

# RNA-BINDING PROTEINS DIRECT MYOGENIC CELL FATE DECISIONS

**Authors:** Joshua R. Wheeler<sup>1,2,12,13,15</sup>, Oscar N. Whitney<sup>3,10,15</sup>, Thomas O. Vogler<sup>2,3</sup>, Eric D. Nguyen<sup>2,4</sup>, Bradley Pawlikowski<sup>3</sup>, Evan Lester<sup>1,2</sup>, Alicia Cutler<sup>3</sup>, Tiffany Elston<sup>3</sup>, Nicole Dalla Betta<sup>3</sup>, Kevin R. Parker<sup>6</sup>, Kathryn E. Yost<sup>6</sup>, Hannes Vogel<sup>13,14</sup>, Thomas A. Rando<sup>8,9,11</sup>, Howard Y. Chang<sup>6,7</sup>, Aaron M. Johnson<sup>4,5</sup>, Roy Parker<sup>1,12,16\*</sup>, Bradley B. Olwin<sup>3,16\*</sup>

## Affiliations:

1. Department of Biochemistry, University of Colorado Boulder, CO, USA
2. Medical Scientist Training Program, University of Colorado Anschutz Medical Campus Aurora, CO, USA
3. Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, CO, USA
4. Molecular Biology Program and Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA.
5. University of Colorado School of Medicine, RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO, USA.
6. Center for Personal and Dynamic Regulomes, Stanford University, CA, USA
7. Howard Hughes Medical Institute, Stanford University, CA, USA
8. Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA.
9. Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA.
10. Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA
11. Center for Tissue Regeneration, Repair, and Restoration, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA.
12. Howard Hughes Medical Institute, University of Colorado, Boulder, CO, USA
13. Department of Pathology, Stanford University
14. Department of Neuropathology, Stanford University
15. These authors contributed equally
16. Co-senior author

\* Corresponding author (Email: [roy.parker@colorado.edu](mailto:roy.parker@colorado.edu) (R.P.); [olwin@colorado.edu](mailto:olwin@colorado.edu) (B.B.O.))

# **ABSTRACT**

RNA-binding proteins (RBPs) are essential for skeletal muscle regeneration and RBP dysfunction causes muscle degeneration and neuromuscular disease. How ubiquitously expressed RBPs orchestrate complex tissue regeneration and direct cell fate decisions in skeletal muscle remains poorly understood. Single cell RNA-sequencing of regenerating skeletal muscle reveals that RBP expression, including numerous neuromuscular disease-associated RBPs, is temporally regulated in skeletal muscle stem cells and correlates to stages of myogenic differentiation. By combining machine learning with RBP engagement scoring, we discover that the neuromuscular disease associated RBP Hnrnpa2b1 is a differentiation-specifying regulator of myogenesis controlling myogenic cell fate transitions during terminal differentiation. The timing of RBP expression specifies cell fate transitions by providing a layer of post-transcriptional regulation needed to coordinate stem cell fate decisions during complex tissue regeneration.

# **KEYWORDS**

Skeletal muscle regeneration; myogenesis; single cell sequencing; machine learning; RNA-binding protein; Hnrnpa2b1; neuromuscular disease

# INTRODUCTION

Skeletal muscle is among the longest-lived tissues in the human body, is essential for locomotion, respiration, and longevity, and thus requires constant maintenance (Sharples et al., 2015). Skeletal muscle is comprised of post-mitotic myofibers that house resident muscle stem cells (MuSCs), which have the remarkable ability to repair skeletal muscle following damage (Lepper et al., 2011; Murphy et al., 2011; Sambasivan et al., 2011; Shi and Garry, 2006). MuSCs typically reside in a quiescent state (Feige et al., 2018). However, in response to muscle injury, MuSCs become activated, expand and then either self-renew or differentiate into progenitors capable of repairing myofibers (Baghdadi and Tajbakhsh, 2018; Dumont et al., 2015).

Single cell analyses of muscle regeneration demonstrate the rich cellular complexity governing myogenesis and make two key observations (De Micheli et al., 2020a; Dell’Orso et al., 2019a; Giordani et al., 2019). First, in response to damage, MuSCs exit quiescence and progress through a hierarchical, dynamic myogenic program as activated MuSCs either commit to terminal differentiation or self-renew and re-acquire a quiescent state. Second, as activated MuSCs progress through myogenesis, MuSCs experience dramatic global changes in gene expression (Barruet et al., 2020). This rapid change in gene expression requires MuSCs to regulate the vast amount of newly transcribed RNA encoding fate-specifying transcription factors and skeletal muscle contractile apparatus constituents, among many other myogenic proteins.

RNA is regulated by an arsenal of abundant RNA-binding proteins (RBPs). Previous work on RBPs in myogenesis focused on specific RBP proteins, often associated with disease, and their function in regulating key myogenic transcription factors that define myogenesis (Apponi et al., 2011). However, more recently identified roles for RBPs in myogenesis include:

(1) maintenance of MuSCs quiescence (Morrée et al., 2017), (2) MuSC activation and expansion (Cho and Doles, 2017; Farina et al., 2012), (3) myogenic differentiation (Hausburg et al., 2015; Vogler et al., 2018) and (4) MuSC-self-renewal (Chenette et al., 2016; Hausburg et al., 2015), demonstrating the critical role of RNA regulation throughout muscle regeneration.

One major RBP function is the regulation of RNA splicing, where RNA is targeted and bound by RBPs co-transcriptionally to guarantee correct splicing of nascent RNA transcripts (Dassi, 2017; Hentze et al., 2018). RBP-mediated splicing yields a rich diversity of RNA isoforms critical for translating proteins necessary for myogenesis (Brinegar et al., n.d.; Imbriano and Molinari, 2018; Nakka et al., 2018; Weskamp et al., 2020). The loss of specific RBPs and the resultant effects on splicing diversely affects quiescence, activation, or muscle differentiation and thus, mutations in RBPs are associated with muscular dystrophies and age-related neuromuscular diseases (Calado et al., 2000; Hinkle et al., 2019; Xue et al., 2020; Yu et al., 2009).

Amyotrophic lateral sclerosis, inclusion body myopathy, and muscular dystrophies are caused by RBP mutations leading to progressive muscle degeneration (Lukong et al., 2008; Xue et al., 2020). In these disorders, disease causing mutations frequently impair RBP splicing function (Cortese et al., 2014; Singh et al., 2018; Taylor et al., 2016) and restoring RBP splicing function prevents or delays the onset of these disorders, identifying splicing dysregulation as a key driver of pathology (Naryshkin et al., 2014; Scotti and Swanson, 2016; Wirth et al., 2020).

Despite the fact that RBPs are essential for skeletal muscle regeneration and are frequently mutated in neuromuscular disease, little is known regarding the mechanisms regulating the timing of RBP function during myogenesis or the mechanisms by which RBPs influence myogenic cell fate decisions. Using single cell RNA-sequencing we examined

temporal RBP expression of several neuromuscular disease-associated RBPs in MuSCs during myogenic differentiation to clarify the timing of RPB expression during muscle regeneration and identify networks of RBPs involved in myogenic cell fate transitions.

## RESULTS

### MYOGENIC CELL FATE TRANSITIONS REVEALED BY SINGLE CELL RNA SEQUENCING

Skeletal muscle has a remarkable ability to repair following injury. In response to injury, DNA-based lineage tracing and single cell sequencing experiments demonstrate that MuSCs activate, proliferate and differentiate to produce the majority of myonuclei by 7-days post injury (dpi) (De Micheli et al., 2020). MuSCs also self-renew to replenish the quiescent MuSC stem cell population with the majority of self-renewal occurring between 5- and 8-dpi (Pawlikowski et al., 2019). Thus, between 4- and 8-dpi the MuSC population traverses all stages of myogenesis as skeletal muscle regenerates. During this time, MuSCs regulate specific networks of RNA to direct myogenic fate decisions, but the role of RBPs in maintaining a malleable myogenic transcriptome remains poorly understood. We first sought to characterize myogenesis in single cells between 4- and 7-dpi to gain insight into the timing of RBP expression during MuSC activation, proliferation, differentiation, and self-renewal.

To induce skeletal muscle regeneration, we first injured adult mouse tibialis anterior (TA) muscles with barium chloride (Caldwell et al., 1990), collected regenerating muscle at 4-dpi and 7-dpi and performed single cell RNA sequencing. Analysis of the regenerating skeletal muscle shows strong reproducibility amongst biological replicates (S1A-B) and defines the cellular composition of the regenerating skeletal muscle (Figure 1A). Following acute muscle

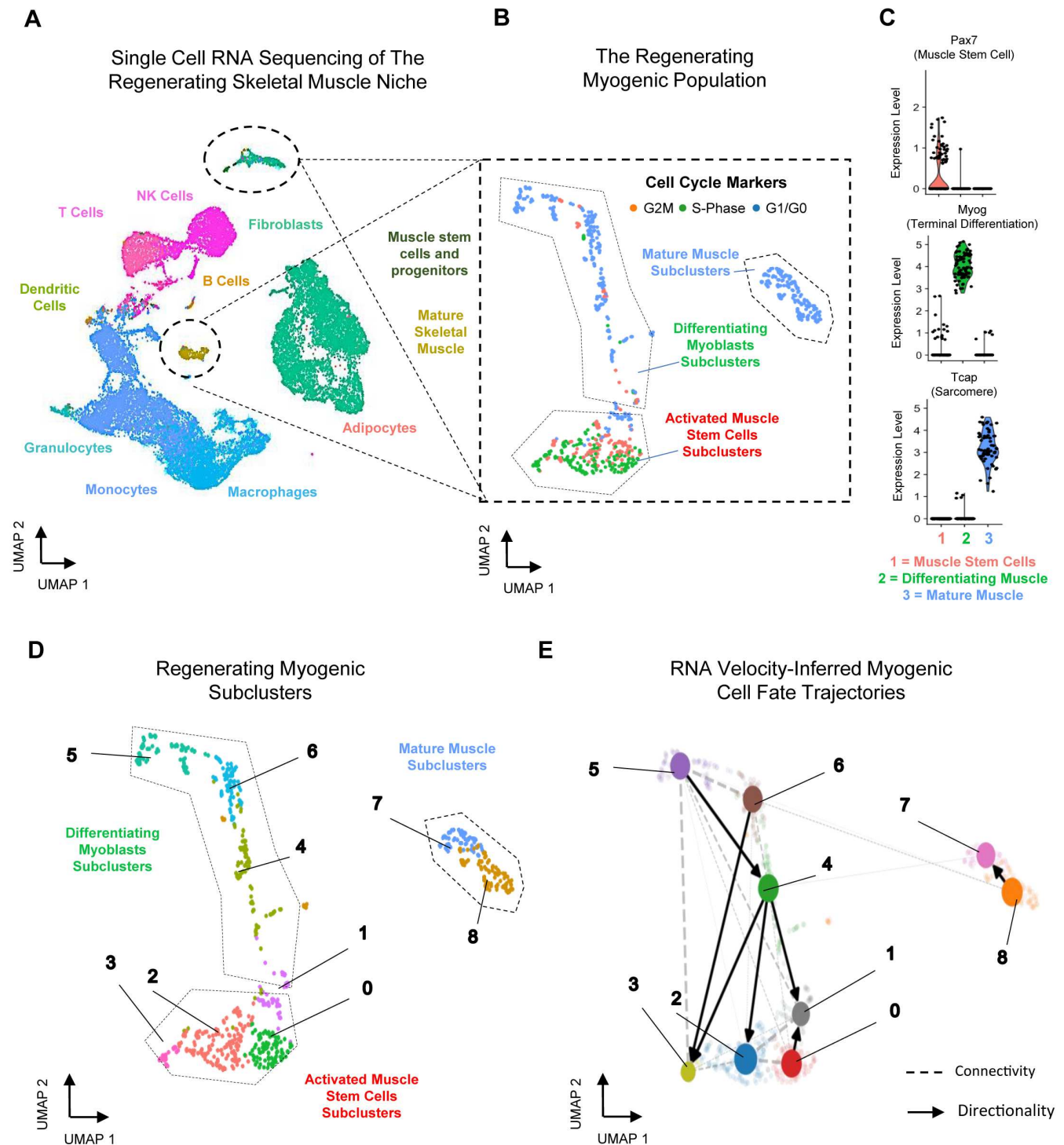
injury, we observe the infiltration of immune cells, dynamic cycling of individual fibro/adipogenic progenitors (FAPs), and a myogenic progenitor population (Figure 1A, S1A-B). The defined cellular populations during skeletal muscle regeneration are consistent with those from published myogenic single cell sequencing datasets, which suggests that acutely injured muscle undergoes similar mechanisms of cellular regeneration irrespective of the inciting injury (Barruet et al., 2020; De Micheli et al., 2020b; Dell’Orso et al., 2019b).

Gene expression analysis and cell cycle scoring analysis of the myogenic populations early during muscle regeneration reveals three dominant myogenic clusters: a proliferative Pax7-positive MuSC population, a differentiating Myogenin-positive population exiting cell cycle, and a post-mitotic, terminally differentiating muscle population expressing RNAs encoding sarcomeric proteins (Figure 1B-C, S1C). These dominant myogenic populations are further subclassified into nine subclusters, which together define the regenerating myogenic population (Figure 1D). It is unclear if these sub-clusters represent specific cell state transition points within the continuum of myogenesis, or whether we are catching cells at various points along a more continuous differentiation trajectory.

Using RNA velocity, we next sought to identify temporal dynamics between each of the nine myogenic subclusters by leveraging directed dynamic information between single cells (Bergen et al., 2020). RNA velocity works on the principle that newly transcribing genes’ pre-mRNAs contain introns that have not yet been removed by RNA splicing, which can be detected in single cell data. By globally comparing pre-mRNA reads to the corresponding processed mRNA, RNA velocity can infer cellular dynamics for a single cell (La Manno et al., 2018). We found that RNA velocity analysis on single cells identified myogenic differentiation directionality within specific subclusters, suggesting that myogenic differentiation trajectories

can be resolved and predicted based on the processed and unprocessed RNA within myogenic cells (Figure S1C).

To further characterize and predict the connectivity between the nine subclusters, we utilized partition-based graph abstraction (PAGA), which estimates connectivity among different sub-clusters while preserving global data topology, inferring cellular trajectories in finer detail than traditional pseudotemporal analyses (Wolf et. al, 2019). PAGA reveals a dynamic connectivity between the myogenic subclusters 0-6 that underlies *in vivo* myogenesis (Figure 1E, S1D) where robust connections are observed between MuSC clusters and differentiating clusters, highlighting the underlying complexity of myogenic cell fate decisions as myogenic cells either undergo self-renewal or commit to terminal differentiation (Figure S1D). While a high degree of connectivity is observed amongst the immature myogenic populations, very few connections lead to terminally differentiated, mature muscle, suggesting that terminal differentiation may follow a singular, universal pathway rather than multiple, diverse pathways. Our analyses suggest that these subclusters likely represent dynamic cell fate transition points within proliferating MuSCs and differentiating progenitor populations, capable of either progressing to terminal differentiation or returning to a more stem-like state; however, the mechanism that drives these cell fate dynamics between myogenic subclusters remains unknown. Although the mechanism of myogenic cell fate change is clearly multifactorial, a likely driver of rapid and dynamic cell fate changes is post-transcriptional control of RNA. Our data reveal a trajectory map of early *in vivo* myogenesis and provide an opportunity to examine temporal RNA regulation during rapid and dynamic myogenic cell fate transitions.



**Figure 1:** Single cell analysis reveals myogenic cell fate transitions in regenerating skeletal muscle (A) Single cell atlas of regenerating skeletal muscle at 4- and 7-days post injury (-dpi) (B) Cell cycle scoring in regenerating myogenic subclusters (C) Violin plots showing expression of myogenic markers of regeneration per myogenic clusters (D) Myogenic subclusters comprising the regenerating myogenic cellular population (E) RNA velocity-inferred myogenic cell fate trajectories. See also Figure S1.



## RNA-BINDING PROTEIN EXPRESSION IS TEMPORALLY DEFINED DURING MYOGENESIS

RNA is regulated in large part through the interaction with an ensemble of RBPs. While RBP expression increases during muscle regeneration, little is known about the timing of RBP expression in individual MuSCs and how specific RBP expression changes over the course of myogenesis *in vivo*. We hypothesized that temporal regulation of RBP expression in MuSCs modulates post-transcriptional gene expression to influence cell fate.

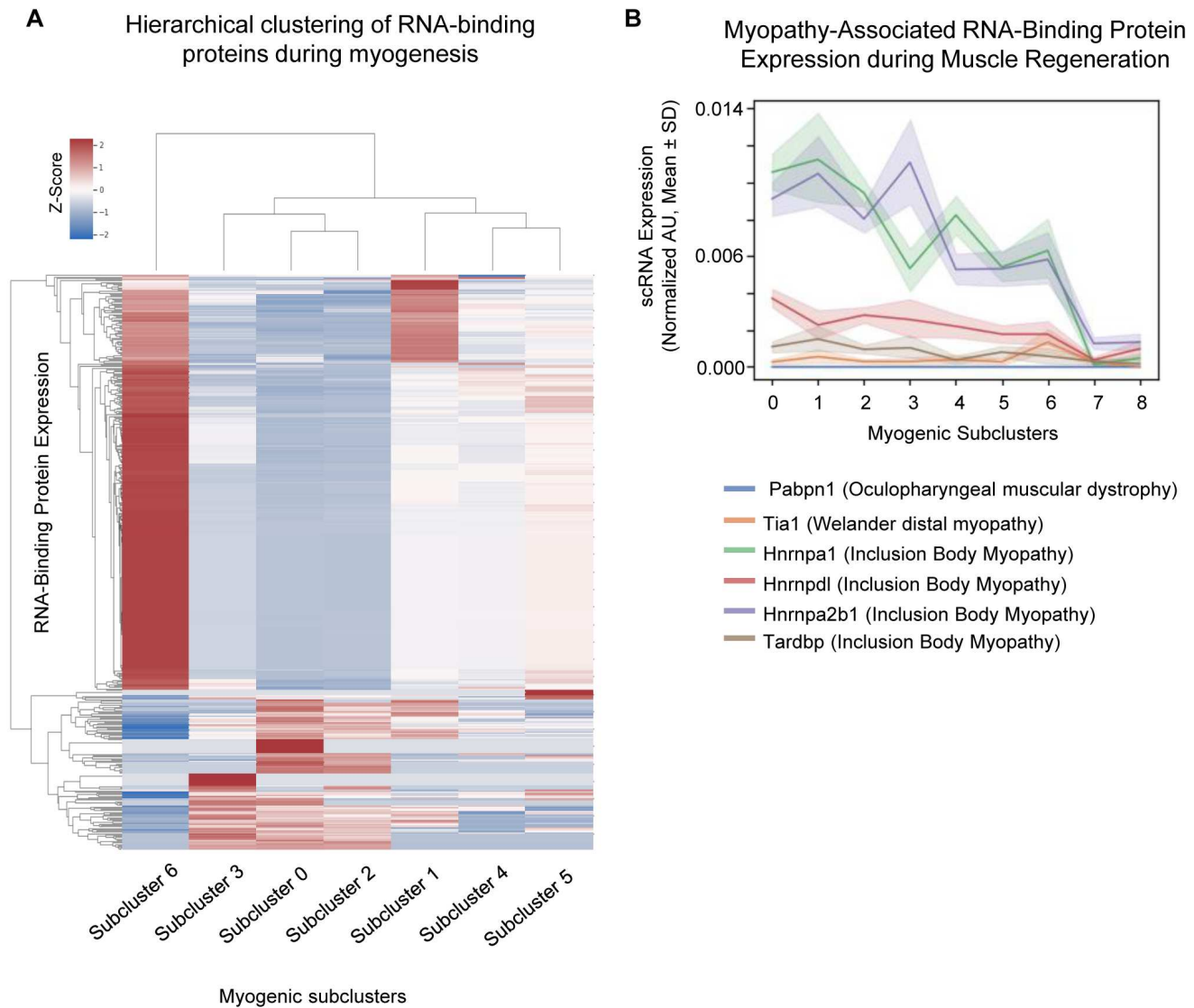
To test this hypothesis, we examined the steady state transcript levels of all RBPs in myogenic subclusters within proliferating MuSC and differentiating progenitor populations (subclusters 0-6). Hierarchical clustering of RBP transcript levels in each myogenic subcluster demonstrates that RBP expression correlates to the underlying myogenic cell state (Figure 2A). Groups of RBPs appear to exhibit similar expression patterns related to the associated myogenic subcluster (Figure 2A). Thus, coordinated upregulation or downregulation of select groups of RBPs serves to post-transcriptionally regulate RNAs specific to dynamic cell state changes over the course of myogenesis.

Dysfunction in neuromuscular disease associated RBPs gives rise to distinctive myogenic phenotypes. For example, Tardbp (TDP-43) dysfunction impairs myogenic proliferation, whereas loss of Hnrnpa1 disrupts terminal myogenic differentiation (Liu et al., 2017; Vogler et al., 2018). Since select RBP dysfunction contributes to degenerative neuromuscular diseases, defining the expression timing of disease related RBPs may help provide mechanistic clues to the contribution of RBPs to muscle pathology.

We observe cell state specific changes in the gene expression for disease associated RBPs (Figure 2B). For example, TDP-43 expression is highest in subcluster 1, Hnrnpa2b1 expression

is highest in subcluster 3, and Tia1 expression is highest in subcluster 6. Expression of specific disease associated RBPs fluctuates and peaks at different time points during myogenesis, which may influence in which cell clusters RBP dysfunction is most deleterious. Intriguingly, some RBPs display marked comparative expression timing changes. For example, both Hnrnpa1 and Hnrnpa2b1 are highly expressed in subcluster 1, while in subcluster 3, Hnrnpa1 decreases and comparatively, Hnrnpa2b1 increases expression. Finally, the amount of RBP RNA expressed varies dramatically between different RBPs. For example, Hnrnpa2b1 expression is five-fold higher in proliferating MuSC and in differentiating MuSC clusters as compared to TDP-43 expression (Figure 2B). The differences in expression between subclusters may reflect functional requirements for specific RBPs at precise times during myogenic cell fate changes.

Individual RBP expression, including the expression of many RBPs linked to human disease, appears to correlate with specific myogenic cell states. Myogenic cells may alter the expression of these individual RBPs within distinct cell stages to post-transcriptionally regulate large families of RNAs. By modulating the expression of key RBPs, and in turn specific subsets of RNA, myogenic cells could more easily transition between different cell fate trajectories without massive transcriptional changes.



**Figure 2:** Cluster-specific RNA-binding protein expression is temporally defined during myogenesis. (A) Hierarchical clustering of RNA-binding protein expression in myogenic subclusters 1-6 during myogenesis. (B) Line plot of select myopathy-associated RNA-binding protein expression in subclusters 0-8 during myogenesis.

## HNRNPA2B1 EXPRESSION DYNAMICS IN REGENERATING SKELETAL MUSCLE

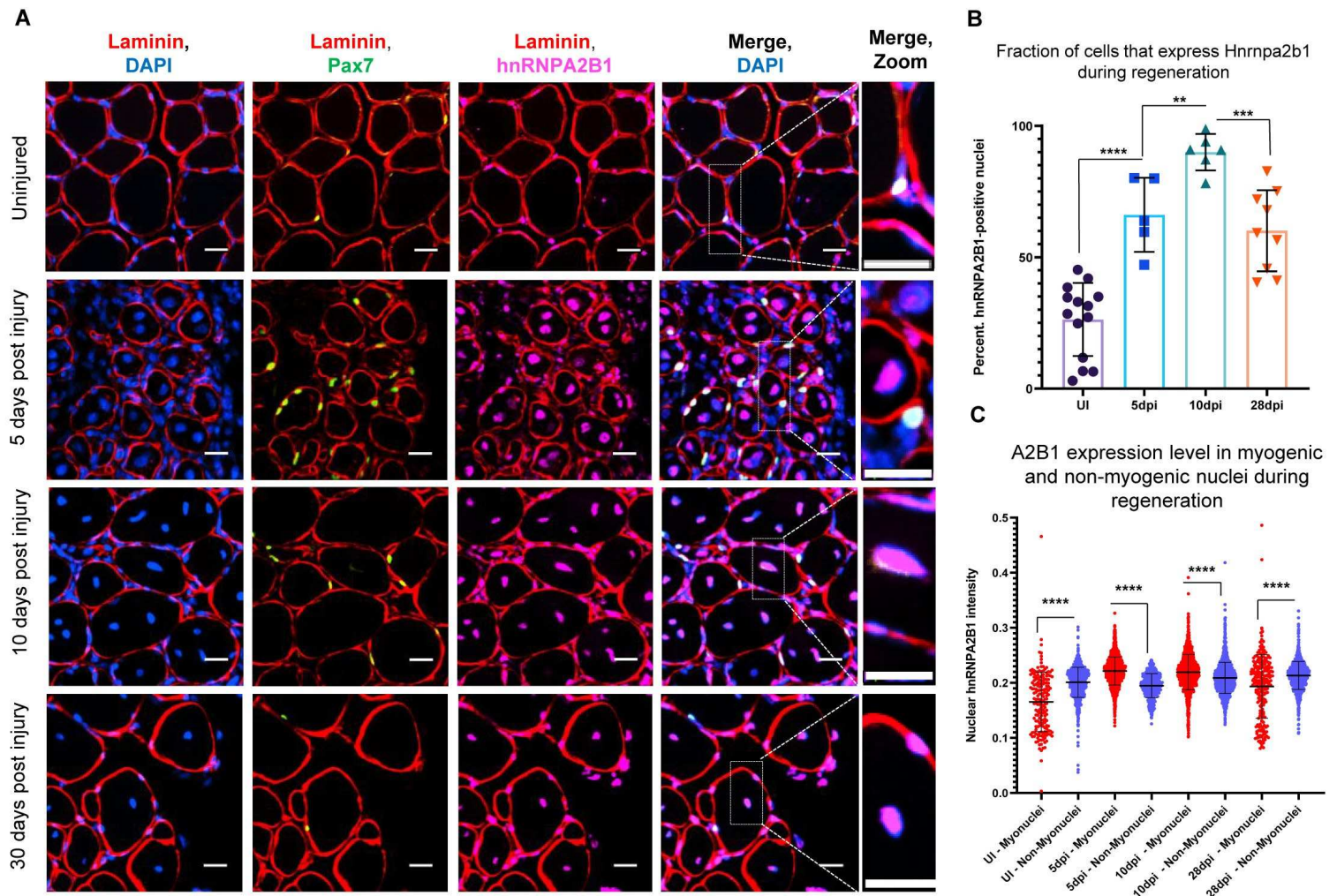
The five-fold expression increase in Hnrnpa2b1 during myoblast differentiation points towards a functional role for Hnrnpa2b1 in regulating muscle regeneration. As an RBP involved in regulating RNA splicing, stability and transport, but also found mutated in neuromuscular diseases such as inclusion body myositis, we hypothesized that Hnrnpa2b1 may function as a pivotal myogenic RNA regulator (Alarcón et al., 2015; Geissler et al., 2016; Percipalle et al., 2002). Here, Hnrnpa2b1 may influence the cycling of myogenic cells between cell fate trajectories to orchestrate successful differentiation. The function of Hnrnpa2b1 during myogenesis is unknown, but in hematopoietic stem cells, Hnrnpa2b1 regulates cell differentiation (Wang et al., 2018) and proliferation (He and Smith, 2009). The mRNA of Hnrnpa2b1 is highly expressed, and it undergoes large relative expression timing changes as compared to other RBPs between myogenic subclusters (Figure 2A-B). We therefore examined the influence of Hnrnpa2b1 on myogenic cell fate.

To validate the dynamics of Hnrnpa2b1 protein during *in vivo* skeletal muscle regeneration, mouse TA muscle was injured with barium chloride, harvested at specific time points following injury, and assayed for Hnrnpa2b1 immunoreactivity. In uninjured muscle, Hnrnpa2b1 is present in a rare subset of Pax7-positive MuSCs and present in some peripherally located myonuclei in mature skeletal muscle fibers (Figure 3A-B, S3A). By five days post injury (5-dpi), Hnrnpa2b1 expression increases and is present in the majority of Pax7-positive MuSCs and centrally located myonuclei of immature regenerating myofibers (Figure 3A-C, S3A-B).

As muscle continues to regenerate, Hnrnpa2b1 levels peak in both Pax7-positive MuSCs and in regenerating myofibers at 10-dpi followed by a decline in expression as myofibers fully regenerate by 28-dpi (Figure 3A-C, S3A, S3C-D). These results identify a global increase in

Hnrnpa2b1 expression early during muscle regeneration in both proliferating Pax7-positive MuSCs and in newly regenerating myofibers, correlating with *in vivo* MuSC Hnrnpa2b1 mRNA expression dynamics.

While an increase in Hnrnpa2b1 expression may be due to a global increase in RNA transcription in cells populating a regenerating muscle, Hnrnpa2b1 immunoreactivity peaks at 10-dpi with a wide, notable variation between muscle progenitor cells (Figure S3C). Thus, Hnrnpa2b1 is likely dynamically regulated during myogenesis. In comparison to Hnrnpa2b1 expression, TDP-43 immunoreactivity peaks in Pax7-positive MuSCs and regenerating myofibers at 5-dpi, before declining to near baseline by 10-dpi (Figure S3E-G). Peak TDP-43 mRNA expression occurs in a different myogenic subcluster than peak Hnrnpa2b1 mRNA expression (Figure 2B). Thus, the timing of TDP-43 mRNA and protein expression differs from that of Hnrnpa2b1 during muscle regeneration, suggesting that individual RBPs influence specific cellular aspects of *in vivo* myogenesis. Here, Hnrnpa2b1 function may be more important at later stages of myogenesis as compared to TDP-43.



**Figure 3:** Hnrnpa2b1 is upregulated in myogenic nuclei during skeletal muscle regeneration. (A) Hnrnpa2b1 immunoreactivity in uninjured (UI), 5-, 10-, and 30-dpi regenerating mouse muscle. All images represent n=3 biological replicates; scale = 20 $\mu$ M. (B) Nuclear Hnrnpa2b1 immunoreactivity in UI, 5-, 10-, 28-dpi regenerating muscle (C) Hnrnpa2b1 nuclear immunoreactivity intensity in either myonuclei or non-myonuclei in UI, 5-, 10-, 28-dpi regenerating muscle. See also Figure S3.



## HNRNPA2B1 IS REQUIRED FOR MYOBLAST DIFFERENTIATION

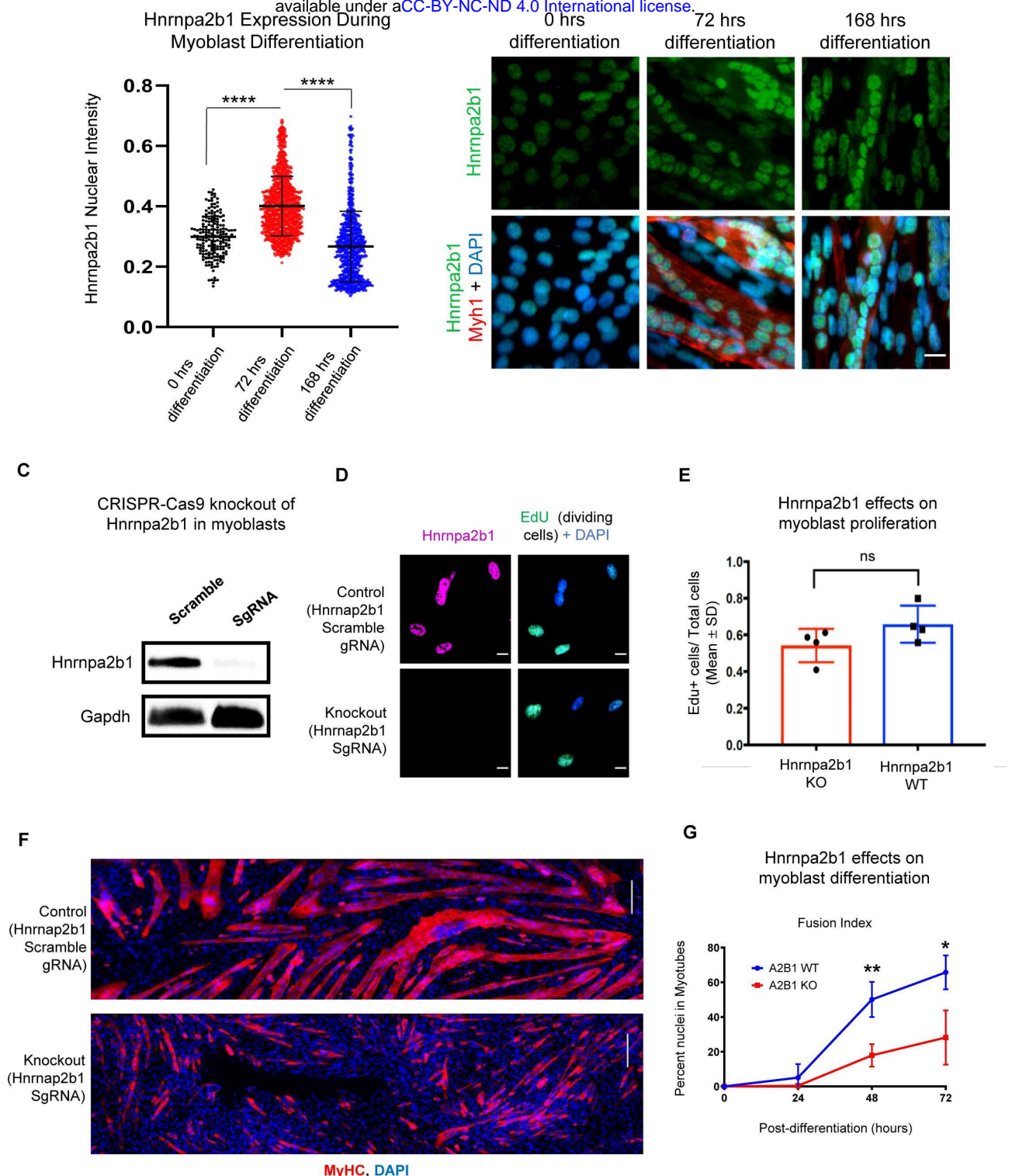
The notable increases in Hnrnpa2b1 mRNA and protein expression at specific points during muscle regeneration suggest that Hnrnpa2b1 may play a functional role at specific myogenic fate transitions. To test this hypothesis, we sought to characterize the function of Hnrnpa2b1 in an *in vitro* model of myogenesis. We find Hnrnpa2b1 expressed in both proliferating myoblasts and differentiating myocytes. Analogous to Hnrnpa2b1 expression *in vivo*, Hnrnpa2b1 expression increases during myoblast differentiation, peaks by 3 days of differentiation and wanes as terminally differentiated, multinucleated myotubes continue to mature in culture (Figure 4A-B). The observation that Hnrnpa2b1 expression peaks at later stages of both *in vivo* and *in vitro* myogenesis points to a possible role for Hnrnpa2b1 in regulating differentiation.

To explore whether Hnrnpa2b1 functions at distinct time points during myogenesis, we knocked out Hnrnpa2b1 in myoblasts using CRISPR/Cas9 and assessed the functional consequences of Hnrnpa2b1 loss on myoblast proliferation and differentiation (Figure 4C, S4A-B). No significant difference between wild type (WT) and Hnrnpa2b1 knockout (KO) proliferating myoblasts was detected when labeled with a timed pulse of the thymidine analog 5-Ethynyl-2'-deoxyuridine (EdU; Figure 4D-E).

Loss of Hnrnpa2b1, however, impaired myoblast differentiation. After 48 hours of differentiation, wild type myoblasts had largely differentiated into multinucleated, myosin heavy chain-positive myotubes (Figure 4F-G), while Hnrnpa2b1 KO myoblasts were unable to form large multinucleated myotubes (Figure 4F-G). The Hnrnpa2b1 KO differentiation defect persisted after extended differentiation in culture, suggesting an inability to terminally differentiate (Figure 4G, S4C). Thus, Hnrnpa2b1 function is required for myotube differentiation

but not myoblast proliferation, despite Hnrnpa2b1 being expressed in both differentiating myotubes and in proliferating myoblasts.





**Figure 4:** Hnrnpa2b1 is required for myogenic differentiation. (A) Hnrnpa2b1 nuclear expression in exponentially growing myoblasts (0 hours) and differentiating myotubes at 72 and 168 hours (B) Immunofluorescence showing Hnrnpa2b1 expression in myoblasts and differentiating myotubes (C) Western blot analysis of CRISPR-Cas9 knockout (KO) and scramble Hnrnpa2b1 sgRNA in C2C12 myoblasts (D) EdU-pulsed wildtype (WT) myoblasts and Hnrnpa2b1 KO myoblasts (scale = 10uM) (E) Quantification of EdU incorporation in WT and KO Hnrnpa2b1 myoblasts (ns = non-significant) (F) Staining for MyHC of differentiating in WT and KO Hnrnpa2b1 myotubes (scale = 200uM) (G) Quantification of Fusion Index (percentage of nuclei fused into myotubes) during differentiation in either WT and KO Hnrnpa2b1 myotubes. All quantified data represent mean  $\pm$  SD, two-tail student t-test p-value: \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , \*\*\*\* =  $p < 0.0001$  unless otherwise stated. See also Figure S4.

## **HNRNPA2B1 IS A MYOGENIC SPLICING REGULATOR CRITICAL FOR TERMINAL MYOGENIC DIFFERENTIATION**

Hnrnpa2b1 is an RNA splicing regulator and therefore, we hypothesized that Hnrnpa2b1 may function to ensure RNA splicing regulation to enable muscle differentiation. To test this hypothesis, we performed high-coverage RNA deep sequencing of Hnrnpa2b1 KO and wild type differentiating myotubes followed by differential gene expression and differential splicing analyses (Figure 5A, Supplementary Table 3 & 4). We identified 2,167 RNAs with significantly altered splicing between Hnrnpa2b1 wild type and Hnrnpa2b1 KO myotubes, of which 40% were differentially expressed, suggesting that alternative splicing may impact RNA stability of these transcripts (Figure 5A, Supplementary Table 3 & 4). These results support the hypothesis that Hnrnpa2b1 splices transcripts regulating myogenic cell fate during muscle differentiation. These widespread splicing aberrations in Hnrnpa2b1 KO cells may also help explain their impaired differentiation.

Given the high burden of splicing alterations identified in Hnrnpa2b1 KO myotubes, and the propensity of RBPs to regulate the splicing of additional RBPs, we intuited that Hnrnpa2b1 may regulate splicing of transcripts encoding additional splicing proteins that in turn affect myogenic cell fates. Synergistic splicing regulation of related RBPs by other splicing associated RBPs cooperatively influences developmental timing and myogenesis (Klinck et al., 2014; Venables et al., 2013). Temporal regulation by splicing associated RBPs of other related RBPs may provide a mechanism to finely tune splicing dynamics and RNA processing during myogenic differentiation.

Differential splicing analysis in Hnrnpa2b1 KO reveals numerous alternatively spliced RNAs that encode splicing related proteins including the myogenic splicing regulator RBPs

Mbnl1, Mbnl2, and Rbfox2 (Figure 5B). As loss of Mbnl1, Mbnl2, or Rbfox2 results in impaired myogenic differentiation, the alternative splicing of these proteins may compromise their splicing function. Consistent with this prediction, we observe a significant overlap in the alternatively spliced RNA populations identified between Hnrnpa2b1 and Mbnl1 & Mbnl2 double knockout differentiating muscle (hypergeometric p-value =  $1.2 \times 10^{-77}$ ), and Hnrnpa2b1 and Rbfox2 knockout differentiating muscle (hypergeometric p-value =  $6.1 \times 10^{-143}$ ) (Figure 5C). Further, comparing the precise splice site and effect of the splicing alteration (inclusion or exclusion of introns) in shared alternatively spliced RNAs identifies many co-regulated splicing changes shared between Mbnl1, Mbnl2, Rbfox2, and Hnrnpa2b1. These RBPs co-regulate the splicing in *-cis* of shared RNA targets (Figure S5A), implying a role for Hnrnpa2b1 as a central regulator within a network of splicing factors that directs specific myogenic cell fate decisions by synergistically regulating a common cadre of target RNAs.

The potential redundancy between RBPs and their splicing functions is consistent with our earlier observations showing aligned expression patterns for related RBPs which correlate with the underlying myogenic subcluster (Figure 2A). As these RBPs are differentially expressed in distinct subclusters corresponding to the stage of myogenic differentiation, these RBPs may regulate the transition from one subcluster to the next. Therefore, we predict that defects in a given RBP may disrupt normal myogenesis and result in myogenic cells arresting in distinct subclusters during differentiation. To test this, we characterized the impact of Hnrnpa2b1 knockout on myogenic subcluster abundances during differentiation, where we employed the machine learning-based methodology CIBERSORTx to computationally impute the abundance of myogenic subclusters present in differentiating Hnrnpa2b1 KO and WT cells. CIBERSORTx estimates the abundance of cell types in a mixed cell population from bulk RNA sequencing

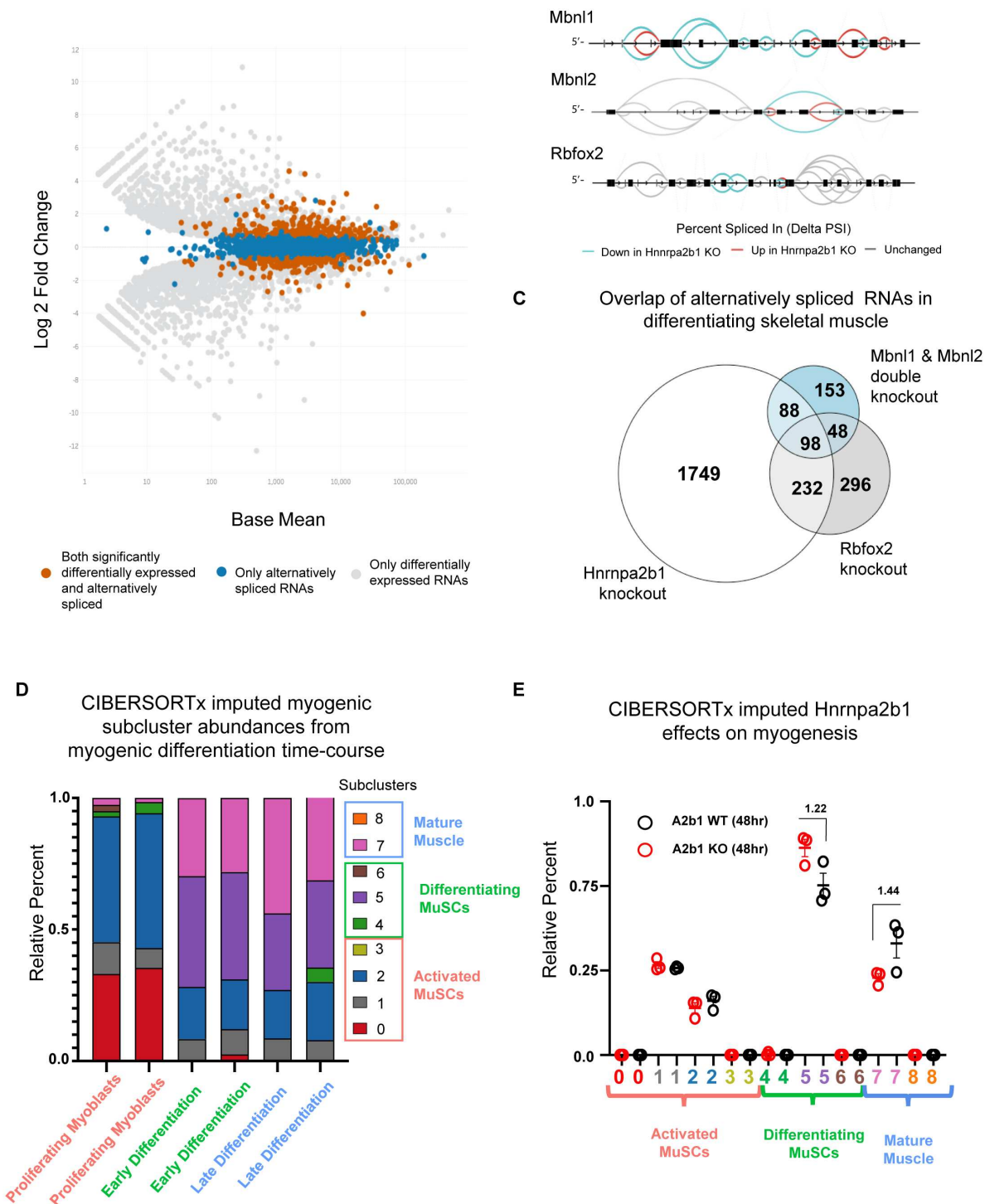
using single cell RNA sequencing data as a training set (Newman et al., 2019). By training this machine learning classifier on our single cell myogenic data, CIBERSORTx could calculate the percentages of cells at different stages of myogenesis from bulk RNA sequencing. If a block in differentiation arises due to RBP dysfunction or knockout, cells would be expected to accumulate into subclusters behind this block. Myogenic-trained CIBERSORTx would then predict the increase in the percentages of these cellular subclusters, pinpointing the timing of differentiation defects due to loss of RBP function.

As a proof-of-concept, we trained CIBERSORTx on our myogenic single cell dataset and imputed the percentages of myogenic cells from an RNA-sequencing dataset of C2C12 cell differentiation performed on exponentially growing myoblasts, and differentiating myotubes at 60 hours and 120 hours after differentiation induction (Trapnell et al., 2010). If myogenic-trained CIBERSORTx is capable of classifying distinct states of myogenic differentiation represented in MuSC single cell sequencing data from bulk C2C12 cell RNA sequencing, then proliferating MuSCs should cluster with proliferating C2C12 cells and MuSCs undergoing differentiation are expected to cluster with differentiating C2C12 cells, proportional to the time of differentiation the RNA sequencing was performed. Indeed, myogenic-trained CIBERSORTx reveals an enrichment for proliferative MuSCs subclusters in the exponentially growing C2C12 myoblast population and a shift towards differentiating subclusters proportional to differentiation duration (Figure 5D, S5B, Supplementary Table 5). The observation that not all myogenic subclusters were identified by CIBERSORTx in the C2C12 differentiation time course (e.g. subcluster 3) suggests that some subclusters are unique to myogenesis *in vivo* or do not comprise a substantial fraction of the *in vitro* C2C12 population at the assessed time points (Figure 5D, S5B). Myogenic-trained CIBERSORTx detects a small percentage of differentiated cells (e.g.

subcluster 7) present in the exponentially growing myoblasts (Figure 5E, S5B-C). Consistent with this observation, we performed immunohistochemistry against MyoG, a marker of differentiation, in exponentially proliferating myoblasts and identify a similar small percentage of MyoG expressing cells in exponentially growing myoblasts as predicted by CIBERSORTx (Figure S5C). These results demonstrate that myogenic-trained CIBERSORTx is capable of accurately imputing myogenic subcluster abundances from bulk RNA sequencing data.

Next, we sought to test the impact of Hnrnpa2b1 loss on cell subclusters during myogenesis. Myogenic-trained CIBERSORTx identifies an enrichment for subcluster 5 myogenic cell signatures and a concomitant decrease in terminally differentiated subcluster 7 in Hnrnpa2b1 KO cells (Figure 5E) showing that Hnrnpa2b1 splicing function ensures MuSCs transition from subcluster 5 to subcluster 6 during terminal differentiation.





**Figure 5:** Hnrnpa2b1 is a myogenic splicing regulator critical for terminal myogenic differentiation. (A) Differential gene expression identified by DESeq2 in WT and Hnrnpa2b1 KO myotubes with alternative splicing changes identified by Leafcutter emphasized (differential gene expression significance p-adjusted= $\leq 0.05$ , splicing FDR p-value= $\leq 0.05$ ) (B) Sashimi plots of significant alternative splicing changes (dPSI) for Mbnl1, Mbnl2, Rbfox2 in differentiating Hnrnpa2b1 KO myotubes (C) Venn diagram of significantly altered spliced transcripts in Hnrnpa2b1 KO, Mbnl1/Mbnl2 double-KO (Thomas et al., 2017), and Rbfox2 KO (Singh et al., 2014) differentiating myogenic cultures (D) Myogenic-trained CIBERSORTx imputed stacked bar chart of myogenic subcluster percentages in C2C12 differentiation time course (Trapnell et al., 2010) (E) Myogenic-trained CIBERSORTx machine learning imputed percentages of myogenic clusters from WT and KO Hnrnpa2b1 differentiating myotubes. Numbers refer to fold change between myogenic cluster percentages.

## **RBP ENGAGEMENT SCORING DELINEATES FUNCTIONAL TIMING OF RBPS DURING MYOGENESIS**

The timeframe in which a splicing RBP performs its function is tied to both the expression timing of the RBP and the expression of the RBP RNA targets. Hnrnpa2b1 promoting the transition of MuSCs from subcluster 5 to subcluster 6 during terminal differentiation suggests that RBP function may be linked to a particular cell state. The necessity of Hnrnpa2b1 for terminal differentiation may reflect both the specific temporal expression of Hnrnpa2b1 as well as Hnrnpa2b1 target RNAs. Therefore, we next identified and defined the expression dynamics of Hnrnpa2b1 RNA targets during myogenesis.

We performed Hnrnpa2b1 enhanced UV crosslinking and immunoprecipitation (eCLIP) in proliferating myoblasts and in differentiating myotubes (Figure 6A, S6A-C) to identify Hnrnpa2b1-associated RNAs. Hnrnpa2b1 eCLIP reveals 808 binding sites across 247 genes for myoblasts and 1,030 binding sites across 137 genes for myotubes, which are significantly enriched compared to size-matched input (reflecting all RNA–protein interactions in the input) (Figure S6D, Supplementary Table 2). Hnrnpa2b1 RNA target eCLIP peaks were highly correlated between biological replicates, had thousands of reproducible eCLIP clusters by irreproducible discovery rate analysis, and capture prior identified Hnrnpa2b1 RNA targets including Hnrnpa2b1’s own 3’ UTR (Figure S6D-G) (Martinez et al., 2016). Analysis of Hnrnpa2b1 RNA binding sites reveals an enrichment for binding sites in the 3’UTR and proximal introns of target RNAs (Figure 6A). Using native Hnrnpa2b1 RNA immunoprecipitation (RIP) followed by RT-qPCR of eCLIP target RNAs, we validated the interaction of Hnrnpa2b1 with several of Hnrnpa2b1 target RNAs identified by eCLIP (Figure 6B, S6H, Primers used detailed in Supplementary Table 1).

Many RBPs with splicing function, including Hnrnpa2b1, engage with target RNAs co-transcriptionally to regulate RNA splicing (Fu and Ares, 2014). Since recent computational methods allow for newly transcribed RNA or pre-mRNA expression to be measured at a single cell level, we may be able to infer the degree of RBP functionality within an individual cell by correlating the expression of a given RBP to the expression of the newly transcribed RBP target pre-mRNAs. We termed this approach RBP engagement scoring (Figure 6C).

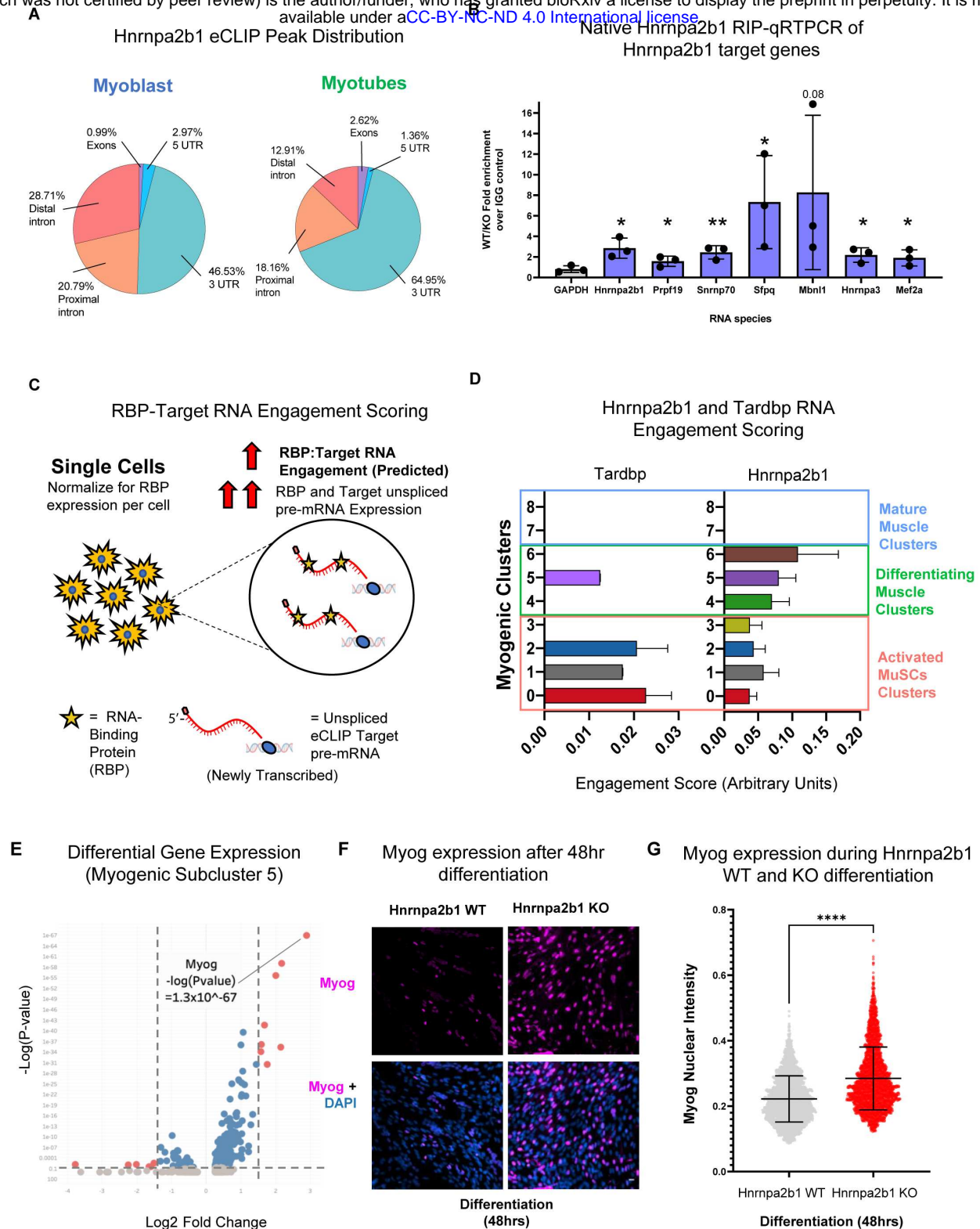
To test if RBP engagement scoring can infer known RBP phenotypes, we first calculated RBP engagement scores for TDP-43, a well characterized splicing RBP that is essential for MuSC proliferation (Vogler et al., 2018). TDP-43 engagement scoring should be higher in proliferating MuSC clusters as compared to differentiating MuSC clusters. The pre-mRNA expression of TDP-43 target RNAs previously identified by eCLIP was quantified across all myogenic cells identified in our single cell sequencing dataset (Vogler et al., 2018). TDP-43 target pre-mRNA expression was then correlated with TDP-43 expression per cell to assign each cell a TDP-43 engagement score. Next, TDP-43 engagement scores were summed across each cell within a myogenic subcluster to determine the relative activity of TDP-43 at each stage of myogenesis. We observe greater TDP-43 engagement scoring in proliferating MuSC clusters as compared to differentiating clusters or mature muscle clusters, consistent with a requirement of TDP-43 for MuSC proliferation (Figure 6D).

We next sought to use RBP engagement scoring to investigate the regeneration stage-specific timing of Hnrnpa2b1 function during myogenesis. In contrast to TDP-43, Hnrnpa2b1 engagement scoring revealed increased engagement scores in differentiating myogenic clusters (e.g. subcluster 6) and to a lesser extent in proliferating MuSCs (e.g. subcluster 1) as compared to mature muscle (Figure 6D). Hnrnpa2b1 expression and Hnrnpa2b1 target RNA expression is



higher in specific differentiating subclusters, suggesting that Hnrnpa2b1 interaction with target RNAs may be increased in those subclusters. Consistent with our prior myogenic CIBERSORTx analysis, Hnrnpa2b1 activity may be increased in select myogenic subclusters during differentiation, which may explain why the loss of Hnrnpa2b1 impairs MuSC transitions from one subcluster to the next (Figure 4D-G). As Hnrnpa2b1 engagement scores are increased in differentiating myogenic subclusters (e.g. subcluster 6), loss of Hnrnpa2b1 may impair these myogenic cells' progress towards mature muscle, leading to a build-up of differentiated but immature myogenic cells. Hnrnpa2b1 engagement scoring predicts that Hnrnpa2b1 targets are most prevalent in myogenic cluster 6 and therefore, cells may accumulate in preceding subcluster 5 in the absence of Hnrnpa2b1.

To identify gene expression signatures unique to each cluster, we performed differential gene expression analysis on myogenic subclusters 5 and 6. The myogenic transcription factor, Myogenin (MyoG) (p-value =  $1 \times 10^{-67}$ ) is significantly enriched in subcluster 5, which is consistent with subcluster 5 representing a key branch point between terminal differentiation and self-renewal (Figure 6E, S6I). When differentiated Hnrnpa2b1 KO and WT myoblasts were differentiated and assessed for MyoG immunoreactivity, MyoG-positive cells were increased in Hnrnpa2b1 KO cells compared WT cells, consistent with accumulation of Hnrnpa2b1 KO cells in subcluster 5 that are unable to progress or delayed in progressing to subcluster 6 (Figure 6F-G). Thus, RBP engagement scoring identifies the precise myogenic subcluster in which Hnrnpa2b1 is most functional and provides insight into how Hnrnpa2b1 loss can disrupt myogenic fate decisions.



**Figure 6:** RNA engagement scoring delineates functional timing of RBPs during myogenesis. (A) Genomic distribution of Hnrnpa2b1 eCLIP peaks in myoblasts and myotubes (B) Native Hnrnpa2b1 RNA immunoprecipitation (RIP)-qRT-PCR of RNA targets identified by eCLIP (C) Schematic for RBP-Target RNA engagement scoring (D) RNA-binding protein (RBP) – Target RNA engagement scoring of Tardbp and Hnrnpa2b1 in myogenic subclusters (median  $\pm$  95% CI for across all RBP-RNA engagement scores in each myogenic subcluster) (E) Differential gene expression analysis of myogenic subcluster 5 (F) Myog staining in WT and Hnrnpa2b1 KO 48 hour differentiated myotubes (G) Myog intensity in differentiating WT and Hnrnpa2b1 KO myotubes (mean  $\pm$  SD, two-tailed student t-test p-value: \*\*\*\*= $<0.0001$ ). See also Figure S6.

# DISCUSSION

RBP are essential for skeletal muscle regeneration and RBP dysfunction is associated with neuromuscular disease (Apponi et al., 2011; Farina et al., 2012). While select RBPs play an important role in myogenesis, the exact differentiation stage-specific roles of RBPs remain poorly understood.

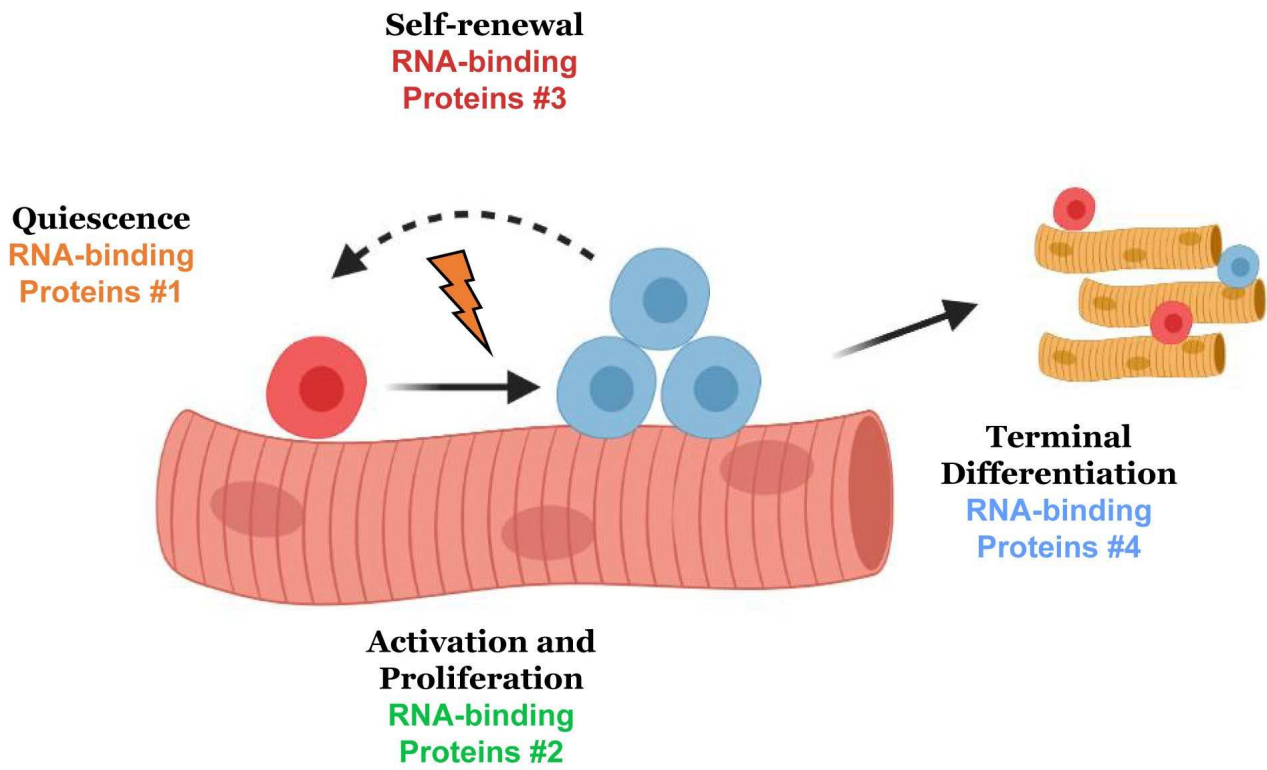
Single cell RNA sequencing of regenerating MuSC population points to MuSCs progressing through a dynamic program to direct cell fates either towards self-renewal or terminal differentiation. To our surprise, global and neuromuscular disease associated RBP expression profiles correlate with the underlying MuSC cell state. Utilizing a combination of machine learning and novel RBP engagement scoring, we predict and experimentally validate the role of Hnrnpa2b1, a neuromuscular disease associated RBP, in orchestrating terminal myogenic differentiation. Our results suggest that RBP splicing proteins coordinately guide myogenic cell fate decisions during muscle regeneration. The coordinated expression of related RBPs may provide an additional layer of post-transcriptional gene expression regulation, which may also occur in other developmental or tissue regenerative contexts.

We propose a model in which temporally defined RBP function is influenced both by the temporal expression of an RBP as well as the temporal expression of RBP RNA targets (Figure 7). In this model, RBP regulation is achieved by the sum of RNA targets available for a given RBP to interact with. During muscle regeneration, MuSCs rapidly respond to differentiation cues. Here, MuSCs may preemptively express cell-stage appropriate RBPs to rapidly regulate a dynamic myogenic transcriptome. We liken this mechanism to the pausing of transcription factors or polymerases at the sites of transcription awaiting the right cellular cues. By having

RBP primed to function, resident MuSCs are capable of more rapidly and flexibly regulating cell fate without the need to globally alter the underlying transcriptome.

Our results further suggest that the function of Hnrnpa2b1 in regulating terminal differentiation may be tied to the splicing regulation of RNAs encoding additional splicing regulatory RBPs. As RBP expression is temporally defined during myogenesis, the regulation of RBP splicing by related RBPs may restrict RBP function to when only absolutely necessary.

Finally, the observation that RBP expression is temporally restricted during myogenesis implies that distinct classes of RBPs may be critical to promote specific cell fate changes during myogenesis. Similar to the stepwise waves of myogenic transcription factor expression, RBP expression waves may globally regulate post-transcriptional gene expression. Further, splicing regulation of RBPs by other RBPs may enable isoform specific RBP functions, potentially to more finely tune the myogenic transcriptome throughout differentiation. Our findings help to resolve the seemingly disparate impact that different RBPs have on myogenesis. As RBP functions may be temporally restricted, dysregulation of an individual RBP may impair myogenesis at a precise cell fate transition. Thus, temporally defined and muscle regeneration stage-specific RBP function may provide ample opportunities for therapeutic intervention in neuromuscular disease and may instruct stem cell fate decisions during tissue regeneration (Ferlini et al., 2021).



**Figure 7:** Discrete RNA-binding protein functional timing and expression as a dynamic post-transcriptional regulatory mechanism for directing myogenic fate decisions.

## REFERENCES

- Alarcón CR, Goodarzi H, Lee H, Liu X, Tavazoie S, Tavazoie SF. 2015. HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* **162**:1299–1308. doi:10.1016/j.cell.2015.08.011
- Apponi LH, Corbett AH, Pavlath GK. 2011. RNA-binding proteins and gene regulation in myogenesis. *Trends Pharmacol Sci* **32**:652–658. doi:10.1016/j.tips.2011.06.004
- Baghdadi MB, Tajbakhsh S. 2018. Regulation and phylogeny of skeletal muscle regeneration. *Dev Biol* **433**:200–209. doi:10.1016/j.ydbio.2017.07.026
- Barruet E, Garcia SM, Striedinger K, Wu J, Lee S, Byrnes L, Wong A, Xuefeng S, Tamaki S, Brack AS, Pomerantz JH. 2020. Functionally heterogeneous human satellite cells identified by single cell RNA sequencing. *eLife* **9**:e51576. doi:10.7554/eLife.51576
- Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**:1408–1414. doi:10.1038/s41587-020-0591-3
- Brinegar AE, Xia Z, Loehr JA, Li W, Rodney GG, Cooper TA. n.d. Extensive alternative splicing transitions during postnatal skeletal muscle development are required for calcium handling functions. *eLife* **6**. doi:10.7554/eLife.27192
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**:411–420. doi:10.1038/nbt.4096
- Calado A, Tomé FM, Brais B, Rouleau GA, Kühn U, Wahle E, Carmo-Fonseca M. 2000. Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. *Hum Mol Genet* **9**:2321–2328. doi:10.1093/oxfordjournals.hmg.a018924
- Caldwell CJ, Matthey DL, Weller RO. 1990. Role of the basement membrane in the regeneration of skeletal muscle. *Neuropathol Appl Neurobiol* **16**:225–238.
- Chenette DM, Cadwallader AB, Antwine TL, Larkin LC, Wang J, Olwin BB, Schneider RJ. 2016. Targeted mRNA Decay by RNA Binding Protein AUF1 Regulates Adult Muscle Stem Cell Fate, Promoting Skeletal Muscle Integrity. *Cell Rep* **16**:1379–1390. doi:10.1016/j.celrep.2016.06.095
- Cho DS, Doles JD. 2017. Single cell transcriptome analysis of muscle satellite cells reveals widespread transcriptional heterogeneity. *Gene* **636**:54–63. doi:10.1016/j.gene.2017.09.014
- Cortese A, Plagnol V, Brady S, Simone R, Lashley T, Acevedo-Arozena A, de Silva R, Greensmith L, Holton J, Hanna MG, Fisher EMC, Fratta P. 2014. Widespread RNA metabolism impairment in sporadic inclusion body myositis TDP43-proteinopathy. *Neurobiol Aging* **35**:1491–1498. doi:10.1016/j.neurobiolaging.2013.12.029
- Dassi E. 2017. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front Mol Biosci* **4**:67. doi:10.3389/fmolb.2017.00067
- De Micheli AJ, Laurillard EJ, Heinke CL, Ravichandran H, Fraczek P, Soueid-Baumgarten S, De Vlaminc I, Elemento O, Cosgrove BD. 2020a. Single-Cell Analysis of the Muscle Stem Cell Hierarchy Identifies Heterotypic Communication Signals Involved in Skeletal Muscle Regeneration. *Cell Rep* **30**:3583–3595.e5. doi:10.1016/j.celrep.2020.02.067
- De Micheli AJ, Laurillard EJ, Heinke CL, Ravichandran H, Fraczek P, Soueid-Baumgarten S, De Vlaminc I, Elemento O, Cosgrove BD. 2020b. Single-Cell Analysis of the Muscle Stem Cell Hierarchy Identifies Heterotypic Communication Signals Involved in Skeletal Muscle Regeneration. *Cell Rep* **30**:3583–3595.e5. doi:10.1016/j.celrep.2020.02.067
- Dell’Orso S, Juan AH, Ko K-D, Naz F, Perovanovic J, Gutierrez-Cruz G, Feng X, Sartorelli V. 2019a. Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions. *Dev Camb Engl* **146**. doi:10.1242/dev.174177
- Dell’Orso S, Juan AH, Ko K-D, Naz F, Perovanovic J, Gutierrez-Cruz G, Feng X, Sartorelli V. 2019b. Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions. *Development* **146**. doi:10.1242/dev.174177
- Dumont NA, Bentzinger CF, Sincennes M-C, Rudnicki MA. 2015. Satellite Cells and Skeletal Muscle Regeneration. *Compr Physiol* **5**:1027–1059. doi:10.1002/cphy.c140068
- Farina NH, Hausburg M, Betta ND, Pulliam C, Srivastava D, Cornelison D, Olwin BB. 2012. A role for RNA post-transcriptional regulation in satellite cell activation. *Skelet Muscle* **2**:21. doi:10.1186/2044-5040-2-21
- Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput Biol* **8**:e1002529. doi:10.1371/journal.pcbi.1002529
- Feige P, Brun CE, Ritso M, Rudnicki MA. 2018. Orienting Muscle Stem Cells for Regeneration in Homeostasis, Aging, and Disease. *Cell Stem Cell* **23**:653–664. doi:10.1016/j.stem.2018.10.006



- Ferlini A, Goyenvalle A, Muntoni F. 2021. RNA-targeted drugs for neuromuscular diseases. *Science* **371**:29–31. doi:10.1126/science.aba4515
- Fu X-D, Ares M. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**:689–701. doi:10.1038/nrg3778
- Geissler R, Simkin A, Floss D, Patel R, Fogarty EA, Scheller J, Grimson A. 2016. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev* **30**:1070–1085. doi:10.1101/gad.277392.116
- Giordani L, He GJ, Negroni E, Sakai H, Law JYC, Siu MM, Wan R, Corneau A, Tajbakhsh S, Cheung TH, Le Grand F. 2019. High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. *Mol Cell* **74**:609–621.e6. doi:10.1016/j.molcel.2019.02.026
- Ha KCH, Blencowe BJ, Morris Q. 2018. QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol* **19**:45. doi:10.1186/s13059-018-1414-4
- Hausburg MA, Doles JD, Clement SL, Cadwallader AB, Hall MN, Blackshear PJ, Lykke-Andersen J, Olwin BB. 2015. Post-transcriptional regulation of satellite cell quiescence by TTP-mediated mRNA decay. *eLife* **4**:e03390. doi:10.7554/eLife.03390
- He Y, Smith R. 2009. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci* **66**:1239–1256. doi:10.1007/s00018-008-8532-1
- Hentze MW, Castello A, Schwarzl T, Preiss T. 2018. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**:327–341. doi:10.1038/nrm.2017.130
- Hinkle ER, Wiedner HJ, Black AJ, Giudice J. 2019. RNA processing in skeletal muscle biology and disease. *Transcription* **10**:1–20. doi:10.1080/21541264.2018.1558677
- Imbriano C, Molinari S. 2018. Alternative Splicing of Transcription Factors Genes in Muscle Physiology and Pathology. *Genes* **9**. doi:10.3390/genes9020107
- Klinck R, Fourrier A, Thibault P, Toutant J, Durand M, Lapointe E, Caillet-Boudin M-L, Sergeant N, Gourdon G, Meola G, Furling D, Puymirat J, Chabot B. 2014. RBFOX1 cooperates with MBNL1 to control splicing in muscle, including events altered in myotonic dystrophy type 1. *PloS One* **9**:e107324. doi:10.1371/journal.pone.0107324
- La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastri ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. 2018. RNA velocity of single cells. *Nature* **560**:494–498. doi:10.1038/s41586-018-0414-6
- Lepper C, Partridge TA, Fan C-M. 2011. An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. *Development* **138**:3639–3646. doi:10.1242/dev.067595
- Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK. 2018. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**:151–158. doi:10.1038/s41588-017-0004-9
- Liu T-Y, Chen Y-C, Jong Y-J, Tsai H-J, Lee C-C, Chang Y-S, Chang J-G, Chang Y-F. 2017. Muscle developmental defects in heterogeneous nuclear Ribonucleoprotein A1 knockout mice. *Open Biol* **7**. doi:10.1098/rsob.160303
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8
- Lukong KE, Chang K, Khandjian EW, Richard S. 2008. RNA-binding proteins in human genetic disease. *Trends Genet* **24**:416–425. doi:10.1016/j.tig.2008.05.004
- Martinez FJ, Pratt GA, Van Nostrand EL, Batra R, Huelga SC, Kapeli K, Freese P, Chun SJ, Ling K, Gelboin-Burkhart C, Fijany L, Wang HC, Nussbacher JK, Broski SM, Kim HJ, Lardelli R, Sundararaman B, Donohue JP, Javaherian A, Lykke-Andersen J, Finkbeiner S, Bennett CF, Ares M, Burge CB, Taylor JP, Rigo F, Yeo GW. 2016. Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron* **92**:780–795. doi:10.1016/j.neuron.2016.09.050
- Morée A de, Velthoven CTJ van, Gan Q, Salvi JS, Klein JDD, Akimenko I, Quarta M, Biressi S, Rando TA. 2017. Stauf1 inhibits MyoD translation to actively maintain muscle stem cell quiescence. *Proc Natl Acad Sci* **114**:E8996–E9005. doi:10.1073/pnas.1708725114
- Murphy MM, Lawson JA, Mathew SJ, Hutcheson DA, Kardon G. 2011. Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138**:3625–3637. doi:10.1242/dev.064162
- Nakka K, Ghigna C, Gabellini D, Dilworth FJ. 2018. Diversification of the muscle proteome through alternative splicing. *Skelet Muscle* **8**:8. doi:10.1186/s13395-018-0152-3

- Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, Ling KKY, Karp GM, Qi H, Woll MG, Chen G, Zhang N, Gabbeta V, Vazirani P, Bhattacharyya A, Furia B, Risher N, Sheedy J, Kong R, Ma J, Turpoff A, Lee C-S, Zhang X, Moon Y-C, Trifillis P, Welch EM, Colacino JM, Babiak J, Almstead NG, Peltz SW, Eng LA, Chen KS, Mull JL, Lynes MS, Rubin LL, Fontoura P, Santarelli L, Haehnke D, McCarthy KD, Schmucki R, Ebeling M, Sivaramakrishnan M, Ko C-P, Paushkin SV, Ratni H, Gerlach I, Ghosh A, Metzger F. 2014. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* **345**:688–693. doi:10.1126/science.1250127
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**:773–782. doi:10.1038/s41587-019-0114-2
- Nguyen ED, Balas MM, Griffin AM, Roberts JT, Johnson AM. 2018. Global profiling of hnRNP A2/B1-RNA binding on chromatin highlights LncRNA interactions. *RNA Biol* **15**:901–913. doi:10.1080/15476286.2018.1474072
- Pawlikowski B, Betta ND, Elston T, O'Rourke R, Jones K, Olwin BB. 2019. A cellular atlas of skeletal muscle regeneration and aging. *bioRxiv* 635805. doi:10.1101/635805
- Percipalle P, Jonsson A, Nashchekin D, Karlsson C, Bergman T, Guialis A, Daneholt B. 2002. Nuclear actin is associated with a specific subset of hnRNP A/B-type proteins. *Nucleic Acids Res* **30**:1725–1734. doi:10.1093/nar/30.8.1725
- Sambasivan R, Yao R, Kissenpfennig A, Wittenberghe LV, Paldi A, Gayraud-Morel B, Guenou H, Malissen B, Tajbakhsh S, Galy A. 2011. Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. *Development* **138**:3647–3656. doi:10.1242/dev.067587
- Scotti MM, Swanson MS. 2016. RNA mis-splicing in disease. *Nat Rev Genet* **17**:19–32. doi:10.1038/nrg.2015.3
- Sharples AP, Hughes DC, Deane CS, Saini A, Selman C, Stewart CE. 2015. Longevity and skeletal muscle mass: the role of IGF signalling, the sirtuins, dietary restriction and protein intake. *Aging Cell* **14**:511–523. doi:10.1111/acer.12342
- Shi X, Garry DJ. 2006. Muscle stem cells in development, regeneration, and disease. *Genes Dev* **20**:1692–708. doi:10.1101/gad.1419406
- Singh RK, Kolonin AM, Fiorotto ML, Cooper TA. 2018. Rbfox Splicing Factors Maintain Skeletal Muscle Mass by Regulating Calpain3 and Proteostasis. *Cell Rep* **24**:197–208. doi:10.1016/j.celrep.2018.06.017
- Singh RK, Xia Z, Bland CS, Kalsotra A, Scavuzzo MA, Curk T, Ule J, Li W, Cooper TA. 2014. Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. *Mol Cell* **55**:592–603. doi:10.1016/j.molcel.2014.06.035
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888–1902.e21. doi:10.1016/j.cell.2019.05.031
- Taylor JP, Brown RH, Cleveland DW. 2016. Decoding ALS: from genes to mechanism. *Nature* **539**:197–206. doi:10.1038/nature20413
- Thomas JD, Sznajder LJ, Bardhi O, Aslam FN, Anastasiadis ZP, Scotti MM, Nishino I, Nakamori M, Wang ET, Swanson MS. 2017. Disrupted prenatal RNA processing and myogenesis in congenital myotonic dystrophy. *Genes Dev* **31**:1122–1133. doi:10.1101/gad.300590.117
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**:511–515. doi:10.1038/nbt.1621
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M, Yeo GW. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**:508–14. doi:10.1038/nmeth.3810
- Venables JP, Lapasset L, Gadea G, Fort P, Klinck R, Irimia M, Vignal E, Thibault P, Prinos P, Chabot B, Abou Elela S, Roux P, Lemaitre J-M, Tazi J. 2013. MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat Commun* **4**:2480. doi:10.1038/ncomms3480
- Vogler TO, Wheeler JR, Nguyen ED, Hughes MP, Britson KA, Lester E, Rao B, Betta ND, Whitney ON, Ewachiw TE, Gomes E, Shorter J, Lloyd TE, Eisenberg DS, Taylor JP, Johnson AM, Olwin BB, Parker R. 2018. TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle. *Nature*. doi:10.1038/s41586-018-0665-2



- Wang H, Liang L, Dong Q, Huan L, He J, Li B, Yang C, Jin H, Wei L, Yu C, Zhao F, Li J, Yao M, Qin W, Qin L, He X. 2018. Long noncoding RNA miR503HG, a prognostic indicator, inhibits tumor metastasis by regulating the HNRNPA2B1/NF- $\kappa$ B pathway in hepatocellular carcinoma. *Theranostics* **8**:2814–2829. doi:10.7150/thno.23012
- Weskamp K, Tank EM, Miguez R, McBride JP, Gómez NB, White M, Lin Z, Gonzalez CM, Serio A, Sreedharan J, Barmada SJ. 2020. Shortened TDP43 isoforms upregulated by neuronal hyperactivity drive TDP43 pathology in ALS. *J Clin Invest* **130**:1139–1155. doi:10.1172/JCI130988
- Wirth B, Karakaya M, Kye MJ, Mendoza-Ferreira N. 2020. Twenty-Five Years of Spinal Muscular Atrophy Research: From Phenotype to Genotype to Therapy, and What Comes Next. *Annu Rev Genomics Hum Genet* **21**:231–261. doi:10.1146/annurev-genom-102319-103602
- Xue YC, Ng CS, Xiang P, Liu H, Zhang K, Mohamud Y, Luo H. 2020. Dysregulation of RNA-Binding Proteins in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* **13**. doi:10.3389/fnmol.2020.00078
- Yu Z, Wang AM, Robins DM, Lieberman AP. 2009. Altered RNA splicing contributes to skeletal muscle pathology in Kennedy disease knock-in mice. *Dis Model Mech* **2**:500–507. doi:10.1242/dmm.003301

**Acknowledgments:** We thank J. Dragavon, J. Orth for help with microscopy and the CU Microscopy Cores (supported by NIST-CU 70NANB15H226, NIH1S10RR026680-01A1) as well as the University of Colorado Cancer Center Genomics Core (supported by NIH-P30CA46934). We thank Theodore E. Ewachiw and Yang Zhao for their technical assistance.

**Funding:** The research was supported by NIH-T32GM008497 (J.R.W., E.D.N., T.O.V. and E.L.), NIH-F30NS093682 (J.R.W.), NIH-F30AR068881 (T.O.V.), NIH-GM045443 (R.P.), NIH-R35GM119575 (A.M.J.), Paul O’Hara II Seed Grant from ACS-IRG Grant Program (A.M.J.), NIH-AR049446 and NIH-AR070360 (B.B.O.), NIH-RM1-HG007735 (H.Y.C), Glenn Foundation for Biomedical Research (B.B.O.), and a Butcher Innovation Award NSF IGERT 1144807 (J.R.W. and T.O.V.), UC Boulder BSI and UROP programs (O.N.W.). R.P. and H.Y.C. are Investigators of the Howard Hughes Medical Institute.

**Author contributions:** J.R.W, O.N.W., T.O.V., B.B.O. and R.P. conceived and designed the research, wrote the manuscript and all authors edited drafts. J.R.W., O.N.W, B.P., N.D.B., T.E. A.C., R.O., and K.J. performed 10X sc-RNA-seq, and analyzed mouse regeneration and myotube formation experiments. E.N. and J.R.W. performed eCLIP and subsequent analysis. A.T., K.R.P., K.E.Y., E.L. provided analytical support for data analysis. T.R., H.Y.C., and A.J. provided scientific insights and materials.

**Declaration of Interests:** H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio, and an advisor of 10x Genomics, Arsenal Biosciences, and Spring Discovery.

**Data and materials availability:** All data is available in the main text or the supplementary materials. Sequencing data deposited as follows: 10X scRNA sequencing (awaiting accession number), Hnrnpa2b1 eCLIP (GSE106553), Hnrnpa2b1 WT and KO RNA sequencing (GSE152467).

## **SUPPLEMENTARY MATERIALS:**

Supplementary figures 1, 3, 4, 5, 6

Materials and Methods

Tables S1-S5

Supplementary References

## **MATERIALS AND METHODS**

**Mice:** Mice were bred and housed according to the National Institutes of Health guidelines for the ethical treatment of animals, in a pathogen free facility at the University of Colorado at Boulder. Wild-type mice were genotype C57BL/6 (Jackson Laboratories). Cells and tibialis anterior (TA) muscles were isolated from 3-6-month-old male and female wild-type mice.

**Mouse injuries:** Male and female mice between 3-6 months of age were anesthetized with isoflurane and the left TA muscle was injected with 50 $\mu$ L of 1.2% BaCl<sub>2</sub>. The injured and contralateral TA muscles were harvested at the indicated time points.

**TA collections and cell isolations:** TA muscles were dissected and placed into 400U/mL collagenase at 37°C for 1h with shaking and then placed into Ham's F-12C supplemented with 15% horse serum to inactivate the collagenase. Cells were passed through three strainers of 100µm, 70µm, and 40µm (BD Falcon) and flow through was centrifuged at 1500×g for 5 min and the cell pellets were re-suspended in Ham's F-12C. To remove dead cells and debris, cells were passed over the Miltenyi, dead cell removal kit columns (Cat# 130-090-101). To remove RBCs, cells were incubated with antiTer119 micro magnetic beads and passed over a Miltenyi column (Cat#130-049-901). For the adult and aged uninjured TAs 6 TA muscles (from 3 mice) were pooled together. For the injured TA muscles 2 TA muscles from 2 different mice were pooled together. Cells were then counted using a BioRad TC20 automated cell counter and processed with the 10X genomics single cell sequencing kit.

**Single Cell Sequencing:** To capture, label, and generate transcriptome libraries of individual cells we used the 10X genomics Chromium Single Cell 3' Library and Gel Bead Kit v2 (Cat #PN-120237) following the manufactures protocols. Briefly, the single cell suspension, RT PCR master mix, gel beads, and partitioning oil were loaded into a Single Cell A Chip 10 genomics chip, placed into the chromium controller, and the chromium Single Cell A program was run to generate GEMs (Gel Bead-In-EMulsion) that contain RT-PCR enzymes, cell lysates and primers for illumine sequencing, barcoding, and poly-DT sequences. GEMs are then transferred to PCR tubes and the RT-PCR reaction is run to generate barcoded single cell identified cDNA. Barcoded cDNA is used to make sequencing libraries for analysis with Illumina sequencing. We captured 1709 cells from young uninjured, 3459 from the aged uninjured, 5077 from the 4-d post injury and 2668 from the 7-d post injury. Sequencing was completed on an

Illumina NovaSeq 6000, using paired end 150 cycle 2x150 reads by the genomics and microarray core at the University of Colorado Anschutz Medical Campus.

**Single Cell Informatics:** Preprocessing was performed using Cellranger v3.0.1 (10X Genomics) count module was used for alignment using cellranger-mm10-3.0.0 refdata, filtering, barcode counting and UMI counting of the single cell FASTQs. Postprocessing was performed using Seurat according to standardized workflows (Butler et al., 2018; Stuart et al., 2019). In brief, using RStudio (v4.0.0), Seurat objects were created for each Cellranger processed sample by importing filtered\_gene\_bc\_matrices. Multiple Seurat objects then were merged, filtered, normalized, feature selected, scaled, and clustered. Non-linear dimensional reduction was performed using UMAP.

**RNA velocity** using the scVelo pipeline was performed according to standardized workflows (La Manno et al., 2018). In brief, Velocity was run in Python v3.6.3 using Samtools v1.8, cellranger-mm10-3.0.0 refdata, and masking repetitive elements to generate unspliced/spliced/ ambiguous read count Loom file for each 10X cellranger preprocessed library. Seurat objects of myogenic clusters (.rds files) were converted in RStudio to loom files using LoomR. In a virtual environment, Velocity Loom files were concatenated and merged with Seurat exported myogenic Loom in Scanpy. The scVelo pipeline was then used to perform log normalization and filtering. RNA velocity was then imputed using stochastic modeling. RNA velocities were projected on pre-computed UMAP embeddings and annotated clusters.

**RBP Engagement Scoring** was performed by first determining the raw unspliced reads for every gene at single cell level using Velocyto as above. Metadata containing raw reads counts and unspliced reads imputed using Velocyto were exported from Scanpy anndata objects for single cells contained within myogenic clusters were exported and reimported into RStudio (v.4.0.0). To calculate an RBP-RNA engagement score for a given RBP, CLIP target genes of a given RBP genes were subset and unspliced read counts and total RBP counts were normalized to total read counts in a given cell. Scoring then represents the fraction of normalized unspliced reads of target RNAs to normalized amounts of a given RBP. To extrapolate to a given cluster, scores were then summed across all cells in a given cluster and data was represented by the median of these scores for a given cluster of cells. Target genes were identified using publicly available eCLIP datasets for TDP-43.

**CIBERSORTx** of myogenic clusters was performed according to published workflows (Newman et al., 2019). In brief, a single cell reference txt file was created from our Seurat processed single cell RNA-seq dataset. A mixture file for bulk RNA sequencing samples was created using TPM values extracted from StringTie serve as input. A single cell reference txt matrix of cells organized by myogenic clusters was then used to train CIBERSORTx using default parameters (Table S3). Cell fractions were then imputed using 100 permutations for significance analysis.

**Immunofluorescence staining of tissue sections:** TA muscles were dissected, fixed on ice for 2h with 4% paraformaldehyde, and then transferred to PBS with 30% sucrose at 4°C overnight. Muscle was mounted in O.C.T. (Tissue-Tek®) and cryo-sectioned on a Leica cryostat to

generate 10µm thick sections. Frozen tissues and sections were stored at -80°C until immunofluorescent staining. Prior to immunofluorescent staining, tissue sections were post-fixed in 4% paraformaldehyde for 10 minutes at room temperature (RT) and washed three times for 5 min in PBS. Staining with the anti-Pax7, Laminin and Hnrnpa2b1 antibodies required heat-induced epitope retrieval, which was performed by placing post-fixed slides in citrate buffer, pH 6.0, and subjected them 6 min of high pressure-cooking in a Cuisinart model CPC-600 pressure cooker. For immunostaining, tissue sections were permeabilized with 0.25% Triton-X100 (Sigma) in PBS containing 2% bovine serum albumin (Sigma) for 60 min at RT. Incubation with primary antibody occurred at 4°C overnight followed by incubation with secondary antibody at room temperature for 1hr. Primary antibodies included anti-Pax7 (Developmental Studies Hybridoma Bank, University of Iowa, USA) at 1:750, rabbit anti-laminin (Sigma-Aldrich) at 1:200 and a mouse anti-Hnrnpa2b1 (ab6102, Abcam) at 1:200. Alexa-fluor secondary antibodies (Molecular Probes) were used at a 1:1000 dilution. For analysis that included EdU detection, EdU staining was completed prior to antibody staining using the Click-iT EdU Alexa fluor 488 detection kit (Molecular Probes) following manufacturer protocols. Sections were incubated with 1 µg/mL DAPI for 10 min at room temperature then mounted in Mowiol supplemented with DABCO (Sigma-Aldrich) or ProLong Gold (Thermo) as an anti-fade agent.

**Cell culture and growth conditions:** C2C12 myoblast cells: Immortalized murine myoblasts (American Type Culture Collection) were maintained on uncoated standard tissue culture plastic or gelatin-coated coverslips for imaging experiments at 37°C with 5% CO<sub>2</sub> in DMEM with 20% fetal bovine serum and 1% penicillin/streptavidin. To induce myoblast differentiation and fusion

into myotubes, C2C12 myoblasts at 80% confluence were switched to DMEM media supplemented with 5% horse serum, 1% penicillin/streptavidin and 1X Insulin-Transferrin-Selenium in DMEM.

**Hnrnpa2b1 CRISPR-Cas9 knockout and EdU incorporation:** CRISPR-Cas9 knockout was performed in C2C12 myoblasts. Single guide RNA (sgRNA) against Hnrnpa2b1 (5'-GAGTCCGCGATGGAG) were designed using (crispr.mit.edu) and cloned into pSpCas9(BB)-2A-Puro (PX459). C2C12 myoblasts were transfected with JetPrime using standard protocols. Myoblasts that integrated the CRISPR construct were selected with puromycin (1  $\mu$ g/mL) for one week. Myoblasts KO and WT clones were isolated using cloning ring and selectively detaching clonal populations via trypsinization. Clonal population KO was validated via immunofluorescence and western blotting against Hnrnpa2b1 with anti-Hnrnpa2b1 antibodies (ab6102 and ab31645). EdU incorporation: C2C12 myoblasts were incubated with 10uM 5-ethynyl-2'-deoxyuridine (EdU – Life Technologies) for three hours. Cells were washed, fixed, and stained using the methods described below.

**Immunofluorescence staining of cells:** C2C12 myoblast cells were washed with PBS in a laminar flow hood and fixed with 4% Paraformaldehyde for 10 min at room temperature in a chemical hood. Cells were permeabilized with 0.25% Triton-X100 in PBS containing 2% bovine serum albumin (Sigma) for 1 hour at RT. Cells were incubated with primary antibody at 4°C overnight, then incubated with secondary antibody at room temperature for 1hr. Primary antibodies included mouse anti-Hnrnpa2b1 (ab6102, Abcam) at 1:200, mouse-anti Myogenin (ab82843, Abcam) and a mouse anti-MHC MF-20 (Developmental Studies



Hybridoma Bank, University of Iowa, USA) at undiluted, neat concentration. Alexa fluor secondary antibodies (Molecular Probes) were used at a 1:1000 dilution.

**Microscopy and image analyses:** Images were captured on a Nikon inverted spinning disk confocal microscope. Objectives used on the Nikon were: 10x/0.45NA Plan Apo, 20x/0.75NA Plan Apo and 40x/0.95 Plan Apo. Confocal stacks were projected as maximum intensity images for each channel and merged into a single image. Brightness and contrast were adjusted for the entire image as necessary against secondary antibody treated control immunofluorescent sections. Cellprofiler was used to quantify IHC and IF images using custom analysis pipelines unless otherwise stated.

**Western blotting of cell and tissue lysates:** Western blot was performed according to standard protocols. Equal volumes (20  $\mu$ l) of fractions were then resolved on a 4–12% Bis-Tris SDS–PAGE gel and transferred to a nitrocellulose membrane (Bio-Rad). Primary antibodies included mouse anti-Hnrnpa2b1 (1:200; ab6102, Abcam) and Monoclonal rabbit anti-GAPDH (14C10) conjugated to HRP (1:2000; Cell Signaling, 3683S).

**Hnrnpa2b1 eCLIP sequencing:** C2C12 myoblasts were seeded at  $6 \times 10^6$  cells per 15 cm plate, grown 24hrs at 37C, 5% CO<sub>2</sub> and either harvested (undifferentiated myoblasts) or differentiated in differentiation media for 4 days. Hnrnpa2b1 enhanced CLIP (eCLIP) was performed according to established protocols.(Nguyen et al., 2018; Van Nostrand et al., 2016) In brief, Hnrnpa2b1-RNA interactions were stabilized with UV crosslinking (254 nm, 150mJ/cm<sup>2</sup>). Cell pellets were collected and snap frozen in liquid N<sub>2</sub>. Cells were thawed, lysed in eCLIP lysis

buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, and 1x protease inhibitor) and sonicated (Bioruptor). Lysate was RNase I (Ambion, 1:25) treated to fragment RNA. Protein-RNA complexes were immunoprecipitated using the rabbit polyclonal anti-Hnrnpa2b1 (ab31645, Abcam) antibody. One size-matched input (SMInput) library was generated per biological replicate using an identical procedure without immunoprecipitation. Stringent washes were performed as described, RNA was dephosphorylated (FastAP, Fermentas), T4 PNK (NEB), and a 3' end RNA adaptor was ligated with T4 RNA ligase (NEB). Protein-RNA complexes were resolved on an SDS-PAGE gel, transferred to nitrocellulose membranes, and RNA was extracted from membrane. After RNA precipitation, RNA was reverse transcribed using SuperScript IV (Thermo Fisher Scientific), free primer was removed, and a 3' DNA adapter was ligated onto cDNA products with T4 RNA ligase (NEB). Libraries were PCR amplified and dual-indexed (Illumina TruSeq HT). Pair-end sequencing was performed on Illumina NextSeq sequencer.

**eCLIP bioinformatic and statistical analysis:** Read processing and cluster analysis for Hnrnpa2b1 eCLIP was performed as previously described (Van Nostrand et al., 2016; Vogler et al., 2018). Read processing and cluster analysis for Hnrnpa2b1 eCLIP was performed as previously described. Briefly, 3' barcodes and adapter sequences were removed using standard eCLIP scripts. Reads were trimmed, filtered for repetitive elements, and aligned to the mm9 reference sequence using STAR. PCR duplicate reads were removed based on the read start positions and random sequence. Bigwig files for genome browser display were generated based on the location of the second of the paired end reads. Peaks were identified using the encode\_branch version of CLIPPER using the parameter “-s mm9.” Peaks were normalized

against size matched input by calculating fold enrichment of reads in IP versus input and were designated as significant if the number of reads in the IP sample was greater than in the input sample, with a Bonferroni corrected Fisher exact p-value less than  $10^{-8}$ .

**RNA-sequencing library preparation and sequencing:** C2C12 Hnrnpa2b1-WT and KO myoblasts were seeded at  $6 \times 10^6$  cells per 15 cm plate, grown for 24hrs at 37C, 5% CO<sub>2</sub> and differentiated via serum withdrawal and ITS supplementation for two days. Differentiated C2C12 myoblasts and myotubes were washed with PBS in a laminar flow hood and scraped from the tissue culture plate. Total RNA was extracted using a Qiagen RNeasy Plus Mini Kit, following manufacturer instructions. Isolated RNA quality was assessed by the CU Boulder BioFrontiers Next-Gen Sequencing core facility using an Agilent tape station. Isolated RNA was sent to the University of Colorado Cancer Center Genomics and Microarray Core for “Ribominus” Ribosomal RNA depletion, NGS library preparation and total RNA-sequencing.

**RNA-sequencing informatics:** All RNA sequencing data preprocessing was carried using standardized Nextflow RNA-sequencing (nf-core/rnaseq) with STAR alignment (genome GRCm39), transcriptome mapping with Salmon, and *in silico* ribosomal depletion using SortMeRNA. Differential gene expression was performed on feature counts tables using DESeq2 (Love et al., 2014). RNA splicing analysis was performed using Leafcutter (Li et al., 2018).

**CLIP/splicing cluster analysis:** The locations of splicing clusters generated by leafcutter and hnRNP A2/B1 eCLIP peaks from myotubes were each compared against genes in the

mm10 refGene table from the UCSC Genome browser to determine the number of clusters or peaks overlapping each gene. The overlap of each cluster or peak was then subdivided into coverage of either UTR or segments of the gene, divided into percentiles along the gene. These were normalized and plotted against each other using R and ggplot2. (Script attached below, R command available on request)

**Relative distance:** Relative distance between splicing clusters generated by leafcutter and hnRNP A2/B1 eCLIP peaks from myotubes were generated using the bedtools reldist command, using eCLIP peaks as the first input, and splicing clusters in genes with eCLIP peaks as the second input (Favorov et al., 2012).

**QAPA Poly-A Analysis:** As previously described, RNA sequencing reads were trimmed, *in silico* ribodepleted, and reads were mapped to an mm10 3'UTR annotation file using salmon v0.13.1. Quantification of Alternative Polyadenylation (QAPA) was then used to estimate relative alternative 3'UTR isoform usage (Ha et al., 2018). Lengthening or shortening of a gene's poly(A) tail was determined by first calculating a proximal poly(A) site usage (PPAU) defined as the percentage of reads mapping to the most proximal poly(A) site relative to reads mapping to the whole 3'UTR.  $\Delta$ PPAU was then calculated as the median PPAU A2B1 KO – median PPAU WT (three replicates per condition). A gene with a  $\Delta$ PPAU > 20 was defined as having a shortened poly(A) tail and a gene with a  $\Delta$ PPAU < -20 was defined as having a lengthened poly(A) tail.

## **Hnrnpa2b1 immunoprecipitation and qRT-PCR:** C2C12 Hnrnpa2b1-WT and KO myoblasts

were seeded at  $6 \times 10^6$  cells per 15 cm plate, grown 24hrs at 37C, 5% CO<sub>2</sub> and harvested as undifferentiated myoblasts. Myoblasts were lysed in a CHAPS-based buffer and pre-cleared using protein-A bound Dynabeads (Thermo, 10001D). Pre-cleared cell lysates were incubated with rabbit polyclonal anti-Hnrnpa2b1 antibody (ab31645, Abcam). Antibody-bound Hnrnpa2b1 was bound to protein-A Dynabeads overnight and magnetically isolated from the whole-cell lysate. Hnrnpa2b1-bound RNA was isolated from the Dynabead-Antibody-Hnrnpa2b1 complex via TRIzol RNA purification. cDNA libraries were created from purified RNA via oligo-DT priming and SuperScriptIII-enzyme reverse transcription. cDNA libraries were probed against Hnrnpa2b1-eCLIP hits chosen as a subset of the significantly enriched gene ontology terms via quantitative, real-time PCR. Primers used target GAPDH, Hnrnpa2b1, Prpf19, Snrnp70, Sfpq, Mbnl1, Hnrnpa3 and Mef2a. (Primer sequence in supplementary table 1). qRT-PCR was performed using SYBR-Green qRT-PCR reagent (BioRad) and fluorescent emission was measured using a BioRad CFX384 Real-Time PCR Detection system.

## **Supplementary References**

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.

Favorov, A., Mularoni, L., Cope, L.M., Medvedeva, Y., Mironov, A.A., Makeev, V.J., and Wheelan, S.J. (2012). Exploring Massive, Genome Scale Datasets with the GenometriCorr Package. *PLOS Comput. Biol.* 8, e1002529.

Ha, K.C.H., Blencowe, B.J., and Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* 19, 45.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastri, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498.

- Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* *50*, 151–158.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* *37*, 773–782.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* *177*, 1888–1902.e21.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* *13*, 508–514.
- Vogler, T.O., Wheeler, J.R., Nguyen, E.D., Hughes, M.P., Britson, K.A., Lester, E., Rao, B., Betta, N.D., Whitney, O.N., Ewachiw, T.E., et al. (2018). TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle. *Nature* *563*, 508–513.
- Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* *20*, 59.