Human Encoded miRNAs that Regulate the Inflenenza Virus Genome

Hao Zhang,Xin Li,Yuaning Liu*
Symbol Computation and Knowledge Engineering of Ministry of Education,
College of Computer Science and Technology,
Jilin University,
Changchun , China
*Corresponding author, e-mail: zhangh@jlu.edu.cn

Minggang Hu
Symbol Computation and Knowledge Engineering of Ministry of Education,
College of Computer Science and Technology,
Jilin University,
Changchun , China

Zhi Li
College of Applied Technique,Changchun University of Science & Technology,
Changchun ,China

Dong Xu*
Computer Science Dept,
Life Sciences Center,
University of Missouri
*Corresponding author,e-mail:zhanhao@missouri.edu

***Abstract*—Motivation:MiRNAs can downregulate gene expression by mRNA cleavage or translational repression. Discovering human encoded miRNAs that regulate the influenza virus genome is important for molecular targets for drug development, and it also plays positive role in influenza control and prevention.**

Methods: We propose a new method based on scoring to discover human encoded miRNAs that regulate the influenza virus genome. The scoring based on the same complementary sites, the secondary structure of the complementary sites and the binding sites of all sequences respectively. Among them, taking the secondary structure as a vital factor is a new attempt.

Results: Has-miR-489, has-miR-325, has-miR-876-3p and has-miR-2117 are targeted HA, PB2, MP and NS of influenza A, respectively.

Keywords-Human encoded miRNA; influenza sequences; binding site; scoring

I. INTRODUCTION

MicroRNAs (miRNAs) are small noncoding and endogenous RNAs, ~22 nucleotides in length, and have been discovered in a variety of organisms. MiRNAs play important roles in many biological processes such as development, apoptosis, and establishment of cell lineage by targeting mRNAs that can direct the RISC (RNA-induced silencing complex) to downregulate gene expression by either of two posttranscriptional mechanisms: mRNA (message RNA) cleavage or translational repression [8]. If complementarity between 3' untranslated region (UTR) of the mRNA and the 5' end of the miRNA (typically between nucleotides 2–7, the so-called 'seed' region [5] is sufficient, the miRNA will specify cleavage, if complementarity is insufficient, it will repress translation. Targets regulated by miRNAs such as transcription factors, signal proteins, and so on are the important molecules in biological networks, although the accurate mechanism by which miRNAs mediate regulation is not completely understood, prediction of targets is critical to advancing our understanding of human diseases, and many proteins controlled by miRNAs might be important drug targets [16,18].

Influenza is an infectious disease caused by RNA viruses of the family Orthomyxoviridae, which are influenza viruses. The most common symptoms are fever, sore throat, muscle pains, etc. Influenza can cause pneumonia in more serious cases, which is fatal. Influenza viruses are classified into three types: A, B and C. Influenza A infect a wide variety of avian and mammalian species including humans, which can be subdivided into different serotypes based on the antibody response to these viruses [1]. Influenza B virus almost exclusively infect humans but are less common than influenza A because there is one known subtype. The illness caused by Influenza C virus has rarely been detected, which usually causes a mild upper respiratory illness [21], and so influenza C is less common than the other types [7]. Influenza A and B genome each contain eight segments of single-stranded RNA, and C contains seven segments of single-stranded RNA. Each RNA segment encodes one or two proteins [15]. Take influenza A for example, the eight RNA segments are HA, NA, NP, M, NS, PA, PB1, and PB2 that code eleven proteins. Influenza virus genome is prone to gene reassortment. The novel influenza emerged in Mexico, the United States, and several other nations in 2009, known as influenza A/H1N1 is a mixed strain [17].

Influenza viruses only replicate within living cells, deliver their genes and accessory proteins into the host cells [2]. Host cells do not passively accept virus infraction, but produce the ability of resistance and neutralization actively. To study the interaction between viruses and host cells is very necessary to understand the mechanism of pathogenic deeply so as to search appropriate anti-virus method. Recent researches demonstrate miRNAs encoded by virus or human may exert an important influence on the interaction between virus and host. Regulatory effect of miRNAs on virus replication and pathogenicity has been studied. Among the effects, some

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

literatures prove the binding mode between human encoded miRNAs and virus. For example, human foamy virus (PFV) genome have a region that can be used as the target site of human encoded miR-32 [6], hepatitis C virus (HCV) replication is regulated by miR-199a* that may serve as a novel antiviral therapy [24] and human encoded miRNAs can target crucial HIV-1 genes [14]. Researcher have also found that human encoded miRNAs could target critical genes involved in the pathogenesis and tropism of Influenza virus A/H5N1 and the target regions in the respective genes were found to be conserved across different viral strains [23]. Consequently, it is of great significance to study on host-virus interactions. Here we propose a method based on scoring to discover human encoded miRNAs which have targets in influenza virus genome. Four segments HA, PB2, MP and NS of influenza A genome are used in this paper. The result shows that has-miR-489, has-miR-325, has-miR-876-3p and has-miR-2117 are targeted HA, PB2, MP and NS of influenza A genome, respectively. Proteins coded by these genes are important components of influenza A infection and replication cycles.

II. DATA

The human coded miRNA sequences were downloaded from miRBase (http://www.mirbase.org/index.shtml) [20]. The data of this database include predicted hairpin portion of a miRNA transcript, information on the location and mature sequences. Mature sequences are used in this paper. The number of mature sequences is 904 and data mode is given in Table I. And the sequences of human encoded miRNA are numbered. The RNA segments of Influence A genome, HA, PB2, MP and NS are chosen to be study data. 50 sequences of each segment from past ten years (2000-2009) are needed, which are randomly selected from NCBI (http://www.ncbi.nlm.nih.gov/) and are complete cds. Data mode is like that of miRNA sequences, which is also FASTA format.

Align all the influenza sequences using ClustalW2 with default parameters (ALINMENT: full, SCORE TYPE: percent, NO END GAPS: yes, ITERATION: none, NUMITER: 1, OUTPUT FAMAT: aln w/numbers, OUTPUT ORDER: aligned, TREE TYPE: none, CORRECT DIST: off, IGNORE GAPS: off, CLUSTERING: NJ, others: def) and get a consensus secondary structure using RNAalifold of Vienna RNA package with default parameters (RNAalifold version: new RNAalifold with RIBOSUM scoring, Fold algorithms and basic options: minimum free energy (MFE) and partition function, avoid isolated base pairs) by which alignment format of ClustalW2 is accepted. Alignment of 50 sequences is convenient to show the position information of nucleotide. The results of alignment and RNA folding are the input data.

III. METHODS

The method proposed in this paper is based on scoring, and the secondary structure of mRNAs is considered as an important factor. Extensive RNA structure constitutes an important component of the genetic code. Higher-order RNA structure has been demonstrated that it directly encodes protein structure and is crucial for the adaptability of virus [11]. It affects molecular function indirectly. For this reason, scoring based on the secondary structure is one requisite part in the method.

TABLE I. THREE OF HUMAN ENCODED MIRNA SEQUENCES USED IN THE STUDY

>hsa-miR-576-3p MIMAT0004796
AAGAUGUGGAAAAAUUGGAAUC
>hsa-miR-194* MIMAT0004671
CCAGUGGGGCUGCUGUUAUCUG
>hsa-miR-140-5p MIMAT0000431
CAGUGGUUUUACCCUAUGGUAG

The data are FASTA format, which is the only format that fits for the algorithm we proposed in this paper. Has-miR-** is ID and MIMAT** is accession number.

There are three steps of scoring in the method which is expressed in terms of formula as follows to find the human coded miRNA sequences that can regulate influenza virus genome. I1 of the formula is the number of sequences from 2009, i2 is the number of the sites, j1 is the number of sequences from 2000 to 2008, N is the number of nucleotides that are completely complementary with the bind sites. The steps are:

$$SCORE = \sum_{i_1=1}^{5}\sum_{i_2=1}^{N_1} k_{i_2} + \sum_{i_2=1}^{N_1} k_{i_2} + \sum_{j_1=1}^{45}\sum_{i_2=1}^{N_1}\left[N*\left(\sum_{i_1=1}^{5}\sum_{i_2=1}^{N_1} k_{i_2} + \sum_{i_2=1}^{N_1} k_{i_2} \right)\right] (4 < N \le 8)$$

(1) The first scoring: based on the same complementary sites. Sliding window method is applied to search complete complementary fragments for each sequence from 2009, which are bound by the seed regions of miRNAs with base 1–8 from 5' end to 3' end. The rule is that giving miRNAs some marks, the seeds of which bind the same complementary sites.

(2) The second scoring: based on the secondary structure of the complementary sites which are bound by the seeds of miRNAs. Begin with the binding sites for every miRNA, if the nucleotide is in the stem, give the miRNA some marks. Then compute the total scores of every miRNA. Add these scores to the corresponding ones got in the first step.

(3) The third scoring: based on the bind sites of sequences from 2000 to 2008. Sliding window method is applied again. For every miRNA, if nucleotides in the widow of current bind site are completely complementary with the seed, the score increases by 8-fold; it is 7-fold if seven nucleotides are complementary, and so on. When four or less than four nucleotides are complementary ($0 \le N \le 4$), the score of the third part is zero, that is the final score is the sum of the first and the second part.

At last, rank scores in ascending order. The miRNA with the highest score and the target subsequence from 2009 are the results. Then use RNAfold of Vienna RNA package to get the information of complementarity and binding energy. If more than one miRNA have the same score, only the one with the lowest binding energy is the last result. Because a single linear RNA sequence are required in the RNAfold program as input, a linker sequence, "LLLLLL", is used to connect the miRNA and target subsequence. The "L" is not an RNA nucleotide, so it does not match with any RNA nucleotide, it is only to guarantee the binding of miRNA and target subsequence.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

To determine the marks of step 1 and step 2 respectively and which step has significant effect on the result, we use the different scores for the test. Firstly, we give step 1 ten marks and step 2 zero. Secondly, give step1 zero and step ten marks. Lastly, give step 1 and step 2 ten marks. The result shows in Fig 2, x-coordinate is in the order of increasing scores and y-coordinate is the number of miRNAs. We can see that when give step 1 ten marks and step 2 zero, the result is different from the others, but when give step 1 zero and step 2 ten marks, the result is almost identical. So step 2 has significant effect on the result, that is to say, mRNA secondary structure is an important factor for discovering human encoded miRNAs that regulate the influenza virus genome. But if only step 2 is used to get the last result, the number of miRNAs may be little, especially for the segment that has the more binding sites. Three miRNAs are missing for HA, five are missing for PB2, one is missing for MP and zero is missing for NS. Consequently, step 1 and step 2 are all necessary. Five marks and ten marks are used in step1 and step 2 respectively.

IV. RESULTS AND DISCUSSION

Has-miR-489, has-miR-325, has-miR-876-3p and has-miR-2117 are targeted HA, PB2, MP and NS of influenza A, respectively. Table II and Table III provide start position of binding and energy, the results of binding pairs. The number of binding sites of has-miR-489 is least, and only the seed region is paired. Because the binding energy of has-miR-325 and PB2 is the lowest, their binding is the most stable. Figure 1 shows the distribution of complementary sites. The horizontal axis is the order number of human encoded miRNAs and the ordinate axis is the site number of segments. The positions of binding are showed in Figure 1. X-coordinate is the number of miRNAs and y-coordinate is the position of binding sites. There are the most complementary sites in PB2 segment and the least in NS segment. This means PB2 may be the most important segment for the binding of human encoded miRNAs and H1N1 segments.

Scoring is the core of the method, and it is based on different things that are important factors to obtain the exact results. That sequences from ten years make the results more convinced, although the end results are indicated with the sequences from 2009. Moreover, RNA secondary structure is considered. The lowest free energy which is usually taken as a key factor is not considered in this paper, which is only used in the process of choosing the end result. But we focus the bind sites are whether in the stem of RNA secondary structure. The secondary structure of any single sequence is not representative; as a result, the consensus one is used. Because the stem is the stable composition, we believe that if the bind sites are in the stem, miRNA will break mRANs directly. This idea is introduced for the first time.

TABLE II. THE INFORMATION OF START POSITION OF BINDING AND BINDING ENERGY

miRNA	Start position of binding	Energy(kcal/mol)
has-miR-489	151	-11.19
has-miR-325	1029	-13.72
has-miR-876-3p	526	-13.23
has-miR-2117	827	-11.74

TABLE III. THE INFORMATION OF BINDING PAIRS

miRNA	Binding pairs
has-miR-489	CACTGTAGACACAGTACTAGAA ||||||| GUGACAUCACAUAUACGGCAGC
has-miR-325	GGATCATCA----GTCAAGAGAGAAGA |||||||| |||| CCUAGUAGGUGUCCAGUAAGUGU----
has-miR-876-3p	ACCACCAATCCACTAATCAGGC |||||||| || ||| UGGUGGUUUACAAAGUAAUUCA
has-miR-2117	ACAAGAGATAAGAGCTTTCTC-- |||||||| ||||| UGUUCUCUUUGCC--AAGGACAG

We use the different scores for the test to determine which step has significant effect on the result. The test confirms that mRNA secondary structure is an important factor for discovering human encoded miRNAs that regulate the influenza virus genome. But the number of miRNAs is affected by step 1, especially for the segment that has the more binding sites.

The chief purpose of our study is to obtain the human encoded miRNAs by a prompt and accurate method. The miRNAs target influenza virus genome to cleave mRNA or repress translation, and virus infection and replication are broken. The binding sites provide the pharmaceutical industry with important molecular targets for drug development. But we are not yet in possession of reliable information about the expression of the miRNAs, and at present our knowledge of interaction of host cell and influenza virus is still uncompleted. It reqires much further research in this field.

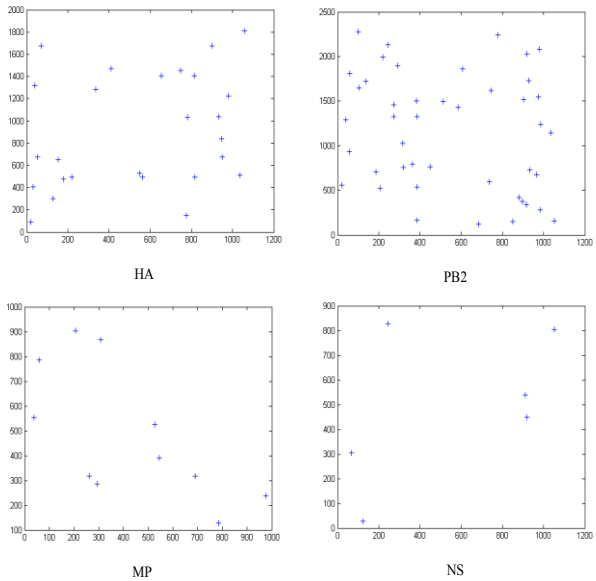

Figure 1. The profile of complementary sites in H1N1 genome

REFERENCES

[1] Alan,J.H. et al. (2001) The evolution of human influenza viruses. Philos Trans R Soc Lond B Biol Sci, 365(1416), 1861-1870.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

[2] Alicia,E. S and Ari,H. (2004) How viruses enter animal cells. Science, 304, 237-242.

[3] Ayae,H. et al. (1990) Purification and Molecular Structure of RNA Polymerase from Influenza Virus A/PRS. J. Biochem, 107, 624-628.

[4] Beatriz, P and Juan,O. (1997) The influenza A virus PB2 polymerase subunit is required for the replication of viral RNA. J.Virol., 71, 1381-1385.

[5] Benjamin ,P.L. et al. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120, 15-20.

[6] Charles-Henri ,L. et al. (2005) A cellular microRNA mediates antiviral defense in human cells. Science, 308(5721), 557- 560.

[7] Clors ,F.A. et al. (2009) Molecular Anatomy of 2009 Influenza Virus A (H1N1). Arch. Medical Res, 40, 643-654.

[8] David, P.B. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281-297.

[9] Detjen ,B.M. et al. (1987) The three influenza virus polymerase (P) proteins not associated with viral nucleocapsids in the infected cell are in the form of a complex. J. Virol, 61, 16-22.

[10] Ismo, U. et al. (1981) Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription. PNAS, 78, 7355-7359.

[11] Joseph,M.W et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. Nature, 460, 711-719.

[12] John,C.K et al. (2006) Hijacking of the host-cell response and translational control during influenza virus infection. Virus Res., 119, 111-120.

[13] Licheng. S. et al. (1995) Influenza A Virus RNA Polymerase Subunit PB2 Is the Endonuclease Which Cleaves Host Cell mRNA and Functions Only as the Trimeric Enzyme. Virology, 208,38-47.

[14] Manoj,H. et al. (2005) Targets for human encoded microRNAs in HIV genes.BBRC, 37(4), 1214-1218.

[15] Nakajima,K. (1997) Influenza virus genome structure and encoded proteins. Nippon Rinsho, 55, 2542-2546.

[16] Noam,S.(2010)MicroRNAs and pharmacogenomics. Pharmacogenomics, 11, 629-632.

[17] Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team (2009). Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med, 360: 2605-2615.

[18] Pierre,M. and Anton ,J.E. (2007) Prediction of miRNA targets. Drug Discovery Today, 12, 11-12.

[19] Rachel ,R.H. et al. (2009) Differential patterns of amantadine-resistance in influenza A (H3N2) and (H1N1) isolates in Toronto, Canada. J. Clin. Vir., 44, 91-93.

[20] Sam ,G.-J. et al. (2007) miRBase: tools for microRNA genomics. Nucleic Acids Res., 1-5.

[21] Susumu, K. et al. (1983) An outbreak of type C influenza in a children's home. J.Inf.Dis, 148, 51-56.

[22] Terrence ,M.T et al. (2005) Diagnostic Approach for Differentiating Infected from Vaccinated Poultry on the Basis of Antibodies to NS1, the Nonstructural Protein of Influenza A Virus. J. Clin. Microbiol., 43, 676-683.

[23] Vinod,S. et al. (2006) Host-virus interaction: a new role for micorRNAs. Retrovirology, 3: 68-76.

[24] Yoshiki,M. et al. (2008) Regulation of the hepatitis C virus genome replication by miR-199a*. J. Hepatol., 50(3), 453-460.

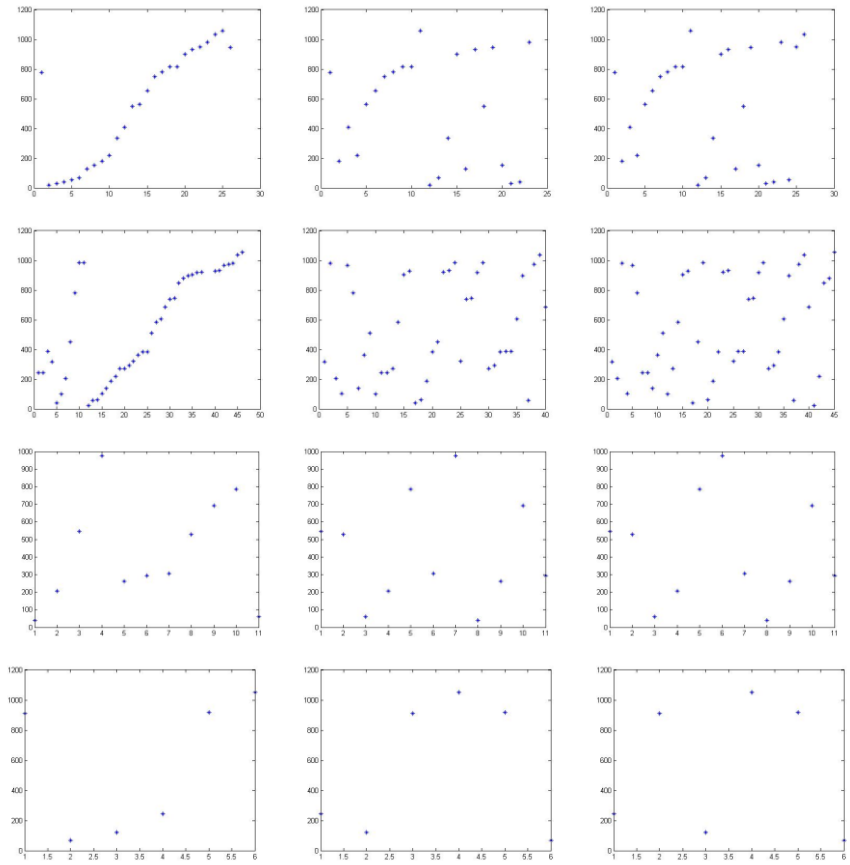

Figure 2. The profile of different scores of HA, PB2, MP and NS respectively

978-1-4673-4398-5/12/$31.00 ©2012 IEEE