Research Paper

Volume : 2 | Issue : 8 | August 2013 • ISSN No 2277 - 8179

Bioinformatics Analysis of Mirna Data for Potential Biomarker Discovery

Biotechnology

KEYWORDS : microRNA (miRNA), biomarker, Gomir, gene

B. Veena

Department of Biotechnology, The Oxford College of Engineering, Bommanahalli, Hosur road, Bangalore-560068, Karnataka state, India.

ABSTRACT

miRNA are small and non coding RNA molecules about 22 nucleotides which play an important role in regulation of gene expression. miRNAs can be one type of potential molecular biomarkers. There is plenty of gene expression data available for many normal and disease conditions. A novel meta analysis approach developed by Shodhaka has been promising. This approach can be combined with the available miRNA data to identify potential novel markers. But the existing miRNA data are scattered across more than 150 web resources, it is not possible to select suitable one among them. In this project, the stand-alone miRNA tools were listed, down-loaded and compared. Potential novel biomarker(s) were identified for a specific condition. The results show that, among about 40 stand-alone tools only one tool (Gomir) is functional, easy to use and compatible with the available system. The procedure to use this tool was standardized. There were 6 different databases within this tool which were also compared. Among these databases, miRanda and RNAhybrid database give more number of predicted targets. All the union miRNA were selected for analysis with different database. There were twenty seven databases, selected by screening post-2010 publications, which can queried with a miRNA. From the comparison result miR-449, miR-380-5p, miR-368, miR-518a, miR-146b and miR-454-3p were identified as novel potential biomarker for testis tissue.

Introduction:

MicroRNAs (miRNAs) are short and non coding RNA molecule about 22 nucleotide RNA molecules that function as negative regulators of gene expression in eukaryotic organisms. RNA molecules found in gene silencing and the pathways have essential roles in development, cell differentiation, cell proliferation, cell death, and chromosome structure and virus resistance. The study shows, from few years have demonstrated that there is altered expression of miRNA genes in many human malignancies [1]. Biomarker is usually a molecule that has a unique association with a specific tissue and condition in terms of its quantitative expression.

A biomarker is a parameter that can be used to measure the progress of disease or the effects of treatment. Biomarkers help in early diagnosis, disease prevention, drug target identification, drug response etc. Several biomarkers have been identified for many diseases which found to be an effective and acceptable.

Usually markers can be identified based on single or multiple samples, aim are to identify the markers by tools and databases. There were several databases and tools which are scattered across.

Testis is a male gonad in animals. Testicular neoplasms are the most common sex cord-stromal tumors and comprise 1-3% of all testicular neoplasms. This tumor is always benign in children and approximately 90% are benign in adults [2].

Materials and methods:

All the tools were listed in (www.startbioinfo.com). Our team was the first team to collect all the information about the tools and databases of the miRNA. The existing miRNA data are scattered across more than 150 web resources, it is possible to select suitable among them. Stand-alone miRNA tools were downloaded and compared. All the tools listed in **Table 1.**

Among 40 stand-alone downloadable tools only GOmir tool [3] is functional, easy to use and compatible with the available system. The procedure to use this tool was standardized.

GENE DATA SET:

The team currently working on this project at Shodhaka developed their own method of selection. They first collected all available screened microarray data from two major databases – GEO and Array. They focused mainly on those studies in which the gene expression was studied in more than one tissue to increase reliability. Thus, they were able to collect data from 40 different tissues on analyzing 12 microarray studies. To identify and categorize the genes and their expression, an in-house scoring program was followed based on three main categories:

1. The genes present in 1 tissue but absent in remaining 39 tissues,
2. The genes present in all tissues with the exception of 1 tissue.
3. The genes present in few tissues and absent in few tissues.

They were able to arrive at a list of 700 odd genes, found to be exclusively present in only testis tissue; and 4000 ubiquitous genes. A reliability score was assigned to each gene based on how many studies validated the existence of the gene in the tissue. The top 200 highest scored genes from both lists were selected as the final list of genes. This whole process was achieved over a period of 5 years.

SELECTION OF DATABASE:

There were three types of selection were followed to select the databases relevant to our needs **Table 2**.

a) Based on the Usage Frequency: This is based on how many times the database has been cited or referred to. The reliability of the resource is high if it has been referred to by scientists the most. The top 37 databases were chosen (Table 3.4).

b) Based on Year of Publication: Some new resources may have been developed in recent years which may be very useful and reliable but since they are relatively new, they may not be cited as often. To avoid missing out any such databases, all databases published in 2011 or later were considered as well. This came up to a total of 54 databases (Table 3.5).

c) Based on Unique Feature: A database specific to some unique application such as used only to find miRNAs in human promoter region, or databases specific to miRNAs found in types of cancers, etc. may not be cited often nor be a fairly new database. Hence to avoid missing out such databases, the remaining 183 databases were screened individually to discover if they had any unique application. 12 such databases were found (Table 3.6).

From the **Table 2** there were twenty seven databases, selected by screening post-2010 publications, which can queried with a miRNA.

To identify the biomarker, microRNA targets are calculated with the databases.

Scores are calculated by using the formula:

Score= A+B+C....

From the comparison result lowest microRNA targets from the databases chosen as a biomarker.

RESULT AND DISCUSSION:

The GOmir is a stand-alone application to study miRNA interaction and targets. It consists of two separate tools like Jtarget and TAGGO. Jtarget integrates miRNA target prediction and functional analysis where as TAGGO application designed for the gene ontology annotation (GO) [3].

JTarget combines the data from six different prediction databases like miRanda, RNAhybrid, PicTar4Way, PicTar5Way, TargetScan and TarBase. TAGGO is for gene ontology (GO) resources of a gene product organism like plants, animals and microbial genomes. So we are concentrating on the JTarget application [3].

The JTarget main goal is to find the miRNA with respect to the given gene from the six different databases. We can obtain the resulting target genes shown below.

- One database only
- Two or more databases
- Combined databases

miRanda Comparative Analysis:

miRanda was also inserted in the JTarget comparative module of GOmir. miRanda gives the predicated targets of the miRNA and also provides information for the validated result. In order to understand the features of miRNA predicated targets performed comparative analysis using other predicated target databases like RNAhybrid, TarBase, PicTar4Way, TargetScan and PicTar5Way with the miRanda database. From the analysis between each database calculated the targets of the miRNA from this analysis got more number of predicates targets for miRanda database. Among these databases, miRanda and RNAhybrid database give more number of predicted targets showed in **Figure 1**.

COLLECTION OF DATABASE:

As told before to collect the databases used three methods shown in **Table 2** are

- a) Based on the Usage Frequency
- b) Based on Year of Publication
- c) Based on Unique Feature

There were twenty seven databases, selected by screening post-2010 publications, which can queried with a miRNA. From the analysis of twenty seven databases only ten databases were selected for target prediction.

Databases are selected in such a way that they belong to the homo-species and should query with a single miRNA. Reason to reject remaining databases is because those databases don't belong to the homo-species and belong to other species, some databases were not working properly and some databases were giving same result to all miRNA.

As shown in the **Figure 2** all the union miRNAs were selected to get a target for miRNA. Got predicted targets from the databases, various databases give different numbers of targets. All most 500 union miRNA were tried with the databases, all the databases gave predicted targets.

All the union miRNA were selected for analysis with different database. There were twenty seven databases, selected by screening post-2010 publications, which can queried with a miRNA. From the **Figure 3** comparison result shows **miR-449, miR-380-5p, miR-368, miR-518a, miR-146b and miR-454-3p** were identified as novel potential biomarker for testis tissue.

CONCLUSION:

MicroRNAs have increasingly been recognized as significant agent to regulate gene silencing and pathways, which have essential roles in development, cell differentiation, cell proliferation, cell death, and chromosome structure and virus resistance. So, miRNAs can be used as a potential tool not only for the understanding of cancer growth and progression, but also as potential cancer biomarkers.

The results show that, only one tool (Gomir) is functional, easy to use and compatible with the available system. We have also compared the existing databases on which Gomir tool is working, and our result shows that miRanda and RNAhybrid database give more number of predicted targets among 6 other databases.

Finally, after comparing the result of all 27 existing databases and Gomir tool we have identified miR-449, miR-380-5p, miR-368, miR-518a, miR-146b and miR-454-3p as novel potential biomarkers for testis tissue.

The new microRNA markers are likely to contribute to the improved early diagnosis, patients, prognosis, or anti-cancer therapy.

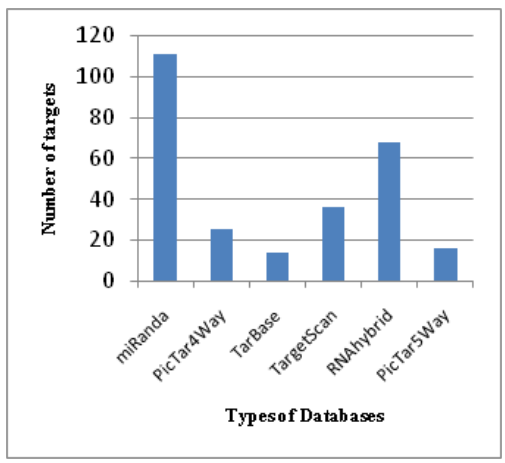

Figure 1: Intersection of each database

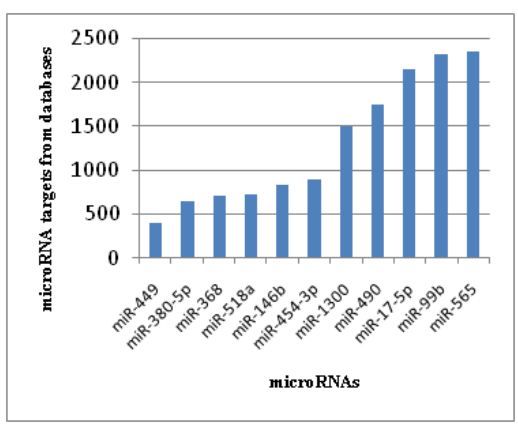

Figure 2: Comparison of all the microRNAs

RNAhybrid	miRanda	miRanda	miRanda	miRanda	miRanda	TargetScan	miRanda	RNAhybrid
LDHC	PRM1	BRDT	ADAM2	MSMB	TULP2	AC5BG2	FLSCR2	PDHA2
miR-130a	miR-1256	miR-136	miR-198	miR-128	miR-421	miR-361/361	miR-188-3p	miR-106a
miR-130b	miR-1285	miR-181b	miR-519b-5p	miR-27a		miR-421	miR-340	miR-130
miR-27a	miR-1305	miR-181d	miR-519c-5p	miR-27b			miR-532-3p	miR-143
miR-27b	miR-135a	miR-328a		miR-655			miR-664	miR-146a
miR-370	miR-135b	miR-328b						miR-146b
miR-449	miR-31	miR-328c						miR-17-5p
miR-449c	miR-34a	miR-328d						miR-20a
miR-454-3p	miR-484	miR-338-5p						miR-215
miR-485-3p	miR-543	miR-452						miR-27a
miR-511	miR-576-3p	miR-509-3p						miR-302d
miR-519e*	miR-940	miR-548a-3p						miR-368
miR-581		miR-548e						miR-380-5p
miR-595		miR-548f						miR-485-5p
miR-603		miR-548g						miR-520b
miR-622		miR-586						miR-587
miR-629		miR-664						miR-591
		miR-944						miR-595
								miR-623
								miR-652
								miR-99c

Figure 3: Total number of targets from the databases

Volume : 2 | Issue : 8 | August 2013 • ISSN No 2277 - 8179

Table 1: List of all down-loadable tools

Expander	GenMIR++	mirExplorer	RNAplex
IntaRNA	Gomir	miRExpress	RNAVLab
miReduce	HHMMiR	miRFam	Seqbuster
miRiam	MicroPred	MiRFinder	SigTerms
NovoMIR	MicroTar	Mermaid	SNMNMF
Triplet-SVM-classifier	miPred	miRNAkey	SplamiR
Bi-targeting	MIR	miRTRAP	SV micro
CorNa	MIRAGAA	Musmirus	yasMIR
Expmicro	Mire	Tool by Xiao J et al	minRcos
FASTH	MIReNA	RNAmicro	SBM-Stacking Binding Matrix

Table 2: Databases list

Table 3.4 : Usage Frequency	Table 3.5 : Based on Year of Publication		Table 3.6 : Based on Unique Features
Babelomics	MsigDB 3.0	miRTour	HMDD
CrossLink	Anta gomir base	miRTrail	CAMI-Finder
DBTSS	DIANA-microT-ANN	miRver	CircuitsDB
EXCERBT	doRiNA	miRvestigator	dbDEMC
Expander	DPO RE-miRNA	miRWalk	FindTar3Online Prediction
FetiGo	EXCERBT	miTALos	Genecopoeia
fRNAdb	Biomedical Text	MultiMiTar	GeneHub-GEPIS
GeneCodis	GARNET	myMIR site	microPIR
MicroCosom	G-DOC	NRDR	mimiRNA
MicroInspector	HOCTAR	PACMIT	miRCancer
microRNA.org	InMiR	PlantMiRNAPred	S-MED
miR2Disease Base	IntmiR	PmmR	TargetRank
miRanda	Lex f senescence database	psRNATarget	
miRBase	MeturePred	RNAimmuno	
miRDB	mESAdb	SNMNMF	
miRecords	microRNA body map	SoMART	
miRGen	MIR@NT@N	SplamiR	
miRNAMap	mirACT	Starbase	
MiRror	miR-ATI	Tool by Xiao Jet al	
miRScan	miRBase		
miRStart	mirConnx		
mirSVR	mirDIP		
MirTarget(I, II)	miR ds NP		
MirZ	MiREE		
MsigDB	MireNVIRONMENT		
PicTar	mirEX		
PITA	mirExplorer		
RNA22	miRFam		
RNAdb	miRNEST		
RNAhybrid	MiRPare		
RNAz	miRstart		
smime DB	miRT		
TarBase	miRTar		
TargetScan	miRTarBase		
Weblogo	MIRTFnet		

REFERENCE

1. Marek Mraz. "Biological Role of microRNAs in Animal Cells and Development and Cancer". | 2. Sönmez N, Ton O, Arısan S, Kılınç F, Eken K, Güney S. "Bilateral Leydig cell tumor of the testis: a case report". | 3. Maria G Roubelakis, Pantelis Zotos, Georgios Papachristoudis, Ioannis Michalopoulos, Kalliopi I Pappa, Nicholas P Anagnou1, and Sophia Kossida. "Human microRNA target analysis and gene ontology clustering by GOmir, a novel stand-alone application". | 4. Pranidhi Sood, Azra Krek, Mihaela Zavolan, Giuseppe Macino, and Nikolaus Rajewsky "Cell-type-specific signatures of microRNAs on target mRNA expression" | 5. Sabah Kadri, Veronica Hinman and Panayiotis V Benos. "HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models" | 6. Rahul Thadani1 and Martti T Tammi "MicroTar: predicting microRNA targets from RNA duplexes" | 7. Hertel J, Stadler PF. "Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data". | 8. Tafer H, Hofacker IL (2008 Nov 15). "RNAplex: a fast tool for RNA-RNA interaction search". | 9. Kawasaki H, Moriyama M, Ohtani Y, Naitoh M, Tanaka A, Nariuchi H (1989 Oct;57) "Analysis of endotoxin fever in rabbits by using a monoclonal antibody to tumor necrosis factor (cachectin)". | 10. Christoph J. Thieme, Lydia Gramzow, Dajana Lobbes and Günter Theißen. "SplamiR – Prediction of spliced miRNAs in plants" | 11. Veksler-Lublinsky I, Shemer-Avni Y, Kedem K, Ziv-Ukelson M (2010 May 13). "Gene bi-targeting by viral and human miRNAs" PMID: 20465802 | 12. Batuwita R, Palade V (2009 Apr 15). "microPred: effective classification of pre-miRNAs for human miRNA gene prediction". | 13. Mathelier A, Carbone A (2010 Sep 15) "MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data". | 14. Creighton CJ, Nagaraja AK, Hanash SM, Matzuk MM, Gunaratne PH (2008 Nov 14). "A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions"