

Identification of Key Claudin Genes Associated With Survival Prognosis and Diagnosis in the Colon Cancer Through Integrated Bioinformatic Analysis.

Rana Alghamdi (✉ raalghamdi3@kau.edu.sa)

King Abdulaziz University

Maryam Al-Zahrani

King Abdulaziz University

Research Article

Keywords: Claudin, Genetics, bioinformatics methods, survival prognosis, immune infiltrations, cancer-associated pathways

Posted Date: November 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1007190/v1>

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

[Read Full License](#)

Identification of Key Claudin Genes Associated with Survival Prognosis and Diagnosis in the Colon Cancer Through Integrated Bioinformatic Analysis.

Rana A. Alghamdi ^{1,2,*}, Maryam H. Al-Zahrani ³,

1 Department of Chemistry, Science and Arts College, King Abdulaziz University, Rabigh, Saudi Arabia; raalghamdi3@kau.edu.sa

2 Regenerative Medicine and Stem Cells Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia.

3 Biochemistry Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia; mhsalzahrani@kau.edu.sa

* Correspondence: raalghamdi3@kau.edu.sa

Abstract

Background: Claudin's gene are associated with various aberrant physiological and cellular signaling. However, the association of claudins with survival prognosis, signaling pathways, and diagnostic efficacy in colon cancer remain lacking. **Methods:** We used various bioinformatics methods, including differential expression analysis, gene set enrichment analysis (GSEA), protein-protein interaction (PPI), survival analysis, single sample gene set enrichment analysis (ssGSEA), mutation analysis, and identifying receiver operating characteristic (ROC) curve of claudins in the TCGA colon adenocarcinoma (COAD). **Results:** We found that: *CLDN2*, *CLDN1*, *CLDN14*, *CLDN16*, *CLDN18*, *CLDN9*, *CLDN12*, and *CLDN6* are elevated in COAD. In contrast, the *CLDN8*, *CLDN23*, *CLDN5*, *CLDN11*, *CLDN7*, and *CLDN15* are downregulated in COAD. Various claudin's genes are mutated and associated with diagnostic efficacy in the COAD. **Conclusions:** Claudin's genes are associated with prognosis, immune regulation, signaling pathway regulations, and diagnosis. These findings may provide new molecular insight into the treatment of colon cancer.

Keywords: Claudin, Genetics, bioinformatics methods, survival prognosis, immune infiltrations, cancer-associated pathways

1. Introduction

Claudins are a multi-gene transmembrane protein family that included at least 27 members ^{1,2}. Claudins proteins are associated with the tight junctions of cell-cell communication between the plasma membranes of two contacting cells for signaling purposes ³. Claudins are related to the various physiological functions, including paracellular ion pores, extracellular loops, ion permeability, cell polarity, affecting regulatory pathways, stabilizing the integrity of the epithelium, etc. ^{3,4}. Claudins are crucially related to human diseases, including ovarian, breast, pancreatic, and prostate cancers ⁵⁻⁷. The deregulated expression level of claudins can be modulators of carcinogenesis ^{5,6}. In the progression to the metastatic cascade of cancer, claudins are functionally associated with the controlling of various steps ⁵. Claudins regulating the intrinsic roles of the tumor through claudin-mediated functions of stromal cells that substantially influencing the metastatic process of the primary tumor ⁵. Recently, it was stated that the claudins are playing major roles in epithelial-to-mesenchymal transition (EMT), the initiation of cancer stem cells, chemoresistance, and recurrence of tumors ⁶. The various epithelial-derived tumor showed an altered level of claudin expressions and some claudins are associated with the prediction of patient prognosis ⁵. Jian Li demonstrated that the aberrant expression of claudins is linked with neoplastic transformation, which indicating that the claudins are related to the diagnosis and prognosis or targets for the treatment of cancer ⁸. Various studies demonstrated the association of claudins with colon cancer initiation, progression, metastasis, and prognosis ^{2, 9-11}. For example, Didi Zuo et al revealed that claudin-1 is a substantial prognostic biomarker in colorectal cancer ⁹. Liguozhu et al demonstrated that the claudin family participates in colitis-associated colorectal cancer ². Qun Huo et al revealed that claudin-1 protein is a crucial factor associated with the tumorigenesis of colorectal cancer ¹². Claudin-2 potentially trans-activating the

EGFR for promoting colon cancer ¹⁰. The decreased expression level of claudin-4 substantially regulating the invasion and metastasis of colorectal cancer ¹¹. These studies providing the information that the claudins are crucially associated with the invasion, metastasis, and prognosis of colon cancer.

This study presents a thorough bioinformatics investigation on how the changes of claudins expression level regulating the key cancerous phenotypes and immunity of the colon adenocarcinoma. Also, we explore the association of claudins in the prognostic and diagnostic efficacy of colon adenocarcinoma. Moreover, we revealed that the claudins are mutated and substantially associated with regulating the activity of cancerous pathways in COAD.

2. Methods

This study was carried out in the computer laboratories at king Abdulaziz University during 2020-2021. The study was approved by research ethics committee (HA-02-J003) at the center of excellence in genomic medicine research (CEGMR). All of the data in this investigation was analyzed in accordance with CEGMR ethical requirements.

2.1 Datasets

We downloaded the TCGA COAD cohort (<https://portal.gdc.cancer.gov/>) and normalized the data into log₂-base transformation. For analyzing the survival differences of two groups of patients in the TCGA COAD dataset (<https://portal.gdc.cancer.gov/>), we used Expression Profiling Interactive Analysis (GEPIA) ¹³ tool (<http://gepia2.cancer-pku.cn/#index>).

2.2 Identification of Differentially Expressed claudins

The R package "limma" was employed for identifying the significant DEGs between COAD (n=287) and normal samples (n=41) ¹⁴. We identified the DEGs with a threshold absolute value of $\text{Log}_2\text{FC} > 0.50$ and $\text{P-value} \leq 0.05$.

2.3 Gene-Set Enrichment Analysis

We performed gene-set enrichment analysis of the DEGs by using the GSEA ¹⁵. We inputted all claudins into the GSEA tool for identifying GO and pathways. The KEGG ¹⁶ pathways and Reactome pathways ¹⁷ that are significantly associated with the claudins were identified. The $\text{P-value} < 0.05$ was considered significant when selecting the GO terms and pathways.

2.4 Construction of Protein-Protein Interaction (PPI) Network of claudins

To better know the relationship among all claudins, the PPI network was established using the STRING tool ¹⁸. To identify the rank of hub genes, we used Cytoscape plug-in tool cytoHubba ¹⁹. The rank of the genes was identified based on the degree of interactions with neighbor genes. We selected the minimum required interaction score is 0.40 for identifying the PPI of claudins. We visualize the PPI networks by utilizing the Cytoscape 3.6.1 software ²⁰.

2.5 Survival Analysis of claudins by using the GEPIA tool

We compared the overall survival (OS) and the disease-free survival (DFS) of colon cancer patients. Kaplan-Meier survival curves were used to show the survival differences between the high expression group and low expression groups (High expression group > median > Low expression group). The survival significance of all differentially expressed claudins in the TCGA COAD cohort was analyzed using GEPIA ¹³ databases. Cox regression $\text{P-value} < 0.05$ was considered as significant when comparing the survival between the two groups.

2.6 ESTIMATE algorithmic for quantifying immune score and stromal score

ESTIMATE is an algorithmic tool based on the R package for predicting tumor purity, Immune Score (predicting the infiltrations of immune cells), and Stromal Score (predicting the infiltrations of stromal cells) which uses the gene expression profiles of 141 immune genes and 141 stromal genes ²¹. The presence of infiltrated immune cells and stromal cells in tumor tissues were calculated using related gene expression matrix data, represented by Immune Score and Stromal Score, respectively ²¹. Then we calculated the correlations of key genes with immune scores and stromal scores. The threshold value of correlation is $R > 0.20$, and P-value is not less than 0.001 (Spearman's correlation test).

2.7 Single sample gene set enrichment analysis (ssGSEA)

One of the extension packages of GSEA, single-sample gene-set enrichment analysis (ssGSEA) was used to identify the enrichment scores of immune cells for each pairing of a sample and gene set in the tumor samples ²². We collected the marker gene set for immune signatures, biological processes, cancer-associated pathways, and utilizing each gene set to quantify the ssGSEA scores of specific immune signatures ²³⁻²⁶. We identified the ssGSEA score of various immune stimulatory and inhibitory signatures, including B cells, CAFs, CD4 Regulatory T cells, CD8 T cells, cytolytic activity, endothelial cell, immune checkpoint genes, M2 macrophages, macrophages, MDSCs, neutrophils, NK cells, pDC, T cell activation, T cell exhaustion, TAM, Tfh, Th17, TILs, and Tregs. Then, we identified the ssGSEA score of angiogenesis, apical junction, apoptosis, epithelial-mesenchymal transition (EMT), hypoxia, proliferation, and stemness. Moreover, we identified the ssGSEA score of cancer-associated pathways ²⁷ that included cell adhesion molecules CAMs, ECM receptor interaction, ERBB signaling pathway, focal adhesion, gap junction, leukocyte transendothelial migration, MAPK signaling pathway,

MTOR signaling, Notch signaling, pathways in cancer, TGF beta signaling, tight junction, VEGF signaling pathway, and Wnt signaling. All of the gene sets are displaying in **Supplementary Table S1**.

2.8 Diagnostic efficacy evaluation of differentially expressed claudins in the COAD

To assess diagnostic values of the prognostic genes, the receiver operating characteristic (ROC) curve was plotted and the area under the ROC curve (AUC) was calculated using the “pROC” R package ²⁸ to evaluate the capability of distinguishing COAD and normal samples. The greater AUC value of individual genes indicated the differences between tumor and normal samples, and the key gene of AUC>0.5 in the CAFs datasets was defined as a diagnostic efficiency of the gene ²⁹.

2.9 Statistical Analysis

We used the R software version 4.0.1 for all statistical analyses. In the Log-rank test, P<0.05 was considered as statistically significant for survival analysis. To investigate the correlation of genes, Spearman’s correlation between the ssGSEA scores and specific genes was performed (P<0.001). We used the Pearson correlation test to identify the correlation between the two genes (P≤0.05). We utilized R package ggplot2 for the graphical presentation of the Heatmap and correlation graph.

3. Results

3.1 Identifying the differentially expressed Claudins gene members in the COAD

We investigated the differential expression analysis of Claudins gene members in the COAD relative to the normal samples (Table 1). We found that the expression level of *CLDN2*, *CLDN1*, *CLDN14*, *CLDN16*, *CLDN18*, *CLDN9*, *CLDN12*, and *CLDN6* are elevated in the COAD. In contrast, the expression level of *CLDN8*, *CLDN23*, *CLDN5*, *CLDN11*, *CLDN7*, and *CLDN15* is downregulated in the COAD. The Heatmap of the expression value of differentially expressed genes is displaying in **Figure 1**. Other claudins are not differentially expressed in the TCGA COAD samples.

Table 1. The claudins gene members are differentially expressed in the TCGA COAD

Entrez ID	Regulatory status	log2FC	Average Expression	P Value	adjusted P Value	Symbols	Name of the genes
9075	Up regulated	5.96	9.85	1.07E-34	4.55E-34	CLDN2	Claudin 2
9076		4.76	10.21	1.00E-76	1.70E-75	CLDN1	Claudin 1
23562		2.23	3.71	1.03E-16	2.19E-16	CLDN14	Claudin 14
10686		1.69	2.03	5.27E-16	9.95E-16	CLDN16	Claudin 16
51208		1.61	2.85	4.16E-03	5.44E-03	CLDN18	Claudin 18
9080		1.49	4.26	1.20E-07	1.86E-07	CLDN9	Claudin 9
9069		0.80	10.84	2.86E-14	4.87E-14	CLDN12	Claudin 12
9074		0.57	0.82	1.92E-03	2.71E-03	CLDN6	Claudin 6
24146	Down regulated	-0.52	8.67	4.11E-02	4.99E-02	CLDN15	Claudin 15
1366		-1.27	12.72	2.19E-28	6.20E-28	CLDN7	Claudin 7
5010		-2.60	4.39	2.60E-20	6.31E-20	CLDN11	Claudin 11
7122		-2.71	6.88	1.06E-30	3.59E-30	CLDN5	Claudin 5
137075		-3.32	8.21	4.09E-58	2.32E-57	CLDN23	Claudin 23
9073		-7.48	2.44	2.51E-58	2.13E-57	CLDN8	Claudin 8

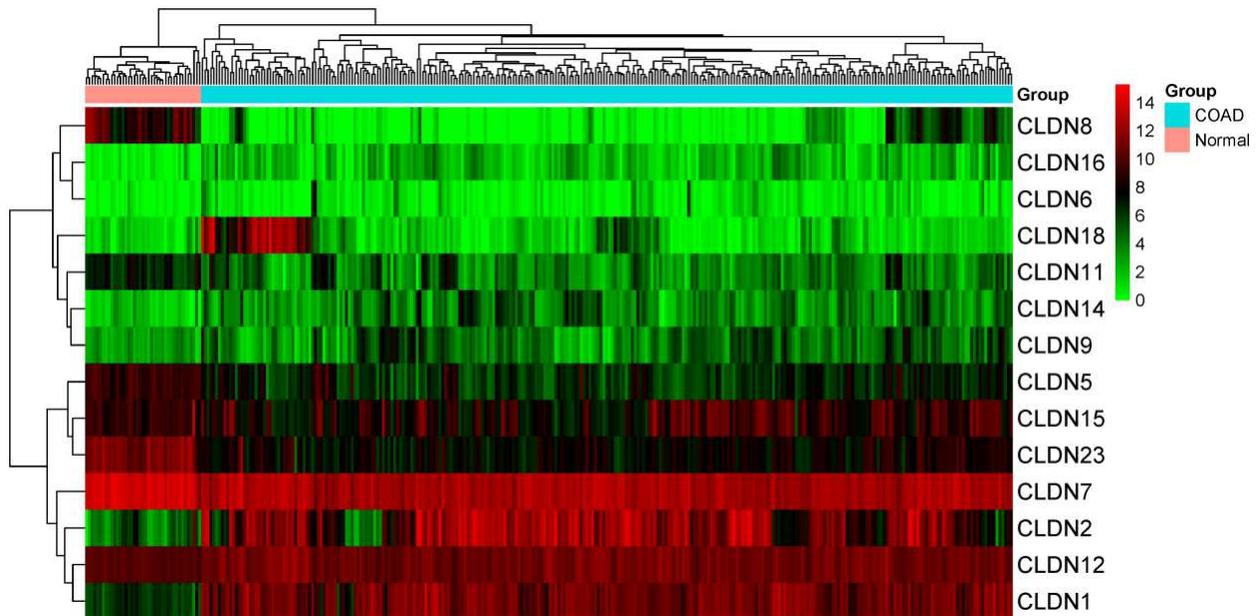


Figure 1. Heatmap of the differentially expressed 14 claudins in the TCGA COAD samples. COAD is colon adenocarcinoma.

3.2 Claudins gene family is associated with functional enrichment and pathways

The enriched gene ontology (GO) terms and pathways were identified by using the GSEA tool (**Figure 2 and Figure 3**). We identified 79 biological processes that are significantly associated with the claudins gene family (**Supplementary Table S2**). the top 20 biological processes are illustrated in Figure 2A, including Calcium independent cell adhesion via plasma membrane cell adhesion molecules, Apical junction assembly, Cell cell junction assembly, Cell cell junction organization, Cell junction assembly, cell cell adhesion via plasma membrane adhesion molecules, Cell junction organization, Biological adhesion, cell cell adhesion, Maintenance of blood brain barrier, and Positive regulation of bicellular tight junction assembly. Besides, we found the 13 significant cellular components terms (Tight junction, Apical junction complex, Cell-cell junction, Anchoring junction, Apicolateral plasma membrane, Lateral plasma membrane, Basal part of cell, Plasma membrane region, Basolateral plasma membrane, Chloride channel complex, Transporter complex, Apical plasma membrane, and Apical part of cell) that are associated with claudins (**Figure 2B**). Moreover, we revealed that the 10 significant molecular functions terms, included Structural molecule activity, Identical protein binding, Exogenous protein binding, Anion channel activity, Chloride transmembrane transporter activity, Inorganic anion transmembrane transporter activity, Ion transmembrane transporter activity, Anion transmembrane transporter activity, Transporter activity, and Passive transmembrane transporter activity are associated with claudins (**Figure 2C**). Furthermore, we found that the three KEGG pathways (Leukocyte transendothelial migration, Tight junction, Cell adhesion molecules (CAMs)) and four Reactome pathways (Tight junction interactions, Cell-cell junction organization, Cell junction organization, and Cell-Cell communication) are significantly enriched that are significantly associated with claudins (**Figure 3**).

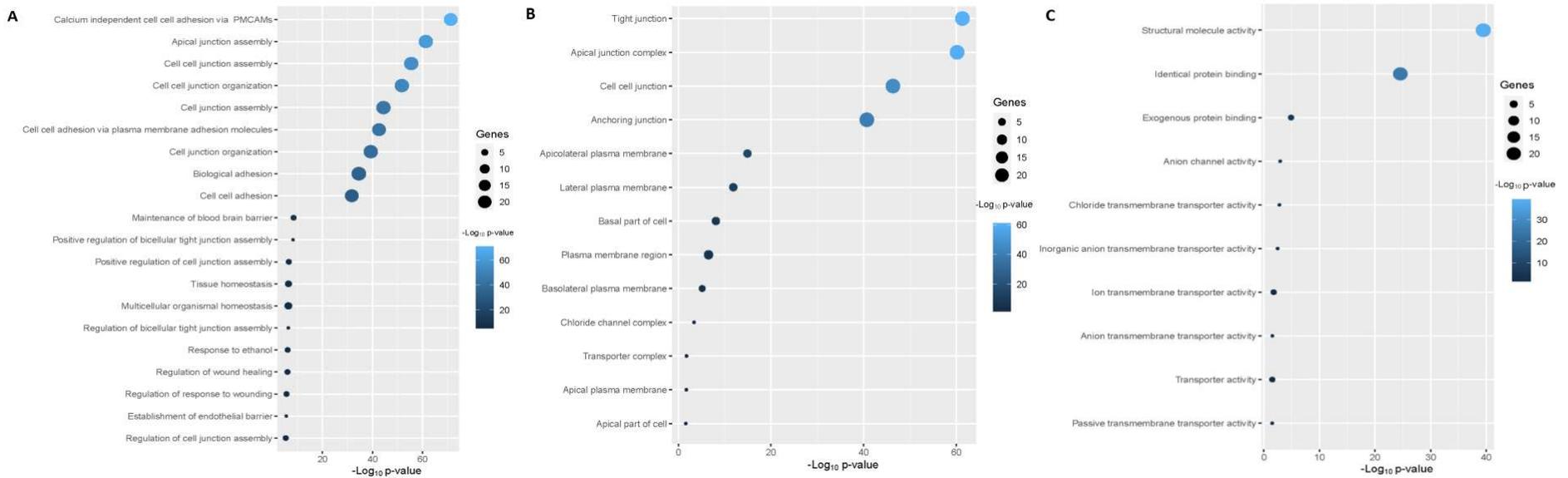


Figure 2. The Claudins family genes are associated with functional enrichment. A. The top 20 significant biological processes that are associated with Claudins B. The 13 cellular components that are associated with Claudins C. The 10 significant molecular functions associated with Claudins.

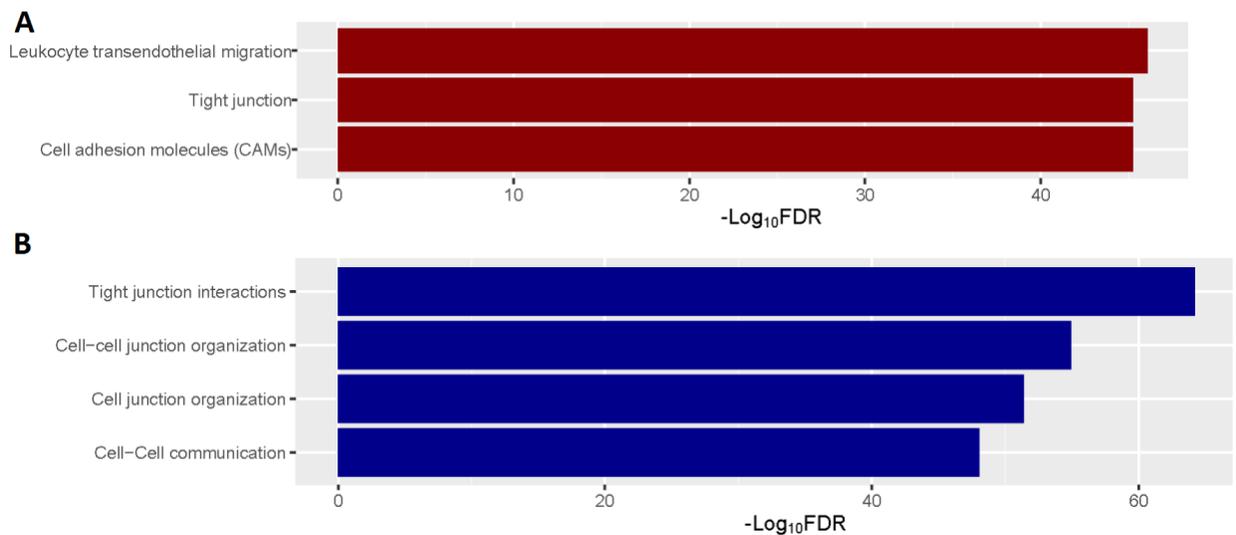


Figure 3. The significantly enriched pathways associated with Claudins. A. The enriched KEGG pathways B. the significantly enriched Reactome pathways.

3.3 Claudins gene family members are involved in the PPI network and correlated with each other

The targeting of protein-protein interactions is substantially relevant to cancer and acting as the tumor-promoting function of several aberrantly expressed proteins in the cancerous conditions is direct influences the capability to interact with a protein-binding regulatory partner ³⁰. For investigating the PPI interaction of all claudins, we used the STRING tool. We revealed that the *CLDN10*, *CLDN2*, *CLDN16*, *CLDN23*, *CLDN12*, *CLDN19*, *CLDN25*, *CLDN5*, *CLDN8*, *CLDN4*, *CLDN15*, *CLDN1*, *CLDN11*, *CLDN9*, *CLDN17*, *CLDN3*, *CLDN22*, *CLDN20*, *CLDN14*, *CLDN6*, *CLDN7*, *CLDN18*, and *CLDND1* are involved in the PPI (**Figure 4A**). The eight claudins (*CLDN10*, *CLDN2*, *CLDN16*, *CLDN23*, *CLDN12*, *CLDN19*, *CLDN25*, and *CLDN5*) are interconnected with other 21 nodes (**Figure 4A**). Besides, the other thirteen claudins (*CLDN8*, *CLDN4*, *CLDN15*, *CLDN1*, *CLDN11*, *CLDN9*, and *CLDN17*) have interacted with other 20 claudins in the PPI (**Figure 4A**). *CLDN18* interconnected with other 7 claudin members and

CLDND1 interacted with only *CLDN25* (**Figure 4A**). After getting the interaction of claudins, we thought that the claudins are correlated with each other (Absolute value of Pearson correlation, $R > 0.2$, $P > 0.05$). Our analysis revealed that the expression level of claudins is correlated with other members (**Figure 4B**). For example, the expression level of *CLDN1* is positively correlated with *CLDN2* and *CLDN12*. Similarly, *CLDN4* is positively correlated with *CLDN1*, *CLDN3*, *CLDN12*, *CLDN7*, and *CLDN15*. In contrast, the expression level of *CLDN18* is negatively correlated with *CLDN15*, *CLDN9*, and *CLDN3*.

3.4 Claudins are associated with poor survival prognosis in the COAD

We investigated the survival significance of all differentially expressed significant claudins (*CLDN2*, *CLDN1*, *CLDN14*, *CLDN16*, *CLDN18*, *CLDN9*, *CLDN12*, *CLDN6*, *CLDN8*, *CLDN23*, *CLDN5*, *CLDN11*, *CLDN7*, and *CLDN15*) in TCGA COAD data. Our analysis revealed that the higher expression group of *CLDN14* and *CLDN11* are significantly correlated with the survival prognosis of colon cancer patients (**Figure 5**) and the low expression group of *CLDN23* is significantly correlated with shorter survival time of colon cancer patients (**Figure 5**). These claudins are associated with the survival prognosis of cancer patients. For example, *CLDN11* and *CLDN14* are correlated with prognostic values in human breast carcinoma ³¹. The risk score models identifying that the *CLDN23* is correlated to the disease prognosis in colon cancer patients ³². the hypermethylated *CLDN11* is related to the metastasis of CRC and also related to the prognosis of poor survival of CRC ³³. It indicates that the expression level of *CLDN14*, *CLDN11*, and *CLDN23* are key regulators in the COAD.

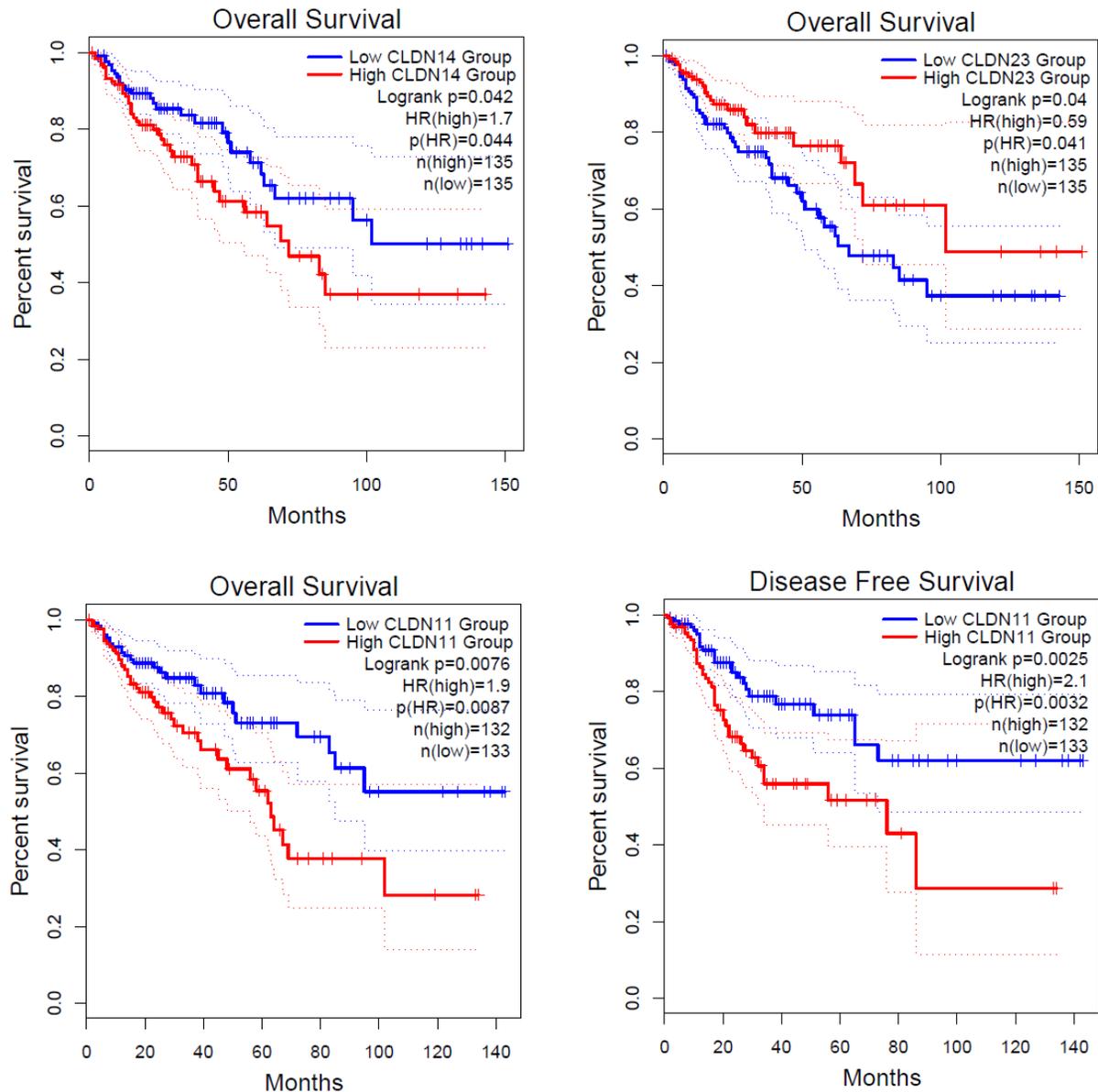


Figure 5. Identification of prognostic DEGs in the COAD. The higher expression group of *CLDN11* and *CLDN14* are significantly correlated with shorter survival time in colon adenocarcinoma. The lower expression group of *CLDN23* is significantly correlated with a shorter survival time in colon adenocarcinoma.

3.5 Claudins are associated with immune infiltrations in the COAD

We investigated the regulation of the tumor microenvironment by Claudins genes. We found that the immune score is positively correlated with *CLDN5*, *CLDN11*, and *CLDN18* and negatively correlated with *CLDN9* (Absolute value of Spearman Correlation, $R > 0.20$ and $P < 0.001$) (**Figure 6A**). Besides, the stromal score is positively correlated with *CLDN5* and *CLDN11* and negatively correlated with *CLDN7* (Absolute value of Spearman Correlation is 0.20 and $P < 0.001$) (**Figure 6A**). Tumor purity, another substantial parameter for the tumor microenvironment, is negatively correlated with *CLDN5*, *CLDN11*, and *CLDN18* (Absolute value of Spearman Correlation is 0.20 and $P < 0.001$) (**Figure 6A**). This result indicates that the expression of *CLDN5*, *CLDN11*, and *CLDN18* in the COAD may be regulating the tumor microenvironment.

Moreover, we analyze the relationship between the immune infiltrations and the expression level of differentially expressed claudins genes members ($R > 0.20$ and $P < 0.001$) (**Figure 6B**). The expression level of *CLDN5* is positively correlated with the various immune signatures, including B cells, CAFs, CD4 Regulatory T cells, CD8 T cell, Endothelial cells, Immune checkpoint genes, M2 Macrophages, Macrophages, MDSCs, Neutrophils, pDC, T cell exhaustion, TAM, Tfh, TILs, and Tregs ($R > 0.20$ and $P < 0.001$) (**Figure 6B**). Also, the expression level of *CLDN11* is positively correlated with the activity of B cells, CAFs, CD8 T cell, Endothelial cells, Immune checkpoint genes, M2 Macrophages, Macrophages, MDSCs, Neutrophils, T cell exhaustion, TAM, Tfh, TILs, and Tregs ($R > 0.20$ and $P < 0.001$) (**Figure 6B**). In addition, the expression level of *CLDN18* is positively correlated with the various immune signatures, including B cells, CD8 T cell, Immune checkpoint genes, M2 Macrophages, Macrophages, MDSCs, Neutrophils, T cell exhaustion, TAM, Tfh, Th17, TILs, and Tregs ($R > 0.20$ and $P < 0.001$) (**Figure 6B**). On the other hand, we found a significant negative correlation of *CLDN9* with some immune signatures, including, CD8

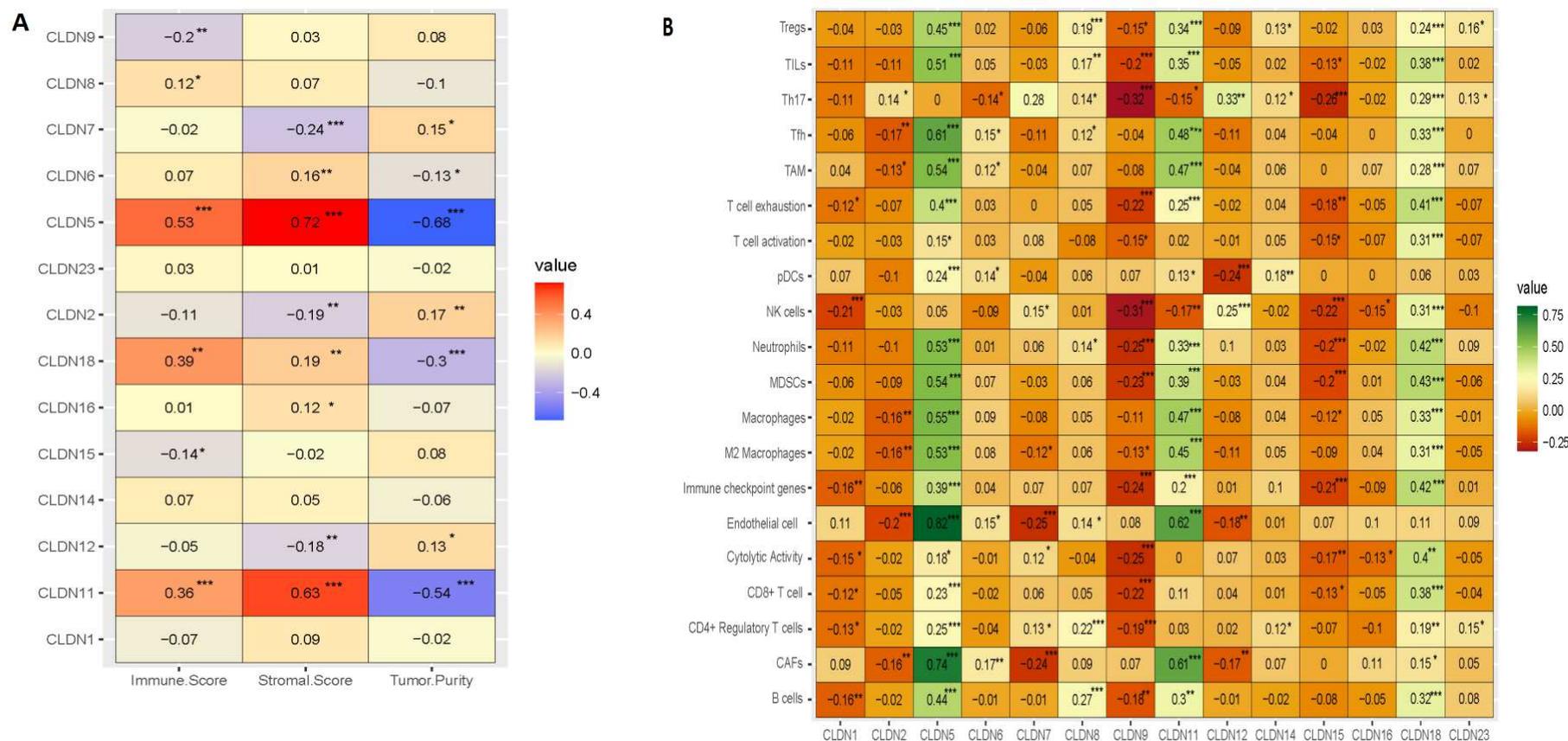


Figure 6. **The association of claudins with the regulation of tumor microenvironment in the COAD.** **A.** Claudins are associated with immune scores, stromal scores, and tumor purity. **B.** The various claudins genes are correlated with immune infiltrations in the COAD.

T cell, cytolytic activity, Immune checkpoint genes, MDSCs, Neutrophils, NK cells, T cell exhaustion, Th17, and TILs ($R > -0.20$ and $P < 0.001$) (**Figure 6B**). Similarly, the expression level of *CLDN15* is negatively correlated with Immune checkpoint genes, MDSCs, Neutrophils, NK cells, and Th17 ($R > -0.20$ and $P < 0.001$) (**Figure 6B**). Altogether, it indicated that the expression level of *CLDN5*, *CLDN9*, *CLDN11*, *CLDN15*, and *CLDN18* are associated with the immune infiltrations in the COAD.

3.6 Claudins are correlated with cancerous biological phenotypes in COAD

Angiogenesis, apical junction, apoptosis, epithelial-mesenchymal transition (EMT), hypoxia, Proliferation, and stemness are major cancerous biological processes in cancers. We investigated the association of these biological processes with the expression level of claudins genes (Absolute value of Spearman Correlation is 0.20 and $P < 0.001$). We revealed that the expression of *CLDN5* and *CLDN11* are positively correlated with angiogenesis, apical junction, apoptosis, epithelial-mesenchymal transition (EMT), and hypoxia, and negatively (**Figure 7**) correlated with proliferation and stemness, indicating that it may be acted as a supportive and protective factor in COAD. The expression level of *CLDN18* is positively correlated with apoptosis and hypoxia (**Figure7**). It indicates that the expression of claudins regulating the aggressive phenotypes of cancers in the COAD.

3.7 Claudins gene family members regulating the cancer-associated pathways

Since our analysis identifying the association of claudins with immune infiltrations and cancerous biological processes, we investigated the correlation of claudins with cancerous pathways activity (Absolute value of Spearman Correlation is 0.20 and $P < 0.001$). Interestingly, the expression level of *CLDN5* and *CLDN11* are positively associated with the activity of cell adhesion molecules CAMs, ECM receptor interaction, ERBB signaling pathway, focal adhesion,

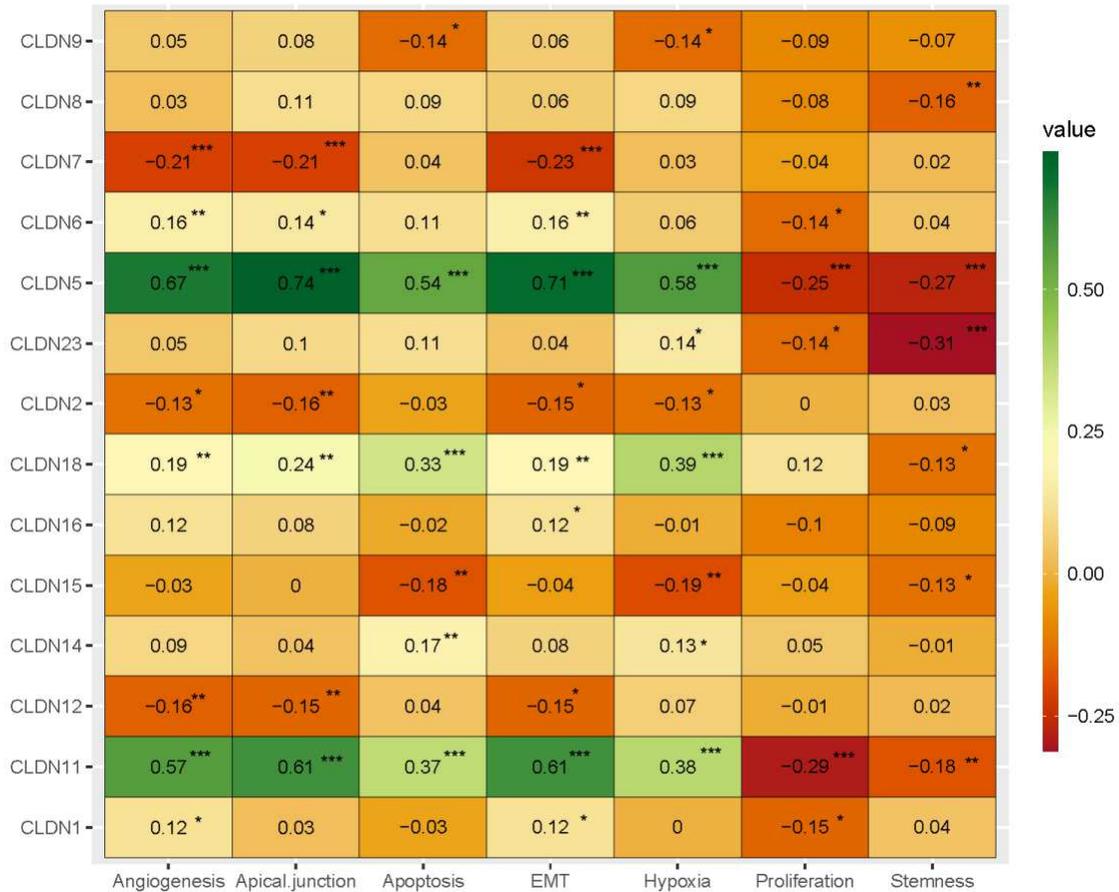


Figure 7. The association of claudins with the regulation of angiogenesis, apical junction, apoptosis, epithelial-mesenchymal transition (EMT), hypoxia, proliferation, and stemness in the COAD.

gap junction, leukocyte transendothelial migration, MAPK signaling pathway, MTOR signaling, Notch signaling, pathways in cancer, TGF beta signaling, tight junction, VEGF signaling pathway, and Wnt signaling (**Figure 8**). Besides, the expression level of *CLDN2*, *CLDN7*, and *CLDN12* are negatively correlated with some of the cancer-associated pathways (**Figure 8**). *CLDN2* is negatively correlated with gap junction and *CLDN12* is negatively correlated with MTOR signaling and Notch signaling pathways (**Figure 8**). Besides, the expression level of *CLDN7* is negatively associated with the activity of ECM receptor interaction, focal adhesion, MTOR

signaling, Notch signaling, TGF beta signaling, and Wnt signaling pathways (**Figure 8**). Altogether, it indicates that the expression of *CLDN5*, *CLDN11*, *CLDN2*, *CLDN7*, and *CLDN12* is associated with regulating the key cancer-associated pathways in the COAD.

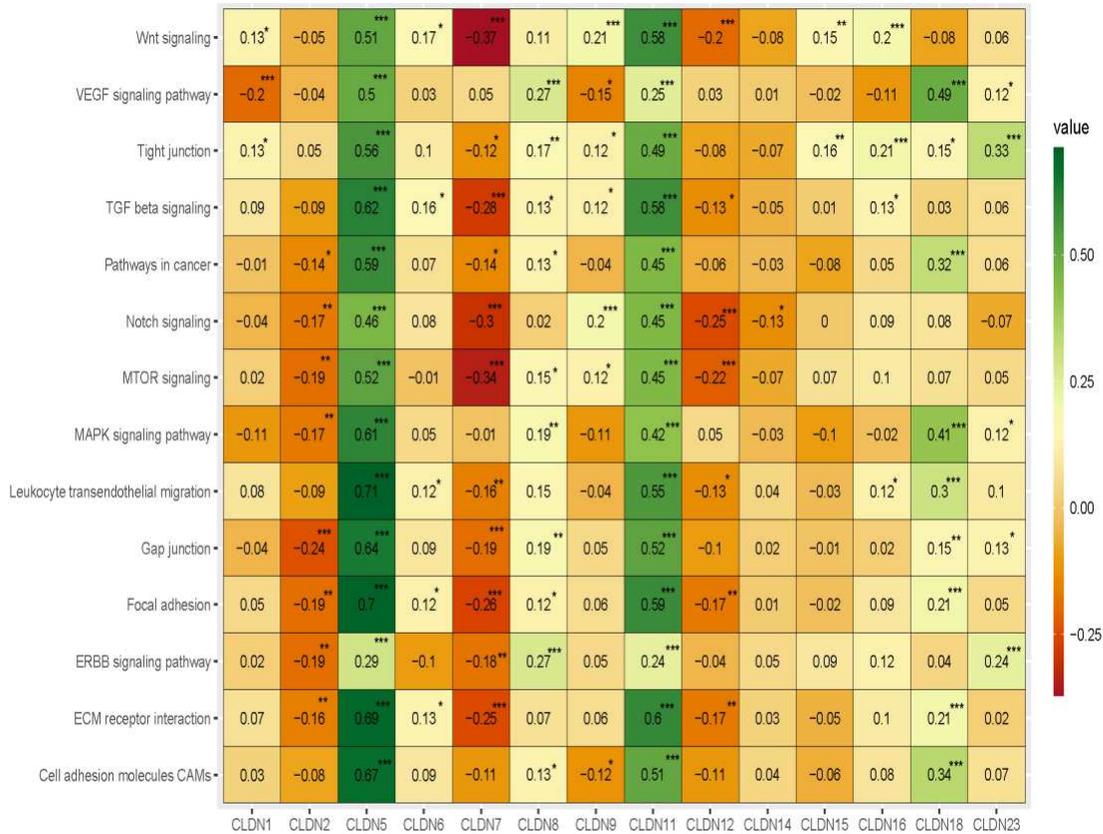


Figure 8. The association of claudins with the regulation of cancer-associated pathways in the COAD.

3.8 Claudins gene family members are mutated in the COAD

We investigated the genetic alterations of all differentially expressed claudins in the COAD. We found that the 14 Queried DEGs genes are altered in 38 (17%) of queried patients out of the 220 patients with mutation and CNA data. The top mutated claudins in the COAD is CLDN23 (7%) (**Figure 9**). The other mutated claudins are *CLDN2* (0.5%), *CLDN14* (1.4%), *CLDN16* (2.3%),

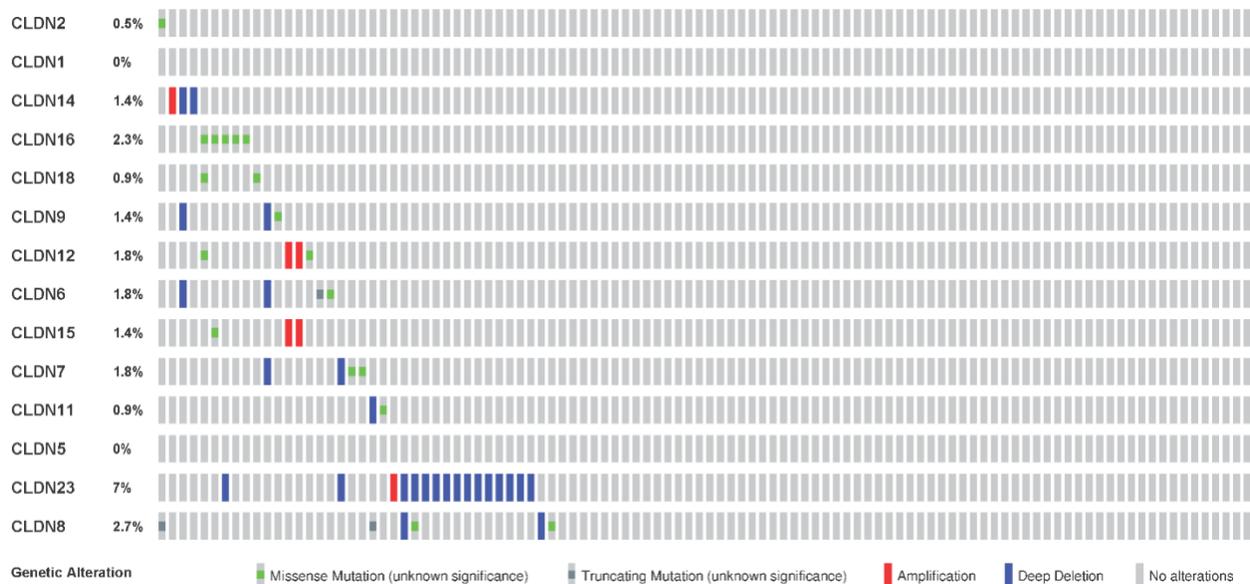


Figure 9. Mutation of differentially expressed claudin genes in the COAD. The mutation was evaluated by using the TCGA COAD dataset in cBioPortal.

CLDN18 (0.9%), *CLDN9* (1.4), *CLDN12* (1.8%), *CLDN8* (2.7%), *CLDN11* (0.9%), *CLDN7* (1.8%) and *CLDN15* (1.4) (**Figure 9**).

3.9 Claudins exhibited the diagnostic efficacy in colon cancers

We speculate that these differentially expressed claudins genes (*CLDN2*, *CLDN1*, *CLDN14*, *CLDN16*, *CLDN18*, *CLDN9*, *CLDN12*, *CLDN6*, *CLDN8*, *CLDN23*, *CLDN5*, *CLDN11*, *CLDN7*, and *CLDN15*) have diagnostic value in colon cancer. We used the TCGA COAD dataset to validate our hypothesis, and the results showed that the ROC curve of the expression levels of these genes showed excellent diagnostic value for colon cancer cases (AUC>0.5) (**Figure 10**). We found that the AUC value of *CLDN2*, *CLDN1*, *CLDN14*, *CLDN16*, *CLDN12*, *CLDN8*, *CLDN23*, *CLDN5*, *CLDN11*, and *CLDN7* is above 0.80, indicating that these claudins have strong diagnostic value for colon cancer patients. Altogether, it can be hypothesized that the claudins are associated with the diagnostic efficacy in COAD.

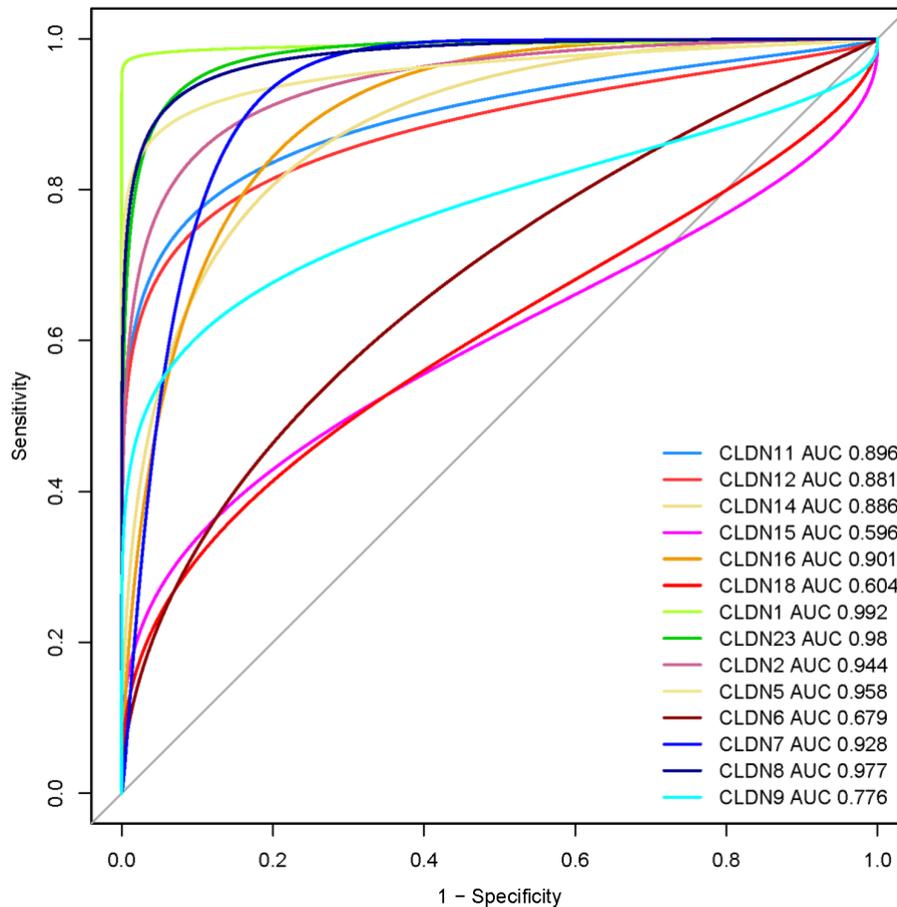


Figure 10. Evaluation of diagnostic efficacy of differentially expressed claudin genes in the COAD. The receiver operating characteristic (ROC) curve of claudin genes in colon cancer (TCGA COAD dataset).

4. Discussion

Since deregulated claudins are associated with neoplastic transformation, tumor biology, tumor development, cancer cells migration and dissemination, and cancer metastasis ^{5,34}, it is necessary to identify aberrantly expressed claudins and their carcinogenic roles in colon cancer. To fill-up this knowledge gap, we identified deregulated 14 claudins in COAD (**Table 1 and Figure 1**). Some previous studies consistently revealed that claudins are associated with the pathogenesis of colon and other cancers. For example, it was demonstrated that the expression of *CLDN1* is a novel

marker for colorectal cancer ³⁵. The upregulated level of *CLDN2* increased the tumorigenicity of colon cancer cells ¹⁰. A higher level of Claudin-2 promotes colorectal cancer liver metastasis and it acted as a crucial biomarker for the replacement type growth pattern ³⁶. Chang-Peng Li et al identified that the *CLDN14* is a potentially direct target for EZH2-mediated H3K27ME3 in human hepatocellular carcinoma ³⁷. The expression level of *CLDN18* is correlated with poor survival in patients with CRC and is associated with the phenotype of gastric cancer ³⁸. In addition to the upregulated claudins, various claudins are downregulated in colon cancer. Sara Cherradi et al demonstrated that *CLDN5*, *CLDN7*, and *CLDN23* were downregulated in CRC samples ³⁹. Altogether, it indicates that deregulated claudins are associated with colon cancer pathogenesis. Moreover, we found that claudins are associated with the enrichment of gene ontology and signaling pathways (**Figure 2 and Figure 3**). GO and pathway analysis revealed that the significant terms are mainly involved with immune regulation and cellular communication (**Figure 2 and Figure 3**). Ryan C Winger et al revealed that claudins are associated with the leukocyte transendothelial migration in a human model of the blood-brain barrier ⁴⁰. The claudins are the backbone of tight junction which controlling the signaling pathways in Inflammation, cell proliferation, transformation, and metastasis ⁴¹. claudins have been identified as crucial cell adhesion molecules working at tight junctions ⁴². Altogether, it indicates that the expression of claudins regulating the various biological signaling pathways in colon cancer. In addition, we identified hub claudins that are interacted with other members (**Figure 4A**). It was demonstrated that the aberrant expression of *CLDN2*, *CLDN4*, *CLDN5*, *CLDN7*, and *CLDN23* are associated with the clinical value in colorectal tumors ³⁹. These findings indicated that the claudins are involved in the PPI network-mediated cellular signaling.

Then, we found that the expression of three deregulated claudins (*CLDN11*, *CLDN14* and *CLDN23*) are significantly correlated with the shorter survival time of colon cancer patients (**Figure 5**). Since the level of immune infiltration is a substantial predictor of a patient's survival in cancer⁴³, we analyzed the correlation of deregulated claudins with the immune infiltration levels in COAD. The previous reports demonstrated that claudins are associated with immune infiltrations in human cancer⁴⁴. Therefore, these consistent findings indicated that claudins are crucial predictors of patient's survival prognosis and immune infiltrations in COAD. Angiogenesis, apical junction, apoptosis, epithelial-mesenchymal transition (EMT), hypoxia, proliferation, and stemness are major cancerous biological processes that influences the pathogenesis of disease. We investigated the association of these biological processes with the expression level of claudins in COAD (**Figure 7**). The expression of *CLDN5* is associated with breast cancer cell motility, indicating the role of *CLDN5* in the metastasis of human breast cancer⁴⁵. The regulatory axis Snail-claudin-11 influences the formation of circulating tumor cell clusters, which are associated with tumor progression⁴⁶. The downregulation of Claudin-7 induces metastasis and invasion in colorectal cancer via the promotion of EMT⁴⁷. For identifying the implications for disease behavior and prevention, it is necessary to identify the correlation of claudins with cancer associated pathways. We found that several deregulated claudins are associated with the activity of cancer associated pathways (**Figure 8**). It was indicated that the claudins are related to the cancerous-associated pathways in cancers. For example, the hypermethylation of the *CLDN11* promoter region in CRC cells committed the metastasis of cells³³. Altogether, it indicates that the expression of claudins is associated with the regulation of cancerous phenotypes and pathways in COAD. Furthermore, we found that the claudins are mutated (**Figure 9**) and it has strong diagnostic value for colon cancer patients (**Figure 10**). Recently, it was stated that the *CLDN15* is

a diagnostic marker for malignant pleural mesothelioma ⁴⁸. *CLDN1*, a gene with diagnostic value, acted as the novel marker in CRC 35. *CLDN7*, with emerging clinical significance, is also a diagnostic marker in the COAD ⁴⁹. *CLDN14*, an upregulated prognostic gene, influence colorectal cancer progression through controlling the PI3K/AKT/mTOR pathway ⁵⁰. Altogether, it can be stated that the claudins are associated with the diagnostic efficacy in COAD.

5. Conclusions

The identification of key claudin's gene associated with prognosis, immune regulation, signaling regulations, and diagnosis may provide insight into the new avenue of colon cancer treatment. In summary, the expression of claudins, especially *CLDN5*, *CLDN11*, and *CLDN18* are substantially associated with immune regulation, cancer-associated pathways, cancerous phenotypes, and diagnostic efficacy in the COAD. The experimental validation of these key claudins would be necessary before applying these present findings into clinical translation for treating colon cancer. Therefore, these comprehensive studies can then be used for identifying the crucial oncogenic functions of claudins in colon cancer.

Data Availability

The TCGA-COAD cohort is downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>).

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

All authors have contributed to the research concept, design, and interpretation of the data. All authors read and approved the final version of the document.

Supplementary materials

Supplementary Table S1. The list of immune markers for identifying ssGSEA scores

Supplementary Table S2. The enriched list of biological processes associated with the claudins.

References

1. Mineta, K. *et al.* Predicted expansion of the claudin multigene family. *FEBS Lett.* **585**, 606–612 (2011).
2. Zhu, L. *et al.* Claudin Family Participates in the Pathogenesis of Inflammatory Bowel Diseases and Colitis-Associated Colorectal Cancer. *Front. Immunol.* **10**, 1441 (2019).
3. Krause, G. *et al.* Structure and function of claudins. *Biochim. Biophys. Acta - Biomembr.* **1778**, 631–645 (2008).
4. Markov, A. G., Aschenbach, J. R. & Amasheh, S. The epithelial barrier and beyond: Claudins as amplifiers of physiological organ functions. *IUBMB Life* **69**, 290–296 (2017).
5. Tabariès, S. & Siegel, P. M. The role of claudins in cancer metastasis. *Oncogene* **36**, 1176–1190 (2016).
6. Kwon, M. J. Emerging roles of claudins in human cancer. *Int. J. Mol. Sci.* **14**, 18148–18180 (2013).
7. Singh, A. B., Sharma, A. & Dhawan, P. Claudin family of proteins and cancer: an overview. *J. Oncol.* **2010**, 541957 (2010).
8. Li, J. Targeting claudins in cancer: diagnosis, prognosis and therapy. *Am. J. Cancer Res.* **11**, 3406 (2021).
9. Zuo, D., Zhang, J., Liu, T., Li, C. & Ning, G. Claudin-1 Is a Valuable Prognostic Biomarker in Colorectal Cancer: A Meta-Analysis. *Gastroenterol. Res. Pract.* **2020**, 4258035 (2020).

10. Dhawan, P. *et al.* Claudin-2 expression increases tumorigenicity of colon cancer cells: role of epidermal growth factor receptor activation. *Oncogene* **30**, 3234–3247 (2011).
11. Ueda, J. *et al.* Heterogeneous Expression of Claudin-4 in Human Colorectal Cancer: Decreased Claudin-4 Expression at the Invasive Front Correlates Cancer Invasion and Metastasis. *Pathobiology* **74**, 32–41 (2007).
12. Huo, Q. *et al.* Claudin-1 protein is a major factor involved in the tumorigenesis of colorectal cancer. *Anticancer Res.* **29**, 851–857 (2009).
13. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).
14. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
15. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
16. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
17. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
18. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
19. Chin, C.-H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex

- interactome. *BMC Syst. Biol.* **8 Suppl 4**, S11–S11 (2014).
20. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
 21. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
 22. Hänzelmann, S., Castelo, R. & Guinney, J. GSV: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
 23. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
 24. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
 25. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, eaaf8399 (2017).
 26. Liu, Z., Li, M., Jiang, Z. & Wang, X. A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. *Transl. Oncol.* **11**, 311–329 (2018).
 27. Chen, B., Li, L., Li, M. & Wang, X. HIF1A expression correlates with increased tumor immune and stromal signatures and aggressive phenotypes in human cancers. *Cell. Oncol.* **43**, 877–888 (2020).
 28. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
 29. Yang, Q. *et al.* Candidate Biomarkers and Molecular Mechanism Investigation for Glioblastoma Multiforme Utilizing WGCNA. *Biomed Res. Int.* **2018**, 4246703 (2018).

30. L. Garner, A. & D. Janda, K. Protein-Protein Interactions and Cancer: Targeting the Central Dogma. *Curr. Top. Med. Chem.* **11**, 258–280 (2011).
31. Jia, H., Chai, X., Li, S., Wu, D. & Fan, Z. Identification of claudin-2,-6,-11 and-14 as prognostic markers in human breast carcinoma. *Int. J. Clin. Exp. Pathol.* **12**, 2195 (2019).
32. Yang, H. *et al.* Association of a novel seven-gene expression signature with the disease prognosis in colon cancer patients. *Aging (Albany. NY).* **11**, 8710–8727 (2019).
33. Li, J. *et al.* Methylated claudin-11 associated with metastasis and poor survival of colorectal cancer. *Oncotarget* **8**, 96249–96262 (2017).
34. Li, J. Context-Dependent Roles of Claudins in Tumorigenesis. *Front. Oncol.* **11**, 676781 (2021).
35. Nakagawa, S. *et al.* Expression of CLDN1 in colorectal cancer: A novel marker for prognosis. *Int. J. Oncol.* **39**, 791–796 (2011).
36. Tabariès, S. *et al.* Claudin-2 promotes colorectal cancer liver metastasis and is a biomarker of the replacement type growth pattern. *Commun. Biol.* **4**, 657 (2021).
37. Li, C.-P. *et al.* CLDN14 is epigenetically silenced by EZH2-mediated H3K27ME3 and is a novel prognostic biomarker in hepatocellular carcinoma. *Carcinogenesis* **37**, 557–566 (2016).
38. Matsuda, M. *et al.* Immunohistochemical analysis of colorectal cancer with gastric phenotype: Claudin-18 is associated with poor prognosis. *Pathol. Int.* **60**, 673–680 (2010).
39. Cherradi, S., Martineau, P., Gongora, C. & Del Rio, M. Claudin gene expression profiles and clinical value in colorectal tumors classified according to their molecular subtype. *Cancer Manag. Res.* **11**, 1337–1348 (2019).
40. Winger, R. C., Koblinski, J. E., Kanda, T., Ransohoff, R. M. & Muller, W. A. Rapid

- remodeling of tight junctions during paracellular diapedesis in a human model of the blood-brain barrier. *J. Immunol.* **193**, 2427–2437 (2014).
41. Bhat, A. A. *et al.* Tight Junction Proteins and Signaling Pathways in Cancer and Inflammation: A Functional Crosstalk. *Front. Physiol.* **9**, 1942 (2019).
 42. Mori, M. Expression of CLDN1 in colorectal cancer: A novel marker for prognosis. *Int. J. Oncol.* (2011). doi:10.3892/ijo.2011.1102
 43. Ohtani, H. Focus on TILs: prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer. *Cancer Immun. Arch.* **7**, (2007).
 44. Gao, P. *et al.* Association of CLDN6 and CLDN10 With Immune Microenvironment in Ovarian Cancer: A Study of the Claudin Family. *Front. Genet.* **12**, 595436 (2021).
 45. Escudero-Esparza, A., Jiang, W. G. & Martin, T. A. Claudin-5 is involved in breast cancer cell motility through the N-WASP and ROCK signalling pathways. *J. Exp. Clin. Cancer Res.* **31**, 43 (2012).
 46. Li, C.-F. *et al.* Snail-induced claudin-11 prompts collective migration for tumour progression. *Nat. Cell Biol.* **21**, 251–262 (2019).
 47. Wang, K. *et al.* Claudin-7 downregulation induces metastasis and invasion in colorectal cancer via the promotion of epithelial-mesenchymal transition. *Biochem. Biophys. Res. Commun.* **508**, 797–804 (2019).
 48. Watanabe, M. *et al.* CLDN15 is a novel diagnostic marker for malignant pleural mesothelioma. *Sci. Rep.* **11**, 12554 (2021).
 49. Wang, K., Xu, C., Li, W. & Ding, L. Emerging clinical significance of claudin-7 in colorectal cancer: a review. *Cancer Manag. Res.* **10**, 3741–3752 (2018).
 50. Qiao, T.-Y. *et al.* Claudin14 promotes colorectal cancer progression via the

PI3K/AKT/mTOR pathway. *Neoplasma* **68**, 947–954 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1and2.pdf](#)