Nucleic Acids Research, 2006, Vol. 34, Database issue **D135–D139**
doi:10.1093/nar/gkj135

miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes

Paul W.C. Hsu[1], Hsien-Da Huang[1,2,*], Sheng-Da Hsu[1], Li-Zen Lin[1], Ann-Ping Tsou[3], Ching-Ping Tseng[2,4], Peter F. Stadler[5], Stefan Washietl[6] and Ivo L. Hofacker[6]

[1]Institute of Bioinformatics and [2]Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu 300, Taiwan, ROC, [3]Institute of Biotechnology in Medicine, National Yang-Ming University, Taipei 112, Taiwan, ROC, [4]Institute of Biochemical Engineering, National Chiao Tung University, Hsin-Chu 300, Taiwan, ROC, [5]Department of Computer Science, Interdisciplinary Center of Bioinformatics, University of Leipzig, Germany and [6]Institute of Theoretical Chemistry, University of Vienna, Austria

Received August 15, 2005; Revised October 17, 2005; Accepted October 27, 2005

ABSTRACT

Recent work has demonstrated that microRNAs (miRNAs) are involved in critical biological processes by suppressing the translation of coding genes. This work develops an integrated database, miRNAMap, to store the known miRNA genes, the putative miRNA genes, the known miRNA targets and the putative miRNA targets. The known miRNA genes in four mammalian genomes such as human, mouse, rat and dog are obtained from miRBase, and experimentally validated miRNA targets are identified in a survey of the literature. Putative miRNA precursors were identified by RNAz, which is a non-coding RNA prediction tool based on comparative sequence analysis. The mature miRNA of the putative miRNA genes is accurately determined using a machine learning approach, mmiRNA. Then, miRanda was applied to predict the miRNA targets within the conserved regions in 3′-UTR of the genes in the four mammalian genomes. The miRNAMap also provides the expression profiles of the known miRNAs, cross-species comparisons, gene annotations and cross-links to other biological databases. Both textual and graphical web interface are provided to facilitate the retrieval of data from the miRNAMap. The database is freely available at http://mirnamap.mbc.nctu.edu.tw/.

INTRODUCTION

MicroRNAs (miRNAs) are small RNA molecules, which are ∼22 nt sequences that have an important role in the translational regulation and degradation of mRNA by the base's pairing to the 3′-untranslated regions (3′-UTR) of the mRNAs. The miRNAs are derived from the precursor transcripts of ∼70–120 nt sequences, which fold to form as stem–loop structures, which are thought to be highly conserved in the evolution of genomes. Previous analyses have suggested that ∼1% of all human genes are miRNA genes, which regulate the production of protein for 10% or more of all human coding genes (1).

Ambros *et al*. developed a uniform system for the identification and annotation of new microRNA in a range of organisms (2). The miRBase (3) supports the sequence and annotation of the published miRNA genes. Washietl *et al*. developed a fast and reliable method for identifying the non-coding RNA structures using comparative sequence analysis (4), which can be incorporated into the proposed resource to identify the conserved miRNA precursors in the four mammalian genomes. TargetScan (5), miRanda (1) and RNAhybrid (6) are three previously developed tools for determining the energetically most favorable hybridization sites of a small RNA to a large RNA. PicTar (7) is a computational method for identifying common targets of known miRNAs. Lu *et al*. developed an miRNA microarray to measure the expression profiles of all known miRNA in various normal tissues and tumors (8).

The investigation of the roles of miRNAs in various research fields depends on having plentiful miRNAs and information about them. This work develops an integrated database, miRNAMap, of miRNA genes, miRNA targets and the relationships between the miRNAs and the miRNA targets. The known miRNA genes in four mammals—humans, mice, rats and dogs—are obtained from miRBase (3). Experimentally validated miRNA targets are identified by surveying the literature. The putative miRNA precursors conserved in the

*To whom correspondence should be addressed: Tel: +886-3-5712121 Ext. 56952; Email: bryan@mail.nctu.edu.tw

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

Table 1. Statistics concerning the miRNA genes in the miRNAMap

	Human		Mouse		Rat		Dog		Total	
	Known	Putative	Known	Putative	Known	Putative	Known	Putative	Known	Putative
miRBase (release 6.0) (3)	131	117	196	54	117	86	6	0	450	257
Putative miRNA genes		464		395		274		336		1469
Subtotal	131	581	196	449	117	360	6	336	450	1726

The amounts of the overlapping miRNAs between the putative miRNAs and miRBase known miRNAs in human, mouse, rat and dog are 122, 97, 44 and 1, respectively.

Table 2. Statistics concerning the known human miRNAs targets detected by miRanda considering different parameters

miRanda MFE (kcal/mol)	miRanda score			
	⩾120	⩾140	⩾160	⩾180
⩽−12	27	23	9	3
⩽−15	21	18	9	3
⩽−20	17	15	9	3
⩽−25	3	3	3	3

Table 3. miRNA genes categorized by genomic location

Genomic locations	Human	Mouse	Rat	Dog	Total
Intergenic	396 (52%)	405 (62%)	359 (76%)	225 (64%)	1385 (62%)
Intronic	308 (40%)	235 (36%)	112 (24%)	107 (31%)	762 (34%)
Exonic	58 (8%)	12 (2%)	0 (0%)	18 (5%)	88 (4%)
Subtotal	762 (100%)	652 (100%)	471 (100%)	350 (100%)	2235 100%)

four mammalian genomes are identified using RNAz (4), which is a tool for predicting non-coding structural RNAs based on comparative sequence analysis. A machine learning method was developed to locate accurately the positions of the mature miRNAs to determine the mature miRNAs that were located at the putative miRNA precursors. miRanda (1) was applied to identify energetically the most probable miRNA targets of the miRNAs against the conserved regions in 3′-UTR of the genes in the four genomes, and thereby predict the target genes of the known miRNAs and the putative miRNAs. Finally, the relationships between the miRNAs and the coding genes are elucidated based on the miRNA targets. The miRNAMap also yields the expression profiles of known miRNAs, cross-species comparisons, gene annotations and cross-links to other biological databases. Both textual and web interface are provided to facilitate the retrieval of data curated in the miRNAMap.

The main contribution of this work is the computational establishment of mapping between the miRNA genes and the miRNA target genes, and the provision of effective annotations, including miRNA expression, cross-species comparisons and the annotations of the miRNA target genes. A comparative sequence analysis yields a series of the putative miRNA precursors and the mature miRNA conserved in human, mouse, rat and dog genomes. Additionally, a variety of search functions and a graphical interface are designed and implemented to help researchers to investigate the microRNA roles in cell regulations.

DATABASE STATISTICS

The miRNAMap presently stores 2176 miRNAs in four mammalian genomes—human, mouse, rat and dog—as presented in Table 1. The number of experimentally validated (known) miRNA genes in humans, mice, rats and dogs obtained from the miRBase (Release 6.0) (3) are 131, 196, 117 and 6, respectively. The number of putative miRNA genes in humans, mice and rats from the miRBase are 117, 54 and 86, respectively. The number of putative miRNA genes in humans, mice, rats and dogs identified in this database are 464, 395, 274 and 336, respectively. By sequence comparison between the putative miRNAs in the proposed resource and known miRNAs from miRBase (3), the amounts of putative miRNAs identical to known miRNAs in human, mouse, rat and dog genomes are 122, 97, 44, and 1, respectively.

The minimum free energy (MFE) of the miRNA–target duplex was determined while predicting the miRNA target sites. The lower MFE values of the miRNAs and the target sites reveal the energetically more probable hybridizations between the miRNAs and the target genes. Table 2 presents statistics concerning the miRanda predicted known human miRNA targets, which are extracted from literature (Table S3). For instance, 27 known miRNA targets can be identified when the miRanda MFE threshold and miRanda score threshold are set to −12 kcal/mol and 120, respectively. However, these parameters are likely to grossly overpredict the number of miRNA targets per gene. Alternatively, the proposed database allows users to consider a set of parameters that is more stringent and gives less likely false positives. All the predictive miRNA/targets are determined and stored in the proposed miRNAMap if the MFE of miRNA/targets is <−12 kcal/mol and the miRanda (1) score of miRNA/targets exceeds 120. For instance, when the MFE threshold is set to −20 kcal/mol and the miRanda score threshold is set to 120, the average number of target genes of each miRNA in humans, mice, rats and dogs are 1419, 1390, 293 and 760, respectively (Table S4). The average number of distinct miRNAs that target each gene in humans, mice, rats and dogs are 131, 116, 80 and 15, respectively. Similarly, when the miRanda (1) score threshold is set to 160 and the miRanda MFE threshold is set to −12 kcal/mol, the average number of the target genes for each miRNA in human, mouse, rat and dog genomes are 183, 175, 40 and 273, respectively. The average number of the distinct miRNAs targeting each gene in human, mouse, rat and dog genomes are 16, 14, 10 and 6, respectively. Additionally, all of the miRNA genes were observed with reference to the annotations of the genes in the four genomes: e.g. ∼52, 40 and 8% of the human miRNA genes were present in the intergenic, intronic and exonic regions, respectively (Table 3). In summary, of all the miRNAs in the proposed

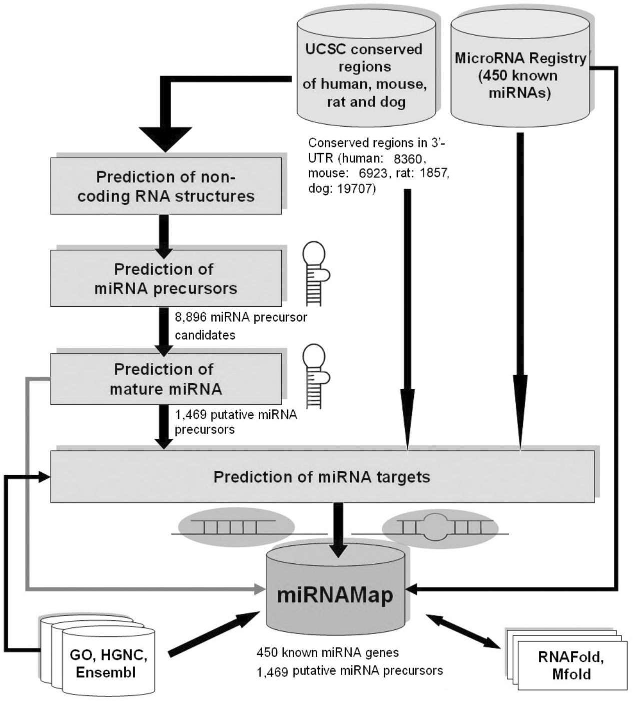

UCSC conserved regions of human, mouse, rat and dog
MicroRNA Registry (450 known miRNAs)
Conserved regions in 3'-UTR (human: 8360, mouse: 6923, rat: 1857, dog: 19707)
Prediction of non-coding RNA structures
Prediction of miRNA precursors
8,896 miRNA precursor candidates
Prediction of mature miRNA
1,469 putative miRNA precursors
Prediction of miRNA targets
miRNAMap
GO, HGNC, Ensembl
450 known miRNA genes
1,469 putative miRNA precursors
RNAFold, Mfold

Figure 1. Generation of data of the miRNAMap resource.

database, ∼62, 34 and 4% were present in the intergenic, intronic and exonic regions, respectively.

DATA GENERATION

Figure 1 presents a flow chart of the generation of data of the miRNAMap database. It comprises the following three main steps: (i) integration of the known miRNA precursors and the miRNA targets; (ii) identification of the miRNA precursors and the mature miRNAs and (iii) prediction of the miRNA targets. Each step is described in detail later.

Integration of known miRNA precursors and miRNA targets

The experimentally validated miRNA genes in the four mammalian genomes—human, mouse, rat and dog—are obtained from the miRBase (release 6.0) (3). Table 1 indicates that 707 miRNAs were extracted from the miRBase. Only 29 experimentally validated miRNA targets, including the miRNAs and their target genes, are taken from the literature (Table S3).

Prediction of miRNA precursors and mature miRNAs

The known miRNA genes were obtained from miRBase (3), and putative miRNA precursors were identified by RNAz (4), along with the four genomes. RNAz is a computational tool for predicting non-coding RNA structures based on comparative sequence analysis. All of the annotated coding exons were removed from the conserved regions obtained from the UCSC Genome Browser (9). A total of 438 788 alignments of the non-coding regions that are conserved in the four mammals—human, mouse, rat and dog—were retained. This number represents 82.64 MBase or 2.88% of the human genome. Then, RNAz was used to detect the non-coding RNA structures against the sequences of the conserved regions above. The following constraints were imposed to filter the non-coding RNA structures of the putative miRNA precursors; (i) the height of the stem of the non-coding RNA structure exceeds 20 nt; (ii) the RNAz *z*-score of the structure is less than −3.5. Hence, 2681, 2231, 1751 and 2233 putative miRNA precursors were present in human, mouse, rat and dog, respectively.

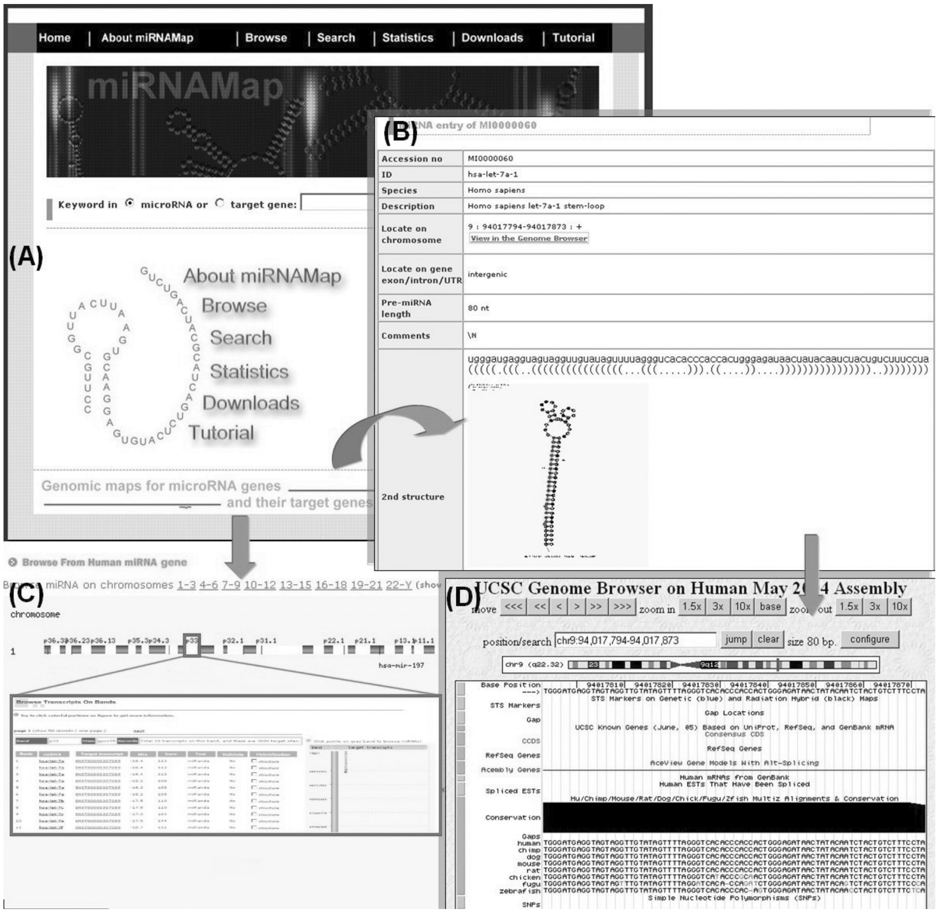
After the putative miRNA precursors had been identified, the mature miRNA of the putative miRNA genes were identified. Therefore, a machine learning method was proposed to determine computationally the mature miRNA in the miRNA precursors. Figure S3 presents the proposed algorithm (Supplementary Data). The detail of the process undergone by the mature miRNA detector is described in the Supplementary Data. In humans, mice, rats and dogs, 464, 395, 274 and 336 putative miRNAs, respectively, are found to contain the putative mature miRNAs. A comparison of sequences of the miRNA precursors demonstrates that 153 of the putative human miRNA precursors are found to be highly homologous to the known miRNA obtained from the miRBase (3).

Prediction of miRNA targets

miRanda (1) was incorporated into the proposed miRNAMap to identify the miRNA targets, which were the most probable miRNA targets of the miRNAs against the conserved regions in 3′-UTR of the genes of the four genomes and thus predict the target genes of the known miRNAs and the putative miRNAs. The conserved regions in 3′-UTR of the genes are derived from the UCSC conserved regions, which are at least 200 bp long and have sequence identities >60% (9); each transcript is effectively annotated in the Ensembl database (10). From human, mouse, rat and dog genomes, 8360, 6923, 1857 and 19 707 3′-UTR conserved regions are extracted, respectively. Reference is made to the literature that records the known miRNA targets (Table S3).

The predictive parameters including miRanda MFE and miRanda score are adjusted for the miRNA target prediction by comparing the predictive results with known miRNA/targets data. It resulted that 27 known miRNA/targets among 29 known miRNA/targets were successfully identified (Table S3). The MFE threshold of the miRNA and target duplex was specified as −12 kcal/mol and the miRanda score was specified as 120. Therefore, the miRNA targets whose MFEs are <−12 kcal/mol and the score exceeds 120 are identified and compiled in the miRNAMap database.

The relationships between the miRNAs and the target genes are determined with reference to the genomic locations and the annotations of the coding genes. The miRNAMap also yields expression profiles of the known miRNAs, cross-species comparisons, gene annotations and various cross-links to other biological databases. The annotations of the coding genes were obtained from the Ensembl database (10), Gene Ontology (11) and HGNC gene grouping/family data (12). The conserved regions among the four selected genomes in the database are obtained from the UCSC Genome Browser (9). The reader should refer to Table S1 in the Supplementary Data for the integrated databases.


Home | About miRNAMap | Browse | Search | Statistics | Downloads | Tutorial
miRNAMap
Keyword in ● microRNA or ○ target gene:
(A)
About miRNAMap
Browse
Search
Statistics
Downloads
Tutorial
Genomic maps for microRNA genes
and their target genes
(B)
(C)
Browse From Human miRNA gene
miRNA on chromosomes 1-3 4-6 7-9 10-12 13-15 16-18 19-21 22-Y
chromosome
Browse Transcripts On Bands
(D)
UCSC Genome Browser on Human May 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr9:94,017,794-94,017,873 jump clear size 80 bp. configure
chr9 (q22.32)
Base Position
STS Markers
Gap
CCDS
RefSeq Genes
Acembly Genes
Spliced ESTs
Conservation
Gaps
human
chimp
dog
mouse
rat
chicken
fugu
zebrafish
SNPs
STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
Gap Locations
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
Consensus CDS
RefSeq Genes
AceView Gene Models With Alt-Splicing
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
Hu/Chimp/Mouse/Rat/Dog/Chick/Fugu/Zfish Multiz Alignments & Conservation
Simple Nucleotide Polymorphisms (SNPs)

miRNA entry of MI0000060

Accession no	MI0000060
ID	hsa-let-7a-1
Species	Homo sapiens
Description	Homo sapiens let-7a-1 stem-loop
Locate on chromosome	9 : 94017794-94017873 : + View in the Genome Browser
Locate on gene exon/intron/UTR	intergenic
Pre-miRNA length	80 nt
Comments	\N
2nd structure	ugggaugagguaguagguuguauaguuuuagggucacacccaccacugggagauaacuauacaaucuacugucuuuccua (((((.(((..((((((((((((((((...(((.....))).((....))....))))))))))))))))..))))))))

Figure 2. Interface of the miRNA genes in miRNAMap.

The expression profiles of the miRNAs are helpful in elucidating the regulatory roles of the miRNAs. Lu *et al*. constructed the miRNA microarray to elucidate the gene expression profiles in various normal tissue and tumors (8). The gene expression data for known miRNAs were integrated into the miRNAMap to determine the tissue-specificity of the known miRNAs.

INTERFACE

Various query interfaces and graphical visualization pages were implemented to facilitate access to data and further analyses to support research on miRNA. The miRNAMap provides two modes for browsing the miRNA information—the miRNA genes browser and the miRNA target browser. As presented in Figure 2, the chromosomal view facilitates the browsing of the miRNA genes. The miRNA gene page shows the sequence, mature miRNA, genomic location, relevant citations from the literature, gene annotations, tissue-specificity and conserved sequences obtained from the UCSC Genome Browser (9). In particular, the diagram revealing the stem–loop RNA structure of the miRNA precursor, which is folded using RNAfold (13), is generated graphically by mfold (14).

Figure S1 (Supplementary Data) presents a graphical visualization tool that is used to present the miRNAs that can target a gene transcript. All the miRNA target sites are also provided in text format: the text states the genomic locations of the target sites, the MFE of the miRNAs, the sequences of the target sites and the alignment of the hybridization structures. Users can change the MFE threshold and score threshold to filter the miRNA targets of interest.

The miRNAMap provides various search criteria for accessing the data, including keyword, accession of the miRNAs, chromosomal locations and names of miRNA target genes. In particular, searching the database by submitting a group of genes helps users elucidate the roles of miRNAs in various gene groupings and to explicate the cooperative or combinatorial control of gene expression by a group of miRNAs (15). Users can also submit gene groups or families using HGNC annotations (12). The tutorial documentation on the miRNAMap website describes in detail the usage of miRNAMap web interfaces.

CONCLUSIONS

An integrated database for miRNAs, miRNAMap, was established to compile the known miRNAs, the putative miRNAs, the miRNA targets and the regulatory relationships between the miRNAs and the coding genes in humans, mice, rats and dogs. The authors hope that this database can provide sufficient information to support any miRNAs-related works. For example, miRNAs might contribute to cancer, miRNA-mediated tumorigenesis results from either down-regulation of tumor suppressor genes or up-regulation of oncogenes (16). Moreover, human miRNA genes are frequently located at fragile sites and genomic regions that are involved in cancer (17). The miRNAMap developed herein can support research in this area by combining the data obtained by comparative genomic hybridization (CGH) experiments, which provide information of the gain and loss chromosomal regions in tumor genomes. A user who is interested in particular extraordinary chromosomes can use miRNAMap to obtain miRNA information by querying the cytobands. Integrated analyses of the CGH data, the expression profiles of the coding genes and the miRNAMap data demonstrate that the miRNA genes and the miRNA targets potentially participate in the regulation of cancer cells and can be systematically identified for further experimental verification.

The database will be developed as follows. (i) It will be made to support miRNA annotations including miRNA genes and miRNA targets for other species and (ii) the miRNAMap data will be further analysed to support the miRNAs involved in the combinatorial control of the gene expression of the coding genes.

AVAILABILITY

The miRNAMap database will be continuously maintained and updated. The database is now freely available at http://miRNAMap.mbc.nctu.edu.tw.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 94-2213-E-009-025. Funding to pay the Open Access publication charges for this article was provided by National Science Council of the Republic of China.

Conflict of interest statement. None declared.

REFERENCES

1. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol*., **2**, e363.
2. Ambros,V., Bartel,B., Bartel,D.P., Burge,C.B., Carrington,J.C., Chen,X., Dreyfuss,G., Eddy,S.R., Griffiths-Jones,S., Marshall,M. *et al*. (2003) A uniform system for microRNA annotation. *RNA*, **9**, 277–279.
3. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res*., **32**, D109–D111.
4. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
5. 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.
6. Rehmsmeier,M., Steffen,P., Hochsmann,M. and Giegerich,R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
7. 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.
8. 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.
9. Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. *et al*. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res*., **31**, 51–54.
10. Hubbard,T., Andrews,D., Caccamo,M., Cameron,G., Chen,Y., Clamp,M., Clarke,L., Coates,G., Cox,T., Cunningham,F. *et al*. (2005) Ensembl 2005. *Nucleic Acids Res*., **33**, D447–D453.
11. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al*. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*., **32**, D258–D261.
12. Cotton,R.G., McKusick,V. and Scriver,C.R. (1998) The HUGO Mutation Database Initiative. *Science*, **279**, 10–11.
13. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res*., **31**, 3429–3431.
14. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*., **31**, 3406–3415.
15. 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.
16. Caldas,C. and Brenton,J.D. (2005) Sizing up miRNAs as cancer genes. *Nat. Med*., **11**, 712–714.
17. Calin,G.A., Sevignani,C., Dumitru,C.D., Hyslop,T., Noch,E., Yendamuri,S., Shimizu,M., Rattan,S., Bullrich,F., Negrini,M. *et al*. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA*, **101**, 2999–3004.