University of Sistan and Baluchestan

Journal of Epigenetics

Online ISSN: 2676 - 4350

Homepage: https://jep.usb.ac.ir

Iranian Genetics Society

Predicting the effect of LncRNAs on different types of cancers in bipartite networks using link prediction

Bahar Ataei [1*], Shahrzad Benvidi [2], Parnaz Soori [2], Shahab Bakhtiari [3]

[1] *Corresponding author*, Department of Genetics, Faculty of Basic Science, Shahrekord University, Shahrekord, Iran, bahar.ataei1373@gmail.com

[2] Department of Biology, Faculty of Basic Science, Islamic Republic Azad University, Tehran, Iran, shahrzadbenvidi1379@gmail.com , parnaz.soori@gmail.com

[3] Department of Biological science, Kurdistan University, Sanandaj, Iran, shahabbakhtiari.gen@gmail.com

ARTICLE INFO

Article type:
Research Article

Article history:
Received: 16 April 2023
Revised: 18 September 2023
Accepted: 7 October 2023

Keywords:
Cancer,
LncRNA,
Link prediction,
AA,
JC,
PA,
CN.

A B S T R A C T

Many complex systems such as the Internet, the World Wide Web, the brain, and the causes of diseases can be described by networks with nodes representing agents and links representing relationships or interactions between nodes. Despite these systems seeming utterly different at first glance, they are all made up of interacting parts. Individual objects in this type of system are not isolated but connected through links or relationships. Long non-coding RNAs (LncRNAs), one of the factors related to many diseases, are specific genes in the human genome that control many different biological processes. lncRNAs have been shown to regulate cancer development and occurrence. We used link prediction bioinformatics tools to identify lncRNAs affecting various types of cancers. To achieve this goal, two-part networks were used using CN (Common Neighbors), AA (Adamic/Adar), PA (Preferential Attachment), and JC (Jaccard's Coefficient) algorithms. The results indicated that all the obtained lncRNAs with a high score had been reviewed in other research articles, which was a sign of the correctness of our algorithms. 4.5% of the obtained results lacked scientific and research studies, all with a high score for future studies and included lncRNA H19 and SPRY4-IT1. In addition, in one case with a high score and first position in the Excel file obtained using the Jaccard coefficient algorithm, lncRNA with the symbol AC09510.3 was associated with adenocarcinoma. Still, this possible link has not been investigated in any article yet. Conclusion: The current study may identify novel lncRNAs implicated in Adenocarcinoma, vulva squamous cell carcinoma, basal cell carcinoma, gestational choriocarcinoma, chromophobe renal cell carcinoma, brain cancer, pancreatic cancer, hepatocellular carcinoma, prostate cancer, embryonal cancer, germ cell cancer, and choriocarcinoma; however, further research is required to determine the potential functions of this lncRNAs in cancers mentioned above.

Introduction

Many complex systems such as the Internet, the World Wide Web, the brain, and the causes of diseases can be described by networks with nodes representing agents and links representing relationships or interactions between nodes. Despite these systems seeming completely

DOI: https://doi.org/10.22111/JEP.2023.45307.1054

Publisher: University of Sistan and Baluchestan

How to Cite: Ataei, B., Benvidi, Sh., Soori, P., Bakhtiari, Sh. (2023). Predicting the effect of LncRNAs on different types of cancers in bipartite networks using link prediction. *Journal of Epigenetics*, 4(1), 44-49. https://doi.org/10.22111/JEP.2023.45307.1054

different at first glance, they are all made up of interacting parts. Individual objects in this type of system are not isolated, but connected through links or relationships (Albert and Barabási 2002, Dorogovtsev, Goltsev et al. 2002, Newman 2003, Boccaletti, Latora et al. 2006, Costa, Rodrigues et al. 2007, Kerrache, Alharbi et al. 2020).

For more than two decades, the use of link prediction methods has been of great interest in complex network research which is essentially too expensive if done in traditional methods and its goal is to understand and develop effective tools to describe and quantify complex systems therefore, the first step in this effort is to observe and record existing interactions to build a network (Ahmad, Akhtar et al. 2020, Jafari, Abdolhosseini-Qomi et al. 2021, Shang and Small 2022).

So far, many indices have been developed for link prediction, which is classified into three main models: models based on the Markov chain, models based on machine learning, and models based on topological structure similarity. Many of the investigated networks are large-scale networks, therefore, due to the complexity of the first two models, they cannot be used, but the third model, relying on the topological structure, provides the information of complex networks in a simple way(Li, Huang et al. 2018). Some widely used methods accurately predict which observed pairs of unrelated nodes should be connected. Therefore, by helping to select more accurate network models, link prediction methods can elucidate the organizing principles of a variety of complex systems (Ghasemian, Hosseinmardi et al. 2020, Qiu, Zhong et al. 2021).

Common Neighbors (CN) (Newman 2001, Li, Huang et al. 2018), Preferred Attachment (PA), Adamic-Adar (AA), and Jaccard's Coefficient (JC) (Jaccard 1901) are types of such indicators that mainly focus on the structural characteristics of nodes (Lü and Zhou 2010, Jafari, Abdolhosseini-Qomi et al. 2021). For better understanding, each algorithm is briefly explained below.

Common Neighbors (CN)

In this algorithm, scores are calculated for each pair of nodes to check the probability of being linked, so that the higher the score, the higher the probability of being linked. CN is a method with low time complexity and easy implementation (10, 16, 17).

In this algorithm, scores are calculated for each pair of nodes to check the probability of being linked, so that the higher the score, the higher the probability of being linked. CN is a method with low time complexity and easy implementation (Cheng, Chen et al. 2016, Li, Huang et al. 2018, Wang and Le 2020).

Score $(x,y) = | \tau(x) \cap \tau(y) |$
(Cheng, Chen et al. 2016, Ahmad, Akhtar et al. 2020)

Jaccard Index (JC)

About a century ago, Paul Jaccard proposed the Jaccard method to compare the similarity and diversity of objects. This measurement evaluates the overlap between the neighbors of two nodes by normalizing the size of the intersection with the size of the union (Wang and Le 2020, Coşkun and Koyutürk 2021), so as the denominator of the fraction increases, the obtained score decreases.

$S_{xy} = (| \tau(x) \cap \tau(y) |)/(| \tau(x) \cup \tau(y) |)$ (Ahmad, Akhtar et al. 2020)

Preferential Attachment Index (PA)

In this algorithm, the higher the degree, the higher the probability of creating a link, and the probability that this new link connects x and y is proportional to $K_x \times K_y$. K_x refers to the degree value of node x (Cheng, Chen et al. 2016, Wang and Le 2020).

$S_{xy} = K x \times K y$ (Cheng, Chen et al. 2016)

Adamic-Adar Index (AA)

An innovative method to calculate the similarity between two web pages was proposed by Adamic and Adar in 2003. In the AA technique, nodes with smaller degrees contribute more (He, Yang et al. 2018, Wang and Le 2020). Then, its similarity is defined as follows:

$S_{xy}=\sum_{z\epsilon| \tau(x) \cap \tau(y) |} \frac{1}{\log kz}$ (He, Yang et al. 2018)

Link prediction is an important and well-studied problem in network biology (Coşkun and Koyutürk 2021). It is possible to use link prediction

to check the probability of a connection between a pathogen and a disease by using two-part networks. For instance, finding the relationship between lncRNAs and different diseases is a controversial issue that can be investigated using the programming tools and algorithms mentioned above.

In the human genome, long non-coding RNAs (LncRNAs) regulate a wide range of biological processes through the regulation of long genes from the non-coding RNA group. A long gene is one that contains at least 200 nucleotides in length. lncRNAs have been shown to regulate cancer development and occurrence (Wang, Su et al. 2018, Ali and Grote 2020, Bridges, Daulagala et al. 2021). LncRNAs are involved in various mechanisms of gene regulation due to their ability to interact with DNA, RNA, or protein. They can act as signals for transcription, decoys for transcription repression, epigenetic regulators, or scaffolds for the assembly of ribonucleoprotein complexes with various protein partners. It is possible for lncRNAs to regulate gene expression at transcriptional and/or posttranscriptional levels, depending on the level of gene expression. Furthermore, considering the importance of cell signaling pathways in the initiation, progression, and metastasis of cancer, lncRNAs that affect these pathways can be expected to influence all aspects of cancer development (Peng, Koirala et al. 2017, Xing, Sun et al. 2021).

In this study, using bipartite networks and programming, we investigate the probability of lncRNAs being associated with various types of cancers. Then we separate cases with a higher probability of connection, which can help researchers in future studies.

Methods

Preliminary data

The LncRNA and Disease Database (http://rnanut.net) was utilized to retrieve lncRNA disease data by integrating comprehensive experimentally supported ncRNA-disease associations curated from manual literature and other resources. Here, 10,565 confirmed links between different lncRNAs with diseases were presented, of which a significant number were assigned to cancers.

Preprocessing of the raw dataset

Due to the close relationship between lncRNAs and types of cancer, which was the aim of our study, the primary data were analyzed after an initial filter and excluding diseases other than cancer in Homo sapiens species. The number of verified links obtained after applying the filter reached 4135 links (Figure 1).

Preparing possible links using link prediction algorithms

The investigated networks in this study were bipartite in the sense that the nodes located in one group cannot communicate with each other and the aim is to detect the possible connections between the nodes of one group with another one. To begin with, using Python programming for each of the Common Neighbors (CN), Jaccard Index (JC), Preferential Attachment Index (PA), and Adamic-Adar Index (AA) algorithms, a program suitable for two-part networks were written to find all the possible links between the two groups. Then the information obtained for each algorithm was classified from the most likely to the least likely, and the top 100 most likely links were selected using the NCBI database (https://pubmed.ncbi.nlm.nih.gov) to ensure the existence or the lack of research in that field was investigated.

Results and Analysis

After Preprocessing the raw dataset part of its network was drawn using the network (Figure1). Red nodes in each network represent LncRNAs and blue nodes represent diseases. The gray lines represent the possible connection of each node of a group with another node of the group. Also, the diameter of each node has a direct relationship with the number of links connected to that node. Jupyter notebook platform, pandas, and matplotlib were used in the experiment for data preprocessing and NetworkX to construct a bipartite network. After programming and processing for each of the 4 selected algorithms, the amount of 1438 types of LncRNA and 108 different cancers was obtained, and their prioritization order was different for each algorithm. According to the said content, it can be concluded that the number of nodes that can be checked is equal to 1546. Using the Jupyter platform, the number of possible links between the two groups was calculated as 152,846.

Figure 2 shows the graph of the frequency of diseases versus lncRNAs degree, according to which it can be said that every disease is associated with a large number of lncRNAs. The top 100 possible links of each algorithm were examined, of which about 96% of the data had already been tested, which confirmed the correctness of the programming and processing. Also, around 4% of the data had not been researched yet (Figure 3). The obtained results are summarized in Table 1.

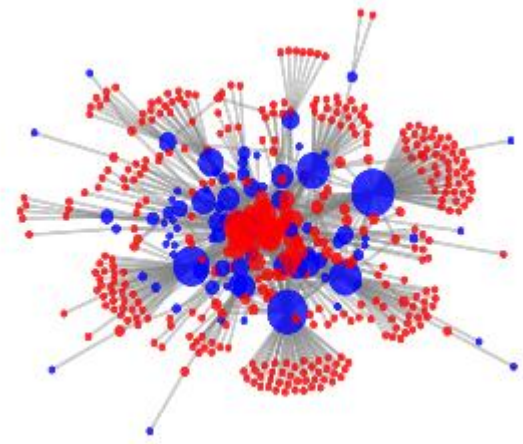

Fig. 1- A bipartite network has been drawn between cancers and different lncRNAs using networkX. Red nodes in each network represent lncRNAs and blue nodes represent diseases. The gray lines represent the possible connection of each node of a group with another node of the group.

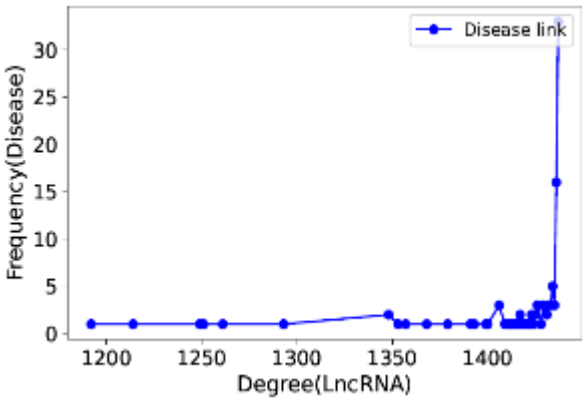

Fig. 2- Frequency of diseases versus lncRNAs degree. every disease is associated with a large number of lncRNAs.

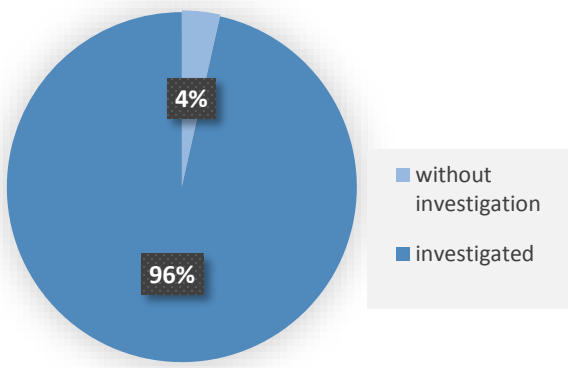

Fig. 3- Link prediction algorithm results for 400 possible links examined. 96% of the data had already been tested. around 4% of the data had not been researched yet

Table 1- The predicted links between 100 top LncRNA and Disease

LncRNA	Disease	Algorithm	Position of prediction
H19	vulva squamous cell carcinoma	AA	38
AC090510.3	colon adenocarcinoma	JC	1
H19	vulva squamous cell carcinoma	JC	73
UCA1	gestational choriocarcinoma	JC	81
SNHG15	chromophobe renal cell carcinoma	JC	38
BCYRN1	pancreatic cancer	PA	14
HOXA11-AS	pancreatic cancer	PA	24
NPTN-IT1	pancreatic cancer	PA	25
RNY1	hepatocellular carcinoma	PA	54
RNY3	hepatocellular carcinoma	PA	55
PTENP1	pancreatic cancer	PA	77
HNF1A-AS1	prostate cancer	PA	96
H19	vulva squamous cell carcinoma	CN	34

UCA1	choriocarcinoma	CN	99

Discussion

Link prediction is the most basic issue in complex networks, which can be simplified by using existing algorithms. In this study, we used the Jupyter platform and Algorithms called CN (Common Neighbors), AA (Adamic/Adar), PA (Preferential Attachment), and JC (Jaccard's Coefficient) algorithms, which are among the types of local approaches and can only be calculated through the local information of the nodes, to write a program suitable for two-part networks to find all the possible links between LncRNAs and Disease groups (Wang and Le 2020). One of the important lncRNAs that was observed in the top 100 data of AA, JC, and CN algorithms was H19 lncRNA. H19 is one of the reported lncRNAs including MALAT1, MEG3, TUG1, and UCA1 in bladder cancer (Hua, Lv et al. 2016). lncRNA H19 can affect the phenotype of vascular endothelial cells and thus cause the growth of liver tumors. In addition, the role of lncRNA H19 in metastasis has been confirmed in prostate cancer as well as endometrial cancer (Zhang, Wang et al. 2018). In the article published by Guang-Yu Wang and his colleagues in 2018, the effect of this LncRNA in Esophageal cancer, gastric cancer, pancreatic cancer, hepatocellular cancer, and colorectal cancer was confirmed (Wang, Zhu et al. 2014). But there was no study regarding the relationship between H19 lncRNA and vulva squamous cell carcinoma.

In a new study in 2021 by Faiza Naz and her colleagues, although effects of lncRNAs including HOTAIR, MALAT1, MIR31HG, NEAT1, ROCK1, and UCA1 in vulvar cancer were mentioned. but the investigation regarding the effect of H19 on this disease was not recorded (Naz, Tariq et al. 2021). As mentioned earlier, UCA1 has been reported as an oncogenic lncRNA in several squamous cell carcinomas, such as oral squamous cell carcinoma and vulval squamous cell carcinoma (Gao, Fang et al. 2021), but there was no study on its relationship with gestational choriocarcinoma.

While the role of HOTAIR lncRNA in this cancer has been identified, the results obtained for pancreatic cancer, which is one of the types of gastrointestinal cancer, emphasize the role of BCYRN1, HOXA11-AS, NPTN-IT1, and PTENP1 (Bhan and Mandal 2015). According to the study conducted by Guo Yu Zhang et al. in 2017 on human NSCLC cell lines, the effect of HNF1A-AS1 lncRNA was shown (Zhang, An et al. 2018), but so far, no research has been reported on the relationship between this lncRNA and prostate cancer.

The role of SNHG15 lncRNA in urological cancers, including bladder cancer, has been investigated and confirmed (Guo, Liu et al. 2019), but the investigation regarding its role in chromophobe renal cell carcinoma, which is one of the rarest kidney cancers, has not been done, which is suggested to be examined due to the relationship between the kidney and urological cancers. Also, RNY1 and RNY3 were among the lncRNAs and a possible link between them and hepatocellular carcinoma was predicted by the PA algorithm, and no article has been published about this until now.

In addition, in one case with a high score and first position in the Excel file obtained using the Jaccard coefficient algorithm, lncRNA with the symbol AC09510.3 was associated with adenocarcinoma which is a type of cancer that starts in cells that form glands making mucus to lubricant the inside of the colon and rectum (Zhang, Qian et al. 2018, Pournoor, Mousavian et al. 2020), but this possible link has not been investigated in any article yet.

Conclusion

The current study may identify novel lncRNAs implicated in Adenocarcinoma, vulva squamous cell carcinoma, basal cell carcinoma, gestational choriocarcinoma, chromophobe renal cell carcinoma, brain cancer, pancreatic cancer, hepatocellular carcinoma, prostate cancer, embryonal cancer, germ cell cancer, and choriocarcinoma; however, further research is required to determine the potential functions of this lncRNAs in cancers mentioned above.

Acknowledgements

The authors are grateful to Geniranlab for creating a gathering space.

References

Ahmad, I., M. U. Akhtar, S. Noor and A. Shahnaz (2020). "Missing link prediction using common neighbor and centrality based parameterized algorithm." Scientific reports **10**(1): 1-9.

Albert, R. and A.-L. Barabási (2002). "Statistical mechanics of complex networks." Reviews of modern physics **74**(1): 47.

Ali, T. and P. Grote (2020). "Beyond the RNA-dependent function of LncRNA genes." Elife **9**: e60583.

Bhan, A. and S. S. Mandal (2015). "LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer." Biochimica et Biophysica Acta (BBA)-Reviews on Cancer **1856**(1): 151-164.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang (2006). "Complex networks: Structure and dynamics." Physics reports **424**(4-5): 175-308.

Bridges, M. C., A. C. Daulagala and A. Kourtidis (2021). "LNCcation: lncRNA localization and function." Journal of Cell Biology **220**(2).

Cheng, C., J. Chen, X. Cao and H. Guo (2016). "Comparison of local information indices applied in resting state functional brain network connectivity prediction." Frontiers in neuroscience **10**: 585.

Coşkun, M. and M. Koyutürk (2021). "Node similarity-based graph convolution for link prediction in biological networks." Bioinformatics **37**(23): 4501-4508.

Costa, L. d. F., F. A. Rodrigues, G. Travieso and P. R. Villas Boas (2007). "Characterization of complex networks: A survey of measurements." Advances in physics **56**(1): 167-242.

Dorogovtsev, S. N., A. V. Goltsev and J. F. F. Mendes (2002). "Pseudofractal scale-free web." Physical review E **65**(6): 066122.

Gao, Q., X. Fang, Y. Chen, Z. Li and M. Wang (2021). "Exosomal lncRNA UCA1 from cancer- associated fibroblasts enhances chemoresistance in vulvar squamous cell carcinoma cells." Journal of Obstetrics and Gynaecology Research **47**(1): 73-87.

Ghasemian, A., H. Hosseinmardi, A. Galstyan, E. M. Airoldi and A. Clauset (2020). "Stacking models for nearly optimal link prediction in complex networks." Proceedings of the National Academy of Sciences **117**(38): 23393-23400.

Guo, X., N. Liu and M. Liu (2019). "Long non-coding RNA LINC00336 as an independent prognostic indicator and an oncogenic lncRNA in bladder cancer." Archives of Medical Science **15**(1).

He, Y., F. Yang, Y. Yu and C. Grebogi (2018). "Link prediction investigation of dynamic information flow in epilepsy." Journal of Healthcare Engineering **2018**.

Hua, Q., X. Lv, X. Gu, Y. Chen, H. Chu, M. Du, W. Gong, M. Wang and Z. Zhang (2016). "Genetic variants in lncRNA H19 are associated with the risk of bladder cancer in a Chinese population." Mutagenesis **31**(5): 531-538.

Jaccard, P. (1901). "Étude comparative de la distribution florale dans une portion des Alpes et des Jura." Bull Soc Vaudoise Sci Nat **37**: 547-579.

Jafari, S. H., A. M. Abdolhosseini-Qomi, M. Asadpour, M. Rahgozar and N. Yazdani (2021). "An information theoretic approach to link prediction in multiplex networks." Scientific Reports **11**(1): 1-21.

Kerrache, S., R. Alharbi and H. Benhidour (2020). "A scalable similarity-popularity link prediction method." Scientific reports **10**(1): 1-14.

Li, S., J. Huang, Z. Zhang, J. Liu, T. Huang and H. Chen (2018). "Similarity-based future common neighbors model for link prediction in complex networks." Scientific reports **8**(1): 1-11.

Lü, L. and T. Zhou (2010). "Link prediction in weighted networks: The role of weak ties." EPL (Europhysics Letters) **89**(1): 18001.

Naz, F., I. Tariq, S. Ali, A. Somaida, E. Preis and U. Bakowsky (2021). "The role of long non-coding RNAs (lncRNAs) in female oriented cancers." Cancers **13**(23): 6102.

Newman, M. E. (2001). "Clustering and preferential attachment in growing networks." Physical review E **64**(2): 025102.

Newman, M. E. (2003). "The structure and function of complex networks." SIAM review **45**(2): 167-256.

Peng, W.-X., P. Koirala and Y.-Y. Mo (2017). "LncRNA-mediated regulation of cell signaling in cancer." Oncogene **36**(41): 5661-5667.

Pournoor, E., Z. Mousavian, A. N. Dalini and A. Masoudi-Nejad (2020). "Identification of key components in colon adenocarcinoma using transcriptome to interactome multilayer framework." Scientific reports **10**(1): 1-14.

Qiu, X., X. Zhong and H. Zhang (2021). "Applied research on the combination of weighted network and supervised learning in acupoints compatibility." Journal of Healthcare Engineering **2021**.

Shang, K.-k. and M. Small (2022). "Link prediction for long-circle-like networks." Physical Review E **105**(2): 024311.

Wang, G.-Y., Y.-Y. Zhu and Y.-Q. Zhang (2014). "The functional role of long non-coding RNA in digestive system carcinomas." Bulletin du Cancer **101**(9): E27-E31.

Wang, H. and Z. Le (2020). "Seven-layer model in complex networks link prediction: A survey." Sensors **20**(22): 6560.

Wang, J., Z. Su, S. Lu, W. Fu, Z. Liu, X. Jiang and S. Tai (2018). "LncRNA HOXA-AS2 and its molecular mechanisms in human cancer." Clinica chimica acta **485**: 229-233.

Xing, C., S.-g. Sun, Z.-Q. Yue and F. Bai (2021). "Role of lncRNA LUCAT1 in cancer." Biomedicine & Pharmacotherapy **134**: 111158.

Zhang, G., X. An, H. Zhao, Q. Zhang and H. Zhao (2018). "Long non-coding RNA HNF1A-AS1 promotes cell proliferation and invasion via regulating miR-17-5p in non-small cell lung cancer." Biomedicine & pharmacotherapy **98**: 594-599.

Zhang, L., D. Wang and P. Yu (2018). "LncRNA H19 regulates the expression of its target gene HOXA10 in endometrial carcinoma through competing with miR-612." Eur Rev Med Pharmacol Sci **22**(15): 4820-4827.

Zhang, Z., W. Qian, S. Wang, D. Ji, Q. Wang, J. Li, W. Peng, J. Gu, T. Hu and B. Ji (2018). "Analysis of lncRNA-associated ceRNA network reveals potential lncRNA biomarkers in human colon adenocarcinoma." Cellular Physiology and Biochemistry **49**(5): 1778-1791.