COLGALT1 is a Prognostic Biomarker in Clear Renal Cell Carcinoma Correlated with Immune Infiltrates: A Study Based on TCGA Data

Bangmin Han (hanbm@163.com)
   Shanghai General Hospital Department of Orthopaedics: Shanghai Jiaotong University First People's Hospital Department of urology

Shiwei Liu
   Shanghai General Hospital Urology Center

Qianwei Xing
   Affiliated Hospital of Nantong University

Yang Yu
   Shanghai General Hospital Urology Center

Yi Wang
   Affiliated Hospital of Nantong University

Research article

Keywords: COLGALT1, Clear cell renal cell carcinoma, Metabolism, Overall survival, Risk score, Immune infiltration, Tumor microenvironment

DOI: https://doi.org/10.21203/rs.3.rs-86622/v1

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

Background: COLGALT1, as one gene enriched in metabolic pathways, which may be related to the tumorigenesis and progression. We aim to explore the potential value of COLGALT1 in clear cell renal cell carcinoma (ccRCC) through the study.

Methods: We searched The Cancer Genome Atlas (TCGA) database to collect ccRCC patients’ information including clinicopathologic parameters and COLGALT1 gene expression. We also validated the COLGALT1 mRNA expression by qRT-PCR. Then, We evaluated the relationship between COLGALT1 and overall survival (OS) by the Cox regression analyses. Gene Set Enrichment Analysis (GSEA) was utilized to compare between tissues with different COLGALT1 expression levels. Microsatellite Instability (MSI), Tumor Mutational Burden (TMB) and Neoantigen were evaluated through different tools. By TIMER, correlations between COLGALT1 and immune cell infiltrations were analyzed. The ESTIMATE algorithm was used to calculate the estimate, stromal and immune scores for ccRCC. Finally, CIBERSORT was carried out to explore the connection between the COLGALT1 and the tumor immune microenvironment.

Results: Significant gene expression of COLGALT1 was identified between normal and ccRCC tissues. Multivariate analysis indicated that high expression of COLGALT1 was linked to poor OS ($P = 0e+00$). GSEA results demonstrated that high COLGALT1 expression was associated with metabolic pathways. COLGALT1 was identified to be one independent prognostic factor through the univariate and multivariate Cox regression analyses. One nomogram was integrated including both the clinicopathologic variables and COLGALT1 expression to provide a quantitative approach to clinicians for predicting prognostic risk. Further, we find out some genes which are significantly correlated with COLGALT1. Besides, MSI and TMB showed strong correlations with COLGALT1 in ccRCC. Also, the correlations between COLGALT1 with immune infiltrations were found in ccRCC. Finally, immune microenvironment including immune checkpoint molecules, immune cells and mismatch repair protein were proved to be linked to COLGALT1 in ccRCC.

Conclusions: Our results revealed that COLGALT1 could act as a favorable prognostic factor for ccRCC. Besides, this study also provided one method to determine the immune infiltration of patients and some signal pathways which are potential regulated by COLGALT1 in ccRCC.

Introduction

Renal cell carcinoma (RCC) is one of the ten most common cancers in both males and female around the world [1]. Clear cell renal cell carcinoma (ccRCC), as the predominant histology of renal cell carcinoma, accounts for 75% of all cases and is the main lethal type, the overall 5-year survival rate of RCC is 74% [2]. Nowadays, the overall prognosis of ccRCC is still poor especially for those who were diagnosed with advanced stages although the 5-year survival rates have been improved [3].
As we all know, Tumor stage (TNM) is one of the important prognostic factors for patients with malignant tumor [4]. According to these reports, the 5-year recurrence free-survival of patients with stage I over 92%, while for those patients with stage II and III is only 40% [5, 6]. The UISS, one authoritative organization has been developed which classify patients into low-risk, medium-risk, and high-risk prognostic groups based on the clinicopathologic parameters [7]. Although, the cancer prognosis has improved significantly over the years through the development of some targeted drugs, none of these therapies is curative [8]. Thus, the early diagnosis and treatment of ccRCC is vital and imperative. Nowadays, some tools have been explored to analyse the association between clinical outcomes and different gene signatures. We plan to discover novel biomarkers for early screening of ccRCC through the cancer database and these methods.

The tumor microenvironment (TME) is a mixture of chemokines, fluids, extracellular matrix molecules, stromal cells, numerous cytokines and immune cells. The molecules and cells of the TME are in a dynamic process, reflecting the evolutionary characteristics of cancer, tumor growth and metastasis [9]. Tumour-infiltrating immune cells (TIICs) including immune cells that migrate from the brink to tumour tissues and exert a negative or positive effect on the growth of cells [10]. Recently, the infiltration of immune cells into different tumors has made significant progress, but the function of these immune cells in immune defense and tumor initiation or tolerance still needs further exploration.

In this study, we discovered the gene COLGALT1 enriched in metabolic pathways by means of gene set enrichment analysis (GSEA). As is known to us, the occurrence and development of tumors are closely related to metabolic pathways. Hence, we want to explore the prognostic value of COLGALT1 in ccRCC based on TCGA database. Also, we evaluate the connection between the COLGALT1 and the TIICs. Through the further analysis, we demonstrated that the COLGALT1 could play one significant role as a favorable prognostic and therapeutic factor for ccRCC.

**Materials And Methods**

**Patient samples**

We obtain the data including the raw RNA-Sequencing reads and corresponding clinical information of kidney renal clear cell carcinoma patients for a total of 539 ccRCC and 72 normal samples in the TCGA database. In this current study, overall survival (OS) was the primary outcome. We use the Bioconductor packages in an R statistical environment to process and normalize the RNA expression levels of the samples. Finally, we conduct further analysis using these samples gene expression data with the Clear cell renal cell carcinoma and the clinical information.

**Identifying genes of differential expression**

All data from the TCGA were analyzed using the R software (https://www.r-project.org/). Through the R/Bioconductor package of edgeR, the different expression of mRNA in Clear cell renal cell carcinoma was calculated [11]. The cut-off standard DEGs was defined as |log2fold-change (log2FC)| > 2 and
adjusted P-value \((P) < 0.05\). And we determine the COLGALT1 met the criteria. Then we compare gene expression data between adjacent non-tumor tissue and tumor tissue. Also, compare the COLGALT1 expression in different cancer clinical stages.

**Cell culture**

Clear cell renal cell carcinoma cell line Caki-1 and Human Kidney-2 cell (Hk-2) were purchased from the Shanghai Enzyme-linked Biotechnology Co., Ltd. (Shanghai, China). Clear cell renal cell carcinoma cell line were maintained in DMEM (Gibco, USA) which contained 10% fetal bovine serum (FBS, HyClone, USA) and 1% penicillin/streptomycin at 37 °C with 5% CO2. Hk-2 were cultured in DF-12 medium (Gibco, USA) containing 10% fetal bovine serum and 1% penicillin/streptomycin.

**RNA Isolation and quantitative real-time PCR**

Total RNA was isolated by Trizol (Takara, Japan) from cells. The ReverTra Ace qPCR RT Kit (Toyobo, China) was utilized to perform the reverse transcription reaction. 1 µl of the resulting cDNAs were utilized as templates for each reaction of the RT-PCR with iQTM SYBR® Green Supermix (Bio-Rad, USA), with a final volume of 20 µl. COLGALT1 primer: 5′- ACTCCACGGAATGGTACAAAC-3′ (forward), 5′- CTACGGACCACAACATGGATAAC-3′ (reverse). GAPDH primer: 5′- GGAGCGAGATCCCTCAAAT-3′ (forward), 5′- GGCTGTTGTNETACTTCTCATGG-3′ (reverse). The annealing temperature of the two primers was 60 °C.

**Functional and pathway enrichment**

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [12] is an encyclopedia of genes and genomes. We performed it for the selected gene to find potential pathways using the database for annotation, visualization and Integrated discovery (http://www.kegg.jp/ or http://www.genome.jp/kegg/). Nominal P-value \((P)\) less than 0.05 and FDR less than 25% was considered statistically significant.

**Relationships between COLGALT1 and adjacent genes**

We plan to evaluate the relationships between the COLGALT1 and adjacent genes. The Ensemble Genome Browser was carried out to identify these genes. We downloaded its expression level from TCGA and expressed it as RPKM value. We then perform differential expression analysis of the adjacent genes. Pearson's correlation coefficient was used to analyze the potential relationships between the COLGALT1 and adjacent genes. All the above statistical analyses were performed using SPSS software [13].

**Evaluation of independent prognostic index (Riskscore)**

Univariate Cox and multivariate Cox regression analysis were employed in the TCGA dataset on the prognostic gene signature and clinicopathological parameters to identify the impact of COLGALT1 as one independent factor correlated with OS. The package of “survivalROC” in R was carried out to complete ROC and AUC for the evaluation of the prognostic ability of RS. Moreover, by using the R “rms” package,
we conduct one nomogram-based model to visualize the relationship between survival rates and individual predictors.

The evaluation of Microsatellite Instability (MSI), Tumor Mutational Burden (TMB) and Neoantigen

The MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html) was applied to identify all the autosomal microsatellite tracts containing five or more repeating subunits 1–5 bp in length. Detailed calculations were performed as mentioned before [14]. The total number of somatic nonsynonymous mutations (NSM) which was determined by comparing sequence data from tumor tissues and matched samples was used to define the Tumor mutation burden through previously described method [15]. We used seq2HLA, version 2.2, 17 with default settings to generate 4-digit typing for various tumors in TCGA. Then, pvac-seq, version 4.0.8,18 was ran to generate Neoantigen on each sample [16, 17].

Correlation analysis between COLGALT1 with tumor microenvironment

The purity-adjusted Spearman correlations between COLGALT1 and the six different kinds of immune cell infiltrations were analyzed by the TIMER. The ESTIMATE algorithm was utilized for the normalized expression matrix to calculate the estimate, stromal and immune scores for ccRCC [18].

CIBERSORT and assessment of TIICs

CIBERSORT is a deconvolution algorithm that was reported to predict the fractions of multiple cell types based on gene expression [19, 20]. According to the standardized gene expression data, the cellular composition of complex tissues can be estimated which indicated the abundances of specific cell types [21, 22]. For this study, the gene composition of each cell was determined through calculating the expression level of each gene in immune checkpoint molecules, immune cells and mismatch repair protein respectively.

Statistical analysis

All statistical data analysis and figures were carried out on SPSS 24.0 (IBM, Chicago, USA), R 3.3.1 (https://www.r-project.org/) and GraphPad Prism 6.0 (San Diego, CA, USA). Pearson's correlation method was used to analyse the correlation between two different genes. Wilcoxon signed rank test and logistic regression were applied to estimate the association between clinicopathologic parameters and COLGALT1. The survival predictive performance of COLGALT1 and risk score (RS) were carried out by using Kaplan-Meier curve and log-rank test. The relationship between variables and OS were evaluated by Univariate and multivariate Cox regression analyses. The rms package of R software was used to create the nomogram. The package of “survivalROC” in R was carried out to complete ROC and AUC for the evaluation of the prognostic ability of RS. Protein-protein interaction (PPI) network was established by STRING and visualized by Cytoscape software. All statistical results with \( P < 0.05 \) were considered statistically significant.
Result

The expression of COLGALT1 in ccRCC

We found that COLGALT1 is highly expressed in carcinoma tissues, including Clear cell renal cell carcinoma through comparing the expression of normal tissue with carcinoma tissue from the TCGA datesets ($P < 0.01$, Fig. 1A). We obtain the GOLGAT1 expression level of the normal tissues ($n = 72$) and the malignant tissues (ccRCC, $n = 539$) from the TCGA datesets (Fig. 1B). It shows that the expression of COLGALT1 of the ccRCC tissue is higher compared to normal tissue ($P < 0.001$). We compare the COLGALT1 expression in Clear cell renal cell carcinoma tissue with the para-cancerous tissue in 72 pairs of samples ($P = 1.097e-21$, Fig. 1C). The results from the cell lines also indicated that COLGALT1 mRNA was increased in ccRCC through qRT-PCR(Fig. 1D).

By using the GOLGAT1 expression as the cut-off point, we classified the patients with ccRCC into high and low risk group. The survival curve evaluated for ccRCC patients in the low-risk group and those in the high-risk group were obviously different indicating that patients with high-COLGALT1 expression had a shorter overall survival time than patients with low-COLGALT1 expression ($P < 0.001$, Fig. 2A). The AUC of a time-dependent ROC curve was used to evaluate the prognostic capacity of the COLGALT1 signature and the AUC of gene biomarker prognostic model was 0.707, see Fig. 2B. The result indicated that the ability of the survival prediction for ccRCC patients was acceptable. The immunohistochemical pictures from the HPA database (http://www.proteinatlas.org/) indicates that COLGALT1 is highly expressed in tumor tissues compared to normal tissues (Fig. 2C-D).

Association between COLGALT1 expression and different clinicopathologic parameters

We compare the COLGALT1 expression in ccRCC patients with different clinical stages. The patient’ data with ccRCC obtained from TCGA are divided into different groups according to the clinicopathologic stage. The corresponding COLGALT1 expression is shown in Fig. 3A-D. We discover advanced tumors have higher expression of COLGALT1 than early tumors in the ($P < 0.01$). However, there is no significant difference of COLGALT1 expression between African and Caucasian, male and female.

GSEA identifies a COLGALT1-related signaling pathway

Gene Set Enrichment Analysis (GSEA) was carried out to find out potential signaling pathways which are differentially activated in ccRCC patients with low and high COLGALT1 expression. We selected the most significantly enriched signaling pathways based on the normalized enrichment score (NES) and FDR q-val ($FDR < 0.01$) (Fig. 4 and Table 2). The result shows that butanoate metabolism, fatty acid metabolism, histidine metabolism, ppar signaling pathway, propanoate metabolism, pyruvate metabolism and tryptophan metabolism are differentially enriched in high COLGALT1 expression phenotype.

Multivariate analysis of OS and establishment of ccRCC prognostic prediction nomogram
Univariate Cox and multivariate Cox regression analysis were employed to identify the impact of COLGALT1 as one independent factor correlated with OS (Table 1). The univariate Cox analysis revealed that the OS of ccRCC patients was significantly associated with the COLGALT1 expression, agepathological stage, grade, T stage and M stage (Fig. 5A). Multivariate Cox regression analysis showed ccRCC patients with high COLGALT1 expression were correlated with a poor OS (HR = 1.024; \( P < 0.01 \)). Moreover, some other clinicopathologic parameters correlated with poor overall survival consist of low grade, advanced stage and old age (Fig. 5B). Therefore, the COLGALT1 expression might be an independent prognostic factor of OS compared with these variables.

Predictive ability of the COLGALT1 expression and clinicopathologic variables was demonstrated by using the ROC curve analysis (Fig. 5C). The AUC of the COLGALT1 expression (AUC = 0.703) was obviously higher than that of age (AUC = 0.660), gender (AUC = 0.497), race (AUC = 0.528), M stage (AUC = 0.680) and lymph nodes status (AUC = 0.459). However, the AUC for COLGALT1 is lower than tumor grade (AUC = 0.709), pathological stage (AUC = 0.779) and T stage (AUC = 0.779). Thus, using the COLGALT1 expression level alone to predict the survival of ccRCC patients is insufficient. We then integrate these clinicopathologic variables and the COLGALT1 expression constructing one nomogram integrating to provide clinicians with an effective approach for predicting prognostic risk (Fig. 5D). We can evaluate 1-, 3-, and 5-year survival probabilities of ccRCC patients by using the nomogram.

Validation of the predictive function of COLGALT1

We find ten genes (AUH, BMP1, CD276, FDX1, IMPDH1, MICU2, MOCS2, PPP1R18, RBM47 and BRHBDF2) were significant correlated with COLGALT1 expression by the Pearson’s correlation analysis based on TCGA dataset (\( P < 0.05 \)) (Fig. 6). Among them, five genes (AUH, FDX1, MICU2, MOCS2 and RBM47) were negatively correlated with COLGALT1 expression and five genes(BMP1, CD276, IMPDH1, PPP1R18 and RHBFDF2) were possitively correlated with COLGALT1 expression. A PPI network of DEGs was constructed, we can find the relationship between the COLGALT1 with ten mostly relevant other genes (Fig. 7A). Also, based on the TCGA database, we explore whether COLGALT1 is related to different tumors from three aspects including Microsatellite Instability, Tumor Mutational Burden and Neoantigen. And the COLGALT1 is related to Microsatellite Instability, Tumor Mutational Burden in ccRCC (\( P < 0.001 \),Fig. 7B-D).

Correlations between COLGALT1 with immune infiltrations and methylation in ccRCC

The correlations between the COLGALT1 and the six different immune cell infiltration levels were shown through online analysis with TIMER. And the COLGALT1 was indicated to be positively correlated with the immune infiltrations including B cell infiltration, CD8 + T cell infiltration, CD4 + T cell infiltration, macrophage infiltration, neutrophil infiltration, and dendritic cell infiltration, respectively (\( P < 0.01 \), Fig. 8A). As shown in Fig. 8B, the transcripts per million (TPM) is positively related to the immune score indicating the positive effects of the Somatic copy number (of the COLGALT1) on the immune response in ccRCC. Through Fig. 8C, we found that the methyltransferase including DNMT1 (DNA methyltransferase 1), DNMT2 (DNA methyltransferase 2), DNMT3a (DNA methyltransferase 3a) and DNMT3b (DNA
methyltransferase 3b) was positively correlated with the COLGALT1 in different tumors ($P<0.01$). It suggests that COLGALT1 may affect tumor formation through methylation.

**The connection between the COLGALT1 and the immune microenvironment in ccRCC**

We also explored immune microenvironment in tumor tissue based on the TCGA database. As is shown in the Fig. 9A, the COLGALT1 is significantly connected with the immune checkpoint molecules like BTLA, CD28, CD40, NRP1 ect in ccRCC. Also, we explore the interaction between the immune cells and the COLGALT1 in ccRCC. It showed that the COLGALT1 is linked to relevant immune cells including Actived memory B cell, Actived CD4 T cell, Actived CD8 T cell ect (Fig. 9B). Next, we investigate the relationship between the COLGALT1 and mismatch repair protein including MLH1, MSH2, MSH6, PMS2 in ccRCC showing potential connection among them ($P<0.001$, Fig. 9C).

**Discussion**

Clear cell renal cell carcinoma (ccRCC) is the most common histological type of RCC accounting for 90% of kidney neoplasms [23]. We plan to explore new reliable method for survival prediction and treatment of ccRCC. Nowadays, many studies have reported that the prognosis of ccRCC has connections with genetic factors, and some genes may provide method for predicting prognosis and selecting treatments [24–28]. Today, with the development of microarray technology, some potential prognostic and therapeutic targets have been found. We find CLOGAT1 may has connection with some metabolism pathways and immune infiltrations which have impacts on tumor. Hence, our present study focused on the prognostic role of COLGALT1 in ccRCC.

In this study, we identified the COLGALT1 expression between the ccRCC tissues and the adjacent normal tissues from the TCGA database. Based on qRT-PCR, the COLGALT1 mRNA expression was high in ccRCC cell line Caki-1 compared with human kidney cell. These results revealed that COLGALT1 shows obviously higher expression in ccRCC tissues than that in adjacent normal tissues. Furthermore, we explored the association among COLGALT1 expression and patient survival based on the database showing that high expression of COLGALT1 indicated a bad prognosis of ccRCC. Additionally, high COLGALT1 expression was significantly associated with better pathologic stage, histological grade, T stage, M stage, and satisfactory survival time. The COLGALT1 expression level was proved to be an independent predictive factor of overall survival for ccRCC patients. In order to explore how COLGALT1 was involved in ccRCC pathogenesis, we carried out GSEA between tissues with different COLGALT1 expression levels and found that some pathways including butanoate metabolism, fatty acid metabolism, histidine metabolism, ppar signaling pathway, propanoate metabolism, pyruvate metabolism and tryptophan metabolism are differentially enriched in COLGALT1 high expression phenotype. We further investigated the relationship of COLGALT1 and related genes based on TCGA dataset and discovered that the five most posssitively relevant genes are BMP1, CD276, IMPDH1, PPP1R18 and RHBDL2. In contrast, the five most negatively relevant genes are AUH, FDX1, MICU2, MOCS2 and RBM47. Also, we find the relationship between the
COLGALT1 with ten mostly relevant other genes through the PPI network. And the COLGALT1 is related to Microsatellite Instability, Tumor Mutational Burden in ccRCC based on the TCGA database. And, we explore the potential connection between the COLGALT1 with the tumor microenvironment and immune infiltrations in various tumors. Finally, we find the COLGALT1 is significantly related to these aspects including immune cell infiltration, DNA methyltransferase, immune checkpoint molecules immune cells and mismatch repair protein among ccRCC.

We search some relevant literature for the progress of the COLGALT1. The galactosyltransferases COLGALT1 and COLGALT2 in the endoplasmic reticulum initiate the glycogen glycosylation reaction. Mutations in the COLGALT1 gene can cause abnormalities in cerebral small blood vessels and malformations of the nostrils, which are common in type IV collagen deficiency. These advances are focused on muscle and small vessel diseases [29–31]. As we all know, alterations in mitochondrial metabolism have been described as one of the major factor of both ageing cells and cancer [32]. This may be one of the reasons that the COLGALT1 are related to ccRCC. However, We still need further study and analysis to explore the possible relationship.

In our study, we discovered that high COLGALT1 expression had association with butanoate metabolism, fatty acid metabolism, histidine metabolism, ppar signaling pathway, propanoate metabolism, pyruvate metabolism and tryptophan metabolism by GSEA. These pathways are crucial biological processes in tumorigenesis and development of tumor. We can find that these pathways are mainly focused on metabolic pathways. Over these years, many studies have begun to explore the molecular and cellular mechanisms that constitute the interaction between nutrition, amino acid, fat metabolism and effects on cancer [33, 34]. Given the research background of metabolism, there are many articles focused on finding potential ways to treat or even cure cancer [35–37]. COLGALT1 as one gene enriched in these metabolic pathways, which may play a certain role in tumor metabolism and the body’s immune process. However, further experiments are needed to verify it.

Ten most significantly relevant genes were discovered in this study. Five genes (BMP1, CD276, IMPDH1, PPP1R18 and RHBDF2) were positively correlated with GOLGALT1 expression. Recently, outstanding research has provided new clues to the function of CD276 (B7H3) in cancer, and determined that CD276 is the key to tumor cell proliferation, migration, invasion, epithelial-mesenchymal transition, cancer stemness, drug resistance and Warburg effect Promoter.[38]. Also, one article reports that RHBDF2 may be linked to the esophageal cancer [39]. On contrary, Five genes (AUH, FDX1, MICU2, MOCS2 and RBM47) were negatively correlated with GOLGALT1 expression. Among them, three genes (AUH, FDX1, MICU2) are related to mitochondrial energy metabolism [40–42]. As we mentioned earlier, mitochondria are closely related to tumor progression. This again proves that COLGALT1 may act on tumors through metabolic pathways. Also, we find the relationship between the COLGALT1 with ten mostly relevant other genes through the PPI network. Further verification need to be carried out to confirm the relationship between them.
Immune response were shown to be linked with various tumors importantly. In COAD, a mouse model showed depletion of CD25+, CD4 + regulatory T cells able to enhance the anti-tumor immunity induced by interleukin2 [43]. CD4 + T cells were found to play significant role in the progression and metastasis of lung cancer [44]. Tumor-infiltrating naive CD4 + T cells were proved to be connected with poor survival in bladder cancer [45]. Here, We explored the connection between the GOLGALT1 and six different immune cells including B cell infiltration, CD8 + T cell infiltration, CD4 + T cell infiltration, macrophage infiltration, neutrophil infiltration, and dendritic cell infiltration, respectively in ccRCC. The COLGALT1 was indicated to be positively correlated with these immune cells (P < 0.01). The methyltransferase including DNMT1, DNMT2, DNMT3a and DNMT3b was positively correlated with the COLGALT1 in different tumors (P < 0.01).

Immune checkpoints provide a general mechanism for various cancers to avoid immune surveillance and play a significant role in the immune system. In lung cancer, anti-CTLA-4 and anti-PD-1/PD-L1 blocking antibodies have shown to be successful for treatment. In addition, in lung cancer, there are some recognition markers for early response, such as TCR library, CD4 + / CD8 + T cell profile, cytokine markers and the expression of immune checkpoint molecules in tumor cells, macrophages or T cells [46, 47]. Here, we find COLGALT1 is related to MSI, TMB in ccRCC. Also, we explored immune microenvironment in tumor tissue through three aspects including immune checkpoint molecules, immune cells and mismatch repair protein respectively by using the CIBERSORT algorithm. The COLGALT1 is significantly connected with the immune checkpoint molecules like BTLA, CD28, CD40, NRP1 ect in ccRCC. Also, the COLGALT1 is linked to relevant immune cells including Actived memory B cell, Actived CD4 T cell, Actived CD8 T cell ect. The relationship between the mismatch repair protein including MLH1, MSH2, MSH6, PMS2 with the COLGALT1 was investigated showing potential connection among them in ccRCC. These results may offer help in the development of ccRCC treatments.

Our study has some limitations. First, our results come from the TCGA database and generated by bioinformatic analysis. Considering various factors (region, age, gender, race, etc.) and the heterogeneity in the analysis process, the sample size of renal clear cell carcinoma cannot guarantee sufficient. Therefore, the results of our study need to be verified with enough clinical samples. Also, we need further investigations to validate our results based on ccRCC samples and clinical data. Second, the relationship between COLGALT1 expression with the ccRCC and these signaling pathways is the first to be reported, and the regulatory mechanism needs to be further investigated. Therefore, further experiments need to be carried out to explore whether progression in ccRCC should be affected by COLGALT1 through metabolic pathways or other possible pathways.

Conclusions

In conclusion, the COLGALT1 could be a reliable prognostic factor for ccRCC according our study. Moreover, butanoate metabolism, fatty acid metabolism, histidine metabolism, ppar signaling pathway, propanoate metabolism, pyruvate metabolism and tryptophan metabolism might be the primary pathways regulated by COLGALT1. Besides, MSI and TMB showed strong correlations with COLGALT1 in...
Moreover, the COLGALT1 was indicated to be positively correlated with the immune infiltrations and the methyltransferase. Furthermore, we verify the connection between the COLGALT1 and the immune microenvironment including immune checkpoint molecules, immune cells and mismatch repair protein in ccRCC. We still need more evidence and experiments to find the potential molecular mechanisms of COLGALT in ccRCC.

Declarations

Funding

None declared.

Author’s Contribution

SW.L and Y.W: Protocol/project development;
QW.X: Data collection or management;
Y.Y: Data analysis;
SW.L and Y.Y: performed the experiments.
BM.H and W.Y: Manuscript writing/editing.

Availability of data and material

All the data used to support the findings of this study are included within the article. Please contact author for data requests.

Competing interests

None declared.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Acknowledgements

We would like to thank the researchers and study participants for their contributions.

References
1. Siegel, R.L., K.D. Miller and A. Jemal. Cancer statistics, 2018. CA Cancer J Clin. (2018) 68(1):7-30. doi: 10.3322/caac.21442. Epub 2018 Jan 4.

2. Hsieh, J.J., et al. Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. J Pathol. (2018) 244(5):525-537. doi: 10.1002/path.5022. Epub 2018 Feb 14.

3. Huang, Y., et al. Clonal architectures predict clinical outcome in clear cell renal cell carcinoma. Nat Commun. (2019) 10(1):1245. doi: 10.1038/s41467-019-09241-7.

4. Amin, M.B., et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin. (2017) 67(2):93-99. doi: 10.3322/caac.21388. Epub 2017 Jan 17.

5. Janowitz, T., et al. Adjuvant therapy in renal cell carcinoma-past, present, and future. Semin Oncol. (2013) 40(4):482-91. doi: 10.1053/j.seminoncol.2013.05.004.

6. Janzen, N.K., et al. Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. Urol Clin North Am. (2003) 30(4):843-52. doi: 10.1016/s0094-0143(03)00056-9.

7. Lam, J.S., et al. Postoperative surveillance protocol for patients with localized and locally advanced renal cell carcinoma based on a validated prognostic nomogram and risk group stratification system. J Urol. (2005) 174(2):466-72; discussion 472; quiz 801. doi: 10.1097/01.ju.0000165572.38887.da.

8. Figlin, R.A., et al. Adjuvant therapy in renal cell carcinoma: does higher risk for recurrence improve the chance for success? Ann Oncol. (2018) 29(2):324-331. doi: 10.1093/annonc/mdx743.

9. Jiang, X., et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. Mol Cancer. (2019) 18(1):10. doi: 10.1186/s12943-018-0928-4.

10. Cackowski, F.C., et al. Detection and isolation of disseminated tumor cells in bone marrow of patients with clinically localized prostate cancer. Prostate (2019) 79(14):1715-1727. doi: 10.1002/pros.23896. Epub 2019 Aug 26

11. Robinson, M.D., D.J. McCarthy and G.K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (2010) 26(1):139-40. doi: 10.1093/bioinformatics/btp616. Epub 2009 Nov 11.

12. Kanehisa, M., et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res.(2017) 45(D1):D353-D361. doi: 10.1093/nar/gkw1092. Epub 2016 Nov 28.

13. Petrizzo, A., et al. Identification and Validation of HCC-specific Gene Transcriptional Signature for Tumor Antigen Discovery. Sci Rep. (2016) 6:29258. doi: 10.1038/srep29258.

14. Timmermann, B., et al. Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. PLoS One. (2010) 5(12):e15661. doi: 10.1371/journal.pone.0015661.

15. Chalmers, Z.R., et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. (2017) 9(1):34. doi: 10.1186/s13073-017-0424-2.
16. Hundal, J., et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. (2016) 8(1):11. doi: 10.1186/s13073-016-0264-5.

17. Boegel, S., et al. HLA typing from RNA-Seq sequence reads. Genome Med. (2012) 4(12):102. doi: 10.1186/gm403. eCollection 2012.

18. Yoshihara, K., et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. (2013) 4:2612. doi: 10.1038/ncomms3612.

19. Newman, A.M., et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. (2015) 12(5):453-7. doi: 10.1038/nmeth.3337. Epub 2015 Mar 30.

20. Shen-Orr, S.S. and R. Gaujoux. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Curr Opin Immunol. (2013) 25(5):571-8. doi: 10.1016/j.coi.2013.09.015.

21. Ge, P., et al. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. Biomed Pharmacother (2019) 118:109228. doi: 10.1016/j.biopha.2019.109228. Epub 2019 Jul 24.

22. Yang, X., et al. Identification and validation of an immune cell infiltrating score predicting survival in patients with lung adenocarcinoma. J Transl Med. (2019) 17(1):217. doi: 10.1186/s12967-019-1964-6.

23. Cohen, H.T. and F.J. McGovern. Renal-cell carcinoma. N Engl J Med. (2005) 353(23):2477-90. doi: 10.1056/NEJMra043172.

24. Lee, D., et al. GABRQ expression is a potential prognostic marker for patients with clear cell renal cell carcinoma. Oncol Lett. (2019) 18(6):5731-5738. doi: 10.3892/ol.2019.10960. Epub 2019 Oct 4.

25. Bao, Y., et al. High eukaryotic translation elongation factor 1 alpha 1 expression promotes proliferation and predicts poor prognosis in clear cell renal cell carcinoma. Neoplasma. (2019) 67(1):78-84. doi: 10.4149/neo_2019_190224N158. Epub 2019 Nov 26.

26. Fu, Y., et al. Increased expression of RUNX1 in clear cell renal cell carcinoma predicts poor prognosis. PeerJ (2019) 7:e7854. doi: 10.7717/peerj.7854. eCollection 2019.

27. Koh, H.M., et al. Prognostic role of S100A9 expression in patients with clear cell renal cell carcinoma. Medicine (Baltimore). (2019) 98(40):e17188. doi: 10.1097/MD.00000000000017188.

28. Zhao, Z., et al. The mRNA Expression Signature and Prognostic Analysis of Multiple Fatty Acid Metabolic Enzymes in Clear Cell Renal Cell Carcinoma. J Cancer. (2019) 10(26):6599-6607. doi: 10.7150/jca.33024. eCollection 2019.

29. Miyatake, S., et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. Ann Neurol. (2018) 84(6):843-853. doi: 10.1002/ana.25367. Epub 2018 Nov 30.

30. Hennet, T. Collagen glycosylation. Curr Opin Struct Biol. (2019) 56:131-138. doi: 10.1016/j.sbi.2019.01.015. Epub 2019 Feb 26.

31. Geister, K.A., et al. Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. Dis Model Mech. (2019) 12(6):dmm037176. doi:
10.1242/dmm.037176.

32. Porporato, P.E., et al. Mitochondrial metabolism and cancer. Cell Res. (2018) 28(3):265-280. doi: 10.1038/cr.2017.155. Epub 2017 Dec 8.

33. Hoxhaj, G. and B.D. Manning. The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. Nat Rev Cancer. (2019) 20(2):74-88. doi: 10.1038/s41568-019-0216-7.

34. Patel, C.H., et al. Targeting metabolism to regulate immune responses in autoimmunity and cancer. Nat Rev Drug Discov. (2019) 18(9):669-688. doi: 10.1038/s41573-019-0032-5. Epub 2019 Jul 30.

35. Phan, T.K., et al. Combating Human Pathogens and Cancer by Targeting Phosphoinositides and Their Metabolism. Trends Pharmacol Sci. (2019) 40(11):866-882. doi: 10.1016/j.tips.2019.09.006. Epub 2019 Oct 31.

36. Casero, R.J., S.T. Murray and A.E. Pegg. Polyamine metabolism and cancer: treatments, challenges and opportunities. Nat Rev Cancer. (2018) 18(11):681-695. doi: 10.1038/s41568-018-0050-3.

37. Sanderson, S.M., et al. Methionine metabolism in health and cancer: a nexus of diet and precision medicine. Nat Rev Cancer. (2019) 19(11):625-637. doi: 10.1038/s41568-019-0187-8. Epub 2019 Sep 12.

38. Castellanos, J.R., et al. B7-H3 role in the immune landscape of cancer. Am J Clin Exp Immunol. (2017) 6(4):66-75. eCollection 2017.

39. Ellis, A., et al. Tylosis with oesophageal cancer: Diagnosis, management and molecular mechanisms. Orphanet J Rare Dis. (2015) 10:126. doi: 10.1186/s13023-015-0346-2.

40. Sawyer, A. and M. Winkler. Evolution of Chlamydomonas reinhardtii ferredoxins and their interactions with [FeFe]-hydrogenases. Photosynth Res. (2017) 134(3):307-316. doi: 10.1007/s11120-017-0409-4. Epub 2017 Jun 15.

41. Jones, D.E., L. Perez and R.O. Ryan. 3-Methylglutaric acid in energy metabolism. Clin Chim Acta. (2019) 502:233-239. doi: 10.1016/j.cca.2019.11.006. Epub 2019 Nov 12.

42. Cui, C., et al. Progress in understanding mitochondrial calcium uniporter complex-mediated calcium signalling: A potential target for cancer treatment. Br J Pharmacol. (2019) 176(9):1190-1205. doi: 10.1111/bph.14632. Epub 2019 Apr 3.

43. Imai, H., et al. Depletion of CD4+CD25+ regulatory T cells enhances interleukin-2-induced antitumor immunity in a mouse model of colon adenocarcinoma. Cancer Sci. (2007) 98(3):416-23. doi: 10.1111/j.1349-7006.2006.00385.x.

44. Marshall, E.A., et al. Emerging roles of T helper 17 and regulatory T cells in lung cancer progression and metastasis. Mol Cancer. (2016) 15(1):67. doi: 10.1186/s12943-016-0551-1.

45. Zhang, Q., et al. High CD4(+) T cell density is associated with poor prognosis in patients with non-muscle-invasive bladder cancer. Int J Clin Exp Pathol. (2015) 8(9):11510-6. eCollection 2015.

46. Zhong, Q.Y. et al., A gene expression-based study on immune cell subtypes and glioma prognosis. BMC Cancer. (2019) 19(1):1116. doi: 10.1186/s12885-019-6324-7.
Tables

Table 1
Prediction of overall survival by using univariate and multivariate analyses of COLGALT mRNA level and clinicopathological variables;

| Variable | Univariate analysis | | Multivariate analysis |
| --- | --- | --- | --- |
| | HR | 95% CI | p | HR | 95% CI | p |
| Overall survival | | | | | | |
| Age(year) | 1.033 | 1.020–1.047 | **0.000** | 1.041 | 1.026–1.057 | **0.000** |
| Gender | 0.933 | 0.680–1.282 | 0.670 | 0.970 | 0.698–1.348 | 0.855 |
| Race | 1.193 | 0.716–1.988 | 0.498 | 1.325 | 0.782–2.246 | 0.295 |
| Grade | 1.967 | 1.639–2.361 | **0.000** | 1.328 | 1.054–1.673 | 0.016 |
| Stage | 1.856 | 1.644–2.095 | **0.000** | 1.745 | 1.237–2.462 | **0.002** |
| T | 1.998 | 1.689–2.362 | **0.000** | 1.039 | 0.787–1.372 | 0.786 |
| M | 2.100 | 1.661–2.655 | **0.000** | 0.878 | 0.471–1.636 | 0.681 |
| N | 0.863 | 0.739–1.008 | 0.063 | 0.839 | 0.714–0.986 | 0.033 |
| COLGALT | 1.026 | 1.018–1.034 | **0.000** | 1.024 | 1.014–1.035 | **0.000** |

HR: hazard ratio, estimated from Cox proportional hazard regression model; CI: confidence interval of the estimated HR; Multivariate models were adjusted for age, grade, race, stage and T,N,M classification.
Table 2
Gene sets enrichment of high COLGALT mRNA expression level in the ccRCC cohort;

| Gene set name                        | NES   | NOM p-val | FDR q-val |
|--------------------------------------|-------|-----------|-----------|
| KEGG_BUTANOATE_METABOLISM            | 2.256 | 0.000     | 0.001     |
| KEGG_FATTY_ACID_METABOLISM           | 2.258 | 0.000     | 0.001     |
| KEGG_HISTIDINE_METABOLISM            | 2.222 | 0.000     | 0.002     |
| KEGG_PPAR_SIGNALING_PATHWAY          | 1.778 | 0.010     | 0.057     |
| KEGG_PROPANOATE_METABOLISM           | 2.304 | 0.000     | 0.004     |
| KEGG_PYRUVATE_METABOLISM             | 2.262 | 0.000     | 0.002     |
| KEGG_WNT_SIGNALING_PATHWAY           | 2.312 | 0.000     | 0.001     |

NES: normalized enrichment score; NOM: nominal; FDR: false discovery rate. Gene sets with NOM p-val < 0.01 and FDR q-val < 0.01 are considered as significant.