UNIVERSITY OF CALIFORNIA SAN DIEGO

Alternative Polyadenylation in Hematopoiesis and Leukemia

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Amanda Gabrielle Davis

Committee in charge:

Professor Dong-Er Zhang, Chair
Professor Rafael Bejar
Professor Xiang-Dong Fu
Professor Jens Lykke-Andersen
Professor Lorraine Pillus

2020

The Dissertation of Amanda Gabrielle Davis is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

To my husband, who loved and supported me throughout my graduate career. He brightened every day, listened during the hard times, and spent countless hours waiting in the loading dock of Moores Cancer Center when experiments needed tending to. Without him, I would not have made it to the finish line.

To my family, who also loved and supported me by praying diligently for me throughout this journey.

EPIGRAPH

The loss of mystery has led to the loss of majesty.

The more we know, the less we believe.

No wonder there is no wonder.

We think we've figured it all out.

Strange, don't you think?

Knowledge of the workings shouldn't negate wonder.

Knowledge should stir wonder.

Who has more reason to worship than the astronomer who has seen the stars?

Than the surgeon who has held a heart?

Than the oceanographer who has pondered the depths?

[Than a molecular biologist who has studied the language of life?]

The more we know, the more we should be amazed.

Ironically, the more we know, the less we worship.

We are more impressed with our discovery of the light switch than with the one who invented electricity.

Call it cricket-brained logic.

Rather than worship the Creator, we worship the creation.

No wonder there is no wonder.

We've figured it all out.

Max Lucado

In the Grip of Grace

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Dong-Er Zhang for her support and mentorship throughout the duration of my graduate training. She provided an environment in which I learned to think like a scientist and gained confidence in my own ideas. I am also grateful for each and every lab member, past and present, that I had the privilege of working alongside in the lab over the years. I learned so much from each of you and am honored to call you colleagues and friends. I would also like to thank my committee members for their time, scientific insight, and overall career advice. Dr. Rafael Bejar, Dr. Xiang-Dong Fu, Dr. Jens Lykke-Andersen, and Dr. Lorraine Pillus, I am sincerely grateful.

I am also so thankful for my parents, siblings, and extended family for their love and support. Your prayers have carried me through some of the hardest moments of the PhD journey. Thank you to all my friends, near and far, for cheering me on and reminding me of what is truly important on days when I needed a different perspective. Most importantly, I am forever grateful for my husband Chris. Thank you for being by my side every step of the way with patience, love, and encouragement. I would not have made it to the finish line without you. And finally, I am thankful to God who created us all to be so incredibly complex. It has been a true joy catching a glimpse of this creativity through the lens of molecular biology.

Chapter 2, in part, has been submitted for publication by Davis AG, Johnson DT, Zheng D, Wang R, Jayne ND, Liu M, Stoner SA, Zhou JH, Ball ED, Tian B, and Zhang DE. The dissertation author is the primary investigator and writer of this manuscript. These studies were made possible because of a phenomenal collaboration with Dr. Bin

Tian's group at Rutgers New Jersey Medical School. Thank you for your preparation of the 3'READS libraries, data analysis, and scientific expertise of alternative polyadenylation. I would also like to thank the UCSD IGM core facility for RNA-sequencing services provided. Finally, I would like to thank the group at Sanford Consortium Human Embryonic Stem Cell Core Facility (HESCC) for FACS support.

Chapter 3, in part, has also been submitted for publication by Davis AG, Einstein JE, Zheng D, Jayne ND, Wang R, Fu XD, Tian B, Yeo GW, and Zhang DE. The dissertation author is the primary investigator and writer of this manuscript as well. For this project, I want to specifically thank Jaclyn Einstein and Dr. Gene Yeo for their collaboration with the CRISPR screen and eCLIP technical expertise. I will again thank our collaborators in Dr. Bin Tian's group for processing the 3'READS samples for this project. For FACS experiments, I want to thank Dennis in the Moores Cancer Center Flow core facility and again the team at HESCC.

VITA

2013 Bachelor of Science, Pepperdine University

2020 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Stoner, SA; Liu, K; Andrews, ET; Liu, M; Arimoto, K; Yan, M; **Davis, AG**; Weng, S; Dow, M; Xian, S; Dekelver, RC; Carter, H; and DE Zhang. The RUNX1-ETO target gene RASSF2 suppresses t(8;21) AML development and regulates Rac GTPase signaling. 2020. *Blood Cancer Journal* 10(16).

Weng, S; Matsuura, S; Mowery, C; Stoner, S; Lam, K; Ran, D; **Davis, AG**; Lo, M; and DE Zhang. Restoration of MYC-repressed targets mediates the negative effects of GM-CSF on RUNX1-ETO leukemogenicity. 2016. *Leukemia* 31(1): 159-169.

Brewster, J; Lumley, E; Osborn, A; Scott, J; **Scholl, AG**; Mercado, V; McMahan, Y; and Coffman, Z. Moderate endoplasmic reticulum stress activates a Perk and p38-dependent apoptosis. 2016. *Cell Stress and Chaperones* 22(1): 43-54.

Lam, K; Muselman, A; Du, R; Harada, Y; **Scholl, AG**; Yan, M; Matsuura, S; Weng, S; Harada, H; and DE Zhang. Hmga2 is a direct target gene of RUNX1 and regulates expansion of myeloid progenitors in mice. 2014. *Blood* 124(14): 2203-2212.

FIELDS OF STUDY

Unfolded protein response
Professor Jay Brewster, PhD

Hematology and myeloid malignancies
Professor Dong-Er Zhang, PhD

ABSTRACT OF THE DISSERTATION

Alternative Polyadenylation in Hematopoiesis and Leukemia

by

Amanda Gabrielle Davis

Doctor of Philosophy in Biology

University of California San Diego, 2020

Professor Dong-Er Zhang, Chair

Post-transcriptional regulation by RNA-binding proteins (RBPs) is an important layer of gene regulation implicated in both healthy hematopoiesis and hematologic malignancy. Among post-transcriptional mechanisms, alternative polyadenylation (APA) regulates gene expression and function, mediating normal cellular differentiation and malignant transformation across cellular systems. In hematology, APA plays a critical role in lymphocyte maturation and dysregulation contributes to multiple myeloma and lymphocytic leukemia. Despite its documented importance in immune cells, it is unknown whether APA plays a critical role in myeloid malignancy or in healthy

hematopoietic stem cell (HSC) maintenance. Furthermore, RBPs that regulate APA in hematologic systems have not been identified.

Here, we first addressed the prevalence and global function of APA in myeloid malignancy. We compared poly(A) site usage in acute myeloid leukemia (AML) blasts to usage in healthy hematopoietic stem and progenitor cells (HSPCs), uncovering global patterns and individual leukemia-promoting genes altered in malignancy. By targeting the RBP and APA regulator FIP1L1, we reversed the global trends in patients and observed cellular differentiation across diverse AML subtypes by disrupting leukemogenic signaling networks. In t(8;21) AML, we validated APA regulation of AML1-ETO, showing for the first time that expression of a prominent oncofusion is sensitive to this mode of post-transcriptional regulation. Altogether, our work defines a critical role for APA in AML and illuminates a new pathway that may be exploited for differentiation therapy in patients.

We also studied the role of APA in healthy HSC pool maintenance by focusing on APA regulation of the critical hematopoietic transcription factor RUNX1. Polyadenylation upstream of the most distal 3'UTR produces a C-terminally truncated protein that antagonizes the pro-differentiation function of full-length RUNX1 in HSCs. We modeled this relevant APA event using a dual fluorescent minigene reporter and used this reporter in a CRISPR screen targeting RBPs. We identified HNRNPA1 and KHDRBS1 as regulators of RUNX1 APA, assigning a new role to these RBPs in HSC fate. Overall, our work highlights the intersection between post-transcriptional regulation and transcription factor function in healthy hematopoiesis.

Chapter 1: Introduction

1.1 Alternative Polyadenylation

1.1.1 Cleavage and polyadenylation of mRNA

Polyadenylation is a post-transcriptional processing event that consists of endonucleolytic cleavage and the addition of an untemplated string of adenosine nucleotides to the 3'end of almost all eukaryotic mRNA transcripts (1). This string of adenosines is called the poly(A) tail and addition is necessary for mRNA stability, export from the nucleus, and translation in the cytoplasm (2). It has recently been shown that most eukaryotic transcripts can be cleaved and polyadenylated at multiple locations, a phenomenon termed alternative polyadenylation (APA) (3). APA diversifies the transcriptome, producing mRNA isoforms that differ in length of the 3' untranslated region (3'UTR) or the coding sequence (CDS) of the transcript. Considering the prevalence of APA potential in the eukaryotic transcriptome, it is not surprising that poly(A) site usage differs based on cellular context (4, 5) and dysregulation is reported in malignancy (6-8). Consequently, ongoing research is focused on elucidating the functional impact and mechanism of APA in various cellular states.

1.1.2 Polyadenylation cis-acting elements and core polyadenylation machinery

Cleavage and polyadenylation occur co-transcriptionally through RNA polymerase II recruitment of RNA-binding proteins (RBPs) to the appropriate cis-acting elements that define the 3'end of an mRNA transcript (9, 10). Among cis-acting

elements, the poly(A) signal (PAS) hexamer plays a central role in directing cleavage and polyadenylation and is located approximately 15-30 nucleotides upstream of the cleavage site (11). The canonical PAS sequence is AAUAAA (12), with variants seen less frequently throughout the genome and corresponding to weaker cleavage efficiency (13, 14). Notably, the canonical PAS is enriched at distal sites, whereas variants are more frequently seen at proximal sites, suggesting an overall trend in proximal sites being weaker than distal sites (3, 13).

In addition to the PAS hexamer, there are upstream and downstream cis-acting elements that contribute to the positioning of polyadenylation machinery and cleavage at a poly(A) site. U and G/U rich downstream elements (DSE) are located ~40 nucleotides downstream of the cleavage site (15, 16). Additionally, upstream elements (USE) consisting primarily of UGUA motifs are located 40-100 nucleotides upstream of the PAS hexamer (17). While most transcripts have these additional cis-acting elements, some poly(A) sites lack one or the other (18), highlighting the need for further research on sequence variants or mechanisms of cleavage and polyadenylation in their absence.

The three major cis-acting elements (PAS, DSE, and USE) uniquely recruit three protein complexes essential for effective cleavage and polyadenylation. Altogether, the core machinery consists of twenty protein including RBPs, enzymes, and scaffolding proteins. The first complex is the cleavage and polyadenylation specificity factor (CPSF) complex (19). This complex is responsible for PAS recognition (20) and is comprised of six proteins: CPSF160 (CPSF1), CPSF100 (CPSF2), CPSF73 (CPSF3), CPSF30 (CPSF4), WDR33, and FIP1L1. While it was originally thought that CPSF160 directly bound the PAS hexamer (21, 22), recent work implicates WDR33 and CPSF30 in direct

PAS binding (23, 24). The CPSF73 subunit is the endonuclease that cleaves mRNA transcripts prior to poly(A) tail addition (25). FIP1L1 is also an RBP, binding to U-rich regions both upstream and downstream of the PAS hexamer (26-28), though its contribution to poly(A) site selection is still unclear.

The second core complex is the cleavage stimulation factor (CSTF) complex, which binds the U and GU-rich DSE (29). This complex contains three proteins: CSTF50 (CSTF1), CSTF64 (CSTF2) and CSTF77 (CSTF3) (30), held together by homodimer interactions (31). CSTF64, or its paralogue CSTF64T, directly binds to the DSE (29), playing functionally redundant roles in poly(A) site selection (32).

The third core complex, mammalian core factor I (CFIm), binds the UGUA USE (33). CFIm is a tetramer that consists of two RNA-binding CFIm25 (CPSF5) subunits and either two CFIm68 (CPSF6) or CFIm59 (CPSF7) subunits (34). Because CFIm25 dimerizes, it can bind two UGUA motifs simultaneously and regulate poly(A) site selection via RNA looping (33, 34).

Upon assembly of these three main protein complexes, additional proteins are recruited, contributing to proper cleavage and polyadenylation. Among these accessory proteins are CLP1 and PCF11, together classified as the mammalian core factor II (CFIIm) complex. The role of these proteins in cleavage and polyadenylation is still unclear (35). Symplexin functions as a scaffolding protein, physically linking the CPSF and CSTF complexes (36). Additionally, RBBP6 binds to the CSTF complex and regulates polyadenylation of transcripts containing AU-rich elements (37, 38). Finally, poly(A) polymerase (PAP) adds adenosine to a growing poly(A) tail and poly(A) binding proteins (PABPN and PABPC) stabilize the growing tail, regulating its final length (39).

1.1.3 Functional impact of alternative polyadenylation

Considering the prevalence of multiple poly(A) sites in human transcripts (3), the functional impact of differential poly(A) site usage by APA has been a growing area of research. So far, the impact of APA on individual genes is conceptually divided into the two major APA categories: 3'UTR-APA and CDS-APA (40). For transcripts that belong to the former category, two or more poly(A) sites are in the most distal 3'UTR of a gene and changes in usage only affect 3'UTR length, not the CDS of the transcript. Conversely, transcripts regulated by CDS-APA contain poly(A) sites upstream of the most distal 3'UTR. Unlike 3'UTR-APA, usage of an upstream poly(A) site alters the CDS of a transcript.

3'UTR-APA regulates transcripts by remarkably diverse mechanisms, despite having no impact on the protein produced from a regulated transcript. Since the 3'UTR contains numerous sequences that bind to miRNAs and RBPs, changes in 3'UTR length can alter mRNA stability, translation efficiency, localization, and even protein complex formation of the protein encoded by the transcript. Early work implicated 3'UTR length in mediating gene expression predominantly through presence or absence of sequences that bind miRNAs. Shorter transcripts are predicted to have better stability and protein output because they escape miRNA-mediated decay (6, 41). Since RNAs compete with one another for binding of miRNAs, loss of miRNA mediated repression of shortened transcripts promotes downregulation of competing-endogenous RNAs (ceRNAs) that still have the seed sequence for miRNA binding (42). However, this association between 3'UTR length and transcript expression is too simplistic since 3'UTR length is not globally predictive of gene expression (43, 44). Indeed, even miRNA

interactions are not straightforward as functional miRNA target sites are enriched near 3'UTR boundaries (45). Consequently, 3'UTR shortening can enhance miRNA targeting of seed sequences immediately upstream of a proximal poly(A) site (46), making a shorter 3'UTR less stable than a longer 3'UTR due to enhanced miRNA targeting.

In addition to miRNA-based regulation, RBP interactions with the 3'UTR of transcripts can also impact expression (47) along with a host of additional interesting features. Regarding expression, RBP interactions have both positive and negative impacts on transcript stability and protein production. A study in triple negative breast cancer reveals upregulation of the NRAS and c-JUN oncogenes by 3'UTR shortening and removal of PUM binding sequences, an RBP that inhibits translation and enhances mRNA degradation (48). Conversely, the longer 3'UTR of BDNF is bound by HuD, an RBP that stabilizes this transcript in the brain (49). RBPs also regulate mRNA localization, which in turn concentrates protein production. This mechanism is particularly important in neurons where binding of distal 3'UTRs by MBNL1 and MBLN2 mediates movement to the neurites, an essential event for neuronal differentiation (50). Furthermore, RBP-3'UTR interactions allow the 3'UTR to function as a scaffold, mediating protein-protein interactions of the newly translated protein. This specific function is implicated in 3'UTR-dependent CD47 localization to the plasma membrane (51) and 3'UTR-dependent formation of unique BIRC3 protein complexes (52). With the expansion of our understanding of RBP function, we anticipate uncovering additional mechanisms regulated by 3'UTR-APA.

CDS-APA events have a more straightforward impact on affected transcripts. In some cases, CDS-APA diversifies the proteome, generating C-terminally truncated

protein isoforms that lack protein domains and thus have different functions or localization than full-length proteins. A classic example of this is the impact of CDS-APA on IgM heavy chain in activated B cells. A switch from distal poly(A) site usage to an upstream, intronic poly(A) site converts IgM heavy chain from a membrane-bound to a secreted form, a switch that is necessary for immune activation (53). However, CDS-APA does not always produce functional protein isoforms. In some cases, the truncated protein product is non-functional and therefore CDS-APA also regulates overall gene expression (54).

Altogether, our understanding of the functional impact of APA on target transcripts has been expanding, but there is still a lot more that needs to be elucidated. Considering the diverse impacts described here, the importance of this post-transcriptional mechanism in various cellular systems is expected to also expand.

1.1.4 Alternative polyadenylation in normal cellular differentiation and disease

Global trends in poly(A) site usage have been observed in a variety of cellular contexts, suggesting a functional consequence of global transcript shortening or lengthening on cellular phenotypes.

Firstly, poly(A) site usage is linked to normal cellular differentiation. During differentiation, a trend in transcript lengthening has been observed across species (55-57). Furthermore, induction of pluripotency correlates with transcript shortening in somatic cells (58). In agreement with these trends, tissue-specific APA profiles have revealed transcriptomes characterized by proximal poly(A) site usage in embryonic stem cells and immature tissues such as the ovaries and testes (4, 5). Conversely,

tissues consisting mostly of differentiated cells, such as in the brain, have transcriptome profiles characterized by distal poly(A) site usage. Recent studies have added a layer of complexity to these observations as transcript length also correlates with cellular localization, with longer transcripts remaining in the neuron soma and shorter transcripts being transported to neurites (59).

Global APA profiles are also correlated with cellular activation. Transcript shortening is observed upon immune cell activation, with a concurrent increase in cellular proliferation (41). Shortening is also seen upon neuronal cell activation (60, 61) and in response to extracellular stress (62), highlighting a likely role of APA in regulating changes to cellular states in response to external stimuli.

Considering these trends, it is therefore not surprising that global shifts in poly(A) site usage have been documented in disease, particularly in cancer. A global shift from distal to proximal poly(A) site usage has been reported across cancer types (6, 8, 63-65), influencing cellular proliferation and maturation. While these trends are generally accepted in the field, there are notable exceptions in hematologic malignancies. In multiple myeloma (MM) cells, global loss of intronic poly(A) site usage contributes to transcript lengthening and to disease phenotypes (66). In chronic lymphocytic leukemia (CLL), CDS-lengthening is not observed (54), however global 3'UTR lengthening is reported (52). Furthermore, even when global transcript shortening is observed in cancer cells, there are also many transcripts that behave opposite the global trend and represent dysregulated APA in disease (67).

Overall, the observation of these patterns in healthy systems and in malignancy highlight a global role for APA regulation in defining cellular states. However, because

the impact of APA on individual transcripts is so variable and global trends do not represent the entire set of transcripts impacted by APA, more work is needed to understand how global patterns influence cellular biology.

1.1.5 Mechanisms of alternative polyadenylation

Considering these global shifts of poly(A) site usage in healthy and malignant systems, the identification of global APA regulators in different cellular contexts is also an area of active research.

As expected, polyadenylation machinery members were first hypothesized to contribute to global shifts in APA. In line with this hypothesis, 3'UTR shortening in proliferating cells is partly mediated by upregulation of numerous polyadenylation machinery members, commonly regulated by E2F transcription factors (64). This observation supports a model whereby APA is predominantly regulated by the concentration of core machinery in the cell, with low levels correlating with reduced overall polyadenylation and enhanced distal poly(A) site usage. However, a careful study on the impact of each core machinery protein on overall poly(A) site usage reveals divergent roles on APA. While some members promote proximal poly(A) site usage (PCF11 and FIP1L1), others promote distal poly(A) site usage (CFIm25, CFIm68, PABPN1, and PABPC1) (68). Indeed, these ascribed roles have held true in other cellular contexts. PCF11 knockdown induces transcript lengthening and differentiation of neuroblastoma cells (69) and FIP1L1 knockdown similarly promotes transcript lengthening and differentiation of murine embryonic stem cells (28). Conversely, CFIm25 downregulation is most famously implicated in carcinogenesis. Downregulation

in glioblastoma (63), osteosarcoma (70), and lung cancer (71) promotes global 3'UTR shortening and contributes to proliferation of tumor cells. Additionally, PABPN1 downregulation in an osteosarcoma cell line promotes global 3'UTR shortening, a phenotype also mediated by mutant PABPN1 in a mouse model of oculopharyngeal muscular dystrophy (OPMD) (7). Furthermore, members of the CSTF complex, CSTF2 (8) and CSTF3 (48), have also been implicated in cancer development, though their upregulation contributes to 3'UTR shortening and disease. In a healthy context, CSTF2 upregulation similarly increases proximal poly(A) site usage of the IgM heavy chain, mediating the switch from membrane-bound to its secreted isoform during B-cell differentiation into plasma cells (31). Overall, polyadenylation machinery plays some role in global APA regulation.

In addition to core machinery members, splicing factors and additional RBPs are implicated in APA. Among the former category, the U1 snRNP plays a ubiquitous role in protecting transcripts from premature cleavage and polyadenylation near transcription start sites, a phenomenon called telescripting (72, 73). In the latter category, some RBPs modulate poly(A) site usage via competition with core APA machinery for cis-acting element binding: TDP-43 (74) and ELAVL1/HuR (75). Others regulate APA due to position-specific binding near poly(A) sites: CPEB1 (76), MBNL (77), FUS (78), PTBP1 (79), PCBP (80), and HNRNPC (81). Considering the unique mode of regulation for each of these additional RBPs, our understanding of global APA regulation in normal cellular differentiation and disease is still limited. The discovery of additional RBPs linked to APA is expected.

1.1.6 Alternative polyadenylation in hematology

The importance of APA in the field of hematology has long been appreciated. Some of the first examples of individual gene APA events occur in lympohocytes: (1) CDS-APA of IgM heavy chain in B cell differentiation to plasma cells (31, 53) and (2) CDS-APA of NF-ATc in the transition from naïve T to T effector cells (82). In agreement with these early individual gene studies, CDS-APA isoform regulation is prevalent in immune cells (66, 83) and global shifts in APA have also been reported in immune cell activation (41). Finally, APA profiles have been modestly studied in both lymphocytic leukemia (52, 54) and multiple myeloma (MM) (66).

Despite the importance of APA in lymphocytes, studies have not been performed on the role of APA in myeloid leukemia or in hematopoietic differentiation outside of lymphocyte maturation. The following two studies were aimed at filling these knowledge gaps. In the first project, we profiled APA in acute myeloid leukemia (AML) and implicated global APA dysregulation in blocking differentiation of leukemia cells. We targeted the APA regulator FIP1L1 and saw a reversal in the APA trends of patients, followed by differentiation of leukemia cells. We also validated APA regulation of additional transcripts relevant to leukemogenesis, including the prominent oncofusion, AML1-ETO. In the second study, we focused on the gene-specific APA of RUNX1, an event that is regulated in normal hematopoietic differentiation and is specifically important for hematopoietic stem cell (HSC) self-renewal versus differentiation. We identified two RBPs, HNRNPA1 and KHDRBS1, that regulate this event and likely play a more general role in APA regulation and normal hematopoietic differentiation.

Chapter 2: Alternative polyadenylation dysregulation contributes to the differentiation block of acute myeloid leukemia

Post-transcriptional regulation has emerged as a driver for leukemia development and an avenue for therapeutic targeting. Among post-transcriptional processes, alternative polyadenylation (APA) is globally dysregulated across cancer types. However, limited studies have focused on the prevalence and role of APA in leukemia. Furthermore, it is poorly understood how altered poly(A) site usage of individual genes contributes to malignancy or whether targeting global APA patterns might alter oncogenic potential. By performing 3'RNA sequencing on acute myeloid leukemia (AML) patient samples and healthy hematopoietic stem and progenitor cells (HSPCs), we show that patient cells exhibit global 3' untranslated region (UTR) shortening and coding sequence (CDS) lengthening due to differences in poly(A) site usage. Among APA regulators, *FIP1L1* expression correlated with the degree of APA dysregulation and knockdown of this RNA-binding protein (RBP) reversed the global trends seen in patients. Importantly, FIP1L1 knockdown induced differentiation of t(8;21) cells by promoting 3'UTR lengthening and downregulation of fusion oncoprotein AML1-ETO. In non-t(8;21) cells, knockdown also promoted differentiation by attenuating mTORC1 signaling and reducing MYC protein levels. Our study provides mechanistic insight into the role of APA in AML pathogenesis and indicates that targeting global APA patterns can overcome the differentiation block of AML patients.

2.1 Introduction

Post-transcriptional regulation is an emerging field of study in both normal and malignant hematopoiesis (84). Among post-transcriptional processes, splicing regulation has been the predominant focus, owing to the high instance of splice factor mutations across hematologic malignancies (85). Additionally, RNA-editing (86), methylation (87), and miRNA regulation (88) have been linked to the proper balance of hematopoietic stem cell (HSC) self-renewal versus differentiation. Despite the importance of post-transcriptional regulation in hematopoietic systems, there has been comparatively little attention paid to the process of alternative polyadenylation (APA) and it is unknown whether regulators of polyadenylation contribute to healthy hematopoiesis or leukemia.

APA is a widespread post-transcriptional mechanism, with the potential to regulate most human genes (3). Differences in usage of poly(A) sites in the most distal 3’ untranslated region (3’UTR) can alter transcript stability (6, 41), localization (51), translation efficiency (48), and even protein complex formation (52). Poly(A) site usage upstream of the 3'most exon, mostly in introns, additionally diversifies the proteome by producing C-terminally truncated proteins with distinct functions (54, 66). Importantly, APA plays a critical role in normal cellular differentiation and cancer transformation. Global poly(A) site profiling reveals that gene expression signatures shift from more proximal to distal poly(A) site usage during normal cellular differentiation (55-57). The opposite shift is reported during cellular transformation, linking transcript shortening to proliferative cellular states (8, 41, 63-65).

So far, the mechanistic link between APA alteration and oncogenic transformation is incomplete. One common explanation is that 3’UTR shortening

contributes to oncogene upregulation when affected transcripts evade normal miRNA-mediated transcript degradation (6, 41). Additionally, global coding sequence (CDS) shortening by enhanced usage of intronic poly(A) sites is reported to inactivate tumor suppressor genes in cancer cells (54). While these explanations are true for some transcripts, oncogenes do not exclusively have shorter 3'UTRs, nor do tumor suppressors solely have shorter CDS due to APA. These examples have shaped our understanding of APA in cancer transformation, but they only explain a subset of the wide breadth of transcripts affected by APA, both with and against the reported global trends. Consequently, we still do not fully understand the collective function of widespread APA events or how global dysregulation contributes to transformation.

Acute myeloid leukemia (AML) is a cancer characterized by a hallmark block in differentiation (89). As such, there has been a lot of clinical interest in developing therapies that specifically target this disease phenotype and promote differentiation of leukemia cells. Among reported differentiating agents, success has been limited to all-trans retinoic acid (ATRA) (90) and arsenic trioxide (ATO) (91) in acute promyelocytic leukemia (APL) patients, and more recently isocitrate dehydrogenase (IDH) inhibitors in IDH mutant AML patients (92-94). Despite their promising efficacy, the percentage of patients that can benefit from these therapies is limited. Therefore, the identification of additional, targetable pathways that mediate the differentiation block of AML are needed.

Here, we profiled poly(A) site usage in AML patient blasts compared to healthy hematopoietic stem and progenitor cells (HSPCs) and report global dysregulation of APA in myeloid leukemia, affecting genes that contribute to leukemic phenotypes.

Among the core cleavage and polyadenylation factors, we found that *FIP1L1* mRNA expression was most correlated with 3'UTR shortening in our patient cohort, assigning a new role to the gene in AML. Furthermore, we disrupted global APA dysregulation by FIP1L1 knockdown and observed leukemia cell differentiation, supporting an underappreciated role of APA in blocking normal maturation of cancer cells. Additionally, we identified 3'UTR-APA regulation of AML1-ETO, a key AML oncofusion that contributes to this observed differentiation block. Finally, by disrupting global APA patterns, we detected downregulation of c-MYC and mTORC1 carcinogenic pathways across diverse mutational contexts in AML, highlighting a critical role of dysregulated APA in maintenance of leukemia cell phenotypes. Our findings underline the importance of post-transcriptional mechanisms in leukemia development and introduce APA as a putative therapeutic target for inducing differentiation in AML patients.

2.2 Results

2.2.1 Alternative polyadenylation is dysregulated in acute myeloid leukemia (AML) patients, affecting oncogenic pathways.

To assess the prevalence of APA dysregulation in AML, we first performed 3'RNA sequencing (3'READS) (95, 96) and standard RNA-sequencing on enriched AML patient blasts (CD34+) and healthy hematopoietic stem and progenitor cells (HSPCs) (**Figures 2.1A and 2.1B**). Since AML blasts are heterogeneous, we selected four patient samples that commonly carried the t(8;21) translocation generating the AML1-ETO fusion oncoprotein, and four non-t(8;21) samples with variable mutations (**Table 2.1**). Compared to healthy HSPCs, both groups of patients exhibited dysregulated APA of numerous genes, with an overall trend of 3'UTR shortening and CDS lengthening (**Figure 2.2A**). CDS lengthening was similarly seen in multiple myeloma samples (66), but both trends are opposite those in chronic lymphocytic leukemia (CLL) samples (52, 54), highlighting the importance of disease-specific poly(A) profiling. Despite their mutational differences, many genes were similarly shortened or lengthened in t(8;21) and non-t(8;21) patients, suggesting similarity in the groups of genes that are most dynamically regulated by APA in AML (**Figure 2.2B**).

To evaluate whether the observed changes in APA might contribute to pathogenesis, we performed pathway enrichment on the common genes that exhibited 3'UTR shortening or CDS lengthening in our patient cohort. Significantly enriched pathways included those related to cell cycle, differentiation, and oncogenic signaling pathways, supporting a likely role of APA in leukemia (**Tables 2.2 and 2.3**).

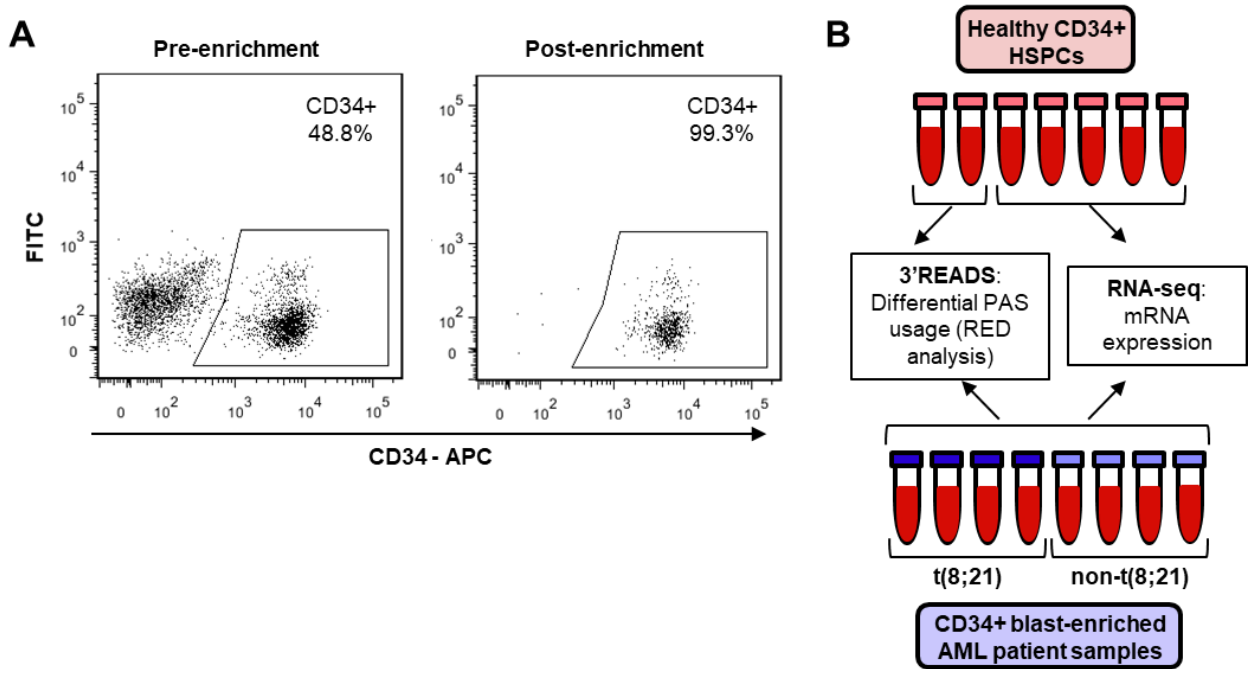


Figure 2.1. Preparation of primary HSPC and AML patient samples for 3'READS and standard RNA-sequencing.

(A) Representative FACS plots of pre- and post-CD34 bead enrichment from one AML patient sample used for 3'READS and RNA-sequencing. **(B)** Schematic of AML patient samples and healthy HSPCs utilized for 3'READS and/or RNA-sequencing.

Table 2.1. Primary HSPC and AML patient sample information.

Sample ID	Karyotype	cKIT	FLT3-ITD	RNA-seq	3'READS
nHSPC 1	Normal	N/A	N/A	No	Yes
nHSPC 2	Normal	N/A	N/A	No	Yes
nHSPC 3	Normal	N/A	N/A	Yes	No
nHSPC 4	Normal	N/A	N/A	Yes	No
nHSPC 5	Normal	N/A	N/A	Yes	No
nHSPC 6	Normal	N/A	N/A	Yes	No
nHSPC 7	Normal	N/A	N/A	Yes	No
AML Pt1	48, +2, +8, inv(16)	(-)	(-)	Yes	Yes
AML Pt2	48, +13, +19, i(17q)	(-)	(-)	Yes	Yes
AML Pt3	46, del(5q), gain(11q)	(-)	(-)	Yes	Yes
AML Pt4	Normal	(-)	(+)	Yes	Yes
t(8;21) Pt1	45, -Y, t(8;21)(q22;q22)	(-)	(-)	Yes	Yes
t(8;21) Pt2	46, t(8;21)(q22;q22)	D816V	(-)	Yes	Yes
t(8;21) Pt3	46, t(8;21)(q21.3;q22)	(-)	(-)	Yes	Yes
t(8;21) Pt4	46, t(8;21)(q22;q22)	(-)	(+)	Yes	Yes

Figure 2.2. Global trends of altered poly(A) site usage in AML patient blasts compared to healthy HSPCs.

(A) Scatter plots showing the change in expression of proximal poly(A) site isoform (x-axis) and distal poly(A) site isoform (y-axis), per gene, in t(8;21) AML patient blasts (top) or non-t(8;21) patients (bottom) compared to healthy HSPCs. Significant APA events ($p < 0.05$, Fisher's exact test) are classified and divided by type: tandem 3'UTR APA (left) and coding sequence APA (right). Each dot corresponds to a single gene. Blue dots indicate significantly more proximal poly(A) site (pPAS) usage. Red dots indicate significantly more distal poly(A) site (dPAS) usage. **(B)** Venn diagrams showing the overlap of genes with 3'UTR shortening and CDS lengthening in t(8;21) AML blasts and non-t(8;21) AML blasts compared to healthy HSPCs.

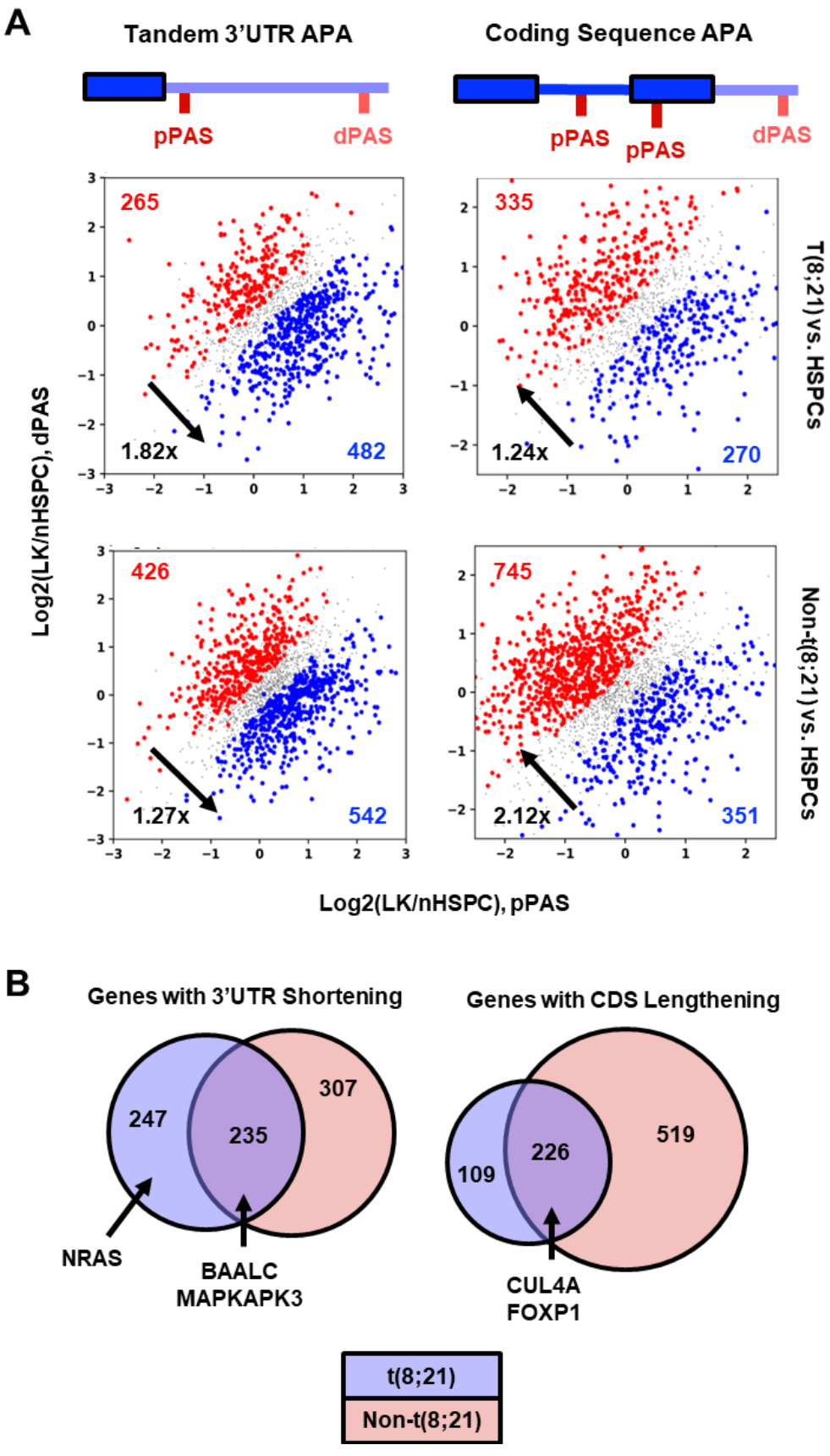

A
Tandem 3'UTR APA
Coding Sequence APA
pPAS
dPAS
pPAS
pPAS
dPAS
265
1.82x
482
335
1.24x
270
426
1.27x
542
745
2.12x
351
T(8;21) vs. HSPCs
Non-t(8;21) vs. HSPCs
Log2(LK/nHSPC), dPAS
Log2(LK/nHSPC), pPAS
B
Genes with 3'UTR Shortening
247
235
307
NRAS
BAALC
MAPKAPK3
Genes with CDS Lengthening
109
226
519
CUL4A
FOXP1
t(8;21)
Non-t(8;21)

Table 2.2. Enriched pathways of genes with 3'UTR shortening in AML.

Term	Description	-log10(p-value)
R-HSA-8953854	Metabolism of RNA	7.38753
GO:0031396	Regulation of protein ubiquitination	6.14594
R-HSA-72187	mRNA 3'-end processing	5.86472
GO:0032984	Protein-containing complex disassembly	5.30595
hsa04141	Protein processing in endoplasmic reticulum	5.30554
GO:0016241	Regulation of macroautophagy	5.19119
GO:0032388	Positive regulation of intracellular transport	5.09806
R-HSA-1640170	Cell Cycle	4.79591
GO:0002757	Immune response-activating signal transduction	4.78658
GO:0031647	Regulation of protein stability	4.64289
R-HSA-5663202	Diseases of signal transduction	4.57846
R-HSA-1257604	PIP3 activates AKT signaling	3.61271

Table 2.3. Enriched pathways of genes with CDS lengthening in AML.

Term	Description	-log10(p-value)
GO:1901987	Regulation of cell cycle phase transition	5.58471
R-HSA-5653656	Vesicle-mediated transport	4.76409
GO:0030522	Intracellular receptor signaling pathway	4.19929
GO:0043124	Negative regulation of I-kappaB kinase/NF-kappaB signaling	4.05632
GO:0043414	Macromolecule methylation	3.8212
GO:0030851	Granulocyte differentiation	3.62986
GO:0006369	Termination of RNA polymerase II transcription	3.52906
GO:0097190	Apoptotic signaling pathway	2.82508

2.2.2 Leukemia-promoting genes are among those with dysregulated APA in AML patients.

We next sought to identify specific genes with altered poly(A) site usage that might directly contribute to disease. 3'UTR-APA regulates gene expression by altering the presence of 3'UTR regions that bind miRNAs or RNA-binding proteins (RBPs), which in turn modulates transcript stability and/or translation efficiency (47, 55). In cancer, global 3'UTR shortening by APA is reported to induce oncogene upregulation by transcript evasion of miRNAs or suppressive RBPs (6, 41, 48, 97). Among genes that exhibited 3'UTR shortening in our dataset, we identified *BAALC*, a negative prognostic marker in AML (98, 99); *MAPKAPK3*, a MAP kinase signaling node that contributes to the differentiation block characteristic of t(8;21) AML (100); and *NRAS*, an oncogene commonly mutated in AML (101) (**Figures 2.3A, 2.3B, and 2.3C**). Since 3'UTR length is not always predictive of protein output (43, 44), we next tested whether 3'UTR length can contribute to expression of these leukemia-promoting genes. We subcloned the short and long 3'UTR variants downstream of renilla luciferase in a dual luciferase reporter and compared the effect on protein output (**Figure 2.3D**). For both genes, the shorter 3'UTR produced significantly more renilla protein than the longer 3'UTR, confirming a direct role of APA on gene expression (**Figure 2.3E**).

Like 3'UTR-APA, CDS-APA can also impact gene expression when premature cleavage and polyadenylation results in a truncated protein product that lacks crucial functional domains (54, 73). In cancer, global CDS shortening by intronic polyadenylation inactivates tumor suppressor genes (54). We observed the opposite overall trend of CDS lengthening in our AML patient 3'READS dataset and reasoned

that activation of oncogenes is also a possibility. Among genes with CDS lengthening in patient blasts compared to healthy HSPCs, we identified *CUL4A*, an oncogenic E3 ubiquitin ligase that promotes proliferation and blocks differentiation in hematopoietic models (102, 103), and *FOXP1*, a transcription factor that suppresses cell cycle inhibitors in leukemia (104) (**Figures 2.4A and 2.4B**). Usage of the proximal, intronic poly(A) site in each case would result in a severely truncated protein that is expected to be dysfunctional (66). Altogether, we saw pronounced differences in poly(A) site usage between patient blasts and healthy controls and identified specific genes that could reasonably contribute to pathogenesis.

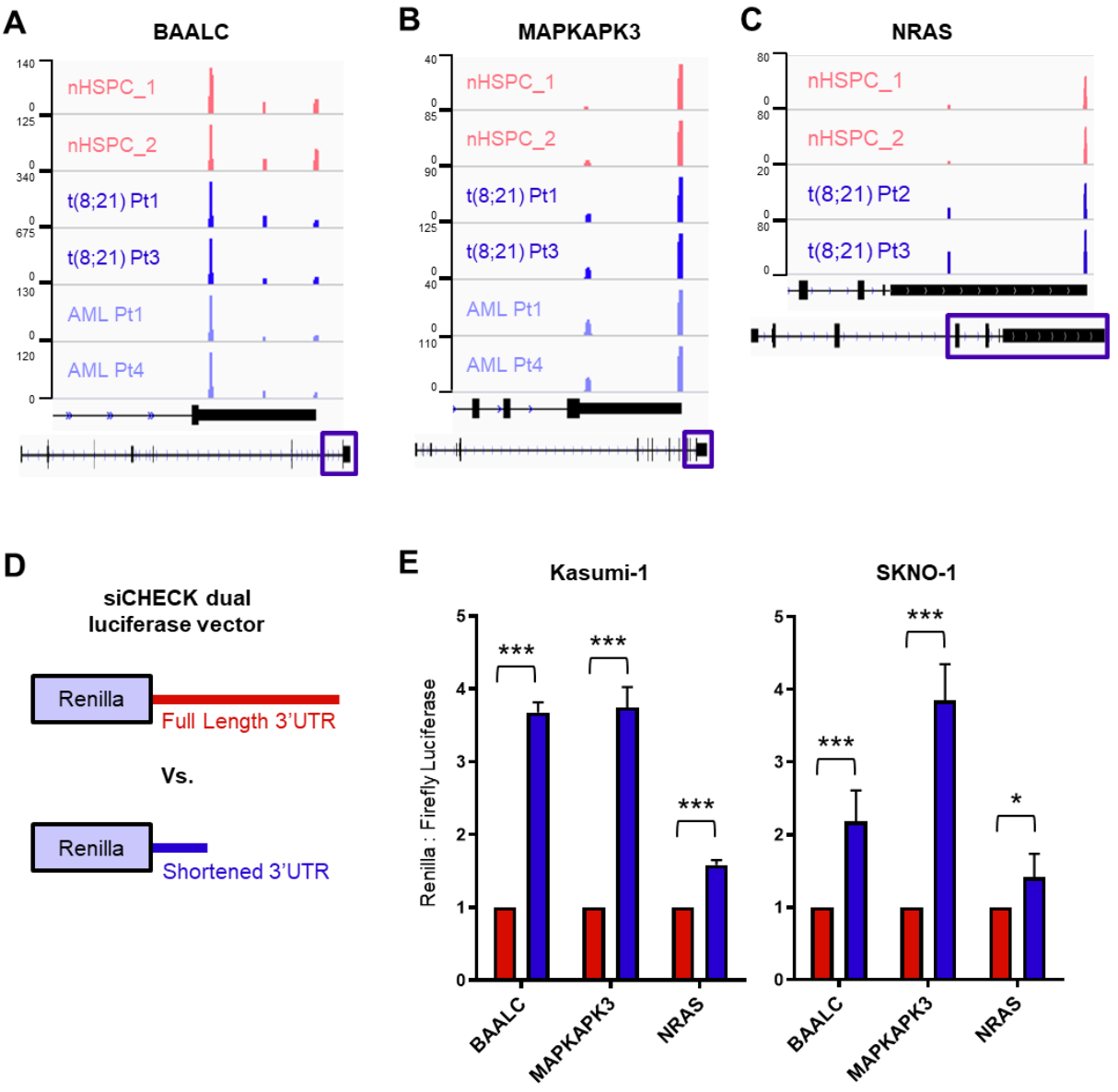
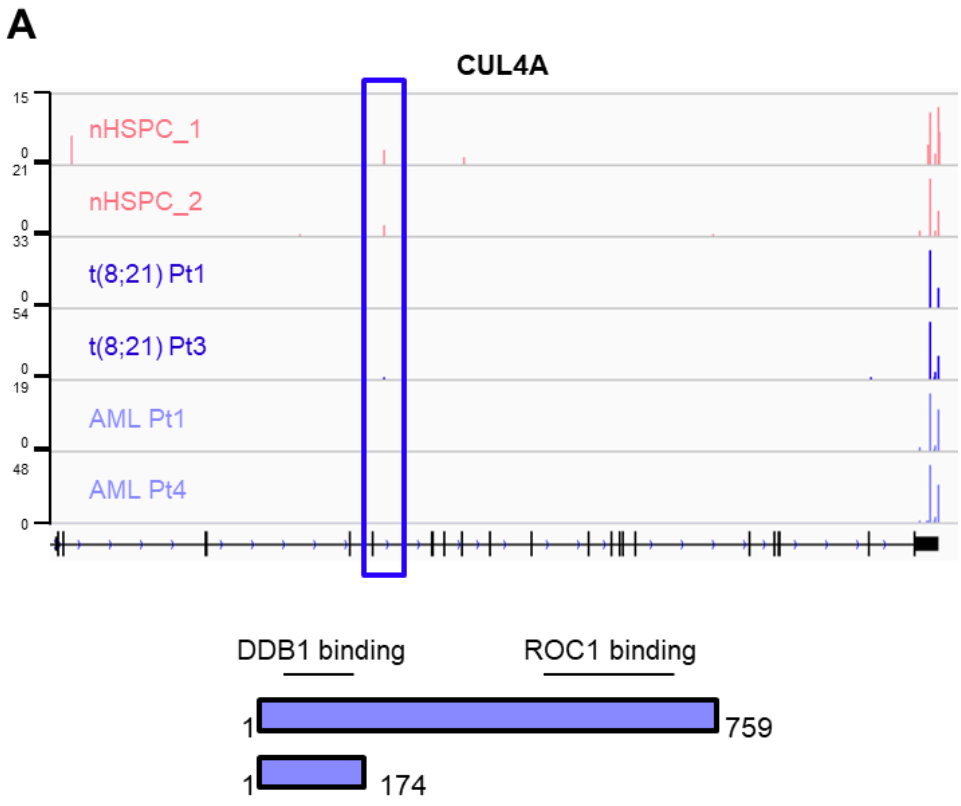


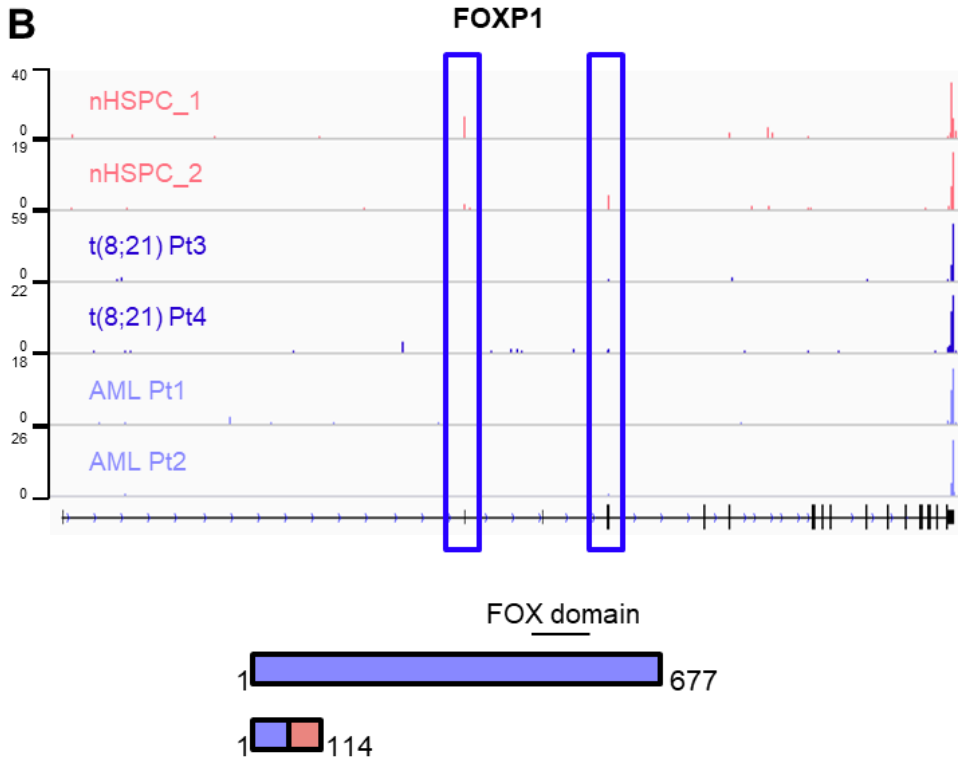
Figure 2.3. 3'UTR shortening contributes to the expression of leukemia-promoting genes.

Genome browser tracks depicting sequencing reads in *BAALC* **(A)**, *MAPKAPK3* **(B)**, and *NRAS* **(C)** obtained from 3'READS of HSPCs and AML patient blasts. The full genomic structure for each gene is shown (bottom) with the purple, boxed region expanded (above). **(D)** Schematic of the psiCHECK dual luciferase reporter vector with either the full length or shortened 3'UTR subcloned downstream of renilla luciferase. **(E)** Relative ratio of renilla to firefly luciferase in Kasumi-1 cells or SKNO-1 cells nucleofected with the indicated dual luciferase reporter. Red bars represent activity when the full 3'UTR of the indicated gene was subcloned downstream of renilla luciferase. Blue bars represent activity when the shortened 3'UTR was present.

Figure 2.4. CDS lengthening activates leukemic oncogenes.

(A) Genome browser tracks depicting sequencing reads in the *CUL4A* gene obtained from 3'READS of HSPCs and AML patient blasts. The proximal, intronic poly(A) site is boxed. **(B)** Genome browser tracks depicting sequencing reads in the *FOXP1* gene obtained from 3'READS of HSPCs and AML patient blasts. Two upstream, intronic poly(A) sites are boxed and numbered. Below each respective gene is a cartoon depicting the protein products from polyadenylation of *CUL4A* and *FOXP1* mRNA at the distal 3'UTR poly(A) site or upstream, intronic poly(A) site. Major protein domains are marked. For *FOXP1*, the most upstream intronic poly(A) site (#1) terminates prior to the start codon and produces no protein product. The product from polyadenylation at the second upstream intronic poly(A) site (#2) is shown, an alternative terminal exon APA event. The blue region is common to the full length and truncated FOXP1 proteins; the pink region is unique to the truncated protein.


A
CUL4A
nHSPC_1
nHSPC_2
t(8;21) Pt1
t(8;21) Pt3
AML Pt1
AML Pt4
DDB1 binding
ROC1 binding
759
174


B
FOXP1
nHSPC_1
nHSPC_2
t(8;21) Pt3
t(8;21) Pt4
AML Pt1
AML Pt2
FOX domain
677
114

2.2.3 Targeting FIP1L1 reverses APA trends in t(8;21) AML.

While altered post-transcriptional regulation of individual, crucial gene targets contributes to oncogenic transformation, the sum of small changes can also collectively promote pathogenesis. Therefore, we wanted to test whether targeting global patterns of polyadenylation might have anti-leukemic effects. To do this, we first selected a common APA regulator to target by utilizing our patient 3'READS and RNA-sequencing datasets. Since 3'UTR shortening is a common feature observed in cancer types (8, 105) and is also seen in our dataset, we focused on the severity of shortening per patient of the overlapping 235 genes (**Figure 2.5A**). We observed noticeable differences in the extent of 3'UTR shortening across this set of genes, with the t(8;21) patients exhibiting shorter 3'UTRs on average compared to non-t(8;21) patients (**Figure 2.5B**). We then correlated the degree of shortening to the expression level of all APA machinery members in each patient and identified *FIP1L1* expression as being most predictive of 3'UTR shortening (**Table 2.4**). Specifically, high *FIP1L1* expression corresponded to more severe 3'UTR shortening (**Figure 2.5C**).

FIP1L1 knockdown has previously been shown to induce 3'UTR lengthening in murine systems (28, 68), therefore we tested whether FIP1L1 knockdown could also induce 3'UTR lengthening in a human context. Indeed, knockdown promoted 3'UTR lengthening of *BAALC* (**Figure 2.6A**), the negative prognostic marker gene that was commonly shortened in our AML patient cohort. Importantly, this lengthening corresponded to significant downregulation of *BAALC* mRNA (**Figure 2.6B**) and protein (**Figure 2.6C**), a change with potential clinical significance. We next tested whether FIP1L1 knockdown could globally alter the APA dysregulation we observed in patients.

We performed shRNA-mediated knockdown in t(8;21) positive Kasumi-1 cells (**Figure 2.7A**) and observed global 3'UTR lengthening but CDS shortening, a reversal of the trends observed in AML patient blasts (**Figure 2.7B**). Importantly, 23% of the genes that had shorter 3'UTRs as detected by 3'READS in our patient samples were lengthened upon FIP1L1 knockdown and 27% of genes with longer CDS-APA isoforms in our patient samples were shortened upon FIP1L1 knockdown (**Figure 2.7C**). *BAALC* was seen among the former group, confirming our previous finding (**Figure 2.6D**). Altogether, these results demonstrate the feasibility of altering global APA profiles by targeting a strategically selected APA regulator.

Figure 2.5. *FIP1L1* expression is correlated with degree of 3'UTR shortening in AML patients.

(A) Heatmap displaying the distal poly(A) site (dPAS) usage of all 235 genes that exhibited significant 3'UTR shortening in both the t(8;21) patients and non-t(8;21) patients compared to healthy HSPCs. Z-scores were calculated per gene (row) from the percent dPAS usage determined by 3'READS. Less dPAS usage is indicative of a greater degree of 3'UTR shortening. The average Z-score was calculated per patient and is shown as a bar graph in **(B)**. Each bar of the graph represents the average Z-score of all 235 genes in the indicated patient or healthy control, a measure of the overall degree of shortening. Data are mean +/- SEM. **(C)** Negative correlation between *FIP1L1* expression, calculated by RNA-sequencing, and the extent of 3'UTR shortening per patient. Among all APA regulators, *FIP1L1* expression is most correlated to 3'UTR shortening. Correlation was determined by linear regression.

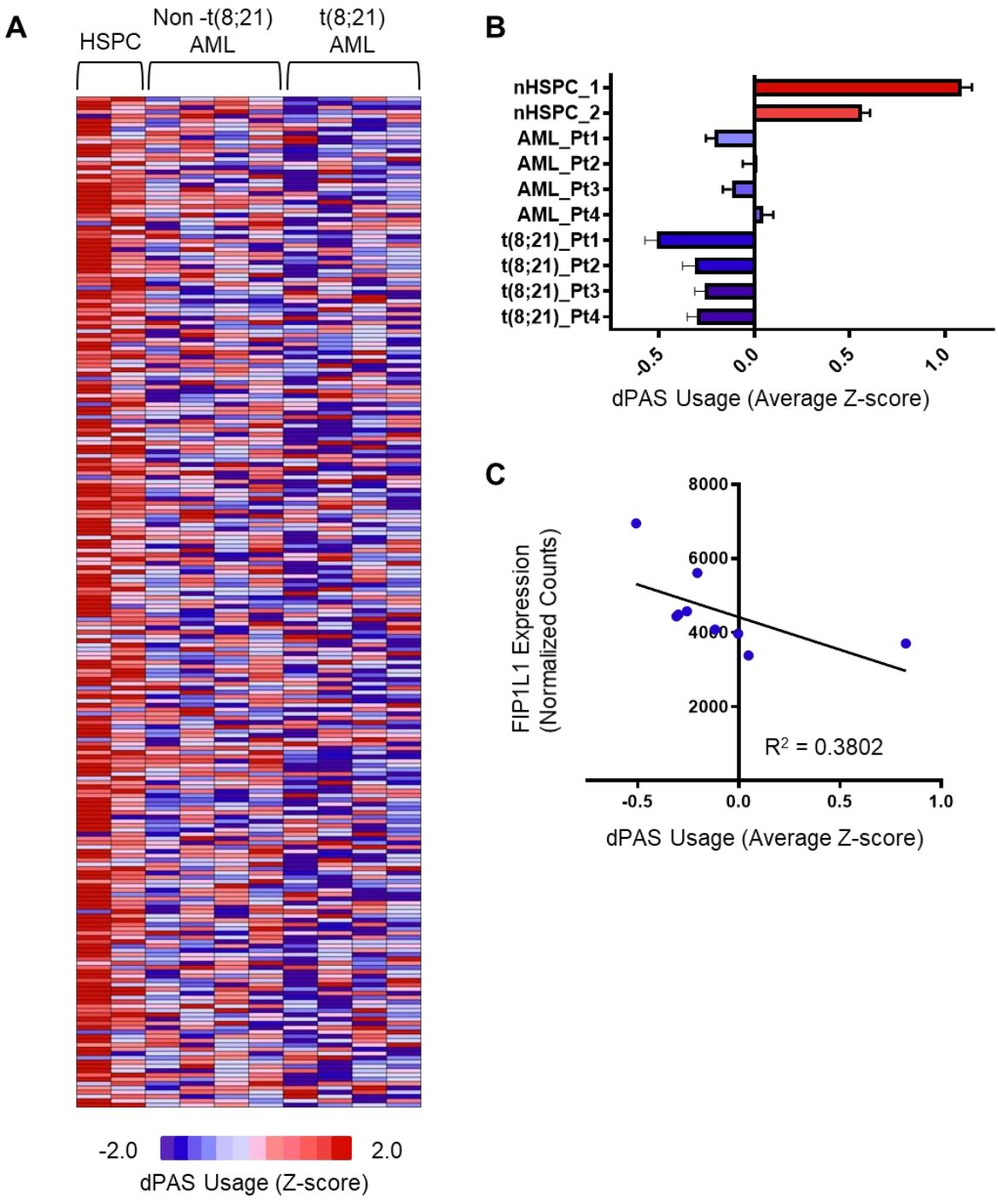

A
HSPC
Non -t(8;21) AML
t(8;21) AML
-2.0
2.0
dPAS Usage (Z-score)
B
nHSPC_1
nHSPC_2
AML_Pt1
AML_Pt2
AML_Pt3
AML_Pt4
t(8;21)_Pt1
t(8;21)_Pt2
t(8;21)_Pt3
t(8;21)_Pt4
-0.5
0.0
0.5
1.0
dPAS Usage (Average Z-score)
C
FIP1L1 Expression (Normalized Counts)
8000
6000
4000
2000
R2 = 0.3802
-0.5
0.0
0.5
1.0
dPAS Usage (Average Z-score)

Table 2.4. APA regulator expression is correlated with 3'UTR shortening in patient samples.

Gene	Average Healthy	AML_Pt1	AML_Pt2	AML_Pt3	AML_Pt4	t(8;21)_ Pt1	t(8;21)_ Pt2	t(8;21)_ Pt3	t(8;21)_ Pt4	RSQ
FIP1L1	3703.473	5607.706	3968.979	4079.743	3377.261	6947.918	4434.57	4572.439	4484.771	**0.380197385**
CSTF1	1311	1602.002	1632.488	1744.376	1368.193	1790.029	1412.087	1946.105	1544.087	**0.352787336**
CSTF3	2693.968	2467.949	2993.103	2543.712	2690.184	1744.858	2388.24	2649.79	1837.764	**0.313817835**
PABPC1	91773.97	235877.7	107971.8	89057.14	128298.7	187550.7	127696.5	132541.7	172186	**0.296967345**
CPSF6	8875.613	15357.99	12882.81	11181.16	9826.251	13745.62	23543.91	12176.78	13199.79	**0.281212461**
PABPC4	10144.57	17652.05	7273.24	8396.427	11967.96	13188.86	20020.53	15886.63	13849.8	**0.209730791**
PCF11	5477.341	7570.054	6516.234	6298.001	5079.306	6005.631	5483.378	7143.537	7126.556	**0.160628011**
PABPN1	2790.514	1634.824	1245.256	968.19	1921.134	2051.6	1853.151	2037.66	2289.374	**0.158327761**
NUDT21	5902.658	7154.539	9803.035	9476.279	6285.045	7218.943	9158.054	8255.813	6393.017	**0.121974952**
CPSF4	2904.622	2286.38	3211.35	935.5085	2481.527	1723.848	3379.778	2601.364	1849.511	**0.098292728**
CPSF3	3131.114	3879.303	3592.971	3498.284	3436.878	3209.236	3678.095	4101.046	3092.09	**0.090069618**
SYMPK	11239.21	10727.27	19573.02	7511.303	10902.32	8647.605	9220.453	10164.09	8499.658	**0.081511684**
CSTF2T	2723.742	3457.503	3470.129	3383.899	2755.762	3705.066	1960.853	3246.787	3255.244	**0.081223936**
RBBP6	7196.625	8165.043	5565.301	8743.668	6748.561	8194.845	8110.101	5749.03	8002.366	**0.065771777**
CPSF7	10570.05	12251.06	11261.55	10485.32	9891.829	10079.42	12740.41	12195.69	10266.94	**0.045152088**
WDR33	5422.061	5051.824	4894.347	5549.05	5363.973	5568.629	4844.007	5524.305	4542.201	**0.035155131**
CPSF1	27826.51	24696.25	13876.15	19050.6	18197.86	30038.66	17392.95	18656.73	17342.59	**0.033503989**
PAPOLG	1487.004	2636.25	1200.983	1850.591	1104.391	1356.178	1056.501	2359.236	1155.129	**0.006749252**
CSTF2	1101.582	1363.168	974.0056	1469.307	810.7812	847.7426	1244.552	1237.88	605.6268	**0.005427696**
CPSF2	2658.28	3900.254	2877.744	3566.371	3232.692	1980.167	3351.571	4854.67	1741.177	**0.001292441**
Average Z-Score of dPAS usage:	**0.82461**	**-0.20506**	**-0.00361**	**-0.11738**	**0.047258**	**-0.50702**	**-0.30813**	**-0.2564**	**-0.29887**	

Figure 2.6. FIP1L1 knockdown induces 3'UTR lengthening and downregulation of BAALC.

(A) RT-qPCR analysis of *BAALC* 3'UTR length in Kasumi-1 cells transduced with shRNAs targeting FIP1L1 or a control shRNA. Usage of either the middle or distal poly(A) site was measured relative to total *BAALC* mRNA using primer pairs upstream of the middle poly(A) site (mPAS) and most proximal poly(A) site (pPAS), respectively. Data are mean +/- s.d. of three independent experiments. * $p < 0.05$, *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(B)** RT-qPCR analysis of total *BAALC* mRNA normalized to *β-Actin* mRNA in the same cell populations. Data are mean +/- s.d. of three independent experiments. *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(C)** Western blot showing FIP1L1, BAALC, and actin (loading control) protein in the same cell populations. BAALC protein was quantified by normalizing BAALC signal intensity to actin signal intensity using LI-COR Image Studio software. Normalized protein quantifications are shown in the bar graph to the right. n = 5 for shControl and shFIP1L1 (2); n = 3 for shFIP1L1(1) and shFIP1L1 (3). *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(D)** Genome browser tracks depicting sequencing reads in the *BAALC* gene obtained from 3'READS of Kasumi-1 cells transduced with a control shRNA or shRNAs targeting *FIP1L1*. The full *BAALC* genomic structure is shown (bottom) with the purple, boxed region expanded (above).

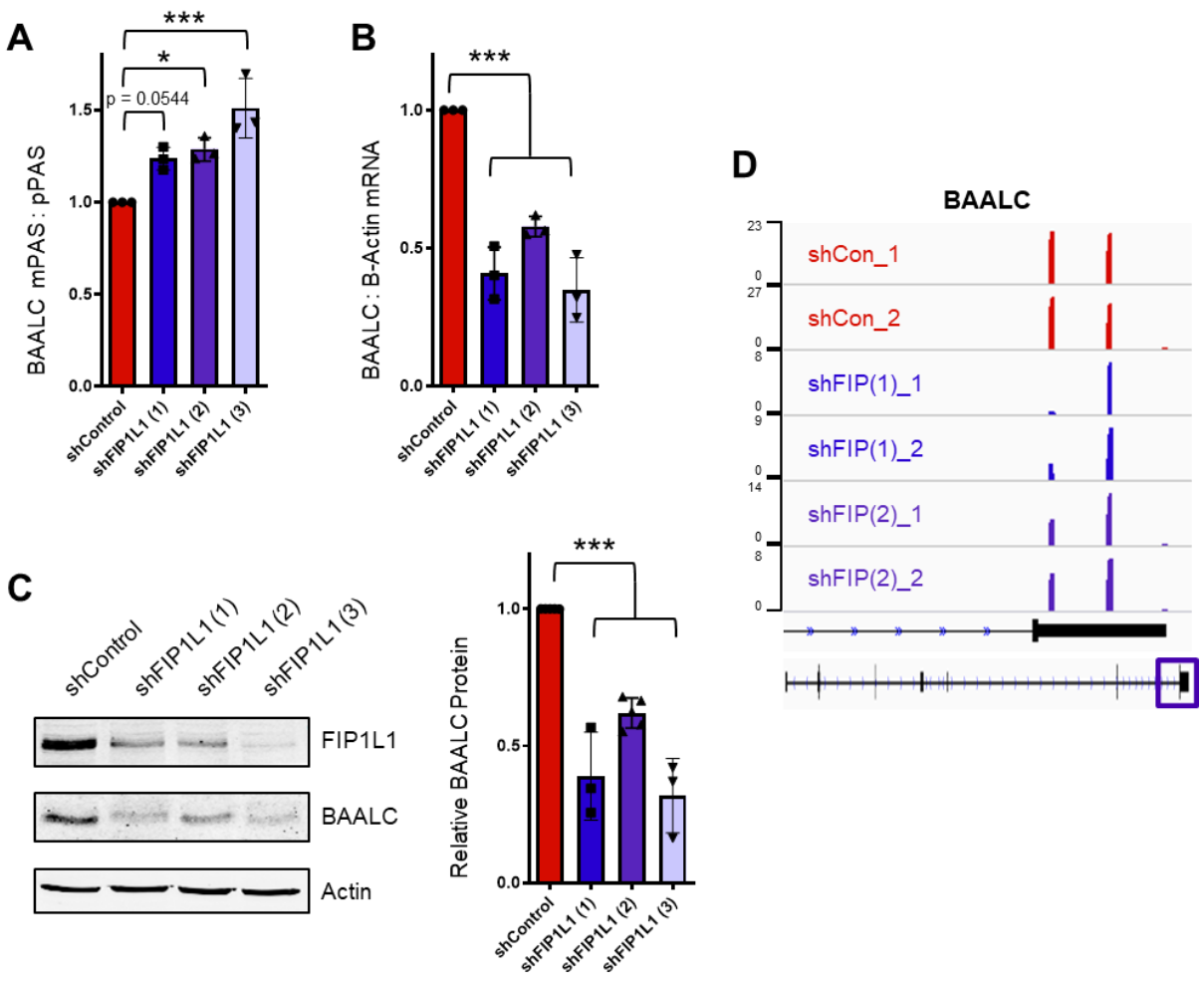
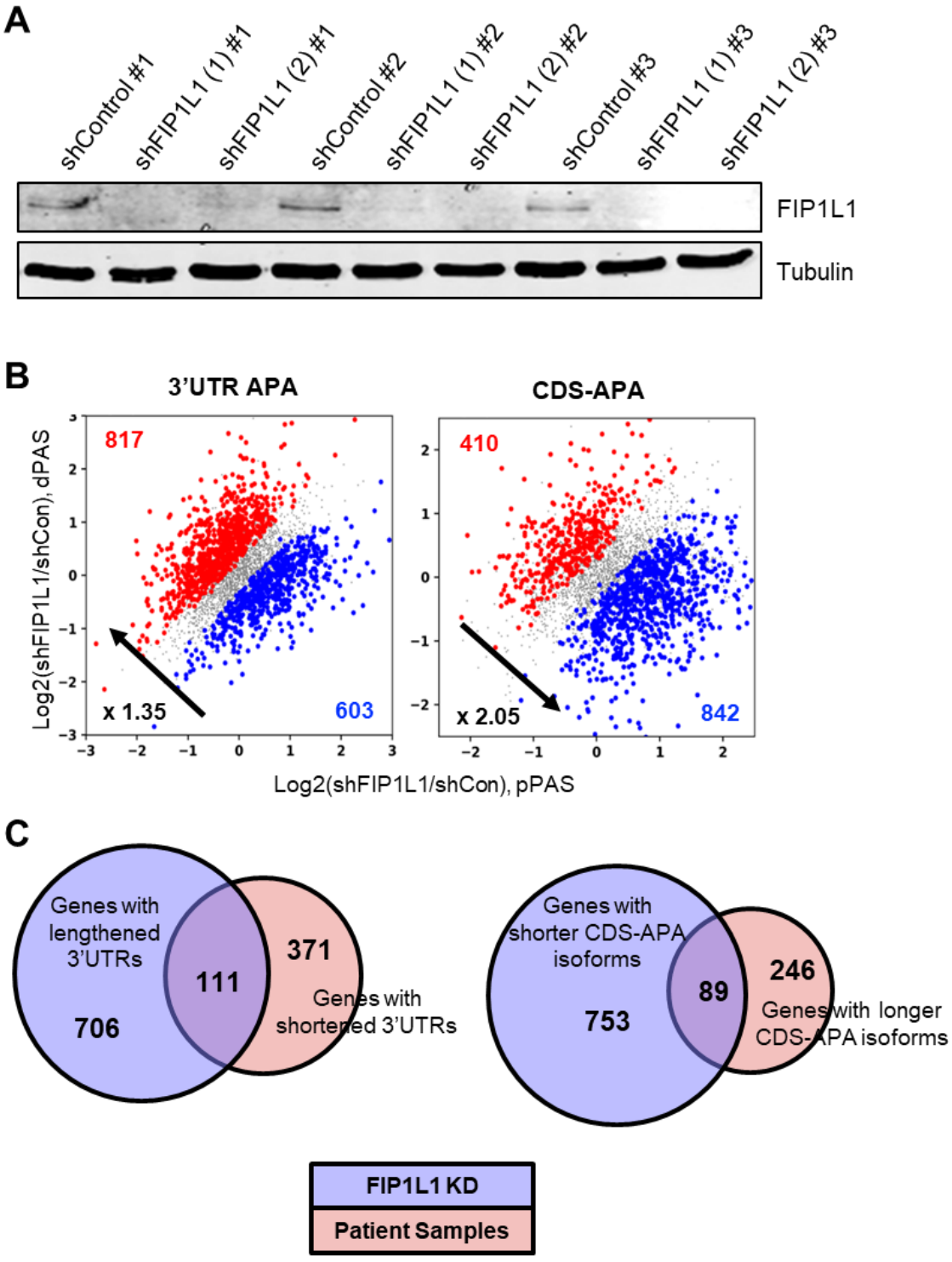

A
***
*
p = 0.0544
BAALC mPAS : pPAS
1.5
1.0
0.5
0.0
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
B
***
BAALC : B-Actin mRNA
1.0
0.5
0.0
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
C
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
FIP1L1
BAALC
Actin
***
Relative BAALC Protein
1.0
0.5
0.0
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
D
BAALC
shCon_1
shCon_2
shFIP(1)_1
shFIP(1)_2
shFIP(2)_1
shFIP(2)_2
23
0
27
0
8
0
9
0
14
0
8
0

Figure 2.7. FIP1L1 knockdown reverses the APA trends seen in AML patients.

(A) Western blot showing FIP1L1 and tubulin (loading control) protein in Kasumi-1 cells transduced with the indicated shRNA. RNA from these cell populations was utilized for 3'READS and RNA-sequencing analyses. **(B)** Scatter plots showing the change in expression of proximal poly(A) isoform and distal poly(A) site isoform, per gene, in Kasumi-1 cells following FIP1L1 knockdown. Significant differences in PAS usage were calculated using a Fisher's exact test, comparing poly(A) site usage in Kasumi-1 cells transduced with both shRNAs targeting FIP1L1 versus the control shRNA. APA events are classified and divided by type: tandem 3'UTR APA (left) and CDS-APA (right). Each dot corresponds to a single gene. Blue dots indicate significantly more proximal poly(A) site (pPAS) usage. Red dots indicate significantly more distal poly(A) site (dPAS) usage. **(C)** Venn diagrams showing: (left) the overlap of genes with 3'UTR shortening in t(8;21) AML blasts that were significantly lengthened upon FIP1L1 knockdown in Kasumi-1 cells and (right) the overlap of genes with CDS lengthening in t(8;21) AML blasts that were significantly shortened upon FIP1L1 knockdown in Kasumi-1 cells.


A
shControl #1
shFIP1L1 (1) #1
shFIP1L1 (2) #1
shControl #2
shFIP1L1 (1) #2
shFIP1L1 (2) #2
shControl #3
shFIP1L1 (1) #3
shFIP1L1 (2) #3
FIP1L1
Tubulin
B
3'UTR APA
CDS-APA
817
410
x 1.35
603
x 2.05
842
Log2(shFIP1L1/shCon), dPAS
Log2(shFIP1L1/shCon), pPAS
C
Genes with lengthened 3'UTRs
706
111
371
Genes with shortened 3'UTRs
Genes with shorter CDS-APA isoforms
753
89
246
Genes with longer CDS-APA isoforms
FIP1L1 KD
Patient Samples

2.2.4 FIP1L1 knockdown promotes t(8;21) leukemic cell differentiation.

We next set out to determine whether this perturbation of global APA patterns by FIP1L1 knockdown had a meaningful anti-leukemic effect. Therefore, we also performed RNA-sequencing of Kasumi-1 cells following shRNA-mediated FIP1L1 knockdown (**Figure 2.8A**). Targeting this APA regulator profoundly changed global gene expression, highlighting the interplay between post-transcriptional gene regulation and transcriptome composition (**Figure 2.8B**). Gene set enrichment analysis (GSEA) revealed that genes which were upregulated upon FIP1L1 knockdown significantly matched those that are upregulated upon myeloid cell development. This included genes such as *ITGAM* (CD11b) (106) and *LST1* (107), classical markers of more mature myeloid cells. Similarly, genes that were downregulated upon FIP1L1 knockdown negatively correlated with those that are typically upregulated in HSCs. Among these genes are reported leukemic oncogenes *BCAT1* (108, 109) and *CBX5* (110) (**Figure 2.9A**). Additionally, flow cytometric analysis of Kasumi-1 cells revealed that FIP1L1 knockdown reduced CD34 cell surface expression, indicative of a more differentiated cellular state (**Figure 2.9B**). This change correlated with a decrease in cell proliferation (**Figure 2.10A**). Overall, our results indicate that targeting APA promotes t(8;21) leukemia cell differentiation.

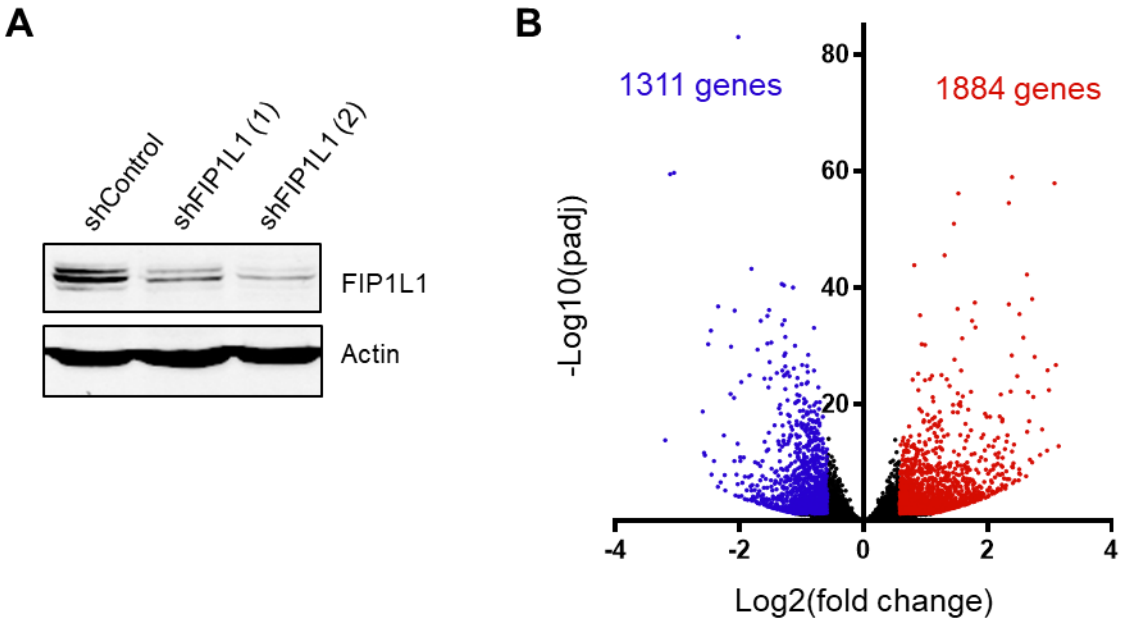
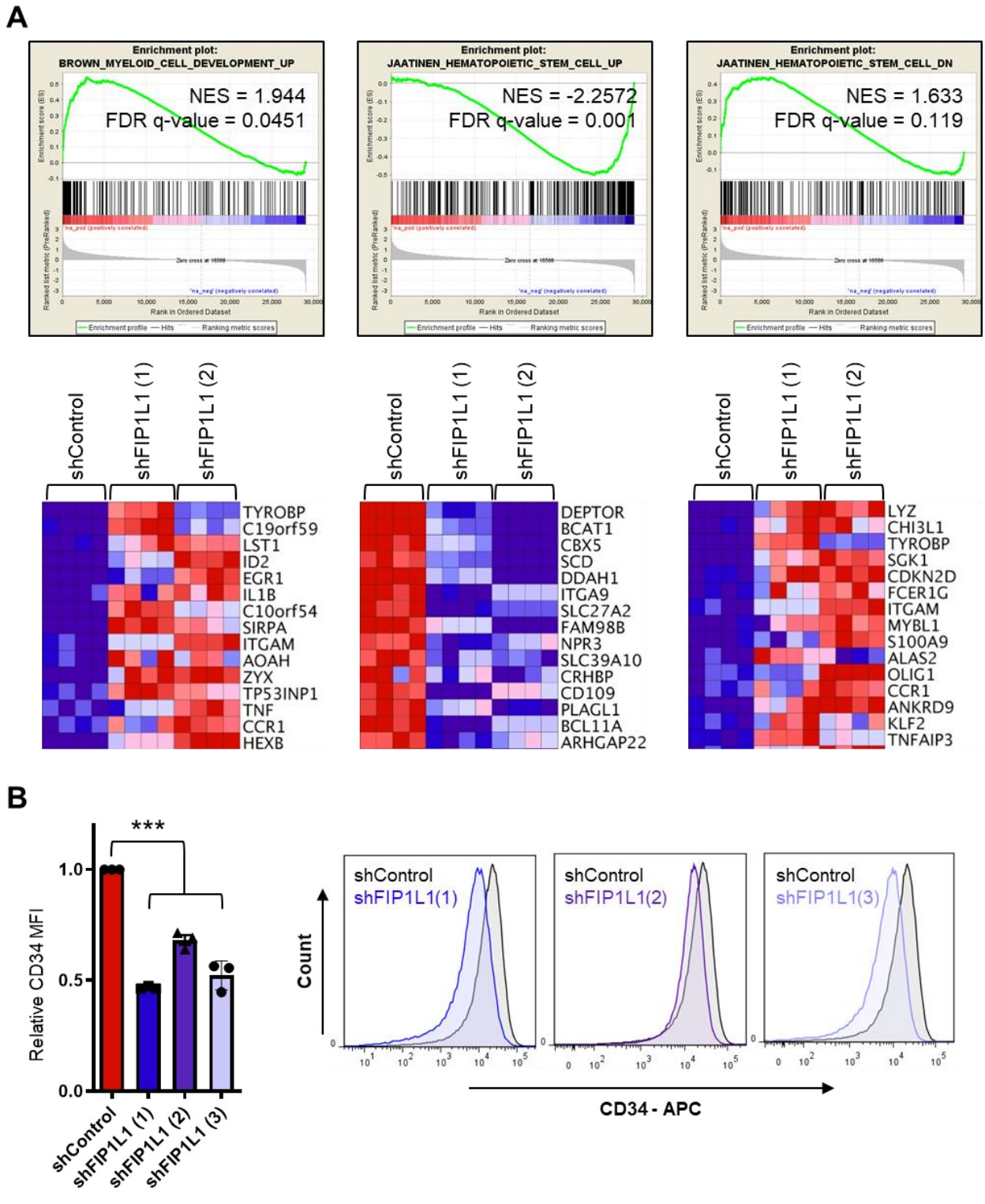


Figure 2.8. Global gene expression signatures are altered by targeting APA.

(A) Western blot showing FIP1L1 and actin (loading control) protein in Kasumi-1 cells following transduction with control shRNAs or shRNAs targeting *FIP1L1*. **(B)** Volcano plot displaying fold-change and significance of all genes measured by RNA-sequencing in Kasumi-1 cells upon FIP1L1 knockdown using two independent shRNAs compared to a control shRNA (n = 4 for each condition). Significantly upregulated (red) and downregulated (blue) genes are highlighted ($p < 0.01$, $FC > 1.5$).

Figure 2.9. FIP1L1 knockdown promotes differentiation of t(8;21) AML cells.

(A) Plots from gene-set enrichment analysis (GSEA) of the RNA-sequencing experiment shown in Figure 2.8. NES, normalized enrichment score; FDR, false discovery rate. Heatmaps display fifteen genes in the leading edge of each gene-set that are most differentially expressed upon FIP1L1 knockdown. **(B)** CD34 cell surface expression measured by flow cytometry of Kasumi-1 cells six days following transduction with shRNAs targeting *FIP1L1* or a control shRNA. MFI, mean fluorescent intensity. Data are mean +/- s.d. of three independent experiments. *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test. Representative histograms (right) depict the CD34-APC fluorescence distribution in the indicated cell populations.


A
Enrichment plot:
BROWN_MYELOID_CELL_DEVELOPMENT_UP
NES = 1.944
FDR q-value = 0.0451
Enrichment plot:
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP
NES = -2.2572
FDR q-value = 0.001
Enrichment plot:
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN
NES = 1.633
FDR q-value = 0.119
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
shControl
shFIP1L1 (1)
shFIP1L1 (2)
TYROBP
C19orf59
LST1
ID2
EGR1
IL1B
C10orf54
SIRPA
ITGAM
AOAH
ZYX
TP53INP1
TNF
CCR1
HEXB
DEPTOR
BCAT1
CBX5
SCD
DDAH1
ITGA9
SLC27A2
FAM98B
NPR3
SLC39A10
CRHBP
CD109
PLAGL1
BCL11A
ARHGAP22
LYZ
CHI3L1
TYROBP
SGK1
CDKN2D
FCER1G
ITGAM
MYBL1
S100A9
ALAS2
OLIG1
CCR1
ANKRD9
KLF2
TNFAIP3
B
***
Relative CD34 MFI
1.0
0.5
0.0
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
Count
shControl
shFIP1L1(1)
shFIP1L1(2)
shFIP1L1(3)
CD34 - APC

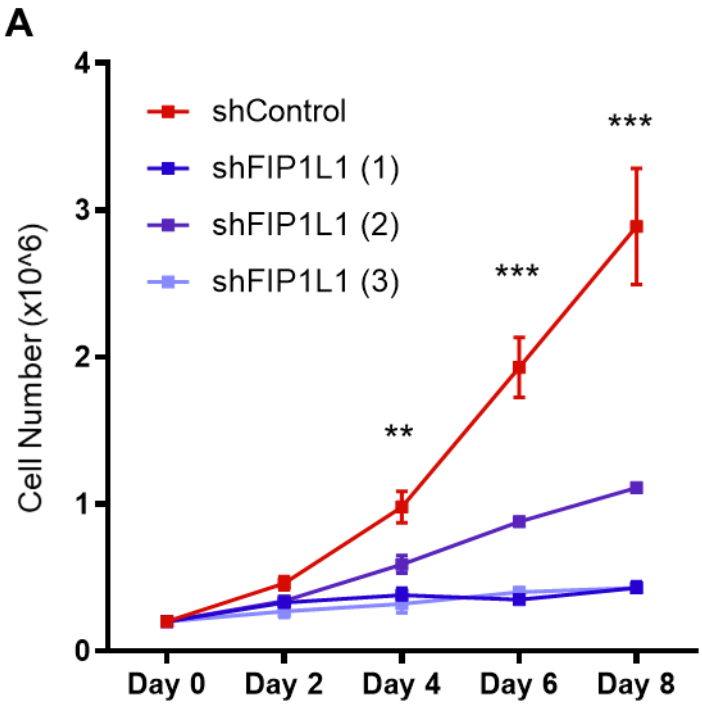


Figure 2.10. FIP1L1 knockdown reduces proliferation of t(8;21) AML cells.

(A) Proliferation of Kasumi-1 cells following transduction with shRNAs targeting *FIP1L1* or a control shRNA. Cells were seeded (day '0') after two days of puromycin selection. The graph displays the mean and s.d. of three technical replicates in one representative experiment of two independent experiments. ** $p < 0.01$, *** $p < 0.001$, multiple t-tests using the Holm-Sidak method to determine statistical significance.

2.2.5 3'UTR APA regulates AML1-ETO expression.

We next wondered whether there was a crucial APA event altered by targeting FIP1L1 that might explain the differentiation phenotype observed. Indeed, among the genes with significantly altered APA was the t(8;21) generated fusion oncoprotein AML1-ETO. FIP1L1 knockdown induced *AML1-ETO* 3'UTR lengthening (**Figures 2.11A and 2.11B**). Specifically, we observed a decrease in usage of the proximal poly(A) site at 1kb, a concurrent increase in usage of the 3.7kb site, and the emergence of usage at the most distal 5.2kb site. 3'UTR lengthening corresponded to downregulation of AML1-ETO protein (**Figure 2.12A**), and the reversal of downstream target gene expression signatures (**Figure 2.12B**). Since AML1-ETO contributes to the differentiation block seen in AML patients (111, 112), this single gene could be a major driver of the overall differentiation phenotype seen upon FIP1L1 knockdown in t(8;21) cells. To confirm that 3'UTR length of the *AML1-ETO* transcript can contribute to protein production, we again performed luciferase assays comparing renilla luciferase production when followed by the 1kb, 3.7kb, or 5.2kb ETO 3'UTR. Indeed, the shortest 1kb 3'UTR produced 7-8 times more renilla protein than either the 3.7 or 5.2 kb 3'UTR (**Figure 2.12C**). This observation is supported by various reports of miRNAs that bind to the 3'UTR region downstream of the 1kb poly(A) site and downregulate *AML1-ETO* (113, 114). Finally, t(8;21) AML patient samples have significantly higher *FIP1L1* expression than non-t(8;21) patients in two large datasets (115, 116) (**Figure 2.13A**). In one of these datasets, *FIP1L1* expression was significantly correlated with *AML1-ETO* mRNA expression, further supporting a role for this specific APA regulator in maintaining expression of AML1-ETO in patients (**Figure 2.13B**). Excitingly, our data reveals that a

fusion oncoprotein is susceptible to APA-mediated gene expression regulation and underlines the importance of miRNA regulation in t(8;21) AML.

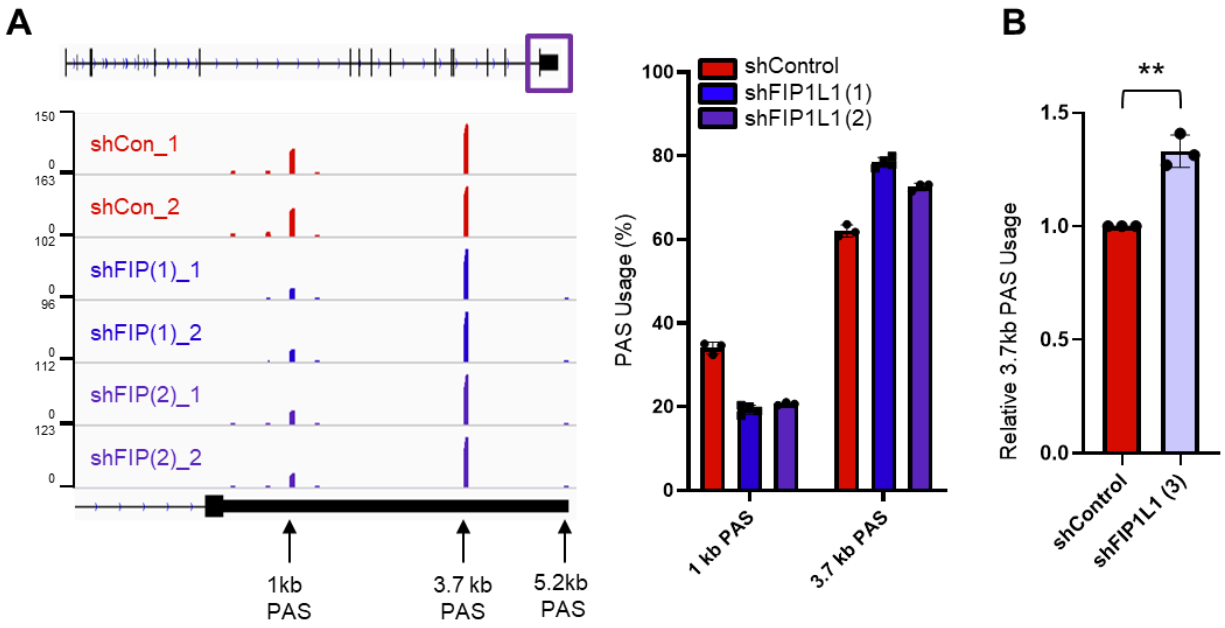


Figure 2.11. The 3'UTR of *AML1-ETO* is lengthened upon FIP1L1 knockdown.

(A) Genome browser tracks depicting sequencing reads in the *RUNX1T1* gene (aka *ETO*) obtained from 3'READS of Kasumi-1 cells transduced with a control shRNA or shRNAs targeting *FIP1L1*. The full *RUNX1T1* genomic structure is shown (top) with the purple, boxed region expanded (below). The percent usage of the polyadenylation signal (PAS) at 1 kb and 3.7 kb, as calculated by 3'READS, are shown in the bar graph to the right. n = 3, shControl and shFIP1L1 (2); n = 4 shFIP1L1 (1). **(B)** RT-qPCR analysis of *AML1-ETO* 3'UTR length in Kasumi-1 cells transduced with a third, unique shRNA targeting *FIP1L1* or a control shRNA. Usage of the PAS at 3.7kb was measured relative to total *AML1-ETO* mRNA using primer pairs upstream of the 3.7kb PAS and 1kb PAS, respectively. Data are mean +/- s.d. of three independent experiments. ** $p < 0.01$, student's t-test.

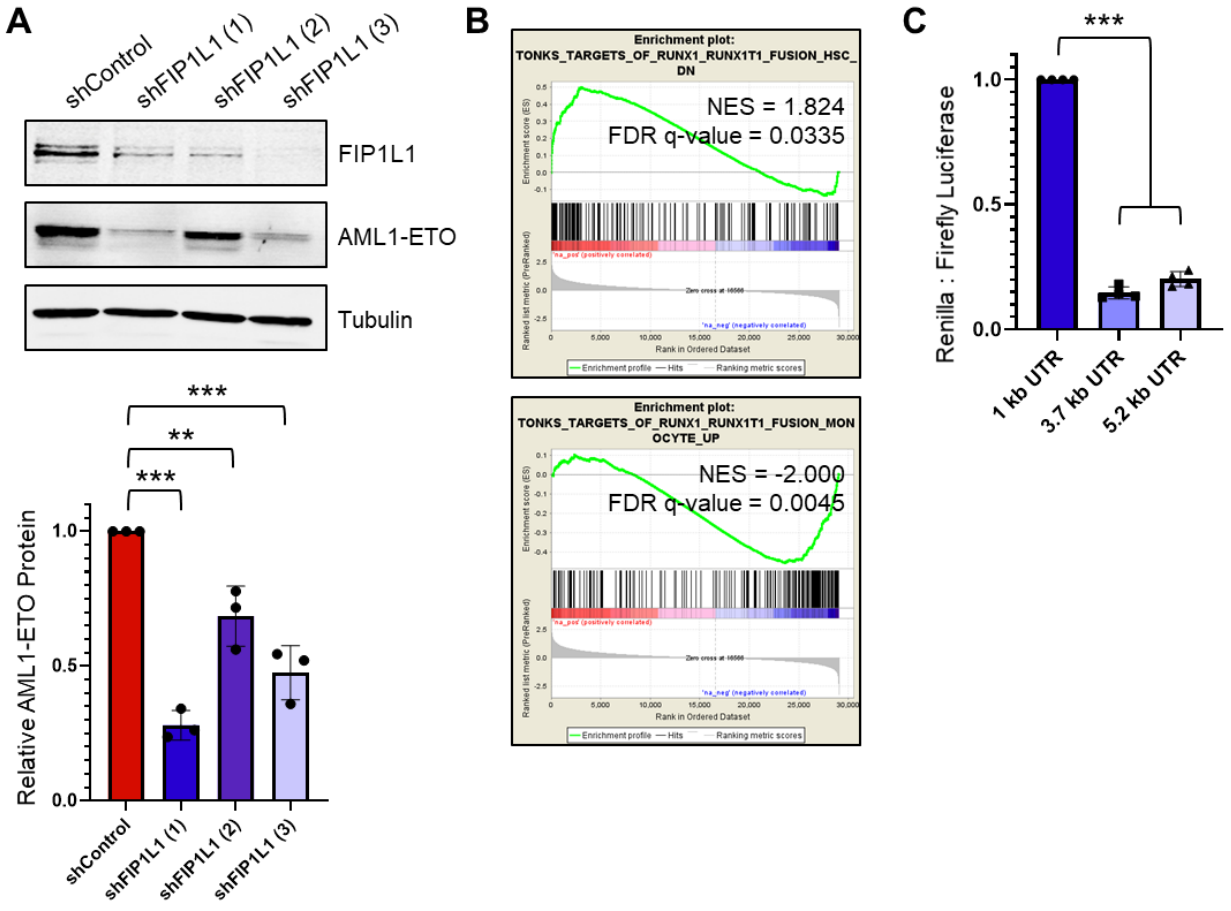


Figure 2.12. *AML1-ETO* 3'UTR length contributes to its expression.

(A) Western blot showing FIP1L1, AML1-ETO, and tubulin (loading control) protein in Kasumi-1 cells following transduction with control shRNAs or shRNAs targeting *FIP1L1*. AML1-ETO protein was quantified by normalizing AML1-ETO signal intensity to tubulin signal intensity using LI-COR Image Studio software. Normalized protein quantifications from three independent experiments are shown in the bar graph below. ** $p < 0.01$, *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(B)** Plots from gene-set enrichment analysis (GSEA) of the RNA-sequencing experiment shown in Figure 2.8. The top plot displays genes that are downregulated by AML1-ETO; the bottom plot those that are upregulated by AML1-ETO. NES, normalized enrichment score; FDR, false discovery rate. **(C)** Relative ratio of renilla to firefly luciferase in Kasumi-1 cells nucleofected with the indicated dual luciferase reporter construct containing variable *RUNX1T1* 3'UTRs. Data are mean +/- s.d. of four independent experiments. *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test.

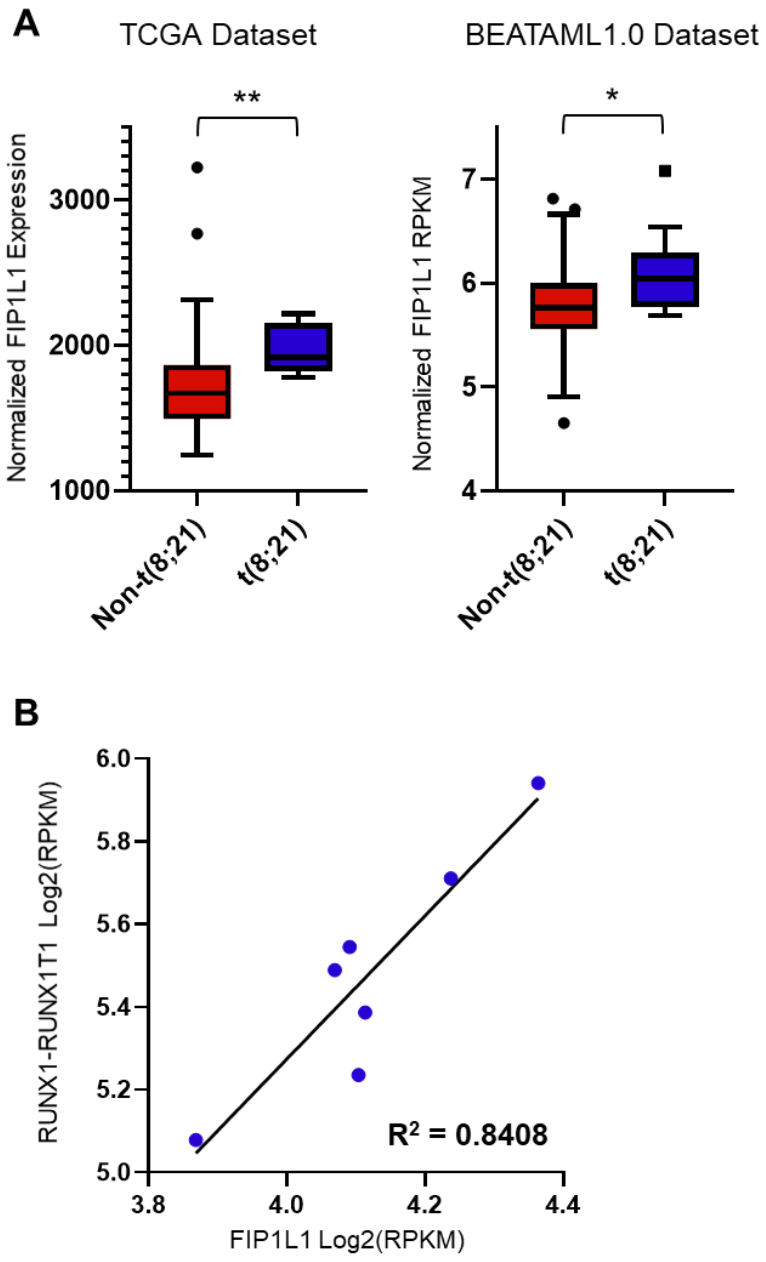


Figure 2.13. *FIP1L1* expression contributes to *AML1-ETO* expression in AML patients.

(A) Box and whisker plots depict normalized *FIP1L1* expression in non-t(8;21) and t(8;21) patients from the TCGA dataset[55] (left) and BEATAML dataset[56] (right). In the TCGA patient cohort, there are 144 non-t(8;21) patients and 7 t(8;21) patients. In the BEATAML patient cohort, there are 472 non-t(8;21) patients and 11 t(8;21) patients. Statistical significance was determined by performing a welch two sample t-test. **(B)** Dot plot showing the correlation between *FIP1L1* and *RUNX1-RUNX1T1* mRNA expression in the TCGA patient cohort. Correlation was calculated by linear regression.

2.2.6 *FIP1L1* expression correlates with stemness signatures across AML.

Though t(8;21) AML patients tended to have higher *FIP1L1* expression, there was also a cohort of non-t(8;21) patients with elevated *FIP1L1* expression. As such, we hypothesized that targeting APA by FIP1L1 knockdown might apply more broadly to AML patients. To address this, we utilized the publicly available RNA-sequencing data of the TCGA patient cohort (115) and performed differential gene expression analysis of the highest and lowest *FIP1L1* expressing patients (**Figure 2.14A**). Gene ontology (GO) analysis of pathways enriched in *FIP1L1* low-expressing patients revealed a striking overlap with those that were enriched upon FIP1L1 knockdown in t(8;21)+ Kasumi-1 cells. Importantly, overlapping enriched pathways support a gene expression signature of leukocyte differentiation (**Figure 2.14B**). GSEA analysis further revealed that *FIP1L1* high patients had gene expression profiles positively correlated with both immature HSCs and leukemic stem cells (**Figure 2.14C**). Furthermore, when stratified by FAB subtype, *FIP1L1* expression was significantly higher in M0-M2 versus M3-M6 leukemia patients (**Figures 2.15A and 2.15B**). Collectively, these data support a role of FIP1L1 in regulating stemness gene signatures in patients and prompted us to evaluate the impact of FIP1L1 knockdown in a non-t(8;21) AML context.

We turned to the HL-60 and NB4 cell lines, classically used to model myeloid cell differentiation. Indeed, FIP1L1 knockdown (**Figures 2.16A and 2.17A**) resulted in the robust emergence of CD11b+ cells (**Figures 2.16B, 2.16C, 2.17B, and 2.17C**) and the corresponding nuclear morphology of differentiated granulocytes (**Figures 2.16D and 2.17D**). These data were further confirmed in HL-60 cells transduced with Cas9 and an sgRNA targeting *FIP1L1* or a control non-targeting sgRNA. Again, knockdown

(**Figure 2.18A**) promoted a robust increase in CD11b+ cells (**Figures 2.18B and 2.18C**). Importantly, this phenotype was induced without the addition of traditional differentiating agents, supporting a profound role of targeting APA and overcoming the hallmark differentiation block of AML. Finally, these Cas9/sgRNA-transduced HL-60 cells were injected into mice via tail vein injection and survival was monitored. This murine xenograft experiment revealed that FIP1L1 knockdown by Cas9/sgRNA significantly extended survival, preliminarily suggesting clinical relevance of targeting this APA regulator (**Figure 2.18D**).

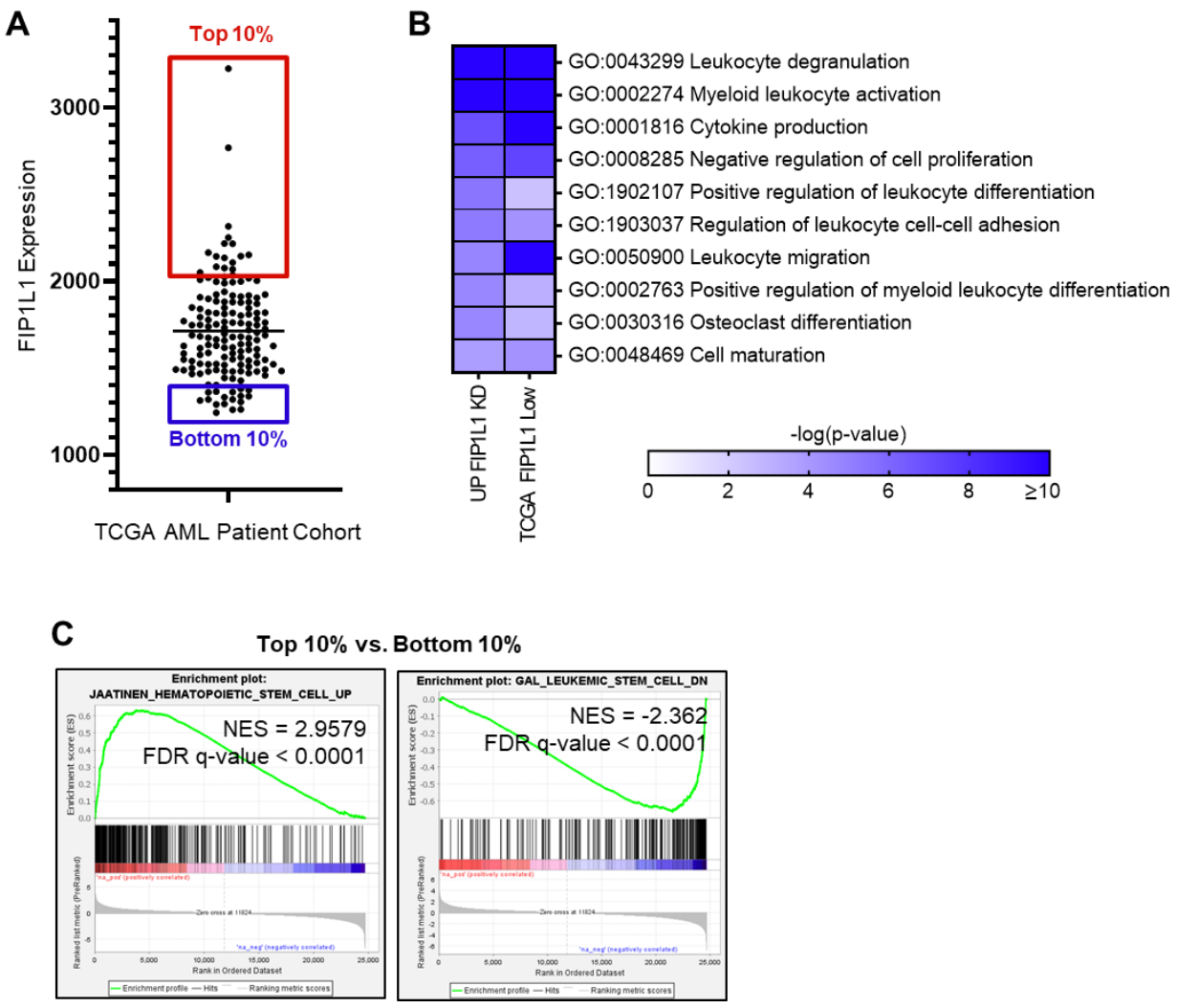


Figure 2.14. *FIP1L1* expression correlates with cell maturity across AML patients.

(A) Box and whisker plot depicting the *FIP1L1* expression of all patients in the AML TCGA patient cohort. Differential gene expression analysis was performed comparing the highest (red) and lowest (blue) *FIP1L1* expressing patients. **(B)** Heatmap depicting significantly enriched gene ontology (GO) terms of upregulated genes in *FIP1L1* low-expressing patients and in Kasumi-1 cells upon FIP1L1 knockdown. **(C)** Plots from gene-set enrichment analysis (GSEA) of the RNA-sequencing analysis described in (a). NES, normalized enrichment score; FDR, false discovery rate.

.

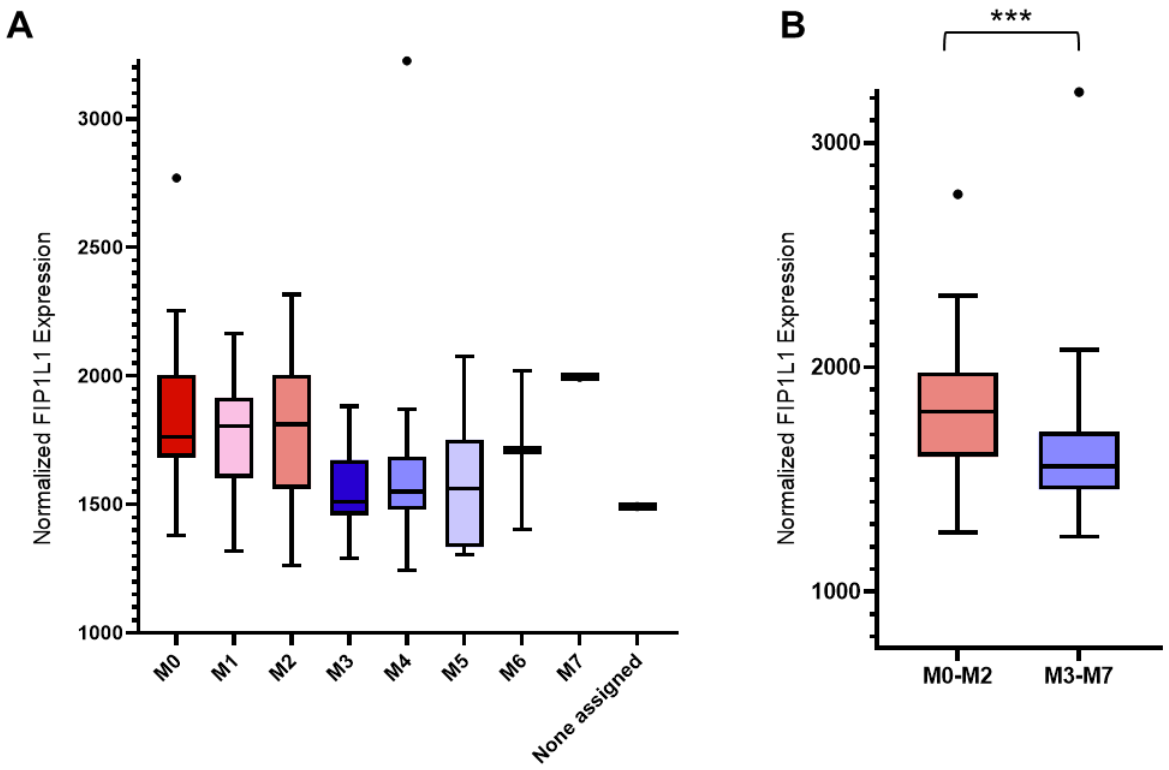


Figure 2.15. *FIP1L1* expression is higher in more immature FAB AML subtypes.

(A) Box and whisker plots depict normalized *FIP1L1* expression in patients of the indicated FAB subtype from the TCGA dataset. M0: n = 15; M1: n = 36; M2: n = 37; M3: n = 15; M4: n = 29; M5 n = 15; M6 n = 2; M7: n = 1; none assigned: n = 1. **(B)** The difference in *FIP1L1* expression between the combined M0-M2 and M3-M7 patients was determined by performing a welch two sample t-test, *** $p < 0.001$.

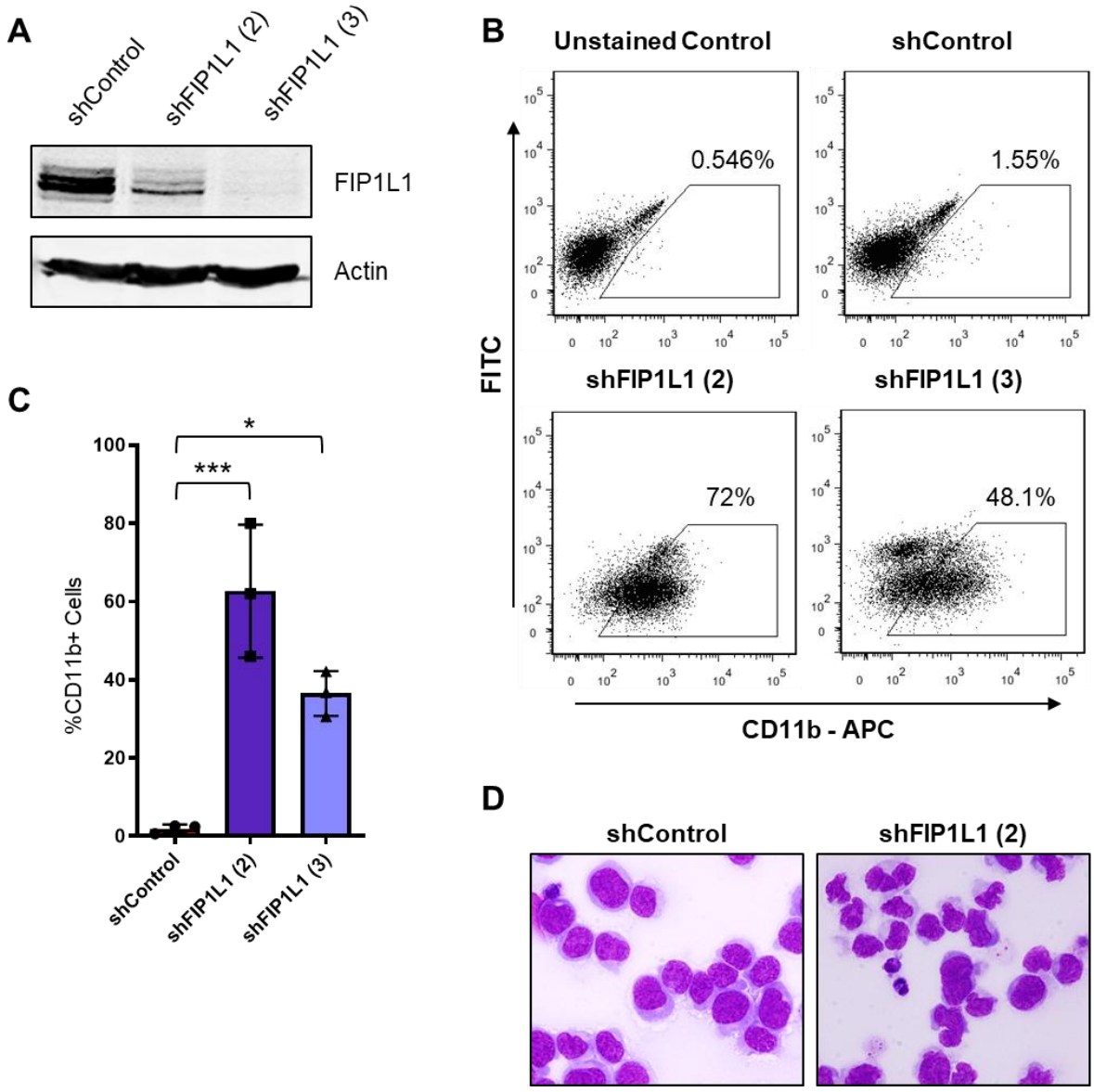


Figure 2.16. FIP1L1 knockdown promotes HL-60 differentiation.

(A) Western blot showing FIP1L1 and actin (loading control) protein in HL-60 cells following transduction with control shRNAs or shRNAs targeting *FIP1L1*. **(B)** Representative FACS plots of CD11b cell-surface expression in HL-60 cells, five days following transduction. **(C)** Percentage of CD11b+ HL-60 cells, measured by flow cytometry, from cells shown in (B). Data are mean +/- s.d. of three independent experiments. * $p < 0.05$, *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(D)** Wright-giemsa staining of HL-60 cell cytospins, five days following transduction with shRNAs targeting *FIP1L1* or a control shRNA (400x). Irregular-shaped nuclei observed in the shFIP1L1 image are indicative of granulocytic differentiation as compared to the spherical, smooth-edged nuclei observed in the shControl image.

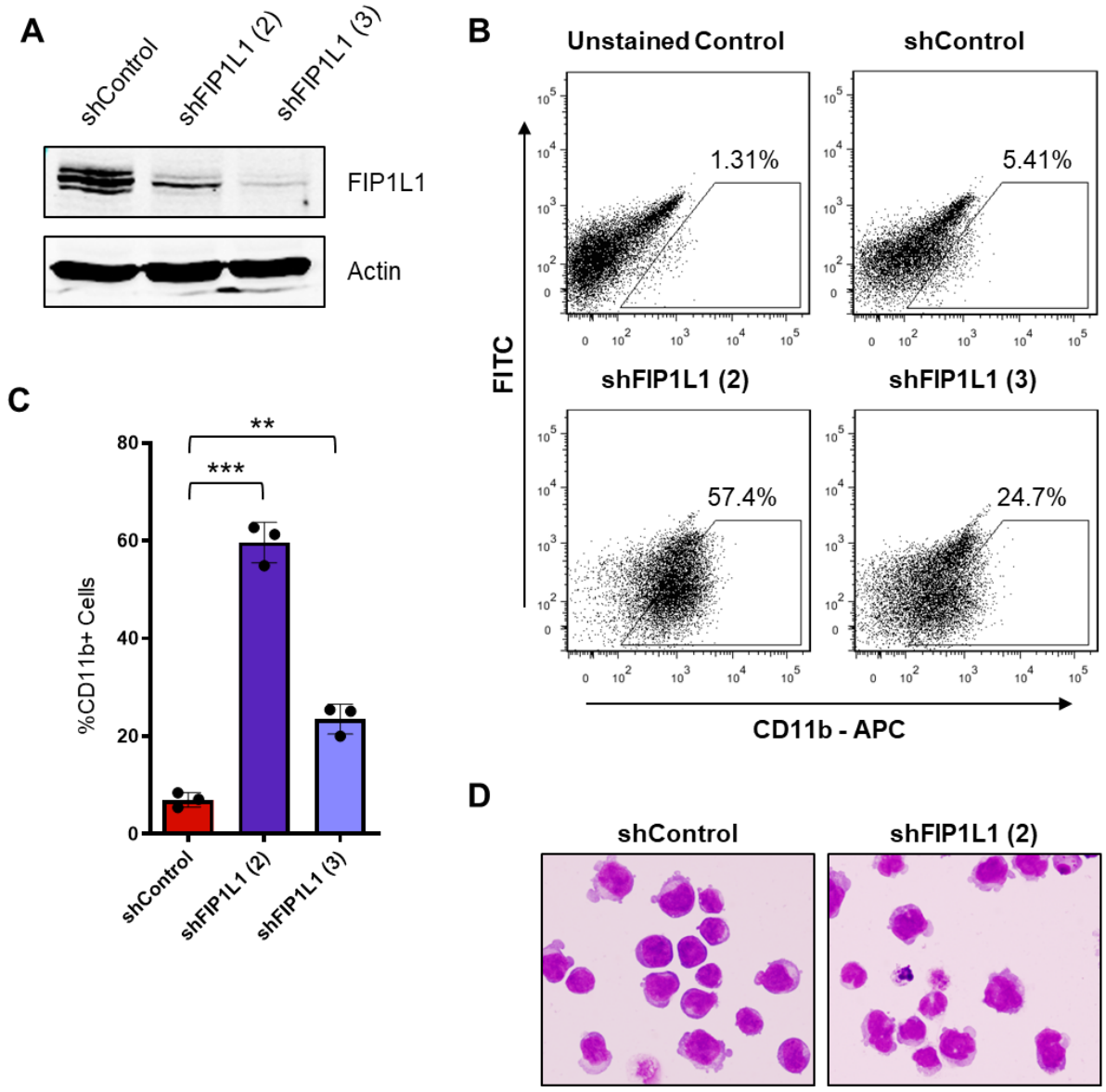


Figure 2.17. FIP1L1 knockdown promotes NB4 differentiation.

(A) Western blot showing FIP1L1 and actin (loading control) protein in NB4 cells following transduction with control shRNAs or shRNAs targeting *FIP1L1*. **(B)** Representative FACS plots of CD11b cell-surface expression in NB4 cells, four days following transduction. **(C)** Percentage of CD11b+ NB4 cells, measured by flow cytometry, from cells shown in (B). Data are mean +/- s.d. of three independent experiments. ** $p < 0.01$, *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test. **(D)** Wright-giemsa staining of NB4 cell cytospins, four days following transduction with shRNAs targeting *FIP1L1* or a control shRNA (400x). Irregular-shaped nuclei observed in the shFIP1L1 image are indicative of granulocytic differentiation as compared to the spherical, smooth-edged nuclei observed in the shControl image.

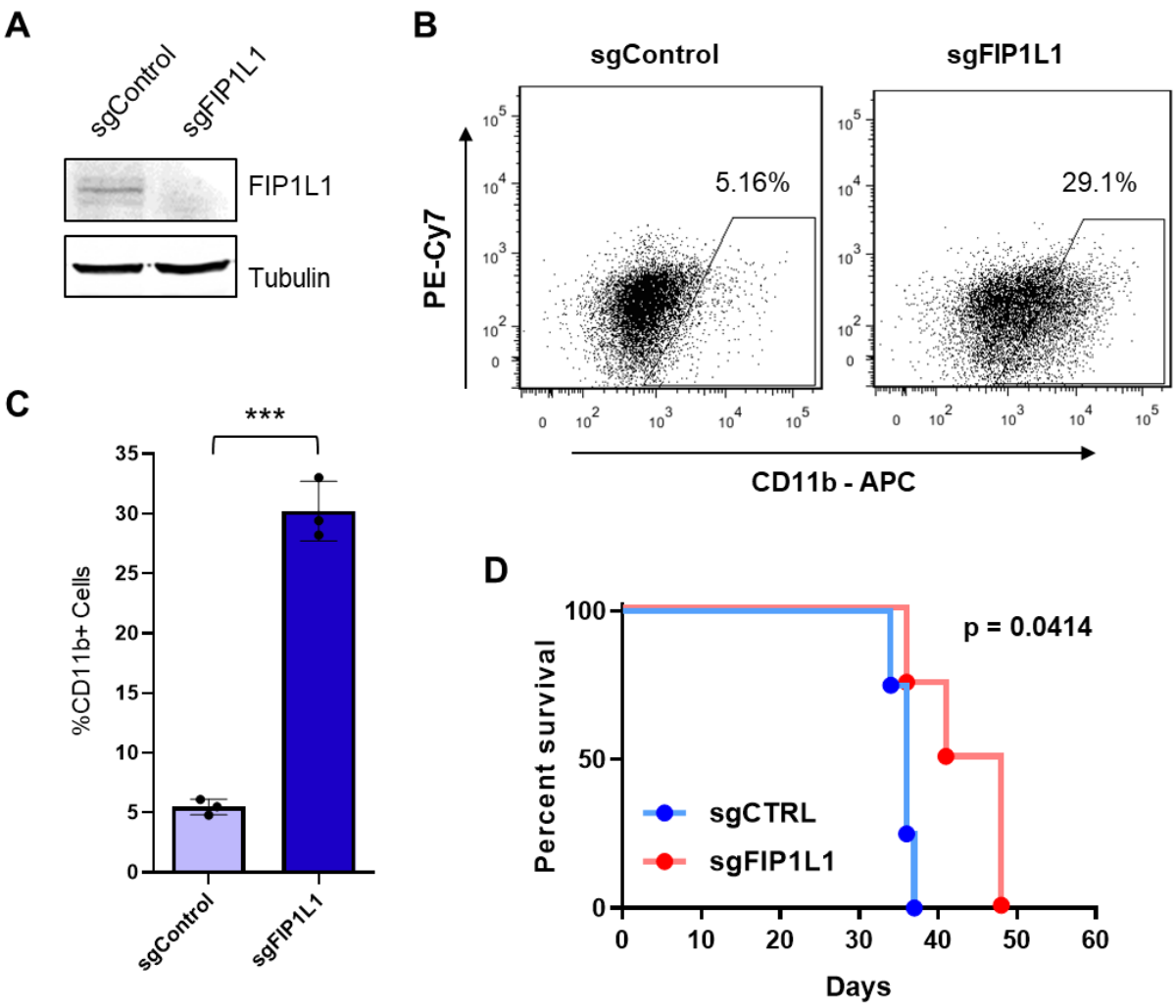


Figure 2.18. CRISPR targeting of FIP1L1 reduces leukemogenic potential of non-t(8;21) cells in vivo.

(A) Western blot showing FIP1L1 and actin (loading control) protein in HL-60 cells following transduction with Cas9 plus control sgRNA or sgRNA targeting *FIP1L1*. **(B)** Representative FACS plots of CD11b cell-surface expression in HL-60 cells, five days following transduction with Cas9/sgRNA targeting *FIP1L1*, or a control sgRNA. **(C)** Percentage of CD11b+ HL-60 cells, measured by flow cytometry, from cells shown in (B). Data are mean +/- s.d. of three independent experiments. *** $p < 0.001$, two-tailed student's t-test. **(D)** Kaplan-meier survival curve of murine cell line xenografts. Mice were transplanted with HL-60 cells transduced with Cas9/sgControl or sgFIP1L1 (n = 4 for each group).

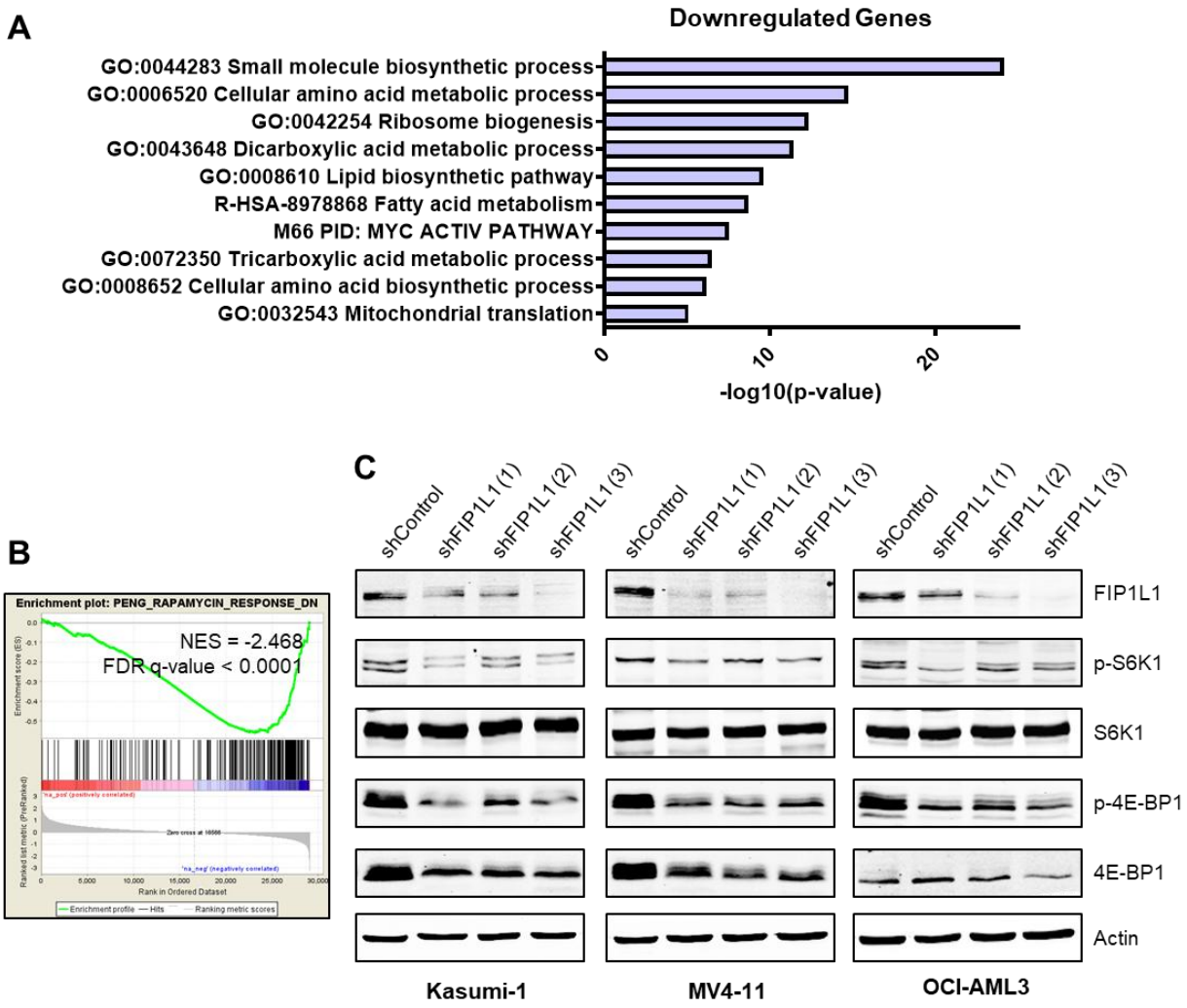
2.2.7 Targeting APA converges on MYC expression and mTORC1 signaling.

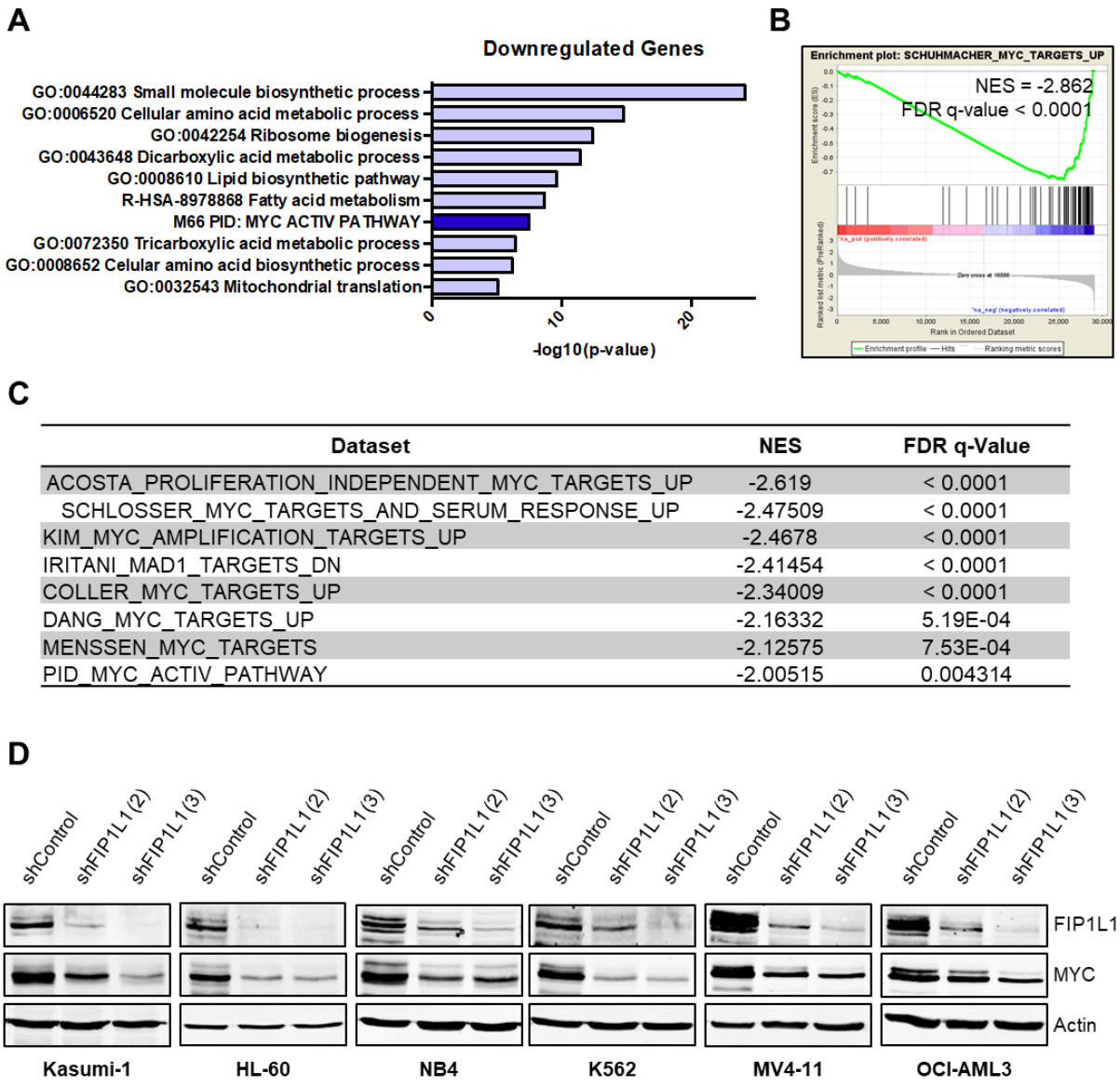
Since non-t(8;21) cells also differentiated upon FIP1L1 knockdown, we reasoned that common leukemia-promoting pathways must be altered that do not exclusively rely on AML1-ETO protein expression. To identify such pathways, we returned to our RNA-sequencing data from FIP1L1 knockdown in Kasumi-1 cells. Pathway analysis of significantly downregulated genes revealed a striking reduction in biosynthetic processes, including ribosome biogenesis, lipid biosynthetic pathways, and mitochondrial translation (**Figure 2.19A**). GSEA clarified this general observation as an overall reduction in mTORC1 signaling, a pathway that is constitutively active in leukemia, contributing to initiation and progression (117-119). Specifically, genes downregulated upon FIP1L1 knockdown matched those that are downregulated upon rapamycin treatment, an mTORC1 inhibitor (**Figure 2.19B**). To test whether FIP1L1 knockdown commonly attenuates mTORC1 signaling across AML, we performed shRNA-mediated knockdown in three mutationally diverse AML cell lines and observed a reduction in phosphorylated p70-S6 Kinase 1 (S6K1), the rapamycin-sensitive direct downstream target of mTOR kinase (120, 121) (**Figure 2.19C**).

Attenuation of mTORC1 signaling regulates myeloid differentiation through translational control of the c-MYC transcription factor (122). Indeed, we also saw potent downregulation of MYC target genes in our FIP1L1 knockdown RNA-sequencing dataset (**Figures 2.20A, 2.20B, and 2.20C**). We confirmed that MYC protein is reduced not only in Kasumi-1 cells, but in six mutationally diverse AML cell lines, explaining the global differentiation phenotype we observed by targeting APA across AML subtypes. (**Figure 2.20D**).

Figure 2.19. Biosynthetic metabolic processes regulated by mTORC1 signaling are reduced by targeting APA.

(A) Metascape analysis of significantly downregulated genes ($p < 0.01$, 1.5 FC) upon FIP1L1 knockdown in Kasumi-1 cells. **(B)** Plot from gene-set enrichment analysis (GSEA) of the RNA-sequencing experiment shown in Figure 2.8 showing genes downregulated upon rapamycin treatment. NES, normalized enrichment score; FDR, false discovery rate. **(C)** Western blot showing FIP1L1, phosphorylated p70-S6 Kinase 1 (p-S6K1), S6K1, p-4E-BP1, 4E-BP1, and actin (loading control) protein in the indicated AML cell lines following transduction with control shRNAs or shRNAs targeting *FIP1L1*. One representative experiment is shown of two independent experiments for each cell line. **(D)** Plots from gene-set enrichment analysis (GSEA) of the RNA-sequencing experiment shown in Figure 2.8 showing genes downregulated upon amino acid deprivation. NES, normalized enrichment score; FDR, false discovery rate.


A
Downregulated Genes
GO:0044283 Small molecule biosynthetic process
GO:0006520 Cellular amino acid metabolic process
GO:0042254 Ribosome biogenesis
GO:0043648 Dicarboxylic acid metabolic process
GO:0008610 Lipid biosynthetic pathway
R-HSA-8978868 Fatty acid metabolism
M66 PID: MYC ACTIV PATHWAY
GO:0072350 Tricarboxylic acid metabolic process
GO:0008652 Cellular amino acid biosynthetic process
GO:0032543 Mitochondrial translation
0
10
20
-log10(p-value)
B
Enrichment plot: PENG_RAPAMYCIN_RESPONSE_DN
NES = -2.468
FDR q-value < 0.0001
C
shControl
shFIP1L1 (1)
shFIP1L1 (2)
shFIP1L1 (3)
FIP1L1
p-S6K1
S6K1
p-4E-BP1
4E-BP1
Actin
Kasumi-1
MV4-11
OCI-AML3
D
Enrichment plot: PENG_GLUTAMINE_DEPRIVATION_DN
NES = -2.782
FDR q-value < 0.0001
Enrichment plot: PENG_LEUCINE_DEPRIVATION_DN
NES = -2.423
FDR q-value < 0.0001



Dataset	NES	FDR q-Value
ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_UP	-2.619	< 0.0001
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_UP	-2.47509	< 0.0001
KIM_MYC_AMPLIFICATION_TARGETS_UP	-2.4678	< 0.0001
IRITANI_MAD1_TARGETS_DN	-2.41454	< 0.0001
COLLER_MYC_TARGETS_UP	-2.34009	< 0.0001
DANG_MYC_TARGETS_UP	-2.16332	5.19E-04
MENSSEN_MYC_TARGETS	-2.12575	7.53E-04
PID_MYC_ACTIV_PATHWAY	-2.00515	0.004314

Figure 2.20. Targeting APA reduces MYC expression across AML cell lines.

(A) Metascape analysis of significantly downregulated genes ($p < 0.01$, 1.5 FC) upon FIP1L1 knockdown in Kasumi-1 cells. **(B)** Plot from gene-set enrichment analysis (GSEA) of the RNA-sequencing experiment shown in Figure 2.8 showing genes upregulated by the oncogenic c-MYC transcription factor. NES, normalized enrichment score; FDR, false discovery rate. **(C)** Table summarizing GSEA results from the same RNA-sequencing experiment, indicating a reduction in MYC downstream targets upon FIP1L1 knockdown. **(D)** Western blot showing FIP1L1, c-MYC, and actin (loading control) protein in the indicated AML cell lines following transduction with control shRNAs or shRNAs targeting *FIP1L1*. One representative experiment is shown of two independent experiments for each cell line.

2.3 Discussion

In summary, we have added to a growing field of knowledge regarding APA dysregulation in cancer. By profiling poly(A) site usage in AML patient blasts compared to healthy HSPCs, we showed that blasts exhibited global 3'UTR shortening, but CDS lengthening due to APA. These trends are opposite those seen in lymphocytic leukemia (52, 54), highlighting the complexity and context-specific nature of APA dysregulation in disease. Though global trends are traditionally reported, many genes exhibited significant differences in poly(A) site usage that counteract the trends, underlining our lack of knowledge regarding the importance of individual gene APA regulation and the global impact of APA in cancer. To address these unresolved questions, we targeted FIP1L1, the APA regulator most correlated with 3'UTR shortening in our AML patient cohort. We revealed a role of global APA dysregulation in blocking differentiation of leukemia cells, defined the direct and indirect impact of APA on the expression of crucial oncogenes, and introduced APA as a putative therapeutic target in AML.

Our results most prominently highlight the contribution of APA in mediating the hallmark differentiation block of leukemia. Despite the importance of APA in modulating healthy cellular differentiation across species (55-57), the role of dysregulated APA in cancer has been predominantly attributed to its impact on cellular proliferation (8, 41, 64, 65). Various studies report 3'UTR shortening and upregulation of cell-cycle and proliferation-related transcripts in cancer, such as *CCND1* (6, 63, 123) and *CDC6* (97). While not excluding the effects on differentiation, these studies merely imply that reduced differentiation is an indirect byproduct of enhanced proliferation in tumorigenesis. However, non-transformed cells with similar proliferative capacity as

transformed cells do not exhibit the same 3'UTR shortening, supporting an oncogenic role for APA that is proliferation-independent (6). Our study addresses this knowledge gap, emphasizing a direct role of APA in mediating the differentiation block of cancer cells. Targeting global APA dysregulation by FIP1L1 knockdown induced differentiation of mutationally diverse AML cell lines. In patients, *FIP1L1* expression was correlated to leukemia cell maturation, supporting the clinical relevance of this phenotypic finding. Mechanistically, we identified 3'UTR-APA of the t(8;21) fusion protein AML1-ETO, reporting for the first time that 3'UTR length can impact expression of a prominent oncofusion. Importantly, AML1-ETO is implicated in the differentiation block, but not the proliferative capacity, of myeloid leukemia cells (111). We also experimentally confirmed 3'UTR-APA mediated regulation of BAALC expression, a negative prognostic factor in leukemia that blocks differentiation in both AML (124) and congenital neutropenia (CN) (125).

In addition to direct APA regulation of leukemic oncogenes, we also demonstrated the indirect effect of targeting APA on leukemogenic pathways that hinder differentiation. Specifically, targeting FIP1L1 prompted downregulation of c-MYC, an oncogenic transcription factor well-known for its role in blocking hematopoietic differentiation (126, 127). Though MYC protein was commonly downregulated across AML mutational contexts, *MYC* polyadenylation was unchanged (**Figure 2.21A**). Indirect MYC downregulation may be mediated by attenuated mTORC1 signaling, since activated mTOR promotes MYC translation (122, 128). Though we observed attenuation of activity, *mTOR* APA was also unchanged (**Figure 2.21B**). Furthermore, MYC indirectly regulates mTORC1 signaling by transcriptional control of cell membrane

transporters that supply amino acids required for mTORC1 activity (129-131). Our data supports this feedback loop as rapamycin signatures were linked specifically to those for amino acid deprivation (**Figure 2.19D**). While the details of this indirect APA mechanism are unclear, we show that targeting APA disrupts this oncogenic positive feedback loop and promotes leukemia cell differentiation (**Figure 2.22A**). Intriguingly, mTORC1 inhibition was recently reported to regulate APA, promoting transcript lengthening (132). This observation highlights a second positive feedback loop between mTORC1 activity and APA that may heighten the anti-leukemic effects of targeting APA. Further work is warranted to identify the precise, possibly context-specific, APA events that disrupt this leukemogenic signaling network.

Altogether, our phenotypic and mechanistic data introduce APA as a putative target for differentiation therapy in AML. Historically, differentiation therapy by all-trans retinoic acid (ATRA) and arsenic trioxide (ATO) has been most effective in acute promyelocytic leukemia (APL) patients (133). Recently, interest in differentiation therapy has been reinvigorated by the efficacy of IDH inhibitors in AML patients with *IDH1/2* mutations (92-94). Our findings illuminate yet another possible target for differentiation therapy, that has the potential to expand the number of patients who can benefit from differentiation-based treatment. Consequently, the identification or synthesis of compounds that can target APA, specifically FIP1L1, are justified.

The identification of *FIP1L1* in leukemogenesis is also clinically intriguing since it is a member of two oncofusions found in chronic eosinophilic leukemia (CEL) and juvenile myelomonocytic leukemia / acute promyelocytic leukemia (JMML / APL): FIP1L1-PDGFRA (134) and FIP1L1-RARA (135, 136), respectively (**Figures 2.23A and**

2.23B). The reported contribution of FIP1L1 in FIP1L1-RARA is limited to receptor dimerization and activation, with no data regarding the impact on global polyadenylation (137). Similarly, reports regarding FIP1L1-PDGFRA describe FIP1L1 as dispensable, claiming that constitutive activation of the receptor is sufficient to induce leukemia in cell line and mouse models (138, 139). Despite this dismissal, not all signaling features dysregulated in CEL are recapitulated with only the C-terminal portion of PDGFRA (140). In both fusions, FIP1L1 retains the major protein-protein interaction Fip1 domain but lacks the C-terminal domain implicated in RNA-binding (26). While our experiments do not directly address the role of FIP1L1 RNA-binding in the observed phenotypes, our data support a likely consequence of FIP1L1 alteration in these fusions and overall contribution to pathogenesis. Consequently, our findings necessitate further study regarding the impact of these two oncofusions on polyadenylation dysregulation in patients.

In conclusion, we have demonstrated the importance of APA dysregulation in contributing to the differentiation block characteristic of AML. Thus, targeting APA may prove to be an effective therapeutic strategy for leukemia patients.

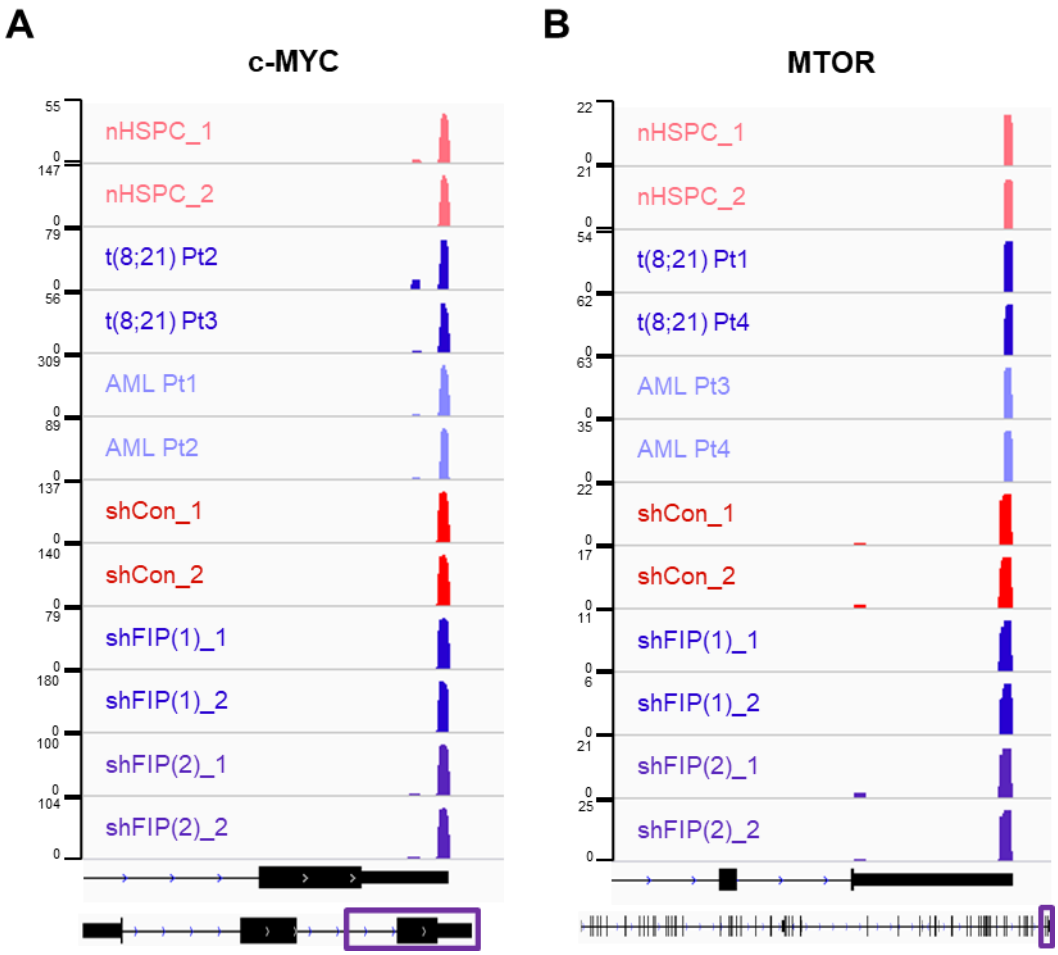


Figure 2.21. Polyadenylation of *MYC* and *MTOR* are not altered.

Genome browser tracks depicting sequencing reads in the *c-MYC* **(A)** and *mTOR* **(B)** genes obtained from 3'READS of HSPCs, AML patient blasts, and Kasumi-1 cells transduced with the indicated shRNA. The full *MYC* and *mTOR* genomic structures are shown (bottom) with the purple, boxed regions expanded (above).

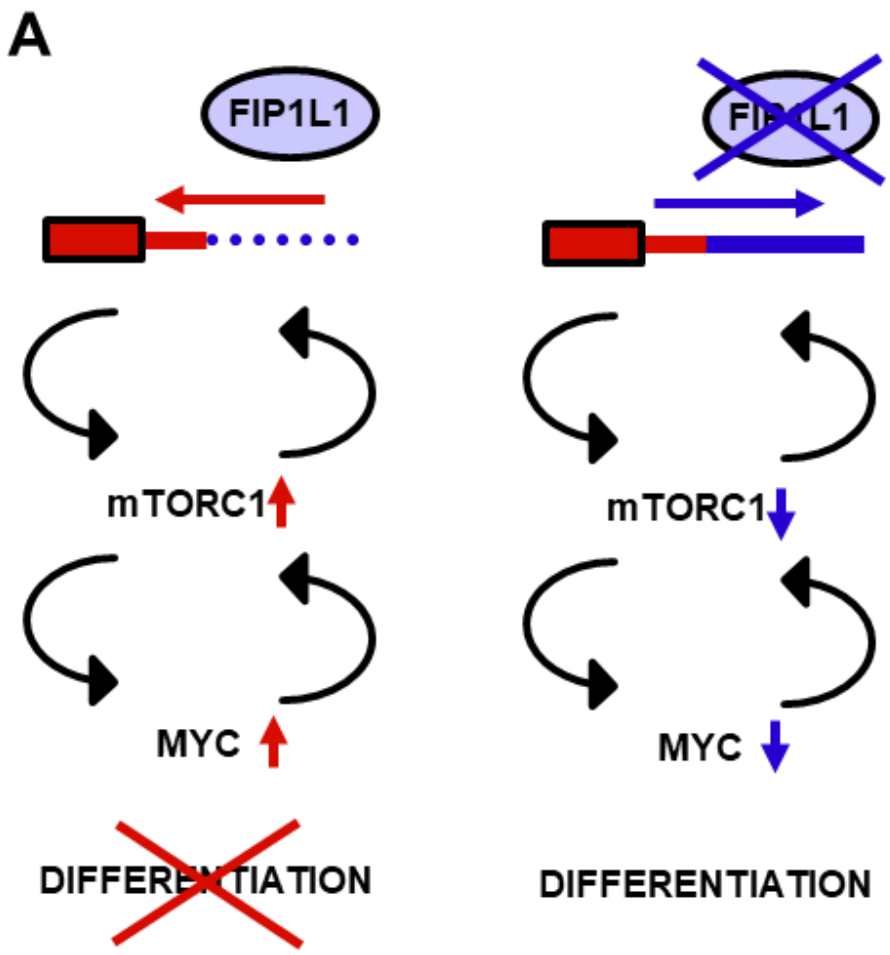


Figure 2.22. Model describing how targeting APA disrupts oncogenic signaling networks.

(A) Cartoon model showing the impact of targeting APA on positive oncogenic feedback loops. (Left) In cancer, mTORC1 and MYC positively regulate each other, blocking cellular differentiation. mTORC1 activity has recently been linked to transcript shortening in cancer, suggesting a second positive feedback loop between post-transcriptional regulation and cellular metabolism. (Right) Targeting APA by FIP1L1 knockdown reduced both mTORC1 signaling and MYC transcriptional networks, promoting leukemia cell differentiation. Further work will elucidate the precise post-transcriptional targets responsible for disrupting this oncogenic signaling network.

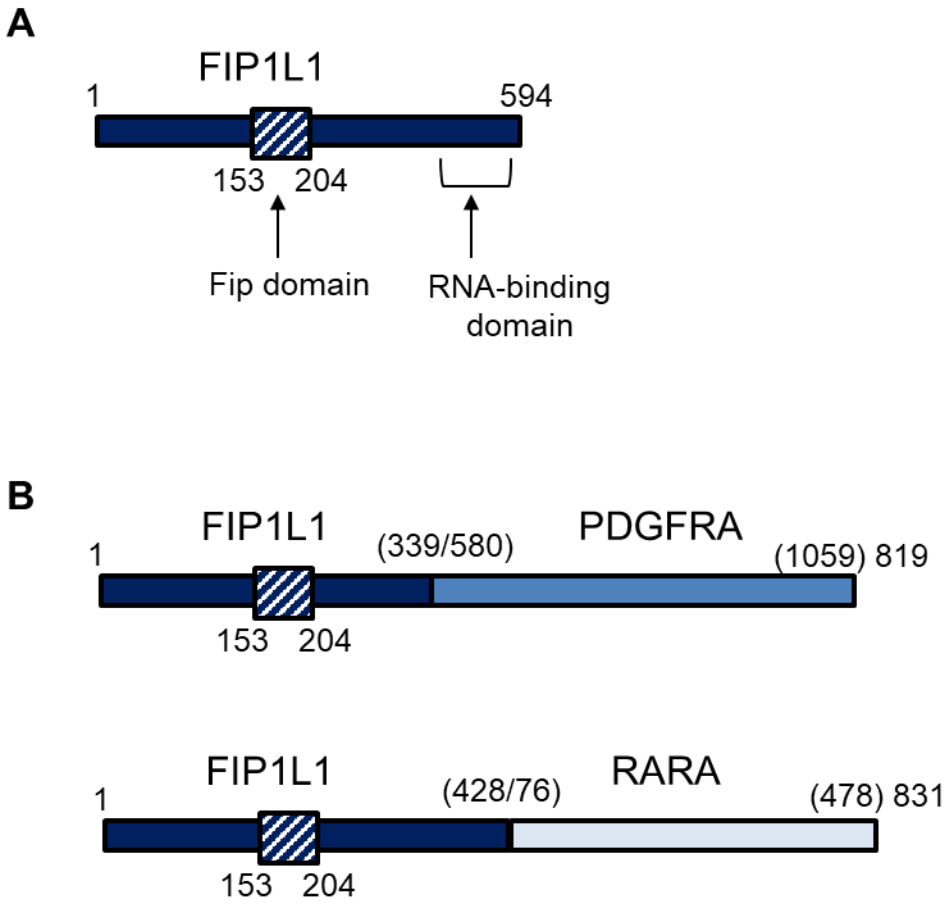


Figure 2.23. FIP1L1 is a member of oncogenic fusions in hematologic malignancy.

(A) Schematic of the full-length FIP1L1 protein. The conserved Fip domain and the arginine-rich RNA-binding domain are annotated. **(B)** Schematic of the two oncogenic fusions that consist of FIP1L1 fused to either PDGFRA or RARA. In both cases, the common breakpoints are represented with the corresponding amino acid contribution of each fusion partner.

2.4 Future Directions

While we have established FIP1L1 as a pertinent APA regulator in AML, contributing to the hallmark differentiation block in patients, many unanswered questions remain on the mechanism of FIP1L1 action and the feasibility of targeting APA regulators in AML. As previously discussed, our data also reinvigorates interest in understanding the role of FIP1L1 in oncogenic fusions seen in patients with CEL and JMML. Finally, we have not addressed the role of FIP1L1 on APA in healthy hematopoiesis.

Across AML cell lines of varying mutational contexts, we confidently established that FIP1L1 knockdown induces differentiation, a phenotype explained by common attenuation of mTOR signaling and downregulation of the c-MYC oncogenic transcription factor. Mechanistically, we only profiled APA changes in the t(8;21) cell line Kasumi-1. Consequently, we have a limited understanding regarding the breadth and overlap of regulated transcripts across AML. The most pressing future experiment is to perform 3'READS (95) and parallel RNA-sequencing in additional AML cell lines shown in this study. These data, when compared to our data in Kasumi-1 cells, would directly address this knowledge gap and allow us to establish a set of commonly regulated transcripts that may disrupt oncogenic mTOR and MYC pathways. These data will also illuminate context-specific APA events, such as APA regulation of AML1-ETO, that uniquely contribute to AML pathogenesis.

In addition to expanding our understanding of target transcripts, we do not understand why FIP1L1 knockdown induces lengthening of some transcripts and

shortening of others. Nor do we have data that explain why we saw global 3'UTR lengthening, but CDS shortening upon knockdown. FIP1L1 is an RNA-binding protein (RBP), that is implicated in binding to U-rich or AU-rich regions (26-28). There is some discrepancy in where binding occurs, with one group reporting binding between the PAS hexamer and cleavage site (27), another reporting upstream of the PAS hexamer (28), and a final group concluding binding of both regions (26). One of these reports suggests that FIP1L1 knockdown induces proximal to distal APA when sites are relatively far apart, and distal to proximal APA when sites are closer together (28). However, this was merely an observation from a study limited to murine embryonic stem cells. Additionally, our data demonstrating potent CDS-shortening upon FIP1L1 knockdown does not support this model since CDS APA sites are always quite far apart. From our 3'READS data, we performed motif enrichment in regions surrounding poly(A) sites with significantly different usage. Interestingly, we noticed sites that were more utilized upon FIP1L1 knockdown tended to consist of poly(A) or poly(U) rich regions downstream of the PAS hexamer (**Figure 2.24A and 2.24B**, blue highlights). By contrast, sites that were less utilized were enriched for poly(A) or poly(U) regions upstream of the PAS hexamer (yellow highlights). These preliminary data intriguingly suggest a location-specific effect of FIP1L1 binding on poly(A) site usage: upstream binding promotes and downstream binding inhibits poly(A) site usage. Ultimately, this computational analysis of our 3'READS data only indirectly addresses FIP1L1 binding and impact on APA specificity. Therefore, we also propose to perform an eCLIP (141) experiment in at least three different AML cell lines to expand upon these preliminary results and directly address the mechanism of APA regulation by FIP1L1. This experiment will reveal a

consensus binding motif, likely consisting of poly(U) and poly(A) elements in agreement with previous studies. When paired with differentially regulated poly(A) sites deduced by 3'READS of the same cell lines, we will observe patterns that help us model how FIP1L1 knockdown variably affects APA.

Our data also supports the therapeutic potential of targeting APA in AML patients. Since differentiation therapy is already an exciting avenue of treatment (94, 133), targeting APA represents a new class of regulation that might have unique advantages compared to current therapies. To lay the foundation for this type of work, we would first test the effect of targeting FIP1L1 in AML mouse models. Our HL-60 cell line xenograft data supports the feasibility of this approach mitigating leukemia progression (**Figure 2.18D**), however it is not sufficiently rigorous to make this conclusion. We propose to design a morpholino antisense oligonucleotide (MAO) that targets *FIP1L1* for degradation or disrupts its normal splicing patterns to ultimately render the transcript non-functional. We will use conventional AML murine models, AE9a (142) and BCR-ABL (143) retroviral transduction and transplantation, and inject the MAO as a treatment modality. By monitoring survival and leukemic burden via peripheral blood and bone marrow parameters, we will glean data that better supports the possibility of targeting APA as a therapeutic avenue. If these experiments are promising, then a small molecule screen for compounds that target FIP1L1 is warranted, followed by further testing of these compounds in cell lines and mouse models.

As previously mentioned, FIP1L1 is also a part of two oncogenic fusion proteins: FIP1L1-PDGFRA (134) and FIP1L1-RARA (135, 136). While previous studies dismiss the importance of FIP1L1 alteration in driving leukemic transformation (138, 139), not all

signaling features dysregulated in these patients are recapitulated with the C-terminal fusion partner alone (140). These observations, when combined with our data showing the critical role of FIP1L1 in mediating blood cell maturity, justify further investigation of the role of FIP1L1 alteration in these fusions on leukemia pathology. In both fusions, FIP1L1 is missing a small portion of the C-terminus, a region containing a putative RNA-binding domain (26) (**Figure 2.22A and 2.22B**). Therefore, we would first test the ability of both FIP1L1 fusions to (1) bind poly(A) machinery members and (2) bind RNA. We would address protein-protein interactions by co-immunoprecipitation and expect that these fusions can still bind to the CPSF complex. We would use RIP-qPCR to determine RNA binding of a subset of targets established from the eCLIP experiments proposed above. Next, we would assess the impact of fusion generation on global APA patterns. FIP1L1-PDGRFA is generated by an interstitial chromosomal deletion, not a chromosomal translocation. A previous group reported how to generate this deletion using CRISPR/Cas9 technology to target the affected introns (139). We would mimic this strategy and generate the fusion in K562 cells followed by 3'READS and RNA-sequencing to directly address whether fusion generation impacts APA and cell maturity. Overall, these experiments would produce exciting new data on the action of these fusions and contribution to malignancy.

Lastly, we would like to perform a study on the role of FIP1L1 in normal hematopoiesis. A study in murine embryonic stem cells supports a role of FIP1L1 in normal cellular differentiation (28). Indeed, transcript lengthening is a widespread trend upon differentiation of various cell contexts (55-57), a mechanism and phenotype that mimic what we saw upon FIP1L1 knockdown in AML cells. Interestingly, *FIP1L1*

expression does not consistently decrease during hematopoiesis (**Figure 2.25A and 2.25B**), so we cannot confidently hypothesize that FIP1L1 will play a similar endogenous role in healthy hematopoiesis. To address this knowledge gap, we would knock down FIP1L1 in human CD34+ HSPCs, followed by colony formation assays, long term culture initiating cell (LTC-IC) assays, and murine xenografts to assess the impact of FIP1L1 expression on HSC self-renewal and lineage output. A mouse model of FIP1L1 has not been reported. Generating a transgenic mouse model and analyzing the impact of homozygous and heterozygous knockout on the hematopoietic system would also address the functional role of FIP1L1 in proper blood development.

Altogether, uncovering FIP1L1 as an important APA regulator in AML has opened a lot of additional, exciting avenues for further study. We expect that pursuing any of these would both enhance mechanistic understanding of APA and illuminate disease-causing mechanisms.

Figure 2.24. Enriched sequence motifs surrounding differentially utilized poly(A) sites upon FIP1L1 knockdown.

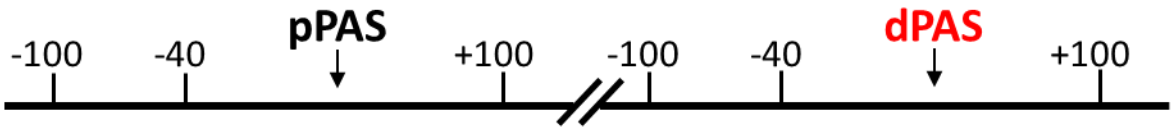
(A) Enriched hexamer sequences in three regions (-100 to -40, -40 to PAS, and PAS to +100) flanking the proximal (pPAS) and distal (dPAS) poly(A) sites differentially utilized upon FIP1L1 knockdown. The top (red) represents sequence features among transcripts that were lengthened. The bottom (blue) represents those among transcripts that were shortened. Blue highlights significantly enriched motifs near more utilized sites. Yellow highlights those near less utilized sites. **(B)** Enriched hexamer sequences flanking intronic poly(A) (IPA) sites of CDS-APA events that were more (red – top) or less (blue – bottom) utilized upon FIP1L1 knockdown. Again, blue highlights enriched motifs in more utilized IPA sites and yellow those in less utilized sites. All numerical values indicate significance of enrichment (-log10(p-value)) using a Fisher's exact test.

A

Prx < Dis (Lengthened)

UCAUCC 3.6	CGUGUG 3.51	GUGUGU 4.44
UGGAGG 3.12	ACAUAU 2.49	CGCGCG 3.03
UUGUAG 2.98	UCUUAA 2.47	ACAUUA 2.85
ACUACU 2.82	CGAGGC 2.46	CCAACC 2.81
UUAUAA 2.75	GCGGAA 2.46	CCCCUU 2.54

ACAAUC 3.33	GGACAA 4.57	UUUUUU 4.69
UAGGUG 3.1	CCUCAA 3.19	AAAAAA 3.58
UACAUG 2.97	UGCAGG 2.91	CCAAUG 3.18
CCACAU 2.82	GCCCUC 2.9	UCACCG 3.03
GUGUGG 2.68	UCUAAU 2.84	CAAAGA 2.91



AAAAAA 5.16	GCUGAU 3.73	AAUAAA 14.18
UACCAG 4.19	GAAUAA 3.06	AAAUAA 7.93
CAGUUG 2.79	UAGGGA 2.7	AUAAAA 5.06
GAGCUA 2.66	CAAUGU 2.58	UGUUGU 4.8
UCCAAC 2.58	AGAAUA 2.49	AUUUUU 4.5

UUUUUU 4.93	AUUUUU 6.1	GUUUAU 3.69
UAAAUG 3.92	UUUUUU 5.58	CGCCCG 2.95
GUUUAU 3.8	UAUUUU 4.34	AGUUUA 2.87
GAUUUU 3.52	UUAUAA 4.27	GAAUUG 2.83
UUCACG 3.49	AAGUUU 3.89	CCUUGA 2.66

Prx > Dis (Shortened)

B

UR PAS up

UUCAUG 3.46	AUGAAC 4.46	AAAAAA 7.11
UCACAG 3.04	AUAAAG 3.56	UUCUUC 3.59
GGGAAC 2.71	UGGUCA 3.32	UGAAUG 3.53
AUAUAU 2.68	UUGGUC 2.68	UCAUUU 3.46
CACAGG 2.66	GAAAGU 2.59	CUAUCA 3.28

AAAAAA 15.64	AAAAAA 82.33	GGGUCC 3.68
CCAGCC 5.38	UGGUAU 3.56	UGGGCC 3.4
AAUAAA 4.51	UCUAAU 3.36	GCACGU 3.37
CAGCCU 4.41	UGCUCU 3.35	AGGAUC 3.27
UUCCAC 4.34	UGUCUA 3.03	GUCUGU 3.19

UR PAS down

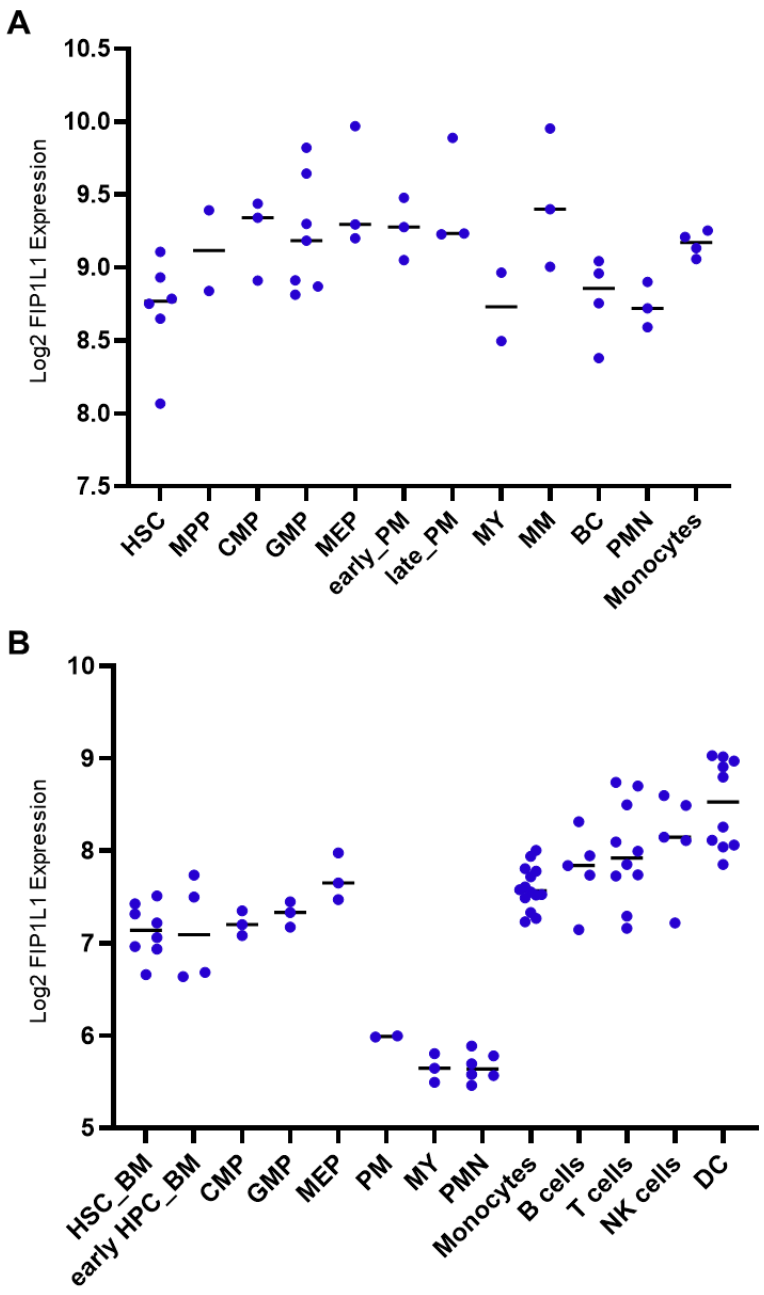


Figure 2.25. *FIP1L1* expression in healthy hematopoiesis.

FIP1L1 expression was downloaded from Bloodspot (bloodspot.eu) (144). Data shown are from Affymetrix probe 221007_s_at. **(A)** Normal hematopoiesis dataset (GSE42519). **(B)** HemaExplorer dataset (GSE17054, GSE19599, GSE11864, E-MEXP-1242). HSC, hematopoietic stem cell; HPC, hematopoietic progenitor cells; MPP, multipotential progenitors; CMP, common myeloid progenitor cells; GMP, granulocyte monocyte progenitors; MEP, megakaryocyte-erythroid progenitor cells; PM, promyelocyte; MY, myelocyte; MM, metamyelocyte; BC, band cell; PMN polymorphonuclear cells; NK, natural killer cells; DC, dendritic cells.

2.5 Materials and Methods

Primary patient samples and healthy HSPCs

Patient samples were obtained from UC San Diego Health with written consent and in accordance with the university-approved Institutional Review Board (IRB) protocol. Bone marrow and peripheral blood patient samples were separated using Ficoll-Pacque (VWR, #17-1440-02) and frozen down until further use. Samples were thawed quickly at 37°C, then immediately diluted in 10 mL of 1x PBS supplemented with 1 mg/mL DNase (Sigma, #11284932001). Cells were washed in 1x PBS supplemented with 2%FBS. Live, mononuclear cells were separated by density gradient centrifugation using Ficoll-Paque and washed again in 1x PBS 2%FBS. Magnetic bead CD34-enrichment was performed using MACS Miltenyi Kit 130-046-702 according to the manufacturer's protocol. An aliquot of CD34-enriched leukemic blasts was analyzed by flow cytometry to confirm that cells were >95% CD34+. RNA was extracted from patient blasts using Trizol reagent (ThermoFisher Scientific, #15596026).

CD34+ HSPCs from healthy donors were obtained from Fred Hutchinson Cooperative Center for Excellence in Hematology (CCEH) (Seattle, Washington). Cells were thawed quickly and serially diluted with 1x PBS supplemented with 2% FBS. Cells were resuspended in Trizol for RNA extraction.

Lentiviral production and transduction

For FIP1L1 shRNA knockdown experiments, lentivirus was produced by transfecting HEK293T cells with 3 µg of the respective pLKO-based lentiviral vector, 5 µg of psPAX2, 2.5 µg of pMD2.G, and 42 µL of polyethylenimine (PEI) in 1 mL of Opti-

MEM I Reduced Serum Medium (Gibco, #31985-070). 24 hours post-transfection, the HEK293T cell supernatant was replaced with 7 mL of RPMI supplemented with 10% FBS. 24 hours later, the cell supernatant containing lentiviral particles was passed through a 0.45 µm syringe filter, supplemented with polybrene (final concentration 4 µg/mL), and added to the appropriate cell line at ~0.3-0.5 million cells per mL. Cells were transduced in 6-well plates by centrifugation (2,000 x g) for 3 hours at 32°C in an Allegra X-12R centrifuge (Beckman Coulter; Brea, CA) on two consecutive days. 24 hours following the second transduction, cells were resuspended in fresh RPMI media supplemented with 1 µg/mL puromycin. After 48 hours of puromycin selection, cells were diluted 1:2 and maintained at 0.5 µg/mL puromycin until flow cytometric analysis or RNA/protein isolation. FIP1L1 pLKO shRNA clones utilized were TRCN0000307316 (shFIP1L1 #1), TRCN0000074418 (shFIP1L1 #2), and TRCN0000074420 (shFIP1L1 #3). The pLKO control shRNA utilized was Addgene plasmid #1864.

3'RNA sequencing

RNA was isolated from healthy CD34+ HSPCs, patient samples, and FIP1L1 knockdown Kasumi-1 cells using Trizol reagent per the manufacturer's instructions. 3'READS library preparation was performed as previously described (95, 96). Libraries were sequenced on an Illumina HiSeq4000 (San Diego, CA) and data were processed as previously described (95, 96). Transcript shortening or lengthening between two experimental groups was calculated by the relative expression difference (RED), defined as the difference in the ratio of dPAS isoform abundance to pPAS isoform abundance (62). Significant APA events were determined by a Fisher's exact test. For differential APA analysis upon FIP1L1 knockdown in Kasumi-1 cells, three biological

replicates of shControl were compared to seven FIP1L1 knockdown replicates (four shFIP1L1 #1 and three shFIP1L1 #2).

RNA-sequencing

Library preparation of total RNA from healthy HSPCs and patient blasts was performed using the TruSeq Stranded mRNA kit (Illumina, 20020594) followed by sequencing on an Illumina HiSeq4000. Raw RNA-seq reads were aligned and mapped using HISAT2 (145), followed by featureCounts quantification, using the usegalaxy public supercomputing platform. Reads were normalized using the DESeq2 software (146) package. Further statistical analyses were performed using R software (v3.6).

For FIP1L1 knockdown in Kasumi-1 cells, RNA from four biological replicates were prepared for shControl, shRNA #1 (TRCN0000307316), and shRNA #2 (TRCN0000074418). For differential gene expression analysis, all eight knockdown replicates were compared to the four control replicates. Libraries were prepared and sequenced by Novogene (Sacramento, CA). Reads were aligned using STAR software (147) and differential expression was performed using DESeq2 (146).

TCGA LAML raw RNA-Seq data were acquired from the Genomic Data Commons (GDC). Sailfish was used to process and align raw reads and quantify feature counts (148). To generate cohorts describing top and bottom 10% of *FIP1L1* expression, feature counts were first normalized using DESeq2. Cutoffs representing the 90th and 10th percentile of *FIP1L1* expression were calculated. DESeq2 (146) was used to perform differential gene expression analysis between the top 10% and bottom 10% cohorts.

Cell proliferation

For Kasumi-1 proliferation curves, cells were transduced as described and seeded at 200,000 cells/mL following 48 hours of puromycin selection (‘day 0’). Cells were then counted using a BioRad TC20 Automated Cell Counter every two days until the conclusion of the experiment.

Gene set enrichment analysis (GSEA)

GSEA (149, 150) was performed using the desktop application and the ‘GSEAPreranked’ function set to 1,000 permutations. For both the FIP1L1 knockdown experiment and TCGA dataset analysis, ranked input data was generated based on the log2(fold change) from differential gene expression analysis.

Gene ontology (GO) and pathway analysis

Gene ontology (GO) and pathway analysis was performed using the Metascape online tool (metascape.org) (151). For pathway enrichment of genes with differential poly(A) site usage, the described groups of genes were compared to the full list of human genes. Similarly, differentially expressed genes (FIP1L1 knockdown experiment or TCGA dataset analysis) were compared to the full list of human genes.

Luciferase assays

The psiCHECK-2 dual luciferase reporter vector (Promega) was utilized for all luciferase assays.

For *BAALC* and *MAPKAPK3*, the 3'UTR that results from polyadenylation at the most distal polyadenylation sequence (PAS) and the most proximal PAS were each

individually subcloned downstream of renilla luciferase in the psiCHECK-2 vector. First, the full length 3'UTR for each was amplified from Kasumi-1 genomic DNA using the following primer pairs: *MAPKAPK3* (5'– CTCATGGGGCCTTGGAGG –3' / 5'– CCAAGAGAAAATGATGAAATGGTG –3') and *BAALC* (5'– CAGAGAGTCCAAGCAGAAGG –3' / 5'– CTGCCCATATATGTTTAAAATTAATGAC –3'). All canonical PAS (AATAAA) were removed by overlap PCR to prevent premature cleavage and to ensure that the full length of the 3'UTR was present downstream of renilla in the resulting transcript. The short 3'UTRs were then subcloned using the aforementioned forward primers paired with the following reverse primers: *MAPKAPK3* (5'–GTGACAAATTAAAACACAACAAAATAATC –3') and *BAALC* (5'– TGAACTGCACATTTGCAGAAC –3').

For *AML1-ETO*, the full-length 5.2kb *ETO* 3'UTR was amplified from KG-1a genomic DNA using the following primer pair: (5'– ACGTGAACTCAGAACTGTCGGAG –3' / 5'– CATGATTAGGCAAACACAAC –3'). All canonical PAS (AATAAA) were removed by overlap PCR. The short 3'UTRs were then subcloned using the aforementioned forward primer paired with the following reverse primers: *ETO* 1kb (5'-TTTTTAATTAAAAATCCAAAC-3'), *ETO* 1.3kb (5'-CTTGAGGACAACCAAAAAGAG-3'), *ETO* 3.7kb (5'-TTTTTTTCAACTTTACACAGTAAAG-3').

Kasumi-1 cells were then nucleofected with 0.5-1 µg of psiCHECK-2 vector containing the longer 3'UTR or an equimolar amount of the shorter 3'UTR vector. The appropriate amount of DNA was mixed with 1.5 million Kasumi-1 cells in 100 µL of nucleofection buffer (140 mM Na_2HPO_4/NaH_2PO_4 (pH 7.2), 5 mM KCl, 15 mM $MgCl_2$). Cells were then nucleofected using program P-019 of the AMAXA II Nucleofector

(Lonza; Basel, Switzerland) and carefully transferred to pre-warmed RPMI supplemented with 20% FBS and no penicillin/streptomycin. Cells were cultured at 37°C and 5% CO_2 for 48 hours prior to performing luciferase assays. The Dual-Luciferase Reporter Assay System (Promega, E1980) was utilized per manufacturer's instructions. Firefly and renilla luciferase fluorescence were measured using the Monolight 3010 luminometer (BD Pharmingen; Franklin Lakes, NJ).

Cell lines

Kasumi-1 and HL-60 cell lines were purchased from ATCC and maintained as low-passage stocks. The NB4 cell line was obtained from Daniel Tenen (Harvard University, Boston). The OCI-AML3 cell line was kindly provided by Suming Huang (University of Florida, Gainesville, FL). The MV4-11 cell line was generously provided by Tannishtha Reya (UC San Diego, La Jolla, CA). AML cell lines were cultured in RPMI supplemented with 10% FBS and 100U/mL penicillin/streptomycin. HEK293T cells were purchased from ATCC and cultured in DMEM supplemented with 10% BCS and 100U/mL penicillin/streptomycin. All cells were cultured at 37°C and 5% CO_2.

Western blotting

Cells were lysed on ice for 10 minutes in cold RIPA buffer (50 mM Tris, 1 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors (Roche, # 11873580001) and phosphatase inhibitors (Roche, #4906837001). Lysates were cleared by centrifugation (10,000 x g) at 4°C, mixed with loading buffer, and denatured at 95°C. Lysates were run using conventional SDS-PAGE techniques.

Primary western antibodies used for this study included: FIP1L1 (C-10) (Santa Cruz Biotechnology, sc-398392), β-Actin (Sigma, A1978), BAALC (Proteintech, #24997-1-AP), RUNX1-ETO (generated by Covance) (142), α-Tubulin (Developmental Studies Hybridoma Bank, #12G10), c-MYC (Abcam, ab32072), Phospho-p70 S6 Kinase (Thr389) (Cell Signaling, #9234), p70 S6 Kinase (Cell Signaling, #2708), Phospho-4E-BP1 (Thr37/46) (Cell Signaling, #2855), and 4E-BP1 (Cell Signaling, #9452). Secondary western antibodies included: IRDye 800CW goat anti-rabbit IgG (LI-COR, #926-32211), IRDye 800CW goat anti-mouse IgG (LI-COR, # 926-32210), and IRDye 680RD goat anti-mouse IgG (LI-COR, #926-68070). Dilutions of all antibodies conformed to the manufacturer's recommendations. Membranes were scanned using a LI-COR Odyssey Classic Infrared Imaging System (Lincoln, Nebraska). Image analysis and densitometry were performed using the LI-COR Application Software Version 3.0.

RT-qPCR

RNA was isolated from transduced Kasumi-1 cells four days following puromycin selection using Trizol Reagent per the manufacturer's instructions. cDNA was prepared from 0.5 – 1 μg of RNA using the Quanta 5x qScript cDNA Supermix (Quanta, #95048) per the manufacturer's instructions. Quantitative PCR (qPCR) was performed using KAPA SYBR FAST qPCR Master Mix (2X) Universal (KAPA Biosystems, #KK4618) on a BioRad CFX Connect machine (Hercules, CA). Reactions were performed in technical duplicate or triplicate. Data were analyzed using the conventional delta-delta Ct method and normalized as described in the figure legends.

Primers used:

BAALC

pPAS: 5'– TCTTCAGTGTCCTTCACGGCA –3' / 5'– CAGTCTTGCCAGGGACTCAG –3'

BAALC mPAS:

5'– AGCACCTGGTTGATGTGTATTC –3' / 5'– TTCCCAATCCCAGGCATGTT –3'

BAALC CDS:

5'– GCCCTCTGACCCAGAAACAG –3' / 5'– CTTTTGCAGGCATTCTCTTAGCA –3'

β-Actin:

5'– TCCCTGGAGAAGAGCTACGA –3' / 5'– AGCACTGTGTTGGCGTACAG –3'

AML1-ETO 1kB PAS:

5'– ATCAGCCAGCTGCCCTAAAT –3' / 5'– GGCTGGAGCTGAAGCCACCAT –3'

AML1-ETO 3.7kB PAS:

5'– GGCATCATGCTTTTGTCAGCA –3' / 5'– AGATTCACAGGTTGACAGACCA –3'

Wright-Giemsa staining

To examine HL-60 and NB4 cell morphology, cytospins (Cytopro 7620 Cytocentrifuge, Wescor) were prepared on shRNA transduced cells three days following puromycin selection. 24 hours post-cytospin, cells were fixed with methanol. Slides were then stained in Wright solution (Sigma, WS16) followed by a PBS wash. Excess

stain was rinsed with water, then slides were stained in Giemsa stain (Sigma, GS500) diluted 1:10 in sodium phosophate buffer (pH 6.4) containing 10% Triton-X. Excess stain was again rinsed, and slides were dried overnight prior to mounting. Microscope images were taken using an Olympus BX51 microscope equipped with a DP71 digital camera using the DP-BSW acquisition software (Olympus Corporation; Shinjuku, Tokyo, Japan).

Flow cytometry

Flow cytometric analyses were performed on a BD FACSCanto flow cytometer using the BD FACSDiva acquisition software. Appropriate single-stained and fluorescence-minus-one (FMO) controls were used to determine compensation and define gates. FlowJo software (FlowJo, LLC) was used for post-acquisition data analysis. CD34-APC (BD Pharmingen, #560940) and CD11b-APC (ThermoFisher Scientific, CD11b05) antibody concentrations were determined by titration and the same number of cells per sample were stained for each experiment. Propidium iodide (PI) was used to stain and exclude dead cells.

AML cell line xenograft

HL-60 cells were transduced with lentivirus expressing Cas9 plus the control sgRNA or sgRNA targeting *FIP1 L1* and then cultured in puromycin using the previously described protocol. Approximately 6 hours prior to transplantation of HL-60 cells, NOD.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ ('NSG') mice were sub-lethally irradiated with 2.5 Gy. HL-60 cells were collected, washed in cold PBS, and resuspended at 0.5 x 10^7 cells / mL in PBS. 1.0 x 10^6 cells (200 µL) were transplanted per mouse via tail vein injection

(n = 4 for each group). Mice were monitored regularly and euthanized when they showed symptoms of AML burden: lethargy, hunched posture, or hind-leg paralysis. Confirmation of AML development was performed by flow cytometric analysis of CD45 expression on cells in the bone marrow or spleen of euthanized mice. All murine experiments were approved by the UCSD Institutional Animal Care and Use Committee (IACUC). NSG mice were obtained from Jackson Lab (#005557).

Statistical analyses

All statistical analyses were performed using GraphPad Prism Software (Version 8.3.1). All tests utilized are documented in the respective figure legend. P values are denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data availability

3'RNA sequencing and standard RNA-sequencing data of patient samples and healthy HSPC controls were deposited to the Gene Expression Omnibus (GEO), accession GSE146657. FIP1L1 knockdown and control Kasumi-1 cell RNA-sequencing and 3'RNA sequencing were also deposited to GEO, accession GSE146475. All DNA constructs generated from this study will be made available by email request or will be submitted to Addgene for purchase.

Acknowledgments

Chapter 2, in part, has been submitted for publication by Davis AG, Johnson DT, Zheng D, Wang R, Jayne ND, Liu M, Stoner SA, Zhou JH, Ball ED, Tian B, and Zhang DE. The dissertation author is the primary investigator and writer of this manuscript.

Chapter 3: A CRISPR RNA-binding protein screen reveals regulators of RUNX1 isoform generation

The proper balance of hematopoietic stem cell (HSC) self-renewal and differentiation is critical for normal hematopoiesis and disrupted in hematologic malignancy. Among regulators of HSC fate, transcription factors have a well-defined, central role and mutations promote malignant transformation. More recently, studies have illuminated the importance of post-transcriptional regulation by RNA-binding proteins (RBPs) in hematopoiesis and leukemia development. However, the RBPs involved and breadth of regulation are only beginning to be elucidated. Furthermore, the intersection between post-transcriptional regulation and hematopoietic transcription factor function is poorly understood. Here, we studied the post-transcriptional regulation of *RUNX1*, a key hematopoietic transcription factor. Alternative polyadenylation (APA) of *RUNX1* produces functionally antagonistic protein isoforms (RUNX1a versus RUNX1b/c) that mediate HSC self-renewal versus differentiation, an RNA-processing event that is dysregulated in malignancy. Consequently, RBPs that regulate this event directly contribute to healthy and aberrant hematopoiesis. We modeled *RUNX1* APA using a split GFP minigene reporter and confirmed the sensitivity of our model to detecting changes in RNA-processing. We utilized this reporter in a CRISPR screen consisting of single guide-RNAs exclusively targeting RBPs and uncovered HNRNPA1 and KHDRBS1 as antagonistic regulators of RUNX1a isoform generation. Overall, our study provides mechanistic insight into the post-transcriptional regulation of a key hematopoietic transcription factor and identifies RBPs which may have a widespread, important function in hematopoiesis.

3.1 Introduction

Hematopoiesis is dependent on the proper balance between hematopoietic stem cell (HSC) self-renewal and differentiation. Perturbation in either direction is the basis of various hematologic malignancies. Among molecular mechanisms governing HSC fate, there is growing interest in the role of post-transcriptional regulation by RNA-binding proteins (RBPs), ignited by the identification of common splice factor mutations in leukemia and myelodysplastic syndrome (MDS) (152-155). In addition to splicing factors, RBPs that regulate RNA methylation (87), editing (156), and translation (157) have been described to directly impact HSC self-renewal and differentiation. However, only a small subset of all RBPs have been implicated in hematopoietic regulation and leukemia development. As such, the mechanistic impact and breadth of RBP involvement in hematopoiesis is only beginning to be understood.

By contrast, transcription factors have well-defined roles in regulating proper hematopoiesis and mutations are common in hematologic malignancies. Interestingly, post-transcriptional regulation of various hematopoietic transcription factors (*GATA1* (158), *IKZF1* (159), *RUNX1* (160), *SCL* (161), and *TEL/ETV6* (162)) produces functionally distinct isoforms, highlighting the intersection between these two regulatory processes. Consequently, elucidating the post-transcriptional mechanism of a key hematopoietic transcription factor will illuminate additional RBPs with a critical role in hematopoiesis.

RUNX1 is a hematopoietic transcription factor that plays a direct role in regulating HSC fate through the antagonistic action of its three major protein isoforms

(163, 164). The long RUNX1b/c isoforms use alternative promoters and differ by 27 amino acids at their N-terminus (163, 165). They both contain the runt homology domain (RHD) for DNA-binding and heterodimerization with CBFβ and the downstream transcriptional regulatory domain (TRD) which recruits essential co-factors (166). RUNX1a is a C-terminally truncated RUNX1 isoform, generated by alternative polyadenylation (APA) (163, 167), that retains the RHD but lacks the TRD (164). This structural difference confers enhanced DNA binding (164, 168) and divergent effects on target gene transcription when compared to RUNX1b/c (164, 169, 170). RUNX1a overexpression expands functional HSCs in vitro and in vivo (170-172), retards hematopoietic differentiation (164, 171), and enhances engraftment potential of murine bone marrow (BM) cells following transplantation (160, 170, 171). Conversely, RUNX1b/c induce HSC quiescence (173), promote differentiation (164, 174), and abrogate engraftment of transplanted murine BM cells (160).

In a healthy hematopoietic system, RUNX1b/c are the dominantly expressed isoforms, whereas RUNX1a represents a minor fraction of the total isoform pool (160, 163). RUNX1a expression is further restricted to immature hematopoietic stem and progenitor cells (HSPC) (160), suggesting that *RUNX1* APA is dynamically regulated during hematopoietic differentiation. Despite this restricted expression, short isoforms of RUNX1 are conserved across species (175-177), playing an important role in healthy HSC pool maintenance (178). Because RUNX1a mediates HSC expansion, overexpression can be leukemogenic (169). Indeed, *RUNX1a* is overexpressed in some patients with acute myeloid leukemia (AML) (164, 169), acute lymphoblastic leukemia (ALL) (169), and MDS (179). Importantly, overexpression is achieved by a change in the

relative ratio of *RUNX1a* to total *RUNX1* transcript, linking post-transcriptional mechanisms to aberrant transcription factor function (164, 179).

Collectively, these observations support the hypothesis that post-transcriptional regulators of RUNX1 isoform generation contribute to HSC biology and protect against leukemia development. So far, the RBPs responsible for this regulation and the cis-acting elements to which they bind are unknown. Here, we devised a fluorescent minigene model which accurately recapitulates RUNX1 isoform generation. We confirmed the sensitivity of our construct that monitors RUNX1a formation and utilized this minigene to perform a CRISPR RBP screen. In doing so, we learned that *RUNX1* proximal poly(A) site usage, and therefore RUNX1a formation, is suppressed primarily due to splicing factor RBPs, and not core APA machinery. We identified HNRNPA1 as a potent suppressor of RUNX1a formation throughout hematopoiesis, by direct binding to alternative, terminal exon 7a. Conversely, KHDRBS1 is an enhancer of RUNX1a production in HSPCs, exhibiting declining expression throughout normal hematopoietic differentiation. Overall, our study highlights the intersection between post-transcriptional regulation and transcription factor function, while uncovering RBPs that play an essential and previously underappreciated role in normal and aberrant hematopoiesis.

3.2 Results

3.2.1 *RUNX1* poly(A) site strength minimally contributes to suppression of the RUNX1a isoform.

RNA-processing events are regulated by both core machinery and unique RBPs which confer cellular context specificity. Therefore, we first assessed the likelihood that *RUNX1* APA is regulated solely by core polyadenylation machinery. Previous studies regarding global poly(A) site usage reveal that genes tend to terminate at distal sites because proximal poly(A) sites generally contain weaker cis-acting elements than their distal counterparts and less efficiently recruit core machinery (3, 13). In these cases, usage of a weaker site is mostly dependent upon the cellular concentration of core polyadenylation machinery (31). Since *RUNX1a* isoform generation is due to polyadenylation at a proximal poly(A) site, we examined whether minimal endogenous *RUNX1a* formation can be attributed to a weak poly(A) site. To this end, we profiled endogenous *RUNX1* poly(A) site usage by performing 3'READS (95) of sorted, common myeloid progenitors (CMP) (CD34+/CD38+/CD123+/CD45RA-) from three healthy human leukapheresis products (**Figure 3.1A**). Among four previously annotated major poly(A) sites (163), we observed that poly(A) site #1, which results in *RUNX1a* formation, was the second most utilized poly(A) site behind only distal poly(A) site #4 (**Figures 3.1B and 3.1C**). Next, we compared the sequences of core polyadenylation cis elements for poly(A) sites #1 and #4 (**Figure 3.2A**). Interestingly, both poly(A) sites have the canonical AAUAAA poly(A) hexamer, predictive of a strong poly(A) site (12, 13, 21). Both also have an upstream UGUA motif which enhances core machinery recruitment (180), and downstream G/U rich elements containing UU dinucleotides that

enhance cleavage efficiency (29, 181) (**Figure 3.2B**). Taken together, *RUNX1* poly(A) site #1 does not have weak cis elements. Finally, we tested the cleavage efficiency of *RUNX1* poly(A) sites using a tandem poly(A) reporter system and RNase protection assays (182) (**Figure 3.3A**). Though poly(A) site #4 had stronger cleavage than poly(A) site #1 (**Figures 3.3B and 3.3C**), the difference was less than 2-fold, supporting the conclusion that *RUNX1a* formation is not limited primarily due to weak polyadenylation cis elements and poor cleavage efficiency. While core machinery may play some role, we hypothesize that there are context specific RBPs that regulate isoform expression of this crucial hematopoietic transcription factor.

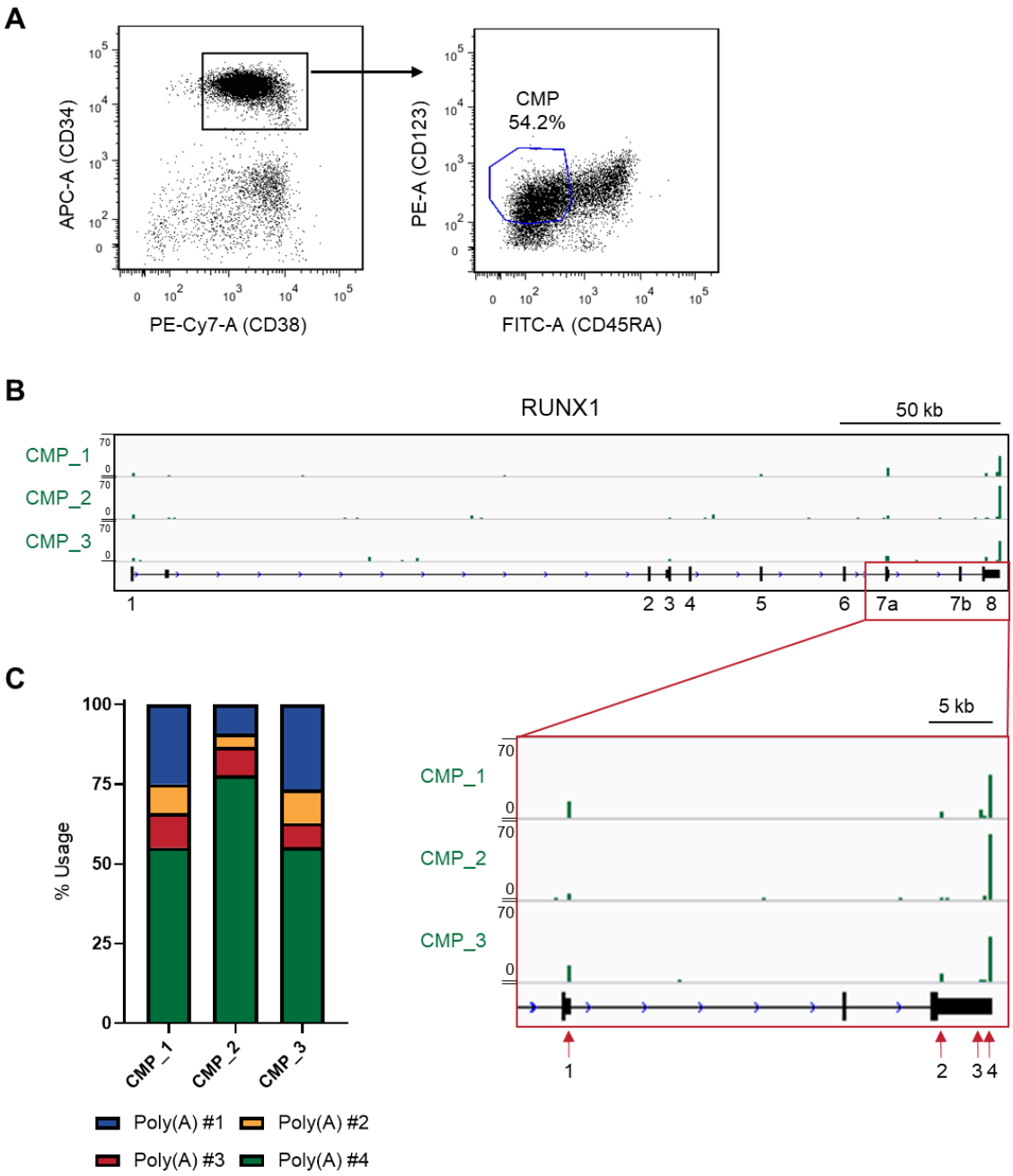


Figure 3.1. *RUNX1* PAS usage in primary CMPs via 3'RNA sequencing.

(A) Fluorescence activated cell sorting (FACS) plots depicting how common myeloid progenitors (CMPs) were identified and sorted from healthy leukapheresis products following CD34 bead enrichment. **(B)** Genome browser tracks depicting sequencing reads in the full *RUNX1* gene obtained from 3'READS of sorted CMPs. The red, boxed region is expanded. The locations of the four major poly(A) sites are marked with red arrows. **(C)** Usage of the four major *RUNX1* poly(A) sites, calculated from 3'READS analysis of sorted CMPs.

Figure 3.2. Sequence comparison of major *RUNX1* poly(A) sites.

(A) mRNA sequences containing the four major *RUNX1* poly(A) sites. The location of each site within the *RUNX1* genomic structure is marked with a red arrow (top). The three major cis-acting elements for polyadenylation are color-coded and underlined. A vertical line indicates the predicted cleavage site. The core polyadenylation complexes that bind to each cis element are depicted in the cartoon. **(B)** Table summarizing the sequence features of the four major *RUNX1* poly(A) sites. A checkmark indicates presence, whereas an 'X' represents absence of the indicated cis element.

A

Poly(A) Site 1

CCUGCAGAAAUCACUUGAUGCACAGCACAUGGGAAGCAUUGUGUGUAUUUAUUAAUCCUUCACAAAGUCUUUGAGAUA
UAUUUUUAUCAAAUAUUUAGCAUGGAUCCCGGUACACUUUCAAUACUUAAUAAAUGGUCAAUGUUAUUCUUUUUCACU
AUU|AUUGCUAUUUAUGUGCUCUUAUUGGCGGCAUUGUUGUAUCACCAGUUAGUUGUACCCAUCGGUCCUCUCCAAGU
CACACAGCUACUCUGCGAUAAUUCCCCCGGUGUCUCCUUAUU

Poly(A) Site 2

GAGGUGUCCGAGGCGACGCACCUCGAGGGUGUCCGCCGGCCCCAGCACCCAGGGGACGCGCUGGAAAGCAAACAGG
AAGAUUCCCGGAGGGAAACUGUGAAUGCUUCUGAUUUAGCAAUGCUGUGAAUAAAAAGAAAGAUUUUAUACCCUUGA|
CUUAACUUUUUAACCAAGUUGUUUAUUCCAAAGAGUGUGGAAUUUUGGUUGGGGUGGGGGGAGAGGAGGGAUGCAAC
UCGCCCUGUUUGGCAUCUAAUUCUUAUUUUUAAUUUUUCCGCACCUUAUCAAUUGCAAAAUGCGUAUUUGCAUUUGGG
UGG

Poly(A) Site 3

ACAGCUAAUAGCAUGGUUCCAAUUUUUUUUAAGUUCACUUUUUUUGUUCUAGGGGAAAUGAAUGUGCAAAAAAAGAAA
AAGAACUGUUGGUUAUUUGUGUUAUUCUGGAUGUAUAAAAAUCAAUGGAAAAAAAUAAACUUUCAAAUU|GAAAUGACG
GUAUAACACAUCUACUGAAAAAGCAACGGGAAAUGUGGUCCUAUUUAAGCCAGCCCCCACCUAGGGUCUAUUUGUGUG
GCAGUUAUUGGGUUUGGUCACAAAACAUCCUGAAAAUUCGUGCGUGGGCUUC

Poly(A) Site 4

CUCUCCUGGGAGCAUUCGUCGUGCCCAGCCUGAGCAGGGCAGCUGGACUGCUGCUGUUCAGGAGCCACCAGAGCCU
UCCUCUCUUUGUACCACAGUUUCUUCUGUAAAUCCAGUGUUACAAUCAGUGUGAAUGGCAAAUAAACAGUUUGACAA
GUACAUACACCAUA|UCCAAUUGGUUUUGUCUUUCUCUGAGAUGGAGGGUUGGGAAACUGAGUAAGGCCCACUGCACA
CACUAGGAAGAUACCUCCAGUGAACAGCUUCAGGCUGAAGCUAUGUCCUCACCCCCGGUGCUCCUAGCACACUCUGG
CUUCUGAAGGC

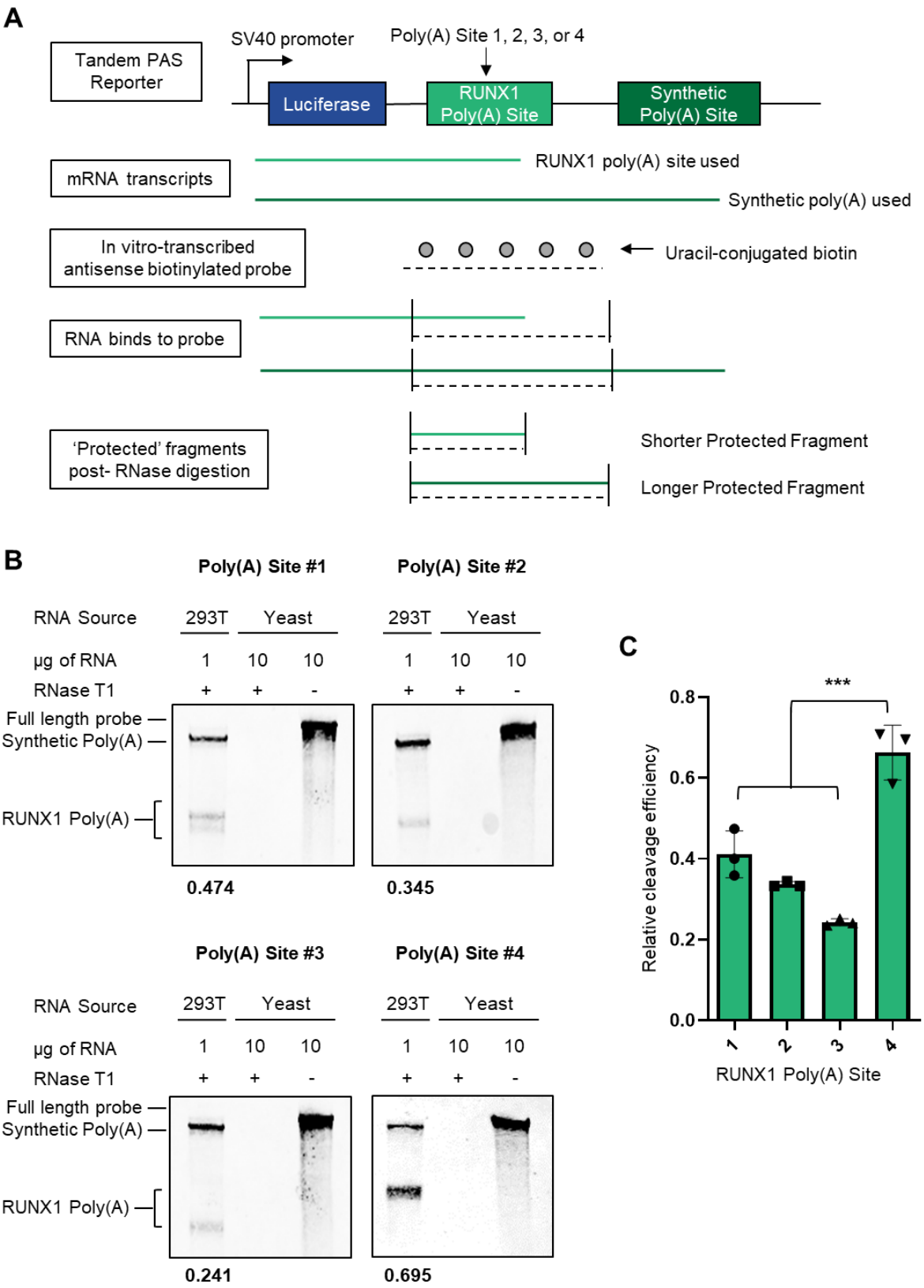
Poly(A) Signal (PAS) Hexamer Downstream G/U-rich elements Upstream UGUA Motif

B

RUNX1 Poly(A) Site ID	Canonical PAS Hexamer (AAUAAA)	Downstream U and G/U-rich Elements	Upstream UGUA motif
1	✓	✓	✓
2	✓	✓	X
3	✓	X	✓
4	✓	✓	✓✓

Figure 3.3. *RUNX1* poly(A) sites differ in cleavage efficiency.

(A) Schematic of the RNase protection assay (RPA). The *RUNX1* poly(A) site sequences cloned between luciferase and the SV40 poly(A) site are shown in Figure 3.2A. **(B)** Representative blots showing the protected fragments generated from all four *RUNX1* poly(A) sites relative to a common synthetic poly(A) site using RPA. For each blot, lane 1 is the experimental lane. Lane 2 is a positive control for effective RNase digestion and a negative control for non-specific binding of the probe. Lane 3 is a control of full-length probe integrity. Yeast RNA is mixed with probe for control lanes (2-3). *RUNX1* poly(A) site and synthetic poly(A) site signal intensities were normalized to each other based on the number of uracil nucleotides present in each protected fragment. Example quantifications are listed below the experimental lanes. **(C)** Quantification of *RUNX1* PAS usage relative to the common synthetic PAS. Data are mean +/- s.d. of three independent experiments. *** $p < 0.001$, one-way ANOVA with a post-hoc Tukey test.


A
Tandem PAS Reporter
SV40 promoter
Poly(A) Site 1, 2, 3, or 4
Luciferase
RUNX1 Poly(A) Site
Synthetic Poly(A) Site
mRNA transcripts
RUNX1 poly(A) site used
Synthetic poly(A) used
In vitro-transcribed antisense biotinylated probe
Uracil-conjugated biotin
RNA binds to probe
'Protected' fragments post- RNase digestion
Shorter Protected Fragment
Longer Protected Fragment
B
Poly(A) Site #1
Poly(A) Site #2
RNA Source
293T
Yeast
µg of RNA
RNase T1
Full length probe
Synthetic Poly(A)
RUNX1 Poly(A)
0.474
0.345
Poly(A) Site #3
Poly(A) Site #4
0.241
0.695
C
***
Relative cleavage efficiency
RUNX1 Poly(A) Site

3.2.2 A split GFP minigene model recapitulates *RUNX1* isoform generation.

Next, we modeled *RUNX1a* formation by including more of its genomic context, capturing regions that bind accessory RBPs that mediate context specific isoform generation. *RUNX1* APA is a unique type of coding sequence APA where proximal poly(A) site usage is coupled with splicing of an alternative, terminal exon (40). For *RUNX1a* formation, exon 6 of the *RUNX1* gene splices exon 7a and terminates at poly(A) site #1 (**Figure 3.1B**). *RUNX1b/c* are formed by skipping exon 7a, splicing exon 7b, and terminating at one of three poly(A) sites in exon 8 (163). Therefore, we devised a minigene model that accounts for both splicing and polyadenylation of *RUNX1* alternative, terminal exon 7a. We cloned exon 7a with ~500bp of flanking intron between the two exons of a split GFP reporter (183, 184) (**Figure 3.4A - left**). Because *RUNX1a* is globally suppressed in hematopoiesis, we expect that exon 7a will be skipped and GFP will be detected in cells expressing the transgene. We also generated an analogous construct containing constitutive exon 7b, which is spliced instead of exon 7a to produce the major *RUNX1b/c* isoforms (**Figure 3.4A - right**). Unlike the exon 7a minigene construct, we expect that exon 7b will be spliced between the two GFP exons, and GFP will not be produced. Indeed, the exon 7a minigene produced a robust GFP signal in cells expressing the transgene, whereas the exon 7b construct produced no detectable GFP (**Figure 3.4B**). We confirmed that GFP fluorescence reflected the expected splice products by performing RT-PCR (**Figure 3.4C - top**). Additionally, we observed low levels of exon 7a spliced and polyadenylated product from cells expressing the exon 7a minigene, effectively modeling low endogenous *RUNX1a* isoform generation (**Figure 3.4C - bottom**). The opposite processing of these

minigenes recapitulates endogenous *RUNX1* post-transcriptional regulation. Consequently, these constructs are suitable for further study of cis-acting elements and RBPs which differentiate these RNA-processing events.

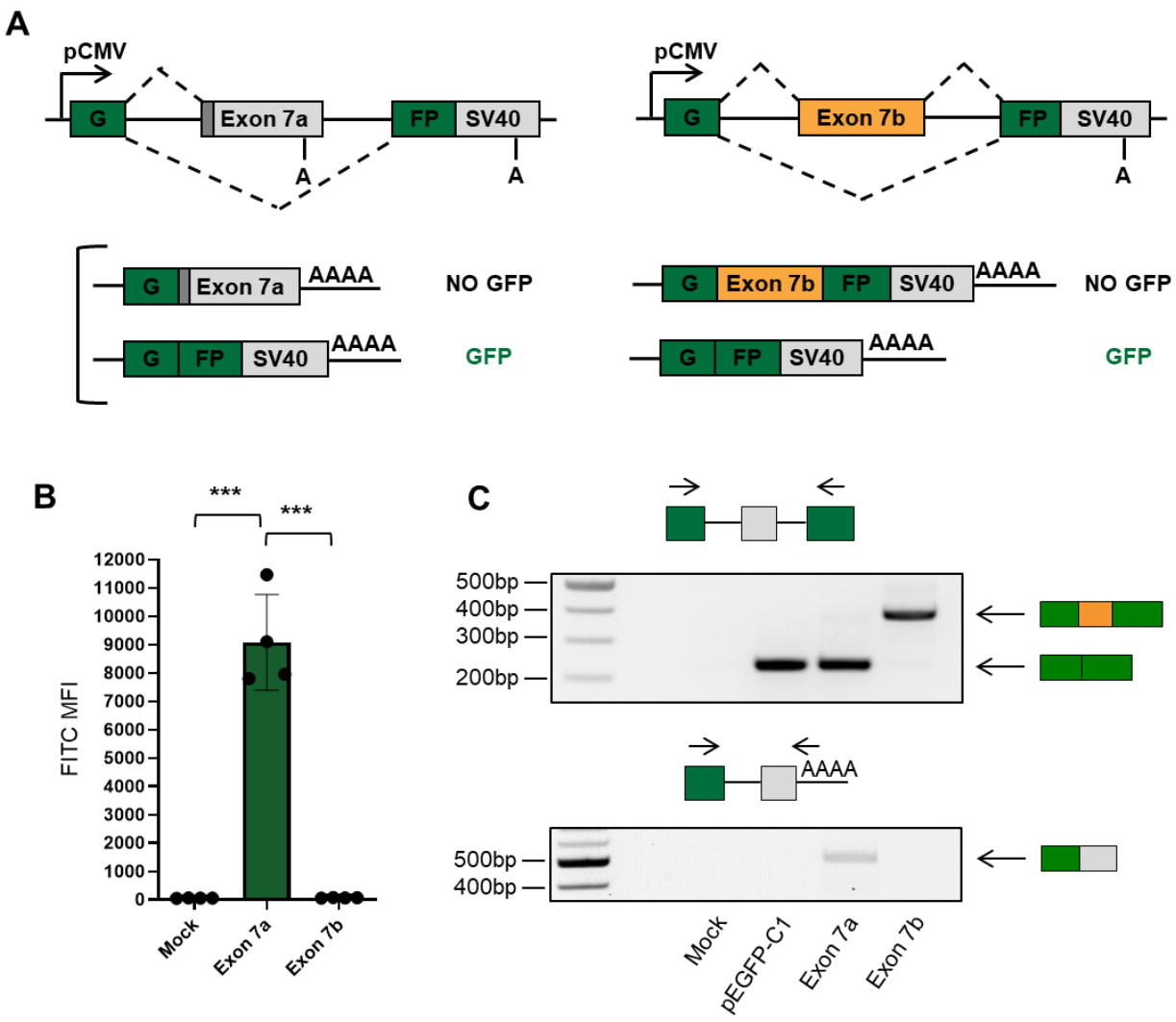


Figure 3.4. A split GFP minigene model recapitulates *RUNX1* post-transcriptional RNA processing.

(A) Schematic of the split GFP minigene constructs containing *RUNX1* exon 7a (top – left) and *RUNX1* exon 7b (top – right). Poly(A) sites are marked with an 'A'. **(B)** FITC mean fluorescent intensity (MFI) measured by flow cytometric analysis of KG-1a cells nucleofected with the *RUNX1* exon 7a or exon 7b minigene construct. FITC MFI is normalized to non-nucleofected KG-1a cells. Data are mean +/- s.d. of four independent experiments. *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test. **(C)** RT-PCR analysis of RNA extracted from KG-1a cells nucleofected with the *RUNX1* minigene constructs, a GFP vector positive control, or mock-nucleofected negative control. Primer sets used for analysis are shown above the corresponding agarose gel image.

3.2.3 Chimeric minigene constructs reveal locations of critical cis-acting elements regulating *RUNX1* exon 7a and 7b inclusion.

To assess whether the disparity in *RUNX1* exon usage is generally defined by exons or flanking intronic regions, we subcloned two series of chimeric minigene constructs. We first modified the split GFP reporter that contained exon 7a by replacing the upstream (bAa), downstream (aAb), or both (bAb) flanking introns with the analogous introns that typically surround exon 7b (**Figure 3.5A**). Cells nucleofected with all four exon 7a minigene constructs still produced GFP, as shown by flow cytometry (**Figure 3.5B**) and RT-PCR (**Figure 3.5C - top**). This observation suggests that major suppressive cis-acting elements are located within exon 7a. Interestingly, we noticed consistent differences in the GFP MFI of cells expressing each of the four constructs. To test whether these differences meaningfully predict changes in inclusion of exon 7a, we performed RT-qPCR to quantify the relative amount of polyadenylated exon 7a product generated by each minigene construct (**Figure 3.5D**). Indeed, decreased GFP MFI (bAa) correlated with increased exon 7a inclusion, whereas increased GFP MFI (aAb and bAb) correlated with exon 7a exclusion. The former observation (bAa) indicates a minor suppressive cis element in the upstream intron of exon 7a. The latter (aAb and bAb) results from ablation of exon 7a polyadenylation by removal of the G/U-rich downstream element (**Figure 3.5C - bottom**). Overall, GFP MFI precisely detected changes in exon 7a inclusion of the minigene reporter.

We next generated a set of minigene constructs that contained constitutive exon 7b. We replaced the upstream (aBb), downstream (bBa), or both (aBa) flanking introns with the analogous introns that typically flank exon 7a (**Figure 3.6A**). When

nucleofected into cells, the wild type construct did not produce GFP and replacing only the downstream intron (bBa) had no effect on GFP production. However, replacing the upstream intron (aBb) produced a slight increase in GFP+ cells and replacing both introns (aBa) produced a dramatic increase in GFP+ cells (**Figure 3.6B**). RT-PCR analysis revealed a striking reversal from complete inclusion of exon 7b in the wild type context to nearly complete exclusion when both introns were replaced (**Figure 3.6C**). Therefore, the constitutive nature of exon 7b is dictated by its intronic context, unlike exon 7a which was modestly affected by alterations to its flanking introns.

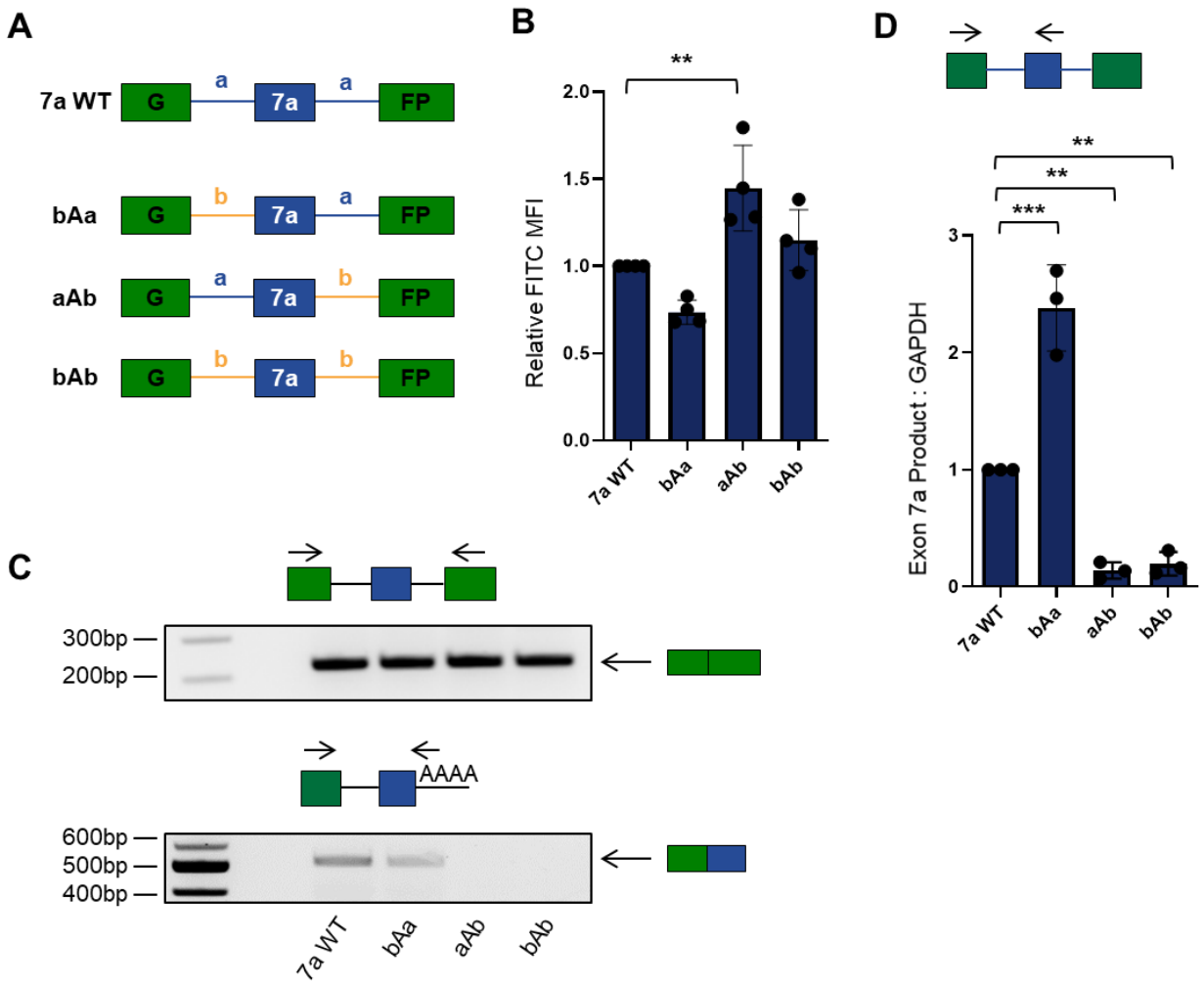


Figure 3.5. Suppressive elements within *RUNX1* exon 7a mediate inclusion.

(A) Schematic of the *RUNX1* exon 7a chimeric split GFP minigene constructs. Changes to the intronic context are indicated by label and color: introns normally flanking exon 7a are blue and labeled 'a'; introns normally flanking exon 7b are yellow and labeled 'b'. **(B)** FITC mean fluorescent intensity (MFI) measured by flow cytometric analysis of KG-1a cells nucleofected with the indicated *RUNX1* exon 7a chimeric minigene constructs. FITC MFI is normalized to KG-1a cells nucleofected with the exon 7a WT minigene construct. Data are mean +/- s.d. of four independent experiments. ** $p < 0.01$, one-way ANOVA with post-hoc Tukey test. **(C)** RT-PCR analysis of RNA extracted from KG-1a cells nucleofected with the *RUNX1* exon 7a chimeric minigene constructs. Primer sets used for analysis are shown above the corresponding agarose gel image. **(D)** RT-qPCR analysis of *exon7a product* mRNA normalized to *GAPDH* mRNA from KG-1a cells nucleofected with the respective constructs. mRNA levels were normalized to that of the *RUNX1* exon 7a WT control. Data are mean +/- s.d. of three independent experiments. ** $p < 0.01$, *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test.

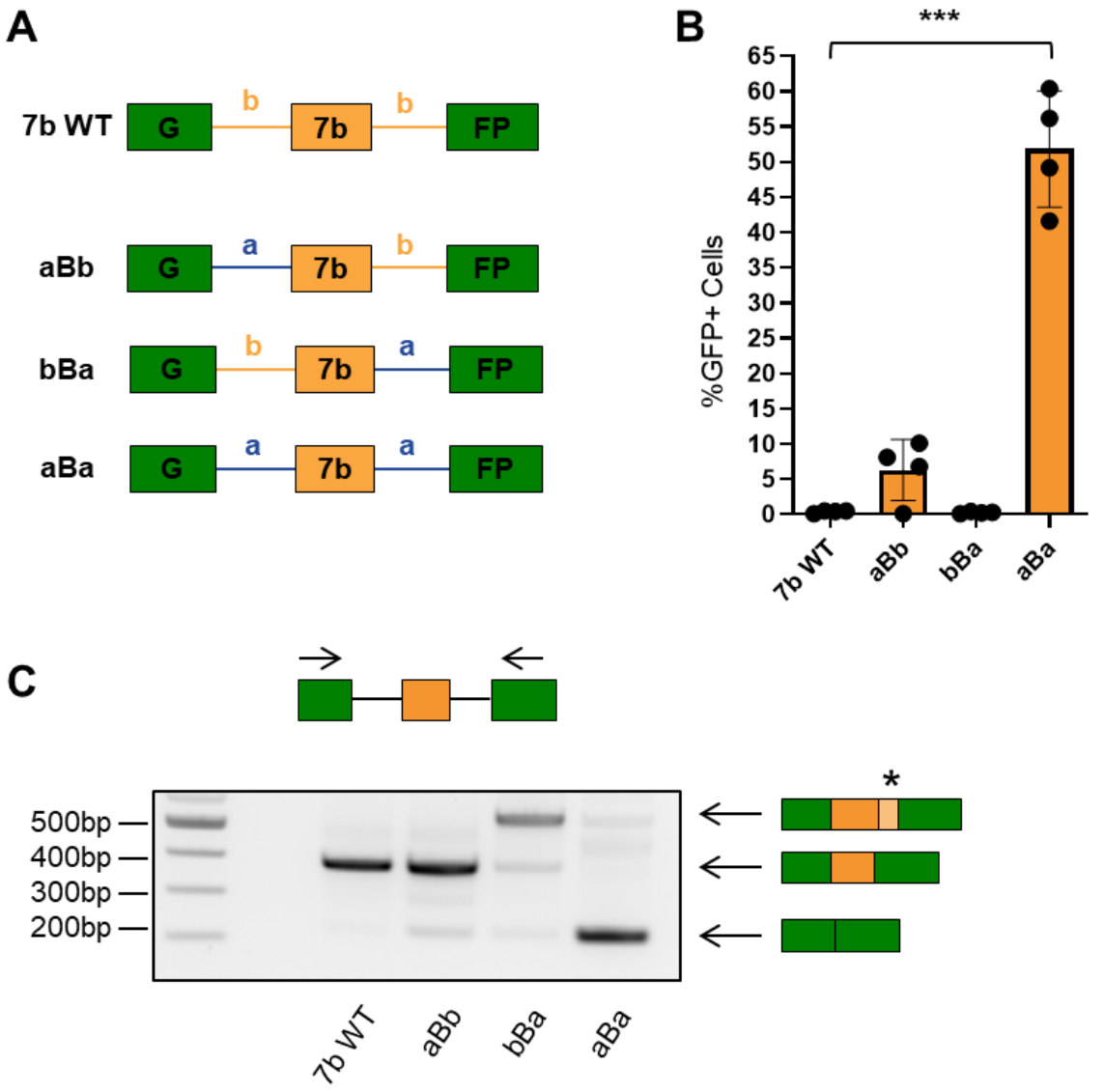


Figure 3.6. Flanking intronic regions direct *RUNX1* exon 7b inclusion.

(A) Schematic of the *RUNX1* exon 7b chimeric split GFP minigene constructs. Changes to the intronic context are indicated by label and color: introns normally flanking exon 7a are blue and labeled ‘a’; introns normally flanking exon 7b are yellow and labeled ‘b’. **(B)** Percentage of GFP+ cells as measured by flow cytometry of KG-1a cells nucleofected with the indicated *RUNX1* exon 7b chimeric minigene constructs. Data are mean +/- s.d. of four independent experiments. *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test. **(C)** RT-PCR analysis of RNA extracted from KG-1a cells nucleofected with the indicated *RUNX1* exon 7b chimeric minigene constructs. The asterisk (*) represents the presence of a cryptic splice that results in a higher molecular weight product.

3.2.4 A CRISPR RNA-binding protein (RBP) screen uncovers putative regulators of *RUNX1a* production.

Motivated by the unique regulation of exon 7a, we next wanted to identify specific RBPs that regulate exon 7a inclusion, and therefore *RUNX1a* formation. We performed a CRISPR/Cas9 screen using a library composed of single guide-RNAs (sgRNAs) exclusively targeting RBPs (185). Since the exon 7a split GFP minigene reporter was sensitive enough to detect bi-directional changes in exon inclusion (**Figure 3.5B**), we modified this reporter for screening. To account for RBPs that may affect transcription, mRNA stability, export, localization, and translation of the reporter independent of changes in inclusion of exon 7a, we added mCherry and a P2A peptide directly upstream of the split GFP (**Figure 3.8A**). We selected MDS-L cells as a diploid, *RUNX1a* intermediate-expressing cell line (**Figures 3.7A and 3.7B**) and generated a stable, clonal line expressing the bicistronic, dual fluorescent minigene reporter. As expected, these cells displayed a tight correlation of mCherry and GFP fluorescence (**Figure 3.8B**).

We infected this reporter cell line with the lentiviral sgRNA library and selected for infected cells with puromycin (**Figure 3.8C**). We collected a population of cells three days post-transduction to confirm representation of the sgRNA library (day 0) (**Figure 3.9A**) and after three weeks (day 21) to ensure complete knockout. Importantly day 0 and day 21 replicates were significantly correlated with each other (**Figure 3.9B**). At the day 21 endpoint, positive control sgRNAs and those targeting essential RBPs were significantly depleted from the pool (**Figure 3.9C**) and there was a clear difference in overall sgRNA representation (**Figure 3.9D**). We also sorted GFP low and GFP high

cells relative to mCherry (**Figure 3.8B**) to identify sgRNAs targeting putative RUNX1a suppressors and enhancers, respectively. We used MaGeCK (186) to calculate beta scores describing the level of enrichment of sgRNAs in these subpopulations. Altogether, we identified 47 putative suppressor RBPs and 55 putative enhancer RBPs from the 1078 RBPs screened (**Figure 3.10A**). Among the ten most significant putative RBP suppressors, eight carry the “RNA splicing” GO annotation (GO:0008380) and none are annotated for “mRNA 3’-end processing” (GO:0031124) (**Table 3.1**). This observation suggests potent splicing regulation of alternative, terminal exon 7a and supports our previous conclusion that APA core machinery plays a minor role in repressing proximal poly(A) site usage (**Figures 3.1-3.3**).

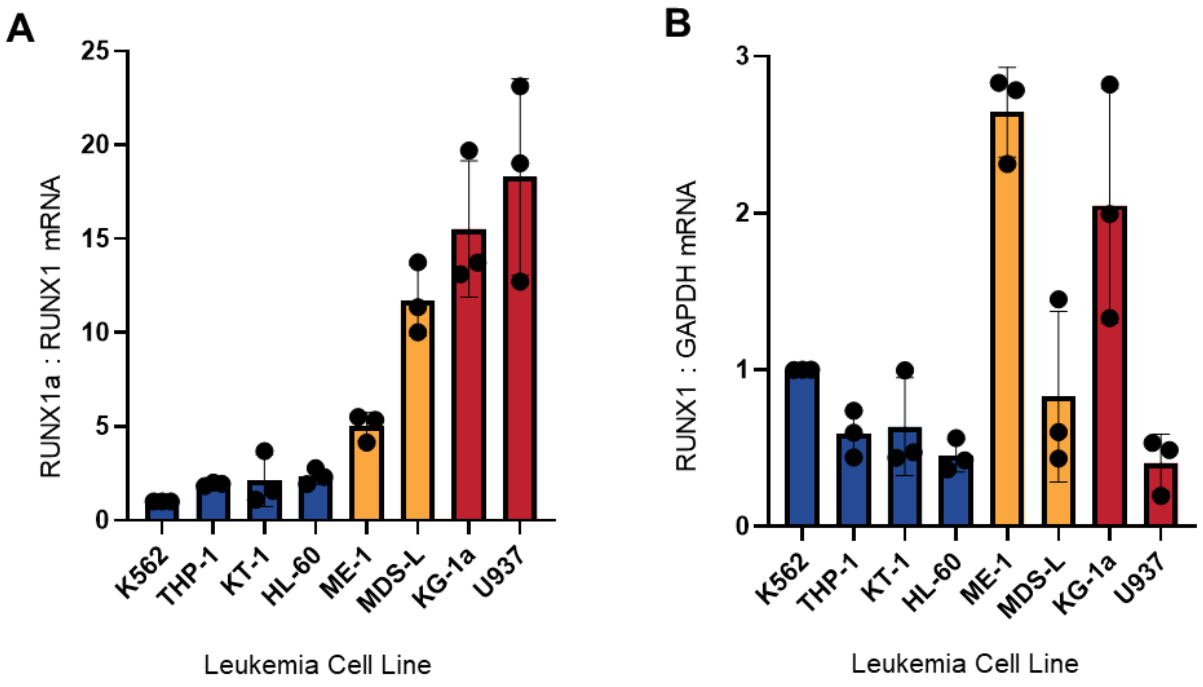


Figure 3.7. Relative *RUNX1a* and total *RUNX1* mRNA in various leukemia cell lines.

(A) RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in the indicated leukemia cell lines. mRNA levels were normalized to K562 cells, which had the lowest relative *RUNX1a* transcript. Blue bars represent '*RUNX1a* low,' yellow bars represent '*RUNX1a* intermediate,' and red bars represent '*RUNX1a* high' cell lines. Data are mean +/- s.d. of three independent batches of RNA. **(B)** RT-qPCR analysis of *RUNX1* mRNA normalized to total *GAPDH* mRNA in the indicated leukemia cell lines. mRNA levels were normalized to K562 cells. Data are mean +/- s.d. of three independent batches of RNA.

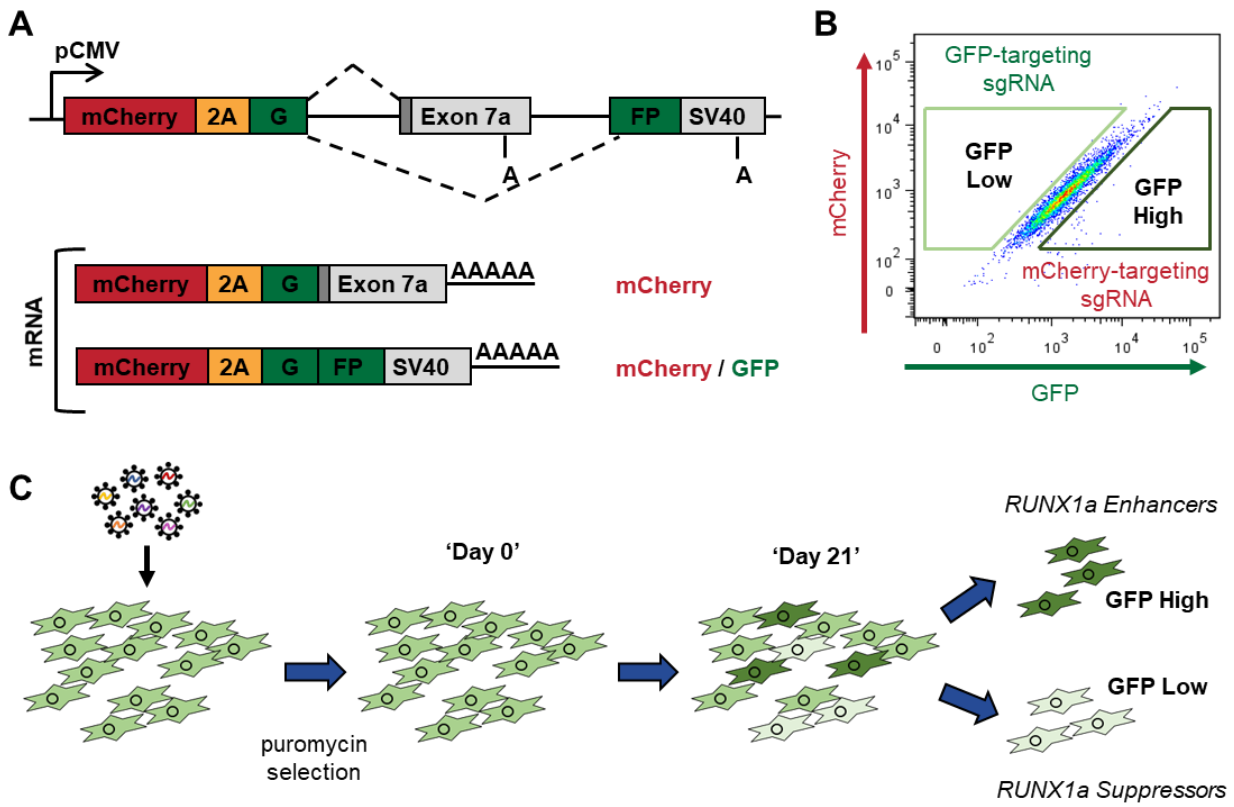
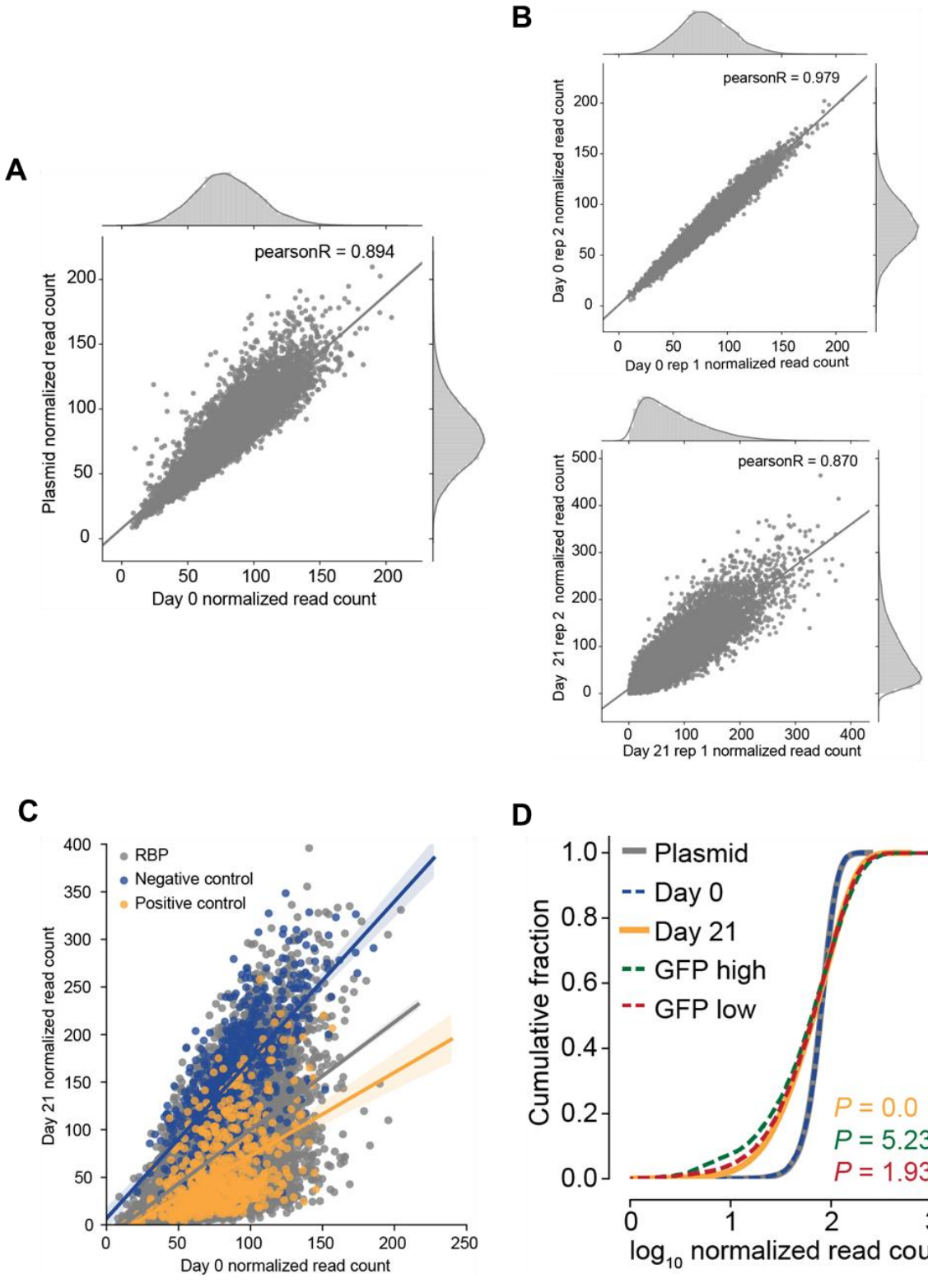


Figure 3.8. A CRISPR RNA-binding protein screen reveals post-transcriptional regulators of *RUNX1*.

(A) Schematic of the bicistronic, dual fluorescent minigene reporter designed for the CRISPR screen. The *RUNX1* exon 7a minigene reporter was modified by subcloning mCherry and a P2A peptide upstream of the first GFP exon. Poly(A) sites are marked with an 'A'. **(B)** Representative flow cytometry plot of the stable, clonal MDS-L cell line containing the dual fluorescent reporter from (A). Boxed regions represent the gating of GFP high and low cell populations that were sorted at the endpoint of the CRISPR screen. The positive control for each sorted population is noted. **(C)** Schematic of the CRISPR screen. MDS-L cells stably expressing the dual fluorescent *RUNX1* exon 7a minigene construct were infected with lentiviral particles containing the RBP sgRNA library. Bulk cell populations were collected at 'Day 0' and 'Day 21.' GFP high and low cells were sorted as depicted in (B) on 'Day 21.'

Figure 3.9. CRISPR screen quality control metrics.

(A) Dot plot showing the correlation between normalized sgRNA read counts in the plasmid library used for lentiviral production and counts in transduced cells on ‘day 0.’ **(B)** Dot plots showing the correlation between normalized sgRNA read counts in ‘day 0’ replicates (top) and ‘day 21’ replicates (bottom). **(C)** Dot plot showing the change in normalized sgRNA read counts from ‘day 0’ to ‘day 21’ of the screen. Negative control sgRNAs (blue) exert no selection pressure. Positive control sgRNAs (yellow) are expected to drop out over time. sgRNAs targeting RBPs are in gray; essential RBPs are expected to drop out over time. **(D)** Cumulative distribution frequency (CDF) plot depicting sgRNA representation in the indicated cell populations and the plasmid library used for lentiviral production. All populations were compared to ‘Day 0.’ Kolmogorov-Smirnov test.


A
pearsonR = 0.894
Plasmid normalized read count
Day 0 normalized read count
B
pearsonR = 0.979
Day 0 rep 2 normalized read count
Day 0 rep 1 normalized read count
pearsonR = 0.870
Day 21 rep 2 normalized read count
Day 21 rep 1 normalized read count
C
RBP
Negative control
Positive control
Day 21 normalized read count
Day 0 normalized read count
D
Plasmid
Day 0
Day 21
GFP high
GFP low
Cumulative fraction
P = 0.0
P = 5.23e-22
P = 1.93e-4
log10 normalized read count

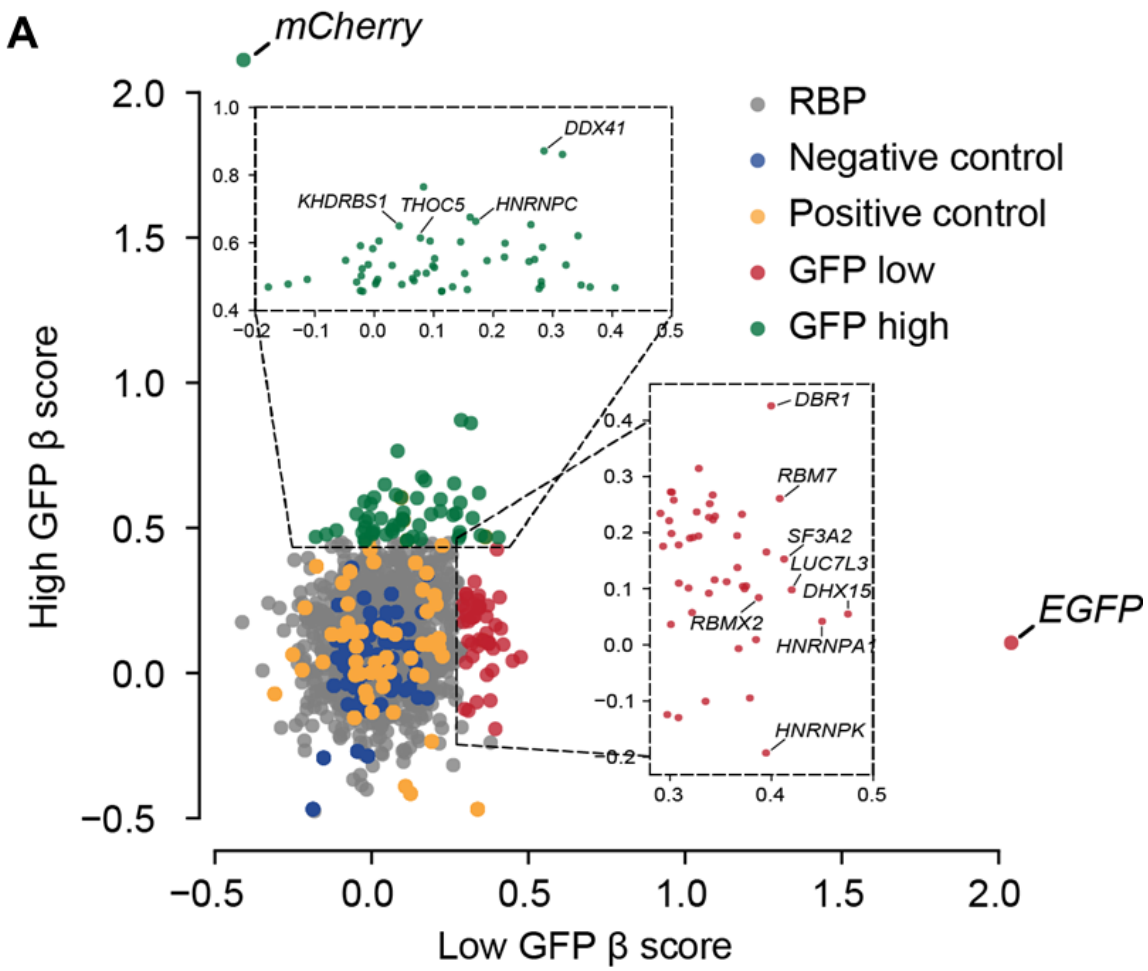


Figure 3.10. RNA-binding proteins identified in the CRISPR screen.

(A) Plot depicting the β-scores of sgRNA enrichment in GFP high versus low populations. The insets show all significantly enriched RBPs for both cell populations ($p < 0.05$). 'RNA splicing' (GO: 0008380) RBPs among the ten most significant RBPs are labeled. Positive controls (mCherry and EGFP) are also labeled.

Table 3.1. Composition of significantly enriched RNA-binding proteins in GFP low and high populations.

The most significantly enriched ($p < 0.01$) RBPs in the GFP low (left) and GFP high (right) populations are listed. EGFP and mCherry targeting sgRNAs, the positive control for each sorted population, were most significantly enriched. 'RNA splicing' (GO: 0008380) and 'mRNA 3'end processing' (GO:0031124) annotations are indicated for each RBP.

Putative RUNX1a Suppressors (GFP Low)				Putative RUNX1a Enhancers (GFP High)			
RBP	P-value	RNA splicing GO:0008380	mRNA 3'-end processing GO:0031124	RBP	P-value	RNA splicing GO:0008380	mRNA 3'-end processing GO:0031124
EGFP	0	N/A	N/A	mCherry	0	N/A	N/A
DHX15	0.00163	Yes	No	DDX41	0	Yes	No
HNRNPA1	0.002445	Yes	No	U2SURP	0	No	No
LUC7L3	0.002445	Yes	No	NARS	0	No	No
SF3A2	0.002445	Yes	No	GARS	0.000815	No	No
RBM7	0.002445	Yes	No	HNRNPC	0.000815	Yes	No
ADAT2	0.00326	No	No	MTPAP	0.000815	No	Yes
DBR1	0.004075	Yes	No	CARS	0.000815	No	No
RRP15	0.00489	No	No	RNF113A	0.00163	No	No
HNRNPK	0.00489	Yes	No	THOC5	0.00163	Yes	No
RBMX2	0.00489	Yes	No	KHDRBS1	0.002445	Yes	No
LSM12	0.00489	No	No	DYNC1H1	0.002445	No	No
RPL7	0.005705	No	No	EIF2S3	0.00326	No	No
DDX21	0.005705	No	No	MRPL14	0.00489	No	No
TIAL1	0.005705	No	No	IPO7	0.00652	No	No
RPL30	0.00652	No	No	RPS15	0.007335	No	No
MARS2	0.00652	No	No	DDX49	0.00815	No	No
RPL26	0.00652	No	No				
XRN2	0.00815	No	No				
RPL19	0.00815	No	No				
RPL23	0.00815	No	No				
PSMD1	0.00978	No	No				

3.2.5 HNRNPA1 is a potent suppressor of RUNX1a formation.

To validate putative RUNX1a suppressors, we performed a secondary siRNA screen. We knocked down the eight "RNA splicing" RBPs and measured endogenous *RUNX1a* mRNA relative to total *RUNX1* transcript by RT-qPCR (**Figure 3.11A**). From this secondary screen, we confidently identified HNRNPA1 as a suppressor of *RUNX1a* isoform generation. To better measure the extent of this suppression, we knocked down HNRNPA1 using stably expressed shRNAs (**Figures 3.12A and 3.12B**). In both K562 and MDS-L leukemia cell lines, HNRNPA1 knockdown led to a robust increase in relative *RUNX1a* to total *RUNX1* mRNA (**Figures 3.12C**) with a concurrent decrease in *RUNX1b/c* transcript (**Figures 3.12D**). In MDS-L cells, HNRNPA1 knockdown also led to the clear detection of RUNX1a protein which was difficult to detect in the parental cell line (**Figures 3.12E**).

Because HNRNPA1 is abundantly expressed in hematopoietic cells, we also generated HNRNPA1 knockout MDS-L cells via CRISPR/Cas9. We validated four unique HNRNPA1 knockout clones and two heterozygous clones by western blot (**Figures 3.13A and 3.13B**) and sequencing of individual alleles (**Figures 3.13C**). HNRNPA1 protein reduction conferred a dose-dependent increase in relative *RUNX1a* mRNA (**Figures 3.14A**) and RUNX1a protein (**Figures 3.14B**), further implicating HNRNPA1 as a potent suppressor of RUNX1a isoform formation.

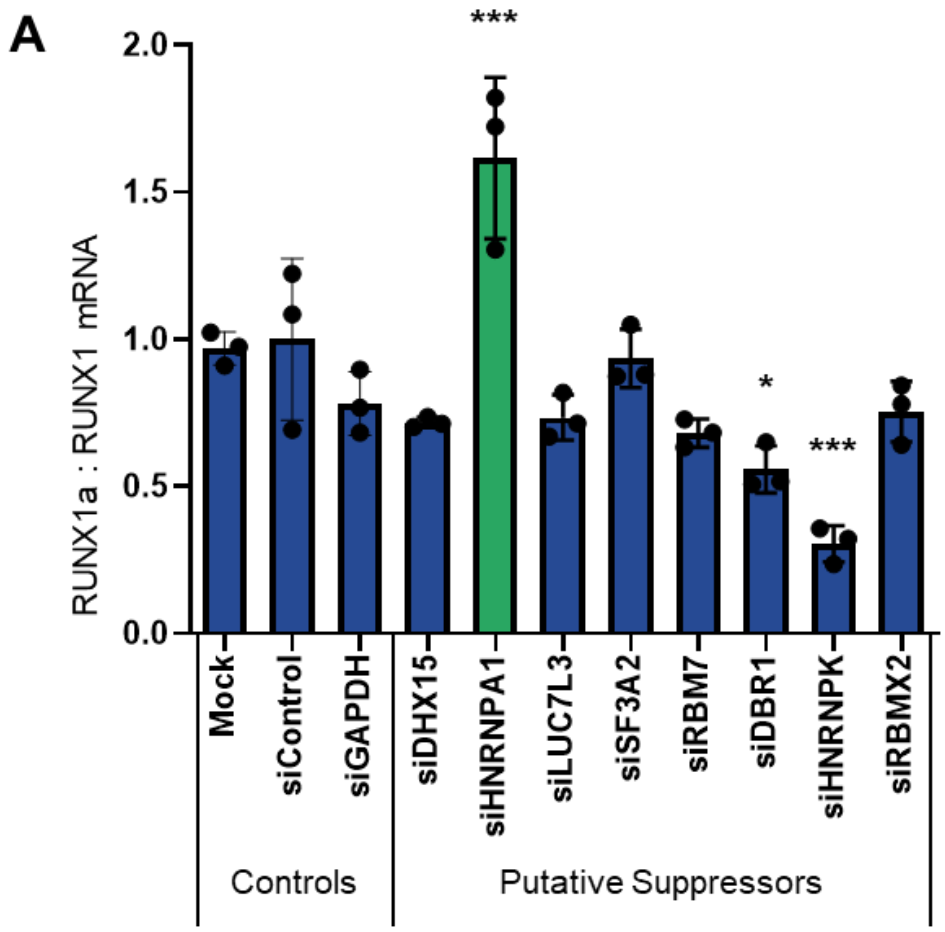
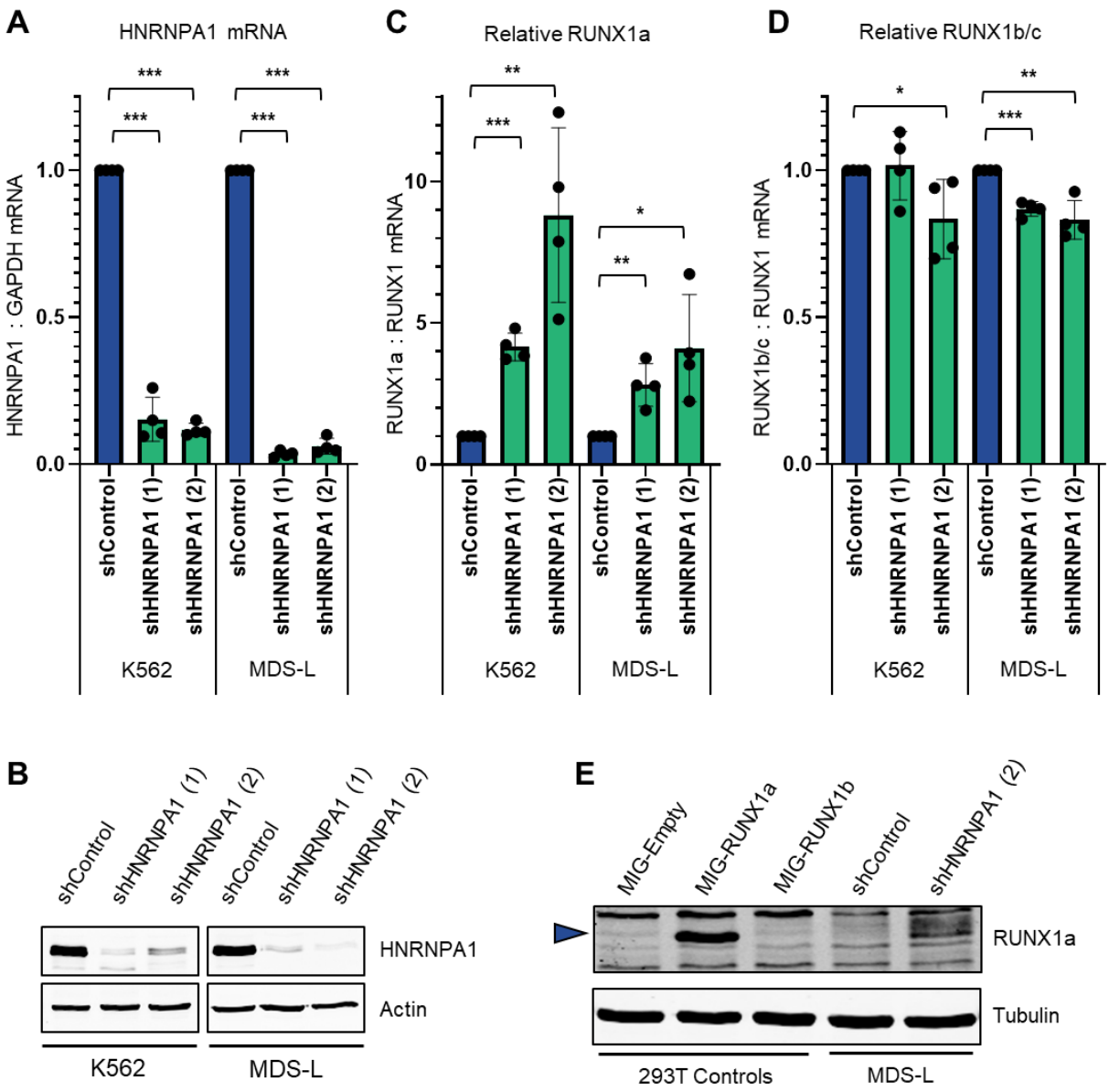


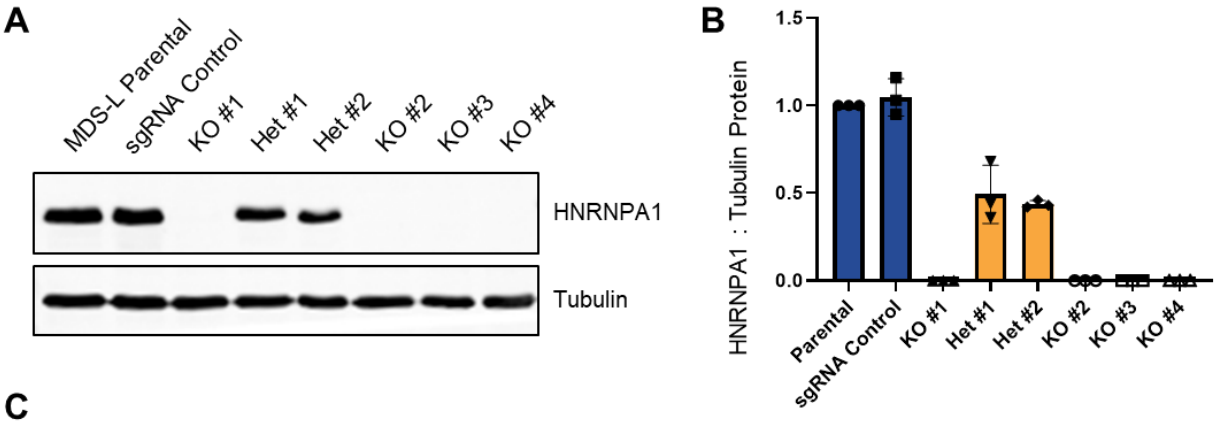
Figure 3.11. A secondary siRNA screen reveals suppressors of *RUNX1a*.

(A) RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in MDS-L cells nucleofected with the indicated siRNAs. mRNA levels were normalized to siControl nucleofected cells. The teal bar marks the suppressor RBP that had significantly increased *RUNX1a* mRNA upon knockdown. Data are mean +/- s.d. of three biological replicates. * $p < 0.05$, *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test.

Figure 3.12. HNRNPA1 knockdown enhances RUNX1a isoform production.

(A) RT-qPCR analysis of *HNRNPA1* mRNA normalized to *GAPDH* mRNA upon knockdown in K562 and MDS-L cells. mRNA levels were normalized to the respective control for each cell line. Data are mean +/- s.d. of four independent experiments. *** $p < 0.001$, two-tailed student's t-tests. **(B)** Western blot showing HNRNPA1 and tubulin (loading control) protein in K562 and MDS-L cells six days following transduction with control shRNAs or shRNAs targeting *HNRNPA1*. **(C)** RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* mRNA upon HNRNPA1 knockdown in K562 and MDS-L cells. mRNA levels were normalized to the respective control for each cell line. Data are mean +/- s.d. of four independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, two-tailed student's t-tests. **(D)** RT-qPCR analysis of *RUNX1b/c* mRNA normalized to total *RUNX1* mRNA upon HNRNPA1 knockdown in K562 and MDS-L cells. mRNA levels were normalized to the respective control for each cell line. Data are mean +/- s.d. of four independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, two-tailed student's t-tests. **(E)** Western blot showing RUNX1a and tubulin (loading control) protein. The first three lanes contain protein from 293T cells transfected with empty vector, vector containing untagged RUNX1a cDNA, or vector containing untagged RUNX1b/c cDNA. Lane 2 is a positive control for RUNX1a protein (marked with the blue arrow). The last two lanes contain protein from MDS-L cells six days following transduction with shControl lentivirus or shHNRNPA1 (2) lentivirus.


A
HNRNPA1 mRNA
HNRNPA1 : GAPDH mRNA
1.0
0.5
0.0
***
***
***
***
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
K562
MDS-L
C
Relative RUNX1a
RUNX1a : RUNX1 mRNA
10
5
0
**
***
*
**
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
K562
MDS-L
D
Relative RUNX1b/c
RUNX1b/c : RUNX1 mRNA
1.0
0.5
0.0
*
**
***
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
K562
MDS-L
B
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
shControl
shHNRNPA1 (1)
shHNRNPA1 (2)
HNRNPA1
Actin
K562
MDS-L
E
MIG-Empty
MIG-RUNX1a
MIG-RUNX1b
shControl
shHNRNPA1 (2)
RUNX1a
Tubulin
293T Controls
MDS-L



C

Clone ID	Individual Allele Sequences	Mutation Analysis	Protein Level
sgRNA Control	...TGCCGTCATGTCTAAGTCAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGTCTAAGTCAGAGGTGAGTTAGGCGCG...	N/A	100%
KO #1	...TGCCGTCATGTCTAAGTGAGTTAGGCGCGTCAGAGGTGAG... ...TGCCGTCATGTCTAAGTTCAGAGGTGAGTTAGGCGCG...	13 bp insertion 1 bp insertion	Undetectable
Het #1	...TGCCGTCATGTCTA - - - CAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGT - - - - - - - - - - - - - GAGTTAGGCGCG...	3 bp deletion 13 bp deletion	~50%
Het #2	...TGCCGTCATGTC - - - - - - - - - GGTGAGTTAGGCGCG... ...- GGTGAGTTAGGCGCG...	9 bp deletion 344 bp deletion	~50%
KO #2	...TGCCGTCATG - - - - - - - CAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGT - - - - - - - - - - - - - GAGTTAGGCGCG...	7 bp deletion 13 bp deletion	Undetectable
KO #3	...TGCCGTCATGTCTAAG - - - - - - - - - - - - - - - - - CG... ...TGCCGTCATGTCTAAGT - AGAGGTGAGTTAGGCGCG...	18 bp deletion 1 bp deletion	Undetectable
KO #4	...TGCCGTCATGTCTAAGT - - - - - - - - - - - TAGGCGCG... ...TGC - - - - - - - - - - - - - - - AGAGGTGAGTTAGGCGCG...	11 bp deletion 15 bp deletion	Undetectable

Figure 3.13. Generation of HNRNPA1 knockout MDS-L cells.

(A) Western blot showing HNRNPA1 and tubulin (loading control) protein in the parental MDS-L cell line, a clonal MDS-L cell line containing empty lenti-v2 Cas9 vector, and individual HNRNPA1 sgRNA-targeted MDS-L clonal cell lines. **(B)** Protein from (A) was quantified by normalizing HNRNPA1 signal intensity to tubulin signal intensity using LI-COR Image Studio software. Clonal cell line HNRNPA1 protein levels were normalized to the parental MDS-L cell line. Data are mean +/- s.d. of three protein lysates from each clonal line. **(C)** Table showing the sequence of each HNRNPA1 allele in the clonal MDS-L cell lines. Mutations are shown in red. Blue nucleotides are in exon 1. Gray nucleotides are in the intron downstream of exon 1. KO, knockout; Het, heterozygous.

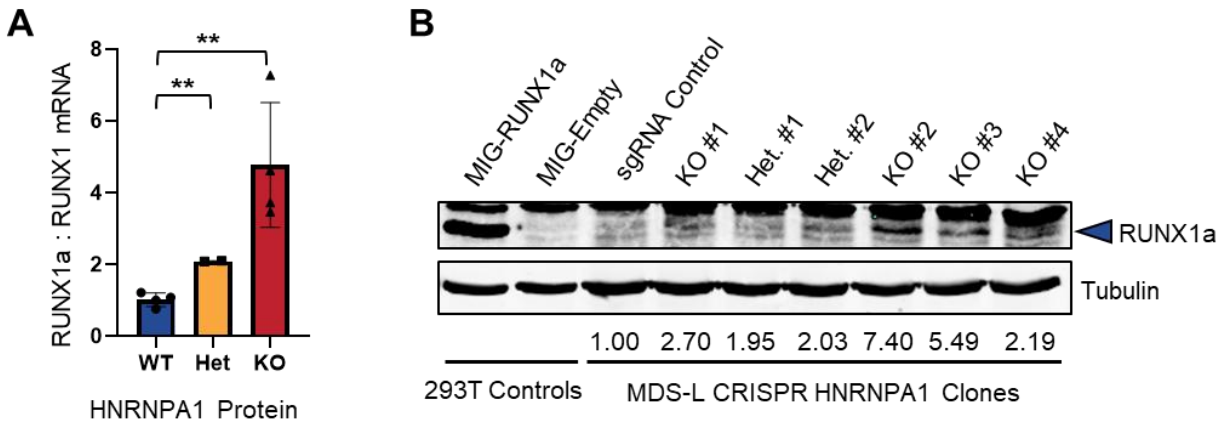


Figure 3.14. HNRNPA1 knockout MDS-L cells have enhanced *RUNX1a* transcript and protein.

(A) RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* mRNA in clonal MDS-L cell lines with normal HNRNPA1 protein levels (WT: n = 4), approximately 50% HNRNPA1 protein (Het: n = 2), and undetectable HNRNPA1 protein (KO: n = 4). Data are mean +/- s.d. ** $p < 0.01$, two-tailed student's t-tests. **(B)** Western blot showing RUNX1a and tubulin (loading control) protein. The first two lanes contain protein lysates from 293T cells that have been transfected with vector containing untagged RUNX1a cDNA or empty vector. Lane one is a positive control for RUNX1a protein (marked with the blue arrow). The remaining lanes contain protein from the clonal MDS-L cell lines. RUNX1a protein was quantified by normalizing RUNX1a signal intensity to tubulin signal intensity using LI-COR Image Studio software. Normalized protein quantifications are shown below the respective lane.

3.2.6 HNRNPA1 suppresses *RUNX1* exon 7a usage via direct binding.

We next addressed whether HNRNPA1 suppresses RUNX1a production through direct binding to exon 7a or its adjacent introns. A consensus binding motif for HNRNPA1 was previously deduced using SELEX (systematic evolution of ligands by exponential enrichment) (187), HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking and immunoprecipitation) (188), and iCLIP (individual nucleotide resolution CLIP) techniques (189). We expanded upon these previous studies and utilized the newest enhanced CLIP (eCLIP) technology (141) in MDS-L cells to deduce a consensus HNRNPA1 binding motif (**Figure 3.15A**). In accordance with previous studies, we uncovered the core 'UAG' motif within the most significantly enriched hexamer: 'UAG(A/G)GG' (**Figure 3.15B**). Most binding sites were in intronic regions, followed by untranslated regions (UTRs) (**Figure 3.15C**). Based on our cis-acting element studies, we hypothesized that suppressive elements are located within exon 7a (**Figures 3.5A, 3.5B, 3.5C, and 3.5D**). Indeed, we identified a putative HNRNPA1 binding motif (UAGAGC) in the 3'UTR region of exon 7a. To disrupt HNRNPA1 binding, we mutated the essential 'AG' dinucleotide to cytosines (UAGAGC → UCCAGC) in the dual fluorescent minigene (190, 191) (**Figure 3.16A**). We observed a significantly lower GFP to mCherry MFI ratio in cells expressing the mutated construct compared to cells expressing the wild type construct (**Figures 3.16B and 3.16C**). The construct containing the mutated HNRNPA1 binding site produced significantly more exon 7a polyadenylated product, confirming that this change in fluorescence resulted from enhanced splicing and polyadenylation of exon 7a (**Figure 3.16D**). We therefore concluded that HNRNPA1 binds directly to *RUNX1* exon 7a to suppress splicing and polyadenylation.

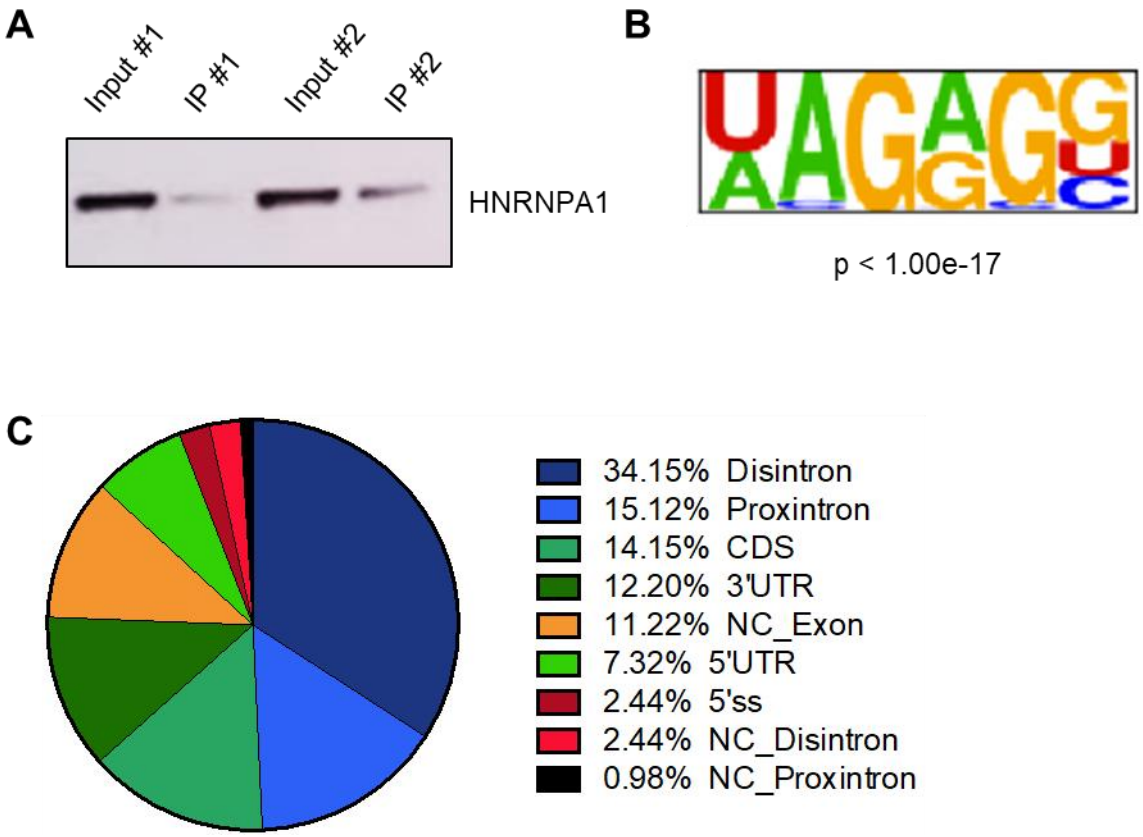


Figure 3.15. eCLIP of HNRNPA1 confirms previously deduced binding motif.

(A) Western blot confirming the success of HNRNPA1 immunoprecipitation in MDS-L cells prior to library preparation of bound RNAs. **(B)** Consensus HNRNPA1 binding motif, generated by Homer (192), as deduced by eCLIP in MDS-L cells. **(C)** Distribution of significant eCLIP peaks, indicative of HNRNPA1 binding locations. Disintron, distal intron; proxintron, proximal intron; CDS, coding sequence; UTR, untranslated region; ss, splice site; NC, non-coding.

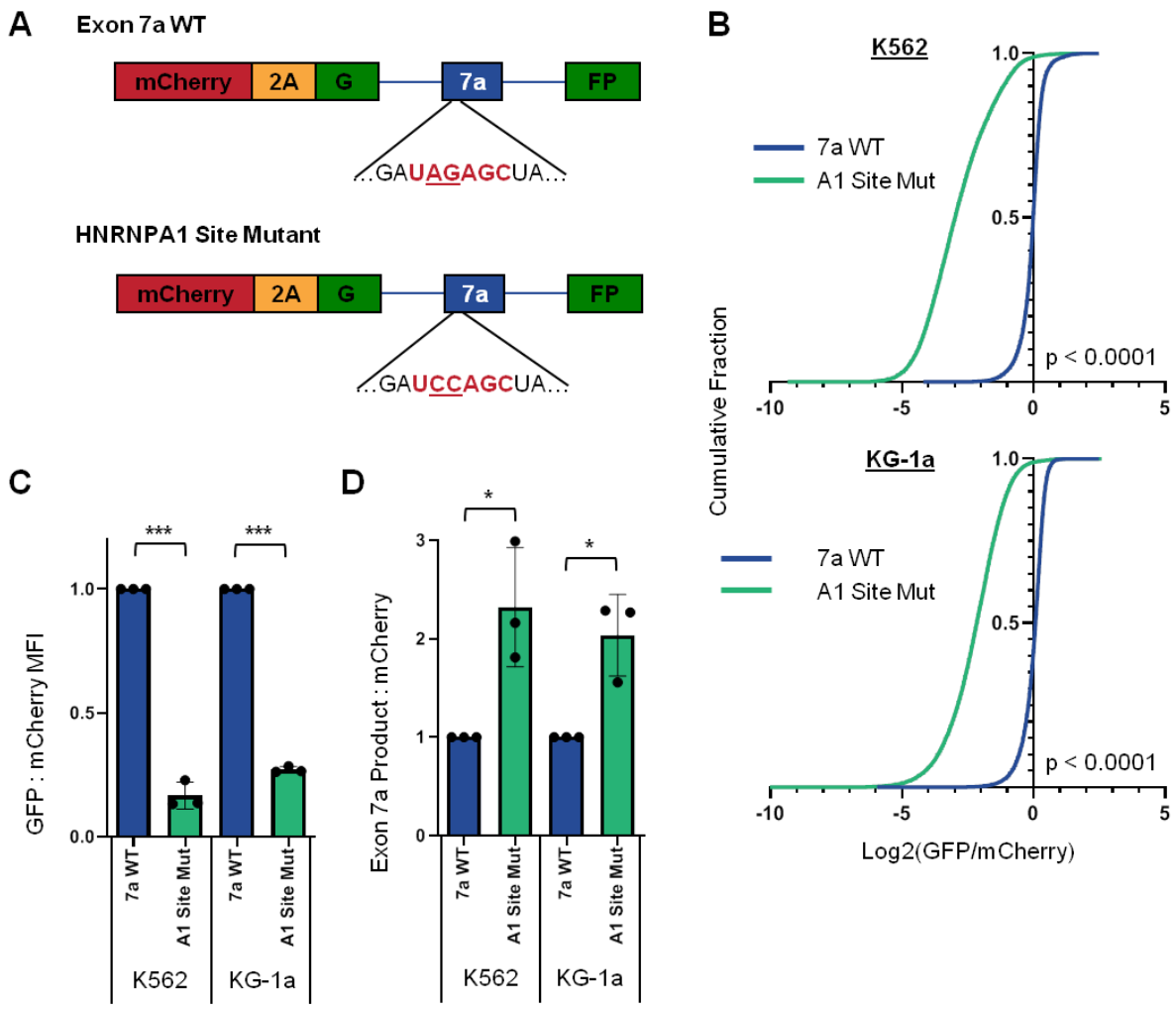


Figure 3.16. HNRNPA1 suppresses *RUNX1a* formation by direct binding to exon 7a.

(A) Schematic of the dual fluorescent *RUNX1* exon 7a wild type and HNRNPA1 site mutant constructs. The underlined 'AG' nucleotides in the wild type construct were mutated to 'CC.' **(B)** CDF plots showing the log2 ratio of GFP to mCherry in individual K562 (top) or KG-1a cells (bottom) nucleofected with the exon 7a wild type construct or HNRNPA1 site mutant construct. Data are from one representative experiment of three independent experiments. *** $p < 0.001$, Kolmogorov-Smirnov test. **(C)** Ratio of GFP mean fluorescent intensity (MFI) to mCherry MFI as measured by bulk flow cytometric analysis of K562 and KG-1a cells nucleofected with the constructs depicted in (A). Data are mean +/- s.d. of three independent experiments. *** $p < 0.001$, two-tailed student's t-test. **(D)** RT-qPCR analysis of *exon 7a product* mRNA normalized to *mCherry* mRNA from KG-1a cells nucleofected with the respective HNRNPA1 site mutant constructs. mRNA levels were normalized to that of the *RUNX1* exon 7a WT control. Data are mean +/- s.d. of three independent experiments. * $p < 0.05$, two-tailed student's t-test.

3.2.7 Additional regulators contribute to the proper balance of RUNX1 isoforms.

HNRNPA1 is abundantly expressed throughout hematopoiesis, playing a pivotal role in repressing RUNX1a isoform production. However, its absence does not confer dominance of the RUNX1a isoform over RUNX1b/c. There must be additional RBPs that contribute to this repression and others that enhance exon 7a inclusion, maintaining low levels of RUNX1a necessary for HSC pool maintenance.

In our CRIPSR RBP screen, we also sorted GFP high cells which contained sgRNAs targeting putative RUNX1a enhancers (**Figures 3.8B and 3.8C**). Among the ten most significantly enriched RBPs, four carry the “RNA splicing” GO annotation (GO:0008380) (**Table 3.1**). We knocked down each of these RBPs with siRNAs followed by RT-qPCR for endogenous *RUNX1a* relative to total *RUNX1* mRNA. From this secondary screen, we identified HNRNPC and KHDRBS1 (SAM68) as putative enhancers of RUNX1a formation (**Figures 3.17A, 3.18A, and 3.18B**). Of these two RBPs, overexpression of KHDRBS1 in K562 cells led to a significant increase in relative *RUNX1a* to total *RUNX1* mRNA (**Figures 3.18C and 3.18D**). Interestingly, the combination of KHDRBS1 overexpression with HNRNPA1 knockdown further enhanced *RUNX1a* transcript production (**Figures 3.19A and 3.19B**). These results highlight the combinatorial role of multiple RBPs in maintaining isoform pools of critical transcription factors such as RUNX1.

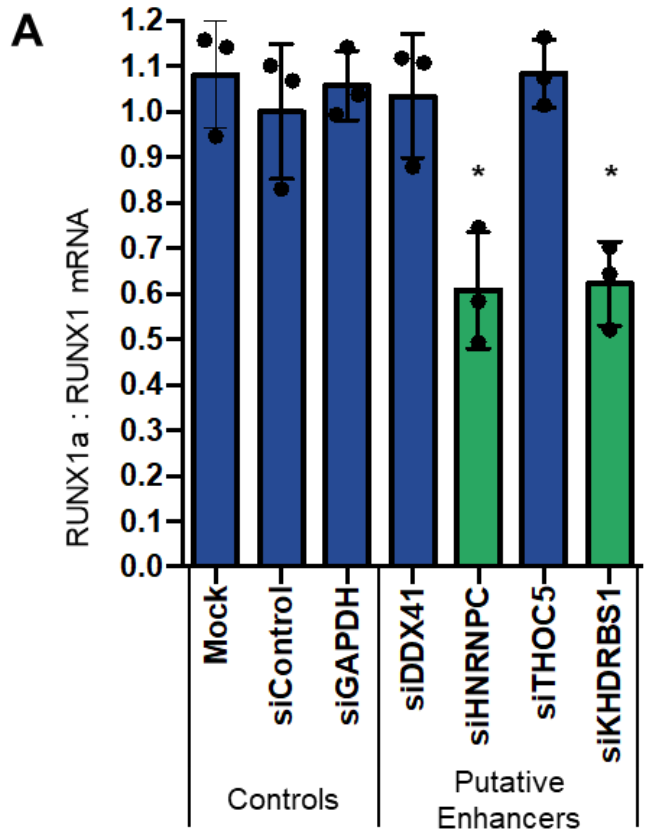


Figure 3.17. A secondary siRNA screen reveals enhancers of *RUNX1a*.

(A) RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in MDS-L cells nucleofected with the indicated siRNAs. mRNA levels were normalized to siControl nucleofected cells. The teal bars mark enhancer RBPs that had significantly decreased *RUNX1a* mRNA upon knockdown. Data are mean +/- s.d. of three biological replicates. * $p < 0.05$, one-way ANOVA with post-hoc Tukey test.

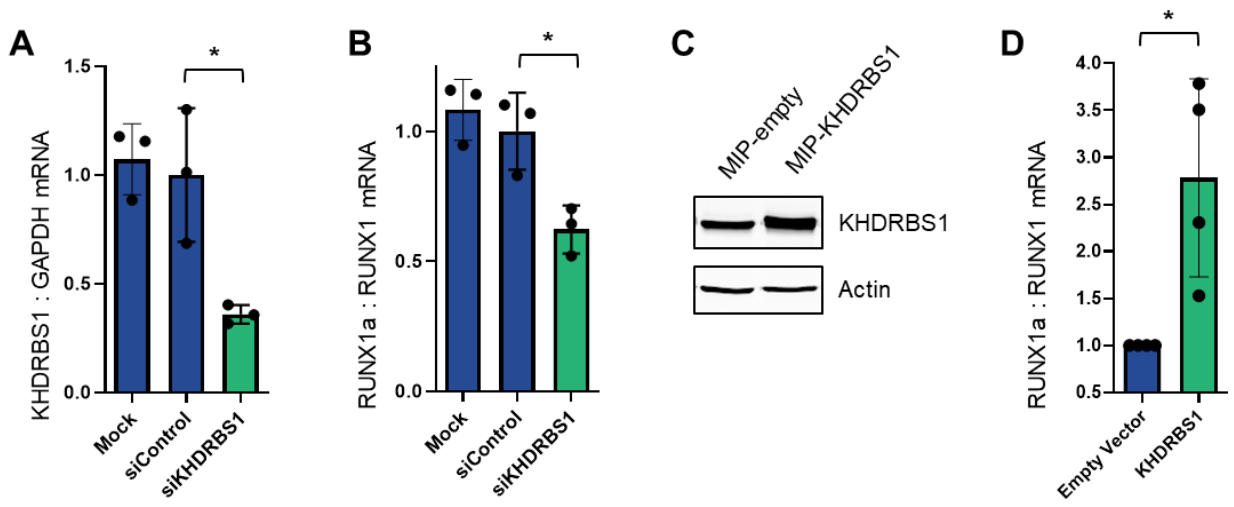


Figure 3.18. KHDRBS1 is an enhancer of *RUNX1a* isoform formation.

(A) RT-qPCR analysis of *KHDRBS1* mRNA normalized to *GAPDH* mRNA upon siRNA knockdown of *KHDRBS1* in MDS-L cells. mRNA levels were normalized to cells treated with control siRNAs. Data are mean +/- s.d. of three independent experiments. * $p < 0.05$, two-tailed student's t-test. **(B)** RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* mRNA upon siRNA knockdown of *KHDRBS1* in MDS-L cells. mRNA levels were normalized to cells treated with control siRNAs. Data are mean +/- s.d. of three independent experiments. * $p < 0.05$, two-tailed student's t-test. **(C)** Western blot showing overexpressed KHDRBS1 protein in K562 cells following transduction with retrovirus expressing KHDRBS1 cDNA. Actin protein is the loading control. **(D)** RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in K562 cells following transduction with retrovirus expressing KHDRBS1 cDNA. mRNA levels were normalized to cells transduced with an empty vector. Data are mean +/- s.d. of four independent experiments. * $p < 0.05$, two-tailed student's t-test.

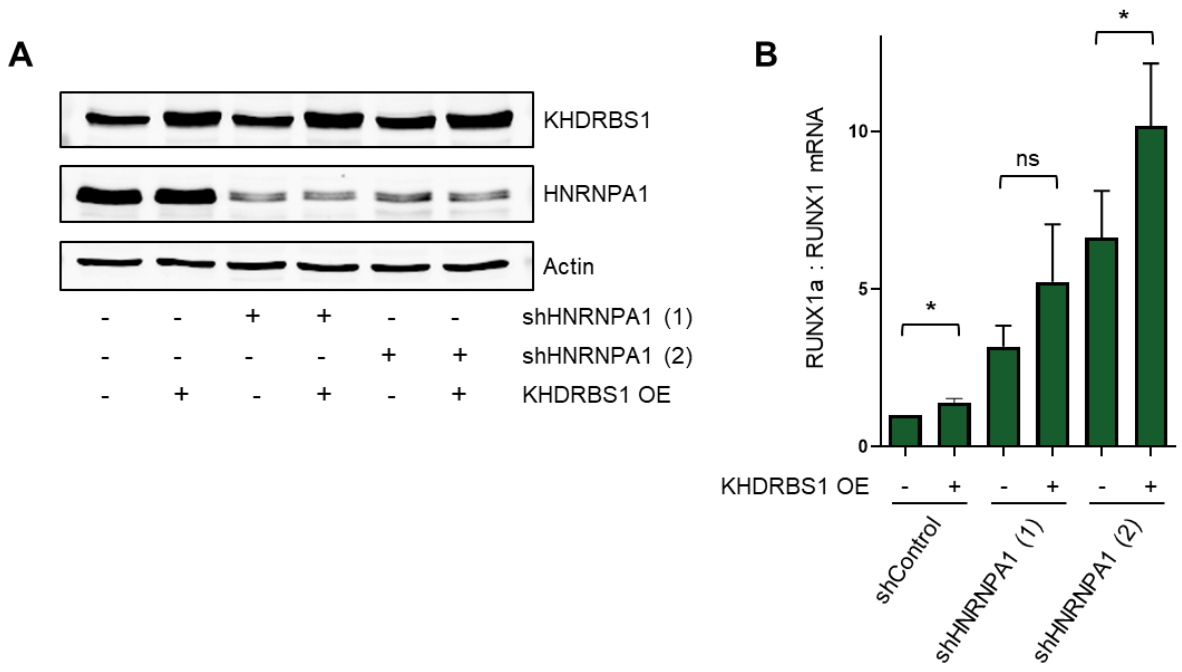


Figure 3.19. Combination of HNRNPA1 knockdown and KHDRBS1 overexpression has an additive effect on *RUNX1a* generation.

(A) Western blot showing HNRNPA1, KHDRBS1, and actin (loading control) proteins in K562 cells with the indicated combinations of shRNA-mediated HNRNPA1 knockdown and KHDRBS1 cDNA retroviral overexpression. **(B)** RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in K562 cells with the indicated combinations of shRNA-mediated HNRNPA1 knockdown and KHDRBS1 cDNA retroviral overexpression. mRNA levels were normalized to cells transduced with control shRNAs and empty vector. Data are mean +/- s.e.m. of five independent experiments. * $p < 0.05$, paired two-tailed student's t-test.

3.3 Discussion

Since transcription factors have well-defined roles in healthy hematopoiesis and leukemogenesis, studying the post-transcriptional processing of a critical hematopoietic transcription factor is a suitable strategy for identifying functionally important RBPs in hematology. Here we studied APA of *RUNX1*, a post-transcriptional event that produces antagonistic isoforms mediating HSC dynamics, that is also dysregulated in hematologic malignancy. By studying this specific event, we uncovered two RBPs, HNRNPA1 and KHDRBS1, with previously unappreciated roles in proper and aberrant hematopoiesis. Additionally, our dual fluorescent model and screening approach can be adapted and utilized to probe RBP regulation of other critical post-transcriptional events. Finally, our study represents one of few reports on the role of APA in normal hematopoiesis and assigns a role of splicing regulators to alternative, terminal exon APA.

To study RBP-RNA interactions that impact RUNX1 isoform generation, we modeled *RUNX1* post-transcriptional processing using a split GFP minigene (183, 184). This fluorescent reporter effectively recapitulated the disparity in expression of RUNX1a (alternative terminal exon 7a) and RUNX1b/c (constitutive exon 7b) in hematopoiesis, and therefore accurately depicted endogenous post-transcriptional processing of *RUNX1* (160). Importantly, this model was sensitive to detect small changes in exon inclusion. When converted to a dual fluorescent, bicistronic model and paired with a CRISPR sgRNA library targeting only RBPs, we demonstrated that this strategy accurately identified RBP regulators of a critical post-transcriptional event. Notably, our strategy can be readily adapted to probe regulation of RNA-processing events beyond this current study.

Using this fluorescent model and screening strategy, we identified RBPs that regulate RUNX1 function, expanding the growing list of post-transcriptional regulators that govern healthy hematopoiesis. First, we identified HNRNPA1 as a potent suppressor of RUNX1a formation by direct binding to the 3'UTR of alternative, terminal exon 7a. Considering the global repression of RUNX1a in hematopoiesis, the discovery of HNRNPA1 as a major player in exon 7a suppression is not surprising. HNRNPA1 has a broadly documented repressive role on exon usage and is one of the most abundantly expressed proteins in the nucleus (193). Indeed, *HNRNPA1* is highly expressed throughout hematopoiesis, maintaining a RUNX1 isoform pool that favors RUNX1b/c (**Figure 3.20A**) (144). Expression remains high in HSPCs and therefore cannot fully explain RUNX1a upregulation in these cells. Expression is lower in differentiated myeloid cells, in agreement with a previous report (194), suggesting the presence of additional suppressors. Our data supports this conclusion because RUNX1a is still not the dominant isoform in HNRNPA1 knockout cells.

HNRNPA1 is overexpressed in AML (194) and BCR-ABL positive CML patients (195), where it contributes to leukemic phenotypes (196). Furthermore, this suppressive RBP is generally reported to be oncogenic (197). Consequently, decreased HNRNPA1 expression is not a pervasive mechanism for RUNX1a upregulation in leukemia. An intriguing alternative possibility is that post-translational modification of HNRNPA1, independent of changes in expression, can impact target gene splicing. In an MDS model of TRAF6 overexpression, ubiquitination of HNRNPA1 alters splicing of *Arhgap1* (198). While our data implicates HNRNPA1 as a directly binding RBP regulator of

RUNX1 exon 7a, we cannot exclude the importance of post-translational modifications on regulation of this event.

We also identified the RUNX1a enhancer KHDRBS1 (aka SAM68), belonging to the STAR (signal transduction and activation of RNA metabolism) family of RBPs that link signal transduction pathways to post-transcriptional regulation (199). KHDRBS1 and HNRNPA1 play cooperative and antagonistic roles on RNA-processing events (200-202). Our data support an antagonistic interaction on RUNX1 isoform generation.

Importantly, KHDRBS1 expression during healthy hematopoiesis and in leukemia support a role for this RBP in fine-tuning the RUNX1 isoform ratio. During normal hematopoiesis, *KHDRBS1* is most highly expressed in HSPCs with decreased expression upon differentiation into lymphoid and myeloid cells (**Figure 3.20B**). Therefore, KHDRBS1 expression contributes to enhanced expression of RUNX1a in HSPCs, and perhaps positively regulates HSC self-renewal. In line with this hypothesis, KHDRBS1 promotes neural progenitor cell (NPC) self-renewal and knockdown induces differentiation, a system analogous to regulation of HSC fate (203). Mechanistically, KHDRBS1 affected NPC fate by regulating coding sequence APA of a glycolytic enzyme, supporting our finding that KHDRBS1 regulates alternative, terminal exon polyadenylation. Furthermore, KHDRBS1 plays a pervasive oncogenic role in various cancer types (204, 205), including leukemia. KHDRBS1 expression is essential for MLL-fusion mediated leukemic transformation (206) and is overexpressed in T-ALL (207) and AML (208). In these studies, the oncogenic role of KHDRBS1 is attributed to protein-protein interactions, not its RNA-binding ability. However, the post-transcriptional impact of KHDRBS1 in other cancer types is well-documented, affecting pertinent genes such

as *CD44*, *CCND1*, *BCL2L1*, and *mTOR* (204, 205). Our data demonstrates for the first time that KHDRBS1 regulates APA of a key hematopoietic transcription factor. This RBP is an attractive candidate for normal HSC regulation and dysregulation in leukemia.

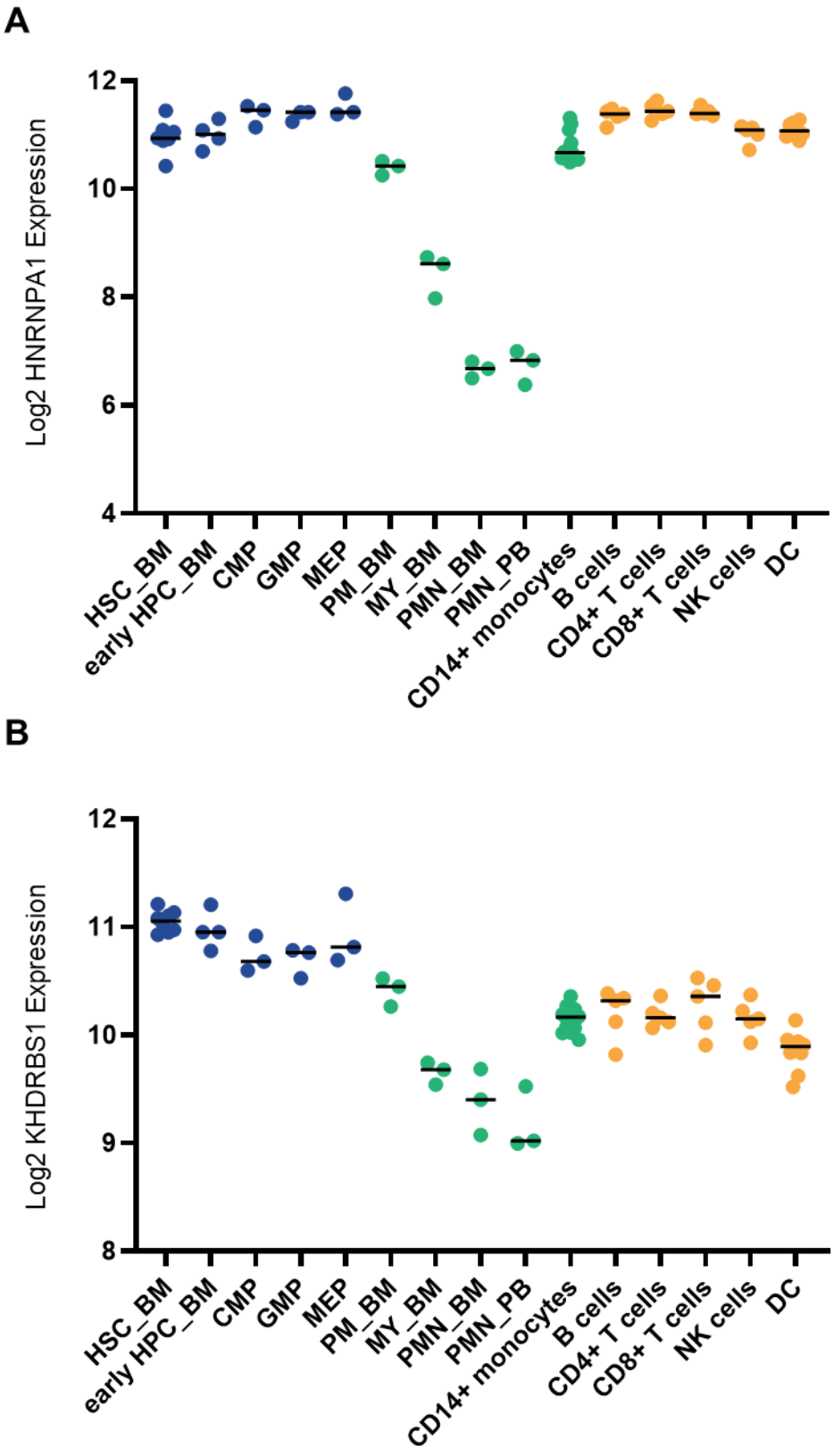
Finally, our study highlights the importance of APA in normal hematopoietic differentiation, a field of post-transcriptional regulation that is relatively understudied in hematology. Global APA trends across species and cellular contexts reveal a shift from proximal to distal poly(A) site usage during differentiation (55-57, 209) and distal to proximal poly(A) site usage during oncogenic transformation (6, 8, 67). In these respects, *RUNX1* APA fits global trends: (1) usage of proximal poly(A) site #1 (*RUNX1a*) is highest in immature HSCs and decreases upon differentiation (160) and (2) usage of proximal poly(A) site #1 is elevated in hematologic malignancy (164, 169, 179). However, a closer analysis reveals unconventional APA regulation. While proximal poly(A) sites tend to be less utilized than their distal counterparts due to inherently weak polyadenylation cis elements (3, 13), this is not the case for *RUNX1*. We saw that proximal poly(A) site #1 has higher endogenous usage than two of three distal sites, canonical cis-acting elements, and relatively strong cleavage. These observations support an evolutionarily conserved role for short RUNX1 isoforms in healthy hematopoiesis (175, 176, 178). Additionally, we implicated splicing RBPs, not APA machinery, in regulation of alternative, terminal exon polyadenylation events. KHDRBS1 has recently been reported to regulate this unique class of APA events (203, 210, 211). Considering its expression in normal hematopoiesis and upregulation in

leukemia, our finding warrants further study on the interaction between KHDRBS1 and APA in these contexts.

In conclusion, we uncovered the RBPs involved in post-transcriptional regulation of *RUNX1*, a mechanism with implications for both normal hematopoiesis and malignancy. Because of their newly assigned role in regulating RUNX1a formation, further studies on the global function of HNRNPA1 and KHDRBS1 in hematology are necessary.

Figure 3.20. *HNRNPA1* and *KHDRBS1* expression during normal hematopoiesis.

(A) *HNRNPA1* mRNA expression in healthy hematopoietic populations. Data are from the HemaExplorer cohort (GSE17054, GSE19599, GSE11864, E-MEXP-1242) obtained from the Bloodspot database. Affymetrix probe 214280_x_at. The blue color labels hematopoietic stem and progenitor populations. Teal indicates myeloid populations. Yellow are lymphoid populations. **(B)** *KHDRBS1* mRNA expression in the same dataset as in (A). Affymetrix probe 200040_at. HSC, hematopoietic stem cells; HPC, hematopoietic progenitor cells; CMP, common myeloid progenitors; GMP, granulocyte monocyte progenitors; MEP, megakaryocyte-erythroid progenitors; PM, promyelocyte; MY, myelocyte; PMN polymorphonuclear cells; NK, natural killer cells; DC, dendritic cells; BM, bone marrow; PB, peripheral blood.


A
Log2 HNRNPA1 Expression
12
10
8
6
4
HSC_BM
early HPC_BM
CMP
GMP
MEP
PM_BM
MY_BM
PMN_BM
PMN_PB
CD14+ monocytes
B cells
CD4+ T cells
CD8+ T cells
NK cells
DC
B
Log2 KHDRBS1 Expression
12
11
10
9
8
HSC_BM
early HPC_BM
CMP
GMP
MEP
PM_BM
MY_BM
PMN_BM
PMN_PB
CD14+ monocytes
B cells
CD4+ T cells
CD8+ T cells
NK cells
DC

3.4 Future Directions

By studying post-transcriptional processing of *RUNX1*, we have identified two RBPs, HNRNPA1 and KHDRBS1, with previously unappreciated roles in normal hematopoiesis. Future studies include phenotypic characterization of these RBPs in healthy hematopoiesis and mechanistic studies on global post-transcriptional regulation, with an emphasis on APA. Finally, our screening strategy paired with our *RUNX1* exon 7b construct will elucidate additional RBP regulators that regulate *RUNX1b/c* formation and can be added to our initial screen findings.

The role of HNRNPA1 in healthy hematopoiesis has been minimally studied, with conflicting functions assigned to normal myelopoiesis. One group reported that HNRNPA1 protein expression decreased over the course of myeloid differentiation and overexpression in primary HSPCs impeded differentiation in vitro (194). By contrast, overexpression of a dominant negative, shuttling deficient HNRNPA1 mutant in the murine 32Dcl3 myeloid precursor cell line induced apoptosis, slowed differentiation, and reduced colony formation (195). In the same system, overexpression induced differentiation. These contradictory results highlight the need for more careful studies on the phenotypic impact of HNRNPA1 in hematopoietic differentiation. Overexpression of a ubiquitously expressed protein is not ideal and reducing HNRNPA1 function through overexpression of a dominant negative mutant does not accurately inhibit the diverse activities of this RBP. Furthermore, there has not been a study on the role of HNRNPA1 in HSPC self-renewal or stem cell function. As such, we propose to perform the classical experiments of HNRNPA1 function in primary, human CD34+ HSPCs. Using the shRNAs that showed good knockdown in our study, we will reduce HNRNPA1

protein in CD34+ HSPCs followed by liquid culture differentiation experiments, colony formation assays, long-term culture initiating cell (LTC-IC) assays, and murine xenografts. This set of experiments will clarify the general role of HNRNPA1 in normal myelopoiesis. Furthermore, mechanistic studies of HNRNPA1 function in hematopoietic systems have been limited to its shuttling activity (195, 196) and regulatory role in translation (212, 213), with minimal focus on post-transcriptional targets (198). Therefore, we also propose to perform RNA-sequencing and 3'READS on RNA from our HNRNPA1 MDS-L knockout clonal lines and wild type controls. Paired with our eCLIP data in MDS-L cells, we will characterize direct HNRNPA1 targets and the consequence of HNRNPA1 occupancy on splicing and polyadenylation.

Our screen also revealed KHDRBS1 as an important RBP regulating RUNX1 isoform generation. KHDRBS1 expression decreases during hematopoietic differentiation and overexpression has been reported in leukemia (207, 208). With these expression profiles, it is likely that KHDRBS1 plays a pivotal role in enhancing RUNX1a formation in primitive HSPCs and upregulating RUNX1a in leukemia. The functional role of KHDRBS1 in normal hematopoiesis has not been studied, however knockdown studies in murine neural progenitor cells reveal a negative role of this RBP on cellular differentiation (203). Collectively, these data suggest that KHDRBS1 plays a positive regulatory role on HSPC expansion and that knockdown will likely promote differentiation in the hematopoietic system. We intend to test this hypothesis by performing the same set of experiments in CD34+ HSPCs as described above for HNRNPA1, using shRNAs targeting *KHDRBS1*.

Like HNRNPA1, the effect of KHDRBS1 on post-transcriptional processing in normal hematopoiesis and leukemia has not been thoroughly examined. The mechanistic contribution of KHDRBS1 to leukemogenesis has been attributed to its scaffolding role mediated by protein-protein interactions (206). Consequently, we also plan to perform knockdown in MDS-L cells followed by RNA-sequencing and 3'READS to directly address this knowledge gap and elucidate mechanisms by which KHDRBS1 contributes to leukemogenesis. Ideally, we would also perform sequencing on RNA from primary HSPCs upon KHDRBS1 knockdown. From these datasets, we would like to focus on KHDRBS1 regulation of alternative, terminal exon APA events, as recent studies indicate that KHDRBS1 may play a broad role in regulating this specific class of APA events (203, 210, 211). Furthermore, we previously tried an eCLIP of KHDRBS1 in MDS-L cells that was unsuccessful because of the antibody selected for the study. We would like to repeat this experiment using a different antibody to uncover direct RNA targets of KHDRBS1 binding. We will focus on binding sites in transcripts with altered processing upon KHDRBS1 knockdown, with an emphasis on transcripts that exhibited APA. These datasets will illuminate post-transcriptional processing events that directly contribute to the phenotypes uncovered from our knockdown studies in primary HSPCs.

Finally, in addition to the two RBPs identified in this study, our data supports another possible mechanism of RUNX1 isoform regulation. Though we mainly focused our studies on the exon 7a minigene model of *RUNX1a* isoform formation, we also generated an analogous construct containing constitutive exon 7b and modeling *RUNX1b/c* formation. When we replaced the adjacent introns of exon 7b in the minigene with the introns typically flanking exon 7a, exon 7b usage dramatically

switched from complete inclusion to almost complete exclusion. There are two possible explanations for this observation: (1) the introns flanking exon 7a contain suppressive cis-acting elements or (2) the introns flanking exon 7b contain enhancer cis elements required for exon 7b inclusion. If the second explanation is true, then a strong enhancer of exon 7b could function as a repressor of exon 7a and therefore *RUNX1a* formation. Our minigene model does not account for competition between exons, which can be affected by RBPs bound to either exon. An analogous CRISPR RBP screen using the exon 7b minigene reporter would likely uncover exon 7b enhancers that could also function as *RUNX1a* suppressors. These RBPs could also play a critical role in normal hematopoiesis and warrant further study.

Altogether, our study has expanded the list of RBPs that could have a critical regulatory function in normal and aberrant hematopoiesis. Further studies clarifying the importance of these RBPs will enhance our understanding of the role of RBPs and post-transcriptional regulation in hematology.

3.5 Materials and Methods

Primary cell sorting and 3'RNA sequencing

Primary GM-CSF mobilized leukapheresis products were obtained from UC San Diego Health Stem Cell Processing Lab (La Jolla, CA) in accordance with a university-approved Institutional Review Board (IRB) protocol. Products were thawed and diluted in 1x PBS supplemented with 1mg/mL DNase I (Sigma, #11284932001). Ficoll-Paque (VWR, #17-1440-02) was utilized for isolation of live, mononuclear cells by density gradient centrifugation. Cells were then resuspended in 1x PBS supplemented with 2%FBS, passed through a cell strainer, and washed again. Magnetic bead CD34-enrichment was performed using MACS Miltenyi Kit 130-046-702 per the manufacturer's protocol. Enriched cells were stained with CD34-APC (BD Biosciences, #555824), CD38-PECy7 (Biolegend, #303516), CD123-PE (Biolegend, #306005), and CD45RA-FITC (Biolegend, #304106) per the manufacturer's instructions. Dead cells were excluded using propidium iodide. Common myeloid progenitors (CD34+/CD38+/CD123+/CD45RA-) were sorted on a FACS Aria II (BD Biosciences, San Diego, CA) using appropriate single stain and fluorescence minus one (FMO) controls. RNA was extracted from sorted cells using Trizol reagent (ThermoFisher Scientific, #15596026) per the manufacturer's protocol. 3'READS libraries were prepared as previously described (95) and sequenced on an Illumina HiSeq4000 (San Diego, CA). Read normalization and mapping were also performed as previously described (95). RUNX1 poly(A) site usage was determined through analysis of bigwig files (.bw) generated via the standard 3'READS pipeline. Using multiBigwigSummary from DeepTools2.0 (214), a count matrix of read coverage at each base pair was

produced for the 3’ region of RUNX1. Each peak (a defined region of 200 bp) was quantified using R software (R version 3.4.4) and calculated as a percentage of all four peaks.

Tandem poly(A) reporters and RNase protection assays

The pGL3 basic empty vector backbone (Promega, E1751) was used to clone the tandem poly(A) reporters for cleavage efficiency assays. First, the synthetic poly(A) site upstream of the luciferase region was removed using the QuikChange II XL Site Mutagenesis Kit (Agilent, #200517) and inserted downstream of luciferase by BamHI/SalI restriction digestion, adding an EcoRI restriction site. The pCMV promoter was inserted upstream of luciferase by XhoI/HindIII restriction enzyme digestion. RUNX1 poly(A) sites 1-4 were individually subcloned between the coding region of luciferase and the synthetic poly(A) site by XbaI/EcoRI restriction digestion. RUNX1 poly(A) fragments were amplified from KG-1a genomic DNA using the following primer sets:

Poly(A) site #1:

5’- CCTGCAGAAATCACTTGATGCAC -3’ / 5’- AATAAGGAGACACCGGGGGAA -3’

Poly(A) site #2:

5’- GAGGTGTCCGAGGCGAC -3’ / 5’- CCACCCAAATGCAAATACGC -3’

Poly(A) site #3:

5’- ACAGCTAATAGCATGGTTCCAA -3’ / 5’- GAAGCCCACGCACGAATTTT -3’

Poly(A) site #4:

5’- CTCTCCTGGGAGCATTCGTC -3’ / 5’- GCCTTCAGAAGCCAGAGTGT-3’

Plasmids for in vitro transcription of antisense biotinylated probes were individually cloned for each RUNX1 poly(A) site using primers

5’- GCCTGAATGGCGAATGGGA -3’ / 5’- TCAGAGAGATCCTCATAAAGGC -3’

to amplify a fragment from each of the tandem poly(A) reporter constructs. Amplified fragments were ligated into the pGEM t-easy vector digested with SacII/SalI. These constructs were linearized by SalI restriction enzyme digestion, purified by gel extraction, and used as a substrate for T7 in vitro transcription using the MEGAscript T7 Transcription Kit (Ambion, AM1334) per the manufacturer’s instructions. Reactions were supplemented with biotin-UTP. Probes were purified using the MEGAclear Transcription Clean-up Kit (Ambion, AM1908) following the manufacturer’s instructions. Full length probes were further purified by gel extraction (5% polyacrylamide/urea gel).

For RNase protection assays, 293T cells were transfected with the respective tandem poly(A) reporter construct. 1 ug of the appropriate construct was mixed with 4 uL of polyethylenimine (PEI) in 100 uL of Opti-MEM I Reduced Serum Medium (Gibco, #31985-070). 24 hours following transfection, RNA was extracted using Trizol reagent (ThermoFisher Scientific, #15596026). 1 ug of 293T total RNA or 10 ug of control yeast RNA was mixed with excess T7 antisense biotinylated probe and RNase protection was performed as described using the RPA III Ribonuclease Protection Assay Kit (Ambion, AM1414). Protected fragments were run on a polyacrylamide/urea gel and transferred to Hybond-N+ nylon membrane (GE). RNA was crosslinked to the membrane using the

Spectrolinker XL-1000 UV Crosslinker set to 120 mJ/cm^2 (Spectronics Corporation; New York, New York). The blot was probed with IRDye 680RD Streptavidin (LI-COR, 926-68079) and visualized on a LI-COR Odyssey Classic Infrared Imaging System (Lincoln, Nebraska). Densitometry was performed using the LI-COR Application Software Version 3.0 and band intensities were normalized based on the number of biotinylated uracil nucleotides in each protected fragment.

Plasmids and minigene reporter cloning

The split GFP vector was kindly provided by Dr. Zefeng Wang (UNC Chapel Hill) (183, 184). RUNX1 exon 7a and 7b wildtype reporters were made by amplifying genomic DNA from KG-1a cells and inserting the resulting fragment into the split GFP reporter by HindIII/KpnI digestion.

Exon 7a: 5'- GTTTTCACGTGACCCAGCAC -3' / 5'- GGGACCTAGCATCTCCCTGA -3'

Exon 7b: 5'- CTTGGGAGAGAATTCGCCTTA -3' / 5'- TGGAACCAGTCCTCCATGGA -3'

Chimeric minigene constructs were generated by amplifying individual intronic / exonic regions from the wild type constructs by PCR, joining fragments by overlap PCR, and inserting the product into the split GFP vector by HindIII/KpnI digestion. For exon 7a chimeric constructs, the 3' splice site was maintained so that splicing can occur normally with the first GFP exon. For exon 7b chimeric constructs, the 3' and 5' splice sites were both maintained so that splicing can occur normally with both flanking GFP exons.

The mCherry-P2A-split GFP RUNX1 exon 7a minigene (dual fluorescent reporter) was generated by cloning mCherry and a P2A peptide directly upstream of GFP. mCherry was amplified from MSCV-IRES-mCherry, using primers to add the P2A peptide. Exon 1 of the split GFP minigene was also amplified and added to the mCherry-P2A fragment by overlap PCR such that translation of the dual fluorescent reporter was in-frame. The combined fragment was reintroduced into the split GFP minigene vector by NheI/SacI restriction enzyme digestion.

The HNRNPA1 binding site mutant construct was made with the QuikChange II XL Site Mutagenesis Kit (Agilent, #200517) per the manufacturer's protocol, using primers:

5'- CCCTCTCCCCAGCCAGGATCCAGCTATCTTTTCCA -3' /

5'- TGGAAAAGATAGCTGGATCCTGGCTGGGGAGAGGG -3'

The pEGFP-C1 plasmid was obtained from Clontech (#6084-1).

Cell Culture

HEK293T cells were purchased from ATCC and cultured in DMEM supplemented with 10% BCS and 100U/mL penicillin/streptomycin. HL-60, K562, KG-1a, THP-1, and U937 leukemia cell lines were all purchased from ATCC and maintained as low-passage stocks. MDS-L cells were kindly provided by Dr. Daniel Starczynowski (Cincinnati Children's Hospital, Ohio). ME-1 cells were obtained from Dr. Paul Liu (NIH, Maryland). KT-1 cells were generously provided by Dr. Ikuya Sakai (Ehime University School of Medicine, Japan). K562, THP-1, KT-1, HL-60, U937, and MDS-L cells were

cultured in RPMI supplemented with 10% FBS and 100U/mL penicillin/streptomycin. MDS-L cells were also supplemented with 10 ng/mL recombinant human IL-3 (Peprotech, #200-03). KG-1a and ME-1 cells were cultured in RPMI supplemented with 20% FBS and 100U/mL penicillin/streptomycin. All cells were cultured at 37°C and 5% CO_2.

Nucleofection of minigene constructs and siRNAs

For all minigene assays, minigene DNA was introduced into KG-1a cells by nucleofection using program V-001 of the AMAXA II Nucleofector (Lonza; Basel, Switzerland). For nucleofection, 3 ug of the indicated minigene DNA was mixed with 1 million KG-1a cells in 100 uL of nucleofection buffer (140 mM Na_2HPO_4/NaH_2PO_4 (pH 7.2), 5 mM KCl, 15 mM $MgCl_2$, 48 mM $NaCl_2$). Following nucleofection, cells were carefully transferred to a 12-well plate containing pre-warmed RPMI supplemented with 20% FBS and no penicillin/streptomycin. Nucleofected cells were analyzed 24 hours post-nucleofection by flow cytometry, RT-PCR, or RT-qPCR.

For the siRNA knockdown secondary validation screen, MDS-L cells were nucleofected with 200 pmol of ON-TARGETplus Human SMARTpool siRNA (Dharmacon) targeting the indicated RBP. 1.5 million MDS-L cells were mixed with siRNA in 100 uL of nucleofection buffer (140 mM Na_2HPO_4/NaH_2PO_4 (pH 7.2), 5 mM KCl, 15 mM $MgCl_2$) and nucleofected using program V-001 of the AMAXA II Nucleofector. Cells were carefully transferred to a 12-well plate containing pre-warmed RPMI supplemented with 20% FBS and 10 ng/mL IL-3. Cells were collected 48 hours post-nucleofection for RNA extraction and RT-qPCR analysis. The ON-TARGETplus

Human SMARTpool siRNAs purchased from Dharmacon are as follows: siControl (Non-targeting siRNA, D-001810-01-05), siGAPDH (D-001830-10-05), siDHX15 (L-011250-01), siHNRNPA1 (L-008221-00), siLUC7L3 (L-015383-00), siSF3A2 (L-018282-02), siRBM7 (L-017936-02), siDBR1 (L-008290-00), siHNRNPK (L-011692-00), siRBMX2 (L-020763-02), siDDX41 (L-010394-00), siHNRNPC (L-011869-03), siTHOC5 (L-015317-01), siKHDRBS1 (L-020019-00). Lyophilized siRNAs were resuspended in siRNA buffer (B-002000-UB-100) diluted to 1x with RNase-free water and stored in aliquots at -20°C.

Retroviral and lentiviral production and transduction

For HNRNPA1 shRNA knockdown experiments, MDS-L and K562 cells were infected with lentivirus produced in HEK293T cells. HEK293T cells were transfected with 3 ug of pLKO-based lentiviral vector, 5 ug of psPAX2, 2.5 ug of pMD2.G, and 42 uL of polyethylenimine (PEI) in 1 mL of Opti-MEM I Reduced Serum Medium. 24 hours post-transfection, the cell supernatant was aspirated and replaced with fresh RPMI supplemented with 10% FBS. 24 hours following the media replacement, the 293T cell supernatant containing lentiviral particles was collected and passed through a 0.45 um syringe filter. Virus was added to MDS-L or K562 cells seeded at 0.3-0.5 million cells per mL and supplemented with 4 ug/mL polybrene. Cells were transduced on two consecutive days in 6-well tissue culture plates by centrifugation (2,000 x g) for 3 hours at 32°C in an Allegra X-12R centrifuge (Beckman Coulter; Brea, CA). 24 hours following the second transduction, cells were resuspended in fresh RPMI media with 1 ug/mL puromycin. After 48 hours of puromycin selection, cells were diluted 1:2 and maintained at 0.5 ug/mL puromycin until RNA or protein was collected on day 6 post-transduction.

The HNRNPA1 pLKO shRNA clones utilized in this study were: TRCN0000235097 (shRNA #1) and TRCN0000235098 (shRNA #2). The pLKO control shRNA was Addgene plasmid #1864.

For KHDRBS1 overexpression experiments, retrovirus was produced in HEK293T cells. 293T cells were transfected with 5 ug of MSCV-IRES-PURO (MIP) empty vector or MIP vector expressing KHDRBS1 cDNA, 5 ug of pCL-10A1, and 40 uL of PEI in 1 mL of Opti-MEM I. Virus was collected and K562 cells were transduced as described above. Puromycin selection of transduced cells was also performed as previously described. The KHDRBS1 cDNA was obtained from transOMIC Technologies (BC019109) and subcloned into the MIP vector.

For HNRNPA1 knockdown and KHDRBS1 overexpression dual treated K562 cells, the same protocol for viral production and transduction was followed as above. Instead of MSCV-IRES-PURO overexpression vectors, MSCV-IRES-NEO plasmids were used. Consequently, double transduced cells were selected with 1 ug/mL puromycin and 1 mg/mL G418. RNA and protein were collected 6 days post-transduction.

CRISPR screening and analysis

The clonal MDS-L dual fluorescent reporter cell line used for the CRISPR screen was generated by nucleofection of MDS-L cells as described with linearized mCherry-P2A-spilt GFP RUNX1 exon 7a minigene construct. Following two weeks of G418 selection, single mCherry+/GFP+ cells were sorted into 96-well plates using a BD FACSAria II (BD Biosciences, San Diego, CA).

For sgRNA CRISPR library lentiviral production, HEK293T cells were seeded in 15 cm plates. One hour prior to transfection, media was replaced with 8 mL of pre-warmed Opti-MEM. Cells were then transfected with a mixture consisting of 62.5 uL Lipofectamine 2000, 125 uL Plus reagent, 12.5 ug lentiCRISPR plasmid library, 6.25 ug of pMD2.G, and 9.375 ug psPAX2. 6 hours following transfection, the media was changed to DMEM supplemented with 10% FBS. Supernatant was collected 48 hours later and filtered through a 0.45 um low protein binding membrane. Virus was ultracentrifuged (24,000 rpm) for 2 hours at 4°C and resuspended in PBS overnight. Virus aliquots were stored in -80°C. For infection of the MDS-L dual reporter line, the volume of virus to achieve a multiplicity of infection ~0.3 was titrated prior to the screen.

For the CRISPR screen, virus was added to 6-well dishes containing 13 million MDS-L reporter cells seeded in duplicate at a density of 0.3-0.5 million cells per mL and supplemented with 4 ug/mL polybrene. Cells were transduced by centrifugation (2,000 x g) on two consecutive days for 3 hours at 32°C in an Allegra X-12R centrifuge (Beckman Coulter; Brea, CA). Cells were resuspended in puromycin 24 hours following the second transduction and cultured in the presence of puromycin for 72 hours, at which time 4 million cells were collected as ‘day 0.’ Following 3 weeks of culture, 4 million cells were collected as a ‘day 21’ endpoint. At this time, GFP low and GFP high cells relative to mCherry were sorted using a FACSAria II. Genomic DNA was extracted from each cell population using the DNeasy Blood and Tissue Kit (Qiagen, #69504). Integrated sgRNA sequences were amplified by PCR using Herculase II Fusion DNA Polymerase (Agilent Technologies, #600679) and the following primers:

Forward (equimolar mixture of the following 4 primers) – the underlined portion anneals to the integrated sgRNA region; the bolded portion is the Illumina compatible overhang for index addition:

5’- **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CTAGTGGAAAGGACGAAACACCG -3’

5’- **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GCTAGTGGAAAGGACGAAACACCG -3’

5’- **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TGCCAGTGGAAAGGACGAAACACCG -3’

5’- **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AAGTCCGTGGAAAGGACGAAACACCG -3’

Reverse:

5’- TGGAAAAGATAGCTGGATCCTGGCTGGGGAGAGGG -3’.

PCR products were agarose gel-purified, then Nextera XT Indexes (Illumina, #15032353) were added to the amplicons using NEB Taq DNA Polymerase (M0273L). Libraries were pooled and sequenced on the Illumina HiSeq4000 (San Diego, CA).

Adapters were trimmed from the raw sequencing reads using Cutadapt (215). Beta scores representing sgRNA enrichment in GFP high and low populations were calculated using the MAGeCK-VISPR pipeline (186).

HNRNPA1 CRISPR knockout cell line generation

The HNRNPA1 sgRNA that was most dramatically enriched in the GFP low population (5’-GCCGTCATGTCTAAGTCAG -3’) was subcloned into the lentiCRISPR v2 vector (Addgene, #52961). HNRNPA1 CRISPR knockout MDS-L clonal cell lines were generated by transducing cells as described with the Cas9/sgRNA vector followed by

puromycin selection. On-target cleavage in the bulk cell population was confirmed by performing the T7 endonuclease assay on DNA amplified from transduced cells by a primer pair that spans the target site (5'- ACGACCGAAGGAATGACGTT -3' / 5'- TTACCACACAGTCCGTGAGC -3'). Single cells were sorted into 96-well plates using a FACSAria II. Clonal lines were screened for HNRNPA1 protein by western blotting and alleles of selected lines were TA-cloned into the pGEM t-easy vector (Promega, A1360) followed by Sanger sequencing.

Flow cytometry

GFP mean fluorescent intensity (MFI) was measured in minigene nucleofected KG-1a cells using a BD FACSCanto flow cytometer (San Diego, CA) and BD FACSDiva Acquisition software. Propidium iodide was used to exclude dead cells.

GFP and mCherry fluorescence were measured in dual fluorescent minigene nucleofected KG-1a cells using a BD LSRFortessa and BD FACSDiva Acquisition software. Sytox Blue Dead Cell Stain (ThermoFisher Scientific, #S34857) was used to exclude dead cells. Post-acquisition data analysis was performed using FlowJo software (FlowJo, LLC).

RT-PCR analysis and imaging

RNA was extracted from KG-1a cells nucleofected with the indicated minigene construct using Trizol reagent. cDNA was prepared from 0.5-1ug of RNA by reverse transcription using the First Strand cDNA Synthesis Kit (Molecular Cloning Laboratories, FSCS-200) following the manufacturer's instructions. PCR of the cDNA was performed

with KOD Hot Start DNA polymerase (EMD Millipore, 71086-3) using the following primer sets:

GFP: 5’- AGTGCTTCAGCCGCTACCC -3’ / 5’- GTTGTACTCCAGCTTGTGCC -3’

Exon 7a spliced and polyadenylated product: 5’- AGTGCTTCAGCCGCTACCC -3’ / 5’- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATAGTG -3’

RT-PCR products were run on 2% agarose gels and imaged with a Biorad Gel Doc XR+ (Hercules, CA) using the Biorad Quantity One 1-D Analysis Software.

RT-qPCR

RNA isolation and cDNA preparation were performed as described above. Quantitative PCR (qPCR) reactions were performed in duplicate or triplicate on a BioRad CFX Connect (Hercules, CA) using KAPA SYBR FAST qPCR Master Mix (2X) Universal (KAPA Biosystems, #KK4618). Data were analyzed using the delta delta Ct method and normalized as described in the figure legends. The following primer pairs were used:

RUNX1 exon 7a minigene product:

5’- AGTGCTTCAGCCGCTACCC -3’ / 5’- GAAAGTGTACCGGGATCCATG -3’

GAPDH:

5’- TCGCTCAGACACCATGGGGAAG -3’ / 5’- GCCTTGACGGTGCCATGGAATTTG -3’

RUNX1a:

5’- CTCCCTGAACCACTCCACTG -3’ / 5’- TTAACATCTCCAGGGTGCTGTGTCTTC -3’

RUNX1b/c:

5’- CTCCCTGAACCACTCCACTG -3’ / 5’- CAGAGAGGGTTGTCATGCCG -3’

RUNX1 (Total):

5’- CCTCAGGTTTGTCGGTCGAA -3’ / 5’- CTGCCGATGTCTTCGAGGTT -3’

HNRNPA1:

5’- AAGGTAGGCTGGCAGATACG -3’ / 5’- CGGGCTCTTTAGGAGACTCTG -3’

mCherry:

5’- CACGAGTTCGAGATCGAGGG -3’ / 5’- CAAGTAGTCGGGGATGTCGG -3’

KHDRBS1:

5’- GAAATTTCTAGTACCGGATATGATG -3’ / 5’- TGGTGTACCACGTACCAAAG -3’

Western blotting

Transduced K562 and MDS-L cells, transfected 293T cells, and MDS-L CRISPR clonal lines were lysed on ice in RIPA buffer (50 mM Tris, 1 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors ((Roche, # 11873580001). Lysates were cleared by centrifugation (10,000 x g), mixed with loading buffer, and denatured at 95°C. Conventional SDS-PAGE and western blotting techniques were used to analyze the lysates. The following primary antibodies were

used in this study, following the manufacturer-recommended dilutions: HNRNPA1 (Santa Cruz, sc-32301), KHDRBS1 (Bethyl Laboratories, A302-110A), RUNX1 (generated by Covance) (142), β-Actin (Sigma, A1978), and α-Tubulin (Developmental Studies Hybridoma Bank, #12G10). The following secondary western antibodies were used in this study: IRDye 800CW goat anti-rabbit IgG (LI-COR, #926-32211), IRDye 800CW goat anti-mouse IgG (LI-COR, # 926-32210), and IRDye 680RD goat anti-mouse IgG (LI-COR, #926-68070). Membranes were scanned on the LI-COR Odyssey Classic Infrared Imaging System and densitometry was performed using the LI-COR Application Software Version 3.0.

eCLIP

Enhanced UV crosslinking and immunoprecipitation (eCLIP) was carried out as previously described (141) on cell lysate from two pools of 20 million MDS-L cells using the HNRNPA1 antibody (Santa Cruz, sc-32301). Size-matched input RNA was prepared as the background control. Libraries were sequence on an Illumina HiSeq4000 (San Diego, CA) and analysis was performed using the established eCLIP bioinformatics pipeline (141).

Expression of *HNRNPA1* and *KHDRBS1* throughout hematopoiesis

HNRNPA1 and *KHDRBS1* mRNA expression data in various hematopoietic populations were obtained from the BloodSpot server (bloodspot.eu) (144). The HemaExplorer dataset (GSE17054, GSE19599, GSE11864, E-MEXP-1242) is shown for each RBP with the following Affymetrix probe IDs: HNRNPA1 (214280_x_at) and KHDRBS1 (200040_at).

Statistical analyses

All statistical analyses were performed using the GraphPad Prism Software (Version 8.3.1). All statistical tests are described in the respective figure legends. P-values are consistently annotated throughout: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data availability

3'RNA sequencing data from healthy common myeloid progenitors (CMP) and results from the CRISPR RBP screen were deposited to the Gene Expression Omnibus (GEO), accession GSE145968. All DNA constructs cloned for this study will either be deposited to Addgene for purchase, or available by email request to the corresponding author.

Acknowledgments

Chapter 3, in part, has also been submitted for publication by Davis AG, Einstein JE, Zheng D, Jayne ND, Wang R, Fu XD, Tian B, Yeo GW, and Zhang DE. The dissertation author is the primary investigator and writer of this manuscript.

References

1. Colgan, D. F., and J. L. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11: 2755-2766.

2. Sachs, A. 1990. The role of poly(A) in the translation and stability of mRNA. *Curr Opin Cell Biol* 2: 1092-1098.

3. Tian, B., J. Hu, H. Zhang, and C. S. Lutz. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33: 201-212.

4. Lianoglou, S., V. Garg, J. L. Yang, C. S. Leslie, and C. Mayr. 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* 27: 2380-2396.

5. Zhang, H., J. Y. Lee, and B. Tian. 2005. Biased alternative polyadenylation in human tissues. *Genome Biol* 6: R100.

6. Mayr, C., and D. P. Bartel. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138: 673-684.

7. Jenal, M., R. Elkon, F. Loayza-Puch, G. van Haaften, U. Kuhn, F. M. Menzies, J. A. Oude Vrielink, A. J. Bos, J. Drost, K. Rooijers, D. C. Rubinsztein, and R. Agami. 2012. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149: 538-553.

8. Xia, Z., L. A. Donehower, T. A. Cooper, J. R. Neilson, D. A. Wheeler, E. J. Wagner, and W. Li. 2014. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* 5: 5274.

9. Hirose, Y., and J. L. Manley. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* 14: 1415-1429.

10. Proudfoot, N. J., A. Furger, and M. J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* 108: 501-512.

11. Tian, B., and J. H. Graber. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 3: 385-396.

12. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263: 211-214.

13. Beaudoing, E., S. Freier, J. R. Wyatt, J. M. Claverie, and D. Gautheret. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 10: 1001-1010.

14. Sheets, M. D., S. C. Ogg, and M. P. Wickens. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* 18: 5799-5805.

15. McDevitt, M. A., R. P. Hart, W. W. Wong, and J. R. Nevins. 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. *Embo j* 5: 2907-2913.

16. Gil, A., and N. J. Proudfoot. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 49: 399-406.

17. Venkataraman, K., K. M. Brown, and G. M. Gilmartin. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* 19: 1315-1327.

18. Zarudnaya, M. I., I. M. Kolomiets, A. L. Potyahaylo, and D. M. Hovorun. 2003. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res* 31: 1375-1386.

19. Murthy, K. G., and J. L. Manley. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* 267: 14804-14811.

20. Bienroth, S., E. Wahle, C. Suter-Crazzolara, and W. Keller. 1991. Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J Biol Chem* 266: 19768-19776.

21. Keller, W., S. Bienroth, K. M. Lang, and G. Christofori. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *Embo j* 10: 4241-4249.

22. Murthy, K. G., and J. L. Manley. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* 9: 2672-2683.

23. Chan, S. L., I. Huppertz, C. Yao, L. Weng, J. J. Moresco, J. R. Yates, 3rd, J. Ule, J. L. Manley, and Y. Shi. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* 28: 2370-2380.

24. Schonemann, L., U. Kuhn, G. Martin, P. Schafer, A. R. Gruber, W. Keller, M. Zavolan, and E. Wahle. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* 28: 2381-2393.

25. Mandel, C. R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley, and L. Tong. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444: 953-956.

26. Kaufmann, I., G. Martin, A. Friedlein, H. Langen, and W. Keller. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *Embo j* 23: 616-626.

27. Martin, G., A. R. Gruber, W. Keller, and M. Zavolan. 2012. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* 1: 753-763.

28. Lackford, B., C. Yao, G. M. Charles, L. Weng, X. Zheng, E. A. Choi, X. Xie, J. Wan, Y. Xing, J. M. Freudenberg, P. Yang, R. Jothi, G. Hu, and Y. Shi. 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *Embo j* 33: 878-889.

29. MacDonald, C. C., J. Wilusz, and T. Shenk. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* 14: 6647-6654.

30. Takagaki, Y., J. L. Manley, C. C. MacDonald, J. Wilusz, and T. Shenk. 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* 4: 2112-2120.

31. Takagaki, Y., R. L. Seipelt, M. L. Peterson, and J. L. Manley. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* 87: 941-952.

32. Yao, C., J. Biesinger, J. Wan, L. Weng, Y. Xing, X. Xie, and Y. Shi. 2012. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A* 109: 18773-18778.

33. Yang, Q., G. M. Gilmartin, and S. Doublie. 2010. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A* 107: 10062-10067.

34. Yang, Q., M. Coseno, G. M. Gilmartin, and S. Doublie. 2011. Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure* 19: 368-377.

35. Schafer, P., C. Tuting, L. Schonemann, U. Kuhn, T. Treiber, N. Treiber, C. Ihling, A. Graber, W. Keller, G. Meister, A. Sinz, and E. Wahle. 2018. Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. *Rna* 24: 1721-1737.

36. Takagaki, Y., and J. L. Manley. 2000. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* 20: 1515-1525.

37. Di Giammartino, D. C., W. Li, K. Ogami, J. J. Yashinskie, M. Hoque, B. Tian, and J. L. Manley. 2014. RBBP6 isoforms regulate the human polyadenylation

machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* 28: 2248-2260.

38. Lee, S. D., and C. L. Moore. 2014. Efficient mRNA polyadenylation requires a ubiquitin-like domain, a zinc knuckle, and a RING finger domain, all contained in the Mpe1 protein. *Mol Cell Biol* 34: 3955-3967.

39. Eckmann, C. R., C. Rammelt, and E. Wahle. 2011. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* 2: 348-361.

40. Elkon, R., A. P. Ugalde, and R. Agami. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. In *Nat Rev Genet*, England. 496-506.

41. Sandberg, R., J. R. Neilson, A. Sarma, P. A. Sharp, and C. B. Burge. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320: 1643-1647.

42. Park, H. J., P. Ji, S. Kim, Z. Xia, B. Rodriguez, L. Li, J. Su, K. Chen, C. P. Masamha, D. Baillat, C. R. Fontes-Garfias, A. B. Shyu, J. R. Neilson, E. J. Wagner, and W. Li. 2018. 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. *Nat Genet* 50: 783-789.

43. Gruber, A. R., G. Martin, P. Muller, A. Schmidt, A. J. Gruber, R. Gumienny, N. Mittal, R. Jayachandran, J. Pieters, W. Keller, E. van Nimwegen, and M. Zavolan. 2014. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun* 5: 5465.

44. Spies, N., C. B. Burge, and D. P. Bartel. 2013. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res* 23: 2078-2090.

45. Gaidatzis, D., E. van Nimwegen, J. Hausser, and M. Zavolan. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* 8: 69.

46. Hoffman, Y., D. R. Bublik, A. P. Ugalde, R. Elkon, T. Biniashvili, R. Agami, M. Oren, and Y. Pilpel. 2016. 3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-differentiation Genes in Proliferating Human Cells. *PLoS Genet* 12: e1005879.

47. Gupta, I., S. Clauder-Munster, B. Klaus, A. I. Jarvelin, R. S. Aiyar, V. Benes, S. Wilkening, W. Huber, V. Pelechano, and L. M. Steinmetz. 2014. Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Mol Syst Biol* 10: 719.

48. Miles, W. O., A. Lembo, A. Volorio, E. Brachtel, B. Tian, D. Sgroi, P. Provero, and N. Dyson. 2016. Alternative Polyadenylation in Triple-Negative Breast

Tumors Allows NRAS and c-JUN to Bypass PUMILIO Posttranscriptional Regulation. *Cancer Res* 76: 7231-7241.

49. Allen, M., C. Bird, W. Feng, G. Liu, W. Li, N. I. Perrone-Bizzozero, and Y. Feng. 2013. HuD promotes BDNF expression in brain neurons via selective stabilization of the BDNF long 3'UTR mRNA. *PLoS One* 8: e55718.

50. Taliaferro, J. M., M. Vidaki, R. Oliveira, S. Olson, L. Zhan, T. Saxena, E. T. Wang, B. R. Graveley, F. B. Gertler, M. S. Swanson, and C. B. Burge. 2016. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol Cell* 61: 821-833.

51. Berkovits, B. D., and C. Mayr. 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522: 363-367.

52. Lee, S. H., and C. Mayr. 2019. Gain of Additional BIRC3 Protein Functions through 3'-UTR-Mediated Protein Complex Formation. *Mol Cell* 74: 701-712.e709.

53. Alt, F. W., A. L. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore. 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* 20: 293-301.

54. Lee, S. H., I. Singh, S. Tisdale, O. Abdel-Wahab, C. S. Leslie, and C. Mayr. 2018. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 561: 127-131.

55. Ji, Z., J. Y. Lee, Z. Pan, B. Jiang, and B. Tian. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 106: 7028-7033.

56. Hilgers, V., M. W. Perry, D. Hendrix, A. Stark, M. Levine, and B. Haley. 2011. Neural-specific elongation of 3' UTRs during Drosophila development. *Proc Natl Acad Sci U S A* 108: 15864-15869.

57. Ulitsky, I., A. Shkumatava, C. H. Jan, A. O. Subtelny, D. Koppstein, G. W. Bell, H. Sive, and D. P. Bartel. 2012. Extensive alternative polyadenylation during zebrafish development. *Genome Res* 22: 2054-2066.

58. Ji, Z., and B. Tian. 2009. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* 4: e8419.

59. Ciolli Mattioli, C., A. Rom, V. Franke, K. Imami, G. Arrey, M. Terne, A. Woehler, A. Akalin, I. Ulitsky, and M. Chekulaeva. 2019. Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. *Nucleic Acids Res* 47: 2560-2573.

60. Flavell, S. W., T. K. Kim, J. M. Gray, D. A. Harmin, M. Hemberg, E. J. Hong, E. Markenscoff-Papadimitriou, D. M. Bear, and M. E. Greenberg. 2008. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* 60: 1022-1038.

61. Niibori, Y., F. Hayashi, K. Hirai, M. Matsui, and K. Inokuchi. 2007. Alternative poly(A) site-selection regulates the production of alternatively spliced vesl-1/homer1 isoforms that encode postsynaptic scaffolding proteins. *Neurosci Res* 57: 399-410.

62. Zheng, D., R. Wang, Q. Ding, T. Wang, B. Xie, L. Wei, Z. Zhong, and B. Tian. 2018. Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nat Commun* 9: 2268.

63. Masamha, C. P., Z. Xia, J. Yang, T. R. Albrecht, M. Li, A. B. Shyu, W. Li, and E. J. Wagner. 2014. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 510: 412-416.

64. Elkon, R., J. Drost, G. van Haaften, M. Jenal, M. Schrier, J. A. Oude Vrielink, and R. Agami. 2012. E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol* 13: R59.

65. Akman, H. B., M. Oyken, T. Tuncer, T. Can, and A. E. Erson-Bensan. 2015. 3'UTR shortening and EGF signaling: implications for breast cancer. *Hum Mol Genet* 24: 6910-6920.

66. Singh, I., S. H. Lee, A. S. Sperling, M. K. Samur, Y. T. Tai, M. Fulciniti, N. C. Munshi, C. Mayr, and C. S. Leslie. 2018. Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat Commun* 9: 1716.

67. Singh, P., T. L. Alley, S. M. Wright, S. Kamdar, W. Schott, R. Y. Wilpan, K. D. Mills, and J. H. Graber. 2009. Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res* 69: 9422-9430.

68. Li, W., B. You, M. Hoque, D. Zheng, W. Luo, Z. Ji, J. Y. Park, S. I. Gunderson, A. Kalsotra, J. L. Manley, and B. Tian. 2015. Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet* 11: e1005166.

69. Ogorodnikov, A., M. Levin, S. Tattikota, S. Tokalov, M. Hoque, D. Scherzinger, F. Marini, A. Poetsch, H. Binder, S. Macher-Goppinger, H. C. Probst, B. Tian, M. Schaefer, K. J. Lackner, F. Westermann, and S. Danckwardt. 2018. Transcriptome 3'end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat Commun* 9: 5331.

70. Zhu, Z. J., P. Huang, Y. X. Chong, L. X. Kang, X. Huang, Z. X. Zhu, and L. Nie. 2016. MicroRNA-181a promotes proliferation and inhibits apoptosis by suppressing CFIm25 in osteosarcoma. *Mol Med Rep* 14: 4271-4278.

71. Huang, J., T. Weng, J. Ko, N. Y. Chen, Y. Xiang, K. Volcik, L. Han, M. R. Blackburn, and X. Lu. 2018. Suppression of cleavage factor Im 25 promotes the proliferation of lung cancer cells through alternative polyadenylation. *Biochem Biophys Res Commun* 503: 856-862.

72. Oh, J. M., C. Di, C. C. Venters, J. Guo, C. Arai, B. R. So, A. M. Pinto, Z. Zhang, L. Wan, I. Younis, and G. Dreyfuss. 2017. U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* 24: 993-999.

73. Kaida, D., M. G. Berg, I. Younis, M. Kasim, L. N. Singh, L. Wan, and G. Dreyfuss. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468: 664-668.

74. Avendano-Vazquez, S. E., A. Dhir, S. Bembich, E. Buratti, N. Proudfoot, and F. E. Baralle. 2012. Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev* 26: 1679-1684.

75. Zhu, H., H. L. Zhou, R. A. Hasman, and H. Lou. 2007. Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J Biol Chem* 282: 2203-2210.

76. Bava, F. A., C. Eliscovich, P. G. Ferreira, B. Minana, C. Ben-Dov, R. Guigo, J. Valcarcel, and R. Mendez. 2013. CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* 495: 121-125.

77. Batra, R., K. Charizanis, M. Manchanda, A. Mohan, M. Li, D. J. Finn, M. Goodwin, C. Zhang, K. Sobczak, C. A. Thornton, and M. S. Swanson. 2014. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol Cell* 56: 311-322.

78. Masuda, A., J. Takeda, T. Okuno, T. Okamoto, B. Ohkawara, M. Ito, S. Ishigaki, G. Sobue, and K. Ohno. 2015. Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev* 29: 1045-1057.

79. Lou, H., R. F. Gagel, and S. M. Berget. 1996. An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev* 10: 208-219.

80. Ji, X., J. Wan, M. Vishnu, Y. Xing, and S. A. Liebhaber. 2013. alphaCP Poly(C) binding proteins act as global regulators of alternative polyadenylation. *Mol Cell Biol* 33: 2560-2573.

81. Gruber, A. J., R. Schmidt, A. R. Gruber, G. Martin, S. Ghosh, M. Belmadani, W. Keller, and M. Zavolan. 2016. A comprehensive analysis of 3' end sequencing

data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res* 26: 1145-1159.

82. Chuvpilo, S., M. Zimmer, A. Kerstan, J. Glockner, A. Avots, C. Escher, C. Fischer, I. Inashkina, E. Jankevics, F. Berberich-Siebelt, E. Schmitt, and E. Serfling. 1999. Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells. *Immunity* 10: 261-269.

83. Gruber, A. J., F. Gypas, A. Riba, R. Schmidt, and M. Zavolan. 2018. Terminal exon characterization with TECtool reveals an abundance of cell-specific isoforms. *Nat Methods* 15: 832-836.

84. de Rooij, L., D. C. H. Chan, A. Keyvani Chahi, and K. J. Hope. 2019. Post-transcriptional regulation in hematopoiesis: RNA binding proteins take control (1). *Biochem Cell Biol* 97: 10-20.

85. Saez, B., M. J. Walter, and T. A. Graubert. 2017. Splicing factor gene mutations in hematologic malignancies. *Blood* 129: 1260-1269.

86. Zipeto, M. A., A. C. Court, A. Sadarangani, N. P. Delos Santos, L. Balaian, H. J. Chun, G. Pineda, S. R. Morris, C. N. Mason, I. Geron, C. Barrett, D. J. Goff, R. Wall, M. Pellecchia, M. Minden, K. A. Frazer, M. A. Marra, L. A. Crews, Q. Jiang, and C. H. M. Jamieson. 2016. ADAR1 Activation Drives Leukemia Stem Cell Self-Renewal by Impairing Let-7 Biogenesis. *Cell Stem Cell* 19: 177-191.

87. Vu, L. P., B. F. Pickering, Y. Cheng, S. Zaccara, D. Nguyen, G. Minuesa, T. Chou, A. Chow, Y. Saletore, M. MacKay, J. Schulman, C. Famulare, M. Patel, V. M. Klimek, F. E. Garrett-Bakelman, A. Melnick, M. Carroll, C. E. Mason, S. R. Jaffrey, and M. G. Kharas. 2017. The N(6)-methyladenosine (m(6)A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat Med* 23: 1369-1376.

88. Guo, S., J. Lu, R. Schlanger, H. Zhang, J. Y. Wang, M. C. Fox, L. E. Purton, H. H. Fleming, B. Cobb, M. Merkenschlager, T. R. Golub, and D. T. Scadden. 2010. MicroRNA miR-125a controls hematopoietic stem cell number. *Proc Natl Acad Sci U S A* 107: 14229-14234.

89. Tenen, D. G. 2003. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* 3: 89-101.

90. Huang, M. E., Y. C. Ye, S. R. Chen, J. R. Chai, J. X. Lu, L. Zhoa, L. J. Gu, and Z. Y. Wang. 1988. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* 72: 567-572.

91. Niu, C., H. Yan, T. Yu, H. P. Sun, J. X. Liu, X. S. Li, W. Wu, F. Q. Zhang, Y. Chen, L. Zhou, J. M. Li, X. Y. Zeng, R. R. Yang, M. M. Yuan, M. Y. Ren, F. Y. Gu, Q. Cao, B. W. Gu, X. Y. Su, G. Q. Chen, S. M. Xiong, T. D. Zhang, S.

Waxman, Z. Y. Wang, Z. Chen, J. Hu, Z. X. Shen, and S. J. Chen. 1999. Studies on treatment of acute promyelocytic leukemia with arsenic trioxide: remission induction, follow-up, and molecular monitoring in 11 newly diagnosed and 47 relapsed acute promyelocytic leukemia patients. *Blood* 94: 3315-3324.

92. Wang, F., J. Travins, B. DeLaBarre, V. Penard-Lacronique, S. Schalm, E. Hansen, K. Straley, A. Kernytsky, W. Liu, C. Gliser, H. Yang, S. Gross, E. Artin, V. Saada, E. Mylonas, C. Quivoron, J. Popovici-Muller, J. O. Saunders, F. G. Salituro, S. Yan, S. Murray, W. Wei, Y. Gao, L. Dang, M. Dorsch, S. Agresta, D. P. Schenkein, S. A. Biller, S. M. Su, S. de Botton, and K. E. Yen. 2013. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* 340: 622-626.

93. Golub, D., N. Iyengar, S. Dogra, T. Wong, D. Bready, K. Tang, A. S. Modrek, and D. G. Placantonakis. 2019. Mutant Isocitrate Dehydrogenase Inhibitors as Targeted Cancer Therapeutics. *Front Oncol* 9: 417.

94. Mugoni, V., R. Panella, G. Cheloni, M. Chen, O. Pozdnyakova, D. Stroopinsky, J. Guarnerio, E. Monteleone, J. D. Lee, L. Mendez, A. V. Menon, J. C. Aster, A. A. Lane, R. M. Stone, I. Galinsky, J. C. Zamora, F. Lo-Coco, M. K. Bhasin, D. Avigan, L. Longo, J. G. Clohessy, and P. P. Pandolfi. 2019. Vulnerabilities in mIDH2 AML confer sensitivity to APL-like targeted combination therapy. *Cell Res* 29: 446-459.

95. Hoque, M., Z. Ji, D. Zheng, W. Luo, W. Li, B. You, J. Y. Park, G. Yehia, and B. Tian. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* 10: 133-139.

96. Zheng, D., X. Liu, and B. Tian. 2016. 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. *Rna* 22: 1631-1639.

97. Akman, B. H., T. Can, and A. E. Erson-Bensan. 2012. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res* 40: 10679-10688.

98. Langer, C., M. D. Radmacher, A. S. Ruppert, S. P. Whitman, P. Paschka, K. Mrozek, C. D. Baldus, T. Vukosavljevic, C. G. Liu, M. E. Ross, B. L. Powell, A. de la Chapelle, J. E. Kolitz, R. A. Larson, G. Marcucci, and C. D. Bloomfield. 2008. High BAALC expression associates with other molecular prognostic markers, poor outcome, and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. *Blood* 111: 5371-5379.

99. Damiani, D., M. Tiribelli, A. Franzoni, A. Michelutti, D. Fabbro, M. Cavallin, E. Toffoletti, E. Simeone, R. Fanin, and G. Damante. 2013. BAALC overexpression retains its negative prognostic role across all cytogenetic risk groups in acute myeloid leukemia patients. *Am J Hematol* 88: 848-852.

100. Ray, D., S. Y. Kwon, H. Tagoh, O. Heidenreich, A. Ptasinska, and C. Bonifer. 2013. Lineage-inappropriate PAX5 expression in t(8;21) acute myeloid leukemia requires signaling-mediated abrogation of polycomb repression. *Blood* 122: 759-769.

101. Bacher, U., T. Haferlach, C. Schoch, W. Kern, and S. Schnittger. 2006. Implications of NRAS mutations in AML: a study of 2502 patients. *Blood* 107: 3847-3853.

102. Li, B., F. C. Yang, D. W. Clapp, and K. T. Chun. 2003. Enforced expression of CUL-4A interferes with granulocytic differentiation and exit from the cell cycle. *Blood* 101: 1769-1776.

103. Li, B., N. Jia, R. Kapur, and K. T. Chun. 2006. Cul4A targets p27 for degradation and regulates proliferation, cell cycle exit, and differentiation during erythropoiesis. *Blood* 107: 4291-4299.

104. Naudin, C., A. Hattabi, F. Michelet, A. Miri-Nezhad, A. Benyoucef, F. Pflumio, F. Guillonneau, S. Fichelson, I. Vigon, I. Dusanter-Fourt, and E. Lauret. 2017. PUMILIO/FOXP1 signaling drives expansion of hematopoietic stem/progenitor and leukemia cells. *Blood* 129: 2493-2506.

105. Lin, Y., Z. Li, F. Ozsolak, S. W. Kim, G. Arango-Argoty, T. T. Liu, S. A. Tenenbaum, T. Bailey, A. P. Monaghan, P. M. Milos, and B. John. 2012. An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res* 40: 8460-8471.

106. Gupta, D., H. P. Shah, K. Malu, N. Berliner, and P. Gaines. 2014. Differentiation and characterization of myeloid cells. *Curr Protoc Immunol* 104: 22f.25.21-22f.25.28.

107. Draber, P., O. Stepanek, M. Hrdinka, A. Drobek, L. Chmatal, L. Mala, T. Ormsby, P. Angelisova, V. Horejsi, and T. Brdicka. 2012. LST1/A is a myeloid leukocyte-specific transmembrane adaptor protein recruiting protein tyrosine phosphatases SHP-1 and SHP-2 to the plasma membrane. *J Biol Chem* 287: 22812-22821.

108. Raffel, S., M. Falcone, N. Kneisel, J. Hansson, W. Wang, C. Lutz, L. Bullinger, G. Poschet, Y. Nonnenmacher, A. Barnert, C. Bahr, P. Zeisberger, A. Przybylla, M. Sohn, M. Tonjes, A. Erez, L. Adler, P. Jensen, C. Scholl, S. Frohling, S. Cocciardi, P. Wuchter, C. Thiede, A. Florcken, J. Westermann, G. Ehninger, P. Lichter, K. Hiller, R. Hell, C. Herrmann, A. D. Ho, J. Krijgsveld, B. Radlwimmer, and A. Trumpp. 2017. BCAT1 restricts alphaKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature* 551: 384-388.

109. Gu, Z., Y. Liu, F. Cai, M. Patrick, J. Zmajkovic, H. Cao, Y. Zhang, A. Tasdogan, M. Chen, L. Qi, X. Liu, K. Li, J. Lyu, K. E. Dickerson, W. Chen, M. Ni, M. E. Merritt, S. J. Morrison, R. C. Skoda, R. J. DeBerardinis, and J. Xu. 2019. Loss of EZH2 Reprograms BCAA Metabolism to Drive Leukemic Transformation. *Cancer Discov* 9: 1228-1247.

110. Somervaille, T. C., C. J. Matheny, G. J. Spencer, M. Iwasaki, J. L. Rinn, D. M. Witten, H. Y. Chang, S. A. Shurtleff, J. R. Downing, and M. L. Cleary. 2009. Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* 4: 129-140.

111. Burel, S. A., N. Harakawa, L. Zhou, T. Pabst, D. G. Tenen, and D. E. Zhang. 2001. Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. *Mol Cell Biol* 21: 5577-5590.

112. de Guzman, C. G., A. J. Warren, Z. Zhang, L. Gartland, P. Erickson, H. Drabkin, S. W. Hiebert, and C. A. Klug. 2002. Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. *Mol Cell Biol* 22: 5506-5517.

113. Zaidi, S. K., A. W. Perez, E. S. White, J. B. Lian, J. L. Stein, and G. S. Stein. 2017. An AML1-ETO/miR-29b-1 regulatory circuit modulates phenotypic properties of acute myeloid leukemia cells. *Oncotarget* 8: 39994-40005.

114. Fu, L., J. Shi, A. Liu, L. Zhou, M. Jiang, H. Fu, K. Xu, D. Li, A. Deng, Q. Zhang, Y. Pang, Y. Guo, K. Hu, J. Zhou, Y. Wang, W. Huang, Y. Jing, L. Dou, L. Wang, X. Ke, C. Nervi, Y. Li, and L. Yu. 2017. A minicircuitry of microRNA-9-1 and RUNX1-RUNX1T1 contributes to leukemogenesis in t(8;21) acute myeloid leukemia. *Int J Cancer* 140: 653-661.

115. Ley, T. J., C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, and G. Eley. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368: 2059-2074.

116. Tyner, J. W., C. E. Tognon, D. Bottomly, B. Wilmot, S. E. Kurtz, S. L. Savage, N. Long, A. R. Schultz, E. Traer, M. Abel, A. Agarwal, A. Blucher, U. Borate, J. Bryant, R. Burke, A. Carlos, R. Carpenter, J. Carroll, B. H. Chang, C. Coblentz, A. d'Almeida, R. Cook, A. Danilov, K. T. Dao, M. Degnin, D. Devine, J. Dibb, D.

K. t. Edwards, C. A. Eide, I. English, J. Glover, R. Henson, H. Ho, A. Jemal, K. Johnson, R. Johnson, B. Junio, A. Kaempf, J. Leonard, C. Lin, S. Q. Liu, P. Lo, M. M. Loriaux, S. Luty, T. Macey, J. MacManiman, J. Martinez, M. Mori, D. Nelson, C. Nichols, J. Peters, J. Ramsdill, A. Rofelty, R. Schuff, R. Searles, E. Segerdell, R. L. Smith, S. E. Spurgeon, T. Sweeney, A. Thapa, C. Visser, J. Wagner, K. Watanabe-Smith, K. Werth, J. Wolf, L. White, A. Yates, H. Zhang, C. R. Cogle, R. H. Collins, D. C. Connolly, M. W. Deininger, L. Drusbosky, C. S. Hourigan, C. T. Jordan, P. Kropf, T. L. Lin, M. E. Martinez, B. C. Medeiros, R. R. Pallapati, D. A. Pollyea, R. T. Swords, J. M. Watts, S. J. Weir, D. L. Wiest, R. M. Winters, S. K. McWeeney, and B. J. Druker. 2018. Functional genomic landscape of acute myeloid leukaemia. *Nature* 562: 526-531.

117. Yilmaz, O. H., R. Valdez, B. K. Theisen, W. Guo, D. O. Ferguson, H. Wu, and S. J. Morrison. 2006. Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* 441: 475-482.

118. Hoshii, T., Y. Tadokoro, K. Naka, T. Ooshio, T. Muraguchi, N. Sugiyama, T. Soga, K. Araki, K. Yamamura, and A. Hirao. 2012. mTORC1 is essential for leukemia propagation but not stem cell self-renewal. *J Clin Invest* 122: 2114-2129.

119. Martelli, A. M., C. Evangelisti, F. Chiarini, and J. A. McCubrey. 2010. The phosphatidylinositol 3-kinase/Akt/mTOR signaling network as a therapeutic target in acute myelogenous leukemia patients. *Oncotarget* 1: 89-103.

120. Choo, A. Y., S. O. Yoon, S. G. Kim, P. P. Roux, and J. Blenis. 2008. Rapamycin differentially inhibits S6Ks and 4E-BP1 to mediate cell-type-specific repression of mRNA translation. *Proc Natl Acad Sci U S A* 105: 17414-17419.

121. Jiang, Y. P., L. M. Ballou, and R. Z. Lin. 2001. Rapamycin-insensitive regulation of 4e-BP1 in regenerating rat liver. *J Biol Chem* 276: 10943-10951.

122. Wall, M., G. Poortinga, K. M. Hannan, R. B. Pearson, R. D. Hannan, and G. A. McArthur. 2008. Translational control of c-MYC by rapamycin promotes terminal myeloid differentiation. *Blood* 112: 2305-2317.

123. Wiestner, A., M. Tehrani, M. Chiorazzi, G. Wright, F. Gibellini, K. Nakayama, H. Liu, A. Rosenwald, H. K. Muller-Hermelink, G. Ott, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. Vose, J. O. Armitage, R. D. Gascoyne, J. M. Connors, E. Campo, E. Montserrat, F. Bosch, E. B. Smeland, S. Kvaloy, H. Holte, J. Delabie, R. I. Fisher, T. M. Grogan, T. P. Miller, W. H. Wilson, E. S. Jaffe, and L. M. Staudt. 2007. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* 109: 4599-4606.

124. Morita, K., Y. Masamoto, K. Kataoka, J. Koya, Y. Kagoya, H. Yashiroda, T. Sato, S. Murata, and M. Kurokawa. 2015. BAALC potentiates oncogenic ERK pathway through interactions with MEKK1 and KLF4. *Leukemia* 29: 2248-2256.

125. Dannenmann, Benjamin, Klimiankou, Maksim, Solovyeva, Anna, Zahabi, Azadeh, Mir, Perihan, Nasri, Masoud, Bernhard, Regine, Lindner, Christian, Morishima, Tatsuya, Steinemann, Doris, Zeidler, Cornelia, Kanz, Lothar, Welte, Karl, Skokowa, and Julia. 2017. *BAALC* is a Key Mediator of Leukemia Development in Congenital Neutropenia. *Blood*. 541.

126. Leon, J., N. Ferrandiz, J. C. Acosta, and M. D. Delgado. 2009. Inhibition of cell differentiation: a critical mechanism for MYC-mediated carcinogenesis? *Cell Cycle* 8: 1148-1157.

127. Delgado, M. D., and J. Leon. 2010. Myc roles in hematopoiesis and leukemia. *Genes Cancer* 1: 605-616.

128. Gera, J. F., I. K. Mellinghoff, Y. Shi, M. B. Rettig, C. Tran, J. H. Hsu, C. L. Sawyers, and A. K. Lichtenstein. 2004. AKT activity determines sensitivity to mammalian target of rapamycin (mTOR) inhibitors by regulating cyclin D1 and c-myc expression. *J Biol Chem* 279: 2737-2746.

129. Yue, M., J. Jiang, P. Gao, H. Liu, and G. Qing. 2017. Oncogenic MYC Activates a Feedforward Regulatory Loop Promoting Essential Amino Acid Metabolism and Tumorigenesis. *Cell Rep* 21: 3819-3832.

130. Liu, P., M. Ge, J. Hu, X. Li, L. Che, K. Sun, L. Cheng, Y. Huang, M. G. Pilo, A. Cigliano, G. M. Pes, R. M. Pascale, S. Brozzetti, G. Vidili, A. Porcu, A. Cossu, G. Palmieri, M. C. Sini, S. Ribback, F. Dombrowski, J. Tao, D. F. Calvisi, L. Chen, and X. Chen. 2017. A functional mammalian target of rapamycin complex 1 signaling is indispensable for c-Myc-driven hepatocarcinogenesis. *Hepatology* 66: 167-181.

131. Zhao, X., A. P. Petrashen, J. A. Sanders, A. L. Peterson, and J. M. Sedivy. 2019. SLC1A5 glutamine transporter is a target of MYC and mediates reduced mTORC1 signaling and increased fatty acid oxidation in long-lived Myc hypomorphic mice. *Aging Cell* 18: e12947.

132. Chang, J. W., W. Zhang, H. S. Yeh, E. P. de Jong, S. Jun, K. H. Kim, S. S. Bae, K. Beckman, T. H. Hwang, K. S. Kim, D. H. Kim, T. J. Griffin, R. Kuang, and J. Yong. 2015. mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nat Commun* 6: 7218.

133. Nowak, D., D. Stewart, and H. P. Koeffler. 2009. Differentiation therapy of leukemia: 3 decades of development. *Blood* 113: 3655-3665.

134. Cools, J., D. J. DeAngelo, J. Gotlib, E. H. Stover, R. D. Legare, J. Cortes, J. Kutok, J. Clark, I. Galinsky, J. D. Griffin, N. C. Cross, A. Tefferi, J. Malone, R. Alam, S. L. Schrier, J. Schmid, M. Rose, P. Vandenberghe, G. Verhoef, M. Boogaerts, I. Wlodarska, H. Kantarjian, P. Marynen, S. E. Coutre, R. Stone, and D. G. Gilliland. 2003. A tyrosine kinase created by fusion of the PDGFRA and

FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med* 348: 1201-1214.

135. Buijs, A., and M. Bruin. 2007. Fusion of FIP1L1 and RARA as a result of a novel t(4;17)(q12;q21) in a case of juvenile myelomonocytic leukemia. In *Leukemia*, England. 1104-1108.

136. Kondo, T., A. Mori, S. Darmanin, S. Hashino, J. Tanaka, and M. Asaka. 2008. The seventh pathogenic fusion gene FIP1L1-RARA was isolated from a t(4;17)-positive acute promyelocytic leukemia. In *Haematologica*, Italy. 1414-1416.

137. Iwasaki, J., T. Kondo, S. Darmanin, M. Ibata, M. Onozawa, D. Hashimoto, N. Sakamoto, and T. Teshima. 2014. FIP1L1 presence in FIP1L1-RARA or FIP1L1-PDGFRA differentially contributes to the pathogenesis of distinct types of leukemia. *Ann Hematol* 93: 1473-1481.

138. Stover, E. H., J. Chen, C. Folens, B. H. Lee, N. Mentens, P. Marynen, I. R. Williams, D. G. Gilliland, and J. Cools. 2006. Activation of FIP1L1-PDGFRalpha requires disruption of the juxtamembrane domain of PDGFRalpha and is FIP1L1-independent. *Proc Natl Acad Sci U S A* 103: 8078-8083.

139. Vanden Bempt, M., S. Demeyer, N. Mentens, E. Geerdens, C. E. De Bock, I. Wlodarska, and J. Cools. 2016. Generation of the Fip1l1-Pdgfra fusion gene using CRISPR/Cas genome editing. In *Leukemia*, England. 1913-1916.

140. Buitenhuis, M., L. P. Verhagen, J. Cools, and P. J. Coffer. 2007. Molecular mechanisms underlying FIP1L1-PDGFRA-mediated myeloproliferation. *Cancer Res* 67: 3759-3766.

141. Van Nostrand, E. L., G. A. Pratt, A. A. Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman, and G. W. Yeo. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13: 508-514.

142. Yan, M., E. Kanbe, L. F. Peterson, A. Boyapati, Y. Miao, Y. Wang, I. M. Chen, Z. Chen, J. D. Rowley, C. L. Willman, and D. E. Zhang. 2006. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med* 12: 945-949.

143. Daley, G. Q., R. A. Van Etten, and D. Baltimore. 1990. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* 247: 824-830.

144. Bagger, F. O., S. Kinalis, and N. Rapin. 2019. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res* 47: D881-d885.

145. Kim, D., B. Langmead, and S. L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12: 357-360.

146. Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550.

147. Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21.

148. Patro, R., S. M. Mount, and C. Kingsford. 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* 32: 462-464.

149. Mootha, V. K., C. M. Lindgren, K. F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267-273.

150. Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. 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.

151. Zhou, Y., B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, and S. K. Chanda. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523.

152. Yoshida, K., M. Sanada, Y. Shiraishi, D. Nowak, Y. Nagata, R. Yamamoto, Y. Sato, A. Sato-Otsubo, A. Kon, M. Nagasaki, G. Chalkidis, Y. Suzuki, M. Shiosaka, R. Kawahata, T. Yamaguchi, M. Otsu, N. Obara, M. Sakata-Yanagimoto, K. Ishiyama, H. Mori, F. Nolte, W. K. Hofmann, S. Miyawaki, S. Sugano, C. Haferlach, H. P. Koeffler, L. Y. Shih, T. Haferlach, S. Chiba, H. Nakauchi, S. Miyano, and S. Ogawa. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478: 64-69.

153. Graubert, T. A., D. Shen, L. Ding, T. Okeyo-Owuor, C. L. Lunn, J. Shao, K. Krysiak, C. C. Harris, D. C. Koboldt, D. E. Larson, M. D. McLellan, D. J. Dooling, R. M. Abbott, R. S. Fulton, H. Schmidt, J. Kalicki-Veizer, M. O'Laughlin, M. Grillot, J. Baty, S. Heath, J. L. Frater, T. Nasim, D. C. Link, M. H. Tomasson, P. Westervelt, J. F. DiPersio, E. R. Mardis, T. J. Ley, R. K. Wilson, and M. J. Walter. 2011. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* 44: 53-57.

154. Ilagan, J. O., A. Ramakrishnan, B. Hayes, M. E. Murphy, A. S. Zebari, P. Bradley, and R. K. Bradley. 2015. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res* 25: 14-26.

155. Kim, E., J. O. Ilagan, Y. Liang, G. M. Daubner, S. C. Lee, A. Ramakrishnan, Y. Li, Y. R. Chung, J. B. Micol, M. E. Murphy, H. Cho, M. K. Kim, A. S. Zebari, S. Aumann, C. Y. Park, S. Buonamici, P. G. Smith, H. J. Deeg, C. Lobry, I. Aifantis, Y. Modis, F. H. Allain, S. Halene, R. K. Bradley, and O. Abdel-Wahab. 2015. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* 27: 617-630.

156. XuFeng, R., M. J. Boyer, H. Shen, Y. Li, H. Yu, Y. Gao, Q. Yang, Q. Wang, and T. Cheng. 2009. ADAR1 is required for hematopoietic progenitor cell survival via RNA editing. *Proc Natl Acad Sci U S A* 106: 17763-17768.

157. Park, S. M., M. Gonen, L. Vu, G. Minuesa, P. Tivnan, T. S. Barlowe, J. Taggart, Y. Lu, R. P. Deering, N. Hacohen, M. E. Figueroa, E. Paietta, H. F. Fernandez, M. S. Tallman, A. Melnick, R. Levine, C. Leslie, C. J. Lengner, and M. G. Kharas. 2015. Musashi2 sustains the mixed-lineage leukemia-driven stem cell regulatory program. *J Clin Invest* 125: 1286-1298.

158. Halsey, C., M. Docherty, M. McNeill, D. Gilchrist, M. Le Brocq, B. Gibson, and G. Graham. 2012. The GATA1s isoform is normally down-regulated during terminal haematopoietic differentiation and over-expression leads to failure to repress MYB, CCND2 and SKI during erythroid differentiation of K562 cells. *J Hematol Oncol* 5: 45.

159. Sun, L., N. Heerema, L. Crotty, X. Wu, C. Navara, A. Vassilev, M. Sensel, G. H. Reaman, and F. M. Uckun. 1999. Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* 96: 680-685.

160. Tsuzuki, S., D. Hong, R. Gupta, K. Matsuo, M. Seto, and T. Enver. 2007. Isoform-specific potentiation of stem and progenitor cell engraftment by AML1/RUNX1. In *PLoS Med*, United States. e172.

161. Calkhoven, C. F., C. Muller, R. Martin, G. Krosl, H. Pietsch, T. Hoang, and A. Leutz. 2003. Translational control of SCL-isoform expression in hematopoietic lineage choice. *Genes Dev* 17: 959-964.

162. Sasaki, K., Y. Nakamura, K. Maki, K. Waga, F. Nakamura, H. Arai, Y. Imai, H. Hirai, and K. Mitani. 2004. Functional analysis of a dominant-negative DeltaETS TEL/ETV6 isoform. *Biochem Biophys Res Commun* 317: 1128-1137.

163. Miyoshi, H., M. Ohira, K. Shimizu, K. Mitani, H. Hirai, T. Imai, K. Yokoyama, E. Soeda, and M. Ohki. 1995. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. In *Nucleic Acids Res*, England. 2762-2769.

164. Tanaka, T., K. Tanaka, S. Ogawa, M. Kurokawa, K. Mitani, J. Nishida, Y. Shibata, Y. Yazaki, and H. Hirai. 1995. An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. *EMBO J* 14: 341-350.

165. Ghozi, M. C., Y. Bernstein, V. Negreanu, D. Levanon, and Y. Groner. 1996. Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. *Proc Natl Acad Sci U S A* 93: 1935-1940.

166. Lam, K., and D. Zhang. 2012. RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis. *Frontiers in Bioscience-Landmark* 17: 1120-1139.

167. Miyoshi, H., K. Shimizu, T. Kozu, N. Maseki, Y. Kaneko, and M. Ohki. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc Natl Acad Sci U S A* 88: 10431-10434.

168. Gu, T. L., T. L. Goetz, B. J. Graves, and N. A. Speck. 2000. Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol Cell Biol* 20: 91-103.

169. Liu, X., Q. Zhang, D. E. Zhang, C. Zhou, H. Xing, Z. Tian, Q. Rao, M. Wang, and J. Wang. 2009. Overexpression of an isoform of AML1 in acute leukemia and its potential role in leukemogenesis. In *Leukemia*, England. 739-745.

170. Guo, H., O. Ma, N. A. Speck, and A. D. Friedman. 2012. Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood* 119: 4408-4418.

171. Tsuzuki, S., and M. Seto. 2012. Expansion of functionally defined mouse hematopoietic stem and progenitor cells by a short isoform of RUNX1/AML1. In *Blood*, United States. 727-735.

172. Ran, D., W. J. Shia, M. C. Lo, J. B. Fan, D. A. Knorr, P. I. Ferrell, Z. Ye, M. Yan, L. Cheng, D. S. Kaufman, and D. E. Zhang. 2013. RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. In *Blood*, United States. 2882-2890.

173. Challen, G. A., and M. A. Goodell. 2010. Runx1 Isoforms Show Differential Expression Patterns During Hematopoietic Development But Have Similar Functional Effects in Adult Hematopoietic Stem Cells. *Exp Hematol* 38: 403-416.

174. Goyama, S., J. Schibler, L. Cunningham, Y. Zhang, Y. Rao, N. Nishimoto, M. Nakagawa, A. Olsson, M. Wunderlich, K. A. Link, B. Mizukawa, H. L. Grimes, M. Kurokawa, P. P. Liu, G. Huang, and J. C. Mulloy. 2013. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest* 123: 3876-3888.

175. Komeno, Y., M. Yan, S. Matsuura, K. Lam, M. Lo, Y. Huang, D. Tenen, J. Downing, and D. Zhang. 2014. Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood* 123: 3760-3769.

176. Ng, C. E., M. Osato, B. H. Tay, B. Venkatesh, and Y. Ito. 2007. cDNA cloning of Runx family genes from the pufferfish (Fugu rubripes). In *Gene*, Netherlands. 162-173.

177. Tsuji, K., and M. Noda. 2000. Identification and expression of a novel 3'-exon of mouse Runx1/Pebp2alphaB/Cbfa2/AML1 gene. In *Biochem Biophys Res Commun*. 2000 Academic Press., United States. 171-176.

178. Osato, M. 2014. An unsung runt 6e isoform for HSC expansion.

179. Sakurai, H., Y. Harada, Y. Ogata, Y. Kagiyama, N. Shingai, N. Doki, K. Ohashi, T. Kitamura, N. Komatsu, and H. Harada. 2017. Overexpression of RUNX1 short isoform has an important role in the development of myelodysplastic/myeloproliferative neoplasms. *Blood Adv* 1: 1382-1386.

180. Brown, K. M., and G. M. Gilmartin. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* 12: 1467-1476.

181. Perez Canadillas, J. M., and G. Varani. 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *Embo j* 22: 2821-2830.

182. Hall-Pogar, T., H. Zhang, B. Tian, and C. S. Lutz. 2005. Alternative polyadenylation of cyclooxygenase-2. *Nucleic Acids Res* 33: 2565-2579.

183. Wang, Y., and Z. Wang. 2014. Systematical identification of splicing regulatory cis-elements and cognate trans-factors. *Methods* 65: 350-358.

184. Wang, Z., M. E. Rolish, G. Yeo, V. Tung, M. Mawson, and C. B. Burge. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* 119: 831-845.

185. Wheeler, E. C., A. Q. Vu, M. DiSalvo, J. M. Einstein, N. Ahmed, E. L. Van Nostrand, A. A. Shishkin, W. Jin, N. L. Allbritton, and G. W. Yeo. Pooled CRISPR/Cas9 screening with image-based phenotyping on microRaft arrays reveals stress granule-regulatory factors. *Nature Methods* In submission.

186. Li, W., J. Koster, H. Xu, C. H. Chen, T. Xiao, J. S. Liu, M. Brown, and X. S. Liu. 2015. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* 16: 281.

187. Burd, C. G., and G. Dreyfuss. 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo j* 13: 1197-1204.

188. Huelga, S. C., A. Q. Vu, J. D. Arnold, T. Y. Liang, P. P. Liu, B. Y. Yan, J. P. Donohue, L. Shiue, S. Hoon, S. Brenner, M. Ares, Jr., and G. W. Yeo. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* 1: 167-178.

189. Bruun, G. H., T. K. Doktor, J. Borch-Jensen, A. Masuda, A. R. Krainer, K. Ohno, and B. S. Andresen. 2016. Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biol* 14: 54.

190. Morgan, C. E., J. L. Meagher, J. D. Levengood, J. Delproposto, C. Rollins, J. A. Stuckey, and B. S. Tolbert. 2015. The First Crystal Structure of the UP1 Domain of hnRNP A1 Bound to RNA Reveals a New Look for an Old RNA Binding Protein. *J Mol Biol* 427: 3241-3257.

191. Rollins, C., J. D. Levengood, B. D. Rife, M. Salemi, and B. S. Tolbert. 2014. Thermodynamic and phylogenetic insights into hnRNP A1 recognition of the HIV-1 exon splicing silencer 3 element. *Biochemistry* 53: 2172-2184.

192. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576-589.

193. Jean-Philippe, J., S. Paz, and M. Caputi. 2013. hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci* 14: 18999-19024.

194. Song, L., H. S. Lin, J. N. Gong, H. Han, X. S. Wang, R. Su, M. T. Chen, C. Shen, Y. N. Ma, J. Yu, and J. W. Zhang. 2017. microRNA-451-modulated hnRNP A1 takes a part in granulocytic differentiation regulation and acute myeloid leukemia. *Oncotarget* 8: 55453-55466.

195. Iervolino, A., G. Santilli, R. Trotta, C. Guerzoni, V. Cesi, A. Bergamaschi, C. Gambacorti-Passerini, B. Calabretta, and D. Perrotti. 2002. hnRNP A1 nucleocytoplasmic shuttling activity is required for normal myelopoiesis and BCR/ABL leukemogenesis. *Mol Cell Biol* 22: 2255-2266.

196. Eiring, A. M., P. Neviani, R. Santhanam, J. J. Oaks, J. S. Chang, M. Notari, W. Willis, C. Gambacorti-Passerini, S. Volinia, G. Marcucci, M. A. Caligiuri, G. W. Leone, and D. Perrotti. 2008. Identification of novel posttranscriptional targets of the BCR/ABL oncoprotein by ribonomics: requirement of E2F3 for BCR/ABL leukemogenesis. *Blood* 111: 816-828.

197. Roy, R., Y. Huang, M. J. Seckl, and O. E. Pardo. 2017. Emerging roles of hnRNPA1 in modulating malignant transformation. *Wiley Interdiscip Rev RNA* 8.

198. Fang, J., L. C. Bolanos, K. Choi, X. Liu, S. Christie, S. Akunuru, R. Kumar, D. Wang, X. Chen, K. D. Greis, P. Stoilov, M. D. Filippi, J. P. Maciejewski, G. Garcia-Manero, M. T. Weirauch, N. Salomonis, H. Geiger, Y. Zheng, and D. T.

Starczynowski. 2017. Ubiquitination of hnRNPA1 by TRAF6 links chronic innate immune signaling with myelodysplasia. *Nat Immunol* 18: 236-245.

199. Vernet, C., and K. Artzt. 1997. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet* 13: 479-484.

200. Paronetto, M. P., T. Achsel, A. Massiello, C. E. Chalfant, and C. Sette. 2007. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J Cell Biol* 176: 929-939.

201. Pedrotti, S., P. Bielli, M. P. Paronetto, F. Ciccosanti, G. M. Fimia, S. Stamm, J. L. Manley, and C. Sette. 2010. The splicing regulator Sam68 binds to a novel exonic splicing silencer and functions in SMN2 alternative splicing in spinal muscular atrophy. *Embo j* 29: 1235-1247.

202. Singh, R., and J. Valcarcel. 2005. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* 12: 645-653.

203. La Rosa, P., P. Bielli, C. Compagnucci, E. Cesari, E. Volpe, S. Farioli Vecchioli, and C. Sette. 2016. Sam68 promotes self-renewal and glycolytic metabolism in mouse neural progenitor cells by modulating Aldh1a3 pre-mRNA 3'-end processing. *Elife* 5.

204. Frisone, P., D. Pradella, A. Di Matteo, E. Belloni, C. Ghigna, and M. P. Paronetto. 2015. SAM68: Signal Transduction and RNA Metabolism in Human Cancer. *Biomed Res Int* 2015: 528954.

205. Bielli, P., R. Busa, M. P. Paronetto, and C. Sette. 2011. The RNA-binding protein Sam68 is a multifunctional player in human cancer. *Endocr Relat Cancer* 18: R91-r102.

206. Cheung, N., L. C. Chan, A. Thompson, M. L. Cleary, and C. W. So. 2007. Protein arginine-methyltransferase-dependent oncogenesis. *Nat Cell Biol* 9: 1208-1215.

207. Wang, Q., Y. Li, J. Cheng, L. Chen, H. Xu, Q. Li, and T. Pang. 2016. Sam68 affects cell proliferation and apoptosis of human adult T-acute lymphoblastic leukemia cells via AKT/mTOR signal pathway. *Leuk Res* 46: 1-9.

208. Benoit, Y. D., R. R. Mitchell, R. M. Risueno, L. Orlando, B. Tanasijevic, A. L. Boyd, L. Aslostovar, K. R. Salci, Z. Shapovalova, J. Russell, M. Eguchi, D. Golubeva, M. Graham, A. Xenocostas, M. R. Trus, R. Foley, B. Leber, T. J. Collins, and M. Bhatia. 2017. Sam68 Allows Selective Targeting of Human Cancer Stem Cells. *Cell Chem Biol* 24: 833-844.e839.

209. Shepard, P. J., E. A. Choi, J. Lu, L. A. Flanagan, K. J. Hertel, and Y. Shi. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *Rna* 17: 761-772.

210. Naro, C., L. Pellegrini, A. Jolly, D. Farini, E. Cesari, P. Bielli, P. de la Grange, and C. Sette. 2019. Functional Interaction between U1snRNP and Sam68 Insures Proper 3' End Pre-mRNA Processing during Germ Cell Differentiation. *Cell Rep* 26: 2929-2941.e2925.

211. Iijima, Y., M. Tanaka, S. Suzuki, D. Hauser, C. Okada, M. Ito, N. Ayukawa, Y. Sato, M. Ohtsuka, P. Scheiffele, and T. Iijima. 2019. SAM68-Specific Splicing Is Required for Proper Selection of Alternative 3' UTR Isoforms in the Nervous System. *iScience* 22: 318-335.

212. Shi, Y., P. Frost, B. Hoang, A. Benavides, J. Gera, and A. Lichtenstein. 2011. IL-6-induced enhancement of c-Myc translation in multiple myeloma cells: critical role of cytoplasmic localization of the rna-binding protein hnRNP A1. *J Biol Chem* 286: 67-78.

213. Shi, Y., Y. Yang, B. Hoang, C. Bardeleben, B. Holmes, J. Gera, and A. Lichtenstein. 2016. Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. *Oncogene* 35: 1015-1024.

214. Ramirez, F., D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dundar, and T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160-165.

215. *Martin, M.* 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet.journal. 10-12.