

# UCbase & miRfunc: a database of ultraconserved sequences and microRNA function

Cristian Taccioli<sup>1</sup>, Enrica Fabbri<sup>2</sup>, Rosa Visone<sup>1</sup>, Stefano Volinia<sup>1</sup>, George A. Calin<sup>3</sup>, Louise Y. Fong<sup>1</sup>, Roberto Gambari<sup>2</sup>, Arianna Bottoni<sup>1</sup>, Mario Acunzo<sup>4</sup>, John Hagan<sup>1</sup>, Marilena V. Iorio<sup>5</sup>, Claudia Piovano<sup>5</sup>, Giulia Romano<sup>4</sup> and Carlo Maria Croce<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210, USA, <sup>2</sup>Biochemistry and Molecular Biology Department, University of Ferrara, 44100 Ferrara, Italy, <sup>3</sup>Department of Experimental Therapeutics, University of Texas, MD Anderson Cancer Center, Houston, TX 77030, USA, <sup>4</sup>Molecular Pathology and Biology Department 'L.Califano', 'Federico II' Naples University, 80100 Naples and <sup>5</sup>IRCCS Foundation, National Tumor Institute, 20100 Milan, Italy

Received July 9, 2008; Revised September 3, 2008; Accepted September 29, 2008

## ABSTRACT

Four hundred and eighty-one ultraconserved sequences (UCRs) longer than 200 bases were discovered in the genomes of human, mouse and rat. These are DNA sequences showing 100% identity among the three species. UCRs are frequently located at genomic regions involved in cancer, differentially expressed in human leukemias and carcinomas and in some instances regulated by microRNAs (miRNAs). Here we present UCbase & miRfunc, the first database which provides ultraconserved sequences data and shows miRNA function. Also, it links UCRs and miRNAs with the related human disorders and genomic properties. The current release contains over 2000 sequences from three species (human, mouse and rat). As a web application, UCbase & miRfunc is platform independent and it is accessible at <http://microrna.osu.edu/UCbase4>.

## INTRODUCTION

UCbase & miRfunc is a database of (i) human, mouse and rat ultraconserved elements and (ii) microRNAs (miRNAs).

The database has three main functions:

- (i) It provides ultraconserved sequence data, annotation, secondary structure and information about the disorder related to the band. Furthermore, it

shows references and links to other resources for all UCRs published by Bejerano *et al.* (1).

- (ii) It provides miRNA sequence data, annotation, secondary structure, information about disorder related to the band, references and links to other resources for all published miRNAs in the genomes of human, mouse and rat. Primary features of the nomenclature scheme, sequences and genomics coordinates are retrieved from miRBase (2).
- (iii) It provides information about miRNAs and ultraconserved elements function and expression profiles.

This is made possible by a few automated Perl scripts linked to miRBase (<http://microrna.sanger.ac.uk>) (2), NCBI (<http://www.ncbi.nlm.nih.gov>) (3) and UCSC server (<http://genome.ucsc.edu>) (4).

## ULTRA-CONSERVED SEQUENCES

Ultraconserved elements (UCRs) were discovered in 2004 by Bejerano and colleagues by bioinformatics comparisons of the genomes of mouse, rat and human (1). Four hundred and eighty-one UCR sequences are 200–779 bp in length showing 100% identity among the three species. Some of them contain protein coding sequences, but over half are not predicted to codify any protein (1).

Previous studies have suggested an important role for these noncoding sequences both in promoting the expression of several genes and in regulation of alternative splicing (5). Probably many of the UCRs date from a very early period in vertebrates evolution, as they have no orthologous counterparts in sea squirts, flies or worms

\*To whom correspondence should be addressed. Tel: +1 614 292 3063; Fax: +1 614 292 3558; Email: [carlo.croce@osumc.edu](mailto:carlo.croce@osumc.edu)

even if at least one group of these UCRs evolved from a novel retrotransposon family that was active in lobe-finned fishes, and is still active today in the 'living fossil' Coelacanth (*Latimeria chalumnae*), the ancient link between marine and land vertebrates (5).

Recently, Calin *et al.* (6) identified a functional role for miRNAs in the transcriptional regulation of cancer-associated UCRs. They proved in tumors that differentially expressed UCRs could alter the functional characteristics of malignant cells. By combining this data with the much more elaborate model involving miRNAs in human tumorigenesis, they propose that alteration in both coding and noncoding RNAs cooperate in the initiation and/or progression of malignancy.

### miRNAs

miRNAs were first described in 1993 by Lee and colleagues (7), yet the term 'microRNA' was only introduced years later in 2001 in an article published in *Science* (8). Findings over the past 5 years supported a role for miRNAs in the regulation of crucial processes such as cell proliferation (9), apoptosis (10), development (11), differentiation (12) and metabolism (13). Each miRNA is supposed to target several hundreds of transcripts (14), making miRNAs one of the main genome regulators.

miRNAs regulate their targets by direct cleavage of the mRNA and by inhibition of protein synthesis, depending on the degree of complementarities with their target 3'-UTR regions (15). miRNAs are processed from primary transcripts known as pri-miRNAs (16) but not translated into protein. A portion of this primary transcript is recognized and cleaved by the enzyme Drosha into a miRNA precursor (pre-miRNA) (17) and finally processed to functional miRNA.

These pre-miRNAs are then processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiates the formation of the RNA-induced silencing complex (RISC) (18). This complex is responsible for the gene silencing observed due to miRNA expression and RNAs interference. The pathway in plants varies slightly due to their lack of Drosha homologs (19).

At the time of this writing, the miRBase server version 11 contains 678 human pre-miRNAs (48–150 nt long) and 847 mature miRNAs (17–28 nt long).

Pre-miRNAs do not have a perfect double-stranded RNA structure and they are topped by a terminal loop (hairpin shape). Most of them are conserved between classes, the free energy is often less than  $-25$  Kcal/mol, the GC-Ratio is usually 30–70% and the entropy is between 0 and 2 (20).

Recently, miRNA expression has been linked to cancer. The first evidence came from the finding that miR-15a and miR-16-1 are downregulated or deleted in most patients with chronic lymphocytic leukemia (CLL) (21). Several other groups have then studied the miRNA expression in cancer patients and found that miRNAs are differentially expressed in normal and tumor tissues (22–26) and, in some cases, are associated with prognosis (27).

### QUERIES

The UCbase & miRfunc database contains three principal types of search:

- (i) *Multiple queries* can be performed typing the names of selected miRNAs or UCRs and obtaining as output the band, the gene name, the coordinates, the human diseases (28) related to the band and the secondary structure performed using UNAFold (29) (Figure 1). These tables can be also downloaded in Excel format. A complete list of UCRs and miRNAs gene names is available clicking on the link above the multiple queries text box.
- (ii) *miRNA and UCRs information* can also be retrieved searching for band, human Mendelian inheritance disorders (28) related to that band or just typing the name of a gene (Figure 2).
- (iii) *The use of an interactive map of biological systems and diseases* allows to link miRNAs and ultraconserved sequences to the related disorders (Figure 2).

### FUNCTION

One of the most useful peculiarities of UCbase & miRfunc is that it provides tables showing ultraconserved elements properties (enhancer activity, alternative splicing, splicing NMD regulation and transcription evidence) and miRNA 'experimentally proved' function, addressing a number of common questions pertaining to these noncoding RNA classes (Figure 3). We decided to show the experiments that provide molecular biology evidence of a particular miRNA function. Paper references supplying only miRNA data were not included in the table.

All of this information is periodically retrieved (every 2 months) using a Perl scripts implemented with WWW::Search::Pubmed (<http://search.cpan.org/~gwilliams/WWW-Search-PubMed-1.004/>) and WWW::Mechanize (<http://search.cpan.org/dist/WWW-Mechanize/>) modules which provide an API search engine linked to the Pubmed database (30).

### SEQUENCE COMPARISON

The sequence comparison page allows the researchers to match selected sequences against miRNAs and ultraconserved elements (exact/500/1000/2000/5000/10 000 bp up/downstream) (Figure 4) using The Basic Local Alignment Search Tool (BLAST) which finds regions of local similarity between sequences (31).

The researchers can also align miRNAs and ultraconserved elements versus TRANSFAC (32) and CpG Island Searcher databases (33). Checking the parsing option it is possible to turn the output into a single-line summary. This function is invoked internally by a Perl script which uses the Bio::SearchIO module (<http://search.cpan.org/~birney/bioperl-1.2.3/Bio/SearchIO.pm>).

Moreover it is possible to download a RepeatMasker (34) miRNAs/UCRs table showing the repetitive miRNAs and ultraconserved sequences in human genome.

**Multi Search function (?)**

uc  mir

UC.1  
UC.77A

Submit Query Reset

Home Blast Search Function Expression Resources Contact us

NAME	UCSC	TYPE	GENE	BAND	LENGTH	ORIGIN BAND RELATED	II STRUCTURE	II STRUCTURE (STRIP)
uc.1	chr1:17007243-10307649	p	PDX14	1p36.32	207		uc.1	
uc.77A	chr2:145294534-145296829	c	ZFXK1B	2q32.3	296		uc.77A	

**UCbase info**

**UCbase and MIRFunc version 1.0 Released**

**Ultraconserved database at microrna.org**

Background: 481 Ultraconserved sequences (UCRs) longer than 200 bases were discovered in the genomes of human, mice and rats. These are DNA sequences showing 100 percent identity among the human, mouse and rat genomes. UCRs are frequently located at genomic regions involved in cancer, differentially expressed in human leukemias and carcinomas and in some instances regulated by microRNAs (miRNAs), the most extensively studied category of non-coding RNAs (ncRNAs). Here we present the first database which links UCRs and miRNAs with the related human disorders and genomic properties.

Description: UCbase & MIRfunc is a database of (i) human microRNAs and (ii) Ultraconserved elements providing information about function, expression and correlation between these classes of non-coding RNAs and the disorders related to their aberrant expression. The genomics interface allows the user to explore where whole-genome collections of miRNAs and UCRs are located with respect to annotation sets such as band, disorders and known genes. Finally, the Blast interface provides a web tool for matching miRNAs/UCRs elements against any given sequence and providing specific functional information on the results. All of these unique features are designed to facilitate investigations into miRNAs and UCRs genomic organization.

[Read More](#)

**Breaking News**

**Ultraconserved and miRNAs: News and papers**

- [Deletion of ultraconserved elements yields viable mice](#)
- [Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas](#)
- [Human genome ultraconserved elements are ultraselected](#)
- [Ultraconserved non-coding sequence element controls a subset of spatiotemporal GLI3 expression](#)

**Other Non-coding RNAs:**

- [Non coding Rna database](#)
- [Mammalian noncoding RNA](#)
- [Functional RNA database](#)
- [Noncoding RNA families](#)
- [MicroRNA database](#)
- [MicroRNA gene expression](#)
- [H/ACA and C/D box snoRNA](#)
- [methylation guide snoRNA dB](#)
- [tRNA database](#)
- [database of tmRNA seq.](#)
- [RNase P RNA and protein dB](#)
- [Signal recognition particle dB](#)

**Web SSH Login**

Username

Password

Login

**Figure 1.** Multi-queries text box (red square). Multiple queries can be performed typing the names of selected miRNAs or UCRs and obtaining as output the band, the gene name, the coordinates, the human diseases related to the band and the secondary structure up to 500 bp up/downstream.

## EXPRESSION

UCbase & miRfunc database is linked to microrna.org (35) to retrieve miRNA gene expression data sets; whereas, it collects ultraconserved sequences microarray expression data from ArrayExpress (36), GEO (37) and microrna.osu.edu server.

## DATABASE INTERFACE

Cascading Style Sheets (CSS) were used for UCbase & miRfunc web development to style web pages written in HTML. The CSS specifications are maintained by the World Wide Web Consortium (W3C). Internet media type (MIME type) text/css is registered for use with CSS by RFC 2318 (March 1998).

UCbase & miRfunc was created using MySQL database under Debian Etch Linux OS installed on a Quad Core Processor machine with 32 GB RAM.

UCbase & miRfunc database can be accessed by Perl programs running on the web server. The MySQL Perl API is mainly provided with the use of CGI (Common Gateway Interface) which is a standard for external gateway programs to interface with information servers such as HTTP servers (Apache2 in this particular case). A CGI object (cgi.pm) was used to handle POST and GET methods correctly, and distinguish between scripts called from ISINDEX documents and form-based documents.

## DATA SOURCES AND IMPLEMENTATION

Almost all the code used to develop the database and obtain miRNAs and ultraconserved elements information was written in Perl.

Perl (www.perl.org) is a programming language originally created for text manipulation and for a wide range of tasks including system administration, web development,

**UCbase & miRfunc**  
ULTRA-CONSERVED SEQUENCES AND MIRNA FUNCTION DATABASE

Multi Search function (?)  
uc  mir   
  
Submit Query Reset

Search Function Expression Resources Contact us

Human Chromosome Location

1-1	2-8	9-12	13-16	17-20	21-XY
-----	-----	------	-------	-------	-------

miRNA and UC signature in cancer and type of disease

System or Type of Disease

Cancer	Nervous	Muscle	Immune
Metabolism	Signals	Transport	

**miRNA Search by band or OMIM disorder**

1p13.1

17,20-lyase deficiency, isolated

Search by GENE (Ex. DLEU2 or intergenic)

**UC Search by band or OMIM disorder**

1p13.1

17,20-lyase deficiency, isolated

Search by GENE (Ex. PEX14 or NA)

**Help**

**miRNA and Ultraconserved sequences Searching : How to...**

This database is constructed allowing the user to find microRNA or ultraconserved sequences searching them by band, by gene or by OMIM disorder correlated to that band.

**Figure 2.** Search Page. miRNA and UCRs information can be retrieved searching for band, human Mendelian inheritance disorders related to that band or just typing the name of a gene (red highlighted boxes). The use of an interactive map of biological systems and diseases allows to link miRNAs and ultraconserved sequences to the related disorders (blue arrow).

network programming and GUI development. It is one of the most common language programming in bioinformatics, where it is valued for rapid application development and deployment, and the ability to handle large data sets. Perl is also used for the web automation.

Web automation can automate all the web processes from simple filling of forms to more complicated tasks for data transfer, web data extraction, image recognition and performing tasks based on it, scheduling tasks, batch process and file management. A Perl module called

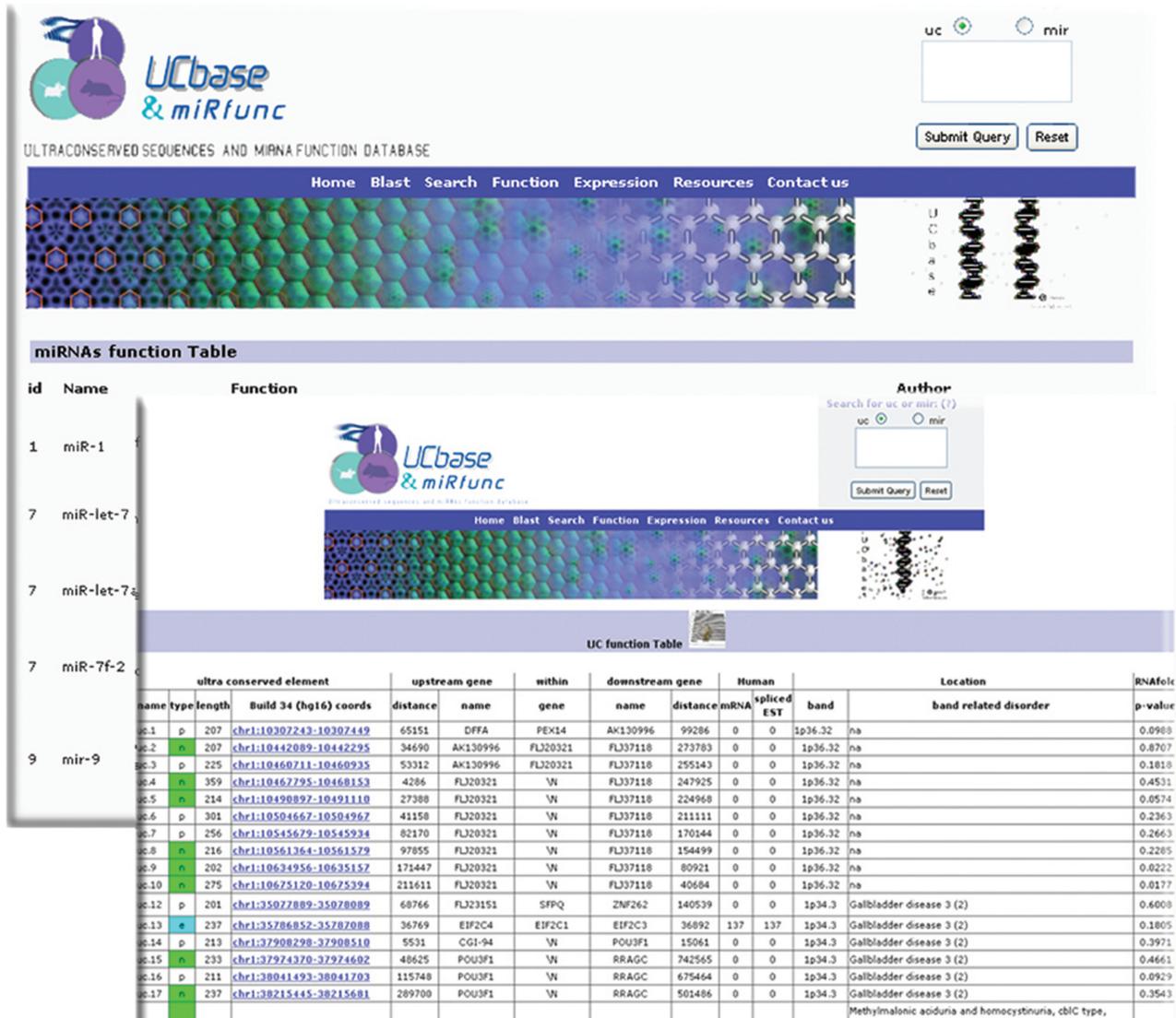


Figure 3. Function page. It provides tables showing ultraconserved elements properties (enhancer activity, alternative splicing, splicing NMD regulation and transcription evidence) and miRNA ‘experimentally proved’ function.

WWW::Mechanize was used to extract the data from NCBI (search\_pubmed\_nuovo.pl) and miRBase (www\_mechanize\_sanger\_total.pl) pages; whereas, a script written in Pascal (http://www.newbie.com) was used to extract miRNAs and ultraconserved sequences (500/1000/2000/5000/10 000 bp up/downstream) from UCSC server (ucsc.nbl).

In particular, all the utilities that come with WWW::Mechanize print the names and elements of every form and provide all the needed information when searching for form using regular expressions (http://www.opengroup.org/onlinepubs/007908799/xbd/re.html).

However, using that data in the code means to cut and paste multiple entire blocks of code. For this reason, it is useful to install HTTP::Recorder to set up an HTTP proxy. HTTP::Recorder saves each action as WWW::Mechanize code. Before running the recording script (recorder.pl), the browser proxy has to be

configured properly. Running recorder.pl starts an HTTP proxy daemon that the browser uses to make requests. The proxy uses HTTP::Recorder agent to log these requests. It saves a logfile as a .t file, which is specified when creating the HTTP::Recorder object.

Moreover, a window will appear displaying the content of .t file which should be a series of statements involving a hypothetical WWW::Mechanize object. These scripts were added to the final Perl code file called www\_mechanize\_sanger\_total.pl.

WWW::Search::PubMed version 1.004 provides instead a backend to the WWW::Search module allowing searches of the National Library of Medicine’s PubMed biomedical citation database (30) (pubmed.pl).

All the miRNAs genomic information are obtained using an automatic script linked to miRBase ftp server (ftp://ftp.sanger.ac.uk/pub/mirbase/sequences/CURRENT/database\_files) which runs periodically

The screenshot displays the UCbase & miRfunc database interface. At the top left is the logo and the text 'ULTRACONSERVED SEQUENCES AND MIRNA FUNCTION DATABASE'. On the right, there is a 'Multi Search function (?)' section with radio buttons for 'uc' (selected) and 'mir', and a search box. Below this is a navigation bar with links: 'Home Blast Browse Expression Resources Contact us'. The main content area shows search results for 'uc' with four hits: 'Hit = uc.1A, Length = 1207, Percent\_id = 97.5144987572494', 'Hit = uc.1, Length = 1207, Percent\_id = 97.5144987572494', 'Hit = uc.483A, Length = 16, Percent\_id = 100', and 'Hit = uc.428A, Length = 16, Percent\_id = 100'. Below the results is a section titled 'Blast your sequences against ultraconserved and human miRNA'. This section contains a form with 'Type' set to 'ultraconserved', 'Downstream/Upstream' set to '+/-2000', and 'Parsing' checked. A dropdown menu for 'Category' is open, showing options: 'ultraconserved', 'micromiata\_precursor', and 'micromiata\_mature'. A blue arrow points to the 'Parsing' checkbox. Below the form are 'Submit' and 'Reset' buttons. To the right of the form is a 'UCbase Download' section with a 'Join Now' button and a list of other non-coding RNAs: 'Non coding Rna database', 'Mammalian noncoding RNA', 'Functional RNA database', 'Noncoding RNA families', 'MicroRNA database', 'MicroRNA gene expression', 'H/ACA and C/D box snoRNA', 'methylation guide snoRNA dB', 'tRNA database', 'database of tmRNA seq.', 'RNase P RNA and protein dB', and 'Signal recognition particle dB'.

**Figure 4.** Sequence comparison page. This page allows the researchers to match selected sequences against miRNAs and ultraconserved elements (exact/500/1000/2000/5000/10 000 bp up/downstream). Checking the parsing option it is possible to turn the output into a single-line summary (blue arrow).

using CRON, a time-based scheduling service in Unix-like computer operating systems (<http://packages.debian.org/etch/cron>).

## CONCLUSIONS

UCbase & miRfunc is the first database containing the long 481 ultraconserved sequences discovered in the genomes of human, mouse and rat by Bejerano *et al.* (1) and identified as dysregulated in cancer by Calin *et al.* (6). In addition, UCbase & miRfunc is the first database that provides miRNA function, which is a particularly important feature due to the increasing output of miRNA data. The goal of this study is to offer the researchers an advanced set of tools that supply information about correlation between miRNAs, UCRs and the disorders related to their aberrant expression.

The first version of UCbase & miRfunc includes new methods of queries such as search for band, disorders and genes. It has the capability of using multiple queries and combining the results from several miRNA/UCRs input. In addition, the system has a new tool for the RNA secondary structure prediction (up to 500 bp up/downstream), which allows innovative visualization

of miRNAs and ultraconserved elements. This feature improves presentation of data output that is especially useful when several structures are obtained using multiple queries.

The sequence comparison tool, which is also available, allows researchers to match selected sequences that are 500/1000/2000/5000/10 000 bp upstream/downstream of all defined miRNAs and UCRs in the database.

Additionally, UCbase & miRfunc has adopted several strategies to integrate search tools. It provides updated data by using automate web scripts.

The system will be regularly upgraded with new human, mouse and rat miRNAs/UCRs by providing the corresponding reference sequences and annotations, thus allowing the data to be refined continuously with every new miRBase and UCSC database version. It also provides complete human, mouse and rat miRBase table as a single file (available from the resource web page), a feature that is particularly useful if one wants to use regular expressions to identify the segment of the input file associated with sub-field of interest.

Although other alternatives are available for retrieval of miRNA nomenclature, sequence data and annotation, UCbase & miRfunc is unique combination of features that provide biologists many options for data analysis

and discovery of relationships between miRNAs and ultraconserved sequences.

## FUTURE DIRECTIONS

The improvements to a new set of OMIM (28) and HGMD (38) databases needed for an overall plan for collection (39), hopefully should enable UCbase & miRfunc users, in the next year, to fully exploit the relation between miRNAs/UCRs and the mutation or chromosome aberrations related to their genomic location.

## SUPPLEMENTARY DATA

Supplementary data and programming codes are available at <http://web.unife.it/utenti/cristian.taccioli/software/>. The user has to rename the codes files into .pl format because of security reasons we have uploaded them as .txt.

Free web tools written for miRNA and UCRs microarray analysis can also be downloaded at <http://web.unife.it/utenti/cristian.taccioli/mix/examples.html>.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the support of all members of the Department of Molecular Virology, Immunology and Medical Genetics of The Ohio State University. We also appreciate the technical support of Melissa Dickman, Daniela Taccioli, Ivan Tassani, Stefan Costinean, Nicola Zanasi, Pierluigi Gasparini and Gianpiero Di Leva.

## FUNDING

Funding for open access charge: OSU Human Cancer Genetics Program.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
2. Griffiths-Jones, S. (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.
3. 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.
4. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
5. Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J. and Haussler, D. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, **441**, 87–90.
6. Calin, G.A., Liu, C.G., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M., Cimmino, A., Lee, E.J., Wojcik, S.E. et al. (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*, **12**, 215–229.
7. 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.
8. Ruvkun, G. (2001) Molecular biology. Glimpses of a tiny RNA world. *Science*, **294**, 797–799.
9. Cheng, A.M., Byrom, M.W., Shelton, J. and Ford, L.P. (2005) Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.*, **33**, 1290–1297.
10. Xu, P., Guo, M. and Hay, B.A. (2004) MicroRNAs and the regulation of cell death. *Trends Genet.*, **20**, 617–624.
11. Karp, X. and Ambros, V. (2005) Developmental biology. Encountering microRNAs in cell fate signaling. *Science*, **310**, 1288–1289.
12. Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83–86.
13. Poy, M.N., Eliasson, L., Krutzfeldt, 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.
14. Rajewsky, N. and Succi, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.*, **267**, 529–535.
15. Lai, E.C. (2002) Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.*, **30**, 363–364.
16. Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
17. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
18. Kurihara, Y. and Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA*, **101**, 12753–12758.
19. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
20. Nam, J.W., Kim, J., Kim, S.K. and Zhang, B.T. (2006) ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Res.*, **34**, W455–W458.
21. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. et al. (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA*, **99**, 15524–15529.
22. Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. et al. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA*, **103**, 2257–2261.
23. Calin, G.A., Cimmino, A., Fabbri, M., Ferracin, M., Wojcik, S.E., Shimizu, M., Taccioli, C., Zanasi, N., Garzon, R., Aqeilan, R.I. et al. (2008) MiR-15a and miR-16-1 cluster functions in human leukemia. *Proc. Natl Acad. Sci. USA*, **105**, 5166–5171.
24. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. et al. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.
25. Iorio, M.V., Visone, R., Di Leva, G., Donati, V., Petrocca, F., Casalini, P., Taccioli, C., Volinia, S., Liu, C.G., Alder, H. et al. (2007) MicroRNA signatures in human ovarian cancer. *Cancer Res.*, **67**, 8699–8707.
26. Visone, R., Pallante, P., Vecchione, A., Cirombella, R., Ferracin, M., Ferraro, A., Volinia, S., Coluzzi, S., Leone, V., Borbone, E. et al. (2007) Specific microRNAs are downregulated in human thyroid anaplastic carcinomas. *Oncogene*, **26**, 7590–7595.
27. Calin, G.A. and Croce, C.M. (2006) MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res.*, **66**, 7390–7394.
28. Hamosh, A., Scott, A.F., Amberger, J., Valle, D. and McKusick, V.A. (2000) Online Mendelian Inheritance in Man (OMIM). *Hum. Mutat.*, **15**, 57–61.
29. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.

30. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Edgar,R., Federhen,S. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36**, D13–D21.
31. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
32. Wingender,E., Dietze,P., Karas,H. and Knuppel,R. (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.*, **24**, 238–241.
33. Takai,D. and Jones,P.A. (2003) The CpG island searcher: a new WWW resource. *In Silico Biol.*, **3**, 235–240.
34. Zhi,D., Raphael,B.J., Price,A.L., Tang,H. and Pevzner,P.A. (2006) Identifying repeat domains in large genomes. *Genome Biol.*, **7**, R7.
35. 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.
36. Brazma,A. and Parkinson,H. (2006) ArrayExpress service for reviewers/editors of DNA microarray papers. *Nat. Biotechnol.*, **24**, 1321–1322.
37. Barrett,T., Suzek,T.O., Troup,D.B., Wilhite,S.E., Ngau,W.C., Ledoux,P., Rudnev,D., Lash,A.E., Fujibuchi,W. and Edgar,R. (2005) NCBI GEO: mining millions of expression profiles–database and tools. *Nucleic Acids Res.*, **33**, D562–D566.
38. Stenson,P.D., Ball,E.V., Mort,M., Phillips,A.D., Shiel,J.A., Thomas,N.S., Abeyasinghe,S., Krawczak,M. and Cooper,D.N. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.*, **21**, 577–581.
39. George,R.A., Smith,T.D., Callaghan,S., Hardman,L., Pierides,C., Horaitis,O., Wouters,M.A. and Cotton,R.G. (2008) General mutation databases: analysis and review. *J. Med. Genet.*, **45**, 65–70.