Systems Biomedicine of Colorectal Cancer Reveals Potential Targets for CRC treatment

Mehran Piran[1], 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. Patient 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 therapeutic approaches. In the present study, we conducted transcriptomic data analysis employing a systems biology method to identify potential molecular targets for colorectal cancer treatment. These targets went under investigation one by one. We proposed some genes that their expression induction or suppression alone or in combination with each other would inhibit tumor progression or metastasis based on their biological activity. They are involved in cell proliferation, energy production under hypoxic conditions, epithelial to mesenchymal transition (EMT) and angiogenesis. Employing a network analysis viewpoint, some genes were discovered that has not been reported noticeably in any kind of cancer so far. As a result, they might have some roles in progression of colorectal cancer.

Introduction

Colorectal cancer (CRC) is a major global medical burden worldwide. It is identified as the third most frequently diagnosed cancer and the forth cause of cancer death [1]. Approximately 1235108 people being diagnosed with CRC each year, and about 609051 die of CRC annually [2]. Accumulation of complex genetic interactions and environmental risks are recruited 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. Metastasis is characterized as a dynamic process in which migration of transformed tumor cells from their primary site to other regions in the body occurs [3]. One determinant function associated with the metastasis of early stage is tumor cells ability to detach from the stroma as well as motility acquisition [4]. This event happens during a process called EMT in which cells lose their epithelial characteristic including adhesion and subsequently dedifferentiate into mesenchymal mobile cells [5]. Genes that are differentially expressed (DEGs) between Primary and Metastatic groups in a gene expression dataset, would be the ones that their effect would cause the EMT and metastasis.

A large number molecular and pathway targets have been identified for the treatment of CRC through the past decades and a growing progress has been made in the development of chemotherapy and antibody drugs for CRC treatment [6]. The first priority in the treatment procedure is to determine whether tumor is primary or metastatic. Besides, its required to follow a standard approach for performing molecular testing to determine mutation status for targeted therapy. For instance, mutations in RAS/RAF pathway leads to resistance to anti-epidermal growth factor receptor antibodies in CRC [7].

Tyrosine kinases (TKs) targeting using strategies such as monoclonal antibodies and small molecule tyrosine kinase inhibitors have been identified as an effective treatment [8]. Targeting cancer-related inflammation hallmarks like IL-6/JAK/STAT3 pathway which inhibits the progression of solid tumors is another beneficial therapeutic strategy [9]. In addition, cytosolic β-catenin induction via hyperactive Wnt/β-catenin signaling pathway that mediates transactivation of T cell factor (Tcf)-dependent genes, could be a major cause of colorectal and many other types of cancer. Accordingly, devising some strategies for

disruption of this signaling pathway holds a great promise for the development of new anti-cancer treatments [10]. Inhibition of matrix metalloproteinases (MMPs) and TGFβ signaling pathway are a therapeutic approach to prevent liver metastasis [11] [12, 13]. Furthermore, PI3K inhibition suppresses lung metastasis growth in CRC patients [14, 15]. Among the known chemotherapy drugs, Cetuximab is one of the popular one which is a monoclonal antibody against epidermal growth factor receptor (EGFR) having activity against colorectal cancers that express EGFR [16]. Furthermore, Endothelial growth factor (VEGF) antibody bevacizumab is standard first-line treatment for metastatic colorectal cancer [17]. Moreover, FOLFOX and FOLFIRI are examples of the combination therapies (leucovorin, fluorouracil, and irinotecan) commonly used for treatment of CRC [18].

One way to identify molecular mechanism of pathogenesis in a biological context is to analyze transcriptomic data which is the expression values of all genes present in a biological tissue. In this study, transcriptomic data from two microarray studies each containing three groups of samples were excavated to recognize differentially expressed genes (DEGs) in any two-pair comparison between groups. The goal was to determine DEGs specific to CRC that can be targeted for cancer therapy. Healthy groups came from healthy colon tissue from people with colorectal cancer. Primary group contained gene expression data from a benign tumor and metastatic group contained expression values from metastatic tumor samples in colon and liver and lung metastasis in CRC patients. These DEGs are involved in cancer progression where proteins, mRNA, or genes themselves can be targeted (suppressed or induced) individually or in combination with one another that contributes to cancer recovery. To this end, a systems biology method with focus on network analysis was employed to obtain a number of CRC DEGs from the results of transcriptomic analysis. Then role of DEGs were interpreted biologically to see whether they could be potential targets for CRC treatment.

Methods

Database Searching and recognizing pertinent experiments

Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database was searched to detect experiments containing high-quality transcriptomic samples concordance to the study design. Searches were filtered for Homo sapiens and colorectal/colon cancer, EMT and metastasis are the keywords utilized in the search. Microarray raw data with accession numbers GSE18105 and GSE41258 were downloaded from GEO database. In the first study, all samples were from Metastatic (Stage III and IV of the disease), Primary (non-metastatic, stage II of the disease) and normal colon tissue. In the second experiment, normal, primary tumor samples (samples in stage I and II of cancer) were selected from colon tissue. Moreover, metastatic samples from liver and lung in colon cancer patients were selected.

Meta-Analysis and Identifying Differential Expressed Genes

R software was used to import and analyze the data for each dataset separately. Preprocessing step involving background correction and probe summarization was done using RMA method in "affy" package [19]. Absent probesets were also 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 method. Next, data were normalized using Quantile normalization method [20]. Then, standard deviation (SD) for each gene was computed and SDs third quartile was used as a cut off to remove low variant genes. Many to Many problem [21] which is mapping multiple probesets to the same gene symbol was solved using nsFilter function in "genefilter" package [22]. This function selects the probeset with the highest Interquartile range (IQR) to map to the gene symbol. "LIMMA" package which applies linear models on the expression matrix was utilized to discover DEGs between three groups of samples [23]. Genes with absolute log fold change larger than 0.5 and Benjamini Hochberg adjusted p-value [24] less than 0.05 were selected as the DEGs. Three sets of DEGs were obtained for each dataset (experiment). For each set, common DEGs between two datasets were selected as the final DEGs.

Network Construction

STRING database was used to generate the Interactions from all the DEGs. Using "igraph" package in R software [25], 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 [26]. To remain with a distance matrix, the new adjacency matrix was subtracted from one.

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 [27, 28]. 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 normalized 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 [28, 29].

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

Here S is the set of seed nodes and NS is the set of non-seed nodes. Therefore, a score greater than zero implies that node j falls closer on average to the seed nodes than it does on average to the rest of the network. 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.

Enrichment Analysis

Enrichment analysis was performed using Enrichr online tool [30]. 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.

Results

Data Preprocessing for Microarray Experiments

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. Figure1A illustrates the PCA plot for the samples in GSE18105 study. The same plot was created for the samples in GSE41258 study. Between the three groups of samples, a number of them are away from their sets and should be considered as outliers. To be more specific, a hierarchical clustering method introduced by Oldham MC, et al [31] 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. 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 distances average in all samples then, results of these subtractions are normalized (divided) by the standard deviation of distance averages [31]. Samples with Number-SD less than negative two which were in a distance from their cluster set in the PCA plot were regarded as an outlier. Fourteen outlier samples for GSE18105 and sixteen outlier samples for GSE41258 were found. Supplementary file1 contains information about groups of samples.

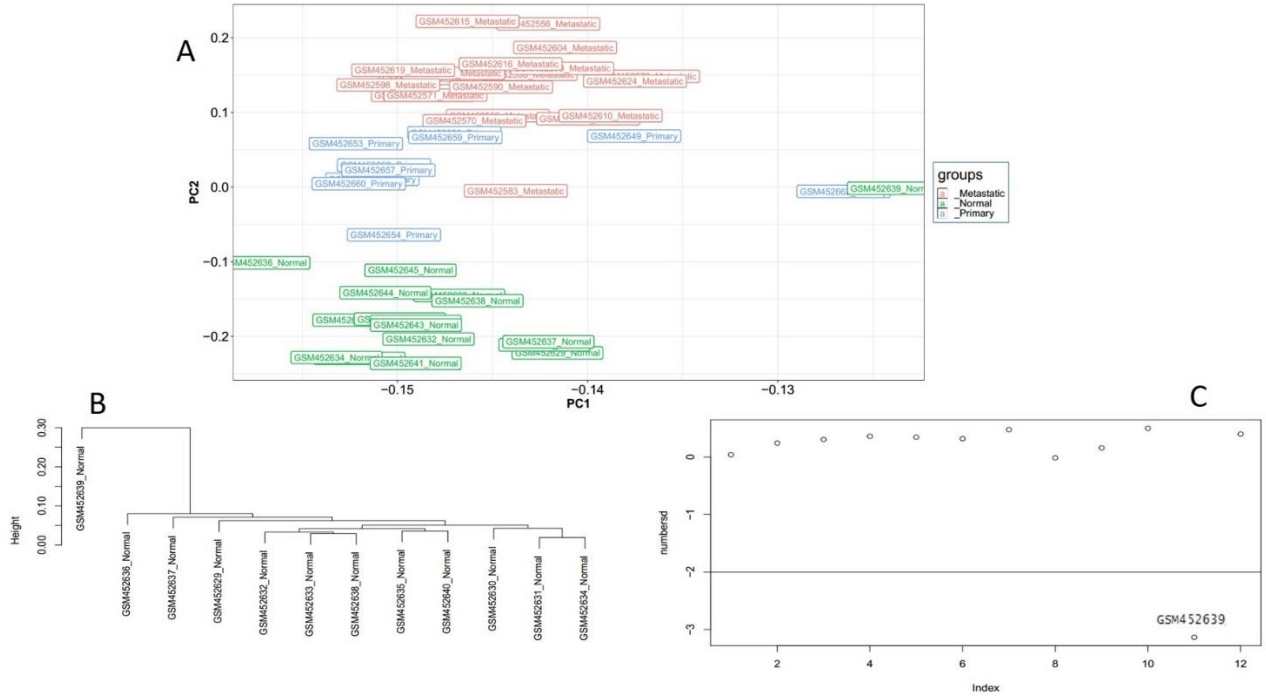


Figure 1: Illustration of outlier samples in GSE18105 study. A, is the PCA plot, B is the dendrogram for the normal samples and C is the Number-SD plot for same samples.

Figure2 illustrates the expression values for some housekeeping genes and several DEGs with positive and negative log-fold changes in GSE18105 dataset. Housekeeping genes are situated on the diagonal of the plot whilst some of the DEGs are located above or under the diagonal. This demonstrates that the preprocessed dataset is of sufficient quality for the analysis. The same was true of another dataset.

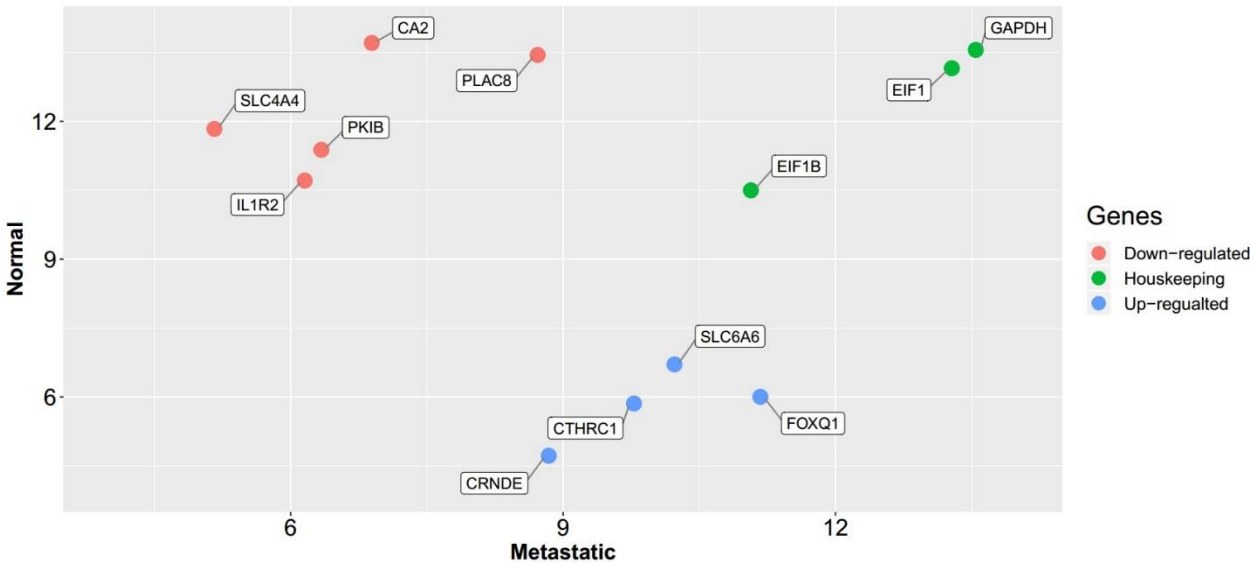


Figure2: The quality Control plot. The housekeeping genes are near the diameter of plot presenting close average values between three groups.

Undirected Protein-Protein Interaction Network

4118 unique DEGs (differentially expressed genes) with adjusted p-value < 0.05 and absolute log fold change > 5 were achieved from seven sets of DEGs. Sets are liver metastasis vs normal, liver metastasis vs primar, lung metastasis vs normal, lung metastasis vs primary and primary vs normal for GSE41258. For GSE18105 they are metastasis vs normal, metastasis vs primary and primary vs normal. In the next step, common DEGs between all metastasis vs normal analyses in two datasets were 123 number. Common DEGs between all metastasis vs primary analyses in two datasets were 13 and Common DEGs between all primary vs normal analyses in two datasets were 300. There were 242 common DEGs between these three group of DEGs which were considered as the final DEG. All DEG sets are presented in supplementary file 2. These DEGs were utilized to construct the Protein-Protein-Interaction (PPI) network. STRING database was used 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.

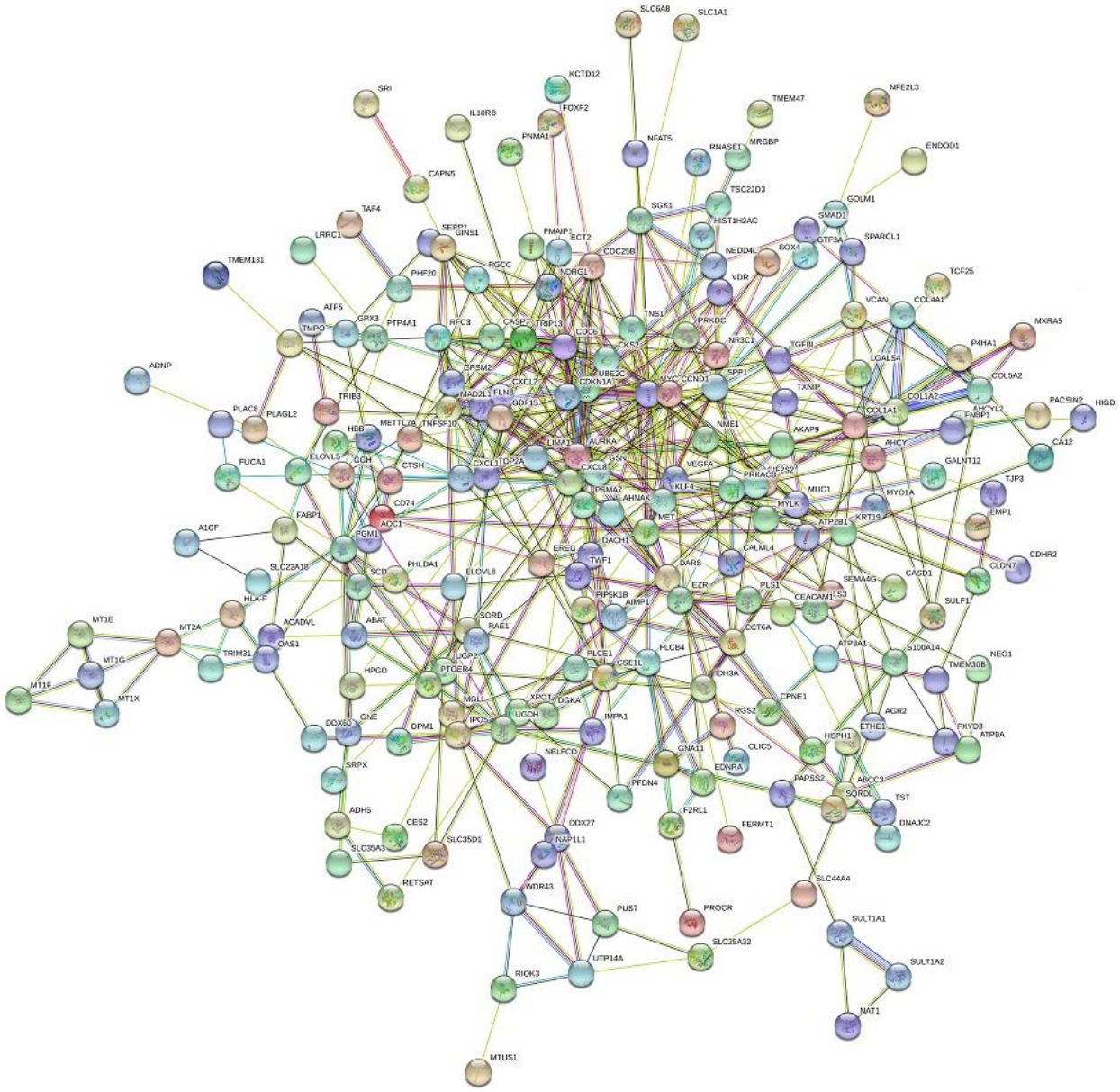


Figure3: The whole network giant component. Labels are protein/gene symbols and edges with stronger evidence are thicker. This is a scale free network [32] which follows a power law distribution (most of the network nodes have a low degree while there are few nodes with high degree).

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

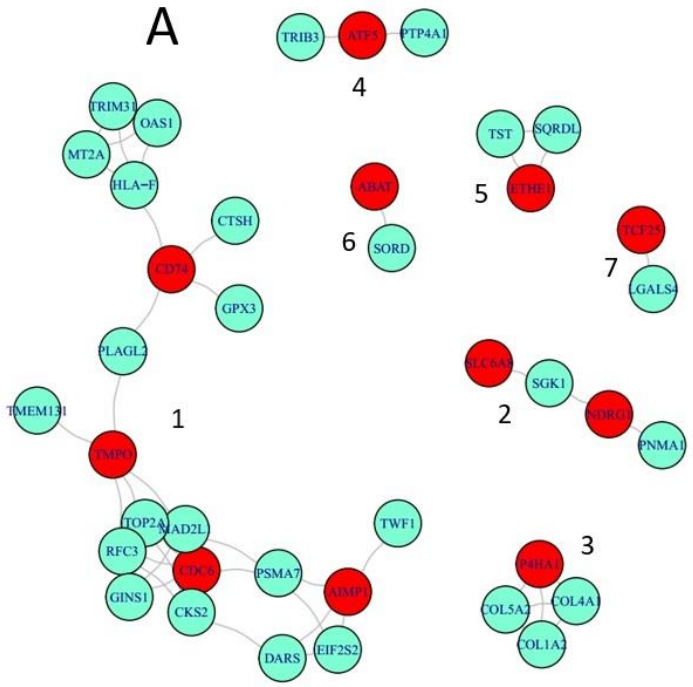
In this step, we used weights between nodes as the estimation of distances in the weighted adjacency matrix. Nodes with shorter distances from the seed genes were selected and a smaller network was extracted from the main network. Computing the shortest path score for the non-seed genes led to a network with 39 nodes comprising 12 seed nodes and 27 neighbors. This multi-component graph is called Core network and is illustrated in figure 4A. The expression states for these genes between any pair-wise comparison are depicted in Table1. For the three Metastatic-Normal comparisons (MvsN set) in the two microarray datasets, only TWF1status is not the same between the three groups. For the Primary-Normal analyses (PvsN set), all the genes have a similar status between the two datasets. Status for the three Metastatic-Primary comparisons (MvsP set) are a bit more challenging. In this set, the liver-primary and lung-primary are more similar to one another than the metastatic-primary comparison. They differ only in CD74 and ABAT genes.

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	Down
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. The State column refers to the logFC (log fold change) column and Up means gene is upregulated and Down means gene is Downregulated.



B

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

Figure4: The Core network and centralities. This network presented in A contains seven components and seed genes are in red. B, Illustrates the Degree and Betweenness centralities for the nodes in the giant component. DEGs are sorted based on the highest Degree.

Network Descriptive

The network diameter is eight containing TRIM31, HLA-F, CD74, PLAGL2, TMPO, MAD2L1 PSMA7, AIMP1 and TWF1. In the following, the giant component descriptive is explained. Transitivity is around 60%, edge density is about 18% and the mean distance is 3.48. Two important centralities, Degree and Betweenness, for the giant component, are provided in figure 4B. MAD2L1 gene has the highest Degree and a relatively high betweenness. TMPO has the highest Betweenness and a rather high degree. Similar to TMPO, its direct neighbour, PLAGL2 has a pretty high Betweenness and links the two parts of the giant component together. A common feature of the graph components is 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 file3 (S3).

Enrichment Analysis

Gene sets in the components 1 to 4 were given to *Enrichr* online tool. Tables 2A and 2B depict the enrichment analysis for the core network giant component. Enriched terms are sorted based on the highest p-value. Gastric Cancer Network2 is 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 are present in the highest number of pathways. Top2A also is involved in Gastric Cancer Network1.

In 2B, PSMA7 and MAD2L1 genes have monopolized the most GO terms to its self. After them, OSA1 and HLA_F possess the highest number of terms. GO:0060333, GO:0071346 and GO:0019221 terms are contained the highest number of genes. Since the terms are related to interferon gamma and cytokine-mediated signaling, there would be some roles 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.

Component 2, SGK1 contains the most number of pathways. NDRG1 and SGK1 are of the maximum number of GO terms and all four genes in component 2 are engaged in the regulation of apoptosis process. In component 3, all the collagen genes are engaged in Focal Adhesion-PI3K-Akt-mTOR-signaling pathway and extracellular matrix organization biological process. Component 4 gene, TRIB3, is engaged in three enriched pathways especially insulin signaling and together with ATF5 enriches the negative regulation of transcription GO terms.

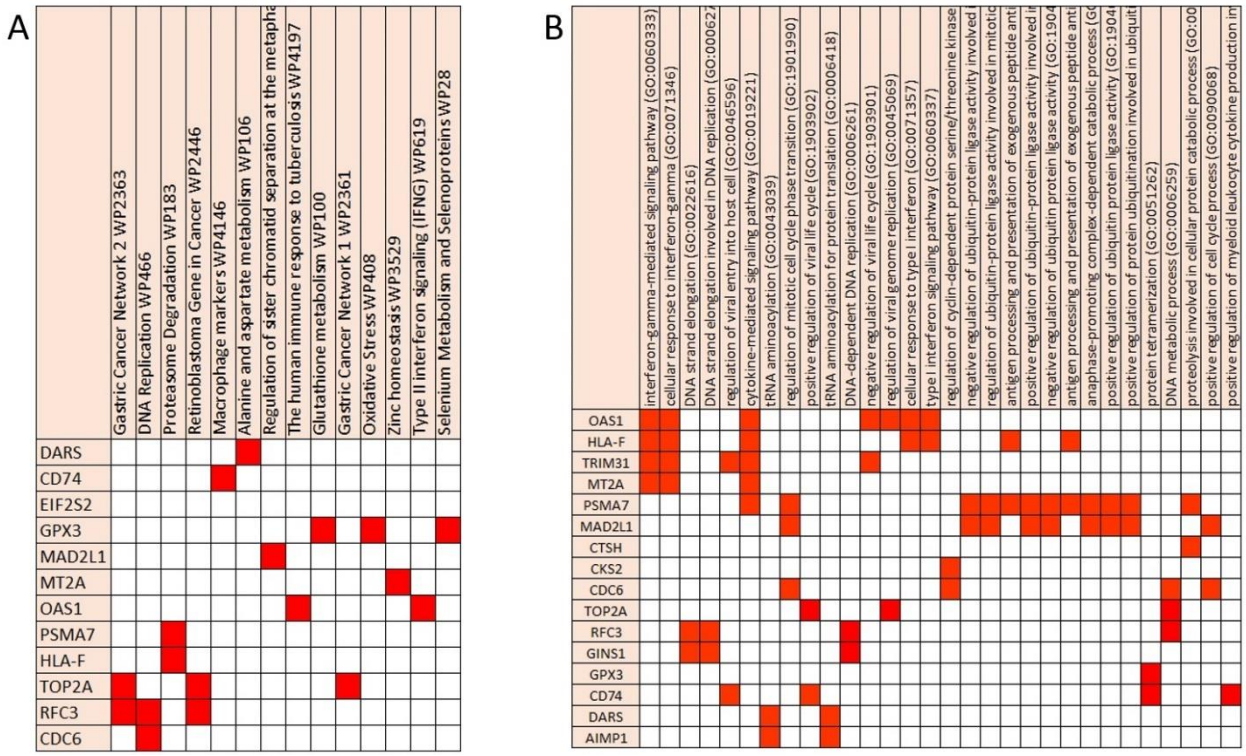


Table2: Enrichment terms for the giant component. 2A illustrates the enriched pathways from WikiPathways signaling datasets. 2B shows the enriched terms for biological processes from GO repository. Red cells reveal the involvement of the genes in the enriched elements.

Discussion

The core network of GC is composed of an up and a down part attached via PLAGL2 transcription factor (TF). The lower part is engaged mostly in cell cycle and DNA replication and the upper part is engaged in cellular apoptosis which will be explained in the following. Component two contains genes involved in ECM remodeling, component 3 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. Table 3 illustrates the expression results for obtained genes in all components of GC.

The expression of PLAGL2 gene in Balb/c3T3 cell lines leads to induction of Nip3 gene which can bind to Bcl-2 causing cellular apoptosis [33]. However, it is considered as an oncogene in different cancers, PLAGL2 anti-apoptotic role is carried out when it binds and prevents

Pirh2 proteasomal degradation. Because Pirh2 promotes proteasomal degradation of P53 protein [34]. PLAGL2 suppresses neural stem cell (NSC) differentiation by interfering in Wnt/β-catenin signaling resulting in glioblastoma [35]. Besides, PLAGL2 regulates actin cytoskeleton and cell migration by promoting actin stress fibers and focal adhesion [36]. 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 GC, it would be a pertinent target for cancer therapy.

TRIM31 (a ubiquitin ligase) was downregulated in MvsN sets while there are contradictory results in different studies where it is shown to be reduced in lung cancer cells [37] and stepped up in gastric cancer [38] and colorectal cancer [39]. 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 [40] [41]. Inhibition of this gene expression results in proliferation inhibition of CRC cells [42]; However, it was downregulated in PvsN analyses. OAS1expression is downregulated in breast ductal carcinoma, prostate adenocarcinoma, and prostate hyperplasia, at both mRNA and protein level. In addition, OAS1 expression is negatively correlated with the progression of breast and prostate cancer [43]. This gene was downregulated in MvsN analysis. HLA-F 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] [46]. HLA-F is inducible by NF-κB and is responsive to IFN-γ similar to OAS1. This gene was suppressed in MvsP analysis.

CD74 (HLA class II histocompatibility antigen gamma chain) one of the seed genes exhibited a mixed trend of expression in MvsP analysis in table 3. CD74 serves as a survival receptor expressed on normal colon epithelial cells (CEC) and colon carcinoma cells. CD74 stimulation by MIF results in enhanced cell survival by upregulation of Bcl-2 and Akt phosphorylation [47].

CTSH gene is a lysosomal cysteine protease important in the overall degradation of lysosomal proteins and was upregulated in PvsN analysis. This protease plays an important role in cytoskeletal protein Talin maturation. Talin promotes integrin activation and focal

adhesion formation leading to cell migration [PMID: 23204516]. As a result, suppression of CTSH expression could be a choice of metastasis inhibition. Glutathione peroxidase GPX3 is an important antioxidant enzyme that defends 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 [48]. It reduces H2O2 and lipid peroxides to water and lipid alcohols, respectively, and in turn oxidizes glutathione to glutathione disulfide [49]. Downregulation of GPX3, happened in PvsN analysis, leading to ascending of H2O2 level which is positively correlated with the tumor progression [50]. As a result, induction of GPX gene families would be a therapeutic approach.

Thymopoetin (TMPO/ Lamina-associated polypeptide 2 (LAP2)) gene was downregulated in liver and lung metastasis in MvsP analysis. This gene encodes a number of separate LEM domain containing protein isoforms via alternative splicing [51, 52]. TMPO is a protein that is located in the nucleus of the cells which helps to form nuclear lamina and maintenance of the nucleus membrane structure [53]. TMPO gene 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. Expression of this gene is positively correlated with pathological features of the patients such as lymph node and distant metastasis [54]. TMPO-AS1 overexpression results in depolymerization of nuclear laminas and amplification of mitosis pathways [55]. Regarding the fact that TMPO is one of the seed genes that has the greatest Betweenness centrality, it connects the top part of the GC engaged in cellular apoptosis to the bottom part of the GC engaged in cell cycle and DNA replication. Based on the given information it would be suggested that inhibition of TMPO-AS1 may result in the induction of TMPO gene and this would help to hinder the progression of cancer.

TMEM131 is a transmembrane protein which was downregulated in MvsN analysis*. No documentation was found to connect this gene to cancer.

TOP2A gene (Topoisomerase II alpha) was upregulated in PvsN analysis. In breast cancer, HER-2 along with its adjacent gene TOP2A the molecular target for several anticancer drugs, are frequently coamplified together [56]. Moreover, Copy Number Variations (CNVs) in this

gene have been identified as biomarkers in colorectal cancer [57]. DNA topological structure is controlled by this nuclear enzyme and upregulation of this gene is a cell proliferation marker in both normal and neoplastic tissues [58]. 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 [59].Therefore, rather than a target for cancer therapy, this gene could be a possible prognosis biomarker.

Mitotic Arrest Deficient 2 Like1 (MAD2L1) was upregulated in PvsN and is a mitotic spindle assembly checkpoint component that is responsible for preventing anaphase initiation until precise and complete metaphase alignment of all chromosomes takes place. Increasement in the level of BUB1B and MAD2L1 transcripts was detected in a large number of samples of the ductal breast carcinoma [60]. MAD2L1 inhibits APC (anaphase-promoting complex), but is absent when chromosomes are correctly aligned on the metaphase plate [61]. Its upregulation in our analysis would provide evidence that cancerous cells were dealing with mitotic deficiencies.

Replication Factor C subunit 3 (RFC3) roles in DNA replication, DNA damage repair, and cell cycle checkpoint control indicating the importance of this factor in all eukaryotes. Hepatocellular carcinoma and cell proliferation of ovarian tumors are suppressed by shRNA-mediated silencing of RFC3 gene [62] [63]. This gene was upregulated in PvsN analysis and it is also upregulated in Triple-negative breast cancer (TNBC) [64]. Since expression inhibition of this gene at both mRNA and protein level suppresses the migratory and invasive ability of MCF-7 cell lines [65], this gene would be a therapeutic target for colorectal cancer treatment.

The GINS complex is a DNA replication machinery component in the eukaryotes and is a requirement both for the initiation of chromosomal DNA replication and normal progression of DNA replication forks [66]. GINS1 (PSF1) expression is found to be increased in CRC patients and is a prognostic marker of CRC. GINS1 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

meaning that expression inhibition of this gene promotes apoptosis [68]. It was upregulated in PvsN analysis, therefore, would be a potential target for inhibition of colorectal cancer progression by preventing initiation of 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]. CDC6 is a key component of the DNA replication initiation machinery, and its transcription is regulated by E2F or androgen receptor (AR) alone or in combination, in PCa cells [70]. Transfection of CDC6 siRNA results in not only decreased level of ovarian cancer SKOV3 cell proliferation but also increased apoptosis rates [71]. Expression of Cdc6 and Cdt1 is upregulated in ER-positive and ER-negative breast cancer (BC) cells. Cdc6 is considered as a potential therapeutic target in in BC patients [72]. Parallel to these findings, silencing the CDC6 gene by miR-26b in gastric cancer cell line SGC7901 leads to growth inhibition but resistance to paclitaxel-based chemotherapy [73]. Our results for expression of this gene in MvsP analysis are totally contradictory to the mentioned information although, no study directly measured the expression level of this gene in CRC samples, so further investigation is needed.

CKS2 (CDC28, Cyclin-Dependent Kinases Regulatory Subunit 2) protein interacts with the catalytic subunit of the cyclin dependent kinases and its downregulation results in suppressing the expression of p-Akt and p-mTOR. Therefore, one of CSK2 oncogenic roles would be played by Akt/mTOR oncogenic pathway [74]. Compared with normal tissues, CKS2 is expressed at much higher level in CRC tissues and it has revealed that increased CKS2 expression is highly correlated with enhanced metastatic stage [75]. CKS2 upregulation has been found in BC tissues at both mRNA and protein levels. Importantly based on functional assays, 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 decreases tumor growth in vivo [76]. In the PvN analysis, this gene was upregulated which would be a therapeutic target for CRC treatment.

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 their anchorage-independent growth, cell invasion, and migration. Consequently, inhibition of PSMA7 could be a beneficial therapeutic strategy for colorectal cancer patients [79].This gene was upregulated in PvsN analysis.

AIMP proteins form a complex with aminoacyl-tRNA synthetases. Recent studies have determined that AIMP1 is the precursor of multifunctional 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 compared with normal ones suggests the tumor-suppressive roles of AIMP proteins that might hinder CRC progression [81]. This gene depicted a mixed expression trend in MvsP analysis.

DARS was found to be upregulated in MvsN analyses. This gene encodes a member of a multi-enzyme complex that its role has been proved in mediating the amino acids attachment to their cognate tRNAs. So far no study has investigated the role of this gene in cancer progression. 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].

Eukaryotic Translation Initiation Factor 2 Subunit Beta (EIF2S2/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 has been 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 analysis.

Twinfilin Actin Binding Protein 1 (TWF1/PK9) gene encodes twinfilin, an actin monomer-binding protein that promotes EMT in pancreatic cancer tissues [87]. siRNAs of TWF1 dramatically inhibited 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 (TWF1 inhibition) action [88]. In MvsN analysis, this gene was upregulated although it was suppressed in the metastasized liver and lung tissues. This may suggest that metastatic biopsies from colon are undergoing the EMT while biopsies from other tissues are not. Based on the given information, inhibition of this gene would be a good therapeutic approach since this inhibition bolsters the sensitivity to chemotherapy drugs and prevents EMT.

Creatine metabolism has been implicated in colon cancer progression and metastatic colonization of the liver. CRC cells induce and release creatine kinase-B (CKB) into the extracellular space where it phosphorylates creatine to generate the high-energy metabolite phosphocreatine. Phosphocreatine is then actively imported into metastatic cells via a specific creatine transporter SLC6A8, one of the seed genes. Intracellular phosphocreatine converts to ATP to provide the fuel for the survival of metastatic cancer cells within the hypoxic microenvironment. Consistent with these findings, a significant reduced liver colonization in mouse xenograft models has been observed due to genetic depletion of SLC6A8 in colon and pancreatic cancer cell lines. Regarding important function of this gene for cancer cell survival, expression inhibition of this gene or preventing import of phosphocreatine into the cells would be a therapeutic method to inhibit energy production of cancer cells [89]. This gene has a mixed trend in MvsP analysis.

AGC protein kinases consist of more than 60 protein kinases in the human genome, classified into 14 families. SGK1, one of the DEGs in the core network, and AKT are two families of this superfamily. SGK1 is a serine/threonine kinase that activates certain potassium, sodium, and

chloride channels [90]. SGK1 is a downstream effector of PI3K, which tries to separate targets and pathways of SGK1-dependent and independent pathways shared with AKT. The two kinases are phosphorylated and activated by PDK1 and mTORC2 complex [91, 92]. PI3K-dependent survival signals can be mediated by either Akt or SGK1 that inactivates the pro-apoptotic proteins, Bad and FKHRL1 (FOXO3a) [93]. In a study on A-498 kidney cancer cells, they found that survival signal promoted by IL-2 is mediated by SGK1 activation [94]. IL-2 activates Sgk1 but not Akt at the transcriptional level [94], Moreover, the promoter of Sgk1 is under tight control of the tumor suppressor protein p53 [95]. Sgk1 has been shown to mediate cell survival and drug resistance to platinoid and taxane compounds in women with breast cancer [96]. This gene was downregulated in the PvsN analysis and it has been still under discussion to propose a potential curative approach regarding this gene for CRC treatment.

One of the seed genes N-Myc Downstream Regulated 1 (NDRG1) is connected to SGK1 gene. This gene which has a mixed expression trend in MvsP analysis is one of the repressors of the EMT in metastatic CRC by attenuating NF-κB signaling [97]. Inhibition of NDRG1 results in the increase in NF-kB and the reduction in E-cadherin [98]. 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 [99]. 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, binds to hypoxia response element (HRE) in the promoter of NDRG1 [100]. 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 [100]. Furthermore, NDRG1 interferes with WNT/ β-catenin pathway by downregulating nuclear β-catenin and prevent metastasis in CRC [101]. In addition, NDRG1 Controls gastric cancer migration and invasion through regulating MMP-9 [102]. As a result, Enhancing NDRG1 signaling would be a promising therapeutic strategy in CRC.

Increase in expression of PNMA1 is observed in pancreatic ductal adenocarcinoma (PDAC) patients. 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. Members of the anti-apoptotic Bcl-2 family and the PI3K/AKT, MAPK/ERK pathways are likely to be involved in the pro-survival and anti-apoptotic effect of PNMA1 on PDAC patients [103]. When miR-33a-5p regulates PNMA1, subsequently proliferation and EMT are promoted in hepatocellular carcinoma by activation of the Wnt/β-catenin pathway. Upregulation of PNMA1 occurs in HCC and higher expression of PNMA1 is associated with more aggressive phenotypes [104]. This gene was downregulated in PvsN, so it is a struggle to decide whether this gene could be a therapeutic target in CRC or not.

Component 3 contains collagen (COL1A2, COL5A2 and COL4A1) and P4HA1 (a collagen hydroxylase) genes inter-connected in the process of ECM remodeling based on enrichment results. COL1A2 and COL5A2 were upregulated in MvsN and PvsN analyses. COL4A1 was upregulated in MvsN analysis and P4HA1 which is one the seed genes upregulated in MvsP analysis. P4HA1 is implicated in metastasis of breast and pancreatic cancers [105] [106]. In triple-negative Breast Cancer (TNBC), P4HA1 induces HIF-1α expression at the protein level by modulating α-ketoglutarate (α-KG) and succinate levels [106]. Correspondingly, under hypoxic conditions this gene would also trigger HIF1 pathway, cell survival and progression of colorectal cancer cells. There is a positive feedback between P4HA1 and HIF-1 in modulation of ECM. HIF-1 induces expression of genes that are encoding 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 extracellular matrix (ECM) through attachment of integrins to collagens [108]. As a result, targeting P4HA1 and P4HA2 expressions would inhibit the progression of EMT via HIF1-Collagen pathway.

PTP4A1 a member of component 4, is a protein phosphatase engaged in 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 p21-activated kinases

(PAKs) leading to formation of membrane ruffles, stress fibers and remodeling of focal adhesion complexes [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 analysis. As a result, it is challenging to regard this gene as a target for CRC treatment. Antiapoptotic gene BCL-2 is a target of the transcription factor ATF5 one of the seed genes [112]. ATF5 was induced in the liver and lung metastasis analyses and there are many evidences linking the role of ATF5 in mitochondrial dysfunction in cancer and cancer progression [113]. Survival pathways stimulated by this TF would be essential to ignore anchorage-dependent and niche-dependent cell death signals in malignant glioma, so targeting this TF may have therapeutic implications [114]. TRIB3 which is a colorectal cancer prognosis marker is a scaffold protein activated under hypoxic conditions [115]. Gastric cancerous tissues exhibit higher expression of TRIB3 than that in adjacent normal tissues. Silencing of TRIB3 downregulates VEGF-A expression in gastric cancerous cells, leading to suppression of endothelial cell recruitment and vessel formation. This gene was upregulated in MvsN analysis, 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. These genes are highly expressed in normal colon tissue compared to other tissues based on RNA-seq expression information in the Gene database of NCBI [117] because of the presence of anaerobic bacteria in the digestive tract [118]. In normal conditions, ETHE1 (persulfide dioxygenase) and SQOR are antioxidants that convert hydrogen sulfide (H2S) to persulfide and persulfide to sulfite respectively. Therefore, they defend 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. Their downregulation is essential for cancer cells proliferation and survival, since it has proved that in the hypoxic environment of CRC tumors, sulfid 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 upregulation in component 6, is one of the hallmarks of estrogen positive breast cancer. Tamoxifen is the main element in treatment of estrogen receptor (ER)-positive breast cancer as it is an agonist of ER. Two novel biomarkers for tamoxifen resistance in advanced breast cancer are ABAT and STC2 [121]. This gene showed a mix expression trend in MvsP analysis. SORD is another element of component 6 upregulated in MvsN and PvsN analyses. The connection of this gene with cancer has not efficiently investigated. TCF25 in component 7 witnessed a mixed expression trend in MvsP analysis. This gene is a member of the basic helix-loop-helix (bHLH) family of transcription factors that are vital in embryonic development. LGALS4 is implicated in regulating cell-cell and cell-matrix interactions. This gene is mostly expressed in small intestine, colon, and rectum, which is suppressed in CRC [122]. However, it is also a blood marker of CRC [123]. Like SORD gene, the expression and implications of TCF25 and LGALS4 need to be more investigated to see whether they could be therapeutic targets or not.

Summary

In this study we identified targets that alone or in combination with each other would hinder cancer progression. We also presented some DEGs as seed genes that are different in expression between primary samples and metastatic samples. The intension was to specifically find genes that is responsible for metastasis in CRC, however except for ETHE1, CDC6 and HLA-F the rest of the genes in MvsP had a mixed trend in expression. ETHE1 was the only gene that was presented in three analyses and its expression was suppressed proportionally to cancer progression. It was proposed that induction of ETHE1 would inhibit cancer cells from producing ATP from H_2S. PLAG2 is a potential target that promotes focal adhesion formation. CTSH is a protease that plays role in integrin activation via Talin and focal adhesion formation. GPx3 is an antioxidant that reduces level of stimulator of cell division H2O2. Therefore, its induction would have the therapeutic effects on CRC. TMPO is a tumor suppressor which is halted by lncRNA TMPO-AS1 in CRC. as a result, suppression of

TMPO-AS1 would pause cancer progression. TOP2A is a target in anticancer drugs. Inhibition of RFC3 as a target, would disturb cancer cell proliferation. GINS1 is another target that its inhibition would pause initiation of DNA replication machinery. CKS2 could be another target for CRC treatment which previously has been reported for BC treatment. PSMA7 is another possible target for CRC treatment that its inhibition could halt anchorage independent growth of cells. we also proposed that induction of active GSK-3β and suppression of EIF2B would have effective influence of preventing tumor growth. SLC6A8 gene plays a role in ATP production for cancer cell survival, therefore, expression suppression of this gene or preventing import of phosphocreatine into cells would be a therapeutic method to inhibit energy production of cancer cells. In spite of that NDRG1 has a mixed trend in MvsP analysis, Enhancing NDRG1 signaling would be a promising therapeutic strategy in CRC. this gene attenuates NF-κB and WNT/ β-catenin pathway. Moreover, targeting P4HA1 and P4HA2 expressions, would inhibit the progression of EMT via HIF1-Collagen pathway. TRIB3 would be pertinent target for inhibition of angiogenesis in CRC. ETHE1 and SQOR are mitochondrial genes that their expression exhibited a reduction in three kind of analyses. induction of these genes would inhibit the survival of CRC cells under hypoxic conditions. SORD upregulated in MvsN and PvsN analyses that its connection to cancer has not efficiently investigated. TCF25, DARS and TMEM131 are the genes that differentially expressed between groups normal and CRC samples. the connection of these genes to any kind cancer has not reported yet so their expression may have role in induction of CRC specially DARS and TMEM131 that have uniform expression pattern in MvsN analysis.

Conclusion

Based on the findings in this study we propose a combination targeting of the following genes would prevent CRC progression: TRIB3 for preventing angiogenesis and SLC6A8 for halting ATP production, induction of ETHE1 and SQOR to avoid cancer cells to survive from hypoxic environment, GPx induction to inhibit H_2O_2 production and NDRG1 induction to attenuate NF-κB and WNT/ β-catenin pathways to inhibit cell proliferation.

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. Costi, R., et al., *Palliative care and end-stage colorectal cancer management: the surgeon meets the oncologist.* World Journal of Gastroenterology: WJG, 2014. **20**(24): p. 7602.
4. 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.
5. Jamora, C. and E. Fuchs, *Intercellular adhesion, signalling and the cytoskeleton.* Nature cell biology, 2002. **4**(4): p. E101.
6. Al-Sohaily, S., et al., *Molecular pathways in colorectal cancer.* Journal of gastroenterology and hepatology, 2012. **27**(9): p. 1423-1431.
7. 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.
8. 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.
9. 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.
10. 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.
11. Krüger, A., et al., *Hydroxamate-type matrix metalloproteinase inhibitor batimastat promotes liver metastasis.* Cancer research, 2001. **61**(4): p. 1272-1275.
12. Katz, L.H., et al., *Targeting TGF-β signaling in cancer.* Expert opinion on therapeutic targets, 2013. **17**(7): p. 743-760.
13. 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.
14. Rychahou, P., et al., *Colorectal cancer lung metastasis treatment with polymer–drug nanoparticles.* Journal of controlled release, 2018. **275**: p. 85-91.
15. Rychahou, P., et al., *Colorectal cancer pulmonary metastasis treatment with lung-selective delivery of PAN-class I PI3K inhibitors*. 2016, AACR.
16. Jonker, D.J., et al., *Cetuximab for the treatment of colorectal cancer.* New England Journal of Medicine, 2007. **357**(20): p. 2040-2048.
17. Tol, J., et al., *Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer.* New England Journal of Medicine, 2009. **360**(6): p. 563-572.
18. Martin, M., A. VH, and T. Lu, *Colorectal Cancer Therapeutics: Present and the Future.* 2017.
19. Gautier, L., et al., *affy—analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-315.
20. 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).

21. Ramasamy, A., et al., *Key issues in conducting a meta-analysis of gene expression microarray datasets.* PLoS medicine, 2008. **5**(9): p. e184.
22. Gentleman, R., et al., *Genefilter: Methods for filtering genes from microarray experiments.* R package version, 2011. **1**(0).
23. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 2005, Springer. p. 397-420.
24. Ferreira, J. and A. Zwinderman, *On the Benjamini–Hochberg method.* The Annals of Statistics, 2006. **34**(4): p. 1827-1849.
25. Luke, D.A., *A user's guide to network analysis in R*. 2015: Springer.
26. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC bioinformatics, 2008. **9**(1): p. 559.
27. 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.
28. Nomaguchi, M. and A. Adachi, *Highly Mutable Animal RNA Viruses: Adaptation and Evolution.* Frontiers in microbiology, 2017. **8**: p. 1785.
29. Berger, S.I., A. Ma'ayan, and R. Iyengar, *Systems pharmacology of arrhythmias.* Sci. Signal., 2010. **3**(118): p. ra30-ra30.
30. 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.
31. 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.
32. Khanin, R. and E. Wit, *How scale-free are biological networks.* Journal of computational biology, 2006. **13**(3): p. 810-818.
33. Mizutani, A., et al., *A zinc-finger protein, PLAGL2, induces the expression of a proapoptotic protein Nip3, leading to cellular apoptosis.* Journal of Biological Chemistry, 2002. **277**(18): p. 15851-15858.
34. 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.
35. 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.
36. Sekiya, R., et al., *PLAGL2 regulates actin cytoskeletal architecture and cell migration.* Carcinogenesis, 2014. **35**(9): p. 1993-2001.
37. 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.
38. 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.
39. 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.
40. Ma, H., et al., *HMBOX1 interacts with MT2A to regulate autophagy and apoptosis in vascular endothelial cells.* Scientific reports, 2015. **5**: p. 15121.

41. 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.
42. 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.
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. Egawa, K., *Cancer cell-specific HLA-F antigen and a diagnostic method of cancer by using thereof*. 2004, Google Patents.
47. Maharshak, N., et al., *CD74 is a survival receptor on colon epithelial cells.* World journal of gastroenterology: WJG, 2010. **16**(26): p. 3258.
48. 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.
49. Sarsour, E.H., et al., *Redox control of the cell cycle in health and disease.* Antioxidants & redox signaling, 2009. **11**(12): p. 2985-3011.
50. Moloney, J.N. and T.G. Cotter. *ROS signalling in the biology of cancer*. in *Seminars in cell & developmental biology*. 2018. Elsevier.
51. 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.
52. 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.
53. 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.
54. 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.
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. Simon, R., et al., *HER-2 and TOP2A coamplification in urinary bladder cancer.* International journal of cancer, 2003. **107**(5): p. 764-772.
57. 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.
58. 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.

59. 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.
60. 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.
61. Percy, M.J., et al., *Expression and mutational analyses of the human MAD2L1 gene in breast cancer cells.* Genes, Chromosomes and Cancer, 2000. **29**(4): p. 356-362.
62. 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.
63. 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.
64. 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.
65. 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.
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. Zhao, B., et al., *Mir-26b inhibits growth and resistance to paclitaxel chemotherapy by silencing the CDC6 gene in gastric cancer.* Archives of medical science: AMS, 2019. **15**(2): p. 498.
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. 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.
90. 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.
91. Di Cristofano, A., *SGK1: the dark side of PI3K signaling*, in *Current topics in developmental biology*. 2017, Elsevier. p. 49-71.
92. 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.

93. 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.
94. 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.
95. 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.
96. 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.
97. 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.
98. 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.
99. Stein, S., et al., *NDRG1 is necessary for p53-dependent apoptosis.* Journal of Biological Chemistry, 2004. **279**(47): p. 48930-48940.
100. Park, K.C., et al., *Pharmacological targeting and the diverse functions of the metastasis suppressor, NDRG1, in cancer.* Free Radical Biology and Medicine, 2019.
101. 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.
102. Chang, X., et al., *NDRG1 controls gastric cancer migration and invasion through regulating MMP-9.* Pathology & Oncology Research, 2016. **22**(4): p. 789-796.
103. Jiang, S.-H., et al., *PNMA1 promotes cell growth in human pancreatic ductal adenocarcinoma.* International journal of clinical and experimental pathology, 2014. **7**(7): p. 3827.
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. 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.
118. Helmy, N., et al., *Oxidation of hydrogen sulfide by human liver mitochondria.* Nitric Oxide, 2014. **41**: p. 105-112.
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. Burger, P.M., et al., *Synaptic vesicles immunoisolated from rat cerebral cortex contain high levels of glutamate.* Neuron, 1989. **3**(6): p. 715-720.
122. 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.
123. 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).