Systems Biomedicine of Colorectal Cancer Reveals Potential Targets for CRC Treatment.

Mehran Piran[1,5*], Mehrdad Piran[2,3], Neda Sepahi[4,5], Amir Rahimi[1*]

[1] Bioinformatics and Computational Biology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

[2] Department of Tissue Engineering and Applied Cell Sciences, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[3] Department of Biology, East Tehran Branch, Islamic Azad University, Tehran, Iran

[4] Department of Medical Biotechnology, Fasa University of Medical Sciences, Fasa, Iran

[5] Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran

Abstract

Colorectal cancer is one of the leading causes of cancer death worldwide. Patients' survival at the time of diagnosis depends largely on the stage of the tumor. Therefore, understanding the molecular mechanisms promoting cancer progression from early stages to high-grade stages is essential for implementing therapeutic approaches. To this end, we performed a unique meta-analysis flowchart by identifying differentially expressed genes (DEGs) in some test datasets verified by the validation datasets. DEGs were employed to construct a protein-protein interaction network (PPI). Then, a smaller network was extracted from the PPI network based on a shortest-path scoring algorithm. 12 genes were recognized that their expression is different between benign and metastatic tumors. Utilizing a network analysis point of view 27 genes around them were selected. Among the total number of 39 identified DEGs, we proposed a number of them that their expression induction or suppression alone or in combination with each other would inhibit tumor progression or metastasis. Furthermore, expressions of a number of them were presented as diagnostic biomarkers for colorectal cancer. They are involved in cell proliferation, energy production under hypoxic conditions, epithelial to mesenchymal transition (EMT) and angiogenesis. Finally, some

genes were discovered that have not been reported noticeably in any kind of cancer neither as a biomarker nor curative target. As a result, the present transcriptomic meta-analysis flowchart would help to identify potential molecular targets for treatment of not only colorectal cancer but also different types of cancer.

1. Introduction

Colorectal cancer (CRC) is a major global medical burden worldwide [1]. Approximately more than one million people are diagnosed with CRC each year, and about half of them die of CRC annually [2]. Complex genetic interactions are combined with environmental factors to trigger a cell to become cancerous. Among them, aberrant growth factor signals contribute to the uncontrolled proliferation of cells which ultimately leads to metastasis. Contrary to early-stage tumor cells, malignant cells have the ability to detach from the stroma as well as acquire motility [3]. This event happens during a process called EMT in which cells lose their epithelial characteristic including adhesion and subsequently dedifferentiate into mesenchymal mobile cells [4]. Genes that are differentially expressed (DEGs) between benign and malignant tumors in gene expression datasets, would be the ones that their effects cause different steps of metastasis.

A large number of molecular and pathway targets have been identified for treatment of CRC during the past decades. Growing progresses have been made in the development of chemotherapy and antibody drugs [5]. The first priority in the therapy procedure is to determine whether tumor is primary or metastatic. Besides, it is required to follow a standard procedure to determine mutation status for targeted therapies. For instance, mutations in RAS/RAF/MAPK pathway lead to resistance to anti-epidermal growth factor receptor antibodies [6].

Tyrosine kinases (TKs) targeting using monoclonal antibodies and small molecule tyrosine kinase inhibitors are effective strategies [7]. Targeting cancer-related inflammation biomarkers like IL-6/JAK/STAT3 pathway which inhibits the progression of solid tumors is

another beneficial therapeutic strategy [8]. In addition, restraint of cytosolic β-catenin via disturbing hyperactive Wnt/β-catenin signaling pathway could be another treatment approach for colorectal and many other types of cancer [9]. Inhibition of matrix metalloproteinases (MMPs) and TGFβ signaling pathways is a therapeutic approach to prevent liver metastasis [10] [11, 12]. Furthermore, PI3K inhibition suppresses lung metastasis in CRC patients [13, 14]. Among the known anticancer drugs, Cetuximab is one of the popular ones which is a monoclonal antibody against epidermal growth factor receptor (EGFR) [15]. Furthermore, Vascular Endothelial Growth Factor (VEGF) antibody, bevacizumab, is the standard treatment for metastatic colorectal cancer [16].

One way to identify molecular mechanism of pathogenesis in a biological context is to analyze transcriptomic data. In this study, we conducted a unique meta-analysis flowchart which contained two datasets for retrieving differentially expressed genes considered as the test set and five datasets to validate selected DEGs considered as the validation set. The test datasets were constructed from two microarray studies each containing three to four groups of samples and DEGs were excavated in any two-pair comparison between the groups in each dataset. Common DEGs between the analyses of two datasets were regarded as the final DEGs. Datasets containing appropriate high-quality samples were chosen and many restrict rules were practiced on the datasets for finding real DEGs. As a result, with three datasets, no common DEGs were found with the desired log-fold change and p-value thresholds. To compensate for the small number of datasets, five RNAseq and microarray datasets were employed from different databases to validate selected DEGs. We obtained some DEGs involved in cancer progression that their expression could be targeted (suppressed or induced) individually or in combination with one another for CRC treatment. To this purpose, a network analysis method with focus on shortest path scoring system was employed to retrieve a small network containing the potential genetic targets. Moreover, expressions of some genes were proposed to be CRC biomarkers.

2. Methods

2.1 Database Searching and recognizing pertinent experiments

Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) repositories were searched to detect test set experiments containing high-quality transcriptomic samples concordance to the study design. Colorectal/colon cancer, EMT and metastasis are the keywords utilized in the search and it was filtered for Homo sapiens. Microarray raw data with accession numbers GSE18105 and GSE41258 were downloaded from GEO and ArrayExpress databases respectively. In the first study (GSE18105), all samples were from Metastatic (Stage III and IV of the disease), Primary (non-metastatic, stage I and II of the disease) and normal colon tissue. In the second experiment (GSE41258) normal and primary samples (samples in stage I and II of cancer) were selected from colon tissue while metastatic ones were selected from liver and lung metastasis tissues in CRC patients. In the validation set, a dataset was constructed from count RNAseq files in The Cancer Genome Atlas (TCGA) database (TCGA dataset). Three RNAseq datasets were selected from GEO (GSE50760, GSE144259, GSE89393). Moreover, a validation dataset was constructed from samples in GSE41258 study which were not present in the test dataset.

2.2 Identifying Differential Expressed Genes in Microarray Datasets

R software was used to import and analyze data for each dataset separately. Preprocessing step involving background correction and probe summarization was done using RMA method in "affy" package [17]. Absent probesets were identified using "mas5calls" function in this package. If a probeset contained more than two absent values in each group of samples, that probeset was regarded as absent and removed from the expression matrix. Besides, Outlier samples were identified and removed using PCA and hierarchical clustering methods. Next, data were normalized using Quantile normalization approach [18]. Then, standard deviation (SD) for each gene was computed and SDs median was used as a cut off to remove low variant genes. Therefore, low-variant genes can no longer influence the significance of the high-variant genes. Many to Many problem [19] which is mapping multiple probesets to the same gene symbol was solved using nsFilter function in "genefilter"

package [20]. This function selects the probeset with the highest Interquartile range (IQR) to map to the gene symbol. “LIMMA” R package which applies linear models on the expression matrix was utilized to discover DEGs between three groups of samples [21]. Genes with absolute log fold change larger than 0.5 and Benjamini Hochberg adjusted p-value [22] less than 0.05 were selected as the DEGs.

2.3 Identifying Differential Expressed Genes in RNAseq Datasets

Count files for all metastatic carcinoma and adenocarcinoma CRC samples (four files) and five primary samples containing more than 90 percent tumor cells as well as five normal samples involving 100 percent normal cells were downloaded from TCGA database. They were imported into R and merged with each other to construct the TCGA dataset. Then, using “Deseq2” R package [23] data were normalized and DEGs were identified between the groups. For RNAseq datasets in GEO, FPKM normalized data were downloaded and imported into R. data were log-2 transformed and using “LIMMA” package in R, DEGs were identified between the groups.

2.4 Network Construction

STRING database was used to generate Protein-Protein Interaction (PPI) network from all final DEGs. Using “igraph” package in R software [24], the giant component of the weighted network was extracted from the whole network. Next, the weighted adjacency matrix was transformed into a symmetric matrix to get modified into a new adjacency matrix using topological overlapping measure (TOM) function in “WGCNA” R package [25]. To remain with a distance matrix, the new adjacency matrix was subtracted from one.

2.5 Neighbourhood Ranking to the Seed Genes

Using R software, a matrix of all shortest paths between all pairs of nodes in a weighted network was constructed using using Dijkstra algorithm [26, 27]. Next, a distance score, Dj, for each node in the PPI network was computed. Dj is the average of the shortest paths from

all the non-seed genes to reach the node j subtracted from the average of the shortest paths from the seed genes to reach the node j divided by the average of the all shortest paths to reach the node j from the whole network. This scoring system implies how much close each node is to the seed nodes [27, 28].

$$D_j = \frac{\frac{\sum_{i \notin S} SP_{ij}}{|NS|} - \frac{\sum_{i \in S} SP_{ij}}{|S|}}{\frac{\sum_i SP_{ij}}{|S| + |NS|}}$$

Here S is the set of seed nodes and NS is the set of non-seed nodes. A positive score implies that node j is closer to the seed nodes compare to the remaining of the nodes. Nodes with positive scores were kept and the rest were removed from the network. It should be noted that the D scores were calculated without imposing any threshold on edge weights.

2.6 Enrichment Analysis

Enrichment analysis was performed using Enrichr online tool [29]. Enriched terms for biological processes were obtained from *GO* repository. For pathway enrichment analysis, *wikiPathways* signaling repository version 2019 for humans was used. Enriched terms with the top score and the p-value less than 0.05 were selected.

3. Results

3.1 Data Preprocessing in Test Datasets

Each dataset was imported into R separately. Almost 75% of probesets were regarded as absent and left out from the expression matrix to reduce the number of false positives in multiple testing. To be more precise in preprocessing step, outlier sample detection was conducted using PCA and hierarchical clustering approaches. Figure1A illustrates the PCA plot for the samples in GSE18105 study. The same plot was created for the GSE41258 dataset. Between the three groups of samples, a number of them are separated from their sets along the PC1 axis and they should be considered as the outliers. To be more specific, a

hierarchical clustering method introduced by Oldham MC, et al [30] was used. Pearson correlation coefficients between samples were subtracted from one for measurement of the distances. Figure 1B depicts the dendrogram for the normal samples. In figure 1C normal samples are plotted based on their Number-SD scores. To get this number for each sample, the average of whole distances is subtracted from the distance average of each sample. Then, results of these subtractions are normalized by the standard deviation of distance averages [30]. Samples with Number-SD less than negative two are usually far apart from their cluster set in the PCA plane and they were regarded as outliers. Fourteen outlier samples for GSE18105 dataset and sixteen outlier samples for GSE41258 test dataset and one for GSE41258 validation dataset were found. Supplementary file1 contains information about the groups of samples.

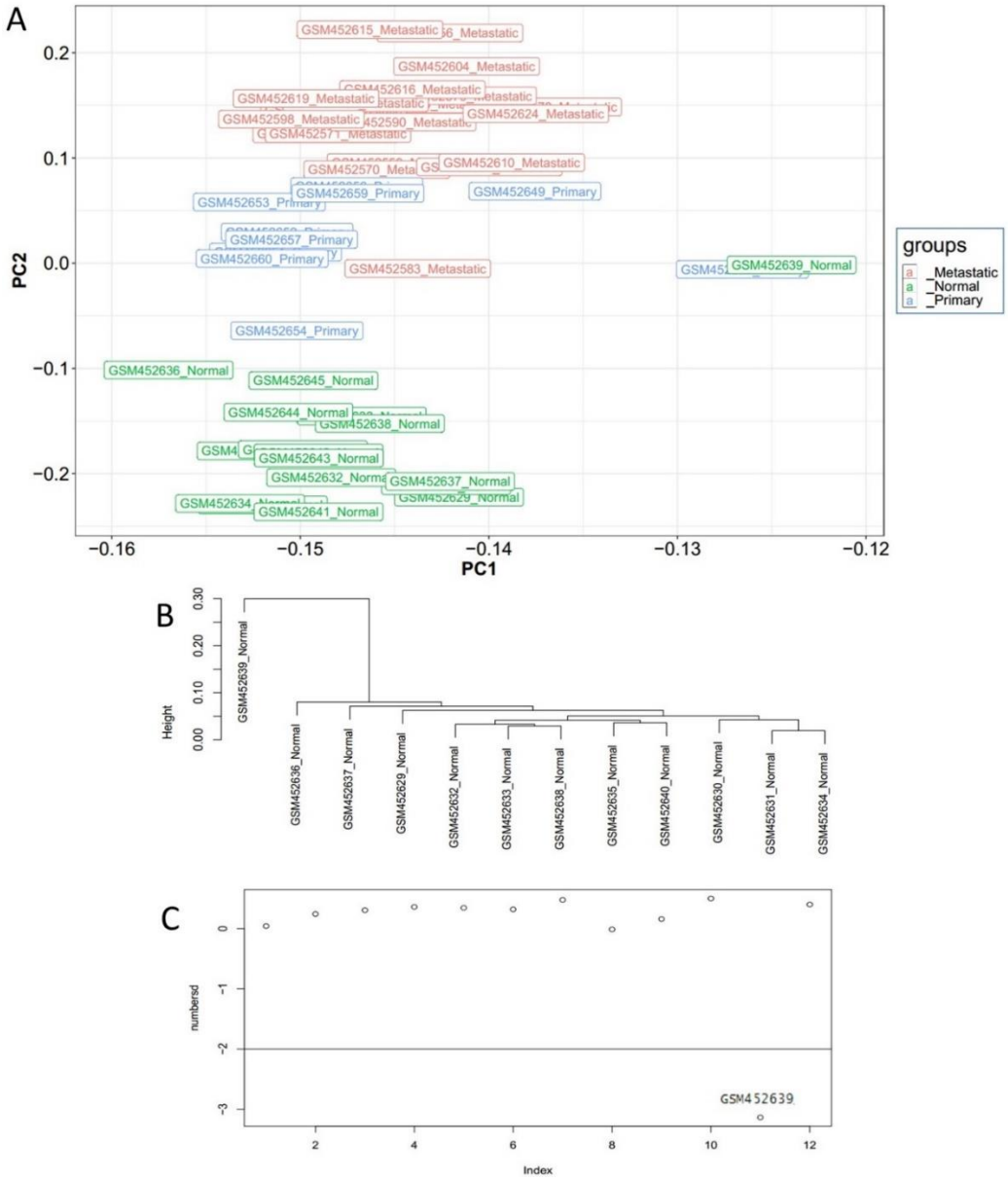

Figure 1: Illustration of outlier samples in GSE18105 study. A, is the PCA plot, B is the dendrogram for the normal samples and C is the Number-SD plot for the normal samples. GSM452639 is the outlier sample in normal group which is illustrated by the three plots.

Figure2 illustrates the average expression values for some housekeeping genes and DEGs between Primary and Metastatic colon samples. A, illustrates DEGs with absolute log fold change (LogFC) larger than 2 in GSE18105 dataset. Common DEGs between lung-primary analysis and liver-primary analysis (refer to the next session) with absolute LogFC larger than 1 in GSE41258 dataset are illustrated in B as well. Housekeeping genes are situated on

the diagonal of the plots whilst DEGs are located above or under the diagonal. This demonstrates that the preprocessed dataset is of sufficient quality for the analysis. SPP1 gene is not DEG in metastatic colon samples while it had a high LogFC in metastatic liver and metastatic lung samples comparing to the primary group, 2.6 and 3.2 respectively.

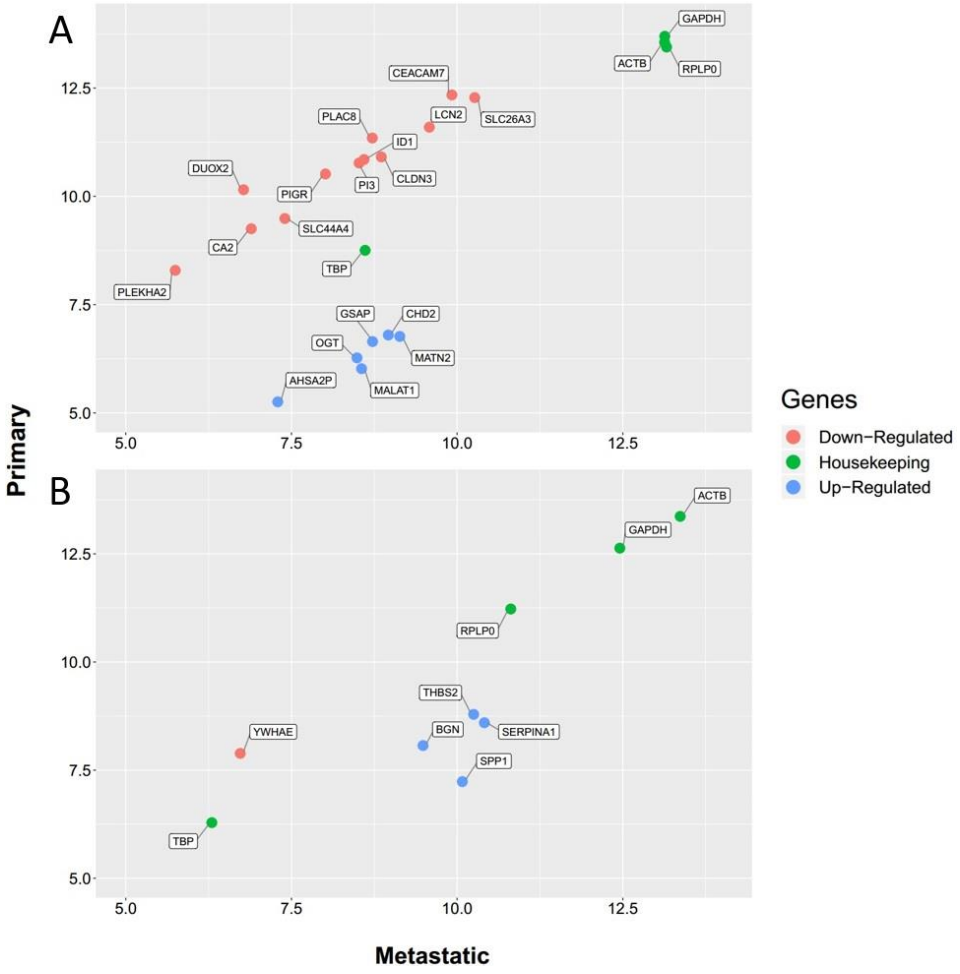

Figure 2: The quality Control plots. A, illustrates DEGs and housekeeping average values between Primary and Metastatic colon samples. B, illustrates common DEGs and housekeeping average values between Primary and Metastatic lung and liver samples. These charts exhibit metastatic group versus primary group.

3.2 Meta-Analysis and Identifying Differential Expressed Genes

4118 unique DEGs (differentially expressed genes) with adjusted p-value < 0.05 and absolute log fold change > 5 were achieved from eight groups of DEGs. Groups are liver metastasis versus normal, liver metastasis versus primary, lung metastasis versus normal, lung metastasis versus primary and primary versus normal for GSE41258 dataset. Liver metastasis contains metastatic colorectal samples acquired from liver tissue and lung metastasis involves metastatic colorectal samples obtained from lung tissue. Primary groups contain samples from primary colon tumor and normal groups contain adjacent normal colon samples. For GSE18105 they are metastasis versus normal, metastasis versus primary and primary versus normal. Common DEGs between all metastasis vs normal analyses in two datasets were 123 genes. Common DEGs between all metastasis vs primary analyses in two datasets were 12 regarded as the seed genes. Common DEGs between all primary vs normal analyses in two datasets were 210 genes. There were 242 unique DEGs between these three groups and they were considered as the final DEGs. All DEG sets and their LogFC are presented in Supplementary file 2.

3.3 Undirected Protein-Protein Interaction Network

242 final DEGs were utilized to construct the Protein-Protein-Interaction (PPI) network. STRING database was employed to generate the Interactions based on seven filtrations namely, Neighborhood, Text mining, Experiments, Databases, Co-expression, Gene fusion and Co-occurrence. STRING combined scores were used as the edge weights. The giant component of the weighted network with 205 nodes and 554 edges is presented in figure 3.

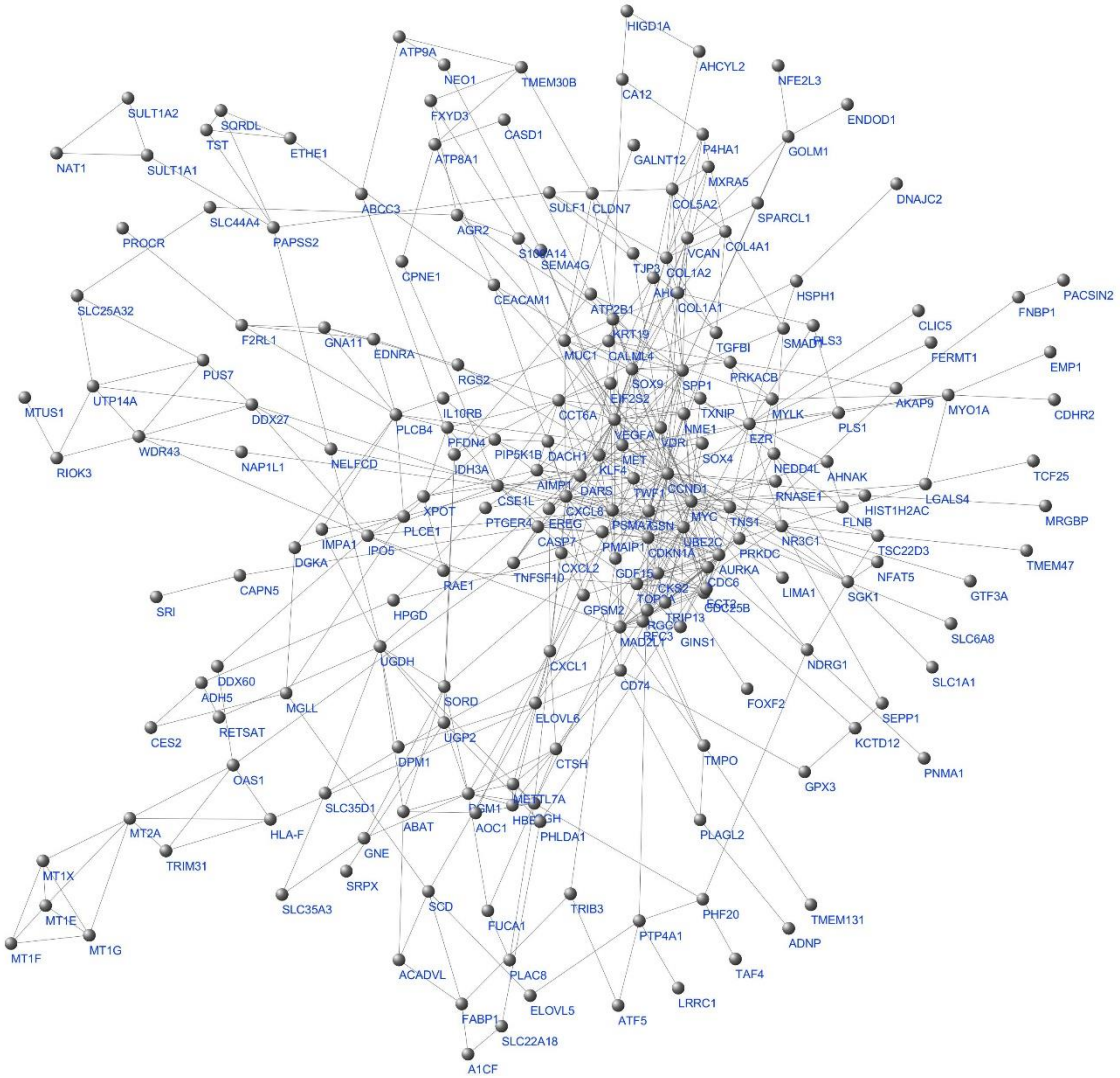

Figure3: The whole network giant component. Labels are protein/gene symbols. This is a scale free network [31] which follows a power law distribution (most of the network nodes have a low degree while there are few nodes with high degree).

3.4 Determination of Seed Genes Neighborhood through Shortest Path-Based Scoring System

In this step, we used weights between nodes as the estimation of distances in the weighted adjacency matrix. Nodes with shorter distances from the seed genes were selected and a smaller network was extracted from the main network. Computing the shortest path score

for the non-seed genes led to a network with 39 nodes comprising 12 seed nodes and 27 neighbors. This multi-component graph is called core network illustrated in figure 4.

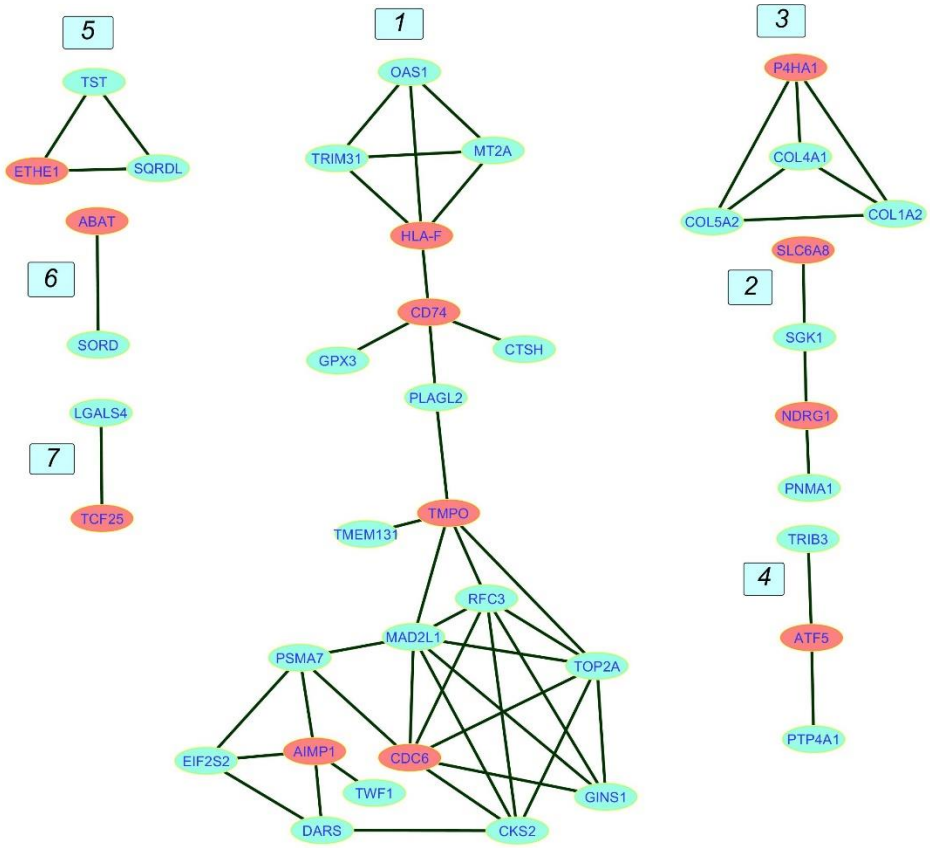

Figure 4. The Core network. The network contains seven components numbered from 1 to 7. Component 1 is the giant component. Seed genes are in red and non-seed genes are in blue.

The expression states for these genes between any pair-wise comparison are depicted in Table1. For the three Metastatic-Normal comparisons (MvsN) in the two microarray datasets, only TWF1 status was not the same between the three analyses. For the Primary-Normal analyses (PvsN), all the genes witness a similar status between the two datasets. Status for the three Metastatic-Primary comparisons (MvsP) were more challenging. In this set, the liver-primary and lung-primary were more similar to one another than the metastatic-primary comparison.

MvsN			
DEGs	GSE18105	GSE41258	
	metastatic-normal	liver-normal	lung-normal
ETHE1	Down	Down	Down
DARS	Up	Up	Up
SGK1	Down	Down	Down
TMEM131	Down	Down	Down
TST	Down	Down	Down
LGALS4	Down	Down	Down
TRIB3	Up	Up	Up
COL5A2	Up	Up	Up
COL4A1	Up	Up	Up
PTP4A1	Up	Up	Up
COL1A2	Up	Up	Up
SORD	Up	Up	Up
SQOR	Down	Down	Down
OAS1	Down	Down	Down
TRIM31	Down	Down	Down
TWF1	Up	Down	Down

MvsP			
DEGs	GSE18105	GSE41258	
	metastatic-primary	liver-primary	lung-primary
TCF25	Down	Up	Up
CD74	Down	Down	Up
ETHE1	Down	Down	Down
ATF5	Down	Up	Up
AIMP1	Up	Down	Down
CDC6	Down	Down	Down
ABAT	Up	Up	Down
SLC6A8	Down	Up	Up
HLA-F	Down	Down	Up
TMPO	Up	Down	Down
NDRG1	Down	Up	Up
P4HA1	Up	Up	Up

PvsN		
DEGs	GSE18105	GSE41258
	primary-normal	
SGK1	Down	Down
EIF2S2	Up	Up
TRIB3	Up	Up
DARS	Up	Up
RFC3	Up	Up
ETHE1	Down	Down
TOP2A	Up	Up
CKS2	Up	Up
SORD	Up	Up
PSMA7	Up	Up
GPX3	Down	Down
MAD2L1	Up	Up
SQOR	Down	Down
TST	Down	Down
GINS1	Up	Up
LGALS4	Down	Down
PNMA1	Down	Down
PTP4A1	Down	Down
PLAGL2	Up	Up
COL1A2	Up	Up
CTSH	Up	Up
MT2A	Down	Down
COL5A2	Up	Up

Tabe1: The practical information for the core network components. Up means gene was upregulated and Down means gene was Downregulated. MvsN contains all the metastatic versus normal analyses, PvsN contains all the primary versus normal analyses and MvsP contains all the metastatic versus primary analyses. Some genes are present in more than one analysis.

3.5 Network Descriptive

The network diameter was eight containing TRIM31, HLA-F, CD74, PLAGL2, TMPO, MAD2L1 PSMA7, AIMP1 and TWF1. In the following, the giant component descriptive is explained. Transitivity was around 60%, edge density was about 18% and the mean distance was 3.48. Two important centralities, Degree and Betweenness, for the giant component, are provided in Table 2. MAD2L1 had the highest Degree and a relatively high betweenness.

TMPO had the highest Betweenness and a pretty high degree. Similar to TMPO, its direct neighbour, PLAGL2 had a rather high Betweenness. This gene linked the two parts of the giant component together. A common feature of the graph components was that all of them contain at least one seed gene. This would suggest that each component might have a role in metastasis process. Other centralities for the nodes and the average distances between each node and the other nodes are provided in Supplementary file 3 (S3).

DEGs	Degree	Betweenness
MAD2L1	7	59.33
CDC6	6	6
TOP2A	6	13.83
RFC3	6	13.83
TMPO	5	107
CKS2	5	17
AIMP1	4	19.5
HLA-F	4	51
CD74	4	93
PSMA7	4	43
GINS1	4	0
TRIM31	3	0
OAS1	3	0
EIF2S2	3	0.5
DARS	3	5
MT2A	3	0
PLAGL2	2	91
CTSH	1	0
TWF1	1	0
GPX3	1	0
TMEM131	1	0

Table 2. Illustration of Degree and Betweenness centralities for the nodes in component 1. DEGs are sorted based on the highest Degree.

3.6 Processing Validation Datasets

Core network DEGs were identified in the five validation datasets. They are presented in Table 3. Most of the DEGs illustrated similar results in both Table1 and Table3 in MvsN and MvsP analyses which proved the accuracy of obtained DEGs from the test datasets. Expressions of genes that were totally homogeneous in all five to six analyses in each of MvsN or MvsP or PvsN are presented in green and the ones that differed only in one analysis are presented in yellow. Expression of Genes that were different in more than one dataset are in

white. LogFCs less than 0.2 or -0.2 were not reported in the table and are presented with dots. All DEGs identified in validation datasets are presented in Supplementary file 2.

MvsN						
DEGs	TCGA	GSE50760	GSE41258		GSE144259	GSE89393
	metastatic-normal	metastatic-normal	liver-normal	lung-normal	liver-normal	liver-normal
ETHE1	Down	Down	Down	Down	Down	Down
DARS	Up	Up	Up	Up	Up	Up
SGK1	Down	Down	Down	Down	Down	Up
TMEM131	Up	Down	Down	Down	Down	Down
TST	Down	Down	Down	Down	Down	Down
LGALS4	Up	Down	Down	Down	Down	Down
TRIB3	Up	Up	Up	Up	Up	Up
COL5A2	Down	Up	Up	Up	Up	Down
COL4A1	...	Up	Up	Up	Up	Down
PTP4A1	...	Down	Down	Down	Down	Down
COL1A2	Down	Up	Up	Up	Up	...
SORD	Up	Up	Up	Up	Up	Up
SQOR	Down	Down	Down	Down	Down	Down
OAS1	...	Down	...	Down	...	Down
TRIM31	Up	Down	Down	Down	Down	Down
TWF1	Up	...	...	Down	Up	Down

MvsP						
DEGs	TCGA	GSE50760	GSE41258		GSE144259	GSE89393
	metastatic-primary	metastatic-primary	liver-primary	lung-primary	liver-primary	liver-primary
TCF25	Down	...	...	...	Down	Down
CD74	Down	Up	...	Up	...	Down
ETHE1	Down	Down	...	Down	Down	...
ATF5	Down	Up	Up	...	Up	...
AIMP1	Down	...	Down	Down	...	Down
CDC6	Up	Up	Down	Down	Down	Down
ABAT	Up	Up	Up	Up	Up	Down
SLC6A8	Up	Down	Up	Up	Up	Up
HLA-F	Down	Down	...	...	Down	Down
TMPO	Down	...	...	Down	Down	Down
NDRG1	Up	Up	Up	Up	Up	Up
P4HA1	...	Up	Up	Up	Up	Down

PvsN					
DEGs	TCGA	GSE50760	GSE41258	GSE144259	GSE89393
	primary-normal				
SGK1	Down	Down	Down	Down	Down
EIF2S2	Up	Up	Up	Up	Up
TRIB3	Up	Up	Up	Up	Up
DARS	Up	Up	Up	Up	Up
RFC3	Up	Up	Up	Up	Up
ETHE1	Down	Down	Down	Down	Down
TOP2A	Up	Up	Up	Up	Up
CKS2	Up	Up	Up	Up	Up
SORD	Up	Up	Up	Up	Up
PSMA7	Up	Up	Up	Up	Up
GPX3	Up	Down	Down	Down	Down
MAD2L1	Up	Up	Up	Up	Up
SQOR	Down	Down	Down	Down	Down
TST	Down	Down	Down	Down	Down
GINS1	Up	Up	Up	Up	Up
LGALS4	...	Down	Down	Down	Down
PNMA1	...	Down	Down	Down	Down
PTP4A1	Up	Down	Down	Down	Down
PLAGL2	Down	Up	Up	Up	Up
COL1A2	Down	Up	Up	Up	Up
CTSH	...	Up	Up	Up	Up
MT2A	Down	Down	Down	Down	Down
COL5A2	Down	Up	Up	Up	Up

Table 3. Illustration of core network DEGs in validation datasets. Up means gene was upregulated and Down means gene was Downregulated. MvsN contains all the metastatic versus normal analyses, PvsN contains all the primary versus normal analyses and MvsP contains all the metastatic versus primary analyses. Some genes are present in more than one analysis.

3.7 Enrichment Analysis

Gene sets in the components 1 to 4 were given to *Enrichr* online tool. Tables 4A and 4B depict the enrichment analysis for the core network giant component. Enriched terms were sorted based on the highest p-value. Gastric Cancer Network2 was of the highest p-value containing TOP2A and RFC3 genes which are involved in DNA replication process. Involvement of the same genes in retinoblastoma cancer (WP2446) proposes the potential importance of these genes in different cancers. GPX3, TOP2A, and RFC3 were present in the highest number of pathways. Top2A was involved in Gastric Cancer Network1 as well.

In 4B, PSMA7 and MAD2L1 genes have monopolized the most GO terms to themselves. After them, OSA1 and HLA_F posseed the highest number of terms. GO:0060333, GO:0071346 and

GO:0019221 terms contained the highest number of genes. These terms are related to interferon gamma and cytokine-mediated signaling which presents the role of cytokine mediated signaling cascades in colon cancer progression. The enrichment analysis for the core network components of 2, 3 and 4 are presented in Supplementary file 4.

In component 2, SGK1 contained the most number of pathways. NDRG1 and SGK1 were of the maximum number of GO terms and all four genes in component 2 were engaged in the regulation of apoptosis processes. In component 3, all the collagen genes were involved in Focal Adhesion-PI3K-Akt-mTOR-signaling pathway and extracellular matrix organization biological process. TRIB3 in component 4, was implicated in three enriched pathways especially insulin signaling and together with ATF5 were enriched for the negative regulation of transcription GO terms.

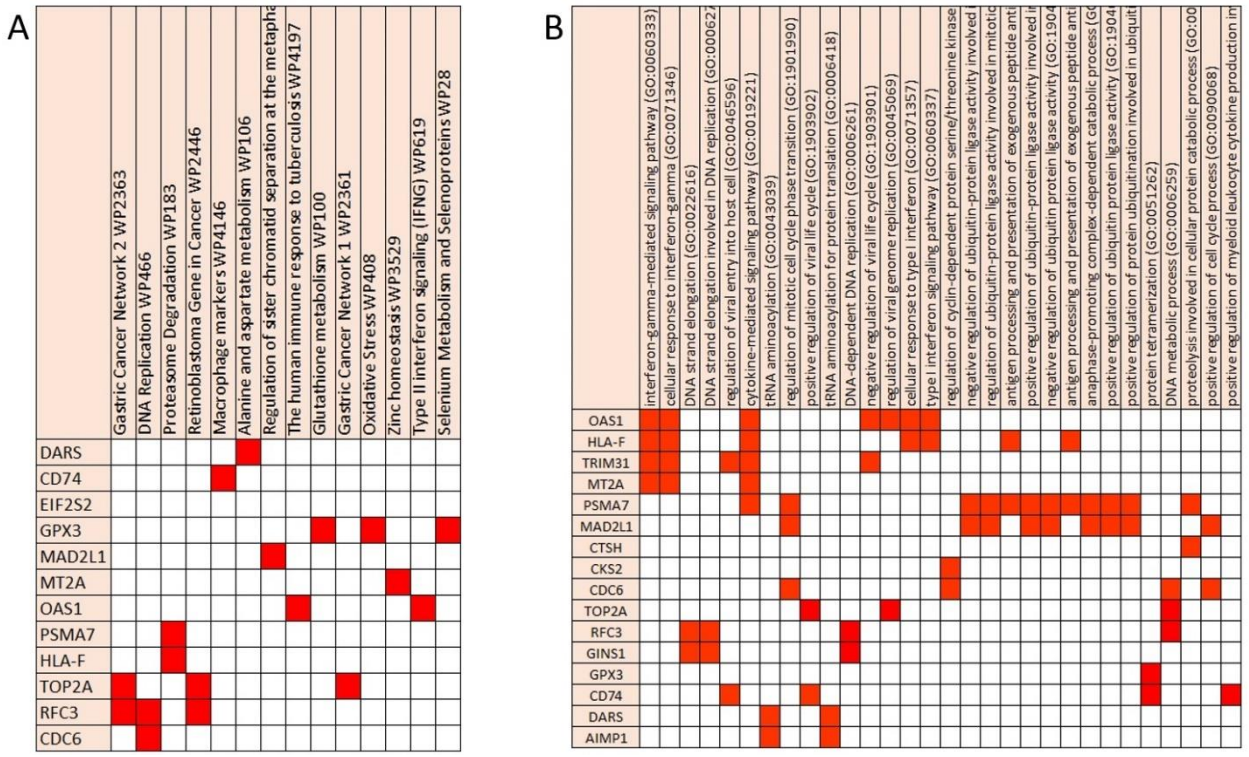

Table 4. Enrichment terms for the giant component. A, illustrates the enriched pathways from WikiPathways signaling datasets. B, shows the enriched terms for biological processes from GO repository. Red cells reveal the involvement of the genes in the enriched elements.

4. Discussion

The core network giant component is composed of an up and a down part attached via PLAGL2 transcription factor (TF). The lower part is engaged mainly in cell cycle and DNA replication . Component 2 contains genes involved in ECM remodeling, component 3 nodes produce ECM proteins, component 4 is composed of genes involved in transcription inhibition, and Component 5 is composed of mitochondrial genes playing important roles in controlling cellular redox homeostasis.

PLAGL2 is considered as an oncogene in different cancers. It binds to and prevents Pirh2 proteasomal degradation and in turn Pirh2 promotes proteasomal degradation of P53 protein [32]. PLAGL2 suppresses neural stem cell (NSC) differentiation by interfering in Wnt/β-catenin signaling resulting in glioblastoma [33]. Besides, PLAGL2 regulates actin cytoskeleton and cell migration through promoting actin stress fibers and focal adhesion [34]. Results of PvsN analysis manifests that this gene is induced in colon cancer and have a high betweenness centrality. Since this gene connects the two parts of the giant component, it would be a pertinent target for disturbing colon cancer network. Its expression is supported by the majority of validation datasets in Table 3.

TRIM31 (a ubiquitin ligase) was downregulated in MvsN in all test datasets and a large number of validation datasets. However, there are contradictory results in different studies where it is shown to be reduced in lung cancer cells [35] and stepped up in gastric [36] and colorectal cancer [37]. MT2A gene is an antioxidant involved in the process of autophagy and apoptosis which protects cells against hydroxyl radicals and plays an important role in detoxification of heavy metals [38, 39]. Inhibition of this gene expression results in proliferation inhibition of CRC cells [40] and its silencing promotes the effects of chemotherapy drugs on osteosarcoma primary tumors [41]. However, this gene was downregulated in PvsN analyses which are supported by the results in Table 3. It is also downregulated in pancreatic cancer [42]. Therefore, this downregulation needs to go under more investigation. OAS1 expression is downregulated in breast ductal carcinoma and prostate cancer (PCa), at both mRNA and protein levels. In addition, OAS1 expression is negatively correlated with the progression these cancers [43]. Our analysis shows that expression of this gene is suppressed in CRC like the other two cancers which is endorsed by

validation datasets. HLA-F, one of the seed genes, is a MCH class 1b gene a surface marker of activated lymphocytes [44]. Presence of anti-HLA-F antibodies in cancer patients serums is a diagnostic method that implies the activated lymphocytes against cancerous cells [45] . HLA-F is inducible by NF-κB and is responsive to IFN-γ similar to OAS1. Enriched GO terms back up the fact that these genes are engaged in interferon gamma and cytokine-mediated signaling (Table 2B). Although this gene saw mixed expression trend in MvsP, validation datasets proves its downregulation that could be a hallmark of CRC.

CD74 (HLA class II histocompatibility antigen gamma chain) one of the seed genes exhibited a mixed trend of expression in MvsP analyses. CD74 is a survival receptor expressed on normal colon epithelial cells and colon carcinoma cells. CD74 stimulation by MIF results in enhanced cell survival by induction of Bcl-2 and Akt phosphorylation [46, 47]. CTSH gene is a lysosomal cysteine protease upregulated in PvsN. This protease plays an important role in cytoskeletal protein Talin maturation. Talin promotes integrin activation and focal adhesion formation leading to cell migration [48]. Upregulation of this gene in CRC is more verified by validation datasets. As a result, suppression of CTSH expression could be a choice of metastasis inhibition. Glutathione peroxidase 3 (GPX3) is an important antioxidant enzyme that protects cells against Reactive Oxygen Species (ROS) and its downregulation occurs in many cancers. For instance, its expression is suppressed in human colon carcinoma Caco2 cell lines, resulting in augmented ROS production [49]. It reduces H_2O_2 and lipid peroxides to water and lipid alcohols, respectively, and in turn oxidizes glutathione to glutathione disulfide [50]. Downregulation of GPX3, happened in PvsN analyses, leading to ascending of H_2O_2 level which is positively correlated with the tumor progression [51]. Its downregulation is further supported by the majority of datasets in Table3. As a result, induction of GPX gene families would be a therapeutic approach.

TMPO gene presented a mixed expression trend in MvsP although Its downregulation is supported by validation datasets. This gene produces different protein isoforms via alternative splicing [52, 53]. The proteins are located in the nucleus of the cells which help to form nuclear lamina and maintenance of the nucleus membrane structure [54]. TMPO prevents the depolymerization of nuclear laminas and excessive activation of the mitosis pathway. TMPO-AS1 is an antisense lncRNA for TMPO that is overexpressed in CRC patients.

Its overexpression results in depolymerization of nuclear laminas and augmentation of mitosis pathways [55]. TMPO-AS1 expression is positively correlated with pathological features of the patients and distant metastasis as well [56]. Regarding the fact that TMPO is one of the seed genes that has the greatest Betweenness centrality, inhibition of TMPO-AS1 may result in the induction of TMPO gene and that would help hinder the tumor growth by disturbing CRC signaling network.

TMEM131 is a transmembrane protein that was downregulated in MvsN analyses and the large number of the validation datasets. No documentation was found to connect this gene to cancer. Therefore, this gene would be CRC specific. Enrichment analysis shows that this gene is engaged in IFN-γ signaling pathway.

TOP2A gene was upregulated in PvsN analyses which is completely endorsed by the validation results. In breast cancer (BC) HER-2 and TOP2A are the molecular targets for several anticancer medicines frequently coamplified together [57]. Moreover, Copy Number Variations (CNVs) in TOP2A gene have been identified as biomarkers in colorectal cancer [58]. This enzyme controls DNA topological structure and its upregulation is a hallmark of aberrant cell growth in tumors [59]. TOP2A mRNA expression is an independent prognostic factor in patients with (Estrogen Receptor) ER-positive breast cancer and could be useful in the assessment of the ER-positive breast cancer risk [60]. Therefore, rather than a target for cancer therapy, this gene could be either a possible prognostic or diagnostic marker for CRC.

Replication Factor C subunit 3 (RFC3) plays a role in DNA replication, DNA damage repair, and cell cycle checkpoint control. Hepatocellular carcinoma (HCC) and cell proliferation of ovarian tumors are suppressed by shRNA-mediated silencing of RFC3 gene [61, 62] . This gene was upregulated in PvsN and it is upregulated in Triple-negative breast cancer (TNBC) as well [63]. Its upregulation is supported by validation datasets. Since expression inhibition of this gene at both mRNA and protein level suppresses the migratory and invasive ability of MCF-7 cell lines [64], this gene would be a therapeutic target for colorectal cancer treatment. Moreover, TOP2A and RFC3 were shown to be engaged in the Gastric Cancer Network2 pathway in the enrichment analysis which could indicate the importance of these two genes in cancer progression.

Mitotic Arrest Deficient 2 Like1 (MAD2L1) upregulated in PvsN in both test and validation analyses is a mitotic spindle assembly checkpoint molecule that is responsible for preventing anaphase initiation until precise and complete metaphase alignment of all chromosomes takes place. An increase in the level of MAD2L1 transcripts was detected in a large number of samples of the ductal breast carcinoma [65]. Its upregulation in our analysis would provide evidence that cancerous cells were dealing with mitotic deficiencies. The GINS complex is a DNA replication machinery component in the eukaryotes and is a requirement for the replication initiation and progression of DNA replication forks [66]. GINS1 (PSF1) mRNA level is positively correlated with tumor size in CRC patients and is a prognostic marker of CRC [67]. This gene has been recently introduced as a targeted oncogenic agent for inhibition of synovial sarcoma [68]. It was totally upregulated PvsN analyses in both Tables 1 and 3. therefore, its expression inhibition would be a potential target for inhibition of tumor growth by disturbing DNA replication machinery.

CDC6, one of the seed genes, plays a critical role in regulation of the eukaryotic DNA replication onset and its downregulation has been demonstrated in prostate cancer [69]. Expression of this gene is regulated by E2F and androgen receptor (AR) in PCa cells [70]. Transfection of CDC6 siRNA leads to not only decreased level of ovarian cancer cell proliferation but also increased apoptosis rates [71]. Cdc6 is considered as a potent therapeutic target in BC patients [72]. It is a regulator of cell cycle in S phase and its expression and nucleus translocation is mediated by E_2F TF and Cdks respectively in Saccharomyces cerevisiae [73]. Results for this gene in MvsP analysis are totally contradictory to the BC data. Supporting results in Table 3 presents a vague trend as well. Since this gene is more engaged in aberrant cell growth, It was upregulated in primary vs normal analysis in all validation datasets and GSE18105 test dataset. No study directly measured the expression level of this gene in CRC samples.

CKS2 protein interacts with the catalytic subunit of the cyclin-dependent kinases and its downregulation contributes to suppression of p-Akt and p-mTOR. Therefore, one of CSK2 oncogenic roles is played by Akt/mTOR oncogenic pathway [74]. CKS2 is expressed at high level in CRC tissues and it has revealed that increased CKS2 expression is highly correlated with enhanced metastatic stage [75]. Importantly, CKS2 is considered as a potential

biomarker and therapeutic target for the BC treatment due to the fact that its inhibition suppresses cell proliferation and invasion ability in vitro and in vivo [76]. In the PvN analysis, this gene was upregulated which would be a therapeutic target for CRC treatment because validation results completely support this upregulation.

PSMA7 gene encodes a protein that is one of the essential subunits of 20S proteasome complex [77]. Overexpression of PSMA7 both at the mRNA and protein levels has been reported in gastric cancer [78]. Depletion of PSMA7 by shRNA-transfected RKO CRC cell lines mediates inhibition of cell growth and migration. Consequently, inhibition of PSMA7 could be a beneficial therapeutic strategy for colorectal cancer patients [79]. This gene was upregulated in PvsN analyses in test and validation datasets. AIMP1 is a precursor of inflammatory cytokine EMAPII which mediates growth inhibition of PCa cells. In addition, it sensitizes tumor vessels to tumor necrosis factor α (TNFα) [80]. Furthermore, Down-expression of AIMP1, AIMP2 and AIMP3 genes in CRC tissues has been reported which suggests the tumor-suppressive roles of AIMP proteins [81]. AIMP1 gene depicted a mixed expression trend in MvsP analysis.

DARS was found to be upregulated in MvsN analyses in all test and validation datasets. This gene encodes a member of a multi-enzyme complex that its role has been proved in mediating attachment of amino acids to their cognate tRNAs. No study has investigated the role of this gene in cancer progression so far. Only two studies have reported that DARS-AS1 gene is positively associated with the pathological stages in thyroid and ovarian cancer by targeting mir-129 and mir-532-3p respectively [82, 83]. This gene might be a CRC prognostic marker or a curative target.

EIF2S2 or EIF2B acts in the early steps of protein synthesis. GTP-bound EIF-2 transfers Met-tRNAi to the 40S ribosomal subunit. EIF-2 consists of three subunits, alpha, beta, and gamma. The hydrolysis of GTP to GDP takes place at the end of the initiation process that leads to release of the inactive eIF2·GDP from the ribosome. Exchange of GDP for GTP is performed by beta subunit so that active EIF-2 is ready for another round of initiation [84]. In one study, EIF2B was proposed as a potent therapeutic target in lung cancer [76]. Moreover, elimination of natural killer cell cytotoxicity via promoted expression of natural

killer (NK) cell ligands is done by pSer535-eIF2B following by expression of pSer9-GSK-3β (inactive GSK3β) and generation of ROS which promotes breast cancer growth and metastasis [85]. Since Tyr216-GSK-3β (Active GSK3β) has the inhibitory effects on the EMT process contrary to the inactive form, induction of active GSK-3β together with EIF2B would be a therapeutic approach to prevent EMT [86]. EIF2B stepped up in PvsN analyses which is supported by validation results.

TWF1 gene encodes twinfilin, an actin monomer-binding protein that promotes EMT in pancreatic cancer tissues [87]. siRNAs of TWF1 dramatically inhibits F-actin organization and focal adhesions formation and promotes the mesenchymal-to-epithelial transition (MET) with a round cell shape in BC MDA-MB-231 cell lines. Inhibition of TWF1 expression by both miR-30c and siRNA in MDA-MB-231 cells increases sensitivity of the cancer cells to doxorubicin and paclitaxel drugs. Furthermore, expression levels of EMT markers, VIM and SNAI2, are reduced as a result of miR-30c action [88]. However, in MvsN analyses, this gene witnessed a mixed expression trend in both test and validation datasets. As a results, it is a struggle to propose it as target for CRC treatment.

Creatine is involved in CRC enhancement and liver metastasis of colorectal cancer. Creatine kinase-B is secreted by cancerous cells into the ECM to generates phosphocreatine [89]. Phosphocreatine is then imported into metastatic cells via a specific creatine transporter SLC6A8, one of the seed genes and a member of component 2. Intracellular phosphocreatine is converted into ATP to provide the fuel for the survival of metastatic cancer cells within the hypoxic microenvironment [89]. Consistent with these findings, depletion of SLC6A8 gene in colon and PCa metastatic cell lines hinders the liver colonization of these cells in mouse [90]. Although this gene was upregulated in MvsP in the majority of validation and test datasets, it was downregulated in all seven datasets in PvsN analyses. This might suggest that once there is the lack of oxygen, CRC cells acquire the necessary energy via other biochemical pathways which would be triggered by the nodes in component 5.

SGK1, a member of component 2, and AKT are two families of AGC protein superfamily. SGK1 is a serine/threonine kinase that activates certain potassium, sodium, and chloride

channels [91]. SGK1 is a downstream effector of PI3K, which run pathways independent of pathways shared with AKT. The two kinases are phosphorylated and activated by PDK1 and mTORC2 complex [92, 93]. In general, PI3K-dependent survival signals can be mediated by either Akt or SGK1 that inactivates the pro-apoptotic proteins Bad and FKHRL1 [94]. In a study on A498 kidney cancer cells, they found that survival signals promoted by IL-2 is mediated by SGK1 activation [95]. Moreover, the promoter of SGK1 is under tight control of the p53 protein [96]. SGK1 has been shown to mediate cell survival and drug resistance to platinoid and taxane compounds in breast cancer patients [97]. On the contrary, This gene was totally downregulated in all MvsN and PvsN of validation and test datasets which are a total of 16 different analyses. This overall downregulation might be specific to CRC so it could be a diagnostic hallmark of CRC.

One of the seed genes NDRG1 is connected to SGK1 gene. This gene which has a mixed expression trend in MvsP is one of the repressors of the EMT in metastatic CRC by attenuating NF-κB signaling [98]. Inhibition of NDRG1 results in the increase in NF-kB and the reduction in cell-cell junction protein E-cadherin [99]. The promoter region of the NDRG1 gene contains a p53 binding site which is necessary in caspase activation and apoptosis in the case of DNA damage [100]. A transcription factor that plays an important role in response to low oxygen concentrations, or hypoxia is Hypoxia-Inducible Factor (HIF)-1. HIF-1α is translocated from cytoplasm to nucleus, where it binds to HIF-1β to form HIF-1 complex. This complex works as a transcription factor that binds to the hypoxia response element (HRE) in the promoter of NDRG1 [101]. Consequently, under hypoxic environment of tumor, induction of NDRG1 by HIF-1 complex is a defensive mechanism against cancer progression. HIF-1α also removes the inhibitory effect of myc complex on this gene [101]. Furthermore, NDRG1 interferes with WNT/ β-catenin pathway by downregulating nuclear β-catenin and preventing metastasis in CRC [102]. In addition, NDRG1 Controls gastric cancer migration and invasion through regulating MMP-9 [103]. However, our results would suggest that the tumor suppressive role of this gene could be targeted in benign tumors than metastatic sites since expression of this gene was downregulated in PvsN analyses of validation datasets while it was upregulated in their all MvsP analyses.

An increase in expression of PNMA1 is observed in pancreatic ductal adenocarcinoma patients (PDAC). In addition, the higher expression of PNMA1 gene, the larger tumor size is observed. Suppressing PNMA1 expression causes a reduction in cell viability which promotes apoptosis. Upregulation of PNMA1 occurs in HCC and higher expression of PNMA1 is associated with more aggressive phenotypes [104]. This gene was downregulated in all MvsN, MvsP and PvsN analyses in validation datasets and MvsN and PvsN analyses of the test datasets. These similarities in a large number of analyses is not accidental therefore, downregulation of this gene would be CRC specific which would diagnostic hallmark. What's more, all the members of component 2, were downregulated in PvsN analyses.

Component 3 contains collagen (COL1A2, COL5A2 and COL4A1) and P4HA1 (a collagen hydroxylase) genes interconnected in the process of ECM remodeling based on enrichment results. All the members presented a similar trend in expression. COL1A2 and COL5A2 were upregulated in MvsN and PvsN analyses. COL4A1 was upregulated in MvsN and P4HA1 which is one of the seed genes upregulated in MvsP. The majority of validation datasets support these upregulations. P4HA1 is implicated in metastasis of breast and pancreatic cancers [105, 106] . Under hypoxic condition, this gene would trigger HIF-1 pathway, cell survival and progression of colorectal cancer cells [107]. HIF-1 induces expression of genes that encodes collagen prolyl (P4HA1 and P4HA2) and lysyl (PLOD2) hydroxylases. P4HA1 and P4HA2 are required for collagen deposition, whereas PLOD2 is required for ECM stiffening and collagen fiber alignment [107]. These changes in ECM triggered by HIF-1 are necessary for motility and invasion because in focal adhesion junctions, actin cytoskeleton is connected to ECM through attachment of integrins to collagens [108]. Besides, there is a positive feedback between P4HA1 and HIF-1 in modulation of ECM. As a result, targeting P4HA1 and P4HA2 expressions would inhibit the progression of cell migration via HIF1-Collagen pathway.

PTP4A1 a member of component 4, is a protein phosphatase engaged in p21-activated kinase (PAK) signaling pathway. Inhibition of PTP4A1 gene in MDA-MB-231 breast cancer cells by an increase in miR-944 expression impairs cell invasion [109]. Rac/Cdc42 activates PAKs leading to formation of membrane ruffles, stress fibers and remodeling of focal adhesion junctions [110]. This phosphatase might have a relationship with Src by triggering

TGFB1 induction and cancer progression [111]. However, this gene was down-regulated in PvsN in the majority of datasets and its expression in MvsN and MvsP follows a mixed trend. BCL-2 is a target of ATF5 gene one of the seed genes [112]. ATF5 follows a mixed expression trend in test and validation datasets in MvsP analyses. There are pieces of evidence that link the role of ATF5 in mitochondrial dysfunction in cancer progression [113]. In malignant glioma, metastatic cells take advantage of survival signals triggered by ATF5 gene which is important to ignore anchorage-dependent and niche-dependent cell death signals [114]. TRIB3 is a prognosis hallmark of colorectal cancer activated under hypoxic conditions [115]. TRIB3 silencing suppresses VEGF-A expression in gastric cancerous cells which inhibits endothelial cell migration and vessel formation. This gene was upregulated in MvsN analyses in all test and validation datasets, therefore, it would be a favorable target for anti-angiogenic therapy [116].

Genes in component 5 are mitochondrial which their role in cancer progression has not been sufficiently investigated so far. All three genes were downregulated in our analysis. Because of the presence of anaerobic bacteria in the digestive tract these genes are highly expressed in normal colon tissue compared to other tissues [117]. These findings are supported by the RNA-seq expression information in the Gene database of NCBI [118]. ETHE1 (persulfide dioxygenase) and SQOR are antioxidants that convert hydrogen sulfide (H2S) to persulfide and persulfide to sulfite respectively. Therefore, they protect cells against toxic concentrations of sulfide. ETHE1 is the only gene in our analysis which was downregulated in the three analyses while SQOR was downregulated in MvsN and PvsN analyses. All these expressions are totally verified by the validation datasets. Their downregulation is essential for cancer cells proliferation and survival since it has been proved that under the hypoxic environment of CRC tumors, sulfide is a supplementary tool that provides electron for mitochondrial ETC to generate ATP [119]. As a result, helping in expression induction or activation of ETHE1 and SQOR proteins would hinder CRC tumor growth. TST thiosulfate sulfurtransferase encodes a protein that is localized to the mitochondria and catalyzes the conversion of thiosulfate and cyanide to thiocyanate and sulfite. Therefore, like the previous two mitochondrial enzymes, it acts in Hydrogen sulfide metabolism [120].

ABAT gene expression suppression is a biomarker for tamoxifen resistance in advanced breast cancer however, there are contradictory results on the expression of this gene in ER+ breast cancer [121, 122] Another study has reported the expression suppression of ABAT gene due to the activation of SNAIL transcription factor in aggressive breast cancer [123]. This gene showed a mixed expression trend in MvsP although the majority of validation datasets illustrate its upregulation. SORD is another element of component 6 upregulated in MvsN and PvsN analyses. The connection of this gene with cancer has not been efficiently investigated. Since it exhibited a totally ascending trend in all validation and test datasets, it might be a potential target and biomarker for CRC treatment. TCF25 in component 7 witnessed a mixed expression trend in MvsP analysis. This gene is a basic helix-loop-helix transcription factor that is vital in embryonic [124]. LGALS4 is implicated in regulating cell-cell and cell-matrix interactions so its induction might have positive curative impacts on CRC cells. This gene is mostly expressed in small intestine, colon, and rectum, which is suppressed in CRC [125]. It was downregulated in MvsN and PvsN analyses in both validation and test datasets. It is also a blood marker of CRC [126].

5. Conclusion

In this study, we illustrated some therapeutic targets and biomarkers for CRC. A combination of these targets would beneficially disturb progression of colorectal cancer. Generally, all the discovered antioxidants were downregulated in different stages of CRC namely ETHE1, SQOR, TST, GPX3, and MT2A. CDC6 Downregulation in benign colon tumors is another finding of this study which its upregulation would inhibit tumor growth. PSMA7 is a previously reported target for CRC treatment. All members of component 2 were downregulated in PvsN analyses but SGK1 and PNMA1 witnessed a significant downregulation in all the analyses of the test and validation datasets. Therefore, these downregulations would be CRC specific. Upregulation of RFC3 and TOP2A together might be a prognostic marker of CRC and their expression regulation might hinder tumor growth. The combination of genetic targets that inhibit angiogenesis (TRIB3 suppression), aberrant

proliferation (GINS1 and CSK2 suppression), EMT (suppression of Pirh2, CTSH and P4HA1 and induction of GSK-3β together with EIF2B), migratory behavior (suppression of RFC3, P4HA1), mitotic pathways (TMPO-AS1 inhibition), energy production under hypoxic conditions (induction of ETHE1 and SQOR) and ROS generation (induction of GPX3) could be a pertinent CRC treatment strategy. Furthermore, Expressions for ETHE1 and P4HA1 genes in MvsP analyses depict a pretty homogenous trend. What's more, based on the evidence and our findings, suppression of oncogenic targets of HIF-1 (P4HA1) and induction of tumor-suppressive targets of HIF-1 complex (NDRG1) could hinder the progression of CRC tumors. LGALS4 downregulation and its presence in blood would be the diagnostic and prognostic markers as well as a curative target. The performance of ABAT and TCF25 genes in CRC needs to go under investigation. TMEM131, DARS and SORD genes have specific expression trends as cells go from normal to metastatic. The relation of these three genes to colorectal cancer progression has been reported for the first time in this study so more investigation is required to find their molecular mechanism causing cancer promotion.

References

1. Favoriti, P., et al., *Worldwide burden of colorectal cancer: a review.* Updates in surgery, 2016. **68**(1): p. 7-11.
2. Ferlay, J., et al., *v2. 0, Cancer Incidence and Mortality Worldwide.* IARC Cancerbase No, 2008. **10**.
3. Chambers, A., A. Groom, and I. MacDonald, *ā€ Dissemination and Growth of Cancer Cells in Metastatic Sites, ā€ Nat.* Rev. Cancer, 2002. **2**: p. 563ā.
4. Jamora, C. and E. Fuchs, *Intercellular adhesion, signalling and the cytoskeleton.* Nature cell biology, 2002. **4**(4): p. E101.
5. Al-Sohaily, S., et al., *Molecular pathways in colorectal cancer.* Journal of gastroenterology and hepatology, 2012. **27**(9): p. 1423-1431.

6. Thompson, E.W. and D.F. Newgreen, *Carcinoma invasion and metastasis: a role for epithelial-mesenchymal transition?* Cancer research, 2005. **65**(14): p. 5991-5995.
7. Sierra, J.R., V. Cepero, and S. Giordano, *Molecular mechanisms of acquired resistance to tyrosine kinase targeted therapy.* Molecular cancer, 2010. **9**(1): p. 75.
8. Wang, S.-W. and Y.-M. Sun, *The IL-6/JAK/STAT3 pathway: potential therapeutic strategies in treating colorectal cancer.* International journal of oncology, 2014. **44**(4): p. 1032-1040.
9. Dihlmann, S. and M. von Knebel Doeberitz, *Wnt/β-catenin-pathway as a molecular target for future anti-cancer therapeutics.* International journal of cancer, 2005. **113**(4): p. 515-524.
10. Krüger, A., et al., *Hydroxamate-type matrix metalloproteinase inhibitor batimastat promotes liver metastasis.* Cancer research, 2001. **61**(4): p. 1272-1275.
11. Katz, L.H., et al., *Targeting TGF-β signaling in cancer.* Expert opinion on therapeutic targets, 2013. **17**(7): p. 743-760.
12. Pickup, M., S. Novitskiy, and H.L. Moses, *The roles of TGFβ in the tumour microenvironment.* Nature Reviews Cancer, 2013. **13**(11): p. 788-799.
13. Rychahou, P., et al., *Colorectal cancer lung metastasis treatment with polymer–drug nanoparticles.* Journal of controlled release, 2018. **275**: p. 85-91.
14. Rychahou, P., et al., *Colorectal cancer pulmonary metastasis treatment with lung-selective delivery of PAN-class I PI3K inhibitors*. 2016, AACR.
15. Jonker, D.J., et al., *Cetuximab for the treatment of colorectal cancer.* New England Journal of Medicine, 2007. **357**(20): p. 2040-2048.
16. Tol, J., et al., *Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer.* New England Journal of Medicine, 2009. **360**(6): p. 563-572.
17. Gautier, L., et al., *affy—analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-315.
18. Do, J.H. and D.-K. Choi, *Normalization of microarray data: single-labeled and dual-labeled arrays.* Molecules & Cells (Springer Science & Business Media BV), 2006. **22**(3).
19. Ramasamy, A., et al., *Key issues in conducting a meta-analysis of gene expression microarray datasets.* PLoS medicine, 2008. **5**(9): p. e184.
20. Gentleman, R., et al., *Genefilter: Methods for filtering genes from microarray experiments.* R package version, 2011. **1**(0).
21. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 2005, Springer. p. 397-420.
22. Ferreira, J. and A. Zwinderman, *On the Benjamini–Hochberg method.* The Annals of Statistics, 2006. **34**(4): p. 1827-1849.
23. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.
24. Luke, D.A., *A user's guide to network analysis in R*. 2015: Springer.
25. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC bioinformatics, 2008. **9**(1): p. 559.
26. Jasika, N., et al. *Dijkstra's shortest path algorithm serial and parallel execution performance analysis*. in *2012 proceedings of the 35th international convention MIPRO*. 2012. IEEE.
27. Nomaguchi, M. and A. Adachi, *Highly Mutable Animal RNA Viruses: Adaptation and Evolution.* Frontiers in microbiology, 2017. **8**: p. 1785.
28. Berger, S.I., A. Ma'ayan, and R. Iyengar, *Systems pharmacology of arrhythmias.* Sci. Signal., 2010. **3**(118): p. ra30-ra30.
29. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic acids research, 2016. **44**(W1): p. W90-W97.

30. Oldham, M.C., et al., *Identification and Removal of Outlier Samples Supplement for:" Functional Organization of the Transcriptome in Human Brain.* dim (dat1). **1**(18631): p. 105.
31. Khanin, R. and E. Wit, *How scale-free are biological networks.* Journal of computational biology, 2006. **13**(3): p. 810-818.
32. Zheng, G., J. Ning, and Y.-C. Yang, *PLAGL2 controls the stability of Pirh2, an E3 ubiquitin ligase for p53.* Biochemical and biophysical research communications, 2007. **364**(2): p. 344-350.
33. Zheng, H., et al., *PLAGL2 regulates Wnt signaling to impede differentiation in neural stem cells and gliomas.* Cancer cell, 2010. **17**(5): p. 497-509.
34. Sekiya, R., et al., *PLAGL2 regulates actin cytoskeletal architecture and cell migration.* Carcinogenesis, 2014. **35**(9): p. 1993-2001.
35. Li, H., et al., *TRIM31 is downregulated in non-small cell lung cancer and serves as a potential tumor suppressor.* Tumor Biology, 2014. **35**(6): p. 5747-5752.
36. Sugiura, T., *The cellular level of TRIM31, an RBCC protein overexpressed in gastric cancer, is regulated by multiple mechanisms including the ubiquitin—proteasome system.* Cell biology international, 2011. **35**(7): p. 657-661.
37. Wang, H., et al., *TRIM31 regulates chronic inflammation via NF-κB signal pathway to promote invasion and metastasis in colorectal cancer.* American journal of translational research, 2018. **10**(4): p. 1247.
38. Ma, H., et al., *HMBOX1 interacts with MT2A to regulate autophagy and apoptosis in vascular endothelial cells.* Scientific reports, 2015. **5**: p. 15121.
39. Liu, Y., et al., *EOLA1 protects lipopolysaccharide induced IL-6 production and apoptosis by regulation of MT2A in human umbilical vein endothelial cells.* Molecular and cellular biochemistry, 2014. **395**(1-2): p. 45-51.
40. Xu, Y., et al., *Tea polyphenol inhibits colorectal cancer with microsatellite instability by regulating the expressions of HES1, JAG1, MT2A and MAFA.* Zhong xi yi jie he xue bao= Journal of Chinese integrative medicine, 2010. **8**(9): p. 870-876.
41. Mangelinck, A., et al., *MT2A is an early predictive biomarker of response to chemotherapy and a potential therapeutic target in osteosarcoma.* Scientific reports, 2019. **9**(1): p. 1-12.
42. Hui, B., et al., *Overexpressed long noncoding RNA TUG1 affects the cell cycle, proliferation, and apoptosis of pancreatic cancer partly through suppressing RND3 and MT2A.* OncoTargets and therapy, 2019. **12**: p. 1043.
43. Maia, C.J., et al., *Oligoadenylate synthetase 1 (OAS1) expression in human breast and prostate cancer cases, and its regulation by sex steroid hormones.* Advances in Modern Oncology Research, 2016. **2**(2): p. 97-110.
44. Lee, N., A. Ishitani, and D.E. Geraghty, *HLA-F is a surface marker on activated lymphocytes.* European journal of immunology, 2010. **40**(8): p. 2308-2318.
45. NOGUCHI, K., et al., *Detection of anti-HLA-F antibodies in sera from cancer patients.* Anticancer research, 2004. **24**(5C): p. 3387-3392.
46. Maharshak, N., et al., *CD74 is a survival receptor on colon epithelial cells.* World journal of gastroenterology: WJG, 2010. **16**(26): p. 3258.
47. Egawa, K., *Cancer cell-specific HLA-F antigen and a diagnostic method of cancer by using thereof*. 2004, Google Patents.
48. Jevnikar, Z., et al., *Cathepsin H mediates the processing of talin and regulates migration of prostate cancer cells.* Journal of Biological Chemistry, 2013. **288**(4): p. 2201-2209.
49. Barrett, C.W., et al., *Tumor suppressor function of the plasma glutathione peroxidase gpx3 in colitis-associated carcinoma.* Cancer research, 2013. **73**(3): p. 1245-1255.
50. Sarsour, E.H., et al., *Redox control of the cell cycle in health and disease.* Antioxidants & redox signaling, 2009. **11**(12): p. 2985-3011.

51. Moloney, J.N. and T.G. Cotter. *ROS signalling in the biology of cancer*. in *Seminars in cell & developmental biology*. 2018. Elsevier.
52. Holmer, L. and H. Worman, *Inner nuclear membrane proteins: functions and targeting.* Cellular and Molecular Life Sciences CMLS, 2001. **58**(12-13): p. 1741-1747.
53. Harris, C.A., et al., *Three distinct human thymopoietins are derived from alternatively spliced mRNAs.* Proceedings of the National Academy of Sciences, 1994. **91**(14): p. 6283-6287.
54. Huang, W., et al., *Overexpression of AR-regulated lncRNA TMPO-AS1 correlates with tumor progression and poor prognosis in prostate cancer.* The Prostate, 2018. **78**(16): p. 1248-1261.
55. Brachner, A. and R. Foisner, *Lamina-associated polypeptide (LAP) 2α and other LEM proteins in cancer biology*, in *Cancer Biology and the Nuclear Envelope*. 2014, Springer. p. 143-163.
56. Mohammadrezakhani, H., et al., *Overexpression and Clinicopathological Correlation of Long Noncoding RNA TMPO-AS1 in Colorectal Cancer Patients.* Journal of gastrointestinal cancer, 2019: p. 1-5.
57. Simon, R., et al., *HER-2 and TOP2A coamplification in urinary bladder cancer.* International journal of cancer, 2003. **107**(5): p. 764-772.
58. Vang Nielsen, K., et al., *The value of TOP2A gene copy number variation as a biomarker in breast cancer: Update of DBCG trial 89D.* Acta oncologica, 2008. **47**(4): p. 725-734.
59. De Resende, M.F., et al., *Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer.* Journal of translational medicine, 2013. **11**(1): p. 36.
60. Rody, A., et al., *Gene expression of topoisomerase II alpha (TOP2A) by microarray analysis is highly prognostic in estrogen receptor (ER) positive breast cancer.* Breast cancer research and treatment, 2009. **113**(3): p. 457-466.
61. Shen, H., et al., *ShRNA-mediated silencing of the RFC3 gene suppress ovarian tumor cells proliferation.* International journal of clinical and experimental pathology, 2015. **8**(8): p. 8968.
62. Yao, Z., et al., *shRNA-mediated silencing of the RFC3 gene suppresses hepatocellular carcinoma cell proliferation.* International journal of molecular medicine, 2015. **36**(5): p. 1393-1399.
63. He, Z.-Y., et al., *Up-Regulation of RFC3 promotes triple negative breast cancer metastasis and is associated with poor prognosis via EMT.* Translational oncology, 2017. **10**(1): p. 1-9.
64. Zhou, J., et al., *Downregulation of hsa_circ_0011946 suppresses the migration and invasion of the breast cancer cell line MCF-7 by targeting RFC3.* Cancer management and research, 2018. **10**: p. 535.
65. Scintu, M., et al., *Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma.* Cancer letters, 2007. **254**(2): p. 298-307.
66. Labib, K. and A. Gambus, *A key role for the GINS complex at DNA replication forks.* Trends in cell biology, 2007. **17**(6): p. 271-278.
67. Wei, H.-B., et al., *Expression and clinical significance of GINS complex in colorectal cancer.* Zhonghua wei chang wai ke za zhi= Chinese journal of gastrointestinal surgery, 2011. **14**(6): p. 443-447.
68. Tang, L., et al., *Anlotinib inhibits synovial sarcoma by targeting GINS1: a novel downstream target oncogene in progression of synovial sarcoma.* Clinical and Translational Oncology, 2019: p. 1-10.
69. Robles, L.D., et al., *Down-regulation of Cdc6, a cell cycle regulatory gene, in prostate cancer.* Journal of Biological Chemistry, 2002. **277**(28): p. 25431-25438.
70. Liu, Y., et al., *FOXM1 and androgen receptor co-regulate CDC6 gene transcription and DNA replication in prostate cancer cells.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2014. **1839**(4): p. 297-305.
71. Sun, T.-Y., et al., *Expression of CDC6 in ovarian cancer and its effect on proliferation of ovarian cancer cells.* Int J Clin Exp Med, 2016. **9**(6): p. 10544-10550.

72. Mahadevappa, R., et al., *The prognostic significance of Cdc6 and Cdt1 in breast cancer.* Scientific reports, 2017. **7**(1): p. 985.
73. Santocanale, C. and J. Diffley, *ORC-and Cdc6-dependent complexes at active and inactive chromosomal replication origins in Saccharomyces cerevisiae.* The EMBO journal, 1996. **15**(23): p. 6671-6679.
74. Xu, J., Y. Wang, and D. Xu, *CKS2 promotes tumor progression and metastasis and is an independent predictor of poor prognosis in epithelial ovarian cancer.* European review for medical and pharmacological sciences, 2019. **23**(8): p. 3225-3234.
75. Yu, M.-H., et al., *Up-regulated CKS2 promotes tumor progression and predicts a poor prognosis in human colorectal cancer.* American journal of cancer research, 2015. **5**(9): p. 2708.
76. Huang, N., et al., *Overexpression of CKS2 is associated with a poor prognosis and promotes cell proliferation and invasion in breast cancer.* Molecular medicine reports, 2019. **19**(6): p. 4761-4769.
77. Coux, O., K. Tanaka, and A.L.J.A.r.o.b. Goldberg, *Structure and functions of the 20S and 26S proteasomes.* 1996. **65**(1): p. 801-847.
78. Xia, S., et al., *Overexpression of PSMA7 predicts poor prognosis in patients with gastric cancer.* Oncology letters, 2019. **18**(5): p. 5341-5349.
79. Hu, X.-T., et al., *Depletion of the proteasome subunit PSMA7 inhibits colorectal cancer cell tumorigenicity and migration.* Oncology reports, 2009. **22**(5): p. 1247-1252.
80. Schwarz, R.E., et al., *Antitumor effects of EMAP II against pancreatic cancer through inhibition of fibronectin-dependent proliferation.* Cancer biology & therapy, 2010. **9**(8): p. 632-639.
81. Kim, S.S., et al., *Expression of AIMP1, 2 and 3, the scaffolds for the multi-tRNA synthetase complex, is downregulated in gastric and colorectal cancer.* Tumori Journal, 2011. **97**(3): p. 380-385.
82. Zheng, W., et al., *LncRNA DARS-AS1 regulates microRNA-129 to promote malignant progression of thyroid cancer.* European Review for Medical and Pharmacological Sciences, 2019. **23**: p. 10443-10452.
83. Huang, K., et al., *Long noncoding RNA DARS-AS1 acts as an oncogene by targeting miR-532-3p in ovarian cancer.* European review for medical and pharmacological sciences, 2019. **23**(6): p. 2353-2359.
84. Kimball, S.R., *Eukaryotic initiation factor eIF2.* The international journal of biochemistry & cell biology, 1999. **31**(1): p. 25-29.
85. Jin, F., et al., *The PI3K/Akt/GSK-3β/ROS/eIF2B pathway promotes breast cancer growth and metastasis via suppression of NK cell cytotoxicity and tumor cell susceptibility.* Cancer biology & medicine, 2019. **16**(1): p. 38.
86. Goto, D., et al., *P1. 03-13 eIF2β, A Subunit of Translation-Initiation Factor EIF2, as a Potential Therapeutic Target for Non-Small Cell Lung Cancer.* Journal of Thoracic Oncology, 2018. **13**(10): p. S516-S517.
87. Jin, X., et al., *MicroRNA-486-5p improves nonsmall-cell lung cancer chemotherapy sensitivity and inhibits epithelial–mesenchymal transition by targeting twinfilin actin binding protein 1.* Journal of International Medical Research, 2019: p. 0300060519850739.
88. Bockhorn, J., et al., *MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11.* Nature communications, 2013. **4**: p. 1393.
89. Loo, J.M., et al., *Extracellular metabolic energetics can promote cancer progression.* Cell, 2015. **160**(3): p. 393-406.
90. Kurth, I., et al., *RGX-202, a first-in-class small-molecule inhibitor of the creatine transporter SLC6a8, is a robust suppressor of cancer growth and metastatic progression*. 2018, AACR.

91. Arencibia, J.M., et al., *AGC protein kinases: from structural mechanism of regulation to allosteric drug development for the treatment of human diseases.* Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics, 2013. **1834**(7): p. 1302-1321.
92. Di Cristofano, A., *SGK1: the dark side of PI3K signaling*, in *Current topics in developmental biology*. 2017, Elsevier. p. 49-71.
93. Park, J., et al., *Serum and glucocorticoid-inducible kinase (SGK) is a target of the PI 3-kinase-stimulated signaling pathway.* The EMBO journal, 1999. **18**(11): p. 3024-3033.
94. Brunet, A., et al., *Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a).* Molecular and cellular biology, 2001. **21**(3): p. 952-965.
95. Amato, R., et al., *IL-2 signals through Sgk1 and inhibits proliferation and apoptosis in kidney cancer cells.* Journal of molecular medicine, 2007. **85**(7): p. 707-721.
96. Maiyar, A.C., et al., *p53 stimulates promoter activity of the sgk serum/glucocorticoid-inducible serine/threonine protein kinase gene in rodent mammary epithelial cells.* Journal of Biological Chemistry, 1996. **271**(21): p. 12414-12422.
97. Wu, W., et al., *Microarray analysis reveals glucocorticoid-regulated survival genes that are associated with inhibition of apoptosis in breast epithelial cells.* Cancer research, 2004. **64**(5): p. 1757-1764.
98. Mi, L., et al., *The metastatic suppressor NDRG1 inhibits EMT, migration and invasion through interaction and promotion of caveolin-1 ubiquitylation in human colorectal cancer cells.* Oncogene, 2017. **36**(30): p. 4323.
99. Ma, J., et al., *Knockdown of NDRG1 promote epithelial–mesenchymal transition of colorectal cancer via NF-κB signaling.* Journal of surgical oncology, 2016. **114**(4): p. 520-527.
100. Stein, S., et al., *NDRG1 is necessary for p53-dependent apoptosis.* Journal of Biological Chemistry, 2004. **279**(47): p. 48930-48940.
101. Park, K.C., et al., *Pharmacological targeting and the diverse functions of the metastasis suppressor, NDRG1, in cancer.* Free Radical Biology and Medicine, 2019.
102. Wangpu, X., et al., *The metastasis suppressor, NDRG1, inhibits "stemness" of colorectal cancer via down-regulation of nuclear β-catenin and CD44.* Oncotarget, 2015. **6**(32): p. 33893.
103. Chang, X., et al., *NDRG1 controls gastric cancer migration and invasion through regulating MMP-9.* Pathology & Oncology Research, 2016. **22**(4): p. 789-796.
104. Liu, P., et al., *PNMA1, regulated by miR-33a-5p, promotes proliferation and EMT in hepatocellular carcinoma by activating the Wnt/β-catenin pathway.* Biomedicine & Pharmacotherapy, 2018. **108**: p. 492-499.
105. Cao, X., et al., *P4HA1/HIF1α feedback loop drives the glycolytic and malignant phenotypes of pancreatic cancer.* Biochemical and biophysical research communications, 2019.
106. Xu, R., *P4HA1 is a new regulator of the HIF-1 pathway in breast cancer.* Cell stress, 2019. **3**(1): p. 27.
107. Gilkes, D.M., et al., *Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts.* Journal of Biological Chemistry, 2013. **288**(15): p. 10819-10829.
108. Burridge, K., et al., *Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton.* Annual review of cell biology, 1988. **4**(1): p. 487-525.
109. Flores-Pérez, A., et al., *Suppression of cell migration is promoted by miR-944 through targeting of SIAH1 and PTP4A1 in breast cancer cells.* BMC cancer, 2016. **16**(1): p. 379.
110. Szczepanowska, J., *Involvement of Rac/Cdc42/PAK pathway in cytoskeletal rearrangements.* Acta biochimica Polonica, 2009. **56**(2).
111. Sacchetti, C., et al., *PTP4A1 promotes TGFβ signaling and fibrosis in systemic sclerosis.* Nature communications, 2017. **8**(1): p. 1060.

112. Dluzen, D., et al., *BCL-2 is a downstream target of ATF5 that mediates the prosurvival function of ATF5 in a cell type-dependent manner.* Journal of Biological Chemistry, 2011. **286**(9): p. 7705-7713.
113. Greene, L.A., H.Y. Lee, and J.M. Angelastro, *The transcription factor ATF5: role in neurodevelopment and neural tumors.* Journal of neurochemistry, 2009. **108**(1): p. 11-22.
114. Sheng, Z., S.K. Evans, and M.R. Green, *An activating transcription factor 5-mediated survival pathway as a target for cancer therapy.* Oncotarget, 2010. **1**(6): p. 457.
115. Miyoshi, N., et al., *Abnormal expression of TRIB3 in colorectal cancer: a novel marker for prognosis.* British journal of cancer, 2009. **101**(10): p. 1664.
116. Dong, S., et al., *Overexpression of TRIB3 promotes angiogenesis in human gastric cancer.* Oncology reports, 2016. **36**(4): p. 2339-2348.
117. Helmy, N., et al., *Oxidation of hydrogen sulfide by human liver mitochondria.* Nitric Oxide, 2014. **41**: p. 105-112.
118. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.* Molecular & Cellular Proteomics, 2014. **13**(2): p. 397-406.
119. Szabo, C., et al., *Regulation of mitochondrial bioenergetic function by hydrogen sulfide. Part I. Biochemical and physiological mechanisms.* British journal of pharmacology, 2014. **171**(8): p. 2099-2122.
120. Gibbins, M.T.G., *Metabolic and vascular effects of thiosulfate sulfurtransferase deletion.* 2018.
121. Budczies, J., et al., *Hormone receptor-dependent regulation of ABAT and beta-alanine metabolism in breast cancer*. 2013, American Society of Clinical Oncology.
122. Jansen, M.P., et al., *Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease.* Molecular oncology, 2015. **9**(6): p. 1218-1233.
123. Chen, X., et al., *Loss of ABAT-Mediated GABAergic System Promotes Basal-Like Breast Cancer Progression by Activating Ca2+-NFAT1 Axis.* Theranostics, 2019. **9**(1): p. 34.
124. Steen, H. and D. Lindholm, *Nuclear localized protein-1 (Nulp1) increases cell death of human osteosarcoma cells and binds the X-linked inhibitor of apoptosis protein.* Biochemical and biophysical research communications, 2008. **366**(2): p. 432-437.
125. Liao, A. and Y. Xiao, *Clinic significance of downexpression of LGALS4 gene in colorectal cancer.* Cancer Research and Clinic, 2011. **23**(10): p. 661-663.
126. Rodia, M.T., R. Solmi, and L. Mattia, *TSPAN8 and LGALS4 combination as blood biomarkers for colorectal cancer detection.* Cancer Cell & Microenvironment, 2016. **3**(3).