

This information has not been peer-reviewed. Responsibility for the findings rests solely with the author(s).

Deposited research article

## MicroRNA targets in *Drosophila*

Anton Enright<sup>1</sup>, Bino John<sup>1</sup>, Ulrike Gaul<sup>2</sup>, Thomas Tuschl<sup>3</sup>, Chris Sander<sup>1</sup>  
and Debora Marks<sup>4</sup>

Addresses: <sup>1</sup>Computational Biology Center, Memorial Sloane Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA.

<sup>2</sup>Developmental Neurogenetics and <sup>3</sup>RNA Molecular Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

<sup>4</sup>Columbia Genome Center, Columbia University, Russ Berrie Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA.

Correspondence: Debora Marks. E-mail: [mirnatargets@cbio.mskcc.org](mailto:mirnatargets@cbio.mskcc.org)

Posted: 13 October 2003

Received: 13 October 2003

*Genome Biology* 2003, 4:P8

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/11/P8>

This is the first version of this article to be made available publicly. A peer-reviewed and modified version is now available in full at <http://genomebiology.com/2003/5/1/R1>

© 2003 BioMed Central Ltd

comment

reviews

reports

deposited research

refereed research

interactions

information



deposited research

AS A SERVICE TO THE RESEARCH COMMUNITY, GENOME **BIOLOGY** PROVIDES A 'PREPRINT' DEPOSITORY TO WHICH ANY ORIGINAL RESEARCH CAN BE SUBMITTED AND WHICH ALL INDIVIDUALS CAN ACCESS FREE OF CHARGE. ANY ARTICLE CAN BE SUBMITTED BY AUTHORS, WHO HAVE SOLE RESPONSIBILITY FOR THE ARTICLE'S CONTENT. THE ONLY SCREENING IS TO ENSURE RELEVANCE OF THE PREPRINT TO GENOME **BIOLOGY**'S SCOPE AND TO AVOID ABUSIVE, LIBELLOUS OR INDECENT ARTICLES. ARTICLES IN THIS SECTION OF THE JOURNAL HAVE **NOT** BEEN PEER-REVIEWED. EACH PREPRINT HAS A PERMANENT URL, BY WHICH IT CAN BE CITED. RESEARCH SUBMITTED TO THE PREPRINT DEPOSITORY MAY BE SIMULTANEOUSLY OR SUBSEQUENTLY SUBMITTED TO GENOME **BIOLOGY** OR ANY OTHER PUBLICATION FOR PEER REVIEW; THE ONLY REQUIREMENT IS AN EXPLICIT CITATION OF, AND LINK TO, THE PREPRINT IN ANY VERSION OF THE ARTICLE THAT IS EVENTUALLY PUBLISHED. IF POSSIBLE, GENOME **BIOLOGY** WILL PROVIDE A RECIPROCAL LINK FROM THE PREPRINT TO THE PUBLISHED ARTICLE.



# MicroRNA Targets in *Drosophila*

Anton Enright (1), Bino John (1), Ulrike Gaul (2),  
Thomas Tuschl (3), Chris Sander (1), Debora Marks (4)

(1) Computational Biology Center, MSKCC

(2) Developmental Neurogenetics, Rockefeller University

(3) RNA Molecular Biology, Rockefeller University

(4) Columbia Genome Center, Columbia University

## Abstract

The recent discoveries of microRNAs (miRNAs) and characterization of the first few targets of their gene products in *Caenorhabditis elegans* and *Drosophila melanogaster* have set the stage for elucidation of a novel network of regulatory control. Here, we present a novel three-step method for whole-genome prediction of miRNA target genes, validated using known examples. We apply the method to discover hundreds of potential target genes in *D. melanogaster*. For each miRNA, target genes are selected based on (a) pattern of sequence complementarity using a position-weighted local alignment algorithm, (b) energy calculation of RNA-RNA duplex formation, and (c) conservation of target sites in related genomes. Application to the *D. melanogaster*, *D. pseudoobscura* and *Anopheles gambiae* genomes in this manner, identifies several hundred target genes potentially regulated by one or more known miRNAs.

These potential targets are enriched for genes that are expressed at specific developmental stages and are involved in cell fate specification, morphogenesis and the coordination of developmental processes, as well as the function of the nervous system in the mature organism. High-ranking targets are two-fold enriched in transcription factors and include genes already known to be under translational regulation. Our results reaffirm the thesis that miRNAs play an important role in establishing the complex spatial and temporal patterns of gene activity necessary for the orderly progression of development and point to additional roles in the function of the mature organism.

The emerging combinatorics of miRNA target sites in the 3' UTRs of messenger RNAs are reminiscent of transcriptional regulation in promoter regions of DNA, with both one-to-many and many-to-one relationships between regulator and regulated target. Typically, more than one miRNA regulates one message, indicative of cooperative control of translation. Conversely, one miRNAs may have several targets, reflecting target multiplicity.

As a guide to targeted experiments, we provide detailed online information [1] about target genes and binding sites for each miRNA and about miRNAs for each gene, ranked by likelihood of match. The target prediction tool can be applied to any similar pair of genomes with identified miRNA sequences.

## Introduction

MicroRNAs (miRNAs) are a novel class of gene products that repress mRNA translation or mediate mRNA degradation in a sequence-specific manner in animals and plants [2-5]. To date, several hundred different miRNAs have been identified from various organisms and their sequences are archived and accessible at the miRNA registry website [6]. Currently, that database contains 21 miRNAs from *Arabidopsis thaliana*, 48 from *Caenorhabditis briggsae*, 106 from *C. elegans*, 73 from *Drosophila melanogaster*, 122 from *Mus musculus*, and 130 from *Homo sapiens*.

With few exceptions, the targets and the mechanism of target suppression are currently unknown because reliable experimental methods for comprehensively identifying the miRNA targets are yet to be developed. Founding members of the miRNA family, *lin-4* and *let-7* in *C. elegans*, have a central role as key regulators of developmental timing through cell fate decisions [7,8]. Because these miRNA genes are also conserved in other animals and mammals [9,10], it is not surprising to find that homologous genes, which were initially identified by genetic interaction, also comprise conserved miRNA binding sites [11]. In insects, the bantam miRNA has been found to regulate cell proliferation and cell death by targeting the antiapoptotic gene *hid* [12]. *D. melanogaster* miR-14 has been implicated in fat metabolism and stress resistance as well as cell death, however the precise target of this miRNA was not identified [13]. The identification of animal miRNA targets is difficult because animal miRNAs are only partially complementary to their mRNA targets, possibly because of additional interactions involving RNA binding proteins. As a result, it is computationally challenging to define an algorithm and thresholds to reliably predict such target sites.

In contrast to animal miRNAs, plant miRNA targets are more readily identified because of near-perfect complementarity to the target sequence [14]. Many of the targeted mRNAs encode transcription factors that regulate plant morphogenesis [15-19]. As a consequence of near-perfect complementarity, plant miRNAs predominantly act as small interfering RNAs (siRNAs) guiding destruction of the mRNA. Another particularity of plant miRNAs is that the targeting sites are predominantly found within the protein-coding segment of the target mRNAs

[14], while animal miRNAs are targeting the 3' untranslated region (3' UTR) [5,12,20-25].

miRNA and siRNA pathways overlap at several points. siRNAs as well as miRNAs are processed from double-stranded RNA precursors requiring dsRNA-specific RNase III enzymes [26-30]. By an unknown molecular mechanism, the excised small RNAs become associated with Argonaute member proteins to form a RNA-induced silencing complex (RISC) that is able to target near-perfect complementary RNAs for degradation or controls translation [31-34]. In animal systems, it was shown that the introduction of a certain number of mismatches at centrally located positions allows for a switch from targeted mRNA degradation to translational repression [35,36]. In general however, mutations in siRNAs typically abolish gene silencing without switching to translational repression [37], although, if siRNA specificity is evaluated at a genome-wide level, siRNA-guided cleavage activity can be detected with sometimes rather distantly related complementary sequences [38].

About 10% of miRNAs identified in invertebrates are also conserved in mammals, indicating that regulatory function of these genes is likely to be conserved cross-species. Since miRNA-containing species have been separated hundreds of millions of years by evolution it is striking that many 22-nt miRNAs were unable to acquire sequence changes. The absence of sequence-evolution in so many miRNAs suggests that these miRNAs have many more than a single targeting site and that evolution by compensatory base-pair changes has become extremely unlikely. Therefore, depending on the birth date of a specific miRNA, miRNAs could regulate a few genes or many. It is also conceivable that additional evolutionary constraints, such as the presence of certain protein binding sites within the miRNA-targeted mRNAs, are conferring specificity to the small RNA regulated processes.

In order to address the question of miRNA target identification in animals, we have developed a computational method to rank the likelihood for each gene to be a miRNA target and conversely for each miRNA to target a gene. The target prediction method relies on the maintenance of evolutionary relationships between miRNAs and their targets, using three completely sequenced insect genomes in this work. We identify distinct networks of gene regulation, such as control of cell fate, morphogenesis and nervous system function, which appear to be preferentially targeted by miRNAs.

## Target Prediction

Every gene in *D. melanogaster* is a potential target for one or more of the characterized *D. melanogaster* miRNAs [39-42].

Reliable identification of miRNA targets is challenging and different to standard sequence similarity analysis. In traditional sequence analysis one tries to assess the likelihood that similarity between two sequences is due to their sharing of a common ancestor, rather than that similarity having arisen by chance.

Here, we aim to assess the likelihood that one sequence (miRNA) has the potential to bind to another sequence (in a 3' UTR). This is complicated by the fact that (i) the sequence of the miRNA is small (22 nt), so standard sequence analysis statistics may not be meaningful, (ii) we do not necessarily know whether both miRNA and target are temporally and spatially co-expressed and (iii) very little knowledge regarding protein complexes which may facilitate this interaction. Because of these problems we have developed a novel three-step method for target identification (Figure 1) that uses (a) sequence-matching, to first assess whether two sequences are complementary and possibly bind (b) free energy calculation (thermodynamics) to assess the likelihood of this physical interaction and (c) evolutionary conservation for further confirmation. We have validated this method using experimentally verified targets from the literature [5,20-25] and against a randomized background model (see methods).

#### **(a) Sequence match**

Using each of the 73 available *D. melanogaster* miRNAs as probes, we scan the 3' untranslated regions (UTR's) of 9,805 *D. melanogaster* genes for possible complementarity matches using a novel dynamic programming algorithm (see methods). For the remaining genes of *D. melanogaster*, accurate 3' UTR sequences were not available. This algorithm uses a weighting scheme that rewards complementarity at the 5' end of the miRNA. This is consistent with known miRNA:target duplexes. Finally, we obtain a score (S) for each detected complementarity match between a miRNA and a potential target gene.

#### **(b) Free energy calculation**

For each match, the free energy ( $\Delta G$ ) of optimal strand-strand interaction between miRNA and UTR is calculated using the Vienna [43]. We cannot however, take into account the energy of interaction with possible protein components of this complex, as the details are at present largely unknown [3].

#### **(c) Evolutionary conservation**

Given imperfect rules for sequence pairing and energy estimation the conservation of predicted miRNA-target pairs in closely related species is an important additional criterion for this analysis. Given the surprisingly high level of sequence conservation of miRNAs across phyla [9,44], we assume that the set of miRNAs in *D. melanogaster* is shared identically with *D. pseudoobscura* and *A.*

*gambiae*. We only consider a miRNA target pair to be conserved across species if the following criteria are met: (i) A specific miRNA independently matches orthologous UTRs in both species, (ii) sequences of detected target sites in both species exhibit more than 80% nucleotide identity (ID) with each other and (iii) the positions of both target sites are equivalent according to a cross-species UTR alignment [45] (see methods).

For this three-step assessment of miRNA-target matches, we use cut-off values that provide a reasonable balance between false-positives and false-negatives, by inspection of known targets (see Methods). The thresholds for sequence conservation ( $\geq 80\%$  for *D. pseudoobscura*;  $\geq 60\%$  for *A. gambiae*) were chosen after extensive analysis of alignments between *D. melanogaster*, *D. pseudoobscura* and *A. gambiae* 3'UTR sequences. To maximize predictive power, we have kept the number of adjustable parameters and cutoffs small (see Methods).

Next, we take into account potential many-to-one relationships between miRNAs and their targets using an additive scoring scheme. This system allocates a score to a miRNA (or multiple miRNAs) and target gene by summing over all scores for all conserved target sites detected for that pair. All predictions are then sorted and ranked according to this scheme, meaning that miRNA target predictions with high-scores, multiple detected sites, or both, are ranked preferentially. Finally, for each miRNA its ten highest scoring target genes are selected for further analysis. The results of our validation and genomic analysis are described in detail below.

## Results: Validation Using Known miRNA Targets

Application of this novel target prediction method to experimentally verified targets serves as our initial validation. However, it should be noted that our method was developed using known targets as a guide [5,12,20-25], so independent validation and refinement of the method will depend on future experiments.

Using a relatively high threshold (see Methods), the method correctly identifies 9 of the known 10 target genes (Table 1) for the 3 miRNAs *lin-4* and *let-7* in *C. elegans* and *bantam* in *D. melanogaster*. Using a lower threshold, the details of target sites are largely reproduced, but interesting alternative sites on these targets are also found (see tables in supplementary material). There is a strong overlap between the predicted and observed target sites for each target gene (Table 1), but this comparison is as weaker criterion of predictive success, as not all reported target sites in the literature have been individually verified by experiment. The missed duplex between the *lin-4* miRNA and its reported target

gene (*lin-14*) contains an unusually long loop structure in the target site sequence, which cannot easily be detected without adversely affecting the rate of false-positive detection. It is also encouraging to note that we not only detect the majority of known miRNA targets, but that the rankings obtained from our additive scoring scheme for these targets are also consistently high (Table 1). For example, two targets of the *let-7* miRNA (*hbl-1* and *lin-41*) are detected as the number 1 and number 2 ranked hits respectively from a scan against 1014 *C. elegans* 3' UTR sequences.

Independent validation comes from the clear tendency of predicted target sites to be more strongly conserved than target sites identified using control sets of randomized miRNA sequences (see methods). Using this approach we show that predicted and conserved target sites occur 2.9 times more frequently than one would expect by chance. Furthermore, target predictions with two or more conserved sites occur 11 times more frequently than one would expect by chance. These results indicate that our method has the potential to accurately detect miRNA target genes.

MiRNA	Organism	Target Gene (3'UTR)	Expt # sites	Pred # sites	Rank	Pred w/ consrvtn # sites	Match Expt-Pred # sites	Match Expt-Pred % of sites
<i>lin-4</i>	<i>cel/cbr</i>	<b>lin-14</b> - Abnormal cell-lineage protein 14	7	1		0	<b>0</b>	<b>0%</b>
<i>lin-4</i>	<i>cel/cbr</i>	<b>lin-28</b>	1	1	4/1014	1	<b>1</b>	<b>100%</b>
<i>lin-4</i>	<i>cel/cbr</i>	<b>lin-41a Lin41b</b>	1	1	5/1014	N/A	<b>1 †</b>	<b>100% †</b>
<i>let-7</i>	<i>cel/cbr</i>	<b>lin-14</b> - Abnormal cell-lineage protein 14	2	6	9/1014	2	<b>2</b>	<b>100%</b>
<i>let-7</i>	<i>cel/cbr</i>	<b>lin-28</b>	1	1	12/1014	1	<b>1</b>	<b>100%</b>
<i>let-7</i>	<i>cel/cbr</i>	<b>lin-41a Lin41b</b>	2	6	2/1014	N/A	<b>2 †</b>	<b>100% †</b>
<i>let-7</i>	<i>cel/cbr</i>	<b>daf-12</b>	3	10	7/1014	1	<b>1</b>	<b>33%</b>
<i>let-7</i>	<i>cel/cbr</i>	<b>hbl-1</b> - hunchback-related protein	8	14	1/1014	8	<b>5</b>	<b>63%</b>
<i>bantam</i>	<i>dme/dps</i>	<b>hid</b> - Head involution defective (wrinkled)	2	2	1/11318	2	<b>2</b>	<b>100%</b>
<i>miR-13</i>	<i>dme/dps</i>	<b>CG10222</b>	1	1	4/11318	1	<b>1</b>	<b>100%</b>

**Table 1:** Validation of prediction method on experimentally known miRNA targets, using intermediate thresholds (S: 80;  $\Delta G$ : -14 kcal/mol). For each known miRNA and target gene pair (in either *C. elegans* or *D. melanogaster*), we list the number of known experimental target sites (**Expt # sites**), the number of sites detected here, both raw (**Pred # sites**) and conserved (**Pred w/ consrvtn # sites**) in *C. briggsae* or *D. pseudoobscura*; and, the number (**Match Expt-Pred # sites**) and percentage (**Match Expt-Pred % of sites**) of known sites that correspond to computationally detected conserved sites, with larger values indication more successful (retrospective) prediction ( † and 'N/A' indicate that no 3'UTR was available to scan against in *C. briggsae*, hence no conservation analysis was possible, results assume conservation).

## Results: miRNA targets and processes

## 1. Potential miRNA targets

We report potential targets for the 73 known *D. melanogaster* miRNAs in decreasing order of match score. All results from this analysis have been made available online [1]. All of these targets have passed the filters for free energy estimates, as well as conservation of target site between *D. melanogaster* and *D. pseudoobscura*, and, in some cases, *A. gambiae* (see methods). Of the 701 predicted target genes, 348 have some functional annotation [46,47], and 129 of these have more than one predicted target interaction site in their 3' UTR. About one-third of all miRNA-target pairs identified have two or more target sites per mRNA, for which our random model suggests are highly reliable. Perhaps the most interesting candidates for target validation experiments are the 117 target genes (with or without annotation) with multiple sites.

Cooperative binding of miRNAs to a single target can involve multiple hits by one or several distinct miRNAs. Specific examples are the genes (1) The eye pigmentation gene *brown* (*bw*) is hit by the miRNAs *bantam* (3x) and *mir-314* (2x); (2) the apoptosis gene *hid/Wrinkle* (*W*) is hit by *bantam* (2x), *mir-309* and *mir-286*; and the eye development gene *seven-up* (*svp*) is hit by *mir-33* (2x), *mir-124*, *mir-277*, and *mir-312*.

A more stringent filter on target candidates is the requirement of conservation of target sites in the mosquito *Anopheles gambiae*. Searching for targets of the same 73 miRNAs (given strong conservation of miRNA sequences between *D. melanogaster* and *A. gambiae*, [41]) in the UTRs of *A. gambiae* genes using the same procedure (Figure 1), we find 150 potential targets in corresponding genes. Of these about 40% had the target site conserved relative to *D. melanogaster* at over 60% identical residues. Notable examples are *scr* (*mir-10*), *netrin-B* (*mir-184*, *mir-284*) *sticks and stones* (*mir-28 2- 2 hits*) , and *VACht* (*mir-9*); notation is *target gene* (*miRNA*).

Having essentially all protein coding genes and (possibly) nearly all miRNAs in *D. melanogaster*, one can identify both the biological processes and molecular functions more likely to be targeted by miRNAs than expected by chance. These include cell differentiation, larval and embryonic development, morphogenesis, cell-death (Fig. 2). Separate groups of miRNAs appear to be specific to particular functional classes of target genes: a group of seven miRNAs, *mir-281*, *mir-311*, *mir-79*, *mir-92*, *mir-305*, *mir-131* and *mir-31a*, all enriched between two and four times for targets in larval development; a group of five, *bantam*, *mir-286*, *mir-309*, *mir-14* and *mir-306* are enriched between three and six times for targets implicated in death or cell death. *bantam*, *mir-286* and the *mir-2/mir-13* family have targets five to six times enriched for apoptotic regulation. A group of nine miRNAs is also two to three times enriched in genes involved in pattern specification. Remarkably, target genes annotated as transcription factors

(Supplementary material) are detected twice as frequently as one would expect by chance (21% of annotated identified target genes, compared to 9.5% of all annotated Flybase [48] genes). Translation factors are increased four times over expectation (mir-318, mir-304, mir-276b). This could represent a feedback loop for miRNA regulation of translation itself, as well as the translation of individual genes. Interestingly, for the well characterized (in *C.elegans*) miRNA let-7 only two of the top ten let-7 targets are annotated - *tamo* and *lar*. The gene *tamo* is thought to be required for the nuclear import of dorsal and recent work has connected it to the expression of a small RNA regulated by ecdysone.

Investigating possible connections between genomic location and function, we analyzed at least 12 clusters of miRNAs in the *D. melanogaster* genome which are potentially co-expressed, e.g., *let 7*, *mir-125*, *mir-100* [39-42]. Contrary to expectation, we did not find any obvious links between the co-location on the genome and the functional classes of the predicted targets. (Except in the obvious case of co-location of miRNAs of similar sequence). A possible exception is the link between the position of three of the 5 copies of the *mir-2* family in the intron of the gene *spitz* (involved in growth) and one of its top targets, *reaper* (involved in cell death).

In the following sections we detail three particularly interesting biological processes that are potentially targets of miRNAs.

## **2. Body axis determination**

A multi-tiered hierarchy of transcription factors establishes the morphological segmentation and diversification of the anterior-posterior body axis of *Drosophila* embryos [49]. The *Hox* genes (*lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A*, and *Abd-B*) play a key role in the diversification by switching the fates of embryonic segments between alternative developmental pathways [50]. The genes are organized in two separate clusters on chromosome 3R in *Drosophila*, the Antennapedia (*lab*, *pb*, *Dfd*, *Scr*, *Antp*) and Bithorax (*Ubx*, *abd-A*, *Abd-B*) complex. Both the genes and their relative order within the complexes are conserved in vertebrates [51].

Our predictions indicate that 5 of the 8 *Hox* genes are regulated by microRNAs (Table 2). *Scr* is a potential target of mir-10, which is located between the *Dfd* and the *Scr* gene. *mir-10* is also located in near the homolog of *Dfd* (*hox4*) in the *Hox* gene cluster in *A. gambiae*, *Tribolium castaneum* as well as in zebrafish, pufferfish, mouse and human [52]. *Scr* is also a strong hit for *bantam*, the microRNA associated with the pro-apoptotic gene *hid* [12] and for *mir-125*, the putative *Drosophila* homolog of the miRNA *lin-4* in *C. elegans*. All three of the Bithorax complex genes are likely to be regulated by multiple microRNAs. Interestingly,

*abd-A*, and *Abd-B* are both targeted by *mir-iab-4-3-p*, which is located within the complex between *abdA* and *abdB*.

In addition, not only Hox genes but several regulators of Hox gene function appear to be miRNA targets. These include members of the trithorax activator (*trx*, *trr*) and the *Polycomb* (Pc) repressor groups, which control the spatial patterns of *Hox* gene expression by maintaining chromatin structure, and homothorax (*hth*), which is required for the nuclear translocation of the Hox cofactor extradenticle (*exd*).

Gene	CG Identifier	miRNA
abd-A	CG10325	mir-281 <sup>†</sup> mir-263a mir-iab-4-3p <sup>a</sup>
abd-B	CG11648	mir-3 mir-306 mir-iab-4-3p <sup>a</sup>
Antp	CG1028	mir-304
hth	CG17117	mir-276a mir-276b mir 279
Pc	CG32443	mir-100
Scr	CG1030	bantam mir-10 mir-125 mir-275 mir-315
trr	CG3848	mir-124 <sup>†</sup>
trx	CG8651	mir-283
Ubx	CG10388	mir-280 mir-315 mir-317

Table 2: **Potential miRNA targets of the hox cluster genes and their regulators.**

<sup>a</sup>Target gene based on top 20 hits of each miRNA; all other are based on top 10 hits. <sup>†</sup>Conserved in *D. pseudoobscura* and *A. gambiae*.

### 3. Ecdysone signaling and developmental timing

Ecdysone signaling triggers and coordinates many of the developmental transitions in the life cycle of *Drosophila*. Ecdysone pulses occur during embryonic, the three larval instar, prepupal, pupal and adult stages and regulate numerous physiological processes, including morphogenetic cell shape changes, differentiation and death. [53-55]. The regulation of these diverse processes by ecdysone is achieved by a complex gene regulatory hierarchy. At the top acts the ecdysone receptor (*EcR*), a member of the nuclear hormone receptor family [56], which regulates the expression of different sets of transcription factors, including the zinc finger proteins of the Broad-Complex and many other nuclear hormone receptors, which in turn control key regulators of the different physiological processes.

Our predictions show many potential miRNA targets at several levels of the ecdysone cascade. These include *EcR* and several of the downstream transcription factors and co-factors (eg *B $\rho$  C*, *E74*, *E75*, *E93*, *crol*, *fkf*, *ftz-F1*, *bonus*). Specifically, *broad* (*br*) has 7 different splice forms with 5 different UTRs, whose expression is exquisitely timed and differentially controlled in different tissues. All five different UTRs of *br* are predicted to be high-ranking targets for miRNA regulation (Table 3). The regulation of *broad* by mir-9, mir-14, and mir-210 at the level of multiple transcripts (Z2, Z3, and Z4; Z1 and Z2; Z2, and Z4, respectively) suggests combinatorial mechanisms to achieve specificity in targeting genes.

In addition to the core transcription factors of the Ecdysone cascade, several of its effector pathways are likely to be directly targeted by miRNAs. These include genes in morphogenetic/stress signaling signaling (*aop*, *msn*, *slpr*, *hep*), biogenesis (*rab6*) and the cell death pathway (*hid*, *rpr*, *parcas*, *Rep2*). Finally, several miRNAs target genes involved in the biosynthesis of ecdysone (*woc*, *CypP450s*) and of other hormones triggering developmental transitions (*amon/ETH*, *Eh*). Despite

their synchronous expression with ecdysone pulses in late larvae and pre-pupae [9], let-7 and miR-125 are not prominently targeting the core factors of the ecdysone cascade.

Gene	CG Identifier	miRNA
amon	CG6438	mir-2a mir-2c
aop	CG3166	mir-7 mir-92b
bon	CG5206	mir-iab-4-5p
br <sup>a</sup>	CG11491 <sup>a</sup>	Z1 - mir-14 Z2 - mir-9 mir-14 mir-210 Z3 - let-7 mir-9 Z4 - mir-9 mir-210 Z5 - mir-316
crol	CG14938	mir-210 <sup>†</sup> mir-79 mir-313
Cyp314a	CG13478	mir-308
Eh	CG5400	mir-279 mir-100
EcR	CG1765	mir-14 <sup>†</sup>
Eip63F-1	CG15855 <sup>c</sup>	mir-277 mir-184 <sup>†</sup> mir-31b <sup>†</sup>
Eip63E	CG10579 <sup>c</sup>	mir-263b
Eip71CD	CG7266	mir-34
Eip74EF	CG32180	mir-306
Eip75B	CG8127 <sup>b</sup>	mir-263a
Eip93F	CG18389	mir-14 mir-286
fkh	CG10002	mir-281
ftz-f1	CG4059 <sup>b</sup>	mir-286
Heph	CG3100	mir-274 <sup>b</sup>
Hr38	CG1864	mir-308
Hr46	CG33183 <sup>b</sup>	mir-1 mir-9a mir-9c mir-11 mir-124 mir-318
Hr96	CG11783 <sup>b</sup>	mir-92a mir-281
ImpE1	CG32356 <sup>c</sup>	mir-210
ImpE2	CG1934 <sup>c</sup>	mir-184 mir-7
ImpL2	CG15009 <sup>c</sup>	mir-9a
msn	CG16973	mir-307 <sup>b</sup>
pcs	CG7761	mir-308
Rab6	CG6601	mir-317
reaper	CG4319	mir-13a
Rep2	CG1975	mir-210 <sup>†</sup>
slpr	CG2272	mir-3 <sup>b</sup>
W	CG5123	bantam mir-286 mir-309
woc	CG5965	mir-100

Table 3: **Potential miRNA targets of ecdysone induction.** <sup>a</sup>The 5 splice variants of the gene br are indicated as Z1, Z2, Z3, Z4, and Z5. <sup>b,c</sup>Target gene based on the top 20 and top 30 hits of each miRNA, respectively; all others

based on top 10 hits. † [23]Indicates conservation in *D. pseudoobscura* and *A. gambiae*.

#### **4. Development and function of the nervous system**

We predict a large number of miRNA target genes involved in cell fate decisions in the developing nervous system (Table 4). These include the neurogenic genes of the E(spl) complex and the Notch pathway regulators of the Brd complex (*Brd, Bob, Tom*) [57] genes of the two complexes are known to share motifs for translational regulation in their 3'UTR (Bearded- and K box), some of which have been previously predicted as miRNA target [58]. Our targets further include factors involved in the asymmetric cell division of neuroblasts (*insc, par 6*) and transcription factors regulating different aspects of neuronal differentiation (*vvl, svp, SoxN, nerfin, Dr, unc4, Lyra, jim, cpo*).

The establishment of neural connectivity is a complex morphogenetic process comprising the growth and guidance of axons and dendrites, and the formation of synapses. Many miRNAs target these processes at several different levels. Figure 3 illustrates a number of genes and the repertoire of miRNAs which target them. The targets comprise a surprisingly large number of secreted and transmembrane factors that mediate cell-ECM and cell-cell interactions during axon guidance (*Slit, Netrin A and B; Robos, Semas, Dscam Eph, Drl, PTP; Fasl, beatIV*). In addition to these cell surface factors, miRNAs target the cellular machinery that effects cell shape change and adhesion, including regulators and components of the cytoskeleton (e.g., *trio, dock, Wasp, Abl, tricornered, RhoGAP, Rho1, Khc-73, Tm2, •amma-tubulin*) and of the cell junctions (e.g., *skiff, mbc, crumbs*).

Many of these developmental factors are re-employed in the mature nervous system to control synaptic function by effecting morphogenetic changes in their size, shape and strength. Additional miRNA targets in the mature nervous system include neurotransmitter receptors, ion channels as well as synaptic vesicle components and their regulators (e.g., *Shaker, DopR, Cirl, SerT, nAChR*).

Why does translational regulation by miRNAs feature so prominently in the development and function of the nervous system? Transcriptional regulation is problematic on several accounts: the distances between dendrites / soma(nucleus)/axon are relatively large, making efficient nuclear regulation difficult. Compartmentalization of different parts of neurons (e.g., dendrites *vs* axon) and even within a given compartment (e.g., within the region of a growth cone or different branches of a dendritic tree) are crucial for their function. Specifically, in axon guidance, the composition of adhesion molecules and chemotropic receptors on the surface of the growth cone changes in response to

external cues presented by intermediate targets through post-transcriptional mechanisms.

Gene	CGIdentifier	miRNA
Abl	CG4032	mir-9a mir-9b mir-318
Beat-1a	CG4846 <sup>a</sup>	mir-263b
Dock	CG3727	mir-33
Drl-2	CG3915	mir-274 mir-275 mir-304
Eph	CG1511	mir-282 mir-283 mir-306
Lar	CG10443	let-7
NetA	CG18657	mir-275 mir-216 mir-288 mir-314 mir-14 mir-iab-4-3 <sup>p</sup>
NetB	CG10521 <sup>a</sup>	mir-1 mir-31b <sup>†</sup> mir-184 <sup>†</sup> mir-311
Ptp99A	CG2005	mir-8
Robo	CG13521 <sup>a</sup>	mir-4 mir-10 mir-13a <sup>†</sup> mir-125 <sup>†</sup> mir-133 mir-282 mir-284 mir-307
Sema-1a	CG18405 <sup>a</sup>	mir-281 mir-iab-4-5p
Sema-1b	CG6446	mir-184
Sema-2a	CG4700 <sup>a</sup>	mir-3 mir-4
Sli	CG8355	mir-33 <sup>†</sup> mir-278 <sup>†</sup>
Spn	CG18497 <sup>a</sup>	let-7
Trio	CG18214	mir-184 <sup>†</sup> mir-79

Table 4: **Potential miRNA targets of the axon guidance pathway.** <sup>a</sup>Target genes based on top 20 hits of each miRNA; all others based on top 10 hits. <sup>†</sup> Target in *D.melanogaster* conserved in *D. pseudoobscura* and *A. gambiae*.

## Discussion

The precise rules and energetics for pairing between a miRNA and its mRNA target, with probable involvement of a protein complex, are not known and cannot easily be deduced from the few experimentally proven examples. Therefore, currently any computational methods for the identification of potential miRNA targets are at risk of having a substantial rate of false positives and false negatives. Based on analysis of the known examples, we have biased our method toward stronger matches at the 5' end of the miRNA, and used energy calculation plus conservation of target site sequence to provide our current best estimate of biologically functional matches. Overall, we find that conservation is a crucial filter and reduces the rate of prediction error, as assessed using randomized miRNA sequences. The identification of preferred biological processes likely to be regulated by miRNAs (Figure 2), some of which require precise definition of boundaries in space and time, consistent with the pioneering work on the role of miRNAs in developmental processes [41,59] is additional support for the utility of the target groups, as presented here, for the design of experiments.

As miRNA and mRNA have to be simultaneously present at minimum levels in the same cellular compartment for a biologically meaningful interaction, more precise expression data as a function, e.g., of developmental stage [41], will be extremely useful and will be incorporated in future versions of target prediction methods. Similarly, further work will include the analysis of potential target sites in coding regions and 5' UTRs, as well as conservation and adaptation of target sites in many species and modeling these across evolutionary history.

The genome-wide scan for potential miRNA targets is giving us a first glimpse of the complexity of the emerging network of regulatory interactions involving small RNAs (Supplementary material). Both multiplicity (one miRNA targets several genes) and cooperativity (one gene is targeted by several distinct miRNAs) appear to be general features for many miRNAs, as already apparent with the discovery of the targets for *lin-4* and *let-7*. The analogy of these many-to-one and one-to-many relationships to those of transcription factors and promoter regions is tempting, with quantitative detail and predictive modeling to be worked out. Elucidation of the network of regulation by miRNAs will make a major contribution to cellular systems biology.

In the meantime, we would not be surprised if experiments focusing on target candidates filtered in this way will have a high rate of success in unraveling the fascinating biology of regulation by miRNA-mRNA interaction.

## Acknowledgements

We would like to thank Gary Bader and Boris Reva for relevant discussions, Ulrich Unnerstall for support and Kerstin Dose for assistance.

## Methods

### *MiRNA Sequences*

An initial set of *D. melanogaster* miRNA sequences was built using the RFAM miRNA database [44]. Mature miRNA sequences were placed in a FASTA formatted sequence file. In total the final file contained 73 unique miRNA sequences. All sequences used for this analysis are available from RFAM and at [1].

### *3' UTR Sequences*

Sequences for *D. melanogaster* 3' UTRs were obtained from the Berkeley Drosophila Genome Project (BDGP). In total, 3' UTR sequences were available for 14,287 transcripts, representing 9,805 individual *D. melanogaster* genes. A corresponding set of *D. pseudoobscura* 3' UTR sequences was then built from the March 2003 first freeze of the *D. pseudoobscura* genome project at Baylor College of Medicine. Each *D.*

*melanogaster* 3' UTR was mapped to *D. pseudoobscura* contigs by searching both the actual *D. melanogaster* 3' UTR sequence (using NCBI BLASTn [60]) and the peptide sequence of each gene (using NCBI tBLASTn [60]) against *D. pseudoobscura* contigs [60]. Results from these two scans were then used to identify candidate 2000bp regions of *D. pseudoobscura* contigs, within which we believe an orthologous *D. pseudoobscura* 3'UTR is present. The AVID [45] alignment tool is used to align the real *D. melanogaster* 3'UTR and a candidate *D. pseudoobscura* region. Finally, this alignment is used to trim each candidate region, leaving the predicted *D. pseudoobscura* 3' UTR. In total 12,416 transcripts and 8,282 genes from *D. pseudoobscura* were mapped to orthologous *D. melanogaster* UTRs in this fashion. The Ensembl database [46] API was used to construct *A. gambiae* predicted 3'UTRs by taking 2000bp downstream from the last exon of each transcript. Orthology mappings between *A. gambiae* and *D. melanogaster* UTR's were then obtained by searching all Ensembl *A. gambiae* peptides against all *D. melanogaster* peptides using BLASTp. In total 9,823 *A. gambiae* genes were mapped to *D. melanogaster* genes in this manner.

### **Scanning Algorithm**

This algorithm is similar to the Smith-Waterman algorithm [61], however, instead of building alignments based on matching nucleotides (i.e. **A-A**, or **T-T**), it scores based on the complementarity of nucleotides (i.e. **A=T** or **G=C**). The scoring matrix used for this analysis allows **G=U** 'wobble' pairs which are important for the accurate detection of RNA:RNA duplexes. The algorithm uses affine penalties for gap-opening and gap-extension. The algorithm is given two sequences, typically a miRNA sequence and a 3' UTR sequence, and using this scoring scheme extracts all non-overlapping hybridization alignments in order of optimality from the dynamic programming matrices. This method is similar to sub-optimal alignment heuristics previously used in sequence alignment [62,63]. An important modification to the algorithm is based on observation of match patterns in known miRNA-mRNA target pairs: the scoring system is weighted so that complementarity to the 5' end (first 10bp) of the miRNA is more heavily rewarded (in terms of score), than for the rest of the alignment, and penalties for non-complementarity are more heavily penalized. The scaling factor used for this analysis was 2.0.

In order to estimate the thermodynamic properties of a predicted duplex, the algorithm uses folding routines from the Vienna 1.3 RNA secondary structure programming library (RNALib) [43]. The expanded thermodynamic parameters used [64] are more computationally intensive than the initial scan, but allow potential hybridization sites to be scored according their respective folding energies. The miRNA sequence and 3'UTR sequence from a hybridization alignment are joined into a single sequence with an 8 base sequence linker. This strand-linker-strand configuration, assumes the phase space entropy of strand-strand association is constant for all miRNA-target matches [43,64]. The minimum energy of this structure, with the last matching base pair constrained is then calculated using RNALib.

### **Conservation of Target Sites**

All miRNA sequences are scanned against the 3' UTR datasets of *D. melanogaster*, *D. pseudoobscura* and *A. gambiae*. The thresholds used for

hit detection are: (i) Each initial Smith-Waterman hybridization alignment must have a score  $\geq 80$  and (ii) The minimum energy of the duplex structure  $\leq -14$  kCal/Mol. Each hit between a miRNA and a UTR sequence is then scored according to the total energy and total score of all hits between those two sequences. Hits are deemed to be conserved in *D. pseudoobscura* or *A. gambiae* if a target site equivalent to that detected in a *D. melanogaster* UTR can be found in the orthologous *D. pseudoobscura* or *A. gambiae* UTR at the same position in the UTR alignments. Our definition of equivalence between target sites is that their sequences are more than 80% identical. All results from the scan are then ranked and sorted according to total score of conserved target sites detected. For each miRNA the ten highest ranked genes are selected as its candidate targets in this way.

### **Validation**

For the initial validation 3' UTR sequences for *C. elegans* and *C. briggsae* were obtained if possible from UTRdb [65]. If unavailable, UTR sequences were estimated by taking 2000bp of flanking nucleotide sequence downstream of the last exon of the gene in question using the ensembl database [46].

Control sequences for the randomized experiment were constructed by assembling 1000 sets of 73 miRNAs each generated randomly using individual *D. melanogaster* miRNA nucleotide frequencies. Each of these sets of 73 random miRNAs was independently searched against all *D. melanogaster* and *D. pseudoobscura* 3' UTRs as in the reference experiment. Results and counts are then averaged over all 1000 random sets, and are compared with the results of the real miRNA scan.

## **Figures**

### **Figure 1 - Computational Method**

**Flowchart of algorithm and analysis pipeline.** Source data consisting of miRNAs and 3'UTRs are shown in the top left. These data are processed initially by the rsearch algorithm (bottom left) which searches for complementarity matches between miRNAs and 3'UTRs using dynamic programming alignment (Phase 1) and thermodynamic calculations (Phase 2). All results are then post-processed (bottom right), by first filtering out results not consistently conserved with *D. pseudoobscura*, and *A. gambiae* (Phase 3), then by sorting and ranking all remaining results. Finally (top right), all miRNA target predictions are annotated using data from Flybase and stored for further analysis.

### **Figure 2 - A group of miRNAs targeting a process**

**MiRNA target for genes involved in Axon Guidance.** Representation of 3' UTRs for miRNA target genes involved in axon guidance. Each individual conserved hit between an miRNA and a target gene is represented by an annotated triangle superimposed on a conservation plot (*D. melanogaster* vs *D. pseudoobscura*) for that UTR. Red triangles indicate target site locations that are illustrated in more detail (alignment and 2° structure) below. Multiple target

sites on a 3' UTR for one or more miRNAs are not uncommon and reflect cooperative regulation of transcription.

***Figure 3 - Functional Classes of Target Genes***

**Functional map of miRNAs and their targets.** The left axis represents selected over-represented Flybase [47] derived GO [66] classifications from the 'biological process' hierarchy. The bottom axis represents each of the 73 miRNAs used for the analysis. Each cell in the matrix is color coded to represent the degree of over-representation for a miRNA hitting a specific functional class. For example, a bright red box indicates that a given miRNA hits 12x-14x more targets in a particular class than one would expect by chance. The matrix is built by 2D hierarchical clustering (see methods).

## SUPPLEMENTARY MATERIAL

All results, data and supplementary material are available online at the following website [1]:

<http://www.cbio.mskcc.org/research/sander/data/miRNA2003/>

## References

1. :  
<http://www.cbio.mskcc.org/research/sander/data/miRNA2003/>.
2. Pasquinelli, AE: **MicroRNAs: deviants no longer**. Trends Genet 2002.18, 171-173.
3. Ambros, V: **MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing**. Cell 2003.113, 673-676.
4. Carrington, JC, and Ambros, V: **Role of microRNAs in plant and animal development**. Science 2003.301, 336-338.
5. Lin, SY, Johnson, SM, Abraham, M, Vella, MC, Pasquinelli, A, Gamberi, C, Gottlieb, E, and Slack, FJ: **The C. elegans hunchback Homolog, hbl-1, Controls Temporal Patterning and Is a Probable MicroRNA Target**. Dev Cell 2003.4, 639-650.
6. Ambros, V, Bartel, B, Bartel, DP, Burge, CB, Carrington, JC, Chen, X, Dreyfuss, G, Eddy, SR, Griffiths-Jones, S, Marshall, M, et al.: **A uniform system for microRNA annotation**. Rna 2003.9, 277-279.
7. Slack, F, and Ruvkun, G: **Temporal pattern formation by heterochronic genes**. Annu Rev Genet 1997.31, 611-634.
8. Pasquinelli, AE, and Ruvkun, G: **Control of developmental timing by microRNAs and their targets**. Annu Rev Cell Dev Biol 2002.18, 495-513.
9. Pasquinelli, AE, Reinhart, BJ, Slack, F, Martindale, MQ, Kuroda, MI, Maller, B, Hayward, DC, Ball, EE, Degnan, B, Muller, P, et al.: **Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA**. Nature 2000.408, 86-89.
10. Lagos-Quintana, M, Rauhut, R, Yalcin, A, Meyer, J, Lendeckel, W, and Tuschl, T: **Identification of tissue-specific microRNAs from mouse**. Curr Biol 2002.12, 735-739.
11. Moss, EG, and Tang, L: **Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites**. Dev Biol 2003.258, 432-442.
12. Brennecke, J, Hipfner, DR, Stark, A, Russell, RB, and Cohen, SM: **bantam Encodes a Developmentally Regulated**

- microRNA that Controls Cell Proliferation and Regulates the Proapoptotic Gene hid in Drosophila.** Cell 2003.113, 25-36.
13. Xu, P, Vernooy, SY, Guo, M, and Hay, BA: **The Drosophila MicroRNA Mir-14 Suppresses Cell Death and Is Required for Normal Fat Metabolism.** Curr Biol 2003.13, 790-795.
14. Rhoades, MW, Reinhart, BJ, Lim, LP, Burge, CB, Bartel, B, and Bartel, DP: **Prediction of plant microRNA targets.** Cell 2002.110, 513-520.
15. Llave, C, Xie, Z, Kasschau, KD, and Carrington, JC: **Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA.** Science 2002.297, 2053-2056.
16. Palatnik, JF, Allen, E, Wu, X, Schommer, C, Schwab, R, Carrington, JC, and Weigel, D: **Control of leaf morphogenesis by microRNAs.** Nature 2003.425, 257-263.
17. Chen, X: **A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development.** Science 2003.
18. Tang, G, Reinhart, BJ, Bartel, DP, and Zamore, PD: **A biochemical framework for RNA silencing in plants.** Genes Dev 2003.17, 49-63.
19. Xie, Z, Kasschau, KD, and Carrington, JC: **Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation.** Curr Biol 2003.13, 784-789.
20. Lee, RC, Feinbaum, RL, and Ambros, V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.** Cell 1993.75, 843-854.
21. Wightman, B, Ha, I, and Ruvkun, G: **Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.** Cell 1993.75, 855-862.
22. Moss, EG, Lee, RC, and Ambros, V: **The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA.** Cell 1997.88, 637-646.
23. Slack, FJ, Basson, M, Liu, Z, Ambros, V, Horvitz, HR, and Ruvkun, G: **The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor.** Mol Cell 2000.5, 659-669.
24. Reinhart, BJ, Slack, FJ, Basson, M, Pasquinelli, AE, Bettinger, JC, Rougvie, AE, Horvitz, HR, and Ruvkun, G: **The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.** Nature 2000.403, 901-906.
25. Abrahamte, JE, Daul, AL, Li, M, Volk, ML, Tennessen, JM, Miller, EA, and Rougvie, AE: **The Caenorhabditis elegans hunchback-like Gene lin-57/hbl-1 Controls Developmental Time and Is Regulated by MicroRNAs.** Dev Cell 2003.4, 625-637.

26. Bernstein, E, Caudy, AA, Hammond, SM, and Hannon, GJ: **Role for a bidentate ribonuclease in the initiation step of RNA interference.** Nature 2001.409, 363-366.
27. Hutvagner, G, McLachlan, J, Pasquinelli, AE, Balint, E, Tuschl, T, and Zamore, PD: **A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.** Science 2001.293, 834-838.
28. Park, W, Li, J, Song, R, Messing, J, and Chen, X: **CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana.** Curr Biol 2002.12, 1484-1495.
29. Lee, Y, Ahn, C, Han, J, Choi, H, Kim, J, Yim, J, Lee, J, Provost, P, Radmark, O, Kim, S, and Kim, VN: **The nuclear RNase III Drosha initiates microRNA processing.** Nature 2003.425, 415-419.
30. Lee, Y, Jeon, K, Lee, JT, Kim, S, and Kim, VN: **MicroRNA maturation: stepwise processing and subcellular localization.** Embo J 2002.21, 4663-4670.
31. Hammond, SM, Bernstein, E, Beach, D, and Hannon, GJ: **An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells.** Nature 2000.404, 293-296.
32. Martinez, LA, Naguibneva, I, Lehrmann, H, Vervisch, A, Tchenio, T, Lozano, G, and Harel-Bellan, A: **Synthetic small inhibiting RNAs: efficient tools to inactivate oncogenic mutations and restore p53 pathways.** Proc Natl Acad Sci U S A 2002.99, 14849-14854.
33. Hutvagner, G, and Zamore, PD: **A microRNA in a multiple-turnover RNAi enzyme complex.** Science 2002.297, 2056-2060.
34. Zeng, Y, Yi, R, and Cullen, BR: **MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms.** Proc Natl Acad Sci U S A 2003.100, 9779-9784.
35. Doench, JG, Petersen, CP, and Sharp, PA: **siRNAs can function as miRNAs.** Genes Dev 2003.17, 438-442.
36. Saxena, S, Jonsson, ZO, and Dutta, A: **Small RNAs with imperfect match to endogenous mRNA repress translation: implications for off-target activity of siRNA in mammalian cells.** J Biol Chem 2003.2, 2.
37. Elbashir, SM, Martinez, J, Patkaniowska, A, Lendeckel, W, and Tuschl, T: **Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate.** Embo J 2001.20, 6877-6888.
38. Jackson, AL, Bartz, SR, Schelter, J, Kobayashi, SV, Burchard, J, Mao, M, Li, B, Cavet, G, and Linsley, PS: **Expression profiling reveals off-target gene regulation by RNAi.** Nat Biotechnol 2003.21, 635-637.

39. Lagos-Quintana, M, Rauhut, R, Lendeckel, W, and Tuschl, T: **Identification of novel genes coding for small expressed RNAs**. Science 2001.294, 853-858.
40. Sempere, LF, Sokol, NS, Dubrovsky, EB, Berger, EM, and Ambros, V: **Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity**. Dev Biol 2003.259, 9-18.
41. Aravin, AA, Lagos-Quintana, M, Yalcin, A, Zavolan, M, Marks, D, Snyder, B, Gaasterland, T, Meyer, J, and Tuschl, T: **The small RNA profile during Drosophila melanogaster development**. Dev Cell 2003.5, 337-350.
42. Lai, EC, Tomancak, P, Williams, RW, and Rubin, GM: **Computational identification of Drosophila microRNA genes**. Genome Biol 2003.4, R42.
43. Wuchty, S, Fontana, W, Hofacker, IL, and Schuster, P: **Complete suboptimal folding of RNA and the stability of secondary structures**. Biopolymers 1999.49, 145-165.
44. Griffiths-Jones, S, Bateman, A, Marshall, M, Khanna, A, and Eddy, SR: **Rfam: an RNA family database**. Nucleic Acids Res 2003.31, 439-441.
45. Bray, N, Dubchak, I, and Pachter, L: **AVID: A global alignment program**. Genome Res 2003.13, 97-102.
46. Hubbard, T, Barker, D, Birney, E, Cameron, G, Chen, Y, Clark, L, Cox, T, Cuff, J, Curwen, V, Down, T, et al.: **The Ensembl genome database project**. Nucleic Acids Res 2002.30, 38-41.
47. : **The FlyBase database of the Drosophila genome projects and community literature**. Nucleic Acids Res 2002.30, 106-108.
48. Gelbart, WM, Crosby, M, Matthews, B, Rindone, WP, Chillemi, J, Russo Twombly, S, Emmert, D, Ashburner, M, Drysdale, RA, Whitfield, E, et al.: **FlyBase: a Drosophila database. The FlyBase consortium**. Nucleic Acids Res 1997.25, 63-66.
49. Akam, M: **Hox genes: from master genes to micromanagers**. Curr Biol 1998.8, R676-678.
50. McGinnis, W, and Krumlauf, R: **Homeobox genes and axial patterning**. Cell 1992.68, 283-302.
51. Mann, RS: **Why are Hox genes clustered?** Bioessays 1997.19, 661-664.
52. Lagos-Quintana, M, Rauhut, R, Meyer, J, Borkhardt, A, and Tuschl, T: **New microRNAs from mouse and human**. Rna 2003.9, 175-179.
53. Jiang, C, Baehrecke, EH, and Thummel, CS: **Steroid regulated programmed cell death during Drosophila metamorphosis**. Development 1997.124, 4673-4683.

54. Jiang, C, Lamblin, AF, Steller, H, and Thummel, CS: **A steroid-triggered transcriptional hierarchy controls salivary gland cell death during Drosophila metamorphosis.** Mol Cell 2000.5, 445-455.
55. Thummel, CS: **Molecular mechanisms of developmental timing in C. elegans and Drosophila.** Dev Cell 2001.1, 453-465.
56. Simon, AF, Shih, C, Mack, A, and Benzer, S: **Steroid control of longevity in Drosophila melanogaster.** Science 2003.299, 1407-1410.
57. Lai, EC, Burks, C, and Posakony, JW: **The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts.** Development 1998.125, 4077-4088.
58. Lai, EC, and Rubin, GM: **Neuralized is essential for a subset of Notch pathway-dependent cell fate decisions during Drosophila eye development.** Proc Natl Acad Sci U S A 2001.98, 5637-5642.
59. Ambros, V, and Moss, EG: **Heterochronic genes and the temporal control of C. elegans development.** Trends Genet 1994.10, 123-127.
60. Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W, and Lipman, DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** Nucleic Acids Res 1997.25, 3389-3402.
61. Smith, TF, and Waterman, MS: **Identification of common molecular subsequences.** J Mol Biol 1981.147, 195-197.
62. Waterman, MS, and Eggert, M: **A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons.** J Mol Biol 1987.197, 723-728.
63. Schneider, R, and Sander, C: **The HSSP database of protein structure-sequence alignments.** Nucleic Acids Res 1996.24, 201-205.
64. Mathews, DH, Sabina, J, Zuker, M, and Turner, DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** J Mol Biol 1999.288, 911-940.
65. Pesole, G, Liuni, S, Grillo, G, Licciulli, F, Mignone, F, Gissi, C, and Saccone, C: **UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs.** Update 2002. Nucleic Acids Res 2002.30, 335-340.
66. Ashburner, M, and Lewis, S: **On ontologies for biologists: the Gene Ontology--untangling the web.** Novartis Found Symp 2002.247, 66-80; discussion 80-63, 84-90, 244-252.

# Figure 1

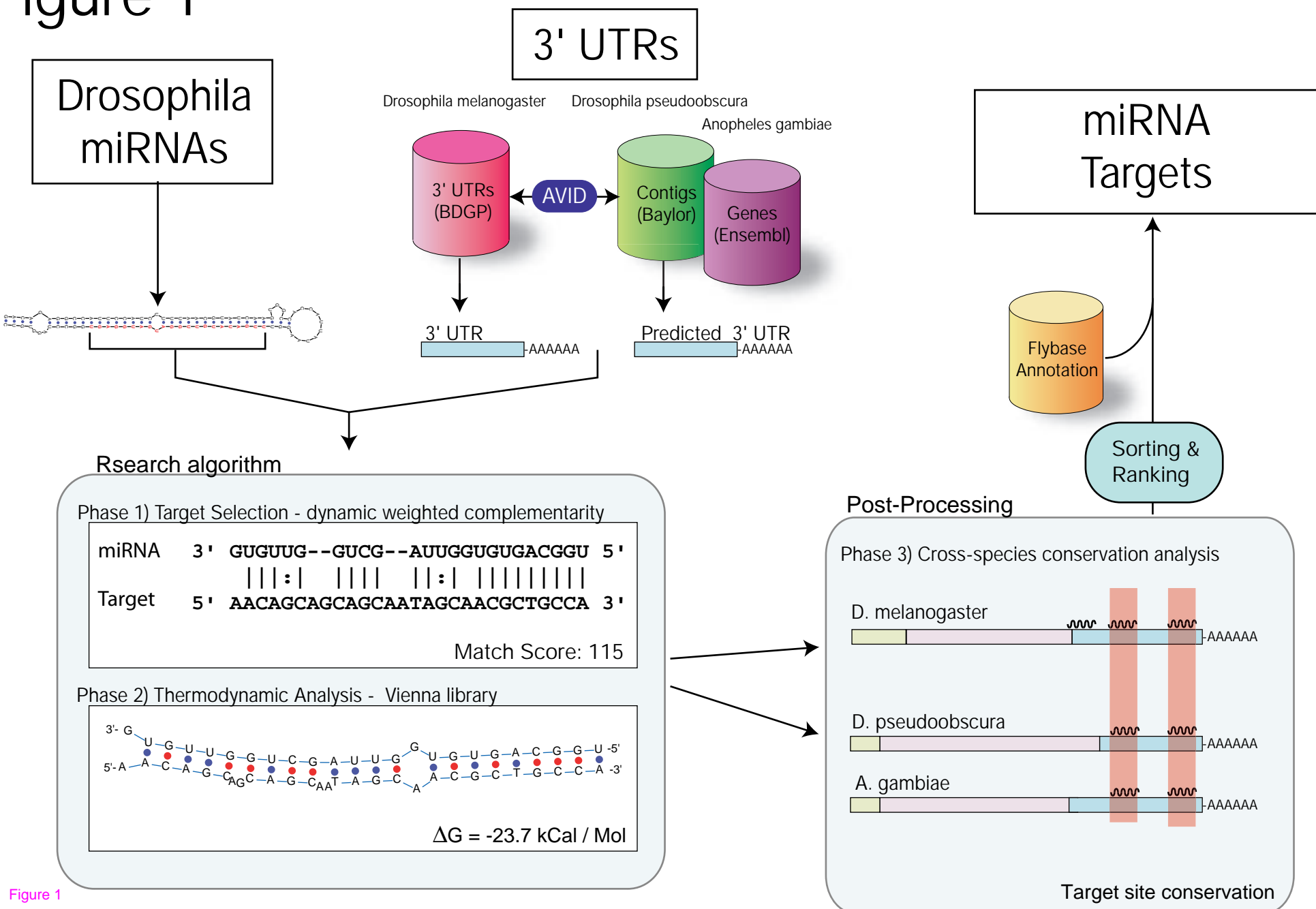


Figure 1

# Figure 2

Functional Class (GO)

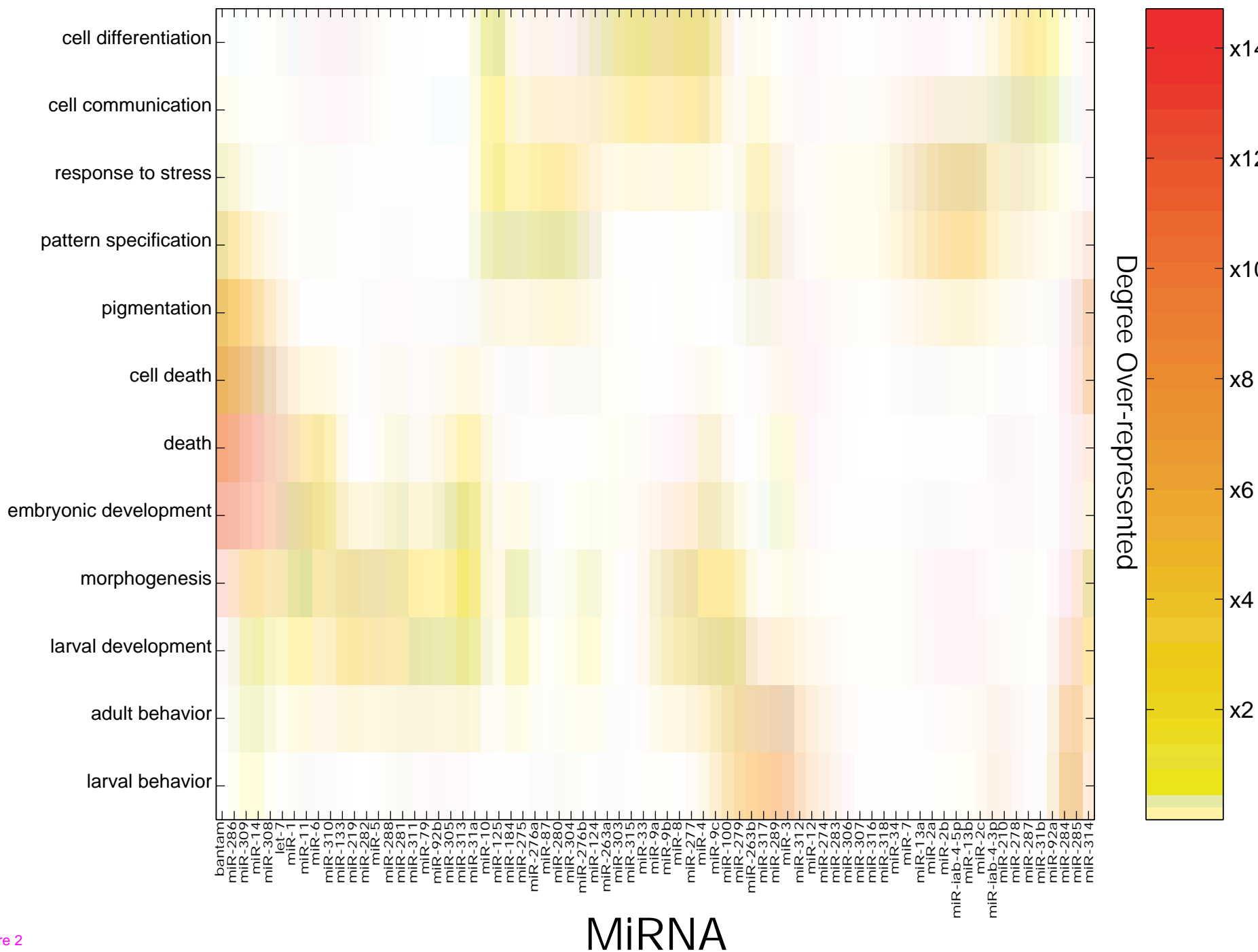


Figure 2

# Figure 3

