RESA identifies mRNA-regulatory sequences at high resolution

Valeria Yartseva[1,5], Carter M Takacs[1,5], Charles E Vejnar[1], Miler T Lee[1,2] & Antonio J Giraldez[1,3,4]

Gene expression is extensively regulated at the levels of mRNA stability, localization and translation. However, decoding functional RNA-regulatory features remains a limitation to understanding post-transcriptional regulation *in vivo*. Here, we developed RNA-element selection assay (RESA), a method that selects RNA elements on the basis of their activity *in vivo* and uses high-throughput sequencing to provide a quantitative measurement of their regulatory functions at near-nucleotide resolution. We implemented RESA to identify sequence elements modulating mRNA stability during zebrafish embryogenesis. RESA provides a sensitive and quantitative measure of microRNA activity *in vivo* and also identifies novel regulatory sequences. To uncover specific sequence requirements within regulatory elements, we developed a bisulfite-mediated nucleotide-conversion strategy for large-scale mutational analysis (RESA–bisulfite). Finally, we used the versatile RESA platform to map candidate protein–RNA interactions *in vivo* (RESA–CLIP).

Post-transcriptional control is critical in the regulation of cellular protein levels. Mutations disrupting mRNA-regulatory pathways have been linked to developmental defects and human diseases[1–4], thus underscoring the importance of mRNA-regulatory elements. Post-transcriptional regulation occurs through mRNA processing, nuclear export, translation, localization and stability, all of which strongly affect gene expression. This control is achieved by specificity factors that typically recognize elements within mRNA 5′ and 3′ untranslated regions (UTRs).

Decoding mRNA regulation transcriptome wide remains a challenge. Pioneering work assaying sequence fragments ('bashing') combined with mutational analysis has identified regulatory sequences within individual transcripts[4–8], thus laying the groundwork for high-throughput methods that have expanded understanding of mRNA regulation. RNA profiling approaches have illuminated the microRNA (miRNA) regulatory landscape by identifying miRNA targets transcriptome wide[9–12], refining miRNA–mRNA targeting rules[12–14] and mapping miRNA-binding sites on target mRNAs[15,16]. Additionally, alternative 3′-UTR sequences promoting mRNA stability have been discovered through measurements of isoform half-life differences[17].

Recent approaches have combined plasmid reporters with custom oligonucleotides to discover sequences with regulatory potential[18–22]. However, these methods have disadvantages for comprehensively identifying discrete regulatory elements across a transcriptome (**Supplementary Table 1**). Oligonucleotide libraries represent a tradeoff between sequence diversity and length. Short oligonucleotides[22] cannot represent regulatory elements with longer sequence requirements, whereas longer random oligonucleotide libraries[19] can contain only a subset of the exponentially large number of possible sequences and therefore may not adequately represent all biologically relevant elements. Targeted libraries that curate sequences according to criteria such as conservation[18,21] are limited in scope by design, whereas approaches that build reporters from 3′ UTRs in full[23] or in large windows[20] have limited resolution to identify discrete elements. Finally, plasmid delivery constrains library sequence diversity and can require normalization during analysis.

To achieve a comprehensive map of the mRNA-regulatory landscape, we engineered an *in vivo* reporter system with flexible sequence input and the capacity to select functional RNA elements across large candidate regions, on the basis of biological function. Here, we describe RESA, a high-throughput technique that measures regulatory activities of mRNA sequences at near-nucleotide resolution. Rather than relying on custom oligonucleotide synthesis, RESA couples sequence amplification with random fragmentation, thereby generating highly overlapping fragments spanning RNA regions of interest. This library is introduced into an *in vivo* system, subjected to experimental treatment and then assayed for differential selection through RNA-seq. This flexible framework also enabled us to develop RESA–bisulfite (RESA-B), which introduces mutations across the library and allows detection of regulatory-sequence nucleotide requirements, and RESA coupled with cross-linking and immunoprecipitation (RESA–CLIP), which selects for specific protein–RNA interactions.

RESULTS

RESA design and implementation

RESA measures the regulatory strength of diverse sequence fragments by quantifying position-specific changes in abundance

[1]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA. [2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [3]Yale Stem Cell Center, Yale University School of Medicine, New Haven, Connecticut, USA. [4]Yale Cancer Center, Yale University School of Medicine, New Haven, Connecticut, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to A.J.G. (antonio.giraldez@yale.edu) or M.T.L. (miler@pitt.edu).

RECEIVED 29 SEPTEMBER; ACCEPTED 2 DECEMBER; PUBLISHED ONLINE 26 DECEMBER 2016; DOI:10.1038/NMETH.4121

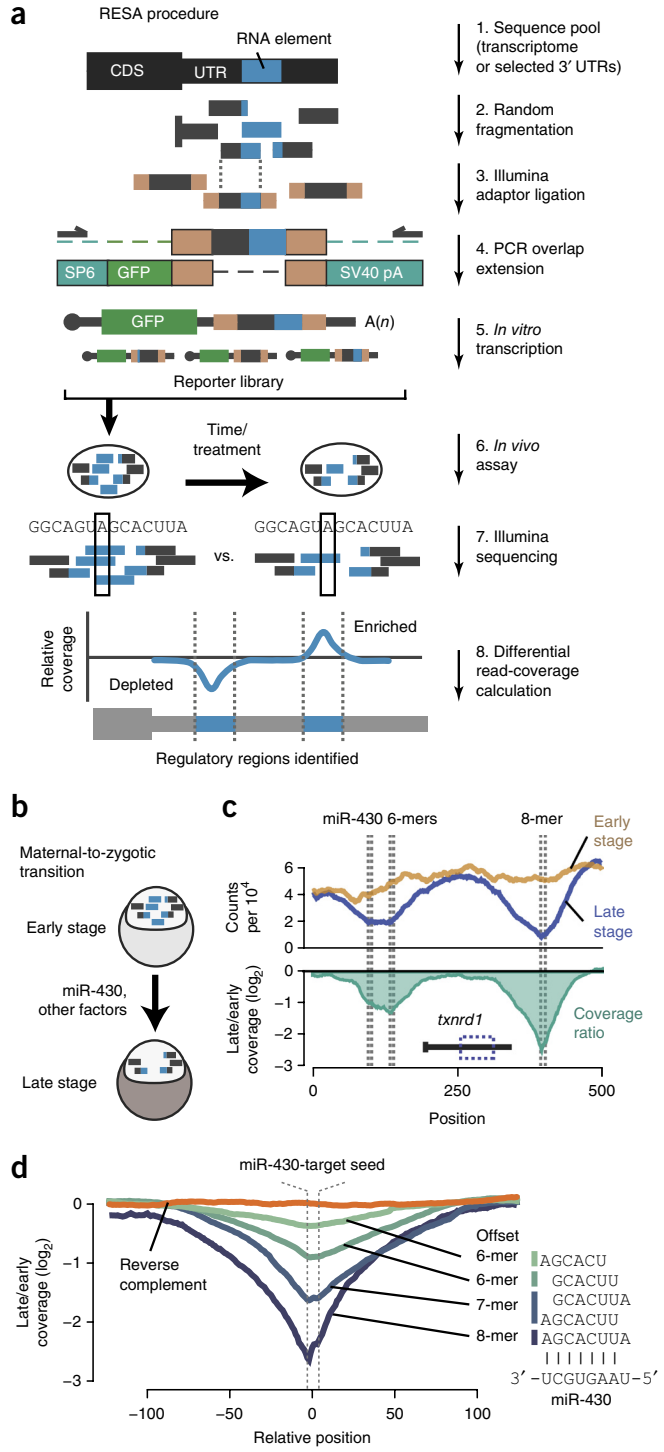

Figure 1 | RESA is a high-throughput method to systematically map mRNA-regulatory sequences *in vivo*. (**a**) Pipeline for RESA library generation. CDS, coding sequence; pA, poly(A). (**b**) RESA application to mRNA stability during the zebrafish MZT, when maternal mRNAs are destabilized by miR-430 and unknown factors. (**c**) Top, coverage profile for regions overlapping miR-430-target seed sequences (dashed lines) within the *txnrd1* locus, at the early stage (yellow) and late stage (blue). Bottom, ratio of late-stage versus early-stage profiles (green). (**d**) Metaprofiles of late-stage versus early-stage coverage ratios centered on the miR-430-target seed sequence, grouped on the basis of seed complementarity: 8-mers (n = 8), 7-mers (n = 31), 6-mers (n = 94), offset 6-mers (n = 95) and reverse-complement 8-mers and 7-mers (n = 39).

by using high-throughput sequencing (**Fig. 1a**). To generate the RESA library, we isolated 3′-UTR sequences through PCR amplification, induced random fragmentation through sonication, ligated Illumina sequencing adaptors and used overlap-extension PCR to introduce the adaptors into the 3′ UTRs of a GFP open reading frame with an SP6 promoter and an SV40 polyadenylation signal. This strategy avoids cloning and generates a library of high sequence complexity. As a byproduct of library preparation, reverse-complement inserts are also generated and can serve as internal controls for their forward-direction counterparts (**Supplementary Fig. 1a**,**b**). Alternatively, this method can use existing RNA-seq libraries as source DNA for the overlap-extension PCR, thus enabling interrogation of regulatory activity across a transcriptome. In both cases, *in vitro* transcription generates a diverse mRNA reporter pool with identical 5′ UTRs, coding sequences and polyadenylation signals but variable 3′ UTRs (**Fig. 1a**).

Embryonic development in all animals requires global remodeling of the transcriptome (reviewed in ref. 24), thus providing an ideal context to study post-transcriptional regulation. To interrogate mRNA regulation during the maternal-to-zygotic transition (MZT) in zebrafish, we built a RESA library of 3′-UTR fragments from maternally provided mRNAs (**Fig. 1b**), spanning 434 genes and 502,728 nucleotides (**Supplementary Data 1** and **2**).

Identifying regulatory RNA elements at high resolution requires extensive sampling of fragments throughout the mRNA. To assess sequence diversity, we sequenced the *in vitro*–transcribed RESA library by using the built-in Illumina adapters. We found that sonication yielded a highly diverse population of overlapping fragments of 56–129 nt in length, which mapped throughout the UTRs (**Supplementary Fig. 1c**). This diversity was maintained after injection into embryos (**Supplementary Fig. 1d**). 97% of the UTRs had ≥20× median sequencing coverage across their length in the injected library, whereas 87% had ≥100× coverage (**Supplementary Fig. 1e**). Unique read start sites occurred at 69% of all nucleotide positions in the library, and per UTR, a median 83% of nucleotides overlapped read start sites (**Supplementary Fig. 1f**), thus indicating a high abundance of unique fragments. Together, these results demonstrate that RESA efficiently generates an mRNA library of extensive sequence complexity.

RESA detects and quantifies discrete regulatory RNA elements at high resolution

To assess how effectively RESA detects RNA-regulatory potential *in vivo*, we injected the library into replicate sets of one-cell embryos (**Supplementary Table 2**) and performed RNA-seq at 2 ('early stage') and 8 ('late stage') hours post fertilization (hpf) to identify fragments that became enriched or depleted as a measure of their regulatory activities in the cognate mRNAs (**Fig. 1b**). We first evaluated RESA's ability to identify miR-430-induced mRNA destabilization. miR-430 is zygotically transcribed[3,25] and canonically triggers the translation repression, deadenylation and destabilization of several hundred maternal mRNAs by recognizing complementary target sequences (GCACUU) within target 3′ UTRs[3,9,26]. The RESA library includes 29 validated miR-430-target 3′ UTRs[9]. Whereas sequencing-read abundance across the library was highly correlated between replicate time points (R ≥0.97) (**Supplementary Fig. 1h**,**i**), we observed strong valleys of depletion between different time points in fragments overlapping

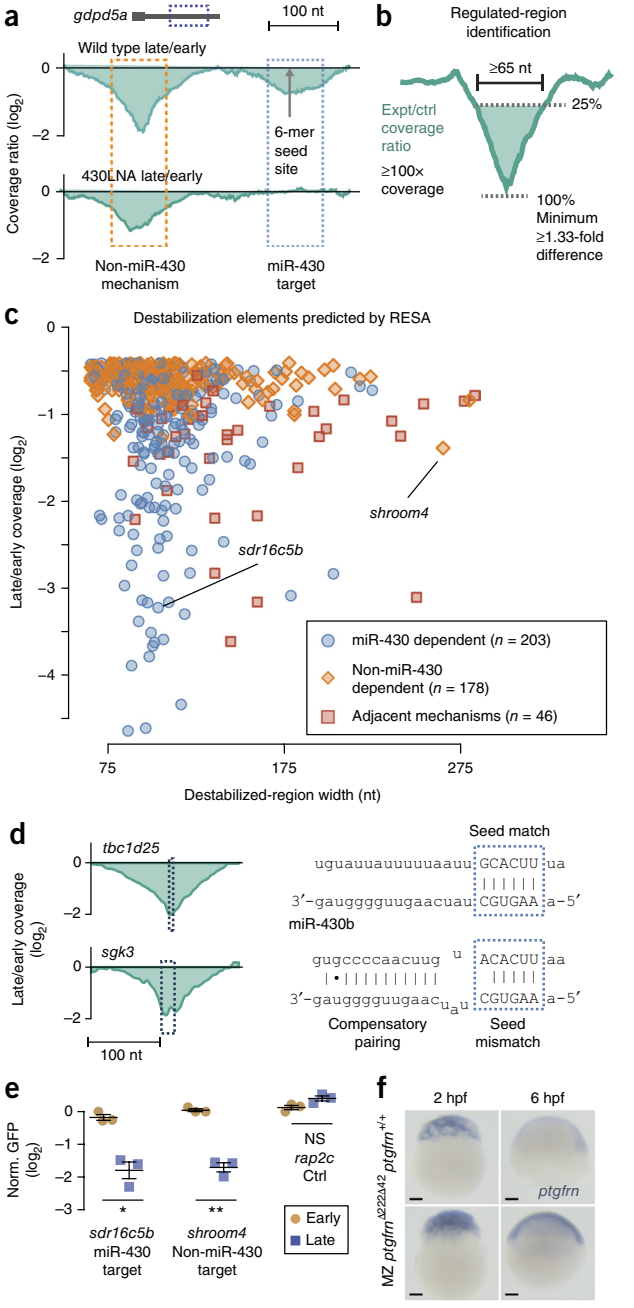

Figure 2 | RESA identifies sequences that promote mRNA destabilization. (**a**) Coverage depletion over time in the *gdpd5a* 3′-UTR (top) and in the presence of 430LNA (bottom). (**b**) Regulated regions were identified by searching for a ≥1.33-fold depletion in late versus early stages, within a ≥65-nt region with ≥100× sequencing coverage. Expt, experimental; ctrl, control. (**c**) Biplot of RESA-predicted destabilization regions, separated on the basis of mode of regulation. (**d**) Wild-type coverage depletion around a canonical miR-430 6-mer site (*tbc1d25*) and noncanonical seed mismatch site with 3′ compensatory pairing (*sgk3*). (**e**) GFP reporter analysis of RESA-predicted regulatory elements with different regulatory behaviors. GFP mRNA levels are assayed at early (2 hpf; yellow circles) and late (8 hpf; blue squares) stages by qRT–PCR and normalized (norm.) to a control dsRed mRNA. Plot shows mean ± s.e.m. $^{**}P < 0.01$; $^{*}P < 0.05$; NS, not significant by Welch's *t* test; additional statistical information in Online Methods. (**f**) *In situ* hybridization to detect endogenous *ptgfrn* mRNA levels in MZ mutants for the predicted destabilization element (MZ $ptgfrn^{\Delta 222\Delta 42}$) compared with $ptgfrn^{+/+}$ at 2 and 6 hpf. Scale bars, 100 µm.

validated miR-430-target seeds, such as in the *txnrd1* 3′ UTR[9] (**Fig. 1c** and **Supplementary Fig. 1g**). The greatest magnitude of depletion precisely coincided with the nucleotide positions of the target seeds (**Fig. 1c**), thus suggesting that RESA is able to specifically map the effects of miR-430-mediated destabilization.

The number of contiguous base pairs between the miRNA seed and target site correlates with targeting efficacy[14,26]. Accordingly, we found that the magnitude of fragment depletion was significantly stronger for greater matches to the miR-430 seed (8-mers > 7-mers > 6-mers > offset 6-mers) ($P < 3.8 \times 10^{-39}$, Kruskal–Wallis test) (**Fig. 1d** and **Supplementary Fig. 2a**,**b**), thus demonstrating RESA's ability to quantify the strength of *in vivo* regulation.

RESA identifies RNA elements regulating mRNA stability

Next, we searched the entire RESA library for differentially regulated regions. To distinguish miR-430-dependent and miR-430-independent effects, we repeated RESA in the absence of miR-430 function by using a tiny locked nucleic acid complementary to the miR-430 seed (430LNA)[27]. Under these conditions, we found that miR-430-mediated regulation was specifically blocked (**Fig. 2a** and **Supplementary Fig. 2c**,**d**) and were consequently able to isolate the effects of non-miR-430-dependent destabilization.

To identify regulated regions, we parameterized a search strategy based on features of miR-430 targets. We performed simulations on a known miR-430-regulated mRNA to measure the effects of varying read coverage on target-site discovery and determined that we were able to confidently recover destabilized regions with a ≥1.33-fold decrease in sequencing coverage from early to late stage, given ≥100× coverage (**Supplementary Fig. 3**). Using these parameters on pooled replicates, with a region-width threshold of ≥65 nt, as measured at 25% of the valley minimum (**Fig. 2b** and **Supplementary Fig. 4**), we found 427 destabilized regions across 273 3′ UTRs (**Fig. 2c**). Of these, 52% showed evidence of regulation in the absence of miR-430 function (**Fig. 2c**, **Supplementary Figs. 5** and **6** and **Supplementary Data 3**), including 46 regions that appeared to be composed of adjacent miR-430-dependent and miR-430-independent elements (**Fig. 2c** and **Supplementary Fig. 5c**). The majority of destabilized regions (86%) were robustly identified even in individual replicates in isolation (**Supplementary Fig. 7a–c**), and all 29 of the previously validated miR-430 targets[9] were correctly classified, thus demonstrating the efficacy of this method. An additional 400 lower-confidence destabilizing elements were identified, as well as 361 regions that appeared to enhance stability in late-stage embryos (**Supplementary Figs. 6b** and **7d**, and **Supplementary Data 3**), thus suggesting that multiple mechanisms regulate mRNA levels during embryogenesis.

miR-430-target identification was highly specific. 93% of 8-mer target sequences and 86% of 7-mers fell within high-confidence miR-430-dependent destabilized regions. Weaker seed matches were less likely to be destabilized, and 35% of offset 6-mer sequences showed no evidence of regulation, a result reflecting the existence of other features that modulate targeting efficacy[14] (**Supplementary Table 3**). 82% of target-seed-containing regions classified by RESA as miR-430 dependent had valley minima within 10 nt of the seed (**Supplementary Fig. 7e**), thus indicating the high nucleotide precision of predicting regulatory-element location with RESA. In addition to canonical seed-based regulation, RESA identified 46 miR-430-dependent regions

lacking a seed match (**Fig. 2d** and **Supplementary Fig. 8a**,**b**), thus suggesting the prevalence of noncanonical miR-430 targeting. These regions were highly enriched in partial 5-nt matches to the miR-430 seed ($P = 4.3 \times 10^{-13}$, chi-squared test), as well as complementarity to the 3′ end of mature miR-430 ($P = 7.3 \times 10^{-5}$, Wilcoxon rank-sum test) (**Fig. 2d** and **Supplementary Fig. 8c–e**). These results are consistent with previous observations that noncanonical target-miRNA base-pairing can also induce destabilization[1,14,28,29].

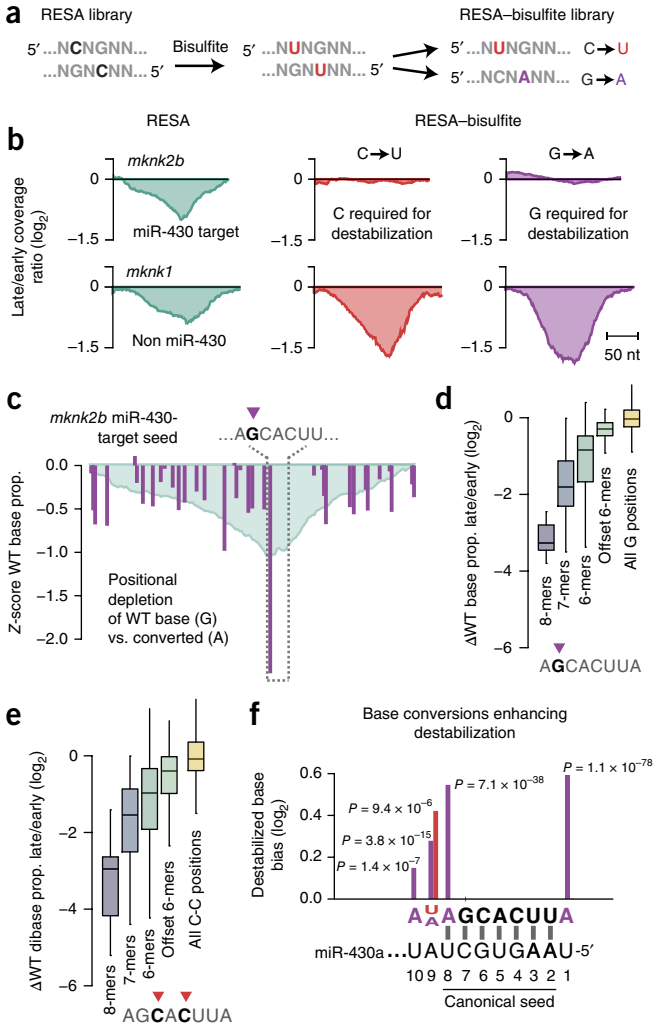

Figure 3 | RESA-B implements high-throughput mutational analysis of regulatory sequences. (**a**) RESA-B mutational strategy. (**b**) Coverage-ratio plots for reads with wild-type bases (left), C-to-U-converted bases (middle) and G-to-A-converted bases (right) for a region within *mknk2b* that requires C and G bases for destabilization (left) and a region within *mknk1* that does not require C or G bases for destabilization (right). (**c**) Positional analysis of G-to-A conversions over the *mknk2b* miR-430-target seed at the late stage, showing positions with G depletion (purple). The *y* axis shows the *Z* score of the ratio of the wild type (WT) base count to the mutated base count at each position, compared with all positions. The RESA signal is overlaid in blue. Prop., proportion. (**d**,**e**) Depletion of wild-type G bases versus those converted to A (**d**) and C-C dibases versus those converted to U-C, C-U and U-U (**e**) in late-stage versus early-stage embryos across all miR-430 targets, grouped on the basis of seed complementarity: 8-mers ($n = 13$), 7-mers ($n = 40$), 6-mers ($n = 136$) and offset 6-mers ($n = 115$). All G ($n = 99,157$) and C-C ($n = 92,696$) positions across the library are also shown. Boxes show medians and quartiles, and whiskers show 1.5× the interquartile range. Outliers are not shown. (**f**) Pooled base-conversion biases adjacent to miR-430-target seeds showing enrichment in wild-type bases relative to converted bases. The *y* axis shows the magnitude of change in the proportion of converted bases observed (U, red; A, purple) from the early to the late stage. The *x* axis is oriented sense to the target. *P* values from *G* tests of independence are shown.

We independently verified the regulatory activity of RESA-identified regions by generating reporters and assaying mRNA levels in embryos by qRT–PCR. We observed reporter destabilization consistent with RESA predictions (**Fig. 2e** and **Supplementary Fig. 6b**), thus indicating that these regions are sufficient to confer destabilization *in vivo*. To test the necessity of RESA-identified regions for endogenous destabilization, we used CRISPR–Cas9 (ref. 30) to generate deletions of the RESA-predicted destabilizing regions in *ptgfrn*, a gene that encodes prostaglandin F2 receptor inhibitor and is maternally provided and decays during the MZT (**Fig. 2f** and **Supplementary Fig. 6d**). Maternal-zygotic (MZ) mutants failed to clear *ptgfrn* mRNA (**Fig. 2f**), thus indicating that the sequences identified by RESA are required for mRNA destabilization *in vivo*. Together, these results demonstrate that RESA effectively identifies regulatory elements that confer differential mRNA stability.

RESA-B identifies nucleotide requirements in regulatory elements

Elucidating specific sequence requirements driving regulatory-element activity typically involves measuring the effects of mutations within the element. However, conventional targeted mutational analysis is a challenge in analysis of large numbers of elements. To address this challenge, we developed an extension to RESA called RESA-B, which introduces mutations across a RESA library to assay the effect of disrupting regulatory elements. Bisulfite treatment induces cytosine deamination at a proportion of sites. After PCR, each RESA insert encodes either C-to-T or G-to-A mutations in the forward orientation, at random positions (**Fig. 3a** and **Supplementary Fig. 9a**), thereby generating a highly complex library of mutated sequences that can be assayed for activity and compared with nonmutated libraries.

We created a RESA-B library from the same UTR fragments as in the original RESA experiment, injected it into one-cell embryos and sequenced RNA extracted at 2 h and 8 h. After read alignment with a procedure to accommodate mutated bases (**Supplementary Fig. 9**), we observed a 68% conversion rate of C and G positions (**Supplementary Fig. 10a**,**b**), thus suggesting that a large proportion of fragments would contain mutated, and hence potentially ineffective, regulatory elements. Indeed, many destabilized regions exhibited decreased activity in RESA-B, a result suggesting that they require C and G bases (**Fig. 3b** and **Supplementary Data 3**).

To identify individual C and G positions required for regulatory-element activity, we quantified the proportion of wild-type versus converted bases at each position. For discrete elements such as miR-430-target sites (GCACUU), single-base changes can affect function. In a mixed population of partially converted reporters, molecules encoding wild-type target sites would become destabilized and depleted, but those with converted sites, for example, ACACUU (with converted site underlined), would be more stable and would become over-represented (**Fig. 3c**). Indeed, G bases were significantly depleted in miR-430-target seeds over time, and this depletion was correlated with seed strength ($P = 2.1 \times 10^{-95}$,

Kruskal–Wallis test; **Fig. 3d**), thus demonstrating that RESA-B is able to detect selective stabilization of miR-430 targets mutated at single nucleotides. Similarly, the proportion of unconverted C-C dibases was significantly decreased in miR-430-target seeds ($P = 3.5 \times 10^{-63}$, Kruskal–Wallis test; **Fig. 3e**), thus reflecting their combined regulatory contribution (Online Methods). Across the entire library, we found 2,060 positions with significantly depleted wild-type bases from the early stage to the late stage (individual *G* tests of independence, false discovery rate (FDR) <0.05) (**Supplementary Fig. 10d**, and **Supplementary Data 4**). Together, our results indicate that RESA-B facilitates large-scale mutational analyses that can reveal regulatory-element base requirements.

RESA-B identifies base changes that enhance regulation

Mutations can also create sites conferring stronger regulation (**Supplementary Fig. 9a** and **Supplementary Data 4**). We found that offset 6-mer miR-430 targets adjacent to a wild-type C (denoted in lowercase bold, AGCACU**c**) were significantly less depleted than mutated sequences that converted the target into a stronger 7-mer (AGCACU**u**) ($P = 2.1 \times 10^{-20}$, *G* test of independence) (**Supplementary Fig. 10e**). We also identified base changes outside the miR-430 seed region that enhanced destabilization. Converted bases were strongly depleted at nonseed positions 9 and 10 (**Fig. 3f**), thus suggesting that A and U strengthen miR-430 targeting. We confirmed this effect in the original RESA library (**Supplementary Fig. 10f**) and, using reporters, we found that changing position 9 to an A in an endogenous target increased destabilization by 35% (**Supplementary Fig. 10g**), thus demonstrating additional targeting determinants for miRNA-mediated regulation.

RESA–CLIP maps RNA–protein interactions

We envision that RESA might be extended beyond analysis of mRNA stability by combining RESA libraries with various selection strategies (**Supplementary Fig. 11a**). To demonstrate one such application, we developed RESA–CLIP to map RNA–protein interactions (**Supplementary Fig. 11b**). mRNA regulation is often mediated by specific *trans*-acting RNA-binding proteins, whose target sequences can be identified through CLIP–seq. Present methods are labor intensive, require high sequencing depth and involve challenging downstream computational analyses (**Supplementary Fig. 11c**). We reasoned that RESA might be an ideal template for quickly identifying RNA–protein interactions *in vivo* because of two key advantages. First, because the RESA library is prebuilt with Illumina adaptors, RNA fragmentation and inefficient adaptor ligation are not necessary to generate the sequencing library. Second, the results can be analyzed with existing CLIP–seq pipelines or with the RESA framework.

To perform RESA–CLIP, we co-injected the RESA library with Flag-tagged Ago2 to assay binding to miR-430 targets (**Supplementary Fig. 11b**). This entire procedure, from injection to sequencing library, requires only 10 h, in contrast to other CLIP–seq protocols that require several days (**Supplementary Fig. 11c**). We observed an enrichment in Ago2 CLIP fragments overlapping miR-430-target sequences, corresponding to the strength of seed complementarity (**Fig. 4a**,**b** and **Supplementary Fig. 12a**).

At high stringency, 94% of CLIP peaks overlapped with RESA-predicted miR-430-dependent destabilized regions, whereas at lower stringency, peaks were still highly specific for miR-430-regulated regions, and we observed higher CLIP enrichment for stronger seed matches ($P = 8.5 \times 10^{-37}$, Kruskal–Wallis test) (**Fig. 4c**, **Supplementary Fig. 12c–f** and **Supplementary Data 5**). Notably, even RESA-predicted miR-430-dependent regions lacking complementary target seeds were significantly enriched in Ago2 binding over non-miR-430-target regions ($P = 3.6 \times 10^{-5}$, Wilcoxon rank-sum test) (**Supplementary Fig. 12g**), thus demonstrating that these were likely bona fide miR-430 targets. Among the CLIP peaks not overlapping RESA-predicted destabilized regions, we did not find any significant sequence motifs indicative of other miRNAs; however, we did find extensive partial complementarity to miR-430 (**Supplementary Fig. 12e**), thus suggesting that the Ago2 CLIP indeed reflects the activity of only miR-430.

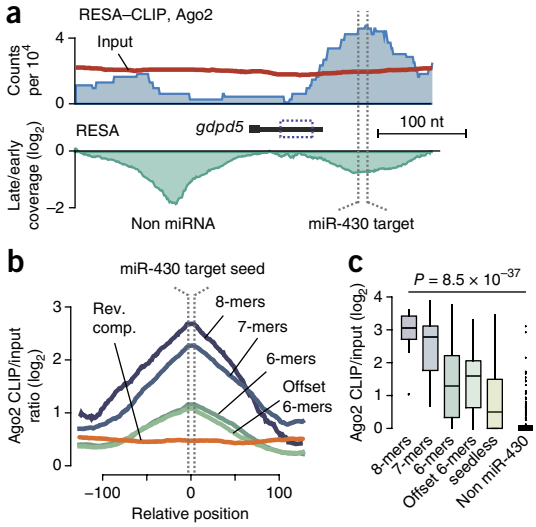

Figure 4 | The RESA framework can be adapted to map RNA–protein interactions and other RNA biology *in vivo*. (**a**) RESA coverage ratio for the *gdpd5a* locus (bottom) and RESA Ago2 CLIP (blue region) and input library (red line) counts (top). The RESA-verified miR-430 6-mer target site is marked with dashed lines. (**b**) Metaprofiles of Ago2 CLIP versus input ratios centered on the miR-430-target seed sequence, grouped on the basis of seed complementarity and compared with reverse-complement (rev. comp.) inserts overlapping 8-, 7- and 6-mers. (**c**) Box plots showing Ago2 CLIP enrichment over input across miR-430-dependent RESA destabilized regions containing 8-mer target sites ($n = 11$), 7-mers ($n = 33$), 6-mers ($n = 74$), offset 6-mers ($n = 24$) and targets lacking a canonical target seed ('seedless') ($n = 61$) compared with non-miR-430-dependent regions ($n = 178$). The *P* value from a Kruskal–Wallis test comparing all groups is shown. Boxes show median and quartiles; whiskers show 1.5× the interquartile range; points are outliers.

DISCUSSION

Sequence-selection methods (SELEX)[31–33] have made it possible to determine sequence binding specificity *in vitro*. *In vivo* adaptations have relied on custom-made oligonucleotide libraries[18–22], which can suffer from limited sequence complexity and resolution. RESA provides a versatile and quantitative platform to identify RNA-regulatory activity *in vivo*, by using a direct PCR assembly approach that maximizes the molecular complexity of the query sequence pool. *In vitro* library transcription eliminates the need to normalize RNA levels to plasmid DNA quantities, which can substantially vary among cells. RESA enables the

analysis of regulatory-sequence potential in diverse organisms amenable to introduction of exogenous RNA, including fish, frog, fly, mouse and nematode, and it can also be adapted to cell culture with RNA transfection methods. RESA also streamlines sequencing-library preparation by building in Illumina adapters, thereby avoiding inefficient adaptor ligation and allowing for robust library recovery. These innovations yield high sensitivity with broad applicability and enable transcriptome-wide surveys of regulatory-sequence activity.

We also demonstrated that RESA–CLIP can interrogate RNA–protein interactions without the cost and labor associated with other pulldown protocols; however, because RESA–CLIP assesses protein binding with exogenously supplied sequence fragments, detection is limited to elements that maintain activity outside of their native context. Despite this caveat, RESA–CLIP can provide a simple and rapid strategy to identify sequences that underlie the binding specificities of many known RNA-binding proteins.

RESA applies a selection principle that is based on the differential accumulation of specific sequences to identify regulated regions and can easily be applied to interrogate diverse levels of post-transcriptional control (**Supplementary Fig. 11a**). For example, RESA can improve upon the identification of mRNA-localization elements[34,35], in combination with existing methods to isolate specific embryonic compartments, cell types or subcellular structures. Similarly, polysome fractionation can distinguish strongly and weakly translated mRNAs and allow for high-throughput identification of sequences directing translation activation or repression. Finally, RESA can be adapted to investigate how different codons[36–38] or regulatory elements within coding regions[39,40] affect mRNA stability and translation. We envision widespread usage of RESA variations to decode post-transcriptional mechanisms of gene expression across biological settings.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank K. Bilguvar, S. Mane, C. Castaldi and I. Tikhonova for sequencing support; H. Codore for technical assistance; A. Bazzini, P. Oikonomou and S. Tavazoie for discussions; and all the members of the Giraldez laboratory for intellectual and technical support. This research was supported by the US National Institutes of Health (R01 HD074078, GM103789, GM102251, GM101108 and GM081602), the Pew Scholars Program in the Biomedical Sciences, the March of Dimes (1-FY12-230), the Yale Scholars Program, the HHMI Faculty Scholars Program, Whitman fellowship funds provided by E.E. Just, Lucy B. Lemann, and Evelyn and Melvin Spiegel, and the H. Keffer Hartline and Edward F. MacNichol, Jr. Fellowship Fund of the Marine Biological Laboratory (Woods Hole, Massachusetts, USA) to A.J.G.; the Swiss National Science Foundation (grant P2GEP3_148600) to C.E.V.; NIH Fellowship F32 HD061194 to C.M.T.; NIH Fellowship F32 HD071697 and start-up funds from the University of Pittsburgh to M.T.L.; and NIH Training Grants T32 GM007223 and T32 HD007149, an Edward L. Tatum Fellowship (Yale University) and the Yale MRSP to V.Y.

AUTHOR CONTRIBUTIONS

V.Y., C.M.T. and A.J.G. designed and conceived the project. V.Y. generated the RESA UTR, RESA-B and RESA–CLIP libraries, and performed the validation experiments. V.Y. and M.T.L. developed the RESA and RESA-B analysis. M.T.L. and C.E.V. performed the computational analyses. All authors interpreted and analyzed the data. V.Y., M.T.L. and A.J.G. wrote the manuscript with input from the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Reinhart, B.J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
2. Meola, N., Gennarino, V.A. & Banfi, S. microRNAs and genetic diseases. *PathoGenetics* **2**, 7 (2009).
3. Giraldez, A.J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).
4. Lai, E.C. & Posakony, J.W. The Bearded box, a novel 3′ UTR sequence motif, mediates negative post-transcriptional regulation of Bearded and Enhancer of split Complex gene expression. *Development* **124**, 4847–4856 (1997).
5. Treisman, R. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-fos 3′ sequences. *Cell* **42**, 889–902 (1985).
6. Piqué, M., López, J.M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).
7. Mowry, K.L. & Melton, D.A. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science* **255**, 991–994 (1992).
8. Zubiaga, A.M., Belasco, J.G. & Greenberg, M.E. The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Mol. Cell. Biol.* **15**, 2219–2230 (1995).
9. Giraldez, A.J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
10. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
11. Nielsen, C.B. *et al.* Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).
12. Lai, E.C. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363–364 (2002).
13. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
14. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
15. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460**, 479–486 (2009).
16. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
17. Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F. & Struhl, K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**, 812–824 (2014).
18. Oikonomou, P., Goodarzi, H. & Tavazoie, S. Systematic identification of regulatory elements in conserved 3′ UTRs of human transcripts. *Cell Rep.* **7**, 281–292 (2014).
19. Rosenberg, A.B., Patwardhan, R.P., Shendure, J. & Seelig, G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698–711 (2015).
20. Weingarten-Gabbay, S. *et al.* Comparative genetics: systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*. **351**, aad4939 (2016).
21. Zhao, W. *et al.* Massively parallel functional annotation of 3′ untranslated regions. *Nat. Biotechnol.* **32**, 387–391 (2014).
22. Wissink, E.M., Fogarty, E.A. & Grimson, A. High-throughput discovery of post-transcriptional *cis*-regulatory elements. *BMC Genomics* **17**, 177 (2016).
23. Wolter, J.M., Kotagama, K., Babb, C.S. & Mangone, M. Detection of miRNA targets in high-throughput using the 3′LIFE assay. *J. Vis. Exp.* e52647 (2015).
24. Yartseva, V. & Giraldez, A.J. The maternal-to-zygotic transition during vertebrate development: a model for reprogramming. *Curr. Top. Dev. Biol.* **113**, 191–232 (2015).

25. Lee, M.T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
26. Bazzini, A.A., Lee, M.T. & Giraldez, A.J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
27. Staton, A.A., Knaut, H. & Giraldez, A.J. Reply to: "On the robustness of germ cell migration and microRNA-mediated regulation of chemokine signaling". *Nat. Genet.* **45**, 1266–1267 (2013).
28. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
29. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. & Slack, F.J. The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev.* **18**, 132–137 (2004).
30. Moreno-Mateos, M.A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *Nat. Methods* **12**, 982–988 (2015).
31. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
32. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
33. Blackwell, T.K. & Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104–1110 (1990).
34. Martin, K.C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
35. Rabani, M., Kertesz, M. & Segal, E. Computational prediction of RNA structural motifs involved in post-transcriptional regulatory processes. *Methods Mol. Biol.* **714**, 467–479 (2011).
36. Bazzini, A.A. *et al.* Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).
37. Mishima, Y. & Tomari, Y. Codon usage and 3′ UTR length determine maternal mRNA stability in zebrafish. *Mol. Cell* **61**, 874–885 (2016).
38. Presnyak, V. *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
39. Davis, C.A., Monnier, J.M. & Nick, H.S. A coding region determinant of instability regulates levels of manganese superoxide dismutase mRNA. *J. Biol. Chem.* **276**, 37317–37326 (2001).
40. Wellington, C.L., Greenberg, M.E. & Belasco, J.G. The destabilizing elements in the coding region of c-fos mRNA are recognized as RNA. *Mol. Cell. Biol.* **13**, 5034–5042 (1993).

ONLINE METHODS

RESA reporter library construction. 3′-UTR regions of approximately 850- to 1,300-bp fragments were PCR-amplified from 2-hpf-embryo cDNA with Taq polymerase for 36 cycles and an annealing temperature of 55 °C. Gene definitions and corresponding gene symbols were obtained from Ensembl r79 on the zebrafish Zv9 genome (in figures, *gdpd5a* is *gdpd5*, and *sgk3* is C24H8orf44). Primers were designed with Primer3 (ref. 41) and are listed in (**Supplementary Data 1**). Half of each PCR reaction was pooled and purified on PCR cleanup columns (Qiagen, cat. no. 28104). To generate the PCR library, approximately 4 µg DNA (50 µL) was sheared with a Covaris sonicator at full power for 20 min, ends were repaired, and Illumina adapters were ligated. After adaptor ligation, fragments were selected on the basis of sizes ranging from 130 to 250 bp to exclude adaptor-dimer species.

The reporter library was generated by overlap-extension PCR with Phusion High Fidelity DNA polymerase (NEB, cat. no. M0530) for 5 cycles of 98 °C, 61 °C and 72 °C for 30 s each. Then 24 reactions with 2 µL of elongated product were amplified for 10 cycles, with annealing at 57 °C, with a forward primer of GCTTGATTTAGGTGACACTATA and a reverse primer of GAATTAAAAAACCTCCCACACC, mapping to the SP6 promoter and downstream of the SV40 polyadenylation signal. The reporter library template of 1,270 bp was purified on 1% agarose, and 400 ng was used for *in vitro* RNA synthesis with an Sp6 mMessage mMachine kit (Thermo Fisher, cat. no. AM1340). Zebrafish embryos were injected with 10 pg mRNA library and collected at the indicated stages.

miR-430 inhibition. To inhibit miR-430 function, zebrafish embryos were injected with 1 nL of 10 µM tiny locked nucleic acid complementary to the miR-430 seed (430LNA), 5′-TAGCACT T-3′ (Exiqon), as described in ref. 27.

Sequencing-library construction. Total RNA was extracted from 30–75 zebrafish embryos injected with the reporter library by using TRIzol reagent according to the manufacturer instructions. Poly(A)$^+$ mRNA was selected with NEB magnetic beads (NEB, cat.no. S1419S). Briefly, 250 µg or 50 µL of beads was washed with 1× wash buffer (0.5 M NaCl, 20 mM Tris-HCl, pH 7.5, and 1 mM EDTA) for each sample. 3 µg total RNA from each sample was diluted in 25 µL water, incubated at 65 °C for 2 min to denature the secondary structure and then placed on ice. 25 µL of 2× wash/binding buffer (1 M NaCl, 40 mM Tris-HCl, pH 7.5, and 2 mM EDTA) was added to each RNA sample, mixed with precleared beads, vortexed gently and incubated at 40 °C for 10 min with occasional agitation. After a 10-min binding reaction, the supernatants were discarded, and the beads were washed twice with 100 µL 1× wash buffer. A third wash was performed with 100 µL ice-cold low-salt buffer (0.15 M NaCl, 20 mM Tris-HCl, pH 7.5, and 1 mM EDTA). RNA was eluted with 25 µL elution buffer (10 mM Tris-HCl, pH 7.5, and 1 mM EDTA) prewarmed to 65 °C for 2 min. A second elution was performed under the same conditions. Poly(A)-selected RNA was recovered with ethanol precipitation and resuspended in 15 µL of water. 6 µL of RNA was reverse-transcribed with a reporter-specific primer (CAT CAATGTATCTTATCATGTCTGGATC) with SuperScript III. The sequencing library was prepared with 16 cycles of Fusion PCR with the forward primer AATGATACGGCGACCACCGA GATCTACACTCTTTCCCTACACGACGCTC and the reverse primer CAAGCAGAAGACGGCATACGAGAT(barcode)GTGA CTGGAGTTCAGACGTGTGCTCTTCCGATCT.

Computational and statistical analysis. All analyses were performed with custom scripts written in R and Python 3, except as noted below. All box plots show median, lower and upper quartiles and whiskers of 1.5× the interquartile range; individual points represent outliers beyond the whiskers.

The following statistical tests were performed: **Figure 2e**, two-sided Welch's *t* test on $n = 3$ independent sets of embryos ($t = 6.0454$, $P = 0.015$ for *sdr16c5b*; $t = 11.7406$, $P = 0.0035$ for *shroom4*; $t = -2.7369$, $P = 0.053$ for *rap2c*). **Figure 3d**,**e**, Kruskal–Wallis rank-sum test (statistic = 446.8683, $P = 2.1 \times 10^{-95}$; statistic = 297.6536, $P = 3.5 \times 10^{-63}$). **Figure 3f**, individual *G* tests for independence; results reported in Source Data. **Figure 4c**, Kruskal–Wallis rank-sum test (statistic = 179.0614, $P = 8.5 \times 10^{-37}$) and two-sided Wilcoxon rank-sum test comparing seedless and non-miR-430 targets ($W = 7{,}388.5$, $P = 3.6 \times 10^{-7}$). **Supplementary Figure 2b**,**c**, Kruskal–Wallis rank-sum tests comparing targets with background (statistic = 186.0093, $P = 3.8 \times 10^{-39}$ for late/early; statistic = 209.2651, $P = 3.8 \times 10^{-44}$ for late/430LNA); two-sided Wilcoxon signed-rank tests ($V = 2{,}908.5$, $P = 7.1 \times 10^{-40}$ for late/early; $V = 1{,}604.5$, $P = 9.4 \times 10^{-45}$ for late/430LNA). **Supplementary Figure 8c**, chi-squared test (statistic = 52.4877, df = 1, $P = 4.3 \times 10^{-13}$). **Supplementary Figure 8d**, two-sided Wilcoxon rank-sum test ($W = 19{,}897$, $P = 7.3 \times 10^{-5}$). **Supplementary Figure 10e**, *G* tests of independence ($G = 85.7295$, df = 1, $P = 2.1 \times 10^{-20}$ for C bias; $G = 33.6585$, df = 1, $P = 6.6 \times 10^{-9}$ for G bias). **Supplementary Figure 10f**, two-sided Wilcoxon rank-sum tests ($W = 2{,}613.5$, $P = 0.0017$ for offset 6-mers; $W = 2{,}158.5$, $P = 0.0025$ for 6-mers; $W = 333$, $P = 0.0038$ for 7- and 8-mers). **Supplementary Figure 10g**, two-sided Welch's *t* tests on $n = 3$ independent sets of embryos ($t = 4.1776$, $P = 0.017$ for *pam* A9 late versus early; $t = 13.9267$, $P = 0.00066$ for *pam* G9 late versus early; $t = 6.9051$, $P = 0.0054$ for *itpr3* A9 late versus early; $t = 35.9421$, $P = 4.2 \times 10^{-6}$ for *itpr3* G9 late versus early; $t = -5.2482$, $P = 0.027$ for *pam* A9 late versus G9 late; $t = -6.5417$, $P = 0.019$ for *itpr3* A9 late versus G9 late). **Supplementary Figure 12e**, chi-squared test (statistic = 18.1649, df = 1, $P = 2.0 \times 10^{-5}$) and two-sided Wilcoxon rank-sum test ($W = 18{,}745$, $P = 2.4 \times 10^{-8}$). **Supplementary Figure 12f**, two-sided Wilcoxon rank-sum test ($W = 723$, $P = 6.1 \times 10^{-8}$). **Supplementary Figure 12g**, chi-squared test (statistic = 17.0887, df = 1, $P = 3.6 \times 10^{-5}$).

Sequencing-read processing and alignment. Paired-end sequencing reads were preprocessed to trim any Illumina sequencing adapters from the 3′ ends[42], then aligned to the zebrafish Zv9 genome with TopHat v2.0.13 (ref. 43), allowing only alignments consistent with the gene isoforms used to design the amplicons (**Supplementary Data 2**) by providing amplicon-specific transcriptome GTF definitions. The TopHat parameters used were --no-discordant --no-mixed --library-type fr-firststrand --transcriptome-only --read-mismatches 7 --read-gap-length 7 --read-edit-dist 7 --segment-mismatches 3 --b2-very-sensitive.

Before alignment, the genome sequence was modified to maximize concordance with RESA library design: (i) all base positions outside of the amplicon regions were masked with '*n*', and

doi:10.1038/nmeth.4121

(ii) because certain primer pairs used to amplify the library encode nontemplate restriction-site overhangs, the genome sequence was changed to match the overhanging primer sequence for each such primer, to ensure maximal read alignment at the ends of these amplicons.

An alternative to this approach would be to simply extract the spliced transcript sequences for each UTR, exclusively align reads to a transcriptome index, then back-convert the transcriptome-relative coordinates to genome coordinates for downstream analysis.

Calculating read coverage. Positional sequencing coverage (i.e., pileup) was calculated individually for each UTR in pooled biological replicates, such that each nucleotide position encoded the total number of sequence fragments that overlapped it. In detail: first, the samtools version 1.2 (ref. 44) view command was used to obtain all sequencing reads overlapping the genomic coordinates of a UTR. Each read pair was merged into a single sequencing-read fragment spanning the start and end coordinates of the 5′ and 3′ read, respectively. Only read fragments with end-to-end lengths between 56 and 129 nt, inclusive, were used, on the basis of the observed distribution of fragment lengths (**Supplementary Fig. 1c**,**d**), to avoid the effects of aberrant read alignments that would artificially produce shorter or longer genomic spans. Genomic coordinates were converted to UTR-relative coordinates. Read coverage was incremented by one at each base position between the start and end coordinates of the fragment, inclusive, only if the strand of the read fragment (defined by the strand of read 1) was sense relative to the UTR. X coverage, relative to an experimental condition, is the raw count at a given nucleotide position. UTRs with 0 median sequencing coverage in the pooled early stage samples were considered to be failed PCRs and were excluded from further analysis. For antisense coverage (**Supplementary Fig. 1a**), only read fragments in the opposite-strand orientation of the UTR were counted.

The $\log_2$ ratio of normalized coverage between two experimental conditions was used to identify position-specific differential regulation. Raw counts in each experiment were normalized by dividing by the total count in that experiment (sum of all counts in all positions across all UTRs). A smoothing factor (0.5) was added to each count to prevent division by zero. Finally, the $\log_2$ ratio was calculated.

Analysis of miR-430-target sites. Putative miR-430-target sites were identified by using the Zv9 genome sequence. UTR sequence matches to AGCACTTA were considered to be 8-mers, those to AGCACTT and GCACTTA were considered to be 7-mers, those to GCACTT were considered to be 6-mers, and those to AGCACT were considered to be offset 6-mers. Seed mismatch sites were considered to be those with any 1-nt difference from the GCACTT 6-mer. 3′ complementarity was measured through Smith–Waterman local alignment to the reverse complement of each mature miR-430 sequence excluding the seed (430a, CTACCCCAACAAAT; 430b, CTACCCCAACTTAAT; 430c, CTACCCCAAAGAGA). Parameters were match = 1, mismatch = 0, gap = −100. The complementarity score was the number of dinucleotide matches in the best local alignment. Sequences used to estimate background frequencies were obtained from the UTR regions in the library not predicted to be regulated: for each such segment ≥65 nt, the middle 50 nt were used. For all population and meta-analyses, only miR-430-target sites with at least 100× sequencing coverage in the pooled early-stage libraries were used. Meta-analyses additionally excluded target sites less than 91 nt from each other, to avoid measuring the overlapping influences of two independent targets; and excluded target sites too close to the ends of the UTRs, depending on the size of the region window being queried, to avoid artifacts due to truncation. To detect polymorphisms at miR-430 seeds in the libraries compared with the reference genome, samtools[5] mpileup was run on the early-stage alignments followed by bcftools and vcftools commands to obtain the consensus sequence from the sequencing reads (samtools mpileup -uf Zv9.fa <bamfiles> | bcftools call -c | vcfutils.pl vcf2fq > utrs.fq).

Regulated-region identification. Destabilized and stabilized regions were identified separately. For destabilized regions, positions at local minima of $\log_2$ coverage ratio were identified in each UTR as sites of destabilization. Broader regulated regions were defined as the minimal window around a local minimum such that each position in the window had a $\log_2$ coverage ratio within 0.25 times the local minimum. Stabilized regions were identified analogously by using local maxima. In detail, with a sliding window of 100 nt across a UTR, the position of the minimal $\log_2$ coverage ratio relative to all other positions in the window was recorded (local minimum). If this minimum value occurred at the last position of the window, thus suggesting that an even lower value would occur downstream, no local minimum was called, and instead the window was shifted to the right by half a window length to continue searching. For each local minimum identified, the boundary of the regulated region was extended to the left and right, such that each position encountered had a $\log_2$ coverage ratio <0.25 times the local minimum. Regulated regions that overlapped were merged.

From this list of candidate destabilized regions, *post hoc* criteria were applied to classify the region as high confidence (width 65 nt, coverage at early stage 100× at the minimum, value $-\log_2(4/3)$ at the minimum) or low confidence (width 40 nt, value $-\log_2(5/4)$ at the minimum, no threshold for coverage). These parameters were selected on the basis of the width of miR-430-associated regions (**Supplementary Fig. 4a**) and FDR simulations (described below).

For each region, the subregion most likely to contain the responsive sequence element was defined to be bounded by the leftmost and rightmost positions, with at least 0.9 times the local minimum (or maximum). For the purpose of motif analysis, this subregion was extended 15 nt left and right, without exceeding the boundaries of the regulated region.

Regulatory-element width analysis. To identify the determinant of the RESA region width, we reasoned that this property was the combined effect of regulatory-element length requirements and sequence-fragment length. The characteristic 'valleys' centered on miR-430-target seed sequences spanned 150–200 nt, approximately twice the insert-fragment length obtained during RESA library construction plus the 6–8 nt of miR-430 complementarity. To test the effects of sequence-fragment length, aligned read pairs were restricted to short (56–80 nt) or long (105–129 nt) length classes on the basis of the end-to-end distance of each read pair, as reported

by TopHat. RESA regulated regions were identified by using both sets of reads individually and were compared with RESA analysis using all reads (56–129 nt) (**Supplementary Fig. 4**).

False-discovery-rate estimation. To simulate the effects of sequencing coverage on the FDR, the *acat1* UTR was selected to perform read sampling (**Supplementary Fig. 3**). At full coverage, *acat1* encodes a single well-defined miR-430-target site. For each simulated coverage level (for example, 100× or 40×), a random subset of the overlapping reads was obtained for both the early stage and the late stage. The $\log_2$ coverage ratio was calculated, and destabilized regions were predicted, with no threshold for size or magnitude. Among the regions (if any) not corresponding to the known miR-430-target site, the minimal $\log_2$ coverage ratio was recorded as the false-positive destabilization magnitude. This process was repeated 500 times per coverage depth, thus inducing a distribution of false-positive coverage ratios. At that coverage depth, an FDR of 0.1 corresponded to setting a threshold equal to the tenth percentile of the false-positive distribution.

Motif analysis. To identify overrepresented sequences in the regulated regions, the MEME Web tool v4.11.2 (ref. 45) was used in discriminative mode. The control sequences consisted of all UTR regions in the library not predicted to be regulated (UTR regions minus regulated regions, with any segments less than 25 nt long discarded). Search was limited to the sense strand, allowing a motif width from 6 to 12 nt.

RESA–bisulfite (RESA-B). To preserve the integrity of sequencing adaptors, 1 ng RESA library was amplified for 10 cycles with primers containing methylated cytosines, a modification that disfavors cytosine deamination after bisulfite exposure. The library with hemimethylated adapters (1.4 µg) was deaminated with an EZ DNA Methylation-Direct kit (Zymo Research, cat. no. D5020) according to the manufacturer's directions. The only modification to the protocol was that the deamination reaction was performed at 98 °C for 8 min followed by 64 °C for 25 min to decrease the deamination rate to ~50%, according to a pilot assay of individual sequences with Topo cloning. The deaminated library was amplified for 7 cycles at a 68 °C annealing temperature with nonmethylated primers to fix mutations. A pool of 12 PCR reactions was ethanol-precipitated and separated on 8% PAGE, and a 180- to 220-bp region was purified. The mutated product was built into a GFP reporter library and injected into zebrafish as described above.

Analysis of bisulfite-treated libraries. The alignment strategy is outlined in **Supplementary Figure 9**. Read pairs were first subjected to wild-type genome alignment to filter out fragments that could not be unambiguously classified as C-to-T or G-to-A converted. Read pairs unaligned during this filtering step were then subjected to each of the four mutated alignment protocols in succession, with the expectation that only one combination of read conversion and genome conversion would yield successful alignment. For each UTR, both C-to-T- and G-to-A-converted alignments were expected to exist and were analyzed separately.

Wild-type and converted base counts were enumerated at each position in the library where the reference genome sequence encoded either C or G. Base identities not conforming to the expected wild-type or converted base at a position were not counted and may have been due to sequencing errors or polymorphisms.

Conversion biases in a single condition were identified by (i) calculating a Z score of WT base proportion at each informative position for an experimental condition (for example, late stage) (separate mean and s.d. were used for C and G bases) and (ii) calculating a G test of independence for the number of WT and converted bases in the experimental condition versus the control condition. A Benjamini–Hochberg control for FDR was applied to the resulting P values. Positions with significant WT base depletion biases were defined to have $Z < -1$ and an adjusted G test P value < 0.05. For positions with significant WT base enrichment, a $Z > 1$ threshold was used.

Some bases, owing to their proximity, were effectively 'linked' in their contribution to regulatory activity, for example, for the two Cs in the miR-430 seed (GCACUU), both GUACUU and GCAUUU should exhibit diminished destabilization efficacy despite encoding one unconverted C. For linked bases, consecutive (but not necessarily adjacent) C or G positions were considered. Counts for fully wild-type (C-C or G-G for dibases; C-C-C or G-G-G for tribases) versus mutated (at any one of the positions) alleles were enumerated across the library, then analyzed as above.

Meta-analyses for destabilization-enhancing base conversions were performed by using only the subset of sequencing reads encoding wild-type miR-430 motifs of interest, then counting base conversions at adjacent positions, when relevant (i.e., the reference sequence encoded a C or G base at that position). Base counts at the same relative positions were pooled across all instances of the motif in the library, and a G test of independence was performed for each position comparing late and early stages, for each of C-to-T and G-to-A conversions.

RESA–CLIP procedure. 800 zebrafish embryos were injected with 80 pg RESA library and 80 pg Flag-tagged Ago2, cross-linked with 254 nm UV at 30% epiboly stage (3.3 hfp) for 4 min on ice, lysed in 750 µL lysis buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitor (Roche, cat. no. 04693159001; 1 tablet/10 mL), 1:1,000 volume of RNase inhibitor (Thermo Fisher, cat. no. AM2696)), immunoprecipitated with 100 µL anti-Flag M2 magnetic beads (Sigma, cat. no. M8823) for 2 h at 4 °C and washed five times with lysis buffer without protease inhibitor. To recover bound mRNAs, beads were incubated with 200 µL proteinase K buffer (100 mM Tris-HCl, pH 7.4, 50 mM NaCl and 10 mM EDTA) with 10 µL proteinase K (Thermo Fisher) for 20 min at 37 °C with shaking at 1,100 r.p.m. RNA was extracted with TRIzol reagent. The entire RNA sample was reverse transcribed. Sequencing libraries were built as described above by using 8 PCR cycles for the input sample and 21 cycles for Flag IP.

Read alignment was performed as above for CLIP and input. To avoid PCR artifacts, reads with identical start positions were counted at most twice. However, reads starting at the first or last position in each UTR were not subjected to this restriction. The $\log_2$ CLIP/input ratio was calculated as above, but all $\log_2$ ratios < 0 were set to 0.

doi:10.1038/nmeth.4121

Genetic deletion of 3′-UTR regions. CRISPR-mediated mutagenesis was performed as described in ref. 30. Genotyping oligos used to amplify wild-type and mutant alleles were forward, AAAGCTGGAGGTTTGCAGAG and reverse, CAAACT GAGGGACCTGGAGA. *In situ* hybridization was performed as in ref. 46 with 20 ng of DIG-labeled RNA probe per 200-µL hybridization reaction. RNA probes were synthesized from PCR products with a T7 promoter on the reverse-orientation oligo: forward, CTGGGTGTCTCCTGGGTTTA; reverse, AGGCTGA GGGTGAAGCTGTA.

Reporter validation with qRT–PCR. All validation reporter constructs were generated with overlap-extension PCR by combining templates encoding the SP6 promoter, GFP, an SV40 polyadenylation signal and the tested insert with 15-nt overhangs complementary to GFP on the 5′ end and SV40 on the 3′ end. RNA was *in vitro* synthesized with an Sp6 mMessage mMachine kit (Thermo Fisher, cat. no. AM1340).

Zebrafish embryos were injected with 4 pg of each reporter and dsRed control mRNAs and collected at 2 hpf and 8 hpf. Total RNA (250 ng) was reverse-transcribed with a SuperScript III kit (Invitrogen, cat. no. 18080-051) with random hexamers. cDNA was diluted 1:20 and used in 10-µL reactions (5 µL SYBR Green master mix (Thermo Fisher, cat. no. 4472908), 0.5 µL 10 µM forward and reverse primer mix, 1 µL 1:20-diluted cDNA and 3.5 µL water). All GFP reporters used a common forward primer (CATGGTCCTGCTGGAGTTCGTGAC) and a reporter-specific reverse primer (*sdr16c5b*, CTAACTGTGTCT CTCTTCTTCACAG; *shroom4*, CCTCTCCCATCTGAACCAAC; *rap2c*, CACTGCTCTTGCTTCTCGGCTCG; *pam*, CGGTCGA AGCCATGTCG; and *itpr3*, TTGAGGGTGTGTAGATCCAA). dsRed was amplified with the forward primer GAAGGGCGAG ATCCACAAG and reverse primer GGACTTGAACTCCACCA GGTA. Biological and technical triplicates were performed for each sample, and relative expression with the $\Delta\Delta C_T$ method was measured with ViiA 7 software v1.2.2, with dsRed mRNA as the reference control.

Zebrafish maintenance. Zebrafish embryos were obtained through natural mating of the TU/AB strain of mixed ages (5–18 months). Mating pairs were selected randomly from a set of fish previously allocated to be used for the week(s) when embryos were collected. The number of mating pairs was determined according to the total number of embryos required for analysis on that day, given approximately 100 embryos per mating pair. Embryos from multiple pairs were pooled before random allocation into experimental groups. No experimenter blinding was used. Fish were maintained in accordance with the research guidelines of the International Association for Assessment and Accreditation of Laboratory Animal Care, under a protocol approved by the Yale University Institutional Animal Care and Use Committee (IACUC).

Code availability. Source code and executables for RESA are available at https://github.com/MTLeeLab/RESA.

Data availability statement. Raw reads are publicly accessible in the Sequence Read Archive under accession code SRP090954. Source data for **Figures 1**–4 is available online.

41. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
42. Jiang, H., Lei, R., Ding, S.W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
43. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
46. Thisse, C. & Thisse, B. High-resolution *in situ* hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008).