VNU Journal of Science:
Computer Science and Communication Engineering
Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

A General Computational Framework for Prediction of Disease-associated Non-coding RNAs

Duc-Hau Le*

School of Computer Science and Engineering, Thuy Loi University,
175 Tay Son, Dong Da, Hanoi, Vietnam

Received 10 February 2019; Accepted 11 October 2019

Abstract: Since last decade, we have been witnessing the raise of non-coding RNAs (ncRNAs) in biomedical research. Many ncRNAs have been identified and classified into different classes based on their length in number of base pairs (bp). In parallel, our understanding about functions of ncRNAs is gradually increased. However, only small set among tens of thousands of ncRNAs have been well studied about their functions and their roles in development of diseases. This raises a pressing need to develop computational methods to associate diseases and ncRNAs. Two most widely studied ncRNAs are microRNA (miRNA) and long non-coding RNA (lncRNA), since miRNAs are the regulators of most protein-coding genes and lncRNAs are the most ubiquitously found in mammalian. To date, many computational methods have been also proposed for prediction of disease-associated miRNAs and lncRNAs, and recently comprehensively reviewed. However, in the previous reviews, these computational methods were described separately, thus this limits our understanding about their underlying computational aspects. Therefore, in this study, we propose a general computational framework for prediction of disease-associated ncRNAs. The framework demonstrates a whole computational process from data preparation to computational models.

Keywords: MicroRNA, long non-coding RNA, disease-miRNA association, disease-lncRNA association, non-coding RNA similarity, disease similarity, network-based method, machine learning-based method.

1. Introduction

Our understanding of noncoding RNAs (ncRNAs) and their functions in a variety of physiological processes has been significantly improved for last decade [1]. The knowledge about noncoding RNAs has shifted from a hypothesis "one gene-one enzyme" [2] to "~80% of the genome is transcribing ncRNAs" [3]. Several types of ncRNAs have been discovered and classified by their length (in number of base pairs (bp)) into short, mid-size and long ncRNAs. Short ncRNAs are a class of ncRNAs having length less than 30bp long,

* Corresponding author.
E-mail address: duchaule@tlu.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.224

mid-size ncRNAs have length in range of 20bp to 200bp, and long ncRNAs are remainders (length > 200bp) [4]. Beside difference in size, they also have different functions related to diseases in general [4], to cancer development [5], and to therapeutically regulate gene expression [6]. For instance, when ncRNAs plays as therapeutic targets, they can be either tumor suppressor or oncogene [5].

Although, tens of thousands of ncRNAs have been discovered, yet our understanding in their functions, especially in disease development, is still limited. Therefore, a number of computational methods have been proposed to predict novel disease-associated ncRNAs [7-9]. Among ncRNAs, microRNA (miRNA) is the most widely studied, which are small ncRNAs of ~22 bp long that mediate post-transcriptional gene silencing by controlling the translation of mRNA into more than 60% proteins. They are also involved in regulating many processes, including splicing, editing, mRNA stability, and translation initiation [6]. Meanwhile, long non-coding RNA (lncRNA) is the largest portion of the mammalian non-coding transcriptome including transcripts more than 200bp long that are involved in many biological processes such as chromatin modification, poll activity regulation, and transcriptional interference [6]. Therefore, in this study, we focus on reviewing computational methods proposed for predicting disease-associated miRNAs and lncRNAs.

H

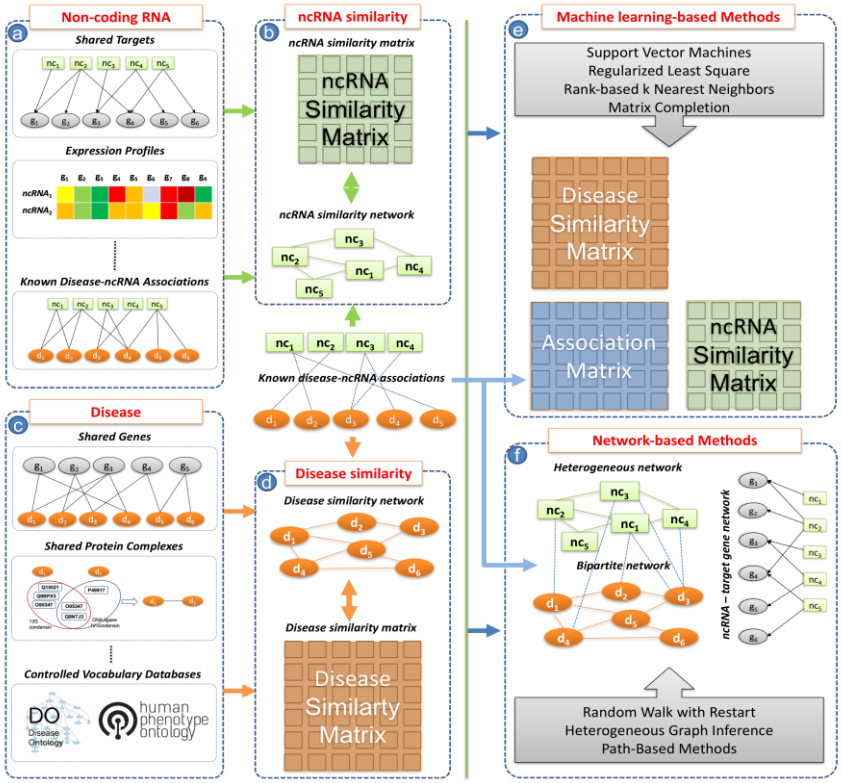

Figure 1. A general computational framework for predicting disease-associated ncRNAs. **(a)** Data sources for calculating similarity between ncRNAs. **(b)** Data sources for calculating similarity between diseases. **(c)** Similarity among ncRNAs represented in similarity network/matrix. **(d)** Similarity among ncRNAs represented in similarity network/matrix. (e) Machine learning-based methods proposed based on matrix representation of the similarities. **(f)** Network-based methods proposed based on network representation of the similarities.

Many proposed computational methods for prediction of disease-associated have been reviewed separately in [8, 9] for miRNAs and [7] for lncRNAs. Although, details of them were described for each type of ncRNAs, however a general computational framework has not been proposed irrespective of that prediction of disease-associated miRNAs and lncRNAs are very similar in the view of algorithm. Roughly, two main approaches have been proposed using machine learning techniques (i.e., machine learning-based) or methods on biological networks (i.e., network-based). In general, network-based methods formulated the prediction of disease-associated ncRNAs as a ranking problem, where candidate ncRNAs are ranked according to their relevance to a disease of interest. Meanwhile some of machine learning-based methods considered the problem as a binary classification, where candidate ncRNAs are determined to be associated/not associated with the disease of interest. Even though, they usually use similar input data such as disease similarity, ncRNA similarity and known disease-ncRNA information, but in different forms. More specifically, similarities of diseases and ncRNAs were embedded as networks in network-based methods, meanwhile these similarities are represented by matrices in some machine learning-based methods. In addition, known disease-ncRNA associations were represented as a bipartite network and an adjacency matrix in network- and machine learning-based methods, respectively. Figure 1 illustrates a general computational framework for predicting disease-associated ncRNAs. In following sections, we are going to summarized detail about common methods to build similarity networks/matrices of diseases and ncRNAs (focus on miRNAs and lncRNAs). Then, network- and machine learning-based methods commonly proposed for predicting both disease-associated miRNAs and lncRNAs are also reviewed. In addition, some methods proposed separately to miRNAs and lncRNAs are described.

Table 1. Disease-miRNA association databases

Database	Description	URL
miR2Disease [22]	Contains 270 manually curated disease phenotype–miRNAs associations between 53 disease phenotypes and 118 miRNAs	http://www.mir2disease.org/
HMDD [23]	Manually collected 32.281 miRNA-disease association entries which include 1102 miRNA genes, 850 diseases from 17.412 papers	http://www.cuilab.cn/hmdd
MiRCancer [24]	Provides 6.642 miRNA–cancer associations, 57.984 miRNAs and 193 human cancers curated from 5.138 papers,	http://mircancer.ecu.edu/
DbDEMC [25]	Contains 2.224 differentially expressed miRNAs in 36 cancer types, curated from 436 experiments.	http://www.picb.ac.cn/dbDEMC/
OncomiRDB [26]	A database for the experimentally verified oncogenic and tumor-suppressive microRNAs. It contains 2259 entries, 328 miRNAs and 829 targets	http://lifeome.net/database/oncomirdb/
OncomiRdbB [27]	Contains microRNAs which are known to be deregulated in various cancers	http://tdb.ccmb.res.in/OncomiRdbB/index.htm

2. Construction of similarity networks/matrices

Computational methods proposed for prediction of disease-associated ncRNAs are commonly based on an assumption that functionally similar ncRNAs are associated with similar diseases. Thus, functional similarity among diseases and ncRNAs are widely used in predicting novel disease-associated ncRNAs. Here, we summarize methods to construct ncRNA and disease similarity networks/matrices.

2.1. Construction of a ncRNA functional similarity network/matrix

A common way to construct a ncRNA functional similarity network/matrix is relying on shared targets such as target genes of miRNAs [10-14], interacting miRNAs of lncRNAs [15]. Then, weight of an interaction between two ncRNAs can be proportional to number of shared targets [11-14] or a correlation efficient between two interacting score profiles of targets [10]. Expression profiles of ncRNAs were also used to calculate similarity between lncRNAs [16] and between miRNAs [17] by correlating two expression profiles of ncRNAs. Finally, similarity between ncRNAs was also estimated using known ncRNA-disease associations. For instance, similarity matrices were generated using Gaussian interaction profile kernel similarity on known lncRNA-disease associations [16, 18], known miRNA-disease associations [17, 19-21]. Figure 1(a) and (b) demonstrate the source information used for calculating similarities between ncRNAs and network/matrix representations of these similarities

2.2. Construction of disease similarity networks/matrices

To explore human diseasome, a number of computational methods have been proposed to construct a "human disease network" [28]. The simplest way to build such the network is based on shared genes [29]. More specifically, two diseases are connected to each other if they share at least one gene in which mutations are associated with both diseases. In similar way, a miRNA-associated disease network is constructed if any two diseases share one common associated miRNAs [30]. In addition to shared single cellular components, the disease similarity networks were also constructed based on functional modules such as pathways [31] and protein complexes [32]. Moreover, controlled vocabulary databases describing diseases such as disease ontology (DO) [33], human phenotype ontology (HPO) [34] and medical subject headings (MeSH) [35] were used to build disease similarity network using semantic similarity measures [36-38]. Finally, disease-disease associations can be estimated by fusing molecular data [39, 40]. Figure 1 (c) and (d) demonstrate different ways to calculate disease similarity network/matrix.

K

Table 2. Disease-lncRNA association databases

Database	Description	URL
lncRNADisease [41]	Integrate nearly 3.000 lncRNA-disease entries and 475 lncRNA interaction entries, including 914 lncRNAs and 329 diseases from ~2.000 publications. it also provides the predicted associated diseases of 1.564 human lncRNAs.	http://www.cuilab.cn/lncrnadisease
Lnc2Cancer [42]	Contains 4.989 entries of associations between 1.614 human lncRNAs and 165 human cancer subtypes through review of more than 6.500 published papers	http://www.bio-bigdata.net/lnc2cancer

3. Known disease-ncRNA association databases

In addition to disease and ncRNA similarity networks/matrices, known disease-ncRNA associations were used. For network-based methods, these associations were represented as a bipartite network and used to connect the similarity networks. For machine learning-based methods, they are labeled data for training or represented by an association matrix in computational models (see Figure 1(e) and (f)). Table 1 and 2 show known disease-miRNA association and known disease-lncRNA association databases, respectively.

4. Computational methods

From the algorithmic view, prediction of disease-associated miRNAs and lncRNAs is very similar. This can be formulated as a ranking problem where candidate miRNAs/lncRNAs are ranked based on their relevance to a disease of interest. Meanwhile, these candidates can be determined as associated/not-associated in some classification models. In addition, they can be considered as a link prediction problem in network-based models. Therefore, a number of machine learning- and network-based methods have been commonly proposed for the two problems. When prediction of disease-associated ncRNAs is formulated as a classification problem, Naïve Bayesian technique was proposed for miRNAs [43] and lncRNAs [44]. Candidate miRNAs and lncRNAs were also classified as associated/not-associated using Support Vector Machines [45, 46]. In addition, ensemble learning model such as Random Forest, which are considered more advanced than single learning models, was proposed for miRNAs [47] and for lncRNAs [48]. A limitation of the binary classification models is that negative samples (i.e., ncRNAs not associated with the disease of interest) must be defined, thus semi-supervised learning models such as Regularized Least Squares (RLS) was used for miRNAs [49] and lncRNAs [18]. Some of the machine learning-based using similarity matrices in their models such as kernels in Support Vector Machines, similarity matrices in Regularized Least Square. More recently, inductive matrix completion has been proposed for both miRNAs and lncRNAs [50, 51]. Figure 1(e) demonstrates some of the machine learning-based methods which made use of similarity matrix to predict disease-associated ncRNAs.

Similarly, a number of network-based methods were commonly proposed. A typical network propagation model, random walk with restart (RWR), which has been successfully applied for disease gene prediction [52-57], was proposed to rank candidate miRNAs [58] and lncRNAs [59] on miRNA and lncRNA similarity networks, respectively. When these ncRNA similarity networks are integrated with a disease similarity network to form a heterogeneous network of diseases and ncRNAs, then a variant of RWR, namely RWRH, was applied to better exploit the assumption "similar ncRNAs are associated with similar diseases" in predicting promising miRNAs [13] and lncRNAs [60, 61]. Another extension of RWR is to force it run on bipartite network of ncRNAs and targets genes, e.g., miRNA-target gene interaction network [11] and lncRNA-protein-coding gene network [62]. Finally, a method based on hypergeometric distribution was applied to predict both disease-associated miRNAs [10] and lncRNAs [15] using bipartite networks representing known associations. In addition to the commonly proposed network-based methods, other network propagation methods were also proposed for prediction of disease-lncRNA associations on a coding-non-coding gene-disease bipartite network [63], and on the heterogeneous network of diseases and lncRNAs using KATZ measure [64]. Figure 1(f) illustrates some of the network-based methods for prediction of disease-associated ncRNAs.

5. Conclusion

With development of next generation sequencing and high-throughput technologies in recent years, there have been great advances in not only understanding coding regions in the chromosome, but also identifying and understanding ncRNAs. Ten of thousands ncRNAs have been identified and freely accessed in public databases. However, only small set of ncRNAs have well studied about their functions, especially their roles in disease development. Therefore, computational methods to predict novel disease-associated ncRNAs are highly needed to understand roles of ncRNAs in underlying molecular mechanism of diseases. Many computational methods have been proposed for this problem and comprehensively reviewed. However, these computational methods were described separately with less connection to others, thus limits our understanding on intrinsic of the methods. In this study, we proposed a general computational framework for prediction of disease-associated ncRNAs. The framework described steps from general methods for constructing similarity network/matrices from various data sources to commonly proposed network- and machine learning-based methods. This framework could pave a way for development of more advanced computational methods for the problem in future. Moreover, unlike prediction of disease-associated protein-coding genes which have been well studied for decades, the prediction disease-associated non-coding RNAs has been focused since last few years. However, it is interesting that the two problems are very similar in the algorithmic view. Therefore, computational methods, which have been successfully applied for protein-coding genes [65-69], can be used for non-coding RNAs.

Acknowledgements

Funding: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2017.14.

References

[1] K.V. Morris, J.S. Mattick, The rise of regulatory RNA, Nature Reviews Genetics, 2014, p. 423- 437.

[2] G.W. Beadle, E.L. Tatum, Genetic Control of Biochemical Reactions in Neurospora, Proceedings of the National Academy of Sciences 27 (11) (1941) 499-506.

[3] K.R. Rosenbloom, et al., ENCODE whole-genome data in the UCSC Genome Browser: update 2012, Nucleic Acids Research 40 (D1) (2012) D912-D917.

[4] M. Esteller, Non-coding RNAs in human disease, Nat Rev Genet 12 (12) (2011) 861-874.

[5] C.-P. Lin, L. He, Noncoding RNAs in Cancer Development, Annual Review of Cancer Biology 1 (1) (2017) 163-184.

[6] C. Wahlestedt, Targeting long non-coding RNA to therapeutically upregulate gene expression, Nature Reviews Drug Discovery 12 (2013) 433-446.

[7] X. Chen et al., Long non-coding RNAs and complex diseases: from experimental results to computational models, Briefings in Bioinformatics 18 (4) (2017) 558-576.

[8] X. Zeng, X. Zhang, Q. Zou, Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks, Briefings in Bioinformatics 17 (2) (2016) 193-203.

[9] X. Chen et al., MicroRNAs and complex diseases: from experimental results to computational models, Briefings in Bioinformatics, 2017, pp. bbx130-bbx130.

[10] Q. Jiang et al., Prioritization of disease microRNAs through a human phenome-microRNAome network, BMC Systems Biology (2010) **4**(Suppl 1):S2.

[11] D.-H. Le et al., Random walks on mutual microRNA-target gene interaction network improve the prediction of disease-associated microRNAs, BMC Bioinformatics 18 (2017) 479. https://doi.org./10.1186/s12859-017-1924-1.

[12] D.H. Le, Network-based ranking methods for prediction of novel disease associated microRNAs, Computational Biology and Chemistry 58 (2015) 139-148.

[13] D.-H. Le, Disease phenotype similarity improves the prediction of novel disease-associated microRNAs, In Information and Computer Science (NICS), 2015 2nd National Foundation for Science and Technology Development Conference on 2015.

[14] D.-H. Le, K. Marchal, Integration of miRNA-miRNA networks improves the prediction of novel disease associated miRNAs, In The First NAFOSTED Conference on Information and Computer Science, Hanoi, 2014, pp. 438-448.

[15] X. Chen, Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, Scientific Reports (2015) **5**:13186.

[16] X. Chen et al., Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity, Scientific Reports (2015) **5**:11338.

[17] D. Wang et al., Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases, Bioinformatics 26 (13) (2010) 1644-1650.

[18] X. Chen, G.-Y. Yan, Novel human lncRNA–disease association inference based on lncRNA expression profiles, Bioinformatics 29 (20) (2013) 2617-2624.

[19] X. Chen et al., HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction, Oncotarget 7 (40) (2016) 65257-65269.

[20] D. Sun et al., NTSMDA: prediction of miRNA–disease associations by integrating network topological similarity, Molecular BioSystems 12 (7) (2016) 2224-2232.

[21] P. Xuan et al., Prediction of potential disease-associated microRNAs based on random walk, Bioinformatics 31 (11) (2015) 1805-1815.

[22] Q. Jiang et al., miR2Disease: A manually curated database for microRNA deregulation in human disease, Nucleic acids research 37 (suppl 1) (2009) D98-D104.

[23] Y. Li et al., HMDD v2.0: A database for experimentally supported human microRNA and disease associations, Nucleic Acids Research 42 (D1) (2014) D1070-D1074.

[24] B. Xie et al., MiRCancer: A microRNA-cancer association database constructed by text mining on literature, Bioinformatics 29 (5) (2013) 638-644.

[25] Z. Yang et al., dbDEMC 2.0: Updated database of differentially expressed miRNAs in human cancers, Nucleic Acids Research 45 (D1) (2017) D812-D818.

[26] D. Wang et al., OncomiRDB: A database for the experimentally verified oncogenic and tumor-suppressive microRNAs, Bioinformatics 30 (15) (2014) 2237-2238.

[27] R. Khurana et al., OncomiRdbB: A comprehensive database of microRNAs and their targets in breast cancer, BMC Bioinformatics (2014) **15**(1):15.

[28] K.-I. Goh, I.-G. Choi, Exploring the human diseasome: The human disease network, Briefings in Functional Genomics 11 (6) (2012) 533-542.

[29] K.-I. Goh et al., The human disease network, Proceedings of the National Academy of Sciences 104 (21) (2007) 8685-8690.

[30] M. Lu et al., An Analysis of Human MicroRNA and Disease Associations, PLoS ONE (2008) 3 (10): e3420.

[31] Y. Li, P. Agarwal, A Pathway-Based View of Human Diseases and Disease Relationships, PLoS ONE (2009) 4 (2):e4346.

[32] Q. Wang et al., Community of protein complexes impacts disease association, Eur J Hum Genet 20 (11) (2012) 1162-1167.

[33] W.A. Kibbe et al., Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data, Nucleic Acids Research 43 (D1) (2015) D1071-D1078.

[34] S. Köhler et al., The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data, Nucleic acids research 42 (D1) (2014) D966-D974.

[35] C.E. Lipscomb, Medical Subject Headings (MeSH), Bull Med Libr Assoc 88 (3) (2000) 265-266.

[36] D.-H. Le, L.T.M. Dao, Annotating Diseases Using Human Phenotype Ontology Improves Prediction of Disease-Associated Long Non-coding RNAs, Journal of Molecular Biology 430 (15) (2018) 2219-2230.

[37] Y.-A. Huang et al., ILNCSIM: improved lncRNA functional similarity calculation model, Oncotarget 7 (18) (2016) 25902-25914.

[38] X. Chen et al., FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model, Oncotarget 7 (29) (2016) 45948-45958.

[39] M. Žitnik et al., Discovering disease-disease associations by fusing systems-level molecular data, Sci, Rep., 2013, pp.3202.

[40] E. Oerton et al., Understanding and predicting disease relationships through similarity fusion, Bioinformatics, 2018, pp. bty754-bty754.

[41] G. Chen et al., LncRNADisease: A database for long-non-coding RNA-associated diseases. Nucleic Acids Research 41 (D1) (2013) D983-D986.

[42] S. Ning et al., Lnc2Cancer: A manually curated database of experimentally supported lncRNAs associated with various human cancers, Nucleic Acids Research 44 (D1) (2016) D980-D985.

[43] Q. Jiang, G. Wang, Y. Wang, An approach for prioritizing disease-related microRNAs based on genomic data integration, In Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on 2010, IEEE.

[44] T. Zhao et al., Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features, Molecular BioSystems 11 (1) (2015) 126-136.

[45] J. Qinghua et al. Predicting human microRNA-disease associations based on support vector machine, In Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on, 2010.

[46] W. Lan et al., LDAP: a web server for lncRNA-disease association prediction, Bioinformatics 33 (3) (2017) 458-460.

[47] D. Le, V. Pham, T.T. Nguyen, An ensemble learning-based method for prediction of novel disease-microRNA associations, In 2017 9th International Conference on Knowledge and Systems Engineering (KSE), 2017.

[48] X. Pan, L.J. Jensen, J. Gorodkin, Inferring disease-associated long non-coding RNAs using genome-wide tissue expression profiles, Bioinformatics, 2018, pp. bty859-bty859.

[49] X. Chen, G.-Y. Yan, Semi-supervised learning for potential human microRNA-disease associations inference, Scientific Reports (2014) 4:5501.

[50] C. Lu et al., Prediction of lncRNA–disease associations based on inductive matrix completion, Bioinformatics 34 (19) (2018) 3357-3364.

[51] X. Chenet al., Predicting miRNA–disease association based on inductive matrix completion, Bioinformatics, 2018, pp. bty503-bty503.

[52] D.-H. Le, V.-T. Dang, Ontology-based disease similarity network for disease gene prediction, Vietnam Journal of Computer Science, 2016, pp. 1-9.

[53] D.-H. Le, Y.-K. Kwon, Neighbor-favoring weight reinforcement to improve random walk-based disease gene prioritization, Computational Biology and Chemistry 44 (0) (2013) 1-8.

[54] D.-H. Le, V.-H. Pham, HGPEC: A Cytoscape app for prediction of novel disease-gene and disease-disease associations and evidence collection based on a random walk on heterogeneous network, BMC Systems Biology (2017) 11 (1):61.

[55] D.-H. Le, Y.-K. Kwon, GPEC: A Cytoscape plug-in for random walk-based gene prioritization and biomedical evidence collection, Computational Biology and Chemistry 37 (0) (2013) 17-23.

[56] S. Kohler et al., Walking the Interactome for Prioritization of Candidate Disease Genes, The American Journal of Human Genetics 82 (4) (2008) 949-958.

[57] Y. Li, J.C. Patra, Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network, Bioinformatics 26 (9) (2010) 1219-1224.

[58] X. Chen, M.-X. Liu, G.-Y. Yan, RWRMDA: predicting novel human microRNA-disease associations, Molecular BioSystems 8 (10) (2012) 2792-2798.

[59] J. Sun et al., Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network, Molecular BioSystems 10 (8) (2014) 2074-2081.

[60] M. Zhou et al., Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network, Molecular BioSystems 11 (3) (2015) 760-769.

[61] G.U. Ganegoda et al., Heterogeneous Network Model to Infer Human Disease-Long Intergenic Non-Coding RNA Associations, NanoBioscience, IEEE Transactions on 14 (2) (2015) 175-183.

[62] Y. Liu et al., Construction of a lncRNA-PCG bipartite network and identification of cancer-related lncRNAs: a case study in prostate cancer, Molecular BioSystems 11 (2) (2015) 384-393.

[63] X. Yang et al., A Network Based Method for Analysis of lncRNA-Disease Associations and Prediction of lncRNAs Implicated in Diseases, PLOS ONE (2014) 9(1):e87797.

[64] X. Chen, KATZLDA: KATZ measure for the lncRNA-disease association prediction, Scientific Reports 5 (2015) 16840.

[65] X. Wang, N. Gulbahce, H. Yu, Network-based methods for human disease gene prediction. Briefings in Functional Genomics, 10 (5) (2011) 280-293.

[66] D.-H, Le, N. Xuan Hoai, Y.-K. Kwon, A Comparative Study of Classification-Based Machine Learning Methods for Novel Disease

Gene Prediction, in Knowledge and Systems Engineering, V.-H. Nguyen, A.-C. Le, and V.-N. Huynh, Editors, Springer International Publishing, 2015, p. 577-588.

[67] D.-H. Le, M.-H. Nguyen, Towards more realistic machine learning techniques for prediction of disease-associated genes, in Proceedings of the Sixth International Symposium on Information and Communication Technology. 2015, ACM: Hue City, Viet Nam. p. 116-120.

[68] M.G. Kann, Advances in translational bioinformatics: computational approaches for the hunting of disease genes, Briefings in Bioinformatics 11 (1) (2009) 96-110.

[69] R.M. Piro, F. Di Cunto, Computational approaches to disease-gene prediction: rationale, classification and successes, FEBS Journal 279 (5) (2012) 678-696.