Supplementary Information for:

The role of site accessibility in microRNA target recognition

Michael Kertesz[1*], Nicola Iovino[2*], Ulrich Unnerstall[2], Ulrike Gaul[2†], Eran Segal[1†]

[1] Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.
[2] Laboratory of Developmental Neurogenetics, Rockefeller University, New York, NY 10021, USA.
[*] These authors contributed equally to this work.
[†] These authors contributed equally to this work. Correspondence should be addressed to U.G. (gaul@mail.rockefeller.edu) or E.S. (eran@weizmann.ac.il)

Supplementary Note: Discussion of other approaches that use secondary structure in predicting microRNA-mRNA interactions

In the following, we provide a brief discussion of previously published papers that considered the secondary structure of the target mRNA as part of the interaction between the target and the microRNA. A detailed description of our own approach is provided in the Methods section of the paper.

Zhao et al., Nature 2005[1], Cell 2007[2]

Zhao et al. were the first to make use of secondary structure considerations in identifying microRNA targets, and applied their approach to the case of miR-1. They initially filter for perfect seed matches to the first 8 nucleotides of the microRNA (allowing for G:U wobble at the last nucleotide) and require conservation across four species (chicken, mouse, rat, and human). They then examine the mRNA structure in which the sites are embedded as follows: Adding 70 nucleotides upstream and downstream of the target site, they use mFold[3] to predict the minimum free energy (MFE) structure, which is then examined for unstable structure elements, such as hairpin loops, interior loops etc.; the authors define a (species-specific) threshold for the minimum size of each such element by examining the distribution of unstable elements in secondary structures obtained from arbitrary 3'UTR regions. In addition, they consider the overall stability of the entire (target +/- 70 bp) region as computed by the total folding energy ΔG, again using randomly selected 3'UTR regions for comparison.

Although the authors showed the utility of this procedure for explaining interactions between miR-1 and its targets, the approach has several limitations. First, by considering only a single structure (the MFE structure), the method is susceptible to errors that secondary structure prediction algorithms, such as mFold, make in the pairing of individual nucleotides; this problem becomes particularly acute when specific secondary structure elements, with defined size requirements, are to be predicted. One solution to this problem is to consider larger sets of structures or the entire ensemble of possible structures. In addition, when choosing thresholds by comparing with 'random' UTR sequences, the base composition should be controlled, since G:C pairs have

higher binding energy compared to A:U pairs, which of course affects the total free energy and thus, indirectly, the MFE structure prediction. Finally, the choice of 70 bp flanking sequence for computing the overall stability of the target region is not justified or evaluated. The authors do not experimentally validate their approach or show its utility for genome-wide predictions.

Robins et al., PNAS 2005[4]

Independently of Zhao et al, Robins et al. have early on proposed a method for predicting microRNA-target interactions that incorporates certain features of the secondary structure of the target mRNA. The authors define microRNA target sites as those that meet three criteria: (1) 5' pairing: perfect sequence complementarity between bases 1-7, 2-8, or 3-9 of the microRNA and 7 consecutive nucleotides in the target; (2) 3' pairing: the number of A:U and G:C matches are counted (G:U wobbles are ignored), with each type of match having a potentially different weight; a threshold for this measure is defined by fitting to a list of known target sites; (3) The MFE structure of the target site is required to contain at least 3 consecutive nucleotides within the 5' seed that are unpaired.

In contrast to nearly all other microRNA-target prediction methods, this approach scores the microRNA-target duplex on the basis of counting the number of paired nucleotides, i.e., without explicitly using the binding energy of the microRNA-target duplex. Such a computation does not make use of stacking energies between adjacent bases and thus provides additional degrees of freedom in determining the relative weights of A:U and G:C matches, which are then fitted using a small set of validated *C. elegans* and *Drosophila* targets. In addition, the proposed approach integrates secondary structure by way of a specific demand for three consecutive unpaired nucleotides in the MFE structure, a requirement that is not well motivated. Moreover, as discussed above, MFE structure predictions are generally susceptible to errors in predicting the exact pairing of particular nucleotides. The authors test 19 targets predicted by their algorithm and achieve good correlation for the very top predictions, but do not perform well for targets ranking outside the top 4, which leads them to an unusually low estimate of 10 as the median number of targets per microRNA. Their

predictions thus miss many targets that have in the meantime been validated experimentally.

Long et al., Nature Structural Molecular Biology 2007[5]

Motivated by the native paradigms in the worm (let7:lin41), in which tandem target sites retain *in vivo* functionality despite G:U wobbles and mismatches in the 5' seed, these authors do not enforce any constraints on the 5' pairing of the microRNA on the target, in contrast to most other prediction methods. Instead, the authors model microRNA-target interactions as a two-step process that consists of a rate-limiting nucleation step followed by an elongation step in which the microRNA-target interaction is extended to encompass the full length of the microRNA.

For the nucleation step, which extends a criterion used by Robins et al.[4], the authors examine the secondary structure of the target and search for blocks of four consecutive nucleotides that have a probability of 0.5 or greater of being unpaired; they compute the binding free energy between the microRNA and each such block, and then take the best such binding energy as the nucleation potential for the considered hybridization. To overcome the limitation of MFE structure predictions, the authors perform this calculation by averaging over 1000 structures sampled from the entire Boltzmann ensemble. They then apply a threshold of -4.09[6] kcal mol^{-1} or -5.2[7] kcal mol^{-1}, for the nucleation potential; once a target site has passed this threshold, the authors determine the total interaction energy between microRNA and target site in a second step. This is done by computing the difference between the free energy of the microRNA-target duplex (ΔG_{hybrid}) and the free energy of disrupting the structure of the target site ($\Delta G_{disruption}$). ΔG_{hybrid} is computed as the binding free energy of the MFE structure, forcing a threshold of -14 kcal mol^{-1}. $\Delta G_{disruption}$ is determined by computing the difference between the native mRNA structure and the same structure with the target site unpaired, again averaging over 1000 samples from the ensemble of structures. Finally, for multiple sites, the scores of the individual binding sites are summed, and another threshold is introduced for the total combined free energy (-10 kcal mol^{-1}).

Sampling and averaging over many predicted secondary structures as Long et al. do represents a clear improvement over considering only the MFE structure. However, it is unclear whether the concept of a nucleation step, which is derived from kinetic measurements of freely hybridizing RNA or DNA molecules[7], is appropriate for microRNA-target interaction, since microRNAs bind as part of a large protein complex (RISC). Overall, the authors employ four different numerical thresholds that a target site must meet in order to be deemed functional. These parameters are derived from different sources and fitted to a small and biased dataset (only one microRNA-mRNA pair), which is based on a non-linear experimental readout. Finally, the authors do not provide genome-wide prediction or comparisons to other published approaches that would permit gauging their method's success.

We generated such a comparison using the web-server provided by the authors (http://sfold.wadsworth.org/starmir.pl). Each microRNA-target pairs tested in our own experiments and from the set of interactions we collected from the literature was scored, using the STarMir default parameter threshold values (ΔG_{hybrid} = -14 kcal/mol; $\Delta G_{initiation}$ = 4.09 kcal/mol) and the procedure described by the authors of linearly summing multiple target sites for a microRNA on the same 3'UTR to arrive at a total predicted degree of repression $\sum \Delta G_{total}$. We also followed the authors approach in adding the last 200bp of the coding sequence to the UTR. We found that STarMir does not perform better than expected by chance in predicting validated microRNA-target interactions in *Drosophila* (**Fig. 4a**). When applying the algorithm to our own experimental data, we find that many targets do not pass the prediction thresholds that are applied (**Supplementary Fig. 3**), and the ones that do achieve poor correlation to the measured degree of repression (r=0.44, p=0.23 for the set of target constructs from *hid*, *grim*, and *rpr*, and r=0.13, p=0.62 for the *miR-184* target constructs). Importantly, in the converse test, our own method, without any fitting of thresholds or parameters, predicts the experimental data used to validate STarMir with good success (correlation of 0.8 between our interaction energy $\Delta\Delta G$ and the measured expression levels reported in ref[5]).

Muckstein et al., Bioinformatics 2006[8]

This computational study extends the Vienna RNA package to compute the probability that a given sequence interval is unpaired. Using this extension, the authors predict binding energies of RNA-RNA interactions in a fashion that takes into account the secondary structure of the binding site. First, the partition function of the possible secondary structures of the target RNA is calculated, subject to the constraint that a certain sequence interval (the binding site) remains unpaired. Next, they compute the interaction energy between the target and interacting RNA given that the binding site of the target is unpaired. In this computation, they assume that the region of interaction between the two RNAs may contain bulge loops and mismatches. Finally, the total interaction probability at a possible binding site is obtained as the sum over all possible types of binding. The authors apply their method to published siRNA experimental data and show that it can quantitatively predict siRNA binding to the target mRNA.

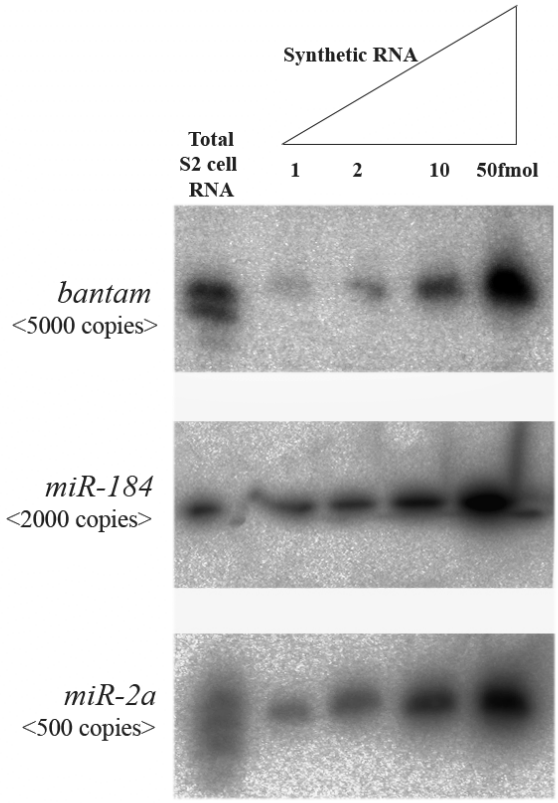
Overall, the thermodynamic modeling of the binding event employed in this approach is similar to the one we describe. However, the authors do not apply their method to microRNAs, and thus their approach does not model several features of the microRNA-target interaction that we find as critical for explaining the extant experimental data for microRNAs. These important additions include the constraints on the target seed that further constrain the partition function, the combination of multiple sites into a total interaction energy score, and the option to include an additional requirement to unpair the local region flanking the binding site.

REFERENCES

1. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214-20 (2005).
2. Zhao, Y. et al. Dysregulation of Cardiogenesis, Cardiac Conduction, and Cell Cycle in Mice Lacking miRNA-1-2. *Cell* **129**, 303-317 (2007).
3. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429-31 (2003).

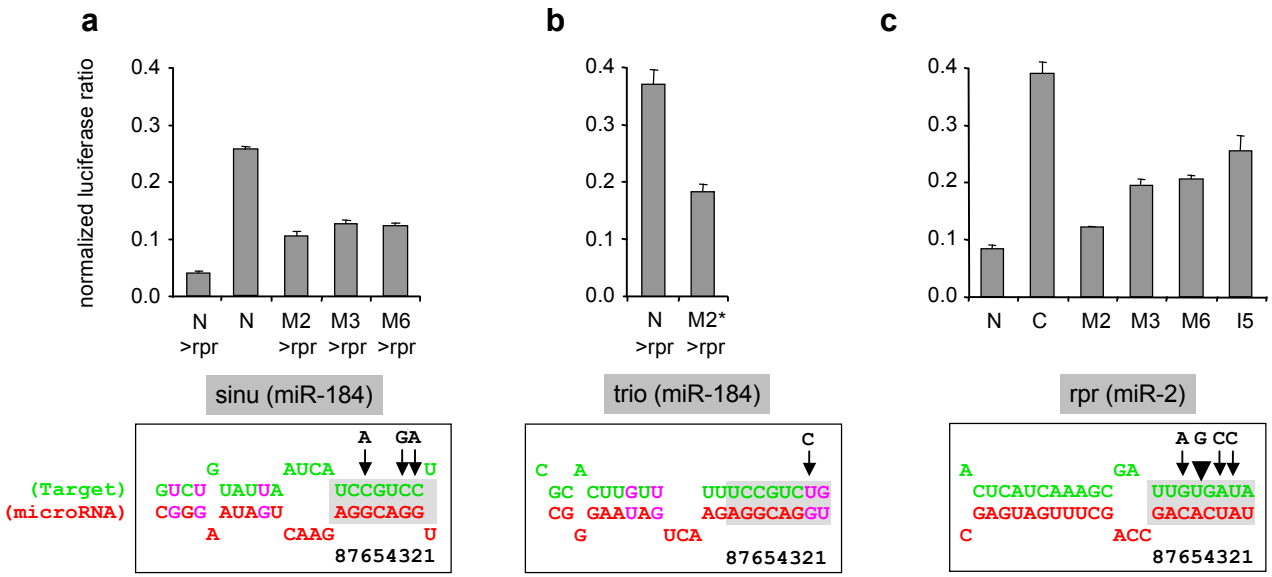
4. Robins, H., Li, Y. & Padgett, R. W. Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci U S A* **102**, 4006-9 (2005).
5. Long, D. et al. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* **14**, 287-94 (2007).
6. Xia, T. et al. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719-35 (1998).
7. Hargittai, M. R., Gorelick, R. J., Rouzina, I. & Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol* **337**, 951-68 (2004).
8. Muckstein, U. et al. Thermodynamics of RNA-RNA binding. *Bioinformatics* **22**, 1177-82 (2006).

Supplementary Figure 1



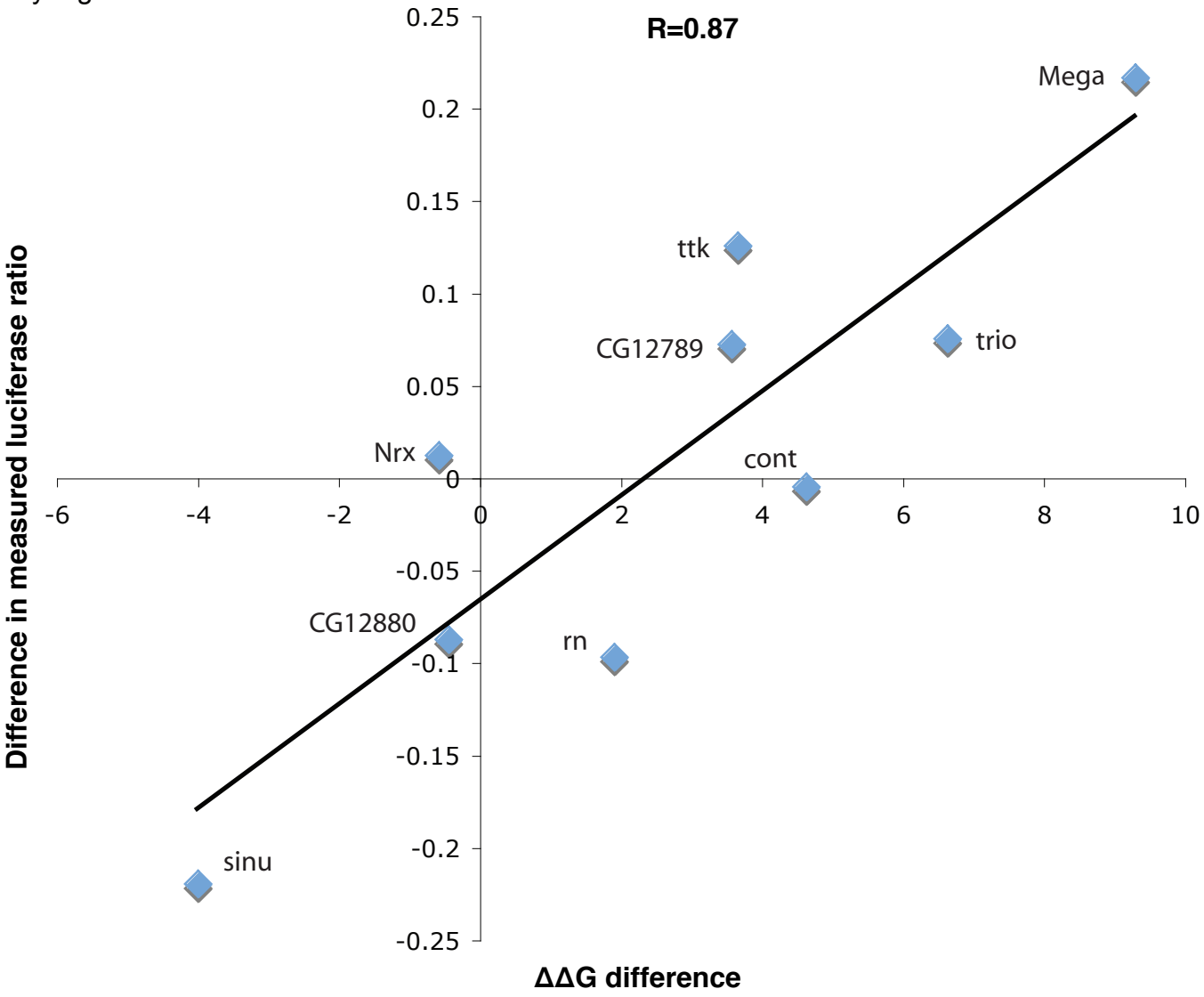
Supplementary Figure 1. Quantitative Northern analysis of *bantam, miR-184* and *miR-2* expression in S2 cells. 5 μg of total RNA were extracted from 300,000 S2 cells using Trizma (Sigma-Aldrich), electrophoresed in a 10% polyacrylamide-8M urea gel and electroblotted onto Hybond-N^+ membrane (Amersham). To quantify the amount of microRNA within the extract, synthetic oligoribonucleotides (Sigma-Genosys) were loaded in different amounts (1,2,10,50 fmol) onto the gel for comparison. The blots were hybridized (5X SSC, 20mM Na_2HPO_4, 7% SDS, 1X Denhardt's, 100μg/ml salmon sperm DNA ; 50°C) using ^{32}P-endlabeled deoxyribonucleotides, and exposed on a phosphoimager screen using a Fuji BAS100 Bioimager for appriopriate lengths of time. Copy number was determined by comparing signal intensities using *image gauge* software.

Supplementary Figure 2



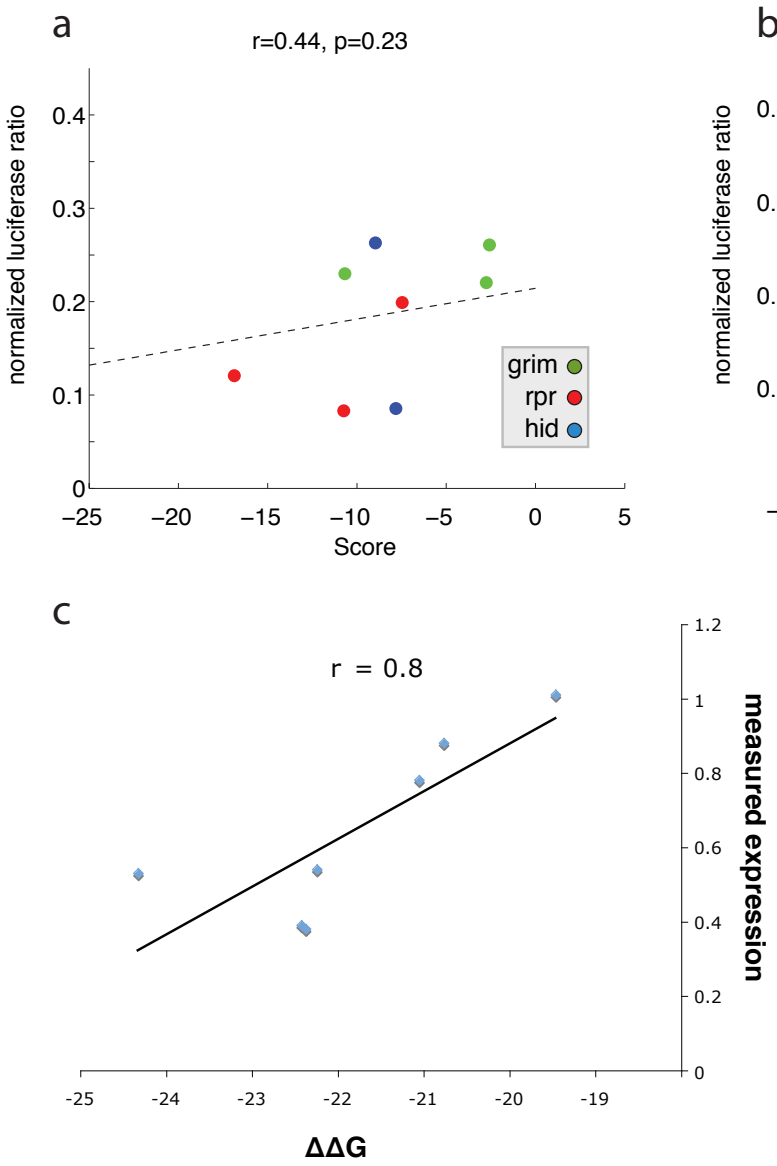
Supplementary Figure 2. Mutational analysis of *miR-184* targets. (a) Expression levels mediated by ~200 bp 3' UTR constructs containing different versions of the *miR-184* target site from *sinu*, embedded in the *rpr* sequence context (*>rpr*). Shown are results for the native site (*N>rpr*) and for constructs containing different point mutations in the 5' seed of the site (*M>rpr*); for comparison, the result for the native site in the *sinu* context (*N*), which has a more closed secondary structure, is also shown. The miRNA:target sequence pairing is depicted below the graph, 5' seed is highlighted in grey; G:U wobbles in pink, arrows point to the mutated bases (2, 3 and 6). Comparison with the equivalent mutations for the miR-2 target site in *rpr* in (c) shows similar scale of effects. (b) Similar to (a), but for constructs containing versions of the *miR-184* target site from *trio*, and a mutation designed to improve site complementarity (the G:U wobble at position 2 is replaced by a G:C match); the mutated site indeed shows significantly stronger repression. (c) Similar to (a), but for the *miR-2* target site from *rpr*; data repeated from Figure 1 for comparison. All results shown are average values (± SEM) of normalized Renilla/Firefly ratios obtained from four to eight replicates.

Supplementary Figure 3



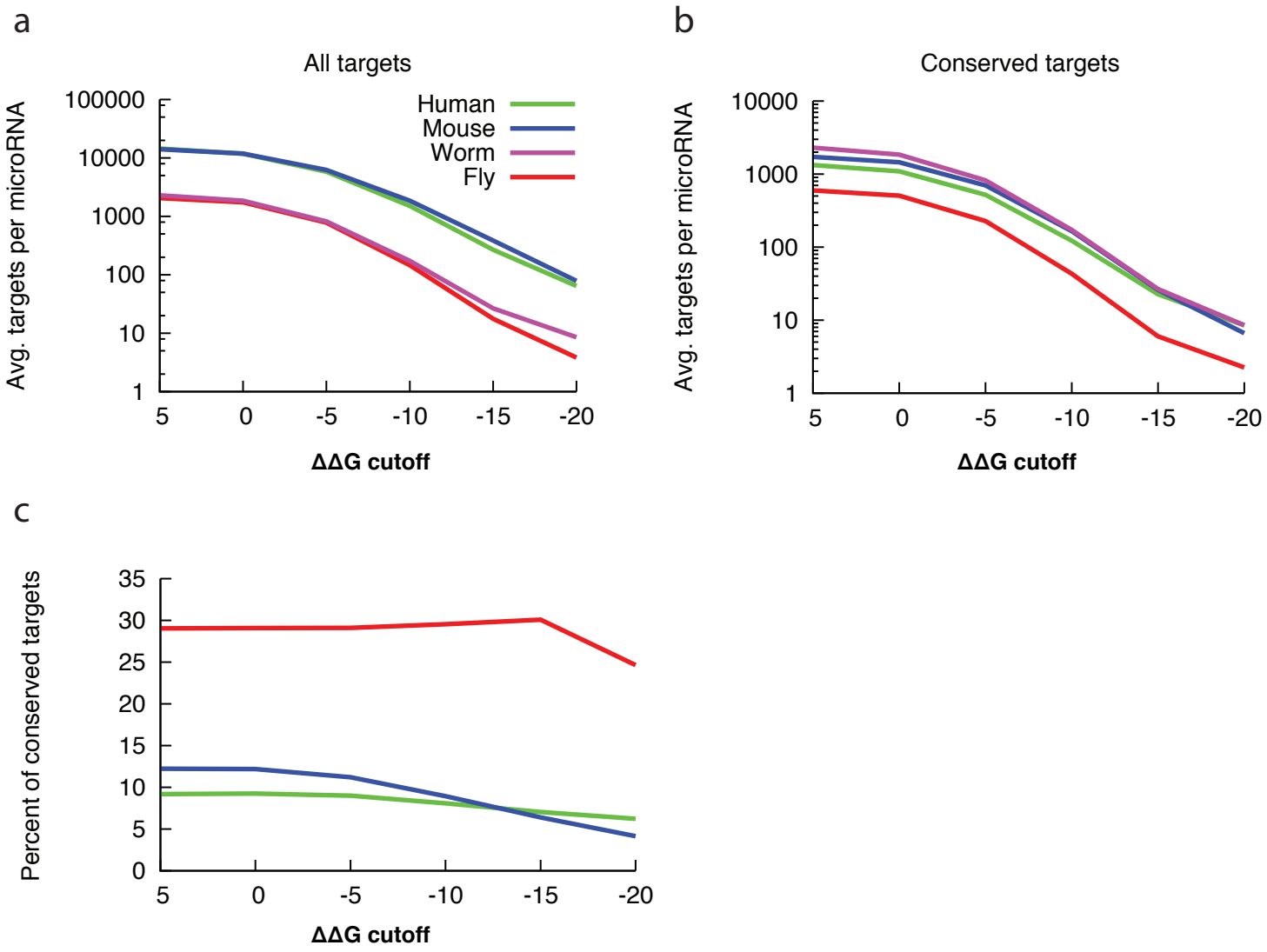
Supplementary Figure 3. Differentials in site repression between native and *rpr* sequence context are predicted by our model. For each of the miR-184 sites that were tested in both the native and *rpr* sequence context (**Fig. 3a**), shown is the difference between the interaction energy ΔΔG predicted by the model in the *rpr* sequence context and the native context (x-axis), and the difference in the measured expression for the site between the *rpr* sequence context and the native context (y-axis). All pairs from **Fig. 3a** are shown and labeled, except for *Gli*, whose site does not pass our seed criteria. The correlation between the difference in predicted interaction energy and the difference in measured expression is shown (top). ΔΔG values were calculated using the same flank settings as in **Fig. 3** (14 upstream, 15 downstream).

Supplementary Figure 4



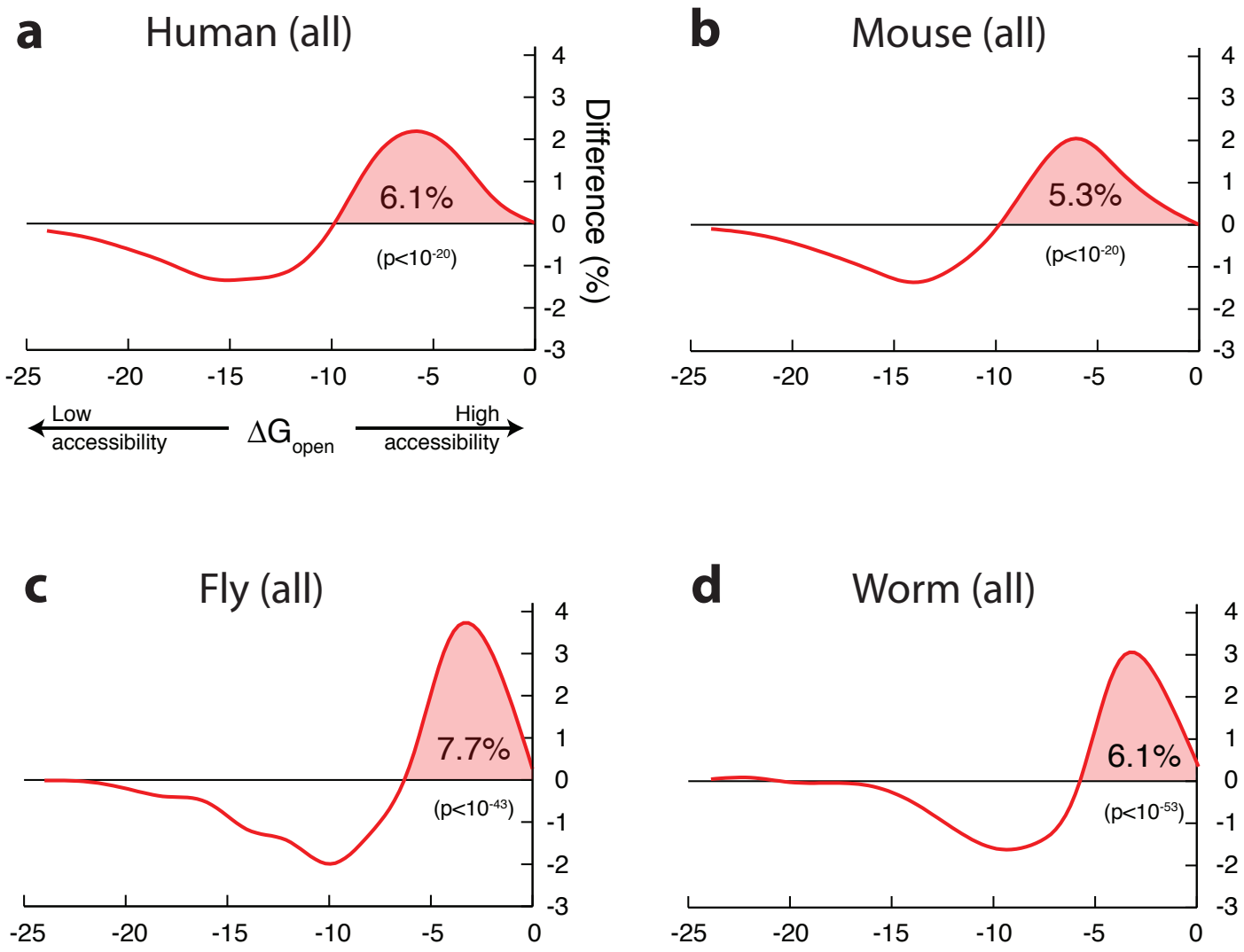
Supplementary Figure 4. Predictive power of STarMir on our experimental data and of our method on STarMir experiments. (a-b) Scatter plot of the score assigned by STarMir to each of our microRNA-target pairs (x-axis) and the observed expression level (y-axis). The correlation between the score given by STarMir and the actual measured expression level of each of our constructs is given along with its p-value (top) for (a) our constructs for *grim*, *hid*, and *rpr* from **Fig. 1**, and (b) our constructs for miR-184 from **Fig. 3a**. In (a), only 8 of our 22 constructs pass the threshold of prediction by STarMir, and in (b) only 16 of our 22 constructs pass the STarMir prediction threshold. (c) plot of the interaction energy assigned by our model, ΔΔG (x-axis), and the reported expression level of the constructs from Long et al. (y-axis). The correlation between our score and the reported expression level is given (top). All eight constructs reported by Long et al. are shown, except for one construct (pVT712) in which we made no predictions as there are no sites that passed our minimal seed requirements. ΔΔG values were calculated with no flank.

Supplementary Figure 5



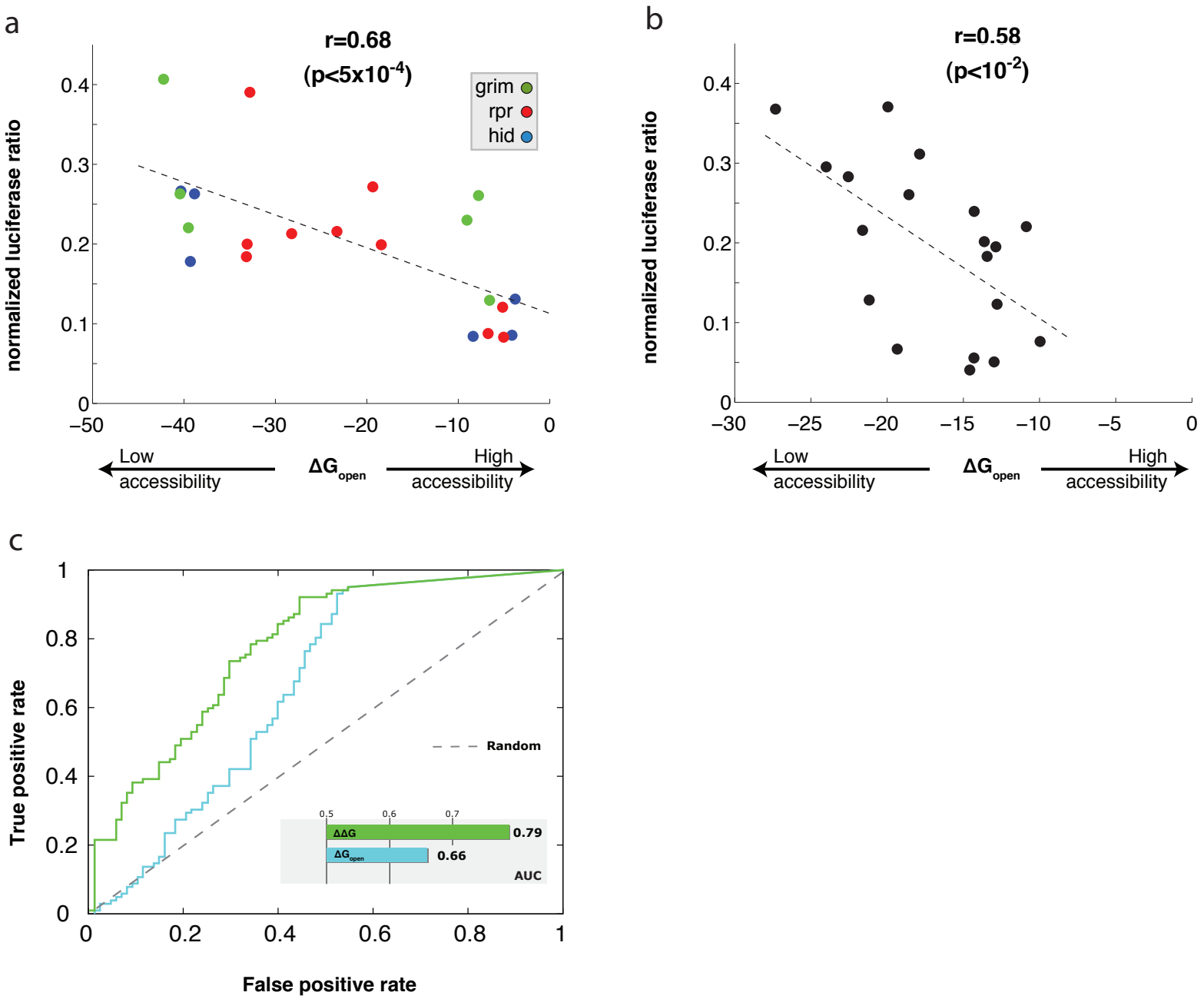
Supplementary Figure 5. Average number of targets per microRNA. Average number of target sites per microRNA predicted by PITA (y-axis) as a function of ΔΔG cutoff (x-axis). In (a) all targets are shown, and in (b) only targets with conservation above 0.9 are shown. (c) Percentage of conserved (>0.9) targets (y-axis) as a function of the ΔΔG cutoff (x-axis).

Supplementary Figure 6



Supplementary Figure 6. microRNA targets in animal genomes are preferentially located in regions of high accessibility. Shown is the difference between the distribution of target accessibility scores, ΔG_{open}, for all microRNA seeds in the 3'UTRs of human (a), mouse (b), fly (c), and worm (d), and the distribution of an equal-sized set of seeds whose locations were chosen at random from each of the respective organisms. For each accessibility value *x* (x-axis), we plot the difference between the fraction of real seeds and the fraction of randomly placed seeds whose accessibility is *x*. Since the ΔG_{open} measure depends on the number of G and C nucleotides, random seed locations were selected such that they have the same G/C content distribution as that of the real seeds. The fraction of real seeds that are more accessible than the random seeds is indicated (shaded area), along with the significance level of the difference between the distributions, as measured by the Kolmogorov-Smirnov test. ΔG_{open} values were calculated with no flank.

Supplementary Figure 7



Supplementary Figure 7. Predictive power of ΔG_{open}. (a) Shown is a scatter plot of the energy required to unpair the target site and make it accessible to the microRNA (x-axis), ΔG_{open}, and the observed expression level (y-axis) of the microRNA target constructs that we measured in **Fig. 1**. ΔG_{open} alone achieves a high correlation score (shown at the top together with its associated p-value). (b) Same as in (a), for the performance on the miR-184 target constructs. (c) Graphs showing the sensitivity and specificity of ΔG_{open} and $\Delta\Delta G$ on all 190 microRNA-target interaction pairs that were experimentally tested in *Drosophila* to date (**Table S3**). In each case, scores were calculated with flank, using the flank sizes specified for the respective data set in **Fig. 2, 3**, and **4**.

Supplementary Table 1: Quantification of transcript abundance of luciferase-UTR fusions. For qPCR, RNA was extracted from S2 cells ($5x10^5$) 20 h after transfection using RNeasy (Quiagen); to avoid contamination with genomic DNA, samples were vortexed (1') prior to loading onto the RNeasy columns. 1µg of total RNA was reverse transcribed using an oligo dT primer and Superscript III (Invitrogen; 20µl reaction volume; 30' at 42°C; termination by heat inactivation 10' at 99°C). qPCR was carried out following the Biorad iQ SYBR Green Supermix protocol using 100 ng cDNA template and 800mM primers designed with VectorNTI software (20-22 bp; Tm 60-62°C; length of amplification product 100-200 bp). We compared the transcript levels of Renilla (FW: TGAGGAG TTCGCTGCCTACC; REV: TGCGGACAATCTGGACGACG) with Firefly (FW: ATCGTGGACCGCCTGAAGTC; REV: ACGACGGCGGCAGGCAGC) luciferase and RP49 (FW:GATGCCCAACATCGGTTACG; REV: TTGTGCACCAGG AACTTCTT) as internal control. Each data point is the result of two independent experiments done in triplicate. Relative changes in gene expression were calculated using the $2^{-\Delta\Delta Ct}$ method, see Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402-8 (2001).

Backbone	Mean Fold changes
Psicheck empty	1.00 (1.02 – 0.98)
RPR_N	0.98 (1.02 – 0.93)
RPR_C	1.02 (1.04 – 1.00)
HID_N	1.01 (1.03 - 1.00)
HID_C	1.01 (1.04 – 1.00)
GRIM_N	1.02 (1.04 – 1.00)
GRIM_C	1.01 (1.04 – 0.98)
HID_N>RPR	0.98 (1.04 – 0.93)
HID_N>GRIM	0.97 (0.99 – 0.94)
HID_Δ5	1.11 (1.17 – 1.05)

Supplementary Table 2

Set of 190 experimentally tested *Drosophila* microRNA:mRNA pairs, with references and indication of functionality (1=functional, 0=nonfunctional)

no.	miRNA	gene name	CG number	functional	UTR length	miRNA promoter	UTR promoter	Assay	Reference
1	let-7	ab	CG4807	1	1671	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Burgler et al, 2005
2	miR-92b	ab	CG4807	1	1671	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Burgler et al, 2005
3	miR-92b	CrebA	CG7450	1	1940	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Burgler et al, 2005
4	miR-34	Eip74EF	CG32180	1	1612	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Burgler et al, 2005
5	miR-92b	ttk	CG1856	1	1697	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Burgler et al, 2005
6	bantam	Mad	CG12399	1	935	actin ~ 200bp	metallothionein	S2 in vitro (Luciferase)	Robins et al, 2005
7	miR-287	CRMP	CG1411	1	2775	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
8	miR-279	SP555	CG14041	1	758	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
9	miR-310	imd	CG5576	1	381	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
10	miR-1	tutl	CG15427	1	1421	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
11	miR-34	Su(z)12	CG8013	1	1129	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
12	miR-12	rt	CG6097	1	174	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
13	miR-124	Gli	CG3903	1	1095	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
14	miR-7	fng	CG10580	1	517	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
15	miR-287	DIP1	CG17686	0	1259	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
16	miR-303	CG14991	CG14991	0	735	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
17	miR-278	tup	CG10619	0	1457	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
18	miR-317	yellow-c	CG4182	0	114	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
19	miR-318	CG13380	CG13380	0	570	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
20	miR-286	boss	CG8285	0	150	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
21	miR-288	CG32057	CG32057	0	1888	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
22	miR-276b	kel	CG7210	0	4284	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
23	miR-316	ia2	CG31795	0	1314	actin ~ 200bp	actin	S2 in vitro (Luciferase)	Robins et al, 2005
24	miR-7	hairy	CG6494	1	816	UAS-DsRed, wing disc	tubulin	in vivo (GFP) wing	Stark et al, 2003
25	miR-2b	grim	CG4345	1	964	UAS-DsRed, wing disc	tubulin	in vivo (GFP) wing	Stark et al, 2003
26	miR-2b	skl	CG13701	1	945	UAS-DsRed, wing disc	tubulin	in vivo (GFP) wing	Stark et al, 2003
27	miR-2	rpr	CG4319	1	485	endogenous	tubulin	S2 in vitro (Luciferase)	Stark et al, 2003
28	bantam	hid	CG5123	1	2269	p-Uast	tubulin	in vivo (GFP) wing	Brennecke et al, 2003
29	miR-7	HLHm3	CG8346	1	442	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
30	miR-7	HLHmg	CG8333	1	139	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
31	miR-7	Brd	CG3096	1	217	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
32	miR-7	m4	CG6099	1		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
33	miR-7	Tom	CG5185	1	414	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
34	miR-7	BobA	CG12487	1	333	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
35	miR-7	HLHm5	CG6096	1	266	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
36	miR-7	Ocho	CG3396	0		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2006
37	miR-4	Tom	CG5185	1	414	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
38	miR-4	HLHmd	CG8328	1	383	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
39	miR-4	HLHmg	CG8333	1	139	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
40	miR-4	ma	CG8337	1		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
41	miR-4	m4	CG6099	1		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
42	miR-4	Brd	CG3096	1	217	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
43	miR-4	HLHm5	CG6096	1	266	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
44	miR-2a-1	ma	CG8337	1		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
45	miR-2a-1	HLHmd	CG8328	1	383	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
46	miR-11	m4	CG6099	1		UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
47	miR-11	BobA	CG12487	1	334	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
48	miR-7	HLHmd	CG8328	1	383	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
49	miR-4	BobA	CG12487	1	335	UAS-DsRed, wing disc	tubulin	in vivo (GFP or lacZ)	Lai et al, 2005
50	miR-4	bap	CG7902	1	165	UAS-DsRed, wing disc	tubulin	in vivo (GFP)	Brennecke et al, 2005
51	miR-6	hid	CG5123	1	2269	endogenous	tubulin	in vivo (GFP and Luciferase)	Leaman et al, 2005
52	miR-6	grim	CG4345	0	964	endogenous	tubulin	in vivo (GFP and Luciferase)	Leaman et al, 2005
53	miR-2	hid	CG5123	0	2269	endogenous	tubulin	in vivo (GFP and Luciferase)	Leaman et al, 2005
54	miR-308	rpr	CG4319	0	485	endogenous	tubulin	in vivo (GFP and Luciferase)	Leaman et al, 2005
55	miR-1	Delta	CG3619	1	2097	UAS-Gal4	tubulin	in vitro (Luciferase)	Kwon et al, 2005
56	iab-4-5p	Ubx	CG10388	1	2396	UAS-DsRed	tubulin	in vivo (GFP or lacZ)	Ronshaugen et al, 2005
57	miR-7	aop	CG3166	1	1681	miRNA depletion	tubulin	in vivo (GFP)	Li et al, 2005
58	miR-278	expanded	CG4114	1	1452	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Teleman et al, 2006
59	miR-9a	sens	CG32120	1	606	H1 promoter, ~ 500bp	SV40	in vitro (Luciferase)	Li et al, 2006
60	bantam	-	CG12187	0	2220	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
61	bantam	S	CG4385	1	1161	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
62	bantam	-	CG5919	0	365	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
63	miR-1	Tom	CG5185	0	414	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
64	miR-1	par-6	CG5884	1	1677	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
65	miR-11	bap	CG7902	1	165	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
66	miR-14	-	CG10473	0	537	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
67	miR-14	dpld	CG1624	0	1524	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
68	miR-14	EcR	CG1765	1	1808	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
69	miR-14	rst	CG4125	0	1608	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
70	miR-14	wg	CG4889	1	1069	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
71	miR-14	par-6	CG5884	0	1677	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
72	miR-14	Eip75B	CG8127	0	1678	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
73	miR-278	rtGEF	CG10043	1	771	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
74	miR-278	Lar	CG10443	0	1400	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
75	miR-278	-	CG11539	0		tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
76	miR-278	-	CG1154	0	691	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
77	miR-278	-	CG12187	0	2221	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005

no.	miRNA	gene name	CG number	functional	UTR length	miRNA promoter	UTR promoter	Assay	Reference
78	miR-278	-	CG14960	1	211	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
79	miR-278	-	CG15097	1	782	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
80	miR-278	-	CG15715	0	281	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
81	miR-278	-	CG15861	0	272	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
82	miR-278	-	CG16975	0	809	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
83	miR-278	Vha100-1	CG1709	0	296	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
84	miR-278	lmg	CG18042	0	1199	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
85	miR-278	-	CG18815	0	392	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
86	miR-278	-	CG1919	0	514	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
87	miR-278	-	CG30089	0	1309	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
88	miR-278	-	CG3026	0	755	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
89	miR-278	Prosap	CG30483	0	847	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
90	miR-278	-	CG31651	0	1727	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
91	miR-278	-	CG31841	0		tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
92	miR-278	-	CG32057	0	1888	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
93	miR-278	Hk	CG32688	0	1200	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
94	miR-278	-	CG33006	0	2333	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
95	miR-278	Gg30A	CG3694	0	639	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
96	miR-278	px	CG4444	0	305	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
97	miR-278	-	CG6930	0	227	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
98	miR-278	-	CG7368	0		tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
99	miR-278	Rab3	CG7576	0	2482	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
100	miR-278	-	CG7656	0	403	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
101	miR-278	-	CG9339	0	2420	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
102	miR-278	DopR	CG9652	0	637	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
103	miR-279	nerfin-1	CG13906	1	1622	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
104	miR-279	JIL-1	CG6297	0	1874	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
105	miR-279	-	CG6520	1	238	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
106	miR-2b	-	CG11293	0	150	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
107	miR-2b	-	CG1969	0	516	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
108	miR-2b	-	CG4269	1	149	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
109	miR-2b	GLaz	CG4604	1	161	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
110	miR-2b	-	CG7956	0	484	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
111	miR-6	dpld	CG1624	0	1524	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
112	miR-6	E(spl)	CG8365	1		tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
113	miR-7	-	CG6520	0	238	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
114	miR-8	fj	CG10917	1	453	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
115	miR-8	pygo	CG11518	0	1064	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
116	miR-8	Cf2	CG11924	0	1732	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
117	miR-8	-	CG18622	0	314	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
118	miR-8	-	CG1882	1	235	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
119	miR-8	disp	CG2019	1	827	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
120	miR-8	rn	CG32466	1	304	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
121	miR-8	rst	CG4125	0	1608	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
122	miR-8	-	CG4484	0	252	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
123	miR-8	wg	CG4889	1	1069	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
124	miR-8	-	CG5735	1	599	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
125	miR-8	-	CG6210	1	614	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
126	miR-8	MIP	CG6456	1	471	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
127	miR-8	-	CG6634	1	897	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
128	miR-8	Tie	CG7525	1	538	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
129	miR-8	Nedd4	CG7555	0	1050	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
130	miR-8	sd	CG8544	0	1288	tubulin, ~ 500bp	tubulin	in vivo (GFP), wing	Stark et al, 2005
131	miR-92b	Tom	CG5185	0	414	tubulin, ~ 500bp	tubulin	in vitro (Luciferase)	Stark et al, 2005
132	miR-2	crq	CG4280	1	767	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
133	miR-2	hep	CG4353	1	1747	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
134	miR-2	ice	CG7788	0	668	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
135	miR-2	scyl	CG7590	1	1046	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
136	miR-2	drpr	CG2086	0	710	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
137	miR-2	kay	CG15509	1	1091	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
138	miR-2	pnr	CG3978	1	605	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, unpublished
139	miR-184	Gli	CG3903	1	1095	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
140	miR-184	Nrx	CG6827	1	765	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
141	miR-184	cora	CG11949	0	438	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
142	miR-184	Lac	CG12369	1	768	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
143	miR-184	ttk	CG1856	1	1697	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
144	miR-184	Cont	CG1084	1	254	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
145	miR-184	nrv2	CG9261	0	878	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
146	miR-184		CG7161	1	102	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
147	miR-184	trio	CG18214	0	2107	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
148	miR-184		CG12789	1	284	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
149	miR-184	activin-beta	CG11062	1	567	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
150	miR-184	chp	CG1744	1	455	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
151	miR-184		CG12880	1	679	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
152	miR-184		CG6583	1	517	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
153	miR-184	rn	CG32466	1	304	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
154	miR-184	sinu	CG10624	1	1282	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
155	miR-184		CG1298	1	168	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
156	miR-184	Mega	CG14779	1	386	endogenous	SV40	S2 in vitro (Luciferase)	Gaul lab, this study
157	miR-9a	bTub60D	CG3401	1	956	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006

no.	miRNA	gene name	CG number	functional	UTR length	miRNA promoter	UTR promoter	Assay	Reference
158	miR-2b		CG4851	1	259	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
159	miR-13a		CG4851	0	259	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
160	miR-2b	Sema-1b	CG6446	0	1473	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
161	miR-14	Sema-1b	CG6446	0	1473	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
162	miR-2b		CG10011	0	1464	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
163	miR-9a		CG10011	1	1464	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
164	miR-9b		CG10011	0	1464	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
165	miR-12		CG10011	1	1464	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
166	miR-9b	Vha68-1	CG12403	1	647	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
167	miR-9a		CG12505	1	1513	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
168	miR-9b		CG12505	0	1513	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
169	miR-14		CG12505	0	1513	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
170	miR-2b		CG15385	0	677	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
171	miR-9b		CG30337	0	302	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
172	miR-12		CG30337	0	302	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
173	miR-2b		CG31886	1	541	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
174	miR-12		CG31886	1	541	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
175	miR-9a		CG33087	1	323	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
176	miR-9b		CG33087	0	323	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
177	miR-9a	nerfin-1	CG13906	0	1622	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
178	miR-9b	nerfin-1	CG13906	1	1622	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
179	miR-9a		CG1172	1	1009	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
180	miR-9b		CG1172	0	1009	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
181	miR-13a	Reg-5	CG2928	0	285	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
182	miR-13a		CG3799	1	260	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
183	miR-13a		CG5977	0	1030	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
184	miR-13a	Tab2	CG7417	0	391	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
185	miR-9b	htt	CG9995	0	737	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
186	miR-9a	brat	CG10719	0	1313	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
187	miR-13a		CG18604	0	1017	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
188	miR-14		CG18604	0	1017	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
189	miR-9a		CG31712	1	507	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006
190	miR-9b		CG31712	0	507	actin ~ 200bp	tubulin	S2 in vitro (Luciferase)	Rehwinkel et al, 2006

References

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113, 25-36 (2003).

Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol* 3, e85 (2005).

Burgler, C. & Macdonald, P. M. Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. *BMC Genomics* 6, 88 (2005).

Kwon, C., Han, Z., Olson, E. N. & Srivastava, D. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. *Proc Natl Acad Sci U S A* 102, 18986-91 (2005).

Lai, E. C., Tam, B. & Rubin, G. M. Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev* 19, 1067-80 (2005).

Leaman, D., Chen, P.Y., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., Marks, D.S., Sander, C., Tuschl, T. & Gaul, U. Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. *Cell* 121, 1097-108 (2005).

Li, X. & Carthew, R. W. A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. *Cell* 123, 1267-77 (2005).

Li, Y., Wang, F., Lee, J. A. & Gao, F. B. MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. *Genes Dev* 20, 2793-805 (2006).

Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S. M. & Izaurralde, E. Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. *Mol Cell Biol* 26, 2965-75 (2006).

Robins, H., Li, Y. & Padgett, R. W. Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci U S A* 102, 4006-9 (2005).

Ronshaugen, M., Biemar, F., Piel, J., Levine, M. & Lai, E. C. The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev* 19, 2947-52 (2005).

Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of Drosophila MicroRNA targets. *PLoS Biol* 1, E60 (2003).

Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* 123, 1133-46 (2005).

Teleman, A. A., Maitra, S. & Cohen, S. M. Drosophila lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev* 20, 417-22 (2006).

Supplementary Table 3: Overlap between PITA predictions and other microRNA target prediction methods, and among other methods. (a) Comparison between PITA predictions and other methods. For each comparison, shown is the number of targets predicted by the compared method, and the number of these targets that are also predicted by PITA (overlap column). For the PITA predictions, we selected its k highest ranking targets, where k in each comparison is equal to the number of targets predicted by the compared method. (b) Comparison between existing microRNA target prediction methods. For each pair of compared methods, shown is the number of targets of each method, and the overlap. The overlap percentage is given in respect to each of the compared prediction set sizes.

(a)

Organism	Method	Num. Targets	Overlap	Overlap (%)
Human	PicTar	55,840	14,362	25
	TargetScan	40,482	7,435	18
	miRanda	5,895	386	6
Mouse	PicTar	51,895	8,573	16
	TargetScan	31,683	2,983	9
Fly	Stark	12,619	3,877	30
	PicTar (s1)	10,105	2,869	28
	PicTar (s3)	5,224	1,023	19
	miRanda	10,230	2,660	26

(b)

Organism	Method I	N. Targets I	Method II	N. Targets II	Overlap	Overlap (%)
Human	PicTar	55,840	TargetScan	40,482	13,010	23 / 32
	PicTar	55,840	miRanda	5,895	2,148	3 / 36
	TargetScan	40,482	miRanda	5,895	880	2 / 14
Mouse	PicTar	51,895	TargetScan	31,683	9,594	14 / 23
Fly	Stark	12,619	PicTar (s1)	10,105	3,455	51 / 64
	Stark	12,619	PicTar (s3)	5,224	3,773	29 / 72
	Stark	12,619	miRanda	10,230	3,164	25 / 30
	PicTar (s1)	10,105	miRanda	10,230	3,455	34 / 33
	PicTar (s3)	5,224	miRanda	10,230	2,178	41 / 21

Supplementary Table 4: Sequences of constructs used in our experiments.
Sequences are given in FASTA format, with putative microRNA binding site marked in green.

>grim_UTR
GGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTGGGAAAGGCAGGCTCAA
TCAAAGCGCATTGTGATTTCTTTTTTGGGTATGGACACATGAGGAGGAACGAGTTTTGTAAAAC
ATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTAAACTATTTTTGAGAGAGAGAAAATT
CAAAATTTGTCATCGACTTGCTTAAAGTGATTTTGGTTGTTTGTTACTTTTTTGGATGTCAAAT
TTTCCGATATAAGTTTTGAAAAAATGTTTAATTTTGCAAGCAAAGGTTTCCCGGTTGGTTCAAA
TACATAAACATACTTAACAACATTTAGAGAGCATAAAGAGAAATGTGTAAAATTGTCATAAACC
TTATTAGTGTCTATACGTATGAGCAAAACCACAAGTATCGTACTCAACTAACAAATAATTGTCA
AATTTTCTACTTCAACTACTTCAACTACATTTCAACTACTTACAAATATAAATCGTAAACTAAT
ATTACAGAACTACAATACAGAACCAACAACAAAACAAAGATGCAACCGCAGCAACAAGAATAAC
AAAAGACAACAACTCAATTGTACAAAATAGATTTTAAGCAAACAGCAAATTTTATATGTTTACA
ACAAAGAGAGAAAATACAAAACGTCTACAAAAAATAGAATACAAGAACCGCCGAACAGAATACA
TTTTAACAAAAATTAAACAAAACAACAACAAACATTTAAGCAATTTTGATACGAATAATCAAGA
ATAACAAAATTTGAAATGTATCTTATCATTATTATTGATTTTAGCAAAATTCAAAACATATTAT
TTTAATTTTGGCATTGCAAATTAAACAAGAAATATTGAAATTAATCACATGTAAAGATAGCATT
TGTAAGAAATTAAAATTTTTATAAGACAAACAAGCAAAAAATAAATACAATTTTATAAAAAATT

>grim_N
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGAAGGCAGGC
TCAATCAAAGCGCATTGTGATTTCTTTTTTTGGGTGCGGCCGCATGAGGAGGAACGAGTTTTGTA
AAACATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTAAACTATTTTTGAGAGAGAGAA
AATTCAAAATTTGT

>grim_C
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGAAGGCAAAT
CACAATGCGCTTTGATTGAGCCGGTTTGGCTCAATCAAAGCGCATTGTGATTTCTTTTTTTGGGT
GCGGCCGCATGAGGAGGAACGAGTTTTGTAAAACATGGGATATTTCCACAACAAAATAAGAAAA
CTATTTTTAAACTA

>grim_Δ5
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGAAGGCAGGC
TCAATCAAAGCGCATTCTTTTTTGGGTGCGGCCGCATGAGGAGGAACGAGTTTTGTAAAACATG
GGATATTTCCACAACAAAATAAGAAAACTATTTTTAAACTATTTTTGAGAGAGAGAAAATTCAA
AATTTGT

>grim_Δ
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGGCGGCCGCA
TGAGGAGGAACGAGTTTTGTAAAACATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTA
AACTATTTTTGAGAGAGAGAAAATTCAAAATTTGT

>hid_UTR
AAGCGCAGGAGACGTGTAATCGAATGATCTATAGTGAAATCAGCTAGCCCTTAAGATATATGCC
GATCTAAACATAGTTGTAGTTAAACCGTACATAAGTGCAACGAATTTATTGAACTGCAGGAGCG
AAAGCAGAAAGTCATTAATTCGTAAACGGATTGTTAGATACACAAACAGCCAACATACACGAAG
AGTGTGCCTAAGATTAAGAAGGTTGACGGGACACAAGAACAATATATTCTATCTGTCTATGGTA

ACTGCATTTGTATTTCTAAAACGAAACGAAAGATAACAATCTTAACTGCTCAAAGTAATGAAAA
CTCTTAGACTGGCAAGAGACTCAAATCACACTTATTTTTTTGCTGATCCATATTTTTGTACAAC
CTTTTGAGCGATATTTACAAATTATACTAGTACAAAAAAAAGAGAGAGAGAGATAAGCAAAAGA
AAACTGCCACTTTTGAGATACTTTTGATAATCTTTGATTTGCATTTAATCATTTCCACACTTGC
ATTTTTTATAAACAACAAACAAAATTACTTCCATTGTAGAACAAAGTAAACTGCAATTTCAATG
TCTTCGCATTTGTAATTCCGAATTGCAAGAAAAACAAAAATATTTTAAATATGTTTAACTAGTA
GAATTTTTTAAACGTAAGTCCACAAAAACAAGCACATCTAGCTTTAATTGTTGAAACAAAAGCA
GAAAAAACGCAACAAAAAAATGAATGAAAATCATTAAATTAATTTTGTATATAGTTTTTATGCC
ATTTTTGTGATGTTTTGTGTCTACGGTTTATGTCATGTTATTTTAGTTAAATTTCTTATGATTT
ATGTTTATTTGTAATATTTTTTGTCATTGTTTGTTCATCATCATATTCAAATTGGTCTCACAAT
ATAATAGTTTTAAGCTCCACGCCCGGGAGATTGATGGCAAAACGATTGAAATTTGGCCAGAAGA
GAGATAGTTTTCCCCATTCGTACACAGTCTTTTTTGGAATGCACATTAATGATCTCTCACAATG
GAAATTAATGAAAATTGATCTCCGCAGCTAGCCAAAGTTAAAAAAGAAATGAAGAGGAAAACAT
ATTCTATAGGCAATTTTCACTATATGCTAGAATTTCCCGGGCGTTTCAATGCTAATCGAATACA
GTGACATGAAAGCAAACATAGCGAAAATATTAAGAAAATCAATCAAAAAGAAAGAAAAACCAAT
TCCCAAAAATCGCATTGATCTCATGGATTTATACAATACAATTACATCAACCGTTTTTTTACAA
TGAGAAATGTTATAAAAAGCAGAAAGTGAAACACAGAAACATAAACAAAAATTAACGAAAAGCT
TAGATATAAGTTCGCCAAGCGTTTTAGTTCTATTTTCTAGAATGTCTAAGTTGGTTTAGTGAGT
TTATTAAGCTGTCTTCGGACACAAGTTTATTTGTATATAAGCAATATTATTTGTGTAGCCTAAG
TGACAGTCCCAATCAAATCCAATCCAATATCACCCAGTCCCGGACATTTCCCAGCAAAACAATA
GACTATTCTCGCGTTCACATGTATCAATCTTAATTTGAATTACCACAAAATGAAATGAAATACT
AAAACCATACACAAATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTA
GATTAAGGAATCATCCTTGCTTTATCCCCTGCTTATTGCTAATTAGTTTTCACAATGATCTCGG
TAAAGTTTTGTGGCCTTGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTCGAAC
AAAAAGTTAATGAAAAACGAAGCAAATGGAAAAAAAATCAGAATGAAACACAAGAAATTTATAT
TTTTGACCCAATGCTACTTAATCCGTTTTTGTAATTTAAGTATCTTTACTCGACCTTGTATATA
GCGCAGTTCGAATCACAGAATCAAATGCCATTTTTGTATAGAATTTTATTTGGTGCCAAAACAG
TGACAGATAATTAAATGTCTATGAACCCGTGTATTTCGCATATTATACATTTATACATATATCG
TAACTTCAATGATAAGTTTGATTCTGAAATTTTGTCAACTCAATTTAAGAAACATTTCTGTTGT
AGTTTAGTGATTGCTAGCAGAAAGCACTTTGTTTAATTGTACATTTTATATTATGCTGTAATAT
TTTAATATACATAAATATCATTATTGATCTCATGAATATGTTCATAAGACAACAAAAATTATAT
ATATGAATACATCTATGTGTATGTGTAAA

>hid_N
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGTTATCCCCTGCTTATTGCTAATTAGTTTTCACAATGATCTCGGTAAAGTTTTGCGG
CCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTCGAACAAAAAGTTAATGA
AAAACGAAGCAAA

>hid_C
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGTTATCCCCTGCTTATTGCTAATTAGTTTTCACAATGATCTCGGACTACCGAGATCA
TTGTGAAAACTAATTAGTAAAGTTTTGCGGCCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTT
GCCATAAATACTC

>hid_Δ5
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGTTATCCCCTGCTTATTGCTAATTAGTTTTCACGTAAAGTTTTGCGGCCGCGCGCCC

AAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTCGAACAAAAAGTTAATGAAAAACGAAGC
AAA

>hid_Δ
TCGACAATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGG
AATCATCCTCGAGGCGGCCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTC
GAACAAAAAGTTAATGAAAAACGAAGCAAA

>rpr_N
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGATAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_C
ACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAGCCACAACACCCACCCAATTTTAGTTTT
TATCACAATCGCTTTGATGAGTAATAACTCATCAAAGCGATTGTGATAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_C3
TCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAGCCACAACACCCACCCAA
TTTTAGTTTCGCTTTGATGAGTTCTTACTCATCAAAGCGATTGTGATAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_C3+
ACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAGCCACAACACCCACCCACGCTTTGATGA
GTAAACTAAAATTCTTATTTTAGTTTACTCATCAAAGCGATTGTGATAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_C5
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTACCGGACTCATCAAAGCGATTGTGATAATTCTTTATCACAATC
CTGCGTTCCCAATGGTTTTGTTTCTACAAAAAAGCGGCCGCAAAAGGCCAAAAATAACGTTTTT
ATAATCACCCC

>rpr_C5+
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGATAATGGTTTTGTTTTTCT
AAACAAAACCATTATCACAATCTACAAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTA
TAAAGTCCCCA

>rpr_M2
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGACAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_M3
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGCTAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_M6
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTATGATAATGGTTTTGTTTCTAC
AAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAA
TGTTTCAATGA

>rpr_I5
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGGATAATGGTTTTGTTTCTA
CAAAAAAGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAA
ATGTTTCAATGA

>rpr_Δ5
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATGGTTTTGTTTCTACAAAAAAGCG
GCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAAT
GA

>rpr_Δ
TCGTCGACCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGA
CCTCGAGGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAA
ATGTTTCAATGA

>grim_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACcTCGAG
AAGGCAGGCTCAATCAAAGCGCATTGTGATTTCTTTTTTTGGGTGCGGCCGCAAAATTTGAAAAA
ATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>grim_C>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACcTCGAG
AAGGCAAATCACAATGCGCTTTGATTGAGCCGGTTTGGCTCAATCAAAGCGCATTGTGATTTCT
TTTTTGGGTGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAA
AAATGTTTCAATGA

>grim_Δ5>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
AAGGCAGGCTCAATCAAAGCGCATTCTTTTTTGGGTGCGGCCGCAAAATTTGAAAAAATAACGT
TTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>grim_N>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGAAGGCAGGCTCAATCAAAGCGCATTGTGATTTCTTTTTTTGGGTGCGGCCGCGCGCC

CAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTCGAACAAAAAGTTAATGAAAAACGAAG
CAAA

>grim_C>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGAAGGCAAATCACAATGCGCTTTGATTGAGCCGGTTTGGCTCAATCAAAGCGCATTG
TGATTTCTTTTTTGGGTGCGGCCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAAT
ACTCGAACAAAAAGTTAATGAAAAACGAAGCAAA

>grim_Δ5>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGAAGGCAGGCTCAATCAAAGCGCATTCTTTTTTGGGTGCGGCCGCGCGCCCAAAAGT
CGTACAGATTTTTGGTTTGCCATAAATACTCGAACAAAAAGTTAATGAAAAACGAAGCAA

>hid_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACcTCGAG
TTATCCCCTGCTTATTGCTAATTAGTTTTCACAATGATCTCGGTAAAGTTTTGCGGCCGCAAAA
TTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>hid_C>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACcTCGAG
TTATCCCCTGCTTATTGCTAATTAGTTTTCACAATGATCTCGGACTACCGAGATCATTGTGAAA
ACTAATTAGTAAAGTTTTGCGGCCGCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAAT
TTTTTACAAAAATGTTTCAATGA

>hid_Δ5> rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
TTATCCCCTGCTTATTGCTAATTAGTTTTCACGTAAAGTTTTGCGGCCGCAAAATTTGAAAAAA
TAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>hid_N>grim
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGTTATCCCCT
GCTTATTGCTAATTAGTTTTCACAATGATCTCGGTAAAGTTTTGCGGCCGCATGAGGAGGAACG
AGTTTTGTAAAACATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTAAACTATTTTTGA
GAGAGAGAAAATTCAAAATTTGT

>hid_C>grim
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGTTATCCCCT
GCTTATTGCTAATTAGTTTTCACAATGATCTCGGACTACCGAGATCATTGTGAAAACTAATTAG
TAAAGTTTTGCGGCCGCATGAGGAGGAACGAGTTTTGTAAAACATGGGATATTTCCACAACAAA
ATAAGAAAACTATTTTTAAACTATTTTTGAGAGAGAGAAAATTCAAAATTTGT

>hid_Δ5> grim
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGTTATCCCCT
GCTTATTGCTAATTAGTTTTCACGTAAAGTTTTGCGGCCGCATGAGGAGGAACGAGTTTTGTAA
AACATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTAAACT

>rpr_N>grim

ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGCCACAACAC
CCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGATAATGGTTTTGTTTCTACAAAAAAGCG
GCCGCATGAGGAGGAACGAGTTTTGTAAAACATGGGATATTTCCACAACAAAATAAGAAAACTA
TTTTTAAACTATTTTTGAGAGAGAGAAAATTCAAAATTTGT

>rpr_C>grim
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGCCACAACAC
CCACCCAATTTTAGTTTTTATCACAATCGCTTTGATGAGTAATAACTCATCAAAGCGATTGTGA
TAATGGTTTTGTTTCTACAAAAAAGCGGCCGCATGAGGAGGAACGAGTTTTGTAAAACATGGGA
TATTTCCACAACAAAATAAGAAAACTATTTTTAAACTATTTTTGAGAGAGAGAAAATTCAAAAT
TTGT

>rpr_Δ5>grim
ATTTGGTTTCCATATTTCATCCTGGTGGAGAGAAAATCTTTGGGATTTTCTCGAGCCACAACAC
CCACCCAATTTTAGTTTACTCATCAAAGCGATGGTTTTGTTTCTACAAAAAAGCGGCCGCATGA
GGAGGAACGAGTTTTGTAAAACATGGGATATTTCCACAACAAAATAAGAAAACTATTTTTAAAC
T

>rpr_N>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGCCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATTGTGATAATGGTTTTT
GTTTCTACAAAAAAGCGGCCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACT
CGAACAAAAAGTTAATGAAAAACGAAGCAAA

>rpr_C>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGCCACAACACCCACCCAATTTTAGTTTTTATCACAATCGCTTTGATGAGTAATAACT
CATCAAAGCGATTGTGATAATGGTTTTGTTTCTACAAAAAAGCGGCCGCGCGCCCAAAAGTCGT
ACAGATTTTTGGTTTGCCATAAATACTCGAACAAAAAGTTAATGAAAAACGAAGCAAA

>rpr_Δ5>hid
AATGAAAAATTATTTTTGTAAATTGTTTGCATCAAGTGAGCAAGGGGATTAGATTAAGGAATCA
TCCTCGAGCCACAACACCCACCCAATTTTAGTTTACTCATCAAAGCGATGGTTTTGTTTCTACA
AAAAAGCGGCCGCGCGCCCAAAAGTCGTACAGATTTTTGGTTTGCCATAAATACTCGAACAAAA
AGTTAATGAAAAACGAAGCAA

>gli_N
CCTCAGACATCCGTATAAGTAACAGCTCTCTAAGCGCAGGATGCGCTTAGGGAGTCAGCCGAAT
CGTCCAATTACAGAGCCAATCAACAATCCAGGAGGAAGTATGCCAGTTTCTCCTGGAATTTTTC
TTTAAAGTAAGTTAGTTATTTTCGAAAAAAGTAGTTTATGTTTTGTATTTATGATAGGTTTGTA
AACAGATCTTTGAG

>nrx_N
TTATGTTTTTTTTTTTAGTCTTTGGTTCTGTCTGACGGCATTATTGAAATTTCTTCTATCACAG
AATACCGTCCATTTTTAGTTAATTTTATGAACTTTTTGTTTGCAAAATTGAGTTAAAAACTTAT
TTAAGGGAGCTGCCCGAGTAGCTCAACAAAAGCCACCTATTTTATGTATTTCTA

>ttk_N
CTCTGAACTGTAGAATACATCATAGCAACATAAGCAATCGTGGCACTCAAACGCAACACACCAG
ACTCAAAACGATTTTTAGTTAATGTAAACCCGTAAGCAATATTAATTCTAACACTTGCTCCGTC
CAGAACGCCTTTTTTAAGCAACACTAGTATTTAATTTACCTAGGACATAAGTTTTGCCCAATT

>cont_N
AATGAATGCAGCACAAATTAAGATCCATAACTTTAGGCAGTAAGAACATACATATACTAAGAAG
TGACCTTATATTTGAGTAAGTGAAGTTATAAAAAGAACTACATTCCATCCGTGCTCTCTTGCAA
GCCCCTTAATTGATACAAATGCAATTCCGTCCATTTTATAGACTTTTATCAAAAAATAATTTCT
AAAATTTTATCAACAAAATAAATATGAATACCATTATATGGTCCAGTATGTTGCTGAACTAAAA
TACCAAGGG

>trio_N
TTCCAAATGATAAGGCAGGAAGCTTAAAGGCAAGCAAGAACAACAAACGAAAGTATGAGAAAAT
GATATTCAAATGTGTTTTGGGCAAAGGTGTTCCGCCACCGAAGTGAGCATTAGCCTCGCACTTG
TTTTTCCGTCTGTGCGTTTGTTTACGGATAGTTTTGCTCACTCACACCCAGAAACACTCGCATA
CTCAATCCTAATTTGTATGTAACTTGTAATTGTAATTTATTTATTCGTCATCAATTGCGTTTCC
GTTCAGTCCACC

>CG12789_N
AGCCATCAGCGACAATAAGCAGGATTGAACAGAACCGCCACATTCACGCCATTTAGATAGTTAA
GCCGTCCTAGTGCTTTAGTCGTAGCTCATAGTTTAGTGACACTCTAAAACACCCAACTGAAAAT
TG

>CG12880_N
GTTCTACCAGAACATCATCTCGTAATCCAACGGCATACCCTCGGTATACCCAGGGAATACCCTT
GAAAGACTGCGGTGCCGGGCGTTCTATTTTCAATATTCTAACGAAGGGGAGCTTCTCCCCTCTT
CCGTCCAGTTTCCACTTGGAATACCCAATCCTGGCGAATCCCGATCACATCTATAGCATCCTGG
CCAAATGTTTAACAATTGCGATG

>rn_N
TTTGTAATTTTTGTAGATCTTCTTGGGTTAAGTTAAGGGTGTTCGTTCTCGCGCACTCGATAAG
CACTATGTAAATGGCCAACGATACGTACTCTAATTTGACAAACACCGTCCATACCTATGTTTAC
ACAATGTCTTGGAATACAAGCACCAGTTCAGGTCCAGATA

>sinu_N
CCTTAATTAAGTGACCCAGAGAAAGGTAAGCGAAACTCCGCCCATTAGTTTTAAGAACAAAACA
TTCCCTTTTGAACTCGAATATTCCGCACAAGTTACCTTCTTTTAAAGGTAACAATTAGGTCTGT
ATTAATCATCCGTCCTTTGCACGTATTTGTAACATTTTCTAAACCGAGAATTTAGCAATGTATT
AAAGCAATACATACGTCTTTTACCGTCCCATATAC

>mega_N
GTCCGCCCGACTTCCATACATTCCGTTCCGATTGGATCTTAGCTTAAGTTTTCAGACGCAATCT
CGTGCCTCGAATCTGCGCAATCGACGCCGTCCATAGTGCCTTTAACAATCGACCATATGTATCT
ATATACACGCCGACTCAGCCGAGATCAGCTCGTATCCGTTTGTAATTGTATTA

>gli_N>rpr

CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
GCGCAGGATGCGCTTAGGGAGTCAGCCGAATCGTCCAATTACGCGGCCGCAAAATTTGAAAAAA
TAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>nrx_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
TATTGAAATTTCTTCTATCACAGAATACCGTCCATTTTTAGCGGCCGCAAAATTTGAAAAAATA
ACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>cora_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
AAATTCCCAATTATCCGTTACCGCAACCGTCTACGAGAAAATGCGGCCGCAAAATTTGAAAAAA
TAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>ttk_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
AATGTAAACCCGTAAGCAATATTAATTCTAACACTTGCTCCGTCCAGGCGGCCGCAAAATTTGA
AAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>cont_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CAAGCCCCTTAATTGATACAAATGCAATTCCGTCCATTTTAGCGGCCGCAAAATTTGAAAAAAT
AACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>trio_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CATTAGCCTCGCACTTGTTTTTCCGTCTGTGCGTTTGTTTGCGGCCGCAAAATTTGAAAAAATA
ACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>CG12789_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CCGCCACATTCACGCCATTTAGATAGTTAAGCCGTCCTAGTGCTTTAGTCGCGGCCGCAAAATT
TGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>vvl_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
ATCCAGGAGTCGAACTCAGCTGCAGCTGCGTCCACTCCTGCGGCCGCAAAATTTGAAAAAATAA
CGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>CG12880_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
TCAATATTCTAACGAAGGGGAGCTTCTCCCCTCTTCCGTCCAGTTTCCACGCGGCCGCAAAATT
TGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>sinu_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
GTAACAATTAGGTCTGTATTAATCATCCGTCCTTTGCGGCCGCAAAATTTGAAAAAATAACGTT
TTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>mega_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
ATCTCGTGCCTCGAATCTGCGCAATCGACGCCGTCCATAGTGCGGCCGCAAAATTTGAAAAAAT
AACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>rn_N>rpr
CCAAACACAGTCACCACTCATCAGCCGACGGCACTCGATTTCTACTGCAGTCAAGGACCTCGAG
CACTATGTAAATGGCCAACGATACGTACTCTAATTTGACAAACACCGTCCATACCTATGCGGCC
GCAAAATTTGAAAAAATAACGTTTTTATAAAGTCCCCAATTTTTTACAAAAATGTTTCAATGA

>gli_UTR
CCTCAGACATCCGTATAAGTAACAGCTCTCTAAGCGCAGGATGCGCTTAGGGAGTCAGCCGAAT
CGTCCAATTACAGAGCCAATCAACAATCCAGGAGGAAGTATGCCAGTTTCTCCTGGAATTTTTC
TTTAAAGTAAGTTAGTTATTTTCGAAAAAAGTAGTTTATGTTTTGTATTTATGATAGGTTTGTA
AACAGATCTTTGAGATCTCAGTTTTGAGTTGTAGTTTAGGCAAACCTGTATTTCTAATGCCTTA
CTTTAAGTTAAATTGTTCGTAAGTTTGATAACCCAACGAATACAAATCCATATCAATGACATTC
CATAGCCTTATCATCGGCCTTTACGTGTATATATAGAATTTAAATGTATAATAAGAATTATTTA
AATCTAGGAAATTGCCTGCCTTAACTGAAGGAAATTAGCTATATAATACAAGTATTTTTATAAC
AAATGCTGAGCTTAAATTTTTAAAACTCACTTTTAGTGGCAGTAATATCGAATGAAATTGTAAA
GAGAGTATATCTTTTAAATCAAATATTAAAATACAAAAATAAATTAACCTTATAAAAAATAATG
CATTACATTATTTAAGACCTGCTATTATTTGATAACAAACTCATAACATTGTCACGTCCCGTAT
TATAAGAATATCATTAGAATACCATTAGATCGTAGCCACTCCCACACATCATCAAGAGCAATAC
ACCTACAGTGCCTTAATAATATTTGGTAATTGTCCTCGATGGACCCACCCACCCCATCCATGAT
GTACCAACTAAGTGTGCCTTTCATACATACATAAATGTACTCCTAGTTATATACTATATATGTT
CTATATGTTCTGTGTGTCGCATATAAGCAAAATGACCATACAAAATCAGTGTGAACCGTCCAAT
AACTTATGTTTTCAAAAGTGTACATTGTTTTTCCTGTATATATCATCATTTGTTACACGATTGA
AAGCATATTGTTTTTACCACATTATATGTTTATATGTACTTTAGATAAGTTTTATGTATTTGTC
CATAA

>CG7161_UTR
TGATGGAAGACTTTTGAAATCGACATAAGAAGTGTGTGCAAATATATAATTTCCGTCCAAAACT
GTACAAATAGGAACTATACGCAAATAAATACGGTGTTACGCCCAGGTTTCCTCCTCCCTTCA

>trio_UTR
ACCAAAAACAAATGAACGCCGTTGGCGAAAGAATTCCAAATGATAAGGCAGGAAGCTTAAAGGC
AAGCAAGAACAACAAACGAAAGTATGAGAAAATGATATTCAAATGTGTTTTGGGCAAAGGTGTT
CCGCCACCGAAGTGAGCATTAGCCTCGCACTTGTTTTTCCGTCTGTGCGTTTGTTTACGGATAG
TTTTGCTCACTCACACCCAGAAACACTCGCATACTCAATCCTAATTTGTATGTAACTTGTAATT
GTAATTTATTTATTCGTCATCAATTGCGTTTCCGTTCAGTCCACCTTTAGATTTTAGTTGGGCC
GCCTTTCTGTTCGTATAGTTTTTGTTTTCTATTCCGGATTTGTGTGTATGAAACGCAAAGTATT
ATATGCTCAAGATTACCTTTGACATTGTGTGCATTTTTAACTTATTTAGACTCCTTTGCTCTTA
GTTTTACTTTTCCGAGTCTGATTTACTAACGATCCGCAGCCGAGAACTACTCAGAATCAAAAGC
AATTTACATTTCGTTTTCGTTTCGTACTCGTTTTATTCGCATTGAACGCAGAGAGTAAGAGTAT
AACTATTCTAGCCTTAGTTTGGTACATAGTTTTAGTTAAACTGCTAAATGTTAATCATAGTTTA
TACTTGAGTTACTGGATTTGCTTTGCAAAACTAAAGCAGAGATTTGTGTAATGGCACTCGTTCG
AGTGTGACAACTTCGTGAAATAATTTAAGCTTAACTTTAGTTTATGAATGCAATGCAAAGCTTA
TTTGTACTTTATAATTTACAAATAAAATATGTAAGCTTTAAACGTAATCAAAACCAGAAACAAT
AAGTTTTTAATTATTCTAATTTTAAACTATGCCCGAAATAAACATAATGCCTGCTCGGATGTAA
TTATTATTTTTTGGTTTTCGTATTCTTTGCCCCATTTTCGGTTGTTTTGTTCACAAAAAAGTCC

ACAACACTTATAGGCTAAGCAGTCAAGTTTGAAGCCCGAACACCCAGATGATAGCACTTAAGCC
AATCTAAACTGTATCCGAGACGATTTTCGAAGCATATTTCTCGTTTTGGTGTCTGTGCGTTTAT
GTTAGTTGGTAAGCAGGAGCCAGTTGGGCCAATGTACAGTGTAGCATACGAGGATAAGTCAGAG
TCACTCAAATGCATGTGTCTCTTGCAAACAAATTACTCAAACAGGACCAGGATTTGATTAAACG
ATTATATGTATAGAACACGAGACGAAGTTACTAAGCGGTTGTCCATCCAATGCTTTCAATCCGA
TTTTGTCTCTTAGAAGCACCCACACACACACACCGAGACGCCGGTTACAGTCTGGAGCATGAAT
ATAAATGCCAGATGCATATATCAATGCGAGGAAAACAATAATGATTTAGCAAAGCTAAATGTAG
GAAGCAAACATGAAATGTATTTTTTTACAACATTTGTGAAACGAACCTCCGGCATTATTCGTAA
TTAAATGTTAAAGTCGCGACGACGGATCGGATCGGATGGTAGGATCGTAGGTCGAAGGATCGAA
GGATCGATGGATCAAATGATCCGATGATCAGATGACAAGCAAACGCACAGTGAGGGGAACCCAA
TAGCCACAACATTTTCTACATTTTTGCACTTTACTCGTGGAGGGAGCTCTACTCAACTCACAAC
CGATCGACCGTAACATGTAACATACTAAATTCTTTTCGATTGTATTAATGGCACTTACCGATTA
ACTAATTGTTGTACTTAAGCCCTGTCGAAATTCTAGATAATTGTGGTAATTAAAGTGTGTGTTT
TAGTTTATCTAATTACTCTGACTAACTATAAATATATTAATTACAACTGCGACATCATGTGCGT
CAGTGCAATGCGAACAGCATAAAGTAAACGTGTTAATCTGTGTTTTAGTCAAGTAAGCATCAGC
TTTCCCTCGGAAAATGCCCGGAAAAGCCTGGCGAGATAAACTTTTGGCCCGGGCGCAGGCTGCA
AATAAATATAATTATATCCAGAAACCACTATGCAAATTGTATGTAACCAACAAAGCATGTAATC
TGAATTTAAATTTTTCAAAGTGAATAATAACAAAGCAAAAGTAGCTGAATAAACTTAACATACT
CAAA

>CG12789_UTR
AGCCATCAGCGACAATAAGCAGGATTGAACAGAACCGCCACATTCACGCCATTTAGATAGTTAA
GCCGTCCTAGTGCTTTAGTCGTAGCTCATAGTTTAGTGACACTCTAAAACACCCAACTGAAAAT
TGCAAGCATTCCCGTGCCTATGCCAAAGATTAGATAGCCTTCGAAGTCGAAACGTCTGGTGGAC
TTCGCTTAGCTAACGAATACTCTTTTTACATTCCTTAGTTATTAGATATGTTTTGTATTAAAAT
GATTTCTTGACCCATGGCTTTCGTGTCAAATAAAATCTTCAACTAA

>activin-beta_UTR
TGATGAGTGTGGTTGCCCTTAGCCTTGCCTTACAATTTTATATTTTCCGTCCGATAGAAATAAA
AATATATGTGTCCTAACTGAGCGCCAATCTCTTAACGAAATCTTTTACTTTTAAGTTAAAAGTA
CATAAATATATACATAAAGTAGACTCAATTTTATTTTATACTTAGCTATGCTGGTGACAATATT
TGTATATTTACGAACAAATCCAAATTGAGGAAGTGCCTAAATTACGTTAATGAAATATTTGTAC
ATTTTAAGAATATTTCAAAAATCACTAAAATATCTTTGTACTAATAAAATTCGATATTTTTAAT
TATTACTTTATTTAATATGTTCAAATTAACATAAGTATAAAACGAAATGATTTAATAACCTATA
ACATGAGCAAAGCGTCGCGCTTTTTTTATTGTCAAAACATTAATTTTACTAACTTGAAAAGCTT
ATATCACAGACATGTAATAAATATTTCATATTACAGTTTAATAAAGTATTAATA

>vvl_UTR
GGCGGCCCACTAGCAATCAGAAATCCAGGAGTCGAACTCAGCTGCAGCTGCGTCCACTCCTGCA
TCCCTCAATAATCTAAGCCAGCAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAACACCAGC
AACAGCAGCAACATCAGCAGCAGCAGCACACACCGAATACACCATCCTCCAGTGCGGGATCATC
GGCTACGATGACCAGCCAGGTGATGTCGCCGCAGAGTCCGCTGGGCAGCTCCTCCGCTGGCAAT
CAGGCTAACAACAATAATAATAATAATAACAACAACAATAGCAGCACCAACAATAACAATAATA
ACAATAACAACGAGGAGCAAGTTAAACACCACCAGCAGCAACAGCAACAACAGCAGGCATCCAG
CATAGCAGCCGCTGCAGCAGCCGCCATGTACGTGGATCCCATGCGCTACCAGCATCCGCACCAC
CCGCATCCGCATCCGCACCAGCATCCGCACCTGAATCCCGCCCACCATCCGCACCTGTTCCACA
CGAGCGATCAGTTGCAGCACTCGCAGCAGCAGACCGTCGGCGGCACAGCAGCAACTCCCAATTG
TCACCGGGTGCCGGCAGCAACAGTGGCCTGCCACTCCAGCCGCCCAGTTCCGCCTCGCCGGCCG
GATCCGCATCCTCGGTGCTTGGTGGCCATCTCAATCTGAATCCGCTGCACCAGCACCACCACTA

TCAGCACCACCATCACCAGCAGCACCACCACCACCAGCAGCAGCAGCAGCAGCTCCACGCCCGC
CGACTCTTCGAGTTGAGTCTGCAGTTTGGAGCGGCGGCGGCACCGGGAGCGGTACGGCACTTTA
ACAGCTTTGGCGGGGCGGGCGGAGCGGGTGGCGAATAGCATTCGTCGGCCGCCTGGTTCGGCTA
TGAGTCCTCCTAGGATTCGGCGTCGAATGGGTGGCCGCATTTCAAGCGGGAGCAGCCCTAGAAT
CGTACAGTCGAACTTCTCTAACGCGAACTCAGTTGGTTCGATTCGAATGATCGCAAAAGTAGCT
ATCTAAAATTTAACTAGCTGTGAGCATTTCAGAATATTTCGTAAGTCATAAGAATATTTTTGAA
ATTTTTCAATATTTTATTTCTGCGGTGTTTTGAGGTTTTTTAAAATTTGAAAATAGCATTTCAC
AGTGTTCTTAAAATCTGTTACTAAATTGGCGACTGAAAAAAATCATTTTAAACATATATACTTA
AACAATATGAACCATATTGAGTTCCACGTTAGACAAGCTCGACTGTATGCAAATCTTTAAAGAA
TAGAAATAAGATGAGGTCCGTCGATGAGGTTGAGAAATCATCTTTTTTACCAGTGCATTAATCT
GTGTTAAACTTCAGCTTAAATAGGAGCAGGGAGTGGAGGAGGTGGAGAAGATGGGGAGGCTGCC
ATCCACCATTTTAGTTTGTTA

>chp_UTR
TGACTTCTGCACTTTACATATAAGACGTCCACCGACTTTTATACCATTCTTGAGCAAGTCTACG
GCAGACAGGTCATATGAGTCCTGCCGTCTAGAGTTTAAGCGTGATGATGAGACTTTGGCCACTT
TGATACCTTTTGCAGTTAGTTGATTCTCTTCCGACAACCTTCTCTCACCGATCCGCTGGTCCTT
GGTAACGCCTTATATGTAGTACTCCTCACAATTCACCACCGAAGCACTCCATTCTTCGCCTCGC
CTCACTGTCTATCTCCATGTATGTATATTCTCGTAGGCCTTAACTAATCCAGCTAATCCAGCTA
TCTAGTAAACCGCCTGACCTGACCCACGAATACTTTCTGTACTTCTTCTATAAGTTAGGTCACT
GCGCTGCTTAAACAATTCAAAATGAAAGTCTGTAAAATGTTTCATAAACA

>CG12880_UTR
GTTCTACCAGAACATCATCTCGTAATCCAACGGCATACCCTCGGTATACCCAGGGAATACCCTT
GAAAGACTGCGGTGCCGGGCGTTCTATTTTCAATATTCTAACGAAGGGGAGCTTCTCCCCTCTT
CCGTCCAGTTTCCACTTGGAATACCCAATCCTGGCGAATCCCGATCACATCTATAGCATCCTGG
CCAAATGTTTAACAATTGCGATGGGTGCGTTCTATTTTTATATAACGGCAATGAATACAATAAC
AGGTGGGCTGGTCCAGGTGGGGTGCATCTCATATCCATCCATGTGGTCAGCCATATCAGCACAA
AGAACAAAGTGTCAAAGTGACATGACTTCATTCGAGATTTCAATGTAAAAACAAAGAGTTTAAA
GTTCAGAAAGGGGTTTTGATTTGAGCACAAACTGAGTTGAAACTAATCGGATTATATATTTAAT
TGAAATTTAATATCTAAAGTTTAATCAGCTTAGCAGCAGAGTGTGAAATATTCGAGTGGCACTT
CTTCATACACATATCGTTTTGATCGGAAATCCCAAAATCCTTTTCCTATGAAGGACACCCCATC
CGAAGGACCAGTCCAAACTGCATTTAGCTAAGCAACTGCGGCAAGCAATAAATTATGGCATAAT
CAAGCTAGGTTTAAGTGCATGAACGACTTTGTTTTGATAAATAAAAATAA

>CG6583_UTR
CAGAATATTTAGAAGTTAATAGGAAAACCAGGCTTATCCTCACCACAAACGAAACACGCAATAA
GTGCCTTTTCGCCACTCCACATGATATCTAAGCTAGGATTGACTTTTGATATCAAAAACACATT
CCGTCCACCTTTTTCCTGCCTATAATTAAGTCGAAAAGTTGAAAACATTTTAATCGTATAGGTA
CATTAGTTATCTACAATATCCAAACAAGAGGCAGGTAAATTTACAGACCAAAACTAGCTAAAAT
ATTTAAATGAAAAGTAACGTTAAACAATTGTAAATTTCCTATGTCGCTTGTACTTTTTATATAG
TTAACTTAATGAACATGTTTTCAAAATGATTTTACTACATTTGATTTTTAAAAAATGACCATTC
CGAAAATGCTTACGTTTGACATATTTATTACTAGTATTTACCTGCCATATCCGAATAACCACTA
CCATATCGAAAAACCAGTAAATTGTACACACAGAATATGAAAATATATACTATATTAAGAGTAT
TAAACAATTCAATTTTCATTTA

>nrx_UTR
GGTCGCAAGAATATGCACATAAATCGTTTTTAAATCGCCTTAGGCACGCAACAGCGACCACTCA
GACACACCAAAAGATAAAAAAAAAATAGATGAGAACGAAGTGAAAATCAAATAGACACGGAAAA

GGCTCACCACGCATAACAGACAGAAAATACTTTTCCATGCACTGCCAATTGTTTTGCGCAAGTA
AATGCTAAGATTTGATTTATAATTGGTTGGCTACATTTTGCACTTATGTTTTTTTTTTTAGTCT
TTGGTTCTGTCTGACGGCATTATTGAAATTTCTTCTATCACAGAATACCGTCCATTTTTAGTTA
ATTTTATGAACTTTTTGTTTGCAAAATTGAGTTAAAAACTTATTTAAGGGAGCTGCCCGAGTAG
CTCAACAAAAGCCACCTATTTTATGTATTTCTATTGCGCGAGTATGTAAGAACGAACATTCCAT
TATTTTTAATATGTTTTCTCGCTTTACGTTTAATTTAAAAATGTCTTCTACTTTACACACAAAA
GAACATGAAAATCATTTTGTGGCTAAACAAAATCGGCTACTTTAGTAACACAAAGCATTAAAAA
CAAATTGTATTTTTGGAAAATGCATTTTGAACGTCTTGACTTGAAATTTTCAAATATACATTAG
CATGTGTATATAGATATATATAGCAATAAAAAAATATATATAATTACAAAAAAGAGACAATGAA
AAAATTAAAATTTTTATCTTTCAATAAAAGAGAACTACAAATAAATGAAATTGGA

>rn_UTR
GTCCTGGGAAGGACAAGGGATAGCTAGGGGCCTACTTCTAGATGGAAGGTAGTTCATCCGTTGT
TTCTCCAAGCGGAGACGAGGCGTGCATTCCTGTACAGTATTAATTTGTAGTTAGTCGTAGCGAG
TAAAGGTCATTTTGTAATTTTTGTAGATCTTCTTGGGTTAAGTTAAGGGTGTTCGTTCTCGCGC
ACTCGATAAGCACTATGTAAATGGCCAACGATACGTACTCTAATTTGACAAACACCGTCCATAC
CTATGTTTACACAATGTCTTGGAATACAAGCACCAGTTCAGGTCCAGATAA

>sinu_UTR
TCGTTTTATATTTTATTCAATGGTTTCCCTGTGCATAGGTTAAAGTATTCATGTCGGTAACATT
CGTTTTATATTACCTTAAATCGTTTAAATAACTCCCCAACTGGCGGAGATTTCGAAAAATTTTA
ACGAACCAAAGTTATTTTGGATAAAAATAATTTTCGCCATATATTCATCGCTTCTCTCTAACAT
ACCTAAAGAAATTAATTTTAGTTCAGAGAAGTTCGACTGTACTCGTAGTTTTTAATTTAAAAAG
TTTATGCAATTGACGGCCGCACCGATTCACAAGCTTTCAGTTTTTGGTACTTATAATCATCCAT
CCAATCGACCAGATCAGTTCCAAGAGAAGACTGAATGTTTTTCCGAATGAAAACCAGAACCAAG
CACAGCTAACATCAAACGAAGACGCTCAACTAGATGACTCATGTTCATGTAGAGTATTCTACAG
TATTTGCATATCAATTTTCGCTCAATTAACGCCCAGACACAAAGAGCATACCCCTCCGCAATTG
TACGTTTATTTTTTTCTTTAGCAAAAGTCGCAGGCTAGTCGGAACTAGCATCGAAATATTACAC
TCGTGCCTTAATTAGATAGAACTATTATGGTGTAGTTCGTAGTGCCTGAAAACAAATTATTGTT
ATAAACATACCTGAACGACTGACCTGCGACTAACCATAGAGATCATTGAATTGCATAAACTTCA
GCTAAGCTAGTGCCTTAATTAAGTGACCCAGAGAAAGGTAAGCGAAACTCCGCCCATTAGTTTT
AAGAACAAAACATTCCCTTTTGAACTCGAATATTCCGCACAAGTTACCTTCTTTTAAAGGTAAC
AATTAGGTCTGTATTAATCATCCGTCCTTTGCACGTATTTGTAACATTTTCTAAACCGAGAATT
TAGCAATGTATTAAAGCAATACATACGTCTTTTACCGTCCCATATACAATAATATGTGTAACAT
TTTTATGTACAAACGACTTTTAGCTAAAGCGAAGAATAATAAAGTTA

>CG1298_UTR
CCGCTGTACAAGGAGGAATTCAAACAGATATATAAATCAAAAATGAATTCGAAGGAATTATCCG
TCCATTTACACAATATTTAATAATAATTAACAAATATACAAATATATACGATGTAAGGTAGAAT
AGTCATTCCACTAAATCCACTACAAATGTTTTTAAA

>mega_UTR
GTCCGCCCGACTTCCATACATTCCGTTCCGATTGGATCTTAGCTTAAGTTTTCAGACGCAATCT
CGTGCCTCGAATCTGCGCAATCGACGCCGTCCATAGTGCCTTTAACAATCGACCATATGTATCT
ATATACACGCCGACTCAGCCGAGATCAGCTCGTATCCGTTTGTAATTGTATTAAATAGCACGTA
CAGCTAAACGTAAGATCAGTGCGACCAAAAAGATCCACTAACTTTTTACACATCTACCATTATA
CTTTTTATATAGGCTCAAAAATGTTTACAGCTTAAATTTTTAGACCATATTCCACCTTTTATAA
GCAAACTGGTGTAGAGACATATCAACATTGAATTAAATGTATTATTTGAAAA

>cora_UTR
AGAAGATTGAAATTCAGCAGCAAACGCAGTAGATTTTATAATTATTTTTAAACCATTCGGATGA
TTTTCTAAGTAAACCCGCACACACACTCACACACATGAACACCAGCCCACGCATACGCAGCACT
TGCAGTACACCTTTTATACAACCTATATCTAAATATAATTTAAACAAATATATATATATATACA
CACGAGTGTATATTTGAGCTCGAGCTGATGACCCACATGACATAACACTAAGAAACATAAAAAG
ATAAATTCCCAATTATCCGTTACCGCAACCGTCTACGAGAAAATGCATCTACATGTTTATTATT
TTTATTACAACTGTTGTTTATAATTACTTTTTATTACGTATATTATATATTGAATGTGCAAGGC
AACGTTGAAATGAAAACTCTATATATTGTAAACCACGATTGCATTCAAATACAAGAATAAAACC
TAAATGTACACACACAGAAAAA

>lac_UTR
CGGCGTTGCTCTTCGCCAGATAAGCCAATGGGCCCACGGCCGTCGACTCCAAGCGCTTCAGGTC
CAATCCATCAACCAGCCCTCGAATGCATAAAACATATATAAAATAGATATATCGAAAACAAACA
TCGTGTAACCGCGCAATAGATGTCTCAAACGATTTCGGACACACCAACCCACATCCAGTTCTCA
GTTCGTTTTGTAGCTTCTGCTTCCGTTTCCGTATAGTCCAGCCCAATTTTTGTACGTCCAATAT
TATGTCTTAAGTTAAGTATGATAAGTGAAGGGCCAACGATAACTAGACTATAATTTATTAGCAC
CCTAGAGCCGAACTCAAACCGAAATTAGACTCCAGCCCAGCCGTAAACGATGCATTTCACCTTG
ATCCCGTCCGTTTTGCTTTCTGATCTGGCACTGTATATGTTTTAAATATTTTAATCTAAGCAAG
ATACATCACAAGCATTCCGATTTTGCTTCGACAATTGAGAGGATATTTTATTTTTATTTAATAA
TAATAAAGTATTTTCGTTACGCAAACCTATCAGCTCGACACACTATTCCCTAAATAATATCTCA
ACCCCCATAATCGAGGCAAAATCCATATTCGAATAAAAGGCAACAAAAGTTGTGTACATTATTA
GCTCATAAGCACAATGAGAAATTCTTTTGAATAACAAGACATTTTTACCCAAATTTTTGTGTGG
GTGTGAGTAAAATCAATTTCGACTTTGTAACAAAAAAAAAAAAAGACAAACGAAAAACTAAAAT
TAACA

>ttk_UTR
TGAAGGACCAGAGTTACGGAAGGACTACGAACGCACAACAACGTTTACCTTAAGCAGCGAGCTC
TGAACTGTAGAATACATCATAGCAACATAAGCAATCGTGGCACTCAAACGCAACACACCAGACT
CAAAACGATTTTTAGTTAATGTAAACCCGTAAGCAATATTAATTCTAACACTTGCTCCGTCCAG
AACGCCTTTTTTAAGCAACACTAGTATTTAATTTACCTAGGACATAAGTTTTGCCCAATTGAGC
AACGCAATTGAGCAGTTCAGTGAACTTTAATGACTTTCGGCTTTAATTATAAACAATTCGACTG
GCCGCTTACTGGATCCGCACACGATCCGTAAGGCAAGCAACTTTAGAACTTTTCATTTCTCTTT
GTGATTTTGTTAAACCTTTTTTATATAGACCTTTGTTTTTTATTTTTTGTTGTACATAGCTATAT
TTTGTTTAATACTCAAAGTCCGTAGTATAATTTTGTTCATTTCCTTGGCAACAAGTTGGAATCA
AATGTATTGGCTAATTGTGCAATCTTCAAAACGTGTTGGTGGCTTCGGAGGAACGCGACAACGC
TGTTCTGCGGCCCCACGAAATGTTGTAAACTTAATTAACTACTAAGGCCACTTAACTAGGATCT
AGGTGTAGTCATAAGGCATTGGATCAATGTTCCAGCGTACAACAAAACGGACTTCCAGTGAAGA
GACAACTGATATTTATACGGAATCCCAATCTGGTATTATGTATTATGATTTTTACAGTTTACTT
AACGCAACAAATACGACTTAAGACTTGTGCGCTCAATGACTAGAAATTATAGTTGCGATTTTAG
GATAACTATTCATGGAAAAACCACAAACGAGTTTTTAATAAAGA

>cont_UTR
CTTTACACAAGCCAATGAATGCAGCACAAATTAAGATCCATAACTTTAGGCAGTAAGAACATAC
ATATACTAAGAAGTGACCTTATATTTGAGTAAGTGAAGTTATAAAAAGAACTACATTCCATCCG
TGCTCTCTTGCAAGCCCCTTAATTGATACAAATGCAATTCCGTCCATTTTATAGACTTTTATCA
AAAAATAATTTCTAAAATTTTATCAACAAAATAAATATGAATACCATTATATGGTCCAGTATGT
TGCTGAACTAAAATA

>nrv2_UTR

GAGCTCCTCATTGATTAAATTGTACCACACTATTACAACATAACATAATAGCAAACACAATAAC
AAGAATCTCCGATTAACAAGAATCTCGGTATCAGAGGTGTGTGAGGACGATAAGTCTCGTCTAA
GACAATCCTACCAGAGGCAAACCACCAGATGTTGTAGTGAAGATTGTGAAAGGTTCAGATTTTC
GAATGTCATTAAAAATCCAAATTATTTAAACTTGTTTTCCGCGTTATAAAGCATACAAAATATA
TATTCAAGACTTGCAACGCCTATAAACATACAAATTTTGAATTCCAATAACTTTTTATGCACTT
TACGATGAAAGTTAGACGTAAGATTGTAAAGCTAAACCACCAAGATGGGAACTCAATGGACAGG
ATGGAACTAAACCAGCCTGAATCTCGTTTGGCGCTGACCAGTAAATTTGGCGCTGACCAGTAAA
TAGTAAATCAATCAAATGCAACAAATTTACAGCAATTAATCCAAACTGCCATTTCCAATGAATT
AACAAACGAATCCCTAGATATTACAAATTTATAAAGAAAATTTAAGAATCATTTTTGAGTCTGC
CGCTATGTAACATGTAACCGAAAACAGACCAATTTTGTCTACCAAAAATCTAAATATTTACAGC
TAGCTTCAGAATTTACCATATTTATTAATAACTTAGTCAAAAGTGAAGCCTATATACATATACA
TCACATTTTGATAGACGACAAAAGTGCAACATTTGGAAAACAGAAAATCGTAACTTCGGTTAAC
TTGTAAGCTTATGATATAATAATTGCGAAGATTGAAATTTAAGGAAATAAATTGAGACGGCTAA
GAGACACAAAACACTTTCAATTCCTTTTCAATAAAAATCAAATGAAAGTAGTCA