Detecting Composite Functional Module in miRNA Regulation and mRNA Interaction Network

Sakshi Sharma[1], Sangmeshwar Yewange[2], Ms. Shweta Tyagi[3], Dr.Abhimanyu Kumar Jha[1,4*]
[1,3,4] *Department of Biotechnology, Faculty of Life Science, Institute of Applied Medicines and Research, Ghaziabad, Uttar Pradesh, India*
[2]*Latur College of Pharmacy Hasegaon, Tq. Ausa, Dist. Latur- 413512, Maharashtra, India*
[4]*Dr.A.P.J Abdul Kalam Technical University, Lucknow,Uttar Pradesh, India*

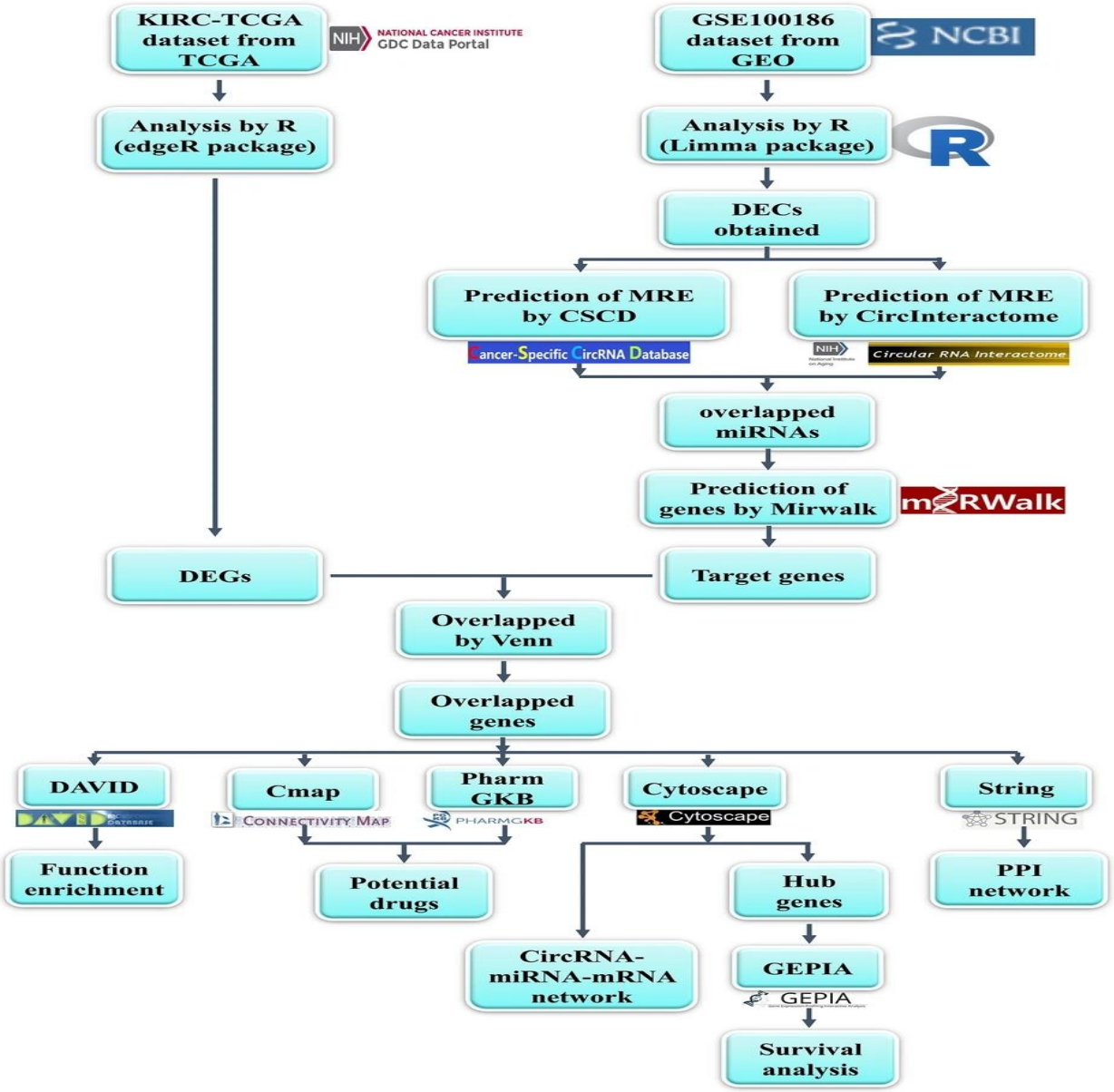
Abstract **- MicroRNAs (miRNAs) play an essential role in the post-transcriptional gene regulation in plants and animals. They regulate a wide range of biological processes by targeting messenger RNAs (mRNAs). Evidence suggests that miRNAs and mRNAs interact collectively in gene regulatory networks. The collective relationships between groups of miRNAs and groups of mRNAs may be more readily interpreted than those between individual miRNAs and mRNAs, and thus are useful for gaining insight into gene regulation and cell functions. Several computational approaches have been developed to discover miRNA-mRNA regulatory modules (MMRMs) with a common aim to elucidate miRNA-mRNA regulatory relationships. However, most existing methods do not consider the collective relationships between a group of miRNAs and the group of targeted mRNAs in the process of discovering MMRMs. Our aim is to develop a framework to discover MMRMs and reveal miRNA-mRNA regulatory relationships from the heterogeneous expression data based on the collective relationships.**

Index Terms **- miRNA-mRNA regulatory modules, Collective group relationships, Group pair, Canonical correlations.**

INTRODUCTION

Renal cancer is common cancer, and the incidence rates in males and females are 5% and 3%, respectively1. Clear cell renal cell carcinoma (ccRCC) accounts for 70–80% of renal cancer, which is the most representative subtype, and the incidence rate increased year by year. Compared with other cancers, kidney cancer-related clinical symptoms and biomarkers are less, so early diagnosis is difficult. Moreover, ccRCC has poor responses to conventional chemotherapy and radiation therapy, leading to a low 5-year survival rate of advanced patients, which is only 10–20%2,3. Nowadays, VEGF tyrosine kinase inhibitor monotherapy had been one type of standard therapy. Moreover, with the advances in immunotherapy and the more newly discovered therapeutic target, a combination of the immunotherapy and the targeted therapies could be the next standard of treatment4. Therefore, it is especially necessary to explore the internal mechanism of ccRCC to find some new therapeutic targets.

In this study, some novel RNAs may act as RNA to regulate gene expression in RCC, and their potential mechanisms have been investigated by utilizing gene chip and bioinformatics methods. The process digraph is showed in Fig. 1: Firstly, circRNAs related microarray datasets of ccRCC were obtained from GEO database, and differential expressed circRNAs (DECs) were also acquired. Then, to demonstrate whether the DECs function as ceRNAs in ccRCC, their related miRNAs and miRNA target genes have been collected, and a circRNA/miRNA/mRNA network also has been constructed. Furthermore, a protein–protein interaction (PPI) network was successfully built, and the hub-genes were also obtained. Functional enrichment and pathway enrichment analyses were performed to reveal the potential pathogenesis of ccRCC. Furthermore, connectivity map (CMap) analysis and pharmacogenomics analysis were conducted to predict bio-active compounds and potential drugs for the treatment of ccRCC, which may provide a new method in the latent therapeutic capacity of circRNAs in ccRCC.



Circular RNA (circRNA), derived from the exon or intron region of a gene, is a particular type of non-coding RNA molecule that is different from linear RNA. Compared with linear RNA structure, circRNA has no 5′–3′ polarity and no polyA tail, making it a closed circular structure. Therefore, it is more stable than linear RNA, and it is not easily degraded by RNA exonuclease or RNase5. CircRNAs' function can generalize as below (1) miRNA can regulate the post-transcriptional expression of target genes, and circRNA can act as a competing endogenous RNA (ceRNA) to bind to miRNA like a sponge to regulate the function of miRNA, thus indirectly regulating the expression of genes (2) It can affect gene expression through interacting with RNA binding protein and modulating the stability of mRNAs (3) It also can function as protein scaffolds and encode functional proteins in some cancer cells lines6,7,8,9. Recently, some studies have demonstrated that circRNA not only acts as a molecular marker but can also participate in cancer proliferation and invasion by regulating miRNA in colorectal cancer, lung cancer, and bladder cancer10.

PROCEDURE

MicroRNAs (miRNAs) are a family of small (i.e. with typical length of 19–25 nucleotides) non-protein-

coding RNA molecules that can play important regulatory roles in animals and plants [1, 2]. They base-pair with messenger RNAs (mRNAs) of protein-coding genes to induce mRNA degradation or translational repression. The mature human miRNAs potentially target majority of the human mRNAs. It has been demonstrated that miRNAs regulate a wide range of biological or cellular processes such as proliferation, metabolism, differentiation, development [9], apoptosis, cellular signaling, and cancer development and progression.

In the data pre-processing step, DICORE first creates a weighted bipartite graph representation of the relationships among the individual variables of the given miRNA and mRNA expression profiles. Taking the variables as the vertices of a weighted bipartite graph G, a weighted edge is introduced between a miRNA variable and a mRNA variable to represent their interaction. Referring to Fig. Fig.1,1, given p miRNAs and q mRNAs, let W denote the (p×q) miRNA-mRNA interaction weights matrix, where wij is the interaction weight for miRNA i targeting mRNA j. To compute miRNA-mRNA interaction weights, we calculate the Pearson correlation coefficient (PCC) [25] between each pair of miRNA and mRNA using the R built-in function, cor. The obtained PCCs are within the range of [−1,1], and the signed correlation coefficients provide two types of valuable information: the absolute values implying the strength of the miRNA-mRNA interactions (the higher the values, the stronger the interactions), and the signs indicating the directions of the associations. However, as the aim of the paper is to identify MMRMs (and thus to uncover miRNA-mRNA regulatory relationships), the collaboration score (explained in the next section) defined for discovering the modules considers the sum of the miRNA-mRNA correlations. In order to cater for both up and down miRNA regulations when calculating the total strength of the interactions, we use absolute values of the PCCs in the interaction weights matrix W, otherwise the signed PCCs or interaction weights will cancel out in Eq.

Due to the higher possibility of dense interactions in the expression profile datasets, complete weighted graph mining may not be able to distinguish correct group structure. Accordingly we used a cutoff threshold η to tradeoff between the two extreme approaches namely complete unweighted graph mining and complete weighted graph mining.

There is a growing body of literature showing that multiple miRNAs are coordinated by forming cohesive groups to collectively regulate one or more pathways [16, 17]. The collective relationships yielded between a group of miRNAs and a group of mRNAs due to the tendency of the group formation act as a vital force in catering similar functioning miRNAs and mRNAs together. Therefore, the collective relationships between cohesive groups of miRNAs and their targeted mRNAs may provide better understandings on robust and potent regulatory relationships of miRNA-mRNA regulatory modules (MMRMs).

Several algorithms have been proposed to identify MMRMs from expression data using different approaches including Bayesian network learning [18], rule induction [19], association rule mining [20], population-based probabilistic learning [21], probabilistic graphical model [22–24], matrix factorization [25], and graph mining [17, 26]. Most of these existing methods do not consider the collective relationships between a group of miRNAs and the group of targeted mRNAs in the process of identifying MMRMs. In addition, many of them are either stochastic, or require prior knowledge such as number of modules to be identified, confirmed interactions, target site information.

Adapting a greedy overlapping neighborhood expansion clustering method, ClusterONE, which was developed to discover protein complexes from protein-protein interactions networks, Li et al. [27] proposed a clustering algorithm, Mirsynergy to detect synergistic miRNA regulatory modules. However, it requires and depends on the prior knowledge of confirmed gene-gene interactions. Recently Karim et al. [28] coined the notion of collective group relationships and developed a method by integrating unweighted graphing mining concept and canonical correlation analysis to explore miRNA-mRNA regulatory relationships. However, it is noted that unweighted graph mining techniques are associated with limitation in representing the true interactions, and sometimes fail to capture correct regulatory relationships. Whereas weighted graph mining approaches can greatly improve the detection of the module structures [29], and hence regulatory relationships.

In this paper, we propose an effective computational framework, DIscovering COllective group

RElationships (DICORE) to identify MMRMs and hence reveal miRNA-mRNA regulatory relationships from heterogeneous data. In order to extract MMRMs from the given gene expression datasets, we utilize the notion of collective group relationships, which provide MMRMs with additional quantitative strength information. The method finds a deterministic solution to the problem of discovering MMRMs from weighted bipartite graph representation of the given datasets, and rank the collective group relationships based on their strength of collective relationships. We apply DICORE to a dataset for Epithelial to Mesenchymal Transition, a breast cancer dataset, and a multi-class cancer dataset. Based on the knowledge from the literature, it is observed that the identified MMRMs exhibit enriched functionality with biological significance.

METHODS

Problem statement

Consider two sets of variables $X=\{X1,\ldots,Xp\}$ and $Y=\{Y1,\ldots,Yq\}$ such that $X \cap Y = \emptyset$, representing the attributes of two different types of objects. In this paper, X and Y refer to the expression levels of a set of miRNAs and a set of mRNAs, respectively. With their given datasets, DX and DY, having n matching miRNA and mRNA expression samples, our goal is to identify any $Cx \subset X$ and $Cy \subset Y$, such that Cx and Cy are related, as a result of miRNAs in Cx collaboratively interacting with mRNAs in Cy and vice versa. We call (Cx,Cy) a group pair, and the relationship between Cx and Cy a COllective group RElationship (in short, CORE). The COREs are characterized by both group pairs and the collective relationships among the two cohesive groups in group pairs. Then the group pair (Cx,Cy) is an MMRM if the strength of the CORE between Cx and Cy is significant.

In order to discover COREs, and thus to identify MMRMs, we develop a two stages method, DIscovering CORE (DICORE). Two measures, collaboration score and canonical correlations, are employed in the two stages respectively. In the following, we firstly overview the workflow of DICORE, and then present the details of DICORE, including the definition of the collaboration score and the calculation of canonical correlations.

OVERVIEW OF DICORE

The workflow of DICORE. The overall workflow comprises a data pre-processing step and two main stages: (1) forming separate miRNA and mRNA groups and (2) searching for COREs.

CONCLUSIONS

Analysis of the results demonstrates that a large portion of the regulatory relationships discovered by DICORE is consistent with the experimentally confirmed databases. Furthermore, it is observed that the top mRNAs that are regulated by the miRNAs in the identified MMRMs are highly relevant to the biological conditions of the given datasets. It is also shown that the MMRMs identified by DICORE are more biologically significant and functionally enriched.

RESULTS

We propose DIscovering COllective group RElationships (DICORE), an effective computational framework for revealing miRNA-mRNA regulatory relationships. We utilize the notation of collective group relationships to build the computational framework. The method computes the collaboration scores of the miRNAs and mRNAs on the basis of their interactions with mRNAs and miRNAs, respectively. Then it determines the groups of miRNAs and groups of mRNAs separately based on their respective collaboration scores. Next, it calculates the strength of the collective relationship between each pair of miRNA group and mRNA group using canonical correlation analysis, and the group pairs with significant canonical correlations are considered as the MMRMs. We applied this method to three gene expression datasets and validated the computational discoveries.

REFERENCES

[1] Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. CA Cancer J. Clin.69, 7–34.

[2] Hsieh, J. J. et al. Chromosome 3p loss-orchestrated VHL, HIF, and epigenetic deregulation in clear cell renal cell carcinoma. J. Clin. Oncol.

[3] Li, X. et al. Dissecting LncRNA roles in renal cell carcinoma metastasis and characterizing genomic heterogeneity by single-cell RNA-seq. Mol. Cancer Res.16, 1879–1888.

[4] Lalani, A. A. et al. Systemic Treatment of Metastatic Clear Cell Renal Cell Carcinoma in 2018: Current Paradigms, Use of Immunotherapy, and Future Directions. Eur Urol75, 100–110.
[5] Memczak, S. et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature495, 333–338.
[6] Qu, S. et al. The emerging landscape of circular RNA in life processes. RNA Biol.14, 992–999.
[7] Qian, L. et al. The emerging role of circRNAs and their clinical significance in human cancers. Biochim. Biophys. Acta Rev. Cancer247–260, 2018.
[8] Qu, S. et al. The emerging functions and roles of circular RNAs in cancer. Cancer Lett414, 301–309.
[9] Zhong, Y. et al. Circular RNAs function as ceRNAs to regulate and control human cancer progression. Mol. Cancer17, 79.
[10] Meng, S. et al. CircRNA: functions and properties of a novel potential biomarker for cancer. Mol. Cancer16, 94.
[11] Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res.41, D991-995.
[12] Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.43, e47.
[13] Xia, S. et al. CSCD: a database for cancer-specific circular RNAs. Nucleic Acids Res.46,
[14] Dudekula, D. B. et al. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol.13, 34–42.
[15] Vlachos, I. S. et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res.43, W460–W466.
https://doi.org/10.1093/nar/gkv403 (2015).