FULL LENGTH	Iranian Biomedical Journal 23 (1): 34-46 January 2019

Differential Connectivity in Colorectal Cancer Gene Expression Network

Fereshteh Izadi

Sari Agricultural Sciences and Natural Resources University (SANRU), Farah Abad Road, Mazandaran 4818168984, Iran

Received 10 February 2018; revised 17 March 2018; accepted 19 March 2018

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]

[DOR: 20.1001.1.1028852.2019.23.1.5.5]

[DOI: 10.29252/ibj.23.1.34]

ABSTRACT

Background: Colorectal cancer (CRC) is one of the challenging types of cancers; thus, exploring effective biomarkers related to colorectal could lead to significant progresses toward the treatment of this disease. **Methods:** In the present study, CRC gene expression datasets have been reanalyzed. Mutual differentially expressed genes across 294 normal mucosa and adjacent tumoral samples were then utilized in order to build two independent transcriptional regulatory networks. By analyzing the networks topologically, genes with differential global connectivity related to cancer state were determined for which the potential transcriptional regulators including transcription factors were identified. **Results:** The majority of differentially connected genes (DCGs) were up-regulated in colorectal transcriptome experiments. Moreover, a number of these genes have been experimentally validated as cancer or CRC-associated genes. The DCGs, including *GART*, *TGFB1*, *ITGA2*, *SLC16A5*, *SOX9*, and *MMP7*, were investigated across 12 cancer types. Functional enrichment analysis followed by detailed data mining exhibited that these candidate genes could be related to CRC by mediating in metastatic cascade in addition to shared pathways with 12 cancer types by triggering the inflammatory events. **Conclusion:** Our study uncovered correlated alterations in gene expression related to CRC susceptibility and progression that the potent candidate biomarkers could provide a link to disease. ***DOI: 10.29252/ibj.23.1.34***

Keywords: Colon cancer, Colorectal cancer, Transcriptional networks

Corresponding Author: Fereshteh Izadi
Sari Agricultural Sciences and Natural Resources University (SANRU), Farah Abad Road, Mazandaran 4818168984, Iran; Mobile: (+98-918) 6291302; E-mail: izadi1991@yahoo.com

INTRODUCTION

Colorectal cancer (CRC) is a fatal malignancy with estimated 1.4 million cases yearly[1]. In spite of conducting leading researches to elucidate the molecular processes that advocate the normal colorectal cells toward cancer, the rate and average years of survival have not profoundly changed over decades. Experimental evidence has demonstrated the function of a certain number of genes such as *HMGA1*, *TACSTD2*[2], *SLC6A4*[3], *COL3A1*[4], *ITGA2*[5], *TXNDC17*[6], and *PPP2R5A*[7] in CRC. Although the association of genes in CRC has been presented in a number of research works[8,9], employing robust algorithms in network mining and topology analysis offers an unprecedented opportunity in depicting the etiology of cancers. Rewiring of the biological networks, to detect differentially co-regulated (DRGs) and co-expressed genes (DCG), could simplify the network's components observation and assist to depict the relationships between interconnected genes. Gene co-expression networks enable to highlight molecular mechanisms underlying diseases[10] and can be accounted as an efficient way to assess CRC. Generally, tools designed for recovering gene regulatory interactions rely on similarity matrices indirectly measured by correlation matrices or mutual information. These matrices usually include many indirect links that should be identified and removed for increasing the reliability of gene regulatory network (GRN) inference algorithms. Hence, several sophisticated approaches have attempted to remove

indirect interactions and to detect the causal relationships between gene pairs. Differential co-expression analysis aids to detect gene with different connectivity in the disease state and offers a powerful approach for elucidating transcriptome patterns and dysfunction of gene expression underlying phenotypic changes[11]. A plenty number of differentially co-expression network methods have been proposed in the literature[12-17]. For instance, DCGs and links (DCGL) attempts to identify DRGs and its links (DRLs) by comparing the expression datasets of disease and normal states[11,12]. Weighted gene co-expression network analysis (WGCNA) is a relatively new statistical method not only infers correlation patterns between two genes but also covers neighborhoods across expression data[13].

In this work, instead of DRGs, we focused on genes with differential connectivity in cancer state versus normal condition. These genes, indicating hubs within the network, supposedly to be key units controlling a wide range of essential cellular functions in a specific process like cancers. Thus, the presence of potential differential interactions through CRC genes expression datasets has been investigated. We mainly aimed to uncover the mediated relationships between genes using *in silico* approaches. The differentially expressed connected genes and molecular pathways, which we previously thought to influence the pathogenesis of CRC, were subsequently prioritized.

MATERIALS AND METHODS

Used datasets and pre-processing

In this work, we collected samples of normal human colorectal mucosa and adjacent CRC of four independent whole genome expression series (single-color Affymetrix Human Genome U133 Plus 2.0 Array). After a comprehensive search in NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/gds/?term=), series with accession numbers GSE4183[18], GSE8671[19], GSE9348[20], and GSE18105[21] consisting of 84 normal and 210 CRC cases were collected. Raw CEL files of these samples, based on the platform of GPL6244, were normalized with robust multi-array average (RMA) expression measure[22] method by using the linear models for microarray data (LIMMA) R package[23] (R software v. 3.2.5). After removing ambiguous probes, the extracted probe IDs were transformed into 21654 unique and validated official gene symbols. After normalization, differentially expressed genes (DEGs) were identified between cancer and normal mucosa if the expression level alteration was above the defined threshold (fold-change >2.0 or <0.5 and adjusted *p* value <0.01) by employing LIMMA R package. The defined threshold prevents withdrawing genes with lower differential alteration.

Assessing regulatory interactions and topological analysis

The expression values of mutual DEGs among the GSE4183, GSE8671, GSE9348, and GSE18105 series were used to construct two independent GRNs, one from 84 normal samples and another from 210 CRC samples, by employing Graphical Gaussian Models (GGM), as implemented in GeneNet R package[24]. To cover all of the mutual DEGs between control and CRC samples and more sparsity, only the 1500 top ranking edges were visualized in Cytoscape (v. 3.4.0). Using Cytoscape's built-in Network Analyzer, we set the nodes with higher degrees and betweenness with a darker shade and bigger size, respectively. In order to analyze the topology of two independent constructed GRNs, betweenness centrality (the percentage of times a node appears on the shortest path between all pairs of nodes in the network), as a network centrality parameter, was calculated. Genes with higher betweenness centrality score, as globally connected genes, were then identified, through the CytoNCA[25]. Globally connected nodes were determined if they were in the top 40% of betweenness centrality score distribution; genes ranked between top 1–40%. Based on Yu *et al.*'s[26] report, the selection of genes in the range of 10-40% of this distribution did not have a significant effect on the results. These genes thereof were remarked as differential connected genes (DCGs) between normal and CRC conditions. Afterward, in order to extract transcription factor (TF)-gene regulatory interactions among the DCGs, a list of 9905 regulatory links between human gene TF was obtained from TRRUST database[27]. These regulatory connections were collected from 11,237 over 20 million PubMed articles and experimentally validated transcriptional regulations consisting of 821 human TFs and 2,159 target genes of TFs.

Pathway, gene ontology (GO) enrichment, and expression pattern analysis

Mutual DEGs were separately classified by utilizing KEGG[28] to underlying pathways and to GO molecular labels by Enrichr[29]. DCGs were finally fed into pathwAX web server[30] to find a network crosstalk of significant pathways. PathwAX contains KEGG pathway information in addition to networks of gene-gene links in model organisms. The expression patterns of DCGs was ultimately sought by GEMMA database[31]. To examine the extent to which the DCGs

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

are CRC specific in contrast to being generally expressed in several cancer types, we obtained DEGs from bladder, breast, colon, esophagus, kidney, liver, lung, skin, ovary, prostate, sarcoma, and thyroid cancers with accession numbers GSE11545[32], GSE20437[33], GSE25071[34], GSE34619[35], GSE20602[36], GSE49515[37], GSE43346[38], GSE6887[39], GSE14407[40], GSE45016[41], GSE2719[42], and GSE53072[43], respectively. As described before, the threshold cut-off of fold-change >2.0 or <0.5 and adjusted *p* value < 0.01 was employed to extract DEGs in 12 cancer types. The power analysis of the selected genes was conducted by calculating the survival time statistics based on the log-rank test and visualized as Kaplan-Meier survival curve[44].

RESULTS

Differential expression of CRC-related genes across healthy mucosa and adjacent tumor tissues

Our major criteria for selecting the four aforementioned expression arrays was the avoidance of pooling transcriptome data of cell lines, *in vitro* assays, or gene expression measurement under any treatment. Genes whose expression level in normal colon mucosa displayed 2> or < 0.5 fold-change at adjusted *p* value <0.01, in comparison to tumour tissues, were selected as statistically significant DEGs. The range of DEGs through these series was unequal from 392 genes in GSE4183 to 11591 in GSE18105 (Fig. 1A). In addition, among the expression values of DEGs among

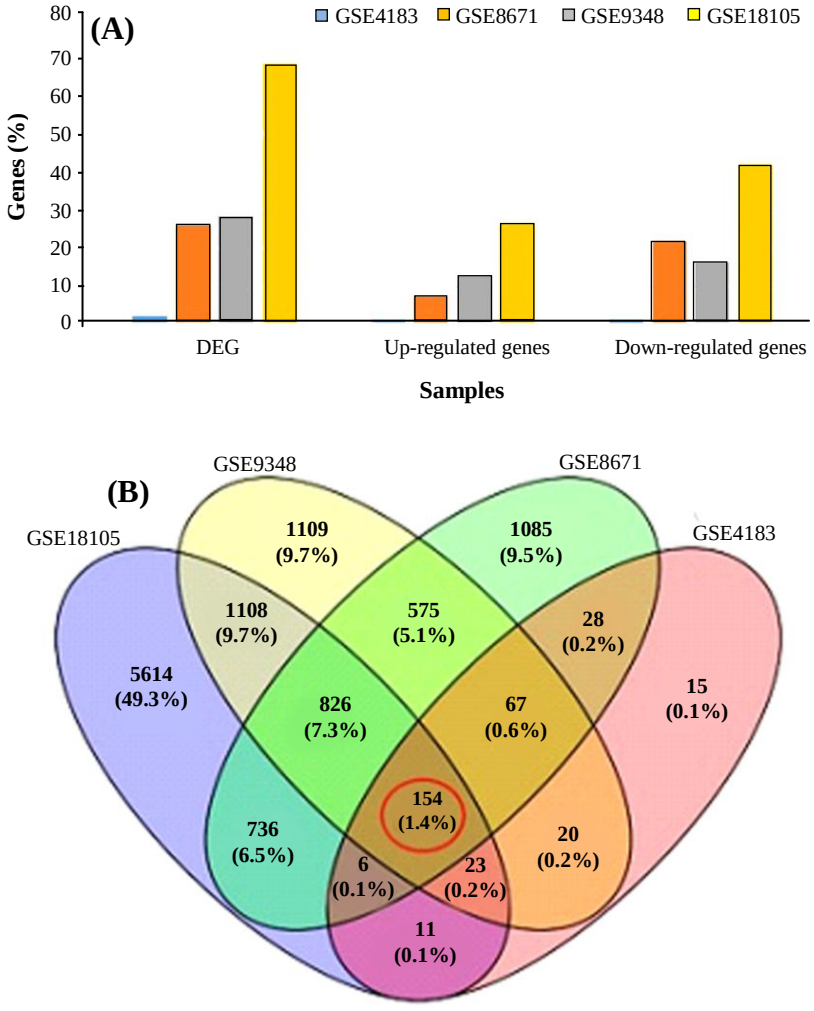


Fig. 1. Differentially expressed genes (DEGs) in colorectal cancer-associated datasets. (A) The distribution of DEGs as well as up- and down-regulated genes in four used CRC experiments. The bars have been arranged to illustrate the percent of genes assigned to each experiment. (B) Venn diagram of intersection among the expression values of DEGs across GSE4183, GSE8671, GSE9348, and GSE18105 series (absolute log fold-change >1, absolute log fold-change <0.5, and adjusted *p* value <0.01). Ultimately, 154 genes were taken as mutual DEGs among these datasets.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

GSE4183, GSE8671, GSE9348, and GSE18105 series, 154 genes were identified as mutual DEGs (Fig. 1B).

As shown in Fig. 2A, gene products of mutual DEGs were enriched in significantly over-represented GO molecular relevance, including cytokine and chemokine processes, in addition to tumor necrosis factor (TNF), oligomerization domain (NOD), and epithelial cell signaling pathways (Fig. 2B).

Networking of CRC-associated datasets

Here, we employed R implementation of GGM as GRN inference algorithm to recognize and remove indirect links between shared DEGs. To this end, we reconstructed two independent GRNs, from 84 normal and 210 cancer cases, but both of GRNs were composed of the same expression values of 154 mutual DEGs across healthy and diseased conditions. To cover all the 154 nodes and more sparsity, we only selected the 1500 highly ranked edges between mutual DEGs (Fig. 3). In order to identify genes with differential connectivity in CRC, by exploiting CytoNCA Cytoscape plugin, 40% of the top globally connected genes (a number of shortest paths with other nodes namely betweenness) from CRC and normal networks were selected separately. Betweenness characteristic, as a centrality measurement, indicates how significant a node would be in healthy and diseased GRNs. Identifying the central nodes by these measures seemingly provides genes that modulate responses to various cellular conditions. From 154 mutual DEGs in each of CRC and healthy GRNs, 61 genes (40%) were selected



Fig. 2. Enrichment analysis of differentially expressed genes (DEGs) in colorectal cancer datasets. (A) Functional classification of biological processes; (B) biological pathways in which mutual DEGs are involved by Enrichr and KEGG databases, respectively with default setting. The bars have been arranged top to down illustrating the number of DCGs, and significance level assigned to each GO molecular terminology and biological pathway.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

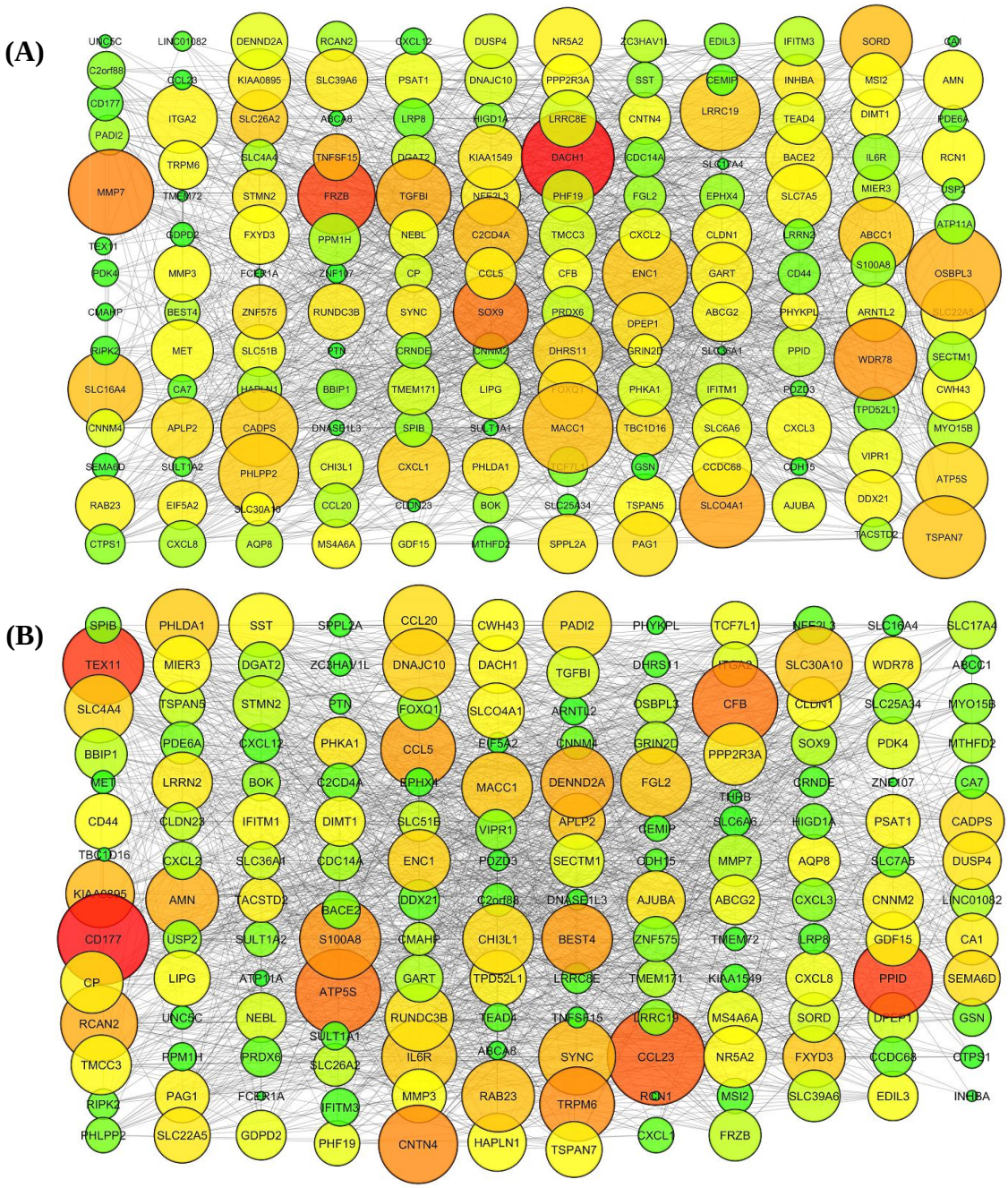


Fig. 3. Reconstructing two independent differential regulatory networks using normal colorectal mucosa and adjacent tumor tissues, by utilizing the expression values of mutual differentially expressed genes (DEGs) across datasets. (A) Cancer and (B) healthy transcriptional regulatory networks derived by GGM algorithm. Using NetworkAnalyzer Cytoscape plug-in, degree and betweenness parameters have been mapped to node size and color so that darker and bigger nodes show higher degree and betweenness centrality.

based on their importance in the network by calculating the betweenness as centrality parameter. From these 61 genes, 21 genes indicated an overlap in two GRNs that were removed from the analysis as we strictly wanted to evaluate the genes that are ranked in the network by connection type variations. Ultimately, 40 genes with more relevance to CRC were selected for further analysis as DCGs.

Filtering the CRC-related candidates across experts-curated databases

For the identification of CRC-related candidates among the DCGs, we intersected the DCGs with two lists of genes: 1572 genes from Network of Cancer Genes (NCG) database[45] and 3265 CRC associated genes from DisGeNET v. 2.0 database[46]. Consequently, *INHBA*, *FOXQ1*, *MET*, *SLC16A4*,

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

SOX9, and *MET* were found to be common genes (Fig. 4A). DisGeNET contains a list of diseases-associated genes collection based on the presence of genetic overlaps between diseases collected from UNIPROT, human CTD, PsyGeNET, Orphanet, and the HPO. From this collection, only genes with at least one evidence from Pfam 31.0 were selected for intersection. Lower coverage of DCGs with NCG genes (17.5%) fuels the efforts in the validation of DCGs associations with CRC by experimentalists. The expression pattern of DCGs was finally checked in 33 transcriptomic experiments via GEMMA database. The expression pattern analysis through several CRC transcriptome datasets exhibited the up-regulation of the majority of DCGs. Expect for *ZNF575*, *STMN2*, *SPPL2A*, *SLC26A2*, *PHYKPL*, *PHLPP2*, *LRRC19*, *FRZB*, *DHRS1*, and *CCDC68* with a relative down-regulation (fold-change <2 at *p* values 0.01-0.005), the rest of DCGs indicated up-regulation (fold-change >2 at *p* value 0.01-0.005), as indicated in Figure 5.

Identifying biological regulators of differentially connected genes in CRC

The complex molecular interactions underlying cancer genesis warrants the identification of biological entities viz. Therefore, inferring regulatory links between TFs, as transcriptional regulators, promisingly will reveal interesting aspects of DCGs. In the next step, we sought potential TFs associated with a circuits of 40 arbitrary DCGs obtained by TRRUST database. Among the DCGs, ABCC1, BACE2, CXCL1, DDX21, ITGA2, MMP7, SLC7A5, SOX9, STMN2, and TGFBI were found to be regulated by 30 TFs, mainly in an activating way. Except for the CXCL1, MMP7, and SOX9 that were regulated by 8, 9, and 10 TFs, respectively (SP1 TF was the common regulator of MMP7 and SOX9), the rest of DCGs were being regulated by distinct TFs. However, SP1, JUN, and SF1 TFs regulated more than one DCG, and the rest of TFs acted as a regulator of just one DCG (Fig. 6).

Expression pattern of selected genes across different cancer types

DCGs were compared to statistically significant DEGs from 12 cancer types (bladder, breast, colon, esophagus, kidney, liver, lung, skin, ovary, prostate, sarcoma, and thyroid cancers) to identify the extent to which DCGs are CRC specific (supplementary S 13-1, 13-12). It would be a strong support if DEGs identified in this study are specific to CRC, then it shows that the computational methods have likely discovered CRC-associated biomarkers correctly. For this purpose, we took the intersection of DEGs obtained from each cancer type and DCGs separately enriched the shared genes to biological pathways (data not shown). Liver and skin (>80%) as well as lung, prostate, sarcoma, and esophagus cancers (<20%) shared the most and the least genes with DCGs, respectively (Fig. 7). The shared genes principally enriched in cytokine and TNF signaling (bladder, breast, colon, kidney, liver, melanoma, and sarcoma), epithelial cell signaling, carbon metabolism (esophagus, lung, prostate, and thyroid), and PI3K-Akt signaling pathway (ovary). As the most cancers shared a few number of genes with DCGs, the finest explanation could be the roughly specific roles that the DCGs play as central nodes in CRC by mediating in PathwAX-derived pathways like extracellular matrix (ECM)-receptor interaction and

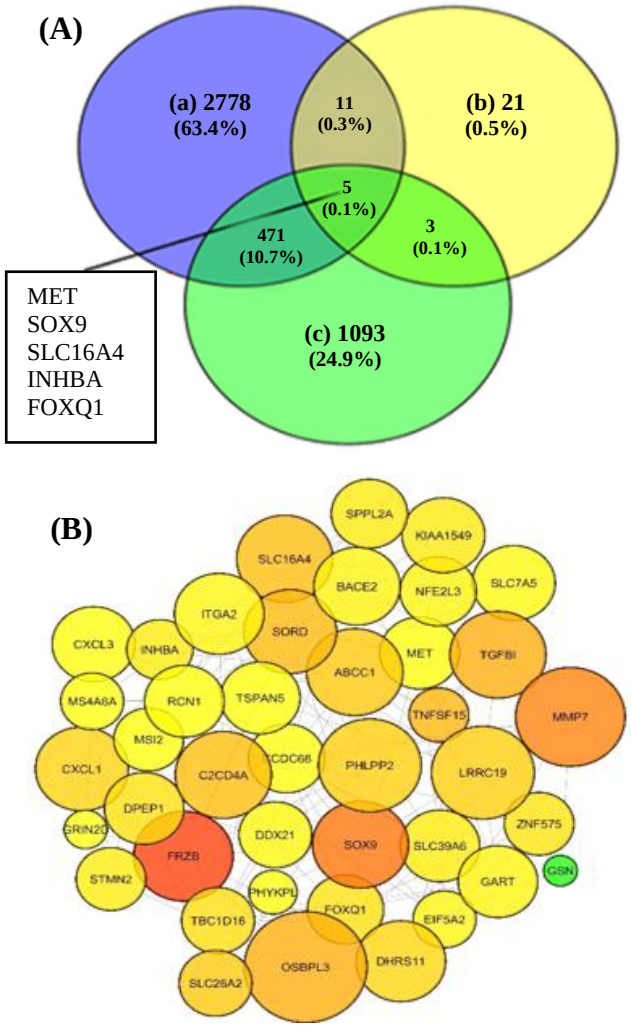


Fig. 4. Characterizing a network of differentially connected genes (DCGs) along with topology feature analysis. (A) Venn diagrams of a, 3265 gene from DisGeNET database; b, DCGs; c, 1572 genes from Network of Cancer Genes (NCG); five genes were taken mutual colorectal cancer-associated genes. (B) Regulatory interactions of DCGs, using NetworkAnalyzer Cytoscape plug-in, degree and betweenness parameters have been mapped to node size and color so that darker and bigger nodes show higher degree and betweenness centrality.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]



Fig. 5. Expression profile of differentially connected genes (DCGs) across 32 colorectal cancer transcriptome datasets by GEMMA database. Dark purple, light purple, dark green, and light green show up-regulation >2 at *p* value 0.005, up-regulation >2 at *p* value 0.01-0.005, down-regulation < 2 at *p* value 0.005, down-regulation >2 at *p* value 0.01-0.005, respectively.

synaptic vesicle cycle (Fig. 8). In sum, *CXCL3*, *SLC7A5*, *SLC26A2*, *GART*, and *CCDC68* genes were differentially expressed in three or more cancer types (Table 1).

Validating the differentially connected genes by power analysis

Since having been differentially expressed in cancer state in comparison to normal tissues, 14 DCGs can represent potential genes for CRC prognosis. We therefore checked the importance of 14 DCGs in CRC progression by plotting Kaplan-Meier survival curves. The survival curves was plotted by feeding DCGs in Kaplan Meier-plotter. As a result, the DCGs were predictive of CRC at *p* value = 0.004 with hazard ratio of 2.81 (Fig. 9).

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

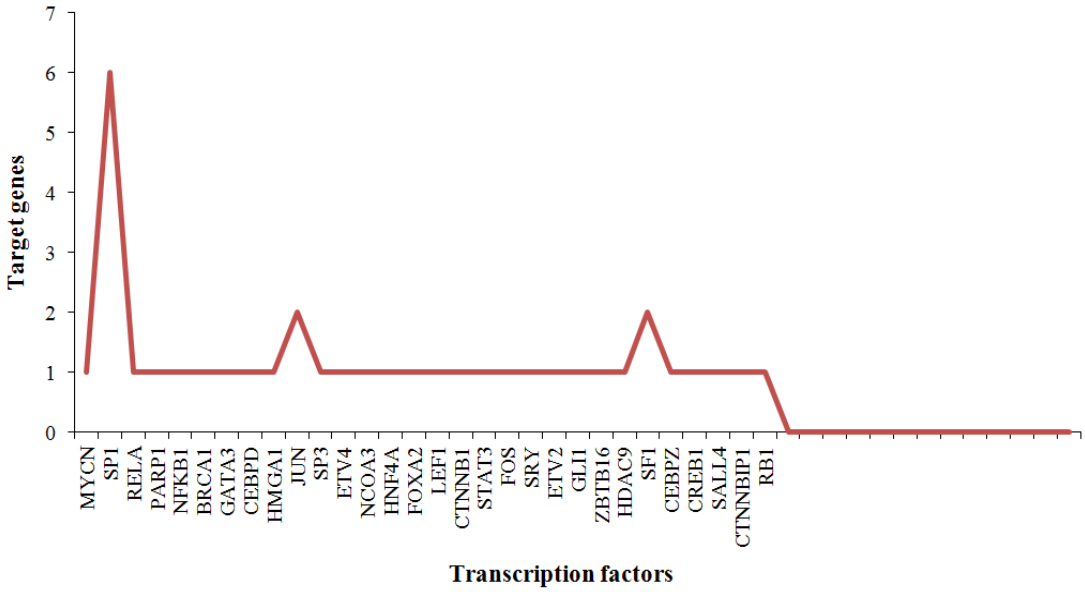


Fig. 6. Potential regulators of differentially connected genes (DCGs). The distribution of transcription factors that modulate the expression of DCGs based on TRRUST database.

DISCUSSION

By capturing the molecular roadmap underlying human diseases, systems biology paves the path to better understanding of diseases mechanism, biomarker identification and drug discovery. CRC is still the major causes of cancer death worldwide; hence, discovering a system of biomarkers triggering the initiation and progression of CRC is a challenging topic in cancer biology[47]. The identification of co-expressed genes related to cancers presumably provides new insights in networks underlying cancers. In other words, a combination of gene effects likely holds promise as a more effective approach for detecting disease-associated genes[48]. In fact, examining co-expressed genes in contrast to the individual genes, could be more informative to explore new biomarkers[49,50]. Hereby, varied correlation between two genes in distinct states such as healthy and diseased conditions is recovered as differential co-expression. As the correlation between two genes may alter free from the expression levels of two genes, transcriptome analysis exclusively based on the differential transcript profiling impedes the structured description of regulatory patterns[51]. Several studies have therefore conducted differential co-expression analysis to facilitate the deconvolution of cellular networks in cancers[52,53]. In the present work, freely available data sources and bioinformatics tools have been exploited to infer DRGs whose interactions hypothetically promote colorectal cancer. Furthermore, attempt was made to disclose the likely relevant molecular pathways implicated by central nodes through a network of these genes. To increase the statistical power, we performed a meta-analysis, combined of multiple Affymetrix experiments. On the other hand, to decrease the experimental specific batch effects, each experiment was processed independently. Of note, we focused on the colon mucosa samples from which nearly all CRC starts. Consequently, 154 statistically significant mutual DEGs across the healthy mucosa and adjacent tumor tissues were extracted that overrepresented with chemokine and cytokine processes, signaling pathways, and transporters. In this

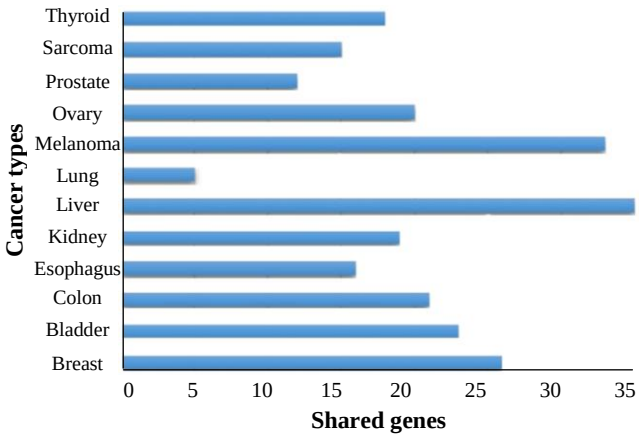


Fig. 7. The expression pattern of differentially connected genes (DCGs) within the statistically significant differentially expressed genes (DEGs) across different cancer types (absolute log fold-change >1, absolute log fold-change <0.5, and adjusted p <0.01). The bars have been arranged to illustrate the number of genes shared between DCGs and each cancer type.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

Pathway class	Significant pathways
Environmental information processing	1
Cellular processes	1
Organismal systems	1



Pathway (enriched/depleted)	FWER	Network connectivity of query genes (pathway gene)
ECM-receptor interaction	2.44e-2	ITGA2 4links; TGFBI 7 links
Cell cycle	3.34e-2	ABCC1 20 links; DDX21 26 links; EIF5A2 16 links; RCN1 3 links; SLC7A5 1 link; GART 7 links
Synaptic vesicle cycle	3.53e-2	ABCC1 23 links; STMN2 3 links; EIF5A2 20 links; SLC7A5 1 link; GART 1 link

Fig. 8. PathwAX results for the differentially connected genes (DCGs) in the colorectal cancer datasets. The Table and Pie chart summarize the pathway distribution. The Table shows enriched (blue) and depleted (red) pathways at *q* value of 0.05 as defined cut-off threshold. Darker shades in colored boxes within the table indicate higher connectivity (links) that a query gene has.

context, we reconstructed two independent GRNs by employing GGM algorithm, using the expression values of 154 mutual DEGs. GGM algorithm produces a high-fidelity representation of the cellular network topology as a graph by recognizing regulatory interactions from non-regulatory interactions and removing non-causal links. Indeed, the casual impact of a TF on its target genes is being inferred. The critical idea behind this algorithms is the modeling of partial correlation as a measure of independence of any two genes. The assumption of inferring gene network using GGM algorithm is that the selected 154 genes are in the same pathway (or network), and they are interacting and regulating each other. However, based on the definition of DEG, these 154 genes are not necessarily related to each other. The expression correlation observed may be due to indirect regulation. In this context, the betweenness analysis will highlight genes that being regulated by the most number of other genes, and those genes with least number of connections are true important regulators. Then two inferred GRNs were topologically analyzed to find DCGs. To achieve this, 61 genes, by in rank ordering of betweennees centrality scores, were selected from which 21 genes were shared through normal and cancer GRNs that removed from further investigation. Finally, 40 genes were selected as potential key connectors considered as DCGs specific to cancer samples. Within a network, DCGs, FRZB, SOX9, MMP7, and WDR78 were ordered as the highest strongly connected genes; all up-regulated across different CRC transcriptome experiments in GEMMA database (Fig. 5). Pathway annotation is usually performed by taking overlap of a gene set with a pathway that increases false positives and false negatives[30]. However, PathwAX web server in addition to network crosstalk enrichment can perform depletion analysis. A significant depleted pathway suggests that the links between genes is not much significant to be affected by a certain pathway. Here, DCGs were enriched with ECM-receptor interactionby ITGA2 and TGFB1 and synaptic vesicle cycle by GART, ABCC1, *STMN2*, *EIF5A2* and *SLC7A5*, while cell cycle pathway depleted significantly at *q* value of 0.05 (Fig. 8). ECM-

Table 1. The expression pattern of DCGs across different cancer types

Shared gene	Cancer type	Expression pattern	Fold-change
CXCL3	Breast	Up	< 2
SLC7A5	Bladder	Up	< 2
SLC26A2	Sarcoma	Down	>2
	Thyroid		
	Prostate		
	Ovary	Up	< 2
GART	Melanoma		
	Lung		
	Liver		>2
CCDC68	Kidney	Down	
	Esophagus		
	Colon		

Up, upregulation; Down, downregulation

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

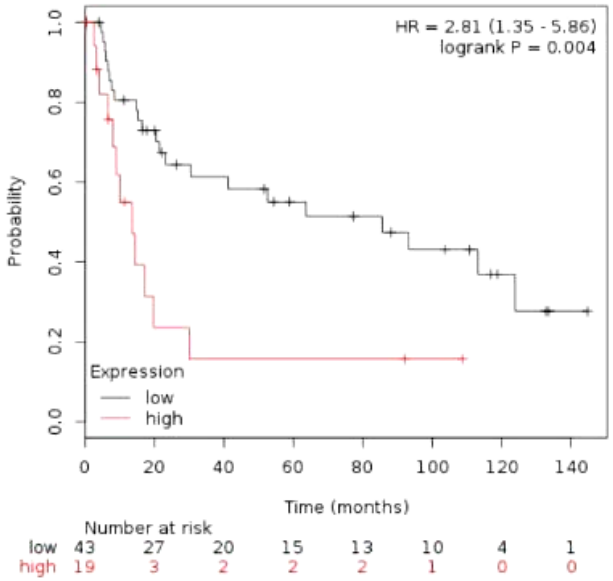


Fig. 9. Power analysis of differentially connected genes (DCGs) in colorectal cancer datasets and Kaplan-Meier analysis of colorectal cancer dataset using the DCGs. The *p* values are computed using log-rank.

receptor proteins have been dysregulated in the progression from an isolated tumor to metastatic phase[54] and shown to be related to CRC[55]. This is likely emphasizing the role of DCGs in CRC invasion by disseminating the tumor to secondary sites of body. Screening the DCGs across 12 cancer types (bladder, breast, colon, esophagus, kidney, liver, lung, skin, ovary, prostate, sarcoma, and thyroid cancers), suggested that these genes could be speculated as metastatic-related genes as DCGs-shared GART and SLC7A5 with 12 tumor types. The over-representation of DCGs across multiple cancer types with cytokines suggests the explosion of inflammatory events as a similarity among these cancers. DEGs extracted from 12 cancers types shares pathways with DEGs such as cytokines, metabolism, and signaling pathways. Therefore, DCGs likewise are oriented to distinct cascade running the CRC metastasis, while cell cycle is depleted with DCGs. Thus, the specific roles of DCGs in CRC seemto be triggering the metastasis in contrast to their common roles as mediators in TNF, epithelial cell signaling, metabolism, and transportations that was shared with different cancer types.

To evaluate whether the DCGs have any relevance to diseases, we obtained an intersection of DCGs with DisGeNET genes, thereof 11 DGCs (25%) were found to be shared with CRC-related genes in DisGeNET. In fact, this amount of similarity highlights the DCGs with higher ranks in centrality measure, which potentially play remarkable roles in CRC compared to the other genes. However, the low amount of coverage of DCGs with experimentally validated NCG suggests potential unknown genes related to the CRC as targets for future studies. The DCGs were regulated with distinct TFs from which SP1, SF1, and JUN were found to regulate the majority of DCGs. MMP7, SOX9, and CXCL1 found to be regulated by diverse TFs with SP1 as a common regulator (Fig. 6A). This observation may imply that TFs play specific roles in modulating highly interconnected nodes in CRC network. MMP7 and SOX9 are shown to be nodes with the highest connectivity in DCGs (Fig. 4B). MMP family has proven to act in metastatic phase by degrading ECM structures, thus paving the way for the cells through the dense environment[56,57]. Interestingly, SPA1 was the regulator of DCGs implicated in EMC-receptor like ITGA2, ABCC1, and SLC7A5. Integrins, including TGA5, ITGA5, ITGB5, ITGA11, and ITGBL1 elevated in cancer tissues. In accordance with our study, ITGA2 showed up-regulation of 2> fold. Sp1 has been acknowledged to enhance or repress gene expression that in turn plays pivotal roles in metastasis of various tumors[58,59]. Keeping with this analysis, SP1 is likely implicated in CRC progression by regulating EMC components like SLC7A5. It has also been recovered as mutual differentially expressed in 12 screened cancers.

We aimed to delineate prognostic biomarkers underlying CRC; therefore, in the frame of *in silico* analysis, a certain number of genes were explored whose reciprocal interplays are supposedly associated with CRC. Taken together, interactions of these genes which majorly occurred through the metastatic cascade could be considered as the mediators of CRC aggression. Indeed, the identifying these genes exhibits the importance of network topology analysis to rank more important genes as disease-related biomarkers against a set of exclusively DEGs in a meaningful way. These ranks measure the relative importance of a protein in a biological network and could identify strongly correlated genes with specific states. However, this analysis is challenged by the disadvantage of inevitable overestimation in computational approaches; thus, applying more stringent parameters in predicting the regulatory links would be apparently helpful in acquiring more reliable results and overcoming any inaccuracy coming from the nature of reverse engineering methods. Moreover, we employed unweighted networks during GRN reconstruction and topology analysis. We then should be cautious about dynamic nature of cancers via strictly analysis of statics networks.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

The main goal of this analysis was exploiting differential connectivity that is thought to rank the influential genes in the pathogenesis of CRC. Utilizing gene expression data with pooling information of TFs in cancers can help to discover crucial findings to identify underlying mechanisms and enlighten more molecular underpinnings of different cancers. We observed that the identified genes and TFs are mainly guided to cytokine signaling pathway and metabolism implicated in CRC. To summarize, selected genes viz. *GART*, *TGFB1*, *ITGA2*, *SLC16A5*, *SOX9*, and *MMP7* with differential connectivity across normal and CRC samples along with SP1, SF1, and JUN TFs could be taken into account for future detection and therapeutic targets by experimental investments.

ACKNOWLEDGEMENTS

We thank Dr. Nooshin Omranian, a scientific staff in Systems Biology and Mathematical Modelling Group, Max Planck Institute for Molecular Plant Physiology, Potsdam (Germany) for her precious guidance.

CONFLICT OF INTEREST. None declared.

REFERENCES

1. Quan L, Stassen APM, Ruivenkamp CAL, Wezel T, Fijneman RJA, Hutson A, Kakarlapudi N, Hart AAM, Demant P. Most lung and colon cancer susceptibility genes are pair-wise linked in mice, humans and rats. *PLoS one* 2011; **6**(2): e14727.
2. Grade M, Hummon AB, Camps J, Emons G, Spitzner M, Gaedcke J, Hoermann P, Ebner R, Becker H, Difilippantonio MJ, Ghadimi BM, Beissbarth T, Caplen NJ, Ried T. A genomic strategy for the functional validation of colorectal cancer genes identifies potential therapeutic targets. *International journal of cancer* 2011; **128**(5): 1069-1079.
3. Savas S, Hyde A, Stuckless SN, Parfrey P, Younghusband HB, Green R. Serotonin transporter gene (SLC6A4) variations are associated with poor survival in colorectal cancer patients. *PLoS one* 2012; **7**(7): e38953.
4. Wu Y, Wang X, Wu F, Huang R, Xue F, Liang G, Tao M, Cai P, Huang Y. Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing. *PLoS one* 2012; **7**(8): e41001.
5. Viana LS, Affonso RJ Jr, Silva SR, Denadai MV, Matos D, Salinas de Souza C, Waisberg J. Relationship between the expression of the extracellular matrix genes SPARC, SPP1, FN1, ITGA5 and ITGAV and clinicopathological parameters of tumor progression and colorectal cancer dissemination. *Oncology* 2013; **84**(2): 81-91.
6. Zhang Z, Wang A, Li H, Zhi H, Lu F. STAT3-dependent TXNDC17 expression mediates Taxol resistance through inducing autophagy in human colorectal cancer cells. *Gene* 2016; **584**(1): 75-82.
7. Balkwill F. Tumour necrosis factor and cancer. *Nature reviews cancer* 2009; **9**(5): 361-371.
8. Sells E, Pandey R, Chen H, Skovan BA, Cui H, Ignatenko NA. (2017). Specific microRNA–mRNA regulatory network of colon cancer invasion mediated by tissue kallikrein-related peptidase 6. *Neoplasia* 2017; **19**(5): 396-411.
9. Wang H, Luo J, Liu C, Niu H, Wang J, Liu Q, Zhao Z, Xu H, Ding Y, Sun J, Zhang Q. Investigating MicroRNA and transcription factor co-regulatory networks in colorectal cancer. *BMC bioinformatics* 2017; **18**: 388.
10. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: the potential of gene co-expression networks for investigating molecular mechanisms of complex brain disorders. *Genes, brain and behavior* 2014; **13**(1): 13-24.
11. Yang J, Yu H, Liu BH, Zhao Z, Liu L, Ma LX, Li YX, Li YY. DCGL v2.0: an R package for unveiling differential regulation from differential co-expression. *PLoS one* 2013; **8**(11): e79729.
12. Liu BH, Yu H, Tu K, Li C, Li YX, Li YY. DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. *Bioinformatics* 2010; **26**(20): 2637-2638.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008; **9**: 559.
14. Wang T, Ren T, Ding Y, Fang Z, Sun Z, MacDonald ML, Sweet RA, Wang J, Chen W. FastGGM: an efficient algorithm for the inference of gaussian graphical model in biological networks. *PLoS computational biology* 2012; **12**(2): e1004755.
15. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* 2006; **7**(Suppl): S7.
16. Faith FF, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS biology* 2007; **5**(1): e8.
17. Song WM, Zhang B. Multiscale embedded gene co-expression network analysis. *PLoS computational biology* 2015; **11**(11): e1004574.
18. Gyorffy B, Molnar B, Lage H, Szallasi Z, Eklund AC. Evaluation of microarray preprocessing algorithms based on concordance with RT-PCR in clinical samples. *PLoS one* 2009; **4**(5): e5645.
19. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, Laczko E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Gomes V, Pastorelli A, Faggiani R, Anti M, Jiricny J,

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

Clevers H, Marra G. Transcriptome profile of human colorectal adenomas. *Molecular cancer research* **5**(12): 1263-1275.
20. Hong Y, Downey T, Eu KW, Koh PK, Cheah PY. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clinical and experimental metastasis* 2010; **27**(2): 83-90.
21. Matsuyama T, Ishikawa T, Mogushi K, Yoshida T, Lida S, Uetake H, Mizushima H, Tanaka H, Sugihara K. MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. *International journal of cancer* 2010; **127**(10): 2292-2299.
22. Bolstad BM, Irizarry RA, Astrand M, Speed TPA. Comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19**(2): 185-193.
23. Smyth GK. Limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. eds. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health. New York: Springer; 2005.
24. Schäfer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 2005; **21**(6): 754-764.
25. Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* 2015; **127**: 67-72.
26. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology* 2007; **3**(4): e59.
27. Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, Kim H, Kim K, Yang S, Bae D, Yun A, Kim S, Kim CY, Cho HJ, Kang B, Shin S, Lee I. TRRUST: a reference database of human transcriptional regulatory interactions. *Scientific reports* 2015; 5: 11432.
28. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research* 2014; **42**(Database issue): 199-205.
29. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* 2016; **44**(W1): W90-W97.
30. Ogris C, Helleday T, Sonnhammer ELL. PathwAX: a web server for network crosstalk based pathway annotation. *Nucleic acids research* 2016; **44**(Web Server issue): W105-W109.
31. Zoubarev A, Hamer KM, Keshav KD, McCarthy EL, Santos JR,. Van Rossum T, McDonald C, Hall A, Wan X, Lim R, Gillis J, Pavlidis P. Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics* 2012; **28**(17): 2272-2273.
32. Stathopoulos GP, Armakolas A. Differences in gene expression between individuals with multiple primary and single primary malignancies. *International journal of molecular medicine* 2009; **24**(5): 613-622.
33. Graham K, de las Morenas A, Tripathi A, King C, Kavanah M, Mendez J, Stone M, Slama J, Miller M, Antoine G, Willers H, Sebastiani P, Rosenberg CL. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *British journal of cancer* 2010; **102**(8): 1284-1293.
34. Ågesen TH, Berg M, Clancy T, Thiis-Evensen E, Cekaite L, Lind GE, Nesland JM, Bakka A, Mala T, Hauss HJ, Fetveit T, Vatn MH, Hovig E, Nesbakken A, Lothe RA, Skotheim RI. CLC and IFNAR1 are differentially expressed and a global immunity score is distinct between early- and late-onset colorectal cancer. *Genes and immunity* 2011; **12**(8): 653-662.
35. di Pietro M, Lao-Sirieix P, Boyle S, Cassidy A, Castillo D, Saadi A, Eskeland R, Fitzgerald RC. Evidence for a functional role of epigenetically regulated midcluster HOXB genes in the development of Barrett esophagus. *Proceedings of the national academy of sciences of the United States of America* 2012; **109**(23):9077-9082.
36. Neusser MA, Lindenmeyer MT, Moll AG, Segerer S, Edenhofer I, Sen K, Stiehl DP, Kretzler M, Gröne HJ, Schlöndorff D, Cohen CD. Human nephrosclerosis triggers a hypoxia-related glomerulopathy. *The American journal of pathology*; **176**(2): 594-607.
37. Shi M, Chen MS, Sekar K, Tan CK, Ooi LL, Hui KM. A blood-based three-gene signature for the non-invasive detection of early human hepatocellular carcinoma. *European journal of cancer* 2014; **50**(5):928-936.
38. Sato T, Kaneda A, Tsuji S, Isagawa T, Yamamoto S, Fujita T, Yamanaka R, Tanaka Y, Nukiwa T, Marquez VE, Ishikawa Y, Ichinose M, Aburatani H. PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. *Scientific reports* 2013; **3**: 1911.
39. Critchley-Thorne RJ, Yan N, Nacu S, Weber J, Holmes SP, Lee PP. Down-regulation of the interferon signaling pathway in T lymphocytes from patients with metastatic melanoma. *PLoS medicine* 2017; **4**(5):e176.
40. Bowen NJ, Walker LD, Matyunina LV, Logani S, Totten KA, Benigno BB, McDonald JF. Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC medical genomics* 2009; **2**: 71.
41. Satake H, Tamura K, Furihata M, Anchi T, Sakoda H, Kawada C, Liyama T, Ashida S, Shuin T. The ubiquitin-like molecule interferon-stimulated gene 15 is overexpressed in human prostate cancer. *Oncology reports* 2010; **23**(1): 11-16.
42. Detwiller KY, Fernando NT, Segal NH, Ryeom SW, D'Amore PA, Yoon SS. Analysis of hypoxia-related gene expression in sarcomas and effect of hypoxia on RNA interference of vascular endothelial cell growth

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]

factor A. *Cancer research* 2005; **65**(13): 5881-5889.
43. Pita JM, Figueiredo IF, Moura MM, Leite V, Cavaco BM. Cell cycle deregulation and TP53 and RAS mutations are major events in poorly differentiated and undifferentiated thyroid carcinomas. *The journal of clinical endocrinology and metabolism* 2014; **99**(3): E497-507.
44. Szász AM, Lánczky A, Nagy Á, Förster S, Hark K, Green JE, Boussioutas A, Busuttil R, Szabó A, Győrffy B. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* 2016; **7**(31): 49322-49333.
45. Omer A, Dall'Olio GM, Mourikis TP, Ciccarelli FD. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic acids research* 2016; **44**(Database issue): D992-D999.
46. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* 2015; **2015**: bav028.
47. Izadi F, Zamanian-Azodi M, Mansouri V, Khodadoostan M, Naderi N. Exploring conserved mRNA-miRNA interactions in colon and lung cancers. *Gastroenterology and hepatology from bed to bench* 2017; **10**(3): 184-193.
48. Xu Y, Yue W, Shugart YY, Li S, Cai L, Li Q, Cheng Z, Wang G, Zhou Z, Jin C, Yuan J, Tian L, Wang J, Zhang K, Zhang K, Liu S, Song Y, Zhang F. Exploring Transcription Factors-microRNAs Co-regulation Networks in Schizophrenia. *Schizophrenia bulletin* 2016; **42**(4): 1037-1045.
49. Okamura Y, Aoki Y, Obayashi T, Tadaka S, Ito S, Narise T, Kinoshita K. COXPRESdb in 2015: co-expression database for animal species by DNA microarray and RNA-Seq-based expression data with multiple quality assessment systems. *Nucleic acids research* 2014; **43**(Database issue): D82-D86.
50. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene co-expression networks in human and chimpanzee brains. *Proceedings of the national academy of sciences of the United States of America* 2006; **103**(47): 17973-17978.
51. de la Fuente A. From 'differential expression' to 'differential networking'- identification of dysfunctional regulatory networks in diseases. Trends in genetics 2010; **26**(7): 326-333.
52. Lin CC, Mitra R, Cheng F, Zhao Z. Cross-cancer differential co-expression network reveals microRNA-regulated oncogenic functional modules. *Molecular biosystems* 2015; **11**(12): 3244-3252.
53. Li J, Li YX, Li YY. Differential regulatory analysis based on coexpression network in cancer research. *BioMed research international* 2016; **2016**: 4241293.
54. Venning FA, Wullkopf L, Erler JT. Targeting ECM disrupts cancer progression. *Frontiers in oncology* 2015; **5**: 224.
55. Chen S, Wang Y, Zhang L, Yinan S, Zhang M, Wang J, Zhang X. Exploration of the mechanism of colorectal cancer metastasis using microarray analysis. *Oncology letters* 2017; **14**(6): 6671-6677.
56. Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. *Annual review of cell and developmental biology* 2001; **17**: 463-516.
57. Mook OR, Frederiks WM, Van Noorden CJ The role of gelatinases in colorectal cancer progression and metastasis. *Biochim biophys acta* 2004; **1705**(2): 69-89.
58. Beishline K, Azizkhan-Clifford J. Sp1 and the 'hallmarks of cancer. *FEBS journal* 2015; **282**(2): 224-258.
59. Vizcaíno C, Mansilla S, Portugal J. Sp1 transcription factor: A long-standing target in cancer chemotherapy. *Pharmacology and therapeutics* 2015; **152**: 111-124.

[Downloaded from ibj.pasteur.ac.ir on 2024-03-28]
[DOR: 20.1001.1.1028852.2019.23.1.5.5]
[DOI: 10.29252/ibj.23.1.34]