

mRNA SPLICING-MEDIATED GENE
EXPRESSION REGULATION IN INNATE
IMMUNITY

Thesis by
Luke Frankiw

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
(Defended May 17th, 2019)

© 2019

Luke Frankiw

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest thanks to my advisor, David Baltimore. David has a genuine love of science, a love that from the perspective of a young scientist, is inspiring. I struggle to put into words the significant impact he has had on my life. When I first sat down with David to discuss joining his lab, it was apparent that I likely would not finish before the lab shut-down date, which had been set for some time. While most would have told me it wouldn't work, David went out of his way to help me return to Caltech. He delayed closing the lab to allow time for me to finish this work, which speaks to the type of person David is. For this, I will forever be grateful. Having had the chance to meet lab alumni from the last half century, what stands out is the sheer number of lives David has been able to impact as a mentor. In summarizing his career, David expressed that having been given the chance to guide his students through the enterprise of science, was the gift of a lifetime. I am sure I can say on behalf of all his students, that the mentorship David provided is a gift that will go on for many lifetimes. David, it is an honor to have been your last student. My time in your lab is something I will cherish for the rest of my life.

I also wish to acknowledge my thesis committee, Mitch Guttman, Sarkis Mazmanian, and Michael Elowitz.

Mitch, you have been a major figure throughout my Ph.D. Having had the opportunity to rotate in your lab helped me develop some of the computational skills I needed to complete this work. You are a remarkable scientist and I appreciate that you always made time for me when I needed help or advice.

Sarkis, thank you for pushing me to be both thorough and focused. It was extremely helpful running through my future directions with you at committee meetings. You were exceptional at helping me parse through the experiments to determine those that were necessary and unnecessary for a publication.

Michael, I feel lucky to have you on my committee. You provided a unique and refreshing perspective, especially with respect to splicing and splicing kinetics.

I would like to thank those in the Baltimore lab who have played an instrumental role in my graduate school experience, particularly the following people:

Devdoot Majumdar, who mentored me both as an undergraduate and graduate student in the lab.

Alok Joglekar, for keeping me in line at the lab, which I can admit, I needed at times. More importantly, Alok taught me how to think like a scientist, and for that I am extremely grateful.

Mati Mann, who for the latter half of my time in grad school has been isolated with me on the far side of the lab. While I'm sure from his perspective it may have been extraordinarily annoying, Mati always made time for my consistent and frequent "quick questions".

Guideng Li, whom I have collaborated with more recently, taught me a great deal on how to be both practical and efficient in the lab. I've enjoyed my time working with him, and hope we can continue to collaborate in the future.

Christian Burns, who I was lucky enough to work beside at the start of grad school, and who had a fundamental role in setting up many of the techniques used in this thesis.

Jocelyn Kim and Arnav Mehta, for offering insight and guidance with respect to the physician-scientist career path.

Michael Leonard, for both his help with the many computational problems I encountered and his unique sense of humor.

Beyond the lab, there were a number of people who greatly contributed to my graduate school experience.

Prashant Bhat, who was always willing to take a quick trip to Red Door to hear me complain about failed experiments or listen to me pitch a project.

Yapeng Su, who provided me with company and advice many late nights in lab.

My Caltech gym group, Bill Ling, Constantine Sideris, Nikolay Dedushenko, Szilard Szoke, and Sheel Shah. Our time in the gym was primarily spent talking and joking, with few weights actually being lifted; but nonetheless, that time provided a great release from the lab.

Additionally, I would like to acknowledge Jennifer Chan at the University of Calgary, who was my first mentor and who introduced me to the joy of experimental biology.

Lastly, I would like to thank my family, who has provided unconditional support. To my parents, Marsha and Ken, you have been the greatest source of motivation through the years. You provided my brothers and me the opportunity to pursue our own goals and supported us through our journeys. To my brothers, Mark and Andrew, thank you for always being there for me and pushing me to be the best I can be, even if it was something as trivial as mini hockey in the basement. Finally, to my grandparents, you set great examples for your grandchildren on how to live life and for that, we are all thankful.

ABSTRACT

At the heart of an inflammatory response lies a tightly regulated gene expression program. Perturbations to this finely tuned response can result in unchecked or inappropriately scaled inflammation, shifting the balance from protective to destructive immunity. A variety of post-transcriptional mechanisms play a role in the fine-tuning of an inflammatory gene expression program. One such mechanism involves unproductive RNA splicing, whereby alternative splicing can frameshift the transcript or introduce a premature termination codon (PTC). These effects render the transcript nonfunctional and/or subject it to nonsense-mediated decay.

We observed such an event in *Irf7*, the master regulator of the type I interferon response. We found a single intron was consistently retained at a level much greater than other introns in the *Irf7* transcript. In an effort to understand *trans*-acting factors that regulate this retention, we used RNA-antisense purification followed by mass spectrometry (RAP-MS) to identify the factor BUD13 as a highly enriched protein on *Irf7* transcripts. Deficiency in BUD13 was associated with increased retention, decreased mature *Irf7* transcript and protein levels, and consequently a dampened type I interferon response, which compromised the ability of BUD13-deficient macrophages to withstand vesicular stomatitis virus (VSV) infection.

Beyond this intron retention event in *Irf7*, we identified a variety of other unproductive splicing events in a number of important genes involved with the innate immune response. This unproductive splicing was not restricted to intron retention events. For example, we identified a frequently used alternative splice site in the crucial murine antiviral response gene, oligoadenylate synthetase 1g (*Oas1g*) that led to both a frameshift and incorporation of a PTC. Genome editing was used to remove the alternative splice site in a macrophage cell line, which led to both increased *Oas1g* expression and improved viral clearance. We hypothesize these events exist as a means of mitigation for what might otherwise be an inappropriately scaled response. In doing so, they represent a previously underappreciated layer of gene expression regulation in innate immunity.

PUBLISHED CONTENT AND CONTRIBUTIONS

- Majumdar, D., Frankiw, L., Burns, C. H., Garcia-Flores, Y. & Baltimore, D. Programmed Delayed Splicing: A Mechanism for Timed Inflammatory Gene Expression. *bioRxiv* 443796 (2018).

DSM/LF/CB/DB designed the experiments and wrote the manuscript. DSM/LF/CB/YG designed and conducted experiments, and DSM/LF performed bioinformatics analysis on the data.

- Frankiw, L. *et al.* Bud13 Promotes a Type I Interferon Response by Countering Intron Retention in Irf7. *Molecular cell* **73**, 803–814 (2019).

L.F., D.M., and D.B., conceived and designed experiments. L.F. conducted experiments. C.B. helped develop RAP-MS and knockdown experiments. L.F. and D.M. analyzed sequencing data. A.M. oversaw mass spectrometry and M.J.S. performed mass spectrometry analysis. L.S.F, D.M., and D.B wrote the manuscript with input from all authors.

- Frankiw, L. *et al.* Alternative Splicing Coupled with NMD Acts to Mitigate OAS1 Antiviral Activity. *Manuscript in Preparation*

L.F. and D.B. conceived and designed experiments. L.F. conducted experiments and analyzed data. The manuscript was written by L.F and D.B.

- Frankiw, L. *et al.* Alternative mRNA Processing in Cancer Immunotherapy. *In Review*

L.F. D.B. and G.L. wrote and prepared the manuscript.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract	vi
Published Content and Contributions.....	vii
Table of Contents.....	viii
Chapter 1: Introduction.....	1
Introduction to Splicing and Post-Transcriptional Regulation.....	2
Figure 1.1: Mechanisms of Post-Transcriptional Regulation	4
Figure 1.2: Unproductive Alternative Splicing.....	4
Figure 1.3: Autoregulatory Splicing-Mediated Negative Feedback.....	5
Overview of Thesis	15
References	17
Chapter 2: A Study of the Kinetics of mRNA Splicing in the NF-κB Transcriptome.....	25
Abstract.....	26
Introduction	27
Results	28
Discussion	36
Experimental Procedures	38
Figure Legends.....	41
References	44
Figures	47
Figure 1: RNA Antisense Purification (RAP).....	47
Figure 2: Enrichment of targeted reads following RAP	48
Figure 3: CoSI splicing analysis.....	49
Figure 4: <i>Irf7</i> contains an inefficiently spliced intron.....	50
Figure 5: Poorly spliced introns in <i>CD40</i> and <i>Daxx</i>	51
Figure 6: Retained introns contain conserved, weak 5' splice sites	52
Figure 7: Heterogeneity in the rates of intron splicing.	53
Figure 8: Splicing correlates with the timing of expression	54
Chapter 3: BUD13 Promotes a Type I Interferon Response by Countering Intron Retention in <i>Irf7</i>	55
Abstract.....	56
Introduction	57
Results	58
Discussion	69
Acknowledgements	74
Experimental Procedures	75

Figure Legends	88
References	93
Supplemental Figure Legends.....	97
Figures	101
Figure 1: <i>Irf7</i> contains a weak intron that is retained following many forms of stimulation.....	101
Figure 2: RAP-MS and RIP identify BUD13 as an RNA binding protein that interacts with <i>Irf7</i> mRNA	103
Figure 3: BUD13 knockdown leads to increased retention in the Poorly spliced intron of <i>Irf7</i>	104
Figure 4: BUD13 knockdown alters the type I interferon response ...	105
Figure 5: Global analysis of the role of BUD13	107
Figure 6: BUD13 interacts primarily near the 3' splice site of small, GC rich introns.....	108
Figure 7: BUD13 knockdown alters the BMDM response to VSV ...	110
Figure S1: Splicing Ratios across all junctions in <i>Irf7</i>	111
Figure S2: shBUD13 knocks down BUD13 protein and mRNA	112
Figure S3: <i>Irf7</i> intron 4 is the most BUD13 knockdown affected junction of all ISGs	113
Figure S4: BUD13 knockdown alters the type I interferon response to CpG.....	114
Figure S5: Supplemental global analysis of BUD13	115
Figure S6: BUD13 knockdown alters BMDM infection via VSV	116
Figure S7: Knockdown of other RES complex proteins.....	117
Chapter 4: Alternative Splicing Coupled with NMD Acts to Mitigate OAS1 Antiviral Activity.....	118
Abstract.....	119
Introduction	120
Results	121
Discussion	126
Acknowledgements	129
Experimental Procedures	130
Figure Legends.....	135
References	137
Supplemental Figure Legends.....	131
Figures	143
Figure 1: <i>Oas1g</i> has a frequent AS-NMD event.....	143
Figure 2: Removal of alternative splice site alters <i>Oas1g</i> expression and macrophage response to EMCV.....	144
Figure 3: A similar AS-NMD event occurs in human macrophages..	145
Figure 4: AS-NMD events are common in transcripts related to Innate immunity.....	146
Figure S1: The alternative 5' splice site mediating the AS event is of similar strength to the consensus 5' splice site	147

Figure S2: <i>Oas1g</i> Macrophage cell line genotyping.....	148
Figure S3: <i>Oas1a</i> has a similar frequently utilized AS-NMD event..	149
Figure S4: <i>Oas1a</i> macrophage cell line genotyping.....	150
Chapter 5: Conclusions and Future Directions	151
Concluding Remarks	152
Summary	153
References	158

Chapter 1

**INTRODUCTION TO mRNA SPLICING AND POST-
TRANSCRIPTIONAL REGULATION**

mRNA Splicing and Post-Transcriptional Regulation

While transcription is the most well-scrutinized area of gene expression regulation^{1,2}, there have emerged a variety of post-transcriptional mechanisms that play a role in the fine-tuning of a gene expression program (Figure 1.1). These post-transcriptional mechanisms act in concert to ensure proper expression of individual transcripts given the cell-type and environment. In general, such regulation involves the alteration of either lifespan, localization, or translational efficiency of a given RNA molecule. In doing so, a cell can effectively adjust at the post-transcriptional level how many functional gene products are produced during a given period of time.

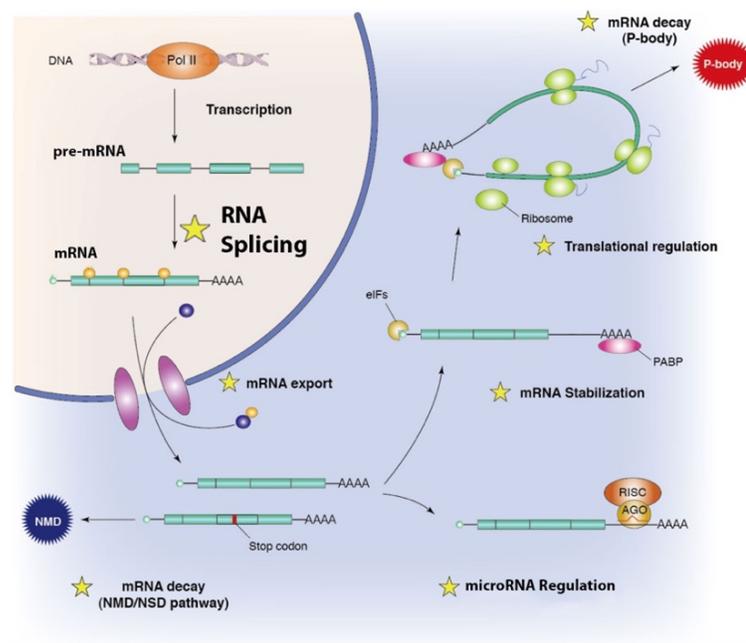


Figure 1.1: Common mechanisms of post-transcriptional gene expression regulation. (Adapted from *ruo.mbl.co*)

One mechanism that in recent years has emerged as an important mediator of post-transcriptional gene expression regulation is RNA splicing. The vast majority of human

genes contain multiple exons, with adjoining intronic sequences that need to be spliced from a transcribed pre-mRNA to form the mature mRNA. Through a process known as alternative splicing (AS) a single pre-mRNA can be variably spliced into unique mature transcripts. The advent of next generation sequencing has led to a wealth of transcriptomic data over the past decade. This, coupled with the development of computational tools that allow proper analysis of splicing events (Box 1), has shed light on the widespread nature of AS. This process is extensively controlled in different tissues, cell types, and differentiation stages³⁻⁸, and dysregulation is believed to be significantly contribute to the development of human^{9,10} disease¹¹⁻¹³.

Despite the fact most mammalian genes exhibit alternative splicing^{14,15}, not all of the produced transcripts encode functional proteins (Figure 1.2). In cases where translation does occur, different splice variants can lead to different protein functions. However, in addition to these cases, which we define as productive splicing events, AS can also generate unproductive isoforms that are not translated and are either subjected to decay or have their expression restricted to the nucleus¹⁶. Starting with the latter, this chapter will cover RNA splicing as a mechanism of post-transcriptional gene expression and further, discuss the relevance such regulation has with respect to human physiology and disease.

Unproductive alternative splicing

While its evident alternative splicing in certain cases acts to increase proteomic diversity, it can also generate isoforms that either shift the frame of the transcript, altering the coding region in the process, and/or lead to the incorporation of a premature termination codon (PTC). The vast majority of these isoforms are subjected to decay^{17,18}, primarily by the

Nonsense-Mediated Decay (NMD) pathway¹⁹ in the case of PTC-containing transcripts. However, there exists a nuclear exosome, which is responsible for degradation of unspliced RNAs in the nucleus^{20,21}. In essence, by preventing the production of the final gene product, these events allow a cell to functionally down-regulate expression of a given gene at the post-transcriptional level. This mechanism of regulation has been called AS coupled to NMD (AS-NMD) or Regulated Unproductive Splicing and Translation (RUST)²², the latter being the term we will use to broadly describe all unproductive splicing events (with the exception of the fourth chapter, where the term AS-NMD is used due to likely degradation of *Oas1g* transcripts by the NMD machinery). It has been estimated between 10-30% of mammalian genes may be regulated post-transcriptionally through RUST, potentially in a context specific manner, through unproductive splicing^{17,18,23,24}.

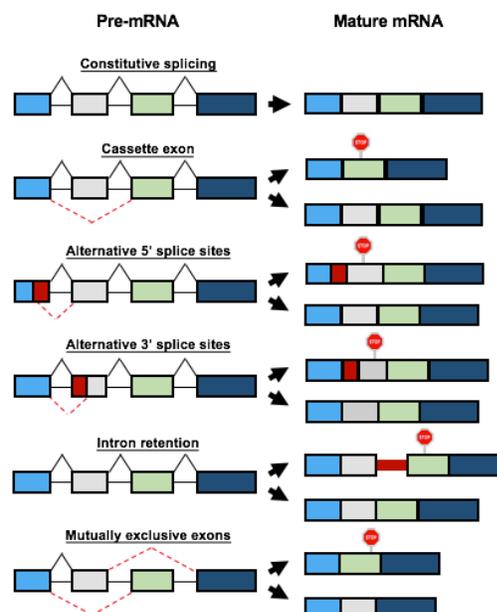


Figure 1.2. A schematic depiction of constitutive splicing as well as the different forms of alternative splicing. Depicted on the left are the pre-mRNA with the possible splice choices represented. Depicted on the right are the possible mature mRNA isoforms from each alternative splicing event. The stop sign represents the inclusion of a premature termination codon (PTC), which thus subjects the transcript to decay.

The coupling of alternative splicing to decay has been shown to affect a variety of biological processes. Perhaps most well studied are the consequences of RUST in the autoregulation of splicing factor genes. Initially discovered to impact the autoregulation of serine/arginine-rich (SR) proteins in *C. elegans*²⁵, RUST is now recognized as a wide-spread mechanisms of splicing factor autoregulation^{24,26-30}. With respect to SR protein autoregulation, this occurs through the establishment of a negative feed-back loop. SR proteins bind to *cis* elements of a pre-mRNA molecule and can promote the inclusion of an exon at a nearby junction³¹. An increase concentration of a given SR protein can lead to the inclusion of a cassette exon in its own transcript, which contains a PTC and thus subjects the transcript to NMD decay (Figure 1.3). Such exons are termed poison cassette exon³². This leads to less production of the SR protein, in turn reducing the inclusion of the poison cassette exon. In addition to autoregulation, cross-regulation can occur where RUST is used to regulate the expression of other splice factors, exemplified by the negative regulation of PTBP2 by PTBP1³³. However, the impact of RUST extends beyond splicing factor regulation. More recently, RUST has been implicated in the regulation of transcripts that play a role in cellular differentiation^{4-6,34-36}, chromatin modification³⁷, and inflammation^{7,38}.

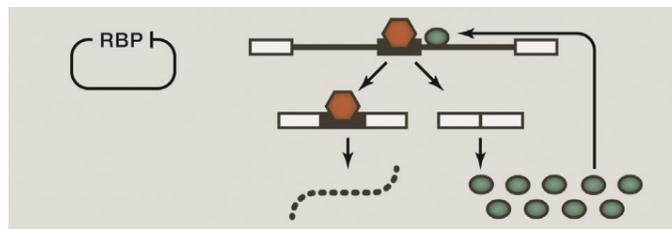


Figure 1.3. Schematic depiction of auto-regulatory negative feedback. The coded splicing factor promotes the inclusion of a poison cassette exon, thus regulating its own NMD event. When levels of the splicing factor are high, the unproductive transcript incorporating the poison cassette is frequently used, leading to a reduction in splicing factor production. When levels of the splicing factor are low, the productive transcript is frequently used, leading to an increase in splicing factor production. Adapted from Jangi and Sharp²⁴.

In the last decade, with respect to unproductive splicing, IR has emerged as a previously underappreciated form of AS that mediates post-transcriptional gene expression regulation. The transient nature of an intron, as well as transcripts that retain introns due to swift degradation, has made it hard to identify and quantify retention³⁹. However, improvements in next-generation sequencing have played a large role in changing this. While there are cases of IR containing transcripts being translated⁴⁰⁻⁴², the vast majority of these isoforms are subjected to decay^{17,18} either via the NMD pathway¹⁹ or the nuclear exosome^{20,21}. Thus, as was the case for RUST mediated by other AS events, IR can functionally “tune” the transcriptome of a cell⁴³. IR as a mechanism of gene expression regulation is maybe best exemplified through the study of cellular differentiation. One notable study involved the impact of IR events on granulopoiesis⁵. As a granulocyte develops from a promyelocyte, IR is found to be dramatically upregulated in a subset of junctions, ultimately leading to a decrease in the corresponding proteins expression. Accompanying this IR is the downregulation of a number of important spliceosome and splicing factor genes. Many of these IR events are found in genes coding for proteins that play an important role in nuclear structure, an interesting finding considering granulocytes have a multi-lobed, highly deformable nuclear morphology which allows them to move through tissue interstitial spaces⁴⁴. One particular example involved LaminB1 (*Lmnb1*), whereby retention in a number of introns in the transcript was upregulated >100-fold and mRNA expression was down-regulated 100-fold. Expression of intronless *Lmnb1* reduced granulocyte numbers and altered nuclear morphology. Thus, it was concluded that

orchestrated IR coupled with degradation could act as a physiological mechanism of gene expression control, ultimately affecting an important cellular differentiation pathway.

In contrast to regulation via degradation, a scenario where these IR events are spliced in a delayed or regulated manner is also possible. For example, in LPS induced macrophages, certain introns were found to represent rate-limiting intermediates, thus acting as a timing mechanism altering the kinetics of gene expression^{38,45-47}. Another scenario involves a group of introns that have been defined as “detained introns” (DIs)⁴⁸. Such introns are defined as being unspliced in otherwise fully spliced polyadenylated transcripts, which in turn leads to nuclear retention of the transcript^{44,48}. While it is possible some degradation occurs in the nucleus via the nuclear exosome, these transcripts are insensitive to NMD due to their nuclear localization. Regardless, they negatively regulate protein expression as they are not translated. This negative regulation can be removed through post-transcriptional splicing, which has been shown to occur in response to certain stimuli^{4,34,48,49}. Alteration of DI splicing can have major physiological consequences. It was recently shown that in glioblastomas (GBM), the arginine methyltransferase PRMT5, upregulated in high grade gliomas^{50,51}, regulates the splicing of detained introns through the modulation of snRNP biogenesis⁴⁹. Inhibition of PRMT5 had an antitumor effect, believed to be mediated through an increase in DIs. Many of these DIs were located in genes that were predominantly associated with proliferation and neurogenesis. It was concluded that GBMs, through PRMT5, assume control of DI splicing allowing them to upregulate a gene expression program suited for proliferation. With respect to slow-splicing junctions and DIs, a great deal of work is needed to understand the exact nature of such junctions, to determine the

extent to which they are post-transcriptionally spliced as compared to degraded, and to understand the factors that regulate them.

Other forms of AS mediated post-transcriptional regulation.

In contrast to unproductive AS events, productive AS events do not shift the frame of the transcript and/or lead to the incorporation of a premature termination codon (PTC). As such, the resulting isoforms have the potential to be translated into functional protein products. Still, these AS events have the ability to affect post-transcriptional regulation. Productive AS can lead to a protein product that has an altered function. In many cases, this new function is not consistent with the original function, or even further acts to negatively regulate the original function. This form of regulation is particularly ubiquitous in the innate immune response⁵²⁻⁶⁰. For example, there exists an alternative splice variant of the toll-receptor gene *Tlr4* that introduces an extra exon that contains an in-frame stop codon⁵². *Tlr4* is a pattern recognition receptor (PRR) that is most well-known for recognizing the Gram-negative bacteria component lipopolysaccharide, leading to activation of an innate immune response⁶¹. This in-frame stop codon leads to the generation of a soluble form of TLR4 that still binds LPS, but no longer has the ability to signal to downstream components. Thus, it acts to negatively regulate a *Tlr4* response. Interestingly, many of the AS events that lead to negative regulation of the original protein product in the innate immune response are induced by LPS stimulation, suggesting that such negative regulation is needed to ensure responses are self-limiting^{52,53,60,62}.

Beyond these AS events that lead to productive transcripts, AS in non-coding regions of a transcript can have a major effect on the stability of a transcript. One such case involves the regulation of transcript stability through AS at junctions located in the 3' UTR⁶³. The 3' UTR is an important determinant of transcript stability. Both miRNAs and RNA binding proteins can bind to the 3' UTR and alter transcript stability. With respect to miRNAs, binding at the 3' UTR negatively regulates gene expression⁶⁴. However, in order for this regulation to occur, the miRNA binding site must be present in the transcript. It has been estimated that one third of miRNA binding sites are controlled by AS events⁶⁵. This is exemplified by a case involving the divalent metal transporter 1 gene (*DMT1*), which contains two alternative 3' terminal exons⁶⁶. One isoform contains an iron response element (IRE), the other does not. The isoform lacking this IRE carries a binding site for the miRNA let-7d, which in turn limits expression of this isoform but has no effect on the isoform containing the IRE. Down-regulation of let-7d during erythroid differentiation allows the DMT1 isoform without the IRE element to become more prevalent. While the 3' UTR plays an important role in transcript stability via miRNA regulation, it also has an effect on transcript stability due to binding of RNA-binding proteins (RBPs) to the region. A notable example of this type of regulation involves 3'UTR-enriched AU-rich elements (AREs) that recruit corresponding RBPs⁶⁷⁻⁷⁰ and modulate transcript stability and translational activity⁶³. As was the case with miRNA regulation, ARE mediate regulation can be affected by AS⁷¹. For example, in the human parathyroid hormone-related protein (PTHrP), three alternative 3' terminal exons exist, one of which contains an ARE in its 3' UTR, making it the least stable⁷².

More recently, work in *S. cerevisiae* has identified a new mechanism by which an AS event can post-transcriptionally regulate a gene expression program. A new class of “spliceosome-sequestering” introns were found to play a key role in the cellular response of *S. cerevisiae* to nutrient deprivation^{73,74}. These introns are either retained, forming a hairpin with the 5' UTR⁷³, or they are first excised and then stabilized through interactions with spliceosomal proteins⁷⁴. Then, under stress, these introns modulate splicing by sequestering specific spliceosome components. In the case of *S. cerevisiae* cells under nutrient deprivation, this modulation of splicing leads to dampen the expression of highly expressed genes as a means of energy conservation during starvation.⁷⁵ It will be interesting to see if a similar mechanism of gene expression regulation occurs in both higher eukaryotes and other physiological conditions.

Regulation of AS Events

While a large body of work has identified and classified a number of the AS events that mediated post-transcriptional regulation, it remains unknown the extent to which these events are actively regulated by an external input. A simpler scenario involves transcripts being split between isoforms at a constant ratio. As such, the combined effect of the AS on a gene expression program remains relatively constant. For example, in cases of unproductive splicing where AS is coupled with NMD, the reduction of mRNA abundance occurs by a relatively constant factor. There are a number of examples of such events, including in the widely expressed protease Calpain-10 which has four isoforms that incorporate PTCs and

have shown to be consistently downregulated^{76,77}. However, in all of these situations, such a conclusion is quite difficult to make due to the potential for undiscovered regulatory inputs.

Contrary to constitutive unproductive splicing, it is also possible for external factors to regulate splicing mediated post-transcriptional gene expression regulation. For example, a change in the abundance of splicing factors can promote the production of an unproductive splicing event. This in turn will reduce the number of transcripts and the amount of functional protein products. This is exactly what happens in the aforementioned work involving orchestrated IR during granulocyte differentiation as the observed IR is accompanied by downregulation of spliceosomal components and *trans*-acting splicing factors⁵. In contrast, the factors that regulate other seemingly orchestrated AS events that occur in response to a given stimuli remain poorly understood. During an innate immune response, there is a significant increase in isoform diversity⁷⁸. While some factors have been identified that play a role in either “safe-guarding” proper splicing or regulating a given AS event^{7,62,79}, it remains unknown why the majority of these AS events occur in the presence but not absence of stimulation. This is epitomized by the aforementioned case with *Tlr4*. It is possible that some of this increase in AS is due to burden on the spliceosome, akin to what is seen upon oncogenic MYC activation⁸⁰. Stimulation leads to an orchestrated transcriptional response which in turn increases the amount of RNA that needs to be spliced. Regardless, new RNA centric methods have been developed that allow for the discovery of RNA/protein interactions. Methods like RAP-MS⁸¹ and ChIRP⁸², promise to help identify *trans*-acting factors that regulate these AS events and thus, contribute to post-transcriptional gene expression regulation. In fact, the third chapter of this thesis describes use of RAP-MS to discover an RNA-protein interaction that aids in the splicing efficiency of an IR event in

*Irf7*⁷. This protein, BUD13, was found to be enriched on *Irf7* transcripts using RAP-MS, and deficiency of BUD13 led to increased IR in *Irf7*. This increased IR was associated with impaired induction, a dampened type I interferon response, and consequently an inability to clear virus. As these methods continue to improve, our understanding of external regulation of AS events affecting gene expression should continue to blossom.

Conclusion

Advancements in next-generation sequencing technologies have drastically improved our understanding of AS. This in turn has led to the development of AS as a mechanism of gene expression regulation. Increased wealth of transcriptomic data from a wider variety of cell-types and more physiological conditions promises to further improve our understanding of such regulation. Further, seeing as dysregulated AS has been shown to be widespread in a variety of diseases, from cancer to autoimmunity, our understanding of AS mediated gene expression regulation could shed light on the relationship between alterations in isoform abundance and disease pathogenesis^{13,83-86}.

Inherently, regulation at the post-transcriptional level might seem inefficient. Why spend the resources to transcribe a transcript if it is destined for degradation? First and foremost, the very fact introns exist and are transient in nature argues against the idea that the cost of transcription is prohibitive²². A significant majority of transcribed sequence (~90% in humans⁸⁷) is spliced and discarded. As such, it can be argued the fine-tuning capabilities inherent to splicing based post-transcriptional regulation far outweigh the cellular cost of additional transcription. Additionally, it is well understood that transcriptional regulation is largely a cooperative venture⁸⁸, epitomized by complexes like

the interferon- β (IFN- β) enhanceosome⁸⁹. As transcriptional regulation is not simply one protein interacting with one DNA sequence, but instead a multitude of proteins interacting with a host of other proteins and a variety of DNA sequences, it is quite possible that once transcriptional control has been placed on a system, changing it quantitatively is difficult. Thus, secondary mechanisms, like mRNA splicing, are needed to fine-tune the gene expression levels of select transcripts.

Box 1 – Computational Analysis of RNA Splicing

There are a number of widely used computational tools designed to analyze and quantify differential splicing in RNA-seq data. Broadly, they fall into two main categories. The first category involves tools that utilize a transcript-based approach, whereby an expectation-maximization algorithm is used to estimate isoform abundance^{90,91}. Recent progress in pseudo-alignment has allowed for the development of tools like kallisto⁹² and Salmon⁹³, which perform alignment free isoform quantification and are computationally extraordinarily efficient. Still, the identification and quantification of full-length transcripts from short reads is non-trivial and further, in the case of pseudo-alignment, the approach is inherently sensitive to the input transcript annotations (i.e. an input recording the precise location of intron and exon boundaries), which may be incomplete or inconsistent^{94,95}.

The second category involves tools that utilize an event-based approach. These tools ignore the estimation of isoform expression and instead, detect alternative splicing events by comparing reads at a given junction between multiple samples and quantifies using a metric like percent spliced in (PSI/ Ψ). There are a variety of commonly used tools in this category (i.e. MISO⁹⁶, rMATS⁹⁷, MAJIQ⁹⁸, Leafcutter⁹⁹, SplAdder¹⁰⁰, JUM¹⁰¹ and Whippet¹⁰²). Each has its own intricacies, most notably the statistical methods used to quantify differences between data-sets and the extent to which they rely on a pre-annotation of known alternative splicing events. The latter can alter the number of events detected as programs that can augment an annotation or are annotation free are inherently able to detect a wider range of alternative splicing events. However, for a given set of alternative splicing events, the tools produce very similar PSI values⁹⁵.

Overview of thesis

The overarching theme behind this thesis is the study of RNA splicing as a means of post-transcriptional gene expression regulation. In particular, this work focuses on how splicing-mediated post-transcriptional regulation affects an inflammatory response. Central to an inflammatory response is a robust and coordinated gene expression program. Tight regulation of this gene expression program is essential as small alterations can shift the balance from protective to destructive immunity. While transcriptional control certainly drives such a regulated gene expression program, post-transcriptional regulation has been shown to be essential to aid in the fine-tuning of expression. Inherently, post-transcriptional regulation appears inefficient. Why transcribe a transcript simply to throw it away? But as will be discussed, transcriptional regulation is largely a cooperative venture, involving a host of proteins and a range of *cis* elements. Thus, secondary mechanisms are needed to fine-tune the gene expression levels of select transcripts. The benefit of quickly fine-tuning expression of select transcripts post-transcriptionally far outweighs the cellular cost of additional transcription, especially in the context of a tightly regulated gene expression program like inflammation.

The first chapter will present background information with respect to mRNA splicing and its ability to mediate post-transcriptional regulation, particularly in the fine-tuning of an inflammatory response. Chapter 2 will cover our work quantifying the rates of intron splicing throughout the NF- κ B transcriptome, and our identification of intron retention events that remain extraordinarily unspliced throughout a stimulation time-course. In Chapter 3, we focus on one such retention event in *Irf7*, the master regulator of the type I interferon

response, and attempt to find *trans*-acting factors that regulate the level of retention. Using RNA Antisense Purification followed by Mass Spectrometry (RAP-MS), we identified the RNA binding protein BUD13 as enriched on Irf7, and further perform *Bud13* knockdown studies to show altered Irf7 induction and an impaired anti-viral response in macrophages deficient of *Bud13*. Chapter 4 expands our study of splicing as a post-transcriptional mechanisms of gene expression regulation during inflammation by focusing on alternative splicing events beyond intron retention. While it is true that alternative splicing (AS) can act to increase proteomic diversity, it can also generate unproductive isoforms that incorporate a premature termination codon (PTC), and are thus subjected to NMD or exosomal decay. We focus on one frequently used unproductive splicing event in oligoadenylate synthetase 1g (*Oas1g*), an important murine anti-viral response factor, and show that removal of the alternative splice site mediating this AS event increases expression of *Oas1g* and consequently, improves the ability for macrophages to clear viruses. Finally, Chapter 5 concludes this thesis by offering insight regarding the future directions and potential implications of this work.

References

1. Lee, T. I. & Young, R. A. Transcriptional Regulation and its Misregulation in Disease. *Cell* **152**, 1237–1251 (2013).
2. Smale, S. T. & Natoli, G. Transcriptional control of inflammatory responses. *Cold Spring Harb. Perspect. Biol.* **6**, a016261 (2014).
3. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci.* **105**, 20333–20338 (2008).
4. Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B. & Makeyev, E. V. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.* **26**, 1209–1223 (2012).
5. Wong, J. J.-L. *et al.* Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**, 583–595 (2013).
6. Pimentel, H. *et al.* A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **42**, 4031–4042 (2014).
7. Frankiw, L. *et al.* Bud13 promotes a type I interferon response by countering intron retention in Irf7. *Mol. Cell* **73**, 803–814 (2019).
8. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
9. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269 (2013).
10. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153 (2013).
11. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta BBA-Mol. Basis Dis.* **1792**, 14–26 (2009).
12. Oltean, S. & Bates, D. O. Hallmarks of alternative splicing in cancer. *Oncogene* **33**, 5311 (2014).
13. Coltri, P. P., dos Santos, M. G. & da Silva, G. H. Splicing and cancer: challenges and opportunities. *Wiley Interdiscip. Rev. RNA* e1527 (2019).
14. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470 (2008).

15. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413 (2008).
16. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
17. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci.* **100**, 189–192 (2003).
18. Weischenfeldt, J. *et al.* Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* **13**, R35 (2012).
19. Kurosaki, T., Popp, M. W. & Maquat, L. E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* (2019). doi:10.1038/s41580-019-0126-2
20. Bousquet-Antonelli, C., Presutti, C. & Tollervey, D. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**, 765–775 (2000).
21. Fan, J. *et al.* Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J.* **36**, 2870–2886 (2017).
22. Lareau, L. F., Brooks, A. N., Soergel, D. A., Meng, Q. & Brenner, S. E. The coupling of alternative splicing and nonsense-mediated mRNA decay. (2007).
23. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **36**, 1073 (2004).
24. Jangi, M. & Sharp, P. A. Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498 (2014).
25. Morrison, M., Harris, K. S. & Roth, M. B. smg mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **94**, 9782–9785 (1997).
26. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).

27. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926 (2007).
28. Saltzman, A. L. *et al.* Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell. Biol.* **28**, 4320–4330 (2008).
29. Han, S. P., Tang, Y. H. & Smith, R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.* **430**, 379–392 (2010).
30. Saltzman, A. L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* **25**, 373–384 (2011).
31. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).
32. Lareau, L. F. & Brenner, S. E. Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol. Biol. Evol.* **32**, 1072–1079 (2015).
33. Boutz, P. L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
34. Mauger, O., Lemoine, F. & Scheiffele, P. Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* **92**, 1266–1278 (2016).
35. Edwards, C. R. *et al.* A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood* **127**, e24–e34 (2016).
36. Schmitz, U. *et al.* Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* **18**, 216 (2017).
37. Yan, Q. *et al.* Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc. Natl. Acad. Sci.* **112**, 3445–3450 (2015).
38. Majumdar, D., Frankiw, L., Burns, C. H., Garcia-Flores, Y. & Baltimore, D. Programmed Delayed Splicing: A Mechanism for Timed Inflammatory Gene Expression. *bioRxiv* 443796 (2018).
39. Vanichkina, D. P., Schmitz, U., Wong, J. J.-L. & Rasko, J. E. Challenges in defining the role of intron retention in normal biology and disease. in *Seminars in cell & developmental biology* **75**, 40–49 (Elsevier, 2018).

40. Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A. & Kalyna, M. Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* **25**, 995–1007 (2015).
41. Li, Y. *et al.* An NXF1 mRNA with a retained intron is expressed in hippocampal and neocortical neurons and is translated into a protein that functions as an Nxf1 cofactor. *Mol. Biol. Cell* **27**, 3903–3912 (2016).
42. Rekosh, D. & Hammarskjöld, M.-L. Intron retention in viruses and cellular genes: Detention, border controls and passports. *Wiley Interdiscip. Rev. RNA* **9**, e1470 (2018).
43. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
44. Jacob, A. G. & Smith, C. W. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **136**, 1043–1057 (2017).
45. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
46. Pandya-Jones, A. *et al.* Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *Rna* **19**, 811–827 (2013).
47. Hao, S. & Baltimore, D. RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci.* **110**, 11934–11939 (2013).
48. Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80 (2015).
49. Braun, C. J. *et al.* Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell* **32**, 411–426 (2017).
50. Yan, F. *et al.* Genetic validation of the protein arginine methyltransferase PRMT5 as a candidate therapeutic target in glioblastoma. *Cancer Res.* **74**, 1752–1765 (2014).
51. Han, X. *et al.* Expression of PRMT5 correlates with malignant grade in gliomas and plays a pivotal role in tumor growth in vitro. *J. Neurooncol.* **118**, 61–72 (2014).
52. Iwami, K. *et al.* Cutting edge: naturally occurring soluble form of mouse Toll-like receptor 4 inhibits lipopolysaccharide signaling. *J. Immunol.* **165**, 6682–6686 (2000).
53. Janssens, S., Burns, K., Tschopp, J. & Beyaert, R. Regulation of interleukin-1 and lipopolysaccharide-induced NF- κ B activation by alternative splicing of MyD88. *Curr. Biol.* **12**, 467–471 (2002).

54. Burns, K. *et al.* Inhibition of interleukin 1 receptor/Toll-like receptor signaling through the alternatively spliced, short form of MyD88 is due to its failure to recruit IRAK-4. *J. Exp. Med.* **197**, 263–268 (2003).
55. Janssens, S., Burns, K., Vercammen, E., Tschopp, J. & Beyaert, R. MyD88S, a splice variant of MyD88, differentially modulates NF- κ B-and AP-1-dependent gene expression. *FEBS Lett.* **548**, 103–107 (2003).
56. Lynch, K. W. Consequences of regulated pre-mRNA splicing in the immune system. *Nat. Rev. Immunol.* **4**, 931 (2004).
57. Ohta, S., Bahrun, U., Tanaka, M. & Kimoto, M. Identification of a novel isoform of MD-2 that downregulates lipopolysaccharide signaling. *Biochem. Biophys. Res. Commun.* **323**, 1103–1108 (2004).
58. Rao, N., Nguyen, S., Ngo, K. & Fung-Leung, W.-P. A novel splice variant of interleukin-1 receptor (IL-1R)-associated kinase 1 plays a negative regulatory role in Toll/IL-1R-induced inflammatory signaling. *Mol. Cell. Biol.* **25**, 6521–6532 (2005).
59. Leeman, J. R. & Gilmore, T. D. Alternative splicing in the NF- κ B signaling pathway. *Gene* **423**, 97–107 (2008).
60. Gray, P. *et al.* Identification of a novel human MD-2 splice variant that negatively regulates Lipopolysaccharide-induced TLR4 signaling. *J. Immunol.* **184**, 6359–6366 (2010).
61. Brubaker, S. W., Bonham, K. S., Zanoni, I. & Kagan, J. C. Innate immune pattern recognition: a cell biological perspective. *Annu. Rev. Immunol.* **33**, 257–290 (2015).
62. O'Connor, B. P. *et al.* Regulation of toll-like receptor signaling by the SF3a mRNA splicing complex. *PLoS Genet.* **11**, e1004932 (2015).
63. Mockenhaupt, S. & Makeyev, E. V. Non-coding functions of alternative pre-mRNA splicing in development. in *Seminars in cell & developmental biology* **47**, 32–39 (Elsevier, 2015).
64. Mehta, A. & Baltimore, D. MicroRNAs as regulatory elements in immune system logic. *Nat. Rev. Immunol.* **16**, 279–294 (2016).
65. Wu, C.-T., Chiou, C.-Y., Chiu, H.-C. & Yang, U.-C. Fine-tuning of microRNA-mediated repression of mRNA by splicing-regulated and highly repressive microRNA recognition element. *BMC Genomics* **14**, 438 (2013).
66. Andolfo, I. *et al.* Regulation of divalent metal transporter 1 (DMT1) non-IRE isoform by the microRNA Let-7d in erythroid cells. *Haematologica* **95**, 1244–1252 (2010).

67. Barreau, C., Paillard, L. & Osborne, H. B. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* **33**, 7138–7150 (2005).
68. Gratacós, F. M. & Brewer, G. The role of AUF1 in regulated mRNA decay. *Wiley Interdiscip. Rev. RNA* **1**, 457–473 (2010).
69. Brooks, S. A. & Blackshear, P. J. Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta BBA- Gene Regul. Mech.* **1829**, 666–679 (2013).
70. Simone, L. E. & Keene, J. D. Mechanisms coordinating ELAV/Hu mRNA regulons. *Curr. Opin. Genet. Dev.* **23**, 35–43 (2013).
71. Hitti, E. & Khabar, K. S. Sequence variations affecting AU-rich element function and disease. *Front Biosci Landmark Ed* **17**, 1846–1860 (2012).
72. Luchin, A. I. *et al.* AU-rich elements in the 3'-UTR regulate the stability of the 141 amino acid isoform of parathyroid hormone-related protein mRNA. *Mol. Cell. Endocrinol.* **364**, 105–112 (2012).
73. Parenteau, J. *et al.* Introns are mediators of cell response to starvation. *Nature* **565**, 612–617 (2019).
74. Morgan, J. T., Fink, G. R. & Bartel, D. P. Excised linear introns regulate growth in yeast. *Nature* **565**, 606 (2019).
75. Zaffagni, M. & Kadener, S. Craving for Introns. *Mol. Cell* **73**, 1095–1096 (2019).
76. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* **26**, 163 (2000).
77. Green, R. E. *et al.* Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* **19**, i118–i121 (2003).
78. Rotival, M., Quach, H. & Quintana-Murci, L. Defining the genetic and evolutionary architecture of alternative splicing in response to infection. *Nat. Commun.* **10**, 1671 (2019).
79. De Arras, L. *et al.* Comparative genomics RNAi screen identifies Eftud2 as a novel regulator of innate immunity. *Genetics* **197**, 485–496 (2014).
80. Hsu, T. Y.-T. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, 384 (2015).

81. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232 (2015).
82. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
83. Evsyukova, I., Somarelli, J. A., Gregory, S. G. & Garcia-Blanco, M. A. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol.* **7**, 462–473 (2010).
84. Padgett, R. A. New connections between splicing and human disease. *Trends Genet.* **28**, 147–154 (2012).
85. Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **16**, 413 (2016).
86. Kahles, A. *et al.* Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224 (2018).
87. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
88. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613 (2012).
89. Thanos, D. & Maniatis, T. Virus induction of human IFN β gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).
90. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
91. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
92. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
93. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
94. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).
95. Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* **102**, 11–26 (2018).

96. Katz, Y., Wang, E. T., Airoidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
97. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.* **111**, E5593–E5601 (2014).
98. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *elife* **5**, e11752 (2016).
99. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
100. Kahles, A., Ong, C. S., Zhong, Y. & Räsch, G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics* **32**, 1840–1847 (2016).
101. Wang, Q. & Rio, D. C. JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns. *Proc. Natl. Acad. Sci.* **115**, E8181–E8190 (2018).
102. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. & Blencowe, B. J. Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol. Cell* **72**, 187–200 (2018).

*Chapter 2***A STUDY OF THE KINETICS OF mRNA SPLICING IN THE NF- κ B
TRANSCRIPTOME**

Adapted from: Majumdar, D., Frankiw, L., Burns, C. H., Garcia-Flores, Y. & Baltimore, D. Programmed Delayed Splicing: A Mechanism for Timed Inflammatory Gene Expression. *bioRxiv* 443796 (2018).

Abstract

The process of inflammation involves a coordinated gene expression program. Tight regulation of this gene expression program is crucial to ensure inflammation remains properly scaled. Here, we developed a RNA “hybrid-capture” purification technique to enrich for specific cDNAs relevant to the NF- κ B pathway, which thus allowed us to investigate the role of pre-mRNA splicing in the regulation of inflammatory gene expression. Our results support the notion that most introns in mRNA are spliced linearly and co-transcriptionally (i.e. the earlier an intron is transcribed, the earlier it will be spliced). However, a number of transcripts contain introns that are spliced at significantly slower rates than neighboring introns. In many instances, poor splicing at such junctions is attributable to evolutionarily conserved poor splice donor sequences. This has led to the idea that specific introns, so called “bottleneck” introns, can regulate the expression of genes through the timing of splicing, as well as decay of intron-containing transcripts.

Introduction

Essential to an effective inflammatory response is a tightly regulated gene expression program. Small alterations to this program can shift the balance from protective to destructive immunity¹. While transcription and protein turnover are the most well-scrutinized area of gene expression regulation²⁻⁵, there have emerged a variety of post-transcriptional mechanisms that play a role in the fine-tuning of an inflammatory gene expression program i.e. RNA stabilization⁶, RNA deadenylation⁷, and microRNA regulation⁸. We and others have recently investigated the role of RNA splicing as a means of post-transcriptional gene expression regulation⁹⁻¹³.

RNA synthesis begins with the initiation of transcription by RNA Polymerase II (Pol II) at the promoter region of a gene. Pol II synthesizes RNA as it elongates through the DNA template until transcription is terminated many kilobases downstream¹⁴. This pre-mRNA contains many intervening sequences (introns) that are excised allowing the remaining sequences (exons) to be concatenated together to form a mature eukaryotic RNA transcript¹⁵. Following the completion of transcription, pre-mRNA molecules must be cleaved from the template DNA and polyadenylated at the 3' end¹⁶. Then, full-length completely spliced transcripts can be released into the cytoplasm. In cases where a transcript retains an intron, such a transcript can remain in the nucleus and undergo delayed splicing¹⁷, or be degraded via the nuclear exosome or the NMD decay machinery in the cytoplasm¹⁸. This allows a cell to fine-tune when and how much message is produced at the post-transcriptional level. While there are exceptions, it is rare for a transcript with a retained intron to contribute to proteomic diversity¹⁹.

In order to understand the role of splicing with respect to the timing and levels of gene expression during an inflammatory response, a proper analysis of the rate of pre-mRNA processing is needed. Current methods, namely qPCR and RNA-seq, have limitations with respect to such analyses. qPCR is inherently limited by the number of junctions with which can be analyzed, and because of the relative number of transcripts involved with the inflammatory response as compared to total mRNA, RNA-seq results in limited resolution at the majority of induced junctions. To overcome this sequencing depth problem, we developed a RNA “hybrid-capture” purification technique to enrich for specific cDNAs relevant to the NF- κ B pathway, which thus allowed us to investigate the role of pre-mRNA splicing in the regulation of inflammatory gene expression. We found that the majority of splicing is co-transcriptional. However, there were junctions in important inflammatory transcripts that were significant outliers, remaining primarily unspliced throughout a stimulation time-course. We hypothesize that these outliers, which we deem “bottleneck” introns, regulate the timing and levels expression of their respective genes through the delay of splicing and decay of intron containing transcripts.

Results

Hybrid-Capture Optimization

The first step in the project involved implementing the “hybrid-capture” purification technique (Figure 1A)²⁰. This technique utilizes biotinylated 120-nucleotide sense strand probes that are able to form strong hybrids with the target, antisense cDNA. These probe-DNA complexes are captured with streptavidin-coated beads and then undergo multiple washes before the captured sequences are eluted into a basic solution. Probes were designed

that were targeted to the last exon of all of the >300 genes in the NF- κ B transcriptome previously shown to be highly induced by an LPS immune response in macrophages¹⁰. Oligo(dT) priming and the choice of the last exon as a capture target enabled us to sequence complete transcripts from the standpoint of the splicing machinery, because all introns will have been transcribed in such transcripts. To ensure efficient pull-down of targeted genes 120bp probes were created that were tiled to sequences from the last exons of transcripts using a process called microarray printing. The oligonucleotides for each fragment were synthesized via microarray printing, a T7 promoter was appended through the use of PCR, and biotinylated RNAs were made with an RNA Pol reaction from the PCR product. qPCR was used to measure efficiency of “pull-down”, which showed significant enrichment for targeted transcripts and correspondingly, depletion of the non-targeted transcript RPL32 (Figure 1B-D).

Upon optimization of the “hybrid-capture” protocol, our next goal was to optimize the library preparation protocol in order to be able to submit the samples for initial Illumina sequencing. In a typical library preparation for Illumina RNA-Seq, samples are reverse transcribed and sheared to ~200 bp. Then, samples are end-repaired and dA-tailed to allow for the ligation of a specific Illumina adaptor. These adaptors allow for further non-biased amplification via PCR and are necessary for sample bridge formation on flow cell, an essential component of Illumina’s sequencing by synthesis (SBS) approach. Despite the optimization we performed on the hybrid-capture, the technique is only able to effectively “pull-down” a limited number of transcripts. This presents a number of issues with respect to sequencing library preparation as a typical Illumina library preparation protocol requires >1ng of input DNA. As we capture much less than 1ng, we needed to develop a way to

amplify these “pulled-down” transcripts in a manner that introduces no length or sequence biases. The molecular biology to do this successfully involves investigation and feasibility studies of several methods of Rapid Amplification of cDNA Ends (RACE).

Our initial approach utilized T4 RNA ligase to perform a single-stranded ligation that adds a specialized adaptor sequence onto samples. As the samples had been reverse transcribed with a special 5' poly-A primer that contained a unique flanking handle, the ligation of the second handle to the 3' end of the samples would allow for non-biased whole sample PCR enrichment. Two PCR primers specific for the 5' and 3' handles could be designed to amplify the entire sample. At the extremely low concentration of DNA that we were working with, this ligation proved to be problematic. Our analysis showed that there was only a 0.01% alignment between our reads and the reference genome. Further inspection of the sequencing reads showed a plethora of reads that contained the sequence of the adaptor that was used in the T4 RNA ligase reaction. We concluded that the extremely low concentration of the sample hindered the efficiency of the ligation reaction, in turn preventing us from amplifying our sample.

Next, we tried to use terminal deoxynucleotidyl transferase (TdT) to add the 3' handle to the captured DNA sample. TdT is a template independent polymerase that is able to add specific deoxynucleotides to the 3' end of DNA molecules²¹. TdT plays an important role in V(D)J recombination as it is able to add N-nucleotides during antibody gene recombination, allowing for junctional diversity²¹. We attempted to first add a polycytosine handle by adding TdT and a specific concentration of deoxycytidine triphosphate

(dCTP). At 2.5 uM, a ~5-25 cytosine handle was shown to be added. However, the subsequent PCR amplification reaction involving the complementary poly-guanine PCR primer proved to be extremely inefficient. Attempts made using different nucleotide base handles and PCR primer combinations had similar amplification issues.

Finally, our last attempt at RACE exploited the ability of MMLV reverse transcriptase to “template switch”. Prior to the hybrid capture, the extracted RNA needs to be reverse transcribed to cDNA. Because the MMLV reverse transcriptase enzyme has terminal transferase activity²², when it reaches the 5’ end of an RNA molecule, it adds a few additional nucleotides (primarily deoxycytidine). A special “template-switch” oligonucleotide that contains a poly(G) sequence at its 3’ end is simultaneously added to the reverse transcription reaction. The 3’ end of this oligonucleotide base pairs with the deoxycytidine stretch that the MMLV reverse transcriptase added to the synthesized cDNA. Reverse transcriptase then is able to switch template, adding the complementary sequence to the “template switch” oligonucleotide to the end of the synthesized cDNA molecule. This “template switch” method of RACE provides the second handle needed for the non-biased amplification. We tested the method by performing the “template switch” reverse transcription on RNA from LPS stimulated bone-marrow derived macrophages (BMDMs). Initial libraries showed read alignments >70% (data not shown). Comparison of the “hybrid-capture” derived sequencing data to data from LPS stimulated BMDMs that had not been captured¹⁰ highlights the effectiveness of the capture. Counting reads that map to NF-κB transcriptome (i.e. reads of interesting transcripts induced by LPS stimulation), ~70% of the reads from the capture data set map to the transcriptome as compared to ~2% of the reads

from the non-capture data-set (Figure 2A-C). This makes our protocol more cost-efficient, as greater depth can be achieved for transcripts of interest without an increase in the total number of reads sequenced.

Splicing Analysis

Following optimization, chromatin-associated captured transcripts from BMDMs induced with TNF α were sequenced. To quantify splicing, we modified what is known as the Completed Splicing Index (CoSI) that was originally published as a part of the 2012 ENCODE project²³. The CoSI metric from the ENCODE project offers insight into the splicing of introns around a single exon (i.e. the extent of splicing of the introns upstream and downstream of a given exon). As we wanted to narrow in on intron specific splicing, we modified this CoSI value so that it would only provide us with a ratio for the extent of splicing around a single intron (Figure 3A). CoSI values of 1 and 0 imply near-complete splicing and virtually unspliced states, respectively. Using the CoSI metric, quick induction of I κ B α can be seen by the significant drop in the CoSI values within the first 5 minutes of TNF α stimulation (Figure 3B). This drop corresponds to active transcription of I κ B α . Further, this CoSI data again supports the conclusion that splicing occurs primarily co-transcriptionally. The closer an intron is to the 3' end of the transcript, the lower its CoSI value and thus, the less likely it is to be completed splicing upon completion of transcription. Strikingly, the final intron deviates significantly in its kinetic trajectory, as its read density does not obey a similar relative reduction. This might be due to a lag in terminal intron splicing²⁴ or a feature of splicing that accompanies transcript release from chromatin. A global analysis of CoSI data using a single time point (12 minutes) from the TNF α stimulation time course showed

this statement more conclusively. Introns were grouped based on their location to the 3' end of their respective transcripts. From these groups, we calculated the average CoSI values. As shown in Figure 3C, terminal (3') introns have lower CoSI values on average than 5' introns. This agrees with a recently published co-transcriptional splicing model developed by the Hoffmann lab at UCLA²⁵. Hoffmann models splicing as a series of sequential reactions where the time of each reaction is an independent, exponentially distributed random variable (with an associated rate constant). In turn, the probability that an intron has been spliced at a certain time is dependent only on the time following synthesis of the intron and the rate at which the intron is spliced. Assuming similarity in the rates of splicing between introns, an assumption which will later be shown to not always be valid, this model predicts terminal introns will have much lower co-transcriptional splicing efficiencies as they are synthesized last and thus, there is less time between synthesis and completion of transcription / polyadenylation.

We next used the CoSI metric to study splicing on a global scale across all time points. Figure 4A represents the CoSI values in Tukey boxplot format for all introns in the NF- κ B transcriptome at different TNF α stimulation time points. The majority of introns splice relatively quickly and consistently. Their CoSI values dip with the initiation of transcription following stimulation; however, the introns transcribed during this induction phase are spliced out and the CoSI values return to their non-stimulated levels in well under 60 minutes. In saying this, there are numerous splicing outliers that reveal delayed inflammatory introns. In particular, the fourth intron of IRF7, shown in Figure 4B, epitomizes a delayed intron that is spliced at a much slower rate than would be expected and can be seen as an outlier on the plot. IRF7 is one of many genes among our list of NF- κ B

induced genes that have an intron that splice with notably slow kinetics. Other inflammatory genes with a conspicuously and consistently unspliced intron include: CD40 (Figure 5A), DAXX (Figure 5B), A20, CXCL2, BAX, *I κ B ϵ* , IL27, DUSP2, MX1, XBP1, and CD79a. We hypothesize that the timing of expression of certain of these genes may represent a biological mechanism delaying or limiting their expression and that regulation of intron splicing may be an important avenue in understanding the processes limiting inflammation. Thus, we have given them the name “bottleneck” introns as they presumably prevent the completion of the full-length spliced transcript well past the completion of transcription. It is interesting to note that many of these “bottleneck introns” contain evolutionarily conserved splice donors or acceptors that differ from the consensus donor/acceptor sequences. Figure 6 depicts the case for the poorly spliced fourth intron of IL27, which has a splice donor that both differs from the consensus sequence and is conserved across a variety of species. We hypothesize that this non-consensus, “poor” splice donor plays a role in the regulation of expression of the IL27 gene, thus the reason it is evolutionarily conserved.

Next, we measured the rates of splicing across the transcriptome. As ongoing transcription complicates the process of quantitatively analyzing splicing rates and intron half-lives, we decided to again stimulate BMDMs with TNF α but halt transcription using actinomycin D (ActD) to get at meaningful quantification of the splicing pattern. ActD is able to stop transcription by binding DNA at the transcription initiation complex and preventing the elongation of the RNA chain by RNA polymerase²⁶. Once transcription is stopped, it becomes feasible to accurately measure the rate of intron disappearance. BMDMs were first stimulated with TNF α and then, 8 minutes after stimulation, were treated with ActD. Whole cell RNA was enriched for mRNAs from the NF- κ B transcriptome and

sequenced. Whole cell RNA was used as opposed to chromatin-fractionated RNA to prevent biasing for introns in transcripts and thus, artificially increasing their half-lives. Figure 7A shows the IκBα sequencing reads for our TNFα / ActD time course. Following the addition of ActD, intron levels rapidly decreased indicating splicing without simultaneous transcription. To quantify intron half-lives, we performed an exponential fit of the CoSI data following the addition of the ActD (Figure 7B, C). Then, by using the fit to determine the time it takes for half the intron to disappear (in other words, the time it takes for the CoSI value to get to the midpoint of its maximally unspliced and maximally spliced levels), we could quantify the rate with which each intron was being spliced. The majority of introns had relatively small (<250 seconds) half-lives (Figure 7B), indicating the efficiency with which they were being spliced. However, there was a great deal of heterogeneity in these half-life values. The distribution of half-lives for introns in the NF-κB transcriptome ranged from 30s to 10 minutes.

Finally, we looked at the correlation between splicing rates and transcriptional gene expression groups. Previous work by the Smale group at UCLA used RNA-Seq transcript expression levels to group genes in the NF-κB transcriptome as either early, intermediate, or late induced (Figure 8). With our TNFα stimulation time course data, we used Tukey boxplots to globally represent the CoSI values for the introns within each of these groupings (Figure 8A). As seen in Figure 8A, we are able to show that early genes tend to splice quicker than intermediate genes, which tend to splice quicker than late genes. Although there are outliers, the mean and the 25th percentile or the lower edge of the IQR box shifts lower the later the expression of the grouping. This result naturally fits with a hypothesis whereby the

rate at which introns of given transcript splice have the ability to effectively regulate the timing of the expression of a given gene.

Discussion

In this study, we sought to investigate the role of pre-mRNA splicing in the regulation of inflammatory gene expression. We developed a “hybrid-capture” method which allowed us to enrich for transcripts in the NF- κ B transcriptome. We found that the majority of junctions splice efficiently. Further, the closer an intron is to the 3' end of the transcript, the lower its CoSI value and thus, the less likely it is to be completed splicing upon completion of transcription, agreeing with a largely co-transcriptional model of splicing^{23,27,28}. In saying this, we found considerable heterogeneity in splicing efficiency among these introns, with a large number of splicing “outliers” which remain significantly unspliced throughout a stimulation time-course. One explanation for such outliers comes from analysis of splice donor sequences. A number of poorly spliced introns contained evolutionarily conserved weak 5' splice donors. However, there were a significant number of introns that did not contain such sequences, which suggests splice donor strength is just one of the regulatory mechanisms responsible for poor splicing. It is likely factors like intron GC content, size, branch point strength, and 3' splice site strength, which have been shown to correlate with poor splicing in other work²⁹, contribute to the observed poor splicing efficiency of given junctions. To determine more specifically the rates of slowly splicing introns, we used ActD to stall transcription and monitor intron half-lives. Supporting the idea of heterogeneous rates of splicing, while most introns spliced within 20-40s, some were delayed significantly and took >5 minutes to splice.

The fact that these inefficiently spliced introns occur in important inflammatory genes, some with the seemingly negative characteristic of having an evolutionarily conserved poor splice donor sequence, begs the question as to why they exist? We hypothesize that these junctions act to either slow-down, or limit gene expression. Again, at the heart of an inflammatory response is a tightly regulated gene expression program. Regulation of this gene expression program is crucial, as small changes can shift the balance away from protective immunity toward either nonexistent or destructive immunity. Regulation at the post-transcriptional level through splicing can allow a cell the ability to quickly fine-tune when and how much of a given gene product is produced without altering the transcriptional landscape of a cell. With respect to the transcripts containing poorly spliced introns, it stands to reason the weak introns acts to dampen protein output, perhaps as a means to mitigate what otherwise would be an unchecked or inappropriately scaled response. As splicing inherently involves a variety of *trans*-acting factors, future work aimed at identifying such factors that potentially regulate such splicing events will be important to our understanding of this form of post-transcriptional regulation.

In conclusion, the hybrid capture approach provides a large number of junctional sequencing reads, which permitted unique insight into the efficiency and kinetics of splicing of mature transcripts, and revealed surprising heterogeneity. We suggest that this methodology and analysis could have wider applicability for other gene induction situations.

Materials and Methods

Animals

The California Institute of Technology Institutional Animal Care and Use Committee approved all experiments. C57BL/6 WT mice were bred and housed in the Caltech Office of Laboratory Animal Resources (OLAR) facility. C56BL6/J mice were sacrificed via CO₂ euthanasia and sterilized with 70% ethanol. Femur and tibia bones harvested and stripped of muscle tissue. Bone marrow cells were resuspended in 20mL of fresh DMEM. 2.5 *10⁶ bone-marrow cells plated in a 150mm dish in 20mL of BMDM Media (DMEM, 20% FBS, 30% L929 condition media, and 1% Pen/Strep) and grown at 5% CO₂ and 37°C. BMDM media was completely replaced on day 3 as well as a supplemental addition of 5mL L929 condition media on day 5.

Cell Culture

Human embryonic kidney cells (HEK293T) from ATCC were cultured in DMEM supplemented with 10% FBS and 1% Pen/Strep. Cell lines were maintained at 37°C in 5% CO₂.

RNA Fractionation

Confluent 10-cm dish of mature BMDMs were scraped into 400mL cold NP-40 lysis buffer, APJ1 (10mM Tris-HCl (pH 7.5), 0.08% NP-40, 150mM NaCl). Lysed cells layered onto 1mL cold sucrose 322 cushion, APJ2 (10mM Tris-HCl (pH 7.5), 150mM NaCl, 24% w/v sucrose) and centrifuged for 10min at 4°C and 13000 rpm. The supernatant from this spin

represents the cytoplasmic RNA fraction, which is immediately added to 3 volumes of 100% ethanol and 2 volumes of buffer RLT (4M GuSCN, 325 0.1M β -mercaptoethanol, 0.5% N-lauroyl sarcosine, 25mM Na-citrate, pH7.2) and stored at 80°C until ready to purify RNA. Pellet, containing intact nuclei, is resuspended in 500mL TRIzol reagent. If the pellet was difficult to dissolve, it was heated at 50°C with occasional vortexing. 100mL chloroform added and shaken vigorously for 15-20 s; allowed to phase separate at room temperature for 5min. Tube centrifuged at 4°C and 12000 x g for 15min. Clear upper aqueous phase removed to a new tube, ensuring white DNA mid-phase is not removed, and is immediately added to 3 volumes of 100% ethanol and 2 volumes of buffer RLT and stored at 80°C until ready to purify RNA. RNA is purified according to QIAGEN RNeasy column protocol and eluted in 30mL nuclease-free H₂O. RNA samples are DNase treated with Turbo-DNase and stored at 80°C.

cDNA Pulldown

cDNA is added to biotinylated RNA probes, generated by Ampliscribe T7-Flash Biotin Kit (Epicentre), and incubated at 74°C for 4.5min to denature followed by addition of 1 volume of 2X hybridization (HYB) buffer (1M LiCl, 40mM Tris-HCl (pH 7.5), 20mM EDTA (pH 8.0), 4M Urea, 0.5% Triton X-100, 1% SDS, 0.2% Na-deoxycholate). Reaction incubated at 70°C for 30min. 0.3mg BioMag streptavidin beads (Bang Laboratories Inc.), washed 3 times in 1X HYB buffer, added and reaction incubated at 70°C and 1100rpm for 20min to capture cDNA-probe complex. Beads pelleted on magnet, followed by 2 washes of 150 μ L with preheated 1X HYB at 70°C, 1 wash of 150 μ L with wash #4 (160mM LiCl, 20mM Tris-HCl (pH 7.5), 10mM EDTA (pH 8.0), 2M Urea, 0.25% Triton X-100, 0.5% SDS, 0.1% Na-

deoxycholate), and 1 final wash with wash #5 (40mM LiCl, 20mM Tris-HCl (pH 7.5), 10mM EDTA (pH 8.0), 2M Urea, 0.25% Triton X-100, 0.5% SDS, 0.1% Na-deoxycholate). Beads resuspended in 35 μ L of base elution buffer (125mM NaOH, 10mM EDTA (pH 8.0), 10mM Tris-HCl (pH 7.5) and incubated at 74°C and 1100rpm for 5min. Beads pelleted and 30 μ L cDNA containing supernatant removed to a new tube. Solution neutralized with 6.25 μ L neutralization buffer (800mM HCl, 160mM Tris-HCl (pH 7.5), 20mM EDTA (pH 8.0)). Immediately after neutralization, cDNA purified by 1.0X Sera-Mag treatment as described previously above and eluted in 45 μ L and stored at -80°C.

RNA-Sequencing Analysis

Sequencing was performed on a HiSeq 2500 High Throughput Sequencer (Illumina). Single-end 50-mer reads were aligned using Tophat v2.1.1³⁰. Gene expression was determined using Cufflinks v2.2.1 and the FPKM (Fragments Per Kilobase Million) metric³¹.

Figure Legends

Figure 1. (A) Schematic of RAP. Biotinylated nucleotide probes are first hybridized to cDNA. This probe-cDNA complex is then bound to streptavidin magnetic beads, which allows for non-hybridized sequences to be washed away. The “captured” cDNA is then eluted from the beads and ultimately sequenced. (B) A comparison of the total transcripts pulled down by the A20, I κ B ϵ , TNF α , and I κ B α probes as compared to the non-targeted RPL32. (C) Fold depletion of transcripts following “hybrid-capture” in log scale. (D) As in (C) but in linear scale.

Figure 2. (A) Comparison of reads that map to NF- κ B transcriptome with hybrid capture or without hybrid capture. (B) Number of reads that map to the NF- κ B transcriptome (left) and number of reads sequenced (right). (C) Percent of reads that map to the NF- κ B transcriptome with hybrid capture (red) or without hybrid capture (blue).

Figure 3. (A) CoSI metric is a ratio comparing the number of reads that map across a splice junction to the number of reads that contain partial intron and exon sequences. A CoSI value of 1 indicates that a junction is completely spliced, whereas a CoSI value of 0 indicates that no splicing has occurred. (B) A plot depicting the change in CoSI for chromatin-associated RNA at different timepoints of TNF α stimulation. (C) The RefSeq annotation and the sequencing reads for the 20 minute timepoint are shown below the CoSI plot. (D) Plot depicting CoSI with respect to intron position.

Figure 4. (A) Tukey boxplot representing CoSI values at different TNF α stimulation time points. The top of the ‘box’ represents the 75th percentile value; the bottom of the ‘box’ represents the 25th percentile value. The space between represents the interquartile difference (IQR) and the line inside the box represents the mean CoSI value. The ‘whisker’ is calculated by subtracting 1.5*IQR from the 25th percentile. Values lower than the whisker are classified as ‘outliers’. **(B)** Histogram of RNA-seq reads for IRF7 throughout a TNF α stimulation time-course. The retained fourth intron is highlighted in yellow.

Figure 5. Histogram of RNA-seq reads for CD40 **(A)** and DAXX **(B)** throughout a TNF α stimulation time-course. The retained intron is highlighted in yellow.

Figure 6. (A) A depiction of the sequencing reads that map to the IL27 gene at 40 minutes of TNF α stimulation. **(B)** Consensus sequence for splice donors. The first four bases are usually gtaag or gtgag. **(C)** The mouse NCBI37/mm9 RefSeq annotation for IL27. The fourth intron has the sequence gtaga, which differs from the consensus sequence. **(D)** A comparison of the splice donor sequence for the fourth intron of IL27 shows the sequence is conserved across a variety of species.

Figure 7. (A) Histogram of RNA-seq reads showing induction of I κ B α in response to TNF α and ActD. Time at which TNF α and ActD are added is depicted. **(B)** Histogram of intron half-lives. **(C)** Scatter plot representing the intron half-life on the x-axis and the minimum CoSI value during the time-course on the y-axis.

Figure 8. (A) Tukey boxplot representing CoSI values at different TNF α stimulation time points for early, intermediate, and late induced genes. (B) Expression data from the Smale group depicting the timing of expression for genes classified as early, intermediate, or late.

References

1. Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F. & Kollias, G. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398 (1999).
2. Gautier, E. L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118 (2012).
3. Chen, J. & Chen, Z. J. Regulation of NF- κ B by ubiquitination. *Curr. Opin. Immunol.* **25**, 4–12 (2013).
4. Smale, S. T. & Natoli, G. Transcriptional control of inflammatory responses. *Cold Spring Harb. Perspect. Biol.* **6**, a016261 (2014).
5. Smale, S. T., Tarakhovsky, A. & Natoli, G. Chromatin contributions to the regulation of innate immunity. *Annu. Rev. Immunol.* **32**, 489–511 (2014).
6. Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* **10**, 281 (2009).
7. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).
8. O’Connell, R. M., Rao, D. S. & Baltimore, D. microRNA regulation of inflammatory responses. *Annu. Rev. Immunol.* **30**, 295–312 (2012).
9. Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**, 436 (2011).
10. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
11. Hao, S. & Baltimore, D. RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci.* **110**, 11934–11939 (2013).
12. Pandya-Jones, A. *et al.* Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *Rna* **19**, 811–827 (2013).
13. Rabani, M. *et al.* High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014).
14. Selth, L. A., Sigurdsson, S. & Svejstrup, J. Q. Transcript elongation by RNA polymerase II. *Annu. Rev. Biochem.* **79**, 271–293 (2010).

15. Darnell, J. E. Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *Rna* **19**, 443–460 (2013).
16. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163 (2014).
17. Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80 (2015).
18. Sayani, S. & Chanfreau, G. F. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in *Saccharomyces cerevisiae*. *Rna* **18**, 1563–1572 (2012).
19. Schmitz, U. *et al.* Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* **18**, 216 (2017).
20. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
21. Chang, L. M., Bollum, F. J. & Gallo, R. C. Molecular Biology of Terminal Transferase. *Crit. Rev. Biochem.* **21**, 27–52 (1986).
22. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: A SMARTTM approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
23. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
24. Oesterreich, F. C., Preibisch, S. & Neugebauer, K. M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* **40**, 571–581 (2010).
25. Davis-Turak, J. C. *et al.* Considering the kinetics of mRNA synthesis in the analysis of the genome and epigenome reveals determinants of co-transcriptional splicing. *Nucleic Acids Res.* **43**, 699–707 (2014).
26. Sobell, H. M. Actinomycin and DNA transcription. *Proc. Natl. Acad. Sci.* **82**, 5328–5331 (1985).
27. Ameer, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18**, 1435 (2011).
28. Brugiolo, M., Herzel, L. & Neugebauer, K. M. Counting on co-transcriptional splicing. *F1000prime Rep.* **5**, (2013).

29. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
30. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
31. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

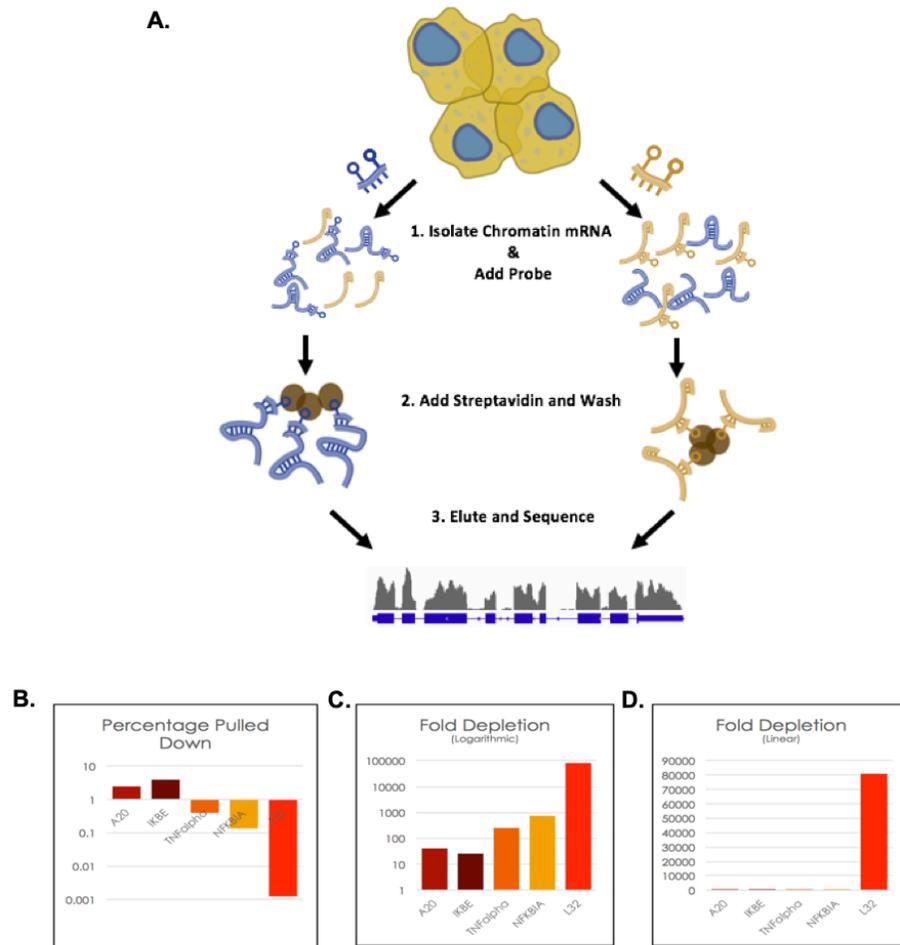


Figure 1

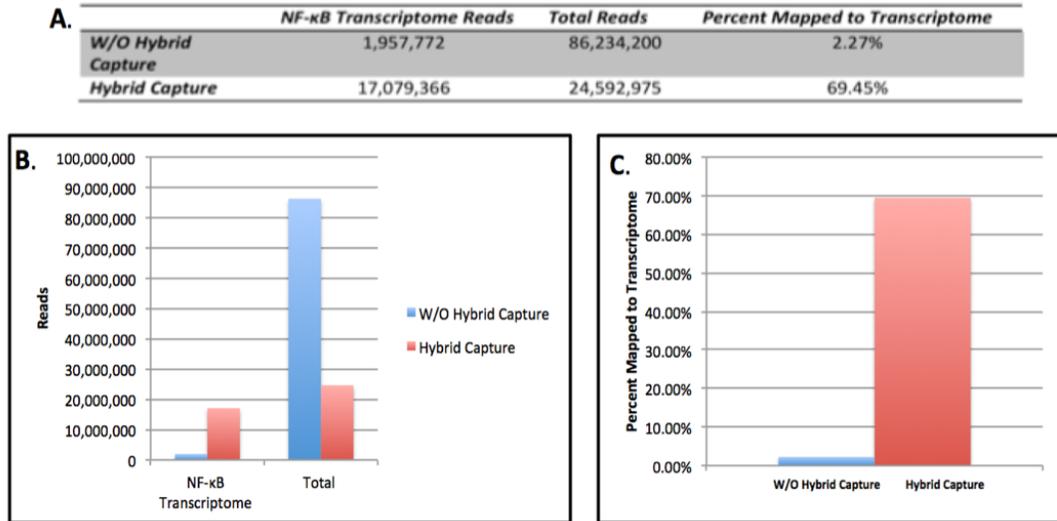
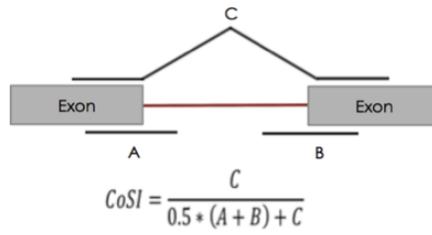
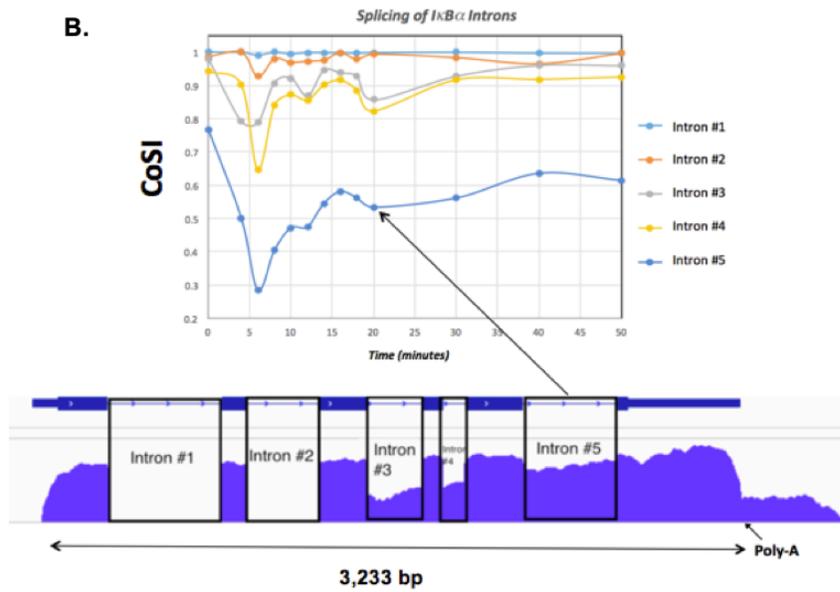


Figure 2

A.



B.



D.

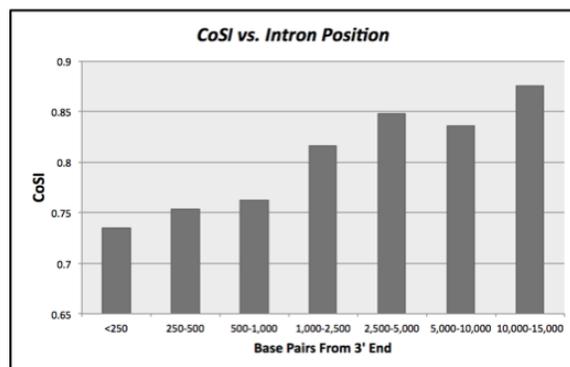


Figure 3

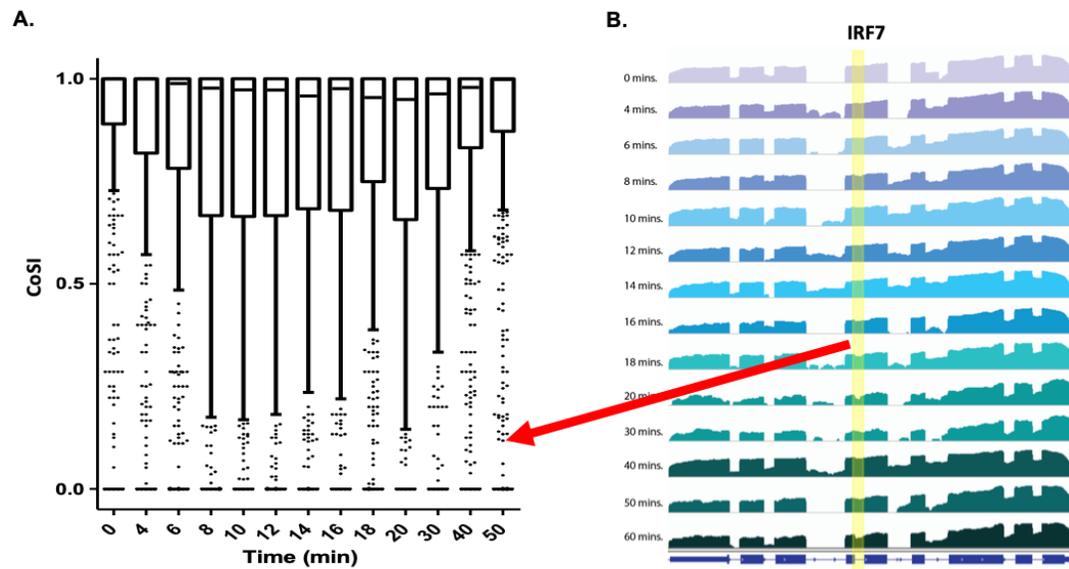


Figure 4

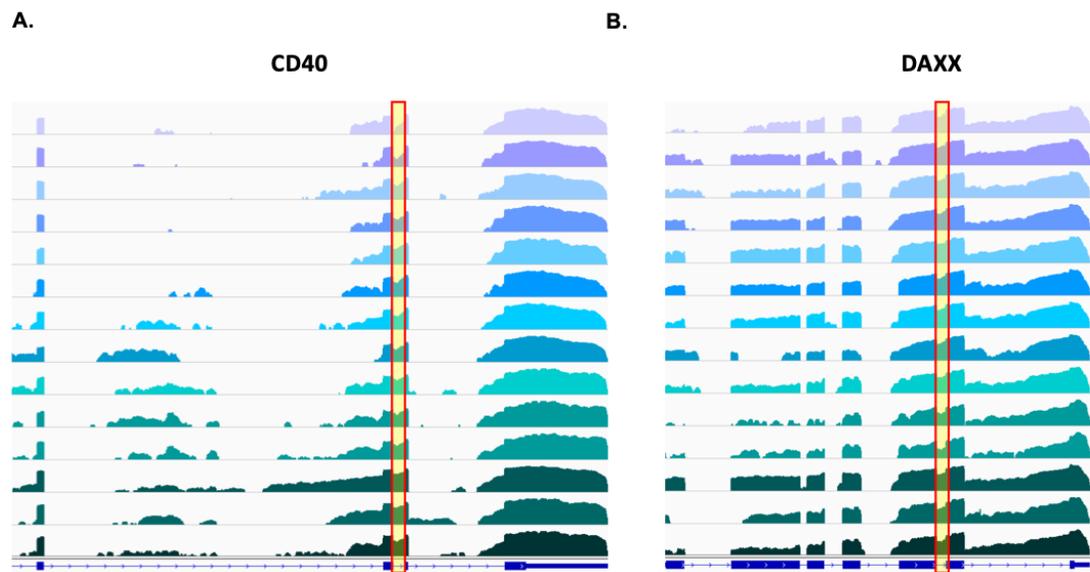


Figure 5

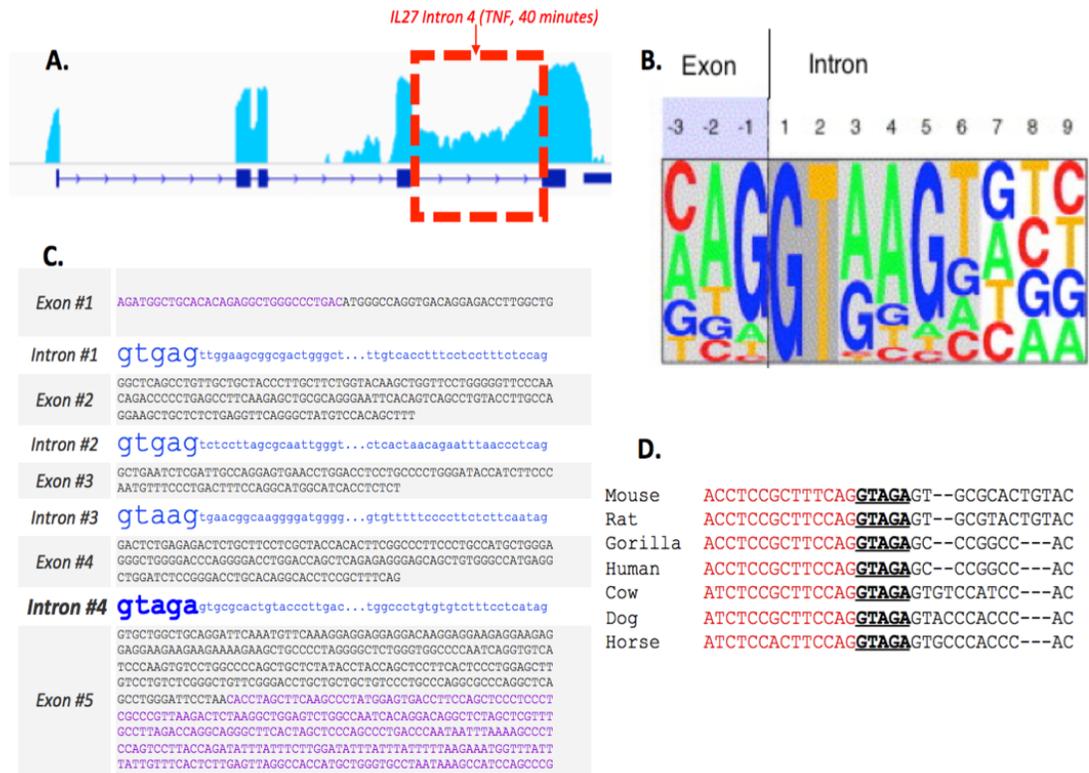


Figure 6

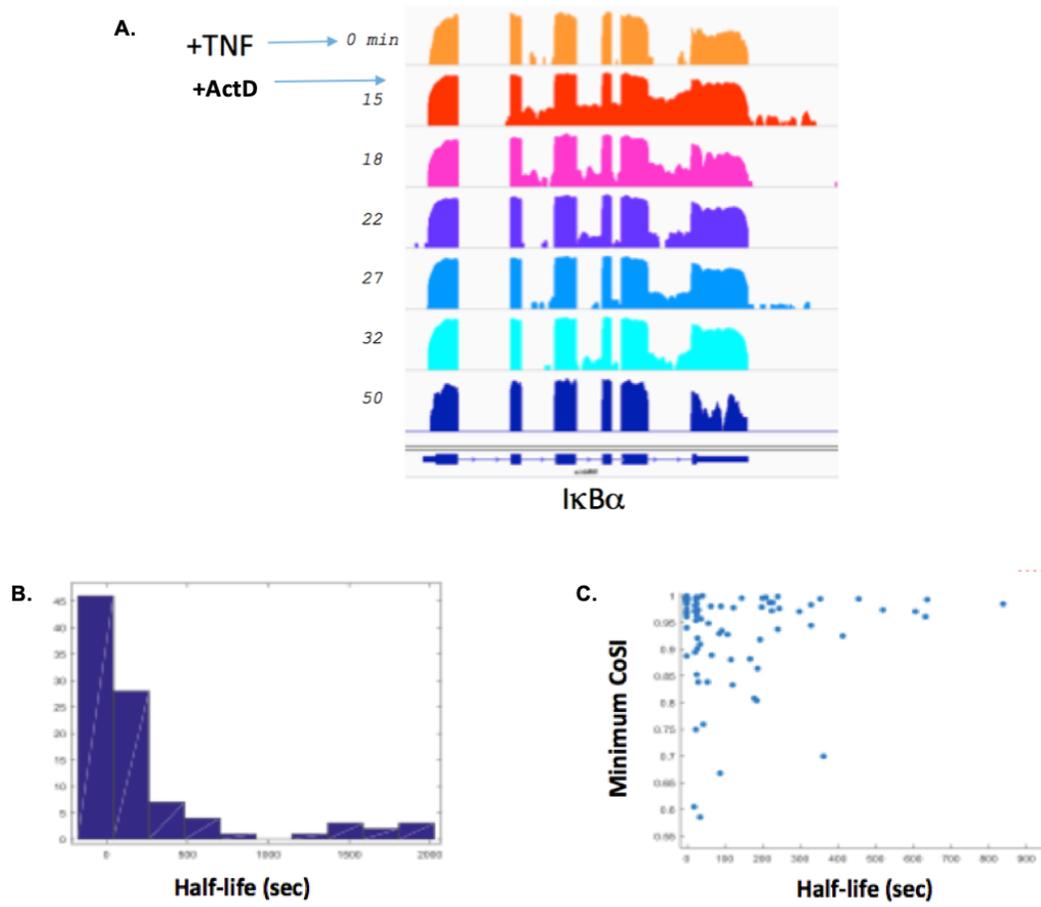


Figure 7

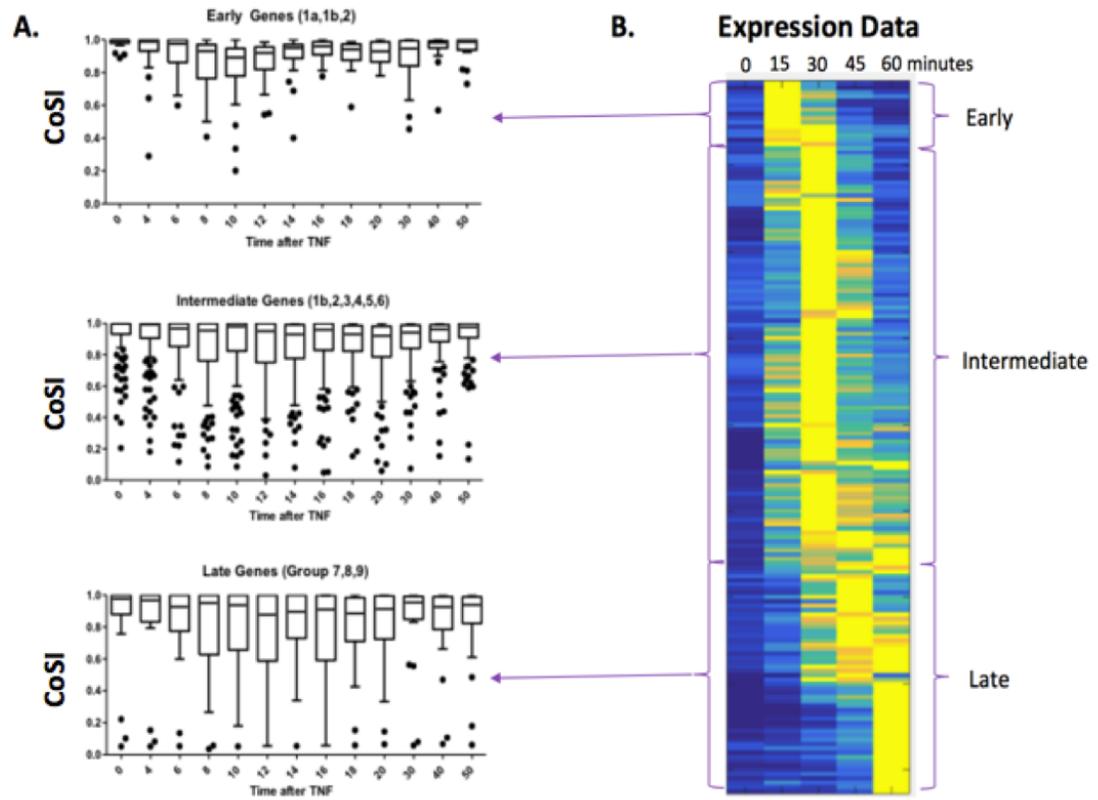


Figure 8

*Chapter 3***BUD13 Promotes a Type I Interferon Response by Countering Intron Retention in *Irf7***

Published as: Frankiw, L. *et al.* Bud13 Promotes a Type I Interferon Response by Countering Intron Retention in *Irf7*. *Molecular cell* **73**, 803–814 (2019).

Abstract

Intron retention (IR) has emerged as an important mechanism of gene expression control, but the factors controlling IR events remain poorly understood. We observed consistent IR in one intron of the *Irf7* gene and identified BUD13 as an RNA-binding protein that acts at this intron to increase the amount of successful splicing. Deficiency in BUD13 was associated with increased IR, decreased mature *Irf7* transcript and protein levels, and consequently a dampened type I interferon response which compromised the ability of BUD13-deficient macrophages to withstand VSV infection. Global analysis of BUD13 knockdown and BUD13 cross-linking to RNA revealed a subset of introns that share many characteristics with the one found in *Irf7* and are spliced in a BUD13-dependent manner. Deficiency of BUD13 led to decreased mature transcript from genes containing such introns. Thus, by acting as an antagonist to IR, BUD13 facilitates the expression of genes at which IR occurs.

INTRODUCTION

Three forms of alternative processing of a pre-mRNA have been described: differential inclusion of an exon, alternative splice site selection, and intron retention (IR). The latter, IR, has emerged as a previously underappreciated mechanism of post-transcriptional gene regulation. Unlike the two alternative splicing events, IR rarely contributes to proteomic diversity (Schmitz et al., 2017). However, IR events have the ability to act as negative regulators of gene expression by: (1) delaying onset of gene expression by slowing down splicing kinetics (Hao and Baltimore, 2013), (2) increasing potential nuclear degradation by nuclear exosomes, (3) increasing potential cytoplasmic degradation by nonsense mediated decay (Wong et al., 2016).

Recent genomic studies suggest IR plays an important role in the regulation of gene expression in a wide range of processes including cellular differentiation (Wong et al., 2013; Yap et al., 2012) and tumorigenesis (Dvinge and Bradley, 2015). Further, widespread IR throughout mouse and human cell and tissue types has led to the idea that IR events act to functionally “tune” the transcriptome of a cell (Braunschweig et al., 2014). However, with few exceptions, the factors that control IR events and thus potentially shape gene expression programs of cells, remain poorly understood.

Irf7 is an interferon-inducible master regulator of the type-I interferon-dependent immune response and is crucial to the production of interferon- α and β (Honda et al., 2005). Aberrant IRF7 production is linked to a wide range of pathologies, from life-threatening influenza (Ciancanelli et al., 2015) to autoimmunity (Harley et al., 2008), because precise regulation of *Irf7* ensures a proper immune response. Notably, intron 4 of *Irf7* is short, GC-

rich, and has a poor splice donor sequence, characteristics shared by many poorly spliced introns. We and others have previously shown that intron 4 of *Irf7* splices inefficiently (Shalek et al., 2013), affecting gene expression and opening a new line of inquiry as to the mechanism of IR regulation in *Irf7*.

Using RNA antisense purification-mass spectrometry (RAP-MS) (McHugh et al., 2015), we identified the protein BUD13 as one that regulates IR in *Irf7*. BUD13 was found to aid splicing efficiency and expression of the *Irf7* mature transcript and protein, thus promoting the downstream type-I interferon-dependent immune response. We show that IRF7 is able to trigger a robust interferon response in the presence but not in the absence of BUD13. Further, BUD13 was found to increase the splicing efficiency of a multitude of other junctions with similar characteristics to the one found in *Irf7*. By aiding in splicing efficiency, BUD13 limits intron retention and increases gene expression levels of transcripts containing BUD13 dependent junctions.

RESULTS

***Irf7* contains an intron that splices poorly following stimulation.**

To study the role of mRNA splicing during an innate immune response, we sequenced the RNA from mouse bone marrow-derived macrophages (BMDMs) stimulated with TNF α . From this sequencing, we identified an increased number of reads in the fourth intron of the most abundant transcript of *Irf7* as compared to other introns in the transcript (Figure 1A). A variety of features of this intron make it a likely candidate for retention (Braunschweig et al., 2014). It is extremely small at 69 nucleotides and has a high G/C content in both the flanking exons and within the intron itself (Figures 1B-E). Furthermore, the intron contains a ‘weak’

5' splice site, one that deviates from a consensus splice site sequence. This is quantified using a maximum entropy model to calculate the splice site quality score (Figure 1F) (Yeo and Burge, 2004). This increase in the number of intron reads at the fourth intron as compared to neighboring introns is also seen upon induction of *Irf7* with Poly(I:C) and IFN α (Figures 1G, H).

To quantify the extent of retention across RNA-seq data-sets, we use a metric we designate the “splicing ratio” (SR) (Figure 1I; see methods), which is a length normalized ratio of intronic reads to total reads at each junction. Low SR values indicate a junction is primarily spliced, whereas high SR values indicate a junction is primarily unspliced. Of note, junction reads that map across intron-exon boundaries are called unspliced as they can only be derived from a transcript that has yet to splice at the given junction. This can lead to some discrepancy when comparing SR with the histogram of RNA-seq reads as reads that map only a few base pairs into the intron do not easily appear unspliced but come from unspliced transcripts and count exclusively as unspliced in the SR value calculation (see methods). Using this metric, we quantified the extent of retention for all junctions in the most abundant *Irf7* transcript. We observed that for all types of stimulation, the retention of the fourth intron of the transcript is much greater than that seen for any of the other introns (Figures 1J and S1A, B). This intron remains poorly spliced despite the fact that there is clear excision of neighboring introns. It is worth noting that quantitation of the IFN α stimulation shows increased intronic signal throughout the *Irf7* transcript. This increased intronic signal is due to faster and stronger induction of *Irf7* via stimulation with IFN α and as such, an increase in the amount of pre-mRNA at a given stimulation time-point. Despite this increase in intronic signal throughout the transcript, we observed a corresponding increase in the level of

retention for the poorly spliced fourth intron (Figures 1J, S1B). Thus, we conclude this intron of *Irf7* splices poorly following many forms of stimulation.

RAP-MS identifies BUD13 as an RNA binding protein that interacts with *Irf7* mRNA.

To understand how cells handle a retained intron, we sought to identify trans-acting proteins that might affect the process using RNA Antisense Purification followed by Mass Spectrometry (RAP-MS) (Figure 2A) (McHugh et al., 2015). RAP-MS employs antisense biotin-containing ssDNAs complementary to *Irf7* exons to purify the proteins associated with the total pool of *Irf7* transcripts, containing both nascent pre-mRNAs and mature mRNA. Using this proteomic approach, we identified the RNA-binding protein BUD13 to be highly enriched (~6-fold) on *Irf7* transcripts as compared to β -*actin* transcripts, which were used as a control (Figure 2B, Table S2). Bud13 has been characterized in yeast as a member of the Retention and Splicing complex (RES) (Dziembowski et al., 2004), forming a trimeric complex with Pml1p and Snu17p, and aids in the splicing and nuclear retention of a subset of transcripts. It is not well characterized in mammalian systems. We captured a variety of other known RNA-binding proteins (PUM2, PRPF40a, SON); however, no other protein was enriched greater than two fold on *Irf7* transcripts. We observed specificity in the RNA antisense purification for the intended transcripts (Figure 2C).

Following RAP-MS, we confirmed BUD13 enrichment on *Irf7* transcripts by performing RNA Immunoprecipitation (RIP) followed by qPCR. Using formaldehyde cross-linked, BMDMs stimulated with TNF α for 30 minutes or Poly(I:C) for 12 hours, we observed >7-fold enrichment of *Irf7* transcripts associated with Bud13 immunoprecipitates as compared to Rabbit IgG control immunoprecipitates (Figures 2D, E). Of note, despite using

two different stimuli, we found similar levels of enrichment. In contrast, no differential enrichment of *Rpl32* was observed. Thus, isolating the proteins associated with *Irf7* mRNA transcripts led to the identification of BUD13, and immunoprecipitation of BUD13 protein confirmed enrichment of *Irf7* mRNA.

Bud13 knockdown leads to increased retention in the weak *Irf7* intron.

To determine whether the enrichment of Bud13 had an effect on *Irf7* mRNA processing, we used an shRNA approach to knockdown Bud13 protein levels in BMDMs (Figures S2A, B). To quantify differences in splicing between the shBud13 sample and the scrambled control sample, we calculated the difference in the previously mentioned splicing ratio (SR) metric between shBud13 and control for each junction at each time point. This resulting value was designated Δ SR. A positive Δ SR indicates a junction is more unspliced in the shBud13 sample while a negative Δ SR indicates a junction is more unspliced in the control sample. RNA-seq was performed on RNA from unstimulated BMDMs, as well as macrophages stimulated with TNF α for 0, 30, 60, and 120 minutes. Bud13 knockdown led to a further increased retention of the fourth intron in *Irf7* (Figure 3A – highlighted intron, S2C). Further, the sequencing coverage plots showed little variation in splicing for the other seven introns in the transcript. This was confirmed when splicing was quantified using the Δ SR metric (Figure 3B). At all stimulation time-points, the Δ SR value for the fourth intron was significantly greater than zero, indicating an increase in retention when BUD13 levels were reduced. There is a significant difference in the Δ SR of intron 4 as compared to every other junction in the *Irf7* transcript ($p < 0.001$, Student's t-test). All other pairwise comparisons are insignificant. This splicing difference at the fourth intron was confirmed via RT-PCR

(Figure 3C). It appears that BUD13 plays a specific role of aiding in the excision of the poorly spliced junction but is not required for total splicing of other introns in the transcript, at least as indicated by the partial knockdown with an shRNA. We next looked at how this retention affected the induction kinetics of *Irf7*. We observed decreased induction of *Irf7* mRNA in response to TNF α stimulation in shBUD13 BMDMs as compared to control BMDMs (Figure 3D), consistent with the idea that intron retention leads to transcript degradation (Jacob and Smith, 2017). Other TNF α induced transcripts that lacked a *BUD13* dependent splicing defect showed similar induction between the time-courses (Figures S2 D-F).

BUD13 knockdown alters the type I interferon response.

Because *Irf7* is known as a ‘master regulator’ for robust type I interferon production (Honda et al., 2005), we next investigated the effect of BUD13 knockdown on a type I interferon response. To do so, we stimulated BMDMs with the TLR3 agonist Poly(I:C) for up to 24 hours. Activation of TLR3 leads to the production of type I interferons followed by the downstream induction and activation of *Irf7*, which serves to amplify the type I interferon response via positive feedback (Ciancanelli et al., 2015). We again observed differential splicing between the shBUD13 samples and the control samples in intron 4 of *Irf7* (Figures 4A, S3A). As before, there is a significant difference in the Δ SR of intron 4 as compared to every other junction in the *Irf7* transcript (Figure 4B. $p < 0.001$, Student’s t-test), whereas all other pairwise comparisons are insignificant. As is the case with TNF α , knocking down BUD13 altered *Irf7* induction kinetics. Less *Irf7* mRNA is induced at 240, 720, and 1440

mins of poly(I:C) stimulation (Figure 4C). This reduction in *Irf7* mRNA leads to a decrease in the amount of IRF7 protein produced (Figure 4D).

Next we looked at how this reduction in IRF7 would alter the production of RNA from interferon signature genes (ISGs). Expression of 119 ISGs (selected based on upregulation in response to IFN α ; see methods) (Mostafavi et al., 2016) was examined. In unstimulated BMDMs, used as a baseline, the median log₂ expression fold change (FPKM shBUD13/ FPKM control) is 0.1655 (Figure 4D). In contrast, at 720 mins of stimulation, the median log₂ expression fold change shifts to -0.1007 (Figure 4E), indicating a significant decrease in ISG expression in the shBUD13 sample compared to the control sample at this time-point compared to the baseline (Wilcoxon rank-sum, $P < 0.001$). This significant decrease in ISG expression remained true when comparing any of the ‘early’ timepoints (0, 15, 60 mins) to any of the ‘late’ timepoints (240, 720, 1440 mins) (Figure 4G, Wilcoxon rank-sum, $P < 0.001$). qPCR was used to monitor expression of both IFN α and IFN β following 720 and 1440 mins of Poly(I:C) stimulation. We observed significant reduction in both when comparing the shBUD13 samples to the control samples (Figures 4H, I). To ensure differential expression of ISGs was not due to splicing defects from BUD13 knockdown, we quantified the Δ SR for every ISG junction at 720 mins. The fourth intron of *Irf7* has the greatest Δ SR at 0.227. Only four other junctions of the 375 that were examined have Δ SRs greater than 0.1, and the majority of junctions have Δ SRs close to 0 (Figure S3C and Table S3; mean = 0.002, median = 0). Similar results were obtained when BMDMs were stimulated with the TLR9 agonist CpG (Figures S4A-E). Taken together, we conclude that Bud13 deficiency results in a highly compromised type I interferon response.

We next examined whether *Irf7* pre-mRNA with a retained fourth intron was able to exit the nucleus and enter the cytoplasm. BMDMs were stimulated with poly(I:C) and fractionated into a nuclear fraction (containing the nucleoplasm and chromatin) and a cytoplasmic fraction. RNA-seq was performed on both fractions. In the cytoplasm, we found *Irf7* mRNA to be completely spliced (Figures 4J, K). Thus, unspliced *Irf7* is either being degraded in the nucleus, or it makes it to the cytoplasm and is degraded extremely quickly, such that virtually no signal can be detected via RNA-seq. Furthermore, in support of our whole cell sequencing data and IRF7 immunoblots, we observed less *Irf7* mRNA in the cytoplasm in shBUD13 samples as compared to control samples (Figure 4L). Finally, although we notice a large number of unspliced reads in the nucleus at all junctions in both samples, the fourth intron had a greater nuclear RPKM in shBUD13 compared to control BMDMs across the stimulation time-course (Figure 4K) and had a significantly greater nuclear Δ SR as compared to any other junction in the transcript (Figure S4G).

Global analysis of the role of BUD13 in BMDMs.

We next investigated global splicing differences caused by Bud13 knockdown. Using the TNF α stimulated data-set, Δ SR was calculated for every junction in every expressed gene. We found that a number of other transcripts had a Bud13 dependent junction (Figure 5A). Of note, the fourth intron of *Irf7* is among the most BUD13 dependent junctions in both the TNF α and Poly(I:C) data-sets (Figures 5A and S5H, see methods for analysis details). Similar to the case with *Irf7*, almost all transcripts contain only a single Bud13 dependent junction, even when low thresholds are used to quantify dependency (Figure 5B). To

determine whether splicing differences caused by BUD13 knockdown led to altered gene expression, we compared the effect of the BUD13 knockdown on genes that contained a BUD13 dependent junction to those that did not. (see methods). The median \log_2 expression fold change (FPKM shBUD13/ FPKM control) for genes containing a BUD13 dependent junction was -0.5084. In contrast, the median \log_2 expression fold change (FPKM shBUD13/ FPKM control) for genes without any junctions affected by Bud13 knockdown was -0.2170. Thus, we conclude there is an inverse relationship between IR due to BUD13 knockdown and gene expression (Wilcoxon rank-sum, $P < .01$) (Figure 5C).

Next, it was of interest to us to identify sequence elements that led BUD13 to have its specific splicing effect. The most evident element to explore was the effect of splice site strength on BUD13 dependent splicing. Previous work has shown that the yeast orthologue of *Bud13* plays a role in efficient splicing for a junction with a weak 5' splice site (Dziembowski et al., 2004). Further, the junction affected in *Irf7* has a non-consensus 5' splice site. To investigate this issue, we first quantified every 5' and 3' splice site using a maximum entropy model (Yeo and Burge, 2004). Then, we took progressively weaker splice site thresholds, and compared the mean Δ SR for every junction below that threshold to the mean Δ SR of every junction in the data-set (Figure 5D). We saw that as the splice site threshold for the 5' splice site became progressively weaker, the mean Δ SR for junctions weaker than that threshold increased and thus there was a greater BUD13 splicing effect. This result was not seen when the same analysis was applied to the 3' splice site. In support of a 5' splice site dependency for a BUD13 effect, we noticed differences in the 5' splice site motif of BUD13 dependent junctions as compared to to all expressed junctions (Figure 5E).

We then analyzed the BUD13 splicing effect with respect to other features known to correlate with IR (Braunschweig et al., 2014). Across all time-points for both TNF α (Figures 5F-H) and Poly(I:C) (Figures S5 A-C), BUD13 dependent introns were dramatically smaller and had increased G/C content in both the intron and in the flanking exons. We also noticed that the distance from the branch point to the 3' splice site was smaller in the BUD13 dependent introns than in the total data-set (Figures 5I and S5D). This could be a byproduct of the smaller intron length; however, it is of interest because BUD13 has been shown in yeast to bind just downstream of the branch point (Schneider et al., 2015). A significant difference was not seen in branch point strength and BUD13 splicing effect (Figures S5 E, F).

Finally, as IR is only one form of alternative splicing, we looked at whether other forms of alternative splicing were affected by Bud13 knockdown. We found that the majority of statistically significant alternative splicing events involved intron retention (Figure S5G, see methods). Of the 42 alternative splicing events that were significant in multiple data-sets upon BUD13 depletion, 27 involved intron retention, 9 involved a skipped exon, and the remaining 6 involved either an alternative 3' or 5' splice site. IR at intron 4 in *Irf7* was the only alternative splicing event that occurred in transcripts related to the type I interferon response.

eCLIP shows enrichment on Bud13 dependent junctions near the 3' splice site.

Next we used enhanced crosslinking and immunoprecipitation (eCLIP)-seq data from the ENCODE Project Consortium (Consortium, 2012) to investigate BUD13 binding specificity

across the genome. We found that in K562 and Hep G2 cells, the majority of *Bud13* eCLIP-sequencing reads were located downstream of the branchpoint near the 3' splice site (Figures 6A, B), consistent with what is seen in yeast (Schneider et al., 2015). Plots are shown as a measure of binding over input. There is some read density near the 5' splice site, which we hypothesize is due to BUD13's association with the spliceosome. Although BUD13 may not bind near the 5' splice site, factors in the spliceosome that interact with BUD13 may immunoprecipitate with it, leading to 5' signal. Data for SF3B4 and PRPF8, known RBPs that interact with the 3' and 5' splice site respectively, is also shown for comparison (Figures 6A, B). Additionally, it was expected that BUD13 binding would correlate with BUD13 activity. To test this hypothesis, we used knockdown data from the ENCODE Project Consortium to determine BUD13 dependent junctions in K562 and Hep G2 cells. In K562 cells, we noticed that there was a significant increase in BUD13 binding over input at BUD13 dependent junctions (Figure 6A). In Hep G2 cells, this increase was less pronounced (Figure 6B); however, we note that we found BUD13 knockdown had a much greater impact in K562 cells as compared to Hep G2 cells (Figure S5I). In order to survey a large enough selection of junctions in Hep G2's, we had to significantly lower our threshold for what was deemed a BUD13 dependent junction (see methods), which in turn might explain the dampened BUD13 binding/activity relationship in Hep G2. We conclude that Bud13 either preferentially associates with these BUD13 dependent junctions or associates with them for a longer period of time.

We then performed peak calling to determine the location of significant peaks. We found the majority of peaks are in intronic regions or intron-exon junctions and that most of the peaks that lie in intron-exon junction are located at the 3' junction (Figures 6C, D). As

might be expected from knockdown data, when comparing introns that have an overlapping eCLIP peak to all introns from expressed transcripts, we see both a length and G/C% bias (Figures 6E, F). BUD13 peaks tend to fall in smaller introns that are GC rich, a finding consistent with the Δ SR data. Lastly, a list of the GO biological processes most enriched from the list of peaks in K562 and Hep G2 cells is shown (Figure 6G).

Bud13 knockdown alters the BMDM response to VSV.

Vesicular stomatitis virus (VSV) is a (-)ssRNA virus known to induce type I IFN through TLR7 (Lund et al., 2004). To test whether impairment of IRF7 due to BUD13 knockdown was present in VSV stimulated BMDMs, we infected both shBUD13 and control BMDMs at an MOI of 5 and 10. At all time-points throughout infection in both MOIs, there was dampened IRF7 induction (Figures 7A, B) as quantified by Taqman qPCR. Next, in order to determine the consequences of impaired IRF7 induction, we determined the yield of virus from BMDMs following a period of infection with a given input MOI. shBUD13 BMDMs produce significantly more VSV as compared to control BMDMs (Figure 7C). This difference in viral production is presumably due to decreased production of IRF7 associated with depletion of BUD13 and the corresponding dampened type I interferon response. To test the extent to which this increase in viral production following BUD13 depletion was due to impairment of IRF7, we rescued IRF7 levels by expressing *Irf7* cDNA either in the context of the BUD13 knockdown or the control. We found overexpression of *Irf7* cDNA effectively rescues the ability for a cell to clear virus (Figure 7C). As such, we conclude that the viral susceptibility associated with BUD13 knockdown is due primarily to the inefficient production of mature IRF7 associated with BUD13 depletion.

DISCUSSION

In this study, we sought proteins that might relate to the poor splicing of an intron in *Irf7* transcripts. Using RAP-MS, we identified BUD13 as a protein that has the ability to increase splicing of the *Irf7* intron. In the absence of BUD13, in response to inflammatory stimulus, macrophages produced *Irf7* with increased intron retention (IR) and notably less mature *Irf7* transcript and protein (Figures 3C, 4C, 4D, S4C). *Irf7* is the interferon-inducible master regulator of the type-I interferon-dependent immune response (Honda et al., 2005). Correspondingly, depletion of BUD13 led to a general reduction in ISG and cytokine production, implying a compromised type I interferon response (Figures 4E-J, S4D-G). This splicing and corresponding expression defect upon BUD13 depletion was observed under various stimulation regimens and times. We found that macrophages deficient for BUD13 were strikingly more susceptible to infection by VSV, presumably owing to the reduction in *Irf7* transcript levels (Figure 7).

We observed the BUD13 splicing dependence in other introns of other genes. A number of short, GC-rich introns with non-consensus splice donor sites were excised inefficiently when BUD13 levels were depleted (Figure 5). As was the case with *Irf7*, this increased IR reduced mature transcript levels (Figure 5A). Transcripts containing retained introns have been shown in the literature to be degraded by two mechanisms: (1) nuclear degradation via the RNA exosome, (2) cytoplasmic degradation upon detection of a pre-termination codon (PTC) via the NMD decay machinery. Although the majority of these introns contain a PTC, it remains to be determined whether degradation is occurring in the nucleus or cytoplasm (Jacob and Smith, 2017; Sayani and Chanfreau, 2012)

Bud13 was originally identified as a part of a “Retention and Splicing” (RES) complex (Dziembowski et al., 2004) in yeast. However, yeast Bud13 (ScBud13) and mammalian BUD13 are significantly different lengths (266 vs. 637 amino acids) (Na et al., 2016), with only the mammalian protein containing a large, disordered arginine-rich N-terminal domain. ScBud13 counteracts IR in introns within the mediator complex, mating genes, and tRNA modifying genes (Ni and Snyder, 2001; Zhou et al., 2013; Zhou and Johansson, 2017), which in turn impair yeast budding. In connection with the RES complex, ScBud13 is thought to safeguard formation of the ‘B^{act} complex’ of the spliceosome (Bao et al., 2018). In the B^{act} stage, the 5’ splice donor and branch point are recognized by the spliceosome. However, progression to catalysis of the first step of the splicing reaction requires remodeling of several spliceosome components (Ohrt et al., 2012). Lack of the RES complex has been shown to lead to premature binding of Prp2, a quality control factor that is responsible for spliceosome remodeling as well as the disassembly of suboptimal substrates. It has been hypothesized that ScBud13 and the RES complex temporally regulate Prp2 binding (Bao et al., 2018). In the mammalian context, short, GC-rich introns with weak donor sites may be particularly susceptible to Prp2-mediated disassembly, which may explain the specificity of IR events upon BUD13 depletion.

In yeast, differential studies using mass spectrometry (Fabrizio et al., 2009) and cross-linking have established that some ScBud13 is detectable in preparations of stalled B, B^{act}, and B* complexes. One cryo-EM structure of the yeast spliceosome found density corresponding to ScBud13 in a stalled B^{act} pre-catalytic complex, although a structure of the stalled B complex found only weak density for ScBud13 (Plaschka et al., 2017; Zhou and Johansson, 2017). In mammals, structural evidence of BUD13 is limited. Given the partial

sequence homology between all members of the yeast RES complex and their mammalian counterparts, it is not surprising that BUD13 (and other RES complex members) are often undetectable in preparations of stalled spliceosomes using cross-linking and mass-spectrometry. Furthermore, BUD13 was not detected in a recent human cryo-EM structure of a stalled B complex (Bertram et al., 2017). Taken together, it is not yet possible to determine if the sub-stoichiometric nature of BUD13 in mammalian spliceosome complexes is because it is constitutively associated but highly transient or because it serves as a non-essential accessory to spliceosome function. Cryo-EM studies, as well as single molecule studies, would seem to suggest compositional heterogeneity of the spliceosome, and that the BUD13-endowed spliceosome may catalyze the splicing reaction in a fundamentally different way than is used in its absence (Bao et al., 2018; Blanco et al., 2015; Hoskins and Moore, 2012).

Recently, the RES complex in zebrafish was shown to regulate levels of IR in short, GC-rich introns in knockout studies (Fernandez et al., 2018). Indeed, both in Zebrafish (Fernandez et al., 2018) and *C. Elegans* (Jiang et al., 2001), deficiency of RES components has been reported to lead to embryonic lethality. Our results show that mammalian BUD13 shares this splicing fidelity function, and deficiency may prevent proper development. Despite this, knockdown and knockout cell lines have displayed no overt growth defects, suggesting a developmental but not immune-cell intrinsic dependence on BUD13 for survival. Of note, we did knockdown the other components of the RES complex, RBMX2 and SNIP1 (Figure S7). We found that the fourth intron of *Irf7* has a largest Δ SR as compared to other introns in the transcripts (Figure S7B) and that there is some global IR upon knockdown (Figure S7C); however, these effects are very slight as compared to what

is observed with BUD13 depletion. It might be that more efficient knockdowns or a total knockout of these components is needed to replicate the strong effect seen with BUD13 depletion, or alternatively that there is some redundancy with these components in mammalian cells.

With respect to *Irf7*, the fact that a crucial immunological gene has such an intron, with its variety of seemingly negative characteristics that make it difficult for the spliceosome to excise, begs the question as to why it exists. At the heart of an inflammatory response is a tightly regulated gene expression program. Regulation of this gene expression program is crucial as small changes can shift the balance away from protective immunity towards either nonexistent or destructive immunity (Kontoyiannis et al., 1999). Here we've shown that alterations to the splicing efficiency of the fourth intron have the ability to significantly alter the functional output of IRF7. Thus, by existing in the *Irf7* transcript and commonly being retained, it stands to reason the weak fourth intron acts to dampen IRF7 output, perhaps as a means to mitigate what otherwise would be an unchecked or inappropriately scaled response. Whether a cell actively controls this splicing event and thus, the intron serves as a regulatory control point, remains unknown. Further, it remains unknown whether BUD13 plays a role in this regulation or whether it simply represents a mechanism that evolved to counter intron retention in a subset of introns that require splicing but happen to be inherently weak.

In summary, we found that BUD13 modulates gene expression through its ability to alter IR, often in notably small, GC-rich introns with weak splice sites. Deficiency of BUD13 results in IR and concomitant decreased gene expression in transcripts such as *Irf7*, dampening the type I interferon response and increasing viral susceptibility. Therefore, in mediating *Irf7* gene expression, BUD13 presents a potential therapeutic target for the

treatment of infections or autoimmune conditions. Future studies should seek to understand why BUD13 is vital for the efficient splicing of only a subset of junctions and whether or not this junction specificity plays an active role in regulating gene expression. If modulated, this strategy by which components associated with the spliceosome rescue transcripts from intron retention and degradation may represent a previously underappreciated layer of regulation in many gene expression programs.

ACKNOWLEDGEMENTS

The authors would like to thank Mario Blanco and Mitchell Guttman (Dept. of Biology, Caltech) for assistance with eCLIP analysis; Patricia Turpin, Mati Mann, Guideng Li, and Alok Joglekar (Dept. of Biology, California Institute of Technology) for experimental and computational assistance. Additionally, the authors would like to thank Genhong Cheng (Department of Microbiology, Immunology & Molecular Genetics, UCLA) for VSV and Jae Jung (Department of Molecular Microbiology & Immunology, USC) for VSV-GFP. This work was funded from a grant from NIH (5R21AI126344) and from an endowment provided by the Raymond and Beverly Sackler Foundation.

AUTHOR CONTRIBUTIONS

L.F., D.M., and D.B., conceived and designed experiments. L.F. conducted experiments. C.B. helped develop RAP-MS and knockdown experiments. L.F. and D.M. analyzed sequencing data. A.M. oversaw mass spectrometry and M.J.S. performed mass spectrometry analysis. L.S.F, D.M., and D.B wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests with respect to this manuscript.

EXPERIMENTAL METHODS

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Baltimore (baltimo@caltech.edu).

Experimental Model and Subject Detail

Animals

The California Institute of Technology Institutional Animal Care and Use Committee approved all experiments. C57BL/6 WT mice were bred and housed in the Caltech Office of Laboratory Animal Resources (OLAR) facility. C56BL6/J mice were sacrificed via CO₂ euthanasia and sterilized with 70% ethanol. Femur and tibia bones harvested and stripped of muscle tissue. Bone marrow cells were resuspended in 20mL of fresh DMEM. 2.5×10^6 bone-marrow cells plated in a 150mm dish in 20mL of BMDM Media (DMEM, 20% FBS, 30% L929 condition media, and 1% Pen/Strep) and grown at 5% CO₂ and 37°C. BMDM media was completely replaced on day 3 as well as a supplemental addition of 5mL L929 condition media on day 5.

Cell Culture

Human embryonic kidney cells (HEK293T) from ATCC were cultured in DMEM supplemented with 10% FBS and 1% Pen/Strep. Cell line was maintained at 37°C in 5% CO₂.

Method Detail

Knockdown Experiments

BMDMs for knockdown experiments were grown as described above with a few additions. On days 3 and 4, retrovirus encoding shRNAs were added to cells. On day 5, cells were selected with puromycin (5ug/mL). On day 8, following ~72 hours of puromycin treatment, media was removed and 10mL of PBS w/ 2mM EDTA was added. Depending on the experiment, cells were stimulated directly or lightly scraped and replated in 6-well plates or 10-cm dishes for stimulation the following day. Stimulation involved with either 20ng/mL of TNF α , 5ug/mL Poly(I:C) (Sigma), 5 μ M ODN 1585 (InvivoGen), or the indicated MOI of VSV.

RNA Isolation:

Total RNA was purified from BMDMs using TRIzol reagent (Ambion) as per the manufacturer's instructions. Genomic DNA in RNA purifications was eliminated through treatment with Turbo DNase (Thermo Fisher Scientific) for 30 min at 37°C. 0.1-1 μ g RNA and 1 μ M dT(30) oligo (d14-954: 5'-AAGCAGTGGTATCAACGCAGAGTACT(30)) was heated at 80°C for 2.5min followed by snap cooling on ice. 10 μ L template-switch RT mix added (10 μ M template-switch oligo (TSO: 5'-AAGCAGTGGTATCAACGCAGAGTACACArGrGrG), 20mM DTT, 2X ProtoScript II Reverse Transcriptase Reaction Buffer (NEB), 1mM dNTPs, 40U Murine RNase Inhibitor (NEB), and 200U ProtoScript II (NEB) Reverse Transcriptase. Reaction incubated in thermocycler with the following program: 1. 42°C for 30min, 2. 45°C for 30min, 3. 50°C for 10min, followed by deactivation of RT for 10min at 80°C.

RNA Fractionation:

Confluent 10-cm dish of mature BMDMs were scraped into 400 μ L cold NP-40 lysis buffer, APJ1 (10mM Tris-HCl (pH 7.5), 0.08% NP-40, 150mM NaCl). Lysed cells layered onto 1mL cold sucrose 322 cushion, APJ2 (10mM Tris-HCl (pH 7.5), 150mM NaCl, 24% w/v sucrose) and centrifuged for 10min at 4°C and 13000 rpm. The supernatant from this spin represents the cytoplasmic RNA fraction, which is immediately added to 3 volumes of 100% ethanol and 2 volumes of buffer RLT (4M GuSCN, 325 0.1M β -mercaptoethanol, 0.5% N-lauroyl sarcosine, 25mM Na-citrate, pH7.2) and stored at -80°C until ready to purify RNA. Pellet, containing intact nuclei, is resuspended in 500 μ L TRIzol reagent. If the pellet was difficult to dissolve, it was heated at 50°C with occasional vortexing. 100 μ L chloroform added and shaken vigorously for 15-20s; allowed to phase separate at room temperature for 5min. Tube centrifuged at 4°C and 12000 x g for 15min. Clear upper aqueous phase removed to a new tube, ensuring white DNA mid-phase is not removed, and is immediately added to 3 volumes of 100% ethanol and 2 volumes of buffer RLT and stored at -80°C until ready to purify RNA. RNA is purified according to Qiagen RNeasy column protocol and eluted in 30 μ L nuclease-free H₂O. RNA samples are DNase treated with Turbo-DNase and stored at -80°C.

Library preparation and RNA-Seq Analysis

Limited PCR amplifications was performed prior to library preparation. PCR reaction done with KAPA HiFi HotStart 2x ReadyMix, 5% cDNA, and 1 μ M primer (d14-955: 5'-AAGCAGTGGTATCAACGCAGAGTACT). Thermal cycler programmed for 120 seconds at 95°C as initial denaturation, followed by 14 cycles of 30sec at 95°C for

denaturation, 30sec at 62.5°C as annealing, 150sec at 72°C for extension, and final extension at 72°C for 5 min. PCR reactions 0.9X SeraMag and eluted in 25µL. Concentrations of purified library determined using Qubit High Sensitivity dsDNA kit (Invitrogen) as described. Full length cDNA libraries were barcoded using the Nextera XT Tagmentation protocol (Illumina).

RNA-Antisense Purification

RNA antisense purification-mass spectrometry (RAP-MS) was performed as described in McHugh et al. with a few alterations. Briefly, we designed three 90-mer DNA oligonucleotide probes that were antisense to the complementary target RNA sequence in both *Irf7* and *Actb* transcripts. Each probe was targeted to a different location on the transcript and modified with a biotin in order to enable capture of DNA:RNA hybrids on streptavidin coated magnetic beads.

RNA Prep and Lysis: ~250million cells, or 25 150mm plates of BMDMs were used for each capture. Following stimulation with TNF α (20ng/ml) for 30 minutes, ~5-10 mL of PBS w/ 2mM EDTA was added to each plate and cells were removed by lightly scraping. Cells were pelleted, resuspended in PBS, and poured into a new 150mm plate. The cells were then crosslinked in Spectrolinker at 254 nm wavelength with 0.8 J/cm² (instrument setting: 8000 x 100 uJ/cm²). Following crosslinking, cells were again pelleted, at which point the pellet could be frozen and stored at -80°C. Cells were lysed in 2mL of lysis buffer per capture (10 mM Tris pH 7.5, 500 mM LiCl, 0.5% Triton X-100, 0.2% sodium dodecyl sulphate, 0.1% sodium deoxycholate) supplemented with Protease Inhibitor Cocktail (EMD Millipore) and 1000 U of Murine RNase Inhibitor (New England Biolabs). We found the smaller the

volume used per sample, the more efficient the capture was downstream and thus the minimum volume needed to lyse cells should be optimized. Samples were incubated for 10 min on ice to allow lysis. Following lysis, sample was passed through 20-gauge needle once and then 26-gauge needle 3-5 times to disrupt the pellet and shear genomic DNA. In between passing the sample through the 26-gauge needle, the sample was sonicated on ice with a microtip set at 5W power for a total of 30 s in intermittent pulses (0.7 s on, 1.3 s off). Samples were then mixed with twice the lysate volume of 1.5x LiCl/Urea Buffer (the final buffer contains 10 mM Tris pH 7.5, 500 mM LiCl, 0.5% Triton X-100, 0.2% SDS, 0.1% deoxycholate, 4 M urea). Lysates were incubated on ice for 10 min then cleared by centrifugation for 10 min at 4,000g.

Pre-clearing lysate: BioMag streptavidin beads (Bang Laboratories Inc.) were first washed 3x in 0.25-0.5ml of 500mM LiCl/4M Urea buffer (10 mM Tris pH 7.5, 500 mM LiCl, 0.5% Triton X-100, 0.2% SDS, 0.1% deoxycholate, 4 M urea). 50ul of beads were added to each sample and the samples were incubated at 37°C for 30 min with shaking. Streptavidin beads were then magnetically separated from lysate samples using a magnet. The beads used for pre-clearing lysate were discarded and the lysate sample was transferred to fresh tubes twice to remove all traces of magnetic beads. Input for quality control experiments can be removed at this point.

Hybridization, Capture of Probes and Elution of Associated Protein: Following pre-clearing, the biotinylated 90-mer DNA oligonucleotide probes specific for the RNA target of interest (vary per sample but ~5ul of 25uM per probe) were heat-denatured at 85°C for 3 min and then snap-cooled on ice. Probes and pre-cleared lysate were mixed and incubated at 55°C with shaking for 2 h to hybridize probes to the capture target RNA. 500mL of washed

streptavidin beads (Bang Laboratories Inc.) were then added to each sample at 55°C with shaking for 30 mins. Beads with captured hybrids were washed 6 times with LiCl/Urea Hybridization Buffer. If needed, 1% of the beads can be removed for qPCR quality control experiment. TRIzol reagent can be added directly to beads to elute RNA. Beads were then resuspended in Benzonase Elution Buffer (20 mM Tris pH 8.0, 2 mM MgCl₂, 0.05% NLS, 0.5 mM TCEP) and 125 U of Benzonase nonspecific RNA/DNA nuclease was added. Incubation occurred for 1-2 h at 37°C. Beads were then separated from the sample using a magnet. Supernatant was collected. Contaminant beads were removed by 5 rounds of magnetic separation on supernatant. Protein was precipitated overnight at 4°C with 10% trichloroacetic acid (TCA). TCA treated protein elution samples were pelleted by centrifugation for 30 min at 20,000g, then washed with 1 ml cold acetone and recentrifuged. Final protein elution pellets were air dried to remove acetone, resuspended in fresh 8 M urea dissolved in 40 ml of 100 mM Tris-HCl pH 8.5, and stored at -20°C.

Mass Spec Prep. and Analysis Performed as in McHugh et al. with few exceptions. Instead of SILAC we label proteins at the mass spec prep step using TMT (Thermo). After desalting on a Microm Bioresources C8 peptide MicroTrap column and lyophilization of peptide fraction, lyophilized protein pellets were resuspended in 100mM TEAB at a concentration of 1ug/ul. We then added 1.64ul of TMT labelling reagent to each ug of sample. The reaction was incubated for one hour at room temperature. The reaction was quenched with 0.32ul of 5% hydroxylamine per ug of protein used and incubated for 15 mins at room temperature. Following quenching, the samples were mixed, desalted as before, lyophilized, and mass spec was performed on Orbitrap Fusion mass spectrometer using a TMT instrument method as described in Liu et al (Liu et al., 2016).

Raw files were searched using MaxQuant (v. 1.5.3.30) against the UniProt mouse database (59550 sequences) and a contaminant database (248 sequences). TMT 6plex was selected as the quantitation method with a reporter mass tolerance of 0.3. Oxidation of methionine and protein N-terminal acetylation were variable modifications and carbamidomethylation of cysteine was fixed modification. A 1% protein and peptide false discovery rate as estimated by the target-decoy approach was used for identification.

RNA Immunoprecipitation

RNA immunoprecipitations were performed as previously described. Between 5-10 confluent 15 cm² dishes of BMDMs per sample were stimulated with either 20ng/mL of TNF α for 30 minutes or 5ug/mL Poly(I:C) for 12 hours. Following stimulation, proteins were cross-linked to DNA by adding formaldehyde directly to the media to a final concentration of 0.75%, with light shaking at room temperature for 10 mins. To quench the crosslinking reaction, glycine to a final concentration of 125 mM was added to the media and incubated with shaking for 5 mins at room temp. Media was then aspirated and cells were rinsed twice with 10 mL of cold PBS. Following the second wash, cells were scraped into 10mL of PBS and spun down gently (5 min, 4°C, 1,000xg). Final cell pellet was resuspended in 0.1-1mL of polysome lysis buffer (100 mM KCl, 5 mM MgCl₂, 10 mM HEPES (pH 7.0), 0.5% NP40, 1 mM DTT, 100 U.ml RNase Inhibitor (NEB)) supplemented with Protease Inhibitor Cocktail (EMD Millipore). At this point the mRNP lysate was frozen. If needed, passing the lysate through a small gauge needle can help with lysate. Protein-G beads were pre-treated at 4°C with NT2 (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM MgCl₂, 0.05% NP40) supplemented with 5% BSA to a final ratio of 1:5 for at least 1h before

use. Appropriate amount of antibody per sample (optimized based on antibody used but typically ~1-10ug) was added to 250-500ul of bead/BSA slurry and incubated at 4°C. Following incubation, beads were spun down and washed with 1 ml of ice-cold NT2 buffer 4–5 times. Following final wash, beads were resuspended in 850ul of NT2 and supplemented with 200U of RNase inhibitor, 10 µl of 100 mM DTT and EDTA to 20 mM. Frozen lysate was thawed and centrifuged at 15,000*g for 15 mins. The cleared supernatant was removed and 100ul was added to the prepared beads. Input removed at this step. Beads and lysate were incubated for 4h at 4°C with mixing. The beads were washed 4-5 times with ice-cold NT2 and then resuspended in 100ul of NT2 buffer. 4ul of 5M NaCl was added incubated with shaking at 65°C for 2 hours. NT2 buffer can also be supplemented with 30 µg of proteinase K to release the RNP component. RNA was isolated by adding TRIzol reagent (Ambion) as per the manufacturer's instructions. RNA was reverse transcribed and quantification was performed using TaqMan qPCR.

Immunoblot

BMDM samples were prepared as described previously. BMDMs were stimulated with either TNF α or Poly(I:C) for the indicated period of time. Cells extracts were collected using RIPA lysis buffer (Sigma cat: R0278-50ML), and were subjected to gel electrophoresis and transfer onto a nitrocellulose membrane. pRroteins were analyzed by immunoblot using the following reagents: anti-IRF7 (abcam, ab215326) and anti-beta Actin (Cell Signalling, 13E5). For nuclear fractionation, cells were scraped into subcellular fractionation buffer (20mM HEPES (pH 7.4), 10 mM KCl, 2 mM MgCl₂, 1mM EDTA, 1 mM EGTA). The cells were then passed through a 27-gauge needle 10 times, incubated on ice for 10 mins, and spun

down at 720xg for 5 min. The pellet contained the nuclei, which was washed with fractionation buffer, passed through a 25-gauge needle 10 times, and centrifuged again at 720xg for 10 mins. The resulting pellet was resuspended in RIPA lysis buffer. Equal amounts of proteins were analyzed by immunoblot using the following reagents: anti-IRF7 (Millipore, ABF130), anti-Lamin B1 HRP conjugate (Cell Signalling, D9V6H), and anti-rabbit IgG HRP conjugate (Cell Signalling).

Viral Plaque Assays

Plaque assays were done on Vero cells. 2.5×10^5 Vero cells were plated in a 12 well plate the night before infection. Prior to infection, cells were checked to ensure confluence. VSV was serially diluted and infected in 12 well plate for 1 h. VSV was then removed and cells were layered carefully with DMEM supplemented with 2% FBS and 0.4% agarose. Plate was incubated for 2 days, and then fixed with 10% formaldehyde, for 1 h to overnight. Finally, agarose plugs were removed carefully and cells were stained with crystal violet.

VSV-GFP Infection Experiment

BMDMs were grown as described above in 150mm dishes. On day 8, following ~72 hours of puromycin treatment, media was removed and 10mL of PBS w/ 2mM EDTA was added. Cells were lightly scraped and 250,000 cells/well were replated in 12 well plates in BMDM media. Cells were left for 12 hours to adhere. Following adherence, VSV-GFP was added at the specified MOI for the specified amount of time. Following the time-course, cells were lightly scraped, washed and spun down, and resuspended in PBS. Samples were analyzed on

a MACSQuant10 Flow Cytometry machine (Miltenyi). Gating strategy depicted in Figure S7.

VSV-GFP Viral Supernatant Experiment

BMDMs were grown as described above in 150mm dishes. On day 8, following ~72 hours of puromycin treatment, media was removed and 10mL of PBS w/ 2mM EDTA was added. Cells were lightly scraped and 400,000 cells/well were replated in 12 well plates in BMDM media. Cells were left to adhere for 12 hours, before being infected at an MOI of 25 for 8 hours. Following infection, virus was removed and the cells were washed with PBS three times. Then, 500ul of BMDM media (DMEM, 20% FBS, 30% L929 condition media, and 1% Pen/Strep) was added to each well. 18 hours later, media was collected and stored at -80°C. To titer viral supernatant, Vero cells were plated in a 96-well plate at 30,000 cells per well in 90ul of D10 media. 12 hours after plating, 90ul supernatant was added to the 90ul of D10 at different dilutions. PFU/mL was calculated from a standard curve with a virus of known concentration.

Quantification and Statistical Analysis

All statistical analysis was performed in Python (version 2.7.9). Unless otherwise indicated in figure legends, statistical significance measurements were marked as follows: * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$, and n.s. denotes not significant. RNA-Seq expression and splicing analysis as well as eCLIP analysis is described in more detail below.

RNA-Sequencing Analysis

Sequencing was performed on a HiSeq 2500 High Throughput Sequencer (Illumina). Single-end 50-mer reads were aligned using Tophat v2.1.1 (Kim et al., 2013). Gene expression was determined using Cufflinks v2.2.1 and the FPKM (Fragments Per Kilobase Million) metric (Trapnell et al., 2010).

Splicing Ratio and Δ SR Calculation

A custom script was written in Python using the HTSeq (Anders et al., 2015) library to calculate Splicing Ratio. First, reads that map to an intron or exon feature are summed. To map to a feature, reads must have >1 bp overlap with the feature. If a read maps to more than one feature, such as in the case of a splice junction read, the read is split between the features. SR is calculated by taking the length normalized number of reads that map to each intron, divided by the average length normalized number of exon reads plus the length normalized intron value. When SR is equal to 0, this indicates a junction is completely spliced. In contrast, large SR values indicate intron retention. We use the SR value to calculate Δ SR, which is equal to $SR(\text{shBUD13}) - SR(\text{Ctl})$. Values greater than 0 indicate the junction is more unspliced in the shBud13 sample, whereas values less than 0 indicate the junction is more unspliced in the Ctl sample. For each stimulation (TNF α , Poly(I:C), and CpG), Δ SR was calculated for each individual junction of the Irf7 transcript. Bar graphs represent the mean (error bars indicate s.d.) Δ SR for stimulated time-points (non-zero time-points). For the global analysis, in order for the Δ SR of a junction to be considered, it must pass through a number of filters. To account for transcripts that are annotated in Ensembl version 67, but not expressed, we set an FPKM threshold of 15. Further, a local normalized

read count threshold on the upstream/downstream exons was implemented to ensure a level of sequencing depth needed to get accurate splicing values. To pass this threshold, the sum of the reads that map to the the upstream/downstream exons divided by the length of these exons must be ≥ 0.25 .

ISG and Genome-Wide Analysis

ISGs used in Figures 4 E-H were selected based on induction 2 hours after *in vivo* IFN α injection (Mostafavi et al., 2016). We classified ISGs to be any gene with a fold change ≥ 3.5 following 2 hours of induction. Intron RPKM was calculated using a custom python script with the HTSeq library. In Figure 5a, transcripts from the 30 min. TNF α data-set that had a junction with a Δ SR value above 0.15 were sorted into an ‘increased IR’ category (Δ SR >0.15), whereas all other transcripts were sorted into an unaffected category (Δ SR <-0.15). The selected data-set is representative of all time-points from the TNF α , Poly(I:C) and CpG datasets. A maximum entropy model was used to calculate 3’ and 5’ splice site strengths (Yeo and Burge, 2004). To determine differences in 5’ splice site sequence for Bud13 dependent junctions, the nine base pair sequence near the 5’ splice site junctions for junctions that had a Δ SR >0.15 was compared to all expressed junctions (FPKM >1). The top Bud13 dependent junctions were plotted based on the average Δ SR value across all time-points from the TNF α data-set (Figure 5D) as well as the Poly (I:C) data-set (Figure S5H). Junctions that had a Δ SR value <0.15 in a time-point were filtered out in the TNF α data-set, while junctions that had a Δ SR value <0.15 in two time-points were filtered out in the Poly (I:C) data-set. The zero time-point was removed for the transcripts induced by the stimulant (Irf7 and Cd14).

For the comparison of alternative splicing events, rMATs (S. Shen et al., 2014) was used on the TNF α data-set. Splicing events were deemed significant if $p < 0.05$ and $FDR < 0.1$ for all time-points. SVMBPfinder was used to determine BP related features (BP strength and distance from BP to 3' splice site) (Corvelo et al., 2010).

eCLIP

Data for eCLIP experiments were downloaded from ENCODE Project Consortium (Consortium, 2012). Analysis of eCLIP data is the same as has been described previously (Van Nostrand et al., 2016). Fold change of eCLIP read density compared to input read density along a normalized intron was calculated using `ngs.plot` (L. Shen et al., 2014). Bud13 dependent junctions were calculated using ΔSR . In K562 cells, any junction that had a $\Delta SR > 0.1$ for all pairwise comparison of replicates was considered Bud13 dependent. In Hep G2, the ΔSR was lowered to 0.03. Peaks were called using CLIPper (Lovci et al., 2013). Peaks were deemed significant if they were >3-fold enriched and had $p\text{-value} < 10^{-5}$. Peak locations were determined using a custom python script with the HTSeq library. Enriched GO terms were determined using Seten (Budak et al., 2017).

Data and Software Availability

All next-generation sequencing data reported in this study is deposited in the Gene Expression Omnibus database under accession number GSE122543.

FIGURE LEGENDS

Figure 1: *Irf7* contains a weak intron that is retained following many forms of stimulation. (A) Histogram of mapped reads corresponding to the TNF α -induced expression of *Irf7*. The poorly spliced fourth intron is highlighted. For all read density plots, reads are histogrammed in log₁₀ scale and normalized to the maximum value across the stimulation. (B) Comparison of *Irf7* splice donor and acceptor sites in mice, rats, and humans. (C-F) Histogram representing the intron length (C), intron GC content (D), flanking exon GC content (E), or 5' splice site strength of introns of expressed in BMDMs. Red represents location of *Irf7* intron 4 (C, D, F) or upstream exon (E). Black line represents downstream exon (E). (G, H) Histogram of mapped reads corresponding to the IFN α (G) and poly(I:C) (H) induced expression of *Irf7* focused on the slow splicing fourth intron. (I) Outline of Splicing Ratio (SR) metric. (J) Splicing ratio for all introns in *Irf7* plotted against time stimulated with TNF α .

Figure 2: RAP-MS and RIP identifies BUD13 as an RNA binding protein that interacts with *Irf7* mRNA. (A) Outline of the RAP-MS procedure used to identify RNA-binding proteins on transcripts of interest. (B) TMT ratio (*Irf7*/*Actb*) for proteins identified as enriched on either *Irf7* (TMT ratio >1) or *ActB* (TMT ratio <1) transcripts. (C) RT-qPCR analysis of transcripts captured via RAP for *Irf7* (blue) and *ActB* (gold) probes. (D) RIP followed by RT-qPCR for *Irf7* and *Rpl32* in TNF α stimulated BMDMs. Shown is the relative enrichment of transcripts captured in BUD13 RIP as compared to Rabbit IgG RIP. (E) Same as (d) except stimulation with poly(I:C). Data are representative of two

independent experiments ((C-E), mean, error bars indicate s.d.). *P < 0.05, **P < 0.01 and ***P < 0.001 (t-test).

Figure 3: BUD13 knockdown leads to increased retention in the poorly splicing intron of *Irf7*. (A) Histogram of mapped reads corresponding to the TNF α -induced expression of *Irf7*. The poorly spliced fourth intron is highlighted. shBUD13 samples are shown in green. Control samples are shown in grey. (B) Δ SRs calculated for each junction in the *Irf7* transcript for all stimulated time-points. The Δ SR of intron 4 as compared to all other junctions is significant (Student's t-test, p<0.001). No other pairwise comparison is significant. (C) Splicing gel from RNA extracted from BMDMS stimulated for 30 mins. TNF α (top). Quantification of splicing gel (bottom). (D) *Irf7* FPKM fold change with respect to time stimulated. shBUD13 is shown in green, control is shown in grey. Data is representative of two independent experiments (C) and is represented as mean (error bars indicate s.d.). * denotes p < 0.05, ** denotes p < 0.01, and *** denotes p < 0.001 using a Student's t test.

Figure 4: BUD13 knockdown alters the type I interferon response. (A) Histogram of mapped reads corresponding to the TNF α -induced expression of *Irf7*. The poorly spliced fourth intron is highlighted. shBUD13 samples are shown in blue. Control samples are shown in grey. (B) Δ SRs calculated for each junction in the *Irf7* transcript for all stimulated time-points. The Δ SR of intron 4 as compared to all other junctions is significant (Student's t-test, p<0.001). No other pairwise comparison is significant. (C) *Irf7* FPKM fold change with respect to time stimulated. shBUD13 is shown in blue, control is shown

in grey. **(D)** Immunoblot analysis of IRF7 protein following 720 mins. poly(I:C) stimulation (left). Quantification relative to ActB (right). **(E)** Log₂ expression fold change (shBUD13/control) for 119 ISGs in unstimulated BMDMs (median = 0.1655). **(F)** As in **(E)** for stimulated BMDMs (720 mins poly(I:C) (median = -0.1007). Wilcoxon rank-sum between **(E)** and **(F)**, $P < .001$. **(G)** Median log₂ expression fold change (shBUD13/control) for ISGs in unstimulated BMDMs, and BMDMs stimulated with Poly(I:C) 15, 60, 240, 720, and 1440 mins. Bars represent 95% CI. (Wilcoxon rank-sum, $P < .001$, for any of the ‘early’ time-points (0, 15, 60 mins) compared to any of the ‘late’ time-points (240, 720, 1440 mins). **(H)** RT-qPCR analysis of IFN α mRNA levels in unstimulated BMDMs and BMDMs stimulated with poly(I:C) for 720 mins and 1440 mins. **(I)** Same as **(H)** for IFN β . **(J)** Nuclear fraction (top) and cytoplasmic fraction (bottom) histograms of mapped reads corresponding to the poly(I:C)-induced expression of *Irf7* (720 mins). The poorly spliced fourth intron is highlighted. shBUD13 samples are shown in blue. Control samples are shown in grey. Nuclear Δ SR = 0.35. **(K)** Nuclear and cytoplasmic RPKM for *Irf7* intron 4 from fractionated BMDMs stimulated with poly(I:C). **(L)** Cytoplasmic *Irf7* FPKM for control (grey) and shBUD13 BMDMs stimulated with poly(I:C). Data is representative of three **(D)** or four **(H,I)** independent experiments and is represented as mean (error bars indicate s.d.). * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student’s t test. Results are presented relative to those of *Rpl32* **(H,I)**.

Figure 5: Global analysis of the role of BUD13. **(A)** Ranked bar chart showing genes with a junction most affected by BUD13 knock-down in all samples during TNF α stimulation. See S7 for histograms relating to most affected junctions. **(B)** Grouped bar chart depicting

the number of genes that have a single BUD13 affected junction vs. multiple BUD13 affected junctions using three different Δ SR thresholds. **(C)** Transcripts were classified as ‘BUD13 dependent’ if they had a junction with a Δ SR. >0.15 . The \log_2 expression fold change (FPKM shBUD13/ FPKM control) for each gene represented by the transcripts in the ‘BUD13 dependent’ category as well as all other genes is shown. Median ‘increased IR’ = -0.5084. Median ‘decreased IR’ = -0.2170. (Wilcoxon rank-sum, $P < .01$). **(D)** Mean Δ SR. for junctions below the indicated threshold (x-axis) vs. mean Δ SR. for all junctions. Threshold applied for the 5’ splice site (blue) and the 3’ splice site (green). **(E)** 5’SS motif for all expressed junctions as compared to junctions that show retention upon Bud13 knockdown (Δ SR. > 0.15). **(F)** Size of intron for introns retained upon BUD13 knockdown (Δ SR. > 0.15) (blue), in introns located in the same transcript as those affected by BUD13 (green), and in introns from all expressed transcripts (orange). **(G)** Same as **(F)** for GC content. **(H)** Flanking exon GC content for exons that flank introns retained upon BUD13 knockdown (Δ SR. > 0.15) (dark green) as compared to exons that flank introns from all expressed transcripts (light green). **(I)** Distance from the branch point to the 3’ splice site for introns retained upon BUD13 knockdown (Δ SR. > 0.15) (dark blue) as compared to introns from all expressed transcripts (light blue). **(F-I)** data from BMDM TNF α stimulation. Box plots show median (center line), interquartile range (box) and tenth and ninetieth percentiles. $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$ (Mann-Whitney U -test).

Figure 6: BUD13 interacts primarily near the 3’ splice site of small, GC rich introns.

(A) eCLIP-seq read density plots in K562 cells. BUD13 density plot over all expressed junctions shown in blue (top), BUD13 density plot over BUD13 dependent junctions shown

in red (top). SF3B4 density plot over all expressed junctions shown in maroon (middle), and PRPF8 density plot over all expressed junctions is shown in green (bottom). **(B)** Same as in **(A)** but for Hep G2 cells. **(C)** BUD13 eCLIP-seq peak distribution. Peaks fell within either intronic regions, intron-exon junctions, or exonic regions. Peaks that fell within intron-exon junction were further classified as 5' junction peaks or 3' junction peaks (bottom). **(D)** Same as **(C)** but for Hep G2. **(E)** Size of all introns in expressed transcripts for the given cell line (dark blue) vs size of introns with overlapping eCLIP peak (maroon). Shown in K562 (left) and Hep G2 (right) cells. Box plots show median (center line), interquartile range (box) and tenth and ninetieth percentiles (whiskers). * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ (Mann-Whitney U -test). **(F)** Same as **(E)** for GC content. **(G)** GO terms (biological process) enriched among BUD13 eCLIP peaks in K562(dark blue) and Hep G2 (maroon) cells.

Figure 7: BUD13 knockdown alters the BMDM response to VSV. **(A)** RT-qPCR analysis of *Irf7* mRNA levels in infected control or shBUD13 BMDMs stimulated with VSV (MOI 5) across 24 hours. **(B)** Same as in **(A)** except stimulated at an MOI of 10. Results are presented relative to those of *Rpl32*. **(C)** PFU/mL for viral supernatant from infected shBUD13 (blue), control (red), shBUD13 with *Irf7* overexpression (yellow), or control with *Irf7* overexpression (maroon) BMDMs. Data is representative of two **(A, B)** or three independent experiments **(C)** and is shown as mean (error bars indicate s.d.). * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student's t test.

References:

- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Bao, P., Will, C.L., Urlaub, H., Boon, K.-L., Lührmann, R., 2018. The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated Bact complex. *Genes Dev.*
- Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., Stark, H., 2017. Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* 170, 701–713.
- Blanco, M.R., Martin, J.S., Kahlscheuer, M.L., Krishnan, R., Abelson, J., Laederach, A., Walter, N.G., 2015. Single Molecule Cluster Analysis dissects splicing pathway conformational dynamics. *Nat. Methods* 12, 1077.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., Blencowe, B.J., 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786.
- Budak, G., Srivastava, R., Janga, S.C., 2017. Seten: a tool for systematic identification and comparison of processes, phenotypes, and diseases associated with RNA-binding proteins from condition-specific CLIP-seq profiles. *RNA* 23, 836–846.
- Ciancanelli, M.J., Huang, S.X., Luthra, P., Garner, H., Itan, Y., Volpi, S., Lafaille, F.G., Trouillet, C., Schmolke, M., Albrecht, R.A., 2015. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science* 348, 448–453.
- Consortium, E.P., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Corvelo, A., Hallegger, M., Smith, C.W., Eyras, E., 2010. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput. Biol.* 6, e1001016.
- Dvinge, H., Bradley, R.K., 2015. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* 7, 45.
- Dziembowski, A., Ventura, A.-P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprévote, O., Séraphin, B., 2004. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J.* 23, 4847–4856.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., Lührmann, R., 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* 36, 593–608.

- Fernandez, J.P., Moreno-Mateos, M.A., Gohr, A., Miao, L., Chan, S.H., Irimia, M., Giraldez, A.J., 2018. RES complex is associated with intron definition and required for zebrafish early embryogenesis. *PLoS Genet.* 14, e1007473.
- Hao, S., Baltimore, D., 2013. RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci.* 110, 11934–11939.
- Harley, J.B., Alarcón-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., Genetics (SLEGEN, I.C. for S.L.E., 2008. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.* 40, 204.
- Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., Shimada, N., Ohba, Y., Takaoka, A., Yoshida, N., Taniguchi, T., 2005. IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* 434, 772–777. <https://doi.org/10.1038/nature03464>
- Hoskins, A.A., Moore, M.J., 2012. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.* 37, 179–188. <https://doi.org/10.1016/j.tibs.2012.02.009>
- Jacob, A.G., Smith, C.W., 2017. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 1–15.
- Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V., Kim, S.K., 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* 98, 218–223.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kontoyiannis, D., Pasparakis, M., Pizarro, T.T., Cominelli, F., Kollias, G., 1999. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* 10, 387–398.
- Liu, J.M., Sweredoski, M.J., Hess, S., 2016. Improved 6-Plex Tandem Mass Tags Quantification Throughput Using a Linear Ion Trap–High-Energy Collision Induced Dissociation MS3 Scan. *Anal. Chem.* 88, 7471–7475.
- Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S., 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* 20, 1434–1442.
- Lund, J.M., Alexopoulou, L., Sato, A., Karow, M., Adams, N.C., Gale, N.W., Iwasaki, A., Flavell, R.A., 2004. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5598–5603.

- McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236.
- Mostafavi, S., Yoshida, H., Moodley, D., LeBoité, H., Rothamel, K., Raj, T., Ye, C.J., Chevrier, N., Zhang, S.-Y., Feng, T., 2016. Parsing the interferon transcriptional network and its disease associations. *Cell* 164, 564–578.
- Na, I., Meng, F., Kurgan, L., Uversky, V.N., 2016. Autophagy-related intrinsically disordered proteins in intra-nuclear compartments. *Mol. Biosyst.* 12, 2798–2817.
- Ni, L., Snyder, M., 2001. A Genomic Study of the Bipolar Bud Site Selection Pattern in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 12, 2147–2170. <https://doi.org/10.1091/mbc.12.7.2147>
- Ohr, T., Prior, M., Dannenberg, J., Odenwälder, P., Dybkov, O., Rasche, N., Schmitzová, J., Gregor, I., Fabrizio, P., Enderlein, J., 2012. Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18, 1244–1256.
- Plaschka, C., Lin, P.-C., Nagai, K., 2017. Structure of a pre-catalytic spliceosome. *Nature* 546, 617. <https://doi.org/10.1038/nature22799>
- Sayani, S., Chanfreau, G.F., 2012. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in *Saccharomyces cerevisiae*. *Rna* 18, 1563–1572.
- Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M.-C., Shini, S., Lieschke, G.J., Wong, J.J., Rasko, J.E., 2017. Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* 18, 216.
- Schneider, C., Agafonov, D.E., Schmitzová, J., Hartmuth, K., Fabrizio, P., Lührmann, R., 2015. Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. *PLOS Genet.* 11, e1005539. <https://doi.org/10.1371/journal.pgen.1005539>
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J.J., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J.Z., Park, H., Regev, A., 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236. <https://doi.org/10.1038/nature12172>
- Shen, L., Shao, N., Liu, X., Nestler, E., 2014. ngs. plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284.

- Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.* 111, E5593–E5601.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514.
- Wong, J.J.-L., Au, A.Y., Ritchie, W., Rasko, J.E., 2016. Intron retention in mRNA: No longer nonsense. *Bioessays* 38, 41–49.
- Wong, J.J.-L., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., 2013. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154, 583–595.
- Yap, K., Lim, Z.Q., Khandelia, P., Friedman, B., Makeyev, E.V., 2012. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.* 26, 1209–1223.
- Yeo, G., Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394.
- Zhou, Y., Chen, C., Johansson, M.J., 2013. The pre-mRNA retention and splicing complex controls tRNA maturation by promoting TAN1 expression. *Nucleic Acids Res.* 41, 5669–5678.
- Zhou, Y., Johansson, M.J., 2017. The pre-mRNA retention and splicing complex controls expression of the Mediator subunit Med20. *RNA Biol.* 1–7.

Supplemental Figures

Figure S1: Splicing Ratios across all junctions in *Irf7*. Related to Figure 1. Splicing ratios calculated for all junctions in the most abundant transcript of *Irf7*. Color represents time-point indicated in legend. **(A)** Poly(I:C) **(B)** IFN α .

Figure S2: shBUD13 knocks down BUD13 protein and mRNA. Related to Figure 3. **(A)** Immunoblot analysis of BUD13 in BMDMs infected with control or shBUD13. ActB serves as loading control. **(B)** qRT-PCR analysis of *Bud13* mRNA in BMDMs infected with control or shBUD13. **(C)** Immunoblot analysis of IRF7 protein following 120 mins. TNF α stimulation. **(D-F)** FPKM fold change with respect to time stimulated **(C)** Zfp36, **(D)** I κ B ϵ , and **(E)** CD83. shBUD13 is shown in green, control is shown in grey. Data is representative of two individual experiments **(A, B)** and is shown as mean (error bars indicate s.d.) **(B)**. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

Figure S3: *Irf7* Intron 4 is the most BUD13 knockdown affected junction of all ISGs. Related to Figure 4. **(A)** Normalized levels of spliced vs. unspliced intron 4 in control and shBUD13 BMDMs as measure through quantitative RT-qPCR. **(B)** Normalized FPKM expression levels in shBUD13 and control samples at 720 mins poly(I:C) stimulation for select ISGs). **(C)** Δ SR was calculated at 720 mins of poly(I:C) stimulation for each ISG junction that passed the transcript and local read count threshold (see methods). Mean Δ SR = 0.002, Median Δ SR = 0. **(D)** Immunoblot analysis of *Irf7* protein after nuclear fractionation from BMDMs left untreated (UT) or treated with poly(I:C) (PIC) or CpG for 12h. Lamin B1 serves as loading control.

Figure S4: BUD13 knockdown alters the type I interferon response in response to CpG.

(A) Histogram of mapped reads corresponding to the CpG-induced expression of *Irf7*. The poorly spliced fourth intron is highlighted. shBUD13 samples are shown in pink. Control samples are shown in grey. (B) *Irf7* FPKM fold change with respect to time stimulated. shBUD13 is shown in pink, control is shown in grey. (C) RT-qPCR analysis of IFN α mRNA levels in unstimulated BMDMs and BMDMs stimulated with CpG for 720 and 1440 mins. (D) Log₂ expression fold change (shBUD13/Control) for 119 ISGs (selected based on upregulation in response to IFN α) in unstimulated BMDMs (median = -0.2442). (E) As in (D) for stimulated BMDMs (720 mins CpG (median = -0.4776). Wilcoxon rank-sum between (D) and (E), $P < .001$. (F) Ratio of cytoplasmic FPKM levels to cytoplasmic and nuclear FPKM levels for transcripts that are primarily nuclear (*Malat1*, *Neat1*, *Xist*, *U2*; left), and primarily cytoplasmic (*Rpl32*, *Rps5*, *Actb*, *Rpl5*; right) (BMDMs – 720 mins poly(I:C) stimulation). (G) Nuclear Δ SR calculated for each junction in the *Irf7* transcript. Unless indicated, comparison of intron 4 Δ SR to any other junction is significant (Student's t-test, $p < 0.001$). No other pairwise comparison is significant. Data is represented as mean (error bars indicate s.d.). * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student's t test. Results are presented relative to those of *Rpl32*.

Figure S5: Supplemental global analysis of BUD13. Related to Figure 5 and 6.

(A) Size of intron for introns retained upon BUD13 knockdown (Δ SR. > 0.15) (blue), in introns located in the same transcript as those affected by BUD13 (green), and in introns from all expressed transcripts (orange). (B) Same as (A) for GC content. (C) Flanking exon GC content for exons that flank introns retained upon BUD13 knockdown (Δ SR. > 0.15) (dark

green) as compared to exons that flank introns from all expressed transcripts (light green). **(D)** Distance from the branch point to the 3' splice site for introns retained upon BUD13 knockdown ($\Delta SR. > 0.15$) (dark blue) as compared to introns from all expressed transcripts (light blue). **(A-D)** data from BMDM poly(I:C) stimulation. **(E)** Branch point score for introns retained upon Bud13 knockdown ($\Delta SR. > 0.15$) (beige) as compared to introns from all expressed transcripts (dark brown) in TNF α stimulated BMDMs. **(F)** Same as **(E)** but for poly(I:C) stimulated BMDMs. **(G)** Significant number of alternative splicing events across the TNF α time-course as calculated by rMATs. **(H)** Ranked bar chart showing genes with a junction most affected by BUD13 knock-down in all samples during PIC stimulation. **(I)** Box plot showing the number of retention events across replicates at the indicated ΔSR . Box plots show median (center line), interquartile range (box) and tenth and ninetieth percentiles. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ (Mann-Whitney *U*-test).

Figure S6: BUD13 knockdown alters the BMDM infection via VSV. Related to Figure 7. **(A)** FSC/SSC plot showing the gating of live BMDMs in an uninfected control sample and the subsequent threshold used to calculate infectivity. **(B)** Same as in **(A)** but for a control sample infected with VSV-GFP for 12 hours. **(C)** Percent of live cells infected with VSV-GFP (MOI 10) in both control and shBUD13 BMDMs across a 24-hour time-course. **(D)** Percent of live cells infected in both control and shBUD13 BMDMs at 12 hours across a range of VSV-GFP MOIs. Data is representative of three **(C, D)** independent experiments and is represented as mean (error bars indicate s.d.). * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student's *t* test.

Figure S7: Knockdown of other RES complex proteins. Related to Figure 4. (A) Bargraph indicating knockdown efficiency for RBMX2 shRNA (green) and SNIP1 shRNA (orange) as compared to BUD13 shRNA (blue). (B) Δ SRs calculated for each junction in the *Irf7* transcript for shBUD13 (blue), shRBMX2 (green) and shSNIP1 (orange). The shBUD13 data is from figure 4 and is shown for perspective. Comparison of intron Δ SR at intron 4 to all other junctions is significant (Student's t-test, $p < 0.001$). No other pairwise comparison is significant for shBUD13. No pairwise comparison is significant for other knockdown constructs. (C) Box plot showing the number of retention events across replicates at the indicated Δ SR for shBUD13 (blue), shRBMX2 (green), and shSNIP1 (orange).

Supplemental Table S1: shRNA Sequences. Related to STAR Methods.

Supplemental Table S2: List of *Irf7*/ActB Associated Proteins Detected with RAP-MS. Related to Figure 2

Supplemental Table S3: Δ SR for ISGs at 720 mins. Poly(I:C). Related to Figure 4.

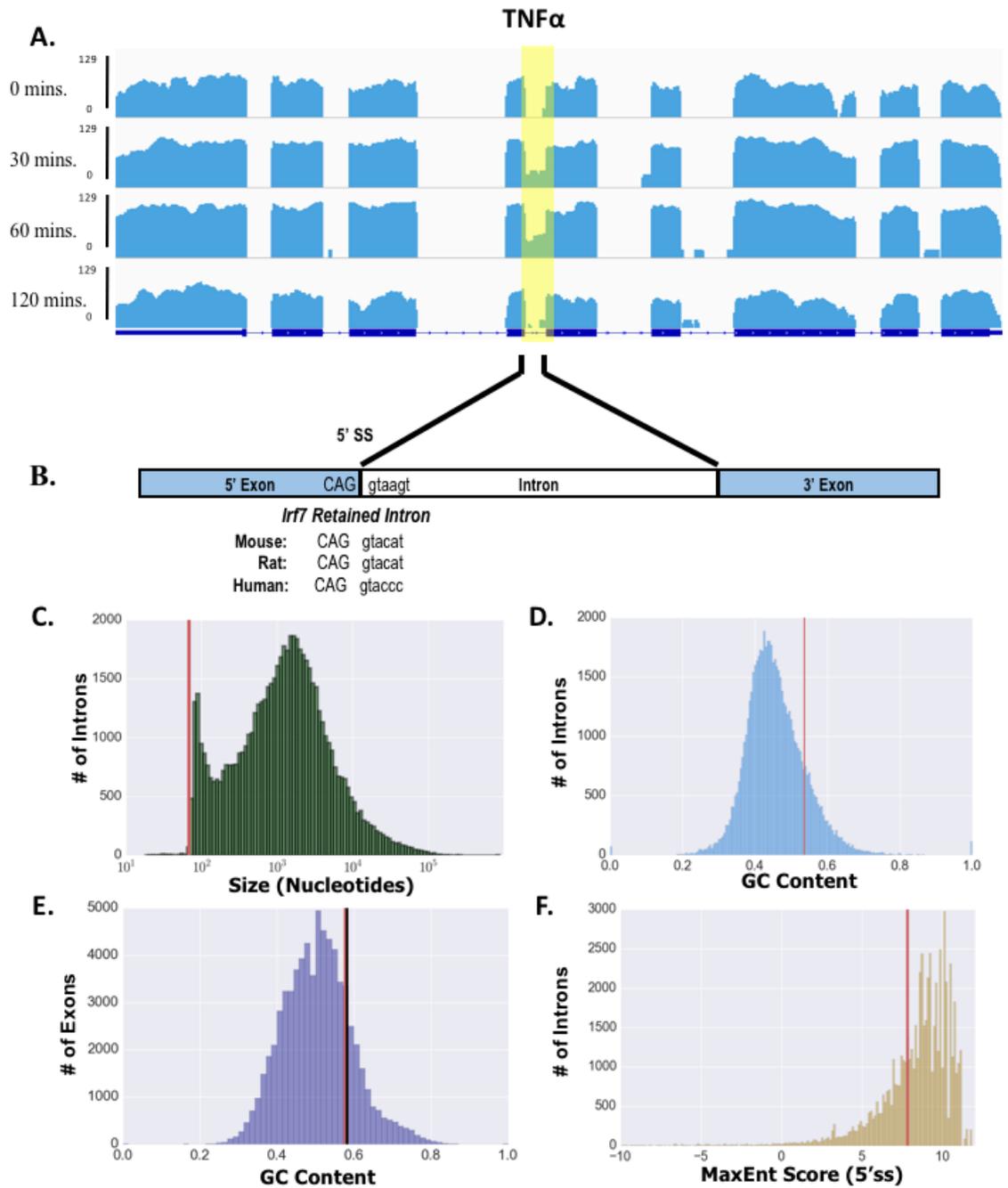


Figure 1-1. *Irf7* contains a weak intron that is retained following many forms of stimulation.

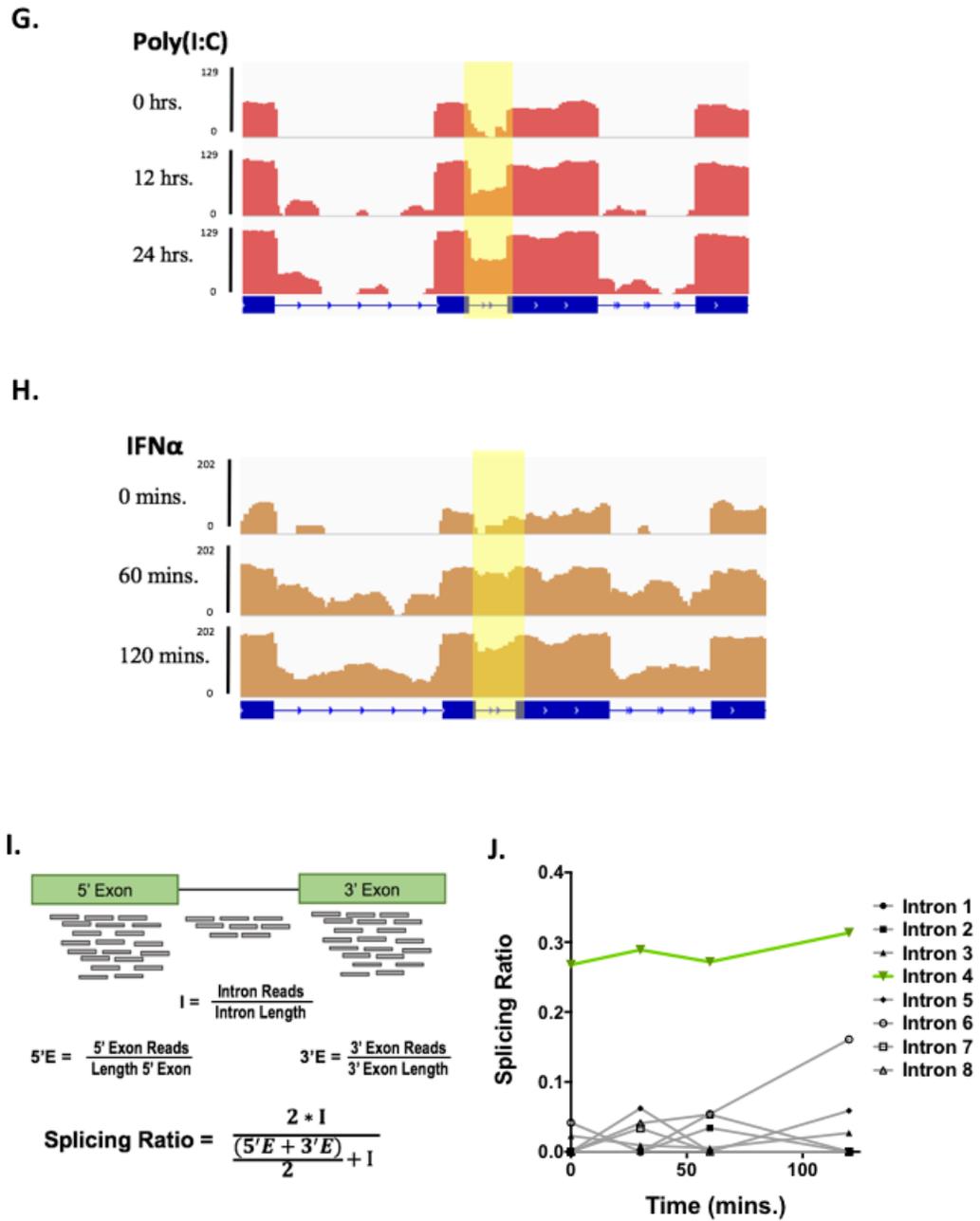


Figure 1-2: *Irf7* contains a weak intron that is retained following many forms of stimulation.

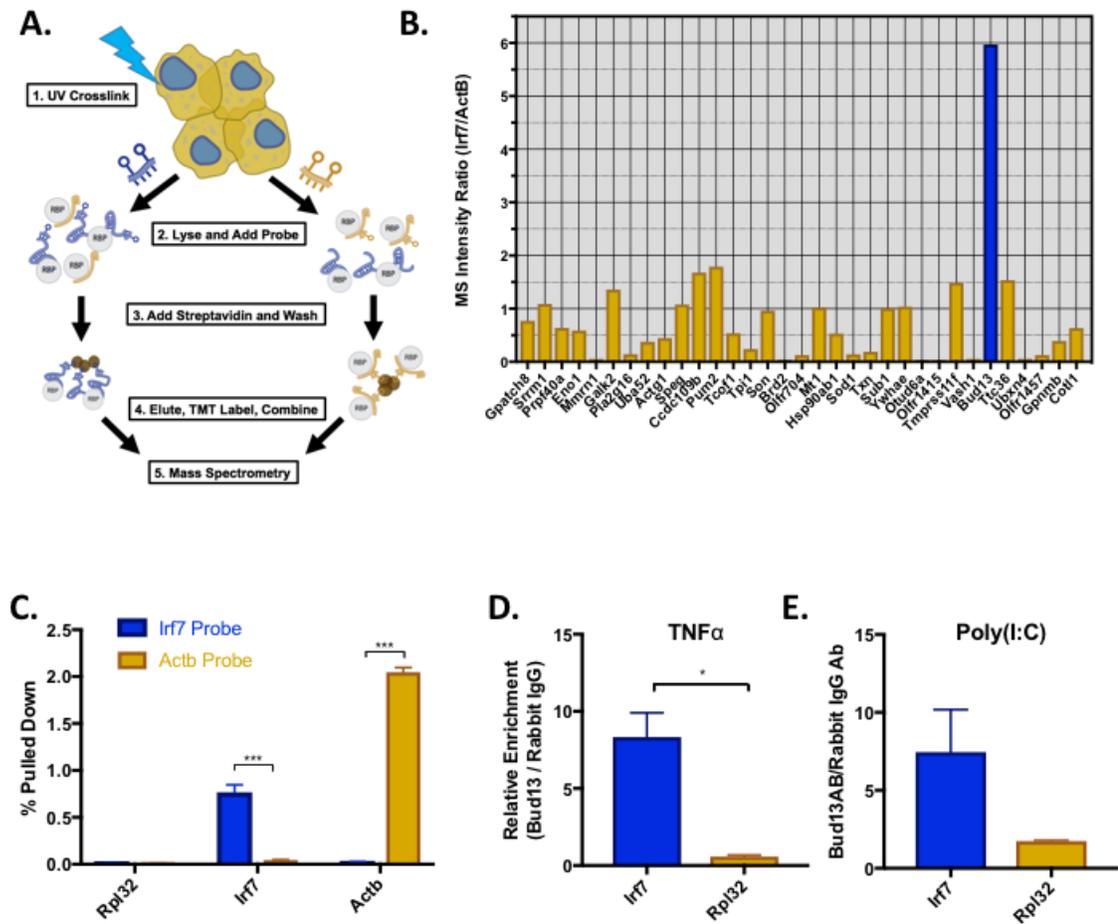


Figure 2: RAP-MS and RIP identify BUD13 as an RNA binding protein that interacts with *Irf7* mRNA

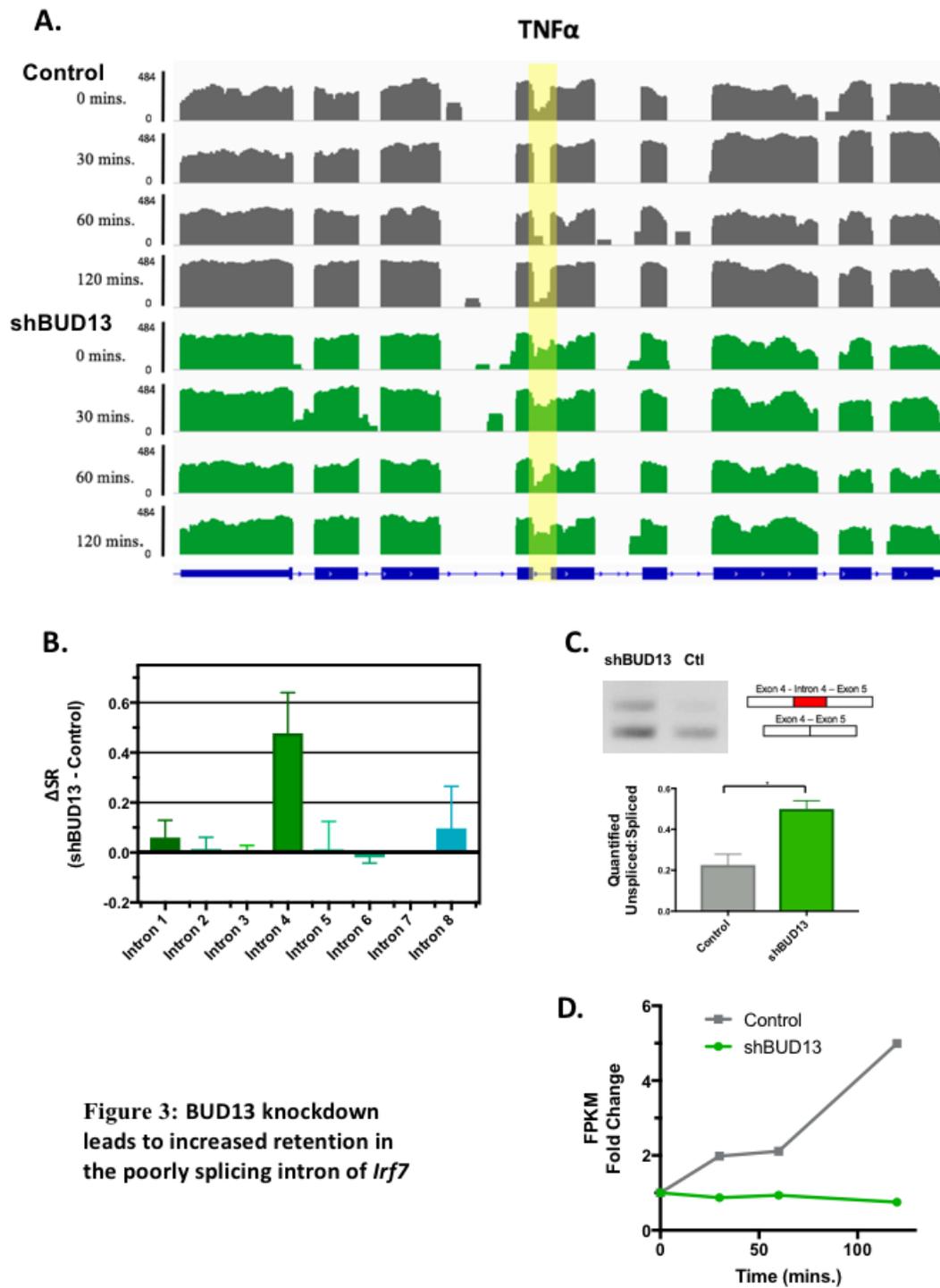


Figure 3: BUD13 knockdown leads to increased retention in the poorly splicing intron of *Irf7*

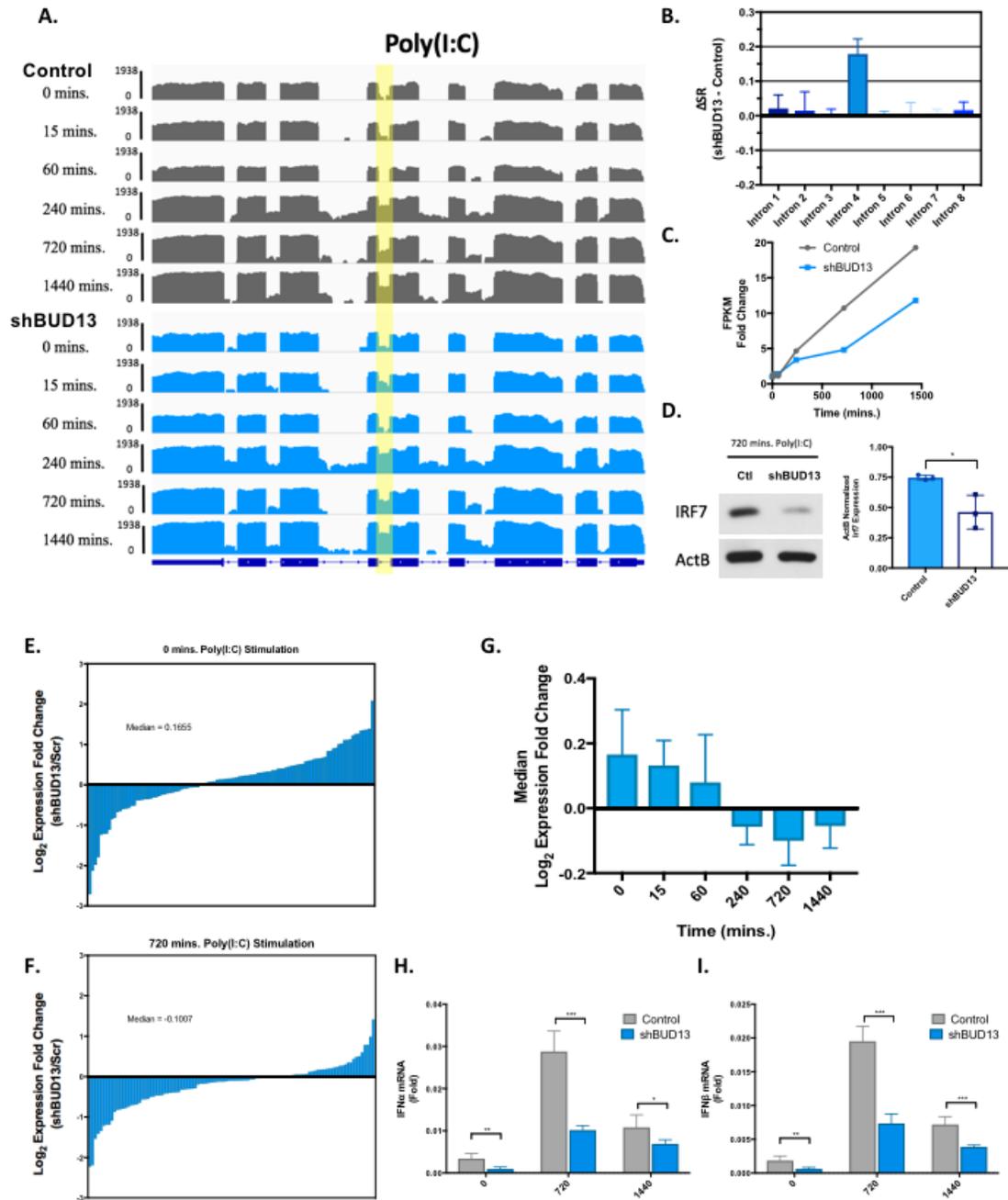


Figure 4-1: BUD13 knockdown alters the type I interferon response

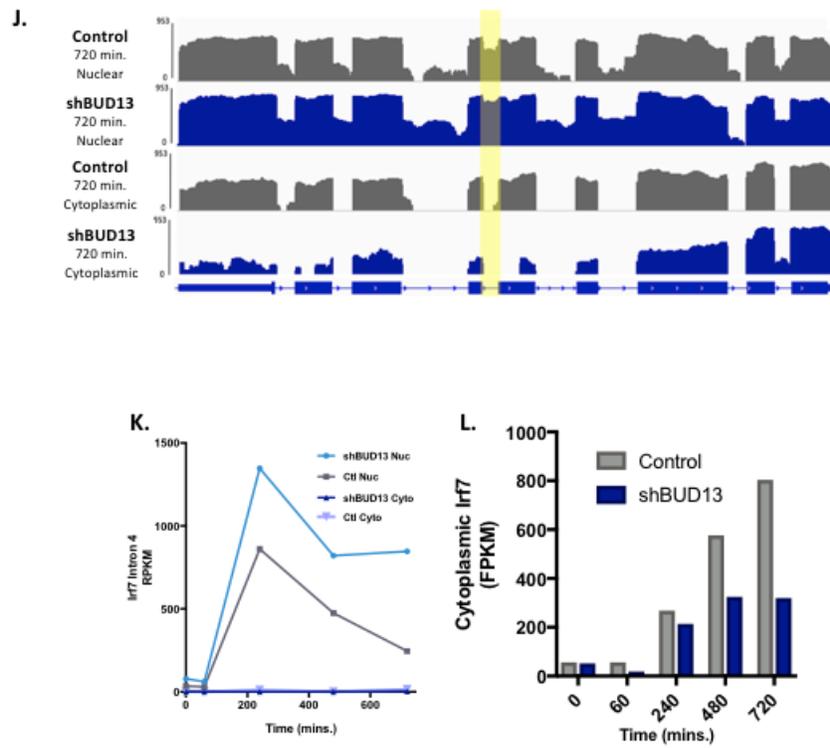


Figure 4-2: BUD13 knockdown alters the type I interferon response

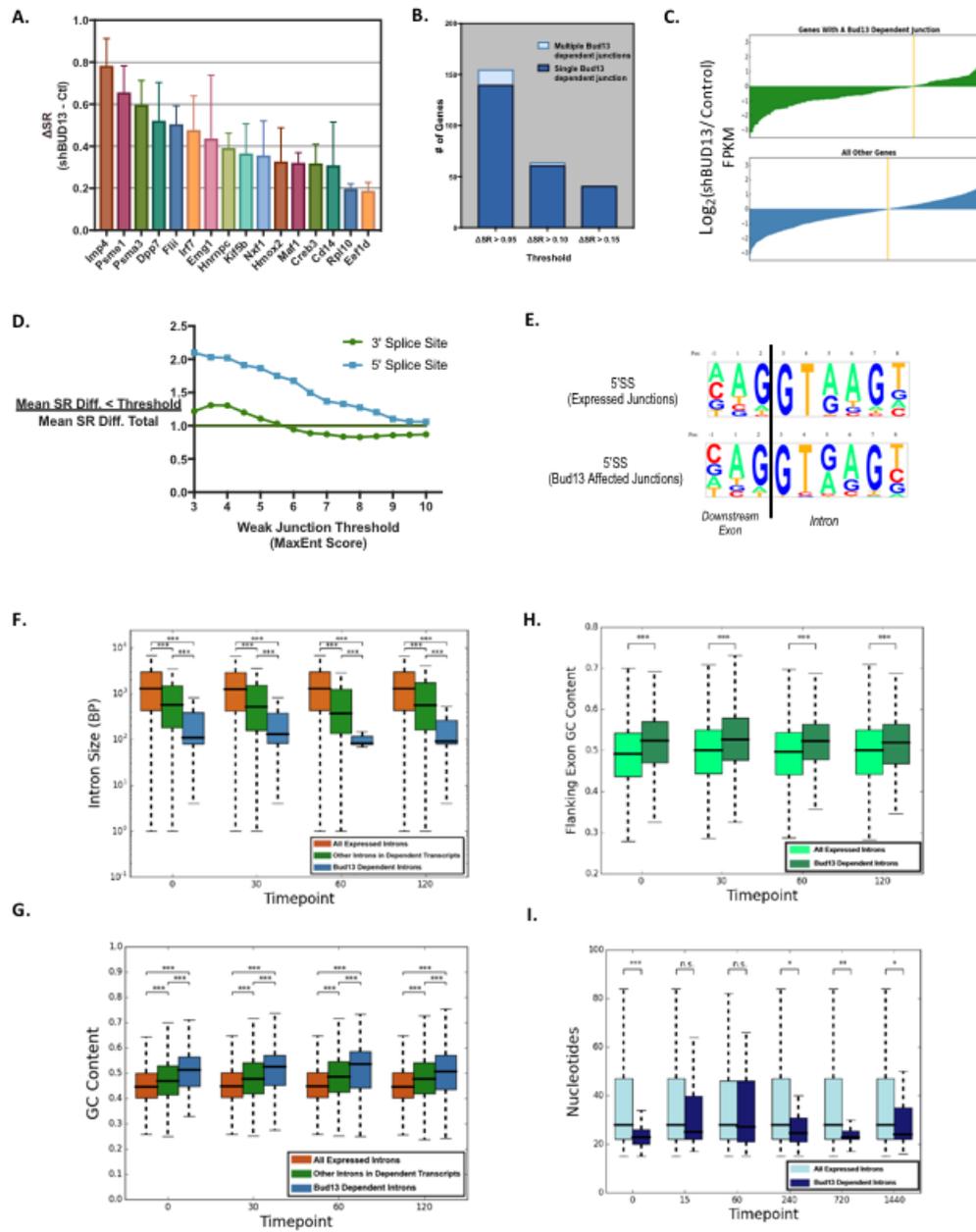


Figure 5: Global analysis of the role of BUD13.

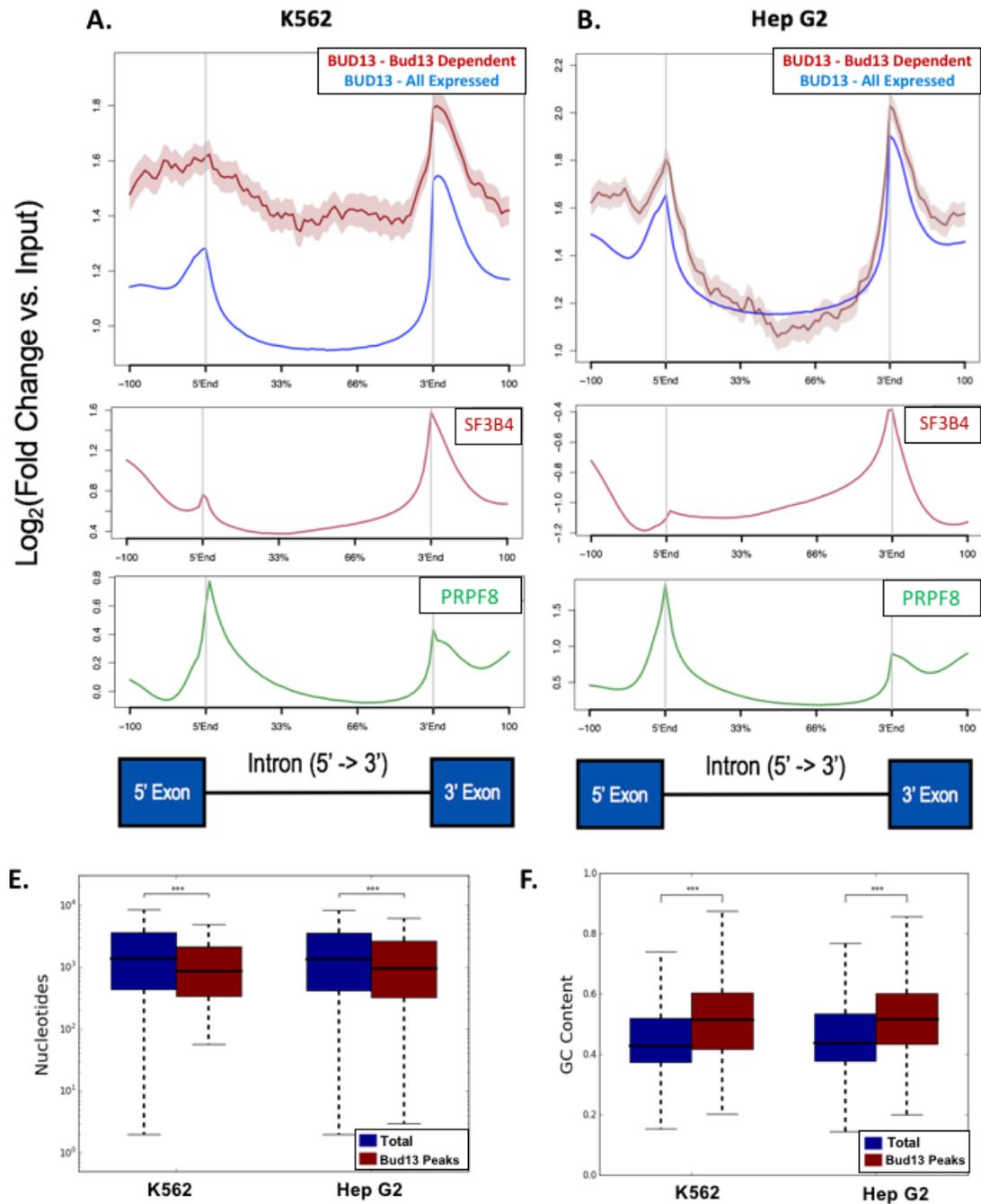


Figure 6-1: BUD13 interacts with BUD13 dependent junctions near the 3' splice site.

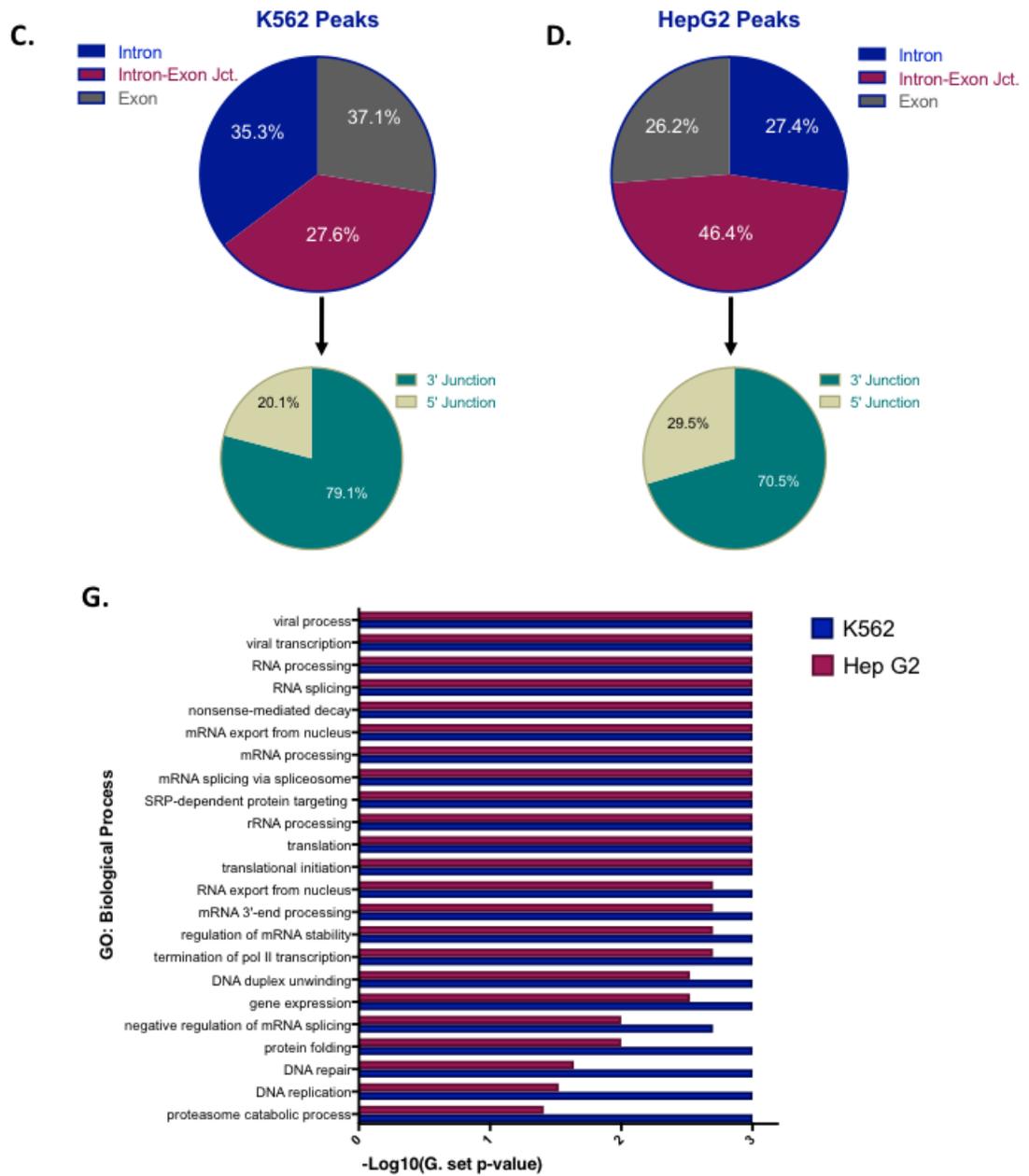


Figure 6-2: BUD13 interacts with BUD13 dependent junctions near the 3' splice site.

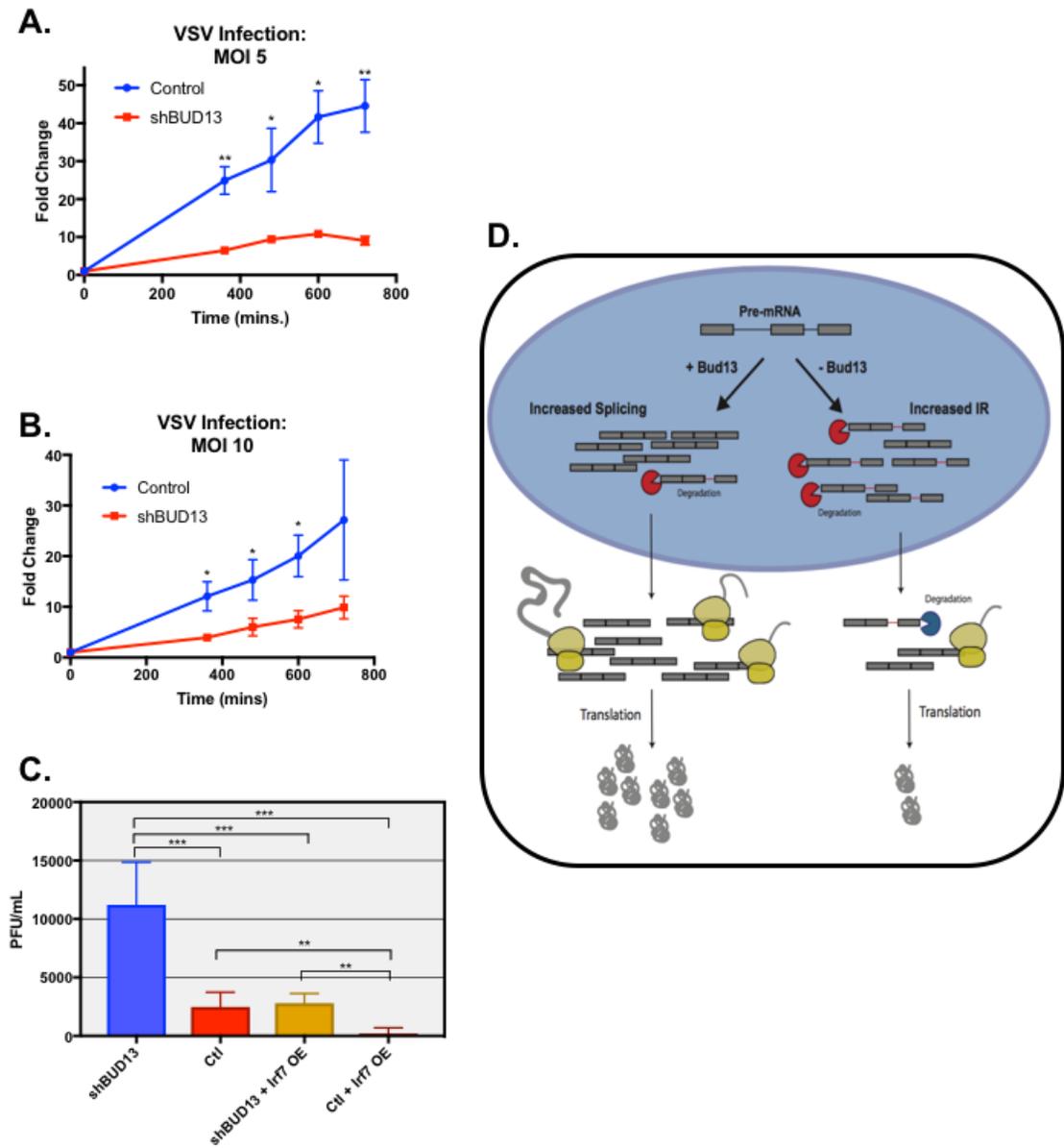


Figure 7: BUD13 knockdown alters the BMDM response to VSV.

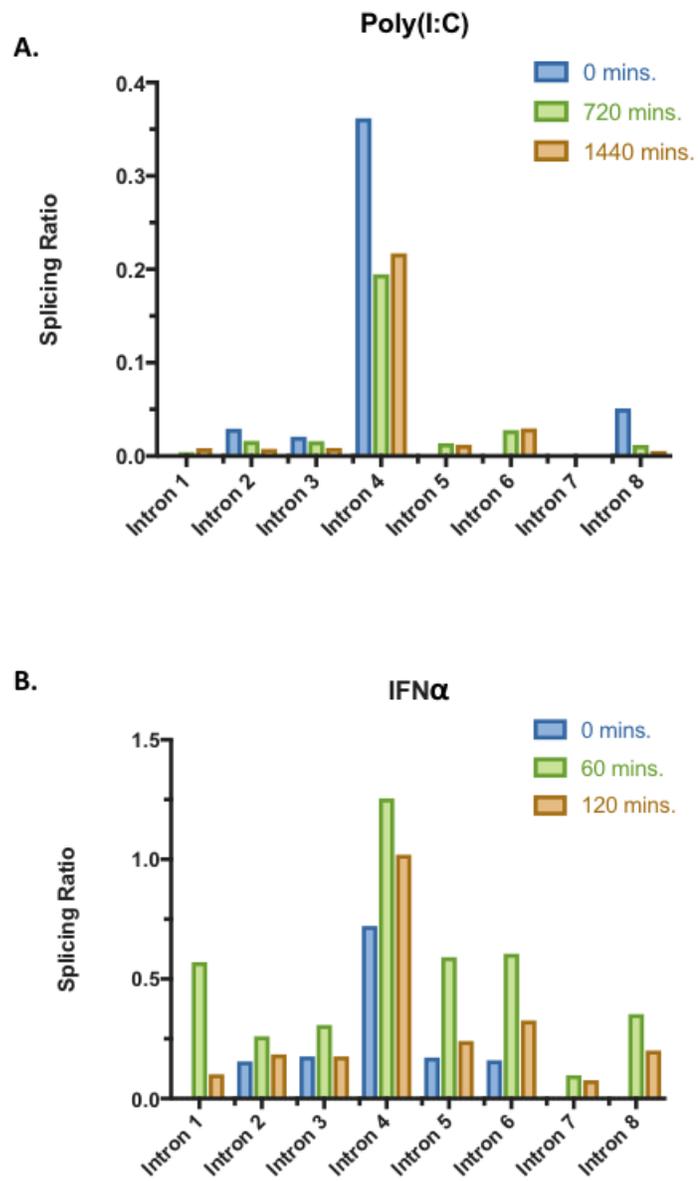


Figure S1: Splicing Ratios across all junctions in *Irf7*. Related to Figure 1.

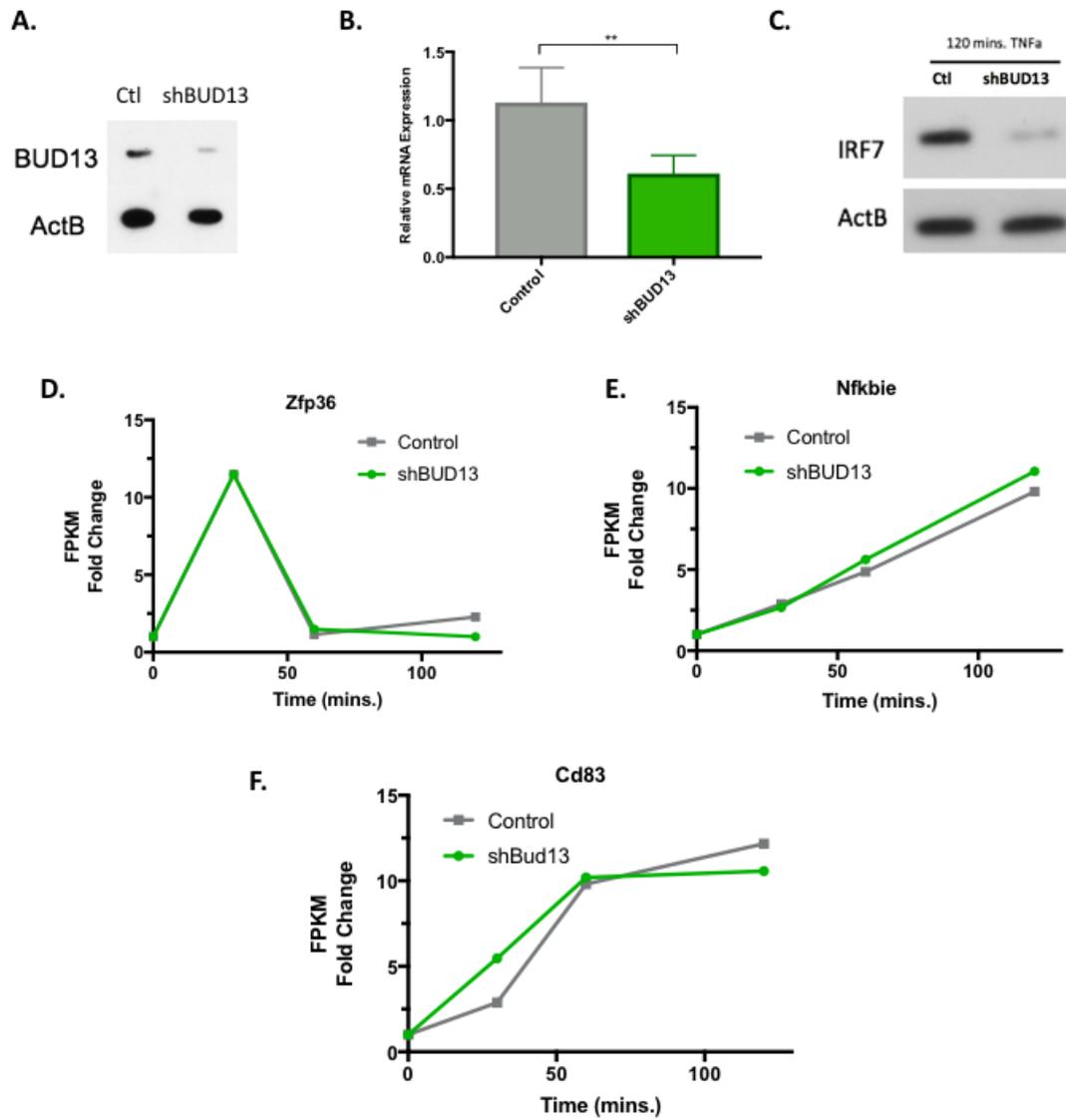


Figure S2: shBUD13 knocks down BUD13 protein and mRNA. Related to Figure 3.

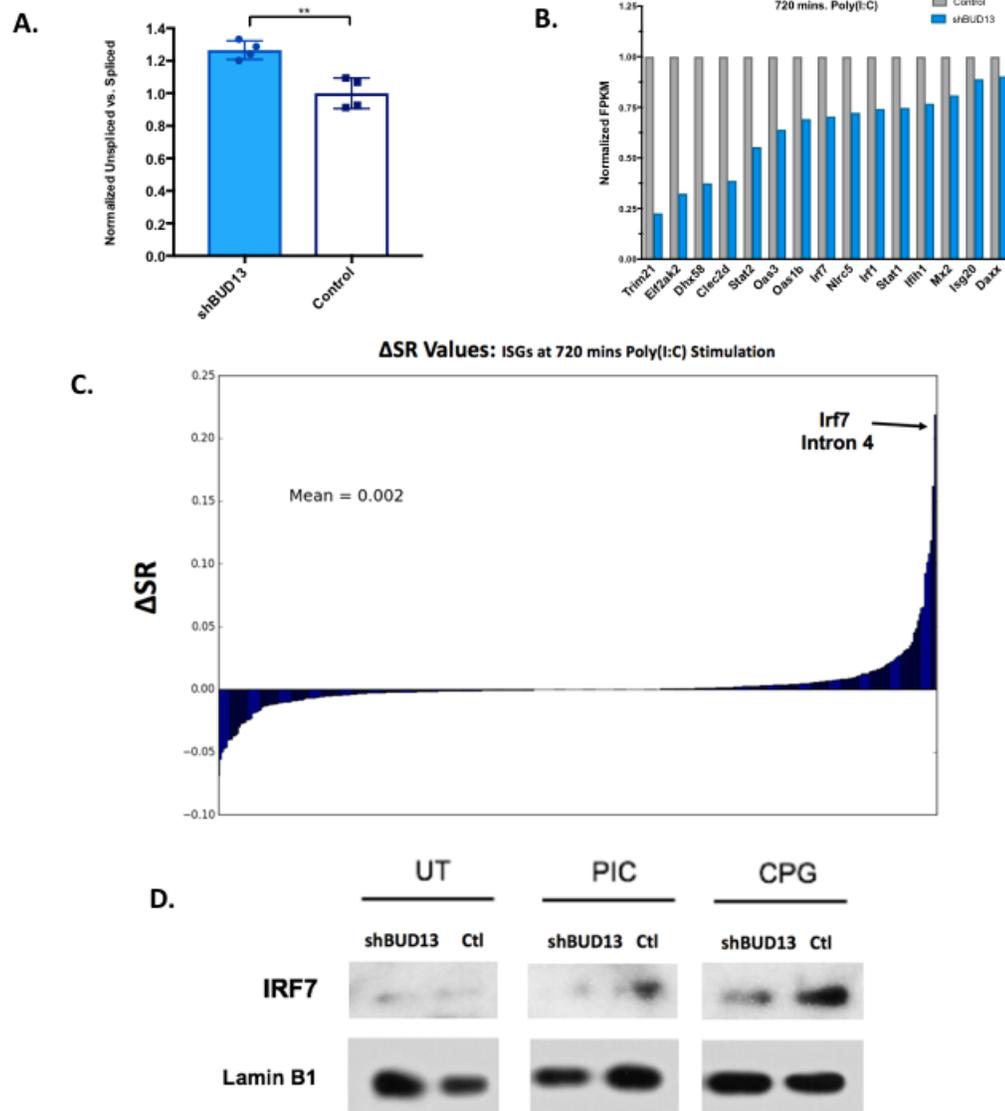


Figure S3: *Irf7* Intron 4 is the most BUD13 knockdown affected junction of all ISGs. Related to Figure 4.

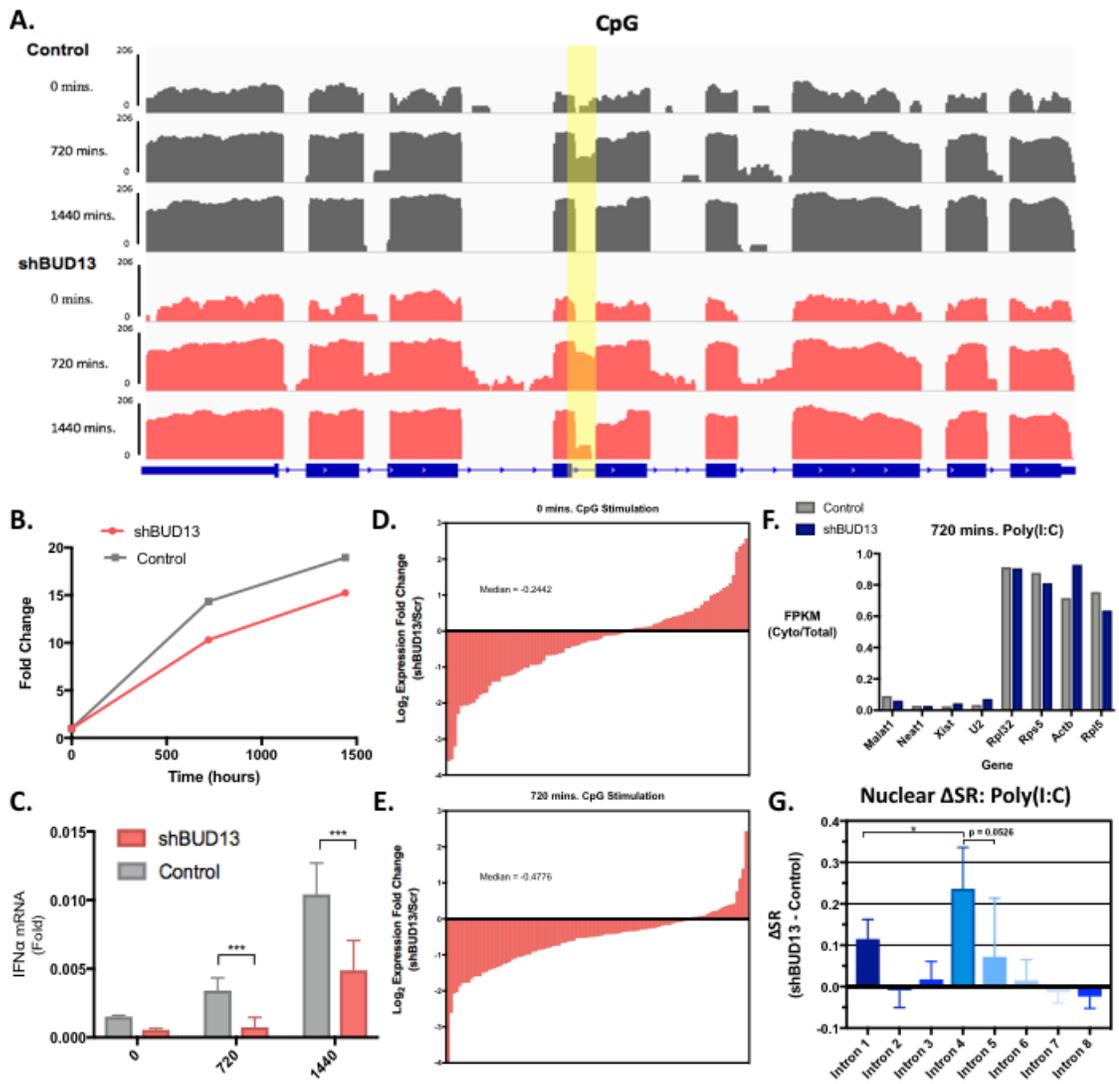


Figure S4: BUD13 knockdown alters the type I interferon response in response to CpG. Related to Figure 4.

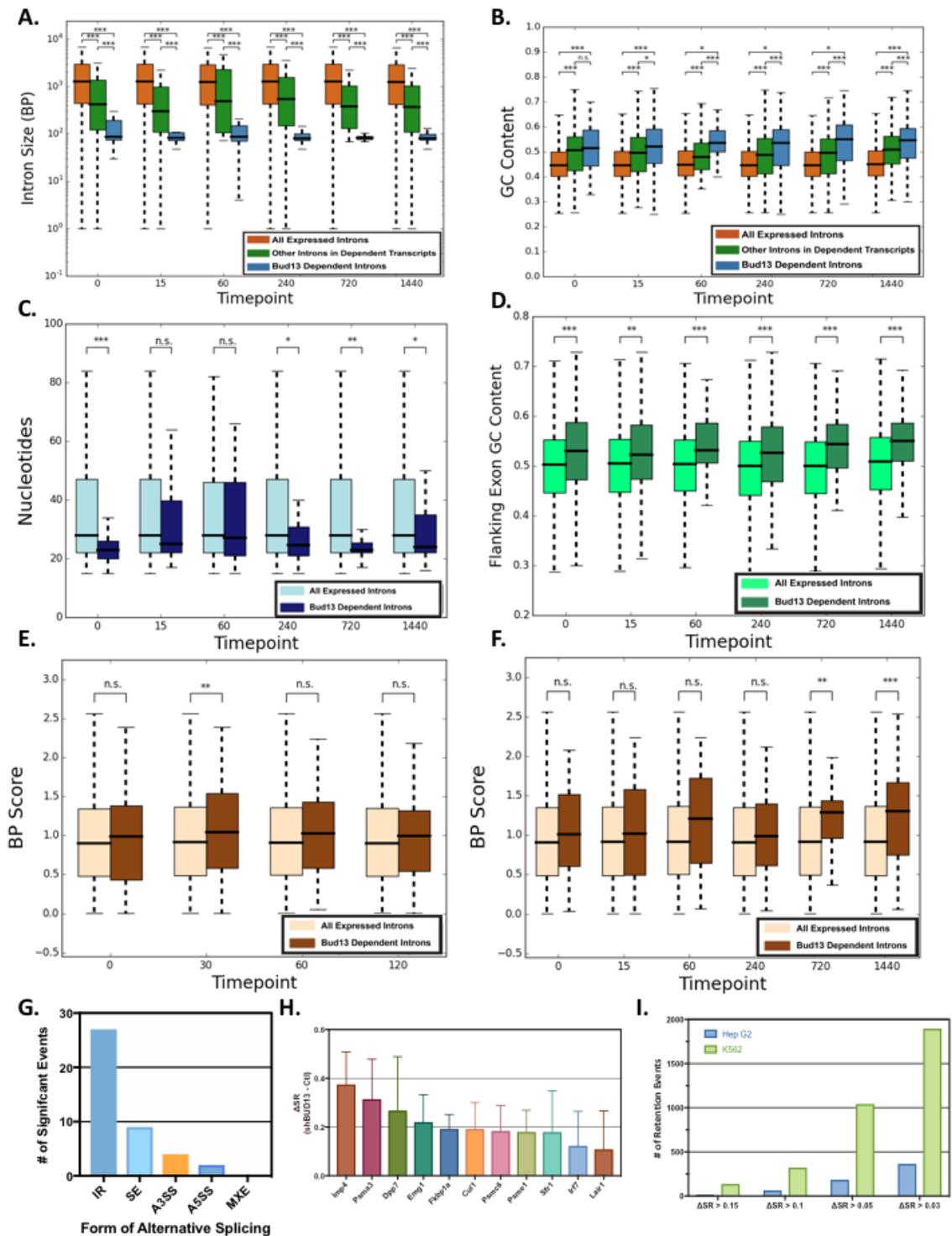


Figure S5: Supplemental global analysis of BUD13. Related to Figure 5 and 6.

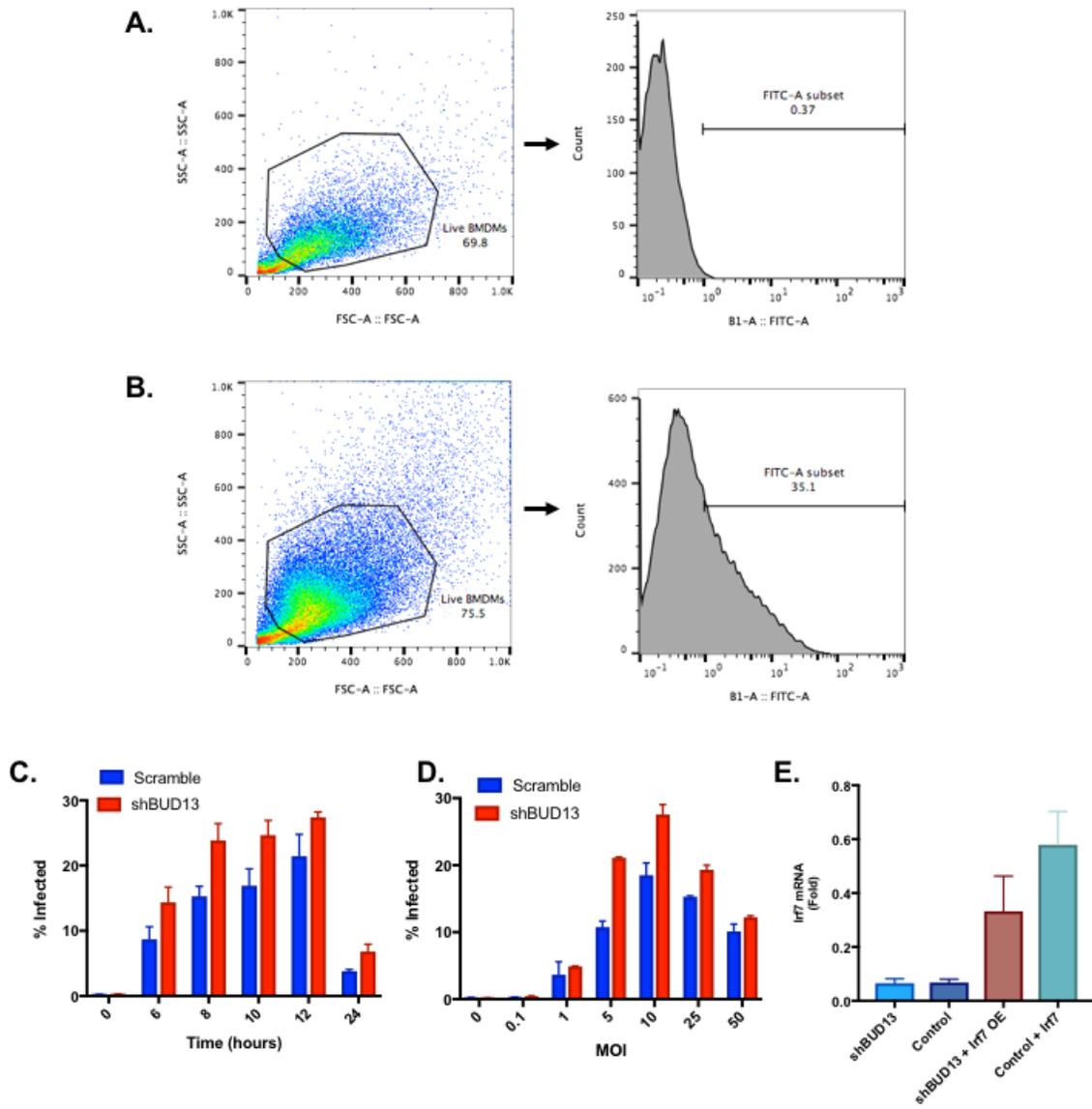


Figure S6: BUD13 knockdown alters the BMDM infection via VSV. Related to Figure 7.

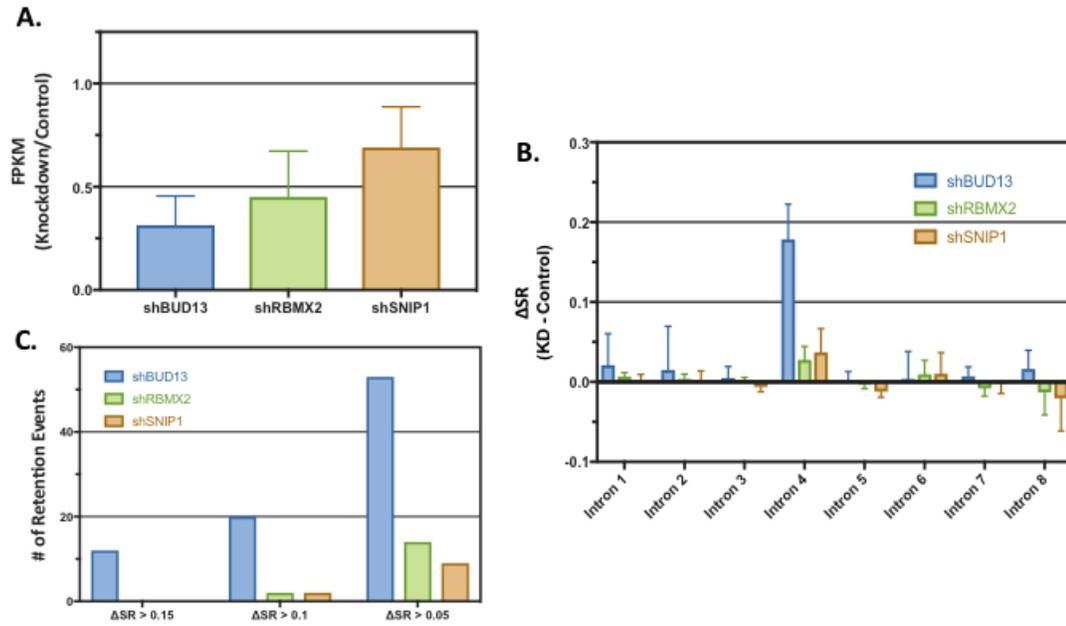


Figure S7: Knockdown of other RES complex proteins. Related to Figure 4.

*Chapter 4***Alternative Splicing Coupled with NMD Acts to Mitigate OAS1 Antiviral Activity**

Manuscript in Preparation: Frankiw, L. *et al.* Alternative Splicing Coupled with NMD Acts to Mitigate OAS1 Antiviral Activity.

Abstract

At the heart of an innate immune response lies a tightly regulated gene expression program. This precise regulation is crucial because small changes can shift the balance from protective to destructive immunity. Here we identify a frequently used alternative splice site in the gene oligoadenylate synthetase 1g (*Oas1g*), a key component of the 2-5A antiviral system. Usage of this splice site leads to the generation of a transcript subject to decay, and removal of the site leads to increased expression of *Oas1g* and an improved antiviral response. However, removal of the splice site also leads to an increase in apoptotic cell death, suggesting this splicing event exists as a compromise between the pathogen protective benefits and collateral damage associated with OAS1g activity. Across the innate immune response, we show similar alternative splicing events coupled with decay are widespread and represent a previously underappreciated mechanism of gene expression regulation in innate immunity.

Introduction

Central to an inflammatory response is a robust and coordinated gene expression program. Precise regulation of this gene expression program is essential because small alterations can shift the balance from protective to destructive immunity¹. While transcription and protein turnover are the best-examined areas of gene expression regulation²⁻⁵, a variety of post-transcriptional mechanisms have emerged that play a role in the fine-tuning of an inflammatory response. Well-studied examples include mRNA stabilization⁶, mRNA deadenylation⁷, and microRNA regulation⁸.

More recently, the wealth of transcriptomic data generated over the last decade has shed light on the widespread nature of alternative mRNA splicing of mammalian genes. While most mammalian genes exhibit alternative splicing^{9,10}, not all of the produced transcripts encode functional proteins. It is true that alternative splicing can act to increase proteomic diversity; however, it can also generate unproductive isoforms that incorporate a premature termination codon (PTC), thus subjecting the transcript to either cytoplasmic NMD decay¹¹. Coupling of alternative splicing to NMD decay (AS-NMD) provides cells with a mode of downregulation of expression of a given gene. It has been estimated that 10-30% of mammalian genes may be regulated post-transcriptionally, potentially in a context-specific manner, through AS-NMD¹²⁻¹⁵.

While several AS-NMD events have been identified and have been shown to play an important role in a variety of biological processes, most notably the autoregulation of splicing factor genes^{15,16}, little is known with respect to the role of AS-NMD during the finely-tuned inflammatory response. Here we identify a frequently used unproductive splicing event in oligoadenylate synthetase 1g (*Oas1g*), an important murine anti-viral response factor. Upon

binding viral dsRNA, OAS1g acts to convert ATP into 2-5 linked oligoadenylates (2-5A), which in turn activate RNase L. Although humans have a single *Oas1* gene, in mice the *Oas1* gene locus underwent a series of duplication events leading to the existence of eight *Oas1* paralogues. However, only OAS1a and OAS1g have been shown to be enzymatically active^{17,18}. Activated RNase L degrades viral RNA, in turn inhibiting viral replication and propagation¹⁹. Removal of the *Oas1g* alternative splice site in a murine macrophage cell line led to increased expression of *Oas1g*, both in stimulated and unstimulated conditions. Further, this increased expression of *Oas1g* improved the ability of macrophages that lack the unproductive splice site to withstand infection with Encephalomyocarditis virus (EMCV). However, removal of the *Oas1g* alternative splice site led to an increase in apoptotic cell death in uninfected cells, a finding consistent with the idea that activation of the 2-5A system can be detrimental to host fitness²⁰⁻²². Beyond *Oas1g*, AS-NMD events were found in a number of other important transcripts involved with the innate immune response. Thus, evolution of splice sites in such transcripts, with a consequent dampening of gene output, is a means of mitigating what might otherwise be an unchecked or inappropriately scaled response.

Results

***Oas1g* has a Frequent AS-NMD Event**

AS events have the potential to generate both productive isoforms coding for functional proteins as well as unproductive isoforms subject to degradation (Figure 1A). The latter allows for the use of AS as a post-transcriptional mechanism of gene-expression regulation. To investigate the extent to which unproductive splicing acts as a post-transcriptional

regulator of gene expression during inflammation, we analyzed nuclear fractionation RNA-sequencing data from mouse bone marrow-derived macrophages (BMDMs) stimulated with the TLR3 agonist poly(I:C) for up to 12 hrs²³. Activation of TLR3 leads to activation of interferon regulator factors, production of interferon- α and β , and induction of a type I interferon response²⁴. From this data, we identified frequent usage of an alternative 5' splice site at the third junction of *Oas1g* (Figure 1B). In each time-point, this alternative “unproductive” splice site is frequently selected over the consensus “productive” splice site (Figure 1C, left). This is evident by simply comparing the number of reads that map across the two different junctions, as well as through the use of the computational program MISO, which utilizes a probabilistic framework to estimate the expression of alternatively spliced isoforms²⁵ (Figure 1C, right). The expression metric is represented by the value Percent Spliced In (PSI; ϕ), which is an estimate of the fraction of transcripts that utilize the alternative splice site. Of interest was the strength of the productive and unproductive splice site, which can be quantified using a maximum entropy model²⁶. We find the productive and unproductive 5' splice sites are similar in strength, and are fairly strong with respect to all expressed junctions (Figure S1).

Next we looked at this alternative splicing event in the context of all expressed junctions. To do this, we calculated the alternative junction usage at each expressed junction from the BMDM data-set stimulated with poly(I:C) for 4 hours (Figure 1D, see methods). From this junction-centric viewpoint, the sequencing data supports the conclusion that most expressed junctions splice with high fidelity (Figure 1E). Still, there is some alternative junction usage, which can be attributed to both regulated AS events as well as splicing noise. With respect to the alternatively spliced junction of *Oas1g*, it ranks among the top percentile

of alternative junction usage, supporting the conclusion that this AS event is among the most frequently utilized in poly(I:C) stimulated BMDMs (Figure 1E).

Removal of Alternative Splice Site Alters *Oas1g* Expression and Macrophage Response to EMCV.

In order to explore the effect of this alternative splicing event on *Oas1g* expression, and correspondingly the antiviral response, we used clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein-9 nuclease (Cas9) technology to engineer murine RAW 264.7 cell lines devoid of this unproductive splice site (Figure 2A). In parallel, cell lines expressing Cas9 and a non-targeting guide were generated. We selected seven clones that had the splice site removed in both alleles, which we deem as “fixed” clones (Figure 2B, S2). RT-PCR upon stimulation with poly(I:C) both confirmed alternative splice site usage in control populations, and showed forced productive splicing in these fixed clones (Figure 2C).

To determine what effect this forced productive splicing has on *Oas1g* expression, we used Taqman qPCR to monitor levels of *Oas1g* in both unstimulated and stimulated (8 hrs poly(I:C)) conditions. In each case, the engineered lines lacking the unproductive *Oas1g* splice site had significantly higher levels of expression, presumably due to lack of AS-NMD associated with selection of the unproductive splice site (Figure 2D). Of interest, levels of *Oas1g* in unstimulated *Oas1g* splice site engineered cells were similar to levels of *Oas1g* in stimulated control cells. Next, to determine the effect of removal of the unproductive splice site with respect to the anti-viral response, we used EMCV to infect both groups of macrophages. EMCV is a (+)ssRNA member of the *Picornaviridae* family that replicate

through partially dsRNA intermediates²⁷. Infection has been shown to cause accumulation of 2-5A, and viral replication is sensitive to the OAS/RNase L pathway^{20,28}. As oligoadenylate synthetases bind viral dsRNA, the RNA activators in EMCV-infected cells are believed to be the viral replicative intermediates¹⁹. Upon 18 hrs of infection with EMCV, we again observed significantly higher levels of *Oas1g* expression in the engineered lines lacking the unproductive *Oas1g* splice site. Using qPCR to measure levels of EMCV following 18 hrs of infection, we found the engineered lines controlled viral replication more efficiently than the control lines. Thus, we conclude that forced productive splicing of *Oas1g* improves the antiviral defense through increased expression of *Oas1g*. Next, as activation of the 2-5A system can affect apoptosis in host cells²⁹, we were interested in determining whether removal of the unproductive *Oas1g* splice site altered the levels of apoptotic cells. These cells were detected with annexin V, which binds to phosphatidylserine exposed on the outer leaflet of cells undergoing apoptosis. We observed ~2 fold increase in the fraction of cells positive for annexin V in the engineered lines lacking the unproductive *Oas1g* splice site as compared to the control cells (Figure 2G) in unstimulated conditions. We conclude that the increased *Oas1g* observed with the removal of the unproductive splice site has the effect of increasing levels of apoptosis in a cell population.

Of note, the other enzymatically active member of the murine *Oas1* family, *Oas1a*, has a highly homologous junction with an identical unproductive splice site. However, despite nearly complete similarity of sequence at and nearby this splice-site (Figure S3A), it is used less frequently than that of *Oas1g* (Figure S3B-D). Because of this similarity, our guide targeted to the unproductive splice site of *Oas1g* also cut at *Oas1a* (Figure S4), and genotyping confirmed all selected clones deleted the *Oas1a* unproductive splice site in

addition to the *Oas1g* unproductive splice site. Again, RT-PCR upon stimulation with poly(I:C) confirmed alternative splice site usage in control populations, and showed forced productive splicing in edited clones (Figure S3C). To determine what effect this forced productive splicing has on *Oas1a* expression, we again used Taqman qPCR to monitor levels of *Oas1a* in both unstimulated and stimulated (8 hrs poly(I:C)) conditions. In this case, we found that while the mean expression of *Oas1a* in both unstimulated and stimulated conditions was greater in non-engineered clones, the effect lacked significance (Figure S3D). We hypothesize the dampened effect with respect to *Oas1a* as compared to *Oas1g* is likely due to decreased usage of the unproductive splice site to begin with, but also note that the small differences observed in *Oas1a* expression levels could play a role in the aforementioned antiviral and apoptosis effects.

A Similar AS-NMD Event Occurs in Human Monocytes

Human Oas1 differs quite significantly from the mouse Oas1 paralogues, a finding that is perhaps not surprising given the volatile evolutionary history of the gene³⁰⁻³². The human Oas1 orthologue contains six exons, and alternative splicing gives rise to five isoforms (p42, p44, p46, p48, and p52) which differ at the C-terminal region. Genetic variation that alters isoform abundance has been shown to lead to altered OAS1 activity and further, viral susceptibility^{33,34}. There exists a single G/A SNP in the OAS1 exon 6 splice-acceptor (rs10774671) that accounts for some of this variability. Those with the G allele predominantly produce p46, while the A allele leads to production of p42, p44, p48, and p52. The p46 isoform has been shown to have increased activity, an effect mediated at least in part by defects in protein accumulation of the other alleles^{22,34,35}.

However, in addition to the productive splicing events that lead to the generation of multiple isoforms, we also find a previously unreported unproductive splicing event at the third splice junction of *Oas1*. Human monocytes consistently and frequently use an alternative splice site that leads to an NMD substrate (Figure 3B). This 3' alternative splice both shifts the frame of the transcript and incorporates a PTC. While this alternative splice site is used less frequently than the one found in murine *Oas1g*, it is worth noting that the human sequencing samples are derived from whole cell RNA as compared to nuclear RNA. As such, the fraction of human *Oas1* transcripts that utilize the unproductive splice site due to efficient degradation of transcripts targeted by the NMD decay machinery in the cytoplasm. Interestingly enough, while this splice site is used frequently in a variety of stimuli, frequency of usage does differ with the cell-type and stimulation. Monocytes stimulated with LPS use the alternative splice site much more frequently as compared to HIV infected CD4+ T cells (Figure 3B, C). This points to stimulation and/or cell-type specific regulation of this splicing event. Regardless, this AS-NMD, coupled with the altered activity from productive splicing events, supports the conclusion that despite differences between human *Oas1* and mouse *Oas1g*, human *Oas1* is extensively regulated at the post-transcriptional level.

AS-NMD Events Are Common in Transcripts Related to Innate Immunity

While *Oas1g* contained one of the most frequently used AS-NMD events, it was not the only AS-NMD event found in genes related to the innate immune response. For example, in nuclear fractionation RNA-sequencing data from mouse BMDMs stimulated with poly(I:C), we found significantly utilized skipped exon events that led to a frameshift and incorporation

of a PTC in the important inflammatory transcripts *Mx1*, *IKKε*, and *Oasl2* (Figure 4A-C). In each case, the event is utilized in all, or nearly all of the sequenced time-points. These events were confirmed in a macrophage cell line with RT-PCR upon stimulation with poly(I:C) for 4, 8, and 12 hours (Figure 4E-F). To classify AS-NMD events globally, we utilized the tool SplAdder to predict and quantify AS events supported by an input sample³⁶. A stringent confidence criteria was required to avoid including AS events derived from splicing noise (see methods). Then, a custom Python script was used to select only events that led to frameshifts and/or PTC inclusion. Among the list of AS-NMD events, as compared to a background of expressed genes, we observed significant enrichment for GO terms associated with the innate immune response (Figure 4G). With respect to the viral pathogen response, which is tasked with limiting viral replication through degradation of viral (as well as non-viral) mRNA and establishment of a cellular antiviral state, a host of factors involved with the response contain AS-NMD events identified here or in other published work²³ (Figure 4H).

Discussion

The robust and coordinated gene expression program involved in the defense against pathogens requires extraordinarily tight regulation. In this study, we sought to shed light on the role of AS-NMD in this regulation. We identified a frequently used unproductive splicing event in *Oas1g*, an important murine anti-viral response factor, and show that forced productive splicing leads to increased *Oas1g* expression and further, an increased ability to clear virus. Additionally, we identify a number of other examples of unproductive splicing

events in the innate immune response which could subject the corresponding transcript to decay via the NMD pathway.

With respect to *Oas1g*, it is fair to ask what benefit such an alternative splicing event offers? The alternative splice site mediating this AS-NMD event is of comparable strength to the consensus 5' splice site (Figure S1). If possession of the greatest pathogen defense were the only goal of an organism, it seems unlikely this splice site would be retained. However, while pathogen defense systems can provide a protective benefit, they also can cause collateral damage to a host. With respect to *Oas1*, its pathogen defense effects are repeatedly forfeited by a host due to the fact its activity can be so detrimental^{22,37}. This is exemplified by the surprisingly high frequency of loss-of-function mutations in primates²², and the fact OAS1 activity has been completely lost in several animal lineages, including teleost fish and insects³⁷. Moreover, while mice deficient for RNase L, the downstream effector of *Oas1* in the 2-5A system, exhibit susceptibility to viral infection²⁰, in the absence of infection they display significantly increased longevity²¹. Given the fact that host RNAs have been shown to be able to activate OAS enzymes, its reasonable to hypothesize that the longevity effect is mediated, at least in part, by chronic 2-5A production^{22,38-40}. With respect to the AS-NMD event we observed in *Oas1g*, we found removal of the unproductive splice site significantly increased the number of cells undergoing apoptosis. From this, it stands to reason that removal of the unproductive splice site, while improving the ability to limit viral infection, could negatively impact host fitness. In turn, we believe this splice site represents a compromise between the pathogen protective benefits and collateral damage associated with OAS1g activity.

A second question has to do with the manner with which this mitigation occurs. Innately, regulation at the post-transcriptional level through AS-NMD seems appears inefficient. Why spend the resources to transcribe a transcript if it is destined for degradation? For one, the very fact introns exist and are transient in nature argues against the idea that the cost of transcription is prohibitive⁴¹. A significant majority of transcribed sequence (~90% in humans⁴²) is spliced and discarded. Additionally, it is well understood that transcriptional regulation is largely a cooperative venture⁴³, epitomized by complexes like the interferon- β (IFN- β) enhanceosome⁴⁴. As transcriptional regulation is not simply one protein interacting with one DNA sequence, but instead a multitude of proteins interacting with a host of other proteins and a variety of DNA sequences, it is quite possible that once transcriptional control has been placed on a system, changing it quantitatively is difficult. Thus, secondary mechanisms are needed to fine-tune the gene expression levels of select transcripts. As such, we argue the fine-tuning capabilities inherent to splicing based post-transcriptional regulation far outweigh the cellular cost of additional transcription, especially in the context of a tightly regulated gene expression program like inflammation.

It remains unknown whether the AS-NMD event in *Oas1g* is regulated by an external input or whether a constant fraction of transcripts is discarded. While we do see a trend whereby increases in stimulation time accompany decreases in ϕ (Fig. 1C), in essence arguing this AS-NMD event acts as a break released upon *Oas1g* induction, we are hesitant to draw such a conclusion without both increased sequencing depth at *Oas1g* and more sequencing time-points. Regardless, the fact that the alternative splice site for *Oas1g* and *Oas1a* is identical in sequence, and further both junctions are nearly identical, strongly supports the idea that *trans*-acting proteins might affect the process. Newly developed

methods like RAP-MS⁴⁵ and ChIRP⁴⁶, which identify RNA binding proteins bound to RNAs of interest, could help discover interactions that have the ability to affect splice site selection. In summary, we found a frequently used AS-NMD event in *Oas1g*. When the splice site that mediates this event is removed, we observed increased expression of *Oas1g* and an improved antiviral response. A similar AS-NMD event was found in human *Oas1*. Indeed, genetic variation that dampens OAS1 activity in humans has been shown to lead to susceptibility to viral infection, particularly to West Nile virus³³ and Epstein Barr virus³⁴. Forced productive splicing using an antisense oligonucleotide could limit viral propagation and thus, has a potential therapeutic role in the treatment of infection. Similar unproductive splicing events were found throughout the innate immune response. While future studies should seek to understand the functional significance of individual events, this form of unproductive splicing represents a previously underappreciated mechanism of gene expression regulation in innate immunity.

ACKNOWLEDGEMENTS

The authors would like to thank Bert Semler (Department of Microbiology, Immunology & Molecular Genetics, UCLA) and Frank van Kuppeveld (Molecular Virology, Faculty of Veterinary Medicine, Utrecht University) for providing EMCV. This work was funded from an endowment provided by the Raymond and Beverly Sackler Foundation.

AUTHOR CONTRIBUTIONS

L.F. and D.B. conceived and designed experiments. L.F. conducted experiments and analyzed data. The manuscript was written by L.F and D.B. M.M., A.J., and G.L. provided advice and assistance with experiments and data analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests with respect to this manuscript.

Figure Legends

Figure 1. *Oas1g* has a Frequent AS-NMD Event. (A) A schematic depiction an alternative splicing event leading to either a productive isoform destined for translation or an unproductive isoform destined for degradation. (B) Sashimi plot for the entire gene body of *Oas1g* from BMDMs stimulated with poly(I:C) for 12 hrs. *Oas1g* is a negative strand gene and is depicted right to left. (C) (left) Sashimi plots centered at the third junction of *Oas1g* from BMDMs stimulated with poly(I:C) for 0, 1, 4, 8, and 12 hrs. (right) ϕ estimates (red line), as well as confidence intervals over estimates (histogram) for each time point. (D) Schematic representation of the alternative junction usage calculation. (E) Pie chart representing alternative junction usage for all expressed junctions upon 4 hrs. of poly(I:C) stimulation. The slice including the alternatively spliced third junction of *Oas1g* is depicted by the arrow.

Figure 2. Removal of Alternative Splice Site Alters *Oas1g* Expression and Macrophage Response to EMCV. (A) Schematic representation of the two alternative splice isoforms, and the gRNA/Cas9 targeting of the alternative splice site. (B) Sanger sequencing gDNA from a control sample (top) and an *Oas1* SS KO sample (bottom). Sequencing is oriented such that the negative strand runs left to right. The alternative splice site is represented by the yellow highlighted region. (C) RT-PCR upon stimulation with poly(I:C) confirming alternative splice site usage in control populations and forced productive splicing in fixed clones. (D) RT-qPCR analysis of *Oas1g* mRNA levels in unstimulated and stimulated (8 hrs poly(I:C)) macrophages. Control samples are represented in light blue, SS KO clones are represented in dark blue. (E) RT-qPCR analysis of *Oas1g* mRNA levels in EMCV

infected (18 hrs) macrophages. Control samples are represented in light blue, SS KO clones are represented in dark blue. **(F)** RT-qPCR measurement of EMCV viral load following 18 hrs of infection at 1 MOI. Control samples are represented in light blue, SS KO clones are represented in dark blue. **(G)** Annexin V staining for apoptotic cells under unstimulated conditions. Control samples are represented in light blue, SS KO clones are represented in dark blue. Data is representative of two independent experiments **(D-G)** and is shown as mean (error bars indicate SEM). * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student's t test. Results are presented relative to those of *Rpl32* **(D-F)**.

Figure 3. A Similar AS-NMD Event Occurs in Human Macrophages. **(A)** Depiction of the mRNA splice isoforms found in human *Oas1*. There exists a single G/A SNP in the OAS1 exon 6 splice-acceptor (rs10774671), with the G variant producing the more active p46 isoform. **(B)** Sashimi plots for an AS-NMD event identified in exon 3 of human *Oas1* from human monocytes stimulated with LPS for 0, 1 and 6 hrs. **(C)** Same as (B) but for patient derived HIV infected CD4+ T cells.

Figure 4. AS-NMD Events Are Common in Transcripts Related to Innate Immunity. **(A)** Sashimi plots for an AS-NMD event identified in *Mx1* from BMDMs stimulated with poly(I:C) for 0, 1, 4, 8, and 12 hrs. **(B)** Same as (A) for *IKKε*. **(C)** Same as (A) for *Oasl2*. **(D)** RT-PCR of *Mx1* upon stimulation with poly(I:C) for 4, 8, and 12 hrs. **(E)** Same as (D) for *IKKε*. **(F)** Same as (D) for *Oasl2* **(G)** GO terms enriched for AS-NMD events, as compared to a background of expressed genes. **(H)** Schematic representation of major

pathways in the viral pathogen response. Red arrows are shown above factors containing AS-NMD events. Data is representative of two independent experiments (**D-F**).

Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Baltimore.

Experimental Model and Subject Detail

Animals

The California Institute of Technology Institutional Animal Care and Use Committee approved all experiments. Bone marrow derived macrophages were isolated from mixed-sex C57BL/6 mice and cultured and stimulated as previously described²³.

Cell Culture

human embryonic kidney cells (HEK293T) were cultured in DMEM supplemented with 10% FBS. All cell lines were maintained at 37°C. Human embryonic kidney cells (HEK293T) from ATCC were cultured in DMEM supplemented with 10% FBS and 1% Pen/Strep. RAW 264.7 murine macrophages from ATCC were cultured in DMEM supplemented with 10% FBS and 1% Pen/Strep. Cell lines were maintained at 37°C in 5% CO₂.

Method Detail

RNA Isolation

Total RNA was purified from BMDMs using TRIzol reagent (Ambion) as per the manufacturer's instructions. Genomic DNA in RNA purifications was eliminated through

treatment with Turbo DNase (Thermo Fisher Scientific) for 30 min at 37°C. RT reactions were performed in 20µL (20mM DTT, 2X ProtoScript II Reverse Transcriptase Reaction Buffer (NEB), 1mM dNTPs, 40U Murine RNase Inhibitor (NEB), and 200U ProtoScript II (NEB) Reverse Transcriptase) with 500-1000ng RNA. Reaction incubated in thermocycler with the following program: 1. 42°C for 60min, 2. 65°C for 20min.

RT-PCR of Splice Isoforms

Total RNA was isolated using Tri reagent solution and digested with DNase I (Invitrogen). RT reactions were performed in 20µL (20mM DTT, 2X ProtoScript II Reverse Transcriptase Reaction Buffer (NEB), 1mM dNTPs, 40U Murine RNase Inhibitor (NEB), and 200U ProtoScript II (NEB) Reverse Transcriptase) with 500-1000ng RNA. Reaction incubated in thermocycler with the following program: 1. 42°C for 60min, 2. 65°C for 20min. PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB) and gene specific primers.

CRISPR Experiments

RAW 264.7 cell lines were grown in individual 10cm plates. 20,000 cells were plated in a cell well plate and left overnight to adhere. Following adherence, lentivirus expressing Cas9 and either the *Oas1a/g* guide or a scramble control guide was added to the cells. 48 hours later, infected cells were selected with puromycin, which was added at a concentration of 3.75ug/mL. Following 72 hours of selection, cells infected with the *Oas1a/g* guide were single-cell plated in 96 well plates. Clones were passed to 6-well plates following 5 days of growth, at which point genotyping was performed. Cells infected with the scramble control

guide were passaged as a bulk infected population, with independent biological replicates representing cells independently infected and puromycin selected.

Poly(I:C) Stimulations

RAW 264.7 cell lines were grown in individual 10cm plates. 12 hours prior to infection cells were counted and plated at a density of 350,000 cells/well in 6 well plate. Following adherence, 5ug/mL of Poly(I:C) (Sigma) was added. Following the infection, cells were lysed in TRIzol. Viral RNA was quantified as described above. Cellular RNA was quantified as described above.

EMCV Infection Experiment

RAW 264.7 cell lines were grown in individual 10cm plates. 12 hours prior to infection, cells were counted and plated at a density of 350,000 cells/well in 6 well plate. Following adherence, EMCV was added at the specified MOI for the specified amount of time. Following the infection, cells were lysed in TRIzol. Cellular and viral RNA was quantified as described above.

Annexin V Experiment

RAW 264.7 cell lines were grown in individual 10cm plates. 12 hours prior to infection, cells were counted and plated at a density of 125,000 cells/well in 12 well plate. 24 hours after plating, cells were stained with Annexin V APC Ready Flow Conjugate (Thermo Fischer). Samples were analyzed on a MACSQuant10 Flow Cytometry machine (Miltenyi). Data was analyzed with FlowJo 10.2.

Data Download

Raw RNA-sequencing samples in FASTQ format were downloaded from the Gene Expression Omnibus (*GEO*) database. Mouse bone-marrow macrophages derived data can be found under accession number GSE122543. Human monocyte data can be found under accession number GSE60216

RNA-Sequencing Analysis

Sequencing was performed on a HiSeq 2500 High Throughput Sequencer (Illumina). All previously downloaded RNA-seq samples were individually aligned using a uniform processing pipeline based on the *STAR aligner*⁴⁷. The STAR software (version 2.6.0a) was used in a 2-pass mode. The first pass identifies non-annotated junctions in the input, allowing for the construction of a genome index containing non-annotated junctions. The second pass alignment is then performed against the junction-aware index. The command line parameters were as follows:

```
STAR --runThreadN 12 --genomeDir GENOME --genomeLoad NoSharedMemory --sjdbGTFfile GTF --twopassMode Basic --readFilesIn FASTQ --outFileNamePrefix NAME -outSAMattributes NM --outStd BAM_Unsorted --outSAMtype BAM Unsorted --outSAMmode Full --outSAMstrandField intronMotif --outSAMunmapped None --outFilterType BySJout > OUTPUT.bam
```

Following alignment, Portcullis⁴⁸ was used to filter invalid splice junctions from the aligned BAM file. Isoform expression was quantified using the raw fastq files and the mouse reference transcriptome mm9 as input for Kallisto (v.0.45.0)⁴⁹. The resultant normalized transcript frequencies were provided to the R package Sleuth for differential

analysis (v.0.30.0)⁵⁰. Alternative splicing events were detected and quantified on produced BAM files using the *SplAdder* toolkit³⁶ as per Kahles et al., 2018⁵¹. Unproductive splicing events were determined using a custom Python script. Alternative splicing sashimi plots across an entire gene were generated with ggsashimi⁵² whereby only junctions with greater than 5 supporting RNA-seq reads were plotted, while alternative splicing sashimi plots centered at an individual junction were generated with MISO²⁵.

Alternative Junction Usage Analysis

Isoform expression was quantified using the raw fastq files and the mouse reference transcriptome mm9 as input for Kallisto (v.0.45.0). Only isoforms with a TPM greater than 10 were considered. Junctions from the most abundant isoform for each gene were selected, as long as there were 8 supporting reads for the junction. Alternative junction usage was calculated by comparing the number of reads that overlap a given selected junction, but did not utilize the same 5' and/or 3' splice site, to the total number of reads at a junction.

Quantification and Statistical Analysis

All statistical analysis was performed in Python (version 2.7.9). Unless otherwise indicated in figure legends, statistical significance measurements were marked as follows: * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$, and n.s. denotes not significant.

References

1. Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F. & Kollias, G. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398 (1999).
2. Gautier, E. L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118 (2012).
3. Chen, J. & Chen, Z. J. Regulation of NF- κ B by ubiquitination. *Curr. Opin. Immunol.* **25**, 4–12 (2013).
4. Smale, S. T. & Natoli, G. Transcriptional control of inflammatory responses. *Cold Spring Harb. Perspect. Biol.* **6**, a016261 (2014).
5. Smale, S. T., Tarakhovskiy, A. & Natoli, G. Chromatin contributions to the regulation of innate immunity. *Annu. Rev. Immunol.* **32**, 489–511 (2014).
6. Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* **10**, 281 (2009).
7. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).
8. O’Connell, R. M., Rao, D. S. & Baltimore, D. microRNA regulation of inflammatory responses. *Annu. Rev. Immunol.* **30**, 295–312 (2012).
9. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470 (2008).
10. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413 (2008).
11. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
12. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci.* **100**, 189–192 (2003).
13. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **36**, 1073 (2004).

14. Weischenfeldt, J. *et al.* Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* **13**, R35 (2012).
15. Jangi, M. & Sharp, P. A. Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498 (2014).
16. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926 (2007).
17. Kakuta, S., Shibata, S. & Iwakura, Y. Genomic structure of the mouse 2', 5'-oligoadenylate synthetase gene family. *J. Interferon Cytokine Res.* **22**, 981–993 (2002).
18. Elkhateeb, E. *et al.* The role of mouse 2', 5'-oligoadenylate synthetase 1 paralogs. *Infect. Genet. Evol.* **45**, 393–401 (2016).
19. Silverman, R. H. Viral encounters with 2', 5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. *J. Virol.* **81**, 12720–12729 (2007).
20. Zhou, A. *et al.* Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997).
21. Andersen, J. B. *et al.* Role of 2-5A-dependent RNase-L in senescence and longevity. *Oncogene* **26**, 3081 (2007).
22. Carey, C. M. *et al.* Recurrent loss-of-function mutations reveal costs to OAS1 antiviral activity in primates. *Cell Host Microbe* **25**, 336–343 (2019).
23. Frankiw, L. *et al.* Bud13 Promotes a Type I Interferon Response By Countering Intron Retention in Irf7. *Mol. Cell* **73**, 803–814 (2019).
24. Perales-Linares, R. & Navas-Martin, S. Toll-like receptor 3 in viral pathogenesis: friend or foe? *Immunology* **140**, 153–167 (2013).
25. Katz, Y., Wang, E. T., Airoidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
26. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
27. Carocci, M. & Bakkali-Kassimi, L. The encephalomyocarditis virus. *Virulence* **3**, 351–367 (2012).
28. Hearl, W. G. & Johnston, M. I. Accumulation of 2',5'-oligoadenylates in encephalomyocarditis virus-infected mice. *J. Virol.* **61**, 1586–1592 (1987).

29. Castelli, J. C. *et al.* A study of the interferon antiviral mechanism: apoptosis activation by the 2–5A system. *J. Exp. Med.* **186**, 967–972 (1997).
30. Kumar, S., Mitnik, C., Valente, G. & Floyd-Smith, G. Expansion and molecular evolution of the interferon-induced 2'–5' oligoadenylate synthetase gene family. *Mol. Biol. Evol.* **17**, 738–750 (2000).
31. Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L. & Elde, N. C. Overlapping patterns of rapid evolution in the nucleic acid sensors cGAS and OAS1 suggest a common mechanism of pathogen antagonism and escape. *PLoS Genet.* **11**, e1005203 (2015).
32. Fish, I. & Boissinot, S. Functional evolution of the OAS1 viral sensor: insights from old world primates. *Infect. Genet. Evol.* **44**, 341–350 (2016).
33. Lim, J. K. *et al.* Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. *PLoS Pathog.* **5**, e1000321 (2009).
34. Liu, X. *et al.* A functional variant in the OAS1 gene is associated with Sjögren's syndrome complicated with HBV infection. *Sci. Rep.* **7**, 17571 (2017).
35. Bonnevie-Nielsen, V. *et al.* Variation in Antiviral 2',5'-Oligoadenylate Synthetase (2'5'AS) Enzyme Activity Is Controlled by a Single-Nucleotide Polymorphism at a Splice-Acceptor Site in the OAS1 Gene. *Am. J. Hum. Genet.* **76**, 623–633 (2005).
36. Kahles, A., Ong, C. S., Zhong, Y. & Rättsch, G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics* **32**, 1840–1847 (2016).
37. Kjaer, K. H. *et al.* Evolution of the 2'-5'-oligoadenylate synthetase family in eukaryotes and bacteria. *J. Mol. Evol.* **69**, 612 (2009).
38. Molinaro, R. J. *et al.* Selection and cloning of poly (rC)-binding protein 2 and Raf kinase inhibitor protein RNA activators of 2', 5'-oligoadenylate synthetase from prostate cancer cells. *Nucleic Acids Res.* **34**, 6684–6695 (2006).
39. Dan, M., Zheng, D., Field, L. L. & Bonnevie-Nielsen, V. Induction and activation of antiviral enzyme 2', 5'-oligoadenylate synthetase by in vitro transcribed insulin mRNA and other cellular RNAs. *Mol. Biol. Rep.* **39**, 7813–7822 (2012).
40. Li, Y. *et al.* Ribonuclease L mediates the cell-lethal phenotype of double-stranded RNA editing enzyme ADAR1 deficiency in a human cell line. *Elife* **6**, e25687 (2017).
41. Lareau, L. F., Brooks, A. N., Soergel, D. A., Meng, Q. & Brenner, S. E. The coupling of alternative splicing and nonsense-mediated mRNA decay. (2007).
42. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).

43. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613 (2012).
44. Thanos, D. & Maniatis, T. Virus induction of human IFN β gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).
45. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232 (2015).
46. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
47. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Mapleson, D., Venturini, L., Kaithakottil, G. & Swarbreck, D. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience* **7**, (2018).
49. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525 (2016).
50. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687 (2017).
51. Kahles, A. *et al.* Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224 (2018).
52. Garrido-Martín, D., Palumbo, E., Guigó, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser-and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360 (2018).

Supplemental Figure Legends

Figure S1. The Alternative 5' Splice Site Mediating the AS Event is of Similar Strength to the Consensus 5' Splice Site. (A) Histogram representing the 5' splice site strength (MaxEntScore) of introns of expressed in BMDMs. The bin with which the consensus splice site falls is shown by the light blue line. The bin with which the alternative splice site falls is shown by the dark blue line. (C) MaxEntScore quantifications of the consensus and alternative 5' splice site.

Figure S2. *Oas1g* Macrophage Cell Line Genotyping. Sanger sequencing gDNA from a control sample (very top) and the *Oas1g* SS KO clones. Sequencing is centered around the *Oas1g* alternative splice site. Sequencing is oriented such that the negative strand runs left to right.

Figure S3. *Oas1a* has a Similar Frequently Utilized AS-NMD Event. (A) Schematic depiction showing the homology between *Oas1a* and *Oas1g* at the alternatively spliced third junction. (B) (left) Sashimi plots centered at the third junction of *Oas1g* from BMDMs stimulated with poly(I:C) for 0, 1, 4, 8, and 12 hrs. (right) ϕ estimates (red line), as well as confidence intervals over estimates (histogram) for each time point. (C) RT-PCR upon stimulation with poly(I:C) confirming alternative splice site usage in control populations and forced productive splicing in fixed clones. (D) RT-qPCR analysis of *Oas1g* mRNA levels in unstimulated and stimulated (8 hrs poly(I:C)) macrophages. Control samples are represented in light blue, SS KO clones are represented in dark blue. Data is representative of two independent experiments (D-F) and is shown as mean (error bars indicate SEM). *

denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ using a Student's t test.

Results are presented relative to those of *Rpl32* (**D**)

Figure S4. *Oas1a* Macrophage Cell Line Genotyping. Sanger sequencing gDNA from a control sample (very top) and the *Oas1a* SS KO clones. Sequencing is centered around the *Oas1a* alternative splice site. Sequencing is oriented such that the negative strand runs left to right.

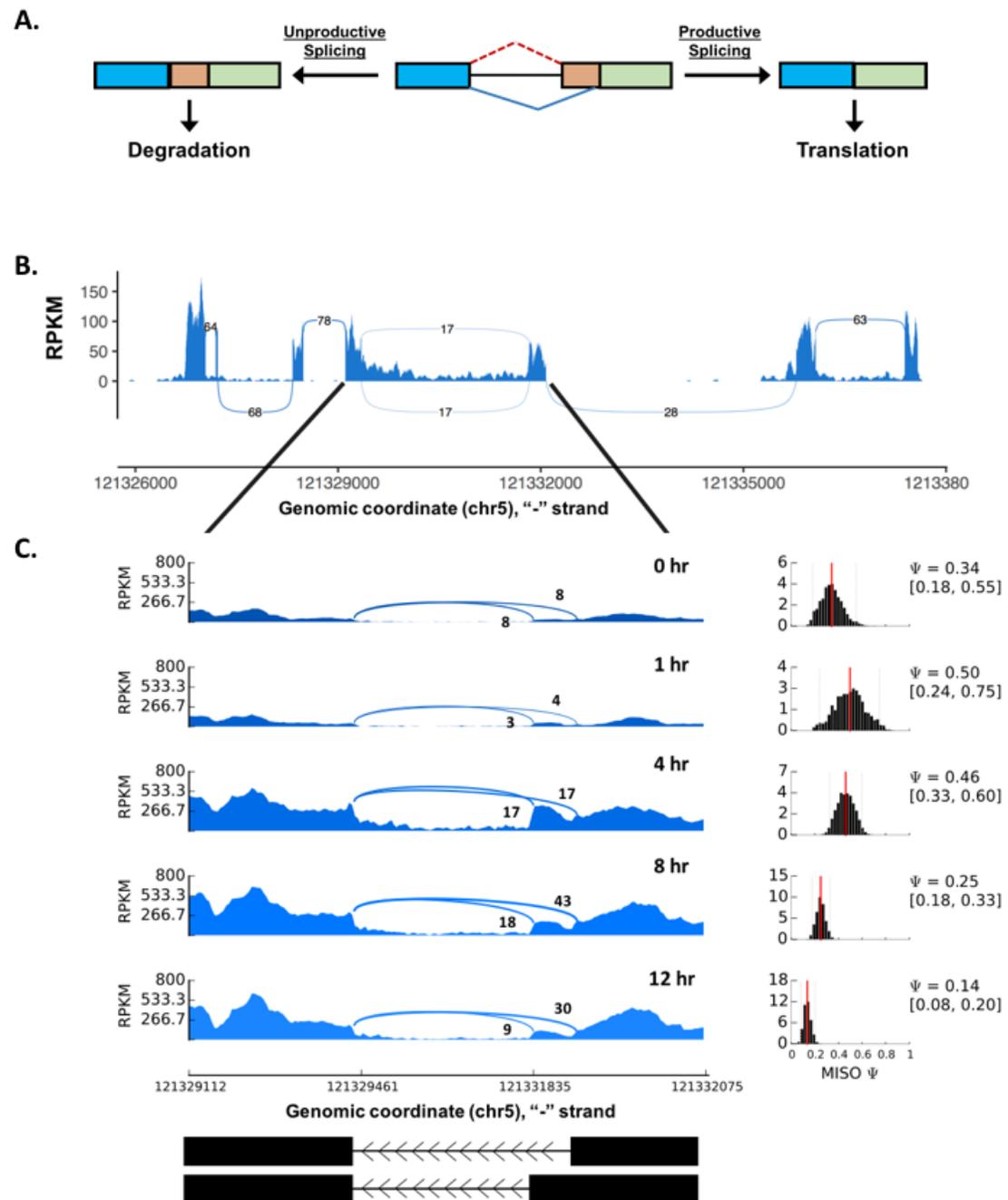


Figure 1-1

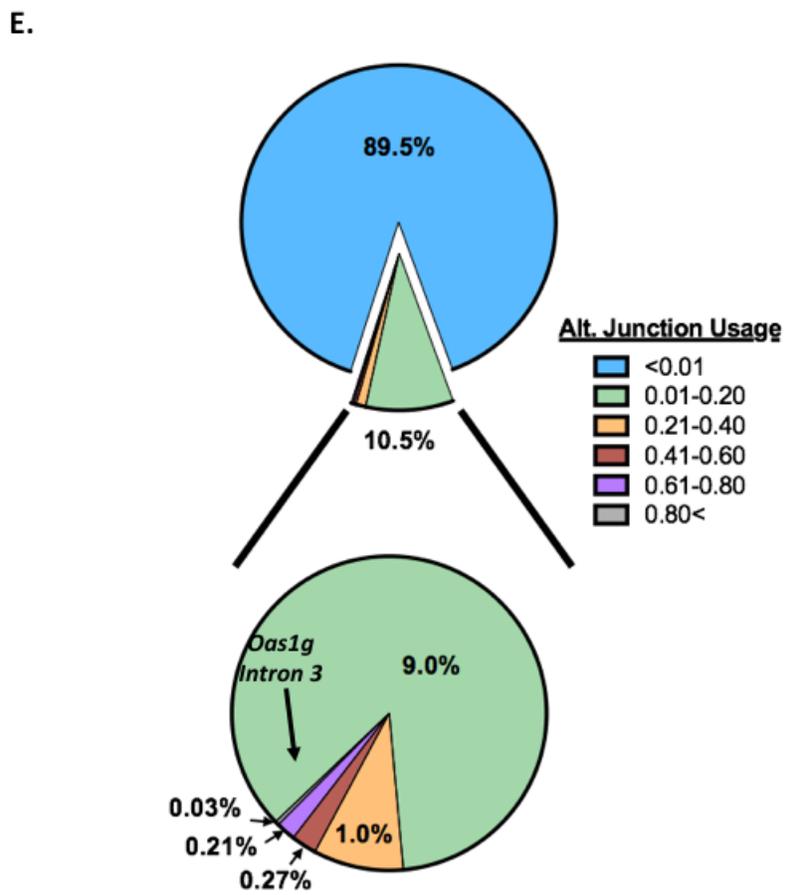
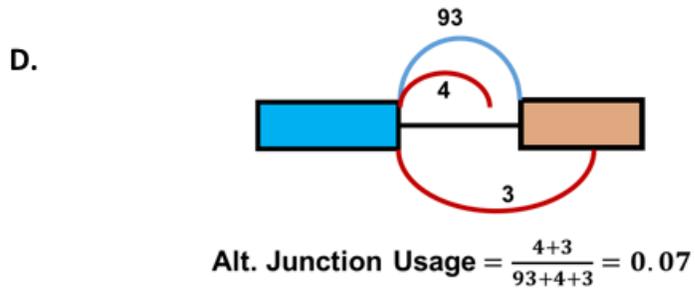


Figure 1-2

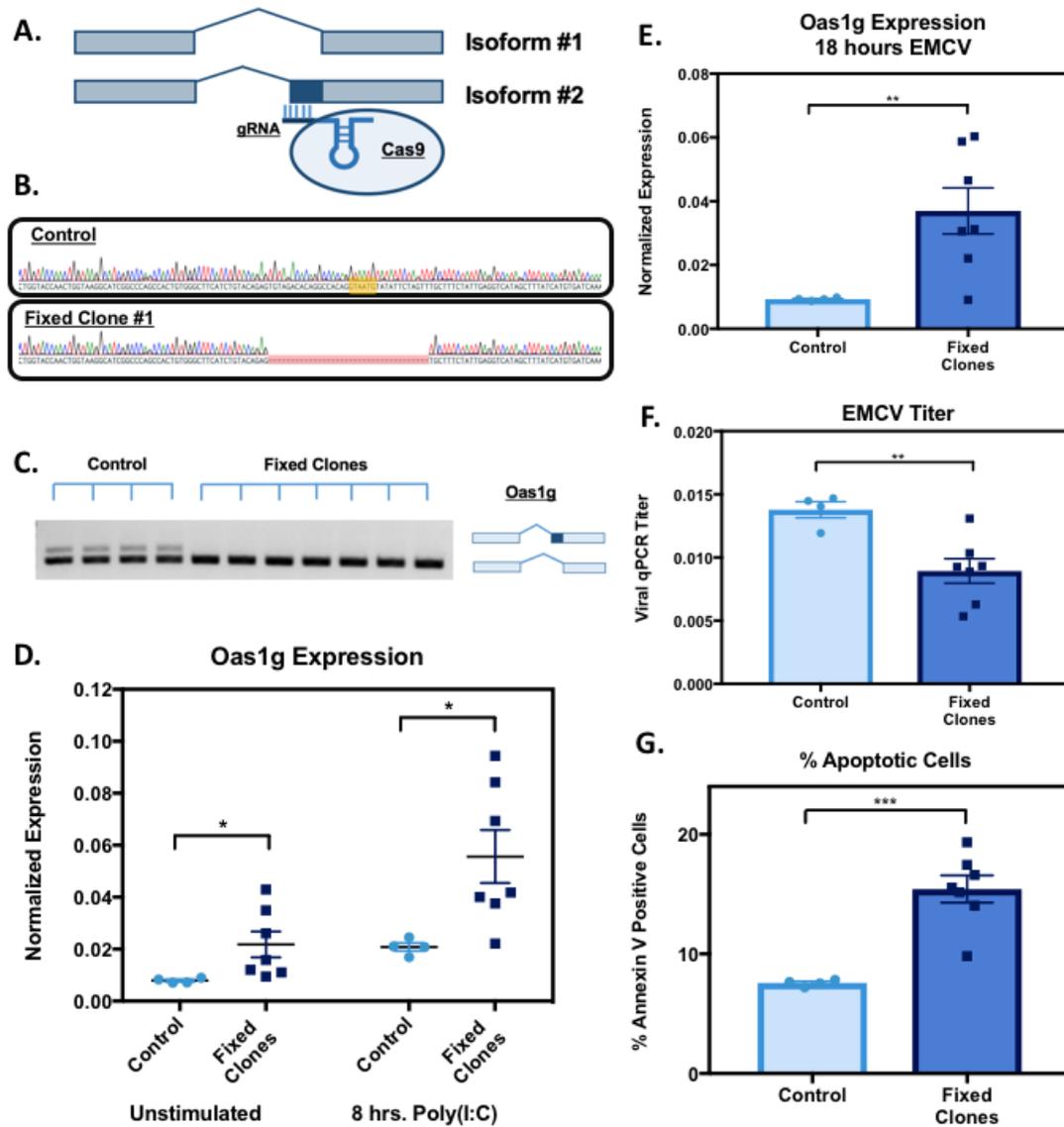


Figure 2

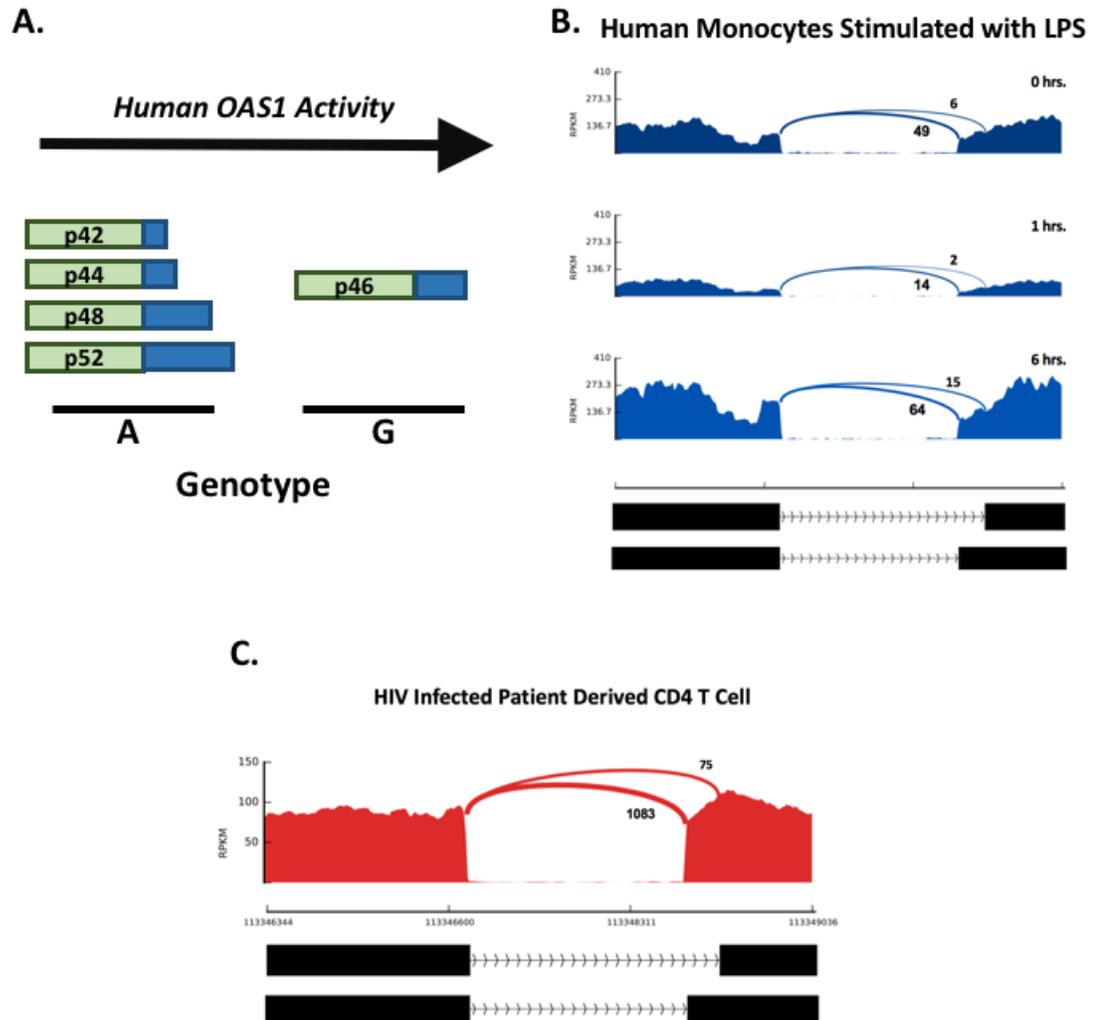


Figure 3

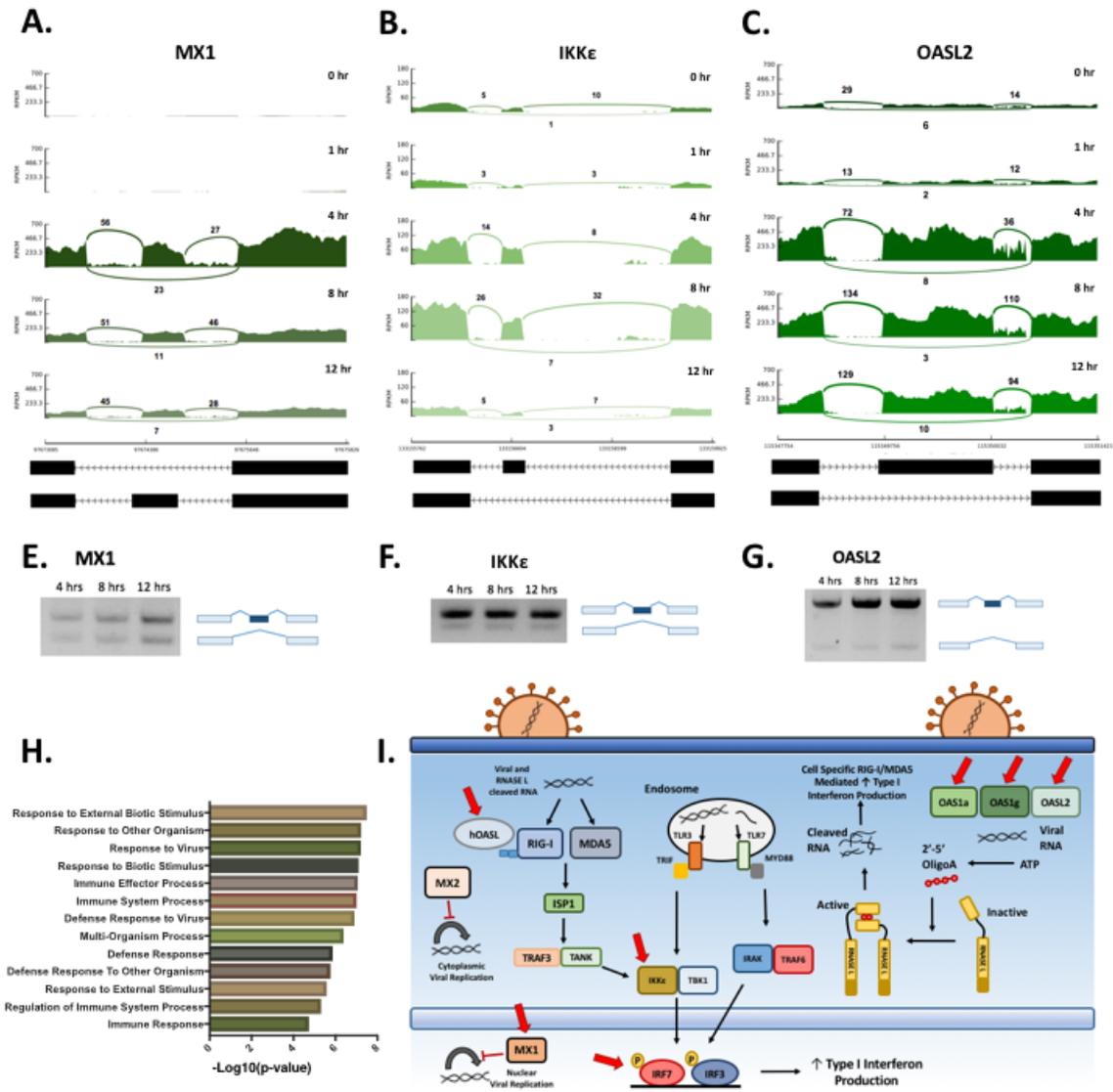


Figure 4

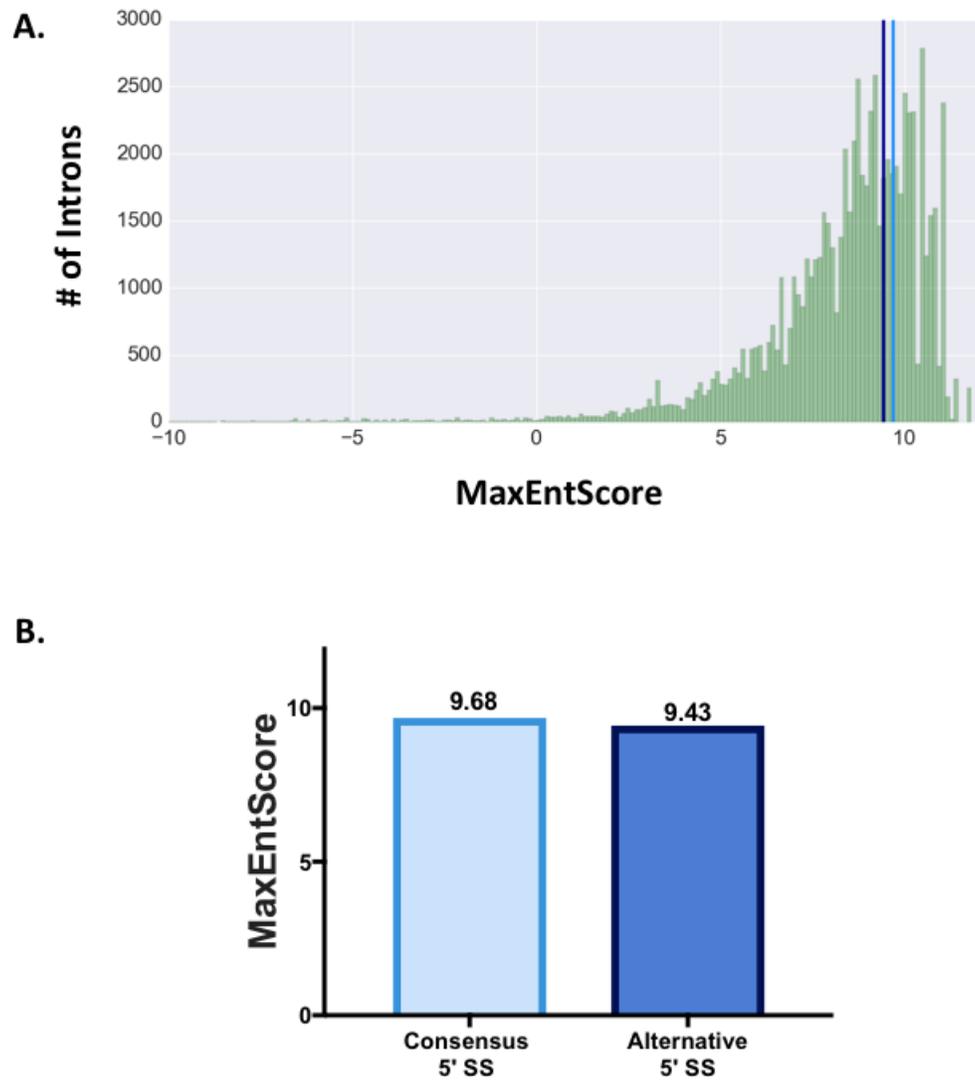


Figure S1

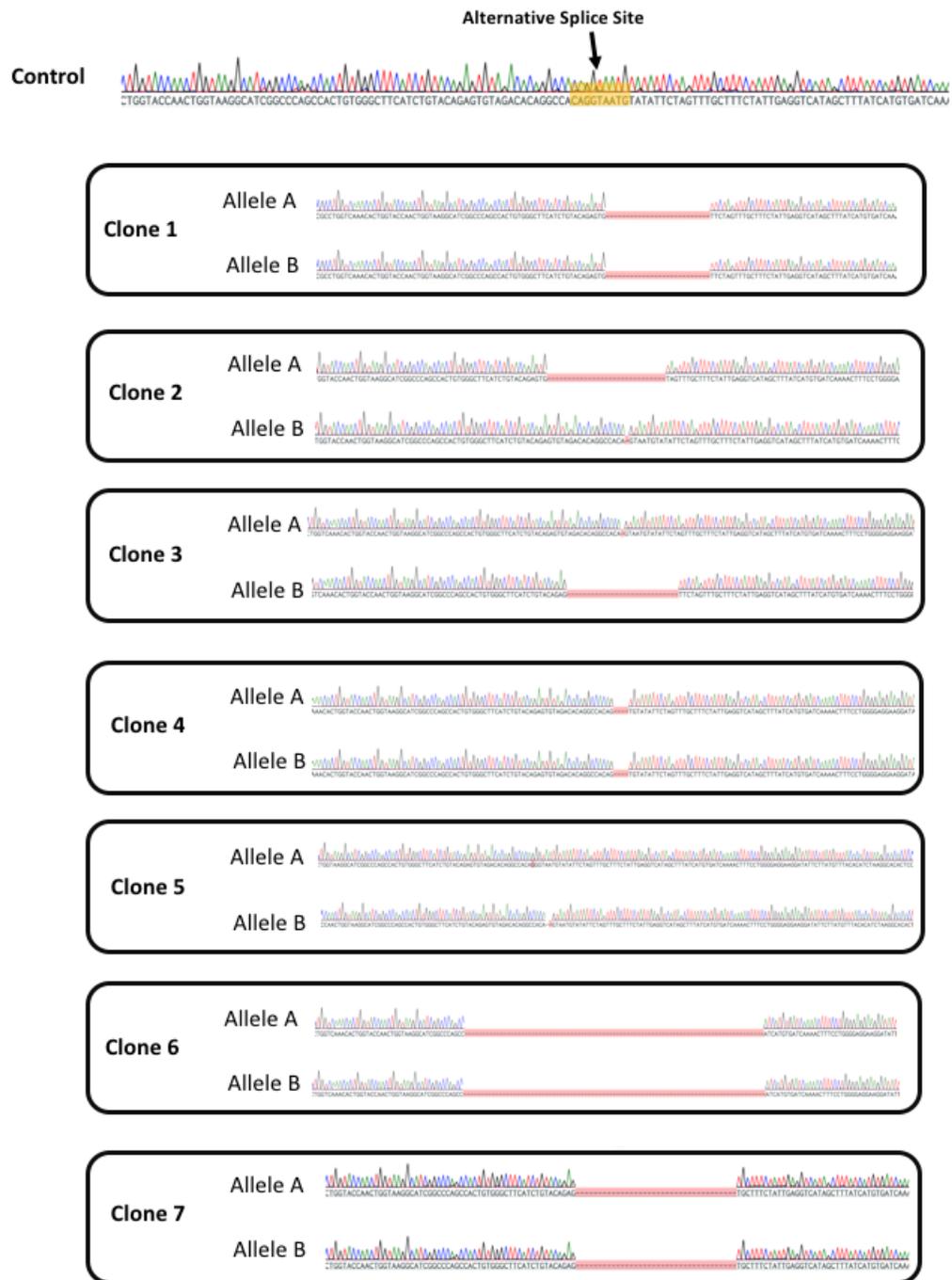


Figure S2

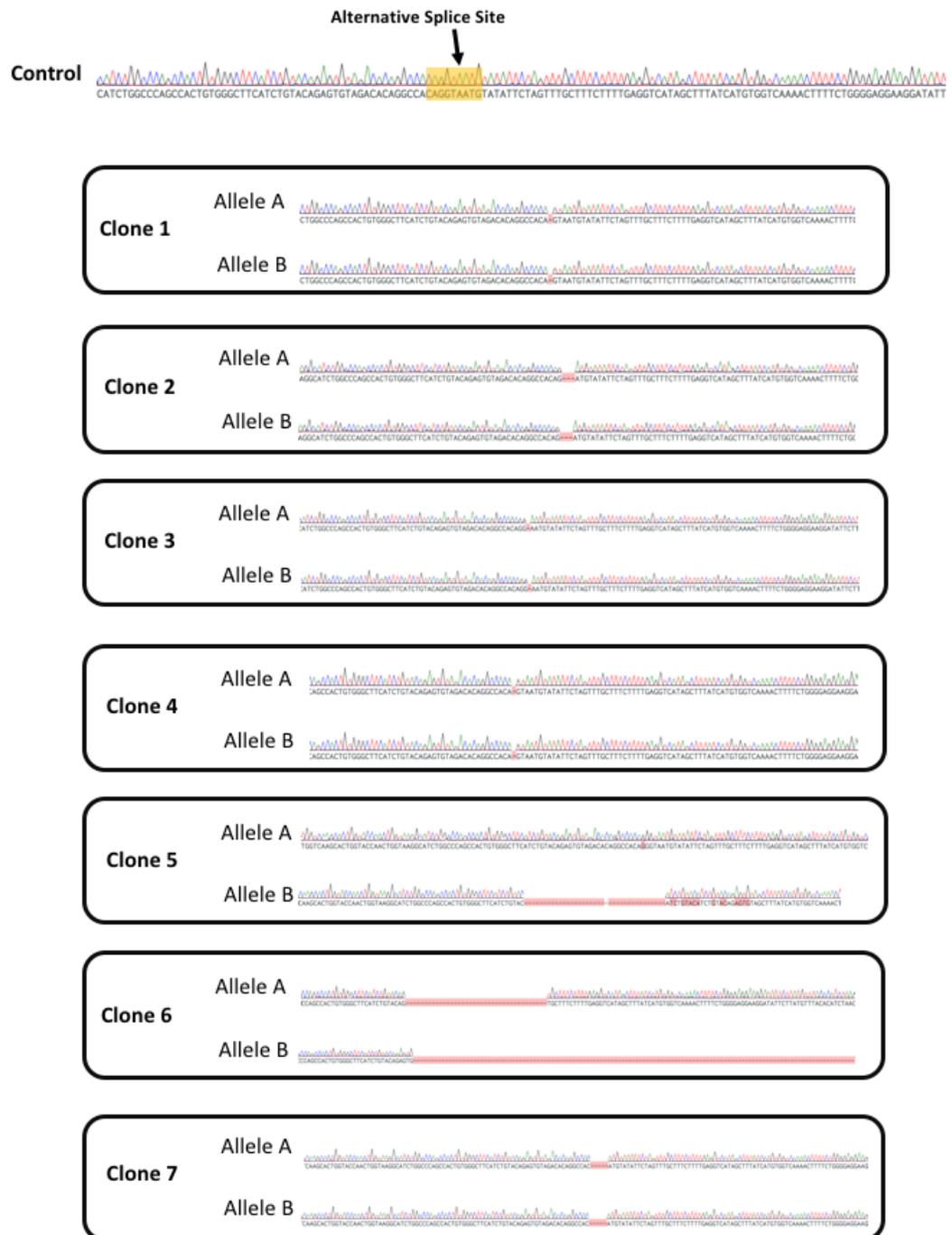


Figure S4

*Chapter 5***Conclusion and Future Directions**

Concluding Remarks

This thesis highlights the importance of post-transcriptional regulation mediated by mRNA splicing in the control of an inflammatory gene expression program. In this thesis, I have identified unproductive splicing events that affect gene expression levels of transcripts involved in the innate immune response, as well as a *trans*-acting factor that regulates one such event. This type of regulation has shown to be extraordinarily important for the fine-tuning of the tightly regulated inflammatory gene expression program.

While seemingly an inefficient method of regulation, post-transcriptional regulation is a key mechanism by which a cell can fine-tune the levels of specific transcripts. While transcriptional regulation has been the focus in the study of gene expression regulation, it has some characteristics that make the process of fine-tuning difficult. In contrast to post-transcriptional regulation, transcriptional regulation is largely a cooperative venture¹. It is not simply one protein interacting with one DNA sequence, but instead a multitude of proteins interacting with a host of other proteins and a variety of DNA sequences. Thus, we believe that once transcriptional control has been placed on a system, changing it quantitatively is difficult. In turn, secondary mechanisms, like the one highlighted in this thesis, are needed for the fine-tuning of gene expression levels. While there are cellular costs due to unneeded transcription, we argue the fine-tuning capabilities inherent to splicing based post-transcriptional regulation far outweigh this increased burden. Furthermore, the very fact introns exist and are transient in nature argues against the idea that the cost of transcription is prohibitive as a significant majority of transcribed sequence are spliced and discarded².

An innate immune response offers an exceptional system to study such regulation. The innate immune response is crucial in the fight against infection; however, there are inherent costs to such a response. If this response is left unscaled or unchecked, it can do damage to an organism and contribute to a variety of diseases³. Thus, this response needs to be highly regulated and turned on at the right scale only in response to pathogens. If not, the balance can be shifted from protective to destructive immunity. Relatively small changes to this tightly regulated gene expression program can have fairly drastic effects⁴. Thus, this post-transcriptionally mediated fine-tuning is exceptionally important in such a system, and dysregulation has easily observable effects.

Future Directions

While it is apparent such post-transcriptional regulation mediated by mRNA splicing is frequent and wide-spread during an innate immune response, several open questions remain regarding the role of this form of regulation. The most obvious questions have to do with the regulation of these unproductive splicing events. With respect to the third chapter, it remains unknown whether or not the retention event in *Irf7* is actively regulated. We showed decreasing the splicing efficiency of the fourth intron has the ability to significantly dampen the functional output of IRF7. We hypothesize this dampening acts to mitigate what otherwise might be an unchecked or inappropriately scaled response. However, whether a cell actively controls this splicing event and thus the intron serves as a regulatory control point remains unknown. Further, despite the fact we identified the factor BUD13 as a component involved with the splicing efficiency of the intron, it remains unknown whether BUD13 plays a role in this regulation. An alternative hypothesis is that BUD13 simply

represents a mechanism that has evolved to counter IR in a subset of inherently weak introns that require splicing⁵.

Regulation is also a question affecting *Oas1g* and *Oas1a* in the fourth chapter. Are these alternative splicing events regulated by an external input or is some constant fraction of transcripts discarded? With respect to *Oas1g*, we do notice a trend whereby increases in stimulation time accompany decreases in PSI (ϕ). It is tempting to speculate regarding this trend. For example, it is possible this unproductive splicing event acts as a break released upon *Oas1g* induction. However, we are hesitant to draw such a conclusion without both increased sequencing depth at *Oas1g* and more sequencing time-points. Regardless, the fact that the alternative splice sites and much of the junction for *Oas1g* and *Oas1a* are identical in sequence, yet the two junction significantly differ in their usage of the unproductive splice site, strongly suggests a *trans*-acting proteins might affect this process.

Beyond regulation, an important future direction involves the determination of functional relevance of many of the other identified unproductive splicing events. While *Irf7* and *Oas1g* contain two of the most frequently used unproductive splicing events, they are not the the only events found in genes related to the innate immune response. For example, in RNA-sequencing data from mouse BMDMs stimulated with poly(I:C), we found significantly utilized skipped exon events that led to a frameshift and incorporation of a PTC in the important inflammatory transcripts *Mx1*, *IKK ϵ* , and *Oas12*. In each case, the event is utilized in all, or nearly all of the sequenced time-points (Chapter 4). The computational tool SplAdder was used to predict and quantify AS events⁶. A stringent confidence criteria was utilized to avoid including AS events derived from splicing noise. However, mRNA splicing can be a noisy process as the dynamic nature of the spliceosome can be a source of stochastic

fluctuation⁷. In order to prove these splicing events play a role in post-transcriptional regulation, and are not just a byproduct of this noise, it needs to be shown that removal of the event has some functional effect.

Furthermore, little work has been done regarding the role of unproductive splicing during inflammation at the organismal level. While we notice significant effects upon alteration of unproductive splicing events in cells grown in culture (Chapters 3 and 4), to gain a proper understanding of the physiological implications of the removal of such events, experiments need to be done at the level of an animal. Creating a mouse lacking the alternative splice sites found in *Oas1a* and *Oas1g* with CRISPR-Cas9 technology should be relatively straightforward. Further, it is potentially feasible to remove introns, which could either eliminate the potential for a skipped cassette exon or intron retention event. Of note, deficiency of Bud13 has been shown to be embryonic lethal both in zebrafish⁸ and in *C. elegans*⁹, findings that are consistent with preliminary work of ours in mice. Thus, the organism wide alteration of *Irf7* retention through removal of Bud13 appears not to be feasible.

It is important to note that all of the aforementioned work has been studied at the bulk level. Cell-level insight regarding unproductive splicing could provide meaningful insight into the purpose and regulation of individual unproductive splicing events. It has previously been reported that there is a great deal of variation in splicing patterns among single cells^{10,11}. Can such heterogeneity be found among unproductive splicing events and if so, what regulates this heterogeneity and why does it exist? In saying this, it remains uncertain how feasible it is to perform such analyses. Inherent to the sequencing of a single cell is the low amount of starting material, which can restrict analysis to only highly abundant transcripts¹².

While this can be challenging in typical gene expression analyses, it is even more problematic when studying isoform abundance in non mutually-exclusive cases as non-dominant isoforms tend to be expressed at low levels and thus, are especially susceptible to “drop-out”¹³. Further, low sequence coverage common in scRNA-seq data makes it difficult to accurately characterize splicing variations in low abundant transcripts. Such a problem might be alleviated through the development of machine learning algorithms like the recently published DARTS¹⁴, which offers the ability to better characterize splicing variations in transcripts with minimal coverage. Technological advances in both library preparation and sequencing methods, as well as new computational strategies that are tailored to the intricacies of scRNA-seq (namely high technical noise, high processing requirements, and misquantification of poorly expressed isoforms due to lack of coverage) offer a great deal of promise¹⁵⁻¹⁷.

Finally, there are potential therapeutic consequences to unproductive splicing. This can be exemplified through our work with *Oas1g*, which shows the unproductive splicing event in *Oas1g* acts to mitigate *Oas1g* gene expression, and thus the antiviral response. While the evolutionary history of *Oas1* is quite volatile, epitomized by the fact that in humans there is only one copy of *Oas1* that differs significantly in sequence from the eight *Oas1* mouse paralogues, human *Oas1* contains a similar unproductive splicing event. It might be possible to force productive splicing of this human *Oas1* transcript using an antisense oligonucleotide (ASO) to block the unproductive splice site. ASOs are short, synthetic, single-stranded oligodeoxynucleotides that can alter RNA splicing by base pairing to *cis* sequences¹⁸. The clinical success of drugs like Eteplirsen, an ASO which negates a frameshift mutation in *DMD* by causing the skipping of exon 51 and thus shifting the transcript back into frame, has

created a great deal of interest in ASOs as potential therapeutics¹⁹. As in mice, genetic variation that dampens OAS1 activity has been shown to lead to susceptibility to viruses like West Nile virus²⁰ and Epstein Barr virus²¹. Forcing productive splicing with an ASO could improve an anti-viral response and thus, potentially be used as an anti-viral agent.

References

1. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613 (2012).
2. Lareau, L. F., Brooks, A. N., Soergel, D. A., Meng, Q. & Brenner, S. E. The coupling of alternative splicing and nonsense-mediated mRNA decay. (2007).
3. O'Connor, B. P. *et al.* Regulation of toll-like receptor signaling by the SF3a mRNA splicing complex. *PLoS Genet.* **11**, e1004932 (2015).
4. Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F. & Kollias, G. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398 (1999).
5. Frankiw, L. *et al.* Bud13 Promotes a Type I Interferon Response By Countering Intron Retention in Irf7. *Mol. Cell* **73**, 803–814 (2019).
6. Kahles, A., Ong, C. S., Zhong, Y. & Ratsch, G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics* **32**, 1840–1847 (2016).
7. Wan, Y. & Larson, D. R. Splicing heterogeneity: separating signal from noise. *Genome Biol.* **19**, 86 (2018).
8. Fernandez, J. P. *et al.* RES complex is associated with intron definition and required for zebrafish early embryogenesis. *PLoS Genet.* **14**, e1007473 (2018).
9. Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **98**, 218–223 (2001).
10. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
11. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
12. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
13. Arzalluz-Luque, Á. & Conesa, A. Single-cell RNAseq for the study of isoforms—how is that possible? *Genome Biol.* **19**, 110 (2018).
14. Zhang, Z. *et al.* Deep-learning augmented RNA-seq analysis of transcript splicing. *Nat. Methods* **16**, 307–310 (2019).

15. Song, Y. *et al.* Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol. Cell* **67**, 148–161 (2017).
16. Huang, Y. & Sanguinetti, G. BRIE: transcriptome-wide splicing quantification in single cells. *Genome Biol.* **18**, 123 (2017).
17. Welch, J. D., Hu, Y. & Prins, J. F. Robust detection of alternative splicing in a population of single cells. *Nucleic Acids Res.* **44**, e73 (2016).
18. Rinaldi, C. & Wood, M. J. A. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* **14**, 9–21 (2018).
19. Levin, A. A. Treating Disease at the RNA Level with Oligonucleotides. *N. Engl. J. Med.* **380**, 57–70 (2019).
20. Lim, J. K. *et al.* Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. *PLoS Pathog.* **5**, e1000321 (2009).
21. Liu, X. *et al.* A functional variant in the OAS1 gene is associated with Sjögren's syndrome complicated with HBV infection. *Sci. Rep.* **7**, 17571 (2017).