

mESAdb: microRNA Expression and Sequence Analysis Database

Koray D. Kaya¹, Gökhan Karakulah², Cengiz M. Yakıcıer^{1,3}, Aybar C. Acar^{4,*} and Özlen Konu^{1,5,*}

¹Department of Molecular Biology and Genetics, Bilkent University, 06800 Ankara, Turkey, ²Department of Medical Informatics, Health Sciences Institute, Dokuz Eylül University, 35340 Inciralti, Izmir, ³Department of Medical Biology, Acibadem University, 34848 Maltepe, Istanbul, ⁴Department of Computer Engineering and ⁵BilGen, Bilkent University Genetics and Biotechnology Research Center, Bilkent University, 06800 Ankara, Turkey

Received October 1, 2010; Revised November 19, 2010; Accepted November 20, 2010

ABSTRACT

microRNA expression and sequence analysis database (<http://konulab.fen.bilkent.edu.tr/mirna/>) (**mESAdb**) is a regularly updated database for the multivariate analysis of sequences and expression of microRNAs from multiple taxa. mESAdb is modular and has a user interface implemented in PHP and JavaScript and coupled with statistical analysis and visualization packages written for the R language. The database primarily comprises mature microRNA sequences and their target data, along with selected human, mouse and zebrafish expression data sets. mESAdb analysis modules allow (i) mining of microRNA expression data sets for subsets of microRNAs selected manually or by motif; (ii) pair-wise multivariate analysis of expression data sets within and between taxa; and (iii) association of microRNA subsets with annotation databases, HUGE Navigator, KEGG and GO. The use of existing and customized R packages facilitates future addition of data sets and analysis tools. Furthermore, the ability to upload and analyze user-specified data sets makes mESAdb an interactive and expandable analysis tool for microRNA sequence and expression data.

INTRODUCTION

microRNAs are small (19–22 nt) RNAs that play crucial roles in many cellular processes via targeting mRNAs for translational repression or cleavage thus regulating gene expression (1). microRNAs, through their compatible 5'-seed sequences, exert regulatory functions primarily on the 3'-untranslated regions (UTRs) of targeted

mRNAs (2–4). microRNAs that target the same mRNAs may share common motifs due to duplication events and/or common evolutionary ancestry (5–7). Previous studies using genome search and target prediction algorithms have provided lists of common genomic regulatory nucleotide motifs, some of which are also shared by microRNA sequences (8). However, whether microRNAs that are similar in sequence exhibit similarities also in function and/or expression is not yet well understood and warrants further study.

Based on large-scale studies, microRNAs have been annotated for their specificity for particular tissues, developmental stages and/or pathologies such as cancer (9–12). For example, Bargaje *et al.* (13) compiled and normalized multiple data sets from different sources to determine the tissue-specific and tissue-invariant consensus expression profiles. Others have surveyed microRNA expression profiles in large numbers of normal and cancerous tissues to decipher microRNA networks and conserved expression clusters in disease (14). There also is evidence suggesting that expression patterns of microRNAs are conserved at the species level (5). Nevertheless, the conserved associations between microRNA sequence motifs and expression profiles across taxa still remain relatively unexplored (15–17). Similarly, there is a growing need for tools developed for multivariate comparison of expression patterns between different data sets (18).

In recent years, several databases and analysis tools also have been published that feature high-throughput analysis results of microRNA sequence or expression. Among these, miRBase functions as a central repository for microRNA genomics for a variety of organisms and thus serves the community with up-to-date microRNA sequence, chromosome location and transcript information (19). mSigDB, using motif lists from Xie *et al.* (8), provides microRNA target gene lists that could be tested

*To whom correspondence should be addressed. Tel: +90 312 290 2123; Fax: +90 312 266 5097; Email: konu@fen.bilkent.edu.tr
Correspondence should also be addressed to Aybar C. Acar. Tel: +90 312 290 2094; Fax: +90 312 266 4047; Email: aacar@cs.bilkent.edu.tr

for enrichment with Gene Ontology (GO) functional terms, KEGG signaling pathways or other gene lists (20). Similarly, a manually curated database, called Mir2DiseaseBase, can be used for extracting associations between diseases and microRNAs (21). Most recently, mirBridge has been developed to predict microRNA function and link microRNAs with cellular pathways using network algorithms (22). Among the expression analysis focused databases, miRgator is a comprehensive repository and analysis tool for microRNA expression, target and ontology data providing a graphical transcriptional evaluation of selected microRNA types for mice or humans (23). microRNA.org is another source of microRNA expression and functional data for understanding microRNA expression regulation through target prediction and examination of tissue transcript abundance (24). Accordingly, using a database approach has been fruitful in allowing the users to interactively query and make associations among large-scale data sets and thus is highly suited for exploring the association between sequence motifs and expression profiles of microRNAs and meta-analysis of microRNA expression data sets.

In the present study, we have developed the microRNA Expression and Sequence Analysis Database, mESAdb, to provide a series of interactive analysis tools for testing the association of microRNA sequence characteristics with target gene function, human diseases and microRNA expression patterns using multivariate analyses. mESAdb is also a meta-analysis tool for comparative analysis of function and expression for microRNA lists across different taxa, including human, mouse and zebrafish. Complementing existing databases, mESAdb takes advantage of the available sequence information to search for microRNAs with common motifs (e.g. dinucleotide frequencies or conserved seed sequences) after which these microRNA sets can be analyzed for determination of the extent of coordinate expression and target gene enrichment using terms from GO, KEGG and HUGE Navigator databases (25–27). mESAdb is compatible and periodically updated in an automated way with data from related large external repositories. It also allows upload and analysis of user-specified data sets, and makes extensive use of existing and customized R packages (28). Overall, we believe that mESAdb, by specifically addressing the need for comparative and multivariate analysis of microRNA sequence and expression profiles, is likely to significantly enhance our understanding of the role of microRNAs in biological processes.

DATABASE DESIGN AND STRUCTURE

mESAdb enables access and retrieval of microRNAs with specified motifs to associate and analyze them functionally as well as based on expression profiles (Figure 1). An initial version of this work was presented in abstract form in BioSysBio 2007: Systems Biology, Bioinformatics, Synthetic Biology (29).

Data collection and storage

Data used in mESAdb are obtained periodically from multiple sources and processed for integration into the underlying MySQL database using a series of routines which download, parse and integrate these data from relevant sources (Ensembl, miRBase, microCosm, HUGE, KEGG and GO) either directly or through the Biomart integration service (Figures 1 and 2) (19,25–27,30).

Mature microRNA ID and sequences were downloaded from miRBase Release 15 (31); each microRNA was associated with a species-specific sequence and stored in a table. microRNA microarray experiment data sets for human, mouse and zebrafish, primarily focusing on expression from different tissues and developmental stages, were stored separately as default data sets (14,32–38) (Table 1; Supplementary Data). Tables containing the normalized expression values were associated with sequence data linked with the corresponding miRBase names for these microRNAs (Figure 2). Where available, the probe sequences printed on microarrays that match exactly with the species-specific reverse complementary sequences in miRBase were included resulting in increased stringency; thus the number of microRNAs from each microarray study incorporated into mESAdb might be smaller than that reported in the original study. Expression data were logarithmically transformed where necessary, and quantile normalized (39). To link sequence and expression properties with functional information, the predicted human targets were retrieved from MicroCosm Targets (Figure 2) (19). These targets were further processed on the R environment (Version 2.11.1) (www.bioconductor.org); transcript IDs were matched with Ensembl Gene IDs (Ensembl Release 59) using the package *biomaRt* (30). Only a single Ensembl ID was retrieved for each target gene with multiple transcript entries. Species-specific microRNAs were paired with target gene IDs associated with ontology terms and these matched pairs were stored in mESAdb's underlying DBMS (Figure 2; MySQL). KEGG and Gene Ontology terms associated with microRNA targets were extracted and matched with the corresponding microRNA IDs (25,26). The disease terms associated with microRNA targets were obtained from the phenopedia view of HUGE Navigator, an integrated knowledge base of genetic associations and human genome epidemiology (27). These terms were parsed and matched with microRNA targets and stored in the MySQL tables underlying mESAdb. Target and associated terms are updated periodically (Figure 2).

User-specified expression data set management

mESAdb incorporates a tool for the upload of user-specified expression data sets provided as comma separated files (Figure 3). The user is free to add, view and remove expression data sets having expression data for arbitrary numbers of microRNAs against arbitrary number of expression classes (e.g. tissues, developmental stages, disease states). The format for the input file is straightforward: a comma delimited file with the first row giving the names of the classes, the subsequent lines

Figure 1. Screenshot of the mESAdb main page. The modules, 'motif-expression', 'expression-expression', 'motif-function' and 'microRNA search', are shown.

of the file each beginning with the name of the microRNA (e.g. the miRBase ID) and, optionally, the probe sequence, followed by the measured expression for each of the classes given in the header line. The file uploaded is preprocessed line-by-line; for each, if the reverse complement of the probe sequence given for that line contains the mature sequence for the corresponding microRNA as given in the latest miRBase, the line is verified. For lines that cannot be thus verified, the system searches for a match in the miRBase sequences for the relevant species. If found, the microRNA is renamed to its miRBase standard name, if not, the line is discarded. Subsequently, the lines for duplicate microRNAs are averaged. The upload module generates a downloadable text file listing the actions performed while parsing and processing the .csv file. mESAdb uses nomenclature by miRBase for cross taxa comparisons performed in 'expression-expression' module where microRNAs with the same name from two different species are matched. Most microRNAs carrying the same name exhibit high sequence similarity across species while ~5% are relatively divergent. Data

Upload utility also warns users for such cases. The module also provides utilities to log transform, center and scale or quantile normalize the expression data upon verification by miRbase (Figure 3).

A user-uploaded data set is tied to the specific user account that creates it and may be retrieved from another source or may be the product of the users' own research. We protect privacy of proprietary data by keeping uploaded sets visible only to the account that owns it and no data is retained once a user removes a data set. An exemplary data set was provided in the current version of the mESAdb (40) (i.e. GSE2564NORMAL_seq.csv; mESAdb supporting material; http://konulab.fen.bilkent.edu.tr/mirna/supplementary_files.php). Accordingly, we downloaded GSE2564 expression series matrix from GEO (41). This data set includes normal tissues from stomach ($n = 5$), colon ($n = 5$), pancreas ($n = 1$), liver ($n = 3$), kidney ($n = 3$), bladder ($n = 2$), prostate ($n = 8$), uterus ($n = 9$), lung ($n = 4$), breast ($n = 3$) and brain ($n = 2$), together with cancer samples for different tissues. For the example used herein, only the expression data on the normal tissues were obtained; linked with microRNA IDs and probe

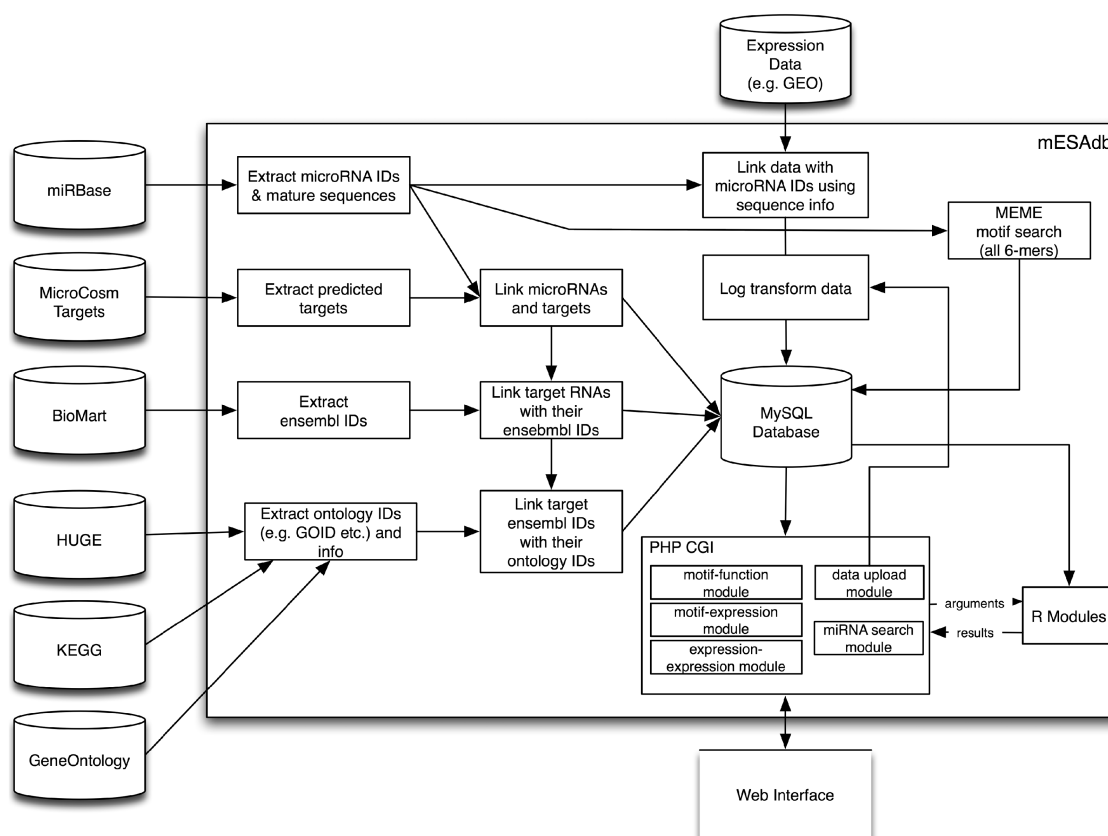


Figure 2. Workflow diagram of mESAdB. MESAdB combines data from a variety of external data sources. For example, microRNA mature sequences and IDs are retrieved from miRBase and matched with microRNA data sets (e.g. from GEO). microRNA sequences are processed by the MEME motif finder for conserved motifs. The microRNA targets are fetched from EBI's MicroCosm Targets for each species; BioMart is used to get the ENSEMBL Gene IDs of the targets' transcript IDs. These ENSEMBL Gene IDs are then linked to HUGE Navigator Disease IDs, KEGG Pathway IDs and GO IDs. A user-friendly interface has been developed in PHP for accessing data in the system and allowing versatile analysis via various R scripts (<http://php.net>; <http://www.r-project.org/>; <http://www.mysql.com/>).

Table 1. Default data sets provided in mESAdB (processed data sets; supporting material at http://konulab.fen.bilkent.edu.tr/mirna/supplementary_files.php)

References	Species	GSE No.	Platform	Pubmed ID	Tissues
Meiri <i>et al.</i> (36)	<i>Homo sapiens</i>	GSE20414	GPL10067	20483914	Ly, K, En, Lu, Bl, B, H, Li
Navon <i>et al.</i> (14)	<i>Homo sapiens</i>	GSE14985	GPL8227	19946373	Br, Pr, Ly, O, Co, Li, Te, Lu
Ach <i>et al.</i> (32)	<i>Homo sapiens</i>	GSE11806	GPL6955	18783629	Pl, B, Br, H, Th, Li, O, SM, Te
Barad <i>et al.</i> (34)	<i>Homo sapiens</i>	–	MOE-ER array	15574827	HeLa, B, Li, Th, Te, Pl
Baskerville and Bartel (33)	<i>Homo sapiens</i>	–	MWG biotech	15701730	BM, B, H, K, Li, Lu, Pa, Pr, SM, Sp, Th, FC, Ly, Co, HeLaS3, Ce, Bl, Te, A, U, Br, F, SI, Pl, O
Wienholds <i>et al.</i> (38)	<i>Danio rerio</i>	GSE2628	GPL2023	15919954	B, Ey, SM, H, Gi, Fi, Sk, G, Li, Te, O
Thomson <i>et al.</i> (37)	<i>Mus musculus</i>	GSE1635	GPL1391	15782152	Li, K, Lu, O, H, B, Th, ES, EBD3, EBD28, E7, E11, E15, E17
Beuvink <i>et al.</i> (35)	<i>Mus musculus</i>	–	Custom	17355992	B, Li, Lu, SI, SM, H, K, Sp

F, Fallopian tube; U, Uterus; Ly, Lymph node; Pl, Placenta; Br, Breast; Pa, Pancreas; Li, Liver; B, Brain; Th, Thymus; H, Heart; Lu, Lungs; Sp, Spleen; Te, Testicle; O, Ovary; K, Kidney; SM, Skeletal muscle; SI, Small intestine; Co, Colon; Pr, Prostate; Bl, Bladder; Ce, Cervix; A, Adrenal gland; St, Stomach; BM, Bone Marrow; FC, Frontal Cortex; Ey, Eye; Gi, Gill; Fi, Fin; Sk, Skin; G, Gut; HeLa, HeLa Cells; HL3S, HeLa S3; En, Endometrium. (<http://php.net>; <http://www.r-project.org/>; <http://www.mysql.com/>)

sequences in the GPL1986 description file; and a comma separated file was formed for upload. An account of processing of the microRNAs in the .csv file was generated by mESAdB. The data set, called GSE2564_normal, could be uploaded using the 'Manage Datasets' facility of mESAdB (Figure 3) and compared with the existing data sets listed in Table 1.

Integration of R-packages

mESAdB uses a hybrid of PHP and R as a computational environment. The basic operations and the web interface elements are coded in PHP whereas more significant statistical analyses are performed in R (Figure 2). The web interface has been made as responsive and user-friendly as

Figure 3. Screenshot of the data upload module. User can select from species and microarray normalization options and then browse to upload a data set.

possible with the addition of dynamic elements created with Javacript and the JQuery UI (<http://jqueryui.com/>) library. The communication between the PHP and R environments is performed using the common underlying MySQL database and Unix pipes. Briefly, a PHP script creates a child R process to which command line arguments are passed onto. The R process uses this information to retrieve the relevant information from the MySQL database and subsequently prepares the output (e.g. graphics; the bar plot, correspondence plots) which it passes onto the calling PHP script to display on the page. If the output is mostly textual (e.g. tabular data), it is passed on the output stream of the R program. If it is a larger result like an image, the R program saves it under a predetermined filename in a temporary location, which the PHP script retrieves from once the child R process is finished. This two-way communication between the PHP code and its R child processes has been implemented as a simple but effective API, which allows new R scripts to be easily integrated into the mESAdb tool as needed. This enables mESAdb to build on well-designed and verified analysis packages such as MADE4 (28) available for the R environment and use them to leverage its analysis tasks.

DESCRIPTIONS OF ANALYSIS MODULES

Motif-expression

mESAdb has a motif selection tool with a pulldown menu in which users might select from different options to group retrieved microRNAs with a given motif, i.e. dinucleotide motifs or motifs up to 6 nt long using the IUPAC code (42). It is also possible to upload user-specified microRNA lists. ‘Motif-expression’ module integrates the motif selection tool with default microarray data sets found in mESAdb as well as those uploaded by the user (Figures 1 and 3; Table 1). Accordingly, mESAdb provides a platform for visualization of microRNA expression in humans, mouse and zebrafish. Once a microRNA list is selected, expression of this set of microRNAs can be investigated using three different analysis options: ‘expression analysis’, ‘correspondence analysis’ and ‘co-intertia analysis’.

The ‘expression analysis’ option enables the user to compare, using bar plots, the amount of mean expression of the selected microRNAs with those of the remaining microRNAs across the studied expression classes, i.e. tissues or developmental stages. Expression data (Table 1) for the selected microRNAs and those for the

unselected microRNAs are extracted from the quantile normalized log transformed expression tables that have been generated by the mESAdb Data Upload facility. The class (e.g. tissue) specific mean values for the selected and unselected microRNAs then are plotted separately for each column of the data set (e.g. each tissue) using a bar plot. Bars are color coded by the value of the ϕ -coefficient to assess the association of the selected microRNAs with the tissue in consideration (also called the Yule- ϕ ; Supplementary Data) (43). A dynamic hover feature has been implemented for user to see exact information about each column by hovering the mouse pointer over it in the barplot. Expression data sets are accessible in the html format, and the χ^2 and P -values for the ϕ -coefficient also are generated. Help boxes are made available for data plots and analysis tools.

mESAdb performs multivariate analysis of expression using the R package *MADE4* customized for visualization and analysis in mESAdb (28). ‘Correspondence analysis’ of the selected set of microRNAs produce three graphical outputs, allowing for visualization of the expression patterns across classes (e.g. tissues), or microRNAs, or both the classes and microRNAs. ‘Co-inertia analysis’ (28) of the selected set of microRNAs helps visualize the similarities between microRNA expression and occurrence of common 6-mer MEME motifs (44) found among the microRNA sequences housed in mESAdb (45). Users can link from a motif to back to the ‘expression analysis’ module explained above to visualize the expression data as bar plots per expression class (e.g. tissue), of the group of microRNAs used in the coinertia plot containing the specified motif. MEME motif outputs we generated for the human, mouse and zebrafish microRNAs can be

accessed from the supporting material (http://konulab.fen.bilkent.edu.tr/mirna/supplementary_files.php) found at the mESAdb.

Expression-expression

This module provides a tool for meta-analysis of microRNA expression data sets. Selected sets of microRNAs can be investigated with regard to the data sets listed in Table 1 in a pair-wise fashion; other user-defined data sets can be uploaded and analyzed as well (Figure 1). ‘Expression-expression module’ outputs coinertia graphics for (i) classes (e.g. tissues) and (ii) microRNAs, and also a heatmap of both data sets using customized *MADE4* (28) and *heatplus* (<http://bioconductor.org/packages/2.6/bioc/vignettes/Heatplus/inst/doc/Heatplus.pdf>) packages in R (www.bioconductor.org). The output has been customized for better visualization; and the degree of association, indicated by the RV coefficient (28,46) between two different microarray data sets also is provided. A high RV score suggests better correlation among data sets. For the microRNA oriented coinertia graph, several utilities are provided in order to facilitate the visualization of potentially high numbers of datapoints. It is possible to visualize the microRNA datapoints with or without labels on the coinertia graph. The coinertia tool also provides an automatic clustering of the microRNAs based on the similarity of their expressions in both data sets using k -means clustering (47); the default clustering displayed is the clustering with the maximum silhouette coefficient (48). Since k -means clustering is not deterministic, for each k -value the module performs 20 runs of the algorithm and the best clustering for each k is selected using highest

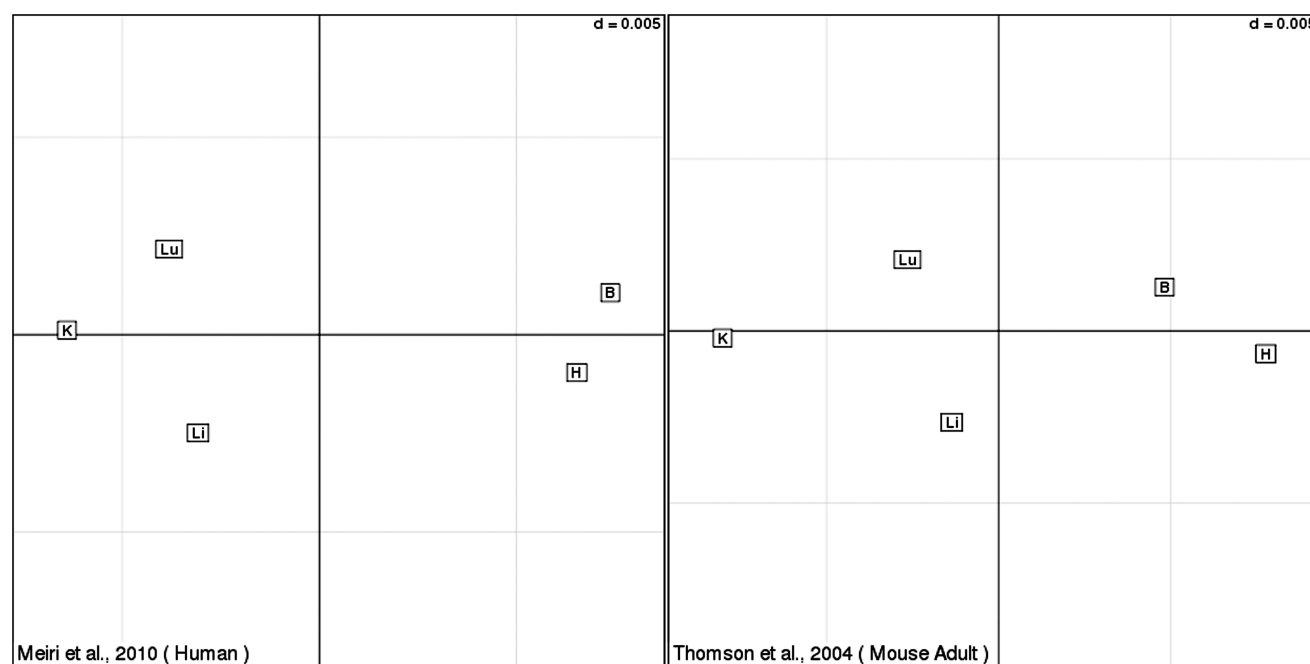


Figure 4. Coinertia plot of Meiri and Thomson expression data sets for a set of microRNA clusters with sequence similarity (mESAdb supporting material; http://konulab.fen.bilkent.edu.tr/mirna/supplementary_files.php). Similarity of microRNA expression patterns between mice and humans are shown for brain (B), heart (H), kidney (K), liver (Li), and lung (Lu).

silhouette. The clustering with the overall best silhouette is displayed by default. The user can manually set a cluster number between 2 and 10 clusters (i.e. $2 \leq k \leq 10$) if desired. These clusters can further be investigated to visualize the expression profiles for the given data sets using expression bar plots of in-cluster and out-of-cluster microRNAs, by clicking on the cluster centroids.

Motif-function

This function may be useful for functional analysis of, for example, a set of differentially expressed microRNAs (Figures 1 and 2). In the present study, information from HUGE Navigator, in addition to GO and KEGG databases can be associated with the selected microRNAs (25–27). For any selected subset of microRNAs, mESAdb

then can be used to retrieve the mappings of the selected functional terms, with the targets of these microRNAs and subsequently to calculate a probability value based on the hypergeometric distribution (49).

microRNA search

Functional and expression correlates of a single microRNA can be assessed using this module to enable a quick search involving multiple modules of mESAdb (Figure 1). Terms from GO, HUGE, KEGG and target genes associated with the given microRNA can be extracted; and the observed and expected counts as well as hypergeometric *P*-values can be downloaded. Expression profile of the selected microRNA also can be visualized using the aforementioned bar plots and downloaded as .txt files.

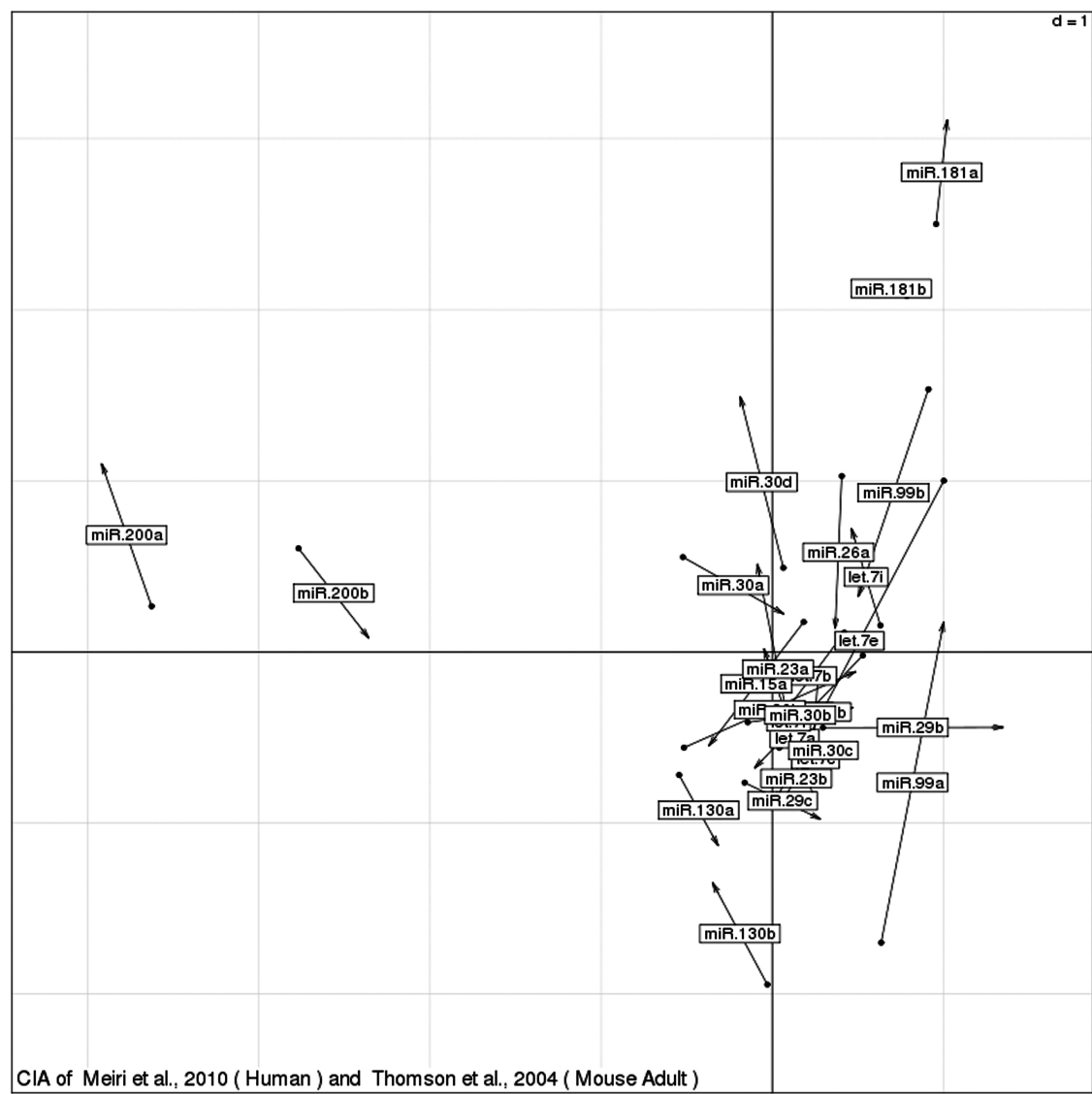


Figure 5. Distribution of microRNAs after dimension reduction by co inertia analysis. microRNAs related in expression clustered together. The length of an arrow correlates with the amount of expression divergence for a particular microRNA between the two data sets, i.e. human versus mouse.

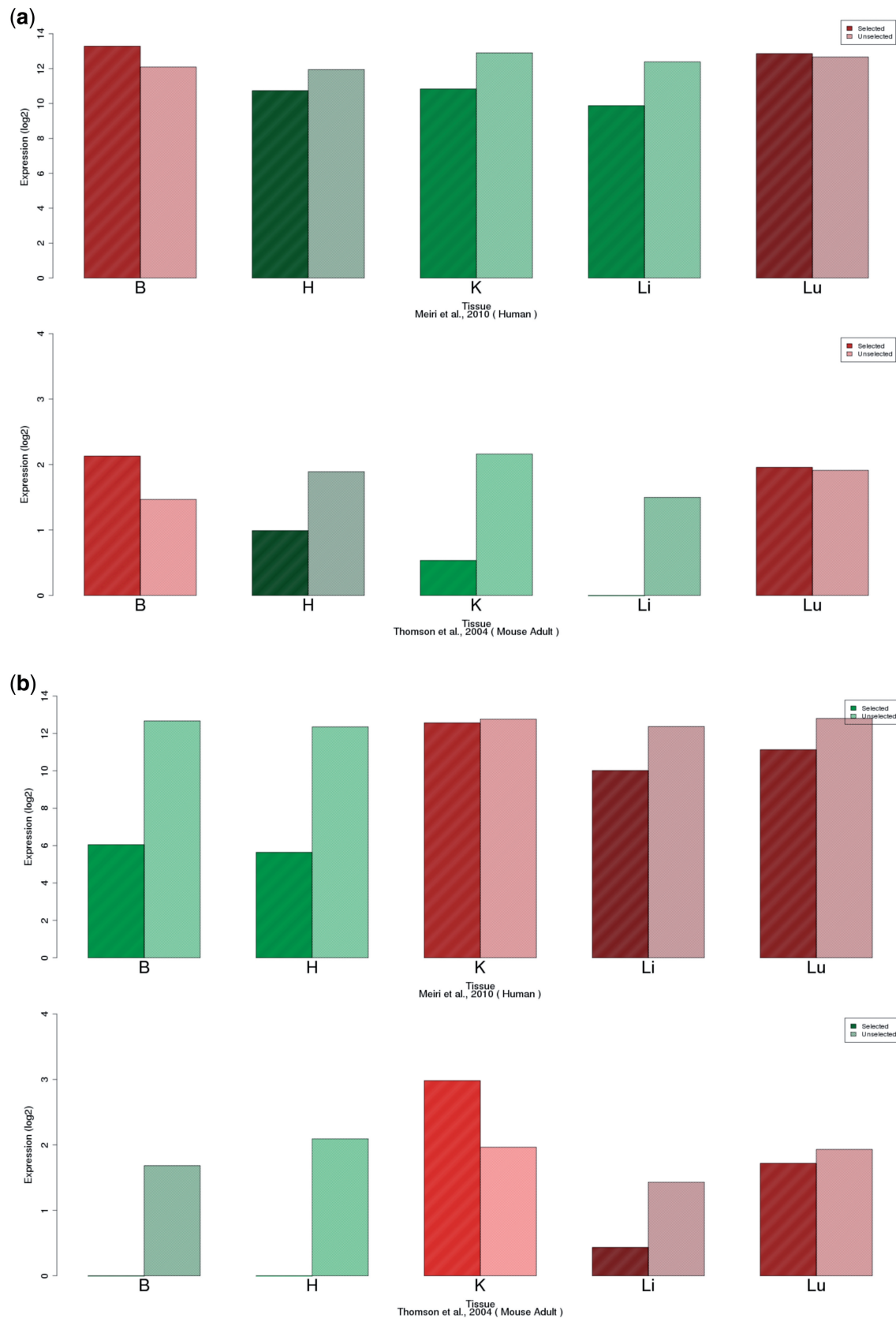


Figure 6. Similarity of expression of microRNA expression from Meiri and Thomson. Expression bar plot of (a) *mir-181a* and *mir-181b* cluster (b) *miR-200a* and *miR-200b* for five different tissues [i.e. brain (B), heart (H), kidney (K), liver (Li), and lung (Lu), respectively.]. For each tissue, the bar on the left indicates the mean expression of the members of the cluster and the right hand bar indicates the mean expression of the remainder of the data set.

DATABASE USAGE

mESAdb is a highly interactive and flexible database with an ability to analyze and visualize selected expression profiles for a given subset of microRNAs in a multivariate manner using correspondence and co-inertia analyses. One can also study a single microRNA of interest using bar plots associated with a gene expression enrichment index, based on the ϕ -coefficient. This index provides a significance value for the relative enrichment of a microRNA(s) in a particular class with respect to others (Supplementary Data). Furthermore, the user can obtain information about the functional enrichment of a microRNA or a group of microRNAs using different databases, including GO, KEGG and HUGE Navigator.

The default expression data sets currently focus on tissue- and stage-specificity; however, users can add any microarray data containing other types of expression classes, e.g. cancer versus normal, treatment versus control (Figure 3). This allows for great flexibility in analyzing one's own research data.

As an example, we demonstrate that the user can compare two data sets with respect to a list of microRNA clusters that are common to both mice and humans. Using the 'expression-expression' module of mESAdb, we have chosen a human (36) and a mouse (37) data set (Table 1) and uploaded a microRNA list (mESAdb supporting material; http://konulab.fen.bilkent.edu.tr/mirna/supplementary_files.php; the list included *let-7a-i*, *mir-130a-b*, *mir-15a-b*, *mir-181a-b*, *mir-200a-b*, *mir-23a-b*, *mir-26a-b*, *mir-29a-c*, *mir-30a-d* and *mir-99a-b* clusters). We then performed the coinertia analysis using only the tissues common to both data sets, namely, brain (B), liver (Li), lung (Lu), kidney (K) and heart (H) (Figure 4). mESAdb through coinertia analysis allows for comparison and visualization of two expression data sets by plotting them side by side in terms of the expression of selected microRNAs for the given tissues. Accordingly, we found that microRNAs in our list were expressed similarly in human and mouse data sets because the location of the projected tissues closely corresponded between the two plots (Figure 4). mESAdb also enables visualization of the expression of selected microRNAs from both data sets by simultaneously overlaying them on a two-dimensional plot. In this microRNA-oriented view, similarly expressed microRNAs are found closer in space. The analysis of our microRNA list indicated that several microRNAs formed clusters based on their expression, in particular, *mir-181a* and *mir-181b*, and *mir-200a* and *mir-200b* (Figure 5 and Supplementary Data). Indeed, *mir-181a* and *mir-181b* that are similar in sequence and diverging only with 3 nt exhibit a common sequence motif (i.e. AACATTCA) in their first 8 nt. Similarly, *mir-200a* and *mir-200b* are similar in their sequences containing a common motif (i.e. TAA[C][T]ACTG) in their first 8 nt. Using the 'expression analysis' module, *miR-181a* and *miR-181b* were found to be expressed primarily in the brain and lung (Figure 6a) whereas the *miR-200a-b* cluster was clearly expressed mostly in the kidney and lung both in mice and humans (Figure 6b).

Our findings suggested that expression patterns of *mir-181a-b* and *mir-200a-b* were highly conserved between human and mice.

In conclusion, mESAdb focuses on providing a meta-analysis tool/database to enhance our understanding in an important field in microRNA biology, i.e. discovery of associations between microRNA sequence and expression. mESAdb is advantageous because it allows interactive analysis of selected subsets of microRNAs in addition to analysis of single microRNA types. Its modular and expandable nature makes mESAdb a unique and functional database for comparative analysis of microRNA sequence and expression.

AVAILABILITY

mESAdb is freely available at <http://konulab.fen.bilkent.edu.tr/mirna/>. mESAdb is located on a Linux server (Apache/2.2.4; Ubuntu 8.04 LTS, Kernel: 2.6.24-24-server; PHP 5.2.3-1; R-2.11.1) equipped with four Intel® Xeon® CPU E5335, 2.00GHz processors and 8 GB RAM. Microarray data sets incorporated into the mESAdb, as well as R codes used in correspondence analysis are available for download at the mESAdb site.

FUTURE EXTENSIONS

Modular nature of mESAdb allows for incorporation of additional data sets and statistical tools. Future extensions to mESAdb will include addition of microarray data sets from GEO particularly focusing on different aspects of human pathogenesis. Use of R packages enhances the modular nature of the mESAdb thus future addition of statistical and visual tools for sequence/expression/function analysis of microRNAs is planned.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Alper Tolga Kocatas for help in optimizing MySQL queries for faster execution, Sergen Eren for proofreading microarray data set processing, Rengul Cetin-Atalay for providing rack space for the server and Michelle Adams for her helpful comments on the article.

FUNDING

The Scientific and Technological Research Council of Turkey (TUBITAK) and Bilkent University, Ankara. Funding for open access charge: Partially waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

- Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **23**, 281–297.
- 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.
- 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.
- Iwama,H., Masaki,T. and Kuriyama,S. (2007) Abundance of microRNA target motifs in the 3'-UTRs of 20527 human genes. *FEBS Lett.*, **581**, 1805–1810.
- Hertel,J., Lindemeyer,M., Missal,K., Fried,C., Tanzer,A., Flamm,C., Hofacker,I.L. and Stadler,P.F. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics*, **7**, 25.
- Bentwich,I., Avniel,A., Karov,Y., Aharonov,R., Gilad,S., Barad,O., Barzilai,A., Einat,P., Einav,U., Meiri,E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
- Yu,J., Wang,F., Yang,G.H., Wang,F.L., Ma,Y.N., Du,Z.W. and Zhang,J.W. (2006) Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochem. Biophys. Res. Commun.*, **349**, 59–68.
- 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.
- Sun,Y., Koo,S., White,N., Peralta,E., Esau,C., Dean,N.M. and Perera,R.J. (2004) Development of a micro-array to detect human and mouse microRNAs and characterization of expression in human organs. *Nucleic Acids Res.*, **32**, e188.
- Sempere,L.F., Freemantle,S., Pitha-Rowe,I., Moss,E., Dmitrovsky,E. and Ambros,V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.*, **5**, R13.
- Houbaviy,H.B., Murray,M.F. and Sharp,P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, **5**, 351–358.
- Liu,C.G., Calin,G.A., Meloon,B., Gamlie,N., Sevignani,C., Ferracin,M., Dumitru,C.D., Shimizu,M., Zupo,S., Dono,M. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA*, **101**, 9740–9744.
- Bargaje,R., Hariharan,M., Scaria,V. and Pillai,B. (2010) Consensus miRNA expression profiles derived from interplatform normalization of microarray data. *RNA*, **16**, 16–25.
- Navon,R., Wang,H., Steinfeld,I., Tsalenko,A., Ben-Dor,A. and Yakhini,Z. (2009) Novel rank-based statistical methods reveal microRNAs with differential expression in multiple cancer types. *PLoS One*, **4**, e8003.
- Smith,D.D., Saetrom,P., Snove,O. Jr, Lundberg,C., Rivas,G.E., Glackin,C. and Larson,G.P. (2008) Meta-analysis of breast cancer microarray studies in conjunction with conserved cis-elements suggest patterns for coordinate regulation. *BMC Bioinformatics*, **9**, 63.
- Shalgi,R., Lieber,D., Oren,M. and Pilpel,Y. (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput. Biol.*, **3**, e131.
- Sood,P., Krek,A., Zavolan,M., Macino,G. and Rajewsky,N. (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA*, **103**, 2746–2751.
- Madden,S.F., Carpenter,S.B., Jeffery,I.B., Bjorkbacka,H., Fitzgerald,K.A., O'Neill,L.A. and Higgins,D.G. (2010) Detecting microRNA activity from gene expression data. *BMC Bioinformatics*, **11**, 257.
- Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
- Subramanian,A., Kuehn,H., Gould,J., Tamayo,P. and Mesirov,J.P. (2007) GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*, **23**, 3251–3253.
- Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
- Tsang,J.S., Ebert,M.S. and van Oudenaarden,A. (2010) Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol. Cell*, **38**, 140–153.
- Nam,S., Kim,B., Shin,S. and Lee,S. (2008) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.
- Betel,D., Wilson,M., Gabow,A., Marks,D.S. and Sander,C. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
- Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
- Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Yu,W., Gwinn,M., Clyne,M., Yesupriya,A. and Khourey,M.J. (2008) A navigator for human genome epidemiology. *Nat. Genet.*, **40**, 124–125.
- Culhane,A.C., Thioulouse,J., Perriere,G. and Higgins,D.G. (2005) MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics*, **21**, 2789–2790.
- Kaya,K.D., Karakulah,G., Yalciner,C. and Konu,O. (2007) MicroRNA sequence and expression database. *BMC Syst. Biol.*, **1**, P29.
- Durinc,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
- Griffiths-Jones,S. (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.
- Ach,R.A., Wang,H. and Curry,B. (2008) Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. *BMC Biotechnol.*, **8**, 69.
- Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
- Barad,O., Meiri,E., Avniel,A., Aharonov,R., Barzilai,A., Bentwich,I., Einav,U., Gilad,S., Hurban,P., Karov,Y. *et al.* (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.*, **14**, 2486–2494.
- Beuvink,I., Kolb,F.A., Budach,W., Garnier,A., Lange,J., Natt,F., Dengler,U., Hall,J., Filipowicz,W. and Weiler,J. (2007) A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs. *Nucleic Acids Res.*, **35**, e52.
- Meiri,E., Levy,A., Benjamin,H., Ben-David,M., Cohen,L., Dov,A., Dromi,N., Elyakim,E., Yerushalmi,N., Zion,O. *et al.* (2010) Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic Acids Res.*, **38**, 6234–6246.
- Thomson,J.M., Parker,J., Perou,C.M. and Hammond,S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nat. Methods*, **1**, 47–53.
- Wienholds,E., Kloosterman,W.P., Miska,E., Alvarez-Saavedra,E., Berezikov,E., de Bruijn,E., Horvitz,H.R., Kauppinen,S. and Plasterk,R.H. (2005) MicroRNA expression in zebrafish embryonic development. *Science*, **309**, 310–311.
- Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
- 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.
- Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Rudnev,D., Evangelista,C., Kim,I.F., Soboleva,A., Tomashevsky,M., Marshall,K.A. *et al.* (2009) NCBI GEO: archive for

- high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
42. Panico,R., Powell,W.H. and Richer,J.C. (eds), (1993) *A Guide to IUPAC Nomenclature of Organic Compounds*. Blackwell Scientific Publications, Oxford.
43. Guilford,J. (1941) The phi coefficient and chi square as indices of item validity. *Psychometrika*, **6**, 11–19.
44. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
45. Hennig,C. and Hausdorf,B. (2004) Distance-based parametric bootstrap tests for clustering of species ranges. *Comput. Stat. Data Anal.*, **45**, 875–895.
46. Robert,P. and Escoufier,Y. (1976) A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Stat.*, **25**, 257–265.
47. Hartigan,J.A. and Wong,M.A. (1979) Algorithm AS 136: a K-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.*, **28**, 100–108.
48. Lovmar,L., Ahlford,A., Jonsson,M. and Syvanen,A.C. (2005) Silhouette scores for assessment of SNP genotype clusters. *BMC Genomics*, **6**, 35.
49. Kachitvichyanukul,V. and Schmeiser,B. (1985) Computer generation of hypergeometric random variates. *J. Stat. Comput. Simul.*, **22**, 127–145.