ISSN:2153-0602

Journal of Data Mining in Genomics & Proteomics

The International Open Access
Journal of Data Mining in Genomics & Proteomics

Available online at: OMICS Publishing Group (www.omicsonline.org)

Digital Object Identifier: http://dx.doi.org/10.4172/2153-0602.1000104

Journal of Data Mining in Genomics & Proteomics - Open Access

www.omicsonline.org

Research Article
OPEN ACCESS Freely available online
doi:10.4172/2153-0602.1000104

Prediction of micro RNAs against H5N1 and H1N1 NS1 Protein: a Window to Sequence Specific Therapeutic Development

Pankaj Koparde[1] and Shailza Singh[2]*

[1]*Institute of Bioinformatics and Biotechnology, University of Pune, Pune 411007, India*
[2]*National Centre for Cell Science, NCCS Complex, Pune University Campus, Pune-411007, India*

Abstract

Motivation: Conventional Influenza drugs might result in failure of action due to highly variable nature of influenza antigenic proteins. *in silico* prediction of sequence specific therapeutics is less time consuming and cost effective. Most of the miRNA target prediction platforms differ in their specificity and sensitivity in prediction. Here we propose a bioinformatics approach to predict micro RNAs against non-structural protein 1 of avian influenza H5N1 virus and against segment 8 swine influenza H1N1 virus.

Results: hsa-miR-138, hsa-miR-525-5p and hsa-miR-124 were sorted out as potential sequence specific therapeutic agents against NS1 protein of H5N1 virus. Similar studies on different H1N1 segment 8 genomes resulted in the prediction of miRNAs which can be used as anti-influenza agents. The predicted micro RNAs also possess roles in regulatory processes of host cell, stress related pathways including MAPK pathway, mTOR pathway and are found to be involved in pathways related to cancer.

Introduction

Influenza is a disease of concern world-wide. Recent pandemic of swine influenza (H1N1) virus has raised questions regarding efficacy of current chemotherapy against influenza viruses. It is known that birds especially wetland birds such as ducks and geese are the reservoirs of influenza viruses [1-3]. The spread of influenza viruses to humans is from wetland birds to poultry birds via poultry animals such as pigs [3]. 1996 avian influenza (AI) H5N1 virus was a highly pathogenic influenza virus but it was largely restricted to poultry birds. The potential of AI H5N1 virus to cause a pandemic was speculated earlier and hence many groups started working on H5N1 inhibitory chemotherapy largely focusing on specific therapeutics. Current chemotherapeutic agents used as anti-influenza medicines are protein inhibitors such as neuraminidase inhibitor and M2 ion channel inhibitors.

These drugs show good results as an anti-influenza drug but they are disadvantageous to the patient in terms of severe side effects they might cause [4,5]. Sequence specific therapeutics (SSTs) is the latest addition to anti-influenza drugs. Advantages of SSTs include reliability and specificity in their action. Previously small interfering RNAs (siRNAs) have been shown to be effective against influenza [5-8]. Although siRNAs and miRNAs do not differ much in their physicochemical properties, the source of siRNAs is generally exogenous as compared to the endogenous source of miRNA biogenesis [9,10]. These drugs show good results as an anti-influenza drug but they are disadvantageous to the patient in terms of severe side effects they might cause [4,5]. Endogenous sources of action are useful because they are key elements in gene regulation. miRNAs are important in gene regulation, as they act at sequence level either by degrading mRNA of target protein or by causing translational repression of target protein by inhibiting mRNA of target protein.

NS1 protein

Influenza A virus contains eight segments of negative sense linear RNA genome [11,12]. These 8 segments code for 11 different proteins [12]. Hemagglutinin (HA) and Neuraminidase (NA) are the major antigens of the virus. Other than these proteins non-structural protein 1 (NS1) of segment 8 of AI H5N1 virus have been discovered to effect viral propagation by inhibiting host cell's antiviral pathways [13-18]. NS1 protein binds double stranded (ds) RNA of virus during its replication thereby preventing it from undergoing degradation mediated by RNA interference pathway of host cell (supplementary files) [19]. It also inhibits synthesis of interferons, which are key molecules of antiviral pathway. Palese et al. [20] found that repression of NS1 protein leads to reduction in viral propagation. This makes NS1 a very good target to raise sequence specific therapeutics against for.

miRNA prediction and filtering

A variety of web platforms can be used for prediction of miRNAs and their target genes. Although recent developments in the field of sequence specific therapeutics and bioinformatics has resulted in the release of miRNA prediction and target prediction web platforms, many of these possess innate problems with the low specificity and sensitivity in prediction [21]. Also, a variety of factors such as miRNA binding site location (3'UTR, CDS, 5'UTR), number of G:U wobble base pairs in the seed region of miRNA, the kinetics and thermodynamics of miRNA and its target interaction are important to understand the scope of miRNA functionality [22]. Common miRNAs and common targets can be found out by using filtering methods. Thermodynamic stability of miRNA molecules can be deduced from minimum folding energy (MFE) calculations. MFE values and thermodynamic stability of miRNA and target RNA duplex formation can be used to understand

***Corresponding author:** Shailza Singh, National Centre for Cell Science, NCCS Complex, Pune University Campus, Pune-411007, India, E-mail: shailza_iitd@yahoo.com

Received September 20, 2010; **Accepted** November 13, 2010; **Published** November 15, 2010

Citation: Koparde P, Singh S (2010) Prediction of micro RNAs against H5N1 and H1N1 NS1 Protein: a Window to Sequence Specific Therapeutic Development. J Data Mining in Genom Proteomics 1:104. doi:10.4172/2153-0602.1000104

efficacy of miRNAs. This reduces noise in the prediction and time consumption. The experimentally validated targets of miRNAs provide reliable data to build up regulatory networks of miRNAs. Although *in sillico* prediction reduces time in SST development, experimental validation of predicted miRNAs is important to fully understand miRNA action on target genes [23].

Regulatory networks

miRNAs are diverse in terms of their action in gene regulation. They show multiplicity and cooperativity in their action; this means a single miRNA may possess multiple targets, at the same time; multiple miRNAs may regulate a single gene [24]. This diverse nature of action of miRNAs is important in developing regulatory networks on miRNA actions. Filtering at target prediction level is useful to avoid false positives. The regulatory networks may be qualitative, commenting only on the interactions of miRNA and its targets or they may be quantitative, commenting on kinetics and other details about the miRNA and its targets.

Methods

Selection of genomes

For H5N1 (A/Goose/Guangdong/1/96 (H5N1)) segment 8 was selected as it was the only complete sequence found in Genbank. For H1N1 14 sequences from different strains world-wide were selected randomly, most of them being from Asiatic countries. The complete list of genomes and their accession is provided as supplementary information.

Prediction of miRNAs

Softberry findmiRNA (Softberry webserver) and miReval Ritchie et al. [25] web platforms were used to predict sequences from segment 8 of H5N1 and H1N1 viruses. The predicted sequences were searched in miRBase for getting homologous metazoan miRNAs. These miRNAs were sorted out according to their e values. Only *Mus musculus* and *Homo sapiens* miRNAs were selected. Common miRNAs from predictions were sorted out for further analysis. Sequences of pre-cursors of miRNAs were obtained from miRBase [26].

Analysis of miRNAs and pre-miRNAs

Length distribution and % sequence position analysis for miRNAs and pre-miRNAs were done. To avoid noise caused by lengths of pre-miRNAs, optimal lengths were considered for sequence analysis. For each sequence cumulative data of A, U, G, C content was obtained from RNAfold web platform. Statistical analysis including mean, maximal, minimal, standard deviation, covariance and correlation was carried out.

Filtering of miRNA prediction

Total of five filters were used to sort out miRNAs with therapeutic potential. The first filter used is determination of thermodynamic stability of miRNAs based on MFE values [27,28]. miRNAs falling below threshold value were removed. The second filter used is genomic locations of pre-miRNAs [29] and miRNA expression profile data [30]. According to this filter, miRNA genes located on sex chromosomes were removed as they cannot be used therapeutics. UCSC miRNA gene location maps are provided in the supplementary information files. Third filter setup was that of analysis of binding site. Here, target sites predicted by miRNA and target alignment using ClustalW were subjected to MFE value determination. To avoid variation in result due to variable lengths of target site bindings, AMFE values were considered. AMFE was calculated using following equation Zhang *et al.* [31] :AMFE = (MFE/Length of pre-miRNA sequence)*100 According to the third filter, relatively less stable binding sites and their respective miRNA pairs were removed. Fourth and fifth filters consisted of miRNA and target interaction analysis. Fourth filter is PAIRFOLD analysis Andronescu et al. [32] and fifth filter is RNAHybrid analysis [33]. The third, fourth and fifth filters were applied to both the target sequences NS1 and segment 8 for H5N1 predicted miRNA binding. Whereas filters 3-5 were applied only for segment 8 sequences of different H1N1 strains for H1N1 predicted miRNA binding, because complete sequences of NS1 of chosen H1N1 strains are not available.

Filtering of target prediction

Total of six filters were used to sort out possible targets of predicted miRNAs. These include DIANA micro T 3.0 strict, DIANA micro T 3.0 loose, DIANA micro T 4.0 beta, PicTar, Target ScanS Papadopoulos et al. [34] and miRGen (Intersection of miRanda, PicTar and TargetScanS) [35]. Genes common to 4 or more filters were considered as significant targets. The filtering was also setup according to the –ln (p-value) of prediction web platforms.

Qualitative miRNA network construction

Sorted out genes after target prediction filtering was considered for miRNA regulatory network construction. Common target genes for predicted miRNAs were sorted out and their involvement in various pathways was searched accordingly. Following is the flowchart made by our group to predict miRNAs against virulent proteins.

Results

Segment 8 of H5N1 shows distant homology with human genes

The source genome of H5N1 selected was Influenza A virus (A/Goose/Guangdong/1/96 (H5N1)). Segment 8 when aligned with non redundant protein database of human, shows distant homology with Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), heparin sulfate N-deacetylase/N-sulfotransferase (NDST3) and PRKCA-binding protein (PICK-1) proteins. Out of which GAPDH is homologous with NS2, while NDST3 and PICK-1 proteins are found to be distantly homologous with NS1 protein of H5N1. These proteins hence were used for miRNA prediction against segment 8. Homologous sequences showed presence of putative domain of NS superfamily protein.

Prediction of micro RNAs

Total of 22 miRNAs were predicted using various miRNA prediction algorithms, out of which 16 were metazoan and 6 were viral. Predicted miRNAs showed involvement in various cell signaling and regulatory pathways.

Percentage sequence composition and length distribution of miRNAs

Predicted miRNAs show significant skew of 25% each. In case of metazoan miRNAs, Guanine (G) dominates over other bases followed by adenine (A) (Figure 1a). Guanine was found to be dominant (30.38%) at fifth (50%) and thirteenth (63%) as well as all the positions from 21-23. Whereas adenine was found to be dominant (24.70%) at positions 1-3 and at fifteenth (56%) position. Highest enrichment of guanine was found to be at positions 5, 13 and 23. Highest enrichment of adenine is at positions 3 and 15. In case of viral miRNAs (Figure 1b), Uracil (32.60%) was found to be dominant at positions 10, 11, 15-

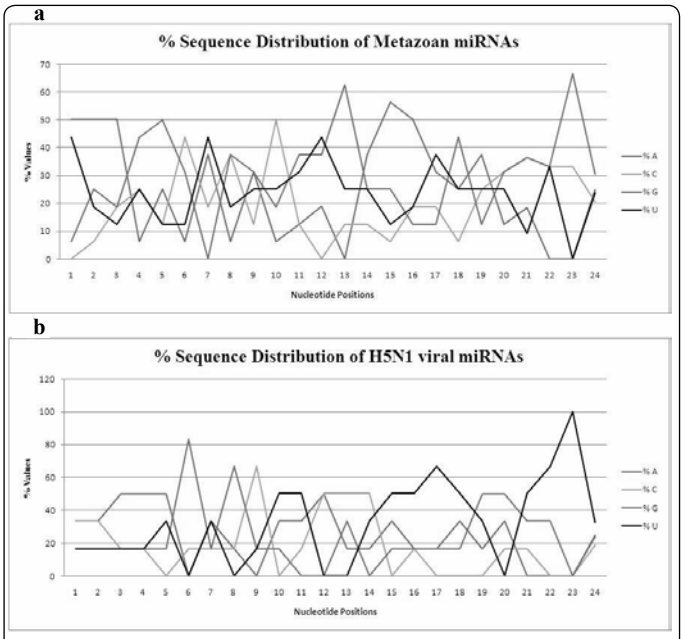

Figure 1: Percentage sequence distribution of miRNAs.

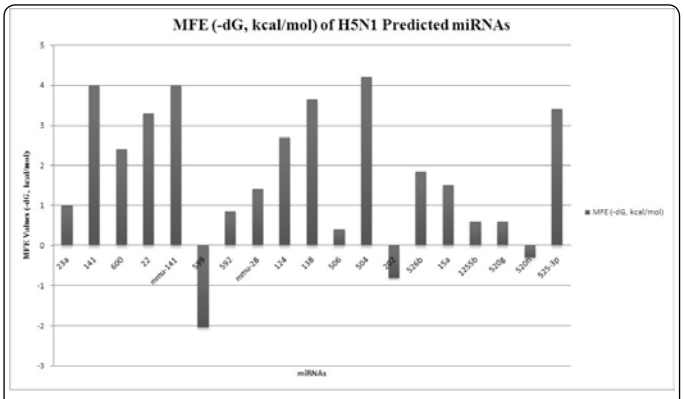

Figure 2: Minimum Folding Energy values for predicted miRNAs.

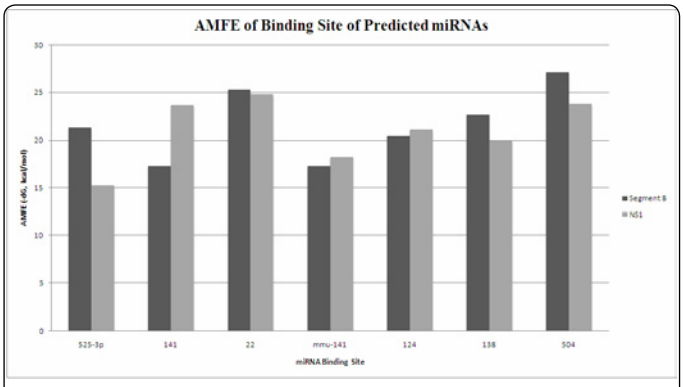

Figure 3: Adjusted MFE values of binding sites of predicted miRNAs.

18 and positions 21, 22 followed by adenine (24.63%) at positions 3-5. Highest enrichment of Uracil was found to be at seventeenth nucleotide (67%) and that of adenine at fifth nucleotide position (50%). The length of predicted miRNAs varied from 19-23 nucleotides with an average of 22 nt. The cumulative average value of 22 nt length was found to be 43.75% (supplementary files).

Percentage sequence composition and length distribution of pre-miRNAs

Precursors of predicted metazoan miRNAs showed a length distribution varying from 72 to 124 nucleotides. To avoid the noise during percentage sequence distribution analysis sequence length of 85 nucleotides was considered. For precursors of predicted viral miRNAs the sequence length of 83 nucleotides was considered for sequence analysis. Both the viral and metazoan pre-miRNA sequences showed Uracil as the predominant base, for which percentage values were found to be 26.91% for metazoan pre-miRNAs and 30.63% for viral pre-miRNAs (supplementary files). The next predominant base found was guanine which differed slightly from average Uracil composition. For metazoan pre-miRNAs all the bases were distributed evenly with a standard deviation of 2.28 and that of viral pre-miRNAs were less evenly distributed with a standard deviation of 7.84. The percentage values of AU and GC content varied largely for metazoan and viral pre-miRNAs. For metazoan pre-miRNAs GC content (51.10%) was higher than AU content (49.04%); whereas for viral pre-miRNAs GC content (42.16%) was found to be lower than AU content (57.22%). Although GC and AU content showed variation in percentage content value, it was found to follow a certain range for both metazoan [(66.91±0.06)%] and viral (83.33%) pre-miRNAs. The overall % GC and % AU content was found to be (49.91±0.3) %. The variation in dinucleotide content and in nucleotide base composition relates with stability and thermodynamic parameters of pre-miRNA sequences, in terms of MFE and AMFE values.

MFE values of miRNAs

MFE values of miRNAs acts as the first filter to sort out miRNAs. MFE values of predicted miRNAs range from -4.2 to 2.04 kcal/mol (Figure 2.0). Average value of MFE given by these miRNAs is -1.844 kcal/mol. The maximum value of MFE is given by hsa-miR-504 (-4.20 kcal/mol). Hsa-miR-504, hsa-miR-141, mmu-miR-141, hsa-miR-22, hsa-miR-138, hsa-miR-124 and hsa-miR-525-3p can be considered for further analysis, as according to the MFE data, their structures are relatively stable than other miRNAs. All other miRNAs show less MFE values suggesting that these miRNAs are less stable as compared to the others.

miRNA locations and their tissue level expression

Although hsa-miR-504 forms the most stable structure, the gene coding for hsa-miR-504 is located of X chromosome. This restricts action of hsa-miR-504 as a therapeutic agent. Studies using mimiRNA web server showed that hsa-miR-504 is maximally expressed in testis, placenta and heart tissue. A wide distribution of expression of predicted miRNAs was found in a variety of cell types and disease conditions such as teratocarcinoma and other types of cancer. Heart cells are also a common location for many miRNAs. mimiRNA could not provide information on hsa-miR-22 and hsa-miR-124.

Analysis of binding site

For a miRNA to work efficiently the binding site of the targets sequence should be stable enough. Analysis of binding site (Figure 3.0) showed the the stability of binding site. All the sorted miRNAs showed relatively stable binding sites. Most of the binding site AMFEs followed a threshold of -15.0 kcal/mol; suggesting that most of the binding sites for sorted miRNAs are stable in nature. The most

stable binding site was shown by hsa-miR-504 target site (-27.14 kcal/mol for segment 8 binding and -23.81 kcal/mol for NS1 mRNA binding)) followed by has-miR-22 target site (-25.27 kcal/mol for segment 8 binding and -24.75 kcal/mol for NS1 mRNA binding). Mmu-miR-141 showed lowest values of binding site AMFE (-17.2 kcal/mol for segment 8 binding and -18.20 kcal/mol for NS1 mRNA binding). All the miRNAs other than mmu-miR-141 hence were sorted for further filtering.

PAIRFOLD and RNAHybrid analysis

miRNA and its target interactions were studied by using PAIRFOLD, Softberry TargetmiRNA and RNAHybrid web servers (Figure 4a, 4b). These web tools were used as the third filter to sort out miRNAs for therapeutic use. PAIRFOLD energies showed that ΔMFE values are maximum for hsa-miR-138 and its target pairs (-162.2 kcal/mol) followed by hsa-miR-124 (-160.5 Kcal/mol for segment 8, -159.5 kcal/mol for NS1). RNA Hybrid showed maximal values of pairing for hsa-miR-124 (-27.4 kcal/mol) followed by hsa-miR-525-3p (-26.2 kcal/mol) and hsa-miR-138 (-23.2 kcal/mol). This suggests hsa-miR-138, hsa-miR-124 and hsa-miR-525-3p show stable interaction with their target pairs.

Qualitative miRNA network construction

Selected miRNAs with sequence specific therapeutic potential shows interactions with many genes. miRNAs show cooperativity and multiplicity in their function; hence a single miRNA may possess more than one target as well as a single gene may get regulated by multiple miRNAs. To understand the scope of selected miRNA action, a qualitative network of miRNA involvement in gene regulation was constructed (Figure 5.0). Hsa-miR-138 shows targets in axon guidance and small cell lung cancer, whereas hsa-miR-124 shows targets in variety of cancer types such as chronic myeloid leukemia, glioma, acute myeloid leukemia, prostate cancer, melanoma, non small cell lung cancer, pancreatic cancer, colorectal cancer and endometrial cancer as well as in gap junction, insulin signaling pathway and MAP kinase pathway. Hsa-miR-525-3p possesses targets in axon guidance and melanoma. miRNA network construction of predicted miRNAs thus suggests roles of miRNAs in cell regulatory pathways and pathways related to cancer. Upregulation or downregulation of hsa-miR-124 may result in drastic change in cell physiology at the same time the role of hsa-miR-124 in variety of cancers implies that hsa-miR-124 might possess important roles in cancer down regulation of genes involved in cancer.

Studies on H1N1

Segment 8 do not show homology with human genes: Considering world-wide impact of recent H1N1 virus, 14 segment 8 genomes from 14 different H1N1 strains isolated from different parts world-wide were used for analysis.

Prediction of micro RNAs: Total of 5 miRNAs which were found common in predictions from all the 14 sequences were used for further filtering. These include hsa-miR-520c-3p, hsa-miR-520f, hsa-miR-520b, hsa-miR-548d-3p and hsa-miR-1302.

Percentage sequence composition and length distribution of miRNAs

Sequence distribution data showed that Uracil predominates at position 13 and its maximum value found is 100%. Also, cumulative data of percentage values of nucleotide sequences shows that Uracil occurs most commonly with an average value of 35.75% followed by Adenine (25.45%). This is in favor of (A+U)%; where it is found to be 61.21% as compared to (G+C)% which is 38.78%. Highest enrichment of Uracil was found to be at position 13 (100%), whereas that of adenine was found to be at position 2 (60%). The length of predicted miRNAs varied between 21-22 nucleotides with an average of 22 nt. The cumulative average value of 22 nt length was found to be 60% (supplementary files).

Percentage sequence composition and length distribution of pre-miRNAs

Precursors of predicted miRNAs of 1302 family showed a length distribution varying from 72 to 151 nucleotides with an average value of 126 nucleotides, whereas precursors of predicted miRNAs of other than 1302 family showed length distribution varying from 61-97 with an average value of 86 nucleotides (supplementary files). To avoid the noise during percentage sequence distribution analysis, sequence length of 86 and 126 nucleotides was considered for pre-miRNAs other than 1302 family and for 1302 family pre-miRNAs respectively. 1302 family pre-miRNAs showed adenine as the predominant base (34.60%) followed by uracil (34.50%), whereas pre-miRNAs other than 1302 family showed Uracil as the predominant base (30.70%) followed by adenine (27.23%). For pre-miRNAs other than 1302 family all the bases were distributed relatively evenly with a standard deviation of 5.68 whereas those of 1302 family pre-miRNAs were less evenly distributed with a standard deviation of 11.17. The percentage values of AU and GC content did not show large variation for both the types. %AU content was found to be higher for both the types as compared to %GC content. 1302 family pre-miRNAs showed higher % value of AU content (69.10%) than other pre-miRNAs (57.94%); whereas %GC content for pre-miRNAs other than 1302 family showed higher %

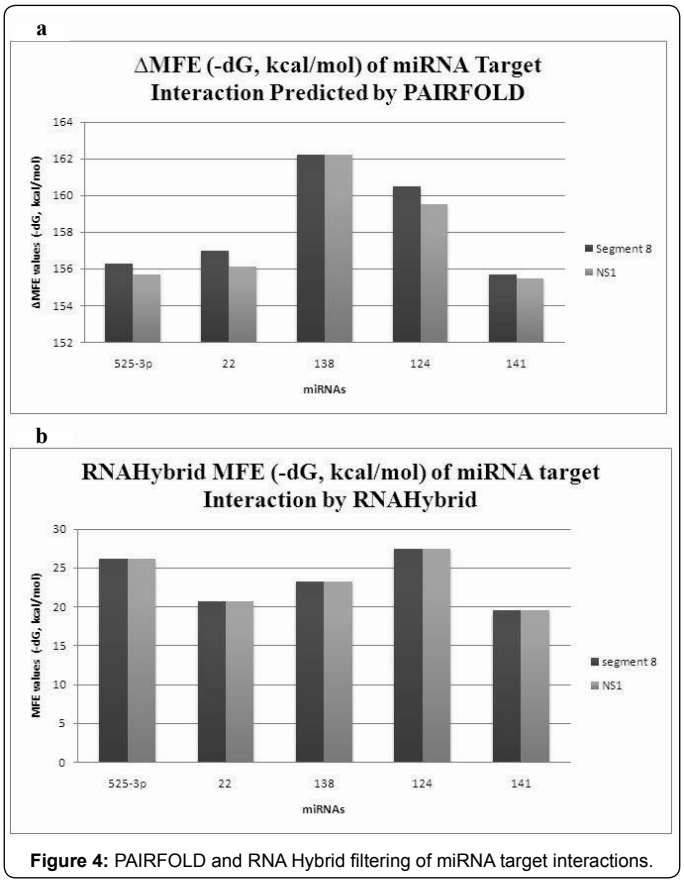

Figure 4: PAIRFOLD and RNA Hybrid filtering of miRNA target interactions.

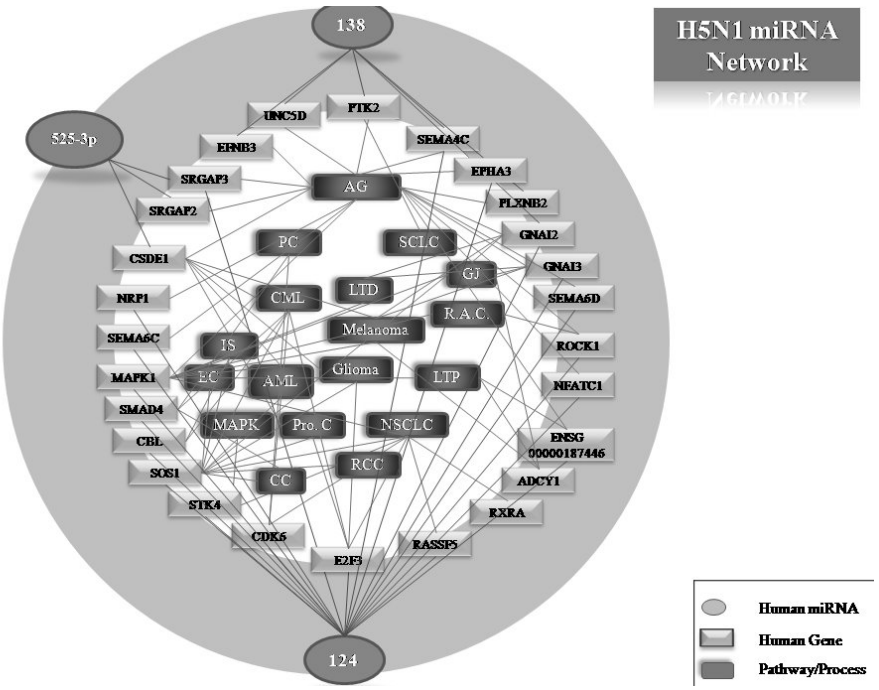

Figure 5: Qualitative miRNA network of predicted miRNAs. (AG: Axon guidance, PC: Pancreatic cancer, SCLC: Small cell lung cancer, CML: Chronic myeloid leukemia, GJ: Gap junction, R.A.C.: Regulation of actin cytoskeleton, LTD: long term depression, Pro.C: Prostate cancer, IS: Insulin signaling, AML: Acute myeloid leukemia, NSCLC: Non small cell lung cancer, EC: Endometrial cancer, CC: Colorectal cancer, LTP: Long term potentiation, RCC: Renal cell carcinoma).

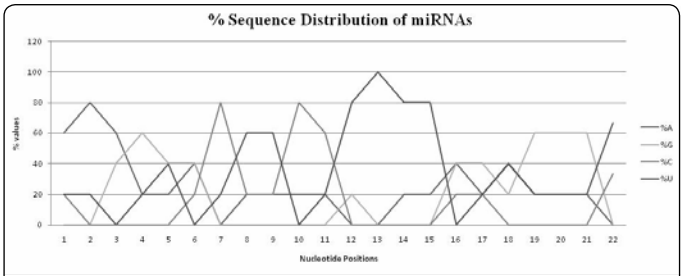

Figure 6: Percentage sequence distribution and length distribution of predicted miRNAs.

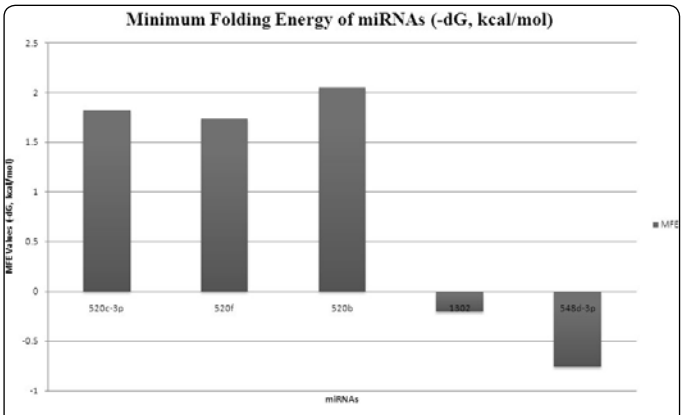

Figure 7: MFE values of predicted miRNAs.

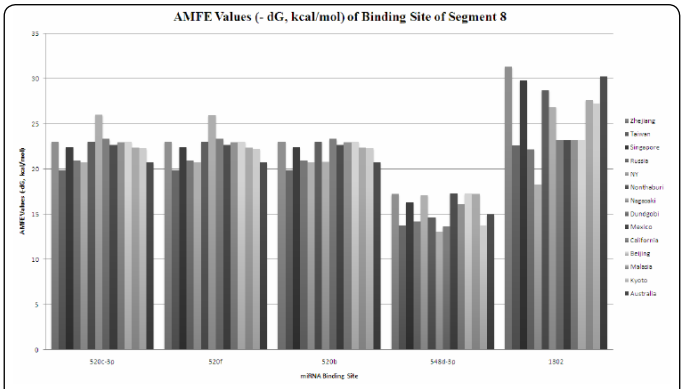

Figure 8: AMFE values of binding sites of miRNAs.

value (42.059%) than 1302 family pre-miRNAs (30.46%). Although GC and AU content showed variation in % value, it was found to follow a certain range for both 1302 family [(73.21±1.79) %] and pre-miRNAs other than 1302 family (100%).

MFE values of miRNAs

MFE values of predicted miRNAs range from -2.05 to 0.76 kcal/mol (Figure 7.0). Average value of MFE given by these miRNAs is -0.93 kcal/mol. The maximum value of MFE is given by hsa-miR-520b (-2.05 kcal/mol), followed by hsa-miR-520c-3p (-1.82 kcal/mol) and hsa-miR-520f (-1.74 kcal/mol). Hsa-miR-1302 and hsa-miR-548d-3p both show positive values of MFE. In general 520 series miRNAs are more stable than other miRNAs.

miRNA locations and their tissue level expression

All the predicted miRNAs show moderate expression levels for various cell types and tissues such as breast tissue, brain, heart tissue and liver. Hsa-miR-520f shows high expression in breast malignant tumor cells as well as moderate expression level in thymus, whereas other miRNAs show poor to moderate expression in various tissues. 520 series of miRNAs are located on chromosome 19, that of 548

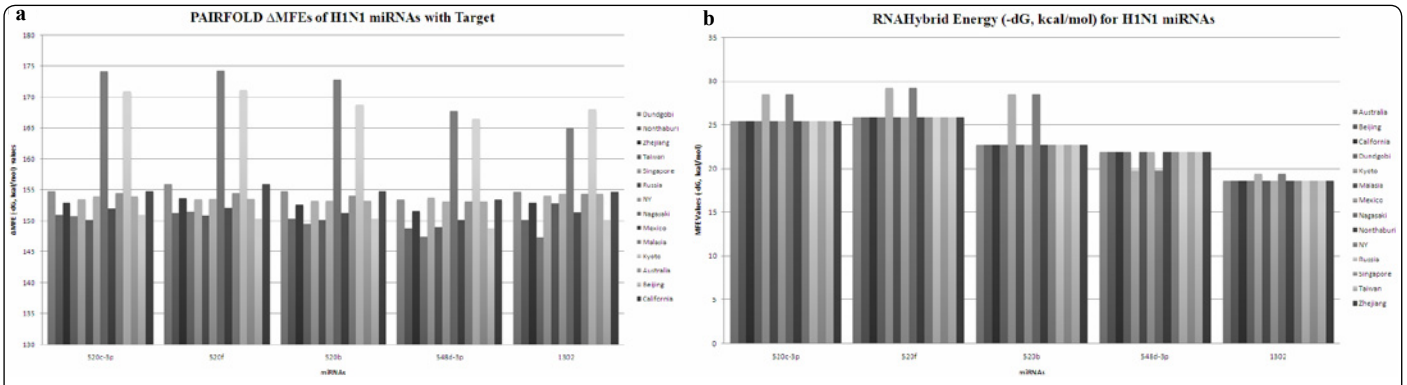

Figure 9: PAIRFOLD and RNA Hybrid filtering of miRNA target interactions.

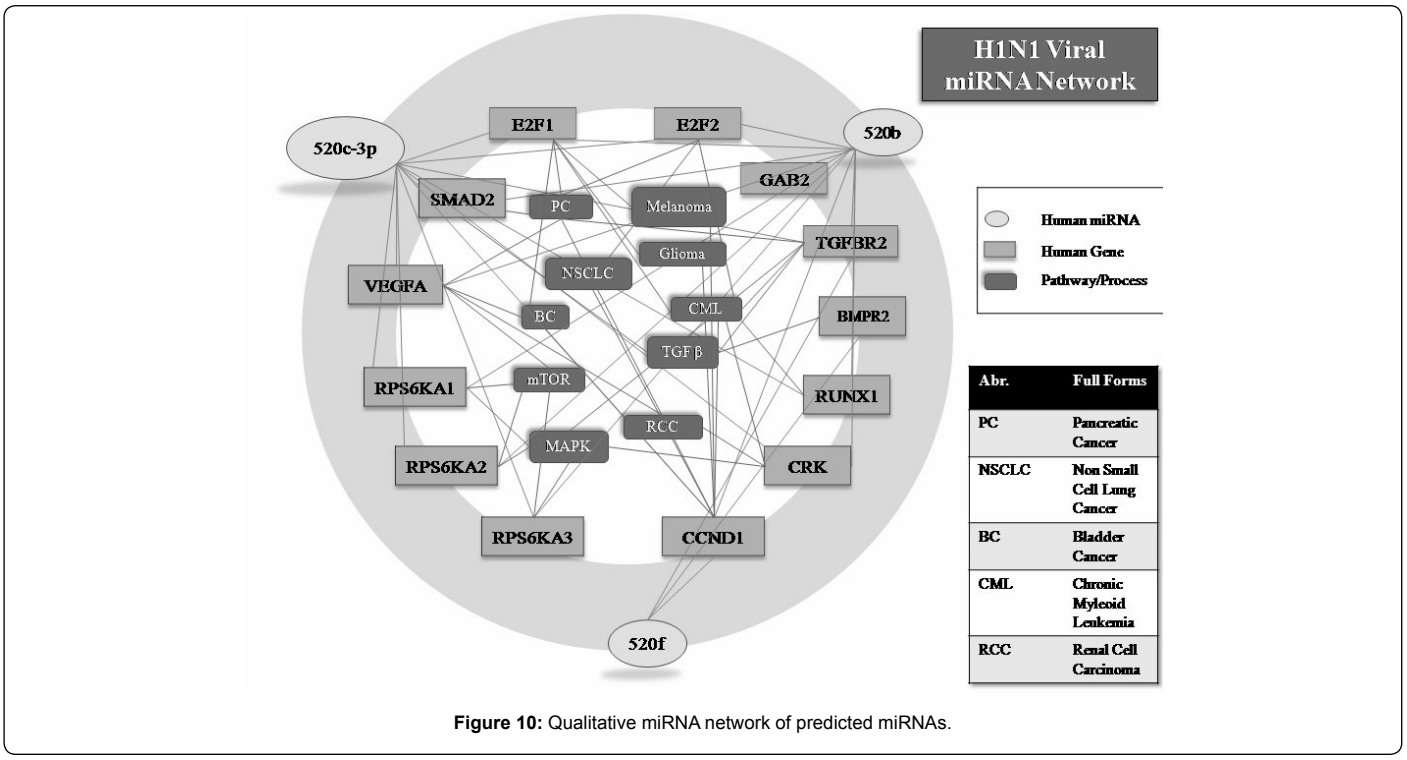

Figure 10: Qualitative miRNA network of predicted miRNAs.

are distributed on chromosome 8 and 17; whereas the 1302 family of miRNA genes are located on chromosomes 2, 8, 9, 7, 17 and 20. No miRNAs from predicted list is located on sex chromosomes. Expression data for hsa-miR-1302 is lacking.

Analysis of binding site

For 520 series miRNAs the AMFE values of binding sites coincided for all the strains (Figure 8.0). The maximum value was found to be -31.3 kcal/mol for hsa-miR-1302 for Zhejiang strain. The AMFE values varied from -13.1 to -31.3 kcal/mol. lowest values of AMFE are given by hsa-miR-548d-3p.

PAIRFOLD and RNA hybrid analysis

PAIRFOLD (Figure 9a) and RNA Hybrid (Figure 9b) MFE values of miRNAs and their respective targets pairs showed that 520 series of miRNAs show maximum values, specifically for Nagasaki and Kyoto strains. The maximum value of MFE (RNA hybrid) is given by hsa-miR-520f (-29.2 kcal/mol) for both Nagasaki and Kyoto strains, followed by hsa-miR-520c-3p and hsa-miR-520b (-28.5 kcal/mol). Whereas maximum ΔMFE values (PAIRFOLD) are given by hsa- miR-520f (-174.2 kcal/mol) followed by hsa-miR-520c-3p (-174.1 kcal/mol) for Nagasaki strain. This infers that hsa-miR-520f and hsa-miR-520c-3p can be used as the potential sequence specific therapeutics.

Qualitative miRNA network construction

The potential sequence therapeutic miRNAs predicted against NS1 of H1N1 possess roles in gene regulation. To understand a qualitative miRNA network was constructed (Figure 10), which showed that 520 series miRNAs, possess targets in various cell regulatory pathways such as Map Kinase pathway, mTOR pathway and TGF β pathway. These miRNAs possess targets in various cancer types such as glioma, melanoma, chronic myeloid leukemia, pancreatic cancer etc. miRNA network construction provides an insight regarding miRNA and target gene interaction as well as possible impact on gene regulation and host physiology when miRNA expression is altered or changed due to artificial conditions or by natural cellular processes.

Correlations between H5N1 and H1N1 predicted miRNAs

All the predicted miRNAs show optimal value of length of 22

nucleotides. Although guanine predominates in H5N1 metazoan miRNAs (30.38%), Uracil predominates in all other miRNAs with a value of 32.60% for H5N1 viral miRNAs and 35.75% for H1N1 miRNAs. Adenine is the second most abundant base in all the miRNAs and its enrichment is found to be at positions 1-3 and 5 for metazoan miRNAs, 3-5 for viral miRNAs and position 2 for H1N1 miRNAs. This nucleotide position falls in the seed region of miRNAs, which is the primary determinant of miRNA functionality. This suggests that adenine might be important base for determining miRNA functionality in case of predicted miRNAs. Uracil was found predominant base in all the precursors of predicted miRNAs. The second most abundant base found was guanine for H5N1 pre-miRNAs and adenine for H1N1 pre-miRNAs. This has reflected in the variable GC and AU content of the predicted pre-miRNAs. Also, it was found that standard deviation for pre-miRNA sequence composition was highest for 1302 family pre-miRNAs (11.17) , followed by H5N1 viral pre-miRNAs (7.84). To understand the correlation of Uracil predominance and MFE values Karl pearson's correlation coefficient was calculated for %U content and MFE values. %U content and AMFE values show a significant negative correlation (-0.5082) and covariance (-25.02) for all predicted pre-miRNAs from both H5N1 and H1N1 viruses, whereas %GC and AMFE values show significant positive correlation (0.49) and covariance (48.97) values. This suggests that thermodynamic stability of a pre-miRNA molecule may decrease with increasing amount of Uracil and with decreasing amount of GC content.

Discussion

miRNAs are key regulators in gene regulation networks found in metazoan and viruses. miRNAs essentially play roles in regulation of viral replication and propagation by regulating host gene regulatory mechanisms. The action of miRNAs is specific as they act at sequence level which is the crucial stage in the central dogma. Viruses, especially RNA viruses such as influenza viruses show highly variable nature of antigenic proteins, which is due to antigenic shift, antigenic drift and erroneous replication of viral genome by viral RNA polymerase. NS1 protein and segment 8 genome of influenza A viruses which code for it are the major targets as these sequences are fairly conserved across different strains and also they are not variable in nature as conferred by ClustalW analysis.

Filtering by RNAfold, PAIRFOLD and RNAHybrid web platforms reduces number of false positives in the prediction. Filtering by Target ScanS, DIANA micro T, PicTar, miRanda web platforms reduces number of false positives in case of miRNA targets. Construction of miRNA regulatory networks provides insights into possible roles of predicted miRNAs in gene regulation. The target genes of predicted miRNAs are related with stress related pathways and various cancer types. This suggests that action of NS1 protein inside the host cell might be analogous to that of proteins involved in cancer induction. Correlation between predicted miRNAs clearly shows that Uracil is the predominant base in pre-miRNAs followed by adenine. Also, adenine predominates in the seed region of miRNAs, suggesting importance of adenine in miRNA functionality. Also correlation values show that Uracil contributes to thermodynamic stability of miRNAs. These observations are important in developing therapeutic miRNAs considering delivery miRNAs which might pose problems in their action *in vivo*.

References

1. Brown IH (2000) The epidemiology and evolution of influenza viruses in pigs. Vet Microbiol 74: 29-46.
2. Trampuz A, Prabhu RM, Smith TF, Baddour LM (2004) Avian influenza: a new pandemic threat? Mayo Clin Proc 79: 523-530.
3. Peiris JS, de Jong MD, Guan Y (2007) Avian influenza virus (H5N1): a threat to human health. Clin Microbiol Rev 20: 243-267.
4. Brahmachari SK (2008) Targets for human microRNAs in avian influenza virus (H5N1) genome. US 2008/0045472.
5. De Clercq E, Neyts J (2007) Avian influenza A (H5N1) infection: targets and strategies for chemotherapeutic intervention. Trends Pharmacol Sci 28: 280-285.
6. Zhou H, Jin M, Yu Z, Xu X, Peng Y, et al. (2007) Effective small interfering RNAs targeting matrix and nucleocapsid protein inhibit influenza A virus replication in cells and mice. Antiviral Res 76: 186-193.
7. Ge Q, McManus MT, Nguyen T, Shen CH, Sharp PA, et al. (2003) RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. Proc Natl Acad Sci U S A 100: 2718-2723.
8. Tompkins SM, Lo CY, Tumpey TM, Epstein SL (2004) Protection against lethal influenza virus challenge by RNA interference in vivo. Proc Natl Acad Sci U S A 101: 8682-8686.
9. Bushati N, Cohen MS (2007) miRNA functions. Annu Rev Cell Dev Biol 23: 175-205.
10. Sontheimer EJ, Carthew RW (2005) Silence from within: endogenous siRNAs and miRNAs. Cell 122: 9-12.
11. Lamb RA, Krug RM (2001) Orthomyxoviridae: the viruses and their replication. In Fields, B.N., Peter, M., Howley, M.D., Griffin, D.E., Robert, A., Lamb, R.A., Malcolm, A., Martin, M.D., Roizman, B., Straus, M.D., Knipe, D.M. (eds.), Fields Virology, Volume 1, 4th Edition, Lippincott Williams and Wilkins.
12. Cheung KWT, Poon LLM (2007) Biology of influenza A virus. Annals of NewYrok Academy of Sciences 1102: 1-25.
13. Geiss GK, Salvatore M, Tumpey TM, Carter VS, Wang X, et al. (2002) Cellular transcriptional profiling in influenza A virus infected lung epithelial cells: the role of nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. Proc Natl Acad Sci U S A 99: 10736-10741.
14. García-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, et al. (1998) Influenza A virus lacking the NS1 gene replicates in interferon deficient systems. Virology 252: 324-330.
15. Bergmann M, Garcia-Sastre A, Carnero E, Pehamberger H, Wolff K, et al. (2000) Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. J Virol 74: 6203-6206.
16. Wang X, Li M, Zheng H, Muster T, Palese P, et al. (2000) Influenza A virus NS1 protein prevents activation of NF-κB and induction of alpha/beta interferon. J Virol 74: 11566-11573.
17. García-Sastre A (2001) Inhibition of interferon-mediated antiviral responses by influenza A viruses and other negative strand RNA viruses. Virology 279: 375-384.
18. Ludwig S, Wang X, Ehrhardt C, Zheng H, Donelan N, et al. (2002) The influenza A virus NS1 protein inhibits activation of Jun-N-terminal kinase and AP-1 transcription factors. J Virol 76: 11166-11171.
19. Koparde P et al. Avian Influenza and micro RNA: Role of bioinformatics. J Biomed Biotechnol (in pipeline).
20. Palese P, Basler CF, García-Sastre A (2002) The makings of a killer. Nat Med 8: 927-928.
21. Martin G, Schouest K, Kovvuru P, Spillane C (2007) Prediction and validation of microRNA targets in animal genomes. J Biosci 32: 1049-1052.
22. Ghosh Z, Chakrabarti J, Mallick B (2007) miRNomics-The bioinformatics of microRNA genes. Biochem Biophys Res Commun 363: 6-11.
23. Barbato C, Arisi I, Frizzo ME, Brandi R, Da Sacco L, et al. (2009) Computational challenges in miRNA target predictions: to be or not to be a true target? J Biomed Biotechnol 2009: 1-9.
24. Xiao C, Rajewsky K (2009) MicroRNA control in the immune system: basic principles. Cell 136: 26-36.
25. Ritchie W, Théodule FX, Gautheret D (2008) Mireval: A web tool for micro RNA prediction in genome sequences. Bioinformatics 24: 1394-1396.
26. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: 140-144.

27. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

28. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. J Mol Biol 288: 911-940.

29. Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, et al. (2008) miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Res 36: 165-169.

30. Ritchie W, Flamant S, Rasko JE (2009) mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. Bioinformatics 26: 223-227.

31. Zhang B, Stellwag EJ, Pan X (2009) Large scale genome analysis reveals unique features of microRNAs. Gene 443: 100-109.

32. Andronescu M, Aguirre-Hernández R, Condon A, Hoos HH (2003) RNAsoft: a suite of RNA secondary structure prediction and design software tools. Nucleic Acids Res 31: 3416-3422.

33. Krüger J, Rehmsmeier M (2006) RNAHybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res 34: 451-454.

34. Papadopoulos GL, Alexiou P, Maragkakis M, Reczko M, Hatzigeorgiou AG (2009) DIANA-mirPath: Integrating human and mouse microRNAs in pathway. Bioinformatics 25: 1991-1993.

35. Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG (2006) miRGen: A database for the study of animal microRNA genomic organization and function. Nucleic Acids Res 35: 149-155.

36. softberryfindmiRNA http://linux1.softberry.com/berry.phtml?topic=findmirna&group=programs&subgroup=rnastruct