

Hui, Zhang; Yanchun, Liang; Cheng, Peng; Siyu, Han; Wei, Du; Ying, Li

Predicting lncRNA-disease associations using network topological similarity based on deep mining heterogeneous networks. (English) Zbl 1425.92088

Math. Biosci. 315, Article ID 108229, 9 p. (2019).

Summary: A kind of noncoding RNA with length more than 200 nucleotides named long noncoding RNA (lncRNA) has gained considerable attention in recent decades. Many studies have confirmed that human genome contains many thousands of lncRNAs. LncRNAs play significant roles in many important biological processes, including complex disease diagnosis, prognosis, prevention and treatment. For some important diseases such as cancer, lncRNAs have been novel candidate biomarkers. However, the role of lncRNAs in human diseases is still in its infancy, and only a small part of lncRNA-disease associations have been experimentally verified. Predicting lncRNA-disease association is an important way to understand the mechanism and function of lncRNA involved in diseases to enrich the annotations of lncRNA. Therefore, it is urgent to prioritize lncRNAs potentially associated with diseases. Biological system is a highly complex heterogeneous network involved different molecules. Therefore, the algorithms based on network methods have been extensively applied in information fields which can provide a quantifiable characterization for the networks characterizing multifarious biological systems. A heterogeneous network topology possessing abundant interactions between biomedical entities is rarely utilized in similarity-based methods for predicting lncRNA-disease associations based on the array of varying features of lncRNAs and diseases. DeepWalk, encoding the relations of nodes in a continuous vector space, is an extension of language model and unsupervised learning from sequence-based word to network. In this article, we present a novel lncRNA-disease association prediction method based on DeepWalk, which enhances the existing association discovery methods through a topology-based similarity measure. We integrate the heterogeneous data to construct a linked tripartite network which is a heterogeneous network containing three types of nodes which generated from bioinformatics linked datasets and use DeepWalk method to extract topological structure features of the nodes in the linked tripartite network for calculating similarities. Our proposed method can be separated into the following steps: Firstly, we integrate heterogeneous data to construct a linked tripartite network: containing the topological interactions of known lncRNA-disease, lncRNA-microRNA and microRNA-disease. Secondly, the topological structure features of the nodes are extracted based on DeepWalk. Thirdly, similarity scores of disease-disease pairs and lncRNA-lncRNA pairs are computed based on the topology of this network. Finally, new lncRNA and disease associations are discovered by rule-based inference method with lncRNA-lncRNA similarities. Our proposed method shows superior predictive performance for prediction of lncRNA-disease associations based on topological similarity from heterogeneous network. The AUC value is used to show the performance of our method. The similarity measurement using network topology based on DeepWalk provide a novel perspective which is different from the similarity derived from sequence or structure information.

Availability: All the data and codes are freely availability at: <https://github.com/Pengeace/lncRNA-disease-link>.

MSC:

- 92C42 Systems biology, networks
- 92C40 Biochemistry, molecular biology
- 68T05 Learning and adaptive systems in artificial intelligence

Cited in 1 Document

Keywords:

deep learning; lncRNA-disease association prediction; similarity measure; biological network science

Software:

ncPred; word2vec; DeepWalk; LINE; LncADeep; LncDisease; GitHub; node2vec; starBase

Full Text: [DOI](#)

References:

- [1] Guttman, M.; Russell, P.; Ingolia, N. T.; Weissman, J. S.; Lander, E. S., Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins, *Cell*, 154, 240-251, (2013)
- [2] Esteller, M., Non-coding RNAs in human disease, *Nat. Rev. Genet.*, 12, 861-874, (2011)
- [3] Wang, K. C.; Chang, H. Y., Molecular mechanisms of long noncoding RNAs, *Mol. Cell*, 43, 904-914, (2011)
- [4] Wapinski, O.; Chang, H. Y., Long noncoding RNAs and human disease, *Trends Cell Biol.*, 21, 354-361, (2011)
- [5] Chen, X.; Yan, G. Y., Novel human lncRNA-disease association inference based on lncRNA expression profiles, *Bioinformatics*, 29, 2617-2624, (2013)
- [6] Mercer, T. R.; Dinger, M. E.; Mattick, J. S., Insights into functions, *Nat. Rev. Genet.*, 10, 155-159, (2009)
- [7] Johnson, R., Long non-coding RNAs in Huntington's disease neurodegeneration, *Neurobiol. Dis.*, 46, 245-254, (2012)
- [8] Ouimet, M.; Drouin, S.; Lajoie, M.; Caron, M.; St-Onge, P.; Gioia, R.; Richer, C.; Sinnett, D., A childhood acute lymphoblastic leukemia-specific lncRNA implicated in prednisolone resistance, cell proliferation, and migration, *Oncotarget*, 8, 7477-7488, (2017)
- [9] Congrains, A.; Kamide, K.; Oguro, R.; Yasuda, O.; Miyata, K.; Yamamoto, E.; Kawai, T.; Kusunoki, H.; Yamamoto, H.; Takeya, Y.; Yamamoto, K.; Onishi, M.; Sugimoto, K.; Katsuya, T.; Awata, N.; Ikebe, K.; Gondo, Y.; Oike, Y.; Ohishi, M.; Rakugi, H., Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B, *Atherosclerosis*, 220, 449-455, (2012)
- [10] Chen, X.; Yan, C. C.; Zhang, X.; You, Z.-H., Long non-coding RNAs and complex diseases: from experimental results to computational models, *Brief. Bioinform.*, 18, (2016), bbw060
- [11] Chen, X.; Yan, C. C.; Zhang, X.; You, Z.-H., Long non-coding RNAs and complex diseases: from experimental results to computational models, *Brief. Bioinform.*, 18, (2016), bbw060
- [12] Zhang, Y.; Tao, Y.; Liao, Q., Long noncoding RNA: a crosslink in biological regulatory network, *Brief. Bioinform.*, 1-16, (2017)
- [13] Gu, Y.; Chen, T.; Li, G.; Yu, X.; Lu, Y.; Wang, H.; Teng, L., LncRNAs: emerging biomarkers in gastric cancer, *Future Oncol.*, 11, 2427-2441, (2015)
- [14] Chen, G.; Wang, Z.; Wang, D.; Qiu, C.; Liu, M.; Chen, X.; Zhang, Q.; Yan, G.; Cui, Q., LncRNADisease: a database for long-non-coding RNA-associated diseases, *Nucleic Acids Res.*, 41, 983-986, (2013)
- [15] Fang, S.; Zhang, L.; Guo, J.; Niu, Y.; Wu, Y.; Li, H.; Zhao, L.; Li, X.; Teng, X.; Sun, X.; Sun, L.; Zhang, M. Q.; Chen, R.; Zhao, Y., NONCODEV5: a comprehensive annotation database for long non-coding RNAs, *Nucleic Acids Res.*, 46, D308-D314, (2018)
- [16] Zhou, M.; Wang, X.; Li, J.; Hao, D.; Wang, Z.; Shi, H.; Han, L.; Zhou, H.; Sun, J., Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network, *Mol. BioSyst.*, 11, 760-769, (2015)
- [17] Chen, X.; You, Z.-H.; Yan, G.-Y.; Gong, D.-W., IRWRLDA: improved random walk with restart for lncRNA-disease association prediction, *Oncotarget*, 7, 57919-57931, (2016)
- [18] Peng, W.; Lan, W.; Yu, Z.; Wang, J.; Pan, Y., A framework for integrating multiple biological networks to predict microRNA-disease associations, *IEEE Trans. Nanobiosci.*, 14, (2016), 1-1
- [19] Chen, X.; Yan, C. C.; Zhang, X.; You, Z.-H., Long non-coding RNAs and complex diseases: from experimental results to computational models, *Brief. Bioinform.*, 18, (2016), bbw060
- [20] Chen, X., Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, *Sci. Rep.*, 5, 1-11, (2015)
- [21] Huang, Y.-A.; Chen, X.; You, Z.-H.; Huang, D.-S.; Chan, K. C.C., ILNCSIM: improved lncRNA functional similarity calculation model, *Oncotarget*, 7, 7-14, (2016)
- [22] Zhao, T.; Xu, J.; Liu, L.; Bai, J.; Xu, C.; Xiao, Y.; Li, X.; Zhang, L., Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features, *Mol. Biosyst.*, 11, 126-136, (2015)
- [23] Q. Wang, Junyi, Ma, Ruixia, Cui, LncDisease: a sequence based bioinformatics tool for predicting lncRNA-disease association, (2016). doi:10.1093/nar/gkw093.
- [24] W. Lan, M. Li, K. Zhao, J. Liu, F. Wu, Y. Pan, Subject section LDAP : a web server for lncRNA-disease association prediction, (2016) 3-5.
- [25] Yu, G.; Fu, G.; Lu, C.; Ren, Y.; Wang, J., BRWLDA: bi-random walks for predicting lncRNA-disease associations, *Oncotarget*, (2017)
- [26] Sun, J.; Shi, H.; Wang, Z.; Zhang, C.; Liu, L.; Wang, L.; He, W.; Hao, D.; Liu, S.; Zhou, M., Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network, *Mol. BioSyst.*, 10, 2074-2081, (2014)
- [27] Ganegoda, G. U.; Li, M.; Wang, W.; Feng, Q., Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations, *IEEE Trans. Nanobiosci.*, 14, 175-183, (2015)
- [28] Chen, X., KATZLDA: KATZ measure for the lncRNA-disease association prediction, *Sci. Rep.*, 5, 1-11, (2015)
- [29] Yao, Q.; Wu, L.; Li, J.; Yang, L. G.; Sun, Y.; Li, Z.; He, S.; Feng, F.; Li, H.; Li, Y., Global prioritizing disease candidate lncRNAs via a multi-level composite network, *Sci. Rep.*, 7, 1-14, (2017)
- [30] Fu, G.; Wang, J.; Domeniconi, C.; Yu, G., Matrix factorization-based data fusion for the prediction of lncRNA - disease associations, *Bioinformatics*, 34, 1-9, (2017)
- [31] Wang, P.; Guo, Q.; Gao, Y.; Zhi, H.; Zhang, Y.; Liu, Y., Improved method for prioritization of disease associated lncRNAs based on ceRNA theory and functional genomics data, *Oncotarget*, 8, 4642-4655, (2017)

- [32] Yang, X.; Gao, L.; Guo, X.; Shi, X.; Wu, H.; Song, F.; Wang, B., A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases, *PLoS One*, 9, (2014)
- [33] Alaimo, S.; Giugno, R.; Pulvirenti, A., ncPred : ncRNA-disease association prediction through tripartite network-based inference, *Front. Bioeng. Biotechnol.*, 2, 1-8, (2014)
- [34] Perozzi, B.; Al-Rfou, R.; Skiena, S., DeepWalk: online learning of social representations Bryan, (Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14, (2014)), 701-710
- [35] Grover, A.; Leskovec, J., Node2Vec, (Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, (2016)), 855-864
- [36] Tang, J.; Qu, M., LINE : large-scale information network embedding categories and subject descriptors, (*ACM World Wide Web.*, (2015)), 1067-1077
- [37] Chen, X., Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, *Sci. Rep.*, 5, 1-12, (2015)
- [38] Ning, S.; Zhang, J.; Wang, P.; Zhi, H.; Wang, J.; Liu, Y.; Gao, Y.; Guo, M.; Yue, M.; Wang, L.; Li, X., Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers, *Nucleic Acids Res.*, 44, D980-D985, (2016)
- [39] Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q.; v2, HMDD, A database for experimentally supported human microRNA and disease associations, *Nucleic Acids Res.*, 42, 1070-1074, (2014)
- [40] Xie, B.; Ding, Q.; Han, H.; Wu, D., MiRCancer: a microRNA-cancer association database constructed by text mining on literature, *Bioinformatics*, 29, 638-644, (2013)
- [41] Jiang, Q.; Wang, Y.; Hao, Y.; Juan, L.; Teng, M.; Zhang, X.; Li, M.; Wang, G.; Liu, Y., miR2Disease: a manually curated database for microRNA deregulation in human disease, *Nucleic Acids Res.*, 37, 98-104, (2009)
- [42] You, Z. H.; Huang, Z. A.; Zhu, Z.; Yan, G. Y.; Li, Z. W.; Wen, Z.; Chen, X., PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction, *PLoS Comput. Biol.*, 13, (2017)
- [43] Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H., starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, *Nucleic Acids Res.*, 42, D92-D97, (2014)
- [44] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, (2013) 1-12. doi:10.1162/153244303322533223.
- [45] Mnih, A.; Hinton, G. E., A scalable hierarchical distributed language model, *Adv. Neural Inf. Process. Syst.*, 1-8, (2008), doi:10.1.1.205.5467
- [46] Yang, C.; Yang, L.; Zhou, M.; Xie, H.; Zhang, C.; Wang, M. D.; Zhu, H., LncADeep : an ab initio lncRNA identification and functional annotation tool based on deep learning, *Bioinformatics*, 34, 3825-3834, (2018)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.