ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2018, 19(3), pp. 565 - 573, DOI: 10.18038/aubtda.345972

COMPUTATIONAL IDENTIFICATION OF MICRORNAS FROM SSDNA VIRUSES

Müşerref Duygu SAÇAR DEMİRCİ *

Bioinformatics, Faculty of Life and Natural Sciences, Abdullah Gül University, Kayseri, Turkey

ABSTRACT

MicroRNAs (miRNAs) are post-transcriptional regulators of gene expression and the fact that they are associated with various disease phenotypes is one of the main reasons for their importance. The complexity of experimental detection of miRNAs due to their characteristics led to the development of computational methods. In this work, a machine learning based approach was applied to identify and analyze potential miRNAs that might be originated from 60 single strand DNA (ssDNA) viruses' genomes. The results suggest that 53 of these viruses may possibly produce proper miRNA precursors. Moreover, the possibility of these candidate miRNA precursors' ability to generate mature miRNAs that could target human genes and viral genomes has been tested. Overall, the outcomes of this research indicate that there might be another level of host-virus interaction through miRNAs which requires further experimental confirmation.

Keywords: MicroRNA, Bioinformatics, Machine learning, Virus, Computational biology

1. INTRODUCTION

MicroRNAs (miRNAs) act as the regulators of post-transcriptional regulation in many organisms. They are a class of non-coding RNAs that are small and single-stranded. Their regulatory function involves in a vast variety of processes e.g. developmental timing and crop yield in agricultural plants. The latest version of miRBase (Release 21) [1] lists miRNA information from 223 organisms showing that miRNA based regulation mechanism is shared from viruses to higher eukaryotes. Various miRNAs have been associated with many human diseases [2]. In addition, miRNA mediated regulation is assumed to be an important process in viral pathogenesis [3]. Moreover, the role of some miRNAs in complicated cross-talk between host and pathogen has been revealed [4,5].

Though more investigation is needed to solve the overall host-miRNA communications, recent evidence suggested that viral miRNAs have the capacity to affect the host cell [6,7]. Furthermore, there are many reasons that make miRNAs a beneficial tool for viruses such as gaining the ability to modulate host gene expression to form a more suitable environment for virus replication, high evolution rates of miRNAs increase the adaptability to new hosts and most importantly the presence of host miRNAs that might make virus encoded miRNAs possibly less immunogenic [6].

Among many challenges for the identification of new miRNAs, the most demanding one is the capacity of a genome to produce huge numbers of candidate miRNAs, e.g. 60 million hairpins found in human genome [8]. This problem led to the creation of sophisticated bioinformatics methods for miRNA prediction. To solve this issue, one of the two main methods are usually chosen; homology based methods using evolutionary relations and *ab initio* approaches predominantly depending on machine learning [9]. In this work, izMiR software is used since it performs better than previous approaches [8].

Previously, it has been demonstrated that the Torque Teno virus could encode a miRNA through host miRNA biogenesis elements and this miRNA targets human genes related to the host antiviral

*Corresponding Author: duygusacar@gmail.com
Received: 23.10. 2017 Accepted: 20.06.2018

response [10]. In this study, to test if it is also possible for other single strand DNA (ssDNA) viruses to encode miRNAs, all 60 ssDNA viral genomes with human host available in NCBI genome server were analyzed (Table 1). Furthermore, the targets of these miRNAs in the human genome and virus genomes as well as the targets of known human miRNAs on the virus genomes were investigated. The obtained results imply that 53 of these viruses might generate 239 functional miRNA precursors and 226 mature sequences that might target genes of host and other viruses which should be further validated with wet-lab experiments.

2. MATERIAL AND METHODS

In this study, all the data analysis tasks were performed using KNIME [11] which is a workflow management and data analytics platform. The miRNA analysis workflows known as izMiR [8,12] were applied on the generated virus data.

2.1. Datasets

The genomes of ssDNA viruses (Table 1) were obtained from NCBI (https://www.ncbi.nlm.nih.gov/genome/viruses/) split into overlapping fragments (500 nt long with 250 nt overlaps) and transcribed into RNA sequences (T => U as - strand, the complement as + strand) (Figure, step 1).

Secondary structures of all fragments were calculated by using RNAfold [13] with default settings and based on the resulting structures hairpins were extracted (Figure, steps 2 and 3). After filtering the hairpins according to their length distribution (min: 36, max: 180) and removing the duplicate sequences, for the remaining 2165 (+) strand and 1793 (-) strand hairpins, required features are calculated (Figure, step 4). Features defining hairpins were calculated by using an in-house java package; but it was also possible using online services (http://jlab.iyte.edu.tr/software/mirna).

Table 1. The list of virus genomes used in the study and their NCBI accessions. Size indicates number of nucleotides. (hp: number of hairpins passing threshold, M: number of matures passing threshold)

Accession	Definition	Size	hp	M
NC_000883	Human parvovirus B19, complete genome	5596	5	
NC_001401	Adeno-associated virus - 2, complete genome	4679	2	1
NC_001729	Adeno-associated virus - 3, complete genome	4726	3	
NC_002076	Torque Teno virus 1, complete genome	3852	8	9
NC_002195	Torque Teno mini virus 9, complete genome	2915	2	
NC_006152	Adeno-associated virus 5, complete genome	4642	6	3
NC_007013	Small anellovirus 1, complete genome	2249	4	1
NC_007014	Small anellovirus 2, complete genome	2635	1	1
NC_007455	Human bocavirus, complete genome	5299	6	9
NC_009225	Torque Teno midi virus 1, complete genome	3245		
NC_012126	California sea lion anellovirus, complete genome	2140		
NC_012564	Human bocavirus 3, complete genome	5242	4	
NC_014068	Torque Teno mini virus 8, complete genome	2910		
NC_014069	Torque Teno virus 4, complete genome	3690	5	12
NC_014070	Torque Teno sus virus 1a, complete genome	2878	2	1
NC_014071	Torque Teno canis virus, complete genome	2797	9	11
NC_014072	Torque Teno felis virus, complete genome	2064	4	3
NC_014073	Torque Teno virus 28, complete genome	3629	5	3
NC_014074	Torque Teno virus 27, complete genome	3729	3	3
NC_014075	Torque Teno virus 12, complete genome	3759	6	8
NC_014076	Torque Teno virus 10, complete genome	3770	3	5
NC_014077	Torque Teno virus 14, complete genome	3899	6	11
NC_014078	Torque Teno virus 19, complete genome	3808	5	6
NC_014079	Torque Teno virus 26, complete genome	3798	5	3
NC_014080	Torque Teno virus 7, complete genome	3736	6	6
NC_014081	Torque Teno virus 3, complete genome	3748	8	6
NC_014082	Torque Teno mini virus 7, complete genome	2952	2	3
NC_014083	Torque Teno virus 25, complete genome	3763	2	1
NC_014084	Torque Teno virus 8, complete genome	3790	10	15
NC_014085	Torque Teno tamarin virus, complete genome	3371	1	1
NC_014086	Torque Teno mini virus 2, complete genome	2765		
NC_014087	Torque Teno douroucouli virus, complete genome	3718	5	1
NC_014088	Torque Teno mini virus 3, complete genome	2897		
NC_014089	Torque Teno mini virus 5, complete genome	2908	4	1
NC_014090	Torque Teno mini virus 4, complete genome	2785	1	
NC_014091	Torque Teno virus 16, complete genome	3818	9	11
NC_014092	Torque Teno sus virus k2 isolate 2p, complete genome	2735	1	1
NC_014093	Torque Teno midi virus 2, complete genome	3253		
NC_014094	Torque Teno virus 6, complete genome	3705	4	3
NC_014095	Torque Teno mini virus 6, complete genome	2897		
NC_014096	Torque Teno virus 15, complete genome	3787	10	11
NC_014097	Torque Teno mini virus 1, complete genome	2856	1	1
NC_014480	Torque Teno virus 2, complete genome	3322	4	5
NC_015212	Seal anellovirus TFFN/USA/2006, complete genome	2164	2	2
NC_015783	Torque Teno virus, complete genome	3725	7	7
NC_020498	TTV-like mini virus isolate TTMV_LY1, complete genome	2912	1	2
NC_024691	MSSI2.225 virus complete sequence	2259	3	1
NC_024890	Seal anellovirus 3, complete genome	2169	2	
NC_024891	Seal anellovirus 2, complete genome	2149	2	1
NC_025726	Torque Teno mini virus ALA22, complete genome	2914	1	
NC_025727	Torque Teno mini virus ALH8, complete genome	2855	2	
NC_025966	Rodent Torque Teno virus 2 isolate RN_2_Se15, complete genome	2572	7	5
NC_026662	Simian Torque Teno virus 31 isolate VGA00123.3, complete genome	3907	7	6
NC_026663	Simian Torque Teno virus 30 isolate VWP00522.2, complete genome	3612	8	6
NC_026664	Simian Torque Teno virus 32 isolate VGA00154.2, complete genome	3812	6	14
NC_026764	Simian Torque Teno virus 33 isolate VWP00522.11, complete genome	3774	8	9
NC_026765	Simian Torque Teno virus 34 isolate VGA00120.1, complete genome	3621	6	11
NC_027059	Torque Teno sus virus 1b isolate 1p, complete genome	2872	5	1
NC_028101	Human bocavirus 2 isolate PK-2255, complete genome	5134	6	2
NC_028753	Anellovirus PRA4 strain chat-13 ORF1 gene, complete cds	2065	4	3

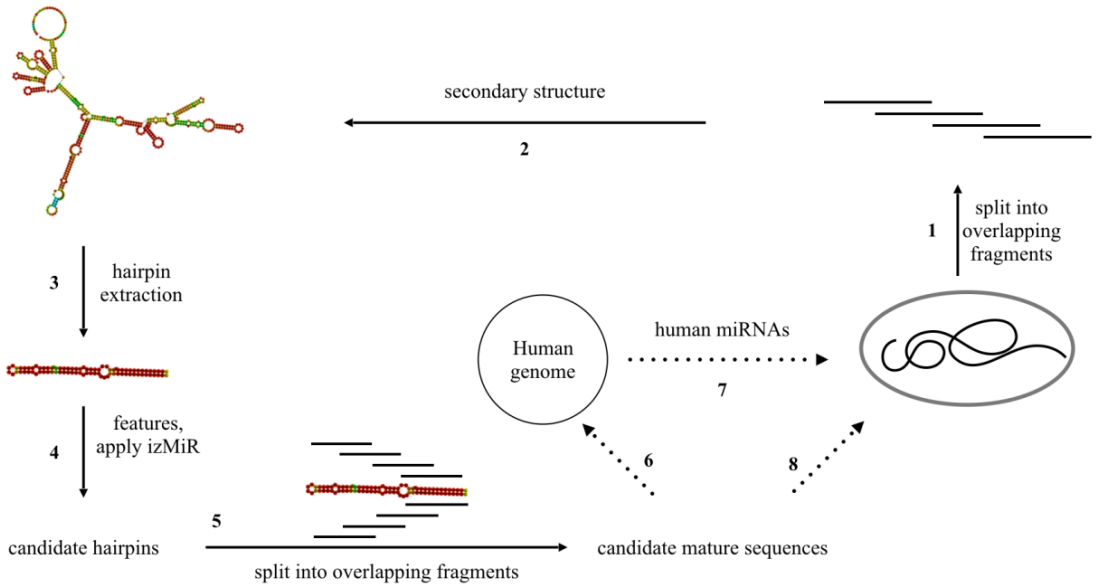

Figure. Schematic representation of the miRNA analysis process. Numbers indicate the order of steps.

2.1. miRNA prediction

For identifying miRNA precursors, the izMiR framework was used (Figure step 4). After this step, for hairpins passing the defined thresholds, overlapping fragments (24 nt length, 6 nt overlap) were generated and for those sequences mature miRNA detection was performed (Figure, step 5). In order to achieve this, a model was created by using 4316 mature miRNA sequences from miRBase which were also listed in miRTarBase (Release 6.0)[14]. To generate a negative data set, the mature sequences were shifted by half of their length within the hairpin sequence [15]. To train the classifier for mature miRNA prediction, 101 features were calculated: start and end positions of mature sequence in pre-miRNA (2), central loop start and end points (2), length of hairpin without flanking ends, miRBase hairpin length, stem length, mature length, maximum loop length (5), number of matches and mismatches in the mature sequence region (2), single nucleotide counts (4), dinucleotide counts (16), trinucleotide counts (64), distances of start and end positions to 3', 5', loop start and loop end (6). Then, these data sets with features were used to train a Random Forest learner using 1000 fold Monte Carlo Cross Validation [16] (70% (learning) - 30% (testing) ratio [17]). The model with the highest accuracy score (0.932) was selected and applied to mature sequence candidates.

2.2. miRNA target prediction

There are various computational tools, websites and databases that are designed for miRNA target prediction. However, in most of the cases such tools are limited for analysis in certain organisms [18]. In this work, psRNATarget (2017 release, [19]) was used to find possible miRNA - mRNA interactions, by using default settings since it is possible to upload mature sequences and genes for their targets while using psRNATarget online analysis. For finding virus miRNAs' targets in human genome, available target library from human genomic sequencing project (downloadable at psRNATarget) was used (Figure, step 6). Human miRNA mature sequences from miRBase were tested against viral genomes to search for candidate targets (Figure, step 7). In the latter case, for finding targets of virus miRNAs in the virus genomes, genomes of 60 viruses (Table 1) were uploaded as the target data (Figure, step 8).

3. RESULTS

Applying izMiR prediction workflows on the 3958 hairpins (both strands included), provided various scores for further analysis. Based on the prediction score value of AverageDT (average of decision tree based models' prediction scores, for more detailed explanation please see [8]) a threshold of 0.9 was set. This cut-off value led to the prediction of 239 hairpins which were then used for mature miRNA prediction.

Since hairpins need to be further processed into mature miRNAs in order to be functional, by fragmenting the 239 hairpins into 24 nt long sequences with 6 nt overlaps, a mature miRNA candidate pool was created. Mature candidates were filtered based on their model prediction score of at least 1.0, leading to 226 mature sequences.

In target prediction analysis, 226 virus originated candidate mature sequences were tested if they had any targets in the human genome and in the viral genomes listed in Table 1. According to the results shown in Table 2, viral miRNAs could take part in 17 miRNA – human mRNA interactions. Moreover, 2 viral miRNAs (both (-) strand) seemed to be able to target 16 locations in some viral genomes (Table 3).

Table 2. Targets of candidate virus miRNAs in human genome. Target descriptions were obtained from Amigo2 [20,21]. (S: strand, Origin: origin of miRNA)

MiRNA	S	Origin	Target Accession	Target Description
AUUCAUCGACUGCAGU UCCUAACU	+	Torque Teno virus 1	NM_033426 KIAA1737	CLOCK-interacting pacemaker
CCUCUCCCUGCACAUG GAGGUUCC	+	Adeno-associated virus 5	NM_001082968 TOM1L2	TOM1-like protein 2
UCACAUCUUUUGUGAU GGAGAGCA	-	Human bocavirus	NM_002725 PRELP	Prolargin
ACUAGGAGGUUUGCCU ACAGACCG	+	Human bocavirus	NM_016282 AK3	GTP:AMP phosphotransferase AK3, mitochondrial
GUUCACAUCUUUUGUG AUGGAGAG	-	Human bocavirus	NM_001256 CDC27	Cell division cycle protein 27 homolog
CUUUUUCUCCUUCAGC GAGACUAA	+	Torque Teno canis virus	NM_001845 COL4A1	Collagen alpha-1(IV) chain
CUUUUUCUCCUUCAGC GAGACUAA	+	Torque Teno canis virus	NM_001136265 IFFO2	Intermediate filament family orphan 2
CUUUUUCUCCUUCAGC GAGACUAA	+	Torque Teno canis virus	NM_001256585 MGLL	Monoglyceride lipase
CUACCUCUGCUGCUGG AGCUUGAC	+	Torque Teno felis virus	NM_182679 GPATCH4	G patch domain-containing protein 4
CUGGGCAUUUUUCUUA UUCUUAUU	+	Torque Teno virus 7	NM_017553 INO80	DNA helicase INO80
AUCCGGAAGGCCUGAU GGUUUUAC	+	Torque Teno virus 8	NM_001080485 ZNF275	Zinc finger protein 275
CGGCCAUUUUAGAUUG GCGCAGAG	-	Torque Teno virus 15	NM_005171 ATF1	Cyclic AMP-dependent transcription factor ATF-1
GGAGGUUUGGGGGCU GGGGGCCCU	+	Torque Teno virus 2	NM_001144952 SDK2	Protein sidekick-2
UUGGGGGCUGGGGGCC CUCGCGGC	+	Torque Teno virus 2	NM_018969 GPR173	Probable G-protein coupled receptor 173
AGGAGGAGGGAGUCU UGGGGUCGG	-	Simian Torque Teno virus 33 isolate VWP00522.11	NM_001143854 RPH3A	Rabphilin-3A
AGGAGGAGGGAGUCU UGGGGUCGG	-	Simian Torque Teno virus 33 isolate VWP00522.11	NM_001004342 TRIM67	Tripartite motif-containing protein 67
CCUCCUCCCUCAGAAC CCCAGCCU	+	Simian Torque Teno virus 33 isolate VWP00522.11	NM_173481 C19orf21	Mitotic interactor and substrate of PLK1

Table 3. Targets of candidate virus miRNAs in virus genomes. (Origin: origin of miRNA, Start: starting position of miRNA-target binding in target)

MiRNA	Origin	Start	Target
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3714	Torque Teno virus 3
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3673	Torque Teno virus 6
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3757	Torque Teno virus 8
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3655	Torque Teno virus 27
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3728	Torque Teno virus 25
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3757	Torque Teno virus 15
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3787	Torque Teno virus 16
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3651	Torque Teno virus
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3781	Simian Torque Teno virus 32 isolate VGA00154.2
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	3656	Torque Teno virus 4
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	2748	Torque Teno canis virus
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	2660	Torque Teno canis virus
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	2717	Torque Teno sus virus 1a
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	2874	Torque Teno mini virus 8
CAGGGGGGCGGAGCCCCC CCCGCA	Rodent Torque Teno virus 2 isolate RN_2_Se15	2411	Rodent Torque Teno virus 2 isolate RN_2_Se15
GGCCCCGUCACGUGACUU ACCACG	Torque Teno virus 1	3642	Torque Teno virus 14

For the last stage of targeting analysis, 2588 human miRNA mature sequences from miRBase were used to find if they had targets in virus genomes under consideration. This resulted in 8 miRNA – target interactions between 7 miRNAs and 6 virus genomes (Table 4).

Table 4. Targets of human miRNAs in virus genomes. (Start: starting position of miRNA-target binding in target)

MiRNA	Target Accession	Start	Target
hsa-miR-4668-3p	NC_015783	2813	Torque Teno virus
hsa-miR-5007-5p	NC_007013	552	Small anellovirus 1
hsa-miR-519c-3p	NC_014091	247	Torque Teno virus 16
hsa-miR-6750-3p	NC_014078	2605	Torque Teno virus 19
hsa-miR-6786-5p	NC_001729	74	Adeno associated virus 3
hsa-miR-6786-5p	NC_001729	4675	Adeno associated virus 3
hsa-miR-6808-3p	NC_014077	1045	Torque Teno virus 14
hsa-miR-6844	NC_007013	968	Small anellovirus 1

Out of the 124 hairpin sequences provided in an earlier study [10] only 6 were identical to the hairpins detected in this work and 4 of them passed the defined thresholds (Data not shown).

4. DISCUSSION

Since a wide range of organisms use miRNA mediated post-transcriptional regulation of gene expression, searching for miRNAs has been an essential step to understand this regulatory mechanism. Especially, due to the fact that many miRNAs are associated with various disease phenotypes, identifying new miRNAs efficiently is very important. However, there are challenging issues for a successful analysis. For instance, a miRNA might target hundreds of different mRNAs and one mRNA might be targeted by hundreds of different miRNAs, e.g. hsa-mir-155-5p has 173 known targets and PTEN gene is targeted by 83 miRNA (MiRTarBase Releae 7.0)[14]. Thus, to reduce the search space for experimental miRNA - mRNA interactions, using a reliable computational method is essential. Due to this, one of the most comprehensive approaches for miRNA detection, izMiR was used in this study.

Growing evidence suggest that there might be host-parasite communications through miRNAs. Previously, we showed that such an interaction could be possible in *Toxoplasma gondii* infection and between retro-transcribing viruses and their human host [4,5,22]. Since there are only 29 viruses out of 223 organisms listed in miRBase, and none of them are ssDNA viruses, in this work, the capacity of 60 ssDNA viruses with human hosts to produce functional miRNAs were analyzed.

Since it has been shown that at least one ssDNA virus strain was able to encode miRNAs [10] and based on the findings presented in Table 2 and Table 3, there is evidence that these viruses have the capacity to produce functional miRNAs. Moreover, the outcomes of this analysis suggest that these miRNAs not only target essential human genes taking part in various networks but also could alter the expression of other viruses' genes as well as their own (Table 3, miRNA by Rodent Torque Teno virus 2 isolate RN_2_Se15 targets itself). There are examples of autoregulation of viral mRNAs by the miRNAs produced by the virus itself [23] but for a virus miRNA to be able to target another virus's mRNAs, coinfection is required. Torque Teno virus strains are shown to be co-infecting with other viruses [24] but given the fact that all the virus miRNA and virus mRNA interactions presented in Table 3 are between Torque Teno virus strains, it might be due to their sequence similarity even though they are known with their high level of genetic diversity (between 5-90%) [25].

In addition, whether an innate miRNA mediated defense mechanism against infecting viruses exists in human was searched by looking for the targets of human miRNAs in viral genomes (Table 4). According to the results, some of the known human miRNAs might indeed have the potential to act as antiviral agents. This could lead to the development of new strategies for usage of miRNAs in therapeutic applications.

4. CONCLUSION

In conclusion, miRNAs appear to be taking part in ssDNA viruses' interactions with their human host. Further experimental evidence is required to fully grasp the specifics of such cross-kingdom regulation. On another level, communications among different viruses through miRNAs are still not clear. However, here it is shown that such a mechanism could be possible between ssDNA viruses.

ACKNOWLEDGEMENTS

The author is indebted to Dr. Jens Allmer for critical feedback on the manuscript.

REFERENCES

[1] Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic acids research 2011; 39: D152-7.

[2] Tüfekci KU, Oner MG, Meuwissen RLJ, Genç S. The role of microRNAs in human diseases. Methods in molecular biology (Clifton, N.J.) 2014; 1107: 33–50.

[3] Scaria V, Hariharan M, Maiti S et al. Host-virus interaction: a new role for microRNAs. Retrovirology 2006; 3: 68.

[4] Saçar MD, Bağcı C, Allmer J. Computational prediction of MicroRNAs from toxoplasma gondii potentially regulating the hosts' gene expression. Genomics, Proteomics and Bioinformatics 2014; 12: 228–38.

[5] Saçar Demirci MD, Bağcı C, Allmer J. Differential Expression of Toxoplasma gondii MicroRNAs in Murine and Human Hosts. Non-Coding RNAs and Inter-Kingdom Communication. Cham: Springer International Publishing, 2016: 143–59.

[6] Skalsky RL, Cullen BR. Viruses, microRNAs, and host interactions. Annual review of microbiology 2010; 64: 123–41.

[7] Grundhoff A, Sullivan CS. Virus-encoded microRNAs. Virology 2011; 411: 325–43.

[8] Saçar Demirci MD, Baumbach J, Allmer J. On the performance of pre-microRNA detection algorithms. Nature Communications 2017; 8.

[9] Saçar MD, Allmer J. Machine Learning Methods for MicroRNA Gene Prediction. In: Yousef M, Allmer J, eds. MiRNomics: MicroRNA Biology and Computational Analysis SE - 10. Vol1107. Humana Press, 2014: 177–87.

[10] Kincaid RP, Burke JM, Cox JC, de Villiers E-M, Sullivan CS. A human torque teno virus encodes a microRNA that inhibits interferon signaling. PLoS pathogens 2013; 9: e1003818.

[11] Berthold MR, Cebron N, Dill F et al. KNIME - The Konstanz Information Miner. SIGKDD Explorations. Vol11. 2009: 26–31.

[12] Allmer J, Saçar Demirci MD. izMiR: computational ab initio microRNA detection. 2016 Jul 15.

[13] Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Research 2003; 31: 3429–31.

14] Chou CH, Chang NW, Shrestha S et al. miRTarBase 2016: Updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Research 2016; 44: D239–47.

[15] Gkirtzou K, Tsamardinos I, Tsakalides P, Poirazi P. MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. PloS one 2010; 5: e11843.

[16] Xu Q-S, Liang Y-Z. Monte Carlo cross validation. Chemometrics and Intelligent Laboratory Systems 2001; 56: 1–11.

[17] Saçar Demirci MD, Allmer J. Delineating the impact of machine learning elements in pre-microRNA detection. PeerJ 2017; 5: e3131.

[18] Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD. Bioinformatic tools for microRNA dissection. Nucleic Acids Research 2016; 44: 24–44.

[19] Dai X, Zhao PX. PsRNATarget: A plant small RNA target analysis server. Nucleic Acids Research 2011; 39: W155-9.

[20] Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. AmiGO: online access to ontology and annotation data. BIOINFORMATICS APPLICATIONS NOTE 2009; 25: 288–28910.

[21] The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research 2017; 45: D331–8.

[22] Saçar Demirci MD, Toprak M, Allmer J. A Machine Learning Approach for MicroRNA Precursor Prediction in Retro-transcribing Virus Genomes. Journal of integrative bioinformatics 2016; 13.

[23] He Y, Yang K, Zhang X. Viral microRNAs targeting virus genes promote virus infection in shrimp in vivo. Journal of virology 2014; 88: 1104–12.

[24] García-Álvarez M, Berenguer J, Álvarez E et al. Association of torque teno virus (TTV) and torque teno mini virus (TTMV) with liver disease among patients coinfected with human immunodeficiency virus and hepatitis C virus. European Journal of Clinical Microbiology & Infectious Diseases 2013; 32: 289–97.

[25] Ssemadaali MA, Effertz K, Singh P, Kolyvushko O, Ramamoorthy S. Identification of heterologous Torque Teno Viruses in humans and swine. Nature Publishing Group2016; doi 10.1038/srep26655.