ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

EMPLOYING COMPUTATIONAL SYSTEMS AND NETWORKS IN PREDICTION AND MODIFICATION OF DISEASE GENES TO INCREASE IMMUNITY

Ekambir Singh
Ajmera Global School
EkambirS@ags.org.in

Abstract

This paper reviews modern computational network-based techniques, specifically network diffusion methods, in prediction of disease genes, as well as mechanisms being used to modify disease genes to ultimately increase immunity against fatal diseases and disorders. It explains computational network diffusion approaches in disease gene prediction, analyzing their benefits and limitations. The paper next examines the Clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins(CRISPR-Cas) system, a breakthrough in the genetic modification sector, and its use in genetic modification of disease genes. Showing why network-based strategies and the CRISPR-Cas system are currently the vanguard of this sector, it concludes by suggesting further unique possible routes for study andimprovement.

Keywords:*network diffusion methods, disease genes, genetic modification, CRISPR-Cas*

INTRODUCTION

The possibility of being equipped with the ability to completely curb life-threatening diseases to humans is extremely inspiring. However, the prediction and modification of disease genes has been an onerous mission for researchers. This paper explains achieving this efficiently in two steps: 1. Disease gene prediction and 2. Use of genetic modification in correcting disease genes. A disease gene can be defined as the 'defective' gene that goes through a mutation to cause disruptions in the cell system and so the disease. Usually, a group of disease genes malfunction to cause a disease. Nearly all diseases involve some genetic malfunction. Essentially, biomolecules in cells always carry out their functions in convoluted pathways, involving many chemical interactions and reactions. Portraying the pathways in a network format would not only be easier to understand, but also track disruptions in them because of various influences. Hence, the use of network-based systems with the assistance of computational systems and software at our disposal is a preferable option. These networks basically consist of 'nodes' representing the object and 'edges' representing the links between them. Complex algorithms in the system would allow missing nodes or edges to be deduced, leading to the prediction of unknown disease genes that cause x diseases(x being the name/type of disease). Protein-Protein Interaction(PPI) networks are currently most used to study the relationship between them. This paper goes on to examine and compare modern computational network diffusion type techniques in disease gene prediction.The second part of this research paper examines the modification of disease genes,

ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

combining concepts of genetic modification and gene therapy, to essentially 'correct' them and prevent life-threatening diseases.Genetic modification is the editing of the genetic composition of an organism. Arguably the cheapest, quickest and most accurate system for editing genes is the Clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins(CRISPR-Cas) system, explained in section 'CRISPR-Cas for genetic modification' later in the paper. CRISPR-Cas was first found in prokaryotes, as an active immune system to defend against foreign virus enemies. This paper explains how the CRISPR-Cas systems work and how computational systems are supporting them to contribute to advancement in genetic modification to enhance human immunity.

Theory

Analysing PPI networks to predict disease genes

Proteins play a vital role in the cellular and molecular reactions of our body to keep us healthy. The nature of these protein "interactions" is such that they can be displayed as a network structure to show how a certain biological function is being performed. So, a particular PPI not taking place because of a faulty protein or other disease-causing factor would be identified by the system. This could be shown by an edge/node removal in the network[1]. Additionally, the topology i.e. the way different protein molecules are connected in the network and the overall patterns that emerge out of this, is an integral feature of PPI networks. Hence with the help of computational systems and network theory, PPI networks can be integrated with disease-gene networks to unearth new disease-gene associations.

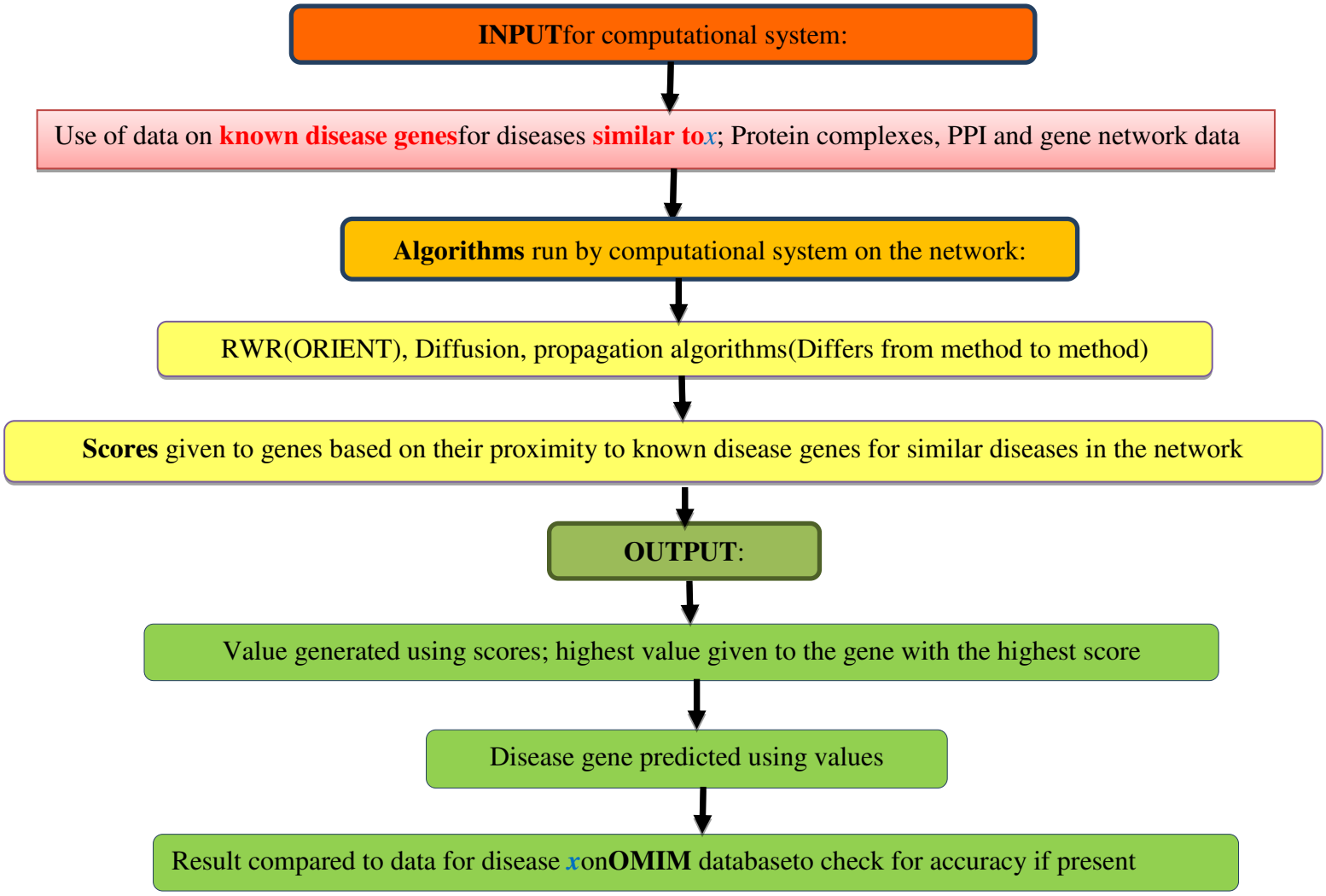
Existing network diffusion methods for disease gene prediction

Diffusion or propagation in a network can be thought of as spreading or travelling of biological data across its edges. Computational systems are arguably the best choice to construct and monitor them due to the vast amounts of data. There have been various network diffusion-type techniques created in this decade that try and predict disease genes using this principle. PRINCE[2] is one such method that uses PPI networks to infer disease genes. Designed by Vanunu O et al., PRINCE uses prior data on known disease-causing genes for similar phenotypical diseases to the one in question. Then using a diffusion-based algorithm it gives scores to the candidate genes in the network. This scoring is done based on the proximity of the candidate gene to a known disease-causing gene for a similar disease. So, the closer the candidate gene to the disease-causing gene, the higher the score. It also can infer the protein complexes involved in that disease. VAVIEN is another method that assigns scores to proteins based on their topological properties using a random walk algorithm. A 'walk' in network theory is going along a sequence of adjacent nodes through the edges connecting them to reach a desired destination. ORIENTuses a RWR(Random walk with restart) algorithm(walks but with the ability to stop and restart from another node) to assign scores in the form of 'weights' to the edges through the shortest path from a candidate gene to the known disease gene. Fig 1. shows the basic steps and framework that all these network systems follow to obtain a result for the disease-causing genes of the disease *x* in question. These results were compared to those in OMIM(Online Mendelian Inheritance in man), the global catalog for diseases and genes.

ISSN 2350-109X
www.indianscholar.co.in
Indian Scholar
An International Multidisciplinary Research e-Journal
MISA MEMBERS OF INTERNATIONAL SCHOOLS' ASSOCIATION

Limitations and further possible areas for research

Network diffusion methods are all highly dependent on prior data of known disease genes for similar diseases(highlighted in red in Fig 1.). So, in cases where the disease in question has no other phenotypically similar diseases, it may be much more challenging to predict disease genes for the disease. The data that is actually there to work with is quite limited, which poses study limitations[1]. The reliability and "completeness"(they are still dependent on known disease-gene associations giving rise to bias) of the PPI and gene networks is also influential in determining the position and scores of genes in the network, so it has a direct impact on the results obtained. It is necessary for scientists to address these hurdles for improving accuracy of these models. A future prospect could be to try and use a combination of PPI and gene networks and machine learning to predict disease genes more accurately. By doing this it would be expected ideally that problems of incomplete networks get eliminated and all methods are combined into one. Such a method could be devised that current existing knowledge of known disease genes is used to the fullest by the system without any excessive bias[1] shown in the results because of them.



ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

Fig 1. A flowchart summarising the common key steps taken by network diffusion-based computational systems to discover disease genes for disease *x* (the disease in question).

CRISPR-Cas for genetic modification

The working of the CRISPR-Cas immune system in prokaryotes is explained in detail in [4].However, recent research has proved that CRISPR-Cas systems can be used to serve other purposes too. The four main steps to modify disease genes using CRISPR-Cas9 are listed below:

1) Scientists construct the custom guide RNA, complementary to the DNA sequence in the disease gene they want to edit.

2) The guide RNA is made to bind with extracted Cas9 nuclease and form an artificial effector complex.

3)Effector complex is inserted in the organism's body and makes the cut at the required site.

4) Using HDR (Homology Directed Repair - a template is given by scientists)[3] the cut is repaired and new DNA made; disease gene corrected into a healthy gene and immunity is enhanced.

Novel computational systems have been freshly createdto aid scientists in steps 1) and 4) above. CHOPCHOP, CRISPOR and CRISPRscan are three guide RNA designing web-based computational tools created to:

1)Design, evaluate and clone guide RNA sequences as asked by scientists

2)Create predictive models showing positions of targeted genes

3)Searching for off-target impacts and sites where unintended changes may happen

4)Be compatible with a wide variety of genomes, easily create genome-scale "libraries".

CONCLUSION

To conclude, disease gene prediction and modification is of utmost importance to be able to upgrade human immunity against severe diseases that threaten our population. Computational systems in disease gene prediction and genetic modification have tremendously played a key role in augmenting our knowledge. But there is still a long way to go in improving PPI networks' reliability and network diffusion models' accuracy and efficiency**.** With the assistance of rapidly advancing computational resources, it is hoped that we will soon be able to perfect and use these methods as a way of saving lives affected unfortunately by disease.

Acknowledgements

I would like to thank my teachers Ms. NishaBhatia and Ms. Pooja Shah for their valuable guidance.

REFERENCES

1. Sezin Kircali Ata, Min Wu, Yuan Fang, Le Ou-Yang, Chee Keong Kwoh, Xiao-Li Li, "Recent advances in network-based methods for disease gene prediction", in *Briefings in Bioinformatics*, Volume 22, Issue 4, Oxford University Press, July 2021

ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

2. Oron Vanunu, Oded Magger, Eytan Ruppin, Tomer Shlomi, Roded Sharan, "Associating Genes and Protein Complexes with Disease via Network Propagation", in *PLoS Computational Biology*, Volume 6, Issue 1, January 2010
3. C. Li, W. Chu, R.A. Gill, S. Sang, Y. Shi, X. Hu, Y. Yang, Q.U. Zaman, B. Zhang, "Computational tools and resources for CRISPR/Cas genome editing", in *Genomics, Proteomics & Bioinformatics*,February 2022
4. Unknown. "What is CRISPR-Cas9".yourgenome.org(Accessed June 4,2022)