

# Identification of Novel *Drosophila melanogaster* MicroRNAs

Thomas Sandmann<sup>1,2</sup>, Stephen M. Cohen<sup>1,2\*</sup>

**1** Temasek Lifesciences Laboratory, Department of Biological Sciences, National University of Singapore, Singapore, Singapore, **2** European Molecular Biology Laboratory, Heidelberg, Germany

MicroRNAs (miRNAs) are small non-coding RNAs with important regulatory roles in post-transcriptional regulation of metazoan development, homeostasis and disease. The full set of miRNAs is not known for any species and it is believed that many await discovery. The recent assembly of 15 insect genomes has provided the opportunity to identify novel miRNAs in the fruit fly, *Drosophila melanogaster*. We have performed a computational screen for novel microRNAs in *Drosophila melanogaster* by searching for phylogenetically conserved putative pre-miRNA structures. The ability of predicted novel miRNA precursors to be processed to produce miRNAs was experimentally verified in S2 cells and in several cases their endogenous expression was validated by Northern blots. After experimental validation, the predictions were cross-checked with reference to a newly released set of small RNA sequences. Combining both datasets allowed us to identify 53 novel miRNA loci in the fruit fly genome 22 of which we had predicted computationally. This significantly expands the set of known miRNAs in *Drosophila melanogaster*. Most novel miRNAs contain unique seed sequences not found in other *Drosophila* miRNAs and are therefore expected to regulate novel sets of target genes. This data provides the basis for future genetic analysis of miRNA function and will aid the discovery of orthologous sequences in other species.

Citation: Sandmann T, Cohen SM (2007) Identification of Novel *Drosophila melanogaster* MicroRNAs. PLoS ONE 2(11): e1265. doi:10.1371/journal.pone.0001265

## INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs, typically 21–23 nt long, which serve as post-transcriptional regulators of gene expression. The founding members of this class, *lin-4* and *let-7* [1–3], were discovered using forward genetic screens in *C. elegans*. Subsequently large numbers of miRNAs have been identified by cloning and sequencing of short RNAs isolated from animals, plants and even the unicellular green alga *Chlamydomonas reinhardtii* [4]. To date, evidence for 533 human, 442 mouse and 93 *Drosophila* miRNAs has been deposited in mirBase (version 10.0) [5]. Based on comparison of miRNA predictions and sequence data it has been suggested that these genomes might contain many more, yet undiscovered, miRNA loci, with estimates of ~1000 miRNAs in the human and ~120 in the *Drosophila melanogaster* genome [6,7].

MicroRNA genes have been found as independent transcription units as well as within introns of protein coding genes. More than half of the known loci are part of tandem arrays within operon-like clusters [8], containing up to 40 individual miRNAs [9]. Each primary miRNA transcript (pri-miRNA) contains an extended stem-loop structure, from which a precursor pre-miRNA is released through the action of the RNaseIII enzyme Drosha and the double-stranded RNA-binding domain (dsRBD) protein DGCR8/Pasha (reviewed in [10]). Additionally, a novel mechanism of pre-miRNA generation from very short introns through the splicing process itself has been reported recently [11,12]. Animal pre-miRNAs are exported from the nucleus to the cytoplasm and processed further by the Dicer enzyme to yield a characteristic ~22-nt miRNA duplex. The strand with the lower base-pairing energy at its 5' end is then loaded onto the RNA-induced silencing complex (RISC) and guides it towards its regulatory targets [13,14].

miRNAs mediate post-transcriptional inhibition by base-pairing with their cognate target mRNAs. Extended complementarity, as is typically found in plants, leads to cleavage of the target. Limited pairing, including but not limited to the 5' seed region of the microRNA causes translational inhibition, and in some cases increased RNA turnover [15–18]. Evidence from microarray

studies suggests that single miRNAs can target hundreds of messenger RNAs [19] and recent *in vivo* studies of loss-of-function phenotypes demonstrate the important regulatory role of miRNAs e.g. in development [20–23], homeostasis [24] or disease [25].

The majority of miRNAs known to date has been identified through cloning and sequencing of small RNA libraries. Additionally, complementary computational tools for predicting miRNAs have been developed. Two basic strategies have been described to predict novel microRNAs: Sequence similarity searches led to the identification of novel members of previously established miRNA families, many of which are conserved across the animal kingdom [26,27]. Other approaches have successfully taken advantage of the well-defined secondary-structure constraints of pre-miRNA sequences as well as evolutionary conservation in related species to detect members of completely novel miRNA families [28,29].

The publication of many newly sequenced and assembled genomes presents us with an unprecedented wealth of information about evolutionary conservation of coding and non-coding sequences. Based on the observation that the evolution of many non-coding RNAs (ncRNAs) is constrained by functionally impor-

.....  
**Academic Editor:** Cayetano Gonzalez, Institute for Research in Biomedicine, Spain

**Received** October 31, 2007; **Accepted** November 10, 2007; **Published** November 28, 2007

**Copyright:** © 2007 Sandmann, Cohen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the European Molecular Biology Laboratory, Heidelberg Germany and by the Temasek Life Sciences Laboratory. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* **To whom correspondence should be addressed.** E-mail: cohen@tll.org.sg

tant secondary structure elements, novel approaches have been developed to scan whole metazoan genomes for putative ncRNAs. The RNAz program employs a support vector machine (SVM) to evaluate both thermodynamic stability of candidate sequences as well as the conservation of the predicted secondary structure [30,31]. Evofold uses stochastic context free grammars (SCFG) to identify conservation patterns indicating folded RNA structures [32]. Both programs are not limited to any particular class of ncRNA and have been used to predict several tens of thousands of putative ncRNAs in the human genome. To identify novel miRNAs in *Drosophila melanogaster*, we have combined a computational screen for generic ncRNAs using RNAz with stringent filtering criteria specifically targeting miRNA precursors. We have validated 22 of the newly predicted loci experimentally. 31 additional loci were identified from recently published small RNA sequence libraries.

## RESULTS AND DISCUSSION

The fruit fly *Drosophila melanogaster* is a powerful model system to study metazoan development. Its compact genome is well annotated, owing to many years of genetic experimentation [33], and has been mined extensively for coding as well as non-coding gene loci [28,34,35]. With the assemblies of a range of additional insect genomes, information about twelve *Drosophila* species, the honeybee (*A. mellifera*), the flour beetle (*T. castaneum*) and the mosquito (*A. gambiae*) has recently become available. To identify novel microRNAs, we took advantage of the whole-genome alignments of 9 and 15 insect species provided by the UCSC genome browser [36]. Multiple sequence alignments for all “most conserved” regions, as identified by phastcons analysis [37], were extracted and scanned with RNAz for putative conserved non-coding sequences and several thousand candidate regions were recovered at a p-value cutoff of  $>0.9$  (4003 with multiz9 and 17727 with multiz15 input alignments). At this threshold, 51 of the 65 previously known *Drosophila* miRNAs covered by the multiz9 and 51 of the 76 covered by the multiz15 input alignments were recovered (37/51 miRNAs were detected in both inputs).

To identify novel miRNA precursor structures, we applied a stringent filtering strategy [38] to the RNAz output, selecting only RNAz predictions with i) a stem-loop structure exceeding 20 nt, and ii) high predicted thermodynamic stability, as indicated by a z-score of less than  $-3.5$ . To enrich further for true miRNA precursors, we selected sequences displaying the bimodal conservation pattern typical of microRNAs [6] by visually inspecting the “conservation” track in the UCSC genome browser (Figure 1). As a number of the known miRNA loci in the *Drosophila melanogaster* genome show high conservation across the full sequence of the pre-miRNA, we also included regions of constantly high conservation (Data S1, e.g. locus19, manual3). This strategy yielded 47 novel miRNA candidates, an example prediction is shown in Figure 1.

In addition to the RNAz predicted stem-loop structures, we noticed the presence of several additional bimodally-conserved sequences close to some of the identified candidates. As many known miRNAs are found in clusters in animal genomes, we tested these sequences for their potential to form stable hairpin structures using RNAalifold [39]. In this way, nine additional sequences were added as “manual” predictions to the list of candidates, giving rise to several predicted clusters of miRNAs (Figure 2 and Data S1). In one of them, intriguingly miRNAs are predicted in two groups flanking an exon of the protein-coding gene CG31646 (Figure 2C, cluster 3).

### Experimental validation of predictions

Like protein-coding genes, miRNAs are often expressed only at specific developmental stages or in a subset of tissues. To test if

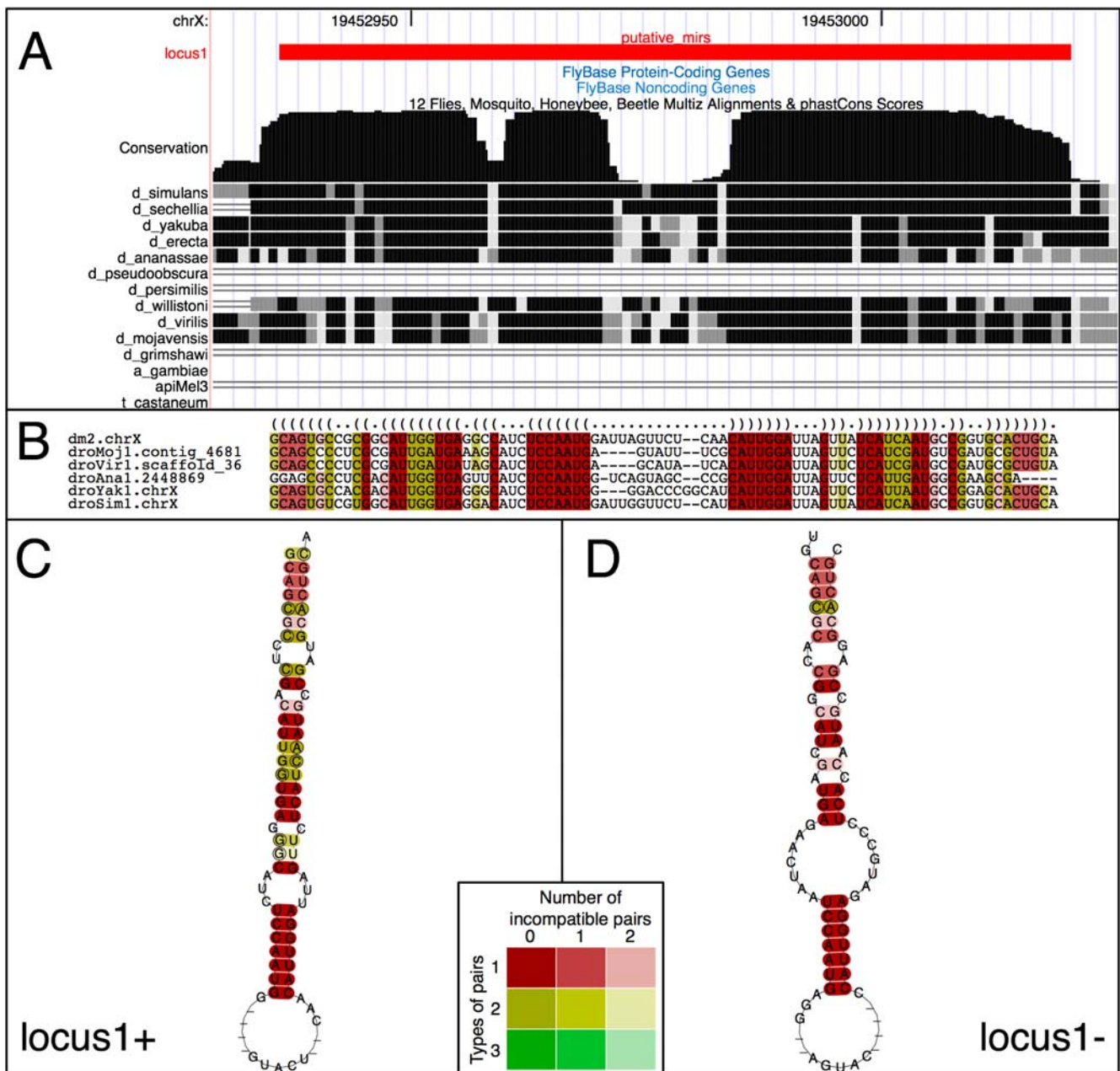
mature microRNAs can be processed from primary transcripts originating from the predicted genomic sequences in a single experimental system, we cloned the genomic regions covering a number of putative novel miRNA loci and expressed them under their control of the actin 5C promoter by transfection into *Drosophila* S2 cells. The majority of candidate sequences had significant RNAz scores ( $>0.9$ ) in both reading directions (53/56). The predicted hairpins were often symmetrical. These candidates were therefore expressed in both orientations to direct production of primary transcripts from both strands. Processing of these long primary transcripts was assayed by Northern blotting, and detected by hybridization with oligonucleotide probes directed against each arm of the predicted hairpin. Out of 33 candidates tested, 16 gave rise to a distinct small RNA (Table 1). Five of these proved to be endogenously expressed in S2 cells and could therefore be detected in untransfected cells.

While this experimental analysis was nearing completion, high-throughput sequencing data was published for small RNA libraries cloned from ten different samples, including the major developmental stages of *Drosophila melanogaster* development [12] [GEO:GSE7448]. This allowed us to compare our experimental data with an independent unbiased global expression analysis. We mapped these small RNAs to the non-repetitive part of the *Drosophila* genome and used RNAfold [40] to screen for sequenced RNAs falling into regions predicted to fold into an extended hairpin structure. This way, 53 novel miRNA loci were identified. 19 of these corresponded to one of our 56 microRNA predictions (Table 1). 13 of these were miRNAs independently validated by our Northern blot analysis. In each case small RNA sequences mapped only to a single strand, corresponding to the previously identified direction of transcription. In agreement with the stepwise processing of pre-miRNA precursors, 21–25 long oligonucleotides often originated from both arms of the hairpin, albeit at different frequencies: the mature miRNA outnumbers its cognate miRNA\* sequence with a ratio up to 1713:3 (locus12, Figure 3A).

We have used available sequence data from ten different anatomies or developmental stages to validate and complement our predictions, pushing the total number of annotated miRNAs to 146 - beyond the previous expectation of  $\sim 120$ . Although there is considerable overlap between the prediction/validation and sequence data sets, comparing them showed that neither set is complete. 19/53 of the novel candidates (38.2%) were identified both computationally and by sequencing. In addition, the small RNA libraries provided evidence for 31 additional loci that we had not predicted. Three predicted miRNAs (locus26, locus7 and manual3) could be detected in untransfected S2 cells and in developing embryos by Northern blot analysis (Figure 4), but were not found in the sequencing data. Locus7 and manual3 are expressed strongly in early stages and decline in abundance toward the end of embryogenesis, whereas locus26 expression begins low and increases during mid- to late embryogenesis (Figure 4E–G). In aggregate, this illustrates that any single approach aimed at identifying miRNAs continues to underestimate the miRNA complement of the *Drosophila melanogaster* genome.

### Mature miRNA sequences

By identifying the small RNA sequences detected most frequently in the small RNA libraries, we could pinpoint the mature microRNAs. Interestingly, eight of the novel miRNAs share identical seed sequences with known miRNAs (Table S1). The first eight bases of the mature manual7 sequence correspond to that of another *Drosophila melanogaster* miRNA, *mir-12*, suggesting that both miRNAs could regulate a similar set of targets (Figure 5A). While *mir-12* was detected in samples from all developmental stages/



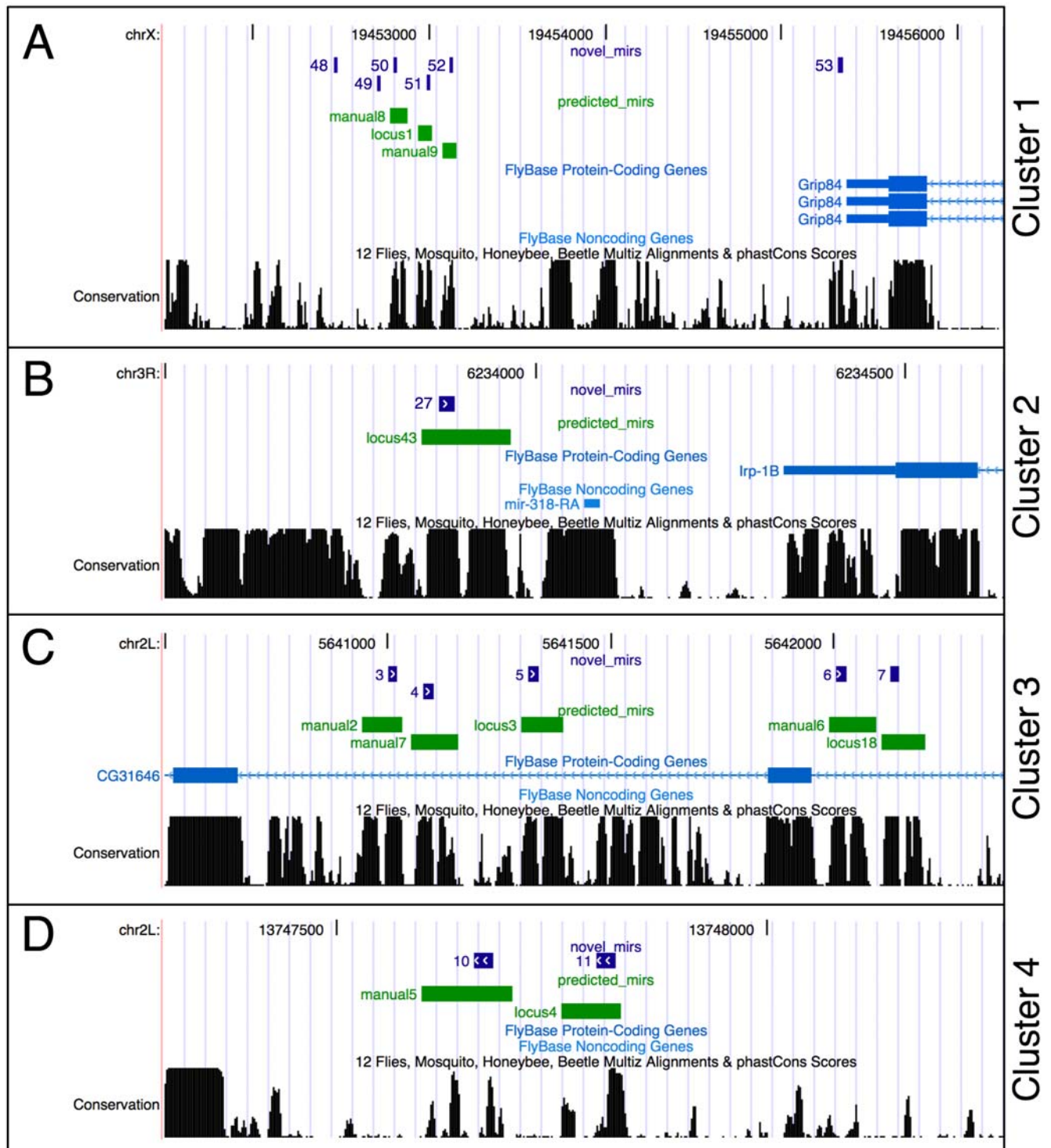
**Figure 1. Example of a pre-miRNA predicted by RNAz. A)** Locus1 shows the bimodal conservation pattern typical of a conserved miRNA hairpin structure in the phastcons track of the UCSC genome browser. Sequences from *Drosophilids* as distant as *D. mojavensis* contribute to the multiz alignment at this genomic position. **B)** Sequences from six species were chosen as input for RNAz and lend different levels of support to stabilizing selection of the predicted secondary structure prediction (first row). The color code indicates the number of different base-pairs (green = 3 pairs to red = 1 pair) and the number of pair incompatible with the predicted structure (dark color = 0 to faint color = 2) at each position. **C, D)** Locus1 is predicted to fold into symmetrical hairpins in both possible directions of transcription (color code as in B). doi:10.1371/journal.pone.0001265.g001

tissues with the exception of the early embryo (Figure 5B), mature manual7 is highly enriched in imaginal disks (80.9% of all sequence hits, Figure 5C), where it might tighten the repression of the common set of targets specifically at this stage of development. Similarly, “seed paralogs” were found for *Drosophila mir-279*, *mir-285* and *mir-286*. Also, several of the newly identified miRNAs have unique seeds in this model organism, but similar seeds are found in miRNAs of other species, including vertebrates (Table S1). In some cases, the alignment between novel and previously described miRNAs extends well beyond the seed sequences,

suggesting a common evolutionary ancestry. Locus11 shares e.g. 14 out of 22 bases with vertebrate *mir-22*, a microRNA highly conserved in vertebrates but until now lacking a *Drosophila* ortholog (Figure 5D, Table S1).

**Clusters**

Several of the predicted miRNAs were found in distinct experimentally validated genomic clusters (Figure 2). To investigate whether miRNAs from each genomic cluster are likely to be processed from the same primary transcript, we compared



**Figure 2. Prediction of four novel miRNA clusters.** A) locus1, manual8, manual9 form a tight cluster downstream of the *Grip84* locus (green). Small RNAs detected by sequencing map to each of the novel miRNA loci (matches 50–52, blue). In addition, sequences map to two loci in the vicinity, giving rise to identical mature miRNAs (matches 48,49). Finally, a sixth locus is detected immediately downstream of *Grip84* (match 53), located downstream of the *Irp1B* locus, clusters together with novel locus41 (green and small RNA sequence match 28, blue) B) *mir-318*, located downstream of the *Irp1B* locus, clusters together with novel locus41 (green and small RNA sequence match 28, blue) C) Five novel miRNAs were predicted (green) and validated (blue) on either side of an exon of the CG31646 locus. D) Novel locus4 and manual5 (green) give rise to mature miRNAs (blue) and form cluster4.  
 doi:10.1371/journal.pone.0001265.g002

the relative frequencies of the mature miRNA from these validated clusters in different RNA libraries (Figure 6B–E). Although the absolute number of sequence hits varies between

the different members of the same cluster (e.g. 19 hits for locus1 and 162 hits for manual9 in cluster1), their relative abundance in different sequenced tissues is remarkably similar. Comparable

**Table 1.** Validation of predicted miRNAs

ID	Chr	Start	Stop	Strand	processed in S2 cells	validated by sequencing
locus1	X	19452936	19453020	+	Y	Y
locus2	2L	6830211	6830286		N	N
locus3	2L	5641300	5641395	+	endogenous	Y
locus4	2L	13747762	13747830	–	Y	Y
locus6	3L	10312110	10312214		N	N
locus7	3R	24628381	24628452	–	endogenous	N
locus8	3R	6842228	6842321		N	N
locus9	3R	10816276	10816378		N	N
locus11	X	174206	174307	–	Y	Y
locus12	X	1682883	1682986	–	Y	Y
locus13	3R	21414577	21414677	–	Y	Y
locus14	3R	20313365	20313451		N	N
locus15	X	15859994	15860084	–	Y	Y
locus17	2L	243033	243125	–	N	Y
locus18	2L	5642109	5642206	+	Y	Y
locus20	2L	6902058	6902149	+	Y	Y
locus23	2L	15479411	15479482		N	N
locus26	2R	1705689	1705773	+	endogenous	N
locus27	2R	3456100	3456167		N	N
locus39	3R	11559905	11559991		N	N
locus40	3L	10153900	10153966		N	N
locus43	3R	6233847	6233967	+	Y	Y
locus45	3L	17699260	17699368		N	N
locus46	3R	4357258	4357332		N	N
locus47	3R	121092	121163	+	n.a.	Y
locus48	3R	2602056	2602170	+	n.a.	Y
manual1	X	16362566	16362654		N	N
manual2	2L	5640941	5641032	+	N	Y
manual3	2L	6829891	6829990	+	endogenous	N
manual4	X	16362736	16362841		N	N
manual5	2L	13747600	13747705	–	Y	Y
manual6	2L	5641992	5642096	+	N	Y
manual7	2L	5641054	5641158	+	endogenous	Y
manual8	X	19452776	19452876	+	N	Y
manual9	X	19453074	19453153	+	Y	Y

The processing of primary transcripts ectopically expressed in S2 cells under the control of the actin 5C promoter was analyzed by Northern blot using probes designed against both arms of the predicted hairpins. In 16 cases, mature sequences could be detected (column 6, Y), five of which were present in untransfected S2 cells. For the remaining predictions, no processing was observed (column 6, N). In most cases, small RNAs corresponding to the probed sequences could also be detected in published small RNA libraries (13/16). Six miRNAs either not assayed (column 5, n.a.) or not detected by Northern blot could be validated, too. No matching sequences were found in the libraries for locus7, locus25 and manual3, which can be detected as endogenous transcripts (column 6, endogenous) by Northern blot.

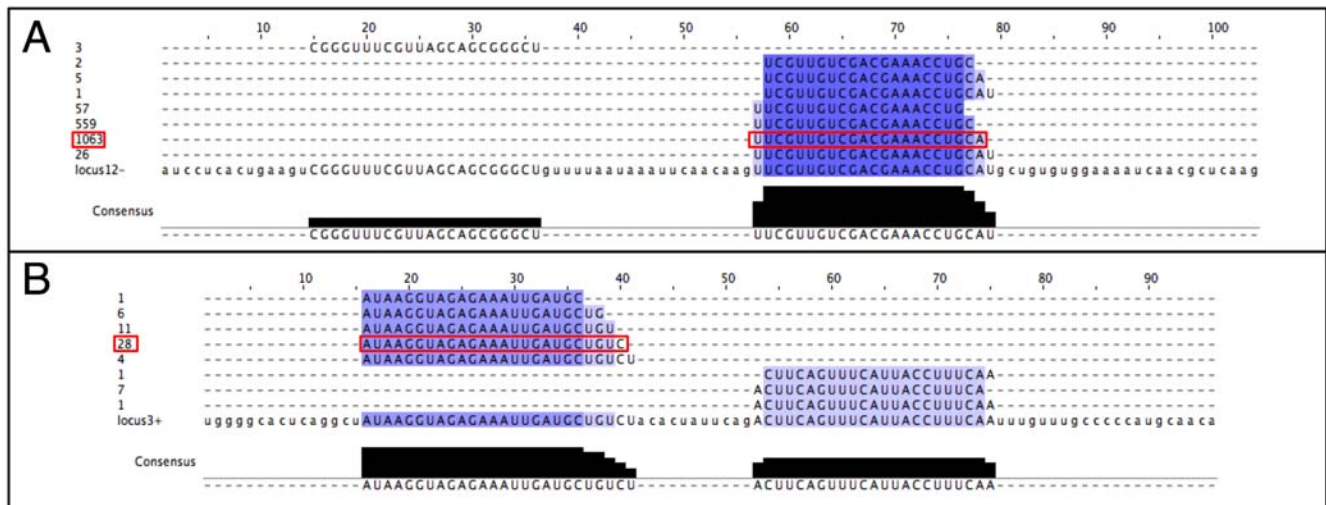
doi:10.1371/journal.pone.0001265.t001

results were obtained with a known miRNA cluster (*mir-310*, *mir-311*, *mir-312* and *mir-313*, Figure 6A), suggesting that the novel clusters may also encode polycistronic transcripts. While the miRNAs of cluster1, cluster2 and cluster3 seem to be unrelated in sequence, cluster4 gives rise to two very similar mature sequences (Figure 6F). Manual5 carries an additional U at its 5' end compared to locus4, giving rise to a novel putative seed sequence. As both mature sequences clearly outnumber any other matching oligonucleotides mapped to the same locus (Table 1), locus4 and manual5 most likely represent distinct, recently diverged micro-RNAs.

## Conclusions

By combining RNAz, a “generic” tool to predict non-coding RNAs, with specific filters to identify miRNA precursors, we have taken advantage of the recent sequencing of a large number of insect species. The fact that 22 out of our 56 predictions (39%) could be verified by Northern blot analysis or sequencing, but were missed in previous studies, demonstrates the power of this approach.

The degree of sequence conservation of miRNA precursors between related species varies widely (compare e.g. the phastcons tracks for manual5 and locus48, Data S1). The genome-wide



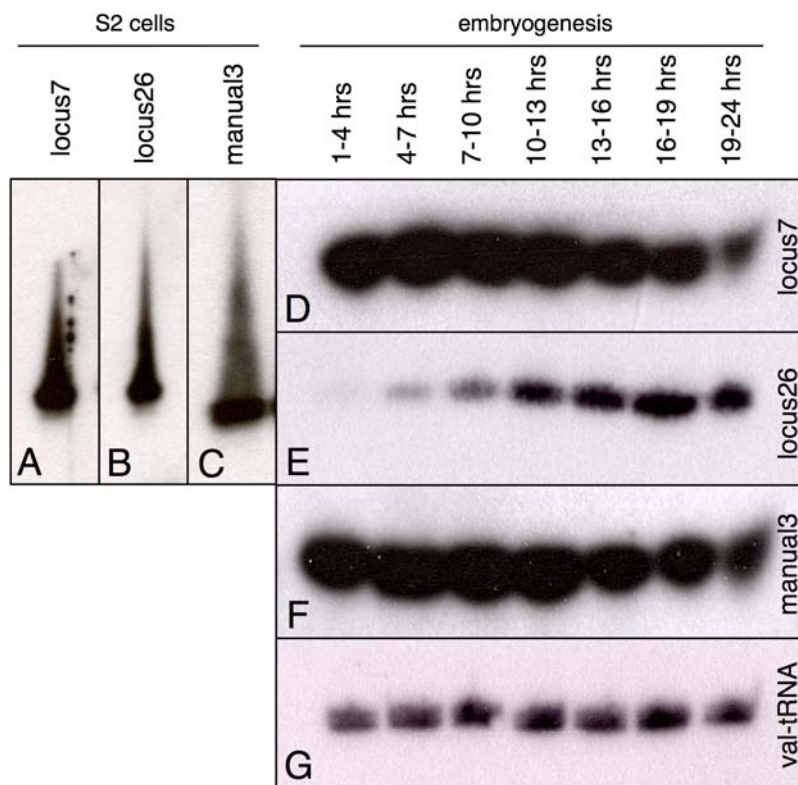
**Figure 3. Examples of small RNA sequences mapping to predicted miRNA loci.** Small RNA sequences recovered by sequencing [12] were mapped to the predicted loci using Megablast. **A)** locus12 is transcribed from the ‘-’ strand and multiple overlapping sequences map to its coordinates. Numbers on the left indicate how often each sequence was identified in the small RNA libraries. The most abundant species (red box) most likely represents the mature miRNA. **B)** locus3 maps to the ‘+’ strand. Details for all predicted loci is available as supplemental data. doi:10.1371/journal.pone.0001265.g003

alignments of 9 or 15 insects enabled us to select suitable sequences for each candidate region separately, allowing for sequence variation of rapidly as well as slowly diverging miRNAs. This approach is likely to prove useful for the analysis of other animal genomes, as more related genomes are sequenced.

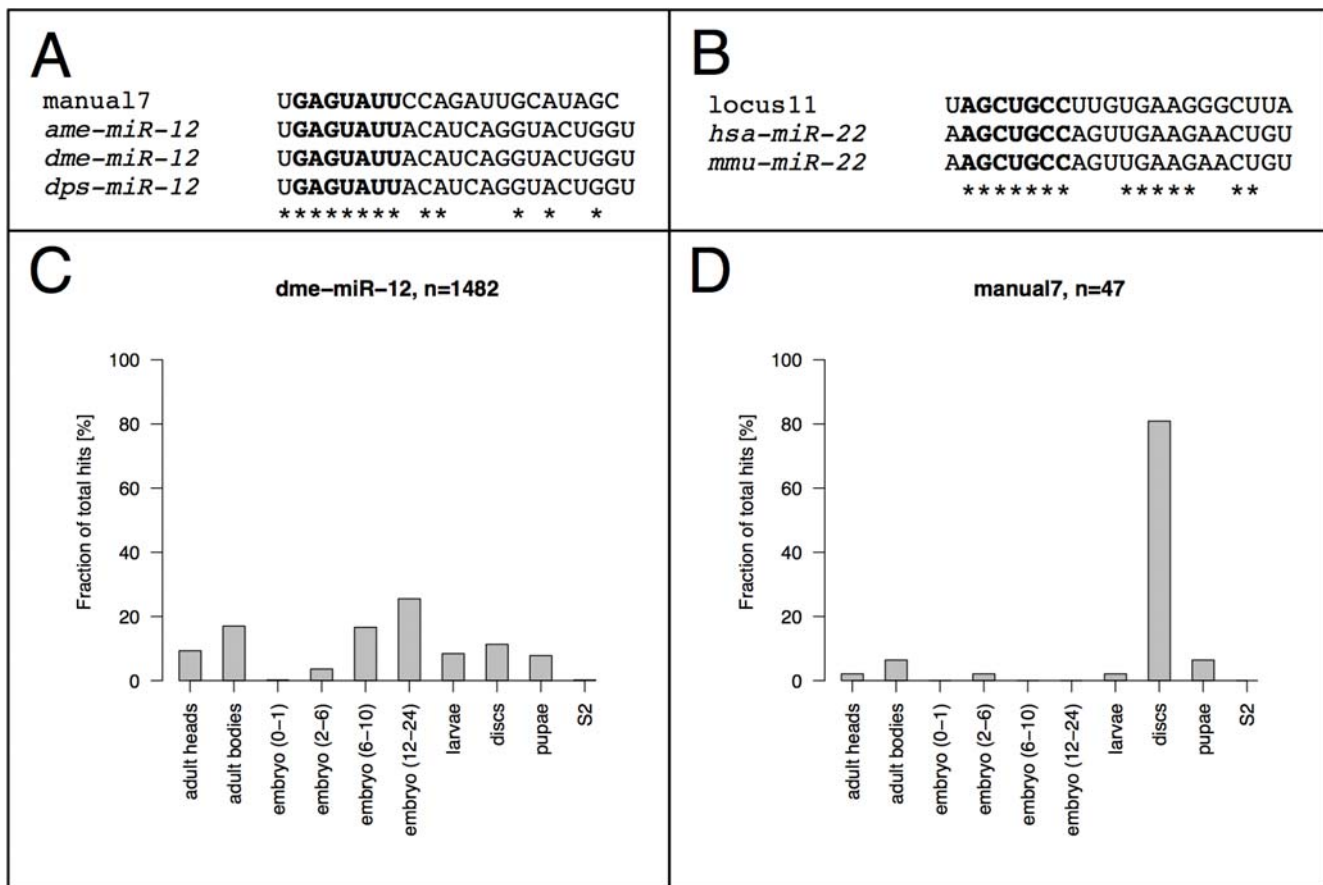
## MATERIALS AND METHODS

### Computational miRNA predictions

Genome-wide alignments of 9 (multiz9) and 15 (multiz15) insect species were downloaded from the UCSC genome browser.



**Figure 4. Northern blot validation of three miRNAs not found in small RNA libraries.** Three predicted miRNAs were not represented in the small RNA libraries, yet could be detected by Northern blot analysis. **A–C)** Endogenous small RNAs were detected with probes directed against locus7, locus26 and manual3 in total RNA extracted from untransfected S2 cells. **D–F)** All three transcripts were also detected in RNA from different embryonic stages (time after egg-laying). **F)** A probe against valine tRNA was used as a loading control. doi:10.1371/journal.pone.0001265.g004



**Figure 5. miRNAs sharing seed sequences.** A) The novel miRNA manual7 contains the same seed sequence as the *D. melanogaster mir-12* miRNA and its orthologs. B) Sequence similarity between the novel locus11 and vertebrate *miR-22* extends beyond the seed region, hinting at a possible common ancestry of these miRNAs. C) The mature sequence of *Drosophila mir-12* was detected more than 1400 times in the small RNA libraries in samples from all developmental stages/tissues with the exception of the early embryo. D) The most abundant sequence mapping to the novel manual7 locus is distributed differently over the 10 sequenced samples, with RNA from discs contributing >80% of hits. doi:10.1371/journal.pone.0001265.g005

Starting from the “most conserved” track for each alignment generated by the Phastcons program, we padded the genomic coordinates by 25 bps on both sides. Neighboring sequences closer than 50 bps to each other were combined into a single region. All sequences overlapping annotated genes according to the “Flybase genes” and “Refseq genes” tracks of the UCSC genome browser were removed.

Input windows were prepared with the rnazWindow.pl script of the RNaz package with default settings. This resulted in alignments with a maximum length of 200 bps (longer alignments were split into overlapping windows of 120 bps with a step size of 40) and a minimum number of aligned sequences (>50% identity) from at least 4 different species. RNaz (version 1.0) was used to score the alignments and hits with a probability score  $p > 0.9$  were stored. Overlapping predictions were joined using the rnazCluster.pl script of the RNaz package and filtered for hairpin length (>20 bps, total length of the prediction >65 bps), thermodynamic stability (z-score < -3.5). The remaining predictions were uploaded to a local installation of the UCSC genome browser and their multiz15 phastcons conservation profile inspected manually.

### Cloning and expression of putative miRNAs in S2 cells

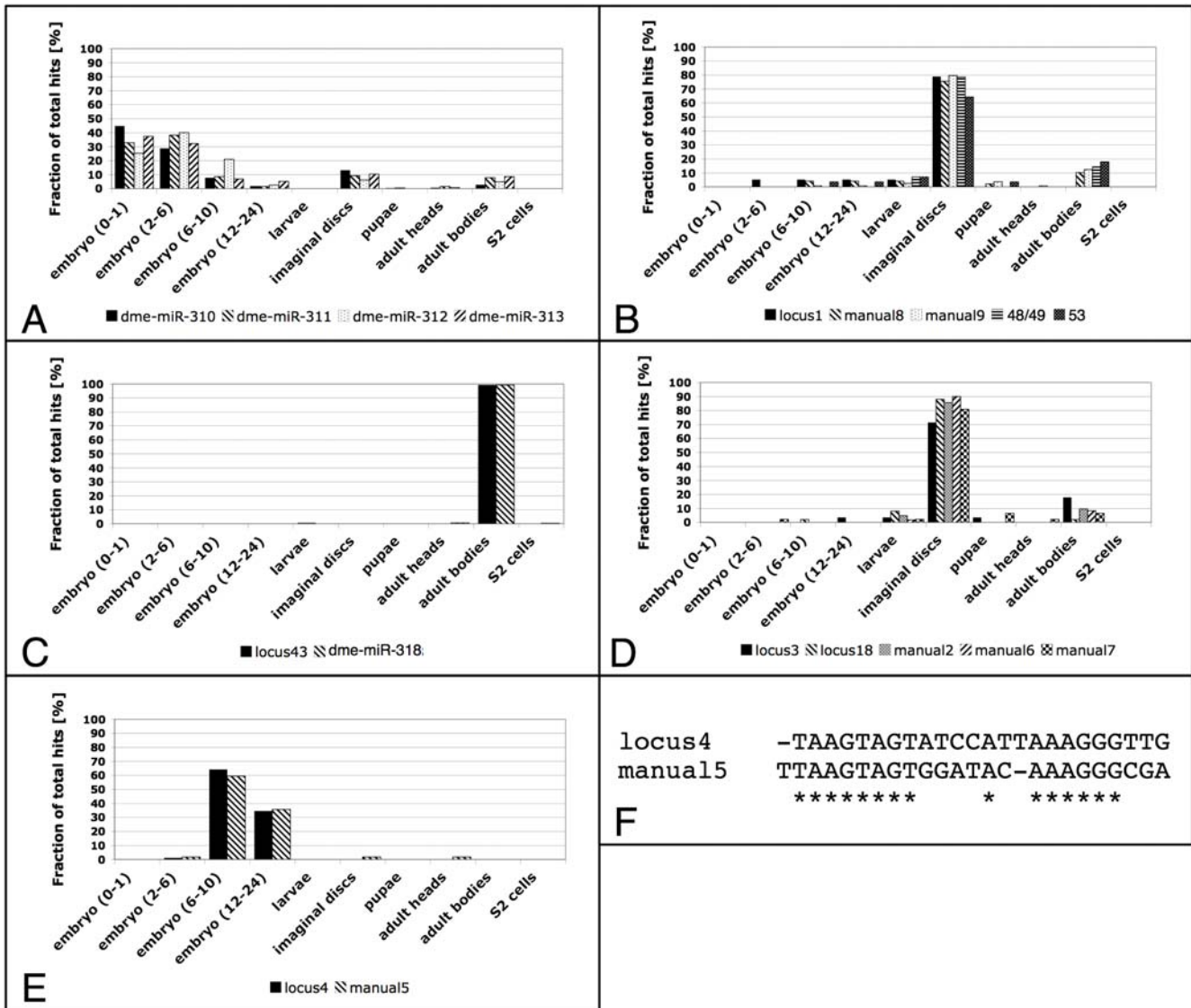
Amplicons of ca. 0.5 kb length, centered on the predicted pre-miRNA were generated by PCR and cloned into vector pAc5.1B

(Invitrogen). S2 cells were grown in the absence of serum, transfected using Cellfectin (Invitrogen) and incubated for 3 days at 25°C. Total RNA was extracted from transfected or control cells, as well as from staged collections of *Drosophila* embryos using Trizol (Invitrogen) and precipitated with isopropanol overnight at -20°C. 15 µg of total RNA was separated on a 15% denaturing polyacrylamide gel and blotted using a chemical cross-linking procedure [41]. Oligonucleotide probes were designed separately against each full arms of the predicted hairpin and end-labeled with <sup>32</sup>P-ATP using T4 polynucleotide kinase (NEB).

### Mapping small RNA sequencing data to predicted loci

Sequences detected in *Drosophila* small RNA libraries [12] were obtained from GEO [GEO:GSE7448] and mapped to the putative pre-miRNA sequences using megablast [42] with wordsize 14 and a score cutoff of 16. Only perfectly matched sequence matches were retained. The resulting alignments were displayed with Jalview [43].

miRNA loci were identified from the small RNA libraries *de novo* by moving a 110 bp window (stepsize 5) over all bases recovered at least 10 times. RNAfold [40] was used to predict the secondary structure of each sequence window and hairpins with more than 30 base-paired nucleotides and bulges of less than 15 nucleotides



**Figure 6. Four novel clusters encoding tightly co-expressed miRNAs.** Though the individual number of sequence hits recovered in the small RNA libraries differs greatly between different miRNAs, the relative abundance in the different sequenced samples is highly correlated for miRNAs originating from the same cluster. **A)** Similar fractions of the four members of the *Drosophila* *mir-310-313* cluster are recovered in embryonic, imaginal disc and adult body samples, but virtually absent from the other samples. **B–E)** Members of the four novel miRNA clusters are distributed equally tightly over the small RNA libraries, demonstrating tight co-regulation. **F)** The mature sequences of locus4 and manual5 are highly similar and map to neighboring genomic coordinates (E), suggesting a duplication event that gave rise to two independent loci.  
doi:10.1371/journal.pone.0001265.g006

were retained. Repetitive regions or sites overlapping with annotated loci were removed.

**SUPPORTING INFORMATION**

**Data S1** Pdf document containing information about all validated novel miRNAs, including RNaz alignment and structure models.  
Found at: doi:10.1371/journal.pone.0001265.s001 (2.17 MB PDF)

**Table S1** Genome data on 53 new miRNAs. Information about all 53 validated novel miRNAs, including their genomic coordinates (genome version 5) and their mature sequences.

Found at: doi:10.1371/journal.pone.0001265.s002 (0.02 MB PDF)

**ACKNOWLEDGMENTS**

We would like to thank Natascha Bushati, Lars J. Jensen, Charles Girardot and Julian Gagneur (EMBL Heidelberg, Germany) for helpful discussion and valuable advice. Thomas Sandmann is a Fellow of the Singapore Millennium Foundation.

**Author Contributions**

Conceived and designed the experiments: SC TS. Performed the experiments: TS. Analyzed the data: TS. Wrote the paper: SC TS.



## REFERENCES

- Lee R, Feinbaum R, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843–854.
- Reinhart B, Slack F, Basson M, Pasquinelli A, Bettinger J, et al. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901–906.
- Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855–862.
- Zhao T, Li G, Mi S, Li S, Hannon GJ, et al. (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev* 21: 1190–1203.
- Griffiths-Jones S, Grocock R, van Dongen S, Bateman A, Enright A (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140–144.
- Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk R, et al. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 120: 21–24.
- Brennecke J, Cohen S (2003) Towards a complete description of the microRNA complement of animal genomes. *Genome Biol* 4: R28.
- Kim V, Nam J (2006) Genomics of microRNA. *Trends Genet* 22: 165–173.
- Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, et al. (2004) A large imprinted microRNA gene cluster at the mouse *Dlk1-Gtl2* domain. *Genome Res* 14: 1741–1748.
- Bushati N, Cohen S (2007) microRNA Functions. *Annu Rev Cell Dev Biol*.
- Okamura K, Hagen J, Duan H, Tyler D, Lai E (2007) The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*. *Cell*.
- Ruby J, Jan C, Bartel D (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*.
- Khvorovaya A, Reynolds A, Jayasena S (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115: 209–216.
- Schwarz D, Hutvagner G, Du T, Xu Z, Aronin N, et al. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115: 199–208.
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115: 787–798.
- Stark A, Brennecke J, Russell RB, Cohen SM (2003) Identification of *Drosophila* MicroRNA targets. *PLoS Biol* 1: E60.
- Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. *PLoS Biol* 3: e85.
- Kim V (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 6: 376–385.
- Lim L, Lau N, Garrett-Engle P, Grimson A, Schelter J, et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433: 769–773.
- Tang F, Kaneda M, O'Carroll D, Hajkova P, Barton S, et al. (2007) Maternal microRNAs are essential for mouse zygotic development. *Genes Dev* 21: 644–648.
- Cao X, Pfaff S, Gage F (2007) A functional study of miR-124 in the developing neural tube. *Genes Dev* 21: 531–536.
- Neilson J, Zheng G, Burge C, Sharp P (2007) Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev* 21: 578–589.
- Thompson B, Cohen S (2006) The Hippo pathway regulates thebantam microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell* 126: 767–774.
- Teleman AA, Maitra S, Cohen SM (2006) *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev* 20: 417–422.
- He L, He X, Lim L, de Stanchina E, Xuan Z, et al. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*.
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, et al. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 33: D121–124.
- Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, et al. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7: 25.
- Lai E, Tomancak P, Williams R, Rubin G (2003) Computational identification of *Drosophila* microRNA genes. *Genome Biol* 4: R42.
- Wang X, Zhang J, Li F, Gu J, He T, et al. (2005) MicroRNA identification based on sequence and structure alignment. *Bioinformatics* 21: 3610–3614.
- Washietl S, Hofacker I, Lukasser M, Huttenhofer A, Stadler P (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 23: 1383–1390.
- Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 23: 1383–1390.
- Pedersen J, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, et al. (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* 2: e33.
- Celniker SE, Rubin GM (2003) The *Drosophila melanogaster* genome. *Annu Rev Genomics Hum Genet* 4: 89–117.
- Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, et al. (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 5: 337–350.
- Brennecke J, Aravin A, Stark A, Dus M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103.
- Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, et al. (2007) The UCSC genome browser database: update 2007. *Nucleic Acids Res* 35: D668–673.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034–1050.
- Hsu PW, Huang HD, Hsu SD, Lin LZ, Tsou AP, et al. (2006) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res* 34: D135–139.
- Hofacker IL, Fekete M, Stadler PF (2002) Secondary structure prediction for aligned RNA sequences. *J Mol Biol* 319: 1059–1066.
- Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31: 3429–3431.
- Pall G, Codony-Servat C, Byrne J, Ritchie L, Hamilton A (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res* 35: e60.
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7: 203–214.
- Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* 20: 426–427.