Arch Virol (2012) 157:497–501
DOI 10.1007/s00705-011-1181-y

BRIEF REPORT

vHoT: a database for predicting interspecies interactions between viral microRNA and host genomes

Hanjoo Kim · Seunghyun Park · Hyeyoung Min · Sungroh Yoon

Received: 22 June 2011 / Accepted: 19 November 2011 / Published online: 8 December 2011

Abstract Some viruses have been reported to transcribe microRNAs, implying complex relationships between the host and the pathogen at the post-transcriptional level through microRNAs in virus-infected cells. Although many computational algorithms have been developed for microRNA target prediction, few have been designed exclusively to find cellular or viral mRNA targets of viral microRNAs in a user-friendly manner. To address this, we introduce the viral microRNA host target (vHoT) database for predicting interspecies interactions between viral microRNA and host genomes. vHoT supports target prediction of 271 viral microRNAs from human, mouse, rat, rhesus monkey, cow, and virus genomes. vHoT is freely available at http://dna.korea.ac.kr/vhot.

MicroRNAs (miRNAs) are small (19–24 nt in length) noncoding RNAs that down-regulate gene expression by binding to the 3′ untranslated region (3′-UTR) of the target mRNA bearing complementary target sequences [1]. MiRNAs have been reported to control a wide range of biological processes such as hematopoiesis [2], neurogenesis [3], cell cycle control [4], and oncogenesis [5], indicating that miRNAs are core elements of the complete gene regulatory network, together with transcription factors.

Since the fundamental role of miRNA is gene regulation, the final goal of functional research on miRNA is often to find cognate targets of the miRNA, to elucidate the regulatory mechanism mediated by the miRNA, and to characterize its interactions with the target mRNA. To this end, many target prediction methods [6–11] have been developed and have provided valuable tools for predicting the candidate targets of the miRNA of interest before wet-lab experiments are initiated for mechanism studies. However, those informatics tools are limited to predicting intra-species interactions between miRNA and mRNA targets in vertebrates and some lower animals such as flies and worms. Little has been done to predict interspecies interactions and especially to find the interactions between host targets and miRNAs of parasites such as viruses.

A number of viruses (mostly DNA viruses) have been shown to encode miRNAs [4, 12–14], and it is intriguing to see the effects of viral miRNAs on viral pathogenesis. Viral miRNAs have some distinct features compared to conventional miRNAs. Although most animal miRNAs down-regulate gene expression through incomplete binding to target sequences that is not accompanied by target cleavage, some virus-encoded miRNAs such as miR-BART2 from Epstein-Barr virus (EBV) have been shown to act in an siRNA-like manner [13]. In addition, in contrast to common cellular miRNA-mRNA interactions, there exist complex relationships between miRNAs and mRNA targets in virus-infected host cells, because infected cells possess two different (i.e., host and viral) genomes. Cellular miRNAs may affect the expression of viral mRNA targets as well as cellular targets, and similarly, viral

Electronic supplementary material The online version of this article (doi:10.1007/s00705-011-1181-y) contains supplementary material, which is available to authorized users.

H. Kim · S. Park · S. Yoon (✉)
School of Electrical Engineering,
Korea University, #207 Engineering Bldg., Anamdong,
Seongbukgu, Seoul 136-713, Korea
e-mail: sryoon@korea.ac.kr

H. Min (✉)
College of Pharmacy, Chung-Ang University,
84 Heukseokro, Dongjakgu, Seoul 156-756, Korea
e-mail: hymin@cau.ac.kr

miRNAs may inhibit the expression of host mRNAs and virus mRNAs. In other words, two additional interspecies interactions, namely the interaction between viral miRNA and host mRNA, and the interaction between host miRNA and viral mRNA, are present in virus-infected host cells. These interspecies interactions are critical for understanding the life cycle of a virus and its pathogenesis. Additionally, given that there are examples of viral miRNAs whose target regions are present not in the 3′ UTR but in the coding region [4], the 5′ UTR and coding sequences as well as the 3′ UTR need to be considered when putative mRNA targets are computationally predicted.

Although many computational algorithms have been developed to predict candidate miRNA targets, few have been designed exclusively to find cellular or viral mRNA targets of viral miRNAs. Recently, two computational methods, Reptar [15] and miRiam [16], have been developed to predict host mRNA targets of viral miRNAs, but there are limitations to using those tools in a viral miRNA study. miRiam is not based on a user-friendly web application but instead requires the Python interpreter to run the program, and thus, users who are unfamiliar with Python may not have easy access to this method. RepTar has been reported to predict putative cellular targets of viral miRNAs, but queries are limited to viral miRNAs derived from Epstein-Barr virus (EBV), human cytomegalovirus (HCMV), Kaposi's sarcoma virus (KSHV), mouse cytomegalovirus (MCMV), and mouse gammaherpesvirus (MGHV).

To address these limitations of the existing approaches, we introduce the viral miRNA host target (vHoT) database, a web-based tool that can search for mRNA targets of viral miRNAs. The vHoT database allows interspecies analysis and can thus be used to predict both cellular mRNA targets and viral mRNA targets of virus-derived miRNAs. The current version of vHoT can predict the targets of 271 viral miRNAs within the human, mouse, rat, rhesus monkey, cow, and virus genomes. Either a user-specified set of genes or the whole genome of an organism can be used for prediction. Five widely used algorithms that are known to be relatively accurate for target prediction, TargetScan [9, 10], miRanda [17, 18], RNAhybrid [19], DIANA-microT [8, 20] and PITA [21], were customized and used as the search engines of vHoT, and the user can specify the key parameters of these algorithms to fine-tune search results. Several aspects of the internal database and the user interface of vHoT were optimized to maximize the operational efficiency as well as the quality of the user experience.

Figure 1 shows the overall architecture of the proposed vHoT database. It takes three types of sequences as input: viral miRNA sequences; human, mouse, rat, rhesus monkey, and cow 3′ UTR sequences; and viral gene sequences. For the current release of vHoT, we obtained a total of 271 viral miRNA sequences from release 15 of miRBase [22–25]. For human and mouse, we downloaded approximately 89,530 mRNA 3′ UTR sequences from the TargetScan web site (release 5.1). We also utilized 20 viral genes available in the 2006 release of the ViTa database [7]. As the prediction engine of vHoT, we customized five widely used miRNA target prediction tools: TargetScan, miRanda, RNAhybrid, DIANAmicroT and PITA. When making a decision on which tools to use, we considered various aspects of the tool, such as its quality of prediction, how well it is maintained, the ease of data extraction and customization, and its popularity and acceptance in the community.

The target prediction algorithms and their prediction principles are depicted in Fig. 2. TargetScan is software to predict miRNA targets by searching for conserved sites that match the miRNA seed region. As of June 2011, the TargetScan database lists 3,088 miRNAs and 410,320 mRNA 3′ UTR sequences for 10 species. miRanda utilizes three key principles, namely sequence complementarity, the free energy of an RNA-RNA duplex, and conservation of target sites in related genomes. As of October 2011,

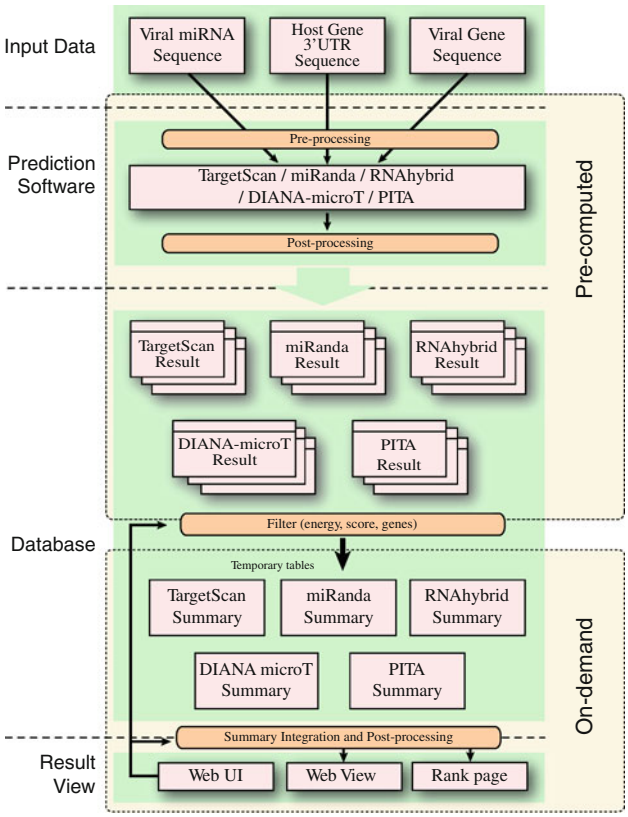

Fig. 1 Overall system architecture of the proposed vHoT database. There are two phases depicted in the figure: the pre-computation and the on-demand phases. The pre-computation phase occurs during the initial development of the database or whenever updates for prediction algorithms become available. The activities in the on-demand phase occur when the database is up and running

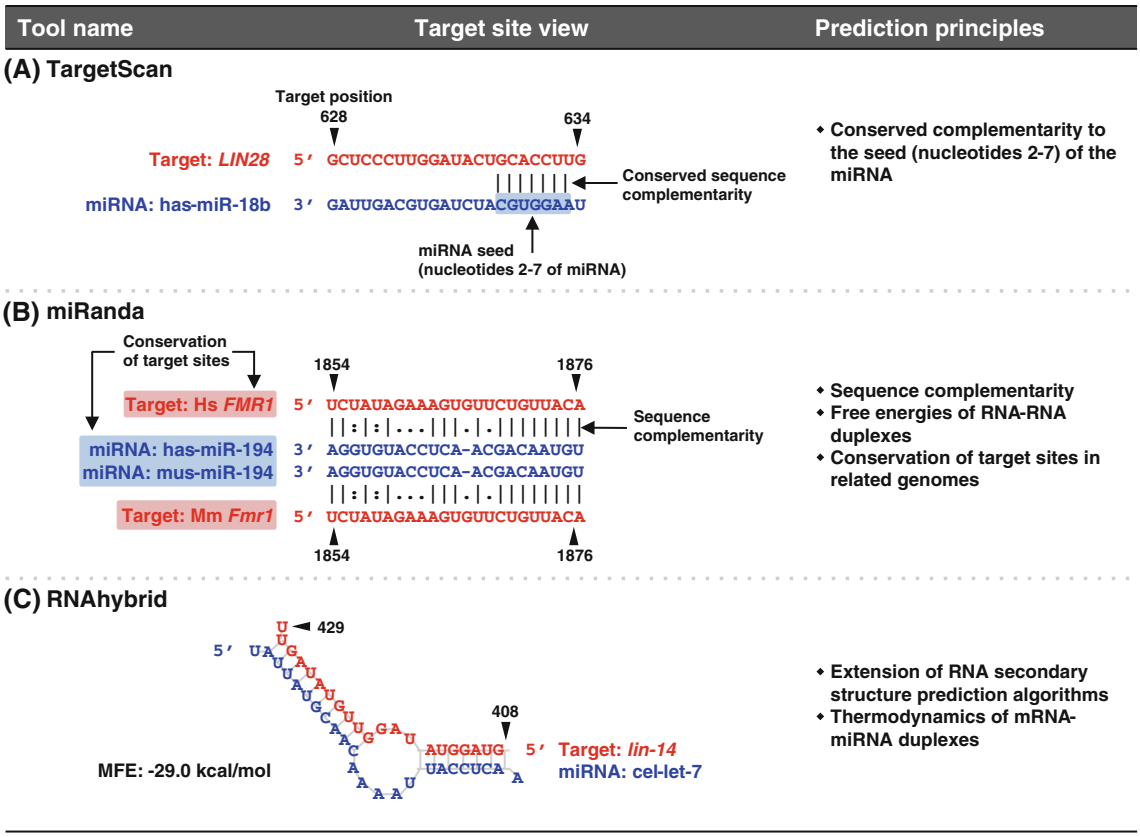

Fig. 2 Target-prediction algorithms and their prediction principles. (**A**) TargetScan predicts miRNA targets by searching for conserved sites that match the miRNA seed region (nucleotides 2–7). Shown are miRNA has-miR-18b and its *LIN28* target. (**B**) miRanda relies on three key prediction principles, namely sequence complementarity, the free energy of an RNA-RNA duplex and conservation of target sites in related genomes. Shown are conserved miRNAs has-miR-194 and mus-miR-194 and their targets Hs *FMR1* and Mm *Fmr1*. (**C**) RNAhybrid predicts miRNA targets by finding the minimum free energy (MFE) of hybridization of miRNA-mRNA duplexes. Shown are miRNA cel-let-7 and its target *lin-14*, which form a duplex with an MFE of −29.0 kcal/mol

miRanda provides 851 human, 793 mouse and 698 rat miRNAs and many more miRNAs derived from 19 species other than viruses. RNAhybrid predicts miRNA targets by finding the minimum free energy (MFE) of hybridization of miRNA-mRNA duplexes. RNAhybrid was originally developed as an extension of an RNA secondary structure prediction algorithm called BiBiServ RNA Studio [26]. DIANA-microT uses a 38-nt window for progressively scanning 3′ UTRs of candidate targets and calculates the minimum binding energy of miRNA-target duplexes by dynamic programming. DIANA-microT has been reported to have high precision levels greater than 66% [27]. PITA employs a parameter-free interaction model that computes the difference between the free energy of a miRNA-target duplex and the energetic cost of unpairing the target for making it accessible to the miRNA. PITA demonstrates that target accessibility is key in miRNA function.

For implementation, there are two phases in vHoT, as depicted in Fig. 1: the pre-computation and on-demand phases. The pre-computation phase occurs during the initial development of the database, or whenever updates for prediction algorithms become available. We first appropriately preprocess the three types of input sequences mentioned above so that the five prediction tools used can take them as input. Using the preprocessed sequence information, we then execute the five miRNA target prediction tools. To provide the user of vHoT with near-real-time response we compute a vast number of search results for each prediction software in advance and insert them into the MySQL database management system [28]. For efficient insertion, we can use the low-level IO functionality of MySQL and/or the database interface (DBI) module of the PERL language [29]. As a database engine, we employ the MyISAM storage engine, which is optimized for read-dominant applications such as vHoT, over the InnoDB engine [30].

The activities in the on-demand phase occur when the database is up and running. The user can issue a query after customizing it with various search options and conditions that vHoT supports. For instance, the user can adjust the parameters of individual prediction tools and decide

whether to use the union or intersection operation to combine the results from individual tools.

According to the user-specified filtering conditions and parameters, vHoT then creates a set of intermediate tables from the pre-computed results and processes these tables to generate the search result. To highlight the most important findings for the user, vHoT supports various options to rank the search result. Further details on vHoT can be found on the website (http://dna.korea.ac.kr/vhot/) or in the user manual provided as Supplementary Data.

The identification of miRNAs from viruses brought in a new layer of gene regulation affecting virus-host interactions. Many reports have shown that viruses use miRNAs to regulate their life cycles and evade host immune surveillance [31], but studies on viral miRNAs still lag behind those on human miRNAs or mouse miRNAs. With the development of vHoT, we expect that researchers investigating viral miRNAs will find it much easier to search for putative cellular and viral targets of their viral miRNAs of interest. We further anticipate that vHoT will contribute to elucidating the mechanism of viral pathogenesis by revealing interactions between miRNAs and cellular mRNAs. One limitation of the current version of vHoT is the number of species supported: vHoT considers humans, mice, rats, rhesus monkeys, cows and viruses as target species. The next release of vHoT will include more species, enabling the analysis of host-specific interspecies interactions of viruses with more types of host genomes.

For user convenience, it would also be possible to highlight among the search results the experimentally validated targets of viral miRNAs by utilizing the information available in the TarBase 5.0 [32] and miRecords [33] databases.

Acknowledgments Funding: This work was supported by the National Research Foundation (NRF) funded by the Korean Government Ministry of Education, Science and Technology (MEST) [2011-0009963 to S.Y., 2011-0005462 to H.M.].

References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297
2. Chen CZ, Li L, Lodish HF, Bartel DP (2004) MicroRNAs modulate hematopoietic lineage differentiation. Science 303:83–86
3. Liu C, Teng ZQ, Santistevan NJ, Szulwach KE, Guo W, Jin P, Zhao X (2010) Epigenetic regulation of mir-184 by mbd1 governs neural stem cell proliferation and differentiation. Cell Stem Cell 6:433–444
4. Grey F, Tirabassi R, Meyers H, Wu G, McWeeney S, Hook L, Nelson JA (2010) A viral microRNA down-regulates multiple cell cycle genes through mRNA 5′UTRs. PLoS Pathog 6:e1000967
5. Jiang S, Zhang HW, Lu MH, He XH, Li Y, Gu H, Liu MF, Wang ED (2010) MicroRNA-155 functions as an OncomiR in breast cancer by targeting the suppressor of cytokine signaling 1 gene. Cancer Res 70:3119–3127
6. Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N (2005) microRNA target predictions across seven drosophila species and comparison to mammalian targets. PLoS Comput Biol 1:e13
7. Hsu PW, Lin LZ, Hsu SD, Hsu JB, Huang HD (2007) Vita: prediction of host microRNAs targets on viruses. Nucleic Acids Res 35:D381–D385
8. Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A (2004) A combined computational-experimental approach predicts human microRNA targets. Genes Dev 18:1165–1178
9. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798
10. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20
11. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 126:1203–1217
12. Gottwein E, Mukherjee N, Sachse C, Frenzel C, Majoros WH, Chi JT, Braich R, Manoharan M, Soutschek J, Ohler U, Cullen BR (2007) A viral microRNA functions as an orthologue of cellular mir-155. Nature 450:1096–1099
13. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T (2004) Identification of virus-encoded microRNAs. Science 304:734–736
14. Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR (2008) MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. Nature 454:780–783
15. Lagana A, Forte S, Russo F, Giugno R, Pulvirenti A, Ferro A (2010) Prediction of human targets for viral-encoded microRNAs by thermodynamics and empirical constraints. J RNAi Gene Silencing 6:379–385
16. Elefant N, Berger A, Shein H, Hofree M, Margalit H, Altuvia Y (2011) RepTar: a database of predicted cellular targets of host and viral miRNAs. Nucleic Acids Res 39:D188–D194
17. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human microRNA targets. PLoS Biol 2:e363
18. Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microrna.org resource: targets and expression. Nucleic Acids Res 36:D149–D153
19. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10:1507–1517
20. Maragkakis M, Alexiou P, Papadopoulos GL, Reczko M, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Simossis VA, Sethupathy P, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG (2009) Accurate microRNA target prediction correlates with protein repression levels. BMC Bioinformatics 10:295
21. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39:1278–1284
22. Griffiths-Jones S (2004) The microRNA registry. Nucleic Acids Res 32:D109–D111
23. Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods Mol Biol 342:129–138
24. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microrna sequences, targets and gene nomenclature. Nucleic Acids Res 34:D140–D144
25. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158

26. Sczyrba A, Kruger J, Mersch H, Kurtz S, Giegerich R (2003) RNA-related tools on the bielefeld bioinformatics server. Nucleic Acids Res 31:3767–3770
27. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. Nature 455:58–63
28. Bessant C, Shadforth I, Oakley D (2009) Building bioinformatics solutions: with Perl, R and MySQL. Oxford biology. Oxford University Press, Oxford
29. Descartes A, Bunce T (2000) Programming the Perl DBI. O'Reilly, Cambridge
30. Zawodny JD, Balling DJ (2004) High performance MySQL: optimization, backups, replication, and load balancing, 1st edn. O'Reilly, Beijing
31. Sarnow P, Jopling C, Norman K, Schutz S, Wehner K (2006) MicroRNAs: expression, avoidance and subversion by vertebrate viruses. Nat Rev Microbiol 4:651–659
32. Sethupathy P, Corda B, Hatzigeorgiou A (2006) Tarbase: a comprehensive database of experimentally supported animal microRNA targets. RNA 12:192–197
33. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 37:D105–D110