HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

Committee on Immunology

have examined a dissertation entitled

The ALS-linked gene TDP-43 regulates IFNβ expression through a novel mechanism of 3' UTR-mediated promoter cis-regulation

presented by Raquel Payzant Deering

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: _____

Typed Name: Dr. Ramnik Xavier

Signature: _____

Typed Name: Dr. Paul Anderson

Signature: _____

Typed Name: Dr. Katherine Fitzgerald

Dr. Michael Carroll, Program Head

Date: November 16, 2012

Dr. David Lopes Cardozo, Director of Graduate Studies

The ALS-linked gene TDP-43 regulates IFNβ expression through a novel mechanism of 3' UTR-mediated promoter cis-regulation

A dissertation presented

by

Raquel Payzant Deering

to

The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Immunology

Harvard University
Cambridge, Massachusetts

November 2012

Dissertation Advisor: Dr. Nir Hacohen   Raquel P. Deering

The ALS-linked gene TDP-43 regulates IFNβ expression through a novel mechanism of 3' UTR-mediated promoter cis-regulation

Abstract

The TAR DNA-binding protein (TDP-43) is a heterogeneous nuclear ribonucleprotein that is involved in multiple stages of RNA processing. Mutations in the TDP-43 gene and mislocalization of TDP-43 protein have been implicated in a growing number of neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). Here, we show that TDP-43 negatively regulates innate immune gene expression in response to RNA virus sensing. Perturbation of TDP-43 protein expression leads to an increase in antiviral gene expression in a variety of human and mouse cells. Crosslinked RNA immunoprecipitation (CLIP) experiments revealed that TDP-43 binds to type I interferon (IFN) and interferon stimulated gene (ISGs) transcripts. Using massively parallel 3' UTR reporter assays coupled with high throughput sequencing (MPRA-seq), we identified polyadenylation signal sequences in the 3' UTRs of innate immune genes to be specifically regulated by TDP-43. Surprisingly, IFN and ISG mRNA decay rates are faster in TDP-43-perturbed cells. Using a metabolic labeling approach to measure nascent transcript generation, we found that perturbation of TDP-43 expression leads to an increase in antiviral gene transcription rates. Additionally, RNA polymerase II (pol II) chromatin immunoprecipitation (ChIP) confirmed that there is greater pol II occupancy

on innate immune genes when TDP-43 is depleted. Although TDP-43 perturbation has no effect on an isolated IFNβ promoter reporter, we found that TDP-43 inhibits IFNβ promoter activity when the IFNβ 3' UTR sequence is inserted downstream of the IFNβ promoter element, suggesting a novel mechanism of 3'UTR-mediated promoter cis-regulation.

Table of Contents

Experimental Aims:

CHAPTER 1: The ALS-linked gene TDP-43 regulates IFNβ expression through a novel mechanism of 3'UTR-mediated promoter cis-regulation……………………………………………………..42

CHAPTER 2: The RNA binding proteins Fus and TDP-43 play roles in innate immunity and ALS disease pathophysiology……………………………………………………...99

CHAPTER 3: Msi2 maintains self-renewal through TGFβ signaling in hematopoietic stem cells………………....125

List of Figures and Tables

Tables:

Figures:

Introduction:

Chapter 1:

Chapter 2:

Chapter 3:

Appendix A:

Appendix B:

Isabel H.P. Deering, PhD
(1923-1997)

Dedication: I would like to dedicate this doctoral dissertation to my late grandmother, Dr. Isabel Deering. I attribute my early and continued interest in science to her influence and life-long love for learning. In a generation when few women pursued doctoral degrees, she managed the challenges of her time with a gentle deliberation. As a biochemist, a mother to five children, and a devoted caretaker of her goat farm, her life continues to inspire me.

Dr. Deering died from complications of amyotrophic lateral sclerosis in 1997. Having watched several of her siblings and other family members succumb to the disease before her, she accepted her fate with her typical understated grace. This nonetheless imprinted me with a responsibility to try to improve the lives of those suffering from devastating diseases.

Acknowledgements

I would first like to thank my advisor, Dr. Nir Hacohen, for his support over the course of my PhD work. Nir is unique in that he inspires his students to tackle big and exciting questions, even at the beginning of their training. I have learned so much from him and am a much better scientist as a result of his mentorship. On a personal note, Nir has shown tremendous understanding of and encouragement for my life decisions – in particular to have children in graduate school. It is this sort of support that makes all the difference for young scientists that are trying to balance a family and career in science.

I would also like to thank all past and present members of the Hacohen lab, as well as the larger Broad community, for their technical and personal support over the course of this work. In particular, thanks to Raktima Raychowhurdy, Thomas Eisenhaure, Tarjei Mikkelsen, Li Wang, Xiaolan Wang, Alexandre Melnikov, Schraga Schwartz, Mordecai Choder, Aviv Regev, Alexander Shishkin, Mitchell Guttman, Phil DeJager, Kirsten Lindblad-Toh, and The RNAi Consortium,

Very special thanks to my thesis committee members, Drs. Jon Kagan, Phillip Sharp, and Kai Wucherpfennig for their guidance and suggestions.

Lastly, I am particularly thankful to my husband, Dr. Benjamin Olenchock, and my two wonderful sons, Winston and Adam, for supporting me in their assorted ways – and for continually reminding me what is really important in life.

Introduction

This work explores the regulation of the immune response to viral pathogens by RNA binding proteins. The larger goal of the Hacohen lab is to identify and dissect gene networks that are responsible for innate immune regulation. A specific effort has been made to explore RNA binding proteins that have previously unknown regulatory functions in our system. My thesis work focuses on one such RNA binding protein, TDP-43, that was also identified to be an amyotrophic lateral sclerosis (ALS) disease allele during the course of my studies. This introduction will present the viral sensing pathway that our work is focused on, as well as provide a background for the role of TDP-43 and innate immunity in ALS. Finally, the RNA biochemistry approaches that were developed in order to adequately study the mechanisms of function of RNA binding proteins are introduced.

Section I: The innate immune response to RNA viruses

The human body is equipped with several layers of defenses to evade infection by bacteria and viruses. Broadly, there are two main arms of the immune system. The innate immune system utilizes invariant, germ-line encoded receptors that recognize broad molecular features on pathogens and can respond very rapidly, within minutes, to pathogen invasion. Conversely, adaptive immunity uses a complex system of antibody selection to elicit very specific responses against all potential pathogen components, and in turn develops a long-term immunological memory. This latter process, though more effective at eradicating infectious

agents, functions over a course of days rather than minutes. Therefore, both innate and adaptive immune players are required to generate a seamless response to pathogen insults and to curb the damage caused by future encounters with a given pathogen (Janeway's Immunobiology for review).

The work in this thesis will focus on the innate immune response to viral infections specifically. While there are specialized cell types that are categorized as either innate (macrophages, dendritic cells) or adaptive (T cells, B cells) by the roles that they play in maintaining healthy immunity, the cellular response to virus is interesting because all cells of the body must be equipped with viral sensing machinery and downstream signaling components that are required to combat a viral infection and limit the spread of viral replication.

How viruses are sensed

One of the fastest immune surveillance interactions is the recognition of conserved pathogen-associated molecular patterns (PAMPs) by the host's germline-encoded pattern recognition receptors (PRRs). PAMPs tend to be molecular structures of a pathogen that are so critical for the species' survival that they cannot be easily mutated without compromising reproductive success. PRRs, on the other hand, are invariant receptors on the host side that recognize these pathogen components and rapidly activate a complex signaling pathway that leads to the engagement of either an inflammatory or an antiviral gene program. The resulting production of cytokines and chemokines eliminates

infected cells and signals to the adaptive immune system to initiate long-lasting immunity specific for a given pathogen (Medzhitov 2007).

The two arms of an innate immune response are 1) **pro-inflammatory** and 2) **antiviral**. Pro-inflammatory cytokines primarily function to initiate and coordinate innate immune responses by recruiting professional antigen presentation cells, particularly macrophages and dendritic cells, to attack infected cells, process pathogenic antigens, and prime T cells to engage antibody-mediated immunity. Conversely, the antiviral pathway is initiated by the production of type I interferons, specifically IFNα and IFNβ (IFN). These cytokines bind to neighboring cells directly through IFNAR1/2, which signals to activate the induction of potent interferon stimulated genes (ISGs). Both IFNs and ISGs then locally induce transcriptional and post-transcriptional changes in all neighboring cells to augment an "antiviral state." By changing cellular processes such as mRNA translation rates, viral replication is inhibited. Furthermore, IFNs and ISGs induce upregulation of MHC class I molecules and directly activate various components of adaptive immunity (Kawai and Akira 2006).

Toll-like receptors

Toll like receptors (TLRs) are essential pathogen sensors that are, at least minimally, expressed by all cells of the body to provide an initial barrier against infectious agents. TLRs are type I membrane glycoproteins and constructed of an extracellular leucine rich repeat (LRR) region for the recognition of PAMPs, linked to a cytoplasmic Toll/interleukine-1 receptor (TIR) domain that mediates

downstream signaling. There are currently 13 known TLRs in mice and 10 in humans. Each TLR recognizes a unique component of bacteria, virus, or fungi. TLR1, 2, and 4-6 are cell surface receptors that recognize PAMPs from bacteria and fungi, while TLR3, 7-9 reside within endosomes and recognize nucleic acids from viruses (Takeuchi, 2010). Single stranded (ss) and double stranded (ds) RNA and DNA from viruses are recognized by distinct TLRs. For example, dsDNA from the herpes simplex virus (HSV) is sensed by TLR9 (Krug 2004), whereas TLR3 is a critical viral dsRNA sensor (Alexopoulou 2001).

The recognition of ligands by TLRs leads to a rapid induction of either pro-inflammatory or antiviral gene network responses. Activation of TLRs1, 2, 4-6 leads to the production of primarily pro-inflammatory cytokines, such as tumor necrosis factor (TNF). Conversely, triggering TLR3, 7, and 9 results in an antiviral response that leads to type I interferon production (Kumar 2011). TLR signaling is mediated by the recruitment of a limited set of TIR domain-containing adaptor molecules to TLR cytosolic TIR domains. Key adaptor molecules include MyD88, TRIF, TIRAP, and TRAM. The dependency on specific adaptor molecules by each TLR for signaling is complex, but can be distilled down to the activation of a core set of master transcription factors, such as IRF3/7 and NF-kB, that leads to the engagement of either a pro-inflammatory or antiviral response (Kumar 2009). Professional immune surveillance cells, such as dendritic cells, express high levels of all TLRs to mediate the rapid induction of cytokines and increase

antigen presentation of microbial peptides to T cells, which in turn activates antigen-specific immune responses (Hemmi and Akira 2005).

RIG-I-like receptors

RIG-I-like receptors (RLR) include RIG-I, MDA5, and LGP2 and, unlike TLRs, are localized to the cell cytosol to detect RNA from infecting RNA viruses. Again, activating RLRs can lead to the induction of both pro-inflammatory and antiviral gene network responses. RIG-I and MDA5 contain N-terminal CARD domains that are required for downstream signaling. LGP2 lacks a CARD domain, but all three have a DExD/H-box RNA helicase domain that is used for ligand recognition (Wilkins and Gale 2010).

RIG-I recognizes members of the paramyxoviridae family (Newcastle Disease virus, vesicular stomatitis virus), the flaviviridae family (Hepatitis C virus), as well as short dsRNA and ssRNA with a 5' triphosphate. MDA5 recognizes chemically synthesized dsRNA and works with RIG-I to elicit appropriate responses to certain viral pathogens, such as West Nile Virus (Kumar 2011). LGP2 is thought to improve the accessibility of viral RNA to RIG-I and MDA5 recognition to induce a more robust antiviral response (Satoh 2010).

Table 0.1: Innate immune sensors and their ligands

Receptor	Location	Ligand
TLRs		
TLR1/2	Cell surface	Triacyl lipopeptides
TLR2	Cell surface	Peptidoglycan, zymosan
TLR3	Endosome	ssRNA, dsRNA, poly I:C
TLR4 (CD14)	Cell surface	lipopolysaccride
TLR5	Cell surface	Flagellin
TLR6/2	Cell surface	Diacyl lipopeptides
TLR7	Endosome	ssRNA, imiquimod
TLR8	Endosome	ssRNA, imiquimod
TLR9	Endosome	dsDNA, CpG motifs
TLR11 (mice)	Endosome	Uropathogenic bacteria,
TLR 13 (mice)	Endosome	23S rRNA
RLRs		
RIG-I	Cytoplasm	Short dsRNA (up to 1 kb), 5'triphosphate, short poly I:C
MDA5	Cytoplasm	Long poly I:C, ssRNA,
LGP2	Cytoplasm	

See Kumar et al., 2011 for review

During an active viral infection, viral RNA enters the cytosol and is sensed by RLRs. The CARD domains of RIG-I and MDA5 are conformationally activated and interact with the CARD-containing adaptor molecule, IPS-1. IPS-1 then localizes to mitochondria and peroxisomes to induce downstream signaling. The master regulatory transcription factors, IRF1, 3, 7 and NF-kB, are then activated and turn on the antiviral gene expression network (Kumar 2011).

Type I interferon and interferon stimulated gene regulation

The innate immune response to viral PAMPs by TLRs and RLRs involves the critical activation of type I interferon and interferon stimulated genes. In the

absence of intact interferon signaling, such as in the interferon alpha receptor (IFNAR) knockout mouse, the ability to achieve an antiviral state is abolished and susceptibility to viral infection is a significant problem (Muller 1994). When viral sensors are triggered by pathogen components, the initial cytosolic signaling events culminate in the activation of a core set of critical transcription factors, namely IRF3/7, NF-kB, and ATF-2/c-Jun. These transcription factors normally reside in the cytosol but, when activated, translocate to the nucleus and assemble onto the virus-inducible enhancer of the IFNβ gene promoter. Through a coordinated mechanism of recruitment, these transcription factors are joined by the transcriptional co-activators, p300 and CREB-binding protein (CBP) (Thanos and Maniatis 1995, Wathelet 1998). IFNβ gene transcription is initiated and IFNβ protein is secreted from the cell. In both an autocrine and paracrine fashion, type I interferons then signal through the IFNAR1/2 to amplify the antiviral response and activate genes that contain an interferon-stimulated response element (ISRE; sequence: 5'-GAAANNGAAAG/CT/C-3'). Approximately fifty genes contain an ISRE and are activated in response to IFN (Darnell 1994, Honda 2005).
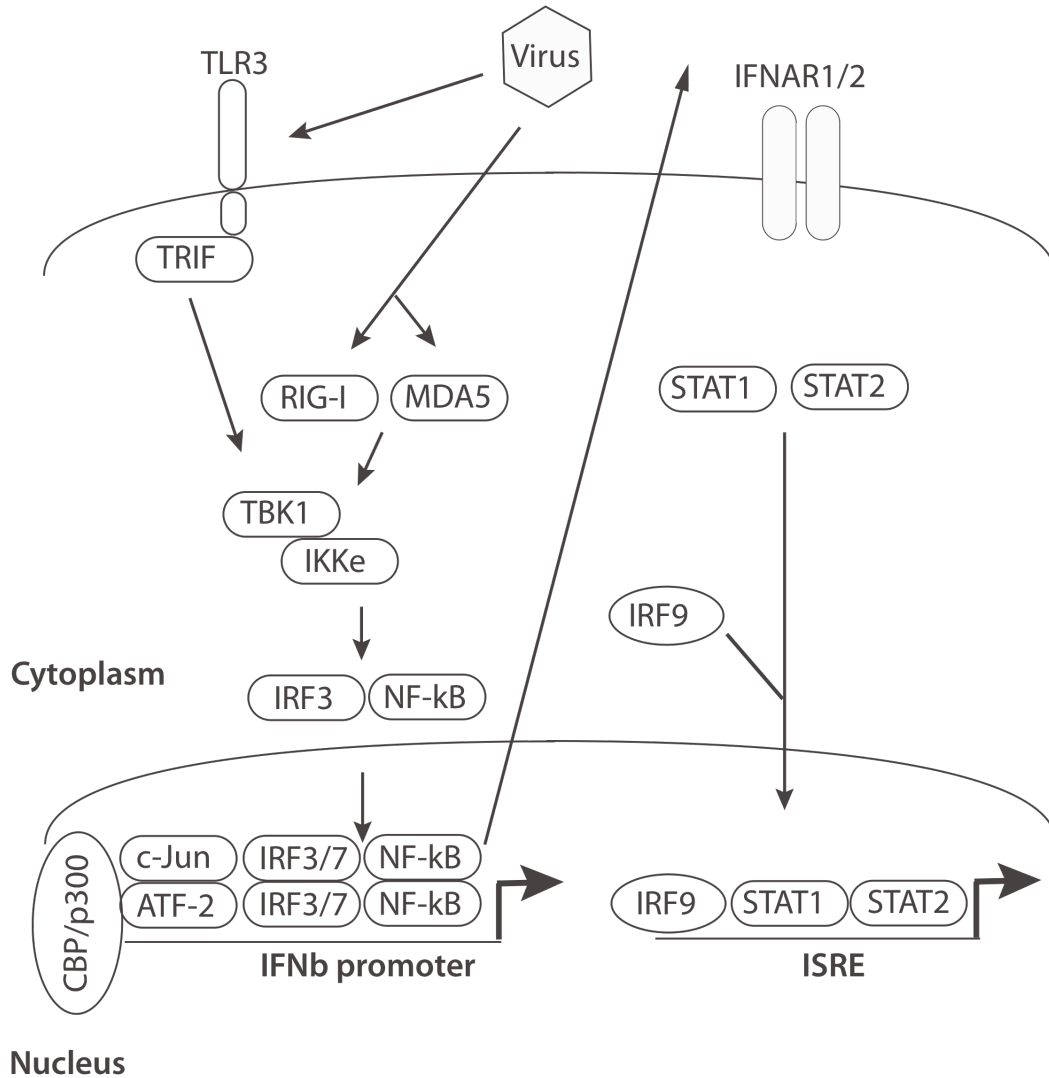
Figure 0.1: Type I interferon signaling pathway in response to viral RNA sensing. Toll-like receptors (TLRs) or cytosolic nucleic acid receptors (RIG-I, MDA5) sense viral RNA and signal through adaptor molecules to activate key transcription factors (IRFs, NF-kB) that are required for IFNβ gene expression. IFNβ protein is secreted from the cell and signals though interferon alpha-receptors 1 and 2 (IFNAR1/2) to amplify the antiviral gene response by activation transcription factors that turn on interferon-stimulated response element- (ISRE)-containing genes.

Negative regulators of the type I interferon response

Interferons and ISGs are inherently toxic to cells and their expression is limited to

an active antiviral response. Many essential negative regulators of the antiviral

signaling pathway exist to prevent uncontrolled activation of this pathway, which

8

could lead to irreparable tissue damage or autoimmunity. Known negative regulators are active at all stages of interferon induction and signaling. For example, ISG56 (IFIT1) is one of the first genes to be upregulated in response to type I interferon or viral infection. ISG56 was found to interrupt the physical association of the cytosolic signaling mediators IPS-1, STING, and TBK1 with each other, thus preventing the activation of a key transcription factor for type I interferon genes expression, IRF3 (Li 2009). Similarly, several other factors that interrupt antiviral signaling upstream of transcription have been identified, such as SIRPα (Dong 2008), TRIM30α (Shi 2008), and A20 (Saitoh 2005).

Signal-dependent phosphorylation events are another level that the induction of type I interferons is regulated. IκB-kinase-ε (IKKε) and TANK-binding kinase-1 (TBK1) function to control the phosphorylation of two transcription factors that are required to augment IFNβ gene activation, IRF3 and NF-kB (Fitzgerald 2003). The NEMO-related protein and ALS disease gene, optineurin (OPTN), negatively regulates the function of TBK1 by physically associating with TBK1 and preventing the phosphorylation of IRF3 (Mankouri 2010).

Transcriptional repression of IFNβ is not the only mechanism used to control the activity of the antiviral gene expression network. The IFNβ mRNA itself contains sequences in its coding region and 3' untranslated region (UTR) that are targeted by RNA binding proteins for decay. Both sequences are approximately 20 nucleotides in length and are significantly enriched in the nucleotides A and U,

9

specifically AUUUA (Raj and Pitha 1993). Such A-U rich elements, or AREs, are found in the 3' UTRs of several other cytokines as well, including TNF and IL-1β, and are part of a critical mechanism to rapidly modulate cytokine protein levels (Stoecklin and Anderson 2006).

Section I summary

The antiviral signaling pathway is a key defense against viral pathogens. Despite the identification of many negatively regulators of type I interferon induction, it is clear than more remain to be discovered. This thesis presents a project that describes a novel negative regulator of the IFN signaling pathway, TDP-43. This work defines where TDP-43 functions as an inhibitor of the antiviral pathway as well as the mechanism by which TDP-43 regulates type I interferon gene activity.

Section II: The RNA binding protein TDP-43

TDP-43 is a multi-functional RNA binding protein

In 2009, our lab published the results of an RNA interference (RNAi) screen in which approximately 150 regulators of innate immunity were systematically knocked down in primary bone marrow dendritic cells (BMDC). The set of targeted known and putative regulators included transcription factors, chromatin modifiers, and RNA binding proteins. Gene expression profiles following knockdown and challenge with TLR ligands were assessed, and the resulting regulatory networks identified the RNA binding protein Fus to be a negative regulator of antiviral gene expression. Depletion of Fus resulted in an increase in

the expression of approximately 50 antiviral genes but had no effect on the pro-inflammatory gene pathway (Amit 2009). Based on this work, we became interested in another RNA binding protein that shares sequence homology with Fus, called TDP43. TDP-43 (gene: Transactivation response DNA binding protein of 43 bp, TARDBP) is a 43 kD nuclear RNA binding protein associated with a variety of cellular functions, particularly those related to RNA metabolism. Subsequent to the publishing of the screen, I found that, like Fus, TDP-43 also regulates antiviral gene expression. Interestingly, my work and others have demonstrated that TDP-43 and Fus proteins associate with each other and exist in a large ribonucleoprotein complex (Kim 2010, Freibaum 2010). My thesis work describes how TDP-43 mechanistically functions to control the antiviral response.

TDP-43 is an amyotrophic lateral sclerosis disease gene

TDP-43 was found to be the most highly enriched protein in cytosolic ubiquitinated protein aggregates in the central nervous system of individuals with sporadic and familial amyotrophic lateral sclerosis (ALS) and frontotemporal lobar dementia (FTLD) (Arai 2006, Neumann 2006). In addition to the wild type protein, truncated forms of TDP-43 and hyperphosphorylated forms were also detected. Direct sequencing of the TARDBP gene in patients with these neurodegenerative diseases revealed approximately fifty causative point mutations (Kabashi 2008, Sreedharan 2008). Now, roughly 5% of sporadic and familial ALS is attributed to mutations in TDP-43 (Mackenzie 2010). TDP-43 protein aggregates have more recently been identified in other

neurodegenerative diseases, including Alzheimer's Disease (AD) (Galimberti and

Scarpini 2010) and chronic traumatic encephalopathy (McKee 2010). An

increase in TDP-43 expression has also been found in the central nervous

system of horses that spontaneously develop an ALS-like neurodegenerative

disease (El-Assaad 2012). Collectively, neurodegenerative diseases with TDP-43

protein pathology are now referred to as "TDP-43 proteinopathies."

TDP-43 has a variety of cellular functions and has the potential to regulate

immunity at a number of locations along the antiviral signaling and gene

expression pathway. Previously, TDP-43 has been reported to repress

transcription, direct splice site selection, traffic mRNA to and from the nucleus,

control gene expression by stabilization or destabilization of RNA transcripts, and

regulate microRNA expression. However, the significance of these functions with

regard to disease pathophysiology is unclear.

TDP-43 is a predominantly nuclear protein (414 amino acids in length) that is

encoded by the gene TARDBP, located on chromosome 1. TARDBP has six

exons and encodes up to eleven isoforms, with the most abundant being the 43

kD isoform (Wang 2004). TDP-43 expression is ubiquitous across cells of the

body in both rodents and humans (Buratti 2001). TARDBP is also a very

evolutionarily conserved gene and has found to be required for pluripotency (Wu

Genesis 2010). The N-terminus of TDP-43 has been shown to interact with

transcription factors (Swarup 2011) and contains a nuclear localization

sequence. TDP-43 contains three putative RNA binding domains: the RNA

recognition motif 1 (RRM1), RRM2, and a C-terminal arginine-glycine-glycine-rich

repeat region (RGG) (Da Cruz and Cleveland 2011). Most of the ALS dominant

missense mutations are localized in the RGG (See Figure 1). The RRM1 of TDP-

43 is indispensible for its binding to UG repeat sequences in RNA transcripts

(Ayala 2006).
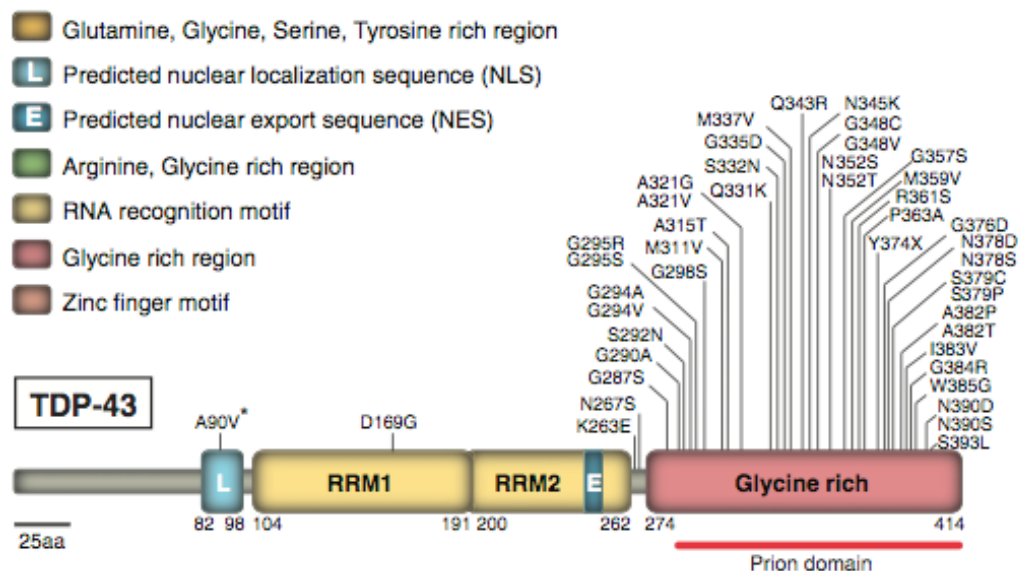
Figure 0.2: TDP-43 structural domains and mutations linked to ALS. (From Da Cruz and Cleveland, Neurobiology of Disease, 2011),

Proteomic analyses of TDP-43 interaction partners show that TDP-43 is heavily

enriched for associations with many RNA splicing factors (e.g. SF2, TIA-1),

translation machinery (e.g. PABP2, eEF1A1 and eIF3F), and other proteins that

regulate mRNA abundance including those in the nonsense mediated decay

pathway (e.g. UPF1) and mRNA destabilization (e.g. Musashi 2). Importantly, the

protein association map for two key TDP-43 ALS disease alleles (A315T and

M337) is the same as for the wild type TDP-43 protein (Kim 2010, Freibaum 2010).

TDP-43 is largely nuclear, although low levels of the protein can be found in the cytosol under normal conditions (Ayala 2008). In the nucleus, TDP-43 is concentrated in perichromatin fibrils around euchromatin, which are sites of active transcription and co-transcriptional splicing, but is absent from transcription-free Cajal bodies (Lafarga 2009). It has been noted that TDP-43 can aggregate in cytosolic stress granules (SG) under induced conditions of stress, and that the expression of disease-linked TDP-43 mutants increased the accumulation of TDP-43 in SGs (Liu-Yesucevitz 2010).

Some of the known functions of TDP-43 are outlined here:

TDP-43 is a transcriptional repressor

Although most of the more recent papers on TDP-43 focus on its RNA binding functions, TDP-43 was originally described as a DNA binding protein that regulates the human immunodeficiency virus (HIV) type I transactivation response (TAR) regulatory element (Ou 1995). TAR is an activation domain in the HIV-1 long terminal repeat promoter region that is required for the activation of gene expression. In Ou et. al., proteins in a HeLa cell extract were UV crosslinked to a labeled TAR DNA probe and TDP-43 was found to bind to the TAR DNA sequence. Interestingly, RNA that is transcribed from the TAR region is also essential for HIV-1 gene expression but TDP-43 was not found to be

specifically associated with TAR RNA. An *in vitro* transcription assay confirmed that TDP-43 was able to repress transcription from the HIV-1 long terminal repeat, but not from an adenovirus promoter.

Conversely, a more recent study found that TDP-43 binds to and is a co-activator of the NF-kB subunit, p65 (Swarup 2011). When TDP-43 was overexpressed, the activity of an NF-kB-dependent promoter reporter was increased significantly, both at baseline and in response to LPS stimulation. The N-terminus of TDP-43 was required for this activity, but none of the RNA binding regions of the protein had an effect on transcriptional activation. To rule out a dominant negative effect of overexpressing TDP-43, the protein was depleted with siRNA and the NF-kB-dependent promoter reporter activity was reduced in the absence of TDP-43. The authors correlate this function for TDP-43 with the observation that NF-kB-dependent genes, such as TNF, are abnormally upregulated in the spinal cords of ALS patients.

TDP-43 was also identified to be a key protein, along with Fus/TLS, in the Microprocessor complex that includes Drosha and DGCR8 (Gregory 2004). This complex mediates microRNA (miRNA) biogenesis. Perturbation of TDP-43 is associated with changes in miRNA expression, and TDP-43 is directly involved in the production of a subset of precursor miRNAs (pre-mirRNAs) (Kawahara 2012).

TDP-43 directs splice site selection

One of the first roles ascribed to TDP-43 was its ability to bind to the $(TG)_m T_n$ polymorphism at the 3' splice site of the cystic fibrosis (CF) transmembrane conductance regulator (CFTR) gene. One of the major causes of CF is an aberrant skipping of exon 9 in the CFTR gene, thought to be partially caused by a higher frequency of TG repeats at the 3' splice site among diseased individuals. TDP-43 has been show by systematic evolution of ligands by exponential enrichment (SELEX) and RNA pull down experiments to preferentially bind to RNA transcripts containing UG repeats (Buratti 2001, Ayala 2005). In the presence of "unfavorable" CFTR $(UG)_m T_n$ sequences, TDP-43 was found to promote exon 9 skipping, and the depletion of TDP-43 protein from CF patient cells resulted in improved exon 9 inclusion (Ayala 2006).

Fiesel et. al. recently published a paper that explored the effects of TDP-43 depletion on splicing globally using Affymetrix exon arrays. In TDP-43-depleted neural cell lines, a key exon junction complex protein, ribosomal S6 kinase 1 (S6K1) Aly/REF-like target (SKAR) is alternatively spliced (Fiesel 2012). SKAR normally functions to recruit S6K1 and facilitates the first round of translation. When TDP-43 is absent, the alternatively spliced form of SKAR enhances this function, causing a global increase in translation rates and cell size. We have likewise observed a significant increase in total cell size and protein content per

cell in our own work. The authors postulate that a general defect in regulating protein translation might contribute to the marked protein aggregation that is observed in ALS and FTLD.

A more broad role for TDP-43 in regulating splicing was reported in a study in which TDP-43 was depleted *in vivo* from mouse striatum and the resulting transcriptome was analyzed by RNA sequencing (Polymenidou 2011). The depletion of TDP-43 resulted in a downregulation of ~240 genes. Additionally, this RNA was applied to splicing-sensitive microarrays and ~800 genes were found to be misspliced when TDP-43 was depleted. Rather than directing exon skipping, the depletion of TDP-43 from brain cells in this studied largely resulted in aberrant exon inclusion. The authors also performed crosslinked RNA immunoprecipitation (CLIP) followed by deep sequencing for endogenous TDP-43 protein from mouse brains to capture the RNA transcript sequences that TDP-43 binds. When the transcriptome and CLIP sequencing data sets were correlated with each other, it was observed that ~100 of the most downregulated genes from TDP-43-depleted cells are direct TDP-43 targets, and that TDP-43 binds these genes primarily at intronic regions. Interestingly, these TDP-43 target genes contained very long introns, suggesting a relationship between intron length and stabilization by TDP-43.

This paper focused on the effect of TDP-43 depletion on downregulated genes. Significantly, antiviral genes are overrepresented in the list of genes that were

upregulated *in vivo* when TDP-43 was depleted from the brain. Since there was less than 20% overlap between the upregulated genes and the TDP-43 CLIP binding associations (compared to 75% overlap between downregulated genes and TDP-43 binding), the authors postulate that many of the these genes are upregulated in the absence of TDP-43 by an indirect mechanism. As a side point, TDP-43 primarily bound these upregulated genes at their 3' UTRs, in contrast to the intronic binding observed in the downregulated gene set. Looking at these data from an innate immune perspective, we realized that the downstream effect of dysregulated type I interferon expression would be an increase in the activity of the entire antiviral pathway. Since our work has found that TDP-43 negatively regulates IFNβ transcript levels directly, the findings in this paper are particularly relevant to understanding how TDP-43-mediated dysregulation of innate immune pathways might contribute to neurodegeneration.

TDP-43 binds to mRNA and regulates gene expression

Several groups have used a variety of biochemical methods to assess which RNA transcripts are directly bound by TDP-43 in an attempt to better understand how TDP-43 normally functions in the cell. RNA immunoprecipitation (RIP) and UV-crosslinked RNA immunoprecipitation (CLIP) followed by quantitative real time PCR (qPCR) or massively parallel sequencing (seq) have been the most informative methods used to identify TDP-43 RNA targets.

Initial gene expression observations have lead to several gene-focused investigations about TDP-43:mRNA binding. For example, depletion of TDP-43 was found to decrease histone deacetylase 6 (HDAC6) and low molecular weight neurofilament (NFL) expression (Fiesel 2009, Volkening 2009). RIP experiments targeting endogenous TDP-43 found that HDAC6 and NFL mRNA is associated with TDP-43 protein and that TDP-43 binding is enriched at the 3' UTRs (Kim 2010, Volkening 2009).

To evaluate the RNA targets of TDP-43 more broadly, several groups have published RIP- and CLIP-seq datasets. Don Cleveland's lab performed CLIP-seq on endogenous TDP-43 protein from normal mouse brains and found a significant correlation between transcripts that are directly bound by TDP-43 and genes that are stabilized by TDP-43 expression (Polymenidou 2011). The TDP-43 CLIP associations were enriched for genes that are critical for neural physiology. The most enriched binding motif for TDP-43 consisted of GU-repeats, and there was a correlation between the number of GU repeats and the strength of TDP-43 binding. In this study, 93% of the TDP-43 binding sites were within introns along the length of target pre-mRNAs. Conversely, a similar RIP-seq library demonstrates a high enrichment for transcripts that encode genes related to RNA processing, such as polyA binding protein cytoplasmic 1 (PABPC1) and splicing factor arginine/serine-rich 1 (SFRS1), and that TDP-43 preferentially binds these transcripts at their 3' UTRs (Sephton 2011).

Several studies have found that TDP-43 associates with the 3' UTR of its own

RNA transcript (Polymenidou 2011, Sephton 2011, Avendano-Vasquez 2012,

Ayala 2010). Given the toxicity of overexpressing TDP-43, it was suggested that

TDP-43 might autoregulate its expression post-transcriptionally. Indeed, the C-

terimal domain of the TDP-43 protein was found to bind to the TDP-43 transcript

3' UTR and target it for depletion via the exosome-mediated degradation

pathway (Ayala 2011). Later work has delved deeper into the mechanism of

TDP-43 autoregulation. Francisco Baralle's group noted that the overexpression

of TDP-43 results in the activation of a cryptic polyadenylation signal through a

process of alternative splice site selection. The inclusion of this alternative PAS

causes a nuclear retention of the TDP-43 transcript, thereby subjecting it to a

nuclear mRNA decay mechanism. Furthermore, overexpressed TDP-43 sterically

inhibits the polyadenylation complex from recognizing the normally favored PAS,

adding another level to the mechanism of autoregulation (Avendano-Vasquez

2012). Clearly, TDP-43 is capable of functioning at multiple levels of RNA

metabolism to regulate gene expression, even at one locus.

Rodent models of TDP-43 function

To begin to unravel how missense mutations in the gene encoding TDP-43

protein might contribute to ALS disease pathophysiology, several TDP-43

transgenic and knockout animal model systems have been developed over the

past few years. Both gain- and loss-of-function rodent systems have added to our

understanding of the normal and disease mechanisms of TDP-43 function.

The initial rodent models of TDP-43 involved constitutive overexpression of wild type human TDP-43 or ALS disease alleles. The first animals created express TDP-43 under the control of the mouse prion protein promoter (PrP), which is largely active in the brain, spinal cord, and heart (Wegorzewska 2009, Xu 2010, and Stallings 2010). Each group also generated mice that express the A315T and M33V ALS-linked mutants. The level of expression of TDP-43 in the brain in these mice is significantly higher than endogenous levels. All of the mice demonstrated motor deficits and early lethality. Interestingly, overexpressing wild type TDP-43 leads to identical motor deficits and pathology as overexpression of the disease forms. Other groups created similar mouse models using different promoter systems: Thy1 promoter (Shan 2010, Wils 2010), Ca^{2+} /calmodulin-dependent kinase II promoter (Tsai 2010), and a low expression level bacterial artificial chromosome (BAC) transgenic approach (Swarup 2011). A rat model using a BAC for TDP-43 overexpression was also created (Zhou 2010). Despite variations in wild type or ALS-linked TDP-43 expression that was conferred by these different transgenic models, all of the lines resulted in the development of motor neuron disease.

An important observation from transgenic TDP-43 rodent models is that the localization of TDP-43 protein does not appear to directly correlate with disease progression or severity. Mislocalized TDP-43 was initially thought to be required for the development of ALS. Immunohistochemical staining of motor neurons

21

from sporadic and familial ALS individuals with anti-TDP-43 antibodies revealed a cytoplasmic aggregation of normally nuclear TDP-43 protein (Sasaki 2010). However, neural cells in animals that developed ALS-like disease in the aforementioned rodent models largely maintained nuclear TDP-43. Therefore, the toxic effects of TDP-43 that contribute to motor neuron disease are likely due to a nuclear function.

Although TDP-43 remains nuclear in most of the neurons from the various transgenic TDP-43 animal models, large TDP-43-free aggregates still formed in the cell bodies of motor neurons. Morphological analysis of these cytoplasmic inclusions revealed that they contain large numbers of mislocalized mitochondria (Shan 2010, Xu 2010).

Several groups also attempted to create viable constitutive *Tardbp* loss of function mouse lines (Wu 2009, Kraemer 2010, Sephton 2011), but in all instances the animals died in very early embryogenesis. My own work with the Jaenisch lab determined that TDP-43 knock out blastocysts perished around the time of implantation, coinciding with a loss of maternal mRNA (unpublished). Since TDP-43 seems to be important for development, condition knock out models were generated. Philip Wong's lab at Johns Hopkins University has developed several useful inducible TDP-43 knock out lines (Chiang 2010). To test the physiological requirements for normal TDP-43 function, the group crossed *floxed Tardbp* mice with a *Rosa26-ErCre* line in which the Rosa 26

promoter is tamoxifen responsive. The animals became increasingly ill and died

within nine days of tamoxifen induction, showing significant weight loss,

specifically mesenteric fat loss. Further experimentation determined that TDP-43

depletion lead to a rapid reduction in the amount of fat stored within adipocytes,

likely through a TDP-43-directed mechanism of increased fat oxidation.

In order to properly assess the effects of deleting TDP-43 from brain cells, and

how this perturbation might contribute to ALS, several groups are now creating

lines to test the effects of knocking out *Tardbp* in CNS-specific cell types. Again,

these models will need to be conditional as the animals from constitutive Hb9-

ErCre;Tardbp$^{F/F}$ lines are embryonic lethal (Tsao 2012).

An excellent review of the published TDP-43 ALS rodent models to date is

referenced here: Tsao, et al. 2012. Rodent models of TDP-43: Recent advances.

Brain Research, 1462; 26-39.

Section II Summary:

Recently, much attention has been turned to TDP-43 given its apparent role in

multiple debilitating neurodegenerative diseases. TDP-43 mediates a variety of

cellular processes, but as yet it is unclear which of these functions causes

neuron death – or if disease is a result of a combination of dysregulated

functions. My thesis work finds that TDP-43 negatively regulates antiviral gene

expression and, more specifically, that it does so through 3'UTR-mediated

mechanism of transcriptional control. We wondered whether the effect of TDP-43

on immune regulation contributes to ALS pathophysiology. In the next section, the history and key features of ALS will be presented and evidence of innate immune dysregulation in ALS will be explored.

Section III: Amyotrophic lateral sclerosis

In 1869, the revolutionary French neurologist Jean-Martin Charcot first described distinct pathological spinal cord lesions in autopsy material from adult patients that had succumbed to a progressive paralytic disease (Goetz 2000). This disease, Amyotrophic lateral sclerosis (ALS), or Lou Gehrig's disease as it is commonly known in the United States, is the most common adult onset form of motor neuron disease (MND) and has an estimated prevalence of 2 in 100,000 Americans. ALS is associated with selective degeneration and death of upper and lower motor neurons and is characterized by the resulting progressive, severe disabling features including broad muscle paralysis and spasticity. Despite the devastating nature of this paralysis, ALS individuals typically remain cognitively intact throughout disease progression. The time between clinical diagnosis of ALS and patient fatality is, on average, three years, with respiratory muscle failure being the main cause of death. Motor neuron loss is not the only pathologic abnormality in ALS. Advanced disease is also typically accompanied by reactive gliosis, intracytoplasmic neurofilament abnormalities, and axonal spheroids (Leigh and Swash 1991, Gonatas 1992, Hirano 1984). Unfortunately, there is currently no treatment or cure for the ALS despite the causative pathology having been identified almost 150 years ago.

Approximately 10% of ALS cases are inherited in an autosomal dominant fashion (familial ALS; fALS), but the vast majority has no known genetic association (sporadic; sALS). There is no discernible difference in pathophysiology or disease course between familial and sporadic forms of ALS. As the population ages, men and women are seemingly statistically equally affected with the disease, though it was initially thought that men were more over represented. First symptoms usually present between 50-60 years of age and individuals from all races and ethnic backgrounds can be affected. (Reviewed in Cleveland and Rothstein 2001).

SOD1 and ALS

Why motor neurons are specifically targeted in ALS remains unknown, but progress has been made toward understanding which cell types contribute to disease pathology. In 1993, Cu/Zn binding superoxide dismutase 1 (SOD1) was identified as the first ALS disease gene (Rosen 1993). Using linkage analysis, eleven distinct mutations in SOD1 were found in individuals from thirteen families involved in the study. Now, more than 150 ALS-causing mutations have been found in the SOD1 gene and largely involve single amino acid substitutions. SOD1 mutations account for approximately 20% of FALS cases.

SOD1 is a highly conserved antioxidant enzyme located throughout the cell. Structurally, it contains one copper and one zinc binding site, and is critical in catalyzing the disproportiation of superoxide to yield hydrogen peroxide and

dioxygen. In the absence of a conversion of the superoxide reactive oxygen species (or "free radical") into these less harmful products, superoxide becomes highly toxic by aberrantly oxidizing lipids, proteins, and nucleic acids (Valentine and Hart 2003). In fact, superoxide is so effective at functioning as a cellular damaging agent that the superoxide anion is enriched in the lysosomes of phagocytic immune cells to destroy invading pathogens. Furthermore, type I interferons induce the production of reactive oxygen species (Seifert 2010), which suggests a closer connection between the control of interferon activity and the oxidative stress pathway. This connection will be explored further in this thesis in Chapter 2.

Although SOD1 is classically thought of as a very different protein from TDP-43 since it does not have defined RNA binding domains or known to be an inhibitor of innate immunity, it does have some functional features that overlap with TDP-43. For example, SOD1 has been shown to bind the 3'UTR of NFL mRNA transcripts and inhibit expression (Volkening 2009), SOD1 ALS mice overexpress ISGs in their spinal cords (Wang 2011), and we have found that SOD1 negatively regulates type I interferon expression (Chapter 2).

Motor neuron disease is not cell autonomous
One of the greatest developments in understanding ALS pathophysiology using transgenic mice was the finding that motor neuron disease is non-cell autonomous. Chimeric mouse models demonstrated that, when motor neurons

expressing mutant SOD1^{G93A} are surrounded by WT SOD1-expressing cells, neurodegeneration is delayed or avoided entirely (Clement 2003). Removal of mutant SOD1 from motor neurons delays disease onset but is not sufficient to bypass neurodegeneration (Boillee 2006). In an effort to dissect precisely which cell types are responsible for conferring motor neuron death, Yamanaka et. al. created a transgenic system in which mice carrying an SOD1^{G37R} gene flanked by loxP sites were crossed with mice expressing the Cre recombinase in a variety of cell types, allowing for the systematic excision of the mutant alleles from individual brain cell types. This work determined that mutant SOD1 expression in motor neurons modulates disease onset, but the mutant allele expression in microglia conferred the greatest toxicity post-onset (Yamanaka 2008). *In vitro* co-culture experiments using mouse cell lines have found that astrocytes are also toxic to wild type motor neurons when they express SOD1 disease alleles, and this neurotoxicity is more pronounced when both astrocytes and motor neurons harbor the SOD1 mutation (Di Giorgio 2007). Elegant follow-up of this work demonstrated that human embryonic cell line-derived motor neurons are similarly subject to *in vitro* cell death by microglial cells that were derived from ALS patient fibroblasts harboring either SOD1 or TDP-43 mutations using induced pluripotent stem cell (iPS) technology (Di Giorgio 2008, Bilican 2012).

Other genes linked to ALS

There is a long gap in time between the identification of SOD1 as an ALS disease gene and the discovery of other ALS-linked alleles. However, several

new ALS genes have been reported in rapid succession, in particular C9ORF72, Fus/TLS, TARDBP (TDP-43), OPTN (optineurin), and ATXN2 (ataxin-2) (See Table II). Whereas mutations in SOD1 are associated exclusively with fALS, dominant missense mutations in the gene encoding TDP-43 have been identified in both sporadic and familial ALS cases (Kabashi 2008, Sreedharan 2008, Lagier-Tourenne and Cleveland 2009).

Many of the genes that are mutated in ALS involve RNA processing. For this reason, the neurodegeneration field is questioning whether motor neuron death is, in part, the result of aberrant RNA metabolism. Furthermore, we noted that several of the key ALS disease genes are also involved in innate immune regulation. Our lab has found that TDP-43 and Fus are negative regulators of the antiviral gene program. Optineurin is a well-characterized negative regulator of interferons, and ALS-related optineurin disease alleles are defective in their ability to inhibit IRF3 activation, leading to uncontrolled type I interferon activation (Sakaguchi 2011). Interestingly, high levels of antiviral gene expression was identified in spinal cords from the G93A SOD1 mouse model of ALS (Wang 2011), and I have found that perturbation of SOD1 expression in HEK293T cells leads to a spontaneous increase in type I interferon gene activity (See Chapter 2).

Table 0.2: Known ALS disease genes

Gene	Protein	Proportion of inherited ALS	RNA processing role?	Innate immune role?	Reference
SOD1	Superoxide dismutase-1	20%	Yes	Yes	Rosen 1993
C9ORF72	C9ORF72	30-40%	Yes	No	DeJesus-Hernandez 2011, Renton 2011
TLS	Fus/TLS	5%	Yes	Yes	Kwiatkowski 2009, Vance 2009
TARDBP	TDP-43	5%	Yes	Yes	Sreedharan 2008
OPTN	Optineurin	1-2%	No	Yes	Maruyama 2010
VCP	Valosin-containing protein	1-2%	No	No	Johnson 2010
TAF15	TAF15	1%	Yes	No	Couthouis 2012
UBQLN2	Ubiquilin-2	<1%	No	No	Deng 2011
ANG	Angiogenin	<1%	Yes	No	Greenway 2006
ATXN2 expansions	Ataxin-2	?	Yes	No	Elden 2010
PFN1	Profilin1	?	No	No	Wu 2012
RBM45	RBM45	?	Yes	No	Collins 2012

Links between innate immune alterations and neurodegenerative diseases

Although the brain is an "immune privileged" tissue, the neurons and structural cells of the brain maintain innate immune properties that can generate an inflammatory environment. For example, glial cells have been shown to confer neurodegeneration in mouse models of ALS. Brain microglia are considered the "macrophage/dendritic cell of the central nervous system," are self-renewing, and

arise from fetal myeloid precursors (Bjorkqvist 2009). Activated brain microglia produce inflammatory cytokines that have been shown to play both neuroprotective and neurotoxic roles, which has complicated progress toward developing drugs that inhibit immune functions in the CNS (Ransohoff and Perry 2009). TLRs 1-9 and cytosolic nucleic acid receptors are also expressed at varying levels on all cell types of the brain, including neurons, microglia, and astrocytes (Okun 2009). Interestingly, there is a connection between the age of the brain and the degree of TLR expression: older mouse and human brains increase the expression of TLRs, particularly TLR 3 (Letiembre 2007). Additionally, the cytokine environment in the brain shifts with age. The CNS of younger individuals expresses protective cytokines, such as IL-10, but with age there is a regulated increase in the transcription of pro-inflammatory cytokines (Godbout 2004, Ye 2001, Lu 2004, Letiembre 2007). It is possible that individuals with a genetic defect in a negative regulator of innate immunity (such as TDP-43 or optineurin) might be particularly susceptible to neuroinflammation when this switch in immune gene expression occurs. This might explain why many neurodegenerative diseases present late in life.

Some have questioned whether ALS is the result of a viral infection. Transcipt levels of the nucleic acid sensors TLR3 and RIG-I are markedly upregulated in the brains of ALS individuals (Jackson 2006). In fact, the entire antiviral pathway seems to hyperactivated in ALS since levels of active, phosphorylated protein kinase R (PKR) are over 3000% higher in the spinal cords of ALS individuals (Hu

2003). Several groups have shown levels of retroviral reverse transcriptase to be as high in the serum and CSF of ALS patients as in individuals infected with the HIV virus (Steele 2005, MacGowan 2007, Verma 2006). However, follow-up studies have failed to find a pathogenic link. Additionally, a recent profiling of RNA from ALS spinal cords revealed that transcript levels corresponding to inherited active human endogenous retroviruses (HERVs) are significantly higher in brain tissue from ALS cadavers (Douville 2011).

The contribution of defects in the immune system to neurodegenerative disease has become more appreciated in the past few years. Pathologic microglial activation has been observed in the brains of individuals with Alzheimer's disease, Parkinson's disease, Huntington's disease, and ALS (Lobsiger and Cleveland 2007), and overactive innate immune pathways in ALS are increasingly reported (Butovsky 2012, Wang 2011, Polymenidou 2011, Swarup 2011, Papadimitriou 2010). There is a growing effort to understand how dysregulated immunity contributes to neurodegeneration since 1) existing drugs that target various aspects of the immune system can be tested in clinical trials for disease like ALS – which currently has no effective treatment, and 2) immune biomarkers can serve as clinical indications of disease severity and progression.

Elevated levels of the pro-inflammatory cytokines TNF and IL-6 have been identified in the cerebrospinal fluid of humans and mice with ALS, and in circulating blood serum of patients with ALS (Moreau 2005, Weydt 2004, Poloni

2000, Hensley 2003). Although drugs targeting TNF pathway, including thalidomide, showed promising results in SOD1 mouse models of ALS (Kiaei 2006), they failed in phase II clinical trials (http://www.clinicaltrials.gov/ct/show/NCT00231140). The antiviral arm of the innate immune signaling pathway is also dysregulated in ALS, but we are not aware of any trials that are specifically targeting antiviral mediators. There was a mild improvement in survivorship of an SOD1 mouse model when these animals were crossed to IFNAR1/2 knockout mice (Wang 2011). One important piece of work was recently published showing that TDP-43 functions as a co-activator of the p65 subunit of the transcription factor NF-kB. Many innate immune genes are regulated by NF-kB, including type I interferons. The authors found that treating ALS mice with the NF-kB inhibitor Withaferon A reduced ALS disease symptoms (Swarup 2011).

Section III Summary:

ALS is a devastating disease with no known treatment or cure. However, 150 years after being described, the disease still holds more questions than answers. There are many innate immune features of ALS, although the antiviral pathway has not been deeply explored. On the pro-inflammatory side of investigation, high levels of the cytokine TNF has been found in the serum of ALS patients and in mouse models of ALS. However, there is evidence indicating that antiviral responses are more active in ALS patients as well. For example, there is higher expression of two key viral sensors, TLR3 and RIG-I, in the brain cells of ALS patients. Additionally, a microarray analysis of PBMC from sporadic ALS patients

found the expression of several interferon stimulated genes (ISGs) to be higher in patient cells.

Section IV: Functional coupling among RNA processing machinery

RNA processing factors are highly functionally coupled

TDP-43 has been associated with a variety of cellular functions including regulating transcription, alternative splicing, and mRNA stability (see Section II of this Introduction for details and references). The commonality among these functions is RNA processing. Although the three main steps of gene expression – transcription, pre-mRNA splicing, and 3' end processing – are biochemically distinct processes, increasing experimental evidence shows that they are coordinated and often interdependent.

This section will present a synopsis of links associating all stages of RNA processing. In Chapter 1, we will present data showing that TDP-43 binds to mRNA transcripts that encode innate immune genes, often at their 3' UTRs. Furthermore, TDP-43 exerts an effect on transcriptional elongation that is dependent on the polyadenylation signals that are present in these 3' UTRs. We propose that TDP-43 is a factor that links transcription to 3' end processing. Although this is an unusual mechanism, it is not inconceivable given the level of coordination among RNA processing factors.

Pre-mRNA splicing factors are concentrated in nuclear speckles, particularly

perichromatin fibrils (also known as the nuclear location of nascent transcripts),

throughout the nucleoplasm. This collection of splicing apparatus is so large that

it can be visualized by electron microscopy without antibody labeling (Monneron

and Bernhard 1969). Poly(A)-binding protein II, one of the principle binding

partners of TDP-43, is present in these nuclear speckles, as are RNA

polymerase II (pol II), SR-proteins, and the cleavage factor CstF. In fact, all

nuclear mRNA processing events appear to occur at the site of transcription

(Krause 1994, Wansink 1993, Iborra 1996, Mortillaro 1996, Carmo-Fonseca

1991, Neugebauer 1997, Schul 1996, Misteli 1997).

RNA processing is coordinated by transcriptional elongation

Coordination between transcription and processing begins at initiation. One of

the most well characterized mechanisms of *cis*-regulation is promoter structure.

Cell specificity is in part maintained by the coordinated recruitment of distinct

combinations of factors to gene promoters. Promoter structure can also dictate

downstream alternative splicing events. For example, promoter architecture can

be involved in the recruitment of splicing factors to exonic splicing enhancers

(Cramer 1999). More recent findings in yeast have shown that a gene's promoter

sequences can also regulate mRNA decay of the transcript when it is transported

to the cytosol (Bregman 2011). This is thought to provide a mechanism for the

cell to coordinate the rates of mRNA synthesis and decay. RNA binding proteins

that mediate this process are referred to as "synthedegradases."

Once transcription is activated, factors associated with all levels of RNA processing interact. Transcription and pre-mRNA processing machinery are not only colocalized, but also physically associate with each other. The carboxy-terminal domain (CTD) of RNA pol II is highly phosphorylated at YSPTSPS repeat sites to engage active transcriptional elongation (Dahmus 1996). The RS domains of SR splicing factors bind to this hyperphosphorylated CTD and allow for a local accumulation of splicing and polyadenylation machinery co-transcriptionally. A form of pol II lacking the CTD results in deficient pre-mRNA splicing and 3' end processing (Mortillaro 1996, Yuryev 1996, Vincent 1996, Du 1997, Kim 1997, Dantonel 1997). The CTD is also involved in directing 5' methylguanosyl (5'mG) cap addition to pre-mRNAs. The two key enzymes that mediate the 5'mG cap mRNA modification, RNA guanylyltransferase and guanine-7-methyltransferase, physically associate with phosphorylated CTD, and without this association 5' cap formation is severely impaired (McCracken 1997, Cho 1997, Yue 1997). CTD regions have been dissected and distinct segments of the CTD have been found to regulate the individual pre-mRNA processing events - capping, splicing, and polyadenylation (Fong and Bentley 2001).

Transcription has also been shown to remotely control translation initiation in yeast. The RNA binding proteins Rbp4p and Rbp7p interact with the processing pol II, then travel with the mRNA transcript to the cytosol where they are involved in either directing the transcript to be decayed or they augment translation

35

initiation (Harel-Sharvit 2010). However, their initial communication with the polymerase is required for these later functions. Therefore, the authors suggest that some RNA binding proteins function as "mRNA coordinators" that function at all levels of mRNA processing, from promoter to decay or translation, to directly regulate gene expression.

Transcriptional elongation is affected by RNA processing

Transcriptional elongation is directly regulated by splicing factor recruitment. The elongation factor P-TEFb is recruited to promoters by transcription factors. It, in turn, recruits TAT-SF1, another elongation factor, that then recruits SR proteins to nascent transcripts (Fong and Zhou 2001). The presence of these splicing factors increases the efficiency of transcription elongation, and genes with introns are more efficiently transcribed than their intron-less counterparts (Ares 1999).

Interestingly, pre-mRNA processing machinery interdependency is taken to another degree as 3' end processing is found to be dependent on adequate 5' cap formation (Lewis 1995, Flaherty 1997). Conversely, deleting the AAUAAA polyadenylation signal from a 3'UTR results in defects in alternative splicing and transcriptional termination (Niwa and Berget 1991, Connelly 1988). In general, failures in 3' end formation results in an inability of the pre-mRNA to be released from transcription sites and undergo nuclear export (Vinciguerra and Stutz 2004).

Biochemistry of polyadenylation

The formation of the 3' end of a RNA transcript is a multi-step process involving a massive, 85-protein complex (Shi 2009). All protein-coding pol II-transcribed mRNAs contain a 3' end consisting of ~200 adenosine residues (except for replication-dependent histone genes). The presence of the poly(A) tail aids in nuclear export, translation, and maintenance of mRNA stability. For example, shortening the stretch of poly(A)'s leads to enzymatic degradation of the transcript (Guhaniyogi and Brewer 2001). *Cis* elements in the sequence at the 3' end of the pre-mRNA direct the polyadenylation process. Hexameric polyadenylation signal sequences (PAS), in particular the classic sequence encoded by AAUAAA, signals the site for the first step of 3' RNA end processing: cleavage. Cleavage occurs between the PAS and a downstream sequence element (DSE), which tends to be a GU-rich motif. The cleavage and polyadenylation specificity factor (CPSF), in a complex with other related proteins, then docks at the PAS. The cleavage stimulatory factor (CstF) interacts with the DSE, and then cleavage factors (CF) - I and -II are recruited to the pre-mRNA substrate. The cleavage factors stimulate cleavage but the poly(A) polymerase (PAP) is actually required for productive cleavage of the transcript and release of the polymerase to recycle back to the promoter. The CPSF and PAP work together to add the poly(A) tail, with the help of PABP II which increases PAP processivity. See Figure 3 and Proudfoot et. al., 2002 for details on this process.
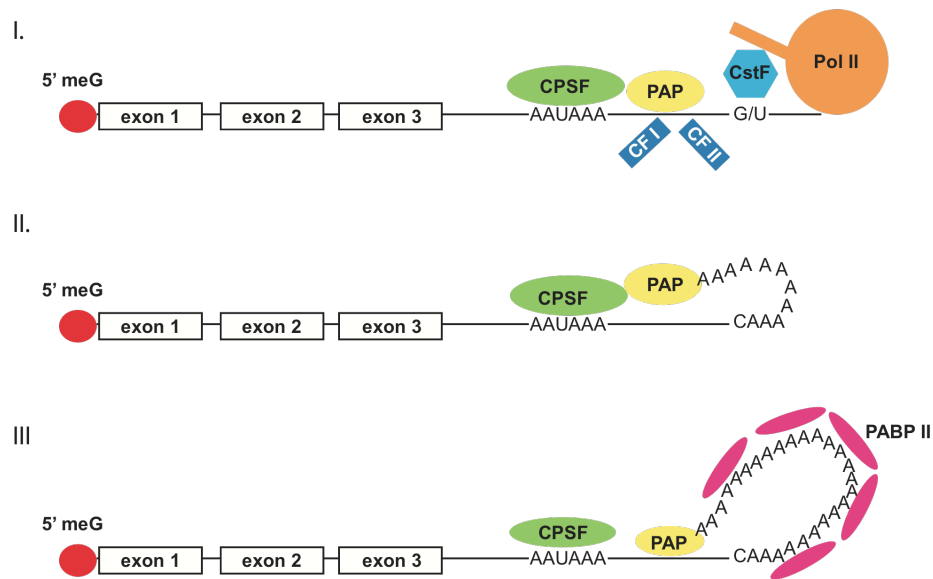
Figure 0.3: Polyadenylation is a multi-step process involving RNA cleavage and poly(A) synthesis. (Adapted from Proudfoot 2002).

Genes often contain more than one PAS and the mechanism of PAS selection

remains a subject of for more research. Just as alternative splicing increases the

diversity of gene expression, alternative polyadenylation can change the

dynamics of expression as well. For example, there is a very nice literature

demonstrating that rapidly-cycling cells (such as cancer cells) express transcripts

with shorter 3'UTRs, using more proximal PAS, effectively "hiding" *cis* regulatory

sequences in the 3'UTR that would otherwise target the transcripts for depletion

(Sandberg 2008).

It is clear that there is cross talk between splicing and polyadenylation. For

example, the IgM gene encodes a "weak" internal PAS in an intron. The usage of

this PAS is repressed by the presence of an upstream strong 5' splice site. Therefore, the PAS-containing intron is spliced out to inhibit termination at that site and thus force the usage of a downstream PAS (Zhao 1999).

Recommended reviews on this literature are: Bentley 2002; Maniatis and Reed 2002; Neugebauer 2002; Proudfoot 2002.

Section IV summary: Is TDP-43 an "mRNA coordinator?"

Why are these complex mechanisms relevant to TDP-43? TDP-43 is a known splicing factor. In addition to the obvious splicing factor interactions, TDP-43 has been shown to bind PABP II and translation initiation factors, to bind mRNAs at intron-exon junctions and at 3' UTRs, to regulate translation rates, and to regulate alternative PAS usage (Jernej Ule lab, data presented at a conference in October of 2012). Furthermore, we have noted that TDP-43 often associates with AU-rich sequences in 3'UTRs, as well as GU-rich downstream elements that are involved with polyadenylation.

Just as a promoter can affect alternative splicing and transcription elongation can affect efficient 3' pre-mRNA end processing, we wondered whether sequences in the 3' UTR can effect transcriptional elongation. The work in this thesis finds that 1) TDP-43 binds to the 3' UTR of the IFNβ transcript 2) TDP-43 regulates the activity of an IFNβ 3' UTR reporter and that the PAS sequence is required for this regulation and 3) the PAS in the IFNβ 3' UTR is required for TDP-43 to inhibit IFNβ promoter activity. We also noted that IFNβ mRNA decay is faster in the

absence of TDP-43, suggesting that TDP-43 also regulates the stability of the transcript.

There is only one instance of 3'UTR-mediated promoter regulation in mammalian cells in the literature. (Le Cam 1995). Two related proteins, serine protease inhibitor (*spi*) 2.1 and 2.3 are inversely expressed during an inflammatory response. A major difference in the sequences of these two genes is that the *spi* 2.3 3' UTR contains a 384-bp extension. By generating constructs that localized the *spi* 2.3 3' UTR extension to positions upstream or downstream of the *spi* 2.3 promoter, the authors found that this 3' UTR sequence repressed *spi* 2.3 promoter activity.

I introduced the term "mRNA coordinator" earlier. TDP-43 has been shown to function along all stages of gene expression – transcription, splicing, polyadenylation, translation, and decay. Perhaps its ability to regulate IFNβ transcription is part of a larger function that regulates gene expression. This might be even more important in an immune context. In response to a viral infection, the cell must quickly respond and engage an antiviral state. However, it is equally important for the cell to downregulate this response to prevent long-term damage. One way that cytokine levels are regulated is through a mechanism of mRNA decay using 3'UTR-mediated signals. We found that the PAS is required for the regulation of IFNβ by TDP-43. Perhaps productive 3' pre-

mRNA end processing is required for TDP-43 to dampen transcriptional

elongation.

CHAPTER 1: The ALS-linked gene TDP-43 regulates IFNβ expression through a novel mechanism of 3'UTR-mediated promoter cis-regulation

Raquel P. Deering, Schraga Schwartz, Raktima Raychowhurdy, Jacob

Dockterman, Nancy Kedersha, Alexander Melnikov, Li Wang, Xiaolan Zhang,

Aviv Regev, Tarjei Mikkelsen, and Nir Hacohen

Abstract

The TAR DNA-binding protein (TDP-43) is a heterogeneous nuclear ribonucleprotein that is involved in multiple stages of RNA processing. Mutations in the TDP-43 gene and mislocalization of TDP-43 protein have been implicated in a growing number of neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). Here, we show that TDP-43 negatively regulates innate immune gene expression in response to RNA virus sensing. Perturbation of TDP-43 protein expression leads to an increase in antiviral gene expression in a variety of human and mouse cells. Crosslinked RNA immunoprecipitation (CLIP) experiments revealed that TDP-43 binds to type I interferon (IFN) and interferon stimulated gene (ISGs) transcripts. Using massively parallel 3' UTR reporter assays coupled with high throughput sequencing (MPRA-seq), we identified polyadenylation signal sequences in the 3' UTRs of innate immune genes to be specifically regulated by TDP-43. Surprisingly, IFN and ISG mRNA decay rates are faster in TDP-43-perturbed cells. Using a metabolic labeling approach to measure nascent transcript generation, we found that perturbation of TDP-43 expression leads to an increase in antiviral gene transcription rates. Additionally, RNA polymerase II (pol II) chromatin immunoprecipitation (ChIP) confirmed that there is greater pol II occupancy on innate immune genes when TDP-43 is depleted. Although TDP-43 perturbation has no effect on an isolated IFNβ promoter reporter, we found that TDP-43 inhibits IFNβ promoter activity when the IFNβ 3' UTR sequence is inserted downstream of the IFNβ promoter element, suggesting a novel mechanism of 3'UTR-mediated promoter cis-regulation.

Introduction

In 2006, the 43 kD transactive response DNA-binding protein (TDP-43) was found to be the most abundant protein in cytosolic inclusions that characterize amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar dementia (FTLD) (Neumann 2006). Since then, genomic sequencing of the gene that encodes TDP-43 (TARDBP) in patients with ALS and FTLD has identified several dozen dominant missense mutations that cause neurodegenerative disease (Gitcho 2008, Sreedharan 2008, Kabashi 2008, Yokoseki 2008, Van Deerlin 2008). TDP-43 immunoreactive inclusions have also been identified in other neurodegenerative diseases, including Alzheimer's, Huntington's, and Parkinson's diseases as well as chronic traumatic encephalopathy (Amador-Ortiz 2007, Nakashima-Yasuda 2007, Schwab 2008, McKee 2010).

Disease-causing mutations are almost exclusively found in the RNA binding domains of TDP-43. Therefore, understanding how TDP-43 normally functions has become a priority. TDP-43 has already been shown to play a variety of roles in regulating gene expression, including transcriptional repression, alternative slice site selection, and mRNA stability. Early work found that TDP-43 repressed transcription from the HIV-1 long terminal repeat region (TAR). TDP-43 also regulates spermatogenesis by binding to 5'-GTGTGT-repeat sequences on the antisense strand of the SP-10 promoter and preventing premature expression of the gene. Conversely, TDP-43 is a co-activator of the NF-kB subunit, p65, and promotes the expression of many immune genes (Swarup 2011).

The best-characterized functions for TDP-43, however, are its roles in RNA processing. TDP-43 promotes exon skipping in the cyctic fibrosis transmembrance conductance (CFTR) gene (Ayala 2006) and was found to globally regulate the splicing of neuronal genes that contain very long introns (Polymenidou). In addition to binding to the 3' UTR of its own transcript for autoregulation (Avendano-Vasquez 2012), TDP-43 also promotes the stabilization of histone deacetylase-6 (HDAC6) and neurofilament (NFL) transcripts by binding to their 3' UTRs (Fiesel 2009, Kim 2010, Volkening 2009).

Our lab found that the expression of type I interferons and other interferon-stimulated antiviral genes (ISGs) is upregulated in TDP-43-depleted cells. We previously identified another ALS disease gene, Fus/TLS, to be a potent negative regulator of antiviral responses (Amit 2009). Optineurin is a known negative regulator of antiviral gene expression and mutations in the optineurin gene are causative of a percentage of familial ALS cases (Maruyama 2010). Furthermore, increased ISG expression has been noted in human ALS and an SOD1 ALS mouse model (Wang 2011). The misregulation of antiviral genes, therefore, might contribute to ALS pathogenesis.

All cells of the body are equipped with intracellular viral sensing machinery and can elicit an antiviral response to the nucleic acid components of invading viruses. Following viral sensing, a series of signaling events activate key

transcription factors and transcriptional co-activators that culminates in the transcription of type I interferons, specifically interferon (IFN) β (Thanos and Maniatis 1995). Although levels of IFNβ in most virally infected tissues remain relatively low, IFNβ protein is a potent activator and amplifier of the entire downstream antiviral response. Hence, understanding how type I interferon production is regulated is critical for understanding the mechanism of misregulation in disease processes.

In order to determine how TDP-43 functions as an inhibitor of the antiviral pathway, we used crosslinked RNA immunoprecipitation (CLIP) experiments to identify transcripts that are bound by TDP-43 during an active antiviral response. RNA target analysis determined that TDP-43 binds to the 3' UTR of the critical type I interferon, IFNβ. We also found that TDP-43 binds directly to other antiviral gene transcripts. Massively parallel 3'UTR reporter assays revealed that polyadenylation signal sequences (PAS) are required for TDP-43-mediated regulation of antiviral genes. Rather than regulating mRNA stability, however, TDP-43 functions as a repressor of IFNβ transcription. We report a surprising mechanism in which the PAS in the IFNβ 3' UTR is required for TDP-43 to appropriately repress IFNβ promoter activity. In the absence of TDP-43, pol II accumulates on the transcription start site and along the gene body of IFNβ and interferon-stimulated genes (ISGs). It has been previously published that the splicing factor SC35 functions as a positive regulator of transcriptional elongation (Lin 2008), and that appropriate 3' end processing is required for alternative

splicing and transcriptional termination (Niwa and Berget 1991, Connelly and Manley 1988).

We propose an analogous mechanism in which a splicing factor negatively regulates transcriptional elongation, but requires 3' end processing signals to regulate promoter activity.

Results

Perturbation of TDP-43 expression induces the antiviral gene program

Our lab found that Fus, a protein with regions of sequence homology to TDP-43, is a strong negative regulator of the antiviral gene program in response to viral pathogen sensing (Amit et al. 2009). Given the sequence similarities between Fus and TDP-43, we tested whether TDP-43 similarly regulates antiviral gene expression. Type I interferons, namely IFNβ and IFNa, are potent antiviral genes that function to induce and amplify the entire antiviral gene program. Therefore, monitoring the expression of IFNβ is a good proximal measure of the activity of an antiviral response. Puromycin-selectable lentiviruses encoding small hairpin RNAs (shRNA) that target TDP-43 were used to deplete protein levels in primary bone marrow dendritic cells (BMDC) and IFNβ gene expression was assessed by quantitative real-time polymerase chain reaction (qPCR) following Sendai virus (SeV) infection. Like Fus perturbation, the depletion of TDP-43 led to a marked increase in IFNβ gene expression (Figure 1A). Since all cells of the body are

equipped with the cellular machinery to sense and respond to viral infection, we tested whether the depletion of TDP-43 in a non-immune cell demonstrated a similar phenotype. siRNA oligonucleotides targeting TDP-43 were transfected into mouse lung fibroblasts (MLF) or a human osteosarcoma cell line (U2OS). Indeed, the depletion of TDP-43 resulted in an increase in both IFNβ expression as well as the transcript levels for a key interferon-stimulated gene (ISGs), CXCL10 (Figure 1B). We tested the effect of depleting TDP-43 from a variety of human and mouse cells, including primary human lung fibroblasts and primary mouse microglia, and found a consistent interferon and ISG phenotype (S. Figure 1A-C). Additionally, others have published that the depletion of TDP-43 *in vivo* from mouse brains using anti-sense oligonucleotides results in a spontaneous upregulation in the expression of many key ISGs (Polymenidou 2011).

To confirm that greater functional interferon protein is produced when TDP-43 is depleted, we used a type I interferon-sensitive bioassay to test the supernatants of cells that were infected with either a control shRNA or a hairpin targeting TDP-43. In this assay, a reporter cell line that stably expresses an interferon stimulated response element (ISRE) fused to a luciferase reporter (ISRE-luc) detects all type I interferons present in cell supernatants. Following shRNA infection and puromycin selection, cells were infected with the negative-sense ssRNA PR8 strain of the influenza A virus (influenza A/PR8). Supernatants from TDP-43-depleted cells induced higher total ISRE activity than cells that were

infected with a control lentiviral vector (Figure 1D), indicating a greater

accumulation of functional type I interferon proteins.

We next tested whether overexpression of TDP-43 would have the reciprocal

effect of TDP-43 depletion. TDP-43 overexpression is rapidly toxic to cells.

Therefore, we wanted to use an inducible system to control the level and duration

of TDP-43 expression. We generated a U2OS cell line in which human TDP-43

containing a GFP tag is tetracycline-inducible. Upon the addition of tetracycline,

GFP fused to either wild type human TDP-43 (GFP-TDP-43) or an ALS-linked

TDP-43 mutation (GFP-TDP-43^{A315T}) was expressed. After 24 hours,

overexpressed TDP-43 protein remained restricted to the nucleus (S. Figure 2A).

Surprisingly, the overexpression of both wild type and disease-associated TDP-

43 induced higher antiviral gene activity (Figure 1F).

Figure 1.1: Perturbation of TDP-43 expression leads to increased type I interferon and interferon-stimulated gene expression.

(A) mRNA levels (qPCR: relative to GAPDH abundance) of IFNb in TDP-43-depleted BMDCs. Cells were infected with SeV (MOI 1) for 6 hrs. Error bars represent the standard
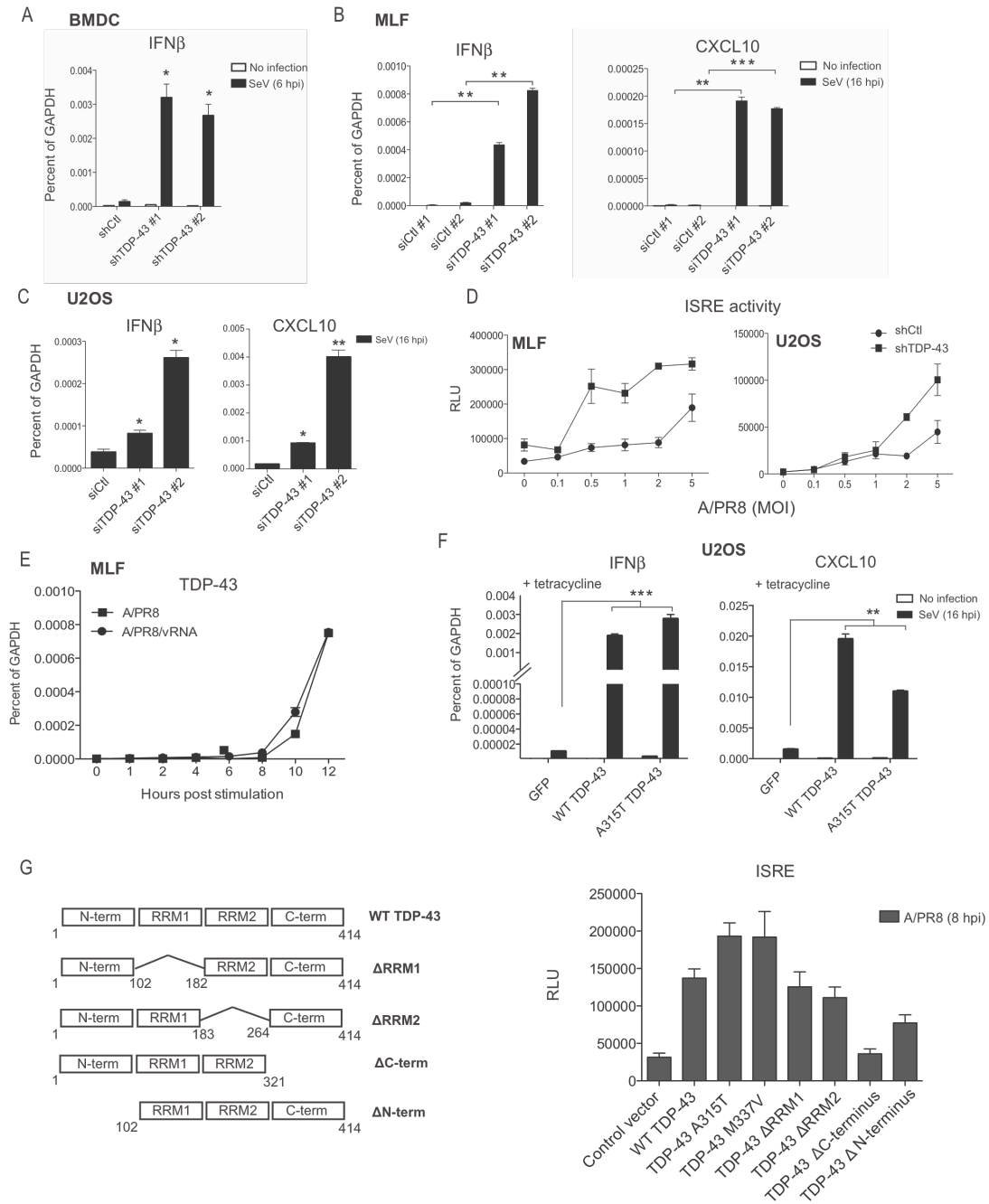
Figure 1.1 (Continued):
error of the mean (SEM). (*p<0.05, **p<0.01, ***p<0.001). Data represents six biological and two technical replicates from one mouse.

(B-C) mRNA levels (qPCR) of IFNb (left) and CXCL10 (right) in TDP-43-depleted mouse lung fibroblasts (MLF)(B) and U2OS cells (C). Cells were infected with SeV (MOI 2) for 16 hrs. Data represents three independent experiments.

(D) Luciferase bioassay testing type I interferon production of TDP-43-depleted cells. Culture supernatants of MLF (left) and U2OS cells (right) that were infected with influenza A/PR8 (A/PR8) virus at various MOI (0, 0.1, 0.5, 1, 2, 5) for 24 hrs were tested on type I interferon-sensitive HEK293T cells expressing an ISRE-luciferase reporter (ISRE-luc). Error bars represent the SEM. Data represents three independent experiments.

(E) mRNA level (qPCR) of TDP-43 in MLF infected with A/PR8 (squares) or transfected with viral RNA (vRNA) isolated from A/PR8 (circles) over a time course (0, 1, 2, 4, 6, 8, 10, 12 hrs). Data represents two independent experiments in MLF and experiments in multiple cell types.

(F) mRNA levels (qPCR) of IFNb (left) and CXCL10 (right) in U2OS cells when TDP-43 is overexpressed. U2OS cells transduced with a gene encoding wild type TDP-43 (WT TDP-43) or an ALS-linked TDP-43 disease allele (TDP-43^{A315T}) under the control of a tetracycline inducible promoter were treated with tetracycline (0.25 ug/ml) for 24 hrs to induce TDP-43 overexpression. Cells were infected with SeV (MOI 2) for 16 hrs. Data represents three independent experiments.

(G) Type I interferon production of cells overexpressing TDP-43 ALS disease alleles and domain mutants. A schematic of TDP-43 domain mutations (left). HEK239T cells were transfected with plasmid DNA encoding WT TDP-43, ALS disease alleles (TDP-43^{A315T}, TDP-43^{M337V}), or TDP-43 domain mutants. Cells were infected with SeV for 16 hrs and culture supernatants were tested on ISRE-luc cells. Error bars represent the SEM. Data represents three independent experiments.

The dominant negative TDP-43 overexpression phenotypes were puzzling to us. TDP-43 has been found to be highly ubiquitinated and phosphorylated in ALS and FTLD-U cytosolic aggregates (Neumann 2006), and the Q/N-rich TDP-43 C-terminal domain causes it to aggregate in vitro and in vivo (Udan and Baloh 2011). We wondered whether overexpressed TDP-43 was similarly modified and aggregating, thus preventing TDP-43 from functioning normally. Proteins from lysates of HEK293T that were transfected with wild type TDP-43 or two ALS disease alleles were separated on SDS-PAGE gels and TDP-43 protein was probed using specific antibodies by Western blot. We found that the TDP-43 antibody recognized a high molecular weight protein smear indicating that overexpressed TDP-43 is highly ubiquitylated (S. Figure 2B). Furthermore, the induction of TDP-43 overexpression in the U2OS cell line resulted in a depletion of the endogenous TDP-43 protein band migrating at 43 kD, consistent with the finding that TDP-43 regulates its own transcript stability by its 3' untranslated region (3' UTR) (S. Figure 2C) (Polymenidou 2011, Avendano-Vasquez 2012). It is possible that the overexpression of TDP-43 not only depletes endogenous TDP-43 transcript, but also is rendered non-functional by post-translational modifications that target the protein for degradation and by aggregation. This might explain why TDP-43 overexpression recapitulates the knock down phenotypes.

Expression of TDP-43 is induced following viral pathogen sensing

We wondered whether TDP-43 is regulated in response to viral pathogen sensing. MLF were either infected with influenza A/PR8 or were transfected with total viral RNA that was isolated from influenza A/PR8 (vRNA). The expression of TDP-43 increased following sensing of both whole influenza virus and isolated viral RNA (Figure 1E). Transfection of immunostimulatory DNA (ISD) also led to similar induction of TDP-43 gene expression (S. Figure 3A).

TDP-43 expression in the human U2OS cell line was also tested in response to RNA virus sensing. Cytosolic levels of TDP-43 did not change over time but nuclear TDP-43 increased approximately 20% in response to influenza A/PR8 infection (S. Figure 3B) and SeV (data not shown). Although the protein levels do not match the transcriptional upregulation of TDP-43, it is likely that virally mediated translational repression and TDP-43 autoregulation effect total protein levels. Furthermore, we and others find that increased TDP-43 protein levels are toxic and are likely tightly regulated. On the other hand, in an inducible TDP-43 knock out system TDP-43 heterozygous mice maintain the same TDP-43 protein levels as homozygotes, suggesting that TDP-43 levels are carefully controlled and compensated in heterozygotes (Chiang 2010).

TDP-43 is a largely nuclear protein, although it has been shown to shuttle between the cytosol and nucleus, and can relocalize to cytosolic stress granules

under conditions of induced stress (Liu-Yesucevitz 2010). However, we did not

see a redistribution of TDP-43 to TIAR-positive stress granules following a time

course of viral sensing (S. Figure 3B). Our lab has shown that TDP-43

undergoes post-translational changes in response to pathogen sensing,

however. Phosphoproteomic analysis of TDP-43 peptides from primary bone

marrow dendritic cells (BMDC) stimulated with lipopolysaccharide (LPS)

demonstrated that serines 292 and 273 of TDP-43 are highly phosphorylated at

time points between 120-360 minutes following pathogen sensing (Chevrier

2011). Other sites of TDP-43 phosphorylation have been reported in the context

of neurodegenerative disease, but these sites appear to be unique to pathogen

sensing.

TDP-43 functions downstream of IFNβ transcription

In order to determine which domains are required for TDP-43 to inhibit type I

interferon activity, we generated TDP-43 expression constructs in which we

systematically deleted sequence regions that encode for key functional domains

of the human TDP-43 protein fused to an mCherry tag. HEK293T cells were

transfected with plasmids encoding TDP-43 disease alleles or domain mutants.

All of the tagged TDP-43 proteins were retained in the nucleus, except for the N-

terminal domain mutant, which lacks a nuclear localization signal (S. Figure 4).

The supernatants from these cells were tested for type I interferon activity using

the interferon-sensitive ISRE-luc reporter cell line. Overexpression of WT TDP-43

or ALS disease alleles resulted in an increase in type I interferon expression following A/PR8 infection (Figure 1F). Deleting either the RNA recognition motif (RRM)-1 or RRM2 did not fully reduce the level of interferon activity. However, type I interferon production was brought down to normal levels when the C-terminal arginine-glycine-glycine rich (RGG) domain was absent. Presumably, the presence of the C-terminal RGG domain is required for TDP-43 to maintain its toxic effects.

In order to determine where TDP-43 is functioning as an inhibitor of antiviral signaling, we disrupted various points along the innate immune signaling pathway. TDP-43 function is dependent of intact cytosolic viral sensing since depleting TDP-43 from MEFs that lack the signaling adaptor MAVS completely abolishes the induction of IFNβ (S. Figure 5). Interferon regulatory factor 3 (IRF3) is a master transcription factor for type I interferon expression. Normally, no type I interferon is produced in response to SeV infection in IRF3-/- cells. However, the depletion of TDP-43 from IRF3-/- cells restores type I interferon gene induction in an IRF3-independent mechanism (Figure 2A). TDP-43 does not appear to involve an increase in cytosolic nucleic acid sensing or signaling since IRF3 phosphorylation does not increase in TDP-43-depleted cells (Figure 2F). IRF7 is another master regulator of type I interferon induction. We tested the effects of TDP-43 depletion in IRF3/IRF7 double knockout cells and found that, although the absence of both of these transcription factors significantly reduces type I interferon expression, the depletion of TDP-43 induces type I interferon in

an IRF3- and IRF7- independent fashion (Figure 2B). Depletion of TDP-43 in IFNR-/- cells completely abolishes ISG expression, indicating that type I interferon expression is required for TDP-43 to regulate the expression of ISGs (Figure 2C).

It was recently observed that TDP-43 functions as an activator of the p65 subunit of NF-kB (Swarup 2011). Since the IFNβ promoter contains an NF-kB-dependent element, we wondered if TDP-43 was directly acting on the IFNβ promoter. To explore this, we tested the effect of TDP-43 perturbation on a series of transfected IFNβ promoter elements (PRD II, NF-kB-dependent; PRD III/I, IRF3/7 dependent; and the entire minimal enhancer element) driving a luciferase reporter. Neither the depletion of TDP-43 nor the overexpression of WT TDP-43 or TDP-43 disease alleles resulted in an increase in the NF-kb- dependent or independent IFNβ promoter elements in our system (Figure 2D,G). However, TDP-43 depletion does increase the activity of a promoter reporter containing an ISRE (Figure 2E). Additionally, depletion of TDP-43 results in both a baseline and viral sensing-induced increase in STAT1 phosphorylation (Figure 2F). This is consistent with our observation that perturbation of TDP-43 expression increases type I interferon production. To confirm that TDP-43 is not directly associated with any relevant promoter DNA elements, we performed genome-wide TDP-43 chromatin immunoprecipitation of both endogenous TDP-43 and overexpressed TDP-43 protein fused to a 3x flag tag followed by high throughput sequencing (ChIP-seq), but found no significant DNA enrichment compared to positive

controls (data not shown). From these experiments, we concluded that TDP-43 is not binding to the IFNβ promoter directly. The increase in IFNβ transcript levels lead us to interrogate whether TDP-43 was regulating RNA transcripts directly.
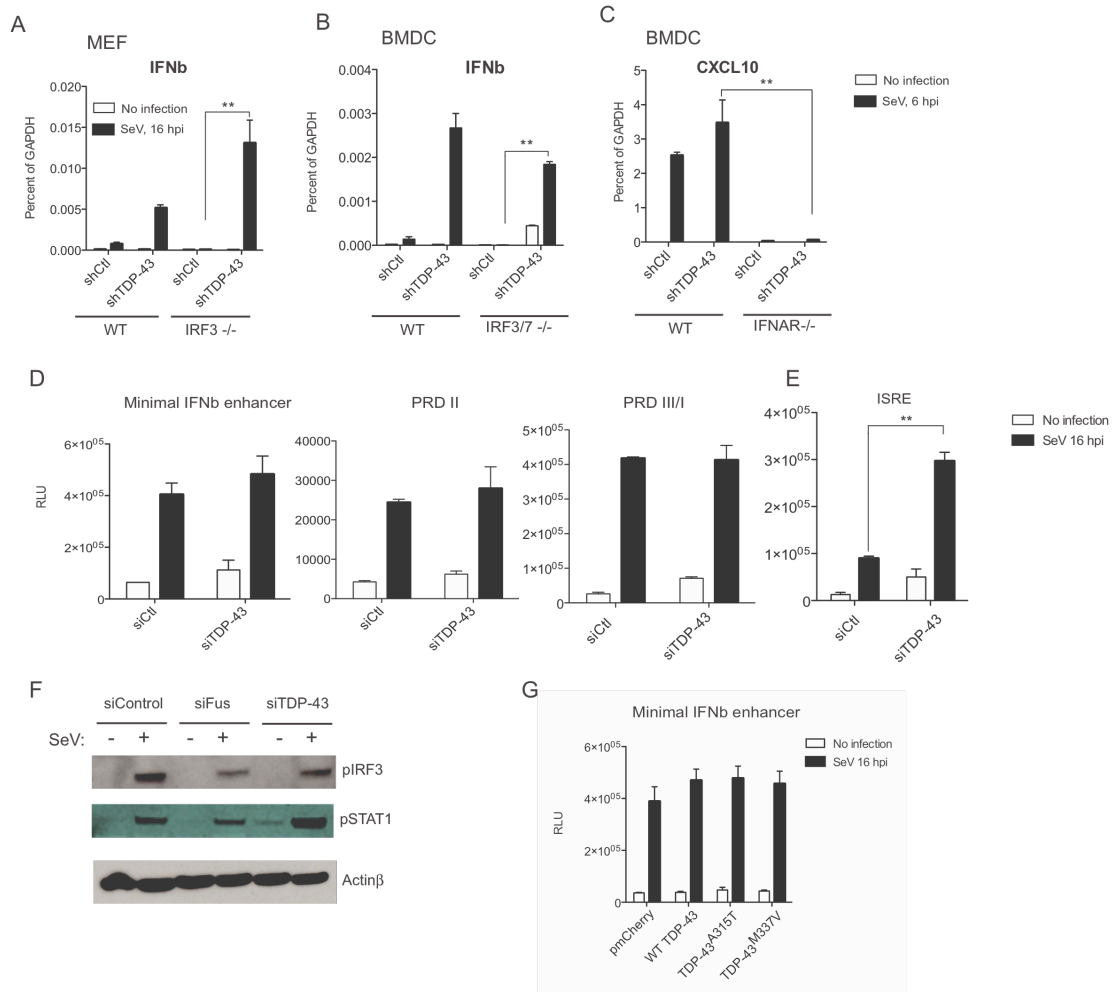
Figure 1.2: TDP-43 acts as an inhibitor of type I interferon expression downstream of cytosolic antiviral signaling.

(A) mRNA levels (qPCR) of IFNb in IRF3-/- mouse embryonic fibroblasts (MEF). TDP-43 was depleted from wild type (WT) MEFs and IRF3-/- MEFs with shRNA and cells were infected with SeV (MOI 2) for 16 hrs. Data represents six biological and two technical replicates from one mouse. Error bars represent the SEM. *p<0.05, **p<0.01.
(B) mRNA levels (qPCR) of IFNb in IRF3/7-/- BMDCs. Methods and statistics same as in (A) but cells were infected with SeV (MOI 1) for 6 hrs.
(C) mRNA levels (qPCR) of CXCL10 in IFNR-/- BMDCs. Methods and statistics same as in (B).

Figure 1.2 (Continued):
(D) Luminescent IFNb promoter enhancer activity in TDP-43-depleted cells. HEK293T cells were transfected with siRNA targeting a control sequence or TDP-43 and then cells were transfected with expression plasmids encoding either the complete minimal IFNb enhancer (left), the PRD II enhancer domain (middle), or the PRD III/I enhancer domain (right), all driving a firefly luciferase reporter. Cells were infected with SeV (MOI 5) for 18 hrs to induce antiviral gene expression. All luminescence values are normalized to an internal firefly renilla control. Data is representative of five experiments.
(E) Luminescent ISRE promoter activity in TDP-43-depleted cells. HEK293T cells were transfected with siRNA targeting a control sequence or TDP-43 and 48 hrs later transfected with a plasmid encoding an ISRE-containing promoter driving a luciferase. Cells were infected with SeV (MOI 5) for 18 hrs to induce antiviral gene expression. Data is representative of six experiments.
(F) Western blot for pIRF3 and pSTAT1 in TDP-43-depleted cells. HEK293T cells were transfected with siRNA targeting a control sequence, Fus, or TDP-43. Seventy-two hours later, cells were infected with SeV (MOI 5) for 6 hrs. Lysates were processed, run on SDS-PAGE gels, and probed with antibodies specific for pIRF3 (top), pSTAT1 (middle), or actinβ (bottom). Data is representative of three experiments.
(G) Luminescent IFNb promoter enhancer activity in TDP-43 overexpressing cells. Methods same as in (D) but HEK293T cells were transfected with expression plasmids encoding an empty vector control (pmCherry), WT TDP-43, TDP-43^{A315T}, or TDP-43^{M337V}. Data is representative of four experiments.

TDP-43 associates with the 3' UTRs of type I interferon and ISG transcripts

Several reports have shown that TDP-43 directly binds to mRNA (Polymenidou 2011, Sephton 2011, Ayala 2011). There is some evidence demonstrating that TDP-43 preferentially binds to GU-rich RNA sequences (Buratti 2001, Ayala 2005), but the presence of this putative consensus sequence is not required for TDP-43 to interact with the many of its RNA binding partners (Polymenidou 2011, Sephton 2011). We wondered if TDP-43 regulates type I interferon and ISG expression by directly binding to these genes' mRNA transcripts. To find *in vivo* RNA targets of TDP-43 in our system, we performed ultraviolet (UV) irradiation crosslinked RNA immunoprecipitation followed by high throughput sequencing (CLIP-seq) on both endogenous and overexpressed human TDP-43 protein. Technical limitations associated with the UV crosslinking efficiency of the CLIP protocol and the activity of the commercially available TDP-43 antibody required us to overexpress TDP-43 in order to enrich for associations with low-expression antiviral genes. To bypass the toxicity of overexpressing TDP-43, we induced its expression in U2OS cell lines expressing GFP-tagged wild type TDP-43 or a GFP-tagged TDP-43[A315T] disease variant in a tetracycline-responsive system. Viral RNA was transfected into cells to induce the RNA sensing pathway and resulting RNA-protein interactions were immobilized by crosslinking the cells with UV radiation (265 nm). We then performed immunoprecipitations using antibodies that targeted either endogenous TDP-43 or TDP-43-GFP under conditions of high RNase A digestion to isolate RNA fragments that are

specifically bound to the TDP-43 protein. The sizes of RNA isolated from each

TDP-43 CLIP ranged from 40-150 bases in length. RNA sequencing libraries

were generated for both the isolated CLIP RNA as well as the total input RNA

and single-end Illumina sequencing was performed.

We determined differential enrichment by comparing TDP-43 CLIP library

enrichment to total input RNA abundance and subtracting any background signal

from the GFP control CLIP. By selecting a threshold of 4-fold enrichment or

greater, we identified many coding and non-coding genes to be enriched for

TDP-43 binding (Table 1), with significant overlap in binding sites between the

wild type TDP-43 CLIP and TDP-43^{A315T} disease allele CLIP. The majority of

TDP-43 interactions occurred in the 3' UTR of target transcripts (A complete list

of CLIP targets is provided in S. Table 1 online). Strikingly, the most enriched

binding motif for the overexpressed forms of TDP-43 was not the canonical GU-

rich binding motif but rather a more AT-rich motif (Figure 3B). The endogenous

TDP-43 CLIP hexameric binding motif, however, was the preferred GTGTGT

sequence. We noticed that overexpressed forms of TDP-43 associated with 3'

UTRs more frequently (61.8-62.4%) than the endogenous form (26.4%). We

wonder if this shift in motif enrichment is due to a higher occupancy of

overexpressed TDP-43 on 3' UTRs, which are AU-rich in nature. Given the

overlap in binding associations between WT and ALS-linked TDP-43, the

functional differences conferred by mutations in the TDP-43 protein are likely

independent of binding site recognition and 3' UTR occupancy.

Table 1.1: CLIP binding enrichment

	IP # reads	# Peaks	Coding	Non-coding	5' UTR	CDS	3' UTR
Flag-WT TDP-43	42,717,178	3085	2974	111	1.0%	36.5%	62.4%
Flag-TDP-43^{A315T}	46,745,673	1949	1818	131	0.9%	37.3%	61.8%
Endogenous TDP-43	20,599,946	291	231	60	16.0%	57.6%	26.4%

We found enrichment for GFP-TDP-43 binding in the 3'UTRs of several antiviral genes including IFIT1 and CXCL10 (Figure 3A). The reads associated with the endogenous TDP-43 CLIP for these genes were many fold fewer, however, so we were not able to determine enrichment. We also confirmed that TDP-43 binds its own 3'UTR in our system as well as to the transcript of the related ALS-linked protein, Fus/TLS. IFNβ was too lowly expressed in these cells to achieve adequate CLIP-seq enrichment to meet our cut-off for statistical significance. Although all cells are equipped with the machinery to respond to virus, type I interferon expression is carefully regulated during the course of an antiviral response to prevent toxicity.

We found that qPCR was a more sensitive assay for assessing relative CLIP enrichments for low-expression genes. To validate that the GFP-TDP-43 CLIP-seq findings were relevant for endogenous TDP-43, we performed CLIP experiments under slightly lower RNase A conditions and used qPCR as a read-out. Based on binding site enrichment from the GFP-WTTDP-43 CLIP-seq, primers were designed against the 3' UTRs of several genes of interest. We also tested primers that targeted internal regions of the genes. We confirmed that

endogenous TDP-43 specifically associates with IFNβ and ISG 3' UTRs

compared to the RNA binding proteins Ezh2 and Fus (Figure 3C). Furthermore,

TDP-43 was more enriched in 3' UTR associations than for internal regions of the

transcripts, indicating that even under lower RNase A conditions the CLIP

protocol was robust.
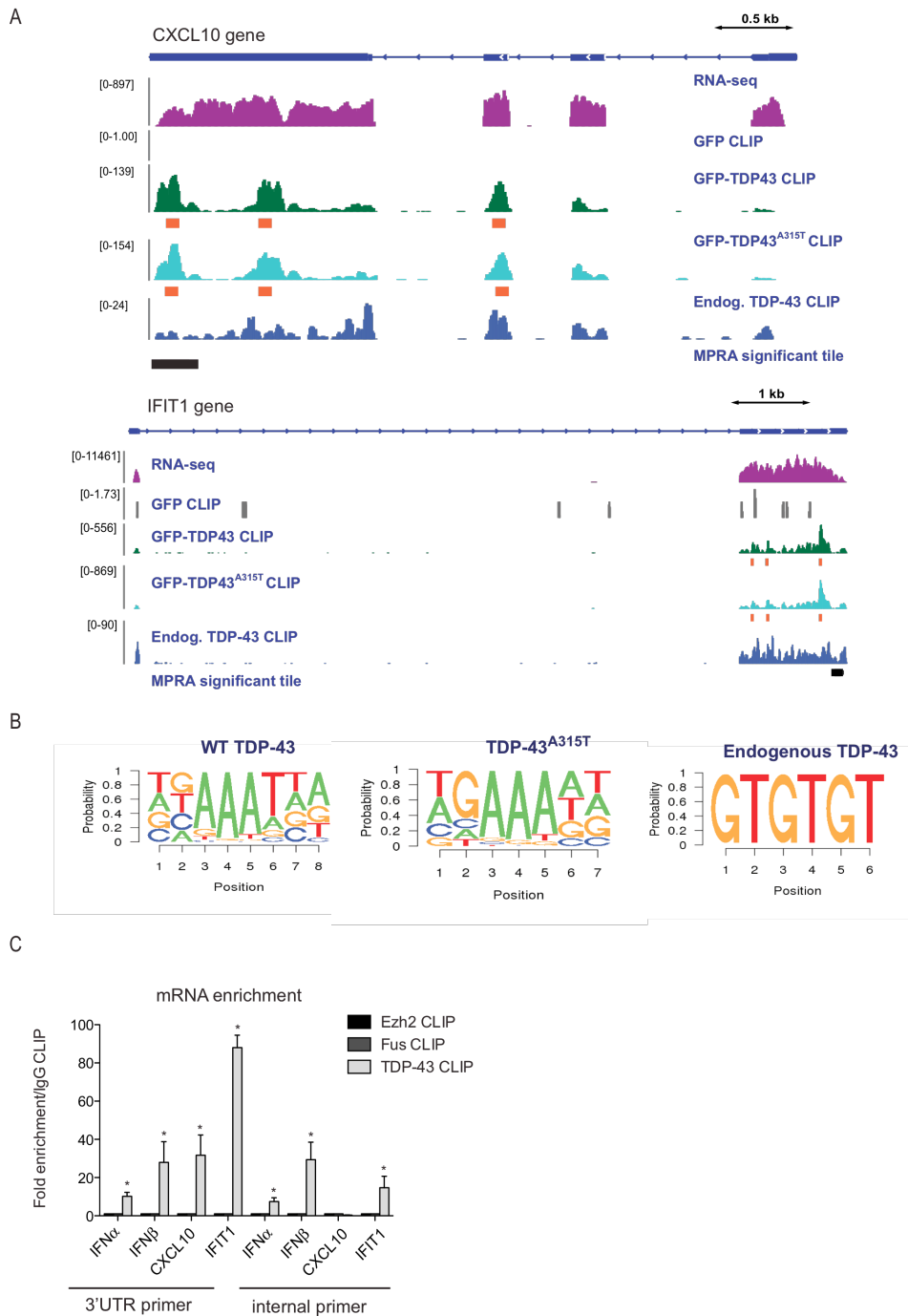
Figure 1.3: TDP-43 binds to type I interferon and ISG transcripts *in vivo*.

(A) Intergrative Genomics Viewer screen shot of TDP-43 UV-crosslinked immunoprecipitation (CLIP) binding peaks on CXCL10 and IFIT1 transcripts. U2OS cells expressing GFP, GFP-WTTDP-43, or GFP-TDP-43A315T under tetracycline control were transfected with influenza A/PR8 viral RNA (A/PR8/vRNA) for 18 hrs to induce antiviral gene expression. The cells were UV crosslinked (265 nm) and either total RNA was isolated for RNA sequencing (purple) or

Figure 1.3 (Continued):
CLIP-seq was performed. Sequencing reads corresponding to GFP (gray), GFP-WTTDP-43 (green), GFP-TDP-43^{A315T} (aqua) or endogenous TDP-43 (blue) CLIP enrichments are shown by orange boxes. Massively parallel reporter assay significant tiles are shown by black boxes.
(B) Enriched sequence motifs corresponding to WT TDP-43, TDP-43^{A315T}, and endogenous TDP-43 CLIP associations.
(C) qPCR for TDP-43 CLIP enrichment. U2OS cells were transfected with A/PR8/vRNA for 18 hrs and CLIP with antibodies targeting Ezh2, Fus, and TDP-43 was performed under low RNase A conditions. qPCR was performed using primers targeting 3' UTR and internal sequences of IFNα, IFNβ, CXCL10, and IFIT1. CLIP enrichments are normalized to an IgG CLIP background control and significance is relative to Ezh2 and Fus CLIP binding. Data is representative of three experiments. *p<0.05.

As further validation that TDP-43 specifically associates with some of the 3'

UTRs of transcripts that it regulates, we performed crosslinked RNA pull down

assays (XL-RPD) for in vitro transcribed RNA sequences. The 3' UTRs of

GAPDH, beta actin, and several genes encoding innate immune cytokines were

in vitro transcribed and biotin was added to their 3' ends. The biotin-labeled 3'

UTR baits were then individually incubated with lysates from U2OS cells that

were stimulated with transfected vRNA, and RNA-protein interactions were

stabilized by UV irradiation. To reduce nonspecific RNA interactions, the lysates

were treated with high-concentration RNase A. The biotin-labeled RNA baits

bound to interacting proteins were then isolated with streptavidin-coated

magnetic beads and TDP-43 protein was probed with a specific antibody. TDP-

43 was not recovered by a nonsense probe or a GAPDH 3'UTR probe, but was

associated with all of the other immune cytokine 3' UTRs testeds, indicating that

TDP-43 has broad RNA binding activity in an *in vitro* assay but maintains some

specificity (S. Figure 6A).

TDP-43 affects 3'UTR transcripts at alternative polyadenylation signals

The 3'UTRs of type I interferons and ISGs do not contain the known GU-rich

TDP-43 consensus binding sequence and our CLIP experiments indicated that

AU-rich sequences were enriched in TDP-43 binding. In order to examine which

3' UTR RNA sequences are susceptible to TDP-43 regulation, we used a

massively parallel reporter assay (MPRA) system (Melnikov 2012) to sequence

the relative abundance of barcoded 3' UTR elements from fifty genes, including

interferons, ISGs, and several other genes shown to be regulated by TDP-43

perturbation (S. Table 2). We first used custom arrays to synthesize

approximately fifty thousand oligonucleotides that contained a library of 3' UTR

regulatory elements coupled to unique ten base pair DNA tags for individual

identification. The oligonucleotide pool was then cloned into a vector that

contained an invariant CMV promoter, an arbitrary luciferase open reading frame

(ORF), an SV40 polyadenylation signal, and an identifying sequence tag. This

entire plasmid pool was transfected into HEK293T cells and total RNA was

isolated thirty-six hours later. To determine the relative activities of each 3' UTR

element, we sequenced and counted the tags on the recovered mRNAs and the

original plasmid pool, and then generated ratios of these two counts (Figure 4A).

We calculated the distribution of medians from the control samples and the TDP-

43-depleted samples to generate a relative abundance for a given 3' UTR

element. By focusing on barcoded elements that were increased at least 1.5 fold

in the TDP-43-depleted samples compared to controls, we identified RNA

elements corresponding to twenty-seven genes that were significantly stabilized

in the TDP-43-depleted cells (Figure 4B, S. Table 2). Interestingly, all of the 3'

UTR elements that were stabilized in the absence of TDP-43 were enriched for

AU-rich elements (ARE) and often contained polyadenylation signals (PAS),

particularly the canonical PAS (AAUAAA). For example, the 3' UTR of TDP-43

has four known PAS and all of the reporter elements that were enriched in the TDP-43-depleted sample contained one of each of these PAS (Figure 4D). Likewise, the only PAS present in the shorter CXCL10, IFNa, and IFNβ 3' UTRs were stabilized in the absence of TDP-43 (Figure 4C and S. Figure 7B). Destabilization was specific, however, since not all 3' UTR elements containing PAS were affected. The PAS-containing sequence in the actinb 3' UTR was not differentially affected by TDP-43 (Figure 4C). TDP-43 is known to regulate the stability of cyclin-dependent kinase-6 expression (Cdk6)(Ayala 2008), presumably through a mechanism of mRNA destabilization. Tiles corresponding to regions of the Cdk6 3' UTR were the most significantly destabilized sequence elements when TDP-43 was depleted in the MPRA. Although the preferred binding site for TDP-43 is thought to be GU-repeats, AREs were enriched in the Cdk6 tiles that were stabilized by TDP-43.

Many of the genes that were dissected in the MPRA experiments were not expressed in the U2OS cells that were used for the CLIP-seq analysis. However, among these genes, there was an overlap between TDP-43-CLIP occupancy on their 3'UTRs and the RNA sequence that was stabilized by TDP-43 in the MPRA experiments. Ten out of the twenty-seven 3' UTRs that showed significant signal in the MPRA were also expressed in U2OS cells and were significantly enriched in TDP-43-CLIP binding. Six of the 3' UTRs corresponding to the ten expressed genes showed a direct overlap between the CLIP binding site and the sequence tile that is destabilized in the absence of TDP-43 (S. Table 2).

Figure 4: TDP-43 affects RNA sequences containing polyadenylation signals
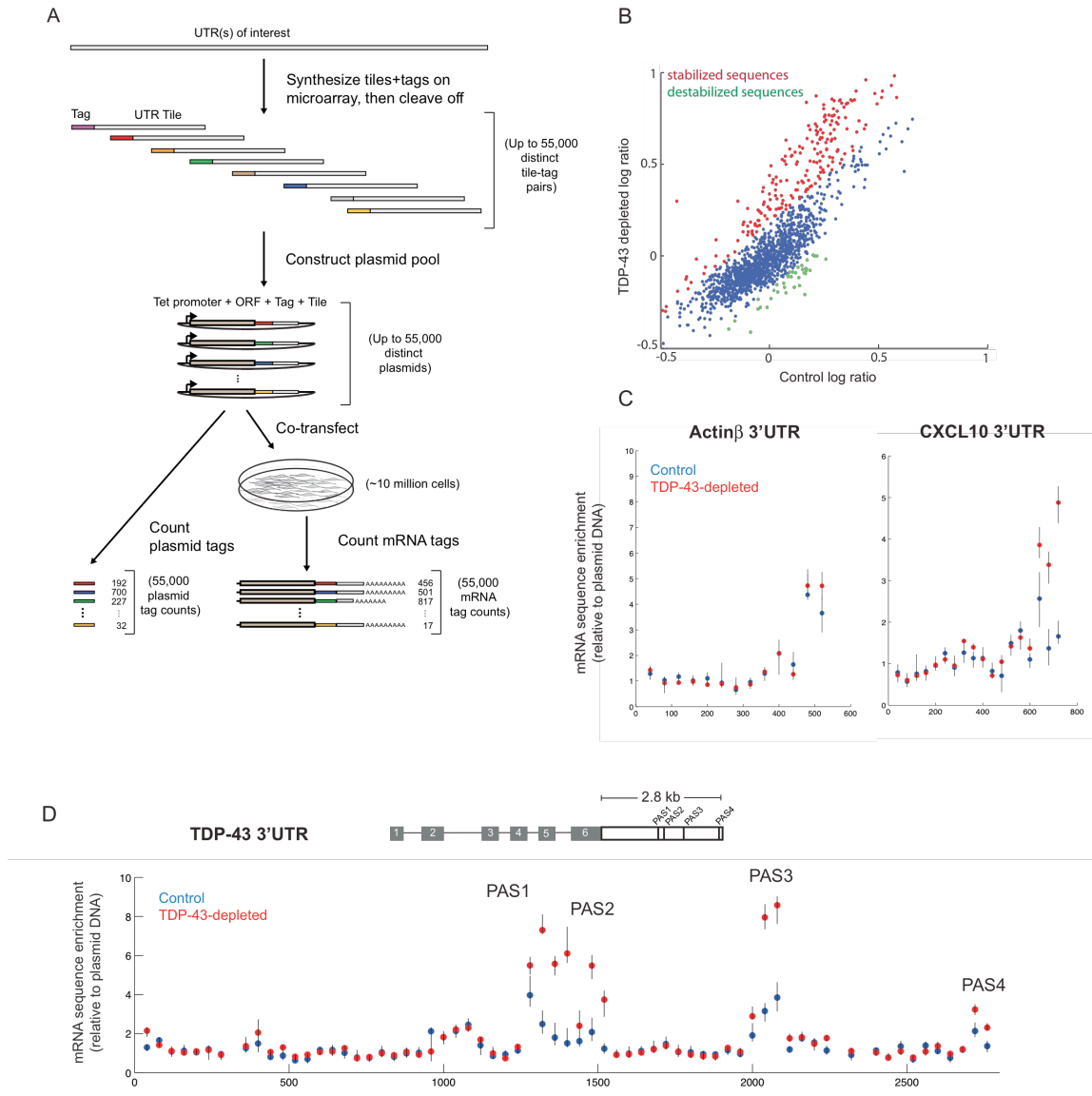
Figure 1.4: TDP-43 affects RNA sequences containing polyadenylation signals.
(A) Schematic of massively parallel reporter assay (MPRA) system. HEK293T cells were transfected with siRNA targeting a control sequence or TDP-43. 24 hrs later, cells were transfected with a plasmid pool encoding 55,000 DNA-tagged 200mer sequence tiles corresponding to 3' UTR sequences of 49 genes of interest.
(B) The log ratio of MPRA mRNA sequence tags that were differentially enriched (red) versus depleted (green) in the absence of TDP-43. The mRNA/plasmid tag ratios were normalized by multiplying by the ratio of the total number of plasmid and mRNA tag counts from the corresponding sequencing libraries. To estimate the relative abundance of each distinct UTR tile, the median of 32 mRNA/plasmid tag ratios were compared.

Figure 1.4 (Continued):
(C) mRNA sequence enrichment for MPRA tiles corresponding to actinβ and CXCL10 3' UTR sequences in TDP-43-depleted cells. Control cells (blue dots), TDP-43-depleted cells (red dots).
(D) mRNA sequence enrichment for MPRA tiles corresponding to TDP-43 3' UTR sequences. A schematic of the TDP-43 gene structure showing the arrangement of exons (gray boxes), the 3' UTR (white box), and polyadenylation signals (black lines).

Alternative PAS usage has been observed to be a major mechanism of gene expression regulation, particularly in rapidly proliferating cells (Mayr 2009, Sandberg 2008). In order to evaluate whether TDP-43 is directly mediating alternative PAS usage or if it is regulating the cell cycle, which in turn regulates PAS selection, we repeated the reporter assay with cells that were cycling at different rates. We found that there was no significant stabilization of 3'UTR elements in the MPRA for lowly proliferating versus highly proliferating cells (S. Figure 7A). Therefore, we can attribute the stabilization of RNA elements containing PAS in TDP-43-depleted cells to some function of the TDP-43 gene.

We wondered if TDP-43 functions to affect alternative PAS usage. To test this, we generated 3'-tag digital gene expression (DGE) RNA sequencing libraries and looked at the global effects of TDP-43 perturbation on PAS usage. Depleting TDP-43 had no apparent effect on the 3' lengths of mRNA transcripts (data not shown). TDP-43 apparently uses the sequence encoded in the 3' UTRs of genes it regulates to perform a function independent of 3' end processing.

TDP-43 destabilizes mRNA transcripts at their 3' UTRs

Although TDP-43 is a known splicing factor, it was unlikely that TDP-43 regulates IFNβ expression through a splicing mechanism since IFNβ is a single-exon gene. Additionally, the CLIP-seq and MPRA data pointed toward the 3' UTR as the site

of TDP-43 innate immune gene regulation. There are several examples of microRNA-independent innate immune cytokine regulation by 3' UTR destabilization (Brown 1996, Deleault 2008). To test if TDP-43 is regulating type I interferon and ISG transcript stability by 3' UTR destabilization, we cloned the 3' UTRs of several relevant genes into a vector driving a destabilized luciferase reporter. These 3'UTR reporter plasmids were individually transfected into HEK293T cells and their relative activity was measured. The depletion of TDP-43 had no effect on the stability of a control GAPDH 3' UTR. Therefore, in addition to renilla transfection normalization, the relative expression of all other 3' UTR reporter plasmids were normalized to GAPDH 3' UTR reporter activity. Cells depleted of TDP-43 or overexpressing WT TDP-43 or TDP-43 ALS disease alleles demonstrated two- to three-fold greater IFNβ and ISG reporter stabilization (Figure 5A).
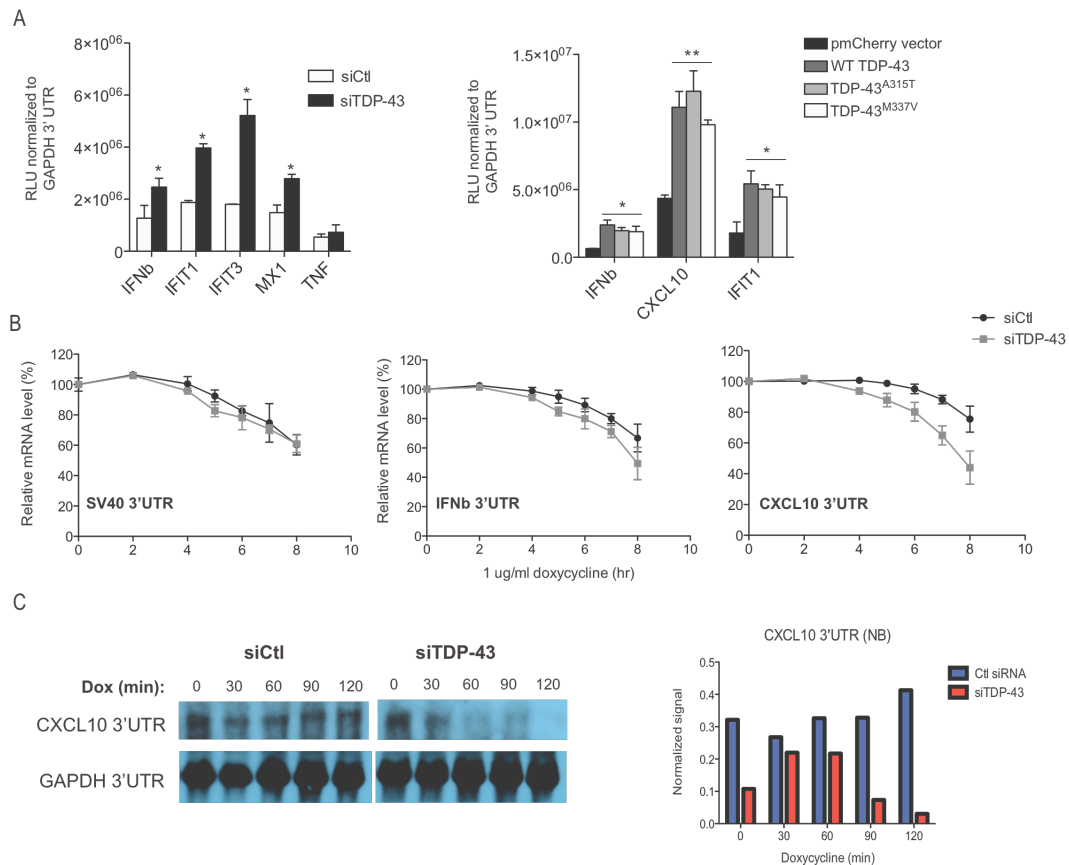
Figure 1.5: Perturbation of TDP-43 affects IFN and ISG transcript stability.

(A) Luminescent 3' UTR reporter assays in TDP-43-depleted and TDP-43-overexpressing cells. HEK293T cells were transfected with siRNA targeting a control sequence or TDP-43 (left), or were transfected with expression plasmids encoding an empty vector (pmCherry), WT TDP-43, TDP-43^{A315T}, or TDP-43^{M337V}. 24 hrs later, cells were transfected with firefly luciferase expression plasmids encoding 3' UTR sequences for innate immune genes. Cells were infected with SeV for 18 hrs. Values are normalized to an internal firefly renilla control and relative to GAPDH 3' UTR reporter plasmid activity. Data are representative of three experiments. *p<0.05, *p<0.01.

(B) Relative IFNβ and CXCL10 mRNA levels in TDP-43-depleted cells. Stable HEK293T were co-transfected with plasmids encoding 3'UTR sequences driving a luciferase reporter under tetracycline control and with siRNA targeting a control sequence or TDP-43. Doxycycline (1 ug/ml) was added to culture media over a time course (0, 2, 4, 6, 8, 10 hrs) to stop transcription from the reporter plasmids and relative luminescence was measured. Relative mRNA levels are given as a percent of total RNA for each condition. Data are representative of three experiments.

(C) Northern blot for CXCL10 3'UTR RNA decay. Same methods as in (B) but total RNA was isolated from cells and biotinylated RNA probes targeting the CXCL10 3'UTR transcript were used to quantify RNA levels by RNA hybridization. CXCL10 3' UTR signal normalized to total GAPDH 3'UTR transcript levels (bar graph, right).

The 3' UTR reporter assay data coupled with the CLIP-seq binding associations

suggested that TDP-43 binds to the 3' UTRs of IFN and ISGs and targets these

transcripts for decay. To test this hypothesis, we wanted to decouple

transcription from mRNA decay and measure the effects of TDP-43 on type I

interferon and ISG RNA decay rates. We cloned the 3' UTRs of several antiviral

genes into a vector encoding a destabilized luciferase reporter driven by a

tetracycline responsive promoter. The addition of doxycycline turned off

transcription and allowed us to measure relative RNA decay rates across

conditions. To our surprise, the rates of mRNA decay for IFNβ and ISG 3' UTRs

were faster in the absence of TDP-43, which argues against a model of 3'UTR-

mediated transcript destabilization (Figure 5B). To rule out any effects that TDP-

43 might have on the translation rate of the luciferase reporter protein, we

repeated the experiment but measured mRNA stability by RNA hybridization

using biotinylated RNA probes. By measuring RNA directly, we still observed a

faster rate of CXCL10 mRNA decay in TDP-43-depleted cells (Figure 5C). These

data indicated that TDP-43 normally stabilizes mRNA transcripts, a finding that

conflicts with the increase in type I interferon and ISG transcript levels that is

observed when TDP-43 expression is perturbed.

TDP-43 affects the transcription rates of type I interferons and ISGs

There have been reports of RNA binding proteins that regulate gene expression

both transcriptionally and by decay mechanisms (Harel-Sharvit 2010). Such

proteins, or "synthedegradases," are thought to couple together RNA metabolism

processes to more efficiently control the kinetics of a transcriptional response.

Although we did not see direct enrichment for TDP-43 at promoters by genome-

wide ChIP-sequencing experiments, the rapid IFNβ and ISG mRNA decay rates

in TDP-43-depleted cells suggests that TDP-43 might be exerting some

unexpected transcriptional effects on antiviral genes. To test this in a nuclear run-

on assay (NRO), 4-thiouridine (4SU) was pulsed into virally-infected U2OS cells

and incorporated into nascent mRNA transcripts. The 4SU cysteine sulfhydryl

groups were then labeled with biotin and the nascent transcripts were isolated

out of total cellular RNA using streptavidin-coated magnetic beads (Figure 6A).

The abundance of biotinylated transcripts was compared by qPCR to generate

ratios of nascent transcript enrichment in normal versus TDP-43-depleted cells.

Surprisingly, we observed a 5-10 fold increase in IFNβ and CXCL10 transcription

in the absence of TDP-43 (Figure 6B).

The NRO experiments demonstrated greater IFNβ and ISG transcription rates in

TDP-43-depleted cells. To validate that there is a true increase in transcription for

these genes specifically, we performed genome-wide polymerase (pol) II ChIP-

seq using an antibody that recognizes active, serine 2- and serine 5-

hyperphosphorylated pol II and analyzed pol II occupancy on antiviral genes in

normal versus TDP-43-depleted U2OS cells. We first performed a genome-wide

comparison of the median ratios of pol II enrichment at transcription start sites

(TSS) and across gene bodies for TDP-43-depleted cells versus normal cells.

Then, we compared the ratios of pol II enrichment for a set of antiviral genes (S.

Table 3) versus all other genes in normal versus TDP-43-depleted cells.

Phosphorylation of pol II serine 5 is typically associated with transcription

initiation at promoter regions and pol II serine 2 phosphorylation is enriched for

gene bodies (Komarnitsky 2000). There was a marginal trend toward increased

Pol II serine 5 on the TSS of antiviral genes ($p < 0.06$), but no change over the

remaining transcription units (TU; $p < 0.3$). However, we found approximately

30% higher pol II serine 2 occupancy on both the TSS and gene bodies of our

antiviral gene set when TDP-43 is depleted (Figure 6C). This confirms that the

increase in nascent transcript abundance observed in the NRO experiments is

associated with greater pol II recruitment and transcriptional elongation.
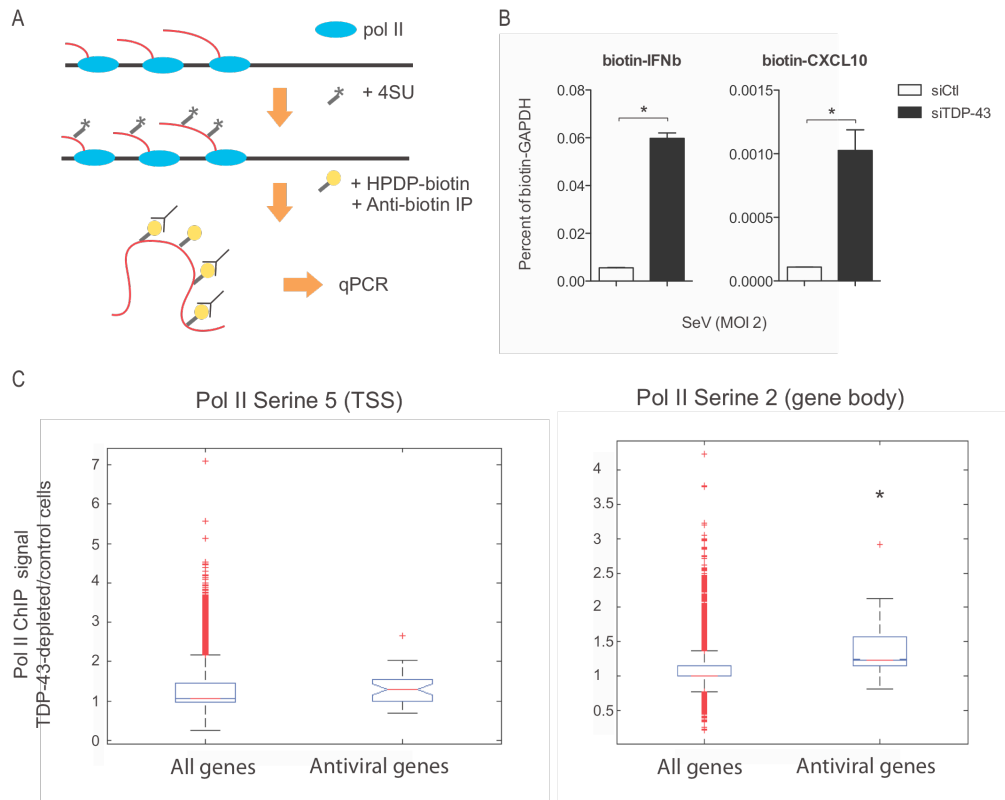
Figure 1.6: Depletion of TDP-43 results in greater IFNβ and ISG transcription.

(A) Schematic of 4-thiouridine (4SU) nascent transcript labeling protocol.　　**(B)** qPCR of IFNβ and CXCL10 nascent transcript levels in TDP-43-depleted cells. U2OS cells were transfected with siRNA targeting a control sequence or TDP-43, then infected with SeV (MOI 2) for 16 hrs. Cells were pulsed with 4SU (0.5 mM) for 5 minutes to label nascent transcripts and immediately lysed. Total RNA was isolated and sites of 4SU incorporation were labeled with biotin. Biotinylated nascent transcripts were isolated using streptavidin magnetic beads. Error bars represent the standard error of the mean (SEM). *$p < 0.001$.
(C) Box plots for phosphorylated polymerase II (pol II) chromatin immunoprecipitation (ChIP) enrichment for antiviral genes in TDP-43-depleted cells. U2OS cells were transfected with siRNA and infected with SeV as in (B). ChIP was performed using a antibodies that recognize hyperphosphorylated pol II at serine 2 (left) and serine 5 (right). Genome-wide comparison of the median ratios of pol II enrichment at transcription start sites (TSS) and across gene bodies for TDP-43-depleted cells versus normal cells was performed and the ratios of pol II enrichment for a set of antiviral genes versus all other genes in normal versus TDP-43-depleted cells was assessed.

TDP-43 regulates IFNβ promoter activity through the IFNβ 3' UTR

polyadenylation signal

Our CLIP-seq data analysis demonstrated that TDP-43 associates with the 3'
UTRs of antiviral genes. Surprisingly, we also observed that TDP-43 affects both
transcription and decay rates of regulated genes. We performed multiple ChIP-
seq experiments in an attempt enrich for DNA sequences that bind to
endogenous or overexpressed TDP-43 protein, but were not able to uncover any
specific interactions between TDP-43 and DNA. However, the MPRA results
suggest that TDP-43 functions on polyadenylation signals in mRNA transcripts.

TDP-43 perturbation had no significant effect on the activity of a luciferase IFNβ
promoter reporter with a downstream SV40 3' UTR (Figure 2). We wondered if
the effect of TDP-43 on transcription is dependent on the presence of the specific
3' UTR of the gene that it is regulating. To test this, we modified the original IFNβ
promoter reporter by inserting the human IFNβ 3' UTR sequence between the
luciferase ORF and the SV40 3' UTR (Figure 7a). We then compared the relative
activities of these reporter constructs in HEK293T cells that expressed
endogenous levels of TDP-43 or overexpressed forms of TDP-43. Although
overexpressing TDP-43 had no effect on the original IFNβ promoter reporter
construct, overexpression stabilized the IFNβ promoter-driven luciferase activity
when the IFNβ 3' UTR sequence was inserted downstream of the luciferase ORF
(Figure 7A). Deleting the polyadenylation signal from the IFNβ 3' UTR sequence,

however, rescued the derepression of the IFNβ promoter activity by TDP-43. TDP-43 does not appear to act directly on DNA since TDP-43 had no effect on the reporter when the IFNβ 3' UTR was placed upstream of the promoter sequence. Similarly, the depletion of TDP-43 led to a PAS-dependent regulation of the IFNβ promoter activity (Figure 7B).
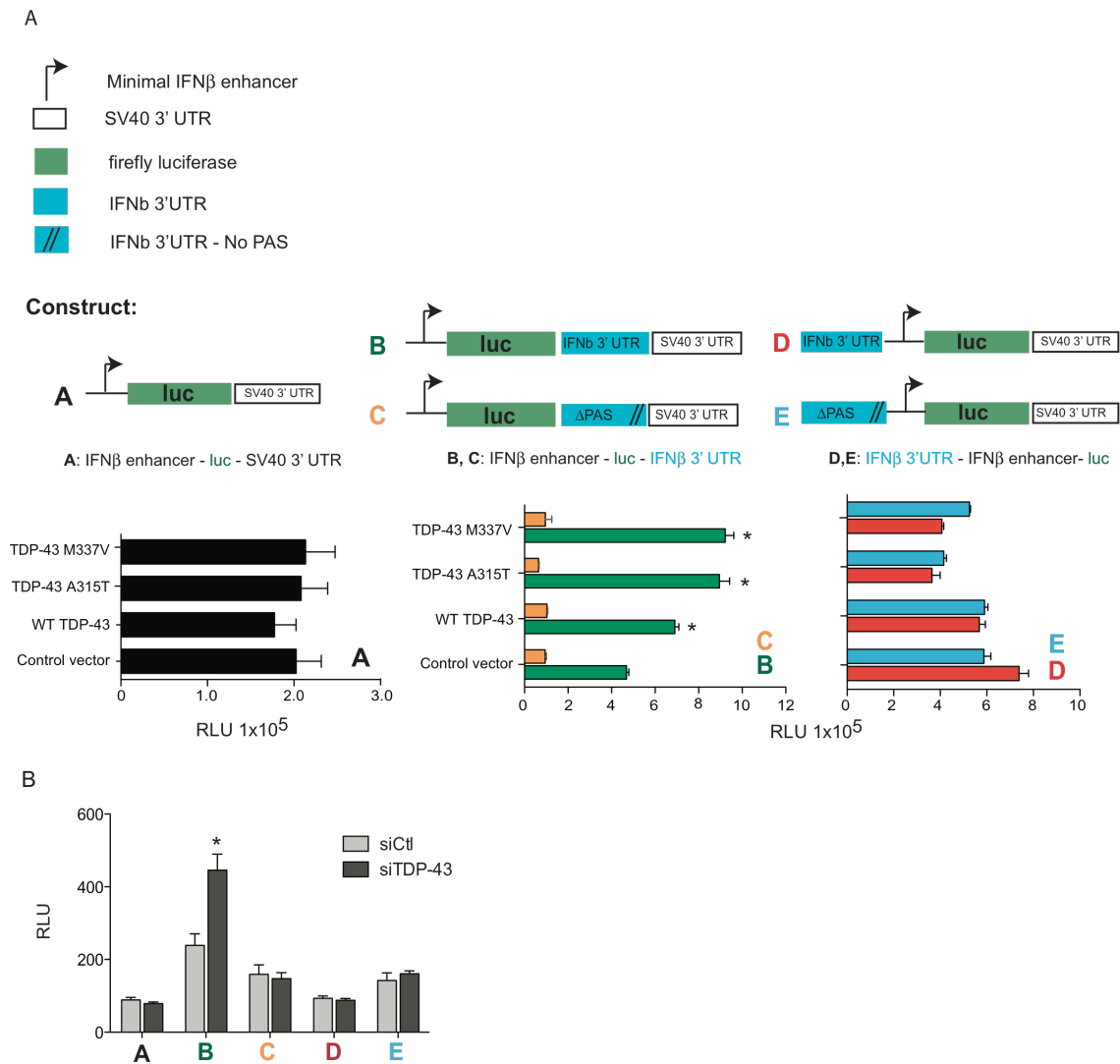
Figure 1.7: TDP-43 requires a polyadenylation signal in the IFNβ 3' UTR to inhibit IFNβ promoter activity.

(A) Luminescent reporter assays in TDP-43-overexpressing cells for various IFNβ enhancer reporter constructs. HEK293T cells were co-transfected with plasmids encoding an empty vector, WT TDP-43, TDP-43[A315T], or TDP-43[M337V] and the indicated IFNb enhancer driving a luciferase reporter: the minimal IFNb enhancer with a standard SV40 3' UTR (left, black bars), the minimal enhancer with the IFNb 3'UTR (middle) with the PAS (green bars) or without the PAS (yellow bars), or the minimal IFNb enhancer with the IFNb 3'UTR upstream of the promoter region (right) with the PAS (red) or without the PAS (blue).
(B) The same constructs as in (A) but with HEK293T cells that were transfected with siRNA targeting a control sequence or TDP-43.

79

Discussion

Understanding defects in both RNA processing factors and misregulated innate immune pathways are becoming key areas of focus in neurodegeneration research. The work presented here links TDP-43 to the regulation of the antiviral gene pathway. TDP-43 is a multifunctional RNA binding protein that has already been assigned a variety of roles across all stages of gene expression, but we find that it can modulate IFNβ expression by a novel mechanism of *cis* promoter control directed by IFNβ 3' UTR sequences. More specifically, the PAS in the IFNβ 3' UTR is sufficient to mediate the regulation of the IFNβ promoter by TDP-43.

IFNβ is a short, single-exon gene that is not spliced and in theory would not recruit splicing factors. Nonetheless, we were not surprised to find enrichment for the 3' UTR of IFNβ with TDP-43 protein since levels of IFNβ can be destabilized by an ARE located in its 3' UTR (Paste 2003) and TDP-43 regulates the transcript levels of other genes through 3' UTR associations (Fiesel 2009, Volkening 2009, Kim 2010). However, we were very surprised to find that the rate of IFNβ and CXCL10 degradation was actually faster in the absence of TDP-43 (Figure 5c). This finding contradicted our initial hypothesis that TDP-43 normally destabilizes IFNβ transcripts and pointed to transcriptional regulation. Nuclear run on assays confirmed that more nascent IFNβ transcripts are produced in the absence of TDP-43, and pol II ChIP experiments demonstrated that there is

significantly more pol II on the TSS and gene body of IFNβ in TDP-43-depleted

cells. We also see greater transcriptional activity on many ISGs but presume that

this is a downstream consequence of higher IFNβ levels.

3' UTRs are better known for their ability to regulate gene expression by

mechanisms of transcript destabilization. We are only aware of one other

instance in which a 3' UTR sequence is required for the repression of promoter

activity. Sequences encoded in the rat serine protease inhibitor (*spi*) 2.3 3' UTR

were shown to specifically repress transcription from the *spi* 2.3 gene promoter

(Le Cam and Legraverend 1995). Although we find a minimal requirement for the

IFNβ 3' UTR PAS to confer the effect of TDP-43 on IFNβ enhancer activity

(Figure 7), no specific sequence motifs in the *spi* 2.3 3' UTR were identified to be

critical for *spi* 2.3 transcriptional regulation. In fact, multiple fragments of the *spi*

2.3 3' UTR were equally capable of repressing *spi* 2.3 transcription.

Direct binding of the IFNβ promoter sequence by TDP-43 seems unlikely. NF-κB

is one of the transcription factors that drives IFNβ promoter activity and TDP-43

has been found to function as a co-activator of the p65 subunit of NF-κB, but the

RNA binding domains of TDP-43 are dispensable for this function (Swarup 2011)

and we find that the c-terminal domain of TDP-43 contributes to IFNβ

misregulation (Figure 1g). Furthermore, we performed multiple endogenous and

overexpressed and tagged TDP-43 ChIP-seq experiments to identify DNA

targets of TDP-43 but were unsuccessful in recovering any specific interactions

81

(data not shown). Based on our data, one very likely possibility is that TDP-43 does indeed regulate the IFNβ promoter indirectly via a larger complex that is tethered to the nascent IFNβ transcript and is not captured by standard ChIP crosslinking protocols.

We only investigated the direct effect of TDP-43 on the IFNβ 3' UTR, but our CLIP and MPRA data suggest that TDP-43 might directly regulate other genes using a similar 3' UTR *cis* mechanism. We found enrichment for TDP-43 binding on the 3' UTRs of key ISGs including IFIT1 and CXCL10. Additionally, MPRA sequence tiles corresponding to regions that contain PAS in the IFIT1 and CXCL10 3' UTRs were stabilized in the absence of TDP-43. Further work to dissect how global the direct function of TDP-43 on the 3' UTRs of genes that it regulates is warranted.

Proteomic analyses of TDP-43 interaction partners show that TDP-43 is heavily enriched for associations with many RNA splicing factors (e.g. SF2, TIA-1), translation machinery (e.g. PABP2, eEF1A1 and eIF3F), and other proteins that regulate mRNA abundance including those in the nonsense mediated decay pathway (e.g. UPF1) and mRNA destabilization (e.g. Musashi 2) (Kim 2010, Freibaum 2010). It has been known for some time that extensive functional coupling exists among RNA processing factors that mediate biochemically distinct stages of gene expression (Maniatis and Reed 2002). In fact, all nuclear mRNA processing events appear to occur at the site of transcription (Krause

1994, Wansink 1993, Iborra 1996, Mortillaro 1996, Carmo-Fonseca 1991, Neugebauer 1997, Schul 1996, Misteli 1997). Not only are gene expression processes coupled, they are interdependent. For example, 3' end processing is dependent on adequate 5' cap formation (Lewis 1995, Flaherty 1997). Conversely, deleting the AAUAAA polyadenylation signal from a 3'UTR results in defects in alternative splicing and transcriptional termination (Niwa and Berget 1991, Connelly 1988). Like TDP-43, the splicing factor SC35 has been shown to affect pol II processivity and transcriptional elongation (Lin 2008).

Additionally, there is an emerging literature demonstrating that individual RNA binding proteins can regulate multiple stages of gene expression, from promoter regulation and transcription to mRNA decay and translation. Such proteins are termed "RNA coordinators" (Choder 2011). For example, the yeast RNA pol II subunits Rbp4p and Rbp7p interact with processing pol II, then bind to and travel with the nascent mRNA transcript to the cytosol where they are involved in either directing the transcript to be decayed or they augment translation initiation (Harel-Sharvit 2010). However, their initial communication with the polymerase and "imprinting" is required for these later functions.

We wonder whether TDP-43 functions as an mRNA coordinator. In our model, TDP-43 binds directly to the 3' UTR of nascent mRNA transcripts and uses the PAS, likely with 3' end processing factors, to signal back to the promoter and dampen pol II activity. In this way, TDP-43 might function as an mRNA "counter"

to rapidly and locally modulate the number of transcripts that are generated from the IFNβ locus. IFNβ is a critical regulator of antiviral immunity and this layer of local regulation might serve as another mechanism to limit the toxicity associated with an antiviral response.

Since the overexpression and depletion of TDP-43 result in similar phenotypes, and both the RNA and protein interaction partners of disease forms of TDP-43 are similar to WT TDP-43 (Kim 2010), there is a key mechanistic link that is missing. Although sequestration of aggregated TDP-43 to the cytosol, as is the case in some instances of TDP-43-linked neurodegeneration, would prevent TDP-43 from directly interacting with target genes, TDP-43 remains nuclear in our experimental system (S. Figure 3). Though much of the overexpressed TDP-43 protein aggregates and appears to be ubiquitinated (S. Figure 2), we can still recover specific interactions by CLIP. TDP-43 has putative prion domains that might cause nuclear aggregation and prevent TDP-43 from appropriately interacting with other proteins that are recruited to the 3' UTR PAS (Udan and Baloh 2011). The most likely TDP-43 interaction partners would be members of the massive eight-five-protein 3' end processing complex (Shi 2009). Although TDP-43 was not identified as a member of this complex, many of its binding partners are. Furthermore, TAF15 is a core component of the polyadenylation complex, shares sequence homology with TDP-43 and Fus/TLS, and has been linked to ALS (Couthouis 2012), suggesting that defects in genes like TDP-43 that regulate RNA 3' end processing contribute to neurodegeneration.

Our data suggest that if TDP-43 is not able to interact with 3' end processing machinery, than it cannot repress IFNβ promoter activity. This mechanism is not totally surprising since the 3' UTR PAS has been shown to regulate both alternative splicing and transcriptional termination (Niwa and Berget 1991, Connelly and Manley 1988). These data present a new instance in which the PAS also regulates transcriptional elongation (Figure 8). Defective TDP-43-mediated transcriptional repression might have significant consequences for neurodegeneration. A compromise in the ability to downregulate type I interferon can be toxic to any tissue, but the large, post-mitotic motor neurons of the central nervous system might be even more susceptible to elevated levels of antiviral cytokines. There is some evidence that repressing the type I interferon signaling in an ALS mouse model slows disease progression (Wang 2011). The pro-inflammatory arm of innate immune signaling has been targeted for ALS therapeutics in the past, but perhaps attention should be turned to the antiviral pathways.

Figure 8: Model
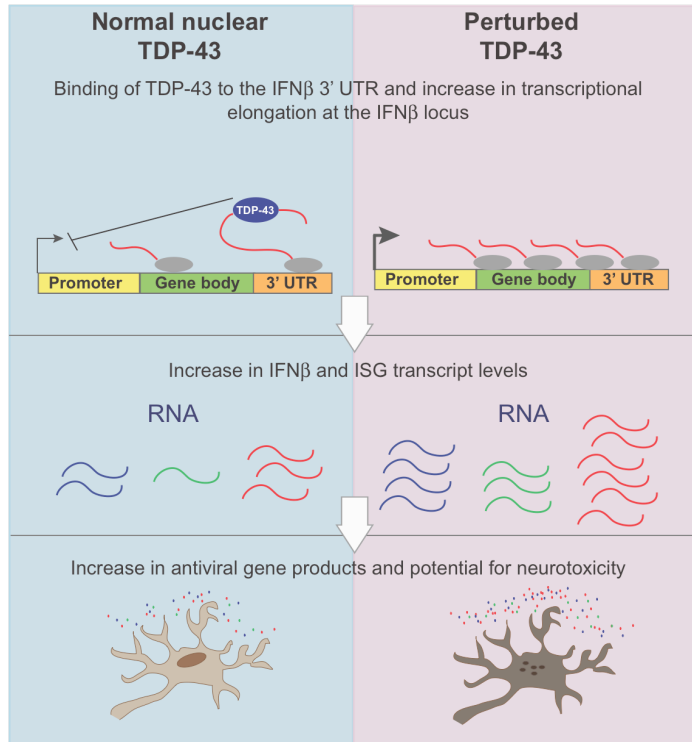
Figure 1.8: Model: TDP-43 regulates IFNβ transcriptional elongation through a mechanism that involved the polyadenylation sequence in the IFNβ 3' UTR. TDP-43 normally associates with the 3'UTR of IFNb transcripts. In the presence of an intact polydenylation signal in the IFNb 3' UTR, TDP-43 is able to repress IFNb promoter activity and transcriptional elongation *in vivo* (top panel). When normal TDP-43 function is perturbed, the ability of TDP-43 to repress transcription and IFNb mRNA transcript levels increase, leading to an increase in the expression of interferon-stimulated genes as well (middle panel). Increases in toxic antiviral immune mediators might contribute to neurodegeneration (bottom panel).

Experimental Procedures

Mouse dendritic cells

Six- to eight- week-old female C57BL/6J mice were obtained from the Jackson Laboratories. Bone marrow was collected from femora and tibiae. Bone marrow dendritic cells (BMDCs) were derived from plating total bone marrow onto non-tissue culture treated plastic dishes (Corning) and growing the cells for one week in RPMI medium (Gibco, Carlsbad, CA, Invitrogen, Carlsbad, CA) supplemented with 10% FBS and 1% L-glutamine, penicillin/streptomycin, MEM non-essential amino acids, HEPES, sodium pyruvate, 0.1% beta-mercaptoethanol, and 15 ng/mL GM-CSF (Peprotech, Rocky Hill, NJ). BMDCs were also generated from IRF3/7-/- and IFNR-/- mice (gifts from K. Fitzgerald).

Cell lines

HEK293T cells, U2OS cells, and mouse lung fibroblasts were cultured in tissue culture-treated plastic dishes in DMEM media (Gibco, Carlsbad, CA) supplemented with 10% FBS and 1% penicillin/streptomycin. The cells were grown in an incubator at 37°C. Primary MEFs were generated from C57BL/6J, IRF3-/-, and MAVS-/- mice (gifts from J. Kagan).

Viruses

Sendai virus, Cantell strain (ATCC, Manassas, VA) was used for infections. Influenza A virus strain A/PR/8/34 was grown in Vero cells, and virus titers from

cell supernatants were quantified using 293T cells transfected with a vRNA luciferase reporter plasmid.

Antibodies and reagents

Anti-TDP-43 (ProteinTech Group, Chicago, IL), Fus (AbCam, Cambridge, MA), pIRF3 (Cell Signaling, Danvers, MA), pSTAT1 (AbCam) antibodies were used for immunoprecipitations and Western blotting.

pmCherry plasmids encoding WT TDP-43, TDP-43A315T, and TDP-43M337V were a gift from R. Baloh.

RNA isolation

mRNA was isolated using the RNeasy mini kit according to the manufacturer's instructions (Qiagen, Valencia, CA). RNA was reverse transcribed with the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA).

Total RNA was isolated using phenol:chloroform:isoamylalchohol (Invitrogen, Carlsbad, CA). RNA was precipitated overnight at -80°C from the aqueous phase by adding an equal volume of 100% isopropanol and 1 ul of glycogen (Thermo Scientific, Rockland, Il). The RNA was pelleted by centrifuging for 45 minutes at

4°C and then washed with 80% ethanol to retain small RNA species. RNA

sample quality was assessed on a 2100 Bioanalyzer (Agilent, Palo Alto, CA).

For experiments with very low RNA yield, RNA was reverse transcribed using the

Sensiscript RT kit (Qiagen).

qPCR measurements

Quantitative real time PCR (qPCR) reactions were performed on the LightCycler

480 system (Roche, Indianapolis, IN) using the FastStart Universal SYBR Green

Master Mix (Roche). Every reaction was run in duplicate and endogenous

GAPDH levels were used for normalization.

For bar graphs, the unpaired 2-tailed student's T-test was used to compute p

values, except where stated otherwise. Error bars reflect the standard error of the

mean (SEM), except where stated otherwise. All statistical analyses were carried

out using GraphPad Prism 4.0 and the R statistical environment.

shRNA infection

High-titer lentiviruses expressing shRNAs were obtained from The Broad RNAi

Platform and used to infect cells as previously described (Moffatt et al., 2006).

Primary BMDC were infected according to previously published protocols (Amit

2009). For all other cells, 5000 cells were plated on flat bottom 96-well plates (Corning) and 24 hrs later spin-infected with shRNA encoding lentiviruses (MOI 2-10) supplemented with polybrene (8 ug/ml). 24 hours after infection, the cells were selected using puromycin (2 ug/ml) in culture medium. Cells were used in experiments after 4 days of selection.

siRNA transfection

siRNAs targeting a negative control sequence or TDP-43 (Dharmacon siGenome) were transfected into cells using Lipofectamine RNAiMax (Invitrogen) according to manufacturer's instructions.

Plasmid transfection

Plasmid DNA was transfected using TransIT LT1 reagent (Mirus, Madison, WI) according to manufacturer's protocols.

Luminescent reporter assays

Plasmid DNA encoding luminescent reporter proteins and ORFs or 3' UTRs of interest were transfected into cells using TransIT LT1 (Mirus). The Dual Glo Luciferase Assay System (Promega, Madison, WI) was used to read luminescence using the Envision Multilabel Reader (Perkin Elmer, Waltham, MA).

4-thiouridine metabolic labeling of nascent transcripts

Nascent RNA transcripts were isolated from cells that were metabolically labeled using 4-thiouridine (4SU). We labeled U2OS cells following infection with SeV with 4SU (final concentration of 500 uM) for pre-defined labeling times. The cells were then washed three times in cold PBS and lysed in buffer RLT (Qiagen). We isolated mRNA, used EZ-Link HPDP-Biotin (Thermo Scientific, Rockford, II) to covalently link biotin to 4SU residues, and isolated the 4SU-labeled RNA (RNA-4SU) using biotin capture with streptavidin magnetic beads according to previously published protocols (Rabani et al.). Total mRNA and captured biotinylated RNA was quantified either by standard qPCR or by Illumina sequencing.

UV-crosslinking and Immunoprecipitation and high throughput RNA sequencing

Protocol based on UV-crosslinking and immunoprecipitation with high throughput RNA sequencing (HITS-CLIP), (Chi et al., 2009). In brief, 100 million U2OS cells or U2OS cells stably expressing a tetracycline-inducible vector encoding GFP or GFP-WTTDP-43 or GFP-TDP-43^{A315T} were treated with tetracycline (0.25 ug/ml) for 24 hrs. Then, the cells were infected with SeV (MOI 5) for 18 hrs to induce an antiviral state. The cell media was removed and replaced with PBS and the cells

were UV-crosslinked (265 nm) for two rouns of 4000 J, on ice. Cells were

washed with PBS containing protease inhibitors three times and the cell pellets

were resuspended in a low salt buffer (100 mM NaCl, 0.1% sodium dodecyl

sulfate (SDS), 0.5% sodium deoxycholate, 0.5% Nonident P-40) with protease

inhibitors. Cells were sonicated with a Branson probe sonicator for three cycles

of 10 seconds (0.7 sec on/1.3 sec off) at 20% power on ice. The lysates were

then rested on ice for 20 mins. RQ1 DNase (Promega) was added to the lysates

and incubated at 37° C for 5 min. RNAse A (USB) was added at a concentration

of 1:1000 and incubated at 37° C for 10 minutes. Lysates were spun and

supernatants were incubated with protein A Dynabeads that were pre-incubated

either with anti-TDP-43 (ProteinTechGroup) to immunoprecipitate endogenous

TDP-43 or with GFP-TRAP beads (Chromo Tek) in low salt buffer. Beads were

washed with low salt wash buffer three times followed by two washes with high

salt buffer (500 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% Nonident

P-40). Proteinase K (5 ug) was added to the washed beads and incubated at 70°

C for 1.5 hrs. RNA was extracted using standard Phenol/Chloroform extraction

methods and the resulting RNA pellets were washed with 80% ethanol to retain

short RNA fragments. RNA sizes were assessed on a 2100 Bioanalyzer (pico

chip) and the sizes of extracted RNA ranged between 40 – 150 bases. RNA

libraries were prepared and barcoded according to standard Broad RNA

sequencing protocols. Single-end Illumina sequencing was performed.

Detection of TDP-43 binding sites

We focused our CLIP analysis on the 5' UTR, exons, and 3' UTR of genes but did not consider intronic regions. The search for enriched peaks in the TDP-43 immunoprecipitation sample compared to the input control was performed by scanning each gene using sliding windows of 100 nucleotides with 50 nucleotides overlap. The mean coverage for each window was calculated for the immunoprecipitated and control samples (MeanWinIP and MeanWinControl, respectively). Gene median coverages for immunoprecipitation and input control were determined (MedianGeneIP and MedianGeneControl, respectively) to robustly estimate background levels. Every window was assigned the following unit-less metric, representing the enrichment fold change of immunoprecipitated over input control samples after normalization by background.

A robust estimation signal was ensured by setting a limit of mean window coverage >20 for the immunoprecipitated sample. The FDR of this method was estimated by comparing the number of enriched windows exceeding a set threshold using this approach to the number that is obtained if the experiments were computationally reversed with input being treated as immunoprecipitate and vice versa. For a threshold of 2, corresponding to a fourfold increase with respect to control, the FDR was <7%; thus, the winscore threshold was set to 2. All overlapping enriched windows were merged and positions with maximal immunoprecipitate-positive and immunoprecipitate-negative coverage were

identified. If the distance between these two maxima was <100 nucleotides and the ratio between the amplitude of these two maxima was less than twofold, a putative TDP-43 binding site was defined in the centre of these two maxima, and a score was allocated to this peak, peakscore, corresponding to the maximal winscore from among all the overlapping windows. When these conditions were not met, windows were classified as non-resolved peaks and excluded from the analysis, to avoid PCR/sequencing artifacts. The negative control peaks were detected as described above: All windows with winscore <0 were selected. Control windows were selected from the same genes as the detected TDP-43 binding peaks (limited to a mean window coverage >20). All overlapping control windows matching these criteria were merged and a negative control peak was defined in their centre. GFP peaks were subtracted from TDP-43 peaks to remove background artifacts.

For genome-based analysis, the peak-caller MACS[51] was used for peak detection; the 'effective genome size' parameter was adjusted to the calculated transcriptome size (1.35×10^8). Peaks were considered if their MACS-assigned fold change was >4 and individual FDR value <5%.

Identification and clustering of enriched motifs

In-house method for detecting motifs using all identified peaks. Motifs enriched within TDP-43 binding peaks compared with control peaks were identified by counting the occurrence of 4–6-nucleotide *k*-mers in the immunoprecipitate and

its corresponding control group. The total number of k-mers of each length within every group was counted and the ratio between their prevalence was used to calculate the fold change between the two groups. Exact Fisher test was used to evaluate the differences in the prevalence of each k-mer between the groups. Analysis was limited to motifs enriched more than twofold and with an associated Bonferonni-corrected P value < 0.05. To correct for the underlying base composition we repeated these analyses using a second set of control sequences based on randomly permuting the sequences of the TDP-43 binding sites, validating the significance of the found motif.

Massively parallel reporter assays (MPRA):

Oligonucleotide library design and synthesis

200mer oligonucleotides were designed to contain, in order, the universal primer site ACTGGCCGCTTCACTG, a 10 nt variable tag sequence, the Tag-Seq primer site AGATCGGAAGAGCGTCG, a 145 nt UTR tile and the universal primer site GACAATGCAGGCCTCG. The tile sequences were selected to cover the annotated UTRs of 49 RefSeq genes with a step size of 40 nt, and each of the resulting 1,683 tiles were linked to 32 distinct tags, for a total oligo library complexity of 53,856. The distinct tags were selected from randomly generated 10 nt sequences, with the following constraints: (1) must contain all four nucleotides, (2) must not contain a run of more than four identical nucleotides, and (3) must not contain a known mammalian microRNA seed sequence

(obtained from http://www.targetscan.org, April 2009). The resulting

oligonucleotide libraries were synthesized by Agilent, Inc. as previously

described [8].

Plasmid construction

To generate the UTR MPRA vector, we obtained a commercially synthesized

DNA sequence (Genscript) containing, in order, a transcriptional terminator and

synthetic polyA site derived from pGL4.10 (Promega), the TRE3G Dox-inducible

promoter (Clontech), the luc2P destabilized luciferase gene, two SfiI sites

separated by a 38 nt stuffer and the SV40 UTR/polyA signal. This sequence was

cloned into the EcoRI and BamHI sites of pUC57.

To insert UTR tiles into the MPRA vector, full-length oligonucleotides were

isolated using PAGE and then amplified and *SfiI*-tailed using 20-26 cycles of

emulsion PCR as described in [Schütze et al. Analytical Biochemistry 410 (2011)

155-157] using Herculase II Fusion DNA Polymerase (Agilent) and primers

GCTAAGGGCCTAACTGGCCGCTTCACTG and

GTTTAAGGCCACCGAGGCCAGCTTTGTC to add SfiI sites. The resulting PCR

products were digested with SfiI (NEB) and directionally cloned into the SfiI

digested MPRA plasmid using One Shot® TOP10 Electrocomp™ E. coli cells

(Invitrogen). To preserve library complexity, the efficiency of transformation was

maintained at >3×10^8 cfu/µg.

Tag-Seq

mRNA was extracted from total RNA using MicroPoly(A)Purist™ kits (Ambion) and treated with DNase I using the Turbo DNA-free™ kit (Ambion). First-strand cDNA was synthesized from 400-700 ng mRNA using High Capacity RNA-to-cDNA kits (Applied Biosystems).

Tag-Seq sequencing libraries were generated directly from 12% of a cDNA reaction or 50 ng plasmid DNA by 26 cycle PCR using Pfu Ultra HS DNA polymerase 2x master mix (Agilent) and primers AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT and CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGAGGTGCCTAAAGG (where XXXXXXXX is a library-specific index sequence). The resultant PCR products were size-selected using 2% agarose E-Gel EX (Invitrogen). The libraries were sequenced in indexed pools of eight, or individually, using 36 nt single-end reads on Illumina HiSeq 2000 instruments.

To infer the tag copy numbers in each Tag-Seq library, all sequence reads were examined, regardless of their quality scores. If the first ten nucleotides of a read perfectly matched one of the 53,856 designed tags and the remaining nucleotides matched the expected upstream MPRA construct sequence, this was counted as one occurrence of that tag. All reads that did not meet this criterion were discarded. All tags that did not have a count of at least 20 in every

sequenced plasmid pool were also discarded. The mRNA/plasmid tag ratios

were normalized by multiplying by the ratio of the total number of plasmid and

mRNA tag counts from the corresponding Tag-Seq libraries. To estimate the

relative abundance of each distinct UTR tile, we compared the median of their 32

mRNA/plasmid tag ratios.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) for endogenous and overexpressed

3xFlag-TDP-43 were performed according to published protocols (Mikkelsen

2007). The commercially available antibody was used for endogenous TDP-43

(ProteinTech Group). Polymerase II ChIP protocol and reagents used were as

per published protocols (Rahl 2010).

CHAPTER 2: The RNA binding proteins Fus and TDP-43 play roles in innate immunity and ALS disease pathophysiology

Introduction

Two independent RNAi screens performed in the Hacohen lab identified the nuclear RNA binding protein, Fus, to be a negative regulator of the type I interferon response (Amit 2009, Shapira 2009). Subsequently, I identified a protein with similar structural motifs as Fus, TDP-43, to also be a negative regulator of innate immune genes (See Chapter 1). Mutations in both of these proteins have been linked to a variety of human neurodegenerative diseases, particularly familial amoytrophic lateral sclerosis (ALS). Here, we show that Fus and TDP-43 are both transcriptionally and post-translationally regulated in response to viral pathogen sensing. Fus and TDP-43 were found to negatively regulate type I interferon production as well as RNA polymerase (pol) I and III gene expression. Furthermore, the proteins associate with IFN and pol III mRNA in vivo. Despite this increase in antiviral gene activity, Fus- and TDP-43- depleted cells are impaired in their ability to control influenza A virus, vesicular stomatitis virus, and reovirus replication. The defect in controlling influenza virus replication is reversed upon the administration of a pol III inhibitor. The overexpression of Fus and TDP-43 also fails to inhibit viral replication. The transcriptional regulation and physiological functions of Fus and TDP-43 in viral immunity were not previously reported, and we believe these genes play a critical role in controlling viral immunity through a novel and unknown mechanism.

Interestingly, fibroblasts derived from a familial ALS patient harboring a TDP-43

mutation produced more type I interferon in response to influenza virus infection than control cells. Additionally, RNA from the frontal cortex of deceased ALS individuals' brains also have a striking antiviral and pol III gene phenotype. As in human ALS, a missense mutation in the canine SOD1 protein (E40K) causes an upper and lower motorneuron disease called canine degenerative myelopathy (DM). We found that the spinal cords of diseased dogs also exhibit higher expression of antiviral genes.

There are many connections between the innate immune system and neurodegenerative diseases. Our preliminary data suggests that the innate immune functions of Fus and TDP-43 might contribute to ALS disease pathophysiology.

Background

Fus (human: TLS (translocated in sarcoma); hnRNP P2) is a ubiquitously expressed, nuclear RNA binding protein in the TET protein family. Structurally, Fus has an N-terminal activation domain, three arginine-glycine-glycine- rich repeats, and an RNA recognition motif. This 74 kD protein has been shown to bind both DNA and GGUG-rich RNA sequences (Lerga 2001), as well as regulate gene transcription (Wang 2008) and alternative splice site selection (Hallier 1998). The Manley lab demonstrated that Fus negatively regulates RNA polymerase (pol) III activity by competing for binding to the TATA binding protein (TBP) (Tan 2009). A Fus knockout mouse on an inbred background did not

survive the first 24 hours of life and showed an absence of the B cell compartment (Hicks 2000). Mouse embryonic fibroblasts derived from these mice were associated with chromosomal abnormalities and high sensitivity to ionizing radiation. Fus is a target of the DNA damage response kinase, ATM (Gardiner 2008).

Mutations in both Fus and TDP-43 have been associated with the development of familial ALS (Kwiatkowski 2009; Sreedharan 2008). The TDP-43 protein is also associated with other neurodegenerative diseases, including frontotemporal lobar degeneration (FTLD), Alzheimer's Disease (AD) (Galimberti 2010), and chronic traumatic encephalopathy (McKee 2010). In 1993, superoxide dismutase 1 (SOD1) was identified by genome wide association to be the first familial ALS disease gene (Rosen 1993). In 2010, the fourth familial ALS gene, optineurin, was identified (Maruyama 2010). Optineurin has been independently shown to function as both a negative regulator of type I interferon production and as an antiviral factor (Mankouri 2010). Very recently, polyglutamine expansions in the translational regulator ataxin-2 were also shown to be associated with an increased risk for ALS (Elden 2010). Point mutations in the RNA binding domains of both Fus and TDP-43 have been linked to familial ALS pathophysiology. Furthermore, in many cases, Fus and TDP-43 are mislocalized to the cytosol in diseased brain cells. Independent of mutations, a hyperphosphorylated and ubiquitinated form of TDP-43 is the primary protein associated with RNA-rich cytosolic inclusions observed in ALS and FTLD (Buratti and Baralle 2009).

In 2009, Ido Amit from the Hacohen lab published an RNAi screen in which ~150 transcription factors, chromatin modifiers, and RNA binding proteins that are regulated in primary bone marrow dendritic cells in response to pathogen sensing, including Fus, were individually knocked down (Amit 2009). The innate immune gene profile for each knock down condition was then assessed and an innate immune gene network was generated. Fus was found to negatively regulate ~50 genes in the antiviral response pathway, including type I interferons. To our knowledge, this is the only report demonstrating a role for Fus in the innate immune response.

The Manley lab found that Fus regulates RNA pol III gene expression (Tan 2010). Subsequently, I have observed that TDP-43 also negatively regulates both innate immune genes and pol III gene expression, and that cells have an increased influenza viral load when Fus or TDP-43 expression is knocked down. TDP-43 has not been previously reported to play a role in either the innate immune system or in pol III gene regulation.

Results

Fus and TDP-43 are regulated transcriptionally and post-translationally following pathogen sensing.

The mouse lung fibroblast (MLF) cell line is a good model system to study innate immune responses to influenza virus infection. MLFs were either infected with influenza A/PR8 (A/PR8) at an MOI of 5 or transfected with RNA isolated from A/PR8 for various time points. Gene expression was analyzed by quantitative RT-PCR (qPCR). Transcription of Fus and TDP-43 is upregulated in response to viral pathogen sensing (Figure 1a). Fus is also regulated in response to LPS treatment (100 ng/mL) and poly I:C (50 ug/mL) or immunostimulatory DNA (ISD) transfection (120 ng/mL). TDP-43 is regulated only in response to poly I:C and ISD transfection (Figure 1b-c). However, the total protein levels of Fus and TDP-43 do not appear to change over time following pathogen sensing (Figure 1d). Since Fus is a target of the kinase ATM, we decided to test if Fus is more phosphorylated by ATM following pathogen sensing. MLFs were infected with A/PR8 for various time points and immunoprecipitation (IP) of total Fus protein was performed. Both total Fus and ATM-phosphorylated Fus was probed using anti-Fus and anti-phospho-ATM target antibodies, respectively. Indeed, Fus is more highly phosphorylated by ATM following pathogen sensing (Figure 1e). Furthermore, phosphorylated peptide mass spectrometry data from another

member of the Hacohen lab found that TDP-43 is highly phosphorylated following

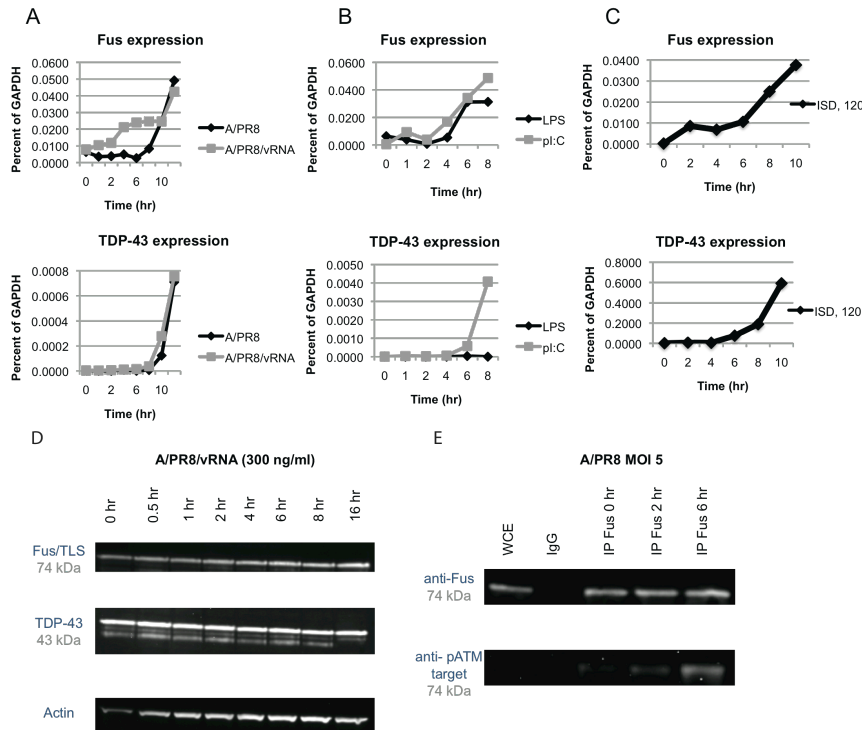innate immune sensing in primary bone marrow dendritic cells (data not shown).

Figure 2.1: Fus and TDP-43 are regulated in response to pathogen sensing. (**A**) Mouse lung fibroblasts (MLF) are infected with influenza A virus/PR8 strain (A/PR8) at an MOI 5 or transfected with RNA isolated from influenza A virus/PR8 (A/PR8/vRNA) over a time course. cDNA libraries were made and gene expression was assessed by quantitative RT-PCR (qPCR). (**B**) MLFs were treated with LPS (100 ng/mL) or transfected with poly I:C (50 ug/mL) over time and gene expression was assessed by qPCR . (**C**) MLFs were transfected with immunostimulatory DNA (ISD, 120 ng/mL) over time and gene expression was assessed by qPCR . (**D**) MLFs were transfected with A/PR8/vRNA (300 ng/mL) over a time course. The levels of Fus and TDP-43 protein was determined from these cell lysates by Western blotting. (**E**) MLFs were infected with A/PR8 at an MOI 5 for 0, 2, and 6 hours. An immunoprecipitation of an isotype control (Rabbit IgG) or Fus was performed. Both the total Fus protein levels and levels of Fus that were phosphorylated by ATM was determined by Western blotting and compared to whole cell extract (WCE).

Fus and TDP-43 are negative regulators of antiviral gene expression

Knockdown

To test how innate immune gene expression changes when Fus is depleted, Fus

was knocked down in MLF using siRNA. Then, cells were infected with A/PR8 (MOI 5) for 18 hours and gene expression was analyzed by qPCR. Fus knockdown cells expressed higher levels of antiviral genes, monitored here by IFNβ and CXCL10 expression (Figure 2a). (See Chapter 1 for details on TDP-43 perturbation). To test whether more type I interferon protein is produced in these cells, Fus and TDP-43 were knocked down in MLF using shRNAs, then infected with either A/PR8 or influenza A/deltaNS1 (A/dNS1) at various MOIs for 24 hours. The supernatants from this experiment were tested on a type I interferon-sensitive ISRE-luciferase reporter cell line. Fus and TDP-43 knockdown cells had higher ISRE activity than controls (Figure 2b).

Overexpression

Overexpression of Fus has been challenging since the protein is very toxic to cells if expressed at levels higher than baseline. However, transient transfection of HEK293T cells with plasmids encoding either wild type human Fus or ALS-linked Fus disease alleles has been successful. HEK293T cells were transfected with 0.5 ug of plasmid DNA/10^5 cells. Twenty-four hours post transfection, cells were infected with A/PR8 for 8 hour and gene expression was analyzed by qPCR. Overexpression of wild type Fus results in more IFNβ gene expression and the overexpression of ALS-linked Fus disease alleles resulted in significantly higher IFNβ expression (Figure 2c). Like TDP-43, overexpressing Fus results in a dominant negative phenotype that matches the knock down phenotype. This

106

effect is not completely understood for either protein. Perhaps Fus autoregulates its own expression via a 3'UTR destabilizing mechanism like TDP-43 does, or perhaps the overexpressed protein is heavily modified rendering it non-functional as TDP-43 appears to be. These possibilities remain to be explored.
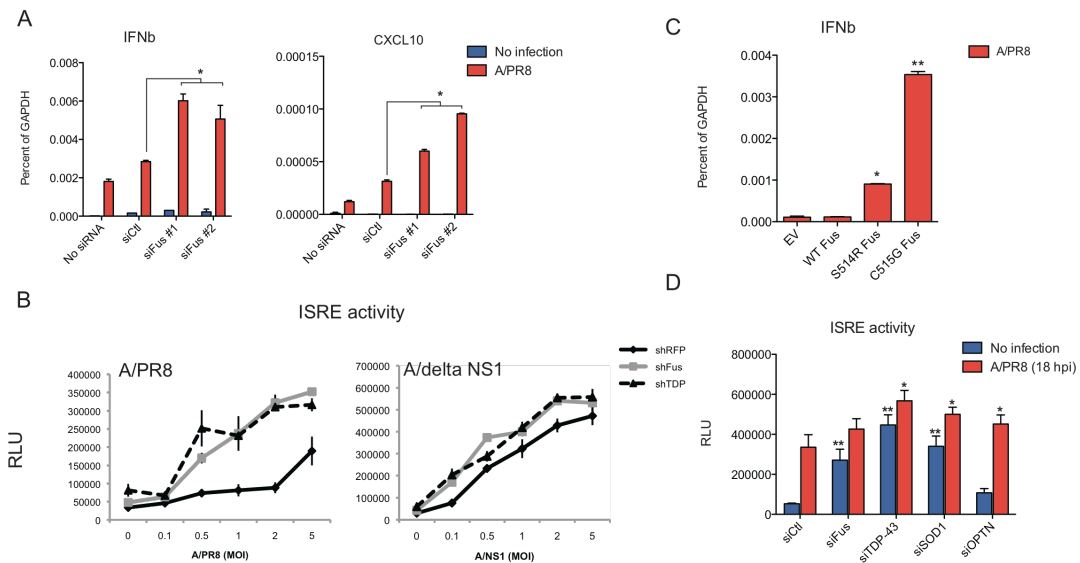

Figure 2.2: Fus and TDP-43 negatively regulate the expression of antiviral genes. (**A**) Mouse lung fibroblasts (MLFs) were transfected with Lipofectamine RNAiMax transfection reagent only (no siRNA), an siRNA targeting an irrelevant transcript (siCtl), or Fus (siFus#1 & 2). After 72 hours, the cells were infected with A/PR8 (MOI 5) for 18 hours. Then, IFNβ and CXCL10 mRNA levels were measured by qPCR. *p<0.05, **p<0.01. Data are representative of three experiments. (**B**) MLFs were infected with puromycin-resistant shRNA lentiviral vectors targeting RFP, Fus, or TDP-43. After 5 days of puromycin selection, cell were infected with A/PR8 (left) or A/NS1 (right) at MOI 0, 0.1, 0.5, 1, 2, or 5 for 24 hours. The supernatants were then tested on an ISRE-luciferase reporter cell line to measure relative type I interferon production. Data are representative of three experiments. (**C**) HEK293T cells were transfected with plasmid DNA (0.5 ug/10^5 cells) encoding the empty pmCherry vector backbone only (EV), WT Fus, or Fus with ALS-associated point mutations (S514R and C515G) for 24 hours. Cells were then infected with A/NS1 at an MOI of 5 for 8 hours. Gene expression was assessed by qPCR. Data are representative of five experiments. (**D**) Mouse embryonic fibroblasts (MEF) stably expressing an ISRE promoter driving a luciferase reporter (ISRE-luc) were transfected with siRNA targeting Fus (siFus), TDP-43 (siTDP-43), SOD1 (siSOD1), and optineurin (siOPTN). Cells were infected with A/PR8 (MOI 5) for 18 hrs and luminescence was read. Data are representative of three experiments.

SOD1 is a negative regulator of antiviral gene expression

There is an increasing understanding that innate immune pathways are involved

in neurodegeneration (See Section III of the Introduction). We noted that several of the ALS disease genes, namely TDP-43, Fus, and optineurin, were negative regulators of type I interferons. Additionally, a transgenic SOD1 ALS mouse line was found to have greater interferon-stimulated gene expression in its central nervous system (Wang 2010). Therefore, we wanted to test if knock down of SOD1 would result in an increase in interferon activity. We knocked down Fus, TDP-43, optineurin, and SOD1 proteins using siRNAs from mouse embryonic fibroblasts that stably express an interferon-sensitive ISRE-luciferase reporter construct (ISRE-luc). We then tested ISRE-luc activity in these cells in response to A/PR8 infection. Surprisingly, we found that the depletion of SOD1 resulted in both a baseline and viral sensing-induced increase in type I interferon, similar to the perturbation of Fus and TDP-43 (Figure 2d). We are not aware of any reports showing that SOD1 is a direct regulator of innate immune signaling but more work in this area might be informative to the field.

Fus and TDP-43 are negative regulators of viral replication

Since the absence of Fus and TDP-43 is associated with greater type I interferon activity, we postulated that Fus and TDP-43 knock down cells would control viral replication better than control cells. Surprisingly, there is more A/PR8 viral replication when Fus and TDP-43 are depleted (Figure 3a). There is also more vesicular stomatitis virus (VSV) when Fus and TDP43 are depleted and slightly more reovirus replication when Fus and TDP-43 are depleted (Figure 3b-c).

The antiviral effect of Fus and TDP-43 seems to be somewhat independent of interferon activity, as the knock down of Fus using siRNAs targeting the human transcript in a cell that does not produce interferons, the African Green Monkey cell line Vero, was also associated with more A/PR8 replication (the human siRNAs targeting TDP-43 in this cell type did not successfully knock down the TDP-43 transcript) (Figure 3d).

The overexpression of wild type or disease forms of Fus and TDP-43 was not able to rescue the viral phenotype. Overexpression of these genes results in an increase of A/PR8 virus replication (Figure 3e-f). We were not surprised that the knockdown and overexpression phenotypes were consistent with each other since we found a dominant negative phenotype for type I interferon and ISG expression.
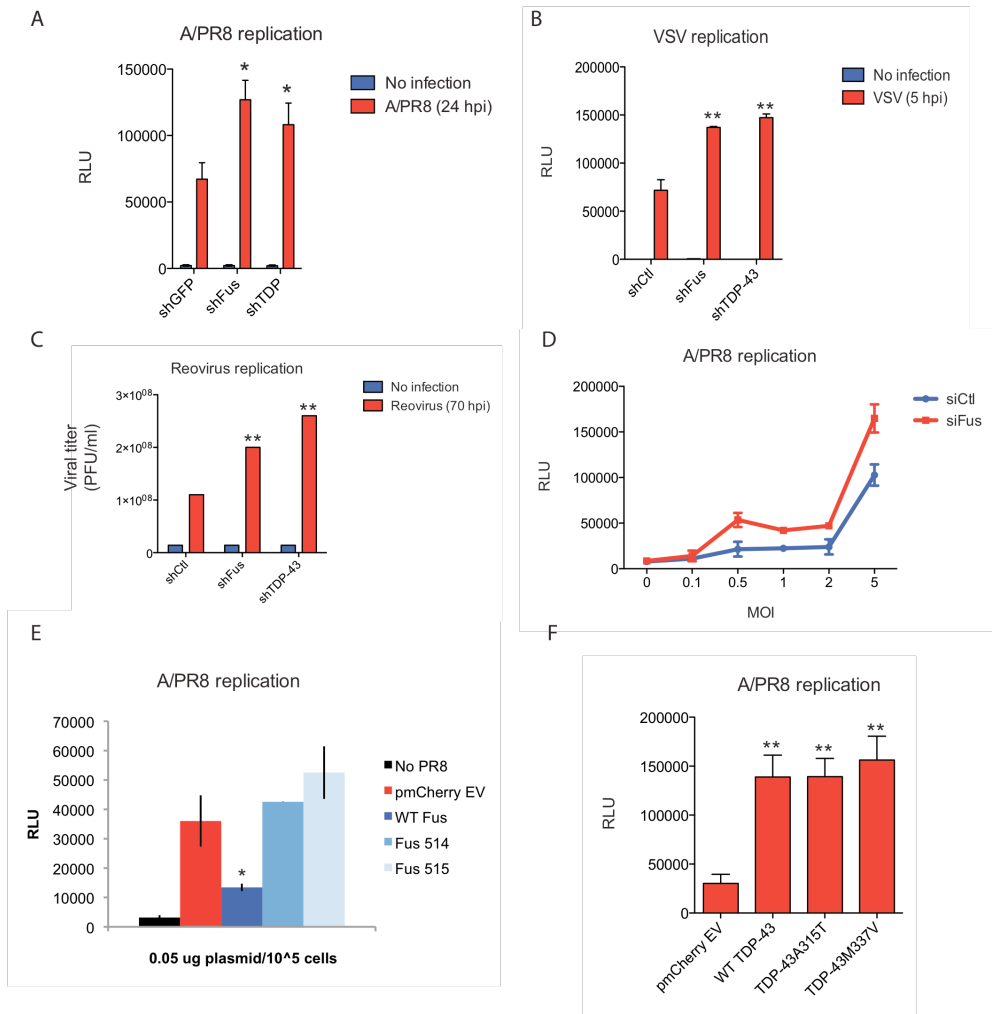
Figure 2.3: Fus and TDP-43 have antiviral effects. (**A**) MLFs were infected with puromycin-resistant shRNA lentiviral vectors targeting GFP, Fus, or TDP-43. After 5 days of puromycin selection, cells were infected with A/PR8 (MOI 5) for 24 hours. The supernatants were then tested on in an influenza A/PR8 vRNA-luciferase (vRNA-luc) reporter assay to quantify the presence of replication-competent viral particles. *p<0.05, **p<0.01. Data are representative of five experiments. (**B**) MLFs were infected with shRNA lentiviral vectors targeting the indicated genes. After selection, cells were infected with vesicular stomatitis virus (VSV) expressing a luciferase reporter (MOI 10). Five hours post-infection, total luciferase activity was measured to quantify VSV replication. Data are representative of three experiments. (**C**) MLFs were infected with shRNA lentiviral vectors targeting the indicated genes. After selection, cells were infected with reovirus (MOI 3). Seventy hours post infection, the supernatants were applied to reovirus plaque assay to measure total reovirus replication. Data are representative of two experiments. (**D**) African Green Monkey Vero cells (Vero) were transfected with human siRNA constructs targeting either a negative control (Mission negative, MN) or Fus (siFus). After 72 hours, the cells were infected with A/PR8 (MOI of 0, 0.1, 0.5, 1, 2, or 5) for 24 hours. Influenza viral replication was then measured using the vRNA-luc reporter assay. Data are representative of two experiments. (**E**-**F**) HEK293T cells were transfected with plasmid DNA (0.5 ug/10^5 cells) encoding the empty pmCherry vector backbone only (EV) or (**E**) Fus contructs: WT Fus, Fus 514, Fus 515; (**F**) WT TDP-43, TDP-43 A315T, or TDP-43 M337V for 24 hours. Cells were then infected with A/PR8 (MOI 5) for 24 hours. Influenza viral replication was then measured from the supernatants using the vRNA-luc reporter assay. Data are representative of three experiments.

110

At this point, we are not sure how Fus and TDP-43 are functioning as antivirals. In the next section, we present data regarding the regulation of biosynthesis genes by both Fus and TDP-43. One possibility is that the roles of Fus and TDP-43 in regulation of RNA transcripts contribute to viral replication.

TDP-43 is a negative regulators of pol I and III transcription

Tan and Manley recently demonstrated that Fus negatively regulates RNA pol III gene expression by associating with the TATA binding protein (TBP) and inhibiting its interaction with the RNA polymerase (Tan et al., 2010). The tumor suppressor gene, p53, likewise inhibits biosynthesis by controlling pol III activity using the same mechanism (Stein et al., 2002). Pol III-transcribed genes are not translated and encode non-coding RNAs that are critical for splicing, translation, and other aspects of mRNA processing. We wondered if TDP-43 likewise regulates pol III-transcribed genes. Pol III-transcribed genes utilize four different promoter configurations. To be complete, we monitored how perturbation of TDP-43 affects the expression of genes using each of these promoter systems. In the absence of a stimulus, this pol III inhibition phenotype is subtle. However, following pathogen sensing, the absence of TDP-43 results in a tremendous upregulation of pol III gene expression (Figure 4a). We also tested whether perturbation of TDP-43 affects transcription from polymerase I (pol-I) -transcribed genes and found that depleting TDP-43 results in a baseline and viral infection-induced increase in 18S expression (Figure 4b). The depletion of Fus, however,

does not appear to affect pol I-transcribed gene expression.

We wondered whether this increased biosynthesis might be responsible for the increase in viral replication when Fus and TDP-43 are knocked down. To test this, we titrated a specific RNA pol III chemical inhibitor into MLFs that received siRNA targeting Fus and TDP-43. Interestingly, by inhibiting pol III gene activity, influenza replication in Fus and TDP-43 knockdown cells was brought down to control levels (Figure 4c).

Some work was done to test if TDP-43 regulates pol III transcription by the same mechanism that Fus does. However, I was not able to find a direct binding association between TDP-43 and TBP, the pol III coactivator that Fus inhibits (data not shown). However, I did note that TDP-43 was directly associating with pol III-transcribed RNA transcripts. Using a modified low-RNase CLIP protocol, I pulled down RNAs that are specifically associated with Fus and TDP-43 in resting cells and in cells that were transfected with A/PR8/vRNA to activate the antiviral pathway, and then identified the RNA binding partners by qPCR. From these experiments, both genes were enriched in pol III transcript binding over controls, but that TDP-43 was much more enriched for binding to key pol III-transcribed genes, such as U6 and 7SL (Figure 4d). We see that TDP-43 specifically binds to many non-coding RNAs from CLIP experiments under more stringent conditions with an Illumina sequencing read out as well. Last week, a

112

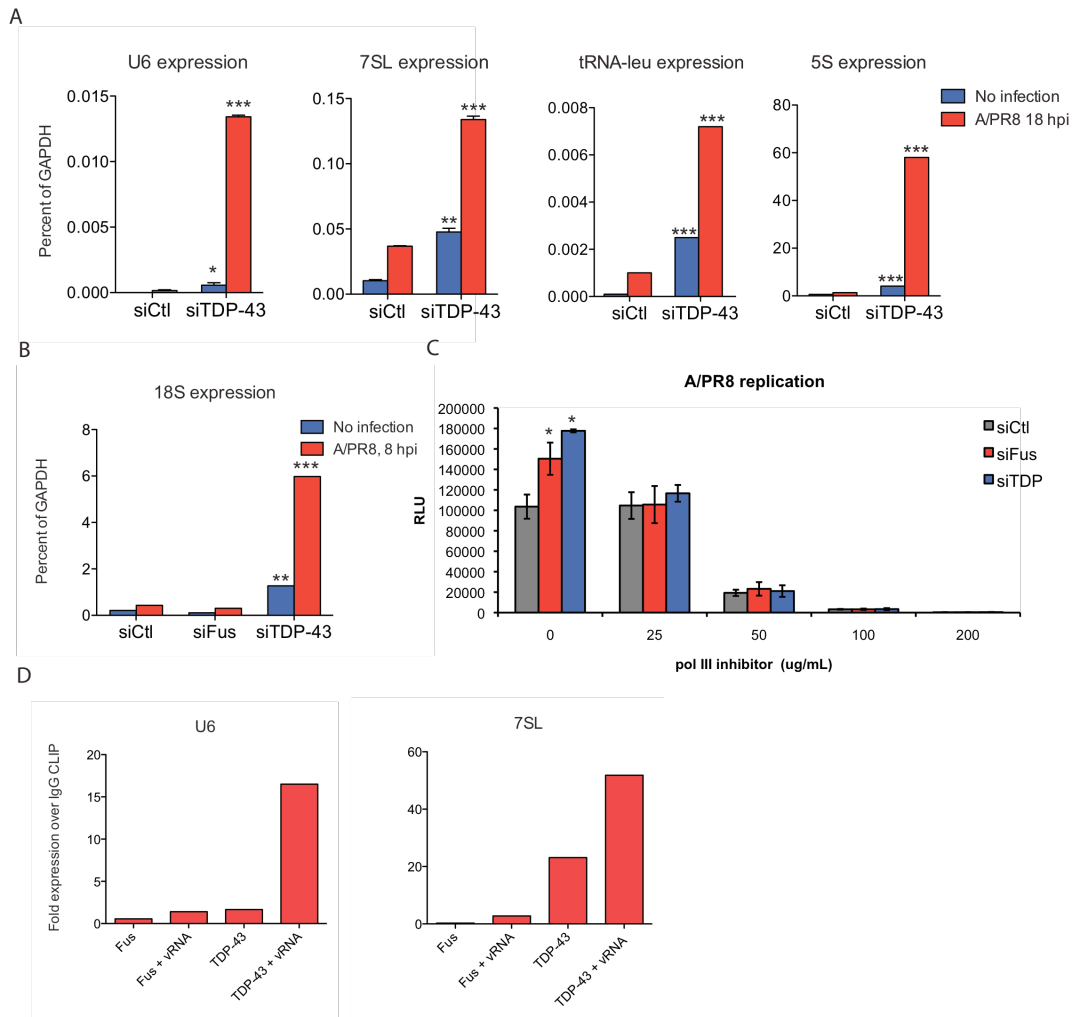paper from the Cleveland lab reported that TDP-43 binds to these transcripts as well (Lagier-Tourenne 2012).

Figure 2.4: TDP-43 regulates RNA polymerase I- and III-transcribed genes. (A-B) MLFs were

Figure 2.4 (Continued):
transfected with siRNA targeting TDP-43. After 72 hours, the cells were infected with A/PR8 (MOI 5) for 18 hours and gene expression was assessed by qPCR. *p<0.05, **p<0.01, ***p<0.001. Data are representative of five experiments. (**C**) MLFs were transfected with siRNA constructs targeting Fus and TDP-43. After 72 hours, the cells were infected with A/PR8 (MOI 5) for 24 hours in the presence of increasing concentrations of an RNA polymerase III inhibitor. Data are representative of two experiments. (**D**) Resting MLFs or MLFs that were transfected with vRNA/A/PR8 for 18 hours were UV-crosslinked (265 nm) to immobilize RNA:protein interactions in vivo. CLIP under low RNAse conditions was performed using control antibodies and antibodies targeting Fus and TDP-43. Relative RNA abundance was assessed by qPCR. Values are expressed relative to IgG CLIP RNA enrichment. Data are representative of five experiments.

We are not sure how TDP-43 functions mechanistically as an antiviral factor. One possibility is that TDP-43 regulates the stability of biosynthesis RNA transcripts and in its absence these transcripts accumulate. This would result in greater translation of proteins in the cell globally, which has recently been reported (Fiesel 2012). We also noted that total RNA and protein content per cell is ~10% higher in TDP-43-depleted cells. Typically, viral infection is associated with an almost global reduction in protein translation through activation of PKR. In the absence of TDP-43 this balance might be disrupted and thus provide a more favorable environment for viral replication. I think that this is one of the more interesting aspects of this unfinished thesis work and merits further investigation.

Connections between ALS and antiviral gene expression

Amyotrophic lateral sclerosis (ALS), or Lou Gehrig's disease, is a progressive, fatal neurodegenerative disease that is characterized by upper and lower motor deficits (See Introduction). There are many innate immune features of neurodegeneration, although the antiviral arm of the pathway has not been deeply explored. We see distinct innate immune phenotypes associated with Fus

and TDP-43, and are interested in testing if 1) ALS patient brains have an interferon phenotype, 2) if type I interferons play a role in motor neuron cell death.

Primary cells from ALS patients have greater antiviral gene expression

Primary skin fibroblasts were derived from two living familial ALS patients. One patient has a point mutation in Fus (R521C) and the other has a point mutation in TDP-43 (M337V). Control and patient fibroblasts were infected with A/PR8 (MOI 5) for 18 hours and gene expression was analyzed by qPCR. The ALS patient cells with a TDP-43 point mutation had higher interferon gene expression in response to viral infection than the Fus ALS patient or control cells (Figure 5a).

RNA was generated from fixed and frozen frontal cortex brains samples from deceased familial ALS patients and control patients. Complementary DNA libraries were constructed for each RNA sample using random nonamers and gene expression was analyzed by qPCR analysis. Several of the ALS patient brain samples highly expressed antiviral genes, including IFNa, IFNβ, and IFIT1, as well as pol III-transcribed genes, including U6 and 7SL (Figure 5b). However, not all innate immune genes are upregulated in ALS brains as mRNA levels for the pro-inflammatory gene IL-6 do not vary significantly between patients and controls.
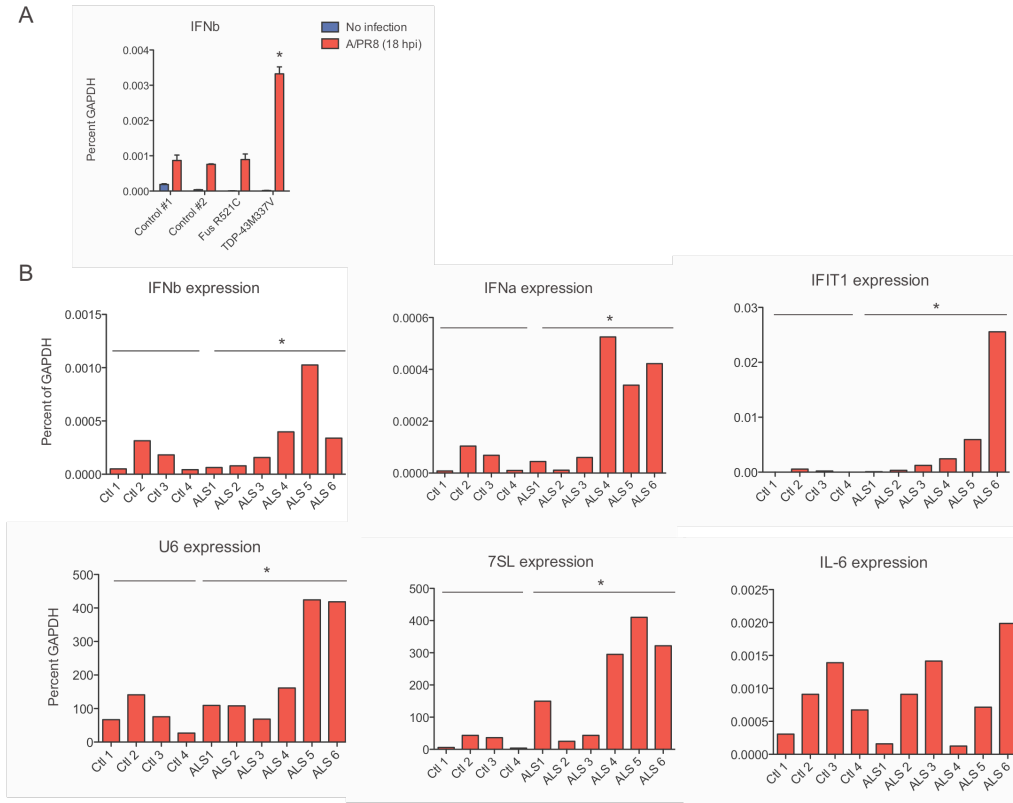
Figure 2.5: ALS brains express higher levels of antiviral- and biosynthesis- related genes. (**A**) Primary human skin fibroblasts (HFs) were derived from normal aged individuals (Control #1 and #2) and two ALS patients, one with a point mutation in TDP-43 (M337V) one with a point mutation in Fus (R521C). HFs were infected with A/PR8 (MOI 5) for 18 hours and gene expression was assessed by qPCR. (**B**) Purified RNA from frontal cortex samples of deceased control cadavers (Ctr l-4) and ALS patient cadavers (ALS 1-6) was obtained from the Massachusetts General Hospital Brain Bank. Randomly primed cDNA libraries were generated for each sample and gene expression was assessed by qPCR. *p<0.05. All experiments were performed once due to limited samples.

ALS dog spinal cords have an antiviral gene expression phenotype

As in human ALS, a missense mutation in the canine SOD1 protein (E40K) causes an upper and lower motorneuron disease called canine degenerative myelopathy (DM), which pathologically mimics human ALS (Awano 2009). We wondered whether the spinal cords of these animals also show an ISG phenotype. DM severity is graded according to disease progression (Figure 6a).

116

We profiled innate immune gene expression in the spinal cords of eight DM dogs across the range of disease severity. Interestingly, IFNβ, CXCL10, and IFIT1 expression is upregulated in the spinal cords of dogs with mild to intermediate disease, while MX1 and IL-6 expression is higher in dogs with late-stage disease compared to controls (Figure 6b). The observation that innate immune gene expression varies significantly across disease progression in a dog model of ALS suggests that gene profiles might change over time in human ALS patients as well. Not only is this important for understanding disease pathophysiology but might also have significant implications for biomarker development.

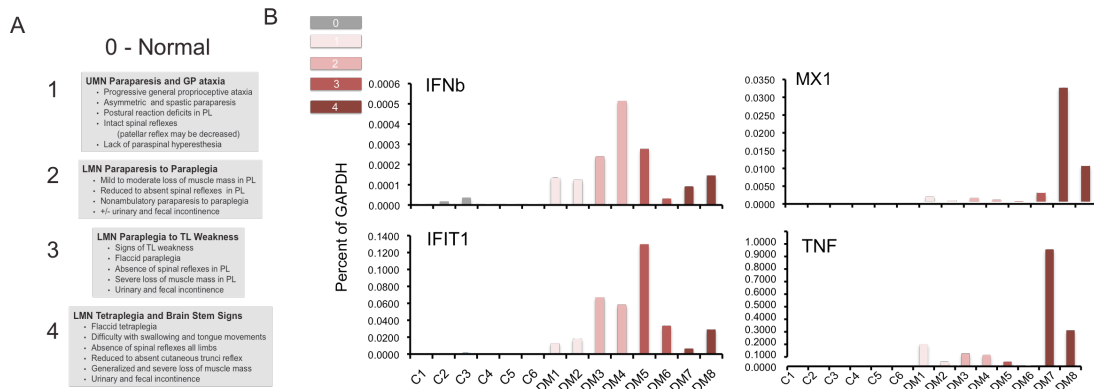**Figure 2.6: ALS dog spinal cords express higher levels of antiviral genes. (A)** Canine ALS, or degenerative myelopathy (DM), disease severity is graded on a scale from 0 (Normal) to 4. **(B)** RNA was isolated from frozen spinal cord samples from dogs that were euthanized at various stages of DM progression (DM1-8) and from age/sex matched canine controls (C1-6). Gene expression was assessed by qPCR.

Chapter 2 Summary:

There are two major findings in this body of unfinished work: 1) Fus and TDP-43 are somehow functioning as broad antiviral proteins and 2) there is a clear antiviral gene expression phenotype in primary tissues from ALS patient brains. I decided to spend the majority of my thesis effort studying the biochemical mechanism of function of TDP-43 because that was the work that I found to be more technically challenging and conceptually interesting. However, the observations presented in Chapter 2 have potentially important clinical implications. In particular, there is a potential to develop a therapeutic that can inhibit type I interferons in the CNS of ALS individuals and could delay disease progression.

Intrinsic defects in innate immune responses in ALS patients should not be restricted to the brain and an evaluation of systemic defects might shed light onto the role of intrinsic innate immune defects in disease progression and potential treatment options. One on-going effort in the lab stemming from the work presented here is to interrogate the innate immune responses of ALS patient blood cells to bacterial and viral pathogens. Over the past two years, we have worked with Drs. Merit Cudkowicz and James Berry at the MGH ALS clinic to collect peripheral blood from ALS individuals. Presently, we have banked peripheral blood from ~25 Caucasian non-Hispanic ALS individuals equally representing both sexes (average age: 60 yo). Three of the patients have

confirmed SOD1 mutations and the remaining 22 are either sporadic cases or have not been genotyped. We also worked with Dr. Phil DeJager at the Brigham and Women's Hospital to collect blood from a total of 50 healthy, aged donors (average age: 60 yo). Although others have profiled resting ALS PBMC on microarrays, our proposed study will monitor dynamic innate immune gene expression changes over time in response to pathogens. Furthermore, our lab is generating a large data set in which the immune responses of PBMC from 600 healthy individuals are being profiled to generate a range of "normal" responses to pathogens to compare to disease cells.

In our proposed study, blood cells will be 1) cultured *in vitro* for a week to return cells to a resting state and "erase" prior immune memory, 2) exposed to defined components of several viral and bacterial pathogens, 3) profiled for hundreds of genes that are important for the immune response using the Nanostring high-throughput gene expression profiling method. As a final analysis, immune phenotypes will be correlated with clinical information regarding disease progression and we will determine if there is phenotypic correspondence of intrinsic immune defects among instances of deliberate ALS disease allele expression, naturally occurring disease alleles, and sporadic ALS cases.

Expected outcomes:

We expect to see a correlation in innate immune phenotypes between deliberate and naturally occurring disease allele expression in ALS tissues. Furthermore, we

predict that the extent of immune activation will be correlated with disease progression. If our hypotheses are correct, then finding the mechanism that controls innate immune responses in ALS, and potentially a direct treatment to interrupt this pathway, would be an appropriate experimental follow-up in mouse models.

Experimental Procedures

Cell lines

HEK293T cells, U2OS cells, and mouse lung fibroblasts were cultured in tissue culture-treated plastic dishes in DMEM media (Gibco, Carlsbad, CA) supplemented with 10% FBS and 1% penicillin/streptomycin. The cells were grown in an incubator at 37°C.

Viruses

Sendai virus, Cantell and 52 strains (ATCC, Manassas, VA), were used for infections. Influenza A virus strain A/PR/8/34 was grown in Vero cells, and virus titers from cell supernatants were quantified using 293T cells transfected with a vRNA luciferase reporter plasmid. Vesicular stomatitis virus (VSV) expressing luciferase was a gift from S. Whelan. Reovirus and viral titering was performed by the Nibert lab at Harvard Medical School.

Virus infection

Influenza A, Sendai virus, VSV, and reovirus was added directly to cell culture media for the stated time points and viral replication was measured according to virus-appropriate assays.

Antibodies and reagents

Anti-TDP-43 (ProteinTech Group, Chicago, IL), Fus (AbCam, Cambridge, MA), pIRF3 (Cell Signaling, Danvers, MA), pSTAT1 (AbCam), pATM target (Cell Signaling), and b-actin (AbCam) antibodies were used for immunoprecipitations and Western blotting.

TLR ligands were from Invivogen (ultra-pure E. coli K12 LPS) and Enzo Life Sciences (poly(I:C)).

RNA polymerase III inhibitor (EMD Millipore, Billerica, MA).

pmCherry plasmids encoding WT TDP-43, TDP-43A315T, and TDP-43M337V were a gift from R. Baloh.

Influenza virus titering

$2x10^6$ 293T cells (filtered through a 0.4mm filter) were seeded in 10cm dishes andtransfected with a vRNA luciferase reporter plasmid based on prior design (Lutz et al.,2005a) using Transit-LT1. Cells were trypsinized at 24h post-transfection and 10^4transfected reporter cells were re-seeded in white Costar plates. Supernatants (frozen with trypsin) from PR8-infected HBECs were added to reporter cells and incubated for 24 hours. Reporter activity was measured with firefly luciferase substrate (Steady-Glo,Promega, Madison, WI). Luminescence in 96-well plates was quantified using the Envision Multilabel reader (Perkin Elmer, Waltham, MA) fitted with an automated plate stacker.

RNA isolation

mRNA was isolated using the RNeasy mini kit according to the manufacturer's instructions (Qiagen, Valencia, CA). RNA was reverse transcribed with the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA).

Total RNA was isolated using phenol:chloroform:isoamylalchohol (Invitrogen, Carlsbad, CA). RNA was precipitated overnight at -80°C from the aqueous phase by adding an equal volume of 100% isopropanol and 1 ul of glycogen (Thermo Scientific, Rockland, Il). The RNA was pelleted by centrifuging for 45 minutes at

4°C and then washed with 80% ethanol to retain small RNA species. RNA sample quality was assessed on a 2100 Bioanalyzer (Agilent, Palo Alto, CA).

For experiments with very low RNA yield, RNA was reverse transcribed using the Sensiscript RT kit (Qiagen).

qPCR measurements

Quantitative real time PCR (qPCR) reactions were performed on the LightCycler 480 system (Roche, Indianapolis, IN) using the FastStart Universal SYBR Green Master Mix (Roche). Every reaction was run in duplicate and endogenous GAPDH levels were used for normalization.

For bar graphs, the unpaired 2-tailed student's T-test was used to compute p values, except where stated otherwise. Error bars reflect the standard error of the mean (SEM), except where stated otherwise. All statistical analyses were carried out using GraphPad Prism 4.0 and the R statistical environment.

shRNA infection

High-titer lentiviruses expressing shRNAs were obtained from The Broad RNAi Platform and used to infect cells as previously described (Moffatt et al., 2006). Primary BMDC were infected according to previously published protocols (Amit 2009). For all other cells, 5000 cells were plated on flat bottom 96-well plates (Corning) and 24 hrs later spin-infected with shRNA encoding lentiviruses (MOI 2-10) supplemented with polybrene (8 ug/ml). 24 hours after infection, the cells were selected using puromycin (2 ug/ml) in culture medium. Cells were used in experiments after 4 days of selection.

Plasmid transfection

Plasmid DNA was transfected using TransIT LT1 reagent (Mirus, Madison, WI) according to manufacturer's protocols.

siRNA transfection

siRNAs targeting a negative control sequence or TDP-43 (Dharmacon siGenome Smartpool) were transfected into cells using Lipofectamine RNAiMax (Invitrogen) according to manufacturer's instructions.

Luminescent reporter assays

Plasmid DNA encoding luminescent reporter proteins and ORFs or 3' UTRs of interest were transfected into cells using TransIT LT1 (Mirus). The Dual Glo Luciferase Assay System (Promega, Madison, WI) was used to read luminescence using the Envision Multilabel Reader (Perkin Elmer, Waltham, MA).

CHAPTER 3: Msi2 maintains self-renewal through TGFb signaling in hematopoietic stem cells

One of the primary goals for my PhD work was to develop key RNA biochemistry techniques with which to study RNA binding proteins, identify their interaction partners, and determine their mechanisms of function more broadly. In addition to my thesis project, I have been able to apply some of the methods that I developed at the Broad to collaborations with other investigators. In particular, I have worked on projects investigating the RNA targets of the RNA binding proteins Musashi 1 (Jaenisch lab, Whitehead Institute), Musashi 2 (Kharas lab, Memorial Sloan-Kettering), and Lin28 (Daley lab, Harvard Medical School). These projects are in various stages of manuscript preparation. Chapter 3 of this thesis presents work on Musashi 2 that was recently accepted to the journal *Cell Stem Cell*.

Msi2 maintains self-renewal through modulation of TGF-β signaling in hematopoietic stem cells

Sun-Mi Park[1], Raquel P. Deering[2], Yuheng Lu[3], Patrick Tivnan[1], Steve Lianoglou[3] Fatima Al-Shahrour[4], Benjamin L. Ebert[2,5], Christina Leslie[3], George Q. Daley [5,7], Christopher J. Lengner [6] and Michael G Kharas[1]*

[1]Molecular Pharmacology and Chemistry Program and Center for Cell Engineering, Memorial Sloan-Kettering Cancer Center, New York, NY; [2]Harvard Medical School, Boston, MA; [3]Computational Biology Program Memorial Sloan-Kettering Cancer Center, New York NY; [4]Translational Bioinformatics Unit, Clinical Research Programme, Spanish National Cancer Research Centre, Madrid, Spain [5]Brigham & Women's Hospital, Boston, MA, Division of Hematology, Karp, Brigham and Women's Hospital, Boston, MA; [6]Department of Animal Biology, Department of Cell and Developmental Biology and Institute for Regenerative Medicine, Schools of Veterinary Medicine and Medicine, University of Pennsylvania, Philadelphia, PA; [7]Hematology Stem Cell Program and Division of Hematology/Oncology, Children's Hospital Boston, Howard Hughes Medical Institute, Boston, MA.

*Correspondence: Michael G Kharas (M.G.K.), Kharasm@mskcc.org

ABSTRACT

Hematopoietic stem cells are maintained through the regulation of symmetric and asymmetric divisions. We report that conditional ablation of the Msi2 RNA binding protein results in reduced HSC numbers due to a loss of quiescence and increased commitment divisions. Surprisingly, this was not mediated through a general increase in translation of the Msi2 binding target Numb. Instead, global transcriptome profiling and RNA target analysis identified direct MSI2 binding targets that govern HSC self-renewal, cell cycle and differentiation. Additionally, we revealed complex control of the TGF-β pathway by MSI2 in hematopoietic cells. Correspondingly, we found HSCs without Msi2 have reduced TGF-β signaling and are insensitive to TGF-β-mediated expansion. Msi2 deficiency reduced the frequency of myeloid-biased HSCs and transplants exhibited diminished myeloid reconstitution. Taken together, these results provide insight into the role of Msi2 in hematopoiesis and uncover a link between hematopoietic stem cell self-renewal and Msi2-mediated control of developmental pathways.

INTRODUCTION

Hematopoiesis is a tightly regulated process in which different cell lineages are generated from hematopoietic stem cells (HSC). HSCs can self-renew and give rise to more differentiated progenitor cells through symmetric and asymmetric cell division, respectively(Beckmann et al., 2007; Wu et al., 2007). Progenitors can further differentiate into more committed cell subsets that generate the mature lymphoid and myeloid compartments. In order to support a normal hematopoietic system, HSCs must maintain a delicate balance between self-renewal and differentiation. However, little is known about the developmentally conserved programs that regulate this balance.

The Musashi (Msi) family of RNA binding proteins, including Msi1 and Msi2, contribute to the control of symmetric and asymmetric stem cell division, regulate stem cell function, and play a role in cell fate determination (Okano et al., 2005). In *Drosophila,* Musashi was first identified to be important for sensory neuron development, where it is essential for asymmetric division of neuroblasts into daughter cells with neuronal and non-neuronal potential(Nakamura et al., 1994). Loss of Musashi function results in defective asymmetric division and the formation of two non-neuronal daughter cells at the expense of commitment to the neuronal lineage(Nakamura et al., 1994). In vertebrates, Msi1 and Msi2 are evolutionarily conserved, harboring two tandem RNA recognition motifs (RRMs) and a carboxyl-terminal poly-A binding protein association domain(Kawahara et al., 2008; Okano et al., 2005).

Msi proteins are thought to function by binding to the 3' UTRs of target mRNAs at a consensus sequence and then blocking translation by hindering access of the poly-A binding protein to the elongation initiation complex(Kawahara et al., 2008). An inverse relationship between Msi and Numb expression as a result of Msi-mediated repression of Numb has been demonstrated in a variety of systems, including myeloid leukemia cells(Imai et al., 2001; Ito et al., 2010; Kharas et al., 2010). We and others recently identified Msi2 as an important modulator of proliferation and differentiation in both normal hematopoietic stem cells and in myeloid malignancies (de Andrés-Aguayo et al., 2011; Hope et al., 2010; Ito et al., 2010; Kharas et al., 2010). Analysis of the hematopoietic compartment of *Msi2* gene trap mice revealed a reduced number of short term HSC (ST-HSC) and lymphoid primed myeloid progenitor (LMPP) cells, but no significant defect was found in long term HSCs(de Andrés-Aguayo et al., 2011). Although *Msi2* is most highly expressed in HSCs, and *MSI2* overexpression drives quiescent HSCs out of G_0 and into cycle (Kharas et al., 2010), it remains unclear if and how Msi2 affects HSC self-renewal and commitment under homeostatic conditions. Furthermore, the critical RNA binding targets of Msi2 in hematopoietic cells that regulate self-renewal and lineage commitment remain to be uncovered.

To determine the role of Msi2 in HSCs and to avoid potentially confounding compensatory mechanisms by germline *Msi2* loss, we generated *Msi2* conditional knockout mice that allow us to study Msi2 function in a cell

autonomous manner in adult tissues using spatiotemporally controlled deletion. Here, analysis of microarray data of *Msi2* conditional knockout mice coupled with a MSI2 HITS-CLIP (CrossLinking and ImmunoPrecipitation followed by HIgh Throughput Sequencing) profiling data allowed us to identify novel regulatory pathways downstream of Msi2 in hematopoietic stem cells.

RESULTS

In order to assess the role of Msi2 in the hematopoietic compartment, we developed a conditional knockout mouse model to study the role of Msi2 in adult hematopoiesis. We targeted the *Msi2* locus in embryonic stem cells with a construct containing loxP sites flanking the first four exons (Figure 1A and S1). After deletion of the neomycin resistance selection cassette, an *Msi2^{flox/flox}* mouse colony was established and crossed with Mx1-Cre mice to generate an inducible Msi2 loss of function strain (*Msi2^{flox/flox}::Mx1-Cre*). To delete the Msi2 gene from HSCs, we induced the Cre transgene in *Msi2^{flox/flox}::Mx1-Cre* mice by three polyI:polyC (pIpC) injections, which efficiently excised the Msi2 gene from the bone marrow and spleen assessed by southern blot and by qRT-PCR analysis within the LSK (Lineage^{lo} c-kit^{+}, Sca^{+}) compartment (Figure 1B-C). Msi2 deleted mice will be referred to as *Msi2^{Δ/Δ}* and control mice either *Msi2^{f/f} or Msi2^{Δ/wt}*

Msi2^{Δ/Δ} mice had normal peripheral blood counts after 6 and 18 weeks post-pIpC injections (Figure S1). However, after 12-18 weeks the mice had reduced spleen

weights and cellularity (Figure 1D and S1). We previously observed alterations in

myeloid differentiation upon *MSI2* overexpression *in vivo* (Kharas et al., 2010). In

contrast, we found equivalent frequencies of immature myeloid cell types in the

bone marrow and in the spleen (Figure S1). Also, we found the frequencies of B-

cells or T-cells in the bone marrow and spleen to be normal compared to controls

(Figure S1). Furthermore, we tested the progenitor activity using a colony assay

with $Msi2^{\Delta/\Delta}$bone marrow cells (Figure 1E). We found a significant decrease in

myeloid colony formation with reductions in multilineage progenitors and serial

replating efficiency (Figure 1E). The decreased cellularity in the spleen and

decreased colony forming ability indicated a defect in either early stem or

progenitor function. Thus, we examined the overall frequency and numbers of

HSCs in these mice. We found reduced frequency and absolute number of LSKs

and reduced overall numbers of HSCs after 12wks (Figure 1F-G and S1).

Decreases in LSK numbers were seen at as early as 4 weeks after deletion of
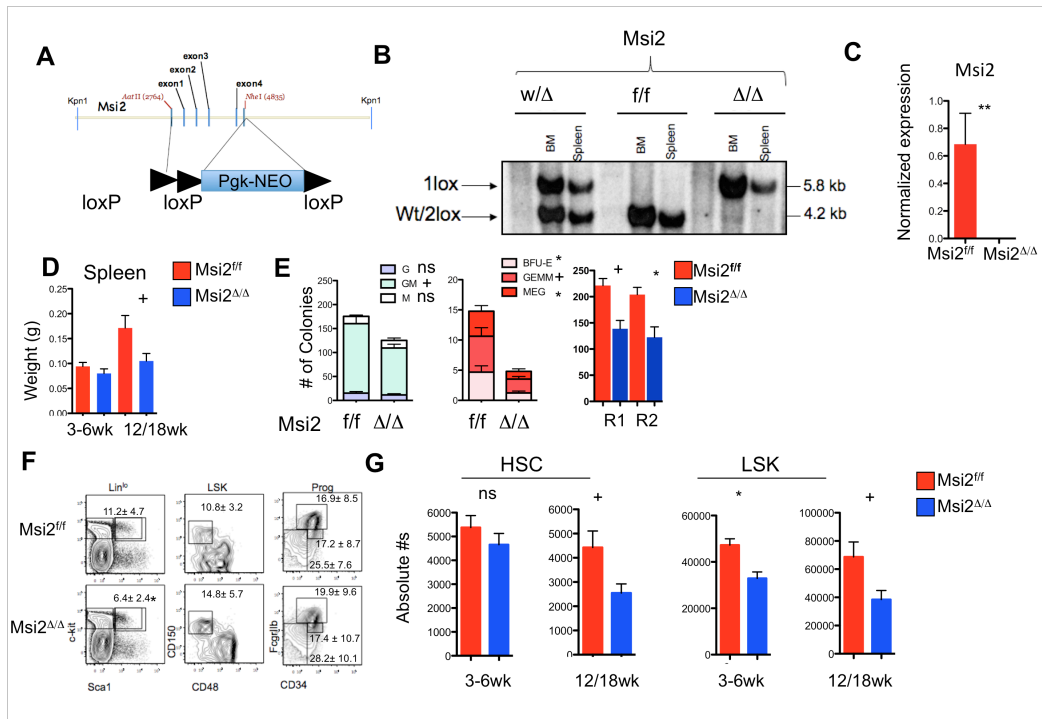
Msi2 (Figure 1F-G and S1).

Figure 3.1: Msi2 conditional knockout have reduced HSC numbers. (A) Targeting schematic for Msi2 conditional knockout mice. (B) Southern blot of indicated genotypes after 4 weeks post pIpC treatment *in vivo.* (C) Quantitative PCR of Msi2 normalized to Gapdh from LSK (lineage low, Sca$^+$ Kit$^+$) sorted cells from 4 week post-pIpC injected mice (n=3 per group). (D) Spleen weights from indicated mice at indicated time points (3-6wk, n=12, 12-18wk, Msi2$^{f/f}$ n=7 Msi2$^{\Delta/\Delta}$n=8). (E) M3434 colony formation assay from BM cells counted at Day 7 CFU-granulocyte (G; granulocyte), mixed Granulocyte and monocyte colony (GM), Burst forming units erythroid lineage (BFU-E), mixed granulocyte, erythroid monocyte and megakaryocyte (GEMM), megakaryocyte (MEG), (n=4 per group) and Serial replating colony assay (right panel), round of replating (R1 or R2), (R1; n=8,10 and R2; 8,6). (F) Representative flow cytometric analysis from mice 3-6-weeks post pIpC (n=12). (G) Absolute number of HSCs and LSKs from indicated mice post pIpC, (3-6wk; n=12, 12/18wk; n=6). Data error bars are SEM and (+p<0.05, *p<0.01, **p<0.001).

132

In order to test if $Msi2^{\Delta/\Delta}$ HSCs are functionally normal, we transplanted the bone

marrow of Msi2 deleted mice into congenic recipients. After engraftment, we

observed reduced chimerism at 6 and 16 weeks in the peripheral blood (Figure

S2). Similarly, 20 weeks post-transplantation we observed decreased

engraftment with a more pronounced defect in the myeloid compartment

observed in the bone marrow and spleen (Figure 2A and S2). At 20 weeks these

mice had reduced spleen size, decreased bone marrow and splenic cellularity

after normalizing for chimerism in each mouse (Figure 2B). Also, we observed a

significant decrease in chimerism in the HSC and progenitor cell compartments

(Figure 2C-D). In order to quantitate the number of HSCs in $Msi2^{\Delta/\Delta}$ mice, we

performed a cobblestone area forming cell assay with limiting dilutions of non-

transplanted $Msi2^{\Delta/\Delta}$ bone marrow (Figure 2E). At 4 weeks we observed a

significant decrease in the frequency of cobblestone formation in $Msi2^{\Delta/\Delta}$

cultures. These findings indicate that Msi2 maintains the function and normal

numbers of HSCs.

To further delineate if this phenotype is cell autonomous, we non-competitively

transplanted cells from both control and $Msi2^{flox/flox}::Mx1-Cre$ bone marrow, which

was then allowed to engraft normally into congenic recipients. The mice were

then injected with pIpC and the peripheral blood was assessed 6 weeks later.

We observed a reduction in the overall white blood cell count that consisted of

decreased myeloid, lymphoid and red blood numbers (Figure 2F). Mice

sacrificed at 18 weeks had reduced chimerism in the blood, which was more

pronounced in the myeloid compared to the lymphoid lineages (Figure 2G and

S2). Similar to the transplants performed with Msi2 deleted cell, the decrease in

chimerism was also observed at the level of the HSC (Figure 2H). This relative

defect in both HSCs and myeloid chimerism prompted us to examine the

frequency of myeloid bias HSCs based assessed by CD150 marker expression.

In accordance with the myeloid defect we found reduced CD150 high myeloid

biased HSCs in the primary Msi2 deleted animals compared to control mice

(Figure 2I). Taken together these results indicate that Msi2 is critical for both

normal HSC maintenance and maintaining the myeloid biased HSC compartment
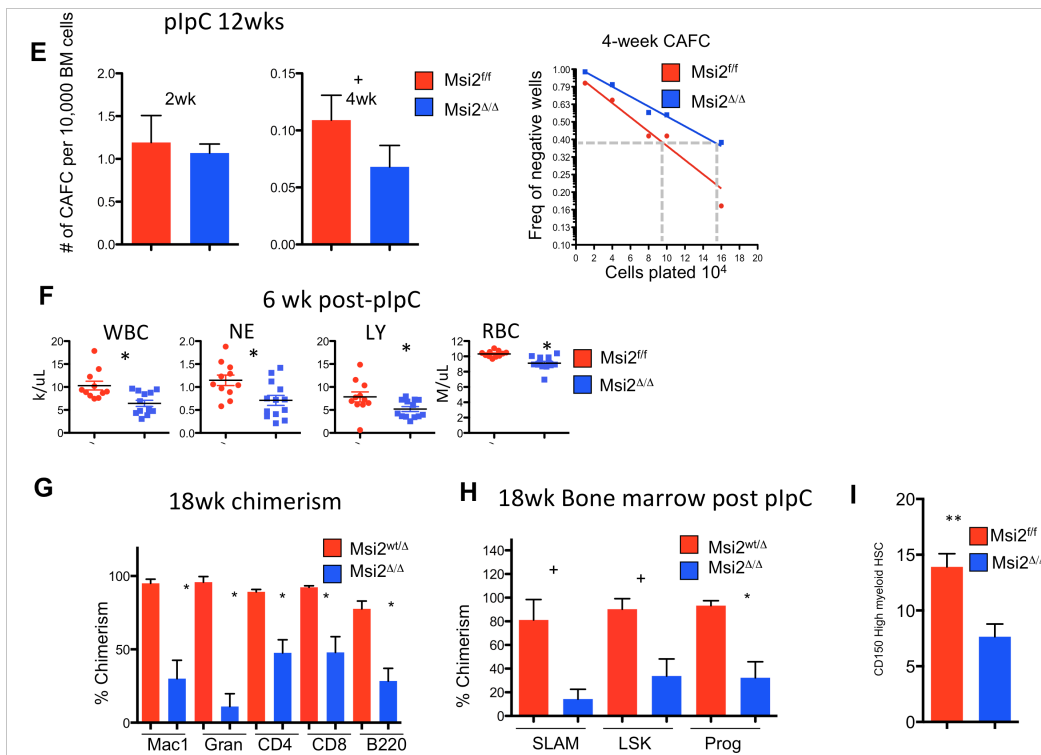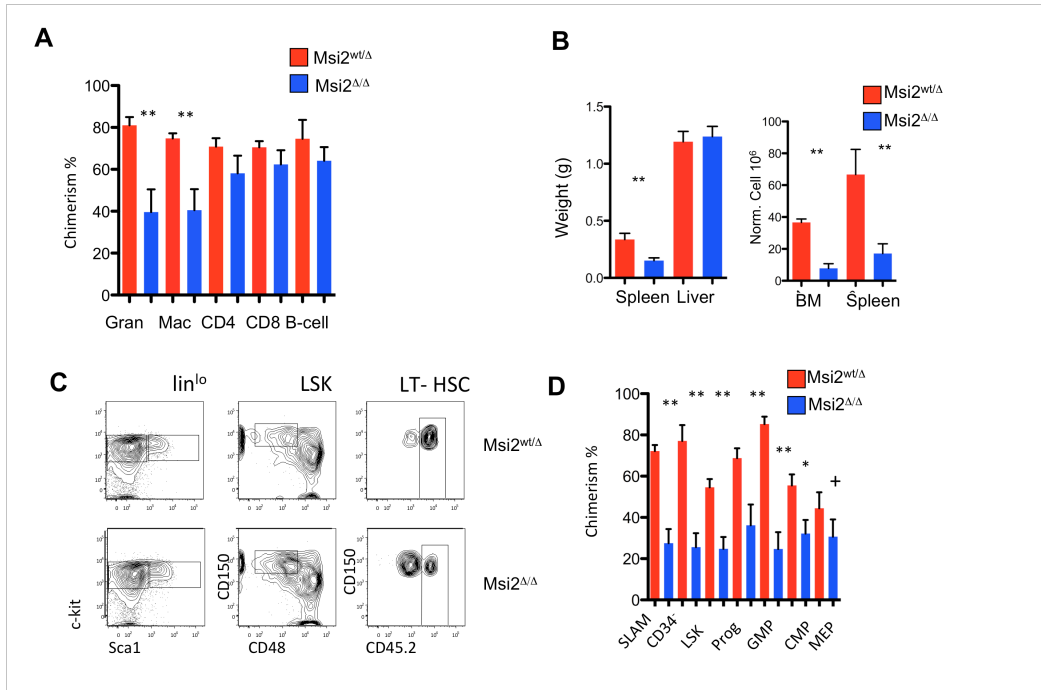
(Beerman et al., 2010).

Figure 3.2: Msi2 deficiency decreases repopulating activity in the myeloid compartment and HSC cell maintenance. (A) Msi2 deleted bone marrow was transplanted non-competitively

Figure 3.2 (Continued):

and analyzed at 20 weeks for donor chimerism in splenic lymphocytes and bone marrow myeloid cells (n=9). Data error bars are SEM and (**p<0.001). (B) Primary bone marrow transplanted animals were sacrificed at 20 weeks post pIpC and indicated organ weights (left panel) and (CD45.2$^+$) cell counts from bone marrow (BM) and spleen which have been normalized for chimerism, are shown (right panel), (n=9). (C) Representative flow cytometric analysis from mice in (B). (D) Chimerism analysis from mice in (B) measured by CD45.2 staining on indicated populations (SLAM; LSK, CD48$^-$ CD150$^+$), (n=8,9). (E) Cobblestone forming assay frequency analysis and limiting dilution assay from indicated time points, (n=2,5). (F) Non-competitively transplanted mice were allowed to normally engraft before pIpC injections with 6-week peripheral blood analysis, (n=11,13). (G) Mice were analyzed at 18 weeks post pIpC for mature marker donor chimerism in the blood (n=4). (H) HSC and progenitor donor chimerism as in (G), (n=4). (I) Frequency of myeloid biased HSC in control and Msi2 deleted bone marrows. (n=15). Data error bars are SEM, (+p<0.05, *p<0.01, **p<0.001).

Consistent with a loss of stem cell self-renewal, we found that $Msi2^{\Delta/\Delta}$ mice had a

reduction in the G_0 quiescent population that was identified in both the

hematopoietic progenitor cells (HPC) and HSCs (Figure 3A). The quiescent cells

within the HSC compartment are considered to have higher engraftment potential

in transplant assays (Passegue et al., 2005). To test if the loss of Msi2 would

sensitize the cells to proliferative stress, we challenged mice with 5-FU. There

was a clear reduction of chimerism in the peripheral blood in the myeloid

compartment whereas the lymphoid compartment was not significantly affected

(Figure 3B and S3). We also found loss of chimerism within the myeloid

progenitors and HSCs in the 5-FU treated mice (Figure 3C-D). Furthermore, we

detected reduced recovery after 5-FU injection within the monocytes and

neutrophils in $Msi2^{\Delta/\Delta}$ compared to controls (Figure S3). Altogether, loss of Msi2

leads to reduced hematopoietic self-renewal during homeostasis, upon

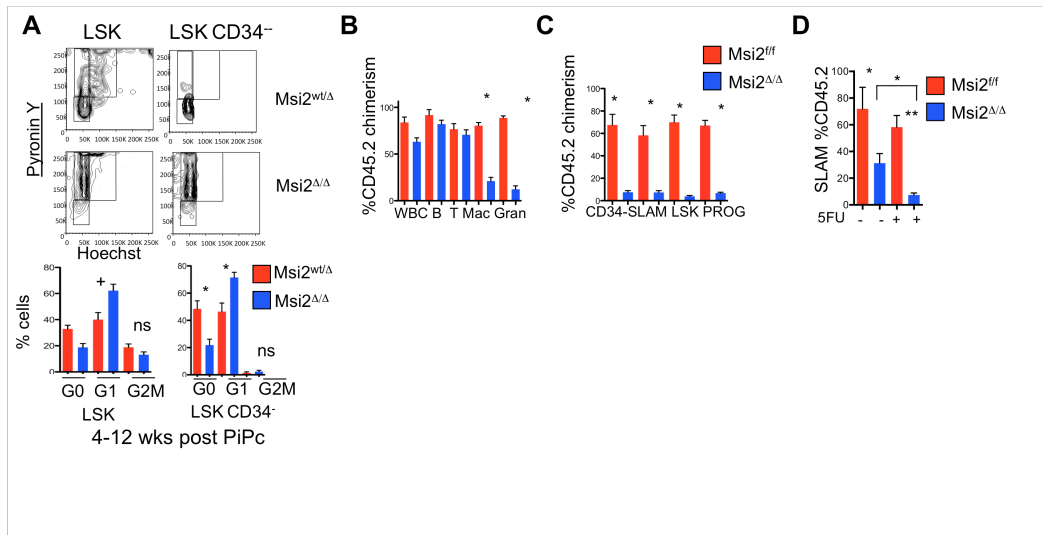transplantation and after replicative stress.

Figure 3.3: Msi2 deficiency reduced quiescence and proliferative stress further depletes HSCs. (A) Representative flow cytometry for cell cycle analysis with statistical analysis (lower panels), (n=6, 3). (B) Indicated mice were transplanted non-competitively (Figure 2F), injected with pIpC, and after 3-weeks were then injected with 5-FU. Peripheral blood chimerism for mature populations was analyzed after 40 days post-treatment, (n=5, 8). (C) Donor chimerism of HSC and progenitor populations from the bone marrow of mice in (B), (n=6, 5). (D) HSC (SLAM) chimerism with or without 5FU injections analyzed at 8-9 weeks same as (C), (n=6, 8) Data error bars are SEM and (+$p<0.05$, *$p<0.01$, **$p<0.001$).

The reduced quiescence and increased sensitivity to stress prompted us to examine if $Msi2^{\Delta/\Delta}$ HSCs were defective in their ability to undergo lineage commitment in a cell-autonomous manner. Indeed, in sorted HSCs and HPCs we found decreased colony formation with reductions in multilineage differentiation (GEMMs), (Figure 4A). We performed *in vitro* proliferation assays on highly purified HSCs and HPCs and found reduced cellular output and increased commitment based on myeloid differentiation (Figure 4B-C). The reduced output and colony formation indicates that Msi2 is required for proper stem cell self-renewal.

To assess if there was a defect in cell fate determination that could compromise self-renewal in $Msi2^{\Delta/\Delta}$ HSCs, we examined Numb protein levels and distribution in HSC daughter pairs as a surrogate readout for asymmetric cell division and as a potential direct target of Msi2 binding. Surprisingly, in the Msi2 null HSC or HPCs we observed equal levels of Numb. This result indicated that Msi2 deficiency does not alter the global expression of Numb within an individual cell (Figure 4D). However, we did observe a decrease in the percentage of cells that underwent asymmetric Numb segregation in the HPCs and an increase in the percentage of cells with symmetric Numb staining in both the HSCs and HPCs (Figure 4E-F). These results suggested that loss of Msi2 altered the cell fate decision towards commitment.
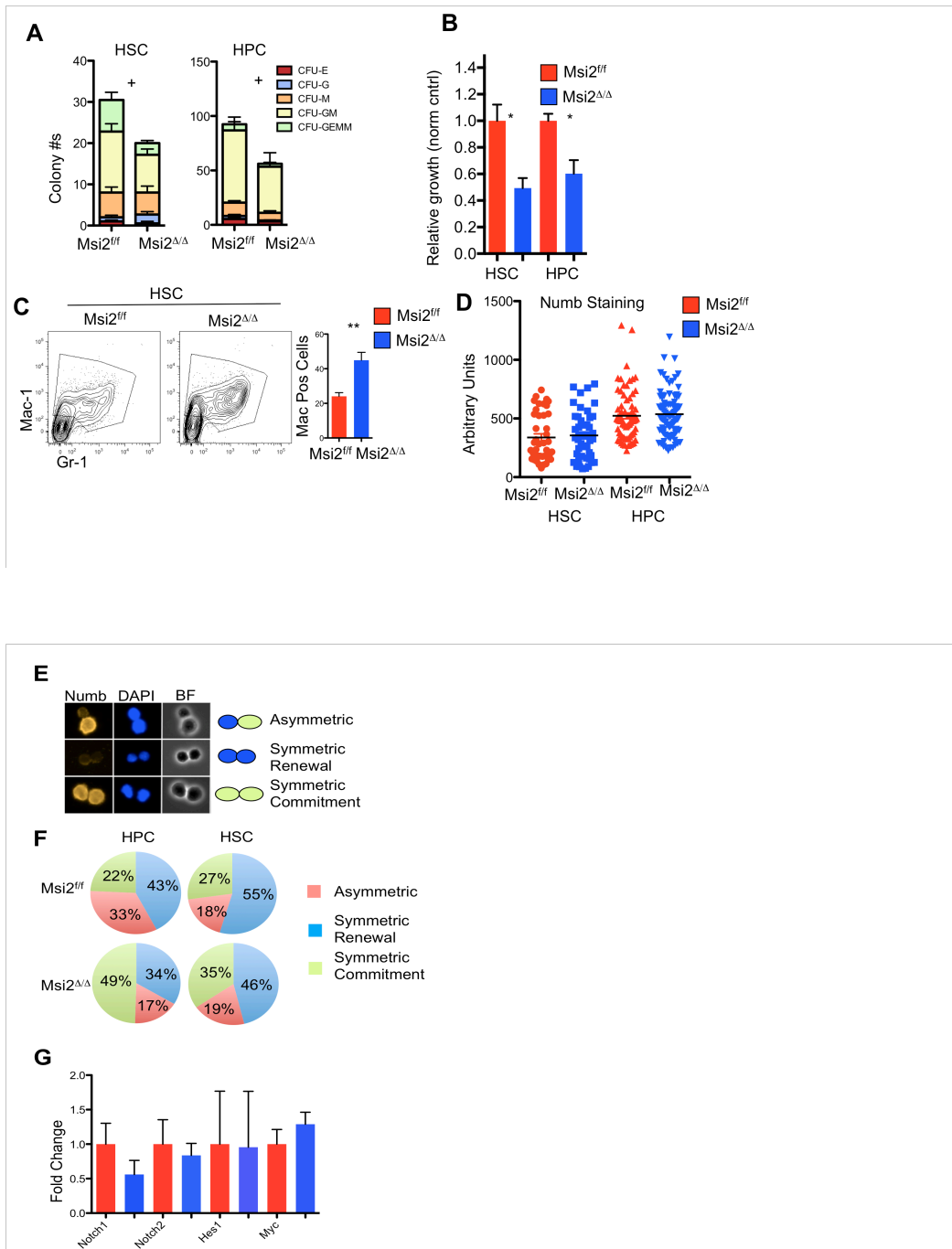
Figure 3.4: HSCs and HPCs from Msi2 deficient mice have reduced self-renewal divisions with differentiation independent of the Numb/Notch axis.

Figure 3.4 (Continued):
(A) Indicated sorted populations were plated in methylcellulose assays from indicated mice HSC (left panel), (n=3, +p<0.05 in GEMM), and HPC (right pane), (n=3, *p<0.05 for Gran), colony designation is the same as in (Figure 1E). (B) Normalized cell growth after in vitro culture of indicated sorted populations from Msi2 deleted donors (6-9 weeks), (HSC; n=14-15, HPC; n=5). (C) Representative flow for mature myeloid markers (left panel) from HSCs grown in (B), Mac positive cells from six independent experiments (right panel), (n=13, 14). (D) Sorted control and Msi2 deleted HSCs and HPCs from 4 week pIpC mice cultured in vitro for 36 hours and stained for expression of Numb and DAPI, (n=4-5 mice combined HSC; number of cells analyzed, 42, 46, HPC; 76, 94). (E) Representative images for paired Numb daughter assay with DAPI (middle panel) and bright-field images (right panel) and experimental scheme for studying asymmetric division in HSC/HPCs. (F) Percentage of doublet cells in each type of cell division as in (E), (n=4-5 independent mice, HSC; 22, 72, HPC; 84, 107). (G) LSK cells analyzed for the expression of indicated Notch targets normalized to Gapdh (n=3-4).

Although controversial, Notch signaling has been implicated in self-renewal of HSCs and as a critical downstream target of the Msi family through its ability to block Numb translation. Using quantitative PCR for Notch target genes in sorted LSK cells from Msi2 deleted mice, we found no statistical difference in the expression of genes of the Notch pathway (Notch1, Notch2, Hes1 or Myc) (Figure S4). These results prompted us to take a more global approach in characterizing the consequence of the loss of Msi2 in hematopoietic cells.

In order to understand which pathways were altered in the $Msi2^{\Delta/\Delta}$ hematopoietic stem cells, we performed transcriptome profiling 4 weeks after Msi2 deletion. We found several statistically significant gene set enrichment analysis (GSEA) signatures that included a loss of self renewal and increased differentiation of stem cells toward CMPs, GMPs and MEPs, and pathways signaling pathways including Cell cycle, Myc, Ras, Wnt and Tgfb (Subramanian et al., 2005) (Krivtsov et al., 2006) (Figure 5A and Supplementary Table 1) These signatures were consistent with the phenotypes we observed in our previous *in vitro* and *in vivo* experiments. Although we found no difference in the small selected targets of Notch signaling by qPCR, we found a statistically significant reduction in the overall pathway based on the GSEA analysis.

Besides alterations in gene expression programs, it remained unclear how Msi2 was directly maintaining self renewal. Therefore, we decided to assess Msi2's global RNA binding targets. Due to the requirement for a large number of cells

and the importance of MSI2 in myeloid leukemias, we utilized K562 (CML-BC cell line) cells overexpressing a FLAG-tagged version of MSI2 to identify its direct RNA-binding targets. Using recently developed technologies, we performed experiments with UV-cross-linked and immunoprecipitated MSI2 which was then followed by RNA-sequencing (HITS-CLIP) to identify the global direct binding associations (Chi et al., 2009). Our analysis characterized the distribution of binding across the genome with a majority of binding found in UTR-specific regions. We then calculated a log fold change of binding over the empty vector control to quantify the specific regions of interest. Using the entire ranked RNA target list we identified GSEA signatures including Cell Cycle, ƳSelf-renewalФ and ƳHSC to CMP differentiationФ that matched pathways we already identified using transcriptome analysis from Msi2 deleted LSKs (Figure 5B).

We re-examined the microarray data to identify a specific pathway or genes that may provide further insight into Msi2 function. We ranked and plotted a heat map of the top and bottom 40 genes that showed differences in expression (Figure S4). In Msi2-deleted HCS, the most depleted gene based on fold change was Tgfbr1 (Figure S4). The TGFbeta pathway has been previously shown to be critical for self renewal, quiescence and maintenance of HSCs. Based on Ingenuity Pathway Analysis and the TGFb gene set signature we found a reduction of expression of upstream and downstream components of the pathway (Figure 5C-D). Additionally, when we examined our HITS-CLIP data, we found Msi2 also binding directly to multiple components of the TGFb pathway as

143

observed by Ingenuity and GSEA analysis (Figure 5E-F and S5). Although the

GSEAs and Ingenuity analysis of the microarray suggested differences in the

transcript abundance of pathway components, we only observed a trend toward

reductions with no statistical difference in expression except for an increase in

Tgfb1 (Figure S5). The pathway analysis suggested that Msi2 binding acts

through a post-transcriptional mechanism and modulating multiple nodes within a

pathway.

Based on the expression and HITS-CLIP data, we examined the functional

signaling of the Tgfb pathway. An established readout for TGFb signaling is the

phosphorylation of downstream components such as Smad2/3. We found

reduced phosphorylation of Smad2/Smad3 in HSCs grown *in vitro* or directly

isolated from bone marrow (Figure 5G and data not shown). Another readout for

TGFb signaling is the expression of cell cycle regulators p21, p27 and p57.

Consistent with reduced signaling output of Tgfb we observed a significant

decrease in p57 expression (Figure 5H). The role of p57 has been recently

highlighted as a critical regulator of HSC quiescence (Zou et al., 2011)

(Matsumoto et al., 2011). Interestingly, we also observed a small but significant

increase in p27 levels in consistent with a Tgfb pathway-independent mechanism

and probably a compensatory mechanism for p57 reduction. Loss of p21 or p27

was not required for growth inhibition of Tgfb in hematopoietic cells(Cheng et al.,
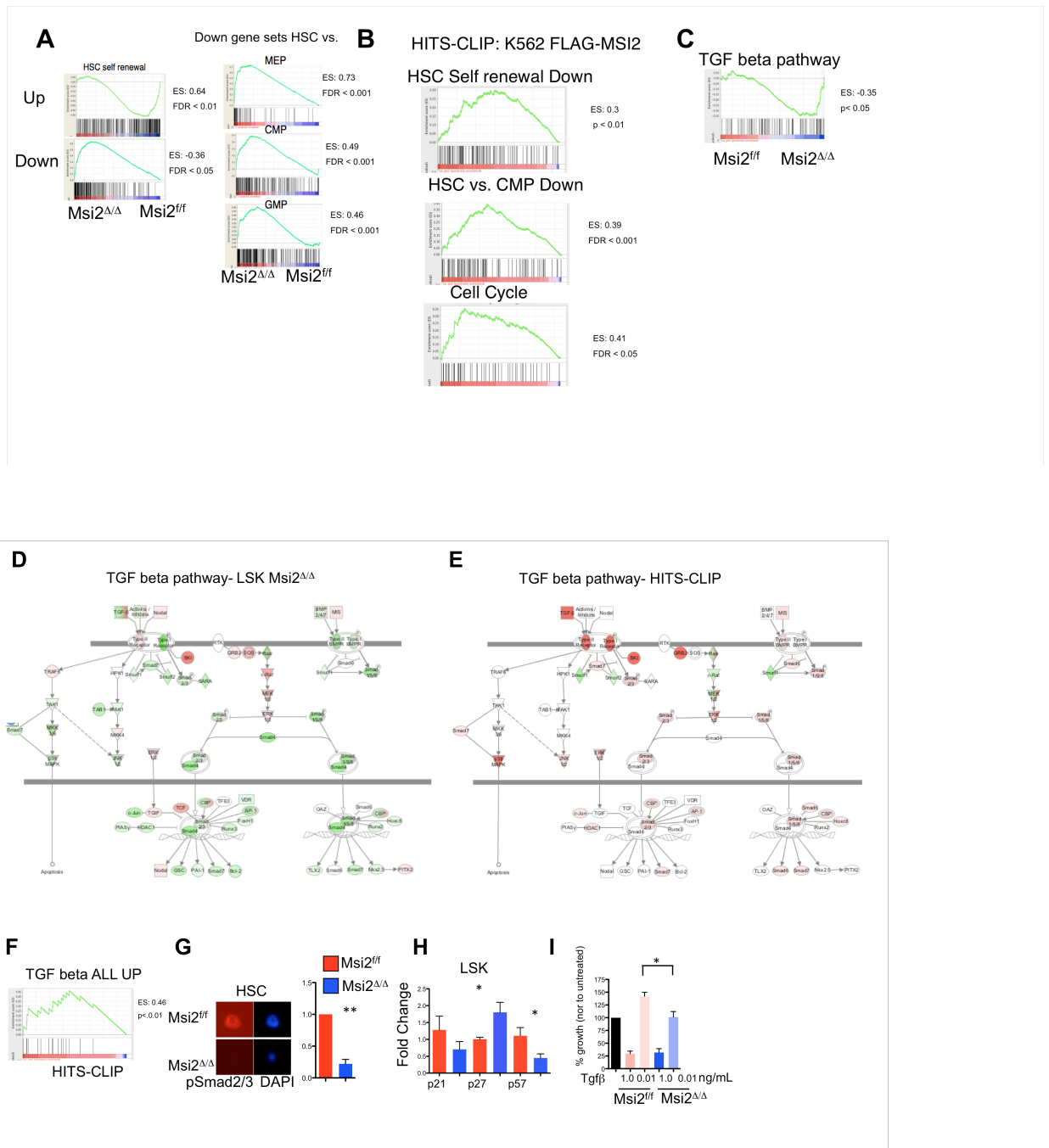
2001).

Figure 3.5: Msi2 modulates global gene expression and direct RNA targets uncovering a role for self-renewal, differentiation and Tgfb signaling. (A) Control and Msi2 deleted LSKs were sorted, RNA was isolated and then analyzed by microarray. Ranked fold change between control and Msi2 deleted LSKs analyzed by GSEAs for self renewal gene signatures were decreased in Msi2 deleted LSKs while differentiation signatures were increased in Megakaryocyte erythroid progenitor (MEP), common myeloid progenitor (CMP) and granulocyte monocyte progenitor (GMP) gene signatures (B) GSEA of the pre-ranked list of all the RNA binding targets mapping that have been mapped and log fold was calculated over the control to genes analyzed against the indicated gene sets. (C) GSEA analysis of the Tgfb pathway as in

Figure 3.5 (Continued): (A). (D) Ingenuity pathway analysis overlapping ranked fold change gene expression profiling from (A) over the Tgfb pathway. Green color indicates downregulated genes and red color indicated upregulated genes. (E) Ingenuity pathway analysis of all log fold-change ranked HITS-CLIP RNA targets overlapped with TGFB pathway with red color indicating enriched binding over the background and green color indicating negatively enriched binding. (F) GSEA analysis of the HITS-CLIP ranked list over the TGFb upregulated gene signature. (G) Sorted HSCs cultured as in Figure 4E were stained with pSmad2/3 specific antibodies and DAPI (representative image left panel) and fluorescence was quantified and normalized to the control (right panel), (n=3 independent experiments, 38-40 cells quantified per group in total). (H) QPCR expression profiling of TGFb gene targets from control and Msi2 deleted LSKs, (n=5). (I) Proliferation assay of HSCs sorted and grown for seven days as in (Figure 4B) and Tgfb1 was added at indicated concentrations. Absolute cell numbers were normalized by each samples untreated growth. Data error bars are SEM and (*p<0.01, **p<0.001).

Tgfb addition has been shown to be biphasic with high concentrations blocking

proliferation and low doses leading to increased proliferation in a subset of

myeloid-biased HSCs (Challen et al., 2010; Kale, 2004; Kale and Vaidya, 2004).

Finally, we tested if there was a difference in the functional response to TGFb

signaling in vitro using two different doses of TGFb. The suboptimal Tgfb

signaling found in steady state conditions in Msi2 deleted HSCs was overcome

by the addition of high dose of exogenous TGFb based on its equivalent growth

suppressive effects compared to the control (Figure 5I). In contrast, at the lower

dose when normal HSCs expand, Msi2 deleted HSCs failed to respond to the

stimulation indicating diminished sensitivity to proliferative effect of Tgfb1 (Figure

5J). Overall, this defect in Tgfb signaling and self-renewal pathways corresponds

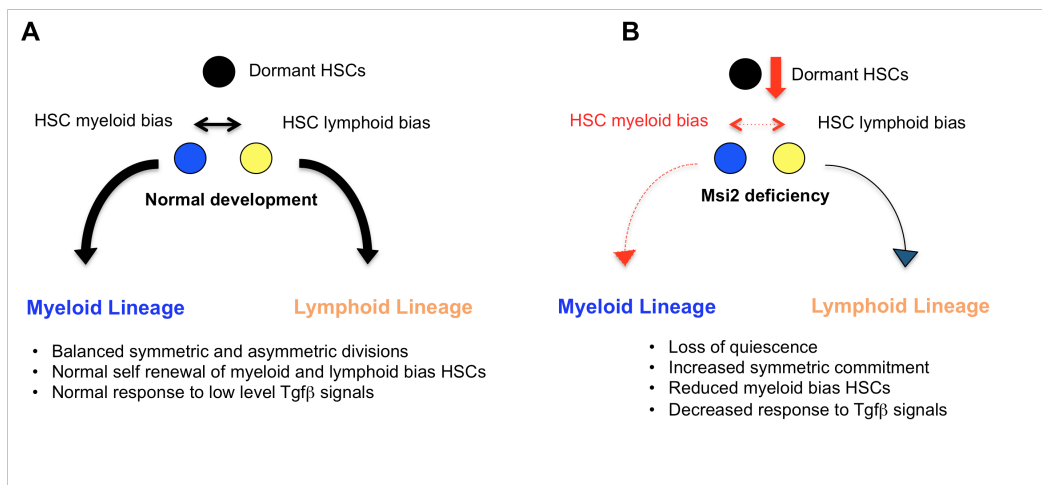to the reduced overall numbers and function of Msi2 deleted HSCs.

Figure 3.6. Model of Msi2 function in HSC. (A) Myeloid and lymphoid biased HSCs undergo normal self renewal with balanced symmetric and asymmetric divisions giving rise to differentiated progeny.. Normal HSCs proliferate in response to low levels of Tgfb signals. (B) Deletion of Msi2 in HSCs leads to loss of quiescence and increased symmetric commitment towards differentiated progeny. Additionally Msi2 loss leads to reduced myeloid biased HSCs, which are unresponsive to Tgfb-induced proliferative signals.

DISCUSSION

In this report, we studied the role of *Msi2* in HSCs using a conditional knockout murine model. Conditional ablation of *Msi2* in the hematopoietic system results in dramatic reductions in steady state HSC and HSPC numbers with no changes in the maturation of myeloid or lymphoid hematopoietic lineages. We show that *Msi2* is required for HSC quiescence and for maintaining the balance of self-renewal and lineage commitment needed to regulate normal hematopoiesis. Notably, studies of single HSCs allowed us to uncover how Msi2 maintains stem cell identity and to show this was not due solely to regulation of Numb protein levels. Applying global RNA target analysis of MSI2 binding in leukemic cells combined with transcriptional analysis in mouse HSCs, we identified common self-renewal and differentiation pathways. Additionally, using this approach we uncovered a novel role for Msi2 in the regulation of TGF-β signaling by demonstrating reduced TGF-β signaling output and impaired response to TGF-β stimulation in *Msi2* deficient HSCs.

Previous studies from our group and others have identified a role for *Msi2* in hematopoietic development, but it remained unclear how Msi2 functions and what pathways Msi2 regulates in hematopoietic cells(de Andrés-Aguayo et al., 2011; Hope et al., 2010; Ito et al., 2010; Kharas et al., 2010). Our studies are consistent with experiments utilizing shRNAs specific to *Msi2* that revealed reduced engraftment of Msi2 knock-down cells(Hope et al., 2010; Kharas et al., 2010). The analysis of hematopoietic cells from mice harboring an insertional

gene trap at the *Msi2* locus demonstrated reduced repopulating activity *in vivo* without an apparent defect in HSCs with hypomorphic levels of Msi2(de Andrés-Aguayo et al., 2011). Also, the same study suggested that there was a proliferative defect in the LMPP compartment. In accordance with this recent study, we find reductions in HSPCs, but also find a significant defect in the HSCs after genetic ablation of *Msi2*. Moreover, we demonstrate that this is due to a failure in maintaining quiescence and skewing symmetric self-renewal divisions towards differentiation. The apparent discrepancy in the defect within the HSCs between our two mouse models suggests that the loss of MSI2 may be compensated during development or could be due to differences between gene trap-mediated insertional disruption and the conditional deletion approach *in vivo*.

Studies on the Msi family have considered Numb to be a critical downstream target of Msi mRNA binding and translational suppression(Imai et al., 2001). Numb modulates a variety of pathways, most notably the inhibition of Notch and hedgehog pathway components as well as stabilization of p53(Gulino et al., 2010). However, we found equivalent expression of canonical Notch target genes in HSPCs in either our previously described inducible *MSI2* knock-in model or in *Msi2*$^{\Delta/\Delta}$HSCs (Figure S4 and(Kharas et al., 2010)). Previous studies examined the relationship between Msi1 and Numb, predominantly in the neural system; more recently the inverse relationship between Numb and Msi2 expression was demonstrated in blast crisis leukemic cells (Imai et al., 2001; Ito et al., 2010; Kharas et al., 2010). Specifically, *MSI2* overexpression resulted in reduced Numb

149

staining in CML-BC murine cells(Ito et al., 2010). We also previously found that

shRNA depletion of *MSI2* led to increased NUMB levels in human CML-BC and

AML cell lines(Kharas et al., 2010). However, this correlation was not clearly

seen in HSPCs overexpressing MSI2 (data not shown and (Kharas et al., 2010)

nor in *Msi2$^{\Delta/\Delta}$*HSCs (Figure 4D). We found that overexpression of *MSI2* in

HSPCs did not decrease Numb staining, but did, however, increase the

asymmetric segregation of Numb protein into daughter cells after mitosis(Kharas

et al., 2010). Conversely, we demonstrated that *Msi2$^{\Delta/\Delta}$* HSPCs had reduced

asymmetric segregation of Numb, and in both HSPCs and HSCs we observed

increased symmetric commitment divisions (Figure 4E-F). This apparent

difference between the correlation of staining of Msi2 and Numb in both normal

and leukemic cells, and between the effects of *MSI2* deficiency and

overexpression was also found recently in human samples(Pereira et al., 2012).

Studies of normal CD34^{+} cells from healthy bone marrow displayed no

correlation between *MSI2* and *NUMB* expression, while the same study found

reduced expression of *NUMB* in MDS/AML samples with the highest level of

MSI2 expression(Pereira et al., 2012). It is also unclear if MSI2 regulates *NUMB*

expression through indirect transcriptional effects, or through a direct post-

transcriptional mechanism. Our HITS-CLIP data demonstrates that MSI2 does

not bind *NUMB* nor does Msi2 knockdown lead to increased protein expression

after shRNA depletion of *MSI2* in K562 cells (data not shown). Nevertheless, the

relationship between mRNA/protein expression of *NUMB, and MSI2* in both

normal and leukemic cells should be further examined.

To gain an unbiased understanding of the pathways downstream of Msi2 in

maintaining proper HSC self-renewal, we employed transcriptome-wide MSI2

RNA-binding target analysis. A previous global study examining Msi1 using

native RNA immunoprecipitation (RIP-CHIP) identified targets known to

contribute to oncogenic transformation(de Sousa Abreu et al., 2009). In our

study, we utilized a recently developed approach for studying RNA binding

proteins (HITS-CLIP) that involves UV cross-linking to generate more specificity

in binding targets through direct interactions between target proteins and

RNA(Krivtsov et al., 2006). This allowed us to define a global map of MSI2 direct

RNA targets for the first time. Moreover, integration of our HITS-CLIP data with

expression data performed in *Msi2* deficient hematopoietic stem and progenitor

cells identified a set of developmental pathways, which are altered through direct

RNA-protein interactions.

One of the most significant pathways we identified was the TGF-β pathway,

which has been linked to maintaining both normal and leukemia stem cell self-

renewal. However, similar to the Notch pathway, a clear understanding of how

TGF signaling alters normal HSC behavior remains elusive. Initial studies *in vitro*

with TGF ligands suggested that TGF-β signaling could block proliferation of

HSCs and maintain quiescence (Sitnicka et al., 1996). Conditional deletion

studies of mice deficient of TGF-β1 signaling revealed no detectable alterations

in the HSC compartment *in vivo,* but increased proliferation was observed in vitro

at low concentrations of stimulation within Tgfb1 deficient HSCs(Larsson et al.,

2003; Larsson et al., 2005)*.* In contrast, loss of *TGF-βrII* or *Smad4* resulted in

reduced chimerism in competitive transplants(Yamazaki et al., 2011). Other

studies have demonstrated that myeloid biased HSCs and myeloid cell lines

treated with low levels of TGF *in vitro* exhibit increased proliferation(Challen et

al., 2010; Kale, 2004; Kale and Vaidya, 2004). Here we link Msi2 to the TGF-β

pathway by demonstrating that it can directly interact with mRNAs encoding the

components and targets of the pathway. We validated these results by

demonstrating a functional reduction in Smad2/3 phosphorylation, decreased *p57*

expression, and insensitivity to the growth stimulatory effects of TGF-β.

Furthermore, the reduced TGF-β signaling ultimately results in a decrease in

HSC self renewal and a dramatic reduction in myeloid biased HSCs. This

observation explains the significant reduction in the mature cells of the myeloid

lineage compared to the lymphoid lineage (Figure 6). Furthermore, p57 a critical

downstream target of TGF-β signaling which is important in maintaining

quiescence in HSCs, is significantly suppressed upon *Msi2* deletion(Cheng et al.,

2001; Matsumoto et al., 2011; Scandura et al., 2004).

Our report characterizes *Msi2* deletion in the hematopoietic system and uses

global RNA binding analysis to identify a novel link between Msi2 and TGF-β

signaling in regulating quiescence, self-renewal and myeloid/lymphoid lineage

determination of HSCs. It will be interesting to test if this pathway is retained and functionally important for leukemic stem cells or if Msi2 modulates other critical pathways in these cells. Taken together, these results have important implications for understanding how the activity of RNA binding proteins contribute to normal tissue homeostasis and stem cell self-renewal, and how MSI2 and its targets contribute to oncogenic transformation.

EXPERIMENTAL PROCEDURES

Quantitative RT-PCR

mRNA was isolated from FACS-purified populations using TRIZOL (Life

Technologies) and RNEASY RNA extraction kit (Qiagen), and reverse

transcribed into cDNA with iSCRIPT (BioRad). Quantitative PCR was performed

as previously described with indicated primers (Supplementary Table 4) for p21,

p27, p57, smad2, smad3, smad4, smad7, TGF-βr1, TGF-βr2, TGF-β1, TGF-β2,

Notch1, Notch2, Hes1, Hey, Myc and Actin.

Msi2 conditional knockout mice

We cloned one lox P site 3' of the transcriptional start site and a floxed PGK-

neomycin-resistance cassette 3' of the fifth intron using a Kpn1 fragment

encompassing the 3' region of the *Msi2* locus cloned into the pBS backbone.

(Figure 1A and S1A). The targeting vector was electroporated into V6.5 ES cells

and G418 resistant colonies were subcloned. Homologous recombinants was

detected by Southern blotting using 5' and 3' external probes, and correctly

targeted ES clones were transiently treated with Cre to remove the Neo

resistance cassette (verified by Southern blotting), prior to blastocyst injection of

conditional 2lox clones. Resulting chimeras were backcrossed to Bl/6 mice and

the F1 generation was bred to Mx1 Cre mice on a Bl/6 background. Deletion was

initiated using pIpC (Amersham) by intraperitoneal injections every 3-4 days as

indicated (150mg/kg) per mouse three times. Cre recombination of the

conditional 2-lox (floxed) locus to the inactive, 1 lox (null) locus was verified by

Southern blotting after cutting genomic DNA with Xba I and hybridization with a 5'
external probe.

Non-competitive transplants

Non-competitive transplants were performed with 1×10^6 bone marrow cells from
6-10 week old $Msi2^{flox/flox}$ or littermate control mice, transplanted into lethally
irradiated B6SJL congenic CD45.1 recipients. In the primary transplants with
$Msi2^{\Delta/\Delta}$ bone marrow, transplants were performed 12-16 weeks after $Msi2$
deletion. In the cell autonomous transplants, transplanted marrow was allowed to
engraft for 4-6 weeks before pIpC was administered. Peripheral blood chimerism
was assessed post-engraftment prior to deletion to verify successful engraftment.

Peripheral blood analysis

Peripheral blood was collected from the retro-orbital cavity using a heparinized
glass capillary tube. Complete peripheral blood count analysis including a
differential blood count was obtained by using Hemavet (Drew Scientific).

Flow Cytometry

Bone marrow cells or splenocytes were harvested and subjected to red blood cell
lysis. Fresh or frozen cells were stained with the following antibodies: CD45.2-
FITC and CD45.1-APC, Mac1-PE, Gr1-APC, c-Kit-APC, CD71-PE, Ter119-APC,
B220-PE, and CD3-APC (Becton-Dickinson) and analyzed on the BD
FACSCalibur instrument. Staining for multiparameter flow cytometry was

performed after a c-kit enrichment using 10 uL of MACS beads (CD117) per mouse and then run on an AutoMACs per manufacturer instructions. The cells were then stained with the following cocktail: (Lineage; CD3, CD4, CD8, Gr1, B220, CD19, TER119 all conjugated with PeCy5), Sca-Pac Blue, CD34-FITC or CD45.2 FITC, SLAM-APC, CD48-PE, c-KIT-AlexaFluor 780, FcgRIIb-PeCy7, Figure 2D; HPC(Lin^{lo} c-Kit$^+$ Sca1$^-$), GMP (LK, FcgRIIb^{hi} CD34$^+$), CMP (LK, FcgRIIb^{mid} CD34$^+$), MEP (LK, FcgRIIb^{low} CD34$^-$), B-cells (B220$^+$) and T-cells (CD3$^+$) from the spleen were also sorted.

Colony-Forming Assays

Bone marrow cells were harvested, subjected to red blood cell lysis, and resuspended in RPMI/10%FBS/5% penicillin-streptomycin. Cells were plated in duplicate in M3434 methylcellulose media (Stem Cell Technologies) at 1x10^4 cells/dish for bone marrow and 5x10^4 cells/dish for spleen cells. Colonies were scored after 7 days. If at least 25 colonies were observed per dish, the cells were replated in M3434 media at 1x10^4 cells/dish, and then counted and replated every 7 days. 200 sorted HSCs or 500 sorted HSPCs were plated and scored after ten and seven days respectively.

Microarray Analysis of *Msi2* expression in conditional knockout mice. Hematopoietic stem and progenitor cells (Lineage^{lo} ScaI$^+$ c-Kit$^{+)}$ were sorted 4 week post-pIpC injection. RNA was extracted using TRIZOL and RNEASY RNA extraction kit. RNA was then amplified using NUGEN Pico amplification kit,

156

fragmented and hybridized on Mouse Expression Array 430 2.0. Signal normalization was performed by RMA method. Data was analyzed using GSEA across the complete list of genes ranked by signal-to-noise ratio. (Microarray data is representative of n=4 $Msi2^{flox/flox}$ mice and n=4 $Msi2^{\Delta/\Delta}$ mice.

Immunofluorescence

HSCs or HSPCs were sorted from primary mice 6 weeks post-pIpC and cultured with STIF media in 96 round bottom wells for 16 hours and then treated with 10nM of Nocodazole for 24 hours. Cells were then cytospun, fixed and methanol permeabilized and then stained on slides with anti-Numb (Ab4147 Abcam) and secondary antibody (donkey anti-goat Alexa fluor 546, Molecular Probes) with DAPI counterstaining(Kharas et al., 2010). Symmetric versus asymmetric percentages were assessed based on quantification of the signal intensity of each cell (divided by surface area) normalized for background staining using a Zeiss Microscope followed by a confirmatory visual inspection that determined symmetric commitment versus symmetric renewal. Daughter cells with elevated equivalent staining of Numb were counted as symmetric commitment whereas if one daughter cell contained more staining than the other cell then the division was considered as asymmetric division. If there was low or no staining in the daughter pairs, this was scored as a symmetric renewal division. Cells were also stained for phospho-Smad2/3 and staining was quantified as above in all cells including non-dividing cells. Phospho-Smad2/3 was also performed on HSCs directly isolated from animals with similar results.

Cobblestone area-forming Assay

We performed CAFC assays as described in(Moore et al., 1997). In short, we plated bone-marrow cells in 96 flat bottom wells in limiting dilutions on top of confluent OP9 stroma at the indicated doses of bone marrow cells: $1x10^4$, $4x10^4$, $8x10^4$, $16x10^4$ with 12 replicates per dose. The number of cobblestone positive wells per dose were enumerated and at 2 weeks post-plating. L-Calc software (Stem Cell Technologies) was used to calculate frequency and p-values between groups.

In vitro HSC and HSPC proliferation Assay

We sorted 100 HSCs (LSK $CD48^-$ $CD150^+$) and 1000 HSPCs (LSK $CD150^-$) into SFEM medium containing 10 ng/ml IL-3, 10 ng/ml IL-6, 50 ng/ml SCF, 10 ng/ml TPO and 20 ng/ml Flt3L in a round bottom plate and counted at day 7 and 5, respectively. Then cells were analyzed using counting beads and indicated markers for differentiation by flow cytometry.

Statistical Analyses

For bar graphs, the unpaired 2-tailed student's T-test was used to compute p values, except where stated otherwise. Error bars reflect the standard error of the mean (SEM), except where stated otherwise. All statistical analyses were carried out using Graph Pad Prism 4.0 and the R statistical environment.

UV-crosslinking, immunoprecipitation and high throughput RNA sequencing (HITS-CLIP)

The protocol used is mainly based on UV-crosslinking and immunoprecipitation with high throughput RNA sequencing (HITS-CLIP) published by (Chi et al., 2009; Krivtsov et al., 2006). In brief, 100 million empty vector control- or Flag-MSI2- overexpressing K562 cells were UV-crosslinked with 4000 J at 265 nm on ice. Cells were washed three times with cold PBS containing protease inhibitors and the cell pellets were resuspended in a low salt buffer that is optimized for reducing non-specific RNA interactions (100 mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 0.5% sodium deoxycholate, 0.5% Nonident P-40) with protease inhibitors. Cells were sonicated with a Branson probe sonicator for three cycles of 10 seconds (0.7 sec on/1.3 sec off) at 20% power on ice. The lysates were then rested on ice for 20 minutes. RQ1 DNase (Promega, Madison, WI) was added to the lysates and incubated at 37°C for 5 minutes. RNAse A (USB Products, Cleveland, OH) was added at a concentration of 1:1000 and incubated at 37°C for 10 minutes. Lysates were spun and supernatants were incubated with protein G Dynabeads (Invitrogen, Grand Island, NY) that were pre-incubated with anti-Flag M2 Ab in low salt buffer (Sigma, St. Louis, MO). Beads were washed with low salt wash buffer three times followed by two washes with high salt buffer (500 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% Nonident P-40) to reduce non-specific protein interactions. Proteinase K (Invitrogen) (5 ug) was added to the washed beads and incubated at 70°C for 2 hours. RNA was

extracted using standard Phenol/Chloroform extraction methods, the RNA was precipitated at 80°C overnight, and the resulting RNA pellets were washed with 80% ethanol to retain short RNA fragments. RNA sizes were assessed on a 2100 Bioanalyzer (pico chip) and the sizes of extracted RNA ranged between 40 – 200 bases in length. RNA libraries were prepared and barcoded at the Broad Institute Sequencing core facility and then sent to the core facility at Sloan Kettering Institute for single-end Illumina sequencing. A first round of sequencing for one control and two replicates produced ~20 million 36-bp alignable reads for each library. The second replicate was sequenced for an additional 60 million reads to enhance coverage.

Statistical analysis and plots of HITS-CLIP data

To identify potential binding sites, we used a custom R package to call peaks in read coverage in the Msi2 HITS-CLIP experiments ((Loeb et al., 2012) and see Supplemental Online Methods), and we tested for differential binding between Msi2 overexpressing cells and control while correcting for differential expression using a joint generalized linear model of HITS-CLIP and RNA-sequencing read counts in the candidate peak regions (Supplemental Online Methods). Peak windows that were differentially bound at an FDR threshold of 0.1 were reported unless otherwise noted. Plots of HITS-CLIP maps provide positional read coverage after scaling for library size. For visualization purposes, positional read counts are then set to the highest peak within each coverage plot. Peaks identified by the peak calling procedure are shown in color if they are significantly

160

bound (red, FDR < 0.05; magenta, FDR < 0.1, dark grey FDR < 0.12) in MSi2 pulldown versus control. We identified enriched motifs by running differentially bound peak regions in MEME separately on 3'UTR and CDS binding site sequences (Supplementary Online Methods).

ACKNOWLEDGMENTS

References:

Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* 413, 732–738 (2001).

Amador-Ortiz, C. *et al.* TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann Neurol.* 61, 435–445 (2007).

Amit et al. Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses. Science (2009) vol. 326 (5950) pp. 257-263

Arai et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun.351(3):602-11 (2006).

Ares M et al. A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA. 1999 Sep;5(9):1138-9.

Avendano-Vazquez, S. E. *et al.* Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev* 26, 1679–1684 (2012).

Ayala, Y. M., Pagani, F. & Baralle, F. E. TDP43 depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett.* 580, 1339–1344 (2006).

Ayala, Y. M. *et al.* Structural determinants of the cellular localization and shuttling of TDP-43. *Journal of Cell Science* 121, 3778–3785 (2008).

Ayala, Y. M. et al. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. Proc Natl Acad Sci 105(10):3785-9 (2008).

Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* 30, 277–288 (2010).

Beckmann, J., Scheitza, S., Wernet, P., Fischer, J.C., and Giebel, B. (2007). Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. Blood *109*, 5494-5501.

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I.L., Bryder, D., and Rossi, D.J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proceedings of the National Academy of Sciences *107*, 5465-5470.

Bigas, A., and Espinosa, L. (2012). Hematopoietic stem cells: to be or Notch to be. Blood *119*, 3226-3235.

Byers, R.J., Currie, T., Tholouli, E., Rodig, S.J., and Kutok, J.L. (2011). MSI2 protein expression predicts unfavorable outcome in acute myeloid leukemia. Blood *118*, 2857-2867.

Bilican B et al. Mutant induced pluripotent stem cell lines recapitulate aspects of TDP-43 proteinopathies and reveal cell-specific vulnerability. PNAS 109(15):5803-8 (2012).

Bjorkqvist, M., Wild, E. J. & Tabrizi, S. J. Harnessing Immune Alterations in Neurodegenerative Diseases. *Neuron* 64, 21–24 (2009).

Boillee, S., VANDEVELDE, C. & CLEVELAND, D. ALS: A Disease of Motor Neurons and Their Nonneuronal Neighbors. *Neuron* 52, 39–59 (2006).

Bregman, A. *et al.* Promoter Elements Regulate Cytoplasmic mRNA Decay. *Cell* 1–18 (2011).doi:10.1016/j.cell.2011.12.005

Buratti, E. & Baralle, F. E. *Chapter 1 - The Molecular Links Between TDP-43 Dysfunction and Neurodegeneration. Advances in Genetics* **66**, 1–34 (Elsevier Inc.: 2009).

Butovsky, O. *et al.* Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS. *J. Clin. Invest.* (2012).doi:10.1172/JCI62636DS1

Buratti, E. Characterization and Functional Implications of the RNA Binding Properties of Nuclear Factor TDP-43, a Novel Splicing Regulator of CFTR Exon 9. *Journal of Biological Chemistry* 276, 36337–36343 (2001).

Carmo-Fonesca M et al. Mammalian nuclei contain foci which are highly enriched in components of the pre-mRNA splicing machinery. EMBO J 1991, 10:195-206.

Challen, G.A., Boles, N.C., Chambers, S.M., and Goodell, M.A. (2010). Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGF-[beta] 1. Stem Cell *6*, 265-278.

Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A., Vinup, K.E., Mrass, P., Oliaro, J*., et al.* (2007).

Asymmetric T lymphocyte division in the initiation of adaptive immune responses. Science *315*, 1687-1691.

Cheng, T., Shen, H., Rodrigues, N., Stier, S., and Scadden, D.T. (2001). Transforming growth factor beta 1 mediates cell-cycle arrest of primitive hematopoietic cells independent of p21(Cip1/Waf1) or p27(Kip1). Blood *98*, 3643-3649.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Cho EJ et al. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev 1997, 11:3319-3326.

Choder M. mRNA imprinting. Cellular Logistics 2011, 1:1, 37-40.

Clement, A. M. *et al.* Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice. *Science* 302, 113–117 (2003).

Cleveland, D.W. and Rothstein, J.D. (2001) From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nat. Neurosci. Rev. 2, 806-819.

Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* 2, 440–452 (1988).

Couthouis, J. *et al.* Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet* 21, 2899–2911 (2012).

Da Cruz, S. & Cleveland, D. W. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Curr. Opin. Neurobiol.* 21, 904–919 (2011).

Dantonel JC. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature. 1997 (6649):399-402.

Dahmus ME. Reversible phosphorylation of the c-terminal domain of RNA polymerase II. J Biol Chem 1996, 27:19009-19012.

Darnell et al. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*. 264 (5164):1415-21 (1994).

164

de Andrés-Aguayo, L., Varas, F., Kallin, E.M., Infante, J.F., Wurst, W., Floss, T., and Graf, T. (2011). Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood *118*, 554-564.

De Sousa Abreu, R., Sanchez-Diaz, P.C., Vogel, C., Burns, S.C., Ko, D., Burton, T.L., Vo, D.T., Chennasamudaram, S., Le, S.-Y., Shapiro, B.A.*, et al.* (2009a). Genomic Analyses of Musashi1 Downstream Targets Show a Strong Association with Cancer-related Processes. Journal of Biological Chemistry *284*, 12125-12135.

Di Giorgio, F. P., Carrasco, M. A., Siao, M. C., Maniatis, T. & Eggan, K. Non–cell autonomous effect of glia on motor neurons in an embryonic stem cell–based ALS model. *Nat Neurosci* 10, 608–614 (2007).

Di Giorgio, F. P., Boulting, G. L., Bobrowicz, S. & Eggan, K. C. Human Embryonic Stem Cell-Derived Motor Neurons Are Sensitive to the Toxic Effect of Glial Cells Carrying an ALS-Causing Mutation. *Stem Cell* 3, 637–648 (2008).

Dong, L. *et al.* Signal regulatory protein α negatively regulates both TLR3 and cytoplasmic pathways in type I interferon induction. *Molecular Immunology* 45, 3025–3035 (2008).

Douville, R., Liu, J., Rothstein, J. & Nath, A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Ann Neurol.* 69, 141–151 (2011).

Du L et al. A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. J Cell Biol 1997, 136: 5-18.

Duncan, A.W., Rattis, F.M., DiMascio, L.N., Congdon, K.L., Pazianos, G., Zhao, C., Yoon, K., Cook, J.M., Willert, K., Gaiano, N.*, et al.* (2005). Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. Nature Immunology *6*, 314-322.

El-Assaad, I. *et al.* Differential expression of TAR DNA-binding protein (TDP-43) in the central nervous system of horses afflicted with equine motor neuron disease (EMND): a preliminary study of a potential pathologic marker. *Vet Res Commun* (2012).doi:10.1007/s11259-012-9533-y

Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069–1075 (2010).

Fiesel, F. C., Weber, S. S., Supper, J., Zell, A. & Kahle, P. J. TDP-43 regulates

global translational yield by splicing of exon junction complex component SKAR. *Nucleic Acids Res* 40, 2668–2682 (2012).

Fitzgerald, K. A. *et al.* IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol* 4, 491–496 (2003).

Flaherty SM et al. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Nat; Acad Sci USA 1997, 94:11893-11898.

Fong and Bentley. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. 2001 (14):1783-95.

Fong and Zhou. Stimulatory effect of splicing factors on transcriptional elongation. Nature. 2001;414(6866):929-33.

Freibaum, B. D., Chitta, R. K., High, A. A. & Taylor, J. P. Global Analysis of TDP-43 Interacting Proteins Reveals Strong Association with RNA Splicing and Translation Machinery. *J. Proteome Res.* 9, 1104–1120 (2010).

Galimberti and Scarpini. Genetics and biology of Alzheimer's disease and frontotemporal lobar degenerationInt J Clin Exp Med. 3(2): 129–143 (2010).

Gardiner, M., Toth, R., Vandermoere, F., Morrice, N. A. & Rouse, J. Identification and characterization of FUS/TLS as a new target of ATM. *Biochem J* 415, 297–307 (2008).

Gulino, A., Marcotullio, L.D., and Screpanti, I. (2010). The multiple functions of Numb. Experimental Cell Research *316*, 900-906.

Gitcho, M.A., Baloh, R.H., Chakraverty, S., Mayo, K., Norton, J.B., Levitch, D., Hatanpaa, K.J., White, C.L. 3rd, Bigio, E.H., Caselli, R. et al. (2008) TDP-43 A315T mutation in familial motor neuron disease. Ann. Neurol., 63, 535–538.

Godbout JP, Johnson RW. Interleukin-6 in the aging brain. J Neuroimmunol 2004;147:141–144.

Goetz CG. Amyotrophic lateral sclerosis: early contributions of Jean-Martin Charcot. Muscle Nerve. Mar;23(3):336-43 (2000).

Gonatas et a. Fragmentation of the Golgi apparatus of motor neurons in amyotrophic lateral sclerosis. Am J Pathol. 140(3): 731–737 (1992).

Gregory R.I. et al. The microprocessor complex mediates the genesis of microRNAs. Nature. 432(7014):235-40 (2004).

Guhaniyogi J and Brewer G. Regulation of mRNA stability in mammalian cells. Gene. 2001 265(1-2):11-23.

Hadland, B.K., Huppert, S.S., Kanungo, J., Xue, Y., Jiang, R., Gridley, T., Conlon, R.A., Cheng, A.M., Kopan, R., and Longmore, G.D. (2004). A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development.

Hallier, M., Lerga, A., Barnache, S., Tavitian, A. & Moreau-Gachelin, F. The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *J Biol Chem* 273, 4838–4842 (1998).

Harel-Sharvit, L. *et al.* RNA Polymerase II Subunits Link Transcription and mRNA Decay to Translation. *Cell* 143, 552–563 (2010).

Hemmi H, Akira S. TLR signaling and functioning in dendritic cells. Chem Immunol Allergy 2005; 86: 120-35.

Hensley et al. Message and protein level elevation of TNFα and TNFα-modulating cytokines in spinal cords of the G93A-SOD1 mouse model for amyotrophic lateral sclerosis. Neurobiol Dis 2003;14:74-80.

Hicks, G. G. *et al.* Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nat Genet* 24, 175–179 (2000).

Honda, K., Yanai, H., Takaoka, A. & Taniguchi, T. Regulation of the type I IFN induction: a current view. *Int Immunol* 17, 1367–1378 (2005).

Hope, K.J., Cellot, S., Ting, S.B., Macrae, T., Mayotte, N., Iscove, N.N., and Sauvageau, G. (2010). An RNAi Screen Identifies Msi2 and Prox1 as Having Opposite Roles in the Regulation of Hematopoietic Stem Cell Activity. Stem Cell *7*, 101-113.

Iborra FJ et al. Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. J Cell Sci 1996; 109:1427-1436.

Imai, T., Tokunaga, A., Yoshida, T., Hashimoto, M., Mikoshiba, K., Weinmaster, G., Nakafuku, M., and Okano, H. (2001). The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. Molecular and Cellular Biology *21*, 3888-3900.

Ito, T., Kwon, H.Y., Zimdahl, B., Congdon, K.L., Blum, J., Lento, W.E., Zhao, C., Lagoo, A., Gerrard, G., Foroni, L*., et al.* (2010). Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature *466*, 765-768.

Jackson AC, et al. Expression of Toll-like receptor 3 in the human cerebellar cortex in rabies, herpes simplex encephalitis, and other neurological diseases. J Neurovirol 2006;12:229–234.

Ji, Z. *et al.* Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol Syst Biol* 7, 1–13 (2011).

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F. et al. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet., 40, 572–574.

Kale, V.P. (2004). Differential activation of MAPK signaling pathways by TGF-beta1 forms the molecular mechanism behind its dose-dependent bidirectional effects on hematopoiesis. Stem cells and development *13*, 27-38.

Kale, V.P., and Vaidya, A.A. (2004). Molecular mechanisms behind the dose-dependent differential activation of MAPK pathways induced by transforming growth factor-beta1 in hematopoietic cells. Stem cells and development *13*, 536-547.

Karlsson, G., Blank, U., Moody, J.L., Ehinger, M., Singbrant, S., Deng, C.-X., and Karlsson, S. (2007). Smad4 is critical for self-renewal of hematopoietic stem cells. The Journal of experimental medicine *204*, 467-474.

Kawahara, H., Imai, T., Imataka, H., Tsujimoto, M., Matsumoto, K., and Okano, H. (2008). Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. J Cell Biol *181*, 639-653.

Kawahara and Mieda-Sato. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. PNAS. 109(9):3347-52 (2010).

Kawai, T. & Akira, S. Antiviral Signaling Through Pattern Recognition Receptors. *Journal of Biochemistry* 141, 137–145 (2006).

Kharas, M.G., Lengner, C.J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., Morgan, K., Tam, W., Paktinat, M., Okabe, R.*, et al.* (2010). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nature medicine *16*, 903-908.

Kim E et al. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J Cell Biol 1997, 136: 19-28.

Kim, S. H., Shanware, N. P., Bowler, M. J. & Tibbetts, R. S. Amyotrophic Lateral Sclerosis-associated Proteins TDP-43 and FUS/TLS Function in a Common Biochemical Complex to Co-regulate HDAC6 mRNA. *Journal of Biological Chemistry* 285, 34097–34105 (2010).

Komarnitsky, P. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14, 2452–2460 (2000).

Kornblihtt, A. R. Multiple links between transcription and splicing. *RNA* 10, 1489–1498 (2004).

Kraemer, B. C. *et al.* Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol* 119, 409–419 (2010).

Krause S et al. Immunodetection of poly(A) binding protein II in the cell nucleus. Exp Cell Res. 1994, 214: 75-82.

Krivtsov, A.V., Twomey, D., Feng, Z., Stubbs, M.C., Wang, Y., Faber, J., Levine, J.E., Wang, J., Hahn, W.C., Gilliland, D.G.*, et al.* (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature *442*, 818-822.

Krug, A. Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood* 103, 1433–1437 (2003).

Kumar, H., Kawai, T. & Akira, S. Pathogen recognition in the innate immune response. *Biochem J* 420, 1–16 (2009).

Kumar, H., Kawai, T. & Akira, S. Pathogen Recognition by the Innate Immune System. *Int Rev Immunol* 30, 16–34 (2011).

Kwiatkowski et al. Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. Science (2009) vol. 323 (5918) pp. 1205-1208

Larsson, J., Blank, U., Klintman, J., Magnusson, M., and Karlsson, S. (2005). Quiescence of hematopoietic stem cells and maintenance of the stem cell pool is not dependent on TGF-β signaling in vivo. Experimental Hematology *33*, 592-596.

LeCam A & Legraverend C. Transcriptional repression, a novel function for 3' untranslated regions. Eur J Biochem. 1995;231(3):620-7.

Lafarga, M., Casafont, I., Bengoechea, R., Tapia, O. & Berciano, M. T. Cajal's contribution to the knowledge of the neuronal cell nucleus. *Chromosoma* 118, 437–443 (2009).

Lagier-Tourenne, C. & Cleveland, D. W. Rethinking ALS: The FUS about TDP-43. *Cell* 136, 1001–1004 (2009).

Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci* **15**, 1488–1497 (2012).

Leigh and Swash. Cytoskeletal pathology in motor neuron diseases. Adv Neurol. 56:115-24 (1991).

Lerga, A. *et al.* Identification of an RNA binding specificity for the potential splicing factor TLS. *J Biol Chem* 276, 6807–6816 (2001).

Letiembre M, et al. Innate immune receptor expression in normal brain aging. Neuroscience 2007;146:248–254.

Lewis JD et al. The influence of 5' and 3' end structures on pre-mRNA metabolism. J Cell Sci Suppl 1995, 19:13-19.

Li, Y. *et al.* ISG56 is a negative-feedback regulator of virus-triggered signaling and cellular antiviral response. *Proc Natl Acad Sci USA* 106, 7945–7950 (2009).

Liu-Yesucevitz, L. *et al.* Tar DNA Binding Protein-43 (TDP-43) Associates with Stress Granules: Analysis of Cultured Cells and Pathological Brain Tissue. *PLoS ONE* 5, e13250 (2010).

Lobsiger, C. S. & Cleveland, D. W. Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease. *Nat Neurosci* 10, 1355–1360 (2007).

Lu T, et al. Gene regulation and DNA damage in the ageing human brain. Nature 2004;429:883–891.

MacGowan DJ, Scelsa SN, Imperato TE, et al. A controlled study of reverse transcriptase in serum and CSF of HIV-negative patients with ALS. Neurology 2007;68:1944–1946.

Macnicol, M.C., Cragle, C.E., and Macnicol, A.M. (2011). Context-dependent regulation of Musashi-mediated mRNA translation and cell cycle regulation. Cell cycle (Georgetown, Tex) *10*, 39-44.

Maillard, I., Koch, U., Dumortier, A., Shestova, O., Xu, L., Sai, H., Pross, S.E., Aster, J.C., Bhandoola, A., Radtke, F.*, et al.* (2008). Canonical notch signaling is dispensable for the maintenance of adult hematopoietic stem cells. Cell Stem Cell *2*, 356-366.

Mancini, S.J.C., Mantei, N., Dumortier, A., Suter, U., MacDonald, H.R., and Radtke, F. (2005). Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. Blood *105*, 2340-2342.

Maniatis T and Reed R. An extensive network of coupling among gene expression machines. Nature. 2002;416(6880):499-506.

Mankouri et al. Optineurin Negatively Regulates the Induction of IFNβ in Response to RNA Virus Infection. PLoS Pathog (2010) vol. 6 (2) pp. e1000778

Matsumoto, A., Takeishi, S., Kanie, T., Susaki, E., Onoyama, I., Tateishi, Y., Nakayama, K., and Nakayama, K.I. (2011). p57 is required for quiescence and maintenance of adult hematopoietic stem cells. Cell Stem Cell *9*, 262-271.

Maruyama et al. Mutations of optineurin in amyotrophic lateral sclerosis. Nature (2010) pp. 1-5

McCormick AL, Brown RH Jr, Cudkowicz ME, et al. Quantification of reverse transcriptase in ALS and elimination of a novel retroviral candidate. Neurology 2008;70:278–283.

McCracken S et al. The c-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 1997, 385:357-361.

McCracken S et al. 5'-capping enzymed are targeting to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev 1997, 22:3306-3318.

Mckee et al. TDP-43 Proteinopathy and Motor Neuron Disease in Chronic Traumatic Encephalopathy. Journal of Neuropathology and Experimental Neurology (2010) vol. 69 (9) pp. 918-929

Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* 449, 819–826 (2007).

Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 1–9 (2012).doi:10.1038/nbt.2137

Misteli and Spector. Protein phosphorylation and the nuclear organization of pre-mRNA splicing. Trends Cell Biol 1997, 7:135-138.

Monneron and Bernhard. Fine structural organization of the interphase nucleus in some mammalian cells. J Ultrastruct Res.27(3):266-88 (1969).

Moore, K.A., Ema, H., and Lemischka, I.R. (1997). In vitro maintenance of highly purified, transplantable hematopoietic stem cells. Blood *89*, 4337-4347.

Moreau, C. *et al.* Elevated IL-6 and TNF- levels in patients with ALS: Inflammation or hypoxia? *Neurology* 65, 1958–1960 (2005).

Mortillaro MJ et al. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proc Natl Acad Sci USA 1996, 93:8253-57.

Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560 (2007).

Muller et al. Functional role of type I and type II interferons in antiviral defense. *Science* 264 (5167): 1918-21 (1994).

Murphy K, Travers P, Walport M, Janeway C. Janeway's Immunobiology, 7th Edition. New York: Garland Science; 2008.

Nakashima-Yasuda, H. *et al.* Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathol* 114, 221–229 (2007).

Neugebauer, K. M. On the importance of being co-transcriptional. *Journal of Cell Science* 115, 3865–3871 (2002).

Neumann et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314(5796):130-3 (2006).

Niwa, M. & Berget, S. M. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* 5, 2086–2095 (1991).

Okabe, M., Imai, T., Kurusu, M., Hiromi, Y., and Okano, H. (2001). Translational repression determines a neuronal potential in Drosophila asymmetric cell division. Nature *411*, 94-98.

Okano, H., Kawahara, H., Toriya, M., Nakao, K., Shibata, S., and Imai, T. (2005). Function of RNA-binding protein Musashi-1 in stem cells. Exp Cell Res *306*, 349-356.

Okun et al. Toll-like receptors in neurodegeneration. Brain Research Reviews (2009) vol. 59 (2) pp. 278-292.

Ou S. H. et al. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type I TAR DNA sequence motifs. J Virol. 69:6; 3584 (1995).

Papadimitriou, D. *et al.* Inflammation in ALS and SMA: Sorting out the good from the evil. *Neurobiol. Dis.* 37, 493–502 (2010).

Passegue, E., Wagers, A.J., Giuriato, S., Anderson, W.C., and Weissman, I.L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J Exp Med *202*, 1599-1611.

Pereira, J.K.N., Traina, F., Machado-Neto, J.A., Duarte, A.d.S.S., Lopes, M.R., Saad, S.T.O., and Favaro, P. (2012). Distinct expression profiles of MSI2 and NUMB genes in myelodysplastic syndromes and acute myeloid leukemia patients. Leukemia research.

Poloni et al. Circulating levels of tumour necrosis factor and its soluble receptors are increased in the blood of patients with amyotrophic lateral sclerosis. Neurosci Let (2000) vol. 287 pp. 211-14

Polymenidou et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci (2011) vol. 14 (4) pp. 459-468

Proudfoot NJ et al. Integrating mRNA processing with transcription. Cell. 2002 108(4):501-12.

Rahl, P. B. *et al.* c-Myc Regulates Transcriptional Pause Release. *Cell* 141, 432–445 (2010).

Raj and Pitha. 65-kDa protein binds to destabilizing sequences in the IFNb mRNA coding and 3' UTR. *FASEB J.* 7(8):702-10 (1993).

Ransohoff, R. M. & Perry, V. H. Microglial Physiology: Unique Stimuli, Specialized Responses. *Annu Rev Immunol* 27, 119–145 (2009).

Rosen et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nat (1993) vol. 4 (6415) pp. 59-62.

Saitoh, T. *et al.* A20 is a negative regulator of IFN regulatory factor 3 signaling. *J Immunol* 174, 1507–1512 (2005).

Sakaguchi et al. Optineurin with amyotrophic lateral sclerosis-related mutations abrogates inhibition of interferon regulatory factor-3 activation. Neuroscience Letters (2011) pp. 1-3

Sakakibara, S., Imai, T., Hamaguchi, K., Okabe, M., Aruga, J., Nakajima, K., Yasutomi, D., Nagata, T., Kurihara, Y., Uesugi, S*., et al.* (1996). Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. Dev Biol *176*, 230-242.

Sandberg R et al. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science. 2008, 320(5883):1643-7.

Sasaki, S., Takeda, T., Shibata, N. & Kobayashi, M. Alterations in subcellular localization of TDP-43 immunoreactivity in the anterior horns in sporadic amyotrophic lateral sclerosis. *Neuroscience Letters* 478, 72–76 (2010).

Satoh, T. *et al.* From the Cover: LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences* 107, 1512–1517 (2010).

Schul W et al. The RNA 3' cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. EMBO J 1996, 15:2883-2892.

Schwab C et al. Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease. J Neuropathol Exp Neurol. 2008;67(12):1159-65.

Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16, 990–995 (2009).

Seifert, U. *et al.* Immunoproteasomes Preserve Protein Homeostasis upon Interferon-Induced Oxidative Stress. *Cell* 142, 613–624 (2031).

Sephton C. F. et al. Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. J Biol Chem. 286(2):1204-15 (2011).

Shan X et al. Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. PROC NATL ACAD SCI USA. 107(37):16325-30 (2010).

Shapira et al. A Physical and Regulatory Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection. Cell (2009) vol. 139 (7) pp. 1255-1267

Shi, M. *et al.* TRIM30 alpha negatively regulates TLR-mediated NF-kappa B activation by targeting TAB2 and TAB3 for degradation. *Nat Immunol* 9, 369–377 (2008).

Sitnicka, E., Ruscetti, F.W., Priestley, G.V., Wolf, N.S., and Bartelmez, S.H. (1996). Transforming growth factor beta 1 directly and reversibly inhibits the initial cell divisions
of long-term repopulating hematopoietic stem cells. Blood *88*, 82-88.

Sreedharan et al. TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. Science (2008) vol. 319 (5870) pp. 1668-1672

Stallings, N. R., Puttaparthi, K., Luther, C. M., Burns, D. K. & Elliott, J. L. Progressive motor weakness in transgenic mice expressing human TDP-43. *Neurobiol. Dis.* 40, 404–414 (2010).

Steele AJ, Al-Chalabi A, Ferrante K, et al. Detection of serum reverse transcriptase activity in patients with ALS and unaffected blood relatives. Neurology 2005;64:454–458.

Stoecklin, G. & Anderson, P. *Advances in Immunology*. 89, 1–37 (Elsevier: 2006).

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Swarup et al. Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor kB-mediated pathogenic pathways. JEM (2011).

Swarup, V. & Julien, J.-P. ALS pathogenesis: recent insights from genetics and mouse models. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 363–369 (2011).

Taghon, T., Michor, F., Levine, R.L., and Aifantis, I. (2011). A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. Nature.

Tan, A. Y. & Manley, J. L. TLS Inhibits RNA Polymerase III Transcription. *Mol Cell Biol* 30, 186–196 (2009).

Thanos, D. & Maniatis, T. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100 (1995).

Tsai, K. J. *et al.* Elevated expression of TDP-43 in the forebrain of mice is sufficient to cause neurological and pathological phenotypes mimicking FTLD-U. *Journal of Experimental Medicine* 207, 1661–1673 (2010).

Tsao, W. *et al.* Rodent models of TDP-43: Recent advances. *Brain Research* 1462, 26–39 (2012).

Udan, M. & Baloh, R. H. Implications of the prion-related Q/N domains in TDP-43 and FUS. *prion* 5, 1–5 (2011).

Valentine and Hart. Misfolded CuZnSOD and amyotrophic lateral sclerosis. Proc Natl Acad Sci USA. 100(7): 3617–3622 (2003).

Van Deerlin, V.M., Leverenz, J.B., Bekris, L.M., Bird, T.D., Yuan, W., Elman, L.B., Clay, D., Wood, E.M., Chen-Plotkin, A.S., Martinez-Lage, M. et al. (2008) TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. Lancet Neurol., 7, 409–416).

Vance, C. *et al.* Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science* 323, 1208–1211 (2009).

Verma et al. ALS syndrome in patients with HIV-1 infection. J Neurol Sci. 240(1-2):59-64 (2006).

Vincent M et al. The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res 1996, 24:4649-4652.

Vinciguerra, P. & Stutz, F. mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol* 16, 285–292 (2004).

Volkening, K., Leystra-Lantz, C., Yang, W., Jaffee, H. & Strong, M. J. Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). *Brain Research* 1305, 168–182 (2009).

Wang, H.-Y., Wang, I.-F., Bose, J. & Shen, C. K. J. Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics* 83, 130–139 (2004).

Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454, 126–130 (2008).

Wang et al. Activation of interferon signaling pathways in spinal cord astrocytes from an ALS mouse model. Glia (2011) vol. 59 (6) pp. 946-958.

Wansink DG et al. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. J Cell Biol. 1993, 122:282-293.

Wathalet et al. Virus infection induces the assembly of coordinately activated transcription factors on the IFNb enhancer in vivo. *Molecular Cell*, Vol. 1, 507–518(1998)

Wegorzewska et al. TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. Proc Natl Acad Sci USA (2009)

Weydt, P., Yuen, E. C., Ransom, B. R. & M ller, T. Increased cytotoxic potential of microglia from ALS-transgenic mice. *Glia* 48, 179–182 (2004).

Wils, H. *et al.* TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences* 107, 3858–3863 (2010).

Wilkins, C. & Gale, M., Jr Recognition of viruses by cytoplasmic sensors. *Current Opinion in Immunology* 22, 41–47 (2010).

Wu, L.-S. *et al.* TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *genesis* n/a–n/adoi:10.1002/dvg.20584

Wu, M., Kwon, H.Y., Rattis, F., Blum, J., Zhao, C., Ashkenazi, R., Jackson, T.L., Gaiano, N., Oliver, T., and Reya, T. (2007a). Imaging hematopoietic precursor division in real time. Cell Stem Cell *1*, 541-554.

Wu, M., Kwon, H.Y., Rattis, F., Blum, J., Zhao, C., Ashkenazi, R., Jackson, T.L., Gaiano, N., Oliver, T., and Reya, T. (2007b). Imaging hematopoietic precursor division in real time. Stem Cell *1*, 541-554.

Xu, Y. F. *et al.* Wild-Type Human TDP-43 Expression Causes TDP-43 Phosphorylation, Mitochondrial Aggregation, Motor Deficits, and Early Mortality in Transgenic Mice. *Journal of Neuroscience* 30, 10851–10859 (2010).

Yamanaka K et al. Mutant SOD1 in cell types other than motor neurons and oligodendrocytes accelerates onset of disease in ALS mice. 27;105(21):7594-9 (2008).

Yamazaki, S., Ema, H., Karlsson, G., Yamaguchi, T., Miyoshi, H., Shioda, S., Taketo, M.M., Karlsson, S., Iwama, A., and Nakauchi, H. (2011). Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. Cell *147*, 1146-1158.

Ye SM, Johnson RW. An age-related decline in interleukin-10 may contribute to the increased expression of interleukin-6 in brain of aged mice. Neuroimmunomodulation 2001;9:183–192.

Yokoseki, A., Shiga, A., Tan, C.F., Tagawa, A., Kaneko, H., Koyama, A., Eguchi, H., Tsujino, A., Ikeuchi, T., Kakita, A. et al. (2008) TDP-43 mutation in familial amyotrophic lateral sclerosis. Ann. Neurol., 63, 538–542.

Yue Z et al. Mammalian capping enzymed complements mutant Saccharomyces cerecisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. Proc Natl Acad Sci 1997, 94: 12898-12903.

Yuryev A et al. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc Natl Acad Sci 1996, 93: 6975-6980.

Zhang et al. Gene expression profiling in peripheral blood mononuclear cells from patients with sporadic amyotrophic lateral sclerosis (sALS). Journal of neuroimmunology (2011) vol. 230 (1-2) pp. 114-123

Zhao J et al. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. <u>Microbiol Mol Biol Rev.</u> 1999;63(2):405-45.

Zhou et al. Transgenic Rat Model of Neurodegeneration Caused by Mutation in the TDP Gene. PLoS Genet (2010) vol. 6 (3) pp. e1000887

Zou, P., Yoshihara, H., Hosokawa, K., Tai, I., Shinmyozu, K., Tsukahara, F., Maru, Y., Nakayama, K., Nakayama, K.I., and Suda, T. (2011). p57(Kip2) and p27(Kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. Cell Stem Cell *9*, 247-261.

Supplementary Figure 1:
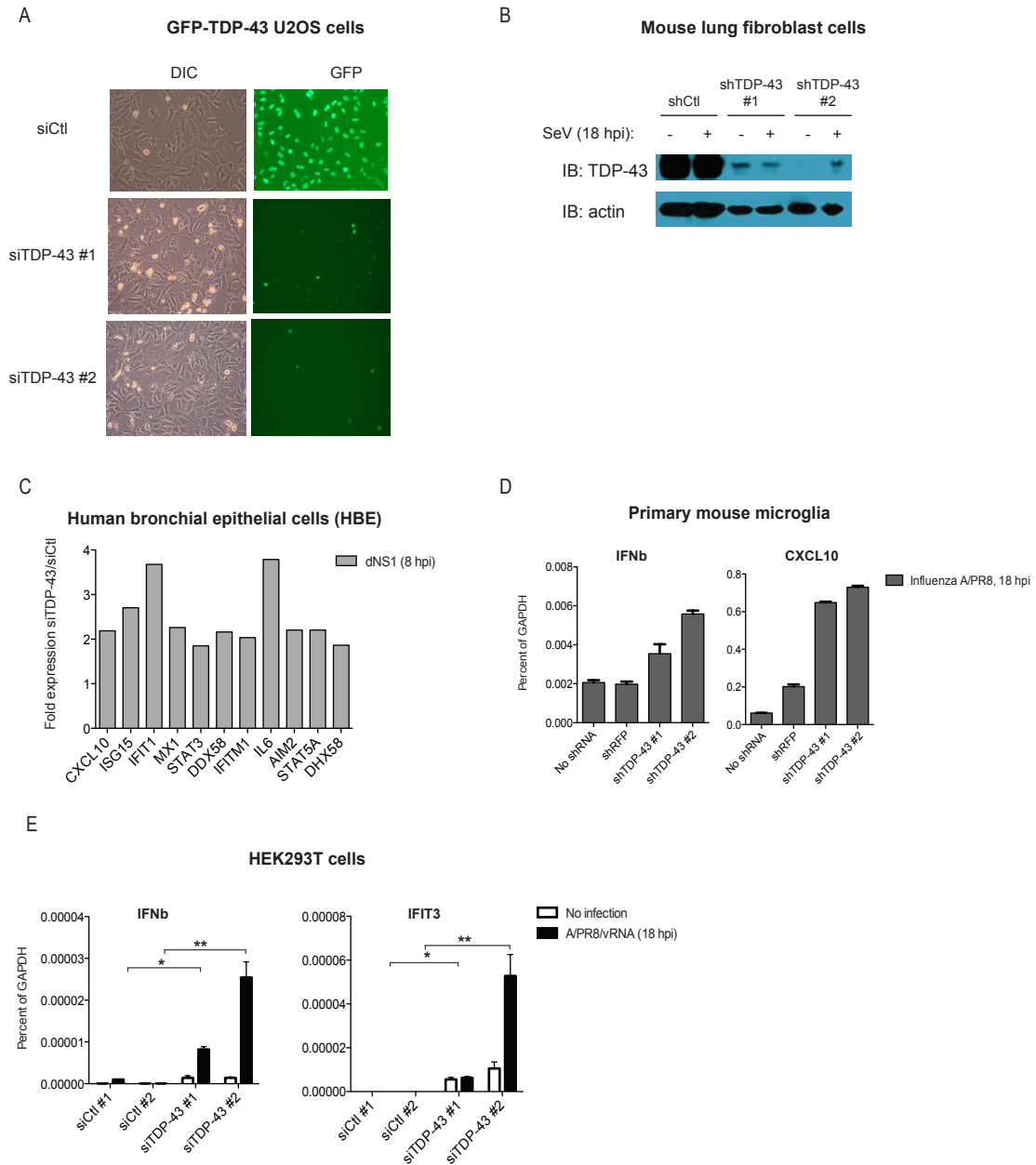
S. Figure 1.1: TDP-43 negatively regulates antiviral gene expression.

(A) Fluorescent images of GFP-TDP-43 protein in TDP-43-depleted cells. U2OS cells expressing GFP-TDP-43 were transfected with a control siRNA or two independent siRNA sequences targeting TDP-43 and fluorescent images were taken 72 hrs post transfection.
(B) Western blot for TDP-43 protein in TDP-43-depleted cells. Mouse lung fibroblasts were were transfected with a control siRNA or two independent siRNA sequences targeting TDP-43 and **S.**

Figure 1.1 (Continued): TDP-43 protein levels in cell lysates were assessed in resting and Sendai virus (SeV)-stimulated cells 72 hours later.

(C) mRNA levels for interferon-stimulated genes in TDP-43-depleted human lung cells. Primary human bronchial epithelial (HBE) cells were transfected with siRNA targeting a control sequence or TDP-43. HBE were infected with infected with Influenza A/dNS1 (A/dNS1) (MOI 2) for 8 hrs and the relative expression of antiviral genes was assessed by Nanostring nCounter.

(D) mRNA levels (qPCR) for IFNb and CXCL10 in primary mouse microglia. Primary mouse microglia were infected with shRNA targeting a control sequence (shRFP) or with two independent shRNA targeting TDP-43. Following selection with puromycin, cells were infected with Influenza A/PR8 (A/PR8) (MOI 2) for 18 hrs and gene expression was assessed by qPCR.

(E) mRNA levels (qPCR) for IFNb and IFIT3 in TDP-43-depleted HEK293T cells. HEK293T cells were transfected with siRNA targeting control sequences or TDP-43. Cells were transfected with Influenza A/PR8 viral RNA (A/PR8/vRNA) for 18 hrs and gene expression was assessed by qPCR. *p<0.05, **p<0.01.
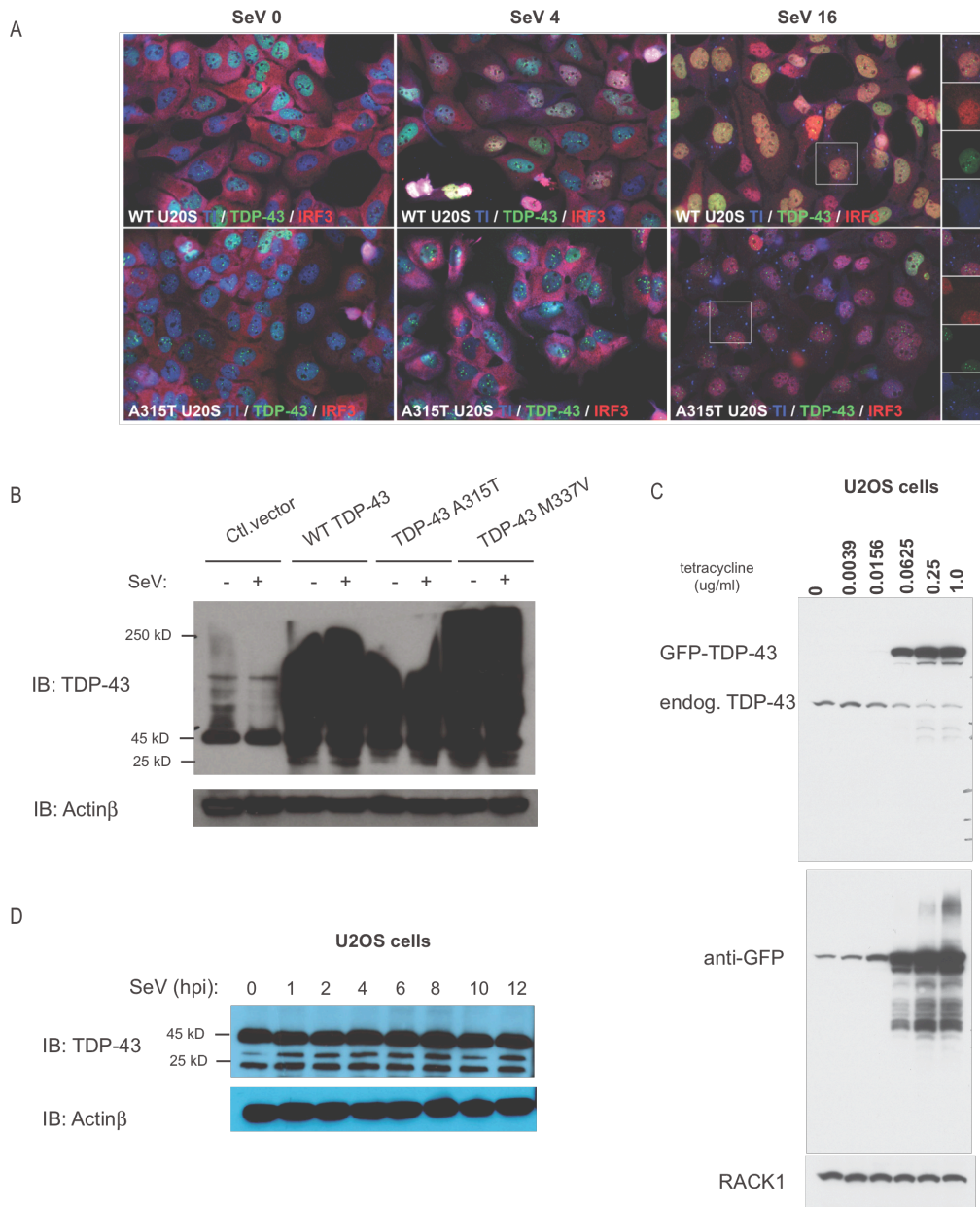
**S. Figure 1.2: TDP-43 protein levels are carefully regulated.**

(A) Fluorescent images of TDP-43 localization in GFP-TDP-43 overexpressing cells. U2OS cells stably expressing GFP-WTTDP-43 (top panel) or GFP-TDP-43A315T (bottom panel) under a tetracycline-inducible promoter were treated with tetracycline for 24 hrs. Then, cells were infected with SeV (MOI 5) over a time course (0, 4, 16 hr). Cells were fixed and stained with antibodies targeting GFP (green), nuclei (blue), or IRF3 (red) and imaged.
(B) Western blot for TDP-43 protein in cells overexpressing TDP-43. HEK293T cells were transfected with plasmids encoding an empty vector, WT TDP-43, TDP-43A315T, or TDP- **S.**

Figure 1.2 (Continued): 43M337V. 24 hrs later, cells were infected with SeV (MOI 5) for 18 hrs and TDP-43 protein was measured from cell lysates by Western blotting.
(C) Western blot for TDP-43 protein in cells overexpressing TDP-43. U2OS cells stably expressing GFP-WTTDP-43 were treated with tetracycline for 24 hrs at several doses (0, 0.0039, 0.0156, 0.0625, 0.25, and 1 ug/ml) and TDP-43 protein levels in cell lysates were assessed by Western blotting.
(D) Western blot for TDP-43 protein in U2OS cells infected with SeV. U2OS cells were infected with SeV (MOI 5) over a time course (0,1,2,4,6,8,10,12 hr) and TDP-43 protein levels in cell lysates was assessed by Western blotting.
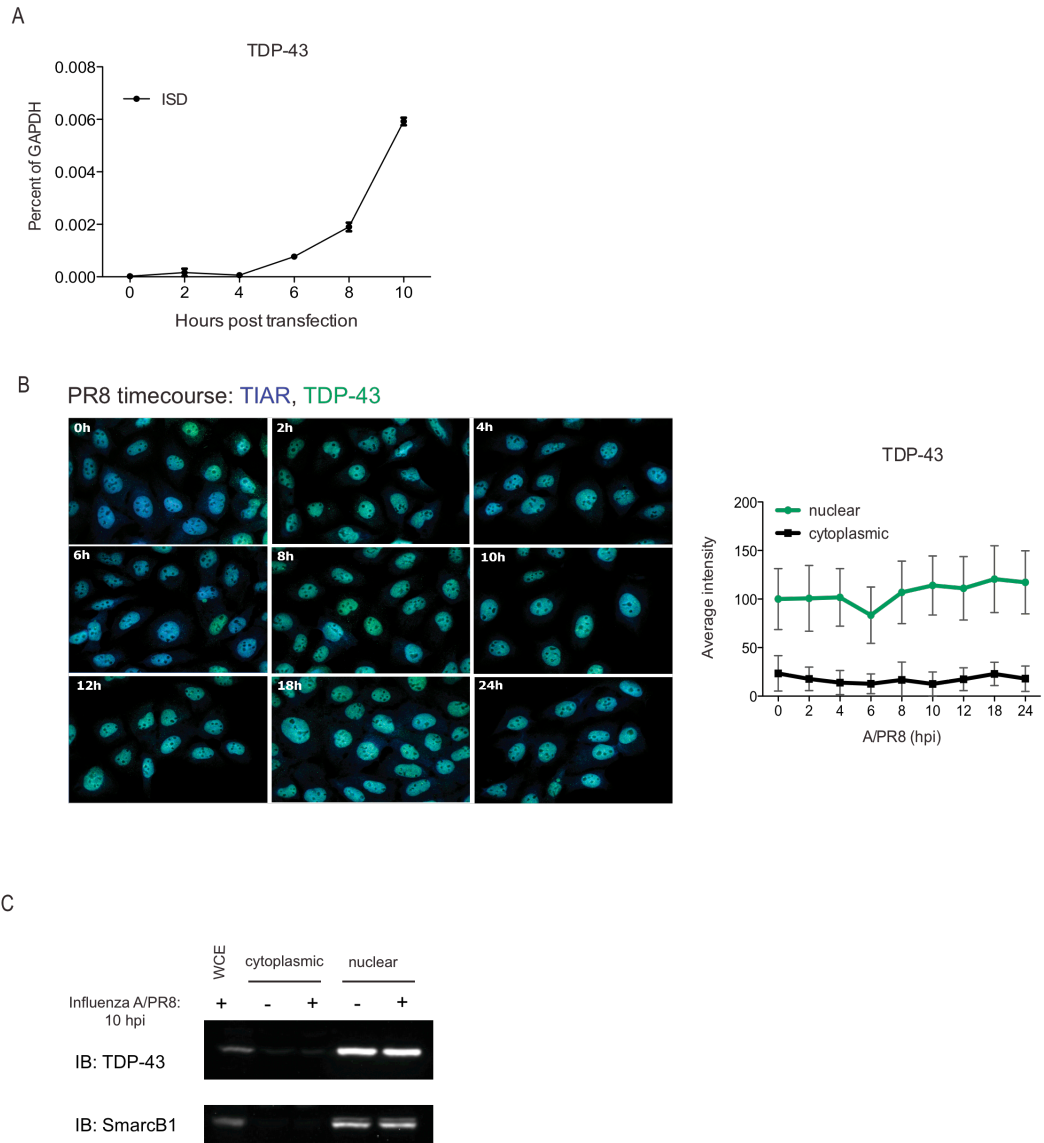
S. Figure 1.3: TDP-43 remains nuclear in response to viral sensing.

(A) mRNA levels (qPCR) of TDP-43 in response to DNA sensing. U2OS cells were transfected with immunostimulatory DNA (ISD)(120 ng/ml) over a time course (0,2,4,6,8,10 hr) and TDP-43 transcript levels were assessed by qPCR.
**(B)** Fluorescent imaging showing nuclear versus cytosolic localization of TDP-43 protein in response to A/PR8 infection in human cells. U2OS cells were infected with A/PR8 (MOI 5) over a time course (0,2,4,6,8,10,12,18 hr). Cells were fixed and stained with fluorescent antibodies targeting TDP-43 (green) or TIAR (blue). Quantification of cytoplasmic versus nuclear TDP-43 fluorescent signal (right).
(C) Western blot showing nuclear versus cytoplasmic localization of TDP-43 in response to A/PR8 infection in mouse cells. MLF were infected with A/PR8 (MOI 5) for 10 hrs. Cytoplasmic and nuclear fractionation was performed and TDP-43 protein levels were assessed by Western blot.

184
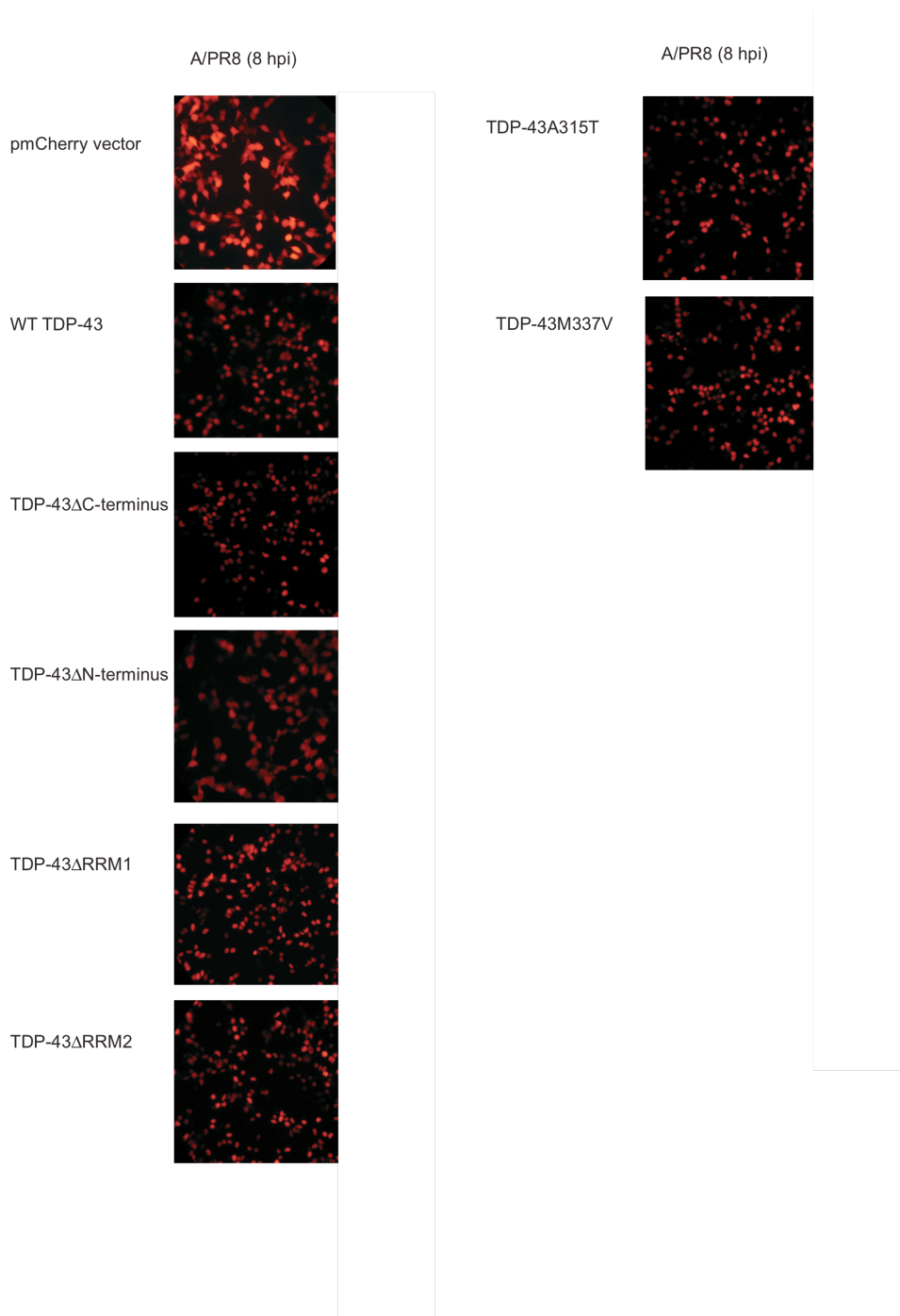
S. Figure 1.4: Localization of TDP-43 domain mutant expression.

(A) Fluorescent imaging of Cherry-tagged TDP-43 protein in HEK293T cells in response to A/PR8 infection. HEK293T cells were transiently transfected with plasmids encoding an empty vector (pmCherry vector), or pmCherry fused to WTTDP-43, TDP-43A315T, TDP-43M337V, TDP-43ΔC-terminus, TDP-43ΔN-terminus, TDP-43ΔRRM1, or TDP-43ΔRRM2. Cells were infected with A/PR8 for 8 hrs, then fixed and imaged.

A   MEF
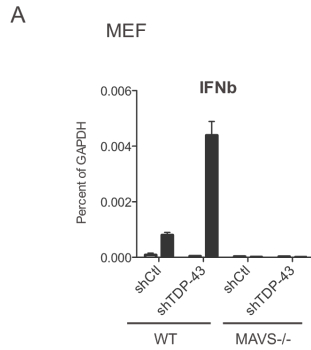
S. Figure 1.5: Intact antiviral signaling is required for the TDP-43-mediated antiviral phenotype.

(A) mRNA levels (qPCR) for IFNb transcript in MAVS-/- cells. MAVS+/+ or MAVS-/- mouse embryonic fibroblasts (MEF) were infected with shRNA targeting a control sequence or TDP-43. After selection with puromycin, cells were infected with SeV (MOI 2) for 18 hrs. IFNb transcript levels were assessed by qPCR.
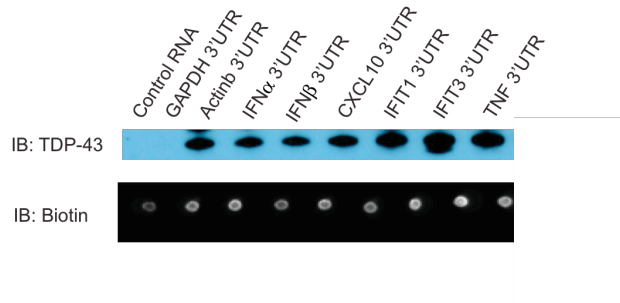
S. Figure 1.6: TDP-43 associates with the 3' UTR of innate immune transcripts in vitro.

(A) Western blot for crosslinked RNA pull down of innate immune 3' UTR. HEK293T cells were infected with SeV (MOI 5) for 18 hrs. Lysates were prepared and biotinylated 3'UTR sequences corresponding to a control sequence (Control RNA, GAPDH) or TDP-43-target transcripts (actinb, IFNa, IFNb, CXCL10, IFIT1, IFIT3, TNF) were incubated with the lysates for 2 hrs. The protein:RNA interactions were immobilized by UV crosslinking and the lysates were RNase A digested. Proteins were resolved on an SDS-PAGE gel and TDP-43 protein was probed with a specific antibody.

186
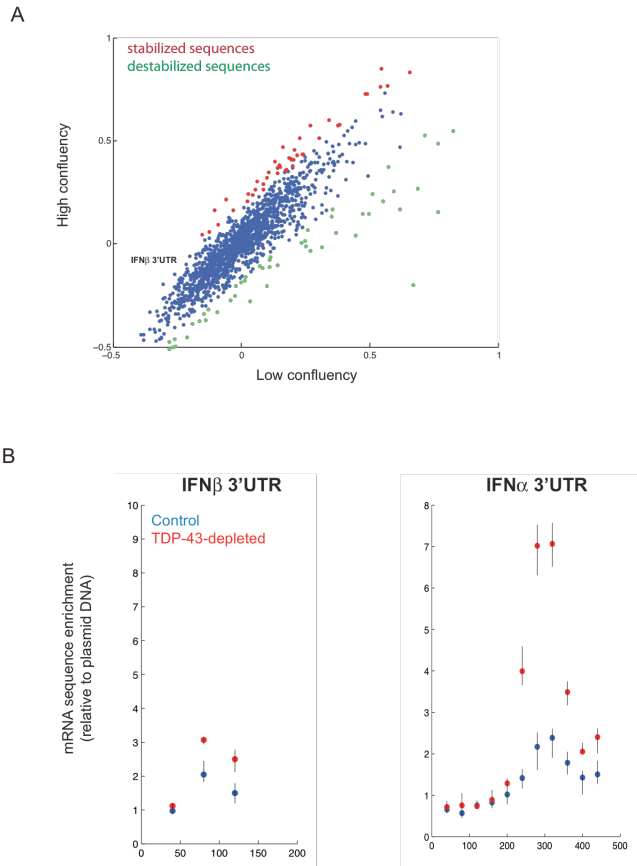
S. Figure 1.7: Perturbation of TDP-43 stabilizes polyadenylation sequences in the 3' UTR of IFNb and IFNa.

(A) Distribution of massively parallel reporter assay (MPRA) tagged sequences in cells grown at high versus low confluency. HEK293T cells at high versus low confluency were transfected with our MPRA plasmid pool. Relative enrichment for stabilized (red) and destabilized (green) tile tags is shown.
(B) MPRA mRNA sequence enrichment for IFNb and IFNa 3' UTR. HEK293T cells were tranfected with siRNA targeting a control sequence or TDP-43, then transfected with our MPRA plasmid pool. Relatively stabilized tile tags (red dot) are shown.

Supplementary Table 2:

**MPRA
gene list:**

Gene:	Protein name:	RefSeq ID	MPRA peaks?
ACTB	ACTB	NM_001101.3	
ADAR	Double-stranded RNA-specific adenosine deaminase	NM_001111.4	Y
ATF3	ATF3	NM_001040619.1	Y
Axl	axl receptor tyrosine kinase	NM_021913.3	Y
bst2	tetherin	NM_004335.2	
C3	complement component 3	NM_000064.2	
CCND1	cyclin D1	NM_053056.2	
cdk6	cyclin dept kinase 6	NM_001145306.1	Y
CNN1	calponin	NM_001299.4	Y
CXCL10	CXCL10	NM_001565.3	Y
DDX58	RIG-I	NM_014314.3	Y
Eif2ak2	Eif2ak2	NM_002759.2	Y
Fus	Fus/TLS	NM_004960.3	Y
GAPDH	GAPDH	NM_002046.3	
GRN	granulin	NM_002087.2	
HDAC1	HDAC1	NM_004964.2	Y
HDAC4	HDAC4	NM_006037.3	Y
HDAC5	HDAC5	NM_005474.4	Y
HDAC6	HDAC6	NM_006044.2	
HDAC8	HDAC8	NM_018486.2	
IFI16	IFI16	NM_005531	Y
Ifih1	interferon induced helicase c	NM_022168.2	Y
IFIT1	IFIT1	NM_001548.3	Y
IFIT3	IFIT3	NM_001549.4	Y
IFITM3	IFITM3	NM_021034.2	
IFNA2	interferon alpha	NM_000605.3	Y
IFNB1	interferon beta	NM_002176.2	Y
IFNg	interferon gamma	NM_000619.2	Y
IL-6	IL-6	NM_000600.3	Y
IL12B	IL-12p40	NM_002187	Y
IRF3	IRF3	NM_001197123.1	
IRF7	IRF7	NM_004031.2	
IRF9	IRF9	NM_006084.4	
IRGM	immunity-related GTPase family, M	NM_001145805.1	
ISG 15	ISG 15	NM_005101.3	
lrrc47	lrrc47	NM_020710.2	
MAVS	MAVS	NM_020746.4	Y
MX1	MX1	NM_001144925.1	
NEFL	neurofilament	NM_006158.3	
OAS1	OAS1	NM_016816.2	

188

OASL	OASL	NM_003733.2	
Plk1	polo like kinase 1	NM_005030.3	
SHISA5	SHISA5	NM_025858.2	Y
SOD1	SOD1	NM_000454.4	
STAT1	STAT1	NM_009283.3	Y
STAT2	STAT2	NM_005419.3	
TARDBP	TDP-43	NM_007375.3	Y
TNF	TNF	NM_000594.2	Y
VIM	vimentin	NM_003380.3	Y
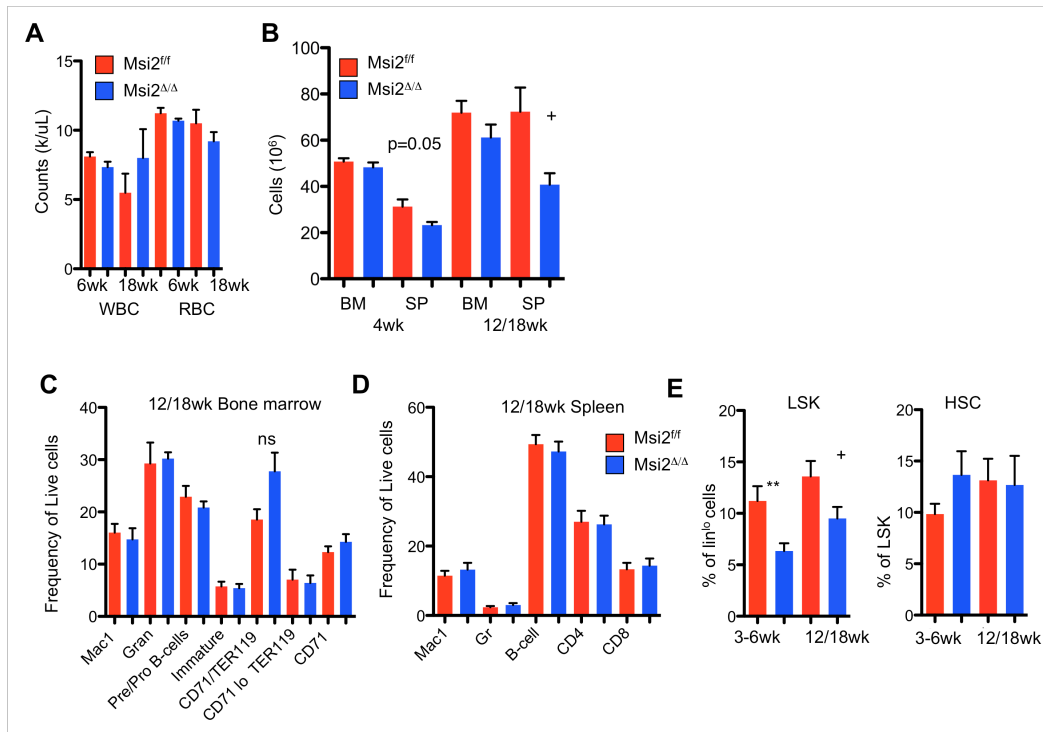
S. Figure 3.1: Characterization of Msi2 deleted hematopoietic cells. (A) Peripheral blood analysis from control and conditional Msi2 deleted mice. White blood cells (WBC) and red blood cells (RBC) at indicated time points post pIpC, (n=6-7) are shown. (B) Overall cell counts in mice from (A) in the bone marrow (BM) and spleen (SP), (4wk; n=4, 12/18wk; n=9,10). (C) Combined flow cytometric analysis of the frequency of different mature hematopoietic populations from indicated live cells from bone marrow at 12wks or 18wks, Mac1$^+$ Gr1$^-$ (Mac1), Mac1$^+$ Gr1$^+$ (Gran), IgM$^-$ B220$^+$ (Immature), CD71$^+$ Ter119$^+$ (CD71/Ter119), CD71$^-$ Ter119$^+$ (CD71lo TER119), CD71$^+$ Ter119$^-$ (CD71), (n=6-9). (D) Flow cytometric analysis as in (C) from the spleen, IgM$^+$ B220$^+$ (B-cell), **S.**

Figure 3.1 (Continued): CD4$^+$ (CD4 T-cells), CD8$^+$ (CD8, T-cells), (n=6-9). (E)

Frequencies from bone marrow of indicated mice injected with pIpC at indicated

time points (LSK; 4wk, n=4, 12/18wk; 6-8), (HSC; 4wk, n=4, 12/18wk; 8-10).

Data error bars are SEM and (ns not significant, +p<0.05, *p<0.01, **p<0.001).


Supplemental Figure 2
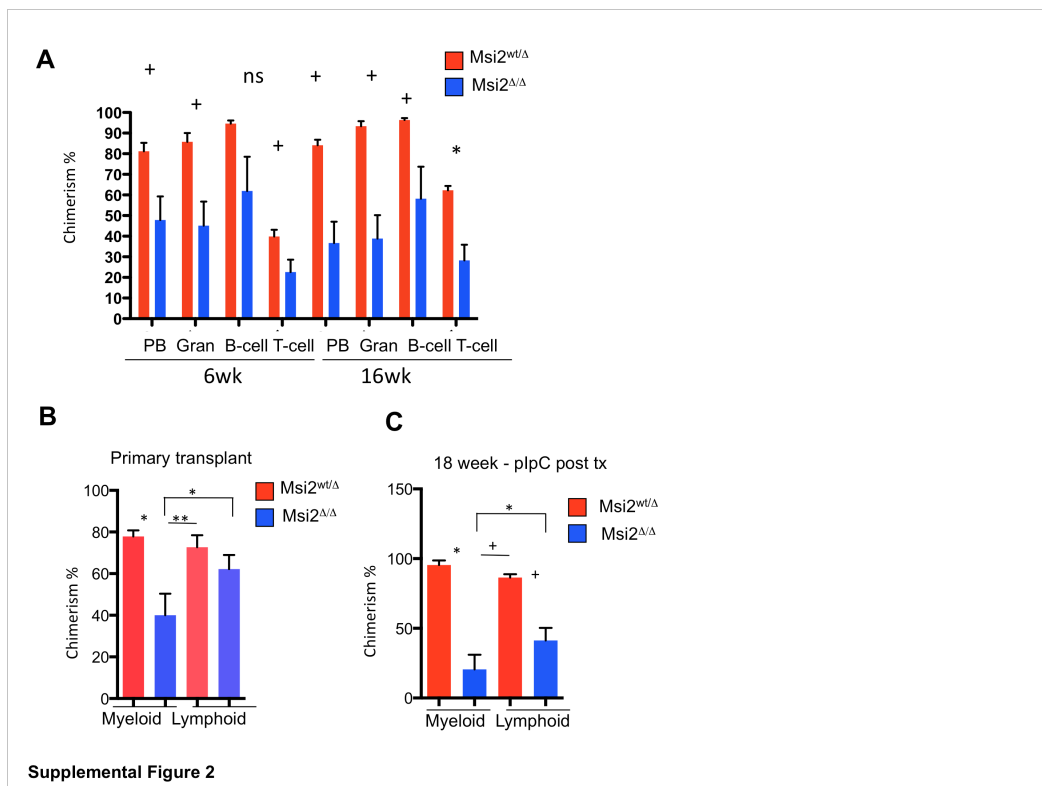
S. Figure 3.2: Msi2 deleted cells have reduced repopulating activity in the

myeloid compartment. (A) Chimerism analysis of the mature populations in

peripheral blood of primary transplanted animals at indicated time points (n=4,5).

Statistics--. (B) Chimerism of myeloid and lymphoid lineages in spleen of primary

transplanted mice shown in Figure 2A. (C) Chimerism of myeloid and lymphoid

lineages in bone marrows of transplanted mice that were allowed to engraft and
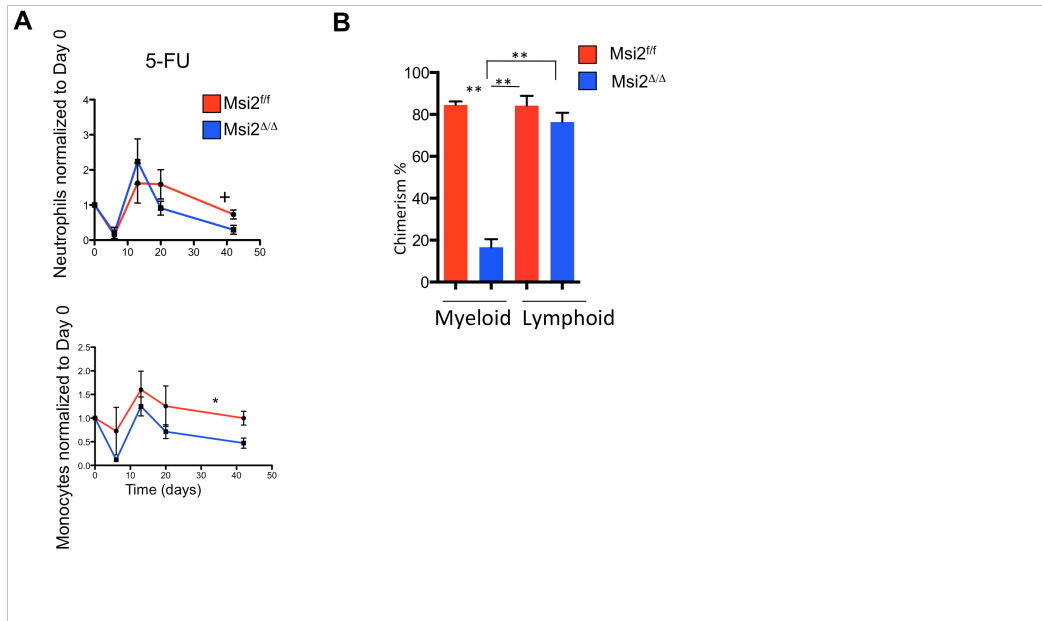
S. Figure 3.3: Reduced myeloid recovery after replicative stress in Msi2

deleted animals. (A) Blood analysis of non-competitively transplanted animals.

Measurements of neutrophils (upper panel) and monocytes (lower panel) at

indicated time points were normalized back to the same animal before 5-

Flourouracil (5-FU), (n=5, 8) and after 6-weeks post pIpC injections (see also

Figure 3B-C). (Data error bars are SEM and (ns not significant, +p<0.05,

*p<0.01). (B) Chimerism analysis of myeloid and lymphoid lineages in peripheral

blood of mice shown in Figure 3B. Data error bars are SEM and (**p<0.001).
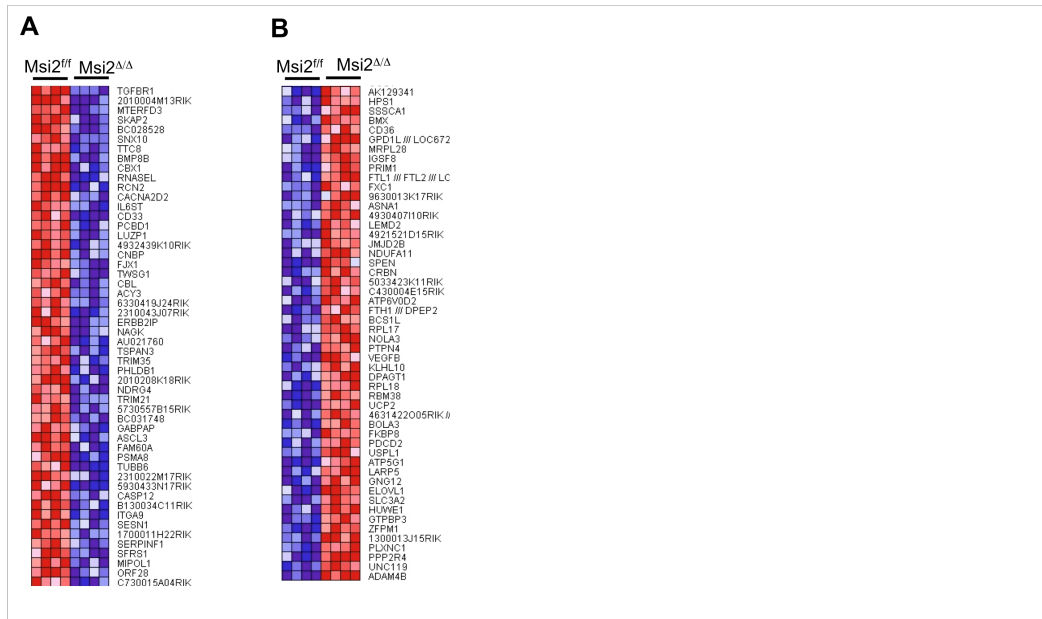
S. Figure 3.4: Differentially expressed genes in Msi2 deleted Hematopoietic stem and progenitor cells. (A) Microarray heat map of top 40 differentially downregulated expressed genes based on ranked t-test values (n=4), (4 weeks post pIpC) sorted LSKs. (B) Same as in (A) except selected heat map is for genes unregulated in Msi2 deleted LSKs.
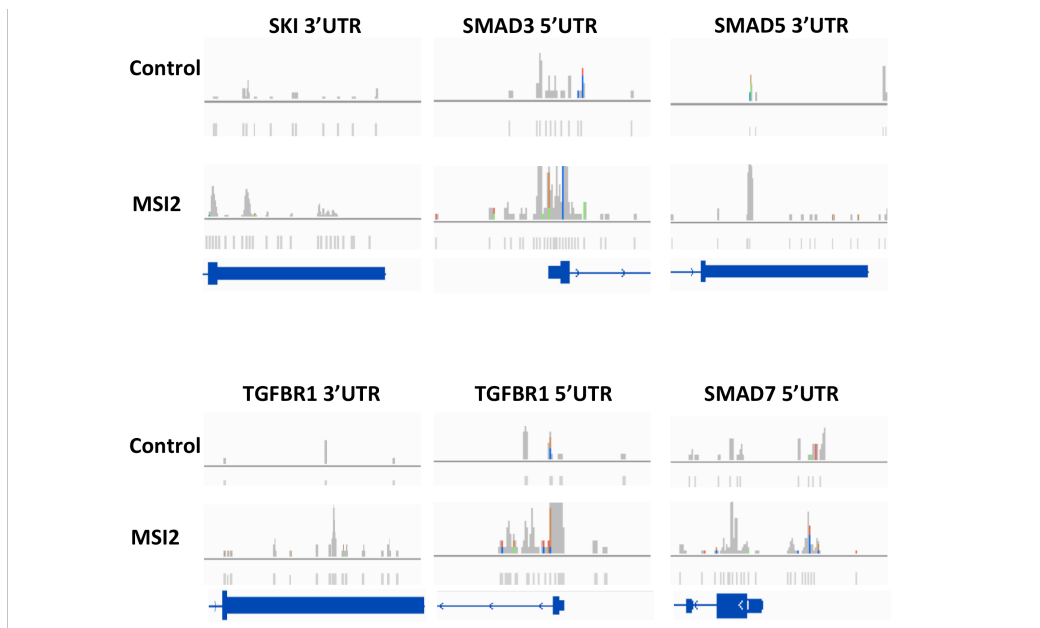
S. Figure 3.5: MSI2 binds to multiple components of the TGFb signaling pathway. Integrative Genomics Viewer (IGV) mapped sequence reads from the HITS-CLIP experiment with empty vector infected K562s (Control) and FLAG-MSI2 (MSI2) transduced cells. Peak reads are in the same scale and comparable between samples.
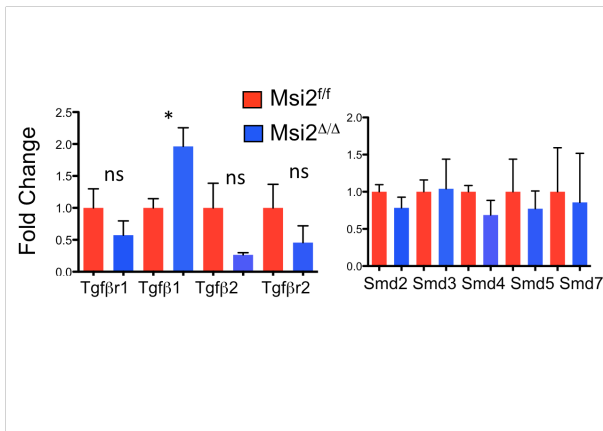
S. Figure 3.6: Comparable RNA expression of Tgfb pathway components in Msi2 deleted hematopoietic stem and progenitor cells. LSK cells analyzed for the expression of indicated Tgfb pathway member i.e Smads (Smd) normalized to Gapdh, (n=4-6). Data error bars are SEM and (ns not significant, *p<0.01).