Identification of Hub Genes to Control Cancer Stem cell Characteristics in Colon Adenocarcinoma Through Bioinformatics Analysis and Validation by Western Blotting

Yu Li

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Ye Zhang

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Yuhao Teng

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Junyi Wang

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Jingyan Zhang

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Bofan Ling

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Jie Shao

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Youran Li

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine

Baoxinzi Liu (✉ 240136491@qq.com)

Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine <https://orcid.org/0000-0001-5197-0056>

Primary research

Keywords: Colon adenocarcinoma, Cancer stem cell, mRNAsi, bioinformatics, TCGA, Biomarker

Posted Date: July 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-41765/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Cancer stem cells (CSCs) are involved in the development of cancer. This study aimed to identify hallmark genes associated with the adjustment of CSC properties.

Methods: The COAD data from The Cancer Genome Atlas (TCGA) database were assessed based on the mRNA stemness index (mRNAsi) and corrected mRNAsi. Then, both of them were analyzed with differentially expressed genes (DEGs). The gene modules were pinpointed through weighted gene co-expression network analysis (WGCNA). The genes of the most interest module were functionally annotated through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). The Search Tool for the Retrieval of Interacting Genes (STRING) was adopted to obtain protein-protein interaction (PPI) networks, and the hub genes were identified in the light of the Molecular Complex Detection (MCODE) and Cytoscape plugin cytoHubba. Upstream genes were analyzed via the DisNor. In addition, the associations between clinical variables and hub genes were estimated. The bioinformatic results were verified based on Gene Expression Profiling Integrative Analysis (GEPIA), Oncomine and Gene Expression Omnibus (GEO). Finally, the protein expression of the hub genes was verified by western blotting.

Results: Both the corrected mRNAsi and mRNAsi increased within COAD tissues relative to those within normal colon tissues, which showed a significantly decreasing trend in COAD as the clinical stage. GO indicated the biological processes, including muscle cell differentiation, multicellular organismal signaling, muscle system process and muscle contraction. KEGG revealed Tight junction, Dilated cardiomyopathy (DCM), Vascular smooth muscle contraction, and the cGMP PKG signal transduction pathway. The seven hub genes (ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN and TPM2) were identified in the brown module. Most of them were associated with clinical stage; MYL9, TAGLN and TPM2 were associated with overall survival. TGFB1, SRF, ROCK1 and PRKCA were the upstream genes through the DisNor. In the Oncomine, GEO and GEPIA databases, the expression of seven hub genes were down-regulated. In TCGA and GEPIA databases, the tendency of the seven hub genes was decreased with the advance in stage.

Conclusions: The hub genes identified in the present study might play a vital role in the preservation of COAD stem cells.

Introduction

Colorectal cancer (CRC) represents the frequently seen cancer type, which is also a cause of cancer-associated death in the world, and colon adenocarcinoma (COAD) is one of the various pathological types of CRC. About 21% CRC cases develop distant metastasis (DM) when they are diagnosed, with the 5-year survival rate of approximately 14% [1, 2]. With the rapid development of technology, many cancer treatments are available at present, such as molecular targeted drugs, surgery, radiotherapy and

chemotherapy[3]; however, disease invasion, migration and drug resistance are still the common causes of treatment failure.

A majority of cells in tumor with a great size may be non-tumorigenic end cells originating from stem cells and progenitor cells. But a small portion of tumor cells, known as cancer stem cells (CSCs)[4], have been selected from certain cancer types, and they are suggested to promote tumorigenesis, proliferation, metastasis and resistance to treatment[5–9]. Therefore, it has become a significant and profound subject to reveal the hidden molecular mechanism of CSCs in the treatment for CRC[10].

The novel cancer stemness index used to estimate the degree of oncogenic dedifferentiation was proposed by Tathiane et al. by using the innovative machine learnign algorithm referred to as the one-class logistic regression (OCLR)[11]. In this algorithm, two independent stemness indices are obtained through multi-platform analyses, of which, one (mDNAsi) provides the reflection of epigenetic features, while the other one (mRNAsi) provides the reflection of gene expression. When the activeness of CSCs and the degree of tumor cell dedifferentiation are higher, the stemness index values are higher accordingly.

The weighted gene co-expression network analysis (WGCNA) has been utilized to be the analysis means to identify clinical traits-associated co-expressed gene modules and to screen the hub genes in the network[12]. The WGCNA algorithm establishes the gene co-expression matrix with reference to the scale-free distribution of gene network. Later, a hierarchical clustering tree will be established according to the gene network adjacency coefficient, with the diverse cluster tree branches representing the diverse gene modules. Finally, this algorithm explores relationships between modules and specific phenotypes or diseases, and finally identifies target genes from the gene modules for disease treatment. The WGCNA algorithm highlights the classification of DEGs into the similarly expressed modules, so as to further analyze the signal network possibly regulating related phenotypic characteristics. Therefore, this algorithm has been widely adopted in various biological processes to examine the correlations of disease traits with gene modules, and to recognize the candidate biomarkers and therapeutic targets.

In the present study, a novel approach was adopted to analyze the COAD stemness-related genes by combining WGCNA with mRNAsi, which helped to screen the candidate targets and might provide a brand-new vision for the treatment of COAD.

Materials And Methods

Processing of data

Collection as well as pre-processing of data

Outcomes of RNA-sequencing (RNA-seq) from 514 tissue samples and 452 patients with COAD were obtained from TCGA database (<https://portal.gdc.cancer.gov>), which was accessed on March 10, 2020. The Perl language (<http://www.perl.org/>) was employed to merge the RNA-seq outcomes from 41 non-

carcinoma together with 473 cancer samples to the matrix file. Subsequently, gene names were converted from Ensembl IDs into the gene symbol matrix based on Ensembl database (<http://asia.ensembl.org/index.html>).

Acquisition of corrected mRNAsi values

Tumor tissues consisted of tumour cells, immunocytes and stromal cells, implying that tumor purity might serve as the influencing factor in evaluating mRNAsi. One article reported the application of the ESTIMATE method in assessing tumor purity, and the tumor purity in TCGA might be acquired. The corrected mRNAsi values (mRNAsi/tumor purity) were calculated according to the method reported in literature.

Combination of mRNAsi and corrected mRNAsi with COAD clinical significance

The samples correlated with mRNAsi or corrected mRNAsi of COAD were obtained. For investigating the significance of mRNAsi or corrected mRNAsi score in prognosis prediction, overall survival (OS) was analyzed based on the mRNAsi or corrected mRNAsi score by adopting the survival and survminer package in R. The statistical significance was determined through long-rank test. Then, the mRNAsi and corrected mRNAsi values within non-carcinoma tissues were compared with those in cancer tissues through the unpaired t-test. All data were incorporated in clinical stage, tumor, node and metastasis (TNM) classification. Besides, a box chart was plotted using median mRNAsi or corrected mRNAsi value for monitoring the alterations of mRNAsi or corrected mRNAsi in the presence of diverse clinical parameters.

DEGs screening

Data on DEGs expression between non-carcinoma and cancer tissues were screened using edgeR of R package. Then, the pheatmap of R package was used to draw the volcano plot and heat map. Later, the ggpubr of R package was utilized to draw the box-plot regarding DEGs for validation. For genes that had identical names, the average was calculated, whereas those with the expression of < 1 were removed.

Confirmation of significant modules

For those genes filtered, the gene co-expression network was constructed by performing WGCNA using the WGCNA of R package. DEGs with top 25% variance were screened, so as to ensure heterogeneities and the bioinformatic analysis accuracy in subsequent co-expression network analyses. Firstly, this study established one Pearson correlation matrix for those paired genes. Then, the power function $amn = |cmn|^\beta$ (where amn stands for the adjacency of gene m with gene n; while cmn stands for the Pearson correlation of gene m with gene n) was utilized to construct the weighted adjacency matrix. Typically, the soft threshold β highlighted the strong and weak associations across genes. Therefore, for classifying genes that showed akin expression in the modules, TOM-based dissimilarity was utilized for mean linkage hierarchical clustering, and the minimal module size was set at 50 in gene dendrogram. The dissimilarity was determined by further analyzing the modules, and later the module dendrograms were

established. In addition, gene significance (GS) was computed based on p-value after log₁₀ transformation ($GS = \lg p$), and it indicated linear regression of EREG-mRNAsi or mRNAsi with the gene expression. Besides, module significance (MS) represented the mean GS for a certain module, and it indicated the association of module with sample characteristics. Thereafter, those similar modules were merged according to the threshold of < 0.25 , among which, modules with the highest MS value were deemed to be most closely related to sample characteristics. When the interest modules were identified, GS value, together with the module membership (MM, association of genes within one module with expression of gene), was determined for all genes. The thresholds of $GS.mRNAsi < -0.5$ and $MM > 0.8$ were used to select key genes from one specific module.

Gene ontology (GO) along with Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis for DEGs

GO and KEGG functional analyses were performed to examine gene biological functions in the interested module by using the clusterProfiler R package[13]. The thresholds were set at $FDR < 0.05$ and $P < 0.01$.

Construction of protein-protein interaction (PPI) network, as well as identification and co-expression analysis of hub genes

For determining co-expression associations across key genes, a PPI network was constructed based on DEGs from the interested module by the database Search Tool for the Retrieval of Interacting Genes (STRING) (<https://www.string-db.org>), and the minimal required interaction score should be over 0.4. The disconnected nodes were excluded, and genes with linked nodes were retained. Later, the network relationship file was downloaded and hub genes were identified from the interacting genes in accordance with the Cytoscape plugin cytoHubba[14] and Molecular Complex Detection (MCODE)[15]. Besides, for determining co-expression associations among hub genes, the R package corplot function was adopted for calculating the transcriptional Pearson correlations. The thresholds were set at $FDR < 0.05$ and $P < 0.01$.

Data verification

Differences in hub gene expression at mRNA level between non-carcinoma and cancer samples were examined by Oncomine (<http://www.oncomine.org>). The following criteria were used to screen against Oncomine, the top 10% genes were adopted, the fold change (FC) was 2, and the p-value was $1E-4$. Generally, the Gene Expression Profiling Interactive Analysis (GEPIA) (<http://gepia.cancer-pku.cn/>) is the online approach used to analyze gene levels between non-carcinoma and cancer data collected based on the Genotype Tissue Expression (GTEx) and TCGA database. The GEPIA database was used in analyzing the survival as well as stage plots for the putative hub genes. One-way ANOVA was employed to analyze the differential expression of genes, with pathological stage being utilized to be the parameter to calculate the differential expression. In addition, $Pr(> F) < 0.05$ indicated statistical significance. For better examining hub genes, two datasets, namely, GSE25070 and GSE44076, were selected from Gene Expression Omnibus (GEO). Thereafter, Mann-Whitney U-test or t-test was conducted for statistical analyses.

Protein causality and interactions

DisNor (<https://disnor.uniroma2.it/>) was utilized for generating and investigating the PPI networks that linked disease genes from data on protein interactions annotated in Mentha and those of protein causality in SIGNOR. In this study, we chose the Multi-Protein Search module. Meanwhile, complexity level was set as the first neighbor, and it examined the protein causality.

Cell culture

The SW480 LoVo and HT-29 CRC cell lines, together with the normal human colonic epithelial NCM460 cell line, were provided by Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (Shanghai, China) and cultured within the RPMI-1640 medium containing 10% fetal bovine serum (FBS, TermoScientific HyClone, Logan, UT, USA). Then, all cells were cultivated under 5% humid CO₂ and 37 °C conditions, the medium was exchanged at an interval of 3–4 days.

Western blotting

First of all, the CRC cell lines were rinsed by the cold PBS, then RIPA buffer supplemented with the protease inhibitor cocktail (P8340, Sigma-Aldrich) was adopted for 15 min of cell lysis on ice. Later, the resultant lysate was subjected to 20 min of centrifugation at 12,000 g and 4 °C. The Bradford approach was utilized to calculate protein content, and all samples were preserved under – 80 °C. Subsequently, 20 µg protein was added in every group, which was separated through 10% or 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Afterwards, protein was transferred onto the polyvinylidene fluoride membranes (Millipore, Bedford, MA, USA), and the membranes were then blocked using bovine serum albumin under ambient temperature for 1 hour. Later, primary antibodies (1:1000) were used to incubate those membranes overnight under 4 °C, followed by another 1 h incubation with horseradish peroxidase-conjugated secondary antibodies (1:2000) under ambient temperature. At last, protein signals were detected by the ECL detection kit (Millipore, Billerica, MA), with β-actin as an internal reference.

Statistical methods

All thresholds adopted in the present work, such as corrected mRNAsi, mRNAsi, together with expression of key genes, represented the median values. Differences in corrected mRNAsi or mRNAsi score between non-carcinoma and cancer groups were evaluated through R package Wilcox test function. Differences in survival between both groups were assessed by two-sided long-rank test using the R package survival function. The associations of corrected mRNAsi or mRNAsi score with clinical factors were tested by Kruskal test using the R package. The associations between clinical factors and expression of hub genes were evaluated using Kruskal test or Wilcoxon signed-rank test. Besides, Kaplan-Meier analysis was employed to identify the OS-related clinical factors. The R software was adopted for statistical analysis.

Results

mRNAsi and corrected mRNAsi in Clinical Characteristics in COAD

Excluding stage and T, differences in mRNAsi among different N and M classifications of COAD were statistically significant (Fig. 1b-e). In addition, Cases having increased mRNAsi scores showed superior OS to those having decreased mRNAsi scores, which was contrary to our conventional understanding (Fig. 1f). After obtaining the corrected mRNAsi, its value within COAD tissues was analyzed again, and the new finding was obtained, which was that the mRNAsi index in COAD tissues apparently increased following purity correction (Fig. 2a). In addition, the associations of clinical traits with corrected mRNAsi were re-associated, which also came to changed results. In addition to T, the corrected mRNAsi at different stages, N and M classifications of COAD showed distinct differences (Fig. 2b-e), which suggested that COAD cells exhibited strong CSC characteristics at the earlier stage and were prone to invasion and metastasis. At the same time, the OS was shorter for patients with higher corrected mRNAsi scores than that for those with lower scores, although the P-value was not significant (Fig. 2f).

DEGs screening

Since mRNAsi showed significant differences between non-carcinoma and cancer tissues, it was suspected that there might be certain DEGs regulating cancer cell stemness. Based on the analysis results, 6,477 DEGs were screened, among which 1,916 were down-regulated, and 4,561 were up-regulated (Fig. 3a). Additionally, the top 20 up-regulated or down-regulated DEGs were extracted and displayed in a heat-map (Fig. 3b).

WGCNA: Screening of genes and modules with the highest significance

The co-expression network of genes was constructed through WGCNA by using the module eigengenes (MEs), the leading part of principal component analysis for every gene module that sheds more lights on the candidate COAD stemness-related genes. After eliminating the outliers, DEGs showing the top 25% variance identified through cluster analysis were incorporated into the module. The soft threshold of $\beta = 4$ (scale-free $R^2 = 0.950$) was selected in the present work for guaranteeing the scale-free network, finally, 11 modules were obtained in later analyses (Fig. 3c). For examining the associations of modules with sample stemness indexes, MS was utilized as gene level for related module, so as to determine the associations with the clinical phenotypes (Fig. 3d). The red and brown modules showed significantly negative associations with mRNAsi, whereas the green module displayed significantly positive correlations with mRNAsi (Fig. 3e-g). In the meantime, the module – trait relationship score of brown module was -0.72 . Thus, the brown module was regarded as the most interested module and utilized for later analyses.

Moreover, the following criteria were used to screen key genes from mRNAsi group, including $\text{cor. GS} < -0.5$ and $\text{cor. MM} > 0.8$. As a result, altogether 124 key genes were screened.

GO together with KEGG pathway enrichment analyses for DEGs

The clusterProfiler R package was used to illuminate the functional resemblances of the module genes. For brown module genes, GO as well as KEGG enrichment analyses were carried out (Fig. 4a, b). Notably, GO terms can be classified as biological processes (BPs), molecular functions (MFs) and cellular components (CCs). For BPs of GO term, those DEGs were mainly enriched into the multicellular organismal signaling, muscle contraction, muscle cell differentiation, and muscle system process. With regard to MFs of GO term, those DEGs were remarkably enriched into the chemokine binding, muscle alpha actinin binding, actin binding, and calcium ion transmembrane transporter activity. As for CCs of GO term, those DEGs were mainly enriched into the sarcoplasmic reticulum, contractile fiber, sarcomere and myofibril. Results of GO analyses are presented. According to KEGG analysis results, DEGs were significantly enriched into the Vascular smooth muscle contraction, cGMP PKG signal transduction pathway, Tight junction and Dilated cardiomyopathy (DCM).

PPI network construction and hub gene certification

The interactions between key genes were determined by the online tool STRING and the Pearson correlation. A total of 124 DEGs were filtered into the PPI network, which comprised 65 nodes and 91 edges (Fig. 4c). Then, the PPI network was visualized by using the Cytoscape software. In addition, the cytoHubba app was employed to extract those 10 most significant hub genes (Fig. 4d), meanwhile, three significant modules (modules 1, 2 and 3) were screened by MCODE (Fig. 4e-g). Based on this result, seven common network genes (including CALD1, ACTA2, MYL9, LMOD1, TAGLN, MYLK, TPM2) were recognized to be hub genes for later analysis and verification, the analysis showed that those screened genes were distinctly lowexpressed in COAD samples (Fig. 4h).

Associations between hub gene expression and clinical parameters

The clinical data from 452 COAD patients in TCGA were analyzed. It was observed from the figure that, almost all hub genes were notably up-regulated, which were associated with the clinical stage, T, N and M classifications. As shown in figures, ACTA2, LMOD1, MYL9, TAGLN and TPM2 had a link with tumor stages, ACTA2 and TAGLN were related to tumor classification; ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN and TPM2 were also connected with node classification; LMOD1 and TPM2 were related to metastasis classification; However, MYL9, TAGLN and TPM2 were associated with OS (Fig. 5) (Fig. 6).

Hub gene analysis and verification

Hub gene expression was observed based on Oncomine to validate the data. By comparison with non-carcinoma samples, it was found that most of the seven hub genes were down-regulated not only in COAD but also in many other cancer types (Fig. 7a). However, there was no difference in CALD1 expression between non-carcinoma and cancer issues in GSE44076. To systematically understand hub gene expression within multiple cancer types, the hub gene expression matrix was constructed using GEPIA, which revealed the down-regulation of these hub genes at pan-cancer scale (Fig. 7b), including COAD (Fig. 7c-i). To explore the candidate hub genes, the expression profiles of hub genes were validated using another two GEO datasets (GSE25070 and GSE44076). The results showed that all the above

seven hub genes were down-regulated in cancer tissues (Fig. 7j, k), which was consistent with the results in TCGA database. At the same time, hub gene expression was further analyzed in COAD samples of diverse pathological tumor stages and OS concurrently. According to our results, the expression of ACTA2, LMOD1, MYL9, TAGLN and TPM2 was significantly up-regulated with the advance in tumor stage, as shown in the violin plot; besides, and MYL9, TAGLN and TPM2 were associated with the OS (Fig. 8).

Protein interactions and causality

The significant and strong correlation between the seven hub genes were displayed at transcriptional level with the high correlation score (Fig. 9a). The hub gene first neighbors were revealed by DisNor, and Serum response factor (SRF), Transforming growth factor beta-1 (TGFB1), Protein kinase C alpha type (PRKCA), and Rho-associated protein kinase 1 (ROCK1) were identified as the upstream genes that up-regulated or down-regulated hub genes. In addition, STRING was utilized to visualize the broad and strong relationships between hub genes and those upstream genes (Fig. 9b, c).

Validation of hub genes by Western blotting

According to bioinformatics analysis of retrospective data that potentially exerted considerable parts during the maintenance of COAD stem cells, the seven hub genes (ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN and TPM2) were finally selected as the potential targets in Western blotting analysis. As shown in Fig. 10, the expression levels of ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN and TPM2 were obviously down-regulated within CRC cells compared with those in NCM460 ($P < 0.001$), indicating that bioinformatics analysis results were cogent.

Discussion

COAD is a disease that is prone to relapse and metastasis, which is characterized by its high incidence and mortality rates, as well as treatment resistance. There have been many studies on the diagnosis and treatment of COAD, but the therapeutic effect is still unsatisfactory yet, and the disease is likely to develop drug resistance, relapse and metastasis. In recent years, some literature reports the vital part of CSCs in treatment resistance, tumor recurrence and metastasis. As a result, it will be significant to explore the CSC mechanism at molecular level and to develop targeted intervention within stem cells. Identifying core genes prevents stem cells from turning from the resting state to the active state. In the present study, the mRNAsi values between non-carcinoma and cancer tissues were explored, which suggested that the tumor tissue stemness increased compared with that in non-carcinoma samples.

Considering that the tumor tissue contained multiple cell types, including malignant tumor cells, stromal cells and immune cells, thus, for eliminating mRNAsi deviation resulted from normal cells, the tumor purity method was used to modify the mRNAsi scores. As shown by our results, those corrected mRNAsi scores within cancer samples increased relative to those in non-carcinoma samples. Besides, those corrected mRNAsi scores decreased with the increases in tumor stage, N and M classifications, and the differences were statistically significant, but no significant differences were detected in terms of survival.

The results indicated that COAD cells showed higher stem cell properties at an early stage, and they were prone to invasion, metastasis and drug resistance.

In addition, WGCNA was further applied in classifying those DEGs to multiple clusters according to the similar expression patterns, so as to better examine the associations of diverse gene clusters with the clinical features. Using this algorithm, it was discovered that more than one gene module was closely related to mRNAsi. Besides, genes in the green, red, and brown modules were closely related to mRNAsi, and the differences were statistically significant. Thus, the brown module was selected as the interested module for subsequent analysis. Then, those DEGs in brown module were carried out GO as well as KEGG analyses for subsequent research.

For BPs of GO term, those DEGs were mostly enriched into the muscle contraction, muscle system process, muscle cell differentiation and multicellular organismal signaling. It is reported in literature that, these BPs are related to stem cell properties. For instance, Christoph Patsch et al reported that human pluripotent stem cells rapidly and efficiently differentiated to vascular endothelial as well as smooth muscle cells[16]. In addition, Gang Wang et al. mentioned that cachexia caused muscle cell differentiation in patients with metastatic cancer, resulting in the severe loss of skeletal muscle mass and function[17]. Besides, Alejandro Sanchez Alvarado proposed that, differentiation of cells was the process required in the growth, development, longevity and reproduction for multicellular organisms, which was a natural biological process in stem cells[18]. For MFs of GO term, those DEGs were mostly enriched into calcium ion transmembrane transporter activity, actin binding, chemokine binding, and muscle alpha actinin binding. A previous study shows that, the actin-binding proteins are involved in cell migration and micrometastasis[19]. Moreover, the aberrant TRPV6 expression, which plays a critical role in calcium uptake in epithelial tissues, is associated with a variety of diseases, including cancers[20]. Chemokines are the multifunctional mediators that exert vital parts in controlling angiogenesis and inflammation in the context of neoplastic and chronic inflammatory disorders[21], Chemokine expression in melanoma metastases is associated with T-cell recruitment[22].

In addition, the KEGG pathway analysis results, including Vascular smooth muscle contraction, cGMP PKG signal transduction pathway, Tight junction and Dilated cardiomyopathy (DCM), were also obtained. A review reports that the cGMP signaling might be utilized to be the target for preventing and treating breast cancer[23]. Besides, experimental results demonstrate that regulating the tight junction integrity restricts colitis and tumorigenesis[24]. In this study, through the analyses of STRING online database and Cytoscape software, seven common network hub genes were identified. Among them, four upstream genes (TGFB1, SRF, PRKCA, ROCK1) were identified using DisNor.

Many studies have confirmed that the Hedgehog, Wnt/ β -catenin, and Notch signal transduction pathways are frequently involved in regulating the CSCs self-renewal. For instance, TGFB1 in the Wnt/ β -catenin pathway confers chemoresistance-associated metastasis in NSCLC[25]. SRF in the Notch pathway was identified to play a hub role in invasive melanoma cells[26], SRF in the Hedgehog pathway promotes drug resistance in basal cell carcinomas[27], and associated with breast cancer stemness[28]. Biologically,

ROCK1 has been recently implicated in tumor progression[29]. The parallel activation of PDPK1 and PRKCA results in treatment resistance in CRC[30, 31]. Such genes may serve as the potential therapeutic targets in regulating CSC proliferation, differentiation and self-renewal. In TCGA database, expression of those seven hub genes decreased within cancer tissues relative to that within non-carcinoma tissues, as shown by our results. Using the Oncomine and GEPIA databases, it was validated that the hub genes related to stemness were down-regulated within multiple cancer tissues. For further verification, those seven hub gene expression levels were verified with two GEO datasets, including GSE25070 and GSE44076. As expected, all of these hub genes were down-regulated in COAD tissues compared with that in normal tissues. Subsequently, the associations of hub gene expression with clinical data (including the clinical stage, tumor, node and metastasis (TNM) classifications of the 452 COAD patients) were analyzed. According to our results, ACTA2, LMOD1, MYL9, TAGLN and TPM2 were notably associated with clinical stage, whereas ACTA2 and TAGLN were related to tumor classification; ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN and TPM2 were also linked with node classification; and LMOD1 and TPM2 were related to metastasis classification. For those seven hub genes, their expression gradually decreased with the advance in clinical stage, which demonstrated that COAD cells showed a strong stemness at an earlier stage. Moreover, the expression of TAGLN, TPM2 and MYL9 was associated with OS. All these relationships were further validated through the GEPIA database; surprisingly, the results were similar to the previous results. Six of the seven genes are suggested to participate in malignancy; however, LMOD1 has not been reported. Patients with lung adenocarcinoma who have up-regulated ACTA2 expression are more likely to develop distant metastasis and have detrimental prognosis[32]. Besides, the expression level of CALD1 is higher in patients with rectal cancer who do not respond to the neoadjuvant radiochemotherapy[33]. Blockade of MYL9 function is a critical factor for reducing the matrix-remodelling and invasion-promoting abilities of cancer-associated fibroblasts[34]. Tumor-associated fibroblasts promote the resistance of tumor cells to chemotherapy[35]. In addition, the high expression level of MYLK promotes the migration and invasion of gastric cancer cells[36]. TGF- β induces the up-regulation of TAGLN, which is associated with CRC progression. Additionally, Mona Elsafadi et al. suggested TAGLN as a potential prognostic marker associated with the advanced pathological stage of CRC[37]. Serum testing demonstrates that TPM2 significantly increased in ovarian cancer patients compared with non-cancer controls[38]. Interestingly, it was discovered in the present study that, these gene expression within tumor samples decreased relative to that within non-carcinoma samples, and the expression levels of these genes gradually increased with the progression in STAGE, T, N, and M classifications. Probably, these genes in COAD possessed similar functions to those of TGFB, which inhibited tumor progression in normal tissues but promoted tumor progression in tumor tissues[39].

Conclusions

In conclusion, seven hub genes (ACTA2, CALD1, LMOD1, MYL9, MYLK, TAGLN, TPM2) possibly exert vital parts during COAD stem cell maintenance, which may become the therapeutic targets for suppressing the stem characteristics. In addition, regulation of upstream genes TGFB1, SRF, PRKCA and ROCK1 may be helpful for the treatment of COAD. However, results in this study are obtained from bioinformatic analysis

and superficial experiments, and further biological studies are warranted to elucidate the internal mechanisms of COAD stemness.

Abbreviations

CSC

Cancer stem cell; COAD:colon adenocarcinoma; TCGA:The Cancer Genome Atlas; mRNAsi:mRNA stemness index; DEGs:differentially expressed genes; GO:Gene Ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; WGCNA:Weighted Gene Co-expression Network Analysis; STRING:Search Tool for the Retrieval of Interacting Genes; PPI:Protein-Protein Interaction; MCODE:Molecular Complex Detection; GEPIA:Gene Expression Profiling Integrative Analysis; GEO:Gene Expression Omnibus; BP:biological process; CC:cellular component; MF:molecular function; DCM:Dilated Cardiomyopathy; OS:overall survival; CRC:Colorectal cancer; OCLR:one-class logistic regression; RNA-seq:RNA-sequencing; GS:gene significance; MS:module significance; MM:module membership; GTEX:Genotype Tissue Expression; FBS:fetal bovine serum; SRF:Serum response factor; TGFβ1:Transforming growth factor beta-1; PRKCA:Protein kinase C alpha type; ROCK1:Rho-associated protein kinase 1.

Declarations

Authors contributions

YL, BL and YT collected the data; YZ and JW performed the statistical analysis; JZ, JS and BL wrote the paper; YL and BL conceived the study. All authors read and approved the final manuscript.

Author details

¹ Department of Medical Oncology, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, 155# Han Zhong Street, Qinhuai District, 210029, Nanjing, China.

² Department of Anorectal, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, 155# Han Zhong Street, Qinhuai District, 210029, Nanjing, China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The authors declare that the data supporting the findings of this study are available within the article.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the National Natural Science Foundation for Youth of China (81904204)

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
2. Siegel RL, Miller KD, Goding SA, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A. **Colorectal cancer statistics, 2020.** *CA Cancer J Clin* 2020.
3. Wu C. Systemic Therapy for Colon Cancer. *SURG ONCOL CLIN N AM*. 2018;27(2):235–42.
4. Jordan CT, Guzman ML, Noble M. Cancer stem cells. *N Engl J Med*. 2006;355(12):1253–61.
5. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *NAT REV CANCER*. 2008;8(10):755–68.
6. Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. *ANNU REV MED*. 2007;58:267–84.
7. Colak S, Medema JP. Cancer stem cells—important players in tumor therapy resistance. *FEBS J*. 2014;281(21):4779–91.
8. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *NAT REV CLIN ONCOL*. 2017;14(10):611–29.
9. Lytle NK, Barber AG, Reya T. Stem cell fate in cancer growth, progression and therapy resistance. *NAT REV CANCER*. 2018;18(11):669–80.
10. Vaiopoulos AG, Kostakis ID, Koutsilieris M, Papavassiliou AG. Colorectal cancer stem cells. *STEM CELLS*. 2012;30(3):363–71.
11. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O et al: **Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation.** *CELL* 2018, **173(2):338–354.**
12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC BIOINFORMATICS*. 2008;9:559.
13. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
14. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC SYST BIOL*. 2014;8(Suppl 4):11.

15. Kellman P, Chung YC, Simonetti OP, McVeigh ER, Arai AE. Multi-contrast delayed enhancement provides improved contrast between myocardial infarction and blood pool. *J MAGN RESON IMAGING*. 2005;22(5):605–13.
16. Patsch C, Challet-Meylan L, Thoma EC, Urich E, Heckel T, O'Sullivan JF, Grainger SJ, Kapp FG, Sun L, Christensen K et al: **Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells**. *NAT CELL BIOL* 2015, **17(8):994–1003**.
17. Wang G, Biswas AK, Ma W, Kandpal M, Coker C, Grandgenett PM, Hollingsworth MA, Jain R, Tanji K, Lomicronpez-Pintado S et al: **Metastatic cancers promote cachexia through ZIP14 upregulation in skeletal muscle**. *NAT MED* 2018, **24(6):770–781**.
18. Sanchez AA, Yamanaka S. Rethinking differentiation: stem cells, regeneration, and plasticity. *CELL*. 2014;157(1):110–9.
19. Haffner MC, Esopi DM, Chaux A, Gurel M, Ghosh S, Vaghasia AM, Tsai H, Kim K, Castagna N, Lam H et al: **AIM1 is an actin-binding protein that suppresses cell migration and micrometastatic dissemination**. *NAT COMMUN* 2017, **8(1):142**.
20. McGoldrick LL, Singh AK, Saotome K, Yelshanskaya MV, Twomey EC, Grassucci RA, Sobolevsky AI: **Opening of the human epithelial calcium channel TRPV6**. *NATURE* 2018, **553(7687):233–237**.
21. Romagnani P, Lasagni L, Annunziato F, Serio M, Romagnani S. CXC chemokines: the regulatory link between inflammation and angiogenesis. *TRENDS IMMUNOL*. 2004;25(4):201–9.
22. Harlin H, Meng Y, Peterson AC, Zha Y, Tretiakova M, Slingluff C, McKee M, Gajewski TF. Chemokine expression in melanoma metastases associated with CD8 + T-cell recruitment. *CANCER RES*. 2009;69(7):3077–85.
23. Windham PF, Tinsley HN. cGMP signaling as a target for the prevention and treatment of breast cancer. *SEMIN CANCER BIOL*. 2015;31:106–10.
24. Sharma D, Malik A, Guy CS, Karki R, Vogel P, Kanneganti TD: **Pyrin Inflammasome Regulates Tight Junction Integrity to Restrict Colitis and Tumorigenesis**. *GASTROENTEROLOGY* 2018, **154(4):948–964**.
25. Cai J, Fang L, Huang Y, Li R, Xu X, Hu Z, Zhang L, Yang Y, Zhu X, Zhang H et al: **Simultaneous overactivation of Wnt/ β -catenin and TGF β signalling by miR-128-3p confers chemoresistance-associated metastasis in NSCLC**. *NAT COMMUN* 2017, **8(1)**.
26. Manning CS, Hooper S, Sahai EA. Intravital imaging of SRF and Notch signalling identifies a key role for EZH2 in invasive melanoma cells. *ONCOGENE*. 2015;34(33):4320–32.
27. Whitson RJ, Lee A, Urman NM, Mirza A, Yao CY, Brown AS, Li JR, Shankar G, Fry MA, Atwood SX et al: **Noncanonical hedgehog pathway activation through SRF-MKL1 promotes drug resistance in basal cell carcinomas**. *NAT MED* 2018, **24(3):271–281**.
28. Kim T, Yang SJ, Hwang D, Song J, Kim M, Kyum KS, Kang K, Ahn J, Lee D, Kim MY et al: **A basal-like breast cancer-specific role for SRF-IL6 in YAP-induced cancer stemness**. *NAT COMMUN* 2015, **6:10186**.

29. Lefort K, Mandinova A, Ostano P, Kolev V, Calpini V, Kolfshoten I, Devgan V, Lieb J, Raffoul W, Hohl D et al: **Notch1 is a p53 target gene involved in human keratinocyte tumor suppression through negative regulation of ROCK1/2 and MRCKalpha kinases.** *Genes Dev* 2007, **21(5):562–577**.
30. Coppe JP, Mori M, Pan B, Yau C, Wolf DM, Ruiz-Saenz A, Brunen D, Prahallad A, Cornelissen-Steijger P, Kemper K et al: **Mapping phospho-catalytic dependencies of therapy-resistant tumours reveals actionable vulnerabilities.** *NAT CELL BIOL* 2019, **21(6):778–790**.
31. Lonne GK, Cornmark L, Zahirovic IO, Landberg G, Jirstrom K, Larsson C. PKCalpha expression is a marker for breast cancer aggressiveness. *MOL CANCER*. 2010;9:76.
32. Lee HW, Park YM, Lee SJ, Cho HJ, Kim DH, Lee JI, Kang MS, Seol HJ, Shim YM, Nam DH et al: **Alpha-smooth muscle actin (ACTA2) is required for metastatic potential of human lung adenocarcinoma.** *CLIN CANCER RES* 2013, **19(21):5879–5889**.
33. Chauvin A, Wang CS, Geha S, Garde-Granger P, Mathieu AA, Lacasse V, Boisvert FM. The response to neoadjuvant chemoradiotherapy with 5-fluorouracil in locally advanced rectal cancer patients: a predictive proteomic signature. *Clin Proteomics*. 2018;15:16.
34. Calvo F, Ege N, Grande-Garcia A, Hooper S, Jenkins RP, Chaudhry SI, Harrington K, Williamson P, Moeendarbary E, Charras G et al: **Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts.** *NAT CELL BIOL* 2013, **15(6):637–646**.
35. Hu JL, Wang W, Lan XL, Zeng ZC, Liang YS, Yan YR, Song FY, Wang FF, Zhu XH, Liao WJ et al: **CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer.** *MOL CANCER* 2019, **18(1):91**.
36. Xia N, Cui J, Zhu M, Xing R, Lu Y. Androgen receptor variant 12 promotes migration and invasion by regulating MYLK in gastric cancer. *J PATHOL*. 2019;248(3):304–15.
37. Elsafadi M, Manikandan M, Almalki S, Mahmood A, Shinwari T, Vishnubalaji R, Mobarak M, Alfayez M, Aldahmash A, Kassem M et al: **Transgelin is a poor prognostic factor associated with advanced colorectal cancer (CRC) stage promoting tumor growth and migration in a TGFbeta-dependent manner.** *CELL DEATH DIS* 2020, **11(5):341**.
38. Tang HY, Beer LA, Tanyi JL, Zhang R, Liu Q, Speicher DW. Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer. *J PROTEOMICS*. 2013;89:165–78.
39. Derynck R, Akhurst RJ, Balmain A. TGF-beta signaling in tumor suppression and cancer progression. *NAT GENET*. 2001;29(2):117–29.

Figures

Figure 1

Figure 1

Figure 2

Figure 2

Figure 3

Figure 3

Figure 4

Figure 4

Figure 5

Figure 5

Figure 6

Figure 6

Figure 7

Figure 7

Figure 8

Figure 8

Figure 9

Figure 9

a

b

Figure 10

Figure 10