

MirZ: an integrated microRNA expression atlas and target prediction resource

Jean Hausser¹, Philipp Berninger¹, Christoph Rodak^{1,2}, Yvonne Jantscher³, Stefan Wirth⁴ and Mihaela Zavolan^{1,*}

¹Biozentrum, Universität Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, 4056 Basel, ²Fachhochschule Nordwestschweiz, Schulthess-Allee 1, 5201 Brugg, Switzerland, ³Fachhochschule Hagenberg, Softwarepark 11, 4232 Hagenberg, Austria and ⁴Institut für Informatik, Universität Leipzig, Härtelstr. 16-18 04107 Leipzig, Germany

Received March 3, 2009; Revised April 25, 2009; Accepted May 2, 2009

ABSTRACT

MicroRNAs (miRNAs) are short RNAs that act as guides for the degradation and translational repression of protein-coding mRNAs. A large body of work showed that miRNAs are involved in the regulation of a broad range of biological functions, from development to cardiac and immune system function, to metabolism, to cancer. For most of the over 500 miRNAs that are encoded in the human genome the functions still remain to be uncovered. Identifying miRNAs whose expression changes between cell types or between normal and pathological conditions is an important step towards characterizing their function as is the prediction of mRNAs that could be targeted by these miRNAs. To provide the community the possibility of exploring interactively miRNA expression patterns and the candidate targets of miRNAs in an integrated environment, we developed the MirZ web server, which is accessible at www.mirz.unibas.ch. The server provides experimental and computational biologists with statistical analysis and data mining tools operating on up-to-date databases of sequencing-based miRNA expression profiles and of predicted miRNA target sites in species ranging from *Caenorhabditis elegans* to *Homo sapiens*.

INTRODUCTION

MicroRNAs (miRNAs) are a continuously growing class of small RNAs that act as guides in the translational silencing and degradation of target mRNAs (1). Many miRNAs are conserved over large evolutionary distances such as between human and worm (2). Fundamental biological processes such as development (3–6), metabolism (7–9), cardiac (10) and immune system function (11) have

been shown to be regulated by miRNAs, and aberrant miRNA expression has been associated with cancers (12,13).

There are various approaches to miRNA expression profiling, one of which is small RNA sequencing. Classical cloning and sequencing of size-separated small RNAs have been used to generate a large atlas of miRNA expression profiles (14), and this approach can be scaled up considerably through deep sequencing technologies (15). Microarray-based expression profiling is also a popular approach, which has been used for instance to characterize the miRNA expression cancer samples (12). In contrast to sequencing, microarray-based profiling does not allow identification of novel miRNAs.

Numerous approaches have also been proposed for miRNA target prediction. Because the 5'-end of miRNAs (known as 'seed') appears to be important for target recognition, a number of tools focus on the evolutionary conservation of miRNA seed-complementary regions in 3'-UTRs (16–19). Other approaches emphasize the energy of hybridization between miRNA and target (20–22), the expected anti-correlation between the expression level of miRNAs and their mRNA targets (23,24), the properties of the environment of the miRNA target site (25,26) or combine various features of the miRNA target site itself (27,28).

Studies in both native expression (29) as well as transfection-induced miRNA overexpression situations (30,31) indicate that within a given tissue, the miRNAs that are most strongly expressed have the largest impact on mRNA targets. For this reason, deciphering the miRNA-dependent post-transcriptional regulatory layer in a given tissue or cell type needs to start from the miRNA expression profile of that tissue or cell type. Conversely, it is very common that one identifies differences in miRNA expression between cells at various stages of differentiation or between normal and malignant cells, and the natural question is what mRNAs are most likely to be affected by the change in miRNA expression.

*To whom correspondence should be addressed. Tel: +41 61 267 1577; Fax: +41 61 267 1584; Email: mihaela.zavolan@unibas.ch

To address these types of questions, we developed MirZ (www.mirz.unibas.ch), a web service that integrates two resources that we developed in the context of previous research projects: the smiRNadb miRNA expression atlas (14), and the EIMMo miRNA target prediction algorithm (18).

MATERIALS AND METHODS

The smiRNadb miRNA expression atlas

smiRNadb (14) is a unique, web-accessible and widely used resource of miRNA profiles determined by sequencing from hundreds of *Homo sapiens*, *Mus musculus* and *Rattus norvegicus* samples. The web interface of smiRNadb features an extended repertoire of on-line analyses such as visualization and hierarchical clustering of miRNA expression profiles, principal component analysis, comparison of miRNA expression between two (sets of) samples with the aim of identifying the miRNAs whose expression differs most between the samples. We used the Brenda tissue ontology (32,33) as a guide in organizing the samples such that the user can readily identify related cell lineages or normal and pathological samples derived from a given tissue type. Our tissue hierarchy has four levels: the organ/system (e.g. hematopoietic system), subsystem (e.g. lymphoid lineage), cell type (e.g. B cell) and further cell type classification (e.g. B lymphocyte). MiRNAs themselves can be analyzed independently, grouped by their 2–7 subsequence, or grouped in precursor clusters. Two miRNAs are placed in the same precursor cluster if their loci are within 50 kb of each other in the genome, or if they share a mature form.

As an example, one may be interested in comparing miRNA expression between effector and naive human CD4⁺ T lymphocytes. SmiRNadb features a ‘Sample comparison’ tool, which was specifically designed for the pairwise comparison of miRNA (sets of) samples. The user would select to compare the sample named ‘hsa_T-cell-CD4-effector’ to the sample named ‘hsa_T-cell-CD4-naive’. Because the naive CD4⁺ T cell sample and the effector CD4⁺ T cell sample differ widely in the total number of sequenced miRNAs (1374 versus 89), the precision of the miRNA frequency estimates in the two samples will also be very different. This situation is common in sequencing-based datasets making the identification of miRNAs whose expression is *significantly* different a non-trivial problem. At the heart of the tools offered by smiRNadb, however, is a Bayesian model for computing the posterior probability that the *frequency* of a miRNA in the total miRNA population differs between two (sets of) samples. We compute this probability assuming a binomial sampling model and integrating over the unknown miRNA frequencies in the samples. This approach—described in details in Berninger *et al.* (34)—takes into account both the variability between sample sizes and the absolute miRNA counts.

Figure 1 shows the results of comparing the miRNA expression profiles of naive versus effector CD4⁺ cells. The names and sizes of the samples being compared are shown at the top of the page, followed by the log-likelihood

ratio $\log(P_{\text{same}}/P_{\text{diff}})$ of two models, one which assumes that the frequencies of miRNAs are the same and one that assumes that they can be different between the samples. The log-likelihood ratio takes positive values when the miRNA frequencies are similar and negative values when they are different. In this case, the log-likelihood ratio is positive, indicating that overall, the frequencies of miRNAs in these samples are more likely to have been the same. The list of miRNAs ranked from most dissimilar to most similar expression follows. Each row contains the name of an miRNA, the direction of regulation (up or down), the cloning counts and frequencies in both samples and provides a direct link to the predicted targets of the miRNA. The model indicates that with a 18% versus 54% cloning frequency, and despite the small size of the effector CD4⁺ T-cell sample, miR-142-5p is very likely to be downregulated in effector cells. Again, this can be inferred from the negative value of $\log(P_{\text{same}}/P_{\text{diff}})$ for miR-142-5p. From this page, the user can select one or several miRNAs that came out differentially expressed and can browse the list of predicted targets (Figure 2). In the case of miR-142-5p, the top 10 predicted targets include four transcription factors (AFF4, ONECUT2, ZFPM2 and ZNF148), and a kinase (PRPF4B) involved in pre-RNA splicing. These genes could provide a starting point for experimental studies on the function of miR-142-5p in T lymphocytes.

Since the original release of smiRNadb, we have implemented an additional tool for performing principal component analysis on the miRNA expression profiles, we added more possibilities for the user to download miRNA profile data for further processing and we started to incorporate other publicly available small RNA sequencing datasets from *Danio rerio*, *Drosophila melanogaster* and *Caenorhabditis elegans*. We reimplemented the software that was originally written in Perl CGI to use Java Server Faces technology and Apache/Tomcat. The computations are now performed on a computing cluster, with job distribution managed by the Sun Grid Engine queuing system. Finally, we enhanced the result screens of our on-line analysis tools with hyperlinks that directly take the user to the miRNA target predictions within the context of the smiRNadb query, i.e. preserving the selected organism, miRNAs and tissue (if available). Refer to the web connectivity map in the Supplementary Material for an overview of the new links between smiRNadb and EIMMo, as well as of the external resources that we use in performing various analyses.

The EIMMo miRNA target prediction algorithm based on comparative genomic analysis

To be able to address the question of what mRNA is most likely affected by the change in expression of a miRNA, we coupled smiRNadb to a PHP-based web interface to the EIMMo miRNA target predictions (18).

Returning to the example of the hsa-miR-142-5p miRNA that was highlighted in ‘The smiRNadb miRNA expression atlas’ section, the web interface allows aside from browsing the predicted targets, a number of other queries. For instance, given an organism (*H. sapiens* in

Comparison Results

Download Results(raw file)

Pool1: 89.0 miRNA clones

- T-cell-CD4-effector

vs

Pool2: 1374.0 miRNA clones

- T-cell-CD4-naive

Log(Psame/Pdiff): 88.64

	miRNA Name	Direction	Pool1 count	Pool1 frequency	Pool2 count	Pool2 frequency	Log(Psame/Pdiff):	EIMMo Target Prediction
1	hsa-miR-142-5p	↓	16	0.18	739	0.538	-20.5082	<input checked="" type="checkbox"/>
2	hsa-miR-374a	↑	3	0.034	0	0.0	-5.66016	<input type="checkbox"/>
3	hsa-miR-99b	↑	2	0.023	0	0.0	-2.83377	<input type="checkbox"/>
4	hsa-miR-124	↑	2	0.023	0	0.0	-2.83377	<input type="checkbox"/>
5	hsa-miR-16	↑	16	0.18	102	0.075	-2.44487	<input type="checkbox"/>
6	hsa-miR-32	↑	3	0.034	3	0.0030	-2.0095	<input type="checkbox"/>
7	hsa-miR-126	↑	3	0.034	3	0.0030	-2.0095	<input type="checkbox"/>
8	hsa-miR-21	↑	5	0.057	17	0.013	-0.907366	<input type="checkbox"/>
9	hsa-miR-29b	↑	6	0.068	29	0.022	-0.228505	<input type="checkbox"/>
10	hsa-miR-199a-5p	↑	1	0.012	0	0.0	-0.0181883	<input type="checkbox"/>
11	hsa-miR-144	↑	1	0.012	0	0.0	-0.0181883	<input type="checkbox"/>
12	hsa-miR-99a	↑	1	0.012	0	0.0	-0.0181883	<input type="checkbox"/>
13	hsa-miR-9	↑	1	0.012	0	0.0	-0.0181883	<input type="checkbox"/>
14	hsa-miR-423-3p	↑	1	0.012	0	0.0	-0.0181883	<input type="checkbox"/>
15	hsa-miR-29a	↑	3	0.034	10	0.0080	0.421428	<input type="checkbox"/>
16	hsa-miR-150	↓	0	0.0	36	0.027	0.498061	<input type="checkbox"/>
17	hsa-miR-140-3p	↑	2	0.023	5	0.0040	0.722947	<input type="checkbox"/>

Figure 1. Screenshot of the web page showing the result from comparing miRNA expression of human CD4⁺ effector T cells with the CD4⁺ naive T cells. Details are provided in the text.

this example), the user can choose to scan for predicted miRNA target sites not only the default set of transcripts, which is all known RefSeq (35) mRNAs in the chosen organism, but also subsets of transcripts. The SymAtlas project (36) of the Genomics Institute of the Novartis Research Foundation (GNF) generated microarray-based mRNA expression profiles for a wide range of tissues. These profiles are incorporated in MirZ, giving the user the possibility to restrict target prediction to mRNAs that are expressed in a given cell type. The web interface further allows to scan an arbitrary number of mRNAs for up to 20 miRNAs simultaneously. Alternatively, the user can limit the number of mRNAs to scan to 20 mRNAs and then retrieve predicted target sites in these mRNAs for an arbitrary number of miRNAs.

MiRNAs exert their effector function through ribonucleoprotein complexes (miRNP) that contain, aside from the guiding miRNA a member of the Argonaute family of proteins. The determinants of productive miRNA–target site interactions are not entirely known, but a large body of work (16,37–41) established that perfect complementarity of the 7–8 ns from the 5'-end of the miRNA—the so-called miRNA 'seed'—is critical for target recognition. Although miRNA target sites that do not satisfy this constraint have been described, at the genome-wide level the

accuracy of predicting such sites is low (16,18). Other than perfect seed complementarity, the location of the putative target site within the 3'-UTR (18,25,42), structural accessibility (21,22,43), the nucleotide composition in its vicinity (25,26) and the complementarity of specific positions in the miRNA 3'-end to the target site (25) have all been reported to improve the accuracy of miRNA target prediction, yet the relative importance of these features remains unknown. The EIMMo miRNA target prediction method that we developed is based on a Bayesian model that only uses comparative genomics information. Yet it has as high an accuracy as other widely used target prediction programs that incorporate additional constraints, and measures of predictive performance on a set of experimentally validated miRNA targets in *D. melanogaster* can be found in the article describing the EIMMo method (18). Importantly, our model does not have any free parameters, and can easily accommodate additional species whose genome sequence becomes available.

Going back to our example, Figure 2 shows the EIMMo predictions for miR-142-5p in *H. sapiens*. This result screen is organized in two sections: (i) a miRNA-centric summary featuring per-miRNA target prediction statistics and a figure showing the smiRNAdb tissues where the selected miRNAs are mostly expressed, and (ii) a mRNA-centric

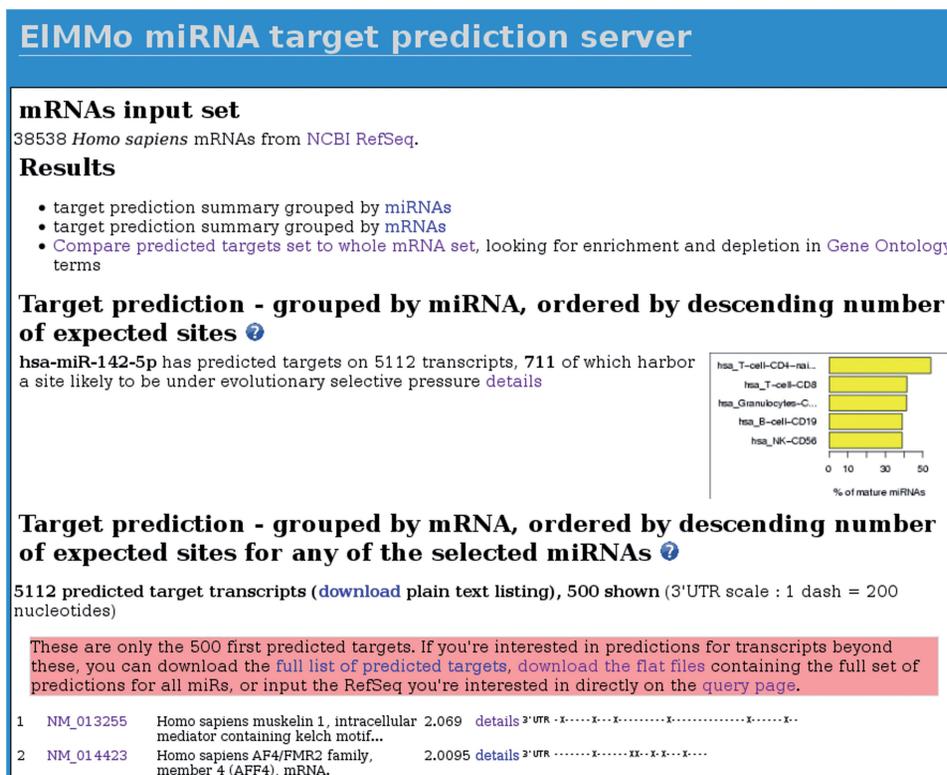


Figure 2. Screenshot of the web page showing the EIMMo miRNA target predictions for miR-142-5p in all *H. sapiens* RefSeq mRNAs. The target predictions results are organized in two sections. The first section—located on the upper part of the web page—is miRNA-centric and features miRNA target predictions statistics as well as a figure showing the smiRNAdb tissues where the miRNA is mostly expressed. The second, mRNA-centric section is located on the lower part of the web page and provides a ranked list of mRNA predicted to be targeted by miR-142-5p.

summary that ranks all mRNAs predicted to be targeted by the selected miRNAs. In this later section, mRNAs are ordered by decreasing expected number of miRNA target sites under selective pressure, defined as the sum of all target site posterior probabilities for the selected miRNAs. The location of the putative target sites in the 3'-UTR is also indicated.

From the result screen, the user has the possibility to zoom onto a specific transcript to visualize the multiple genome alignments in the regions of the predicted target sites, and to find additional information about the targeted mRNAs from the Genbank database of the National Center for Biomedical Information (NCBI). Our web service also offers the possibility to run a Gene Ontology (GO) analysis searching for GO terms that are significantly over- or under-represented in the predicted miRNA targets through a modified version of the GeneMerge software (44). For instance, in the case of miR-142-5p, the most significantly enriched Biological Process GO term is 'regulation of transcription, DNA-dependent' (hypergeometric P -value $<10^{-10}$, after Bonferroni multiple testing correction), followed by two 'muscarinic acetylcholine receptor'-associated GO terms ($P < 10^{-10}$). The muscarinic acetylcholine receptor has been shown to be involved in autocrine control of cell proliferation, including the proliferation of immune cells (45). This type of analyses could thus provide experimental scientists with clues to the function of miR-142-5p in the naive CD4⁺ T cells.

The current release of EIMMo features miRNA target predictions for *H. sapiens*, *M. musculus*, *R. norvegicus*, *D. rerio*, *C. elegans* and *D. melanogaster*. Of these, the *M. musculus* and *R. norvegicus* predictions were not present in our initial publication (18). Furthermore, for the remaining organisms, the current predictions are based on the genome sequences of a larger set of species, because more fully sequenced genomes became available since 2007. We further based our predictions on the most recent mRNA sequences and 3'-UTR annotations provided by the RefSeq database (35). Concerning the microarray profiles that the user can use to guide miRNA target discovery in specific tissues and aside from the *H. sapiens* profiles that were used in our original EIMMo release (18), we incorporated similar mRNA expression profiles for *M. musculus* and *R. norvegicus*. Finally, the EIMMo web interface now informs the user about the smiRNAdb samples in which the selected miRNAs are most strongly expressed.

Experimental data

The miRNA sequences that were used for miRNA sample annotation and for miRNA target prediction were obtained from the miRBase release 12.0 (46). For the miRNA profiles, MirZ includes a total of 297 samples: 173 for *H. sapiens* (14), 88 for *M. musculus* (14), 16 for *R. norvegicus* (14), 10 for *D. melanogaster* (47), 9 for *D. rerio* (48) and 1 for *C. elegans* (49).

For miRNA target predictions, we used the most recent genome assemblies available at the University of California Santa Cruz (UCSC) (50): hg18 for *H. sapiens*, mm9 for *M. musculus*, rn4 for *R. norvegicus*, danRer5 for *D. rerio*, ce6 for *C. elegans* and dm3 for *D. melanogaster*. We further used the following UCSC genome assemblies in the pairwise genome alignments: panTro2, rheMac2, mm9, rn4, canFam2, monDom4, bosTau4 and galGal3 for *H. sapiens*; panTro2, rheMac2, hg18, rn4, canFam2, monDom4, bosTau4 and galGal3 for *M. musculus*; panTro2, rheMac2, hg18, mm9, canFam2, monDom4, bosTau3 and galGal3 for *R. norvegicus*; tetNig1, fr2 and oryLat2 for *D. rerio*; caeJap1, caePb2, caeRem3, cb3 and priPacl for *C. elegans*; and dp4, droAna3, droEre2, droGri2, droMoj3, droPer1, droSec1, droSim1, droVir3, droWill and droYak2 for *D. melanogaster*. mRNAs for all organisms were downloaded from the RefSeq database on 21 January 2009.

The links between sequence entities in various databases were made by mapping them all to the Gene database of NCBI (35). MiRNA expression profiles, microarray mRNA profiles and miRNA target predictions are stored as relational databases managed by a PostgreSQL server (www.postgresql.org).

CONCLUSION AND FUTURE DIRECTIONS

Using a concrete example comparing effector to naive CD4⁺ T-cells, we showed how MirZ can help isolating miRNAs that may be involved in a given biological function, and then provide clues into which molecular pathways may be controlled by these miRNAs to achieve their biological function. The integration of miRNA expression profiles with genome-wide miRNA target prediction combined with the tools we implemented—a Bayesian model for sample comparison, multivariate exploratory statistics, GO-term enrichment analysis—makes MirZ a powerful tool for studying miRNA-based regulation.

Since its publication, the miRNA expression atlas has been a valuable resource to the research community, and with the more general availability of deep sequencing technologies, more miRNA expression datasets are expected to emerge. Being able to explore and compare these datasets in a unified framework is highly desirable, and we plan to further support such analyses by updating MirZ as new datasets become available. Particularly for *D. melanogaster*, we currently only incorporate small-sized samples, and for *C. elegans* a whole-worm sample.

The target prediction methods also continue to evolve. In particular, additional determinants of miRNA targeting specificity must exist because not all transcripts that contain miRNA seed matches respond in a given experiment, but what these determinants are is still an open question (25,26). If a significantly better target prediction method emerges, this could be incorporated in our server.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge the contribution of Robin Vobruba and Philip Handschin from Fachhochschule Nordwestschweiz (Brugg, Switzerland) to the re-implementation of smiRNADB in Java; International Chicken Genome Sequencing Consortium (51) for the *Gallus gallus* genome; Chimpanzee Genome Sequencing Consortium for the *Pan troglodytes* genome; Baylor College of Medicine Human Genome Sequencing Center and the Rhesus Macaque Genome Sequencing Consortium for the *Macaca mulatta* genome; Mouse Genome Sequencing Consortium for the *M. musculus* genome (52); Rat Genome Project at the Baylor College of Medicine Human Genome Sequencing Center for the *R. norvegicus* genome (53,54); Dog Genome Sequencing Project for the *Canis familiaris* genome (55); Broad Institute for the *Monodelphis domestica* (opossum) genome; Baylor College of Medicine Human Genome Sequencing Center for the *Bos taurus* and *Drosophila pseudoobscura* genomes (56); Genoscope for the *Tetraodon nigroviridis* genome; Morishita laboratory for the *Oryzias latipes* (Medaka) genome (57); WormBase for the *C. elegans* genome (58); Agencourt Bioscience Corporation for the *D. ananassae* and *D. erecta* genomes; and the Drosophila 12 Genomes Consortium for all other *Drosophila* genomes (59). The *Takifugu rubripes* genome was provided freely by the Fugu Genome Consortium for use in this publication only. The *C. brenneri*, *C. briggsae*, *C. japonica*, *C. remanei* and *P. pacificus* genomes were produced by the Genome Sequencing Center at Washington University School of Medicine in St Louis and can be obtained from <ftp://genome.wustl.edu/pub/organism/Invertebrates/>.

FUNDING

Swiss National Fund (grant #3100A0-114001 to M.Z.); Swiss Institute of Bioinformatics. Funding for open access charge: Swiss Institute of Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

- Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
- Pasquinelli,A.E., Reinhart,B.J., Slack,F., Martindale,M.Q., Kuroda,M.I., Maller,B., Hayward,D.C., Ball,E.E., Degan,B., Miller,P. *et al.* (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, **408**, 86–89.
- Wightman,B., Ha,I. and 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.
- Brennecke,J., Hipfner,D.R., Stark,A., Russell,R.B. and Cohen,S.M. (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell*, **113**, 25–36.
- Giraldez,A.J., Cinalli,R.M., Glasner,M.E., Enright,A.J., Thomson,J.M., Baskerville,S., Hammond,S.M., Bartel,D.P. and Schier,A.F. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, **308**, 833–838.
- Giraldez,A.J., Mishima,Y., Rihel,J., Grocock,R.J., Van Dongen,S., Inoue,K., Enright,A.J. and Schier,A.F. (2006) Zebrafish MiR-430

- promotes deadenylation and clearance of maternal mRNAs. *Science*, **312**, 75–79.
7. Xu, P., Vernooy, S.Y., Guo, M. and Hay, B.A. (2003) The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.*, **13**, 790–795.
 8. Poy, M.N., Eliasson, L., Krützfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P. *et al.* (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature*, **432**, 226–230.
 9. Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.
 10. Zhao, Y., Samal, E. and Srivastava, D. (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, **436**, 214–220.
 11. Xiao, C., Calado, D.P., Galler, G., Thai, T.H., Patterson, H.C., Wang, J., Rajewsky, N., Bender, T.P. and Rajewsky, K. (2007) MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell*, **131**, 146–159.
 12. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
 13. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.
 14. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. *et al.* (2007) A mammalian microRNA expression atlas based on small rna library sequencing. *Cell*, **129**, 1401–1414.
 15. Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics*, **5**, 433–438.
 16. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
 17. Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
 18. Gaidatzis, D., van Nimwegen, E., Hausser, J. and Zavolan, M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
 19. Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92105.
 20. Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
 21. Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V. and Ding, Y. (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
 22. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
 23. Huang, J.C., Babak, T., Corson, T.W., Chua, G., Khan, S., Gallie, B.L., Hughes, T.R., Blencowe, B.J., Frey, B.J. and Morris, Q.D. (2007) Using expression profiling data to identify human microRNA targets. *Nat. Methods*, **4**, 1045–1049.
 24. Gennarino, V.A., Sardiello, M., Avellino, R., Meola, N., Maselli, V., Anand, S., Cuttillo, L., Ballabio, A. and Banfi, S. (2009) MicroRNA target prediction by expression analysis of host genes. *Genome Res.*, **19**, 481–490.
 25. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engle, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
 26. Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J. and Burge, C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
 27. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. and Cohen, S.M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, **123**, 1133–1146.
 28. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
 29. Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M. and Tuschl, T. (2008) Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA*, **14**, 2580–2596.
 30. Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
 31. Linsley, P.S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M.M., Bartz, S.R., Johnson, J.M., Cummins, J.M., Raymond, C.K., Dai, H. *et al.* (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol. Cell Biol.*, **27**, 2240–2252.
 32. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004) Brenda, the enzyme database: updates and major new developments. *Nucleic Acids Res.*, **32**, D431–D433.
 33. Bard, J., Rhee, S.Y. and Ashburner, M. (2005) An ontology for cell types. *Genome Biol.*, **6**, R21.
 34. Berninger, P., Gaidatzis, D., van Nimwegen, E. and Zavolan, M. (2008) Computational analysis of small RNA cloning data. *Methods*, **44**, 13–21.
 35. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
 36. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
 37. Lai, E.C. (2002) Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.*, **30**, 363–364.
 38. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
 39. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
 40. Rajewsky, N. and Succi, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.*, **267**, 529–535.
 41. Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
 42. Majoros, W.H. and Ohler, U. (2007) Spatial preferences of microRNA targets in 3' untranslated regions. *BMC Genomics*, **8**, 152.
 43. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J. and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nature Biotech.*, **26**, 578–583.
 44. Castillo-Davis, C.I. and Hartl, D.L. (2003) Genemerge – post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics*, **19**, 891–892.
 45. Eglén, R.M. (2006) Muscarinic receptor subtypes in neuronal and non-neuronal cholinergic function. *Auton. Autacoid Pharmacol.*, **26**, 219–233.
 46. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
 47. Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J. and Tuschl, T. (2003) The small RNA profile during Drosophila melanogaster development. *Dev. Cell*, **5**, 337–350.
 48. Chen, P.Y., Manninga, H., Slanchev, K., Chien, M., Russo, J.J., Ju, J., Sheridan, R., John, B., Marks, D.S., Gaidatzis, D. *et al.* (2005) The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.*, **19**, 1288–1293.
 49. Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H. and Bartel, D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, **127**, 1193–207.
 50. Karolchik, D., Kuhn, R.M., Baertsch, R., Barber, G.P., Clawson, H., Diekhans, M., Gardine, B., Harte, R.A., Hinrichs, A.S., Hsu, F. *et al.*

- (2008) The UCSC genome browser database: 2008 update. *Nucleic Acids Res.*, **36**, D773–D779.
51. International Chicken Genome Sequencing Consortium. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **423**, 7018.
52. Mouse Genome Sequencing Consortium. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
53. Havlak, P., Chen, R., Durbin, K.J., Egan, A., Yanru, R., Song, X.-Z., Weinstock, G.M. and Gibbs, R.A. (2004) The atlas genome assembly system. *Genome Res.*, **14**, 721–732.
54. Rat Genome Sequencing Project Consortium. (2004) Genome sequence of the brown norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
55. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J. 3rd, Zody, M.C. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
56. The Bovine Genome Sequencing and Analysis Consortium; Elsik, C.G., Tellam, R.L. and Worley, K.C. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324**, 522–528.
57. Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y. *et al.* (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **446**, 714–719.
58. Bieri, T., Blasiar, D., Ozersky, P., Antoshechkin, I., Bastiani, C., Canaran, P., Chan, J., Chen, N., Chen, W.J., Davis, P. *et al.* (2007) WormBase: new content and better access. *Nucleic Acids Res.*, **35**, D506–D510.
59. Drosophila 12 Genomes Consortium. (2007) Evolution of genes and genomes on the drosophila phylogeny. *Nature*, **450**, 203–218.