage of junction reads was carried out as described previously (Davy et al., 2017).

Laboratory 3

RNA-sequencing of normal breast and fimbria tissue was carried out as previously described [(23) **Supplemental Material** section 1.2 therein]. Briefly, fresh breast tissue was preserved in RNAlater (Ambion) and fimbriae tissue was fresh frozen. RNA was extracted using AllPrep (Qiagen) and libraries prepared with a modified version of the dUTP (Deoxyuridine Triphosphate) method (breast samples) or the TruSeq Stranded mRNA Library Prep Kit (fimbriae samples, Illumina, San Diego, CA). Libraries were paired-end sequenced on an Illumina HiSeq 2000 (2x50 bp, two breast samples) or a NextSeq 500 (2x75 bp, remaining seven samples). Sequence reads were analyzed as described previously (Lopez-Perolio et al., 2019).

Sequencing Data Availability

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

Reverse Transcription Polymerase Chain Reaction Assays

Laboratory 4

RT-PCR analysis was carried out on 10 LCLs, breast tissue (Clontech 636576), and one pool of three non-malignant ovarian tissues (Clontech 636555) as previously described (Lopez-Perolio et al., 2019). Primer sequence and details as shown in **Supplementary Table S1**.

Annotation of Alternative Splicing Events

Alternative splicing events were annotated according to the Human Genome Variation Society (HGVS) guidelines, using the Ensembl transcript ENST00000260947.8 (NCBI RefSeq NM_000465.3) as a reference. Splicing events were also coded as described previously (Lopez-Perolio et al., 2019) using the following symbols: Δ (skipping of reference exonic sequences), ▼ (inclusion of reference intronic sequences), E (exon), I

(intron), p (acceptor shift), q (donor shift), and int (interstitial deletion within an exon). Where possible, the exact number of nucleotides skipped (or retained) is indicated. All *BARD1* alternative splicing events reported are predicted to alter the encoded protein. To decide if the truncated/altering region is critical to protein function, we considered the RING, ARD (ankyrin repeat domain), and BRCT domains as shown in **Supplementary Table S2** and **Figure 1**.

Classification of Splice Site Variants Using American College of Medical Genetics and Genomics and the Association for Molecular Pathology Guidelines

Adaptation of the ACMG/AMP PVS1 decision tree (Abou Tayoun et al., 2018) to *BARD1* donor and acceptor “consensus” dinucleotide (IVS+/- 1,2) variants is detailed in the **Supplementary Methods**.

RESULTS AND DISCUSSION

BARD1 Isoform Discovery and Annotation

We present a comprehensive *BARD1* mRNA splicing catalogue from splicing assays of 12 tissue types (normal and cancer tissue) derived from this study and seven publications (**Table 1**). Targeted and whole RNA-seq performed by contributing laboratories produced 299,479 reads aligned to exon-exon junctions at the *BARD1* locus. Targeted RNA-seq of the 36 LCLs by laboratory 2 yielded 292,143 *BARD1* junction reads, whole RNA-Seq yielded

6,656 junction reads of a single LCL (laboratory 1), and 573 and 107 junction reads respectively from 9 breast and 2 fimbria samples (laboratory 3). A total of 62 alternative *BARD1* splicing events were identified in this study. Of these, 19 novel splicing events were found in this study by four contributing laboratories using nanopore sequencing, short-read RNA-seq, and/or RT-PCR. The most commonly found alternative splicing event across studies was the out-of-frame $\Delta(E4)$, identified by all technologies in all but one of the tissues assayed. Furthermore, skipping events that included exon 4 were observed in 28 isoforms, suggesting that the absence of this 950 nucleotide exon in a small fraction (up to 3%; **Figure 1**) of *BARD1* transcripts is tolerated by different cell types. We observed no *BARD1* splicing events that were expressed exclusively in breast and/or ovarian tissue. LCLs have been a common cell type used for *in vitro* assays assessing splicing changes in patients with potential spliceogenic variants. Our data showed that there were 12 splicing events [(IVS1+4279 ∇ 98, $\Delta(E2q)$, $\Delta(E2_E7)$, $\Delta(E3_E4,E7)$, $\Delta(E4q137)$, $\Delta(E4int104)$, $\Delta(E5)$, $\Delta(E5_E9)$, IVS6+4684 ∇ 67, $\nabla(I7q4)$, IVS9-6318 ∇ 92, IVS9+5946 ∇ 1015) specific to LCLs. Three of these events were detected exclusively by nanopore sequencing ($\Delta(E2_E7)$, $\Delta(E3_E4,E7)$, $\Delta(E5)$) and eight were detected by short-read RNA-seq which was used for more LCL samples and with a greater depth of coverage (higher number of junction reads) than for any other tissue type (**Table 1**).

To compare the splicing data by assay used (nanopore sequencing/short-read RNA-seq vs. RT-PCR), we examined *BARD1* mRNA isoforms detected exclusively by one technology. In addition to the 12 splicing events detected by sequencing but not RT-PCR (listed above), 17 alternative splicing events were detected by RT-PCR but not characterized by long- or short-read

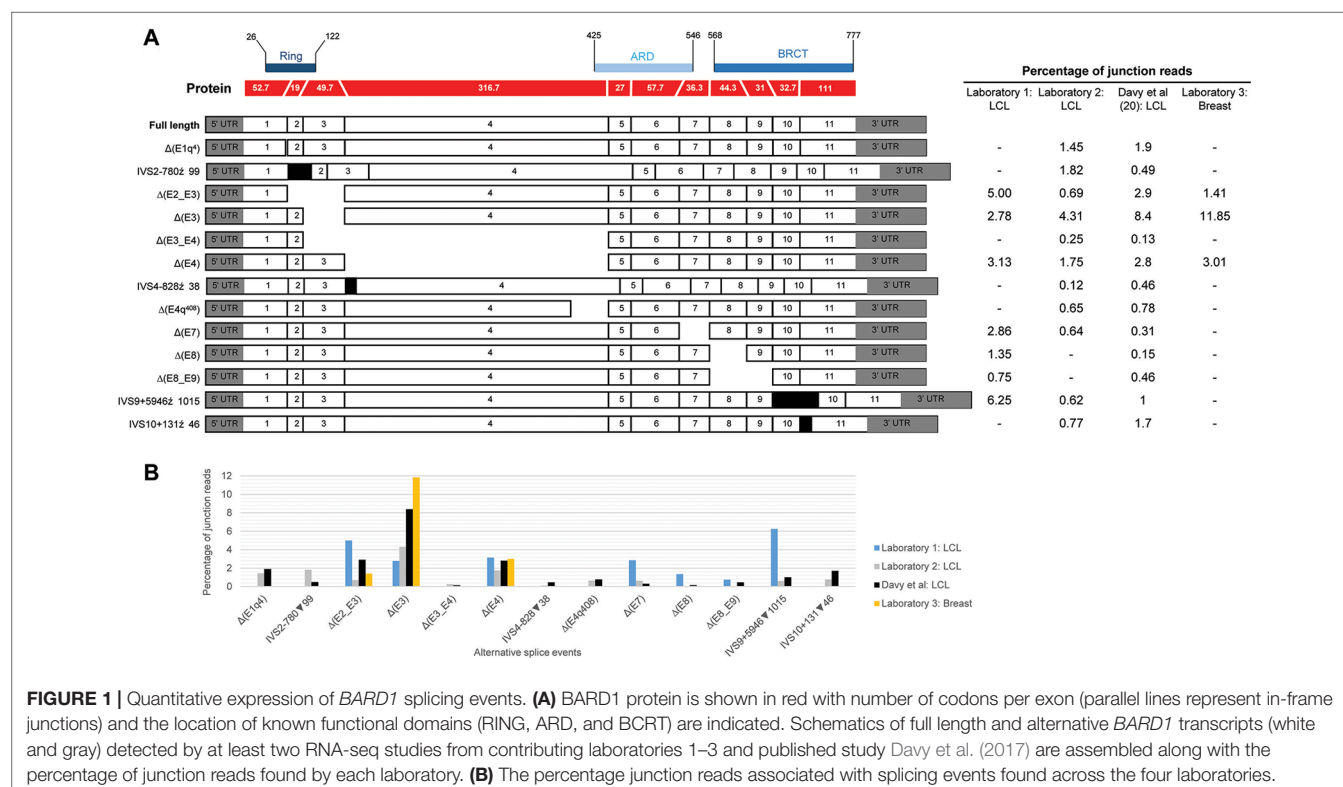


FIGURE 1 | Quantitative expression of *BARD1* splicing events. **(A)** *BARD1* protein is shown in red with number of codons per exon (parallel lines represent in-frame junctions) and the location of known functional domains (RING, ARD, and BRCT) are indicated. Schematics of full length and alternative *BARD1* transcripts (white and gray) detected by at least two RNA-seq studies from contributing laboratories 1–3 and published study Davy et al. (2017) are assembled along with the percentage of junction reads found by each laboratory. **(B)** The percentage junction reads associated with splicing events found across the four laboratories.

RNA-seq assays (Table 1). Our short-read RNA-seq analyses were not able to characterize the complete exon structure for 12 of these 17 because they are compound events combining multiple non-contiguous splicing events. These results highlight a key limitation with assays that are unable to examine the entire transcript.

Many of the detected *BARD1* splicing events occurred at low levels (<7% of the transcript pool) but were identified using newer and older technologies. The differences we have observed across different laboratories and published studies are possibly due to multiple factors including, different technologies with differing sensitivities, different sample types, different culture conditions, and different study cohort sizes. Such variability has also been observed between laboratories that used different cell processing and assay protocols for *BRCA1* and *BRCA2* isoform detection (Whiley et al., 2014). Events detected by only one laboratory or study are most likely due to reduced sensitivity of others methods to detect that particular event. However, we cannot exclude the possibility that some of these events maybe artifacts.

Co-Occurring *BARD1* Splicing Events

Most RNA-seq technologies derive partial information about transcript structure due to targeting relatively short transcript sequences. Determining whether *BARD1* transcripts lead to abnormal and potentially deleterious proteins requires knowledge relating to the complete sequence structure of the coding isoforms. Using MinION (nanopore) sequencing of PCR amplified *BARD1* mRNA transcripts, we were able to sequence the full-length isoform along with 16 alternatively spliced isoforms accounting for 18 of the 62 individual splicing events (Table 1, Supplementary Figure S2). Two of the three novel isoforms found exclusively using this technology were out-of-frame [$\Delta(E2_E7)$ and $\Delta(E3_E4_E7)$] and one was in-frame [$\Delta(E5)$]. *BARD1* exon splicing events, such as $\Delta(E2_E4)$, $\Delta(E4)$, and $\Delta(E8)$, have been shown to co-occur independently in single transcripts as well as combined with other events to generate more complex isoforms.

Based on available data, the most complex *BARD1* transcript structures identified involved two alternative splicing events and was observed in 15 of the alternative transcripts (Table 1).

Although nanopore sequencing was conducted on PCR products generated from an LCL treated with an NMD inhibitor, we were not able to identify all junctions identified by short-read sequencing. This is likely a limitation of only sequencing amplicons derived from PCR assays using a single cell line. It is also important to note that we sequenced targeted amplicons which included exons 1 and 11, leaving the possibility that we excluded transcripts that do not contain these regions, such as $\Delta(E1-E4p)$ (Table 1). Analysis of truncated nanopore reads that do not contain exons 1 and 11 gave rise to several additional low confidence splicing events (Supplementary Figure S2). Results from the FLAIR bioinformatic analysis tool were presented in this study as this method has previously been shown to identify high-confidence spliced isoforms compared to other tools, such as Genomic Mapping and Alignment Program (GMAP) (Tang et al., 2018). Our re-analysis of nanopore sequence reads using the GMAP tool generated a list of 49 alternative *BARD1* transcripts including 11 splicing events that were not detected using the FLAIR analysis or

by short-read RNA-seq and/or RT-PCR methods (Supplementary Table S3). These results suggest that the GMAP tool may be more sensitive than FLAIR, although the large number of novel splicing events detected also suggests a higher rate of false positive results, as previously reported (Tang et al., 2018).

Relative Levels of *BARD1* Splicing

Relative expression levels of splicing events were determined using short-read RNA-seq analysis of LCLs cultured with and without an inhibitor of nonsense mediated decay (NMD). The most highly expressed alternative splicing events identified both in this study and that published by Davy et al. (Davy et al., 2017), using cells not treated with NMD inhibitors, produced out-of-frame transcripts and are shown in Figure 1. To assess the effect of NMD inhibitors on expression of splicing events we compared the percentage of sequenced junction reads corresponding to alternative splicing in treated cells with alternative splicing in non-treated cells. Results showed variable expression of splice junctions between the two groups (Supplementary Figure S3). For example, $\Delta(E4)$ is predicted to lead to the activation of a premature stop codon in exon 5 leading to NMD, however both laboratory 1 and 2 found that the percentage of junction reads for this event was greater in non-treated cells. Relatively low expression variability of *BARD1* splice junctions was observed between LCLs from laboratory 2 suggesting greater inter-laboratory variability than intra-laboratory variability (Supplementary Figure S3).

With the exception of $\Delta(E4)$, there was noticeable variability in the levels of splicing events detected across laboratories (Figure 1). None of these events exceeded 9% of the overlapping natural junctions in LCLs. However, $\Delta(E3)$ was expressed in breast tissue at ~12% relative to the overlapping natural junctions. Since $\Delta(E3)$, and the other most highly expressed events, produce out-of-frame transcripts, this suggests that aberrant splicing is tolerated to at least this level. Interestingly, the level of $\Delta(E3)$ expression in colorectal tumor tissue has been shown to be associated with tumorigenesis and progression (Zhang et al., 2012), although it is unclear whether $\Delta(E3)$ expression levels in normal cells is associated with risk. Each exon deleted from the alternative transcripts overlapped a known functional domain of *BARD1*. The possible function of most isoforms identified to date remains unknown. However, several studies have shown that $\Delta(E2_E3)$ uses an alternative in-frame start codon and encodes a protein which has a proliferative function despite losing the RING domain and therefore the ability to bind to *BRCA1* (Li et al., 2007; Ryser et al., 2009; Zhang et al., 2012). Furthermore, evidence suggests that the in-frame isoforms $\Delta(E3_E6)$ and $\Delta(E3_E7)$ also have a role in cellular proliferation (Li et al., 2007). Apart from the $\Delta(E2_E3)$ isoform, there is little evidence to suggest that other out-of-frame transcripts [e.g., $\Delta(E3)$] use an alternative open reading frame to encode functional proteins.

BARD1 Splicing and Interpretation for Variant Classification

Abou Tayoun et al. recently proposed a decision tree for interpreting the loss of function PVS1 ACMG/AMP criterion

TABLE 1 | List of *BARD1* isoforms across 12 tissue types.

Splicing event description	Transcript structure	Functional annotation	LCL					PBL	Breast		Breast tumor	Fimbria	Ovary	Cell line	Fetal sympathetic ganglia	Neuroblastoma	HeLa cells	Colon cancer		Trophoblasts
			NS	RS	TRS	RP	TRS	RP	RS	RP	RP	RS	RP	RP	RP	RP	RP	RP	RP	RP
			Laboratory 1 (n=1)	Laboratory 1 (n=1)	Laboratory 2 (n=36)	Laboratory 4 (n=10)	Davy et al., 2017	Lombardi et al., 2007	Laboratory 3 (n=9)	Laboratory 4 (n=1)	Lombardi et al., 2007	Laboratory 3 (n=2)	Laboratory 4 (n=3)	Pilyugin and Irminger-Finger, 2014	Bosse et al., 2012	Bosse et al., 2012	Li et al., 2007	Sporn et al., 2011	Zhang et al., 2012	Li et al., 2007
Full length	-	IF	+	N	N	+	N	+	N		+	N		+	+	+	+	+	+	+
Δ(E1_E4p)	r.155_?del	?												+						
Δ(E1q4)	r.155_158del	FS		+	+	+	+		+	+			+		+					
Δ(E1q4_E4)	r.155_1314del	IF		N	N		N		N			N			+					
IVS1+4279▼98	r.158_159ins158+4279_158+4376	FS		+	+		+													
Δ(E2)	r.159_215del	IF			+												+	+		
Δ(E2q)	r.212_215del	FS			+		+													
Δ(E2_E3)	r.159_364del	FS		+	+	+	+		+	+		+	+	+	+	+	+		+	+
Δ(E2_E4-E9)	r.159_215+365_1903del	IF		N	N	+	N		N	+		N	+	+						
Δ(E2_E4)	r.159_1314del	FS	+	+	+	+		+		+	+				+	+		+		+
Δ(E2_E4,E7)	r.159_1314+1569_1677del	FS		N	N		N		N			N						+		
Δ(E2_E4_E8)	r.159_1314+1678_1810del	FS	+	N	N		N		N			N						+		
Δ(E2_E4,E8_E9)	r.159_1314+1678_1903del	FS		N	N		N		N			N						+		
Δ(E2_E4,E8p)	r.159_1314+1678_?del	?		N	N		N		N			N			+					
Δ(E2_E5)	r.159_1395del	FS													+					
Δ(E2_E6)	r.159_1568del	IF	+		+			+			+			+	+	+	+	+	+	+
Δ(E2_E6,E8)	r.159_1568+1678_1810del	FS		N	N		N	+	N		+	N			+	+	+	+	+	+
Δ(E2_E6,E8_E9)	r.159_1568+1678_1903del	FS		N	N	+	N		N			N			+					
Δ(E2_E7)	r.159_1677del	FS	+																	
Δ(E2_E9)	r.159_1903del	FS	+		+	+	+	+		+	+		+	+	+	+	+	+	+	+
IVS2-780▼99	r.215_216ins216-780_216-682	IF		+	+	+	+			+				+	+					
Δ(E3)	r.216_364del	FS	+	+	+	+	+		+	+		+	+			+		+	+	
Δ(E3_E4)	r.216_1314del	FS	+	+	+	+	+	+		+	+		+					+		
Δ(E3_E4,E7)	r.216_1314+1569_1677del	FS	+	N	N		N		N			N								
Δ(E3_E4,E8)	r.216_1314+1678_1810del	FS		N	N		N		N			N				+				
Δ(E3_E5)	r.216_1395del	FS			+										+					
Δ(E3_E5_E9)	r.216_364+1315_1903del	IF		N	N		N		N			N				+	+	+	+	+
Δ(E3_E6)	r.216_1568del	IF	+	+	+	+		+		+	+			+	+	+	+	+	+	+
Δ(E3_E6,E8)	r.216_1568+1678_1810del	FS		N	N		N	+	N		+	N				+	+	+	+	+
Δ(E3_E7)	r.216_1677del	FS	+		+	+							+			+				
Δ(E3_E9)	r.216_1903del	FS	+		+	+										+		+		
Δ(E3q_E10)	r.361_2001del	IF													+					
Δ(E4)	r.365_1314del	FS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Δ(E4q137)	r.1178_1314del	FS		+	+											+	+	+	+	+
Δ(E4q408)	r.907_1314del	IF		+	+	+	+			+				+	+			+	+	
Δ(E4int104)	r.718_821del	FS		+	+					+						+		+	+	
Δ(E4int695)	r.472_1166del	FS		+	+	+														
Δ(E4_E6p)	r.365_?	?		N	N		N		N			N				+				
Δ(E4_E6)	r.365_1568del	FS		+	+	+							+		+					
Δ(E4q_E8q)	r.7_?	?		N	N		N		N			N				+				
Δ(E4_E7)	r.365_1677del	FS			+		+									+			+	
Δ(E4_E8_E9)	r.365_1314+1678_1903del	IF	+	N	N		N		N			N				+		+	+	+
Δ(E4_E9)	r.365_1903del	IF	+		+	+				+				+		+	+	+	+	+
IVS4-828▼38	r.1314_1315ins1315-828_1315-791	FS		+	+	+				+			+							
Δ(E5)	r.1315_1395del	IF	+																	
Δ(E5_E9)	r.1315_1903del	FS			+															
Δ(E5_E10)	r.1315_2001del	IF				+														
▼(I5q3)	r.1395_1396ins1395+1_1395+3	IF		+	+					+										
Δ(E6q2)	r.1567_1568del	FS		+		+				+										
IVS6+4684▼67	r.1568_1569ins1568+4684_1568+4752	FS		+																
Δ(E7)	r.1569_1677del	FS	+	+	+	+	+			+			+							
Δ(E7p35)	r.1569_1603del	FS		+	+	+				+			+							
Δ(E7_E9)	r.1569_1903del	FS			+		+		+			+							+	
▼(I7q4)	r.1677_1678ins1677+1_1677+4	FS		+	+															
Δ(E8)	r.1678_1810del	FS		+		+	+			+									+	
Δ(E8_E9)	r.1678_1903del	FS		+	+		+												+	
Δ(E9)	r.1811_1903del	IF				+				+										
Δ(E9p30)	r.1811_1840del	IF		+	+	+				+			+							
▼(I9q4)	r.1903_1904ins1903+1_1903+4	FS		+	+	+				+										
IVS9+5946▼1015	r.1903_1904ins1903+5946_1904+6960	FS			+		+													
IVS9-6318▼92	r.1903_1904ins1904-6318_1904-6227	FS		+	+															
Δ(E10q4)	r.1998_2001del	FS				+	+			+			+							
IVS10+131▼46	r.2001_2002+131_2001+176	FS		+	+	+	+			+			+							

+, splicing events detected; N, splicing events not detectable from our analyses of short-read sequences; NS, nanopore sequencing; RP, RT-PCR; RS, whole RNA-seq; TRS, large

¹splicing events identified in this study are shown in bold.

TABLE 2 | Classification of canonical *BARD1* splice site variants using American College of Medical Genetics and Genomics and the Association for Molecular Pathology guidelines.

Exon	Splice site	Splice site variant (Ensembl/CinVar)	In vitro splicing data?	gnomAD (alleles)	ClinVar (review status)	PVS1	PM2 ^a	Proposed ACMG/AMP classification ^b
1	158+1,+2	c.158+1G > T	No	0	Likely pathogenic (*)	PVS1	Yes	Likely pathogenic
2	159-1,-2	c.159-2A > G	No	1	Not reported	PVS1_strong	Yes	Likely pathogenic
2	159-1,-2	c.159-1G > T	No	0	Likely pathogenic (**)	PVS1_strong	Yes	Likely pathogenic
2	215+1,+2	215+2T > C	No	0	Likely pathogenic (*)	PVS1_strong	Yes	Likely pathogenic
3	216-1,-2	c.216-1G > A	No	0	Not reported	PVS1	Yes	Likely pathogenic
3	364+1,+2	Not reported	No	0	Not reported	PVS1	Yes	Likely pathogenic
4	365-1,-2	c.365-1G > T	No	0	Likely pathogenic (*)	PVS1_strong	Yes	Likely pathogenic
4	365-1,-2	c.365-2A > G	No	1	Not reported	PVS1_strong	Yes	Likely pathogenic
4	1314+1,+2	c.1314+1G > A	No	1	Likely pathogenic (**)	PVS1_strong	Yes	Likely pathogenic
5	1315-1,-2	c.1315-2A > G	Δ(E5) with RT-PCR E4-E6	1	Not reported	PVS1_strong	Yes	Likely pathogenic
5	1395+1,+2	c.1395+1dup	No	0	Likely pathogenic (**)	PVS1_moderate	Yes	Uncertain significance
6	1396-1,-2	Not reported	No	0	Not reported	PVS1	Yes	Likely pathogenic
6	1568+1,+2	c.1568+2T > C	No	0	Likely pathogenic (*)	PVS1	Yes	Likely pathogenic
7	1569-1,-2	Not reported	No	0	Not reported	PVS1	Yes	Likely pathogenic
7	1677+1,+2	c.1677+1G > C	No	0	Likely pathogenic (*)	PVS1	Yes	Likely pathogenic
7	1677+1,+2	c.1677+1G > A	No	1	Not reported	PVS1	Yes	Likely pathogenic
8	1678-1,-2	c.1678-1G > T	No	2	Not reported	PVS1	Yes	Likely pathogenic
8	1810+1,+2	c.1810+1G > A	No	0	Likely pathogenic (*)	PVS1	Yes	Likely pathogenic
8	1810+1,+2	c.1810+2T > G	No	0	Likely pathogenic (**)	PVS1	Yes	Likely pathogenic
9	1811-1,-2	c.1811-1G > A	No	0	Pathogenic (*)	PVS1_strong	Yes	Likely pathogenic
9	1903+1,+2	c.1903+1G > A	No	1	Not reported	PVS1_strong	Yes	Likely pathogenic
9	1903+1,+2	c.1903+1G > T	No	0	Likely pathogenic (*)	PVS1_strong	Yes	Likely pathogenic
10	1904-1,-2	c.1904-2A > T	No	0	Likely pathogenic (*)	PVS1_strong	Yes	Likely pathogenic
10	2001+1,+2	c.2001+1G > T	No	1	Uncertain significance (*)	PVS1_strong	Yes	Likely pathogenic
10	2001+1,+2	c.2001+1G > C	No	0	Likely pathogenic (**)	PVS1_strong	Yes	Likely pathogenic
10	2001+1,+2	c.2001+1G > A	No	0	Uncertain significance (*)	PVS1_strong	Yes	Likely pathogenic
10	2001+1,+2	c.2001+2T > C	No	1	Not reported	PVS1_strong	Yes	Likely pathogenic
11	2002-1,-2	c.2002-2A > G	No	1	Not reported	PVS1_strong	Yes	Likely pathogenic
11	2002-1,-2	c.2002-2A > C	No	0	Likely pathogenic(1); uncertain significance(1) (*)	1 PVS1_strong	2 Yes	3 Likely pathogenic
4 11	5 2002-1,-2	6 c.2002-2A > T	7 No	8 0	9 Likely pathogenic(1); uncertain significance(1) (*)	10 PVS1_strong	11 Yes	12 Likely pathogenic
13 11	14 2002-1,-2	15 c.2002-1G > A	16 No	17 1	18 Likely pathogenic (*)	19 PVS1_strong	20 Yes	21 Likely pathogenic
22 11	23 2002-1,-2	24 c.2002-1G > C	25 No	26 0	27 Likely pathogenic (*)	28 PVS1_strong	29 Yes	30 Likely pathogenic

^aAbsent.^bPM2 corresponds to the variant being absent from controls (or at extremely low frequency) as per ACMG/AMP guidelines (Richards et al., 2015). Gene specific ACMG/AMP guidelines have yet to be developed for *BARD1*, so it is possible that the use of PM2 and the weight of this criterion (e.g., moderate or supporting) may change in the future.

*criteria provided/single submitter -or- conflicting interpretation, **multiple submitters/no conflicts.

(Abou Tayoun et al., 2018). Regarding premature translation-termination codons (PTC-NMD variants) the guidelines suggest that they should be considered PVS1, unless located in an exon absent from biologically relevant transcript(s). For any PTC-NMD variants located in such exons, the PVS1 criterion is not applicable (N/A). This is a conservative rule introduced to cope with the possibility of rescue transcripts (i.e., alternatively spliced transcripts that skip the PTC-NMD variant providing haplo-sufficiency). Rescue transcripts overcoming the damaging effect of a PTC-NMD variant have been described for cancer predisposition genes such as *APC* (Nieuwenhuis and Vasen, 2007) and *BRCA1* (de la Hoya et al., 2016). However, we did not identify any candidate rescue transcript (no transcript other than the reference is predicted to code for functional RING, ARD, and BRCT domains) in our study. Therefore, we conclude that any PTC-NMD variant identified in *BARD1* should be considered PVS1. Regarding splice site (IVS ± 1,2) variants, ACMG/AMP

guidelines are more complex, and splice site variants may be considered PVS1, PVS1_Strong, PVS1_moderate, or PVS1_not applicable depending on several factors, such as: 1) the predicted outcome of the splice alteration being in-frame or truncating; 2) the predicted impact on clinically functional domains of the protein; and 3) the presence of candidate rescue transcripts (Supplementary Figure S4). According to our analysis, *BARD1* variants located at consensus splice sites can be considered PVS1 (n = 9 sites), or PVS1_strong (n = 10 sites). Only variants located at the donor site of exon 5 should be considered PVS1_moderate (Supplementary Table S4 and Figure S4). The presumed role of naturally occurring *BARD1* donor/acceptor shifts as predictors of cryptic site activation is based on a number of observations that we and others have made in other cancer susceptibility genes, including *PALB2*, *BRCA1*, and *BRCA2*. For example, *PALB2* c.48G > A (last nucleotide of exon 1) inactivates the donor site, leading to activation of a cryptic donor site to increasing levels

of the naturally occurring $\Delta(E1q17)$ alternative splicing event (Lopez-Perolio et al., 2019). It is important to note that caution maybe warranted when assessing variants for potential associated donor/acceptor shifts in genes that have not been thoroughly investigated for alternative transcripts.

Thirty four variants located at *BARD1* canonical splice sites (gnomAD, ClinVar; accessed June 2019) were identified to assess their clinical significance using ACMG/AMP criteria adapted for *BARD1* as described in **Table 2**. In absence of *in vitro* studies, we conclude that these variants (all them absent or extremely rare in control populations, and therefore accounting for PM2) can be reported as likely pathogenic, with the exception of variants targeting the donor site of *BARD1* exon 5, for which we suggest a more conservative classification of uncertain significance (**Table 2**).

In summary, we have conducted the most comprehensive assessment of *BARD1* mRNA splicing to date, and propose appropriate ACMG/AMP PVS1 evidence strengths to assist with classification of *BARD1* sequence variants in a modified version of the Abou Tayoun et al. (2018) decision tree. To our knowledge, we have conducted the first sequence analysis of whole *BARD1* mRNA transcript isoforms using nanopore sequencing, however further investigation of whole transcripts is required to account for all splicing events identified using other methods. This study did not identify *BARD1* candidate rescue transcripts, indicating that all premature translation-termination codons (PTC)_NMD variants can be assigned PVS1 at nominal strength. Moreover, donor and acceptor “consensus” dinucleotide variants (IVS+/-1,2) can be considered PVS1 or PVS1_strong, with the possible exception of variants targeting the exon 5 donor site, which we recommend assigning PVS1_moderate.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Otago Human Ethics Committee (Health) - H14/131. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., et al. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* 39 (11), 1517–1524. doi: 10.1002/humu.23626
- Bosse, K. R., Diskin, S. J., Cole, K. A., Wood, A. C., Schnepf, R. W., Norris, G., et al. (2012). Common variation at *BARD1* results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. *Cancer Res.* 72 (8), 2068–2078. doi: 10.1158/0008-5472.CAN-11-3703
- Couch, F. J., Shimelis, H., Hu, C., Hart, S. N., Polley, E. C., Na, J., et al. (2017). Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol.* 3 (9), 1190–1196. doi: 10.1001/jamaoncol.2017.0424

AUTHOR CONTRIBUTIONS

MH conceived and supervised the study. All authors performed the experiments, conducted data analysis, and/or interpreted the experimental results. kConFab provided LCLs to Laboratories 1 and 4. LW wrote the manuscript. All authors made manuscript revisions.

ACKNOWLEDGMENTS

We thank the Cancer Society of New Zealand Canterbury/West Coast Division for funding. LW was supported by the Royal Society of New Zealand Rutherford Discovery Fellowship. VL was supported by the Mackenzie Familial Breast Cancer Post-doctoral Fellowship. MH have received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 634935 MH, and from Spanish Instituto de Salud Carlos III (grants PI15/00059). We thank the Jim and Mary Carney Charitable Trust (Whangarei, New Zealand) for support (MK, LJ). AS is supported by an NHMRC Senior Research Fellowship (ID1061779). We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. PK and PZ were supported by the grant of Ministry of Health (www.mzcr.cz) NV18-03-00024, and the SVV 2019/260367 project. PK would like to thank Marketa Safarikova for technical support and for a providing a sequencer purchased from the project MH CZ – DRO VFN 64165. We would also like to thank the Sweden Cancerome Analysis Network - Breast (SCAN-B) and Ingrid Hedenfalk for providing access to healthy breast and fimbria tissues.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01139/full#supplementary-material>

- Davy, G., Rousselin, A., Goardon, N., Castera, L., Harter, V., Legros, A., et al. (2017). Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur. J. Hum. Genet.* 25 (10), 1147–1154. doi: 10.1038/ejhg.2017.116
- De Brakeleer, S., De Greve, J., Loris, R., Janin, N., Lissens, W., Sermijn, E., et al. (2010). Cancer predisposing missense and protein truncating *BARD1* mutations in non-*BRCA1* or *BRCA2* breast cancer families. *Hum. Mutat.* 31 (3), E1175–E1185. doi: 10.1002/humu.21200
- de Jong, L. C., Cree, S., Lattimore, V., Wiggins, G. A. R., Spurdle, A. B., kConFab, I., et al. (2017). Nanopore sequencing of full-length *BRCA1* mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res.* 19 (1), 127. doi: 10.1186/s13058-017-0919-1

- de la Hoya, M., Soukari, O., Lopez-Perolio, I., Vega, A., Walker, L. C., van Ierland, Y., et al. (2016). Combined genetic and splicing analysis of BRCA1 c.[594-2A > C; 641A > G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum. Mol. Genet.* 25 (11), 2256–2268. doi: 10.1093/hmg/ddw094
- Hojny, J., Zemankova, P., Lhota, F., Sevcik, J., Stranecky, V., Hartmannova, H., et al. (2017). Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model. *Gene* 637, 41–49. doi: 10.1016/j.gene.2017.09.025
- Irminger-Finger, I., Leung, W. C., Li, J., Dubois-Dauphin, M., Harb, J., Feki, A., et al. (2001). Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis. *Mol. Cell* 8 (6), 1255–1266. doi: 10.1016/s1097-2765(01)00406-3
- Ishitobi, M., Miyoshi, Y., Hasegawa, S., Egawa, C., Tamaki, Y., Monden, M., et al. (2003). Mutational analysis of BARD1 in familial breast cancer patients in Japan. *Cancer Lett.* 200 (1), 1–7. doi: 10.1016/s0304-3835(03)00387-2
- Karpinen, S. M., Heikinen, K., Rapakko, K., and Winqvist, R. (2004). Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. *J. Med. Genet.* 41 (9), e114. doi: 10.1136/jmg.2004.020669
- Kurian, A. W., Hughes, E., Handorf, E. A., Gutin, A., Allen, B., Hartman, A. R., et al. (2017). Breast and ovarian cancer penetrance estimates derived from germline multiple-gene sequencing results in women. *JCO Precis. Oncol.* 1, 1–12. doi: 10.1200/PO.16.00066
- Lattimore, V. L., Pearson, J. F., Currie, M. J., Spurdle, A. B., kConFab, I., Robinson, B. A., et al. (2018). Investigation of experimental factors that underlie brca1/2 mRNA isoform expression variation: recommendations for utilizing targeted RNA sequencing to evaluate potential spliceogenic variants. *Front. Oncol.* 8, 140. doi: 10.3389/fonc.2018.00140
- Li, L., Ryser, S., Dizin, E., Pils, D., Krainer, M., Jefford, C. E., et al. (2007). Oncogenic BARD1 isoforms expressed in gynecological cancers. *Cancer Res.* 67 (24), 11876–11885. doi: 10.1158/0008-5472.CAN-07-2370
- Lombardi, G., Falaschi, E., Di Cristofano, C., Naccarato, A. G., Sensi, E., Aretini, P., et al. (2007). Identification of novel alternatively spliced BRCA1-associated RING domain (BARD1) messenger RNAs in human peripheral blood lymphocytes and in sporadic breast cancer tissues. *Genes Chromosomes Cancer* 46 (9), 791–795. doi: 10.1002/gcc.20460
- Lopez-Perolio, I., Leman, R., Behar, R., Lattimore, V., Pearson, J. F., Castera, L., et al. (2019). Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. *J. Med. Genet.* 56 (7), 453–460. doi: 10.1136/jmedgenet-2018-105834
- McCarthy, E. E., Celebi, J. T., Baer, R., and Ludwig, T. (2003). Loss of Bard1, the heterodimeric partner of the Brca1 tumor suppressor, results in early embryonic lethality and chromosomal instability. *Mol. Cell Biol.* 23 (14), 5056–5063. doi: 10.1128/mcb.23.14.5056-5063.2003
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266 (5182), 66–71. doi: 10.1126/science.7545954
- Nieuwenhuis, M. H., and Vasen, H. F. (2007). Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Crit. Rev. Oncol. Hematol.* 61 (2), 153–161. doi: 10.1016/j.critrevonc.2006.07.004
- Pilyugin, M., and Irminger-Finger, I. (2014). Long non-coding RNA and microRNAs might act in regulating the expression of BARD1 mRNAs. *Int. J. Biochem. Cell Biol.* 54, 356–367. doi: 10.1016/j.biocel.2014.06.018
- Ratajska, M., Antoszewska, E., Piskorz, A., Brozek, I., Borg, A., Kusmierek, H., et al. (2012). Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res. Treat.* 131 (1), 89–97. doi: 10.1007/s10549-011-1403-8
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi: 10.1038/gim.2015.30
- Ryser, S., Dizin, E., Jefford, C. E., Delaval, B., Gagos, S., Christodoulidou, A., et al. (2009). Distinct roles of BARD1 isoforms in mitosis: full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2. *Cancer Res.* 69 (3), 1125–1134. doi: 10.1158/0008-5472.CAN-08-2134
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17 (1), 19–32. doi: 10.1038/nrg.2015.3
- Shimelis, H., LaDuca, H., Hu, C., Hart, S. N., Na, J., Thomas, A., et al. (2018). Triple-negative breast cancer risk genes identified by multigene hereditary cancer panel testing. *J. Natl. Cancer Inst.* 110 (8), 855–862. doi: 10.1093/jnci/djy106
- Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., et al. (2017). The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer* 3, 22. doi: 10.1038/s41523-017-0024-8
- Soukupova, J., Zemankova, P., Lhotova, K., Janatova, M., Borecka, M., Stolarova, L., et al. (2018). Validation of CZECA (CZEch Cancer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes. *PLoS One* 13 (4), e0195761. doi: 10.1371/journal.pone.0195761
- Sporn, J. C., Hothorn, T., and Jung, B. (2011). BARD1 expression predicts outcome in colon cancer. *Clin. Cancer Res.* 17 (16), 5451–5462. doi: 10.1158/1078-0432.CCR-11-0263
- Tang, A. D., Soulette, C. M., van Baren, M. J., Hart, K., Hrabeta-Robinson, E., Wu, C. J., et al. (2018). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *bioRxiv*, 410183. doi: 10.1101/410183
- Tesoriero, A. A., Wong, E. M., Jenkins, M. A., Hopper, J. L., Brown, M. A., Chenevix-Trench, G., et al. (2005). Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum. Mutat.* 26 (5), 495. doi: 10.1002/humu.9379
- Walker, L. C., Whitley, P. J., Couch, F. J., Farrugia, D. J., Healey, S., Eccles, D. M., et al. (2010). Detection of splicing aberrations caused by BRCA1 and BRCA2 sequence variants encoding missense substitutions: implications for prediction of pathogenicity. *Hum. Mutat.* 31 (6), E1484–E1505. doi: 10.1002/humu.21267
- Walker, L. C., Whitley, P. J., Houdayer, C., Hansen, T. V., Vega, A., Santamarina, M., et al. (2013). Evaluation of a 5-tier scheme proposed for classification of sequence variants using bioinformatic and splicing assay data: inter-reviewer variability and promotion of minimum reporting guidelines. *Hum. Mutat.* 34 (10), 1424–1431. doi: 10.1002/humu.22388
- Whitley, P. J., Pettigrew, C. A., Brewster, B. L., Walker, L. C., Spurdle, A. B., and Brown, M. A. (2010). Effect of BRCA2 sequence variants predicted to disrupt exonic splice enhancers on BRCA2 transcripts. *BMC Med. Genet.* 11, 80. doi: 10.1186/1471-2350-11-80
- Whitley, P. J., Guidugli, L., Walker, L. C., Healey, S., Thompson, B. A., Lakhani, S. R., et al. (2011). Splicing and multifactorial analysis of intronic BRCA1 and BRCA2 sequence variants identifies clinically significant splicing aberrations up to 12 nucleotides from the intron/exon boundary. *Hum. Mutat.* 32 (6), 678–687. doi: 10.1002/humu.21495
- Whitley, P. J., de la Hoya, M., Thomassen, M., Becker, A., Brandao, R., Pedersen, I. S., et al. (2014). Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin. Chem.* 60 (2), 341–352. doi: 10.1373/clinchem.2013.210658
- Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., et al. (1996). Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product. *Nat. Genet.* 14 (4), 430–440. doi: 10.1038/ng1296-430
- Zhang, Y. Q., Pilyugin, M., Kuester, D., Leoni, V. P., Li, L., Casula, G., et al. (2012). Expression of oncogenic BARD1 isoforms affects colon cancer progression and correlates with clinical outcome. *Br. J. Cancer* 107 (4), 675–683. doi: 10.1038/bjc.2012.297

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

Copyright © 2019 Walker, Lattimore, Kvist, Kleiblova, Zemankova, de Jong, Wiggins, Hakkaart, Cree, Behar, Houdayer, Investigators, Parsons, Kennedy, Spurdle and de la Hoya. 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.



Exploring the RNA Gap for Improving Diagnostic Yield in Primary Immunodeficiencies

Jed J. Lye¹, Anthony Williams^{1,2*} and Diana Baralle^{1,3*}

¹ University of Southampton Medical School, University of Southampton, Southampton, United Kingdom, ² Wessex Investigational Sciences Hub Laboratory (WISH Lab), Faculty of Medicine, University of Southampton, Southampton, United Kingdom, ³ Faculty of Medicine, Highfield Campus, University of Southampton, Southampton, United Kingdom

OPEN ACCESS

Edited by:

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

Reviewed by:

Sebastian Oltean,
University of Exeter,
United Kingdom
Franco Pagani,
International Centre for Genetic
Engineering and Biotechnology,
Italy

*Correspondence:

Anthony Williams
apw2@soton.ac.uk
Diana Baralle
D.Baralle@soton.ac.uk

Specialty section:

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

Received: 02 September 2019

Accepted: 31 October 2019

Published: 11 December 2019

Citation:

Lye JJ, Williams A and Baralle D
(2019) Exploring the RNA Gap
for Improving Diagnostic Yield in
Primary Immunodeficiencies.
Front. Genet. 10:1204.
doi: 10.3389/fgene.2019.01204

Challenges in diagnosing primary immunodeficiency are numerous and diverse, with current whole-exome and whole-genome sequencing approaches only able to reach a molecular diagnosis in 25–60% of cases. We assess these problems and discuss how RNA-focused analysis has expanded and improved in recent years and may now be utilized to gain an unparalleled insight into cellular immunology. We review how investigation into RNA biology can give information regarding the differential expression, monoallelic expression, and alternative splicing—which have important roles in immune regulation and function. We show how this information can inform bioinformatic analysis pipelines and aid in the variant filtering process, expediting the identification of causal variants—especially those affecting splicing—and enhance overall diagnostic ability. We also demonstrate the challenges, which remain in the design of this type of investigation, regarding technological limitation and biological considerations and suggest potential directions for the clinical applications.

Keywords: primary immunodeficiency disorders, clinical diagnostics, RNASeq, RNA, RNAseq analysis

INTRODUCTION

Primary immunodeficiency disorders (PID) result from altered, poor, or absent function in one or more components of the immune system, rendering the affected individuals with increased susceptibility to immune-related ailments including increased frequency and severity of infection, autoimmunity, aberrant inflammation, and malignancy (McCusker et al., 2018). The understanding of the genetic heterogeneity of PID has expanded greatly over the last decade, now encompassing a list of over 350 distinct disorders arising from at least 344 gene defects, demonstrative of the complexity of the immune system (Bousfiha et al., 2018; Picard et al., 2018). This plethora of genetic causes has brought about a need to categorize the disorders for expedited diagnosis and treatment protocols. Some broader methods simply classify the disorders into groups of innate and adaptive immunity linked to the clinical phenotype (McCusker and Warrington, 2011). The Inborn Errors of Immunity Committee (previously the International Union of Immunological Societies PID Expert Committee) has now devised a precise and useful system, which classifies disorders by the immunological pathway affected. In addition, it now has corresponding phenotypical classification systems for clinicians at the bedside to help identify the disorders. These briefly comprise nine categories: immunodeficiencies affecting cellular and humoral immunity, combined immunodeficiency disorder (CID) with associated or syndromic features, predominantly antibody deficiencies, diseases of immune dysregulation, congenital defects of phagocyte, defects in intrinsic

and innate immunity, auto-inflammatory disorders, complement deficiencies, and phenocopies of PID (Bousfiha et al., 2018).

The most common form of PID is selective immunoglobulin A deficiency, which is usually typically asymptomatic but can manifest with a variety of clinical presentations including coeliac disease, type 1 diabetes mellitus, and increased infections. With an estimated prevalence of 1 in 300–500 persons (Boyle and Buckley, 2007), and while individually rare (>1 in 2,000), the remaining disorders considered in the wider scope of PID together represent a significant burden on the health and economy of a nation. Current diagnostic levels suggest an incidence of 5.90/100,000 (Shillitoe et al., 2018); however, underdiagnosis of PID may mean the true incidence is as high as 1:250 (Europe PIPDDfOCi.).

The importance of early diagnosis in PID cases is high, with relation to both the patient's qualitative experience and the economic cost to healthcare services. Sources vary in cost analysis of undiagnosed PID. Some say that while a diagnosed US patient costs healthcare services over US\$250,000 per annum, largely due to treatment costs, an early diagnosis of the disorder can save as much as US\$6,500 per patient, per annum (Abolhassani et al., 2015). An alternate source suggests an undiagnosed patient might cost the healthcare system US\$102,552 annually. Once diagnosed, these costs may drop by as much as US\$79,942 (Condino-Neto and Espinosa-Rosales, 2018). In a patient survey, 45% of patients reported a diagnostic wait time of between 1 and 6 years; around 1/6th reported waiting 10–20 years. Other key findings of the same survey confirmed undiagnosed patients bring about a dramatically increased burden on National Health Service (NHS) resources (UK). Identification of the precise molecular origins for each patient's case of PID leads to improved patient care (Walter et al., 2016) and improved prognosis. The importance of correct genetic cause for a PID phenotype is demonstrated by the different treatment preferences which exist for conditions which may present with similar clinical phenotypes (Heimall et al., 2012). Precision therapeutic diagnostics can help to achieve this in part, by allowing targeted intervention to the specific molecular causes (Bonilla et al., 2015; Lenardo et al., 2016; Ramakrishnan et al., 2016).

Diagnostic Challenges in PID

Challenges in diagnosing primary immunodeficiency are numerous and diverse. Studies which correlate the phenotype and genotype have been useful in diagnostics, developing an understanding of various PID disorders (Fischer, 1993). Additionally, these correlation studies have been useful for deconvoluting the pleiotropic nature of the involved genes, through which a single variant can bring about a variety of clinical phenotypes (Meyts et al., 2016). However, the development of a universal diagnostic pipeline for PID is hindered by the heterogeneity in the presentation of disease, even among patients with what appears to be the same pathogenic genetic variant (Richardson et al., 2018). Conversely, a number of genotypes can bring about even the most well-characterized phenotype (Meyts et al., 2016). Once a clinical diagnosis of PID is suspected, mainly based upon a compatible phenotype, a family history

is usually taken and a number of subsequent laboratory tests performed to confirm the type of immune mechanism affected (Hernandez-Trujillo and Ballou, 2015). With the emergence of targeted sequencing of larger PID gene panels, clinical exomes, and complete exomes through short-read next-generation sequencing technologies, the inclusion of genetic testing within a PID diagnostic workup has become more widespread. This approach to both adult- and paediatric-onset disease has consolidated the importance of protein-based functional immune testing (cytokines, antibodies, etc.) for characterizing the nature of the phenotypic presentation, but furthermore to evaluate candidate genetic variants in such pathways that have been identified through parallel germline DNA testing.

DNA sequencing-based genetic testing is used where possible, as it provides the best diagnostic capability of existing clinically adapted methods (Condino-Neto and Espinosa-Rosales, 2018). Whole-exome sequencing (WES) has the highest success rate of the clinically adapted diagnostic methods (Gilissen et al., 2011; Boycott et al., 2017), which it achieves despite the exome comprising only ~2% of the human genome (Bamshad et al., 2011). This is in part due to 85% of currently annotated variants existing within the transcribed portion of the genome (Majewski et al., 2011). It has been hypothesized that this focus had likely led to the underestimation of the contribution to disease of non-coding variants (Kremer et al., 2018).

Due to the improvement that WES and whole-genome sequencing (WGS) bring to diagnostics, researchers are calling for universal molecular gene testing for the diagnosis of primary immune deficiencies (Heimall, 2019). Evidence from existing literature, however, suggests that even this may be inadequate; currently, WES and WGS are only able to produce reliable diagnosis in 25–60% of cases (Yang et al., 2013; Moens et al., 2014; Yang et al., 2014; Taylor et al., 2015; Meyts et al., 2016; Stray-Pedersen et al., 2017). Although many countries have undertaken whole-genome sequencing projects to evaluate this approach (Philippidis, 2018), the development of WGS as a clinically validated routine testing modality is still in its infancy. Within the UK's 100,000 Genomes Project, PID was accepted as an indication for inclusion, and plans to incorporate WGS for PID into routine clinical pathways have been approved following the transition phase of 100K Project to WGS sequencing in routine NHS care across England.

Formal confirmed genetic diagnosis of PID relies heavily on existing knowledge pertaining to consequences of the variants in the genes of relevance to the presenting phenotype and assumed mechanism of disease resulting from such variants in a dominant or recessive manner genomic sequence (Rae et al., 2018). The key to this task is the ability of bioinformatics tools to predict the significance of such variants. WES delivers around 20,000–23,000 variants per individual, and WGS produces 3–5 million per individual (Kremer et al., 2018), which makes the task of identifying a Mendelian disease variant vanishingly unlikely without a series of bioinformatics filters. Problems with the WGS/WES sequencing diagnostic methods arise when no variant, identified through a patient's genome sequencing, can be reliably linked to the clinical presentation and cytological/molecular manifestation of the disorder. Failure to identify a

definitive molecular cause occurs in about 70–75% of Mendelian conditions, according to a 2018 meta-analysis (Schwarze et al., 2018), mirrored by examples from PID (Gallo et al., 2016). The types of variants which are not always identified by current next-generation sequencing (NGS) approaches include exonic variants of unknown significance, variants in intronic and intergenic non-coding DNA (Scacheri and Scacheri, 2015), variants in the *cis*-acting regulatory elements of transcription (Bryois et al., 2014) imprinting disorders, and repeat expansions (Kremer et al., 2018).

Conventional clinical diagnostics, utilizing human phenotype ontology for integration of cases into specific diagnostic groups, and traditional genetic sequencing methods for diagnostics are still currently inadequate. While proteomic diagnostic methods are in development, they exist at a relatively early stage of development and can miss the potentially valuable RNA regulatory phenomena.

Variants Affecting Differential Expression

Identification of definitive disease-causing mutations is confounded in some cases by expression levels being modulated by variants occurring in non-coding segments and those hiding in plain sight in genes not currently understood to be linked to the disease or phenotype. Often, these can be lost during the filtering process because of a lack of integrative understanding or supporting evidence (Thormann et al., 2019).

These expression quantitative trait loci (eQTLs) elicit a powerful, sometimes synergistic effect on the expression of a large number of genes. Single-nucleotide polymorphisms (SNPs) on eQTLs affect the transcriptional level of other RNAs, modifying protein expression and causing phenotypic changes to the abilities and behaviors of cells in some immunological cases (Fairfax et al., 2014). These eQTLs individually explain a fraction of the genetic expression of specific genes. The vast majority do not exist in the coding regions of genes and are predicted to be involved in gene regulation (Casamassimi et al., 2017). It is now understood that these eQTLs have a more pronounced effect on immune regulation than the effects of age and sex, and more interestingly exclusive effects only observable during immune stimulation have been identified for some of these eQTL variants (Piasecka et al., 2018). Epigenomic studies have helped to highlight the *cis*-regulatory nature of some non-coding regions of the genome. These suggest that the enrichment of disease-risk variants in cell-specific regulatory sequences is indicative of their cell type and contextual effects (Roadmap Epigenomics et al., 2015). Large-scale investigation into the association between genetic variants and expression of genes in a tissue-specific manner (including whole blood) was carried out by the Genotype—Tissue Expression Consortium (The Genotype-Tissue (GTEx), 2015). This research did not extend to immune tissues specifically, although links between immune cell-specific gene expression levels and eQTLs have been investigated by the DICE Project (Database of Immune Cell Expression, Expression Quantitative Trait Loci and Epigenomics) (Schmiedel et al., 2018). The researchers on this project were able to positively identify a range of *cis*-eQTLs for 12,254 genes, demonstrative of the high abundance of these sites. Interestingly, many of these eQTLs had

effects which were cell type-specific. The identification of these sites and interrogation for the existence of variants will likely play a crucial role in explaining the changes in expression of key genes which lead to PID.

The Role of RNA Splicing in the Immune System and PID

Alternative splicing is the method through which the cell can produce an array of transcript isoforms derived from a single gene or multiple genes spliced together (Ward and Cooper, 2010). Introns are spliced out and exons are either ligated through transesterification reaction or, in many cases, spliced out in different combinations, leaving the remaining exons to form a mature mRNA (Ward and Cooper, 2010).

Deep surveying on alternative splicing has shown that 95% of genes which contain multiple exons undergo alternative splicing, and even when only considering moderate to high abundance events, there are reportedly 100,000 individual splicing events in major tissues (Pan et al., 2008).

Alternative splicing occurs both co-transcriptionally and post-transcriptionally, and the action of transcription factors as well as splicing factors regulates and influences splicing events in some of the most crucial mechanisms of the adaptive immune system (Heyd et al., 2006; Orvain et al., 2008; Alkhatib et al., 2012). Important examples include RNA-polymerase II as a facilitator of splicing factor recruitment (Bentley, 2014), the alternative splicing of CD45 which is necessary for the production of a range of tyrosine phosphatases, imperative for the diverse set of lineage and stage-specific receptor signal transduction thresholds in immune tissues (Zikherman and Weiss, 2008), and FOXO1-induced Ikaros splicing, essential for the recombination of immunoglobulin genes. FOXO1 is a transcription factor which, through its effects on alternative splicing, allows the immune system to produce its diverse range of antibodies/immunoglobulins (Reynaud et al., 2008).

Activation of lymphocytes is a key component of the adaptive immune response to pathogens (Bonilla and Oettgen, 2010). Part of the central activation of these cells is the degradation of I κ B α and release of NF- κ B, which translocates to the nucleus to initiate maturation and activation of the cell. The “CBM” complex, which brings about the degradation of I κ B α , is formed by *CARMA1*, *BCL10*, and *MALT1* (Oeckinghaus et al., 2007). *MALT1*, a crucial component of this complex, undergoes alternative splicing of EXON 7 to produce mRNA isoforms with a differential function. The activation strength of CD4⁺ T cells is mediated by the relative abundance of the alternatively spliced isoforms of *MALT1*, which is in part controlled by the molarity of phosphorylated splicing factor hnRNPU in the nucleus (Meininger et al., 2016). Alternative splicing, then, is a key component of the normally functioning immune system, and perturbations in canonical function can likely lead to pathology.

Variants Affecting Alternative Splicing

The impact of mutations that affect RNA processing/splicing is currently providing a diagnostic revolution. Variants which

affect splicing either occur in active splice sites, those which occur in regulatory elements, and those which occur in intronic or intergenic regions (Grodecká et al., 2017) see **Figure 1**.

Studies comparing variants affecting splicing in PID have determined that the variants which directly influence splice sites are more robustly linked to disease phenotypes than those which effect splicing regulatory elements (Grodecká et al., 2017). *Cis*-mutations in the genome can affect splicing through altering the splice site recognition or altering exon splicing enhancer or silencer sites (Ward and Cooper, 2010). Splice sites usually comprise GT and AG dinucleotides at 5' and 3' sites, respectively. If a variant changes this sequence, or causes another one to appear, it can affect the ability of the splicing machinery to detect the canonical splice site (Krawczak et al., 1992; Krawczak et al., 2007). Additionally, mutations in *trans*-acting splice factors—the splicing machinery of the cell—can also bring about disease by preventing these factors from performing their function of generating the required isoforms (Ward and Cooper, 2010), although these are not covered in this review.

Due to the impact of these findings, interest in the detection of splice-altering variants and activated cryptic splice site has spurred on the development of a number of *in silico* tools for prediction of splice site usage (Grodecká et al., 2017; Ohno et al., 2018). Unfortunately, these tools are often unable to discern the resulting transcripts exon use patterns (Jaganathan et al., 2019), and while their predictive ability can be enhanced by other orthogonal investigations such as mini-gene assays (Grodecká et al., 2017), the multiple facets of splicing control involve more than just the sequence of the splice site in question, as evidenced by the temporal and spatial differences in splicing patterns. Briefly, these include the activation of other splice sites within the gene, splicing quantitative trait loci, the relative abundance, phosphorylation status, and localization of different and often competing *trans*-acting factors (Wang et al., 2015).

Further complicating this process, seemingly benign, synonymous exonic variants can disrupt splicing to cause disease. Using RNA sequencing (RNASeq) to complement genomic sequencing, Cummings et al. evidenced this in the *POMGNT1* and *RYR1* genes, finding variants which were demonstrated to be causative of Mendelian diseases in muscle (Cummings et al., 2017). Part of the normal filtering process which many bioinformaticians adopt is to filter out synonymous variants very early on, but investigation using deep learning has led to the understanding that between 9% and 11% of rare genetic disorders are caused by synonymous or intronic splice-altering mutations (Jaganathan et al., 2019). Indeed, much as gene expression can be influenced by multiple loci, so too can multiple loci contribute to the occurrence of splicing events. These loci are appropriately termed splicing quantitative trait loci (sQTLs) (Jia et al., 2015). Analysis of sQTLs has been improved by RNASeq methodologies, but remains a difficult challenge as the isoform expression has to be estimated using statistical methods (Patro et al., 2017). These sQTLs are not necessarily in close proximity to the splice junction. Characterization of these sites in humans has shown

SNPs demonstrating tangible sQTL activity at 100 kb from the relative splice site (Takata et al., 2017).

Non-protein-coding genes are a significant source of disease-causing variation (Scacheri and Scacheri, 2015). Examples within the PID research and diagnosis space include a recently discovered variant occurring in coding regions for genes comprising RNA components of the minor spliceosome, which is used for the splicing of at least one exon in ~800 genes (Turunen et al., 2013). Specifically, the non-coding gene *RNU4ATAC* that produces a small nuclear RNA (snRNA) termed U4atac was discovered to cause Roifman syndrome (Merico et al., 2015; Heremans et al., 2018) by preventing canonical minor intron splicing. Compound heterozygous variants were first discovered in an affected family after traditional filtering methods had not detected viable variants; the link was confirmed by the detection of intron retention during curated splicing analysis of RNASeq data (Merico et al., 2015).

The importance of alternative splicing in the immune system has been further demonstrated in mouse models. The ImmGen Project was set up specifically to investigate gene expression and regulation in mice using microarray profiling. It found that, in mice, around 60% of genes are expressed as multiple isoforms in T or B cells, and 70% of these had an impact on the lineage differentiation (Ergun et al., 2013). Compound heterozygous mutations in *MALT1*, mentioned earlier, which is heavily implicated in the activation of T cells, have been shown to bring about profound combined immunodeficiency. One of these variants was indeed a splice site acceptor change from the consensus AG to GG, identified by whole-exome sequencing (Punwani et al., 2015).

To further complicate the already complex nexus of control mechanisms contributing to PID, a range of epigenetic mechanisms leading to primary Immunodeficiencies have been observed and reviewed (Campos-Sanchez et al., 2019). In principle, the majority of genes identified to be susceptible to variants in PID may also be subject to heritable epigenetic modifications which could lead to the same or similar symptoms, acting as a further coefficient value when calculating the potential number of disorders, including those disorders affected by splicing, which can cause PID (Zhu et al., 2018).

RNA in Diagnostics

RNA investigation technology and its literature has experienced great leaps forward in recent years in terms of technological advancement and cost reduction (Muir et al., 2016). RNA sequencing is now largely replacing microarrays as the most used quantitative method of mapping gene expression profiles (Lowe et al., 2017). The transcriptome—or RNA expression profile—of a given tissue can give unparalleled insights into the elegant inner workings of the cell. Through capture of all internal RNA species, it characterizes the cellular gene transcription architecture and can deliver an instantaneous picture of environment–cell interaction or response program (Lowe et al., 2017; Wirka et al., 2018).

A range of technologies exist for conducting RNA sequencing, each with its own strengths and limitations.

Long-read sequencing provides reliable structural information, but can have suboptimal reliability in base calling (Feng et al., 2015) or is more expensive for high-throughput analysis (Rhoads and Au, 2015). Short-read NGS RNASeq involves sonication or enzymatic degradation of RNA into smaller fragments, selection of fragments using one of a number of methods, cDNA synthesis, the construction of a library, and subsequent sequencing followed by realignment (Kukurba and Montgomery, 2015). This technology has been the currently favored approach for high-throughput analysis.

Currently, this technology generates a mixture of both quantitative and qualitative analysis opportunities of RNA species: qualitative transcriptome profiling outcomes include identification of sequence variants at the level of the genome (Neums et al., 2017), somatic cell mosaics, non-canonical splice variants, occurring either due to *cis*- or *trans*-acting factor aberrations (Ward and Cooper, 2010). Quantitative outcomes of transcriptional profiling include differentially expressed genes, alternative splicing events, and allele-specific expression quantification (Kukurba and Montgomery, 2015). Previous studies have demonstrated that when compared with large control datasets, identification of expression outliers in peripheral whole blood can contribute to the detection of disease-causing variants (Zeng et al., 2015; Zhao et al., 2016). As well as gene expression levels, perturbations in the relative abundance of specific isoforms is a driving force in the genesis of many diseases (Chen et al., 2010; Kim et al., 2018) as isoforms can have differential function (Takeda et al., 2010), or, in some cases, can be antagonistic (Eshel et al., 2008). Through RNASeq or exon junction spanning probe-based capture, changes in isoform balance can also be resolved. The sensitivity suitability of RNASeq in transcriptomic investigation and splicing was demonstrated in mouse and human models and has enabled the discovery of ~7,600 novel isoforms in mouse immune cells (Ergun et al., 2013) and detected 100,000 splicing events with at least moderate abundance (Pan et al., 2008).

Transcriptome profiling can also give insights into control mechanisms exhibited by the non-coding RNA species, such as long non-coding RNA (lncRNA) and microRNA (miRNA), the significance of which is continually being elucidated in the molecular pathology of disease (Kramer et al., 2015; DiStefano, 2018). Indeed, such examples exist in PID; miR-6891-5p accumulation is demonstrated to contribute to selective IgA deficiency, the most common form of PID (Chitnis et al., 2017). Thanks to the increasing ability of technology and steady reduction in costs, we are also able to cast a wider net. Through RNASeq-based investigation, instead of concentrating on *a priori*, system-specific gene panels that many studies target, it is possible to examine all the mRNA species destined for translation. Through these hypothesis-free methods, it is possible to create profiles of normal transcription and disease transcription in a tissue-specific manner (Gonorazky et al., 2019). Subsequent use of follow-up analysis tools can be used to generate filtering process for causal variants or for biomarker identification (Han and Jiang, 2014). It is also possible to quantify the relative expression of those genes coding for the

splice factors themselves, which can directly bring about pathological processes specific to PID, such as those observed in Roifman's syndrome, mentioned earlier (Merico et al., 2015; Heremans et al., 2018).

Microfluidic technology adaptations have allowed the development of robust, single-cell transcriptomic profiling (Kimmerling et al., 2016). In combination with NGS-based technologies, the single-cell technology provides a method for profiling the transcriptomes of individual cells, giving unparalleled insights into the heterogeneity of cell populations and their transcriptional profiles (Hwang et al., 2018). Adaptations such as the SMART-seq2 or Fluidigm C1 library preparation methods also now allow the production of full-length cDNAs, giving transcript isoform-level resolution. However, these methods do not yet allow multiplexing, massively increasing overall costs and labor in large cohorts (See et al., 2018). The ability to profile the entire transcriptome of a peripheral blood mononuclear cell (PBMC) culture individually would give a dramatically increased ability to understand the cell-cell interactions taking place in an immune challenge, and this approach could be utilized in those patients suspected to be genetic mosaics.

DISCUSSION

The early and accurate diagnosis of primary immunodeficiencies is important to ensure the attainment of positive patient outcomes, through minimizing the time to diagnosis, identifying molecular pathways for targeted therapy, and reducing the economic cost of ill health or inappropriate treatment options. Diagnosis of the disorders remains difficult due to clinical challenges in identifying the presence of a primary immune system disorder, stratifying the phenotype to a myriad of overlapping candidate genes and then the laborious task of variant filtering, interpretation, and lack of knowledge pertaining to variants, especially those residing in the non-coding segments of the DNA. Functional validation of a candidate variant is currently undertaken with protein-based *ex vivo* tests, which are difficult to standardize and mostly available in research laboratories. RNA profiling to identify alternative splicing, gene expression-level variation monoallelic expression may contribute a further insight into candidate variants derived from proband or family-based WES/WGS sequencing results. We propose the introduction of RNASeq-based analysis for patients who have a clinical presentation of PID, but who, despite normal baseline immune testing, cellular analysis, and having undergone WES/WGS remain undiagnosed (see Figure 2).

RNASeq is an emerging technology which, when combined with WES/WGS, provides unprecedented insights into differential gene expression, splicing activity, and allelic specific expression and can inform regarding other phenomena such as genetic mosaicism. However, RNASeq remains relatively novel as a diagnostic testing tool in rare diseases and the control datasets and cellular contributions to complex tissue profiles (i.e., whole blood) will require further dissection.

Utilizing candidate gene lists and large control datasets for comparison enhances the power of the transcriptional

The effects of variants on the mRNA splicing process

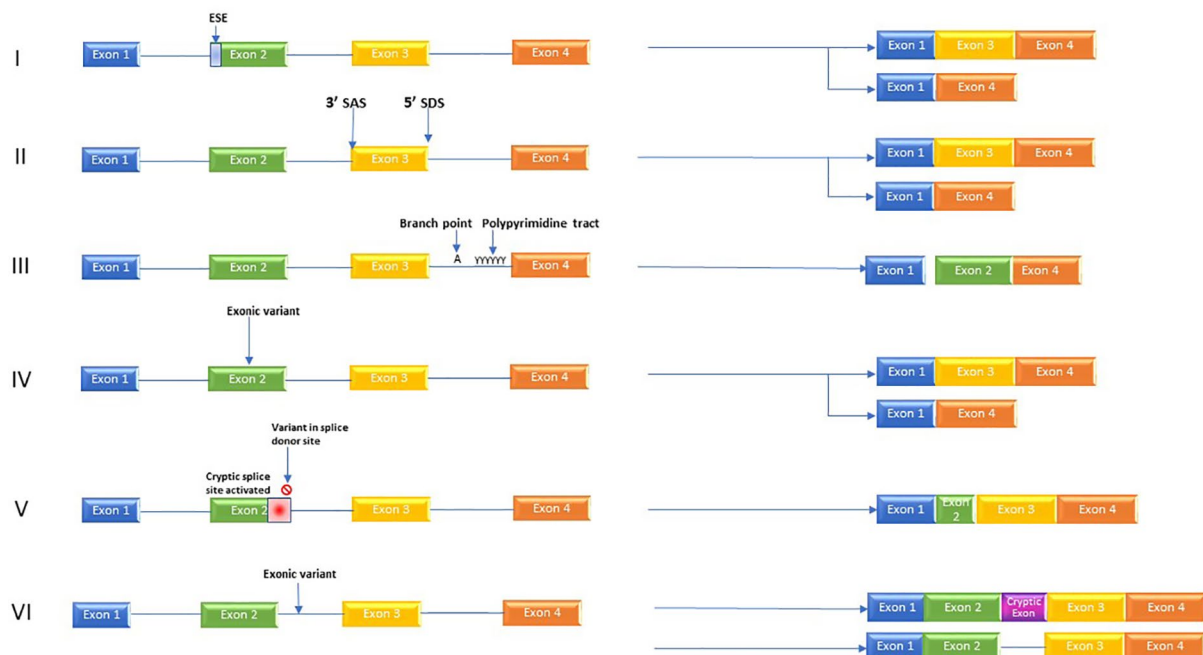


FIGURE 1 | Shows the various effects which variants can have on the splicing process. I) Variants in regulator elements such as exon splicing enhancers resulting in for example, exon skipping. II) Variant in splice acceptor OR splice donor site causes skipping of one or more exons. III) Branch point and Poly pyrimidine tract sequence variants causing exon skipping. IV) Exonic variants causing exon skipping. V) Variant in splice donor site induces activation of an alternative, cryptic splice donor site in exon. VI) Intronic variants producing new cryptic exon or retained intron.

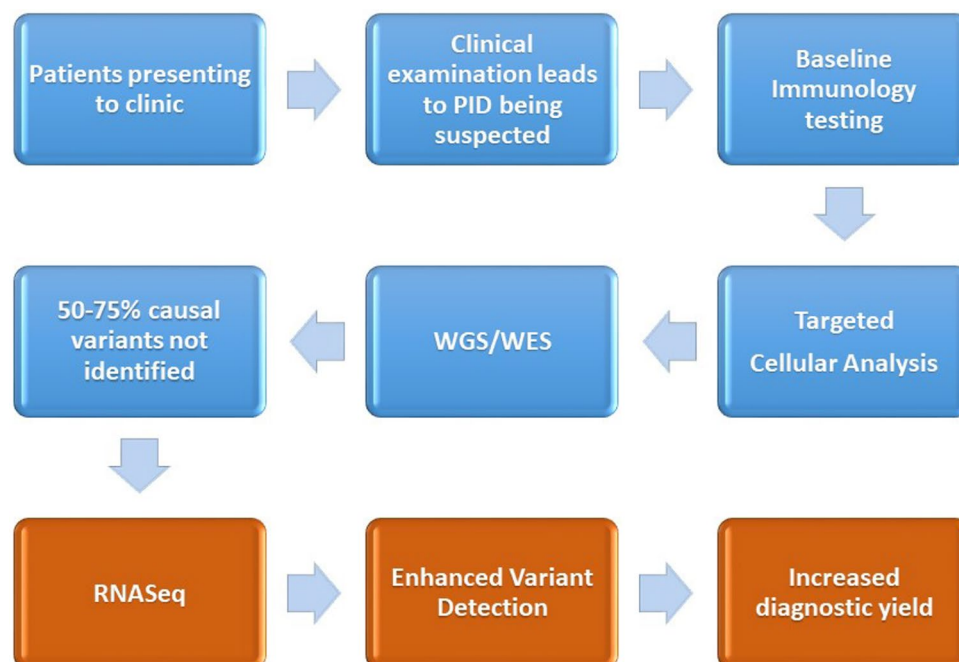


FIGURE 2 | Demonstration of the current diagnostic pathway (blue) for the majority of patients after a diagnosis of PID is suspected. We also outline the proposed intervention point of RNASeq and the associated enhanced variant detection (coming about through assessment of differential expression, changes to alternative splicing) and the increased diagnostic yield.

profiling through RNASeq and improves resolution for differential gene expression. Existing projects have developed these datasets for whole blood and immune cells, which provide a starting point for the interrogation of clinical samples for diagnostic research.

Immune responses to pathogenic challenges are exceptionally variable, and the variability in these responses is not easily elucidated. Environmental influences such as age, sex, seasonality, nutrition, and lifestyle all have effects on the specific response profile exhibited by individuals (Piasecka et al., 2018). These factors that influence responses can have a greater degree of significance in specific cell types. CD8⁺ T cells, for example, show a high degree of heterogeneity in the context of temporal changes through the life course of the individual, and CD4⁺ T cells and monocytes are heavily influenced by sex (Piasecka et al., 2018). It is therefore useful to be able to discern transcripts from different cell types within a culture. Utilizing flow cytometry to separate cell types or utilizing single-cell RNASeq is becoming an attractive option.

In order to assess the impact of genomic variation on the unstimulated immune system, the normal immune response and the immune-deficient responses, it is important to experimentally “tune out” the variations in signal arising from environmental factors. It has been established that a high degree of the cellular variation in CD8⁺ cell populations can be attributed to environmental factors, which makes them a poor model for genetic variant impact. CD4⁺ T cells display a large degree of heritability in these assays, and as such should provide a good level of transcriptomic heritability also. This will allow for clearer elucidation of the effects of variants on differential gene expression (Brodin et al., 2015).

The immune system's response to pathogen-based challenges is highly dynamic, and observing this response is more informative when identifying impaired response (Duffy et al., 2017). Indeed, it has been shown in innate immune system studies that the effects on differential expression of some variants can only be observed in a dynamic fashion (Fairfax et al., 2014; Lee et al., 2014). Co-culture of PBMCs provides a greater insight into activation pathways as it allows for the cell–cell communication response programs and produces similar results in terms of ranked gene expression response networks, with a few notable exceptions (Duffy et al., 2017). Studies of dynamic immune responses to challenges, in concert with machine learning, can be used to identify small groups of stimulation pathway-specific genes (Urrutia et al., 2016). Comparing the expression profiles of these genes in healthy cohorts with PID patients can potentially be utilized to identify candidate genes, which may then harbor a disease-causing variant or indicate some anomaly in the pathway for further investigation.

The transcriptomic landscape provides an excellent opportunity for advancement of diagnostic yield, and transcriptional profiling is already being utilized across a range of disorders to help build a “molecular fingerprint” of disease and better inform variant-filtering processes. The immunology community has made a case for PID diagnosis to be supported using transcriptional profiling using whole-transcriptome sequencing (Moens et al.,

2014), and these are being answered with examples in primary immunodeficiency cases such as Dock8 CID, GATA2 deficiency, and X-linked reticulate pigmentary disorder (XLPDR) (Hsu et al., 2013; Khan et al., 2016; Starokadomskyy et al., 2016). Over the coming years, an extended diagnostic approach to PID testing may develop that builds on a clinical module of phenotype, family history, and baseline immunological testing. This will be complemented by a DNA module of coding and non-coding variant analysis, utilizing sophisticated bioinformatic pipelines to prioritize candidate genetic variants of new loci that would be consistent with the clinical phenotype and family segregation. These candidate variants for monogenic disease may then be functionally interrogated *via* RNASeq for an influence within the gene itself and possibly the network within which it operates. In parallel, functional testing of candidate genes through protein-based assays may be undertaken to characterize the impact of a putative monogenic pathogenic variant within a reductionist model at the protein level. The sharing of these modular assessments across the international community will incrementally improve the standardized analysis of novel variants that will continue to grow over the next few years.

AUTHOR CONTRIBUTIONS

DB and AW devised and planned the project and co-supervised JL. JL wrote the manuscript. DB and AW reviewed, edited and contributed to the manuscript.

FUNDING

This work was funded by a NIHR research (grant number: NIHR RP2026-07-011) professorship to DB.

GLOSSARY OF TERMS

Term	Description
NGS (next-generation sequencing)	Next-generation sequencing describes the massively parallel, high-throughput modern evolution of sequencing.
WES (whole-exome sequencing)	Whole-exome sequencing is the process of reading and recording the nucleotide sequence of all the protein coding regions of a subject's genome (termed the exome).
WGS (whole-genome sequencing)	Whole-genome sequencing is the process of reading and recording the nucleotide sequence of the entirety of a subject's genome.
RNASeq (RNA sequencing)	RNA sequencing is the process of capturing and sequencing the available RNA from a tissue or cell. This can include all transcribed RNAs, or just the messenger RNA, depending on the capture method.
Short-read sequencing	Short-read sequencing is the process of sequencing a pre-fragmented library of nucleic acid, typically 75–300 bp in size.
Long-read sequencing	Long-read sequencing is the process of sequencing full-length nucleic acid fragments.

REFERENCES

- Abolhassani, H., Naseri, A., Rezaei, N., and Aghamohammadi, A. (2015). Economic burden of common variable immunodeficiency: annual cost of disease AU - Sadeghi, Bamdad. *Expert Rev. Clin. Immunol.* 11 (5), 681–688. doi: 10.1586/1744666X.2015.1029457
- Alkhatib, A., Werner, M., Hug, E., Herzog, S., Eschbach, C., and Faraidun, H. (2012). FoxO1 induces Ikaros splicing to promote immunoglobulin gene recombination. *J. Exp. Med.* 209 (2), 395–406. doi: 10.1084/jem.20110216
- Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* 59 (3), 253–268. doi: 10.1007/s13353-018-0444-7
- Bamshad, M. J., Ng, S. B., Bigam, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. et al. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745. doi: 10.1038/nrg3031
- Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15 (3), 163–175. doi: 10.1038/nrg3662
- Bonilla, F. A., and Oettgen, H. C. (2010). Adaptive immunity. *J. Allergy Clin. Immunol.* 125 (2 Suppl 2), S33–S40. doi: 10.1016/j.jaci.2009.09.017
- Bonilla, F. A., Khan, D. A., Ballas, Z. K., Chinen, J., Frank, M. M., Hsu, J. T. et al. (2015). Practice parameter for the diagnosis and management of primary immunodeficiency. *J. Allergy Clin. Immunol.* 136 (5), 1186–205.e1–78. doi: 10.1016/j.jaci.2015.04.049
- Bousifha, A., Jeddane, L., Picard, C., Ailal, F., Bobby Gaspar, H., Al-Herz, W. et al. (2018). The 2017 IUIS Phenotypic classification for primary immunodeficiencies. *J. Clin. Immunol.* 38 (1), 129–143. doi: 10.1007/s10875-017-0465-8
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G. et al. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* 100 (5), 695–705. doi: 10.1016/j.ajhg.2017.04.003
- Boyle, J. M., and Buckley, R. H. (2007). Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J. Clin. Immunol.* 27 (5), 497–502. doi: 10.1007/s10875-007-9103-1
- Brodin, P., Pojic, V., Gao, T., Bhattacharya, S., Angel, C. J., and Furman, D. et al. (2015). Variation in the human immune system is largely driven by non-heritable influences. *Cell* 160 (1–2), 37–47. doi: 10.1016/j.cell.2014.12.020
- Bryois, J., Buil, A., Evans, D. M., Kemp, J. P., Montgomery, S. B., Conrad, D. F. et al. (2014). Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 10 (7), e1004461. doi: 10.1371/journal.pgen.1004461
- Campos-Sanchez, E., Martínez-Cano, J., del Pino Molina, L., López-Granados, E., and Cobaleda, C. (2019). Epigenetic Deregulation in Human Primary Immunodeficiencies. *Trends In Immunol.* 40 (1), 49–65. doi: 10.1016/j.it.2018.11.005
- Casamassimi, A., Federico, A., Rienzo, M., Esposito, S., and Ciccodicola, A. (2017). Transcriptome Profiling in human diseases: new advances and perspectives. *Int. J. Mol. Sci.* 18 (8), 1652. doi: 10.3390/ijms18081652
- Chen, S., Townsend, K., Goldberg, T. E., Davies, P., and Conejero-Goldberg, C. (2010). MAPI2 isoforms: differential transcriptional profiles related to 3R and 4R splice variants. *J. Alzheimers Dis.* 22 (4), 1313–1329. doi: 10.3233/JAD-2010-101155
- Chitnis, N., Clark, P. M., Kamoun, M., Stolle, C., Brad Johnson, F., and Monos, D. S. (2017). An expanded role for HLA genes: HLA-B encodes a microRNA that regulates IGA and other immune response transcripts. *Front. Immunol.* 8 (583). doi: 10.3389/fimmu.2017.00583
- Condino-Neto, A., and Espinosa-Rosales, F. J. (2018). Changing the Lives of people with primary immunodeficiencies (pi) with early testing and diagnosis. *Front. In Immunol.* 9 (1439). doi: 10.3389/fimmu.2018.01439
- Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R. et al. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9 (386), eaa5209. doi: 10.1126/scitranslmed.aal5209
- DiStefano, J. K. (2018). The Emerging Role of Long Noncoding RNAs in Human Disease. *Methods Mol. Biol. (Clifton NJ)* 1706, 91–110. doi: 10.1007/978-1-4939-7471-9_6
- Duffy, D., Rouilly, V., Braudeau, C., Corbiere, V., Djebali, R., Ungeheuer, M. N. et al. (2017). Standardized whole blood stimulation improves immunomonitoring of induced immune responses in multi-center study. *Clin. Immunol. (Orlando Fla)* 183, 325–335. doi: 10.1016/j.clim.2017.09.019
- Ergun, A., Doran, G., Costello, J. C., Paik, H. H., Collins, J. J., Mathis, D. et al. (2013). Differential splicing across immune system lineages. *Proc. Natl. Acad. Sci. U. S. A.* 110 (35), 14324–14329. doi: 10.1073/pnas.1311839110
- Eshel, D., Toporik, A., Efrati, T., Nakav, S., Chen, A., and Douvdevani, A. (2008). Characterization of natural human antagonistic soluble CD40 isoforms produced through alternative splicing. *Mol. Immunol.* 46 (2), 250–257. doi: 10.1016/j.molimm.2008.08.280
- Europe PIPDDfOCi. European Reference Paper. P1-16 worldpiweek.org.
- Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E. et al. (2014). Innate Immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343 (6175), 1246949. doi: 10.1126/science.1246949
- Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology. *Genomics Proteomics Bioinf.* 13 (1), 4–16. doi: 10.1016/j.gpb.2015.01.009
- Fischer, A. (1993). Primary T-cell immunodeficiencies. *Curr. Opin. In Immunol.* 5 (4), 569–578. doi: 10.1016/0952-7915(93)90040-Y
- Gallo, V., Dotta, L., Giardino, G., Cirillo, E., Lougaris, V., D'Assante, R. et al. (2016). Diagnostics of primary immunodeficiencies through next-generation sequencing. *Front. In Immunol.* 7, 466. doi: 10.3389/fimmu.2016.00466
- Gilissen, C., Hoischen, A., Brunner, H. G., and Veltman, J. A. (2011). Unlocking Mendelian disease using exome sequencing. *Genome Biol.* 12 (9), 228. doi: 10.1186/gb-2011-12-9-228
- Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P. et al. (2019). Expanding the boundaries of rna sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* 104 (3), 466–483. doi: 10.1016/j.ajhg.2019.01.012
- Grodecká, L., Hujová, P., Kramárek, M., Kršjaková, T., Kováčková, T., Vondrášková, K. et al. (2017). Systematic analysis of splicing defects in selected primary immunodeficiencies-related genes. *Clin. Immunol.* 180, 33–44. doi: 10.1016/j.clim.2017.03.010
- Han, H., and Jiang, X. (2014). Disease Biomarker Query from RNA-Seq Data. *Cancer Inform.* 13 (Suppl 1), 81–94. doi: 10.4137/CIN.S13876
- Heimall, J., Keller, M., Saltzman, R., Bunin, N., McDonald-McGinn, D., Zakai, E. et al. (2012). Diagnosis of 22q11.2 deletion syndrome and artemis deficiency in two children with T-B-NK+ immunodeficiency. *J. Clin. Immunol.* 32 (5), 1141–1144. doi: 10.1007/s10875-012-9741-9
- Heimall, J. (2019). Now is the time to use molecular gene testing for the diagnosis of primary immune deficiencies. *J. Allergy Clin. Immunol. In Pract.* 39 (1), 129–140. doi: 10.1016/j.jaip.2018.12.022
- Heremans, J., Garcia-Perez, J. E., Turro, E., Schlenner, S. M., Casteels, I., Collin, R. et al. (2018). Abnormal differentiation of B cells and megakaryocytes in patients with Roifman syndrome. *J. Allergy Clin. Immunol.* 142 (2), 630–646. doi: 10.1016/j.jaci.2017.11.061
- Hernandez-Trujillo, V., and Ballow, M. (2015). Diagnosing primary immunodeficiency: a practical approach for the non-immunologist AU - Lehman, Heather. *Curr. Med. Res. Opin.* 31 (4), 697–706. doi: 10.1185/03007995.2014.1001063
- Heyd, F., ten Dam, G., and Moroy, T. (2006). Auxiliary splice factor U2AF26 and transcription factor Gfi1 cooperate directly in regulating CD45 alternative splicing. *Nat. Immunol.* 7 (8), 859–867. doi: 10.1038/ni1361
- Hsu, A. P., Johnson, K. D., Falcone, E. L., Sanalkumar, R., Sanchez, L., Hickstein, D. D. et al. (2013). GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121 (19), 3830–7, s1–7. doi: 10.1182/blood-2012-08-452763
- Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (8), 96. doi: 10.1038/s12276-018-0071-8
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., and Li, Y. I. et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176 (3), 535–48.e24. doi: 10.1016/j.cell.2018.12.015
- Jia, C., Hu, Y., Liu, Y., and Li, M. (2015). Mapping Splicing Quantitative Trait Loci in RNA-Seq. *Cancer Inform.* 14 (Suppl 1), 45–53. doi: 10.4137/CIN.S24832
- Khan, S., Kuruvilla, M., Hagin, D., Wakeland, B., Liang, C., Vishwanathan, K. et al. (2016). RNA sequencing reveals the consequences of a novel insertion in dedicator of cytokinesis-8. *J. Allergy Clin. Immunol.* 138 (1), 289–92.e6. doi: 10.1016/j.jaci.2015.11.033
- Kim, H. K., Pham, M. H. C., Ko, K. S., Rhee, B. D., and Han, J. (2018). Alternative splicing isoforms in health and disease. *Pflügers Arch. Eur. J. Physiol.* 470 (7), 995–1016. doi: 10.1007/s00424-018-2136-x

- Kimmerling, R. J., Lee Szeto, G., Li, J. W., Genshaft, A. S., Kazer, S. W., Payer, K. R. et al. (2016). A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat. Commun.* 7 (1), 10220. doi: 10.1038/ncomms10220
- Kramer, N. J., Wang, W.-L., Reyes, E. Y., Kumar, B., Chen, C.-C., Ramakrishna, C. et al. (2015). Altered lymphopoiesis and immunodeficiency in c-miR-142 null mice. *Blood* 125 (24), 3720. doi: 10.1182/blood-2014-10-603951
- Krawczak, M., Reiss, J., and Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* 90 (1–2), 41–54. doi: 10.1007/BF00210743
- Krawczak, M., Thomas, N. S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J. et al. (2007). Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* 28 (2), 150–158. doi: 10.1002/humu.20400
- Kremer, L. S., Wortmann, S. B., and Prokisch, H. (2018). "Transcriptomics": molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *J. Inher. Metab. Dis.* 41 (3), 525–532. doi: 10.1007/s10545-017-0133-4
- Kukurba, K. R., and Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.* 2015 (11), 951–969. doi: 10.1101/pdb.top084970
- Lee, M. N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T. M. et al. (2014). Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science* 343 (6175), 1246980. doi: 10.1126/science.1246980
- Lenardo, M., Lo, B., and Lucas, C. L. (2016). Genomics of Immune Diseases and New Therapies. *Annu. Rev. Immunol.* 34, 121–149. doi: 10.1146/annurev-immunol-041015-055620
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput. Biol.* 13 (5), e1005457. doi: 10.1371/journal.pcbi.1005457
- Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011). What can exome sequencing do for you? *J. Med. Genet.* 48 (9), 580–589. doi: 10.1136/jmedgenet-2011-100223
- McCusker, C., and Warrington, R. (2011). Primary immunodeficiency. *Allergy Asthma Clin. Immunol.* 7 (Suppl 1), S11–S18. doi: 10.1186/1710-1492-7-S1-S11
- McCusker, C., Upton, J., and Warrington, R. (2018). Primary immunodeficiency. *Allergy Asthma Clin. Immunol.* 14 (2), 61. doi: 10.1186/s13223-018-0290-5
- Meininger, I., Griesbach, R. A., Hu, D., Gehring, T., Seeholzer, T., Bertossi, A. et al. (2016). Alternative splicing of MALT1 controls signalling and activation of CD4(+) T cells. *Nat. Commun.* 7, 11292. doi: 10.1038/ncomms11292
- Merico, D., Roifman, M., Braunschweig, U., Yuen, R. K., Alexandrova, R., Bates, A. et al. (2015). Compound heterozygous mutations in the noncoding RNU4ATAC cause Roifman Syndrome by disrupting minor intron splicing. *Nat. Commun.* 6, 8718. doi: 10.1038/ncomms9718
- Meyts, I., Bosch, B., Bolze, A., Boisson, B., Itan, Y., Belkadi, A. et al. (2016). Exome and genome sequencing for inborn errors of immunity. *J. Allergy Clin. Immunol.* 138 (4), 957–969. doi: 10.1016/j.jaci.2016.08.003
- Moen, L. N., Falk-Sörqvist, E., Asplund, A. C., Bernatowska, E., Smith, C. I. E., and Nilsson, M. (2014). Diagnostics of primary immunodeficiency diseases: a sequencing capture approach. *PLoS One* 9 (12), e114901. doi: 10.1371/journal.pone.0114901
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., and Salichos, L. et al. (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol.* 17, 53. doi: 10.1186/s13059-016-0917-0
- Neums, L., Suenaga, S., Beyerlein, P., Anders, S., Koestler, D., Mariani, A. et al. (2017). VaDiR: an integrated approach to Variant Detection in RNA. *GigaScience* 7 (2), 1–13. doi: 10.1093/gigascience/gix122
- Oeckinghaus, A., Wegener, E., Welteke, V., Ferch, U., Arslan, S. C., Ruland, J. et al. (2007). Malt1 ubiquitination triggers NF-kappaB signaling upon T-cell activation. *EMBO J.* 26 (22), 4634–4645. doi: 10.1038/sj.emboj.7601897
- 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
- Orvain, C., Matre, V., and Gabrielsen, O. S. (2008). The transcription factor c-Myb affects pre-mRNA splicing. *Biochem. Biophys. Res. Commun.* 372 (2), 309–313. doi: 10.1016/j.bbrc.2008.05.054
- 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 (12), 1413–1415. doi: 10.1038/ng.259
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14 (4), 417–419. doi: 10.1038/nmeth.4197
- Philippidis, A. (2018). The 100,000 Genomes club www.genenews.com [Available from: <https://www.genenews.com/insights/the-100000-genomes-club/>].
- Piasecka, B., Duffy, D., Urrutia, A., Quach, H., Patin, E., Posseme, C. et al. (2018). Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *Proc. Natl. Acad. Sci. U. S. A.* 115 (3), e488–e497. doi: 10.1073/pnas.1714765115
- Picard, C., Bobby Gaspar, H., Al-Herz, W., Bousfiha, A., Casanova, J.-L., Chatila, T. et al. (2018). International union of immunological societies: 2017 primary immunodeficiency diseases committee report on inborn errors of immunity. *J. Clin. Immunol.* 38 (1), 96–128. doi: 10.1007/s10875-017-0464-9
- Punwani, D., Wang, H., Chan, A. Y., Cowan, M. J., Mallott, J., Sunderam, U. et al. (2015). Combined immunodeficiency due to MALT1 mutations, treated by hematopoietic cell transplantation. *J. Clin. Immunol.* 35 (2), 135–146. doi: 10.1007/s10875-014-0125-1
- Rae, W., Ward, D., Mattocks, C., Pengelly, R. J., Eren, E., Patel, S. V. et al. (2018). Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnostics. *Clin. Genet.* 93 (3), 647–655. doi: 10.1111/cge.13163
- Ramakrishnan, K. A., Pengelly, R. J., Gao, Y., Morgan, M., Patel, S. V., Davies, E. G. (2016). Precision molecular diagnosis defines specific therapy in combined immunodeficiency with megaloblastic anemia secondary to MTHFD1 deficiency. *J. Allergy Clin. Immunol. In Pract.* 4 (6), 1160–e6. doi: 10.1016/j.jaip.2016.07.014
- Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerve, H., Bertolino, E., Chen, Z. et al. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat. Immunol.* 9 (8), 927–936. doi: 10.1038/ni.1626
- Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinf.* 13 (5), 278–289. doi: 10.1016/j.gpb.2015.08.002
- Richardson, A. M., Moyer, A. M., Hasadsri, L., and Abraham, R. S. (2018). Diagnostic tools for inborn errors of human immunity (primary immunodeficiencies and immune dysregulatory diseases). *Curr. Allergy Asthma Rep.* 18 (3), 19. doi: 10.1007/s11882-018-0770-1
- Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A. et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317. doi: 10.1038/nature14248
- Scacheri, C. A., and Scacheri, P. C. (2015). Mutations in the noncoding genome. *Curr. Opin. Pediatr.* 27 (6), 659–664. doi: 10.1097/MOP.0000000000000283
- Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J. et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* 175 (6), 1701–15.e16. doi: 10.1016/j.cell.2018.10.022
- Schwarze, K., Buchanan, J., Taylor, J. C., and Wordsworth, S. (2018). Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet. In Med.* 20 (10), 1122–1130. doi: 10.1038/gim.2017.247
- See, P., Lum, J., and Chen, J. (2018). Ginhoux F A single-cell sequencing guide for immunologists. *Front. Immunol.* 9, 2425. doi: 10.3389/fimmu.2018.02425
- Shillitoe, B., Bangs, C., Guzman, D., Gennery, A. R., Longhurst, H. J., Slatter, M. et al. (2018). The United Kingdom Primary Immune Deficiency (UKPID) registry 2012 to 2017. *Clin. Exp. Immunol.* 192 (3), 284–291. doi: 10.1111/cei.13125
- Starokadomskyy, P., Gemelli, T., Rios, J. J., Xing, C., Wang, R. C., and Li, H. et al. (2016). DNA polymerase- α regulates the activation of type I interferons through cytosolic RNA:DNA synthesis. *Nat. Immunol.* 17 (5), 495–504. doi: 10.1038/ni.3409
- Stray-Pedersen, A., Sorte, H. S., Samarakoon, P., Gambin, T., Chinn, I. K., Coban Akdemir, Z. H. et al. (2017). Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. *J. Allergy Clin. Immunol.* 139 (1), 232–245. doi: 10.1016/j.jaci.2016.05.042
- Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among

- schizophrenia-associated loci. *Nat. Commun.* 8 (1), 14519. doi: 10.1038/ncomms14519
- Takeda, N., O'Dea, E. L., Doedens, A., J-w, Kim, Weidemann, A., Stockmann, C. et al. (2010). Differential activation and antagonistic function of HIF- α isoforms in macrophages are essential for NO homeostasis. *Genes Dev.* 24 (5), 491–501. doi: 10.1101/gad.1881410
- Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A. et al. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.* 47 (7), 717–726. doi: 10.1038/ng.3304
- The Genotype-Tissue Expression (GTEx) (2015) Pilot analysis: Multitissue gene regulation in humans. *Science* 348 (6235), 648. doi: 10.1126/science.1262110
- Thormann, A., Halachev, M., McLaren, W., Moore, D. J., Svinti, V., Campbell, A. et al. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat. Commun.* 10 (1), 2373. doi: 10.1038/s41467-019-10016-3
- Turunen, J. J., Niemela, E. H., Verma, B., and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4 (1), 61–76. doi: 10.1002/wrna.1141
- UK, P. Patients' experience survey of Primary Immunodeficiency Disorders services. www.piduk.org; 2016 September 2016.
- Urrutia, A., Duffy, D., Rouilly, V., Posseme, C., Djebali, R., Illanes, G. et al. (2016). Standardized Whole-Blood Transcriptional Profiling Enables the Deconvolution of Complex Induced Immune Responses. *Cell Rep.* 16 (10), 2777–2791. doi: 10.1016/j.celrep.2016.08.011
- Walter, J. E., Farmer, J. R., Foldvari, Z., Torgerson, T. R., and Cooper, M. A. (2016). Mechanism-Based strategies for the management of autoimmunity and immune dysregulation in primary immunodeficiencies. *J. Allergy Clin. Immunol. In Pract.* 4 (6), 1089–1100. doi: 10.1016/j.jaip.2016.08.004
- Wang, Y., Liu, J., Huang, B. O., Xu, Y.-M., Li, J., Huang, L.-F. et al. (2015). Mechanism of alternative splicing and its regulation. *BioMed. Rep.* 3 (2), 152–158. doi: 10.3892/br.2014.407
- Ward, A. J., and Cooper, T. A. (2010). The pathobiology of splicing. *J. Pathol.* 220 (2), 152–163. doi: 10.1002/path.2649
- Wirka, R. C., Pjanic, M., and Quertermous, T. (2018). Advances in Transcriptomics: Investigating Cardiovascular Disease at Unprecedented Resolution. *Circ. Res.* 122 (9), 1200–1220. doi: 10.1161/CIRCRESAHA.117.310910
- Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A. et al. (2013). Clinical Whole-exome sequencing for the diagnosis of mendelian disorders. *New Engl. J. Med.* 369 (16), 1502–1511. doi: 10.1056/NEJMoa1306555
- Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y. et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312 (18), 1870–1879. doi: 10.1001/jama.2014.14601
- Zeng, Y., Wang, G., Yang, E., Ji, G., Brinkmeyer-Langford, C. L., Cai, J. J. (2015). Aberrant Gene Expression in Humans. *PLoS Genet.* 11 (1), e1004942. doi: 10.1371/journal.pgen.1004942
- Zhao, J., Akisanmi, I., Arafat, D., Cradick, T. J., Lee, C. M., Banskota, S. et al. (2016). A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am. J. Hum. Genet.* 98 (2), 299–309. doi: 10.1016/j.ajhg.2015.12.023
- Zhu, L.-Y., Zhu, Y.-R., Dai, D.-J., Wang, X., and Jin, H.-C. (2018). Epigenetic regulation of alternative splicing. *Am. J. Cancer Res.* 8 (12), 2346–2358.
- Zikherman, J., and Weiss, A. (2008). Alternative splicing of CD45: the tip of the iceberg. *Immunity* 29 (6), 839–841. doi: 10.1016/j.immuni.2008.12.005

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

Copyright © 2019 Lye, Williams and Baralle. 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.



Consequences of Making the Inactive Active Through Changes in Antisense Oligonucleotide Chemistries

Khine Zaw^{1,2}, Kane Greer^{1,3}, May Thandar Aung-Htut^{1,3}, Chalermchai Mitrpant^{2,3}, Rakesh N. Veedu^{1,3}, Sue Fletcher^{1,3} and Steve D. Wilton^{1,3*}

¹ Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA, Australia, ² Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, ³ Perron Institute for Neurological and Translational Science and Centre for Neuromuscular and Neurological Disorders, The University of Western Australia, Perth, WA, Australia

OPEN ACCESS

Edited by:

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

Reviewed by:

Argyris Papantonis,
University Medical Center Göttingen,
Germany
Mainá Bitar,
Federal University of Minas Gerais,
Brazil

*Correspondence:

Steve D. Wilton
s.wilton@murdoch.edu.au

Specialty section:

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

Received: 30 August 2019

Accepted: 13 November 2019

Published: 20 December 2019

Citation:

Zaw K, Greer K, Aung-Htut MT,
Mitrapant C, Veedu RN, Fletcher S and
Wilton SD (2019) Consequences of
Making the Inactive Active Through
Changes in Antisense
Oligonucleotide Chemistries.
Front. Genet. 10:1249.
doi: 10.3389/fgene.2019.01249

Antisense oligonucleotides are short, single-stranded nucleic acid analogues that can interfere with pre-messenger RNA (pre-mRNA) processing and induce excision of a targeted exon from the mature transcript. When developing a panel of antisense oligonucleotides to skip every dystrophin exon, we found great variation in splice switching efficiencies, with some antisense oligonucleotides ineffective, even when directed to canonical splice sites and transfected into cells at high concentrations. In this study, we re-evaluated some of these ineffective antisense oligonucleotide sequences after incorporation of locked nucleic acid residues to increase annealing potential. Antisense oligonucleotides targeting exons 16, 23, and 51 of human *DMD* transcripts were synthesized as two different chemistries, 2'-O-methyl modified bases on a phosphorothioate backbone or mixmers containing several locked nucleic acid residues, which were then transfected into primary human myotubes, and *DMD* transcripts were analyzed for exon skipping. The ineffective 2'-O-methyl modified antisense oligonucleotides induced no detectable exon skipping, while all corresponding mixmers did induce excision of the targeted exons. Interestingly, the mixmer targeting exon 51 induced two unexpected transcripts arising from partial skipping of exon 51 with retention of 95 or 188 bases from the 5' region of exon 51. These results indicated that locked nucleic acid/2'-O-methyl mixmers are more effective at inducing exon skipping, however, this improvement may come at the cost of activating alternative cryptic splice sites and off-target effects on gene expression.

Keywords: *DMD*, antisense oligonucleotide, locked nucleic acid, locked nucleic acid/2'-O-methyl mixmer, cryptic splice site

INTRODUCTION

Mutations in the *DMD* gene are responsible for Duchenne (DMD) and Becker (BMD) muscular dystrophies. The more severe DMD is typically associated with frameshifting deletions, duplications, or insertions, or nonsense mutations that cause disruption of the open reading frame (ORF). Most mutations that do not disrupt the ORF produce an internally truncated but partially functional

protein, resulting in the milder BMD phenotype (Monaco et al., 1988). This spectrum of disease severity underlies the development of therapeutic interventions such as antisense oligonucleotide (AO) induced targeted exon skipping to treat DMD (Fletcher et al., 2017; Stein and Castanotto, 2017).

AOs are short, single-stranded nucleic acid analogues that are designed to anneal to a messenger RNA (mRNA) or pre-mRNA through Watson–Crick base pairing interactions and, depending on the base and backbone chemistries, induce a variety of mechanisms to alter the gene expression. AO induced splice-switching strategies to skip one or more specific exons, with restoration of the ORF and expression of dystrophin with improved function has been explored as a treatment for DMD. *Exondys 51* (Sarepta Therapeutics), a phosphorodiamidate morpholino oligomer targeting exon 51, has been given accelerated approval by the US Food and Drug Administration (FDA) for the treatment of DMD (Syed, 2016).

Strategies to improve AO potency through more efficient cellular uptake or increased stability and specificity of AOs are continually being explored to develop compounds that would confer optimal therapeutic effects. Approaches to enhance AO potency involve chemical modifications of the phosphate backbone or at the 2' position of the ribose sugar, such as 2'-O-methyl (2'-OMe) or locked nucleic acid (LNA) in order to increase binding affinity and resistance against nuclease degradation (Figure 1A). LNA-modified oligonucleotides show a high binding affinity and increased stability against nuclease degradation compared to only the 2'-OMe modification (Kierzek et al., 2009).

When developing a panel of AOs to skip all dystrophin exons, remarkably we found that two out of three compounds could induce some level of exon skipping, albeit at variable efficiencies. During development of splice switching AOs, targeting many of the canonical donor or acceptor splice sites in the dystrophin primary transcript appeared to be largely ineffective, especially when compared to targeting intra-exonic splice enhancer motifs (Adams et al., 2007). The current study aimed to ascertain if increasing the annealing potential of an AO by incorporating LNA residues at selected positions could influence splicing. We selected three examples of acceptor or donor splice sites that were ineffective as 2'-OMe AO target sites. We found that the LNA/2'-OMe AO mixmers with increased annealing potential were capable of modifying pre-mRNA processing, indicating that these AOs could act as splice switching compounds if the strength of annealing was sufficiently increased. However, in one case, multiple transcripts were induced due to the activation of intra-exonic cryptic splice sites, suggesting that enhanced annealing may compromise splice switching specificity.

MATERIALS AND METHODS

Design and Synthesis of Chemically Modified AOs

2'-OMe AOs, previously designed, evaluated, and found to be inactive with respect to inducing skipping of human dystrophin exons 16, 23, and 51 (Harding et al., 2007;

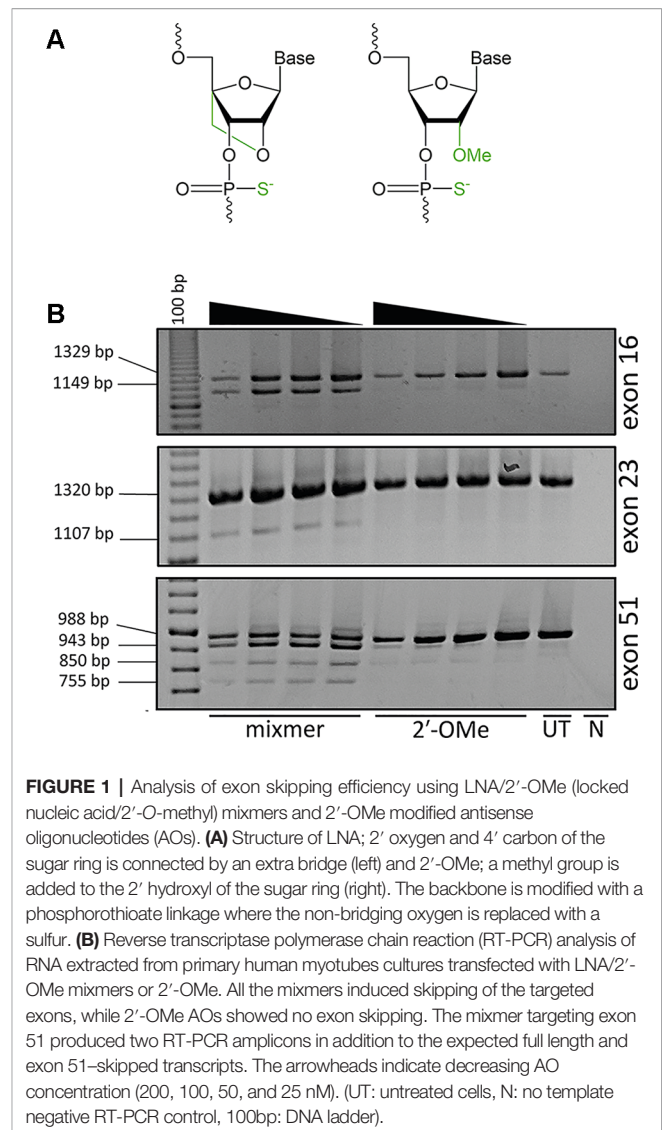


FIGURE 1 | Analysis of exon skipping efficiency using LNA/2'-OMe (locked nucleic acid/2'-O-methyl) mixmers and 2'-OMe modified antisense oligonucleotides (AOs). **(A)** Structure of LNA; 2' oxygen and 4' carbon of the sugar ring is connected by an extra bridge (left) and 2'-OMe; a methyl group is added to the 2' hydroxyl of the sugar ring (right). The backbone is modified with a phosphorothioate linkage where the non-bridging oxygen is replaced with a sulfur. **(B)** Reverse transcriptase polymerase chain reaction (RT-PCR) analysis of RNA extracted from primary human myotubes cultures transfected with LNA/2'-OMe mixmers or 2'-OMe. All the mixmers induced skipping of the targeted exons, while 2'-OMe AOs showed no exon skipping. The mixmer targeting exon 51 produced two RT-PCR amplicons in addition to the expected full length and exon 51-skipped transcripts. The arrowheads indicate decreasing AO concentration (200, 100, 50, and 25 nM). (UT: untreated cells, N: no template negative RT-PCR control, 100bp: DNA ladder).

Mitrapant et al., 2009), were resynthesized in house on a GE AKTA Oligopilot plus 10 (GE Healthcare Life Sciences) oligonucleotide synthesizer, as described previously using the 1 μ mol thioate protocol (Le et al., 2017). The mixmers incorporating the LNAs are described in Table 1 and were also synthesized in house (Le et al., 2017).

Cell Culture and Transfection

Primary human myoblasts, obtained after informed consent and approved by the Murdoch University human ethics committee (#2013_156), were cultured and differentiated into myotubes as described by Rando and Blau (Rando and Blau, 1994) with minor modifications (Harding et al., 2007). Briefly, myoblasts were seeded on 24-well plates coated with poly-D-lysine (Merck Millipore), and Matrigel (Corning, supplied through In Vitro Technologies) at a density of 30,000 cells/well. Cells were differentiated into myotubes in 5% horse serum low glucose

TABLE 1 | Sequences of AOs used in this study.

Sequence	LNA/2'-OMe Mixmers	2'-OMe
H16A (-07+13)	C CGCUUUUAA A ACCUGUUA	CCGCUUUUAA A ACCUGUUA
H23D (+07-18)	AGU A AAAU CU UGAAU U ACCUGAAU U	AGU A AAAU CU UGAAU U ACCUGAAU U
H51D (+07-18)	UA U CAU UUUUU CUCAU ACC U UCUG C	UA U CAU UUUUU CUCAU ACC U UCUG C

LNA nucleotide monomers are represented as bold characters.

Dulbecco's modified Eagle medium (DMEM) (Thermo Fisher Scientific) by incubating at 37°C in 5% CO₂ for 48 h. All 2'-OMe AOs and mixmers were transfected into differentiated myotubes in Opti-MEM (Invitrogen) as cationic lipoplexes with Lipofectamine 2000 reagent at 1:1 w:w ratio, according to the manufacturer's instructions (Invitrogen) in a final transfection volume of 500 µl/well in a 24-well plate. Transfected cells were incubated for 48 h before total RNA extraction.

RNA Extraction and RT-PCR

RNA extraction was carried out using the MagMAX-96 Total RNA Isolation Kit (Life Technologies), according to the manufacturer's instructions. RT-PCRs were performed using the One-Step SuperScript III RT-PCR kit (Life Technologies) as per manufacturer's instructions. The temperature profile was 55°C for 30 min, 94°C for 2 min, followed by 35 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 1 min 10 s. RT-PCR products were separated on 2% agarose gels in Tris-acetate-ethylenediaminetetraacetic acid (EDTA) buffer, and the images were captured on a Fusion FX gel documentation system (Vilber Lourmat, Marne-la-Vallee, France).

Band-Stab PCR and Sequencing

Individual bands representing RT-PCR amplicons were purified and amplified by band-stab PCR (Wilton et al., 1997) using AmpliTaq Gold DNA Polymerases (Thermo Fisher Scientific) with the thermal profile of 94°C for 5 min followed by 25 cycles of 94°C for 15 s, 50°C

for 15 s, 72°C for 1 min with the final extension of 72°C for 5 min. Amplicon sequences were identified by Sanger sequencing at the Australian Genome Research Facility (AGRF, Perth, Australia). The nucleotide sequences were deposited at GenBank and available as accession numbers MN490082–MN490085.

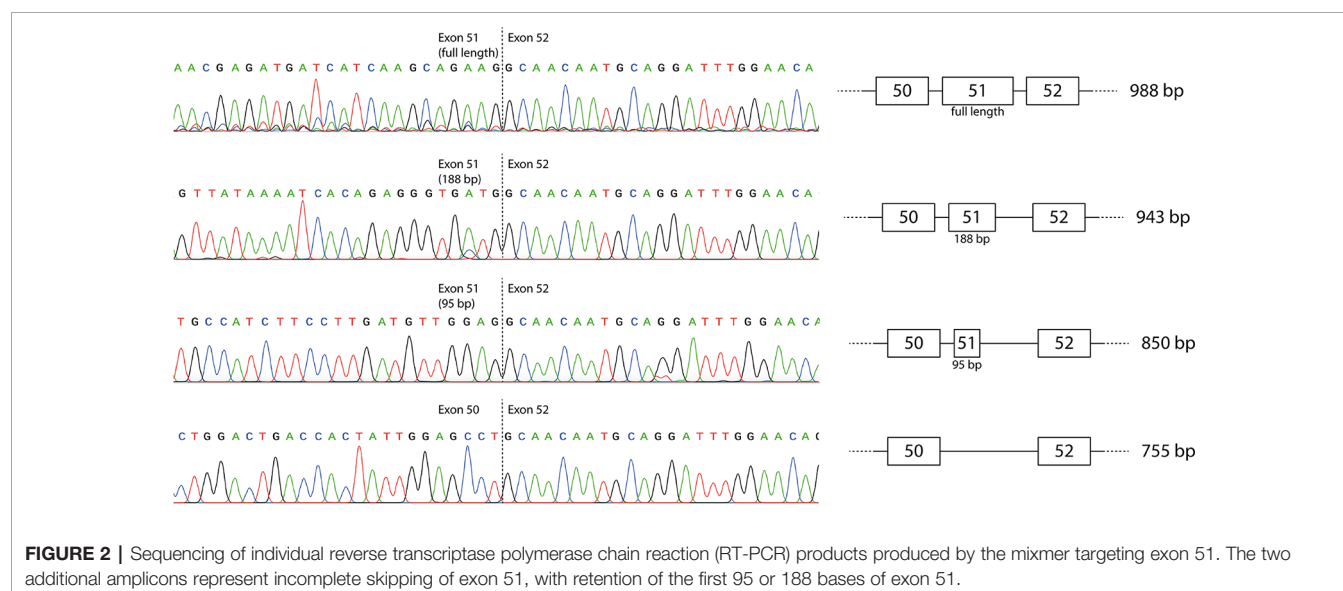
In Silico Analysis

The natural and potential donor splice sites of *DMD* exon 51 were analyzed using Human Splicing Finder 3.1 (<http://www.umd.be/HSF3/HSF.shtml>) (Desmet et al., 2009).

RESULTS

AO Synthesis and Transfection

In this study, we used AO sequences that had been designed to induce skipping of exons 16, 23, and 51 from the human dystrophin gene transcript, but were previously reported to be largely ineffective when transfected into cells as 2'-OMe modified compounds on a phosphorothioate backbone. These sequences were re-synthesized as 2'-OMe AOs and as respective LNA-2'-OMe "mixmers" on a phosphorothioate backbone, a 20-mer for exon 16 and 25-mers for exons 23 and 51, as shown in **Table 1**. Primary human myoblasts obtained after informed consent and approved by the Murdoch University human ethics committee (#2013_156) were cultured on 24-well plates, differentiated over 48 h, and transfected with AO cationic lipoplexes at 200, 100, 50, and 25 nM concentrations. The cells transfected with either 2'-



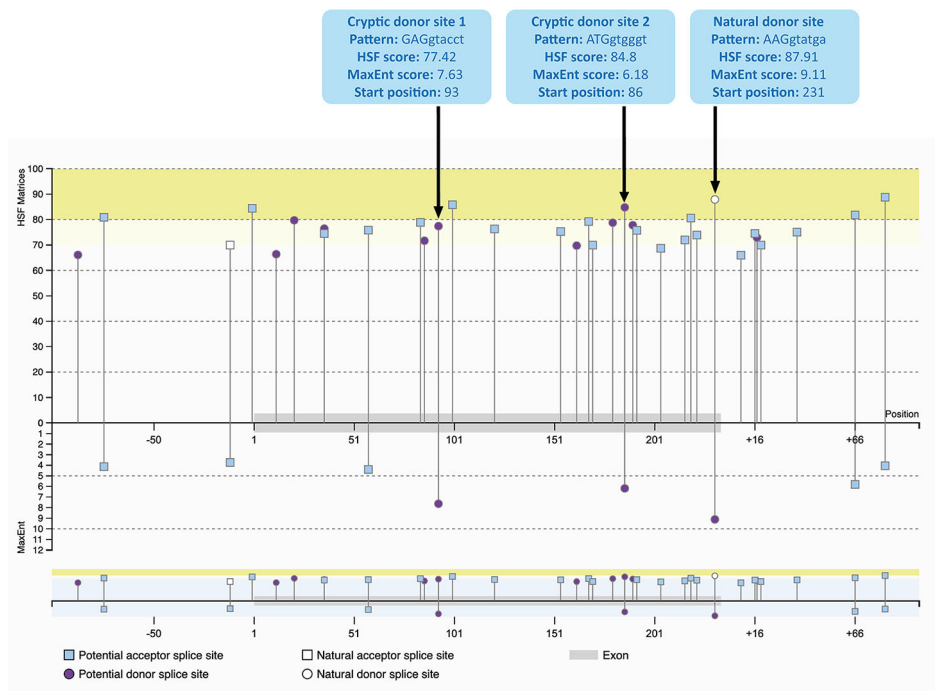


FIGURE 3 | Potential splice sites for human *DMD* exon 51 predicted by Human Splicing Finder 3.1 (HSF) indicating the natural donor splice site and the two cryptic donor splice sites activated by the mixmers. The splice sites scores are predicted for each site (Yeo and Burge, 2004; Desmet et al., 2009). The splice site score for the cryptic donor site 2 is higher than the cryptic donor site 1 in HSF score, although the cryptic donor site 1 has a higher score in the maximum entropy model.

OMe or LNA/2'-OMe mixer cationic lipoplexes were healthy and showed no obvious signs of toxicity or cell death at these concentrations.

Evaluation of AO Efficiency

The cells were collected 48 h after transfection and analyzed for respective *DMD* exon skipping by reverse transcriptase polymerase chain reaction (RT-PCR). As anticipated, the 2'-OMe AOs did not induce exon skipping while all the LNA/2'-OMe mixmers induced consistent skipping of the targeted exons at all tested concentrations (**Figure 1B**). Interestingly, the LNA/2'-OMe mixer targeting exon 51 induced two transcript products in addition to the predicted full-length products (988 bp) and exon 51-skipped (755 bp) transcript product. The 2'-OMe AO directed at dystrophin exon 51 donor splice site induced very weak exon skipping and also low levels of the 943 and 850 bp amplicons.

Identification of Full-Length and Skipped Transcripts

The individual products were purified by band-stab PCR (Wilton et al., 1997) and identified by direct DNA sequencing that confirmed the identity of the full-length and induced exon-skipped amplicons (**Figure 2**). The additional products generated predominantly by the exon 51 LNA/2'-OMe mixer arose from activation of cryptic donor splice sites, with retention of 95 or 188 bases from the beginning of exon 51 (850 and 943 bp amplicons respectively). The sequences are available at GenBank with the accession numbers MN490082–MN490085.

Analysis of Activated Cryptic Splice Sites

The cryptic splice sites activated by the mixer targeting exon 51 are potential splice sites within *DMD* exon 51 (**Figure 3**) as predicted by Human Splicing Finder 3.1 (HSF) (<http://www.umd.be/HSF3/HSF.shtml>) (Desmet et al., 2009). Both cryptic donor splice sites included the canonical donor splice site sequence, 'GT' at the 5' end of the "intron." The HSF scores for cryptic donor sites 1 and 2 are 77.42 and 84.8 respectively, which are high and close to the natural splice site score, 87.91. In the maximum entropy model, both cryptic sites are also the only positions predicted to be potential splice sites in addition to the natural donor site (Yeo and Burge, 2004). Interestingly, the splice site score for the cryptic donor site 2 is higher than the cryptic donor site 1 in HSF score, although the cryptic donor site 1 has a higher score in the maximum entropy model.

DISCUSSION

The most common consequence of mutations of the canonical donor or acceptor splice sites is exon skipping. However, the majority of dystrophin exons appeared unresponsive to splice switching AOs targeting these motifs, and only 2 exons out of the 77 were identified with the donor splice sites being optimal targets (Wilton et al., 2007). In this study, we modified three previously reported inactive AO sequences identified by ourselves and others, with LNA incorporation to improve the annealing affinity, and revisited their ability to induce exon

skipping. Consistent with the previous results, newly synthesized 2'-OMe AOs showed no or poor exon skipping after transfection, while all mixmers induced readily detectable skipping of the targeted exons.

During optimization of AOs to excise *DMD* exon 16, we previously demonstrated that AOs targeting the acceptor site and adjacent putative exonic splicing enhance sites could induce marked exon skipping, as long as the AOs were 25-mers or longer (Harding et al., 2007). While overlapping 25-mer and 31-mers induced robust exon 16 skipping, a shorter 20-mer common to all sequences was ineffective. Since shorter AOs are more efficiently and economically synthesized, we aim to identify the shortest sequence capable of inducing robust exon skipping, but in this instance, the 20-mer did not have sufficient annealing strength to the target because of the low 'GC' content (Harding et al., 2007). Incorporation of LNAs into this short and inactive 20-mer resulted in increased annealing potential and relatively robust exon skipping. This suggested that the limitation of the original 20-mer was due to weak hybridization.

A similar trend was observed for AO induced human dystrophin exon 23 skipping. We previously reported refining 2'-OMe AOs designed to induce exon 23 skipping and restore dystrophin expression in the *mdx* mouse (Mann et al., 2002). In subsequent studies, we found that the optimal annealing coordinates to skip mouse dystrophin exon 23, H23D (+07–18), were not an effective target for excising human dystrophin exon 23 (Mitropant et al., 2009). In this study, we synthesized the same sequence as a LNA/2'-OMe mixmer, and exon skipping was observed when AO was applied to cultured human myoblasts. Nevertheless, this increased annealing potential only resulted in modest levels of exon skipping and would not be considered for clinical application.

Dystrophin exon 51 was selected as the exon target for the first *DMD* exon skipping clinical trials since excising this exon would be relevant to the largest subset of *DMD* patients (Aartsma-Rus et al., 2009; Bladen et al., 2015). During the pre-clinical development of *Eteplirsen*, all the oligomers targeting the donor splice site of human dystrophin exon 51 were found to be largely ineffective (Arechavala-Gomez et al., 2007). A 25-mer 2'-OMe AO, H51D(+08–17), failed to skip exon 51, whereas another AO, a 23-mer with higher GC content, H51D(+16–07), induced only weak exon skipping (Harding et al., 2007). We resynthesized the AO sequence H51D (+07–18) as a mixmer and showed induction of three different transcripts: complete skipping of exon 51 together with two other transcripts produced by the activation of intra-exonic cryptic donor sites. The levels of complete exon 51 skipping were modest, and thus, such an approach would not be clinically applicable to restore the reading

frame. One of the few examples of activation of cryptic splice sites in the dystrophin gene transcript was in exon 53 of the mouse dystrophin gene, 77 bases upstream of the normal donor splice site (Mitropant et al., 2009). We anticipated that cryptic splice site activation would be far more common than we have actually encountered to date.

Our results show that the incorporation of LNA bases into the AOs increased the annealing potential and transformed inactive antisense sequences into active AOs. However, one significant concern that must be investigated is the extent to which cryptic splice sites are activated, leading to partial exon skipping. LNA-fully-modified AOs containing up to three mismatches induced dystrophin exon 46 skipping in *DMD* patient cells, indicating low sequence specificity, with the possibility of off-target binding potential (Aartsma-Rus et al., 2004). In addition, gapmers modified with LNAs show acute hepatotoxicity in mice (Kasuya et al., 2016). We and others have previously reported that AOs on a fully modified phosphorothioate backbone recruited nuclear proteins involved in RNA processing and induced global disturbance of cellular processes (Shen et al., 2014; Flynn et al., 2018).

In conclusion, we show that the incorporation of LNAs into 2'-OMe antisense sequences increased their potency as steric blockers of splicing, thereby making the inactive active. However, this enhancement came at a cost in efficiency and specificity due to activation of cryptic splicing, raising the risk of adverse and off-target effects elsewhere in the human transcriptome.

DATA AVAILABILITY STATEMENT

The nucleotide sequences were deposited at GenBank and available as accession number MN490082-MN490085.

AUTHOR CONTRIBUTIONS

Conceptualization, SW, SF, and RV. Experiments, KZ and KG. Writing—original draft preparation, KZ. Writing—review and editing, KZ, MA-H, SW, and SF. Supervision, MA-H, CM, SW, and SF.

FUNDING

This work is supported by the National Health and Medical Research Council (Australia) grant APP1144791.

REFERENCES

- Aartsma-Rus, A., Kaman, W. E., Bremmer-Bout, M., Janson, A. a. M., den Dunnen, J. T., van Ommen, G.-J. B., et al. (2004). Comparative analysis of antisense oligonucleotide analogs for targeted *DMD* exon 46 skipping in muscle cells. *Gene Ther.* 11, 1391. doi: 10.1038/sj.gt.3302313
- Aartsma-Rus, A., Fokkema, I., Verschuuren, J., Ginjaar, I., Deutekom, J., van Ommen, G.-J., et al. (2009). Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. *Hum. Mutat.* 30, 293–299. doi: 10.1002/humu.20918
- Adams, A. M., Harding, P. L., Iversen, P. L., Coleman, C., Fletcher, S., and Wilton, S. D. (2007). Antisense oligonucleotide induced exon skipping and the dystrophin gene transcript: cocktails and chemistries. *BMC Mol. Biol.* 8, 57. doi: 10.1186/1471-2199-8-57
- Arechavala-Gomez, V., Graham, I. r., Popplewell, L. j., Adams, A. m., Aartsma-Rus, A., Kinali, M., et al. (2007). Comparative analysis of antisense

- oligonucleotide sequences for targeted skipping of exon 51 during dystrophin pre-mRNA splicing in human muscle. *Hum. Gene Ther.* 18, 798–810. doi: 10.1089/hum.2006.061
- Bladen, C. L., Salgado, D., Monges, S., Foncuberta, M. E., Kekou, K., Kosma, K., et al. (2015). The TREAT-NMD DMD global database: analysis of more than 7,000 duchenne muscular dystrophy mutations. *Hum. Mutat.* 36, 395–402. doi: 10.1002/humu.22758
- Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., and Bérout, C. (2009). Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37, e67. doi: 10.1093/nar/gkp215
- Fletcher, S., Bellgard, M. I., Price, L., Akkari, A. P., and Wilton, S. D. (2017). Translational development of splice-modifying antisense oligomers. *Expert Opin. Biol. Ther.* 17, 15–30. doi: 10.1080/14712598.2017.1250880
- Flynn, L. L., Ruohan, L., Aung-Htut, M. T., Pitout, I. L., Cooper, J., Hubbard, A., et al. (2018). Interaction of modified oligonucleotides with nuclear proteins, formation of novel nuclear structures and sequence-independent effects on RNA processing. *bioRxiv*, 446773. doi: 10.1101/446773
- Harding, P. L., Fall, A. M., Honeyman, K., Fletcher, S., and Wilton, S. D. (2007). The influence of antisense oligonucleotide length on dystrophin exon skipping. *Mol. Ther.* 15, 157–166. doi: 10.1038/sj.mt.6300006
- Kasuya, T., Hori, S., Watanabe, A., Nakajima, M., Gahara, Y., Rokushima, M., et al. (2016). Ribonuclease H1-dependent hepatotoxicity caused by locked nucleic acid-modified gapmer antisense oligonucleotides. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep30377
- Kierzek, E., Pasternak, A., Pasternak, K., Gdaniec, Z., Yildirim, I., Turner, D. H., et al. (2009). Contributions of stacking, preorganization, and hydrogen bonding to the thermodynamic stability of duplexes between RNA and 2'-O-methyl RNA with locked nucleic acids. *Biochemistry* 48, 4377–4387. doi: 10.1021/bi9002056
- Le, B. T., Adams, A. M., Fletcher, S., Wilton, S. D., and Veedu, R. N. (2017). Rational design of short locked nucleic acid-modified 2'-O-methyl antisense oligonucleotides for efficient exon-skipping *in vitro*. *Mol. Ther. Nucleic Acids* 9, 155–161. doi: 10.1016/j.omtn.2017.09.002
- Mann, C. J., Honeyman, K., McClorey, G., Fletcher, S., and Wilton, S. D. (2002). Improved antisense oligonucleotide induced exon skipping in the mdx mouse model of muscular dystrophy. *J. Gene Med.* 4, 644–654. doi: 10.1002/jgm.295
- Mitrant, C., Adams, A. M., Meloni, P. L., Muntoni, F., Fletcher, S., and Wilton, S. D. (2009). Rational design of antisense oligomers to induce dystrophin exon skipping. *Mol. Ther.* 17, 1418–1426. doi: 10.1038/mt.2009.49
- Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H., and Kunkel, L. M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* 2, 90–95. doi: 10.1083/jcb.125.6.1275
- Rando, T. A., and Blau, H. M. (1994). Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J. Cell Biol.* 125, 1275–1287. doi: 10.1083/jcb.125.6.1275
- Shen, W., Liang, X., and Crooke, S. T. (2014). Phosphorothioate oligonucleotides can displace NEAT1 RNA and form nuclear paraspeckle-like structures. *Nucleic Acids Res.* 42, 8648–8662. doi: 10.1093/nar/gku579
- Stein, C. A., and Castanotto, D. (2017). FDA-approved oligonucleotide therapies in 2017. *Mol. Ther.* 25, 1069–1075. doi: 10.1016/j.ymthe.2017.03.023
- Syed, Y. Y. (2016). Eteplirsen: first global approval. *Drugs* 76, 1699–1704. doi: 10.1007/s40265-016-0657-1
- Wilton, S. D., Lim, L., Dye, D., and Laing, N. (1997). Bandstab: a PCR-based alternative to cloning PCR products. *BioTechniques* 22, 642–645. doi: 10.2144/97224bm14
- Wilton, S. D., Fall, A. M., Harding, P. L., McClorey, G., Coleman, C., and Fletcher, S. (2007). Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript. *Mol. Ther.* 15, 1288–1296. doi: 10.1038/sj.mt.6300095
- 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

Conflict of Interest: SW, KG, and SF are named inventors on exon skipping patents and as such are entitled to royalty and milestone payments as they arise.

The remaining 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 Zaw, Greer, Aung-Htut, Mitrant, Veedu, Fletcher and Wilton. 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.



Circular RNAs: Potential Regulators of Treatment Resistance in Human Cancers

Shivapriya Jeyaraman, Ezanee Azlina Mohamad Hanif, Nurul Syakima Ab Mutalib, Rahman Jamal and Nadiyah Abu*

UKM Medical Molecular Biology Institute (UMBI), UKM Medical Center, Kuala Lumpur, Malaysia

OPEN ACCESS

Edited by:

Stefano Duga,
Humanitas University, Italy

Reviewed by:

Marianna Aprile,
Italian National Research Council
(CNR), Italy
Mohammadreza Hajjari,
Shahid Chamran University of
Ahvaz, Iran

*Correspondence:

Nadiyah Abu
nadiyah.abu@ppukm.ukm.edu.my

Specialty section:

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

Received: 31 May 2019

Accepted: 16 December 2019

Published: 28 January 2020

Citation:

Jeyaraman S, Hanif EAM, Ab
Mutalib NS, Jamal R and Abu N (2020)
Circular RNAs: Potential
Regulators of Treatment
Resistance in Human Cancers.
Front. Genet. 10:1369.
doi: 10.3389/fgene.2019.01369

Circular RNAs (circRNAs) which were once considered as “junk” are now in the spotlight as a potential player in regulating human diseases, especially cancer. With the development of high throughput technologies in recent years, the full potential of circRNAs is being uncovered. CircRNAs possess some unique characteristics and advantageous properties that could benefit medical research and clinical applications. CircRNAs are stable with covalently closed loops that are resistant to ribonucleases, have disease stage-specific expressions and are selectively abundant in different types of tissues. Interestingly, the presence of circRNAs in different types of treatment resistance in human cancers was recently observed with the involvement of a few key pathways. The activation of certain pathways by circRNAs may give new insights to treatment resistance management. The potential usage of circRNAs from this aspect is very much in its infancy stage and has not been fully validated. This mini-review attempts to highlight the possible role of circRNAs as regulators of treatment resistance in human cancers based on its intersection molecules and cancer-related regulatory networks.

Keywords: non-coding RNAs, biomarker, RNA splicing, chemoresistance, radioresistance

INTRODUCTION

Increasing evidence has shown that circular RNAs (circRNAs) a form of non-coding RNA is involved in important biological processes and cellular functions (Kristensen et al., 2018). Though the concept of circRNAs first emerged in 1976, it became particularly interesting in recent years when it was found to have regulatory roles in cancer biology. During the early years of discovery, the occurrence of circRNAs was considered as post-transcriptional errors; however, with the advent of various biomedical technologies, the biogenesis and functions of circRNAs have been extensively explored (Wang et al., 2017). As of 2018, there were approximately 30000 circRNAs identified and the numbers are increasing (Si-Tu et al., 2019). There are four categories of circRNAs, exonic circRNA (ecircRNA), circular intronic RNA (ciRNA), exon-intron circRNAs, and intergenic circRNAs (Li Z. et al., 2015; Memczak et al., 2013; Zhang et al., 2013). CircRNAs have covalently closed loops without a free 3' or 5' ends (Lasda and Parker, 2014). It is postulated that the formation of circRNAs are *via* back-splicing where the downstream exons are spliced to upstream exons in reverse order in the primary transcript (Chen and Yang, 2015). Furthermore, several unique properties make circRNAs a promising entity in providing key insights into human diseases. Besides

being abundant both in normal and cancer cells, it was also found that circRNAs are specifically expressed at every stage of cell development (Li J. et al., 2015). It was further confirmed that different isoforms of circRNAs from the same gene are expressed differently in different cell types. In several types of cancers such as hepatocellular carcinoma and colorectal cancer, it was noted that the expression level of circRNAs varies according to TNM stage, presence of metastasis and size of tumor (Szabo and Salzman, 2016). Unlike linear RNAs, circRNAs are more stable and are not easily degraded by ribonucleases such as exonuclease or RNase R due to the unexposed 3' and 5' terminals (Wang et al., 2017). Moreover, most circRNAs have an average half-life of over 48 h compared to linear mRNA with an average half-life of 10 h, thus making it more available for both research and clinical purposes. In addition to its advantageous properties, studies have found that circRNAs are involved in several biological activities as competing endogenous RNA by sponging miRNAs (Lin ADF and Chen, 2018), RNA binding proteins (RBPs) (Wang et al., 2015) and translating peptides (Granados-Riveron and Aquino-Jarquín, 2016; Du et al., 2017). Of particular interest is the role of circRNAs as miRNA sponge in tumor pathogenesis, and there have been many publications related to this (Wang et al., 2015; Zhang et al., 2017; Kun-Peng et al., 2018). By serving as a miRNA sponge with many binding sites, circRNAs can regulate the expression of miRNA as a competitive inhibitor that suppresses the ability of the miRNA to bind to its target genes. This event can, in turn, increase the levels of the miRNA target causing dysregulation of gene expression and pathological effects on tumor environment (Huang et al., 2015; Palmieri et al., 2018; Zeng et al., 2018). Some of these potential miRNA targets have been reported to function as important regulators of various cellular processes including apoptosis, invasion, migration, and drug resistance in several cancers.

Recently, much evidence was published on the role of circRNAs in disease progression and activation of key pathways like EMT and Wnt (Shen et al., 2019; Wu et al., 2019). Cancers that are gaining popularity like gastric, hepatocellular, lung, and breast are being studied closely with the hope to target the specific circRNAs that are involved in the development of tumor (Shang et al., 2019). Accumulating data on the association between circRNA and tumorigenesis shows promising results. However, little is known about its role in cancer therapy resistance. As therapy resistance remains one of the major clinical hurdles in cancer management, this mini-review aims to explore the potential of circRNAs as a regulator of treatment resistance. We reviewed recent relevant publications focusing on circRNAs in treatment resistance, particularly regarding drug therapy and radiotherapy. We also looked at studies at the network level to explain the relationship of circRNAs with the potential targets and pathways that could influence disease progression.

CircRNA Effects Radiotherapy Receptivity *via* WNT Pathway

Non-coding RNAs have been linked to tumorigenesis, metastasis, and the development of resistance to treatment (Gong et al., 2014).

Radiation therapy is one of the main treatment solutions for esophageal squamous cell carcinoma (ESCC) patients, especially with unresectable esophageal cancer. Unfortunately, radioresistance has been one of the reasons for failed treatments and local tumor recurrence in ESCC (Chen et al., 2017). In a study conducted by Su et al, hsa_circ_001059 and hsa_circ_000167 levels were shown to be dysregulated in radioresistant ESCC cell line as compared to the parental cell line (Su et al., 2016). The analysis showed that circRNA_001059 could sponge to multiple miRNAs including miR-30c-1, miR-30c-2, miR-122, miR-139-3p, miR-339-5p, and miR-1912. In support of this finding, miR-30 and miR-122 were found to be dysregulated in chemoresistant prostate cancer and miR-30 in radiosensitive leukemia cells (Ni et al., 2017; Liamina et al., 2017). These dysregulated circRNAs were mapped to their target genes and were found to be mainly involved in the Wnt signaling pathway and the PI3K/Akt pathway. The crosstalk mechanisms between the Wnt pathway and other non-coding RNAs such as the long non-coding RNA (lncRNAs) have been observed in cancer pathogenesis (Yang et al., 2018) including in breast cancer (Liu et al., 2016; Koval and Katanaev, 2018) and lung cancer (Wan et al., 2016; Zeng et al., 2017). Wnt signaling pathway was shown to be an important regulator in cancer cell responsiveness to radiotherapy (Takebe et al., 2015). The activation of the pathway increases the rate of DNA double-strand break repair and is also involved in breast cancer and colorectal cancer chemotherapy resistance (Flanagan et al., 2015; Pohl et al., 2017). Moreover, Li et al. have found that the expression of circITCH was low in ESCC tissues. It is possible that circITCH could influence the radioresistance of ESCC by inhibiting the target gene *via* suppression of the Wnt/ β signaling pathway as in a previous radioresistance study (Li J. et al., 2015). Besides ESCC, circITCH has also been associated with other cancers as well such as hepatocellular, bladder, lung, and colorectal where it acts as miRNA sponges particularly to miR-7, miR-17 and miR-214 (Li J. et al., 2015; Guo et al., 2018; Shang et al., 2019). Due to the sponging effect, the expression of the protein-coding gene ITCH increases as well (Fang, 2018). It was discovered that the ITCH gene is an important player in tumor formation and responsiveness to chemotherapy (Li J. et al., 2015). Phosphorylation of the mediator protein Dvl2 is required for activation of Wnt signaling and previous research have found that ITCH gene can disrupt Dvl2 phosphorylation and subsequently lead to the inhibition of the signaling pathway (Li J. et al., 2015). It has also been shown that parallel circITCH expression level increases with *ITCH* mRNA level in colorectal cancer (Huang et al., 2015). Sponging effects of circITCH to miR-7 and miR-20a lead to the downregulation of *ITCH* *via* bindings of these miRNAs to the 3'UTR of *ITCH*, ultimately attenuating the proliferative rate of CRC cell lines (Huang et al., 2015).

Multiple CircRNAs Regulate Chemoresistance *via* MAPK Pathway

Chemoresistance poses as one of the biggest challenges for cancer patients receiving neoadjuvant, adjuvant or palliative chemotherapy. Similar to radioresistance, the possibility of circRNAs as chemoresistance regulators were recently

explored. Evidence on this is still limited but seems promising. It is well known that the Mitogen-Activated Protein Kinase (MAPK) signaling pathway is dysregulated in many cancers such as breast, pancreatic, colon, and leukemia (Li et al., 2014). MAPK pathway is important in controlling cellular activities including but not limited to proliferation, differentiation, apoptosis, survival, inflammation, and influence gene expression. This pathway has also shown potential in regulating treatment resistance involving non-coding RNAs in non-small cell lung cancer and malignant melanoma (Li et al., 2014). A study on EGFR therapy in colorectal cancer showed that global circRNAs were significantly downregulated in mutant Kras cell line, particularly circFAT1 and circARHGAP5 (Dou et al., 2016). Kras functions in activating the MAP3K tier which in turn activates the ERK1/2 cascade while EGFR is an important growth factor which can act as an oncogene by hyper-activating the signaling pathway (Hymowitz and Malek, 2018). In line with this observation, some studies showed that overexpression of EGFR can promote radio and chemotherapy resistance by activating the second and third tiers of the MAPK pathway (Pozzi et al., 2016). The MEK/ERK cascade was shown to be exclusively deactivated when EGFR expression is inhibited. These findings suggest that dysregulation in any tier of this cascade affects the levels of circRNA and may contribute to treatment resistance. However, it is difficult to conclude from this study that the lower levels of circRNA is a regulator of the oncogenic factors or it directly protects the cancer cells from chemoradiation treatment. Thus, a further functional investigation is required to define the role of circRNAs in this instance. Dysregulation of the MAPK pathway was also found in lung adenocarcinoma (LUAD) treatment resistance whereby the increased level of circRNA CCDC66 contributes to the overexpression of EGFR (Joseph et al., 2018). CircCCDC66 acts as a sponge for multiple miRNAs and therefore can have a wider effect on its targeted genes (Weng et al., 2017). Besides the MAPK signaling pathway, HGF-MET pathway was also reported to be involved. The role of HGF-MET pathway in EGFR targeted therapy resistance was first reported in 2007 and since then many studies have shown that increase in its activity causes an elevation in RAS-RAF-MEK-ERK or the MAPK pathways (Ko et al., 2017). Thus, current studies are trying to approach the dual inhibition of MET and EGFR in managing resistance in LUAD patients (Ko et al., 2017). Another notable circRNA that plays a role in MAPK pathway is ciRS-7, which is known to sponge to miR-7 and elevates the expression of target genes (Hansen et al., 2013a; Weng et al., 2017). ciRS-7 is known as a super sponge for miR-7 with more than 70 binding sites and act as an expression inhibitor (Hansen et al., 2013a; Hansen et al., 2013b). MiR-7 which is a well-researched miRNA is known to be a player in the progression of many types of cancers in human and directly targets key oncogenes such as EFGR (Kefas et al., 2008), c-KIT in brain cancer (Tamim et al., 2014), PAX6 in colorectal cancer (Needham et al., 2014) and AKT in hepatocellular carcinoma (Fang et al., 2012). Thus, lower miR-7 level due to the sponging effect of ciRS-7 causes an increase in target genes, in this instance

EGFR and causes an imbalance to the signaling pathway (Fang et al., 2012).

Several Reported CircRNAs Contribute to Treatment Resistance *via* PI3K/AKT Pathway

Aberrant fusion-circRNA (f-circRNAs), a new form of circRNA generated by chromosomal translocation, was recently linked to treatment resistance in leukemic cells. Lehmann et al. found that f-circRNAs may potentially contribute to the development of resistance to therapy by protecting the leukemic cells from arsenic trioxide (Lehmann et al., 2001). Arsenic oxide is used to treat newly diagnosed and relapsed leukemic patients where it is known to induce apoptotic and cytotoxic effects in blast cells (Dong et al., 2015). Similarly, another study found that by knocking out the expression of f-circM9, leukemic cells can resume apoptosis and resistance toward Ara-C can be reversed. The team has proved that this type of f-circRNA can contribute to tumorigenesis and resistance to therapy both *in vitro* and *in vivo*. Besides the MAPK pathway, the PI3K/AKT signaling pathway was triggered simultaneously by the presence of f-circM9. Similar to MAPK, the PI3K/AKT pathway is another common pathway often associated with cancer. This complex cascade influences cell apoptosis, cell cycle, DNA repair, glucose metabolism, and cell transformation. Mainly it is known to play a critical role in transducing signal between oncogenes and cellular functions (Gao et al., 2017). Inhibition of these two pathways caused drug induced apoptosis, hence the high presence of f-circRNAs in leukemic cells will have a negative effect on treatment response. This could be an indicator of possible treatment failure for leukemic patients (Nanba and Toyooka, 2008).

The hsa_circ_0006528 was found to play a vital role in mediating chemoresistance in breast cancer (Carrle and Bielack, 2006). It was observed that higher expression of circ_0006528 is significantly associated with adriamycin (ADM)-resistance in breast cancers (Miao et al., 2017). Circ_0006528 is derived from exons 2 to 5 of the PRELID gene and is exclusively related to miR-7-5p. Previously, it was reported that low levels of miR-7 contribute to the resistance to chemotherapy, and the findings of this study further supported this (Callaghan et al., 2014). Upregulated circRNAs triggered the MAPK signaling pathway and were shown to regulate the cancer cells response to ADM treatment (Nanba and Toyooka, 2008). PI3K/AKT could also play a part in mediating the drug resistance *via* regulation of hsa-miR-130b (Miao et al., 2017). Current evidence showed that up-regulation of miRNA-130b mediates chemoresistance and increases the proliferation of breast cancer cells. It also reduces the expression of PTEN target which is a common tumor suppressor gene responsible for cell growth and apoptosis (Di Cristofano and Pandolfi, 2000). It is worth noting that the MAPK and PI3K/AKT pathways are complex and multilayered in nature, thus we could see that different types and levels of circRNAs trigger these pathways with opposing effects. In support of this, activation of AKT pathway was shown to have a better prognosis in luminal breast cancer in contrast to

its detrimental effect on the above-mentioned study involving circ_0006528 (Sonnenblick et al., 2019).

PI3K/AKT signaling pathway was also predicted to regulate response to 5-Fluorouracil (5-FU) in colorectal cancer patients. 5-FU, a pyrimidine antagonist, is widely used in managing advanced stage colorectal chemotherapy patients but has a recurrence rate of more than 50% (Yu et al., 2009). Upregulation of hsa_circ_0000504 was observed in 5-FU resistance colorectal cancer, where the circRNA interacts with miR-485-5p which targets STAT3 gene (Xiong et al., 2017). It was demonstrated that by silencing the STAT3 gene, clonogenic survival of cancer cells was significantly reduced (Spitzner et al., 2014). From the KEGG pathway analysis, miR-485-5p was predicted to activate the AKT signaling pathway which is associated with chemoradiotherapy treatment response (Xiong et al., 2017). In another recent study, it was reported that hsa_circ_0004015 was highly expressed in non-small cell lung cancer (NSCLC) which significantly contributes to disease progression and EGFR-Tyrosine Kinase Inhibitors (TKI) resistance (Zhou et al., 2019). TKIs are used as first line drug for NSCLC patients with EGFR mutations. Subsequently, *via* bioinformatics prediction, miR-1183 and target gene PDPK1 were further studied. It is known that PDPK1 a crucial component of the Akt-mTOR pathway, can intervene in cell proliferation and apoptosis (Li et al., 2018). In vitro results concluded that knockdown of hsa_circ_0004015 significantly sensitized NSCLC resistant cell line to TKIs (Zhou et al., 2019). Similarly, another group of investigators also proved the relationship of circRNA in resisting EGF-TKIs treatment by activating AKT/mTOR in EGFR-mutant NSCLC (Cheng et al., 2015).

Interestingly this pathway was also seen activated in cisplatin resistance in gastric cancer (GC) (Huang et al., 2019). Cisplatin is among the main chemotherapy drugs given for GC patients. In this study circRNA AKT3 upregulates PIK3R1 which, in turn increases the treatment resistance *via* miR-198 suppression and activation of PI3K/AKT signaling pathway. Similarly, cisplatin resistance was also found in human thyroid carcinoma cells in which circRNA EIF6 was studied (Liu et al., 2018). By bioinformatics analysis, miR-144-3p was found to regulate the expression of circRNA EIF6 and Transforming growth factor- α (TGF- α). This growth factor could promote tumor growth *via* various signaling pathways, such as PI3K/AKT and MEK/VEGF. Besides, a transcription factor known as Forkhead box O (FOXO) was reported to interact with the PI3K/AKT pathway as well and is mostly targeted for cancer treatment therapy (Farhan et al., 2017). In support of this, a group of researchers has found that overexpression of circRNA 0067835 promotes the expression of FOXO3a gene which causes detrimental effect to temporal lobe epilepsy (TLE) (Guo et al., 2018). Researchers discovered that circRNA-0067835 acts as sponge to miR-155 which has binding sites to FOXO3a gene. From these findings, aberrant activation of this pathway not only contributes to the progression of disease but also the cell receptivity to chemotherapy and radiotherapy treatment (Toulany and Rodemann, 2015). This multiplex pathway consisting of many regulatory molecules are seen hyper activated in different types of

cancers. Looking in a different lens, this could open a new path for research and therapy management especially in terms of chemoradioresistance.

CircRNA Influences Uptake of Drugs *via* Drug Transporter Pathway

A group of researchers has found that downregulation of circPVT1 by small inhibiting RNA (siRNA) might weaken resistance to doxorubicin and cisplatin treatment in osteosarcoma (Kun-Peng et al., 2018). Doxorubicin or ADM is an essential component in the treatment of osteosarcoma and cisplatin is the second most commonly used drug given in combination with doxorubicin (Carrle and Bielack, 2006). This study found that circPVT1 was significantly higher in the chemoresistance group as compared to the normal control group and knockdown of circPVT1 could partly reverse the resistance of doxorubicin and cisplatin (Kun-Peng et al., 2018). Overexpression of circPVT1 is also present in chemoresistant lung cancer as compared to chemosensitive patients. Knockdown of circPVT1 decreased the expression level of a well-known drug transporter gene, *ABCB1*. From previous discoveries, it was found that the gene promotes chemoresistance *via* removing the intracellular drugs by P-glycoprotein (P-GP) an ATP-dependent efflux pump that may have to metabolize the drugs in cancer cells (Callaghan et al., 2014).

CircRNA Interrupt Chemotherapy Acceptance by Regulating VEGF Pathway

CircRNA has also been shown to be a potential regulator for gemcitabine resistance in pancreatic cancer (Shao et al., 2018). Gemcitabine, a pro drug used as one of the key drugs in treating pancreatic cancer is able to kill cells with active DNA synthesis and block cell cycle progression at G1/S phase (Plunkett et al., 1991; Plunkett et al., 1995). Researchers found that when chr14:101402109-101464448+ and chr4:52729603-52780244+ were silenced in gemcitabine resistant cell lines, the sensitivity towards the drug was enhanced (Shao et al., 2018). Likewise, when the circRNAs were overexpressed, the gemcitabine resistance towards the cell lines was further intensified. By further exploring the binding site for miRNA, it was shown that miR-145-5p was down-regulated in the tumor tissues and plasma of chemotherapy resistant patients (Shao et al., 2018). MiR-145 is known to be associated with gemcitabine resistance in pancreatic cancer and it is believed that ErbB and VEGF signaling biological pathways were activated (Skrypek et al., 2015; Zhou et al., 2016). The ErbB signaling pathway was reported to be involved in gemcitabine resistance in pancreatic cancer while the VEGF pathway was implicated in the progression of the disease which could also cause drug resistance (Shao et al., 2018). Not much literature is available with regards to VEGF and non-coding RNA in treatment resistance however it is worth to note that there is a link between non-coding RNAs particularly circRNAs, in tumor angiogenesis involving VEGF pathway (Abhinand et al., 2016). As pancreatic cancer is known for its poor prognosis, any avenue that could predict treatment response can ease the burden of

unnecessary treatment hassle for patients. Potentially, the circRNAs can be used as biomarkers for planning the treatment of patients' in a more personalized approach.

CircRNA Causes Treatment Failure via Hypoxia-Inducible Factor-1 Regulatory Pathway

Hypoxia-inducible factor (HIF) pathway is increasingly studied due to its therapeutic potential in disease management. A group of circRNA researchers has joined the bandwagon in exploring the relationship between this pathway and cisplatin resistance in bladder cancer (Su et al., 2019). HIF 1 protein plays an integral part in angiogenesis and responding directly to hypoxia. However, due to this property, the enhancement of this gene can also allow the proliferation of cancer cells. Thus inhibition of this pathway potentially could prevent the metastasis of cancer cells (Ziello et al., 2007). In this study, circELP3 was elevated in hypoxia condition, was found to contribute to treatment resistance in bladder cancer cells (Su et al., 2019). Elevation of circELP3 was also in accordance with the severity of disease in human cancer patients. Stimulatingly, the hypoxia elevated circELP3 was independent to ELP3. In theory, it is expected that the hypoxia-related circRNA is correlated to this angiogenesis pathway however the findings in this study were contradictory. This unexpected finding gives a new perspective

to the function of circRNA in regulating cisplatin resistance and further test should be carried out to confirm the independence. Another group of researchers has identified circ_103470 and circ_101102 was dysregulated in endometriosis and HIF-1 may be associated with the pathogenesis (Zhang et al., 2018). However, this was only a profiling study thus the space to explore remains large. Compared to other mentioned pathways, HIF-1 has just commenced. As angiogenesis is one of the key target approaches in cancer study, it is worth to explore this area of the pathway with a combination of circRNA.

FUTURE DIRECTIONS AND CONCLUSIONS

From the evidence compiled as shown in **Table 1**, it was apparent that there are still many unexplored circRNAs that might be involved in cancer treatment resistance. At this juncture it is difficult to conclude the impact of circRNAs in chemoresistance and radioresistance; however, they have already exhibited enormous potential for further exploration. As treatment resistance in newly diagnosed patients or in recurrence state remains a major challenge, circRNAs could be an ideal marker to predict for treatment failure or even in reversing the resistance. Although the evidence is rather limited and incomplete, it is clear that there are certain interactions between circRNAs and

TABLE 1 | List of CircRNAs involved in treatment resistance in human cancers.

Type of cancer	Sequence name	Expression	Intersection molecules and/pathway	Sample type	Resistant to (chemotherapy/radiotherapy)	References
Esophageal squamous cell cancer	circRNA_000167 circRNA_001059	Down	Wnt/ β signaling pathway	Cell line	Radioresistant	(Su et al., 2016)
Colorectal cancer	circFAT1	Down	EGFR/MEK/ERK signaling pathway	Cell line	Chemoradioresistant	(Dou et al., 2016)
	circARHGAP5	Up	miR-7/EGFR/RAF1/MAPK	Tissue	Chemoresistant	(Weng et al., 2017)
	hsa_circ_0000504	Up	miR-485-5p/AKT pathway/STAT3	Cell line	Chemoresistant (5-Fluorouracil drug)	(Xiong et al., 2017)
Lung adenocarcinoma	CCDC66	Up	HGF/c-Met pathway and MAPK	Tissue and cell line	Chemoresistant (Gefitinib drug)	(Weng et al., 2017)
Leukemia	f-circM9	Up	PI3K/AKT1	Cell line and xenograft model	Chemoresistant (Ara-C drug and arsenic oxide)	(Lehmann et al., 2001)
Osteosarcoma	circPVT1	Up	Drug transporter pathway ABCB1 gene	Tissue	Chemoresistant (Doxorubicin, cisplatin drug)	(Zeng et al., 2018)
Pancreatic cancer	chr14:101402109-101464448+ and chr4:52729603-52780244+	Up	mir-145-5p/ErbB and VEGF	Tissue and cell line	Chemoresistant (Gemcitabine drug)	(Shao et al., 2018)
Breast cancer	Circ_0006528	Up	miR-7-5p/MAPK and PI3K signaling pathway	Tissue and cell line	Chemoresistant (Doxorubicin drug)	(Nanba and Toyooka, 2008; Miao et al., 2017)
Non-small cell lung cancer (NSCLC)	hsa_circ_0004015	Up	miR-1183/PDPK1/Akt-mTOR pathway	Tissue and cell line	Chemoresistant (EGFR-TKIs drug)	(Zhou et al., 2019)
Gastric cancer	circRNA AKT3	Up	miR-198/PI3K/AKT	Tissue and cell line	Chemoresistant (cisplatin drug)	(Huang et al., 2019)
Thyroid cancer	circRNA EIF6	Up	miR-144-3p/TGF- α /PI3K/AKT	Tissue and cell line	Chemoresistant (cisplatin drug)	(Liu et al., 2018)
Temporal lobe epilepsy	circRNA 0067835	Up	miR-155/FOXO3a/PI3K/AKT	Tissue and cell line	Antiepileptic drugs (AEDs)	(Guo et al., 2018)
Bladder cancer	circELP3	Up	ELP3/HIF-1	Tissue and cell line	Chemoresistant (cisplatin drug)	(Su et al., 2019)

miRNAs which act as contributing factors in disease progression and therapy resistance. By identifying the interactions between circRNA/miRNA/gene/pathway, more opportunities on targeted therapies can be created as shown in **Figure 1**. In this review, many circRNAs were shown to activate the MAPK pathway and the PI3K signaling pathway. By studying the complex interaction between circRNAs, targeted gene, and the pathways involved, we may be able to detect responses to treatment, early resistance or recurrence in real time. However, analyzing and targeting specific circRNAs persist as a challenge to researchers. Besides than optimizing the feature of circRNA as competing endogenous RNA, limited computational approaches, such as microarray and RNA sequencing, are used for

differential expression analysis. Subsequently, inhibition and overexpression of circRNAs are carried out to functionally characterize the targeted circRNAs based on the bioinformatics analysis. These pipelines are rather complex and require tedious preparation for reliable results and remain as a great challenge for novice researchers. It is apparent that there are still gaps to be filled in terms of predicting the target genes and the functional mechanisms in modulating diseases. For validation of the high throughput sequencing, current approach mostly focuses on proving the presence of circRNA by carrying out qPCR and Sanger sequencing. To date, only a few circRNAs have been characterized and validated fully. The unpredictable character of

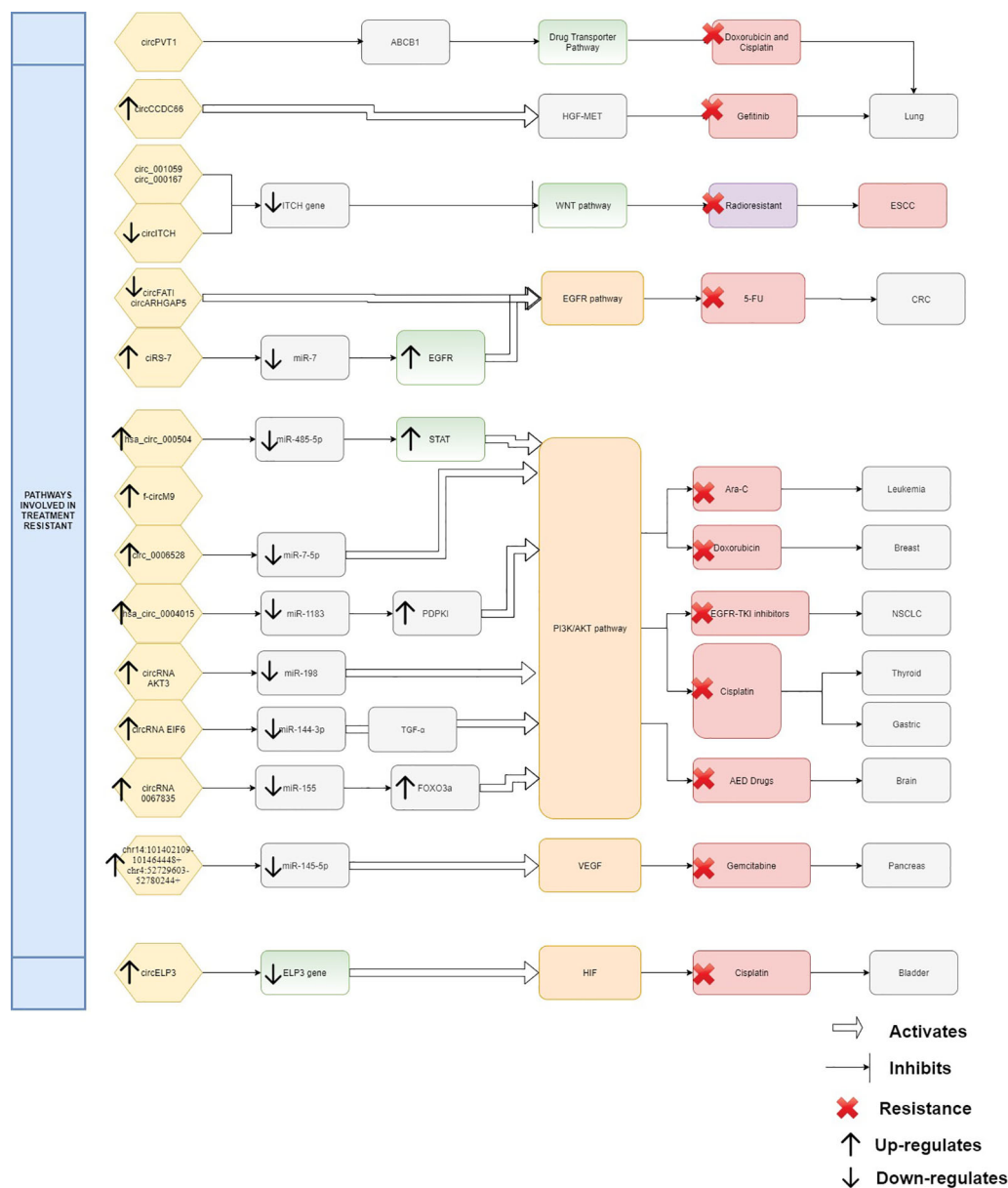


FIGURE 1 | Interaction between circRNAs, targeted gene, and pathways.

circRNA as a pro-oncogene and anti-tumor molecule is thought-provoking and a limiting factor for immediate clinical application. It is more challenging when the same circRNA exhibit multiple roles in different cancers. Thus, it is difficult to deduce and predict a single role for circRNA. Nonetheless, this shows the flexibility of circRNAs as a biomarker and when manipulated could contribute to prevention or delaying of disease progression. CircRNAs as a predictive marker and potential regulator of treatment resistance is still in the early phase of research and application, nevertheless, the future of circRNAs is certainly promising, and it will be likely that they will become part of the armamentarium to treat cancers in future.

REFERENCES

- Abhinand, C. S., Raju, R., Soumya, S. J., Arya, P. S., and Sudhakaran, P. R. (2016). VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. *J. Cell Commun. Signal.* 10 (4), 347–354. doi: 10.1007/s12079-016-0352-8
- Callaghan, R., Luk, F., and Bebawy, M. (2014). Inhibition of the multidrug resistance P-glycoprotein: Time for a change of strategy? *Drug Metab. Disposition* 42 (4), 623–631. doi: 10.1124/dmd.113.056176
- Carrle, D., and Bielack, S. S. (2006). Current strategies of chemotherapy in osteosarcoma. *Int. Orthopaedics* 30 (6), 445–451. doi: 10.1007/s00264-006-0192-x
- Chen, L. L., and Yang, L. (2015). Regulation of circRNA biogenesis. *RNA Biol.* 12 (4), 381–388. doi: 10.1080/15476286.2015.1020271
- Chen, G.-Z., Zhu, H.-C., Dai, W.-S., Zeng, X.-N., Luo, J.-H., and Sun, X.-C. (2017). The mechanisms of radioresistance in esophageal squamous cell carcinoma and current strategies in radiosensitivity. *J. Thorac. Dis.* 9 (3), 849–859. doi: 10.21037/jtd.2017.03.23
- Cheng, N., Cai, W., Ren, S., Li, X., Wang, Q., Pan, H., et al. (2015). Long non-coding RNA UCA1 induces non-T790M acquired resistance to EGFR-TKIs by activating the AKT/mTOR pathway in EGFR-mutant non-small cell lung cancer. *Oncotarget*. 6 (27), 23582–23593. doi: 10.18632/oncotarget.4361
- Di Cristofano, A., and Pandolfi, P. P. (2000). The multiple roles of PTEN in tumor suppression. *Cell*. 100 (4), 387–390. doi: 10.1016/s0092-8674(00)80674-1
- Dong, L., Deng, J., Sun, Z. M., Pan, A. P., Xiang, X. J., Zhang, L., et al. (2015). Interference with the β -catenin gene in gastric cancer induces changes to the miRNA expression profile. *Tumor Biol.* 36 (9), 6973–6983. doi: 10.1007/s13277-015-3415-1
- Dou, Y., Cha, D. J., Franklin, J. L., Higginbotham, J. N., Jeppesen, D. K., Weaver, A. M., et al. (2016). Circular RNAs are down-regulated in KRAS mutant colon cancer cells and can be transferred to exosomes. *Sci. Rep.* 6 (1), 37892–37903. doi: 10.1038/srep37982
- Du, W. W., Zhang, C., Yang, W., Yong, T., Awan, F. M., and Yang, B. B. (2017). Identifying and characterizing circRNA-protein interaction. *Theranostics*. 7 (17), 4183–4191. doi: 10.7150/thno.21299
- Fang, Y., Xue, J. L., Shen, Q., Chen, J., and Tian, L. (2012). MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/Akt pathway in hepatocellular carcinoma. *Hepatology* 55 (6), 1852–1862. doi: 10.1002/hep.25576
- Fang, Y. (2018). Circular RNAs as novel biomarkers with regulatory potency in human diseases. *Futur. Sci. OA* 4 (7), FSO314. doi: 10.4155/fsoa-2018-0036
- Farhan, M., Wang, H., Gaur, U., Little, P. J., Xu, J., and Zheng, W. (2017). FOXO signaling pathways as therapeutic targets in cancer. *Int. J. Biol. Sci.* 13 (7), 815–827. doi: 10.7150/ijbs.20052
- Flanagan, L., Lindner, A. U., de Chaumont, C., Kehoe, J., Fay, J., Bacon, O., et al. (2015). BCL2 protein signalling determines acute responses to neoadjuvant chemoradiotherapy in rectal cancer. *J. Mol. Med.* 93 (3), 315–326. doi: 10.1007/s00109-014-1221-7
- Gao, D., Zhang, X., Liu, B., Meng, D., Fang, K., Guo, Z., et al. (2017). Screening circular RNA related to chemotherapeutic resistance in breast cancer. *Epigenomics* 9 (9), 1175–1188. doi: 10.2217/epi-2017-0055
- Gong, Z., Lin, H. K., Yang, J., Li, J., Yang, L., Le, Y., et al. (2014). Novel insights into the role of microRNA in lung cancer resistance to treatment and targeted therapy. *Curr. Cancer Drug Targets*. 14 (3), 241–258. doi: 10.2174/1568009614666140305104845
- Granados-Riveron, J. T., and Aquino-Jarquín, G. (2016). The complexity of the translation ability of circRNAs. *Biochim. Biophys. Acta - Gene Regul. Mechanisms*. 1859 (10), 1245–1251. doi: 10.1016/j.bbagr.2016.07.009
- Gong, G. H., An, F. M., Wang, Y., Bian, M., Wang, D., and Wei, C. X. (2018). Comprehensive circular RNA profiling reveals the regulatory role of the CircRNA-0067835/miR-155 pathway in temporal lobe epilepsy. *Cell. Physiol. Biochem*. doi: 10.1159/000495589
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., and Kjems, J. (2013a). Natural RNA circles function as efficient microRNA sponges. *Nature* 495 (7441), 384–388. doi: 10.1038/nature11993
- Hansen, T. B., Kjems, J., and Damgaard, C. K. (2013b). Circular RNA and miR-7 in Cancer. *Cancer Res.* 73 (18), 5609–5612. doi: 10.1158/0008-5472.CAN-13-1568
- Huang, G., Zhu, H., Shi, Y., Wu, W., Cai, H., and Chen, X. (2015). Cir-ITCH plays an inhibitory role in colorectal cancer by regulating the Wnt/ β -Catenin Pathway. *PLoS One*. 10 (6), e0131225. doi: 10.1371/journal.pone.0131225
- Huang, X., Xu, Z., Li, Z., Zhang, Q., Wang, W., Li, B., et al. (2019). Circular RNA AKT3 upregulates PIK3R1 to enhance cisplatin resistance in gastric cancer via miR-198 suppression. *Mol. Cancer*. 18 (71), 1–20. doi: 10.1186/s12943-019-0969-3
- Hymowitz, S. G., and Malek, S. (2018). Targeting the MAPK Pathway in RAS Mutant Cancers. *Cold Spring Harb. Perspect. Med.* 8 (11), 1–16. doi: 10.1101/cshperspect.a031492
- Joseph, N. A., Chiou, S. H., Lung, Z., Yang, C. L., Lin, T. Y., Chang, H. W., et al. (2018). The role of HGF-MET pathway and CCDC66 circRNA expression in EGFR resistance and epithelial-to-mesenchymal transition of lung adenocarcinoma cells. *J. Hematol. Oncol.* 11 (1), 74–88. doi: 10.1186/s13045-018-0557-9
- Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M., et al. (2008). microRNA-7 inhibits the epidermal growth factor receptor and the akt pathway and is down-regulated in glioblastoma. *Cancer Res.* 68 (10), 3566–3572. doi: 10.1158/0008-5472.CAN-07-6639
- Ko, B., He, T., Gadgil, S., and Halmos, B. (2017). MET/HGF pathway activation as a paradigm of resistance to targeted therapies. *Ann. Transl. Med.* 5 (1), 4–4. doi: 10.21037/atm.2016.12.09
- Koval, A., and Katanaev, V. L. (2018). Dramatic dysbalancing of the Wnt pathway in breast cancers. *Sci. Rep.* 8 (1), 7329–7339. doi: 10.1038/s41598-018-25672-6
- Kristensen, L. S., Hansen, T. B., Venø, M. T., and Kjems, J. (2018). Circular RNAs in cancer: Opportunities and challenges in the field. *Oncogene* 37 (5), 555–565. doi: 10.1038/onc.2017.361
- Kun-Peng, Z., Xiao-Long, M., and Chun-Lin, Z. (2018). Overexpressed circPVT1, a potential new circular RNA biomarker, contributes to doxorubicin and cisplatin resistance of osteosarcoma cells by regulating ABCB1. *Int. J. Biol. Sci.* 14 (3), 321–330. doi: 10.7150/ijbs.24360
- Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA*. 20 (12), –. doi: 10.1261/rna.047126.114
- Lehmann, S., Bengtzen, S., Paul, A., Christensson, B., and Paul, C. (2001). Effects of arsenic trioxide (As₂O₃) on leukemic cells from patients with non-M3 acute

AUTHOR CONTRIBUTIONS

SJ and NA drafted the article. NA conceived the idea. NM, EH, and RJ provided feedback and critical input.

FUNDING

This manuscript was funded by the Fundamental Research Grant Scheme (FRGS/1/2017/SKK08/UKM/03/3) awarded by the Ministry of Higher Education Malaysia.

- myelogenous leukemia: studies of cytotoxicity, apoptosis and the pattern of resistance. *Eur. J. Haematol.* 66 (6), 357–364. doi: 10.1034/j.1600-0609.2001.066006357.x
- Li, R., Zhang, L., Jia, L., Duan, Y., Li, Y., Bao, L., et al. (2014). Long non-coding RNA BANCER promotes proliferation in malignant melanoma by regulating MAPK pathway activation. *PLoS One*. 9 (6), e100893. doi: 10.1371/journal.pone.0100893
- Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22 (3), 256–264. doi: 10.1038/nsmb.2959
- Li, J., Yang, J., Zhou, P., Le, Y., Zhou, C., Wang, S., et al. (2015). Circular RNAs in cancer: novel insights into origins, properties, functions and implications. *Am. J. Cancer Res.* 5 (2), 472–480.
- Li, F., Zhang, L., Li, W., Deng, J., Zheng, J., An, M., et al. (2015). Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/ β -catenin pathway. *Oncotarget* 6 (8), 6001–6013.
- Li, D., Mullinax, J. E., Aiken, T., Xin, H., Wiegand, G., Anderson, A., et al. (2018). Loss of PDPK1 abrogates resistance to gemcitabine in label-retaining pancreatic cancer cells. *BMC Cancer*. 18 (1), 772–787. doi: 10.1186/s12885-018-4690-1
- Liamina, D., Saenko, Y., Sibirnyj, W., Khokhlova, A., Saenko, V., Rastorgueva, E., et al. (2017). Radiation-induced changes of microRNA expression profiles in radiosensitive and radioresistant leukemia cell lines with different levels of chromosome abnormalities. *Cancers (Basel)*. 9 (10), 136–152. doi: 10.3390/cancers9100136
- Lin ADF, X., and Chen, Y. (2018). Identification of potentially functional CircRNA-miRNA-mRNA regulatory network in hepatocellular carcinoma by integrated microarray analysis. *Research* 24, 70–78. doi: 10.12659/MSMBR.909737
- Liu, H., Wang, G., Yang, L., Qu, J., Yang, Z., and Zhou, X. (2016). Knockdown of long non-coding RNA UCA1 Increases the tamoxifen sensitivity of breast cancer cells through inhibition of wnt/ β -catenin pathway. *PLoS One*. 11 (12), e0168406. doi: 10.1371/journal.pone.0168406
- Liu, F., Zhang, J., Qin, L., Yang, Z., Xiong, J., Liu, J., et al. (2018). Circular RNA EIF6 (Hsa_circ_0060060) sponges miR-144-3p to promote the cisplatin-resistance of human thyroid carcinoma cells by autophagy regulation. *Aging (Albany, NY)*. 10 (12), 3806–3820. doi: 10.18632/aging.101674
- 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 (7441), 333–338. doi: 10.1038/nature11928
- Miao, Y., Zheng, W., Li, N., Su, Z., Zhao, L., Zhou, H., et al. (2017). MicroRNA-130b targets PTEN to mediate drug resistance and proliferation of breast cancer cells via the PI3K/Akt signaling pathway. *Sci. Rep.* 7, 1–12. doi: 10.1038/srep41942
- Nanba, K., and Toyooka, S. (2008). The allelic distribution of a single nucleotide polymorphism in the PDCD5 gene locus of Japanese non-small cell lung cancer patients. *Mol. Med. Rep.* 1 (5), 667–671. doi: 10.3892/mmr.00000010
- Needham, M., White, R. B., Giles, K. M., Dunlop, S. A., and Thomas, M. G. (2014). Regulation of human PAX6 expression by miR-7. *Evol. Bioinforma.* 10 107–113. doi: 10.4137/EBO.S13739
- Ni, J., Bucci, J., Chang, L., Malouf, D., Graham, P., and Li, Y. (2017). Targeting microRNAs in prostate cancer radiotherapy. *Theranostics* 7 (13), 3243–3259. doi: 10.7150/thno.19934
- Palmieri, D., Langenberger, D., Dai, Y., Calin, G. A., and Dragomir, M. (2018). Circular RNAs in cancer – lessons learned from microRNAs. *Front. Oncol.* 8 (8). doi: 10.3389/fonc.2018.00179
- Plunkett, W., Grindley, G. B., and Plunkett, W. (1991). Action of 2',2'-difluorodeoxycytidine on DNA synthesis. *Cancer Res.* 51 (22), 6110–6117.
- Plunkett, W., Huang, P., Xu, Y. Z., Heinemann, V., Grunewald, R., and Gandhi, V. (1995). Gemcitabine: metabolism, mechanisms of action, and self-potential. *Semin. Oncol.* 22 (11), 3–10.
- Pohl, S. G., Brook, N., Agostino, M., Arfuso, F., Kumar, A. P., and Dharmarajan, A. (2017). Wnt signaling in triple-negative breast cancer. *Oncogenesis*. 6, e310–e322. doi: 10.1038/oncsis.2017
- Pozzi, C., Cuomo, A., Spadoni, I., Magni, E., Silvola, A., Conte, A., et al. (2016). The EGFR-specific antibody cetuximab combined with chemotherapy triggers immunogenic cell death. *Nat. Med.* 22 (6), 624–631. doi: 10.1038/nm.4078
- Shang, Q., Yang, Z., Jia, R., and Ge, S. (2019). The novel roles of circRNAs in human cancer. *Mol. Cancer* 18 (1), 1–10. doi: 10.1186/s12943-018-0934-6
- Shao, F., Huang, M., Meng, F., and Huang, Q. (2018). Circular RNA Signature Predicts Gemcitabine Resistance of Pancreatic Ductal Adenocarcinoma. *Front. Pharmacol.* 9, 1–11. doi: 10.3389/fphar.2018.00584
- Shen, B., Wang, Z., Li, Z., Song, H., and Ding, X. (2019). Circular RNAs: an emerging landscape in tumor metastasis. *Am. J. Cancer Res.* 9 (4), 630–643.
- Si-Tu, J., Cai, Y., Feng, T., Yang, D., Yuan, S., Yang, X., et al. (2019). Upregulated circular RNA circ-102004 that promotes cell proliferation in prostate cancer. *Int. J. Biol. Macromol.* 122, 1235–1243. doi: 10.1016/j.ijbiomac.2018.09.076
- Skrypek, N., Vasseur, R., Vincent, A., Duchêne, B., Van Seuningen, I., and Jonckheere, N. (2015). The oncogenic receptor ErbB2 modulates gemcitabine and irinotecan/SN-38 chemoresistance of human pancreatic cancer cells via hCNT1 transporter and multidrug-resistance associated protein MRP-2. *Oncotarget* 6 (13), 10853–10867. doi: 10.18632/oncotarget.3414
- Sonnenblick, A., Venet, D., Brohée, S., Pondé, N., and Sotiriou, C. (2019). pAKT pathway activation is associated with PIK3CA mutations and good prognosis in luminal breast cancer in contrast to p-mTOR pathway activation. *NPJ Breast Cancer*. 8 (7), 1–7. doi: 10.1038/s41523-019-0102-1
- Spitzner, M., Roesler, B., Bielfeld, C., Emons, G., Gaedcke, J., Wolff, H. A., et al. (2014). STAT3 inhibition sensitizes colorectal cancer to chemoradiotherapy in vitro and in vivo. *Int. J. Cancer*. 134 (4), 997–1007. doi: 10.1002/ijc.28429
- Su, H., Jin, X., Xie, C., Lin, F., Deng, X., Shen, L., et al. (2016). Profiling and bioinformatics analyses reveal differential circular RNA expression in radioresistant esophageal cancer cells. *J. Transl. Med.* 14 (1), 1–10. doi: 10.1186/s12967-016-0977-7
- Su, Y., Yang, W., Jiang, N., Shi, J., Wang, C., and Lin, T. (2019). Hypoxia-elevated circELP3 contributes to bladder cancer progression and cisplatin resistance. *Int. J. Biol. Sci.* 15 (2), 441–452. doi: 10.7150/ijbs.26826
- Szabo, L., and Salzman, J. (2016). Detecting circular RNAs: Bioinformatic and experimental challenges. *Nat. Rev. Genet.* 17, 679–692. doi: 10.1038/nrg.2016.114
- Takebe, N., Miele, L., Harris, P. J., Jeong, W., Bando, H., Kahn, M., et al. (2015). Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: Clinical update. *Nat. Rev. Clin. Oncol.* 12 (8), 445–464. doi: 10.1038/nrclinonc.2015.61
- Tamim, S., Vo, D. T., Uren, P. J., Qiao, M., Bindewald, E., Kasprzak, W. K., et al. (2014). Genomic analyses reveal broad impact of miR-137 on genes associated with malignant transformation and neuronal differentiation in glioblastoma cells. *PLoS One*. 9 (1), e85591. doi: 10.1371/journal.pone.0085591
- Toulany, M., and Rodemann, H. P. (2015). Phosphatidylinositol 3-kinase/Akt signaling as a key mediator of tumor cell responsiveness to radiation. *Semin. Cancer Biol.* 35, 180–190. doi: 10.1016/j.semcancer.2015.07.003
- Wan, L., Zhang, L., Fan, K., Cheng, Z. X., Sun, Q. C., and Wang, J. J. (2016). Circular RNA-ITCH suppresses lung cancer proliferation via inhibiting the Wnt/ β -catenin pathway. *BioMed. Res. Int.* 2016, 1–11. doi: 10.1155/2016/1579490
- Wang, X., Zhang, Y., Huang, L., Zhang, J., Pan, F., Li, B., et al. (2015). Decreased expression of hsa_circ_001988 in colorectal cancer and its clinical significances. *Int. J. Clin. Exp. Pathol.* 8 (12), 16020–16025.
- Wang, Y., Mo, Y., Gong, Z., Yang, X., Yang, M., Zhang, S., et al. (2017). Circular RNAs in human cancer. *Mol. Cancer*. 16 (25), 1–8. doi: 10.1186/s12943-017-0598-7
- Weng, W., Wei, Q., Toden, S., Yoshida, K., Nagasaka, T., Fujiwara, T., et al. (2017). Circular RNA ciRS-7 — A promising prognostic biomarker and a potential therapeutic target in colorectal cancer. *Clin. Cancer Res.* 23 (14), 3918–3928. doi: 10.1158/1078-0432.CCR-16-2541
- Wu, J., Qi, X., Liu, L., Hu, X., Liu, J., Yang, J., et al. (2019). Emerging epigenetic regulation of circular RNAs in human cancer. *Mol. Ther. - Nucleic Acids* 16, 589–596. doi: 10.1016/j.omtn.2019.04.011
- Xiong, W., Ai, Y. Q., Li, Y. F., Ye, Q., Chen, Z. T., Qin, J. Y., et al. (2017). Microarray analysis of circular RNA expression profile associated with 5-fluorouracil-based chemoradiation resistance in colorectal cancer cells. *BioMed. Res. Int.* 2017, 1–8. doi: 10.1155/2017/8421614
- Yang, G., Shen, T., Yi, X., Zhang, Z., Tang, C., Wang, L., et al. (2018). Crosstalk between long non-coding RNAs and Wnt/ β -catenin signalling in cancer. *J. Cell. Mol. Med.* 22 (4), 2062–2070.
- Yu, Y., Kanwar, S. S., Patel, B. B., Nautiyal, J., Sarkar, F. H., and Majumdar, A. P. N. (2009). Elimination of colon cancer stem-like cells by the combination of curcumin and FOLFOX. *Transl. Oncol.* 2 (4), 321–328. doi: 10.1593/tlo.09193

- Zeng, Y., Xu, Y., Shu, R., Sun, L., Tian, Y., Shi, C., et al. (2017). Altered expression profiles of circular RNA in colorectal cancer tissues from patients with lung metastasis. *Int. J. Mol. Med.* 40 (6), 1818–1828. doi: 10.3892/ijmm.2017.3189
- Zeng, K., Chen, X., Xu, M., Liu, X., Hu, X., Xu, T., et al. (2018). CircHIPK3 promotes colorectal cancer growth and metastasis by sponging miR-7 article. *Cell Death Dis.* 9 (4). doi: 10.1038/s41419-018-0454-8
- Zhang, Y., Zhang, X. O., Chen, T., Xiang, J. F., Yin, Q. F., Xing, Y. H., et al. (2013). Circular Intronic Long Noncoding RNAs. *Mol. Cell.* 51 (6), –. doi: 10.1016/j.molcel.2013.08.017
- Zhang, Y., Qi, X., Li, G., Li, J., Yu, J., Liu, H., et al. (2017). Circular RNAs signature predicts the early recurrence of stage III gastric cancer after radical surgery. *Oncotarget.* 8 (14), 22936–22943. doi: 10.18632/oncotarget.15288
- Zhang, M., Ren, C., Xiao, Y., Xia, X., and Fang, X. (2018). Expression profile analysis of circular RNAs in ovarian endometriosis by microarray and bioinformatics. *Med. Sci. Monit.* 24, 9240–9250. doi: 10.12659/MSM.913885
- Zhou, R., Curry, J. M., Roy, L. D., Grover, P., Haider, J., Moore, L. J., et al. (2016). A novel association of neuropilin-1 and MUC1 in pancreatic ductal adenocarcinoma: role in induction of VEGF signaling and angiogenesis. *Oncogene.* 35 (43), 5608–5618. doi: 10.1038/ncr.2015.516
- Zhou, Y., Zheng, X., Xu, B., Chen, L., Wang, Q., Deng, H., et al. (2019). Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem. Biophys. Res. Commun.* 508 (2), 527–535. doi: 10.1016/j.bbrc.2018.11.157
- Ziello, J. E., Jovin, I. S., and Huang, Y. (2007). Hypoxia-Inducible Factor (HIF)-1 regulatory pathway and its potential for therapeutic intervention in malignancy and ischemia. *Yale J. Biol. Med.* 80 (2), 51–60.

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

Copyright © 2020 Jeyaraman, Hanif, Ab Mutalib, Jamal and Abu. 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.



Expression Changes Confirm Genomic Variants Predicted to Result in Allele-Specific, Alternative mRNA Splicing

Eliseos J. Mucaki¹, Ben C. Shirley² and Peter K. Rogan^{1,2,3,4*}

¹ Department of Biochemistry, University of Western Ontario, London, ON, Canada, ² CytoGnomix, London, ON, Canada, ³ Department of Oncology University of Western Ontario, London, ON, Canada, ⁴ Department of Computer Science, University of Western Ontario, London, ON, Canada

OPEN ACCESS

Edited by:

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

Reviewed by:

Stefan Stamm,
University of Kentucky,
United States
Lucie Grodecká, Center of
Cardiovascular and Transplant
Surgery, Czechia

*Correspondence:

Peter K. Rogan
progan@uwo.ca

Specialty section:

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

Received: 13 September 2019

Accepted: 30 January 2020

Published: 05 March 2020

Citation:

Mucaki EJ, Shirley BC and
Rogan PK (2020) Expression
Changes Confirm Genomic
Variants Predicted to
Result in Allele-Specific,
Alternative mRNA Splicing.
Front. Genet. 11:109.
doi: 10.3389/fgene.2020.00109

Splice isoform structure and abundance can be affected by either noncoding or masquerading coding variants that alter the structure or abundance of transcripts. When these variants are common in the population, these nonconstitutive transcripts are sufficiently frequent so as to resemble naturally occurring, alternative mRNA splicing. Prediction of the effects of such variants has been shown to be accurate using information theory-based methods. Single nucleotide polymorphisms (SNPs) predicted to significantly alter natural and/or cryptic splice site strength were shown to affect gene expression. Splicing changes for known SNP genotypes were confirmed in HapMap lymphoblastoid cell lines with gene expression microarrays and custom designed q-RT-PCR or TaqMan assays. The majority of these SNPs (15 of 22) as well as an independent set of 24 variants were then subjected to RNAseq analysis using the ValidSpliceMut web beacon (<http://validsplicemut.cytognomix.com>), which is based on data from the Cancer Genome Atlas and International Cancer Genome Consortium. SNPs from different genes analyzed with gene expression microarray and q-RT-PCR exhibited significant changes in affected splice site use. Thirteen SNPs directly affected exon inclusion and 10 altered cryptic site use. Homozygous SNP genotypes resulting in stronger splice sites exhibited higher levels of processed mRNA than alleles associated with weaker sites. Four SNPs exhibited variable expression among individuals with the same genotypes, masking statistically significant expression differences between alleles. Genome-wide information theory and expression analyses (RNAseq) in tumor exomes and genomes confirmed splicing effects for 7 of the HapMap SNP and 14 SNPs identified from tumor genomes. q-RT-PCR resolved rare splice isoforms with read abundance too low for statistical significance in ValidSpliceMut. Nevertheless, the web-beacon provides evidence of unanticipated splicing outcomes, for example, intron retention due to compromised recognition of constitutive splice sites. Thus, ValidSpliceMut and q-RT-PCR represent complementary resources for identification of allele-specific, alternative splicing.

Keywords: allele-specific gene expression, mRNA splicing, single nucleotide polymorphism, mutation, cryptic splicing, intron retention, alternative splicing, information theory

INTRODUCTION

Accurate and comprehensive methods are needed for predicting impact of noncoding mutations, in particular, mRNA splicing defects, which are prevalent in genetic disease (Krawczak et al., 1992; Teraoka et al., 1999; Ars et al., 2003; Spielmann and Mundlos, 2016; Gloss and Dinger, 2018). This class of mutations may account for as much as 62% of point mutations (López-Bigas et al., 2005). Large transcriptome studies have suggested that a large fraction of genome-wide association studies (GWAS) signals for disease and complex traits are due to single nucleotide polymorphisms (SNPs) affecting mRNA splicing (Park et al., 2018). ValidSpliceMut (Shirley et al., 2019) presents evidence of altered splicing (Viner et al., 2014; Dorman et al., 2014) for 309,848 validated genome splice-variant predictions (Shirley et al., 2013). The majority of mutations were associated with exon skipping, cryptic site use, or intron retention, and in these cases ValidSpliceMut assigns a molecular phenotype classification to all variants as either aberrant, likely aberrant or inducing alternative isoforms.

While allele-specific alternative splicing can predispose for disease susceptibility (Park et al., 2018), these genetic variations also are associated with common phenotypic variability in populations (Hull et al., 2007). Soemedi et al. (2017) determined that 10% of a set of published disease-causing exonic mutations ($N = 4,964$) altered splicing. Their analysis of a control set of exonic SNPs common among those without disease phenotypes revealed a smaller proportion (3%) that altered splicing ($N = 228$). However, we recently showed that splice-altering, common SNPs are considerably more abundant in tumor genomes in the ValidSpliceMut web-beacon (<http://validsplicemut.cytogenomix.com>; Shirley et al., 2019). Variants with higher germline population frequencies which impact splicing are less likely than rare mutations with direct splicing effects to be involved in Mendelian diseases or cancer. The present study analyzes predicted splice-altering polymorphic variants in genotyped lymphoblastoid cell lines by q-RT-PCR, expression microarrays of samples of known SNP genotypes, and high throughput expression data corresponding to sequenced tumor exomes and genomes. The relatively high frequencies of these variants enable comparisons of expressed transcripts in multiple individuals and genotypes. Effects of these SNPs are confirmed by multiple methods, although the supporting evidence from these distinct approaches is often complementary, rather than entirely concordant.

An estimated 90% to 95% of all multiexon genes are alternatively spliced (Pan et al., 2008; Wang et al., 2008; Baralle and Giudice, 2017). The selection of splicing signals involves exon and intron sequences, complementarity with snRNAs, RNA secondary structure, and competition between spliceosomal recognition sites (Moore and Sharp, 1993; Berget, 1995; Park et al., 2018). U1 snRNP interacts with the donor (or 5') splice site (Zhuang and Weiner, 1986; Séraphin et al., 1988) and U2 (and U6) snRNP with the acceptor and branch sites of pre-mRNA (Parker et al., 1987; Wu and Manley, 1989). The majority of human splice donors (5') and acceptors (3') base pair

with the U1 and U2 RNAs in spliceosomes, but are generally not precisely complementary to these sequences (Rogan et al., 2003). Additional exonic and intronic cis-regulatory elements can promote or suppress splice site recognition through recruitment of trans-acting splicing factors. SR proteins are positive trans-acting splicing factors which contain RNA-recognition motifs (RRM) and a carboxy-terminal domain enriched in Arg/Ser dipeptides (SR domain; Birney et al., 1993). Binding of RRM in pre-mRNA enhances exon recognition by promoting interactions with spliceosomal and other proteins (Fu and Maniatis, 1992). SR proteins function in splice site communication by forming an intron bridge needed for exon recognition (Zuo and Maniatis, 1996). Factors that negatively impact splicing include heterogeneous nuclear ribonucleoproteins (hnRNPs; Martinez-Contreras et al., 2007).

Splicing mutations affect normal exon recognition by altering the strengths of natural donor or acceptor sites and proximate cryptic sites, either independently or simultaneously. Weakened splice sites reduce of kinetics of mRNA processing, leading to an overall decrease in full length transcripts, increased exon skipping, cryptic splice site activation within exons or within adjacent introns, intron retention, and inclusion of cryptic, pseudo-exons (Talerico and Berget, 1990; Carothers et al., 1993; Buratti et al., 2006; Park et al., 2018). The kinetics of splicing at weaker cryptic sites is also slower than at natural sites (Domenjoud et al., 1993). Mutations strengthen cryptic sites either by increasing resemblance to “consensus sequences” (Nelson and Green, 1990) or by modulating the levels of SR proteins contributing to splice site recognition (Mayeda and Krainer, 1992; Cáceres et al., 1994). Mutations affecting splicing regulatory elements (Dietz et al., 1993; Richard and Beckmann, 1995) disrupt trans-acting SR protein interactions (Staknis and Reed, 1994) with distinct exonic and intronic cis-regulatory elements (Black, 2003).

Information theory-based (IT-based) models of donor and acceptor mRNA splice sites reveal the effects of changes in strengths of individual sites (termed R_i ; Rogan et al., 1998; Rogan et al., 2003). This facilitates prediction of phenotypic severity (Rogan and Schneider, 1995; von Kodolitsch et al., 1999; von Kodolitsch et al., 2006). The effects of splicing mutations can be predicted *in silico* by information theory (Rogan and Schneider, 1995; Kannabiran et al., 1998; Rogan et al., 1998; Svojanovsky et al., 2000; Rogan et al., 2003; Caminsky et al., 2014; Dorman et al., 2014; Viner et al., 2014; Caminsky et al., 2016; Mucaki et al., 2016; Shirley et al., 2019), and these predictions can be confirmed by *in vitro* experimental studies (Vockley et al., 2000; Lamba et al., 2003; Rogan et al., 2003; Khan et al., 2004; Susani et al., 2004; Hobson et al., 2006; Caux-Moncoutier et al., 2009; Olsen et al., 2014; Vemula et al., 2014; Peterlongo et al., 2015). Strengths of one or more splice sites may be altered and, in some instances, concomitant with amino acid changes in coding sequences (Rogan et al., 1998; Peterlongo et al., 2015). Information analysis has been a successful approach for recognizing nondeleterious, sometimes polymorphic variants (Rogan and Schneider, 1995; Colombo et al., 2013), and for distinguishing of milder from severe mutations (Rogan et al., 1998; von Kodolitsch et al., 1999; Lacroix et al., 2012).

Predicting the relative abundance of various transcripts by information analysis requires integration of the contributions of all pertinent cis-acting regulatory elements. We have applied quantitative methods to prioritize inferences as to which SNPs impact gene expression levels and transcript structure. Effects of mutations on combinations of splicing signals reveal changes in isoform structure and abundance (Mucaki et al., 2013; Caminsky et al., 2014). Multisite information theory-based models have also been used to detect and analyze SNP effects on cis-acting promoter modules that contribute to establishing transcript levels (Bi and Rogan, 2004; Vyhldal et al., 2004; Lu et al., 2017; Lu and Rogan, 2019).

The robustness of this approach for predicting rare, deleterious splicing mutations justifies efforts to identify common SNPs that impact mRNA splicing. We previously described SNPs from dbSNP that affect splicing (Rogan et al., 1998; Nalla and Rogan, 2005). Here, we explicitly predict and validate SNPs that influence mRNA structure and levels of expression of the genes containing them in immortalized lymphoblastoid cell lines and tumors. Since constitutive splicing mutations can arise at other locations within pre-mRNA sequences that elicit cryptic splicing, we examined whether more common genomic polymorphisms might frequently affect the abundance and structure of splice isoforms.

METHODS

Information Analysis

The protein-nucleic acid interactions intrinsic to splicing can be analyzed using information theory, which comprehensively and quantitatively models functional sequence variation based on a thermodynamic framework (Schneider, 1997). Donor and acceptor splice site strength can be predicted by the use of IT-based weight matrices derived from known functional sites (Rogan et al., 2003). The Automated Splice Site and Exon Definition server (ASSEDA) is an online resource based on the hg19 coordinate system to determine splice site information changes associated with genetic diseases (Mucaki et al., 2013). ASSEDA is now part of the MutationForecaster (<http://www.mutationforecaster.com>) variant interpretation system.

Creation of Exon Array Database

Exon-level microarrays have been used to compare abnormal expression for different cellular states, which can then be confirmed by q-RT-PCR (Thorsen et al., 2008). We hypothesized that the predicted effect of SNPs on expression of the proximate exon would correspond to the expression of exon microarray probes of genotyped individuals in the HapMap cohort. We used the dose-dependent expression of the minor allele to qualify SNPs for subsequent information analysis consistent with alterations of mRNA splicing. Additional SNPs predicted by information analysis were also tested for effects on splicing (Nalla and Rogan, 2005).

Expression data were normalized using the PLIER (Probe Logarithmic Intensity Error) method on Affymetrix Human

Exon 1.0 ST microarray data for 176 genotyped HapMap cell lines (Huang et al., 2007, Gene Expression Omnibus accession no. GSE 7792; Nembaware et al., 2008). Microarray probes which overlap SNPs, that were subsequently removed, were identified by intersecting dbSNP129 with probe coordinates [obtained from X:MAP (Yates et al., 2008) using the Galaxy Browser (Giardine et al., 2005)]. A MySQL database containing the PLIER normalized intensities and CEU (Utah residents with Northern and Western European ancestry) and YRI (Yoruba in Ibadan) genotypes for Phase I+II HapMap SNPs was created. Tables were derived to link SNPs to their nearest like-stranded probeset (to within 500 nt), and to associate probesets to the exons they may overlap (transcript and exon tables from Ensembl version 51). A MySQL query was used to create a table containing the splicing index (SI; intensity of a probeset divided by the overall gene intensity) of each probeset for each HapMap individual.

The database was queried to identify significant SI changes of an exonic probeset based on the genotype of a SNP the probeset was associated with (SNP within natural donor/acceptor region of exon). Probesets displaying a stepwise change in mean SI (where the mean SI of the heterozygous group is in between the mean SI values of the two homozygous groups) were identified using a different program script (criteria: the mean SI of homozygous rare and heterozygous groups are < 90% of the homozygous common group). Splicing Index boxplots were created with R, where the x- and y-axis are genotype and SI, respectively (**Supplementary Image 1**). These boxplots analyze the effect a SNP has on a particular probeset across all individuals.

SNPs with effects on splicing were validated by q-RT-PCR of lymphoblastoid cell lines. Where available, results were also compared to abnormal splicing patterns present in RNAseq data from tumors carrying these same SNPs (in the ValidSpliceMut database; Shirley et al., 2019). SNPs predicted to exhibit nominal effects on splicing ($\Delta R_i < 1$ bit) were included to determine minimal detectable changes by q-RT-PCR.

Cell Culture and RNA Extraction

EBV-transformed lymphoblastoid cell lines of HapMap individuals with our SNPs of interest (homozygous common, heterozygous and homozygous rare when available) were ordered from the Coriell Cell Repositories (CEU: GM07000, GM07019, GM07022, GM07056, GM11992, GM11994, GM11995, GM12872; YRI: GM18855, GM18858, GM18859, GM18860, GM19092, GM19093, GM19094, GM19140, GM19159). Cells were grown in HyClone RPMI-1640 medium [15% FBS (HyClone), 1% L-Glutamine, and 1% Penicillin: streptomycin (Invitrogen); 37°C, 5% CO₂]. RNA was extracted with Trizol LS (Invitrogen) from 10⁶ cells and treated with DNAase [20 mM MgCl₂ (Invitrogen), 2 mM DTT (Sigma-Aldrich), 0.4 U/μL RNasin (Promega), 10 μg/ml DNase (Worthington Biochemical) in 1x TE buffer] at 37°C for 15 min. The reaction was stopped with EDTA (0.05 M; 2.5% v/v), and heated to 65°C for 20 min, followed by ethanol precipitation (resuspended in 0.1% v/v DEPC-treated 1x TE buffer). DNA was extracted using a Puregene Tissue Core Kit B (Qiagen).

Design of Real-Time Expression Assays

Sequences were obtained from UCSC and Ensembl. DNA primers used to amplify a known splice form, or one predicted by information analysis, were designed using Primer Express (ABI). DNA primers (**Supplementary Table 1**) were obtained from IDT (Coralville, IA, USA), and dissolved to 200 μ M. Primers were placed over junctions of interest to amplify a single splice form. T_m ranged from 58°C–65°C, and amplicon lengths varied from 69–136 nt. BLASTn (Refseq_RNA database) was used to reduce possible cross-hybridization. Primers were designed to amplify the wildtype splice form, exon skipping (if a natural site is weakened), and cryptic site splice forms which were either previously reported (UCSC mRNA and EST tracks) or predicted by information analysis (where R_i cryptic site $\geq R_i$ weakened natural site).

Two types of reference amplicons were used to quantify allele specific splice forms. These consisted of intrinsic products derived from constitutively spliced exons with the same gene and external genes with high uniformity of expression among HapMap cell lines. Reference primers internal to the genes of interest were designed 1–4 exons adjacent from the affected exon (exons without any evidence of variation in the UCSC Genome Browser; Kent et al., 2002), placed upstream of the SNP of interest whenever possible. Two advantages to including an internal reference in the q-RT-PCR experiment include: potential detection of changes in total mRNA levels; and account for inter-individual variation of expression.

External reference genes (excluding the SNP of interest) were chosen based on consistent PLIER intensities with low coefficients of variation in expression among all 176 HapMap individuals. The following external controls were selected: exon 39 of *SI* (PLIER intensity 11.4 ± 1.7), exon 9 of *FRMPD1* (22 ± 2.81), exon 46 of *DNAH1* (78.5 ± 9.54), exon 3 of *CCDC137* (224 ± 25), and exon 25 of *VPS39* (497 ± 76). The external reference chosen for an experiment was matched to the intensity of the probeset within the exon of interest. This decreased potential errors in $\Delta\Delta C_T$ values and proved to be accurate and reproducible for most genes.

To control for interindividual variation in expression, we compared expression in HapMap individuals based on their SNP genotypes and familial relatedness. Families with all three possible genotypes were available (homozygous common, rare, and heterozygous) for 12 of these SNPs (rs1805377, rs2243187, rs2070573, rs2835655, rs2835585, rs2072049, rs1893592, rs6003906, rs1018448, rs13076750, rs16802, and rs8130564). For those families in which all genotypes were not represented, samples from the same ethnic background (YRI or CEU populations) were compared for the missing genotype ($N = 8$; rs17002806, rs2266988, rs1333973, rs743920, rs2285141, rs2838010, rs10190751, rs16994182; individuals with homozygous common and rare genotypes were from the same families for the latter two SNPs). Two SNPs were tested using homozygous individuals from different ethnic backgrounds: rs3747107 (*GUSBP11*) and rs2252576 (*BACE2*). While the splicing impact of rs3747107 was clearly observable by q-RT-PCR, either background or data noise did impact the interpretation of effects of rs2252576.

PCR and Quantitative RT-PCR

M-MLV reverse transcriptase (Invitrogen) converted 1 μ g of DNase-treated RNA to cDNA with 20 nt Oligo-dT (25 μ g/ml; IDT) and rRNasin (Promega). Precipitated cDNA was resuspended in water at 20 ng/ μ l of original RNA concentration. All designed primer sets were tested with conventional PCR to ensure a single product at the expected size. PCR reactions were prepared with 1.0 M Betaine (Sigma-Aldrich), and were heated to 80°C before adding Taq Polymerase (Invitrogen). Optimal T_m for each primer set was determined to obtain maximum yield.

Quantitative PCR was performed with an Eppendorf Mastercycler ep Realplex 4, a Bio-Rad CFX96, as well as a Stratagene Mx3005P. SYBR Green assays were performed using the KAPA SYBR FAST qPCR kit (Kapa Biosystems) in 10 μ l reactions using 200 μ M of each primer and 24 ng total of cDNA per reaction. For some tests, SsoFast EvaGreen supermix (Bio-Rad) was used with 500 μ M of each primer instead.

When testing the effect of a SNP, amplification reactions with all primers designed to detect all relevant isoforms (as well as the gene internal reference and external reference) were run simultaneously, in triplicate. C_t values obtained from these experiments were normalized against the same external reference using the Relative Expression Software Tool (REST; <http://www.gene-quantification.de/rest.html>; Pfaffl et al., 2002).

Taqman Assay

Two dual-labeled Taqman probes were designed to detect the two splice forms of *XRCC4* (detecting alternative forms of exon 8 either with or without a 6 nt deletion at the 5' end). Probes were placed over the sequence junction of interest where variation would be near the probe middle (**Supplementary Table 1**). The assay was performed on an ABI StepOne Real-Time PCR system using ABI Genotyping Master Mix. Experiment was run in 25 μ l reactions (300 nM each primer, 400nM probe [5'-FAM or TET fluorophore with a 3' Black Hole quencher; IDT], and 80 ng cDNA total). Probes were tested in separate reactions.

RNAseq Analyses

The previous analyses were extended to include 24 additional, common SNPs for their potential influence on splicing. All SNVs present in ICGC (International Cancer Genome Consortium) patients (Shirley et al., 2019) were evaluated by the Shannon Pipeline (SP; Shirley et al., 2013) to identify those altering splice site strength. Common SNPs (average heterozygosity > 10% in dbSNP 150) predicted to decrease natural splice site strength by SP (where $\Delta R_i < -1$ bit) were selected. ICGC patients carrying these flagged SNPs were identified, and the expression of the corresponding SNP-containing region in RNAseq was visualized with IGV (Integrated Genome Viewer; <https://igv.org>; Robinson et al., 2017). Similar RNAseq reads were grouped using IGV collapse and sort commands, which caused nonconstitutive spliced reads to cosegregate to the top of the viewing window. IGV images which did not meet our gene expression criteria (exon affected by the SNP must have ≥ 5 RNAseq reads present) were eliminated. As this generated thousands of images, we

report the analysis of two ICGC patients [DO47132 (Renal Cell Cancer) and DO52711 (Chronic Lymphocytic Leukemia)], chosen randomly, preselecting tissues to increase the likelihood of finding expression in these regions. Images were evaluated sequentially (in order of rsID value) and only concluded once the first 24 SNPs meeting these criteria were found. This type of analysis could not reveal a splicing event to be more abundant in these patients when compared to noncarriers. Nevertheless, splicing information changes resulting from SNPs corresponded to observed alternative and/or other novel splice isoforms. We then queried the ValidSpliceMut database for these SNPs, as abnormal splicing was only flagged in the database when the junction-read or read-abundance counts significantly exceeded corresponding evidence type in a large set of normal control samples (Shirley et al., 2019).

RESULTS

Selection of Candidate SNPs Affecting Splicing

A publicly available exon microarray dataset was initially used to locate exons affected by SNPs altering splice site strength. A change in the mean SI of a diagnostic probeset in individuals of differing genotypes at the same variant can suggest altered splicing. The increase or decrease in SI is related to the expected impact of the SNP on splicing. For example, an exonic probe which detects a normally spliced mRNA will have decreased SI in the event of skipping. Mean SI may be increased when a probe detects the use of an intronic cryptic splice site. SNPs with strong impact on splicing will distinguish mean SI levels of individuals homozygous for the major versus minor alleles (and with heterozygous genotypes).

There were 9,328 HapMap-annotated SNPs within donor/acceptor regions of known exons which contained at least one probeset. Of 987 SNPs that are associated to exonic probesets which differ in mean SI between the homozygous common and rare HapMap individuals, 573 caused a decrease in natural site R_i value. Inactivating and leaky splicing variants (reduction in information content where final $R_i \geq R_{i, \text{minimum}}$ [minimum functional splice site strength]) both exhibit reduced SI values and were similarly abundant. Thus, both severe and moderate splicing mutations with reduced penetrance and milder molecular phenotypes were detected, consistent with Mendelian disorders (von Kodolitsch et al., 1999; von Kodolitsch et al., 2006).

Of the SNPs associated with significant changes in R_i (termed ΔR_i), 9,328 occurred within the natural splice sites of exons detectable with microarray probesets. We initially focused on 21 SNPs on chromosome 21 (0.23% total, 18.8% of chr21) and 34 on chromosome 22 (0.36% of total, 14.5% of chr22) associated with stepwise decreases in probeset intensity at each genotype. Seven of the chr21 SNPs and nine of the chr22 SNPs caused information changes with either natural splice site $\Delta R_i \geq 0.1$ bits, or cryptic site(s) with an R_i value comparable to a neighbouring natural site, and in which mRNA or EST data

supported use of the cryptic site. These SNPs included: rs2075276 [MGC16703], rs2838010 [FAM3B], rs3747107 [GUSBP11], rs2070573 [C21orf2], rs17002806 [WBP2NL], rs3950176 [EMID1], rs1018448 [ARFGAP3], rs6003906 [DERL3], rs2266988 [PRAME], rs2072049 [PRAME], rs2285141 [CYB5R3], rs2252576 [BACE2], rs16802 [BCR], rs17357592 [COL6A2], rs16994182 [CLDN14], and rs8130564 [TMPRSS3].

The minimum information change for detecting a splicing effect by expression microarray is constrained by several factors. Detection of splice isoforms can be limited by genomic probeset coverage, which cannot distinguish alternative splicing events in close proximity (see **Figure 1A**). Even where genotype-directed SI changes are very distinct, some individuals with the common allele have equivalent SI values to individuals with the rare allele [rs2070573 (**Figure 2**) and rs1333973 (**Figure 3**)]. In some cases, the number of individuals with a particular genotype is insufficient for statistical significance (rs2243187; **Supplementary Image 1.4**). Although exon microarrays can be used to find potential alternate splicing and give support to our predictions, it became necessary to validate the microarray predictions by q-RT-PCR, TaqMan assays, and with RNAseq data from SNP carriers.

We report q-RT-PCR validation studies for 13 of the 16 SNPs (q-RT-PCR primers could not be designed for rs16994182, rs2075276, and rs3950176), along with nine other candidate SNPs from our previous information theory-based analyses (Nalla and Rogan, 2005): rs1805377 [XRCC4], rs2243187 [IL19], rs2835585 [TTC3], rs2865655 [TTC3], rs1893592 [UBASH3A], rs743920 [EMID1], rs13076750 [LPP], rs1333973 [IFI44L], and rs10190751 [CFLAR].

After amplification of known and predicted splice forms (**Supplementary Table 1**), 15 SNPs showed measurable changes in splicing consistent with information-theory predictions. Ten increase alternate splice site use (two of which increased strength of cryptic site, eight activated an unaffected pre-existing cryptic site), six affect exon inclusion (five increased exon skipping), three increased activation of an alternative exon, and four decreased overall expression levels. Altered splicing could not be validated for six SNPs, however experimental analyses of three of the five SNPs where $\Delta R_i < 1$ bit were hampered by high interindividual variability in expression.

Changes in splice site information were used to predict observed differences in splice isoform levels (**Table 1**). **Figures 1–3** and **Supplementary Image 1** indicate the experimentally-determined splicing effects for each SNP, a modified UCSC Genome Browser image of the relevant region, boxplots showing exon microarray expression levels of each allele for the relevant probesets, and an IGV image of the RNAseq results for an individual tumor carrying the SNP. Abundance of the aberrant splice forms measured by q-RT-PCR (relative to an internal gene reference) is indicated in **Table 2**. Changes in predicted splice site strength were consistent with results measured by q-RT-PCR for 12 out of the 15 SNP (exceptions were rs2070573, rs17002806, and rs2835585). Variants predicted to reduce strength ≥ 100 -fold

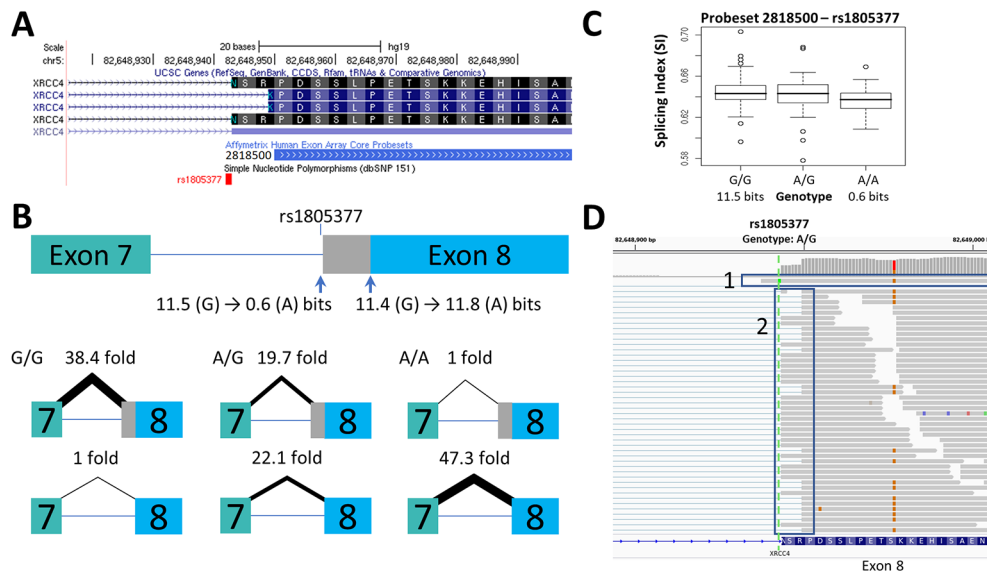


FIGURE 1 | Splicing Impact of rs1805377 (*XRCC4*). The natural acceptor of *XRCC4* exon 8 is abolished by rs1805377 (11.5 → 0.6 bits) while simultaneously strengthening a second exononic cryptic acceptor 6nt downstream (11.4 to 11.8 bits), resulting in a 6nt deletion in the mRNA. **(A)** Both of these acceptor sites have been validated in GenBank mRNAs, i.e., NM_022406 and NM_003401 (UCSC panel derived from <http://genome.ucsc.edu>). **(B)** The relative abundance of the two splice forms was determined by q-RT-PCR. The weaker rs1805377 A/A genotype (0.6 bit acceptor) was used ~47-fold less frequently than the cryptic downstream acceptor (11.8 bits). **(C)** The two splice isoforms cannot be distinguished by the exon microarray as the upstream probeset (ID 2818500) does not overlap the variable region, though the average expression of the rs1805377 A/A genotype is reduced. **(D)** ValidSpliceMut flagged this mutation for intron retention, which can be observed in the RNAseq of heterozygous ICGC patient DO27779 [Box 1]. Use of both acceptor sites is also evident [Box 2]. For more detail for this and all of the other single nucleotide polymorphisms (SNPs) analyzed, refer to **Supplementary Image 1**.

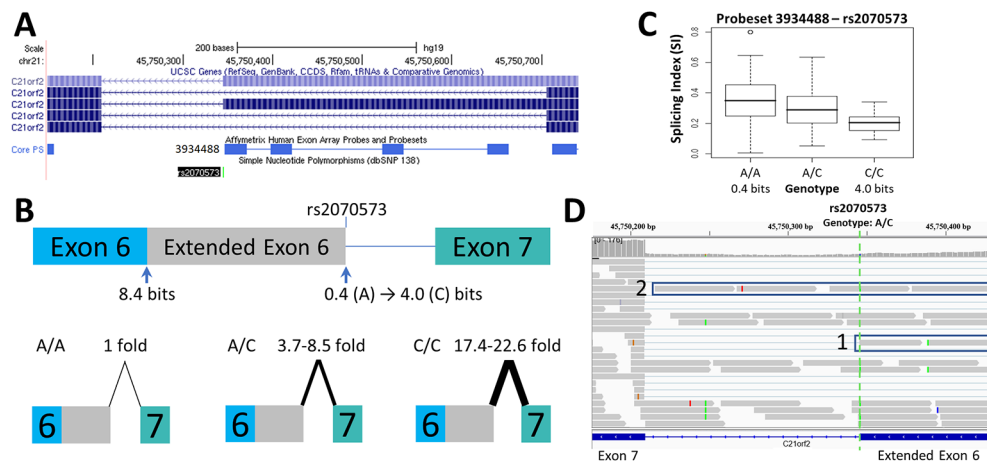


FIGURE 2 | Splicing Impact of rs2070573 (*C21orf2*). **(A)** The single nucleotide polymorphisms (SNP) rs2070573 is a common polymorphism which alters the first nucleotide of the extended form of *C21orf2* exon 6. **(B)** The donor site is strengthened by the presence of the C-allele (R; 0.4 to 4.0 bits; A > C) and its use extends the exon by 360 nt. Q-RT-PCR found a ~4-9-fold and ~17-23-fold increase in the extended exon 6 splice form in the A/C and C/C cell lines tested, respectively. **(C)** The exon microarray probeset which detects the extension (ID 3934488) shows a stepwise increase in SI with C-allele individuals which supports the q-RT-PCR result. **(D)** The variant was present in ValidSpliceMut, which associated the A-allele with an increase in total intron retention [six patients flagged for total intron retention read abundance; $p=0.019$ (average over all patients)]. This image displays sequence read distributions in the RNAseq data of TCGA BRCA patient, TCGA-BH-A0H0, who is heterozygous for rs2070573. The IGV panel indicates reads corresponding to total intron 6 retention [Box 1] and which extend beyond the constitutive donor splice site of exon 6 into the adjacent intron [Box 2]. All 4 reads which extend over the exon splice junction are derived from the G-allele (strong binding site; not all visible in panel D).

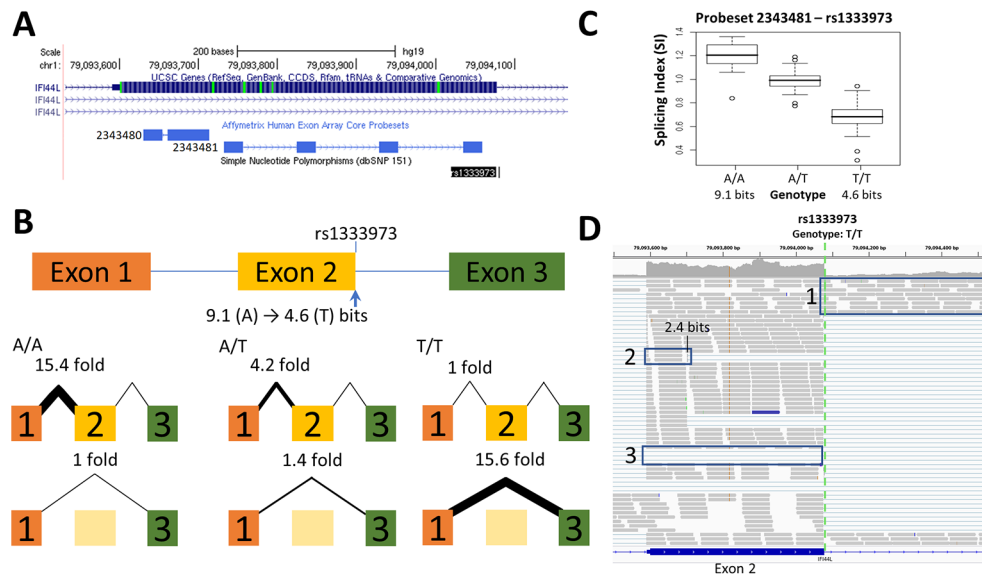


FIGURE 3 | Splicing Impact of rs1333973 (*IFI44L*). **(A)** The natural donor of *IFI44L* exon 2 is weakened from 9.1 to 4.6 bits ($A > T$) by rs1333973, increasing the frequency of exon 2 skipping and other events. **(B)** By q-RT-PCR, skipping was found to be 15.6-fold higher while normal splicing was 15.4-fold lower in A homozygotes (relative to T homozygotes). **(C)** Exon microarray data strongly supports the q-RT-PCR findings. **(D)** ICGC patient DO6354 is homozygous for the T-allele, which resulted in exon 2 skipping [Box 3], failure to recognize the exon 2 donor causing total retention of intron 2 [Box 1], and activation of an upstream exonic 2.4 bit cryptic donor 375 nt from the affected site [Box 2].

were found to reduce expression by 38- to 58-fold, the variance falling within the margin of measurement error. Modest natural splice site affinity changes predicted to be $< \text{eightfold}$ ($\Delta R_i < 3.0$) did not consistently result in detectable changes in splicing. In some instances, lower abundance splice forms were observed (i.e. rs2835585 altered exon skipping levels by up to 8.8-fold; nevertheless, the normal splice form predominated).

SNPs Affecting Cryptic Site Strength and Activity

Increased cryptic site use coinciding with a decrease in natural site strength (Table 1) was validated for: rs1805377 (Figure 1); rs2243187 (Supplementary Image 1.4); rs3747107 (Supplementary Image 1.8); rs17002806 (Supplementary Image 1.13); rs6003906 (Supplementary Image 1.15); and rs13076750 (Supplementary Image 1.12). rs2070573 (Figure 2) and rs743920 (Supplementary Image 1.7) strengthened cryptic splice sites resulting in increased use of these sites. Despite the difference in strength between the natural and cryptic sites affected by rs743920, the upstream 2.4 bit site was used more frequently (Table 2). Both *IL19* and *XRCC4* regions tested showed preference to the upstream acceptor as well, which is consistent with the processive mechanism documented to recognize acceptor splice sites (Robberson et al., 1990).

SNPs Affecting Exon Inclusion

SNPs that reduced natural site strength (ΔR_i from 1.6 to 10.9 bits) increased exon skipping from 2- to 27-fold for homozygotes

of differing genotypes of rs2835585 (Supplementary Image 1.21), rs1018448 (Supplementary Image 1.3), rs1333973 (Figure 3), rs2266988 (Supplementary Image 1.9), and rs13076750 (Supplementary Image 1.12). The exon microarray probesets for rs1018448 and rs1333973 detect decreased expression by genotype, which is consistent with increased exon skipping. Changes of average SI values did not correspond as well to specific genotypes for rs2835585 (*TTC3*), rs2266988 (*PRAME*), and rs13076750 (*LPP*), possibly due to increased cryptic site use (*LPP*) or large differences in the abundance of constitutive and skipped isoforms (*PRAME*, *TTC3*).

SNPs Promoting Alternate Exon Use

SNP-related decreases in natural splice site strength may promote the use of alternative exons up or downstream of the affected exon. rs10190751 (Supplementary Image 1.2) is known to modulate the presence of the shorter c-FLIP(S) splice form of *CFLAR* (Ueffing et al., 2009). The use of this exon differed by 2^{17} -fold between the strong and weak homozygotes tested, which was reflected by the expression microarray result. By q-RT-PCR, the *CFLAR* (L) form using an alternate downstream exon was found to be 2.1-fold more abundant in the homozygote with the weaker splice site. rs3747107 (Supplementary Image 1.8) and rs2285141 (Supplementary Image 1.20) exhibit evidence of an increased preference by q-RT-PCR for activation of an alternate exon, though the microarray results for the corresponding genotypes for both SNPs were not significantly different.

TABLE 1 | Summary of q-RT-PCR results.

Summary of q-RT-PCR results					SNP effects (Increase/Decrease in fold change of homozygotes)						Additional Expression Evidence	
Gene	Supplementary Image 1 panel / rsID / HGVS notation	Splice type	Information Change (n-natural, c-cryptic site)	Fold change	Natural site	Cryptic site	Exon skipping	Alternate exon	Total mRNA	In-frame ^a	Exon microarray	RNAseq (Valid-SpliceMut)
XRCC4	S1.6/ rs1805377 /	A	11.5 (G) -> 0.6 (A) (n)	1886	38.4	–	–	–	n/a ^b	Y	N	Y
	NM_003401: c.894-7G>A	A	11.4 (G) -> 11.8 (A) (n)	1.3	–	47.3 ^c	–	–	n/a ^b	Y	N	Y
	NM_022406: c.894-1G>A											
IL19	S1.4/ rs2243187 / NM_013371: c.364-1G>A	A	4.7 (G) -> -6.2 (A) (n)	26.3	1.8 ^d	1.8 ^d	2.1 ^d	–	n/a ^b	Y	Y	N
C21orf2	S1.1/ rs2070573 / NM_004928: c.643-137A>C	D	0.4 (A) -> 4.0 (C) (c)	12.1	NC	22.6	–	–	n/a ^b	Y	Y	Y
TTC3	S1.21/ rs2835655 / NM_003316: c.5115G>A	D	11.2 (G) -> 8.2 (A) (n)	8.1	1.5	–	1.3 ^e	–	1.5	–	N	N
TTC3 ^f	S1.5/ rs2835585 / NM_003316: c.188-8T>A	A	7.6 (T) -> 5.4 (A) (n)	4.5	NC	–	8.8	–	n/a ^b	Y	N	Y
WBP2NL	S1.13/ rs17002806 / NM_152613: c.*86G>A	D	9.2 (G) -> 6.0 (A) (n)	9.1	23.8	34+ ^g	–	–	n/a ^b	N	Y	N
GUSBP11	S1.8/ rs3747107 / NR_024448: n.2961+6512 C>G	A	4.7 (C) -> -7.0 (G) (n)	25.9	98.3	31/ 42.8 ^h	–	1.6	n/a ^b	Y/N ⁱ	N	N
PRAME ^j	S1.9/ rs2266988 ^k / NM_006115: c.19G>A	D	7.8 (G) -> 6.2 (A) (n)	3.0	NC	–	8.8	–	n/a ^b	N	N	N
PRAME ^j	S1.14/ rs2072049 / NM_006115: c.954-16G>T	A	8.2 (G) -> 7.1 (T) (n)	2.2	2.6 ^d	–	–	–	3.1 ^d	–	Y	Y
UBASH3A	S1.10/ rs1893592 / NM_018961: c.1393+3A>C	D	8.7 (A) -> 4.2 (C) (n)	22.9	3.0	2.0/1.7 ^l	N/A ^m	–	1.4	Y	N	Y
DERL3	S1.15/ rs6003906 / NM_198440: c.328-8A>T	A	-2.1 (A) -> -4.3 (T) (n)	4.6	NC	2.0	–	–	n/a ^b	N	Y	N
ARFGAP3	S1.3/ rs1018448 / NM_014570: c.1065C>A	A	10.5 (C) -> 8.3 (A) (n)	4.5	2.2	–	1.4	–	2.0	Y	Y	N
CFLAR	S1.2/ rs10190751 / NM_003879: c.606 +934G>A NM_001127184:c.607-1G>A	A	16.1 (G) -> 5.2 (A) (n)	1885	10000+	–	–	2.1	n/a ^b	Y	Y	Y
IFI44L	S1.11/ rs1333973 / NM_006820: c.478+3A>T	D	9.1 (A) -> 4.6 (T) (n)	22.6	15.4	–	15.6	–	3.9	N	Y	N
LPP	S1.12/ rs13076750 / NM_001167671:c.-66-8G>A	A	9.3 (G) -> -1.6 (A) (n)	625	5000+	15.8	26.6	–	–	Y	N	Y
EMID1	S1.7/ rs743920 / NM_133455: c.326C>G	A	6.0 (C) -> 7.9 (G) (c)	3.6	n/a ^b	5.8	–	–	n/a ^b	–	Y	N
CLDN14	S1.22/ rs16994182 / NM_001146078:c.-82 +4C>G	D	7.4 (C) -> 6.8 (G) (n)	1.6	5.9	–	2.1	–	n/a ^b	Y	N/A	N
BCR	S1.17/ rs16802 / NM_004327:c.2708-13A>G	A	5.6 (A) -> 5.8 (G) (n)	1.1	NC ⁿ	–	–	–	n/a ^b	–	N	N
TMPRSS3	S1.19 / rs8130564 / NM_032404: c.66-13T>C	A	4.2 (T) -> 4.4 (C) (n)	1.1	NC ⁿ	–	–	–	n/a ^b	–	N	N
BACE2	S1.18/ rs2252576 / NM_012105: c.748-10C>T	A	7.2 (C) -> 8.0 (T) (n)	1.5	NC ⁿ	–	–	–	n/a ^b	–	N	N
CYB5R3	S1.20/ rs2285141 / NM_007326: c.-48-18G>T	A	2.0 (G) -> 1.2 (T) (n)	1.7	NC	–	–	1.8	n/a ^b	–	N	N
FAM3B	S1.16/ rs2838010 / NM_058186: c.20-217A>T	D	-10.8 (A) -> 7.8 (T) (c)	228	–	–	–	N/A ⁱ	n/a ^b	Y	N	N

Red text indicates a decrease in the abundance of a particular splice form, while green text indicates an increase in abundance. A – Acceptor Splice Site Affected; D – Donor Splice Site Affected; NC – Not detectable (abolished). ^a Splicing events which alter reading frame may induce nonsense-mediated decay; ^b No allele specific difference in expression and splicing; ^c complete discrimination of both isoforms using a custom designed TaqMan probe; ^d Values from comparing heterozygote with homozygote common; ^e Change in splicing likely related to change in RNA level; ^f Intron 2-3 retention of TTC3 amplified by PCR, but no allele specific change detected; ^g This splice form not at detectable levels in homozygote; ^h Cryptic acceptor 114nt upstream of affected site / cryptic acceptor 118nt upstream of affected site; ⁱ mRNA in-frame when alternate exon is used, and out of frame due to cryptic site use; ^j PRAME is a special case where two SNPs affect splicing of two separate exons; ^k rs2266988 and rs1129172 are identical SNPs on opposite strands; ^l Cryptic donor 55nt downstream of affected site / cryptic donor 29nt downstream of affected site; ^m Splice form not detected by PCR; ⁿ High variability between individuals with the same genotype by q-RT-PCR.

TABLE 2 | Abundance of mRNA splice forms relative to internal gene reference.

Gene	rsID	mRNA Splice Form	Homozygotes strong allele (%) [# Patients Tested]	Homozygotes weak allele (%) [# Patients Tested]
<i>C21orf2</i>	rs2070573	Extended Exon 6 Splice Form	45.3 ± 16.9 [2] ¹	1.5 ± 0.2 [2]
<i>TTC3</i>	rs2835585	Exon 3 Skipping	<0.1 ± 0.0 [2]	0.3 ± 0.1 [2]
<i>IL19</i> ²	rs2243187	3 nt Inclusion of Exon 5	~100 [1]	53.2 ± 3.0 [1; Het.]
<i>IL19</i> ²	rs2243187	3 nt Exclusion of Exon 5	61.0 ± 3.4 [1]	~100 [1; Het.]
<i>IL19</i> ²	rs2243187	Exon 5 Skipping	14.1 ± 1.4 [1]	6.8 ± 1.1 [1; Het.]
<i>UBASH3A</i>	rs1893592	29 nt Retention of Intron 10	8.6 ± 4.6 [1]	4.2 ± 5.8 [1]
<i>XRCC4</i>	rs1805377	6 nt Inclusion of Exon 8	~100 [1]	3.1 ± 0.5 [1]
<i>XRCC4</i>	rs1805377	6 nt Exclusion of Exon 8	2.6 ± 1.1 [1]	61.4 ± 26.3 [1]
<i>PRAME</i>	rs2266988	Normal Exon 3 Splicing	32.4 ± 3.6 [1]	20.1 ± 3.9 [1]
<i>PRAME</i>	rs2266988	Exon 3 Skipping	1.2 ± 1.2 [1]	3.5 ± 3.0 [1]
<i>GUSBP11</i>	rs3747107	Exon 8 Splicing	34.0 ± 0.6 [1]	0.3 ± 0.0 [2]
<i>GUSBP11</i>	rs3747107	Alternative Exon 8	37.3 ± 7.3 [1]	39.8 ± 3.7 [2]
<i>GUSBP11</i>	rs3747107	114 nt Retention of Intron 7	0.1 ± 0.0 [1]	1.4 ± 0.1 [2]
<i>GUSBP11</i>	rs3747107	118 nt Retention of Intron 7	0.2 ± 0.0 [1]	8.0 ± 0.5 [2]
<i>DERL3</i>	rs6003906	Normal Exon 5 Splicing	~100 [1]	55.8 ± 0.0 [1]
<i>DERL3</i>	rs6003906	Extended Exon 4; Short Exon 5	3.2 ± 0.0 [1]	4.0 ± 0.0 [1]
<i>ARFGAP3</i>	rs1018448	Exon 12 Skipping	12.3 ± 5.7 [3]	23.1 ± 10.1 [1]
<i>IFI44L</i>	rs1333973	Normal Exon 2 Splicing	57.0 ± 0.0 [1]	14.7 ± 0.0 [1]
<i>IFI44L</i>	rs1333973	Exon 2 Skipping	0.8 ± 0.0 [1]	48.0 ± 0.0 [1]
<i>CFLAR</i>	rs10190751	Upstream Exon 7 Use	~100 [1]	<0.1 ± 0.0 [1]
<i>CFLAR</i>	rs10190751	Downstream Exon 7 Use	39.0 ± 0.0 [1]	87.1 ± 0.0 [1]
<i>WBP2NL</i>	rs17002806	25 nt Intron 6 Retention	N.D. ²	2.3 ± 0.0 [1]
<i>CYB5R3</i>	rs2285141	Alternate Exon 2 Use	<0.1 ± 0.0 [1]	<0.1 ± 0.0 [1]
<i>EMID1</i>	rs743920	6 nt deletion of Exon 4	69.0 ± 9.7 [1]	9.2 ± 1.4 [1]
<i>CLDN14</i>	rs16994182	Exon 2 Skipping	1.9 ± 0.0 [1]	5.3 ± 0.0 [1]

¹Average expression was computed by comparing qPCR C_t values across multiple experimental runs and normalized against C_t of internal gene reference. SNPs tested in multiple experiments with one individual of each genotype will have a standard deviation of 0.0. ²Heterozygote; Individuals who are homozygous for IL19 SNP rs2243187 were not available for testing. N.D., Not detected. C_t values were not available for LPP rs13076750.

SNP-Directed Effects on mRNA Levels

A change in the strength of a natural site of an exon can affect the quantity of the processed mRNA (Caminsky et al., 2014). This decrease in mRNA could be caused by nonsense mediated decay (NMD), which degrades aberrant transcripts that would result in premature protein truncation (Cartegni et al., 2002). Of the 22 SNPs tested, 2 showed a direct correlation between a decrease in natural splice site strength, reduced amplification of the internal reference by q-RT-PCR (of multiple individuals) and a decreasing trend in expression by genotype by microarray: rs2072049 (**Supplementary Image 1.9**) and rs1018448 (**Supplementary Image 1.3**), although these differences do not meet statistical significance.

SNPs With Pertinent Splicing Effects Detected by RNAseq

Evidence for impact on splicing of the previously described SNPs was also assessed in TCGA and ICGC tumors by high throughput expression analyses. Splicing effects of these variants detected by q-RT-PCR and RNAseq were concordant in 80% of cases ($N = 16$ of 20 SNPs), while impacts of 10% of SNPs ($N = 2$) were partially concordant as a result of inconsistent activation of cryptic splice sites (**Supplementary Images 1.8D** and **1.10D**). Several isoforms predicted by information analysis of these SNPs were present in complete transcriptomes, but were undetectable by q-RT-PCR or expression microarrays. Examples include a 4.9 bit cryptic site activated by rs3747107 located 2

nucleotides from the natural splice site (**Supplementary Image 1.8D**), exon skipping by rs1893592 (**Supplementary Image 1.10D**), and a cryptic exon activated by rs2838010 (**Supplementary Image 1.16C**). Processed mRNAs that were not detected by q-RT-PCR may have arisen as a result of a lack of sensitivity of the assay, to NMD (which could mask detection of mis-splicing), to a deficiency of an undefined trans-acting splicing factor, or to design limitations in the experimental design. Another possibility is that the discordant splicing patterns of these two SNPs based could potentially be related to differences in tissue origin, since only the RNAseq findings were tumor-derived, whereas results obtained by the other approaches were generated from RNA extracted from lymphoblastoid cell lines. Cell culture conditions such as cell density and phosphorylation status can affect alternative splicing patterns (Li et al., 2006; Szafranski et al., 2014). These conditions, however, have not been studied in cases of allele-specific, sequence differences at splice sites or cis-acting regulatory sites that impact splice site selection. Considering the high level of concordance of splicing effects for the same SNPs in uncultured and cultured cells, it seems unlikely that culture conditions significantly impacts the majority of allele-specific, alternatively spliced isoforms. Our information theory-based analyses show that the dominant effect of SNP genotypes is to dictate common changes in splice site strength regardless of cell origin.

The results obtained from q-RT-PCR and RNAseq data for rs2070573, rs10190751, rs13076750, rs2072049, rs2835585,

rs1893592, and rs1805377 were complementary to findings based on RNAseq (**Supplementary Table 2**). RNAseq data can reveal potential allele-specific alternate splicing events that were not considered at the primer design phase of the study, while q-RT-PCR is more sensitive and can reveal less abundant alternative splice forms. A weak 0.4 bit splice site associated with rs2070573 was less abundant than the extended isoform (**Figure 2**) by both q-RT-PCR and exon microarray, however ValidSpliceMut also revealed increased total *C21orf2* intron 6 retention in five tumors with this allele. Similarly, rs10190751 was flagged for intron retention in 29 tumors, which was not evident by the other approaches. The long form of this transcript (c-FLIP[L]) in homozygous carriers of this SNP was twice as abundant by q-RT-PCR than the shorter allele, associated with the weak splice site. rs13076750 activates an alternate acceptor site for a rare exon that extends the original exon length by seven nucleotides. The exon boundary can also extend into an adjacent exon, based on RNAseq of eight tumors carrying this SNP. Expression was decreased in the presence of a 6.2 bit splice site derived from a rs2072049 allele that weakens the natural acceptor site of the terminal exon of *PRAME*. The actual cause of diminished expression is likely to have been related to NMD from intron retention. ValidSpliceMut showed intron retention to be increased in rs2835585, whereas increased exon skipping for the allele with the weaker splice site was demonstrated by q-RT-PCR. rs1893592 caused significant intron retention in all tumors ($N = 9$), with exon skipping present in 3 diffuse large B-cell lymphoma patients, which was not detected by q-RT-PCR.

Finally, rs1805377 was associated with the significant abundance of read sequences indicating *XRCC4* intron 7 retention by RNAseq ($N = 32$), however this isoform could not be distinguished by the primers designed for q-RT-PCR and by TaqMan assay.

Alternative splicing events detected by RNAseq that were not evident in either q-RT-PCR or microarray studies included exon skipping induced by rs743920 (**Supplementary Image 1.7D**), activation of a preexisting cryptic splice site by rs1333973 (**Figure 3**), and intron retention by rs6003906 (**Supplementary Image 1.15D**). rs743920 creates an exonic hnRNP A1 site ($R_i = 2.8$ bits) distant from the natural site which may compromise exon definition (Mucaki et al., 2013; Peterlongo et al., 2015) and may explain the SNP-associated increase in exon skipping. Exon definition analyses of total exon information ($R_{i,total}$) also predicted the cryptic isoform arising from rs1333973 to be the most abundant ($R_{i,total} = 9.4$ bits).

Allele-Specific mRNA Splicing for Other SNPs Identified Through RNAseq

A distinct set of 24 high population frequency SNPs were also evaluated for their potential impact on mRNA splicing by RNAseq analysis of ICGC patients. Those resulting in significantly decreased natural splice site strength ($\Delta R_i < -1$ bit) were analyzed for SNP-derived alternative splicing events. SNPs fulfilling these criteria expressed at sufficient levels over the region of interest were: rs6467, rs36135, rs154290, rs166062, rs171632, rs232790, rs246391, rs324137, rs324726, rs448580,

TABLE 3 | RNAseq analysis of natural splice sites weakened by common single nucleotide polymorphisms (SNPs).

Gene	rsID ¹	HGVS Notation (HG19) ²	R_i initial	R_i final	ΔR_i	Alternative Splicing Observed
<i>CYP21A2</i>	rs6467	6:32006858C > A	6.1	4.5	-1.6	Intron Retention; Cryptic Site Use
<i>TRIM23</i>	rs36135	5:64890479A > C	10.0	7.5	-2.5	Intron Retention
<i>ZFYVE16</i>	rs166062	5:79773028T > G	14.1	11.8	-2.4	Intron Retention
<i>APBB3</i>	rs171632	5:139941318A > G	2.5	1.4	-1.1	Intron Retention; Cryptic Site Use
<i>SMIM8</i>	rs448580	6:88040399T > G	6.9	4.4	-2.5	Exon Skipping
<i>FCHSD1</i>	rs469074	5:141024136T > G	10.4	7.1	-3.3	Intron Retention
<i>KIFAP3</i>	rs518928	1:169890933A > G	15.6	14.5	-1.1	Intron Retention
<i>CEPT1</i>	rs694180	1:111726213A > G	9.6	7.0	-2.6	Intron Retention
<i>ADCY10P1</i>	rs722442	6:41089681A > G	7.5	4.9	-2.5	Exon Skipping
<i>MICAL1</i>	rs752262	6:109770999G > C	4.0	2.7	-1.4	Intron Retention
<i>MAP3K1</i>	rs832567	5:56152416C > A	6.8	5.0	-1.7	Intron Retention; Exon Skipping
<i>METTL13</i>	rs909958	1:171763522C > A	11.4	9.9	-1.4	Intron Retention
<i>DDX39B</i>	rs933208	6:31506648G > T	4.8	3.7	-1.1	Intron Retention; Cryptic Site Use
<i>CCT7</i>	rs1018342	2:73471653T > G	5.9	4.4	-1.5	Intron Retention
<i>PPIP5K2</i>	rs154290	5:102537200T > G	12.5	11.3	-1.3	Wildtype Only
<i>MYSM1</i>	rs232790	1:59131311G > T	10.4	8.7	-1.7	Wildtype Only
<i>PDGFRB</i>	rs246391	5:149497177T > C	6.2	3.6	-2.6	Wildtype Only
<i>AARS2</i>	rs324137	6:44273546A > C	10.2	8.9	-1.3	Wildtype Only
<i>USO1</i>	rs324726	4:76722353G > A	11.8	8.8	-3.0	Wildtype Only
<i>KIF13A</i>	rs624105	6:17855864G > C	14.1	13.0	-1.1	Wildtype Only
<i>TNFRSF1B</i>	rs653667	1:12251808T > G	3.7	2.4	-1.3	Wildtype Only
<i>CCDC93</i>	rs748767	2:118731573G > A	4.6	3.4	-1.1	Wildtype Only
<i>CAPN2</i>	rs751128	1:223951841T > C	5.3	4.2	-1.1	Wildtype Only
<i>FYCO1</i>	rs751552	3:46016851A > T	6.9	4.7	-2.2	Wildtype Only

¹rsIDs are hyperlinked to their associated dbSNP page; ²If present, variant coordinates are hyperlinked to the ValidSpliceMut database; Thick bars separate SNP-affected exons with and without RNAseq-observed alternate splicing events.

rs469074, rs518928, rs624105, rs653667, rs694180, rs722442, rs748767, rs751128, rs751552, rs752262, rs832567, rs909958, rs933208, and rs1018342 (**Table 3**). Splicing was predicted to be leaky for all natural splice sites affected by these SNPs (Rogan et al., 1998; $R_{i,final} \geq 1.6$ bits), where reduction in R_i values ranged from 1.1 to 3.3 bits.

Alternative mRNA splicing was observed in 14 SNPs: rs6467, rs36135, rs166062, rs171632, rs448580, rs469074, rs518928, rs694180, rs722442, rs752262, rs832567, rs909958, rs933208, rs1018342; **Table 3**). Reads spanning these regions revealed intron retention ($N = 12$), activation of cryptic splicing ($N = 4$), and complete exon skipping ($N = 3$). Eleven of these SNPs (79%) exhibited splicing patterns that significantly differed from the control alleles, and were therefore present in ValidSpliceMut. Interestingly, ValidSpliceMut contained entries for 7 of 10 SNPs where alternative splicing had not been found in the two patients reported in **Table 3** (rs246391, rs324137, rs624105, rs653667, rs748767, rs751128, rs751552). The observed significant splicing differences for these SNPs occurred in distinct tumor types, consistent with tissue-specific effects of these SNPs on splicing.

Instances of Limited Corroboration of SNP-Related Predictions

Anticipated effects of the SNPs on splicing were not always confirmed by expression studies. Aside from incomplete or incorrect predictions, both design and execution of these studies as well as uncharacterized tissue specific effects could provide an explanation for these discrepancies. Furthermore, these undetected splicing events may have been targeted for NMD, however expression was not compared with mRNA levels from cells cultured with an inhibitor of protein translation. Stronger preexisting cryptic sites were, in some instances, not recognized nor was isoform abundance changed. These include: rs1893592 (6.4 and 5.2 bit cryptic donor sites 29 and 555 nt downstream of the affected donor); rs17002806 (a 5.7 bit site 67 nt downstream of the natural site); rs3747107 [creates a 4.9 bit cryptic site 2 nt downstream (observed by RNAseq; **Supplementary Image 1.8 D**)] and rs2835585 (5.8 and 5.9 bit cryptic sites 60 and 87 nt upstream of the natural site). SNPs with modestly decreased natural site strength (0.2 to 4.5 bits) did not consistently result in exon skipping (for example, rs1893592, rs17002806, and rs2835585).

Six SNPs predicted to disrupt natural splice sites could not be confirmed. Splicing effects were not identified in the 4 SNPs where the information change was < 1 bit (< 2 -fold). Genetic variability masked potential splicing effects of three of these SNPs, including rs16802, rs2252576, and rs8130564 (**Table 1**). PCR primer sets designed for *COL6A2* exon 21 (affected by rs17357592) did not produce the expected amplicon. Interpreting the results for rs16994182 (*CLDN14*) was complicated by the lack of a suitable internal reference. As *CLDN14* consists of three exons, any internal reference covering the affected second exon cannot parse whether differences in exon 2 expression were caused by the SNP or by general expression changes.

The T-allele rs2838010 was predicted to activate a donor splice site of a rare exon in IVS1 of *FAM3B* (GenBank Accession AJ409094). The cryptic pseudoexon was neither detected by RT-PCR nor expression microarray (**Supplementary Image 1.16**). Interestingly, this exon is expressed in a malignant lymphoma patient who is a carrier for this genotype [ICGC ID: DO27769; (**Supplementary Image 1.16C**)]. Although the T-allele is probably required to activate the pseudoexon, additional unknown splicing-related factors appear to be necessary.

DISCUSSION

Predicted SNP alleles that alter constitutive mRNA splicing are confirmed by expression data, and appear to be a common cause of alternative splicing. The preponderance of leaky splicing mutations and cryptic splice sites, which often produce both normal and mutant transcripts, is consistent with balancing selection (Nuzhdin et al., 2004) or possibly with mutant loci that contribute to multifactorial disease. Minor SNP alleles are often found in $> 1\%$ of populations (Janosiková et al., 2005). This would be consistent with a bias against finding mutations that abolish splice site recognition in dbSNP. Such mutations are more typical in rare Mendelian disorders (Rogan et al., 1998).

Exon-based expression microarrays and q-RT-PCR techniques were initially used to confirm the predicted impact of common and rare SNPs on splicing. Results were subsequently confirmed using RNAseq data for some of these SNPs (Dorman et al., 2014; Viner et al., 2014; Shirley et al., 2019). However, exon skipping due to rs1893592 was not consistently seen in all carriers. Although detected only in one type of tumor, this event may not be tissue specific, since five patients with the same genotype did not exhibit this isoform. Nevertheless, exon skipping was also observed in malignant lymphoma (**Supplementary Image 1.10D**). Intron retention in rs1805377 carriers was evident in only 22% of tumors. Increased total intron retention may be due to failure to recognize exons due to overlapping strong splice sites (Vockley et al., 2000; Rogan et al., 2003).

The splicing impacts of several of these SNPs have also been implicated in other studies. rs10190751 modulates the FLICE-inhibitory protein (c-FLIP) from its S-form to its R-form, with the latter having been linked to increased lymphoma risk (Ueffing et al., 2009). We observed the R-form to be twice as abundant for one of the rs10190751 alleles. Increased exon skipping attributed to rs1333973 has been reported in RNAseq analysis of *IFI44L* (Zhao et al., 2013a), which has been implicated in reduced antibody response to measles vaccine (Haralambieva et al., 2017). The splicing impact of *XRCC4* rs1805377 has been noted previously (Nalla and Rogan, 2005). This SNP has been implicated with an increased risk of gastric cancer (Chiu et al., 2008), pancreatic cancer (Ding and Li, 2015) and glioma (Zhao et al., 2013b). Similarly, the potential impact of rs1893592 in *UBASH3A* has been recognized (Kim et al., 2015) and is associated with arthritis (Liu et al., 2017) and type 1 diabetes (Ge and Concannon, 2018). Hiller et al. (2006) described the 3nt deletion caused by rs2243187 in *IL19* but did not report increased exon skipping.

rs743920 was associated with change in *EMID1* expression (Ge et al., 2005), however its splicing impact was not recognized. Conversely, studies linking *TMPRSS3* variants to hearing loss did not report rs8130564 to be significant (Lee et al., 2013; Chung et al., 2014). Interestingly, rs2252576 (in which we did not find a splicing alteration) has been associated to Alzheimer's dementia in Down syndrome (Mok et al., 2014).

rs2835585 significantly increased exon skipping in *TTC3*, however normal expression levels at the affected exon junction were not significantly altered. This was most likely due to the large difference in abundance between the constitutive and skipped splice isoforms (**Table 2**). The skipped isoform does not disrupt the reading frame and the affected coding region has not been assigned to any known protein domain (Tsukahara et al., 1996; Suizu et al., 2009). It is unclear whether allele-specific, exon skipping in this instance would impact *TTC3* protein function or activity.

Why are so few natural splice sites strengthened by SNP-induced information changes? Most such changes would be thought to be neutral mutations, which are ultimately lost by chance (Fisher, 1930). Those variants which are retained are more likely to confer a selective advantage (Li, 1967). Indeed, the minor allele in rs2266988, which strengthens a donor splice site by 1.6 bits at the 5' end of the open reading frame in *PRAME* and occurs in 25% of the overall population (~50% in Europeans). Several instances of modest changes in splice site strength that would be expected to have little or no impact, in fact, alter the degree of exon skipping.

Allele frequency can significantly vary across different populations, which can be indicative of gene flow and migration of a population (Cavalli-Sforza and Bodmer, 1971) as well as, in the case of splicing variants, genetic load and fitness in a population (Rogan and Mucaki, 2011). The frequencies of several of the variants presented here are significantly different between ethnic and geographically defined populations (**Tables 1** and **3**). We examined allele frequencies of these variants in sub-populations in both HapMap and dbSNP version 153. For example, representation of the alleles of the *XRCC4* variant rs1805377 (where its A-allele leads to a 6 nt deletion of the gene's terminal exon) differs between Caucasians and Asians (for the G- and A-alleles, respectively). Different linkage disequilibrium patterns of this variant occur in Han Chinese (CHB) and Utah residents with Northern and Western European ancestry (CEU) populations (Zhao et al., 2013b). Similar differences in SNP population frequency include *EMID1* rs743920 (in HapMap: G-allele frequency is 47% in CHB, but only 7% in CEU and 10% in northern Swedish cohorts). This is consistent with dbSNP (version 153) where it is present in 72% in Vietnamese, but only 16% of a northern Sweden cohort). In *BACE2*, rs2252576, the T-allele is most prevalent at 84% in Yoruba in Ibadan, Nigeria populations (YRI), but only 8% in CHB. In *FCHSD1*, rs469074 the frequency of the G-allele is 37% in YRI and <1% in CHB. Some SNPs were exclusively present in a single population in the HapMap cohort (e.g. only the YRI population is polymorphic for *IL19* rs2243187, *WBP2NL* rs17002806, *DERL3* rs6003906, and *CLDN14* rs16994182).

Because of their effects on mRNA splicing, these differences in allele frequency would be expected to alter the relative abundance of certain

protein isoforms in these populations. We speculate about whether isoform-specific representation among populations influences disease predisposition, other common phenotypic differences, or whether they are neutral. We suggest that SNPs decreasing constitutive splicing while increasing mRNA isoforms which alter the reading frame would be more likely to result in a distinct phenotype. q-RT-PCR experiments confirmed five SNPs which increased the fraction of mRNA splice forms causing a frameshift (**Table 1**), three of which simultaneously decrease constitutive splicing by ≥ 10 -fold (*WBP2NL* rs17002806; *GUSBP11* rs3747107; and *IFI44L* rs1333973). rs3747107 and rs17002806 are much more common in YRI populations in HapMap (rs3747107 G-allele is present in 64% in YRI but only 23% in CEU; rs17002806 A-allele was not identified in any CHB or CEU individuals), while the A-allele of rs1333973 is much more common in CHB (76%, compared to 31% and 35% in CEU and YRI populations, respectively). These common variants are likely to change the function of these proteins and may influence individual phenotypes. A somewhat comprehensive catalog of DNA polymorphisms with splicing effects—confined or with increased prevalence in specific ethnic or geographically identifiable groups—could be derived from combining ValidSpliceMut with population-specific SNP databases. Aside from those phenotypes described earlier, genes implicated by GWAS or other analyses for specific disorders represent reasonable candidates for further detailed or replication studies aimed at identification of the risk alleles in these cohorts.

The extent to which SNP-related sequence variation accounts for the heterogeneity in mRNA transcript structures has been somewhat unappreciated, given the relatively high proportion of genes that exhibit tissue-specific alternative splicing (Pan et al., 2008; Wang et al., 2008; Baralle and Giudice, 2017). This and our previous study (Shirley et al., 2019) raise questions regarding the degree to which apparent alternative splicing is the result of genomic polymorphism rather than splicing regulation alone. Because much of the information required for splice site recognition resides within neighboring introns, it would be prudent to consider contributions from intronic and exonic polymorphism that produce structural mRNA variation, since these changes might be associated with disease or predisposition.

Individual information corresponds to a continuous molecular phenotypic measure that is well suited to the analysis of contributions of multiple, incompletely penetrant SNPs in different genes, as typically seen in genetically complex diseases (Cooper et al., 2013). Our protocol identifies low or nonpenetrant allele-specific alternative splicing events through bioinformatic analysis, and either q-RT-PCR, exon microarrays or RNAseq data analysis. Allele-specific splicing can also be determined by full-length alternative isoform analysis of RNA [or FLAIR (Workman et al., 2019)]. Differentiated splice forms are associated with specific alleles in heterozygotes with exonic SNPs. However, combining genome-based information with FLAIR may enable identification of intronic SNPs influencing splicing and low abundance alternative splice forms, which might

otherwise be missed by FLAIR.

Targeted splicing analysis generally reproduces the results of our multi-genome-wide surveys of sequence variations affecting mRNA splicing. As splicing mutations and their effects were often observed in multiple tumor types, the impact of these mutations may be pleiotropic. Some events were only detected in q-RT-PCR data and not by RNAseq (and vice versa), highlighting the complementarity of these techniques for splicing mutation analyses. Results of this study increase confidence that the publicly (<https://ValidSpliceMut.cytognomix.com>) and commercially (<https://MutationForecaster.com>) available resources for information-theory based variant analysis and validation can distinguish mutations contributing to aberrant molecular phenotypes from allele-specific alternative splicing.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

Controlled-access TCGA and ICGC sequence data was approved by NCBI at the US National Institutes of Health (dbGaP Project #988: “Predicting common genetic variants that alter the splicing of human gene transcripts”; Approval Number #13930-11; PI: PK Rogan) and by the International Cancer Genome Consortium (ICGC Project #DACO-1056047; “Validation of mutations that alter gene expression”).

REFERENCES

- Ars, E., Kruyer, H., Morell, M., Pros, E., Serra, E., Ravello, A., et al. (2003). Recurrent mutations in the NF1 gene are common among neurofibromatosis type 1 patients. *J. Med. Genet.* 40 (6), e82. doi: 10.1136/jmg.40.6.e82
- 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
- Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270 (6), 2411–2414. doi: 10.1074/jbc.270.6.2411
- Bi, C., and Rogan, P. K. (2004). Bipartite pattern discovery by entropy minimization-based multiple local alignment. *Nucleic Acids Res.* 32 (17), 4979–4991. doi: 10.1093/nar/gkh825
- Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* 21 (25), 5803–5816. doi: 10.1093/nar/21.25.5803
- Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336. doi: 10.1146/annurev.biochem.72.121801.161720
- Buratti, E., Baralle, M., and Baralle, F. E. (2006). Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res.* 34 (12), 3494–3510. doi: 10.1093/nar/gkl498
- Cáceres, J. F., Stamm, S., Helfman, D. M., and Krainer, A. R. (1994). Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* 265 (5179), 1706–1709. doi: 10.1126/science.8085156
- Caminsky, N., Mucaki, E. J., and Rogan, P. K. (2014). Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *Fl000Res* 3, 282. doi: 10.12688/fl000research.5654.1

AUTHOR CONTRIBUTIONS

EM designed and performed all q-RT-PCR experiments, processed and analyzed publicly exon microarray data, and performed all formal analysis. BS performed data curation and software development. PR conceptualized the project and was the project administrator. EM and PR prepared the original draft of the manuscript, while EM, BS, and PR reviewed and edited the document.

ACKNOWLEDGMENTS

PR acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) [371758-09 and RGPIN-2015-06290], Canadian Foundation for Innovation, Canada Research Chairs, and CytoGnomix. Cell lines were obtained from the NIGMS Human Genetic Cell Repository under a Materials Transfer Agreement with the Coriell Institute (Camden, NJ). An earlier version of this article is available from bioRxiv: <https://www.biorxiv.org/content/10.1101/549089v2> (Mucaki and Rogan, 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00109/full#supplementary-material>

- Caminsky, N. G., Mucaki, E. J., Perri, A. M., Lu, R., Knoll, J. H., and Rogan, P. K. (2016). Prioritizing variants in complete hereditary breast and ovarian cancer genes in patients lacking known BRCA mutations. *Hum. Mutat.* 37 (7), 640–652. doi: 10.1002/humu.22972
- Carothers, A. M., Urlaub, G., Grunberger, D., and Chasin, L. A. (1993). Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell Biol.* 13 (8), 5085–5098. doi: 10.1128/MCB.13.8.5085
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3 (4), 285–298. doi: 10.1038/nrg775
- Caux-Moncoutier, V., Pagès-Berhouet, S., Michaux, D., Asselain, B., Castéra, L., De Pauw, A., et al. (2009). Impact of BRCA1 and BRCA2 variants on splicing: clues from an allelic imbalance study. *Eur. J. Hum. Genet.* 17 (11), 1471–1480. doi: 10.1038/ejhg.2009.89
- Cavalli-Sforza, L. L., and Bodmer, W. F. (1971). *The Genetics of Human Populations* (San Francisco: W.H. Freeman and Company).
- Chiu, C. F., Wang, C. H., Wang, C. L., Lin, C. C., Hsu, N. Y., Weng, J. R., et al. (2008). A novel single nucleotide polymorphism in XRCC4 gene is associated with gastric cancer susceptibility in Taiwan. *Ann. Surg. Oncol.* 15 (2), 514–518. doi: 10.1245/s10434-007-9674-3
- Chung, J., Park, S. M., Chang, S. O., Chung, T., Lee, K. Y., Kim, A. R., et al. (2014). A novel mutation of TMPRSS3 related to milder auditory phenotype in Korean postlingual deafness: a possible future implication for a personalized auditory rehabilitation. *J. Mol. Med. (Berl)* 92 (6), 651–663. doi: 10.1007/s00109-014-1128-3
- Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., et al. (2013). Comparative *in vitro* and *in silico* analyses of variants in splicing

- regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. *PLoS One* 8 (2), e57173. doi: 10.1371/journal.pone.0057173
- Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* 132(10), 1077–1130. doi: 10.1007/s00439-013-1331-2
- Dietz, H. C., Valle, D., Francomano, C. A., Kendzior, R. J. Jr, Pyeritz, R. E., and Cutting, G. R. (1993). The skipping of constitutive exons *in vivo* induced by nonsense mutations. *Science* 259 (5095), 680–683. doi: 10.1126/science.8430317
- Ding, Y., and Li, L. N. (2015). Association between single nucleotide polymorphisms of X-ray repair cross-complementing protein 4 gene and development of pancreatic cancer. *Genet. Mol. Res.* 14 (3), 9626–9632. doi: 10.4238/2015.August.14.25
- Domenjoud, L., Kister, L., Gallinaro, H., and Jacob, M. (1993). Selection between a natural and a cryptic 5' splice site: a kinetic study of the effect of upstream exon sequences. *Gene Expr.* 3 (1), 83–94.
- Dorman, S. N., Viner, C., and Rogan, P. K. (2014). Splicing mutation analysis reveals previously unrecognized pathways in lymph node-invasive breast cancer. *Sci. Rep.* 4, 7063. doi: 10.1038/srep07063
- Eliseos, E. J., Shirley, B. S., and Rogan, P. K. (2019). Expression changes confirm predicted single nucleotide variants affecting mRNA splicing. *bioRxiv*, 549089. doi: 10.1101/549089
- Fisher, R. A. (1930). *The genetical theory of natural selection* (Oxford New York: Clarendon Press).
- Fu, X. D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc. Natl. Acad. Sci. U.S.A.* 89 (5), 1725–1729. doi: 10.1073/pnas.89.5.1725
- Ge, Y., and Concannon, P. (2018). Molecular-genetic characterization of common, noncoding UBASH3A variants associated with type 1 diabetes. *Eur. J. Hum. Genet.* 26 (7), 1060–1064. doi: 10.1038/s41431-018-0123-5
- Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., et al. (2005). Survey of allelic expression using EST mining. *Genome Res.* 15 (11), 1584–1591. doi: 10.1101/gr.4023805
- Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15 (10), 1451–1455. doi: 10.1101/gr.4086505
- Gloss, B. S., and Dinger, M. E. (2018). Realizing the significance of noncoding functionality in clinical genomics. *Exp. Mol. Med.* 50 (8), 97. doi: 10.1038/s12276-018-0087-0
- Haralambieva, I. H., Ovsyannikova, I. G., Kennedy, R. B., Larrabee, B. R., Zimmermann, M. T., Grill, D. E., et al. (2017). Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine. *Hum. Genet.* 136 (4), 421–435. doi: 10.1007/s00439-017-1768-9
- Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., et al. (2006). Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *Am. J. Hum. Genet.* 78 (2), 291–302. doi: 10.1086/500151
- Hobson, G. M., Huang, Z., Sperle, K., Sistermans, E., Rogan, P. K., Garbern, J. Y., et al. (2006). Splice-site contribution in alternative splicing of PLP1 and DM20: molecular studies in oligodendrocytes. *Hum. Mutat.* 27 (1), 69–77. doi: 10.1002/humu.20276
- Huang, R. S., Duan, S., Bleibel, W. K., Kistner, E. O., Zhang, W., Clark, T. A., et al. (2007). A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc. Natl. Acad. Sci. U. S. A.* 104 (23), 9758–9763. doi: 10.1073/pnas.0703736104
- Hull, J., Campino, S., Rowlands, K., Chan, M. S., Copley, R. R., Taylor, M. S., et al. (2007). Identification of common genetic variation that modulates alternative splicing. *PLoS Genet.* 3 (6), e99. doi: 10.1371/journal.pgen.0030099
- Janosíková, B., Zavadáková, P., and Kozich, V. (2005). Single-nucleotide polymorphisms in genes relating to homocysteine metabolism: how applicable are public SNP databases to a typical European population? *Eur. J. Hum. Genet.* 13 (1), 86–95. doi: 10.1038/sj.ejhg.5201282
- Kannabiran, C., Rogan, P. K., Olmos, L., Basti, S., Rao, G. N., Kaiser-Kupfer, M., et al. (1998). Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene. *Mol. Vis.* 4, 21.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12(6), 996–1006. doi: 10.1101/gr.229102
- Khan, S. G., Metin, A., Gozukara, E., Inui, H., Shahavi, T., Muniz-Medina, V., et al. (2004). Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum. Mol. Genet.* 13 (3), 343–352. doi: 10.1093/hmg/ddh026
- Kim, K., Bang, S. Y., Lee, H. S., Cho, S. K., Choi, C. B., Sung, Y. K., et al. (2015). High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Ann. Rheumatol. Dis.* 74 (3), e13. doi: 10.1136/annrheumdis-2013-204749
- Krawczak, M., Reiss, J., and Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* 90 (1-2), 41–54. doi: 10.1007/BF00210743
- López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579 (9), 1900–1903. doi: 10.1016/j.febslet.2005.02.047
- Lacroix, M., Lacaze-Buzy, L., Furio, L., Tron, E., Valari, M., Van der Wier, G., et al. (2012). Clinical expression and new SPINK5 splicing defects in Netherton syndrome: unmasking a frequent founder synonymous mutation and unconventional intronic mutations. *J. Invest. Dermatol.* 132 (3 Pt 1), 575–582. doi: 10.1038/jid.2011.366
- Lamba, V., Lamba, J., Yasuda, K., Strom, S., Davila, J., Hancock, M. L., et al. (2003). Hepatic CYP2B6 expression: gender and ethnic differences and relationship to CYP2B6 genotype and CAR (constitutive androstane receptor) expression. *J. Pharmacol. Exp. Ther.* 307 (3), 906–922. doi: 10.1124/jpet.103.054866
- Lee, J., Baek, J. I., Choi, J. Y., Kim, U. K., Lee, S. H., and Lee, K. Y. (2013). Genetic analysis of TMRSS3 gene in the Korean population with autosomal recessive nonsyndromic hearing loss. *Gene* 532 (2), 276–280. doi: 10.1016/j.gene.2013.07.108
- Li, C., Kato, M., Shiue, L., Shively, J. E., Ares, M., and Lin, R. J. (2006). Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. *Cancer Res.* 66 (4), 1990–1999. doi: 10.1158/0008-5472.CAN-05-2593
- Li, C. C. (1967). Genetic equilibrium under selection. *Biometrics* 23, 397–484. doi: 10.2307/2528009
- Liu, D., Liu, J., Cui, G., Yang, H., Cao, T., and Wang, L. (2017). Evaluation of the association of UBASH3A and SYNGR1 with rheumatoid arthritis and disease activity and severity in Han Chinese. *Oncotarget* 8 (61), 103385–103392. doi: 10.18632/oncotarget.21875
- Lu, R., and Rogan, P. K. (2019). Transcription factor binding site clusters identify target genes with similar tissue-wide expression and buffer against mutations [version 2; peer review: 2 approved]. *F1000Research* 7, 1933. doi: 10.12688/f1000research.17363.2
- Lu, R., Mucaki, E. J., and Rogan, P. K. (2017). Discovery and validation of information theory-based transcription factor and cofactor binding site motifs. *Nucleic Acids Res.* 45 (5), e27. doi: 10.1093/nar/gkw1036
- Margulis, V., Lin, J., Yang, H., Wang, W., Wood, C. G., and Wu, X. (2008). Genetic susceptibility to renal cell carcinoma: the role of DNA double-strand break repair pathway. *Cancer Epidemiol. Biomarkers Prev.* 17 (9), 2366–2373. doi: 10.1158/1055-9965.EPI-08-0259
- Martínez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J. F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* 623, 123–147. doi: 10.1007/978-0-387-77374-2_8
- Mayeda, A., and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68 (2), 365–375. doi: 10.1016/0092-8674(92)90477-T
- Mok, K. Y., Jones, E. L., Hanney, M., Harold, D., Sims, R., Williams, J., et al. (2014). Polymorphisms in BACE2 may affect the age of onset Alzheimer's dementia in Down syndrome. *Neurobiol. Aging* 35 (6), 1513.e1–5. doi: 10.1016/j.neurobiolaging.2013.12.022
- Moore, M. J., and Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* 365 (6444), 364–368. doi: 10.1038/365364a0
- Mucaki, E. J., Shirley, B. C., and Rogan, P. K. (2013). Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum. Mutat.* 34 (4), 557–565. doi: 10.1002/humu.22277
- Mucaki, E. J., Caminsky, N. G., Perri, A. M., Lu, R., Laederach, A., Halvorsen, M., et al. (2016). A unified analytic framework for prioritization of non-coding variants of uncertain significance in heritable breast and ovarian cancer. *BMC Med. Genomics* 9, 19. doi: 10.1186/s12920-016-0178-5

- Nalla, V. K., and Rogan, P. K. (2005). Automated splicing mutation analysis by information theory. *Hum. Mutat.* 25 (4), 334–342. doi: 10.1002/humu.20151
- Nelson, K. K., and Green, M. R. (1990). Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* 87 (16), 6253–6257. doi: 10.1073/pnas.87.16.6253
- Nembaware, V., Lupindo, B., Schouet, K., Spillane, C., Scheffler, K., and Seoighe, C. (2008). Genome-wide survey of allele-specific splicing in humans. *BMC Genomics* 9, 265. doi: 10.1186/1471-2164-9-265
- Nuzhdin, S. V., Wayne, M. L., Harmon, K. L., and McIntyre, L. M. (2004). Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* 21 (7), 1308–1317. doi: 10.1093/molbev/msh128
- Olsen, R. K., Brøner, S., Sabaratnam, R., Doktor, T. K., Andersen, H. S., Bruun, G. H., et al. (2014). The ETFDH c.158A>G variation disrupts the balanced interplay of ESE- and ESS-binding proteins thereby causing missplicing and multiple Acyl-CoA dehydrogenation deficiency. *Hum. Mutat.* 35 (1), 86–95. doi: 10.1002/humu.22455
- 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
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* 102 (1), 11–26. doi: 10.1016/j.ajhg.2017.11.002
- Parker, R., Siliciano, P. G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49 (2), 229–239. doi: 10.1016/0092-8674(87)90564-2
- Peterlongo, P., Catucci, I., Colombo, M., Caleca, L., Mucaki, E., Bogliolo, M., et al. (2015). FANCM c.5791C > T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum. Mol. Genet.* 24 (18), 5345–5355. doi: 10.1093/hmg/ddv251
- Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for groupwise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30 (9), e36. doi: 10.1093/nar/30.9.e36
- Richard, I., and Beckmann, J. S. (1995). How neutral are synonymous codon mutations? *Nat. Genet.* 10 (3), 259. doi: 10.1038/ng0795-259
- Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell Biol.* 10 (1), 84–94. doi: 10.1128/MCB.10.1.84
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Res.* 77 (21), e31–e34. doi: 10.1158/0008-5472.CAN-17-0337
- Rogan, P. K., and Mucaki, E. J. (2011). Population fitness and genetic load of single nucleotide polymorphisms affecting mRNA splicing. arXiv:1107.0716 [q-bio.GN].
- Rogan, P. K., and Schneider, T. D. (1995). Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum. Mutat.* 6 (1), 74–76. doi: 10.1002/humu.1380060114
- Rogan, P. K., Faux, B. M., and Schneider, T. D. (1998). Information analysis of human splice site mutations. *Hum. Mutat.* 12, 153–171. doi: 10.1002/(SICI)1098-1004(1998)12:3<153::AID-HUMU3>3.0.CO;2-I
- Rogan, P. K., Svojanovsky, S. R., and Leeder, J. S. (2003). Information theory-based analysis of CYP219, CYP2D6 and CYP3A5 splicing mutations. *Pharmacogenetics* 13 (4), 207–218. doi: 10.1097/00008571-200304000-00005
- Séraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J.* 7 (8), 2533–2538. doi: 10.1002/j.1460-2075.1988.tb03101.x
- Schneider, T. D. (1997). Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res.* 25, 4408–4415. doi: 10.1093/nar/25.21.4408
- Shirley, B. C., Mucaki, E. J., Whitehead, T., Costea, P. I., Akan, P., and Rogan, P. K. (2013). Interpretation, stratification and evidence for sequence variants affecting mRNA splicing in complete human genome sequences. *Genomics Proteomics Bioinf.* 11 (2), 77–85. doi: 10.1016/j.gpb.2013.01.008
- Shirley, B. C., Mucaki, E. J., and Rogan, P. K. (2019). Pan-cancer repository of validated natural and cryptic mRNA splicing mutations [version 3; peer review: 1 approved, 2 approved with reservations]. *F1000Research* 7, 1908. doi: 10.12688/f1000research.17204.3
- Soemedi, R., Cygan, K. J., Rhine, C. L., Wang, J., Bulacan, C., Yang, J., et al. (2017). Pathogenic variants that alter protein code often disrupt splicing. *Nat. Genet.* 49 (6), 848–855. doi: 10.1038/ng.3837
- Spielmann, M., and Mundlos, S. (2016). Looking beyond the genes: the role of non-coding variants in human disease. *Hum. Mol. Genet.* 25 (R2), R157–R165. doi: 10.1093/hmg/ddw205
- Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell Biol.* 14 (11), 7670–7682. doi: 10.1128/MCB.14.11.7670
- Suizu, F., Hiramaki, Y., Okumura, F., Matsuda, M., Okumura, A. J., Hirata, N., et al. (2009). The E3 ligase TTC3 facilitates ubiquitination and degradation of phosphorylated Akt. *Dev. Cell.* 17 (6), 800–810. doi: 10.1016/j.devcel.2009.09.007
- Susani, L., Pangrazio, A., Sobacchi, C., Taranta, A., Mortier, G., Savarirayan, R., et al. (2004). TCIRG1-dependent recessive osteopetrosis: mutation analysis, functional identification of the splicing defects, and *in vitro* rescue by U1 snRNA. *Hum. Mutat.* 24 (3), 225–235. doi: 10.1002/humu.20076
- Svojanovsky, S. R., Schneider, T. D., and Rogan, P. K. (2000). Redundant designations of BRCA1 intron 11 splicing mutation; c. 4216-2A > G; IVS11-2A > G; L78833, 37698, A > G. *Hum. Mutat.* 16 (3), 264. doi: 10.1002/1098-1004(200009)16:3<264::AID-HUMU9>3.0.CO;2-I
- Szafarski, K., Fritsch, C., Schumann, F., Siebel, L., Sinha, R., Hampe, J., et al. (2014). Physiological state co-regulates thousands of mammalian mRNA splicing events at tandem splice sites and alternative exons. *Nucleic Acids Res.* 42 (14), 8895–8904. doi: 10.1093/nar/gku532
- Talerico, M., and Berget, S. M. (1990). Effect of 5' splice site mutations on splicing of the preceding intron. *Mol. Cell Biol.* 10 (12), 6299–6305. doi: 10.1128/MCB.10.12.6299
- Teraoka, S. N., Telatar, M., Becker-Catania, S., Liang, T., Onengüt, S., Tolun, A., et al. (1999). Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am. J. Hum. Genet.* 64 (6), 1617–1631. doi: 10.1086/302418
- Thorsen, K., Sørensen, K. D., Brems-Eskildsen, A. S., Modin, C., Gaustadnes, M., Hein, A. M., et al. (2008). Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol. Cell Proteomics* 7 (7), 1214–1224. doi: 10.1074/mcp.M700590-MCP200
- Tsukahara, F., Hattori, M., Muraki, T., and Sakaki, Y. (1996). Identification and cloning of a novel cDNA belonging to tetratricopeptide repeat gene family from Down syndrome-critical region 21q22.2. *J. Biochem.* 120 (4), 820–827. doi: 10.1093/oxfordjournals.jbchem.a021485
- Ueffing, N., Singh, K. K., Christians, A., Thorns, C., Feller, A. C., Nagl, F., et al. (2009). A single nucleotide polymorphism determines protein isoform production of the human c-FLIP protein. *Blood* 114 (3), 572–579. doi: 10.1182/blood-2009-02-204230
- Vemula, S. R., Xiao, J., Zhao, Y., Bastian, R. W., Perlmutter, J. S., Racette, B. A., et al. (2014). A rare sequence variant in intron 1 of THAP1 is associated with primary dystonia. *Mol. Genet. Genom. Med.* 2 (3), 261–272. doi: 10.1002/mgg.3.67
- Viner, C., Dorman, S. N., Shirley, B. C., and Rogan, P. K. (2014). Validation of predicted mRNA splicing mutations using high-throughput transcriptome data [version 2; referees: 4 approved]. *F1000Res* 3, 8. doi: 10.12688/f1000research.3-8.v2
- Vockley, J., Rogan, P. K., Anderson, B. D., Willard, J., Seelan, R. S., Smith, D. I., et al. (2000). Exon skipping in IVD RNA processing in isovaleric acidemia caused by point mutations in the coding region of the IVD gene. *Am. J. Hum. Genet.* 66 (2), 356–367. doi: 10.1086/302751
- von Kodolitsch, Y., Pyeritz, R. E., and Rogan, P. K. (1999). Splice-site mutations in atherosclerosis candidate genes: relating individual information to phenotype. *Circulation* 100 (7), 693–699. doi: 10.1161/01.CIR.100.7.693
- von Kodolitsch, Y., Berger, J., and Rogan, P. K. (2006). Predicting severity of haemophilia A and B splicing mutations by information analysis. *Haemophilia* 12 (3), 258–262. doi: 10.1111/j.1365-2516.2006.01216.x
- Vyhldal, C. A., Rogan, P. K., and Leeder, J. S. (2004). Development and refinement of pregnane X receptor (PXR) DNA binding site model using information theory: insights into PXR-mediated gene regulation. *J. Biol. Chem.* 279 (45), 46779–46786. doi: 10.1074/jbc.M408395200
- 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 (7221), 470–476. doi: 10.1038/nature07509

- Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., et al. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods*. doi: 10.1038/s41592-019-0617-2
- Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.* 3 (10), 1553–1561. doi: 10.1101/gad.3.10.1553
- Wu, X., Gu, J., Grossman, H. B., Amos, C. I., Etzel, C., Huang, M., et al. (2006). Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am. J. Hum. Genet.* 78 (3), 464–479. doi: 10.1086/500848
- Yates, T., Okoniewski, M. J., and Miller, C. J. (2008). X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Res.* 36 (Database issue), D780–D786. doi: 10.1093/nar/gkm779
- Zhao, L., Lu, Z., Park, J. W., Zhou, Q., and Xing, Y. (2013a). GLiMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biol.* 14 (7), R74. doi: 10.1186/gb-2013-14-7-r74
- Zhao, P., Zou, P., Zhao, L., Yan, W., Kang, C., Jiang, T., et al. (2013b). Genetic polymorphisms of DNA double-strand break repair pathway genes and glioma susceptibility. *BMC Cancer* 13, 234. doi: 10.1186/1471-2407-13-234
- Zhuang, Y., and Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46 (6), 827–835. doi: 10.1016/0092-8674(86)90064-4
- Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* 10 (11), 1356–1368. doi: 10.1101/gad.10.11.1356

Conflict of Interest: PR founded and BS is an employee of CytoGnomix. The company holds intellectual property related to information theory-based mutation analysis and validation.

EM declares 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 © 2020 Mucaki, Shirley and Rogan. 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.



Dynamic Supraspliceosomes Are Assembled on Different Transcripts Regardless of Their Intron Number and Splicing State

Naama Sebbag-Sznajder¹, Yehuda Brody², Hodaya Hochberg-Laufer², Yaron Shav-Tal², Joseph Sperling³ and Ruth Sperling^{1*}

¹ Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel, ² The Mina and Everard Goodman Faculty of Life Sciences and The Institute of Nanotechnology and Advanced Materials, Bar Ilan University, Ramat Gan, Israel,

³ Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot, Israel

OPEN ACCESS

Edited by:

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

Reviewed by:

Maurizio Romano,
University of Trieste, Italy
Xuexiu Zheng,
Gwangju Institute of Science
and Technology, South Korea

*Correspondence:

Ruth Sperling
r.sperling@mail.huji.ac.il

Specialty section:

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

Received: 16 September 2019

Accepted: 31 March 2020

Published: 15 May 2020

Citation:

Sebbag-Sznajder N, Brody Y,
Hochberg-Laufer H, Shav-Tal Y,
Sperling J and Sperling R (2020)
Dynamic Supraspliceosomes Are
Assembled on Different Transcripts
Regardless of Their Intron Number
and Splicing State.
Front. Genet. 11:409.
doi: 10.3389/fgene.2020.00409

Splicing and alternative splicing of pre-mRNA are key sources in the formation of diversity in the human proteome. These processes have a central role in the regulation of the gene expression pathway. Yet, how spliceosomes are assembled on a multi-intronic pre-mRNA is at present not well understood. To study the spliceosomes assembled *in vivo* on transcripts with variable number of introns, we examined a series of three related transcripts derived from the β -globin gene, where two transcript types contained increasing number of introns, while one had only an exon. Each transcript had multiple MS2 sequence repeats that can be bound by the MS2 coat protein. Using our protocol for isolation of endogenous spliceosomes under native conditions from cell nuclei, we show that all three transcripts are found in supraspliceosomes – 21 MDa dynamic complexes, sedimenting at 200S in glycerol gradients, and composed of four native spliceosomes connected by the transcript. Affinity purification of complexes assembled on the transcript with most introns (termed E6), using the MS2 tag, confirmed the assembly of E6 in supraspliceosomes with components such as Sm proteins and PSF. Furthermore, splicing inhibition by spliceostatin A did not inhibit the assembly of supraspliceosomes on the E6 transcript, yet increased the percentage of E6 pre-mRNA supraspliceosomes. These findings were corroborated in intact cells, using RNA FISH to detect the MS2-tagged E6 mRNA, together with GFP-tagged splicing factors, showing the assembly of splicing factors SRSF2, U1-70K, and PRP8 onto the E6 transcripts under normal conditions and also when splicing was inhibited. This study shows that different transcripts with different number of introns, or lacking an intron, are assembled in supraspliceosomes even when splicing is inhibited. This assembly starts at the site of transcription and can continue during the life of the transcript in the nucleoplasm. This study further confirms the dynamic and universal nature of supraspliceosomes that package RNA polymerase II transcribed pre-mRNAs into complexes composed of four native spliceosomes connected by the transcript, independent of their length, number of introns, or splicing state.

Keywords: pre-mRNA splicing, specific supraspliceosomes, MS2-tagged supraspliceosomes, splicing inhibition, splicing factors

INTRODUCTION

To generate an mRNA, RNA polymerase II (Pol II) transcribed pre-mRNAs must go through nuclear processing events prior to their export into the cytoplasm. RNA processing is composed of 5'-end capping, 3'-end processing, splicing, and RNA editing. Pre-mRNA splicing takes place in a dynamic ribonucleoprotein complex (RNP) – the spliceosome. The splicing machinery engages with *cis* elements in the pre-mRNA such as the 5' and 3' splice sites (SSs) consensus sequences, a branch site, a polypyrimidine tract, and exonic and intronic splicing enhancers and silencers (reviewed in Wahl et al., 2009; Will and Luhrmann, 2011; Papasaikas and Valcarcel, 2016). The *cis* elements in the pre-mRNA are identified by *trans* factors, such as the U1, U2, U4, U5, and U6 snRNPs, and many splicing factors, including the hnRNP proteins and the serine/arginine (SR)-rich protein family.

The splicing reaction is a two-step transesterification process that is performed by the spliceosome. Spliceosome assembly can be monitored *in vitro*, showing that this is a process occurring in a stepwise manner, generating intermediate complexes (reviewed in Brow, 2002; Will and Luhrmann, 2011). The spliceosomal U snRNPs, which are key players in pre-mRNA splicing, go through major dynamic alterations in their RNA:RNA contacts during the assembly of the spliceosome and the splicing reaction. The first step is the base pairing of U1 snRNP with the 5' splice site and is followed by the assembly of additional snRNPs. The U snRNPs also interact with numerous splicing factors during the assembly of the spliceosome and the splicing reaction. The interaction of snRNPs with the pre-mRNA is supported by proteins from the SR protein family. These are SR-rich proteins (Valcarcel and Green, 1996; Shepard and Hertel, 2009), and they are required for the stabilization of the early spliceosomal complex. For instance, the interaction of SRSF2 with U1 snRNP is assisted by the U1 snRNP protein U1-70K (Wu and Maniatis, 1993; Zahler and Roth, 1995) and can determine transcript fate (Fu et al., 1992). Recent subnanometric structures of splicing complexes determined by high-resolution cryo-EM have portrayed the catalytic center of the spliceosome and have revealed the dynamic alterations in U snRNA:U snRNA and U snRNA:pre-mRNA interactions taking place during the assembly of the spliceosomes and the splicing reaction, which is reflected in alterations in the structures of spliceosome intermediates. A key protein, present at the heart of the spliceosome, is the U5 snRNP protein PRP8 (reviewed in Shi, 2017a,b; Fica et al., 2017; Wilkinson et al., 2018; Plaschka et al., 2019; Yan et al., 2019).

The majority of Pol II transcribed pre-mRNAs have multiple introns, and they can thus undergo alternative splicing (AS), which is a key element in the regulation of gene expression (reviewed in Kelemen et al., 2013; Akerman et al., 2015; Lee and Rio, 2015; Naftelberg et al., 2015). Furthermore, errors in AS are at the heart of numerous human diseases, as well as in cancer (Irimia and Blencowe, 2012; Singh and Cooper, 2012; Chabot and Shkreta, 2016). Splicing regulation requires multiple interactions between sequences present in the pre-mRNA and *trans* factors that target these positive and negative signals. Among the *trans* factors are the SR proteins (Lin and Fu, 2007; Long and Caceres, 2009; Shepard and Hertel, 2009; Han et al., 2011) and the hnRNP

proteins (Han et al., 2010; Busch and Hertel, 2012). The accuracy of splice site selection is accomplished through the blending of numerous weak interactions between RNA:RNA, protein:RNA, and protein:protein.

The endogenous spliceosome assembles individual transcripts of Pol II in a giant RNP (21 MDa)—called the supraspliceosome. All nuclear pre-mRNAs, regardless of their intron number and length, are packaged in supraspliceosomes. The latter can be isolated from cell nuclei under physiological conditions and remain active in splicing (reviewed in Sperling et al., 2008; Shefer et al., 2014; Sperling, 2017). Supraspliceosomes are composed of the five spliceosomal U snRNPs and additional splicing factors (Mirami et al., 1995; Azubel et al., 2006). The five spliceosomal U snRNPs are associated with the supraspliceosome at all splicing steps, as revealed by examining affinity-purified specific supraspliceosomes at different splicing stages (Kotzer-Nevo et al., 2014). The supraspliceosome harbors splicing factors such as all phosphorylated SR proteins (Yitzhaki et al., 1996), hnRNP G (Heinrich et al., 2009), and the alternative splicing factors RBM4 and WT1 (Markus et al., 2006) and ZRANB2 (Yang et al., 2013). Mass spectrometry (MS) analysis of supraspliceosomes has revealed further splicing factor components (Chen et al., 2007) as did MS analysis of specific supraspliceosomes analyzed at distinct functional states (Kotzer-Nevo et al., 2014). The presence of regulatory splicing factors in supraspliceosomes is in accordance with their task in splicing and AS (Heinrich et al., 2009; Sebbag-Sznajder et al., 2012). Additional components found in supraspliceosomes are pre-mRNA processing factors, among them are the cap-binding proteins, 3'-end processing components (Raitskin et al., 2002), and the ADAR1 and ADAR2 editing enzymes (Raitskin et al., 2001). These findings portray the supraspliceosome as the nuclear pre-mRNA processing machine.

The supraspliceosome is formed of four active native spliceosomes joined together by the pre-mRNA (Sperling et al., 1997; Müller et al., 1998; Medalia et al., 2002; Azubel et al., 2004; Azubel et al., 2006; Cohen-Krausz et al., 2007). The native spliceosome, which is similar to an *in vitro* assembled spliceosome, is an elongated globular particle made of large and small substructures, as resolved by single particle cryo-electron microscopy (cryo-EM) at a resolution of 20 Å (Azubel et al., 2004). *In silico* studies have localized the spliceosomal U snRNPs within the native spliceosome in a single layout, mainly within the large substructure, thereby protecting the elements of the active center in the cleft within the spliceosome (Frankenstein et al., 2012). The native spliceosomes are placed within the supraspliceosome with their small substructures facing its center, an arrangement that enables interactions between them. Communication between the native spliceosomes within the supraspliceosomes is likely an essential aspect of splicing control, also required for quality control of the mRNAs (Azubel et al., 2006; Cohen-Krausz et al., 2007). The supraspliceosome thus emerges as a principal controller of pre-mRNA processing important in the regulation of multiple pre-mRNA processing steps.

Although most RNA Pol II transcribed transcripts are multi-intronic, at present, it is not well understood how spliceosomes are assembled on a multi-intronic pre-mRNA. One view that has

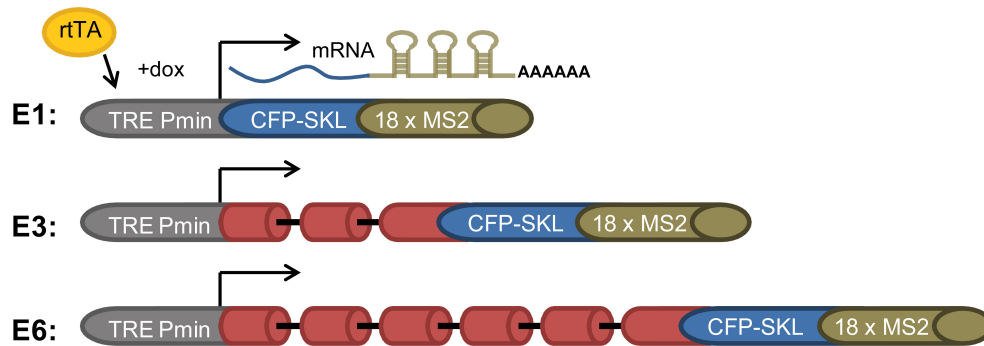


FIGURE 1 | E1, E3, and E6 globin constructs. (i) E1—a gene with no introns containing part of exon 3 and the CFP-SKL signal. The 3'UTR has 18 MS2 repeats; (ii) E3—having two introns and three exons, and identical 3'-end to that of E1; and (iii) E6—in which part of exon 2 intron two and part of exon 3 were duplicated three times and fused to the 5' part of exon 3, with the 3'-end identical to E1 and E3, generating a construct with six exons and five introns. All constructs were Tet-On inducible and stably integrated into human U2OS Tet-On cells (Brody et al., 2011). Transcription was induced by the addition of doxycycline (dox) to the medium in the presence of the rTA (Tet-On) transactivator. The different-sized gene constructs produced the correctly spliced mRNAs.

emerged from *in vitro* studies has suggested that a spliceosome is assembled onto each of the synthesized introns of the pre-mRNA, which disassembles after splicing is performed in order to get ready for the next round of splicing (Staley and Guthrie, 1998). This model suggests that splicing of a multi-intronic pre-mRNA requires the assembly of multiple spliceosomes, whose number equals the number of introns. Another view, developed from studies of the endogenous spliceosome, demonstrated that the pre-mRNA is assembled into supraspliceosomes, composed of four active native spliceosomes, which are linked by the transcript. Notably, transcripts having diverse number of introns or length are found in supraspliceosomes, indicating that the four spliceosomes of the supraspliceosome are adequate for splicing of every transcript. Furthermore, the distinctive size and hydrodynamic assets of supraspliceosomes signify their universal nature (reviewed in Sperling et al., 2008; Shefer et al., 2014; Sperling, 2017; Sperling and Sperling, 2017). To study the spliceosomes assembled *in vivo* on transcripts with variable number of introns, we examined herein a series of three related transcripts: two with rising number of introns originated from the β -globin gene, while one had only an exon. Each transcript had multiple MS2 sequence repeats that can be bound by the MS2 coat protein (Brody et al., 2011). We show here that pre-mRNA transcripts with no intron (termed E1), with two introns (E3), and with five introns (E6) are found in supraspliceosomes. Affinity purification of complexes assembled on the transcript with most introns (termed E6), using the MS2 tag, confirmed the assembly of E6 mRNA in supraspliceosomes. Furthermore, splicing inhibition by spliceostatin A did not inhibit the assembly of supraspliceosomes on the E6 transcript, yet, increased the percentage of E6 pre-mRNA supraspliceosomes. These findings were corroborated in intact cells, using RNA FISH to detect the MS2-tagged E6 mRNA, together with GFP-tagged splicing factors, showing the assembly of splicing factors SRSF2, U1-70K, and PRP8 onto the E6 transcripts under normal conditions and also when splicing was inhibited. This study shows that different transcripts with different number of introns, or lacking one, are assembled in supraspliceosomes even when splicing is

inhibited. This assembly starts at the site of transcription and continues during the life of the transcript in the nucleoplasm. This study further confirms the dynamic and universal nature of supraspliceosomes that package Pol II transcribed pre-mRNAs into complexes composed of four native spliceosomes connected by the transcript, independent of their length, number of introns, or splicing state.

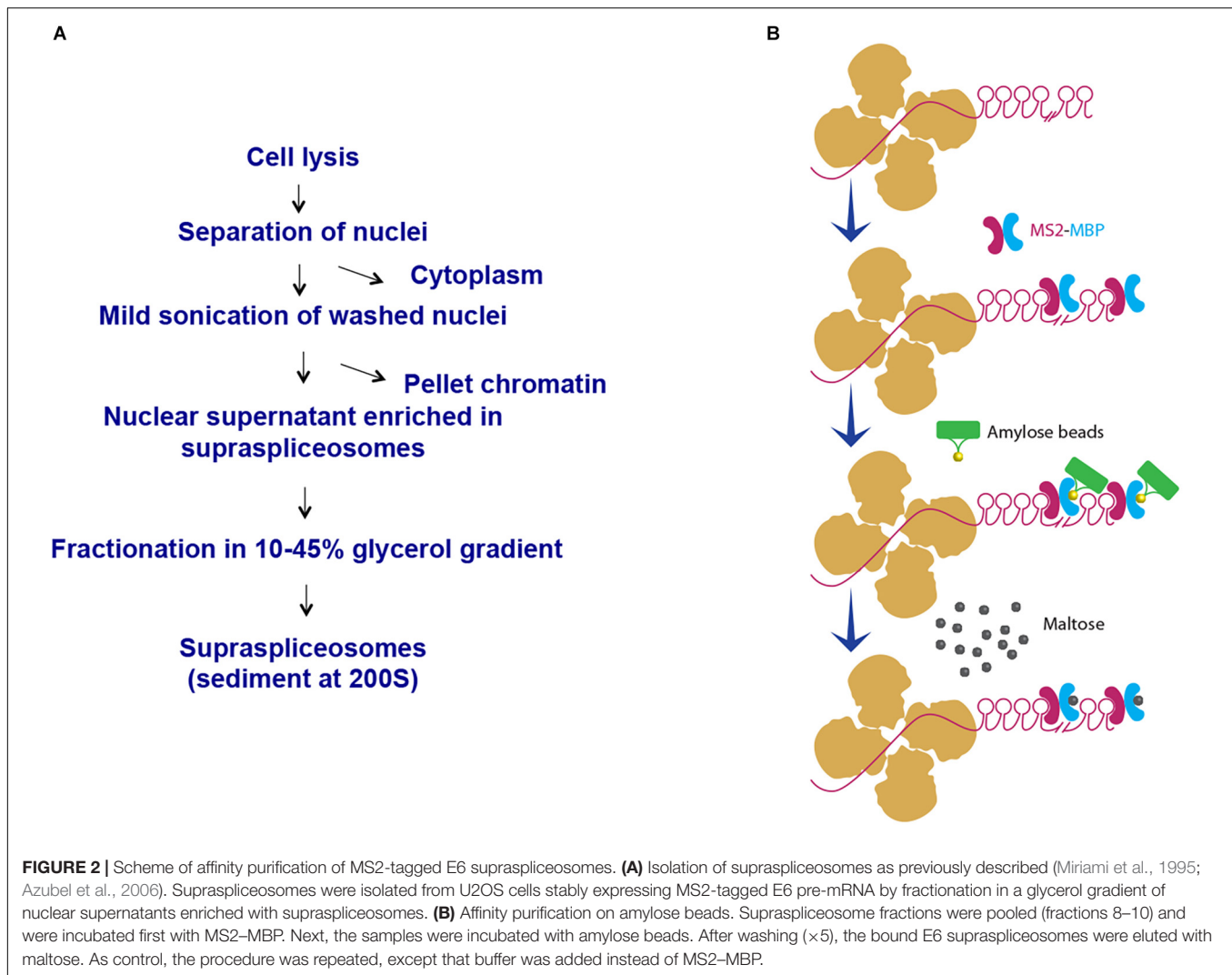
MATERIALS AND METHODS

Cells

U2OS Tet-On human osteosarcoma cells were grown in low-glucose Dulbecco's modified Eagle's medium (DMEM, Biological Industries, Israel) containing 10% fetal bovine serum (FBS, HyClone). A series of U2OS stable cell lines were used, as described in Brody et al. (2011). The cells used were U2OS Tet-On cells containing a stable integration of a Tet-inducible β -globin mini-gene termed E1, E3, or E6, where the number denotes the number of exons in the gene. The genes were integrated as tandem gene arrays in one locus that forms a detectable site of transcription upon gene activation. Induction of transcription was obtained by doxycycline (dox, 1 μ g/mL, Sigma) and results in the expression of a transcript encoding β -globin fused to a CFP protein that contains an SKL tripeptide for peroxisomal targeting, and in the 3'UTR, a series of 18 \times MS2 sequence repeats. For imaging, U2OS E6 cells carrying a stable integration of BACs were used. The BACs were C-terminally GFP-tagged SC35 (SRSF2), U1-70K, and PRP8 that were previously described (Poser et al., 2008; Huranova et al., 2010; Hochberg-Laufer et al., 2019b). For splicing inhibition, U2OS E6 cells were incubated with either 10 or 100 ng/mL Spliceostatin (SSA) (a kind gift from Dr. Yoshida's lab) for 5 h, or with Pladienolide B (10 μ M, Santa Cruz) for 6 h.

Supraspliceosome Isolation

For isolation of supraspliceosomes, first, nuclear supernatants, which were enriched in supraspliceosomes were prepared



as previously described (Miriami et al., 1995; Azubel et al., 2006), from the U2OS cell clones. Briefly, nuclear supernatants enriched for supraspliceosomes were prepared from purified nuclei of the above-described U2OS cells, by microsonication of the nuclei and precipitation of the chromatin in the presence of tRNA. After fractionation of the nuclear supernatants in 10–45% glycerol gradients in an SW41 rotor, at 11,700 rpm for 18 h, the gradients were analyzed by EM visualization of aliquots from fractions corresponding to the 200S region of the gradient (tobacco mosaic virus served as a sedimentation coefficient marker). RNA extraction from gradient fractions was performed as previously described (Azubel et al., 2006).

RNA Isolation and Analysis

RNA was isolated from each gradient fraction (520 μ L) by adding 150 μ L of extraction buffer (50 mM Tris-HCl, pH 7.5, 30 mM NaCl, and 1 mM EDTA) and 50 μ L of 10% SDS. RNA isolated from gradient fractions was analyzed by RT-PCR as previously described (Sebbag-Sznajder et al., 2012).

For total RNA analysis, RNA was extracted (Sperling et al., 1985) and RT-PCR was performed with the relevant primers corresponding to the E1, E3, and E6 mRNA, and actin as a control.

For analysis of E1:

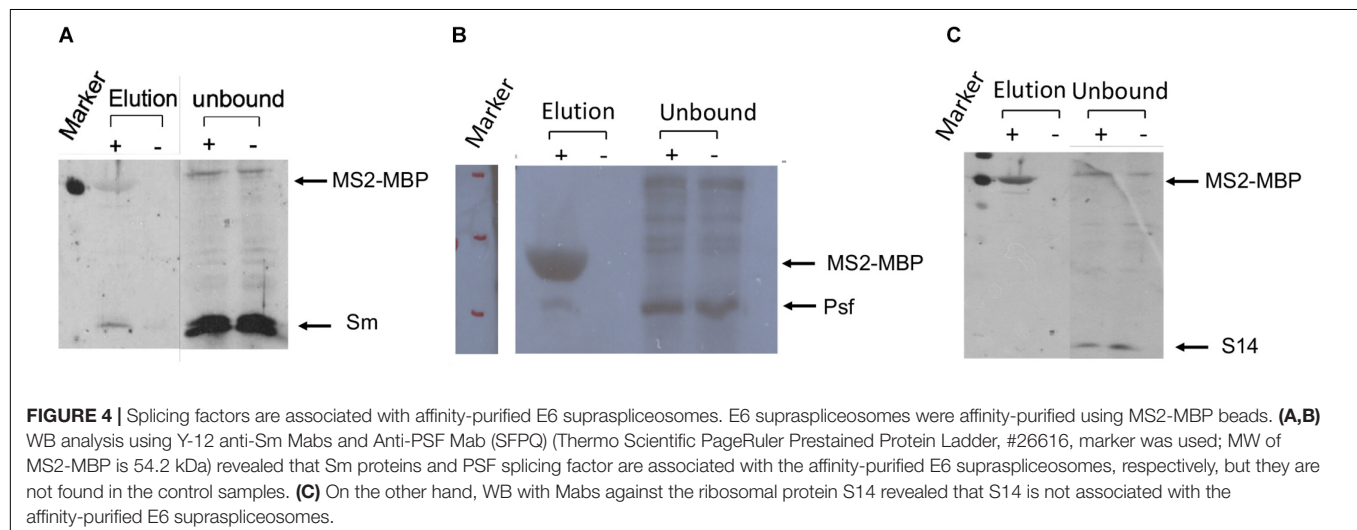
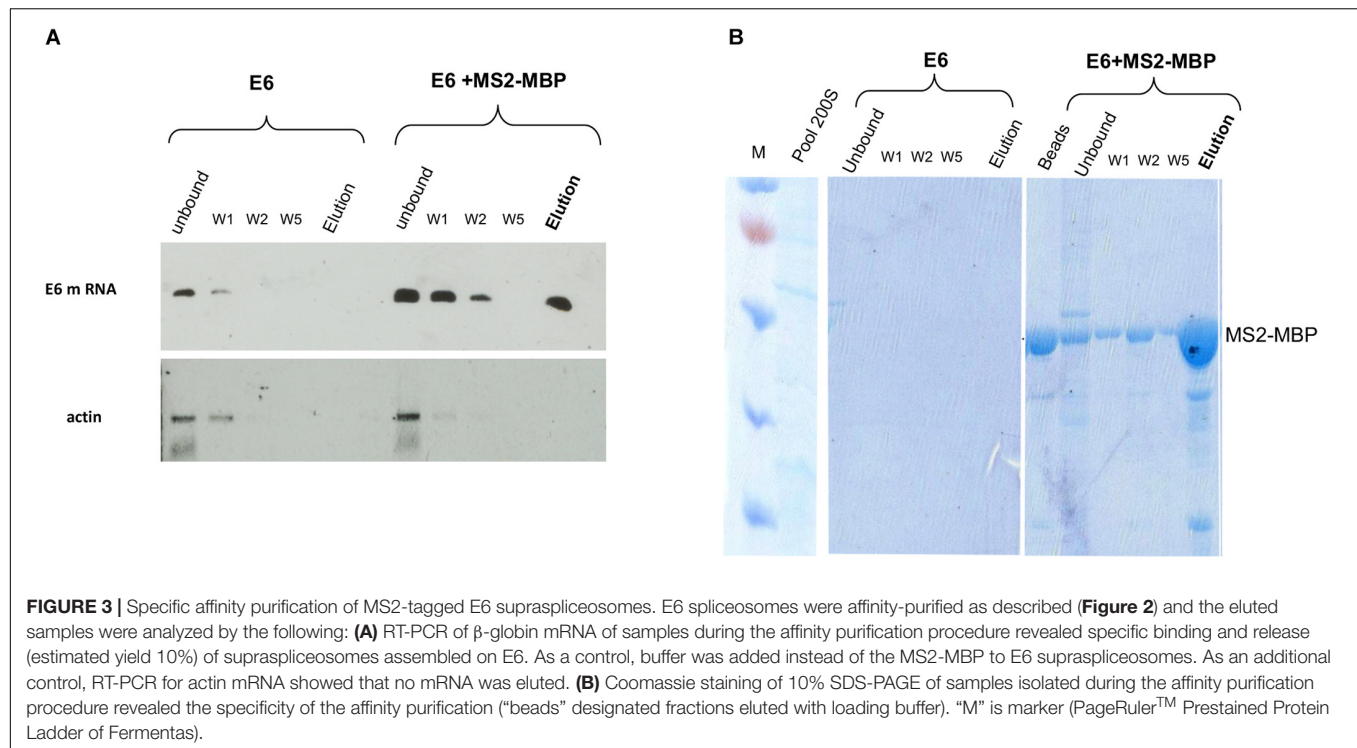
CFP-SKL forward: 5'-GCAAGCTGACCCTGAAGTTC-3'
CFP-SKL reverse: 5'-GTCTTGTAGTTGCCGTCGT-3'

For analysis of E3 and E6:

E6 β -globin forward Ex1: 5'-GCAACCTCAAACAGACA CCA-3'
E6 β -globin reverse Ex2: 5'-CAGCATCAGGAGTGGAC AGA-3'
E6 β -globin reverse CFP: 5'-GCCCTTGCTCACCATGAAT-3'

For analysis of β -actin:

Actin sense: 5'-CAAGGCCAACCGCGAGAAGATGAC-3'
Actin antisense: 5'-AGGAAGGAAGGCTGGAAGAGTGC-3'



Affinity Purification of MS2-Tagged E6 Supraspliceosomes

1.5 mL of supraspliceosomes isolated from U2OS E6 cells (from 10 plates of 15 cm, collecting fractions 8–10 of the glycerol gradients, see above) were incubated with 300 μ g of MS2-MBP [prepared as described (Das et al., 2000)] for 1 h, at 4°C. Next, washed Amylose beads (200 μ L of Amylose resin 50% in 20% EtOH) were added to the supraspliceosomes/MS2-MBP sample and left for overnight incubation at 4°C with shaking at 15 rpm. As a control, we used supraspliceosomes without MS2-MBP. After centrifugation of the beads with supraspliceosomes (supernatant termed unbound), and washing

($\times 5$), supraspliceosomes were eluted with maltose by incubating with 400 μ L of the elution buffer (20 mM maltose) for 30 min at 15 rpm, at 4°C. The tubes were spun down and the supernatant was kept for analysis.

Western Blotting

For WB analysis, proteins were precipitated in 80% cold acetone, using 1 μ L of Quick Precip (EdgeBio, cat No 14201) as carrier. The pelleted proteins were dissolved in SDS sample buffer and analyzed by 12% SDS PAGE. Gels were either stained with Coomassie Blue G250, or analyzed by Western blots using the anti-Sm antibody [Y12, 1: 10,000 dilution in NET (150 mM

NaCl, 50 mM Tris, pH 7.5, 0.05% Triton X-100/or NP-40)]; anti-S14 Mab (1:3000 dilution); and anti-PSF Mab (SFPQ) (Sigma, 1:3000 dilution), visualized with horseradish peroxidase conjugated to goat anti-mouse antibody (1:3000 dilution of goat anti-mouse Fab2-HRP, Jackson), as previously described (Sebbag-Sznajder et al., 2012).

EM Visualization

Aliquots (10 μ L) from the samples were absorbed on glow-discharged carbon-coated copper EM grids, washed with water, and negatively stained with 1% (w/v) uranyl-acetate. A Tecnai 12 TEM (FEI), operating at an acceleration voltage of 100 kV, equipped with a CCD camera was used.

Fluorescence *in situ* Hybridization

Cells grown on coverslips were fixed for 20 min in 4% PFA and then transferred to 70% ethanol at 4°C for overnight. The next day, cells were washed with 1 \times PBS and treated for 2.5 min with 0.5% Triton X-100. Cells were washed with 1 \times PBS and incubated for 10 min in 40% formamide (4% SSC). Cells were transferred to 40% formamide at 37°C and hybridized overnight with a specific Cy5 fluorescently labeled DNA probe (\sim 10 ng probe, 50 mer) that binds to the MS2 region of the MS2 repeats. The intron probe was described in Brody et al. (2011). The next day, cells were washed twice with 40% formamide for 15 min and then washed for 2 h in 1 \times PBS. Nuclei were counterstained with Hoechst 33342 (Sigma) and coverslips were mounted in mounting medium. Wide-field fluorescence images were obtained using the Cell'R system based on an Olympus IX81 fully motorized inverted microscope (60 \times PlanApo objective, 1.42 NA) fitted with an Orca-AG CCD camera (Hamamatsu) driven by the Cell'R software. When imaging, the focus was on the active transcription sites that usually produced a strong signal in the nucleus. Presentation of this transcription site signal without saturation usually reduced the ability to present the weaker mRNA signal in the rest of the cell.

RESULTS

A Series of Transcripts With Variable Number of Introns

Splicing and alternative splicing of pre-mRNA play a major role in regulating gene expression. Yet, how spliceosomes are assembled on a multi-intronic pre-mRNA is at present not well understood. To study the spliceosomes assembled *in vivo* on transcripts with variable number of introns, we examined a series of three related transcripts derived from the β -globin gene. Two of the genes contained increasing number of introns, while one gene encoded an exon only. The transcripts were each expressed in U2OS Tet-On stable cell clones that expressed the β -globin mini-genes that contain a series of MS2 sequence repeats in their 3'UTR (Brody et al., 2011). The three genes were under the transcriptional control of the inducible Tet-On system, and transcription was induced in the presence of the rtTA (Tet-On) transactivator expressed by the cells and the addition of

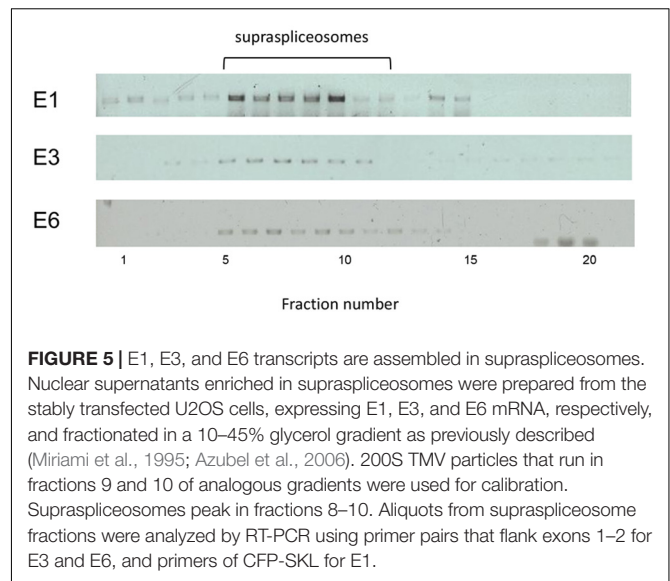


FIGURE 5 | E1, E3, and E6 transcripts are assembled in supraspliceosomes. Nuclear supernatants enriched in supraspliceosomes were prepared from the stably transfected U2OS cells, expressing E1, E3, and E6 mRNA, respectively, and fractionated in a 10–45% glycerol gradient as previously described (Miriami et al., 1995; Azubel et al., 2006). 200S TMV particles that run in fractions 9 and 10 of analogous gradients were used for calibration. Supraspliceosomes peak in fractions 8–10. Aliquots from supraspliceosome fractions were analyzed by RT-PCR using primer pairs that flank exons 1–2 for E3 and E6, and primers of CFP-SKL for E1.

doxycycline (dox) to the medium. The cell clones used were as follows: (i) E3, consisting of a β -globin mini-gene with three exons and two introns (**Figure 1**). Specifically, exon 3 was truncated and fused in-frame to a cyan fluorescent protein (CFP) coding region containing in its C-terminus the peroxisomal targeting tripeptide Ser-Lys-Leu (SKL). The mRNA therefore finally generated cytoplasmic cyan fluorescing peroxisomes. At the 3'-end of the gene, a series of 18 MS2 sequence repeats were added, thus providing high-affinity binding sites for the MS2 coat protein in the 3'UTR of the mRNA. (ii) E1, an intronless version of the mini-gene containing part of exon 3 only + CFP-SKL, together forming a single exon; and (iii) E6 with six exons and five introns, in which intron 2, flanked by the splice sites and part of exons 2 and 3, was multiplied (**Figure 1**). These cell clones were previously used in a study that followed transcription in living cells, showing that the transcriptionally active E3 and E6 genes recruited splicing factors and that the E3 and E6 mRNAs were co-transcriptionally spliced (Brody et al., 2011).

E6 Transcripts Are Assembled in Supraspliceosomes

We have chosen to focus first on the E6-expressing cells as the E6 transcript has the highest number of introns relative to the E1- and E3-expressing cells. Also, preliminary analysis of the cells revealed that the E6-expressing cells had the highest level of expression of the β -globin MS2-tagged transcript (data not shown). We first used our protocol for isolation of endogenous spliceosomes under native conditions from cell nuclei (**Figure 2A**) to isolate complexes assembled in the E6 expressing U2OS cells. For this purpose, we prepared nuclear supernatants enriched with RNA Polymerase II transcripts, under physiological conditions and fractionated them in 10–45% glycerol gradients, as described previously (Miriami et al., 1995; Azubel et al., 2006). This method conserves higher-order splicing complexes, as formerly demonstrated by electron microscopy (Spann et al., 1989; Azubel et al., 2004; Azubel et al., 2006),

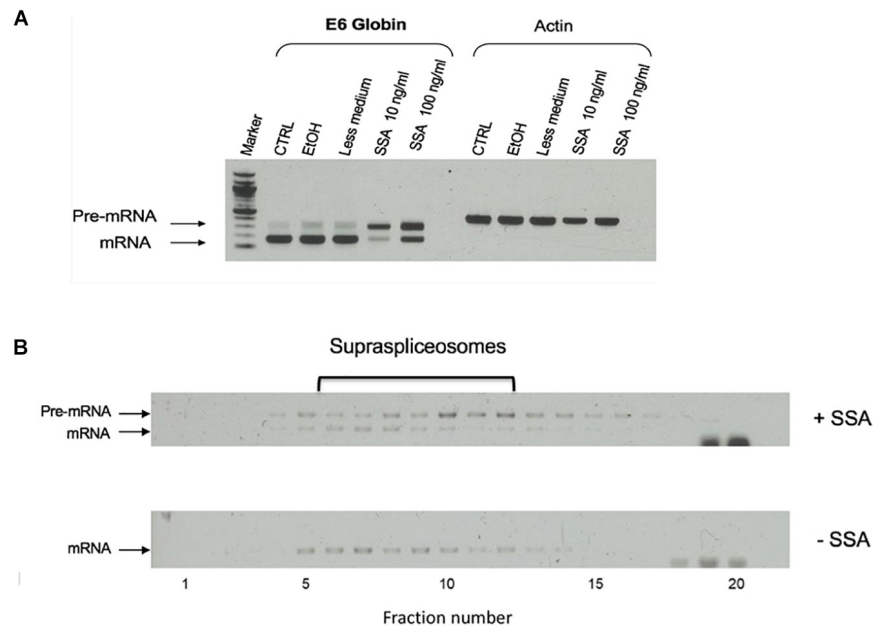


FIGURE 6 | Effect of spliceostatin A (SSA) on E6 expression. **(A)** Effect of SSA on U2OS cells stably expressing the E6 transcript. Total RNA was extracted from E6 expressing U2OS cells treated with SSA (5 h), and the RNA was investigated by RT-PCR using primers from exons 1 and 2 for E6. Actin was used as a control. For comparison, RNA was extracted from untreated cells. Untreated cells incubated with EtOH was used as an additional control, as SSA was solubilized in ethanol. **(B)** Effect of SSA on E6 splicing in supraspliceosomes. Nuclear supernatants enriched in supraspliceosomes were prepared from U2OS E6 cells, treated or not with SSA (100 ng/mL, for 5 h), and were fractionated in a 10–45% glycerol gradient and gathered (bottom to top) in 20 fractions. 200S TMV particles that run in fractions 9 and 10 of analogous gradients were used for calibration. Aliquots from each gradient fraction were examined by RT-PCR with primer pairs that flank exons 1–2 of E6.

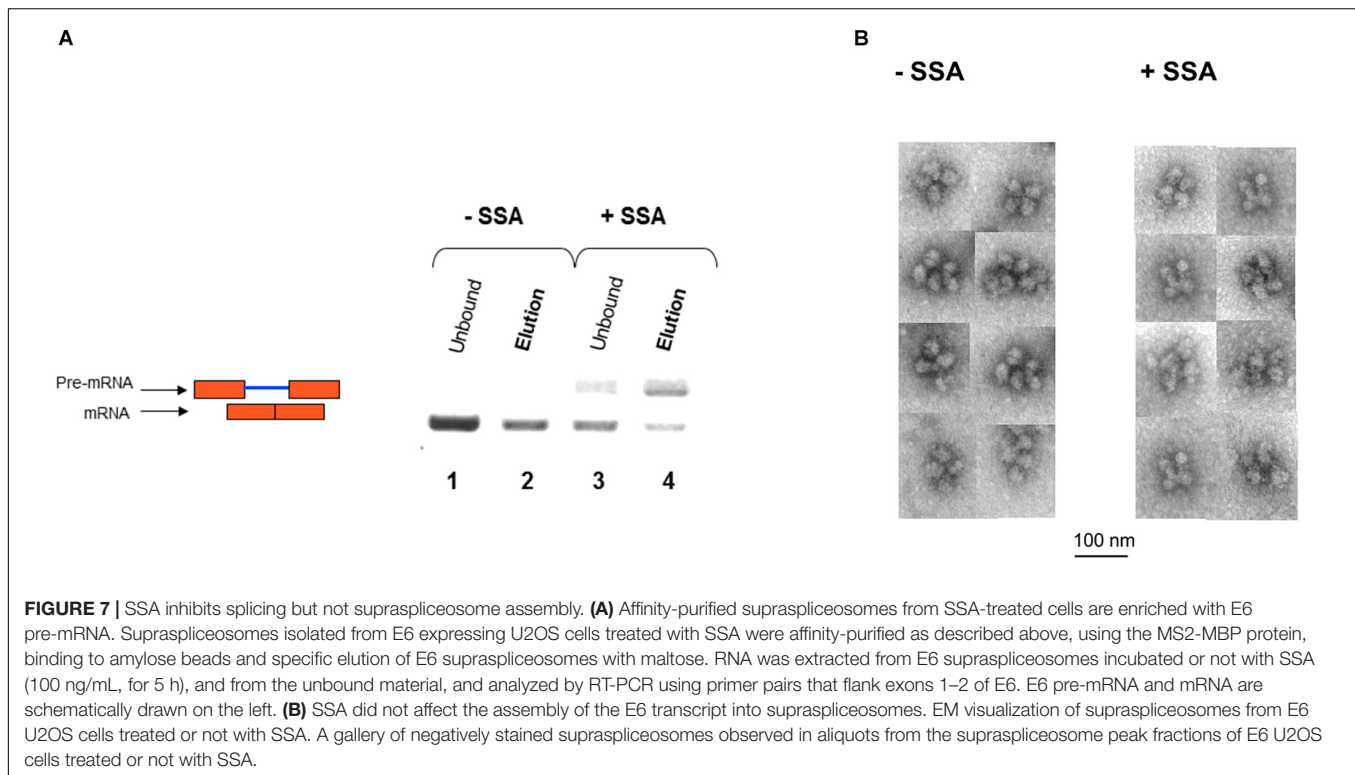
which are associated with splicing factors (Miriami et al., 1995; Yitzhaki et al., 1996; Azubel et al., 2006; Markus et al., 2006; Heinrich et al., 2009; Sebbag-Sznajder et al., 2012; Yang et al., 2013; Kotzer-Nevo et al., 2014). To affinity purify complexes assembled on E6 transcripts, we followed the scheme detailed in **Figure 2B**. Recombinant MS2-MBP protein was added to the pooled supraspliceosome fractions (see section Materials and Methods), and the bound supraspliceosomes were further incubated overnight with amylose beads at 4°C. After washing, the bound supraspliceosomes were eluted by maltose. As a control, we repeated the same protocol with buffer only instead of MS2-MBP.

RT-PCR analysis of the affinity-purified supraspliceosomes revealed the association of E6 transcripts with supraspliceosomes (**Figure 3A**). The affinity purification is specific as E6 mRNA was specifically released by maltose from the amylose beads bound by MS2-MBP, while no eluted E6 mRNA was observed in the control experiment (**Figure 3A**). The binding to the amylose beads is specific to the MS2-tagged E6 mRNA, as RT-PCR analysis of endogenous actin mRNA showed no binding or elution. The specificity of the affinity purification was also confirmed by analysis of the eluted proteins by SDS PAGE followed by staining of the gel with Coomassie (**Figure 3B**), where eluted proteins were observed only in the samples bound by MS2-MBP and amylose. As expected, no MS2-MBP was observed in the control sample without it, yet no other proteins were observed in the eluted sample from the control experiment.

The association of E6 with supraspliceosomes was further confirmed by WB analysis using anti-Sm Mabs (Y12), revealing that splicing components such as Sm proteins were associated with the affinity-purified E6 supraspliceosomes (**Figure 4A**). It should be noted that, as expected, the level of Sm proteins in the unbound sample was much higher than in the eluted sample. This is because the affinity purification is specific to the MS2-tagged supraspliceosomes, which constitute only a small fraction of the entire endogenous nuclear transcripts' population, the majority lacking the MS2 tag. This was also exemplified in **Figure 3A** where endogenous actin supraspliceosomes were not affinity-purified. The association of E6 with supraspliceosomes was further confirmed by the association of E6 with the regulatory splicing factor PTB-associated splicing factor (PSF) (**Figure 4B**), also termed SFPQ (Shav-Tal and Zipori, 2002). On the other hand, WB analysis with antibodies directed against the ribosomal protein S14 (**Figure 4C**) revealed that S14 was not present in the affinity-purified E6 supraspliceosomes, further confirming the specificity of the affinity purification. These experiments revealed that E6 transcripts are assembled in supraspliceosomes.

E1, E3, and E6 Transcripts Are Found in Supraspliceosomes

Each of the E1-, E3-, and E6-expressing cells was used to analyze the type of splicing complexes assembled on this series of transcripts containing an increasing number of introns and exons (E1, one exon and no intron; E3, three exons and two introns; E6,



six exons and five introns). For this purpose, we used our protocol for isolation of endogenous spliceosomes under native conditions from cell nuclei and prepared nuclear supernatants enriched with RNA Pol II transcripts and fractionated them in 10–45% glycerol gradients (**Figure 2A**), as previously described (Miriami et al., 1995; Azubel et al., 2006). Next, we analyzed the distribution of each of the E1, E3, and E6 transcripts across its respective gradient by RT-PCR. **Figure 5** reveals that despite the change in length and number of introns, E1, E3, and E6 transcripts peaked at the 200S region of the gradient, where supraspliceosomes (21 MDa) sedimented (Miriami et al., 1995; Müller et al., 1998; Azubel et al., 2006). The sedimentation patterns of E1, E3, and E6 were analogous to those of the phosphorylated SR proteins (Yitzhaki et al., 1996; Raitskin et al., 2002; Heinrich et al., 2009) and hnRNP G (Heinrich et al., 2009), which were shown previously to be predominantly associated with supraspliceosomes in these fractions. These results are consistent with our previous finding showing that diverse transcripts regardless of their length or number of introns are assembled in supraspliceosomes (Spann et al., 1989; Kotzer-Nevo et al., 2014). It also shows for the first time that even transcripts that are devoid of introns, like the E1 transcript, are assembled in supraspliceosomes, emphasizing their universal nature.

Spliceostatin A (SSA) Inhibits Splicing but Not the Assembly Into Supraspliceosomes

To analyze how splicing inhibition affects the assembly of the supraspliceosome, we used Spliceostatin A (SSA). SSA is

a methylated derivative of an anticancer bacterial metabolite FR901464. It inhibits splicing *in vitro* and *in vivo* by binding to SF3b, a component of U2 snRNP (Kaida et al., 2007). Previous studies showed that SSA inhibits spliceosome assembly *in vitro*, yet, all five spliceosomal U snRNPs and SSA were found associated with the inhibited spliceosomes (Roybal and Jurica, 2010). It was shown that SSA inhibits the binding of the SF3b 155-kDa protein to the pre-mRNA, resulting in reduced binding specificity of the U2 snRNP to the branch point, and causing some changes in alternative splicing (Corrionero et al., 2011). RNA-seq of transcripts after SSA treatment revealed that intron retention, namely, splicing inhibition, is the major effect of SSA on splicing (Carvalho et al., 2017; Yoshimoto et al., 2017). Furthermore, previous analysis of the effect of SSA on E6 transcripts in intact cells revealed that SSA treatment affects splicing, but not the rate of transcription. Yet, SSA obliterated the retention of E6 transcripts at the transcription site, resulting in rapid release of the transcript to the nucleoplasm (Brody et al., 2011), a release that can also be generated by the availability of splicing factors in the nucleoplasm (Hochberg-Laufer et al., 2019a). Treatment with SSA also results in partial pre-mRNA leakage (Kaida et al., 2007; Brody et al., 2011; Martins et al., 2011; Schmidt et al., 2011; Takemura et al., 2011; Carvalho et al., 2017; Yoshimoto et al., 2017).

To test the effect of SSA on the expression of E6 transcripts, we first incubated the cells for 5 h with SSA at 10 or 100 ng/mL and prepared total RNA from the treated cells. As controls, we used untreated cells and cells incubated with ethanol, since SSA is dissolved in ethanol. **Figure 6A** shows that while untreated cells expressed mainly mature E6 mRNA,

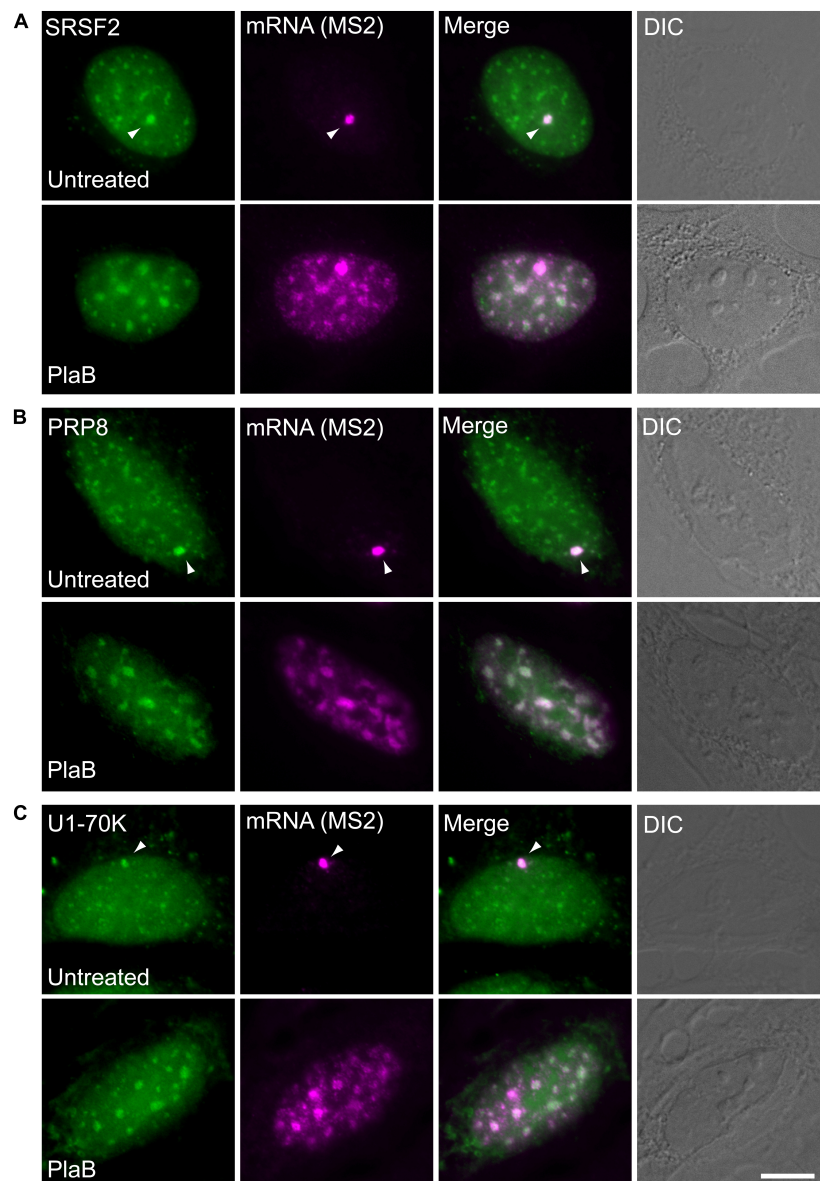


FIGURE 8 | Splicing factors continue to be recruited to the E6 mRNA under splicing inhibition conditions. The recruitment of several RNA processing factors to the actively transcribing E6 gene was examined in cells stably expressing BACs that transcribe GFP fusions of either **(A)** SRSF2, or **(B)** Prp8, or **(C)** U1-70K (green) under normal and splicing inhibition by Pladienolide B (PlaB). The active site of transcription of the E6 gene was detected by RNA FISH with a Cy5-labeled probe that hybridizes to the MS2 region of the E6 mRNA (magenta). Arrowheads point to the actively transcribing genes. DIC is in gray. Bar = 5 μ m.

after treatment with SSA at 10 ng/mL, the E6 mRNA was predominantly unspliced. Treatment with 100 ng/mL SSA decreased the percentage of E6 pre-mRNA. This effect of increasing amount of SSA on splicing is not clear. It is possible that additional effects of SSA on gene expression play a role here, such as the coupling of transcription and splicing (e.g., it has been shown that treatment with SSA at 100 ng/mL decreases the phosphorylation of Ser2 in Pol II CTD, causes early dissociation of Pol II, and decreases phospho-Ser2 level of chromatin-bound Pol II; Koga et al., 2015). Thus, it is possible that high SSA levels affect transcription in addition to

splicing, yet other explanations cannot be ruled out at this stage. Next, we tested the effect of SSA on the E6 supraspliceosomes. Supraspliceosomes were prepared from the E6-expressing U2OS cells, either treated or untreated with 100 ng/mL SSA for 5 h, and fractionated in glycerol gradients. **Figure 6B** shows that E6 supraspliceosomes from SSA-treated and untreated cells were found in supraspliceosomes that sediment at 200S in the glycerol gradients. While E6 supraspliceosomes from untreated cells assembled E6 mRNA, E6 supraspliceosomes from SSA-treated cells portrayed mainly E6 pre-mRNA and a lower percentage of E6 mRNA.

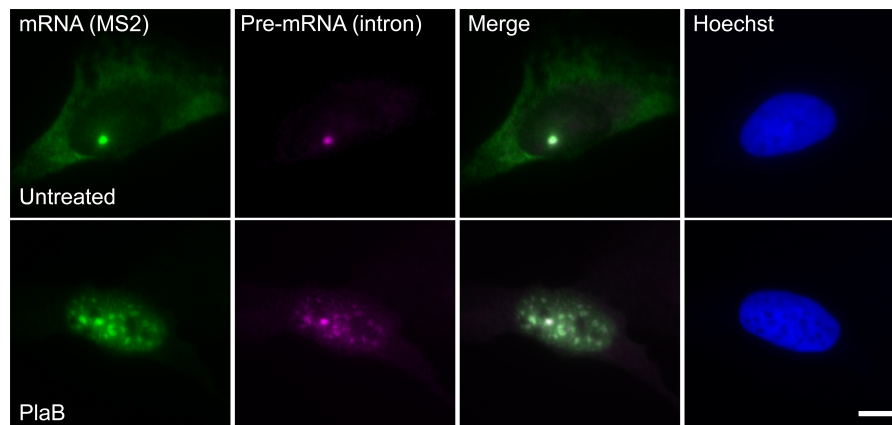


FIGURE 9 | E6 mRNA accumulated in the nucleus under splicing inhibition. The distribution of E6 mRNA under normal (top) and splicing inhibition conditions by Pladenolide B (PlaB) (bottom) was detected by RNA FISH using a Cy5-labeled probe that hybridizes with the MS2 region of the E6 mRNA (green), and a Cy3-labeled probe that binds to the intron of the E6 mini-gene (magenta). Hoechst DNA counterstain is in blue. Bar = 10 μ m.

We next affinity-purified E6 supraspliceosomes from SSA-treated and untreated cells (**Figure 7A**). Supraspliceosomes were found assembled on mature E6 mRNA in untreated cells. After treatment with SSA, affinity-purified E6 supraspliceosomes were mainly assembled on E6 pre-mRNA, while a small percentage was assembled on E6 mRNA. These studies revealed that SSA inhibits splicing, but does not interfere with supraspliceosome assembly. This finding is further confirmed by electron microscopy visualization of aliquots from the 200S peak region of the glycerol gradients where supraspliceosomes sediment. Supraspliceosomes, composed of four native spliceosomes, were visualized in both treated and untreated cells, and no significant structural changes could be visualized in the SSA-treated supraspliceosomes (**Figure 7B**).

These findings were next corroborated in intact cells. We examined whether splicing factors can indeed continue to co-transcriptionally assemble on the E6 mRNAs during transcription under normal and splicing inhibition conditions. We used U2OS Tet-On stable cell lines that contain a stable integration of the E6 gene in one gene locus. The gene integration forms a tandem array of the gene and, upon activation with dox, a single spot of the active E6 gene can be detected by RNA FISH with a probe to the MS2 repeats in the 3'UTR of the transcript (Brody et al., 2011). This is the active site of transcription on which we wanted to detect whether co-transcriptional recruitment of splicing factors occurs, as previously described (Huranova et al., 2010; Brody et al., 2011; Hochberg-Laufer et al., 2019a,b). In order to detect the splicing factors in intact cells, we used E6 cells with additional stable integrations of bacterial artificial chromosomes (BACs) (**Figures 8A–C**) containing the full gene body of either the SR protein SRSF2 (SC35), or two snRNP components, U1-70K (part of the U1 snRNP, which binds to the 5'-splice site) and PRP8 (part of U5 snRNP, which is part of the U4/U6.U5 triple-snRNP) tagged with GFP in the C-terminal (Poser et al., 2008). Using RNA FISH that detects the MS2-tagged E6 mRNA together with the staining of the cells with GFP-tagged SRSF2, or U1-70K or Prp8, we found that these splicing factors

were recruited to the transcriptionally active E6 gene under normal conditions, as expected. When splicing was inhibited using Pladenolide B, which inhibits splicing by interaction with the SF3B complex [similar to SSA, (Kotake et al., 2007)], the E6 pre-mRNAs accumulated in the nucleus (**Figure 9**), specifically in nuclear speckles that are known to contain splicing factors, as is known to occur for unspliced transcripts. Importantly, the splicing factors continued to be recruited to the transcribing genes and so continued to assemble on the pre-mRNAs under splicing inhibition conditions, in agreement with the biochemical data showing that supraspliceosomes assembled on these mRNAs under all conditions.

DISCUSSION

Although most of RNA Pol II transcribed pre-mRNAs are multi-intronic, the process of spliceosome assembly on such pre-mRNAs is at present not well understood. To address this question, we chose here to examine spliceosome assembly *in vitro* and *in vivo* on a series of three related transcripts: two of the transcripts contained increasing number of introns derived from the β -globin gene (E6 with six exons and five introns, E3 with three exons and two introns) and one that had only one exon and no introns. Each transcript had multiple MS2 sequence repeats that can be bound by the MS2 coat protein and therefore be used for both affinity purification and visualization in intact cells. We showed that E6 transcripts are assembled in supraspliceosomes composed of four native spliceosomes joined together by the transcript. This was confirmed by analyzing isolated and affinity-purified complexes and by studies in intact cells showing the association of E6 with splicing factors. We further demonstrated that this series of transcripts, namely, E6 with five introns, E3 with two introns, and E1 with no intron, are assembled in supraspliceosomes. These findings corroborated our previous findings showing that Pol II transcripts are assembled into 21-MDa supraspliceosomes, regardless of their

number of introns or length (reviewed in Sperling et al., 2008; Shefer et al., 2014; Sperling, 2017). For example, SMN transcripts, having eight introns, were shown sedimenting at 200S with supraspliceosomes, and were demonstrated as assembled in supraspliceosomes by immunoprecipitation using anti-Sm antibodies, further showing that both splicing isoforms, with and without exon 7, were assembled in supraspliceosomes (Sebbag-Sznajder et al., 2012). An additional example is the case of PP7-tagged AdML transcripts that contain one intron. Analysis of the affinity-purified PP7 tagged splicing complexes assembled on this AdML transcript *in vivo* demonstrated that they are assembled in supraspliceosomes (Kotzer-Nevo et al., 2014). The supraspliceosome, composed of four native spliceosomes, can splice four introns at one setting. Because all pre-mRNAs are found assembled in supraspliceosomes, independent of their number of introns, we suggest that splicing of a pre-mRNA having more than four introns likely occurs through the movement of the pre-mRNA through the supraspliceosome in a “rolling model” manner. In the case of pre-mRNA with one intron, or with less than four introns, which are also assembled in supraspliceosomes (Azubel et al., 2006; Kotzer-Nevo et al., 2014), or with intronless transcripts, as shown here, it is likely that the interactions of the transcript with the native spliceosomes are appropriate to keep the supraspliceosome structure. Supraspliceosomes harbor all five spliceosomal U snRNPs (Miriami et al., 1995; Azubel et al., 2006; Kotzer-Nevo et al., 2014) and splicing factors (Yitzhaki et al., 1996; Markus et al., 2006; Chen et al., 2007; Heinrich et al., 2009; Yang et al., 2013; Kotzer-Nevo et al., 2014). The finding of regulatory splicing factors within supraspliceosomes is in line with their function in splicing regulation and alternative splicing (Heinrich et al., 2009; Sebbag-Sznajder et al., 2012). Supraspliceosomes also contain all the additional factors required for pre-mRNA processing, including 5'-cap components, 3'-end processing components, and A-to-I RNA processing, in addition to splicing and alternative splicing components (Raitskin et al., 2001; Raitskin et al., 2002), portraying them as the nuclear pre-mRNA processing machine. This likely explains the finding that a transcript with no introns (E1) is also assembled in supraspliceosomes. Here, we not only confirmed our previous findings but also show that a transcript lacking an intron is also assembled in supraspliceosomes. This result is in agreement with our previous finding that a PP7-tagged AdML transcript assembled on mature AdML is assembled in supraspliceosomes (Kotzer-Nevo et al., 2014).

We next examined how splicing inhibition affects the assembly into supraspliceosomes. For this aim, we used spliceostatin A (SSA), which binds to the SF3b part of U2 snRNP and inhibits splicing *in vitro* and *in vivo* (Kaida et al., 2007; Lo et al., 2007). Treatment with SSA reduces the binding specificity of the U2 snRNP to the branch point, resulting in some changes in alternative splicing (Corrionero et al., 2011). Focusing on E6 transcripts, we show that treatment with SSA inhibits splicing and increases the percentage of E6 pre-mRNA. However, this intron retention did not affect the assembly of E6 pre-mRNA into supraspliceosomes composed of four native spliceosomes connected by the transcript. This was confirmed for

isolated E6 supraspliceosomes, using ultracentrifugation, affinity purification, and electron microscopy. These results were further corroborated by studies in intact cells, using Pladienolide B, which inhibits splicing by interaction with the SF3B complex (similar to SSA; Kotake et al., 2007), showing that the association of E6 pre-mRNA at the active gene locus with essential splicing factors was not affected by the inhibition of splicing, yet it resulted in nuclear accumulation of the E6 spliceosomes. These latter findings are consistent with previous observations of nuclear accumulation of pre-mRNA resulting from treatment with SSA probably due to lack of appropriate export signals that are assembled during regular splicing (Carvalho et al., 2017; Yoshimoto et al., 2017).

The splicing complex is dynamic, undergoing chemical changes during the two steps of the splicing reaction, involving dynamic changes in U snRNA:U snRNA, U snRNA:pre-mRNA, and protein:RNA interactions, accompanied by local structural changes as revealed by the recent high-resolution structures of spliceosome intermediates (reviewed in Will and Luhrmann, 2011; Papasaikas and Valcarcel, 2016; Fica et al., 2017; Shi, 2017a,b; Wilkinson et al., 2018; Plaschka et al., 2019; Yan et al., 2019). In agreement with that, the supraspliceosome is a dynamic complex, as splicing and alternative splicing occur in supraspliceosomes (Azubel et al., 2006; Sebbag-Sznajder et al., 2012). We have shown here that both pre-mRNA and spliced transcripts are assembled in supraspliceosomes, as demonstrated for E6 pre-mRNA and mRNA that are assembled in supraspliceosomes composed of four native spliceosomes joined together by the transcript. This finding and the results showing that a series of three related transcripts, two having a growing number of introns derived from the β -globin gene (two and five introns, respectively) and a third having only one exon and lacking an intron, are assembled in supraspliceosomes confirm the generality of the supraspliceosome.

In previous studies, we have shown that the supraspliceosome is assembled on one transcript (Sebbag-Sznajder et al., 2012). We have further demonstrated that the pre-mRNA is linking the four native spliceosomes of the supraspliceosome, as specific cleavage of the pre-mRNA using RNase H yielded native spliceosomes (Azubel et al., 2004). In this study, we further show that the transcript whether spliced or not is connecting the four spliceosomes of the supraspliceosome. Namely, an mRNA can also connect the four native spliceosomes, as in the case of E1 transcripts, which lack an intron, and yet they are also found assembled in supraspliceosomes. This finding is supported by previous studies showing that affinity-purified transcripts of AdML mini-gene having either one intron or spliced were found assembled in supraspliceosomes (Kotzer-Nevo et al., 2014).

It should be noted that our previous studies have shown that both the native spliceosome and the supraspliceosome contained all five spliceosomal U snRNPs (Azubel et al., 2006) and that the supraspliceosome harbors the five spliceosomal U snRNPs throughout all stages of the splicing reaction (Kotzer-Nevo et al., 2014). This is in contrast to changes in composition during spliceosome assembly observed *in vitro* (Wahl et al., 2009; Will and Luhrmann, 2011). It is possible that supraspliceosomes and native spliceosomes harbor additional components to those

of intermediate complexes assembled *in vitro*, which help keep them together. The remodeling of the spliceosome during the splicing reaction is regulated by a vast dynamic network of RNA:RNA, protein:protein, and RNA:protein interactions. These alterations might not require extensive variations in the general shape of the splicing complex, but might be accommodated by limited conformational variations. It should be pointed out that the detailed dynamic changes that happen along the two steps of the splicing reaction cannot be visualized at this resolution of EM visualization of supraspliceosomes, and higher-resolution studies are required for that.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

RS, JS, and YS-T developed the project and designed the experiments. NS-S performed the supraspliceosome experiments.

YB generated the cell clones. HH-L performed the imaging experiments. RS and YS-T wrote the manuscript.

FUNDING

This work was partly supported by the US National Institutes of Health (RO1 GM079549 to RS and JS), the Israel Science Foundation (ISF to RS), the Helen and Milton Kimmelman Center for Biomolecular Structure and Assembly at the Weizmann Institute of Science (to JS), and the German Israel Foundation (GIF to YS-T).

ACKNOWLEDGMENTS

We thank Aviva Petcho for excellent technical assistance and Minna Angenitski for help with the electron microscopy. We also thank Eva Böhnlein and Dr. Karla Neugebauer (Yale University) for the integration of the BACs into the U2OS cell lines. We are grateful to Dr. Minoru Yoshida of The Chemical Genomics Research Group and Chemical Genetics Laboratory, RIKEN, The Institute of Physical and Chemical Research, Wako City, Saitama, Japan, for spliceostatin A.

REFERENCES

- 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
- Azubel, M., Habib, N., Sperling, J., and Sperling, R. (2006). Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. *J. Mol. Biol.* 356, 955–966. doi: 10.1016/j.jmb.2005.11.078
- Azubel, M., Wolf, S. G., Sperling, J., and Sperling, R. (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol. Cell* 15, 833–839. doi: 10.1016/j.molcel.2004.07.022
- Brody, Y., Neufeld, N., Bleberstein, N., Causse, S. Z., Bohnlein, E. M., Neugebauer, K. M., et al. (2011). The *in vivo* kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* 9:e1000573. doi: 10.1371/journal.pbio.1000573
- Brow, D. A. (2002). Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* 36, 333–360. doi: 10.1146/annurev.genet.36.043002.091635
- Busch, A., and Hertel, K. J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscipl. Rev.* 3, 1–12. doi: 10.1002/wrna.100
- Carvalho, T., Martins, S., Rino, J., Marinho, S., and Carmo-Fonseca, M. (2017). Pharmacological inhibition of the spliceosome subunit SF3b triggers exon junction complex-independent nonsense-mediated decay. *J. Cell Sci.* 130, 1519–1531. doi: 10.1242/jcs.202200
- 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
- Chen, Y. I., Moore, R. E., Ge, H. Y., Young, M. K., Lee, T. D., and Stevens, S. W. (2007). Proteomic analysis of *in vivo*-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* 35, 3928–3944. doi: 10.1093/nar/gkm347
- Cohen-Krausz, S., Sperling, R., and Sperling, J. (2007). Exploring the architecture of the intact supraspliceosome using electron microscopy. *J. Mol. Biol.* 368, 319–327. doi: 10.1016/j.jmb.2007.01.090
- Corrionero, A., Minana, B., and Valcarcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev.* 25, 445–459. doi: 10.1101/gad.2014311
- Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell* 5, 779–787. doi: 10.1016/s1097-2765(00)80318-4
- Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X. C., Newman, A. J., et al. (2017). Structure of a spliceosome remodelled for exon ligation. *Nature* 542, 377–380. doi: 10.1038/nature21078
- Frankenstein, Z., Sperling, J., Sperling, R., and Eisenstein, M. (2012). A unique spatial arrangement of the snRNPs within the native spliceosome emerges from *In silico* studies. *Structure* 20, 1097–1106. doi: 10.1016/j.str.2012.03.022
- Fu, X. D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992). General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11224–11228. doi: 10.1073/pnas.89.23.11224
- Han, S. P., Tang, Y. H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.* 430, 379–392. doi: 10.1042/BJ20100396
- 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.* 31, 793–802. doi: 10.1128/MCB.01117-10
- Heinrich, B., Zhang, Z., Raitskin, O., Hiller, M., Benderska, N., Hartmann, A. M., et al. (2009). Heterogeneous nuclear Ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich Regions in Pre-mRNA. *J. Biol. Chem.* 284, 14303–14315. doi: 10.1074/jbc.M901026200
- Hochberg-Laufer, H., Schwed-Gross, A., Neugebauer, K. M., and Shav-Tal, Y. (2019b). Uncoupling of nucleo-cytoplasmic RNA export and localization during stress. *Nucleic Acids Res.* 47, 4778–4797. doi: 10.1093/nar/gkz168
- Hochberg-Laufer, H., Neufeld, N., Brody, Y., Nadav-Eliyah, S., Ben-Yishay, R., and Shav-Tal, Y. (2019a). Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLoS Genet.* 15:e1008459. doi: 10.1371/journal.pgen.1008459
- Huranova, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., et al. (2010). The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J. Cell Biol.* 191, 75–86.

- Irimia, M., and Blencowe, B. J. (2012). Alternative splicing: decoding an expansive regulatory layer. *Curr. Opin. Cell Biol.* 24, 323–332. doi: 10.1016/j.ccb.2012.03.005
- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* 3, 576–583. doi: 10.1038/nchembio.2007.18
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., et al. (2013). Function of alternative splicing. *Gene* 514, 1–30. doi: 10.1016/j.gene.2012.07.083
- Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. *Nucleic Acids Res.* 43, 8258–8267. doi: 10.1093/nar/gkv740
- Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., et al. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.* 3, 570–575. doi: 10.1038/nchembio.2007.16
- Kotzer-Nevo, H., de Lima Alves, F., Rappsilber, J., Sperling, J., and Sperling, R. (2014). Supraspliceosomes at defined functional states present portray the pre-assembled nature of the pre-mRNA processing machine in the cell nucleus. *Int. J. Mol. Sci.* 15, 11637–11664. doi: 10.3390/ijms150711637
- Lee, Y., and Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. doi: 10.1146/annurev-biochem-060614-034316
- Lin, S., and Fu, X. D. (2007). SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.* 623, 107–122.
- Lo, C. W., Kaida, D., Nishimura, S., Matsuyama, A., Yashiroda, Y., Taoka, H., et al. (2007). Inhibition of splicing and nuclear retention of pre-mRNA by spliceostatin A in fission yeast. *Biochem. Biophys. Res. Commun.* 364, 573–577. doi: 10.1016/j.bbrc.2007.10.029
- Long, J. C., and Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* 417, 15–27. doi: 10.1042/BJ20081501
- Markus, M. A., Heinrich, B., Raitskin, O., Adams, D. J., Mangs, H., Goy, C., et al. (2006). WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Exp. Cell Res.* 312, 3379–3388. doi: 10.1016/j.yexcr.2006.07.008
- Martins, S. B., Rino, J., Carvalho, T., Yoshida, M., Klose, J. M., et al. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. *Nat. Struct. Mol. Biol.* 18, 1115–1123. doi: 10.1038/nsmb.2124
- Medalia, O., Typke, D., Hegerl, R., Angenitzki, M., Sperling, J., and Sperling, R. (2002). Cryoelectron microscopy and cryoelectron tomography of the nuclear pre-mRNA processing machine. *J. Struct. Biol.* 138, 74–84. doi: 10.1016/s1047-8477(02)00027-8
- Miriami, E., Angenitzki, M., Sperling, R., and Sperling, J. (1995). Magnesium cations are required for the association of U small nuclear ribonucleoproteins and SR proteins with pre-mRNA in 200 S large nuclear ribonucleoprotein particles. *J. Mol. Biol.* 246, 254–263. doi: 10.1006/jmbi.1994.0081
- Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. (1998). A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. *J. Mol. Biol.* 283, 383–394. doi: 10.1006/jmbi.1998.2078
- Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198. doi: 10.1146/annurev-biochem-060614-034242
- Papasaikas, P., and Valcarcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem. Sci.* 41, 33–45. doi: 10.1016/j.tibs.2015.11.003
- Plaschka, C., Newman, A. J., and Nagai, K. (2019). Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harb. Perspect. Biol.* 11:a032391. doi: 10.1101/cshperspect.a032391
- Poser, I., Sarov, M., Hutchins, J. R., Hériché, J. K., Toyoda, Y., Pozniakovskiy, A., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* 5, 409–415. doi: 10.1038/nmeth.1199
- Raitskin, O., Angenitzki, M., Sperling, J., and Sperling, R. (2002). Large nuclear RNP particles—the nuclear pre-mRNA processing machine. *J. Struct. Biol.* 140, 123–130. doi: 10.1016/s1047-8477(02)00541-5
- Raitskin, O., Cho, D. S., Sperling, J., Nishikura, K., and Sperling, R. (2001). RNA editing activity is associated with splicing factors in InRNP particles: the nuclear pre-mRNA processing machinery. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6571–6576. doi: 10.1073/pnas.111153798
- Roybal, G. A., and Jurica, M. S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic Acids Res.* 38, 6664–6672. doi: 10.1093/nar/gkq494
- Schmidt, U., Basyuk, E., Robert, M. C., Yoshida, M., Villemin, J. P., Auboeuf, D., et al. (2011). Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J. Cell Biol.* 193, 819–829. doi: 10.1083/jcb.201009012
- Sebbag-Sznajder, N., Raitskin, O., Angenitzki, M., Sato, T. A., Sperling, J., and Sperling, R. (2012). Regulation of alternative splicing within the supraspliceosome. *J. Struct. Biol.* 177, 152–159. doi: 10.1016/j.jsb.2011.11.005
- Shav-Tal, Y., and Zipori, D. (2002). PSF and p54(nrb)/NonO—multi-functional nuclear proteins. *FEBS Lett.* 531, 109–114. doi: 10.1016/s0014-5793(02)03447-6
- Shefer, K., Sperling, J., and Sperling, R. (2014). The supraspliceosome—a multi-task-machine for regulated pre-mRNA processing in the cell nucleus. *Comput. Struct. Biotechnol. J.* 11, 113–122. doi: 10.1016/j.csbj.2014.09.008
- Shepard, P. J., and Hertel, K. J. (2009). The SR protein family. *Genome Biol.* 10:242. doi: 10.1186/gb-2009-10-10-242
- Shi, Y. (2017a). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. doi: 10.1038/nrm.2017.86
- Shi, Y. (2017b). The spliceosome: a protein-directed metallo-ribozyme. *J. Mol. Biol.* 429, 2640–2653.
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006
- Spann, P., Feinerman, M., Sperling, J., and Sperling, R. (1989). Isolation and visualization of large compact ribonucleoprotein particles of specific nuclear RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 86, 466–470. doi: 10.1073/pnas.86.2.466
- Sperling, J., Azubel, M., and Sperling, R. (2008). Structure and function of the Pre-mRNA splicing machine. *Structure* 16, 1605–1615. doi: 10.1016/j.str.2008.08.011
- Sperling, J., and Sperling, R. (2017). Structural studies of the endogenous spliceosome - The supraspliceosome. *Methods* 125, 70–83. doi: 10.1016/j.jymeth.2017.04.005
- Sperling, R. (2017). The nuts and bolts of the endogenous spliceosome. *Wiley Interdiscipl. Rev. RNA* 8:e1377. doi: 10.1002/wrna.1377
- Sperling, R., Koster, A. J., Melamed-Bessudo, C., Rubinstein, A., Angenitzki, M., Berkovitch-Yellin, Z., et al. (1997). Three-dimensional image reconstruction of large nuclear RNP (InRNP) particles by automated electron tomography. *J. Mol. Biol.* 267, 570–583. doi: 10.1006/jmbi.1997.0898
- Sperling, R., Sperling, J., Levine, A. D., Spann, P., Stark, G. R., and Kornberg, R. D. (1985). Abundant nuclear ribonucleoprotein form of CAD RNA. *Mol. Cell Biol.* 5, 569–575. doi: 10.1128/mcb.5.3.569
- Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326. doi: 10.1016/s0092-8674(00)80925-3
- Takemura, R., Takeiwa, T., Taniguchi, I., McCloskey, A., and Ohno, M. (2011). Multiple factors in the early splicing complex are involved in the nuclear retention of pre-mRNAs in mammalian cells. *Genes Cells* 16, 1035–1049. doi: 10.1111/j.1365-2443.2011.01548.x
- Valcarcel, J., and Green, M. R. (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends Biochem. Sci.* 21, 296–301.
- Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi: 10.1016/j.cell.2009.02.009
- Wilkinson, M. E., Lin, P. C., Plaschka, C., and Nagai, K. (2018). Cryo-EM studies of Pre-mRNA splicing: from sample preparation to model visualization. *Annu. Rev. Biophys.* 47, 175–199. doi: 10.1146/annurev-biophys-070317-033410

- Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3:a003707. doi: 10.1101/cshperspect.a003707
- Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *C* 75, 1061–1070. doi: 10.1016/0092-8674(93)90316-i
- Yan, C., Wan, R., and Shi, Y. (2019). Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* 11:a032409. doi: 10.1101/cshperspect.a032409
- Yang, Y.-H. J., Markus, A. M., Mangs, H. A., Raitskin, O., Sperling, R., and Morris, B. J. (2013). ZRANB2 localizes to supraspliceosomes and influences the alternative splicing of multiple genes in the transcriptome. *Mol. Biol. Rep.* 40, 5381–5395. doi: 10.1007/s11033-013-2637-9
- Yitzhaki, S., Miriami, E., Sperling, J., and Sperling, R. (1996). Phosphorylated Ser/Arg-rich proteins: limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8830–8835. doi: 10.1073/pnas.93.17.8830
- Yoshimoto, R., Kaida, D., Furuno, M., Burroughs, A. M., Noma, S., Suzuki, H., et al. (2017). Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. *RNA* 23, 47–57. doi: 10.1261/rna.058065.116
- Zahler, A. M., and Roth, M. B. (1995). Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *PNAS* 92, 2642–2646. doi: 10.1073/pnas.92.7.2642

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

Copyright © 2020 Sebbag-Sznajder, Brody, Hochberg-Laufer, Shav-Tal, Sperling and Sperling. 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.



Innovative Therapeutic and Delivery Approaches Using Nanotechnology to Correct Splicing Defects Underlying Disease

Marc Suñé-Pou^{1†}, María J. Limeres^{2†}, Cristina Moreno-Castro³,
Cristina Hernández-Munain⁴, Josep M. Suñé-Negre¹, María L. Cuestas² and
Carlos Suñé^{3*}

¹ Drug Development Service (SDM), Faculty of Pharmacy, University of Barcelona, Barcelona, Spain, ² Institute of Research in Microbiology and Medical Parasitology (IMPaM), Faculty of Medicine, University of Buenos Aires-CONICET, Buenos Aires, Argentina, ³ Department of Molecular Biology, Institute of Parasitology and Biomedicine "López-Neyra" (IPBLN-CSIC), Granada, Spain, ⁴ Department of Cell Biology and Immunology, Institute of Parasitology and Biomedicine "López-Neyra" (IPBLN-CSIC), Granada, Spain

OPEN ACCESS

Edited by:

Rosanna Asselta,
Humanitas University, Italy

Reviewed by:

Dario Balestra,
University of Ferrara, Italy
Mauricio Fernando Budini,
University of Chile, Chile

*Correspondence:

Carlos Suñé
csune@ipb.csic.es

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 17 February 2020

Accepted: 16 June 2020

Published: 14 July 2020

Citation:

Suñé-Pou M, Limeres MJ,
Moreno-Castro C,
Hernández-Munain C,
Suñé-Negre JM, Cuestas ML and
Suñé C (2020) Innovative Therapeutic
and Delivery Approaches Using
Nanotechnology to Correct Splicing
Defects Underlying Disease.
Front. Genet. 11:731.
doi: 10.3389/fgene.2020.00731

Alternative splicing of pre-mRNA contributes strongly to the diversity of cell- and tissue-specific protein expression patterns. Global transcriptome analyses have suggested that >90% of human multiexon genes are alternatively spliced. Alterations in the splicing process cause missplicing events that lead to genetic diseases and pathologies, including various neurological disorders, cancers, and muscular dystrophies. In recent decades, research has helped to elucidate the mechanisms regulating alternative splicing and, in some cases, to reveal how dysregulation of these mechanisms leads to disease. The resulting knowledge has enabled the design of novel therapeutic strategies for correction of splicing-derived pathologies. In this review, we focus primarily on therapeutic approaches targeting splicing, and we highlight nanotechnology-based gene delivery applications that address the challenges and barriers facing nucleic acid-based therapeutics.

Keywords: splicing, RNA, gene therapy and therapeutic delivery, siRNAs, ASOs, SMART, gene editing, nanoparticle

INTRODUCTION

The genome is the complete set of DNA that contains all the information necessary for the development and survival of an organism. In humans, the genome is contained in 23 chromosome pairs, comprising approximately 21,000 protein-coding genes and slightly over 3 billion DNA base pairs in total. Although the human genome was sequenced approximately fifteen years ago, there is still disagreement regarding the number of genes due to inconsistencies in the available databases (Willyard, 2018). Regardless of the final count, the numbers of protein-coding genes are similar between humans and worms (~21,000 and ~19,000, respectively), and the number found in flies is not drastically lower (~14,000); however, humans are more complex than these other organisms. The explanation for this apparent paradox may be related to the large predicted number of proteins encoded by the human genome (possibly more than 100,000) as a result of regulatory processes such as alternative transcription, splicing, 3'-end formation, translation and

posttranslational modifications. To date, there is no solid evidence (based on, for example, mass spectrometry) to support the existence of this level of complexity in the human proteome. Thus, ambitious projects are being designed to identify and characterize protein variant isoforms for each protein-coding gene (Baker et al., 2017).

Alternative splicing (AS), the phenomenon by which a single precursor (pre-) messenger RNA (mRNA) can generate alternative mRNAs to yield proteins with related or different functions, expands the protein information encoded by the genome. Global transcriptome analyses have estimated that 95–100% of multiexon genes undergo AS (Pan et al., 2008; Wang et al., 2008). The best-known example of the considerable transcriptome diversity resulting from this process is the *Drosophila* Down syndrome cell adhesion molecule (*Dscam*) gene, which may give rise to 38,016 cell-surface proteins through AS (Wojtowicz et al., 2007). The functional genome-wide consequences of AS are exemplified by the finding that distinct alternative isoforms encoded by a single gene exhibit distinct protein interaction profiles (Yang et al., 2016). Each of these protein isoforms can be further processed through posttranslational modifications to yield many more distinct proteoforms (Smith et al., 2013) harboring new functions. Recent technological and bioinformatics advances will help to unambiguously decipher the specific sequence and amount of each RNA molecule synthesized by a given cell (Hardwick et al., 2019). In addition, AS approaches specifically designed for single-cell RNA sequencing (scRNAseq) data are emerging, and these approaches may greatly improve our understanding of isoform usage at the single-cell level (Chen et al., 2019).

Alternative splicing in eukaryotes has considerable impacts on a variety of biological pathways; therefore, it is not surprising that AS is a highly orchestrated process involving multiple protein–protein and protein–RNA interactions. The spliceosome, the multiprotein complex that performs the splicing reaction, is composed of a core of five uridine-rich small nuclear RNAs (snRNAs; termed U1, U2, U4, U5, and U6) and 200 other proteins. The spliceosome assembles on pre-mRNA to remove noncoding introns through two sequential transesterification reactions: the branching and exon ligation steps. Detailing all the steps of the AS process is beyond the scope of this review, and comprehensive reviews on the molecular choreography of pre-mRNA splicing have been published elsewhere (Wahl et al., 2009; Matera and Wang, 2014; Wan et al., 2019). In this review, we briefly summarize some aspects of splicing regulation before turning toward advances in therapies and nanodelivery systems targeting splicing for the treatment of human disease.

The major forms of AS include exon skipping, alternative 3' and 5' splice site (SS) usage, intron retention, and mutual exon exclusion (Figure 1). Other events that generate different transcript isoforms include alternation of initial exons due to alternative promoter usage and alternation of terminal exons due to alternative polyadenylation. Recognition of exon/intron boundaries for correct intron removal by the splicing machinery requires the presence of several sequence elements on pre-mRNA, including the 5' and 3' SSs, the branch point sequence (BPS), and the polypyrimidine (Py) tract. In addition to these

core SS motifs, other *cis*-regulatory elements that recruit specific RNA-binding proteins that either activate or repress the use of adjacent SSs contribute to the fine-tuning and specificity of this pre-mRNA processing event. These sites, known as exonic splicing enhancers (ESEs) or silencers (ESSs) and intronic splicing enhancers (ISEs) or silencers (ISSs), recruit specific *trans*-acting proteins such as heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine/arginine (SR) proteins (Wu and Maniatis, 1993). AS can also be regulated at the levels of transcription and chromatin structure, adding complexity to the molecular mechanisms that govern splicing control. Two models have been proposed to explain the link between transcription and splicing. The first model, known as the recruitment model, involves the recruitment of splicing factors to pre-mRNA through RNA polymerase II (RNA Pol II) (McCracken et al., 1997). In the second model, the kinetic model, the transcript elongation rate influences AS (de la Mata et al., 2003; Montes et al., 2012; Dujardin et al., 2014; Fong et al., 2014). Chromatin structure, DNA methylation, histone marks, and nucleosome positioning also impact AS by affecting transcription and/or cotranscriptional splicing. An excellent review covering the different levels at which AS is regulated has been published previously (Naftelberg et al., 2015).

AS AND DISEASE

Defects in core spliceosome components, *trans*-acting splicing regulatory factors, *cis*-regulatory signals, and the transcription rate or changes in chromatin structure or marks can cause multiple pathologies as a result of misprocessing of pre-mRNA, highlighting the importance of RNA processing (Figure 2). To date, 23,868 mutations responsible for human inherited disease that have been reported in the Human Gene Mutation Database (HGMD, accessed in February 2020) have consequences for mRNA splicing (accounting for 8.7% of heritable disease-causing mutations) (Stenson et al., 2012). This number is likely an underestimate, since most of the reported mutations have been identified by genomic DNA sequencing without consideration that missense, nonsense and synonymous changes can affect splicing, as reported previously for coagulation factor IX exon 5 (Tajnik et al., 2016). It has been estimated that up to 50% of mutations that lead to heritable disease occur as a result of errors in the RNA splicing process or its regulation (Lopez-Bigas et al., 2005; Taylor and Lee, 2019). While we wait for functional testing of predicted mutations, the development of new and effective predictive algorithms for splicing effect analysis is critical (Anna and Monika, 2018).

Recent advances in the development of novel technologies and tools, such as next-generation sequencing and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein-9 nuclease (Cas9) genome editing technology, have greatly expanded the available information regarding RNA missplicing events and their association with diseases. These technologies have demonstrated the presence of many naturally occurring genetic variants that affect AS and lead to phenotypic variability and disease susceptibility among humans (Park

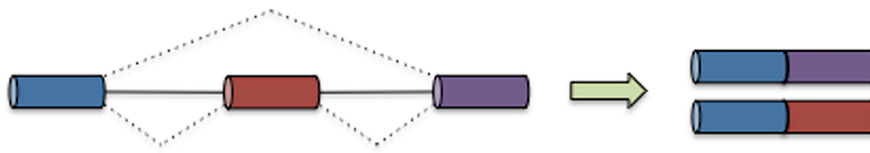
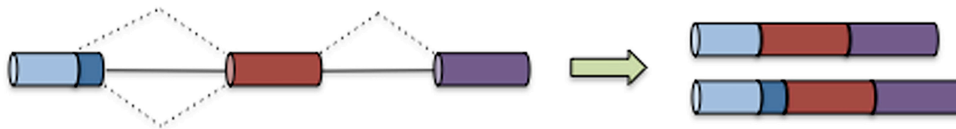
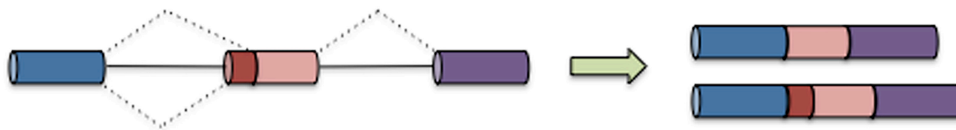
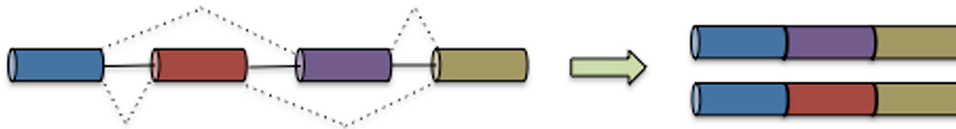
A Exon skipping**B Alternative 5' SS usage****C Alternative 3' SS usage****D Intron retention****E Mutually exclusive exons**

FIGURE 1 | Types of AS. In the graphs, exons are represented by boxes, and introns are represented by lines. Dashed lines indicate AS events. The five main types of AS are illustrated: exon skipping (**A**), alternative 5' SS usage (**B**), alternative 3' SS usage (**C**), intron retention (**D**), and mutually exclusive exons (**E**).

et al., 2018). For example, 371 splicing quantitative trait loci (sQTLs), including sQTLs in known type 2 diabetes-associated genes or in genes associated with beta cell function and glucose metabolism in human pancreatic islets, have been identified; these sQTLs may aid in elucidation of individual susceptibility to type 2 diabetes (Fadista et al., 2014). Recently, Takata et al., analyzed RNA-seq data on human brain tissues from more than 200 individuals in combination with genotype data and identified approximately 1,500 sQTLs throughout the genome. Interestingly, these researchers observed significant enrichment of epigenetic mark variants that may influence transcriptional activation and AS. In a comparative analysis of genome-wide association study (GWAS) data, many of the observed variants were found to be associated with various human diseases, particularly schizophrenia (Takata et al., 2017).

Mutations that occur in genes encoding fundamental components of the splicing machinery have been described in many splicing-related diseases. However, the frequency of these mutations is low, probably because their effects are incompatible

with life. Disease-associated mutations that occur within introns lead to intron retention or to exon skipping upstream or downstream of the mutated SSs without affecting the coding sequence. In contrast, exonic mutations may or may not affect the coding sequence depending on the type of mutation (silent *versus* missense or nonsense) and can also alter the splicing pattern. Thus, mutations in introns or exons may disrupt RNA secondary structure or disrupt or create *de novo* cryptic SSs or *de novo* splicing silencers and enhancers, leading to dysregulation of AS. Mutations or quantitative changes in proteins with regulatory functions during the splicing process can also lead to aberrant splicing, affecting many RNA transcripts at the same time (Havens et al., 2013).

With regard to cancer, genomic studies have identified frequent and recurrent mutations in genes that code for pre-mRNA splicing factors in both hematological malignancies (e.g., myelodysplastic syndrome [MDS], acute myeloid leukemia, and chronic lymphocytic leukemia) (Yoshida et al., 2011) and solid malignancies (e.g., breast cancer, lung cancer, pancreatic cancer

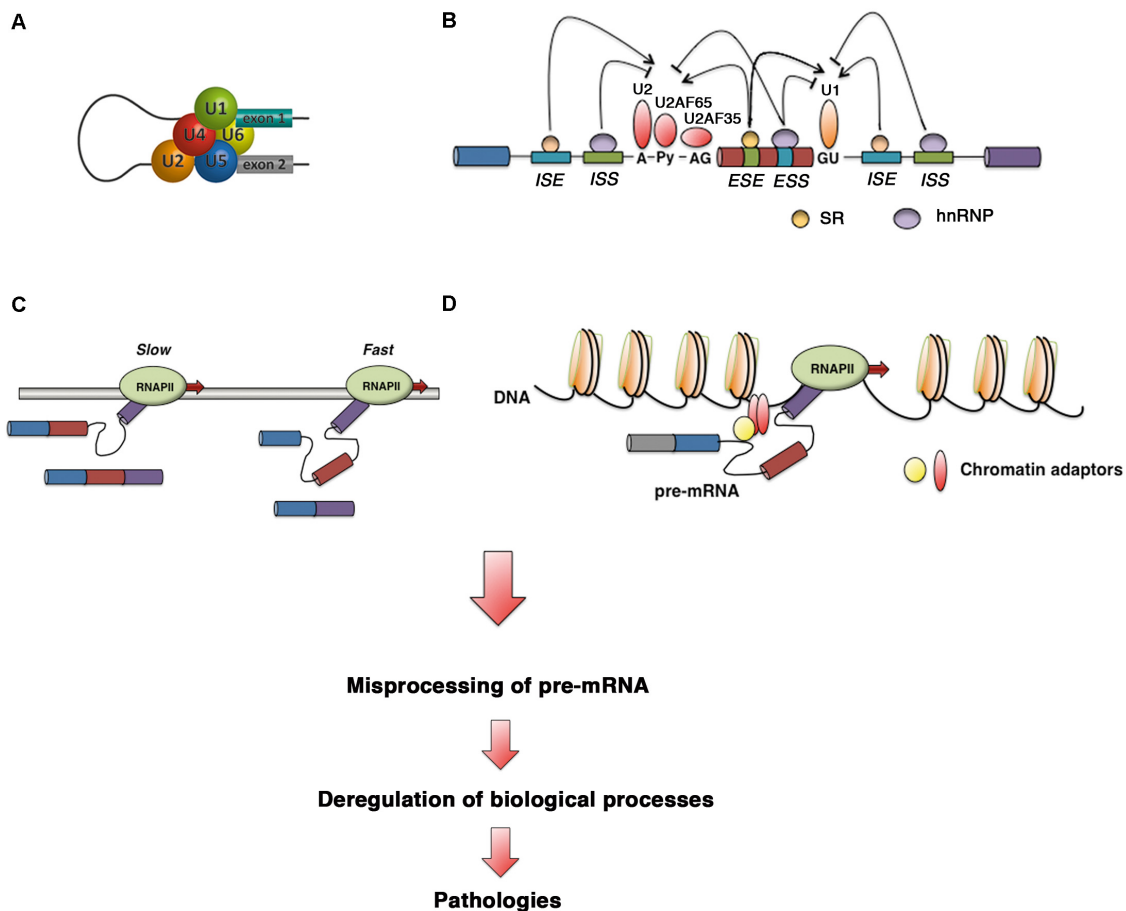


FIGURE 2 | Defective control in pre-mRNA splicing leads to human diseases. Diagram showing how defects in **(A)** core spliceosome components, **(B)** *trans*-acting splicing regulatory factors and *cis*-regulatory signals, **(C)** the transcription rate, and **(D)** chromatin structure or marks can disrupt splicing regulation to cause human disease via misprocessing of pre-mRNA.

and uveal melanoma) (Imielinski et al., 2012; Harbour et al., 2013; Bailey et al., 2016; Nik-Zainal et al., 2016). These findings suggest a potential relationship between certain spliceosome gene mutations and carcinogenesis. For MDS, *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* are the four most commonly mutated splicing factor genes, although mutations in other splicing factor genes have also been observed (Taylor and Lee, 2019). Although the underlying mechanisms and contributions of splicing factors in cancer pathogenesis have not been elucidated, and although more work is needed to understand the splicing alterations observed in cancer cells, these data identify novel opportunities for development of splicing-based cancer therapies.

Recent advances in the treatment of some diseases have led to improvements in patient prognosis and life expectancy. For example, spinal muscular atrophy (SMA) type 1, which is considered to be most serious at an early age, can currently be treated with Zolgensma®. Zolgensma® is a new gene therapy-based drug approved by the United States Food and Drug Administration (FDA) that improves the quality of life and life expectancy of infants with SMA type 1. Although this treatment can cure this deadly inherited disease, the Swiss multinational

corporation Novartis AG has established a sale price of 2.1 million dollars for a single intravenous administration. This drug is by far the most expensive pharmacological treatment in existence today.

Identification of splicing mutations has significantly advanced our understanding of how splicing dysregulation contributes to disease pathogenesis and of how splicing, a key pre-mRNA processing event, can be targeted for therapeutic applications. In **Supplementary Table 1**, we provide a list of the most frequent splicing-related human diseases that could be targeted for gene therapy. To learn more, readers are directed to several comprehensive reviews covering human diseases caused by RNA missplicing that have been published elsewhere (Cieply and Carstens, 2015; Dagueuet et al., 2015; Chabot and Shkreta, 2016; Scotti and Swanson, 2016).

THERAPEUTIC APPROACHES

Designing effective therapeutic strategies to overcome the consequences of aberrant splicing events on disease states

remains a major challenge. Gene therapy has emerged as a promising pharmacotherapeutic option for patients with diseases of genetic origin. Hence, targeting of aberrant RNA splicing is a logical approach for directly correcting disease-associated splicing alterations without affecting the genome. Other approaches, such as targeting splicing reactions to disrupt the expression of disease-related proteins or targeting exon junctions mutated mRNA to disrupt protein coding, can be used to reframe and rescue protein expression (Havens et al., 2013).

Several strategies have been designed to manipulate the splicing process, including spliceosome-mediated RNA trans-splicing (SMaRT) and the use of antisense oligonucleotides (ASOs), bifunctional oligonucleotides, small-molecule compounds, and modified snRNAs (Figure 3). All of these approaches have been used to correct the effects of RNA misprocessing. In recent years, genome editing techniques involving zinc finger (ZF) proteins (ZFPs), transcription activator-like nucleases (TALENs) and CRISPR/Cas9 systems have become new treatment avenues for correction of splicing defects (Hsu et al., 2014; Fernandez et al., 2017; Knott and Doudna, 2018).

RNA Editing Approaches

Antisense Oligonucleotides (ASOs)

Antisense oligonucleotides strategies use short synthetic single-stranded DNA molecules that are complementary to a specific pre-mRNA sequence to alter the splicing process. ASO binding to a target pre-mRNA in the nucleus sterically blocks the recruitment of *trans*-acting proteins to the pre-mRNA sequence at the target site (Figure 3A).

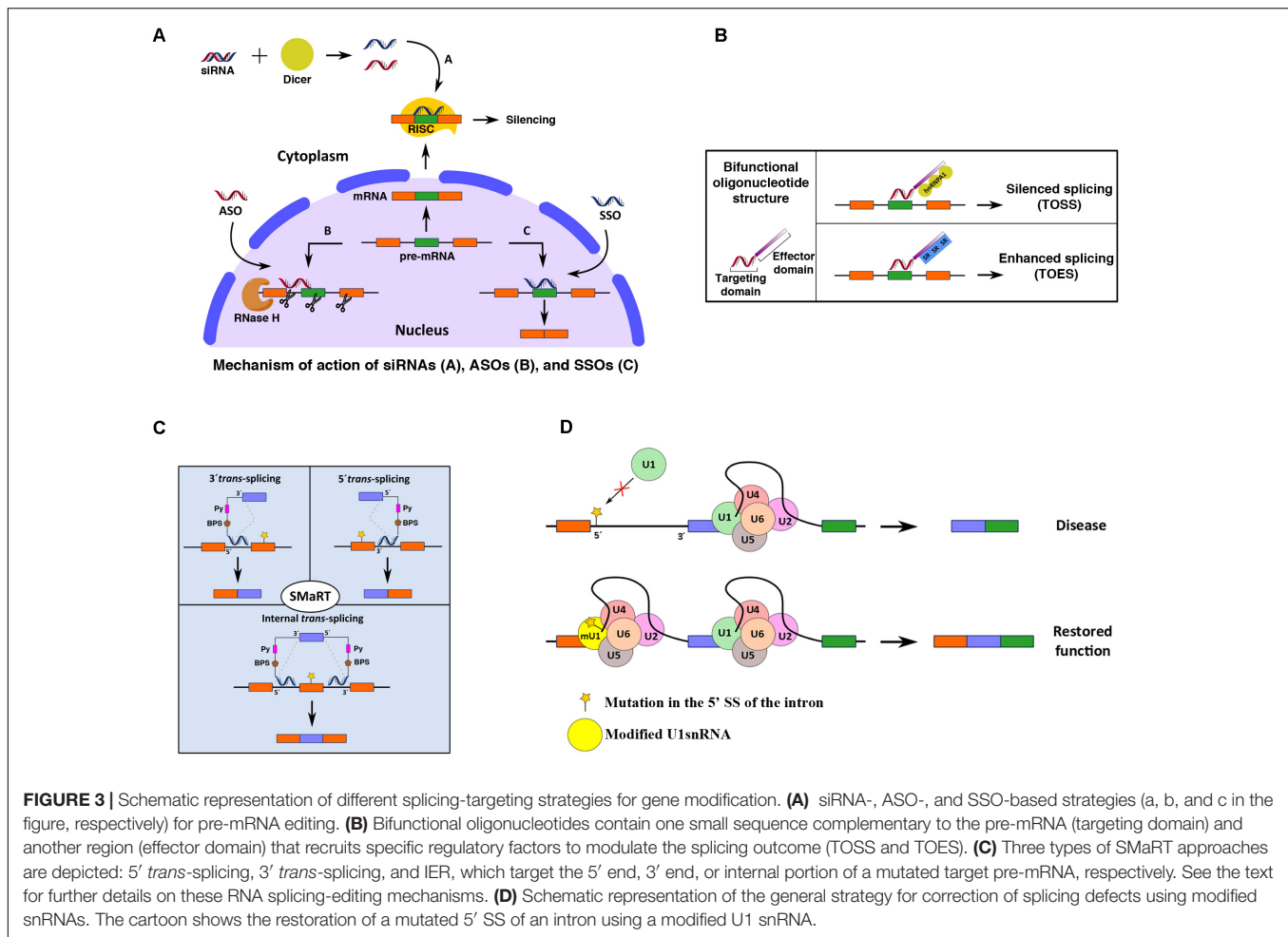
ASOs can be designed to target (i) the SS, thereby redirecting splicing to an adjacent site; (ii) auxiliary sequences (enhancers or silencer elements) within the immature transcript, thereby modifying the outcome of the splicing reaction (by blocking or promoting splicing); and (iii) RNA bases, thereby stabilizing/destabilizing regulatory structures and modifying the splicing outcome (Havens et al., 2013; Sune-Pou et al., 2017).

Splicing-related ASOs that act according to the first two mechanisms mentioned above by promoting or redirecting splicing are also called splice-switching oligonucleotides (SSOs). These short, 15- to 30-nucleotide-long sequences sterically block important motifs in pre-mRNA (i.e., SSs and/or regulatory sequences) to prevent RNA–RNA base-pairing or protein–RNA binding interactions between spliceosome components and pre-mRNA without promoting degradation of the RNA transcript while altering splicing outcomes (Havens and Hastings, 2016). The nucleotides of an SSO are chemically modified (e.g., into morpholino antisense oligomers) such that the RNA-cleaving enzyme RNase H is not recruited to degrade the pre-mRNA–SSO complex (Summerton, 1999). This property differs from the RNase H activity exerted by conventional ASOs, which inhibits gene expression by degrading the target pre-mRNA. Chemical modifications of ASOs are also crucial because they stabilize the ASOs *in vivo* and improve their cellular uptake, release and binding affinity for their targeted RNA sequences; unmodified oligonucleotides are highly susceptible to degradation by circulating nucleases and are excreted by the kidneys. Examples of these chemical modifications include changes to the phosphate

backbones and/or sugar components of the oligonucleotides, such as the use of a phosphorothioate backbone (first-generation ASOs) (Eckstein, 2014), the use of locked nucleic acid chemistry for bridging of the sugar furanose ring (Campbell and Wengel, 2011), alterations at 2' positions of the ribose sugar ring (2'-O-methylation [2'-OMe] and 2'-O-methoxyethylation [2'-MOE]) (second-generation ASOs) (Geary et al., 2001), and the addition of phosphorodiamidate morpholinos (third-generation ASOs) (Summerton, 1999).

The clinical application of this technology has resulted in the commercialization of VitraveneTM (fomivirsen), which, in 1998, became the first ASO approved by the FDA for the treatment of AIDS-related cytomegalovirus retinitis; MacugenTM (pegaptanib), approved by the FDA in 2004 for the treatment of neovascular age-related macular degeneration; and KynamroTM (mipomersen), approved by the FDA in 2013 for the treatment of homozygous familial hypercholesterolemia. These products have been withdrawn from the market for commercial reasons owing to an overall small patient population and competing alternative drugs, such as statins in the case of familial hypercholesterolemia (Sharma and Watts, 2015). The first published report on the use of ASOs as a splicing-targeting therapeutic tool was published by Dominski and Kole in 1993. These authors restored correct splicing in thalassemic pre-mRNA by using a 2'-OMe ASO (Dominski and Kole, 1993). Since then, many ASO strategies have been designed to modify splicing for the treatment of several diseases, and some of them are currently in clinical trials. Only two ASOs are already FDA approved (Supplementary Table 2). Exondys 51TM (eteplirsén) was the first drug in its class to be approved by the FDA (in September 2016) under the Accelerated Approval Program to treat patients with Duchenne muscular dystrophy (DMD). Exondys 51TM belongs to the third generation of phosphorodiamidate morpholino ASOs and is specifically indicated for patients who have a confirmed mutation of the dystrophin gene amenable to exon 51 skipping. This mutation affects approximately 13% of the population with DMD. Recently, Vyondys 53TM (golodirsén), a DMD drug that is highly similar to eteplirsén except that involves exon 53 skipping rather than exon 51 skipping, was denied by the FDA in August 2019 because of the risk of infections related to intravenous infusion ports and renal toxicity seen in preclinical studies. A similar decision was reached in January 2016 for the promising drug KyndrisaTM (drisapersen). This drug was intended for the treatment of patients with DMD amenable to exon 51 skipping but failed to demonstrate substantial effectiveness. The pharmaceutical company BioMarin invested over \$66 million in the development of drisapersen; as a consequence of the denial, BioMarin also discontinued the development of three follow-on products of drisapersen, BMN 044, BMN 045, and BMN 053. These products were in mid-stage trials for specific forms of the muscle-wasting disease.

In December 2016, the FDA approved SpinrazaTM (nusinersen), the first drug approved for the treatment of SMA the first drug approved for the treatment of SMA (Lorson et al., 1999; Monani et al., 1999). SpinrazaTM is a 2'-OMe phosphorothioate ASO that targets an intron 7 internal SS within SMN2 pre-mRNA, inducing exon 7 inclusion and producing a functional SMN protein.



Notably, because ASOs generally do not cross the blood-brain barrier, repeated intrathecal nusinersen delivery is required. This requirement is highly disadvantageous and makes administration challenging, especially for infants (Verma, 2018).

Investigational ASOs for Huntington's disease (HD), amyotrophic lateral sclerosis (ALS) and transthyretin (TTR) amyloidosis, such as RG6042, tofersen and inotersen, respectively, are currently in phase III clinical trials. RG6042 (previously known as IONIS-HTTRx) reduces the concentration of mutant huntingtin (HTT) levels in the cerebrospinal fluid of patients with HD without causing serious adverse events (Tabrizi et al., 2019). Tofersen (previously known as BIIB067) targets superoxide dismutase (SOD1) in ALS patients, reducing SOD1 concentrations in spinal fluid to preserve motor neurons and slow the progression of the disease. Inotersen is a 2'-MOE-modified ASO that reduces the production of TTR and improves disease course and quality of life in early hereditary TTR amyloidosis polyneuropathy (ATTR) (Mathew and Wang, 2019).

The development of the ASO milasen is an example of how cutting-edge medicine can be used with great speed for patient-customized treatment of a rare and fatal neurodegenerative

disease (ceroid lipofuscinosis 7, CLN7, a form of Batten's disease). Researchers at Boston Children's Hospital identified a novel mutation in a 6-year-old girl, designed and produced an ASO, and obtained FDA approval for its clinical deployment in less than one year (Kim et al., 2019).

ASOs are also applicable to cancer treatment. For instance, Dewaele et al., used ASO-mediated exon skipping to decrease the expression of *MDM4*, a splice isoform produced in cancer cells (Dewaele et al., 2016). Similarly, Hong et al. preclinically and clinically evaluated a chemically modified ASO termed AZD9150 that targets the *STAT3* gene, a transcriptional activator and oncogenic mediator of the JAK-STAT signaling pathway (Hong et al., 2015). Ross et al. described the use of an ethyl-containing ASO (AZD4785) to downregulate *KRAS* mRNA, which is mutated in approximately 20% of human cancers, and demonstrated its efficacy in preclinical *KRAS* mutant lung cancer models (Ross et al., 2017).

Antisense oligonucleotides have been validated as therapeutic agents; however, because of the high cost associated with these products, improvements must be made to prevent health insurance companies from denying patients access. For instance, Exondys 51TM is priced at \$300,000 per patient per year even though the efficacy of this drug is controversial.

Bifunctional Oligonucleotides

Bifunctional oligonucleotides are ectopic modulators of AS used to control the patterns of splicing of specific genes. In brief, these oligonucleotides contain two parts: (i) an antisense portion targeting a specific sequence and (ii) a nonhybridizing tail or effector domain that recruits acting factors (targeted oligonucleotide enhancer of splicing [TOES] or targeted oligonucleotide silencer of splicing [TOSS]) (Brousseau et al., 2014) (**Figure 3B**). Oligonucleotides containing binding site sequences for the splicing repressor hnRNP A1/A2 have been used to reprogram AS via TOSS. The oligonucleotides were positioned upstream of a 5' SS to interfere with U1 snRNP binding and repress SS use (Villemaire et al., 2003). Dickson et al. used a TOSS with hnRNPA1 tails to block the inclusion of exon 8 in *SMN2*, thereby favoring exon 7 inclusion and restoring the functionality of the protein (Dickson et al., 2008). Similarly, bifunctional TOES, whose tail of enhancer sequences recruits activating proteins such as positively acting SR proteins, has been used to increase the splicing of refractory exon 7 in *SMN2* in fibroblasts derived from patients with SMA (Skordis et al., 2003; Owen et al., 2011). This approach is mechanistically different than ASO approaches, although both can be used for stimulating the inclusion of exon 7 in SMA. Whereas ASOs such as SpinrazaTM block the binding of splicing factors to intron 7, causing exon 7 inclusion, bifunctional TOSS/TOES is intended to direct exon 7 inclusion and thus restore protein expression.

Small Interfering RNAs (siRNAs)

In 1998, another related technology emerged with the discovery of the siRNA pathway, which can be used to silence the expression of genes (Fire et al., 1998). The discovery of RNA interference (RNAi), in which double-stranded RNA (dsRNA) is hybridized with a specific mRNA sequence to induce its silencing or degradation, was a major scientific breakthrough for which Craig Mello and Andrew Fire were recognized with the Nobel Prize in Physiology and Medicine in 2006. This discovery revolutionized the way scientists study gene function and offered an innovative strategy for the treatment of diseases, particularly those of genetic origin, as demonstrated for the first time by Elbashir et al. (2001). Through this strategy, administration of synthetic 21- to 25-nucleotide duplexes with overhanging 3' ends (siRNAs) can be used to suppress the expression of endogenous and heterologous genes (**Figure 3A**). Although many classes of small RNAs have emerged, three main categories are widely recognized: siRNAs; microRNAs, or miRNAs; and piwi-interacting RNAs, or piRNAs. These RNA types differ in structure, biological roles, associated effector proteins and origins (Dana et al., 2017). Physiologically, in cells, siRNAs help to maintain genomic integrity by preventing the action of foreign nucleic acids, including those of viruses, transposons and retrotransposons and transgenes, while miRNAs act as posttranscriptional endogenous gene regulators (Meister and Tuschl, 2004). piRNAs have been implicated in the silencing of retrotransposons at both the posttranscriptional and epigenetic levels as well as other genetic elements in germlines, particularly those activated during spermatogenesis (Siomi et al., 2011). Upon delivery into cells, siRNAs are bound by a multiprotein component complex, known as the RNA-induced silencing

complex (RISC), in the cytoplasm; the two strands are then separated, and the strand with the RISC hybridizes with the target mRNA. After that step, Argonaute-2 (Ago2), a catalytic component of the RISC, drives mRNA cleavage (Dana et al., 2017). Since siRNA-mediated targeting of aberrant splicing isoforms is widely used as an RNAi technology, only siRNA approaches will be discussed in detail in this section (**Figure 3A**).

These siRNA approaches can be used to target aberrant splicing isoforms for therapeutic applications (Sune-Pou et al., 2017). Indeed, siRNAs targeting exonic/intronic sequences close to alternative exon or exonic/intronic junction sequences can induce degradation of alternatively spliced and aberrant mRNAs without affecting the expression of normal mRNAs. Such targeting approaches have been used for diseases such as Ullrich congenital muscular dystrophy (UCMD) (Bolduc et al., 2014), growth hormone deficiency (GHD) type II diseases (Ryther et al., 2004) and several cancers. In the context of cancer, it has been observed that the occurrence of specific splice variants is increased during tumorigenesis and that the splicing regulatory machinery is abnormal in many malignant cells (Hayes et al., 2006). Bolduc et al. designed different siRNAs targeting the most frequent mutation that causes exon 16 skipping in the *COL6A3* gene and tested the siRNAs *in vitro* in UCMD-derived dermal fibroblasts. These siRNAs resulted in specific knockdown of the mutant allele and increased the abundance and quality of collagen VI matrix production (Bolduc et al., 2014). Similarly, Ryther et al. used siRNAs to specifically degrade exon 3-skipped transcripts in a GHD type II disease (Ryther et al., 2004). Finally, Hayes et al. have shown that siRNA-mediated downregulation of SR protein kinase 1 (SRPK1), which is significantly upregulated in tumors of the pancreas, breast and colon, decreases cell proliferation and increases apoptosis. Moreover, the sensitivity of tumor cells to chemotherapeutic agents such as gemcitabine and cisplatin increases upon treatment with this siRNA (Hayes et al., 2006).

In October 2016, the development of revusiran, an siRNA designed for the treatment of ATTR amyloidosis, was suspended after a randomized double-blind placebo-controlled phase III trial of its efficacy and safety, ENDEAVOUR, demonstrated that the siRNA caused greater mortality than a placebo (18 of 206 enrolled patients). However, in August 2018, the FDA approved OnpatroTM (patisiran), the first siRNA-based drug for the treatment of hereditary TTR-mediated amyloidosis (hATTR) (**Supplementary Table 2**). Hence, siRNA-based technology has shown promising therapeutic results and the possibility of translation into clinical use.

Small-Molecule Compounds

Small-molecule compounds can also be used to modulate RNA expression. Some molecules are capable of binding specific three-dimensional RNA structures, thereby preventing their translation or function. Furthermore, these compounds can also modify splicing factor activity (by affecting posttranslational modifications of splicing factors) or directly alter splicing events. Compared with oligonucleotide-based therapeutics, these compounds are easier to deliver to target sites and normally have lower toxicity profiles. However, small-molecule compounds frequently act through unknown mechanisms,

resulting in a lack of information, and have less target specificity than other therapeutic formulations, thus potentially exhibiting more nonspecific and off-target effects. While oligonucleotides (ASOs, TOSS, TOES and siRNAs) can very specifically and efficiently modulate particular RNA targets upon complementary recognition of the RNA sequences, small molecules can recognize specific three-dimensional structures and overcome known oligonucleotide drawbacks, e.g., poor pharmacological properties.

Some small molecules have already been approved for use in clinical practice for applications other than splicing defect correction. For example, digoxin and other prescribed cardiotonic steroids, routinely used in the treatment of heart failure, have been described as modulators of AS (Stoilov et al., 2008). Pentamidine and Hoechst 33258 also modulate AS in myotonic dystrophy (DM) by disrupting MBNL1 binding to CUG *in vitro* and *in vivo* (Warf et al., 2009; Parkesh et al., 2012). The histone deacetylase sodium butyrate, which is known to upregulate the expression of splicing factors, has been demonstrated to increase *CFTR* transcript levels, leading to activation of *CFTR* channels and restoration of their function in *CFTR*-derived epithelial cells (Nissim-Rafinia and Kerem, 2006). It has also been reported that the plant cytokinin kinetin improves *IKBKAP* mRNA splicing in patients with familial dysautonomia (FD) (Axelrod et al., 2011).

Some small-molecule splicing modulators have been evaluated in clinical trials for the treatment of solid tumors and leukemia. For instance, E7107 (pladienolide B) is a splicing modulator whose preferential cytotoxicity is positively influenced by some antiapoptotic *BCL2* family genes, such as *BCL2A1*, *BCL2L2* and *MCL1*. Furthermore, Aird et al. have reported that combinations of E7107 and *BCLxL* (*BCL2L1*-encoded) inhibitors enhance cytotoxicity in cancer cells (Aird et al., 2019). Notably, E7107 was the first compound of a new class of anticancer agents targeting the spliceosome. Specifically, E7107 interacts with subunit 1 of SF3b to block the normal splicing of oncogenes. Unfortunately, the development of E7107 was suspended after phase I clinical trials due to an unacceptable profile of adverse events. Herboxidienes, spliceostatin A, meayamycin B and sudemycin D6/K have also been shown to exhibit antitumoral activity *in vivo* by targeting the SF3b subunit of the spliceosome (Lin, 2017). Recently, Carabet et al. reported VPC-80051 as the first small-molecule inhibitor of hnRNPA1, which plays an important role in cancer by controlling the transcriptional levels of the oncoprotein c-Myc (Carabet et al., 2019). Thus, small molecules that selectively inhibit hnRNP A1-RNA interactions can be designed for the treatment of tumors expressing cancer-specific alternatively spliced proteins. Similarly, highly specific inhibitors of the RNA helicase Brr2, which is an essential component of the spliceosome, have been designed for therapeutic purposes (Iwatani-Yoshihara et al., 2017).

Recently, the Massachusetts-based pharmaceutical company Skyhawk Therapeutics invested \$100 million in the development of small-molecule compounds capable of correcting misspliced exon skipping related to cancer and neurological diseases. Novartis and Roche have also independently developed two

different splicing-modulating compounds for the treatment of SMA: branaplam and risdiplam, respectively. Both molecules enhance exon 7 inclusion to increase the levels of functional SMN protein. Branaplam, also known as LMI070, is an orally available drug that was designed by Novartis using a high-throughput phenotypic screening approach with approximately 1.4 million compounds. Currently, this molecule is in a phase II clinical trial that is expected to be completed in July 2020.

Risdiplam (Ratni et al., 2018) is a brain-penetrant orally administered drug that is currently being evaluated in patients with SMA in four multicenter clinical trials (NCT02913482, NCT02908685, NCT03032172, and NCT03779334). This drug is a splicing modulator that increases exon 7 inclusion in the *SMN2* gene, thereby increasing the levels of SMN protein throughout the organism. Furthermore, this drug is being studied for use in patients of all age ranges with SMA types 1, 2, and 3.

All these examples demonstrate the applicability of small molecules for splicing event modulation and suggest that these molecules are useful as complements and alternatives to oligonucleotides.

SMaRT

Spliceosome-mediated RNA trans-splicing is a gene-reprogramming system based on the *trans*-splicing process that can be used for therapeutic applications. The *trans*-splicing methodology is designed to correct aberrant mRNAs by replacing the entire coding sequence upstream or downstream of a target SS. Three different components are involved: the target mRNA, the spliceosome machinery and the pre-*trans*-splicing molecule (PTM), also known as the RNA *trans*-splicing molecule (RTM). The first two components are present in cells, while the third must be provided exogenously. *Trans*-splicing is induced between the exogenous RNA and the endogenous pre-mRNA, producing a chimeric RNA with the wild-type sequence (without mutation/s). To achieve successful correction, the PTM must be correctly designed with the following regions: a binding domain (complementary to the pre-mRNA), a splicing domain (incorporating 5' and 3' SS, intronic BPS and Py sequences) and a coding domain (containing the wild-type coding region) (Wally et al., 2012; Berger et al., 2016; **Figure 3C**).

Depending on the targeted region of the pre-mRNA, SMaRT can be divided into (i) 5'-*trans*-splicing, which targets the 5' portion; (ii) 3'-*trans*-splicing, which targets the 3' portion; and (iii) internal exon replacement (IER), which targets an internal portion of the pre-mRNA (**Figure 3C**). SMaRT approaches have been used in models of cystic fibrosis (CF) (Song et al., 2009), hemophilia A (Chao et al., 2003), SMA (Coady and Lorson, 2010), retinitis pigmentosa (RP) (Berger et al., 2015), frontotemporal dementia with parkinsonism-17 (FTDP-17) and tauopathies (Rodriguez-Martin et al., 2005, 2009). The results have shown that this technology can successfully reprogram gene expression and offers promising gene therapy applications. However, *trans*-splicing approaches require the use of vectors for the delivery of the PTM into cells. Therefore, selection of a good delivery vector is critical for future treatment approaches based on this technology (Wally et al., 2012). SMaRT offers multiple advantages

as a gene therapy tool; however, it needs to be better understood and optimized in order to increase its overall efficiency.

Modified snRNAs

Exon-specific U1 snRNAs (ExSpe U1s) are modified U1 snRNAs complementary to intronic regions downstream of the 5' SS that can be used to eliminate the skipping of some exons caused by different mutations (**Figure 3D**). ExSpe U1s have been tested in different models and have shown potential for use in therapeutic applications (Dal Mas et al., 2015). In one study, different ExSpe U1s were tested for the treatment of SMA, CD and hemophilia B in *SMN2* exon 7, *CFTR* exon 12 and *F9* exon 5 models, respectively (Fernandez Alanis et al., 2012). Similar approaches have been reported for ATP8B1 deficiency (van der Woerd et al., 2015), FD (Donadon et al., 2018b), Sanfilippo syndrome type C (Matos et al., 2014), Fanconi anemia (Mattioli et al., 2014), RP (Tanner et al., 2009), thalassemia (Gorman et al., 2000), severe coagulation factor VII (FVII) deficiency (hemophilia A, HA) (Pinotti et al., 2008; Donadon et al., 2018a; Balestra et al., 2019b), *CDKL5*-deficiency disorder (CDD) (Balestra et al., 2019a), Seckel syndrome (Scalet et al., 2017), and hereditary tyrosinemia type I (HT1) (Scalet et al., 2018).

Other modified versions of spliceosomal snRNAs have also been tested for their usefulness in restoration of base pairing to the mutated SS. For example, combined treatment with mutation-adapted U1 and U6 snRNAs has been used to correct mutation-induced splice defects in exon 5 of the *BBS1* gene (Schmid et al., 2013); this gene is implicated in Bardet-Biedl syndrome, which causes retinal degeneration and developmental disabilities. Another example of a modified oligonucleotide is U7 snRNA, which participates in histone pre-mRNA maturation by recognizing the sequence of the histone 3' untranslated region. Changes in the target sequence can be introduced to convert this snRNA into an antisense tool capable of blocking splicing signals and inducing exon skipping or inclusion (Brun et al., 2003). This strategy has been used to design artificial U7 snRNAs to enhance exon 23 skipping of mutated dystrophin in DMD (Brun et al., 2003; Goyenvalle et al., 2004).

The modified snRNA approach is based on engineered variants of small coding genes and has various advantages. Its main advantages are the possible exploitation of virtually any viral vector and the fact that, based on its molecular mechanisms, it does not alter the physiological expression of the target gene.

Genome Editing Approaches

ZFPs

Zinc finger proteins are powerful and widely studied tools for efficient establishment of targeted genetic modifications (Cristea et al., 2011). ZFPs designed for therapeutic use consist of ZF arrays in which every ZF element recognizes three bases of a DNA sequence through an α -helix structure (**Figure 4A**). Examples include zinc finger nucleases (ZFNs) that cleave DNA and zinc finger transcription factors (ZFP-TF) that modulate gene expression (Wild and Tabrizi, 2017).

Zinc finger nucleases were the first endonucleases designed for genome editing. In one such application, the association of ZF domains with the DNA cleavage domain of the restriction

protein *FokI* leads to breakage of a specific region in the DNA sequence (Porteus and Baltimore, 2003). To this end, two different ZFNs must recognize adjacent sequences separated by a spacer sequence where the break will be located. After this step, DNA repair pathways such as nonhomologous end joining (NHEJ) or homologous recombination (HR) with a codelivered exogenous DNA template lead to the establishment of a modified sequence in which the targeted mutation is corrected (Cristea et al., 2011). Unfortunately, the use of ZFNs has some limitations, including high cost, off-target effects due to low specificity, and inappropriate interaction between domains.

Zinc finger nucleases can be used as therapeutic tools to correct genetic mutations associated with splicing-related diseases. The main advantage of ZFNs is that the correction of mutations is permanent; thus, continuous administration is not needed. For instance, Ousterout et al., applied ZFNs for permanent deletion of exon 51 of the dystrophin gene for the treatment of DMD (Ousterout et al., 2015). Other authors have designed ZFPs targeting CAG repeats to decrease the levels of mutant HTT for therapeutic purposes in HD (Garriga-Canut et al., 2012; Zeitler et al., 2019).

The current and completed clinical trials of ZFN therapies (13 in total¹, accessed in February 2020) focus on three major areas: cancer (1 clinical trial), blood disorders (thalassemia, hemophilia B and sickle-cell disease: 3 clinical trials), infectious diseases (human immunodeficiency virus [HIV]: 7 clinical trials) and orphan diseases (mucopolysaccharidosis: 2 clinical trials).

TALENs

These genome editing tools are chimeric nucleases engineered by fusion of the DNA-binding domain of the bacterial protein TALE with the catalytic domain of the restriction endonuclease *FokI* (Schornack et al., 2006; **Figure 4B**). Recognition of a specific DNA sequence is performed by the binding domain, which is composed of monomeric tandem repeats of 33–35 conserved amino acids; within this domain is a variable region known as the repeat variable diresidue (RVD) located at residues 12 and 13. The RVD is responsible for binding to a specific nucleotide. Similar to ZFNs, TALENs work as pairs, recognizing sequences separated by 12–25 bp to promote cleavage of the DNA by the *FokI* endonuclease. The advantage of TALENs compared to ZFNs is decreased cytotoxicity due to reduced off-target effects (Sun and Zhao, 2013).

Fang et al. used TALENs *in vivo* for the treatment of β 654-thalassemia in a mouse model. The TALENs targeted the mutation site, corrected aberrant β -globin RNA splicing, and ameliorated the β -thalassemia phenotype in β 654 mice (Fang et al., 2018). TALENs have also been designed for the treatment of HD (Fink et al., 2016) and DMD (Li et al., 2015). The current trials on TALEN therapies (5 in total¹, accessed in February 2020) are focused entirely on cancer.

CRISPR/Cas9

The development of CRISPR/Cas9 technology is considered one of the most important breakthroughs of the last decade. This

¹<https://clinicaltrials.gov>

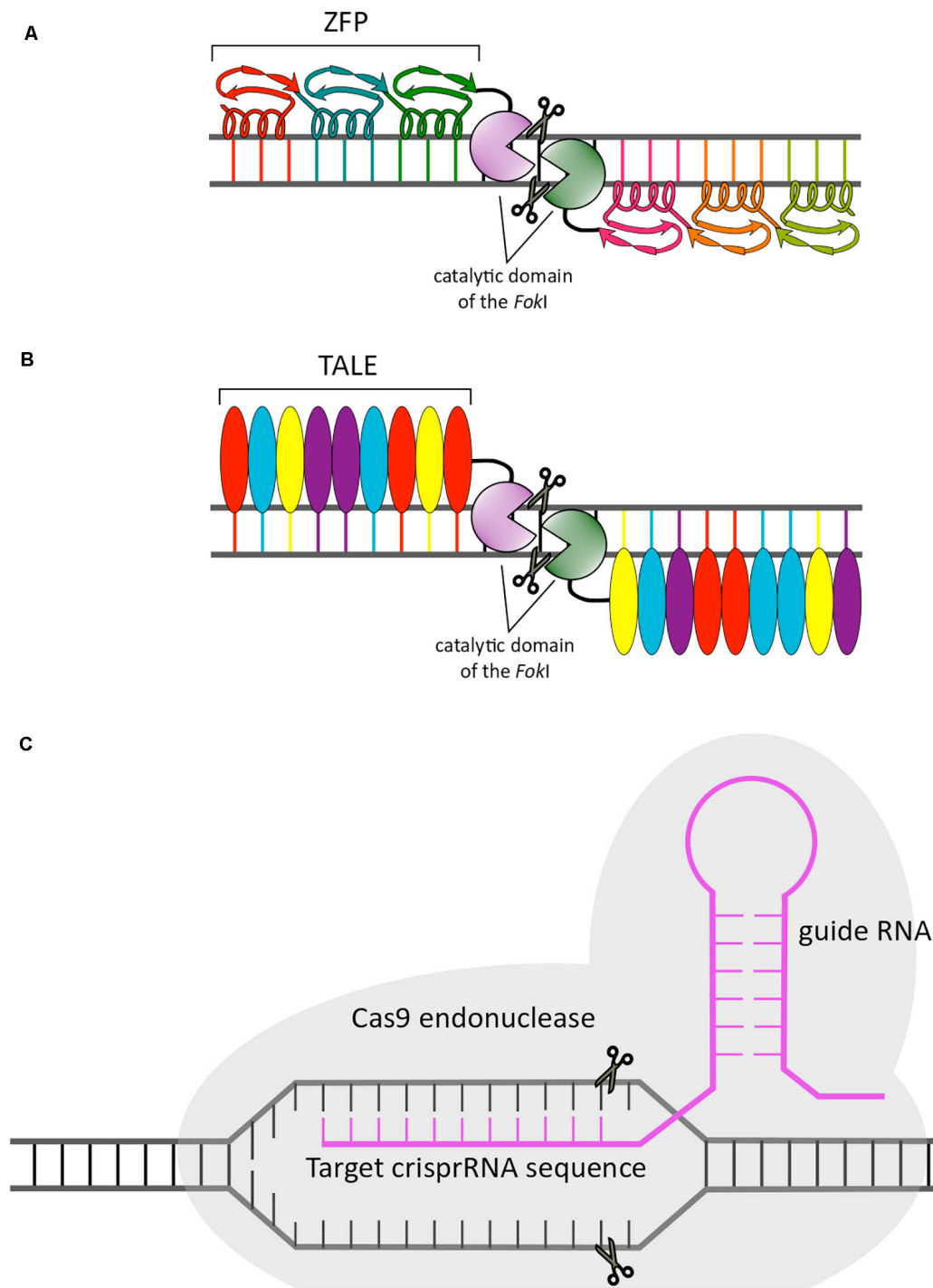


FIGURE 4 | Schematic representation of genome editing approaches. The three gene editing systems described in the main text are illustrated: ZFPs (**A**), TALENs (**B**), and CRISPR/Cas9 (**C**).

nucleic acid immune system was first discovered in bacteria and archaea more than thirty years ago (Ishino et al., 1987; Hermans et al., 1991; Mojica et al., 1993). In 2012, Jennifer Doudna and Emmanuelle Charpentier suggested that this system could be used for RNA-programmable genome editing (Jinek

et al., 2012). After that study, Feng Zhang and George Church performed the first *in vitro* studies in eukaryotes demonstrating the genome editing capacity of the CRISPR/Cas9 system in mouse and human cells (Cong et al., 2013; Mali et al., 2013). Since these seminal reports, many researchers have contributed

to the molecular understanding, technological development and medical applications of this gene editing system (Lander, 2016).

The CRISPR/Cas9 gene editing system requires administration of two main elements into a cell: (i) a small molecule of RNA (guide RNA [gRNA]) with a sequence complementary to that of the target mutation intended for editing and (ii) an endonuclease CRISPR-associated system (Cas) that allows the cleavage of the specific sequence under the guidance of gRNA binding (Figure 4C). Although many Cas9 orthologs have been investigated, the most widely used is Cas9 from *Streptococcus pyogenes*. In most eukaryotic cells and after cleavage, two repair pathways, NHEJ and homology-directed repair (HDR), are used to correct CRISPR-mediated breaks. The application areas of CRISPR/Cas9 technology go beyond genome editing, and many comprehensive reviews discussing the potential and specific applications of this system in science and medicine have been published (Adli, 2018 and references therein).

Several studies have used CRISPR/Cas9 to correct splicing-related defects. Yuan et al. used a CRISPR-guided cytidine deaminase to successfully correct mutations associated with splicing-related diseases. These researchers used this tool to restore the expression and function of the protein dystrophin in DMD patients (Yuan et al., 2018). Foltz et al. used CRISPR/Cas9 to generate corrected induced pluripotent stem cells (iPSCs) from fibroblasts with a mutation in the PRPF8 splicing factor from patients with RP. These researchers were also able to differentiate each of these clones into retinal pigment epithelial cells with a nearly normal phenotype, highlighting the power and utility of this genome editing tool (Foltz et al., 2018). Dastidar et al. used CRISPR/Cas9 to excise a CTG-repeat expansion of the *DMPK* gene, abnormal length of which leads to sequestration of muscle blind-like (MBLN) splicing factors, and achieved correction efficiencies of up to 90% in myotonic dystrophy type-1 (DM1) iPSCs. These results support the use of this tool in developing new therapies for the treatment of DM1 (Dastidar et al., 2018). CRISPR/Cas9 has also been used to treat the neurodegenerative disease X-linked dystonia-parkinsonism (XDP) by excising the SINE-VNTR-Alu (SVA) retrotransposition in intron 32 of the *TAF1* gene in multiple pluripotent stem cell-derived neuronal lineages. An XDP-specific transcriptional signature with normalized TAF1 expression levels was achieved in these cells (Aneichyk et al., 2018). Kemaladewey et al. demonstrated *in vivo* systemic delivery of an adeno-associated virus (AAV) carrying CRISPR/Cas9 genome editing components in a mouse model of congenital muscular dystrophy type 1A (MDC1A) to correct a pathogenic SS mutation in *LAMA2* pre-mRNA in order to include exon 2. The *LAMA2* gene encodes the $\alpha 2$ chain of the most abundant laminin isoform of the basal lamina (Laminin-211), and restoration of full-length *LAMA2* expression by CRISPR/Cas9 improves muscle histopathology and function (Kemaladewey et al., 2017). These observations, together with those described in a follow-up report by the same authors, validate the use of this gene editing technology as a therapeutic strategy for MDC1A (Kemaladewey et al., 2019). Thus far, trials using CRISPR therapies (27 in total¹, accessed in February 2020) have focused mainly on cancer. Very recently, Stadtmauer et al.

reported data from the first phase I clinical trial on cancer immunotherapy combined with CRISPR. The results of the trial demonstrated that it is feasible and safe to apply this technology for cancer immunotherapy (Stadtmauer et al., 2020).

Despite the targeting specificity of Cas9, off-target DNA cleavage activity can occur. In addition, many recent reports have suggested that CRISPR/Cas9 might unintentionally generate alternatively spliced products, large genomic deletions, translocations and inversions; this is a matter of concern that should be further evaluated prior to the clinical application of this technology (Smith et al., 2018).

A modification of the CRISPR/Cas9 system termed the base editing system (BEs) has expanded the arsenal of tools for genome modification. Unlike CRISPR, the BEs can introduce precise base changes without introducing double-strand breaks (DSBs) and, in the case of HDR, without the requirement of a template donor (Hess et al., 2017; Rees and Liu, 2018). Two classes of BEs have been developed: the cytosine BEs (CBEs) (Komor et al., 2016) and the adenine BEs (ABEs) (Gaudelli et al., 2017); these systems include cytidine deaminases and evolved *Escherichia coli* TadA, respectively. Osborn et al. applied the ABEs to an *in vitro* model of primary fibroblasts extracted from recessive dystrophic epidermolysis bullosa (RDEB) patients. Compared with HDR using a donor template, the ABEs efficiently corrected two *COL7A1* mutations with minimal indels (Osborn et al., 2020). Recently, Song et al. also systemically delivered an ABEs and a single gRNA (sgRNA) to edit a point mutation in the *FAH* gene in an adult mouse model of HTI. The results showed improvements in liver histology in ABEs-treated mice, and the correction of the point mutation was confirmed by sequencing, indicating restoration from the diseased to the normal phenotype *in vivo* (Song et al., 2020). The enormous potential of this technique is related to the fact that conversion of A-T to G-C base pairs in genomic DNA makes it possible to correct almost half of the 32,000 point mutations that cause genetic diseases (Gaudelli et al., 2017).

Some of the strategies presented in this section are feasible treatment options for several diseases. Although the potential of nucleic acids (DNA or RNA) as drugs immediately became obvious decades ago, the actual development of nucleic acid-based medicines has faced major and evident hurdles. For instance, nucleic acids are highly susceptible to degradation by endogenous nucleases. Some of these nucleic acids, such as short oligonucleotides in their native form, have a very short half-life, even before they are filtered out through the kidneys. Large DNA/RNA constructs with highly negative charges cannot cross the vascular endothelium, dense extracellular matrix and cell and nuclear membranes to reach their intracellular DNA, pre-mRNA or mRNA targets. Moreover, off-target effects of many DNA/RNA therapeutic tools can lead to devastating adverse reactions. Finally, some of these tools can be immunogenic. Although chemical modifications can improve pharmacokinetic and pharmacodynamic properties, the ability of these promising therapeutic tools to efficiently deliver DNA/RNA in order to modify sufficient numbers of cells for therapeutic benefits is still the limiting factor for the translation of preclinical models to standard clinical care. This ongoing challenge, which

is considered the Achilles heel of gene therapy (Somia and Verma, 2000), is beginning to be overcome through the use of nanotechnologies. These technologies use complexes of nucleic acids or encapsulate the nucleic acids in nonviral vectors, such as liposomes, lipids, and polymeric or inorganic nanoparticles, to enhance safe delivery to the target site. Next, we will focus on the application of nanotechnologies for gene delivery and discuss the advantages and problems associated with nanotechnology-based systems. Addressing the problems will dismantle the barriers facing nucleic acid-based therapeutics.

NANOMEDICINE

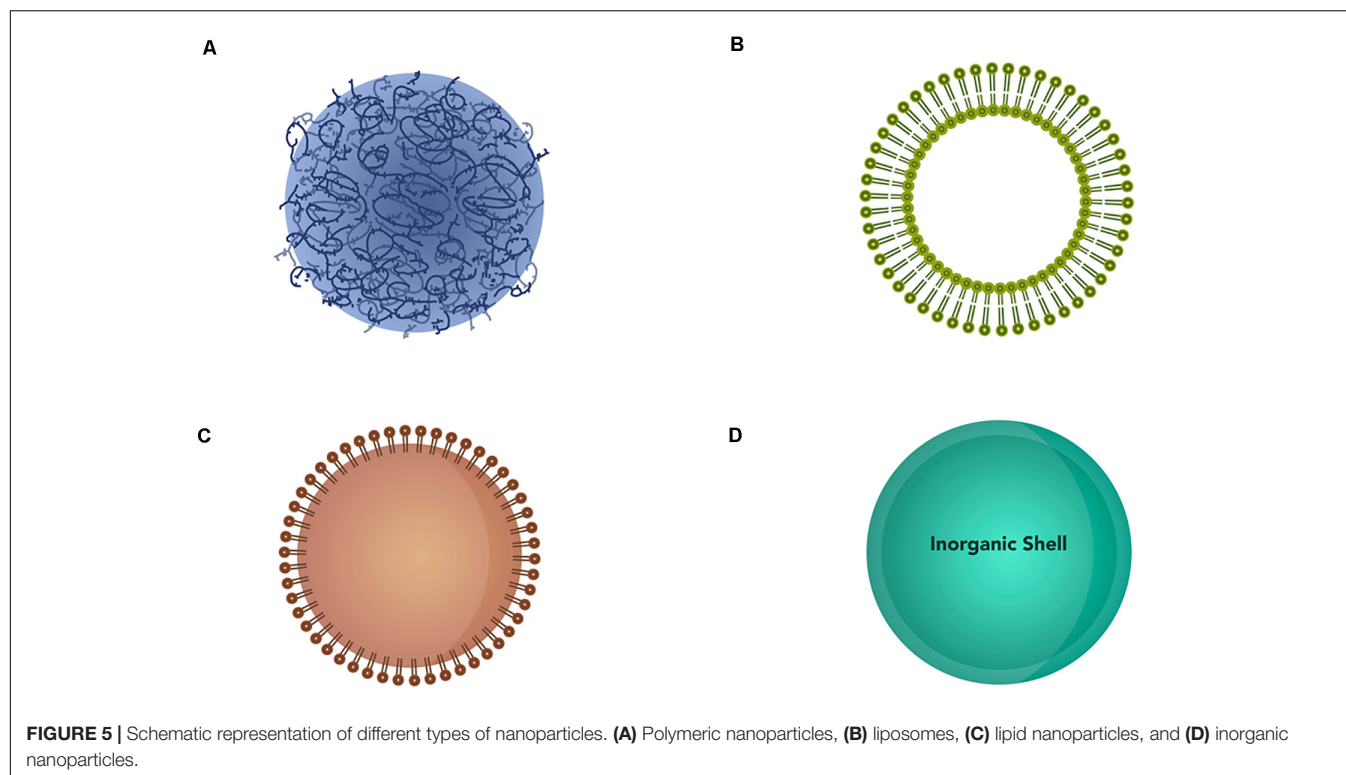
In recent decades, various vectors and tools have been developed for gene therapy. In addition, the advent of new gene editing therapies, such as CRISPR/Cas9, has sparked investigation into appropriate gene delivery systems, including viral vector and transposon-based vector systems. Nanostructures are nanoscale-sized particles capable of transfecting cells and releasing cargoes such as small molecules, DNA, RNA and peptides to exert pharmacological effects. These nonviral vectors have received considerable attention due to their advantages compared to viral systems, which have been the most common choices for gene delivery. Several good reviews have extensively explained the differences between these types of vectors (Chira et al., 2015; Ramamoorthi and Narvekar, 2015; Foldvari et al., 2016). The main advantage of nonviral gene delivery systems is their low immunogenicity, as high immunogenicity can impair viral transduction efficacy. Insertional mutagenesis is also a recognized safety concern associated with viral vectors intended for use in gene therapy (Hacein-Bey-Abina et al., 2003; David and Doherty, 2017), and viral integration is recognized as a common outcome of applications that utilize AAVs for genome editing (Hanlon et al., 2019). The major advantages of viral vectors include strong and prolonged transgene expression, broad cell tropism, and thorough understanding of viral gene function. Compared with viral systems, nanoparticle-mediated nucleic acid delivery systems have the advantages of weak immunogenicity, lack of integration and absence of potential for viral recombination, all of which translate to improved safety (Yin et al., 2014). The development processes and manufacturing capacity for clinical-grade nanoparticles are also advantages of nanoparticle-based methods *versus* viral methods. Nanotechnologies are applicable to a large cohort of patients (Paliwal et al., 2014), and several nanoparticle-based formulations are already on the market. For example, the siRNA-based drug Onpatro®, used for the treatment of the polyneuropathy hATTR in adults via inhibition of the production of the disease-causing protein, is encased in lipid nanoparticles for delivery into the body. However, the transfection efficiency of nanoparticle-based systems is comparatively poor, and poor transfection efficiency is the main limitation for this and other nonviral methods. For this reason, AAV vectors are the most commonly used vectors for nucleic acid delivery. A good overview of the current status of the clinical translation of viral and nonviral systems for gene therapy has recently been published (Kaemmerer, 2018).

Several formulations based on nanoparticles have demonstrated sustained expression of transported cargoes and long-term achievement of biological effects (Cohen et al., 2000; Shi et al., 2014). Nanoparticles can also achieve successful tissue-specific delivery of biomolecules through different strategies. For example, incorporation of specific antibodies into the nanoparticle surface has enabled effective targeting of nanoparticles to the brain and lung endothelium (Kolhar et al., 2013) for the treatment of many types of cancer, inflammation dysfunction, and infectious disease (Cardoso et al., 2012). Specific chemical components have also been incorporated into nanoparticles to increase delivery of biomolecules to targeted cells. Incorporation of phosphatidylserine (PS), cholesteryl-9-carboxynonanooate (9-CCN) (Maiseyeu et al., 2010), or folate into targeted nanoparticles (Krzyszton et al., 2017) has been shown to increase uptake by macrophages to help treat atherosclerosis and rheumatoid arthritis.

Nanoparticles exist in different forms and can be divided into different classes based on their compositions and properties: polymeric nanoparticles, liposomes, lipid nanoparticles, and inorganic nanoparticles (Figure 5).

Polymeric Nanoparticles

Polymeric nanoparticles are based on a polymeric matrix. The most frequently used polymers are poly-lactide-co-glycolide (PLGA), polyhydroxyalkanoates (PHAs) and CDs (CDs) (Zhang et al., 2018). PLGA is a biodegradable and biocompatible polymer formed by units of lactic acid and glycolic acid. This excipient is approved by the FDA and has been extensively used to develop nanoparticles (Zakharova et al., 2017). The grade of the polymer depends on the ratio between lactic acid and polyglycolic acid (PGA), which can affect the final characteristics of the nanoparticle. For example, polymers with higher glycolide contents have shorter deterioration times due to their more hydrophilic and amorphous characteristics. On the other hand, polymers with higher lactic content are more hydrophobic, thus exhibiting longer deterioration times (Schliecker et al., 2003). Frequently, PLGA is mixed with other polymers to improve the characteristics of the resulting nanoparticles. For example, polyethylenimine (PEI) is commonly incorporated to improve the transfection efficiency of nanostructures (Xie et al., 2016; Wang et al., 2018). PEI is a polycationic polymer capable of condensing DNA and RNA into stable nanostructures, primarily via electrostatic interactions. However, several studies have revealed that this polymer is cytotoxic, as elegantly discussed by Hunter more than a decade ago (Hunter, 2006). Thus, the translation of PEI-based nanoparticles to clinical applications is limited. Another polymer frequently mixed with PLGA is polyethylene glycol (PEG) (Zhang et al., 2014). PEG is a nonionic biocompatible polymer coated onto the surfaces of nanoparticles that prevents recognition and destruction of these carriers by the mononuclear phagocyte system (MPS), thereby increasing the plasma half-lives of the nanoparticles (Mustafa et al., 2017). Furthermore, PEGylation of nanoparticles improves their stability by reducing intermolecular aggregation and the accessibility of the target site (Gref et al., 2000). PLGA-based nanoparticles can also be functionalized with ligands such as



antibodies and Fab fragments to improve cellular targeting (Kennedy et al., 2018).

Polyhydroxyalkanoates are polyesters produced by microorganisms through, for example, bacterial fermentation of sugars or lipids. Because they are biodegradable and biocompatible polymers, PHAs have been used as bioimplant materials for medical and therapeutic applications for more than thirty years (reviewed in Zhang et al., 2018). Different types of PHAs can be used for nanoparticle formulation. There are 2 types of PHAs with different chain lengths: short-chain-length PHAs (scl-PHAs), which are composed of 3 to 5 carbon atoms, and medium-chain-length PHAs (mcl-PHAs), which are composed of 6 to 16 carbon atoms. There is also a subtype of PHAs that includes copolymers of scl-mcl PHAs of 4 to 12 carbon atoms (Hazer and Steinbuechel, 2007). PHA-based nanoparticles have been used to deliver biomolecules for anticancer (Lu et al., 2011; Fan et al., 2018; Radu et al., 2019) and antibacterial applications (Castro-Mayorga et al., 2014, 2016; Mukheem et al., 2018). Nanocarriers fabricated from PHA-grafted copolymers have also been developed for efficient siRNA delivery (Zhou et al., 2012); these formulations are safe siRNA carriers for gene therapy. A review discussing the use of PHA-based nanovehicles as therapeutic delivery carriers has been recently published (Lin, 2017).

CDs are cyclic oligosaccharides extensively used in pharmaceutical and biomedical applications. These molecules can be divided into 3 groups: α -CDs (6 subunits of glucose), β -CDs (7 subunits of glucose) and γ -CDs (8 subunits of glucose). CDs are biocompatible products approved by the FDA that are currently present in marketed formulations (Jambhekar and

Breen, 2016). The cyclic structure of CDs results in a hydrophobic lumen and a hydrophilic surface. This characteristic allows the use of CDs for multiple purposes, such as the vectorization of lipophilic drugs (Fine-Shamir et al., 2017). Furthermore, CDs can penetrate cells and release their cargoes through, for example, pH-dependent mechanisms (Tardy et al., 2016). CD-based nanoparticles have also been used for gene delivery. For example, Zuckerman et al. developed CD nanoparticles containing siRNAs for the treatment of chronic kidney diseases. Additionally, these researchers reported the functionalization of these nanoparticles with mannose or transferrin for enhanced nanoparticle uptake (Zuckerman et al., 2015).

Liposomes

Liposomes were some of the first nanostructures to be developed for drug delivery. Liposomes were developed in the 1960s and are currently present in marketed formulations such as Doxyl®, Myocet® and Caelix®. At present, no liposome-based marketed formulations for gene delivery exist. Liposomes are nanoscale particles with a lipid bilayer composition that forms a spherical structure inside an aqueous compartment. In aqueous solution, liposomes form colloidal dispersions. The main components of liposomes are phospholipids, such as phosphatidylcholine (PC), phosphatidylethanolamine (PE), PS, phosphatidylinositol (PI) and phosphatidyl glycerol (PG), and cholesterol, which can be incorporated into the phospholipid membrane to increase liposome stability (Bozzuto and Molinari, 2015). Other excipients can be used to improve the properties of liposomes or to endow them with new characteristics suitable for gene delivery. For example, dioleoylphosphatidylethanolamine (DOPE) is

used to produce pH-sensitive liposomes, and the cationic lipids N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTMA) and N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTAP) are used to formulate liposomes with cationic charges to facilitate the incorporation of DNA and RNA. In addition, polymers and carbohydrates, such as PEG and monosialoganglioside (GM1), can be incorporated into nanoparticle formulations to improve their *in vivo* half-lives and stability (Daraee et al., 2016b).

Many published studies have shown the efficacy of liposomes in delivering RNA or DNA into cells. For example, Dorrani et al. developed a liposome formulation with DOTAP and sodium cholate (NaChol) as edge-activators that is capable of efficiently delivering siRNA through skin layers after topical administration, demonstrating that liposomes are good candidates for the treatment of skin diseases such as melanoma (Dorrani et al., 2016). Qiao et al. developed a formulation incorporating mannosylated zwitterionic-based cationic liposomes (man-ZCLs) that increases the endosomal/lysosomal escape of nanostructures to enable administration of a DNA vaccine for HIV, providing a new, safe and effective HIV vaccination method that can be tested in future studies (Qiao et al., 2016). Recently, a four-component liposome formulation with DOTAP, DOPE, PEG, and cholesterol was used to transfect Cas9/sgRNA with high efficacy in order to knock out targeted genes *in vivo* (Hosseini et al., 2019).

Lipid Nanoparticles

Lipid nanocarriers are some of the most promising nonviral tools for gene therapy. Currently, the only medicine approved by the FDA and European Medicines Agency (EMA) that uses nanostructures to deliver RNA is Onpatro® (mentioned in section “Small Interfering RNAs (siRNAs)”); Onpatro® is a lipid nanoparticle-based drug product that transports patisiran, an siRNA molecule for the treatment of TTR amyloidosis (Rizk and Tuzmen, 2019).

Lipid nanoparticles were developed to address important drawbacks of other lipid-based systems, such as instability and the necessity for use of surfactants and other toxic substances; to increase loading capacity; and to resolve other problems related to manufacturing and scale-up processes (Muller et al., 2000). Lipid nanoparticles can be divided into 2 categories: solid lipid nanoparticles (SLNs) and nanostructured lipid carriers (NLCs) (Gordillo-Galeano and Mora-Huertas, 2018). Both types of nanoparticles use lipid excipients that are biocompatible and biodegradable. These structures are highly attractive for clinical applications due to their simple and inexpensive manufacturing processes that do not require organic solvents and can be easily scaled up; their high stability; and their ability to be administered through different routes, such as the parenteral, pulmonary, oral and topical routes (Uner and Yener, 2007).

Solid lipid nanoparticles have been developed for gene delivery since 2000. The structure of SLNs can help to protect the drug or RNA/DNA against external agents and can enable modified and/or targeted release (Muller et al., 2000). SLNs are manufactured with solid lipid excipients. The most common excipients include stearic acid, cholesterol

derivatives (e.g., cholesteryl oleate), glyceryl monostearate, glyceryl behenate, cetylpalmitate, glycolipids, tripalmitine and tristearin. Other essential excipients that are incorporated into these formulations are surfactants and cosurfactants, such as Pluronic® compounds (i.e., Pluronic F68), Poloxamer® compounds (i.e., Poloxamer 188), Brij®, Tween 80, and Span 20. Cationic molecules can also be incorporated in formulations to provide a positive surface charge (forming cationic SLNs [cSLNs]) in order to facilitate the formation of SLNplexes with DNA/RNA. The most commonly used cationic excipients are octadecylamine, benzalkonium chloride, cetrimide (DTAB), DOTAP, N,N-di-(β -stearoylethyl)-N,N-dimethyl-ammonium chloride (Esterquat 1 [EQ1]) and stearylamine (de Jesus and Zuhorn, 2015). To improve the efficacy of this type of vector, other excipients, such as protamine (Limeres et al., 2019), can be incorporated into the formulation. There are many examples of the use of SLNs for gene therapy. For example, Apaolaza et al. developed a formulation incorporating hyaluronic acid for transfection of cells with a plasmid containing the RS1 gene. Intravitreal administration of this formulation induced the expression of the protein retinoschisin in photoreceptors of Rslh-deficient mice, leading to structural improvements in degenerated retinas (Apaolaza et al., 2016). Rassu et al. designed SLNs capable of carrying *BACE1* siRNA to the brain after nasal administration for the treatment of Alzheimer's disease. These researchers formulated nanoparticles with RVG-9R, a type of cell-penetrating peptide (CPP) that facilitated the transcellular pathway in neuronal cells. Furthermore, the researchers coated the nanoparticles with chitosan, which provided extra protection to the siRNA and increased the mucoadhesiveness of the particles, thereby increasing the residence time in the nasal cavity (Rassu et al., 2017).

Nanostructured lipid carriers were first developed several years after SLNs. In contrast to SLNs, which involve solid lipids, NLCs involve liquid lipids; the use of liquid lipids increases the stability and drug loading capacity of the nanoparticles (Uner and Yener, 2007). The liquid lipids most commonly used to formulate NLCs are oleic acid and caprylic/capric triglycerides. Other lipids used are canola stearin and myristyl myristate. The other excipients are the same as those used for SLNs. A study by Taratula et al. has demonstrated the high loading capacity and gene delivery potential of NLCs. The authors developed a multifunctional NLC-based system containing a drug (paclitaxel or doxorubicin), two different types of siRNA, and a modified synthetic analog of luteinizing hormone-releasing hormone (LHRH) to increase specificity for local targeted delivery to lung tumors (Taratula et al., 2013). Similarly, Chen et al. have shown the capability of NLCs to coencapsulate plasmids and temozolomide, an anticancer drug. Those authors tested the system *in vitro* and *in vivo* for efficient delivery to malignant glioblastoma cells for the treatment of malignant gliomatosis cerebri (Chen et al., 2016).

Inorganic Nanoparticles

Inorganic nanoparticles include nanostructures that are manufactured using inorganic materials, such as gold, silicon

and iron oxide; carbon materials; layered double hydroxide (LDH); or calcium phosphate (Xu et al., 2006). The easy surface functionalization, good target delivery and controlled release of these nanoparticles are their main advantages. Some of the most widely used inorganic nanostructures are gold nanoparticles (AuNPs). This type of nanoparticle is used in the biomedical field for different applications, such as biodetection, biondiagnostics, and bioelectronic or therapeutic agent development. Among the advantages of these nanoparticles are their size- and structure-dependent visual and electronic characteristics and high surface/volume ratios and the capability to functionalize their surfaces due to their high affinities for different functional groups (Giljohann et al., 2010; Daraee et al., 2016a). One example of an AuNP-based gene delivery system was recently developed by Jia et al. (2017). The authors surface-conjugated AuNPs with thiol-modified antago-miR155, an RNA antagonist to a potent promoter of proinflammatory type 1 macrophage polarization (miR155) that plays an important role in diabetic cardiomyopathy. *In vivo* administration of the AuNP complex resulted in the incorporation of nucleic acids into macrophages via phagocytosis and led to reduced inflammation, reduced apoptosis and restoration of cardiac function. AuNPs have also been used to deliver Cas9 ribonucleoprotein and donor DNA *in vitro* and *in vivo* and to correct the DNA mutation in the dystrophin gene that causes DMD (Lee et al., 2017).

Oxide nanoparticles can be classified into two important groups: silicon oxide nanoparticles and iron nanoparticles (Xu et al., 2006). Mesoporous silica nanoparticles (MSNs) have been extensively investigated with regard to gene delivery. Indeed, RNA can be loaded onto the surfaces of small pore-sized MSNs, thereby enabling RNA delivery into cells. Furthermore, RNA and drugs can be loaded onto the same large pore-sized MSNs; thus, large pore-sized MSNs have the capacity to deliver two therapeutic agents at the same time. Sun et al. developed core-shell hierarchical MSNs (H-MSNs); doxorubicin can be loaded into the core mesopores, and siRNAs that downregulate the expression of P-gp to reverse multidrug resistance (MDR) can be loaded into the shell mesopores. The specific release of siRNA into the microtumor environment enables inhibition of MDR, and the subsequent release of doxorubicin enhances this effect (Sun et al., 2017).

Iron nanoparticles can be made of different materials, such as magnetite (Fe_3O_4) and maghemite (Fe_2O_3). The intrinsic magnetic properties of these nanoparticles can be used for gene or drug delivery. For example, an external magnetic field can be used to guide the nanoparticles to a specific zone of the body for drug release. Furthermore, a magnetic field can be applied to a cell culture dish to enhance cell transfection by magnetic nanoparticles in a procedure known as magnetofection (Estelrich et al., 2015). Despite the difficulties encountered in translating the magnetofection technique to clinical applications, advanced studies have demonstrated that iron nanoparticles can be applied for *ex vivo* delivery of chemically modified RNA (cmRNA), opening the door to continuing studies on gene therapy applications (Badiéyan et al., 2017).

CONCLUDING REMARKS

In 1993, the Nobel Prize in Physiology and Medicine was awarded to Phillip Sharp and Richard Roberts for their discovery of adenoviral RNA splicing (Berget et al., 1977; Chow et al., 1977). This discovery had notable consequences for elucidation of gene expression regulation and the evolution of eukaryotic cells. More than forty years after this seminal discovery, we have a deep understanding of the molecular mechanisms that control this important regulatory process, and we have recently begun to unravel the molecular links that connect faulty splicing with many human disorders. This knowledge has enabled the design of innovative therapeutic strategies intended to correct splicing defects. Many tools based on nucleic acid gene repair have been tested with positive results, and many more tools warrant further development. The field is moving notably quickly, but we have attempted to provide a general overview of the main developments. However, as important as it is to decipher the mechanisms that govern the connections between missplicing and pathologies and to apply these findings in the clinic, research on the safe transport of therapeutic biomolecules into cells and to their targets is equally important. Further development, characterization and testing of engineered technologies for targeted delivery and controlled release of DNA and RNA directly into cells with clinical applications are needed, as the demand for innovative nucleic acid delivery systems continues to grow. Nanoparticles possess considerable potential for use in the controlled delivery of therapeutic agents to specific target sites for splicing-based treatments. Conducting related research is a challenging task, as basic scientists must interact and collaborate with nanotechnology experts. Funding opportunities should emphasize such collaboration as a way forward for grant support.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing and revision of the manuscript and approved the submitted version.

FUNDING

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness (grant number BFU2017-89179-R to CS and grant number BFU2016-79699-P to CH-M). Support from the European Region Development Fund (ERDF [FEDER]) is also acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00731/full#supplementary-material>

REFERENCES

- Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9:1911.
- Aird, D., Teng, T., Huang, C. L., Pazolli, E., Banka, D., Cheung-Ong, K., et al. (2019). Sensitivity to splicing modulation of BCL2 family genes defines cancer therapeutic strategies for splicing modulators. *Nat. Commun.* 10:137.
- Alves, C. A., Benito-Vicente, A., Medeiros, A. M., Reeves, K., Martin, C., and Bourbon, M. (2018). Further evidence of novel APOB mutations as a cause of familial hypercholesterolaemia. *Atherosclerosis* 277, 448–456. doi: 10.1016/j.atherosclerosis.2018.06.819
- Anechik, T., Hendriks, W. T., Yadav, R., Shin, D., Gao, D., Vaine, C. A., et al. (2018). Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* 172, 897.e821–909.e281.
- Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268. doi: 10.1007/s13353-018-0444-7
- Apaolaza, P. S., Del Pozo-Rodriguez, A., Solinis, M. A., Rodriguez, J. M., Friedrich, U., Torrecilla, J., et al. (2016). Structural recovery of the retina in a retinoschisin-deficient mouse after gene replacement therapy by solid lipid nanoparticles. *Biomaterials* 90, 40–49. doi: 10.1016/j.biomaterials.2016.03.004
- Arnold, E. S., Ling, S. C., Huelga, S. C., Lagier-Tourenne, C., Polymenidou, M., Ditsworth, D., et al. (2013). ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proc. Natl. Acad. Sci. U.S.A.* 110, E736–E745.
- Axelrod, F. B., Liebes, L., Gold-Von Simson, G., Mendoza, S., Mull, J., Leyne, M., et al. (2011). Kinetin improves IKBKAP mRNA splicing in patients with familial dysautonomia. *Pediatr. Res.* 70, 480–483. doi: 10.1203/pdr.0b013e31822e1825
- Aznarez, I., Chan, E. M., Zielinski, J., Blencowe, B. J., and Tsui, L. C. (2003). Characterization of disease-associated mutations affecting an exonic splicing enhancer and two cryptic splice sites in exon 13 of the cystic fibrosis transmembrane conductance regulator gene. *Hum. Mol. Genet.* 12, 2031–2040. doi: 10.1093/hmg/ddg215
- Badieyan, Z. S., Pasewald, T., Mykhaylyk, O., Rudolph, C., and Plank, C. (2017). Efficient ex vivo delivery of chemically modified messenger RNA using lipofection and magnetofection. *Biochem. Biophys. Res. Commun.* 482, 796–801. doi: 10.1016/j.bbrc.2016.11.113
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531, 47–52.
- Baker, M. S., Ahn, S. B., Mohamedali, A., Islam, M. T., Cantor, D., Verhaert, P. D., et al. (2017). Accelerating the search for the missing proteins in the human proteome. *Nat. Commun.* 8:14271.
- Baker, N. L., Mörgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F., et al. (2005). Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum. Mol. Genet.* 14, 279–293. doi: 10.1093/hmg/ddi025
- Balestra, D., Giorgio, D., Bizzotto, M., Fazzari, M., Ben Zeev, B., Pinotti, M., et al. (2019a). Splicing mutations impairing CDKL5 expression and activity can be efficiently rescued by U1snRNA-based therapy. *Int. J. Mol. Sci.* 20:4130. doi: 10.3390/ijms20174130
- Balestra, D., Maestri, I., Branchini, A., Ferrarese, M., Bernardi, F., and Pinotti, M. (2019b). An altered splicing registry explains the differential ExSpeU1-mediated rescue of splicing mutations causing haemophilia A. *Front. Genet.* 10:974. doi: 10.3389/fgene.2019.00974
- Barboux, S., Niaudet, P., Gubler, M. C., Grünfeld, J. P., Jaubert, F., Kuttann, F., et al. (1997). Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nat. Genet.* 17, 467–470. doi: 10.1038/ng1297-467
- Berger, A., Lorain, S., Josephine, C., Desrosiers, M., Peccate, C., Voit, T., et al. (2015). Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa. *Mol. Ther.* 23, 918–930. doi: 10.1038/mt.2015.11
- Berger, A., Maire, S., Gaillard, M. C., Sahel, J. A., Hantraye, P., and Bemelmans, A. P. (2016). mRNA trans-splicing in gene therapy for genetic diseases. *Wiley Interdiscipl. Rev. RNA* 7, 487–498. doi: 10.1002/wrna.1347
- Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 74, 3171–3175. doi: 10.1073/pnas.74.8.3171
- Bolduc, V., Zou, Y., Ko, D., and Bonnemann, C. G. (2014). siRNA-mediated allele-specific silencing of a COL6A3 mutation in a cellular model of dominant ullrich muscular dystrophy. *Mol. Ther. Nucleic Acids* 3:e147. doi: 10.1038/mtna.2013.74
- Bourbon, M., Duarte, M. A., Alves, A. C., Medeiros, A. M., Marques, L., and Soutar, A. K. (2009). Genetic diagnosis of familial hypercholesterolaemia: the importance of functional analysis of potential splice-site mutations. *J. Med. Genet.* 46, 352–357. doi: 10.1136/jmg.2007.057000
- Bowen, D. J. (2002). Haemophilia A and haemophilia B: molecular insights. *Mol. Pathol.* 55, 127–144. doi: 10.1136/mp.55.2.127
- Bozzuto, G., and Molinari, A. (2015). Liposomes as nanomedical devices. *Int. J. Nanomed.* 10, 975–999.
- 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
- Brousseau, J. P., Lucier, J. F., Lamarche, A. A., Shkreta, L., Gendron, D., Lapointe, E., et al. (2014). Redirecting splicing with bifunctional oligonucleotides. *Nucleic Acids Res.* 42, e40. doi: 10.1093/nar/gkt1287
- Brun, C., Suter, D., Pauli, C., Dunant, P., Lochmuller, H., Burgunder, J. M., et al. (2003). U7 snRNAs induce correction of mutated dystrophin pre-mRNA by exon skipping. *Cell Mol. Life Sci.* 60, 557–566. doi: 10.1007/s000180300047
- Cacace, R., Slegers, K., and Van Broeckhoven, C. (2016). Molecular genetics of early-onset Alzheimer disease revisited. *Alzheimers Dement.* 12, 733–748. doi: 10.1016/j.jalz.2016.01.012
- Camacho Vanegas, O., Bertini, E., Zhang, R. Z., Petrini, S., Minosse, C., Sabatelli, P., et al. (2001). Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7516–7521. doi: 10.1073/pnas.121027598
- Campbell, M. A., and Wengel, J. (2011). Locked vs. unlocked nucleic acids (LNA vs. UNA): contrasting structures work towards common therapeutic goals. *Chem. Soc. Rev.* 40, 5680–5689.
- Carabet, L. A., Leblanc, E., Lallous, N., Morin, H., Ghaidi, F., Lee, J., et al. (2019). Computer-aided discovery of small molecules targeting the RNA splicing activity of hnRNP A1 in castration-resistant prostate cancer. *Molecules* 24, 763. doi: 10.3390/molecules24040763
- Cardoso, M. M., Peca, I. N., and Roque, A. C. (2012). Antibody-conjugated nanoparticles for therapeutic applications. *Curr. Med. Chem.* 19, 3103–3127. doi: 10.2174/092986712800784667
- Castro-Mayorga, J. L., Fabra, M. J., Cabedo, L., and Lagaron, J. M. (2016). On the Use of the electrospinning coating technique to produce antimicrobial polyhydroxyalkanoate materials containing in situ-stabilized silver nanoparticles. *Nanomaterials (Basel)* 7:4. doi: 10.3390/nano7010004
- Castro-Mayorga, J. L., Martinez-Abad, A., Fabra, M. J., Olivera, C., Reis, M., and Lagaron, J. M. (2014). Stabilization of antimicrobial silver nanoparticles by a polyhydroxyalkanoate obtained from mixed bacterial culture. *Int. J. Biol. Macromol.* 71, 103–110. doi: 10.1016/j.ijbiomac.2014.06.059
- 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
- Chao, H., Mansfield, S. G., Bartel, R. C., Hiriyanna, S., Mitchell, L. G., Garcia-Blanco, M. A., et al. (2003). Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing. *Nat. Med.* 9, 1015–1019. doi: 10.1038/nm900
- Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-Seq technologies and related computational data analysis. *Front. Genet.* 10:317. doi: 10.3389/fgene.2019.00317
- Chen, Z., Lai, X., Song, S., Zhu, X., and Zhu, J. (2016). Nanostructured lipid carriers based temozolomide and gene co-encapsulated nanomedicine for gliomatosis cerebri combination therapy. *Drug Deliv.* 23, 1369–1373.
- Chira, S., Jackson, C. S., Oprea, I., Ozturk, F., Pepper, M. S., Diaconu, I., et al. (2015). Progresses towards safe and efficient gene therapy vectors. *Oncotarget* 6, 30675–30703. doi: 10.18632/oncotarget.5169
- Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell. Sep* 12, 1–8. doi: 10.1016/0092-8674(77)90180-5

- Cieply, B., and Carstens, R. P. (2015). Functional roles of alternative splicing factors in human disease. *Wiley Interdiscipl. Rev. RNA* 311–326. doi: 10.1002/wrna.1276
- Cladaras, C., Hadzopoulou-Cladaras, M., Felberg, B. K., Pavlakis, G., and Zannis, V. I. (1987). The molecular basis of a familial ApoE deficiency: an acceptor splice site mutation in the third intron of the deficient apoE gene. *J. Biol. Chem.* 262, 2310–2315.
- Coady, T. H., and Lorson, C. L. (2010). Trans-splicing-mediated improvement in a severe mouse model of spinal muscular atrophy. *J. Neurosci.* 30, 126–130. doi: 10.1523/jneurosci.4489-09.2010
- Cohen, H., Levy, R. J., Gao, J., Fishbein, I., Kousaev, V., Sosnowski, S., et al. (2000). Sustained delivery and expression of DNA encapsulated in polymeric nanoparticles. *Gene Ther.* 7, 1896–1905. doi: 10.1038/sj.gt.3301318
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- Coutinho, M. F., Lacerda, L., Prata, M. J., Ribeiro, H., Lopes, L., Ferreira, C., et al. (2008). Molecular characterization of portuguese patients with mucopolysaccharidosis IIIC: two novel mutations in the HGSNAT gene. *Clin. Genet.* 74, 194–195. doi: 10.1111/j.1399-0004.2008.01040.x
- Cristea, S., Gregory, P. D., Urnov, F. D., and Cost, G. J. (2011). Dissection of splicing regulation at an endogenous locus by zinc-finger nuclease-mediated gene editing. *PLoS ONE* 6:e16961. doi: 10.1371/journal.pone.0016961
- Cuyvers, E., and Sleegers, K. (2016). Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* 15, 857–868. doi: 10.1016/s1474-4422(16)00127-7
- Daguet, E., Dujardin, G., and Valcarcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep.* 16, 1640–1655. doi: 10.15252/embr.201541116
- Daiger, S. P., Bowne, S. J., and Sullivan, L. S. (2014). Genes and mutations causing autosomal dominant retinitis pigmentosa. *Cold Spring Harb. Perspect. Med.* 5:a017129. doi: 10.1101/cshperspect.a017129
- Dal Mas, A., Rogalska, M. E., Bussani, E., and Pagani, F. (2015). Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. *Am. J. Hum. Genet.* 96, 93–103. doi: 10.1016/j.ajhg.2014.12.009
- Dana, H., Chalbatani, G. M., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., et al. (2017). Molecular mechanisms and biological functions of siRNA. *Int. J. Biomed. Sci.* 13, 48–57.
- Daraee, H., Eatemadi, A., Abbasi, E., Fekri Aval, S., Kouhi, M., and Akbarzadeh, A. (2016a). Application of gold nanoparticles in biomedical and drug delivery. *Artif. Cells Nanomed. Biotechnol.* 44, 410–422.
- Daraee, H., Eatemadi, A., Kouhi, M., Alimirzalu, S., and Akbarzadeh, A. (2016b). Application of liposomes in medicine and drug delivery. *Artif Cells Nanomed. Biotechnol.* 44, 381–391.
- Dastidar, S., Ardui, S., Singh, K., Majumdar, D., Nair, N., Fu, Y., et al. (2018). Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPSCs and myogenic cells. *Nucleic Acids Res.* 46, 8275–8298. doi: 10.1093/nar/gky548
- David, R. M., and Doherty, A. T. (2017). Viral vectors: the road to reducing genotoxicity. *Toxicol. Sci.* 155, 315–325. doi: 10.1093/toxsci/kfw220
- de Jesus, M. B., and Zuhorn, I. S. (2015). Solid lipid nanoparticles as nucleic acid delivery system: properties and molecular mechanisms. *J. Control. Release* 201, 1–13. doi: 10.1016/j.jconrel.2015.01.010
- de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., et al. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* 12, 525–532. doi: 10.1016/j.molcel.2003.08.001
- den Hollander, A. I., Koenekeop, R. K., Yzer, S., Lopez, I., Arends, M. L., Voesenek, K. E. J., et al. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of leber congenital amaurosis. *Am. J. Hum. Genet.* 79, 556–561. doi: 10.1086/507318
- Dessein, A. F., Fontaine, M., Andresen, B. S., Gregersen, N., Brivet, M., Rabier, D., et al. (2010). A novel mutation of the ACADM gene (c.145C> G) associated with the common c.985A > G mutation on the other ACADM allele causes mild MCAD deficiency: a case report. *Orphanet J. Rare Dis.* 5, 26. doi: 10.1186/1750-1172-5-26
- Dewaele, M., Tabaglio, T., Willekens, K., Bezzi, M., Teo, S. X., Low, D. H., et al. (2016). Antisense oligonucleotide-mediated MDM4 exon 6 skipping impairs tumor growth. *J. Clin. Invest.* 126, 68–84.
- Dickson, A., Osman, E., and Lorson, C. L. (2008). A negatively acting bifunctional RNA increases survival motor neuron both in vitro and in vivo. *Hum. Gene Ther.* 19, 1307–1315. doi: 10.1089/hum.2008.067
- Disset, A., Bourgeois, C. F., Benmalek, N., Claustres, M., Stevenin, J., and Tuffery-Giraud, S. (2006). An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. *Hum. Mol. Genet.* 15, 999–1013. doi: 10.1093/hmg/ddl015
- Dominski, Z., and Kole, R. (1993). Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8673–8677. doi: 10.1073/pnas.90.18.8673
- Donadon, I., McVey, J. H., Garagiola, I., Branchini, A., Mortarino, M., Peyvandi, F., et al. (2018a). Clustered F8 missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. *Haematologica* 103, 344–350. doi: 10.3324/haematol.2017.178327
- Donadon, I., Pinotti, M., Rajkowska, K., Pianigiani, G., Barbon, E., Morini, E., et al. (2018b). Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. *Hum. Mol. Genet.* 27, 2466–2476. doi: 10.1093/hmg/ddy151
- Dorrani, M., Garbuzenko, O. B., Minko, T., and Michniak-Kohn, B. (2016). Development of edge-activated liposomes for siRNA delivery to human basal epidermis for melanoma therapy. *J. Control. Release* 228, 150–158. doi: 10.1016/j.jconrel.2016.03.010
- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L. E., Munoz, M. J., Le Jossic-Coscos, C., et al. (2014). How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* 54, 683–690. doi: 10.1016/j.molcel.2014.03.044
- Eckstein, F. (2014). Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Therap.* 24, 374–387. doi: 10.1089/nat.2014.0506
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498. doi: 10.1038/35078107
- Estelrich, J., Escribano, E., Queralt, J., and Busquets, M. A. (2015). Iron oxide nanoparticles for magnetically-guided and magnetically-responsive drug delivery. *Int. J. Mol. Sci.* 16, 8070–8101. doi: 10.3390/ijms16048070
- Fadista, J., Vikman, P., Laakso, E. O., Mollet, I. G., Esguerra, J. L., Taneera, J., et al. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13924–13929. doi: 10.1073/pnas.1402665111
- Fan, F., Wang, L., Ouyang, Z., Wen, Y., and Lu, X. (2018). Development and optimization of a tumor targeting system based on microbial synthesized PHA biopolymers and PhaP mediated functional modification. *Appl. Microbiol. Biotechnol.* 102, 3229–3241. doi: 10.1007/s00253-018-8790-2
- Fang, Y., Cheng, Y., Lu, D., Gong, X., Yang, G., Gong, Z., et al. (2018). Treatment of beta(654)-thalassaemia by TALENs in a mouse model. *Cell Prolif.* 51, e12491. doi: 10.1111/cpr.12491
- Fernandez, A., Josa, S., and Montoliu, L. (2017). A history of genome editing in mammals. *Mamm. Genome* 28, 237–246. doi: 10.1007/s00335-017-9699-2
- Fernandez Alanis, E., Pinotti, M., Dal Mas, A., Balestra, D., Cavallari, N., Rogalska, M. E., et al. (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* 21, 2389–2398. doi: 10.1093/hmg/ddo045
- Ferrari, S., Di Iorio, E., Barbaro, V., Ponzin, D., Sorrentino, F. S., and Parmeggiani, F. (2011). Retinitis pigmentosa: genes and disease mechanisms. *Curr. Genomics* 12, 238–249. doi: 10.2174/138920211795860107
- Fine, J. D., and Hintner, H. (2009). *Life with Epidermolysis Bullosa (EB): Etiology, Diagnosis, Multidisciplinary Care and Therapy*. Berlin: Springer.
- Fine-Shamir, N., Beig, A., Zur, M., Lindley, D., Miller, J. M., and Dahan, A. (2017). Toward successful cyclodextrin based solubility-enabling formulations for oral delivery of lipophilic drugs: solubility-permeability trade-off, biorelevant dissolution, and the unstirred water layer. *Mol. Pharm.* 14, 2138–2146. doi: 10.1021/acs.molpharmaceut.7b00275
- Fink, K. D., Deng, P., Gutierrez, J., Anderson, J. S., Torrest, A., Komarla, A., et al. (2016). Allele-specific reduction of the mutant huntingtin allele using transcription activator-like effectors in human huntingtin's disease fibroblasts. *Cell Transplant.* 25, 677–686. doi: 10.3727/096368916x690863
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi: 10.1038/35888

- Foldvari, M., Chen, D. W., Nafissi, N., Calderon, D., Narsineni, L., and Rafiee, A. (2016). Non-viral gene therapy: gains and challenges of non-invasive administration methods. *J. Control. Release* 240, 165–190. doi: 10.1016/j.jconrel.2015.12.012
- Foltz, L. P., Howden, S. E., Thomson, J. A., and Clegg, D. O. (2018). Functional assessment of patient-derived retinal pigment epithelial cells edited by CRISPR/Cas9. *Int. J. Mol. Sci.* 19:4127. doi: 10.3390/ijms19124127
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., et al. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676. doi: 10.1101/gad.252106.114
- Furney, S. J., Pedersen, M., Gentien, D., Dumont, A. G., Rapinat, A., Desjardins, L., et al. (2013). SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* 3, 1122–1129. doi: 10.1158/2159-8290.cd-13-0330
- Garriga-Canut, M., Agustín-Pavón, C., Herrmann, F., Sanchez, A., Dierssen, M., Fillat, C., et al. (2012). Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3136–E3145.
- Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. doi: 10.1038/nature24644
- Geary, R. S., Watanabe, T. A., Truong, L., Freier, S., Lesnik, E. A., Sioufi, N. B., et al. (2001). Pharmacokinetic properties of 2'-O-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J. Pharmacol. Exp. Therap.* 296, 890–897.
- Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., et al. (2007). HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum. Mutat.* 28, 206. doi: 10.1002/humu.9479
- Giljohann, D. A., Seferos, D. S., Daniel, W. L., Massich, M. D., Patel, P. C., and Mirkin, C. A. (2010). Gold nanoparticles for biology and medicine. *Angew. Chem. Int. Ed. Engl.* 49, 3280–3294.
- Gordillo-Galeano, A., and Mora-Huertas, C. E. (2018). Solid lipid nanoparticles and nanostructured lipid carriers: a review emphasizing on particle structure and drug release. *Eur. J. Pharm. Biopharm.* 133, 285–308. doi: 10.1016/j.ejpb.2018.10.017
- Gorman, L., Mercatante, D. R., and Kole, R. (2000). Restoration of correct splicing of thalassemic beta-globin pre-mRNA by modified U1 snRNAs. *J. Biol. Chem.* 275, 35914–35919. doi: 10.1074/jbc.m006259200
- Goyenvalle, A., Vulin, A., Fougereousse, F., Leturcq, F., Kaplan, J. C., Garcia, L., et al. (2004). Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science* 306, 1796–1799. doi: 10.1126/science.1104297
- Gref, R., Luck, M., Quellec, P., Marchand, M., Dellacherie, E., Harnisch, S., et al. (2000). 'Stealth' corona-core nanoparticles surface modified by polyethylene glycol (PEG): influences of the corona (PEG chain length and surface density) and of the core composition on phagocytic uptake and plasma protein adsorption. *Colloids Surf. B Biointerfaces* 18, 301–313. doi: 10.1016/s0927-7765(99)00156-3
- Habara, Y., Takeshima, Y., Awano, H., Okizuka, Y., Zhang, Z., Saiki, K., et al. (2009). In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by (1G–A mutations in introns of the dystrophin gene. *J. Med. Genet.* 46, 542–547. doi: 10.1136/jmg.2008.061259
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415–419. doi: 10.1126/science.1088547
- Hanlon, K. S., Kleinstiver, B. P., Garcia, S. P., Zaborowski, M. P., Volak, A., Spirig, S. E., et al. (2019). High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* 10:4439.
- Harbour, J. W., Roberson, Anbunathan, H., Onken, M. D., Worley, L. A., and Bowcock, A. M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat. Genet.* 45, 133–135. doi: 10.1038/ng.2523
- Hardwick, S. A., Joglekar, A., Flicek, P., Frankish, A., and Tilgner, H. U. (2019). Getting the entire message: progress in isoform sequencing. *Front. Genet.* 10:709. doi: 10.3389/fgene.2019.00709
- Havens, M. A., Duelli, D. M., and Hastings, M. L. (2013). Targeting RNA splicing for disease therapy. *Wiley Interdiscipl. Rev. RNA* 4, 247–266. doi: 10.1002/wrna.1158
- Havens, M. A., and Hastings, M. L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res.* 44, 6549–6563. doi: 10.1093/nar/gkw533
- Hayes, G. M., Carrigan, P. E., Beck, A. M., and Miller, L. J. (2006). Targeting the RNA splicing machinery as a novel treatment strategy for pancreatic carcinoma. *Cancer Res.* 66, 3819–3827. doi: 10.1158/0008-5472.can-05-4065
- Hazer, B., and Steinbuchel, A. (2007). Increased diversification of polyhydroxyalkanoates by modification reactions for industrial and medical applications. *Appl. Microbiol. Biotechnol.* 74, 1–12. doi: 10.1007/s00253-006-0732-8
- He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R. C., et al. (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* 332, 238–240. doi: 10.1126/science.1200587
- Hermans, P. W., van Soolingen, D., Bik, E. M., de Haas, P. E., Dale, J. W., and van Embden, J. D. (1991). Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* 59, 2695–2705. doi: 10.1128/iai.59.8.2695-2705.1991
- Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell.* 68, 26–43. doi: 10.1016/j.molcel.2017.09.029
- Ho, C., Musa, F., Bell, C., and Walker, S. (2015). LDLR gene synonymous mutation c.1813C>T results in mRNA splicing variation in a kindred with familial hypercholesterolaemia. *Ann. Clin. Biochem.* 52, 680–684. doi: 10.1177/0004563215572702
- Hong, D., Kurzrock, R., Kim, Y., Woessner, R., Younes, A., Nemunaitis, J., et al. (2015). AZD9150, a next-generation antisense oligonucleotide inhibitor of STAT3 with early evidence of clinical activity in lymphoma and lung cancer. *Sci. Transl. Med.* 7:314ra185. doi: 10.1126/scitranslmed.aac5272
- Hosseini, E. S., Nikkha, M., and Hosseinkhani, S. (2019). Cholesterol-rich lipid-mediated nanoparticles boost of transfection efficiency, utilized for gene editing by CRISPR-Cas9. *Int. J. Nanomed.* 14, 4353–4366. doi: 10.2147/ijn.s199104
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Hunter, A. C. (2006). Molecular hurdles in polyfectin design and mechanistic background to polycation induced cytotoxicity. *Adv. Drug Deliv. Rev.* 58, 1523–1531. doi: 10.1016/j.addr.2006.09.008
- Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., et al. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 393, 702–705.
- Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107–1120.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433. doi: 10.1128/jb.169.12.5429-5433.1987
- Iwatani-Yoshihara, M., Ito, M., Klein, M. G., Yamamoto, T., Yonemori, K., Tanaka, T., et al. (2017). Discovery of allosteric inhibitors targeting the spliceosomal RNA helicase Brr2. *J. Med. Chem.* 60, 5759–5771. doi: 10.1021/acs.jmedchem.7b00461
- Jambhekar, S. S., and Breen, P. (2016). CDs in pharmaceutical formulations I: structure and physicochemical properties, formation of complexes, and types of complex. *Drug Discov. Today* 21, 356–362. doi: 10.1016/j.drudis.2015.11.017
- Jia, C., Chen, H., Wei, M., Chen, X., Zhang, Y., Cao, L., et al. (2017). Gold nanoparticle-based miR155 antagonist macrophage delivery restores the cardiac function in ovariectomized diabetic mouse model. *Int. J. Nanomed.* 12, 4963–4979. doi: 10.2147/ijn.s138400
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Kaemmerer, W. F. (2018). How will the field of gene therapy survive its success? *Bioeng. Transl. Med.* 3, 166–177. doi: 10.1002/btm2.10090
- Kemaladewi, D. U., Bassi, P. S., Erwood, S., Al-Basha, D., Gawlik, K. I., Lindsay, K., et al. (2019). A mutation-independent approach for muscular dystrophy via

- upregulation of a modifier gene. *Nature* 572, 125–130. doi: 10.1038/s41586-019-1430-x
- Kemaladewi, D. U., Maino, E., Hyatt, E., Hou, H., Ding, M., Place, K. M., et al. (2017). Correction of a splicing defect in a mouse model of congenital muscular dystrophy type 1A using a homology-directed-repair-independent mechanism. *Nat. Med.* 23, 984–989. doi: 10.1038/nm.4367
- Kennedy, P. J., Sousa, F., Ferreira, D., Pereira, C., Nestor, M., Oliveira, C., et al. (2018). Fab-conjugated PLGA nanoparticles effectively target cancer cells expressing human CD44v6. *Acta Biomater.* 81, 208–218. doi: 10.1016/j.actbio.2018.09.043
- Kim, J., Hu, C., Moufawad, El Achkar, C., Black, L. E., Douville, J., et al. (2019). Patient-customized oligonucleotide therapy for a rare genetic disease. *N. Engl. J. Med.* 381, 1644–1652.
- Knott, G. J., and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* 361, 866–869. doi: 10.1126/science.aat5011
- Kolhar, P., Anselmo, A. C., Gupta, V., Pant, K., Prabhakarapandian, B., Ruoslahti, E., et al. (2013). Using shape effects to target antibody-coated nanoparticles to lung and brain endothelium. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10753–10758. doi: 10.1073/pnas.1308345110
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946
- Krzyszton, R., Salem, B., Lee, D. J., Schwake, G., Wagner, E., and Radler, J. O. (2017). Microfluidic self-assembly of folate-targeted monomolecular siRNA-lipid nanoparticles. *Nanoscale* 9, 7442–7453. doi: 10.1039/c7nr01593c
- Lander, E. S. (2016). The heroes of CRISPR. *Cell* 164, 18–28. doi: 10.1016/j.cell.2015.12.041
- Lapuk, A., Marr, H., Jakkula, L., Pedro, H., Bhattacharya, S., Purdom, E., et al. (2010). Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol. Cancer Res.* 8, 961–974. doi: 10.1158/1541-7786.mcr-09-0528
- Lee, K., Conboy, M., Park, H. M., Jiang, F., Kim, H. J., Dewitt, M. A., et al. (2017). Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* 1, 889–901. doi: 10.1038/s41551-017-0137-2
- Lee, S. C., Dvinge, H., Kim, E., Cho, H., Micol, J. B., Chung, Y. R., et al. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat. Med.* 22, 672–678. doi: 10.1038/nm.4097
- Lelli, N., Garuti, R., Ghisellini, M., Tiozzo, R., Rolleri, M., Aimale, V., et al. (1995). Occurrence of multiple aberrantly spliced mRNAs of the LDL-receptor gene upon a donor splice site mutation that causes familial hypercholesterolemia (FHBenevento). *J. Lipid Res.* 36, 1315–1324.
- Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., et al. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Rep.* 4, 143–154. doi: 10.1016/j.stemcr.2014.10.013
- Limeres, M. J., Suné-Pou, M., Prieto-Sanchez, S., Moreno-Castro, C., Nusblat, A. D., Hernandez-Munain, C., et al. (2019). Development and characterization of an improved formulation of cholesteryl oleate-loaded cationic solid-lipid nanoparticles as an efficient non-viral gene delivery system. *Colloids Surf. B Biointerfaces* 184, 110533. doi: 10.1016/j.colsurfb.2019.110533
- Lin, J. C. (2017). Therapeutic applications of targeted alternative splicing to cancer treatment. *Int. J. Mol. Sci.* 19:75. doi: 10.3390/ijms19010075
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903. doi: 10.1016/j.febslet.2005.02.047
- Lorson, C. L., Hahnen, E., Androphy, E. J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6307–6311. doi: 10.1073/pnas.96.11.6307
- Lu, X. Y., Ciraolo, E., Stefanina, R., Chen, G. Q., Zhang, Y., and Hirsch, E. (2011). Sustained release of PI3K inhibitor from PHA nanoparticles and in vitro growth inhibition of cancer cell lines. *Appl. Microbiol. Biotechnol.* 89, 1423–1433. doi: 10.1007/s00253-011-3101-1
- Maiseyue, A., Mihai, G., Roy, S., Kherada, N., Simonetti, O. P., Sen, C. K., et al. (2010). Detection of macrophages via paramagnetic vesicles incorporating oxidatively tailored cholesterol ester: an approach for atherosclerosis imaging. *Nanomedicine (Lond)* 5, 1341–1356. doi: 10.22217/nnm.10.87
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- 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
- Mathew, V., and Wang, A. K. (2019). Inotersen: new promise for the treatment of hereditary transthyretin amyloidosis. *Drug Design Dev. Ther.* 13, 1515–1525. doi: 10.2147/dddt.s162913
- Matos, L., Canals, I., Dridi, L., Choi, Y., Prata, M. J., Jordan, P., et al. (2014). Therapeutic strategies based on modified U1 snRNAs and chaperones for Sanfilippo C splicing mutations. *Orphanet. J. Rare Dis.* 9:180.
- Mattioli, C., Pianigiani, G., De Rocco, D., Bianco, A. M., Cappelli, E., Savoia, A., et al. (2014). Unusual splice site mutations disrupt FANCA exon 8 definition. *Biochim. Biophys. Acta* 1842, 1052–1058. doi: 10.1016/j.bbadis.2014.03.014
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361. doi: 10.1038/385357a0
- Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343–349. doi: 10.1038/nature02873
- Mojica, F. J., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* 9, 613–621. doi: 10.1111/j.1365-2958.1993.tb01721.x
- Monani, U. R., Lorson, C. L., Parsons, D. W., Prior, T. W., Androphy, E. J., Burghes, A. H., et al. (1999). A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* 8, 1177–1183. doi: 10.1093/hmg/8.7.1177
- Montes, M., Cloutier, A., Sánchez-Hernández, N., Michelle, L., Lemieux, B., Blanchette, M., et al. (2012). TCERG1 regulates alternative splicing of Bcl-x gene by modulating the rate of RNAPII transcription. *Mol. Cell. Biol.* 32, 751–762. doi: 10.1128/mcb.06255-11
- Mukheem, A., Muthosamy, K., Manickam, S., Sudesh, K., Shahabuddin, S., Saidur, R., et al. (2018). Fabrication and characterization of an electrospun PHA/graphene silver nanocomposite scaffold for antibacterial applications. *Materials (Basel)* 11:1673. doi: 10.3390/ma11091673
- Muller, R. H., Mader, K., and Gohla, S. (2000). Solid lipid nanoparticles (SLN) for controlled drug delivery – a review of the state of the art. *Eur. J. Pharm. Biopharm.* 50, 161–177. doi: 10.1016/s0939-6411(00)00087-4
- Mustafa, S., Devi, V. K., and Pai, R. S. (2017). Effect of PEG and water-soluble chitosan coating on moxifloxacin-loaded PLGA long-circulating nanoparticles. *Drug Deliv. Transl. Res.* 7, 27–36. doi: 10.1007/s13346-016-0326-7
- Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198. doi: 10.1146/annurev-biochem-060614-034242
- Nielsen, K. B., Sorensen, S., Cartegni, L., Corydon, T. J., Doktor, T. K., Schroeder, L. D., et al. (2007). Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous snp in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am. J. Hum. Genet.* 80, 416–432. doi: 10.1086/511992
- Nielsen, L. R., Schwartz, and Scheibel, E. (1992). Screening for mutations in the gene encoding factor IX. *J. Inher. Metab. Dis.* 15, 339–341. doi: 10.1007/bf02435971
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54.
- Nissim-Rafinia, M., and Kerem, B. (2006). Splicing modulation as a modifier of the CFTR function. *Prog. Mol. Subcell Biol.* 44, 233–254. doi: 10.1007/978-3-540-34449-0_10
- Norcliffe-Kaufmann, L., and Kaufmann, H. (2012). Familial dysautonomia (Riley-Day syndrome): when baroreceptor feedback fails. *Auton. Neurosci.* 172, 26–30. doi: 10.1016/j.autneu.2012.10.012
- Osborn, M. J., Newby, G. A., McElroy, A. N., Knipping, F., Nielsen, S. C., Riddle, M. J., et al. (2020). Base editor correction of COL7A1 in recessive dystrophic epidermolysis bullosa patient-derived fibroblasts and iPSCs. *J. Invest. Dermatol.* 140:e335.

- Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Perez-Pinera, P., Brown, M. T., Majoros, W. H., et al. (2015). Correction of dystrophin expression in cells from Duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Mol. Ther.* 23, 523–532. doi: 10.1038/mt.2014.234
- Owen, N., Zhou, H., Malygin, A. A., Sangha, J., Smith, L. D., Muntoni, F., et al. (2011). Design principles for bifunctional targeted oligonucleotide enhancers of splicing. *Nucleic Acids Res.* 39, 7194–7208. doi: 10.1093/nar/gkr152
- Paliwal, R., Babu, R. J., and Palakurthi, S. (2014). Nanomedicine scale-up technologies: feasibility and challenges. *AAPS PharmSciTech.* 15, 1527–1534. doi: 10.1208/s12249-014-0177-9
- 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
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* 102, 11–26. doi: 10.1016/j.ajhg.2017.11.002
- Parkesh, R., Childs-Disney, J. L., Nakamori, M., Kumar, A., Wang, E., Wang, T., et al. (2012). Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database and chemical similarity searching. *J. Am. Chem. Soc.* 134, 4731–4742. doi: 10.1021/ja210088v
- Pinotti, M., Balestra, D., Rizzotto, L., Maestri, I., Pagani, F., and Bernardi, F. (2009). Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood* 113, 6461–6464. doi: 10.1182/blood-2009-03-207613
- Pinotti, M., Rizzotto, L., Balestra, D., Lewandowska, M. A., Cavallari, N., Marchetti, G., et al. (2008). U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood* 111, 2681–2684. doi: 10.1182/blood-2007-10-117440
- Planté-Bordeneuve, V., and Said, G. (2011). Familial amyloid polyneuropathy. *Lancet Neurol.* 10, 1086–1097.
- Porteus, M. H., and Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* 300:763. doi: 10.1126/science.1078395
- Procter, A. M., Phillips, J. A., and Cooper, D. N. (1998). The molecular genetics of growth hormone deficiency. *Hum. Genet.* 103, 255–272.
- Qiao, C., Liu, J., Yang, J., Li, Y., Weng, J., Shao, Y., et al. (2016). Enhanced non-inflammasome mediated immune responses by mannosylated zwitterionic-based cationic liposomes for HIV DNA vaccines. *Biomaterials* 85, 1–17. doi: 10.1016/j.biomaterials.2016.01.054
- Radu, I. C., Hudita, A., Zaharia, C., Galateanu, B., Iovu, H., Tanasa, E. V., et al. (2019). Poly(3-hydroxybutyrate-CO-3-hydroxyvalerate) PHBHV biocompatible nanocarriers for 5-FU delivery targeting colorectal cancer. *Drug Deliv.* 26, 318–327. doi: 10.1080/10717544.2019.1582729
- Ramamoorthi, M., and Narvekar, A. (2015). Non viral vectors in gene therapy- an overview. *J. Clin. Diagn Res.* 9, GE01–GE06.
- Rassu, G., Soddu, E., Posadino, A. M., Pintus, G., Sarmiento, B., Giunchedi, P., et al. (2017). Nose-to-brain delivery of BACE1 siRNA loaded in solid lipid nanoparticles for Alzheimer's therapy. *Colloids Surf. B Biointerfaces* 152, 296–301. doi: 10.1016/j.colsurfb.2017.01.031
- Ratni, H., Ebeling, M., Baird, J., Bendels, S., Bylund, J., Chen, K. S., et al. (2018). Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.* 61, 6501–6517.
- Rees, H. A., and Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788. doi: 10.1038/s41576-018-0059-1
- Rizk, M., and Tuzmen, S. (2019). Patisiran for the treatment of patients with familial amyloid polyneuropathy. *Drugs Today (Barc)* 55, 315–327.
- Rodriguez-Martin, T., Anthony, K., Garcia-Blanco, M. A., Mansfield, S. G., Anderton, B. H., and Gallo, J. M. (2009). Correction of tau mis-splicing caused by FTDP-17 MAPT mutations by spliceosome-mediated RNA trans-splicing. *Hum. Mol. Genet.* 18, 3266–3273. doi: 10.1093/hmg/ddp264
- Rodriguez-Martin, T., Garcia-Blanco, M. A., Mansfield, S. G., Grover, A. C., Hutton, M., Yu, Q., et al. (2005). Reprogramming of tau alternative splicing by spliceosome-mediated RNA trans-splicing: implications for tauopathies. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15659–15664. doi: 10.1073/pnas.0503150102
- Ross, S. J., Revenko, A. S., Hanson, L. L., Ellston, R., Staniszewska, A., Whalley, N., et al. (2017). Targeting KRAS-dependent tumors with AZD4785, a high-affinity therapeutic antisense oligonucleotide inhibitor of KRAS. *Sci. Transl. Med.* 9:eaa5253. doi: 10.1126/scitranslmed.aal5253
- Ryther, R. C., Flynt, A. S., Harris, B. D., Phillips, J. A. III, and Patton, J. G. (2004). GH1 splicing is regulated by multiple enhancers whose mutation produces a dominant-negative GH isoform that can be degraded by allele-specific small interfering RNA (siRNA). *Endocrinology* 145, 2988–2996. doi: 10.1210/en.2003-1724
- Sathasivam, K., Neueder, A., Gipson, T. A., Landles, C., Benjamin, A. C., Bondulich, M. K., et al. (2013). Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2366–2370. doi: 10.1073/pnas.1221891110
- Scalet, D., Balestra, D., Rohban, S., Bovolenta, M., Perrone, D., Bernardi, F., et al. (2017). Exploring splicing-switching molecules for seckel syndrome therapy. *Biochim. Biophys. Acta Mol. Basis Dis.* 1863, 15–20. doi: 10.1016/j.bbdis.2016.09.011
- Scalet, D., Sacchetto, C., Bernardi, F., Pinotti, M., van de Graaf, S. F. J., and Balestra, D. (2018). The somatic FAH C.1061C> A change counteracts the frequent FAH c.1062+5G > A mutation and permits U1snRNA-based splicing correction. *J. Hum. Genet.* 63, 683–686. doi: 10.1038/s10038-018-0427-x
- Schlieker, G., Schmidt, C., Fuchs, S., and Kissel, T. (2003). Characterization of a homologous series of D,L-lactic acid oligomers; a mechanistic study on the degradation kinetics in vitro. *Biomaterials* 24, 3835–3844. doi: 10.1016/s0142-9612(03)00243-6
- Schmid, F., Hiller, T., Korner, G., Glaus, E., Berger, W., and Neidhardt, J. (2013). A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. *Hum. Gene Ther.* 24, 97–104. doi: 10.1089/hum.2012.110
- Schornack, S., Meyer, A., Romer, P., Jordan, T., and Lahaye, T. (2006). Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J. Plant Physiol.* 163, 256–272. doi: 10.1016/j.jplph.2005.12.001
- 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
- Sharma, V. K., and Watts, J. K. (2015). Oligonucleotide therapeutics: chemistry, delivery and clinical progress. *Future Med. Chem.* 7, 2221–2242. doi: 10.4155/fmc.15.144
- Shi, J., Xu, Y., Xu, X., Zhu, X., Pridgen, E., Wu, J., et al. (2014). Hybrid lipid-polymer nanoparticles for sustained siRNA delivery and gene silencing. *Nanomedicine* 10, 897–900.
- Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258. doi: 10.1038/nrm3089
- Skordis, L. A., Dunkley, M. G., Yue, B., Eperon, I. C., and Muntoni, F. (2003). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4114–4119. doi: 10.1073/pnas.0633863100
- Smith, J. L., Mou, H., and Xue, W. (2018). Understanding and repurposing CRISPR-mediated alternative splicing. *Genome Biol.* 19:184.
- Smith, L. M., Kelleher, N. L., and Consortium for Top Down, P. (2013). Proteoform: a single term describing protein complexity. *Nat. Methods* 10, 186–187. doi: 10.1038/nmeth.2369
- Somia, N., and Verma, I. M. (2000). Gene therapy: trials and tribulations. *Nat. Rev. Genet.* 1, 91–99. doi: 10.1038/35038533
- Song, C. Q., Jiang, T., Richter, M., Rhym, L. H., Koblan, L. W., Zafra, M. P., et al. (2020). Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.* 4, 125–130. doi: 10.1038/s41551-019-0357-8
- Song, Y., Lou, H. H., Boyer, J. L., Limberis, M. P., Vandenbergh, L. H., Hackett, N. R., et al. (2009). Functional cystic fibrosis transmembrane conductance regulator expression in cystic fibrosis airway epithelial cells by AAV6.2-mediated segmental trans-splicing. *Hum. Gene Ther.* 20, 267–281. doi: 10.1089/hum.2008.173
- Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 319, 1668–1672.
- Stadtmauer, E. A., Fraietta, J. A., Davis, M. M., Cohen, A. D., Weber, K. L., Lancaster, E., et al. (2020). CRISPR-engineered T cells in patients with refractory cancer. *Science* 367:eaba7365.
- Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shaw, K., and Cooper, D. N. (2012). The human gene mutation database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr. Protoc. Bioinform.* Chapter 1:Unit113.

- Stoilov, P., Lin, C. H., Damoiseaux, R., Nikolic, J., and Black, D. L. (2008). A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11218–11223. doi: 10.1073/pnas.0801661105
- Strong, T. V., Wilkinson, D. J., Mansoura, M. K., Devor, D. C., Henze, K., Yang, Y., et al. (1993). Expression of an abundant alternatively spliced form of the cystic fibrosis transmembrane conductance regulator (CFTR) gene is not associated with a cAMP-activated chloride conductance. *Hum. Mol. Genet.* 2, 225–230. doi: 10.1093/hmg/2.3.225
- Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim. Biophys. Acta* 1489, 141–158. doi: 10.1016/S0167-4781(99)00150-5
- Sun, L., Wang, D., Chen, Y., Wang, L., Huang, P., Li, Y., et al. (2017). Core-shell hierarchical mesostructured silica nanoparticles for gene/chemo-synergistic stepwise therapy of multidrug-resistant cancer. *Biomaterials* 133, 219–228. doi: 10.1016/j.biomaterials.2017.04.028
- Sun, N., and Zhao, H. (2013). Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol. Bioeng.* 110, 1811–1821. doi: 10.1002/bit.24890
- Sune-Pou, M., Prieto-Sanchez, S., Boyero-Corral, S., Moreno-Castro, C., El Yousfi, Y., Sune-Negre, J. M., et al. (2017). Targeting splicing in the treatment of human disease. *Genes (Basel)* 8:87. doi: 10.3390/genes8030087
- Tabrizi, S. J., Leavitt, B. R., Landwehrmeyer, G. B., Wild, E. J., Saft, C., Barker, R. A., et al. (2019). Targeting huntingtin expression in patients with huntington's disease. *N. Engl. J. Med.* 380, 2307–2316.
- Tajnik, M., Rogalska, M. E., Bussani, E., Barbon, E., Balestra, D., Pinotti, M., et al. (2016). Molecular basis and therapeutic strategies to rescue factor IX variants that affect splicing and protein function. *PLoS Genet.* 12:e1006082. doi: 10.1371/journal.pgen.1006082
- Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* 8:14519.
- Taniguchi-Ikeda, M., Kobayashi, K., Kanagawa, M., Yu, C., Mori, K., Oda, T., et al. (2011). Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature* 478, 127–131. doi: 10.1038/nature10456
- Tanner, G., Glaes, E., Barthelmes, D., Ader, M., Fleischhauer, J., Pagani, F., et al. (2009). Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Hum. Mutat.* 30, 255–263. doi: 10.1002/humu.20861
- Taratula, O., Kuzmov, A., Shah, M., Garbuzenko, O. B., and Minko, T. (2013). Nanostructured lipid carriers as multifunctional nanomedicine platform for pulmonary co-delivery of anticancer drugs and siRNA. *J. Control. Release* 171, 349–357. doi: 10.1016/j.jconrel.2013.04.018
- Tardy, B. L., Tan, S., Dam, H. H., Ejima, H., Blencowe, A., Qiao, G. G., et al. (2016). Nanoparticles assembled via pH-responsive reversible segregation of CDs in polyrotaxanes. *Nanoscale* 8, 15589–15596. doi: 10.1039/c6nr04841b
- Taylor, J., and Lee, S. C. (2019). Mutations in spliceosome genes and therapeutic opportunities in myeloid malignancies. *Genes Chromosomes Cancer* 58, 889–902. doi: 10.1002/gcc.22784
- Thornton, C. A. (2014). Myotonic dystrophy. *Neurol. Clin.* 32, 705–719.
- Tumer, Z. (2013). An overview and update of ATP7A mutations leading to menkes disease and occipital horn syndrome. *Hum. Mutat.* 34, 417–429. doi: 10.1002/humu.22266
- Ullrich, N. J., and Gordon, L. B. (2015). Hutchinson-gilford progeria syndrome. *Handb. Clin. Neurol.* 132, 249–264.
- Uner, M., and Yener, G. (2007). Importance of solid lipid nanoparticles (SLN) in various administration routes and future perspectives. *Int. J. Nanomed.* 2, 289–300.
- van der Woerd, W. L., Mulder, J., Pagani, F., Beuers, U., Houwen, R. H., and van de Graaf, S. F. (2015). Analysis of aberrant pre-messenger RNA splicing resulting from mutations in ATP8B1 and efficient in vitro rescue by adapted U1 small nuclear RNA. *Hepatology* 61, 1382–1391. doi: 10.1002/hep.27620
- Verma, A. (2018). Recent advances in antisense oligonucleotide therapy in genetic neuromuscular diseases. *Ann. Indian Acad. Neurol.* 21, 3–8.
- Villemaire, J., Dion, I., Elela, S. A., and Chabot, B. (2003). Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J. Biol. Chem.* 278, 50031–50039. doi: 10.1074/jbc.m308897200
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi: 10.1016/j.cell.2009.02.009
- Wally, V., Muraier, E. M., and Bauer, J. W. (2012). Spliceosome-mediated trans-splicing: the therapeutic cut and paste. *J. Invest. Dermatol.* 132, 1959–1966. doi: 10.1038/jid.2012.101
- Wan, R., Bai, R., and Shi, Y. (2019). Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr. Opin. Struct. Biol.* 59, 124–133. doi: 10.1016/j.sbi.2019.07.010
- 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, X., Yang, L., Zhang, H., Tian, B., Li, R., Hou, X., et al. (2018). Fluorescent magnetic PEI-PLGA nanoparticles loaded with paclitaxel for concurrent cell imaging, enhanced apoptosis and autophagy in human brain cancer. *Colloids Surf. B Biointerf.* 172, 708–717. doi: 10.1016/j.colsurfb.2018.09.033
- Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A., and Berglund, J. A. (2009). Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18551–18556. doi: 10.1073/pnas.0903234106
- Wild, E. J., and Tabrizi, S. J. (2017). Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* 16, 837–847. doi: 10.1016/S1474-4422(17)30280-6
- Willyard, C. (2018). New human gene tally reignites debate. *Nature* 558, 354–355. doi: 10.1038/d41586-018-05462-w
- Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S. L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* 130, 1134–1145. doi: 10.1016/j.cell.2007.08.026
- Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061–1070. doi: 10.1016/0092-8674(93)90316-i
- Xie, Q., Jia, L. N., Xu, H. Y., Hu, X. G., Wang, W., and Jia, J. (2016). Fabrication of core-shell PEI/pBMP2-PLGA electrospon scaffold for gene delivery to periodontal ligament stem cells. *Stem Cells Int.* 2016:5385137.
- Xu, Z. P., Zeng, Q., Lu, G.-Q. M., and Yu, A. (2006). Inorganic nanoparticles as carriers for efficient cellular delivery. *Chem. Eng. Sci.* 61, 1027–1040. doi: 10.1016/j.ces.2005.06.019
- Yanaizu, M., Sakai, K., Tosaki, Y., Kino, Y., and Satoh, J. (2018). Small nuclear RNA-mediated modulation of splicing reveals a therapeutic strategy for a TREM2 mutation and its post-transcriptional regulation. *Sci. Rep.* 8:6937.
- Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., et al. (2016). Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* 164, 805–817.
- Yin, H., Kanasty, R. L., Eltoukhy, A. A., Vegas, A. J., Dorkin, J. R., and Anderson, D. G. (2014). Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* 15, 541–555.
- 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.
- Yuan, J., Ma, Y., Huang, T., Chen, Y., Peng, Y., Li, B., et al. (2018). Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* 72:e387.
- Zakharova, L., Pashirova, T., Kashapov, R., Gabdrakhmanov, D., and Sinyashin, O. (2017). *Drug Delivery Mediated by Confined Nanosystems: Structure-Activity Relations and Factors Responsible for the Efficacy of Formulations*, eds E.A.A.M. Grumezescu. Amsterdam: Elsevier.
- Zeidler, B., Froelich, S., Marlen, K., Shivak, D. A., Yu, Q., Li, D., et al. (2019). Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. *Nat. Med.* 25, 1131–1142.
- Zhang, J., Shishatskaya, E. I., Volova, T. G., da Silva, L. F., and Chen, G. Q. (2018). Polyhydroxyalkanoates (PHA) for therapeutic applications. *Mater. Sci. Eng. C Mater. Biol. Appl.* 86, 144–150. doi: 10.1016/j.msec.2017.12.035
- Zhang, K., Tang, X., Zhang, J., Lu, W., Lin, X., Zhang, Y., et al. (2014). PEG-PLGA copolymers: their structure and structure-influenced drug delivery applications. *J. Control. Release* 183, 77–86. doi: 10.1016/j.jconrel.2014.03.026

- Zhou, L., Chen, Z., Chi, W., Yang, X., Wang, W., and Zhang, B. (2012). Mono-methoxy-poly(3-hydroxybutyrate-co-4-hydroxybutyrate)-graft-hyper-branched polyethylenimine copolymers for siRNA delivery. *Biomaterials* 33, 2334–2344. doi: 10.1016/j.biomaterials.2011.11.060
- Zhu, H., Tucker, H. M., Gear, K. E., Simpson, J. F., Manning, A. K., Cupples, L. A., et al. (2007). A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. *Hum. Mol. Genet.* 16, 1765–1772. doi: 10.1093/hmg/ddm124
- Zuckerman, J. E., Gale, A., Wu, P., Ma, R., and Davis, M. E. (2015). siRNA delivery to the glomerular mesangium using polycationic cyclodextrin nanoparticles containing siRNA. *Nucleic Acid Therap.* 25, 53–64. doi: 10.1089/nat.2014.0505

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

Copyright © 2020 Suñé-Pou, Limeres, Moreno-Castro, Hernández-Munain, Suñé-Negre, Cuestas and Suñé. 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.



Contribution of mRNA Splicing to Mismatch Repair Gene Sequence Variant Interpretation

Bryony A. Thompson^{1,2*†}, Rhiannon Walters³, Michael T. Parsons^{3†}, Troy Dumenil³, Mark Drost⁴, Yvonne Tiersma⁴, Noralane M. Lindor⁵, Sean V. Tavtigian^{6†}, Niels de Wind⁴, Amanda B. Spurdle^{3†} and the InSiGHT Variant Interpretation Committee

OPEN ACCESS

Edited by:

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

Reviewed by:

Minttu Kansikas,
University of Helsinki, Finland
Christopher Heinen,
University of Connecticut,
United States

*Correspondence:

Bryony A. Thompson
bryony.thompson@mh.org.au

†ORCID:

Bryony A. Thompson
orcid.org/0000-0001-8655-1839
Michael T. Parsons
orcid.org/0000-0003-3242-8477
Sean V. Tavtigian
orcid.org/0000-0002-7543-8221
Amanda B. Spurdle
orcid.org/0000-0003-1337-7897

Specialty section:

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

Received: 20 February 2020

Accepted: 03 July 2020

Published: 27 July 2020

Citation:

Thompson BA, Walters R,
Parsons MT, Dumenil T, Drost M,
Tiersma Y, Lindor NM, Tavtigian SV,
de Wind N, Spurdle AB and the
InSiGHT Variant Interpretation
Committee (2020) Contribution
of mRNA Splicing to Mismatch Repair
Gene Sequence Variant Interpretation.
Front. Genet. 11:798.
doi: 10.3389/fgene.2020.00798

¹ Department of Pathology, The Royal Melbourne Hospital, Melbourne, VIC, Australia, ² Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia, ³ Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, ⁴ Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands, ⁵ Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, United States, ⁶ Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT, United States

Functional assays that assess mRNA splicing can be used in interpretation of the clinical significance of sequence variants, including the Lynch syndrome-associated mismatch repair (MMR) genes. The purpose of this study was to investigate the contribution of splicing assay data to the classification of MMR gene sequence variants. We assayed mRNA splicing for 24 sequence variants in *MLH1*, *MSH2*, and *MSH6*, including 12 missense variants that were also assessed using a cell-free *in vitro* MMR activity (CIMRA) assay. Multifactorial likelihood analysis was conducted for each variant, combining CIMRA outputs and clinical data where available. We collated these results with existing public data to provide a dataset of splicing assay results for a total of 671 MMR gene sequence variants (328 missense/in-frame indel), and published and unpublished repair activity measurements for 154 of these variants. There were 241 variants for which a splicing aberration was detected: 92 complete impact, 33 incomplete impact, and 116 where it was not possible to determine complete versus incomplete splicing impact. Splicing results mostly aided in the interpretation of intronic (72%) and silent (92%) variants and were the least useful for missense substitutions/in-frame indels (10%). MMR protein functional activity assays were more useful in the analysis of these exonic variants but by design they were not able to detect clinically important splicing aberrations identified by parallel mRNA assays. The development of high throughput assays that can quantitatively assess impact on mRNA transcript expression and protein function in parallel will streamline classification of MMR gene sequence variants.

Keywords: mismatch repair genes, splicing aberrations, variant interpretation and classification, variant type, Lynch syndrome, mRNA splicing

INTRODUCTION

Loss of function sequence variants in the mismatch repair (MMR) genes causes the cancer susceptibility syndrome, Lynch syndrome. However, for many sequence variants identified, the clinical significance can only be established after considering further evidence, such as population allele frequencies, tumor pathology, family co-segregation information, *in silico* predictions, and

experimental assays of MMR function (Thompson et al., 2013a,b, 2014). Some variants are “spliceogenic” and confer pathogenicity by an effect on mRNA splicing, either through the disruption of the native splice sites (5'-donor GT and 3'-acceptor AG), gain of *de novo* sites, activation of cryptic splice sites, or altering splicing regulatory elements (e.g., exonic splicing enhancers and silencers, ESEs and ESSs, respectively) (Cartegni et al., 2002). *In vitro* splicing assays using patient RNA or minigenes are thus often used to test if sequence variants cause splicing defects (Thompson et al., 2015). Output of mRNA splicing assays is incorporated into the MMR gene sequence variant classification scheme developed by the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) Variant Interpretation Committee (Thompson et al., 2014), and the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines with minimal specifications (Richards et al., 2015). An important consideration in the InSiGHT classification criteria is that allele-specific assays are required to determine the contribution of the variant allele to the overall transcript profile.

Using mRNA splicing assay results from 24 MMR gene variants, and additional splicing data submitted to the InSiGHT and Universal Mutation Databases (UMD), we investigated the utility of splicing assays in the final interpretation of MMR gene variants, considering variant location and predicted effect. We additionally considered the utility of protein functional assay data, where such information was available, for the classification of predicted missense variants.

METHODS

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. The following GenBank reference sequences were used: *MLH1* – NM_00249.3, *MSH2* – NM_00251.2, *MSH6* – NM_00179.2, and *PMS2* – NM_00535.6.

Sources of MMR Gene Variants

Cases with MMR gene germline variants (24 unique variants, **Supplementary Table S1**) in this study were identified from the Colon Cancer Family Registry (CCFR) and the Australian National Endometrial Cancer Study (ANECS) from participants with lymphoblastoid cell lines (LCLs) available for RNA analyses. Both resources have been described previously (Buchanan et al., 2014; Jenkins et al., 2018). Informed consent was obtained from all study participants. All variants interrogated in this study have been submitted to the InSiGHT MMR gene locus-specific databases¹. Additional clinical data were collected from international sites (through the InSiGHT Variant Interpretation Committee) to aid in variant classification.

mRNA Analysis

Culturing of CCFR/ANECS case-derived ($n = 24$) and healthy Red Cross donor control-derived ($n = 12$) LCLs in the

presence/absence of the nonsense-mediated decay inhibitor puromycin, and RNA extraction and cDNA synthesis were performed as previously described (Whiley et al., 2014). PCR amplification of cDNA from both cases and healthy controls was performed using Mango Taq (Bioline, Eveleigh, NSW, Australia) under the following conditions: 95°C for 2 min followed by 40 cycles of 94°C for 20 s, 60°C for 30 s and 72°C for 1 min and a final extension step at 72°C for 5 min (primer details in **Supplementary Table S2**). PCR products were separated by agarose gel electrophoresis. Three controls were run alongside each case. Cases and controls showing only single transcripts on gel visualization were sequenced at the Australian Genome Research Facility (Brisbane, QLD, Australia). For products that contained multiple transcripts, the individual bands were excised from the gel and purified using the NucleoSpin Gel and PCR clean up kit (Macherey-Nagel, Düren, Germany) per manufacturer's instructions. These purified transcripts were then re-amplified before Sanger sequencing. Sequencing chromatograms were visualized using FinchTV (Geospiza, Seattle, WA, United States). The 24 MMR gene variants were also analyzed using multiple *in silico* splicing tools (outlined in **Supplementary Table S1**).

CIMRA Assays

A subset of predicted missense substitutions were analyzed for this study using the cell-free *in vitro* mismatch repair activity (CIMRA) assay using techniques previously described for *MLH1*, *MSH2* (Drost et al., 2018), and *MSH6* (Drost et al., 2020).

Dataset Used to Assess Utility of Splicing Assay Data for Classification

All records as of July 2019 that have reported splicing analysis using RNA or minigene assays were extracted from the InSiGHT variant classification database (see text footnote 1), UMD-*MLH1/MSH2/MSH6* databases ($n = 162$) (Grandval et al., 2013), and various recent publications from which results have since been submitted to the InSiGHT database (**Supplementary Table S3**). If available, the missense/in-frame indel variants in this set were further annotated with previously generated CIMRA assay data. The five class InSiGHT MMR gene classification scheme was applied if new data were available for previously classified variants and to interpret new variants (Thompson et al., 2014). This incorporated both quantitative (multifactorial likelihood) and qualitative approaches. Multifactorial likelihood analysis was conducted as described previously (Thompson et al., 2013a,b), including the application of recently updated tumor characteristics likelihood ratios (LRs) (Li et al., 2020), and functional LRs. The functional LRs were based on the MMR activity outputs of the *MLH1*, *MSH2* (Drost et al., 2018), and *MSH6* (Drost et al., 2020) missense variants from CIMRA assays, represented as percent of wild-type activity. For the purposes of comparing splicing assay and MMR activity assays for missense/in-frame indels, the CIMRA assay data were categorized into deficient, moderate, or proficient function. The thresholds set for deficient and proficient function were equivalent to the probability of pathogenicity cut-offs used for

¹<https://www.insight-database.org>

Class 4, likely pathogenic (0.95) and Class 2, likely benign (0.05) derived using the CIMRA assay functional LR_s (Thompson et al., 2014; Drost et al., 2018, 2020). The deficient wild-type activity thresholds were set at <23% for MLH1 and MSH2, and <18% for MSH6 and PMS2 (*in lieu* of a calibrated PMS2 functional LR). The proficient wild-type activity thresholds were set at ≥70% for MLH1 and MSH2, and ≥100% for MSH6 and PMS2 (as PMS2 penetrance is closer to MSH6 than MLH1/MSH2 (Dominguez-Valentin et al., 2020)). If no validated CIMRA data was available for a variant, then the highest published MMR activity assay data value (most conservative) from alternative published assay data extracted from the InSiGHT variant classification database was used as qualitative data points to assign an effect on function. To compare splicing predictions to mRNA results, all sequence variants were annotated with MES-SWA and categorized into groups based on predicted potential to alter splicing, according to guidelines in v2.5 of the ENIGMA consortium BRCA1/2 variant interpretation criteria² and shown to have 98.7% sensitivity and 96.5% specificity to detect the correct impact on splicing (Shamsani et al., 2018). The groups were as follows, where diff is the difference between the reference and alternate scores and alt refers to the alternate score: *native loss minimal* is diff < 0, or alt > 8.5, or diff < 1.15 and 6.2 ≤ alt ≤ 8.5; *native loss moderate* is diff ≥ 1.15 and 6.2 ≤ alt ≤ 8.5, or diff < 1.15 and alt < 6.2; *native loss high* is diff ≥ 1.15 and alt < 6.2; *gain minimal* is diff > 0, or alt < 6.2, or diff < 0 and 6.2 ≤ alt ≤ 8.5 alt < closest upstream/downstream native splice site; *gain moderate* is diff < 0 and 6.2 ≤ alt ≤ 8.5 alt > closest upstream/downstream native splice site; *gain high* is diff < 0 and alt > 8.5.

Terminology to Describe Impact of Variants on mRNA Splicing

Variants were placed into one of three categories, determined through Sanger sequencing of cDNA if exonic variant present (method used variants tested in this study) or from other allele-specific techniques:

- Complete impact – variant allele results in expression of only alternatively spliced transcript(s), i.e., no or minimal reference (full-length) transcript is derived from the variant allele,
- Incomplete impact – variant allele results in expression of both reference (full-length) and alternatively spliced transcript(s)
- Extent of impact unknown – a splicing aberration was detected but it was not possible to determine if variant impact was complete or incomplete

RESULTS AND DISCUSSION

mRNA assays were conducted in this study for 24 MMR gene sequence variants. Results are summarized in **Table 1** and detailed in **Supplementary Table S1** (sequence traces are shown in the **Supplementary Figure S1**). Results from the

CIMRA assay for the 12 presumed missense substitutions are shown in **Figure 1**.

We then assessed the contribution of splicing assay results to final variant classification for 671 MMR gene sequence variants, including the 24 variants assayed for mRNA aberrations from this study (see **Supplementary Table S3**: *MLH1*: *n* = 324, 48%; *MSH2*: *n* = 225, 34%; *MSH6*: *n* = 73, 11%; *PMS2*: *n* = 49, 7%). *MLH1* and *MSH2* had the highest proportion of variants assessed, which may be due to their higher penetrance and the increased likelihood of detection using historic Lynch syndrome gene testing guidelines in the clinical setting (Dominguez-Valentin et al., 2020).

There were 156 variants that had not yet been classified by InSiGHT, and 43 variants where new splicing or CIMRA assay data could lead to reclassification from the existing InSiGHT classification. These variants were classified by applying the InSiGHT criteria and have been submitted to the InSiGHT Variant Interpretation Committee for formal classification. Overall, 92 variants caused a splicing aberration designated as complete, 33 variants had incomplete impact (i.e., the full-length transcript was also present), and for 116 variants, it was not possible to determine if impact was complete or not (see **Supplementary Table S3**).

Of the variants in the acceptor (last 20 bases of intron) or donor (first 6 bases of intron) splice site region, or the first/last 3 bases of the exon (see splice category in **Supplementary Table S3**), 168/172 with high predicted native splice site loss showed some sort of splicing aberration (98%, three of these were designated incomplete and one variant was reported as complete and incomplete in two separate studies). Another 12/15 with moderate predicted native splice site loss showed an aberration (80%, impact for one variant was designated complete and incomplete splicing in two separate studies). Splicing impact was seen for 11/52 variants with minimal predicted native loss (21%, three reported as incomplete); 4/11 were exonic variants that led to complete exon skipping events, which may be due to an effect on ESE or ESS that are not predicted by the MES-SWA tool, or otherwise false negative native loss predictions.

For the *de novo* donor/acceptor gain predictions, 13/26 variants with high predicted gain showed effect on mRNA splicing aberration (50%); of these, three had incomplete impact: one was a predicted stop gain variant, and two were confirmed to also have an effect on function due to the predicted missense change. Splicing impact was observed for 3/7 (43%) of variants with moderate predicted gain, one of which demonstrated complete activation of a cryptic splice site (*MSH2* c.2635-1G > T). Of the remaining two variants, one had high predicted native loss (*MLH1* c.1039-2A > T) and the third had no predicted effect on the native splice site (*MSH2* c.1979A > G).

Splicing alterations were reported for 225/638 (35%) of variants with no/minimal predicted gain, with splicing impact due to alternative mechanisms. The vast majority of these (176/225) were located in the splice region (defined as above—last 20 bases of the intron, first 6 bases of the intron, or the first/last 3 bases of the exon) with moderate-high prediction of native site loss, and the remainder were largely exonic variants with incomplete exon skipping events (26/49)—again implying effect on ESE/ESS.

²<https://enigmaconsortium.org/>

TABLE 1 | Summary of splicing assay results from this study and their contribution to variant classification.

Variant	Splicing/transcript expression	Variant splicing impact	InSiGHT class ^a	Splicing contributes to classification?
Missense (<i>n</i> = 12)				
<i>MLH1</i>				
c.299G > C p.(Arg100Pro)	No effect		5	No
c.793C > T p.(Arg265Cys)	r.791_884del	Complete	5	Yes
c.923A > C p.(His308Pro)	r.[791_1038del, 923a > c]	Incomplete	5 ^b	No
c.1136A > G p.(Tyr379Cys)	No effect		1 ^b	No
c.1166G > A p.(Arg389Gln)	r.[1039_1409del, 1166g > a]	Incomplete	3	No
c.1652A > G p.(Asn551Ser)	No effect		1 ^b	No
<i>MSH2</i>				
c.944G > T p.(Gly315Val)	No effect		1 ^b	No
c.1661G > A p.(Ser554Asn)	r.1511_1661del	Complete	5 ^{b,c}	Yes
c.2075G > T p.(Gly692Val)	No effect		5	No
c.2714C > G p.(Thr905Arg)	No effect		1	No
<i>MSH6</i>				
c.2314C > T p.(Arg772Trp)	r.[628_3172del, 2314c > u]	Incomplete	5	No
c.3469G > A p.(Gly1157Ser)	No effect		4	No
Silent (<i>n</i> = 5)				
<i>MLH1</i>				
c.438A > G	No effect		2 ^b	Yes
<i>MSH2</i>				
c.1275A > G	r.[1229_1276del, 1275a > g]	Incomplete	3	No
c.1344C > T	No effect		2 ^b	Yes
c.2154A > G	No effect		2	Yes
<i>MSH6</i>				
c.3246G > T	No effect		1	No
Splice site (<i>n</i> = 3)				
<i>MLH1</i>				
c.117-2A > G	r.117_121del	Unknown	5 ^b	No
c.589-2A > C	r.589_677del	Unknown	4 ^b	No
c.790 + 2T > A	r.678_790del	Unknown	5	No
Intronic (<i>n</i> = 4)				
<i>MLH1</i>				
c.454-13A > G	r.454_545del	Unknown	4	Yes
<i>MSH2</i>				
c.1276 + 11A > G	No effect		2 ^b	Yes
c.1511-9A > G	No effect		2 ^b	Yes
c.1661 + 5G > C	r.1511_1661del	Unknown	4	Yes

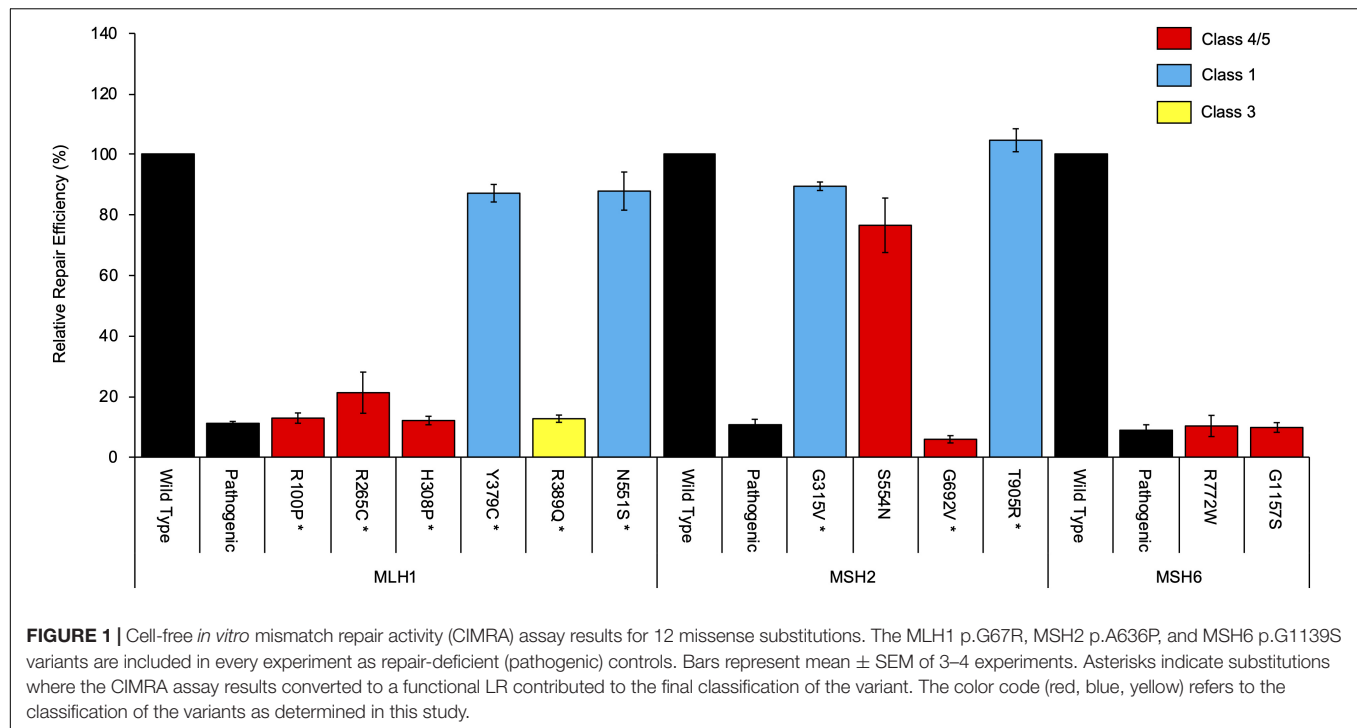
^aUpdated InSiGHT classification. The current InSiGHT database classifications are in **Supplementary Table S3**. ^bNew submissions to the InSiGHT database. ^cVariant would be classified as Class 2, likely benign based on the *in silico* prior probability with the CIMRA-based functional likelihood ratio. See **Supplementary Table S1** for more detail.

Overall, these findings highlight the complexities of using splice site prediction algorithms to prioritize variants for potential splice assays. Prediction relating to both native site loss and *de novo* gain need to be considered in parallel to assess if a variant is potentially spliceogenic, and to consider variant location in/near a splice site. Nevertheless, it is clear that triage of variants based on location in the splice region provides the most efficient method to detect spliceogenic variants. Our findings also emphasize a known deficiency in variant annotation with respect to potential effect on ESE/ESS, due to the poor specificity of currently available prediction tools (Houdayer et al., 2008). This observation stresses the importance of considering all

available points of evidence (clinical and functional) to inform variant interpretation.

All variants were assigned to categories based on variant type. The results are summarized in **Figure 2** (and described in more detail in **Supplementary Table S3**). Bearing in mind that *in vitro* experiments were likely prioritized by splicing predictions for individual variants, the results show that splicing assay results informed classification most for silent variants (92%; 69/75) and intronic variants (72%; 93/129), and least for missense substitutions/in-frame indels (10%; 34/328).

All native splice site dinucleotide variants assessed (*n* = 86) caused splicing aberrations. However, levels of the splicing aberration from the variant allele were reported for only 16



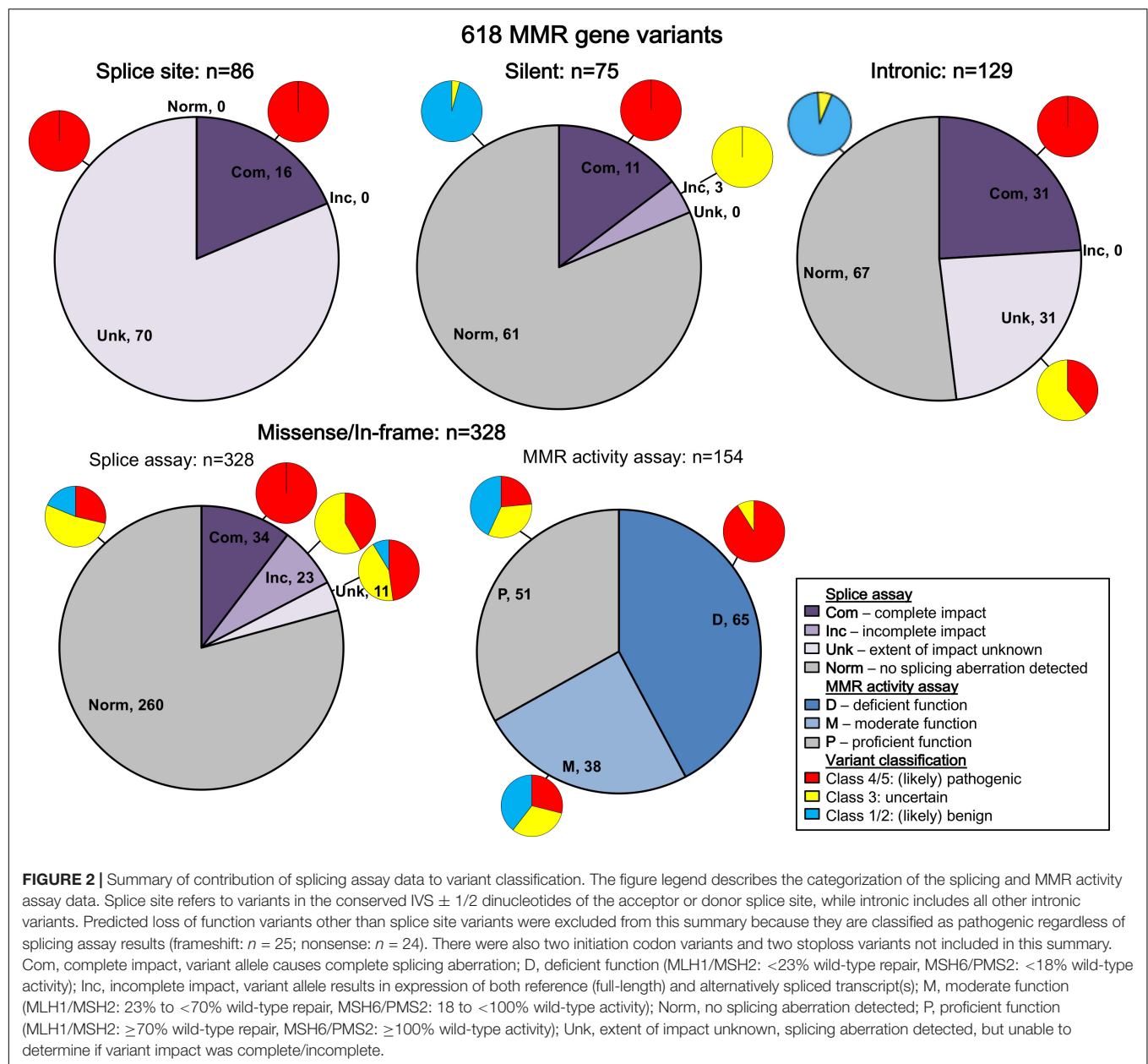
variants, information which alone permits upgrade from likely pathogenic class to pathogenic class, in accordance with InSiGHT classification criteria. Due to their very high likelihood to alter splicing, variants altering the canonical intronic dinucleotides at the native splice sites were traditionally considered pathogenic without the need to conduct splicing assays (Thompson et al., 2014; Abou Tayoun et al., 2018), but this mindset is no longer held given that consideration of naturally occurring splicing, and the predicted mRNA product is now recognized as an important aspect of variant curation (de la Hoya et al., 2016; Abou Tayoun et al., 2018). There are currently no exceptions (due to consideration of naturally occurring “rescue” isoforms) that have been identified in the MMR genes.

Splicing information was most likely to contribute evidence against pathogenicity for synonymous/silent and intronic variants, with 61/75 (81%) and 67/129 (52%) demonstrating the absence of a splicing aberration, respectively. This includes five intronic and three silent variants that demonstrated no impact on splicing, but are classified as VUS because NMD inhibitors were not used in the splicing analysis, which is a requirement for the InSiGHT splicing interpretation criteria. For these variant types, effects on splicing (or perhaps overall transcript expression) are the most likely causes of loss of function (Parmley and Hurst, 2007; Parmley and Huynen, 2009).

We did not find splicing data as useful in the interpretation of predicted missense substitutions; 68/328 (21%) of predicted missense/in-frame alterations altered mRNA splicing. Of these 68 proven spliceogenic variants, the mRNA splicing data contributed to the classification of only 34 variants (50%; due to detection of complete splicing that was considered as evidence toward pathogenicity). Further, this observation likely

overestimates the proportion of predicted missense variants that (also) alter mRNA splicing; bias toward spliceogenic variants having undergone mRNA assays is anticipated given that bioinformatic prediction of potential effect on splicing is commonly used to prioritize selection of variants for splicing assays in the research and clinical setting. Indeed, 37/68 (54%) of spliceogenic missense variants had high-moderate predicted potential to affect splicing using splicing prediction performed here, which focused on impact on native splice sites, or creation of *de novo* or activation of cryptic splice sites (but excluded prediction of effect on exonic splicing regulators, i.e., ESEs and ESSs). As might be expected, MMR activity assays were more useful to support classification of missense substitutions/in-frame indels as pathogenic, with 59/65 (91%) of variants with deficient MMR activity being classified as Class 4/5 (likely) pathogenic (Figure 2 and Supplementary Table S3). Thus, MMR activity functional assays are more useful in the interpretation of missense/in-frame indels, particularly now the output of CIMRA can be used in quantitative multifactorial analysis (Drost et al., 2018, 2020).

The current MMR activity assays do not detect impact on all biological effects; indeed, there were four (likely) pathogenic *MLH1* variants with proficient MMR activity and normal splicing (p.Lys618del, p.Pro640Ser, p.Ala681Thr, and p.Arg687Trp). For these variants, the probable cause of pathogenicity is a defect not measured by either the CIMRA assay or the splice assays reported here, such as that related to cellular localization, protein instability, or DNA damage-response. Further, current MMR activity assays are cDNA-based and cannot detect aberrant splicing; there were seven pathogenic missense variants with proficient MMR activity, where the nucleotide substitution

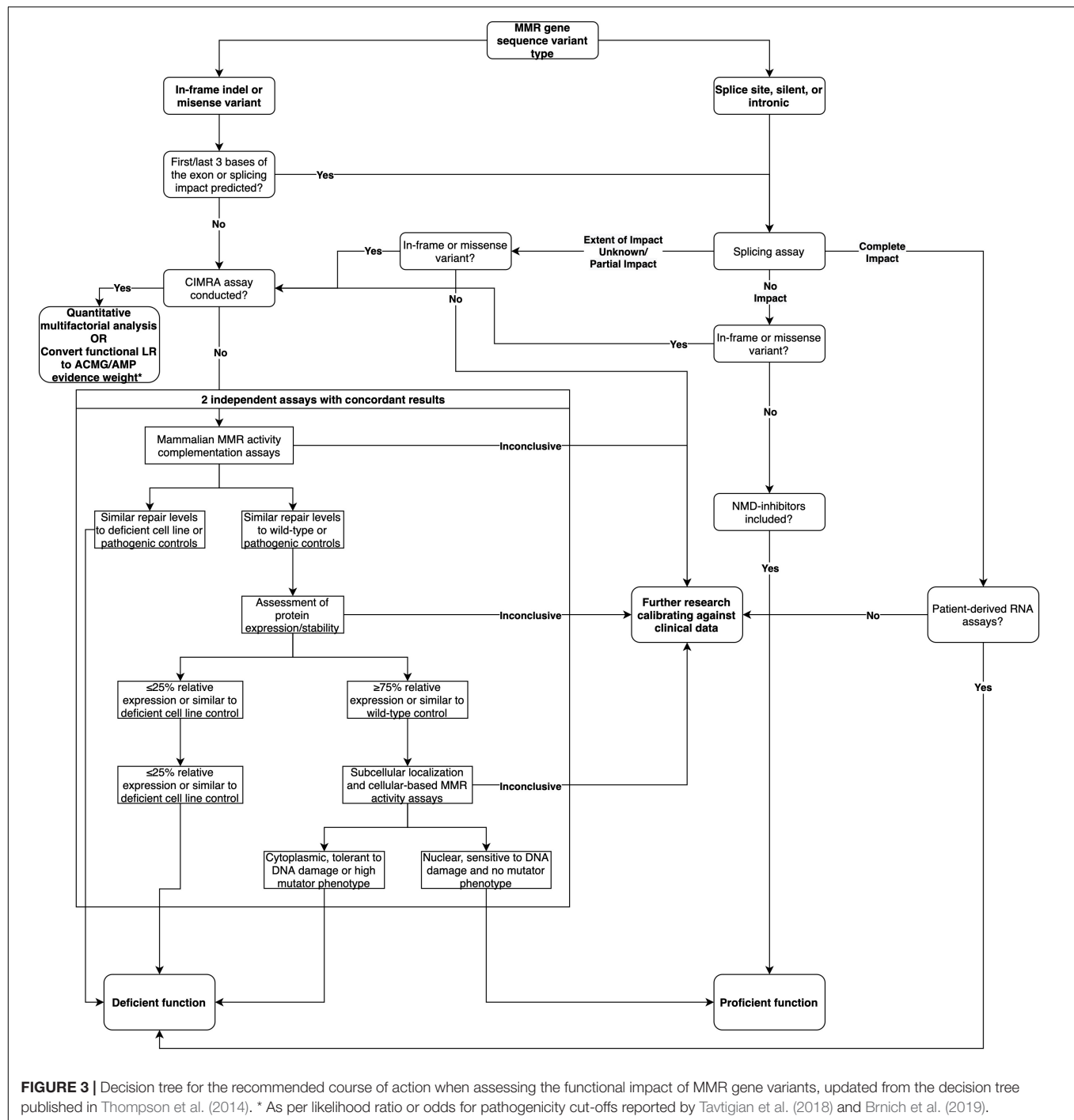


caused complete expression of a splicing aberration. Of the (likely) benign variants, none had deficient MMR activity, and one had moderate MMR activity.

These observations of “conflicting” mRNA splicing and protein functional assays suggest that alternative approaches, which combine assessment of effects at the mRNA and protein level, are required to simplify interpretation on laboratory assay data for MMR gene variant classification. The assay recently developed for BRCA1 (Findlay et al., 2018), saturation genome editing followed by mRNA expression and cellular loss of function, has demonstrated the feasibility and utility of such combined assays for variant interpretation. However, this specific approach would have to be adapted to account for the fact that unlike BRCA1, the MMR genes are not essential

(Blomen et al., 2015). In this regard, an assay based on gene editing of human embryonic stem cells and assessment of both DNA damage response and microsatellite repair was recently developed, holding great promise for the study of variant-induced splicing changes and missense alterations in Lynch syndrome (Rath et al., 2019).

There were 33 variants that demonstrated incomplete impact with respect to expression of aberrant transcripts (see **Figure 2** and **Supplementary Table S3**). Seven of these were frameshift/nonsense variants for which mRNA products are expected to undergo NMD, and thus classification of these variants as pathogenic is unaltered by the mRNA findings. Another 23 were exonic predicted missense/in-frame alterations of the translated protein; protein assay data available for 15/23



variants showed that nine had clear impact on function due to the missense alteration, and another two had moderate function considered to be borderline deficient. That is, protein assay results would inform classification in favor of pathogenicity for 9/15 variants irrespective of the equivocal nature of the mRNA results. Three silent variants (located in the last 3 bp of the exon) and an intronic variant located in the splice donor motif also demonstrated incomplete impact on mRNA splicing, which did not contribute to their classification.

It will be necessary to determine, for variants with incomplete impact on mRNA splicing, what proportion of alternatively spliced MMR gene transcript arising from a variant allele will or will not confer pathogenicity *in vivo*, where a second somatic hit may play a role. It has been shown that a *BRCA1* spliceogenic variant resulting in 70–80% expression of a non-functional transcript (de la Hoya et al., 2016) is not risk-associated. There is some evidence to suggest that the tolerable level of expression may be similar for *MSH2*; *MSH2* c.1275A > G,

reported to be associated with 70% expression of aberrant transcript r.[1229_1276del, 1275a > g] (Morak et al., 2019), is currently classified as a VUS but with accumulating clinical evidence trending toward likely benign. While, evidence from a knock-down study assessing correlation between total mRNA expression levels and MMR protein relative repair activity in human fibroblast cell lines (Kansikas et al., 2014) indicates that ~25% *MLH1* or *MSH2* mRNA expression results in abrogated repair activity. However, it is difficult to interpret the relevance of these apparently conflicting findings in the context of tumorigenesis *in vivo*. We conclude that further research is necessary to elucidate the relationship between MMR gene transcript expression level in human cells and disease risk.

Methods that enable quantification of the proportion of aberrantly spliced transcripts arising from a variant allele, such as recently developed RNA massively parallel sequencing assays (Farber-Katz et al., 2018; Karam et al., 2019), will aid in the interpretation of cases that demonstrate expression of naturally occurring alternatively spliced transcripts and greatly improve the contribution of splicing assays to classification of sequence variants once methods for quantifying transcript expression are routinely instituted. These assays will further increase the use and utility of splicing assay data in variant classification by fulfilling the requirement of quantifying the splicing defect to ensure no full-length transcript is expressed, as currently documented in the InSiGHT MMR gene classification rules (Thompson et al., 2014). This will be particularly useful as supporting clinical data are harder to obtain as more variants of uncertain significance are identified through higher throughput clinical gene panel testing.

In summary, based on the analysis of this dataset, we show that splicing assays are a useful adjunct to the interpretation of intronic and silent variants. While mRNA analysis can contribute to the classification of predicted missense/in-frame indel variants, results have to be considered in parallel with data from MMR activity assays. Based on these findings, we provide a decision tree for the recommended course of action when assessing the functional impact of MMR gene variants (Figure 3). We conclude that there is need to develop and validate different high throughput assays that can measure variant effects on cellular function due to mRNA transcripts and/or protein function—due to a variety of biochemical effects—to streamline future MMR gene variant classification.

MEMBERS OF THE InSiGHT VARIANT INTERPRETATION COMMITTEE

Fahd Al-Mulla, Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait City, Kuwait; Daniel Buchanan, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia, and Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC, Australia; Susan Farrington, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, United Kingdom; Ian Frayling, Institute of Medical

Genetics, University Hospital of Wales, Cardiff, United Kingdom; Maurizio Genuardi, Fondazione Policlinico Universitario A. Gemelli IRCCS, UOC Genetica Medica, Rome, Italy, and Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore, Rome, Italy; Elke Holinski-Feder, Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Munich, Germany, and Center of Medical Genetics, Munich, Germany; Maija R. J. Kohonen-Corish, Woolcock Institute of Medical Research, Sydney, NSW, Australia, and University of Technology Sydney, Sydney, NSW, Australia; Andreas Laner, Medizinisch Genetisches Zentrum, Munich, Germany; Alexandra Martins, INSERM-U1245, UNIROUEN, Normandy Centre for Genomic and Personalized Medicine, Normandie University, Rouen, France; Finlay Macrae, Genetic Medicine, The Royal Melbourne Hospital, Melbourne, VIC, Australia, and Department of Medicine, The University of Melbourne, Melbourne, VIC, Australia; Pål Møller, Department of Tumor Biology, The Norwegian Radium Hospital, Part of Oslo University Hospital, Oslo, Norway; Monika Morak, Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Munich, Germany, and MGZ – Medical Genetics Center, Munich, Germany; Elisabet Ognedal, Haukeland Universitetssjukehus, Bergen, Norway; John-Paul Plazzer, The Royal Melbourne Hospital, Melbourne, VIC, Australia; Lene Juel Rasmussen, Center for Healthy Aging, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; Carli Tops, Department of Clinical Genetics, Leiden University Medical Centre, Netherlands; Ingrid Winship, Genetic Medicine, The Royal Melbourne Hospital, Melbourne, VIC, Australia, and Department of Medicine, The University of Melbourne, Melbourne, VIC, Australia.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by QIMR Berghofer Human Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BT and AS contributed to the conception and design of the study. BT, RW, MP, TD, MD, YT, NL, NW, and ST contributed to the data acquisition and interpretation of the study. BT performed the data analysis and wrote the first draft of the manuscript. BT and AS wrote the sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was funded by US NIH NCI grants R01 CA164944 and UM1 CA167551 and through U01/U24 cooperative agreements from NCI with the following Colon CFR centers: Mayo Clinic (CA074800 to NL), Ontario (OFCCR) (CA074783), and Seattle (SCCFR) (CA074794). The content of this manuscript does not necessarily reflect the views or policies of the NIH or any of the collaborating centers in the CCFR, nor does the mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the Colon CFR. NW, MD, and YT were supported by the Dutch Digestive Foundation (Grant FP 16-01) and the Dutch Cancer Society (Grant UL 2013-5939). BT was supported by an NHMRC CJ Martin Early Career Fellowship (ID1091211). AS was supported by an NHMRC Senior Research Fellowship (ID1061779).

REFERENCES

- Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., et al. (2018). Recommendations for interpreting the loss of function PV51 ACMG/AMP variant criterion. *Hum. Mutat.* 39, 1517–1524. doi: 10.1002/humu.23626
- Blomen, V. A., Májek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096. doi: 10.1126/science.aac7557
- Brnich, S. E., Abou Tayoun, A. N., Couch, F. J., Cutting, G. R., Greenblatt, M. S., Heinen, C. D., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 12:3.
- Buchanan, D. D., Tan, Y. Y., Walsh, M. D., Clendenning, M., Metcalf, A. M., Ferguson, K., et al. (2014). Tumor mismatch repair immunohistochemistry and DNA MLH1 methylation testing of patients with endometrial cancer diagnosed at age younger than 60 years optimizes triage for population-level germline mismatch repair gene mutation testing. *J. Clin. Oncol.* 32, 90–100. doi: 10.1200/JCO.2013.51.2129
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298. doi: 10.1038/nrg775
- de la Hoya, M., Soukarieh, O., López-Perolio, I., Vega, A., Walker, L. C., van Ierland, Y., et al. (2016). Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum. Mol. Genet.* 25, 2256–2268. doi: 10.1093/hmg/ddw094
- Dominguez-Valentin, M., Sampson, J. R., Seppälä, T. T., Ten Broeke, S. W., Plazzer, J. P., Nakken, S., et al. (2020). Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective lynch syndrome database. *Genet. Med.* 22, 15–25. doi: 10.1038/s41436-019-0596-9
- Drost, M., Tiersma, Y., Glubb, D., Kathe, M., van Hees, S., Calléja, F., et al. (2020). Two integrated and highly predictive functional analysis-based procedures for the classification of MSH6 variants in Lynch syndrome. *Genet. Med.* 22, 847–856. doi: 10.1038/s41436-019-0736-2
- Drost, M., Tiersma, Y., Thompson, B. A., Frederiksen, J. H., Keijzers, G., Glubb, D., et al. (2018). A functional assay-based procedure to classify mismatch repair gene variants in Lynch syndrome. *Genet. Med.* 21, 1486–1496. doi: 10.1038/s41436-018-0372-2
- Farber-Katz, S., Hsuan, V., Wu, S., Landrith, T., Vuong, H., Xu, D., et al. (2018). Quantitative analysis of BRCA1 and BRCA2 germline splicing variants using a novel RNA-massively parallel sequencing assay. *Front. Oncol.* 8:286. doi: 10.3389/fonc.2018.00286

ACKNOWLEDGMENTS

We thank Jannah Shamsani for providing MES-SWA scores for the MMR gene variants. We also acknowledge the Australian Red Cross Blood Services (ARCBS) donors who participated as healthy controls in this study. We are grateful to Rachel Morris and the staff at ARCBS for their assistance with the collection of risk factor information and blood samples, and Melanie Higgins, Kimberley Hinze, Felicity Lose, and members of the Molecular Cancer Epidemiology Laboratory for their assistance with collection and processing of blood samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00798/full#supplementary-material>

- Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., et al. (2018). 3 Accurate classification of BRCA1 variants with saturation genome editing. *Nature* 562, 217–222. doi: 10.1038/s41586-018-0461-z
- Grandval, P., Fabre, A. J., Gaildrat, P., Baert-Desurmont, S., Buisine, M. P., Ferrari, A., et al. (2013). UMD-MLH1/MSH2/MSH6 databases: description and analysis of genetic variations in French Lynch syndrome families. *Database* 2013:bat036. doi: 10.1093/database/bat036
- Houdayer, C., Dehainault, C., Mattler, C., Michaux, D., Caux-Moncoutier, V., Pagès-Berhouet, S., et al. (2008). Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum. Mutat.* 29, 975–982. doi: 10.1002/humu.20765
- Jenkins, M. A., Win, A. K., Templeton, A. S., Angelakos, M. S., Buchanan, D. D., Cotterchio, M., et al. (2018). Cohort profile: the colon cancer family registry cohort (CCFRC). *Int. J. Epidemiol.* 47, 387–388. doi: 10.1093/ije/dyy006
- Kansikas, M., Kasela, M., Kantelinen, J., and Nyström, M. (2014). Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. *Hum. Mutat.* 35, 1123–1127. doi: 10.1002/humu.22605
- Karam, R., Conner, B., LaDuca, H., McGoldrick, K., Krempely, K., Richardson, M. E., et al. (2019). Assessment of diagnostic outcomes of RNA genetic testing for hereditary cancer. *JAMA Netw. Open* 2:e1913900. doi: 10.1001/jamanetworkopen.2019.13900
- Li, S., Qian, D., Thompson, B. A., Gutierrez, S., Wu, S., Pesaran, T., et al. (2020). Tumour characteristics provide evidence for germline mismatch repair missense variant pathogenicity. *J. Med. Genet.* 57, 62–69. doi: 10.1136/jmedgenet-2019-106096
- Morak, M., Schaefer, K., Steinke-Lange, V., Koehler, U., Keinath, S., Massdorf, T., et al. (2019). Full-length transcript amplification and sequencing as universal method to test mRNA integrity and allelic expression in mismatch repair genes. *Eur. J. Hum. Genet.* 27, 1808–1820. doi: 10.1038/s41431-019-0472-8
- Parmley, J. L., and Hurst, L. D. (2007). How do synonymous mutations affect fitness? *BioEssays* 29, 515–519. doi: 10.1002/bies.20592
- Parmley, J. L., and Huynen, M. A. (2009). Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. *PLoS Genet.* 5:e1000548. doi: 10.1371/journal.pgen.1000548
- Rath, A., Mishra, A., Ferreira, V. D., Hu, C., Omerza, G., Kelly, K., et al. (2019). Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Hum. Mutat.* 40, 2044–2056. doi: 10.1002/humu.23848
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–423. doi: 10.1038/gim.2015.30

- Shamsani, J., Kazakoff, S. H., Armean, I. M., McLaren, W., Parsons, M. T., Thompson, B. A., et al. (2018). A plugin for the ensembl variant effect predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinformatics* 35, 2315–2317. doi: 10.1093/bioinformatics/bty960
- Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., Boucher, K. M., et al. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* 20, 1054–1060. doi: 10.1038/gim.2017.210
- Thompson, B. A., Goldgar, D. E., Paterson, C., Clendenning, M., Walters, R., Arnold, S., et al. (2013a). A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the colon cancer family registry. *Hum. Mutat.* 34, 200–209. doi: 10.1002/humu.22213
- Thompson, B. A., Greenblatt, M. S., Vallee, M. P., Herkert, J. C., Tessereau, C., Young, E. L., et al. (2013b). Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum. Mutat.* 34, 255–265. doi: 10.1002/humu.22214
- Thompson, B. A., Martins, A., and Spurdle, A. B. (2015). A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin. Genet.* 87, 100–108. doi: 10.1111/cge.12450
- Thompson, B. A., Spurdle, A. B., Plazzer, J. P., Greenblatt, M. S., Akagi, K., Al-Mulla, F., et al. (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.* 46, 107–115. doi: 10.1038/ng.2854
- Whiley, P. J., Parsons, M. T., Leary, J., Tucker, K., Warwick, L., Dopita, B., et al. (2014). Multifactorial likelihood assessment of BRCA1 and BRCA2 missense variants confirms that BRCA1:c.122A>G(p.His41Arg) is a pathogenic mutation. *PLoS One* 9:e86836. doi: 10.1371/journal.pone.0086836

Conflict of Interest: ST holds Illumina stock in a personally managed account.

The remaining 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 CH declared a past co-authorship with one of the author ST to the handling editor.

Copyright © 2020 Thompson, Walters, Parsons, Dumenil, Drost, Tiersma, Lindor, Tavtigian, de Wind, Spurdle and the InSiGHT Variant Interpretation Committee. 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.

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



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