

Full Length Research Paper

Potential drug targets prediction for H1N1 influenza A based on protein-protein interaction networks

Di Chen^{1#}, Xi Liu^{1#}, Chang Chen², Lin Chen¹, Yibo Gao¹, Jianqiang Yi¹, Shihuan Tang², Peng Lu¹ and Hongjun Yang^{2*}

¹Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, People's Republic of China.

²Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, 100070, People's Republic of China.

Accepted 1 October, 2012

In this paper, some proteins were determined as potential drug targets for H1N1 influenza A. These proteins were found using three steps: i) determining a set of human proteins that had interactions with the proteins of H1N1 virus, and assigning a weight to each protein; ii) mapping the set of human proteins into the human protein-protein interaction networks; iii) combining the topological properties of the network and the weights of the proteins to determine the score of each node. Eventually, TRAF2 and MAPK9, two nodes with high scores, were regarded as potential drug targets. The results were consistent when different databases were used, and were in agreement with the KEGG pathways to some degree. However, it still needs further study to ascertain whether these potential targets can be novel practical targets applied in the drug design.

Key words: Protein-protein interactions network, drug targets, H1N1 influenza A.

INTRODUCTION

In March and early April 2009, a new H1N1 virus emerged in Mexico and the United States. During the first few weeks of surveillance, the virus spread worldwide to 30 countries by human-to-human transmission, causing the World Health Organization to raise its pandemic alert level (Smith et al., 2009; Roberts et al., 2012). Currently, two classes (M2 channel blockers and neuraminidase inhibitors) of U.S. Food and Drug Administration (FDA)-approved influenza antiviral drugs are available, but there are great concerns of emergence of viral resistance. Hence, timely identification of suitable targets, which are key elements related with the mechanism of the influenza is crucial (Basler, 2007; Vijayan et al., 2012). Besides, the identification method should be used to predict targets for similar diseases in case of another flu

Typically, a drug target is a key molecule involved in some particular metabolic or signaling pathways that are specific to a disease condition or pathology or to the infectivity or survival of a pathogen (Sams-Doddf, 2005; Knowles and Gromo, 2003; Cheng et al., 2007). Identifications of novel drug targets that activate or inhibit these responses can be broadly divided into studies at the physiological, mechanistic or genetic level (Lindsay, 2005). A great number of databases at genetic or proteomic level make it possible to do some in-silicon predictions on novel targets.

During the process of a H1N1 viral infection, proteins of the virus interfere with many of the host proteins, and some of these proteins interact with each other, or interact with the other proteins simultaneously (Shapira et al., 2009; Zhirnov et al., 2002). Investigating the individual protein is not enough to elucidate the complex variation of the bioprocesses of a body in a disease condition. Instead, mapping these proteins to the protein-protein interaction networks can highlight the complex relationships among them and help to determine the key nodes which can be potential drug targets.

*Corresponding author. E-mail: peng.lu@ia.ac.cn; honjun0420@gmail.com.

#These authors contributed equally to the work outbreak.

This study tried to identify the influenza-related proteins that could be potential drug targets. To achieve this goal, there are mainly three steps, i) find out a set of proteins of human that have interactions with the proteins of H1N1 virus, and assign a weight to each protein according to the number of its interactions with the viral proteins; ii) map the set of human proteins into human protein-protein interaction networks, describing the relationship among these proteins; iii) combine the topological properties of the network and the weights of the proteins to determine the score of each node, the nodes with high scores can be regarded as potential drug targets. Several previous studies had identified potential drug targets through network approaches (Wu et al., 2008; Oti et al., 2006). However, most of these methods used known targets to predict new targets without considering the proteins of virus and human together. By spanning the proteins of influenza virus and human together, our method can have a full perspective about the disease.

MATERIALS AND METHODS

Extracting protein-protein interactions (PPIs) between human and H1N1 influenza virus

When a H1N1 influenza virus invades somebody, the proteins of this virus will have effects on the bioprocesses of human body by interacting with human proteins. So, it is necessary to find out which human proteins have interactions with the virus.

There are many accessible protein-protein interactions databases, such as BioGrid (Stark et al., 2011), IntAct (Kerrien et al., 2012), Human Protein Reference Database (HPRD; Keshava et al., 2009), Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; von Mering et al., 2005), GeneMANIA (Mostafavi et al., 2008). Among these databases, IntAct is a database that can be easily used to retrieve the PPIs of two different species (the results show the interactions of which two interactors are from two different species).

The PPIs between human and H1N1 influenza virus can be investigated using the Molecular Interaction Query Language (MIQL) (Kerrien et al., 2012), which is based on Lucene syntax to limit the species (to get the protein-protein interactions between H1N1 influenza virus and human, the query should be: 'taxid A: 211044 AND taxid B: 9606'). At last, 60 records were given out by the database.

Assigning weights to proteins

Based on the assumption that human proteins which are interfered by more proteins of the virus are more likely to play key roles in the bioprocesses of the disease, the weight of each human protein is defined as:

$$w_p = \frac{c(p)}{N}$$

Where p stands for a protein, $c(p)$ represents the number of the interactions that the protein p has with viral proteins, and N is the total number of interactions. So, a protein that has more neighbors

will be assigned with a higher weight, and the weight can represent the importance of the protein.

Mapping proteins into the human PPI networks

To find out the relationships among the human proteins further, the human proteins that have interactions with virus were mapped into a human PPI network. STRING and GeneMANIA were used to extract sub-networks that span all the 54 proteins and their interactions.

STRING is a database containing known and predicted protein interactions (both physical and functional associations derived from genomic context, high-throughput experiments, conserved co-expression, and previous knowledge). It can also be employed to retrieve the interactions among multiple proteins. Besides, GeneMANIA is another similar database that can be used to find out the interactions among a set of input proteins. On the networks, the nodes are proteins, the edges are interactions between nodes, and each edge is weighted by a confidence score.

Combining the network and the weight of each node to determine a score for each node

Since a node with many neighbors in the PPI network is more likely to play an important role in the bioprocesses of human body, and the node that is a neighbor of an important node is probably another important one, the importance of each node might be related with how many neighbor nodes it has and the importance of its neighbors. Consequently, the score for each node is defined as:

$$score(p) = \sum_{i \in nei(p)} w_i \cdot confidence_i$$

Where p stands for a protein or a node, w_i is the weight of protein i , $nei(p)$ represents the nodes that are connected with p directly in the network, and $confidence_i$ ranges from 0 to 1 (and it is the confidence score given by the PPI database to describe the reliability of the interaction).

Thus this score considers how many neighbors one node has and how important and reliable the neighbors are at the same time. A node with a higher score must be an important protein in the PPI network, as well as a protein have impacts on the interactions between human and virus. Then all the proteins are ranked by the scores.

RESULTS

Protein-protein interactions retrieved from the IntAct

By query with the IntAct, we got 60 interactions between the proteins of H1N1 influenza virus and the proteins of human (Figure 1). Among these 60 interactions, there were 6 proteins from H1N1 influenza, and 54 proteins from human. Set A and Set B were used to represent proteins of the two species respectively (set A contained 6 viral proteins, and set B contained 54 human proteins).

Weights of proteins

Each protein of the set B was weighted based on the

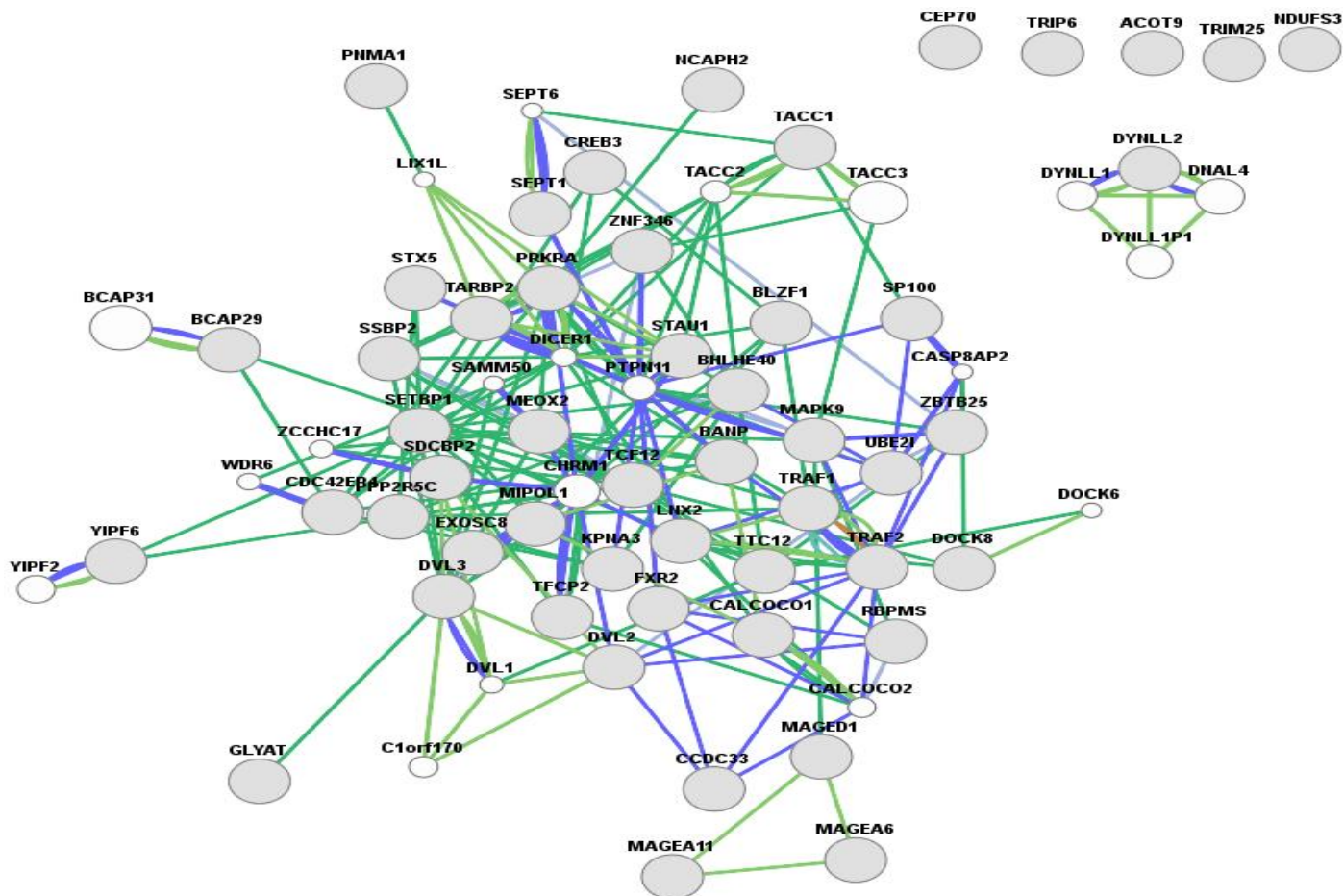


Figure 3. The PPI network derived from GeneMANIA.

Table 2. Top five proteins derived by the PPI network of STRING.

Rank	Protein	Score
1	TRAF2	0.0628
2	MAPK9	0.0376
3	CCDC33	0.0275
4	TRIP6	0.0260
5	UBE2I	0.0259

Table 3. Eight top proteins derived by the PPI network of GeneMANIA.

Rank	Protein	Score
1	ZBTB25	0.0107
2	MAPK9	0.0097
3	TRAF2	0.0093
4	DVL2	0.0039
5	LN2, UBE2I	0.0033
7	BANP	0.0030
8	CCDC33	0.0027

infection, and they may be putative targets. By mapping them into the pathways of KEGG (Kanehisa et al., 2010), it can be found that MAPK9 is a factor in the pathway of Influenza A, and TRAF2 has many interactions with the factors (TNF, TNFRSF1A, TNFRSF10A, TNFRSF10B) of this pathway. So the pathway of influenza A can be disturbed by binding with MAPK9 or TRAF2. These two proteins calculated by our method can be potential drug targets. Other proteins with low scores (EXOSC8, DVL3, BHLHE40, FXR2, GLYAT etc) are not related with the pathway of Influenza A. This shows that our method has the potential to predict novel putative targets. As for other genes found out by our method, which are not on the pathway of influenza, although it can be that they are not potential targets, it is possible they have impacts on the human bioprocesses through other ways or the pathways recorded on the KEGG are not complete enough to contain all related proteins.

Extrapolation to other proteins

The method aforementioned only uses the proteins that

have direct interactions with the virus, but there are many other proteins which can interact with the virus indirectly. Considering that GeneMANIA also determine other proteins that are related to a set of input proteins, it can find out some indirect interactors that are not in set B (the white nodes in Figure 3). In addition, considering GeneMANIA finds other proteins that are related to a set of input proteins, we calculated the scores of proteins that were not in set B by summing up the scores of neighbors that were in set B. Finally, 'PTPN11', 'DICER1', 'STRBP', 'ILF3', 'CHRM1' were among the top 5 proteins. Although they are not related with the viral proteins directly, they interact with many proteins of set B and may perhaps play key roles in the bioprocess during the infection progress.

Further investigation is needed to ascertain the mechanism of the diseases of H1N1 influenza, and more attention should be paid on the proteins (or corresponding genes) ranked high by the method. As they interact with many proteins that may be invaded by the virus, they must play certain roles in the bioprocesses of this disease. Moreover, it is also essential to understand other critical aspects of the drug targets such as underlying toxicity, clinical effects, adverse drug reactions and others (Imming et al., 2006; Hopkins and Groom, 2002). Consequently, it still needs profound investigation to ascertain whether these potential drug targets can be practically applied in drug design (Noble et al., 2004; Lorna and Saad, 2006; Smits et al., 2005).

ACKNOWLEDGEMENTS

This work was supported by the Special Research Foundation for Traditional Chinese Medicine (Grant no. 200907001-5), the National Science Foundation for Post-doctoral Scientists of China (Grant no. 2012M510733), and the National Science Foundation of China (Grant no. 30902020 and 81173463).

REFERENCES

- Basler CF (2007). Influenza viruses: basic biology and potential drug targets. *Infect. Disord. Drug Targets* 7(4):282-293.
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* 25:71-75.
- Hopkins A, Groom C (2002). The druggable genome. *Nature Rev. Drug Discov.* 1:727-730.
- Imming P, Sinning C, Meyer A (2006). Drugs, their targets and the nature and number of drug targets. *Nature Rev. Drug Discov.* 5:821-834.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38:D355-D360.
- Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeifferberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 40(Database issue):D841-846.
- Keshava PTS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadrhan S, Chaerkady R, Pandey A (2009). Human Protein Reference Database - 2009 Update. *Nucleic Acids Res.* 37: D767-772.
- Knowles J, Gromo G (2003). A guide to drug discovery: Target selection in drug discovery. *Nat. Rev. Drug Discov.* 2:63-69.
- Lindsay MA (2005). Finding new drug targets in the 21st century. *Drug Discov. Today* 10(23-24):1683-1687.
- Lorna H, Saad S (2006). Under-reporting of adverse drug reactions: a systematic review. *Drug Saf.* 29(5): 385-396.
- Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* 9(Suppl 1):S4.
- Noble MEM, Endicott JA, Johnson LN (2004). Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* 303 (5665):1800-1805.
- Oti M, Snel B, Huynen MA, Brunner HG (2006). Predicting disease genes using protein-protein interactions. *J. Med Genet.* 43:691-698.
- Roberts KL, Shelton H, Stilwell P, Barclay WS (2012). Transmission of a 2009 H1N1 Pandemic Influenza Virus Occurs before Fever Is Detected, in the Ferret Model. *PLoSOne* 7(8): e43303.
- Sams-Dodd F (2005). Target-based drug discovery: is something wrong? *Drug Discov. Today* 10:139-147.
- Shapira SD, Gat-Viksl, Shum BO, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, Hill DE, Regev A, Hacohen N (2009). A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 139(7):1255-1267.
- Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghvani J, Bhatt S, Peiris JSM, Guan Y, Rambaut A (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459(25):1122-1125.
- Smits KM, Schouten JS, Smits LJ, Stelma FF, Nelemans P, Prins MH (2005). A review on the design and reporting of studies on drug-gene interaction. *J. Clin. Epidemiol.* 58(7):651-654.
- Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M (2011). The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* 39(suppl1):D698-D704.
- Vijayan V, Jing J, Zangwill KM (2012). Evaluation of Diagnostic and Therapeutic Approaches for Suspected Influenza A (H1N1) pdm09 Infection, 2009-2010. *Emerg. Infect. Dis.* 18(9):1414-1421.
- von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P (2005). STRING: known and predicted protein-protein associations integrated and transferred across organisms. *Nucleic Acids Res.* 33:D433-D437.
- Wu X, Jiang R, Zhang MQ, Li S (2008). Network-based global inference of human disease genes. *Mol. Syst. Biol.* 4:189.
- Zhirnov OP, Ksenofontov AL, Kuzmina SG, Klenk HD (2002). Interaction of Influenza A Virus M1 Matrix Protein with Caspases. *Biochemistry* 41(12): 534-539.