ViTa: prediction of host microRNAs targets on viruses

Paul Wei-Che Hsu[1], Li-Zen Lin[1], Sheng-Da Hsu[1], Justin Bo-Kai Hsu[1] and Hsien-Da Huang[1,2,3,*]

[1]Institute of Bioinformatics, [2]Department of Biological Science and Technology and [3]Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsin-Chu 300, Taiwan, ROC

ABSTRACT

MicroRNAs (miRNAs) are involved in various biological processes by suppressing gene expression. A recent work has indicated that host miRNAs are also capable of regulating viral gene expression by targeting the virus genomes. To investigate regulatory relationships between host miRNAs and related viruses, we present a novel database, namely ViTa, to curate the known virus miRNA genes and the known/putative target sites of human, mice, rat and chicken miRNAs. Known miRNAs are obtained from miRBase. Virus data are collected and referred from ICTVdB, VBRC and VirGen. Experimentally validated miRNA targets on viruses were derived from literatures. Then, miRanda and TargetScan are utilized to predict miRNA targets within virus genomes. ViTa also provides the virus annotations, virus-infected tissues and tissue specificity of host miRNAs. This work also facilitates the comparisons between subtypes of viruses, such as influenza viruses, human liver viruses and the conserved regions between viruses. Both textual and graphical web interfaces are provided to facilitate the data retrieves in the ViTa database. The database is now freely available at http://vita.mbc.nctu.edu.tw/.

INTRODUCTION

MicroRNAs (miRNAs) are small RNA molecules, ∼22 nt, that can regulate gene expression by interfering in the post-transcriptional level, which results in the degradation of mRNAs and repression of translation by the pairing of bases to the 3′-untranslated regions (3′-UTRs) of the mRNAs. The miRNAs are derived from hairpin-like precursor transcripts (pre-miRNAs) that are ∼70–120 nt long sequences. These structures are believed to be recognized and taken out of nucleus by exportin 5. Pre-miRNA is then cleaved by Dicer (a ribonuclease III enzyme) to excise the mature miRNAs in the form of siRNA-like duplexes and asymmetrical assembling of the mature miRNA strands, which may be decided upon relative thermodynamic characteristics of the two 5′ termini of strands, combining with the Argonaute proteins into effector complexes (1).

Recent studies show that some miRNAs can target the viral RNA transcripts after viruses enter into cells, and those host miRNAs appear to play varied roles in affecting viruses. For example, human liver-specific miR-122a induces hepatitis C virus (HCV) replication by targeting to the 5′-non-coding region (NCR) (2), and HIV is suppressed by targeting several genes by human miRNAs (3). These indicate that human miRNAs can regulate the life cycle of viruses in hosts. Figure 1 depicts the concept of possible regulatory relationship between host miRNAs and viruses. The interaction between viruses and host miRNAs is worthy of further research (more description). For example, miRNAs may contribute to cancer, miRNA-mediated tumorigenesis results from either down- or up-regulating virus activity.

Compared with related works in animals and plants, research investigations on miRNAs and viruses are in its initial stages. HCV is the major causative viral agent of cirrhosis and hepatocarcinoma. Analysis of the genetic interaction between human miR-122 and the 5′-NCR of the HCV genome via the mutation of the predicted miRNA target site and ectopic expression of miR-122a molecules containing compensatory mutations indicated that human miR-122a is likely to facilitate the replication of the viral RNA, and may present a target for antiviral intervention (2).

Another HIV is also predicted as miRNA targets (3). Human miRNAs that are expressed in T-cells, the natural site of infection by HIV-1 infection, may target HIV-1 genes and related clade sequences at highly conserved target sites. These studies imply that human miRNAs potentially affect the expression of HIV-1 genes and could be utilized in developing therapies to inhibit HIV-1 in future. Lecellier *et al.* (4) demonstrated that cellular miRNAs effectively restrict the accumulation of retrovirus primate foamy virus type 1 (PFV-1) in human cells. Conversely, PFV-1 also encoded a protein named Tas, which suppresses miRNA-directed functions in mammalian cells and displays cross-kingdom anti-silencing activities. Therefore, through fortuitous recognition of foreign nucleic acids, cellular miRNAs have direct antiviral effects in addition to their regulatory functions (4). The various relationships between viruses and miRNAs are worthy of further investigation.

*To whom correspondence should be addressed. Tel: +886 3 5712121, ext. 56952; Fax: +886 3 5729288; Email: bryan@mail.nctu.edu.tw
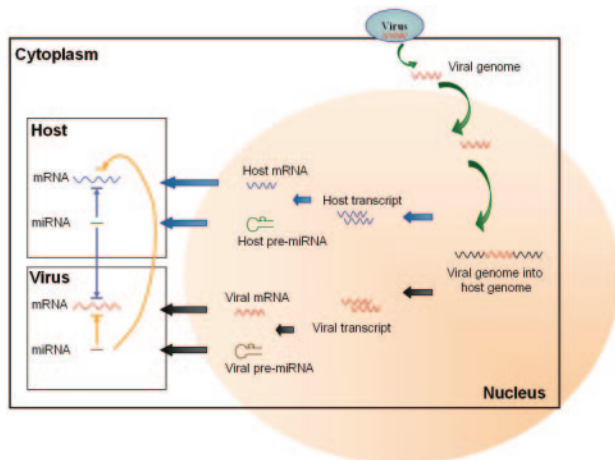
Figure 1. MiRNAs interact with target mRNA in viruses and host cells.

Figure 2. The ViTa flowchart.

MiRNAs are increasingly being shown to play important roles in various biological processes. The miRBase (5) supports the information for published miRNA genes and miRNA targets. miRNAMap (6) is a comprehensive information repository for the miRNAs and their targets in human, mouse, rat and dog genomes. Furthermore, several miRNA target prediction programs were developed previously. TargetScan (7), miRanda (8) and RNAhybrid (9) are commonly used to determine the energetically most favorable hybridization sites of miRNA target prediction. Lu *et al.* (10) developed an miRNA microarray for measuring the expression profiles of all known miRNAs in various normal tissues and tumors, which provides valuable information for accurately predicting miRNA targets on viruses.

To facilitate the research on the roles of host (human, mice, rat and chicken) miRNAs in viral infection, this work develops a database, namely ViTa, which comprises known miRNAs on host, known and putative host miRNA targets on viruses, and relationship among viruses, miRNAs and diseases.

After collecting data, sequences are preprocessed upon life cycle of viruses. Then, the miRNA targets within virus genomes are predicted by miRanda and TargetScan. Furthermore, similarity among sequences is compared with subtypes in every species to find the conserved regions. Finally, the relationships between miRNAs and targeted viruses are identified and integrated by ViTa database, which also contains annotation of viruses, virus-infected tissues and the tissue specificity of miRNAs, conserved regions on virus and statistical results for user. Thus, the ViTa database not only provides miRNAs and miRNA targets, but also expression profiles of known host miRNAs, virus annotation and crosslinks to other biological databases, and the comparison between viruses, such as influenza viruses, hepatitis B virus, HCV and combination of miRNAs.

This work establishes the computational relationship between the host miRNA genes and viruses. Furthermore, this work provides effective annotations, including those for human miRNA expression and virus-infected tissues and annotations of viruses and comparisons between viruses. Additionally, various functions and a graphical web interface are presented that help users in investigating the miRNA roles in viruses.
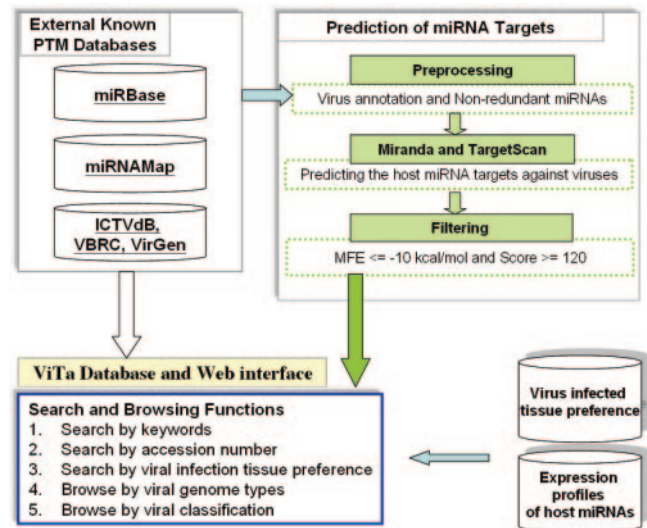
DATA GENERATION

Figure 2 presents the data generation flow of the ViTa database. The flowchart comprises the following three main steps: (i) collect data from databases and the literature; (ii) mine information from collective data and handle sequences; and (iii) predict host miRNA targets on viruses. Each step is described in detail below.

Data collection

Known human, mouse, rat and chicken miRNAs are obtained from miRBase (Release 8.2) (11) and known virus miRNA targets are collected from the literature.

Collecting virus data are difficult as NCBI releases virus genomes only for download (except influenza viruses), and no classifiable, identical virus data packaged exists that can be directly obtained via the FTP site. Additional data were obtained from other databases that are linked to NCBI. Most data are from databases—ICTVdB, VBRC and VirGen—and surfing websites. The ViTa database contains 26 870 virus genomes for 2108 species, including genomes and transcripts. For human-related viruses, there are 15 357 records for 446 species.

The ViTa database currently contains 23 virus families that are classified into six genome types—dsDNA viruses, ssDNA viruses, dsRNA viruses, (+) ssRNA viruses, (−) ssRNA viruses and retro-transcribing virus. The virus genomes were obtained from NCBI, ICTVdB, VBRC and VirGen. Experimentally validated miRNA targets on viruses and the information of virus-infected tissues were extracted by surveying literature.

Expression profiles of miRNAs were obtained from the dataset constructed by Lu *et al.* (10).

After finishing data collection, the second step involves mining information from varied datasets. Because virus data are non-identical, i.e. virus name, strain, serotype, hosts and virus-infected tissue are lacking. To resolve this problem, other virus databases and related literature are accessed. First, the virus names are altered to unite in all datasets. The ViTa

website provides two classifications: genome type and biological classification (order, family, subfamily, genus and species) to facilitate user research. To identify hosts and virus-infected tissues, various sources, such as databases and literatures are accessed. Viral RNA genomes and viral mRNAs are prepared to predict miRNA targets, those RNA sequences were classified as mRNA, positive-strand RNA and negative-strand RNA according to the viral genome types.

Predicting host miRNA targets on viruses

In this work, miRanda (8) and TargetScan (7) were applied to identify the host miRNA target sites in viruses. The minimum free energy (MFE) of the miRNA–target duplex was determined in predicting the miRNA target sites. Low-MFE values of the miRNAs and the target sites reveal the energetically more probable hybridizations between the miRNAs and the target genes. 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. Thus, the predictive parameters, including the cutoff of miRanda MFE and the cutoff of miRanda score, can be adjusted for the miRNA target prediction.

The regulatory relationships between the host miRNAs and the related viruses are determined with reference to the annotations and the organization of viruses. This work supports the expression profiles of known host miRNAs, the cross-species virus comparisons, annotations and various links to biological databases. The conserved regions in each species in the database were obtained using BLAST (12). Supplementary Table S1 gives the list of integrated databases. The expression profiles of the miRNAs, which were obtained from the dataset constructed by Lu *et al.* (10), are useful in elucidating the regulatory roles of miRNAs. Moreover, the tissue specificity of known miRNAs is used to examine the relationship between host miRNAs and viruses, both of which exist in the same tissue.

DATABASE STATISTICS

Table 1 gives the statistics of virus genomes collected in the proposed database. The MFE of the miRNA/target duplex was determined by miRanda and TargetScan to predict miRNA target sites. Low-MFE values of the miRNA/target duplex indicate likely hybridizations between the miRNAs and the targeted viruses. Table 2 presents statistics for the known human miRNA targets on virus, which can also be predicted by both miRanda and TargetScan except that hsa-miR-32/PFV-1 cannot be predicted by TargetScan. The known miRNA targets on HCV can be identified when the miRanda MFE threshold and miRanda score threshold are set at −10 kcal/mol and 120, respectively. However, these parameters are likely to grossly overpredict the number of miRNA targets on viruses.

Alternatively, the proposed database allows users to utilize a set of predictive parameters that is more stringent and obtains fewer false positives than thresholds. All predictive miRNA targets, of which the MFE is <−10 kcal/mol and the score is >120, are determined and stored in ViTa database. Table 3 gives the statistics of miRNA targets on viruses with different predictive parameters. For instance, if the MFE cutoff is set at −20 kcal/mol and score cutoff is set at 120, the average number of targeted viruses for each miRNA is 83.65 and the average number of distinct miRNAs targeting to each virus is 64.50. Similarly, when the MFE cutoff is set to −10 kcal/mol and score cutoff is set at 160, the average number of the targeted viruses for each miRNA is 50.64 and the average number of distinct miRNAs targeting each virus is 33.75. Furthermore, Supplementary Table S4 presents the statistics of miRNA targets predicted using TargetScan. Supplementary Table S5 presents the statistics for miRNA targets on several liver-specific viruses. Supplementary - Table S6 presents the statistics for human miRNA targets on influenza viruses.

INTERFACE

The proposed web interface provides variety of data for both browsing functions and search functions. The database provides two ways for browsing the miRNA targets on viruses: (i) browse by genome types and (ii) browse by taxonomy. It allows users to search the database by keywords, accession numbers, disease names and virus-infected tissue preference (Supplementary Figure S2). Supplementary Figure S1 demonstrates the conserved regions between virus species,

Table 1. Statistics of virus genome collected in ViTa

	Host Human	Mouse	Rat	Avian	Other	Total
Liver-related viruses	285					
Influenza viruses	12 137	414	37	2529	10 333	28 670
Other disease virus	3135					
	15 357					

Table 2. Known host miRNAs targets on viruses predicted by miRanda and TargetScan

Human miRNAs	Target virus	References	By miRanda Score	MFE (kcal/mol)	By TargetScan MFE (kcal/mol)
hsa-miR-122a	HCV 1a strain H77c (AF011751)	(2)	146.00	−14.30	−20.10
has-miR-122a	HCV 1b strain HCV-N (AF139594)	(2)	148.00	−15.70	−24.10
					−20.10
has-miR-29a	HIV-1 isolate BRU (K02013)	(3)	175.00	−22.10	−29.20
has-miR-29b	HIV-1 isolate BRU (K02013)	(3)	179.00	−23.00	−31.30
has-miR-149	HIV-1 isolate BRU (K02013)	(3)	194.00	−26.60	−32.10
has-miR-324-5p	HIV-1 isolate BRU (K02013)	(3)	191.00	−24.50	−32.70
has-miR-378	HIV-1 isolate BRU (K02013)	(3)	177.00	−25.30	−31.10
has-miR-378	HIV-1 isolate ELI (K03454)	(3)	177.00	−25.30	−31.10
has-miR-32	PFV-1	(4)	120.00	−9.91	—

detailed information about targeted viruses and the tissue specificity of the miRNA targets.

Figure 3 presents a visualization tool that is used to present the miRNA targeting the viruses. All miRNAs target sites are also provided in the text format, including the genomic locations of target sites, the MFE of miRNA/target duplex, target site sequences and the alignment of hybridization structures. The detailed documentation for the usage of ViTa is provided on the web interface.

Table 3. Statistics for human miRNA targets on viruses predicted by miRanda

MFE cutoff (kcal/mol)	Average no. of targeted virus for each miRNA	Average no. of distinct miRNAs for each targeted virus	Average no. of target sites for microRNA
Score ≥ 120			
≤−10	229.13	260.93	33 754.86
≤−15	188.96	188.47	11 693.65
≤−20	83.65	64.50	1755.98
≤−25	19.13	9.66	182.86
Score ≥ 140			
≤−10	170.40	156.92	6174.33
≤−15	132.66	113.22	3524.48
≤−20	56.19	38.77	819.14
≤−25	14.06	6.63	117.03
Score ≥ 160			
≤−10	50.64	33.75	620.39
≤−15	45.21	29.52	537.10
≤−20	22.37	12.88	213.28
≤−25	8.41	3.29	61.24
Score ≥ 180			
≤−10	5.81	2.03	44.48
≤−15	5.81	2.03	44.48
≤−20	5.53	1.93	37.68
≤−25	3.91	1.49	28.58
Score ≥ 200			
≤−10	3	1	6.33
≤−15	3	1	6.33
≤−20	3	1	6.33
≤−25	3	1	6.33

CONCLUSIONS

This work presents a novel database for host miRNA targets on virus genomes and virus transcripts. The database comprises known host miRNAs, known viral miRNAs, known host miRNA targets on viruses and putative host miRNA targets on viruses. We believe that the proposed resource can provide sufficient and effective information for the investigation about the interaction between host miRNAs and viral genes. The information about the tissue preferences of viruses is effective to combine the tissue specificity of host miRNAs for further analysis of miRNA targets on viruses in ViTa. For instance, HCV prefers to infect the liver and miR-122a is a liver-specific miRNA. Thus, the miR-122a targets discovered in HCV are significantly interesting.
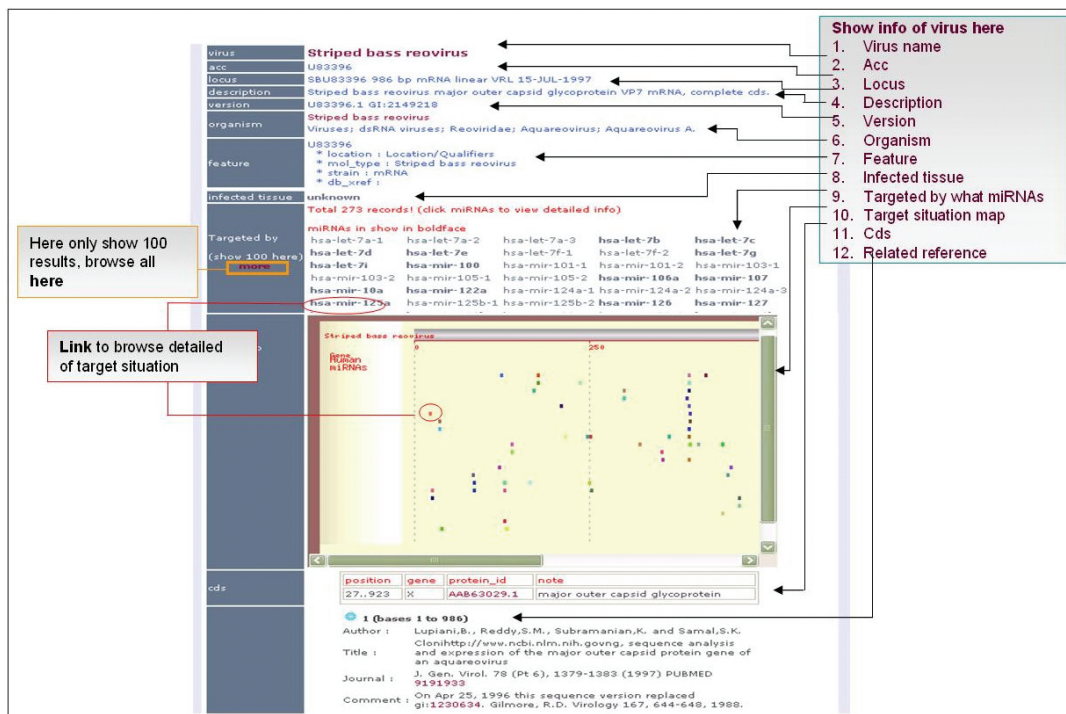
The prospective works of the database are as follows: (i) more expression profiles for host miRNA genes will be supported; (ii) combinatorial miRNA regulation to viral gene expression can be identified by computational analysis and miRNA-related gene expression profiles; and (iii) the association between virus and disease will be explored by clinical data collection.

AVAILABILITY

The ViTa database is continuously maintained and updated. The database is now freely available at http://vita.mbc.nctu. edu.tw.

Figure 3. Interface of putative targeted viruses in ViTa.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Jopling,C.L., Yi,M., Lancaster,A.M., Lemon,S.M. and Sarnow,P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*, **309**, 1577–1581.
3. Hariharan,M., Scaria,V., Pillai,B. and Brahmachari,S.K. (2005) Targets for human encoded microRNAs in HIV genes. *Biochem. Biophys. Res. Commun.*, **337**, 1214–1218.
4. Lecellier,C.H., Dunoyer,P., Arar,K., Lehmann-Che,J., Eyquem,S., Himber,C., Saib,A. and Voinnet,O. (2005) A cellular microRNA mediates antiviral defense in human cells. *Science*, **308**, 557–560.
5. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
6. Hsu,P.W., Huang,H.D., Hsu,S.D., Lin,L.Z., Tsou,A.P., Tseng,C.P., Stadler,P.F., Washietl,S. and Hofacker,I.L. (2006) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res.*, **34**, D135–D139.
7. 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.
8. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
9. Rehmsmeier,M., Steffen,P., Hochsmann,M. and Giegerich,R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
10. 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.
11. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
12. 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.