SCIENTIFIC REPORTS

OPEN

Fusion of multiple heterogeneous networks for predicting circRNA-disease associations

Received: 22 January 2019
Accepted: 18 June 2019
Published online: 03 July 2019

Lei Deng[1], Wei Zhang[1], Yechuan Shi[1] & Yongjun Tang[2]

Circular RNAs (circRNAs) are a newly identified type of non-coding RNA (ncRNA) that plays crucial roles in many cellular processes and human diseases, and are potential disease biomarkers and therapeutic targets in human diseases. However, experimentally verified circRNA-disease associations are very rare. Hence, developing an accurate and efficient method to predict the association between circRNA and disease may be beneficial to disease prevention, diagnosis, and treatment. Here, we propose a computational method named KATZCPDA, which is based on the KATZ method and the integrations among circRNAs, proteins, and diseases to predict circRNA-disease associations. KATZCPDA not only verifies existing circRNA-disease associations but also predicts unknown associations. As demonstrated by leave-one-out and 10-fold cross-validation, KATZCPDA achieves AUC values of 0.959 and 0.958, respectively. The performance of KATZCPDA was substantially higher than those of previously developed network-based methods. To further demonstrate the effectiveness of KATZCPDA, we apply KATZCPDA to predict the associated circRNAs of Colorectal cancer, glioma, breast cancer, and Tuberculosis. The results illustrated that the predicted circRNA-disease associations could rank the top 10 of the experimentally verified associations.

Circular RNA (circRNA) is a class of non-coding RNA recently discovered. Unlike linear RNA, circRNA forms a continuous cycle of covalent closures and is highly represented in the eukaryotic transcriptome. Previous research has found thousands of prototype circRNAs in human, mouse and nematode cells[1–4]. As the report goes, circular RNA in higher organisms were produced by reverse splicing events and synthesized from all regions of the genome, mainly from exons, and a few from antisense, intergenic, intragenic and intron regions[5].

The expression level of circRNA is low, and thus, it was initially thought that circRNA was a by-product of splice-mediated splicing errors or an intermediate that escaped from the intron lariat[6–8]. Therefore, circRNA received little attention in the past. However, with the development of high-throughput sequencing technology and computational analysis techniques, thousands of circRNAs have been discovered in many species ranging from archaea to humans, and the expression level of some circRNAs was ten-fold higher than those obtained from the standard linear transcription of homologous genes[3,4,9–13].

A large number of studies have revealed many circRNA functions, such as serving as scaffolds in the assembly of protein complexes, isolating proteins from their natural subcellular localization, regulating the expression of parental genes, modulating alternative splicing and RNA-protein interactions, and functioning as microRNA (miRNA) sponges[10,14–18]. In addition to their potential function such as significant regulators of gene expression, circRNAs were reported to be related to many different human diseases, including neurodegenerative disorders and cerebrovascular diseases. In particular, experiments have shown that many circRNAs are closely related to cancer[19–21], and some experimental evidence demonstrated that circRNA plays an essential role in atherosclerotic vascular diseases, prion diseases and cancers of the nervous system, especially exhibiting abnormal expression level in colorectal cancer (CRC) and pancreatic ductal adenocarcinoma (PDAC). In this way, the circRNA could act as a biomarker for the diagnosis and prediction of some diseases in the future.

Several circRNA related resource databases have recently been established. The circBase database[22] combines data from several circRNAs, including circular RNA IDs, genomic coordinates, and optimal transcripts, into a standardized database. The CircNet database[23] provides a new circRNA identification tool that offers annotation of genomic circRNA isoforms and circRNA subtype sequences by integrating circRNA-miRNA-mRNA regulatory

[1]School of Computer Science and Engineering, Central South University, Changsha, 410075, China. [2]Department of Pediatrics, Xiangya Hospital, Central South University, Changsha, 410008, China. Correspondence and requests for materials should be addressed to Y.T. (email: tangyj11bhyc@163.com)

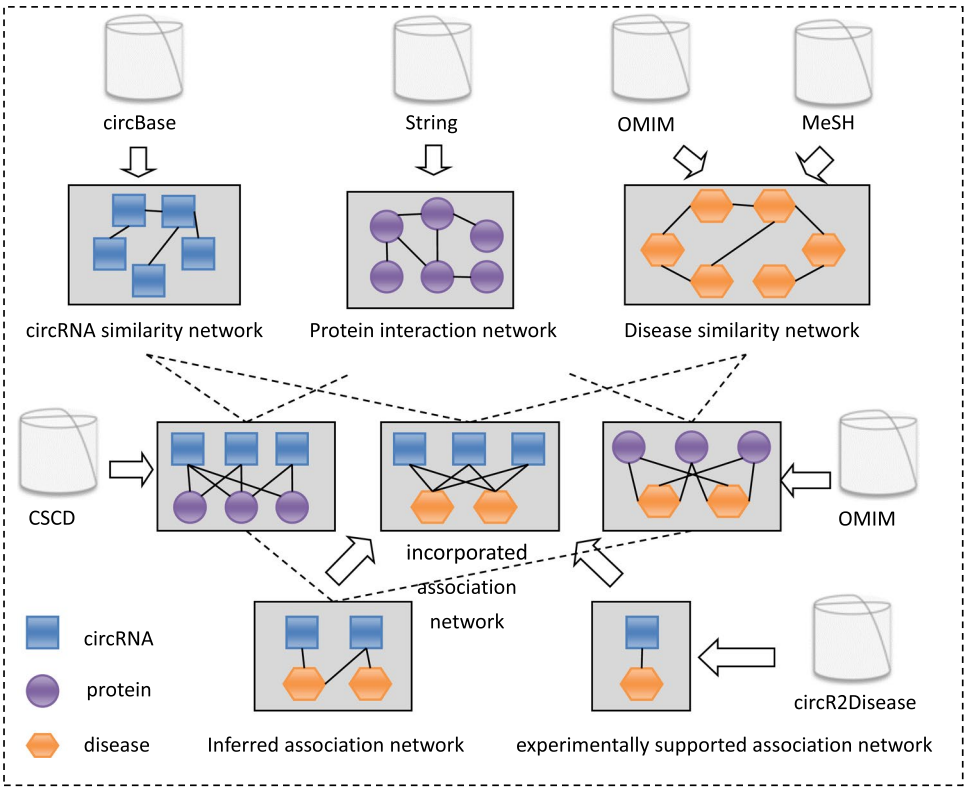

Figure 1. Flowchart for constructing an incorporated circRNA-disease association network.

networks. The Tissue-Specific CircRNA Database (TSCD)[24] provides circRNAs obtained from cancer cells with four algorithms, and the corresponding features of circRNAs, such as cancer-specific circRNAs (CS-circRNAs), RBP-binding sites in CS-circRNAs, cancer-specific alternative splicing associated with CS-circRNAs, miRNA target sites in CS-circRNAs, and possible open reading frames in CS-circRNAs. The CircInteractome database[25] maps RNA-binding protein (RBP) sites on circRNAs, which can be used to search for potential interactions between circRNAs and RBPs or miRNAs, and for potential internal ribosomal entry sites. SomamiR2.0[26] provides target sites for mutations in tumor cells or miRNA cells, while this kind of mutation might alter the interaction of miRNAs with circRNAs. Circ2Traits[27] first validated the interaction between circRNAs and miRNAs, calculating those circRNAs which are likely to be associated with a disease. This database then identified the Argonaute (Ago) interaction site on circRNAs. The Cancer-Specific CircRNA Database (CSCD)[28] provides circRNAs obtained from cancer cells using four algorithms and the corresponding features of circRNAs, such as cancer-specific circRNAs (CS-circRNAs), RBP-binding sites in CS-circRNAs, cancer-specific alternative splicing associated with CS-circRNAs, miRNA target sites in CS-circRNAs, and potential open reading frames in CS-circRNAs. The CircR2Disease database[29] provides experimentally demonstrated circRNA-disease associations and includes detailed information on the associations, such as the disease names, circRNA names, expression patterns, detection methods and simple descriptions. However, the circRNA-disease associations supported by experimental evidence remain relatively rare.

In this study, based on the "gilt-by-association" (GBA) principle, which states that biological entities having the same or related behaviour tend to be associated[30], we assumed that circRNAs associated with the same protein tend to be associated with protein-related diseases. Based on existing resources and previous studies on endogenous non-coding RNAs with disease associations and the GBA principle, we proposed a computational model named KATZCPDA to predict circRNA-disease associations. We first obtained an inferred circRNA-disease association network from a known circRNA-protein association network and a protein-disease association network. Then we gained a known circRNA-disease association network from the circR2Disease database and finally integrated the inferred network with the known network to achieve an incorporated circRNA-disease association network. The KATZCPDA model thus predicts potential circRNA-disease associations using the KATZ method integrated with the combined circRNA-disease association network, disease similarity network, and circRNA similarity network (Fig. 1). Based on the circRNA-disease associations supported by experiments included in the CircR2Disease database[29], we analysed KATZCPDA using leave-one-out cross validation (LOOCV) and 10-fold cross-validation. The results showed that KATZCPDA achieved a significantly higher performance than the existing methods.

Results

Datasets. *circRNA similarity matrix.* We obtained the circRNA expression profiling data from the work of Peng *et al.*[31], which include expression profiles of 2,895 human circRNAs. The circRNA similarity matrix *CS* was built by computing Pearson's correlation coefficient (PCC) between the expression profiles of each pair of circRNAs. If the PCC score between circRNA *i* and circRNA *j* is lower than the threshold, we set *CS(i, j)* to be 0. Otherwise, we updated *CS(i, j)* to be 1.

Matrix of circRNA-protein associations. The circRNA-protein association dataset was downloaded and compiled from the CSCD database[28] (http://gb.whu.edu.cn/CSCD/), which deposit more than 270,000 cancer-specific circRNAs. CSCD also include circRNA binding proteins (RBPs). Based on the circRNA-protein association dataset, we used the adjacency matrix *CP* to describe the association network between circRNAs and proteins: if circRNA *i* is associated with protein *j*, *CP(i,j)* is set to be 1.

Matrix of protein-disease associations. The OMIM database[32] (http://www.omim.org/downloads/) contains information on known Mendelian disorders and over 15,000 genes. Here, we choose to use the associations between proteins and phenotypes updated in October 2018. The adjacency matrix *PD* was used to indicate the functional similarity between proteins and diseases. If protein *i* is associated with disease *j*, $PD(i,j) = 1$; otherwise, $PD(i, j) = 0$.

Disease similarity matrix. The disease similarity matrix consists of integrated phenotypic information. Because the disease names in CircR2Disease[29] are not standardized (the index is not corresponding to the standard database, such as ENSEMBL and RefSeq), we obtained disease-related indexes via manual matching. First, we collected all diseases from the confirmed circRNA-diease association to obtain the list of disease names and then manually searched for each disease in the OMIM database to obtain the closest correlation phenotype ID (In the OMIM database, a prefix of none, % or # usually means that the ID provides a phenotype description). To ensure the accuracy of the data, the diseases that failed to match the phenotype ID in the OMIM database and the corresponding circRNA-disease associations were removed. The disease similarity matrix was obtained using the text mining method developed by Driel *et al.*[33], in which the entity *DS(i,j)* in the *ith* row and the *jth* column represents the disease similarity score between diseases *d(i)* and *d(j)*. According to Oron Vanunu *et al.*[34], similarity scores greater than 0 and less than 0.3 are not informative, while similarity scores greater than 0.6 and less than 1 indicate informative similarity, illustrating a potential similarity between these two diseases. In this study, if the similarity score was less than the threshold 0.4, we replaced the similarity score with 0. If the similarity score was greater than the 0.4, we updated the similarity score to 1.

Matrix of circRNA-disease associations. Seven hundred forty circRNA-disease associations were downloaded from the CircR2Disease database[29] (http://bioinfo.snnu.edu.cn/). We obtained a dataset of 263 high-quality circRNA-disease associations containing 222 circRNAs and 46 diseases. Since the experiment determined circRNA-disease associations in CircR2Disease is limited, we obtained an inferred circRNA-disease network by integrating the collected circRNA-protein associations and protein-disease associations. Based on the inferred circRNA-disease association network, we built an integrated circRNA-disease association network G_{mix} for the KATZCPDA computational model.

Evaluation measures. In this section, we evaluated the performance of the proposed method through leave-one-out cross-validation (LOOCV) and 10-fold cross-validation. In the LOOCV, each circRNA-disease association was individually left out in turn to form the test set, and remaining disease-circRNA associations were used as to train the model. In the 10-fold cross-validation, we randomly divided the circRNA-disease associations into ten subsets. And then we left out one subset as the test set, using the remaining nine subsets to train the model.

With both LOOCV and 10-fold cross-validation, for each query (circRNA) node, its predicted association score with all target (disease) nodes can be obtained. We generated a plot of the ROC curves according to the false positive rate (FPR) and true positive rate (TPR) using each iteration for different thresholds. The simultaneous calculation of the area of the ROC curve yielded the AUC value that can be used to assess overall performance.

Effects of inferred circRNA-disease associations. To demonstrate the effects of the inferred circRNA-disease associations established with protein information, we tested two different networks via LOOCV and 10-fold cross validation: (1) only circRNA disease associations with experimentally confirmed circRNA-disease association networks and (2) circRNA-disease association networks that contain both experimentally validated and inferred circRNA-disease associations (see the Results section for a description of the partial circRNA-disease association network).

As shown in Figs 2 and 3, the LOOCV and 10-fold cross-validation using both experimentally supported and inferred associations showed better performance than that established with only experimentally supported associations. In LOOCV, the use of both experimentally validated and inferred associations yielded an AUC value of 0.95914, and the use of only associations supported by experimental evidence yielded an AUC value of 0.87926. In 10-fold cross-validation, the AUC value obtained from both experimentally supported and inferred association was 0.95874, and that value obtained from the only associations supported by experimental evidence was 0.87246.

Comparison with other methods. To further evaluate the performance of our approach, we compared KATZCPDA with three other predictive approaches (LncRDNetFlow[30], TPGLDA[35], and BiRW[36]), which can predict associations between various biological entities based on integrated network information. We trained

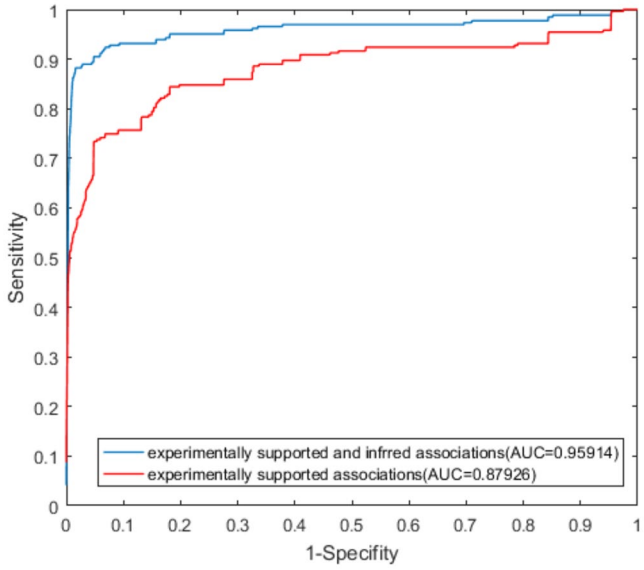

Figure 2. ROC curves from LOOCV using only experimentally validated associations and both experimentally validated and inferred associations.

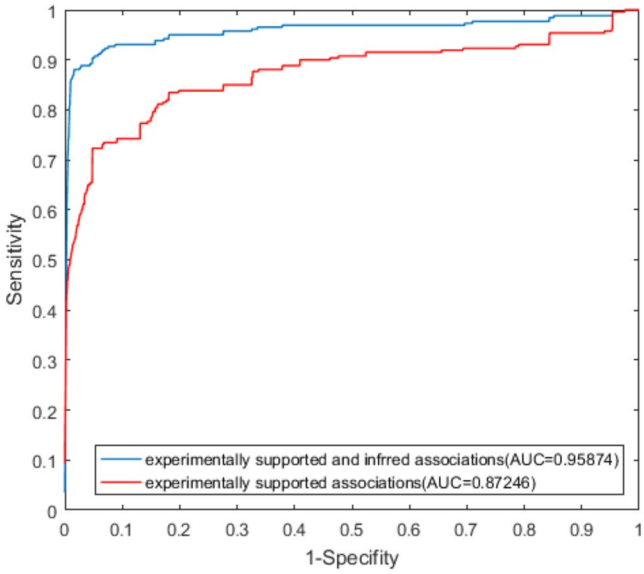

Figure 3. ROC curves from 10-fold cross validation using only experimentally validated associations and both experimentally validated and inferred associations.

and tested all the models on the same dataset. As it shown in Fig. 4, the AUC value for KATZCPDA obtained by LOOCV was 0.95914, and this value was significantly higher than the AUC values of the other three methods (LncRDNetFlow: 0.88437, TPGLDA: 0.6969 and BiRW: 0.725202). Similarly, as shown in Fig. 5, in the 10-fold cross-validation, the AUC value obtained for KATZCPD was 0.95874, which was also higher than the AUC values obtained for the other three methods (LncRDNetFlow: 0.88249, TPGLDA: 0.69686 and BiRW: 0.5784). Therefore, compared with BiRW, KATZCPDA is stable, as proved by both LOOCV and 10-fold cross-validation, because the deletion of the number of edges in the network substantially affects BiRW, i.e., the deletion of some edges in the 10-fold cross validation led to a noticeable decrease in the BiRW performance.

Case studies. To further assess the validity of KATZCPDA, all circRNA-disease connections were utilized as training data for the models, and the diseases that were predicted to be associated with circRNAs were validated using the experimentally confirmed circRNA-disease associations in the CircR2Disease database. Here, we checked the circRNAs associated with three cancers and a diseases (colorectal cancer, glioma, breast cancer, and Tuberculosis), and Table 1 lists the corresponding rankings.

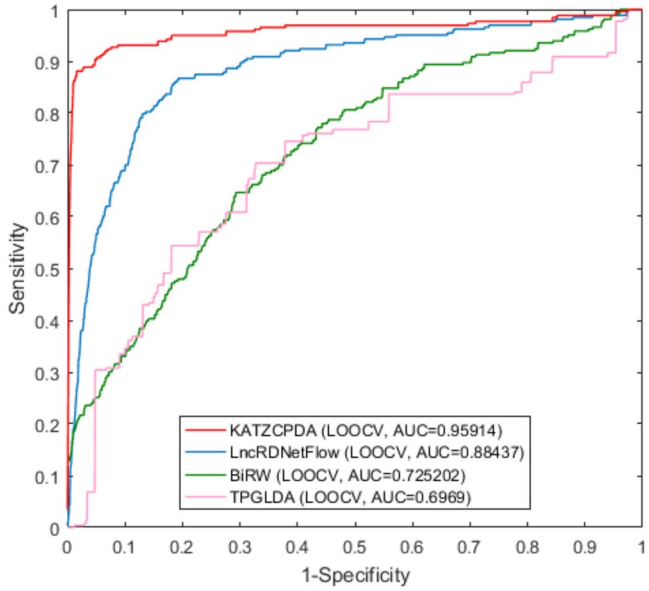

Figure 4. Comparison of the performances of KATZCPDA and other methods in terms of the ROC curve and AUC based on LOOCV.

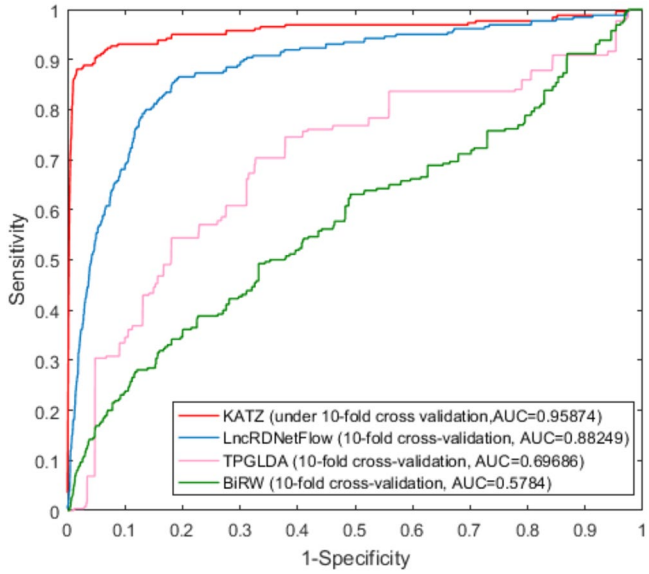

Figure 5. Comparisons of the performances of KATZCPDA and other methods in terms of ROC curve and AUC based on 10-fold cross-validation.

Colon cancer is one of the most common diseases in the world, and even in developed countries, the mortality rate of colon cancer still remains high[37]. In China, the recent prevalence of colon cancer has risen due to unhealthy lifestyles[38]. Some studies have shown that colon cancer and circRNAs are closely related. Based on this conclusion, we predicted the associations of different circRNAs with colon cancer using KATZCPDA. As a result, colon cancer ranked high in the lists of diseases that were predicted to be associated with selected circRNAs. Among the diseases that were predicted to be associated to hsa_circ_0014717, hsa_circ_0000567, hsa_circ_0020397, hsa_circ_0007031, and hsa_circ_0007534, colon cancer ranked 4th, 4th, 4th, 4th, and 6th, respectively. The CircR2Disease database verified the associations between these circRNAs and colon cancer. For example, the database indicates that hsa_circ_0014717acts as a potential tumour suppressor that inhibits CRC growth, at least partly by upregulating p16 expression[39].

Glioma is the most common primary mesenchymal tumour in the central nervous system and is the most common malignant tumour associated with morbidity and mortality. Patients with this cancer have poor prognosis because glioma is strongly invasive and aggressive. Previous studies have proved that circRNA dysregulation might be related to the occurrence and development of glioma and indicated that circRNAs can serve as

circRNAs	Diseases	KATZCPDA rank
hsa_circ_0000567	Colorectal cancer	4
hsa_circ_0008509	Colorectal cancer	4
hsa_circ_0007534	Colorectal cancer	6
hsa_circ_0007031	Colorectal cancer	6
hsa_circ_0000504	Colorectal cancer	6
hsa_circ_0000199	Glioma	3
hsa_circ_0005603	Glioma	4
hsa_circ_0006460	Glioma	5
hsa_circ_0004872	Glioma	5
hsa_circ_0008345	Glioma	7
hsa_circ_0006411	Glioma	7
hsa_circ_0011946	Breast cancer	6
hsa_circ_0001982	Breast cancer	5
hsa_circ_0002874	Breast cancer	6
hsa_circ_0085495	Breast cancer	6
hsa_circ_0001875	Breast cancer	5
hsa_circ_0000681	Tuberculosis	1
hsa_circ_0030045	Tuberculosis	2
hsa_circ_0030569	Tuberculosis	3
hsa_circ_0008797	Tuberculosis	3

Table 1. Ranking of diseases among all diseases predicted to be associated with select circRNAs.

prognostic biomarkers for glioma[40]. We analysed the relevant circRNAs using our KATZCPDA model to predict their associated diseases and found that gliomas ranked very high; specifically, glioma was ranked 5th, 5th, 7th, 5th, 7th, and 7th in the list of diseases associated with hsa_circ_0006460, hsa_circ_0005603, hsa_circ_0008345, hsa_circ_0004872, hsa_circ_0006411, and hsa_circ_0003586, respectively. Zhu *et al.*[41] confirmed that hsa_circ_0006460 is related to gliomas. In addition, circBRAF (hsa_circ_0006460) is an independent biomarker for prognosticating good progression-free survival and overall survival in glioma patients[40].

Breast cancer is the most common cancer among women worldwide. Epidemiological studies have shown that advanced age, oestrogen and progestin use, elderly primiparity, alcohol consumption and lack of physical exercise can increase the risk of breast cancer in women. In addition to genetic mutations, epigenetic mechanisms, including DNA histone modification, methylation and ncRNA, also play crucial roles in breast cancer[42]. circRNAs belonging to ncRNA are also believed to be potentially associated with breast cancer. We analysed the relevant circRNAs using the KATZCPDA calculation model and calculated their related diseases. Among these diseases, breast cancer was ranked at the top of the list of associated diseases. Among the illnesses associated with hsa_circ_0011946, hsa_circ_0001982, hsa_circ_0001785, hsa_circ_0001785, and hsa_circ_0002113, mammary gland cancer was ranked 6th, 6th, 6th, 6th, and 8th, respectively. As detailed in the database, hsa_circ_0001982 has been experimentally proved to be associated with breast cancer, and miR-143 has been demonstrated to be a target of hsa_circ_0001982 through a dual-luciferase reporter assay. In addition, loss-of-function and rescue experiments have indicated that hsa_circ_0001982 could knockdown and suppress breast cancer cells proliferation and invasion, also could induce apoptosis by targeting miR-143[43].

Tuberculosis (TB) is a potentially severe infectious disease and is one of the significant threats to human health. Early correct diagnosis and fast curative treatment help prevent tuberculosis. Studies show that circRNA might serve as a potential new biomarker for tuberculosis infection[44]. We analyzed the relevant circRNAs using the KATZCPDA model and calculated the corresponding diseases. Among these predictions, tuberculosis ranks very high, even in some cases ranks the first. Some research conducted by Qian *et al.*[45] showed that circRNAs such as hsa_circ_0000681 and hsa_circ_0008797 are closely related to tuberculosis.

Discussion and Conclusion

Increasing lines of evidence show that circRNAs are closely related to many different diseases, such as Alzheimer's disease, liver cancer and lung cancer. Some studies have explored the specific dysregulation of circRNA in infections and indicated that circRNA is a promising biomarker for diagnosis, treatment, and prognosis. Because novel experimental approaches have several limitations, models that integrate multiple biological datasets to infer circRNA-disease association can be used as supplementary tools for the detection of disease biomarkers. In this study, we integrated the known associations between circRNAs and proteins, proteins and diseases to infer circRNA-disease associations. Using the inferred circRNA-disease associations and the experimentally supported circRNA-disease associations as predictors, the KATZCPDA algorithm was then developed to predict circRNA-disease associations by integrating known biological information (circRNA similarity, disease similarities, protein-protein interactions, and the associations between these entities). Even in the absence of some associations, our method predicts new circRNA-disease associations successfully. In other words, when constructing a network of circRNA-disease associations, the bioinformatic analysis of the integrated protein information can

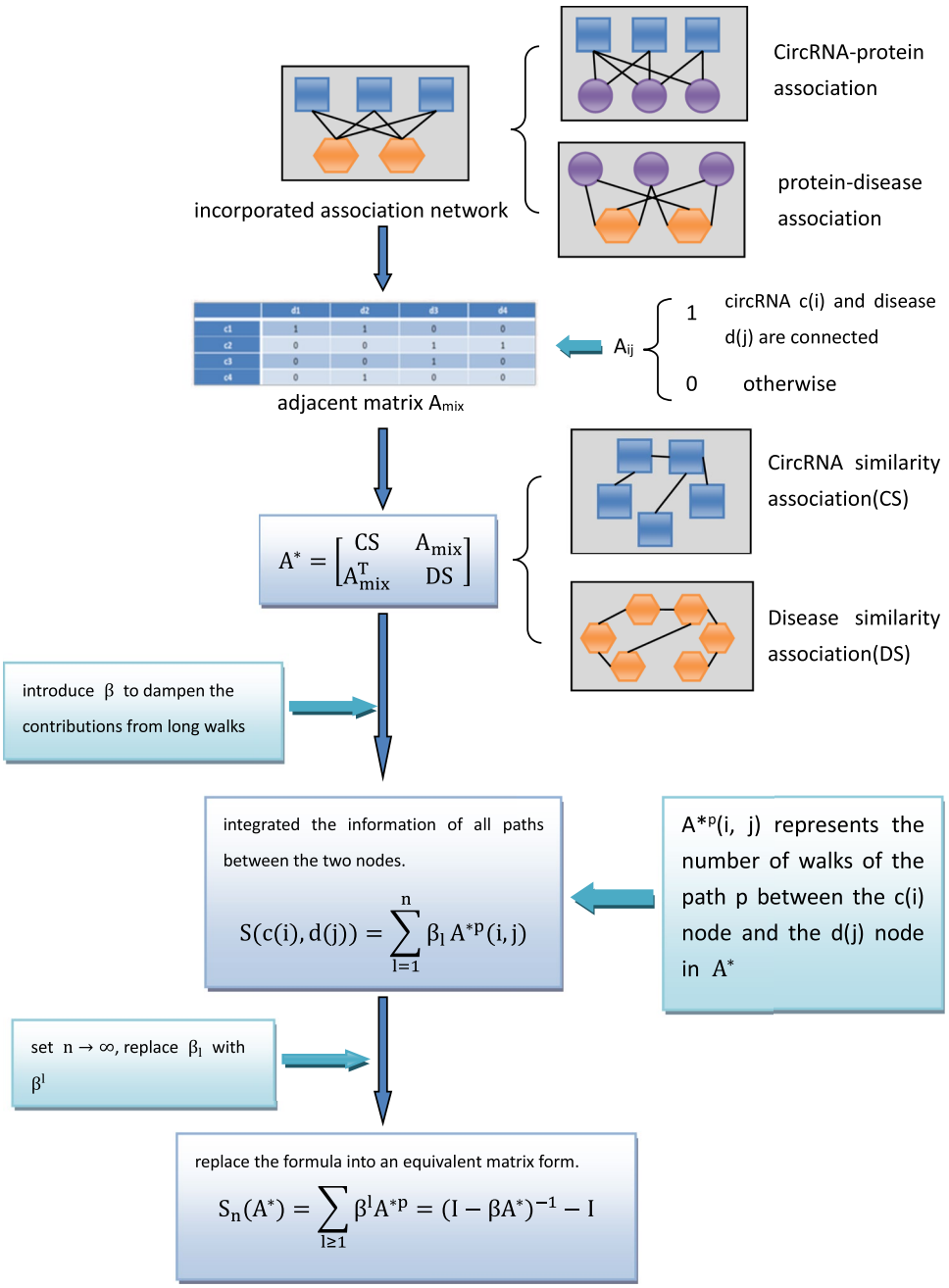

Figure 6. Flowchart of the KATZCPDA method.

infer potential information that cannot be obtained with only circRNA and disease information. In addition, the method is bidirectional because it can predict both circRNA-disease and disease-circRNA associations.

To verify the reliability of the predictive performance of KATZCPDA, we assessed different methods through LOOCV and 10-fold cross-validation using same datasets. The results showed that KATZCPDA has better performance than LncRDNetFlow, TPGLDA, and BiRW. The analysis of KATZCPDA for the prediction of the associations of circRNAs with colon cancer, glioma, breast cancer, and Tuberculosis revealed that the proposed method has excellent performance. Thus, KATZCPDA is likely to play an essential role in the identification of potential circRNA-disease associations in the future.

KATZCPDA can be improved in several aspects in the future. First, the diseases in the CircR2Disease database are not accurate. Although we obtained the most closely related OMIM ID from the OMIM database, a specific deviation in the disease similarity score might still exist. This requires discovering and integrating more reliable data. Second, the identification of circRNA similarity associations and the use of more effective methods to infer circRNA-disease associations can significantly improve the performance of the method. Third, the introduction of a higher number of intermediary entities to identify circRNA-disease associations is beneficial to increase the accuracy of the model prediction.

Methods

Network inference techniques and machine learning approaches have been widely used in many classification fields[46–56]. In this study, we proposed a KATZ measure[57] based approach (KATZCPDA) to predict unknown circRNA-disease associations by measuring the similarities between circRNAs of interest and diseases in the heterogeneous network. KATZ measurements can make successful predictions from social networks, disease-gene association networks, disease-lncRNA association networks, microbe-disease association networks, and disease-miRNA association networks[57–62]. KATZ is a graph-based calculation method that transforms the association prediction problem into the problem of calculating the similarity between nodes in a heterogeneous network. In the constructed global network, the prediction of the association between circRNA and disease nodes, is translated into the calculation of the number of walks and the range of walks connecting the corresponding circRNAs and diseases. The integration of the number of walks and length can yield the potential association probability of each circRNA-disease pair.

Fig. 6 shows a flowchart of KATZCPDA. Heterogeneous data sources were used to construct three interaction/similarity networks (circRNA, disease, and protein) and three different association networks (circRNA-protein, protein-disease, and circRNA-disease). We then generated an incorporated circRNA-disease association network by integrating these three interlinked networks. The network can be represented as an undirected graph $G_i = (V_i, E_i)$, where V_i is the set of nodes and E_i is the set of undirected edges. Each node in the network represents a biological entity (circRNA, protein, or disease), and each undirected edge represents a relationship, similarity, or interaction between the connected objects.

We assumed that W was the adjacency matrix of network G. We normalized W to obtain W' according to the topological information of the network using the normalization formula $W' = D_G^{-1/2} W D_G^{-1/2}$, where the diagonal matrix D_G is defined that $D_G(i, i)$ is the sum of the i^{th} values of W, namely, $D_G(i, i) = \sum_j W_{i,j}$, and W' is the symmetric matrix calculated by the formula $W'(i, j) = W(i, j)/\sqrt{D(i, i)D(j, j)}$.

The primary focus of this study is the identification of circRNA-disease association pairs. Thus, we calculated the number of walks and lengths of the path from the circRNA node c(i) to the disease node d(j). $A^p(i, j)$ represents the number of walks in the path p from the c(i) node to the d(j) node, and the length of the walks in p is 1. The integration of the information about all the paths between two nodes can provide information on the potential association between nodes c(i) and d(j). In this approach, the contribution of the length of the walks to the prediction association probability is inversely proportional on every walk, that is, a shorter walking length l of path p between the nodes means a higher similarity between them. The introduction of the nonnegative coefficient sequence β_l (if the walks of length l_1 are shorter than l_2, β_{l1} is larger than β_{l2}) dampens the contributions from long walks. Therefore, the potential association between circRNA and disease can be predicted using the following formula:

$$S(c(i), d(j)) = \sum_{l=1}^{n} \beta_l A^p(i, j)$$

Because the constructed network is suitable for inclusion in the adjacency matrix, the formula was introduced into an equivalent matrix form. We further set $n \to \infty$ and thus $\beta_l \to 0$. Therefore, β_l was replaced by β^l, $A^p(i,j)$ was replaced by A^l, and the following formula was obtained:

$$S_n(A) = \sum_{l \geq 1} \beta^l A^l = (I - \beta A)^{-1} - I$$

where matrix $S_n(A)$ contains the similarity scores for all circRNAs and all diseases. A higher score indicates a stronger association between the circRNA and disease.

The integration of the adjacency matrix A_{mix} corresponding to the incorporated circRNA-disease association network G_{mix} with the circRNA similarity matrix CS and the disease similarity matrix DS yielded the matrix A*, which provides all the information for the entity. The method then ultimately obtains $S_n(A^*)$ as the prediction. The integrated matrix denotes as followed:

$$A^* = \begin{bmatrix} CS & A_{mix} \\ A_{mix}^T & SD \end{bmatrix}$$

References

1. Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Research* **40**, 3131–3142, https://doi.org/10.1093/nar/gkr1009 (2012).
2. Nigro, J. M. *et al*. Scrambled exons. *Cell* **64**, 607–613 (1991).
3. Jeck, W. R. *et al*. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157, https://doi.org/10.1261/rna.035667.112 (2013).
4. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-type specific features of circular RNA expression. *PLoS Genet* **9**, e1003777, https://doi.org/10.1371/journal.pgen.1003777 (2013).
5. Lan, P. H. *et al*. Landscape of RNAs in human lumbar disc degeneration. *Oncotarget* **7**, 63166–63176, https://doi.org/10.18632/oncotarget.11334 (2016).
6. Qian, L., Vu, M. N., Carter, M. & Wilkinson, M. F. A spliced intron accumulates as a lariat in the nucleus of T cells. *Nucleic Acids Res* **20**, 5345–5350 (1992).
7. Cocquerelle, C., Mascrez, B., Hetuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J* **7**, 155–160 (1993).
8. Kopczynski, C. C. & Muskavitch, M. A. Introns excised from the Delta primary transcript are localized near sites of Delta transcription. *J Cell Biol* **119**, 503–512 (1992).

9. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* **15**, 409, https://doi.org/10.1186/s13059-014-0409-z (2014).
10. Li, Z. *et al*. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* **22**, 256–264, https://doi.org/10.1038/nsmb.2959 (2015).
11. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733, https://doi.org/10.1371/journal.pone.0030733 (2012).
12. Memczak, S. *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338, https://doi.org/10.1038/nature11928 (2013).
13. Zhang, Y. *et al*. Circular intronic long noncoding RNAs. *Mol Cell* **51**, 792–806, https://doi.org/10.1016/j.molcel.2013.08.017 (2013).
14. Armakola, M. *et al*. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat Genet* **44**, 1302–1309, https://doi.org/10.1038/ng.2434 (2012).
15. Du, W. W. *et al*. Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death Differ* **24**, 357–370, https://doi.org/10.1038/cdd.2016.133 (2017).
16. Du, W. W. *et al*. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* **44**, 2846–2858, https://doi.org/10.1093/nar/gkw027 (2016).
17. Li, F. *et al*. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/beta-catenin pathway. *Oncotarget* **6**, 6001–6013, https://doi.org/10.18632/oncotarget.3469 (2015).
18. Ashwal-Fluss, R. *et al*. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**, 55–66, https://doi.org/10.1016/j.molcel.2014.08.019 (2014).
19. Li, P. *et al*. Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin Chim Acta* **444**, 132–136, https://doi.org/10.1016/j.cca.2015.02.018 (2015).
20. Chen, J. *et al*. Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. *Cancer Lett* **388**, 208–219, https://doi.org/10.1016/j.canlet.2016.12.006 (2017).
21. Lukiw, W. J. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front Genet* **4**, 307, https://doi.org/10.3389/fgene.2013.00307 (2013).
22. Glazar, P., Papavasileiou, P. & Rajewsky, N. circBase: a database for circular RNAs. *RNA* **20**, 1666–1670, https://doi.org/10.1261/rna.043687.113 (2014).
23. Liu, Y. C. *et al*. CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* **44**, D209–215, https://doi.org/10.1093/nar/gkv940 (2016).
24. Xia, S. *et al*. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform* **18**, 984–992, https://doi.org/10.1093/bib/bbw081 (2017).
25. Dudekula, D. B. *et al*. CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* **13**, 34–42, https://doi.org/10.1080/15476286.2015.1128065 (2016).
26. Bhattacharya, A. & Cui, Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res* **44**, D1005–1010, https://doi.org/10.1093/nar/gkv1220 (2016).
27. Ghosal, S., Das, S., Sen, R., Basak, P. & Chakrabarti, J. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet* **4**, 283, https://doi.org/10.3389/fgene.2013.00283 (2013).
28. Xia, S. *et al*. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* **46**, D925–D929, https://doi.org/10.1093/nar/gkx863 (2018).
29. Fan, C., Lei, X., Fang, Z., Jiang, Q. & Wu, F. X. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database (Oxford)* **2018**, https://doi.org/10.1093/database/bay044 (2018).
30. Zhang, J., Zhang, Z., Chen, Z. & Deng, L. Integrating Multiple Heterogeneous Networks for Novel LncRNA-disease Association Inference. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **16**, 396–406, https://doi.org/10.1109/TCBB.2017.2701379 (2019).
31. Peng, N. *et al*. Microarray profiling of circular RNAs in human papillary thyroid carcinoma. *PLoS One* **12**, e0170287, https://doi.org/10.1371/journal.pone.0170287 (2017).
32. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic acids research* **43**, D789–D798 (2014).
33. van Driel, M. A., Bruggeman, J., Vriend, G., Brunner, H. G. & Leunissen, J. A. A text-mining analysis of the human phenome. *Eur J Hum Genet* **14**, 535–542, https://doi.org/10.1038/sj.ejhg.5201585 (2006).
34. Huang, Y. F., Yeh, H. Y. & Soo, V. W. Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. *BMC Med Genomics* **6**(Suppl 3), S4, https://doi.org/10.1186/1755-8794-6-S3-S4 (2013).
35. Ding, L., Wang, M., Sun, D. & Li, A. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph. *Sci Rep* **8**, 1065, https://doi.org/10.1038/s41598-018-19357-3 (2018).
36. Xie, M., Hwang, T. H. & Kuang, R. In 2012 Pacific-Asia Conference on Knowledge Discovery and Data Mining. 292–303 (Springer).
37. Han, D. *et al*. Long noncoding RNAs: novel players in colorectal cancer. *Cancer Lett* **361**, 13–21, https://doi.org/10.1016/j.canlet.2015.03.002 (2015).
38. Xue, Y. *et al*. Genome-wide analysis of long noncoding RNA signature in human colorectal cancer. *Gene* **556**, 227–234, https://doi.org/10.1016/j.gene.2014.11.060 (2015).
39. Siegel, R. L. *et al*. Colorectal cancer statistics, 2017. *CA Cancer J Clin* **67**, 177–193, https://doi.org/10.3322/caac.21395 (2017).
40. Zhu, J. *et al*. Differential Expression of Circular RNAs in Glioblastoma Multiforme and Its Correlation with Prognosis. *Transl Oncol* **10**, 271–279, https://doi.org/10.1016/j.tranon.2016.12.006 (2017).
41. Zhu, J. *et al*. Differential expression of circular RNAs in glioblastoma multiforme and its correlation with prognosis. *Translational oncology* **10**, 271–279 (2017).
42. Lu, L. *et al*. Identification of circular RNAs as a promising new class of diagnostic biomarkers for human breast cancer. *Oncotarget* **8**, 44096–44107, https://doi.org/10.18632/oncotarget.17307 (2017).
43. Tang, Y. Y. *et al*. Circular RNA hsa_circ_0001982 Promotes Breast Cancer Cell Carcinogenesis Through Decreasing miR-143. *DNA Cell Biol* **36**, 901–908, https://doi.org/10.1089/dna.2017.3862 (2017).
44. Zhuang, Z. G. *et al*. The circular RNA of peripheral blood mononuclear cells: Hsa_circ_0005836 as a new diagnostic biomarker and therapeutic target of active pulmonary tuberculosis. *Mol Immunol* **90**, 264–272, https://doi.org/10.1016/j.molimm.2017.08.008 (2017).
45. Qian, Z. *et al*. Potential Diagnostic Power of Blood Circular RNA Expression in Active Pulmonary Tuberculosis. *EBioMedicine* **27**, 18–26, https://doi.org/10.1016/j.ebiom.2017.12.007 (2018).
46. Xiaoping Fan, Z. C. *et al* Members Aided Community Structure Detection. *Mobile Networks and Applications*, https://doi.org/10.1007/s11036-018-0994-2 (2018).
47. Zhifang, L. *et al*. A Prediction Model of the Project Life-Span in Open Source Software Ecosystem. *Mobile Networks and Applications*, https://doi.org/10.1007/s11036-018-0993-3 (2018).
48. Zhifang L. *et al*. Healthy or Not: A Way to Predict Ecosystem Health in GitHub. Symmetry **144** (2019).
49. Zhifang, L. *et al*. Identification-Method Research for Open-Source Software Ecosystems. *Symmetry* **182**, https://doi.org/10.3390/sym11020182 (2019).

50. Li, C., Zheng, X., Yang, Z., Kuang, L. J. W. C. & Computing, M. Predicting short-term electricity demand by combining the advantages of arma and xgboost in fog computing environment. **2018** (2018).
51. Kuang, L. *et al*. A personalized qos prediction approach for cps service recommendation based on reputation and location-aware collaborative filtering. **18**, 1556 (2018).
52. Kuang, L. *et al*. A Privacy Protection Model of Data Publication Based on Game Theory. **2018** (2018).
53. Zhu, Y., Yan, X., Li, S., Fan, Y. & Kuang, L. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 1112–1117 (IEEE).
54. Zheng, N., Wang, K., Zhan, W. & Deng, L. Targeting Virus-host Protein Interactions: Feature Extraction and Machine Learning Approaches. *Current drug metabolism* **20**, 177–184 (2019).
55. Zhang, J., Zhang, Z., Wang, Z., Liu, Y. & Deng, L. Ontological function annotation of long non-coding RNAs through hierarchical multi-label classification. *Bioinformatics* **34**, 1750–1757 (2018).
56. Nie, L., Deng, L., Fan, C., Zhan, W. & Tang, Y. Prediction of protein S-sulfenylation sites using a deep belief network. *Current Bioinformatics* **13**, 461–467 (2018).
57. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43, https://doi.org/10.1007/BF02289026 (1953).
58. Chen, X., Huang, Y. A., You, Z. H., Yan, G. Y. & Wang, X. S. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* **34**, 1440, https://doi.org/10.1093/bioinformatics/btx773 (2018).
59. Yang, X. *et al*. A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* **9**, e87797, https://doi.org/10.1371/journal.pone.0087797 (2014).
60. Qu, Y., Zhang, H., Liang, C. & Dong, X. KATZMDA: Prediction of miRNA-disease associations based on KATZ model. *IEEE Access* **PP**, 1–1, https://doi.org/10.1109/ACCESS.2017.2754409 (2017).
61. Chen, X. KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci Rep* **5**, 16840, https://doi.org/10.1038/srep16840 (2015).
62. Zhang, Z., Zhang, J., Fan, C., Tang, Y. & Deng, L. KATZLGO: large-scale prediction of LncRNA functions by using the KATZ measure based on multiple networks. *IEEE/ACM transactions on computational biology and bioinformatics* **16**, 407–416 (2019).

Acknowledgements

This work was funded by National Natural Science Foundation of China under Grant Number 61672541 and Natural Science Foundation of Hunan Province under Grant Number 2017JJ3412.

Author Contributions

L.D., W.Z. and Y.T. conceived this work and designed the experiments. W.Z. carried out the experiments. L.D., W.Z., Y.S. and Y.T. collected the data and analyzed the results. L.D., W.Z., Y.S. and Y.T. wrote, revised, and approved the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.