UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Quaking RNA binding proteins regulate tissue-specific gene expression

Permalink

https://escholarship.org/uc/item/6s15g9b6

Author

Fagg, William Samuel

Publication Date

2016

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SANTA CRUZ

QUAKING RNA BINDING PROTEINS REGULATE TISSUE-SPECIFIC GENE EXPRESSION

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

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELLULAR AND DEVELOPMENTAL BIOLOGY

by

W. Samuel Fagg IV

June 2016

The Dissertation of W. Samuel Fagg is approved by:

_____

Professor Manuel Ares Jr., chair

_____

Professor Susan Strome

_____

Professor Jeremy Sanford

_____

Tyrus Miller

Vice Provost and Dean of Graduate Studies

Copyright © by

W. Samuel Fagg IV

2016

TABLE OF CONTENTS

List of Figures iv

List of Tables viii

Abstract ix

Dedication xi

Acknowledgements xii

Chapter1 – Introduction 1

Chapter 2 – Structural Analysis of the Quaking Homodimerization Interface 27

Chapter 3 – Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation 51

Chapter 4 – Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression 65

Chapter 5 – Autogenous regulation balances Quaking functions in splicing and translation 98

Chapter 6 – Genomewide analysis of Quaking isoforms reveal distinct and opposing functions in regulation of gene expression 163

Chapter 7 – Conclusions and future directions 195

Bibliography 198

LIST OF FIGURES

1-1 Central dogma of molecular biology 3

1-2 Post-transcriptional regulation of gene expression 7

1-3 Alternative pre-mRNA splicing 9

1-4 Examples of alternative splicing that result in different outcomes 12

2-1 The STAR family of RNA-binding proteins. 30

2-2 Overall structure of the QkI Qua1 homodimerization subdomain 33

2-3 Qua1 point mutations that destabilize homodimerization also impair RNA binding 34

2-4 Dimerization-deficient point mutations in the Qua1 domain reduce QkI-5 splicing activity *in vivo* 36

2-5 Comparison of the QkI Qua1 structure (PDB ID 4DNN) to the GLD-1 (PDB ID 3K6T) and Sam68 (PDB ID 2XA6) 38

2-S1 ^{1}H, ^{15}N-HSQC NMR spectra of selected QkI Qua1 constructs 46

2-S2 QkI Qua1 structure 47

2-S3 Dimer interface surface of QkI, GLD1, and Sam68 Qua1 subdomain 48

2-S4 QkI Qua1 homology model from GLD-1 Qua1 template 49

2-S5 *In vivo* splicing activity of QkI-5 in mouse C2C12 myoblast cells 50

3-1 *Capzb* exon 9 is activated during myogenesis and has intronic ACUAA elements 54

3-2 Splicing reporters based on *Capzb* exon 9 55

3-3 Identification of ACUAA-binding proteins in proliferating C2C12 cells 56

3-4 QK and PTB control inclusion of *Capzb* exon 9 58

3-5 The QK-splicing regulatory network in proliferating myoblasts 59

3-6 Regulation of QK protein level, PTB protein level, and alternative splicing during myogenic differentiation 60

4-1 Quaking is expressed in macrophages within atherosclerotic lesions 68

4-2 QKI is highly expressed in macrophages derived from PB monocytes 69

4-3 Characterization of monocyte and macrophages biology in a unique QKI haploinsufficient patient 70

4-4 QKI influences pre-mRNA splicing in naïve PB monocytes and macrophages 72

4-5 QKI influences mRNA transcript abundance during differentiation of THP-1 monocyte-like cells to THP-1 macrophage-like cells 74

4-6 QKI expression levels influence pre-mRNA splicing during THP-1-based monocyte-like to macrophage-like cell differentiation 76

4-7 QKI expression levels have an impact on monocyte adhesion as well as migration and differentiation 78

4-8 QKI regulates the expression of atherosclerosis-related mRNAs and impairs foam cell formation 80

4-9 Schematic depicting how QKI posttranscriptionally regulates monocyte to macrophage differentiation and atherosclerosis development 81

4-S1 RNA-seq shows chr6-derived read coverage on chr5 86

4-S2 FACS analysis of surface marker expression levels in monocyte subsets in the QKI haploinsufficient patient and her sibling 87

4-S3 Schematic representation of transcript stratification 88

4-S4 QKI protein is differentially expressed during THP-1 monocyte-like to macrophage-like differentiation 89

4-S5 GapmeR-mediated reduction of QKI expression regulates the expression of atherosclerosis-related mRNAs 90

4-S6 Experimentally determined and putative transcription factor binding sites in the QKI promoter 91

4-S7 Original immunoblot images utilized for preparation of figures 1e, 2d, 3e, 5b, and Supplementary Fig. 4b 92

4-S8 Original bioanalyzer images utilized for preparation of figures 4c and 6d 93

4-S9 Original bioanalyzer images utilized for preparation of figures 6d and 6e 94

4-S10 Original bioanalyzer images utilized for preparation of figures 4f 95

5-S1 Quaking isoforms expressed in C2C12 myoblasts and myotubes 141

5-1 Qk5 is sufficient for splicing function 143

5-2 Nuclear localization is sufficient for Qk splicing function 145

5-S2 Qk5 is required for Qk6 and Qk7 accumulation and muscle-specific splicing 147

5-3 Qk5 promotes Qk6 expression through Qk6 mRNA 3'UTR 149

5-4 Qk6 inhibits Qk5 protein expression but activates its own protein expression 151

5-5 Qk5 is required for nuclear accumulation of Qk RNA 153

5-6 Qk autogenous regulation is conserved in rat C6 glioma cells 155

5-S4 Western blots supplemental to Figure 6 151

5-7 Model of Qk autogenous regulation 153

6-1 Loss of Qk isoforms results in different changes in mRNA abundance 176

6-S1 Qk5 globally regulates alternative splicing in C2C12 myoblasts 178

6-2 PanQk iCLIP identifies global RNA targets of Qk 180

6-S2 Correlation of number of CLIPper peaks per transcript with magnitude change in mRNA abundance 184

6-S3 Correlation of region of CLIPper peak within a transcript with magnitude change in mRNA abundance 186

6-3 Hmga2 and H19 are directly regulated by Qk5 188

6-4 Loss of Qk5 or Qk6&7 results in opposite changes in RNA abundance 190

LIST OF TABLES

2-1 Data collection and refinement statistics for QkI Qua1 (C35S) 32

2-2 Thermal stability of the QkI Qua1 dimer mutants and RNA affinities of QkI STAR mutants 34

2-S1 QkI-Qua1 C35S constructs cloned 45

3-1 Detection of RNA processing factors in WT- and M-RNA affinity chromatography fractions 57

4-1 IPA assessment of pre-defined canonical pathways affected by changes in QKI expression 77

4-S1 Primers [sic] sets utilized to determine mRNA abundance and pre-mRNA splicing events 96

6-1 Significant enrichment of bipartite sequence elements from iCLIP crosslinked regions 182

ABSTRACT

ANALYSIS OF QUAKING RNA BINDING PROTEINS AND THEIR ROLES IN POST-TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION

W. Samuel Fagg

Alternative RNA processing can robustly influence gene expression through a diverse set of molecular regulatory mechanisms. One of the most upstream RNA processing steps is splicing, which is the removal of non-coding intronic sequences from a pre-mRNA molecule and ligation of the exons together to make a mature mRNA. There are numerous types of alternative splicing that give rise to different quantitative and qualitative changes in the gene expression program, which can drastically impact cell fate and physiology. RNA binding proteins (RBPs) regulate alternative splicing and other types of RNA processing by binding substrate RNA molecules and altering downstream processing steps. How these processes are regulated for many RBPs is not known.

The Quaking (Qk in mouse, QKI in human) family of RBPs regulates many RNA processing steps including splicing, mRNA localization/decay, translation, and microRNA biogenesis. Interestingly, from a single *Qk* gene multiple Qk transcripts are generated by alternative splicing, but the different protein forms share identical dimerization and RNA binding domains. The studies presented here analyze many different aspects of how Qk regulates RNA processing. First, we determine the structure of the Qk dimerization domain and find mutations that disrupt it reduce protein stability and splicing functions *in vivo*. Next we show that Qk and PTB regulate overlapping splicing regulatory networks in myoblasts, but during differentiation to myotubes, Qk

protein increases while PTB decreases, leading to an increase in Qk splicing function and a decrease in PTB splicing. The next study reports that QKI regulates monocyte to macrophage differentiation through tissue-specific regulation of alternative splicing and mRNA abundance. The following study details isoform-specific functions and auto-regulatory interactions of the Qk5 and Qk6 isoforms, and the final study extends these observations genomewide.

The studies presented here identify novel mechanistic details of how Qk regulates RNA targets and shows that Qk regulates tissue specific gene expression during different stem cell types' differentiation.

DEDICATION

This dissertation is dedicated to my grandfather Gilbert Sherwood Callahan, my mom, Pat C. Wheeler and my dad WS Fagg III. This body of work is also, and just as importantly dedicated to the love of my life and my wife Margaret Jenkins Fagg and my son and most recent inspiration WS Fagg V. I thank you all for your unconditional love and support, as well as your patience, during my lengthy graduate career.

ACKNOWLEDGEMENTS

I am firstly and especially grateful for and to all members of the Ares lab for direct and indirect help for all of my work at UCSC. I would like to thank Haller Igel for daily support and guidance in all things lab and project related. Rhonda Perriman provided stringent critiques, advice, technical and academic support; Lily Shiue gave immense assistance particularly on genomics-based analyses including microarray and RNA-seq based assays; John Paul Donohue, Melissa Cline, and Sol Katzman provided computational expertise and data analyses; Jenny Vo provided exceptional support in completing the myotonic dystrophy project; Jason Talkish, Stephanie Nystrom, Pallavi Chowdhuri, and Jian provided excellent and daily feedback on assays both completed and underway. Also a huge thanks to other labs and PI's within the UCSC RNA Center, especially the Sanford Lab, Noller Lab, Jurica Lab, and Zahler Lab. My thesis committee members Jeremy Sanford and Susan Strome, thank you for your attentive feedback and guidance encompassing not only project, but also career, development. Finally, and most importantly, I would like to thank my advisor Manny Ares, who taught invaluable skills encompassing critical thinking, experimental design, science writing, and how to mentor others, in addition to unfailingly offering outstanding original ideas and explanations for my projects.

CHAPTER 1

INTRODUCTION

Metazoan organisms are made up of various cell and tissue types that give rise to critical anatomical structures that meet physiological requirements for multicellular life. However, the cells that make up an organism contain the same genetic information in the form of genes made of DNA. How cells with different functions use the same set of genes to produce the appropriate combination of protein or RNA molecules that influence the development and maintenance of different tissues is a hugely important question for both our basic understanding of biology and to better apply these findings toward therapies for various pathologies. Recent technological advancements have allowed us to approach this question on a larger scale than ever imagined. Together, many bodies of work studying regulation of gene expression using various cell types from different organisms indicate one of the most important factors that influence different cellular identities is the differential regulation of gene expression. The particular quantitative and qualitative output of gene expression gives rise to specific cell and tissue types. Thus regulation of gene expression is central to organ system and organismal identity.

Gene Expression

The underlying concept of the central dogma of molecular biology put forth by Crick nearly 60 years ago, that information flows from nucleic acid to protein, not from protein to nucleic acid (Crick 1970), still holds true. A simple summary of this model is that genetic information is stored as DNA, which is converted to RNA through a process called transcription, and RNA is further processed then converted to protein in a process called translation (Fig 1). However recent refinements to and

Figure 1: Central Dogma of Molecular Biology



Fig 1-1: Central dogma of molecular biology. Genes are stored as DNA, which is transcribed into RNA, which is translated into protein. The flow of information is always from nucleic acid to protein, not backward.

regulatory control points for the central dogma of molecular biology provide a more complicated model for how the flow of genetic information proceeds. These processes have multiple disparate points of regulatory control, and while separated both spatially and temporally, are integrated with one another and culminate in a final protein product.

Transcription is a highly regulated process that occurs in the nucleus of eukaryotic cells. Chromatin organization and structure regulate promoter accessibility, and then initiation by RNA polymerase may occur on an accessible promoter and is also subject to multiple regulatory control points. Once transcription elongation occurs on a protein-coding gene a precursor messenger RNA (pre-mRNA) emerges from RNA polymerase that is immediately subject to further processing steps. These are collectively referred to as co-transcriptional processing, and include addition of a 7-methylguanosine cap (capping), removal of intervening sequences that do not code for protein (introns) and ligation of protein coding sequences (exons) together (splicing), addition of multiple adenosine nucleotides at the 3' end of the substrate RNA (polyadenylation), and cleavage of the RNA substrate from the polymerase (Jonkers and Lis 2015; Naftelberg et al. 2015). These steps contribute not only to the maturation of the substrate RNA, but also protect these molecules from decay and promote export from the nucleus to the cytoplasm by packing them into messenger ribonucleoprotein particles (mRNPs) (Muller-McNicoll and Neugebauer 2013; Sloan et al. 2015). In the cytoplasm the mRNA is also protected from decay by incorporation into the mRNP if it is to be translated (Singh et al. 2015). Multiple

rounds of translation can occur on a single mRNA, but after completion the mRNP is disassembled and the mRNA is degraded (Singh et al. 2015).

A more recent discovery that complicates the simplicity of the central dogma model is identification of a diverse class of non-coding RNAs (ncRNAs). These ncRNAs are also transcribed from a DNA template and subject to a number of additional processing steps (Towler et al. 2015; Quinn and Chang 2016), but unlike mRNAs, do not code for protein. Like proteins, though, these ncRNAs can have diverse functions and influence cellular identity (Quinn and Chang 2016; Virant-Klun et al. 2016).

An important point of distinction regarding the outcomes of different types of regulation of gene expression is whether the consequence is a quantitative or qualitative change. For example, increasing transcription on a gene will lead to a quantitative increase in the amount of product from that gene. On the other hand, alternative splicing of the same gene (assuming a constant rate of transcription) can lead to qualitative changes in the product from that same gene. The downstream influences on the overall molecular and cell biology between these different types of gene expression regulation can vary widely. This distinction further complicates how regulated gene expression contributes to cellular identity, and is outlined in greater detail below.

Post-transcriptional regulation of gene expression

The general overview of gene expression described above does not account for alternative control points of regulation that may be imposed on RNA molecules.

Collectively referred to as post-transcriptional regulation, each step of RNA processing has the potential for immensely complex and varied points of control, some of which can have different outcomes. As noted above, alternative usage and influence of the various regulatory control points on the transcriptome give rise to different gene expression patterns and different cell types. Additionally, many of the proteins translated from mRNAs in a cell themselves then feedback to regulate gene expression.

Direct binding of an RNA binding protein (RBP) to substrate RNAs via an RNA binding domain (RBD) is one possible means by which post-transcriptional regulation can be imposed. Binding may occur if a binding site (sequence motif) is present in a substrate RNA, if the folding/secondary structure of the RNA allows for RBD recognition of the motif, and if the substrate RNA and RBP are in close proximity to one another. The consequences of RBP binding to an RNA can be varied, and depend on the location of the binding site, cellular compartment in which binding occurs, additional protein interaction partners with the RBP, and many additional factors. These alternative RNA processing steps may include alternative splicing, alternative polyAdenylation, differential pre-mRNA or mRNA stability/decay, differential subcellular localization, alternative levels of translation, and alternative processing of non-coding RNAs such as microRNAs (Fig 2), and are explained in greater detail below.

Figure 2: Post-transcriptional regulation of gene expression

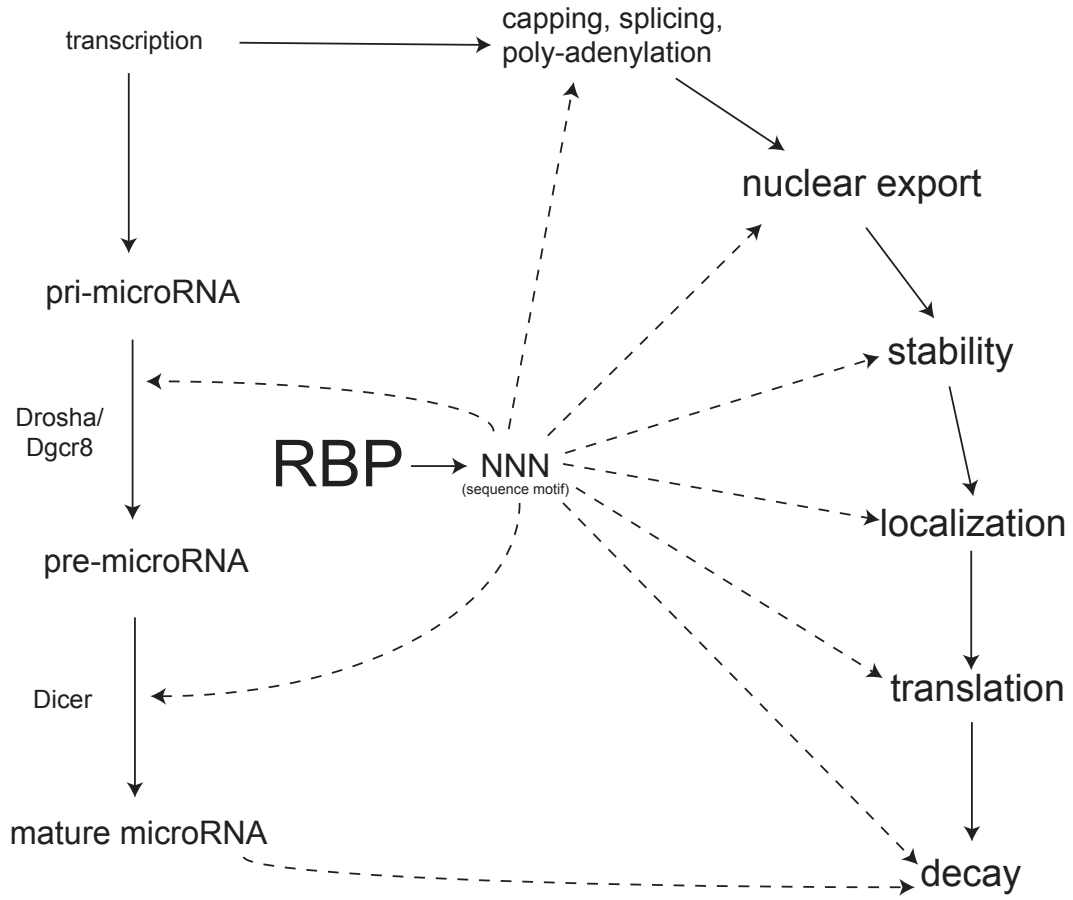


Fig 1-2: Post-transcriptional regulation of gene expression. Production of an RNA molecule from transcription gives gives rise to multiple step-wise processing steps in the lifecycle of both protein coding RNAs (mRNAs, right) and non-coding RNAs (microRNA shown as example on left). Sequence specific RNA binding proteins (RBPs) bind specific sequence elements (NNN) through which they regulate various RNA processing steps (denoted by dashed arrows). microRNAs can also feedback into regulation of mRNA processing as noted by dashed arrows.

Alternative splicing

The term alternative splicing simply refers to the pre-mRNA splicing process of a single substrate RNA in which more than one outcome is possible. Alternative splicing is executed by differential regulation and recruitment of the splicing machinery to a substrate pre-mRNA and often times mediated by RBP direct association with sequence elements in pre-mRNAs. This process, however, is extremely complicated in that both the upstream regulation imposed and the downstream consequences can be numerous and many times interconnected within a broader regulatory network. Furthermore, regulation of alternative splicing is largely context dependent: repeated experimental observations in one or several cell types may not necessarily be applicable in other cell types (Fu and Ares 2014). Therefore understanding how alternative splicing is regulated in diverse cell and tissue types, at various developmental stages, and how it is dysregulated in pathologies remains an open and significant challenge.

An overview of several examples of alternative splicing is shown in Figure 3 and covered in more comprehensive detail elsewhere (Black 2003; Matlin et al. 2005). In most cases, unless the protein-coding reading frame is disrupted, the outcome is an mRNA, and thus protein, containing different sequence. The consequences of these sequence variations range from subtle and/or undetectable to greatly impacting the regulation of the mRNA variants produced or function of resulting protein isoforms. For example, alternative 3' splice site (ss) usage can vary by as little as 3 nucleotides and still maintain a reading frame. Referred to as tandem

Figure 3: Diversity of alternative splicing

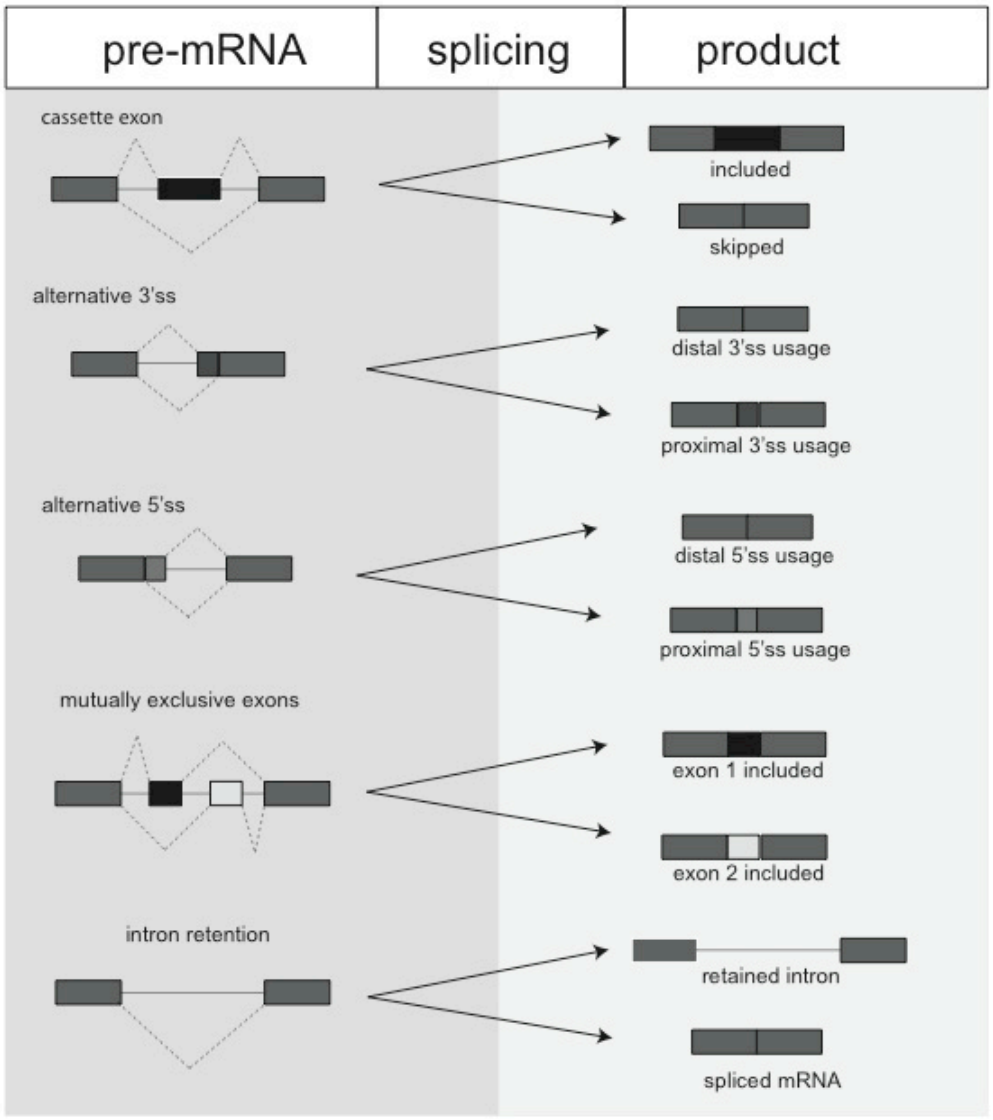


Fig 1-3: Alternative pre-mRNA splicing. Pre-mRNA splicing substrates are shown on the left, the middle arrows represent 2 potential alternative splicing regulatory events, with potential mature mRNA products shown on right.

3’ss or NAGNAG splicing, these sequences occur and are widely utilized in plants and metazoans (Hiller et al. 2004; Akerman and Mandel-Gutfreund 2006; Schindler et al. 2008; Ragle et al. 2015), are differentially regulated in tissues and disease (Hiller et al. 2006; Ragle et al. 2015), and contribute to proteomic diversity and plasticity (Bradley et al. 2012). Therefore, only a 3 nucleotide difference generated by alternative 3’ss usage occurs very frequently throughout the transcriptome of many different organisms and has wide-ranging functional consequences.

An additional intriguing example where alternative splicing greatly impacts the mRNA product is found when the spliced product consists of different last exons. This can be a result of alternative 5’ or 3’ss usage, cassette exon skipping or inclusion, or intron retention. Recently, some SR proteins have been shown to regulate (in some cases opposingly) alternative last exon splicing, which directly impacts the nuclear export, and thus cytoplasmic abundance, or different mRNA forms (Muller-McNicoll et al. 2016). A similar observation has been reported that suggests that alternative last exon splicing that results in different 3’UTR mRNA species significantly influences mRNA localization in neurons ((Taliaferro et al. 2016). Furthermore, alternative 3’UTR containing mRNAs can contain different regulatory motifs such as microRNA binding sites, and thus be differentially expressed due to alternative microRNA-mediated turnover (Boutet et al. 2012; Ji et al. 2013; Hoffman et al. 2016).

Additionally, alternative splicing can produce protein isoforms whose functions differ. The actin filament capping protein isoforms Capzβ1 and Capzβ2 localize to different regions of myocytes and regulate different aspects of muscle cell physiology due to differences in C-terminal protein sequences (Schafer et al. 1994; Hart and Cooper 1999; Yamashita et al. 2003). A recent study found that inclusion of exon 9 in Capzb mRNA is directly promoted by Quaking (Qk) but repressed by poly-pyrimidine tract binding (PTB) RBPs (Hall et al. 2013). Therefore, the overlapping regulatory influences on alternative splicing between these two RBPs, which converge on exon 9 of Capzb (and other alternative exons), critically regulate muscle cell differentiation and physiology.

In summary, alternative splicing serves as a central regulatory point for post-transcriptional regulation of gene expression. The examples outlined above detail several cases where alternative splicing leads not only to changes in protein isoform identity and functional output, but also as a platform for additional forms of downstream alternative RNA processing. As such, a popular but inaccurate representation of RBP sequence motif binding sites as either "splicing enhancers" or "splicing silencers" has been eschewed here, given the context dependence of a specific binding event by an RBP to an RNA molecule. In many cases, the location of a specific sequence motif provides clues to the consequences of RBP binding in the case of alternative splicing. Most tissue-specific RBPs that regulate alternative splicing have a strong correlation with promoting inclusion of cassette exons by binding in the intron downstream of the alternative exon, and promoting skipping by

Figure 4: Consequences of alternative splicing

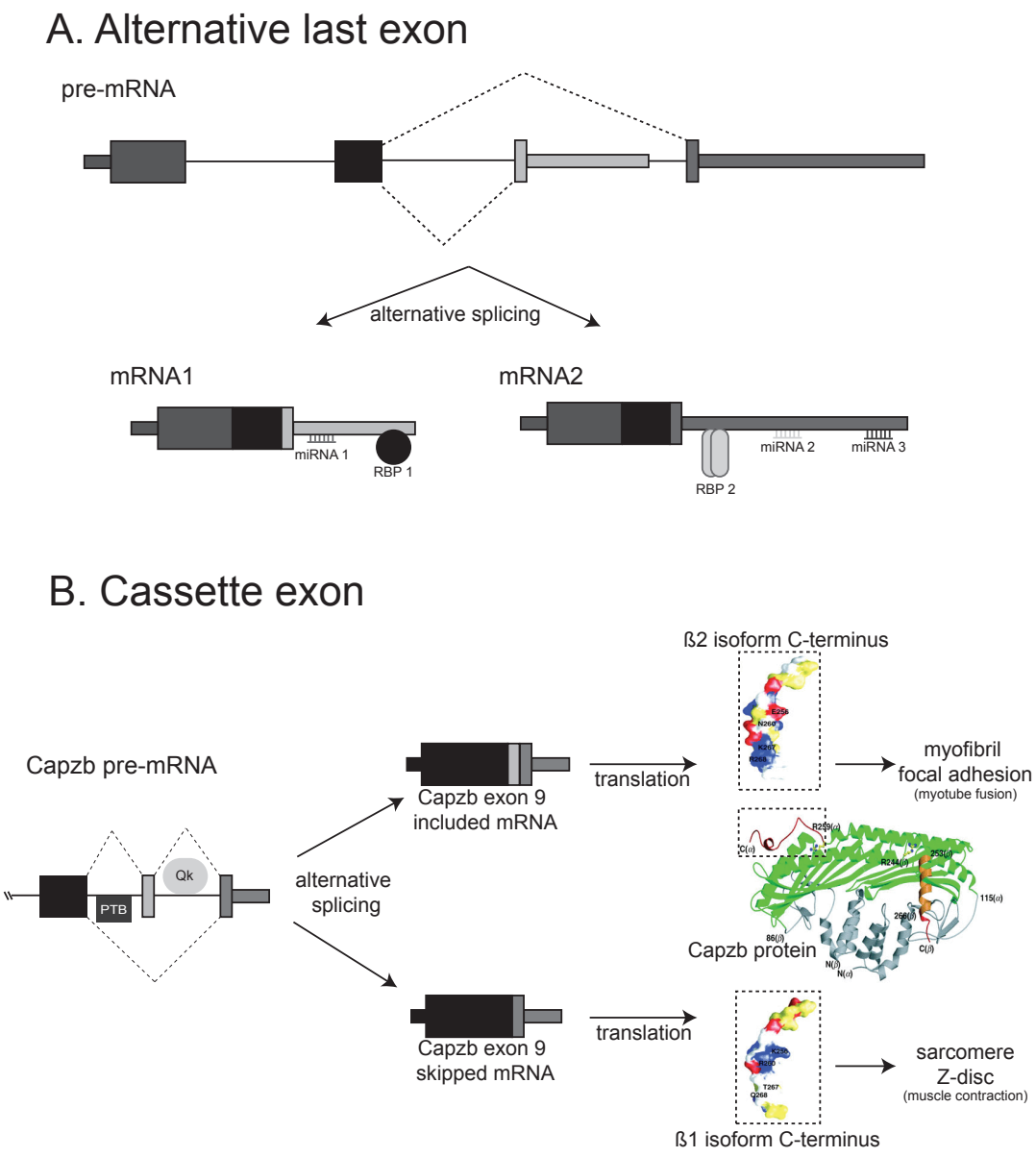


Fig 1-4: Examples of alternative splicing that result in different outcomes. A. Alternative last exons can be produced by alternative 3'ss usage. The different 3'ends contain different protein coding sequence and 3'UTR sequence. The different 3'UTRs contain different regulatory elements that so mRNA1 is subject to different regulatory constraints than mRNA2. B. Alternative splicing of the Capzb exon 9 cassette exon generates Capzb protein isoforms with different C-termini (dotted boxes) which have different localization patterns and functions in muscle cells.

binding in the intron upstream of the alternative exon (Ule et al. 2006; Licatalosi et al. 2008; Xue et al. 2009; Yeo et al. 2009; Du et al. 2010; Llorian et al. 2010; Hall et al. 2013). Additionally popular, yet overly simplistic reference to these RBPs simply as splicing factors is inaccurate, as these RBPs and families of RBPs regulate multiple overlapping processes and steps in RNA metabolism.

RNA binding protein families

In mammals, RBPs are divided up into families (eg, SR protein family, Muscleblind-like (Mbnl) family, Polypyrimidine tract-binding (PTB) family), based on their RBDs. Since the RBD is the very similar among these family members, they usually bind to the same sequence motif, although specific family members tend to function distinctly, in a non-overlapping manner. For example, in the Mbnl family, the members bind a YGCY-like sequence motif, and yet do not functionally overlap. In muscle, Mbnl1 is a splicing factor while Mbnl2 regulates mRNA localization, but in brain Mbnl2 is a splicing factor (Du et al. 2010; Charizanis et al. 2012; Wang et al. 2012). Thus, these two different protein family members demonstrate functional specification within the same tissue (muscle), but also function distinctly in a tissue-specific manner: Mbnl1 is a splicing factor in muscle but Mbnl2 is a splicing factor in brain. How this division of labor between this and other RBP family members is established, maintained, and regulated remains an open question.

This and additional similar observations for other RBP families suggest two major possibilities for how RBP family members may be functionally restricted: 1)

promoter activity/activation between different family members encoded by different genes can produce either a single family member or one that is dominantly expressed in a given tissue, and 2) protein sequence variation outside of the RBD may serve to functionally enhance or restrict different RBP family members in a tissue-specific manner. However sorting out these potential differences within the context of transcription on different genes combined with variable degrees of mRNA and protein sequence differences between RBP family members is a challenge.

Quaking family of RNA binding proteins

From a single *Quaking* (*Qk* gene in mouse, or *QKI* in human) gene at least 3 distinct RBPs (Qk5, Qk6, and Qk7) can be produced (Ebersole et al. 1996; Kondo et al. 1999). Therefore, there is only one promoter, and all protein isoforms share the same 311 amino acids (including the RBD) with only a small amount of isoform-specific protein sequence at each C-terminus.

Although there are multiple Quaking (Qk in mouse) proteins, *Qk* is a member of a larger RBP family called Signal Transduction and Activator of RNA (STAR) metabolism family. These highly conserved proteins respond to cell signaling cues and regulate post-transcriptional gene expression in response to these signals (Vernet and Artzt 1997; Richard 2010). The *Quaking* gene was originally identified in the qk^{viable} mutant mouse, where a 1 megabase deletion around the *Qk* promoter causes lower expression of *Qk* in oligodendrocytes, leading to defective myelination and a shaking phenotype (Sidman et al. 1964; Ebersole et al. 1996; Hardy et al. 1996; Lu et al. 2003). Interestingly, though, knockout mice lacking *Qk* are embryonic lethal,

demonstrating a requirement for Qk protein isoforms in development (Li et al. 2003). The importance of *Qk* function in development is also apparent in model organisms where *Qk* orthologs/relatives are required for muscle and tendon development as well as germ cell fate (Francis et al. 1995; Baehrecke 1997; Zaffran et al. 1997; Ohno et al. 2008; Lobbardi et al. 2011). Consistent with a pro-development/differentiation role, reduced *Quaking* expression is correlated with several types of cancer, demonstrating a role as a tumor suppressor (Novikov et al. 2011; Chen et al. 2012; Ji et al. 2012; Zong et al. 2014). Therefore, expression and regulation of *Qk* and its evolutionarily conserved family members are functionally required for proper development and homeostasis of numerous different tissues.

Collectively, *Qk* regulates RNA processing at several steps including pre-mRNA splicing, mRNA localization, mRNA stability/decay, translation, and miRNA processing (Saccomanno L 1999; Li et al. 2000; Larocque et al. 2002; Larocque et al. 2005; Zhao et al. 2010; Zearfoss et al. 2011; Hall et al. 2013; Zong et al. 2014). These functions are mediated by Qk proteins binding the consensus sequence motif ACUAA in target RNA molecules: regulating mRNA expression, localization, and translation by binding to 3'UTR ACUAA sequence elements, and regulating pre-mRNA splicing by binding to intronic ACUAA elements (Ryder and Williamson 2004; Galarneau and Richard 2005; Hafner et al. 2010; Hall et al. 2013). A bipartite sequence motif has also been identified, and since Qk exists as a dimer, a single Qk dimer is capable of making two points of contact on one RNA molecule (Galarneau and Richard 2005). Underscoring the importance of Qk dimerization is the

observation that a point mutation in the dimerization domain is embryonic lethal, and dimerization domain mutant proteins fail to regulate splicing as efficiently as wild type dimerization domain proteins (Ebersole et al. 1996; Chen T 1998; Beuck et al. 2012). Additionally, all Qk isoforms share the same dimerization domain, suggesting that Qk isoform heterodimers may form. Indeed, studies have demonstrated heterodimerization is possible in yeast-two hybrid, and by overexpressing different epitope tagged versions of Qk isoforms and co-immunoprecipitating different isoforms bound to one another (Wu et al. 1999; Pilotte et al. 2001). How or if these heterodimers bear physiological relevance is still an open question, but heterodimerization between different Qk isoforms may influence the localization of these proteins and therefore their functional capacity (Pilotte et al. 2001).

The observations that *Qk* regulates multiple disparate forms of RNA processing together with the existence of multiple Qk isoforms and the consequences of their mis-regulation, suggests these isoforms may function independently from one another. However, other observations, such as the potential for heterodimer function and that a single mRNA could be regulated sequentially by more than one Qk isoform, suggest the isoforms may functionally synergize with one another. The Qk5 isoform has a non-canonical nuclear localization signal (NLS) within its unique C-terminal amino acids and shuttles from the nucleus to the cytoplasm and back (Wu et al. 1999). Additionally, in Qk^{viable} oligodendroctyes, the Qk5 isoform sequesters mRNAs encoding myelin basic protein (Mbp) in the nucleus, consequentially down regulating their expression (Larocque et al. 2002). The C-terminus specific to the Qk7

isoform has been shown to encode a "killer sequence" that can induce apoptosis when localized to the cytoplasm, a function that can be repressed by heterodimerization mediated nuclear translocation with Qk5 and Qk6 isoforms (Pilotte et al. 2001). Additionally, Qk7 may function to protect target mRNAs from decay in glial cells (Larocque et al. 2005). Functionally, Qk6 has been shown to regulate mRNA stability, localization, and translation, and the Qk^{viable} phenotype can be partially rescued by expression of this single isoform in affected oligodendrocytes (Saccomanno L 1999; Larocque et al. 2005; Zhao et al. 2006; Zhao et al. 2010; Mandler et al. 2014).

Therefore, *Qk* is unique in that a single gene encodes at least 3 distinct RBPs. These 3 RBPs share identical dimerization domains, RBDs, and majority of protein amino acid sequence. The subsequent chapters examine the mechanism of Qk dimerization, how Qk regulates muscle-specific splicing, the contribution of Qk to monocyte and macrophage cell physiology, how the Qk RBPs regulate different functions in the same cells including how they self-regulate, and finally what the consequences of loss of function of different Qk isoforms are on global gene expression in muscle. These novel findings have and will contribute to both mechanistic insights in molecular biology and biochemistry, and allow for refinement of therapies through translational medicine.

References

Akerman M, Mandel-Gutfreund Y. 2006. Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res* **34**: 23-31.

Baehrecke EH. 1997. who encodes a KH RNA binding protein that functions in muscle development. *Development* **124**: 1323-1332.

Beuck C, Qu S, Fagg WS, Ares M, Jr., Williamson JR. 2012. Structural analysis of the Quaking homodimerization interface. *J Mol Biol*.

Black DL. 2003. MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING. *Annual Review of Biochemistry* **72**: 291-336.

Boutet SC, Cheung TH, Quach NL, Liu L, Prescott SL, Edalati A, Iori K, Rando TA. 2012. Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell stem cell* **10**: 327-336.

Bradley RK, Merkin J, Lambert NJ, Burge CB. 2012. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS biology* **10**: e1001229.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G et al. 2012. Muscleblind-like 2-Mediated Alternative Splicing in the Developing Brain and Dysregulation in Myotonic Dystrophy. *Neuron* **75**: 437-450.

Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N et al. 2012. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes & development* **26**: 1459-1472.

Chen T RS. 1998. Structure-function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization. *Molecular and cellular biology* **18**: 4863-4871.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227**: 561-563.

Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA et al. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature structural & molecular biology* **17**: 187-193.

Ebersole TA, Chen Q, Justice MJ, Artzt K. 1996. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genetics* **12**: 260-265.

Francis R, Barton MK, Kimble J, Schedl T. 1995. gld-1, a tumor suppressor gene required for oocyte development in Caenorhabditis elegans. *Genetics* **139**: 579-606.

Fu XD, Ares M, Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature reviews Genetics* **15**: 689-701.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature structural & molecular biology* **12**: 691-698.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP, Ares M, Jr. 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA (New York, NY)* **19**: 627-638.

Hardy RJ, Loushin CL, Friedrich Jr VL, Chen Q, Ebersole TA, Lazzarini RA, Artzt K. 1996. Neural Cell Type-Specific Expression of QKI Proteins Is Altered in quakingviable Mutant Mice. *J Neurosci* **16**: 7941-7949.

Hart MC, Cooper JA. 1999. Vertebrate isoforms of actin capping protein beta have distinct functions In vivo. *The Journal of cell biology* **147**: 1287-1298.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. 2004. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* **36**: 1255-1257.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. 2006. Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *American journal of human genetics* **78**: 291-302.

Hoffman Y, Bublik DR, A PU, Elkon R, Biniashvili T, Agami R, Oren M, Pilpel Y. 2016. 3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-

differentiation Genes in Proliferating Human Cells. *PLoS Genet* **12**: e1005879.

Ji S, Ye G, Zhang J, Wang L, Wang T, Wang Z, Zhang T, Wang G, Guo Z, Luo Y et al. 2012. miR-574-5p negatively regulates Qki6/7 to impact beta-catenin/Wnt signalling and the development of colorectal cancer. *Gut*.

Ji S, Ye G, Zhang J, Wang L, Wang T, Wang Z, Zhang T, Wang G, Guo Z, Luo Y, Cai J, Yang JY. 2013. miR-574-5p negatively regulates Qki6/7 to impact beta-catenin/Wnt signalling and the development of colorectal cancer. *Gut* **62**: 716-726.

Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**: 167-177.

Kondo T, Furuta T, Mitsunaga K, Ebersole TA, Shichiri M, Wu J, Artzt K, Yamamura K, Abe K. 1999. Genomic organization and expression analysis of the mouse qkI locus. *Mammalian genome : official journal of the International Mammalian Genome Society* **10**: 662-669.

Larocque D, Galarneau A, Liu H-N, Scott M, Almazan G, Richard S. 2005. Protection of p27Kip1 mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat Neurosci* **8**: 27-33.

Larocque D, Pilotte J, Chen T, Cloutier F, Massie B, Pedraza L, Couture R, Lasko P, Almazan G, Richard S. 2002. Nuclear Retention of MBP mRNAs in the Quaking Viable Mice. *Neuron* **36**: 815-829.

Li Z, Takakura N, Oike Y, Imanaka T, Araki K, Suda T, Kaname T, Kondo T, Abe K, Yamamura K. 2003. Defective smooth muscle development in qkI-deficient mice. *Development, growth & differentiation* **45**: 449-462.

Li Z, Zhang Y, Li D, Feng Y. 2000. Destabilization and Mislocalization of Myelin Basic Protein mRNAs in quaking Dysmyelination Lacking the QKI RNA-Binding Proteins. *J Neurosci* **20**: 4944-4953.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464-469.

Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, de la Grange P, Ast G et al. 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nature structural & molecular biology* **17**: 1114-1123.

Lobbardi R, Lambert G, Zhao J, Geisler R, Kim HR, Rosa FM. 2011. Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Development* **138**: 1783-1794.

Lu Z, Zhang Y, Ku L, Wang H, Ahmadian A, Feng Y. 2003. The quakingviable mutation affects qkI mRNA expression specifically in myelin-producing cells of the nervous system. *Nucleic acids research* **31**: 4616-4624.

Mandler MD, Ku L, Feng Y. 2014. A cytoplasmic quaking I isoform regulates the hnRNP F/H-dependent alternative splicing pathway in myelinating glia. *Nucleic acids research* **42**: 7319-7329.

Matlin AJ, Clark F, Smith CWJ. 2005. Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology* **6**: 386-398.

Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM. 2016. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* **30**: 553-566.

Muller-McNicoll M, Neugebauer KM. 2013. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet* **14**: 275-287.

Naftelberg S, Schor IE, Ast G, Kornblihtt AR. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**: 165-198.

Novikov L, Park JW, Chen H, Klerman H, Jalloh AS, Gamble MJ. 2011. QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Molecular and cellular biology* **31**: 4244-4255.

Ohno G, Hagiwara M, Kuroyanagi H. 2008. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes & development* **22**: 360-374.

Pilotte J, Larocque D, Richard S. 2001. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes & development* **15**: 845-858.

Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**: 47-62.

Ragle JM, Katzman S, Akers TF, Barberan-Soler S, Zahler AM. 2015. Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in C. elegans. *Genome Res* **25**: 982-994.

Richard S. 2010. Reaching for the stars: Linking RNA binding proteins to diseases. *Advances in experimental medicine and biology* **693**: 142-157.

Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA (New York, NY)* **10**: 1449-1458.

Saccomanno L LC, Jan E, Punkay E, Artzt K, and Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 12605-12610.

Schafer DA, Korshunova YO, Schroer TA, Cooper JA. 1994. Differential localization and sequence analysis of capping protein beta-subunit isoforms of vertebrates. *The Journal of cell biology* **127**: 453-465.

Schindler S, Szafranski K, Hiller M, Ali GS, Palusa SG, Backofen R, Platzer M, Reddy AS. 2008. Alternative splicing at NAGNAG acceptors in Arabidopsis thaliana SR and SR-related protein-coding genes. *BMC Genomics* **9**: 159.

Sidman R, Dickie M, Appel S. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. *Science (New York, NY)* **144**: 309-311.

Singh G, Pratt G, Yeo GW, Moore MJ. 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* **84**: 325-354.

Sloan KE, Gleizes PE, Bohnsack MT. 2015. Nucleocytoplasmic Transport of RNAs and RNA-Protein Complexes. *J Mol Biol*.

Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS et al. 2016. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol Cell* **61**: 821-833.

Towler BP, Jones CI, Newbury SF. 2015. Mechanisms of regulation of mature miRNAs. *Biochem Soc Trans* **43**: 1208-1214.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**: 580-586.

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

Virant-Klun I, Stahlberg A, Kubista M, Skutella T. 2016. MicroRNAs: From Female Fertility, Germ Cells, and Stem Cells to Cancer in Humans. *Stem cells international* **2016**: 3984937.

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S et al. 2012. Transcriptome-wide Regulation of Pre-

mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* **150**: 710-724.

Wu J, Zhou L, Tonissen K, Tee R, Artzt K. 1999. The Quaking I-5 Protein (QKI-5) Has a Novel Nuclear Localization Signal and Shuttles between the Nucleus and the Cytoplasm. *J Biol Chem* **274**: 29202-29210.

Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Molecular cell* **36**: 996-1006.

Yamashita A, Maeda K, Maeda Y. 2003. Crystal structure of CapZ: structural basis for actin filament barbed end capping. *The EMBO journal* **22**: 1529-1538.

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature structural & molecular biology* **16**: 130-137.

Zaffran S, Astier M, Gratecos D, Semeriva M. 1997. The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Development* **124**: 2087-2098.

Zearfoss NR, Clingman CC, Farley BM, McCoig LM, Ryder SP. 2011. Quaking regulates Hnrnpa1 expression through its 3' UTR in oligodendrocyte precursor cells. *PLoS genetics* **7**: e1001269.

Zhao L, Mandler MD, Yi H, Feng Y. 2010. Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein.

Proceedings of the National Academy of Sciences of the United States of America **107**: 19061-19066.

Zhao L, Tian D, Xia M, Macklin WB, Feng Y. 2006. Rescuing qkv Dysmyelination by a Single Isoform of the Selective RNA-Binding Protein QKI. *J Neurosci* **26**: 11278-11286.

Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H et al. 2014. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS genetics* **10**: e1004289.

CHAPTER 2

STRUCTURAL ANALYSIS OF THE QUAKING HOMODIMERIZATION INTERFACE

Reprinted with permission from

Journal of Molecular Biology *423* 766-781 (August 2012)

Copyright © 2012 by Elsevier Ltd.

ARTICLE IN PRESS
IMF YJMBI-63801; No. of pages: 16; 4C: 5, 6, 10

http://dx.doi.org/10.1016/j.jmb.2012.08.027 J. Mol. Biol. (2012) xx, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Analysis of the Quaking Homodimerization Interface

Christine Beuck[1], Song Qu[1], W. Samuel Fagg[2], Manuel Ares Jr.[2] and James R. Williamson[1]*

[1]*Department of Molecular Biology, Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*
[2]*Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, CA 95064, USA*

Received 11 July 2012;
received in revised form
30 August 2012;
accepted 31 August 2012

Edited by C. R. Matthews

Keywords:
STAR/GSG proteins;
Qua1 domain;
crystal structure;
RNA binding;
C2C12 myoblasts

Quaking (QkI) is a prototypical member of the STAR (signal *t*ransducer and *a*ctivator of *R*NA) protein family, which plays key roles in posttranscriptional gene regulation by controlling mRNA translation, stability and splicing. QkI-5 has been shown to regulate mRNA expression in the central nervous system, but little is known about its roles in other tissues. STAR proteins function as dimers and bind to bipartite RNA sequences; however, the structural and functional roles of homodimerization and heterodimerization are still unclear. Here, we present the crystal structure of the QkI dimerization domain, which adopts a similar stacked helix–turn–helix arrangement as its homologs GLD-1 (*g*erm *l*ine development *d*efective-*1*) and Sam68 (*S*rc-*a*ssociated protein during *m*itosis, *68* kDa) but differs by an additional helix inserted in the dimer interface. Variability of the dimer interface residues likely ensures selective homodimerization by preventing association with non-cognate STAR family proteins in the cell. Mutations that inhibit dimerization also significantly impair RNA binding *in vitro*, alter QkI-5 protein levels and impair QkI function in a splicing assay *in vivo*. Together, our results indicate that a functional Qua1 homodimerization domain is required for QkI-5 function in mammalian cells.
© 2012 Elsevier Ltd. All rights reserved.

Introduction

Quaking (QkI) is a conserved multifunctional protein in vertebrates that belongs to the STAR (signal *t*ransducer and *a*ctivator of *R*NA) family of RNA-binding proteins. QkI regulates a multitude of cellular functions, both at different developmental times and in various tissues in mice and humans (reviewed in Ref. 1). STAR family proteins link signaling pathways to various aspects of posttranscriptional regulation of mRNAs. QkI is important for several biological processes including myelination, smooth muscle cell differentiation, vascular development and heart development.[2–4] The cellular processes controlled by QkI are reported to include target mRNA stabilization, translation, subcellular localization and alternative splicing.[5–8] The consensus RNA binding sequence of QkI is A(C/U)UAA(C/U) and is often accompanied by a so-called half-site UAA(C/U) in close proximity.[9–11] The activity of QkI and other STAR family members is regulated via posttranslational modifications including phosphorylation.[12–14] However, the molecular mechanisms linking RNA binding and homodimerization to the diverse regulatory roles of QkI are still poorly understood.

*Corresponding author. E-mail address: jrwill@scripps.edu.

Abbreviations used: CNS, central nervous system; MBP, maltose-binding protein; FP, fluorescence polarization; MR, molecular replacement; PDB, Protein Data Bank; RT, reverse transcription; SeMet, selenomethionine; SSRL, Stanford Synchrotron Radiation Laboratory; wt, wild type.

0022-2836/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

Several alternatively spliced isoforms of Quaking protein differ in their C-terminal 8- to 30-amino-acid sequences, with the three major isoforms being QkI-5, QkI-6 and QkI-7. These unique tails determine the localization within the cell and confer specific functions. QkI-6 and QkI-7 are mainly cytoplasmic, while QkI-5 contains a noncanonical nuclear localization signal within its tail.[15–17]

The most extensively studied aspect of QkI function is its crucial role in myelin formation in the central nervous system (CNS), where QkI controls the expression of many genes important for myelination through a complex mechanism involving all pathways mentioned above. Defects in QkI expression and/or splicing in the CNS have been implicated in several psychiatric diseases including schizophrenia and ataxia.[18,19] In quaking-viable mutation in mice (qk^v mice), myelin fails to undergo compaction in the CNS due to a large deletion upstream of the quaking locus.[20] The alternative splicing of several myelin-specific genes regulated by QkI has been implicated as an important mechanism in regulation of myelin formation.[5,21,22] Recently, QkI expression was correlated with alternative splicing of the histone variant MacroH2A1 in several different human tissues and cancer cell lines, linking defects in MacroH2A1 splicing to a variety of cancers (including testicular, lung, bladder, cervical, breast, colon, ovarian and endometrial) and suggesting that QkI may be an important tumor suppressor.[23,24]

The conserved STAR domain defines the STAR protein family and consists of a central KH (heterogeneous nuclear ribonucleoprotein *K h*omology) RNA binding domain flanked by two conserved subdomains referred to as Qua1 and Qua2 (Fig. 1).[2] STAR proteins bind RNA via the KH and Qua2 regions, while the Qua1 domain is essential for homodimerization.[25–27] The current model for a STAR protein dimer binding to RNA suggests a bipartite binding mode. One STAR protomer is thought to bind one 6- to 7-nt consensus sequence, while the second protomer binds to an additional full consensus site or shorter (3–4 nt) half-site spaced by 1–20 nt.[9,11,27,28] Alternatively, recognition of two distantly spaced consensus sites could organize mRNA structure for regulation by QkI.[29–31]

The structural basis for the recognition of a single RNA consensus site has been described for the splicing factor SF1,[32] indicating that the Qua2 domain extends the RNA-binding surface of the KH domain and that this KH–Qua2 tandem domain is essential for sequence-specific RNA recognition. Interestingly, SF1 is the only STAR family member that lacks the Qua1 dimerization domain. The solution structure of *Xenopus* Quaking KH–Qua2 domain shows that the Qua2 domain does not contact the KH domain in the absence of RNA.[33] Qua1 homodimerization domain structures have been solved for the GLD-1 (*g*erm *l*ine development *d*efective-*1*) and Sam68 (*S*rc-*a*ssociated protein during *m*itosis, *68* kDa) homologs of QkI, revealing a helix–turn–helix fold.[30,31] The hydrophobic dimer interface is located at the top of the hairpin, and the protomers are stacked at a 90° angle. Although the full STAR domain is conserved within the family, the sequence identity in the Qua1 region is only 30% between QkI and its close homolog GLD-1. Interestingly, the hydrophobic "zipper" residues mediating the contact between the two helices within each monomer are highly conserved, while the hydrophobic patch at the top of the hairpin that forms the dimer interface in the GLD-1 and Sam68 Qua1 structures contains significant variation throughout the STAR family (Fig. 1). Furthermore, the Quaking orthologs feature a 3-residue insertion

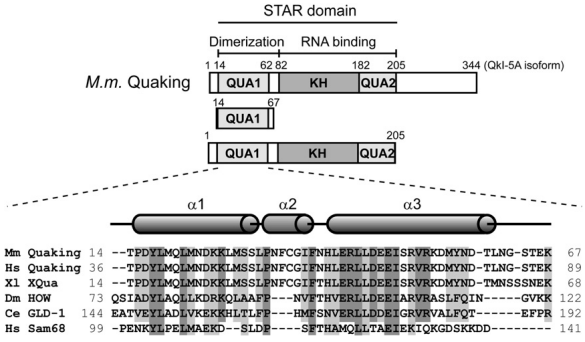


Fig. 1. The STAR family of RNA-binding proteins. Domain structure of QkI and constructs used in this study. The Qua1, KH and Qua2 subdomains are shaded (top). Sequence alignment of the Qua1 domain of representative members of the STAR/GSG protein family (bottom). Identical conserved residues are highlighted in dark gray; similar residues, in light gray. The secondary structure of the Quaking Qua1 domain is shown above. Members of the Quaking subfamily feature a 3-residue insertion between the two large helices that form an additional short helix, instead of a simple turn, which is not present in the Qua1 structures of GLD-1 and Sam68.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

ARTICLE IN PRESS

in this region that is not present in other STAR proteins.

Some studies suggest that STAR protein dimerization is essential for their function, but little is known about how dimerization translates into biological function on a cellular level. In *Caenorhabditis elegans* GLD-1, the KH–Qua2 domain is sufficient for RNA binding, but dimerization enables the bipartite binding mode and deletion of the Qua1 domain reduces the RNA binding affinity by 1 order of magnitude.[27] QkI mutants that abolish dimerization[26] cause an embryonic lethal phenotype in mice, suggesting that homodimerization serves other important functions in addition to facilitating RNA binding. While the splicing factor SF1 lacks the Qua1 dimerization domain and functions as a monomer, Sam68 mutants that impair homodimerization reduce its activity in an alternative splicing assay,[31] suggesting that proper splicing regulation requires dimerization for at least some STAR family members. The role of QkI dimerization for its function as an alternative splicing regulator is unknown.

Here, we present the crystal structure of the QkI Qua1 domain. A fold similar to a helix–turn–helix is expanded by an additional short helix at the top of the hairpin. Based on thermal melting and RNA binding data, we show that failure to dimerize impairs RNA binding. Furthermore, a cell-based splicing assay in mouse C2C12 myoblasts demonstrates that QkI-5 function in alternative splicing requires its ability to homodimerize.

Results

Structure of the QkI Qua1 homodimer

To select a suitable QkI Qua1 construct for crystallography, we cloned 20 C-terminal truncation constructs based on secondary structure prediction (PredictProtein†) and sequence homology to the GLD-1 and Sam68 Qua1 domains. Residue C35 was changed to Ser for the QkI Qua1 and QkI STAR *in vitro* constructs used in this study to avoid nonspecific aggregation.[34] NMR spectra (^{1}H,^{15}N heteronuclear single quantum coherence; Fig. S1) recorded for six constructs, spanning the region from residue 12 or 14 to between residues 55 and 77, indicated that the majority of the protein is structured and that the structured region is present in all constructs analyzed. The only differences in the spectra between the different C-terminal truncations were observed in the unstructured region (7.8–8.5 ppm) of the spectrum, indicating that the C-terminal region comprising residues 55–77 is in fact unstructured in these constructs.

Native QkI Qua1 (14–67) yielded well-diffracting crystals, but molecular replacement (MR) using the homologous proteins GLD-1 and Sam68 failed to produce a satisfactory model for phasing. The phases were ultimately solved using selenomethionine (SeMet) labeling for multiwavelength anomalous dispersion phasing, and the QkI Qua1 structure was solved at 2.1 Å resolution with one homodimer in the asymmetric unit. A summary of the data collection and refinement statistics is given in Table 1.

The QkI Qua1 monomer consists of two α-helices arranged in a hairpin-like geometry resembling the helix–turn–helix fold observed for GLD-1 and Sam68 Qua1. The two antiparallel helices are held together by a hydrophobic "zipper" and a hydrogen bond between the conserved residues Y17 and E48. The QkI Qua1 structure contains an additional short helix, including the 3-residue insertion, in the linker region between the two main helices (Fig. 2a). This helix is not present in the homologous GLD-1 and Sam68 Qua1 domain structures and lies perpendicular to the two main helices. The C-terminal residues 55–61 show weak electron density and do not form a regular secondary structure. No electron density was observed for residues 62–67, consistent with the NMR data indicating that the C-terminal tail is likely unstructured and flexible.

Structural similarity searches with Dali and PISA[35,36] confirm that the most closely related structures in the Protein Data Bank (PDB) are the Qua1 dimerization domain structures of the homologous STAR proteins GLD-1 (PDB ID 3K6T) and Sam68 (PDB ID 2XA6). The overall hairpin-like structure is conserved in all three Qua1 structures, but QkI Qua1 is unique because of the insertion of an additional helix in the turn region of the hairpin, which is present neither in GLD-1 or Sam68 nor in any other of the many helix–turn–helix structures in the PDB.

The Qua1 dimer is formed by stacking the two monomers at the turn region at an angle of 84°. The dimer interface area covers 815 $Å^2$ per monomer, which corresponds to 19% of the total surface area of the monomer. The hydrophobic dimer interface consists of residues L27, L31, F34, I37 and L41 and the highly conserved residues F38 and L44 (Fig. 2). Several hydrogen bonding interactions along the periphery toward the open end of the hairpin reinforce the hydrophobic interface. Residue D24 forms conserved hydrogen bonds to the backbone amides of residues N39′ and H40′, while intermonomer hydrogen bonds between Q20 and H40′ and Y17–R43′ bridge the dimer interface. These interactions are further facilitated by an intramonomer hydrogen bonding network between H40–R43–E47 and Y17–E48–R51, including the highly conserved Y17–E48 monomer "clamp"

† https://www.predictprotein.org/

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

Table 1. Data collection and refinement statistics for QkI Qua1 (C35S)

	SeMet			
	SeMet (peak)	SeMet (remote)	SeMet (inflection)	Merged (solve)
Data collection				
Space group		$P2_12_12_1$		
Cell dimensions (Å)				
a, *b*, *c*		33.98, 36.02, 92.81		
Wavelength (Å)	0.9791358	0.9184018	0.979569	
Resolution (Å)	50–2.15 (2.19–2.15)	50–2.10 (2.14–2.10)	50–2.15 (2.19–2.15)	
Total reflections	50,936?	55,390	51,303	
Unique reflections	11,880 (578)	12,905 (659)	11,969 (593)	
R_{sym} (%)	3.6 (28.8)	3.4 (33.9)	3.6 (32.4)	
$I/\sigma(I)$	35.1 (5.2)	40.0 (4.4)	35.4 (4.4)	
Completeness (%)	99.9 (100)	99.8 (100)	99.9 (100)	
Redundancy	4.3 (4.3)	4.3 (4.2)	4.3 (4.2)	
Wilson *B*-factor	41.06	39.83	42.31	
Refinement				
Resolution (Å)				31.91–2.10 (2.65–2.10)
Unique reflections				7067 (3462)
R_{work}				21.63 (22.03)
R_{free}				25.15 (24.66)
Stereochemistry (%)				
Residues in the favored region				98.02
Residues in the allowed region				1.98
Number of atoms				
Protein				856
Water				44
Ca^{2+}				1
B-factors				
Protein				38.84
Water				36.99
Ca^{2+}				119.53
RMSD				
Bond lengths (Å)				0.008
Bond angles (°)				1.102
Residues not modeled				
A				62–67
B				−2–1 (conformer B), 63–67

between the two long helices that has been shown to be important for monomer stability.[30]

Mutational analysis of the dimer interface and monomer zipper

Key dimer interface and zipper residues in QkI Qua1 were mutated to study their contribution to QkI dimerization. The mutation E48G was chosen in addition to E48A because it is biologically relevant, causing an embryonic lethal phenotype in mice. Residue Tyr 17 was mutated to Phe (Y17F) instead of Ala to selectively disrupt the hydrogen bond to residue E48 without disturbing hydrophobic packing of the aromatic side chain. For the GLD-1 Qua1 domain, the zipper, in addition to the dimer interface residues, proved to be crucial for homodimerization by ensuring the structural integrity of the monomer.[30] The homodimer stability of the set of QkI Qua1 mutants was determined by thermal melting monitored by CD spectroscopy (Fig. 3 and Table 2). The curves for most point mutants show a complete melting profile indicating that these constructs are completely folded at low temperatures. Only the QkI Qua1 L21A and E48G point mutants are still significantly unfolded even at temperatures as low as 5 °C. All curves show a single transition suggesting that QkI Qua1, like GLD-1 Qua1, forms obligate dimers in which dimer dissociation and unfolding of the monomer secondary structure occur simultaneously and a folded monomeric Qua1 domain likely does not exist. The melting temperatures for the set of mutants, as a measure of dimer stability, vary over a large range from <15 °C to >85 °C. Mutants in the monomer zipper and the Y17–E48 clamp have a generally more detrimental effect on dimer stability than residues in the dimer interface. However, the dimer interface mutant L21A confers a large destabilization equaling that of the E48G mutant. Residue L21 plays a dual role in the zipper and dimer interface since it is located at the inner edge of the dimer interface close to the dimer symmetry axis (Fig. 3b) and packs against itself (L21′) in the adjacent protomer. Removal of this pair of L21 side chains leads to a large void, while other dimer interface mutants have a more local effect on the

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

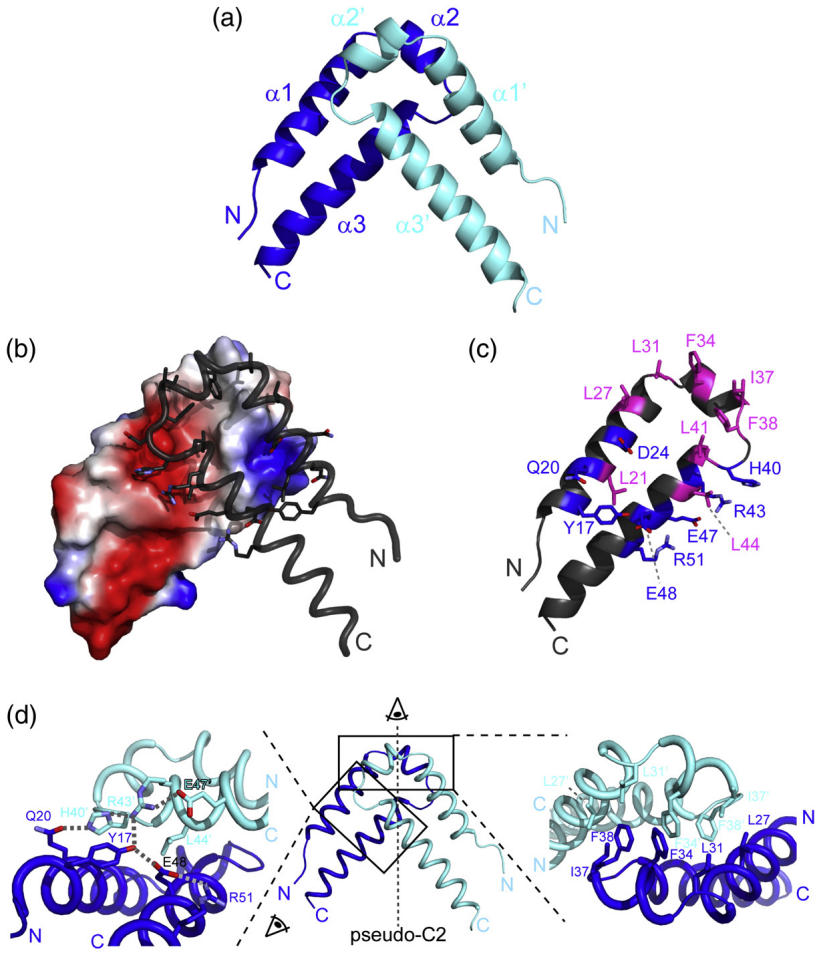


Fig. 2. Overall structure of the QkI Qua1 homodimerization subdomain. (a) Structure of the QkI Qua1 homodimer. Monomers A and B are colored dark blue and light blue, respectively. (b) Dimer interface. Monomer A is shown as electrostatic surface potential, and monomer B is shown as tube with dimer interface residues as sticks. (c) Residues of the QkI Qua1 homodimer interface. The residues that form the hydrophobic core of the interface, containing the conserved Phe34 and Phe38, are highlighted in pink. Residues participating in hydrogen bonds at the edge of the interface are shown in blue. (d) Close-up view on the dimer interface. The prime denotes residues in the other protomer. Side chains of key residues in the dimer interface are shown as sticks, and hydrogen bonds are indicated by gray broken lines.

hydrophobic packing. Furthermore, L21A's bottom face participates in the zipper by packing against I49.

The lack of the Y17–E48 hydrogen bond is associated with a lethal phenotype for mouse Quaking E48G. With the comparison of the E48G and E48A mutants, however, it becomes apparent that the effect of glycine in this position is much greater than simply removing a hydrogen bond.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

With the consideration that glycine residues are known to break α-helices, it is likely that loss of secondary structure in helix α3 in the E48G mutant adds significantly to the dimer destabilization.

To investigate the effect of the point mutations within the Qua1 subdomain in the context of the full QkI STAR domain, we cloned the same point mutations into QkI STAR (1–205) fused to maltose-binding protein (MBP). The C35S mutation that prevents aggregation was also introduced into the QkI STAR mutants. This substitution does not affect the RNA binding affinity of QkI STAR (Table 2). The RNA binding affinity of the mutants was measured using a fluorescence polarization (FP) binding assay.[29] The MBP QkI STAR point mutants were titrated to a constant concentration of the previously identified high-affinity RNA

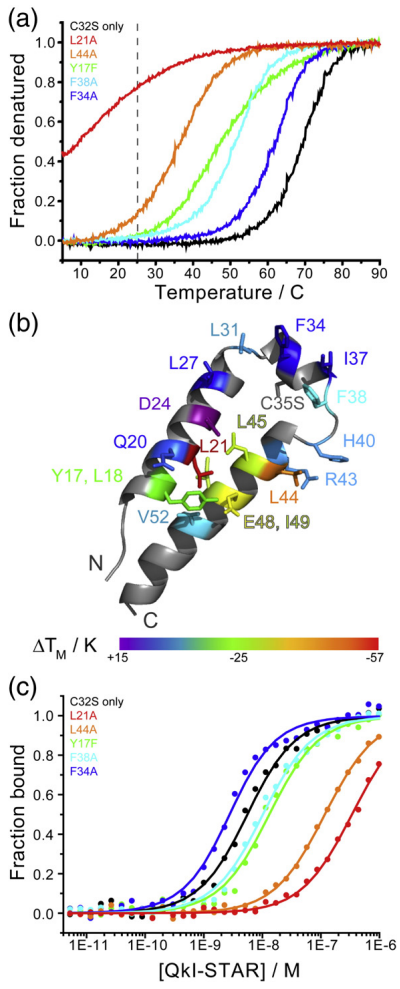


Table 2. Thermal stability of the QkI Qua1 dimer mutants and RNA affinities of QkI STAR mutants

QkI mutant	Qua1 domain (residues 14–67)		Full STAR domain (residues 1–205)	
	T_m (°C)	ΔT_m (K)	K_d (nM)	K_d/K_d (C35S)
wt	n.d.	n.d.	3.4±0.2	0.7
C35S only	69	0	5.1±0.3	1.0
Zipper				
L18A	49	−20	11.6±0.6	2.3
Q20A	65	−4	4.1±0.3	0.8
L44A	37	−32	118±5	23
L45A	40	−29	209±9	41
I49A	38	−31	24±1	4.7
V52A	60	−9	3.8±0.2	0.7
Dimer interface and zipper				
Y17F	45	−24	12.8±0.7	2.5
L21A	<15	<−57	340±30	67
E48A	40	−29	6.7±0.4	1.3
E48G	<15	<−57	22±1	4.3
Dimer interface				
D24A	>85	>+14	4.3±0.2	0.8
L27A	67	−2	3.5±0.2	0.7
L31A	62	−7	n.d.[a]	n.d.[a]
F34A	64	−5	2.3±0.1	0.5
I37A	67	−2	4.4±0.3	0.9
F38A	53	−16	9.8±0.6	1.9
H40A	58	−11	10.8±0.6	2.1
R43A	57	−12	9.0±0.5	1.8

ΔT_m and the K_d ratio are based on the C35S construct since all mutants are within the C35S background.
[a] QkI STAR L31A could not be expressed.

Fig. 3. Qua1 point mutations that destabilize homodimerization also impair RNA binding. All mutants are in the C35S background and are therefore compared to QkI Qua1 C35S. (a) Representative CD melting curves for QkI Qua1 point mutants. Room temperature, at which the RNA binding experiments (c) were performed, is indicated by the gray broken line. (b) ΔT_m mapped onto the monomer structure with residue probes shown as sticks. (c) Representative RNA binding curves for QkI STAR constructs with point mutations in the Qua1 domain. The RNA binding experiments were performed at room temperature [indicated by a gray broken line in (a)]; thus, only those constructs that are significantly destabilized at this temperature are expected to show a significant impact on the RNA binding affinity. The data were fit to the quadratic binding equation for bimolecular binding. The Hill equation does not apply because the RNA concentration cannot be considered "in trace" (probe concentration $\ll 10\ K_d$) for most QkI STAR constructs.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

sequence 5′-UAUUUAAUUUCUUAUC**UACUAA**-UAUCUA-3′ labeled with fluorescein as probe (the high-affinity hexamer consensus is in boldface). The dissociation constants (K_d) for both wild type (wt) and C35S-only mutant agree well with the published values.[29] RNA binding is significantly impaired for only those point mutants that drastically decrease the stability of the QkI Qua1 homodimer, reflected by a melting temperature of the QkI Qua1 construct of $T_m \leq 45\,°C$, namely, Y17F, L21A, L44A, L45A, E48G and I49A (Table 2). Since the RNA binding assay is performed at room temperature, a drastic effect on RNA binding is only observed for those mutants that are already significantly destabilized at this temperature (Fig. 3c). Interestingly, the mutants E48G and L21A that have similar low melting temperatures in the Qua1 construct show quite different RNA binding affinities. Since different length constructs were used in the different assays, it is possible that destabilization caused by any point mutation could be different in the presence of other parts of the protein compared to an isolated domain, especially since the E48G mutation likely combines at least two destabilizing effects: deletion of a stabilizing hydrogen bond and disruption of secondary structure.

All mutant Qua1 and STAR constructs were analyzed on SDS-PAGE and size-exclusion chromatography (data not shown) to ensure intact protein, equal concentrations and homogeneity. All mutants elute in a single peak, with the exception of the D24A mutant in both Qua1 and STAR constructs. The retention time of destabilized mutants slightly increased compared to wt in accordance with the dimer stability measured by CD spectroscopy, but not as much as it would be expected for a globular monomeric form, which might reflect either a fast equilibrium between monomeric and dimeric forms or an increase of hydrodynamic radius due to unfolding. The D24A mutation causes the appearance of an additional higher-molecular-weight species for both Qua1 and STAR domain constructs, indicating that this mutation might cause aggregation, which could explain the surprisingly high melting temperature of the D24A Qua1 construct.

The mutational analysis shows that QkI homodimerization is correlated to its ability to bind RNA, and mutations that disrupt the homodimer also confer a reduced RNA binding affinity. In addition to the QkI Qua1 dimer interface residues themselves, the conserved zipper serves an important function for both homodimer stability and RNA substrate binding. The zipper is well conserved among STAR proteins. It stabilizes the tertiary structure of the α-helical hairpin and ensures the proper display of the dimer interface residues.

A functional Qua1 dimerization domain is required for robust splicing activation by QkI-5 in muscle cells

Many exons that are activated in muscle development are enriched for a motif similar to the QkI consensus downstream of the activated exon, such as the actin capping protein β-subunit, Capzb, which features two ACUAA sites in the 19–80 region of its exon 9 downstream intron.[37] In mouse C2C12 myoblasts, alternative splicing is regulated during differentiation to myotubes,[38,39] and the STAR motif is enriched near exons that change[38] (M. Hall and M. Ares, unpublished results), including Capzb exon 9. An *in vivo* splicing assay using a Capzb exon 9 minigene construct in C2C12 cells (M. Hall and M. Ares, unpublished results) was used to investigate whether a functional Qua1 domain is required for alternative splicing regulation by QkI-5 (Fig. 4 and Fig. S5). Overexpression of wt Myc-tagged QkI-5 (Myc-QkI-5) increases Capzb exon 9 inclusion up to 10-fold in a dose-dependent manner, saturating at a maximum exon 9 inclusion of 77.5%. Four selected mutants (E48G, E48A, Y17F and L21A) that showed significantly impaired homodimerization in the QkI Qua1 thermal melting assay were investigated in the context of full-length Myc-QkI-5 using the Capzb splicing assay in C2C12 cells. Protein expression was reduced for all mutants, with the lowest protein levels being observed for the mutants E48G and L21A. Capzb exon 9 inclusion is significantly reduced by 20–74% for all four mutants compared to wt, with the E48G and L21A mutants showing the most drastic effect. However, since the protein levels for the QkI-5 mutants vary significantly, these data alone do not reveal whether the decrease of exon inclusion is caused by a reduced splicing efficiency of the point mutants or it is simply by the lower protein levels. Varying amounts of wt Myc-QkI-5 expression plasmid were transfected into C2C12 cells along with a constant amount of the Capzb minigene construct. Capzb exon 9 inclusion values were plotted against Myc-QkI-5 expression levels and fit to a simple saturation model, reflecting the per-molecule splicing efficiency. Adding the QkI-5 mutants to the plot allows comparison of their splicing efficiency to wt Myc-QkI-5 (Fig. 4c and Fig. S5). Points falling on the fitted curve represent a splicing efficiency equal to wt protein, while points falling under the curve imply a reduced splicing efficiency and points above the curve imply an enhanced splicing efficiency. Furthermore, the splicing efficiency for each mutant was calculated as the ratio of observed exon inclusion to the expected wt value for the given mutant expression level that was obtained from the wt titration fit. The data points for all four mutants fall under the wt titration curve, and the splicing efficiency is decreased by 17–73% compared to wt,

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

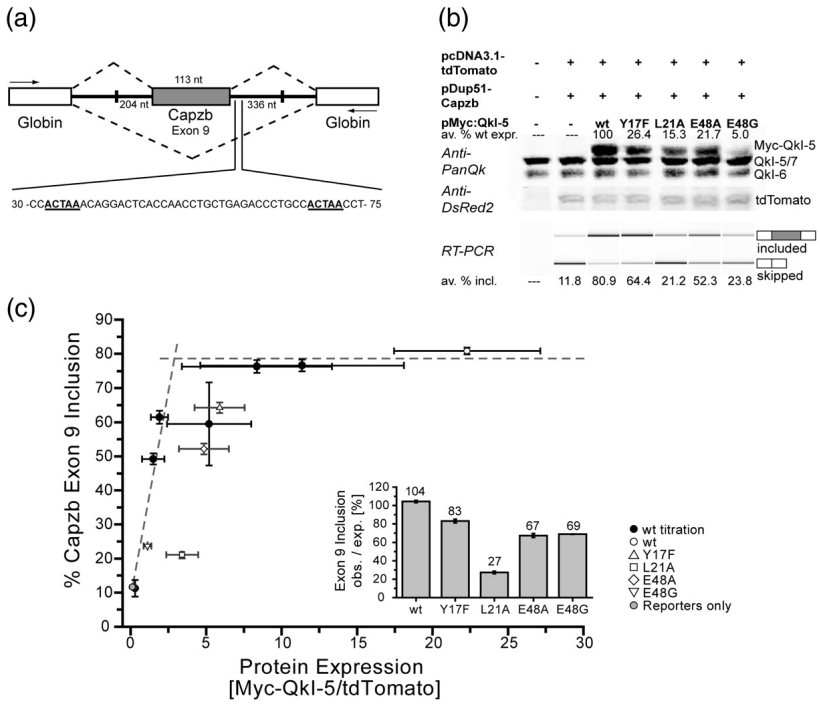


Fig. 4. Dimerization-deficient point mutations in the Qua1 domain reduce QkI-5 splicing activity *in vivo*. (a) Schematic presentation of the Capzb exon9 minigene organization. White boxes represent the globin exons, and the gray box represents the Capzb exon 9. Numbers by the introns flanking the Capzb exon 9 indicate the length of intron sequence that was cloned from the Capzb gene. PCR primer positions are indicated as arrows. The 30- to 75-nt region downstream of the Capzb exon 9 contains two QkI consensus binding sites, shown in boldface and underscored, which are essential for QkI-5 splicing activity in C2C12 myoblast cells. (b) Western blot (top) and RT-PCR analysis of QkI-5 expression levels and minigene splicing patterns from cytoplasmic RNA of C2C12 myoblast cells co-transfected with the Capzb exon 9 minigene construct (pDup51-Capzb), plasmid for expression of wt or mutant QkI-5 protein and a tdTomato expression vector as transfection control. Each QkI-5 construct was assayed in three independent co-transfection experiments. In the Anti-PanQk blot, the two bottom bands represent endogenous QkI, while the top band represents the overexpressed Myc-QkI-5. The Anti-DSRed blot serves as transfection control. Average values for QkI-5 construct expression relative to wt expression are given above the Western blot. The average percentage of Capzb exon 9 inclusion for each construct is given below the RT-PCR gel. Plots of these values including standard deviation can be found in Fig. S5. (c) Capzb exon 9 inclusion plotted *versus* protein expression levels. In order to normalize the splicing efficiencies of the QkI-5 mutants, which show very different protein expression levels, we transfected varying amounts of wt Myc-QkI-5 expression vector in three independent titration experiments. The protein expression level (normalized to tdTomato transfection control) and Capzb minigene splicing efficiency were analyzed as described for the QkI-5 mutants and plotted against each other. Plots of these values including standard deviation can be found in Fig. S5. All individual data points were used to fit a simple saturation model (Fig. S5). This plot shows the wt titration (black circles) and mutant (open symbols) data averaged over the three replicates. Error bars represent one-half standard deviation. The fitted saturation curve is presented as two intersecting broken lines. Points falling below this curve indicate a lower splicing efficiency per QkI-5 molecule compared to wt, while the area above the curve indicates a higher splicing efficiency. The inset shows the observed-to-expected ratio of splicing efficiency (PSI, percent splicing included).

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

indicating that the mutations significantly affect the ability of QkI-5 to activate splicing in C2C12 cells.

Discussion

STAR family members show structural differences in their dimer interface

The STAR domain is highly conserved and defines the family of STAR/GSG proteins in metazoans. The Qua1 subdomains across this family show a high degree of sequence similarity, implying that the structures of the Qua1 homodimers are likely similar; however, the fraction of identical residues between homologs is only 30%. Interestingly, the monomer zipper between the two main α-helices shows high sequence identity, while significant variability is observed within the hydrophobic homodimer interface, including the turn region that links the two main helices.

The Quaking subfamily exhibits a characteristic 3-residue insertion within this turn region that is not present in other paralogs. A homology model of the QkI Qua1 domain using the GLD-1 Qua1 structure[30] as template, consistent with secondary structure prediction from the QkI Qua1 sequence,[1] yielded a model where the two helices in the helix–turn–helix motif overlap exactly and the longer linker simply results in a further extending turn (Fig. S4), prompting the initial attempt using MR to solve the QkI Qua1 structure.

The structure of the QkI Qua1 subdomain shows that the two main helices forming the helical hairpin fold are in fact conserved in the QkI Qua1 structure. Interestingly, the top of the hairpin fold stands out because it does not simply consist of a larger turn lacking secondary structure as predicted but forms a short additional α-helix of about 1.5 turns, which lies perpendicular to the two main helices (Figs. 2 and 5). Formation of this third helix is enabled by the Quaking-subfamily-specific 3-residue insertion that is not present in the GLD-1 and Sam68 Qua1 structures,[30,31] which likely explains why MR using the GLD-1 or Sam68 Qua1 structure was not successful.

The QkI Qua1 dimer interface is enlarged by the inserted helix and covers 815 Å^2 per monomer, compared to 700 Å^2 per monomer for GLD-1[30] and 624 Å^2 per monomer for Sam68.[31] The larger interface allows for a more extended hydrogen bonding network at the edge of the dimer interface that, in contrast to GLD-1 and Sam68, also includes the conserved monomer clamp residues Y17 and E48 (Fig. 2d). The differences in surface area and number of interactions are also reflected in the melting temperatures of the QkI (69 °C), GLD-1 (63 °C)[30] and Sam68 (<50 °C)[31] Qua1 constructs. The quaternary dimer structures of GLD-1 and Sam68 Qua1 domains overlay well in the helical segments for both protomers with the main differences in the turn region (backbone RMSD = 1.1 Å),[31] and both show a 90° angle between the two protomers. In contrast, the protomers in the QkI Qua1 dimer are stacked at an angle of only 84° (Fig. 5c).

Within the QkI Qua1 dimer interface, the conserved F38 residue and F34 are located close to the hairpin turn region and form the center of the hydrophobic dimer interface. F34 stacks face to face with its own counterpart F34′ in the second protomer. The F34–F34′ interaction lies over the dimer symmetry axis and requires one side chain to face toward the center of the dimer interface while the other one has to rotate outward, which breaks the symmetry between the two protomers. Each of the F34 residues also engages in a face-to-edge stacking interaction with F38 within the same protomer (Fig. 2d). A similar Phe π-sandwich structure is also observed in GLD-1 Qua1, but the position of these side chains within the dimer interface is shifted noticeably. Sam68 with only one Phe (F118) per protomer does not use Phe π-stacking interactions at all but achieves hydrophobic packing with Leu (L114) residues. These differences in the dimer interface between STAR family members might serve to enforce homodimerization as opposed to potential heterodimerization of homologous STAR proteins, for example, QkI and Sam68.

Interestingly, the highest degree of sequence identity in the Qua1 subdomain between QkI, GLD-1 and Sam68, as well as other STAR family members, is found in the zipper residues that form the contact between the two main helices in the helical hairpin fold. The zipper is important for dimerization because it stabilizes the tertiary structure of the monomer and ensures a proper display of the dimer interface residues. This stabilizing effect was also observed for the homologous GLD-1 Qua1 homodimer.[30] The hydrophobic dimer interface residues on the other hand show similarity in that the hydrophobic properties are conserved, but only the Pro residue (P32 in QkI), ending helix α1 and inducing the (first) turn, and one Phe residue (F38 in QkI) are invariant throughout the STAR family. Overlaying the QkI, GLD-1 and Sam68 Qua1 structures (Fig. 5b), it becomes apparent that even the invariant dimer interface residues vary in their spatial arrangement resulting in noticeable variations in the details of interface surface shape, even though the general placement of the surface ridges and grooves and the electrostatic surface potential seems similar (Fig. S3).

Dimer stability and its effect on RNA binding *in vitro*

The dimeric STAR proteins bind to bipartite RNA sequences where each monomer is believed to bind

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

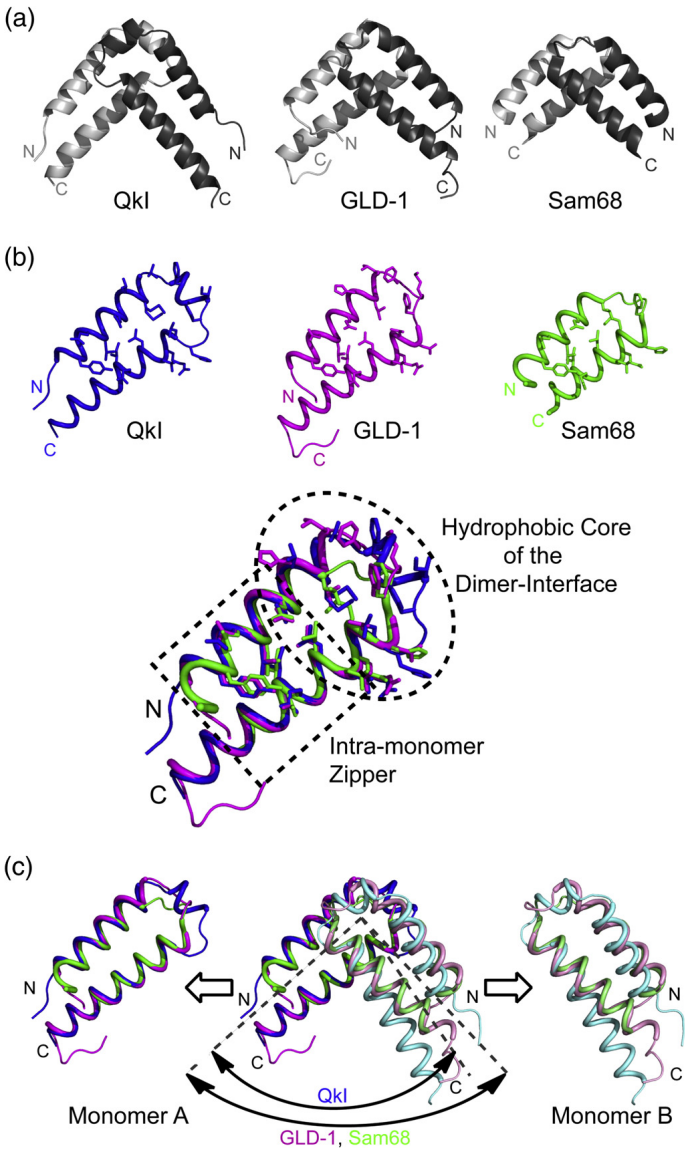


Fig. 5 (*legend on next page*)

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

to one of the two RNA half-sites.[9–11] Homodimerization is essential for enabling this bipartite binding mode. Close proximity of the half-sites, as well as high local concentration, explains the tighter binding of a STAR dimer to a full site *in vitro* compared to binding of a dimerization-deficient monomeric STAR mutant. For GLD-1, the KH–Qua2 subdomain alone is sufficient to bind RNA, but the absence of the Qua1 domain reduces the RNA binding affinity about 1 order of magnitude.[9] This loss in binding affinity can be explained by the fact that each monomeric GLD-1 protein can only bind to one hexamer consensus site, while the dimer binds a bipartite target site that contains at least one high-affinity consensus element.[9,29] Deletion of the QkI Qua1 domain did not abrogate RNA binding despite the lack of dimerization.[26] However, this experiment was of qualitative nature and did not allow quantitating any changes in QkI RNA binding affinity upon Qua1 deletion.

Mutational analysis of the QkI dimer interface and zipper residues shows that, in addition to the dimer interface residues, the hydrophobic zipper that stabilizes the helix–turn–helix fold of the monomer is essential for homodimerization. Zipper mutations have, on average, a larger effect on the stability of the dimer than mutations within the dimer interface itself as they ensure proper tertiary structure and thus present the dimer interface residues in the correct orientation. This effect was also observed for the GLD-1 Qua1 domain[30] and is consistent with the fact that the zipper residues are more highly conserved within the STAR protein family than the dimer interface residues. Mutational analysis of the Qua1 subdomain within the QkI STAR domain context reveals that RNA binding is indeed significantly reduced in those QkI Qua1 mutants that impair dimerization, whether at the dimer interface or in the zipper, linking QkI dimer stability to its biological functions.

What is the role of homodimerization in QkI function *in vivo*?

To test Qk function *in vivo*, we monitored its role as a direct splicing factor in muscle cells (M. P. Hall *et al.*, unpublished results) using a tranfected splicing reporter carrying Capzb exon 9, which contains two QkI consensus sites in its downstream intron. Exon 9 inclusion is promoted by co-transfection of wt Myc-tagged QkI-5 and requires the presence of the QkI binding sequences near exon 9 in the reporter (Hall and Ares, unpublished results), indicating that QkI-5 binding directly to the intron is required to achieve exon inclusion.

The four QkI mutations (Y17F, L21A, E48A and E48G) tested in the *in vivo* splicing assay significantly impact the stability, as well as the ability to homodimerize, of the QkI Qua1 domain and the RNA binding affinity of the QkI STAR domain *in vitro*. Their significantly reduced splicing activity observed in C2C12 myoblast cells correlates well with the *in vitro* data. Interestingly, the reduction of exon inclusion for the mutants is caused by both a diminished splicing efficiency and lower protein expression levels. The observed lower protein levels suggest that the dimerization-deficient mutants with a likely unfolded Qua1 domain are more prone to protein degradation in C2C12 cells. The impaired splicing efficiencies (Fig. 4) for the mutants correlate well with their *in vitro* RNA binding affinities (Table 2). The L21A mutation confers the lowest RNA binding affinity and splicing efficiency, while both mutants E48A and E48G are being equally impaired in both functions and the least destabilizing Y17F mutant confers the least drastic, albeit still significant, loss of function. Overall, the splicing activity of the QkI point mutants correlates well with their ability to homodimerize, indicating that QkI homodimerization is required for QkI splicing activity *in vivo*.

Homodimerization seems to be important for STAR protein biological function beyond a simple increase in RNA binding affinity. Mutants that fail to dimerize show a loss-of-function phenotype, for example, QkI E48G causes embryonic lethality.[26] It is possible that *in vivo* STAR protein dimers bind to two distant sites looping and arranging the RNA to facilitate binding of other RNA-binding proteins of the spliceosome or transcription regulation complexes.[40] The RNA binding and splicing activity of QkI and other STAR proteins is regulated by phosphorylation,[12–14] and it has been proposed that phosphorylation might also be used to disrupt the

Fig. 5. Comparison of the QkI Qua1 structure (PDB ID 4DNN) to the GLD-1 (PDB ID 3K6T) and Sam68 (PDB ID 2XA6). (a) Ribbon presentation. QkI Qua1 features an additional three residues in the turn region that enable to form an additional short helix and expand the surface area of the dimer interface. (b) Comparison of dimer interface and zipper residues (shown as sticks) of QkI, GLD-1 and Sam68 Qua1 monomers and overlay of the three structures. While the monomer zipper is highly conserved, more variation is found in the hydrophobic dimer interface residues. All three proteins use stacking of conserved Phe residues as the core of the dimer interface, but the exact position of these residues varies to some degree. The additional helix in the QkI structure broadens the dimer interface significantly and allows for more hydrogen bond interactions across the edge of the interface (see also Fig. 2). (c) While the two main helices in each monomer overlay precisely for all three Qua1 domains, the two QkI Qua1 monomers are stacked at a narrower angle compared to the GLD-1 and Sam68 homodimers.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

homodimer and switch to a monomeric, inactive form of the protein.[31] Several of the tyrosine residues implicated in phosphorylation are conserved throughout the STAR protein family, suggesting that regulation of dimeric *versus* monomeric forms by phosphorylation might be used as a general switch. It is also possible that other proteins recognize and bind to a STAR protein dimer but not a monomer, or *vice versa*.

QkI exists in multiple alternatively spliced Quaking isoforms, the most prominent ones being QkI-5, QkI-6 and QkI-7. All isoforms are identical in the 1–311 region including the full STAR domain. The C-terminal 8–30 residues differ between isoforms and convey specific functions. QkI-5 (312–341) contains a nuclear localization signal. The 14 C-terminal residues (312–325) of the predominantly cytoplasmic QkI-7 carry an apoptosis-inducing signal. QkI-6 uniquely features an 8-residue tail (residues 312–319) and is found both in the nucleus and in the cytoplasm. Heterodimerization between isoforms can occur,[16] but it is not well understood under what circumstances these isoforms can heterodimerize in cells and how this affects QkI function. When co-transfected and overexpressed in HeLa cells, QkI-5 forms heterodimers with QkI-6 and QkI-7 and self-homodimers. Furthermore, QkI-5 was able to shuttle QkI-7 into the nucleus and suppress the induction of apoptosis by QkI-7. These results are a first indication that heterodimerization between different isoforms might be used to control protein localization and that the balance in the QkI isoforms is critical for the normal function of the QkI proteins and cell viability.

In addition to multiple Quaking isoforms, there are other STAR proteins such as Sam68 present in both the cytoplasm and the nucleus. With the consideration that the topologies of the Qua1 dimer interfaces in the STAR family are similar but not identical, it could be possible that STAR proteins form inter-paralog heterodimers. QkI has been shown to interact with co-transfected GLD-1 but not Sam68 in an immunoprecipitation assay in HeLa cells.[25] The structural differences between STAR family members within the dimer interface likely serve to specify and select dimerization partners. STAR protein heterodimers would be expected to form with a lower affinity, if at all, and thus might be controlled by protein concentrations in the cell, which would add another level of functional regulation. To shed light on these questions, we need further studies addressing the role of STAR protein dimerization *in vivo*, both between isoforms and different family members, and how dimerization affects interaction with RNA and other regulatory proteins. In summary, we have determined the structure of the QkI homodimerization domain and correlated the formation of dimers to RNA binding *in vitro* and to activation of splicing in cells, thus providing insights into how STAR proteins function in regulation of gene expression.

Materials and Methods

Protein expression and purification

MBP-tagged pMAL-QkI-STAR(1–205) was expressed in *Escherichia coli* strain JM109 (New England Biolabs) and purified via affinity chromatography using amylose resin (New England Biolabs), followed by a HiTrap Q anion-exchange column (GE Healthcare), as described previously,[9] but omitting the cation-exchange chromatography step. Purified protein was stored at −80 °C. Cys35 was changed to Ser for consistency with the Qua1 constructs, and point mutants were generated in the C35S background using QuikChange (Agilent).

QkI Qua1 C35S constructs were cloned into pET22b(+) (Novagen), which results in a C-terminal hexahistidine (His) tag, or a pET22b(+)-derived expression vector, introducing an N-terminal His tag and a thrombin cleavage site (MHHHHHLVPRGS). For a complete list of constructs cloned and tested for expression, see Table S1. All QkI Qua1 constructs, including point mutants, were expressed in *E. coli* BL21-Gold (DE3) cells (Stratagene).

The N-terminally His-tagged QkI Qua1 (14–67) was used for all structural studies and dimer stability experiments. Single point mutations in the QkI Qua1 and QkI STAR context were introduced in the C35S background using QuikChange (Agilent). Unlabeled QkI Qua1 protein was expressed in LB media with 100 mg/l ampicillin for 3 h at 37 °C with 1 mM IPTG, added at a cell density of OD_{600} = 0.6–0.8. Six constructs were labeled with ^{15}N for NMR spectroscopy and were expressed for 6 h at 37 °C in M9 minimal media containing $(^{15}NH_4)_2SO_4$ as the sole nitrogen source, unlabeled glucose and supplemental trace metals and vitamins. SeMet-labeled QkI Qua1 was expressed for 6 h at 37 °C in M9 minimal media containing NH_4Cl as nitrogen source, unlabeled glucose, supplemental trace metals and vitamins and was supplemented with the amino acids SeMet, Lys, Thr, Phe, Leu, Ile, Val and thiamine.

Unlabeled and ^{15}N-labeled QkI Qua1 samples for thermal melting experiments were purified in 20 mM sodium phosphate (pH7.5) buffer, while the SeMet sample for crystallography was purified in 20 mM Tris–HCl (pH7.5) buffer. Cells were lysed by sonication in buffers containing 15 mM imidazole and Complete ethylenediaminetetraacetic acid-free protease inhibitor cocktail (Roche). Soluble protein was purified in two steps by Ni-affinity chromatography followed by anion-exchange chromatography. The lysate was passed over a 5-ml HisTrap HP column (GE Healthcare), washed with 25 ml of purification buffer with 15 mM imidazole, followed by 15 ml of purification buffer with 20 mM imidazole and finally eluted in a single step with 25 ml of purification buffer with 200 mM imidazole. Pure unlabeled protein fractions were dialyzed against 20 mM sodium phosphate buffer (pH7.5) at 4 °C overnight. For SeMet-labeled protein fractions, thrombin (5 U/mg of protein) was added, and the protein was dialyzed against 20 mM

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

Tris–HCl (pH7.5) with 0.02% NaN_3 at room temperature for 24 h. Complete thrombin cleavage was confirmed by SDS-*N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine-PAGE. The protein was further purified by anion-exchange chromatography on a 5-ml HiTrap Q HP column (GE Healthcare), where SeMet-labeled QkI Qua1 eluted with a 120-ml gradient from 0 to 1 M NaCl, while unlabeled, still His-tagged, QkI Qua1 constructs were found in the flow through fraction. Unlabeled samples for thermal melting experiments were dialyzed against 20 mM sodium phosphate (pH7.5) and 0.02% NaN_3, while SeMet-labeled protein for crystallization was dialyzed against 20 mM Tris–HCl (pH7.5) and 0.02% NaN_3. All protein samples were concentrated in Centriprep concentrators (3-kDa cutoff; Millipore), and purified protein was stored at 4 °C. Protein concentration was determined by measuring the UV absorption at 280 nm. All protein samples (50 μM for QkI Qua1 and 30 μM for QkI STAR) were run on an SDS-PAGE gel to ensure sample integrity and equal concentrations in the assays.

NMR spectroscopy

The following constructs were labeled with ^{15}N to study by NMR spectroscopy: QkI Qua1 (14–77)-C-His_6, QkI Qua1 (14–77)-N-His_6, QkI Qua1 (12–77)-N-His_6, QkI Qua1 (14–63)-C-His_6, QkI Qua1 (12–60)-C-His_6 and QkI Qua1 (14–55)-C-His_6. NMR samples were prepared in 20 mM sodium phosphate (pH7.5) and 0.02% NaN_3 containing 10% D_2O/90% H_2O with a final protein concentration of 0.5 mM. NMR spectra were recorded at 35 °C on a 750-MHz Bruker Avance spectrometer with 5 mm TXI/HCN triple resonance probes.

Analytical size-exclusion chromatography

All QkI Qua1 and STAR mutants were subjected to analytical size-exclusion chromatography at room temperature to ensure sample integrity and multimerization status. QkI Qua1 proteins were loaded onto a Superdex 75 10/300 gel-filtration column (GE Healthcare) and eluted in 20 mM sodium phosphate buffer (pH7.5). QkI STAR proteins were loaded onto a Superdex 200 10/300 gel-filtration column (GE Healthcare) and eluted in 50 mM Tris–HCl (pH7.9), 20 mM NaCl and 2 mM DTT. Gel-filtration standard (Bio-Rad) with thyroglobulin (bovine, 670,000 Da), γ-globulin (bovine, 158,000 Da), ovalbumin (chicken, 44,000 Da), myoglobin (horse, 17,000 Da) and vitamin B12 (1350 Da) was used to calibrate molecular weight.

Protein crystallography

Protein crystallization

SeMet-QkI-Qua1(14–67) was crystallized by vapor diffusion in 24-well sitting-drop plates, 2 μl drop volume, at 22 °C. Crystals of SeMet-QkI-Qua1 were obtained by mixing 1 μl of protein (5 mg/ml) in 20 mM Tris–HCl (pH7.5), 0.02% NaN_3 with 1 μl of 0.1 M sodium cacodylate (pH6.5), 0.2 M calcium acetate and 29% (v/v) polyethylene glycol 600. Crystals were flash-frozen in liquid nitrogen without additional cryoprotectant.

Data collection and processing

Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) BL 11-1 processed with HKL2000.[41] MR was attempted using the PHENIX[42] AutoSol. With the use of the Sam68 Qua1 (PDB ID 2XA6) monomer or dimer, no solution was obtained. With the use of GLD-1 Qua1 monomer or dimer, solutions in space group $P2_12_12_1$ were found. However, the electron density maps were poorly defined, especially in the turn region. The log-likelihood gains of LLG=34 for GLD-1 Qua1 dimer and LLG=79 for GLD-1 Qua1 monomer as model were both lower than the acceptable value of 100‡, indicating a poor quality of these solutions. The PHENIX package[42] was also used for multiwavelength anomalous dispersion phasing of SeMet-QkI-Qua1. Initial model building after density modification was carried out in PHENIX AutoBuild, followed by alternating rounds of refinement with PHENIX Refine and manual model building and evaluation in Coot.[43] In later rounds of refinement, TLS refinement using each monomer as an independent TLS group was performed in addition to coordinate and atomic *B*-factor refinement. Adding NCS restrains during refinement did not improve the statistic and was therefore not used. Iterative rebuilding and refinement converged on a final model with R=21.6% and R_{free}=25.2% and good stereochemistry (98% of residues in the most favored region/2% of residues in the additionally allowed region). For complete crystallographic statistics, see Table 1.

Thermal melting experiments by CD spectroscopy

Thermal melting experiments were performed using a Jasco J-815 CD spectrometer with Jasco PFD-425S/15 temperature unit and a 1-mm-quartz cuvette. The samples consist of 50 μM protein in 20 mM sodium phosphate (pH7.5). Melting curves were recorded between 5 °C and 90 °C in 0.2 °C steps with a heating rate of 1 °C/min monitoring the signal at 222 nm. The raw CD data were normalized to fraction denatured and corrected for the slopes for folded and unfolded signals according to:

$$F_u(T) = \frac{\text{mdeg}-\text{min}-s_f\cdot(T-T_{min})}{\text{max}-\text{min}+s_u\cdot(T-T_{max})-s_f\cdot(T-T_{min})}$$

with F_u=fraction unfolded, T=temperature in degrees Celsius, mdeg=CD signal in millidegrees, min=minimum CD signal (at 5 °C), max=maximum CD signal (at 90 °C), T_{min}=5 °C, T_{max}=90 °C, s_f=slope for folded state and s_u=slope for unfolded state.

T_m was determined as the inflection point of the normalized melting curves.

Samples were run on an SDS-PAGE gel after the melting experiment to ensure sample integrity and equal concentrations in the assay.

Quantitative RNA binding assay using FP

Quantitative analysis of QkI STAR–RNA interaction was performed using an FP assay as described previously.[9] A

‡ http://www.phenix-online.org/documentation/

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

constant concentration of 1 nM fluorescein (Fluo)-labeled RNA (5′-Fluo-AUUUAAUUUCUUAUCUACUAAUAU-CUA-3′) was equilibrated with varying concentrations of QkI STAR in a total volume of 100 μl in 96-well opaque 200-μl Fluotrak plates (Greiner) for 3 h at room temperature. The blank-corrected polarization (mP) was measured using an EnVision 2104 Multilabel Reader (Perkin Elmer). Each experiment was performed in triplicate. The apparent equilibrium dissociation constant K_d was determined by fitting the FP data from all repeats globally using the quadratic binding equation for a bimolecular complex, keeping the RNA concentration R_0 constant:

$$F_B(P_0) = \frac{mP - mP_{min}}{mP_{max} - mP_{min}} = \frac{P_0 + R_0 + K_d - \sqrt{(P_0 + R_0 + K_d)^2 - 4 \cdot P_0 \cdot R_0}}{2 \cdot R_0}$$

with F_B=RNA fraction bound, mP=polarization value in millipolarization units, mP_{max}=maximum mP value at saturation, mP_{min}=minimum mP value in the absence of protein, P_0=protein concentration, R_0=labeled RNA concentration and K_d=apparent equilibrium dissociation constant.

The Hill equation does not apply because the RNA concentration cannot be considered "in trace" (probe concentration $\ll 10\ K_d$) for most QkI STAR constructs. The data points shown in Fig. 3 are the average of the normalized F_B values.

QkI splicing assay

Plasmid construction

The pDup51 vector was used as a splicing reporter as previously described.[44] Exon 9 of Capzb, including 204 nt 5′ of the exon and 336 nt 3′ of the exon, was PCR-amplified and cloned into the ApaI and BglII sites in pDup51. The pMyc:Qk5 vector was a kind gift from Sean Ryder. Mutations were made by site-directed mutagenesis and verified by sequencing. The pcDNA3.1-tdTomato vector was a kind gift from Rohinton Kamakaka.

Cell culture and transfections

Mouse C2C12 myoblasts (American Type Culture Collection) were cultured in 1× Dulbecco's modified Eagle's media (Gibco/Invitrogen) supplemented with 10% fetal bovine serum (Gibco/Invitrogen) under sub-confluent conditions as recommended by the distributor. Cells were co-transfected in 6-well plates with 1 μg pcDNA3.1-tdTomato, 0.5 μg pDup51-CapzbEx9 and either 2 μg of mutant and wt pMyc:Qk5 (for mutant splicing assays) or varying amounts of wt pMyc:Qk5 (for Qk5 titrations). Transfections were performed using Lipofectamine2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested and RNA was extracted using TriZOL Reagent (Invitrogen), while protein was extracted by lysis in Laemmli buffer.

RT-PCR and bioanalyzer analysis

Reverse transcription (RT) was carried out on 2 μg total RNA using Superscript II (Invitrogen) according to the manufacturer's instructions. Equal volumes of resulting cDNA were used as PCR template for each sample. Reactions were carried out for 30 cycles with primers directed toward the flanking Dup51 reporter exons and visualized by ethidium bromide staining of agarose gel analysis. To measure percent inclusion for splicing, we purified PCR products then separated samples on the Agilent Bioanalyzer. This reported the size and concentration of both skipped and Capzb exon 9 included PCR products.

Western blotting

Equal volumes of cell lysate were loaded onto SDS-polyacrylamide gels and electrophoresed. Gels were transferred to pure nitrocellulose then blocked for 1 h at room temperature. Primary antibodies directed toward Pan-Quaking (Antibodies Incorporated) and DsRed2 (Clontech) were used at respective dilutions of 1:2000 and 1:500 and incubated with membranes overnight at 4 °C. Both IRDye 800CW conjugated goat anti-mouse IgG2b (Li-Cor) and IRDye 680LT conjugated goat anti-mouse IgG1 secondary (Li-Cor) antibodies were diluted to 1:20,000 and simultaneously used to probe the membrane for 60 min at room temperature. The resulting signals were visualized on a Li-Cor Odyssey Infrared Imager. Bands were quantified by corresponding infrared counts in each signal wavelength, and ectopic Myc-tagged Qk5 expression was normalized to the expression of the tdTomato transfection control.

Calculation of normalized splicing efficiency

wt titration data were plotted as Capzb exon 9 inclusion *versus* protein expression and fitted to a simple saturation model keeping the y intercept fixed to the value determined from the average of the reporter only data: $y=(21\pm3)\cdot x+11.8$ for $x\leq3.14$ and $y=(77\pm1)$ for $x>3.14$.

Mutant splicing efficiency relative to wt was calculated as the ratio of the averaged exon inclusion for the mutants divided by the expected exon inclusion for wt protein at the given mutant protein level according to the fit above.

Accession numbers

Coordinates and structure factors for QkI Qua1 (14–67) C35S have been deposited in the PDB with the accession code 4DNN.

Acknowledgements

We thank Ian Wilson for the use of X-ray equipment, Kurt Wüthrich for the use of CD spectrometer and Steven Brown for the use of the fluorescence plate reader. X-ray data were collected at SSRL, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the U.S. Department of Energy, Office of Basic

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

ARTICLE IN PRESS

Energy Sciences, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. This work was supported by the National Institutes of Health grant GM053320 (J.R.W.), a Training Grant 5T32GM008646-14 (W.S.F.) and a postdoctoral fellowship by the Deutsche Forschungsgemeinschaft (C.B.).

Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2012.08.027

References

1. Artzt, K. & Wu, J. I. (2010). STAR trek: an introduction to STAR family proteins and review of quaking (QKI). *Adv. Exp. Med. Biol.* **693**, 1–24.
2. Vernet, C. & Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484.
3. Lobbardi, R., Lambert, G., Zhao, J., Geisler, R., Kim, H. R. & Rosa, F. M. (2011). Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Development*, **138**, 1783–1794.
4. Noveroske, J. K., Lai, L., Gaussin, V., Northrop, J. L., Nakamura, H., Hirschi, K. K. & Justice, M. J. (2002). Quaking is essential for blood vessel development. *Genesis*, **32**, 218–230.
5. Wu, J. I., Reed, R. B., Grabowski, P. J. & Artzt, K. (2002). Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl Acad. Sci. USA*, **99**, 4233–4238.
6. Saccomanno, L., Loushin, C., Jan, E., Punkay, E., Artzt, K. & Goodwin, E. B. (1999). The STAR protein QKI-6 is a translational repressor. *Proc. Natl Acad. Sci. USA*, **96**, 12605–12610.
7. Larocque, D., Galarneau, A., Liu, H. N., Scott, M., Almazan, G. & Richard, S. (2005). Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat. Neurosci.* **8**, 27–33.
8. Larocque, D., Pilotte, J., Chen, T., Cloutier, F., Massie, B., Pedraza, L. *et al.* (2002). Nuclear retention of MBP mRNAs in the quaking viable mice. *Neuron*, **36**, 815–829.
9. Ryder, S. P. & Williamson, J. R. (2004). Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination. *RNA*, **10**, 1449–1458.
10. Galarneau, A. & Richard, S. (2005). Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* **12**, 691–698.
11. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P. *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
12. Zhang, Y., Lu, Z., Ku, L., Chen, Y., Wang, H. & Feng, Y. (2003). Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism. *EMBO J.* **22**, 1801–1810.
13. Matter, N., Herrlich, P. & Konig, H. (2002). Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*, **420**, 691–695.
14. Nir, R., Grossman, R., Paroush, Z. & Volk, T. (2012). Phosphorylation of the *Drosophila melanogaster* RNA-binding protein HOW by MAPK/ERK enhances its dimerization and activity. *PLoS Genet.* **8**, e1002632.
15. Ebersole, T. A., Chen, Q., Justice, M. J. & Artzt, K. (1996). The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nat. Genet.* **12**, 260–265.
16. Pilotte, J., Larocque, D. & Richard, S. (2001). Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes Dev.* **15**, 845–858.
17. Wu, J., Zhou, L., Tonissen, K., Tee, R. & Artzt, K. (1999). The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. *J. Biol. Chem.* **274**, 29202–29210.
18. Haroutunian, V., Katsel, P., Dracheva, S. & Davis, K. L. (2006). The human homolog of the QKI gene affected in the severe dysmyelination "quaking" mouse phenotype: downregulated in multiple brain regions in schizophrenia. *Am. J. Psychiatry*, **163**, 1834–1837.
19. Noveroske, J. K., Hardy, R., Dapper, J. D., Vogel, H. & Justice, M. J. (2005). A new ENU-induced allele of mouse quaking causes severe CNS dysmyelination. *Mamm. Genome*, **16**, 672–682.
20. Ebersole, T., Rho, O. & Artzt, K. (1992). The proximal end of mouse chromosome 17: new molecular markers identify a deletion associated with quaking-viable. *Genetics*, **131**, 183–190.
21. Zhao, L., Mandler, M. D., Yi, H. & Feng, Y. (2010). Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proc. Natl Acad. Sci. USA*, **107**, 19061–19066.
22. Zearfoss, N. R., Clingman, C. C., Farley, B. M., McCoig, L. M. & Ryder, S. P. (2011). Quaking regulates Hnrnpa1 expression through its 3′ UTR in oligodendrocyte precursor cells. *PLoS Genet.* **7**, e1001269.
23. Novikov, L., Park, J. W., Chen, H., Klerman, H., Jalloh, A. S. & Gamble, M. J. (2011). QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Mol. Cell. Biol.* **31**, 4244–4255.
24. Yang, G., Fu, H., Zhang, J., Lu, X., Yu, F., Jin, L. *et al.* (2010). RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology*, **138**, 231–240; e1–e5.
25. Chen, T., Damaj, B. B., Herrera, C., Lasko, P. & Richard, S. (1997). Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol.* **17**, 5707–5718.
26. Chen, T. & Richard, S. (1998). Structure-function analysis of Qk1: a lethal point mutation in mouse

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

quaking prevents homodimerization. *Mol. Cell. Biol.* **18**, 4863–4871.
27. Ryder, S. P., Frater, L. A., Abramovitz, D. L., Goodwin, E. B. & Williamson, J. R. (2004). RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nat. Struct. Mol. Biol.* **11**, 20–28.
28. Galarneau, A. & Richard, S. (2009). The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs. *BMC Mol. Biol.* **10**, 47.
29. Carmel, A. B., Wu, J., Lehmann-Blount, K. A. & Williamson, J. R. (2010). High-affinity consensus binding of target RNAs by the STAR/GSG proteins GLD-1, STAR-2 and Quaking. *BMC Mol. Biol.* **11**, 48.
30. Beuck, C., Szymczyna, B. R., Kerkow, D. E., Carmel, A. B., Columbus, L., Stanfield, R. L. & Williamson, J. R. (2010). Structure of the GLD-1 homodimerization domain: insights into STAR protein-mediated translational regulation. *Structure*, **18**, 377–389.
31. Meyer, N. H., Tripsianes, K., Vincendeau, M., Madl, T., Kateb, F., Brack-Werner, R. & Sattler, M. (2010). Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. *J. Biol. Chem.* **285**, 28893–28901.
32. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R. *et al.* (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, **294**, 1098–1102.
33. Maguire, M. L., Guler-Gane, G., Nietlispach, D., Raine, A. R., Zorn, A. M., Standart, N. & Broadhurst, R. W. (2005). Solution structure and backbone dynamics of the KH-QUA2 region of the *Xenopus* STAR/GSG quaking protein. *J. Mol. Biol.* **348**, 265–279.
34. Ali, M. & Broadhurst, R. W. (2008). Structural studies of Quaking protein Qua1 domain. In International Conference on Magnetic Resonance in Biological Systems, vol. XXIII, pp. 193, San Diego, CA.
35. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v. 3. *Bioinformatics*, **24**, 2780–2781.
36. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
37. Sugnet, C. W., Srinivasan, K., Clark, T. A., O'Brien, G., Cline, M. S., Wang, H. *et al.* (2006). Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4.
38. Bland, C. S., Wang, E. T., Vu, A., David, M. P., Castle, J. C., Johnson, J. M. *et al.* (2010). Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res.* **38**, 7651–7664.
39. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J. *et al.* (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
40. Jungkamp, A. C., Stoeckius, M., Mecenas, D., Grun, D., Mastrobuoni, G., Kempa, S. & Rajewsky, N. (2011). *In vivo* and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell*, **44**, 828–840.
41. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr., Part A*, **276**, 307–326.
42. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
43. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
44. Dominski, Z. & Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* **11**, 6075–6083.

Please cite this article as: Beuck, C. *et al.*, Structural Analysis of the Quaking Homodimerization Interface, *J. Mol. Biol.* (2012), http://dx.doi.org/10.1016/j.jmb.2012.08.027

Supplementary Material

Structural analysis of the Quaking homodimerization interface

Christine Beuck, Song Qu, William Fagg, Manuel Ares, Jr. and James R. Williamson

Table S1: Qki-Qua1 C35S constructs cloned.

construct	tag	expression	solubility	NMR	X-ray structure & point mutants
12-55	N-H_6	+	-		
12-55	C-H_6	+	+		
12-60	C-H_6	+	+	+	
12-63	N-H_6	+	-		
12-67	N-H_6	+	-		
12-71	N-H_6	+	+		
12-77	N-H_6	+	+	+	
14-55	N-H_6	-	-		
14-55	C-H_6	+	+	+	
14-60	N-H_6	+	+		
14-60	C-H_6	+	+		
14-63	N-H_6	+	+		
14-63	C-H_6	+	+	+	
14-67	N-H_6	+	+		+
14-67	C-H_6	+	+		
14-71	N-H_6	+	+		
14-71	C-H_6	+	+		
14-77	N-H_6	+	+	+	
14-77	C-H_6	+	+	+	

Figure S1: $^{1}H,^{15}N$-HSQC NMR spectra of selected QkI Qua1 constructs. Each peak represents a N-H bond (backbone amides and some side chains). All constructs show good dispersion and sharp lines indicating that all of them are folded. Differences between te different length constructs manifest in the unstructured region of the spectrum (7.8-8.5 ppm, marked by a grey box), while the well dispersed structured residues are common in all spectra. .This indicates that the region between residues 55 and 77 is flexible and does not adopt regular secondary structure.

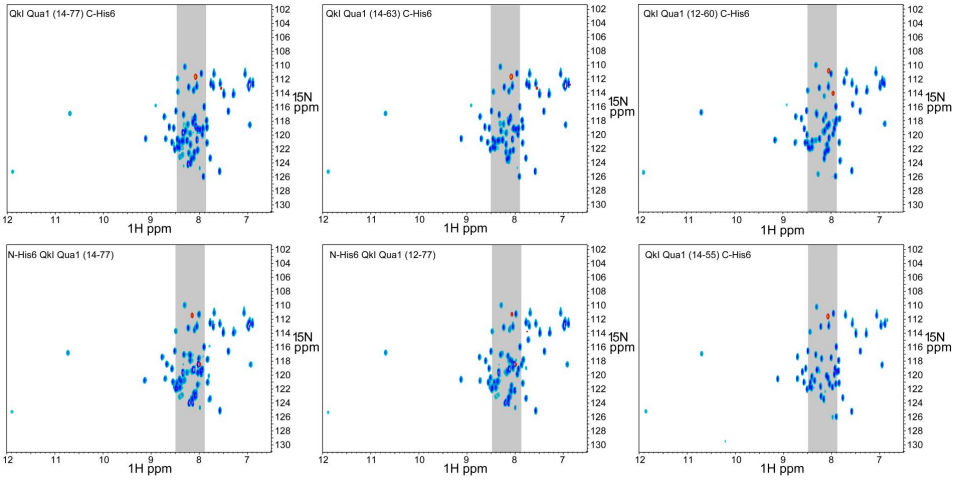


Figure S2: QkI Qua1 structure. (A) QkI Qua1 dimer interface folded open. Convex areas are highlighted in red and concave ones in white. Surface not belonging to the dimer interface is shown in grey. (B) Overlay of the two monomers (chains A and B) that form one dimer in the asymmetric unit. Dimer interface and zipper residues are shown as sticks. The highlighted F34 residue stacks with its counterpart F34' in the dimer interface and lies on top of the pseudo-C2 symmetry axis that defines the symmetry between the two monomers. To achieve this stacking interaction, the F34 side chain in one monomer needs to rotate outwards, while the other on faces inwards towards the interface, breaking the symmetry between the two protomers.

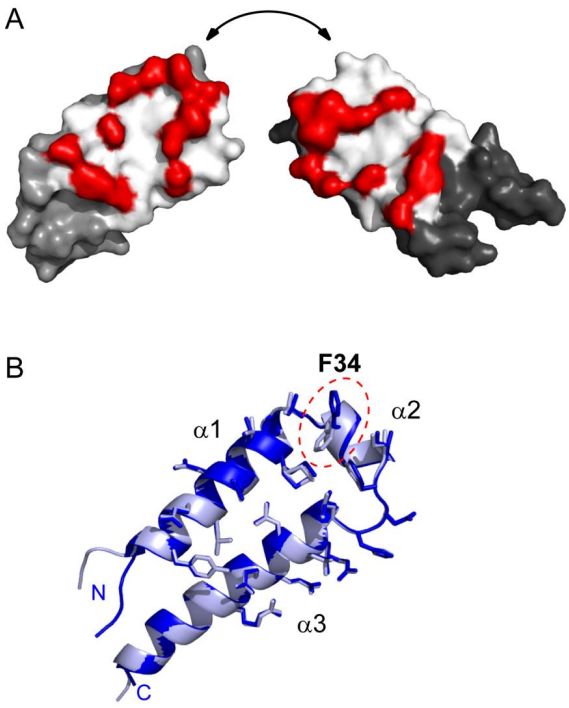


Figure S3: Dimer interface surface of Qkl, GLD1 and Sam68 Qua1 subdomain.

Electrostatic surface potential for positive (blue) and negative (red) charge generated with PyMol.

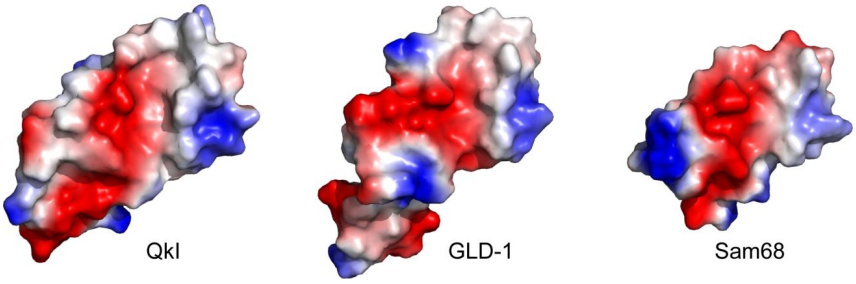


Figure S4: QkI Qua1 homology model from GLD-1 Qua1 template. GLD-1 Qua1 crystal structure (pdb 3K6T) shown in magenta, QkI qua1 homology model in blue. The two main helices of the helix-turn-helix fold are matched well, but the Quaking-specific three residue insertion is not recognized as an additional helix but modeled as an extended loop.

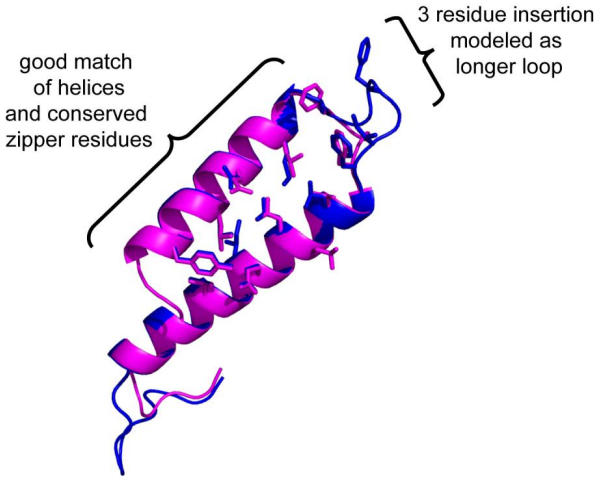
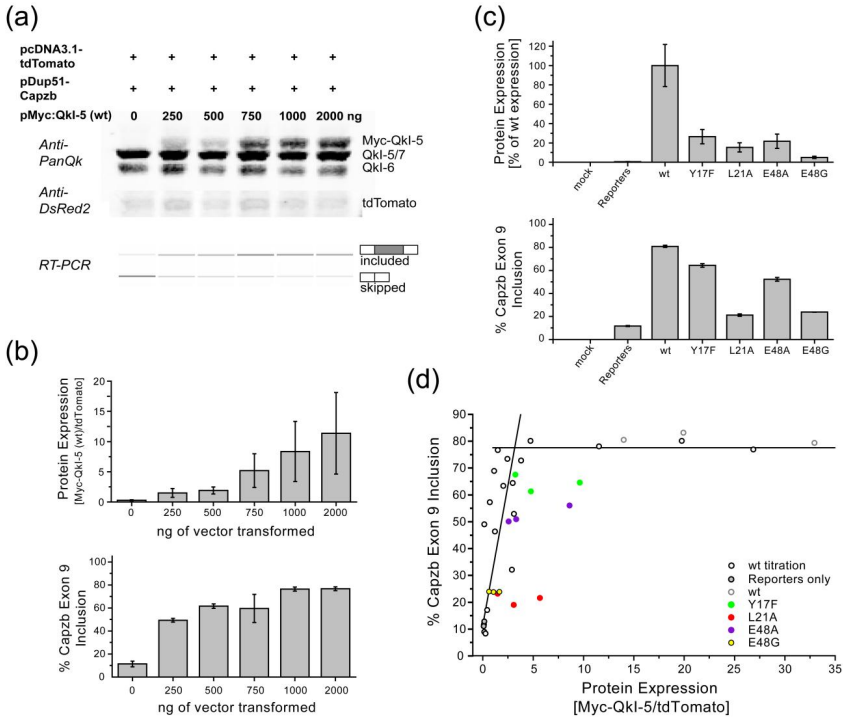


Figure S5: *In vivo* splicing activity of QkI-5 in mouse C2C12 myoblast cells.

(a) Representative QkI-5 (wt) titration, western blot with Anti-Pan-Qk and Anti-DSRed (top) and RT-PCR gel (bottom). Varying amounts of pMyc-QkI-5 plasmid were transformed into C2C12 cells. Protein expression levels and Capzb Exon 9 inclusion were determined. All experiments were performed in triplicate. (b) Averaged protein expression and Capzb Exon 9 inclusion plotted versus amount of vector transformed. Error bars represent 0.5 STD. (c) Mutant protein expression levels compared to wt, and absolute Capzb Exon 9 splicing efficiency. (d) Plot of splicing efficiency vs. protein expression. All single points from the triple replicate experiments are shown without averaging. The un-averaged wt titration data were fit to a simple saturation model. The plot showing the averaged values for wt and mutants is shown in Figure 4.



CHAPTER 3

QUAKING AND PTB CONTROL OVERLAPPING SPLICING REGULATORY NETWORKS DURING MUSCLE CELL DIFFERENTIATION

Reprinted with permission from

RNA *19* 1-12 (August 2012)

Copyright © 2013 by RNA Society

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press



RNA
A PUBLICATION OF THE RNA SOCIETY

Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation

Megan P. Hall, Roland J. Nagel, W. Samuel Fagg, et al.

RNA published online March 22, 2013
Access the most recent version at doi:10.1261/rna.038422.113

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2013/03/14/rna.038422.113.DC1.html
P<P	Published online March 22, 2013 in advance of the print journal.
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**



Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

Copyright © 2013 RNA Society

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation

MEGAN P. HALL, ROLAND J. NAGEL, W. SAMUEL FAGG, LILY SHIUE, MELISSA S. CLINE, RHONDA J. PERRIMAN, JOHN PAUL DONOHUE, and MANUEL ARES, JR.[1]

RNA Center, Department of Molecular, Cell, and Developmental Biology, Sinsheimer Labs, University of California, Santa Cruz, California 95064, USA

ABSTRACT

Alternative splicing contributes to muscle development, but a complete set of muscle-splicing factors and their combinatorial interactions are unknown. Previous work identified ACUAA ("STAR" motif) as an enriched intron sequence near muscle-specific alternative exons such as *Capzb* exon 9. Mass spectrometry of myoblast proteins selected by the *Capzb* exon 9 intron via RNA affinity chromatography identifies Quaking (QK), a protein known to regulate mRNA function through ACUAA motifs in 3′ UTRs. We find that QK promotes inclusion of *Capzb* exon 9 in opposition to repression by polypyrimidine tract-binding protein (PTB). QK depletion alters inclusion of 406 cassette exons whose adjacent intron sequences are also enriched in ACUAA motifs. During differentiation of myoblasts to myotubes, QK levels increase two- to threefold, suggesting a mechanism for QK-responsive exon regulation. Combined analysis of the PTB- and QK-splicing regulatory networks during myogenesis suggests that 39% of regulated exons are under the control of one or both of these splicing factors. This work provides the first evidence that QK is a global regulator of splicing during muscle development in vertebrates and shows how overlapping splicing regulatory networks contribute to gene expression programs during differentiation.

Keywords: alternative splicing; myogenesis; splicing silencer; splicing enhancer

INTRODUCTION

Alternative pre-mRNA splicing is regulated to control both the quantity and the coding potential of mRNAs from a gene. Quantity of mRNA can be controlled by regulating splicing that creates mRNA forms with premature stop codons, turning genes off by the nonsense-mediated decay pathway (Lewis et al. 2003; Lejeune and Maquat 2005; Ni et al. 2007). Coding potential can be controlled by regulating alternative exons to produce mRNAs for distinct proteins from the same gene, expanding the proteome in metazoans (Black 2003; Nilsen and Graveley 2010). Understanding the potential for alternative splicing to regulate important processes is only the first step toward an integrated view of how it is coordinated in complex systems (Kalsotra and Cooper 2011). There is a long but incomplete list of splicing factors, each with a complex set of cooperative and antagonistic relationships with other splicing factors that seem distinct at different regulatory sites and in different tissues. Several splicing factors act at other steps in gene expression, for example, binding to mRNA-untranslated regions to affect nuclear export, localization, stability, and translation (e.g., Sanford et al. 2005). Conversely, many RNA-binding proteins not characterized as splicing factors might yet be shown to regulate splicing.

Muscle cells have evolved alternative splicing for making specialized forms of cytoskeletal and mitochondrial proteins necessary for building, energizing, and controlling the contractile apparatus, such as the α-actinin SM exon, α-tropomyosin exon 2, and cardiac troponin T exon 5 (Gooding et al. 1998; Southby et al. 1999; Charlet-B et al. 2002). Inclusion of α-tropomyosin exon 2 is regulated in part by the antagonistic action of SFRS7 with hnRNPs H and F (Crawford and Patton 2006). Similarly, antagonism between PTB, CUG-BP1, MBNL1, and the CELF proteins regulates α-actinin splicing (Gromak et al. 2003). Cardiac troponin T is regulated by the opposing actions of CUG-BP2 and MBNL1 (Goo and Cooper 2009; Warf et al. 2009). A variety of other alternative splicing factors are expressed in muscle cells (e.g., MBNL2, RBFOX1, RBFOX2). Thus far, none of the proteins known to function as splicing factors in vertebrate muscle recognize the ACUAA motif that we identified near mouse heart and skeletal muscle-specific exons (Sugnet et al. 2006).

A role in splicing regulation is suspected for sequences related to the ACUAA motif since they appear enriched downstream from muscle exons (Sugnet et al. 2006; Das et al. 2007), downstream from alternatively spliced exons in general (Voelker and Berglund 2007), and upstream of exons that are spliced in hESCs but not other ES cell types (Yeo et al. 2007), as

[1]**Corresponding author**
E-mail ares@ucsc.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.038422.113.

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

well as near exons activated during differentiation of C2C12 myoblasts to myotubes in vitro (Bland et al. 2010). This sequence has been called the "STAR" motif because it is contained within a more general binding site consensus for members of the mammalian STAR (Signal Transduction and Activation of RNA) family of proteins (Vernet and Artzt 1997). Despite its bioinformatic identification, direct evidence for its function and the identity of protein factors that might act through it are not available.

We investigated the role of the ACUAA motif in C2C12 myoblasts by searching for proteins that bind a natural intron RNA region containing it. This effort identified the RNA-binding protein QK as a muscle-splicing factor acting directly through ACUAA motifs near the exons that it controls. Unlike in oligodendrocytes, where QK affects splicing by controlling the expression of the direct splicing regulator hnRNPA1 through its mRNA 3′ UTR (Zhao et al. 2010; Zearfoss et al. 2011), QK acts both positively and negatively as a direct splicing regulator through intronic splicing enhancers and silencers in muscle precursor cells. The changes in alternative splicing at QK-regulated exons during myogenesis are likely due to an increase in QK expression. Levels of the splicing factor PTB are reduced during myogenesis (Boutz et al. 2007a), and its regulatory network overlaps that of QK, such that the net regulation of a subset of QK-dependent exons is further refined by relief of PTB repression during myogenesis.

RESULTS

Capzb exon 9 is a muscle-activated exon with conserved ACUAA motifs downstream

To study the role of ACUAA motifs in splicing, we chose exon 9 from the *Capzb* gene as a model. *Capzb* encodes an actin filament capping protein; its C terminus is altered by alternative inclusion of exon 9 in heart and skeletal muscle (Schafer et al. 1994). *Capzb* exon 9 was identified as a muscle- and heart-specific exon with adjacent ACUAA elements in our study on splicing in adult tissues (Sugnet et al. 2006). We confirmed the activation of *Capzb* exon 9 during differentiation of C2C12 myoblasts (Fig. 1A; Bland et al. 2010) by RT-PCR (Fig. 1B). Alignment of sequences downstream from the *Capzb* exon from mouse and other vertebrates reveals conservation of multiple copies of the ACUAA element (Fig. 1C).

ACUAA motifs are required for exon inclusion

To determine whether the ACUAA elements downstream from *Capzb* exon 9 contribute to proper splicing of this exon, we used a β-globin reporter system (Dominski and Kole 1991), inserting *Capzb* exon 9 along with adjacent intron sequences ~200 nucleotides upstream and ~250 nucleotides downstream, between human β-globin exons (Fig. 2A). A series of mutations focusing on the conserved region from 32 to 72 nucleotides downstream from *Capzb* exon 9 containing the ACUAA motifs was tested by transfection into proliferating C2C12 myoblasts, followed by RT-PCR (Fig. 2B). The wild-type reporter includes the *Capzb* exon at ~8% in myoblasts (Fig. 2B, lane 3). Deletion of the conserved region (D1), replacing the conserved region with a nonspecific sequence (R1), or deleting just the ACUAA elements (M-Czb) results in loss of exon inclusion (Fig. 2B, lanes 4–6). These data indicate that the ACUAA motifs downstream from the *Capzb* exon are required for its proper inclusion in myoblasts, as suggested by their enrichment near a large class of regulated exons (Sugnet et al. 2006; Das et al. 2007).

Quaking and SF1 bind specifically to RNA containing ACUAA motifs

To identify proteins that bind ACUAA containing RNA, we used RNA affinity chromatography, followed by multidimensional protein identification technology (MudPIT) (Washburn et al. 2001). RNAs containing either the wild-type *Capzb* sequence from 32 to 72 nucleotides downstream from exon 9 (Fig. 3A, WT) or with mutated ACUAA motifs (Fig. 3A, M) were coupled to agarose beads and incubated with myoblast nuclear extracts. Proteins bound at low salt were released with increasing salt concentrations (Fig. 3B), and the WT and M fractions were compared on a silver-stained gel. The pattern of proteins eluted remains complex; however, specific proteins can be seen in the WT- but not the M-RNA-binding fractions (Fig. 3B, arrow). We chose to analyze the 0.4 M NaCl eluted fraction (Fig. 3B, asterisk) by MudPIT (Yates et al. 2009).

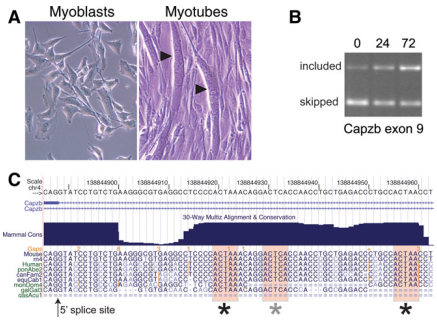


FIGURE 1. *Capzb* exon 9 is activated during myogenesis and has intronic ACUAA elements. (*A*) Phase contrast photographs of proliferating C2C12 or cells treated with low serum for 72 h to induce myogenesis. Multinucleate myotubes are indicated by arrowheads. (*B*) RT-PCR measurement of changes in *Capzb* exon 9 inclusion during differentiation using primers in exons 8 and 10 of the endogenous *Capzb* gene. Exon 9 inclusion increases during differentiation. (*C*) Alignment of intronic sequences downstream from the 5′ splice site bordering *Capzb* exon 9 from several vertebrates. The ACUAA elements are shaded and marked with asterisks.

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

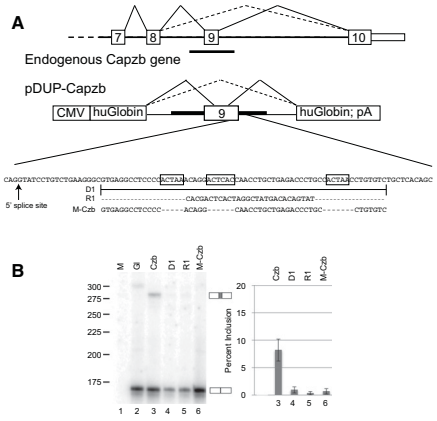


FIGURE 2. Splicing reporters based on *Capzb* exon 9. (*A*) Cartoon of *Capzb* gene structure from exons 7 through 10. The region marked by the black bar *below* the endogenous gene was amplified by PCR and cloned into pDUP51 to create the reporter construct shown. The sequence *below* the reporter diagram represents the downstream intronic region containing the ACUAA elements (boxed). Mutations tested are (D1) deletion of the conserved region; (R1) replacement of the conserved region with a nonspecific sequence; (M-Czb) deletion of the ACUAA elements. (*B*) RNA isolated from proliferating C2C12 cells transiently transfected with the constructs in *A* were analyzed by RT-PCR using primers specific for the flanking β-globin exons; spliced products are indicated to the *right* of the gel. Exon percent inclusion for each lane is graphed, *far right*. Mock-transfected and globin-only controls are shown (M, Gl, lanes *1*,*2*); the band in the Gl lane is nonspecific.

We expected a complex set of proteins to bind to RNA under these conditions; indeed, numerous cytoskeletal proteins, ribosomal proteins, and DNA-binding proteins are found in both fractions (data not shown). To examine the relative association of RNA processing factors with the WT- and M-RNAs, we considered five classes of proteins: splicing factors, hnRNP proteins, spliceosomal snRNP proteins, DEAD/H RNA helicase family proteins, and other processing enzymes (Table 1; Supplemental Table 1). Among proteins detected in the WT-RNA fraction, but absent from the M-RNA fraction, are the known ACUAA-binding factors QK (Ryder and Williamson 2004; Galarneau and Richard 2005) and Splicing Factor 1 (SF1) (Arning et al. 1996; Liu et al. 2001; Corioni et al. 2011). The peptides observed for the QK protein do not distinguish which of the three major QK isoforms (QK-5, QK-6, and QK-7) are bound, since these forms differ only in their C-terminal tails (Ebersole et al. 1996; Chénard and Richard 2008).

We validated several proteins from the MudPIT analysis by Western blot using fractions from the WT- and M-RNA columns. QK and SF1 were well retained by the WT- but not the M-RNA affinity matrix (Fig. 3C). Thus, it appears that QK, SF1, and several other proteins depend on the intact ACUAA elements for their efficient association with the WT-RNA. Whether these bind directly to RNA or indirectly to other proteins bound to RNA is not known. hnRNP K and Ddx5 were retained by both of the matrices, although it appears the M-RNA matrix bound less protein overall. Furthermore, the known splicing regulator PTB is bound by the *Capzb* RNA sequence but is not strongly dependent on the ACUAA sequences (Fig. 3C). We conclude that QK, SF1, PTB, as well as other proteins, bind pre-mRNA sequences downstream from *Capzb* exon 9 and are candidates for factors that regulate its splicing.

Depletion of QK, but not SF1, inhibits *Capzb* exon 9 inclusion in myoblasts

To test the hypothesis that QK activates *Capzb* exon 9 we used siRNA against QK (all isoforms) to deplete the protein in myoblasts. A Western blot using the pan-QK antibody shows substantial reduction in QK protein (Supplemental Fig. 1A), and RT-PCR analysis of the endogenous *Capzb* transcripts reveals that inclusion of exon 9 is greatly inhibited (Fig. 4A, lane 3). In contrast, when we use siRNA to deplete SF1 in myoblasts, endogenous *Capzb* exon inclusion is unaffected (Fig. 4A, lane 4; Supplemental Fig. 1A), suggesting that SF1 does not activate exon inclusion through downstream ACUAA sequences. Thus, QK, not SF1, is the ACUAA-binding protein that contributes to *Capzb* exon 9 inclusion in myoblasts.

Quaking isoform 5 activates *Capzb* exon 9 through the ACUAA motif

Given that reduction in QK protein levels decreases *Capzb* exon 9 inclusion, we asked whether increasing QK protein levels would promote inclusion. We focused on Quaking isoform 5 (QK-5) because it is predominantly nuclear, whereas isoforms 6 and 7 (QK-6, QK-7) are mostly cytoplasmic in cell types that have been tested (Chen and Richard 1998; Wu et al. 1999). We transfected a myc-tagged QK-5 construct (generous gift from Sean Ryder) along with wild-type and mutant *Capzb* reporter into myobloasts (Fig. 4B). Expression of myc-QK-5 results in a nearly eightfold increase in exon 9 inclusion (Fig. 4B, cf. lanes 4 and 5, $P < 0.001$, $n = 4$), while the M-Czb reporter lacking the three ACUAA motifs was not affected (lanes 6 and 7, see Fig. 2 for a description of the M-Czb mutant). Thus, increased expression of QK-5 activates inclusion of *Capzb* exon 9 in a fashion dependent on the downstream intronic ACUAA sequences.

Repression of PTB enhances QK activation of *Capzb* exon 9 inclusion

RNA affinity chromatography detects PTB binding to the intronic sequence downstream from *Capzb* exon 9 (Fig. 3). PTB represses exons in C2C12 myoblasts (Boutz et al.

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

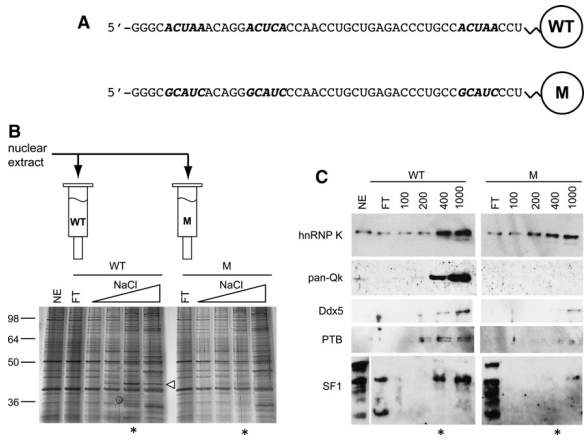


FIGURE 3. Identification of ACUAA-binding proteins in proliferating C2C12 cells. (*A*) Sequences of in vitro-transcribed RNAs used. ACUAA motifs or mutant versions are italicized. (*B*) Silver-stained denaturing gel of proteins bound to the wild-type (WT) or mutant (M) RNAs. Lanes are as follows: NE, nuclear extract; FT, flow through; triangles indicate increasing NaCl elution (0.1, 0.2, 0.4, and 1 M). Arrow points to enriched proteins visible in the WT- but not the M-RNA-binding fractions. Asterisks mark fractions subjected to MudPIT analysis. (*C*) Western blot of hnRNP K, QK, Ddx5, SF1, and PTB in the eluted fractions. Lanes correlate with those in *B*.

2007a), and PTB repression can be mediated through downstream sites in combination with upstream binding sites (Chan and Black 1997; Southby et al. 1999). To test whether *Capzb* exon 9 is under the control of PTB, we depleted PTB from myoblasts with or without concomitant overexpression of QK-5 (Fig. 4C,D). After confirming the effects on protein levels by Western blot (Fig. 4C, lanes 4–7), we analyzed splicing of the *Capzb* exon 9 reporter construct by RT-PCR (Fig. 4D, lanes 4–7). When PTB expression is reduced, inclusion levels increase approximately fourfold (Fig. 4D, cf. lanes 4 and 5, $P < 0.0001$, $n = 4$). As discussed above, increased QK-5 expression activates the *Capzb* exon 9 reporter on its own (Fig. 4D, cf. lanes 4 and 6). However, in combination with reduced PTB expression, increased QK-5 expression leads to greater exon inclusion (Fig. 4D, cf. lanes 3 and 7, $P < 0.002$, $n = 4$). Similar results are obtained for the endogenous *Capzb* exon 9 (Supplemental Fig. 1B). This shows that PTB represses *Capzb* exon 9 antagonistically to QK. A model describing these relationships is shown in Figure 4E.

QK depletion in myoblasts results in widespread alternative splicing changes

To determine the extent of QK function in splicing, we depleted QK in proliferating C2C12 cells and captured changes in alternative splicing using splicing-sensitive microarrays. Western blots revealed that levels of all three isoforms are reduced by a QK-specific siRNA directed at a common region of their mRNAs (Fig. 5A). We detected 406 cassette exons whose inclusion changes in QK-depleted myoblasts as compared with the mock-depleted controls ($q \leq 0.05$ and $|\text{Sepscore}| \geq 0.3$; Supplemental Table 2). We validated a number of these by RT-PCR (Fig. 5B). Among the 162 exons with the most significant changes ($q = 0$, $|\text{Sepscore}| \geq 0.5$), about half show a decrease in inclusion upon QK depletion, indicating that QK promotes their inclusion in myoblasts. In contrast, half show an increase in exon inclusion, indicating that QK represses them. We conclude that QK has a major role in regulating splicing in muscle cells.

Enrichment of intronic ACUAA sequences near QK-regulated exons

If QK functions in muscle cells like other globally active splicing factors such as the MBNL (Du et al. 2010; Charizanis et al. 2012; Wang et al. 2012), RBFOX (Zhang et al. 2008; Yeo et al. 2009), PTB (Xue et al. 2009), and Nova proteins (Ule et al. 2006), we would expect to find QK-binding sites near the exons it controls. We searched for 5-mers enriched in the regions 150 nucleotides upstream of and downstream from the top 162 QK-regulated exons relative to the corresponding regions of 5166 cassette exons with probesets on the microarray. Significant (FDR ≤ 0.05) enrichment for CUAAC and ACUAA (and no other 5-mers) is found upstream of exons repressed by QK and downstream from exons activated by QK (Supplemental Table 3). We mapped the positional occurrence of the ACUAA 5-mer across the entire set of QK-regulated exons relative to cassette exons whose splicing did not change upon QK depletion (Fig. 5C; Supplemental Table 3). In general, when QK binds upstream of an exon it acts as a repressor, and when bound downstream, it acts as an activator (Fig. 5C). This is comparable to other splicing factors that bind intron sequences, such as Nova, MBNL1, and others (Ule et al. 2006; Zhang et al. 2008; Xue et al. 2009; Yeo et al. 2009; Du et al. 2010; Charizanis et al. 2012; Wang et al. 2012). The distribution of ACUAA both upstream at QK silencers and downstream at QK enhancers is notably bimodal. This is likely due to the fact that QK is a dimer with a bipartit-binding site (Zorn and Krieg 1997; Chen and Richard 1998; Galarneau and Richard 2005; Carmel et al. 2010; Beuck et al. 2012).

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

TABLE 1. Detection of RNA processing factors in WT- and M-RNA affinity chromatography fractions

Group	Only WT RNA	Both	Only M RNA
Splicing factors	SF1, KHDRBS1, PTBP2, TIAL1	PTBP1, **SFPQ, SFRS1, SFRS2, SFRS3, SRRM2, U2AF2** KHSRP, SFRS6	CELF2, MBNL1
HnRNP proteins	**HNRNPDL**	**HNRNPK, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPH1, HNRNPH2, PCBP1, PCBP2** HNRNPA0, HNRNPAB, HNRNPD, HNRNPF, HNRNPL, HNRNPL1, HNRNPM, HNRNPR, HNRNPU	HNRNPC
SnRNP and spliceosomal proteins	**SNRPB, SNRPC, SNRPD3, SNRPE**	**SNRPD2, SNRNP70, SNRPA1** PRPF6, PRPF8, SF3A1, SF3B1, SF3B2, SF3B3, SNRNP200	
DEAD/H proteins	DDX3Y, DDX18	**DDX5, DDX17, DDX21** DDX1, DDX4, DHX9, DHX15, DDX23	DDX6, DDX46, DDX58
Other	**QK, CSTF2** ADAR, CPSF7, ELAVL1, EWSR1, EXOSC9, IGF2BP2, IGF2BP3, PABPC2, SART1	ADARB1, CPSF6, CSDE1, DEK, PARP1, PSPC1, RBMS1, RBMS2, SERBP1, STAU1	KHDRBS3, RBM39, SSB

Bold indicates that peptides covering >10% of the coding region were detected. Underlined proteins were validated by Western blot (see Fig. 3).

The PTB and QK splicing regulatory networks overlap in myoblasts

Capzb exon 9 inclusion is controlled in myoblasts by a balance between QK activation and PTB repression and is induced during myoblast differentiation to myotubes. In addition, PTB down-regulation and several PTB-regulated splicing changes take place during myoblast differentiation (Boutz et al. 2007a). To explore the extent of PTB-splicing regulation in myoblasts, we used splicing-sensitive microarrays to analyze splicing changes after PTB depletion by siRNA (Supplemental Fig. 2A). Depletion of PTB results in splicing changes for 485 cassette exons ($q = 0$, $|\text{Sepscore}| \geq 0.5$; Supplemental Table 4), of which 268 show increased inclusion (repressed by PTB), and 217 show decreased inclusion (activated by PTB). A number of 5-mers including UCUCU are significantly enriched in regions upstream of PTB repressed exons (FDR < 0.01; Supplemental Table 5), and a map of the UCUCU frequency over the intronic regions adjacent to PTB repressed exons shows a broad region from ~100 nucleotides upstream and up to the 3′ splice site where this signal is located (Supplemental Fig. 2B; Supplemental Table 5). Our results for myoblasts are similar to analyses of PTB in other cell types and tissues (Boutz et al. 2007b; Xue et al. 2009; Llorian et al. 2010).

When we intersect the QK-splicing network with the PTB-splicing network (Supplemental Table 6), we find 172 exons whose inclusion is significantly affected ($q \leq 0.05$, $|\text{Sepscore}| \geq 0.3$) by either QK or PTB depletion, including *Capzb* exon 9. As expected (Fig. 4), *Capzb* exon 9 is activated by QK and repressed by PTB. The 172 jointly regulated exons represent 42% of the 406 QK-regulated events, but only 17% of the 1012 PTB-regulated events, indicating that while the PTB network detected in this experiment is more than twice as large as the QK network, QK regulation is more strongly influenced by PTB than the other way around. Interestingly, these exons sort out about equally into the four regulatory classes that could be created by the combination of QK and PTB (40 exons are activated by both, 40 are repressed by both, 48 are activated by QK and repressed by PTB, 44 are repressed by QK and activated by PTB). Thus, *Capzb* is not unique; there are substantial numbers of exons controlled by QK and PTB in all possible combinatorial classes.

Contributions of QK and PTB to alternative splicing during myoblast differentiation

The analysis above details QK- and PTB-controlled splicing events in myoblasts. To assess their roles in alternative splicing during muscle cell differentiation, we asked which exons under QK and PTB control in myoblasts are also regulated during myogenesis. Regulation of PTB levels has been implicated in controlling several alternative splicing events in myogenesis (Boutz et al. 2007a). We compared splicing in RNA from myoblasts to that from 72-h differentiated myotubes using the same array platform as for the depletion experiments and detected many robust splicing changes, as also reported by groups using other platforms (Bland et al. 2010; Trapnell et al. 2010). Data for cassette exons showing changes in inclusion in myoblasts depleted of either QK or PTB or after differentiation into myotubes are shown in Supplemental Table 6.

We followed changes in QK protein expression during myogenesis by Western blot using the pan-QK antibody on proteins from undifferentiated myoblasts and 72 h myotube cultures (Fig. 6A). The upper band represents both QK-5 and QK-7, while the lower band represents QK-6. Relative to GAPDH, QK levels increase two- to threefold during differentiation. This increase might be expected to influence inclusion of the many exons controlled by QK in myoblasts in the

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

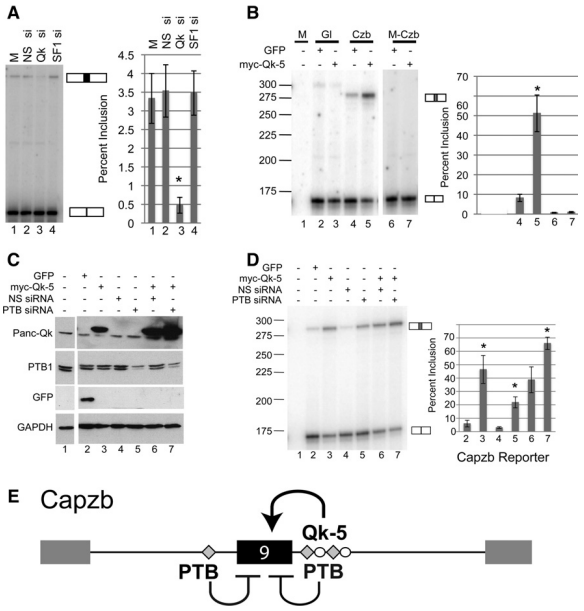


FIGURE 4. QK and PTB control inclusion of *Capzb* exon 9. (*A*) Analysis of RT-PCR products from C2C12 cells mock-transfected (M) or transiently transfected with a nonspecific siRNA (NS si), a QK-specific siRNA (QK si), or an SF1-specific siRNA (SF1 si). Spliced products corresponding to exon 9-included or excluded endogenous *Capzb* mRNAs are indicated to the *right* of the gel. Percent inclusion was quantified and graphed, *far right*. (*B*) Overexpression of QK leads to *Capzb* exon 9 inclusion provided ACUAA elements are intact. Proliferating C2C12 cells were mock transfected (M) or transiently cotransfected with a QK-5 expression construct (myc-QK-5) or a control GFP expression construct (GFP), along with either the wild-type *Capzb* reporter (Czb) or the mutant reporter with deletion of the ACUAA elements (M-Czb) from Figure 2. A globin-only control is also shown (Gl, lanes *2*,*3*); the band in the Gl lanes is a nonspecific RT-PCR product. Inclusion of the WT *Capzb* exon increased with expression of QK by approximately fivefold (cf. lanes *4* and *5*, $P < 0.001$, $n = 4$). Spliced products are indicated and percent inclusion is graphed, *far right*. QK depletion significantly reduces exon inclusion (asterisk, $P < 0.05$, cf. lanes *2* and *3*). (*C*) Western blot of whole-cell extract from cells mock-transfected (lane *1*) or transiently cotransfected with the *Capzb* reporter and either a GFP expression construct (GFP), a QK-5 expression construct (myc-QK-5), a nonspecific siRNA (NS siRNA), or a PTB-specific siRNA (PTB siRNA) as indicated at *top*. Blots were probed with antibodies to GFP, pan-QK, PTB, and GAPDH as a loading control (as indicated to the *left* of the blots). (*D*) Analysis of RT-PCR products from RNA isolated from cells transfected in *C*. Spliced products are indicated to the *right* of the gel. Percent inclusion was quantified and graphed, *far right*. (*) $P < 0.002$. (*E*) Model for regulation of *Capzb* exon 9 by QK and PTB. QK activates inclusion through intronic ACUAA elements downstream, and PTB represses inclusion, possibly through sequences similar to PTB-binding sites upstream and/or through binding downstream, as detected by RNA affinity chromatography (Fig. 3).

opposite direction as observed in QK-depleted myoblasts. Conversely, since PTB is down-regulated during myogenesis (Fig. 6A; Boutz et al. 2007a), the splicing of exons controlled by PTB in myoblasts may change in the same direction as observed in PTB-depleted myoblasts. For exons where QK and PTB appear to act similarly, exon inclusion may not change, as the increase in QK might offset the loss of PTB. Finally, the behavior of some exons may be influenced by other factors in addition to QK or PTB during differentiation.

We classified exons with respect to their dependence on these two splicing factors in myoblasts and their changes during muscle cell differentiation by clustering the Sepscores of exons whose inclusion changed in at least two of the three comparisons (Fig. 6B; Supplemental Table 7). Several expected classes are visible. For example, a subset of the exons activated by QK (decreased inclusion after depletion, blue), but not significantly affected by PTB (black) in myoblasts, is also activated during myotube differentiation (yellow), in part due to increasing QK protein (Fig. 6B, purple circle). A large group of exons is not significantly controlled by QK, but repressed by PTB in myoblasts (increased after depletion, yellow) and activated during differentiation (yellow), likely in part by the decrease in PTB protein (Fig. 6B, green circle; Boutz et al. 2007a). Other expected classes such as would be generated by increased QK repression or loss of PTB activation during differentiation can also be observed (Fig. 6B, white and blue circles). The combined regulatory effect of both factors explains another set of events (e.g., Fig. 6B, yellow circle). In total, 22% of the splicing changes observed during myoblast differentiation occur in exons controlled by QK or PTB and are consistent with the changes in QK and PTB protein levels during differentiation. The observed 72-h splicing changes of another 17% of exons that are also regulated by QK or PTB in myoblasts cannot be explained by the simple expectation that the increase in QK and the reduction in PTB relative to the myoblast state will predict the differentiated cell state; these exons may be influenced by other splicing factors also expressed in C2C12 cells such as RBFOX and MBNL family members.

To identify exons whose inclusion during differentiation is likely altered by changes in activity of both QK and PTB, we

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

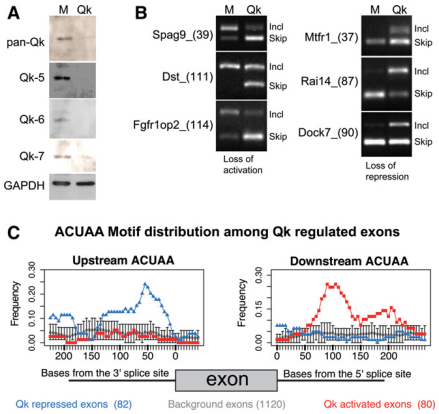


FIGURE 5. The QK-splicing regulatory network in proliferating myoblasts. (*A*) QK siRNA depletes all three forms of QK in myoblasts. Western blot of proteins from proliferating C2C12 cells mock-transfected (M) or transfected with a QK-specific siRNA (QK). Blots were probed with antibodies to pan-QK, QK-5, QK-6, QK-7, and GAPDH as a loading control. (*B*) RT-PCR validation of splicing changes detected by array analysis. Agarose gel analysis of RT-PCR products for alternative cassette exons using RNA from mock-transfected or QK-depleted cells. Exon-included product is always the *upper* band, exon-skipped product always the *lower*. The splicing event is labeled by gene name and exon number and size (nucleotides, in parenthesis). (*C*) The frequency of ACUAA elements upstream of (*left*) and downstream from (*right*) the top 162 QK-regulated cassette exons is mapped. Exons up-regulated by QK depletion are in blue and inferred to be repressed by QK; down-regulated exons are in red and inferred to be activated by QK. Control exons are in gray. Error bars indicate the 95% confidence limit for ACUAA frequency in the control exons.

made a three-dimensional plot of Sepscores for the 66 exons whose splicing changed in all three experiments (Fig. 6C; Supplemental Table 8), allowing the diversity of exon responses to be visualized. There are 10 exons (blue spheres) whose inclusion decreases upon QK depletion and increases upon PTB depletion, which are activated during differentiation, including *Capzb* exon 9 (Capzb_(113)). The behavior of these exons are consonant with the changes in protein level observed during differentiation or depletion in undifferentiated cells. As with the clustering (Fig. 6B; same exons are marked by asterisks), we observe many combinations of splicing changes within the three experimental conditions, many of which are not consonant with the changes in protein level during differentiation. For example, the Slain2_(78) exon (colored in green) is repressed by QK (exon inclusion increases upon QK depletion) and activated by PTB (exon inclusion decreases upon PTB depletion), but is activated during differentiation. Thus the robust activation of Slain2_(78) must be due to other factors that overcome strong repression due to increasing levels of QK and loss of activation with decreasing levels of PTB. This highlights the incomplete nature of our understanding of splicing regulation during muscle cell differentiation and underscores its complexity. Nonetheless, in this more select set of exons, we again observe that the differentiation behavior of a large number of exons (29%) can be explained by changes in QK and PTB protein levels during differentiation (Fig. 6C; Supplemental Table 8).

DISCUSSION

In a previous study we used genome-wide splicing data from mouse tissues to identify motifs associated with heart and skeletal muscle-specific exons, revealing an RNA sequence motif ACUAA similar to those recognized by SF1 and the STAR family of RNA-binding proteins (Sugnet et al. 2006). We have now shown that this motif mediates splicing regulatory control through QK, in particular the nuclear isoform 5 (QK-5). At *Capzb* exon 9 the conserved downstream ACUAA elements are required for exon inclusion, bind QK in a fashion dependent on ACUAA, and mediate QK-5 activation (Figs. 2–4; Supplemental Fig. 1). Furthermore, QK regulates hundreds of splicing events in muscle cells, as a repressor or an activator, depending on the location of nearby ACUAA motifs (Fig. 5; Supplemental Tables 3, 4). We found that QK protein levels change during differentiation of myoblasts to myotubes (Fig. 6), suggesting a mechanism for a part of the alternative splicing program during myotube differentiation. Finally, we intersected the QK splicing regulatory network with that of PTB, a splicing factor that is down-regulated during myogenesis (Boutz et al. 2007a), and found classes of exons controlled by both proteins (Fig. 6). This work identifies QK as a global splicing factor in muscle cells and reveals that QK plays a larger role beyond its known activity as a translational regulator binding to 3′ UTRs.

QK RNA-binding proteins are ancient regulators of fundamental processes

Mouse Quaking was originally identified as a recessive mutation causing a jittery phenotype attributed to dysmyelination in the central nervous system (Sidman et al. 1964). Quaking has been implicated in a striking variety of processes in the mouse, such as embryogenesis, blood vessel development, glial cell fate determination, apoptosis, and smooth muscle development, while the human homolog, *QKI,* has been implicated in a number of diseases, including ataxia, glioblastoma development, and schizophrenia (Chénard and Richard 2008). QK homologs are expressed in genomes as divergent as sea urchin and are critically involved in fundamental cell and developmental processes in deeply diverged metazoans from *C. elegans* (*gld-1*, *asd-2*) (Francis et al. 1995; Ohno et al. 2008) to *Drosophila* (*how*) (Baehrecke 1997; Fyrberg et al. 1997; Zaffran et al. 1997) and zebrafish (*qkA*) (Tanaka et al. 1997; Lobbardi et al. 2011). In vertebrates, a larger family of proteins related to QK also exists, including SAM68 and

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

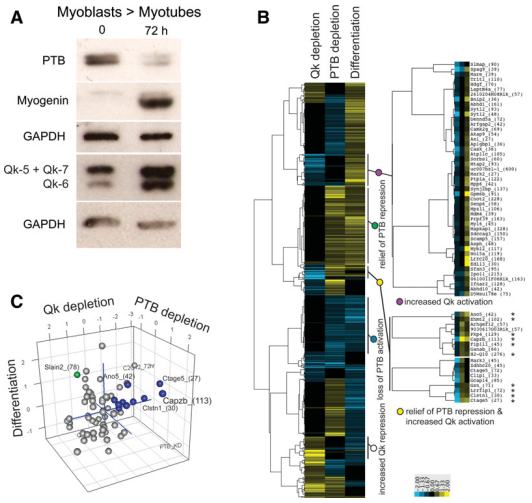


FIGURE 6. Regulation of QK protein level, PTB protein level, and alternative splicing during myogenic differentiation. (*A*) Western blot of proteins during differentiation. Proteins from undifferentiated C2C12 Myoblasts (0) or cells after 72 h of differentiation (Myotubes, 72 h) were detected using antibodies against proteins listed at *left*: PTB, myogenin, pan-QK (co-migration of QK-5 and QK-7 results in one band, QK-6 is the lower band), and GAPDH. (*B*) Clustering of splicing events based on their response to QK or PTB depletion in myoblasts relative to their change in splicing during differentiation. Yellow means that inclusion increased, whereas blue indicates reduction in exon inclusion. Asterisks mark exons colored blue in *C*. A total of 612 cassette exon events with significant changes (q-value ≤ 0.05 and $|\text{Sepscore}| \geq 0.3$) in at least two data sets are shown; Sepscores that did not meet the significance criteria were replaced with zeros for clustering (Supplemental Table 7). (*C*) Graph of exons with splicing changes in all three data sets. The Sepscores of 66 exons with q-value ≤ 0.01 and $|\text{Sepscore}| \geq 0.3$ in all three data sets were plotted using R. Exons activated by QK and repressed by PTB in myoblasts that are activated during differentiation are blue. The Slain2_(78) exon is green.

the less-studied SLM-1 and SLM-2 (Lock et al. 1996; Di Fruscio et al. 1999). Together, this evolutionary depth and breadth of functional contribution implies that QK and its relatives have been integrated into the post-transcriptional regulatory structure of eukaryotic cells for hundreds of millions of years. Here we show that in addition to its previously demonstrated role in translational control, that QK is a global regulator of splicing in vertebrate muscle cells during differentiation.

QK is a splicing factor that directly activates *Capzb* exon 9

The identifying feature of this ancient STAR protein family is a dimeric RNA-binding domain that recognizes bipartite sequences of the form AYUAAY..N2-20..UAAY (Zorn and Krieg 1997; Chen and Richard 1998; Galarneau and Richard 2005; Carmel et al. 2010; Beuck et al. 2012). Sequence elements related to the ACUAA motif are found in many contexts, bound by different proteins, making it difficult to predict the binding partner of the ACUAA motif observed bioinformatically (Sugnet et al. 2006; Das et al. 2007; Bland et al. 2010). For example, the branchpoint sequence URAY (in yeast, UACUAAC), found near exons and utilized in constitutive splicing during the first catalytic step, is bound by SF1 (in yeast, BBP1) to recruit the splicing machinery to the intron (Kramer 1992; Liu et al. 2001). The response element for QK and its *C. elegans* homolog GLD-1 (Ryder and Williamson 2004; Ryder et al. 2004) has been found in 3′ UTRs, where it regulates mRNA stability, localization, and translation (Saccomanno et al. 1999; Li et al. 2000; Lakiza et al. 2005; Zhao et al. 2010; Zearfoss et al. 2011). We found that both QK and SF1 bound the ACUAA-like sequences downstream from *Capzb* exon 9 by RNA affinity chromatography (Fig. 3). We then showed that depletion of SF1 has no detectable effect on *Capzb* exon 9 inclusion, while QK depletion strongly reduced exon inclusion (Fig. 4A,B). Furthermore, overexpression of the QK-5 isoform increased exon inclusion (Fig. 4D; Supplemental Fig. 1A), demonstrating a direct role for QK in regulating alternative splicing of *Capzb* exon 9.

Direct and indirect modes of QK influence on splicing

Several studies of invertebrate QK homologs have identified roles in splicing at specific exons. *C. elegans asd-2* mutations were isolated using a two-color reporter gene carrying mutually exclusive exons from a body wall muscle myosin gene (Ohno et al. 2008). Mutations in an ACUAA-like motif-containing element (two CUAAC repeats) were shown to abrogate the regulation of adult-specific alternative splicing, phenocopying the *asd-2* mutant worms. In flies, the QK homolog *how* promotes inclusion of neurexin IV exon 3, most likely through downstream ACUAA sequences (Rodrigues et al. 2012). It has remained an open question as to whether these single exon studies reflect a larger involvement of QK in splicing at other exons more globally. In vertebrates, QK has been shown to regulate splicing in oligodendrocyte precursors, but not primarily by binding near exons with ACU AA. In these cells, QK stabilizes hnRNPA1 mRNA by binding

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

to its 3′ UTR; thus, a primary effect of depleting QK appears to be down-regulation of hnRNP A1, which causes gene expression and splicing defects (Zearfoss et al. 2011).

In myoblasts, however, introns neighboring the exons most affected by QK depletion are clearly enriched for the QK-binding site (Fig. 5) and exhibit the "upstream repressing, downstream activating" pattern observed for other globally active splicing factors like Nova (Ule et al. 2006), MBNL1 (Du et al. 2010), MBNL2 (Charizanis et al. 2012), RBFOX1 (Jin et al. 2003), and RBFOX2 (Zhang et al. 2008; Yeo et al. 2009). Splicing of *Capzb* exon 9 is influenced by the well-defined binding sites for QK (Ryder and Williamson 2004; Galarneau and Richard 2005) in the downstream intron and by expression of QK (Figs. 2B, 4; Supplemental Fig. 1). Thus, we infer that QK acts directly in myoblast splicing, rather than indirectly through hnRNP A1, as in both the *Quaking*^{viable} mouse (Zhao et al. 2010) and oligodendrocyte precursor cells (Zhao et al. 2010; Zearfoss et al. 2011). This does not exclude the possibility that QK regulates other RNA-binding proteins in myoblasts—a view supported by our observation that a subset of exons is regulated by QK but lacks good matches to the QK response element within 150 nucleotides of their splice sites. These exons could be regulated by other factors that are themselves regulated by QK, or have QK-binding sites farther from the exon, as has been observed for RBFOX regulation (Huh and Hynes 1994; Lim and Sharp 1998; Tang et al. 2009). In addition, we note that a 5-mer related to the RBFOX motif UGCAUG is slightly enriched upstream of exons repressed by QK in myoblasts (Supplemental Table 3), suggesting that either these exons are coregulated by a QK and an RBFOX family member, or perhaps that QK might regulate an RBFOX family member. Clearly, the regulation and overlap of complexity of splicing during muscle differentiation is far from resolved.

QK is a dual regulator of alternative splicing and mRNA function

In mammals, RNA-binding proteins come in families (e.g., PTB/nPTB, RBFOX1 through RBFOX3, MBNL1 through MBBNL3); members of the same family share very similar RNA-binding domains and seem to bind indistinguishable motifs. Yet family members have both shared and distinct functions that are governed by sequence context, cell-type, or tissue-specific information. For example, MBNL1 appears to influence alternative pre-mRNA splicing more significantly in muscle tissue, whereas MBNL2 exerts a similarly predominant influence in brain tissue (Du et al. 2010; Charizanis et al. 2012). In addition, numerous in vivo cross-linking studies on splicing factors show that many also bind to 3′ UTRs of mRNAs as well as to introns near regulated exons (Hafner et al. 2010; Charizanis et al. 2012; Masuda et al. 2012; Wang et al. 2012). This implies that the same protein may function in splicing, localization, or translation within the same cell, or may only participate in some functions in certain cell types. In the case of QK, the three isoforms differ in amino acid sequence only at their C-termini, and this is thought to confer, in part, their distinct localization behaviors. QK-5, the predominantly nuclear form, seems more likely to be involved in direct splicing regulation (this work), whereas the cytoplasmic QK-6 and QK-7 forms may mediate translational control or mRNA stability with indirect effects on splicing (Saccomanno et al. 1999; Zhao et al. 2010; Zearfoss et al. 2011). We currently lack an understanding of the information that specifies which family member or isoform should bind where, or carry out which function. One possibility is that different family members or protein isoforms carry out these different functions, perhaps interacting with separate sets of adaptor proteins that confer specificity. But whether and how these distinct isoforms might be partitioned to different RNA target sites is unknown.

Combinatorial control of the myogenic splicing program by QK and PTB

To understand the relationship between the QK and PTB splicing regulatory networks in myoblasts and during differentiation, we defined the PTB global splicing network in undifferentiated myoblasts and differentiated myotubes using splicing sensitive microarrays. The PTB network is about 2.5 times larger than the QK network; while only 17% of PTB regulated exons are also controlled by QK, PTB coregulates nearly 42% of the QK network. The 172 exons coregulated by QK and PTB are split evenly into four classes: one where both repress, another where both activate, one where QK represses while PTB activates, and one where QK activates and PTB represses. Thus, for just over half of exons regulated by both proteins, they oppose each other and reinforce each other at the remaining exons. Analysis of the genes containing these regulated exons indicates that a diversity of both general and muscle-specific functions are under control of these factors.

During differentiation, the two proteins change levels in opposite directions, with QK increasing and PTB decreasing (Fig. 6) due in part to the induction of miR-133 and its repression of PTB mRNA (Boutz et al. 2007a). Exons such as *Capzb* exon 9 are activated during differentiation by the combined increase in QK activation and decrease of repression by PTB. However, numerous other splicing factors play roles in controlling alternative splicing in muscle, and given the complex interplay of splicing factors at most exons, dissecting a developmental program of alternative splicing will involve understanding not just why exons change inclusion levels, but how those that do not change are made stable (Fig. 6B). Nonetheless, by mapping splicing regulatory networks in myoblasts and following changes in factor level during differentiation, ~22%–29% of the alternative splicing changes during myogenesis occur in a fashion consistent with primary regulation by QK alone, PTB alone, or the two proteins together. The close physical proximity of mapped PTB and

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

QK motifs (Fig. 5, cf. Supplemental Fig. 2) in the sequence upstream of exons repressed by both proteins is notable and may indicate a mechanism whereby the relative occupancy and competition for binding plays a key role in the bimodal relationship between the two proteins—i.e., as one protein leaves another is able to bind. Future work will involve integrating our understanding of the roles of these and other splicing factors during myoblast differentiation, including MBNL1 (which also activates *Capzb* exon 9) (Du et al. 2010), RBFOX proteins, and the CELF proteins (Gromak et al. 2003). In addition, the control of PTB levels by miR-133 during myogenic differentiation (Boutz et al. 2007a) and the interactions between QK and microRNAs in other systems (Chen et al. 2012; Ji et al. 2012; van Mil et al. 2012) suggest that a full understanding of the developmental control of alternative splicing will require expanding our appreciation for other control mechanisms at the same time as we dissect the roles of splicing networks.

MATERIALS AND METHODS

Cell culture

Mouse C2C12 myoblast cells (ATCC) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Invitrogen). To induce myogenesis, 1.5×10^5 cells were plated in growth medium (DMEM + 10% FBS) in 6-well tissue culture-treated plates 12–18 h before addition of differentiation medium (DMEM supplemented with 2% horse serum, Invitrogen). Media was changed every 24 h, and myotube formation was visible after ~48 h. Differentiation was verified visually by myotube formation and Western blot of myogenic markers. RNA and protein were isolated under subconfluent conditions in growth media (for myoblast analyses) or 72 h after addition of differentiation medium (for differentiation analyses).

Microscopy

Live cells were photographed on a Zeiss Axiovert 200 microscope using a Zeiss 20×/0.30 LD A-Plan lens. Images were collected using a SPOT Insight Mosaic 3.2.0 color digital camera and SPOT Advanced software (Diagnostic Instruments, Inc).

Transient transfection

Transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were incubated for 18–24 h post-transfection prior to harvesting RNA and/or protein.

RNA isolation

Whole-cell RNA was isolated using Trizol (Invitrogen) according to manufacturer's instructions. RNA pellets were resuspended in 50 µL of TE (10 mM Tris, pH 8.0; 1 mM EDTA). Cytoplasmic RNA was extracted as described previously (Modafferi and Black 1997).

Protein isolation

Cells were harvested by trypsinization and washed once with 1× PBS, then resuspended in 50 µL of NP-40 lysis buffer (150 mM NaCl, 10 mM tris-HCl at pH 7.8, 0.65% NP-40) with complete protease inhibitor cocktail tablets (Roche), and incubated on ice for 10 min. After spinning for 10 min at 18,000*g* at 4°C, the supernatant was flash frozen and stored at −80°C.

Splicing microarrays

Targets were prepared from three replicate cultures for each sample with the RiboMinus Kit (Invitrogen) and the GeneChip Whole Transcript Sense Target Labeling Kit (Affymetrix) Labeled target was hybridized to the MJAY Chip (Affymetrix #540092). Chips were washed and scanned using the Fluidics Station 450 and GeneChip scanner (Affymetrix). Data were analyzed as in Sugnet et al. (2006). The Sepscore is log2 (Include/Skip ratio) of the experimental sample over the reference sample. When exon inclusion increases the Sepscore is positive. Hierarchical clustering (Eisen et al. 1998) was done with Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm); clusters were displayed with Java TreeView (Saldanha 2004) (http://jtreeview.sourceforge.net/).

Motif analysis

Counts of all 5-mers in the selected region of an exon set are compared with their counts from a background set of sequences using Fisher's exact test, with multiple testing correction. Motif mapping is described in the Supplemental Methods.

RNA affinity chromatography

RNA was bound to beads and incubated with nuclear extract as described in the Supplemental Methods.

Multidimensional Protein Identification Technology (MudPIT)

Mass spectrometry was performed by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley. Peptides were eluted using a 14-step MudPIT procedure (Washburn et al. 2001). The programs SEQUEST and DTASELECT were used to identify peptides and proteins from the mouse database (Eng et al. 1994; Tabb et al. 2002). The false-positive rate for protein identification is calculated to be <0.1% (Elias et al. 2005; Sharma et al. 2008). Detailed methodology is in the Supplemental Methods.

RT-PCR

cDNA synthesis and PCR details are in the Supplemental Methods; primer sequences are in Supplemental Table 9. Percent exon inclusion ([inclusion/(inclusion + skipping)] × 100) is graphed in Excel; error bars are ±SD for $n \geq 3$ independent experiments.

DATA DEPOSITION

Microarray data are deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE40962.

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Doug Black (Dept. of Microbiology, Immunology and Molecular Genetics, UCLA) and Sean Ryder (Dept. of Biochemistry and Molecular Pharmacology, UMass Medical School) for reagents and helpful advice. Dr. Lori Kohlstaedt (NIH grant 1S10RR025622-01) was a great help with MuDPIT. R.N. and M.H. had postdoctoral fellowships from the California Institute of Regenerative Medicine. This work was supported by NIH training grant TG 5T32GM00 8646-14 (W.S.F.), MDA135140 from the Muscular Dystrophy Association, NIH grant GM084317, and NIH grant GM040478 to M.A.

Received January 17, 2013; accepted February 20, 2013.

REFERENCES

Arning S, Gruter P, Bilbe G, Kramer A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* **2:** 794–810.

Baehrecke EH. 1997. *who* encodes a KH RNA binding protein that functions in muscle development. *Development* **124:** 1323–1332.

Beuck C, Qu S, Fagg WS, Ares M Jr, Williamson JR. 2012. Structural analysis of the Quaking homodimerization interface. *J Mol Biol* **423:** 766–781.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291–336.

Bland CS, Wang ET, Vu A, David MP, Castle JC, Johnson JM, Burge CB, Cooper TA. 2010. Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* **38:** 7651–7664.

Boutz PL, Chawla G, Stoilov P, Black DL. 2007a. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev* **21:** 71–84.

Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M Jr, Black DL. 2007b. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* **21:** 1636–1652.

Carmel AB, Wu J, Lehmann-Blount KA, Williamson JR. 2010. High-affinity consensus binding of target RNAs by the STAR/GSG proteins GLD-1, STAR-2 and Quaking. *BMC Mol Biol* **11:** 48.

Chan RC, Black DL. 1997. The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol Cell Biol* **17:** 4667–4676.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75:** 437–450.

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

Chen T, Richard S. 1998. Structure-function analysis of Qk1: A lethal point mutation in mouse quaking prevents homodimerization. *Mol Cell Biol* **18:** 4863–4871.

Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N, et al. 2012. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes Dev* **26:** 1459–1472.

Chénard CA, Richard S. 2008. New implications for the QUAKING RNA binding protein in human disease. *J Neurosci Res* **86:** 233–242.

Corioni M, Antih N, Tanackovic G, Zavolan M, Kramer A. 2011. Analysis of *in situ* pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res* **39:** 1868–1879.

Crawford JB, Patton JG. 2006. Activation of α-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol* **26:** 8791–8802.

Das D, Clark TA, Schweitzer A, Yamamoto M, Marr H, Arribere J, Minovitsky S, Poliakov A, Dubchak I, Blume JE, et al. 2007. A correlation with exon expression approach to identify *cis*-regulatory elements for tissue-specific alternative splicing. *Nucleic Acids Res* **35:** 4845–4857.

Di Fruscio M, Chen T, Richard S. 1999. Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. *Proc Natl Acad Sci* **96:** 2710–2715.

Dominski Z, Kole R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* **11:** 6075–6083.

Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA, et al. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol* **17:** 187–193.

Ebersole TA, Chen Q, Justice MJ, Artzt K. 1996. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nat Genet* **12:** 260–265.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* **95:** 14863–14868.

Elias JE, Haas W, Faherty BK, Gygi SP. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2:** 667–675.

Eng J, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Am Soc Mass Spectrom* **5:** 976–989.

Francis R, Barton MK, Kimble J, Schedl T. 1995. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139:** 579–606.

Fyrberg C, Becker J, Barthmaier P, Mahaffey J, Fyrberg E. 1997. A *Drosophila* muscle-specific gene related to the mouse quaking locus. *Gene* **197:** 315–323.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat Struct Mol Biol* **12:** 691–698.

Goo Y-H, Cooper TA. 2009. CUGBP2 directly interacts with U2 17S snRNP components and promotes U2 snRNA binding to cardiac troponin T pre-mRNA. *Nucleic Acids Res* **37:** 4275–4286.

Gooding C, Roberts GC, Smith CW. 1998. Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated α-tropomyosin exon. *RNA* **4:** 85–100.

Gromak N, Matlin AJ, Cooper TA, Smith CWJ. 2003. Antagonistic regulation of α-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA* **9:** 443–456.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Huh GS, Hynes RO. 1994. Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. *Genes Dev* **8:** 1561–1574.

Ji S, Ye G, Zhang J, Wang L, Wang T, Wang Z, Zhang T, Wang G, Guo Z, Luo Y, et al. 2012. miR-574-5p negatively regulates *Qki6/7* to impact β-*catenin*/Wnt signalling and the development of colorectal cancer. *Gut* doi: 10.1136/gutjnl-2011-301083.

Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, Hashimoto K, Yasuda K, Inoue K. 2003. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J* **22:** 905–912.

Kalsotra A, Cooper TA. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12:** 715–729.

Kramer A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol Cell Biol* **12:** 4545–4552.

Lakiza O, Frater L, Yoo Y, Villavicencio E, Walterhouse D, Goodwin EB, Iannaccone P. 2005. STAR proteins quaking-6 and GLD-1 regulate translation of the homologues *GLI1* and *tra-1* through a conserved RNA 3′UTR-based mechanism. *Dev Biol* **287:** 98–110.

Lejeune F, Maquat LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* **17:** 309–315.

Downloaded from rnajournal.cshlp.org on April 3, 2013 - Published by Cold Spring Harbor Laboratory Press

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci* **100:** 189–192.
Li Z, Zhang Y, Li D, Feng Y. 2000. Destabilization and mislocalization of myelin basic protein mRNAs in quaking dysmyelination lacking the QKI RNA-binding proteins. *J Neurosci* **20:** 4944–4953.
Lim LP, Sharp PA. 1998. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol Cell Biol* **18:** 3900–3906.
Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M. 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294:** 1098–1102.
Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, de la Grange P, Ast G, et al. 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* **17:** 1114–1123.
Lobbardi R, Lambert G, Zhao J, Geisler R, Kim HR, Rosa FM. 2011. Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Development* **138:** 1783–1794.
Lock P, Fumagalli S, Polakis P, McCormick F, Courtneidge SA. 1996. The human p62 cDNA encodes Sam68 and not the RasGAP-associated p62 protein. *Cell* **84:** 23–24.
Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS, Ohno K. 2012. CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and facilitate mRNA decay. *Sci Rep* **2:** 209.
Modafferi EF, Black DL. 1997. A complex intronic splicing enhancer from the c-*src* pre-mRNA activates inclusion of a heterologous exon. *Mol Cell Biol* **17:** 6537–6545.
Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* **21:** 708–718.
Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457–463.
Ohno G, Hagiwara M, Kuroyanagi H. 2008. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes Dev* **22:** 360–374.
Rodrigues F, Thuma L, Klambt C. 2012. The regulation of glial-specific splicing of Neurexin IV requires HOW and Cdk12 activity. *Development* **139:** 1765–1776.
Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA* **10:** 1449–1458.
Ryder SP, Frater L, Abramovitz DL, Goodwin EB, Williamson JR. 2004. RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nat Struct Mol Biol* **11:** 20–28.
Saccomanno L, Loushin C, Jan E, Punkay E, Artzt K, Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proc Natl Acad Sci* **96:** 12605–12610.
Saldanha AJ. 2004. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20:** 3246–3248.
Sanford JR, Ellis J, Caceres JF. 2005. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem Soc Trans* **33:** 443–446.
Schafer DA, Korshunova YO, Schroer TA, Cooper JA. 1994. Differential localization and sequence analysis of capping protein β-subunit isoforms of vertebrates. *J Cell Biol* **127:** 453–465.
Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* **15:** 183–191.
Sidman R, Dickie M, Appel S. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. *Science* **144:** 309–311.
Southby J, Gooding C, Smith CWJ. 1999. Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of α-actinin mutally exclusive exons. *Mol Cell Biol* **19:** 2699–2711.
Sugnet CW, Srinivasan K, Clark TA, O'Brien G, Cline MS, Wang H, Williams A, Kulp D, Blume JE, Haussler D, et al. 2006. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput Biol* **2:** e4.
Tabb D, McDonald W, Yates JR. 2002. DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1:** 21–26.
Tanaka H, Abe K, Kim CH. 1997. Cloning and expression of the quaking gene in the zebrafish embryo. *Mech Dev* **69:** 209–213.
Tang ZZ, Zheng S, Nikolic J, Black DL. 2009. Developmental control of CaV1.2 L-type calcium channel splicing by Fox proteins. *Mol Cell Biol* **29:** 4757–4765.
Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515.
Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444:** 580–586.
van Mil A, Grundmann S, Goumans MJ, Lei Z, Oerlemans MI, Jaksani S, Doevendans PA, Sluijter JP. 2012. MicroRNA-214 inhibits angiogenesis by targeting Quaking and reducing angiogenic growth factor release. *Cardiovasc Res* **93:** 655–665.
Vernet C, Artzt K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet* **13:** 479–484.
Voelker RB, Berglund JA. 2007. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res* **17:** 1023–1033.
Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al. 2012. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150:** 710–724.
Warf MB, Diegel JV, von Hippel PH, Berglund JA. 2009. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci* **106:** 9203–9208.
Washburn MP, Wolters D, Yates JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242–247.
Wu J, Zhou L, Tonissen K, Tee R, Artzt K. 1999. The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. *J Biol Chem* **274:** 29202–29210.
Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H, et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36:** 996–1006.
Yates J, Ruse CI, Nakorchevsky A. 2009. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annu Rev Biomed Eng* **11:** 49–79.
Yeo GW, Nostrand ELV, Liang TY. 2007. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet* **3:** e85.
Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16:** 130–137.
Zaffran S, Astier M, Gratecos D, Semeriva M. 1997. The *held out wings (how) Drosophila* gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Development* **124:** 2087–2098.
Zearfoss NR, Clingman CC, Farley BM, McCoig LM, Ryder SP. 2011. Quaking regulates *Hnrnpa1* expression through its 3′ UTR in oligodendrocyte precursor cells. *PLoS Genet* **7:** e1001269.
Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* **22:** 2550–2563.
Zhao L, Mandler MD, Yi H, Feng Y. 2010. Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proc Natl Acad Sci* **107:** 19061–19066.
Zorn AM, Krieg PA. 1997. The KH domain protein encoded by *quaking* functions as a dimer and is essential for notochord development in *Xenopus* embryos. *Genes Dev* **11:** 2176–2190.

CHAPTER 4

QUAKING PROMOTES MONOCYTE DIFFERENTIATION INTO PRO-ATHEROGENIC MACROPHAGES BY CONTROLLING PRE-MRNA SPLICING ANG GENE EXPRESSION

Reprinted with permission from

Nature Communications 7:10846 (March 2016)

Copyright © 2016 by Nature Publishing Group

nature COMMUNICATIONS

ARTICLE

Received 17 Jun 2015 | Accepted 26 Jan 2016 | Published 31 Mar 2016

DOI: 10.1038/ncomms10846 OPEN

Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression

Ruben G. de Bruin[1,2], Lily Shiue[3], Jurriën Prins[1,2], Hetty C. de Boer[1,2], Anjana Singh[4], W. Samuel Fagg[3], Janine M. van Gils[1,2], Jacques M.G.J. Duijs[1,2], Sol Katzman[3], Adriaan O. Kraaijeveld[5], Stefan Böhringer[6], Wai Y. Leung[7], Szymon M. Kielbasa[6], John P. Donahue[3], Patrick H.J. van der Zande[1,2], Rick Sijbom[1,2], Carla M.A. van Alem[2], Ilze Bot[8], Cees van Kooten[2], J. Wouter Jukema[5,9], Hilde Van Esch[10], Ton J. Rabelink[1,2], Hilal Kazan[11], Erik A.L. Biessen[4,8], Manuel Ares Jr.[3], Anton Jan van Zonneveld[1,2] & Eric P. van der Veer[1,2]

A hallmark of inflammatory diseases is the excessive recruitment and influx of monocytes to sites of tissue damage and their ensuing differentiation into macrophages. Numerous stimuli are known to induce transcriptional changes associated with macrophage phenotype, but posttranscriptional control of human macrophage differentiation is less well understood. Here we show that expression levels of the RNA-binding protein Quaking (QKI) are low in monocytes and early human atherosclerotic lesions, but are abundant in macrophages of advanced plaques. Depletion of QKI protein impairs monocyte adhesion, migration, differentiation into macrophages and foam cell formation *in vitro* and *in vivo*. RNA-seq and microarray analysis of human monocyte and macrophage transcriptomes, including those of a unique QKI haploinsufficient patient, reveal striking changes in QKI-dependent messenger RNA levels and splicing of RNA transcripts. The biological importance of these transcripts and requirement for QKI during differentiation illustrates a central role for QKI in posttranscriptionally guiding macrophage identity and function.

[1] Einthoven Laboratory of Experimental Vascular Medicine, Leiden University Medical Center, Albinusdreef 2, 2300RC Leiden, The Netherlands. [2] Department of Internal Medicine (Nephrology), Leiden University Medical Center, Albinusdreef 2, C7-36, PO Box 9600, 2300RC, Leiden The Netherlands. [3] Center for Molecular Biology of RNA, Department of Molecular, Cell and Developmental Biology, University of California, 1156 High Street, Santa Cruz, California 95064, USA. [4] Department of Pathology, CARIM, Academic University Hospital Maastricht, P. Debyelaan 25, 6229HX Maastricht, The Netherlands. [5] Department of Cardiology, Leiden University Medical Center, Albinusdreef 2, 2300RC Leiden, The Netherlands. [6] Department of Medical Biostatistics, Leiden University Medical Center, Albinusdreef 2, 2300RC Leiden, The Netherlands. [7] Department of Sequencing Analysis Support Core, Leiden University Medical Center, Albinusdreef 2, 2300RC Leiden, The Netherlands. [8] Division of Biopharmaceutics, Leiden/Amsterdam Center for Drug Research, Leiden University, Einsteinweg 55, 2333CC Leiden, The Netherlands. [9] Durrer Center for Cardiogenetic Research, Meiburgdreef 9, 1105AZ Amsterdam, The Netherlands. [10] Department of Human Genetics, University Hospitals Leuven, Herestraat 43, 3000 Leuven, Belgium. [11] Department of Computer Engineering, Antalya International University, Universit Cad. No.2, Antalya 07190, Turkey. Correspondence and requests for materials should be addressed to E.P.v.d.V. (email: e.p.van_der_veer@lumc.nl).

Monocytes serve as danger sensors within the circulation. The activation of blood-borne monocytes by inflammatory stimuli triggers their adhesion and homing to sites of tissue injury, where they differentiate into macrophages and collectively aid in the resolution of damage[1,2]. However, the chronic accumulation of macrophages at these sites of injury is a hallmark of inflammatory diseases such as rheumatoid arthritis[3], Crohn's disease[4] and atherosclerosis[5–7].

Dynamic changes in gene expression are associated with monocyte to macrophage differentiation, where PU.1 (ref. 8), Signal Transducer and Activator of Transcription (STATs)[9] and CCAAT/Enhancer Binding Protein (C/EBP)s[10] are key transcription factors that drive this alteration in cellular phenotype and function[11,12]. Importantly, numerous studies have identified critical roles for both microRNAs (miRNAs) and RNA-binding proteins (RBPs) in posttranscriptionally regulating monocyte[13] and macrophage[14] biology. However, the posttranscriptional regulation of monocyte to macrophage differentiation has generally been limited to studies detailing miRNA-based targeting of individual transcription factors or effector molecules that either stimulate or delay this phenotypic conversion[15–17].

In contrast to miRNAs, RBPs mediate both quantitative and qualitative changes to the transcriptome, interacting with pre-mRNAs to influence (alternative) splicing, transcript stability, editing, subcellular localization and translational activation or repression[18–20]. This broad arsenal of RNA-based control points enables RBPs to modulate the proteome in response to immunogenic stimuli[17], shifting inflammatory cells from an immature or naive state to a mature or activated state, as has previously been established in lymphoid cells[21,22]. In recent times, we discovered that expression of the RBP Quaking (QKI) is induced in human restenotic lesion-resident vascular smooth muscle cells (VSMCs), where it directly mediates a splicing event in the Myocardin pre-mRNA that governs VSMC function[23]. This finding prompted us to investigate whether QKI could similarly serve as an inflammation-sensitive posttranscriptional guide during monocyte to macrophage differentiation. Alternative splicing of the QKI pre-mRNA yields mature transcripts of 5, 6 or 7 kb that encode distinct protein isoforms, namely QKI-5, -6 and -7 (refs 24,25). QKI-5 possesses a nuclear localization signal in the carboxy-terminal region and is found in the nucleus of cells. In contrast, QKI-6 and QKI-7 are found in the cytoplasm. However, QKI-6 and QKI-7 can also translocate to the nucleus[23,26]. The presence of a KH-family homology domain confers QKI with the capacity to bind RNA[27], albeit dimerization is required[26,28] to bind with high affinity to the QKI response element (QRE) sequence (NACUAAY N1-20 UAAY) on target RNAs[29–33]. Importantly, aberrant QKI expression is associated with inflammatory diseases such as schizophrenia[34,35], cancer[36] and restenosis[23].

Here we show that the RBP QKI plays a critical role in regulating the conversion of monocytes into macrophages in, for example, atherosclerotic lesions. Our studies pinpoint QKI as a dynamic regulator of pre-mRNA splicing and expression profiles that drive monocyte activation, adhesion and differentiation into macrophages, and facilitates their conversion into foam cells.

Results

Human atherosclerotic lesion macrophages express QKI. We previously observed QKI expression in VSMCs[23] and in leukocyte foci within human coronary restenotic lesions. Based on this observation, we used laser-capture microdissection to harvest $CD68^+$ macrophages from early and advanced atherosclerotic lesions of human carotid arteries. QKI mRNA was 4.2-fold enriched in macrophages derived from advanced as compared with early atherosclerotic lesions (Fig. 1a). Next, using immunohistochemistry, we assessed QKI protein expression in human tissue sections at various stages of atherosclerotic lesion development, namely early pathological intimal thickening (PIT), fibrous cap atheroma (FCA) and intraplaque haemorrhaging (IPH). Although QKI was detectable in $CD68^+$ myeloid cells of PIT, it was abundantly expressed in macrophage-rich FCA and IPH lesions (Fig. 1b). Furthermore, QKI-5, -6 and -7 were detectable in the nuclear, perinuclear and cytoplasmic regions of intimal macrophages in both FCA and IPH, respectively (Fig. 1c). We conclude that the accumulation of macrophages in human atherosclerotic lesions is associated with increased mRNA and protein expression of all three QKI isoforms within the macrophage.

A reduction in QKI decreases lesional macrophage burden. To investigate whether decreased QKI expression in monocytes and macrophages could influence atherosclerotic lesion formation, we transplanted bone marrow (BM) from QKI viable (qk^v) mice[37], which express reduced levels of QKI protein, or their wild-type (*wt*) littermate controls (LM) into atherogenic $LDLR^{-/-}$ mice. Although *qk* knockout mice die as embryos, the qk^v mouse harbours a spontaneous ~1 Mb deletion in the *QK* promoter region that leads to reduced levels of QKI mRNA and protein[37]. Indeed, macrophage colony-stimulating factor (M-CSF)-mediated conversion of LM and qk^v BM-derived monocytes to macrophages showed subtly reduced QKI-5 mRNA and protein levels, and almost a complete ablation of QKI-6 and -7 protein (Fig. 1d,e). Following BM transplantation, the $LDLR^{-/-}/qk^v$ and $LDLR^{-/-}$/LM mice were fed a high-fat diet for 8 weeks, to induce atherosclerotic lesion formation. Interestingly, the long-term reduction of QKI expression during haematopoietic reconstitution limited neutrophil and monocytic repopulation (Supplementary Data 1). In keeping with this finding, immunohistochemical analysis of the aortic root revealed significantly decreased monocyte/macrophage content within atherosclerotic lesions of $LDLR^{-/-}/qk^v$ mice (Fig. 1f), a finding that immunohistochemical analysis revealed was independent of plaque size or collagen content. These findings suggested that changes in haematopoietic and monocytic QKI expression could influence the macrophage content of atherosclerotic lesions.

QKI is induced on monocyte to macrophage differentiation. Having identified high QKI expression in macrophages in atherosclerotic lesions, we first explored whether QKI mRNA expression levels differ in macrophage precursors, namely classical ($CD14^{++}/CD16^-$), intermediate ($CD14^{++}/CD16^+$) and non-classical ($CD14^+/CD16^+$) monocytes derived from human peripheral blood (PB)[2]. All three monocyte subpopulations abundantly expressed QKI-5, -6 and -7 mRNAs as compared with glyceraldehyde 3-phosphate dehydrogenase (Fig. 2a). Moreover, QKI-5, -6 and -7 mRNA levels increased as classical monocytes progressed towards intermediate or non-classical monocytes. Interestingly, QKI-5 mRNA was the most abundantly expressed transcript in all three subpopulations. (Fig. 2a).

Next, we assessed QKI mRNA and protein levels in human PB monocytes treated with granulocyte–macrophage CSF (GM-CSF) and M-CSF, to stimulate their conversion to pro-inflammatory and anti-inflammatory macrophages, respectively (Fig. 2b). We observed remarkable increases in the expression of all QKI mRNA transcripts in mature macrophages (Fig. 2c). However, despite abundant QKI mRNA, the distinct QKI isoforms

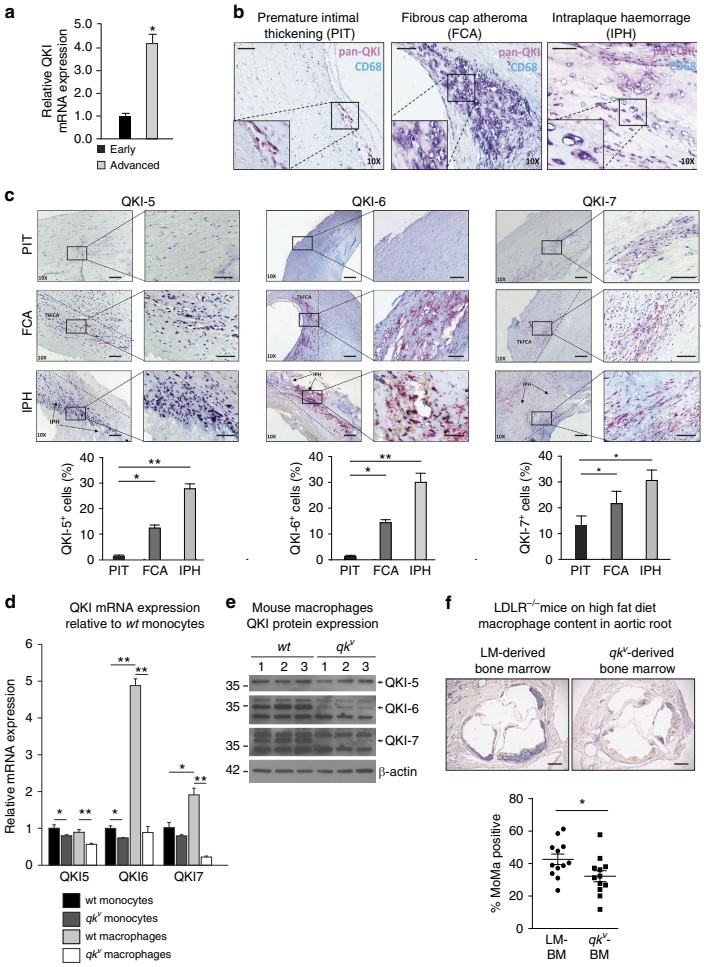


Figure 1 | Quaking is expressed in macrophages within atherosclerotic lesions. (**a**) Pan-QKI mRNA expression levels in CD68$^+$ macrophages of early and advanced atherosclerotic lesions isolated by laser-capture microdissection ($n=4$). Data expressed as mean ± s.e.m.; Student's *t*-test, $^*P<0.05$. Scale bar, 50 μm. (**b**) Immunohistochemical analysis of co-localization of pan-QKI and CD68 expression in preliminary intimal thickening (PIT), FCA and intraplaque haemorrage (PIH). Dashed line denotes intimal/adventitial border. Scale bar, 50 μm. (**c**) Immunohistochemical analysis of QKI-5, -6 and -7 expression in PIT, FCA and IPH (top), and quantification of QKI-positive cells mm^2 per tissue sample ($n=5$). Data expressed as mean ± s.e.m.; one-way analysis of variance (ANOVA), Bonferroni's *post-hoc* test; $^*P<0.05$, $^{**}P<0.01$. (**d**) Quantitative RT–PCR (qRT–PCR) analysis of QKI mRNA expression in naive BM-derived CD115+ mouse monocytes and 7 days M-CSF stimulated macrophages of either *wt*-littermates (LM) or quaking viable (*qk*^{V}) mice ($n=$at least 3 mice per condition). Data expressed as mean ± s.e.m.; one-way ANOVA, Bonferroni's *post-hoc* test; $^*P<0.05$ and $^{**}P<0.01$. (**e**) Western blot analysis of QKI-5, -6 and -7 expression levels in 7 days M-CSF stimulated macrophages derived from BM of *wt* and *qk*^{V} mice. Each lane represents an individual mouse lysate (biological $n=3$). (**f**) Immunohistochemical analysis for atherosclerotic plaque-resident macrophages (% MoMa-positive area) in aortic root sections of γ-irradiated (8 Gy) *LDLR*$^{-/-}$ mice that subsequently were transplanted with BM from either *qk*^{V} mice (*qk*^{V}-BM) or littermates (LM)(LM-BM) and fed a high-fat diet for 8 weeks to develop atherosclerotic lesions ($n=12$ per group). Scale bar, 200 μm. Data expressed as mean ± s.e.m.; Student's *t*-test, with $^*P<0.05$.

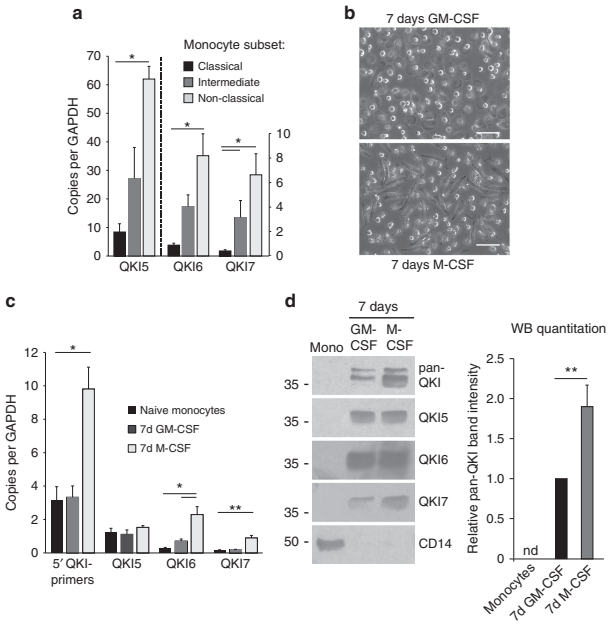


Figure 2 | QKI is highly expressed in macrophages derived from PB monocytes. (**a**) mRNA expression levels of distinct QKI isoforms following negative selection and FACS sorting for blood-derived human monocyte subsets, namely classical ($CD14^{++}$, $CD16^{-}$), intermediate ($CD14^{++}$,$CD16^{+}$) and non-classical ($CD14^{+}$,$CD16^{+}$). Expression is depicted relative to copies per glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Data expressed as mean ± s.e.m.; one-way analysis of variance (ANOVA), Bonferroni's *post-hoc* test; $^{*}P<0.05$ and $^{**}P<0.01$. (**b**) Phase-contrast photomicrographs of human PB monocytes cultured for 7 days in the presence of either GM-CSF or M-CSF. Scale bar, 50 µm. (**c**) Quantitative RT-PCR (qRT-PCR) analysis for QKI mRNA isoforms in naive PB monocytes isolated using $CD14^{+}$ microbeads, 7 days GM-CSF and 7 days M-CSF differentiated macrophages ($n=3$). Expression is depicted relative to copies per GAPDH. Data expressed as mean ± s.e.m.; one-way ANOVA, Bonferroni's *post-hoc* test; $^{*}P<0.05$ and $^{**}P<0.01$. (**d**) Western blot analysis of QKI protein isoforms in naive monocytes, 7 days GM-CSF and 7 days M-CSF differentiated macrophages (pan-QKI and CD14: $n=5$, QKI-5, 6 and 7: $n=1$) with quantification of pan-QKI ($n=5$). Data expressed as mean ± s.e.m.; Student's *t*-test, with $^{**}P<0.01$. Equivalent concentrations of whole-cell lysates were loaded per lane as determined using a BCA protein assay.

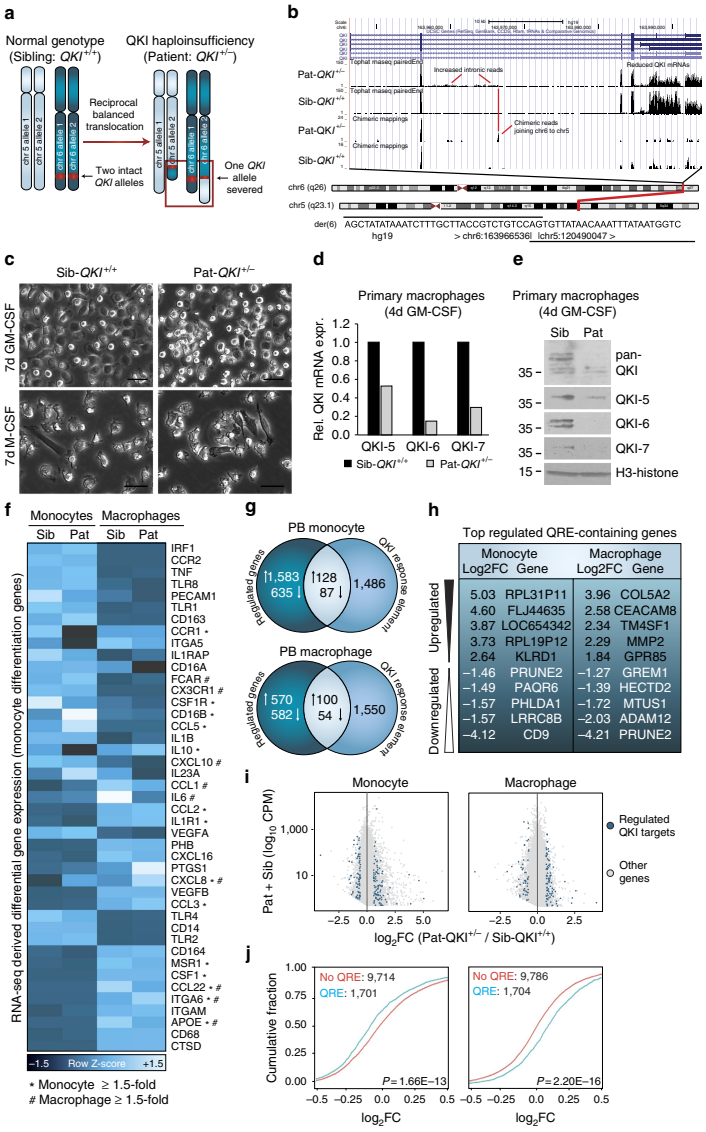
were poorly expressed in freshly isolated PB monocytes as compared with mature macrophages (Fig. 2d). The GM-CSF or M-CSF-induced conversion of monocytes into macrophages was associated with striking increases in QKI-5, -6 and -7 protein levels, with a more pronounced increase in all three isoforms observed with M-CSF treatment (Fig. 2d).

QKI haploinsufficiency perturbs macrophage differentiation. To further assess the role of QKI in monocyte and macrophage biology, we undertook an in-depth analysis of a unique, QKI haploinsufficient individual (Pat-$QKI^{+/-}$) and her sister control (Sib-$QKI^{+/+}$)[38]. This patient is the only known carrier of a balanced reciprocal translocation (t(5;6)(q23.1;q26)), where a breakpoint in one of her *QKI* alleles specifically reduces QKI expression by 50% in both QKI mRNA[38] and QKI protein levels as compared with her sibling (Sib-$QKI^{+/+}$; Fig. 3a,b). RNA sequencing (RNA-seq) analysis (see below) confirmed altered QKI expression and furthermore revealed the precise location of the translocation breakpoint in intron 4 of QKI (Fig. 3b and Supplementary Fig. 1a).

We next compared the circulating monocytes of these two individuals for the expression of well-established monocyte cell surface markers such as CD14, CD16, CX3CR1, CCR2, SELPLG and CSF1R by fluorescence-activated cell sorting (FACS) analysis. Although monocyte subset ratios were not different (Supplementary Fig. 2a), the expression of CSF1R, the receptor that drives macrophage commitment, was elevated in Pat-$QKI^{+/-}$ non-classical monocytes as compared with Sib-$QKI^{+/+}$ (Supplementary Fig. 2b). As CSF1R is normally reduced when monocytes differentiate into macrophages, this observation points towards a potential defect in monocyte maturation in the patient.

Next, we investigated the consequences of decreased QKI expression on monocyte to macrophage differentiation. For this, we obtained freshly isolated Pat-$QKI^{+/-}$ and Sib-$QKI^{+/+}$ monocytes from venous blood and treated the cells for 7 days in the presence of either GM-CSF or M-CSF. Similar to the results in Fig. 2b, Sib-$QKI^{+/+}$ monocytes possessed the capacity to adopt the characteristic pro-inflammatory macrophage morphology, whereas monocytes from Pat-$QKI^{+/-}$ generally retained a monocytic morphology (Fig. 3c top panels). We

harvested RNA and protein from these macrophages and found that both QKI mRNA and protein levels were decreased (Fig. 3d,e). Surprisingly, this reduction in QKI did not appear to have an impact on the conversion of monocytes into anti-inflammatory macrophages (Fig. 3c bottom panels), a finding that prompted us to focus on the role of QKI in monocyte to macrophage differentiation in a pro-inflammatory setting.



QKI impacts transcript abundance in monocytes and macrophages. The observed increase in QKI expression during macrophage differentiation and well-established function as a splicing and translational regulator[23,31,39,40] suggested that QKI is necessary for posttranscriptional control of events that lead to macrophage identity. To identify potential regulatory targets of QKI at a genome-wide level, we characterized the transcriptomes of Sib-$QKI^{+/+}$ and Pat-$QKI^{+/-}$ monocytes and GM-CSF-stimulated macrophages by RNA-seq (Supplementary Data 2). First, we assessed the expression levels of established immune-regulated genes[11,12]. As shown in Fig. 3f, the mRNA levels of many monocyte to macrophage differentiation markers[11,12] were similarly regulated in Sib-$QKI^{+/+}$ and Pat-$QKI^{+/-}$ (*CD68*↑, *ApoE*↑, *ITGAM*↑, *CD14*↓, *CX3CR1*↓ and *CD163*↓). In contrast, the expression levels of several key pro- and anti-inflammatory markers indicated an anti-atherogenic shift in Pat-$QKI^{+/-}$ macrophages (Fig. 3f right; *IL6*↓, *IL23a*↓, *CD16A*↓, *CD16B*↓, *ApoE*↓ and *IL10*↑). At the genome-wide level, QKI haploinsufficiency altered the abundance of 2,433 and 1,306 mRNA species in monocytes and macrophages (Fig. 3g, Supplementary Data 2 and Supplementary Fig. 3 top), respectively. Subsequently, we computationally determined the subset of mature mRNA transcripts in the genome that contain a QKI-binding sequence motif (termed QRE)[30] (Supplementary Data 2). Our data suggested that QKI directly modulates the expression of 215 (128↑ and 87↓) and 154 (100↑ and 54↓) mRNAs in PB monocytes and macrophages, respectively (Fig. 3g, Venn sum of intersect). The five most differentially expressed genes in the patient relative to the sibling that harbour a QRE are shown in Fig. 3h. By selecting genes containing QREs, we identified a substantial number of putative QKI-mediated changes in transcript abundance (Fig. 3i).

Previous genome-wide studies have reported contrasting roles for QKI as both a repressor and stabilizer of target mRNAs[31,33]. Intrigued by this ambiguity, we determined the consequences of QKI haploinsufficiency on mRNA transcript abundance in monocytes and macrophages. For this, we tested whether the presence of a QRE within a target mRNA was associated with increased or decreased mRNA abundance in the patient relative to her sibling (Supplementary Fig. 3 top). For this, we plotted the cumulative distribution fraction (CDF: *y* axis, as a fraction of total genes) against the transcript Log_2FC (*x* axis: Pat-$QKI^{+/-}$/Sib-$QKI^{+/+}$) and stratified for either putative QKI targets (with QRE) or non-targets (no QRE). In PB monocytes, a reduction in QKI was associated with significantly lowered target mRNA expression (Fig. 3j left panel, left shift of cyan line). In contrast, in PB macrophages the expression levels of mature mRNAs containing QREs was strikingly increased in the patient relative to her sibling, as compared with those without QREs (Fig. 3j right panel, right shift of cyan line). Collectively, these studies suggested that QKI potently regulates gene expression during monocyte-to-macrophage differentiation.

QKI controls splicing in monocytes and macrophages. Given previous reports that QKI is involved in splicing of pre-mRNAs[23,39–42], we tested whether QKI acts similarly in monocytes and macrophages. First, we evaluated our RNA-seq analysis of Sib-$QKI^{+/+}$ and Pat-$QKI^{+/-}$ PB monocytes and macrophages for splicing changes (Supplementary Data 3). This analysis uncovered 1,513 alternative splicing events between Pat-$QKI^{+/-}$ and Sib-$QKI^{+/+}$ monocytes and macrophages, revealing events that were unique to either monocytes or macrophages, as well as common events (Supplementary Data 3). Previous observations for QKI and other RBPs suggested that when a splicing factor binds the intron downstream of an alternative exon, it promotes exon inclusion; however, when binding the intron upstream of the alternative exon, the RBP promotes exon skipping[19,43]. We analysed the RNA-seq data for such a trend using the set of splicing events that change between the Pat-$QKI^{+/-}$ and Sib-$QKI^{+/+}$, to determine the frequency of the QKI-binding motif, ACUAA, around these regulated exons, relative to a background set of exons that is expressed, but inclusion is unchanged between the two data sets. The results of these analyses are shown in Fig. 4a and Supplementary Data 4, and demonstrate ACUAA motif enrichment upstream of exons with increasing inclusion in Pat-$QKI^{+/-}$ (QKI repressed exons) relative to background exons, as well as an increase in ACUAA motif frequency downstream of exons with increased skipping (QKI activated) relative to background. This suggested that similar to C2C12 myoblasts[39], QKI promotes exon skipping by binding the upstream intron, while promoting inclusion of alternative exons by binding to the downstream intron, in monocytes and in macrophages. These data strongly support a direct, position-dependent role for QKI in regulating alternative splicing, while also providing additional protein diversity during monocyte-to-macrophage differentiation.

As shown in Fig. 4b, QKI haploinsufficiency triggered alternative splicing events in PB monocytes (orange tracks) and macrophages (blue tracks). Interestingly, the presence of QKI-binding sites, as defined by either experimentally determined QKI PAR-CLIP sites[39] and/or ACUAA motifs clearly

Figure 3 | Characterization of monocyte and macrophage biology in a unique QKI haploinsufficient patient. (**a**) Schematic of chromosomal translocation event in the *qkI* haploinsufficient patient (Pat-$QKI^{+/-}$), reducing QKI expression to ~50% that of her sister control (Sib-$QKI^{+/+}$). (**b**) Top: UCSC Genome Browser display of reference genome QKI locus with standard and chimeric reads for the patient and sibling. The reduced expression levels and altered 3′-untranslated region (UTR) composition in the patient RNA as compared with a sibling control is noteworthy. Patient shows increased intronic RNA extending to the point where chimeric reads map at the breakpoint to chr5. Middle: chromosome diagrams showing normal chromosomes 5 and 6 with the red line, indicating the location of the breakage fusion point. Bottom: sequence across the fusion point. The chromosomal origin of the AG dinucleotide is ambiguous. (**c**) Photomicrographs of sibling and patient macrophages, cultured in GM-CSF or M-CSF for 7 days, respectively. (**d**,**e**) Assessment of QKI isoform mRNA and protein expression in 4-day GM-CSF-stimulated macrophages in the sibling and patient. (**f**) Hierarchical clustering (Euclidean algorithm) of key monocyte differentiation genes depicting changes in RNA-seq-derived mRNA abundance where dark blue = low expression, whereas light blue = high expression (* and/or # indicates ≥1.5-fold expression change in monocytes or macrophages, respectively). (**g**) Venn diagrams with numbers of differentially expressed genes (minimally ±1.5-fold; patient/sibling expression) for unstimulated (top) and GM-CSF stimulated macrophages (bottom). An expression cutoff (Pat + Sib expression ≥1CPM) was applied. (**h**) The most differentially expressed genes, harbouring a QRE are depicted. (**i**) Genome-wide scatterplot of mRNA abundance (*y* axis: Log_{10} CPM) versus the log_2FC (*x* axis: Patient/sibling CPM) after an expression cutoff (Pat + Sib expression ≥1 CPM) in monocytes (left) and GM-CSF-stimulated macrophages (right). Blue dots indicate QRE-containing transcripts minimally ±1.5-fold differentially expressed. Grey dots do not fulfill these criteria. (**j**) CDF (*y* axis) for QKI target (QRE containing: blue line) and non-target (non-QRE containing: cyan line) mRNAs (*x* axis: log_2FC) in monocytes (left) and macrophages (right). Left shift indicates lower expression of QKI target genes, whereas a right shift indicates higher expression of QKI targets in the patient samples. Distributions were compared using a Wilcoxon rank-sum test.

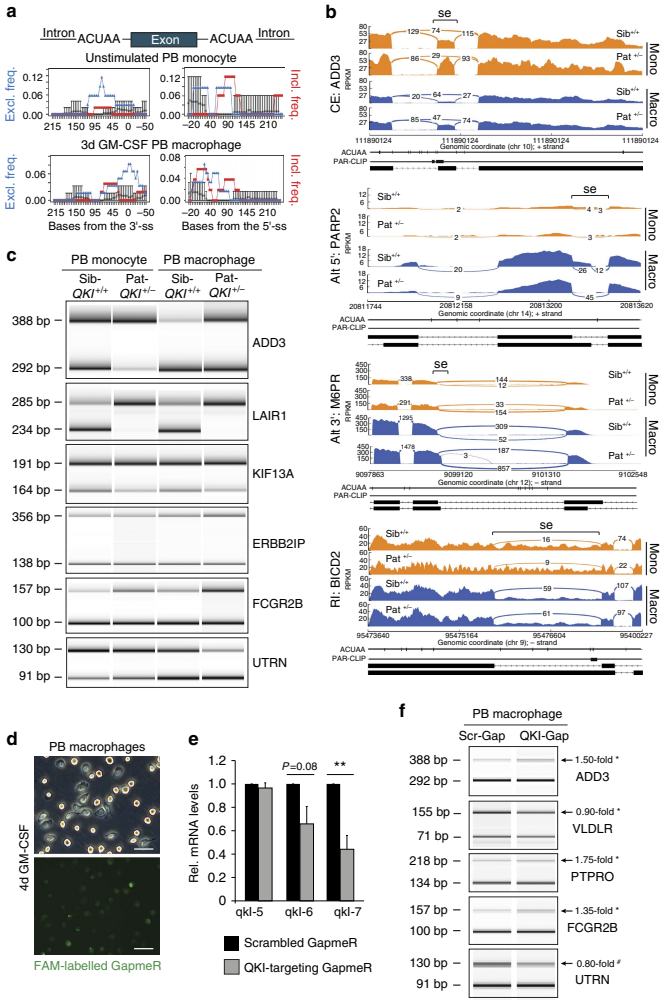


Figure 4 | QKI influences pre-mRNA splicing in naive PB monocytes and macrophages. (**a**) SpliceTrap assessment of the proximal ACUAA RNA motif enrichment in 50-bp windows upstream and downstream of alternatively spliced cassette exons (as compared with a background set of exons; grey circles). The relationship between the frequency of exon exclusion (blue triangles) or exon inclusion (red squares) and ACUAA RNA motif enrichment over the genomic locus are depicted. (**b**) Sashimi plots illustrate RNA-seq read coverage for selected alternative splicing events in Pat-*QKI*$^{+/-}$ versus Sib-*QKI*$^{+/+}$ PB monocytes (orange) and macrophages (blue). Splicing events (se) are highlighted by inverted brackets. The location of ACUAA motifs and QKI PAR-CLIPs are provided below. Splicing events were defined based on the genomic organization of RefSeq transcripts (bottom tracks). Full event details are provided in Supplementary Data 3. (**c**) PCR validation of alternatively spliced cassette exons in Sib-*QKI*$^{+/+}$ and Pat-*QKI*$^{+/-}$ PB-derived monocytes and macrophages. Primers were designed to target constitutive flanking exons. PCR product size for exon inclusion (top) and exclusion (bottom) variants are provided (left). (**d**) Phase-contrast and fluorescence-microscopy photographs (scale bar, 50 μm) of primary human, PB macrophages of healthy controls that have been treated with FAM-labelled GapmeRs, to reduce QKI expression. (**e**) Quantitative RT–PCR (qRT–PCR) of QKI mRNA isoform expression in GapmeR-treated macrophages ($n=3$). Data expressed as mean ± s.e.m.; Student's *t*-test, with $^{**}P<0.01$. (**f**) PCR validation of alternatively spliced cassette exons in GapmeR-treated PB-derived macrophages. Primers were designed to target constitutive flanking exons. PCR product size for exon inclusion (top) and exclusion (bottom) variants are provided (left). A representative illustration is shown of an $n=3$ donors. Data expressed as mean ± s.e.m.; Student's *t*-test, with $^{**}P<0.01$ and $\#P=0.08$.

associated with changes in exon inclusion (for example, *ADD3*), alternative 5′-splice sites (PARP2), alternative 3′-splice sites (M6PR) and intron retention (for example, BICD2), thereby expanding the detection of veritable QKI-regulated events beyond cassette exons (splice event 'se' location defined by brackets). Importantly, strong correlations were observed between QKI expression levels and the magnitude of the splicing event, be it between the patient and sibling control, or between monocytes and macrophages (Fig. 4b).

Finally, we validated several alternatively spliced cassette exons, including events in ADD3, LAIR1 and UTRN by reverse transcriptase–PCR (RT–PCR), using primers in flanking exons (Fig. 4c). Collectively, our RNA-seq data analysis of this unique QKI haploinsufficient individual strongly suggested a direct role for QKI in regulating alternative splicing events that could influence monocyte to macrophage differentiation.

To extend results obtained with the QKI haploinsufficient patient, we abrogated QKI expression in naive primary human monocytes harvested from freshly drawn venous blood of healthy controls. We designed GapmeR antisense oligonucleotides that either targeted QKI for degradation (QKI-Gap), or are scrambled as a control (Scr-Gap), coupled with a 5′-FAM label to track their cellular uptake. The QKI-Gap and Scr-Gap compounds were administered to the freshly isolated monocytes, concomitant with GM-CSF for 96 h, to drive the differentiation to pro-inflammatory macrophages. In contrast to our attempts to reduce QKI mRNA levels using other well-established approaches, we observed virtually no signs of cytotoxicity or apoptosis following GapmeR treatment. Furthermore, the treatment did not hamper the capacity of monocytes to differentiate into macrophages (Fig. 4d top), while uptake efficiency approached 100% (based on FAM^+ cells; Fig. 4d bottom). After 96 h, we harvested RNA from the QKI-Gap- and Scr-Gap-treated macrophages, which yielded a minimal reduction in QKI-5 mRNA levels but remarkable reductions in QKI-6 and QKI-7 mRNAs (Fig. 4e). Albeit that the GapmeR-mediated reduction in QKI expression in primary human macrophages was not as striking as that observed in the QKI haploinsufficient patient, it nonetheless enabled us to visualize and validate significant changes in several of the aforementioned QKI-mediated alternative splicing events, such as *ADD3* and FcγR-IIb (*FCGR2B*) (Fig. 4f; $n=3$ donors). It should be noted that the inability to remarkably reduce the expression of the nuclear QKI isoform, namely QKI-5, could be responsible for the discrepancy between the striking shift in splicing observed in the QKI haploinsufficinet patient as compared with the GapmeR-mediated abrogation of QKI expression. Taken together, these studies clearly pinpoint QKI as a regulator of pre-mRNA splicing during monocyte-to-macrophage differentiation and implicate QKI gene dosage as a determinant of splicing event magnitude.

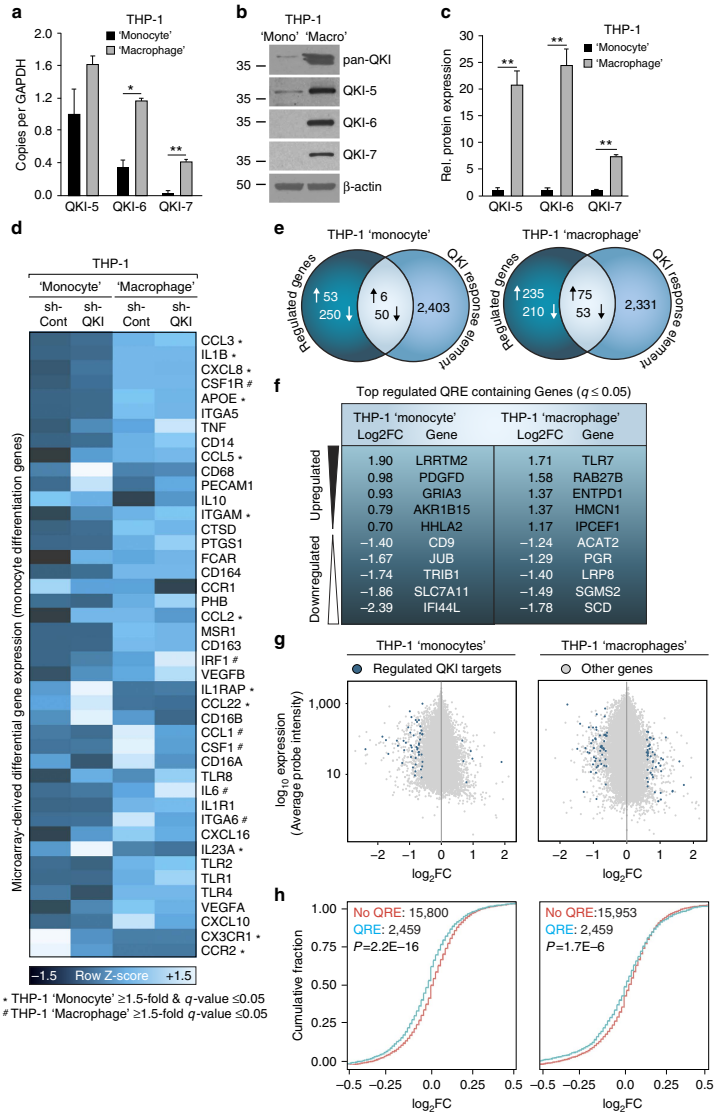
QKI regulates transcript abundance in THP-1 cells. To provide further support for a regulatory role for QKI during monocyte-to-macrophage differentiation, we tested whether QKI could similarly modulate transcript abundance and pre-mRNA splicing in a well-established monocyte cell line, namely THP-1 cells. Similar to GM-CSF-induced differentiation of PB monocytes into macrophages, the 12,13-phorbol myristyl acetate (PMA)-induced transition of THP-1 'monocytes' to 'macrophages' was associated with the following: (1) significantly increased expression of all QKI mRNA transcripts (Fig. 5a); (2) barely detectable levels of QKI protein in THP-1 'monocytes' (Fig. 5b and Supplementary Fig. 4a); and (3) significantly increased expression of QKI protein during THP-1 'monocyte' to 'macrophage' differentiation (Fig. 5b,c). Next, we stably transduced THP-1 'monocytes' with either short-hairpin RNA (shRNA) targeting QKI (sh-QKI) to specifically deplete QKI or with a non-targeting shRNA control (sh-Cont) (Supplementary Fig. 4b). Similar to GM-CSF-stimulated Pat-$QKI^{+/-}$ versus Sib-$QKI^{+/+}$ monocytes, sh-QKI THP-1 'monocytes' displayed an inability to adopt the 'macrophage' morphology following stimulation with PMA as compared with sh-Cont THP-1 'monocytes' (Supplementary Fig. 4c arrows). We subsequently determined mRNA levels using an exon junction microarray[44] analysing RNA isolated from unstimulated and 3 days PMA-stimulated THP-1 sh-Cont and sh-QKI 'monocytes' and 'macrophages' (Supplementary Data 5). Next, as shown in Fig. 5d, we assessed the expression profile of established monocyte differentiation genes (for example, *CD14*↑, *CXCL8*↑, *CSF1R*↑, *ApoE*↑, *CX3CR1*↓, *CCR2*↓ and *CCL22*↓). Similar to QKI haploinsufficient macrophages, several markers in sh-QKI THP-1 'macrophages' displayed an anti-atherogenic phenotypic shift (*IL6*↓, *IL23a*↓, *CD16A*↓, *CD16B*↓, *ApoE*↓ *and IL10*↑) as compared with sh-Cont THP-1 'macrophages' (Fig. 5d).

At the genome-wide level, the reduction of QKI significantly altered the abundance of 359 and 573 mRNAs in THP-1 'monocytes' and 'macrophages', respectively (Fig. 5e, Supplementary Data 5 and Supplementary Fig. 3 bottom). Of these differentially expressed mRNAs, 56 and 128 were computationally predicted QKI targets based on the presence of a QRE in the mature mRNA (Fig. 5e intersect). The most differentially expressed transcripts harbouring a QRE are denoted in Fig. 5f. The expression levels of mRNAs targeted by QKI in THP-1 'monocytes' and 'macrophages' are depicted in Fig. 5g (blue dots) and Fig. 5h (blue lines), relative to those not directly affected by

Figure 5 | QKI influences mRNA transcript abundance during differentiation of THP-1 monocyte-like cells to THP-1 macrophage-like cells. (**a**) mRNA expression of the QKI isoforms as compared with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in THP-1 'monocytes' and 8 days differentiated THP-1 'macrophages' (biological $n=3$). Data expressed as mean ± s.e.m.; Student's *t*-test; $^*P<0.05$ and $^{**}P<0.01$. (**b**) Western blot analysis of whole-cell lysates of THP-1 'monocytes' and THP-1 'macrophages'. (**c**) Western blot quantification of QKI protein isoforms, normalized to β-actin in THP-1 'monocytes' and THP-1 'macrophages' ($n=3$). Data expressed as mean ± s.e.m.; Student's *t*-test; $^{**}P<0.01$. (**d**) Hierarchical clustering (Euclidean algorithm) of key monocyte differentiation genes depicting changes in microarray-derived mRNA abundance THP-1 'monocytes' (left two lanes) and THP-1 'macrophages' (right two lanes), where dark blue = low expression, whereas light blue = high expression (* and/or # beside gene name is indicative of a significant ≥1.5-fold change in expression in monocytes or macrophages, respectively). (**e**) Venn diagrams depicting the number of microarray-derived differentially expressed genes (minimally ±1.5-fold; sh-QKI/sh-Cont expression, q-value ≤ 0.05) for unstimulated THP-1 'monocytes' (left Venn diagram) and THP-1 'macrophages' (right Venn diagram). (**f**) The most significantly differentially expressed genes harbouring a QRE are shown. (**g**) Genome-wide scatterplot of mRNA abundance in THP-1 'monocytes' (left scatterplot) and THP-1 'macrophages' (right scatterplot); *y* axis: Log_{10} probe intensity versus the *x* axis: log2FC: sh-QKI average probe intensity/sh-Cont average probe intensity. Blue dots indicate QRE-containing transcripts that are minimally ±1.5 fold differentially expressed ($q \leq 0.05$). Grey dots do not fulfill these criteria. (**h**) CDF (*y* axis) for QKI target (QRE containing: blue line) and non-target (non-QRE containing: cyan line) mRNAs (*x* axis: log_2FC) in THP-1 'monocytes' (left plot) and THP-1 'macrophages' (right plot). Left shift indicates lower expression of QKI target genes in the sh-QKI samples, whereas a right shift is indicative of higher expression of QKI targets in the sh-QKI samples. Distributions were compared using a Wilcoxon rank-sum test.

changes in QKI levels (Fig. 5g grey dots and Fig. 5h cyan lines). Consistent with our analyses in PB monocytes, putative direct QKI target mRNAs were mostly reduced on a targeted QKI reduction in THP-1 'monocytes' (Fig. 5h, left plot), although a shift towards increased target mRNA abundance in THP-1 'macrophages' was not observed (Fig. 5h, right plot).

QKI modifies pre-mRNA splicing patterns in THP-1 cells. Having identified that QKI haploinsufficiency generates pre-mRNA splicing events that probably have an impact on monocyte and macrophage biology, we also analysed RNA isolated from sh-Cont and sh-QKI THP-1 'monocytes' and 'macrophages' for alternative splicing events using the exon



THP-1 'monocyte' Log2FC	Gene	THP-1 'macrophage' Log2FC	Gene
1.90	LRRTM2	1.71	TLR7
0.98	PDGFD	1.58	RAB27B
0.93	GRIA3	1.37	ENTPD1
0.79	AKR1B15	1.37	HMCN1
0.70	HHLA2	1.17	IPCEF1
−1.40	CD9	−1.24	ACAT2
−1.67	JUB	−1.29	PGR
−1.74	TRIB1	−1.40	LRP8
−1.86	SLC7A11	−1.49	SGMS2
−2.39	IFI44L	−1.78	SCD

junction microarray platform[44]. This highly sensitive technology uses probes that are designed specifically to detect both constitutive exon–exon junctions and alternative exon–exon junctions, enabling one to quantify inclusion ratios for alternative splicing events. These studies uncovered 571 and 629 differentially regulated alternative splicing events in THP-1 'monocytes' and 'macrophages', respectively, including numerous cassette exons, alternative 5′- and 3′-splice sites, and retained introns (Fig. 6a and Supplementary Data 6; $n=3$). Detected splicing events are illustrated in Fig. 6b, where the skip and include intensities (y axis and x axis, respectively) of transcript-specific hybridization probes directed to either the constitutive or alternatively spliced exons are plotted. The separation score, obtained by determining slope differences, indicates the magnitude of the splicing event. Similar to the motif enrichment analyses performed for the RNA-seq of PB monocytes and macrophages, these studies confirmed that exon skipping frequency was significantly correlated with alternative exons that had an ACUAA motif in the upstream intron (Fig. 6c left panels and Supplementary Data 4). In contrast to the subtle enrichment of inclusion frequency observed in Pat-$QKI^{+/-}$ and Sib-$QKI^{+/+}$ monocytes and macrophages (Fig. 3a right panels), exon inclusion frequency in THP-1 'monocytes' and 'macrophages' was clearly associated with the presence of ACUAA motifs in the downstream intron (Fig. 6c and Supplementary Data 4).

Finally, alternative cassette exons in THP-1 'monocytes' and 'macrophages' were PCR validated (Fig. 6d). Importantly, we also selected several top splicing events from THP-1 'monocytes' and 'macrophages' (Supplementary Data 6), and validated these in RNA harvested from *wt* and qk^{v} mice, including REPS1, PTPRO and FGFR1OP2 (Fig. 6e).

QKI targets monocyte activation and differentiation pathways. We subsequently determined how QKI-induced changes in mRNA transcript abundance could have an impact on Gene Ontology (GO) enrichment of coordinately regulated pathways during monocyte-to-macrophage differentiation. As shown in Table 1 and Supplementary Data 7, these GO analyses point towards a central regulatory role for QKI in immune responses to injury, processes that play a critical role in the onset and development of atherosclerosis and other inflammation-based diseases. In both monocytes and macrophages, changes in QKI expression clearly had an impact on Liver X Receptor (LXR)/Retinoid X Receptor (RXR) activation and Peroxisome Proliferator-Activated Receptor (PPAR) activation and signalling, implicating a key role for QKI in regulating cholesterol biosynthesis and metabolism. Furthermore, a reduction in QKI expression also appeared to influence T-cell and Toll-like receptor signalling, biological processes that play prominent roles in the rapid resolution of infection, while in chronic settings exacerbate inflammatory conditions. Finally, the gene enrichment analysis suggested that posttranscriptional processing of factors driving the recruitment, adhesion and diapedesis of immune cells were affected by changes in QKI expression.

QKI facilitates monocyte adhesion and migration. Our experimentally determined changes in (pre)-mRNA splicing and expression, as well as bioinformatically predicted changes in biological processes, prompted us to evaluate whether these QKI-induced posttranscriptional modifications could affect monocyte and macrophage function. To test this, we first assessed whether cell survival is affected by a reduction of QKI expression in THP-1 'monocytes'. Importantly, the cumulative population doublings and apoptotic rates were not affected by decreased QKI levels (Fig. 7a,b). Next, we assessed cell adhesion to glass coverslips treated with effector molecules (collagen and activated platelets) in the presence of fluid shear stress, an experimental design that mimics the response of monocytes to endothelial denudation in the vessel[45]. Live-cell imaging clearly showed that the shRNA-mediated depletion of QKI in THP-1 'monocytes' reduced cellular adhesion under flow conditions, as evidenced by their continued rolling along the substrate and inability to firmly attach (Fig. 7c and Supplementary Movies 1 and 2). This firm adhesion of monocytes is aided by the activation of β1-integrins on the cell surface that mediate high-affinity interactions with the extracellular matrix at sites of injury[36]. We tested whether QKI depletion had an impact on β1-integrin function by incubating sh-Cont and sh-QKI THP-1 'monocytes' with an antibody (TS2/16) that forces β1-integrins into the activated, adhesive conformation[37]. Interestingly, the abrogation of QKI did not affect monocyte adhesion properties in this setting (Fig. 7d), indicating that proper integrin expression and functionality is not dependent on QKI.

We subsequently tested whether QKI expression levels could have an impact on monocyte migration *in vitro* by seeding sh-QKI or sh-Cont THP-1 'monocytes' into transwell migration chambers and assessed their ability to migrate towards the chemoattractant formyl-methionyl-leucyl-phenylalanine (fMLP). Indeed, depletion of QKI in monocytes inhibited migration (Fig. 7e). This finding prompted us to similarly assess the capacity of Pat-$QKI^{+/-}$ and Sib-$QKI^{+/+}$ monocytes freshly isolated from venous blood to migrate to macrophage chemoattractant protein 1, a physiologic recruiter of monocytes at sites of vascular injury. These studies revealed a significant reduction in transwell migration for Pat-$QKI^{+/-}$ monocytes (Fig. 7e), validating our findings in THP-1 'monocytes', and provided evidence that QKI influences monocyte adhesion and migration in inflammatory settings.

QKI drives foam cell formation. As QKI expression remarkably increased during monocyte-tomacrophage differentiation (Fig. 2c–f) and our aforementioned GO analysis revealed a strong association for changes in QKI expression and lipid metabolism (Fig. 7a), we tested whether a reduction in QKI expression influences the handling of lipids. For this, we first assessed the mRNA expression levels of a subset of established lipid-related genes in monocytes and macrophages derived from WT and qk^{v} mice. As shown in Fig. 8a, monocytes from qk^{v} mice are characterized by significant reductions in *NR1H3* (known as *LXRα*) and *PPARG* (*PPARγ*) expression, as well as cholesterol uptake (*CD36* and *LDLR*) and efflux (*ABCG1*) receptors, as compared with WT monocytes. These effects were diminished on conversion to macrophages (Fig. 8a).

We subsequently assessed the expression levels of these lipid metabolism/homeostasis genes in human PB-derived monocytes and macrophages (Fig. 8b and Supplementary Fig. 5). Similar to qk^{v} monocytes, Pat-$QKI^{+/-}$ monocytes were characterized by decreased *NR1H3* and *PPARG* expression, as well as LDLR and *SCARB1* (Fig. 8b). In contrast to qk^{v} monocytes, *ABCG1* expression was potently increased. Similar to qk^{v} macrophages, this differential gene expression profile appeared to normalize in Pat-$QKI^{+/-}$ macrophages as compared with Sib-$QKI^{+/+}$ macrophages (Fig. 8b). Moreover, in primary human macrophages where GapmeR-mediated knockdown of QKI was realized, we observed significant increases in *MYLIP/IDOL* and *ABCG1* expression, whereas *CD36* displayed a trend towards decreased expression (Supplementary Fig. 5).

Having identified that changes in QKI expression levels had an impact on lipid-associated gene expression, we

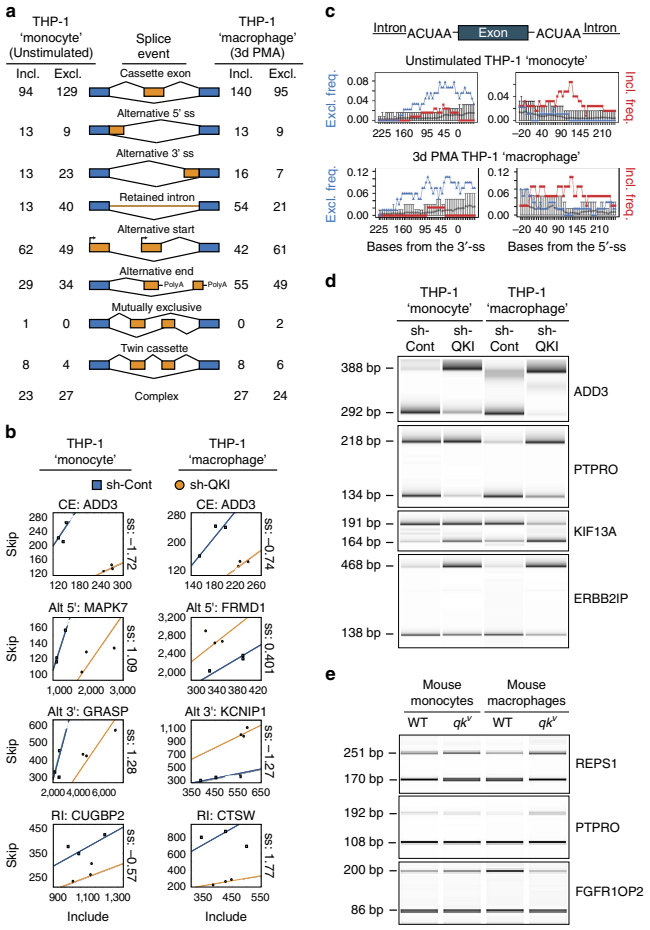


Figure 6 | QKI expression levels influence pre-mRNA splicing during THP-1-based monocyte-like to macrophage-like cell differentiation. (**a**) Schematic depicting detectable alternative splicing events with the splicing-sensitive microarray platform and number of inclusion (incl.; top lines) or exclusion (excl.; bottom lines) events observed in unstimulated THP-1 'monocytes' (left) and 3-day PMA-stimulated THP-1 'macrophages' ($n=3$, $q\leq 0.05$). (**b**) Scatterplots of skip (*y* axis) and include (*x* axis) probe set intensity for selected alternative splicing events in sh-Cont (blue boxes) versus sh-QKI (orange circles) in unstimulated and 3 days PMA-stimulated THP-1 'monocytes' and 'macrophages', respectively. Regression coefficients (constrained to pass the origin) are depicted as solid lines. The $\log_2$ difference in the slopes (termed separation score; ss) are provided to the right of the plots for each event, with for example, an ss of -1.72, indicating a 3.3-fold more inclusion of ADD3 exon 13 in sh-QKI versus sh-Cont THP-1 'monocytes'. Full event details are provided in Supplementary Data 6. CE, cassette exon; Alt 5′ or 3′, alternative 5′ or 3′ splice site; RI, retained intron. (**c**) SpliceTrap assessment of average proximal ACUAA RNA motif enrichment in 50 bp windows upstream and downstream of alternatively spliced cassette exons as compared with a background set of exons (grey circles). The relationship between the frequency of exon exclusion (blue triangles) or exon inclusion (red squares) and ACUAA RNA motif enrichment are depicted. (**d**) PCR validation of alternatively spliced cassette exons in sh-Cont and sh-QKI THP-1 'monocytes' and 'macrophages'. Primers were designed to target constitutive flanking exons. PCR product size for exon inclusion (top) and exclusion (bottom) variants are provided (left). All experiments depict biological $n=3$. (**e**) PCR validation of three splicing events in *wt* and qk^v mouse-derived primary monocytes and 7 days M-CSF-stimulated macrophages. PCR product size for exon inclusion (top) and exclusion (bottom) variants are provided (left). Depicted is a representative PCR for at least a biological $n\geq 3$.

Table 1 | IPA assessment of pre-defined canonical pathways affected by changes in QKI expression.

Monocytes			Macrophages		
THP-1 sh-QKI versus sh-Cont			**THP-1 sh-QKI versus sh-Cont**		
Affected canonical pathway	**−Log (*P*-value)**	**Affected genes**	**Affected canonical pathway**	**−Log (*P*-value)**	**Affected genes**
Atherosclerosis signalling	9.2	*CXCL8, APOE, ICAM1, PDGFA, PLA2, G4C, CCR2, F3, LYZ, CCL2, ORM1, APOC1, IL1B, ORM2, PDGFD, TNF*	Superpathway of cholesterol biosynthesis	10.6	*FDPS, PDFT1, EBP, DHCR7, ACAT2, IDI1, HSD17B7, MSMO1, HMGCS1, CYPS1A1*
Superpathway of cholesterol biosynthesis	8.2	*MVD, FDPS, CHCR7, ACAT2, HSD17B7, MSMO1, HMGCS1,CYPS1A1*	Cholesterol biosynthesis I, II, and III	8.1	*FDFT1, EBP, DHCR7, DHCR24, HSD17B7, MSMO1, CYPS1A1*
LXR/RXR activation	7.4	*SCD, APOE, LYZ, ORM1, CCL2, APOC1, IL1B, ORM2, CD14, PTGS2, IL1RAP, TNF, CYPS1A1*	Superpathway of gernanylgeranylphosphate Biosynthesis I	4.4	*FDPS, ACAT2, IDI1, FNTB, HMGCS1*
Hepatic fibrosis/hepatic stellate cell activation	6.1	*CXCL8, APOE, ICAM1, PDGFA, PLA2, G4C, CCR2, F3, LYZ, CCL2, ORM1, APOC1, IL1B, ORM2, PDGFD, TNF*	LXR/RXR activation	4.4	*SCD, FDFT1, LYZ, IL1A, LDLR, IL36RN, NR1H3, IL6, CLU, CYPS1A1, IL36B, AGT*
PPAR signalling	5.8	*PPARG, JUN, PPARD, PDGFA, MRAS, IL1B, PTGS2,PDGFD, TNF, IL1RAP*	Altered T-cell and B-cell signalling in rheumatoid arthritis	4.3	*IL1A, CSF1, IL36RN, TLR6, TLR8, TLR7, IL6, CSF2, IL36B, IL17A*
RNA-seq Pat-QKI versus Sib-QKI			**RNA-seq Pat-QKI versus Sib-QKI**		
Affected canonical pathway	**−Log (*P*-value)**	**Affected genes**	**Affected canonical pathway**	**−Log (*P*-value)**	**Affected genes**
T-cell receptor signalling	8.9	*CD247, PTPN7, CAMK4, PRKCQ, CD3E, PLCG1, CD8A, CD3D,CD8B, CD28, CD3G, LCK, TXK, ZAP70, ITK*	Granulocyte adhesion and diapedesis	4.9	*CXCL8, IL1A, HRH2, MMP7, SDC1, PPBP, ITGA6, RDX, CCL24, CCL17, MMP2, CCL22, C5, FPR1, CCL13, ICAM2, IL1RN, MMP19, ITGA4*
CCR5 signalling in macrophages	7.8	*CD247, CD3G, CCR5, CAMK4, PRKCQ, CCL4, CD3E, PLCG2, PLCG1, CCL3, CD3D, GNG10*	Agranulocyte adhesion and diapedesis	4	*CXCL8, MMP7, IL1A, PPBP, ITGA6, RDX, CCL24, CCL17, MMP2, CCL22, C5, MYL9, CCL13, ICAM2, IL1RN, PODXL, MMP19, ITGA4*
Role of NFAT in regulation of the immune response	7	*CD247, CAMK4, PRKCQ, CD3E, GCER1A, PLCG1, CD3D, GNG10, CD28, CD3G, LCK, GNAT1, PLCG2, ZAP70, FCGR3A/GCGR3B, FCGR1B, ITK*	Toll-like receptor signalling	3	*MAP2K6, IL1A, TICAM2, IL1RN, TLR7, MAPK13, TLR3, IRAK2, TRAF1*
EIF2 signalling	5.8	*RPL24, RPL36A,* RPS3A, *RPS27, RPL17, RPS18, RPS10, RPL39, RPL12, RPL7A, RPL7, RPL9, RPS28, RPL23A, RPL39L, RPSA*	Cysteine biosynthesis/ homocysteine degradation	2.9	*CBS/CBSL, CTH*
iCOS-iCOSL signalling in T-helper cells	5.7	*CD247, CD3G, CD28, LCK, CAMK4, PRKCQ, CD3E, ZAP70, PLCG1, CD3D, ICOSLG/LOC102723996, ITK*	Axonal guidance signalling	2.9	*SLIT3, ERAP2, MMP7, SLIT1, PDGFA, SEMA6A, BCAR1, TUBB2B, EPHB1, TUBA8, MYSM1, PRKAR1B, GNB1L, WNT5B, ITGA4, SEMA3G, PAK4, ADAM15, TUBA4A, MMP2, KEL, MYL9, FZD4, ADAM12, SEMA4G, SEMA7A, FZD7*

IPA, Ingenuity Pathway Analysis; QKI, Quaking.
The top five affected canonical pathways are shown, along with their respective −log(*P*-value) and the genes that are affected within the particular pathway. Full IPA output is provided in Supplementary Data 7.

investigated whether lipid loading affected QKI expression levels. Indeed, treatment with either acetylated low-density lipoprotein (acLDL) or β-very low-density lipoprotein (β-VLDL) led to significant increases in QKI-5 mRNA levels, while QKI-6 and QKI-7 levels also increased, albeit not significantly (Fig. 8c). In contrast to primary monocytes and macrophages, THP-1 'monocytes' did not display significant changes in lipid metabolism gene expression. However, as shown in Fig. 8d, treatment with modified LDL increased expression of cholesterol uptake genes (*CD36* and *VLDLR*), along with significant increases in cholesterol efflux genes (*ABCA1* and *ABCG1*). Taken together, these studies suggested that changes in QKI expression could have an impact on the net balance of genes that control lipid metabolism and homeostasis.

Finally, we tested whether these QKI-mediated changes in lipid-associated gene expression could translate into consequences for lipid uptake and foam cell formation, a phenomenon tightly associated with pro-inflammatory macrophage function[7]. As shown in Fig. 8e, the impact of decreased QKI expression on foam cell formation on loading with β-VLDL was clear, as sh-QKI THP-1 'macrophages' displayed less extensive lipid staining as compared with sh-Cont THP-1 'macrophages' (Fig. 8e). Similarly, in Pat-$QKI^{+/-}$ macrophages we observed significantly less lipid staining after β-VLDL treatment (Fig. 8f). Even more striking was the potent decrease in oxidized LDL (oxLDL) loading, an atherosclerosis-relevant antigen, in Pat-$QKI^{+/-}$ macrophages (Fig. 8f). Collectively, these studies strongly suggested that the posttranscriptional processing of (pre-) mRNA transcripts by QKI is essential for the physiologic functioning of monocytes and macrophages in disease settings such as atherosclerosis.

Discussion

Genes involved in regulating the transition of monocytes into pro-inflammatory macrophages serve as excellent therapeutic targets for limiting the progression of inflammation-driven diseases such as rheumatoid arthritis and atherosclerosis[3,6]. Our data indicate that alongside wide-ranging changes in gene expression, the differentiation of monocytes to macrophages requires extensive alternative splicing of pre-mRNA species and

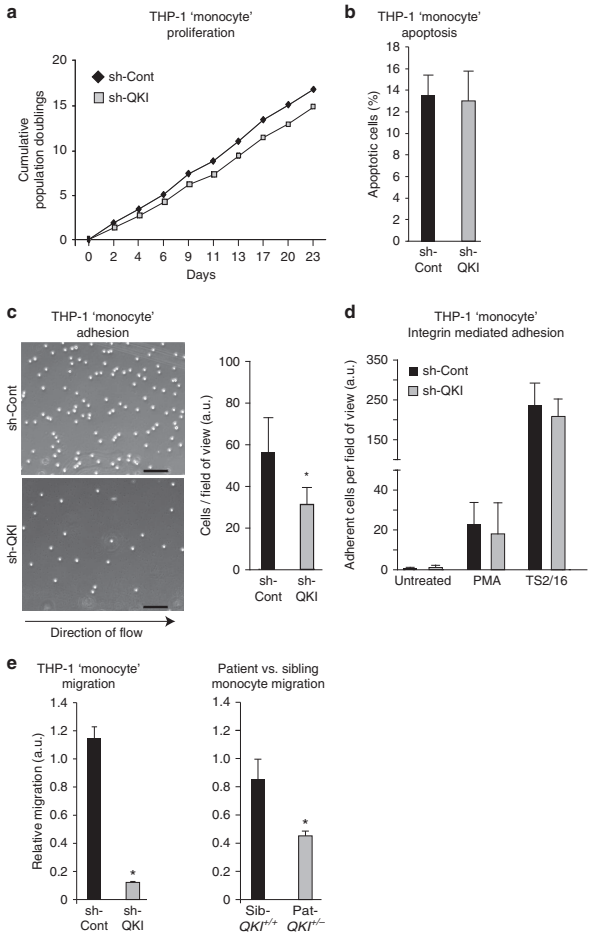


Figure 7 | QKI expression levels have an impact on monocyte adhesion as well as migration and differentiation. (**a**) Cumulative population doublings (*y* axis: CPDs) were counted to assess the effect of QKI reduction on cellular proliferation over time (*x* axis: days). Population growth curves were compared using linear regression analysis.(**b**) Quantification of cellular apoptosis, where annexin V^+ and propidium iodide $^+$ cells were categorized as apoptotic, as determined by FACS analysis. (**c**) Quantification of sh-Cont and sh-QKI THP-1 'monocyte' adhesion to collagen matrix pretreated with platelet-rich plasma under flow, mimicking *in-vivo* endothelial denudation. Direction of flow is indicated below the photomicrographs ($n=3$). Data expressed as mean ± s.e.m.; Student's *t*-test; $^*P<0.05$. Scale bar, 100 μm. (Also see Supplementary Movies 1 and 2). (**d**) Assessment of integrin-mediated adhesion. Quantification of adhesion to collagen for untreated, PMA- or TS2/16-treated sh-Cont and sh-QKI THP-1 'monocytes' are plotted. TS2/16 is an antibody that turns all β1-integrins in the high-affinity conformation, inducing cellular adhesion. (**e**) Quantification of cellular transwell migration towards either fMLP (for THP-1 'monocytes') or macrophage chemoattractant protein 1 (MCP-1; for PB monocytes from either sibling or patient ($n=4$ technical replicates). Data expressed as mean ± s.e.m.; Student's *t*-test; $^*P<0.05$ and $^{**}P<0.01$.

pinpoint QKI as a novel posttranscriptional regulator of both of these processes (Fig. 9).

Expression of the transcription factor PU.1 is associated with the activation of gene expression profiles that drive the differentiation of $CD34^+$ haematopoietic progenitor cells towards a myeloid fate, including monocytes and macrophages[46,47]. Recent work by Pham *et al.*[48] identified that the binding of PU.1 appears to be enhanced by cooperativity with neighbouring transcription factor binding sites, such as KLF4. Importantly, PU.1 induces the expression of critical monocyte

transcription factors, including KLF4 (ref. 49). Of note, KLF4 has been demonstrated to bind to the QKI promoter region of VSMCs, which is GC rich[50]. Furthermore, chromatin immunoprecipitation sequencing data derived from HL-60 cells embedded in the UCSC Encode browser revealed two experimentally determined PU.1-binding sites in the QKI promoter[51] (Supplementary Fig. 6). Collectively, these findings suggest that PU.1, potentially in concert with KLF4, could be responsible for driving QKI expression during monocyte-to-macrophage differentiation.

Our findings also suggest that the abundant expression of QKI mRNAs in naive monocytes could serve to prime these cells with the capacity to rapidly respond to pro-inflammatory triggers at sites of injury. Although we did not assess the translational kinetics of QKI mRNA into protein, our investigation of monocyte activation, adhesion and differentiation strongly suggests that the determination of pre-mRNA fate by QKI critically has an impact on the capacity of the monocyte to aid in response to vascular injury. To date, the genome-wide (alternative) splicing patterns during monocyte-to-macrophage differentiation had not been described. However, the consequences of many splicing events herein described, such as γ-adducin (*ADD3*), *FCGR2B* and *VLDLR* are unknown. However, given that phosphorylation of the γ-adducin C terminus triggers dissociation from spectrin and cortical actin loss[52], it is plausible that the QKI-mediated exclusion of a 13 amino acid coding exon proximal to this region could have an impact on cytoskeletal dynamics in monocytes and macrophages. Furthermore, alternative splicing of exon 6 in *FCGR2B* could potentially have an impact on the inhibitory role of this protein in monocyte- and macrophage-mediated phagocytosis at sites of vascular injury[53]. Therapeutic strategies tailored to target such splicing events in monocytes could potentially deter their conversion to disease-accelerating macrophages.

The expanding repertoire of RNA species have led to the emergence of RNA-based therapeutics as a novel means of treating both rare and common diseases, such as muscular dystrophy and cancer, respectively[54,55]. This is based on extensive efforts geared towards identifying how changes in coding (alternative splicing) and non-coding RNAs (miRNA and long-non-coding RNA) have an impact on cellular pathophysiology[56]. Importantly, the fate of these coding and non-coding RNAs at the cellular level are determined primarily by the more than 500 RBPs that regulate eukaryotic cell biology[32]. Recently, the RNA motifs to which a significant portion of these RBPs bind has been characterized[32], enabling the systematic identification of RNA targets for a given RBP within cells, including QKI[29–32]. Our genome-wide evaluation of posttranscriptional events mediated by QKI implicate both direct and indirect posttranscriptional roles for this protein, where the absence of ACUAA motifs or QREs could nonetheless involve QKI, potentially by tethering to other RBPs, or through QKI-mediated changes in the expression of other RBPs[57]. Moreover, in spite of the presence of QREs within target mRNAs, monocytes and macrophages probably express a large variety of RBPs that compete with QKI for access to either identical or similar motifs with varying affinities within a short stretch of RNA nucleotides[19], which could preclude the observation of a posttranscriptional event.

In conclusion, we have identified QKI as a critical posttranscriptional regulator of pre-mRNA splicing and transcript abundance in monocytes and macrophages. We propose that the RBP-induced reprogramming of the posttranscriptional landscape could generate novel targets for the effective attenuation of inflammatory diseases.

Note added in proof: After the acceptance of our paper, we were informed by Dehghan *et al.*[58] of the CHARGE Consortium's identification that single nucleotide polymorphisms proximal to QKI significantly associated with myocardial infarction and coronary heart disease risk.

Methods

Human immunohistochemistry studies. Early lesions are defined as fatty streaks or PIT, whereas advanced plaques consist of both FCA and IPH (fibroatheroma with early-stage or late-stage necrotic core). Scoring of plaque stage, based on characteristics such as thin cap fibroatheroma (vulnerable or ruptured plaques), vascularization, IPH and/or thrombi/fibrin deposits, were scored by a trained pathologist.

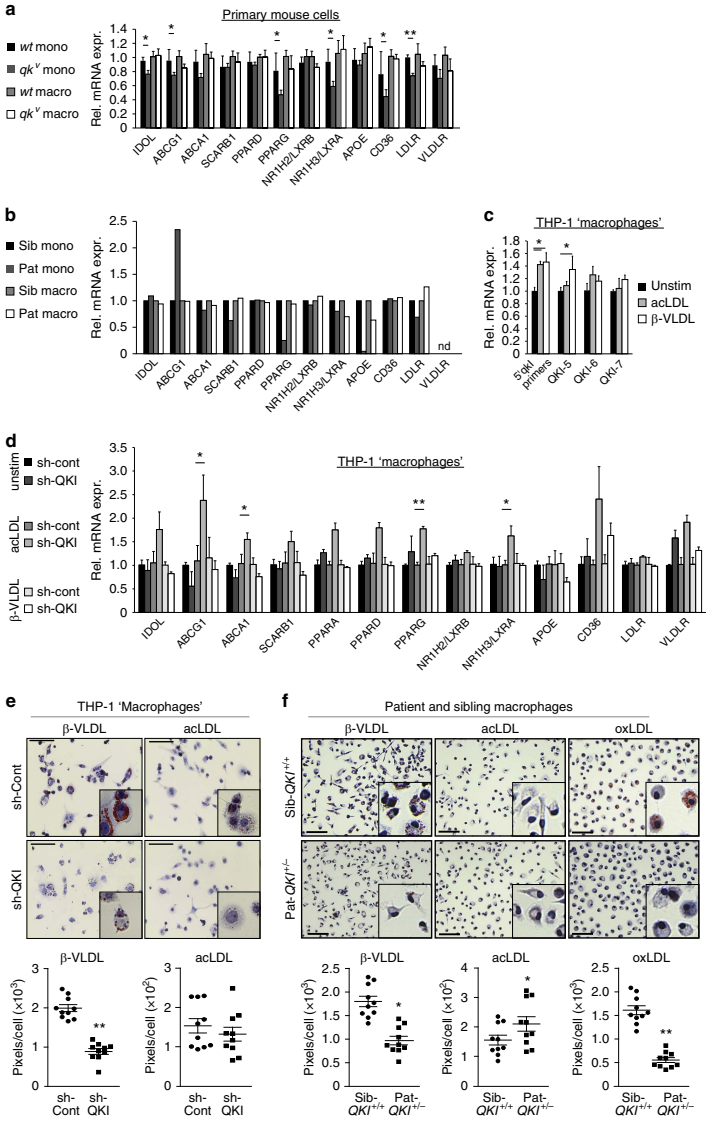
Paraffin tissue sections from human carotid arteries were deparaffinized and rehydrated. After pre-treatment with TE buffer (pH 9 for QKI-5 and QKI-7) or citrate buffer (pH 6 for QKI 6) for antigen retrieval, sections were incubated overnight at 4 °C with primary mouse-anti-human pan-QKI (10 μg ml^{-1}, clone N147/6; UC Davis/NIH NeuroMab Facility, Davis, CA, USA), mouse anti-human QKI-5 (10 μg ml^{-1}, clone N195A/16; NeuroMab/Antibodies, Inc.); mouse anti-human QKI-6 (10 μg ml^{-1}, clone N182/17; NeuroMab/Antibodies Inc.) or mouse anti-human QKI-7 (10 μg ml^{-1}, clone N183/15; NeuroMab/Antibodies Inc.) diluted in Tris-buffered saline containing 1% BSA and 0.1% Tween 20. Subsequently, sections were washed in Tris-buffered saline and incubated with a secondary biotinylated sheep-anti-mouse antibody (GE Healthcare, Eindhoven, The Netherlands). Next, the sections were incubated with streptavidin ABC-alkaline phosphatase (Vector Laboratories, Peterborough, UK) and colour was developed using the Vector Red staining kit (Vector Laboratories), followed by haematoxylin counterstaining. No primary antibody was used for the negative control. QKI/CD68 co-localization immunostaining was achieved using CD68 (Dako-KP1, DakoCytomation) and pan-QKI antibodies, with CD68 and QKI being counterstained with Vector Blue and Vector Red, respectively. For quantification, slides were analysed in a blinded manner using a Leica DM3000 light microscope (Leica Microsystems) coupled to computerized morphometry (Leica Qwin 3.5.1).

BM transplantation. Male *LDLR*$^{-/-}$ mice were housed in sterile filter-top cages and fed a chow diet (Special Diet Services, Witham, Essex, UK). Drinking water was infused with antibiotics (83 mg l^{-1} ciprofloxacin and 67 mg l^{-1} polymyxin B sulfate) and 6.5 g l^{-1} sugar and was provided *ad libitum*. BM transplantation studies were performed as previously described with minor modifications[59]. Briefly, to induce BM aplasia, 10- to 12-week-old recipient mice were exposed to a single dose of 9 Gy (0.19 Gy min^{-1}, 200 kV, 4 mA) total body irradiation, using an Andrex Smart 225 Röntgen source (YXLON International, Copenhagen, Denmark) with a 6-mm aluminium filter, 1 day before transplantation. After 24 h, BM cell suspensions were prepared by flushing the femurs of ~12-week-old *qk*^{v} mice or age-matched LM controls (C57/Bl6-J background; Jackson Laboratories, Bar Harbor, USA) with PBS, after which 5×10^6 cells were injected into the tail vein of recipient mice. After 8 weeks of recovery, the

Figure 8 | QKI regulates the expression of atherosclerosis-related mRNAs and impairs foam cell formation. (**a**) Quantitative RT-PCR (qRT-PCR) analysis of established atherosclerosis-related genes in *wt* of *qk*^{v}-derived monocytes and macrophages. Gene expression in *qk*^{v} samples are relative to either *WT* monocytes or *WT* macrophages ($n \geq 3$). Data expressed as mean ± s.e.m.; Student's *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$. (**b**) RNA-seq-derived expression values to illustrate the expression of established atherosclerosis-related genes in sibling or patient monocytes and macrophages. (**c**) qRT-PCR analysis of QKI isoform mRNA expression in unstimulated THP-1 'macrophages', or treated with β-VLDL or acLDL for 24 h. Data expressed as mean ± s.e.m.; one-way analysis of variance (ANOVA), Bonferroni's *post-hoc* test; $^{*}P < 0.05$. (**d**) qRT-PCR analysis of well-known atherosclerosis-related genes in sh-cont or sh-QKI THP-1 'macrophages' that were either left untreated or treated with acLDL or β-VLDL to induce foam cell formation ($n = 3$). Data expressed as mean ± s.e.m.; Student's *t*-test; $^{*}P < 0.05$ and $^{**}P < 0.01$. (**e**) Photomicrographs of an Oil-red-O staining to assess β-VLDL, acLDL uptake in either sh-Cont or sh-QKI THP-1 'macrophages'. Scale bar, 100 μm. Inset is a high-magnification image of lipid-droplet accumulation. Data expressed as mean ± s.e.m.; Student's *t*-test; $^{**}P < 0.01$. (**f**) Photomicrographs of an Oil-red-O staining to assess β-VLDL, acLDL or oxidized LDL (oxLDL) uptake in either Sib-*QKI*$^{+/+}$ (upper panels) or Pat-*QKI*$^{+/-}$ (lower panels) macrophages that were first differentiated for 7 days with GM-CSF. Scale bar, 100 μm. Inset is a high-magnification image of lipid-droplet accumulation. Data expressed as mean ± s.e.m.; Student's *t*-test; $^{*}P < 0.05$ and $^{**}P < 0.01$.

mice were placed on a Western-type diet, containing 0.25% cholesterol and 15% cacao butter (Special Diet Services) for 8 weeks ($n=12$ per group). Immunohistochemical analysis and quantification of the aortic root using anti-Monocyte and Macrophage-2 (MOMA2) antibody (Sigma-Aldrich) for MOMA2-positive cell area (expressed as a percentage of total plaque area), VSMC content (smooth muscle α-actin-positive cells) and collagen (picosirius red staining) was performed in a blinded manner. Haematologic chimerism of the transplanted $LDLR^{-/-}$ mice was validated using the following primers: *qk* forward primer 5′-TGTGACTTGGGGACTGTCAA-3′; *qk* reverse primer 5′-AAAGGGAAAATTTAGCAACAA-3′.

BM-derived *WT* and qk^{v} monocytes were isolated using $CD115^{+}$ antibody coupled to magnetic beads (Miltenyi Biotech, Leiden, The Netherlands) and differentiated for 7 days to macrophages using mouse recombinant M-CSF (PeproTech, Hamburg, Germany) in RPMI 1640 medium (Gibco, Bleiswijk, The



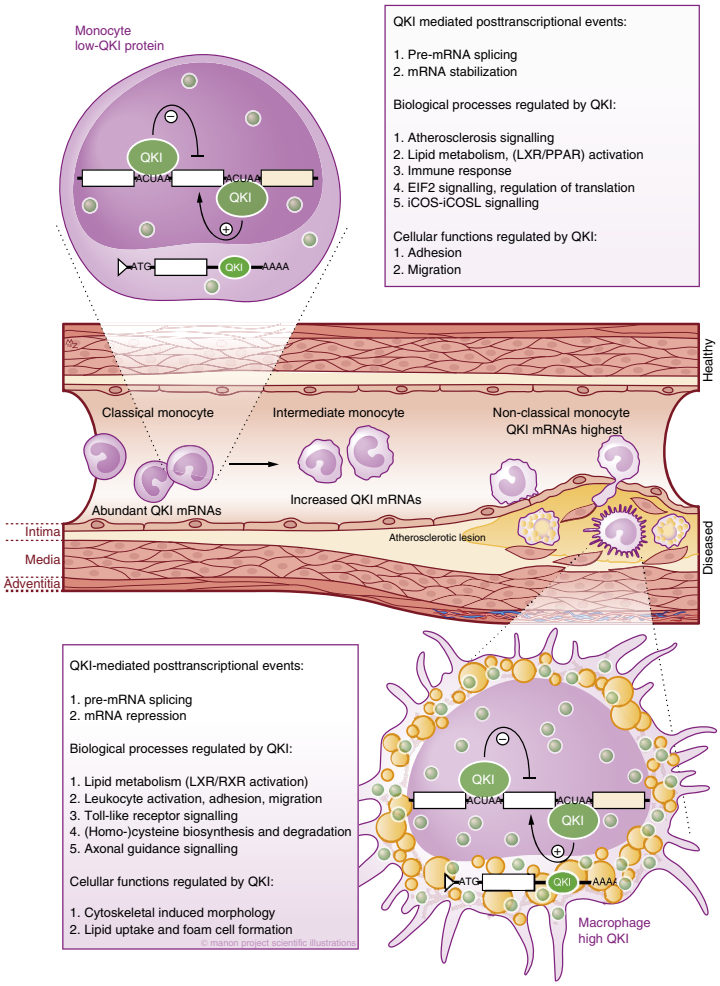


Figure 9 | Schematic depicting how QKI posttranscriptionally regulates monocyte to macrophage differentiation and atherosclerosis development. QKI mRNA expression increases in intermediate monocytes, reaching a peak in the non-classical monocyte (middle). Monocytes adhere to the endothelium at sites of tissue injury, leading to their activation and migration into the subendothelial space. This process requires QKI, as the targeted ablation of QKI impaired monocyte adhesion and migration, and the evident transition in cellular phenotype requires extensive reprogramming of the posttranscriptional landscape. On tissue entrance, the monocyte differentiates into a macrophage, a conversion that was associated with a potent increase in QKI protein. This increase potentiates the interaction of QKI with (pre-)mRNA targets, enhancing splicing and target mRNA repression. The loss of QKI in macrophages results in an inability to adopt the macrophage phenotype and a perturbation of lipid uptake and foam cell formation.

Netherlands) containing 10% FCS (Bodinco, Alkmaar, The Netherlands) and 0.01 μg ml^{-1} glutamine, 50 units per ml penicillin and 50 μg ml^{-1} streptomycin.

Lentiviral transduction of monocytes. Human THP-1 'monocytes' (ATCC, Manassas, VA, USA) were transduced with lentiviral particles encoding either sh-QKI or sh-Cont (catalogue number: SHC202; MISSION library, Sigma-Aldrich). Stable transductants were selected using 3 μg ml^{-1} puromycin (Sigma-Aldrich) for 72 h.

GapmeR design. A single-stranded RNA–DNA hybrid antisense oligonucleotide (GapmeR) was designed to target exon 2 of QKI (Eurogentec, Maastricht, The Netherlands), along with a scrambled GapmeR control. Both

the GapmeR and scrambled control were 22 nucleotides in length and consisted of RNA (A,C,G,U) and DNA (dA, dC, dG, dT) in a 6-10-6 manner, with a phosphorothioate backbone (*) and a 6-FAM label on the 5′-end. GapmeR sequences were as follows:5′-A*C*A*U*G*U*dC*dT*dT*dT*dC*dC*dG*dT*dA*dC*U*C*U*G*C*U-3′ for the QKI-targeting GapmeR and 5′-U*G*C*C*U*C*dT*dC*dT*dC*dG*dT*dA*dC*dC *dG*U*A*U*U*U*A-3′ for the scrambled control.

Monocyte subpopulation analysis. Human monocytes were derived from healthy donor buffy coats (Ethical Approval Number BTL 10.090) following Ficoll density gradient centrifugation and isolated from the PB mononuclear cell fraction using a negative selection cocktail to isolate unlabelled monocytes (StemCell Technologies, Grenoble, France). Purified monocytes were subsequently incubated with 1 μg ml^{-1} CD14-FITC and 1 μg ml^{-1} CD16-Pc5 (Beckman Coulter, Woerden, The Netherlands) for 30 min at 4 °C and FACS sorted using a FACSCalibur (BD Biosciences, Breda, The Netherlands). RNA was isolated from the subpopulations using Trizol reagent (Thermo Fisher Scientific, Bleiswijk, The Netherlands).

Monocyte and macrophage culture. Human monocytes were isolated from buffy coats of healthy donors with an antibody to CD14, conjugated to magnetic beads to allow for MACS sorting (Miltenyi Biotech). Cells were cultured in RPMI media supplemented with 10% FCS and 0.01 μg ml^{-1} glutamine at 37 °C and 5% CO_2. Differentiation of primary CD14^{+} monocytes to pro-inflammatory macrophages was achieved by stimulating with 5 ng ml^{-1} GM-CSF (Thermo Fisher Scientific) or 5 ng ml^{-1} M-CSF (Miltenyi Biotech). Human THP-1 'monocytes' were cultured in RPMI media supplemented with 10% FCS and 0.01 μg ml^{-1} glutamine at 37 °C and 5% CO_2, with differentiation into macrophages being induced by treating with 100 nM PMA[60]. TS/216 for integrin activation experiments was kindly provided by Dr Arnoud Sonneberg, Netherlands Cancer Institute, Amsterdam, The Netherlands. Foam-cell formation was assessed by treating sh-Cont and sh-QKI THP-1 'macrophages' or Sib-QKI$^{+/+}$ and Pat-QKI$^{+/-}$ macrophages for 24 h with either 25 μg ml^{-1} β-VLDL or 50 μg ml^{-1} acLDL, or 10 μg ml^{-1} oxLDL, after which the cells were fixed with 10% Formafix, and Oil-Red-O stained and haematoxylin counterstained. Oil-Red-O area per field of view was divided by the number of cells using ImageJ software.

***In vitro* perfusion assay.** Glass coverslips were coated with type I collagen, after which the system was perfused with platelet-rich plasma for 10 min. Next, the system was flushed with flow buffer (20 mmol l^{-1} HEPES, 132 mmol l^{-1} NaCl, 6 mmol l^{-} KCl, 1 mmol l^{-1} $MgSO_4$, 1.2 mmol l^{-1} KH_2PO_4, 5 mmol l^{-1} glucose, 1.0 mmol l^{-1} CaCl2 and 0.5% BSA) for 2 min. Sh-Cont and sh-QKI THP-1 were resuspended in flow buffer at a concentration of 4×10^6 ml^{-1}, after which the cells were perfused over the substrate for 5 min at 1 dyne cm^{-2}. Cellular adhesion was tracked visually using a Leica DMI5000 microscope. The flow rate was subsequently increased to 2 dynes cm^2 for 2 min, followed by the visual assessment of firm adhesion of perfused monocytes for a duration of 3 min, after which photomicrographs of ten random fields of view in the fluidic chamber were taken and quantified.

Cellular migration assays. Transwell cellular migration studies of sh-Cont and sh-QkI THP-1 'monocytes' towards 1 nM *N*-formyl-methionyl-leucyl-phenylalanine (fMLP, Sigma-Aldrich) or 10 ng ml^{-1} macrophage chemoattractant protein 1 (R&D Systems, Abingdon, UK) for human primary monocytes were performed as previously described[61]. Briefly, cell migration was assessed using Corning Transwell polycarbonate membrane cell culture inserts (6.5 mm transwell with 5.0 μm pore size, Sigma). The lower chamber was loaded with RPMI containing 0.25% BSA and desired chemoattractant. Wells containing no chemoattractant were used as negative controls. Cells were loaded in the upper chamber of the transwell inserts (10^5 cells) and incubated overnight at 37 °C. The following day, cell migration (adherent cells) was quantified by manual counting or ImageJ.

Gene expression microarrays. RNA was isolated from unstimulated sh-Cont and sh-QkI THP-1 'monocytes' (day 0), and from sh-Cont and sh-QkI THP-1 cells stimulated with 100 nM PMA for 3 days using Trizol (Thermo Fisher Scientific) and the RNeasy Mini Kit (Qiagen, Heidelberg, Germany) according to the manufacturer's instructions. RNA quantity and quality was measured using a NanoDrop spectrophotometer (Nanodrop Technologies, Wesington, USA) and an Agilent 2100 bioanalyser (Agilent Technologies, Santa Clara, USA). Samples meeting a RNA integrity number critieria of >8 were used for further analysis.

Splicing microarrays. Microarray data is deposited in GEO under the accession number GSE74887. Targets were prepared from three replicate cultures for each sample. Complementary DNA synthesis and amplification was performed using the WT Expression Kit (Ambion, Bleiswijk, The Netherlands). Samples were enzymatically fragmented and biotinylated using the WT Terminal Labeling Kit (Affymetrix, Santa Clara, California, USA). Labelled target was hybridized to the HJAY Chip (Affymetrix 540091). Chips were washed and scanned using the Fluidics Station 450 and Affymetrix Gene ChIP Scanner 3000 7G (Affymetrix). Data were analysed as previously described[62]. Briefly, in the absence of mismatch probes on these microarrays, probe intensities were first used to construct an empirical CDF, which was subsequently used to calculate an empirical *P*-value that a particular probe's intensity arose from the background of all probes. Probes were stratified for GC content (thermodynamically favourable GC base pairing). For each probe set, the median *P*-value of the set of individual probes in the probe set was used as the *P*-value for that probe set. Before assessing for alternatively spliced transcripts from a particular locus, we first determined whether the gene was expressed. Next, if the expression criteria was met, we determined whether RNA containing two or more alternative splice junctions was detectable using the isoform-specific probes. In situations where the probe sets for two or more alternative isoforms were 'present' in any sample of the data, an alternative splicing event was scored. For these events, the junction probe sets that hybridized to individual isoforms were identified and the probes they contained were used for the Kruskal–Wallis test. Next, we normalized individual probe intensities and grouped the replicate values. Subsequently, using R software, the kruskal.test function was used to test the hypothesis that the probe intensities come from identical populations. If the resulting *P*-value was small enough, the null hypothesis was rejected and the alternative hypothesis that the probes were differentially expressed was accepted. To determine an appropriate value for the 0.01 significance level, 12,740 Kruskal–Wallis tests on randomly selected probe sets were performed, yielding an α-value that associated with the 1% quantile of randomly selected probes (1.975486×10^{-3}). To account for multiple testing, a Bonferroni-corrected α-value of $1.975486 \times 10^{-3}/1.2740 \times 10^4 = 1.550617 \times 10^{-7}$ was used as a *P*-value cutoff for significance. The Sepscore is log2 (Include/Skip ratio) of the experimental sample over the reference sample. Exon inclusion generates a positive Sepscore, whereas exon exclusion generates a negative Sepscore.

Word (5-mer) enrichment and positional mapping. Counts of all 5-mers (ACUAA) in the selected region of an exon set are compared with their counts from a background set of sequences using Fisher's exact test, with multiple testing correction. For motif mapping, we plotted the frequency of specific motifs in 50-nucleotide windows slid along the intron sequences upstream and downstream of each selected exon set with 5-nucleotide sampling intervals. At each interval, we computed the average number of motifs in exons activated or repressed in shRNA-treated cells, and background exons in the same experiment with no splicing change. For the background set, we empirically estimated the 95% confidence interval of motif frequency (error bars). For peaks of consecutive points outside the 95% interval, we applied the Mann–Whitney–Wilcoxon test to estimate a *P*-value that the motif frequency at each point within the peak is greater than background. As the points were not independent, we estimated a *q*-value for each peak by finding the most significant *P*-value for any point in the peak and applying Bonferroni correction for the number of points within the peak. A second Bonferroni correction controlled for the number of possible positions at which a peak might occur, yielding the final *q*-values. To explain further; the effect of QKI depletion on every assessed exon was calculated from the RNA-seq or the splicing-sensitive microarray: that is, whether a reduction in QKI causes inclusion or exclusion from the final mature mRNA transcript.

Next, the intronic regions around differentially spliced exons were analysed for the presence of ACUAA motifs (5-mers). By doing this for every alternatively spliced exon, we could detect an enrichment of ACUAA motifs in the upstream introns of the exons that were included in a 'QKI-deficient' cell: QKI is not binding upstream; thus, the exon is included, as measured by RNA-seq or splicing-sensitive microarray.

In contrast, we could detect an enrichment of ACUAA motifs in the downstream introns of the exons that are included on QKI reduction. QKI binding in the downstream intron would normally give inclusion, but now QKI cannot bind downstream; thus, exclusion is now favoured as assessed by the RNA-seq or splicing-sensitive microarray.

RNA-seq library preparation. RNA-seq data are deposited in GEO under the accession number GSE74979. For each sample, the non-ribosomal fraction of 3 μg of total RNA was isolated using a Ribo-Zero rRNA removal Kit (Epicentre, Madison, Wisconsin, USA). Ribo-Zero-treated RNAs were used to generate barcoded cDNA libraries using the TruSeq RNA Sample Preparation kit (Illumina), with the following additions. Size selections were performed before and after cDNA amplification on an E-gel Safe Imager (Invitrogen) using 2% E-gel SizeSelect gels (Invitrogen). The cDNA fraction of 300 bp in size (including adapters) was isolated and purified. Indexed libraries were pooled and sequenced (paired-end 50 or 100 bp reads) on the Illumina HiSEQ platform to a depth of 41–70 million reads per sample (QB3 Vincent J. Coates Genomics Sequencing Laboratory). After removal of PCR duplicates and repeats, there were 21–26 million uniquely mapping paired-end reads (37–61%).

Mapping and analysis of RNA-seq data. All samples were mapped using Tophat2 (refs 63,64) with Bowtie2 (ref. 64) as the underlying alignment tool. The input Illumina fastq files consisted of paired-end reads with each end containing 100 bp (except for 2 samples with 50 bp of paired-end reads). For equivalency,

100-bp reads were trimmed to 50 bp before mapping. The target genome assembly for the human samples was GRCh37/UCSC-hg19 and Tophat was additionally supplied with a gene model (using its '-GTF' parameter) with data from the hg19 UCSC KnownGenes track[65]. For multiple-mapped fragments, only the highest scoring mapping determined by Bowtie2 was kept. Only mappings with both read ends aligned were kept. Potential PCR duplicates (mappings of more than one fragment with identical positions for both read ends) were removed with the samtools 'rmdup'[66] function, keeping only one of any potential duplicates. The final set of mapped paired-end reads for a sample were converted to position-by-position coverage of the relevant genome assembly using the bedtools 'genomeCoverageBed' function[67]. To determine the count of fragments mapping to a gene, the position-by-position coverage was summed over the exonic positions of the gene. This gene total coverage was divided by a factor of 100, to account for the 100-bp of coverage induced by each mapped paired-end fragment (50 bp from each end) and rounded to an integer. This was calculated for each gene in the UCSC Known Gene set. For input to DESeq[68], all genes with non-zero counts in any sample were considered. Two replicates of each sample were combined per the DESeq methodology. SpliceTrap[69] was used to analyse splicing changes with the parameters j15, ch0.1, ir0.1 and IRMyes.

For mapping of reads that are chimeric to the reference genome[70], to identify the translocation breakpoint on chr6 we used STAR-Fusion https://github.com/STAR-Fusion/. Non-splice junction reads from the macrophage samples that mapped from the QKI gene on chr6 to any location on chr5 were inspected and several were found, which mapped from QKI intron 4 to a site that is strongly transcribed from chr5 in the patient but not at all in her sibling.

Genome-wide computational analysis for RNA motifs. Human mature mRNA sequences were downloaded from UCSC RefSeq database (hg19). Computational screening was performed for the QRE motif, UACUAAY N1-20 UAAY, and counts calculated for the longest annotated transcript. After cross-referencing these transcripts with either the transcripts annotated on the microarray or the RNA-seq, we annotated the number or QREs in the Supplementary Data. Transcripts of which we were unable to assess whether they contain one or more QREs, we annotated as NA and these were excluded from analysis to generate the CDF plots in Figs 3j and 5h. To generate the Venn diagrams and scatterplots for the RNA-seq of monocytes and macrophages (Fig. 3g,i), a ± 1.5-fold change cutoff was applied together with a minimal expression cutoff of patient + sibling ≥ 1 CPM, to avoid artificially large fold changes due to very low expression values. To generate the Venn diagrams and scatterplots for the THP-1-derived expression data (Fig. 5e, g), we applied a ± 1.5-fold change cutoff and applied a DESeq-derived q-value cutoff of $q \leq 0.05$.

MISO analysis. Mixture of isoforms (MISO) analysis was used to assess, quantify and visualize alternative transcripts based on RNA-seq data. Sequences obtained from RNA-seq were aligned using TopHat2 to the human genome v19 transcriptome (annotation-set kindly provided by Dr Christopher Burge, MIT, Cambridge, USA). Using the alignment files (BAM files), MISO analyses was performed on our RNA-seq paired-end sequencing data to identify alternative splicing events, as previously described[71]. Second, an annotation set containing only exons surrounding the splicing event were included in the analysis to generate a more in-depth analysis of select splicing events. A pairwise comparison was performed using the sibling and patient monocyte and macrophage on both the full and selected annotation sets. Additional visualization was performed using the sashimi plot subpackage from MISO, while modifications in the plotting procedure were made to allow visualization of supplementary annotation tracks including ACTAA and QKI PAR-CLIP sites, along with RefSeq transcripts that define the event at (https://github.com/wyleung/rnaveer).

Western blot analysis. Polyacrylamide gel electrophoresis was used to resolve proteins from cellular lysates harvested in RIPA buffer. Protein determinations (BCA) were performed to ensure equal loading of protein on a per-sample basis. QKI-5, -6 and -7 were detected using primary mouse monoclonal antibodies that target pan-QKI (1:1,000, N73/168; UC Davis/NIH NeuroMab Facility), QKI-5 (1:1,000, N195A/16; UC Davis/NIH NeuroMab Facility), QKI-6 (1:1,000, N182/17; UC Davis/NIH NeuroMab Facility) or QKI-7 (1:1,000, N183/15.1; UC Davis/NIH NeuroMab Facility), or polyclonal antibodies targeting QKI-5 (1:1,000, AB9904; Millipore, Amsterdam, The Netherlands), QKI-6 (1:2,000, AB9906; Millipore) and QKI-7 (1:2,000, AB9908; Millipore). For loading references, rabbit polyclonal antibodies were used to detect β-actin or Histone H3 (both 1:4,000; Abcam, Cambridge, UK). All gels were run and blotted with Bio-Rad TGX pre-cast gels and blotted on nitrocellulose 0.2 µm using the Bio-Rad TurboBlot system (Bio-Rad Laboratories). Full blots are shown in Supplementary Fig. 7.

Quantification of pre-mRNA expression levels by PCR. RNA was harvested from monocytes and macrophages using Trizol reagent (Thermo Fisher Scientific). Standard mouse and human cDNA was made using oligo-dT primers (Invitrogen), whereas cDNA for alternative splicing studies was synthesized using random primers (Invitrogen). Primer sets designed for specific pre-mRNA amplification by quantitative RT–PCR analysis are provided in Supplementary Table 1. Quantitative RT–PCR analysis for designated mRNA products was performed using SYBR Green master mix (Bio-Rad, Veenendaal, The Netherlands). For optimal resolution of pre-mRNA splicing patterns, samples were run on an Agilent 2100 bioanalyser (Agilent), with full images provided in Supplementary Figs 8–10.

Statistics. For all experiments, N defined the number of biological replicates. All *in vitro* and *in vivo* results were analysed using GraphPad software with either a Student's t-test or analysis of variance (with a Bonferonni post test being used). All results are expressed as mean ± s.e.m. Differences in P-values < 0.05 or < 0.01 were considered significant and indicated as follows: $^{*}P < 0.05$ or $^{**}P < 0.01$, respectively.

Ethics. Informed consent was obtained for all patient-derived samples. Approval for these studies was provided by the relevant medical ethics committees, namely Maastricht University Medical Center, The Netherlands (Professor Dr E.A.L. Biessen), for immunohistochemical studies, and Leuven University Hospital, Belgium (Professor Dr H. Van Esch), for the QKI haploinsufficient and control materials. All mouse experiments were approved by the regulatory authorities of the Leiden University and were in compliance with the Dutch Government Guidelines.

References

1. Auffray, C., Sieweke, M. H. & Geissmann, F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu. Rev. Immunol.* **27,** 669–692 (2009).
2. Ziegler-Heitbrock, L. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood* **116,** e74–e80 (2010).
3. van Zonneveld, A. J., de Boer, H. C., van der Veer, E. P. & Rabelink, T. J. Inflammation, vascular injury and repair in rheumatoid arthritis. *Ann. Rheum. Dis.* **69**(Suppl 1): i57–i60 (2010).
4. Schwarzmaier, D., Foell, D., Weinhage, T., Varga, G. & Dabritz, J. Peripheral monocyte functions and activation in patients with quiescent Crohn's disease. *PLoS ONE.* **8,** e62761 (2013).
5. Libby, P., Nahrendorf, M., Pittet, M. J. & Swirski, F. K. Diversity of denizens of the atherosclerotic plaque: not all monocytes are created equal. *Circulation* **117,** 3168–3170 (2008).
6. Ross, R. Atherosclerosis—an inflammatory disease. *N. Engl. J. Med.* **340,** 115–126 (1999).
7. Hilgendorf, I., Swirski, F. K. & Robbins, C. S. Monocyte fate in atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **35,** 272–279 (2015).
8. Zhang, D. E., Hetherington, C. J., Chen, H. M. & Tenen, D. G. The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Mol. Cell Biol.* **14,** 373–381 (1994).
9. Coccia, E. M. *et al.* STAT1 activation during monocyte to macrophage maturation: role of adhesion molecules. *Int. Immunol.* **11,** 1075–1083 (1999).
10. Williams, S. C. *et al.* C/EBPepsilon is a myeloid-specific activator of cytokine, chemokine, and macrophage-colony-stimulating factor receptor genes. *J. Biol. Chem.* **273,** 13493–13501 (1998).
11. Martinez, F. O., Gordon, S., Locati, M. & Mantovani, A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *J. Immunol.* **177,** 7303–7311 (2006).
12. Martinez, F. O. The transcriptome of human monocyte subsets begins to emerge. *J. Biol.* **8,** 99 (2009).
13. Lu, J. Y., Sadri, N. & Schneider, R. J. Endotoxic shock in AUF1 knockout mice mediated by failure to degrade proinflammatory cytokine mRNAs. *Genes Dev.* **20,** 3174–3184 (2006).
14. Lu, Y. C. *et al.* ELAVL1 modulates transcriptome-wide miRNA binding in murine macrophages. *Cell Rep.* **9,** 2330–2343 (2014).
15. Eigsti, R. L., Sudan, B., Wilson, M. E. & Graff, J. W. Regulation of activation-associated microRNA accumulation rates during monocyte-to-macrophage differentiation. *J. Biol. Chem.* **289,** 28433–28447 (2014).
16. Lin, H. S. *et al.* miR-199a-5p inhibits monocyte/macrophage differentiation by targeting the activin A type 1B receptor gene and finally reducing C/EBPalpha expression. *J. Leukoc. Biol.* **96,** 1023–1035 (2014).
17. Kafasla, P., Karakasiliotis, I. & Kontoyiannis, D. L. Decoding the functions of post-transcriptional regulators in the determination of inflammatory states: focus on macrophage activation. *Wiley. Interdiscip. Rev. Syst. Biol. Med.* **4,** 509–523 (2012).
18. Chenard, C. A. & Richard, S. New implications for the QUAKING RNA binding protein in human disease. *J. Neurosci. Res.* **86,** 233–242 (2008).
19. Fu, X. D. & Ares, Jr. M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15,** 689–701 (2014).

20. Kafasla, P., Skliris, A. & Kontoyiannis, D. L. Post-transcriptional coordination of immunological responses by RNA-binding proteins. *Nat. Immunol.* **15,** 492–502 (2014).
21. Mukherjee, N., Lager, P. J., Friedersdorf, M. B., Thompson, M. A. & Keene, J. D. Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol. Syst. Biol.* **5,** 288 (2009).
22. Turner, M. & Hodson, D. J. An emerging role of RNA-binding proteins as multifunctional regulators of lymphocyte development and function. *Adv. Immunol.* **115,** 161–185 (2012).
23. van der Veer, E. P. *et al.* Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circ. Res.* **113,** 1065–1075 (2013).
24. Hardy, R. J. *et al.* Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice. *J. Neurosci.* **16,** 7941–7949 (1996).
25. Kondo, T. *et al.* Genomic organization and expression analysis of the mouse qkI locus. *Mamm. Genome* **10,** 662–669 (1999).
26. Pilotte, J., Larocque, D. & Richard, S. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes Dev.* **15,** 845–858 (2001).
27. Zorn, A. M. *et al.* Remarkable sequence conservation of transcripts encoding amphibian and mammalian homologues of quaking, a KH domain RNA-binding protein. *Gene* **188,** 199–206 (1997).
28. Chen, T., Damaj, B. B., Herrera, C., Lasko, P. & Richard, S. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell Biol.* **17,** 5707–5718 (1997).
29. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12,** R79 (2011).
30. Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* **12,** 691–698 (2005).
31. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141,** 129–141 (2010).
32. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499,** 172–177 (2013).
33. Teplova, M. *et al.* Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites. *Genes Dev.* **27,** 928–940 (2013).
34. Jiang, L., Saetre, P., Jazin, E. & Carlstrom, E. L. Haloperidol changes mRNA expression of a QKI splice variant in human astrocytoma cells. *BMC Pharmacol.* **9,** 6 (2009).
35. Saetre, P. *et al.* Inflammation-related genes up-regulated in schizophrenia brains. *BMC Psychiatry* **7,** 46 (2007).
36. Brosseau, J. P. *et al.* Tumor microenvironment-associated modifications of alternative splicing. *RNA* **20,** 189–201 (2014).
37. Sidman, R. L., Dickie, M. M. & Appel, S. H. Mutant mice (Quaking and Jimpy) with deficient myelination in the central nervous system. *Science* **144,** 309–311 (1964).
38. Backx, L. *et al.* Haploinsufficiency of the gene Quaking (QKI) is associated with the 6q terminal deletion syndrome. *Am. J. Med. Genet. A* **152A,** 319–326 (2010).
39. Hall, M. P. *et al.* Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* **19,** 627–638 (2013).
40. Wu, J. I., Reed, R. B., Grabowski, P. J. & Artzt, K. Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl Acad. Sci. USA* **99,** 4233–4238 (2002).
41. Zong, F. Y. *et al.* The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* **10,** e1004289 (2014).
42. Pastuszak, A. W. *et al.* An SF1 affinity model to identify branch point sequences in human introns. *Nucleic Acids Res.* **39,** 2344–2356 (2011).
43. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456,** 464–469 (2008).
44. Srinivasan, K. *et al.* Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* **37,** 345–359 (2005).
45. An, G. *et al.* P-selectin glycoprotein ligand-1 is highly expressed on Ly-6Chi monocytes and a major determinant for Ly-6Chi monocyte recruitment to sites of atherosclerosis in mice. *Circulation* **117,** 3227–3237 (2008).
46. Pacis, A., Nedelec, Y. & Barreiro, L. B. When genetics meets epigenetics: deciphering the mechanisms controlling inter-individual variation in immune responses to infection. *Curr. Opin. Immunol.* **29,** 119–126 (2014).
47. Pham, T. H. *et al.* Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* **119,** e161–e171 (2012).
48. Pham, T. H. *et al.* Mechanisms of *in vivo* binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res.* **41,** 6391–6402 (2013).
49. Feinberg, M. W. *et al.* The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J.* **26,** 4138–4148 (2007).
50. Shankman, L. S. *et al.* KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat. Med.* **21,** 628–637 (2015).
51. Gertz, J. *et al.* Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol. Cell* **52,** 25–36 (2013).
52. Kuhlman, P. A., Hughes, C. A., Bennett, V. & Fowler, V. M. A new function for adducin. Calcium/calmodulin-regulated capping of the barbed ends of actin filaments. *J. Biol. Chem.* **271,** 7986–7991 (1996).
53. Chan, K. R. *et al.* Ligation of Fc gamma receptor IIB inhibits antibody-dependent enhancement of dengue virus infection. *Proc. Natl Acad. Sci. USA* **108,** 12479–12484 (2011).
54. Jirka, S. & Aartsma-Rus, A. An update on RNA-targeting therapies for neuromuscular disorders. *Curr. Opin. Neurol.* **28,** 515–521 (2015).
55. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15,** 541–555 (2014).
56. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12,** 861–874 (2011).
57. Lim, J. *et al.* A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **125,** 801–814 (2006).
58. Dehghan, A. *et al.* Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE consortium. *PLoS ONE* **11,** e0144997 (2016).
59. van Eck, M. *et al.* Bone marrow transplantation in apolipoprotein E-deficient mice. Effect of ApoE gene dosage on serum lipid concentrations, (beta)VLDL catabolism, and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **17,** 3117–3126 (1997).
60. Daigneault, M., Preston, J. A., Marriott, H. M., Whyte, M. K. & Dockrell, D. H. The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages. *PLoS ONE* **5,** e8668 (2010).
61. Bot, I. *et al.* Serine protease inhibitor Serp-1 strongly impairs atherosclerotic lesion formation and induces a stable plaque phenotype in ApoE-/-mice. *Circ. Res.* **93,** 464–471 (2003).
62. Sugnet, C. W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS. Comput. Biol.* **2,** e4 (2006).
63. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12,** R72 (2011).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
65. Hsu, F. *et al.* The UCSC known genes. *Bioinformatics* **22,** 1036–1046 (2006).
66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
67. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
68. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
69. Wu, J. *et al.* SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics* **27,** 3010–3016 (2011).
70. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5,** 4846 (2014).
71. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7,** 1009–1015 (2010).

Acknowledgements

For the studies described in this manuscript, we greatly appreciate the participation and cooperation of the QKI haploinsufficient patient and her family members. R.G.d.B. and E.P.v.d.V. were supported by Netherlands Institute for Regenerative Medicine research grants (Grant No. FES0908). This research was also made possible by a private research endowment to E.P.v.d.V. L.S., S.K., J.P.D. and M.A.J. were supported by US National Institutes of Health grant GM040478 and W.S.F. was supported by training grants CIRM TG2-01157 and NIH T32 GM008646. H.K. was supported by European Union Marie Curie CIG grant (PTRCODE) and TUBITAK grant (113E159). We also gratefully acknowledge Dr Christopher Burge and his laboratory at the Massachusetts Institute of Technology, for assistance with the generation of sashimi plots from our RNA-seq data. In addition, we acknowledge Dr Arnoud Sonneberg at the Netherlands Cancer Institute for the in-depth discussions and integrin antibodies.

Author contributions

R.G.d.B., M.A.J., A.J.v.Z. and E.P.v.d.V. conceived the ideas and experiments within the study. R.G.d.B. and E.P.v.d.V. performed the majority of the experiments detailed in the study. L.S., S.K. and J.P.D. performed the RNA-Seq and splicing-sensitive microarray studies, and analysed the data. J.P., H.C.d.B., A.S., J.M.v.G., J.M.G.J.D., A.O.K., P.H.J.v.d.Z., R.S., C.M.A.v.A, and I.B. assisted with experiments and analyses. R.G.d.B, L.S., S.B., W.Y.L., S.M.K., J.P.D. and H.K. provided bioinformatics and statistical assistance. H.V.E. provided access to the QKI haploinsufficient patient and sibling. R.G.d.B., A.J.v.Z. and E.P.v.d.V. wrote and edited the manuscript, and prepared the figures. J.M.v.G., W.S.F., C.v.K., J.W.J., H.V.E., T.J.R., E.A.L.B., M.A.J. and A.J.v.Z. provided critical experimental commentary, discussion and assistance with manuscript preparation and editing. T.J.R., E.A.L.B., M.A.J., A.J.v.Z. and E.P.v.d.V. provided funding for the studies.

Additional information

Accession code: RNA sequencing and microarray data were deposited in NCBI's Gene Expression Omnibus (GEO) under accession codes GSE74979 (for RNA-seq data) and GSE74887 (for the splicing-sensitive microarray data).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

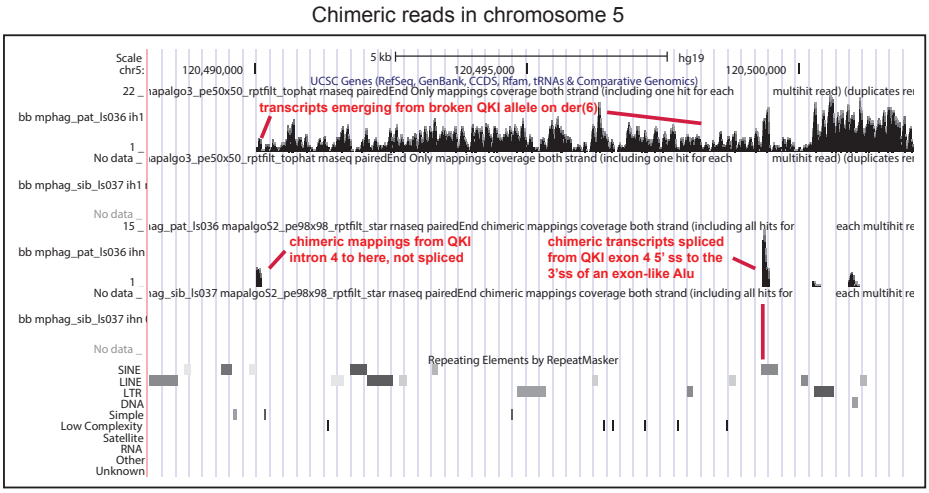
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: de Bruin, R. G. *et al.* Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression. *Nat. Commun.* 7:10846 doi: 10.1038/ncomms10846 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

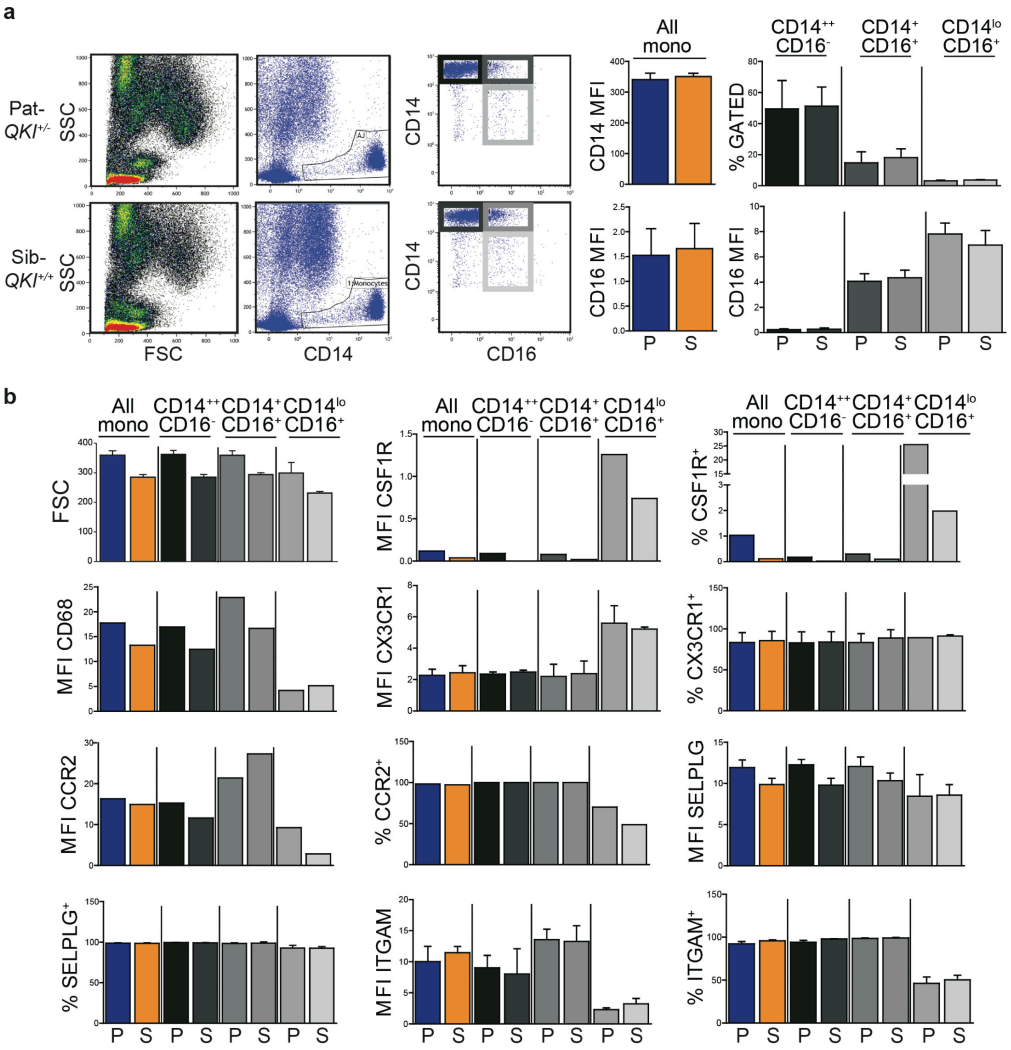
Supplementary Information

Supplementary Figure 1



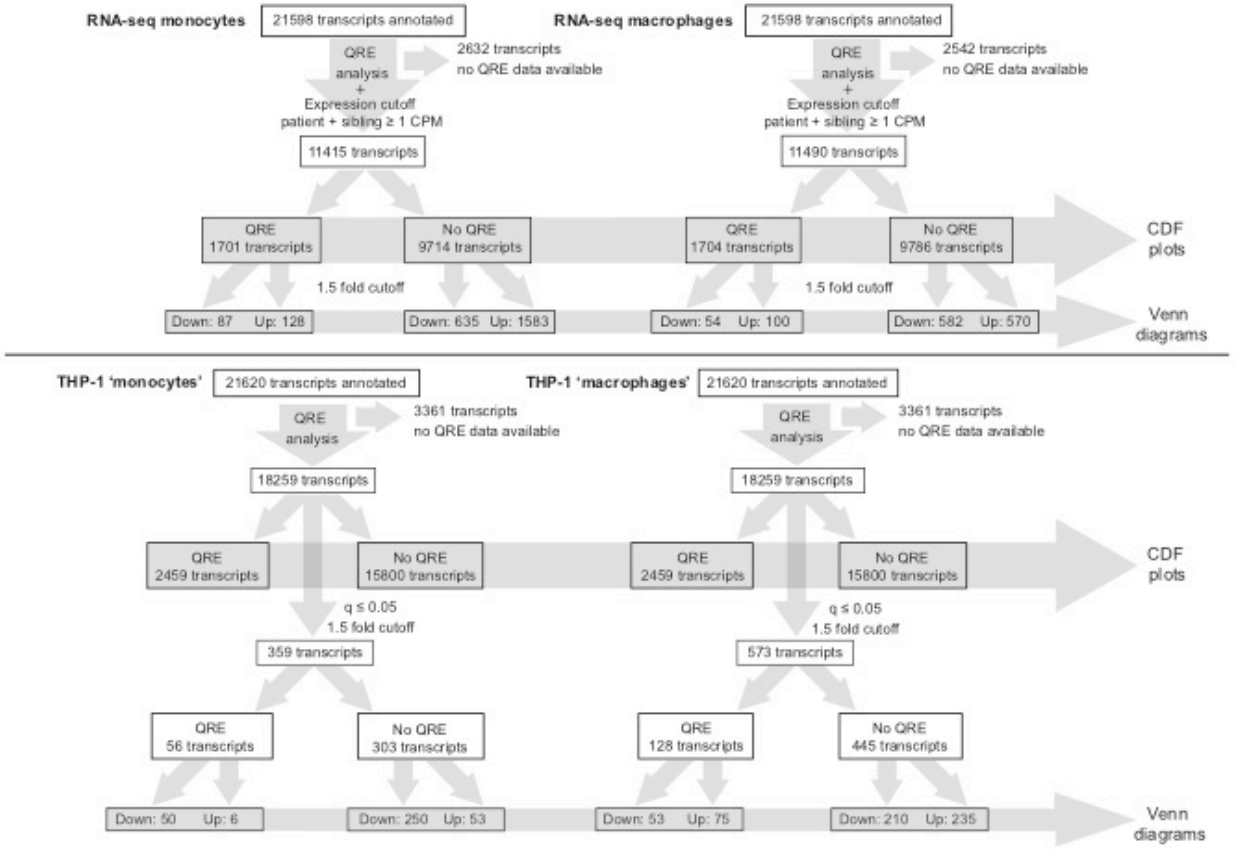
Supplementary Figure 1: RNA-seq shows chr6-derived read coverage on chr5. Genomic location according to UCSC genome browser on chromosome 5. Read coverage is only seen in patient tracks.

Supplementary Figure 2



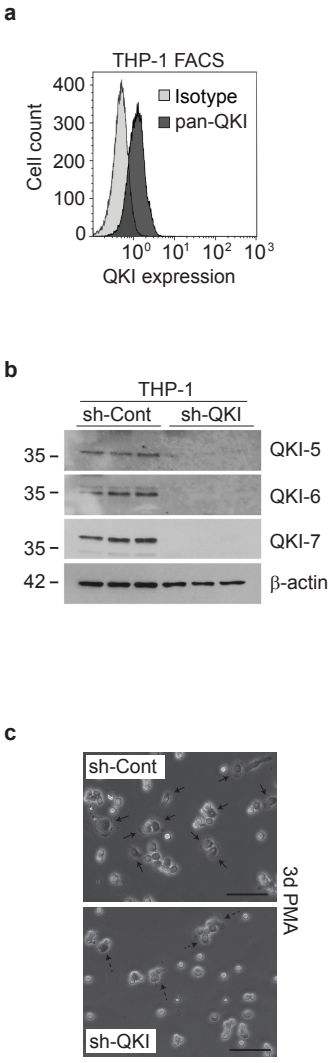
Supplementary Figure 2: FACS analysis of surface marker expression levels in monocyte subsets in the QKI haploinsufficient patient and her sibling. a. FACS analysis of PB harvested from the QKI haploinsufficient patient (P denotes Pat-QKI$^{+/-}$) as compared with an age- and sex-matched sibling control (S denotes Sib-QKI$^{+/+}$). Live cells were first gated in the FSC/SSC gate, after which monocytes were selected based in CD14/SSC expression. Monocyte subpopulations were defined as CD14^{++}/CD16^{-}, CD14^{++}/CD16^{+}, and CD14^{+}/CD16^{+}. MFI denotes mean fluorescent index. (n=1 biological replicate; n=7 technical replicates). **b.** Quantitation of average monocyte size in Pat-QKI$^{+/-}$ and Sib-QKI$^{+/+}$ monocytes based on FSC parameter. All other panels represent quantification of either MFI or percentage of gated monocytes (defined as % gene name^{+}) that express the designated surface markers at levels higher than background signal, such as CSF1R. Data are based on biological n=1, while error bars indicate technical replicates for surface markers included in more than one sample in the phenotype characterization panel (minimal n=2).

Supplementary Figure 3



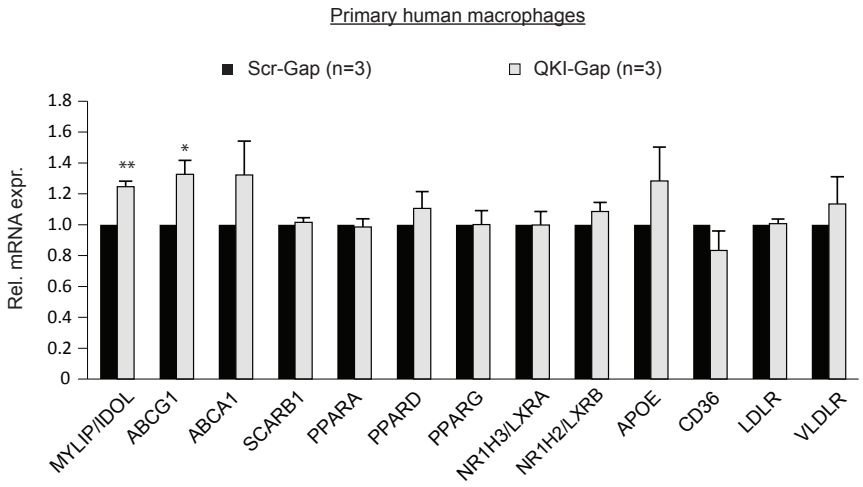
Supplementary Figure 3: Schematic representation of transcript stratification. Flow-diagram illustrating the bioinformatic approach utilized for analysis of RNA-seq data (top) and microarray-based data (bottom). Vertical arrows depict stratification based on the presence or absence of QRE's in transcripts and those achieving expression and/or significance cut-offs. Horizontal arrows indicate transcripts used to generate Venn diagrams (Fig. 3g,5e), tables of most up and downregulated genes (Fig. 3h, 5f), scatterplots (Fig. 3i, 5g) and Cumulative Distribution Fraction plots (Fig. 3j, 5h).

Supplementary Figure 4



Supplementary Figure 4: QKI protein is differentially expressed during THP-1 monocyte-like to macrophage-like differentiation. a. Intracellular FACS analysis of THP-1 'monocytes' for total QKI expression. **b.** Western blot analysis of QKI-5, -6 and -7 expression in cellular lysates harvested from sh-Cont and sh-QKI transduced THP-1 'macrophages' following stimulation with PMA for 3d. **c.** Phase-contrast photomicrographs of 3 days stimulated THP-1 'macrophages'. Scale bar = 50 μm.

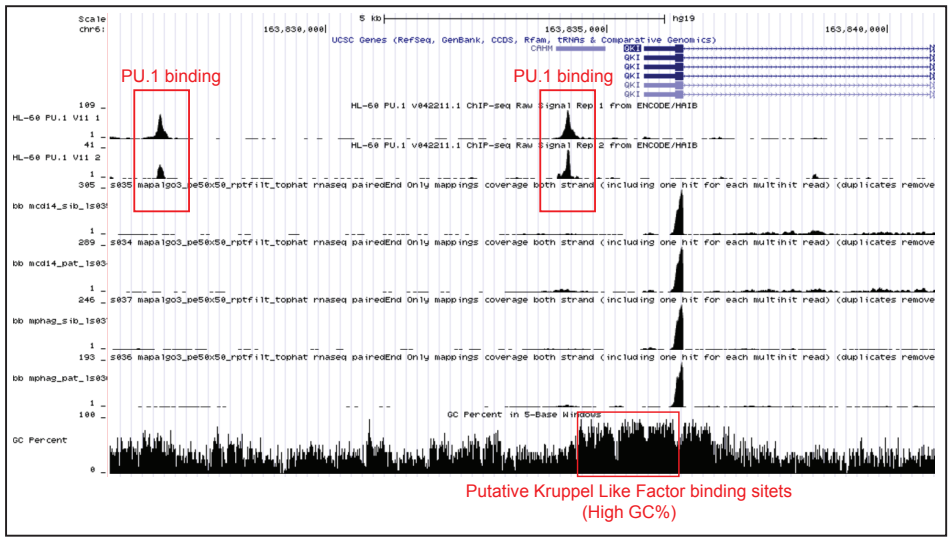
Supplementary Figure 5



Supplementary Figure 5: GapmeR-mediated reduction of QKI expression regulates the expression of atherosclerosis-related mRNAs. qRT-PCR analysis of established atherosclerosis-related genes in QKI-Gap as compared to Scr-Gap treated macrophages. Gene expression in QKI-Gap macrophages are relative to Scr-Gap macrophages (n=3). Data expressed as mean +/- s.e.m; Student's t-test; *p<0.05, ** p<0.01.

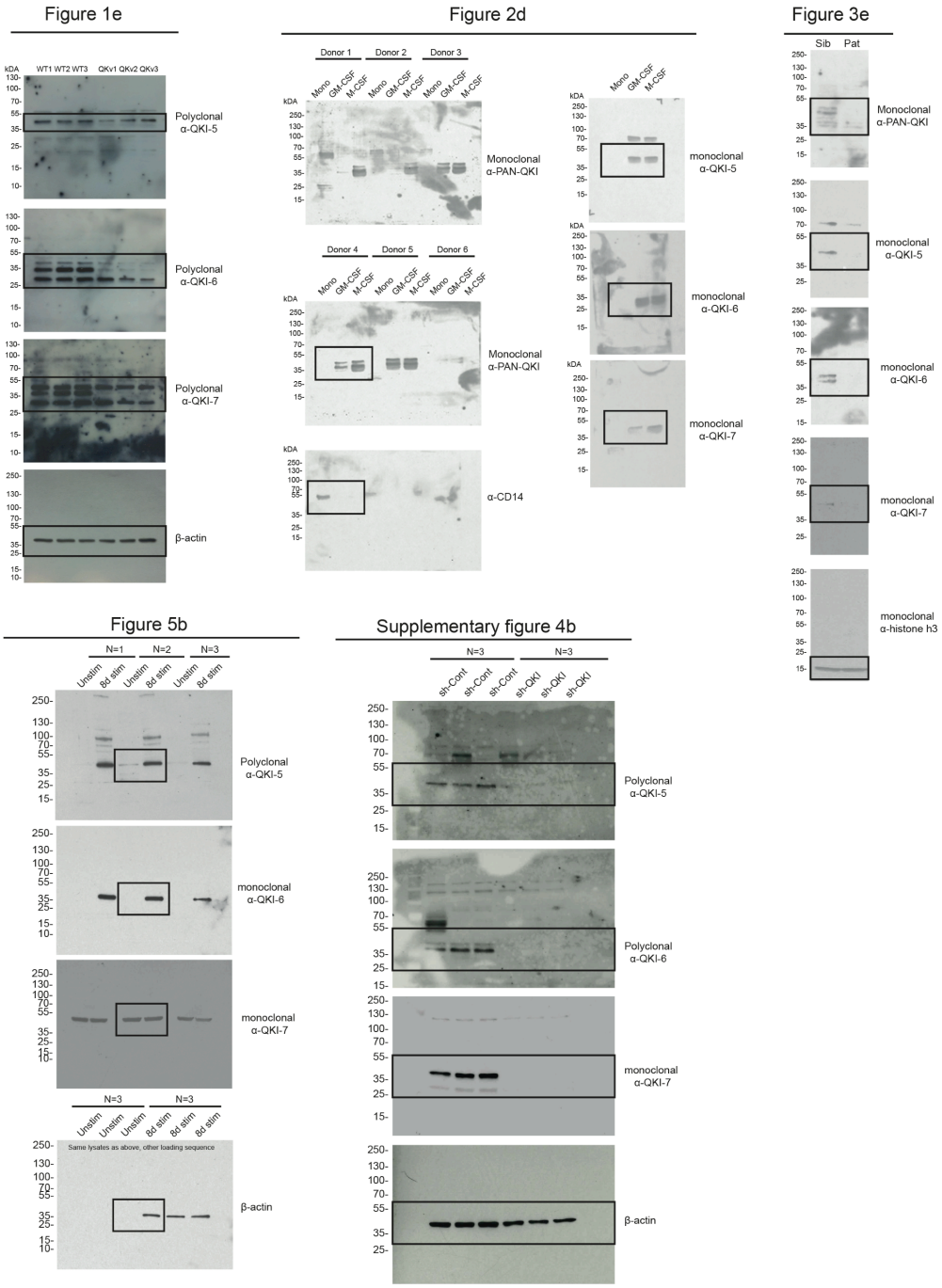
Supplementary Figure 6

Transcription factors that bind to the QKI promoter region



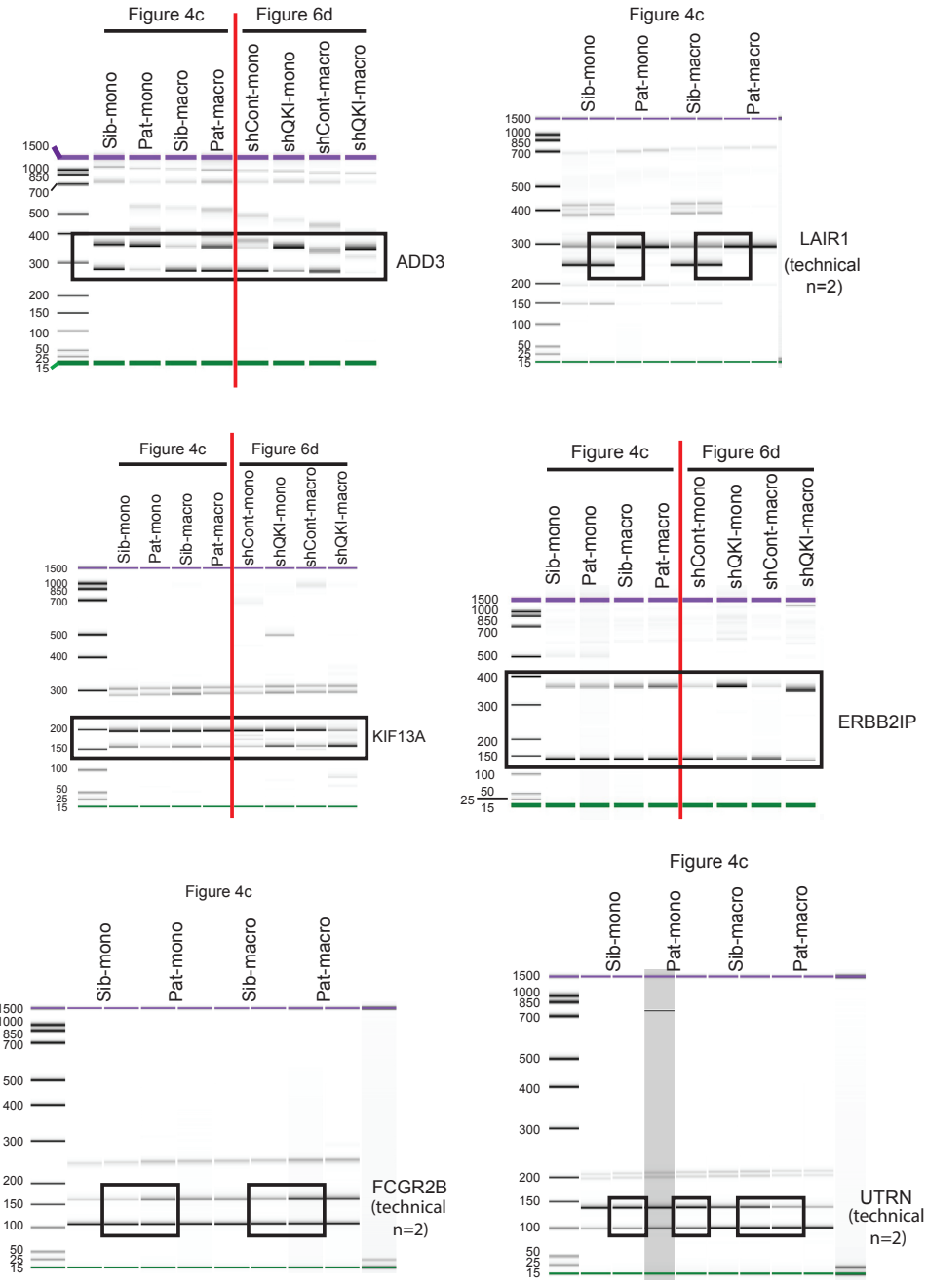
Supplementary Figure 6: Experimentally determined and putative transcription factor binding sites in the QKI promoter region. Top tracks: Genomic location and organization according to UCSC genome browser on chromosome 6 mapping to the QKI locus. Tracks 2 and 3 represent the read coverage for two independent chromatin-immunoprecipitations in HL60 cells for PU.1 [1], a well-known myeloid transcription factor. Tracks 4,5,6,7 represent read coverage for sibling and patient monocyte and macrophages. Track 8 represents percentage GC in 5 bp windows, indicative of putative Kruppel-Like Factor binding sites.

Supplementary Figure 7



Supplementary Figure 7: Original immunoblot images utilized for preparation of figures 1e, 2d, 3e, 5b and Supplementary Fig. 4b.

Supplementary Figure 8



Supplementary Figure 8: Original bioanalyzer images utilized for preparation of figures 4c and 6d.

Supplementary Figure 9

Figure 6d

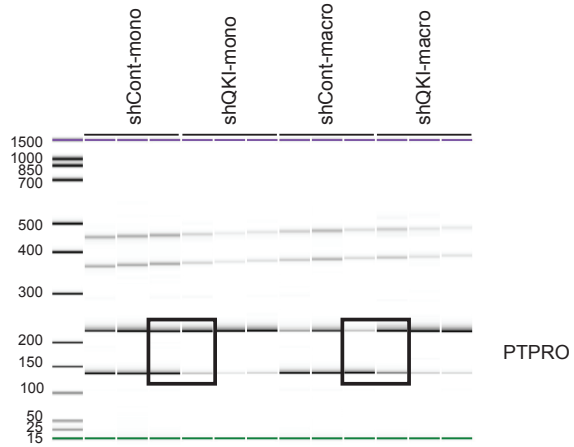


Figure 6e

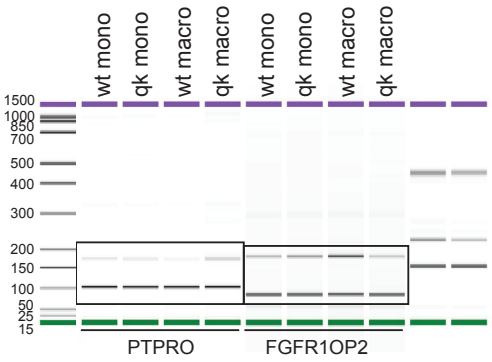
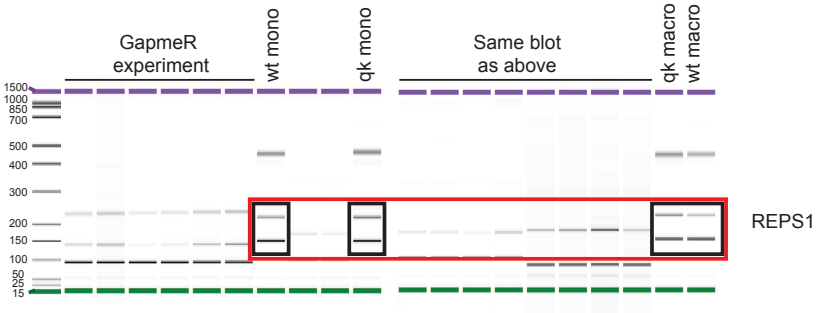


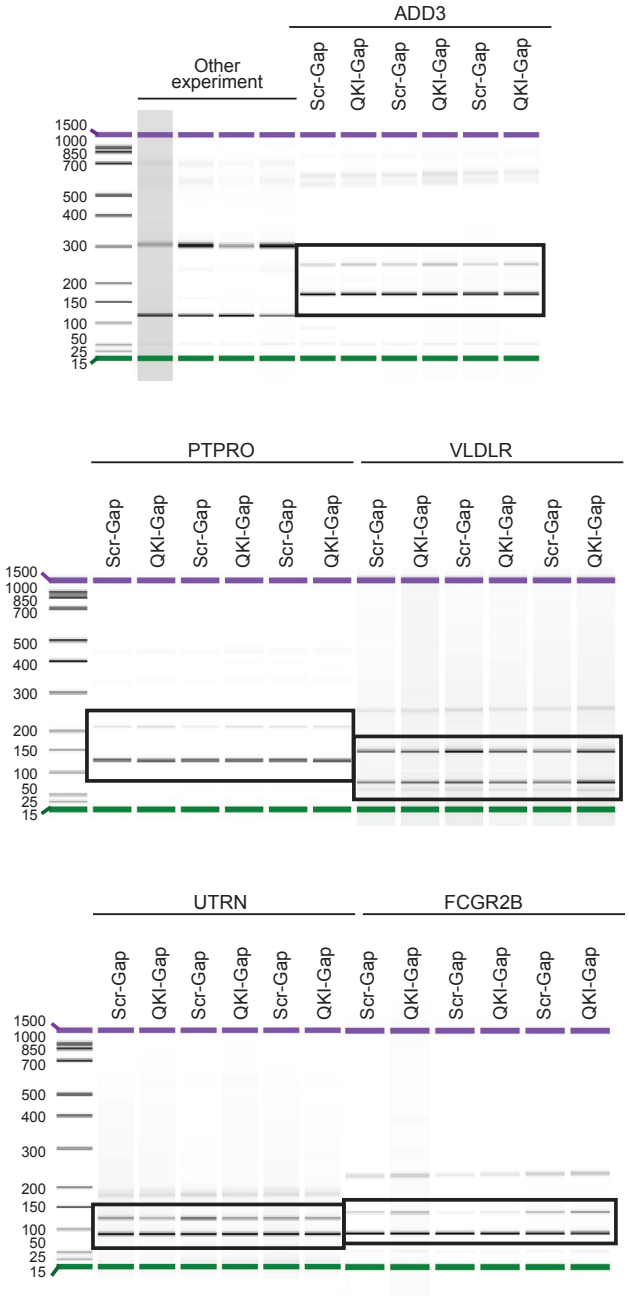
Figure 6e



Note: A pipeting miscue in left gel prompted us to load the two remaining samples in another blot.

Supplementary Figure 9: Original bioanalyzer images utilized for preparation of figures 6d and 6e.

Supplementary Figure 10



Supplementary Figure 10: Original bioanalyzer images utilized for preparation of figures 4f.

Supplementary Table 1

Transcript abundance primers		5'--> 3'
hsa+mmu-QKI-5	Forward	CTGTCATGCCAAACGGAAC
	Reverse	GATGGACACGCATATCGTG
hsa+mmu-QKI-6	Forward	CTGTCATGCCAAACGGAAC
	Reverse	CGTTGGGAAAGCCATAC
hsa+mmu-QKI-7	Forward	CTGTCATGCCAAACGGAAC
	Reverse	GACTGGCATTTCAATCCAC
hsa-GAPDH	Forward	TTCCAGGAGCGAGATCCCT
	Reverse	CACCCATGACGAACATGGG
pre-mRNA splicing primers		5'--> 3'
hsa-ADD3	Forward	ACCAGCTCCTCCTAACCCAT
	Reverse	TCACTCGCTTAGCAAGCTCAT
hsa-ERBB2IP	Forward	AGTTCCTCGTGACTGGAGAGA
	Reverse	AATGGGTTTCCTCTACCCCC
hsa-KIF13A	Forward	GTGCAGCATTCAGGGACACT
	Reverse	GCATCTGACCACCTCTCCCT
hsa-LAIR1	Forward	AAACATTCCGCCTGGAGAGG
	Reverse	CATTGTGACTGTTGTCCGACG
hsa-PTPRO	Forward	GGAGTGTGGAGCTGGTACAT
	Reverse	AGGCATCAAAGTCATCCAGTTG
hsa-FCGR2B	Forward	TCACTGGGATTGCTGTAGCG
	Reverse	GCCTCATCAGGATTAGTGGGA
hsa-UTRN	Forward	TTGCCAAACACCCTCGACTT
	Reverse	AACAGTTGAGGAGATTGTGAGGG
mmu-PTPRO	Forward	ATGTGGAGCTGGCACGTTTG
	Reverse	ACGGGGTTTGTTAGTTTCCTCT
mmu-FGFR1OP2	Forward	CATGGCCAGCAAGAAAGATGAC
	Reverse	TTTGGTCAACATGTGCTTGC
mmu-REPS1	Forward	AGCCAGGTGAGGTAGGTTACT
	Reverse	CTGCATGTGGATTTTGCTTGGA

Atherosclerosis-related genes		5'--> 3'
hsa-MYLIP/IDOL	Forward	CATCTTACAGGAGCAGACTAGGC
	Reverse	TTGGCAGTGTTCTGGTTGTAG
hsa-ABCG1	Forward	ATTCAGGGACCTTTCCTATTCGG
	Reverse	CTCACCACTATTGAACTTCCCG
hsa-ABCA1	Forward	ACCCACCCTATGAACAACATGA
	Reverse	GAGTCGGGTAACGGAAACAGG
hsa-SCARB1	Forward	ACTTCTGGCATTCCGATCAGT
	Reverse	ACGAAGCGATAGGTGGGGAT
hsa-PPARD	Forward	GCCTCTATCGTCAACAAGGAC
	Reverse	GCAATGAATAGGGCCAGGTC
hsa-PPARG	Forward	TACTGTCGGTTTCAGAAATGCC
	Reverse	GTCAGCGGACTCTGGATTCAG
hsa-PPARA	Forward	ATGGTGGACACGGAAAGCC
	Reverse	CGATGGATTGCGAAATCTCTTGG
hsa-NR1H2	Forward	AGAAGATTCGGAAACAACAGCA
	Reverse	GCTGGATCATTAGTTCTTGAGCC
hsa-ApoE	Forward	GTTGCTGGTCACATTCCTGG
	Reverse	GCAGGTAATCCCAAAAGCGAC
hsa-CD36	Forward	CTTTGGCTTAATGAGACTGGGAC
	Reverse	GCAACAAACATCACCACACCA
hsa-LDLR	Forward	ACGGCGTCTCTTCCTATGACA
	Reverse	CCCTTGGTATCCGCAACAGA
hsa-VLDLR	Forward	CTGGGTATGCGACGATGATG
	Reverse	CTTGGTGTGTATGACTGGCTG
hsa-NR1H3	Forward	CTCAATGCCTGATGTTTCTCCT
	Reverse	TCCAACCCTATCCCTAAAGCAA
mmu-PPARD	Forward	TCCATCGTCAACAAAGACGGG
	Reverse	ACTTGGGCTCAATGATGTCAC
mmu-PPARG	Forward	GGAAGACCACTCGCATTCCTT
	Reverse	GTAATCAGCAACCATTGGGTCA
mmu-PPARA	Forward	AGAGCCCCATCTGTCCTCTC
	Reverse	ACTGGTAGTCTGCAAAACCAAA
mmu-NR1H2	Forward	ATGTCTTCCCCCACAAGTTCT
	Reverse	GACCACGATGTAGGCAGAGC
mmu-NR1H3	Forward	CTCAATGCCTGATGTTTCTCCT
	Reverse	TCCAACCCTATCCCTAAAGCAA
mmu-ApoE	Forward	CTGACAGGATGCCTAGCCG
	Reverse	CGCAGGTAATCCCAGAAGC
mmu-CD36	Forward	AGATGACGTGGCAAAGAACAG
	Reverse	CCTTGGCTAGATAACGAACTCTG
mmu-LDLR	Forward	TCAGACGAACAAGGCTGTCC
	Reverse	CCATCTAGGCAATCTCGGTCTC
mmu-VLDLR	Forward	GAGTCTGACTTCGTGTGCAAA
	Reverse	GAACCGTCTTCGCAATCAGGA
mmu-NR1H3	Forward	CTCAATGCCTGATGTTTCTCCT
	Reverse	TCCAACCCTATCCCTAAAGCAA
mmu-MYLIP/IDOL	Forward	ATGCTGTGCTATGTGACGAGG
	Reverse	TCGATGATCCCTAGACGCCTG

Supplementary Table 1: Primers sets utilized to determine mRNA abundance and pre-mRNA splicing events.

Supplementary Reference

1. Gertz et.al. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol. Cell* 52, 25-36 (2013).

CHAPTER 5

AUTOGENOUS REGULATION BALANCES QUAKING FUNCTIONS IN SPLICING AND TRANSLATION

Autogenous regulation balances Quaking functions in splicing and translation

W. Samuel Fagg and Manuel Ares Jr.[1]

Center for Molecular Biology of RNA

Department of Molecular, Cell & Developmental Biology

Sinsheimer Laboratories

UC Santa Cruz. Santa Cruz, CA 95064

[1]Corresponding author. Email ares@ucsc.edu Phone (831) 459-4628)

Running Head: Autoregulatory control of Qk isoforms

Keywords: Quaking, Qk, QKI, RNA binding protein, autoregulation, RNA processing

Abstract

Mammalian RNA binding proteins (RBPs) occur in families whose members bind nearly the same RNA sequence but have different functions in splicing, stability, decay, and localization of mRNA, often within the same cell. Unexplored regulatory control must exist to ensure that the right level of each RBP family protein is expressed. Here we analyze the function and regulation of two protein isoforms expressed from the single *quaking* gene (*Qk*) in mouse muscle cells: Quaking-5 (Qk5) and Quaking-6 (Qk6). We demonstrate the division of labor between Qk5 (in splicing) and Qk6 (translation), and show that these roles are enforced by differential subcellular localization. Alternative RNA processing produces mRNAs encoding a Qk core protein with either a nuclear localization signal (Qk5) or a cytoplasmic retention/nuclear export signal (Qk6), and we show using isoform-specific depletion that a complex auto- and cross-regulatory network exists by which Qk5 and Qk6 control their own and each other's levels. This regulation is manifested through Qk5's roles in transcript processing and Qk6's role in translation, and is itself subject to additional cell type and developmental control to produce characteristic Qk isoform distributions in different cell types.

Introduction

Sequence-specific RNA-binding proteins (RBPs) regulate nearly every step of post-transcriptional gene expression. Control is exerted by recognition of cis-acting regulatory sequences by the RNA-binding domain (RBD), followed by one or more of a number of consequences, including restructuring of the RNA and interaction with other proteins that influence the activity of the spliceosome, export and decay machinery, and the ribosome (Keene 2007; Fu and Ares 2014). Furthermore, many RBPs control their own levels, through the same mechanisms by which they regulate their targets (Wollerton et al. 2004; Damianov and Black 2010; Sun et al. 2010). In many metazoans, multiple proteins are expressed that have the same or highly similar RBDs. These two circumstances, autogenous regulation and functional diversification using the same cis-acting sequences and RBDs, together create a regulatory conundrum: How is functional specificity maintained and how does the cell ensure that the right amount of each functional form of the RBP is made?

An example of this phenomenon occurs in the mammalian Muscleblind-like (Mbnl) family, in which the Mbnl1 and Mbnl2 proteins encoded by different genes possess nearly identical RBDs (Teplova and Patel 2008; He et al. 2009), with parallel abilities to recognize a YGCY sequence motif (Mankodi et al. 2001; Fardaei et al. 2002), while displaying interesting differences in function in different tissues. In skeletal muscle, Mbnl1 is primarily a splicing factor while Mbnl2 is not (Du et al. 2010; Wang et al. 2012), and may instead regulate mRNA localization or stability

(Adereth et al. 2005; Du et al. 2010). In brain, Mbnl2 serves as a splicing factor, whereas Mbnl1 does not (Charizanis et al. 2012; Suenaga et al. 2012). Thus, these two different family members demonstrate distinct tissue-specific divisions of labor. How different RBP family members that recognize the same RNA sequences to effect and regulate splicing, localization, translation and decay remains an open question.

Similar observations for other RBP families (Makeyev et al. 2007; Spellman et al. 2007; Yeo et al. 2009; Gehman et al. 2011; Gehman et al. 2012; Singh et al. 2014) suggest two mechanisms for how isoforms or paralogous RBP family members may be assigned distinct functions: 1) isoform/family member-specific protein features could promote localization to restrict function to processes within a particular compartment (Caceres et al. 1998; Koizumi et al. 1999; Cazalla et al. 2002; Sanford et al. 2004), or 2) isoform/family member-specific protein features might promote interactions with specific elements of the core machinery (Mattaj and De Robertis 1985; Moreland et al. 1985; Query et al. 1989). Because many RBPs are encoded by multi-gene families with complex alternative splicing, whole menageries of RBP isoforms and paralogs can be expressed, even within single cell types. The *Quaking* gene (*QKI* in human, *Qk* in mouse) expresses a limited number of protein isoforms from a single gene, and therefore represents an excellent candidate to address how different RBP isoforms might regulate different functions within a single cell type. Transcription initiates primarily at a single major site, and in most cell types where it is expressed, the gene produces 3 alternatively spliced mRNAs for 3 isoforms [Quaking-5 (Qk5), Quaking-6 (Qk6), and Quaking-7 (Qk7)] that share the same RBD

and differ only in the very C-terminal amino acids (Ebersole et al. 1996; Kondo et al. 1999).

The *Quaking* family is a member of a larger RBP family called Signal Transduction and Activator of RNA metabolism (STAR) family. STAR family proteins respond to cell signaling cues and regulate post-transcriptional gene expression in response to these signals (Vernet and Artzt 1997; Richard 2010). Collectively, *Qk* regulates RNA processing at several steps including pre-mRNA splicing (Hall et al. 2013), mRNA localization (Li et al. 2000; Larocque et al. 2002), mRNA stability/decay (Li et al. 2000; Larocque et al. 2005; Zearfoss et al. 2011), translation (Saccomanno L 1999; Zhao et al. 2010), and miRNA processing (Wang et al. 2013; Zong et al. 2014). Dimeric Qk protein binds an RNA element that includes minimally a "full-site" and a "half-site" of the motif ACUAAY separated by at least one nucleotide (Galarneau and Richard 2005). The location of binding sites within an RNA molecule suggests what step(s) in post-transcriptional regulation a given Qk binding event may regulate: an intronic motif could control splicing (Hall et al. 2013; van der Veer et al. 2013), whereas a 3'UTR motif might control mRNA localization (Li et al. 2000; Larocque et al. 2002; Ryder and Williamson 2004), translation (Saccomanno L 1999; Zhao et al. 2010; Zearfoss et al. 2011), and/or decay (Li et al. 2000; Larocque et al. 2005; Zhao et al. 2006a), for example. However, whether these diverse functions of *Qk* are individually executed by distinct isoforms, or if more than one isoform together is required to carry out a specific function is unknown.

Because the Qk isoforms differ only at their extreme C-termini, clues to their isoform-specific functions likely exist there. The Qk5 isoform has a non-canonical nuclear localization signal (NLS) within its C-terminus and shuttles between the nucleus and cytoplasm and back (Wu et al. 1999). Qk6 has the shortest C-terminus of all three isoforms, is primarily cytoplasmic (Lu et al. 2003), and may regulate mRNA stability (Li et al. 2000; Li et al. 2003b), localization (Li et al. 2000; Larocque et al. 2002), and translation (Saccomanno L 1999; Zhao et al. 2010; Mandler et al. 2014). Qk7 protects target mRNAs from decay in glia (Larocque et al. 2005; Larocque et al. 2009) and is typically cytoplasmic (Lu et al. 2003), although heterodimerization of this isoform with other Qk isoforms can alter its subcellular localization (Pilotte et al. 2001) and perhaps its function. It is unclear if all of these functional observations accurately represent the physiological roles of Qk isoforms, however. Many of these observations come from cell-based overexpression studies and in the *quaking*^{viable} mouse (Sidman et al. 1964), a dysmyelination model in which the expression and isoform ratio of Qk protein isoforms is disrupted in oligodendrocytes (Lu et al. 2003). Furthermore, because Qk isoforms may have distinct, overlapping, opposing, and/or interdependent functions in different contexts, it seems likely that the correct balance of isoform expression is of critical importance for regulation of post-transcriptional gene expression. How this balance of expression is established and maintained in different cell types is unknown.

Here we describe functional analyses of Qk protein isoforms in mouse C2C12 cells. Qk5 and Qk6 are the predominantly expressed isoforms in this cell type. We

find that Qk5, but not Qk6, is a splicing factor. Further, the relative expression of Qk protein isoforms is regulated, at least in part, by Qk protein isoforms themselves. Cell type specific balance of Qk protein isoforms is both auto- and cross-regulated by each isoform, but consistent with a functional division of labor, each isoform imposes this regulation at a different step in the gene expression program. These findings uncover an unexpectedly complex network of cross-isoform and autoregulation within a single family of RBPs, and suggest that homeostatic post-transcriptional regulation by *Qk* in many cell types is tightly regulated by Qk protein levels themselves.

Results

Qk5 and Qk6 are the predominant isoforms in C2C12 cells

We previously showed that Qk controls splicing of hundreds of exons in C2C12 cells (Hall et al. 2013). In other cell types such as oligodendrocytes, Qk controls mRNA localization, abundance, and translation through binding to the 3'UTR of its target mRNAs (Li et al. 2000; Ryder and Williamson 2004; Larocque et al. 2005; Zhao et al. 2006a; Zhao et al. 2010). A single Qk pre-mRNA is transcribed from the *Qk* gene, which is alternatively spliced to make a 5 kb transcript that encodes Qk5, a 6 kb transcript that encodes Qk6, and a 7 kb transcript that encodes Qk7 (Fig S1A). Before dissecting isoform-specific functions in C2C12 cells, though, we determined the Qk isoform-specific expression profiles. Consistent with our previous findings (Hall et al. 2013), we observed that total Qk protein level increases during C2C12 cell differentiation (Fig S1B). Furthermore we found that Qk5 was the

most abundant protein, followed by Qk6, and Qk7 was nearly undetectable (Fig S1B). Additionally, there was no apparent isoform ratio change during differentiation (FigS1B). We verified C2C12 differentiation and Qk protein localization by immunostaining, and found that total Qk is abundantly detected in both the nucleus and cytosol of myoblasts and myotubes (Fig S1C). Using isoform-specific immunostaining, we observed Qk5 is primarily nuclear, although some cytoplasmic localization was observed (Fig S1D). The localization of Qk6 in C2C12 myoblasts varied cell-to-cell: Qk6 staining was robust in some nuclei, while nearly undetectable in others (Fig S1D). Therefore, Qk5 and Qk6 are the abundantly expressed Qk isoforms in C2C12 cells, and there is no isoform ratio change during myoblast to myotube differentiation.

Qk-dependent splicing requires Qk5, but not Qk6

We previously showed Qk protein directly promotes inclusion of Capzb exon 9 by binding to intronic ACUAA sequence elements downstream of this exon (Hall et al. 2013), but it was unclear which Qk isoform(s) bound this sequence. To begin dissecting the roles of the two isoforms, Qk5 and Qk6, expressed in C2C12 myoblasts, we overexpressed one or the other Myc-tagged Qk isoform (Fig 1A), and tested the effect on splicing using the Qk regulated Capzb exon 9 reporter (Hall et al 2013, Fig 1B). Both Myc-Qk5 and Myc-Qk6 activate splicing (Fig 1B), consistent with nuclear localization observed of both Qk5 and Qk6 in untransfected C2C12 myoblasts (Fig S1D).

Based on previous results, overexpression might lead to increased formation of Qk5-Qk6 heterodimers in transfected cells (Pilotte et al. 2001), and increased splicing upon the overexpression of Myc-Qk6 may be an artifact of overexpression. So to ask if either Qk isoform alone is sufficient to function in splicing, we devised a deplete/add-back approach in which siRNA to a common region of Qk mRNA is used to deplete all protein isoforms, and a single isoform is added back by expression of either Myc-Qk5 or Myc-Qk6 containing translationally silent mutations that disrupt siRNA:mRNA base-pairing (a kind gift from Sean Ryder). Although the depletion of endogenous protein is incomplete, the transfected single isoform is the predominant Qk protein expressed in Qk add-back myoblasts (Fig 1C). Inclusion of Capzb exon 9 is reduced after Qk knockdown (Hall et al. 2013), and inclusion is significantly increased by add-back of Myc-Qk5, but not by Myc-Qk6 (Fig 1D). We conclude that the splicing functions of Qk lie with the Qk5 isoform, and that Qk6 is not normally a splicing factor. To address the assumption that nuclear localization of Qk6 is dependent on C2C12 cells also expressing high levels of Qk5, we analyzed the subcellular location of either over-expressed Myc-Qk5 and Myc-Qk6, or in myoblasts depleted of endogenous Qk with Myc-Qk5 or Myc-Qk6 added back (Fig 1E). Whereas Myc-Qk5 is nuclear whether or not endogenous Qk is depleted, Myc-Qk6 is nuclear only when endogenous levels of Qk are present, and is predominantly cytoplasmic in cells that lack other isoforms of Qk. We conclude that Qk5 is predominantly nuclear and serves as a splicing regulator, whereas Qk6 is primarily cytoplasmic and typically does not directly contribute to splicing regulation. These

results also suggest that under some natural conditions yet to be identified, heterodimers of Qk isoforms might play a functional role.

Nuclear localization of Qk is sufficient for splicing function

Qk isoforms differ only in C-terminal amino acids (Fig S1A), offering an opportunity to test the role of isoform specific sequences in specialized function(s). The C-terminal 30 amino acids of Qk5 encode a non-canonical nuclear localization signal (NLS; (Wu et al. 1999)) also found on Sam68 (Ishidate et al. 1997), however additional splicing-specific functions could be contained in this region as well. Thus it is possible that nuclear localization is sufficient for Qk splicing function. Conversely, the Qk5 C-terminus may recruit the splicing machinery to substrate pre-mRNAs to execute splicing function. To address these possibilities, we assessed splicing function of the Qk sequence shared by all isoforms. We made two mutant constructs, both consisting of only the 311 amino acids common to all Qk isoforms (Fig S1A). One of these mutants has an N-terminal HA epitope tag (HA-QkBody), and the other an N-terminal HA tag followed by the canonical SV40 NLS (HA-NLS-QkBody, Fig 2A). We used these mutants to test if the shared Qk body alone, if delivered to nuclei, can perform Qk splicing function. Interestingly, we found that both HA-NLS-QkBody and HA-QkBody were capable of promoting inclusion of Capzb exon 9 either with or without prior depletion of endogenous Qk protein (Fig 2C). Immunolocalization of the proteins using anti-HA antibody analysis showed that HA-NLS-QkBody was concentrated in nuclei, whereas HA-QkBody was present in

both the nucleus and cytoplasm (Fig 2D). Taken together with results in Fig 1, this experiment shows that splicing functions of Qk reside in the body of the protein and not in the isoform-specific tail of Qk5, and that nuclear localization of Qk is sufficient for its splicing function. Furthermore the isoform-specific tail of Qk6 must contain a nuclear export or cytoplasmic retention signal that excludes it from the nucleus, rendering it insufficient on its own for splicing. This is consistent with other analyses of Qk cytoplasmic function in translation and mRNA stability (Saccomanno et al. 1999; Li et al. 2000; Larocque et al. 2005; Zhao et al. 2010; Mandler et al. 2014), and suggests that Qk6 may be sufficient for these functions.

Cross-isoform and autoregulation of Qk expression levels:

Qk5 is required for expression of Qk6 and Qk7

To address downstream consequences for gene expression regulated by individual Qk isoforms in minimally perturbed cells we moved from depletion/add-back to an isoform-specific subtraction approach. We designed siRNAs to specifically target the 3'UTR of either Qk5 mRNA only (siQk5) or Qk6 and Qk7 mRNAs (siQk6&7; see Fig S2A) to uncover the effect of loss of specific Qk isoforms on the expression of genes in C2C12 cells. In testing these siRNAs, we observed loss of Qk5, Qk6, and Qk7 after treatment with siQk5, but no loss of Qk5 after treatment with siQk6&7 (Fig S2B). Strikingly, the depletion of Qk6 protein under the Qk5 knockdown condition is as efficient as targeting Qk6 itself with siRNA (32% relative protein abundance in siQk5 vs. 33% relative protein abundance in siQk6/7; Fig 3A),

suggesting a strong contribution by Qk5 protein to Qk6 expression in C2C12 cells. To determine if the change in Qk6 and Qk7 expression under Qk5 knockdown is at the level of RNA expression, we performed RT-qPCR with a primer set that amplifies both the Qk6 and Qk7 mRNA. Similar to the results observed with protein expression for Qk6 (Fig 3A), we observed a significant and similar magnitude reduction in the relative RNA levels of Qk6 and Qk7 under both the siQk5 and siQk6&7 conditions (Fig 3B). To address the possibility that the original siRNA predicted to only target Qk5 might knockdown Qk6 mRNA via base pairing between Qk6 mRNA and siQk5, we designed an additional 3 siRNAs also predicted to directly base pair with Qk5 but not Qk6 mRNA (Fig S2C). Consistent with results described above, transfection of C2C12 cells with each of these siRNAs reduced both Qk5 and Qk6 protein and mRNA levels (Fig S2D and S2E, respectively) demonstrating that reduction of Qk6 was not due to off-target knockdown by siQk5. These results indicate that Qk5 is required for accumulation of Qk6 mRNA.

Based on previous examples of Qk proteins regulating mRNA expression through 3'UTR binding (Li et al. 2000; Larocque et al. 2002; Larocque et al. 2005) and that many RBPs are known to regulate their own expression and other RBPs expression to maintain homeostatic levels of RNA processing ((Samson 1998; Wollerton et al. 2004; Ni et al. 2007; Damianov and Black 2010; Zearfoss et al. 2011) , we hypothesized that Qk5 may promote Qk6 mRNA abundance by directly binding the Qk6 3'UTR. Evidence in support of direct binding exists in a PAR-CLIP dataset where one sequence read peak from tagged QKI protein immunoprecipitated from

HEK293T cell lysate maps to the Qk6 3'UTR (Fig 3C; (Hafner et al. 2010)). Additionally, RBP Map, a computational tool that predicts RBP sequence motifs, finds 12 putative Qk binding sites in the Qk6 3'UTR mRNA sequence (Fig 3C). Thus Qk protein may directly bind and exert regulation through the Qk6 3'UTR. To address this possibility, we tested if Qk5 is required for accumulation of a GFP reporter gene containing the Qk6 3'UTR. We co-transfected C2C12 myoblasts with this reporter, an RFP transfection control reporter (Fig 3D), and either control (siNT), PanQk (siQk), or Qk5 (siQk5) siRNAs. Western blot analysis shows PanQk and Gapdh protein level abundance under siNT, siQk, and siQk5 treatments, and indicate efficient knockdown of Qk proteins (Fig 3E). If Qk5 promotes Qk6 mRNA-level expression through the Qk6 3'UTR, then C2C12 cells lacking Qk5 expression should result in a lower level of GFP reporter RNA abundance relative to RFP control reporter RNA. The results indeed show relative Qk6 3'UTR reporter mRNA is reduced to ~60% under siQk conditions, and further reduced to ~30% under siQk5 relative to control siRNA (Fig 4F), demonstrating that GFP-Qk6 3'UTR reporter mRNA expression requires Qk5 protein. Interestingly, we observed that depletion of Qk protein is more efficient under siQk5 than siQk treatment (Fig 3E). Reporter mRNA level abundance is also more significantly reduced under the more efficient siQk5 knockdown, thus the Qk6 3'UTR reporter mRNA level shows a dose-dependent response to reduced Qk5 protein levels. Together, these results suggest that Qk5 may directly bind to Qk sequence motifs in the 3'UTR of Qk6 mRNA transcripts to promote expression of this isoform.

Qk6 negatively regulates Qk5 protein expression

The requirement of Qk5 for Qk6 expression prompted us to test the possibility that Qk6 might control Qk5 expression. We performed quantitative western blotting with infrared imaging using the Qk5 specific antibody, and detected a subtle but reproducible increase of about 17% in Qk5 protein when Qk6 was knocked down (Fig 4A). Additionally we observed a significant increase in the level of inclusion of Capzb exon 9 in the Dup51-Capzb splicing reporter (Fig S2F) under this condition, consistent with an increase in Qk5 activity (Hall et al. 2013). We measured the level of Qk5 mRNA by RT-qPCR after Qk6 depletion, and found a modest and statistically insignificant reduction relative to control siRNA, (Fig 4B) however, this slight decrease in Qk5 mRNA, together with the slight increase in Qk5 protein results in a 40% increase in translational efficiency (calculated by dividing relative protein level by relative mRNA level, (Vasudevan and Steitz 2007)) of Qk5 in C2C12 cells. Furthermore, overexpression of Qk6 resulted in a concomitant reduction in Qk5 protein level independent of an mRNA level change (Fig S3A & B). These data show that Qk6 contributes to the translational repression of Qk5, consistent with the known function of Qk6 in other cell types (Saccomanno L 1999; Zhao et al. 2010). These observations indicate a complex regulatory relationship between the different protein products from the same gene, in which the specific functions of each isoform are imposed on the expression of the other in a way that controls both the overall output from the gene, as well as the fraction of the output devoted to each isoform.

Qk6 positively regulates Qk6 protein expression

Since loss of function of different Qk isoforms alters expression of other Qk isoforms, we explored how increasing specific Qk isoform expression might affect Qk gene expression. Loss of Qk5 eliminates Qk6 expression, so we tested whether overexpression of Myc-Qk5 might up-regulate endogenous Qk6. By quantitative western blot analysis, endogenous Qk6 levels remain unchanged when Myc-Qk5 is overexpressed (Fig 4C, middle lane). In contrast overexpression of Myc-Qk6 produces a robust increase of 2.5-fold in endogenous Qk6 protein levels (Fig 4C, right lane). We do not observe a significant change in Qk6 mRNA level upon Myc-Qk6 overexpression. Therefore overexpression of Myc-Qk6 increases the translational efficiency of endogenous Qk6 mRNA by ~2.5 fold (Fig 4D). Thus, Qk6 may positively regulate translational efficiency of its own mRNA even as it reduces translational efficiency of Qk5 mRNA.

An additional possibility for the positive autoregulatory affect of Qk6 is posttranslational: ectopically expressed Qk6 may stabilize endogenous Qk6 protein through increased homo-dimerization. Alternatively or in addition, when highly expressed Qk6 protein might bind the Qk6 3'UTR and promote translation of its own mRNA. As shown in Figure 3C, a PAR-CLIP peak indicates that Qk protein binds to the Qk6 3'UTR when overexpressed in HEK293T cells, so we tested whether overexpression of Qk6 in C2C12 cells might promote translation of our Qk6 3'UTR reporter. We transfected C2C12 cells with either Qk6 3'UTR reporter and RFP control reporter only, or in combination with overexpressed Myc-Qk5 or Myc-Qk6 and measured both reporter RNA and protein levels. Western blotting with Qk5 and

Qk6 specific antibodies showed overexpression of each Qk isoform (Fig 4E). Reporter mRNA levels were measured by RT-qPCR. Reporter protein levels were measured by automated quantitative microscopy, which detects both GFP and RFP levels, in at least 100 total cells from 3 independent biological replicates. Results of these measurements are summarized in Figure 4F, and the relative mRNA and protein levels were used to calculate translational efficiency. Overexpression of Myc-Qk5 did not significantly change mRNA or protein levels (Fig 4F), consistent with analysis of endogenous Qk6 expression after Qk5 overexpression (Fig 4D). Also consistent with Qk6 expression patterns described above, overexpressing Myc-Qk6 resulted in a net positive effect on translational efficiency of the Qk6 3'UTR reporter gene (Fig 4F). Thus, the Qk6 3'UTR is an important regulatory point of control within the *Qk* autoregulatory network. Taken together with the data in Figure 3, these observations suggest that Qk proteins bind to the Qk6 3'UTR to promote expression of Qk6 at both the mRNA (mediated by Qk5) and protein (mediated by Qk6) level.

Qk5 negatively regulates its own mRNA level expression

To determine if increased amounts of Qk5 influenced *Qk* gene expression, we also tested the affect of ectopic Myc-Qk5 expression on Qk isoform abundance. C2C12 myoblasts were transfected with Myc-Qk5 and both protein and RNA were analyzed for changes in abundance of Qk5 and Qk6. No significant changes are observed in Qk6 mRNA and protein levels when Qk5 is overexpressed (Fig 4C and 4D). Additionally, no significant change is observed in endogenous Qk5 protein abundance when Myc-Qk5 is overexpressed (Fig. S3A). However, endogenous Qk5

mRNA levels are significantly reduced when Myc-Qk5 is ectopically expressed in C2C12 myoblasts (Fig S3B). Therefore, when Qk5 protein levels become very high in C2C12 cells, endogenous Qk5 mRNA levels are reduced; which may serve as a negative autoregulatory feedback loop as has been observed for other splicing factor RBPs ((Wollerton et al. 2004; Lareau et al. 2007; Ni et al. 2007).

Qk RNA nuclear accumulation requires Qk5 protein

Having determined that Qk5 specifically regulates splicing (a nuclear function), we tested if nuclear accumulation of Qk RNA was dependent on Qk5. Since siRNA-mediated mRNA degradation is carried out in the cytoplasm (Sen and Blau 2005), we hypothesized that if Qk5 is required for nuclear Qk RNA accumulation, then loss of Qk5 would result in reduction of nuclear Qk RNA levels. To test this, we performed subcellular fractionation of C2C12 cells treated with siNT, siQk5, or siQk6&7, and performed protein and RNA expression analyses. Western blots showed efficient knockdown of Qk protein isoforms under siQk5 treatment in both whole cell lysate (WCE) and nuclear fractions (Fig 5A). Both Gapdh and histone H3 proteins were observed in WCE western blots (Fig 5A, left), however histone H3 was enriched but Gapdh was nearly undetectable in nuclear fractions (Fig 5A, right), indicating efficient enrichment of nuclei in nuclear fractions. Interestingly, loss of Qk5 under siQk5 treatment also resulted in significant reduction of both Qk5 and Qk6&7 mRNAs in nuclear-enriched cell fractions (Fig 5B). Therefore, Qk5 protein is required for nuclear accumulation of Qk RNA, and may promote early post-transcriptional RNA processing events required for pre-mRNA or mRNA biogenesis.

To test the possibility that Qk5 regulates pre-mRNA processing steps required for nuclear transcript accumulation, we devised a RT-qPCR strategy to measure total Qk5 RNA, unspliced Qk5 pre-mRNA, and Qk5 transcript that extends beyond the two putative poly-adenylation sites (referred to here as "read-through Qk5") in nuclear C2C12 RNA fractions. We designed RT-qPCR primers to make the above measurements (Fig 5C), and tested what the consequences of loss of Qk5 or Qk6&7 were on nuclear Qk5 transcript expression. Consistent with a reduction of total nuclear Qk5 transcript, we observe a significant decrease in the amount of unspliced Qk5 relative to nuclear 7SK RNA after siQk5 treatment (Fig 5D, left). Surprisingly we also observe a similar magnitude decrease in unspliced Qk5 relative to 7SK after siQk6&7 treatment (Fig 5D, left). However, we determined nuclear unspliced Qk5 relative to total nuclear Qk5 actually increased by 40% under siQk5 treatment (Fig 5D, right) by normalizing unspliced nuclear Qk5 to total nuclear Qk5. Therefore nuclear Qk5 protein is required for efficient splicing of its own pre-mRNA final intron. Additionally, under siQk6&7 treatment, we observed nearly a 60% decrease in unspliced Qk5 relative to total nuclear Qk5 (Fig 5D, right), indicating an increase in splicing efficiency. Taken together, these results suggest that Qk5 promotes nuclear accumulation of its own transcript by splicing the last intron from this pre-mRNA resulting in a spliced and stable mature mRNA. Furthermore, in the absence of Qk6&7, Qk5 protein level increases (Fig 4A), correlating with an increase in splicing efficiency observed here as removal of the final Qk5 intron (Fig 5D) and also in the

previously described cassette exons Capzb exon 9 (Fig S2F) and Rai14 exon 11 (Fig S2G).

Additionally, Qk5 might promote polyadenylation of its own transcript which could also affect nuclear RNA abundance. To test this we measured Qk5 "read through" RNA (non-polyadenylated Qk5 RNA) level in C2C12 nuclear fractions of RNA under siNT, siQk5 or siQk6&7 treatment. We detected a decrease in Qk5 read through RNA relative to 7SK RNA under siQk5 treatment, but no significant change with siQk6&7 treatment (Fig 5E, left side). However, since total nuclear Qk5 is reduced under these experimental conditions (Fig 5B), we also measured read-through Qk5 normalized to total nuclear Qk5 under the different RNAi conditions. We observed no significant change in Qk5 read-through RNA relative to total nuclear Qk5 RNA with either siQk5 or siQk6&7 treatment (Fig 5E, right side). Therefore, Qk isoforms do not regulate polyadenylation of Qk5 transcript. Taken together, these data indicate that autoregulation of Qk5 nuclear abundance is promoted through splicing, not polyadenylation, and suggest that Qk5 is the master regulator of *Qk* expression in C2C12 myoblasts.

Tissue specific Qk isoform expression patterns and autoregulation

The auto- and cross-isoform regulatory interactions described here are specific to C2C12 myoblasts, where Qk5 is the predominant isoform (Fig S1B). However, we wanted to test if these cross-isoform and autoregulatory interactions are also applicable in different cell types, and in cases where Qk5 levels are not as high and

Qk6 levels are higher. For example, Qk6 and Qk7 expression is higher in adult mouse brain and Qk5 is lower (Hardy et al. 1996), but it is unknown if a similar autogenous regulation network exists in these cell types or cell types other than mouse C2C12 myoblasts. To test this possibility, we first needed to accurately determine the Qk isoform-specific protein expression profiles in various cell and tissue types of interest. We performed quantitative western blotting with infrared imaging comparing C2C12 cell, rat C6 glioma cell, rat CG4 oligodendrocyte precursor cell, adult mouse optic nerve tissue, and adult mouse cerebellum tissue protein lysates. Western blots were simultaneously probed with antibodies directed toward each Qk isoform and PanQk to measure relative isoform abundance. The relative Qk protein isoform levels are shown as a percentage of total Qk protein in each cell or tissue type (Fig 6A). C2C12 cells have a Qk5 high (86%) and Qk6 low (14%) abundance profile (with Qk7 nearly undetectable, as previously described Fig S1). In contrast, adult mouse optic nerve and cerebellum tissue have more abundant Qk6 or Qk7 protein levels than Qk5, respectively (Fig 6A). Interestingly, rat C6 glioma and CG4 oligodendrocyte cells have similar relative levels of Qk5 protein when compared to one another (61% or 56%, respectively; Fig 6A). These relative Qk5 levels are 25-30% lower than those observed in C2C12 cells. Furthermore, in C6 glioma cells, 2.5 times more Qk6 is expressed than in C2C12 cells (35% relative Qk6 versus 14%; Fig 6A). Therefore, Qk5 predominant levels were observed in C2C12 cells, Qk6 or Qk7 predominant abundance was observed in mouse optic nerve and cerebellum respectively, and intermediate levels of Qk5 were observed with moderate levels of either Qk6 (C6

glioma) or Qk7 (CG4 cells). These observations indicate that a variable spectrum of relative Qk protein isoform levels are expressed across different cell and tissue types, and suggest that different Qk isoform ratios may exist to support different cell types' physiological demands.

Analysis of the relative Qk protein isoforms provided us with the opportunity to test if the cross-isoform regulatory interactions observed in C2C12 myoblasts are also conserved in glial cells. Since C6 glioma (a rat glial cell derivative; (Benda et al. 1971)) cells express more Qk6 protein and less Qk5 protein than C2C12, this is an excellent system in which to test the presence and consequences of *Qk* autogenous regulation. In this cellular system, we presume the contribution to the autogenous regulatory circuitry of Qk5 will be less, and Qk6 will be more. Specifically, we predict that Qk6 will have a more significant role in negative regulation of Qk5 protein abundance, and that Qk5 will not be absolutely required for Qk6 mRNA expression. We tested these hypotheses by transfecting rat C6 glioma cells with siNT, siQk5, and siQk6&7 then collected protein and RNA to analyze gene expression of the Qk isoforms under these conditions. Consistent with observations in C2C12 cells, we found that relative Qk5 protein increased when Qk6&7 were knocked down with siRNA (Fig 6B, top left). The protein level increased at a slightly greater magnitude than observed in C2C12 cells (23% increase vs. 17% increase), supporting the hypothesis of a greater regulatory contribution by Qk6. However in C6 glioma cells, Qk5 mRNA level showed a significant decrease under siQk6&7 treatment (Fig 6B, bottom left). Interestingly, this corresponded to an increase in translational efficiency

of 90%, two-fold higher than observed in C2C12 cells under the same experimental conditions. These data indicate that Qk6 inhibits protein-level expression of Qk5 in rat C6 glioma cells as well as in C2C12 cells, and suggest that in the presence of more Qk6, a consequence of translational repression of Qk5 mRNA may be protection of this mRNA from decay.

Analysis of Qk6 and Qk7 expression in siQk5 transfected C6 glioma cells also supported the autogenous regulatory network by correlating with observations in C2C12 cells: both Qk6&7 mRNA (Fig 6E) and protein (Fig 6E and Fig S4B for Qk7) levels were reduced by siQk5 treatment. Unlike in C2C12 cells, however, the magnitude reduction of Qk6 when Qk5 was targeted with siRNA was much less than knocking down Qk6 itself with siRNA. This was determined by calculating the relative Qk6 protein abundance ratio in siQk5 vs. siQk6&7 in C2C12 myoblasts and C6 glioma cells: we observed a 1:1 ratio in C2C12 myoblasts and a 2.4:1 ratio in C6 glioma cells (from Figs 3A and 6B right). Therefore, in C6 glioma cells the absolute requirement of Qk5 for Qk6 accumulation is alleviated. This may potentially be explained by an increase in positive autoregulation of Qk6 due to higher isoform ratio of Qk6 to Qk5 in C6 cells than in C2C12 cells, consistent with the increase in translation of Qk6 in C2C12 cells when Qk6 is overexpressed (Fig 4D&F). Taken together with the above results, these data demonstrate the autogenous regulatory network uncovered in C2C12 cells is present even when Qk isoform ratios are shifted toward higher Qk6 and in cells of glial origin.

Discussion

Here we systematically test the function of Qk isoforms in post-transcriptional regulation of gene expression in C2C12 myoblasts. We find that Qk5 and Qk6 primarily function in distinct, non-overlapping manners, demonstrating a clear division of labor between these proteins. In the context of splicing, this division of labor is enforced by subcellular localization: Qk5 is nuclear and can function as a stand-alone splicing factor whereas Qk6, in the absence of Qk5, is cytoplasmic and can not regulate splicing. Using siRNA to specifically target either the Qk5 or Qk6&7 mRNAs, we find additional evidence for this splicing division of labor (using reporter-based, candidate, and global approaches), and surprisingly uncover a unique and intricate autoregulatory feedback loop (Fig 7) between Qk isoforms that, in turn, regulate the isoforms (and thus the functions) expressed from *Qk*. These data provide novel insight into how RBP families parse out functional regulatory roles to individual family members.

A splicing division of labor: requirement for Qk nuclear localization

We initially tested if different Qk isoforms (Qk5 and Qk6) could regulate splicing, since we had previously shown that collectively, Qk regulated muscle specific splicing (Hall et al. 2013). An overexpression-based approach suggested that heterodimerization of endogenous Qk5 with ectopically expressed Qk6 could result in nuclear localization and splicing function (Fig 1B). This is consistent with previous findings showing Qk heterodimers could form during overexpression-based experiments (Wu et al. 1999) and their ectopic formation can re-direct subcellular

localization of Qk isoforms to cellular compartments not normally occupied by that isoform (Pilotte et al. 2001) Here, we additionally show a functional consequence whereby Qk6 gains a splicing function (Fig 1B) dependent on Qk5 that it cannot carry out on its own (Fig 1D). Consistent with nuclear localization being sufficient for splicing level regulation, we find that nuclear "QkBody" mutant proteins can also function as splicing regulators, but in this case, independent of endogenous Qk5 and thus heterodimerization (Fig 2C). Thus, the functional specificity of Qk5 and Qk6 isoforms in splicing are encoded as localization signals within their unique C-terminal amino acids: Qk5 contains a NLS and can function as a splicing factor, but Qk6 contains a nuclear export signal or cytoplasmic localization signal and can not function alone as a splicing factor. The physiological relevance of Qk heterodimers remains to be determined, but in HITA cells, which express high levels of both Qk5 and Qk6, Qk6 co-localizes with nuclear speckles in untransfected cells, therefore may have some bona fide functional contribution to splicing (van der Veer et al. 2013).

Autogenous regulation of Quaking determines functional output

An unexpected finding when dissecting the Qk isoforms' function in splicing is that Qk5 positively regulates expression of Qk6, and Qk6 negatively regulates expression of Qk5 (Figs 3 and 4). Additionally, Qk6 positively regulates expression of itself (Fig 4). Consistent with our hypothesis that different Qk isoforms regulate different steps in post-transcriptional gene expression, we find that Qk5 regulates mRNA level expression of Qk6 (Fig 3), but Qk6 regulates protein-level expression of Qk5 and Qk6 (Fig 4), consistent with cytoplasmic localization and a function in

translation (Saccomanno L 1999; Zhao et al. 2010). We also demonstrate here that in C2C12 cells loss of Qk5 results in a near absolute loss of expression of Qk6 and Qk7 (Fig 3A&B and Fig S2B&D). Observations in the *Quaking*$^{lethal-1}$ mouse, which reportedly lacks Qk5 expression and is embryonic lethal, suggested a requirement of Qk5 for expression of Qk6 and Qk7 (Shedlovsky et al. 1988; Cox et al. 1999; Bohnsack et al. 2006), but, to our knowledge, this is the first direct molecular evidence that Qk5 is required for Qk6 and Qk7 expression. Our iCLIP-seq analysis shows that Qk protein directly binds to the 3'UTR of Qk6 and Qk7, and Qk5 protein promotes mRNA level expression of both endogenous Qk6 and the Qk6 3'UTR reporter. Taken together, this suggests that Qk5 protein promotes mRNA-level expression of Qk6 via a mechanism involving direct binding to the Qk6 3'UTR.

We also observe that Qk5 positively regulates Qk6 and Qk7 expression in rat C6 glioma cells (Fig 6B). However in this system, the requirement is not absolute and the magnitude change is not as large as compared to C2C12 cells. As mentioned above, the reduction in magnitude of change observed with Qk6 expression when Qk5 is knocked down in C6 cells when compared to C2C12 cells may be due to the higher amount of Qk6 to total Qk protein in C6 cells (Fig 6A), which might positively reinforce its own expression. Consistent with this explanation, we observe an up-regulation of endogenous Qk6 protein when an epitope-tagged version of Qk6 is overexpressed in C2C12 cells (Fig 4C). This suggests that when Qk6 protein levels are higher, positive autoregulation of Qk6 translation may, at least in part, relieve the absolute requirement of Qk5 on Qk6 expression.

Additionally, Qk6 appears to repress the translation of Qk5 in both C2C12 myoblasts and C6 glioma cells. This negative regulation in C2C12 cells, while subtle (protein level expression change is not considered statistically significant), appears to be functionally important, as a significant increase in splicing function is observed under siQk6&7 treatment (Fig S2 F&G) consistent with an increase in Qk5 protein. With higher Qk6 expression, as observed in C6 glioma cells, we observe a larger magnitude increase in translational efficiency in Qk5 (1.8-fold vs 1.4 fold in C2C12 cells) when Qk6 is knocked down (Fig 6B). Thus, with a higher relative amount of Qk6 to total Qk, we also observe an increased magnitude of translational cross-isoform negative regulation. These complicated auto- and cross-isoform regulatory influences culminate to form a bistable switch, which may represent a molecular rheostat that fine tunes the overall functional output of Qk.

Quaking-5 appears to be the most upstream regulator of the Qk RBP family, however, as we show here that Qk5 promotes expression of all Qk RNAs. Consistent with a role in early regulation/RNA biogenesis, Qk5 promotes nuclear accumulation of its own mRNA (Fig 5B) by promoting splicing and removal of the final intron from Qk pre-mRNA (Fig 5D). Therefore, when Qk5 is reduced or absent, the net result is a reduction or absence of other Qk protein isoforms, as shown here with molecular techniques but also described in several interesting mouse mutant models (Hardy et al. 1996; Lu et al. 2003; Bohnsack et al. 2006). The developmental and pathological implications for this regulatory model are wide-ranging. As *QKI* is being identified as a tumor suppressor gene in a number of different cancers (Novikov et al. 2011; Chen

et al. 2012; Lu et al. 2014; Zhao et al. 2014; Zong et al. 2014), it will be important to determine how changes in expression of individual Qk isoforms are altered in both the tumor and tumor microenvironment, relative to the healthy tissue type. The importance of Qk expression for development is well-studied: *Qk* is an essential gene (Li et al. 2003a). Strikingly, additional evidence of the model we present here with Qk5 as master regulator of Qk expression comes from studies of the *Quaking*$^{lethal-1}$ mouse mutant (loss of Qk5 expression) which is also recessive lethal (Shedlovsky et al. 1988; Bohnsack et al. 2006). This underscores the importance not only of overall Qk expression for mammalian development, but suggests that loss of Qk5 is both the molecular and physiological equivalent of complete *Qk* knockout. The identification of this complicated and interconnected RBP regulatory network uncovers a previously unappreciated functional specification of both Qk5 and Qk6, which each uniquely feedback to control overall output from the *Qk* gene and presumably the entire transcriptome.

Materials and Methods

Cell Culture, Lysis, and Transfections

C2C12 cells were routinely cultured in Dulbecco's modified Eagle medium (DMEM) with high glucose (Life Technologies) supplemented with 10% heat inactivated fetal bovine serum (Life Technologies) at 37°C with 5% CO_2. For differentiation experiments, C2C12 cells were allowed to reach ~90% confluency, then the media was changed to DMEM supplemented with 5% horse serum (Life Technologies) and

the cells were harvested 72h post media change. Protein lysates were prepared by directly lysing cells in a buffer consisting of 100 mM Tris-HCl pH 7.4, 0.5% NP-40, 0.5% Triton X-100, 0.1% SDS, 100 mM NaCl, and 2.5 mM $MgCl_2$. RNA extraction was performed using the TriZOL reagent (Life Technologies) according to the manufacturer's suggestions and genomic DNA was degraded using Turbo DNase (Life Technologies) according to the manufacturer's specifications. Transfections were performed using the Lipofectamine2000 reagent (Life Technologies) according to the manufacturer's specifications.

Plasmids, RNAi, and Antibodies

Myc-epitope tagged Qk5 and Qk6 plasmids were a generous gift from Sean Ryder's lab. HA-epitope tagged Qk "Body" and Qk "BodyNLS" mutants were generated by PCR amplification and ligated into the pcDNA3.1 vector (Life Technologies) using BamHI and NotI restriction endonucleases (New England Biolabs). The pcDNA3.1-tdTomato plasmid was constructed by PCR amplifying tdTomato, a gift from Roger Tsien's lab, and ligating into pcDNA3.1 into the BamHI and NotI restriction sites. The Qk6 3'UTR reporter vector was generated by removing the bovine growth hormone 3'UTR by restriction digest and ligating a linker into the vector containing AscI restriction site to mediate 3'UTR cloning. A GFP coding sequence was added by PCR amplification and ligation into BamHI and NotI restriction sites; 6 MS2 stem loops were ligated downstream of the stop codon of GFP, and the Qk6 3'UTR was PCR amplified and ligated into AscI site. Custom siRNAs were designed and purchased from Dharmacon (see supplement for sequences). The splicing reporter

pDup51 has been previously described (Dominski and Kole 1991), and construction of pDup51-Capzb is described by Hall and colleagues (Dominski and Kole 1991; Hall et al. 2013). Mouse monoclonal antibodies for Quaking include: Anti-PanQKI (N147/6; AB_2173149)), Anti-QKI-5 (N195A/16; AB_10674117), Anti-QKI-6 (N182/17; AB_10673512), and Anti-QKI-7 (N183/15; AB_2173148), all from UC Davis/NIH NeuroMab Facility. Rabbit polyclonal antibody to GAPDH (G9545) is from Sigma-Aldrich; mouse monoclonal antibody to Myc-epitope is from Thermo Fisher Scientific (MA1-21316); mouse monoclonal HA antibody clone 16B12 (MMS-101R) is from Covance; rabbit polyclonal antibody to histone H3 (ab1791) is from Abcam; mouse monoclonal antibody to DsRed2 (recognizes tdTomato protein; 632392) is from Clontech.

Subcellular Fractionation

C2C12 cells cultured on 10 cm dishes were washed twice with ice cold PBS then 400 µl Buffer A (10 mM HEPES pH 7.5, 2 mM $MgCl_2$, 15 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM PMSF plus Mini-Protean EDTA-free protease inhibitor tablet (Roche)) was added to the plate, cells were collected by scraping, the pipeted up and down 5 times then incubated on ice 15 minutes. The volume of each sample was determined, then a 1:20 dilution volume of 10% NP-40 was added and samples immediately centrifuged at 5,000 x g for 5 minutes at 4°C. The resulting supernatant was the cytosolic fraction. The nucleoplasm and chromatin fractions were separated as described (Bhatt et al. 2012). Once efficient separation of fractions was

determined, nucleoplasm and chromatin fractions were combined for subsequent analyses.

Western Blotting

Protein lysate was separated by electrophoresis on Tris-Glycine buffered SDS-PAGE. Proteins were then transferred to nitrocellulose membrane, blocked with 5% milk in Tris buffered saline (TBS) and probed with primary antibodies diluted in Tris buffered saline with 0.05% Tween-20 (TBST) and infrared-conjugated secondary antibodies (680 nm and 800 nm IR dye) diluted at 1:20,000 in TBST. Imaging was performed on the Odyssey infrared imaging system (LiCor) with protein abundance calculated by selecting a region of interested and determining the signal intensity relative to loading control protein simultaneously probed on the same membrane.

Reverse Transcription and PCR

Reverse transcription was performed using SuperScript II or SuperScript III reagents (Life Technologies) according to the manufacturer's instructions. PCR was performed as described and splicing analyses for percent included were determined by quantitating included and skipped PCR products on a 2100 BioAnalyzer (Agilent) as described (Hall et al. 2013; Munding et al. 2013). Relative RNA abundance determined by RT-qPCR was also performed as described (Munding et al. 2013).

Immunofluorescence

#1 round glass coverslips were placed into the well of a 24 well plate and coated with Matrigel (Corning) to allow for cell adherence. Indirect immunofluorescence staining was performed by fixing cells in 4% paraformaldehyde (Sigma Aldrich) in PBS pH

8.0 for 30 min at room temperature, then blocked in blocking buffer (5% normal goat serum and 0.3% Triton X-100 in PBS pH 8.0) for 60 min at room temperature, probed with primary antibody in antibody dilution buffer (1% bovine serum albumin and 0.3% Triton X-100 in PBS pH 8.0) overnight at 4°C, and finally probed with secondary antibody (AlexaFluor 568 goat anti-rabbit IgG, AlexaFluor 468 goat anti-mouse IgG2b, or AlexaFluor 568 goat anti-mouse IgG1, each from Life Technologies/Thermo Fisher Scientific) diluted in the same antibody dilution buffer but with the addition of 5% normal goat serum for 1-2 hours in the dark at room temperature. Images were acquired with a Volocity spinning disk confocal system (Perkin-Elmer/Improvision) on a Nikon Eclipse TE2000-E inverted microscope.

Acknowledgements

We thank Sean Ryder, Doug Black, Jeremy Sanford, Rhonda Perriman, David Feldheim, and Susan Strome for reagents, discussion, and advice. This work was supported by National Institutes of Health (NIH) grant R01-GM040478 to M.A., NIH pre-doctoral T32 GM008646 to WSF, and California Institute of Regenerative Medicine (CIRM) pre-doctoral fellowship award TG2-01157 to WSF.

Adereth Y, Dammai V, Kose N, Li R, Hsu T. 2005. RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. *Nature cell biology* **7**: 1240-1247.

Benda P, Someda K, Messer J, Sweet WH. 1971. Morphological and immunochemical studies of rat glial tumors and clonal strains propagated in culture. *Journal of neurosurgery* **34**: 310-323.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**: 279-290.

Bohnsack BL, Lai L, Northrop JL, Justice MJ, Hirschi KK. 2006. Visceral endoderm function is regulated by quaking and required for vascular development. *Genesis (New York, NY : 2000)* **44**: 93-104.

Caceres JF, Screaton GR, Krainer AR. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development* **12**: 55-66.

Cazalla D, Zhu J, Manche L, Huber E, Krainer AR, Caceres JF. 2002. Nuclear export and retention signals in the RS domain of SR proteins. *Molecular and cellular biology* **22**: 6871-6882.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G et al. 2012. Muscleblind-like 2-Mediated Alternative Splicing in the Developing Brain and Dysregulation in Myotonic Dystrophy. *Neuron* **75**: 437-450.

Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N et al. 2012. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes & development* **26**: 1459-1472.

Cox RD, Hugill A, Shedlovsky A, Noveroske JK, Best S, Justice MJ, Lehrach H, Dove WF. 1999. Contrasting effects of ENU induced embryonic lethal mutations of the quaking gene. *Genomics* **57**: 333-341.

Damianov A, Black DL. 2010. Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA (New York, NY)* **16**: 405-416.

Dominski Z, Kole R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* **11**: 6075-6083.

Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA et al. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature structural & molecular biology* **17**: 187-193.

Ebersole TA, Chen Q, Justice MJ, Artzt K. 1996. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genetics* **12**: 260-265.

Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. 2002. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Human molecular genetics* **11**: 805-814.

Fu XD, Ares M, Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature reviews Genetics* **15**: 689-701.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature structural & molecular biology* **12**: 691-698.

Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, Ares M, Jr., Otis TS, Black DL. 2012. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes & development* **26**: 445-460.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M, Jr., Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43**: 706-711.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP, Ares M, Jr. 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA (New York, NY)* **19**: 627-638.

Hardy RJ, Loushin CL, Friedrich Jr VL, Chen Q, Ebersole TA, Lazzarini RA, Artzt K. 1996. Neural Cell Type-Specific Expression of QKI Proteins Is Altered in quakingviable Mutant Mice. *J Neurosci* **16**: 7941-7949.

He F, Dang W, Abe C, Tsuda K, Inoue M, Watanabe S, Kobayashi N, Kigawa T, Matsuda T, Yabuki T et al. 2009. Solution structure of the RNA binding domain in the human muscleblind-like protein 2. *Protein science : a publication of the Protein Society* **18**: 80-91.

Ishidate T, Yoshihara S, Kawasaki Y, Roy BC, Toyoshima K, Akiyama T. 1997. Identification of a novel nuclear localization signal in Sam68. *FEBS letters* **409**: 237-241.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews Genetics* **8**: 533-543.

Koizumi J, Okamoto Y, Onogi H, Mayeda A, Krainer AR, Hagiwara M. 1999. The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). *The Journal of biological chemistry* **274**: 11125-11131.

Kondo T, Furuta T, Mitsunaga K, Ebersole TA, Shichiri M, Wu J, Artzt K, Yamamura K, Abe K. 1999. Genomic organization and expression analysis of the mouse qkI locus. *Mammalian genome : official journal of the International Mammalian Genome Society* **10**: 662-669.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**: 926-929.

Larocque D, Fragoso G, Huang J, Mushynski WE, Loignon M, Richard Sp, Almazan G. 2009. The QKI-6 and QKI-7 RNA Binding Proteins Block Proliferation and Promote Schwann Cell Myelination. *PloS one* **4**: e5867.

Larocque D, Galarneau A, Liu H-N, Scott M, Almazan G, Richard S. 2005. Protection of p27Kip1 mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat Neurosci* **8**: 27-33.

Larocque D, Pilotte J, Chen T, Cloutier F, Massie B, Pedraza L, Couture R, Lasko P, Almazan G, Richard S. 2002. Nuclear Retention of MBP mRNAs in the Quaking Viable Mice. *Neuron* **36**: 815-829.

Li Z, Takakura N, Oike Y, Imanaka T, Araki K, Suda T, Kaname T, Kondo T, Abe K, Yamamura K. 2003a. Defective smooth muscle development in qkI-deficient mice. *Development, growth & differentiation* **45**: 449-462.

Li Z, Takakura N, Oike Y, Imanaka T, Araki K, Suda T, Kaname T, Kondo T, Abe K, Yamamura K-i. 2003b. Defective smooth muscle development in qkI-deficient mice. *Development Growth & Differentiation* **45**: 449-462.

Li Z, Zhang Y, Li D, Feng Y. 2000. Destabilization and Mislocalization of Myelin Basic Protein mRNAs in quaking Dysmyelination Lacking the QKI RNA-Binding Proteins. *J Neurosci* **20**: 4944-4953.

Lu W, Feng F, Xu J, Lu X, Wang S, Wang L, Lu H, Wei M, Yang G, Wang L et al. 2014. QKI impairs self-renewal and tumorigenicity of oral cancer cells via repression of SOX2. *Cancer biology & therapy* **15**: 1174-1184.

Lu Z, Zhang Y, Ku L, Wang H, Ahmadian A, Feng Y. 2003. The quakingviable mutation affects qkI mRNA expression specifically in myelin-producing cells of the nervous system. *Nucleic acids research* **31**: 4616-4624.

Makeyev EV, Zhang J, Carrasco MA, Maniatis T. 2007. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Molecular cell* **27**: 435-448.

Mandler MD, Ku L, Feng Y. 2014. A cytoplasmic quaking I isoform regulates the hnRNP F/H-dependent alternative splicing pathway in myelinating glia. *Nucleic acids research* **42**: 7319-7329.

Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. 2001. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Human molecular genetics* **10**: 2165-2170.

Mattaj IW, De Robertis EM. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell* **40**: 111-118.

Moreland RB, Nam HG, Hereford LM, Fried HM. 1985. Identification of a nuclear localization signal of a yeast ribosomal protein. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 6561-6565.

Munding EM, Shiue L, Katzman S, Donohue JP, Ares M, Jr. 2013. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Molecular cell* **51**: 338-348.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O,ÄôBrien G, Shiue L, Clark TA, Blume JE, Ares M. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & development* **21**: 708-718.

Novikov L, Park JW, Chen H, Klerman H, Jalloh AS, Gamble MJ. 2011. QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Molecular and cellular biology* **31**: 4244-4255.

Pilotte J, Larocque D, Richard S. 2001. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes & development* **15**: 845-858.

Query CC, Bentley RC, Keene JD. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**: 89-101.

Richard S. 2010. Reaching for the stars: Linking RNA binding proteins to diseases. *Advances in experimental medicine and biology* **693**: 142-157.

Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA (New York, NY)* **10**: 1449-1458.

Saccomanno L LC, Jan E, Punkay E, Artzt K, and Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 12605-12610.

Saccomanno L, Loushin C, Jan E, Punkay E, Artzt K, Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 12605-12610.

Samson ML. 1998. Evidence for 3' untranslated region-dependent autoregulation of the Drosophila gene encoding the neuronal nuclear RNA-binding protein ELAV. *Genetics* **150**: 723-733.

Sanford JR, Gray NK, Beckmann K, Caceres JF. 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes & development* **18**: 755-768.

Sen GL, Blau HM. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature cell biology* **7**: 633-636.

Shedlovsky A, King TR, Dove WF. 1988. Saturation germ line mutagenesis of the murine t region including a lethal allele at the quaking locus. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 180-184.

Sidman R, Dickie M, Appel S. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. *Science (New York, NY)* **144**: 309-311.

Singh RK, Xia Z, Bland CS, Kalsotra A, Scavuzzo MA, Curk T, Ule J, Li W, Cooper TA. 2014. Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. *Molecular cell* **55**: 592-603.

Spellman R, Llorian M, Smith CWJ. 2007. Crossregulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Molecular cell* **27**: 420-434.

Suenaga K, Lee KY, Nakamori M, Tatsumi Y, Takahashi MP, Fujimura H, Jinnai K, Yoshikawa H, Du H, Ares M, Jr. et al. 2012. Muscleblind-like 1 knockout mice reveal novel splicing defects in the myotonic dystrophy brain. *PloS one* **7**: e33218.

Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. 2010. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nature structural & molecular biology* **17**: 306-312.

Teplova M, Patel DJ. 2008. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nature structural & molecular biology* **15**: 1343-1351.

van der Veer EP, de Bruin RG, Kraaijeveld AO, de Vries MR, Bot I, Pera T, Segers FM, Trompet S, van Gils JM, Roeten MK et al. 2013. Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circulation research* **113**: 1065-1075.

Vasudevan S, Steitz JA. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* **128**: 1105-1118.

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

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S et al. 2012. Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* **150**: 710-724.

Wang Y, Vogel G, Yu Z, Richard S. 2013. The QKI-5 and QKI-6 RNA binding proteins regulate the expression of microRNA 7 in glial cells. *Molecular and cellular biology* **33**: 1233-1243.

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular cell* **13**: 91-100.

Wu J, Zhou L, Tonissen K, Tee R, Artzt K. 1999. The Quaking I-5 Protein (QKI-5) Has a Novel Nuclear Localization Signal and Shuttles between the Nucleus and the Cytoplasm. *J Biol Chem* **274**: 29202-29210.

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature structural & molecular biology* **16**: 130-137.

Zearfoss NR, Clingman CC, Farley BM, McCoig LM, Ryder SP. 2011. Quaking regulates Hnrnpa1 expression through its 3' UTR in oligodendrocyte precursor cells. *PLoS genetics* **7**: e1001269.

Zhao L, Ku L, Chen Y, Xia M, LoPresti P, Feng Y. 2006. QKI Binds MAP1B mRNA and Enhances MAP1B Expression during Oligodendrocyte Development. *Mol Biol Cell* **17**: 4179-4186.

Zhao L, Mandler MD, Yi H, Feng Y. 2010. Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 19061-19066.

Zhao Y, Zhang G, Wei M, Lu X, Fu H, Feng F, Wang S, Lu W, Wu N, Lu Z et al. 2014. The tumor suppressing effects of QKI-5 in prostate cancer: a novel diagnostic and prognostic protein. *Cancer biology & therapy* **15**: 108-118.

Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H et al. 2014. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS genetics* **10**: e1004289.

Supplemental Figure 1

Fagg_FigS1

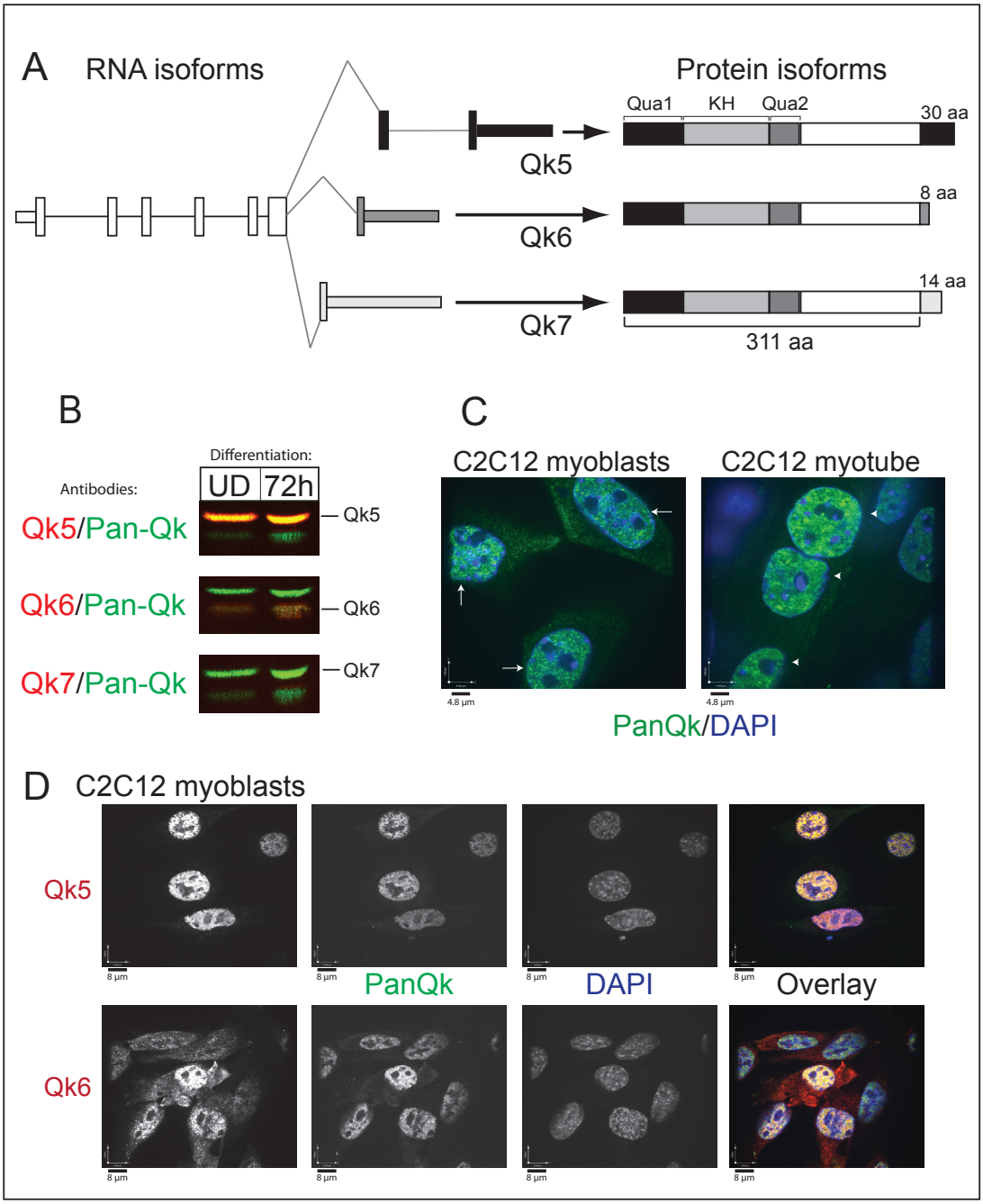


Figure S1: Quaking isoforms expressed in C2C12 myoblasts and myotubes. A. Cartoon of Qk isoforms. Three major Qk RNA isoforms are generated by alternative 3' splice site usage that results in splicing exon 6 to one of three isoform-specific exons (left). These 3 transcripts are translated to Qk5, Qk6, or Qk7 protein. Each of these isoforms share identical 311 amino acids consisting of Quaking homology 1 (QUA1) dimerization domain, hnRNPK homology (KH) and QUA2 domains which mediate RNA binding, and an additional xx amino acids. The number of unique amino acids to each protein isoform is shown: 30 amino acids for Qk5, 8 amino acids for Qk6, and 14 amino acids for Qk7. B. Western blot of undifferentiated C2C12 myoblasts (UD) or C2C12 cells differentiated for 72h (72h) with simultaneously probed with infrared-conjugated secondary antibodies to each Qk isoform (red channel) or PanQk (green channel) which recognizes all Qk isoforms. C. Indirect immunofluorescence showing PanQk staining in green or DAPI staining in blue performed on undifferentiated C2C12 myoblasts (left, arrows show nuclei from individual cells) or differentiated C2C12 myotube (right, arrowheads show three nuclei in single myotube). D. Indirect immunofluorescence in undifferentiated C2C12 myoblasts showing (from left to right) either Qk5 (top) or Qk6 (bottom) grayscale image (red in overlay), PanQk grayscale image (green in overlay), DAPI staining grayscale (blue in overlay), and all three channels as overlay image in color.

Figure 1

Fagg_Fig1

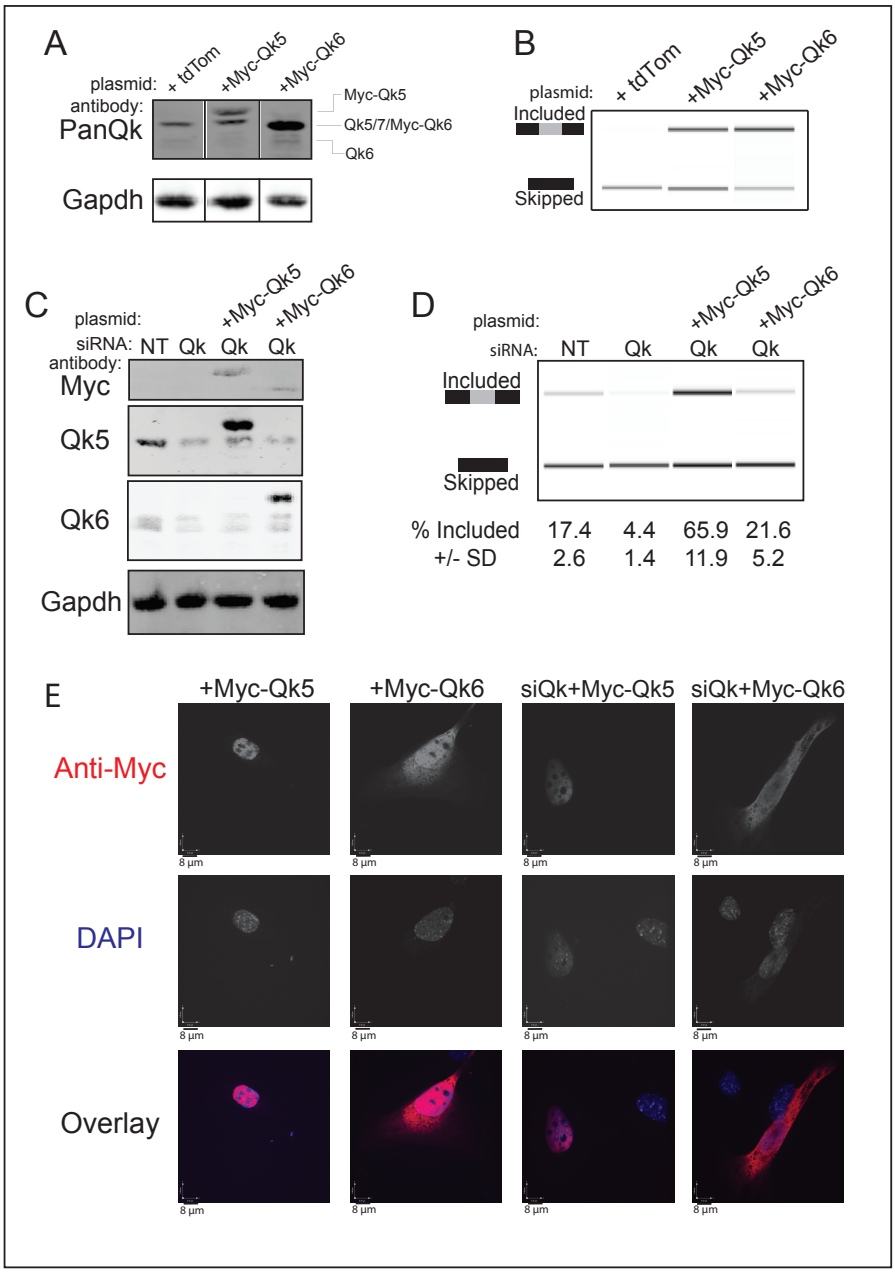


Figure 1: Qk5 is sufficient for splicing function. A. Western blot of protein lysates from C2C12 myoblasts transfected with control vector alone encoding tdTomato, Myc-tagged Qk5, or Myc-tagged Qk6 probed with PanQk antibody or DsRed2 antibody (detects tdTomato, transfection control); migration of different protein forms denoted on right. B. RT-PCR products from Capzb exon 9 splicing reporter analyzed on BioAnalyzer from C2C12 cells transfected as in A. C. Western blot of protein lysates from C2C12 myoblasts transfected with control siRNA (NT), siRNA targeting all Qk isoforms (Qk), siRNA targeting all Qk isoforms co-transfected with plasmid encoding Qk5 (Qk + Myc-Qk5), and siRNA targeting all Qk isoform co-transfected with plasmid encoding Qk6 (Qk + Myc-Qk6) probed with Myc antibody, Qk5 antibody, Qk6 antibody, or Gapdh antibody. D. RT-PCR products from Capzb exon 9 splicing reporter analyzed on BioAnalyzer from C2C12 cells transfected as in C.; mean spliced product percent included is denoted below gel like image +/- standard deviation from 3 independent biological replicates. E. Representative images from transfections described in A and C: indirect immunofluorescence of C2C12 myoblasts showing Myc antibody in grayscale (top; red in overlay), DAPI staining in grayscale (middle; blue in overlay), and channel overlay image at bottom; Myc-Qk5 and Myc-Qk6 overexpression are shown on left respectively, and Qk knockdown with either Myc-Qk5 or Myc-Qk6 add-back is shown on right (respectively).

Figure 2

Fagg_Fig2

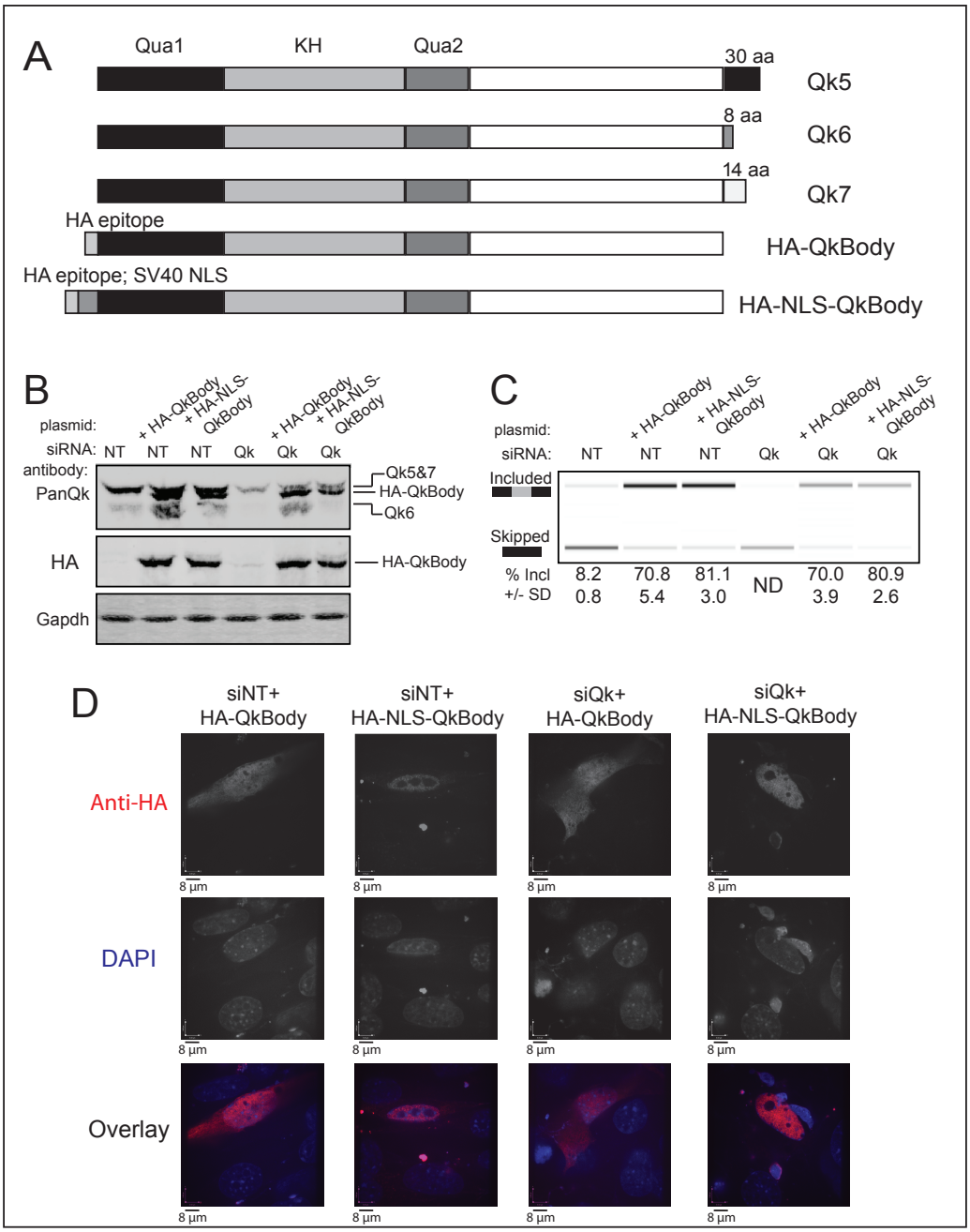


Figure 2: Nuclear localization is sufficient for Qk splicing function. A. Model of Qk protein isoforms and mutant proteins: HA-QkBody consists of N-terminal HA epitope tag and the 311 amino acids common to all Qk proteins and HA-NLS-QkBody consists of N-terminal HA tag and nuclear localization signal from SV40. B. Western blot of protein lysates from C2C12 myoblasts transfected with control siRNA (NT), siNT plus HA-QkBody, siNT plus HA-NLS-QkBody, siRNA targeting all Qk isoforms (Qk), siQk plus HA-QkBody, and siQk plus HA-NLS-QkBody probed with PanQk antibody (top), HA antibody (middle) and Gapdh antibody (bottom). C. RT-PCR products from Capzb exon 9 splicing reporter analyzed on BioAnalyzer from C2C12 myoblasts transfected as noted in B.; mean percent included from three independent biological replicates with standard deviation shown below. D. Indirect immunofluorescence using HA antibody (grayscale at top, red channel in overlay), DAPI staining (grayscale in middle, blue channel in overlay), and overlaid images at bottom of representative C2C12 myoblasts from siNT + HA-QkBody, siNT + HA-NLS-QkBody, siQk + HA-QkBody, and siQk + HA-NLS-QkBody transfections.

Supplemental Figure 2

Fagg_FigS2

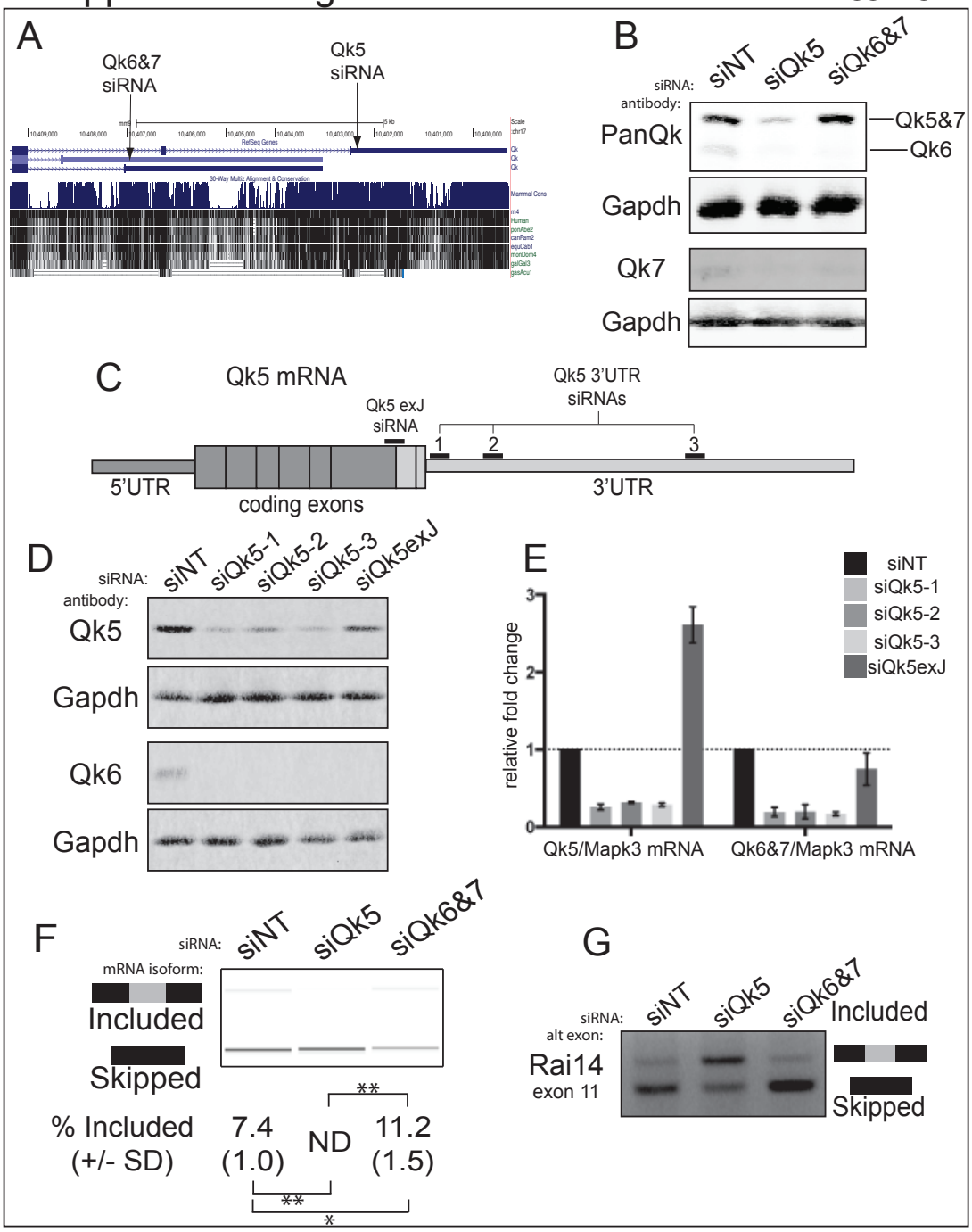


Figure S2: Qk5 is required for Qk6 and Qk7 accumulation and muscle-specific splicing. A. UCSC Genome Browser screen shot showing approximate relative position of where Qk6&7 siRNA and Qk5 siRNA base pair with their respective target transcripts; top track shows RefSeq Genes and bottom track shows Mammalian Conservation. B. Western blot of lysates from C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7 using the PanQk and Gapdh antibodies. C. Cartoon depiction of Qk5 mRNA showing approximate relative positions of base pairing of additional siRNAs against Qk5: Qk5 exJ siRNA (siQk5exJ) base pairs with an exon-exon junction specific to Qk5, Qk5 3'UTR siRNA 1 (siQk5-1) is the siRNA shown in S2A and used in S2B, Qk5 3'UTR siRNA 2 (siQk5-2) targets a region further downstream in the Qk5 3'UTR, and Qk5 3'UTR siRNA 3 (siQk5-3) targets the Qk5 3'UTR most downstream relative to the other 3'UTR siRNAs. D. Western blots of protein lysates from C2C12 mybolasts transfected with siNT, siQk5-1, siQk5-2, siQk5-3, and siQk5exJ siRNAs probed with Qk5, Qk6, and Gapdh antibodies. E. RT-qPCR using RNA extracted from C2C12 myoblasts transfected as in D. showing fold change relative to control siRNA for Qk5 (left) and Qk6&7 (right) normalized to Mapk3. F. RT-PCR products from Capzb exon 9 splicing reporter analyzed on BioAnalyzer from C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7; mean percent included from 3 biological replicates +/- standard deviation is shown below; ** indicates $p < 0.01$, * indicates $p < 0.05$ determined by student's t-test. G. RT-PCR products from RNA extracted from C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7 using primers flanking endogenous Rai14 cassette exon 11 analyzed on agarose gel.

Figure 3

Fagg_Fig3

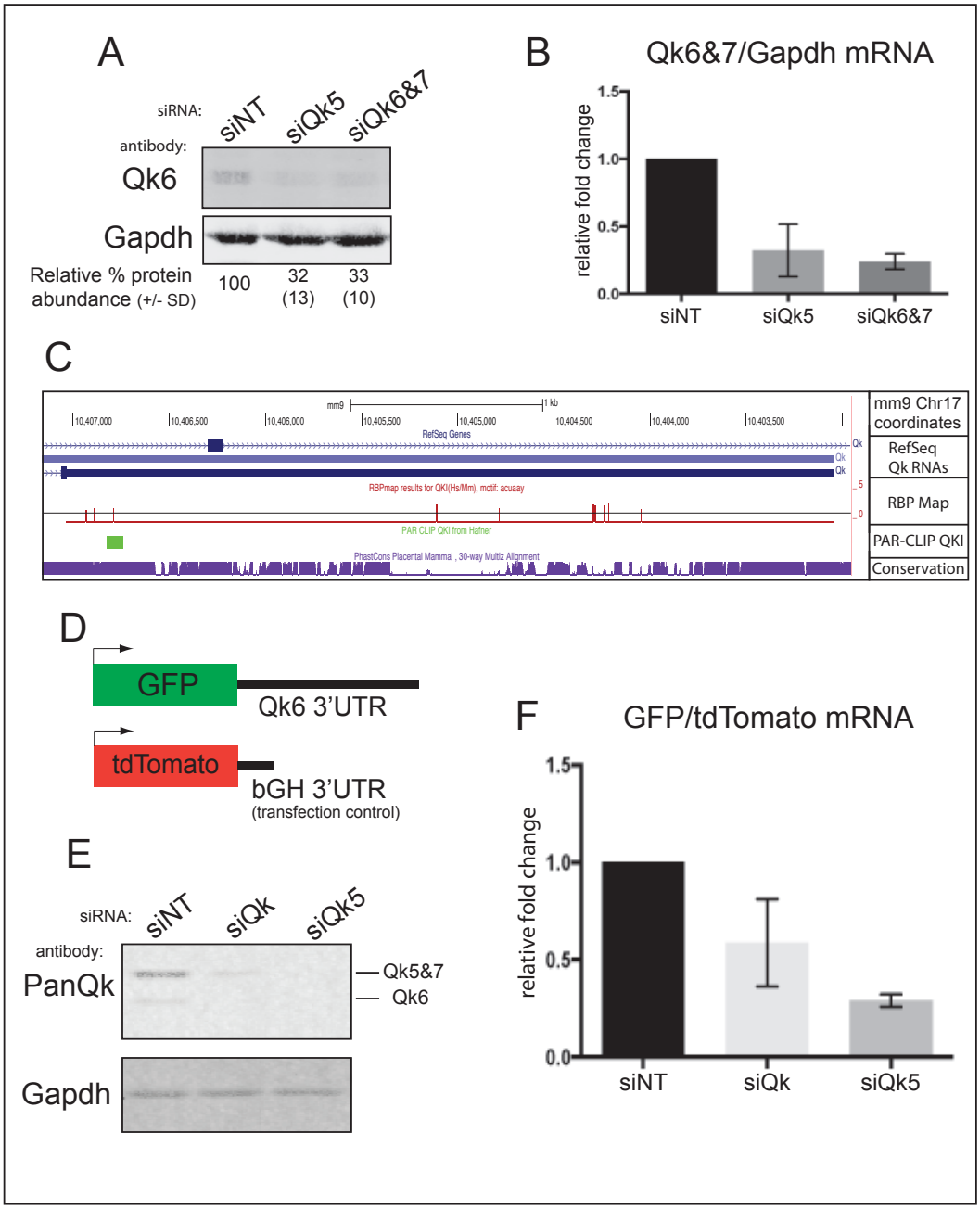
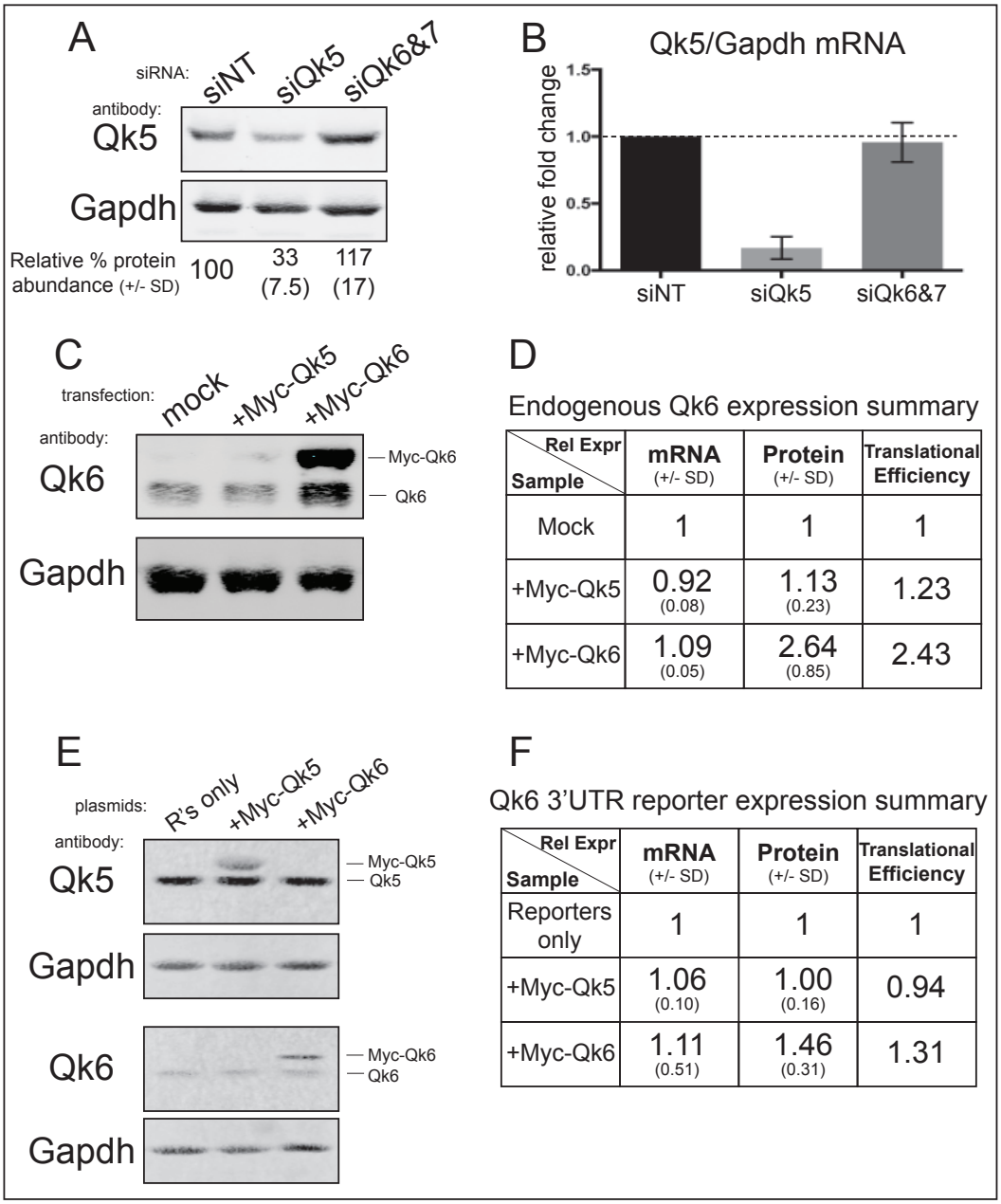


Figure 3: Qk5 promotes Qk6 expression through Qk6 mRNA 3'UTR. A. Western blot of protein lysates extracted from C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7 using Qk6 and Gapdh antibodies; mean percentage of Qk6 protein abundance normalized to Gapdh relative to siNT +/- standard deviation was calculated from 3 independent biological replicates and shown below. B. RT-qPCR using RNA extracted from C2C12 myoblasts transfected as in A for Qk6&7 mRNAs normalized to Gapdh mRNA and shown as fold change relative to siNT; error bars show standard deviation from the mean calculated from 3 independent biological replicates. C. UCSC Genome Browser screen shot shows Qk5 (top), Qk7 (middle), and Qk6 (bottom) transcripts in RefSeq track, RBP Map track shows predicted Qk binding sites, PAR-CLIP QKI track shows a peak calculated by Hafner et al from QKI PAR-CLIP dataset, and Conservation track shows mammalian sequence conservation. D. Cartoon depicting GFP-Qk6 3'UTR and tdTomato reporter gene constructs; tdTomato has a bovine growth hormone (bGH) 3'UTR and is used as a transfection control. E. Western blot analysis of protein extracted from C2C12 myoblasts transfected with siNT, siQk, and siQk5 probed with PanQk and Gapdh antibodies. F. RT-qPCR using RNA extracted from C2C12 myoblasts transfected as in E. for GFP mRNA normalized to tdTomato mRNA displayed as fold change relative to siNT; mean and standard deviations calculated from 3 independent biological replicates.

Figure 4

Fagg_Fig4



Rel Expr / Sample	mRNA (+/- SD)	Protein (+/- SD)	Translational Efficiency
Mock	1	1	1
+Myc-Qk5	0.92 (0.08)	1.13 (0.23)	1.23
+Myc-Qk6	1.09 (0.05)	2.64 (0.85)	2.43

Rel Expr / Sample	mRNA (+/- SD)	Protein (+/- SD)	Translational Efficiency
Reporters only	1	1	1
+Myc-Qk5	1.06 (0.10)	1.00 (0.16)	0.94
+Myc-Qk6	1.11 (0.51)	1.46 (0.31)	1.31

Figure 4: Qk6 inhibits Qk5 protein expression but activates its own protein expression. A. Western blot of protein extracted from C2C12 myoblasts transfected with siNT, siQk5 and siQk6&7 probed with Qk5 or Gapdh antibody; Qk5 protein level is normalized to Gapdh protein and mean percentage protein abundance relative to siNT is calculated from 3 independent biological replicates and is shown at bottom, +/- standard deviation. B. RT-qPCR using RNA extracted from C2C12 myoblasts transfected as in A for Qk5 mRNA normalized to Gapdh mRNA and shown as fold change relative to siNT; mean values and standard deviation calculated from 3 independent biological replicates. C. Western blot of protein lysates from C2C12 myoblasts mock transfected, transfected with Myc-Qk5, and transfected with Myc-Qk6 plasmids probed with antibodies to Qk6 and Gapdh. D. Endogenous Qk6 expression analysis summary of C2C12 myoblasts transfected as in C; RT-qPCR was performed to measure Qk6&7 mRNA and normalized to Gapdh mRNA with values displayed as fold change relative to mock transfected +/- standard deviation; endogenous Qk6 protein abundance measured from western blots using infrared conjugated antibodies also normalized to Gapdh and expressed as fold change relative to mock transfected +/- standard deviation; translational efficiency is calculated by dividing average normalized protein level to average normalized mRNA level; mean and standard deviations all from 3 independent biological replicate samples. E. Western blots of protein lysates from C2C12 myoblasts co-transfected with GFP-Qk6 3'UTR reporter and pCAGGS-mCherry transfection reporter (R's only), both reporters plus Myc-Qk5, and both reporters plus Myc-Qk6 probed with antibodies to Qk5 and Gapdh (top), and Qk6 and Gapdh (bottom). F. GFP-Qk6 3'UTR reporter gene expression analysis summary of C2C12 myoblasts transfected as in E; measurements taken from biological triplicate samples to determine: mean GFP mRNA relative to mCherry mRNA +/- standard deviation in parentheses by RT-qPCR, mean GFP protein relative to mCherry protein by sampling fluorescence intensity of at least 100 individual cells per biological replicate using automated quantitative microscopy with standard deviation shown in parentheses, and mean translational efficiencies; all values are shown as fold change relative to reporters only control transfected cells.

Figure 5

Fagg_Fig5

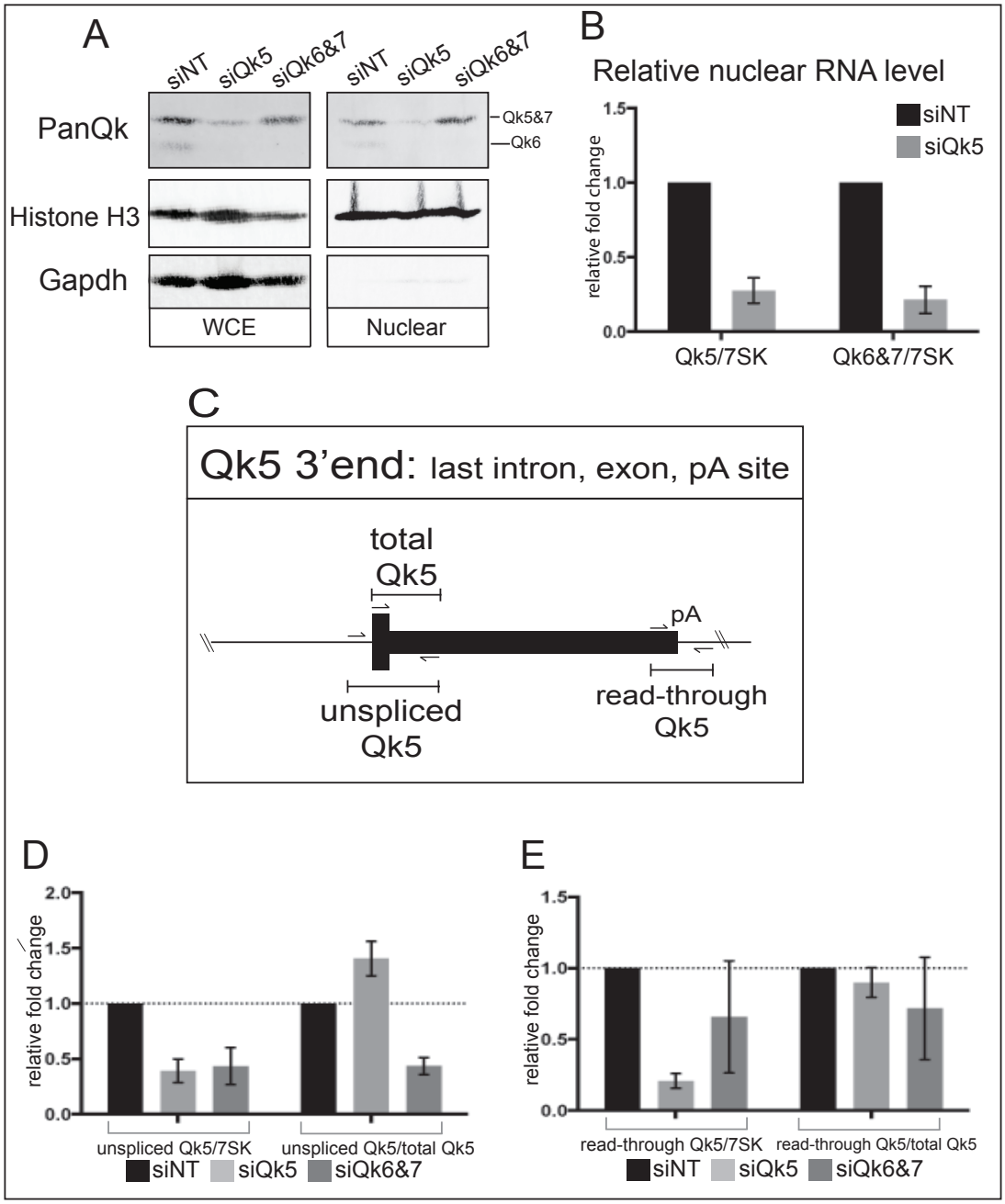


Figure 5: Qk5 is required for nuclear accumulation of Qk RNA. A. Western blot of protein lysates from whole cell extract (WCE, left) or nuclear fraction (nuclear, right) C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7 probed with antibodies against PanQk, histone H3, and Gapdh. B. RT-qPCR measuring nuclear RNA level of C2C12 myoblasts transfected with siNT and siQk5 for Qk5 RNA (left) and Qk6&7 RNA (primer set also amplifies Qk5 last intron, right) relative to 7SK RNA and reported as fold change relative to siNT. C. Schematic showing Qk5 RNA 3'end and primer sets used to measure total Qk5, unspliced Qk5, and read-through (non-polyadenylated) Qk5 reported in D and E. D. RT-qPCR performed on nuclear fraction RNA extracted from C2C12 myoblasts transfected with siNT, siQk5, and siQk6&7; left side shows unspliced Qk5 RNA relative to 7SK RNA and right side shows unspliced Qk5 RNA relative to total Qk5 RNA, and both are reported as fold change relative to siNT. E. RT-qPCR performed on same RNA samples in D; left side shows read-through Qk5 RNA normalized to 7SK RNA and right side shows read-through Qk5 RNA relative to total Qk5 RNA with both reported as fold change relative to siNT.

Figure 6

Fagg_Fig6

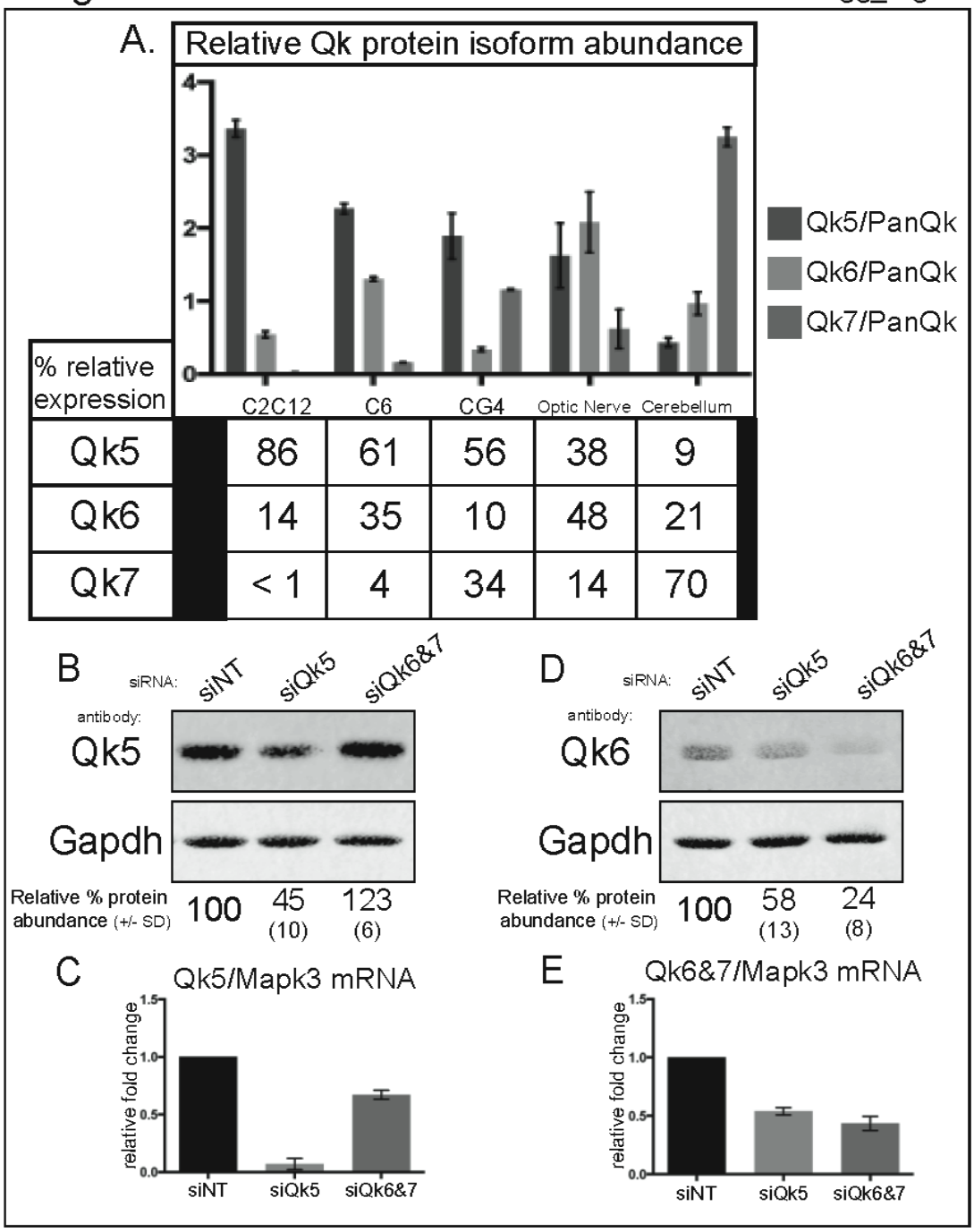


Figure 6: Qk autogenous regulation is conserved in rat C6 glioma cells. A. Quantitation of signal from protein extracted from C2C12 myoblasts, rat C6 glioma cells, rat CG4 oligodendrocyte precursor cells, mouse optic nerve tissue, and mouse cerebellum tissue analyzed on western blots simultaneously probed with infrared conjugated antibodies to Qk5 and PanQk, Qk6 and PanQk, and Qk7 and PanQk, displayed as isoform abundance relative to PanQk in top bar graph, and reported as percentage expression of each isoform relative to total (PanQk) in the table below. B. Western blot of protein extracted from rat C6 glioma cells transfected with siNT, siQk5, and siQk6&7 probed with Qk5 and Gapdh antibodies; quantitation of mean infrared signal reported below as percentage protein abundance relative to siNT and +/- standard deviation of the mean calculated from 3 independent biological replicates. C. RT-qPCR of RNA extracted from rat C6 glioma cells transfected in B measuring Qk5 mRNA normalized to Mapk3 mRNA and displayed as fold change relative to siNT +/- standard deviation. D. Western blot of protein lysates extracted from rat C6 glioma cells described in B and probed with Qk6 and Gapdh antibodies; quantiation of mean infrared signal reported below as percentage protein abundance relative to siNT and +/- standard deviation of the mean calculated from 3 independent biological replicates. E. RT-qPCR of RNA extracted from rat C6 glioma cells described in B measuring Qk6&7 mRNA normalized to Mapk3 mRNA and displayed as fold change relative to siNT +/- standard deviation.

Supplemental Figure 4

Fagg_FigS4

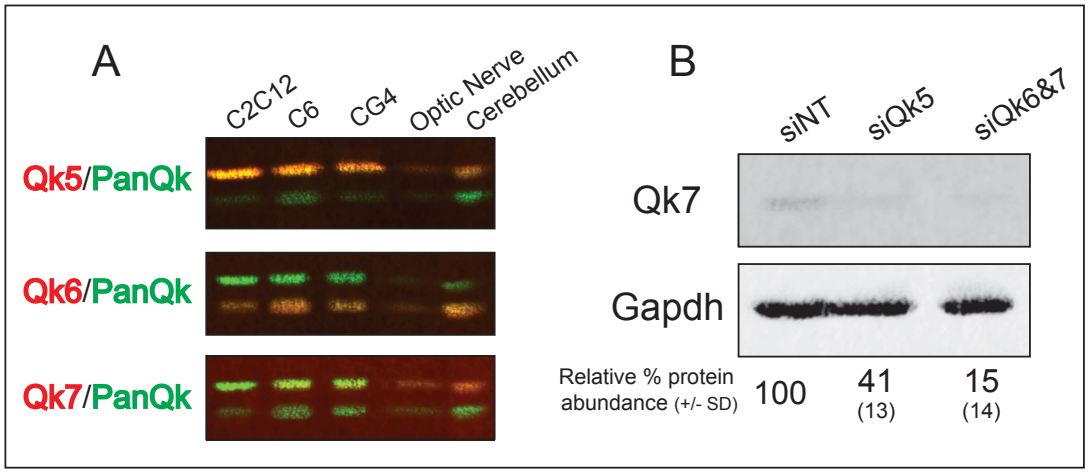


Figure S4: Western blots supplemental to Figure 6. A. Western blots on protein extracted from C2C12 myoblasts, rat C6 glioma cells, rat CG4 oligodendrocyte precursor cells, mouse optic nerve tissue, and mouse cerebellum tissue probed with Qk5 in red channel and PanQk in green channel (top), Qk6 in red channel and PanQk in green channel (middle), and Qk7 in red channel and PanQk in green channel (bottom); representative blots shown from biological triplicate and signal intensities used to calculate data in Fig 6A. B. Western blot using protein extracted from rat C6 glioma cells transfected with siNT, siQk5, and siQk6&7 probed with antibodies to Qk7 and Gapdh; Qk7 signal intensity normalized to Gapdh and reported as mean percentage protein abundance relative to siNT +/- standard deviation of the mean from 3 independent biological replicates.

Figure 7

Fagg_Fig7

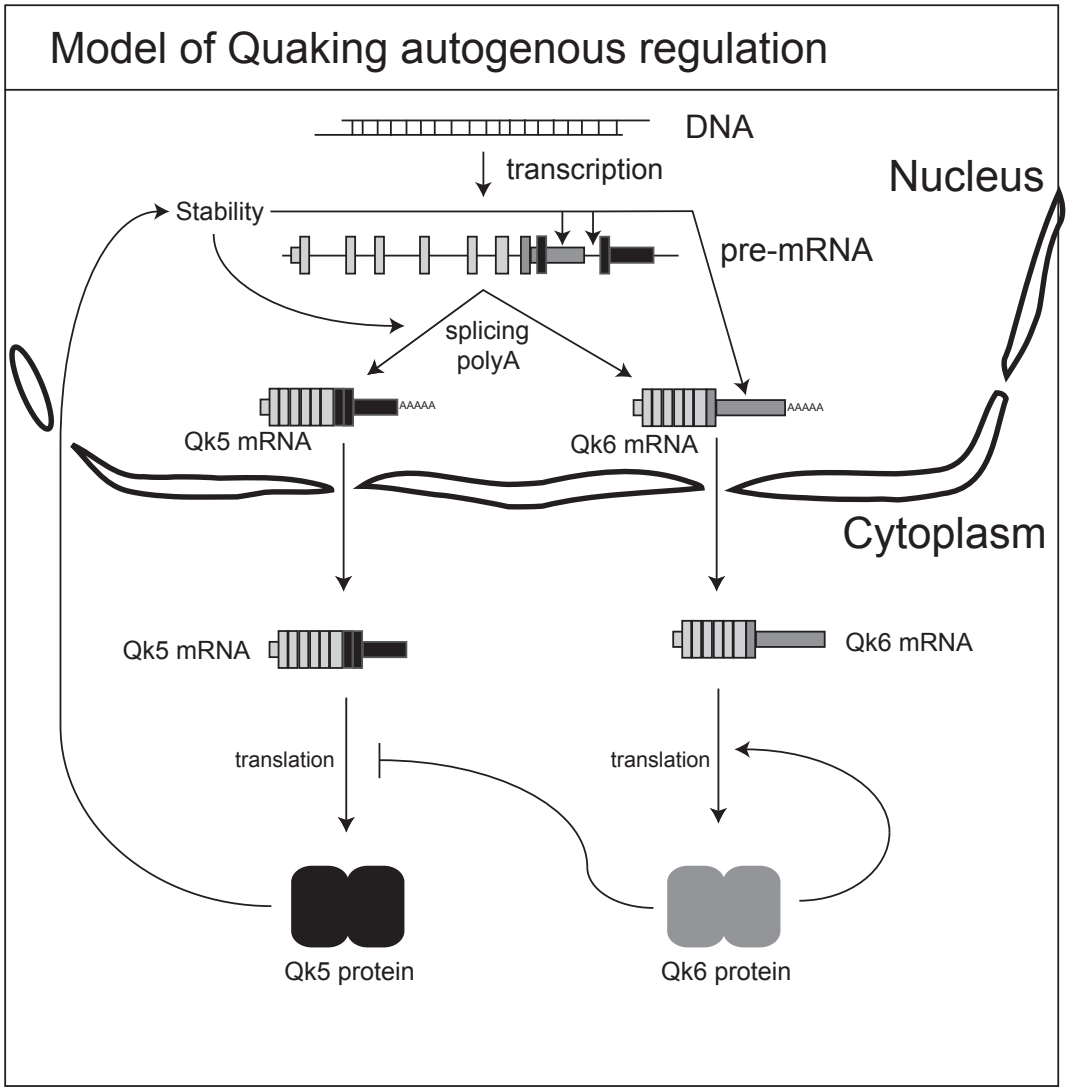


Figure 7: Model of Qk autogenous regulation. The *Qk* gene transcribes a pre-mRNA which is processed to make mature Qk mRNAs which are then exported from the nucleus and translated into Qk5 and Qk6 proteins. The NLS unique to the Qk5 C-terminus mediates nuclear import, where Qk5 promotes nuclear accumulation of Qk pre-mRNA, Qk5 mRNA and Qk6 mRNA through splicing-dependent mechanism and 3'UTR-dependent mechanism for Qk6 mRNA. This allows for accumulation of both Qk mRNAs and proteins. Qk6 protein then negatively regulates the translation of Qk5, and positively regulates its own translation.

Supplemental Information: oligonucleotide sequences

DNA Oligos:

Gene	Purpose	Primer	Sequence
Beta-globin	RT-PCR	F	GCAACCTCAAACAGACACCA
		R	AGCTTGTCACAGTGCAGCTC
Capzb	RT-PCR	F	GGAAAACAAAATCCGAAGCA
		R	CCTCCACCAGGTCGTTCTTA
Rai14	RT-PCR	F	AGTGGTGTAGATGTTATTGATCG
		R	TTCCAAACTCTCGGAAAATGCAGG
Gapdh	RT-qPCR	F	TGAGGCCGGTGCTGAGTATG
		R	TGGTTCACACCCATCACAAACA
Qk-Uni	RT-PCR	F	TGAAGCTGGGTTAATCTACACAC
Qk5	RT-PCR	R	CGGTCACAATCCTTTGGTAAG
Qk6	RT-PCR	R	GCCTTTCGTTGGGAAAGCCAT
Qk7	RT-PCR	R	GCATGACTGGCATTTCAATCC
Qk5(total)	RT-qPCR	F	AGTGGTGTGTTAGGTGCGGTG
		R	GTTCAGAGGTCAGAAGGTCATAGGTTAG
Qk5_Unspl	RT-qPCR	F	TTTCATGTTGTCATTAAGCTGTCA
Qk5_readThr	RT-qPCR	F	TGCTGTTTCTATTAACCCATAGCA
		R	TCAGTGTTAGTCCAGGAACTCAAG
Qk6/7	RT-qPCR	F	TTAACTAGGTGCGGTGGCTAC
		R	GGACACCCAGCACATTGCT
7SK	RT-qPCR	F	GCCTCATTTGGATGTGTCTG
		R	CAAACAAGCTCTCAAGGTCCA
GFP	RT-qPCR	F	GAACAGCGGCAAGTTCTACAG
		R	CGTCCTCCACGTAGGTCTTCT
tdTomato	RT-qPCR	F	CCTCCGAGGACAACAACATGG
		R	ACGTACGCCTTGGAGCCGTAC
mCherry	RT-qPCR	F	CACTACGACGCTGAGGTCAA
		R	TGGTGTAGTCCTCGTTGTGG
GFP	cloning	F	AATTGGATCCACCATGGTGAGCAAGCAGATCCTGAAG
		R	TCGAGCGGCCGCTTACACCCACTCGTGCAGGCTGCCC
tdTomato	cloning	F	AATTGGATCCACCATGGTGAGCAAGGGCGAGGAGGTC
		R	TCGAGCGGCCGCTTACTTGTACAGCTCGTCCAT
HA-Qk:Body	cloning	F	AATCTGGATCCATGTACCCATACGACGTCCCAGACTACGCTGTCGG GGAAATGGAAACGAAGG

HA-Qk:Body$_{NLS}$ Cloning F
AATCTGGATCCACCATGCCTAAGAAGAAGAGGAAGGTTTACCCAT
ACGACGTCCCAGACTACGCTGTCGGGGAAATGGAAACGA
Qk:Body cloning R
GCGGCCGCTTAACCTAACACACCACTGGGTTCA
Qk6_3'UTR cloning F
ACATGGCGCGCCGAATTCAAGAACGGTCTTAACCG
R
ATATAGGCGCGCCTTTATTGCTTTTTCAATTATTCTATTTACAAACA
ACATGGAATTTCTTGGCC

siRNA

siQkEx6M1 sense strand: UCCUUGAGUACCCUAUUGAACCCUU
antisense strand: GGGUUCAAUAGGGUACUCAAGGAUU
siQkEx6M2 sense strand: ACCCUAUUGAACCCAGUGGUGUGUU
antisense strand: CACACCACUGGGUUCAAUAGGGUUU
siQk5-1 sense strand: CUAUGACCUUCUGACCUCUGAAUU
antisense strand: UUCAGAGGUCAGAAGGUCAUAGUU
siQk5-2 sense strand: UCAGUUUAUUUACUAUGCAAUAGUU
antisense strand: CUAUUGCAUAGUAAAUAAACUGAUU
siQk5-3 sense strand: AGUAGAUACUUUUACUAUACAAGGUU
antisense strand: CCUUGUAUAGUAAAAGUAUCUACUUU
siQk5exJ sense strand: UGUUAGGUGCGGUGGCUACUAUU
antisense strand: UAGUAGCCACCGCACCUAACAUU
siQk6&7 sense strand: GAAUUCAAGAACGGUCUUAAUU
antisense strand: UUAAGACCGUUCUUGAAUUCUU

CHAPTER 6:

GENOMEWIDE ANALYSIS OF QUAKING ISOFORMS REVEAL DISTINCT AND OPPOSING FUNCTIONS IN REGULATION OF GENE EXPRESSION

Introduction

RNA binding proteins (RBPs) coordinately regulate multiple RNA processing steps within the gene expression program. Complex regulatory networks have emerged as a theme whereby a single RBP can regulate overlapping and interconnected processing steps on a single substrate RNA molecule. However emerging genomics technologies now allow for identification of all substrate molecules to which an RBP binds, and determination of the consequences on gene expression of RBP loss of function. These powerful techniques have greatly contributed to understanding how single RBPs regulate global gene expression.

The previous chapter demonstrates that Quaking RBP isoforms have different functions in C2C12 myoblasts. Through these specialized functions each isoform is able to feedback into the *Qk* regulatory program to modulate output from the gene. From that study we determined major regulatory targets of the Qk RBPs are Qk pre-mRNA, Qk5 mRNA, Qk6 mRNA, and Qk7 mRNA. These different substrates were regulated in different ways: Qk5 promoted splicing of Qk pre-mRNA and Qk6 mRNA level expression through its 3'UTR, while Qk6 regulated translation of Qk5 and Qk6&7 mRNA. Therefore it is likely that the unique functions of these different Qk isoforms regulate the entire transcriptome in C2C12 myoblasts.

The data present below use a combination of genomics techniques, bioinformatics analysis, and molecular techniques to determine how the Qk5 and Qk6 RBPs regulate different sets of genes and regulatory processes in C2C12 myoblasts. The findings identify both direct and indirect targets of Qk isoform regulation. Of

particular interest is: 1) the observation that Qk isoforms primarily regulate distinct, non-overlapping target sets of transcripts, at the level of transcript abundance and alternative splicing, despite having the same regulatory sequence motif, 2) the identification that Qk5 may regulate Let7 microRNA bioavailability through modulating the level of two competing endogenous RNAs (ceRNAs) that act as "sponges" for Let7, and 3) that Qk5 and Qk6 have opposite effects on regulation of gene expression which are mediated, at least in part, by differential regulation of microRNA biogenesis.

Results

Qk isoforms primarily regulate distinct transcripts and processes

To test the consequences of loss of function of Qk protein isoforms globally, we knocked down individual Qk isoforms in C2C12 myoblasts, or transfected them with control siRNA. We generated RNA-seq libraries from RNA extracted from myoblasts transfected with scrambled siRNA (siNT), siRNA that specifically targeted Qk5 (siQk5), or Qk6&7 (siQk6&7). As noted in the previous chapter, siQk5 treatment results in loss of all Qk isoforms, and siQk6&7 results in specific knockdown of Qk6&7, but also a slight increase in Qk5 protein expression. Initial comparison of changes in mRNA abundance by DESEQ analysis (Anders and Huber 2010) of siQk5 relative to siNT and siQk6&7 relative to siNT find hundreds of changes in each dataset (cutoff $p \leq 0.01$, log2 fold change $\geq |0.5|$, base mean ≥ 50 RPKM; Fig 1A). Both knockdown treatments result in more increases in mRNA

abundance than decreases, but particularly in siQk5 transfected C2C12 myoblasts: over 61% of transcript abundance changes are increasing, suggesting Qk5 is primarily a repressor of transcript abundance.

Since all Qk isoforms share the exact same RNA binding domain, they bind to the same ACUAAY sequence motif. Therefore, one might predict more similarly regulated target transcripts. On the other hand, since Qk isoforms regulate different functions (see previous chapter), the mRNAs regulated at the level of transcript abundance may be very different. To test these possibilities, we investigated the overlap of transcript level abundance changes between the DESEQ. These data indicate that only a small amount of transcripts (9.5% of total) respond to treatment of C2C12 myoblasts with both siQk5 and siQk6&7 (Fig 1B). Thus the majority of transcripts whose abundance changes, changes only with treatment of either siQk5 or siQk6&7, but not both. We validated several of the mRNA abundance changes identified by DESEQ with RT-qPCR and find the transcript level changes agree with bioinformatics findings (Fig 1C). Additional evidence in support of Qk protein isoforms regulating distinct sets of transcripts and thus cellular and molecular processes, is identified by Gene Ontology (GO) analysis. Significant GO terms were only enriched in transcripts that increased in abundance under siQk5 treatment, and also transcripts that increased in abundance under siQk6&7 treatment. Some similar terms were identified between the two enrichment sets, such as “cell differentiation” and “nervous system development” (Fig 1D). However the majority of significantly enriched terms were found in either one dataset or the other, but not both. Particularly

interesting was enrichment of terms associated with muscle contraction and anatomical organization (myofibril assembly, sarcomere organization, actomyosin structure organization) in the siQk5 treated myoblasts, and enrichment of terms associated with transcription and RNA metabolism in the siQk6&7 treated myoblasts (Fig 1D). These term enrichments suggest that Qk5 regulates muscle-specific transcripts and that Qk6&7 may regulate gene expression indirectly, possibly through direct regulation of a transcription factor(s) and RNA binding protein(s) that has broad impacts on global gene expression.

The previous chapter used a deplete/add-back and isoform-specific RNAi approach with the Capzb exon 9 splicing reporter to demonstrate that Qk5 is a splicing regulatory RBP and Qk6 is not. To test if this could be extended transcriptome wide, we performed SpliceTrap analysis (Wu et al. 2011) of siQk5 relative to siNT, and siQk6&7 relative to siNT. We find that under loss of all Qk isoforms (siQk5) 101 cassette exons inclusion changes, but loss of Qk6&7 results in changes in inclusion of only 23 cassette exons (Fig S1A). This supports the notion that Qk5, but not Qk6&7 broadly regulates alternative splicing. Furthermore, inclusion of only 6 alternative exons changes in both siQk5 and siQk6&7 treatments in C2C12 myoblasts (Fig S1B). We also tested if the observed changes in splicing are associated with the known Qk binding motif, ACUAA. In siQk5 treated C2C12 myoblasts, we found enrichment of this sequence in the intron upstream of alternative exons repressed by Qk (events included in siQk5 relative to siNT) and enrichment of this sequence in the intron downstream of exons activated by Qk (events skipped in

siQk5 relative to siNT; Fig S1C). This is consistent with our findings in C2C12 myoblasts transfected with total Qk RNAi analyzed by splicing sensitive microarray (Hall et al. 2013). In contrast, analysis of siQk6&7 alternative splicing events in the context of ACUAA motif enrichment did not produce such a correlation (Fig S1D). Interestingly, the only significant trend observed in this dataset was a slight enrichment of the ACUAA motif in the intron upstream of alternative exons activated by Qk6&7 (more skipping in siQk6&7; Fig S1D). However this could also be due to the observation that under siQk6&7 treatment, Qk5 protein level actually increases (see previous chapter). Increasing Qk5 protein level showed robust gain of function changes in a splicing substrate in a previous study (Beuck et al. 2012). Therefore an increase in Qk5 protein could promote skipping of alternative exons by outcompeting SF1 for the branchpoint sequence (Zong et al. 2014).

Identification of direct targets of Qk proteins

To determine the global RNA binding targets of Qk in C2C12 myoblasts, we performed individual nucleotide resolution crosslinking immunoprecipitation with high throughput sequencing (iCLIP-seq (Huppertz et al. 2014). We confirmed efficient immunoprecipitation and depletion of Qk proteins from C2C12 myoblast lysate (Fig 2A), and association of RNAs with Qk (Fig 2B). We extracted RNA from the Qk immunoprecipitate and negative control HA antibody immunoprecipitate, reverse transcribed, and performed high throughput sequencing of the resulting cDNA libraries. After filtering to remove repeat sequences, unique reads were mapped to the genome, and 330,000 reads from HA library were obtained while

590,000 reads from PanQk library were obtained. The CLIPper program was used to call peaks (https://github.com/YeoLab/clipper/wiki/CLIPper-Home) from these data then after removal of any peaks from PanQk that overlapped with HA peaks, MEME-ChIP was used to identify any significantly enriched sequence motifs (Ma et al. 2014). The only motif identified was ACUAAY (Fig 2D), which is consistent with the known Qk binding motif (Ryder and Williamson 2004; Galarneau and Richard 2005; Hafner et al. 2010). Thus our Qk iCLIP library normalized to control appears to give ample coverage and CLIPper called peaks consistent with the known Qk regulatory sequence motif. To determine, transcriptome wide, the location of Qk binding to target RNAs, we analyzed the genomic position to which Qk peaks (HA overlap removed) mapped as a percentage of total. Consistent with regulation of alternative splicing observed for Qk5, 38% of peaks map to introns (Fig 2E). Additionally, consistent with regulation of mRNA abundance and translation, 33% of peaks map to 3'UTRs and 21% map to coding mRNA sequences (Fig 1E). A small percentage of peaks mapped to intergenic regions and 5'UTRs (6% and 2%, respectively; Fig 1E). Therefore, the binding profile of Qk protein to RNAs is consistent with known functions of Qk in alternative splicing, mRNA abundance, and translation.

A previous study used an *in vitro* selection approach to identify a bipartite sequence motif for Qk protein (Galarneau and Richard 2005). Our iCLIP approach allowed for investigation of evidence of the bipartite site in an *in vivo*, physiologically relevant system independent of Qk overexpression and binding site saturation. To

analyze these iCLIP data to investigate presence of a bipartite Qk binding site, we obtained reads that did not overlap with control HA reads that had ≥ 5 unique crosslinking sites (identical 5'end of read for 5 or more reads) which mapped to RefSeq genes and extended nucleotide sequence 20 nt upstream and 20 nt downstream of this crosslinking site for further analysis. This resulted in 1375 unique 41-mers for further analysis. We find significant enrichment of single Qk motifs (ACUAA, CUAA) and putative bipartite motifs (ACUAA and YAAY, CUAA and YAAY, YAAY and YAAY) in these 41-mers, but no enrichment of the single degenerate sequence YAAY in these 41-mers relative to randomly shuffled sequence of 1375 41-mers (Table 2A; $p \leq 0.05$). Further filtering of these 1375 41-mers to remove any sequence without an additional crosslink site within the 41 nucleotides resulted in 496 41-mers that have one "strong" crosslink site (≥ 5 unique 5'end reads) and one "weak" crosslink site (≥ 1 unique 5'end read) within the 41-mer. Performing the same analysis as above using this list of 496 41-mers compared to randomly shuffled sequence from the same 41-mers showed more significant enrichment of YAAY and YAAY sequence motifs, suggesting the presence of a bipartite sequence motif (Table 2B). However, using this smaller dataset, we observed a less significant p-value for ACUAA and YAAY than in the 1375 41-mers. This may suggest that Qk binds a more degenerate sequence *in vivo* than previously identified by other methods, albeit still a bipartite domain, or that the arbitrary 41-mer window should be extended.

Given the observation that some RBPs can regulate gene expression through a cooperative binding mechanism (Zhu et al. 2001; Damgaard et al. 2002; Park et al. 2006) we tested if there was any correlation between the number of Qk iCLIP peaks to magnitude of transcript abundance change under the various knockdown conditions tested. To test this we plotted the magnitude transcript abundance changes (either positive or negative) relative to number of CLIPper peak per transcript (0, 1, 2, 3, or ≥ 4) under siQk5 (Fig S2A) or siQk6&7 (Fig S2B) relative to siNT, and find no clear correlation with peak number to magnitude RNA change. This suggests that Qk proteins do not exhibit a cooperative binding regulatory affect on RNA abundance. Next we tested if Qk binding within a specific region within an mRNA was correlated with increased magnitude change in mRNA repression or activation under siQk5 or siQk6&7 treatments. The only significant correlation observed in these data was an enrichment of 5'UTR iCLIP peaks in magnitude reduction of transcripts in siQk5 treated C2C12 myoblasts (Fig S3A&B). This suggests that Qk5 binding to 5'UTRs of target transcripts promotes gene expression, and, to our knowledge, is a previously undescribed regulatory function for Qk5 protein.

Qk5 regulates the Let-7/competing endogenous RNA axes

Analysis of RNA-seq data of siQk5 treated C2C12 myoblasts by DESEQ identified differential RNA abundance of both Hmga2 (Fig 3A) and H19 (Fig 3B). Peaks from iCLIP-seq data map to the 3'UTR of Hmga2 and H19 lncRNA, suggesting Qk5 directly regulates abundance of these RNAs. Furthermore, each of these RNAs function as competing endogenous RNAs (ceRNAs) to sequester Let-7

family microRNAs, thereby down-regulating their function (Park et al. 2006; Kallen et al. 2013; Kumar et al. 2014). We validated mRNA level changes observed in RNA-seq by RT-qPCR and find that Qk5 promotes Hmga2 expression while inhibiting H19 expression (Fig 3C). Therefore, Qk5 may act as an indirect regulator that maintains Let-7/ceRNA axes by regulating miRNA bioavailability through promoting Hmga2 mRNA expression and inhibiting H19 expression (Fig 3D). Additional biochemical and molecular approaches should be undertaken to further dissect the this complex regulatory network and the role of Qk5 as an upstream regulator of it.

Differential regulation of miRNA biogenesis by Qk5 and Qk6

DESEQ analysis of mRNA levels changes under siQk5 and siQk6&7 treatment relative to control identified many regulatory target mRNAs that were distinct between the two treatments, but very few that overlapped (Fig 1B). This initially was not a result we predicted, given that Qk isoforms bind identical sequence motifs (Ryder and Williamson 2004; Galarneau and Richard 2005) and that Qk5 and Qk6 might promote opposite affects on a single transcript (Larocque et al. 2002; Nabel-Rosen et al. 2002). Given the autoregulatory network within and between Qk protein isoforms, though, interpretation of these global datasets is not as intuitive and straightforward as single siRNA/single transcript target scenario. Transfection of siQk5 results in loss of all Qk protein isoforms, and siQk6&7 treatment results in loss of Qk6&7 but an increase in Qk5. Thus there are multiple variables that are difficult to reconcile within the DESEQ mRNA abundance datasets as analyzed. In attempt to bypass this caveat, we normalized mRNA abundance changes observed in siQk6&7

treatment (loss of Qk6&7, increase in Qk5 protein) to those observed in siQk5 treatment (loss of all Qk isoforms). Surprisingly we found a much larger number of significant transcript-level changes using the same statistical cutoffs employed previously (Fig 1): siQk5 relative to siNT observed 615 transcript level changes, siQk6&7 relative to siNT observed 775, and siQk6&7 relative to siQk5 observed 1317 significant transcript level changes. Interestingly, the majority of these changes are not observed in either siQk6&7 relative to siNT or siQk5 relative to siNT (Fig 4A, overlap). This suggests that Qk5 and Qk6 may have opposite regulatory affects on global gene expression, and/or that loss of Qk5 results in gain/loss of mRNA abundance and an increase in Qk5 protein (observed in siQk6&7 treated C2C12 myoblasts) results in opposite changes (loss/gain) of the same set of mRNAs.

We tested if the magnitude of transcript level changes in siQk6&7 relative to siQk5 had any correlation with number of iCLIP peaks or location of peaks within the transcript of iCLIP peaks, but found no trend (data not shown). Qk is known to promote miR-20a biogenesis (Chen et al. 2012) and miR-7 biogenesis (Wang et al. 2013), so may regulate gene expression in C2C12 myoblasts through regulation of miRNA biogenesis. By NanoString miRNA analysis, we find that miR-20a is highly expressed in C2C12 myoblasts, but miR-7 is not (data not shown). Analysis of the DESEQ dataset from siQk6&7 relative to siQk5 identifies over 100 instances where the seed site (predicted miRNA base pairing site) for miR-20a (which is shared by miR-17 and miR-106a also) is present in transcripts changing in abundance. Based on this observation we hypothesize that Qk5 and Qk6 might oppositely regulate the

abundance of miR-20a, and/or miR-17 and miR-106a. miR-20a and miR-17 are expressed from the same primary microRNA (pri-miR) cluster (Fig 4B). Analysis of RNA-seq and iCLIP datasets suggest that Qk may directly regulate this cluster: siQk5 treatment results in reduction of RNA-seq reads, and two potential iCLIP binding events are observed in the pri-miRNA (Fig 4B). To determine if pri- or precursor-miRNA 20a (pre-miR) specifically was regulated by Qk5, we performed RT-qPCR to measure transcript level. We find that Qk5 is required for either or both pri- pre-miR-20a expression (Fig 4C), consistent with RNA-seq analysis. We also tested if mature miR-20a abundance was affected by treatment with siQk5 and siQk6&7 and observed reduction of miR-20a under siQk5 treatment, but increased miR-20a under siQk6&7 treatment (Fig 4D&E). We also analyzed mature miR-17 and miR-106a, which share the same seed sequence as miR-20a, and discovered a similar pattern: reduction of these miRNAs under siQk5 treatment, but increased abundance with siQk6&7 treatment. Therefore, the loss of all Qk isoforms (siQk5) results in reduction of miR-20a, -17, and -106a, but the loss of Qk6&7 and increase in Qk5 (siQk6&7) results in an increase in these miRNAs. These opposite affects on miRNA biogenesis provide a partial explanation for the widespread opposite affects on mRNA transcript level observed when altering Qk protein isoform ratios with RNAi. It remains to be determined if this miRNA regulatory affect is mediated entirely through loss/gain of Qk5, or if Qk6&7 are also negative regulators of miR-20a, -17, and -106a biogenesis. Future experiments should be performed to tease this out.

Conclusions

The results above describe the global consequences of both loss of, and altering the ratio of, Qk protein isoforms. We observed distinct sets of transcript level changes of both direct and indirect binding targets of Qk proteins. This suggests that Qk regulates different RNAs through diverse mechanisms. Indeed, we found that Qk directly regulated ncRNAs (H19 and Hmga2 3'UTR) that down-regulate Let-7 in C2C12 and other cell types (Fig 3). Additionally, Qk5 directly regulates pri/pre-miR-20a expression, and possibly miR-17 and miR-106a, which have the potential to themselves regulate 100's of mRNA transcripts. Further work should determine the mechanistic characteristics of these regulatory interactions and their consequences on downstream gene expression.

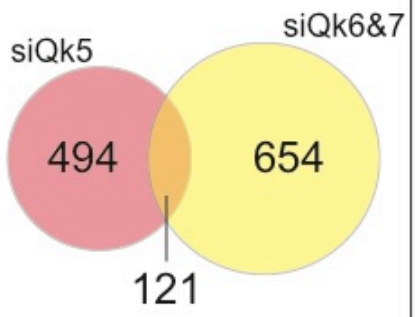
Figure 1

Fagg_Fig1

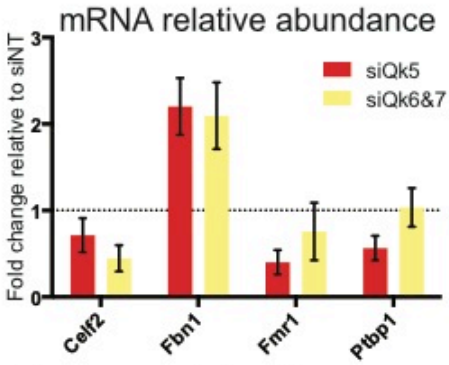
A

DESEQ	siQk5	siQk6&7
Total	615	775
Up (%)	377 (61.3)	435 (56.1)
Down (%)	238 (38.7)	340 (43.9)

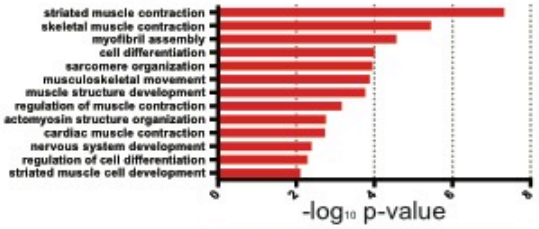
B



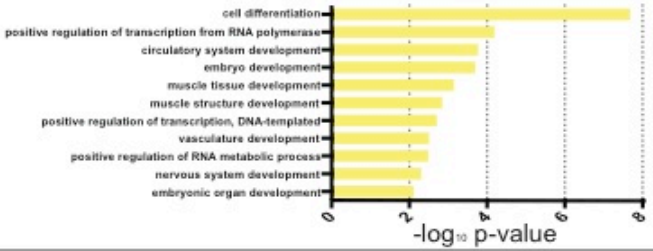
C



D Gene Ontology enriched terms



siQk5 up

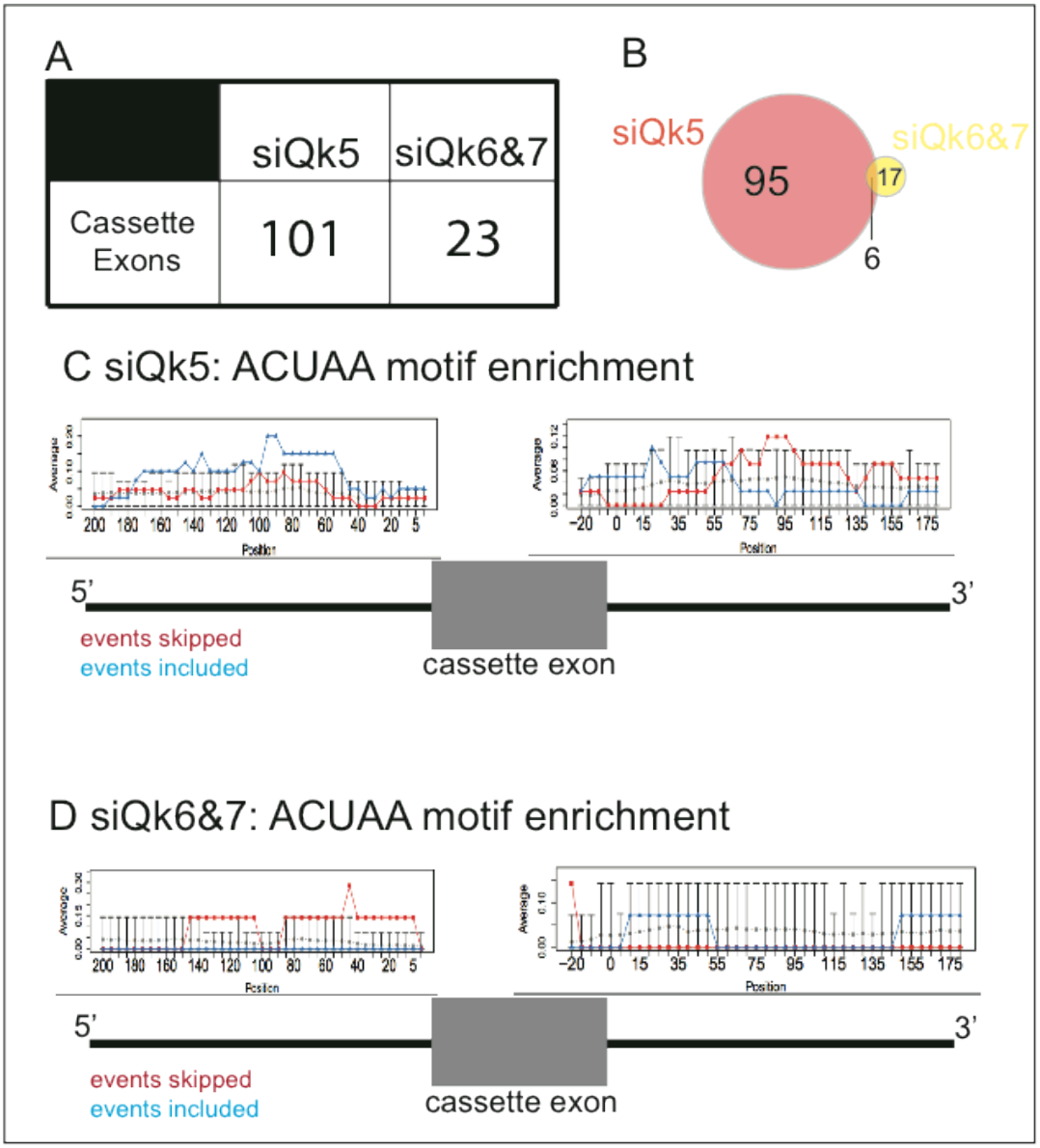


siQk6&7 up

Figure 1: Loss of Qk isoforms results in different changes in mRNA abundance. A. Summary of changes in RNA abundance identified by DESEQ analysis of RNA-seq data from C2C12 myoblasts treated with siQk5 and siQk6&7 normalized to C2C12 cells treated with siNT. Total number of transcript changes is shown and then transcripts whose abundance increases (Up, Qk repressed) or decreases (Down, Qk activated) with percentage of Up or Down shown compared to Total under each treatment. B. Overlap of transcript abundance changes determined by DESEQ in siQk5 relative to siNT and siQk6&7 relative to siNT. C. RT-qPCR validation of selected DESEQ-identified changes in mRNA abundance under siQk5 and siQk6&7 relative to siNT: Celf2, Fbn1, Fmr1, and Ptbp1. D. Enrichment of Gene Ontology (GO) terms within transcripts increasing in abundance under siQk5 transfection relative to siNT (top) and siQk6&7 transfection relative to siNT (bottom); y-axis shows significantly enriched GO terms, x-axis $-\log_{10}$ p-value.

Supplemental Figure 1

Fagg_FigS1



Supplementary Figure 1: Qk5 globally regulates alternative splicing in C2C12 myoblasts. A. Summary of results from SpliceTrap analysis of RNA-seq data that identified significant (cutoff: inclusion ratio change $|\Delta IR| > 0.2$ with at least 20 junction reads representing both skip and include mRNA forms in 4 pair-wise comparisons of each RNA-seq libraries) changes in inclusion of cassette exons from C2C12 myoblasts treated with siQk5 normalized to siNT and siQk6&7 normalized to siNT. B. Overlap of cassette exons whose level of inclusion significantly changed in siQk5 treatment relative to siNT or siQk6&7 treatment relative to siNT. C. Analysis of ACUAA motif enrichment under siQk5 treatment relative to siNT; model cassette exon shown in gray and boxed x-axis represent relative distance from cassette exon with frequency of ACUAA motif observed on y-axis; plotted gray dots represent background exons (expressed but inclusion level unchanged in RNAi), red dots represent observed skipping of cassette exons, and blue dots represent inclusion of cassette exons. D. Analysis of ACUAA motif enrichment under siQk6&7 treatment relative to siNT; frequency of background, included, and skipped exons shown as in C.

Figure 2

Fagg_Fig2

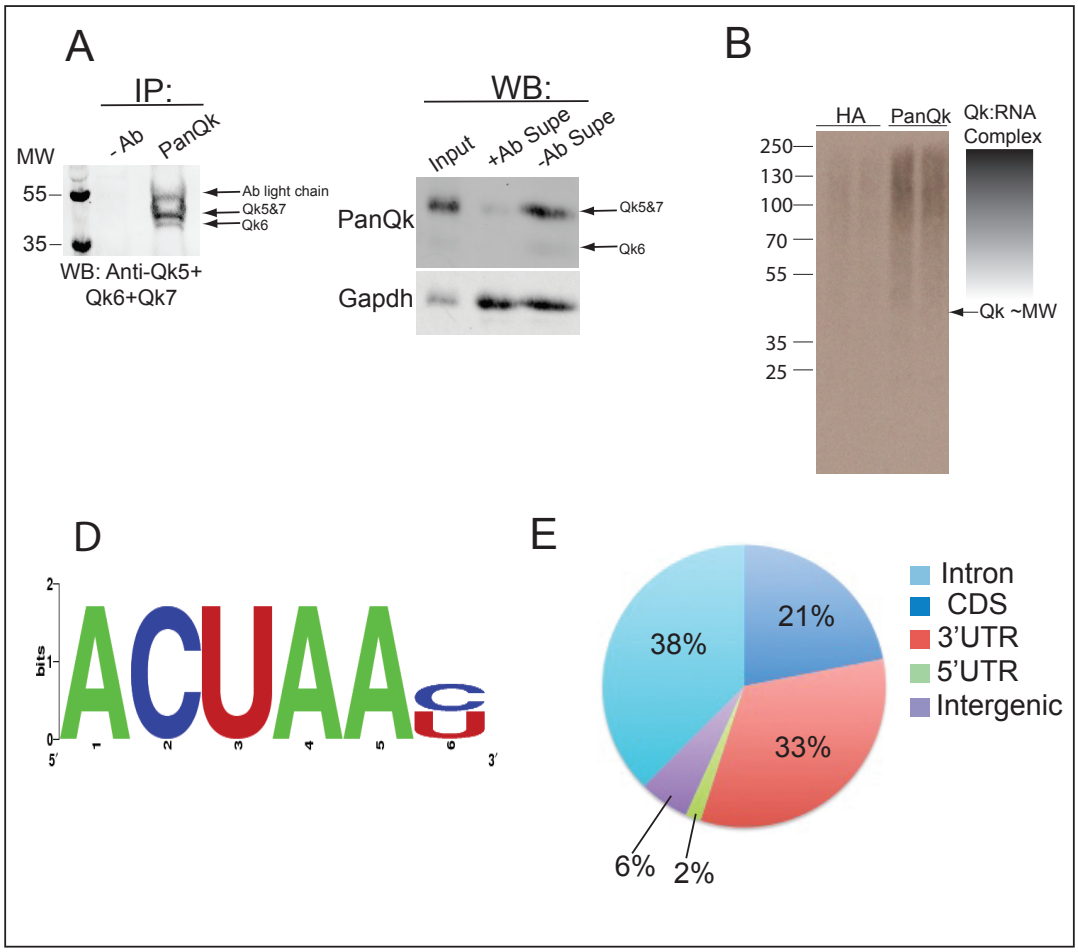


Figure 2: PanQk iCLIP identifies global RNA targets of Qk. A. Western blot analysis of iCLIP experiments performed in C2C12 myoblasts; left is western blot of immunoprecipitates from beads only (- Ab) and PanQk iCLIP probed with isoform-specific antibodies to Qk5, Qk6, and Qk7; right is western blot of input lysate (left) supernatant from PanQk iCLIP (middle), and supernatant from beads only (- Ab supe, right) IP probed with PanQk and Gapdh antibodies. B. Autoradiograph of anti-HA (left two lanes) and anti-PanQk iCLIP from C2C12 myoblasts; MW is shown at left, arrow at right shows approximate migration of Qk proteins (42 kDa for Qk5&7, 38 kDa for Qk6) and box above show region excised from nitrocellulose membrane for RNA extraction and library prep. D. CLIPper called peaks (after removal with any peaks that overlapped with peaks called from anti-HA iCLIP library) were used as input sequences into the MEME-ChIP program and motif finding was performed using the weight matrix-based format. E. The percentage of normalized CLIPper peaks (any overlap with HA peaks removed) from PanQk iCLIP library to annotated genomic regions is shown.

Table 1

Fagg_Table1

Table 1a

motif	n observed	n shuffled	p-value
ACUAA only	173	78	< 0.0001
ACUAA and YAAY	69	41	0.0112
CUAA only	361	237	< 0.0001
CUAA and YAAY	199	118	< 0.0001
YAAY only	689	708	0.6934
YAAY and YAAY	268	212	0.02

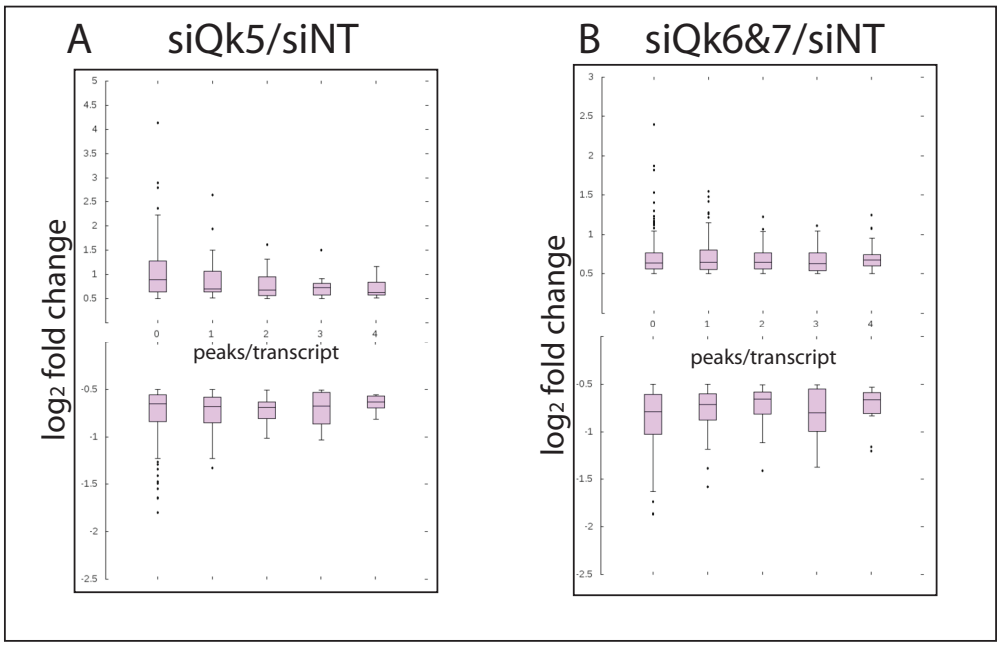
Table 1b

motif	n observed	n shuffled	p-value
ACUAA only	81	29	< 0.0001
ACUAA and YAAY	30	18	0.1043
CUAA only	184	80	< 0.0001
CUAA and YAAY	124	42	< 0.0001
YAAY only	310	272	0.2296
YAAY and YAAY	148	84	0.0002

Tables 1a&b: Significant enrichment of bipartite sequence elements from iCLIP crosslinked regions. 1a Five or more identical crosslink sites (5'end of iCLIP read) that mapped to UCSC Genes and/or RefSeq Genes were used with 20 nucleotide windows extended both upstream and downstream from crosslink site to generate 41-mers for analysis which resulted in 1375 unique sequences; these 41-mers (n observed) or these 41-mer sequences randomly shuffled (n shuffled) were used to determine the number of each motif listed in left column, with two tailed p-value determined by Fisher's exact test using 2 x 2 contingency table for each. 1b Of the 1375 41-mers analyzed in a, 496 had an additional crosslink site (5'end of iCLIP read) that mapped within the 41-mer; these 496 were analyzed as above with number of motifs observed in actual 41-mers (n observed) and number observed in shuffled nucleotide sequence (n shuffled) reported and p-values shown, calculated as above.

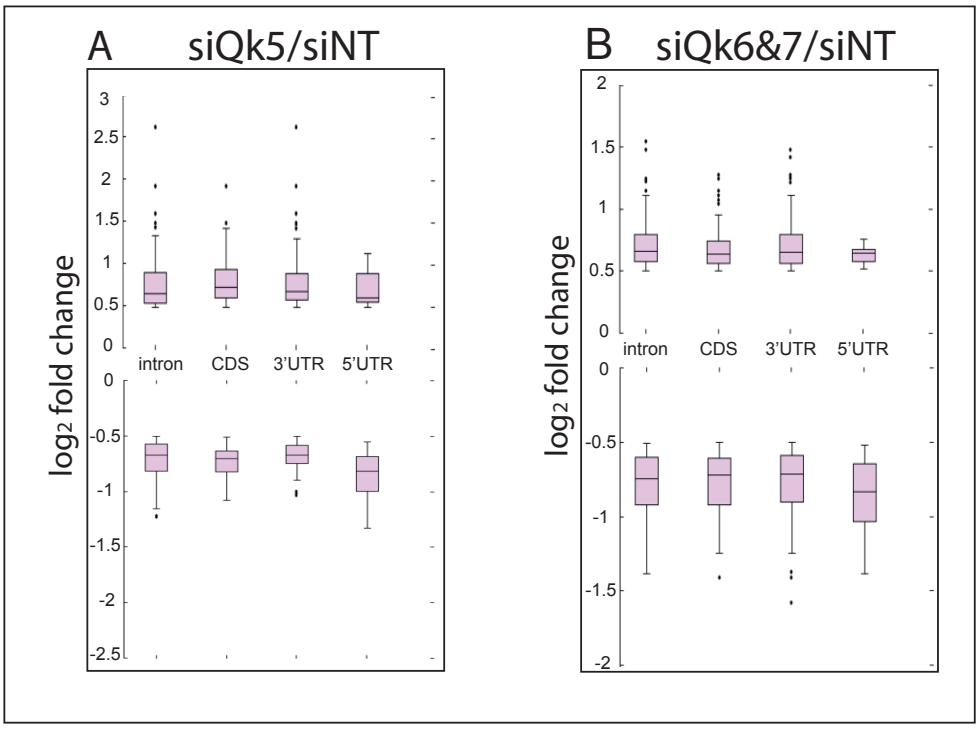
Supplementary Figure 2

Fagg_FigS2



Supplementary Figure 2: Correlation of number of CLIPper peaks per transcript with magnitude change in mRNA abundance. A. Log_2 fold change (lfc) of transcripts significantly changing in DESEQ analysis of siQk5 relative to siNT plotted on y-axis with number of CLIPper peaks per transcript plotted on x-axis (value of 4 peaks equals $\geq$ 4 peaks per transcript). B. Log_2 fold change of transcripts significantly changing in DESEQ analysis of siQk6&7 relative to siNT plotted as in A with number of CLIPper peaks per transcript as in A also.

Supplementary Figure 3 Fagg_FigS3

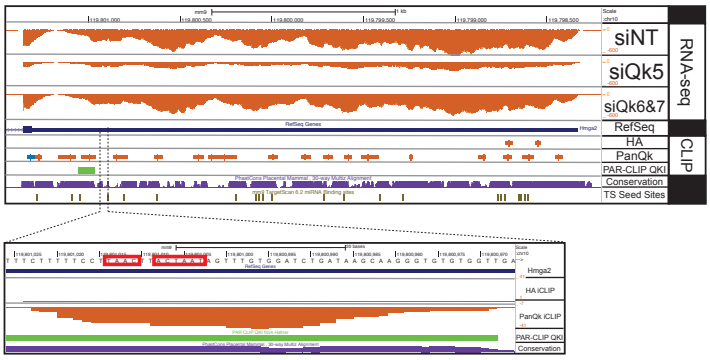


Supplementary Figure 3: Correlation of region of CLIPper peak within a transcript with magnitude change in mRNA abundance. A. Log_2 fold change (lfc) of transcripts significantly changing in DESEQ analysis of siQk5 relative to siNT plotted on y-axis with region of transcript to which CLIPper peak maps within transcript on x-axis. B. Log_2 fold change of transcripts significantly changing in DESEQ analysis of siQk6&7 relative to siNT plotted as in A with region of transcript to which CLIPper peak maps within transcript on x-axis.

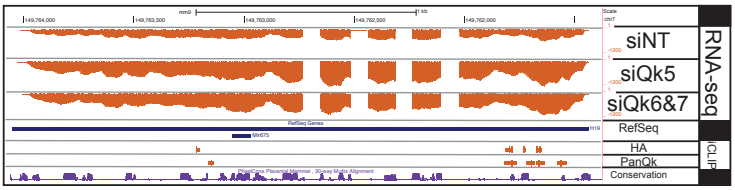
Figure 3

Fagg_Fig3

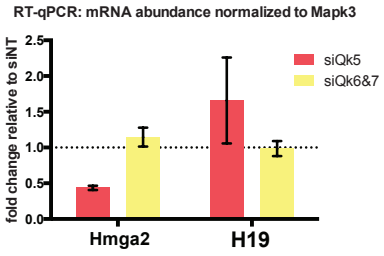
A Hmga2 3'UTR



B H19 lncRNA



C



D

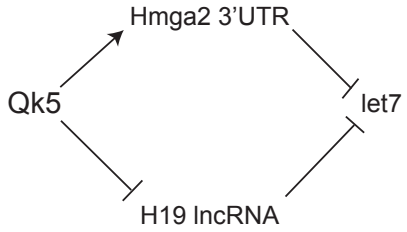


Figure 3: Hmga2 and H19 are directly regulated by Qk5. A. UCSC Genome Browser screen shot showing RNA-seq coverage tracks of siNT, siQk5, siQk6&7, RefSeq Genes, iCLIP peaks for HA in C2C12 myoblasts, iCLIP peaks for PanQk in C2C12 myoblasts, PAR-CLIP peak for QKI in HEK293T cells, PhastCons mammalian conservation track, and mm9 TargetScan 6.2 miRNA Seed Sites within and around the 3'UTR of Hmga2. Inset shows zoomed in region of Hmga2 3'UTR with red boxes around putative Qk bipartite binding motif and HA iCLIP coverage track, PanQk iCLIP coverage track, PAR-CLIP QKI peak, and PhastCons mammalian conservation track. B. UCSC Genome Browser screen shot showing H19 lncRNA with tracks for RNA-seq coverage for siNT, siQk5, siQk6&7, RefSeq Genes, iCLIP peaks for HA control, iCLIP peaks for PanQk, and PhastCons mammalian conservation. C. RT-qPCR for Hmga2 and H19 of RNA extracted from C2C12 myoblasts transfected with siQk5 and siQk6&7 normalized to Mapk3 and shown as fold change relative to siNT. D. Model for Qk5 regulation of Hmga2 and H19 to maintain ceRNA/Let-7 axis in C2C12 cells.

Figure 4

Fagg_Fig4

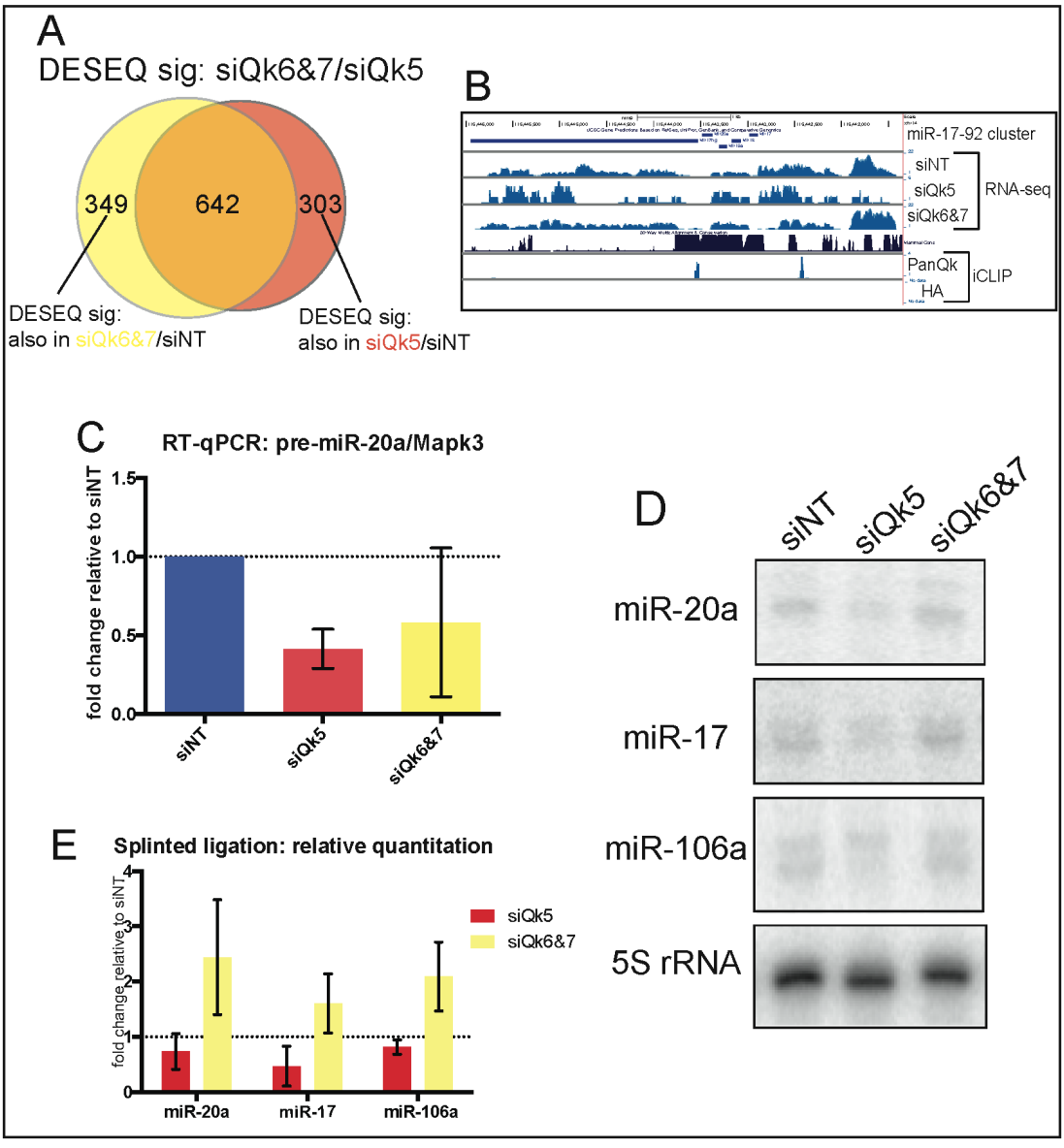


Figure 4: Loss of Qk5 or Qk6&7 results in opposite changes in RNA abundance. A. Venn diagram showing overlap between transcripts significantly changing in abundance by DESEQ when normalizing siQk6&7 to siQk5; overlap shows significantly changing transcripts that are not significantly changing in abundance under siQk6&7 relative to siNT and siQk5 relative to siNT; left side non-overlapping circle (yellow) denotes RNA abundance changes also observed in siQk6&7 normalized to siNT and right side non-overlapping circle (red) denotes RNA abundance changes also observed in siQk5 normalized to siNT.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* **11**: R106.

Beuck C, Qu S, Fagg WS, Ares M, Jr., Williamson JR. 2012. Structural analysis of the Quaking homodimerization interface. *J Mol Biol*.

Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N et al. 2012. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes & development* **26**: 1459-1472.

Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA (New York, NY)* **8**: 1401-1415.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature structural & molecular biology* **12**: 691-698.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP, Ares M, Jr. 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA (New York, NY)* **19**: 627-638.

Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, Tajnik M, Konig J, Ule J. 2014. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods (San Diego, Calif)* **65**: 274-287.

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H et al. 2013. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Molecular cell* **52**: 101-112.

Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, Winslow MM, Downward J. 2014. HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. *Nature* **505**: 212-217.

Larocque D, Pilotte J, Chen T, Cloutier F, Massie B, Pedraza L, Couture R, Lasko P, Almazan G, Richard S. 2002. Nuclear Retention of MBP mRNAs in the Quaking Viable Mice. *Neuron* **36**: 815-829.

Ma W, Noble WS, Bailey TL. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nature protocols* **9**: 1428-1450.

Nabel-Rosen H, Volohonsky G, Reuveny A, Zaidel-Bar R, Volk T. 2002. Two Isoforms of the Drosophila RNA Binding Protein, How, Act in Opposing Directions to Regulate Tendon Cell Differentiation. *Developmental Cell* **2**: 183-193.

Park E, Han J, Son GH, Lee MS, Chung S, Park SH, Park K, Lee KH, Choi S, Seong JY et al. 2006. Cooperative actions of Tra2alpha with 9G8 and SRp30c in the RNA splicing of the gonadotropin-releasing hormone gene transcript. *The Journal of biological chemistry* **281**: 401-409.

Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA (New York, NY)* **10**: 1449-1458.

Wang Y, Vogel G, Yu Z, Richard S. 2013. The QKI-5 and QKI-6 RNA binding proteins regulate the expression of microRNA 7 in glial cells. *Molecular and cellular biology* **33**: 1233-1243.

Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. 2011. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics (Oxford, England)* **27**: 3010-3016.

Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Molecular cell* **8**: 1351-1361.

Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H et al. 2014. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS genetics* **10**: e1004289.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

Here I report findings from diverse studies of Quaking RBPs. These studies ranged from mechanistic biochemical structure/function analysis, to analysis of the molecular processes regulated by Qk in certain stem cell types, to functional analysis of specific Qk isoforms and how they autoregulate and perform cross isoform regulation. The unpublished findings reported in Chapter 5 are of particular significance; as they highlight a previously unknown autoregulatory network within *Qk*. Furthermore this study and the data presented in Chapter 6 demonstrate Qk5 and Qk6 have distinct molecular regulatory functions. Thus it will be of the utmost importance for researchers studying Qk in the future not to treat these proteins generically as "Qk", but to analyze each of the 3 (or more) Qk isoforms as individual and distinct RBPs. This may be particularly challenging to put into practice with translational and clinical researchers, but may be most important with this group, as therapeutic developments targeting *Qk* gene products depend on thorough foundational work.

There are many potential future directions of study analyzing Qk and in various contexts. Thus a challenge will be to focus on the most important questions. It is my opinion that some of the most important of these are: 1) How does Qk5 (at a mechanistic level) regulate alternative splicing? 2) What are the consequences on very early developmental cell fate decisions under Qk loss of function? 3) In cell/tissue types where more Qk6 is expressed than Qk5, how is gene expression regulated globally by Qk? 4) Are there any particular diseases characterized by Qk isoform ratio changes compared to the healthy tissue? It is my hope to continue

working in this area of focus throughout future steps in my career. Furthermore, I hope this current body of work can contribute to future discovery in this and other fields.

Bibliography

Adereth Y, Dammai V, Kose N, Li R, Hsu T. 2005. RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. *Nature cell biology* 7: 1240-1247.

Akerman M, Mandel-Gutfreund Y. 2006. Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res* 34: 23-31.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* 11: R106.

Baehrecke EH. 1997. who encodes a KH RNA binding protein that functions in muscle development. *Development* 124: 1323-1332.

Benda P, Someda K, Messer J, Sweet WH. 1971. Morphological and immunochemical studies of rat glial tumors and clonal strains propagated in culture. *Journal of neurosurgery* 34: 310-323.

Beuck C, Qu S, Fagg WS, Ares M, Jr., Williamson JR. 2012. Structural analysis of the Quaking homodimerization interface. *J Mol Biol*.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150: 279-290.

Black DL. 2003. MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING. *Annual Review of Biochemistry* 72: 291-336.

Bohnsack BL, Lai L, Northrop JL, Justice MJ, Hirschi KK. 2006. Visceral endoderm function is regulated by quaking and required for vascular development. *Genesis (New York, NY : 2000)* 44: 93-104.

Boutet SC, Cheung TH, Quach NL, Liu L, Prescott SL, Edalati A, Iori K, Rando TA. 2012. Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell stem cell* 10: 327-336.

Bradley RK, Merkin J, Lambert NJ, Burge CB. 2012. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS biology* 10: e1001229.

Caceres JF, Screaton GR, Krainer AR. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development* 12: 55-66.

Cazalla D, Zhu J, Manche L, Huber E, Krainer AR, Caceres JF. 2002. Nuclear export and retention signals in the RS domain of SR proteins. *Molecular and cellular biology* 22: 6871-6882.

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G et al. 2012. Muscleblind-like 2-Mediated Alternative Splicing in the Developing Brain and Dysregulation in Myotonic Dystrophy. *Neuron* 75: 437-450.

Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N et al. 2012. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes & development* 26: 1459-1472.

Chen T RS. 1998. Structure-function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization. *Molecular and cellular biology* 18: 4863-4871.

Cox RD, Hugill A, Shedlovsky A, Noveroske JK, Best S, Justice MJ, Lehrach H, Dove WF. 1999. Contrasting effects of ENU induced embryonic lethal mutations of the quaking gene. *Genomics* 57: 333-341.

Crick F. 1970. Central dogma of molecular biology. *Nature* 227: 561-563.

Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA (New York, NY)* 8: 1401-1415.

Damianov A, Black DL. 2010. Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA (New York, NY)* 16: 405-416.

Dominski Z, Kole R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* 11: 6075-6083.

Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA et al. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature structural & molecular biology* 17: 187-193.

Ebersole TA, Chen Q, Justice MJ, Artzt K. 1996. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genetics* 12: 260-265.

Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. 2002. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Human molecular genetics* 11: 805-814.

Francis R, Barton MK, Kimble J, Schedl T. 1995. gld-1, a tumor suppressor gene required for oocyte development in Caenorhabditis elegans. *Genetics* 139: 579-606.

Fu XD, Ares M, Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature reviews Genetics* 15: 689-701.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature structural & molecular biology* 12: 691-698.

Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, Ares M, Jr., Otis TS, Black DL. 2012. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes & development* 26: 445-460.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M, Jr., Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* 43: 706-711.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141: 129-141.

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP, Ares M, Jr. 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA (New York, NY)* 19: 627-638.

Hardy RJ, Loushin CL, Friedrich Jr VL, Chen Q, Ebersole TA, Lazzarini RA, Artzt K. 1996. Neural Cell Type-Specific Expression of QKI Proteins Is Altered in quakingviable Mutant Mice. *J Neurosci* 16: 7941-7949.

Hart MC, Cooper JA. 1999. Vertebrate isoforms of actin capping protein beta have distinct functions In vivo. *The Journal of cell biology* 147: 1287-1298.

He F, Dang W, Abe C, Tsuda K, Inoue M, Watanabe S, Kobayashi N, Kigawa T, Matsuda T, Yabuki T et al. 2009. Solution structure of the RNA binding domain in the human muscleblind-like protein 2. *Protein science : a publication of the Protein Society* 18: 80-91.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. 2004. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* 36: 1255-1257.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. 2006. Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *American journal of human genetics* 78: 291-302.

Hoffman Y, Bublik DR, A PU, Elkon R, Biniashvili T, Agami R, Oren M, Pilpel Y. 2016. 3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-

differentiation Genes in Proliferating Human Cells. *PLoS Genet* 12: e1005879.

Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, Tajnik M, Konig J, Ule J. 2014. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods (San Diego, Calif)* 65: 274-287.

Ishidate T, Yoshihara S, Kawasaki Y, Roy BC, Toyoshima K, Akiyama T. 1997. Identification of a novel nuclear localization signal in Sam68. *FEBS letters* 409: 237-241.

Ji S, Ye G, Zhang J, Wang L, Wang T, Wang Z, Zhang T, Wang G, Guo Z, Luo Y et al. 2012. miR-574-5p negatively regulates Qki6/7 to impact beta-catenin/Wnt signalling and the development of colorectal cancer. *Gut*.

Ji S, Ye G, Zhang J, Wang L, Wang T, Wang Z, Zhang T, Wang G, Guo Z, Luo Y, Cai J, Yang JY. 2013. miR-574-5p negatively regulates Qki6/7 to impact beta-catenin/Wnt signalling and the development of colorectal cancer. *Gut* 62: 716-726.

Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16: 167-177.

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H et al. 2013. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Molecular cell* 52: 101-112.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews Genetics* 8: 533-543.

Koizumi J, Okamoto Y, Onogi H, Mayeda A, Krainer AR, Hagiwara M. 1999. The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). *The Journal of biological chemistry* 274: 11125-11131.

Kondo T, Furuta T, Mitsunaga K, Ebersole TA, Shichiri M, Wu J, Artzt K, Yamamura K, Abe K. 1999. Genomic organization and expression analysis of the mouse qkI locus. *Mammalian genome : official journal of the International Mammalian Genome Society* 10: 662-669.

Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, Winslow MM, Downward J. 2014. HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. *Nature* 505: 212-217.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446: 926-929.

Larocque D, Fragoso G, Huang J, Mushynski WE, Loignon M, Richard Sp, Almazan G. 2009. The QKI-6 and QKI-7 RNA Binding Proteins Block Proliferation and Promote Schwann Cell Myelination. *PloS one* 4: e5867.

Larocque D, Galarneau A, Liu H-N, Scott M, Almazan G, Richard S. 2005. Protection of p27Kip1 mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat Neurosci* 8: 27-33.

Larocque D, Pilotte J, Chen T, Cloutier F, Massie B, Pedraza L, Couture R, Lasko P, Almazan G, Richard S. 2002. Nuclear Retention of MBP mRNAs in the Quaking Viable Mice. *Neuron* 36: 815-829.

Li Z, Takakura N, Oike Y, Imanaka T, Araki K, Suda T, Kaname T, Kondo T, Abe K, Yamamura K. 2003a. Defective smooth muscle development in qkI-deficient mice. *Development, growth & differentiation* 45: 449-462.

Li Z, Takakura N, Oike Y, Imanaka T, Araki K, Suda T, Kaname T, Kondo T, Abe K, Yamamura K-i. 2003b. Defective smooth muscle development in qkI-deficient mice. *Development Growth & Differentiation* 45: 449-462.

Li Z, Zhang Y, Li D, Feng Y. 2000. Destabilization and Mislocalization of Myelin Basic Protein mRNAs in quaking Dysmyelination Lacking the QKI RNA-Binding Proteins. *J Neurosci* 20: 4944-4953.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456: 464-469.

Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, de la Grange P, Ast G et al. 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nature structural & molecular biology* 17: 1114-1123.

Lobbardi R, Lambert G, Zhao J, Geisler R, Kim HR, Rosa FM. 2011. Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Development* 138: 1783-1794.

Lu W, Feng F, Xu J, Lu X, Wang S, Wang L, Lu H, Wei M, Yang G, Wang L et al. 2014. QKI impairs self-renewal and tumorigenicity of oral cancer cells via repression of SOX2. *Cancer biology & therapy* 15: 1174-1184.

Lu Z, Zhang Y, Ku L, Wang H, Ahmadian A, Feng Y. 2003. The quakingviable mutation affects qkI mRNA expression specifically in myelin-producing cells of the nervous system. *Nucleic acids research* 31: 4616-4624.

Ma W, Noble WS, Bailey TL. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nature protocols* 9: 1428-1450.

Makeyev EV, Zhang J, Carrasco MA, Maniatis T. 2007. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Molecular cell* 27: 435-448.

Mandler MD, Ku L, Feng Y. 2014. A cytoplasmic quaking I isoform regulates the hnRNP F/H-dependent alternative splicing pathway in myelinating glia. *Nucleic acids research* 42: 7319-7329.

Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. 2001. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Human molecular genetics* 10: 2165-2170.

Matlin AJ, Clark F, Smith CWJ. 2005. Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology* 6: 386-398.

Mattaj IW, De Robertis EM. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell* 40: 111-118.

Moreland RB, Nam HG, Hereford LM, Fried HM. 1985. Identification of a nuclear localization signal of a yeast ribosomal protein. *Proceedings of the National Academy of Sciences of the United States of America* 82: 6561-6565.

Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM. 2016. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* 30: 553-566.

Muller-McNicoll M, Neugebauer KM. 2013. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet* 14: 275-287.

Munding EM, Shiue L, Katzman S, Donohue JP, Ares M, Jr. 2013. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Molecular cell* 51: 338-348.

Nabel-Rosen H, Volohonsky G, Reuveny A, Zaidel-Bar R, Volk T. 2002. Two Isoforms of the Drosophila RNA Binding Protein, How, Act in Opposing Directions to Regulate Tendon Cell Differentiation. *Developmental Cell* 2: 183-193.

Naftelberg S, Schor IE, Ast G, Kornblihtt AR. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* 84: 165-198.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O,ÄôBrien G, Shiue L, Clark TA, Blume JE, Ares M. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & development* 21: 708-718.

Novikov L, Park JW, Chen H, Klerman H, Jalloh AS, Gamble MJ. 2011. QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Molecular and cellular biology* 31: 4244-4255.

Ohno G, Hagiwara M, Kuroyanagi H. 2008. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes & development* 22: 360-374.

Park E, Han J, Son GH, Lee MS, Chung S, Park SH, Park K, Lee KH, Choi S, Seong JY et al. 2006. Cooperative actions of Tra2alpha with 9G8 and SRp30c in the RNA splicing of the gonadotropin-releasing hormone gene transcript. *The Journal of biological chemistry* 281: 401-409.

Pilotte J, Larocque D, Richard S. 2001. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes & development* 15: 845-858.

Query CC, Bentley RC, Keene JD. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* 57: 89-101.

Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 17: 47-62.

Ragle JM, Katzman S, Akers TF, Barberan-Soler S, Zahler AM. 2015. Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in C. elegans. *Genome Res* 25: 982-994.

Richard S. 2010. Reaching for the stars: Linking RNA binding proteins to diseases. *Advances in experimental medicine and biology* 693: 142-157.

Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA (New York, NY)* 10: 1449-1458.

Saccomanno L LC, Jan E, Punkay E, Artzt K, and Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences of the United States of America* 96: 12605-12610.

Saccomanno L, Loushin C, Jan E, Punkay E, Artzt K, Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences of the United States of America* 96: 12605-12610.

Samson ML. 1998. Evidence for 3' untranslated region-dependent autoregulation of the Drosophila gene encoding the neuronal nuclear RNA-binding protein ELAV. *Genetics* 150: 723-733.

Sanford JR, Gray NK, Beckmann K, Caceres JF. 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes & development* 18: 755-768.

Schafer DA, Korshunova YO, Schroer TA, Cooper JA. 1994. Differential localization and sequence analysis of capping protein beta-subunit isoforms of vertebrates. *The Journal of cell biology* 127: 453-465.

Schindler S, Szafranski K, Hiller M, Ali GS, Palusa SG, Backofen R, Platzer M, Reddy AS. 2008. Alternative splicing at NAGNAG acceptors in Arabidopsis thaliana SR and SR-related protein-coding genes. *BMC Genomics* 9: 159.

Sen GL, Blau HM. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature cell biology* 7: 633-636.

Shedlovsky A, King TR, Dove WF. 1988. Saturation germ line mutagenesis of the murine t region including a lethal allele at the quaking locus. *Proceedings of the National Academy of Sciences of the United States of America* 85: 180-184.

Sidman R, Dickie M, Appel S. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. *Science (New York, NY)* 144: 309-311.

Singh G, Pratt G, Yeo GW, Moore MJ. 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* 84: 325-354.

Singh RK, Xia Z, Bland CS, Kalsotra A, Scavuzzo MA, Curk T, Ule J, Li W, Cooper TA. 2014. Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. *Molecular cell* 55: 592-603.

Sloan KE, Gleizes PE, Bohnsack MT. 2015. Nucleocytoplasmic Transport of RNAs and RNA-Protein Complexes. *J Mol Biol*.

Spellman R, Llorian M, Smith CWJ. 2007. Crossregulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Molecular cell* 27: 420-434.

Suenaga K, Lee KY, Nakamori M, Tatsumi Y, Takahashi MP, Fujimura H, Jinnai K, Yoshikawa H, Du H, Ares M, Jr. et al. 2012. Muscleblind-like 1 knockout mice reveal novel splicing defects in the myotonic dystrophy brain. *PloS one* 7: e33218.

Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. 2010. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nature structural & molecular biology* 17: 306-312.

Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS et al. 2016. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol Cell* 61: 821-833.

Teplova M, Patel DJ. 2008. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nature structural & molecular biology* 15: 1343-1351.

Towler BP, Jones CI, Newbury SF. 2015. Mechanisms of regulation of mature miRNAs. *Biochem Soc Trans* 43: 1208-1214.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* 444: 580-586.

van der Veer EP, de Bruin RG, Kraaijeveld AO, de Vries MR, Bot I, Pera T, Segers FM, Trompet S, van Gils JM, Roeten MK et al. 2013. Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circulation research* 113: 1065-1075.

Vasudevan S, Steitz JA. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 128: 1105-1118.

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

Virant-Klun I, Stahlberg A, Kubista M, Skutella T. 2016. MicroRNAs: From Female Fertility, Germ Cells, and Stem Cells to Cancer in Humans. *Stem cells international* 2016: 3984937.

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S et al. 2012. Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* 150: 710-724.

Wang Y, Vogel G, Yu Z, Richard S. 2013. The QKI-5 and QKI-6 RNA binding proteins regulate the expression of microRNA 7 in glial cells. *Molecular and cellular biology* 33: 1233-1243.

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular cell* 13: 91-100.

Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. 2011. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics (Oxford, England)* 27: 3010-3016.

Wu J, Zhou L, Tonissen K, Tee R, Artzt K. 1999. The Quaking I-5 Protein (QKI-5) Has a Novel Nuclear Localization Signal and Shuttles between the Nucleus and the Cytoplasm. *J Biol Chem* 274: 29202-29210.

Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Molecular cell* 36: 996-1006.

Yamashita A, Maeda K, Maeda Y. 2003. Crystal structure of CapZ: structural basis for actin filament barbed end capping. *The EMBO journal* 22: 1529-1538.

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature structural & molecular biology* 16: 130-137.

Zaffran S, Astier M, Gratecos D, Semeriva M. 1997. The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Development* 124: 2087-2098.

Zearfoss NR, Clingman CC, Farley BM, McCoig LM, Ryder SP. 2011. Quaking regulates Hnrnpa1 expression through its 3' UTR in oligodendrocyte precursor cells. *PLoS genetics* 7: e1001269.

Zhao L, Ku L, Chen Y, Xia M, LoPresti P, Feng Y. 2006a. QKI Binds MAP1B mRNA and Enhances MAP1B Expression during Oligodendrocyte Development. *Mol Biol Cell* 17: 4179-4186.

Zhao L, Mandler MD, Yi H, Feng Y. 2010. Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 107: 19061-19066.

Zhao L, Tian D, Xia M, Macklin WB, Feng Y. 2006b. Rescuing qkv Dysmyelination by a Single Isoform of the Selective RNA-Binding Protein QKI. *J Neurosci* 26: 11278-11286.

Zhao Y, Zhang G, Wei M, Lu X, Fu H, Feng F, Wang S, Lu W, Wu N, Lu Z et al. 2014. The tumor suppressing effects of QKI-5 in prostate cancer: a novel diagnostic and prognostic protein. *Cancer biology & therapy* 15: 108-118.

Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Molecular cell* 8: 1351-1361.

Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H et al. 2014. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS genetics* 10: e1004289.