



Supporting Online Material for

Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs

Antonio J. Giraldez,* Yuichiro Mishima† Jason Rihel, Russell J. Grocock, Stijn Van Dongen, Kunio Inoue, Anton J. Enright and Alexander F. Schier*

*To whom correspondence should be addressed. E-mail: schier@fas.harvard.edu, giraldez@mcb.harvard.edu

Published 16 February 2006 on *Science Express*
DOI: 10.1126/science.1122689

This PDF file includes:

Materials and Methods
Figs. S1 to S12
References

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/1122689/DC1)

Table-S1(1.5fold).zip
SOM-Targets-tested.pdf

SUPPORTING ONLINE MATERIAL

Giraldez et al.

Figure legends

Fig. S1. Time course analysis of miR-430 expression

(A) Northern blot to visualize miR-430 expression between 0 hpf and 4 hpf. Note that precursor accumulation occurs by 2.5 hpf, coinciding with the onset of zygotic transcription. The mature miR-430 is present between 3 and 4 hpf. (B) Northern blot to compare miR-430 expression at 4 hpf with the minimum amount of miR-430 duplex required to rescue the gastrulation and the brain morphogenesis defects in *MZdicer* mutants. (C, D, E) Lateral view of whole mount embryos labeled with a miR-430a probe at 5, 10 and 30 hpf. DIG labeled Lock Nucleic Acids (LNA) oligonucleotides were used to visualize mir-430a expression using in situ hybridization as described in (1). Note the ubiquitous expression of miR-430a.

Fig. S2. miRNAs accelerate target mRNA decay

(A) Northern blot to visualize miR-1 GFP reporter mRNAs, containing either 3 perfect targets (3xPT) or 3 imperfect targets (3xIPT) in the presence or the absence of miR-1 or miR-204 injection. By ten hours after injection, both perfect and imperfect targets decay faster in the presence of the cognate miR-1. A cleavage product was observed specifically in embryos that were co-injected with the miRNA duplex and the perfectly complementary target (shown in fig. S3C in (2)). (B, C) Quantification of the Northern blot shown in (A) and in Fig. 1B respectively. In both cases the cognate miRNA accelerates the decay of the target mRNA.

Fig. S3. Results of microarray experiments and target site predictions

(A) Experimental approach to identify miR-430 targets and graphic representation of mRNA sets whose levels were increased ≥ 1.5 fold in *MZdicer* mutants compared to *MZdicer*^{+miR-430} (light green), wild type (dark green) or both (red). (B) Venn diagram representing the different groups of genes present in the microarray. The gray circle represents the mRNAs in the array that had an experimentally validated 3' UTR (5219, control set*). The darker area represents the fraction of mRNAs that have a 6mer (or more) sequence complementary to the miR-430 seed. The green circle represents mRNAs upregulated ≥ 1.5 -fold in *MZdicer* mutants compared to wild-type embryos (902). The light green circle represents mRNAs upregulated ≥ 1.5 -fold in *MZdicer* compared to *MZdicer*^{+miR-430} (385). The overlap of these two groups is shown in red (328), and the fraction that has sequences complementary to the miR-430 seed is called the putative miR-430 target set (203). Owing to geometric constraints, an area of the red circle representing 38 mRNAs that lack miR-430 target sites and did not overlap with the green group is not pictured. (C) Number of genes with a miR-430 target site in their 3' UTRs in the different groups in the array. mRNAs upregulated in *MZdicer* compared to wild-type and *MZdicer*^{+miR-430} are shown in red, orange and violet depending on the level of upregulation. mRNAs unchanged in all three conditions are shown in blue, and mRNAs of the control set are shown in gray. (D) Number of mRNAs increased more than 1.5-fold in *MZdicer* mutants compared to *MZdicer*^{+mir-430}, wild type, or both. The putative target set refers to the group of genes increased in *MZdicer* compared to both wild type and *MZdicer*^{+miR-430} and contains miR-430 target sites. Probes: indicates the number of probes upregulated ≥ 1.5 -fold. 3' UTR: indicates the number of genes that had an experimentally identified 3' UTR. Target site: indicates the number of genes with a known 3' UTR that had a predicted miR-430 target site in their 3' UTR. (E) Table indicating the number

of probes in each group, the number of genes with an identified 3' UTR, the number of mRNAs with miR-430 target sites (2 target sites, or one 8mer or one 6mer or 7mer target site), and the percentage of mRNAs with target sites (%). Note that a large fraction of mRNAs whose levels are increased ≥ 1.5 -fold in *MZdicer* have predicted miR-430 target sites in their 3' UTR compared to the control group.

Fig. S4. Enrichment for miR-430 seeds in the putative target set

(A, C, E) Quantification of the number of 6mer or 8mer sequences per 100 kb in different groups within the microarray data. The 6mer or 8mer sequences were complementary to (A, B) miR-430 seeds, (C, D) let-7 seeds or (E, F) miR-1 seeds. (B, D, F) Ratio of the number of target sites/100 kb for a given group (F1) divided by the number of sites/100 kb for the (F2) control set (F1/F2). The genome control set includes all the genes in the array for which a 3' UTR was identified (see material and methods for details). Genes increased in *MZdicer* are slightly enriched for miR-430 target sites in the ORF. Complementary sites for let-7 and miR-1 are not significantly enriched in the groups increased in *MZdicer* compared to the control set.

Fig. S5. miR-430 target validation

Expression of GFP reporters in the presence or the absence of miR-430 in wild type and *MZdicer* mutants, respectively. GFP reporter mRNAs were co-injected with a control DsRed mRNA into wild type and *MZdicer* mutant embryos. The 3' UTRs used in this assay corresponded to controls sensors (C), mRNAs unchanged in the array (U), mRNAs increased in *MZdicer* mutants with no apparent miR-430 target sites (0mer), and mRNAs increased in *MZdicer* mutants with a predicted 6mer, 7mer or 8mer target sites in their 3' UTR. T1x indicates whether the base in the 3' position in the target site is either an uridine (complementary to mir-430b; T1U) or an adenosine (complementary to miR-430a, and miR-430c; T1A). The T1A target sites also follow the preference of an adenosine predicted by Lewis and colleagues (3). Genes with two predicted target sites (each site with ≥ 6 mer complementary to the miR-430 seed) were called ≥ 2 sites. The sequences used in this assay are shown in the file "SOM targets tested".

Fig. S6. Summary of miR-430 target validation and extrapolation of number of miR-430 targets

Summary of the results shown in figures S5 and S7. The field called "validation" shows the number of mRNAs for which the expression of the GFP reporter was higher or lower than 2-fold in the absence of miR-430. Each column shows the results for genes with 2 target sites (each ≥ 6 mer), one 8mer, one 6mer or 7mer, or no apparent target sites in their 3' UTR. "Genes 3' UTR", indicates the number of mRNAs that were increased in the array ≥ 1.5 -fold in *MZdicer*. The "extrapolation" field shows an estimate of the number of genes that may be miR-430 targets. This is based on the fraction of genes for which a 3' UTR was identified and the percentage of genes that were experimentally validated using the GFP reporter assay. To calculate the number of validated targets present in the array, we multiplied the validation efficiency for each column by the estimated number of genes in the field called "array". For example, 92% (12/13) of the putative targets that had two or more target sites were validated *in vivo*. If there were 53 genes with two sites in the group of 328 genes with defined 3' UTR, we would expect ~ 121 such genes in the group of the ~ 750 upregulated genes on the array. Thus, by multiplying 0.92×121 we estimate that ~ 111 of the genes in that group could be validated (Array+validation). These calculations lead to an estimate of 200 (2 sites and/or 8mer) to 350 (6mer or more) targets in the

array that are co-expressed with and regulated by miR-430. Since the array covers less than half of the predicted zebrafish genes (e.g. Ensembl Zv5 http://www.ensembl.org/Danio_rerio/index.html), we estimate that miR-430 has several hundred direct targets in the zebrafish genome.

Fig. S7. Regulation of miR-430 targets depends on miR-430 target sites

Expression of GFP reporters for mRNAs within the putative miR-430 target set, in the presence or the absence of miR-430 regulation, in either wild type or *MZdicer* mutants. GFP reporter mRNAs with wild-type or mutant 3' UTRs were co-injected with a control DsRed mRNA into wild type and *MZdicer* mutant embryos. The miR-430 sequence is shown in the 3'-to-5' orientation. The GFP reporter contains either a wild-type 3' UTR (black) or a 3' UTR with mutated miR-430 target sites (mut; red). Fluorescent microscopy shows GFP reporter expression (green) and DsRed expression (red) at 25-30 hpf. Predicted pairing between the target 3' UTR and a member of the miR-430 family is shown. Note that the wild-type reporters were repressed in wild-type embryos compared to *MZdicer* mutants, which lack miR-430. Mutation of the miR-430 target sites in the 3' UTR (mut) abolished repression of the reporter in wild-type embryos. The GFP reporter for *mknk2* serves as a negative control. Despite having a sequence complementary to miR-430 in its 3' UTR, GFP expression is not significantly higher in *MZdicer* mutants compared to wild type. Mutating this sequence does not affect target expression in either wild type or *MZdicer* mutant embryos. Quantification of the fold increase in GFP expression for some of these GFP reporters between wild type and *MZdicer* mutants is shown in Fig. 3. Regulation was considered significant when the GFP reporter was upregulated ≥ 2 fold in *MZdicer* compared to wild-type embryos.

Fig. S8. Conservation of miR-430 targets in other teleosts

(A) Orthologue identification for zebrafish genes in *Fugu rubripes* (Fugu) and *Tetraodon nigroviridis* (Tetra). Venn diagram for the orthologues identified based on the best reciprocal hit using protein alignments. (B) Venn diagram of the genes that have a 6mer match to a miR-430 seed in their 3' UTR for zebrafish or the predicted 3' UTR for their orthologues in Fugu and Tetraodon. (C) Percentage of genes with a 6mer sequence complementary to the miR-430 seed in zebrafish and Tetraodon or Fugu. The groups were defined depending on the fold upregulation in *MZdicer* compared to wild-type and *MZdicer*^{+miR-430} embryos. (D) Fold enrichment for the groups shown in (C) compared to the control set defined as the group containing all the genes in the array for which the 3' UTR was identified. Note that a 6mer sequence is found in all three species for less than 25% of orthologues. Note that the fold enrichment for genes increased ≥ 1.5 in *MZdicer* compared to wild type and *MZdicer*^{+miR-430} is the same as that for unchanged genes in the array. (E, F) Fluorescent microscopy shows GFP reporter expression (green) and control DsRed expression (red) at 25-30 hpf. Validation of four miR-430 targets identified by microarray analysis and increased more than 1.5-fold in *MZdicer*. (E) The miR-430 target sites found in the zebrafish genes were not present in the orthologous genes in Fugu and Tetraodon. (F) The miR-430 target sites found in the zebrafish genes were conserved in the orthologous genes in Fugu and Tetraodon. All four genes were validated as miR-430 targets with the same efficiency, independently of whether these genes had a miR-430 target site in other species.

Fig. S9. Gene ontology analysis of miR-430 target set

Summary of GeneOntology (GO) terms for miR-430 target site i.e. transcripts corresponding to probes that were upregulated ≥ 1.5 -fold in *MZdicer* compared to *MZdicer*^{+miR-430} and wild type containing one or more 'GCACUU' sites in their 3' UTR. The bar chart above illustrates the frequency of each GO term (Level 3, Process Tree) in this group of genes (observed) compared to genome-wide control (expected). The pie chart below shows the distribution of these GO term frequencies for the miR-430 target set.

Fig. S10. miR-430 facilitates the clearance of maternal transcripts

(A-D) Relative expression profile of mRNAs at 1.5 hpf (corresponding to maternally provided mRNAs), 5 hpf (50% epiboly) and 9 hpf (90%epiboly). (A) Expression profile of random set of 50 mRNAs from the control set. (B) Expression profile of 50 mRNAs in the miR-430 target set. Predominantly maternal genes are highlighted in red. Note that the miR-430 target set is 4-fold enriched for predominantly maternal mRNAs when compared to a random set of mRNAs. (C, D) Expression profile of 50 mRNAs in the control set (C) or the miR-430 target set (D) that are present at different levels maternally (yellow) but show stable mRNA levels during the zygotic stages (blue). The mRNAs in the miR-430 target set accumulate to higher levels in the absence of miR-430 (table S1). (E) Number of predominantly maternal, predominantly zygotic, and zygotic stable mRNAs within the control set or the miR-430 target set. (F) Number of 6-mer miR-430 or miR-1 target sites in the control set, the predominantly maternal, or the predominantly zygotic set of genes. 40% of the predominantly maternal mRNAs have a miR-430 target site.

Fig. S11. miRNAs accelerate the deadenylation of target mRNAs

(A) Schematic representation of the experimental setup to analyze poly(A) tail length: 1) injection of the GFP reporter mRNA; 2) isolation of the mRNA; 3) cleavage of the GFP reporter mRNA in the 3' UTR with a specific DNA oligonucleotide and RNase H; 4) incubation of a fraction of the mRNA with oligo dT and RNase H to assay the size of the deadenylated fragment of RNA. Target mRNAs with a polyadenylation signal [poly(A)] were injected at the one-cell stage and assayed for the length of their poly(A) tail 5 hours after injection. The target contained the GFP ORF and the wild type 3' UTR and a poly(A) signal sequence. The length of the poly(A) tail was assayed by a gene specific oligonucleotide and RNase H treatment as described (4)(Fig 5) (B) Deadenylation of the GFP reporter mRNA in the presence (wt, +) or absence of regulation by miR-430 (*MZdicer*). Accelerated deadenylation of the wild type GFP reporter occurs by 5 hours after injection, compared with the same GFP reporter mRNA injected in *MZdicer* mutants. The size of the deadenylated fragment is visualized by incubating the mRNA with oligo dT and RNase H. (C) Method used to detect poly(A) tail length of endogenously expressed target mRNAs, RT-PCR PAT [poly(A) test] (4). The length of the smear reveals the approximate size of the poly(A) tail. (D) Time course RT-PCR PAT of endogenous miR-430 targets in the presence (wt, +) or the absence of miR-430 (*MZdicer*, -). Note the shortening of the poly(A) tail in the presence of miR-430.

Fig. S12 Models for miRNA function

(A) Role of miR-430 in maternal-to-zygotic transition. In the presence of miR-430 (blue) a large set of maternal mRNAs (red) is post-transcriptionally regulated, allowing normal morphogenesis during zygotic stages. In the absence of miR-430 (gray) maternal mRNAs and their products accumulate and interfere with morphogenesis. (B) Model for miRNA function during development. Cell transitions from state A to state B to state C. Normal: miRNA is activated at

the B-to-C transition and post-transcriptionally regulates mRNAs expressed in state B, thus sharpening and accelerating the transition from state B to state C. Mixed State: in the absence of the miRNA, gene products from state B accumulate, leading to a mixed state B/C. Blocked State: in disease states or upon experimental misexpression, the miRNA is expressed prematurely, resulting in the post-transcriptional regulation of genes normally expressed in state B. This prevents the transition from state A to state B, leading to the maintenance of state A. In the case of cancer, state A might correspond to a cancer stem cell. (C) Model for the post-transcriptional regulation of mRNA targets by miRNAs. (1) Target mRNA translation: interaction between poly(A) binding protein (PABP) on poly(A) tail with translation initiation factors eIF4G/eIF4E on Cap stimulates translation (5). (2) miRNA-induced target mRNA deadenylation: miRISC (6) is recruited to the 3' UTR of target mRNA and accelerates deadenylation by a 3'-to-5' exonuclease (7-9). (3) Loss of target mRNA translation: the interaction between poly(A) binding protein and eIF4G/eIF4E is disrupted, resulting in loss of translation initiation (5, 10). (4) Target mRNA degradation: loss of poly(A) tail results in decapping and degradation of target mRNA (10). Steps 3 and 4 are likely to occur in P-bodies (6, 11-18). The suggested role for miRNAs in accelerating mRNA deadenylation provides a conceptual link with the role of siRNAs in inducing mRNA cleavage. In both cases the Cap-poly(A) tail interaction is disrupted and substrates for exonucleases are generated. This results in both translational repression and mRNA decay.

Table S1.

List of Affymetrix probes and corresponding transcripts that were upregulated ≥ 1.5 fold in MZdicer compared to wild type and MZdicer+miR-430 at 9 hpf. Table key: “affi_id”, Affymetrix probe identifier; “UniGene ID”, UniGene cluster identification number; “Entrez Gene; Refseq”, NCBI reference sequence identifier; Gene title; “GO”, Gene ontology; “mean”, mean value on the array for a probe in a given genotype; “MZdicer”, maternal-zygotic dicer mutants; “MZdicer+miR-430”, maternal-zygotic dicer mutants rescued with miR-430 duplex; “MZdicer+miR430_fold”, fold increase in MZdicer_mean compared to MZdicer+miR430_mean, calculated as the ratio between a MZdicer_mean and MZdicer+miR430_mean; “wild type_fold”, fold increase in MZdicer_mean compared to wild type_mean, calculated as the ratio between a MZdicer_mean and wild type_mean; “MZdicer+miR-430_pval”, p-value calculated for MZdicer_miR-430_fold; “wild type_pval”, p-value calculated for wild type_fold ; “call”, indicates whether the signal corresponding to given transcript suggested that this transcript was present (P) or absent (A) in the sample at significant levels over the background signal for a given gene; “Sixteen_cell”, indicates the developmental stage of 16 cells or 1.5 hpf; “50% epiboly”, indicates the developmental stage of 50% epiboly at 5 hpf; “maps_to”, indicates the cDNA that maps to that probe; “UTR_len”, indicates the length of the 3' UTR for a given transcript; “gene_name”, indicates the gene name; “gene_des”, indicates the NCBI gene description; “fugu_map”, indicates the ensembl gene number for the orthologous gene in Fugu; “tetra_map”, indicates the ensembl gene number for the orthologous gene in tetraodon; “fugu GCACUU or fugu AGCACUUA”, indicates the number of 6mer or 8mer target sites complementary to miR-430 seed in the predicted 3' UTR for the orthologous gene in Fugu; “tetra GCACUU” or “tetra AGCACUUA”, indicates the number of 6mer or 8mer target sites complementary to miR-430 seed in the predicted 3' UTR for the orthologous gene in Tetraodon; “ORF-length”, indicates the length of the longest open reading frame for the zebrafish transcript; “ORF GCACUU or ORF AGCACUUA” indicates the number of 6mer or 8mer target sites complementary to miR-430 seed in the open reading frame of the zebrafish transcript; the columns AK to BB, indicate

the number of 6, 7 or 8mer target sites for miR-430, miR-1 and let-7 in the 3' UTR of the transcripts shown.

Data deposition

Microarray was deposited in Gene Expression Omnibus (GEO) database under accession number GSE4201.

Materials and Methods

The oligonucleotides used in this study are listed as numbers in the materials and methods. The sequences can be found at the end of the materials and methods.

GFP reporter constructs

The constructs containing a 3xPT-miR-1, 3xIPT-miR-1 and 3xPT-miR-204 were previously described in (2)

3xIPT-mir-430: The oligonucleotides 1x-1y and 1w-1z were phosphorylated, annealed and subcloned at the 3' end of pCS2+GFP-F Xho-Xba in a three-piece ligation to generate 3xPT. This target is also complementary to miR-17, miR-20, mir-93, miR-302, but these miRNAs are absent or present at much lower levels than miR-430 at the time of the assay from 0 to 10 hpf.

GFP reporter: The 3' UTR of the gene of interest was amplified by PCR from a 0-24 hours cDNA library. The primers used are shown in the document called SOM targets tested. The PCR fragments were purified through a G-25 Sephadex column, digested and cloned into pCS2+GFP-F between Xho-Xba, EcoRI-Xho, or EcoRI-Xba.

Mutant GFP reporter: The 3' UTR of the gene of interest was amplified in two fragments. These two fragments had a ~40 nt overlap in the mutant region. The mutation in the miR-430 target site was included in the bottom primer of the 5' fragment and the top primer of the 3' fragment. The full-length 3' UTR was obtained by PCR of the fragments one and two with the 5' primer of fragment one and the 3' primer of the second fragment.

Microarray analysis: RNA isolation, microarray hybridization and target identification

Transcriptional profiling. Total RNA was isolated from 20 zebrafish embryos using Trizol (Invitrogen). RNA processing, array hybridization and data acquisition were performed at the microarray core facility at Memorial-Sloan Kettering Cancer Center, New York. Quality of RNA was ensured before labeling by analyzing 20–50 ng of each sample using the RNA 6000 NanoAssay and a Bioanalyzer 2100 (Agilent). Samples with a 28S/18S ribosomal peak ratio of 1.8–2.0 were considered suitable for labeling. For samples meeting this standard, 2 mg of total RNA was used for cDNA synthesis using an oligo(dT)-T7 primer and the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). Synthesis, linear amplification, and labeling of cRNA were accomplished by *in vitro* transcription using the MessageAmp aRNA Kit (Ambion) and biotinylated nucleotides (Enzo Diagnostics). Ten micrograms of labeled and fragmented cRNA were then hybridized to the zebrafish genome array (Affymetrix GeneChip® Zebrafish Genome Array ~14,900 oligonucleotide-based probe sets) at 45°C for 16 h. Post hybridization staining and washing were processed according to the manufacturer (Affymetrix). Finally, chips were scanned

with a high-numerical Aperture and flying objective (FOL) lens in a GS3000 scanner (Affymetrix). The .CEL images were normalized across all arrays to the median probe intensity using the invariant set normalization method of dChip software (v.1.3; www.dchip.org). For each individual probe signal, a signal value, standard error, and a present/absent call were calculated based on dChip's PM/MM difference model. Comparisons between arrays were performed with dChip's "Compare" function, which calculated p values for each comparison and assessed a change call (upregulated ≥ 1.5 -fold, downregulated ≤ 0.8 fold, and unchanged between 1.2 and 0.8 fold) between input and baseline based upon a change p value of less than 0.2. The 811 probes upregulated in *MZdicer* mutants compared to wild type and *MZdicer*^{+miR430} fulfilled the following criteria: 1) they were called present (P) in *MZdicer* mutants. 2) They were called upregulated in *MZdicer* mutant compared to both wild type and *MZdicer*^{+miR-430}. 3) The average signal ratio was ≥ 1.5 . 4) The change p value was ≤ 0.2 (617 probes in *MZdicer*+miR430 and 721 probes in wild type had a p value ≤ 0.05). The list of genes was compiled using Netaffix, NCBI, ZFIN, Ensemble and PubMed searches. The list of genes upregulated in *MZdicer* mutants is provided in Table-S1(≥ 1.5 fold).xls. The group of unchanged genes fulfilled the following criteria: 1) they were called present (P) in wild type; 2) the average signal was ≥ 200 units (1/5 of the mean value of the array). 3) The average signal ratio was ≤ 1.2 and ≥ 0.8 . The group of downregulated genes in *MZdicer* fulfilled the following criteria: 1) they were called present (P) in wild type. 2) The average signal ratio was < 0.8 .

miRNA target search: ORF and 3' UTR sequences were obtained for all the genes in the array when available from Refseq or Ensemble transcripts with EST evidence. The control group used in all the analysis contained all the genes in the array for which a 3' UTR was identified. This group is called "control set" or "whole array". Sequences complementary to different miRNA seeds (miR-1, let-7 and miR-430) were quantified in the 3' UTR and the ORF for the different groups. The fold enrichment for each target site was calculated by dividing the average number of sites complementary to seeds/100 kilobases of sequence in each group by the average number of sites complementary to seeds/100 kb in the control set. The 6, 7 and 8mer sequences used in this analysis were based on complete complementarity to the 8nt in the 5'arm of the miRNA (seed) including a variant with an adenosine in the 3' position (3). The 5'-to-3' sequences searched for in the target mRNAs were:

miR-430: GCACUU, AGCACUU, GCACUUU, GCACUUA, AGCACUUA

miR-1: CAUUC, ACAUUC, CAUUC, ACAUUC

let-7: UACCUC, CUACCUC, UACCUCA, CUACCUCA

Transcriptional profiling of miR-430 targets: The stages analyzed in this study were maternal (1.5 hpf; stage1; s1), 50% epiboly (5 hpf; s2) and 90% epiboly (9 hpf; s3). Duplicate samples from 20 zebrafish embryos at each developmental stage were processed for microarray analysis as described above. To obtain the profile of expression for each gene a group of "present" or "absent-low" genes were defined. Genes in the "present" group fulfilled the following criteria: 1) they were called present (P) in wild type at least in one of the stages analyzed, 2) the average signal was ≥ 200 units in any of the three stages (1/5 of the mean value of the array). Genes in the "absent" group fulfilled the following criteria: 1) they were called absent (A) in wild type at all of the three stages analyzed, 2) the average signal was < 200 units (1/5 of the mean value of the array) at all of the three stages analyzed.

The average value for each stage (s) for a gene x was defined as \bar{I}_s^x . The relative expression for each gene at a given stage (e.g. s1; relative expression E_{s1}^x), was calculated as a percentile of the

signal for that gene and stage, relative to all three stages for that gene (i.e. $E^x_{Sn} = 100 * [I^x_{S1} / (I^x_{S1} + I^x_{S2} + I^x_{S3})]$). The sum of the average signal for each stage for a given gene was considered 100% ($\sum(I^x_{S1}, I^x_{S2}, I^x_{S3}) = 100\%$). These values were calculated only for the group of genes called present.

Predominantly maternal genes fulfilled the following criteria: 1) Maternal expression was $\geq 73\%$, 2) the zygotic levels at 5hpf and 9hpf were each $\leq 15\%$, 3) they were in the present group.

Maternal-zygotic genes with steady zygotic levels fulfilled the following criteria: 1) $E^x_{S1} \leq 73$ and $D(E^x_{S2-S3}) \leq 3\%$. The difference of the relative expression between two zygotic stages s2 and s3 is smaller than 3% and predominantly maternal genes were excluded, 2) they were in the present group.

Predominantly zygotic genes fulfilled the following criteria: 1) Maternal expression was $< 15\%$, 2) the zygotic levels at 5hpf and 9hpf were each $> 15\%$, 3) They were in the present group.

The number of genes in each group was calculated for the miR-430 target set and for the present genes in the control set.

A set of 4 genes, *eef1b*, *SMARCA2*, *nme2* and *zgc:103420* were validated at the mRNA level by *in situ* hybridization at 1.5, 5, and 9hpf (data not shown). Their *in situ* hybridization profiles matched the expression profile obtained by microarray analysis.

mRNA synthesis, target validation, miRNA duplexes and injection

mRNA for injection was transcribed using the Message machine kit according to manufacturer's instructions (Ambion). 1000 pl of a 0.15 $\mu\text{g}/\mu\text{l}$ GFP reporter mRNA solution was injected into ~20 wild-type or *MZdicer* embryos at the one cell stage. DsRed control mRNA was co-injected with the GFP reporter mRNA at 0.1 $\mu\text{g}/\mu\text{l}$.

miRNA duplex: The miRNA duplexes were purchased from IDT, resuspended in the manufacturer's buffer to a concentration of 100 μM . Working aliquots were prepared in RNase free water at a 10 μM and stored at -80°C . See below for miRNA:miRNA* sequences. The two most 3' ribonucleotides were replaced by the corresponding deoxynucleotides (e.g. dT). A mismatch was included in the second base of the sense strand of the miRNA to reduce its 5' annealing (19). 500-1000 pl of a 1-10 μM miRNA duplex solution was injected into wild-type and *MZdicer* embryos.

Rescue experiments were performed by injecting 1500 μl of 10 μM concentration of a mixture of miR-430a, miR-430b, and miR-430c in a ratio 2.5:1.7:1 to mimic the stoichiometric concentration observed in *in vivo* cloning approaches (2).

3' UTR identification

All available Refseq Zebrafish mRNAs were assembled and their 3' UTRs identified by isolating sequence downstream of the longest ORF. These sequences were filtered according to whether they contained a valid hexamer within 40bp of the last nucleotide before the poly(A) tail and for length $> 40\text{bp}$, yielding 7,629 3' UTR sequences in total. A second set of 46,219 cDNA and EST sequences was downloaded from NCBI and filtered similarly for length, hexamer sequence and poly(A). These sequences were scanned against all Ensembl Zebrafish transcripts from Ensembl release 34 using BLASTn and used to determine genomic coordinates for the 3' UTR positions of Zebrafish transcripts, yielding 6,213 candidate 3' UTR sequences based on Ensembl sequence and cDNA/EST evidence. Mapping of the Affymetrix probe identification number to our 3' UTR database identified a total of 5219 3' UTR for which we have reasonable evidence based cDNA

sequences. More than 50 of these 3' UTRs were experimentally validated by RT-PCR. Out of the 5219 3' UTRs identified 4797 were obtained from Zebrafish Refseq and 422 were identified from Ensembl and EST evidence. This set of 5219 genes included in the array were used as control set.

***In vivo* target validation**

The GFP reporter mRNA was co-injected into *MZdicer* mutants and wild-type embryos as described above. Wild type and *MZdicer* mutant embryos were photographed side by side at 25-30 hours after injection. The average pixel intensity for GFP and DsRed was measured using Adobe Photoshop CS. To calculate the fold-increase of GFP reporter expression in *MZdicer* compared to wild type the following calculation was performed:

$$\text{Intensity}(I) = I_{\text{total}} - I_{\text{background}}$$

$$\text{Fold increase in intensity} = (I_{\text{GFP-MZdicer}} / I_{\text{GFP-wt}}) / (I_{\text{DsRed-MZdicer}} / I_{\text{DsRed-wt}})$$

The total GFP or DsRed intensities (I_{total}) were measured using a 0.1x0.2 inches rectangle in the trunk of the embryo above the yolk extension. The average background intensity ($I_{\text{background}}$), was measured for each channel and each embryo using a 0.1x0.2 inches rectangle adjacent to the embryo. This was identical to the background intensity of an un-injected embryo. The average intensity (I) was measured by subtracting the background intensity from the total intensity. The fold increase in GFP expression was calculated as the ratio of the GFP intensity in *MZdicer* and wild-type embryos, divided by the ratio of DsRed intensity in *MZdicer* and wild-type embryos, to normalize for any differences due to injection. A similar procedure was followed to calculate the fold-increase in GFP expression for the mutant GFP reporters.

Orthologue identification

We used Ensembl Compara (Release 32) to obtain orthologue mappings from each Zebrafish peptide to peptides in both Fugu and Tetraodon based on best reciprocal hits and synteny conservation. Peptide-peptide mappings were then mapped to transcript-transcript mappings where this could be done unambiguously. 3' UTRs for both Fugu and Tetraodon were assembled by taking annotated Ensembl 3' UTRs where available, or by taking 1.3kb downstream of the last exon. This resulted in a set of 17,561 zebrafish transcripts which could be mapped reliably to Fugu transcripts and 20,686 to Tetraodon transcripts.

GO Analysis

Gene Ontology (GO) terms were obtained for each Zebrafish transcript from ZFIN. For each GO tree and level being considered, terms were mapped up the tree to the appropriate level. Overrepresentation for each term (O_{term}) in a group of sequences with C terms is calculated as follows:

$$O_{\text{term}} = \left(\frac{F_1}{F_2} \right)$$

$$\text{where } F_1 = \frac{N_{\text{group}}^{\text{terms}}}{\sum_{i=1}^C N_{\text{group}}^i} \text{ and } F_2 = \frac{N_{\text{genome}}^{\text{term}}}{\sum_{i=1}^C N_{\text{genome}}^i}$$

where F_1 is the frequency of a term in the group being considered and F_2 is the frequency of a term in the whole genome.

Image acquisition

Embryos were analyzed with Zeiss M2Bio and Axioplan microscopes and photographed with a Zeiss Axiocam digital camera. Images were processed with Zeiss AxioVision 3.0.6 and Adobe Photoshop CS software. Flat mounts for Fig.5 and were prepared by dissecting a dorsal fragment of 90% epiboly embryos.

Translation inhibition of GFP reporter mRNA

Translation inhibition of the GFP reporter mRNA was accomplished by injection of 1000 pl of a 250 μ M solution of a morpholino targeted against the translation start of GFP (GFP-MO).

Northern blot

Between 10 and 30 embryos were collected at different stages during development and frozen in liquid nitrogen. Total RNA was extracted using Trizol (Invitrogene). The samples were resuspended in formamide or water depending on whether the samples would be analyzed by denaturing polyacrylamide or denaturing agarose, respectively.

Northern blots shown in Fig.1, and Fig.5 were probed using a GFP probe or a pCS2+3' UTR probe respectively. To label the probes we used the Random Primed DNA Labeling Kit (Roche). The Northern blot shown in Fig. S1 was probed with an oligonucleotide complementary to miR-430a labeled using T4 PNK and g-P32-ATP.

Polyadenylation assay

To assay the length of the poly(A) tail, we followed the oligo dT/RNase H-Northern assay (4). Briefly, the GFP reporter mRNA was injected at the one cell stage and extracted at 2, 4 and 6 hours after injection. To assay the effect of miRNA regulation on the polyadenylation state of the target mRNA, we injected GFP reporter mRNA in wild type and *MZdicer* mutants, which lack miR-430. We also compared the poly(A) tail of a wild-type GFP reporter and a GFP reporter whose sequences complementary to miR-430 (GCACTT) were mutated to (GCAGAT). mRNA was isolated and incubated with RNase H and a DNA oligonucleotide complementary to the 3' UTR of the target mRNA (oligo #2), liberating a 3' fragment of ~200 bases plus the poly(A) tail. A fraction of these mRNAs was also incubated with an oligo-dT (oligo #3) and RNase H, generating a deadenylated fragment of ~200 bases. The GFP reporter mRNA was visualized by hybridizing the Northern blot with an Xho-NotI fragment of the pCS2+ vector. Changes in poly(A) tail length over time were analyzed by comparing the adenylated and the deadenylated mRNAs.

RT-PCR Poly(A) Tail (PAT)

To visualize the length of the poly(A) tail of endogenously expressed mRNAs we collected wild-type, *MZdicer* and *MZdicer*^{+miR-430} embryos at 2h, 4h, and 6h after fertilization. Total RNA from 3 embryos was reverse transcribed using 1000 ng of the linker primer 5'GCGAGCTCCGCGGCCGCGTTTTTTTTTTTTT 3' and the Retroscript kit (Ambion) following

the manufacturers instructions. The poly(A) tail of several targets was amplified using 1 μ l of RT reaction and a gene specific primer ~200 bp from the poly(A) tail and the linker primer.

In vitro polyadenylation of mRNAs

All GFP reporters contained a SV40 polyadenylation signal except for the GFP reporter mRNAs injected for figure S11, which were polyadenylated *in vitro* using a poly(A) tailing kit (Ambion).

Oligonucleotide sequences

These sequences are described 5'-3'. Restriction sites are in lower case. The 3' UTR sequences and the primers used for target validation are provided in an additional document SOM-Targets-tested.pdf

3xIPT-miR-430

1x:CTAGctcgagaatctagACTACCTGCACTGTGAGCACTTTGataag

1y:ACTACCTGCACTGTGAGCACTTTGataagACTACCTGCACTGTGAGCACTTTG

1w:TGCAGGTAGTcttatCAAAGTGCTCACAGTGCAGGTAGTctagattctcgag

1z:TCGACAAAGTGCTCACAGTGCAGGTAGTcttatCAAAGTGCTCACAG

2: pCS2+ specific primer: TCTACGTAATACGACTCACTATAGT

3: oligodT: TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

miR-1 duplex

UGGAAUGUAAAGAAGUAUGUAUdTdT

AUACAUAUCUUCUUACAUCGAdTdT

mir-204 duplex

UUCCCUUGUCAUCCUAUGCCUdGdT

AGGCAUAGGAUGACAAAGGGUAdTdT

miR-430a duplex

UAAGUGCUAUUUGUUGGGGdTdA

CCCCAACAAAUAGCACUAAdTdT

mir-430b duplex

AAAGUGCUAUCAAGUUGGGdGdT

CCCAACUUGAUAGCACUAUdTdT

miR-430c duplex

UAAGUGCUCUCUUUGGGGdTdA

CCCCAAAGAGAAGCACUAAdTdT

GFP morpholino

ACAGCTCCTCGCCCTTGCTCTCACCAT

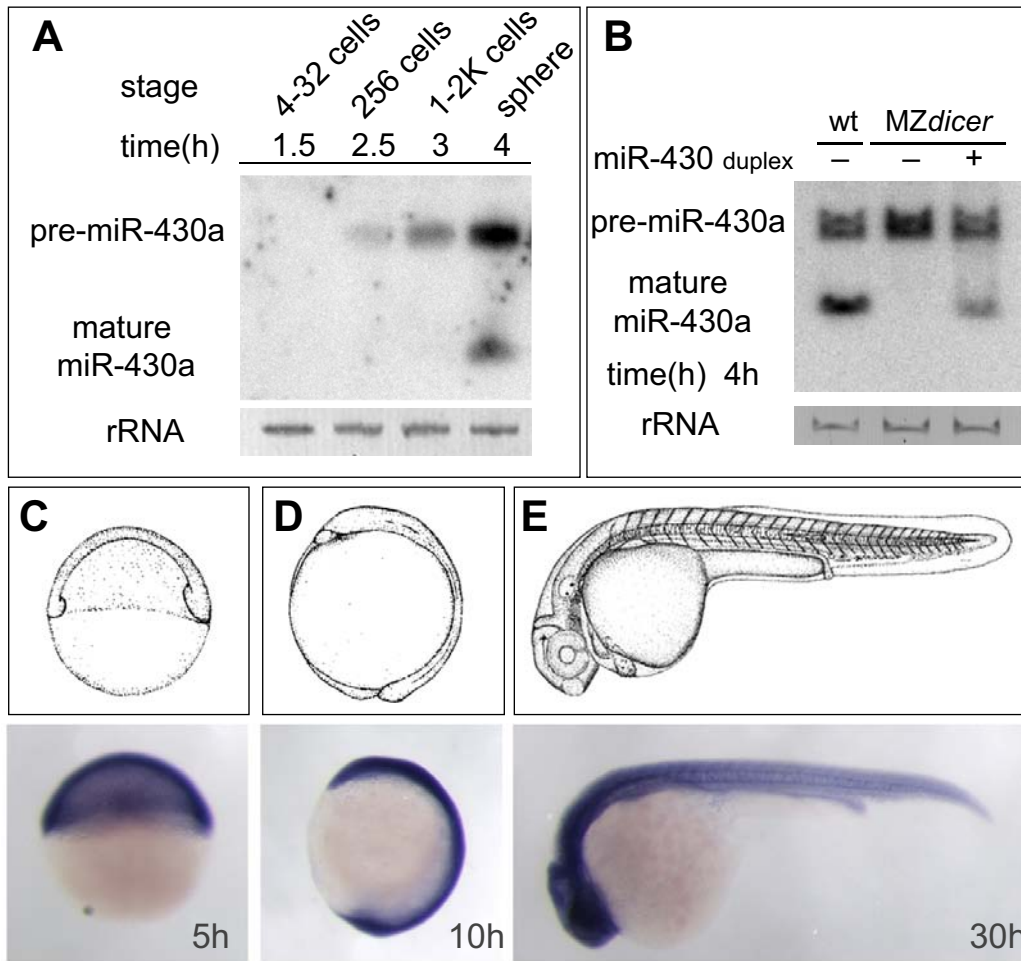
Probes

miR-430a: CTACCCCAACAAATAGCACTTA

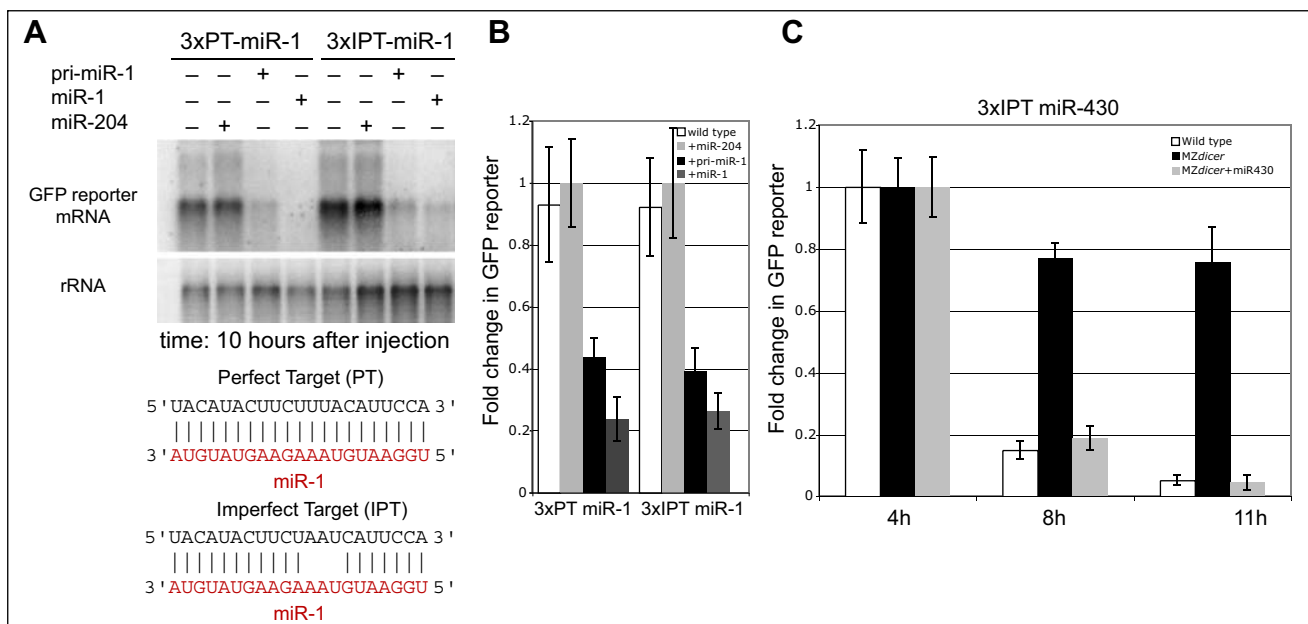
References

1. E. Wienholds *et al.*, *Science* **309**, 310 (2005).
2. A. J. Giraldez *et al.*, *Science* **308**, 833 (2005).
3. B. P. Lewis, C. B. Burge, D. P. Bartel, *Cell* **120**, 15 (2005).
4. F. J. Salles, W. G. Richards, S. Strickland, *Methods* **17**, 38 (1999).
5. L. D. Kapp, J. R. Lorsch, *Annu Rev Biochem* **73**, 657 (2004).
6. R. S. Pillai, *Rna* **11**, 1753 (2005).

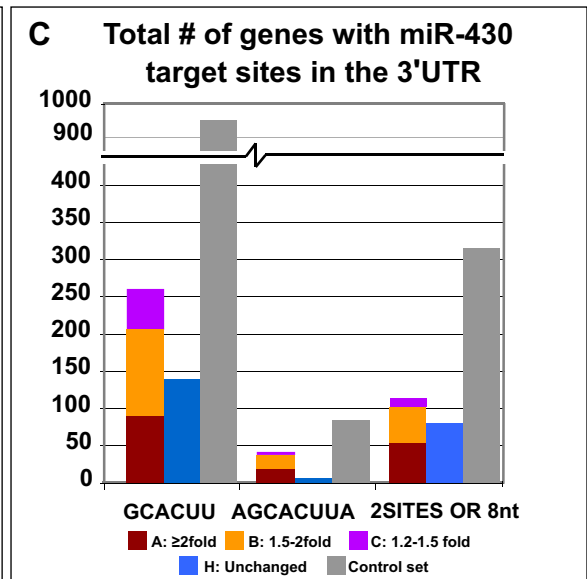
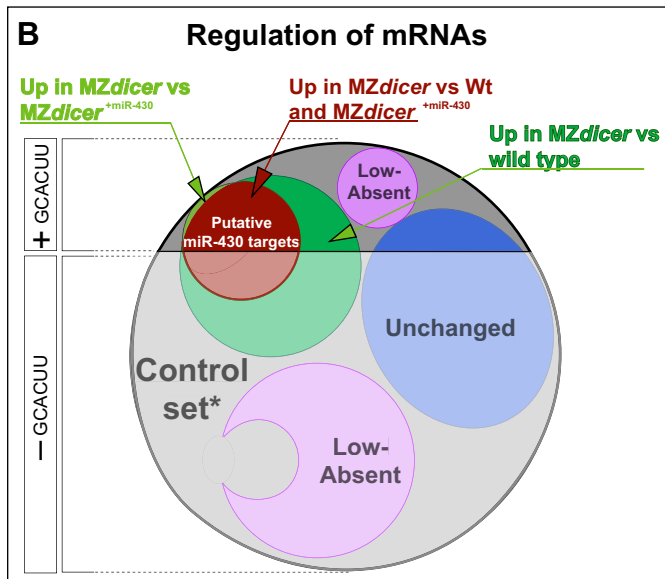
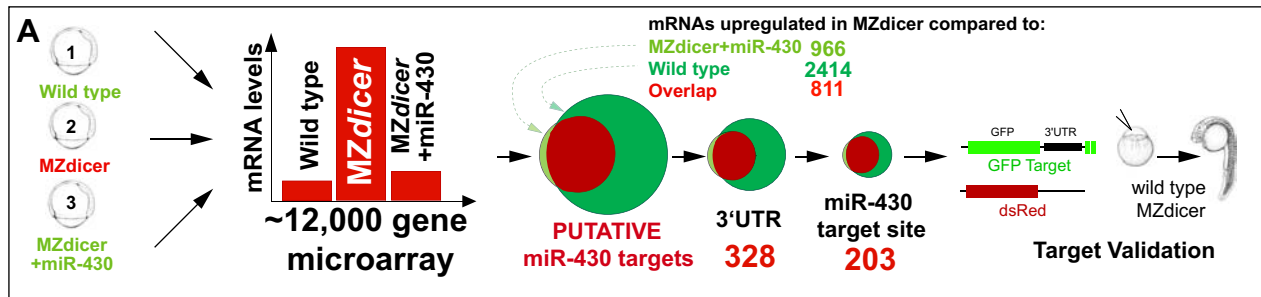
7. J. D. Richter, *Microbiol Mol Biol Rev* **63**, 446 (1999).
8. S. Vasudevan, E. Seli, J. A. Steitz, *Genes Dev* **20**, 138 (2006).
9. C. H. de Moor, H. Meijer, S. Lissenden, *Semin Cell Dev Biol* **16**, 49 (2005).
10. R. Parker, H. Song, *Nat Struct Mol Biol* **11**, 121 (2004).
11. J. Liu *et al.*, *Nat Cell Biol* **7**, 1161 (2005).
12. J. Liu, M. A. Valencia-Sanchez, G. J. Hannon, R. Parker, *Nat Cell Biol* **7**, 719 (2005).
13. J. Rehwinkel, I. Behm-Ansmant, D. Gatfield, E. Izaurralde, *Rna* (2005).
14. U. Sheth, R. Parker, *Science* **300**, 805 (2003).
15. A. Jakymiw *et al.*, *Nat Cell Biol* **7**, 1167 (2005).
16. G. L. Sen, H. M. Blau, *Nat Cell Biol* **7**, 633 (2005).
17. L. Ding, A. Spencer, K. Morita, M. Han, *Mol Cell* **19**, 437 (2005).
18. R. S. Pillai *et al.*, *Science* **309**, 1573 (2005).
19. D. S. Schwarz *et al.*, *Cell* **115**, 199 (2003).



Giraldez et. al, Figure S1



Giraldez et. al, SOM Figure 2



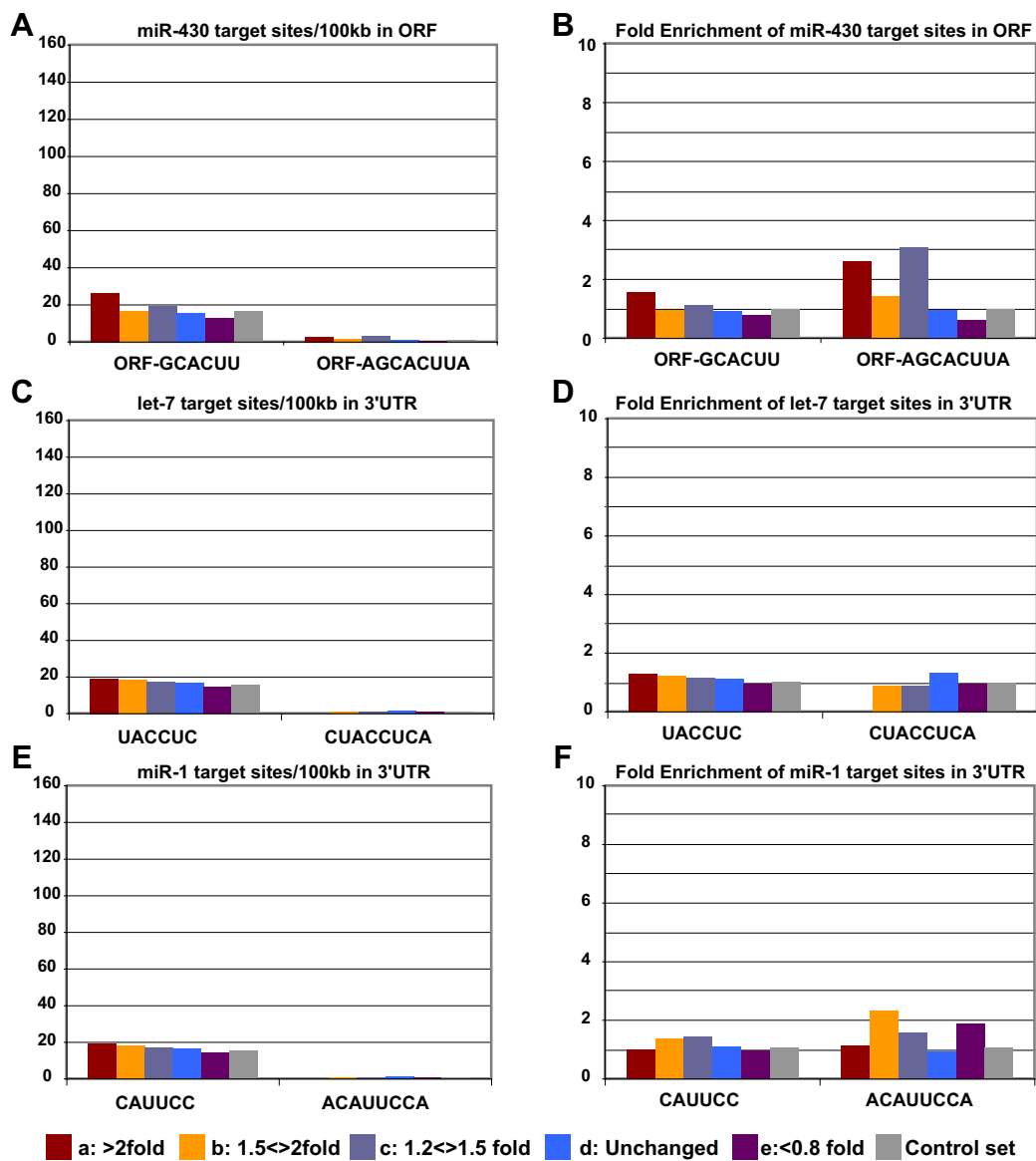
D

	Up in MZdicer vs MZdicer ^{+miR-430}	Up in MZdicer vs wild type	Up in MZdicer vs (wt and MZdicer ^{+miR-430})
≥1.5 fold	n	n	n
Probes	966	2414	811
3'UTR	385	902	328
Target sites	222	359	203

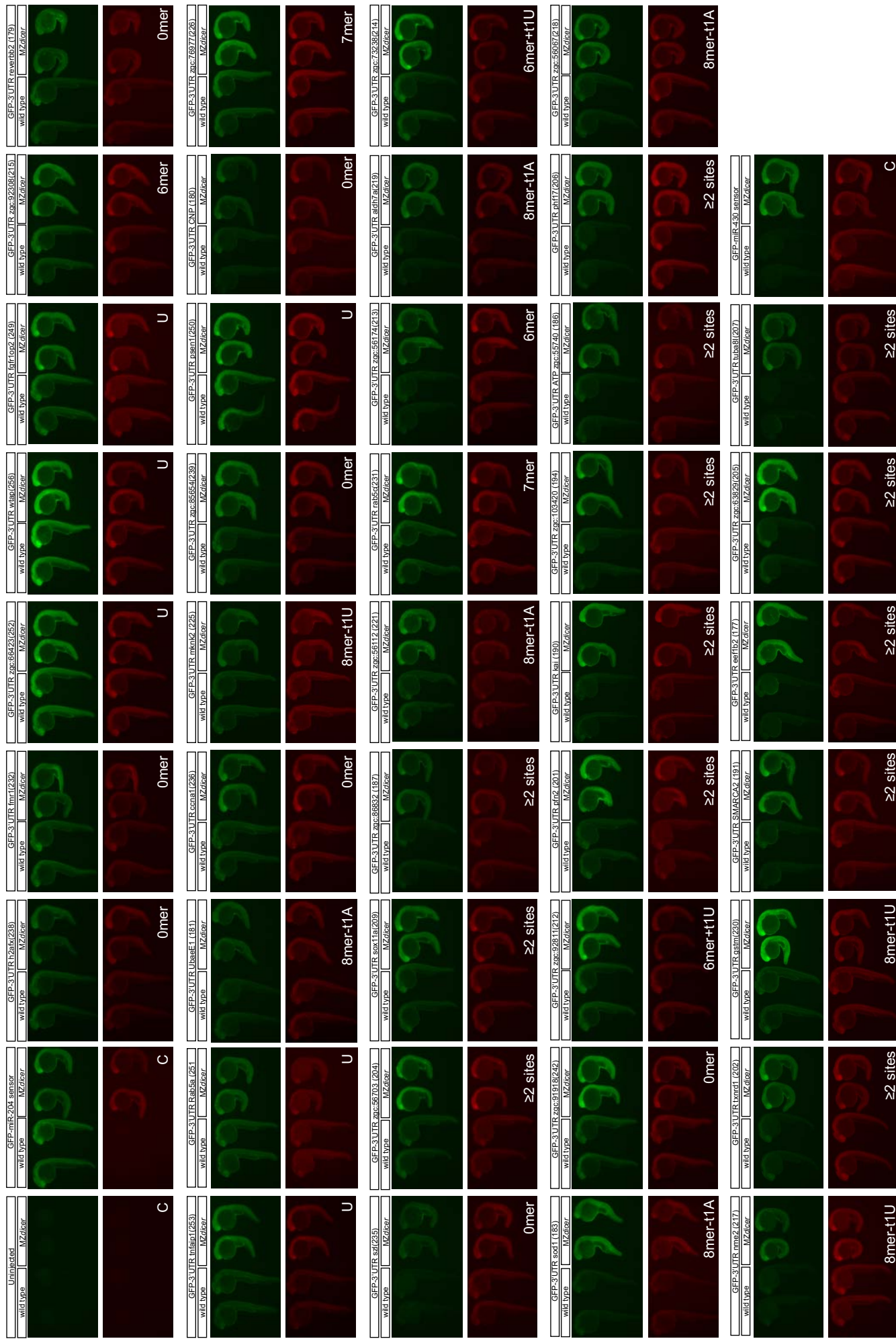
E

	>2fold	1.5-2fold	1.2-1.5fold	Unchanged	<0.8 fold	Whole array	
Probes	275	536	330	2749	713	12754	
3'UTR	118	210	133	1293	359	5219*	
Target sites	Sites	2 8mer 6-7mer	2 8mer 6-7mer	2 8mer 6-7mer	2 8mer 6-7mer	2 8mer 6-7mer	
	#	31 20 37	22 28 65	7 10 33	20 14 105	1 2 14	148 169 635
	%	26 16 31	10 13 31	5 8 25	2 1 8	0 1 4	3 3 12

Giraldez et. al, Figure S3



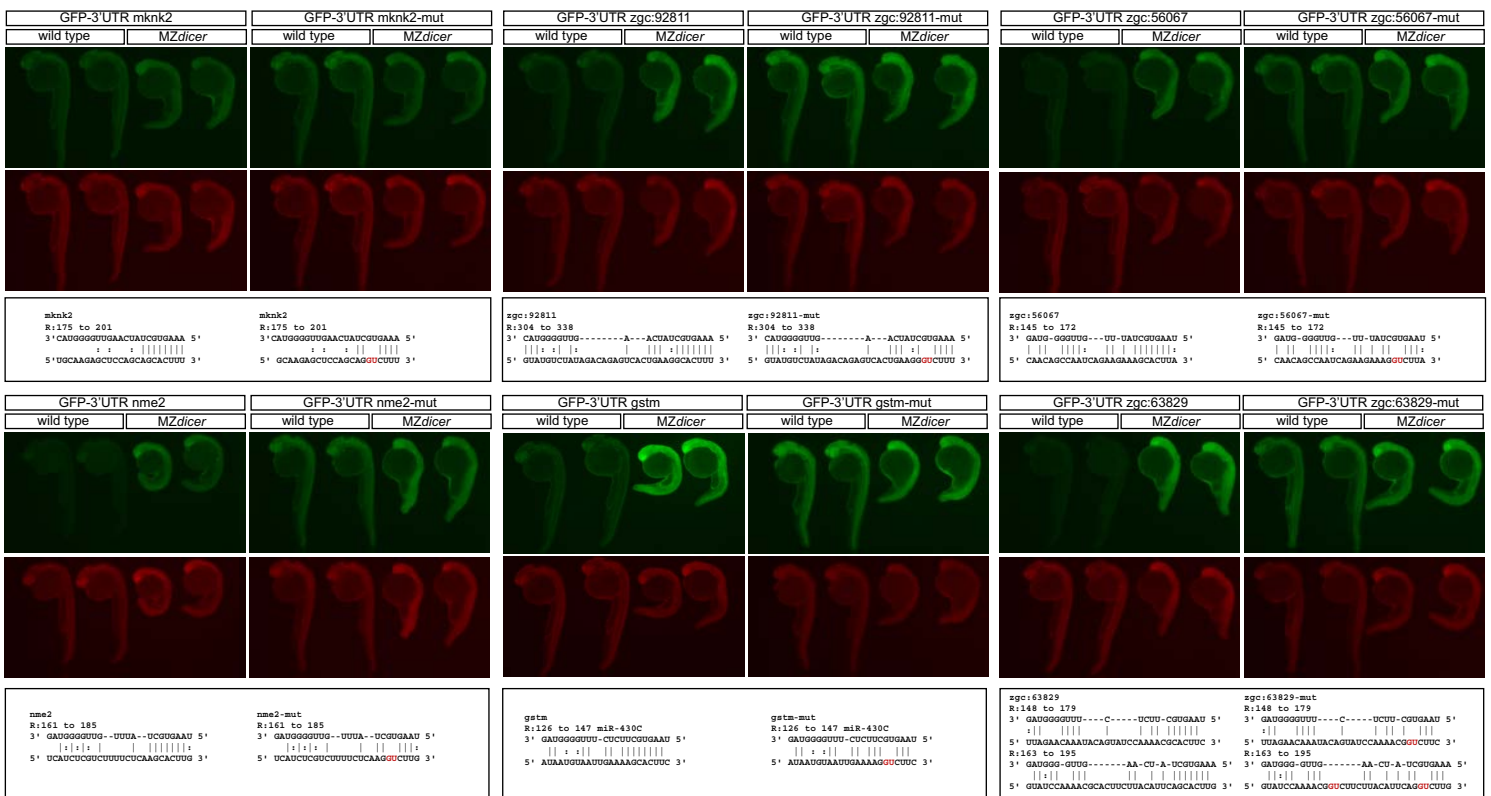
Giraldez et. al, Figure S4



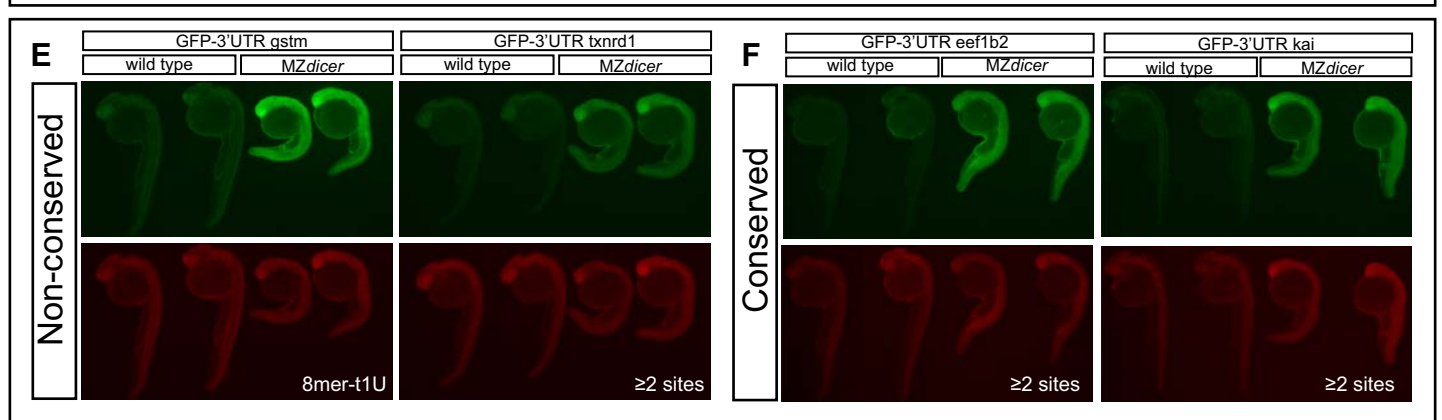
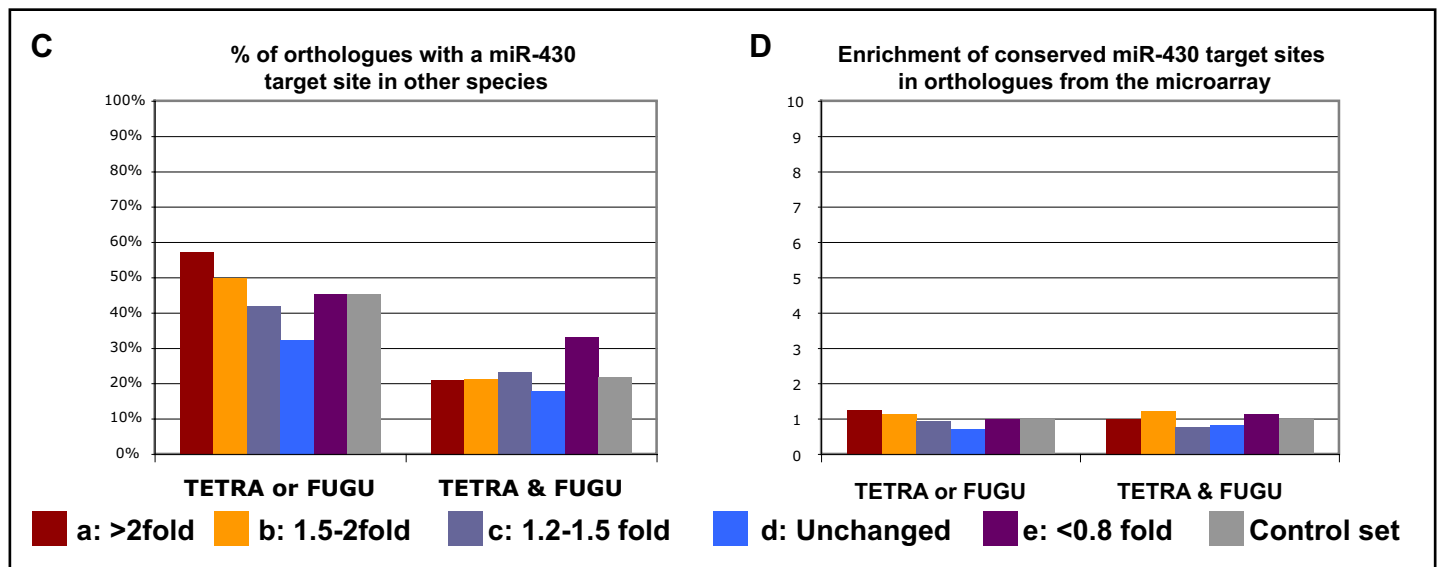
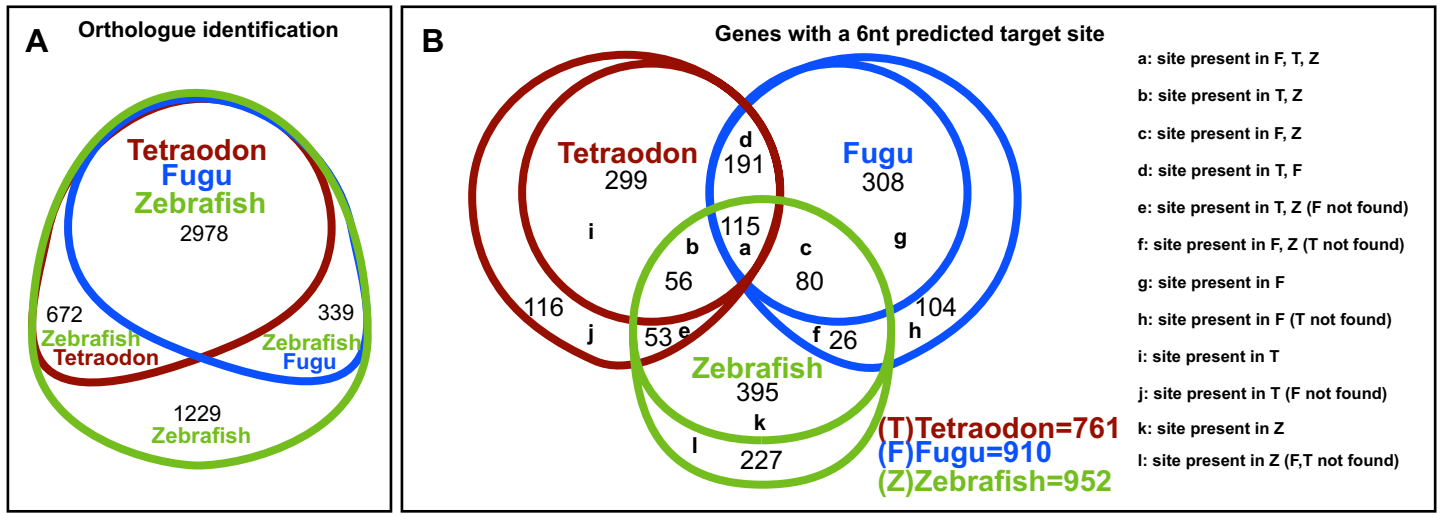
Giraldez et. al, Figure S5

		Target sites			
		2	8mer	6-7mer	0
Validation	Upregulated >2fold	12	6	5	2
	Upregulated <2fold	1	2	2	7
	% validated genes 3'UTR	92	75	71	22
Extrapolation	Array	121	110	233	-
	Array+Validation	111	82	165	-
	Target sites	2 sites or 8mer ≥ 6mer seed			
	Total in 12K array	~200		~350	

Giraldez et. al, Figure S6

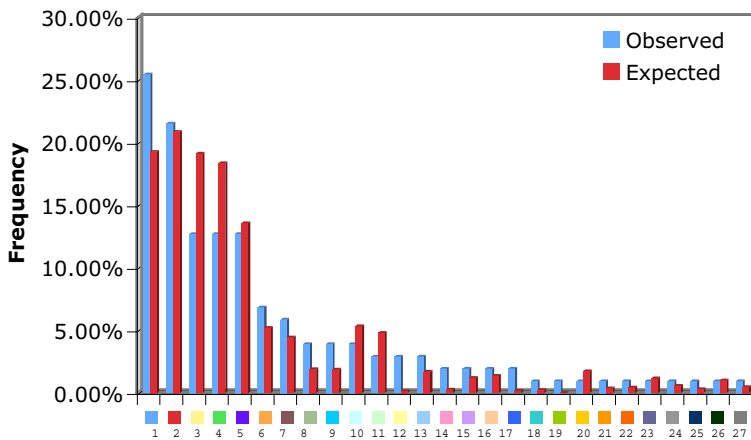


Giraldez et. al, Figure S7



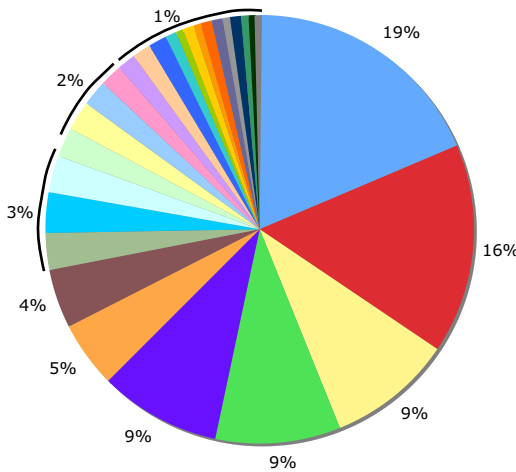
Giraldez et. al, Figure S8

GO Process Terms Level 3

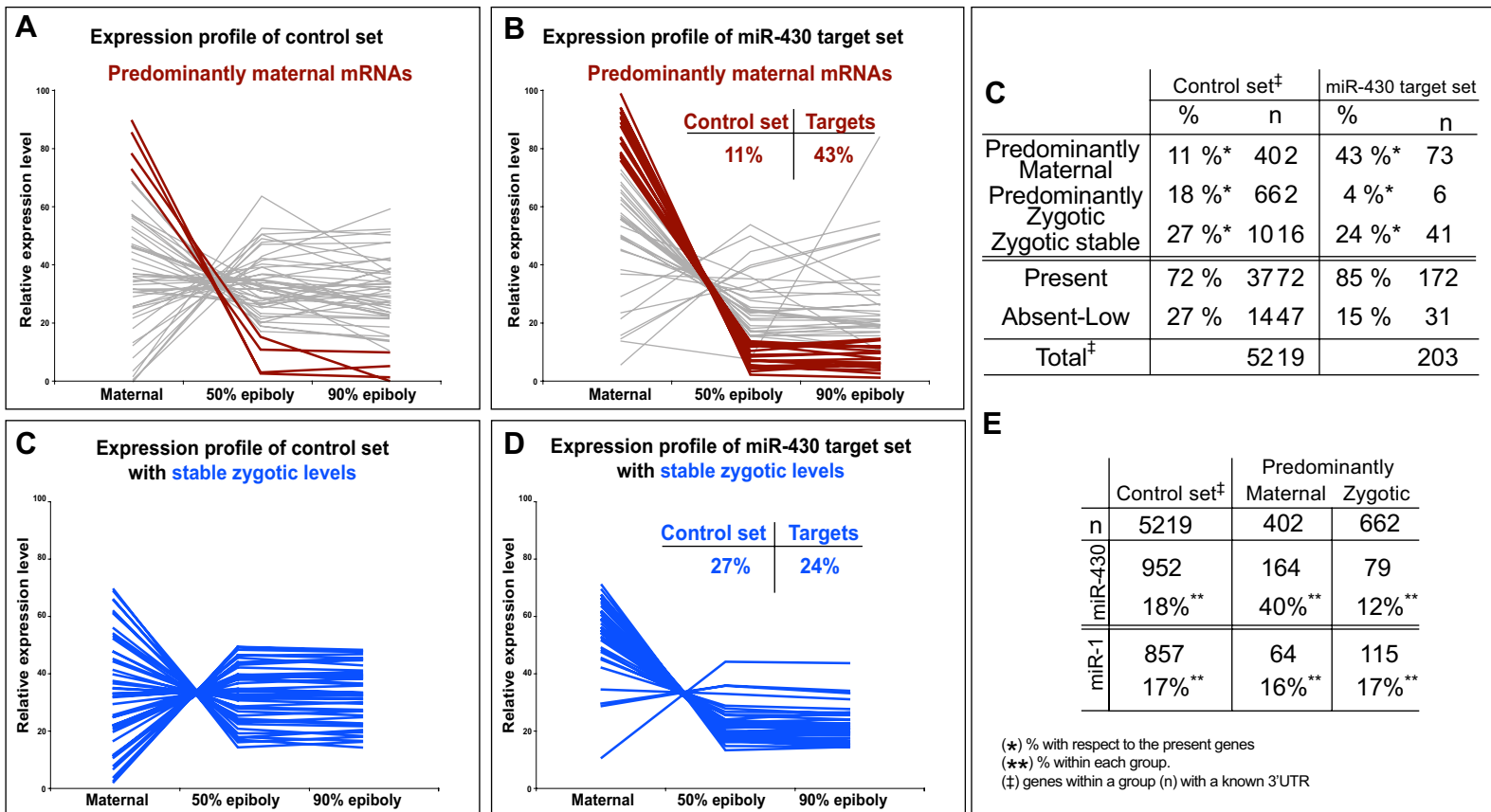


- 1 macromolecule metabolism
- 2 cell growth and/or maintenance
- 3 regulation of metabolism
- 4 nucleobase, nucleoside, nucleotide and nucle..
- 5 signal transduction
- 6 catabolism
- 7 biosynthesis
- 8 amino acid and derivative metabolism
- 9 response to biotic stimulus
- 10 electron transport
- 11 organogenesis
- 12 axis specification
- 13 lipid metabolism
- 14 aromatic compound metabolism
- 15 cofactor metabolism
- 16 cell motility
- 17 sexual reproduction
- 18 one-carbon compound metabolism
- 19 positive regulation of enzyme activity
- 20 response to external stimulus
- 21 regulation of neurogenesis
- 22 embryonic pattern specification
- 23 regulation of cellular physiological process
- 24 embryonic morphogenesis
- 25 oxygen and reactive oxygen species metabolism
- 26 regulation of signal transduction
- 27 cell death

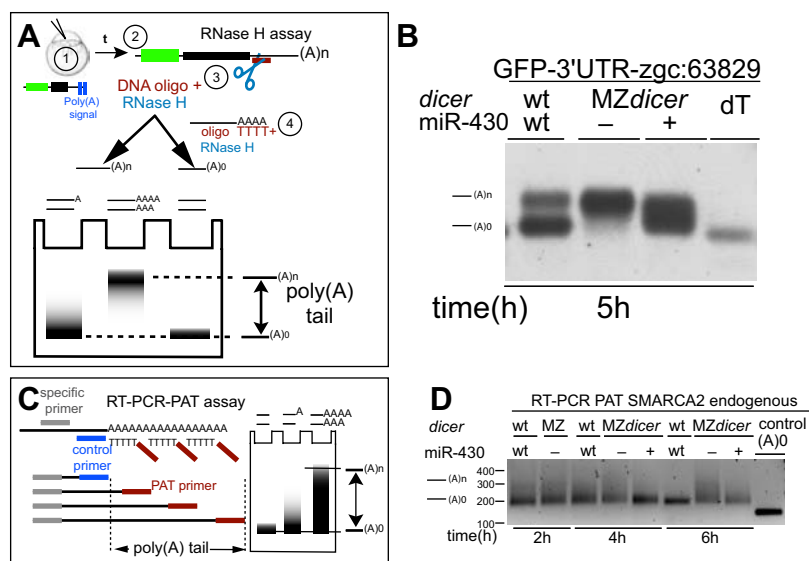
Observed Terms Level 3 GO Process



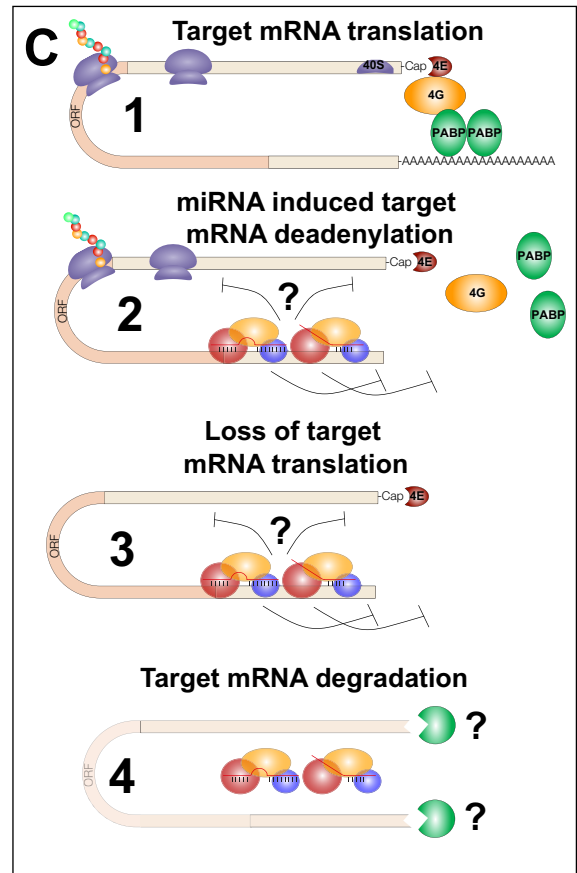
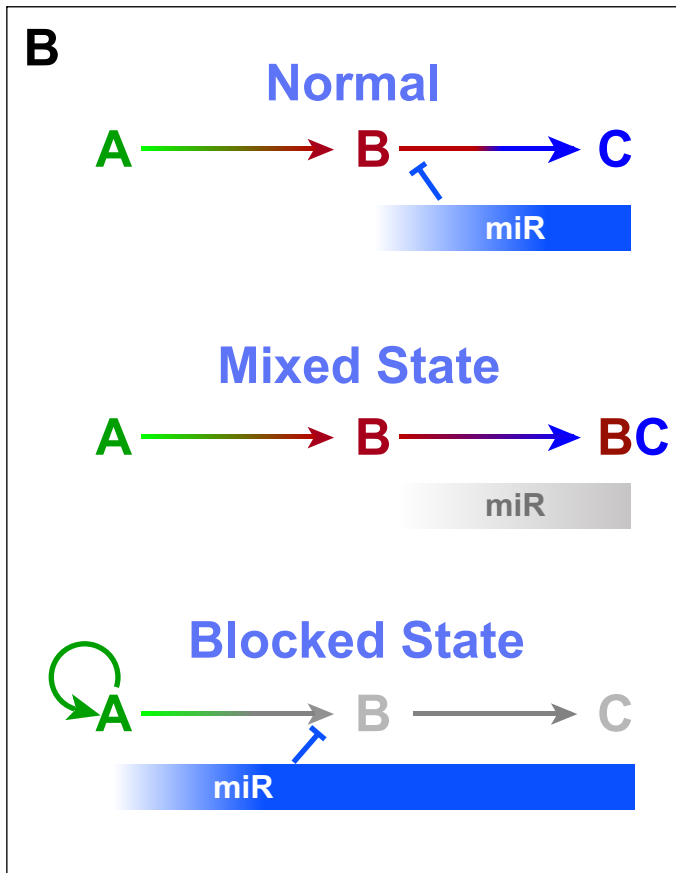
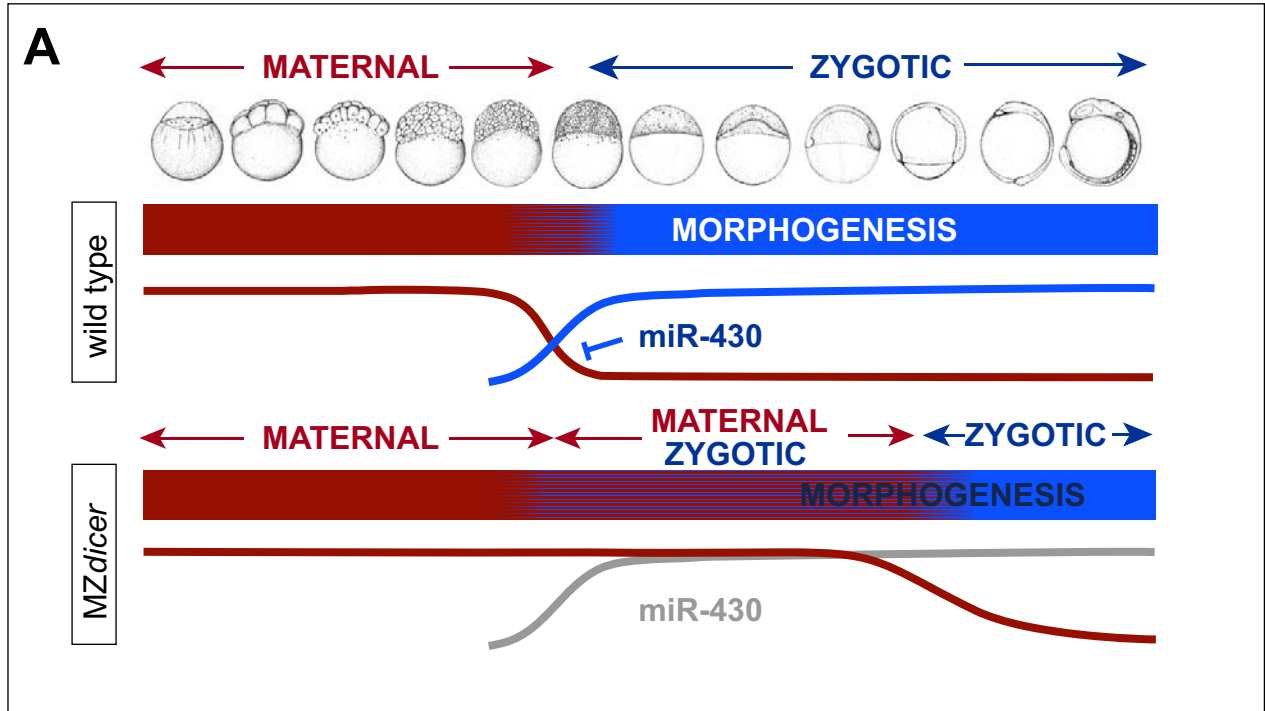
Giraldez et. al, Figure S9



Giraldez et. al, SOM Figure S10



Giraldez et. al., Figure S11



Giraldez et. al, Figure S12