ViTa: a database of host microRNA targets on viruses

Paul Wei-Che Hsu (徐唯哲) [1, †], Li-Zen Lin (林立人) [1, †], Sheng-Da Hsu (許勝達)[1, †], Hsien-Da Huang (黃憲達)[1, 2, 3, *]

[1]*Institute of Bioinformatics, National Chiao Tung University, Hsin-Chu 300, Taiwan*
[2]*Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu 300, Taiwan*
[3]*Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsin-Chu 300, Taiwan*
[*]bryan@mail.nctu.edu.tw

1. ABSTRACT

MicroRNAs (miRNAs) are involved in various biological processes by suppressing gene expression. Recent work has indicated that host-miRNAs also regulate viral gene expression by targeting to virus genomes. To investigate the regulatory relationships between the 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 is collected from ICTVdB, VBRC and VirGen. Experimentally validated miRNA targets on viruses were extracted from literatures. Then, miRanda and TargetScan are utilized to computationally predict the miRNA targets within virus genomes. ViTa also provides the virus annotations, virus infected tissue preferences 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 ViTa database. The database is now freely available at http://vita.mbc.nctu.edu.tw/.

2. INTRODUCTION

MicroRNAs (miRNAs) are small RNA molecules approximately ~22 nts sequences that regulate gene expression by interfering the post-transcriptional level, resulting in degradation of mRNAs and repression of translation by the base-pairing to the 3' untranslated regions (3'-UTR) of mRNAs. The miRNAs are derived from hairpin-like precursor transcripts (pre-miRNAs) approximately 70–120 nts long sequences. These structures are recognized and exported 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 assembly 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 to viral transcripts after viruses enter into the cells, and those host-miRNAs appear to play varied roles in affecting the activities of viruses. For example, human liver-specific miR-122 induces Hepatitis C Virus (HCV) replication by targeting to the 5'-non-coding region (NCR) (2), and HIV is suppressed by human miRNAs, which targets to HIV gene transcripts (3). These indicate that human miRNAs can regulate the life cycle of viruses in hosts. The interaction between viruses and host-miRNAs is worthy for investigation. Figure 1 depicts the conceptual diagram showing the possible regulatory relationships between host-miRNAs and viruses.

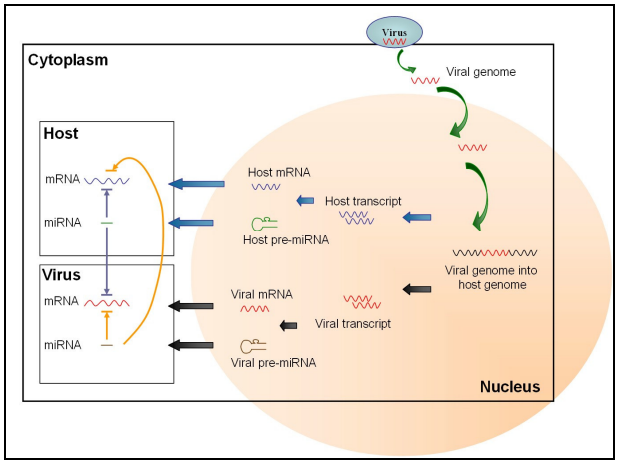

Figure 1. MiRNAs interact with target mRNA in viruses and host cells.

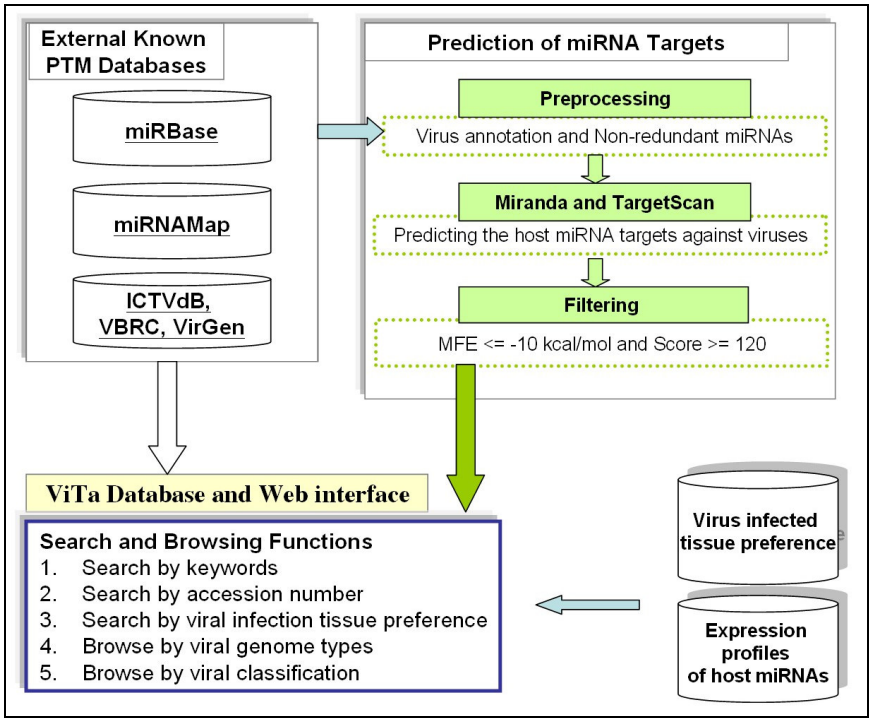

Figure 2. The ViTa flowchart.

HCV is the major causative viral agent of cirrhosis and hepatocarcinoma (HCC). Previous study investigated the interaction between human miR-122 and the 5'-non-coding region of the HCV genome by the mutations of the predicted miRNA target sites and ectopic expression of miR-122 molecules containing compensatory mutations. Their results indicate that human miR-122a is likely to induce the replication of the viral RNA of HCV, and may present a target for antiviral intervention (2).

A variety of human miRNAs expressed in T-cells were identified previously for HIV-1 infection. These miRNAs may target HIV-1 genes at highly conserved target sites (3). The results imply that human miRNAs potentially affect expression of HIV-1 genes and could be utilized in developing therapies to inhibit HIV-1 in the future. Lecellier *et al* (4) demonstrated that a cellular miRNAs, has-miR-32, effectively restricts the accumulation of the retrovirus primate foamy virus type 1 (PFV-1) in human cells. Through fortuitous recognition of foreign nucleic acids, cellular miRNAs have direct antiviral effects in addition to their regulatory functions (4).

The miRBase (5) supports the information for published miRNA genes. 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 targets prediction. Lu *et al.* developed an miRNA microarray for measuring the expression profiles of known miRNAs in various normal tissues and tumors (10), which provides valuable information for predicting miRNAs targets on viruses.

In order to facilitate the research in the association between host-miRNAs and viruses, this work develops a database, namely ViTa, which comprises known host-miRNAs, host-miRNA targets on viruses and known viral miRNAs. The database focuses on the regulatory relationship between the host-miRNAs and viruses. Furthermore, the database provides several effective annotations, including the tissue-specificity of human miRNAs, virus infected tissue preferences and the virus annotations. Additionally, both textual and graphical interfaces are presented to facilitate the data retrieves in ViTa..

3. DATA GENERATION

Figure 2 presents the data generation flow of the ViTa database. The ViTa comprises mainly three components, including the data collection and preprocessing, predicting miRNA targets on viruses by miRanda and TargetScan and filtering the miRNA targets. Each component is described in detail below.

3.1 Data collection

Table S1 (See Supplementary Materials) gives the list of the integrated databases. The ViTa database currently contains 23 virus families that are classified into six genome types, including dsDNA viruses, ssDNA viruses, dsRNA viruses, (+) ssRNA viruses, (-) ssRNA viruses and retro-transcribing virus. The virus

genomes were obtained from ICTVdB (11), VBRC (12) and VirGen (13). The ViTa database contains 26870 virus genomes from 2108 species. For human-related viruses, there are 15357 genomes from 446 species. Furthermore, the ViTa database categorizes the virus genomes by genome type and biological classification (order, family, subfamily, genus and species).

The known human, mouse and rat miRNAs are obtained from miRNAMap (6). Known viral miRNAs in six virus species, including Epstein Barr virus (EBV), Human cytomegalovirus (HCMV), Kaposi sarcoma-associated herpesvirus (KSHV), Mouse gammaherpesvirus 68 (MGHV68), Simian virus 40 (SV40) and Rhesus Lymphocryptovirus (rlcv), were obtained from miRBase (Release 8.2) (14). Chicken miRNAs are also obtained from miRBase. Experimentally validated miRNA targets on viruses and the information of virus infected tissue preferences were extracted by surveying literature.

3.2 Predicting host-miRNA targets on viruses

In this work, miRanda (8) and TargetScan (7) were applied to identify the host-miRNA target sites on viruses (Table S2). The Minimum Free Energy (MFE) of the miRNA/target duplex is calculated when predicting miRNA target sites. The lower MFE values of the miRNAs and the target sites reveal the energetically more probable hybridizations between the miRNAs and the target genes. However, these parameters are likely 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 (6). 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 potentially regulatory relationships between the host-miRNAs and the related viruses are determined in the proposed resource. Besides, this work supports the expression profiles of the known host-miRNAs, the cross-species virus comparisons, virus annotations and cross-links to other biological databases. The sequence conserved regions among virus species were identified by BLAST (15). The expression profiles of the host-miRNAs in human obtained from the data set constructed by Lu et al are useful in elucidating the tissue-specific host-miRNAs and the viruses.

Thus, the ViTa database not only provides host-miRNAs and host-miRNA targets, but also the comparison between viruses, such as influenza viruses, hepatitis B virus and hepatitis C virus.

4. DATABASE STATISTICS

The ViTa database currently contains 73 viral miRNAs in six species, such as EBV (23 entries), HCMV (11 entries), KSHV (13 entries), MGHV68 (9 entries), SV40 (1 entries) and RlCV (16 entries). Table 1 gives the statistics of virus genomes collected in the proposed database.

Table 1. Statistics of virus genome collected in ViTa.

	Host					Total
	Human	Mouse	Rat	Avian	Other	
Liver-related viruses	285	414	37	2,529	10,333	28,670
Influenza viruses	12,137					
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		By TargetScan
			Score	MFE (kcal/mol)	MFE (kcal/mol)
Hsa-miR-122a	HCV 1a strain H77c (AF011751)	(2)	146.00	-14.30	-20.10
Hsa-miR-122a	HCV 1b strain HCV-N (AF139594)	(2)	148.00	-15.70	-24.10
					-20.10
Hsa-miR-29a	HIV-1 isolate BRU (K02013)	(3)	175.00	-22.10	-29.20
Hsa-miR-29b	HIV-1 isolate BRU (K02013)	(3)	179.00	-23.00	-31.30
Hsa-miR-149	HIV-1 isolate BRU (K02013)	(3)	194.00	-26.60	-32.10
Hsa-miR-324-5p	HIV-1 isolate BRU (K02013)	(3)	191.00	-24.50	-32.70
Hsa-miR-378	HIV-1 isolate BRU (K02013)	(3)	177.00	-25.30	-31.10
Hsa-miR-378	HIV-1 isolate ELI (K03454)	(3)	177.00	-25.30	-31.10
Hsa-miR-32	PFV-1	(4)	120.00	-9.91	-

The minimum free energy (MFE) of the miRNA/target duplex was determined by miRanda and TargetScan for predicting miRNA target sites. Table 2 presents statistics for the known human miRNA targets on virus, which can also be predicted by both miRanda and TargetScan except hsa-miR-32/PFV-1 can not 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 predictive parameters are likely to grossly over-predict 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 to obtain fewer false positives that does the thresholds. All predictive miRNA targets, of which the MFE are less than -10 kcal/mol and the score are greater than 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 the distinct miRNAs targeting to each virus is 33.75. Furthermore, Table S4 (See Supplementary Materials) presents the statistics of miRNA targets predicted by TargetScan. Table S5 presents the statistics for miRNA targets on several liver-specific viruses. Table S6 presents the statistics for human miRNA targets on influenza viruses.

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	33754.86
≤ -15	188.96	188.47	11693.65
≤ -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

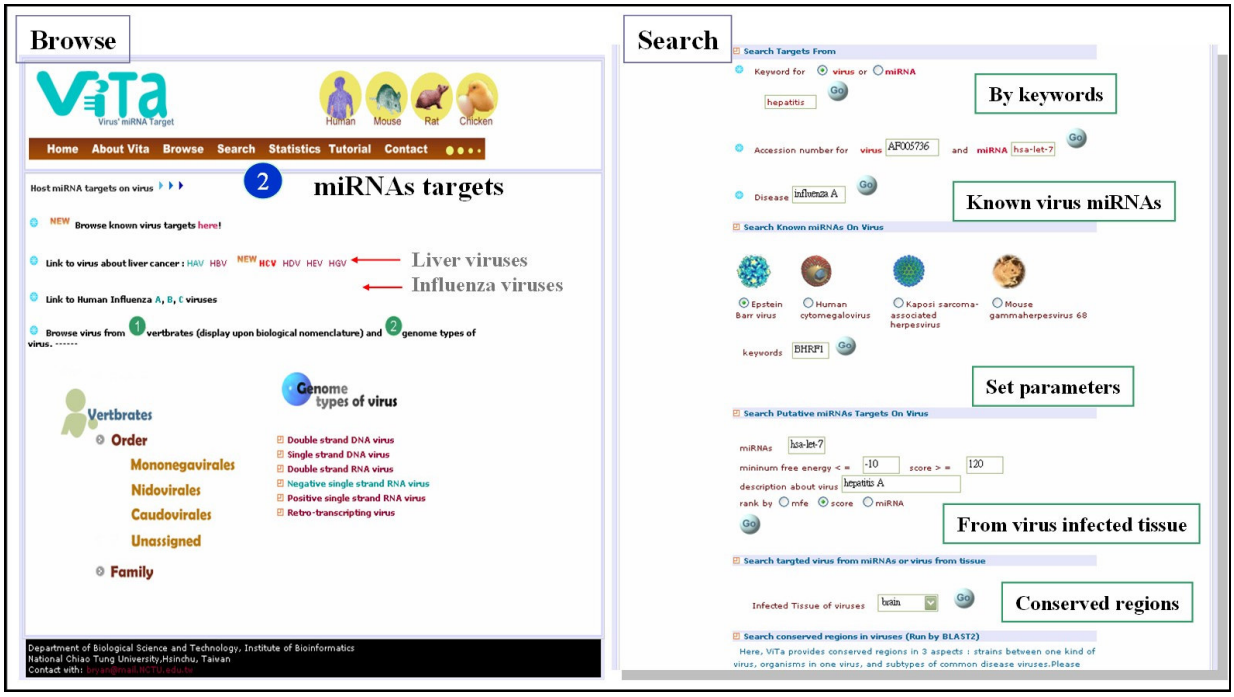

Figure 3. Browse page (left part) and search page (right part) in ViTa web interface

.

5. INTERFACE

The proposed web interface provides both a variety of data browsing functions and search functions, as shown in Fig. 3. For browsing the miRNA targets on viruses, the database provides two ways such as browsing by genome types and browsing by taxonomy (See the left part of Fig. 3). It allows users to search the database by keywords, accession numbers, disease names, and virus infected tissue preference (See the right part of Fig. 3). Figure S1 (See Supplementary Materials) demonstrates the conserved regions between virus species, detailed information about targeted viruses and the tissue-specificity of the miRNA targets. The database also stores known viral miRNAs, including the sequences, mature miRNAs, genomic locations and the literatures. The diagram illustrating the stem-loop RNA structures of the miRNA precursors, which are folded using RNAfold (16) and generated graphically by mfold (17).

Figure S2 (See Supplementary Materials) presents a visualization tool for presenting the host-miRNA targets on viruses. All miRNAs target sites are also provided in 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 documentation for the usage of ViTa is provided on the web interface.

6. 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 in viruses. We believe that the proposed resource can provide sufficient and effective information for investigation about the interaction between the host-miRNAs and viruses. 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. For instance, HCV prefers to infect the liver and miR-122 is a liver-specific miRNAs. Thus, the miR-122 targets discovered in HCV are significantly interesting.

The prospective works of the database are listed as follows. (i) More expression profiles for host-miRNA genes will be supported; (ii) Combinatorial miRNAs regulation to viral gene expression can be investigated; (iii) The association between virus and disease will be explored.

7. AVAILABILITY

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

8. ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council of the Republic of China for

financially supporting this research under Contract No. NSC 95-3112-E-009-002. Special thanks for the financially supports from National Research Program For Genomic Medicine (NRPGM), Taiwan. This work was also partially supported by MOE ATU.

9. 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**, D140-144.
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**, D135-139.
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. Buchen-Osmond, C. (2003) The universal virus database ICTVdB. *COMPUTING IN SCIENCE & ENGINEERING*, **5**, 16-25.
12. Lefkowitz, E.J., Upton, C., Changayil, S.S., Buck, C., Traktman, P. and Buller, R.M. (2005) Poxvirus Bioinformatics Resource Center: a comprehensive Poxviridae informational and analytical resource. *Nucleic Acids Res*, **33**, D311-316.
13. Kulkarni-Kale, U., Bhosle, S., Manjari, G.S. and Kolaskar, A.S. (2004) VirGen: a comprehensive viral genome resource. *Nucleic Acids Res*, **32**, D289-292.
14. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res*, **32**, D109-111.
15. 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.
16. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res*, **31**, 3429-3431.
17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

SUPPLEMENTARY MATERIALS

Table S1. Integrated external data sources.

Integrated Databases	Description	Literature cited
miRBase	Known viral microRNAs	(5)
miRNAMap	Known human, mouse and rat microRNAs	(6)
VirGen	Genomes and viruses	(13)
VBRC	Genomic sequences and gene annotations	(12)
ICTVdB	Vertebrate and plant viruses	(11)

Table S2. Integrated annotated tools.

Integrated Tools	Description	Literature cited
miRanda	Predicting of miRNA targets	(8)
TargetScan	Predicting of miRNA targets	(9)

Table S3. Experimental human miRNA targets on virus.

human miRNAs	Target virus	Literature cited
hsa-miR-122a	Hepatitis C Virus 1a strain H77c (AF011751)	(2)
hsa-miR-122a	Hepatitis C Virus 1b strain HCV-N (AF139594)	(2)
hsa-miR-29a	HIV-1 isolate BRU (K02013)	(3)
hsa-miR-29b	HIV-1 isolate BRU (K02013)	(3)
hsa-miR-149	HIV-1 isolate BRU (K02013)	(3)
hsa-miR-324-5p	HIV-1 isolate BRU (K02013)	(3)
hsa-miR-378	HIV-1 isolate BRU (K02013)	(3)
hsa-miR-378	HIV-1 isolate ELI (K03454)	(3)
has-miR-32	PFV-1	(4)

Table S4. Statistics for human miRNA targets predicted by TargetScan.

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**
≤ -10	87.48	108.45	1056.20
≤ -15	83.34	101.27	972.19
≤ -20	54.30	59.80	522.72
≤-25	18.03	14.90	125.69

Table S5. Statistics for miRNA targets on several liver-specific viruses.

MFE cutoff (kcal/mol)	**HAV (M14707)**	**HBV (X98077)**	**HCV (NC_004102)**	**HDV (NC_001653)**	**HEV (NC_001434)**	**HGV (U44402)**
By miRanda						
≤ -10	129	117	132	82	125	135
≤ -15	79	78	113	44	102	116
≤ -20	24	20	58	15	34	61
≤-25	3	6	11	5	3	13
By TargetScan						
≤ -10	33	0	53	0	25	57
≤ -15	31	0	48	0	25	57
≤ -20	19	0	31	0	14	40
≤-25	6	0	14	0	3	11

Table S6. Statistics for human miRNA targets on influenza viruses.

MFE cutoff (kcal/mol)	**Influenza A**	**Influenza B**	**Influenza C**
By miRanda			
≤ -10	102.96	174.65	107.79
≤ -15	46.94	70.02	37.75
≤ -20	7.16	9.61	6.63
≤-25	0.40	0.19	0.44
By TargetScan			
≤ -10	0	0	0
≤ -15	0	0	0
≤ -20	0	0	0
≤-25	0	0	0

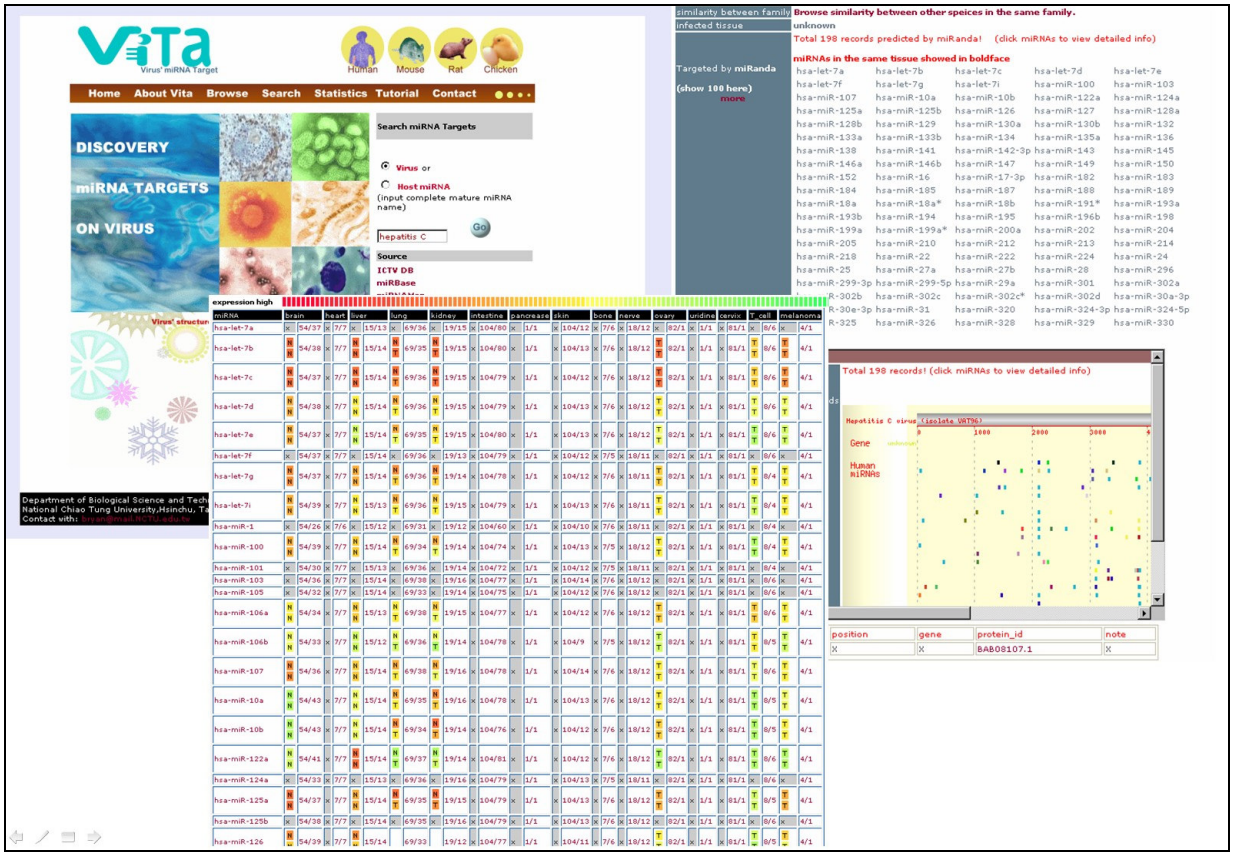

Figure S1. The ViTa web interface (at left is the home page, middle shows relationships between miRNAs and viruses classified by tissues, and at right is the detailed information about targeted viruses)

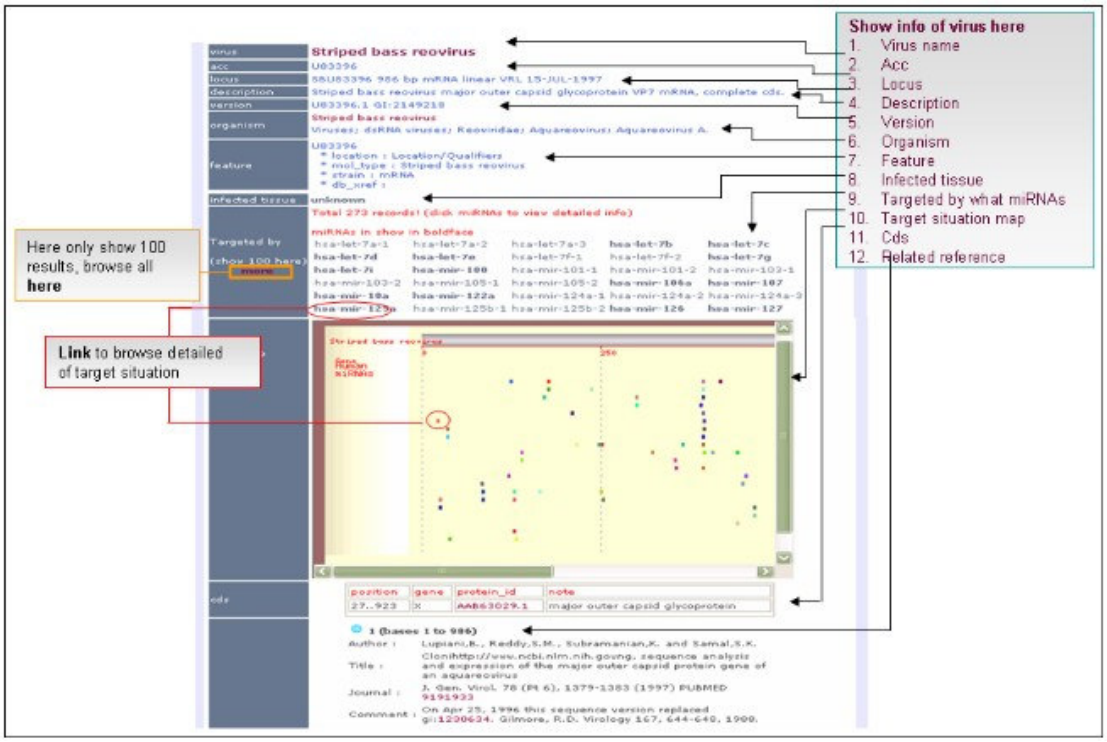

Figure S2. Interface of putative targeted viruses in ViTa.