Iran. J. Biotechnol. April 2022;20(2): e3066 DOI: 10.30498/ijb.2022.289013.3066

Research Article

DGRanker: Cancer Driver Gene Detection in Human Transcriptional Regulatory Network

Majid Rahimi [1], Babak Teimourpour [2,*], Mostafa Akhavan-Safar [3]

[1]Department of information technology, School of Systems and Industrial Engineering, Tarbiat Modares University (TMU), Tehran, Iran.
[2]School of Systems and Industrial Engineering, Tarbiat Modares University (TMU) Chamran/Al-e-Ahmad Highways Intersection, Tehran, Iran.
[3]Department of Computer and Information Technology Engineering, Payame Noor University, Tehran, Iran.

[*] *Corresponding author*: Babak Teimourpour, School of Systems and Industrial Engineering, Tarbiat Modares University (TMU) Chamran/Al-e-Ahmad Highways Intersection, Tehran, Iran. Tel: +98-218288 3987, Fax: +98-218288 4323, E-mail: b.teimourpour@modares.ac.ir

Background: Cancer is a group of diseases that have received much attention in biological research because of its high mortality rate and the lack of accurate identification of its root causes. In such studies, researchers usually try to identify cancer driver genes (CDGs) that start cancer in a cell. The majority of the methods that have ever been proposed for the identification of CDGs are based on gene expression data and the concept of mutation in genomic data. Recently, using networking techniques and the concept of influence maximization, some models have been proposed to identify these genes.
Objectives: We aimed to construct the cancer transcriptional regulatory network and identify cancer driver genes using a network science approach without the use of mutation and genomic data.
Materials and Methods: In this study, we will employ the social influence network theory to identify CDGs in the human gene regulatory network (GRN) that is based on the concept of influence and power of webpages. First, we will create GRN Networks using gene expression data and Existing nodes and edges. Next, we will implement the modified algorithm on GRN networks being studied by weighting the regulatory interaction edges using the influence spread concept. Nodes with the highest ratings will be selected as the CDGs.
Results: The results show our proposed method outperforms most of the other computational and network-based methods and show its superiority in identifying CDGs compared to many other methods. In addition, the proposed method can identify many CDGs that are overlooked by all previously published methods
Conclusions: Our study demonstrated that the Google's PageRank algorithm can be utilized and modified as a network-based method for identifying cancer driver gene in transcriptional regulatory network. Furthermore, the proposed method can be considered as a complementary method to the computational-based cancer driver gene identification tools.

Keywords: Cancer Driver Gene, Diffusion, PageRank, Transcriptional Regulatory Network (TRN)

1. Background

Many studies have been conducted about the detection of cancer driver genes (CDGs) (1-3). The majority of the proposed methods have been designed based on the notion that CDGs are genes that experience more general gene expression changes also known as mutation. Of course, not all mutations in the cancer genome are related to cancer. Hence, most computational methods try to distinguish between cancer-causing mutations and non-cancer-causing ones. Most existing methods rely on transcriptomic or genomic data to identify CDGs. In DawnRank (4) driver genes detected using mutational data and gene interaction network. ActiveDriver method (3) used of

post-translational and mutated data. In e-Driver method (5) and (6) detected driver genes with mutation rate of a protein. In OncodriveFM (7) and OncodriveCLUST (8) proteins functional impact has been examined. In some methods like Dendrix (9), Memos (10), MSEA (11), CoMDP (12) and DriverNet (13) applied of mutation profiles and pathways. iMaxDriver (14) used influence maximization to identify cancer driver gene.

Nevertheless, there are still some limitations and deficiencies in the proposed methods. These methods have a high false-positive rate and a low rate of precision and F-measure in their results (For example, 0.013 to 0.103 in breast cancer). Another point is that these methods are mostly reliant on mutation data which are noisy and may not always be available. Another limitation of previous methods is the low number of detected drivers and high false positive values in the results. For example, the iPac computational method for breast cancer identified 4821 genes as drivers, of which 250 genes were actually cancer drivers. The number of diagnostic drivers for each method is shown in the results section. Given these limitations, we present a simple and computationally lightweight method to identify the CDGs in the network by applying the power of interaction; Without the need for mutation data or time-consuming calculations, this method performs better than all of existing methods (15).

In this research, we employed the concept of influence and spread in the network based on Google's webpage influence algorithm to identify CDGs in TRNs[1] relating to breast, colorectal and lung cancers. We demonstrated that this method can improve accuracy of CDGs identification compared by all of computational methods and even the recently proposed one that is based on influence maximization. _

2. Objectives

In this study, we proposed a network-based algorithm for cancer driver gene detection using transcription regulatory network. To achieve this purpose, the web page ranking algorithm has been modified to be used in weighted transcription regulatory networks.

3. Materials and Methods

3.1. PageRank and Transcriptional Regulatory Network

Social networks connect many people within a short amount of time and have created a revolution in how people communicate. Information is expanded across social networks, ideas and knowledge are shared through social networks and people can influence others by interacting on social networks. Therefore, many problems are studied by analyzing social networks such as social influence and diffusion models. Based on influence and diffusion models, we can evaluate the influence or reputation of a person on social networks which is important for the identification of influential spreaders (16, 17,18). A user's influence and reputation are different. Google's PageRank (19) algorithm is one of the algorithms that calculates the importance and influence of a web page. This algorithm assign a weight as PageRank score to any entity based on the mutual relations between the entities. These ranking scores are applied to the web graph using a mathematical algorithm. An external link to a page is considered as a numerical number for increasing its rank. A page that is linked to by many pages with high PageRank receives a high rank itself. Therefore, in the PageRank algorithm, the page that has the most number of incoming links from other pages will have the highest rank and importance.

For example, consider a small set containing the four pages of A, B, C and D as shown in **Figure 1**.

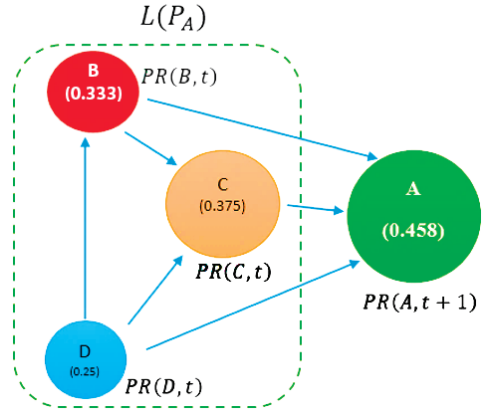

Figure 1. Calculating the rank of each page based on the incoming links to the page from other pages and their PageRank values. At the first stage of the algorithm, page B will give half of its PageRank value namely 0.125 to page A and will give the other 0.125 to page C. Page C will transfer all of its PageRank value that is 0.25 to page A. Also, page D that links to the other three pages will give one third of its PageRank that is almost 0.083 to the other three pages. At the end of this stage, page A will have a PageRank value of almost 0.458.

[1]Transcriptional regulatory network

$$PR(A) = \frac{PR(B)}{d_{out}(B)} + \frac{PR(c)}{d_{out}(C)} + \frac{PR(D)}{d_{out}(D)} = \frac{PR(B)}{2} + \frac{PR(c)}{1} + \frac{PR(D)}{3} \quad (1)$$

In the original version of PageRank, the total PageRank of all pages equals the total number of pages on the web at a particular time, therefore, the initial value of these four pages is equal to 1.In this example, the initial value of PageRank for each page will be 0.25. The calculation of the page rank value of each page is shown in **Figure 1**. In other words, through outgoing links the PageRank value is given to pages containing the document targeted by the links and its value is equally divided among the outgoing links of the page.

In general, the PageRank value for each page *u* can be stated as follows:

$$PR(A) = \sum_{v \in L(P_A)} \frac{PR(v)}{L(v)} \quad (2)$$

The PageRank value for each page *A* depends on the PageRank of each page *V* located in the set $L(P_A)$ (a set containing all pages linked to page *u*) and is divided by the L(v) number that is the number of outgoing links on page *V*.

PageRank assumes a person randomly clicks on clinks and these clicks will eventually stop. In every stage, the probability of the person continuing to click is equal to the value of the damping factor *d*. Based on past studies (20), the best value for the damping factor is 0.85. The value of the damping factor is subtracted from 1 (and in some variations of the algorithm, the result is divided by the number of documents (N) in the collection) and then this value is added to the product of the damping factor and the sum of the incoming PageRank scores. The overall equation is as follows:

$$PR(p_i; t+1) = \frac{1-d}{N} + d \sum_{P_j \in M(P_i)} \frac{a_{ji}, PR(p_j; t)}{L(p_j)} \quad (3)$$

$PR(p_i; 0) = \frac{1}{N}$ Page i at time 0.

$PR(p_i; 0) = 1/N$ Page i at time 0.

$p_1; p_2; \ldots; p_n$: The desired under consideration

d: the damping factor between 0 and 1, usually 0.85.

$L(P_j)$ = the number of outbound links on *Pj*

N: the number of webpages

M (p_i): the set of pages that link to p_i

A: the adjacency matrix

In other words:

$$\boldsymbol{R}(t+1) = \frac{1-d}{N}, \mathbf{1} + d\mathcal{M}\boldsymbol{R}(t) = \frac{1-d}{N}, \mathbf{1} + d(K^{-1}, A)^T \boldsymbol{R}(t) \quad (4)$$

Where:

$$\boldsymbol{R}_i(t) = \boldsymbol{PR}(p_i; t) \ , \ \mathbf{1} = \begin{bmatrix} 1 \\ , \\ , \\ , \\ 1 \end{bmatrix}, \mathcal{M}_{ij} = \begin{cases} \frac{1}{L(p_j)} & if\ j \to i \\ 0 & otherwise \end{cases}$$

A: adjacency matrix

K: the diagonal matrix with the outdegrees in the diagonal

Therefore, the page with the highest number of incoming links from other pages will have a higher importance level as well as a higher PageRank score. But, in a TRN the opposite happens. Based on the theory of spread and influence in social networks, we assume that the gene with the greatest influence on other genes is most probably the cause of the creation and development of abnormality in other genes and leads to cancer. Therefore, we can identify driver genes by extending and using a weighted PageRank algorithm for the human TRN. The concept of the use of the PageRank algorithm in gene networks was also employed in the past. This concept was used for evaluating experimental microarray results with the use of a network constructed using gene ontologies and expression profile correlations (20). In addition, a modified version of the PageRank algorithm was employed to assign a weight to genes in a gene interaction network to identify the subtypes of a cancer where the median absolute deviation (MAD) and gene expression data in the algorithm were employed to assign weight to the genes (21). Also, the concept of webpage ranking was utilized to rank genes in type 2 diabetes and mutation data was used for weighting (22). Although this concept has been used for ranking in other gene networks, in the majority of cases, weight and interaction power were not considered and only nodes were weighted and or the weight of edges was applied to the rank of the target genes. In this research, we will employ the algorithm with the concept of spread in human TRNs to identify CDGs and given the type and nature of a network and whether the interaction power of edges in the gene regulatory network is not equal in every case, we will consider the respective weight

of each edge in weighting each gene in the network and also we will employ the difference in the value of gene expression in healthy and cancer tissues instead of the constant value in the original PageRank algorithm. The results show that the edge weight as utilized in this research, has significantly affected the results. It outperformed 6 out of the 17 methods benchmarked to identify the number of CDGs and also offers higher accuracy.

In **Figure 2**, you can see an illustration of a TRN in humans. As you can observe, in contrast to webpages, in this network, the gene with the highest number of outgoing links to other genes will have a higher influence score. In this example, it is clear that the AATF gene has a higher rank compared to the BAX gene.

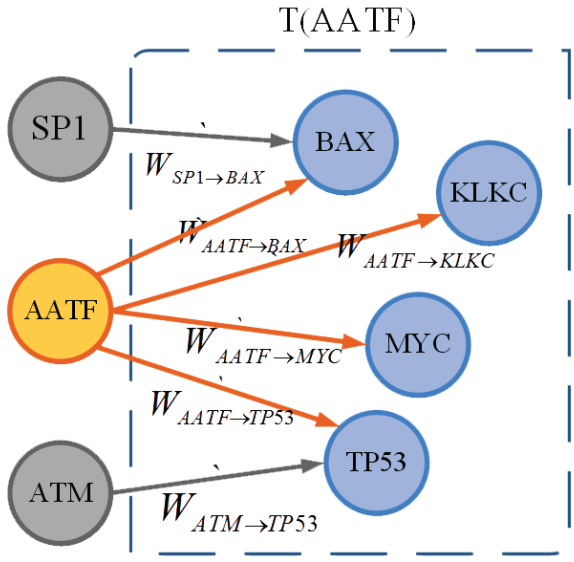

Figure 2. The structure of the human transcriptional regulatory network with the concept of influence to rank genes

According to the **Figure 2**, by reversing the formulas associated with the PageRank algorithm for influence, the formulas will be as follows:

$$GR(gene_i, t+1) = DEXP(gene_i) + d \times \sum_{gene_j \in T(gene_i)} \frac{a_{ij}, GR(gene_j, t)}{w_{gene_i \to gene_j} \times degree_{in}(gene_j)} \quad (5)$$

$$DEXP(gene_i) = \frac{\left|\sum_{i=1}^{3} Normal_{exp_{gene_i}} - \sum_{i=1}^{3} Cancer_{exp_{gene_i}}\right|}{3}$$

$GR(gene_i, 0) = 1$, gene i at time 0.

First, to measure the influence of the expression values and biological properties on the algorithm, the first numerical term is replaced with the absolute value of the gene expression value in the health and cancer tissues of 3 patients. $T(gene_i)$ is a set of genes whose expression value is influenced by gene i, in other words, it represents genes with incoming links from $gene_i$. Also, $Degree_{in}(gene_j)$ is the degree of input of the genes that are members of the $T(gene_i)$ set. In other words, it represents the number of incoming links. To include the power of each connection in the TRN, since the influencing power of each gene on another gene may differ, we have multiplied the incoming edge weight to the gene in the $T(gene_i)$ set by the denominator to apply the existing edge weight to the calculation of the gene rank in the respective outgoing link and not to the calculation of the rank of the target gene in the previous stage. With this definition, the equations in the TRN for the ranking of genes will be as follows:

$$\boldsymbol{GR}(t+1) = \boldsymbol{DEXP} + d, \mathcal{H}, \boldsymbol{GR}(t) = \boldsymbol{DEXP} + d(W, K^{-1}, A)^T \boldsymbol{R}(t) = \boldsymbol{DEXP} + d(W, K^{-1}, A)^T \boldsymbol{R}(t) \quad (6)$$

$$\mathcal{H}_{ji} = \begin{cases} \frac{1}{D^{in}(gene_j) \times w_{gene_i \to gene_j}} & if\ i \to j \\ 0 & otherwise \end{cases}, \quad \boldsymbol{DEXP} = \begin{bmatrix} DEXP(gene_i) \\ , \\ , \\ , \\ DEXP(gene_n) \end{bmatrix}$$

A: adjacency matrix
K: the diagonal matrix with the indegrees in the diagonal
W: the weighted matrix

3.2. Network Construction

To construct transcriptional regulatory networks (TRN) in order to apply the ranking algorithm, RegNetwork (23) database have been used. It is one of the databases used to construct gene regulatory networks (14, 24) .In RegNetwork, gene regulatory interactions have been collected from multiple databases with the use of several methods. These interactions include human and mouse GRNs. We recovered the information related to the human TRN. Of course, in this network, there was some other information about interactions in the regulatory network including the regulatory network of miRNA on that were removed from the final network being studied. The final information included 150202 interactions related to TF-TF and TF-mRNA. In this network, for each interaction, a confidence level has reported. We used it to assign a weight to each interaction. These confidence levels included “low”, “high” and medium. Assigned edge weights According to previous research (0.2, 0.5 and 0.8 for low, medium and high confidence).

In some study like (25) an optimization function is proposed to weighting the interactions.

3.3. Gene Expression Dataset

To apply biological data to the algorithm we required gene expression dataset. Since the study is performed on the breast, colorectal and lung cancers, the gene expression data of these three cancer types were downloaded from the GEO database (GSE3268, GSE32323 and GSE15852). The data is available in files with the CLE format.

In each of the selected GEO datasets, gene expression values are reported for both the tumor tissue and its adjacent normal tissue. The expression values extracted using RMA method implemented in Affy package in R (26). The output files needed to be processed. At first we removed rows with missing gene ID and then combined gene expression values which their gene ID was synonym by averaging gene expression value of respective columns, afterwards we normalized the gene expression value such that by dividing each value by maximum the value of gene expression in the relevant tissue type (tumor or normal).

Based on the extracted gene expression data of the three cancer types being studied and also the gene regulatory interactions extracted from the RegNetwork database, we created three networks to apply the ranking algorithm. So we compared each database's regulatory interactions list with genes possessing expression in each cancer type. The edges whose origin and destination were not both in the respective cancer's gene expression list were removed from the network being studied. Therefore, we constructed three regulatory networks related to breast, colorectal and lung cancers. In each of the constructed networks, the interactions were weighted as described in Section 3.2. The algorithm was applied to each of the three networks, and the genes were ranked.

3.4. Evaluation Methods

We evaluated this method by comparing its results with seventeen popular computational and network-based CDG prediction methods that were used by (14). The list of the genes identified as driver genes by the 15 computational methods mentioned earlier was obtained for evaluation with similar inputs for all the methods with the use of DriverDB v2 (27). In DriverDB v2 there are various lists different cancer datasets, we used breast invasive carcinoma (BRCA), lung squamous cell carcinoma (LUSC) and colon adenocarcinoma (COAD). Moreover, the results were compared with the results obtained using the latest influence-based method for the identification of CDGs introduced in (22). Also, the accuracy of the predicted CDGs was evaluated by comparing each list with the Cancer Gene Census (CGC) (28) gene list as the gold standard .The modified ranking algorithm has been applied to each of the 6 cancer networks being studied and the genes were arranged in descending order. Next, by interpreting the results based on the threshold value, the genes were divided into the two categories of driver and non-driver. To set the accurate value of the threshold used for categorization, the pROC package (29) was utilized in R. Next, the three criteria of Recall, Precision and F-measure have been employed to evaluate recall and precision. The "recall" criterion represents the ratio of "the number of correctly packaged data" in a certain class to the data that must be classified in that same class. A higher "recall" value indicates that there are very few data that are not correctly classified. It is not right to use only this criterion to evaluate system performance and it should be employed in combination with the "precision" criterion.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

The "precision" criterion evaluates the "ratio of correctly performed predictions" for the samples of a certain class to the total "number of predictions" for the samples of the same class (this number includes all the correct and incorrect predictions).

$$Precission = \frac{TP}{TP + FP} \tag{8}$$

The F-measure criterion combines the parameters of "precision" and "recall" to find out how good the performance of a packaging model is. This criterion is also known as the Harmonic Mean of the two criteria of precision and recall. This criterion illustrates a more accurate image of the packaging model's performance on all the classes in the data.

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

4. Results

In this section, the results of the proposed method on different cancer networks are presented.

4.1 Compare of F-measure and number of detected drivers

The results showed, in the breast cancer network, DGRanker performed better than the other methods. **Figures 3** to **5** show the results obtained from the algorithm on the network by considering the power of each incoming edges. The assignment of weight as presented in this study led to a significant improvement in the results. In the network associated with breast cancer, DGRanker has performed better at identifying the number of driver genes than all the methods except iPac. It should be mentioned that DGRanker had the highest F-Measure value among all the methods. In addition, DGRanker with recall = 0.311 had the highest value among all methods (after iPac). The same results have been obtained in the COAD network. DGRanker with F-measure= 0.224 and the number of detected drivers had the highest value among all methods. In addition, it has the highest value among all computational and network-based methods with recall = 0.337 after iPac. In the lung cancer transcription regulation network, the proposed method with Recall = 0.332 and detection of 182 drivers, has the first rank among all computational and network-based methods. In addition, DGRanker with f-measure = 0.225 after iMaxDriver-W has the highest value among all methods.

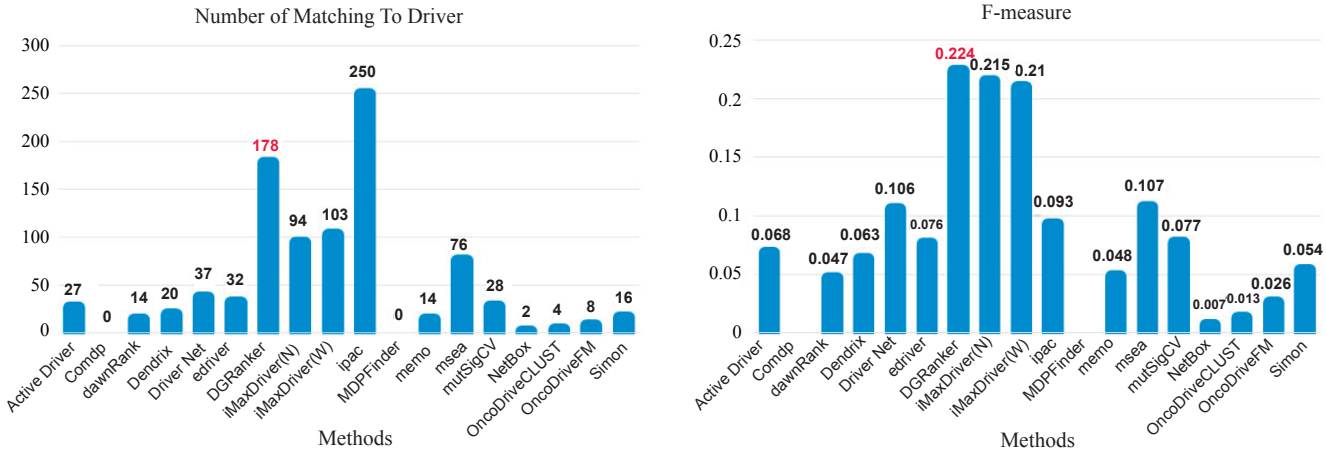

Figure 3. The F-measure of DGRanker and other seventeen computational methods proposed for CDG prediction in breast cancer network

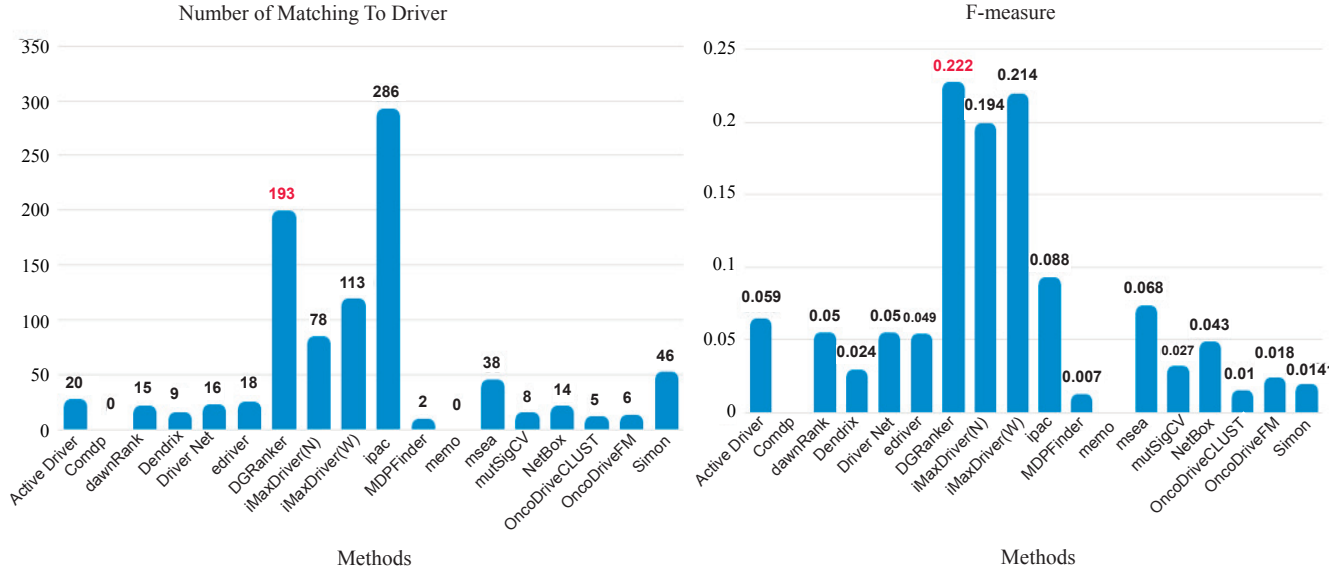

Figure 4. The F-measure of DGRanker and other seventeen computational methods proposed for CDG prediction in colon adenocarcinoma cancer network

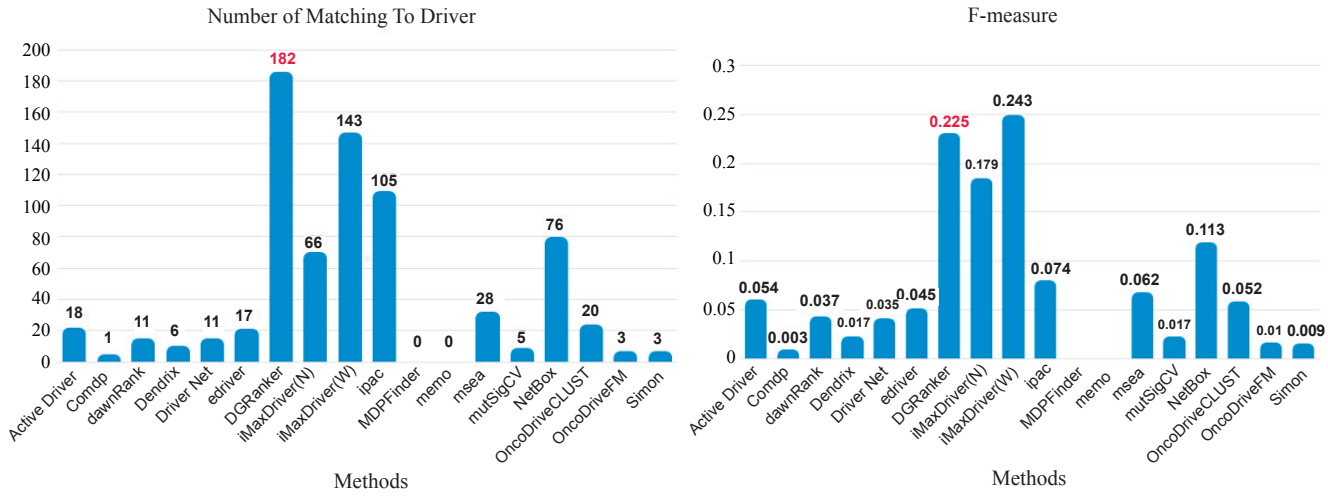

Figure 5. The F-measure of DGRanker and other seventeen computational methods proposed for CDG prediction in lung squamous cell carcinoma cancer network

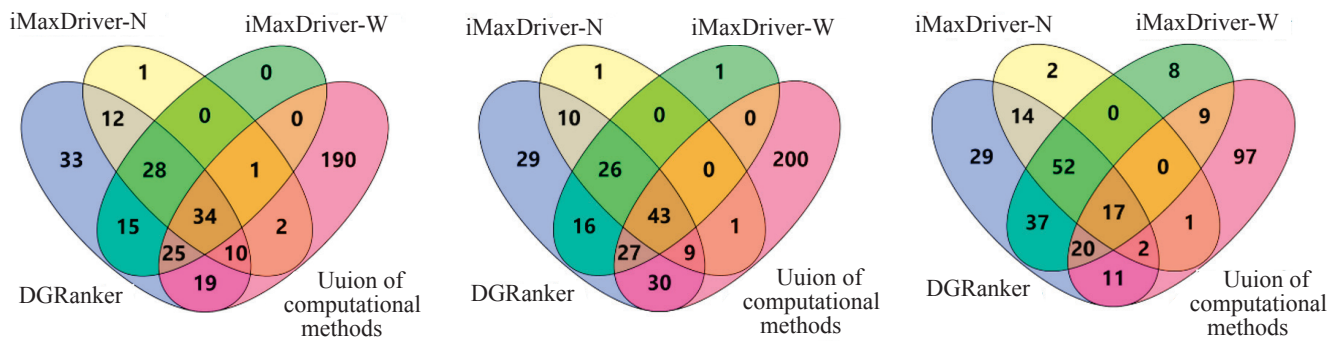

Figure 6. The overlap of driver genes identified by DGRanker and other computational and network-based methods.

4.2. Compare the overlap of detected CDGs

We compared the overlap of driver genes identified by our proposed method and other methods. As shown in **Figure 6**, DGRanker identified 33, 29, and 29 unique genes in three cancers of the breast invasive carcinoma, colon adenocarcinoma, and lung squamous cell carcinoma, respectively, that have not been identified by any of the other methods. In addition, the proposed method detects many genes identified by other methods. This shows that DGRanker can be used as a complementary tool to computational approaches.

5. Discussion

Most CDG identification methods mentioned in introduction employ computational methods and the concept of mutation, genomic data analysis, and pathway data analysis to identify CDGs. These methods have a high computational data size and do not have a good precision level in identifying cancer genes. For this reason, to eliminate these two problems, we employed a network-based method with lighter computations to identify CDGs that uses the concept of spread in social networks and the webpage influence ranking algorithm concept. Given the mechanism for the development of abnormalities in the cell, we hypothesized that genes that had higher Influence were likely to be drivers. To calculate the Influence of each gene, we enriched the PageRank algorithm with biological concepts and modified it based on the concept of Influence. In addition, we entered the weight of regulatory interactions based on the changes made in the algorithm.

The results showed that the use of interaction weights significantly improves the results. As the results show, the proposed method in breast cancer has identified 178 drivers, which is the first among all methods after the iPac method. Although some of these methods, like iPac predict many CDGs in their output, they are not an acceptable precision. For example, in breast cancer network iPac has a precision=0.052. While in DGRanker precision and recall are 0.117 and 0.311, respectively. Also, compared to previous network-based methods, the DGRanker has improved the number of predicted drivers and F-measure by 72.82 and 7.14%, respectively. In colorectal cancer, the DGRanker was able to predict 193 drivers, which is the first among all methods (after the iPac). In this network, the amount of precision for the proposed method was 0.166, while in the iPac method it was 0.048. In addition, compared to previous network-based methods (iMaxDriver methods), the DGRanker has improved the number of predicted drivers and F-measure by 70.79 and 3.73%, respectively. In lung cancer, the proposed method was able to predict 182 drivers, which was the highest value among all previous computational and network-based methods. After the proposed method, the iMaxDriver-W and the iPAC computational method have the best performance. In this network, the DGRanker had precision=0.173 and recall=0.332(the highest value among all methods). However, in terms of harmonic means, the iMaxDriver-W and the proposed method have the best performance, respectively. The results for other methods are shown in **Figures 3 to 5**.

Although the precision and recall values have improved in the proposed method, these two criteria alone cannot show the system performance as well. For this reason, the harmonic mean of these two criteria, F-measure, was used. The F-measure of the proposed method had the highest value in breast (f-measure= 0.222) and colon (F-measure=0.222) cancers and the highest value in lung cancer after the iMaxDriver-W network-based method. The F-measure for other methods are shown in **Figures 3 to 5**. In terms of driver overlap, the proposed method in breast, colon and lung cancers identified 143, 161 and 153 drivers, respectively, identified by other methods. In addition, DGRanker reported 33, 29 and 29 new driver genes in breast, colon and lung cancer, respectively, that have not been reported as drivers by any of the other computational and network-based methods. The results show that using the concept of influence in the transcriptional regulatory network and also applying the weight of interactions in calculating influence scores improves the performance of driver gene detection methods.

6. Conclusion

Identification of cancer driver genes is very important in prevention and treatment. Various methods have been proposed for this purpose, most of which are computational and use mutation data. Using the structure of transcriptional regulator networks and influence-based approaches can improve the efficiency of existing methods. In this study, we proposed a network-based approach, called DGRanker, for CDGs discover in transcriptional regulatory networks. We used the modified PageRank algorithm to calculate the influence scores. The weight of regulatory interactions is also added in this algorithm. Finally, the genes with the highest scores are categorized as drivers. The results show that using the structure of transcriptional regulation networks improve the prediction results and can be used more in the future. As DGRanker improved the results of previous computational and network-based methods. In addition, the results show that DGRanker can find complementary drivers for the most state-of-the-art CDG prediction methods.

References:

1. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, *et al*. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*. 2008;**455**(7216):1069-1075. doi: 10.38/nature07423.
2. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016;**17**(1):1-13. doi: 10.1186/s13059-016-0994-0.
3. Reimand J, Wagih O, Bader GD. The mutational landscape of phosphorylation signaling in cancer. *Sci Rep*. 2013;**3**(1):1-9. doi:10.1038/srep02651.
4. Hou JP, Ma J. DawnRank: discovering personalized driver genes in cancer. *Genome Med*. 2014;**6**(7):1-16. doi: 10.1186/s13073-014-0056-8.
5. Porta-Pardo E, Godzik A. e-Driver: a novel method to identify protein regions driving cancer. *Bioinformatics*. 2014;**30**(21):3109-3114. doi: 10.1093/bioinformatics/btu499.
6. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics*. 2011;**27**(2):175-181. doi: 10.1093/bioinformatics/btq630.
7. Gonzalez-Perez A, Lopez-Bigas N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res*. 2012;**40**(21):e169-e. doi: 10.1093/nar/gks743.

8. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. Oncodrive CLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013;**29**(18):2238-4224. doi: 10.1093/bioinformatics/btt395.
9. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. *Genome Res*. 2012;**22**(2):375-385. doi:10.1101/gr.120477.111.
10. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res*. 2012;**22**(2):398-406. doi:10.1101/gr.125567.111. .
11. Arneson D, Bhattacharya A, Shu L, Mäkinen V-P, Yang X. Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. *BMC genomics*. 2016;**17**(1):1-9. doi: 10.1186/s12864-016-3057-8.
12. Zhang J, Wu L-Y, Zhang X-S, Zhang S. Discovery of co-occurring driver pathways in cancer. *BMC Bioinf*. 2014;**15**(1):1-14. doi: 10.1186/471-2105-15-271.
13. Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, *et al*. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol*. 2012;**13**(12):1-14. doi:10.1186/gb-2012-13-12-r124.
14. Rahimi M, Teimourpour B, Marashi S-A. Cancer driver gene discovery in transcriptional regulatory networks using influence maximization approach. *Comput Biol Med*. 2019;**114**:103362. doi:10.1016/j.compbiomed.2019.
15. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009;**10**(4):252-263. doi: 10.1038/nrg2538.
16. Cataldi M, Di Caro L, Schifanella C, editors. Emerging topic detection on twitter based on temporal and social terms evaluation. *KDD '10*. 2010;**4**:1–10. doi:10.1145/1814245.1814249.
17. Chen W, Wang Y, Yang S, editors. Efficient influence maximization in social networks. *KDD '9*; 2009.**p**.199-208. doi:10.1145/1557019.1557047.
18. Prakash BA, Faloutsos C. Understanding and managing cascades on large graphs. *VLDB Endowment*. 2012;**5**(12):2024-2025. doi:10.14778/2367502.67.
19. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. *Stanford InfoLab*; 1999.
20. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*. 1998;**30**(1-7):107-117. doi: 10.1016/S0169-7552(98)00110-X.
21. Liu Y, Gu Q, Hou JP, Han J, Ma J. A network-assisted co-clustering algorithm to discover cancer subtypes based on gene expression. *BMC Bioinf*. 2014;**15**(1):1-11. doi: 10.1186/471-2105-15-37.
22. Shang H, Liu Z-P. Prioritizing type 2 diabetes genes by weighted PageRank on bilayer heterogeneous networks. *IEEE/ACM Trans. Comput Biol Bioinf*. 2019;**18**(1):336-346. doi: 10.1109/TCBB.2019.2917190.
23. Liu Z-P, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*. 2015;**2015**. doi: 10.1093/database/bav095.
24. Akhavan-Safar M, Teimourpour B. KatzDriver: A network based method to cancer causal genes discovery in gene regulatory network. *Biosystems*. 2021;**201**:104326. doi: 10.1016/j.biosystems.2020.
25. Akhavan-Safar M, Teimourpour B, Kargari M. GenHITS: A network science approach to driver gene detection in human regulatory network using gene's influence evaluation. *J Biomed Inf*. 2021;**114**:103661. doi: 10.1016/j.jbi.2020.
26. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;**20**(3):307-315. doi:10.1093/bioinformatics/btg405.
27. Chung I-F, Chen C-Y, Su S-C, Li C-Y, Wu K-J, Wang H-W, *et al*. DriverDBv2: a database for human cancer driver gene research. *Nucleic Acids Res*. 2016;**44**(D1):D975-D979. doi: 10.1093/nar/gkv314.
28. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, *et al*. A census of human cancer genes. *Nat Rev Cancer*. 2004;**4**(3):177-183. doi:10.1038/nrc299.
29. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf* . 2011;**12**(1):1-8. doi:10.1186/471-2105-12-77.