BMC Bioinformatics

METHODOLOGY

Open Access

MSCFS: inferring circRNA functional similarity based on multiple data sources

Liang Shu[1], Cheng Zhou[1], Xinxu Yuan[2], Jingpu Zhang[3] and Lei Deng[1*]

From The 19th Asia Pacific Bioinformatics Conference (APBC 2021) Tainan, Taiwan. 3-5 February 2021

*Correspondence:
leideng@csu.edu.cn
[1] School of Computer Science and Engineering, Central South University, Lushangnan Road, Changsha, China
Full list of author information is available at the end of the article

Abstract

Background: More and more evidence shows that circRNA plays an important role in various biological processes and human health. Therefore, inferring the circRNA's potential functions and obtaining circRNA functional similarity has become more and more significant. However, there is no effective approach to explore the functional similarity of circRNAs.

Methods: In this paper, we propose a new approach, called MSCFS, to calculate the functional similarity of circRNA by integrating multiple data sources. We combine circRNA-disease association, circRNA-gene-Gene Ontology association, and circRNA sequence information to explore the functional similarity of circRNA. Firstly, we employ different learning representation methods from three data sources to establish three circRNA functional similarity networks. Then we integrate the three networks to obtain the final circRNA functional similarity.

Results: We utilize circRNA–miRNA association similarity and circRNA co-expression similarity to evaluate the performance of MSCFS. The results show a positive correlation with miRNA association ($R = 0.213$) and circRNA co-expression similarity ($R = 0.8991$). Finally, we construct a circRNA functional similarity network and perform case analysis. The result shows our method can be applied to infer new potential functions of circRNA and other associations.

Conclusions: MSCFS combines multiple data sources related to circRNA functions. Correlation analysis and case analyses prove that MSCFS is a useful method to explore circRNA functional similarity.

Keywords: CircRNA functional similarity, Multiple data sources, Multiple representations

Background

Circular RNAs, a class of endogenous non-coding RNAs, are characterized by their covalently closed-loop structures without a 5′ cap or a 3′ Poly A tail [1]. Sanger et al. [2] first found CircRNAs in 1976. However, the circRNAs were thought to be splicing artifact; and were continuously considered as "junk" RNAs for about two decades [3]. More

and more researches have corroborated that circRNAs play an essential role in many cell activities, affecting arteriosclerosis and participating in mRNA expression variable splicing regulation [4–9]. In recent years, circRNAs have been identified as biomarkers and therapeutic targets for various acute diseases. CircRNA is associated with a variety of chronic diseases, such as lung cancer, Alzheimer's disease, diabetes, cardiovascular disease and other [10, 11]. The emerging experimental results have corroborated that circRNA molecules have abundant miRNA binding sites and act as miRNA sponges in cells to releasing the inhibitory effect of miRNA on their target genes and improving the expression level of target genes [12–14]. Identifying the targets of circRNAs helps to understand the functions of circRNAs. Several efforts have been developed to identify circRNA targets [15–17]. For example, Lin et al. [17] designed Analysis of common targets (ACT) to facilitate the identification of potential circRNA targets.

Functional similarity can be defined as an association, such as co-expression similarity, co-Gene Ontology (GO) term similarity, co-similar disease similarity, and co-literature similarity [18]. Analogous to the methods of studying the functional similarity of microRNA, Wang et al. [19] obtained the functional similarity of miRNA through the DAG of disease and microRNA-disease association. The Gene Ontology (GO) project provides the most comprehensive resource currently available for computable knowledge regarding the functions of genes and gene products. Gene Ontology provides the logical structure of the biological functions ('terms') and their relationships to one another, manifested as a directed acyclic graph [20]. Yang et al. [21] obtained the functional similarity of microRNA by calculating GO semantic similarity and miRNA-GO association. There are many methods to acquire the semantic similarity of GO, such as the measures proposed by Resnik et al. [22], Jiang et al. [23], Lin et al. [24], Wang et al. [25], and Wu et al. [26]. Obtaining the functional similarity of RNA can also be obtained through sequence information. Sequence similarity can be calculated by methods such as K-mer [27] or LSTM [28].

However, there is no valid method to calculate the functional similarity of circRNA, and a single circRNA data source can't effectively explore the circRNA functional similarity. In this paper, we propose a novel method called MSCFS by integrating multiple biological data sources to calculate the functional similarity between circRNAs. Firstly, we obtain the circRNA functional similarity matrix by using the DAG graph and association information of the disease. Secondly, we construct the corpus through circRNA-gene-GO associations and GO annotations and employ word2vec to obtain the circRNA functional similarity matrix. Thirdly, we adopt chaos game representation to get circRNA functional similarity by circRNA sequence information. Finally, the circRNA functional similarity is obtained by integrating the three networks. The results show that MSCFS is efficacious and accurate, and it can infer the potential functions of circRNA. The flowchart of our proposed model is shown in Fig. 1.

Methods

Dataset

We downloaded the MeSH descriptor from the National Library of Medicine (http://www.nlm.nih.gov/) [29]. MeSH descriptors are divided into 16 categories: category A is anatomical terms, category B is organisms, category C is diseases, category D

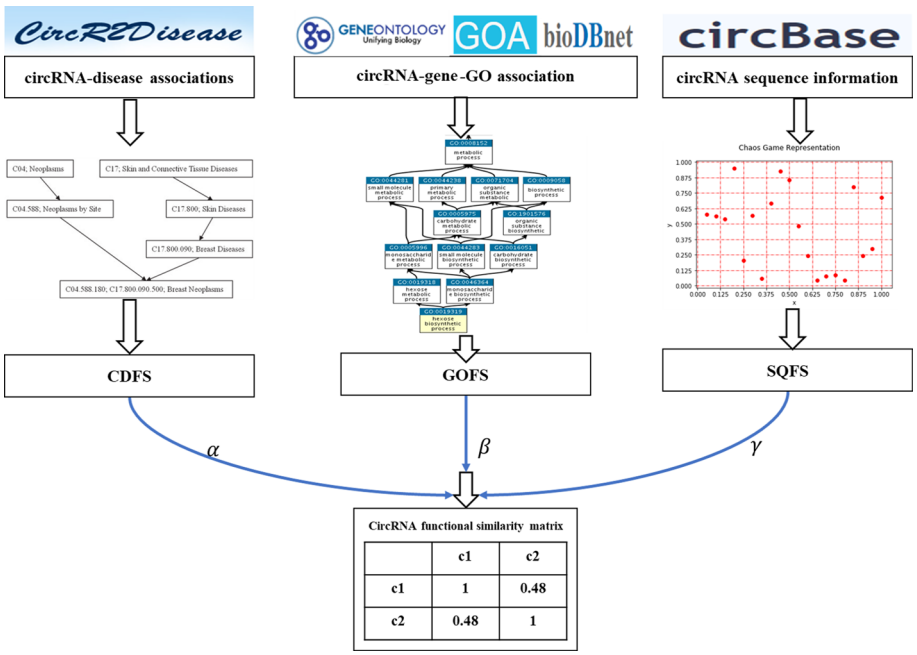

Fig. 1 Overview of MSCFS demonstrating the basic ideas of measuring circRNAs functional similarity

is drugs and chemicals, etc. Then, we obtained the relationship of various diseases based on DAG diseases from the MeSH descriptor of category C.

Many benchmark databases contain circRNA-disease association data, such as circR2Disease [30], circRNADisease [31], circFunBase [32], and Circ2Disease [33], which contain experimentally verified associations between circRNAs and diseases. We utilize circR2Disease as the benchmark data set. Circ2Disease is a database that can manually manage human circRNA supported by experiments and provide the association between circRNA and human diseases. We obtained 418 confirmed circRNA-disease associations consisting of 365 circRNA and 71 diseases after removing the circRNAs in which the gene symbol could not be found.

We downloaded the Gene Ontology (GO) in OWL format from the Gene Ontology Consortium (GOC) [34] and GO annotations in the Gene Ontology Annotation (GOA) Database [35]. We used the OWL API version 4.2.6 to process the GO in OWL format.

We extracted 321 genes associated with circRNA and the circRNA sequence information from the circBase [36]. We obtained 7321 GO-gene associations from multiple versions of the database bioDBnet [37].

Overview of MSCFS

In this article, we combine the three data sources of circRNA to calculate the functional similarity of circRNA. Specifically, we obtain three circRNA functional similarity matrices from the circRNA-disease association, circRNA-gene-GO association, and circRNA sequence information. Finally, we integrate three networks to obtain the final circRNAs functional similarity, and the formula is as follows:

$$\begin{cases} \mathrm{FS} = \alpha * \mathrm{CDFS} + \beta * \mathrm{GOFS} + \gamma * \mathrm{SQFS} \\ \alpha + \beta + \gamma = 1 \end{cases} \tag{1}$$

where CDFS, GOFS, SQFS are circRNA functional similarity matrices obtained through circRNA-disease association, circRNA-gene-GO association, and circRNA sequence information, respectively. α, β, and γ are the weighting coefficients of the three networks severally.

Functional similarity based on circRNA-disease association

Genes with similar functions are known to be associated with similar diseases. A structure of a directed acyclic graph (DAG) can represent the relationship between different diseases. Therefore, we can calculate the functional similarity of circRNA through circRNA-disease association. The process is shown in Fig. 2.

In the MeSH database, the relationship between diseases is described in the form of a directed acyclic graph (*DAG*), where nodes represent diseases and edges represent relationships between diseases. Given a disease D, we have defined a DAG graph $DAG_D = (D, T_a, E_a)$ based on the other diseases it is associated with and related edges, where T_a is the set of ancestor nodes containing itself, and E_a is the set of corresponding edges connecting these diseases. If disease d is in the *DAG*, its contribution to disease A can be calculated as follows:

$$\begin{cases} D_D(D) = 1 \\ D_D(d) = \max\left\{\Delta * D_A\left(d'\right) \mid d' \in \textit{children of } d\right\} \textit{ if } d \neq A \end{cases} \tag{2}$$

where Δ is the semantic contribution factor of disease d and its child nodes. In DAG, the semantic value of disease D itself is defined as 1. Therefore, through the following formula, we calculate the semantic value $DV(D)$ of disease D:

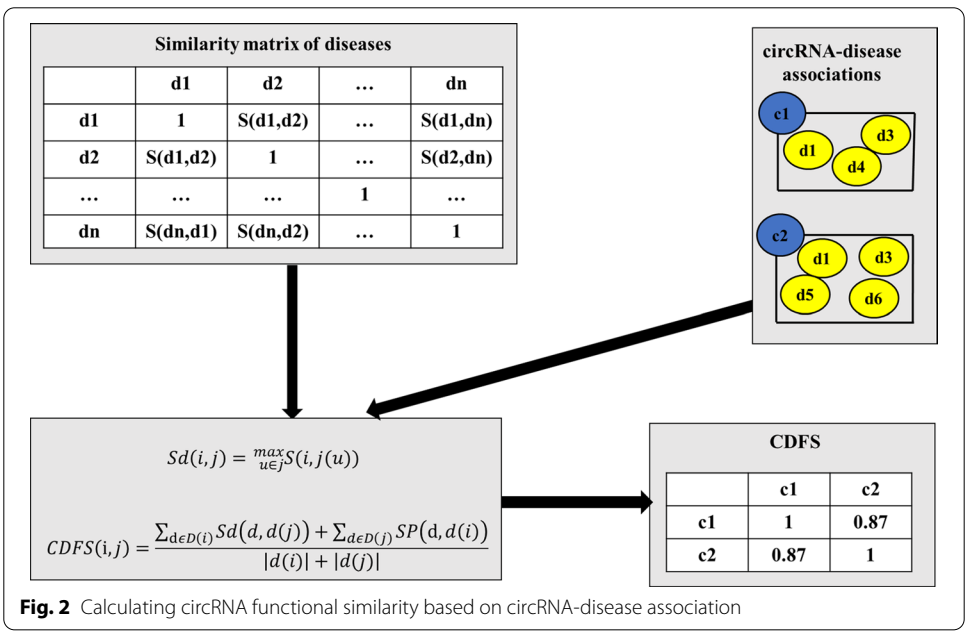

Fig. 2 Calculating circRNA functional similarity based on circRNA-disease association

$$\mathrm{DV(D)} = \sum_{d \in T_A} D_A(d). \tag{3}$$

Here, we assume that the more *DAG* shared parts of the two diseases, the higher the semantic similarity, so according to the position of the two diseases in the *DAG* graph and the semantic relationship with the ancestral diseases, the formula for calculating the semantic similarity of the two diseases M and N is as follows:

$$S(M,N) = \frac{\sum_{d \in T_{M \cap T_N}} (D_M(d) + D_N(d))}{DV(M) + DV(N)}, \tag{4}$$

where $D_M(d)$ is the semantic value of disease d related to disease M, and $D_N(d)$ is the semantic value of disease d related to disease N.

$$Sd(d, DT) = \max_{u \in DT} S(d, DT(u)), \tag{5}$$

where DT is a group of diseases, u is any disease in DT. After obtaining the semantic similarity of disease combinations, we use circRNA-disease correlation to obtain the functional similarity of circRNA, CDFS.

$$CDFS(i,j) = \frac{\sum_{d \in D(i)} Sd(d, d(j)) + \sum_{d \in D(j)} Sd(d, d(i))}{|D(i)| + |D(j)|}, \tag{6}$$

where $CDFS(i, j)$ is the similarity between the ith circRNA and the jth circRNA, $D(i)$ is the ith circRNA associated disease set.

Functional similarity based on circRNA-gene-GO association

Onto2Vec is a measure that combines formal ontology axioms and annotation axioms in ontology metadata to generate a vector representation of biological entities in the ontology [38]. Gene Ontology contains the representation of the essence of the knowledge system in the field of biology. Ontologies are usually composed of a set of categories (or terms or concepts) with relationships between them. In order to explore the functional similarity of circRNA, we use the circRNA-gene-GO association. We add circRNA as new entities and apply the *has−function* relationship to connect them with their functions to generate a corpus. Then we use Onto2Vec to generate a vector representation for each class (using a corpus to only be based on axioms), and further, generate a joint representation of circRNA and classes (using a corpus-based axiom and circRNA and its annotations).

In the end, we constructed 230,699 corpus, with 50,409 categories, using the Skip-gram model in word2vec. Word2Vec is a set of neural network-based tools that can generate vector representations of words from a large corpus. There are two models: the continuous bag of word (CBOW), which uses a context to predict a target word, and the Skip-gram model that tries to maximize the classification of a word based on another word from the same sentence. Figure 3 shows the flow of this section.

The Skip-gram model is chosen because the Skip-gram model generates higher quality rare word representations in the corpus. The Skip-gram model learns more detailed word vectors and has a large number of low-frequency words in the corpus

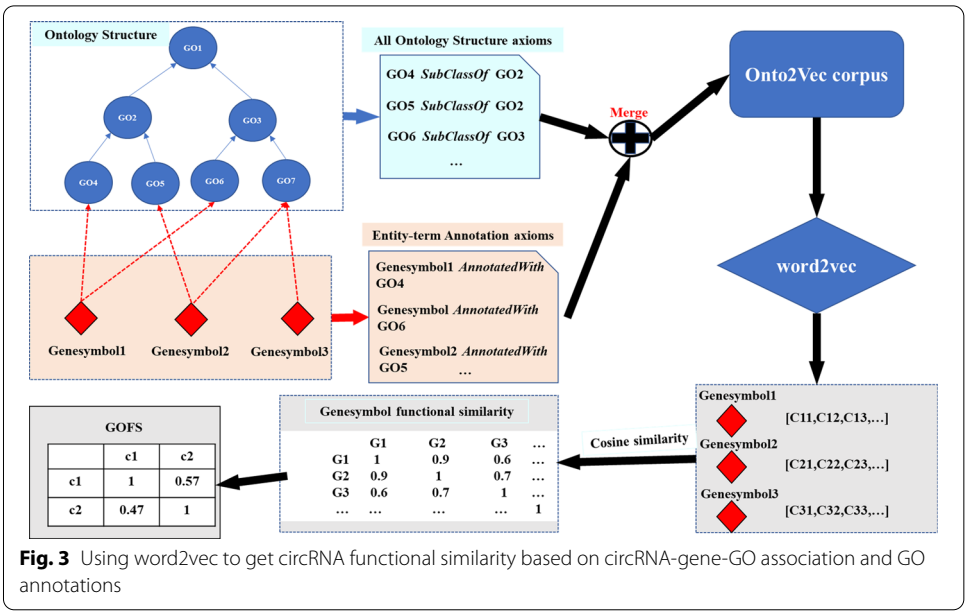

Fig. 3 Using word2vec to get circRNA functional similarity based on circRNA-gene-GO association and GO annotations

Table 1 Parameters of Word2Vec model

Parameter	Definition	Default value
sg	Choice of training algorithm (sg = 1:skip gram; sg = 0:CBOW)	1
min_count	Words with frequency lower than this value will be ignored	1
size	Dimension of the obtained vectors	200
window	Maximum distance between the current and the predicted word	10
iter	Number of iterations	5
negative	Whether negative sampling will be used and how many 'noise words' would be drawn	4

to produce high-quality representations of all biological entities occurring in our large corpus, including uncommon ones. Given a set of training word sequences $w_1, w_2 \ldots, w_N$, Skip-gram the goal is to maximize the following average logarithmic likelihood values:

$$\frac{1}{N}\sum_{t=1}^{N}\sum_{-s\le i\le s, j\neq 0}\log p\left(\omega_{t+j} \mid \omega_t\right), \tag{7}$$

where s means the size of the training context, N means the size of the set of the training words, and w_i is the ith training word in the sequence. In our research, the parameters of word2vec used are shown in Table 1.

Through the training, we get the similarity of the genes, and then through the circRNA-gene relationship GOFS, the calculation formula is as follows:

$$\mathrm{GOFS}(i,j) = \frac{\sum_{g\in G(i)} S(g, g(j)) + \sum_{g\in G(j)} S(g, g(i))}{|g(i)| + |g(j)|}, \tag{8}$$

where $g(i)$ is the set of genes associated with the ith circRNA, and $g(j)$ is the set of genes associated with the jth circRNA.

Functional similarity based on circRNA sequence

Different from the K-mer [27], PSSM method [39], chaos game representation [40] combines position information and nonlinear relationship to obtain vector representation of sequences. Finally, the Pearson correlation is used to quantify their correlation. The advantage of the algorithm is that the original information of the sequence is completely restored in the coordinate system, and the information will not be lost in the mapping. Secondly, the position information will be retained as a mapping. Figure 4 shows the workflow of this section.

The position of each nucleotide in the plane:

$$P_i = 0.5 * (P_{i-1} + S_i) \quad i = 1 \ldots N, \tag{9}$$

where P_0 is any given starting point ($P_0 = (0.5, 0.5)$), N represents the length of the sequence. S_i represents the ith nucleotide in the sequence, which corresponds to the fixed vertex coordinates of $A = (0, 0), C = (1, 0), G = (1, 1)$ and $U = (0, 1)$ respectively.

In this way, the CGR graph is transformed into a N_g grid $\left(N_g = 2^s \times 2^s, \mathrm{s} = 3\right)$ digital matrix, which is called the frequency matrix of CGR graph (FCGR). And grid can be represented as follows:

$$grid_i = (X_i, Y_i, Z_i) \tag{10}$$

We use the x-axis, y-axis direction and their digital features to construct the feature vector of the sequence, the calculation formula is as follows: the abscissa point.x

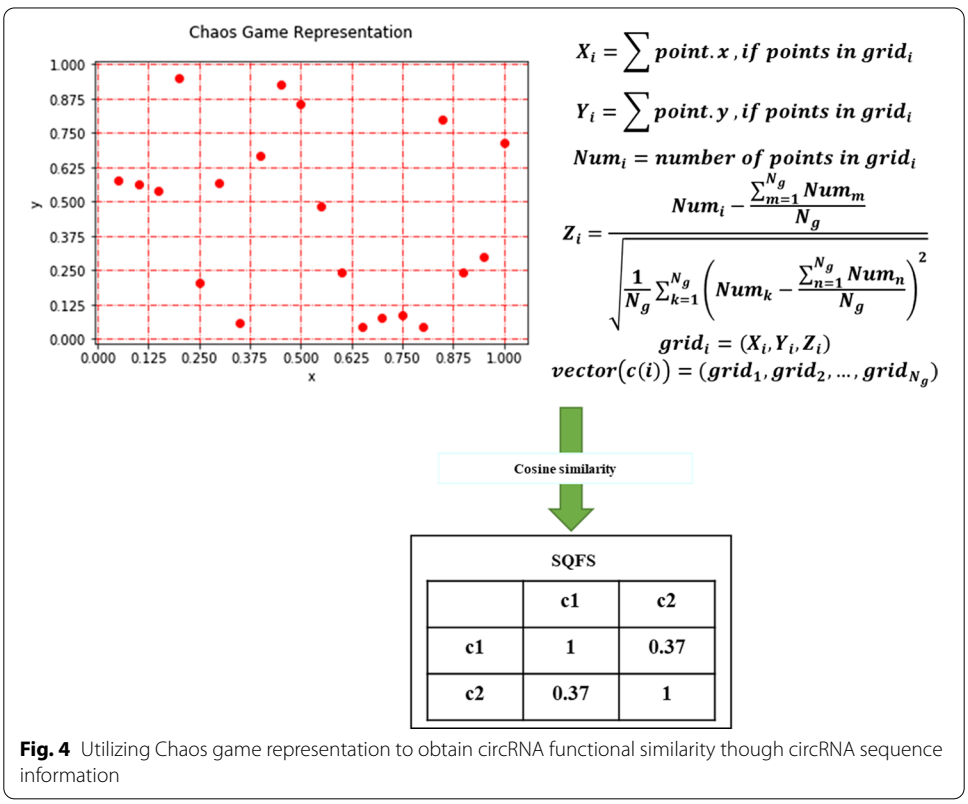

Fig. 4 Utilizing Chaos game representation to obtain circRNA functional similarity though circRNA sequence information

and ordinate point.y in each grid are accumulated respectively to quantify position information.

$$\begin{cases} X_i = \sum point.x & if\ points\ in\ grid_i \\ Y_i = \sum point.y & if\ points\ in\ grid_i \end{cases}. \tag{11}$$

Then, we obtain the z-scores of each grid Z_i to quantify potential features.

$$\begin{cases} Z_i = \dfrac{Num_i - \frac{\sum_{m=1}^{Ng} Num_m}{Ng}}{\sqrt{\frac{1}{Ng}\sum_{k=1}^{Ng}\left(Num_k - \frac{\sum_{n=1}^{Ng} Num_n}{Ng}\right)^2}} \\ Num_i = number\ of\ points\ in\ grid_i \end{cases} \tag{12}$$

Finally, each grid can be represented as three attributes, and we fused the attributes to construct the vectors $vector(c(i))$ to define the sequence functional similarity of circRNAs $SQFS(c(i), c(j))$ by Pearson correlation coefficient. Where $c(i)$ represents the A ith cricRNA.

$$\begin{cases} SQFS(c(i), c(j)) = \frac{vector(c(i)) \cdot vector(c(j))}{\|vector(c(i))\| \|vector(c(j))\|} \\ vevtor(c(i)) = \left(grid_1, grid_2, \ldots, grid_{\mathrm{N_g}}\right) \end{cases}, \tag{13}$$

where $vector(c(i))$ means the sequence feature vector of the ith circRNA, $vector(c(i)) \cdot vector(c(i))$ is the dot product of $vector(c(i))$ and $vector(c(i))$.

Results

Parameters selection

circRNA can be used as the sponge of microRNA to play a role in biological processes [12]. We downloaded circRNA–miRNA association data from the starbase [41]. There are 267 types of circRNAs that match the 365 types of circRNA we calculated. We obtained 267 pairs of circRNA–miRNA association similarities through the Jaccard similarity method, which were compared with the functional similarities we calculated. Denote CMS as the circRNA–miRNA similarity matrix, and its entry $CMS(i, j)$ can be obtained by the following formula:

$$CMS(i,j) = \frac{\left|CM_i \cap CM_j\right|}{\left|CM_i \cup CM_j\right|}, \tag{14}$$

where CM_i is the set of microRNAs associated with the ith circRNA, and CM_j is the set of microRNAs associated with the jth circRNA.

We set the parameter step size to 0.1. Because circRNA has little data associated with diseases, we set α the value range from 0 to 0.2, and the value range for β and γ from 0 to 1.We used the grid search method to obtain the optimal parameters through 30 sets of experiments and selected two groups for display. The results are shown in Fig. 5. The experimental results with parameters of 0.1, 0.4, 0.5 and 0.1, 0.3, 0.6 are $(R = 0.205, P = 8.2e^{-3})$, $(R = 0.213, P = 4.6e^{-4})$. The results prove that our circRNA functional similarity is related to microRNA similarity. We selected the optimal value of

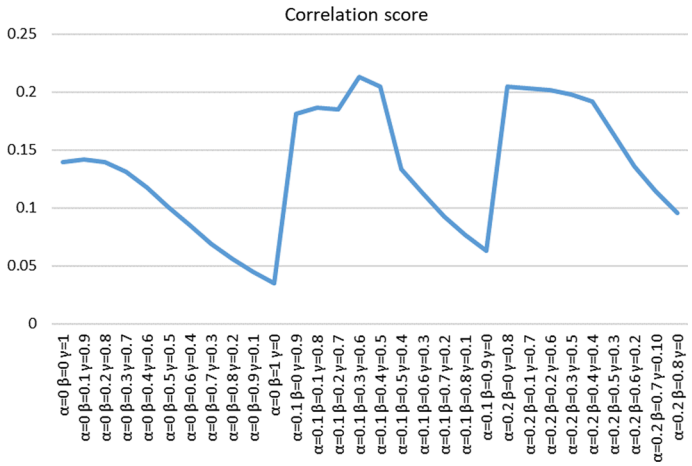

Fig. 5 Parameters selection

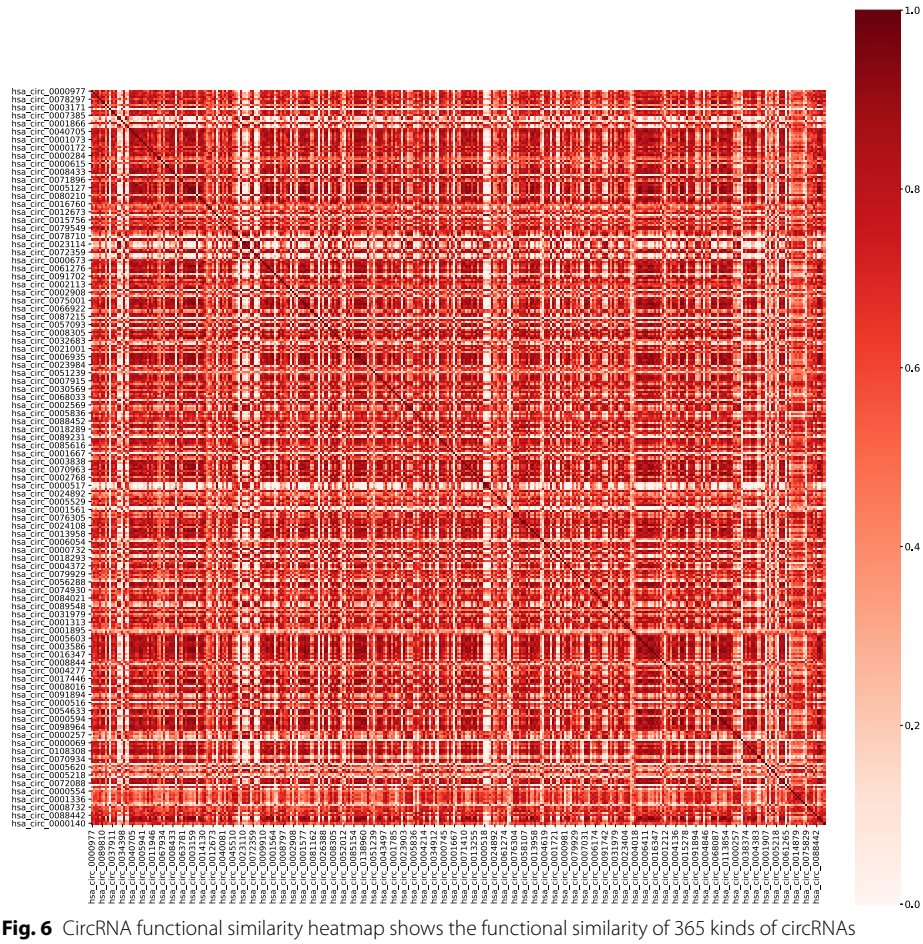

Fig. 6 CircRNA functional similarity heatmap shows the functional similarity of 365 kinds of circRNAs

the parameter α, β, γ are 0.1, 0.3, 0.6. Figure 6 shows the heat map of circRNA functional similarity calculated with parameters of 0.1, 0.3 and 0.6.

CircRNA functional similarity is correlated with expression similarity

CircRNAs with semblable functions squint towards participate in semblable biological processes and interact with semblable cellular components. To verify circRNAs with similar functions may squint towards having similar expression profiles, we seeked the relationship between circRNA functional similarity calculated by MSCFS and expression similarity. In this study, we used the absolute Pearson's correlation coefficient (PCC) to measure circRNA expression similarity.

We finally obtain circRNA expression profiling data from Peng et al.'s work [42], which consists of expression profiles of 2932 circRNAs. Then, we calculated the conducted a comprehensive analysis of circRNA expression in papillary thyroid carcinoma PCC score as the co-expression similarity of each pair of circRNA expression profiles and obtained the co-expression similarity of 8049 pairs of circRNAs, and then performed the correlation analysis of circRNA expression similarity and circRNA functional similarity. As a result, the functional similarity of circRNA confirmed positive correlation with circRNA co-expression similarity ($R = 0.076, P = 9.05e^{-4}$, Pearson correlation). We grouped 8049 pairs of circRNAs into different groups according to functional similarity in steps of 0.1 and calculated the average expression similarity and functional similarity of each group. Clearly, the functional similarity of circRNA is positively correlated with the expression similarity ($R = 0.8991, P = 9.73e^{-4}$, Fig. 7). Results inform that circRNA functional similarity obtained by our method is correlated with circRNA expression similarity, which is well known to be associated with circRNA functional similarity.

A circRNA functional similarity network

Figure 8 shows the distribution of circRNA functional similarity scores. We have structured a partial graph of the circRNA network with a threshold of 0.7 (Fig. 9). Some circRNAs are less associated with other circRNAs, while some circRNAs are more associated with other circRNAs. We can do more research on those circRNAs that are more related and explore the potential functions of circRNAs related to them.

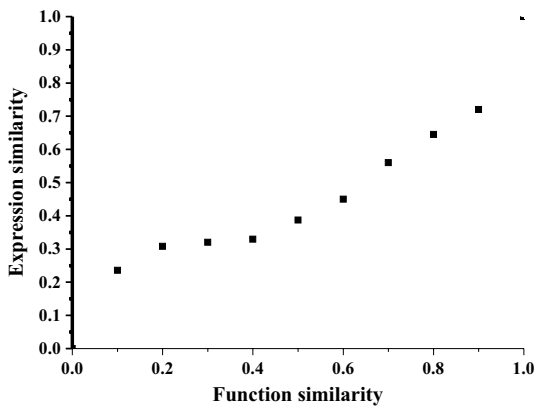

Fig. 7 The relationship between circRNA MSCFS functional similarity and their expression similarity

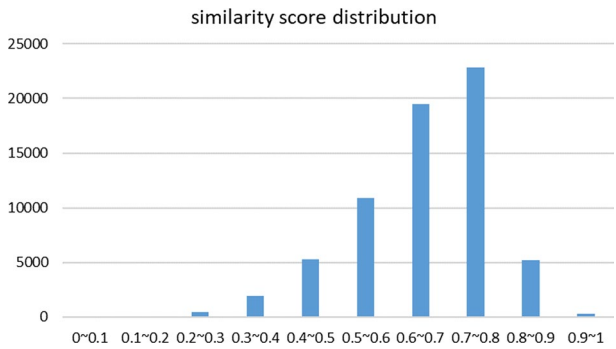

Fig. 8 The circRNA funtional similarity score distribution

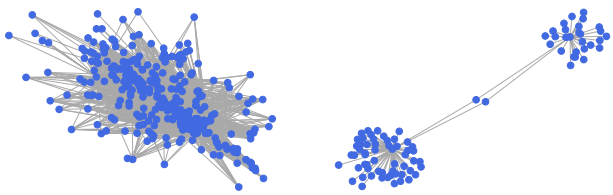

Fig. 9 A circRNA functional network based on the functional similarity of MSCFS. Each node and edge connecting any two nodes (circRNAs) indicates that the functional similarity of the two circRNAs is equal to or greater than the similarity cutoff value

Case study

To verify our results, we executed case analysis on the circRNA function annotation in the CircFunBase database. CircFunBase is a web-accessible database that aims to provide a high-quality functional circRNA resource, including experimentally validated and computationally predicted functions [32].

Amongst the 365 kinds of circRNAs, hsa_circ_0000140 has the highest correlation score with hsa_circ_0001946 in other 364 circRNAs. In the CircFunBase database, the functional annotations of these two circRNAs are related to gastric cancer. Then we select circRNA pairs with high similarity scores for analysis. For example, the functional similarity score of hsa_circ_0043278 and hsa_circ_0006220 is 0.82, and it can be found in the CircFunBase database that both are related to hypertension [differential expression (hypertensive patients and healthy controls)]. The functional similarity score of hsa_circ_0005927 and hsa_circ_0138960 is 0.83, both of which are related to gastric cancer. Through case analysis, we can know the practicality and accuracy of the MSCFS method.

Discussion and conclusion

CircRNAs have peculiar biological structures and have proven to play essential roles in biological processes and human health. Inferring the functional similarity of circRNAs can help analyze the function of circRNAs and predict the association of circRNA-disease. However, due to the lack of functional annotations for circRNAs in public databases, it is not straightforward to calculate the functional similarity of circRNAs using existing single data sources.

This paper proposes a new algorithm, MSCFS, to calculate the functional similarity of circRNA by integrating multiple circRNA associated biological data. The results showed that the circRNAs associated with the same miRNA have a high similarity score. The circRNA co-expression similarity was positively correlated with our calculated results. We also found that circRNAs with a high similarity score were also similar in the function of the disease. By calculating the functional similarity of circRNAs, we can explore more potential functions and associations of circRNAs.

We have integrated multiple biological data sources related to circRNAs, and the results will be somewhat biased due to the quality of some data sources. In the future, we will integrate more and more reliable data to further improve the accuracy of circRNA functional similarity calculations.

Abbreviations
circRNAs: Circular RNAs; GO: Gene Ontology; RBPs: RNA binding proteins; DAG: Directed acyclic graphs; PSSM: Position specific scoring matrix; CMS: CircRNA–miRNA similarity matrix; PCC: Pearson's correlation coefficient.

Acknowledgements
The authors are very grateful to the anonymous reviewers for their constructive comments which have helped significantly in revising this work. We would like to thank the Experimental Center of School of Computer Science and Engineering of Central South University, for providing computing resources.

About this supplement
This article has been published as part of BMC Bioinformatics Volume 22 Supplement 10 2021: Selected articles from the 19th Asia Pacific Bioinformatics Conference (APBC 2021): bioinformatics. The full contents of the supplement are available at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-22-supplement-10.

Authors' contributions
LS, CZ, XY, LD and JP designed the study and conducted experiments. LS, CZ, LD and JP performed statistical analyses. LS, XY and JP drafted the manuscript. JP and LD prepared the experimental materials and benchmarks. All authors have read and approved the final manuscript.

Funding
This work was supported by National Natural Science Foundation of China under Grant No. 61972422. The funding body has not played any roles in the design of the study and collection, analysis and interpretation of data in writing the manuscript.

Availability of data and materials
The Python source codes and the datasets in this work are freely available in the GitHub (https://github.com/CJNabla/MultiSourcCFS).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Computer Science and Engineering, Central South University, Lushangnan Road, Changsha, China. [2]Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA 23284, USA. [3]School of Computer and Data Science, Henan University of Urban Construction, Longxiang Road, Pingdingshan 467000, China.

Received: 20 June 2021 Accepted: 6 July 2021
Published online: 16 July 2021

References

1. Meng S, Zhou H, Feng Z, Xu Z, Tang Y, Li P, Wu M. CircRNA: functions and properties of a novel potential biomarker for cancer. Mol Cancer. 2017;16(1):1–8.
2. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proc Natl Acad Sci. 1976;73(11):3852–6.
3. Zhang Z, Yang T, Xiao J. Circular RNAs: promising biomarkers for human diseases. EBioMedicine. 2018;34:267–74.
4. Du WW, Fang L, Yang W, Wu N, Awan FM, Yang Z, Yang BB. Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. Cell Death Differ. 2017;24(2):357–70.
5. Armakola M, Higgins MJ, Figley MD, Barmada SJ, Scarborough EA, Diaz Z, Fang X, Shorter J, Krogan NJ, Finkbeiner S, et al. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nat Genet. 2012;44(12):1302.
6. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, et al. Exon-intron circular RNAs regulate transcription in the nucleus. Nat Struct Mol Biol. 2015;22(3):256.
7. Zhang Y, Zhang X-O, Chen T, Xiang J-F, Yin Q-F, Xing Y-H, Zhu S, Yang L, Chen L-L. Circular intronic long noncoding RNAs. Mol Cell. 2013;51(6):792–806.
8. Xu H, Guo S, Li W, Yu P. The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells. Sci Rep. 2015;5(1):1–12.
9. Li F, Zhang L, Li W, Deng J, Zheng J, An M, Lu J, Zhou Y. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the WNT/β-catenin pathway. Oncotarget. 2015;6(8):6001.
10. Lukiw W. Circular RNA (circRNA) in Alzheimer's disease (AD). Front Genet. 2013;4:307.
11. Greene J, Baird A-M, Brady L, Lim M, Gray SG, McDermott R, Finn SP. Circular RNAs: biogenesis, function and role in human diseases. Front Mol Biosci. 2017;4:38.
12. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. Nature. 2013;495(7441):384–8.
13. Thomas LF, Sætrom P. Circular RNAs are depleted of polymorphisms at microRNA binding sites. Bioinformatics. 2014;30(16):2243–6.
14. Kulcheski FR, Christoff AP, Margis R. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. J Biotechnol. 2016;238:42–51.
15. Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol. 2016;13(1):34–42.
16. Dori M, Bicciato S. Integration of bioinformatic predictions and experimental data to identify circRNA–miRNA associations. Genes. 2019;10(9):642.
17. Lin Y-C, Lee Y-C, Chang K-L, Hsiao K-Y. Analysis of common targets for circular RNAs. BMC Bioinform. 2019;20(1):372.
18. Li J, Zhang S, Wan Y, Zhao Y, Shi J, Zhou Y, Cui Q. MISIM v2.0: a web server for inferring microRNA functional similarity based on microRNA-disease associations. Nucleic Acids Res. 2019;47(W1):536–41.
19. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26(13):1644–50.
20. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
21. Yang Y, Fu X, Qu W, Xiao Y, Shen H-B. MiRGOFS: a GO-based functional similarity measurement for miRNAs, with applications to the prediction of miRNA subcellular localization and miRNA-disease association. Bioinformatics. 2018;34(20):3547–56.
22. Resnik P. Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. J Artif Intell Res. 1999;11:95–130.
23. Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint arXiv:cmp-lg/9709008. 1997.
24. Lin D, et al. An information-theoretic definition of similarity. Icml. 1998;98:296–304.
25. Wang JZ, Du Z, Payattakool R, Yu PS, Chen C-F. A new method to measure the semantic similarity of GO terms. Bioinformatics. 2007;23(10):1274–81.
26. Wu H, Su Z, Mao F, Olman V, Xu Y. Prediction of functional modules based on comparative genome analysis and gene ontology application. Nucleic Acids Res. 2005;33(9):2822–37.
27. Fletez-Brant C, Lee D, McCallion AS, Beer MA. kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets. Nucleic Acids Res. 2013;41(W1):544–56.
28. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
29. Lord PW, Stevens RD, Brass A, Goble CA. Investigating semantic similarity measures across the gene ontology: the relationship between sequence and annotation. Bioinformatics. 2003;19(10):1275–83.
30. 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. 2018;2018:bay044.
31. Zhao Z, Wang K, Wu F, Wang W, Zhang K, Hu H, Liu Y, Jiang T. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. Cell Death Dis. 2018;9(5):1–2.
32. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. Database. 2019;2019:baz003.
33. Yao D, Zhang L, Zheng M, Sun X, Lu Y, Liu P. Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. Sci Rep. 2018;8(1):1–6.
34. Consortium GO. Gene ontology consortium: going forward. Nucleic Acids Res. 2015;43(D1):1049–56.
35. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R. The gene ontology annotation (GOA) database: sharing knowledge in uniprot with gene ontology. Nucleic Acids Res. 2004;32(suppl–1):262–6.
36. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA. 2014;20(11):1666–70.

37. Mudunuri U, Che A, Yi M, Stephens RM. bioDBnet: the biological database network. Bioinformatics. 2009;25(4):555–6.
38. Smaili FZ, Gao X, Hoehndorf R. Onto2vec: joint vector-based representation of biological entities and their ontology-based annotations. Bioinformatics. 2018;34(13):52–60.
39. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mole Biol. 2000;299(2):501–22.
40. Jeffrey HJ. Chaos game representation of gene structure. Nucleic Acids Res. 1990;18(8):2163–70.
41. 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. 2014;42(D1):92–7.
42. Peng N, Shi L, Zhang Q, Hu Y, Wang N, Ye H. Microarray profiling of circular RNAs in human papillary thyroid carcinoma. PLoS One. 2017;12(3):0170287.

Publisher's Note