

Argonaute—a database for gene regulation by mammalian microRNAs

Priyanka Shahi^{1,2,3}, Serguei Loukianiouk¹, Andreas Bohne-Lang¹, Marc Kenzelmann², Stefan Küffer², Sabine Maertens², Roland Eils³, Herrmann-Josef Gröne², Norbert Gretz¹ and Benedikt Brors^{3,*}

¹Medical Research Center, University Hospital Mannheim, D-68167 Mannheim, Germany, ²Department of Cellular and Molecular Pathology, German Cancer Research Center, D-69120 Heidelberg, Germany and ³Department of Theoretical Bioinformatics, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

Received August 12, 2005; Revised and Accepted October 16, 2005

ABSTRACT

MicroRNAs (miRNAs) constitute a recently discovered class of small non-coding RNAs that regulate expression of target genes either by decreasing the stability of the target mRNA or by translational inhibition. They are involved in diverse processes, including cellular differentiation, proliferation and apoptosis. Recent evidence also suggests their importance for cancerogenesis. By far the most important model systems in cancer research are mammalian organisms. Thus, we decided to compile comprehensive information on mammalian miRNAs, their origin and regulated target genes in an exhaustive, curated database called Argonaute (<http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface>). Argonaute collects latest information from both literature and other databases. In contrast to current databases on miRNAs like miRBase::Sequences, NONCODE or RNadb, Argonaute hosts additional information on the origin of an miRNA, i.e. in which host gene it is encoded, its expression in different tissues and its known or proposed function, its potential target genes including Gene Ontology annotation, as well as miRNA families and proteins known to be involved in miRNA processing. Additionally, target genes are linked to an information retrieval system that provides comprehensive information from sequence databases and a simultaneous search of MEDLINE with all synonyms of a given gene. The web interface allows the user to get information for a single or multiple miRNAs, either selected or uploaded through a text file. Argonaute

currently has information on 839 miRNAs from human, mouse and rat.

INTRODUCTION

MicroRNAs (miRNAs) are small (21–23 nt) non-coding RNAs that are important regulators of gene expression. They act by base pairing with partially complementary sites in the mRNA of their targets, and either inhibit translation into protein or decrease the stability of the transcript (1,2). The number of currently known miRNAs in mammalian systems has risen dramatically in the last year, and predictions about their total number in humans amount up to 1000 (3). miRNAs are transcribed as long primary transcripts (pri-miRNAs), some of them being polycistronic, which are processed in the cell nucleus by an enzyme called Drosha, yielding precursor miRNAs (pre-miRNAs) that exhibit a characteristic stem-loop sequence. These are exported into the cytosol where mature miRNAs are generated by the RNase III-type enzyme Dicer, producing a small double-stranded RNA from which one strand (called miRNA*) is quickly degraded, releasing the small single-stranded miRNA (4). Translational inhibition, which seems to be the major mode of action in animals, is performed by a riboprotein complex called RNA-induced silencing complex (RISC) consisting of the miRNA and proteins of the argonaute family (5,6).

miRNAs are involved in several cellular processes, including cellular differentiation (7,8), organism development (9,10) and apoptosis (11,12). While all of these are conserved in metazoans, the number of conserved miRNAs between mammals suggests that there are additional functions only found in vertebrates (13), e.g. controlling hematopoietic differentiation (14). Recent studies provide growing evidence for the involvement of miRNAs in cancerogenesis (15–18).

*To whom correspondence should be addressed. Tel: +49 6221 42 3614; Fax: +49 6221 42 3620; Email: b.brors@dkfz.de

The database Argonaute was created against this background. Argonaute comprises almost all miRNAs now publicly available from human, mouse and rat. The advantages of Argonaute over current databases on miRNAs are as follows: it has information on (i) origin of a miRNA, i.e. in which host gene it is encoded, (ii) its expression in different tissues and its known or proposed function, (iii) its potential target genes including Gene Ontology annotation as well as (iv) on miRNA families and proteins known to be involved in miRNA processing. The first release of Argonaute contains 839 miRNAs from human, mouse and rat.

In summary, Argonaute is a comprehensive database for mammalian miRNAs collecting all available latest information about miRNA expression, pre-miRNA sequence and length of the stem-loop region, function and potential target genes, and is thus providing a valuable platform for quick and efficient access to all important information on the fast growing field of miRNA research. Through a user-friendly web interface at <http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface>, access is free for all users.

DATABASE IMPLEMENTATION AND DESIGN

As main sources for miRNAs and corresponding annotation, we used the miRBase sequence database (19). We further checked NONCODE (20) and RNAdb (21) for miRNAs that were not contained in miRBase::sequences like hsa-mir-350 and hsa-mir-196-4. These have been reported in literature (22), but the only information given there is that they are contained within intron 6 of ENSG00000143702. Owing to the lack of other information, we also decided not to include them in Argonaute.

For individual miRNAs with direct support in the literature, the details (sequence, accession number, chromosomal location, etc.) were manually curated where possible. Additional annotation, especially on confirmed or predicted targets of miRNAs and on tissue specificity of expression, was entered manually based on primary literature. Literature scanning is done on a regular basis using My NCBI (cubby) in order to keep the contents of the database up-to-date. A list of studies used for the current version of Argonaute is reported in Supplementary Data (Supplementary Table 1).



ARGONAUTE-a database for gene regulation by mammalian micro-RNAs

microRNA is a key element of RNAi machinery that triggers chromosomal modification, post-transcriptional gene silencing and protein translational blockade depending on the source, the RNA and the nature of the interaction with the target nucleic acid.

The result is divided into two tables: first for basic miRNA information and second for target information (when corresponding target options are selected) which can be found by scrolling down.

RESULT FILE FOR miRNA

Name	Accession Number	Chromosome	Strand	Mature Length	Stemloop Length	Mature Sequence	Origin	Organism	Function	Reference
hsa-mir-328	MI0000804	16	-	22	75	cuggccucucugccuuccgu	Within Intron 12 of ENSG00000102890 on opposite strand	Homo sapiens	---	Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Kim

Detail about Origin of miRNA

Name of miRNA	Origin	Gene_ID	Gene_symbol	Detail	Reference
hsa-mir-328	Within Intron 12 of ENSG00000102890 on opposite strand	Q96BJ8	----	ENGULFMENT AND CELL MOTILITY PROTEIN 3. ELMO3 [Q96BJ8]	Identification of mammalian microRNA host genes and transcription units. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Genome Res. 14(10A):1902-10.(2004). PMID: 15364901. Ref

RESULT FILE

Name	Target Gene Symbol	Reference
hsa-mir-328	HOXB8	Human microRNA Targets. John B, Enright AJ, Aravin A, Tuschl T, Sander C, et al. PLoS Biology. 2(11):1862-1879(2004). PMID: 15502875. PubMed (15502875)

Proteins involved in Processing of miRNA

Protein	Motifs	Function	Organism	Reference
Dicer	RNase III, DEAD, PAZ, dsRBD	miRNA precursor processing	H. sapiens	Ref
DCR-1	RNase III, DEAD, PAZ, dsRBD	miRNA precursor processing	C. elegans	Ref
DCR-1	RNase III, HELICc, PAZ, dsRBD	miRNA precursor processing	D. melanogaster	Ref
DC1	RNase III, DEAD,	miRNA precursor	A. thaliana	Ref

Figure 1. Result of query for hsa-mir-328. The upper table provides comprehensive information on the mature miRNA and its precursor, the middle table has information on predicted targets of hsa-mir-328. The inset shows additional information on the origin of this miRNA, i.e. the gene it is encoded in. The inset view is linked to the main view by clicking on the 'origin' link. The bottom table displays information on miRNA processing enzymes.

Data are stored in a relational database based on the MySQL database management system. Supplementary Figure 1 shows an UML diagram of the tables and their relations. Information on miRNAs and their targets is contained in different objects in order to be able to map the many-to-many relationships correctly. For each miRNA–target relation, individual reference to the literature is reported which is kept in a separate table to minimize redundancy.

DATABASE ACCESS AND INTERFACE

Database access is via a web interface based on PHP scripts. Possible queries include exact or wildcard search for an miRNA name, batch retrieval based on a text file with miRNA names and browsing the entire contents of the database. The names of miRNAs comply with miRBase::sequences where appropriate. A selection panel allows to tailor information retrieval by selecting or deselecting fields to be displayed in the output. A comprehensive set of fields is selected by default.

Results of queries are presented in two different tables (Figure 1), one on the miRNAs and one on the targets (provided that target information has been selected for

display). If the box ‘Annotation’ for targets is unchecked, all predicted targets for a given miRNA are presented in one row of the result table, together with all references. If it is checked, there is one row per target gene, including the reference that reported the miRNA–target relationship, so it can be tracked where the evidence on the relationship has come from.

References are linked to PubMed, thus their abstracts can be retrieved immediately. Target gene identifiers are linked to ENSEMBL to get information on the transcript assumed to be regulated by a certain miRNA. The result field ‘Origin’ is linked to a third table (Figure 1, inset) providing comprehensive results on the gene the miRNA is contained in (many miRNAs are hosted in introns of protein-coding genes). Additionally, a detailed list of proteins involved in miRNA processing can also be obtained from the homepage.

TARGET GENE INFORMATION RETRIEVAL SYSTEM

In order to facilitate the characterization of target genes we developed an information retrieval system mainly based on several NCBI sources as well as Unigene, Swiss-Prot and

Affymetrix Probe Set ID	1452493_s_at
Gene Title	homeo box B8
Gene Symbol	Hoxb8
Locus Link	15416
Unigene	Mm.196037 // full length
Chip	Mouse430A_2
Organism	Mus musculus
Chromosome Location	11 56.0 cM
Chromosome Location	11 56.0 cM(Mus musculus); 10q31(Rattus norvegicus)
Synonymous	hoxb8; homeo box b8; hox-2.4
All Names	hoxb8; homeo box b8; hox2r1a; homeobox gene b8; hox-2.4; hox2; hox2d; homeo box 2d; homeobox protein hox-2.4
Entrez	Link
Pubmed/Gene Name	Link
Pubmed/Transgene/Knockout	Link
Locus Link	Link
Unigene	Link
Swissprot	Link
Ensembl	Link
Summary	---
Pubmed/Citations	Link
KEGG	---
Genmapp	---
OMIM	Link
miRNA	Link
Affymetrix Probe Set ID	2212
Gene Title	homeo box B8
Gene Symbol	HOXB8
Locus Link	3218

Figure 2. Result of query for the potential target gene *HOXB8*. The table provides comprehensive information on the gene, including Affymetrix® Probe Set IDs, cross-references to other databases, synonyms and textual summary information. The links to PubMed search by all synonyms of the gene, either alone (inset) or in combination with ‘(transgen* OR knockout OR knock-out) AND (animals[mh] OR animal [mh])’ to find reports on genetically modified organisms where this gene is altered or deleted.

Ensembl. This allows us to perform a comprehensive search for each target gene by clicking on the gene name (Figure 2). The major advantage of our interface is that it generates an automated PubMed search using all synonyms of the individual gene (Figure 2, inset). This search can be tailored to an organ-specific search or a search for transgenes/knock-outs. Furthermore, information on the availability of the target gene on Affymetrix® chips is given. Also a search for pathways is provided in which the target gene is active.

The system is implemented as a PHP-based web interface to an MySQL database. It is capable of generating automatic search strings for the target gene in PubMed using several criteria. Information in the database is synchronized with NCBI sources on a regular basis. It is possible to save the results of the search as an Excel table for later use. Our interface may also be used as a stand-alone query tool: <http://www.ma.uni-heidelberg.de/inst/zmf/affymetrix/evaluation/annotation/ver2/>. Furthermore, the system is able to process a list of genes, which is very convenient when analyzing results of microarray experiments or any experiments yielding a group of related genes.

CURRENT STATUS AND FUTURE DEVELOPMENTS

Currently, Argonaute hosts information on 839 mammalian miRNAs from human, mouse and rat, and reports 312 miRNA–target relationships. We continue to add further information as reported in primary literature and synchronize with miRBase::Sequence DB (microRNA registry) every 4 months. We are currently developing an interface for sequence-based search that allows to find potential target sites for miRNAs in the query sequence based on established miRNA target-prediction algorithms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was funded by the Deutsche Forschungsgemeinschaft through Graduiertenkolleg 886 and by the Federal Ministry of Research and Education through grant 01 GR 0450. Funding to pay the Open Access publication charges for this article was provided by the DFG through Graduiertenkolleg 886.

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

2. He,L. and Hannon,G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.*, **5**, 522–531.
3. Berezikov,E., Guryev,V., van de,B.J., Wienholds,E., Plasterk,R.H. and Cuppen,E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, **120**, 21–24.
4. Cullen,B.R. (2004) Transcription and processing of human microRNA precursors. *Mol. Cell*, **16**, 861–865.
5. Tijsterman,M. and Plasterk,R.H. (2004) Dicers at RISC; the mechanism of RNAi. *Cell*, **117**, 1–3.
6. Tang,G. (2005) siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.*, **30**, 106–114.
7. Johnston,R.J. and Hobert,O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*, **426**, 845–849.
8. Chang,S., Johnston,R.J.Jr, Frokjaer-Jensen,C., Lockery,S. and Hobert,O. (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature*, **430**, 785–789.
9. Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The *C.elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.
10. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
11. 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.
12. 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.
13. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
14. Chen,C.Z., Li,L., Lodish,H.F. and Bartel,D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83–86.
15. Gregory,R.I. and Shiekhattar,R. (2005) MicroRNA biogenesis and cancer. *Cancer Res.*, **65**, 3509–3512.
16. McManus,M.T. (2003) MicroRNAs and cancer. *Semin.Cancer Biol.*, **13**, 253–258.
17. 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.
18. 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.
19. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
20. Liu,C., Bai,B., Skogerbo,G., Cai,L., Deng,W., Zhang,Y., Bu,D., Zhao,Y. and Chen,R. (2005) NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.*, **33**, D112–D115.
21. Pang,K.C., Stephen,S., Engstrom,P.G., Tajul-Arifin,K., Chen,W., Wahlestedt,C., Lenhard,B., Hayashizaki,Y. and Mattick,J.S. (2005) RNADB—a comprehensive mammalian noncoding RNA database. *Nucleic Acids Res.*, **33**, D125–D130.
22. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. and Bradley,A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.