REGULATION AND FUNCTION OF THE RNA-BINDING PROTEIN CELF1

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

DANIEL JOSEPH BEISANG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PAUL R. BOHJANEN MD, PhD

July, 2012

Acknowledgements

I would like thank many people without whom the completion of this thesis would not have been possible. Specifically, I would like to acknowledge the members of my thesis committee: Peter Bitterman, MD, Steve Rice, PhD, Cavan Reilly, PhD, George Karypis, PhD, and Paul Bohjanen, MD PhD for their insight and guidance. Additionally, I appreciate the help and support I have received from my advisor, Paul Bohjanen, throughout my graduate program. It has been a privilege to learn from his insight. Additionally, I thank past and present members of the Bohjanen lab; particularly Bernd Rattenbacher, whose knowledge and expertise served as a springboard for my entry into RNA biology.

I would like to thank the University of Minnesota Medical Scientist Training Program (MSTP) and the Microbiology, Immunology, and Cancer Biology Graduate program for their support. Finally I am grateful for the financial support I have received from the National Institute of General Medical Sciences and National Heart Lung and Blood Institute.

Dedication

This dissertation is dedicated to my wife, Jolene, and our son, William.

Abstract

Post-transcriptional gene regulation is a crucial aspect of the cellular control mechanisms that ensure the accurate expression of genetic information. Coordinate regulation of the rate of decay of networks of functionally related mRNA provides a cell the ability to quickly and precisely alter its transcriptome in response to environmental signals. An important regulator of mRNA decay is the RNA binding protein CUGBP1 and Etr3 Like Factor 1 (CELF1). CELF1 binds to a GU-rich element (GRE) harbored in target transcripts and promotes their rapid decay through the recruitment of RNA decay enzymes. The work presented in this thesis focused on elucidating the function and regulation of CELF1 mediated mRNA decay through the following objectives:

The first objective was to elucidate the network of transcripts that are regulated by CELF1. We utilized RNA-immunoprecipitation targeting CELF1 followed by microarray analysis of the co-precipitated transcripts to determine that CELF1 targets in HeLa cells and primary human T cells. We found that CELF1 targeted a network of approximately 600 transcripts in HeLa cells, 1300 transcripts in resting T cells, and 150 transcripts in stimulated T cells. Functional analysis of these transcripts revealed that CELF1 transcripts code for proteins involved in cellular proliferation, apoptosis, actin-based motility, and post-transcriptional regulation. Bioinformatic analysis of the CELF1 target transcripts revealed enrichment of the previously defined GRE, as well as a GU-repeat motif, in the 3'UTR of transcripts. We further showed that this GU-repeat motif conferred CELF1-mediated rapid decay to otherwise stable reporter transcripts, and thus re-defined the GRE as a UGU[G/U]UGU[G/U]UGU sequence occurring in the context of

the 3'UTR. Through investigation of the decrease in the CELF1 target transcript population, we found that CELF1 underwent an activation dependent phosphorylation event. We found that this phosphorylation event inhibited CELF1 mRNA binding, and correlated with the increased half-life and abundance of CELF1 resting cell targets during T cell activation.

The second objective of this thesis was to investigate whether aberrant regulation of CELF1 mediated mRNA decay may play a role in viral oncogenesis. To investigate this, we utilized a model of KSHV infection whereby HeLa cells were transfected with the oncogenic KSHV encoded GPCR. We found that the vGPCR caused CELF1 phosphorylation, likely in a MEK/ERK dependent fashion. We showed that this phosphorylation correlated with reversal of CELF1 mediated mRNA decay, while maintaining CELF1s ability to bind to the GRE. Investigation into the mechanism of CELF1 inhibition by the vGPCR revealed that phosphorylation of CELF1 through vGPCR signaling may inhibit CELF1s recruitment of the deadenylase PARN, thus inhibiting CELF1 mRNA decay. The inhibition of CELF1 mediated decay is expected to result in the stabilization of a network of proliferation and apoptosis regulatory transcripts, leading to dysregulation of these pathways and potentially contributing to an oncogenic phenotype.

Finally, we studied the affect of alternative polyadenylation (APA) during the early timepoints of T cell stimulation on mRNA decay networks. We utilized high-throughput sequencing technologies to globally quantify the APA landscape at zero, six, and 24 hours of in vitro T cell stimulation. Using this data we were able to confirm

previous reports that T cell stimulation promotes preferential usage of proximal polyA sites, but our results suggest that this occurs much earlier than previously reported. Additionally, we present data suggesting that GREs and AU-rich elements (AREs) are much more likely to be excised than included in a 3'UTR as a result of APA, suggesting preferential regulation of these mRNA decay networks. Finally, we present data suggesting that the temporal pattern of GRE and ARE regulation by APA is different, suggesting functional independence in the ARE and GRE containing networks of transcripts.

Table of Contents

List of Tables

List of Figures

Abbreviations

3'UTR	3' Untranslated Region
5'UTR	5' Untranslated Region
APA	Alternative Polyadenylation
ARE	AU-Rich Element
AREBP	AU-Rich Element Binding Protein
AUF1	AU Rich RNA Binding Factor 1
BRE	Bruno Response Element
BRF	Butyrate Response Factor
Ccr4	Carbon catabolite repression 4
CEBPβ	CCAAT/enhancer-binding protein β
CELF	CUGBP1 and Etr3 Like Factor
COX-2	Cyclooxegynase 2
CPSF	Cleavage and Polyadenylation Specificity Factor
CSTF-64	Cleavage Stimulatory Factor 64kDa
CUGBP	CUG-Binding Protein
DCP	Decapping enzyme
DM1	Myotonic Dystrophy Type 1
DNA	Deoxyribonucleic Acid
DNΔCELF	Dominant Negative CELF1
EDEN	Embryonic Deadenylation Element
EDEN-BP	Embryonic Deadenylation Element Binding Protein
eIF2	Eukaryotic Initiation Factor 2
eIF4E	Eukaryotic Initiation Factor 4E
ELAVL	Embryonic Lethal Abnormal Vision Like
EMSA	Electrophoretic Mobility Shift Assay
Erk	Extracellular Signal-Related Kinase
ETR	ELAV-type RNA Binding Protein
GRE	GU-Rich Element
HA	Hemagglutinin
HHV-8	Human Herpes Virus 8
hnRNP	heterogeneous ribonucleoprotein
HNS	HuR Nucleocytoplasmic Shuttling
HuR	Human Antigen R
IL-2	Interleukin 2
IL-8	Interleukin 8
INF-γ	Interferon-gamma
IRES	Internal Ribosomal Entry Site
KSHV	Kaposi's Sarcoma Herpes Virus

KSRP	KH-Type Splicing Regulatory Protein
Lsm	Like smith antigen
MAPK	Mitogen Activated Protein Kinase
MBNL1	Muscle Blind Like 1
MEK	Mitogen Activated Protein Kinase Kinase
miRNA	micro-RNA
MK2	MAPK-activated Protein Kinase 2
mRNA	Messenger Ribonucleic acid
MS2	viral coat protein
Not	Negative on TATA
NP-40	Nonident P-40
PABP	PolyA Binding Protein
PARN	polyA Ribonuclease
PM/SCL	Polymyositis/scleroderma complex
polyA	Polyadenylation
RISC	RNA Induced Silencing Complex
RNA-IP	RNA Immunoprecipitation
RRM	RNA Recognition Motif
siRNA	Small Interfering RNA
TIA1	T cell Intracellular Antigen 1
TNFα	Tumor Necrosis Factor alpha
TTP	Tristetraprolin
VEGF	Vascular Endothelial Growth Factor
XRN	Exoribonuclease
β-globin	beta globin

Preface

Some of the work presented in the following chapters has been published or submitted for publication and represents contributions from multiple authors. Chapter 1 is a compilation of a book chapter accepted for publication in conjunction with Paul Bohjanen and Irina Vlasova-St. Louis in the book "Binding Protein", published by InTech, a book chapter submitted for publication in conjuction with Paul Bohjanen to the Encyclopedia of Molecular Cell Biology and Molecular Medicine, and a review accepted for publication in WIRES RNA in conjunction with Paul Bohjanen. My role in these three publications was to write the primary manuscript, produce associated figures, and participate in the editing process.

Chapter 3 has been published in the journal Molecular and Cellular Biology [184] in conjunction with Bernd Rattenbacher and me as co-first authors, and Darin Wiesner, Jonathan Jeschke, Maximillian von Hohenberg, Irina Vlasova-St. Louis and Paul Bohjanen as additional co-authors. In conjunction with Bernd Rattenbacher, I performed experiments for figures 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6. Darin Wiesner, Jonathan Jeschke, and Maximillian von Hohenberg assisted in northern blot experiments for figures 3.1, 3.2, 3.3, 3.4, and 3.5, and Irina Vlasova-St. Louis participated in northern blot analysis for the same figures. I performed data analysis and compiled data for tables 3.1, 3.2, and 3.3. Bernd Rattenbacher performed experiments for figures 3.7 and 3.8, and I analyzed northern blots for figure 3.8.

Chapter 4 has been published in the Journal of Biological Chemistry [220] in conjunction with Bernd Rattenbacher, Irina Vlasova-St. Louis and Paul Bohjanen. Data

in this chapter identified phosphorylation as a regulator of CELF1 mRNA decay in T cell stimulation. In conjunction with Bernd Rattenbacher I performed RNAIPs, and I analyzed and compiled data for tables 4.1, 4.2, and 4.3 and figures 4.2, 4.3 and 4.4. I performed experiments for figures 4.1, 4.5, 4.6, and 4.7.

The data in chapters 5 and 6 are unpublished. Chapter 5 represents data supporting the hypothesis that dysregulated CELF1 mediated mRNA decay contributes to oncogenesis. Figures 5.1 and 5.2 were generated by Bernd Rattenbacher in conjunction with Mark Cannon. I generated figures 5.3 and 5.4, and generated figure 5.5 with the help of Michael Snavely. Chapter 6 represents data that has been generated and analyzed solely by myself.

Chapter 1

Introduction

Genetic information is encoded in DNA, and in eukaryotes this information is confined to the membrane bound nucleus. In the nucleus, DNA is transcribed into mRNA and exported to the cytoplasm, where that mRNA is translated into proteins. The physical separation of transcription and translation has necessitated the evolution of complex regulatory mechanisms to control the proper expression of the information coded in mRNA [1]. These regulatory mechanisms allow the cell to precisely and flexibly control the efficiency at which mRNA is translated as well as the lifespan of mRNA in the cytoplasm [2]. The importance of these "post-transcriptional" regulatory mechanisms is highlighted by the observation that many human diseases including some viral infections [3-5], many types of cancer [6-8], and some neuromuscular disorders [9, 10] all share the common feature of disrupted post-transcriptional regulation. The field of post-transcriptional regulation has made great strides in understanding the basic biology of these regulatory mechanisms, and is transitioning its focus onto how to leverage this knowledge to develop therapies for human diseases.

Cells have evolved intricate mechanisms to control the cytoplasmic abundance of mRNAs. The steady-state abundance of an mRNA species is a function of its rate of production (transcription) and its rate of degradation. Classically, it was assumed that much of the regulation of mRNA abundance occurs at the level of transcription, and that upon export from the nucleus, a transcripts lifespan was stochastically determined by the availability of the RNA degradation machinery. A large body of evidence now exists which refutes this assumption. We now understand that the half-life of an mRNA is tightly regulated, and variable depending on the cellular and environmental context [11].

The importance of mRNA decay in the control of genetic information is highlighted by investigations aiming to understand the process of T lymphocyte activation [12, 13]. Within minutes to hours of receiving the appropriate activation signals, T lymphocytes display dramatic changes in their transcriptome [14] which allow the cell to undergo a cellular program characterized by rapid proliferation and altered apoptotic responses [15]. It has been shown that approximately 50% of the changes in T cell gene expression during activation are a result of changes in transcript half-life [16]. These changes have been quantified on a genome wide scale, and individual transcript half-lives can change by orders of magnitude following activation [17].

The components which regulate a transcripts half-life include cis elements and trans-acting factors which recognize those elements [2]. Cis elements are sequences harbored within an mRNA molecule that are recognized by trans factors. Cis elements regulating decay typically occur in the 3' Untranslated Region (3'UTR) of transcripts and function through recruitment of trans factors such as RNA-binding proteins and miRNA [2, 18]. Whether a cis-element promotes transcript stability or instability is determined by the biochemical activity of the trans factor which binds to that element [2].

ARE-Mediated mRNA Decay

The best characterized cis-element involved in mRNA decay is the AU-Rich Element (ARE). AREs were first discovered through a bioinformatic approach investigating evolutionarily conserved sequence elements in the 3'UTR of the TNFα transcript [19]. A short, A- and U- rich sequence was identified which, when inserted into the 3'UTR of an otherwise stable β-globin transcript, caused destabilization of that

transcript [20]. Subsequently, numerous cytokine (TNF-α, IL-2, and INF-γ) and proto-oncogene (c-fos, v-myc) transcripts were identified as containing AREs [21]. Initially the ARE was defined as a sequence containing overlapping AUUUA pentamers, but this definition has since been broadened to include several variations on this sequence. AREs are now classified into three classes depending on their sequence content, and biologic effect on harboring transcripts [22]. In this system, class I AREs contain 1-3 copies of scattered AUUUA motifs with a nearby U-rich region and caused synchronous deadenylation. Class II AREs are composed of at least two overlapping copies of the UUAUUUA(U/A)(U/A) nonamer in a U-rich region and caused asynchronous deadenylation. Finally, class III AREs were said to contain a U-rich region and other "poorly defined features" and caused synchronous deadenylation [22]. This classification system has proved to be useful in understanding the observed variability in the behavior of ARE containing transcripts.

Given the heterogeneity in the structure of AREs, it is not surprising that there are multiple RNA binding proteins capable of binding ARE containing transcripts (AREBPs). AREBPs include over 20 proteins from various families, with approximately 12 well characterized members [23]. AREBPs can promote both transcript stability and transcript decay. The best studied stabilizing AREBPs include the ubiquitously expressed HuR (ELAVL1, HuA) and the predominantly neuronal HuD (ELAVL4) [23-25]. These proteins are both members of the evolutionarily conserved family of Embryonically Lethal Abnormal Vision (ELAV) proteins which are composed of 3 RNA Recognition Motifs (RRMs) with the 2 N-terminal RRMs separated from the C-terminal

domain by the HuR Nucleocytoplasmic Shuttling (HNS) domain involved in determining the proteins subcellular location [26]. These proteins bind to U-rich sequences (class III AREs) and promote transcript stability [26]. It is thought that these proteins promote stability through a competitive binding model where they compete for binding to AREs with other AREBPs which promote decay, and thereby inhibit the transcripts decay [27].

There are several examples of AREBPs which promote transcript decay, and some of these include Tristetraprolin (TTP), BRF1/2, AUF1, and KSRP [28]. The best studied member of this group of proteins is TTP. TTP preferentially binds to AREs with overlapping AUUUA multimers (class II AREs), and promotes the rapid decay of the harboring transcript through the recruitment of enzymes containing RNAse activity [29, 30]. The importance of ARE mediated mRNA decay, and TTP specifically, was shown in mice that had been made genetically deficient in TTP. These mice developed a syndrome of cachexia, arthritis, and autoimmunity shortly after birth [31]. It was shown that this constellation of symptoms was a result of the aberrant upregulation of the pro-inflammatory molecule TNF-α, a target of TTP mediated decay [31].

Studies investigating the mechanism of TTP-mediated mRNA decay have provided a model for understanding mRNA degradation more broadly. Almost all mature, eukaryotic mRNA contain an untemplated string of adenosine residues on their 3' ends, termed a polyA tail [32]. The first step in mRNA decay is the shortening of this polyA tail (deadenylation) by various deadenylase enzymes, including the CCR4-NOT complex and the PolyA Ribonuclease (PARN) enzyme [33, 34]. TTP has been shown to recruit PARN through in vitro decay experiments [35], and to associate with the NOT1

component of the CCR4-NOT complex [36]. Following the shortening of the polyA tail the transcript can enter either a 3'-to-5' or 5'-to-3' decay pathway. In the 3'-to-5' pathway, the deadenylated transcript can then be targeted by the exosome, a large multi-protein complex with RNAse activity [30, 37]. In the 5'-to-3' decay pathway the deadenylated transcript (which now contains a short polyA tail) is bound by the Lsm1-7 complex, which in turn recruits the DCP1 and DCP2 enzymes [38]. DCP1/2 removes the 5'-m^7G cap allowing for the transcript to be degraded in a 5'-to-3' direction by the exoribonuclease 1 (Xrn1) enzyme [39, 40]. TTP has been shown to recruit both XRN1 and the exosome subunit PM/SCL-100 suggesting that it is able to promote both 3'-to-5' and 5'-to-3' degradation [41].

It is interesting to note that in addition to regulating mRNA turnover, AREBPs are also able to regulate the translational state of an mRNA. One example of this is the translational regulation by ELAVL1 of some ARE containing transcripts [42]. ELAVL1 has been shown to increase the translational efficiency of the transcripts coding for p53 [42], cytochrome c [43], XIAP [44], and BCL2 [45], while decreasing the translational efficiency of those coding for p27 [46], MYC [47], and WNT5α [48]. While the biology behind this dichotomous behaviour is poorly understood, the mechanism of HuR mediated control of MYC has been well studied. In this example, ELAVL1 bound to the MYC 3'UTR and recruits the RNA-Induced Silencing Complex (RISC) loaded with the miRNA Let-7 [47]. RISC is able to compete for binding to the 5'-cap structure with eIF4E, and thus inhibit translation [49]. Whether this interaction with the miRNA

pathway is a universal mechanism for HuR, or AREBP, translational regulation remains to be resolved.

GRE-Mediated mRNA Decay

Investigations into the regulation of mRNA decay during T cell activation yielded a genome wide database of mRNA half-lives in resting and activated T cells [17]. It was observed that while many of the short lived transcripts in this database contained AU-rich elements, many more did not contain a known decay element [17]. This observation lead researchers to perform a bioinformatic search for conserved elements in the 3'UTR of short lived transcripts in T cells [50]. This search identified the AU-rich element, as expected, but also identified a G and U rich element (GRE) which was highly enriched in these short lived transcripts [50]. The newly identified motif was of the form UGUUUGUUUGU and was shown to be bound by a protein in the cytoplasm of T cells [50]. The GRE-binding protein was subsequently identified as CUGBP and Etr3 Like Factor 1 (CELF1, CUGBP1) and was shown to be responsible for mediating the decay of transcripts containing the GRE [50].

The CELF protein family is an evolutionarily conserved family of RNA-binding proteins that play essential roles in post-transcriptional gene regulation [51, 52]. These proteins contain three highly conserved RNA-Recognition Motifs (RRM) with the 2 N-terminal RRMs and the C-terminal RRM being separated by a highly divergent linker domain. The RRMs confer RNA binding activity, and it is postulated that the divergent linker domain is an important site for functional regulation. Six members of the CELF family have been identified in humans and mice: CELF1 (CUGBP1) and CELF2

(CUGBP2) proteins are expressed ubiquitously and play vital role in embryogenesis [53-57] whereas CELF proteins 3-6 are restricted to adult tissues and found almost exclusively in the nervous system [58, 59]. This evolutionarily conserved family of proteins, along with its homologs in *Gallus gallus* (chickens), *Zebrafish, Xenopus, Drosophila* and *C. elegans,* have been shown to regulate gene expression at posttranscriptional levels and to control important developmental processes [51, 60-64].

One highly studied CELF1 homologue is the Xenopus protein embryo deadenylation element binding protein (EDEN-BP). This protein binds to a UGU-rich element termed the embryo deadenylation element (EDEN) to promote transcript deadenylation [65, 66]. EDEN-BP is 88% homologous to human CELF1, and CELF1 has been shown to be able to functionally replace EDEN-BP in Xenopus extracts [67]. The deadenylation promoting function of CELF1 homologues is conserved through humans, as CELF1 has been shown to recruit the deadenylase PolyA Ribonuclease (PARN) as an early step in the promotion of mRNA decay [68]. Functionally, EDEN-BP is involved in muscle and bone formation during Xenopus development as evidenced by faulty somatic segregation in EDEN-BP deficient Xenopus embryos [69]. The C. elegans protein ETR-1 is also a homologue of human CELF1. The function of ETR-1 in C. elegans development was investigated by inactivating ETR-1 using interfering RNA. This reduction in ETR-1 levels led to a defect in muscle development which caused failure of embryos to elongate and paralysis ultimately resulting in embryonic lethality [62]. This finding that CELF1 homologues are critical for muscle development in Xenopus and Drosophila closely mirrors findings that dysregulation of CELF1 is

involved in the pathogenesis of the human neuromuscular disease, myotonic dystrophy type 1 (DM1) [70]. Interestingly, CELF1 has been shown to be a major regulator of mRNA decay in muscle tissue through recognition of the GRE, suggesting that dysregulated CELF1 mediated mRNA decay may be involved in the pathogenesis of DM1[71]. These results suggest that in addition to conservation of the biochemical mechanisms of CELF1-GRE regulation, there may be conservation of the function of CELF1-GRE regulation at the organismal level in regulating specific aspects of developmental programs.

Biochemistry of CELF Protein Binding to Target mRNA

CELF 1 and 2 proteins were first isolated and characterized as novel heterogeneous nuclear ribonucleoproteins (hnRNPs). Several studies have since utilized various experimental techniques to investigate the binding specificity of these proteins. Timchenko et.al demonstrated that these proteins bound to RNA containing the sequence $(CUG)_8$ within the 3'UTR of myotonin protein kinase mRNA *in vitro* [72, 73]. Subsequent searches for the RNA-binding specificities of CELF1 and CELF2 using systemic evolution of ligands exponential enrichment (SELEX) revealed that both CELF1 and CELF2 bound preferentially to GU-rich single-stranded RNA sequences [74]. Binding by CELF1 to GU-rich sequences *in vitro* and *in vivo* was abrogated by mutation of G nucleotides to C [50, 75]. Takanashi et. al. used a yeast three hybrid system for evaluating RNA-protein interactions, and found CELF1 bound preferentially to UG repeats rather than to CUG repeats [76]. CELF1 bound with high specificity to $(UG)_{15}$ based on a surface plasmon resonance (SPR) quantitative binding assay [77].

Overall, the results of these experiments suggest that CELF1 binds preferentially to GU-rich sequences, but the field still lacks a formal definition of the *in vitro* binding specificity of CELF1. Orthologues of CELF1 in other species also appear to have preferences for binding to GU-rich sequences. In *Xenopus*, the CELF1 orthologue EDEN-BP binds to the GU-rich EDEN element, which contains the sequence $(UGUA)_{12}$, and functions as a deadenylation signal in *Xenopus* embryos after fertilization [78, 79]. In *Drosophila*, the CELF1 orthologue Bru-3 was found to bind specifically to $(UG)_{15}$ repeats and also was able to bind to the *Xenopus* EDEN element [51]. The *Zebrafish* protein Brul, a homologue of EDEN-BP with 81% identity, was also shown to preferentially bind to GU-rich RNAs [80]. EDEN-BP and Bru-3 can bind to GRE-RNA as dimers [81, 82], and may require GU-rich sequences of sufficient length to allow dimer formation [83]. In addition to the primary GU-rich sequence, adjacent sequence elements may also be important for assembly of CELF proteins on RNA by allowing optimal secondary structure to facilitate the formation of RNA-protein complexes [84, 85]. Further work is needed to formally define the *in vitro* binding specificity of CELF1.

Structural studies have provided valuable insight into the mechanisms underlying the RNA- binding activity of CELF1. CELF proteins contain two N-terminal and one C-terminal RNA recognition motifs (RRMs), separated by a 160-230 residue divergent domain [86, 87]. The highly conserved RRM domains of CELF1 bind to RNA in a sequence-specific manner [88, 89]. Nuclear Magnetic Resonance spectroscopic (NMR)-based solution studies investigating the CELF1 RRM 1 and 2 domains demonstrated that both RRM1 and RRM2 each contributed to binding to a 12-nt target RNA containing two

UUGUU motifs. These experiments showed that tandem RRM1/2 domains showed increased affinity for compared to the binding by each domain separately, indicating binding cooperativity between the two RRMs [90, 91]. Crystallographic studies showed that both RRM2 and RRM1 bind to GRE-RNA, and RRM1 is important for crystal-packing interactions [92]. In addition to RRM1 and RRM2, RRM3 also has RNA-binding activity. According to NMR analysis, RRM3 specifically recognizes the UGU trinucleotide segment of bound $(UG)_3$ RNA through extensive stacking and hydrogen-bonding interactions within a pocket formed by the beta-sheet and the conserved N-terminal extension of RRM3[89].

While NMR experiments have been crucial to understanding the interaction of CELF1 with RNA, insight into the contribution of various CELF1 domains to RNA binding has come from biologic deletion and mutation studies. Experiments investigating CELF1 function through a yeast three hybrid system suggested that deletion/mutation of RRM1 or RRM2 does not abrogate binding to GU-rich RNA, suggesting that RRM3 may recognize GU-repeats more avidly than RRM1 or RRM2 [77]. Additionally, it has been reported that RRM3 is able to recognize a poorly defined G/C-rich sequence from the 5'UTR of Cyclin D1 when combined with the divergent domain [93]. The divergent domain also appears to be important for RNA-binding since the presence of a divergent domain within recombinant CELF1/CELF4 chimeric proteins increased RNA-binding affinity, perhaps by conveying important conformational changes important for RNA-binding [76, 77, 94] . Intriguingly, the divergent domain may also facilitate CELF:CELF

homotypic interactions [81] which may influence its activity. For example, CELF:CELF interactions appear to activate RNA deadenylation in *Xenopus* extracts [83].

Other Cellular Functions of CELF1

CELF1 as a Regulator of mRNA Splicing

Pre-mRNA alternative splicing is a common mechanism for generating transcript and protein diversity. An estimated 90% of human genes produce alternatively spliced transcripts [95, 96]. Alignment of the genomic regions adjacent to mammalian intron-exon splice sites, identified TG-rich motifs (TTCTG and TGTT) as conserved *cis*-elements found at splicing acceptor sites associated with alternative splicing [97, 98]. These C/UG-rich sequences serve as binding sites for CELF proteins which activate or repress the splicing of pre-mRNA targets, depending on the genomic context of the CELF1 binding site [99]. Recent evidence has re-confirmed the position-dependence of CELF1-binding sites in regulating exon inclusion or skipping [100].

CELF1-mediated regulation of alternative splicing is critical for maintenance of normal muscle structure and function [101, 102]. CELF1's function as a regulator of alternative splicing has largely been elucidated through investigation of CELF1's role in the pathogenesis of the neuromuscular disease myotonic dystrophy type 1 (DM1). In this disease, aberrant gain of CELF1 function is combined with a corresponding loss of function of the splicing factor MBNL1. CELF1 and MBNL1 have overlapping binding specificities but opposing effects on exon/intron inclusion. The result of concurrent CELF1 loss and MBNL1 gain of function is the mis-splicing of a number of crucial genes [70]. Minigene reporter systems that contain alternative splice sites were used for the

identification of pre-mRNA targets for CELF1, including genes for cardiac troponin T [103], insulin receptor [104], and chloride channel 1 [105, 106]. Interestingly, these genes were all shown to be mis-regulated in tissues from patients who suffered from DM1. Minigene systems have been particularly useful in demonstrating that individual pre-mRNA splicing events are affected by loss or gain of activities of specific regulatory proteins. Studies performed in cultured cells with transiently transfected minigenes have identified a number of alternative gene regions regulated by CELF1 and other family members [88, 103-114]. Importantly, the results of minigene over-expression experiments may not necessarily reflect the full-length mRNA splicing patterns observed *in vivo*, especially during certain stages of organism development [115].

Clues to the physiologic role of CELF1 mediated alternative splicing come from studies of alternative splicing during development. Through the use of transgenic CELF1 deficient mouse models, it was found that CELF proteins are involved in regulating the switch from fetal to adult splicing patterns of several skeletal muscle transcripts [105, 116]. Additional studies in mice using splicing microarrays to globally characterize the effect of CELF1 on the fetal-to-adult transition found that nearly half of transcripts that undergo fetal-to-adult alternative splicing transitions in heart tissue responded to over-expression of CELF1. These results suggested that the level of CELF1 activity directly regulates the alternative splicing pattern of endogenous transcripts [117]. The development of dominant negative (DN) and tissue specific transgenic mice was advantageous for studying CELF-specific alternative splicing *in vivo* [107, 110, 115, 116]. For example, dominant negative CELF (DNΔCELF) expressed under the control of

a cardiac muscle-specific promoter, promoted the development of dilatated cardiomyopathy and cardiac dysfunction over time [118]. In contrast, when DNΔCELF was expressed under the control of a skeletal muscle-specific promoter, mice exhibit reduction in muscle intersticia and an increase in slow twitch fibers [116]. It would be interesting to see more phenotypic studies using a nucleus-restricted form of the dominant negative CELF protein which would presumably block only the CELF1 nuclear function, leaving the cytoplasmic function intact [119, 120].

CELF1 as a Regulator of Translation

Translation is a critical layer of post-transcriptional control of gene expression that is regulated in response to environmental and developmental changes. CELF proteins have been shown to be involved in the activation of translation of several mRNA species at various stages of development [121]. Additionally, CELF proteins have been shown to function as inhibitors of translation under conditions of stress, where they act as translational silencers in conjunction with other protein binding partners. The involvement of CELF1 in translational regulation is evolutionarily conserved, with several CELF1 homologues having been shown to regulate translation. For example, in the *Drosophila* oocyte, translational repression is mediated by the protein Bruno (CELF1 homolog)**,** that binds specifically to bruno response elements (BREs) within the oskar mRNA 3'UTR. Binding by Bru-3 (CELF1 ortholog) to GU-rich sequences in 3'UTR of gurken, cyclin A and oskar mRNA leads to translational repression [60]. The suggested mechanism underlying Bru-3 mediated translational regulation is through the formation of a Bru-3/eIF4E/5'-cap translational silencing complex during specific stages of embryo

development [122].

CELF1 has also been shown to play a role in translational regulation in mammalian cells, although through a variety of mechanisms. The best studied example of CELF1 mediated translational control involves the ability of CELF1 to promote the translation of alternative isoforms of the transcription factor CCAAT/enhancer-binding protein (CEBPβ) [123-125]. This mode of regulation was first described in a rat model of partial hepatectomy. In this model, CELF1 was phosphorylated following partial hepatectomy, and this phosphorylation promoted the formation of a complex between CELF1 and eIF2. This CELF1-eIF2 complex led to the selective translation of the liver enriched inhibitory protein (LIP) isoform of CCAAT/enhancer-binding protein [126]. It was later shown that in liver, CELF1 undergoes an increasing degree of hyper-phosphorylation during normal ageing through a GSK3beta-cyclin D3-cdk4 kinase pathway[127]. Similar to the partial hepatectomy model, the cdk4-mediated hyper-phosphorylation of CELF1 was involved in the age-associated induction of the CELF1-eIF2 complex [128]. In the rat ageing model, the CELF1-eIF2 complex bound to the 5'UTR of HDAC1 mRNA and increased histone deacetylase 1 protein levels in aging liver [128]. It was further shown that during rat aging, CELF1 phosphorylation promoted its interaction with a GC-rich sequence in the 5'UTR of p21 mRNA causing p21 translational arrest and senescence in fibroblasts [129]. In myocytes, p21 mRNA was stabilized in discrete cytoplasmic structures called stress granules, which served as reversible storage sites for mRNA under conditions of stress. Interestingly, only during late senescence did p21s localization in stress granules interfere with its translation [130].

One important component of stress granules is the RNA-binding protein T cell internal antigen 1 (TIA1). Consistent with CELF1's recruitment to stress granules, CELF1 has been shown to function as a translational silencer through interaction with the TIA1 protein [131]. Further support for this model came from experiments utilizing DM1 cell harboring CUG repeat RNA. The presence of a CUG repeat expansion was found to cause stress and activation of the PKR-phospho-eIF2α–CELF1 pathway leading to stress granule formation and inhibition of mRNA translation [132]. This disruption to physiologic mRNA translation pathways by cellular stress signals might contribute to the progressive muscle loss in DM1 patients. Taken together, this data suggests that CELF proteins may function as activators or repressors of translation, depending on the context.

In addition to the studies outlined above, CELF1 has been implicated in translational regulation through several alternative mechanisms. In human cell lines, tethering of CELF1 to the 3'UTR of mRNA through an interaction with the MS2 coat protein led to decreased steady state levels of reporter transcripts that contained a MS2 RNA-binding site, while reporter protein levels increased [133]. Additionally, CELF1 increased the translation of p21 protein [134], and Mef2a29 [135] during normal muscle cell differentiation via direct interaction with (GC)n repeats located within the 5'UTR of those mRNAs. Recently, an additional mechanism for CELF1 mediated translational regulation through interaction with the 3'UTR was discovered. Binding of CELF1 to the 3'UTR of the mRNAs of Serine hydroxymethyltransferase (SHMT) [136, 137] and cyclin dependent kinase inhibitor p27 (Kip1) [138], was found to regulate an internal ribosome entry site (IRES) translation activation. This implicated CELF1 in participating in an

IRES mode of initiation of mRNA translation. In addition, IRES translation is achieved through CELF1/hnRNPH complex formation, which promotes circularization of RNA transcripts by mediating 5'/3' ends interactions (Fox J. et.al) [136]. Whether CELF1 recruits the translation machinery to the 5'UTR via additional interaction with eIF2 (Eukaryotic Initiation Factor 2) or another initiation factor remains to be determined.

Interestingly, the available data suggests that CELF1 mediates translational regulation through interaction with various motifs including a G- and C-rich motif in the 5'UTR, whereas CELF1 mediates its splicing and degradation effects through interaction with a G- and U-rich motif in introns and 3'UTRs, respectively. Experiments determining CELF1 binding targets through high-throughput means have failed to identify enrichment of GC-rich motifs or 5'UTR binding by CELF1. One explanation of these contradictory observations could be that CELF1 mediated translational regulation is rare, and only occurs in the context of very specific mRNA species and cellular contexts.

Regulation of CELF1 Through Phosphorylation

CELF1 is a known phosphoprotein with multiple predicted phosphorylation sites. It has been shown in a number of model systems that phosphorylation of CELF1 regulates its function in the context of alternative splicing, mRNA decay, and translational regulation [139-141]. One of the pathologic events in DM1 is an increase in the protein abundance of CELF1 and an associated increase in CELF1 mediated alternative splicing activity. This increase in CELF1 protein abundance is a result of increased CELF1 protein stability secondary to hyperphosphorylation [142]. In DM1, the (CUG)n expansion of the DMPK 3'UTR leads to protein kinase C (PKC) activation

through an unknown mechanism. PKC, in turn, hyperphosphorylates CELF1, resulting in increased protein stability and abundance as well as increased splicing activity [143]. In addition to influencing CELF1 mediated splicing efficiency, it has also been shown that the hyperphosphorylated form of CELF1 in DM1 is unable to promote the decay of TNF-α suggesting that phosphorylation may also influence CELF1 mRNA decay [144]. Support for a pathologic role for CELF1 hyperphosphorylation in DM1 comes from transgenic mouse models of DM1, in which mice treated with specific inhibitors of the PKC pathway showed amelioration of cardiac abnormalities associated with the disease phenotype [145].

Phosphorylation of CELF1 also influences its ability to regulate muscle development (rev in [146]). CELF1 phosphorylation by Akt kinase at Ser 28 in normal muscle myoblasts influences its ability to affect the translation of its target transcripts during differentiation. Finally, phosphorylation of CELF1 has been shown to directly impact its RNA-binding activity. Cyclin D3-Cdk4/6, for example, phosphorylates CELF1 at Ser 302, altering the binding specificity of CELF1 to RNA [132]. Overall, it is clear that many aspects of CELF1 function are regulated through phosphorylation, and that phosphorylation of CELF1 is observed in multiple disease states.

Aberrant Regulation of mRNA Decay Networks in Cancer

An important advance in the field of RNA biology in the past decade has been the recognition that altered mRNA metabolism contributes to an oncogenic phenotype [147]. Many aspects of RNA metabolism including splicing patterns [148], polyadenylation site usage [149], RNA degradation [8], and translational efficiency [150] have been shown to

contribute to oncogenesis or malignant transformation. Here we discuss data supporting a role for altered ARE-mediated decay, Kaposi's Sarcoma Herpes Virus infection, and aberrant alternative polyadenylation in influencing mRNA decay patterns to promote oncogenesis.

Aberrant ARE-Mediated Decay in Cancer

Increased levels of the ARE-binding protein HuR are found in several types of cancer [7, 23, 151], and several studies have worked to define mechanisms by which increased HuR may contribute to oncogenesis. HuR has been shown to target the transcript coding for vascular endothelial growth factor (VEGF) [152]. VEGF is expressed in response to hypoxia and serves as a paracrine factor to promote tissue vascularization. In the setting of cancerous tissue, tumor expression of VEGF promotes vascularization of the tumor allowing for increased delivery of oxygen and other nutrients [153]. The interaction between HuR and VEGF serves to stabilize the VEGF transcript and promotes upregulation of the VEGF protein [152]. Another HuR target is the cycloxygenase 2 (COX-2) transcript [6]. COX2 is involved in prostanoid synthesis and tends to be upregulated in cancerous tissue [154]. Similar to VEGF, the interaction of HuR with the COX-2 enzyme promotes increased COX-2 transcript stability and subsequently increased COX-2 protein [6]. Increased HuR expression is correlated with higher tumor grade in colon [155] and breast cancer [156], perhaps through stabilization of these and other transcripts that promote tumor development. In addition to up-regulation of HuR in the cytoplasm of tumors, it has also been shown that down-regulation of TTP is a common finding in a variety of primary cancer tissues [157]. TTP

promotes the decay of many ARE-containing transcripts, and this down-regulation of TTP in cancer promotes the up-regulation of target transcripts that may promote cancer. It was further shown that the level of TTP in cancer tissue is inversely proportional to prognosis [157]. These and other studies suggest that ARE-binding proteins, in particular HuR, may be promising targets for the design of molecular therapies to combat cancer by inducing ARE-mediated mRNA decay. The development of small molecule inhibitors of HuR might present novel treatment options for cancer patients, and would provide researchers with useful tools for understanding the basic biology of cancer.

KSHV Mediated Oncogenesis

In addition to genomic mutations, viral infections are also able to promote the development of cancer. A handful of viruses, including Kaposi's Sarcoma Herpes Virus (HHV-8), have been implicated in human oncogenesis [158]. KSHV is a double-stranded DNA virus with a genome of approximately 165kb in length. The KSHV genome encodes for several proteins with homology to human genes, including the oncogenic, virally encoded G protein-coupled receptor (vGPCR) [159], which is a homologue of the human IL-8 receptor. This ligand-independent vGPCR constitutively activates the MEK/ERK and PI3K pathways through the recruitment of independent G-proteins [160, 161]. Signaling through the vGPCR activates multiple transcription factors including AP-1 and NFκB to promote cellular transformation [160, 161].

Another mechanism utilized by KSHV to promote malignancy is alteration of mRNA decay pathways. KSHV encodes Kaposin-B, a protein which binds and activates the MK2 kinase [162]. The MK2 kinase is involved in activating the p38 MAPK

pathway in response to inflammatory signals [163]. It has been shown that the interaction between Kaposin-B and MK2 promotes the cytoplasmic accumulation of HuR, and subsequent stabilization of the PROX1 mRNA via an ARE present in the 3'UTR of PROX1 [164]. PROX1 is a master regulator of lymphatic reprogramming of blood vascular cells [165, 166], a characteristic feature of KSHV infection [167]. These results suggest that KSHV has evolved mechanisms to hijack the cellular mRNA decay machinery to influence the composition of the cellular transcriptome. It remains to be investigated whether KSHV is able to influence alternative mRNA decay pathways such as GRE mediated decay.

Aberrant Alternative Polyadenylation in Cancer

Newly transcribed pre-mRNA molecules undergo a number of nuclear maturation steps prior to their export into the cytoplasm. Polyadenylation is required for the maturation of nearly all mRNA, and involves the endonucleolytic cleavage of the 3'end followed by the addition of a string of untemplated adenines [168]. Our understanding of the mechanism of polyadenylation is limited to just a few of the almost 100 proteins is involved [168]. Polyadenylation is triggered by the recognition of the polyA signal of the form A[U/A]UAAA by the protein cleavage and polyadenylation specificity factor (CPSF), and the recognition of a poorly defined downstream U/GU rich region by the protein cleavage stimulatory factor of 64kD (CSTF64) [169]. The cooperative interaction between CPSF and CSTF64 serves to recruit cleavage factors I and II, which endonucleolytically cleave the RNA between the polyA signal and the downstream U/GU

sequence [168, 170] . Subsequently, the enzyme polyA polymerase adds a string of single stranded adenines to the 3'end [168].

Polyadenylation can occur at multiple locations within the 3'UTR, and the utilization of non-canonical polyadenylation sites (polyA sites) is termed alternative polyadenylation (APA). It is thought that the levels of CSTF64 may influence the choice of polyA site [171, 172] , but it is likely that additional sequence elements and protein factors are involved in the choice of polyA site [168]. Regardless of how the choice of polyA site is made, it is clear that the usage of APA sites can have functional implications for regulation of affected transcripts through the loss of regulatory sequences residing 3' to the APA site. Recent evidence suggests that APA sites are preferentially utilized under various cellular conditions. T cells, for example, have been shown to utilize proximal polyA sites during proliferation in response to anti-CD3/28 antibody stimulation [173]. Similarly, when analyzing cancer tissue it was found that cancer cells tended to use proximal polyA sites when compared to non-transformed cells [149]. In mouse development, it was found that 3'UTRs tended to use distal polyA sites as developmental progressed, effectively lengthening 3'UTRs [174]. Finally, fully differentiated tissues have been shown to have varying utilization of polyA sites, with certain tissues such as brain and testis transcripts utilizing exceptionally distal polyA sites [175]. The importance of these changes in polyA site utilization are highlighted by the fact that there is evolutionary conservation in polyA site choice, and the variability in polyA site choice is greater between tissues than it is between species [175].

APA regulates transcript fate through altering the 3'UTR sequence content in transcripts. The end result of APA, then, is a function of the specific sequence elements which are included differentially between APA isoforms. A growing body of literature suggests that microRNA binding sites may be some of the sequences which are regulated by APA. It was found in proliferating T cells that the distal, APA excised regions of transcripts were enriched for miRNA binding sites [173]. Additionally, in cancer cells, APA was also found to preferentially regulate miRNA binding sites [149]. In this study, it was found that transcripts whose 3'UTRs were shortened through APA had longer half-lives, and it was shown for some transcripts that a component of the stabilization was a result of loss of miRNA binding sites. Interestingly, it was observed that a large component of the stabilization could not be explained through miRNA-mediated stabilization [149] and was likely the result of additional decay promoting sequence elements. For specific transcripts, it has been shown that AREs are regulated through APA, but it remains to be shown on a genome-wide scale whether ARE-mediated, or GRE-mediated, decay are systematically regulated through APA.

In the chapters that follow, we have addressed some of the knowledge gaps related to CELF1 function and regulation that have been identified above. In chapter three we utilized high-throughput techniques to identify the molecular targets of CELF1 in HeLa cells, and through analysis of these targets better defined the GRE as well as the biologic function of the CELF1-GRE post-transcriptional network. In chapter four we determined the molecular targets of CELF1 in resting and activated primary human T cells leading to the identification of an activation dependent CELF1 phosphorylation

event that causes decreased affinity of CELF1 for the GRE. In chapter five we explored the dysregulation of CELF1 in a model of viral oncogenesis, and found that CELF1 is phosphorylated in the setting of the oncogenic KSHV vGPCR. CELF1 phosphorylation downstream of the vGPCR inhibited CELF1 mediated mRNA decay through the inhibition of CELF1 recruitment of the deadenylase enzyme PARN. Finally, in chapter five, we investigated the impact of APA during T cell stimulation on ARE and GRE mediated mRNA decay using deep sequencing technology. These experiments revealed that AREs and GREs are both preferentially excluded from 3'UTRs as a result of APA during T cell stimulation, although their regulation followed distinct temporal patterns. Overall, the work presented in this thesis has furthered our understanding of the rules and regulation of CELF1 function, while also identifying additional avenues of investigation to further our understanding of the biological importance of CELF1- and GRE- mediated decay.

Chapter 2

Experimental Methods

RNA immunoprecipitation from HeLa cells and microarray analysis. HeLa Tet/off cells (Clontech) were cultured in minimal essential medium alpha (Gibco) containing 10% tetracyclin free FBS (Clontech), 1% L-Glutamine (Gibco) and 100 units/ml Penicillin/Streptomycin (Gibco). Cytoplasmic extractions and RNA-IP were performed as described previously [176, 177] using an anti-HA antibody (F7, Santacruz), anti-CELF1 antibody (3B1, Santacruz), or anti-PABP antibody (Immuquest). Three independent RNA-IP experiments were performed. For each experiment, RNA was purified from the input and immunoprecipitated material from an equivalent number of HeLa cells using the RNeasy kit (Qiagen) following manufacturers recommendations. For the input RNA, 5 ug was used to prepare labeled cRNA for microarray hybridizations. For the IP RNA prepared from an equivalent number of cells, SF9 insect cell RNA was added such that the total amount of RNA totaled 5 ug, and this RNA was used to prepare cRNA using the MessageAMP™ II RNA Amplification kit (Ambion). The cRNA prepared from input RNA and from each IP were hybridized to U133a Plus-2 microarrays. The microarrays were normalized using the gene content robust multi-array average (GCRMA) algorithm using Genespring 9.0 software (Agilent Technologies Inc). Transcripts were considered to be present in the input, anti-CELF1 IP or anti-PABP IP conditions if the log intensity of their average signal was greater than their average log intensity of the signal on the HA chip with a p value less than 0.05 by student t-test.

Identification of conserved sequences of CUGBP1 target transcripts in HeLa Cells.

The UTRs of transcripts found to be present in the anti-CELF1 IP were analyzed by searching for the frequency of the occurrence of all possible 11-mers. The frequency of occurrence of sequences in the 3' UTR of transcripts found in the anti-CELF1 IP were compared to the frequency of occurrence in all the 3'UTRs of 24820 3' UTRs extracted from the NCBI database. In addition, a BioProspector search was performed on the 3'UTRs of the trancripts found in the anti-CELF1 IP (http://ai.stanford.edu/~xsliu/BioProspector/) using the following parameters: motif length of 11 nucleotides, searching the forward strand and background definition human genomic 3' UTR sequences. Consensus motifs were created with Weblogo Software version 2.8.2 (http://weblogo.berkley.edu) [178].

Electrophoretic Mobility Shift and Supershift Assays. Ribo-oligonucleotides were purchased commercially (Sigma). The sequences for each ribo-oligonucleotide are shown as the shaded sequences in Figures 2A. Ribonucleotide oligomers were endlabeled with [γ^{32}P]-ATP (6000Ci/mmol) using T4 polynucleotide kinase (Invitrogen) to produce a radioactively labeled probe with a specific activity of 4 x 10^6 cpm/μg. EMSA was performed as published previously [179]. Each reaction contained 10 ug of HeLa cytoplasmic protein and 10-15 fmol of radiolabeled RNA probe in a total volume of 20–24 ul. For supershift assays, an anti-CELF1 monoclonal antibody (3B1, Santa Cruz Biotechnologies) or an anti-His-tag antibody (SantaCruz Biotechnology Inc.) were added to the reaction mixtures. The mixtures were separated by electrophoresis under nondenaturing conditions on 5% polyacrylamide gels. The gels were dried and analyzed on a Storm 820 phosphorimager (Amersham

Biosciences).

Tet-off mRNA decay assay. The decay of beta-globin reporter constructs was performed as described previously [177, 180]. The Tet-responsive beta-globin expression construct, pTetBBB [181] was a gift from Dr. Ann-Bin Shyu (University of Texas-Houston), and the JUNB-GRE expression construct were described previously [177]. Sequences from the 3' UTR of NDUFS2, GSN or PPIC were inserted at the unique Bgl II restriction site in the beta-globin 3' UTR of pTetBBB to create the NDUFS2-GU, GSN-GU, and PPIC-GU constructs. The NDUFS2-ΔGU, NDUFS2-mGU, GSN-ΔGU , GSN-mGU, GM2, GM3, GM4 or GM8 constructs shown in Figure 1A and Figure 4A respectively were created from the NDUFS2-GU or GSN-GU constructs by site directed mutagenesis using the QickChange® II XL Site-Directed Mutagenesis Kit (Stratagene).

HeLa Tet/off cells (15 cm dish) were transfected with 15 μg of the parental BBB reporter plasmid or BBB plasmids containing 3' UTR inserts. Transfections were performed with 6.25 μg/ml of Lipofectamine 2000 reagent (Invitrogen). Cells were split into four 10cm dishes the next day. After 48 h, 300 ng/ml of doxycycline was added to stop transcription, and total RNA was isolated after 0, 2, 4 or 6 hours. GAPDH probe was generated using the decaprime kit (Ambion). RNA isolation and northern blotting to assess expression of beta-globin and GAPDH transcripts were performed as described previously [180]. For each point, the hybridization intensity of the beta-globin transcript was normalized to the hybridization intensity of the GAPDH transcript, and the normalized values were used to calculate half-lives using

GraphPad Prism 4.03 software based on a one-phase exponential model of decay.

siRNA Transfection. CELF1 knock down experiments were performed using a red fluorescent protein targeting siRNA duplex as control and two CELF1 specific siRNA duplexes described previously [177].

For the Tet-off mRNA decay assays, HeLa Tet/off cells (15 cm dish) were transfected with 1nmol of CELF1 specific siRNAs or a control siRNA. After 24 hours cells were again transfected with 1nmol of siRNA, 15µg of the GSN-GU. Transfections were performed with 6.25 µg/ml of Lipofectamine 2000 reagent (Invitrogen). Cells were split into four 10cm dishes. After 48 h, 300 ng/ml of doxycycline was added to stop transcription, and total RNA was isolated after 0, 2, 4 or 6 hours. GAPDH probe was generated using the decaprime kit (Ambion). RNA isolation and northern blotting to assess expression of beta-globin and GAPDH transcripts were performed as described previously [180]. For each point, the hybridization intensity of the beta-globin transcript was normalized to the hybridization intensity of the GAPDH transcript, and the normalized values were used to calculate half-lives using GraphPad Prism 4.03 software based on a one-phase exponential model of decay. Knockdown of CELF1 was assesed by western blotting for CELF1 with an anti-CELF1 monoclonal antibody (3B1, Santa Cruz Biotechnologies) and an anti-GAPDH polyclonal antibody (FL355, Santa Cruz Biotechnologies) as a loading control.

For apoptosis assays HeLa Tet/off cells (6 well) were transfected with 50 pmol of siRNA of the CELF1 specific or control siRNA on two consecutive days. 48h after

the last transfection apoptosis was induced with 50 ng/ml TNFα (Invitrogen / Biosource) and 50µg/ml Cycloheximide (Acros Organics) for 3 hours. Knockdown of CELF1 was assesed by western blotting for CELF1 with an anti-CELF1 monoclonal antibody (3B1, Santa Cruz Biotechnologies), apoptosis was determined using a rabbit polyclonal antibody against cleaved PARP (Abcam) and an anti-GAPDH polyclonal antibody (FL355, Santa Cruz Biotechnologies) was used as a loading control. FACS analysis was performed on a FACSCanto machine (BD) using a PE-coupled anti-activated caspase 3 antibody (BD Pharmingen). Results were analysed in FlowJo 7.6 (Flowjo.com).

Polysomal gradients. HeLa Tet/off cells (10 cm dish) were transfected with 6 µg of BBB plasmids containing 3' UTR inserts of junB, mutated junB, NDUFS2, mutated NDUFS2, GSN or mutated GSN along with 3 µg of the pTracerC-EF/V5-His/*lacZ* GFP expression plasmid (Invitrogen) to control for transfection efficiency. Transfections were performed with 6.25 µg/ml of Lipofectamine 2000 reagent (Invitrogen). After 48 hours cytoplasmic extracts were prepared and polysomes were separated on a 40% - 15% sucrose gradient as described in [182]. RNA was prepared from each fraction and subjected to northern blotting. Expression of beta-globin and GFP transcripts were assessed by probes specific for the beta-globin and GFP transcripts as described previously [177]. For each point, the hybridization intensity of the beta-globin transcript was normalized to the hybridization intensity of the GFP transcript.

Purification of Human T cells. Primary human T cells were purified from peripheral blood mononuclear cells by negative selection using the CD3+ Rosette-Sep antibody

cocktail from Stem Cell Technologies as previously described [183]. CD3+ lymphocytes were then isolated through a Ficoll-hypaque cushion (GE Healthcare).

T cell Stimulation. Purified human T cells were cultured overnight in RPMI 1640 supplemented with 10% fetal bovine serum, 2mM L-glutamine, 100U/mL penicillin G and 100ug/mL streptomycin. Cells were then incubated for 6 hours in 15cm dishes ($5x10^7$ cells/dish) with medium alone or with a combination of immobilized monoclonal antibodies (1ug/mL) directed against the CD3 component of the TCR complex (R&D Systems) and the CD28 co-stimulatory molecule (R&D Systems) as previously described [50].

RNA-IP and Microarray Analysis for Primary Human T cells. Cytoplasmic extracts were prepared from resting and stimulated T cells and RNA-IP reactions were performed as previously described using antibodies targeting hemagluttinin (HA) (F7, Santa Cruz Biotechnology), CELF1 (3B1, Santa Cruz Biotechnology) or Poly-A binding protein (PABP) (ImmunoQuest) [184]. Three RNA-IP experiments were performed using resting or activated T cells isolated from three individual anonymous donors. For each experiment, RNA was purified from the input and immunoprecipitated material from equal numbers of resting and stimulated T cells using the RNeasy kit (Qiagen) following manufacturers recommendations. For the input RNA, 5ug was used to prepare labeled cRNA for microarray hybridizations. SF9 insect cell RNA was added to the immunoprecipitated RNA such that the total amount of RNA was 5ug. This RNA was then used to prepare cRNA using the MessageAMP III RNA Amplification kit (Ambion). The cRNA from the input and RNA-IP samples from resting and stimulated lysates were

hybridized to Affymetrix U133a Plus-2 microarrays. Microarrays were normalized using the gene content robust mult-array average (GCRMA) algorithm using Genespring 11.0 software (Agilent Technologies Inc). Transcripts were determined to be present if the $\log_2$ normalized signal from the input microarray was greater than the $\log_2$ normalized signal from the HA microarray with $p<0.05$ as determined by Welch's T-test (R) in either the resting or the stimulated conditions.

Target Transcript Identification in Primary Human T cells. Transcripts were determined to be CELF1 targets if the difference between the $\log_2$ normalized signal from the microarrays hybridized with the cRNA from the anti-CELF1 RNA-IP and the anti-PABP RNA-IP was greater than the same value derived from the difference between the anti-HA RNA-IP and anti-PABP RNA-IP, with $p<0.005$, as determined by Welch's T test (R), in either resting or stimulated conditions.

2D-western blotting. 100ug of cytoplasmic lysates were diluted 1:1 with Rehydration buffer from BioRad and dialyzed against 2D gel buffer (7M urea, 2M thiourea, 2% CHAPS, 10mM Tris) overnight at room temperature. The samples were then loaded onto 10cm pH3-10 IPG strips. Samples were focused using a BioRad Protean IEF cell. Subsequently, IPG strips were loaded onto Bis-Tris 4-12% pre-cast gels and run at 175V using MOPS-SDS buffer. Gels were then blotted onto charged PVDF and western blots were performed probing with an anti-CELF1 antibody (3B1) from Santa Cruz Biotechnology. For lambda phosphatase treatment, 100ug of cytoplasmic lysate was incubated with CUGBP1 antibody for 90min at 4C, and subsequently incubated with protein A/G sepharose beads from Pierce for 90min at 4C. Beads were washed 3 times

with RNA-IP lysis buffer and resuspended in 100ul of lambda phosphatase buffer. Samples were then treated with lambda phosphatase (New England Biolabs) or were mock treated, and subsequently, CELF1 was eluted from the beads by the addition of 1% SDS and heating at 65C for 15min. Samples were then diluted 1:1 with rehydration buffer and were separated by 2-D electrophoresis.

EMSA and UV crosslinking assay using T cell cytoplasmic lysates. Biotinylated GRE RNA was ordered from Sigma-Aldrich containing the sequence 5'[Btn]-GAGUGUGUGUGUGUGUGUGUUGUUU-3'. Mutated, biotinylated RNA was also purchased from Sigma-Aldrich and had the sequence 5'[Btn]-GACACAGUGUCACAGUGUCACAUGUUU-3'. Binding reactions were the same as previously described [183] using 10ug of cytoplasmic lysate and 25fmol of biotinylated RNA. For EMSA, electrophoresis was performed using 0.5X TBE buffer and 5% polyacrilamide gels under non-denaturing conditions. For the UV crosslinking assay, reactions were exposed to UV light and were separated by SDS-PAGE on a 10% polyacrylamide gel. In indicated reactions, CELF1 was immunoprecipitated by the addition of an anti-CUGBP1 antibody for 90min at 4C, followed by the addition of protein A/G beads for 90min at 4C. Beads were washed three times, and the material on the beads was separated by SDS-PAGE. Following electrophoresis, gels were blotted onto nylon membranes and UV cross-linked. The blots were then exposed using the Chemiluminescent Nucleic Acid Detection Module from Thermo Scientific, following manufacturer's instructions.

Reverse trancriptase polymerase chain reaction. cDNA used for reverse transcriptase PCR (RT-PCR) amplification was synthesized from total cellular RNA using the StrataScript reverse transcriptase (Stratagene) using oligo(dT). PCR amplifications were then performed as described previously [14]. Quantitative real-time PCR assays were performed using the Roche Universal Probe Library (UPL) technology, and transcript abundance was normalized to the level of the beta-2 microglobulin transcript. Oligonucleotide primers used were ELAVL1: fw-cggctggtgcattttcat rv-caaacggcccaaacatct , FMR1: fw-gggaggaagaggacaaggag rv-atctgttcgggagtgatcgt , HNRNPD: fw- ccaatttgaagattcatttgaagg rv-ttccattcagggacttgataca, PTBP1: fw-tcattgacctgcacaaccac rv-gatggtggacttggagaagg, JunD: fw-cgattctgccctatttatgttt rv-tttttcttcttcccatttcttt, TOP1: fw-agcctcacttgccctcgtgc rv-acaaatttccatccacttgccctct, UBE3A: fw-tggaatgttttaactgcagccctca rv-tggccaaatgcactttccccagt, GAPDH: fw-tgatggtacatgacaaggtgc rv-acagtccatgccatcactgc, and B2M: fw-taggagggctggcaacttag rv-ccaagatgttgatgttggataaga.

UV crosslinking assay using immunopurified CELF1. CELF1 was immunoprecipitated from 20ug of cytoplasmic lysate and samples were treated with lambda phosphatase or were mock treated as described in the 2D-western blotting section above. The immunoprecipitated material was then incubated for 30min with 25fmol of either the biotinylated GRE or mutant GRE probe, as well as 10-fold excess of unlabeled poly-U RNA and 5mg/mL of Heparan Sulfate in 20ul of RBB buffer [183]. Following incubation, the reactions were treated with UV light and material was eluted from the beads by incubation in SDS loading buffer for 5min at 95C. The eluted material was

separated by SDS-PAGE on a 10% acrylamide gel, and the gel was blotted onto nylon membranes. Subsequently, the biotinylated RNA probe was visualized with the Chemiluminescent Nucleic Acid Detection Module from Thermo Scientific, following manufacturer's instructions. The blots were then stripped, and probed by western blot with an anti-CELF1 antibody. Blots were then quantified with ImageJ to determine the RNA:CELF1 ratio.

Preparation of Lysates for KSHV Studies. HeLa cells were grown in 6 well plates to approximately 70% confluency in DMEM supplemented with 10% fetal bovine serum and 100U/mL penicillin G and 100ug/mL streptomycin. Cells were transduced by the addition of 6ul of lentivirus engineered to express the KSHV vGPCR or a vector control (CSGW) to the culture medium. Following transduction cells were allowed to incubate for 48-72 hours upon which time the media was removed, and cells were washed 1X with PBS. Cells were lysed in RNA-IP lysis buffer as previously described and lysates were stored at -80C until usage.

RNA Affinity Purification of Protein Complexes. 500ug of lysate from vGPCR of CSGW conditions was incubated with Biotinylated GRE RNA containing the sequence 5'[Btn]-GAGUGUGUGUGUGUGUGUGUUGUUU-3' or mutated, biotinylated RNA with the sequence 5'[Btn]-GACACAGUGUCACAGUGUCACAUGUUU-3' for 2 hours at 4C. Subsequently 20ul of streptavidin coated agarose beads (Pierce) were washed three times in RNA-IP lysis buffer and added to each reaction. Reactions were then allowed to incubate at 4C with constant agitation for two hours. Following incubation, beads were washed three times in RBB buffer [183], and subsequently complexes were eluted from

the beads by the addition of SDS-loading buffer and boiling for five minutes. Eluted proteins were then interrogated through western blotting using anti-CELF1, anti-PARN, and anti-eIF2 antibodies.

Measurement of CELF1 RNA Affinity. EMSA reactions were performed as described above using the biotinylated GRE and mutated GRE RNA probes described above. For the measurement of CELF1's affinity for the GRE, increasing amounts of non-biotinylated (cold) GRE probe containing the same sequence was added to the reaction.

Determination of Polyadenylation Sites and Comparison between samples. Reads were aligned to the human genome (hg19) using the SHRiMP algorithm [185] with default settings. A list of the coordinates of all human 3'UTR exons was downloaded from the UCSC genome browser [186], and a pileup view of the reads was generated for each 3'UTR exon using samtools [187]. Subsequently, a custom c++ algorithm was written to search for sites which were supported by at least 10 reads. A second script was written to find those 3'UTR exons whose weighted average of read position changed consistently among the three donors at each timepoint, and the sequence between the most distal and proximal polyA site for each of these consistently changed 3'UTRs was extracted using galaxy.

Chapter 3

Analysis of CELF1 targets identifies GU-repeat sequences that mediate rapid mRNA decay

Bernd Rattenbacher†, **Daniel Beisang†**, Darin L. Wiesner, Jonathan C. Jeschke, Maximilian von Hohenberg, Irina A. St. Louis-Vlasova and Paul R. Bohjanen

† Authors contributed equally.

Introduction

RNA-binding proteins that regulate gene expression at posttranscriptional levels do not usually act on a single target transcript, but coordinately regulate multiple transcripts, creating regulatory networks or regulons that are defined by RNA-binding proteins and their target transcripts. Regulons integrate intrinsic and extrinsic signals to coordinately modulate gene expression to regulate distinct cellular processes. The CUG-repeat binding protein 1 (CELF1) regulon coordinately regulates the expression of multiple genes at posttranscriptional levels. CELF1, a member of the CELF (CUGBP and embryonic lethal abnormal vision like factor) family of RNA binding proteins, was first identified as a protein that binds the CUG-repeat sequences of the myotonin protein kinase transcript[73]. In subsequent studies, CELF1 was shown to be multifunctional, regulating many posttranscriptional processes including alternative splicing, deadenylation, mRNA decay and translation reviewed in [188]. For example, alternative splicing events and translational control in muscle development are steered by the action of CELF1 [88, 93, 135, 189, 190]. In *Xenopus laevis* embryonal development, the CELF1 homologue embryo deadenylation element-binding protein (EDEN-BP) regulates translational repression in oocytes and in deadenylation of maternal RNAs in fertilized eggs [78]. Knockout of ETR1, the CELF1 homologue in C. elegans, is lethal and impairs muscle development [62]. Knockout of CELF1 in mice is lethal in most cases, but the few mice that are born display severe fertility defects [191]. Thus, the CELF1 regulon coordinates gene expression at multiple posttranscriptional levels.

To fulfill all these different functions, CELF1 interacts with different RNAs in different contexts and several different RNA target sequences have been proposed. In translational regulation, CELF1 binds to a CUG/CCG sequence in the 5' untranslated region (5'UTR) of the CCAAT enhancer binding protein β (C/EBPβ) to produce a different isoform of C/EBPβ [123]. CELF1 also binds to the AU-rich element of TNFα mRNA [68] and the GC-rich sequence in the 5'UTR of p21 mRNA [129]. In *Xenopus laevis* oocytes EDEN-BP interacts with the a U(A/G) repeat mRNA sequence leading to rapid deadenylation and translational activation [78].

Recently the sequence UGUUUGUUUGU, referred to as a GU-rich element (GRE), was identified as a CELF1 consensus binding sequence that mediated rapid mRNA decay. This UGUUUGUUUGU consensus sequence was significantly enriched in 3'UTRs of unstable mRNAs expressed in primary human T cells and functioned as an mRNA decay element when inserted into the 3' UTR of reporter transcripts through a mechanism that depended on binding by CELF1 [177]. In *Xenopus laevis,* similar GU-rich sequence elements are targets of EDEN-BP, [192], suggesting that this RNA-binding interaction was conserved over evolution. A GU-repeat sequence was also established as a CELF1 binding motif by yeast 3-hybrid screens [76], a SELEX approach [74], in vitro binding studies [77], and NMR structure analysis [89]. Although the GU-repeat sequence has been known to bind to CELF1 for several years, a biological consequence of this binding has not been demonstrated.

Although the RNA-binding activity and posttranscriptional regulatory

functions of CELF1 have been characterized, only a limited number of human CELF1 target transcripts have been identified. Since CELF1 appears to define an evolutionarily conserved posttranscriptional regulatory network that coordinates gene expression in human cells [188], we undertook a systematic approach to identify CELF1 target transcripts in human cells. We performed immunoprecipitation (IP) of CELF1 from HeLa cell cytoplasmic extracts and analyzed the co-immunoprecipitated transcripts using oligonucleotide microarrays. This technique has been used successfully to identify targets of other RNA-binding proteins including HuR [193], AUF1 [194], TIAR [195], TTP [196] and Pum1 [197]. Using this approach, we identified 613 putative targets of CELF1 and found significant enrichment of the consensus GRE sequence UGUUUGUUUGU as well as a GU-repeat sequence in the 3' UTR of the CELF1 target transcripts. We found the GU-repeat sequence functioned as a mRNA decay element and knockdown of CELF1 stabilized GRE-containing messages. This led us to redefine the GRE consensus sequence to include GU repeats. Functional analysis of GRE-containing CELF1 target transcripts revealed a posttranscriptional regulatory network that coordinates the expression of transcripts involved in cell cycle and cell growth regulation, cell motility, and apoptosis.

Results and Discussion

We recently showed that CELF1 mediated the rapid decay of GRE-containing transcripts in HeLa cells [177], and hypothesized that CELF1 coordinately regulates the posttranscriptional expression of a variety of GRE-containing transcripts at the level of RNA-decay [188]. We therefore performed experiments to identify transcripts

associated with CELF1 in human cells. In three independent experiments, cytoplasmic lysates from HeLa cells were immunoprecipitated with immobilized anti-CELF1, anti-hemagglutinin (HA), and anti-poly-A binding protein (PABP) antibodies. Input RNA and IP RNA were analyzed using Affymetrix U133a Plus-2 microarrays. Signal intensities from each microarray were normalized using the gene content robust multi-array average (GCRMA) algorithm. RNA from the anti-HA IP was considered to be a negative control. For microarrays performed on RNA from the input, CELF1 IP, or PABP IP, transcripts were considered to be present if the log intensity of their average signal yielded a p-value less than 0.05 by student T-test when compared to the log intensity values on RNA from the anti-HA IP. We identified 613 probe IDs in the RNA from the anti-CELF1 IP that were overrepresented compared to the HA control (Supplementary table 1). Of these 613 probe IDs, 75 were present in both the input and anti-PABP IP, 19 were present in the anti-PABP IP but not the input, and 59 were present in the input but not the anti-PABP IP. Of note, 477 transcripts were present in the anti-CELF1 IP but not the input or anti-PABP IP, suggesting that these transcripts were enriched during the anti-CELF1 immunoprecipitation procedure. Transcripts present in the anti-CELF1 IP encode proteins that serve various functions, including cell growth and cell cycle regulation, apoptosis, and motility. Table 3.1 lists a subset of the transcripts present in the anti-CELF1 IP that encode proteins that serve important cellular functions and the complete list of transcripts is found in supplemental Table 1. In a recent study in *Xenopus,* RNA IP experiments identified 158 target mRNAs of EDENBP, the CELF1 homologue in *Xenopus* [192]. These

EDENBP target transcripts had only minimal overlap with the transcripts we found associated with CELF1 in human HeLa cells, but GREs were enriched in both the *Xenopus* and human experiments. Interestingly, we found that most of the CELF1 target transcripts that we identified did not contain a GRE as defined previously [177], and we decided to use bioinformatical methods to search for other conserved sequences found in the transcripts from the anti-CELF1 IP.

We extracted the 3'UTR for each of the 613 CELF1 target transcripts and subjected these sequences to a simple overrepresentation algorithm. The overrepresentation algorithm created a list of all possible (4^{11}) 11-mers and counted the occurrence of each of these 11mers in the 3'UTRs of the genes of interest. Each 11-mer was determined to be present in a given 3'UTR if it was found in the sequence, allowing for only one mismatched nucleotide. The statistical significance of the overrepresentation of these 11-mers in the genes of interest was determined through comparison of the rate of appearance of the 11-mers in a background set of 27,472 3'UTRs from transcripts with a known sequence in the NCBI database. The six most frequently encountered 11-mers are shown in Table II. These consensus sequences included the previously elucidated GRE consensus sequence, UGUUUGUUUGU [177], with a prevalence of 18.69% in the 3' UTR from transcripts from the CELF1 IP compared to only 6.7% in all transcripts ($p = 5.35x10^{-26}$). We also identified a GU-repeat sequence that was highly enriched in transcripts from the anti-CUGP1 IP; this sequence was present in the 3' UTR in 17.33% of the transcripts from the anti-CUGBP1 IP, compared to 7.6% of all transcripts ($p = 2.53x10^{-15}$). The GU repeat

sequence was previously reported as a sequence capable of binding to CELF1 in yeast three-hybrid [76], surface plasmon resonance [77] and Selex assays [74]. In addition to the UGUUUGUUUGU and UGUGUGUGUGU sequences, we also identified a poly-U sequence that was highly enriched in transcripts from the anti-CELF1 IP. A GGA-repeat sequence, the sequence TGGGAATGGTC, and a CA-repeat sequence were also enriched in transcripts from the anti-CELF1 IP, with lower statistical significance. As another approach to look for conserved sequences in transcripts from the anti-CELF1 IP, a Bioprospector search was performed on the 3' UTRs of these transcripts. The parameters for the Bioprospector search included searching only on the forward strand, allowing for multiple occurrences of a motif in a given sequence, and using human 3' UTR sequences as the background. The results of the Bioprospector search included the GU-repeat sequence and the CA-repeat sequence shown in Table 2, which were enriched in transcripts found in the anti-CELF1 IP. The Bioprospector search also identified several other motifs, but none of the others were enriched in the anti-CELF1 IP in a statistically significant manner. These results confirmed our previous report that GRE-containing transcripts are targets of CELF1 and suggested that transcripts that contain GU-repeats could also be targets of CUGBP1.

We identified 6 distinct sequence motifs enriched in the 3'UTRs of transcripts found in the anti-CELF1 IP (Table 3.2) and selected two transcripts containing each of these sequences to determine if they could bind to CUGBP1 based on gel shift and supershift assays. For each transcript tested, we designed RNA oligomers that

contained the putative sequence elements and surrounding sequences. Using cytoplasmic extracts from HeLa cells, we found that CELF1 bound to the GRE sequence from JUNB proto-oncogene (JUNB) as well as the GU-repeat sequences from NADH dehydrogenase Fe-S protein 2 (NDUFS2) and Gelsolin (GSN) (Figure 1A and B). Binding was attributed to CELF1 because it was specifically supershifted with an anti-CELF1 antibody (Figure 3.1B, lanes 3, 7, and 11). This binding was sequence specific because binding did not occur when an RNA probe was used that contained mutations in the GU repeats (Figure 3.1B, lanes 14-16). Transcripts containing the poly U sequence, GGA-repeat sequence, CA-repeat sequence or the sequence TGGGAATGGTC did not bind to CELF1 (data not shown). These experiments confirmed our previous result that CELF1 bound to the JUNB GRE [177], and showed that CUGBP1 bound specifically to GU-repeat sequences of NDUFS2 and GSN. These results also confirmed results from others that CELF1 recognizes GU repeats [74, 76, 77]. From the 613 transcripts identified in the CUGBP1 IP, approximately 36% contain UGUUUGUUUGU or GU repeat motifs in their 3' UTRs.

We previously showed that the UGUUUGUUUGU consensus sequence promoted rapid degradation when inserted into the 3'UTR of an otherwise stable beta-globin reporter in a manner that depended on the presence of CELF1 [177]. We hypothesized that the GU-repeat sequence that was also enriched in transcripts found in the anti-CELF1 IP might also function to regulate CELF1-mediated mRNA decay. To test this hypothesis, we inserted GU-repeat containing sequences from the 3'UTRs

of NDUFS2, GSN and PPIC into the 3 'UTR of the beta-globin (BBB) reporter to produce the NDUFS2-GU, GSN-GU, and PPIC-GU reporters, respectively and transfected them into HeLa tet-off cells. Doxycycline was added to the media to stop transcription of the reporter transcripts. Total RNA was extracted after 0, 2, 4 or 6 hours and was analyzed by northern blot for beta-globin and GAPDH mRNA expression in order to measure the decay of the reporter transcripts (Figure 3.2A). We performed 4 independent experiments and graphed the mRNA decay results (Figure 3.2B). The beta-globin signal was normalized to GAPDH and the 0 time point was set to 100%. The percent of RNA remaining was plotted against time, and mRNA decay rates were calculated. As expected, the beta-globin transcript was stable, whereas the JUNB-GRE transcripts showed rapid decay with a half-life of 240 $\pm$ 24 minutes, confirming our previous results [177]. All three GU-repeat containing reporter transcripts, NDUFS2-GU, GSN-GU and PPIC-GU displayed rapid decay relative to the stable beta-globin reporter, suggesting that insertion of the GU repeat sequences was responsible for mediating mRNA decay. To confirm that the GU-repeats were responsible for mediating mRNA decay, we deleted or mutated the GU-repeat sequences to determine the effect on mRNA decay. First, we deleted the GU-repeat of the NDUFS2-GU and GSN-GU reporters to generate the NDUFS2-ΔGU and GSN-ΔGU reporters. Next, we mutated every other GUGU to ACAC to generate the NDUFS2-mGU and GSN-mGU reporters (Figure 3.1A). We tested the effect of these deletions and mutations on the ability of GU repeats to promote mRNA decay. Theses reporter plasmids were transfected into HeLa cells, reporter mRNA decay rates were

assessed by northern blot (Figure 3.3A and B), and results from 4 independent experiments were plotted (Figure 3.3C and D). The NDUFS2-GU and GSN-GU reporters exhibited rapid decay, whereas all of the reporters containing deletions or mutations in the GU repeats were stable. This series of experiments clearly show that the GU-repeat functions to promote decay of the beta-globin reporter transcripts.

The GU-repeats we found were obtained by a very stringent bioinformatical analysis allowing for only one mismatch per 11mer. To further characterize CELF1-GU-repeat binding, we obtained RNA oligomers containing several point mutations into the GSN GU-repeat (Figure 3.4A) corresponding to two mutations per 11mer. We performed shift and supershift experiments with these radiolabeled probes to assess the binding capability of each of the mutant oligomers (Figure 3.4B). As expected, GM1 and GM2 bound tightly to CELF1, as we recreated the GRE [177] or the EDEN-BP binding motif [78]. CELF1 specificity was confirmed by supershift using a specific anti-CELF1 antibody. Mutating the middle guanine to cytidine did not have an effect on CELF1 binding either. Disruption of the UGU binding structure as in GM6 and GM8 resulted in loss of supershift activity. Inserting a single AC-repeat in GM5 did not abrogate CELF1 binding giving rise to the possibility of discontinous GU-repeats capable of CUGBP1 binding as published earlier [68] which were not included in our analysis. Interestingly GM4 retained a slight binding activity and it has to be clarified whether this is sufficient to promote reporter decay. To confirm that the mutated sequences binding to CELF1 are able to promote mRNA decay, these mutations for GM2, GM3, GM4 and GM8 were introduced into the BBB-GSN-GU reporter plasmid

and transfected into Hela tet-off cells. Decay rates of these reporters were assessed by northern blot and the results of 4 independent experiments were analyzed. Against our expectations the G to A mutation (GM2) resulted in marked stabilization of the reporter (1040±12 min), whereas the G to C mutation (GM3) remained unstable (540±10 min). This might be due to the formation of double stranded RNA in GM2 under physiological conditions. The GM4 reporter construct, which showed some binding to CELF1, was very stable over time (half live of 1386±36 minutes). Interestingly, GM8 showed some decay activity (630±21 min), which is not caused by CELF1 binding which has to be investigated further.

We demonstrated previously that CELF1 promotes decay of GRE containing mRNAs [177]. To show that CELF1 is the force driving GU-repeat mediated decay, we performed siRNA knockdown experiments of CELF1. To assess GU-repeat reporter mRNA decay in the absence of CELF1 we performed sequential knockdown of CELF1 using two independent siRNAs and, as a control, red fluorecent protein siRNA followed by transfection of the GSN-GU reporter. Both siRNAs lead to a 90%-99% decrease in CELF1 levels (Figure 5.5B). In the same cells decay was measured over a 6 hour time course by northern blot (Figure 5.5A) and the results of 4 independent experiments was plotted against time (Figure 5.5C). Cells that were transfected with the control siRNA displayed a half life 442±12min for the GSN-GU reporter. Knockdown of CELF1 though significantly stabilized the GU-repeat containing messages ($p<0.01$) resulting in increased half lifes for both CELF1 specific siRNAs of 1040±15min or 1040±20min respectively. This experiment clearly shows

that the GU-repeat sequence and the GRE are regulated by a CELF1 dependent mechanism leading us to redefine the GRE.

We previously defined the GRE as the consensus sequence UGUUUGUUUGU, which was highly enriched in short-lived transcripts expressed in primary human T cells [177]. We showed that CELF1 could bind to this sequence in the context of endogenous transcripts and reporter transcripts and that that this sequence could function as a mediator of mRNA decay in a manner that depended on CELF1 [177]. Our results here suggest that CELF1 binds to a larger subset of transcripts, including transcripts that contain the previously defined GRE as well as transcripts that contain GU repeats. Our finding that GU repeats are able to function as mediators of mRNA decay, suggests that we should broaden our definition of the GRE to include GU repeats. Of the 613 transcripts that were identified in the anti-CELF1 IP, if we allow one mismatch, 123 transcripts contain the previously defined GRE 11-mer and 114 contain the GU repeat sequence. We combined these 237 transcripts into a set of transcripts that contain CELF1 binding sites and used Weblogo 2.8.2 Software [178] to create an new consensus GRE sequence based on all UTRs that contained the UGUUUGUUUGU element or the GU-repeat element (Figure 3.6).

The coordinate regulation of the expression of multiple genes involved in specific functional pathways by specific RNA-binding proteins and cis sequences forms regulatory networks that are known as posttranscriptional regulons. The 237 CELF1 target transcripts that contain the GRE consensus (shown in Figure 6) in their 3' UTRs define a posttranscriptional regulon. These GRE-containing transcripts

encode proteins that participate in specific functional pathways involved in cell growth, apoptosis, and cell motility. For example, transcripts encoding multiple regulators of cell cycle and cell growth were identified as CELF1 targets. Also, CELF1 targets included transcripts encoding multiple regulators of the extrinsic apoptotic pathway such as Akt and TP63 (Figure 3.7A), a function which was described previously [191]. Here the knockout of CELF1 lead to a decrease in male fertility likely caused by increased apoptosis of germ cells. To test this hypothesis we transfected HeLa cells twice within 24 hours with either a red fluorescent protein targeting siRNA duplex as control and two CELF1 specific siRNA duplexes described previously [177]. 48 hours after the second transfection cells were split for FACS analysis and western blot. Apoptosis was induced with TNFα and CHX for 3 hours. For FACS analysis cells were stained with a PE coupled antibody specific for activated caspase 3. Cells transfected with the control siRNA display significantly less apoptosis (19.03±0.19%) than the cells that were transfected the CELF1 specific siRNAs (25.95±0.19 and 27.65±0.45) in four independent experiments (Figure 3.7B). The knockdown of CELF1 resulted in a 90-99% reduction in CELF1 as demonstrated by western blotting (Figure 3.7B). Probing the same western blot for another apoptosis marker, cleaved PARP, further confirmed that CELF1 plays a role in apoptosis.

Besides the apoptosis pathway, we discovered CELF1 target transcripts encoding regulators of cell motility such as the small G-protein rho or gelsolin. Interestingly, cell growth and cell cycle, apoptosis, and cell motility, pathways play important roles in cancer pathogenesis, including cell transformation, abnormal

growth, cancer formation, progression, and metastasis [198]. Thus, the CELF1-GRE regulatory network may play an important role in regulating gene expression during normal cell growth and development by mediating mRNA degradation. Aberrations in the regulation of this network might lead to cancer formation and progression through the abnormal stabilization of critical transcripts.

Overall we have identified 613 new targets for CELF1 and demonstrate that a GU-repeat motif is sufficient to promote rapid decay when present in a transcript's 3'UTR. This GU-repeat and the previously identified UGUUUGUUUGU sequence were used to redefine the GU-rich element (GRE), and we found that GREs were present in the 3' UTR of 36% or 62% of the transcripts found in the anti-CELF1 IP if we allowed one or two mismatches, respectively. For this analysis, we focused on the 3' UTR because that is where known functional GREs are located. We acknowledge, however, that we could miss functional CELF1 target sites in the 5' UTR or coding regions of the transcripts. When we looked for GREs at any location in the target transcripts, allowing two mismatches, we found that 73% of the transcripts in the anti-CUGBP1 immunoprecipitation contained a GRE. Interestingly, although the GRE was highly enriched in the 3' UTR of target transcripts, it was not enriched in the 5' UTR or the coding regions. This led us to investigate the role of the GRE in translation by polysomal gradient analysis (Figure 3.8). Although CELF1 is reported to be involved in translational regulation, we could not find evidence that translational regulation is mediated by the GRE located in the 3'UTR.

Recently CELF1 was found to be a posttranscriptional regulator of TNF alpha

expression via the AU-rich element [199], which contained multiple dispersed UGU sequences that may function as CELF1-binding sites. If dispersed UGU sequences do indeed function as CELF1 binding sites, these sites would not be included in our consensus definition of the GRE. In addition to the GRE, CELF1 is known to bind to CUG repeat sequences [123], although this sequence was not identified in our motif search. Nevertheless, we searched for enrichment of a consensus sequences containing 4 tandem CUG repeats, allowing two mismatches, in the transcripts found in the anti-CELF1 IP. We found this sequences to be enriched in the 5' UTR ($p = 0.02$) and 3' UTR ($p = 6x10^{-4}$) of these transcripts corresponding to 48 and 100 transcripts, respectively. This CUG repeat sequence was not, however, enriched in the coding region. Thus, of the putative target transcripts contained in the anti-CELF1 IP, 83% percent contained GRE or CUG repeat consensus sequences somewhere in the transcript.

Overall, our data suggest that the GREs in the 3' UTR of transcripts are important targets of CELF1 and thereby function to regulate mRNA decay. The CELF1 target transcripts that we identified encode important regulators of cell cycle control, apoptosis and cell motility, as well as proto-oncogenes and other pathways involved in cancer formation and progression. These findings suggest that CELF1 and its target transcripts define a posttranscriptional regulatory network that functions to control cellular growth and homeostasis, and disruptions in this network may play a role in the development of cancer.

Table 3.1 Subset of transcripts present in the anti-CELF1 IP

RefSeq accession no.[a]	Gene name or description	Gene symbol	Sequence(s)[d]	*P* value[b]	GO biological function(s)[c]
NM_000177	Gelsolin (amyloidosis, Finnish type)	GSN	TGAGTGTGTGT, TGTGTGTGTGT	0.003	Actin filament polymerization, cel cycle
NM_002229	jun B proto-oncogene	JUNB	TGTGTTTGTGT	0.002	Cell growth, cell cycl
NM_004550	NADH Dehydrogenase (ubiquinone) Fe-S protein 2	NDUFS2	TGTGTGTGTGT	0.033	Mitochondrial electron transport
NM_000943	Peptidylprolyl isomerase C (cyclophilin C)	PPIC	TGTGTGTGTGT	0.049	Protein folding
NM_005171	Activating transcription factor 1	ATF1	TATGTGTGTGT, TGTGTGTGTGT	0.015	Cell growth, cell cycl
NM_005354	jun D proto-oncogene	JUND	TGTGTGTGTGT	0.0005	Cell growth, cell cycl
NM_001769	CD9 molecule	CD9	TTTTTGTTTGT, TGTTTGTTTTT	0.0001	Motility, cell growth
NM_014989	Regulating synaptic membrane exocytosis 1	RIMS1	TGTTTATTTGT, TGTTTGTTGGT	0.0017	Protein complex assembly
NM_012197	Rab GTPase activating protein 1	RABGAP1	TGTGTGTGTGT	0.008	Cell cycle
NM_004217	Aurora kinase B	AURKB	TGTTTGTATGT	0.045	Cytokinesis, cell cycl
NM_000641	Interleukin-11	IL-11	TGTTTGTTTTT	0.050	Cell-cell signaling
NM_018685	Anillin, actin binding protein	ANLN	TGTTTGTTTGT	0.043	Cytokinesis
NM_001894	Casein kinase 1, epsilon	CSNK1E	TGTGTGTGTGT	0.010	DNA repair
NM_001080421	unc-13 homolog A (*C. elegans*)	UNC13A	TGTGTGTGTGT, TGTTTGTTTTT	0.041	Exocytosis
NM_015083	Cdc2-related kinase, arginine/serine-rich	CRKRS	TGTGTGTGTGT	0.044	Protein amino acid phosphorylation
NM_002578	p21 (CDKN1A)-activated kinase 3	PAK3	TGTTTGTTTTT	0.009	Protein complex assembly
NM_001006610	Seven in absentia homolog 1 (*Drosophila*)	SIAH1	TGTGTGCGTGT	0.007	Proteolysis
NM_002468	Myeloid differentiation primary response gene 88	MYD88	TGGGTGTGTGT, GGTGTGT GTGT	0.022	Response to molecul of fungal origin
NM_005633	Son of sevenless homolog 1 (*Drosophila*)	SOS1	TGTTTGTGTAT	0.042	Signal transduction

[a] From the NCBI Reference Sequence Project.
[b] *P* value indicates the statistical significance of the enrichment of the transcript in the anti-CELF1 IP compared to the anti-HA IP.
[c] GO, gene ontology.
[d] Some transcripts contain more than one enriched GU-rich sequence.

Table 3.2. Prevalence of motifs in the 3'UTR of transcripts in the anti-CELF1 IP compared to all transcripts in the genome

Motif	Prevalence in the anti-CELF IP (%)	Prevalence in the genome (%)	P value
TGTGTGTGTGT	17.33	7.6	2.53×10^{-15}
GGAGGAGGAGG	8.51	5.4	1.02×10^{-2}
TGTTTGTTTGT	18.69	6.7	5.35×10^{-26}
TTTTTTTTTT	38.28	21.8	8.03×10^{-8}
TGGGAATGGTC	2.58	0.8	6.28×10^{-6}
CACACACACAC	7.75	4.9	1.40×10^{-3}

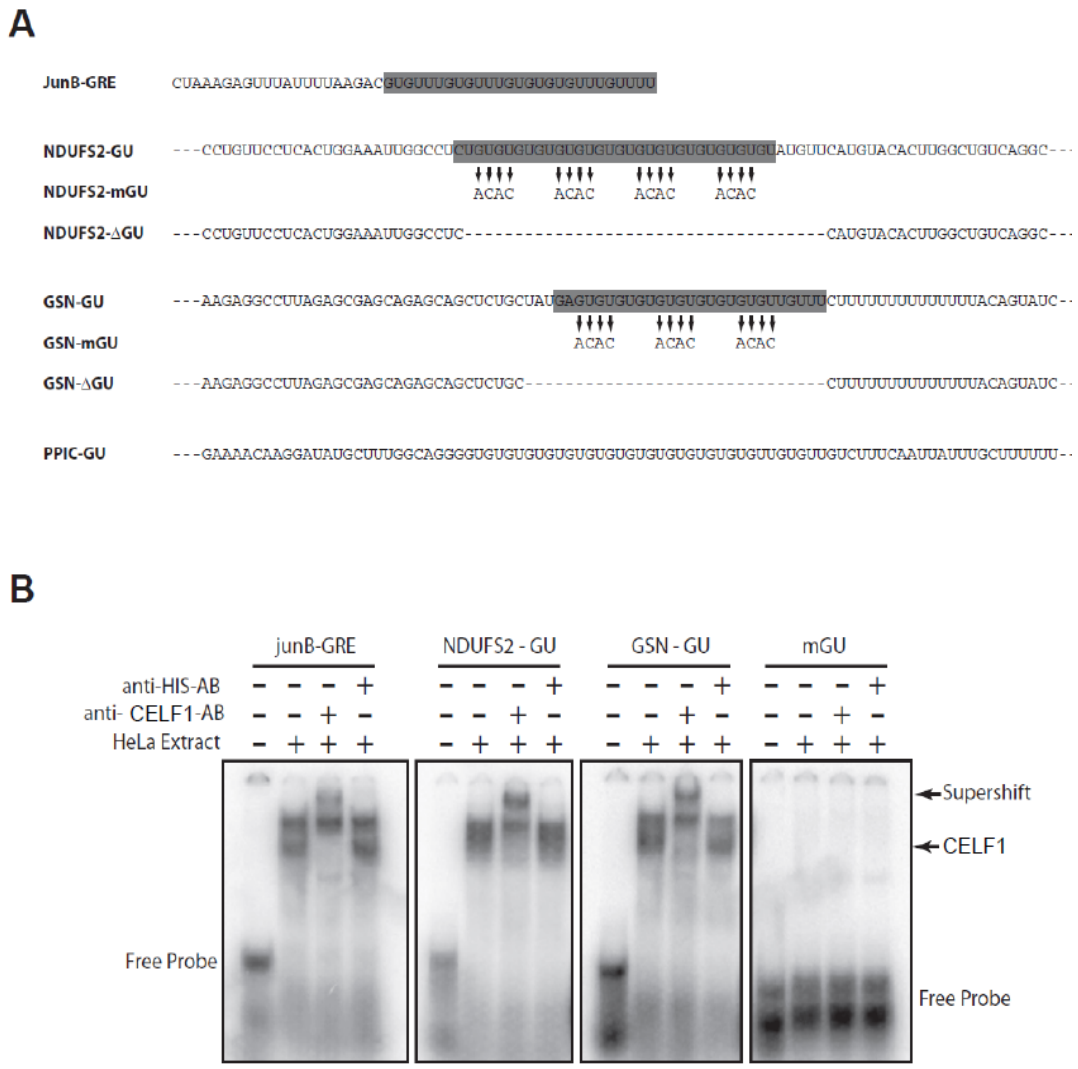

Figure 3.1: CELF1 selectively binds to GU-repeat sequences. A. Sequences from the 3' UTR of transcripts present in the anti-CELF1 IP were cloned into the 3' UTR of the beta-globin reporter to create the JUNB-GRE [177], NDUSF2-GU, GSN-GU, and PPIC reporter transcripts. The indicated mutations and deletions were introduced to create the NDUSF2-mGU, NDUSF-ΔGU GSN-mGU, and GSN-ΔGU transcripts. **B.** The GRE sequence of JUNB (JUNB-GRE), the GU-repeat sequences of NDUFS2 (NDUFS2-GU), GSN (GSN-GU), or a mutated GU-repeat sequence from NDUFS2

(mGRE) shown in the shaded boxes in Figure A were used as radiolabeled RNA probes in EMSA experiments. Anti- CELF1 or anti-His antibodies were added to the indicated reactions. RNA-protein complexes were separated by electrophoresis under nondenaturing conditions. Gels were dried and analyzed on a phosphoimager. The positions of migration of the free probe, CELF1-containing band (CELF1), and the anti-CELF1 supershifted band (Supershift) are indicated.

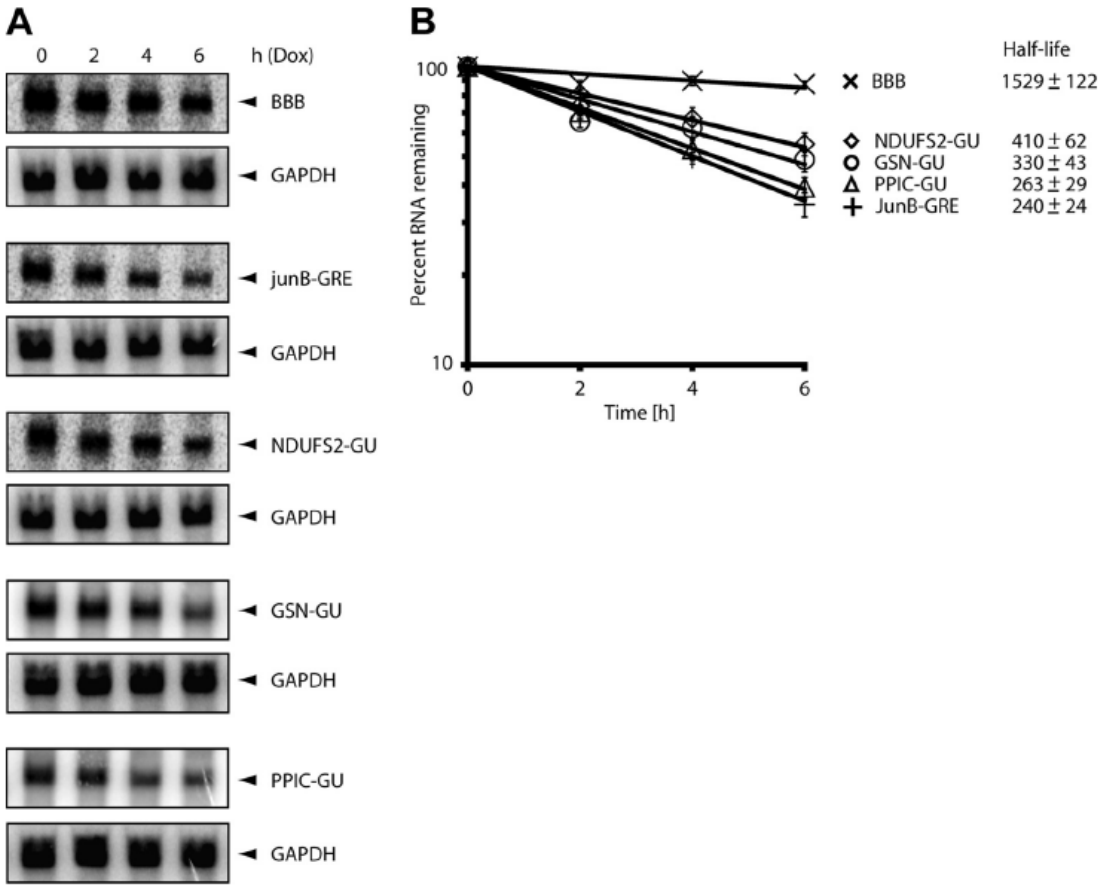

Figure 3.2: The GU-repeat sequence mediates rapid decay of the beta-globin reporter transcript. A. Hela Tet-off cells were transfected with the pTetBBB beta-globin reporter construct or with reporter constructs in which GRE-containing sequences from the 3' UTR of the JUNB transcript (JUNB-GRE) or GU-repeat sequences from the 3' UTRs of the NDUSF2 (NDUSF2-GU), GSN (GSN-GU), or PPIC (PICC-GU) were inserted into the beta-globin 3' UTR. Doxycycline was added to the medium to stop transcription from the tet-responsive promoter and total cellular RNA was collected at 0, 2, 4, and 6 h time points. Northern blot analyses were performed to monitor the levels of the GAPDH transcripts and the beta-globin reporter transcripts. **B.** The experiment shown in (A) was performed 4 times, and the northern

blot signals were quantified using a phosphorimager. For each time point, the intensity of the beta-globin band was normalized to the intensity of the GAPDH band, and the band intensity at the 0 time point was set to 100%. The percent of mRNA remaining was plotted over time. The error bars indicate the standard error of the mean (SEM) from three experiments.

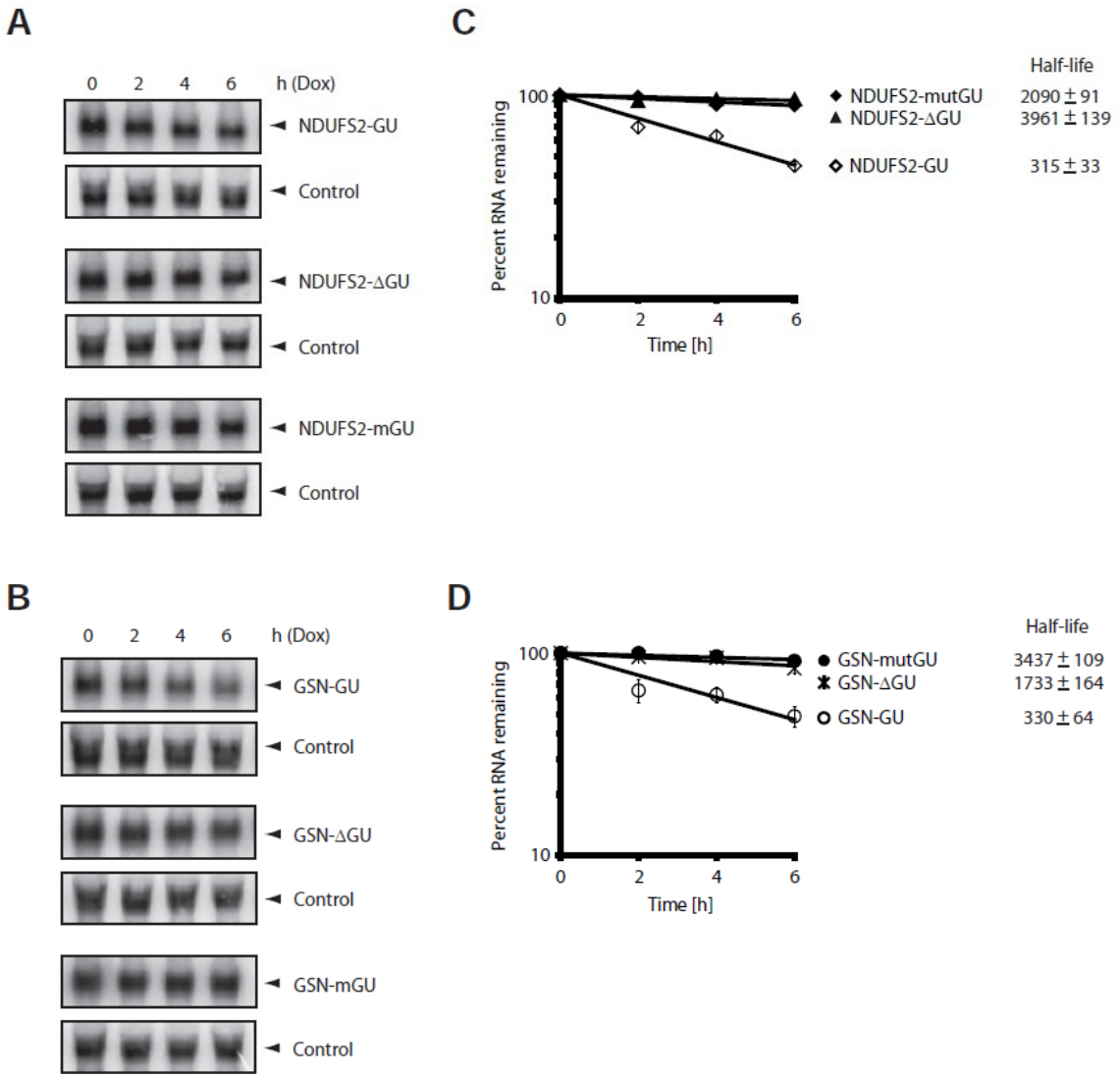

Figure 3.3: Mutation of the GU-repeat sequence abrogates mRNA decay. A and B. The decay of the NDUSF2-GU **(A)** or the GSN-GU **(B)** beta-globin reporter was compared to the decay of these reporter in which the GU-repeats were deleted (NDUSF2-ΔGU and GSN-ΔGU) or mutated (NDUSF2-mGU and GSM-mGU). Hela Tet-off cells were transfected with the indicated beta-globin reporter constructs. Doxycycline was added to the medium to stop transcription from the tet-responsive promoter and total cellular RNA was collected at 0, 2, 4, and 6 h time points.

Northern blot analyses were performed to monitor the levels of the GAPDH transcripts and the beta-globin reporter transcripts. **C and D.** The experiments shown in **(A)** and **(B)** were performed 4 times, and the northern blot signals were quantified using a phosphorimager. For each time point, the intensity of the beta-globin band was normalized to the intensity of the GAPDH band, and the band intensity at the 0 time point was set to 100%. The percent of mRNA remaining was plotted over time. The error bars indicate the standard error of the mean (SEM) from three experiments.

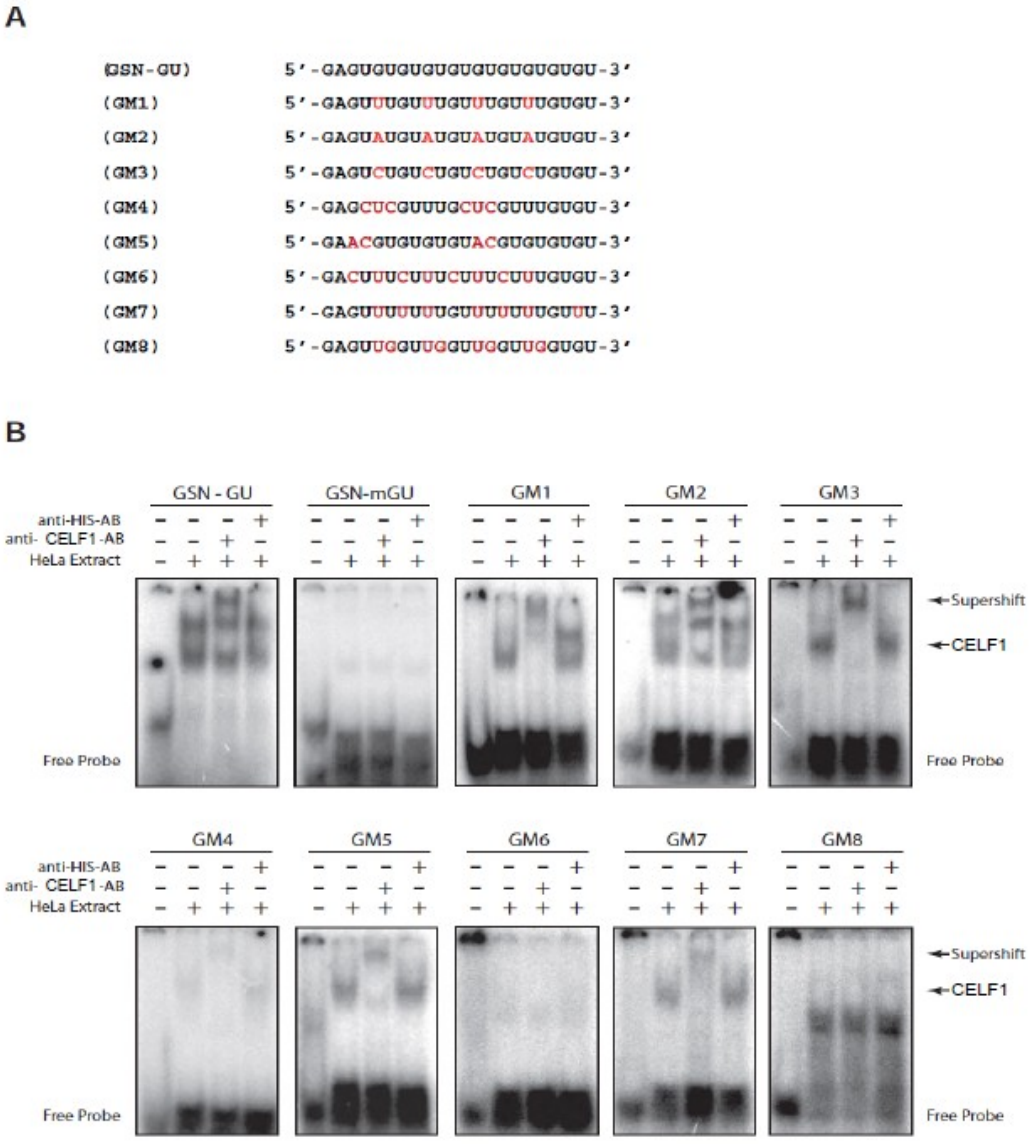

Figure 3.4: CELF1 binding activity allows several mismatches in the GU-repeat sequence. A. Mutations and were introduced into the GSN-GU motif to create GM1, GM2, GM3, GM4, GM5, GM6, GM7 and GM8 as indicated. Mutated nucleotides are depicted in red. **B.** The sequences depicted under A. were used as radiolabeled RNA probes in EMSA experiments. The GSN-mGRE sequence was used as a negativ control. Anti- CELF1 or anti-His antibodies were added to the indicated reactions.

RNA-protein complexes were separated by electrophoresis under nondenaturing conditions. Gels were dried and analyzed on a phosphoimager. The positions of migration of the free probe, CELF1-containing band (CELF1), and the anti-CELF1supershifted band (Supershift) are indicated.

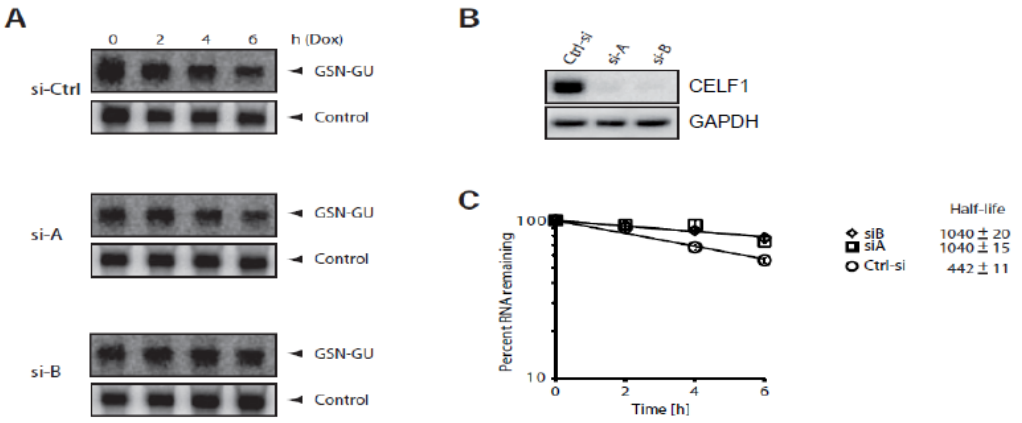

Figure 3.5: CELF1 is responsible for GU-repeat mediated mRNA decay. A. The decay of the GSN-GU beta-globin reporter cotransfected with a control siRNA was compared to the decay of these reporters when CELF1 was knocked down with two different siRNAs (siA and siB). Hela Tet-off cells were transfected with the GSN- GU repeat-containing reporter and the indicated siRNAs. Doxycycline was added to the medium to stop transcription from the tet-responsive promoter and total cellular RNA was collected at 0, 2, 4, and 6 h time points. Northern blot analyses were performed to monitor the levels of the GAPDH transcripts and the beta-globin reporter transcripts. **B**. Knockdown efficiency was monitored for each experiment by western blotting with a specific anti-CELF1 antibody. A GAPDH antibody was used as the loading control. **C.** The experiments shown in **(A)** was performed 4 times, and the northern blot signals were quantified using a phosphorimager. For each time point, the intensity of the beta-globin band was normalized to the intensity of the GAPDH band, and the band intensity at the 0 time point was set to 100%. The percent of mRNA remaining was plotted over time. The error bars indicate the standard error of the mean (SEM) from three experiments.

Figure 3.6: Generation of the GRE consensus sequence. Weblogo was generated by compiling a list of all occurrences of the GU repeat and the GRE found in the 3'UTR of those transcripts identified as present in the CELF1 IP. This list was uploaded into the program Weblogo version 2.8.2 to generate the weblogo [178].

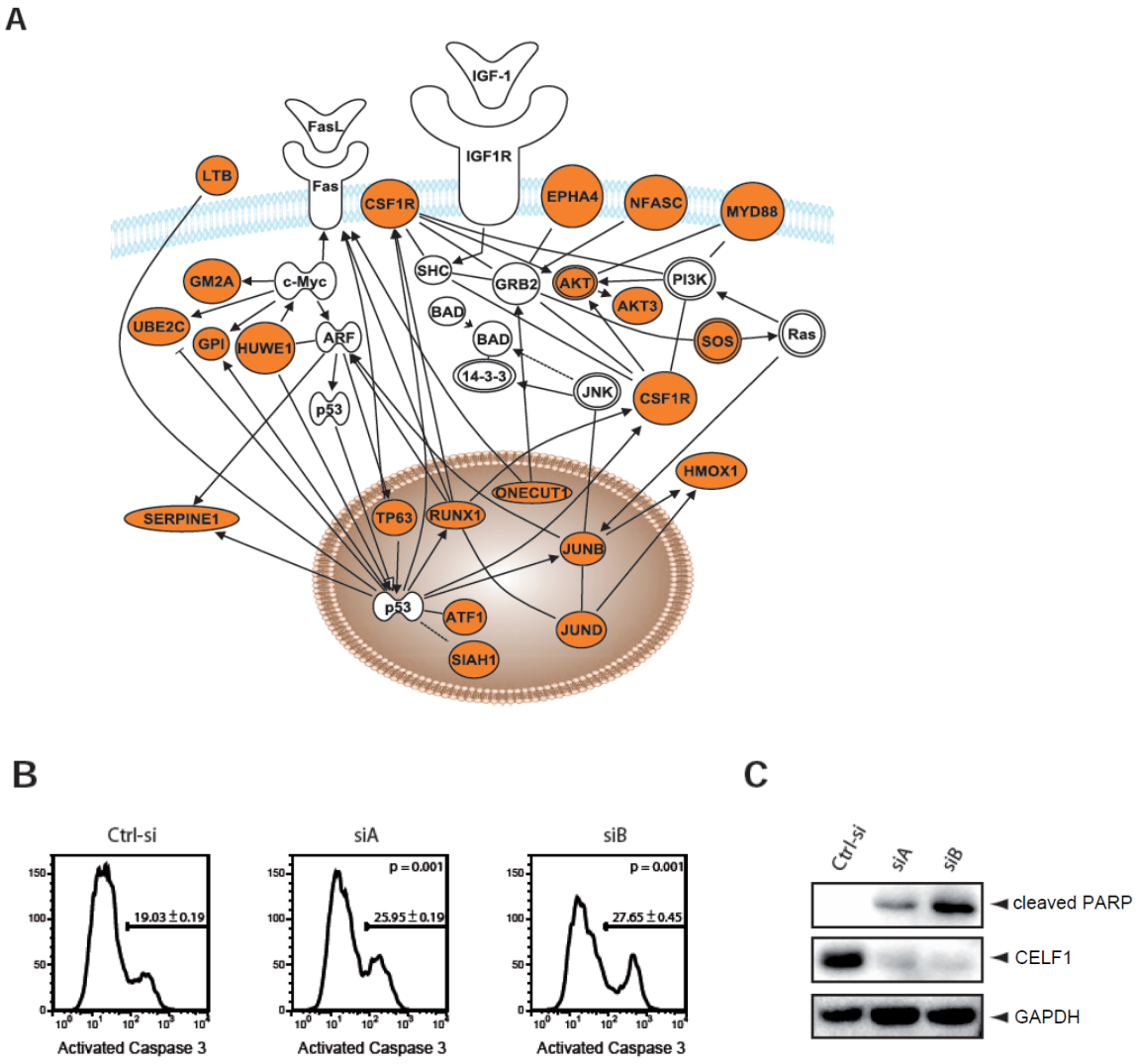

Figure 3.7: Posttranscriptional regulation of the CELF1 target network. A. The network diagram depicts the coordinate regulation of CELF1 target transcripts involved in apoptosis. Transcripts depicted in orange represent GRE-containing transcripts that were enriched in the CELF1 targets. These pathways were identified by Ingenuity Pathway Assistant software (Ingenuity Systems, CA). **B.** Hela Tet-off cells were transfected twice within 24h with either a control siRNA or two different siRNAs specific for CELF1 (siA and SiB). Cells were stained for FACS with a PE-

labeled anti-active Caspase 3 antibody for FACS analysis. A representative experiment is shown. Four independent experiments were analyzed and average percentages of apoptosis, standard error and p-values are indicated. **C.** Knockdown efficiency was monitored by western blotting with a specific anti-CELF1 antibody. A GAPDH antibody was used as the loading control. Apoptosis was reassessed with a specific anti-PARP antibody.

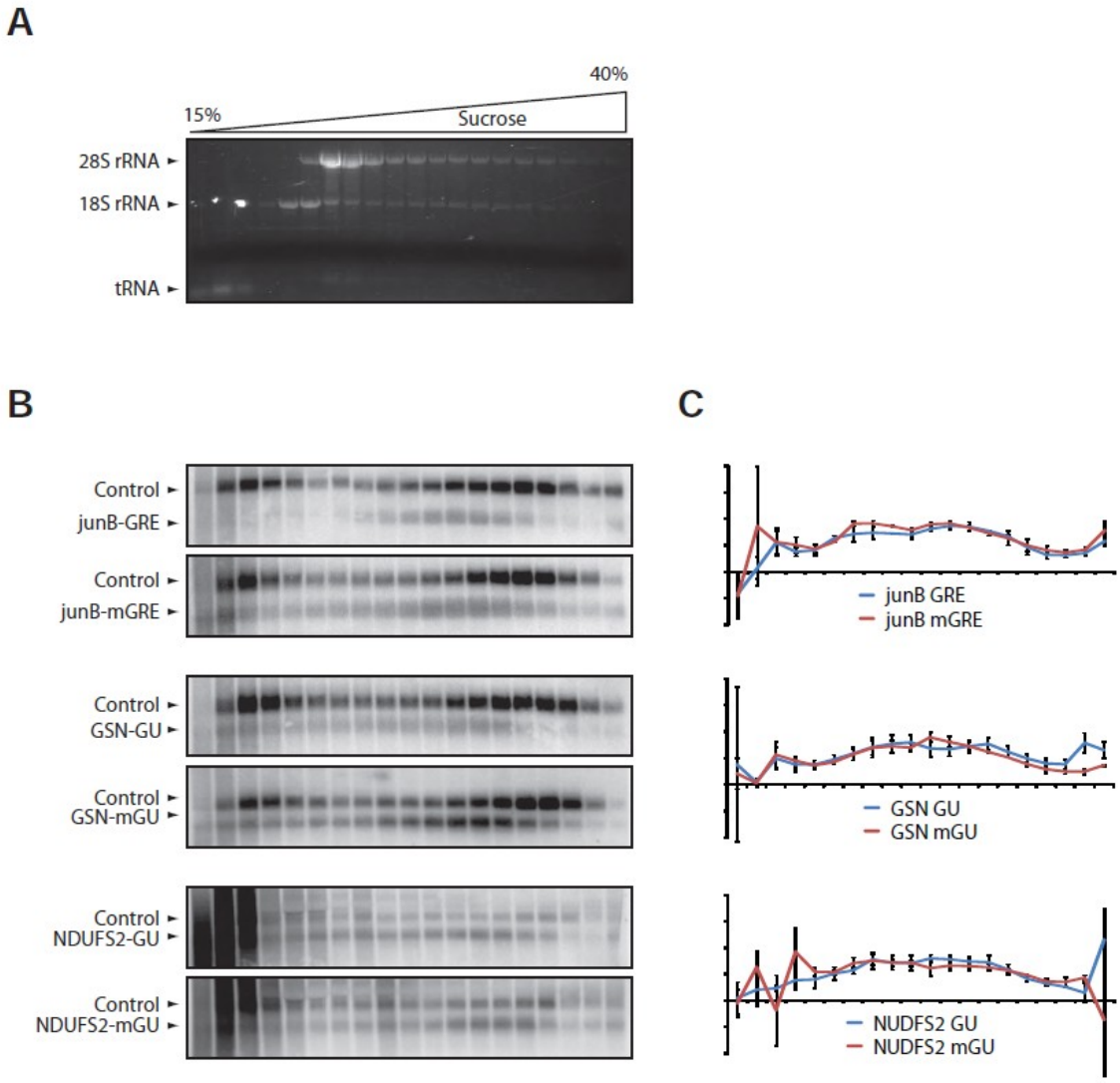

Figure 3.8: CELF1 binding to the GRE does not impact translational control. A. Translation of the junB-GRE, NDUFS2-GU and GSN-GU beta-globin reporter was compared to translation of the respective mutant constructs junB-mGRE, NDUFS2-mGU and GSN-mGU. For this, Hela Tet-off cells were transfected with the indicated reporter genes along with the pTracerC-EF/V5-His/*lacZ* GFP expression plasmid. 48 hours after transfection cytoplasmic extracts were subjected to analysis on a 15%-40% sucrose gradient. RNA was extracted from each of the 20 fractions and RNA was

separated on a 1.8% denaturing agarose gel. A representative gel is shown. **B**. Northern blot analyses were performed to monitor the levels of the control GFP transcripts and the beta-globin reporter transcripts. **C.** The experiments shown in (B) were performed 3 times, and the northern blot signals were quantified using a phosphorimager. For each fraction the intensity of the beta-globin band was normalized to the intensity of the GFP band. The fraction with the highest beta-globin-GFP ratio was set to 1. The distribution of the beta-globin reporter was plotted over the fraction number. The error bars indicate the standard error of the mean (SEM) from three experiments.

Chapter 4

Regulation of CELF1 Binding to Target Transcripts upon T Cell Activation

Daniel Beisang, Bernd Rattenbacher, Irina A. Vlasova-St. Louis, Paul R. Bohjanen

Introduction

The activation and clonal expansion of human T cells during an immune response requires rapid and precise changes in gene expression that are regulated at multiple levels through transcriptional and posttranscriptional mechanisms [200]. The molecular events leading to T cell activation must be tightly controlled to prevent the development of disease states such as autoimmunity or malignancy [201-203], and it is becoming increasingly clear that post-transcriptional gene regulation at the level of mRNA degradation is critical for normal cellular activation, proliferation, and immune effector function [204, 205]. Indeed, over half of the gene expression changes in early T cell activation are a result of changes in mRNA half-life [16]. The significance of mRNA decay regulation is highlighted by the fact that T cell malignancies display abnormal stabilization of numerous transcripts that encode proteins that promote cellular growth and proliferation [8].

RNA-binding proteins or microRNAs bind to specific recognition motifs in mRNA and coordinately regulate the posttranscriptional fate of networks of genes involved in cellular responses [2]. The best characterized example of a posttranscriptional regulatory network that coordinately regulates gene expression during immune responses is AU-rich element (ARE)-mediated mRNA decay. AREs are conserved sequence elements found in the 3' untranslated region (UTR) of transcripts encoding numerous cytokine transcripts and other inflammatory mediators, and AREs function to coordinately regulate mRNA decay during immune responses by interacting with cytoplasmic ARE-binding proteins [183, 206]. Another recently described

posttranscriptional regulatory network involves the RNA-binding protein CUG-triplet repeat binding protein 1 (CUGBP1), also referred to as CUGBP and ELAV Like Family member 1 (CELF1), which binds to a GU-rich element (GRE) residing in the 3'UTR of target transcripts and mediates coordinate degradation of GRE-containing transcripts [50]. The GRE was originally identified as a sequence that was highly enriched in the 3'UTR of transcripts that decayed rapidly in primary human T cells and was shown to function as a regulator of mRNA decay [50]. Based on a bioinformatic analysis of mRNA targets of CELF1 in HeLa cells, the GRE was defined to be the consensus sequence UGU[G/U]UGU[G/U]UGU [184]. Binding by CELF1 to certain GRE-containing transcripts has been shown to promote their rapid degradation, but how this process is regulated during cellular activation is poorly understood.

In addition to regulating mRNA degradation in the cytoplasm, CELF1 has other functions as a regulator of alternative splicing and translation [52]. In the nucleus, CELF1 regulates the alternative splicing of a number of transcripts [94, 207], whereas in the cytoplasm CELF1 binds to the untranslated regions of transcripts and regulates their translation efficiency or stability [52]. In addition to binding to GREs in the 3' UTR of mRNA, CELF1 can bind to some transcripts that contain a GC-rich element in their 5'UTR and promote their translation [141, 208]. The function of CELF1 is regulated by phosphorylation. In a mouse model of myotonic dystrophy it has been shown that CELF1 is phosphorylated by protein kinase C (PKC), resulting in altered cellular distribution and stability of the CELF1 protein, correlating with altered splicing patterns [140, 145]. The CELF1 target transcript TNF-α was stabilized upon chemical activation

of the PKC pathway [144]. In addition, the RNA binding specificity of CELF1 has been shown to be altered via phosphorylation by cyclin D3-cdk4/6 [82]. These results suggest that CELF1 is regulated by phosphorylation which could provide a mechanism for activation-induced changes in CELF1 function during cellular processes such as T cell activation.

Here, we used RNA-immunoprecipitation (RNA-IP) followed by microarray analysis [209] to investigate the cytoplasmic target transcripts of CELF1 in resting and activated primary human T cells. We found that CELF1 target transcripts in resting and activated T cells were highly enriched for the presence of the GRE in their 3'UTRs, but the number of CELF1 target transcripts decreased dramatically following T cell activation. The decrease in the number of CELF1 targets upon T cell activation was caused by activation-dependent phosphorylation of CELF1 and decreased ability of CELF1 to bind to GRE-containing RNA. A large percentage of CELF1 target transcripts exhibited rapid and transient up-regulation and a smaller percentage exhibited transient down-regulation following T cell activation. Many of the transiently up-regulated CELF1 target transcripts encode important regulators necessary for transition from a quiescent state to a state of cellular activation and proliferation. Our data support a model whereby CELF1 phosphorylation regulates a network of transcripts involved in T cell activation. Overall, our results verify that CELF1 binds to GRE-containing target transcripts in primary human T cells and that its ability to bind to mRNA is altered following T cell activation.

Results

GREs are enriched in CELF1 target transcripts in primary human T cells. We recently showed that the GRE was highly enriched in short-lived transcripts expressed in primary human T cells [50] and that CELF1 bound to the GRE and mediated the decay of GRE-containing transcripts [184]. The transcriptional network targeted by CELF1 and the regulation of CELF1 function in T cells remains poorly understood. In order to better understand the role of CELF1 in T cell function, we sought to identify the cytoplasmic mRNA targets of CELF1 in primary human T cells using RNA-IP followed by analysis of targets using Affymetrix U133A plus 2.0 microarrays. Immunoprecipitation experiments were performed in triplicate using antibodies targeting CELF1, poly-A binding protein (PABP), or hemaglutinnin (HA). For each experiment, transcript abundance in input RNA and immunoprciopitated RNA was analyzed using Affymetrix U133A plus 2.0 microarrays. The results of these analyses revealed that CELF1 in resting primary human T cells was associated with 1245 unique transcripts, corresponding to 1309 Affymetrix probe sets. A subset of these target transcripts were verified to be CELF1 targets using traditional RT-PCR since an anti-CELF1 antibody specifically co-immunoprecipitated JunD, topoisomerase I, and ubiquitin-protein ligase E3A transcripts in cytoplasmic extracts from resting T cells (figure 4.1, lane 5). GAPDH transcript, which is not predicted to be a CELF1 target transcript, was not copurified with CELF1. Immunoprecipitation with an antibody against HA did not copurify any of these transcripts (lane 3), whereas an antibody against PABP copurified all of these transcripts (lane 1).

Transcripts determined to be targets of CELF1 encoded proteins serving various functions including mRNA processing, metabolism of RNA, protein turnover, as well as regulators of cell death and proliferation. Transcripts encoding RNA-binding proteins were particularly enriched in the CELF1 targets, suggesting that in resting T cells CELF1 may be functioning as a posttranscriptional "regulator of regulators" whereby CELF1 influences the expression of a network of target transcripts encoding RNA binding proteins which in turn regulate individual sub-networks of transcripts. A partial list of important CELF1 target transcripts is shown in Table 4.1, and a full list can be found in Supplementary Table 2.

CELF1 binds preferentially to a GU-rich RNA motif, UGU[G/U]UGU[G/U]UGU, found in the 3'UTR of target transcripts [184]. We sought to determine whether our experimentally determined CELF1 target transcripts in T cells were enriched for this motif. Genbank files were downloaded from the NCBI database, and custom C++ scripts were written to extract the 5'UTRs, coding sequences, and 3'UTRs from CELF1 target transcripts as well as all transcripts present in the genome (total human transcriptome), and all transcripts present in the resting human T cell transcriptome, determined as described in Materials and Methods. A bioinformatic search of these sequences for the GRE allowing one or two mismatches was performed, and no enrichment of the GRE was found in the 5'UTR or coding region of the target transcripts compared to either the total human transcriptome or the resting T cell transcriptome. In contrast, a significant enrichment of the GRE was found in the 3'UTR of CELF1 target trasncripts compared to both the total human transcriptome and the T cell transcriptome.

These differences were highly statistically significant ($p<0.001$ for both zero and one mismatch, chi squared test). Interestingly, the resting T cell transcriptome exhibited a small but significant enrichment in GRE-containing transcripts ($p<0.001$ for both zero and one mismatches, chi squared test) when compared to the total human transcriptome frequency, suggesting that GRE-containing transcripts may be expressed preferentially in T cells. To control for specificity of the CELF1-GRE interaction, we investigated the prevalence of the 11mer ARE (UAUUUAUUUAU) allowing for no mismatches, and found no enrichment in the CUGBP1 IP compared to the resting T cell transcriptome or the total human transcriptome. Similar to the GRE, we observed a trend toward enrichment of the ARE in the T cell transcriptome compared to the total human transcriptome. It has been previously reported that CELF1 also binds to a CUG/CCG sequence, such as that found in the 5'UTR of the CCAT/enhancer-binding protein β [141]. We did not observe enrichment of this motif in the 5'UTR, coding region, or 3'UTR of CELF1 target transcripts.

As an independent approach to identify CELF1 target sequences, we performed de novo motif searches on the 3'UTRs of the experimentally determined CELF1 target transcripts. First, the 3'UTRs were submitted to a simple over-representation analysis where the frequency of all oligonucleotide sequences of length k (k-mers) in the target 3'UTRs was compared to the frequency of the same k-mers in the total human transcriptome for k values from 8 to 13. Using custom C++ scripts, k-mers were then sorted by the difference of the two frequencies to identify the most highly enriched k-mers. We found that several GRE-like sequences were the most highly enriched motifs

(table 4.2, K-mer Analysis). These results corroborated previous reports showing that both $(UGUU)_n$ motifs and $(UG)_n$ motifs were enriched in the 3'UTRs of CELF1 target transcripts [71, 184]. Additionally, the 3'UTRs from CELF1 target transcripts were submitted to a BioProspector motif search [210], to evaluate enriched sequences. The top result from this search was also a GRE-like sequence (table 4.2, BioProspector Analysis). Overall, these results confirm that CELF1 has preference for binding to GREs.

Only a subset of CELF1 target transcripts exhibit short half-lives. CELF1 has been shown to promote the instability of reporter transcripts that contain the GRE in their 3'UTRs [184]. Furthermore, short-lived transcripts in primary human T lymphocytes were significantly enriched for transcripts containing a GRE in their 3'UTR [50]. We sought to determine whether CELF1 target transcripts in primary human T cells exhibited short half-lives. To answer this question, we utilized the results from a previously published global assessment of mRNA half-lives in resting and activated primary human T lymphocytes [17] . Briefly, primary human T cells were either stimulated for 3 hours with anti-CD3 and anti-CD28 antibodies or allowed to rest, and transcription was halted by administration of actinomycin D. Total cellular RNA was then isolated at 0, 1.5 and 3 hours. Transcript abundance was assessed using U95 Affymetrix microarrays, and half-lives were calculated based on a first order decay model. The U133A+2.0 Affymetrix probeset IDs of the CUGBP1 targets were converted to the corresponding U95 Affymetrix probeset ID using the ID converter tool available from Babylomics. We identified 198 short-lived GRE-containing CUGBP1 target transcripts in resting T cells with half lives less than 180 minutes. Ingenuity Pathway Analyst was used to investigate

the functions of the proteins encoded by these short-lived target transcripts, and this analysis revealed an enrichment of transcripts involved in cell death (p=1.4E-4), protein degradation/ubiquitination (p=0.02), and proliferation (p=0.002). All of these cellular functions must be turned on upon T cell stimulation and play a crucial role in the acquisition of an activated phenotype. CELF1 may be functioning to maintain these transcripts at a low cellular abundance in resting T cells in order to maintain the quiescent phenotype. Some of these short lived transcripts along with their respective half-lives and biological function can be found in table 4.1.

Although a subset of CELF1 target transcripts exhibited rapid decay, most CELF1 target transcripts were relatively stable in resting T cells. The average half-life of CELF1 target transcripts in resting T cells was 1405 minutes, whereas the average half life of all transcripts expressed in resting T cells was 1588 minutes. This data shows that CELF1 targets showed a statistically significant skewing toward shorter half-lives (p=0.001, Welch's t-test) consistent with the observation that CELF1 promotes transcript decay. Interestingly, when comparing CELF1 target transcripts which do and do not contain a consensus GRE in their 3'UTR, it was observed that GRE-containing transcripts had shorter half-lives in resting T cells than those which did not contain a GRE (p=0.029, Welch's t-test). No difference was found in the average half life of those transcripts containing the UGUUUGUUUGU version of the GRE versus the GU-repeat version of the GRE.

Altered CUGBP1 association with target transcripts upon T cell activation. We next sought to compare the targets of CELF1 in resting T cells to T cells that were

stimulated for 6 hours with anti-CD3 and anti-CD28 antibodies in order to determine if CELF1 played a role in the dynamic changes in gene expression that occur following T cell activation. Using the same RNA-IP approach followed by microarray analysis as described above, we discovered that CELF1 associated with only 168 unique transcripts, corresponding to 153 probe sets in activated T cells. Comparing the CELF1 target transcripts in resting T cells to activated T cells we found that there was an overlap of these populations of only 53 probe sets, suggesting that most of the 1309 probe sets that associated with CELF1 in resting T cells were no longer associated with CELF1 after T cell activation. Of the 1256 probesets bound by CELF1 only in resting T cells, 1219 (97%) were present in the transcriptomes of both resting and stimulated T cells, suggesting that the lack of association with CELF1 in activated cells was not due to the disappearance of these transcripts. Using traditional RT-PCR, we verified that a subset of these transcripts showed decreased association with CELF1 in immunoprecipitates from activated T cells (figure 4.1, compare lanes 5 and 6), despite the fact that the association of these transcripts with PABP increased (compare lanes 1 and 2). To more quantitatively examine the effect of T cell activation on the relationship between the change in the abundance of a transcript in the cell relative to its abundance in the CELF1 RNA-IP, we defined a Fold Change in Enrichment (FCE) for each transcript on the microarray as follows:

$$FCE = \frac{(\text{CELF1-IP/input})_{\text{stimulated}}}{(\text{CELF1-IP/input})_{\text{resting}}}$$

where the ratio of the microarray signal from the CELF1 RNA-IP to the microarray signal from input RNA for stimulated T cells is divided by the same ratio for resting T

cells. The average FCE for all microarray probe sets was 1.084 suggesting the association of most transcripts with CELF1 did not change following T cell activation. In contrast, the average FCE for the CELF1 target transcripts was 0.422, suggesting that changes in abundance of these transcripts in the cell could not explain their lack of association with CELF1. The distribution of FCE for both the T cell transcripts and the CELF1 target transcripts is shown in figure 4.3.

Interestingly, although most CELF1 target transcripts bound to CELF1 only in resting T cells, we identified 100 probe sets that were bound by CELF1 only in stimulated T cells but not in resting T cells. Of these, 40 probe sets were induced by T cell activation and were not present in the resting condition. The remaining 60 probe sets were expressed in both the resting and stimulated conditions, but were only bound in the stimulated condition. A list of a subset of CELF1 target transcripts in stimulated cells can be found in table 4.3. We investigated the prevalence of the GRE, allowing for zero or one mismatches, in the 3'UTRs of the CELF1 target transcripts in stimulated T cells as described above. As can be seen in figure 4.2B, we found enrichment of transcripts containing a GRE in the CELF1 targets compared to the activated T cell transcriptome or the total human transcriptome ($p<0.001$ for both zero or one mismatch, chi squared test). Among transcripts bound in both resting and activated T cells, 46 (82%) contained a GRE, whereas among transcripts bound in only stimulated cells, but expressed in both resting and stimulated conditions, 38 (68%) contained a GRE. Finally, in those transcripts bound only in stimulated cells, and newly expressed in the stimulated condition, 20 (54%) contained a GRE. Ingenuity pathway analysis showed the biological

pathways of RNA metabolism, apoptosis, and cell cycle control were enriched in the CELF1 targets in stimulated T cells.

CELF1 target transcripts exhibit changes in gene expression upon T cell activation. We used previously generated data which assessed gene expression using Affymetrix U133A microarrays in primary human T cells stimulated with anti-CD3/28 antibodies for 0 min, 30 min, 1 hr, 3 hr, 6 hr, 12 hr, 24 hr, and 48 hr [14], to determine the patterns of expression of CELF1 targets following T cell activation. As can be seen in figure 4.4A, we found that T cell transcripts exhibited three predominant patterns of expression, which we have termed "Unchanged", "Up", or "Down". Our findings showed that most transcripts in the T cell transcriptome were "Unchanged", consistent with recent studies suggesting that approximately 80% of the transcriptome is ubiquitously expressed [211], and thus would not be expected to change in abundance upon stimulation. As can be seen in figure 4.4B, we found that CELF1 target trasncripts expressed in resting cells were statistically enriched for transcripts that showed significant activation-induced regulation, exhibiting both the "Up" ($p<2.2x10^{-16}$, chi squared test) and "Down" ($p=2.2x10^{-9}$, chi squared test) patterns of expression, with a strong preference for the "Up" pattern. Similarly, CELF1 target transcripts in stimulated cells were enriched for transcripts following the "Up" pattern of expression ($p=1.1x10^{-6}$, chi squared test). Of those transcripts which followed the "Up" pattern of expression, there was a strong enrichment of transcripts coding for proteins which participate in pre-mRNA processing and alternative splicing, including several heterogeneous ribonucleoproteins (hnRNPs). We validated the expression profile of three of the

transcripts encoding RNA-binding proteins following the "Up" pattern and one following the "Down" pattern using qRT-PCR. As can be seen in table 4.4, the fold change in expression after 6 hours of stimulation determined by qRT-PCR correlated closely to that observed in the microarray data set. A list of a subset of CELF1 target transcripts involved in mRNA processing and following the "Up" pattern of expression can be found in table 4.1. Overall, our data suggest that that CELF1 target transcripts were highly enriched for transcripts that show dynamic changes in expression over the course of T cell activation.

Decreased CELF1 binding to RNA is due to activation-induced phosphorylation. The finding that the majority of CELF1 target transcripts had a FCE<1 at 6 hours following T cell activation suggested that an activation-dependent change in CELF1 binding occurred. To test this hypothesis, we performed electromobility shift assays (EMSAs) using a GRE-containing RNA probe and cytoplasmic lysates prepared from resting T cells and T cells that were stimulated for 6 hours with anti-CD3 and anti-CD28 antibodies. As can be seen in figure 4.5A, we observed CELF1 binding to the GRE probe in resting T cell lysates (lane 1). This binding activity was due to CELF1 because it was specifically super-shifted with an anti-CELF1 antibody (lane 3) but was not supershifted with a control anti-HA antibody (data not shown). In stimulated T cell lysates, however, we observed minimal to no binding to the GRE (lane 2). Additionally, we performed an UV crosslinking assay using cytoplasmic lysates from resting and activated T cells and confirmed that CELF1 bound to the GRE probe (figure 4.5B). The identity of the CELF1 band was determined by immunoprecipitation with an anti-CELF1

antibody,(lanes 3 and 4) but not a control antibody (data not shown). The lack of a CELF1 band in the supernatant from the immunoprecipitation reaction (lanes 5 and 6) indicated that the immunoprecipitation depleted CELF1 from the lysates. The intensity of the CELF1 band decreased significantly in the lysates from stimulated T cells, confirming the results of the EMSA showing that binding by CELF1 to the GRE probe decreased following T cell activation. The lack of binding by CELF1 in extracts from stimulated T cells was not due to a decrease in CELF1 protein abundance, because western blot performed on the same extracts showed that similar CELF1 protein levels were present in extracts from resting and activated T cells (Figure 4.5C). These findings suggest that the decrease in CELF1 binding to RNA upon T cell activation was due to a posttranslational effect.

CELF1 is a phosphoprotein, and the binding specificity of CELF1 has been shown to be altered depending on the phosphorylation state [82, 140]. We hypothesized that post-translational phosphorylation of CELF1 may occur in human T cells and may explain the reduced RNA-binding activity of CELF1 upon T cell activation. In order to investigate this hypothesis, 2-dimensional Western blot analysis of CELF1 was performed in resting and activated T lymphocytes (figure 4.6). Compared to the CELF1 signal in extracts prepared from resting T cells, the CELF1 signal in extracts prepared from activated T cells exhibited an acidic shift (figure 4.6, top 2 panels). This shift was a result of phosphorylation because treatment of lysates with lambda phosphatase resulted in the reversion of the CELF1 signal back to its position in resting cells. These results suggest that CELF1 was not-phosphorylated in resting T cells, but upon T cell

stimulation, CELF1 became phosphorylated, and this phosphorylation correlated with reduced binding to GRE-containing RNA.

We next sought to determine directly if the reduction in CELF1 affinity for the GRE was a result of CELF1 phosphorylation. We immunoprecipitated CELF1 from resting and stimulated T cell lysates and then subjected it to lambda phosphatase or mock treatment. The immunopurified CELF1was subsequently incubated with either biotinylated GRE sequences or mutated GRE sequences. These reactions were then UV-crosslinked and the amount of RNA that crosslinked to CELF1 was determined after separating the mixtures by electrophoresis and probing for biotinylated RNA (figure 4.7A top panel). Subsequently, the same membranes were probed with an anti-CELF1 antibody to determine the amount of CELF1 protein immunopurified from resting and activated T cells (figure 4.7A bottom panel). Using this assay, we observed that CELF1 immunopurified from resting T cells bound to the GRE probe (lane 1) but not the mutated GRE probe (data not shown). Phosphatase treatment had no effect on binding by CELF1 from resting T cells to the the GRE probe (compare lanes 1 and 2). In contrast, CELF1 immunopurified from stimulated T cells bound poorly to the GRE RNA probe (figure 4.7A, lane 3, but we observed an increase in CELF1 binding to the GRE probe upon phosphatase treatment (figure 4.7A, compare lanes 3 and 4). The amount of GRE RNA bound, normalized to the amount of CELF1 precipitated was quantified in triplicate experiments for resting or activated T cells (figure 4.7B). The results showed a significant decrease in CELF1 binding to the GRE in the stimulated condition ($p = 0.001$). Upon phosphatase treatment, binding by

immunopurified CELF1 to GRE increased significantly ($p = 0.03$), and was similar to the binding observed for CELF1 immunpurified form resting T cells. Overall, these results indicate that phosphorylation of CELF1 caused a decrease in CELF1 binding to GRE RNA in stimulated T cells, and this effect was reversed by phosphatase treatment.

Discussion

CELF1 regulates posttransciptional gene expression at a number of levels including alternative splicing, translation, and mRNA degradation and functions by binding directly to RNA [212]. A network of GRE-containing genes are coordinately regulated at posttranscriptional levels through the interaction of the CELF1 with its target sequence, the GRE [50]. In the present study we sought to understand the dynamics of the CELF1-GRE interaction throughout the process of T cell activation and performed RNA-IP followed by microarray analysis to identify CELF1 target transcripts. We found approximately 1300 CELF1 target transcripts in resting T cells and approximately 150 target transcripts in activated T cells. CELF1 targets in both resting and activated T cells displayed significant enrichment for GRE-containing transcripts, confirming that the GRE is indeed a CELF1-binding site within cells. We also showed that CELF1 phosphorylation following T cell activation correlated with the decrease in the number of CELF1 target transcripts that we observed as well as a decreased ability of CELF1 to bind to GRE sequences in vitro. We further showed that phosphatase treatment of CELF1 immunopurified from stimulated T cells restored its ability to bind to the GRE by increasing the affinity of CELF1 for the GRE.

Several results suggest that CELF1 function is regulated by phosphorylation, but a unifying theme has not emerged. In some situations, CELF1 phosphorylation is associated with increased binding to RNA. For example, phosphorylation of CELF1 at Ser28 by Akt kinase led to increased affinity of CELF1 for cyclin D1 mRNA [82], and phosphorylation at Ser302 by cyclinD3-cdk4/6 led to increased CELF1 binding to p21 and C/EBPβ mRNA [208]. In a myotonic dystrophy model, CELF1 was hyperphosphorylated by PKCα and βII, causing increased protein half-life and altered splicing patterns [140, 213]. Expression of a $(CUG)_{960}$-expanded DMPK gene caused CELF1 phosphorylation and inhibition of CELF1 mediated decay of the TNF-alpha transcript [144]. It is thought that CELF1 binds to a CG-rich motif in the 5'UTR of certain transcripts and promotes their translation [141, 214]. It has been found in a liver model that following partial hepatectomy, there is phosphorylation of CELF1. It was further shown that this phosphorylation event led to an increase in the association and translational efficiency of specific CELF1 target transcripts. This effect was shown to be mediated by an increased association with the translational initiation factor eIF2 [141]. Here, we found that CELF1 is phosphorylated upon T cell activation, and that this caused a decrease in binding by CELF1 to cytoplasmic mRNA targets. Multiple kinase cascades are triggered upon T cell activation leading to several potential candidates for CELF1 phosphorylation. Further efforts will focus on identifying the phosphorylation site and kinase(s) involved in activation-dependent phosphorylation of CELF1. Our data show that the affinity of CELF1 for the GRE is regulated by phosphorylation, and suggests a powerful mechanism of coordinate posttranscriptional regulation whereby a cell could

quickly alter the stability of a whole network of transcripts. For transcripts exhibiting the "Up" pattern of expression, a possible scenario is that unphosporylated CELF1 promotes the coordinate decay of transcripts in resting T cells, but upon stimulation CELF1 becomes phosphorylated, leading to stabilization of these transcripts and allowing for transient increased expression. It is also possible that CELF1 phosphorylation may play a role in regulating the translation of these transcripts. For transcripts exhibiting the "Down" pattern of expression, CELF1-mediated mRNA decay may play a role in mediating the down-regulation of these transcripts, and CELF1 phosphorylation may regulate their subsequent return to baseline levels.

The targets of CELF1 have been identified in two previous studies. The first study was performed on the HeLa human cervical carcinoma cell line, and identified approximately 600 CELF1 target transcripts, many of which were involved in RNA processing, cellular proliferation and apoptosis [184]. The second study was performed on the mouse C2C12 myoblast cell line and identified 881 CELF1 target transcripts, many of which were also involved in RNA processing, cellular proliferation and apoptosis [71]. These other studies were both performed using transformed cell lines which have been shown to display widespread alterations in 3'UTR isoform [96]. We found relatively little overlap between the actual targets of CELF1 between our study in primary human T cells and these previous studies performed in transformed cell lines. This disparity could be a result of differential regulation of other trans-acting factors as well as differences in the transcriptomes of the different cells types. CELF1 targets showed significant enrichment of GRE-containing transcripts in all three cell types, and

although there was little overlap in specific target transcripts, the biological processes of proliferation, apoptosis, and RNA processing were enriched among target transcripts in all three cell types.

Interestingly, CELF1 target transcripts were enriched for transcripts encoding RNA-binding proteins, including its own transcript. Thus, CELF1 may be similar to other RNA-binding proteins in exhibiting auto-regulatory behavior and function as a "regulator of regulators" [215, 216]. Here we also confirmed the previous reports that CELF1 seems to be targeting transcripts which are involved in cellular proliferation and cell death [56, 184]. One pathway we found to be particularly enriched in CELF1 targets in resting cells was that of mRNA formation. T cell activation induces changes in alternative polyadenylation and alternative splicing [173, 217], and it would be interesting to determine if CELF1 plays a regulatory role in these processes.

Overall, our data is consistent with a model whereby CELF1 is actively involved in coordinately repressing transcripts involved in mRNA maturation, proliferation and cell death to maintain the cell in a quiescent state. Upon stimulation, phosphorylation of CELF1 leads to its inability to bind to certain transcripts, allowing for accumulation of target transcripts necessary for acquisition of the characteristics of activated T cells, such as proliferation and increased sensitivity to apoptotic stimuli. Inactivation of CELF1 plays a role in the development of leukemia through dysregulation of the transcription factor C/EBPbeta [218]. Additionally, a forward genetic screen in mice identified CELF1 loss of function as a potential driving mutation in the development of colorectal cancer [219]. This suggests that CELF1 may be regulating cell survival and proliferation

pathways. CELF1 could regulate these pathways both via directly binding to transcripts coding for proteins involved in these pathways as well as through regulation of splicing factors and other RNA-binding proteins, which in turn regulate mRNA isoform usage, mRNA decay and/or translation patterns which promote survival and proliferation.

In conclusion, we have shown that CELF1 is coordinating networks of transcripts crucial for primary human T cell activation. Upon T cell activation, we observed dramatic reduction in the CELF1 target transcript pool and showed that this change in CELF1 binding to RNA was due to phosphorylation of CELF1. Further work will focus on delineating the signaling pathways involved in CELF1 phosphorylation in primary human T cells and the mechanism of CELF1-mediated regulation of the activated T cell phenotype.

Table 1
CELF1 Target Transcripts (Resting T cell)

Affymetrix Probe ID	Gene Symbol	Resting T cell Half Life (min)	Pattern of Expression	FCE*	Function
mRNA Metabolism					
201586_s_at	SFPQ	68	Up	0.11	RNA splicing
208672_s_at	SFRS3	74	Unchanged	0.32	RNA splicing
200097_s_at	HNRNPK	235	Unchanged	0.33	RNA splicing
211271_x_at	PTBP1	292	Up	0.36	RNA splicing
210093_s_at	MAGOH	406	Up	0.16	RNA splicing, mRNA 3' end processing
208863_s_at	SFRS1	501	Up	0.28	RNA splicing
200687_s_at	SF3B3	675	Up	0.53	RNA splicing
215245_x_at	FMR1	749	Down	0.28	RNA localization
211932_at	HNRNPA3	786	Up	0.18	RNA splicing
201151_s_at	MBNL1	799	Unchanged	0.46	RNA splicing
221481_x_at	HNRNPD	840	Up	0.56	RNA splicing
202157_s_at	CUGBP2	1057	Unchanged	0.45	Regulation of translation, RNA binding
201993_x_at	HNRNPDL	2142	Unchanged	0.46	regulation of transcription
211185_s_at	SF3B1	2581	Down	0.26	RNA splicing
208766_s_at	HNRNPR	2730	Up	0.16	RNA splicing
200594_x_at	HNRNPU	6580	Up	0.29	RNA splicing
222476_at	CNOT6	n/a	Unchanged	0.46	RNA binding, Nuclease activity
221743_at	CUGBP1	n/a	Unchanged	0.44	RNA splicing, Regulation of mRNA stability
216559_x_at	HNRNPA1	n/a	Unchanged	0.66	RNA splicing
201726_at	ELAVL1	n/a	Up	0.45	Regulation of mRNA stability
Regulation of Growth and Survival					
217168_s_at	HERPUD1	33	Unchanged	0.28	Regulation of apoptosis, Regulation of caspase activity
212501_at	CEBPB	38	Unchanged	0.62	Regulation of transcription
212614_at	ARID5B	45	Unchanged	0.35	Regulation of transcription/Growth
205291_at	IL2RB	62	Unchanged	0.48	Regulation of apoptosis
203508_at	TNFRSF1B	63	Up	0.27	Induction of apoptosis
203752_s_at	JUND	71	Unchanged	0.61	Regulation of transcription
202687_s_at	TNFSF10	77	Up	0.33	Apoptosis, RNA destabilization
1554479_a_at	CARD8	294	Unchanged	0.79	Regulation of cell death
1559052_s_at	PAK2	3278	Unchanged	0.35	Apoptosis, Actin cytoskeleton organization
1558233_s_at	ATF1	4110	Unchanged	0.27	Regulation of transcription, Growth, Survival
225585_at	RAP2A	n/a	Unchanged	0.54	Intracellular signaling cascade
Protein Metabolism					
213128_s_at	UBE3A	67	Unchanged	0.33	Protein ubiquitination
201524_x_at	UBE2N	81	Up	0.18	Protein ubiquitination, immune response
211764_s_at	UBE2D1	113	Unchanged	0.19	Protein ubiquitination
204279_at	PSMB9	116	Up	0.36	Peptidase activity, Antigen presentation
201177_s_at	UBA2	143	Unchanged	0.13	Protein SUMOylation
201649_at	UBE2L6	151	Up	0.36	Protein ubiquitination
212513_s_at	USP33	244	Down	0.19	Ubiquitin specific peptidase activity
208770_s_at	EIF4EBP2	372	Unchanged	0.52	Regulation of translation
200668_s_at	UBE2D3	1566	Unchanged	0.20	Protein ubiquitination
217826_s_at	UBE2J1	2053	Unchanged	0.21	Protein ubiquitination

*Fold Change in Enrichment as defined in text

Table 4.2
Sequence Analysis of CELF1 Target Transcripts

K-mer Analysis	BioProspector Analysis
TGTGTGTGTGT	
GTGTGTGTGTG	
TTTGTTTTGTT	
TTGTTTTGTTT	
TGTTTTGTTTT	

Table 3

CELF1 Target Transcripts in Stimulated T cells

Affymetrix Probe ID	Gene Symbol	p-value (0hr vs 6hr)	FC* from 0hr to 6hr	FCE**	Function
Transcripts Targeted in Resting and Stimulated Cells					
201649_at	UBE2L6	0.13	1.77	0.36	Protein ubiquitination
202081_at	IER2	0.99	0.99	0.85	Uncharacterized
208770_s_at	EIF4EBP2	0.42	2.22	0.52	Regulation of translation
212501_at	CEBPB	0.45	2.00	0.62	Regulation of transcription/cell death
209666_s_at	CHUK	0.08	7.78	0.19	Intracellular signaling cascade
226400_at	CDC42	0.31	1.92	0.34	Cytoskeleton organization
225585_at	RAP2A	0.49	1.38	0.54	Intracellular signaling cascade
212751_at	UBE2N	0.04	4.11	0.27	Protein ubiquitination
212060_at	SR140	0.21	3.61	0.29	RNA processing, Splicing
221498_at	SNX27	0.61	0.88	1.12	Protein localization
Transcripts Expressed in Resting and Stimulated Cells, But Only Targeted in Stimulated Cells					
218915_at	NF2	0.47	1.49	0.76	Cytoskeletal organization
213440_at	RAB1A	0.89	1.39	1.13	Regulation of transcription
209457_at	DUSP5	0.01	10.48	0.24	MAP-kinase phosphatase activity
230534_at	ZNF678	0.50	1.16	1.15	Regulation of transcription
201556_s_at	VAMP2	0.09	0.51	1.16	Vesicle-mediated transport
204185_x_at	PPID	0.34	1.80	0.40	Protein folding
224647_at	CCNY	0.43	2.36	0.40	Uncharacterized
205659_at	HDAC9	0.34	1.80	0.65	Regulation of transcription, Cell activation
1552329_at	RBBP6	0.08	8.57	0.31	Protein ubiquitination
Transcripts Newly Expressed in Stimulated Cells					
203820_s_at	IGF2BP3	0.80	1.14	0.96	Regulation of translation, mRNA binding
208712_at	CCND1	0.00	3.27	0.48	Mitotic cell cycle
1568678_s_at	FGFR1OP	0.16	1.27	1.04	Cytoskeleton organization
205227_at	IL1RAP	0.94	1.40	1.01	Cytokine receptor, Inflammatory response
235520_at	ZNF280C	0.28	3.03	0.56	Regulation of transcription
227198_at	AFF3	0.74	1.78	1.09	Regulation of transcription
208916_at	SLC1A5	0.01	2.28	0.48	Organic acid transport
223135_s_at	BBX	0.43	1.69	0.86	Regulation of transcription
225972_at	TMEM64	0.02	6.28	0.38	Uncharacterized
212190_at	SERPINE2	0.02	31.34	0.20	Regulation of proteolysis, Regulation of cell migration

*Fold Change

**Fold Change in Enrichment as defined in text

Table 4.4. Fold change in gene expression at six hours determined by qRT-PCR and Microarray Analysis

	qRT-PCR	Microarray
ELAVL1	1.7±0.7	2.3±1.1
FMR1	0.70±0.3	0.76±0.05
HNRNPD	1.7±1.5	1.3±0.1
PTBP1	2.7±1.5	2.8±0.4

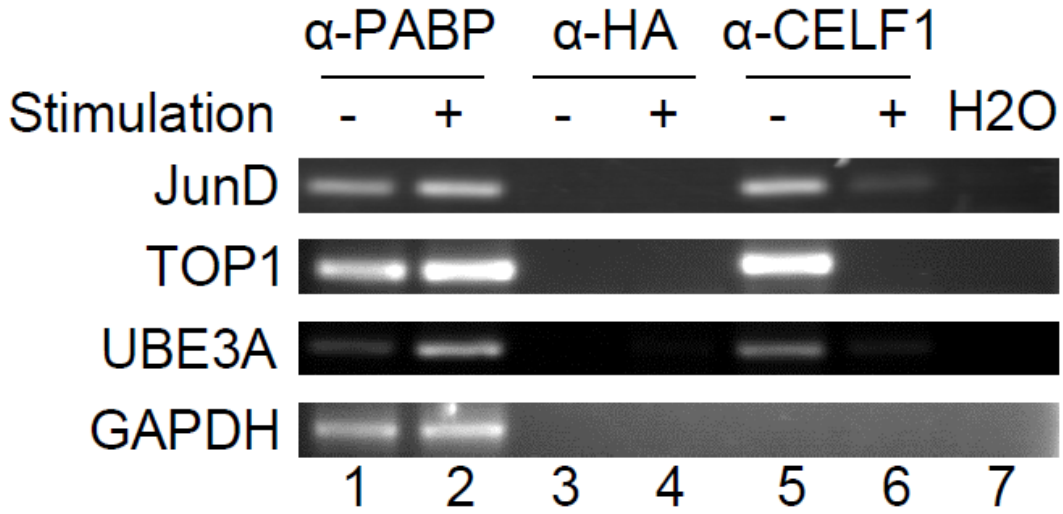

Figure 4.1. Validation of CELF1 Target Transcripts by RT-PCR. RNA-IP was performed using cytoplasmic lysates from unstimulated (-) and stimulated (+) T cells and antibodies targeting PABP (α-PABP), HA (α-HA), and CELF1 (α-CELF1). The levels of Topoisomerase I (TOPI), JunD proto-oncogene (JunD), Ubiquitin-protein ligase E3A (UB3EA) and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) transcripts were measured in the immunopreciptated material by RT-PCR.

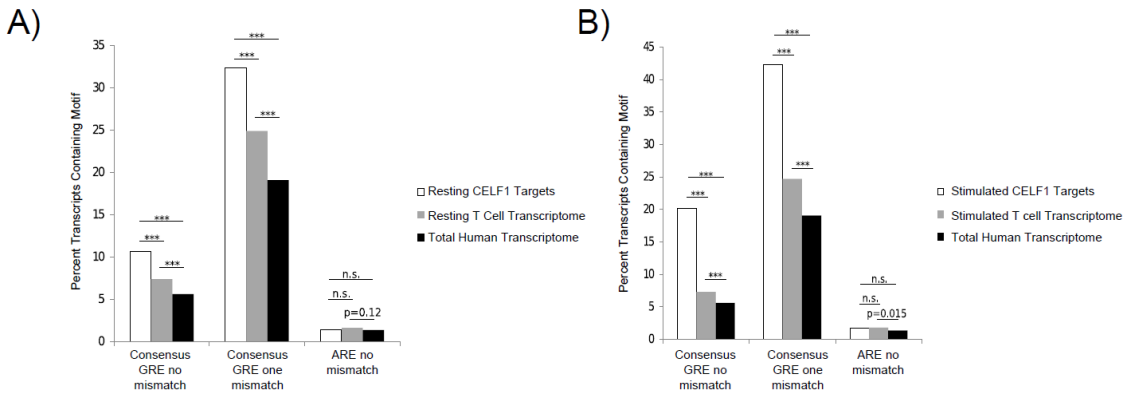

Figure 4.2. GREs were enriched in the 3'UTRs of CELF1 target transcripts expressed in primary human T cells. The 3'UTRs from the total human transcriptome were extracted from NCBI refseq records. A) The prevalence of the consensus GRE or ARE sequence in the 3' UTRs of CELF1 target transcripts from resting T cells (Resting CELF1 Targets) was compared to the presence of these sequences in the Total Human Transcriptome or the Resting T Cell Transcriptome. B) The prevalence of the consensus GRE or ARE sequences in the 3' UTRs of CELF1 target transcripts from T cells that were stimulated for 6 hours with anti-CD3 and anti-CD28 antibodies (Stimulated CELF1 Targets) was compared to the presence of these sequences in the Total Human Transcriptome or the Stimulated T Cell Transcriptome. For the GRE, we allowed zero or one mismatch and for the ARE, we allowed zero mismatches. (***=$p<0.001$, n.s.= not significant, chi squared test)

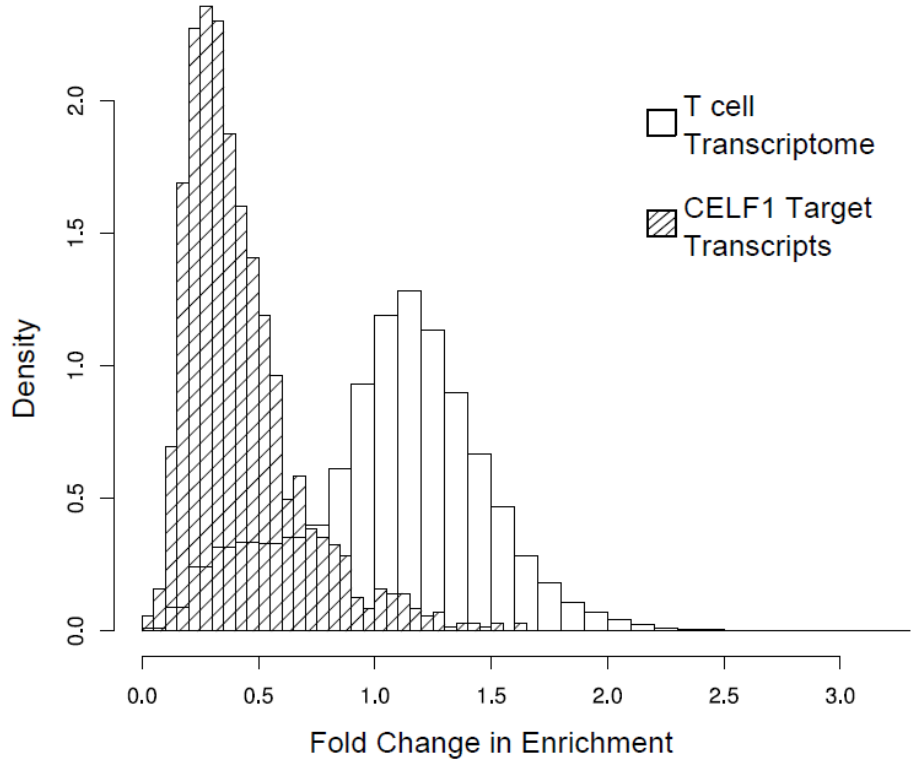

Figure 4.3. Decreased association of CELF1 with target transcripts following T cell activation. The FCE was calculated for all T cell transcripts (blank bars) and CELF1 target transcripts (dashed bars), and a histogram of these values is depicted. On the y-axis, Density is defined as the normalized number of transcripts falling in a given FCE bin, such that the total area of the histogram is one. The CELF1 target transcript population exhibits a reduced FCE compared to all T cell transcripts, suggesting decreased association of these transcripts with CELF1 following T cell activation.

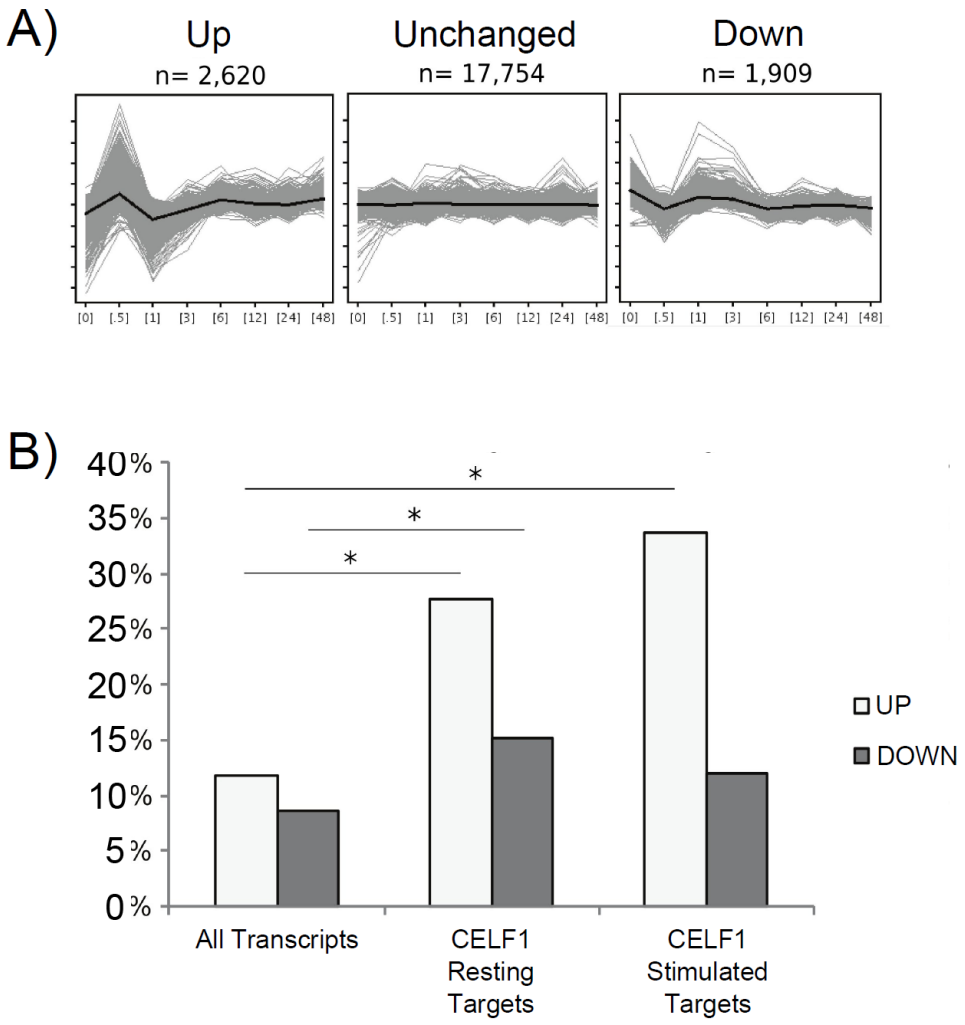

Figure 4.4. Expression patterns of CELF1 target transcripts following T cell activation. A) Gene expression was previously assessed using microarrays over the first 48 hours of T cell stimulation with anti-CD3 and anti-CD28 antibodies [14] . Using this data, the T cell transcriptome was segregated into three common patterns of expression using a K-means clustering approach. The "Up" pattern of expression is shown on the left with a sharp increase in abundance at 30 min followed by a return to baseline at one hour and a subsequent slow increase over the following 47 hours. The "Down" pattern of expression is shown on the right and exhibits an inverse pattern compared to the "Up"

pattern. Finally the majority of transcripts follow the "Unchanged" pattern which shows little change in abundance over T cell stimulation. On the y-axis of each panel is normalized, $\log_2$ microarray expression values, and the x-axis depicts hours of T cell stimulation. For each pattern of gene expression, the average expression pattern of each cluster is depicted by the black line with the range of expression values of each cluster depicted by the gray lines. B) We compared the percentage of transcripts expressing the UP or DOWN patterns for all transcripts expressed in primary T cells at one or more time points throughout the first 48hours of T cell activation (All Transcripts), CELF1 target transcripts in resting T cells (CELF1 Resting Targets) and CELF1 target transcripts in stimulated T cells (CELF1 Stimulated Targets). (*= $p<0.05$, chi squared test)

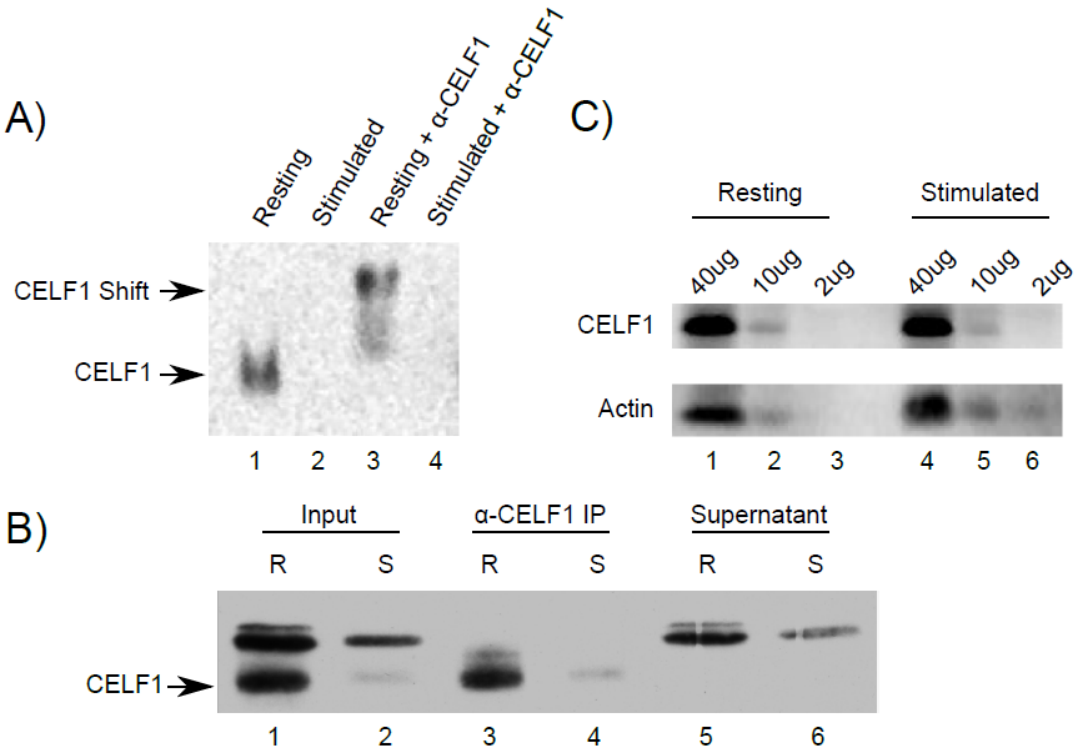

Figure 4.5. Binding by CELF1 to the GRE decreased following T cell activation. A) EMSAs were performed by incubating a biotinylated GRE-containing oligonucleotide with cytoplasmic extracts from resting T cells (Resting) or T cells that had been activated with anti-CD3 and anti-CD28 antibodies for 6 hours (Stimulated). An antibody against CELF1 was added to the indicated reactions (α-CELF1). The CELF1-containing band and the super-shifted band are shown with arrows. B) Cytoplasmic lysates from resting (R) and stimulated (S) T cells were mixed with a biotinylated GRE probe and treated with UV light. Reactions were mixed with an anti-CELF1 antibody, and the non-immunoprecipited reactions (Input), the anti-CELF1 immunoprecipitated fractions (α-CELF1 IP) and the immunoprecipitation supernatants (Supernatant) were separated by SDS-PAGE. The CELF1 band is indicated with an arrow. C) Western blot analysis was

performed on titrated amounts (2-40 ug of protein) of cytoplasmic lysates from resting and activated T cells using anti-CELF1 and anti-actin antibodies.

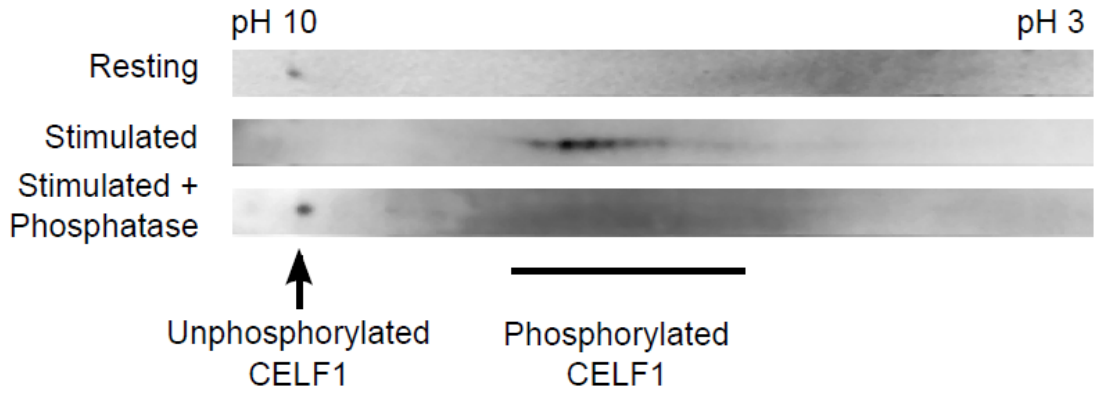

Figure 4.6. CELF1 was phosphorylated upon T cell activation. Cytoplasmic lysates were prepared from resting T cells and T cells that were stimulated for 6 hours with anti-CD3 and anti-CD28 antibodies. The lysates were immunoprecipitated using an anti-CELF1 antibody and the immunprecipitated material was treated with lambda phosphatase or mock treated. Samples were then separated by 2D electrophoresis and CELF1 was identified by western blot analysis.

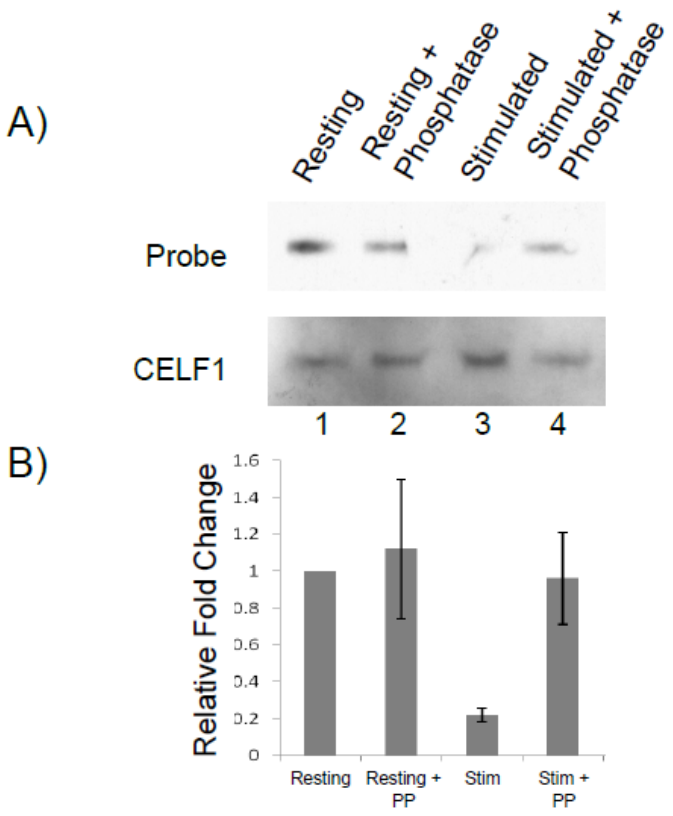

Figure 4.7. CELF1 phosphorylation caused reduced affinity of CELF1 binding to GRE RNA. A) CELF1 was immunopurified from resting (Resting) and stimulated (Stim) T cell lysates and subsequently treated with either lambda phosphatase (PP) or underwent mock treatment. The immunopurified CELF1 was then mixed with a biotinylated GRE probe and treated with UV light. The CELF1:RNA complexes were separated by SDS-PAGE, and the biotinylated GRE probe (Probe) and CELF1 levels (CELF1) levels were visualized. B) The experiment in A was performed three times and the results were quantified. For each reaction, the RNA to CELF1 ratio was determined, and the ratios were normalized to the resting, mock treated condition for each experiment. Error bars indicate the standard deviation of three experiments.

Chapter 5

Aberrant Regulation of CELF1 mRNA Decay by Signaling from the Oncogenic KSHV G-protein Coupled Receptor

Introduction

The accurate expression of genetic information is a complex process involving multiple layers of genetic control. Post-transcriptional regulation at the level of mRNA decay represents a crucial layer of gene control which plays an important role in the maintenance of cellular homeostasis. mRNA decay has been found to regulate networks of transcripts whose protein products are involved in cellular processes such as cellular proliferation and apoptosis [212, 220]. Dysregulation of these and other processes is characteristic of cancer [221]. Mounting evidence suggests that the dysregulation of mRNA decay pathways is in part responsible for the acquisition of these malignant phenotypes [6, 222]. Transcripts involved in oncogenesis such as c-myc [223, 224], VEGF [225], and c-fos [226] have all been shown to be abnormally stable in cancer, and alterations in the expression and function of proteins involved in mRNA decay can be prognostic indicators in some cancers [7, 157, 227].

mRNA decay is regulated by factors such as RNA binding proteins and miRNAs which bind to networks of transcripts through target sequences typically found in the 3'UTR of transcripts [34]. The best characterized mRNA decay network involves the recognition of an AU-rich element (ARE) which is found in the 3'UTR of a number of cytokine and proto-oncogene transcripts [22]. Importantly, a number of ARE-binding proteins, and ARE containing transcripts have been shown to be dysregulated in a number of malignancies [147]. Our group recently identified another mRNA decay network which involves the phosphoprotein CUG-binding protein 1 (CUGBP1, also referred to as CUGBP- and ELAV- like family member 1 (CELF1)) [50]. CELF1 binds

to a GU-rich element (GRE) which resides in the 3'UTR of a number of cellular proliferation and apoptosis regulatory transcripts and promotes their rapid decay [50, 184]. As a regulator of the decay of cellular proliferation and apoptosis related transcripts, CELF1 is poised for involvement in malignancy [184, 220]. One report supporting CELF1s involvement in cancer comes from its identification in a mouse-based transposon based screen for drivers of colorectal cancer [219]. Additionally, CELF1 dysregulation has been found to be involved in leukemogenesis through altering C/EBPβ isoform usage [218]. Finally, we recently found that CELF1's mRNA binding capacity was reduced following a phosphorylation event in primary human T cell activation; a physiologic state of transient proliferation and altered apoptotic responses [220]. These results lead us to hypothesize that CELF1 mediated mRNA decay may be altered in malignancy.

Kaposi's Sarcoma is a malignancy characterized by outgrowth of vascular endothelial cells. The viral cause for this malignancy was discovered in 1994 to be the Kaposi's Sarcoma associated herpesvirus (KSHV) (also known as Human Herpes Virus 8, HHV8) [228]. KSHV is a large double stranded DNA virus which encodes for many proteins with remarkable homology to human proteins [159]. One of these proteins is the oncogenic viral G-Protein Coupled Receptor (vGPCR). This receptor is homologous to the human CXCR1 and CXCR2 receptors, but unlike its human homologues, the vGPCR signals constitutively and does not require ligand binding. The KSHV vGPCR is a viral oncogene due to its ability to transform primary endothelial cells as well as cause a Kaposi's Sarcoma like disease in transgenic mice [229]. Here we show that expression

of the KSHV vGPCR causes inhibition of CELF1 mediated mRNA decay. We further present evidence to support that CELF1 inhibition is a result of an ERK mediated CELF1 phosphorylation event downstream of the vGPCR. This phosphorylation event preserves CELF1s mRNA binding activity, but coincides with a reduction in CELF1s ability to recruit RNA decay factors to target transcripts. Overall, our results support a model whereby KSHV inhibits CELF1 mediated mRNA decay to promote malignant transformation.

Results

KSHV vGPCR reverses CELF1 Mediated Decay

The KSHV vGPCR has previously been shown to activate multiple kinase cascades, including the MAPK/ERK and PI3K pathways, and this signaling promotes cellular transformation [160]. We have previously shown that CELF1 binds to and promotes the decay of transcripts involved in the regulation of cancer related pathways such as cell cycle control and apoptosis [184, 220]. We hypothesized that CELF1 mediated mRNA decay may be inhibited in the setting of the oncogenic KSHV. In order to test this hypothesis, we measured the rate of decay of GRE containing transcripts in the presence and absence of the vGPCR. We cloned a GRE containing portion of the 3'UTR of the JunB transcript downstream of an otherwise stable beta-globin reporter transcript. We have previously shown that this construct rapidly decays in HeLa cells in a CELF1 and GRE dependent manner [50, 184]. Following transfection of HeLa Tet-Off cells with the vGPCR construct, the beta-globin reporter, and a GFP plasmid to serve as a transfection

control, transcription was halted by the addition of Doxycycline to the culture medium. Total cellular RNA was isolated after 1.5, 3, 4.5 and 6 hours and the expression of the beta-globin reporter mRNA and GFP mRNA was analyzed by northern blotting to measure the decay rate of the reporter transcript. We performed three separate experiments, and the results of the decay rates are graphed in figure 5.1. The beta-globin signal was normalized to the GFP signal at each time point, and the 0 hour time point was set to 100%. Consistent with our previous results, it can be seen that the beta-globin transcript without the JunB 3'UTR was relatively stable, but insertion of the GRE containing 3'UTR into the beta-globin transcript caused destabilization of the transcript. Upon co-transfection of the unstable, GRE-containing reporter transcript with the KSHV vGPCR, we observed a stabilization of the reporter transcript such that its decay rate was indistinguishable from the non-GRE containing beta-globin transcript. These results showed that the KSHV vGPCR is able to inhibit CELF1 mediated mRNA decay.

The KSHV vGPCR has been shown to signal strongly through the MAPK/MEK/ERK signal transduction cascade [160, 161]. ERK phosphorylation has been shown to inhibit mRNA decay pathways in other systems [230, 231], and thus we hypothesized that the vGPCR was inhibiting CELF1 mediated mRNA decay in an ERK dependent manner. In order to test this hypothesis, we utilized a similar approach as above to measure the decay rate of a GRE containing beta-globin transcript in the presence and absence of the vGPCR. In addition, we chemically interrupted vGPCR signaling through the MAPK/MEK/ERK pathway by the addition of the MEK inhibitor UO126 to the cell culture media. Similar to the above experiments, we co-transfected

HeLa Tet-Off cells with the vGPCR, the GRE containing beta-globin reporter plasmid and a GFP expressing plasmid to serve as a transfection control. Following transfection, we inhibited transcription by addition of Doxycycline to the cell culture media and collected total cellular RNA after 1.5, 3, 4.5, and 6 hours. We analyzed the abundance of beta-globin mRNA relative to GFP mRNA by northern blot. It can be seen by the representative experiment shown in figure 5.2A that we observed rapid decay of the GRE containing reporter, which was inhibited by the presence of the vGPCR. Upon inhibition of vGPCR signaling by UO126, however, we observed a reversion back to rapid decay of the GRE containing transcript. The MEK inhibitor was added to a concentration of 10uM, and this was sufficient to inhibit signaling as can observed in the western blot shown in figure 5.2B which depicts the inhibition of vGPCR induced ERK phosphorylation in the UO126 containing cultures, despite equal loading as evidenced by equal signal in western blots targeting total ERK. These results show that the inhibition of CELF1 mediated mRNA in the setting of the oncogenic KSHV vGPCR is mediated by the MEK/ERK kinse cascade.

KSHV vGPCR causes CELF1 phosphorylation

Given our results that CELF1 mediated mRNA decay is inhibited by the KSHV vGPCR in a MEK/ERK dependent manner, we hypothesized that CELF1 is directly phosphorylated in the setting of the vGPCR. In order to test this hypothesis, we transduced HeLa cells with lentiviral constructs engineered to express either the KSHV vGPCR or control virus which expressed GFP. 72 hours following transduction,

cytoplasmic lysates were prepared and subjected to 2D gel western blotting to determine the isoelectric point of CELF1 in the two conditions. Briefly, the lysates were dialyzed against a 2D gel buffer, and subsequently separated in the first dimension by isoelectric point. Unmodified CELF1 was predicted to have an isoelectric point of approximately pH 8, and phosphorylation was expected to exert an acidic shift on the CELF1 isoelectric point. Following separation along the first dimension, the lysates were separated by molecular weight and blotted onto a PVDF membrane. The location of CELF1 protein was then interrogated by western blotting using anti-CELF1 antibodies. As can be seen in the top two panels of figure 5.3, CELF1 from lysates prepared from control transduced cells exhibited an isoelectric point of approximately 8. In lysates prepared from cells transduced with vGPCR expressing virus, however, CELF1 exhibited an acidic shift consistent with a phosphorylation event. In order to prove that this acidic shift was a result of phosphorylation and not an alternative form of post-translational modification, CELF1 from the vGPCR expressing lysates was treated with the enzyme lambda phosphatase. As can be seen in the bottom panel of figure 5.3, CELF1 that had been treated with lambda phosphatase reverted to an isoelectric point similar to that seen in the control lysates, suggesting that CELF1 was directly phosphorylated downstream of the vGPCR.

CELF1 mRNA binding is preserved in the setting of vGPCR signaling

We have recently shown that in a primary human T cell model, CELF1 is phosphorylated in an activation dependent manner resulting in reduced CELF1 affinity for the GRE [220]. We predicted that vGPCR dependent phosphorylation similarly reduced CELF1's

ability to bind the GRE. In order to test this prediction we performed in vitro RNA binding reactions to measure the relative affinity of CELF1 for the GRE in lysates prepared from cells transduced with GFP and vGPCR expressing lentivirus. We prepared binding reactions consisting of 20 ug of cytoplasmic lysate, 25 fmol of 25nt RNA probes labeled with a 5'-biotin moiety, increasing amounts of unlabeled ("cold") RNA probes of identical sequence, all in RBB buffer in the presence of appropriate inhibitors to reduce nonspecific binding. Following a brief incubation, these reactions were subjected to electrophoresis on a 5% non-denaturing polyacrylamide gel, transferred to a membrane, and the biotinylated RNA probe was visualized by the addition of a streptavidin-HRP conjugate molecule and substrate compound. Three separate experiments were performed, and representative images can be seen in figure 5.4. It can be seen that in both control (A) and vGPCR transduced (B) cells, multiple cytoplasmic proteins were able bind to the RNA probe as evidenced by the vertical shift of the free probe (lane 1) into multiple bands (lane 3). The CELF1 band is indicated and was identified by a supershift of that band upon the addition of anti-CELF1 antibody to the binding reaction (lane 2). Upon the addition of increasing cold competitor RNA (lanes 3-8), the amount of CELF1 bound labeled probe in both the control and vGCPR conditions decreased to negligible amounts in lane 8 at comparable rates. Importantly, the CELF1 protein levels were also comparable in the two conditions, as can be seen in figure 5.5. These results suggested that in the setting of vGPCR induced CELF1 phosphorylation, CELF1 maintained its ability to recognize and bind the GRE.

CELF1 Recruitment of PARN is reduced by vGPCR signaling

The result that CELF1 is still able to bind to the GRE in the setting of vGPCR dependent phosphorylation encouraged us to explore other hypotheses to explain the vGPCR induced inhibition of CELF1 mediated decay. It had been previously reported that CELF1 recruits the de-adenylase enzyme PolyA ribonuclease in HeLa cells, and that this recruitment is crucial for CELF1 mediated deadenylation and subsequent transcript decay [68]. We thus explored the possibility that the ability of CELF1 to recruit PARN was reduced in the setting of the vGPCR. In order to explore this question, we utilized an RNA-bait approach to purify GRE-binding protein complexes from cytoplasmic lysate. We again transduced HeLa cells with lentivirus expressing either GFP or the vGPCR and prepared cytoplasmic lysates from these samples. We then incubated these lysates with 25nt RNA oligonucleotide "bait" composed of either a GRE or mutated GRE sequence, both of which had been modified on the 5' end with a biotin moiety. The bait was then purified from the reactions by the addition of streptavidin coated agarose beads, and any co-purified proteins were eluted after stringent washing. Eluted proteins were then interrogated by western blotting using anti-PARN, anti-CELF1, and anti-eIF2α antibody. In addition, total lysate, comprising 10% of the amount of lysate used for the RNA bait reactions was analyzed to determine the relative abundance of each of the proteins of interest in each of the transduction conditions. The experiment was performed four separate times and a representative result is shown in figure 5.5. It can be seen that CELF1 and PARN levels were comparable in both the control and vGPCR transduced samples, and none of the proteins analyzed co-purified with the mutated GRE bait in either condition. Consistent with our previous results, we purified equal amounts of

CELF1 protein out of both lysates using the GRE probe. We consistently purified low levels of PARN out of control transduced HeLa cells, but were unable to detect co-purified PARN in the vGPCR transduced condition. In order to control for non-specific binding of proteins to CELF1 in the control condition we examined whether the previously reported CELF1 binding partner eIF2α co-purified with the bait in either condition. As can be seen in the bottom row of figure 5.5, we were unable to detect any co-purified eIF2α in either condition.

Discussion

Post-transcriptional regulatory pathways are crucial for a cell to respond and adapt to the constant influx of environmental signals. Our lab has previously shown that there are widespread changes in mRNA decay rates in leukemic cells [8], and the dysregulation of networks of mRNA decay has been found in a number of independent studies of various malignancies [232]. We have recently shown that the interaction between CELF1 and its target sequence, the GRE, is important for the control of cancer related pathways such as cellular proliferation and regulation of apoptosis [184, 220]. In the present study, we hypothesized that the oncogenic vGPCR encoded by KSHV hijacks cellular signaling pathways leading to dysregulation of CELF1 function which contributes to a malignant phenotype. We found that CELF1 mediated mRNA decay was indeed inhibited in the presence of the vGPCR and that this inhibition was dependent on signaling through the MAPK/ERK pathway. We further showed that this inhibition of CELF1 function corresponded to the phosphorylation of CELF1. Finally, we found that CELF1's affinity for the GRE was preserved following this phosphorylation event, but

CELF1's ability to recruit the de-adenylase PARN was inhibited.

CELF1 is a known phosphoprotein, and several aspects of CELF1 function have been shown to be regulated by phosphorylation [144, 145, 208]. In regards to CELF1 mediated mRNA decay, it has been shown that CELF1 phosphorylation occurs in a cell model of Myotonic Dystrophy type 1 (DM1) [140]. In this model, a (CUG)n-expanded myotonic dystrophy protein kinase gene is expressed and this has been shown to cause CELF1 phosphorylation in a PKC dependent manner [140]. This model was used to show that the PKC mediated phosphorylation of CELF1 causes inhibition of CELF1 mediated decay of the TNF-α transcript, although the mechanism of this inhibition was not explored [144]. We recently showed that in primary human T cell activation, cytoplasmic CELF1 is phosphorylated in an activation dependent manner [220]. We found that this phosphorylation caused a reduction in CELF1s affinity for the GRE, and this correlated with a drastic reduction in the CELF1 cytoplasmic target transcript population [220]. In this study we found that CELF1 phosphorylation by the vGPCR resulted in a reduction of the ability to recruit PARN, but maintained CELF1's RNA binding activity. How do we interpret the seemingly divergent effects of CELF1 phosphorylation in T cell stimulation and KSHV infection? It is likely the phosphorylation of CELF1 in T cell stimulation is a separate event from the phosphorylation in the presence of the vGPCR. We hypothesize that the CELF1 functions of RNA binding and PARN recruitment are able to be independently regulated. This independence is likely the result of the activation of different signalling pathways in the two systems, culminating in different kinases phosphorylating CELF1 at independent

locations. Our results suggest that vGPCR signalling through the MAPK/MEK/ERK pathway results in CELF1 phosphorylation and resultant decrease in PARN recruitment, with minimal alteration in its RNA binding activity. In the setting of T cell stimulation, multiple kinases are activated including the MAPK/MEK/ERK and PKC pathways. It is possible that in T cell activation, multiple pathways converge on CELF1 and it would be interesting to investigate whether CELF1 recruits PARN in T cells and whether this recruitment is regulated during T cell activation.

The ability for the cell to regulate CELF1 RNA binding independently from PARN recruitment would allow the cell to separate CELF1's role in mRNA decay from its other functions. CELF1 is involved in several cellular processes such as alternative splicing and translational regulation in addition to mRNA decay [212]. CELF1 is required to bind to mRNA in all of these functions, but PARN recruitment is presumably only required for CELF1 mediated mRNA decay. By disrupting CELF1 mRNA binding, the cell is effectively shutting down all aspects of CELF1 function. The ability to shut down PARN recruitment independently of mRNA binding would allow the cell to turn off CELF1 mediated mRNA decay while maintaining CELF1's other functions.

What role could inhibition of CELF1 function be playing in KSHV malignancy? CELF1 coordinately regulates the decay of a network of transcripts involved in regulation of cellular proliferation and response to apoptosis [184, 220]. Included in these target transcripts are a number of proto-oncogenes and tumor associated transcripts [220]. We have previously shown that knockdown of CELF1 alters the cellular response to apoptotic stimuli [184]. Additionally, inactivation of CELF1 mediated mRNA decay

was associated with proliferation and altered apoptosis in the setting of T cell activation [220]. Thus, it is possible that CELF1 mediated mRNA decay inhibition by the vGPCR is beneficial for the survival of KSHV infected endothelial cells through altering cellular proliferation and apoptosis pathways.

CELF1 is also known to regulate the expression of a number of RNA binding proteins involved in AU-Rich Element (ARE) mediated mRNA regulation [71, 184, 220]. One specific CELF1 target, ELVAL1, binds to the ARE and stabilizes a number of cytokine transcripts [23, 233]. It was previously found that KSHV expresses a gene product, kaposin B, which stabilizes ARE containing cytokine transcripts through activation of the p38/MK2 signaling cascade [162]. This signaling cascade was shown to increase ELAVL1 cytoplasmic abundance and contribute to increased PROX1, a master regulator of lymphatic endothelial differentiation [164]. It is possible that the inhibition of CELF1 mediated mRNA decay contributes to malignancy by promoting accumulation of ELVAL1 and subsequent stabilization and upregulation of PROX1.

Several lines of evidence point towards altered CELF1 function as a common feature of malignancy. Our group previously reported global alterations in mRNA half-life in malignant versus normal T cells [8]. Since CELF1 is a major determinate of mRNA decay in T cells, it is possible that CELF1 mediated mRNA decay is dysregulated in T cell malignancies. This possibility is supported by evidence that CELF1 inhibition is involved in the development of leukemia through dysregulation of the transcription factor C/EBPβ [218]. CELF1 is ubiquitously expressed, raising the possibility of its involvement in malignancy outside the hematopoietic system. Evidence for this comes

from a transposon based genetic screen which identified CELF1 as a top ten candidate for drivers of colorectal cancer in mice [219]. In the present study we found that CELF1 mediated mRNA decay was inhibited by the oncogenic vGPCR of KSHV, further supporting a role for CELF1 in sarcomas. As a ubiquitously expressed master regulator of a network of proliferation and apoptosis related transcripts, CELF1 represents a central hub in the coordination of these cancer related pathways. Further work remains to determine whether CELF1 represents an attractive candidate as a target for the development of novel therapeutic compounds for the treatment of cancer.

In conclusion, we have shown that CELF1 is inhibited by the oncogenic vGPCR from KSHV. We further determined that this inhibition of CELF1 is mediated through a MEK/ERK phosphorylation event on CELF1 which results in a reduction of CELF1's ability to recruit decay factors to target transcripts. Importantly, this phosphorylation maintains CELF1's ability to bind to mRNA, thus allowing CELF1 to potentially carry out its other cellular roles. Futher work will focus on understanding the downstream effects of CELF1 mediated mRNA decay and how these contribute to the acquisition of an oncogenic phenotype.

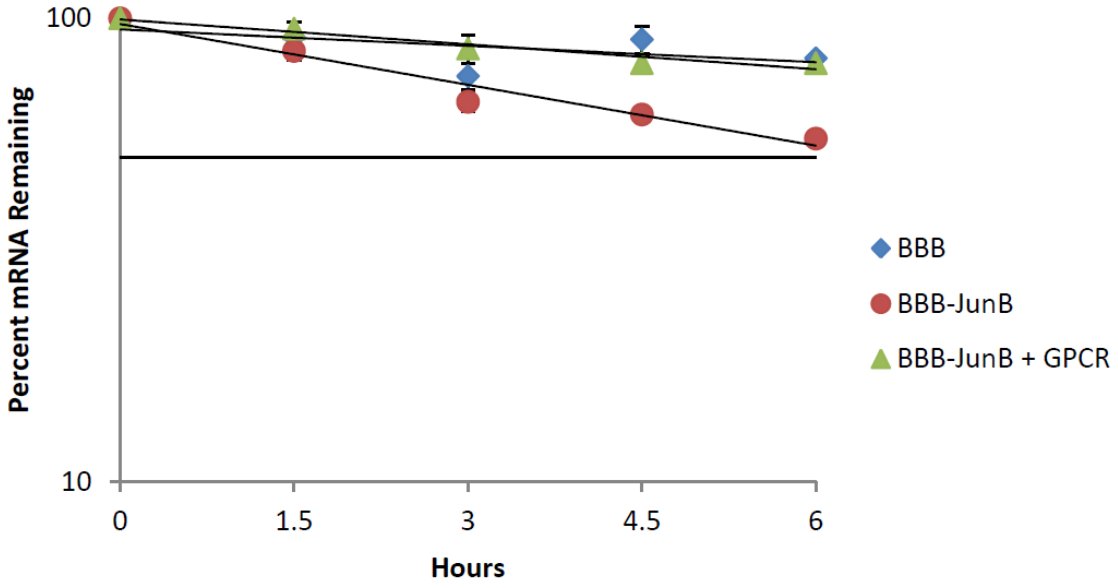

Figure 5.1. CELF1 mediated mRNA decay is inhibited in the setting of the KSHV vGPCR. Two variants of the stable beta-globin transcript were transfected into HeLa tet-off cells. The first variant was comprised of the stable beta-globin transcript. The second variant was comprised of the stable beta-globin transcript with a GRE-containing segment of the JunB 3'UTR inserted into the beta-globin 3'UTR. The second Beta-globin construct was transfected into HeLa Tet-off cells in conjunction with either the vGPCR or a control plasmid. Doxycycline was added to halt transcription, and the abundance of the beta-globin transcript was measured at 0, 1.5, 3, 4.5, and 6 hours via northern blot. The 0 hour time point was set to 100%, and the values percent of transcript present is charted as a function of time. Error bars represent the standard error of independent experiments, and horizontal solid line denotes 50%.

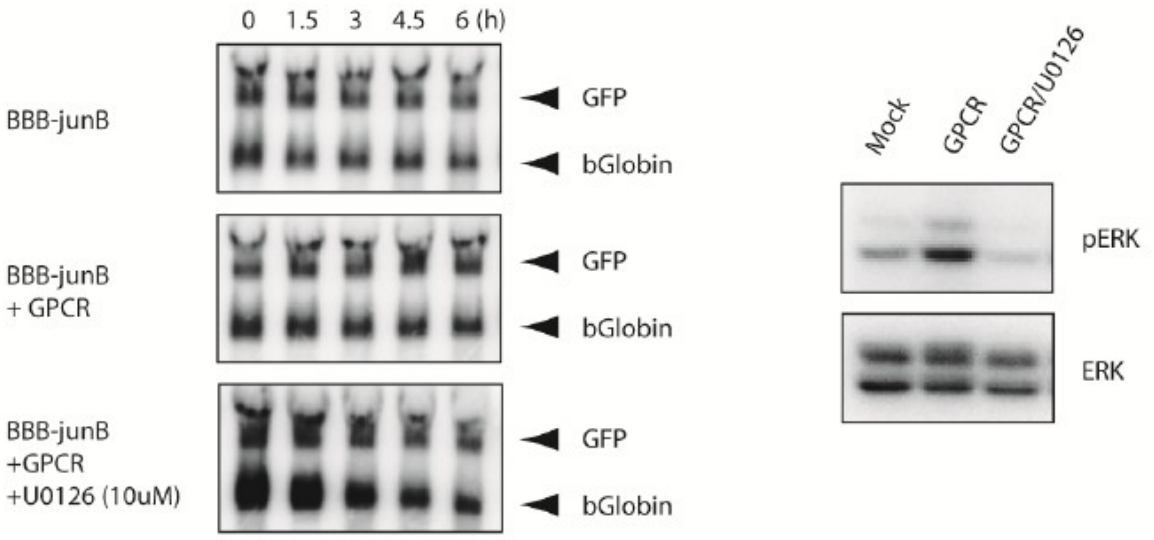

Figure 5.2. KSHV vGPCR inhibits CELF1 mRNA decay by signaling through MEK/ERk pathway. RNA decay experiments were performed with beta-globin transcripts with a GRE-containing segment of the JunB transcript inserted into the beta-globin 3'UTR (BBB-junB). This transcript was transfected into HeLa Tet-off cells and the decay rate of this transcript was measured in the setting of co-transfection with the vGPCR or control plasmid by northern blotting. Additionally, the effect of inhibiting signaling through the MEK/ERK pathway was investigated by measuring the rate of decay of the BBB-junB transcript in the setting of the MEK inhibitor UO126. It can be seen in the top two panels of part A that co-transfection of BBB-junB and the vGPCR reverses the rapid, CELF1-mediated decay of the BBB-junB reporter. In the bottom panel of part A, it can be seen that the addition of 10uM UO126 to the culture medium reversed the stabilizing effect of the vGPCR on the BBB-junB reporter. Western blots against total and phosphor-ERK were performed to assess the degree of MEK inhibition attained by the addition of 10uM UO126 to the culture medium. It can be seen in part B that 10uM UO126 decreased p-ERK down to baseline levels.

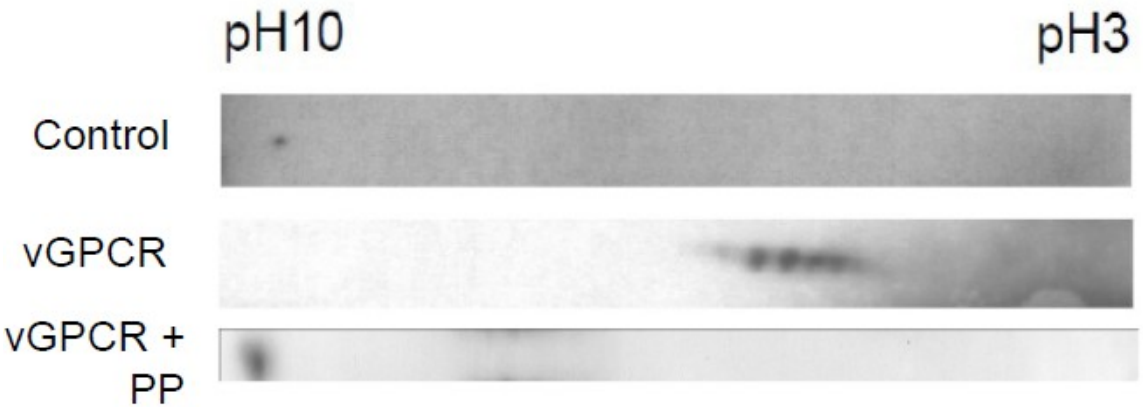

Figure 5.3. CELF1 is phosphorylated in the setting of vGPCR transduction. Cytoplasmic lysates from HeLa cells transduced with control or vGPCR expressing lentivirus were subjected to 2D-gel electrophoresis followed by western blotting for CELF1. It can be seen that in the top panel that in control HeLa cells, CELF1 has an isoelectric point near it's predicted unmodified value of pH 8.0. In the middle panel, it can be seen that HeLa cells transduced with vGPCR expressing lentivirus, the CELF1 isoelectric point exhibited an acidic shift, suggesting post-translational modification. This post-translational modification was confirmed to be phosphorylation due to the return of the CELF1 signal to its original isoelectric point in the bottom panel following treatment of CELF1 from vGPCR transduced HeLa cells with the lambda phosphatase enzyme.

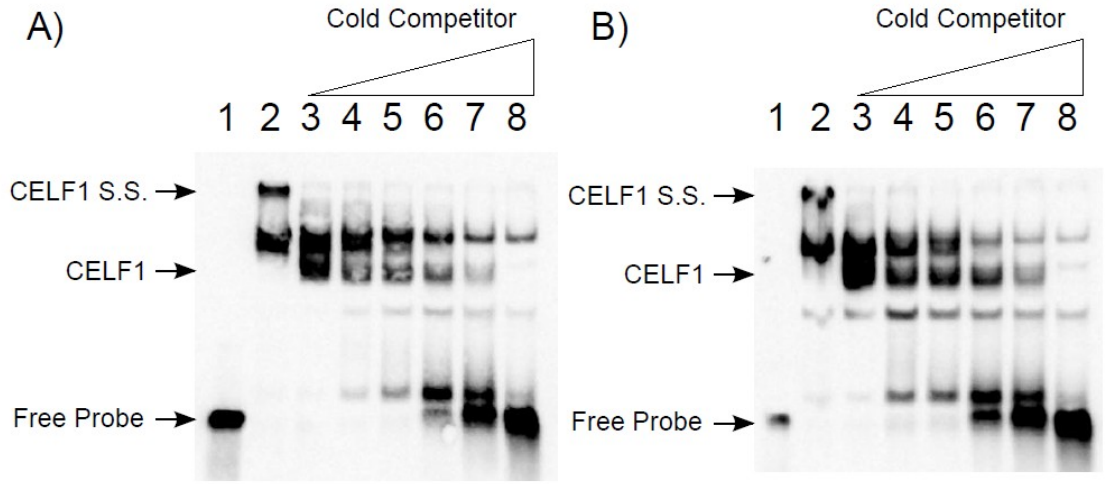

Figure 5.4. CELF1s affinity for the GRE is unchanged in the setting of the KSHV vGPCR. Cytoplasmic lysate from control (A) or vGPCR (B) transduced HeLa cells was incubated with biotinylated RNA containing the GRE, along with appropriate inhibitors. The migration of the free probe can be seen in lane 1. The probe is shifted vertically through the binding of multiple cytoplasmic proteins as can be seen in lane 3. The CELF1 band was identified by its vertical shift upon addition of anti-CELF1 antibody in lane 2. In lanes 3-8, increasing amounts (from 0X to 1024X, in 4 fold increments) of non-labeled GRE probe (Cold Competitor) were added to the reaction. It can be seen that the rate at which CELF1 binding was competed off was nearly identical between the two conditions, suggesting a minimal change in CELF1 binding to RNA in the setting of vGPCR signaling.

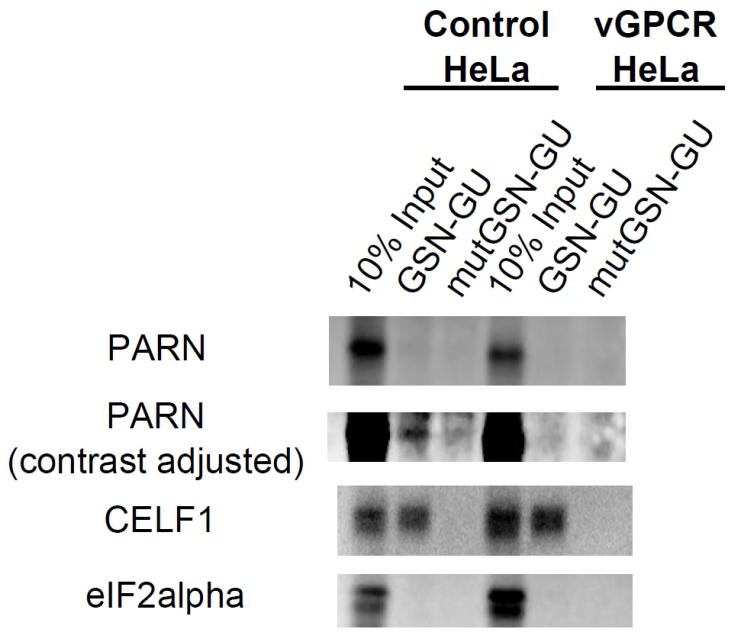

Figure 5.5. CELF1 recruitment of PARN is decreased in the setting of the KSHV vGPCR. GRE binding protein complexes were purified out of cytoplasmic lysates prepared HeLa cells expressing the KSHV vGPCR or an empty plasmid control. Complexes were purified by the addition of a biotinylated bait RNA containing a GRE or a mutated GRE sequence. RNA bait and any bound proteins were isolated by the addition of streptavidin coated agarose beads, and the bound complexes were eluted following stringent washing. The molecular components of the bound complexes were interrogated by western blotting using anti-PARN, anti-CELF1, and anti-eIF2α antibodies. It was observed that CELF1 was purified in both conditions, but PARN was only purified under control conditions. eIF2α served as a negative control, and was not purified under any conditions. None of the investigated proteins co-purified with the mutated GRE bait.

Chapter 6

Regulation of Alternative Polyadenylation by CELF1 During T cell Stimulation.

Introduction

T cell activation is a physiologic process characterized by rapid proliferation and an altered apoptotic response[234]. The T cell transcriptome undergoes dramatic changes within the first few hours of activation[14], and at least half of the gene expression changes following activation are a influenced by post-transcriptional gene regulation[16]. Post-transcriptional gene regulation, and mRNA decay in particular, is a crucial layer of regulation which coordinates the cells rapid and precise changes in gene expression in response to environmental cues. Global quantification of mRNA decay rates in resting and activated T cells revealed widespread alterations in mRNA decay rates following activation[17]. The importance of precisely controlled mRNA decay rates is highlighted by the finding that T cell malignancies have global changes in mRNA decay rates, including changes to the decay rate of several proto-oncogenes and tumor suppressor transcripts[8].

Insight into the mechanism of altered mRNA decay in both T cell activation and malignancy comes from recent studies investigating alternative polyadenylation (APA) in both of these processes. The nuclear maturation of nearly all eukaryotic mRNA involves a process termed polyadenylation, characterized by 3' endonucleolytic cleavage followed by the untemplated addition of a string of adenines to the 3' end of the transcript[168]. In many transcripts, endonucleolytic cleavage and subsequent polyadenylation can occur at multiple sites (polyA sites) within a given 3' untranslated region (3'UTR), resulting in multiple transcript isoforms with variable 3'UTR segments through a process known as alternative polyadenylation (APA) [168]. The utilization of non-canonical polyA sites

for APA has been shown to occur in both the setting of T cell stimulation[173] and cancer[149]. In both systems, activated T cells or cancer cells, it was found that transcripts tended to have shorter 3'UTRs through APA, and the authors found that this resulted in fewer potential microRNA binding sites. It was additionally found that APA in cancer tended to result in stabilization of transcripts, and the authors showed that some, but not all, of this stabilization was due to inhibition of microRNA regulation[149].

The process of alternative polyadenylation is performed by a host of cellular proteins which recognize sequence elements surrounding the polyA site. Upstream of the polyA site lies the polyA signal, which is of the form A[U/A]UAAA and is recognized by the Cleavage and Polyadenylation Specificity Factor (CPSF) protein complex [235]. Downstream of the polyA site lies a U/GU rich sequence which is recognized by the protein Cleavage Stimulatory Factor 64kDa (CSTF-64), or it's related protein tauCSTF-64. Both tau-CSTF-64 and CSTF-64 has recently been shown to bind preferentially to a GU-rich motif lying approximately 25 nucleotides downstream of the polyA site[236, 237]. Through the cooperative binding of these proteins, the remainder of the polyadenylation machinery is recruited to complete the tasks of 3' endonucleolytic cleavage and polyadenylation.

The biological mechanism underlying the choice of alternative polyA sites within a transcript is poorly understood, but some themes have emerged. Non-canonical polyA sites often have "weak" polyA signals containing mutations[238]. Additionally, it is thought that alterations in the functional levels of CSTF-64 promote corresponding changes in the usage of alternative polyA sites. This has been shown in the setting of B-

cell stimulation, in which increases in CSTF-64 lead to alternative polyadenylation of the IgM heavy chain transcript, thus causing a switch from its secreted to membrane bound form[171, 239]. Additionally, LPS stimulation of macrophages results in increased levels of CSTF-64 and a resultant increase in APA[240].

A transcripts rate of decay is determined by short sequence motifs, including miRNA binding sites, harbored typically within the 3'UTR that are recognized by sequence specific RNA binding proteins which serve as adapters to recruit the cellular mRNA decay machinery[34]. The best characterized sequence motif involved in mRNA decay is the AU-Rich Element (ARE) [241]. This element was first discovered as an evolutionarily conserved destabilizing element in the 3'UTR of the TNF-α transcript[19]. It has since been discovered that the ARE represents a family of motifs found in approximately 5-8% of the transcriptome[242]. A number of proteins recognize the ARE, and the biochemical characteristics of the ARE binding protein determine the fate of the harboring transcript[243]. Recently, our group defined another mRNA decay motif through bioinformatic analysis of the 3'UTRs of short lived transcripts in T cells[50]. This motif, termed the GU-Rich Element (GRE), was found to promote the rapid decay of transcripts through binding by CUGBP1 and Etr3 Like Protein 1 (CELF1) [50]. We have recently shown that this ability of CELF1 to promote mRNA decay in T cells is interupted by an activation-dependent phosphorylation event which inhibits CELF1 mRNA binding, corresponding with distinct patterns of expression of GRE containing, CELF1 target transcripts during T cell stimulation[220].

In this study, we utilized a combination of traditional Illumina sequencing and

recently described direct sequencing of mRNA ends to quantify the APA landscape over the first 24 hours of T cell stimulation. We found that CELF1 binds preferentially to transcripts with multiple polyA sites, and present data suggesting that in resting T cells CELF1 binds to GREs immediately downstream of polyA sites. Additionally, we show that CELF1 target transcripts are preferentially shortened by alternative polyadenylation during T cell stimulation. Overall, our findings are consistent with a model of competitive binding between CSTF-64 and CELF1to regulate APA in CELF1 target transcripts.

Results

CELF1 target transcripts shorten by APA at early time points of T cell stimulation

We sought to investigate the impact of alternative polyadenylation on CELF1 target transcripts throughout the early time points of T cell stimulation. Toward this end, we utilized the recently described technique of direct sequencing of RNA ends to quantitatively map polyA sites genome wide in primary human T cells[244]. We performed this sequencing in resting T cells as well as T cells stimulated in vitro with anti-CD3 and anti-CD28 stimulatory antibodies for six and 24 hours from three desperate donors. Reads were mapped to the human genome (GRCh37/HG19) using the SHRiMP 2 algorithm[185]. Reads were allowed to have a maximum of two mismatches and polyA sites were required to be supported by a minimum of 10 reads. Using this criteria, we determined the average number of polyA sites per transcript for each donor at each time point for all transcripts and for the subset of transcripts which have previously been determined to be bound by CELF1 in resting T cells. The results of this analysis are

displayed in figure 6.1. We found that in resting T cells all transcripts had approximately 1.8 detected polyA sites per transcript, compared to CELF1 target transcripts which had approximately 2.8 detected poly a sites per transcript ($p < 0.05$, student T-test). This enrichment in CELF1 target transcripts for detected polyA sites compared to all transcripts increased during T cell stimulation with detection of 3.2 and 3.5 polyA sites per CELF1 target transcript, compared to 2.0 and 2.2 polyA sites per transcript in all transcripts at six and 24 hours, respectively ($p<0.05$ for CELF1 target transcripts compared to all transcripts at six and 24 hours, student T-test).

Given the increased utilization of polyA sites in CELF1 target transcripts, we hypothesized that CELF1 target transcripts would exhibit greater shortening of 3'UTRs compared to all transcripts expressed in T cells. To investigate this we calculated the average 3'UTR length for all transcripts and for CELF1 target transcripts for each donor at zero, six, and 24 hours of T cell stimulation. Each transcripts 3'UTR length was calculated by taking the average length from termination codon to each detected polyA site, with the lengths weighted by the number of reads supporting each polyA site. These values were plotted in figure 6.2, with circles representing the average 3'UTR length from each donor for CELF1 target transcripts and squares representing the average 3'UTR length from each donor for all transcripts. At each time point, CELF1 target transcripts were significantly longer than all transcripts ($p<0.01$ for all time points, student's T-test). Interestingly, a significant trend for decreasing 3'UTR length with stimulation was observed for CELF1 target transcripts (pearson correlation coefficient = -

0.70, p=0.03) but this trend was not observed for all transcripts (pearson's correlation coefficient = -0.52, p=0.15). This data shows that CELF1 target transcripts have longer 3'UTRs than the average transcript present in T cells, and that CELF1 target transcripts displayed a greater degree of 3'UTR shortening upon stimulation than all transcripts. Additionaly, we observed CELF1 transcripts began to undergo significant shortening as a result of APA as early as 6 hours following T cell stimulation.

It has previously been shown that elevated levels of the APA protein CSTF-64 correlate with increased usage of alternative polyA sites[240]. Given the observation that CELF1 target transcripts are shortened by alternative polyadenylation as early as 6 hours following T cell activation, we investigated whether this correlated with increased CSTF-64 expression at these early timepoints. To test this hypothesis, we utilized Illumina high-throughput sequencing to quantify the transcriptome of primary human T cells from three individual donors at 0, 0.5, 1, 3, 6, and 24 hours of stimulation. Reads were mapped to a database of all refseq transcripts using Bowtie2 [245]. The normalized expression of CSTF-64 can be seen in figure 6.3 as a function of time. Each dot represents the expression of an individual donor at each time point. It can be seen that CSTF-64 initially decreases in expression immediately following T cell stimulation, and returns approximately to its baseline by 6 hours of stimulation. By 12 and 24 hours of stimulation, it can be seen that CSTF-64 expression is increased above its expression level in resting T cells, and is approximately two-fold this level by 24 hours of stimulation ($p<0.05$ vs. 0hr, student T-test). This data shows that CSTF-64 expression is

not appreciably different between resting and six hour stimulated T cells, and thus increased CSTF-64 expression is unlikely the reason for the APA mediated shortening of CELF1 target transcripts at this early time point.

CELF1 binds to GREs immediately downstream of polyA sites

Given the result that CELF1 target transcripts utilized exhibited enrichment in polyA site usage, we hypothesized that CELF1 may associate with sequence elements regulating polyA sites. To test this hypothesis, we compiled a database consisting of all experimentally observed polyA sites from all time points of T cell stimulation. Using this database, we then determined the number of unique polyA sites for each CELF1 target transcript. We utilized previously published results from RNA-immunoprecipitation experiments in resting primary human T cells to determine the association of each CELF1 target transcript with CELF1 relative to its association with PolyA Binding Protein (PABP) [220]. This metric was used to assess the association of a given transcript with CELF1 in resting cells irrespective of that transcripts expression level. These values were plotted in figure 6.4, with the number of polyA sites on the x-axis and the CELF1 enrichment on the y-axis and each point representing the respective values for one CELF1 target transcript. The association between these two metrics was assessed by calculating the Pearson's correlation coefficient, and a significant, positive correlation was observed ($r=0.35$, $p<2.2x10^{-16}$). This finding showed that for CELF1 target transcripts, the more polyA sites in the 3'UTR, the more CELF1 associated with that transcript. We investigated whether a similar correlation existed for CELF1 target

transcripts in stimulated T cells. We calculated the correlation coefficient between the enrichment of CELF1stimulated T cell target transcripts in the CELF1 RNA-IP in stimulated T cells relative to the PABP RNA-IP in stimulated T cells, and their number of polyA sites. Interestingly, we found a non-significant trend toward decreasing CELF1 enrichment in stimulated T cells with increasing number of polyA sites ($r=-0.19$, $p=0.11$). This suggests that there may be a decrease in CELF1 association for CELF1 target transcripts in stimulated T cells with increasing numbers of polyA sites. Overall these findings suggest that CELF1 is associating with polyA sites in resting T cells, and that this association may be inhibited in stimulated T cells.

We have previously shown that CELF1 binds to a GU-rich element in the 3'UTR of transcripts. It is also known that a U/GU rich element lies immediately downstream of many polyA sites and binding of this element by CSTF-64 is critical for utilization of the polyA site. We hypothesized that CELF1 binding sites would be enriched immediately downstream of alternative polyA sites in CELF1 target transcripts. To test this hypothesis, we examined the prevalence of the GRE in three regions of 3'UTRs, defined by our database of unique polyA sites. The upstream region was defined as the sequence beginning at the termination codon and ending at the first polyA site. Downstream regions were defined as extending from each polyA site, except the distal most site, to the subsequent polyA site. We examined the prevalence of the GRE within the first 50 nucleotides of the Downstream region(<50 Downstream), as well as from nucleotide 51 to the end of the Downstream region (>50 Downstream). Since each of these sequence

elements from each transcript were of different lengths, the number of GREs in each region was normalized per kilobase of sequence examined. For this analysis, GREs were defined as having the sequence UGU[G/U]UGU[G/U]UGU allowing for 1 mismatch. We performed this analysis on all transcripts and compared these findings to CELF1 target transcripts. It can be seen in figure 6.5a that in the upstream region, there was no enrichment of the GRE in CELF1 target transcripts compared to all transcripts detected in T cells. In the first 50 nucleotides downstream of polyA sites, it can be seen that all transcripts and CELF1 target transcripts were enriched for the GRE compared to the upstream regions. Interestingly, CELF1 target transcripts had a greater enrichment of GREs in the <50 Downstream region compared to all transcripts, suggesting that CELF1 binding sites are enriched in this region in CELF1 target transcripts. Finally, the concentration of GREs in the >50 Downstream regions for all transcripts and CELF1 target transcripts was between that of the upstream region and the <50 Downstream region. We extracted the 100 nucleotides upstream of the GREs detected in the >50 Downstream region of CELF1 target transcripts. We searched for polyA signals of the sequence A[A/U]UAAA allowing for no mismatches in these 100nt segments and a histogram of their locations upstream can be seen in figure 6.5b. It can be seen that there is non-uniform distribution of polyA signals in these segments, with a significant peak at approximately 60nt upstream of GREs. This result suggests that some of these GREs in the >50 downstream region may be contributing to polyA sites which are either not being used, or are being used but at a frequency below our level of detection.

APA sites utilized in CELF1 target transcripts are enriched for downstream GREs

We next sought to determine whether alternative polyA sites utilized in CELF1 target transcripts were enriched for those containing downstream GREs. In order to test this we determined the percentage of GREs downstream of alternative polyA sites which fell within the first 50nt downstream of polyA sites for those sites utilized in all transcripts present in T cells and CELF target transcripts. We determined this percentage for each individual donor at each time point, and the average values of the three donors at each time point are graphed in figure 6.6 with the error bars representing the standard error in the mean of the three donors. It can be seen in resting T cells that there is no statistically significant difference in the distribution of GREs downstream of alternative polyA sites in all transcripts present in T cells compared to CELF1 target transcripts ($p=0.17$, paired T-test). At six and 24 hours of stimulation, however, CELF1 target transcripts had a greater percentage of downstream GREs immediately downstream of utilized polyA sites compared to all transcripts ($p<0.05$ at both six and 24hr of stimulation, paired T-test). These findings suggest that CELF1 target transcripts are preferentially utilizing APA sites with GREs immediately downstream in stimulated T cells, but not in resting T cells.

Discussion

In this study we have utilized high-throughput sequencing approaches to map the alternative polyadenylation landscape with high precision in T cell stimulation. Consistent with previous findings, we showed that the majority of APA in T cell stimulation serves to shorten 3'UTRs. Additionally, we found that APA mediated 3'UTR shortening is detectable in CELF1 target transcripts as early as six hours following

stimulation. We observed exclusion of GREs from 3'UTRs through APA, and the distribution of these GREs in CELF1 target transcripts suggests that the GRE is involved in regulating the process of APA. Overall, our results suggest that APA serves as a modifier of GRE mediated post-transcriptional regulation, and propose a role for CELF1 mediated regulation of APA.

Previous work investigating the role of APA in the proliferation of immune cells found that 3'UTR shortening through APA was a widely observed phenomenon, and that this phenomenon served to regulate miRNA binding sites[173]. Our findings are largely consistent with previous work in that we observed the predominant effect (~75% of all APA was toward shortening transcripts) of APA was to shorten 3'UTRs. Also consistent with previous findings, we observed that the transcriptome as a whole exhibited little 3'UTR shortening at early timepoints following T cell stimulation. This finding was unsurprising as expression of the transcript coding for the APA regulator CSTF-64 was similar between resting T cells and six hour stimulated T cells. Interestingly, despite similar CSTF-64 expression levels, we observed a significant amount of 3'UTR shortening occuring after six hours of stimulation in CELF1 target transcripts, suggesting that these transcripts are regulated in a manner distinct from all transcripts.

Insight into the mechanism underlying the preferential APA of CELF1 target transcripts comes from a previous study of the regulation of APA by the ARE-binding protein HuR. In this study, the authors sought to understand how cells control the expression of

different isoforms of the HuR transcript, and found that HuR bound to regions in the 3'UTR of it's own transcript [246]. It was further shown that HuR bound to an auxillary sequence downstream of an alternative polyA site, and inhibited utilization of the site through competition for binding with factors involved in promoting APA. The authors of this study showed that this served as a negative feedback loop to promote the usage of less efficiently translated HuR isoforms in the setting of abundant HuR protein [246].

The autoregulation of HuR provides a framework for understanding the mechanism behind preferential APA of CELF1 target transcripts, and this postulated mechanism is summarized in figure 6.7. In the present study, we found that the CELF1 binding site, the GRE, was enriched in the region immediately downstream of polyA sites. This biased distribution of the GRE suggests that the recognition of the GRE by CELF1 has a regulatory influence on the process of APA. Additionally, these sites were likely bound by CELF1 in resting T cells, as CELF1 bound more tightly to transcripts with a greater number of polyA sites. Interestingly there was a trend toward CELF1 binding less tightly to transcripts with more polyA sites in stimulated cells, suggesting that the CELF1 binding site was lost through APA in T cell stimulation. Finally, we found that CELF1 target transcripts preferentially utilized polyA sites with downstream GREs in stimulated cells, but not in resting cells. This suggests that CELF1 binding to GREs downstream of these APA sites in resting cells prevented their utilization, a finding consistent with the increased length of CELF1 target transcripts in resting T cells compared to all transcripts. We have previously shown that CELF1 is phosphorylated in stimulated T cells, and this

inhibits CELF1 binding to the GRE. This would be consistent with a model whereby CELF1 releases from the GRE in stimulated cells, and thus allows the utilization of the immediately upstream APA site, resulting in preferential shortening of CELF1 target transcripts. Overall, this mechanism would explain how CELF1 target transcripts are preferentially targeted for APA, and how this occurs in the absence of CSTF-64 upregulation. Further work to investigate the binding pattern of CELF1 in relation to the APA machinery, as well as the kinetics of CELF1 phosphorylation during T cell activation will be important for understanding the interaction between these two regulatory pathways.

Implications for cancer

APA is a broadly observed phenomenon which has been shown to be involved in a number of processes including development[174], long term potentiation of neurons[247], and oncogenesis[149]. Work from our lab has shown that malignancy is associated with global alterations in mRNA decay rates[8]. APA in the setting of cancer was verified in a number of cancer cell lines and primary tumor tissue samples and was found to largely have a stabilizing effect on transcripts[149]. This stabilization as a result of APA was consistent with the loss of cis-acting elements in the distal segments of the regulated 3'UTRs[149]. Importantly, previous work to investigate the impact of miRNA binding site loss through APA on transcript stability in cancer found that miRNA mediated decay was likely a minor contributor to stabilization[149]. The results presented here shed new light on the potential interplay between APA and transcript

stabilization in cancer. GREs are involved in promoting rapid transcript decay, and we have found that the GRE is systematically lost through APA during cellular activation. Since many of the same cellular pathways are activated in cellular activation and malignancy[248, 249], it is possible that the APA program in malignancy is similar to that observed during cellular activation. This suggests that some of the stabilization of mRNA observed in malignancies could be a result of the systematic loss of the GRE. Future experiments to quantify the APA landscape in malignancy will yield insight into the degree to which GRE mediated decay is dysregulated in the setting of cancer and how disruption of these crucial gene control mechanisms may contribute to the oncogenic phenotype.

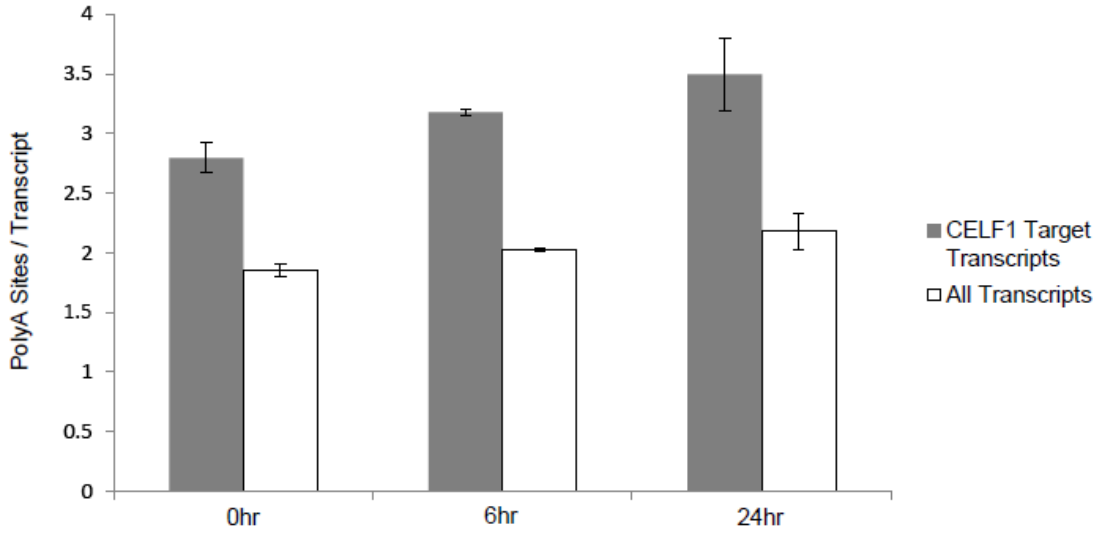

Figure 6.1. CELF1 target transcripts utilize greater numbers of polyA sites than all transcripts. The average number of detected polyA sites per transcript was determined at each time point for CELF target transcripts and compared to all transcripts expressed at each time point. Error bars represent the standard error in the mean between each of three donors, and the difference between CELF1 target transcripts and all transcripts was significant ($p<0.05$, T-test) at all time points. Error bars represent the standard error in the mean from three biological replicates.

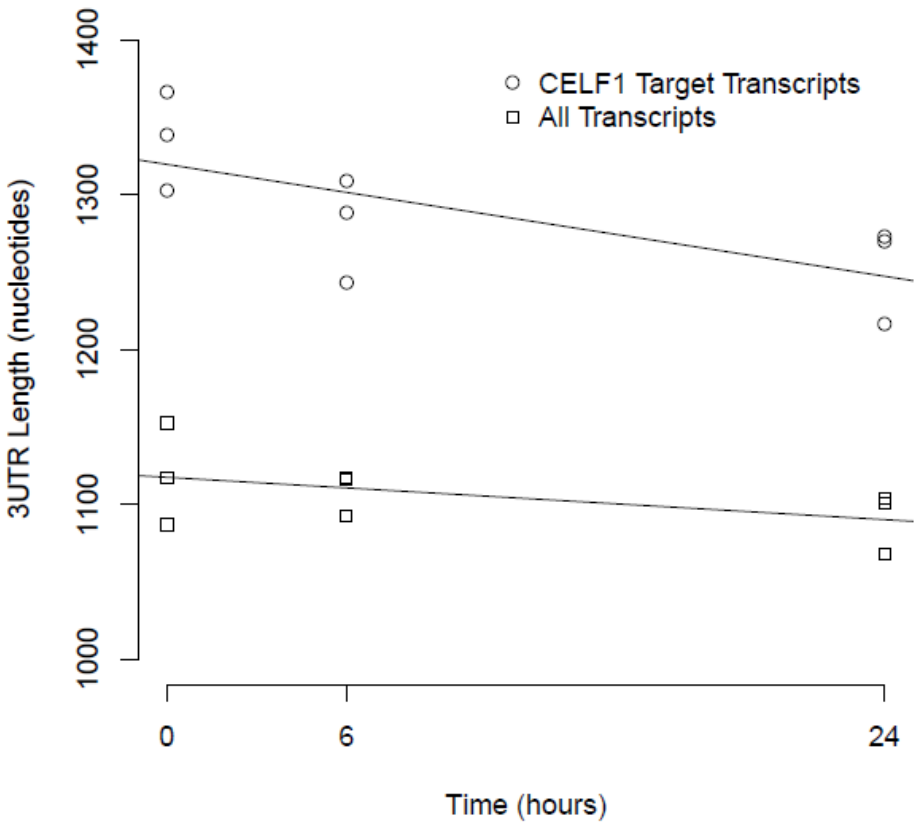

Figure 6.2. CELF1 target transcripts shorten rapidly following T cell stimulation. The average 3'UTR length of each of three donors at each time-point was calculated and the results were plotted for All Transcripts (squares) and CELF1 Target Transcripts (circles). CELF1 Target transcripts were longer than all transcripts at all timepoints ($p<0.01$, Student's T-test). CELF1 target transcript lengths shortened over time (pearson correlation coefficient = -0.70, $p=0.03$), while All Transcripts showed a non-significant trend toward decreased length over time (pearson's correlation coefficient = -0.52, $p=0.15$).

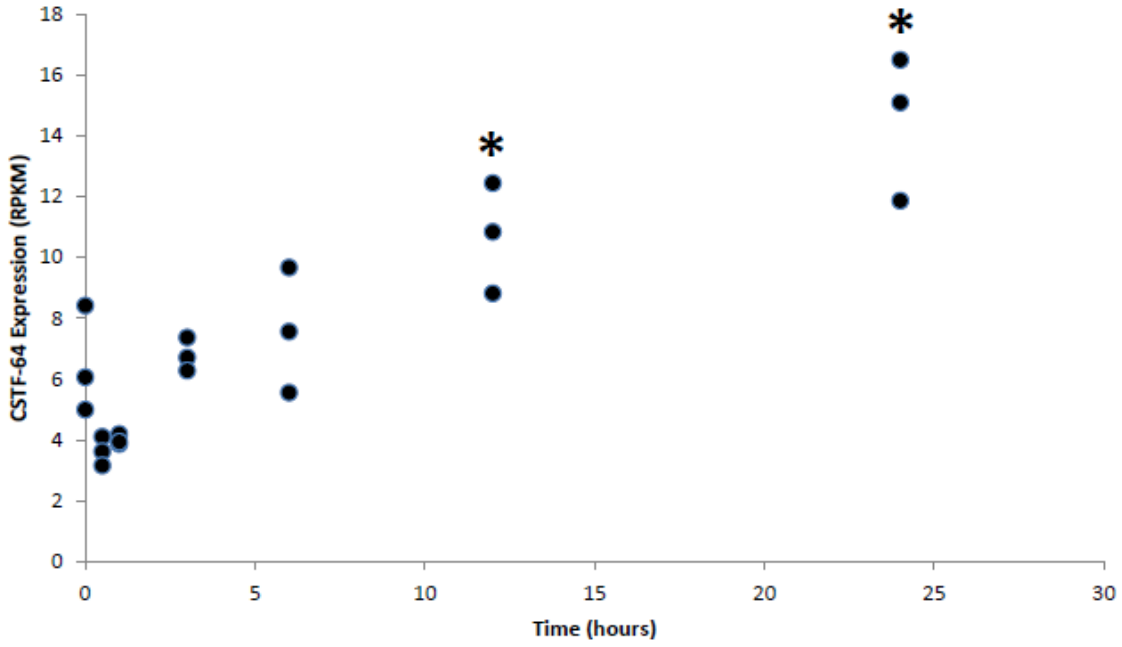

Figure 6.3. CSTF-64 transcript expression is not increased at early time points of T cell stimulation. Deep sequencing data was utilized to assess the expression of the APA regulatory factor CSTF-64 at 0, 0.5, 1, 3, 6, 12, and 24 hours of primary human T cell stimulation from three individual donors. It can be seen that the CSTF-64 transcript is immediately down-regulated upon T cell stimulation, and returns to baseline by approximately six hours. It was found that CSTF-64 expression is increased above baseline by 12 and 24 hours of stimulation ($p<0.05$, T-test).

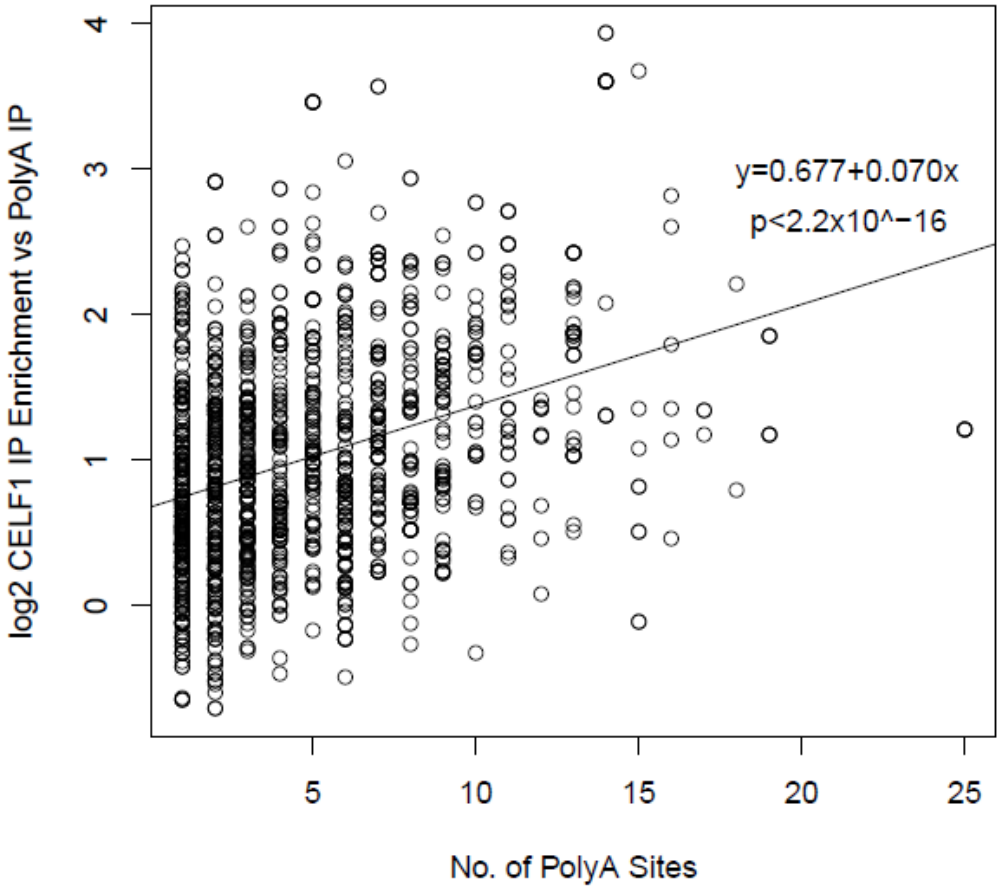

Figure 6.4 CELF1 affinity correlates with increasing number of polyA sites. All polyA sites detected at all time points were compiled into a database, and using this database the number of detected polyA sites per CELF1 target transcript was determined. Utilizing previously generated values for the enrichment of a transcript in CELF1 vs PABP RNA Immunoprecipitations, a significant correlation between CELF1 enrichment in resting T cells and the number of detected polyA sites was observed (Pearson correlation coefficient 0.35, p<2.2x10^-16).

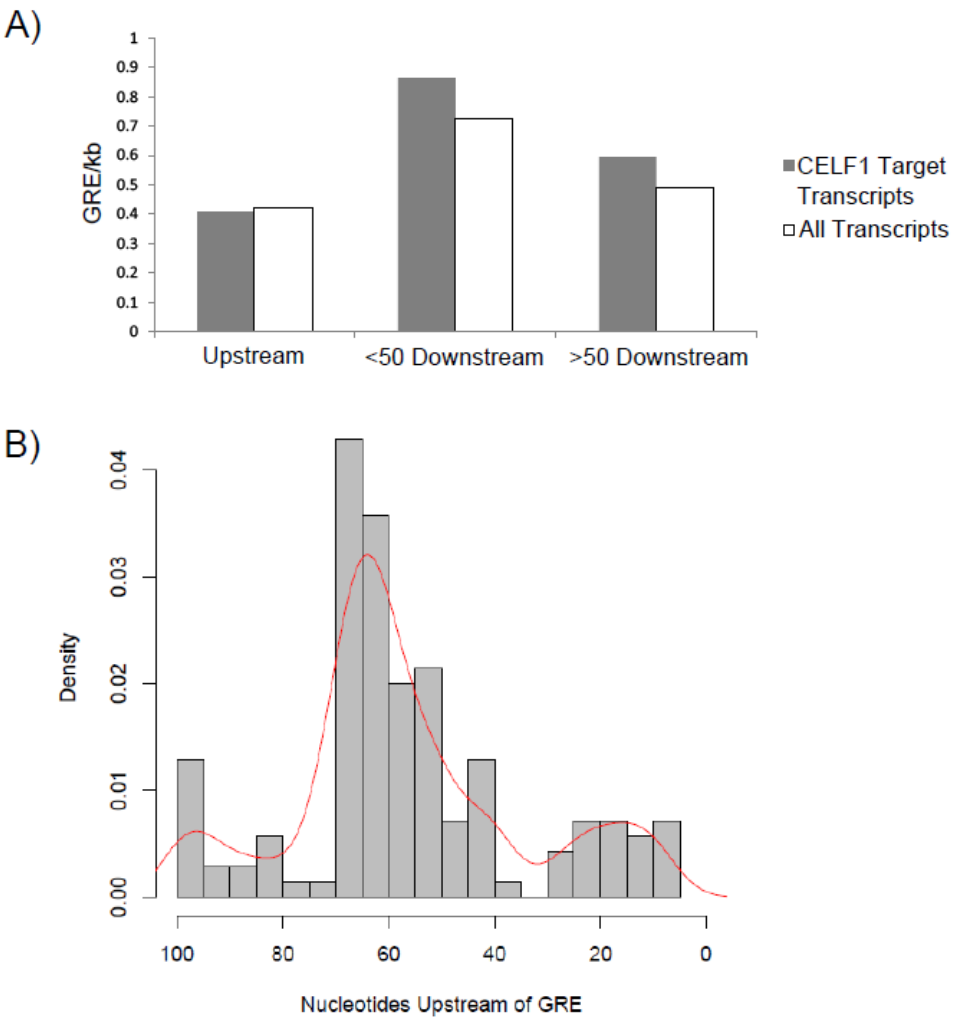

Figure 6.5 GREs are enriched immediately downstream of polyA sites. A) Utilizing the database of all detected polyA sites at all timepoints, 3'UTR regions were extracted from all transcripts and CELF1 transcripts and analyzed for the number of GREs per kilobase of examined sequence (GRE/kb). Upstream regions were from the termination codon to the first polyA site. Downstream regions were from one polyA site to the subsequent 3' polyA site. Downstream regions were separated into the 50 nucleotides immediately 3' of the detected polyA site (<50 Downstream) and the remaining Downstream sequence (>50 Downstream). It can be seen that there was enrichment of

the GRE immediately downstream of polyA sites. B) The GREs detected in the >50 Downstream region were examined for enrichment of polyA signals in the 100nucleotides 5' of the GREs. A significant position specific enrichment of polyA signals was observed approximately 60nt upstream of these GREs, suggesting that these downstream GREs may be associated with unused or undetected polyA sites.

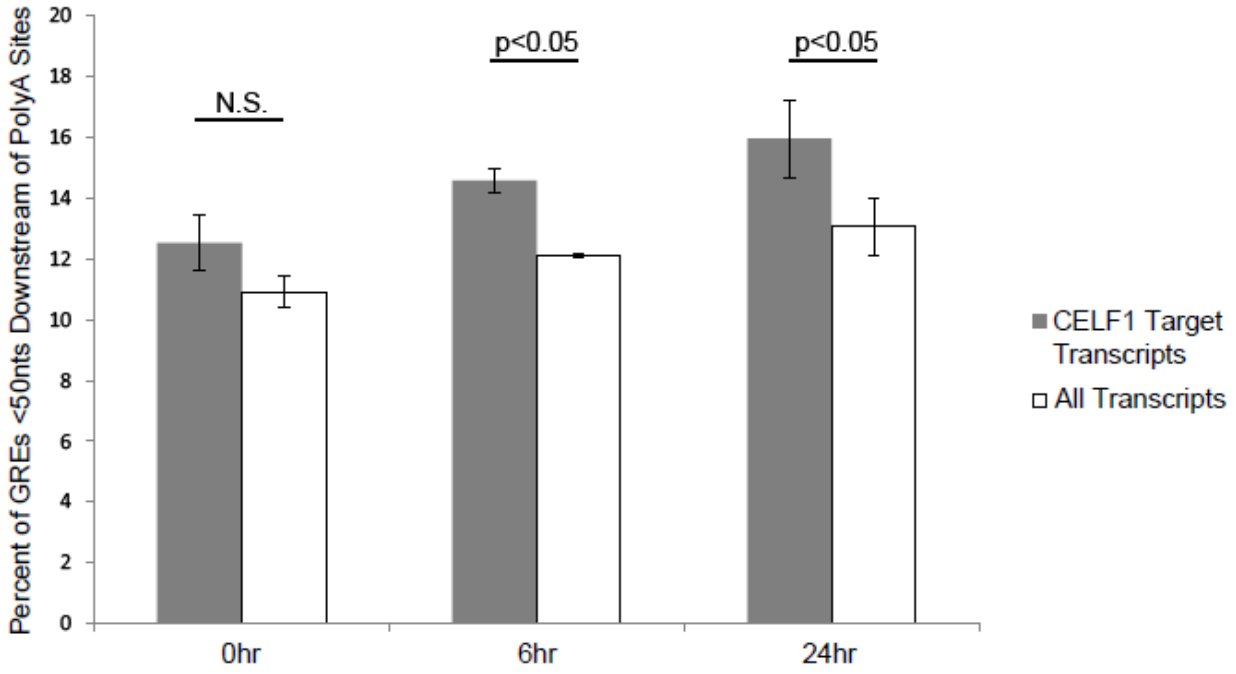

Figure 6.6 CELF1 Target Transcripts Preferentially Utilize PolyA Sites with Downstream GREs in Stimulated T cells. Regions Downstream of PolyA sites in both CELF1 Target Transcripts and All Transcripts detected at each time point from each donor were interrogated for the percent of GREs located in the 50 nucleotides 3' of polyA sites. It was found that in resting T cells there was no difference between the percent of GREs downstream of polyA sites in CELF1 Target Transcripts or All Transcripts. At both 6hrs and 24hrs, polyA sites utilized in CELF1 Target Transcripts had more GREs immediately downstream compared to those polyA sites utilized in All Transcripts. Error bars represent the standard error in the mean from three biological replicates.

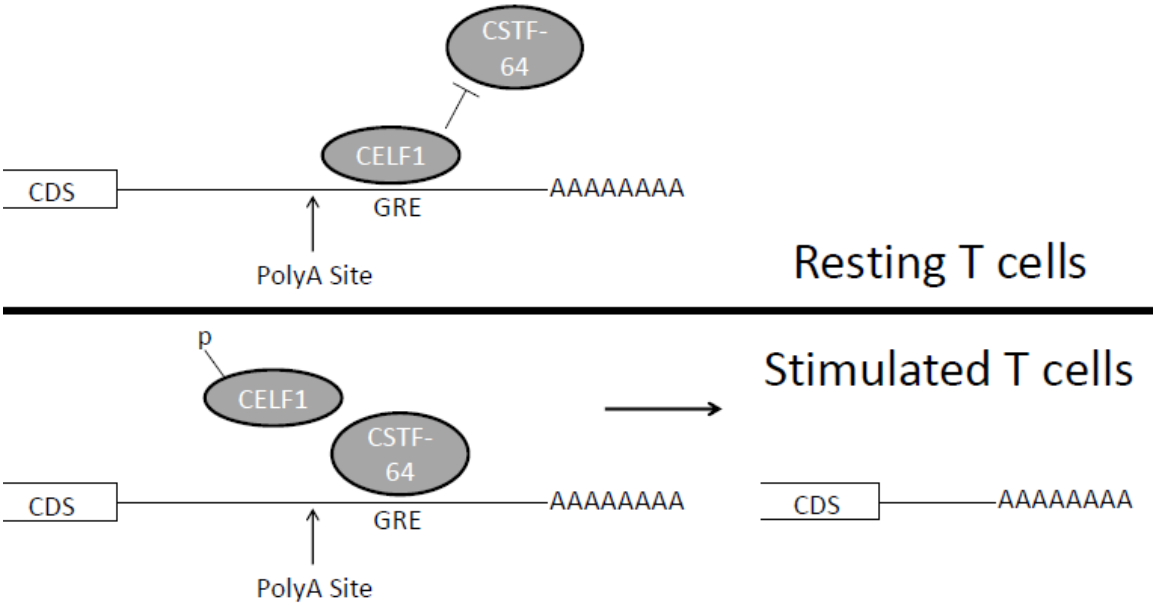

Figure 6.7 Proposed model for CELF1 mediated regulation of APA. In resting T cells, CELF1 binds to GREs immediately downstream of polyA sites and thus prevents the utilization of these sites, resulting in CELF1 target transcripts exhibiting longer 3'UTRs than the average transcript. In Stimulated T cells, CELF1 has been shown to be phosphorylated and release from the GRE. This may then allow for binding by components of the APA machinery, including CSTF-64, to bind to the GRE and promote utilization of the corresponding polyA site. This would result in the preferential shortening of CELF1 target transcripts at early time points following T cell stimulation.

Chapter 7

Discussion & Future Directions

Prokaryotic organisms organize their genetic information into regions of functionally related genes which are under the control of a single promoter and are transcribed en masse, thus ensuring their co-regulation. Eukaryotes, in contrast, lack this organizational scheme and often have functionally related transcripts scattered throughout their genome. How, then, do eukaryotes ensure the proper expression of such a dispersed set of genes? The post-transcriptional operon theory of mRNA decay states that this is achieved partially through regulation of the half-life of the mRNA in the cytoplasm [2]. In this organization, networks of functionally related transcripts contain common cis-elements, often within their 5' or 3' UTRs. These elements are recognized by a trans-acting factor which recognizes the cis-element and promotes transcript stability or de-stability. In this way, the half-life of a whole network of functionally related transcripts are coordinately regulated, and these half-lives can be efficiently altered by changing the abundance or activity of a single factor [2]. Additionally, transcript specific regulation can be obtained through the overlap of two or more individual networks, allowing for combinatorial regulation of those transcripts through co-occurences of distinct cis-elements. CELF1 and the GRE represent one such post-transcriptional operon.

Function of the CELF1 Post-Transcriptional Operon

Through the use of RNA-immunoprecipitation and subsequent microarray analysis of the precipitated mRNA, we were able to define the network of transcripts coordinately regulated by CELF1 in HeLa cells, and primary human T cells in both resting and activated states. Statistical analysis of the annotated functions for the CELF1 bound transcripts in each of the investigated cell types identified enrichment of

transcripts coding for proteins involved in apoptosis, cellular proliferation, and RNA metabolism. These findings are consistent with the results of a similar RNA-immunoprecipitation experiment performed in a murine myoblast cell line [71]. This study identified approximately 900 CELF1 target transcripts, and also found enrichment of the functional categories of cellular proliferation, apoptosis, and RNA processing [71]. The consistent enrichment of these three functional categories in CELF1 target transcripts among multiple cell types suggests that these functions represent a core biological role for the CELF1-GRE interaction. The role of CELF1 in the regulation of apoptosis was highlighted by our finding that the cellular response to apoptotic stimuli was altered under conditions of CELF1 depletion [184]. Furthermore, our finding that CELF1 is inactivated during T cell stimulation, a physiologic state characterized by rapid cellular proliferation, support a role for CELF1 in the regulation of cellular proliferation [220]. CELF1 targets were also enriched for those involved in RNA processing, including several which are themselves involved in post-transcriptional regulation [220]. This phenomenon of post-transcriptional regulatory proteins targeting one another has been observed in studies of other RNA binding proteins [216]. These studies suggest that a complicated regulatory network exists amongst these RNA binding proteins, which may serve to promote the stability of the transcriptome through a series of negative feeback loops [250]. Our findings have identified CELF1 as another "regulator of regulators", and investigation of its interactions with other post-transcriptional networks may provide insight into CELF1s biologic function.

GREs and AREs Represent Orthogonal Regulatory Networks

The development of techniques for the genome-wide identification of target transcripts for RNA-binding proteins such as RNA immunoprecipitation and PAR-CLIP have allowed for the determination of subsets of ARE and GRE-containing transcripts targeted by specific RNA-binding proteins. Analysis of the target transcripts of CELF1 for the prevalence of GRE and AUUUA-pentamer-containing transcripts revealed that the number of transcripts containing both motifs is far less than one would expect by chance alone. This finding suggests that mutually exclusive subsets of transcripts can be defined as transcripts regulated by GREs versus transcripts regulated by AUUUA pentamers. These distinct subsets of transcripts may indicate differential regulation by orthogonal mRNA decay pathways. The fact that these elements are conserved throughout evolution suggests that they co-evolved under evolutionary pressure which served to separate these two pathways. Clues to the source of this evolutionary divergence comes from functional pathway analysis of the networks of transcripts targeted by GRE- and AUUUA-specific ARE-binding proteins. This analysis revealed some overlapping functions, but transcripts targeted by ARE-binding proteins tend to be involved largely in inflammation and cytokine/chemokine pathways [242], whereas GRE-containing target transcripts tend to be largely involved in cell proliferation and apoptosis related pathways [212]. The ability of the cell to regulate these different pathways independently may allow an organism to more precisely respond to environmental changes which could provide an evolutionary advantage.

Regulation of CELF1 Through Phosphorylation

CELF1 is involved in multiple stages of the mRNA lifecycle, and many of these functions of CELF1 have been shown to be regulated through phosphorylation. Evidence that CELF1 mRNA decay could be regulated by phosphorylation comes from studies in a cellular model of myotonic dystrophy type I. In this system, CELF1 was shown to be hyperphosphorylated [140] and this correlated with stabilization of the CELF1 target transcript TNFα [144]. We have shown in two separate systems that phosphorylation inhibits CELF1 mediated mRNA decay. In the setting of T cell stimulation, CELF1 was phosphorylated following activation. This activation dependent CELF1 phosphorylation lead to the inhibition of CELF1 binding to the GRE, which was restored upon phosphatase treatment of CELF1. In the model of KSHV infection and signalling through the oncogenic vGPCR, CELF1 mediated mRNA decay was also inhibited. In this system, CELF1 phosphorylation inhibited its ability to recruit the de-adenylase PARN, but maintained CELF1s ability to bind to the GRE. These observations suggest a model of CELF1 regulation which involves at least two separate kinase cascades. In this model, phosphorylation of CELF1 by one kinase leads to the inhibition of CELF1 binding to RNA, while phosphorylation through another, likely the MEK/ERK pathway, maintains CELF1 binding to RNA but inhibits decay factor recruitment. Both of these pathways result in the inhibition of CELF1 mediated mRNA decay, raising the question of what evolutionary pressure would lead to such redundant regulation. We hypothesize that these two regulatory pathways can be likened to regulation with a hatchet or a scalpel. In the setting of T cell activation, CELF1 phosphorylation inhibits its binding to RNA. As

previously mentioned, CELF1 is involved in multiple stages of the mRNA lifecycle, and CELF1 must bind to RNA to perform its function at every stage [212]. By inhibiting CELF1 mRNA decay through its mRNA binding domain, the cell may be using a figurative hatchet to inhibit all aspects of CELF1 function. In contrast, CELF1 regulation through inhibition of decay factor recruitment, while maintaining RNA binding, may represent a figurative scalpel which serves to inhibit only one of CELF1s many functions. The ability to independently regulate different aspects of CELF1 function may be advantageous for a cell to be able to fine-tune its transcriptome in response to various environmental triggers. Future experiments to delineate the individual kinases responsible for CELF1 phosphorylation, as well as the CELF1 phosphorylation sites will be important for furthering our understanding of this regulatory network.

Regulation of mRNA Decay Networks by Alternative Polyadenylation

Our finding that the GRE is excised as a result of alternative polyadenylation during T cell stimulation reveals an additional layer of complexity to the control of mRNA decay. Previous findings have found specific examples of AREs which are differentially included in transcripts isoforms through APA [215, 246], and this work suggests that ARE-BPs may be involved in regulating excision of the ARE. The work presented here represents the first description of regulation of GRE-mediated mRNA decay through APA. The fidelity of the GRE mediated decay pathway is crucial to the process of T cell activation. Our data shows that APA systematically removes GREs from transcripts and the process very early in T cell activation for CELF1 target transcripts. GREs seem to be preferentially excised from transcripts through APA by six

hours following T cell activation, and these GREs are enriched immediately downstream of APA sites. This biased GRE distribution suggests the GRE, as well as GRE binding proteins such as CELF1, is involved in the regulation of APA. Our data is consistent with a model of competitive binding by CELF1 with the APA machinery to the GRE. When CELF1 is phosphorylated in T cell activation, the APA machinery is able to bind to the GRE and promote preferential APA of CELF1 target transcripts.

While the finding that GRE-mediated decay is inhibited by APA is important for furthering our understanding of how the transcriptome changes during T cell stimulation, greater implications exist for cancer biology. It is known that APA is a feature common to a wide variety of malignant cell lines and tissues [149]. Furthermore, it is well established that aberrant regulation of mRNA decay is involved in the progression of a number of malignancies [6, 251-253]. Our results in T cell stimulation suggest that the APA program preferentially relieves transcripts of regulation from GREs. It would be interesting to study the APA pattern in various malignancies, and determine if a similar preference is observed in a non-physiologic state. If so, this might contribute to the global dysregulation of mRNA decay observed in malignant cells.

Abberant CELF1 Phosphorylation in Cancer

The identity of target transcripts of the CELF1 protein shows that many of these transcripts encode important regulators of cell growth and apoptosis, suggesting that the dysregulation of the GRE/CELF1 network could promote oncogenesis. We found that the ability of cytoplasmic CELF1 to bind to mRNA decreased during early time points of primary T cell stimulation as a result of CELF1 phosphorylation [63]. The decreased

binding activity of CELF1 to mRNA allowed the CELF1 targeted network of transcripts coding for regulators of cell proliferation and apoptosis to rapidly accumulate in the cell. T cells in the early stages of activation share many characteristics with malignant cells, including reduced responses to apoptotic stimuli and rapid proliferation, and many of the same cell pathways activated at early time points in T cell stimulation are aberrantly regulated in cancer (e.g. PKC and MAPK/ERK pathways) [230, 249]. We hypothesized that the activation of these signaling pathways in cancer causes phosphorylation and inactivation of CELF1, and thus contributes to an oncogenic phenotype by preventing the degradation of transcripts that promote cell growth and prevent apoptosis. Our investigations into the mechanism of KSHV mediated oncogenesis supported our hypothesis by revealing that the KSHV encoded vGPCR does indeed cause aberrant CELF1 phosphorylation resulting in the stabilization of CELF1 target transcripts. It will be interesting to investigate whether aberrant CELF1 regulation is observed in other cancer types. Evidence to suggest dysregulated CELF1 function is a broader cancer associated phenomenon comes from observations that loss of CELF1 is associated with the development of certain types of leukemias [218]. Additionally, CELF1 was identified as a potential driver of colorectal cancer in a transposon based screen in mice [219]. These findings are consistent with a model whereby the ubiquitously expressed CELF1 is a master regulator of cellular proliferation and apoptosis which serves as a tumor suppressor. It is possible that in a physiologic setting, CELF1 mRNA decay is regulated in a cell cycle dependent manner. Transient phosphorylation of CELF1 may allow progression through the cell cycle, and its subsequent dephosphorylation may promote

cell cycle exit. The aberrant maintenance of CELF1 phosphorylation would then allow the cell to escape normal cell cycle control and promote uncontrolled proliferation. This would provide a mechanism whereby cancer is able to hijack a physiologic regulatory mechanism to acquire a malignant phenotype. It will be interesting to investigate whether CELF1 phosphorylation occurs in a cell cycle dependent manner, and whether aberrant phosphorylation of CELF1 is observed in a broad set of cancer specimens.

Conclusion

Overall, the work outlined in this thesis has furthered our understanding of the function and regulation of CELF1 post-transcriptional regulation. We have been able to determine the molecular targets of CELF1 and, through these results, understand the functional importance of the CELF1 post-transcriptional operon. Additionally, through bioinformatic analysis of the sequences of CELF1 target transcripts we were able to redefine the GRE in terms of both its sequence content as well as its localization to the 3'UTR. We were further able to show that CELF1s mRNA binding function and its decay factor recruitment are independently regulated, likely through phosphorylation events. Finally, we have preliminary data suggesting that regulation of GRE containing mRNA decay may be regulated through APA during T cell stimulation. Future work will explore the pathways responsible for CELF1 phosphorylation and investigate the phenomenon of aberrant CELF1 regulation in cancer.

References

1. Keene JD, Tenenbaum SA: **Eukaryotic mRNPs may represent posttranscriptional operons**. *Mol Cell* 2002, **9**(6):1161-1167.

2. Keene JD: **RNA regulons: coordination of post-transcriptional events**. *Nat Rev Genet* 2007, **8**(7):533-543.

3. Sandri-Goldin RM, Mendoza GE: **A herpesvirus regulatory protein appears to act post-transcriptionally by affecting mRNA processing**. *Genes Dev* 1992, **6**(5):848-863.

4. Noah DL, Twu KY, Krug RM: **Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAS**. *Virology* 2003, **307**(2):386-395.

5. Dolken L, Perot J, Cognat V, Alioua A, John M, Soutschek J, Ruzsics Z, Koszinowski U, Voinnet O, Pfeffer S: **Mouse cytomegalovirus microRNAs dominate the cellular small RNA profile during lytic infection and show features of posttranscriptional regulation**. *J Virol* 2007, **81**(24):13771-13782.

6. Dixon DA, Tolley ND, King PH, Nabors LB, McIntyre TM, Zimmerman GA, Prescott SM: **Altered expression of the mRNA stability factor HuR promotes cyclooxygenase-2 expression in colon cancer cells**. *J Clin Invest* 2001, **108**(11):1657-1665.

7. Heinonen M, Bono P, Narko K, Chang SH, Lundin J, Joensuu H, Furneaux H, Hla T, Haglund C, Ristimaki A: **Cytoplasmic HuR expression is a prognostic factor in invasive ductal breast carcinoma**. *Cancer research* 2005, **65**(6):2157-2161.

8. Vlasova IA, McNabb J, Raghavan A, Reilly C, Williams DA, Bohjanen KA, Bohjanen PR: **Coordinate stabilization of growth-regulatory transcripts in T cell malignancies**. *Genomics* 2005, **86**(2):159-171.

9. Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, Vrielink JA, Bos AJ, Drost J, Rooijers K *et al*: **The poly(a)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites**. *Cell* 2012, **149**(3):538-553.

10. Bhakar AL, Dolen G, Bear MF: **The Pathophysiology of Fragile X (and What It Teaches Us about Synapses)**. *Annual review of neuroscience* 2012.

11. Wilusz CJ, Wormington M, Peltz SW: **The cap-to-tail guide to mRNA turnover**. *Nature reviews Molecular cell biology* 2001, **2**(4):237-246.

12. Chen CY, Gherzi R, Andersen JS, Gaietta G, Jurchott K, Royer HD, Mann M, Karin M: **Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation**. *Genes Dev* 2000, **14**(10):1236-1248.

13. Anderson P: **Post-transcriptional control of cytokine production**. *Nat Immunol* 2008, **9**(4):353-359.

14. Raghavan A, Dhalla M, Bakheet T, Ogilvie RL, Vlasova IA, Khabar KS, Williams BR, Bohjanen PR: **Patterns of coordinate down-regulation of ARE-containing transcripts following immune cell activation**. *Genomics* 2004, **84**(6):1002-1013.

15. Zhang J, Cado D, Chen A, Kabra NH, Winoto A: **Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1**. *Nature* 1998, **392**(6673):296-300.

16. Cheadle C, Fan J, Cho-Chung YS, Werner T, Ray J, Do L, Gorospe M, Becker KG: **Control of gene expression during T cell activation: alternate regulation of mRNA transcription and mRNA stability**. *BMC Genomics* 2005, **6**:75.

17. Raghavan A, Ogilvie RL, Reilly C, Abelson ML, Raghavan S, Vasdewani J, Krathwohl M, Bohjanen PR: **Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes**. *Nucleic Acids Res* 2002, **30**(24):5529-5538.

18. Shyu AB, Wilkinson MF, van Hoof A: **Messenger RNA regulation: to translate or to degrade**. *Embo J* 2008, **27**(3):471-481.

19. Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A: **Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators**. *Proc Natl Acad Sci U S A* 1986, **83**(6):1670-1674.

20. Shaw G, Kamen R: **A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation**. *Cell* 1986, **46**(5):659-667.

21. Gillis P, Malter JS: **The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine, and oncogene mRNAs**. *J Biol Chem* 1991, **266**(5):3172-3177.

22. Chen CY, Shyu AB: **AU-rich elements: characterization and importance in mRNA degradation**. *Trends in biochemical sciences* 1995, **20**(11):465-470.

23. Baou M, Norton JD, Murphy JJ: **AU-rich RNA binding proteins in hematopoiesis and leukemogenesis**. *Blood* 2011, **118**(22):5732-5740.

24. Deschenes-Furry J, Perrone-Bizzozero N, Jasmin BJ: **The RNA-binding protein HuD: a regulator of neuronal differentiation, maintenance and plasticity**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2006, **28**(8):822-833.

25. Brennan CM, Steitz JA: **HuR and mRNA stability**. *Cellular and molecular life sciences : CMLS* 2001, **58**(2):266-277.

26. Soller M, Li M, Haussmann IU: **Determinants of ELAV gene-specific regulation**. *Biochemical Society transactions* 2010, **38**(4):1122-1124.

27. Raghavan A, Robison RL, McNabb J, Miller CR, Williams DA, Bohjanen PR: **HuA and tristetraprolin are induced following T cell activation and display distinct but overlapping RNA binding specificities**. *J Biol Chem* 2001, **276**(51):47958-47965.

28. Schoenberg DR, Maquat LE: **Regulation of cytoplasmic mRNA decay**. *Nat Rev Genet* 2012, **13**(4):246-259.

29. Lai WS, Carballo E, Strum JR, Kennington EA, Phillips RS, Blackshear PJ: **Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA**. *Mol Cell Biol* 1999, **19**(6):4311-4323.

30. Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M, Karin M: **AU binding proteins recruit the exosome to degrade ARE-containing mRNAs**. *Cell* 2001, **107**(4):451-464.

31. Carballo E, Lai WS, Blackshear PJ: **Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin**. *Science* 1998, **281**(5379):1001-1005.

32. Edmonds M: **A history of poly A sequences: from formation to factors to function**. *Progress in nucleic acid research and molecular biology* 2002, **71**:285-389.

33. Shyu AB, Belasco JG, Greenberg ME: **Two distinct destabilizing elements in the c-fos message trigger deadenylation as a first step in rapid mRNA decay**. *Genes Dev* 1991, **5**(2):221-231.

34. Garneau NL, Wilusz J, Wilusz CJ: **The highways and byways of mRNA decay**. *Nature reviews Molecular cell biology* 2007, **8**(2):113-126.

35. Lai WS, Kennington EA, Blackshear PJ: **Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease**. *Mol Cell Biol* 2003, **23**(11):3798-3812.

36. Sandler H, Kreth J, Timmers HT, Stoecklin G: **Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin**. *Nucleic Acids Res* 2011, **39**(10):4373-4386.

37. Lykke-Andersen S, Tomecki R, Jensen TH, Dziembowski A: **The eukaryotic RNA exosome: same scaffold but variable catalytic subunits**. *RNA Biol* 2011, **8**(1):61-66.

38. Ingelfinger D, Arndt-Jovin DJ, Luhrmann R, Achsel T: **The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci**. *Rna* 2002, **8**(12):1489-1501.

39. Franks TM, Lykke-Andersen J: **The control of mRNA decapping and P-body formation**. *Mol Cell* 2008, **32**(5):605-615.

40. Stoecklin G, Mayo T, Anderson P: **ARE-mRNA degradation requires the 5'-3' decay pathway**. *EMBO reports* 2006, **7**(1):72-77.

41. Hau HH, Walsh RJ, Ogilvie RL, Williams DA, Reilly CS, Bohjanen PR: **Tristetraprolin recruits functional mRNA decay complexes to ARE sequences**. *Journal of cellular biochemistry* 2007, **100**(6):1477-1492.

42. Mazan-Mamczarz K, Galban S, Lopez de Silanes I, Martindale JL, Atasoy U, Keene JD, Gorospe M: **RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8354-8359.

43. Kawai T, Lal A, Yang X, Galban S, Mazan-Mamczarz K, Gorospe M: **Translational control of cytochrome c by RNA-binding proteins TIA-1 and HuR**. *Mol Cell Biol* 2006, **26**(8):3295-3307.

44. Durie D, Lewis SM, Liwak U, Kisilewicz M, Gorospe M, Holcik M: **RNA-binding protein HuR mediates cytoprotection through stimulation of XIAP translation**. *Oncogene* 2011, **30**(12):1460-1469.

45. Ishimaru D, Ramalingam S, Sengupta TK, Bandyopadhyay S, Dellis S, Tholanikunnel BG, Fernandes DJ, Spicer EK: **Regulation of Bcl-2 expression by HuR in HL60 leukemia cells and A431 carcinoma cells**. *Molecular cancer research : MCR* 2009, **7**(8):1354-1366.

46. Kullmann M, Gopfert U, Siewe B, Hengst L: **ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5'UTR**. *Genes Dev* 2002, **16**(23):3087-3099.

47. Kim HH, Kuwano Y, Srikantan S, Lee EK, Martindale JL, Gorospe M: **HuR recruits let-7/RISC to repress c-Myc expression**. *Genes Dev* 2009, **23**(15):1743-1748.

48. Leandersson K, Riesbeck K, Andersson T: **Wnt-5a mRNA translation is suppressed by the Elav-like protein HuR in human breast epithelial cells**. *Nucleic Acids Res* 2006, **34**(14):3988-3999.

49. Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z: **An mRNA m7G cap binding-like motif within human Ago2 represses translation**. *Cell* 2007, **129**(6):1141-1151.

50. Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB *et al*: **Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1**. *Mol Cell* 2008, **29**(2):263-270.

51. Delaunay J, Le Mee G, Ezzeddine N, Labesse G, Terzian C, Capri M, Ait-Ahmed O: **The Drosophila Bruno paralogue Bru-3 specifically binds the EDEN translational repression element**. *Nucleic Acids Res* 2004, **32**(10):3070-3082.

52. Vlasova-St Louis I, Bohjanen PR: **Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1**. *Curr Opin Genet Dev* 2011.

53. Choi DK, Ito T, Tsukahara F, Hirai M, Sakaki Y: **Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein**. *Gene* 1999, **237**(1):135-142.

54. Good PJ, Chen Q, Warner SJ, Herring DC: **A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator**. *J Biol Chem* 2000, **275**(37):28583-28592.

55. Li D, Bachinski LL, Roberts R: **Genomic organization and isoform-specific tissue expression of human NAPOR (CUGBP2) as a candidate gene for familial arrhythmogenic right ventricular dysplasia**. *Genomics* 2001, **74**(3):396-401.

56. Kress C, Gautier-Courteille C, Osborne HB, Babinet C, Paillard L: **Inactivation of CUG-BP1/CELF1 causes growth, viability, and spermatogenesis defects in mice**. *Mol Cell Biol* 2007, **27**(3):1146-1157.

57. Choi DK, Yoo KW, Hong SK, Rhee M, Sakaki Y, Kim CH: **Isolation and expression of Napor/CUG-BP2 in embryo development**. *Biochem Biophys Res Commun* 2003, **305**(3):448-454.

58. Yang Y, Mahaffey CL, Berube N, Maddatu TP, Cox GA, Frankel WN: **Complex seizure disorder caused by Brunol4 deficiency in mice**. *PLoS Genet* 2007, **3**(7):e124.

59. Wu J, Li C, Zhao S, Mao B: **Differential expression of the Brunol/CELF family genes during Xenopus laevis early development**. *Int J Dev Biol* 2010, **54**(1):209-214.

60. Kim-Ha J, Kerr K, Macdonald PM: **Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential**. *Cell* 1995, **81**(3):403-412.

61. Paillard L, Omilli F, Legagneux V, Bassez T, Maniey D, Osborne HB: **EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos**. *Embo J* 1998, **17**(1):278-287.

62. Milne CA, Hodgkin J: **ETR-1, a homologue of a protein linked to myotonic dystrophy, is essential for muscle development in Caenorhabditis elegans**. *Current biology : CB* 1999, **9**(21):1243-1246.

63. Hashimoto Y, Suzuki H, Kageyama Y, Yasuda K, Inoue K: **Bruno-like protein is localized to zebrafish germ plasm during the early cleavage stages**. *Gene Expr Patterns* 2006, **6**(2):201-205.

64. Brimacombe KR, Ladd AN: **Cloning and embryonic expression patterns of the chicken CELF family**. *Dev Dyn* 2007, **236**(8):2216-2224.

65. Graindorge A, Le Tonqueze O, Thuret R, Pollet N, Osborne HB, Audic Y: **Identification of CUG-BP1/EDEN-BP target mRNAs in Xenopus tropicalis**. *Nucleic Acids Res* 2008, **36**(6):1861-1870.

66. Paillard L, Legagneux V, Maniey D, Osborne HB: **c-Jun ARE targets mRNA deadenylation by an EDEN-BP (embryo deadenylation element-binding protein)-dependent pathway**. *J Biol Chem* 2002, **277**(5):3232-3235.

67. Osborne HB, Gautier-Courteille C, Graindorge A, Barreau C, Audic Y, Thuret R, Pollet N, Paillard L: **Post-transcriptional regulation in Xenopus embryos: role and targets of EDEN-BP**. *Biochemical Society transactions* 2005, **33**(Pt 6):1541-1543.

68. Moraes KC, Wilusz CJ, Wilusz J: **CUG-BP binds to RNA substrates and recruits PARN deadenylase**. *Rna* 2006, **12**(6):1084-1091.

69. Gautier-Courteille C, Le Clainche C, Barreau C, Audic Y, Graindorge A, Maniey D, Osborne HB, Paillard L: **EDEN-BP-dependent post-transcriptional regulation of gene expression in Xenopus somitic segmentation**. *Development* 2004, **131**(24):6107-6117.

70. Lee JE, Cooper TA: **Pathogenic mechanisms of myotonic dystrophy**. *Biochemical Society transactions* 2009, **37**(Pt 6):1281-1286.

71. Lee JE, Lee JY, Wilusz J, Tian B, Wilusz CJ: **Systematic analysis of cis-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells**. *PLoS One* 2010, **5**(6):e11201.

72. Timchenko LT, Timchenko NA, Caskey CT, Roberts R: **Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy**. *Hum Mol Genet* 1996, **5**(1):115-121.

73. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, Roberts R, Caskey CT, Swanson MS: **Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy**. *Nucleic Acids Res* 1996, **24**(22):4407-4414.

74. Marquis J, Paillard L, Audic Y, Cosson B, Danos O, Le Bec C, Osborne HB: **CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding**. *The Biochemical journal* 2006, **400**(2):291-301.

75. Goraczniak R, Gunderson SI: **The regulatory element in the 3'-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1**. *J Biol Chem* 2008, **283**(4):2286-2296.

76. Takahashi N, Sasagawa N, Suzuki K, Ishiura S: **The CUG-binding protein binds specifically to UG dinucleotide repeats in a yeast three-hybrid system**. *Biochem Biophys Res Commun* 2000, **277**(2):518-523.

77. Mori D, Sasagawa N, Kino Y, Ishiura S: **Quantitative analysis of CUG-BP1 binding to RNA repeats**. *J Biochem* 2008, **143**(3):377-383.

78. Paillard L, Omilli F, Legagneux V, Bassez T, Maniey D, Osborne HB: **EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos**. *Embo J* 1998, **17**(1):278-287.

79. Paillard L, Legagneux V, Beverley Osborne H: **A functional deadenylation assay identifies human CUG-BP as a deadenylation factor**. *Biol Cell* 2003, **95**(2):107-113.

80. Suzuki H, Jin Y, Otani H, Yasuda K, Inoue K: **Regulation of alternative splicing of alpha-actinin transcript by Bruno-like proteins**. *Genes Cells* 2002, **7**(2):133-141.

81. Bonnet-Corven S, Audic Y, Omilli F, Osborne HB: **An analysis of the sequence requirements of EDEN-BP for specific RNA binding**. *Nucleic Acids Res* 2002, **30**(21):4667-4674.

82. Salisbury E, Sakai K, Schoser B, Huichalaf C, Schneider-Gold C, Nguyen H, Wang GL, Albrecht JH, Timchenko LT: **Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1**. *Exp Cell Res* 2008, **314**(11-12):2266-2278.

83. Cosson B, Gautier-Courteille C, Maniey D, Ait-Ahmed O, Lesimple M, Osborne HB, Paillard L: **Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding**. *Biol Cell* 2006, **98**(11):653-665.

84. Wu C, Alwine JC: **Secondary structure as a functional feature in the downstream region of mammalian polyadenylation signals**. *Mol Cell Biol* 2004, **24**(7):2789-2796.

85. Mooers BH, Logue JS, Berglund JA: **The structural basis of myotonic dystrophy from the crystal structure of CUG repeats**. *Proc Natl Acad Sci U S A* 2005, **102**(46):16626-16631.

86. Han J, Cooper TA: **Identification of CELF splicing activation and repression domains in vivo**. *Nucleic Acids Res* 2005, **33**(9):2769-2780.

87. Ladd AN, Charlet N, Cooper TA: **The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing**. *Mol Cell Biol* 2001, **21**(4):1285-1296.

88. Ladd AN, Charlet N, Cooper TA: **The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing**. *Mol Cell Biol* 2001, **21**(4):1285-1296.

89. Tsuda K, Kuwasako K, Takahashi M, Someya T, Inoue M, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A *et al*: **Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3**. *Nucleic Acids Res* 2009, **37**(15):5151-5166.

90. Tripsianes K, Sattler M: **Repeat recognition**. *Structure* 2010, **18**(10):1228-1229.

91. Edwards J, Malaurie E, Kondrashov A, Long J, de Moor CH, Searle MS, Emsley J: **Sequence determinants for the tandem recognition of UGU and CUG rich RNA elements by the two N--terminal RRMs of CELF1**. *Nucleic Acids Res* 2011, **39**(19):8638-8650.

92. Teplova M, Song J, Gaw HY, Teplov A, Patel DJ: **Structural insights into RNA recognition by the alternate-splicing regulator CUG-binding protein 1**. *Structure*, **18**(10):1364-1377.

93. Salisbury E, Sakai K, Schoser B, Huichalaf C, Schneider-Gold C, Nguyen H, Wang GL, Albrecht JH, Timchenko LT: **Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1**. *Exp Cell Res* 2008, **314**(11-12):2266-2278.

94. Han J, Cooper TA: **Identification of CELF splicing activation and repression domains in vivo**. *Nucleic Acids Res* 2005, **33**(9):2769-2780.

95. Panasenko OM, Chekanov AV, Vlasova, II, Sokolov AV, Ageeva KV, Pulina MO, Cherkalina OS, Vasil'ev VB: **[A study of the effect of ceruloplasmin and lactoferrin on the chlorination activity of leukocytic myeloperoxidase using the chemiluminescence method]**. *Biofizika* 2008, **53**(4):573-581.

96. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes**. *Nature* 2008, **456**(7221):470-476.

97. Ladd AN, Cooper TA: **Finding signals that regulate alternative splicing in the post-genomic era**. *Genome Biol* 2002, **3**(11):reviews0008.

98. Voelker RB, Berglund JA: **A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing**. *Genome Res* 2007, **17**(7):1023-1033.

99. Dembowski JA, Grabowski PJ: **The CUGBP2 splicing factor regulates an ensemble of branchpoints from perimeter binding sites with implications for autoregulation**. *PLoS Genet* 2009, **5**(8):e1000595.

100. Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS, Ohno K: **CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay**. *Sci Rep* 2012, **2**:209.

101. Ladd AN, Stenberg MG, Swanson MS, Cooper TA: **Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development**. *Dev Dyn* 2005, **233**(3):783-793.

102. Ho TH, Bundman D, Armstrong DL, Cooper TA: **Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy**. *Hum Mol Genet* 2005, **14**(11):1539-1547.

103. Philips AV, Timchenko LT, Cooper TA: **Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy**. *Science* 1998, **280**(5364):737-741.

104. Savkur RS, Philips AV, Cooper TA: **Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy**. *Nature genetics* 2001, **29**(1):40-47.

105. Charlet BN, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA: **Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing**. *Mol Cell* 2002, **10**(1):45-53.

106. Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, Cannon SC, Thornton CA: **Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy**. *Mol Cell* 2002, **10**(1):35-44.

107. Ladd AN, Taffet G, Hartley C, Kearney DL, Cooper TA: **Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy**. *Mol Cell Biol* 2005, **25**(14):6267-6278.

108. Ladd AN, Nguyen NH, Malhotra K, Cooper TA: **CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing**. *J Biol Chem* 2004, **279**(17):17756-17764.

109. Gromak N, Matlin AJ, Cooper TA, Smith CW: **Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein**. *Rna* 2003, **9**(4):443-456.

110. Charlet BN, Logan P, Singh G, Cooper TA: **Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing**. *Mol Cell* 2002, **9**(3):649-658.

111. Kino Y, Washizu C, Oma Y, Onishi H, Nezu Y, Sasagawa N, Nukina N, Ishiura S: **MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1**. *Nucleic Acids Res* 2009, **37**(19):6477-6490.

112. Dujardin G, Buratti E, Charlet-Berguerand N, Martins de Araujo M, Mbopda A, Le Jossic-Corcos C, Pagani F, Ferec C, Corcos L: **CELF proteins regulate CFTR pre-mRNA splicing: essential role of the divergent domain of ETR-3**. *Nucleic Acids Res* 2010, **38**(20):7273-7285.

113. Barron VA, Zhu H, Hinman MN, Ladd AN, Lou H: **The neurofibromatosis type I pre-mRNA is a novel target of CELF protein-mediated splicing regulation**. *Nucleic Acids Res* 2010, **38**(1):253-264.

114. Koebis M, Ohsawa N, Kino Y, Sasagawa N, Nishino I, Ishiura S: **Alternative splicing of myomesin 1 gene is aberrantly regulated in myotonic dystrophy type 1**. *Genes Cells* 2011, **16**(9):961-972.

115. Terenzi F, Brimacombe KR, Penn MS, Ladd AN: **CELF-mediated alternative splicing is required for cardiac function during early, but not later, postnatal life**. *J Mol Cell Cardiol* 2009, **46**(3):395-404.

116. Berger DS, Moyer M, Kliment GM, van Lunteren E, Ladd AN: **Expression of a dominant negative CELF protein in vivo leads to altered muscle organization, fiber size, and subtype**. *PLoS One* 2011, **6**(4):e19274.

117. Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA: **A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart**. *Proc Natl Acad Sci U S A* 2008, **105**(51):20333-20338.

118. Ladd AN, Taffet G, Hartley C, Kearney DL, Cooper TA: **Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy**. *Mol Cell Biol* 2005, **25**(14):6267-6278.

119. Dasgupta T, Ladd AN: **The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins**. *Wiley Interdiscip Rev RNA* 2012, **3**(1):104-121.

120. Orengo JP, Ward AJ, Cooper TA: **Alternative splicing dysregulation secondary to skeletal muscle regeneration**. *Ann Neurol* 2011, **69**(4):681-690.

121. Horb LD, Horb ME: **BrunoL1 regulates endoderm proliferation through translational enhancement of cyclin A2 mRNA**. *Dev Biol* 2010, **345**(2):156-169.

122. Nakamura A, Sato K, Hanyu-Nakamura K: **Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis**. *Dev Cell* 2004, **6**(1):69-78.

123. Timchenko NA, Welm AL, Lu X, Timchenko LT: **CUG repeat binding protein (CUGBP1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms**. *Nucleic Acids Res* 1999, **27**(22):4517-4525.

124. Bae EJ, Kim SG: **Enhanced CCAAT/enhancer-binding protein beta-liver-enriched inhibitory protein production by Oltipraz, which accompanies CUG repeat-binding protein-1 (CUGBP1) RNA-binding protein activation, leads to inhibition of preadipocyte differentiation**. *Molecular pharmacology* 2005, **68**(3):660-669.

125. Karagiannides I, Thomou T, Tchkonia T, Pirtskhalava T, Kypreos KE, Cartwright A, Dalagiorgou G, Lash TL, Farmer SR, Timchenko NA *et al*: **Increased CUG triplet repeat-binding protein-1 predisposes to impaired adipogenesis with aging**. *J Biol Chem* 2006, **281**(32):23025-23033.

126. Timchenko NA, Wang GL, Timchenko LT: **RNA CUG-binding protein 1 increases translation of 20-kDa isoform of CCAAT/enhancer-binding protein beta by interacting with the alpha and beta subunits of eukaryotic initiation translation factor 2**. *J Biol Chem* 2005, **280**(21):20549-20557.

127. Jin J, Wang GL, Timchenko L, Timchenko NA: **GSK3beta and aging liver**. *Aging (Albany NY)* 2009, **1**(6):582-585.

128. Timchenko LT, Salisbury E, Wang GL, Nguyen H, Albrecht JH, Hershey JW, Timchenko NA: **Age-specific CUGBP1-eIF2 complex increases translation of CCAAT/enhancer-binding protein beta in old liver**. *J Biol Chem* 2006, **281**(43):32806-32819.

129. Iakova P, Wang GL, Timchenko L, Michalak M, Pereira-Smith OM, Smith JR, Timchenko NA: **Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate**. *Embo J* 2004, **23**(2):406-417.

130. Lian XJ, Gallouzi IE: **Oxidative Stress Increases the Number of Stress Granules in Senescent Cells and Triggers a Rapid Decrease in p21waf1/cip1 Translation**. *J Biol Chem* 2009, **284**(13):8877-8887.

131. Fujimura K, Kano F, Murata M: **Dual localization of the RNA binding protein CUGBP-1 to stress granule and perinucleolar compartment**. *Exp Cell Res* 2008, **314**(3):543-553.

132. Huichalaf C, Sakai K, Jin B, Jones K, Wang GL, Schoser B, Schneider-Gold C, Sarkar P, Pereira-Smith OM, Timchenko N *et al*: **Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells**. *FASEB J* 2010, **24**(10):3706-3719.

133. Barreau C, Watrin T, Beverley Osborne H, Paillard L: **Protein expression is increased by a class III AU-rich element and tethered CUG-BP1**. *Biochem Biophys Res Commun* 2006, **347**(3):723-730.

134. Timchenko NA, Iakova P, Cai ZJ, Smith JR, Timchenko LT: **Molecular basis for impaired muscle differentiation in myotonic dystrophy**. *Mol Cell Biol* 2001, **21**(20):6927-6938.

135. Timchenko NA, Patel R, Iakova P, Cai ZJ, Quan L, Timchenko LT: **Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis**. *J Biol Chem* 2004, **279**(13):13129-13139.

136. Fox JT, Stover PJ: **Mechanism of the internal ribosome entry site-mediated translation of serine hydroxymethyltransferase 1**. *J Biol Chem* 2009, **284**(45):31085-31096.

137. Woeller CF, Fox JT, Perry C, Stover PJ: **A ferritin-responsive internal ribosome entry site regulates folate metabolism**. *J Biol Chem* 2007, **282**(41):29927-29935.

138. Zheng Y, Miskimins WK: **CUG-binding protein represses translation of p27Kip1 mRNA through its internal ribosomal entry site**. *RNA Biol* 2011, **8**(3):365-371.

139. Roberts R, Timchenko NA, Miller JW, Reddy S, Caskey CT, Swanson MS, Timchenko LT: **Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic**

dystrophy and in myotonin protein kinase knockout mice. *Proc Natl Acad Sci U S A* 1997, **94**(24):13221-13226.

140. Kuyumcu-Martinez NM, Wang GS, Cooper TA: **Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation**. *Mol Cell* 2007, **28**(1):68-78.

141. Timchenko NA, Wang GL, Timchenko LT: **RNA CUG-binding protein 1 increases translation of 20-kDa isoform of CCAAT/enhancer-binding protein beta by interacting with the alpha and beta subunits of eukaryotic initiation translation factor 2**. *J Biol Chem* 2005, **280**(21):20549-20557.

142. Kuyumcu-Martinez NM, Wang GS, Cooper TA: **Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation**. *Mol Cell* 2007, **28**(1):68-78.

143. Orengo JP, Chambon P, Metzger D, Mosier DR, Snipes GJ, Cooper TA: **Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy**. *Proc Natl Acad Sci U S A* 2008, **105**(7):2646-2651.

144. Zhang L, Lee JE, Wilusz J, Wilusz CJ: **The RNA-binding protein CUGBP1 regulates stability of tumor necrosis factor mRNA in muscle cells: implications for myotonic dystrophy**. *J Biol Chem* 2008, **283**(33):22457-22463.

145. Wang GS, Kuyumcu-Martinez MN, Sarma S, Mathur N, Wehrens XH, Cooper TA: **PKC inhibition ameliorates the cardiac phenotype in a mouse model of myotonic dystrophy type 1**. *J Clin Invest* 2009, **119**(12):3797-3806.

146. Schoser B, Timchenko L: **Myotonic dystrophies 1 and 2: complex diseases with complex mechanisms**. *Curr Genomics* 2010, **11**(2):77-90.

147. Khabar KS: **Post-transcriptional control during chronic inflammation and cancer: a focus on AU-rich elements**. *Cellular and molecular life sciences : CMLS* 2010, **67**(17):2937-2955.

148. Venables JP: **Aberrant and alternative splicing in cancer**. *Cancer research* 2004, **64**(21):7647-7654.

149. Mayr C, Bartel DP: **Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells**. *Cell* 2009, **138**(4):673-684.

150. Schatz JH, Oricchio E, Wolfe AL, Jiang M, Linkov I, Maragulia J, Shi W, Zhang Z, Rajasekhar VK, Pagano NC *et al*: **Targeting cap-dependent translation blocks converging survival signals by AKT and PIM kinases in lymphoma**. *J Exp Med* 2011, **208**(9):1799-1807.

151. Vazquez-Chantada M, Fernandez-Ramos D, Embade N, Martinez-Lopez N, Varela-Rey M, Woodhoo A, Luka Z, Wagner C, Anglim PP, Finnell RH *et al*: **HuR/methyl-HuR and AUF1 regulate the MAT expressed during liver proliferation, differentiation, and carcinogenesis**. *Gastroenterology* 2010, **138**(5):1943-1953.

152. Levy NS, Chung S, Furneaux H, Levy AP: **Hypoxic stabilization of vascular endothelial growth factor mRNA by the RNA-binding protein HuR**. *J Biol Chem* 1998, **273**(11):6417-6423.

153. Bergers G, Benjamin LE: **Tumorigenesis and the angiogenic switch**. *Nature reviews Cancer* 2003, **3**(6):401-410.

154. Eberhart CE, Coffey RJ, Radhika A, Giardiello FM, Ferrenbach S, DuBois RN: **Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas**. *Gastroenterology* 1994, **107**(4):1183-1188.

155. Denkert C, Koch I, von Keyserlingk N, Noske A, Niesporek S, Dietel M, Weichert W: **Expression of the ELAV-like protein HuR in human colon cancer: association with tumor stage and cyclooxygenase-2**. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2006, **19**(9):1261-1269.

156. Denkert C, Weichert W, Winzer KJ, Muller BM, Noske A, Niesporek S, Kristiansen G, Guski H, Dietel M, Hauptmann S: **Expression of the ELAV-like protein HuR is associated with higher tumor grade and increased cyclooxygenase-2 expression in human breast carcinoma**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004, **10**(16):5580-5586.

157. Brennan SE, Kuwano Y, Alkharouf N, Blackshear PJ, Gorospe M, Wilson GM: **The mRNA-destabilizing protein tristetraprolin is suppressed in many cancers, altering tumorigenic phenotypes and patient prognosis**. *Cancer research* 2009, **69**(12):5168-5176.

158. Bais C, Santomasso B, Coso O, Arvanitakis L, Raaka EG, Gutkind JS, Asch AS, Cesarman E, Gershengorn MC, Mesri EA: **G-protein-coupled receptor of Kaposi's sarcoma-associated herpesvirus is a viral oncogene and angiogenesis activator**. *Nature* 1998, **391**(6662):86-89.

159. Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, Parry JP, Peruzzi D, Edelman IS, Chang Y *et al*: **Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8)**. *Proc Natl Acad Sci U S A* 1996, **93**(25):14862-14867.

160. Cannon ML, Cesarman E: **The KSHV G protein-coupled receptor signals via multiple pathways to induce transcription factor activation in primary effusion lymphoma cells**. *Oncogene* 2004, **23**(2):514-523.

161. Cannon M, Philpott NJ, Cesarman E: **The Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor has broad signaling effects in primary effusion lymphoma cells**. *J Virol* 2003, **77**(1):57-67.

162. McCormick C, Ganem D: **The kaposin B protein of KSHV activates the p38/MK2 pathway and stabilizes cytokine mRNAs**. *Science* 2005, **307**(5710):739-741.

163. Ronkina N, Menon MB, Schwermann J, Tiedje C, Hitti E, Kotlyarov A, Gaestel M: **MAPKAP kinases MK2 and MK3 in inflammation: complex regulation of TNF biosynthesis via expression and phosphorylation of tristetraprolin**. *Biochemical pharmacology* 2010, **80**(12):1915-1920.

164. Yoo J, Kang J, Lee HN, Aguilar B, Kafka D, Lee S, Choi I, Lee J, Ramu S, Haas J *et al*: **Kaposin-B enhances the PROX1 mRNA stability during lymphatic reprogramming of vascular endothelial cells by Kaposi's sarcoma herpes virus**. *PLoS pathogens* 2010, **6**(8):e1001046.

165. Wigle JT, Oliver G: **Prox1 function is required for the development of the murine lymphatic system**. *Cell* 1999, **98**(6):769-778.

166. Wigle JT, Harvey N, Detmar M, Lagutina I, Grosveld G, Gunn MD, Jackson DG, Oliver G: **An essential role for Prox1 in the induction of the lymphatic endothelial cell phenotype**. *Embo J* 2002, **21**(7):1505-1513.

167. Hong YK, Foreman K, Shin JW, Hirakawa S, Curry CL, Sage DR, Libermann T, Dezube BJ, Fingeroth JD, Detmar M: **Lymphatic reprogramming of blood vascular endothelium by Kaposi sarcoma-associated herpesvirus**. *Nature genetics* 2004, **36**(7):683-685.

168. Di Giammartino DC, Nishida K, Manley JL: **Mechanisms and consequences of alternative polyadenylation**. *Mol Cell* 2011, **43**(6):853-866.

169. Tian B, Graber JH: **Signals for pre-mRNA cleavage and polyadenylation**. *Wiley interdisciplinary reviews RNA* 2012, **3**(3):385-396.

170. Beyer K, Dandekar T, Keller W: **RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA**. *J Biol Chem* 1997, **272**(42):26769-26779.

171. Takagaki Y, Seipelt RL, Peterson ML, Manley JL: **The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation**. *Cell* 1996, **87**(5):941-952.

172. Veraldi KL, Arhin GK, Martincic K, Chung-Ganster LH, Wilusz J, Milcarek C: **hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells**. *Mol Cell Biol* 2001, **21**(4):1228-1238.

173. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB: **Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites**. *Science* 2008, **320**(5883):1643-1647.

174. Ji Z, Lee JY, Pan Z, Jiang B, Tian B: **Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development**. *Proc Natl Acad Sci U S A* 2009, **106**(17):7028-7033.

175. Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T: **A quantitative atlas of polyadenylation in five mammals**. *Genome research* 2012.

176. Tenenbaum SA, Lager PJ, Carson CC, Keene JD: **Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays**. *Methods* 2002, **26**(2):191-198.

177. Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB *et al*: **Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1**. *Mol Cell* 2008, **29**(2):263-270.

178. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator**. *Genome Res* 2004, **14**(6):1188-1190.

179. Ogilvie RL, Sternjohn JR, Rattenbacher B, Vlasova IA, Williams DA, Hau HH, Blackshear PJ, Bohjanen PR: **Tristetraprolin mediates interferon-gamma mRNA decay**. *J Biol Chem* 2009, **284**(17):11216-11223.

180. Ogilvie RL, Abelson M, Hau HH, Vlasova I, Blackshear PJ, Bohjanen PR: **Tristetraprolin down-regulates IL-2 gene expression through AU-rich element-mediated mRNA decay**. *J Immunol* 2005, **174**(2):953-961.

181. Peng SS, Chen CY, Shyu AB: **Functional characterization of a non-AUUUA AU-rich element from the c-jun proto-oncogene mRNA: evidence for a novel class of AU-rich elements**. *Mol Cell Biol* 1996, **16**(4):1490-1499.

182. Sauvonnet N, Pradet-Balade B, Garcia-Sanz JA, Cornelis GR: **Regulation of mRNA expression in macrophages after Yersinia enterocolitica infection. Role of different Yop effectors**. *J Biol Chem* 2002, **277**(28):25133-25142.

183. Ogilvie RL, Abelson M, Hau HH, Vlasova I, Blackshear PJ, Bohjanen PR: **Tristetraprolin down-regulates IL-2 gene expression through AU-rich element-mediated mRNA decay**. *J Immunol* 2005, **174**(2):953-961.

184. Rattenbacher B, Beisang D, Wiesner DL, Jeschke JC, von Hohenberg M, St Louis-Vlasova IA, Bohjanen PR: **Analysis of CUGBP1 targets identifies GU-repeat sequences that mediate rapid mRNA decay**. *Mol Cell Biol* 2010, **30**(16):3970-3980.

185. David M, Dzamba M, Lister D, Ilie L, Brudno M: **SHRiMP2: sensitive yet practical SHort Read Mapping**. *Bioinformatics* 2011, **27**(7):1011-1012.

186. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: **The UCSC Table Browser data retrieval tool**. *Nucleic Acids Res* 2004, **32**(Database issue):D493-496.

187. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

188. Vlasova IA, Bohjanen PR: **Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins**. *RNA Biol* 2008, **5**(4):201-207.

189. Huichalaf CH, Sakai K, Wang GL, Timchenko NA, Timchenko L: **Regulation of the promoter of CUG triplet repeat binding protein, Cugbp1, during myogenesis**. *Gene* 2007, **396**(2):391-402.

190. Jin J, Wang GL, Salisbury E, Timchenko L, Timchenko NA: **GSK3beta-cyclin D3-CUGBP1-eIF2 pathway in aging and in myotonic dystrophy**. *Cell Cycle* 2009, **8**(15):2356-2359.

191. Kress C, Gautier-Courteille C, Osborne HB, Babinet C, Paillard L: **Inactivation of CUG-BP1/CELF1 causes growth, viability, and spermatogenesis defects in mice**. *Mol Cell Biol* 2007, **27**(3):1146-1157.

192. Graindorge A, Le Tonqueze O, Thuret R, Pollet N, Osborne HB, Audic Y: **Identification of CUG-BP1/EDEN-BP target mRNAs in Xenopus tropicalis**. *Nucleic Acids Res* 2008, **36**(6):1861-1870.

193. Tenenbaum SA, Carson CC, Lager PJ, Keene JD: **Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays**. *Proc Natl Acad Sci U S A* 2000, **97**(26):14085-14090.

194. Mazan-Mamczarz K, Kuwano Y, Zhan M, White EJ, Martindale JL, Lal A, Gorospe M: **Identification of a signature motif in target mRNAs of RNA-binding protein AUF1**. *Nucleic Acids Res* 2009, **37**(1):204-214.

195. Kim HS, Kuwano Y, Zhan M, Pullmann R, Jr., Mazan-Mamczarz K, Li H, Kedersha N, Anderson P, Wilce MC, Gorospe M *et al*: **Elucidation of a C-rich signature motif in target mRNAs of RNA-binding protein TIAR**. *Mol Cell Biol* 2007, **27**(19):6806-6817.

196. Stoecklin G, Tenenbaum SA, Mayo T, Chittur SV, George AD, Baroni TE, Blackshear PJ, Anderson P: **Genome-wide analysis identifies interleukin-10 mRNA as target of tristetraprolin**. *J Biol Chem* 2008, **283**(17):11689-11699.

197. Morris AR, Mukherjee N, Keene JD: **Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets**. *Mol Cell Biol* 2008, **28**(12):4093-4103.

198. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.

199. Zhang L, Lee JE, Wilusz J, Wilusz CJ: **The RNA-binding protein CUGBP1 regulates stability of tumor necrosis factor mRNA in muscle cells: implications for Myotonic Dystrophy**. *J Biol Chem* 2008.

200. Dustin ML: **The cellular context of T cell signaling**. *Immunity* 2009, **30**(4):482-492.

201. Crispin JC, Tsokos GC: **Transcriptional regulation of IL-2 in health and autoimmunity**. *Autoimmun Rev* 2009, **8**(3):190-195.

202. Merup M: **Genetic abnormalities in non-Hodgkin's lymphomas and chronic lymphocytic leukaemia**. *Med Oncol* 1998, **15**(2):79-88.

203. Pearson PL, Van der Luijt RB: **The genetic analysis of cancer**. *J Intern Med* 1998, **243**(6):413-417.

204. Sanchez-Lockhart M, Marin E, Graf B, Abe R, Harada Y, Sedwick CE, Miller J: **Cutting edge: CD28-mediated transcriptional and posttranscriptional regulation of IL-2 expression are controlled through different signaling pathways**. *J Immunol* 2004, **173**(12):7120-7124.

205. Stellato C, Gubin MM, Magee JD, Fang X, Fan J, Tartar DM, Chen J, Dahm GM, Calaluce R, Mori F *et al*: **Coordinate Regulation of GATA-3 and Th2 Cytokine Gene Expression by the RNA-Binding Protein HuR**. *J Immunol* 2011, **187**(1):441-449.

206. Ogilvie RL, Sternjohn JR, Rattenbacher B, Vlasova IA, Williams DA, Hau HH, Blackshear PJ, Bohjanen PR: **Tristetraprolin mediates interferon-gamma mRNA decay**. *J Biol Chem* 2009, **284**(17):11216-11223.

207. Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA: **A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart**. *Proc Natl Acad Sci U S A* 2008, **105**(51):20333-20338.

208. Timchenko LT, Salisbury E, Wang GL, Nguyen H, Albrecht JH, Hershey JW, Timchenko NA: **Age-specific CUGBP1-eIF2 complex increases translation of CCAAT/enhancer-binding protein beta in old liver**. *J Biol Chem* 2006, **281**(43):32806-32819.

209. Tenenbaum SA, Carson CC, Lager PJ, Keene JD: **Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays**. *Proc Natl Acad Sci U S A* 2000, **97**(26):14085-14090.

210. Liu X, Brutlag DL, Liu JS: **BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes**. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing* 2001:127-138.

211. Ramskold D, Wang ET, Burge CB, Sandberg R: **An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data**. *PLoS Comput Biol* 2009, **5**(12):e1000598.

212. Vlasova-St Louis I, Bohjanen PR: **Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1**. *Curr Opin Genet Dev* 2011, **21**(4):444-451.

213. Koshelev M, Sarma S, Price RE, Wehrens XH, Cooper TA: **Heart-specific overexpression of CUGBP1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1**. *Hum Mol Genet* 2010, **19**(6):1066-1075.

214. Iakova P, Wang GL, Timchenko L, Michalak M, Pereira-Smith OM, Smith JR, Timchenko NA: **Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate**. *Embo J* 2004, **23**(2):406-417.

215. Al-Ahmadi W, Al-Ghamdi M, Al-Haj L, Al-Saif M, Khabar KS: **Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-Regulation**. *Nucleic Acids Res* 2009, **37**(11):3612-3624.

216. Mukherjee N, Lager PJ, Friedersdorf MB, Thompson MA, Keene JD: **Coordinated posttranscriptional mRNA population dynamics during T-cell activation**. *Mol Syst Biol* 2009, **5**:288.

217. Whistler T, Chiang CF, Lonergan W, Hollier M, Unger ER: **Implementation of exon arrays: alternative splicing during T-cell proliferation as determined by whole genome analysis**. *BMC Genomics* 2010, **11**:496.

218. Choi WT, Folsom MR, Azim MF, Meyer C, Kowarz E, Marschalek R, Timchenko NA, Naeem RC, Lee DA: **C/EBPbeta suppression by interruption of CUGBP1 resulting from a complex rearrangement of MLL**. *Cancer Genet Cytogenet* 2007, **177**(2):108-114.

219. Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ *et al*: **A transposon-based genetic screen in mice identifies genes altered in colorectal cancer**. *Science* 2009, **323**(5922):1747-1750.

220. Beisang D, Rattenbacher B, Vlasova-St Louis IA, Bohjanen PR: **Regulation of CUG-binding protein 1 (CUGBP1) binding to target transcripts upon T cell activation**. *J Biol Chem* 2012, **287**(2):950-960.

221. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.

222. Stoecklin G, Gross B, Ming XF, Moroni C: **A novel mechanism of tumor suppression by destabilizing AU-rich growth factor mRNA**. *Oncogene* 2003, **22**(23):3554-3561.

223. Bernasconi NL, Wormhoudt TA, Laird-Offringa IA: **Post-transcriptional deregulation of myc genes in lung cancer cell lines**. *American journal of respiratory cell and molecular biology* 2000, **23**(4):560-565.

224. Gu L, Zhang H, He J, Li J, Huang M, Zhou M: **MDM2 regulates MYCN mRNA stabilization and translation in human neuroblastoma cells**. *Oncogene* 2012, **31**(11):1342-1353.

225. Zhou S, Gu L, He J, Zhang H, Zhou M: **MDM2 regulates vascular endothelial growth factor mRNA stabilization in hypoxia**. *Mol Cell Biol* 2011, **31**(24):4928-4937.

226. Chagnovich D, Cohn SL: **Activity of a 40 kDa RNA-binding protein correlates with MYCN and c-fos mRNA stability in human neuroblastoma**. *Eur J Cancer* 1997, **33**(12):2064-2067.

227. Denkert C, Weichert W, Pest S, Koch I, Licht D, Kobel M, Reles A, Sehouli J, Dietel M, Hauptmann S: **Overexpression of the embryonic-lethal abnormal vision-like protein HuR in ovarian carcinoma is a prognostic factor and is associated with increased cyclooxygenase 2 expression**. *Cancer research* 2004, **64**(1):189-195.

228. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS: **Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma**. *Science* 1994, **266**(5192):1865-1869.

229. Montaner S, Sodhi A, Molinolo A, Bugge TH, Sawai ET, He Y, Li Y, Ray PE, Gutkind JS: **Endothelial infection with KSHV genes in vivo reveals that vGPCR initiates Kaposi's sarcomagenesis and can promote the tumorigenic potential of viral latent genes**. *Cancer cell* 2003, **3**(1):23-36.

230. Yang X, Wang W, Fan J, Lal A, Yang D, Cheng H, Gorospe M: **Prostaglandin A2-mediated stabilization of p21 mRNA through an ERK-dependent pathway requiring the RNA-binding protein HuR**. *J Biol Chem* 2004, **279**(47):49298-49306.

231. Deleault KM, Skinner SJ, Brooks SA: **Tristetraprolin regulates TNF TNF-alpha mRNA stability via a proteasome dependent mechanism involving the combined action of the ERK and p38 pathways**. *Molecular immunology* 2008, **45**(1):13-24.

232. Scholzova E, Malik R, Sevcik J, Kleibl Z: **RNA regulation and cancer development**. *Cancer letters* 2007, **246**(1-2):12-23.

233. Lebedeva S, Jens M, Theil K, Schwanhausser B, Selbach M, Landthaler M, Rajewsky N: **Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR**. *Mol Cell* 2011, **43**(3):340-352.

234. Kaech SM, Wherry EJ, Ahmed R: **Effector and memory T-cell differentiation: implications for vaccine development**. *Nature reviews Immunology* 2002, **2**(4):251-262.

235. Murthy KG, Manley JL: **The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation**. *Genes Dev* 1995, **9**(21):2672-2683.

236. Takagaki Y, Manley JL: **RNA recognition by the human polyadenylation factor CstF**. *Mol Cell Biol* 1997, **17**(7):3907-3914.

237. Monarez RR, MacDonald CC, Dass B: **Polyadenylation proteins CstF-64 and tauCstF-64 exhibit differential binding affinities for RNA polymers**. *The Biochemical journal* 2007, **401**(3):651-658.

238. Hu J, Lutz CS, Wilusz J, Tian B: **Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation**. *Rna* 2005, **11**(10):1485-1493.

239. Takagaki Y, Manley JL: **Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation**. *Mol Cell* 1998, **2**(6):761-771.

240. Shell SA, Hesse C, Morris SM, Jr., Milcarek C: **Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection**. *J Biol Chem* 2005, **280**(48):39950-39961.

241. Khabar KS: **The AU-rich transcriptome: more than interferons and cytokines, and its role in disease**. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* 2005, **25**(1):1-10.

242. Bakheet T, Williams BR, Khabar KS: **ARED 3.0: the large and diverse AU-rich transcriptome**. *Nucleic Acids Res* 2006, **34**(Database issue):D111-114.

243. Barreau C, Paillard L, Osborne HB: **AU-rich elements and associated factors: are there unifying principles?** *Nucleic Acids Res* 2005, **33**(22):7138-7150.

244. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM: **Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation**. *Cell* 2010, **143**(6):1018-1029.

245. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nature methods* 2012, **9**(4):357-359.

246. Dai W, Zhang G, Makeyev EV: **RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage**. *Nucleic Acids Res* 2012, **40**(2):787-800.

247. Niibori Y, Hayashi F, Hirai K, Matsui M, Inokuchi K: **Alternative poly(A) site-selection regulates the production of alternatively spliced vesl-1/homer1 isoforms that encode postsynaptic scaffolding proteins**. *Neuroscience research* 2007, **57**(3):399-410.

248. Jost PJ, Ruland J: **Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications**. *Blood* 2007, **109**(7):2700-2707.

249. Okkenhaug K, Vanhaesebroeck B: **PI3K in lymphocyte development, differentiation and activation**. *Nature reviews Immunology* 2003, **3**(4):317-330.

250. Pullmann R, Jr., Kim HH, Abdelmohsen K, Lal A, Martindale JL, Yang X, Gorospe M: **Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs**. *Mol Cell Biol* 2007, **27**(18):6265-6278.

251. Carrick DM, Blackshear PJ: **Comparative expression of tristetraprolin (TTP) family member transcripts in normal human tissues and cancer cell lines**. *Archives of biochemistry and biophysics* 2007, **462**(2):278-285.

252. Erkinheimo TL, Sivula A, Lassus H, Heinonen M, Furneaux H, Haglund C, Butzow R, Ristimaki A: **Cytoplasmic HuR expression correlates with epithelial cancer cell but not with stromal cell cyclooxygenase-2 expression in mucinous ovarian carcinoma**. *Gynecologic oncology* 2005, **99**(1):14-19.

253. Gubin MM, Calaluce R, Davis JW, Magee JD, Strouse CS, Shaw DP, Ma L, Brown A, Hoffman T, Rold TL *et al*: **Overexpression of the RNA binding protein HuR impairs tumor growth in triple negative breast cancer associated with deficient angiogenesis**. *Cell Cycle* 2010, **9**(16):3337-3346.

Appendix A

Program for enumerating all kmers in a given inputs file of fasta format. Output is a list of the counts of each kmer, as well as the top 200 kmers and their usage. Kmer can be set to a length between 4 and 13.

```
#include <iostream>
#include <fstream>
#include <math.h>
#include <vector>
#include <algorithm>
#include <functional>
#include <string.h>
#include <sstream>
#include <iomanip>
#include <ctime>
#include <list>
#include <stdio.h>
#include <stdlib.h>

using namespace std;

const int kmer=10;

struct SEQUENCE
{
string name;
double k;
string sequence;
double init_prob[30000];
double put_prob[30000];
};

int convert_to_ten(int tempkmer[kmer]);
void motif_convert_kmer(char subst_kmer[kmer], int tempkmer[kmer]);
void convert_to_four(int base_ten, int number_string[kmer]);

int main () {

//initialize variables
int i, j, k, q, r, iterations=0, num_iterations;
string filename, out_filename;
ifstream sequence_file;
ofstream results;
string name_buf;
double ratio_buf, RES_score=0;
string fasta_buf;
SEQUENCE buf;
```

```
int max, num_motifs;
int element_count=0;
int lambda;

//seed the random number generator with the time variable.
srand(time(0));

//Initialize a linked list of SEQUENCE structs and pointer to objects in the list
list<SEQUENCE> sequence_list;
list<SEQUENCE>::iterator list_iter;
list_iter = sequence_list.begin();

list<SEQUENCE> temp_list;
list<SEQUENCE>::iterator temp_iter;
temp_iter = temp_list.begin();

//retrive and open input file
cout<<"Please enter the sequence file name."<<endl;
cin>>filename;
        sequence_file.open(filename.c_str());
                if(sequence_file.is_open())
                cout<<"sequence file open"<<endl;
                else
                {
                cout<<"sequence file was unable to open.  CLOSING PROGRAM"<<endl;
                return 0;
                }
//Open the output file
cout<<"Please enter the output filename."<<endl;
cin>>filename;


results.open(filename.c_str());
                if(results.is_open())
                cout<<"output file open"<<endl;
                else
                {
                cout<<"output file was unable to open.  CLOSING PROGRAM"<<endl;
                return 0;
                }

//Read in file and initialize the linked list.
while(!sequence_file.eof())
{
        sequence_file>>name_buf>>fasta_buf;
        if(fasta_buf == buf.sequence)
                continue;
        if(fasta_buf.length() <=kmer)
                continue;
```

```
        buf.name = name_buf;
        buf.sequence = fasta_buf;
        list_iter= sequence_list.end();
        sequence_list.insert(list_iter, buf);
        element_count++;
}

cout<<"File was read in to the linked list"<<endl;

//initialize variables for exhaustive search
int kmer_baseten=0;
int tempkmer[kmer]={0};
char subst_kmer[kmer];
int *kmer_occur=new int[(int)pow(4,kmer)];
int *kmer_used=new int[(int)pow(4,kmer)];

//pre-algorithm exhaustive search for proper initialization positions.
for(list_iter=sequence_list.begin(); (list_iter)!=sequence_list.end(); list_iter++)
{

buf = *list_iter;
//step through the sequence and enumerate all kmers as well as the overlap.
        //re-set the kmer_used array to zero
        for(i=0; i<pow(4,kmer); i++)
        {
        kmer_used[i] = 0;
        }

        for(i=0; i<((buf.sequence).length()-kmer); i++)
        {
        //pull out the substring of length kmer starting at position i in the sequence.
        memset( subst_kmer, '\0',kmer);
        (buf.sequence).copy( subst_kmer, kmer, i);
        //convert each substring into a representative integer array
        motif_convert_kmer(subst_kmer, tempkmer);
        //convert tempkmer to a base ten number
        kmer_baseten = convert_to_ten(tempkmer);
                if(kmer_baseten != -1)
                {
                kmer_occur[kmer_baseten]++;
                kmer_used[kmer_baseten] = 0;
                }
        }
}

for(i=0; i<pow(4,kmer); i++)
{
results<<kmer_occur[i]<<endl;
}
```

```
//retreive the top !!!200!!! kmers from the list of kmer_occur
int top_kmer[200]= {0};
int top_kmer_count[200] = {0};
int kmer_count=0;

for (i=0; i<200; i++)
{
kmer_baseten =0;
kmer_count = 0;
        for (j=0; j<pow(4,kmer); j++)
        {
                if(kmer_occur[j] > kmer_count)
                {
                kmer_baseten = j;
                kmer_count = kmer_occur[j];
                }
        }
kmer_occur[kmer_baseten] = 0;
top_kmer[i] = kmer_baseten;
top_kmer_count[i] = kmer_count;
}

//print out the top 200 kmers and their number of occurences

for (i=0; i<200;i++)
{
convert_to_four(top_kmer[i], tempkmer);
        for(j=0; j<kmer; j++)
        {
        cout<<tempkmer[j]<<" ";
        }
        cout<<top_kmer_count[i]<<endl;
}

return 0;
}

//*************************************************************************
*
//converts a character array into an array of integers.
//*************************************************************************
*
void motif_convert_kmer(char subst_kmer[], int tempkmer[])
{
int i=0;

for(i=0; i < kmer; i++)
{
```

```
        if(subst_kmer[i] == 'a'|| subst_kmer[i] == 'A')
                tempkmer[i]=0;
        if(subst_kmer[i] == 'c'|| subst_kmer[i] == 'C')
                tempkmer[i]=1;
        if(subst_kmer[i] == 't'|| subst_kmer[i] == 'T')
                tempkmer[i]=2;
        if(subst_kmer[i] == 'g'|| subst_kmer[i] == 'G')
                tempkmer[i]=3;
        if(subst_kmer[i] == 'n' || subst_kmer[i] == 'N')
                tempkmer[i] = -1;
}

return;
}

//*************************************************************************************
**********************
//this subroutine converts a 4Xkmer array representation of a sequence into a base ten sequence.
//*************************************************************************************
**********************
int convert_to_ten(int tempkmer[kmer])
{
int baseten=0;
int i=0;

for(i=0; i<kmer; i++)
{
        if(tempkmer[i] == -1)
                return -1;
        else
                baseten = baseten +pow(4,i)*tempkmer[kmer-i-1];

}

return baseten;
}

//*************************************************************************************
*****
//calculate base four array out of the baseten number
//*************************************************************************************
*****
void convert_to_four(int base_ten, int number_string[kmer])
{
int one=0, two =0, three=0, four=0, five=0, six=0, seven=0, eight=0, nine=0, ten=0, eleven=0,
twelve=0, thirteen=0;
int index;

switch(kmer)
```

```
{
case 4:
        one = base_ten / pow(4,3);
        two = (base_ten - one*pow(4,3))/pow(4,2);
        three = (base_ten - one*pow(4,3) - two*pow(4,2))/pow(4,1);
        four = (base_ten - one*pow(4,3) - two*pow(4,2) - three*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        break;
case 5:
        one = base_ten / pow(4,4);
        two = (base_ten - one*pow(4,4))/pow(4,3);
        three = (base_ten - one*pow(4,4) - two*pow(4,3))/pow(4,2);
        four = (base_ten - one*pow(4,4) - two*pow(4,3) - three*pow(4,2))/pow(4,1);
        five = (base_ten - one*pow(4,4) - two*pow(4,3) - three*pow(4,2)-four*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        break;
case 6:
        one = base_ten / pow(4,5);
        two = (base_ten - one*pow(4,5))/pow(4,4);
        three = (base_ten - one*pow(4,5) - two*pow(4,4))/pow(4,3);
        four = (base_ten - one*pow(4,5) - two*pow(4,4) - three*pow(4,3))/pow(4,2);
        five = (base_ten - one*pow(4,5) - two*pow(4,4) - three*pow(4,3)-
four*pow(4,2))/pow(4,1);
        six = (base_ten - one*pow(4,5) - two*pow(4,4) - three*pow(4,3)-four*pow(4,2)-
five*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        break;
case 7:
        one = base_ten / pow(4,6);
        two = (base_ten - one*pow(4,6))/pow(4,5);
        three = (base_ten - one*pow(4,6) - two*pow(4,5))/pow(4,4);
        four = (base_ten - one*pow(4,6) - two*pow(4,5) - three*pow(4,4))/pow(4,3);
        five = (base_ten - one*pow(4,6) - two*pow(4,5) - three*pow(4,4)-
four*pow(4,3))/pow(4,2);
        six = (base_ten - one*pow(4,6) - two*pow(4,5) - three*pow(4,4)-four*pow(4,3)-
five*pow(4,2))/pow(4,1);
```

```
        seven = (base_ten - one*pow(4,6) - two*pow(4,5) - three*pow(4,4)-four*pow(4,3)-
five*pow(4,2)-six*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        break;
case 8:
        one = base_ten / pow(4,7);
        two = (base_ten - one*pow(4,7))/pow(4,6);
        three = (base_ten - one*pow(4,7) - two*pow(4,6))/pow(4,5);
        four = (base_ten - one*pow(4,7) - two*pow(4,6) - three*pow(4,5))/pow(4,4);
        five = (base_ten - one*pow(4,7) - two*pow(4,6) - three*pow(4,5)-
four*pow(4,4))/pow(4,3);
        six = (base_ten - one*pow(4,7) - two*pow(4,6) - three*pow(4,5)-four*pow(4,4)-
five*pow(4,3))/pow(4,2);
        seven = (base_ten - one*pow(4,7) - two*pow(4,6) - three*pow(4,5)-four*pow(4,4)-
five*pow(4,3)-six*pow(4,2))/pow(4,1);
        eight = (base_ten - one*pow(4,7) - two*pow(4,6) - three*pow(4,5)-four*pow(4,4)-
five*pow(4,3)-six*pow(4,2)-seven*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        number_string[7] = eight;
        break;
case 9:
        one = base_ten / pow(4,8);
        two = (base_ten - one*pow(4,8))/pow(4,7);
        three = (base_ten - one*pow(4,8) - two*pow(4,7))/pow(4,6);
        four = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6))/pow(4,5);
        five = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6)-
four*pow(4,5))/pow(4,4);
        six = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6)-four*pow(4,5)-
five*pow(4,4))/pow(4,3);
        seven = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6)-four*pow(4,5)-
five*pow(4,4)-six*pow(4,3))/pow(4,2);
        eight = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6)-four*pow(4,5)-
five*pow(4,4)-six*pow(4,3)-seven*pow(4,2))/pow(4,1);
        nine = (base_ten - one*pow(4,8) - two*pow(4,7) - three*pow(4,6)-four*pow(4,5)-
five*pow(4,4)-six*pow(4,3)-seven*pow(4,2)-eight*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
```

```
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        number_string[7] = eight;
        number_string[8] = nine;
        break;
case 10:
        one = base_ten / pow(4,9);
        two = (base_ten - one*pow(4,9))/pow(4,8);
        three = (base_ten - one*pow(4,9) - two*pow(4,8))/pow(4,7);
        four = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7))/pow(4,6);
        five = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-
four*pow(4,6))/pow(4,5);
        six = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-four*pow(4,6)-
five*pow(4,5))/pow(4,4);
        seven = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-four*pow(4,6)-
five*pow(4,5)-six*pow(4,4))/pow(4,3);
        eight = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-four*pow(4,6)-
five*pow(4,5)-six*pow(4,4)-seven*pow(4,3))/pow(4,2);
        nine = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-four*pow(4,6)-
five*pow(4,5)-six*pow(4,4)-seven*pow(4,3)-eight*pow(4,2))/pow(4,1);
        ten = (base_ten - one*pow(4,9) - two*pow(4,8) - three*pow(4,7)-four*pow(4,6)-
five*pow(4,5)-six*pow(4,4)-seven*pow(4,3)-eight*pow(4,2)-nine*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        number_string[7] = eight;
        number_string[8] = nine;
        number_string[9] = ten;
        break;
case 11:
        one = base_ten / pow(4,10);
        two = (base_ten - one*pow(4,10))/pow(4,9);
        three = (base_ten - one*pow(4,10) - two*pow(4,9))/pow(4,8);
        four = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8))/pow(4,7);
        five = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-
four*pow(4,7))/pow(4,6);
        six = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6))/pow(4,5);
        seven = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6)-six*pow(4,5))/pow(4,4);
        eight = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6)-six*pow(4,5)-seven*pow(4,4))/pow(4,3);
```

```
        nine = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6)-six*pow(4,5)-seven*pow(4,4)-eight*pow(4,3))/pow(4,2);
        ten = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6)-six*pow(4,5)-seven*pow(4,4)-eight*pow(4,3)-nine*pow(4,2))/pow(4,1);
        eleven = (base_ten - one*pow(4,10) - two*pow(4,9) - three*pow(4,8)-four*pow(4,7)-
five*pow(4,6)-six*pow(4,5)-seven*pow(4,4)-eight*pow(4,3)-nine*pow(4,2)-ten*pow(4,1));
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        number_string[7] = eight;
        number_string[8] = nine;
        number_string[9] = ten;
        number_string[10] = eleven;
        break;
case 12:
        one = base_ten / pow(4,11);
        two = (base_ten - one*pow(4,11))/pow(4,10);
        three = (base_ten - one*pow(4,11) - two*pow(4,10))/pow(4,9);
        four = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9))/pow(4,8);
        five = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-
four*pow(4,8))/pow(4,7);
        six = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7))/pow(4,6);
        seven = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6))/pow(4,5);
        eight = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6)-seven*pow(4,5))/pow(4,4);
        nine = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6)-seven*pow(4,5)-eight*pow(4,4))/pow(4,3);
        ten = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6)-seven*pow(4,5)-eight*pow(4,4)-nine*pow(4,3))/pow(4,2);
        eleven = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6)-seven*pow(4,5)-eight*pow(4,4)-nine*pow(4,3)-
ten*pow(4,2))/pow(4,1);
        twelve = (base_ten - one*pow(4,11) - two*pow(4,10) - three*pow(4,9)-four*pow(4,8)-
five*pow(4,7)-six*pow(4,6)-seven*pow(4,5)-eight*pow(4,4)-nine*pow(4,3)-ten*pow(4,2))-
eleven*pow(4,1);
        number_string[0] = one;
        number_string[1] = two;
        number_string[2] = three;
        number_string[3] = four;
        number_string[4] = five;
        number_string[5] = six;
        number_string[6] = seven;
        number_string[7] = eight;
```

```
            number_string[8] = nine;
            number_string[9] = ten;
            number_string[10] = eleven;
            number_string[11] = twelve;
            break;
case 13:
            one = base_ten / pow(4,12);
            two = (base_ten - one*pow(4,12))/pow(4,11);
            three = (base_ten - one*pow(4,12) - two*pow(4,11))/pow(4,10);
            four = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10))/pow(4,9);
            five = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-
four*pow(4,9))/pow(4,8);
            six = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8))/pow(4,7);
            seven = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7))/pow(4,6);
            eight = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6))/pow(4,5);
            nine = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6)-eight*pow(4,5))/pow(4,4);
            ten = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6)-eight*pow(4,5)-nine*pow(4,4))/pow(4,3);
            eleven = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6)-eight*pow(4,5)-nine*pow(4,4)-
ten*pow(4,3))/pow(4,2);
            twelve = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6)-eight*pow(4,5)-nine*pow(4,4)-ten*pow(4,3)-
eleven*pow(4,2))/pow(4,1);
            thirteen = (base_ten - one*pow(4,12) - two*pow(4,11) - three*pow(4,10)-four*pow(4,9)-
five*pow(4,8)-six*pow(4,7)-seven*pow(4,6)-eight*pow(4,5)-nine*pow(4,4)-ten*pow(4,3)-
eleven*pow(4,2)-twelve*pow(4,1));
            number_string[0] = one;
            number_string[1] = two;
            number_string[2] = three;
            number_string[3] = four;
            number_string[4] = five;
            number_string[5] = six;
            number_string[6] = seven;
            number_string[7] = eight;
            number_string[8] = nine;
            number_string[9] = ten;
            number_string[10] = eleven;
            number_string[11] = twelve;
            number_string[12] = thirteen;
            break;
}
return;
}
```

Program for determining the locations of any instances of the consensus GRE (or any motif with small alteration), from an input file of sequences in fasta format.

```
#include <iostream>
#include <fstream>
#include <math.h>
#include <vector>
#include <algorithm>
#include <functional>
#include <string.h>
#include <sstream>
#include <iomanip>

using namespace std;

int main () {
 ifstream sequence_file;
 ofstream gre_list;
 string sequence;
 string sequence_name;
 string filename;
 int score = 1;
 int x=2;
 int y=0;
 int z=0;
 int q=0;
 int length=0;
 int length1=0;
 int mis;
 string inputone;
 string inputtwo;


 cout<<"Please enter the length of the target sequence"<<endl;
 cin>>length;

 char buf[length];
 char greone[length];
 char gretwo[length];

 memset( buf, '\0', length);
 //cout<<"Please enter the target sequence"<<endl;
 inputone = "TGTGTGTGTGT";
 inputtwo = "TGTTTGTTTGT";

filename="/project/boulware/SHRiMP_2_2_2/24hr_long_overlap_sequences_new_formatted_G
Urep_1mm.txt";
 for(y=0; y<length; y++)
```

```
 greone[y]=inputone[y];
 y=0;

 for(y=0; y<length; y++)
 gretwo[y]=inputtwo[y];
 y=0;

sequence_file.open("/project/boulware/SHRiMP_2_2_2/24hr_long_overlap_sequences_new_for
matted.interval");
if(sequence_file.is_open())
cout<<"3UTR open"<<endl;
else
cout<<"3UTR didn't open"<<endl;

gre_list.open(filename.c_str());

while(!sequence_file.eof())
{
        y=0;
        if((x%2) == 0)
        {
        sequence_file>>sequence_name;
        x++;
        continue;
        }
        else
        {
        sequence_file>>sequence;
        length1=sequence.size();
        if(length1<length)
        {
        x++;
        continue;
        }
                while(y < length1-length)
                {
                z=0;
                memset( buf, '\0',length);
                sequence.copy( buf, length, y);

                mis=0;
                //compare gre sequence to the buf sequence. and get a z score
                for(q=0; q<length; q++)
                {
                if((buf[q] != greone[q]) && (buf[q] != gretwo[q]))
                {
                mis = q+1;
                z++;
                }
```

```
                }
                //print out 11mer if it meets criteria

                if(z<=score)
                {
                gre_list<<sequence_name<<" ";
                for(q=0; q<length; q++)
                        {
                        gre_list<<buf[q];
                        }
                gre_list<<" "<<mis<<" "<<length<<" "<<y<<" "<<length1<<" "<<length1-
y<<endl;
                }

                y++;
                }
        x++;
        }
        cout<<"+";

}
cout<<x<<endl;

return 0;
}
```

Program for determining the locations of polyA sites from Direct RNA sequencing reads which have been aligned to a genome, and placed in pileup format for a series of genomic locations using samtools.

```
#include <iostream>
#include <fstream>
#include <string>
#include <iomanip>
#include <sstream>
#include <stack>
#include <cmath>
using namespace std;

int main(){

int counts[500]={0};
ostringstream convertone, converttwo;
int base_i, base_pi, count_i, count_pi, polyA_start, polyA_end, i, max, pos_buf, count_buf;
ifstream master, in;
ofstream out;
string buf1, buf2, buf3, buf4, buf5, buf6, filebuf, one, two, string_buf;
```

```
stack<int> mystack;

master.open("/project/boulware/SHRiMP_2_2_2/human_3UTR_exons");
//master.open("/project/boulware/SHRiMP_2_2_2/ELAVL1_coordinates");
out.open("/project/boulware/SHRiMP_2_2_2/channel6_462012_merged_polyA_sites.txt");

while(!master.eof()){
master>>buf1>>buf2>>buf3>>buf4>>buf5>>buf6;
/*convertone.str("");
converttwo.str("");
convertone<<buf2;
converttwo<<buf3;
one=convertone.str();
two=converttwo.str();
*/
        filebuf="/project/boulware/SHRiMP_2_2_2/channel6_462012_merged_files/"+buf1+":"
+buf2+"-"+buf3+".txt";
        in.open(filebuf.c_str());

        if(!in.is_open())
        {cout<<"unable to open input file"<<endl;}

        in.seekg(0, ios::end);
        if(in.tellg()==0){
        in.close();
        continue;
        }
        in.close();
        in.open(filebuf.c_str());
        out<<buf1<<":"<<buf2<<"-"<<buf3<<" "<<buf6<<" ";
                //DO POLYA SITE DETECTION ALGORITHM!!!!
        if(buf6=="+")
        {

        while(!mystack.empty()){
        mystack.pop();
        }
        while(!in.eof()){
        in>>pos_buf>>count_buf;
        mystack.push(count_buf);
        mystack.push(pos_buf);
        }

        base_i=mystack.top();
        mystack.pop();
        count_i=mystack.top();
        mystack.pop();
        while(!mystack.empty()){
                if(count_i>10){
```

```
                        base_pi=mystack.top();
                        mystack.pop();
                        count_pi=mystack.top();
                        mystack.pop();

                        polyA_start=base_i;
                        if((base_i-base_pi)>1){
                                out<<base_i<<" "<<count_i<<" ";
                                base_i=base_pi;
                                count_i=count_pi;
                                continue;
                                }
                        for(i=0;i<500;i++)counts[i]=0;
                        i=0;
                        counts[0]=count_pi;
                        while(((base_i-base_pi)==1) && ((double)(count_pi-count_i)/count_i
>0.05)){
                                count_i=count_pi;
                                base_i=base_pi;
                                base_pi=mystack.top();
                                mystack.pop();
                                count_pi=mystack.top();
                                mystack.pop();
                                i++;
                                counts[i]=count_pi;
                                }
                        out<<int((base_i+polyA_start)/2)<<" ";
                        while((base_i-base_pi)==1 && count_pi>10 && !mystack.empty()){
                                count_i=count_pi;
                                base_i=base_pi;
                                base_pi=mystack.top();
                                mystack.pop();
                                count_pi=mystack.top();
                                mystack.pop();
                                i++;
                                if(i>=500){break;}
                                counts[i]=count_pi;
                                }
                        if(i>=500){
                        cout<<buf1<<":"<<buf2<<"-"<<buf3<<" "<<buf6<<" has
indeterminant polyA sites"<<endl;
                        break;
                        }
                        max=counts[0];
                        for(i=1;i<500;i++){
                        if(counts[i]>max) max=counts[i];
                        }
                        out<<max<<" ";
                        }
```

```
                if(!mystack.empty()){
                base_i=mystack.top();
                mystack.pop();
                count_i=mystack.top();
                mystack.pop();
                }
                }
        in.close();
        out<<endl;
        }

        else{
        if(in.eof())
                {
                in.close();
                out<<endl;
                continue;
                }
        in>>base_i>>count_i;
        while(!in.eof()){
                if(count_i>10){
                        in>>base_pi>>count_pi;
                        polyA_start=base_i;
                        if(base_pi-base_i>1){
                                out<<base_i<<" "<<count_i<<" ";
                                base_i=base_pi;
                                count_i=count_pi;
                                continue;
                                }
                        for(i=0;i<500;i++)counts[i]=0;
                        i=0;
                        counts[0]=count_pi;
                        while((base_pi-base_i==1) && ((double)(count_pi-count_i)/count_i
>0.05)){
                                count_i=count_pi;
                                base_i=base_pi;
                                in>>base_pi>>count_pi;
                                i++;
                                counts[i]=count_pi;
                                }
                        out<<int((base_i+polyA_start)/2)<<" ";
                        while((base_pi-base_i)==1 && count_pi>10 && !in.eof()){
                                count_i=count_pi;
                                base_i=base_pi;
                                in>>base_pi>>count_pi;
                                i++;
                                if(i>=500){break;}
                                counts[i]=count_pi;
                                }
```

```
                        if(i>=500){
                        cout<<buf1<<":"<<buf2<<"-"<<buf3<<" "<<buf6<<" has
indeterminant polyA sites"<<endl;
                        break;
                        }

                        max=counts[0];
                        for(i=1;i<500;i++){
                        if(counts[i]>max) max=counts[i];
                        }
                        out<<max<<" ";
                        }
                if(!in.eof()){
                in>>base_i>>count_i;
                }
                }
        in.close();
        out<<endl;
        }
        }
out.close();
master.close();
return 0;
}
```

Program for comparing whether a genomic region got longer or shorter due to alternative polyadenylation by analyzing the output of the previous program.

```
#include <iostream>
#include <fstream>
#include <string>
#include <sstream>
#include <vector>
#include <algorithm>
#include <cmath>

using namespace std;

int main(){
string line, name1, orient, name2, chrom;
ifstream file1, file2, file0;
ofstream short_coord, long_coord, wave_file;
istringstream iss;
int value1, value2, i, totalcounts1, totalcounts2, start_start, end_start, start_int, end_int,length;
vector<int> site1, site2;
vector<int> counts1, counts2;
double wave1, wave2;
bool change;
```

```
int fiveprime, threeprime;
istringstream convert;
string start, end, one, two, three, four, five, six, name0;
short_coord.open("/project/boulware/SHRiMP_2_2_2/donor3_24hr_short_new.txt");
long_coord.open("/project/boulware/SHRiMP_2_2_2/donor3_24hr_long_new.txt");
//wave_file.open("/project/boulware/SHRiMP_2_2_2/cg_stim_donor1_6hr_wave.txt");
///file0.open("/project/boulware/SHRiMP_2_2_2/channel11_polyA_sites_uniq.txt");
//while(!file0.eof()){
//file0>>one>>two>>three>>four>>five>>six;
//name0=one+":"+two+"-"+three;
file1.open("/project/boulware/SHRiMP_2_2_2/channel4_polyA_sites_uniq.txt");
while(!file1.eof()){
getline(file1, line);
iss.str("");
iss.clear();
iss.str(line);
iss>>name1>>orient;
site1.clear();
counts1.clear();
//if(name0!=name1)
//continue;
        while(iss>>value1>>value2){
        site1.push_back(value1);
        counts1.push_back(value2);
        }
        if((int)site1.size()==0){
                continue;
                }
        //extract the length of the exon from the name.
        for(i=0; i<name1.length(); i++){
        if(name1[i]==':')
                start_start=i;
        if(name1[i]=='-')
                end_start=i;
        }
        start=name1.substr(start_start+1, end_start-start_start-1);
        end=name1.substr(end_start+1, name1.length()-end_start-1);
        convert.str("");
        convert.clear();
        convert.str(start);
        convert>>start_int;
        convert.str("");
    convert.clear();
    convert.str(end);
        convert>>end_int;
        length=sqrt((start_int-end_int)*(start_int-end_int));

        file2.open("/project/boulware/SHRiMP_2_2_2/channel6_3292012_polyA_sites_uniq.txt"
);
```

```
while(!file2.eof()){
getline (file2, line);
iss.str("");
iss.clear();
iss.str(line);
iss>>name2;
//cout<<name2<<endl;
        if(name1==name2){
        iss>>orient;
        //extract chromosome name
        for(i=0; i<name1.length(); i++){
        if(name1[i]==':')
                chrom=name1.substr(0,i);
        }

        site2.clear();
        counts2.clear();
        while(iss>>value1>>value2){
        site2.push_back(value1);
        counts2.push_back(value2);
        }
        if((int)site2.size()==0)
                break;
        //compute weighted average of lengths
        totalcounts1=0;
        totalcounts2=0;
        wave1=0;
        wave2=0;
        for(i=0; i<(int)site1.size(); i++){
        totalcounts1+=counts1[i];
        }
        for(i=0; i<(int)site1.size(); i++){
        wave1+=((double)counts1[i]/(double)totalcounts1)*site1[i];
        }

        for(i=0; i<(int)site2.size(); i++){
        totalcounts2+=counts2[i];
        }
        for(i=0; i<(int)site2.size(); i++){
  wave2+=((double)counts2[i]/(double)totalcounts2)*site2[i];
  }

        //decide if transcript got shorter or longer.
//      if(orient=="+"){
//      wave_file<<name1<<" "<<(wave2-wave1)<<endl;
//      }
//      else
//      wave_file<<name1<<" "<<(wave1-wave2)<<endl;
//
```

```
//                  if(sqrt((wave1-wave2)*(wave1-wave2))/length<0.1)
//                          break;


                    if(orient=="+"){
                            //output coordinates to shorter
                            fiveprime=*min_element(site1.begin(), site1.end());
                            if(*min_element(site2.begin(), site2.end())<fiveprime)
                                    fiveprime=*min_element(site2.begin(), site2.end());
                            threeprime=*max_element(site1.begin(), site1.end());
                            if(*max_element(site2.begin(), site2.end())>threeprime)
                                    threeprime=*max_element(site2.begin(), site2.end());
                            if(sqrt((fiveprime-threeprime)*(fiveprime-threeprime))<10)
                                    break;
                            if(wave1>wave2){
                            short_coord<<name1<<" "<<chrom<<" "<<fiveprime<<"
"<<threeprime<<" plus +"<<endl;
                            }

                            if(wave2>wave1){
                            //output coordinates to longer
                            long_coord<<name1<<" "<<chrom<<" "<<fiveprime<<"
"<<threeprime<<" plus +"<<endl;
                            }

                    }

                    if(orient=="-"){
                            fiveprime=*max_element(site1.begin(), site1.end());
                            if(*max_element(site2.begin(), site2.end())> fiveprime)
                                    fiveprime=*max_element(site2.begin(), site2.end());
                            threeprime=*min_element(site1.begin(), site1.end());
                            if(*min_element(site1.begin(), site1.end())<threeprime)
                                    threeprime=*min_element(site1.begin(), site1.end());
                            if(sqrt((fiveprime-threeprime)*(fiveprime-threeprime))<10)
                        break;

                            if(wave1>wave2){
                            long_coord<<name1<<" "<<chrom<<" "<<threeprime<<"
"<<fiveprime<<" minus -"<<endl;
                            }
                            if(wave2>wave1){
                            short_coord<<name2<<" "<<chrom<<" "<<threeprime<<"
"<<fiveprime<<" minus -"<<endl;
                            }

                    }
                    break;
                    }
```

```
            }
            file2.close();
}
file1.close();
//}
//file0.close();
return 0;
}
```