Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP

Markus Hafner,[1,5] Markus Landthaler,[1,4,5] Lukas Burger,[2] Mohsen Khorshid,[2] Jean Hausser,[2] Philipp Berninger,[2] Andrea Rothballer,[1] Manuel Ascano, Jr.,[1] Anna-Carina Jungkamp,[1,4] Mathias Munschauer,[1] Alexander Ulrich,[1] Greg S. Wardle,[1] Scott Dewell,[3] Mihaela Zavolan,[2,*] and Thomas Tuschl[1,*]

[1]Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, NY 10065, USA
[2]Biozentrum der Universität Basel and Swiss Institute of Bioinformatics (SIB), Klingelbergstr. 50-70, CH-4056 Basel, Switzerland
[3]Genomics Resource Center, The Rockefeller University, 1230 York Avenue, Box 241, New York, NY 10065, USA
[4]Present address: Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany
[5]These authors contributed equally to this work
*Correspondence: mihaela.zavolan@unibas.ch (M.Z.), ttuschl@rockefeller.edu (T.T.)
DOI 10.1016/j.cell.2010.03.009

SUMMARY

RNA transcripts are subject to posttranscriptional gene regulation involving hundreds of RNA-binding proteins (RBPs) and microRNA-containing ribonucleoprotein complexes (miRNPs) expressed in a cell-type dependent fashion. We developed a cell-based crosslinking approach to determine at high resolution and transcriptome-wide the binding sites of cellular RBPs and miRNPs. The crosslinked sites are revealed by thymidine to cytidine transitions in the cDNAs prepared from immunopurified RNPs of 4-thiouridine-treated cells. We determined the binding sites and regulatory consequences for several intensely studied RBPs and miRNPs, including PUM2, QKI, IGF2BP1-3, AGO/EIF2C1-4 and TNRC6A-C. Our study revealed that these factors bind thousands of sites containing defined sequence motifs and have distinct preferences for exonic versus intronic or coding versus untranslated transcript regions. The precise mapping of binding sites across the transcriptome will be critical to the interpretation of the rapidly emerging data on genetic variation between individuals and how these variations contribute to complex genetic diseases.

INTRODUCTION

Gene expression in eukaryotes is extensively controlled at the posttranscriptional level by RNA-binding proteins (RBPs) and ribonucleoprotein complexes (RNPs) modulating the maturation, stability, transport, editing and translation of RNA transcripts (Martin and Ephrussi, 2009; Moore and Proudfoot, 2009; Sonenberg and Hinnebusch, 2009). Vertebrate genomes encode several hundred RBPs (McKee et al., 2005), each containing one or more domains able to specifically recognize target tran-

scripts. Furthermore, hundreds of microRNAs (miRNAs) bound by Argonaute (AGO/EIF2C) proteins mediate destabilization and/or inhibition of translation of partially complementary target mRNAs (Bartel, 2009). To understand how the interplay of these RNA-binding factors affects the regulation of individual transcripts, high resolution maps of in vivo protein-RNA interactions are necessary (Keene, 2007).

A combination of genetic, biochemical and computational approaches are typically applied to identify RNA-RBP or RNA-RNP interactions. Microarray profiling of RNAs associated with immunopurified RBPs (RIP-Chip) (Tenenbaum et al., 2000) defines targets at a transcriptome level, but its application is limited to the characterization of kinetically stable interactions and does not directly identify the RBP recognition element (RRE) within the long target RNA. Nevertheless, RREs with higher information content can be derived computationally from RIP-Chip data, e.g., for HuR (Lopez de Silanes et al., 2004) or for Pumilio (Gerber et al., 2006).

More direct RBP target site information is obtained by combining in vivo UV crosslinking (Greenberg, 1979; Wagenmakers et al., 1980) with immunoprecipitation (Dreyfuss et al., 1984; Mayrand et al., 1981) followed by the isolation of crosslinked RNA segments and cDNA sequencing (CLIP) (Ule et al., 2003). CLIP was used to identify targets of the splicing regulators NOVA1 (Licatalosi et al., 2008), FOX2 (Yeo et al., 2009) and SFRS1 (Sanford et al., 2009) as well as U3 snoRNA and pre-rRNA (Granneman et al., 2009), pri-miRNA targets for HNRNPA1 (Guil and Caceres, 2007), EIF2C2/AGO2 protein binding sites (Chi et al., 2009) and ALG-1 target sites in *C. elegans* (Zisoulis et al., 2010). CLIP is limited by the low efficiency of UV 254 nm RNA-protein crosslinking, and the location of the crosslink is not readily identifiable within the sequenced crosslinked fragments, raising the question of how to separate UV-crosslinked target RNA segments from background noncrosslinked RNA fragments also present in the sample.

Here, we describe an improved method for isolation of segments of RNA bound by RBPs or RNPs, referred to as PAR-CLIP (*P*hoto*a*ctivatable-*R*ibonucleoside-Enhanced *C*ross*l*inking
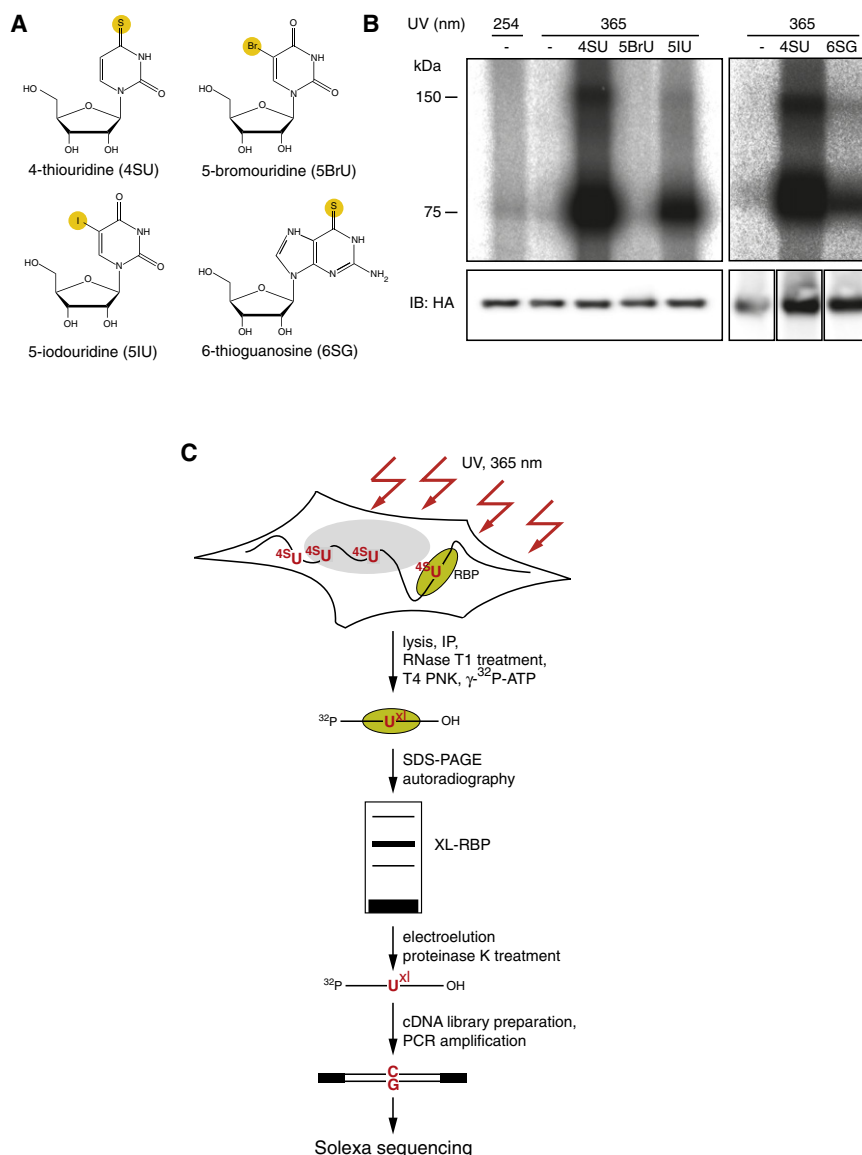
Figure 1. PAR-CLIP Methodology

(A) Structure of photoactivatable nucleosides.

(B) Phosphorimages of SDS-gels that resolved 5'-^{32}P-labeled RNA–FLAG/HA-IGF2BP1 immuno-precipitates (IPs) prepared from lysates from cells that were cultured in media in the absence or presence of 100 μM photoactivatable nucleoside and crosslinked with UV 365 nm. For comparison, a sample prepared from cells crosslinked with UV 254 nm, was included. Lower panels show immunoblots probed with an anti-HA antibody.

(C) Illustration of PAR-CLIP. 4SU-labeled transcripts were crosslinked to RBPs and partially RNase-digested RNA-protein complexes were immunopurified and size-fractionated. RNA molecules were recovered and converted into a cDNA library and deep sequenced.

osides are readily taken up by cells without apparent toxicity and have been used for in vivo crosslinking (Favre et al., 1986). We applied a subset of these nucleoside analogs (Figure 1A) to cultured cells expressing the FLAG/HA-tagged RBP IGF2BP1 followed by UV 365 nm irradiation. The crosslinked RNA-protein complexes were isolated by immunoprecipitation, and the covalently bound RNA was partially digested with RNase T1 and radiolabeled. Separation of the radiolabeled RNPs by SDS-PAGE indicated that 4SU-containing RNA crosslinked most efficiently to IGF2BP1. Compared to conventional UV 254 nm crosslinking, the photoactivatable nucleosides improved RNA recovery 100- to 1000-fold, using the same amount of radiation energy (Figure 1B). We refer to our method as PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) (Figure 1C).

We evaluated the cytotoxic effects upon exposure of HEK293 cells to 100 μM and 1 mM of 4SU or 6SG in tissue culture medium over a period of 12 hr by mRNA microarrays. The mRNA profiles of 4SU or 6SG treated cells were very similar to those of untreated cells (Table S1 available online), suggesting that the conditions for endogenous labeling of transcripts were not toxic.

To guide the development of bioinformatic methods for identification of binding sites, we first studied human Pumilio 2 (PUM2), a member of the Puf-protein family (Figure 2A) known for its highly sequence-specific RNA binding (Wang et al., 2002).

Identification of PUM2 mRNA Targets and Its RRE

PUM2 protein crosslinked well to 4SU-labeled cellular transcripts (Figure 2B). The crosslinked segments were converted into a cDNA library and Solexa sequenced (Hafner et al., 2008). The sequence reads were aligned against the human genome and EST databases. Reads mapping uniquely to the genome
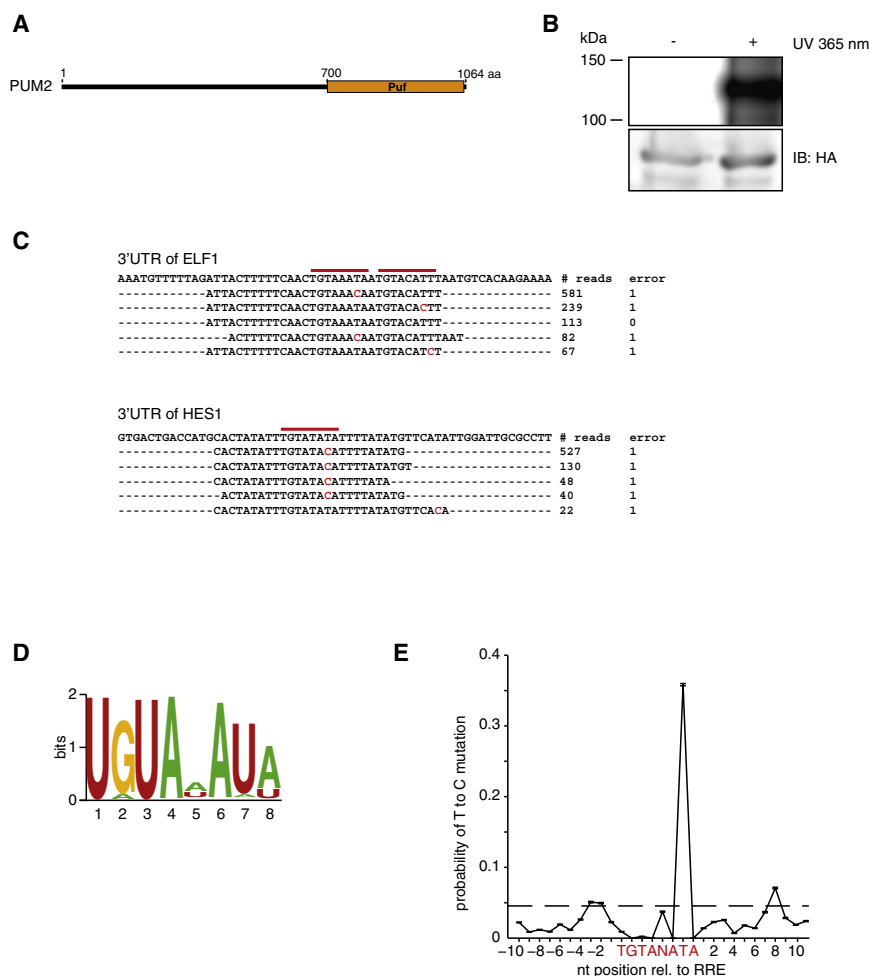
and *I*mmuno*p*recipitation). To facilitate crosslinking, we incorporated 4-thiouridine (4SU) into transcripts of cultured cells and identified precisely the RBP binding sites by scoring for thymidine (T) to cytidine (C) transitions in the sequenced cDNA. We uncovered tens of thousands of binding sites for several important RBPs and RNPs and assessed the regulatory impact of binding on their targets. These findings underscore the complexity of posttranscriptional regulation of cellular systems.

RESULTS

Photoactivatable Nucleosides Facilitate RNA-RBP Crosslinking in Cultured Cells

Random or site-specific incorporation of photoactivatable nucleoside analogs into RNA in vitro has been used to probe RBP- and RNP-RNA interactions (Kirino and Mourelatos, 2008; Meisenheimer and Koch, 1997). Several of these photoactivatable nucle-

A

PUM2 1   700   1064 aa
                                           Puf

B

C

3'UTR of ELF1

```
AAATGTTTTTAGATTACTTTTTCAACTGTAAATAATGTACATTTAATGTCACAAGAAAA   # reads   error
------------ATTACTTTTTCAACTGTAAACAATGTACATTT--------------  581       1
------------ATTACTTTTTCAACTGTAAATAATGTACACTT--------------  239       1
------------ATTACTTTTTCAACTGTAAATAATGTACATTT--------------  113       0
-------------ACTTTTTCAACTGTAAACAATGTACATTTAAT------------  82        1
------------ATTACTTTTTCAACTGTAAATAATGTACATCT--------------  67        1
```

3'UTR of HES1

```
GTGACTGACCATGCACTATATTTGTATATATTTTATATGTTCATATTGGATTGCGCCTT   # reads   error
-------------CACTATATTTGTATACATTTTATATG-------------------  527       1
-------------CACTATATTTGTATACATTTTATATGT------------------  130       1
-------------CACTATATTTGTATACATTTTATA--------------------  48        1
-------------ACTATATTTGTATACATTTTATATG------------------  40        1
-------------CACTATATTTGTATATATTTTATATGTTCACA-------------  22        1
```

D

E

Figure 2. RNA Recognition by PUM2 Protein

(A) Domain structure of PUM2 protein.

(B) Phosphorimage of SDS-gel of radiolabeled FLAG/HA-PUM2-RNA complexes from nonirradiated or UV-irradiated 4SU-labeled cells. The lower panel shows an anti-HA immunoblot.

(C) Alignments of PAR-CLIP cDNA sequence reads to corresponding regions in the 3′UTR of ELF1 and HES1 Refseq transcripts. The number of sequence reads (# reads) and mismatches (errors) are indicated. Red bars indicate the PUM2 recognition motif and red-letter nucleotides indicate T to C sequence changes.

(D) Sequence logo of the PUM2 recognition motif generated by PhyloGibbs analysis of the top 100 sequence read clusters.

(E) T to C positional mutation frequency for PAR-CLIP clusters anchored at the 8-nt recognition motif from all motif-containing clusters (Table S3). The dashed line represents the average T to C mutation frequency within these clusters.

See also Figure S1.

with up to one mismatch, insertion or deletion were used to build clusters of sequence reads (Figure 2C, Extended Experimental Procedures, and Table S2). We obtained 7523 clusters originating from about 3000 unique transcripts, 93% of which were found within the 3′ untranslated region (UTR) (Figure S1) in agreement with previous studies (Wickens et al., 2002). All sequence clusters with mapping and annotation information are available online (http://www.mirz.unibas.ch/restricted/clipdata/RESULTS/index.html).

PhyloGibbs analysis (Siddharthan et al., 2005) of the top 100 most abundantly sequenced clusters (Table S3), as expected, yielded the PUM2 RRE, UGUANAUA (Galgano et al., 2008) (Figure 2D). Unexpectedly, over 70% of all sequence reads that gave rise to clusters showed a T to C mutation compared to the genome (Figure S1). Ranking of sequence read clusters according to the frequency of T to C mutation further enriched for the PUM2 RRE (Figure S1) indicating that the T to C mutation is diagnostic of sequences interacting with the RBP. The T to C changes were not randomly distributed: the T corresponding to U7 of the RRE mutated at higher frequency compared to the Ts corresponding to U1 and U3 (Figure 2E). Our analyses suggest that the reverse transcriptase specifically misincorporated dG across from crosslinked 4SU residues and that local

amino acid environment also affected crosslinking efficiency. Uridines proximal to the RRE also exhibited an increased T to C mutation frequency, indicating that crosslinks also form in close proximity to an RRE and that our method even captured PUM2 binding sites that did not have a U7 in its RRE.

Identification of QKI RNA Targets and Its RRE

To further validate our method, we applied it to the RBP Quaking (QKI), which contains a single heterogeneous nuclear ribonucleoprotein K homology (KH) domain (Figures 3A and 3B). The RRE ACUAAY was determined by SELEX (Galarneau and Richard, 2005), but in vivo targets are largely undefined. Mice with reduced expression of QKI show dysmyelination and develop rapid tremors or "quaking" 10 days after birth. Previous studies suggested that QKI participates in pre-mRNA splicing, mRNA export, mRNA stability and protein translation (Chenard and Richard, 2008).

PhyloGibbs analysis of the 100 most abundantly sequenced clusters (Table S3) yielded the RRE AYUAAY (Figures 3C and 3D), similar to a motif identified by SELEX (Galarneau and Richard, 2005). We found approximately 6000 clusters mapping to 2500 transcripts. Close to 75% of these clusters were derived from intronic sequences, supporting the hypothesis that QKI is a splicing regulator (Chenard and Richard, 2008) and 70% of the remaining exonic clusters fall into 3′UTRs (Figure S2).

Mutation analysis of the clustered sequence reads showed that the T corresponding to U2 in AUUAAY was frequently altered to C whereas the T corresponding to U3 in AUUAAY or ACUAAY remained unaltered (Figure 3E). Crosslinking of 4SU residues located in immediate vicinity to the RRE was mostly responsible for exposing the motif with C2, showing that
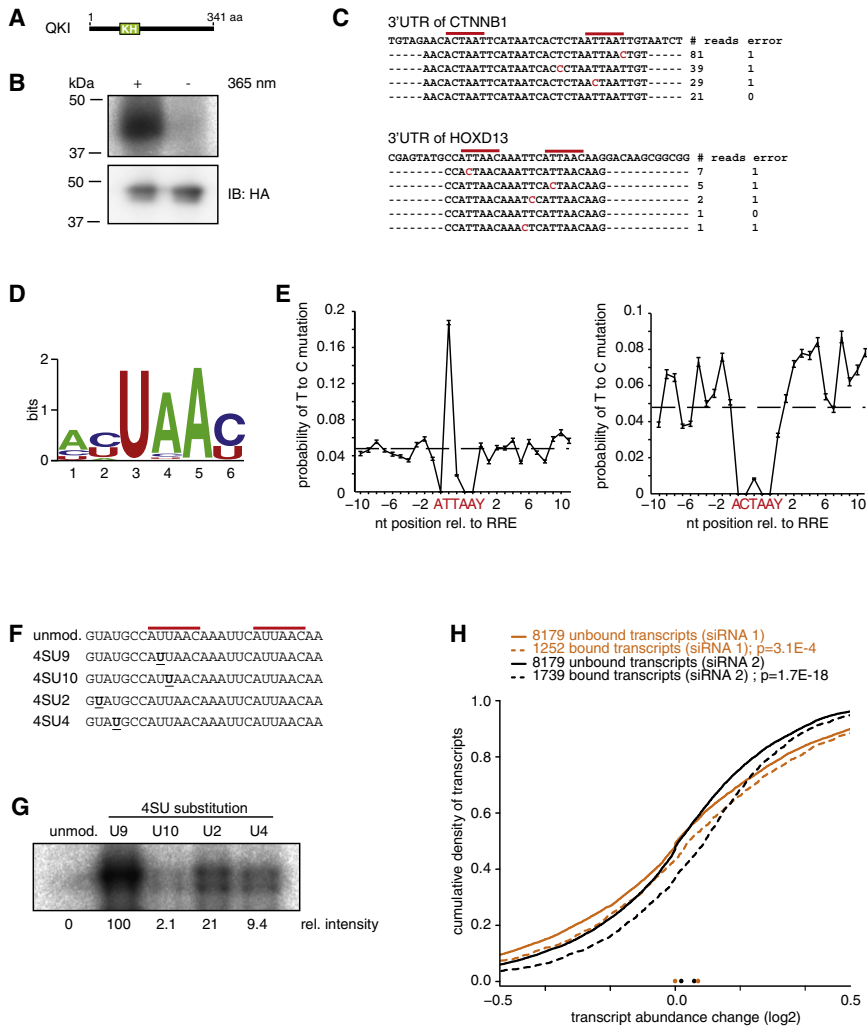
Figure 3. RNA Recognition by QKI Protein

(A) Domain structure of QKI protein.

(B) Phosphorimage of SDS-gel resolving radiolabeled RNA crosslinked to FLAG/HA-QKI IPs from nonirradiated or UV-irradiated 4SU-labeled cells. The lower panel shows the anti-HA immunoblot.

(C) Alignments of PAR-CLIP cDNA sequence reads to the corresponding regions in the 3′UTRs of the CTNNB1 and HOXD13 transcripts. Red bars indicate the QKI recognition motif and red-letter nucleotides indicate T to C sequence changes.

(D) Sequence logo of the QKI recognition motif generated by PhyloGibbs analysis of the top 100 sequence read clusters.

(E) T to C positional mutation frequency for PAR-CLIP clusters anchored at the AUUAAY (left panel) and ACUAAY (right panel) RRE (Table S3); Y = U or C. The dashed line represents the average T to C mutation frequency within these clusters.

(F) Sequences of synthetic 4SU-labeled oligoribonucleotides with QKI recognition motifs, derived from a sequence read cluster aligning to the 3′UTR of HOXD13 shown in (C) 4SU-modified residues are underlined.

(G) Phosphorimage of SDS-gel resolving recombinant QKI protein after crosslinking to radiolabeled synthetic oligoribonucleotides shown in (F). (H) Stabilization of QKI-bound transcripts upon siRNA knockdown. Two distinct siRNA duplexes (1, orange traces and 2, black traces) were used for QKI knockdown and changes in transcript stability relative to mock transfection were inferred from microarray analysis. Shown are the distributions of changes upon siRNA transfection for transcripts that did (dashed lines) or did not (solid lines) contain QKI PAR-CLIP clusters. The p-values obtained in the Wilcoxon rank-sum test comparing the changes in targeted and nontargeted transcripts are indicated.

See also Figure S2.

crosslinking inside the recognition element is not a precondition for its identification. Hence, the discovery of RREs is unlikely to be prevented by sequence-dependent crosslinking biases as long as deep enough sequencing captures these interaction sites at and nearby the RRE.

T to C Mutations Occur at the Crosslinking Sites

To better characterize the T to C transition observed in crosslinked RNA segments, we UV 365 nm crosslinked oligoribonucleotides containing single 4SU substitutions to recombinant QKI (Figures 3F and 3G). The crosslinking efficiency varied 50-fold and mirrored the results of the mutational analysis (Figure 3G). The least effective crosslinking was observed for placement of 4SU at position 3 of the QKI RRE (4SU9), and the most effective crosslinking was found at position 2 of the QKI RRE (4SU10); the crosslinking efficiency for two positions outside of the RRE (4SU2 and 4SU4) was intermediate. Neither of these substitutions affected RNA-binding to recombinant QKI protein as determined by gel-shift analysis, whereas mutations of the recognition element weakened the binding between 2.5- and 9-fold (Table S1).

Next, we sequenced libraries prepared from noncrosslinked as well as QKI-protein-crosslinked oligoribonucleotides containing 4SU at indicated positions (Figure 3F). The fraction of sequence reads with T to C changes obtained from nonirradiated 4SU-containing oligoribonucleotides varied between 10 and 20%, and increased to 50% to 80% upon crosslinking (Table S1). The variation of the degree of T to C changes in the crosslinked samples is most likely determined by background of noncrosslinked oligoribonucleotides. Presumably, the T to C transition frequency is increased upon crosslinking as a direct consequence of a chemical structure change of the 4SU nucleobase upon crosslinking to protein amino acid side chains, resulting in altered stacking or hydrogen bond donor/acceptor properties directing the preferential incorporation of dG rather than dA during reverse transcription (Figure S1). At the doses of 4SU applied to cultured cells, about 1 out of 40 uridines was substituted by 4SU as determined by HPLC analysis of the nucleoside composition of total RNA. Assuming a 20% T to C conversion rate for a noncrosslinked 4SU-labeled site, we estimated that the average T to C conversion rate of 40-nt sequence reads derived from background noncrosslinked sequences will
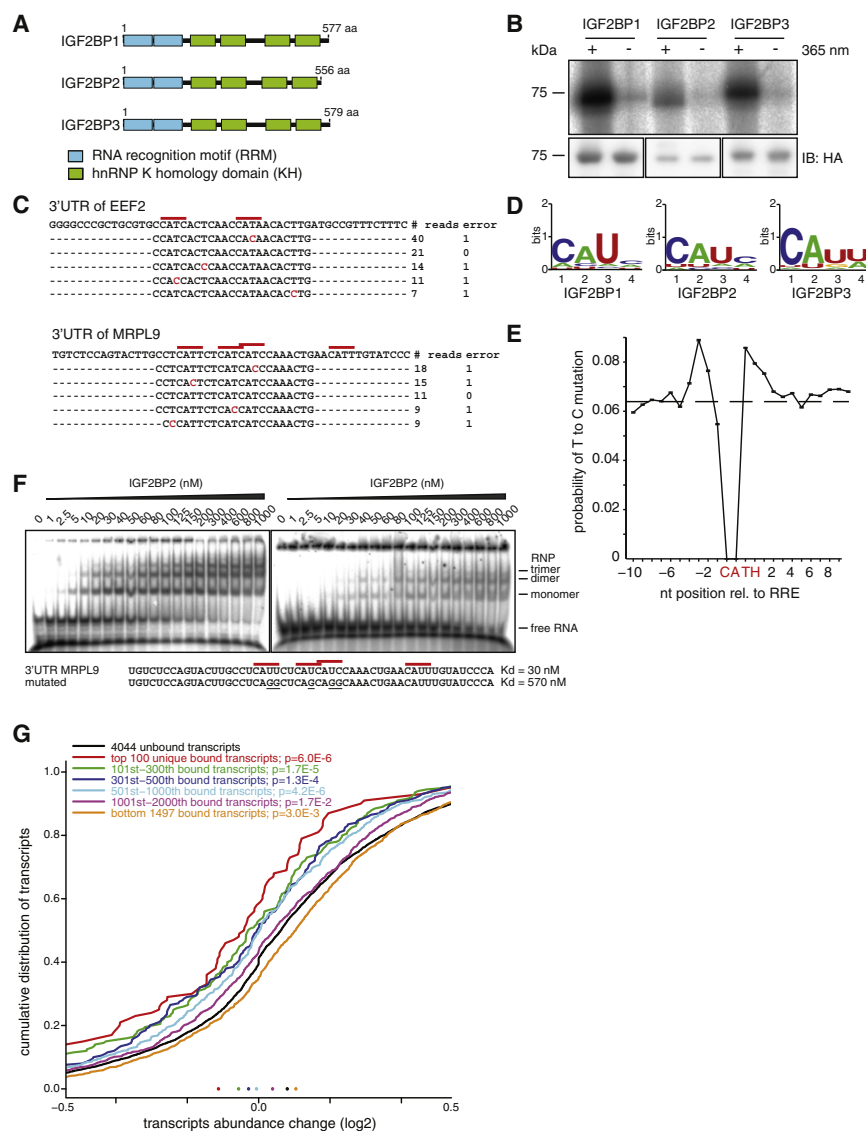
Figure 4. RNA Recognition by the IGF2BP Protein Family

(A) Domain structure of IGF2BP1-3 proteins.

(B) Phosphorimage of an SDS-gel resolving radio-labeled RNA crosslinked to FLAG/HA-IGF2BP1-3 IPs. The lower panel shows anti-HA immunoblots.

(C) Alignments of IGF2BP1 PAR-CLIP cDNA sequence reads to the corresponding regions of the 3′UTRs of EEF2 and MRPL9 transcripts. Red bars indicate the 4-nt IGF2BP1 recognition motif and nucleotides marked in red indicate T to C sequence changes.

(D) Sequence logo of the IGF2BP1-3 RRE generated by PhyloGibbs analysis of the top 100 sequence read clusters.

(E) T to C positional mutation frequency for PAR-CLIP clusters anchored at the 4-nt recognition motif from all motif-containing clusters (Table S3). The dashed line represents the average T to C mutation frequency within these clusters.

(F) Phosphorimage of native PAGE resolving complexes of recombinant IGF2BP2 protein with wild-type (left panel) and mutated target oligoribonucleotide (right panel). Sequences and dissociation constants (K_d) are indicated.

(G) Destabilization of IGF2BP-bound transcripts upon siRNA knockdown. A cocktail of three siRNA duplexes targeting IGF2BP1, 2, and 3 was used, as well as a mock transfection and changes in transcript stability were monitored by microarray analysis. Distributions of transcript level changes for IGF2BP1-3 PAR-CLIP target transcripts versus nontargeted transcripts are shown. IGF2BP1-3 target sequences were ranked and divided into bins. The p-values indicate the significance of the difference between the changes of target versus nontarget transcripts, as given by the Wilcoxon rank-sum test and are corrected for multiple testing.

See also Figure S3 and Figure S4.

be near 5%. Clusters of sequence reads with average T to C conversion above this threshold, irrespective of the number of sequence reads, most certainly represent crosslinking sites. The ability to separate signal from noise by focusing on clusters with a high frequency of T to C mutations rather than clusters with the largest number of reads, represents a major enhancement of our method over UV 254 nm crosslinking methods.

To assess whether the transcripts identified by PAR-CLIP are regulated by QKI, we analyzed the mRNA levels of mock-transfected and QKI-specific siRNA-transfected cells with microarrays. Transcripts crosslinked to QKI were significantly upregulated upon siRNA transfection, indicating that QKI negatively regulates bound mRNAs (Figure 3H), consistent with previous reports of QKI being a repressor (Chenard and Richard, 2008).

Identification of IGF2BP Family RNA Targets and Its RRE

We then applied PAR-CLIP to the FLAG/HA-tagged insulin-like growth factor 2 mRNA-binding proteins 1, 2, and 3 (IGF2BP1-3)

(Figures 4A and 4B), a family of highly conserved proteins that play a role in cell polarity and cell proliferation (Yisraeli, 2005). These proteins are predominantly expressed in the embryo and regulate mRNA stability, transport and translation. They are re-expressed in various cancers (Boyerinas et al., 2008; Dimitriadis et al., 2007) and IGF2BP2 has been associated with type-2 diabetes (Diabetes Genetics Initiative of Broad Institute of Harvard and MIT et al., 2007). The IGF2BPs are highly similar and contain six canonical RNA-binding domains, two RNA recognition motifs (RRMs) and four KH domains (Figure 4A). Therefore, target recognition for this protein family appears complex, with only a small number of coding and noncoding RNA targets being known so far. A precise definition of the RREs is missing (Yisraeli, 2005).

The three IGF2BPs recognized a highly similar set of target transcripts (Table S1), suggesting similar and redundant functions. PhyloGibbs analysis of the clusters derived from mRNAs (Figure 4C and Table S3) yielded the sequence CAUH (H = A,

U, or C) as the only consensus recognition element (Figure 4D), contained in more than 75% of the top 1000 clusters for IGF2BP1, 2 or 3 (Figure S3). In total, we identified over 100,000 sequence clusters recognized by the IGF2BP family that map to about 8,400 protein-coding transcripts. The annotation of the clusters was predominantly exonic (ca. 90%) with a slight preference for 3′UTR relative to coding sequence (CDS) (Figure S3). The mutation frequency of all sequence tags containing the element CAUH (H = A, C, or U) showed that the crosslinked residue was positioned inside the motif, or in the immediate vicinity (Figure 4E). The consensus motif CAUH was found in more than 75% of the top 1000 targeted transcripts, followed in more than 30% by a second motif, predominantly within a distance of three to five nucleotides (Figure S13). In vitro binding assays showed that nucleotide changes of the CAUH motif decreased, but did not abolish the binding affinity (Figure 4F and Table S1).

To test the influence of IGF2BPs on the stability of their interacting mRNAs, as reported previously for some targets (Yisraeli, 2005), we simultaneously depleted all three IGF2BP family members using siRNAs and compared the cellular RNA from knockdown and mock-transfected cells on microarrays. The levels of transcripts identified by PAR-CLIP decreased in IGF2BP-depleted cells, indicating that IGF2BP proteins stabilize their target mRNAs. Moreover, transcripts that yielded clusters with the highest T to C mutation frequency were most destabilized (Figure 4G), indicating that the ranking criterion that we derived based on the analysis of PUM2 and QKI data generalizes to other RBPs.

For comparison to conventional and high-throughput sequencing CLIP (Licatalosi et al., 2008; Ule et al., 2003), we also sequenced cDNA libraries prepared from UV 254 nm crosslinking. Of the 8,226 clusters identified by UV 254 nm crosslinking of IGF2BP1, 4,795 were found in the PAR-CLIP dataset. Although UV 254 nm crosslinking identified the identical segments of a target RNA as PAR-CLIP, the position of the crosslink could not be readily deduced, because no abundant diagnostic mutation was observed (Figure S4).

Identification of miRNA Targets by AGO and TNRC6 Family PAR-CLIP

To test our approach on RNP complexes, we selected the protein components mediating miRNA-guided target RNA recognition. In animal cells, miRNAs recognize their target mRNAs through base-pairing interactions involving mostly 6–8 nucleotides at the 5′ end of the miRNA (the so called "seed") (Bartel, 2009). Target sites were thought to be predominantly located in the 3′UTRs of mRNAs, and computational miRNA target prediction methods frequently resort to identification of evolutionarily conserved sites that are located in 3′UTRs and are complementary to miRNA seed regions (Bartel, 2009; Rajewsky, 2006).

We isolated mRNA fragments bound by miRNPs from HEK293 cell lines stably expressing FLAG/HA-tagged AGO or TNRC6 family proteins (Landthaler et al., 2008). The AGO IPs revealed two prominent RNA-crosslinked bands of 100 and 200 kDa, representing AGO, and likely TNRC6 and/or DICER1 protein.

The TNRC6 IPs showed one prominent RNA-crosslinked protein of 200 kDa (Figure 5A).

From clusters (Figure 5B) formed by at least 5 PAR-CLIP sequence reads and containing more than 20% T to C transitions (Table S2), we extracted 41 nt long regions centered over the predominant T to C transition or crosslinking site. The length of the crosslink-centered regions (CCRs) was selected to include all possible registers of miRNA/target-RNA pairing interactions relative to the crosslinking site.

PAR-CLIP of individual AGO proteins yielded on average about 4000 clusters that overlapped, supporting our earlier observation that AGO1-4 bound similar sets of transcripts (Landthaler et al., 2008). We therefore combined the sequence reads obtained from all AGO experiments, which yielded 17,319 clusters of sequence reads at a cut-off of 5 reads (Table S4). These clusters distributed across 4647 transcripts with defined GeneIDs, corresponding to 21% of the 22,466 unique HEK293 transcripts that we identified by digital gene expression (DGE).

PAR-CLIP of individual TNRC6 proteins yielded on average about 600 clusters that also overlapped substantially, again consistent with our observation that TNRC6 family proteins bind similar transcripts (Landthaler et al., 2008). We therefore combined all sequence reads from all TNRC6 experiments, yielding 1865 clusters and CCRs (Table S4). More than 50% of these TNRC6 CCRs fell within 25 nt of an AGO CCR, and 26% overlapped by at least 75%, indicating that AGO and TNRC6 members bind to the same sites (Figure S5).

Comparison of miRNA Profiles from AGO PAR-CLIP to Noncrosslinked miRNA profiles

To relate the potential miRNA-target-site–containing CCRs to the endogenously expressed miRNAs, we determined the miRNA profiles from total RNA isolated from HEK293 cells, and miRNAs isolated from noncrosslinked AGO1-4 IPs by Solexa sequencing (Hafner et al., 2008), and compared them to the profile from the miRNAs present in the combined AGO1-4 PAR-CLIP library. miRNA profiles obtained from total RNA and IP of the four AGO proteins in noncrosslinked cells correlated well (Figure 5C and Table S5) supporting our observation that AGO1-4 bind the same targets (Landthaler et al., 2008). The most abundant among the 557 identified miRNAs and miRNAs* were miR-103 (7% of miRNA sequence reads), miR-93 (6.5%), and miR-19b (5.5%). The 25 and 100 most abundant miRNAs accounted for 72% and 95% of the total of miRNA sequence reads, respectively. Comparison of the miRNA profile derived from the combined AGO PAR-CLIP library with the combined noncrosslinked libraries showed a good correlation (Spearman correlation coefficient of 0.56, Figure 5C and Figure S5A).

Importantly, in the AGO PAR-CLIP library, the majority of miRNA sequence reads derived from prototypical miRNAs (Landgraf et al., 2007) displayed T to C conversion near or above 50%. The T to C conversion was predominantly concentrated within positions 8 to 13 (Figure 5D), residing in the unpaired regions of the AGO protein ternary complex (Wang et al., 2008). Five of the 100 most abundant miRNAs in HEK293 cells lack uridines at position 8–13, yet only 2 of those miRNAs,
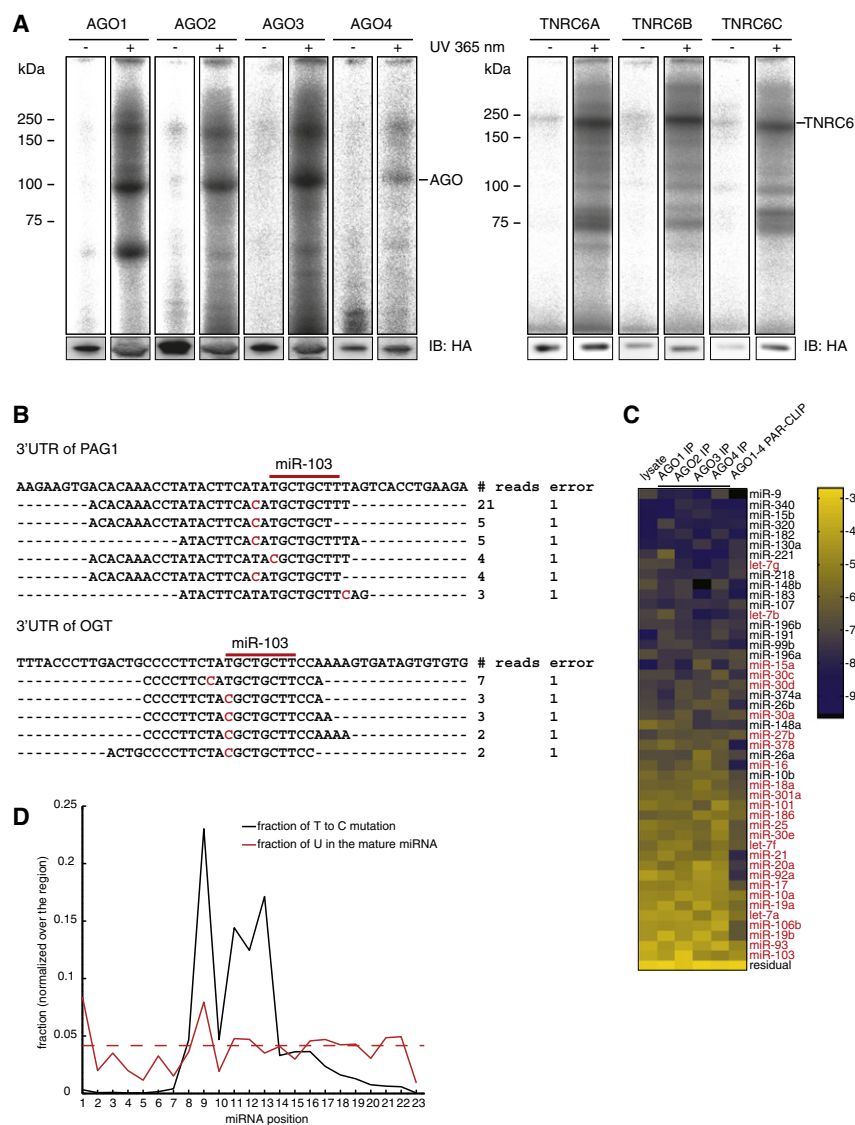
Figure 5. AGO Protein Family and TNRC6 Family PAR-CLIP

(A) Phosphorimage of SDS-gels resolving radiolabeled RNA crosslinked to the FLAG/HA-AGO1-4 and FLAG/HA-TNRC6A-C IPs. The lower panel shows the immunoblot with an anti-HA antibody.

(B) Alignment of AGO PAR-CLIP cDNA sequence reads to the corresponding regions of the 3′UTRs of PAG1 and OGT. Red bars indicate the 8-nt miR-103 seed complementary sequence and nucleotides marked in red indicate T to C mutations.

(C) miRNA profiles from RNA isolated from untreated HEK293 cells, noncrosslinked FLAG/HA-AGO1-4 IPs, and combined AGO1-4 PAR-CLIP libraries. The color code represents relative frequencies determined by sequencing. miRNAs indicated in red were inhibited by antisense oligonucleotides for the transcriptome-wide characterization of the destabilization effect of miRNA binding.

(D) T to C positional mutation frequency for miRNA sequence reads is shown in black, and the normalized frequency of occurrence of uridines within miRNAs is shown in red. The dashed red line represents the normalized mean U frequency in miRNAs.

See also Figure S5.

crosslinking site near the center of the AGO-miRNA-target-RNA ternary complex, where the target RNA is proximal to the Piwi/RNase H domain of the AGO protein (Wang et al., 2008). The polyuridine motif lies within the region of target RNA that may be able to basepair with the 3′ half of miRNA loaded into AGO proteins (Wang et al., 2008, 2009). Therefore, these stretches of uridine may contribute directly to miRNA-target RNA hybridization or, as has been suggested previously, they may represent an independent determinant of miRNA targeting specificity (Grimson et al., 2007; Hausser et al., 2009).

To further examine the positional dependence of target RNA crosslinking, we aligned the CCRs containing 7-mer seed complements to the 100 most abundant miRNAs and plotted the position-dependent frequency of finding a crosslinked position (Figure 6B). This identified two additional crosslinking regions, which correspond to the unpaired 5′ and 3′ ends of the target RNA exiting from the AGO ternary complex, indicating that the window size of 41 nt centered on the predominant crosslink position always included the miRNA-complementary sites.

We then computed the number of occurrences of miRNA-complementary sequences of various lengths in the CCRs and calculated their enrichment (Table S6). The most significant enrichment was generally obtained with 8-mers that were complementary to miRNA seed regions (pos. 1–8). Inspection of the region between 3 nt upstream and 9 nt downstream of the predominant crosslinking site reveals that approximately
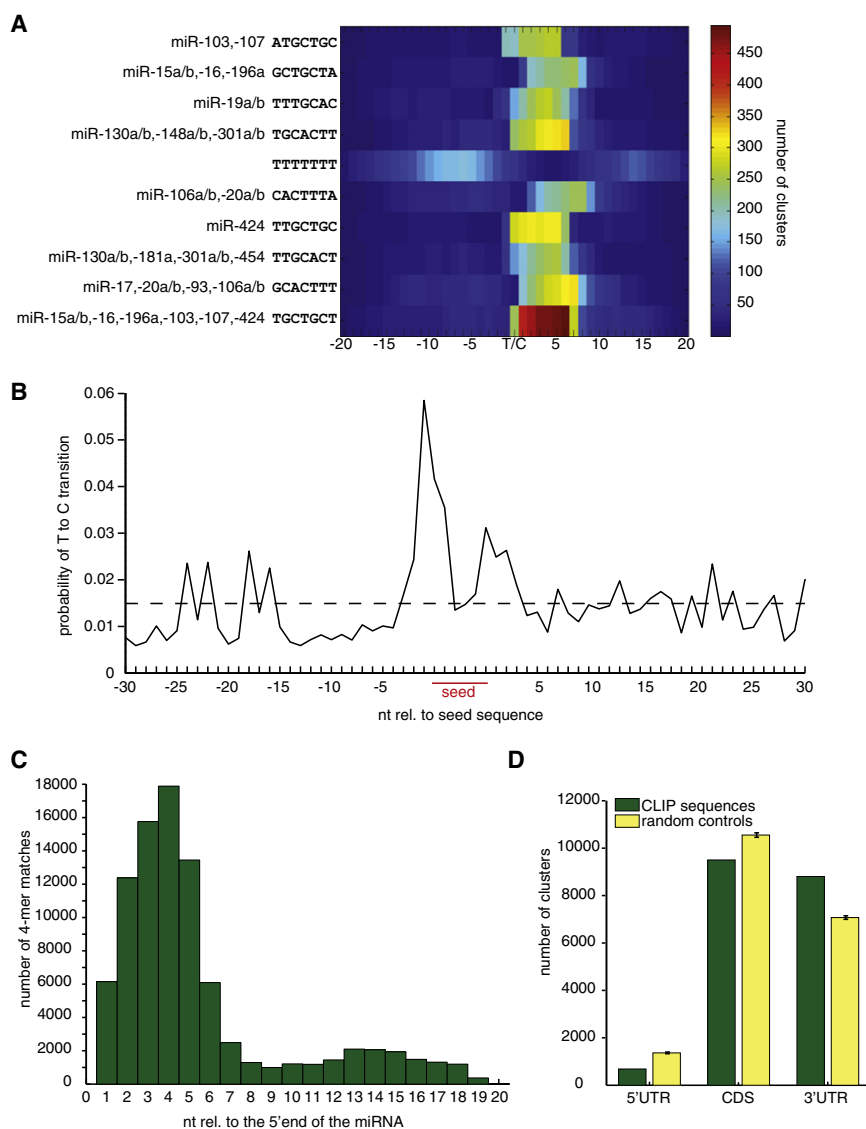
miR-374a and b, showed no crosslinking, because uridines at residues 14 and higher can still be crosslinked (Table S5). This frequency of crosslinks was substantially lower in the miRNAs whose expression did not correlate between AGO-IP and AGO PAR-CLIP samples compared to the miRNAs whose expression correlated well (Figure S5).

mRNAs Interacting with AGOs Contain miRNA Seed Complementary Sequences

Independent of any pairing models for miRNAs and their targets, we first determined the enrichment of all 16,384 possible 7-mers within the 17,319 AGO CCRs, relative to random sequences with the same dinucleotide composition. The most significantly enriched 7-mers, except for a run of uridines, corresponded to the reverse complement of the seed region (position 2–8) of the most abundant HEK293 miRNAs, and they were most frequently positioned 1–2 nt downstream of the predominant crosslinking site within the CCRs (Figure 6A). This places the

A

miR-103,-107 **ATGCTGC**
miR-15a/b,-16,-196a **GCTGCTA**
miR-19a/b **TTTGCAC**
miR-130a/b,-148a/b,-301a/b **TGCACTT**
TTTTTTT
miR-106a/b,-20a/b **CACTTTA**
miR-424 **TTGCTGC**
miR-130a/b,-181a,-301a/b,-454 **TTGCACT**
miR-17,-20a/b,-93,-106a/b **GCACTTT**
miR-15a/b,-16,-196a,-103,-107,-424 **TGCTGCT**

B

C

D

Figure 6. AGO PAR-CLIP Identifies miRNA Seed-Complementary Sequences in HEK293 Cells

(A) Representation of the 10 most significantly enriched 7-mer sequences within PAR-CLIP CCRs. T/C indicates the predominant T to C transition within clusters of sequence reads.

(B) T to C positional mutation frequency for clusters of sequence reads anchored at the 7-mer seed complementary sequence (pos. 2–8 of the miRNA) from all clusters containing seed-complementary sequences to any of the top 100 expressed miRNAs in HEK293 cells. The dashed line represents the average T to C mutation frequency within the clusters.

(C) Identification of 4-nt base-pairing regions contributing to miRNA target recognition. CCRs with at least one 7-mer seed complementary region to one of the top 100 expressed miRNAs were selected. The number of 4-nt contiguous matches in the CCRs relative to the 5′end of the matching miRNA was counted.

(D) Analysis of the positional distribution of CCRs. The number of clusters annotated as derived from the 5′UTR, CDS or 3′UTR of target transcripts is shown (green bars). Yellow bars show the expected location distribution of the crosslinked regions if the AGO proteins bound without regional preference to the target transcript.

See also Figure S6.

Noncanonical and 3′End Pairing of miRNAs to their mRNA Targets Is Limited

Structural and biochemical studies of the ternary complex of *T. thermophilus* Ago, guide and target indicated that small bulges and mismatches could be accommodated in the seed pairing region within the target RNA strand (Wang et al., 2008). We therefore searched for putative target RNA binding sites that did not conform to the model of perfect miRNA seed pairing, but rather contained a discontinuous segment of sequence complementarity to either target or miRNA with a minimum of 6 base pairs. We only considered pairing patterns if they were significantly enriched in CCRs compared to dinucleotide randomized sequences, and if the CCRs containing them did not at the same time contain perfectly pairing seed-type sites. We identified 891 CCRs with mismatches and 256 with bulges in the seed region (Table S7). Mismatches occurred most frequently across from position 5 of the miRNA as G-U or U-G wobbles, U-U mismatches and A-G mismatches (A residing in the miRNA). Therefore, it appears that only a small fraction of the miRNA target sites that we isolated (less than 6.6%), contained bulges or loops in the seed region.

To assess the role of auxiliary base pairing outside of the seed region, we selected CCRs that contained a 7-mer seed match to one of the 100 most abundant miRNAs. Supporting earlier computational results (Grimson et al., 2007), we also detected

50% of the CCRs contain 6-mers corresponding to one of the top 100 expressed miRNAs (Figure S5), with a 1.5-fold enrichment over random 6-mers. Given that 6-mers still showed some degree of excess conservation in comparative genomics studies (Gaidatzis et al., 2007; Lewis et al., 2005) (Table S6) and that our analysis was focused on a narrow window directly downstream of the crosslinking site, our results suggest that the majority of the CCRs represent bona fide miRNA binding sites. Furthermore, the number of miRNA seed complements for all known miRNAs correlated well with the expression levels of miRNAs found in HEK293 cells, and less well with miRNA profiles of other tissue samples (Figure S6B).

The nucleotide composition of CCRs that contained at least one 7-mer seed complementary to one of the top 100 expressed miRNA showed a slightly elevated U-content (approx. 30% U) compared to those CCRs not containing seed matches (Figure S6C), which was expected from previous bioinformatic analyses of functional miRNA-binding sites.
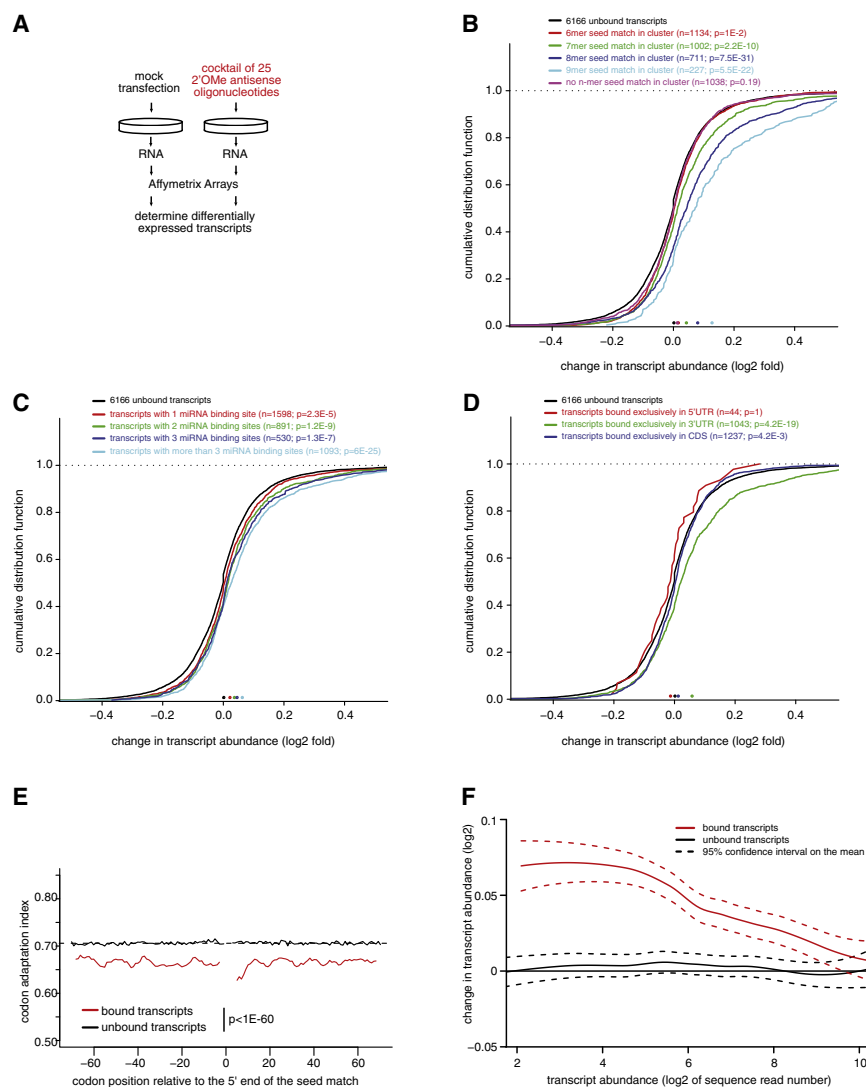
(A) FLAG/HA-AGO2-tagged HEK293 cells were transfected with a cocktail of 25 2′-O-methyl modified antisense oligoribonucleotides, inhibiting miR-NAs marked in red in Figure 5C, or mock transfected, followed by microarray analysis of the change of mRNA expression levels.

(B) Transcripts containing CCRs were categorized according to the presence of n-mer seed complementary matches and the distributions of stability changes upon miRNA inhibition are shown for these categories. The stability change for transcripts harboring CCRs without identifiable miRNA seed-complementary regions is also shown. The p values indicate the significance of the difference between the transcript level changes of transcripts containing CCRs versus transcripts without CCRs, as given by the Wilcoxon rank-sum test and are corrected for multiple testing.

(C) Transcripts were categorized according to the number of CCRs they contained.

(D) Transcripts were categorized according to the positional distribution of CCRs. Only transcripts containing CCRs exclusively in the indicated region are used.

(E) Codon adaptation index (CAI) for transcripts containing 7-mer seed complementary regions (pos. 2-8) in the CDS for the miR-15, miR-19, miR-20, and let-7 miRNA families. The red and the black lines indicate the CAI for seed-complementary sequence containing transcripts bound and not bound by AGO proteins determined by AGO PAR-CLIP.

(F) LOESS regression of total transcript abundance in HEK293 cells (log2 of sequence counts determined by digital gene expression (DGE)) against fold change of transcript abundance (log2) determined by microarrays after transfection of the miRNA antagonist cocktail versus mock transfection of AGO-bound and unbound transcripts.

See also Figure S7.

a weak signal for contiguous 4-nt long matches to positions 13–15 of the miRNA (Figure 6C).

miRNA Binding Sites in CDS and 3′UTR Destabilize Target mRNAs to Different Degrees

The majority (84%) of AGO CCRs originated in exonic regions, with only 14% from intronic, and 2% from undefined regions. Of the exonic CCRs, 4% corresponded to 5′UTRs, 50% to CDS, and 46% to 3′UTRs (Figure 6D).

Evidence of widespread binding of miRNAs to the CDS was reported before (Easow et al., 2007; Lewis et al., 2005). However, miRNAs are believed to predominantly act on 3′UTRs (Bartel, 2009), with relatively few reports providing experimental evidence for miRNA-binding to individual 5′UTRs or CDS (Easow et al., 2007; Forman et al., 2008; Lytle et al., 2007; Orom et al., 2008; Tay et al., 2008).

To obtain evidence that AGO CCRs indeed contain functional miRNA-binding sites, we blocked 25 of the most abundant miRNAs in HEK293 cells (Figure 5C) by transfection of a cocktail

of 2′-O-methyl-modified antisense oligoribonucleotides and monitored the changes in mRNA stability by microarrays (Figure 7A). Consistent with previous studies of individual miRNAs (Grimson et al., 2007), the magnitude of the destabilization effects of transcripts containing at least one CCR depended on the length of the seed-complementary region and dropped from 9-mer to 8-mer to 7-mer to 6-mer matches (Figure 7B). We did not find evidence for significant destabilization of transcripts that only contained imperfectly paired seed regions.

Next, we examined whether the change in stability of CCR-containing transcripts correlated with the number of binding sites. We found that multiple sites were more destabilizing compared to single sites (Figure 7C), and that multiple binding sites may also reside within a single 41-nt CCR (Figure S6). Both of these findings are in agreement with previous observations (Grimson et al., 2007).

Then we analyzed the impact on stability for transcripts with CCRs exclusively present either in the CDS or the 3′UTR; there were not enough transcripts to assess the impact of CCRs

derived from the 5′UTR. CDS-localized sites only marginally reduced mRNA stability (Figure 7D), independent of the extent of seed pairing. To gain more insights into miRNA binding in the CDS, we examined the codon adaptation index (CAI) (Sharp and Li, 1987) around crosslinked seed matches, and found that the sequence environment of crosslinked seed matches differed from that of noncrosslinked seed matches in the CAI. The bias in codon usage extended for at least 70 codons up- as well as downstream of the crosslinked seed matches (Figure 7E), which also correlates well with the marked increase in the A/U content around the binding sites that would lead to a codon usage bias. It was recently reported that miRNA regulation in the CDS was enhanced by inserting rare codons upstream of the miRNA-binding site, presumably due to increased lifetime of miRNA-target-RNA interactions as ribosomes are stalled (Gu et al., 2009). These observations suggest that transcripts with reduced translational efficiency form at least transient miRNP complexes amenable to UV crosslinking.

The abundance of mRNAs expressed in HEK293 cells varied over 5 orders of magnitude as shown by DGE profiling. When we related the expression level of CCR-containing transcripts with the magnitude of transcript stabilization after miRNA inhibition, we found that miRNAs preferentially act on transcripts with low and medium expression levels (Figure 7F). Highly expressed mRNAs appear to avoid miRNA regulation (Stark et al., 2005), at least for those miRNAs expressed in HEK293 cells. However, we cannot fully rule out that the weaker response of highly abundant targets may be due to lower affinity and reduced occupancy of miRNA binding sites in highly abundant transcripts.

Earlier studies defining miRNA target regulation were carried out by transfection of miRNAs into cellular systems originally devoid of these miRNAs (Baek et al., 2008; Lim et al., 2005; Selbach et al., 2008). We transfected miRNA duplexes corresponding to the deeply conserved miR-7 and miR-124 into FLAG/HA-AGO2 cells, performed PAR-CLIP (Figure S7), and also recorded the effect on mRNA stability upon miR-7 and miR-124 transfection by microarray analysis. Transcripts containing miR-7- or miR-124-specific CCRs were destabilized, especially when CCRs were located in the 3′UTR (Figure S7).

Context Dependence of miRNA Binding

Not every seed-complementary sequence in the HEK293 transcriptome yielded a CCR, thereby providing an opportunity to identify sequence context features specifically contributing to miRNA target binding and crosslinking. For seed-complementary sites that were crosslinked and those that were not crosslinked, we computed the evolutionary selection pressure by the ElMMo method (Gaidatzis et al., 2007), the mRNA stability scores by TargetScan context score (Grimson et al., 2007), and sequence composition and structure measures for the regions around the miRNA seed complementary sites. The feature that distinguished most crosslinked from noncrosslinked seed matches was a 25% lower free energy required to resolve local secondary structure involving the miRNA-binding region (Figure S7), associated with a 6% increase in the A/U content within 100 nt around the seed-pairing site. These differences were similar for sites located in the CDS and 3′UTRs. Compared to noncrosslinked sites,

crosslinked sites are under stronger evolutionary selection (ElMMo) and in sequence contexts facilitating miRNA-dependent mRNA degradation (TargetScan context score).

The location of AGO CCRs within transcript regions was nonrandom and 7-mer or 8-mer sites within the 3′UTR were preferentially located near the stop codon or the polyA tail in transcripts with relatively long 3′UTRs (more than 3 kb) (Figure S7). The location of CCRs in the CDS was biased toward the stop codon for the transfected miR-7 and 124, but not for the endogenous miRNAs (Figure S7).

Finally, we wanted to examine how miRNA targets defined by PAR-CLIP compared in regulation of target mRNA stability to those predicted by ElMMo (Gaidatzis et al., 2007), TargetScan context score (Grimson et al., 2007), TargetScan Pct (Friedman et al., 2009) and PicTar (Lall et al., 2006). In each case, we selected the same number of highest-scoring sites containing a 7-mer seed-complement to the top 5 expressed miRNAs (let-7a, miR-103, miR-15a, miR-19a, and miR-20a). The analysis was limited to 3′UTR sites due to restriction by the prediction methods. The effect on mRNA stability, as assessed by miRNA antisense inhibition, was overall equivalent for transcripts harboring CCRs compared to transcripts predicted by ElMMo, TargetScan context score, TargetScan Pct and PicTar (Figure S7).

DISCUSSION

Maturation, localization, decay and translational regulation of mRNAs involve formation of complexes of RBPs and RNPs with their RNA targets (Martin and Ephrussi, 2009; Moore and Proudfoot, 2009). Several hundred RBPs are encoded in the human genome, many of them containing combinations of RNA-binding domains which are drawn from a relatively small repertoire, resulting in diverse structural arrangements and different specificities of target RNA recognition (Lunde et al., 2007). Furthermore hundreds of miRNAs function together with AGO and TNRC6 proteins to destabilize target mRNAs and/or repress their translation (Bartel, 2009). Collectively, these factors and their presumably combinatorial action constitute the code for posttranscriptional gene regulation. Here we describe an approach to directly identify transcriptome-wide mRNA-binding sites of regulatory RBPs and RNPs in live cells.

PAR-CLIP Allows High-Resolution Mapping of RBP and miRNA Target Sites

We showed that application of photoactivatable nucleoside analogs to live cells facilitates RNA-protein crosslinking and transcriptome-wide identification of RBP and RNP binding sites. We concentrated on 4SU after it became apparent that the crosslinking sites in isolated RNAs were revealed upon sequencing by a prominent transition from T to C in the cDNA prepared from the isolated RNA segments. Compared to regular UV 254 nm crosslinking in the absence of photoactivatable nucleosides, our method has two distinct advantages. We obtain higher yields of crosslinked RNAs using similar radiation intensities, and more importantly, we can identify crosslinked regions by mutational analysis. Studies using conventional UV 254 nm CLIP have not reported the incidence of deletions and substitutions (Chi et al., 2009; Licatalosi et al., 2008; Ule et al., 2003;

Zisoulis et al., 2010), except for recent work by Grannemann et al. on the U3 snoRNA that showed an increase of deletions at the RBP binding site (Granneman et al., 2009). Our own analysis indicates that mutations in sequence reads derived from UV 254 nm CLIP were at least one order of magnitude less frequent than T to C transitions observed in PAR-CLIP (Figure S3).

From an experimental perspective, it is important to note that crosslinked RNA segments, irrespective of the methods of isolation, are always contaminated with noncrosslinked RNAs, as shown by consistent identification of rRNAs, tRNAs, and miRNAs (Table S2). Compared to crosslinked RNA fragments, these unmodified RNA molecules are more readily reverse transcribed, which underscores the need for separation of crosslinked signal from noncrosslinked noise. We now provide a method that accomplishes this critical task.

Context Dependence of 4SU Crosslink Sites

It is conceivable that binding sites located in peculiar sequence environments, e.g., those completely devoid of U, may exist and cannot be captured using 4SU-based crosslinking. However, such sites are extremely rare. Only about 0.4% of 32-nt long sequence segments, representative of the length of our Solexa sequence reads, are U-less, corresponding to an occurrence of one such segment in every 8 kb of a transcript.

Nonetheless, to provide a means to resolve such unlikely situations, we explored the use of other photoactivatable nucleosides, such as 6SG to identify IGF2BP1 binding sites. We found a good correlation between the sequence reads obtained from a given gene with 4SU and 6SG (Pearson correlation coefficient 0.65, Table S1). Moreover, the sequence read clusters, representing individual binding sites, overlapped strongly: 59% out of the 47,050 6SG clusters were also identified with 4SU, despite of the fact that the environment of IGF2BP1 binding sites was strongly depleted for guanosine. Interestingly, the sequence reads obtained after 6SG crosslinking were enriched for G to A transitions, pointing to a structural change in 6SG analogous to the situation in PAR-CLIP with 4SU. Because 6SG appears to have lower crosslinking efficiency compared to 4SU, we recommend to first use 4SU and then resort to 6SG when the data indicates that the sites of interest are located in sequence contexts devoid of uridines. It is important to point out that neither of these photoactivatable nucleotides appears to be toxic under our recommended conditions.

miRNA Target Identification

When applying PAR-CLIP to isolate miRNA-binding sites, we were surprised to find nearly 50% of the binding sites located in the CDS. However, miRNA inhibition experiments showed that miRNA binding at these sites only caused small, yet significant mRNA destabilization. In spite of the difference in their efficiency of triggering mRNA degradation, CDS and 3′UTR sites appear to have similar sequence and structure features. The sequence bias around CDS sites is associated with an increased incidence of rare codon usage, which could in principle reduce translational rate, thereby providing an opportunity for transient miRNP binding and regulation. Similar observations were made previously using artificially designed reporter systems (Gu et al., 2009).

The use of the knowledge of the crosslinking site allowed us to narrowly define the miRNA-binding regions for matching the site with the most likely miRNA endogenously co-expressed with its targets, and to assess noncanonical miRNA-binding modes. We were able to explain the majority of PAR-CLIP binding sites by conventional miRNA-mRNA seed-pairing interactions (Grimson et al., 2007), yet found that about 6% of miRNA target sites might best be explained by accepting bulges or mismatches in the seed pairing region, similar to the interaction between let-7 and its target lin-41 (Vella et al., 2004) and those recently observed in biochemical and structural studies of *T. thermophilus* Ago protein (Wang et al., 2008; Wang et al., 2009).

The mRNA Ribonucleoprotein Code and Its Impact on Gene Regulation

We were able to identify all of the crosslinkable RNA-binding sites present in about 9,000 of the top-expressed mRNA in HEK293 cells representing approximately 95% of the total mRNA molecules of a cell. One of the surprising outcomes of our study was that each of the examined RBPs or miRNPs bound and presumably controlled between 5 and 30% of the more than 20,000 transcripts detectable in HEK293 cells. These results demonstrate that a transcript will generally be bound and regulated by multiple RBPs, the combination of which will determine the final gene-specific regulatory outcome. Exhaustive high-resolution mapping of RBP– and RNP–target-RNA interactions is critical, because it may lead to the discovery of specific combination of sites (or modules) that may control distinct cellular processes and pathways. To gain further insights into the dynamics of mRNPs it will be important to also map the sites of RNA-binding factors, such as helicases, nucleases or polymerases, where the specificity determinants are poorly understood. The precise identification of RNA interaction sites will be extremely useful for interrogating the rapidly emerging data on genetic variation between individuals and whether some of these variations possibly contribute to complex genetic diseases by affecting posttranscriptional gene regulation.

EXPERIMENTAL PROCEDURES

PAR-CLIP

Human embryonic kidney (HEK) 293 cells stably expressing FLAG/HA-tagged IGF2BP1-3, QKI, PUM2, AGO1-4, and TNRC6A-C (Landthaler et al., 2008) were grown overnight in medium supplemented with 100 μM 4SU. Living cells were irradiated with 365 nm UV light. Cells were harvested and lysed in NP40 lysis buffer. The cleared cell lysates were treated with RNase T1. FLAG/HA-tagged proteins were immunoprecipitated with anti-FLAG antibodies bound to Protein G Dynabeads. RNase T1 was added to the immunoprecipitate. Beads were washed and resuspended in dephosphorylation buffer. Calf intestinal alkaline phosphatase was added to dephosphorylate the RNA. Beads were washed and incubated with polynucleotide kinase and radioactive ATP to label the crosslinked RNA. The protein-RNA complexes were separated by SDS-PAGE and electroeluted. The electroeluate was proteinase K digested. The RNA was recovered by acidic phenol/chloroform extraction and ethanol precipitation. The recovered RNA was turned into a cDNA library as described (Hafner et al., 2008) and Solexa sequenced. The extracted sequence reads were mapped to the human genome (hg18), human mRNAs and miRNA precursor regions. For a more detailed description of the methods, see the Extended Experimental Procedures. For a video presenting the procedure please visit http://www.jove.com/index/Details.stp?ID=2034.

Oligonucleotide Transfection and mRNA Array Analysis

siRNA, miRNA and 2′-O-methyl oligonucleotide transfections of HEK293 T-REx Flp-In cells were performed in 6-well format using Lipofectamine RNAiMAX (Invitrogen) as described by the manufacturer. Total RNA of transfected cells was extracted using TRIZOL following the instructions of the manufacturer. The RNA was further purified using the RNeasy purification kit (QIAGEN). 2 μg of purified total RNA was used in the One-Cycle Eukaryotic Target Labeling Assay (Affymetrix) according to manufacturer's protocol. Biotinylated cRNA targets were cleaned up, fragmented, and hybridized to Human Genome U133 Plus 2.0 Array (Affymetrix). For details of the analysis, see Bioinformatics section in the Supplementary Material.

Generation of Digital Gene Expression (DGEX) Libraries

1 μg each of total RNA from HEK293 cells inducibly expressing tagged IGF2BP1 before and after induction was converted into cDNA libraries for expression profiling by sequencing using the DpnII DGE kit (Illumina) according to instructions of the manufacturer. For details of the analysis, see Bioinformatics section in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at doi:10.1016/j.cell.2010.03.009.

REFERENCES

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature 455, 64–71.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell 136, 215–233.

Boyerinas, B., Park, S.-M., Shomron, N., Hedegaard, M.M., Vinther, J., Andersen, J.S., Feig, C., Xu, J., Burge, C.B., and Peter, M.E. (2008). Identification of Let-7-Regulated Oncofetal Genes. Cancer Res. 68, 2587–2591.

Chenard, C.A., and Richard, S. (2008). New implications for the QUAKING RNA binding protein in human disease. J. Neurosci. Res. 86, 233–242.

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

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, L.U.a.N.I.o.B.R., Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I.W., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., et al. (2007). Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. Science 316, 1331–1336.

Dimitriadis, E., Trangas, T., Milatos, S., Foukas, P.G., Gioulbasanis, I., Courtis, N., Nielsen, F.C., Pandis, N., Dafni, U., Bardi, G., et al. (2007). Expression of

oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. Int. J. Cancer 121, 486–494.

Dreyfuss, G., Choi, Y.D., and Adam, S.A. (1984). Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. Mol. Cell. Biol. 4, 1104–1114.

Easow, G., Teleman, A.A., and Cohen, S.M. (2007). Isolation of microRNA targets by miRNP immunopurification. RNA 13, 1198–1204.

Favre, A., Moreno, G., Blondel, M.O., Kliber, J., Vinzens, F., and Salet, C. (1986). 4-thiouridine photosensitized RNA-protein crosslinking in mammalian cells. Biochem. Biophys. Res. Commun. 141, 847–854.

Forman, J.J., Legesse-Miller, A., and Coller, H.A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc. Natl. Acad. Sci. USA 105, 14879–14884.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105.

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

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

Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M., and Gerber, A.P. (2008). Comparative Analysis of mRNA Targets for Human PUF-Family Proteins Suggests Extensive Interaction with the miRNA Regulatory System. PLoS ONE 3, e3164.

Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 103, 4487–4492.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Nat. Acad. Sci. 106, 9613–9618.

Greenberg, J.R. (1979). Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res. 6, 715–732.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105.

Gu, S., Jin, L., Zhang, F., Sarnow, P., and Kay, M.A. (2009). Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. Nat. Struct. Mol. Biol. 16, 144–150.

Guil, S., and Caceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat. Struct. Mol. Biol. 14, 591.

Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods 44, 3–12.

Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D., and Zavolan, M. (2009). Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. Genome Res. 19, 2009–2020.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. 8, 533–543.

Kirino, Y., and Mourelatos, Z. (2008). Site-specific crosslinking of human microRNPs to RNA targets. RNA 14, 2254–2259.

Lall, S., Grun, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., MacMenamin, P., et al. (2006). A Genome-Wide Map of Conserved MicroRNA Targets in C. elegans. Curr. Biol. 16, 460–471.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. Cell 129, 1401–1414.

Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA *14*, 2580–2596.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769–773.

Lopez de Silanes, I., Zhan, M., Lal, A., Yang, X., and Gorospe, M. (2004). Identification of a target RNA motif for RNA-binding protein HuR. Proc. Natl. Acad. Sci. USA *101*, 2987–2992.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Lytle, J.R., Yario, T.A., and Steitz, J.A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. Proc. Natl. Acad. Sci. USA *104*, 9667–9672.

Martin, K.C., and Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial Dimension. Cell *136*, 719–730.

Mayrand, S., Setyono, B., Greenberg, J.R., and Pederson, T. (1981). Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)+ heterogeneous nuclear RNA in living HeLa cells. J. Cell Biol. *90*, 380–384.

McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., and Silver, P.A. (2005). A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. BMC Dev. Biol. *5*, 14.

Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-linking of nucleic acids to associated proteins. Crit. Rev. Biochem. Mol. Biol. *32*, 101–140.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. Cell *136*, 688–700.

Orom, U.A., Nielsen, F.C., and Lund, A.H. (2008). MicroRNA-10a Binds the 5′UTR of Ribosomal Protein mRNAs and Enhances Their Translation. Mol. Cell *30*, 460–471.

Rajewsky, N. (2006). microRNA target predictions in animals. Nat. Genet. *38*, S8–S13.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res. *19*, 381–394.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. *15*, 1281–1295.

Siddharthan, R., Siggia, E.D., and van Nimwegen, E. (2005). PhyloGibbs: A Gibbs Sampling Motif Finder That Incorporates Phylogeny. PLoS Comp. Biol. *1*, e67.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell *136*, 731–745.

Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. (2005). Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution. Cell *123*, 1133–1146.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. USA *97*, 14085–14090.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., and Slack, F.J. (2004). The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. Genes Dev. *18*, 132–137.

Wagenmakers, A.J., Reinders, R.J., and van Venrooij, W.J. (1980). Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. Eur. J. Biochem. *112*, 323–330.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M.T. (2002). Modular Recognition of RNA by a Human Pumilio-Homology Domain. Cell *110*, 501–512.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921–926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754–761.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3′UTR regulation as a way of life. Trends Genet. *18*, 150–157.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D., and Gage, F.H. (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.

Yisraeli, J.K. (2005). VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. Biol. Cell *97*, 87–96.

Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat. Struct. Mol. Biol. *17*, 173–179.