Analysis of multi-omics differences in left-side and right-side colon cancer

Yanyi Huang[1,2,*], Jinzhong Duanmu[1,*], Yushu Liu[1,2], Mengyun Yan[1,3], Taiyuan Li[1] and Qunguang Jiang[1]

[1] Department of Gastrointestinal Surgery, First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
[2] Nanchang University, The Second Clinical Medicine College, Nanchang, Jiangxi, China
[3] Nanchang University, The First Clinical Medicine College, Nanchang, Jiangxi, China
* These authors contributed equally to this work.

ABSTRACT

Background: Colon cancer is one of the most common tumors in the digestive tract. Studies of left-side colon cancer (LCC) and right-side colon cancer (RCC) show that these two subtypes have different prognoses, outcomes, and clinical responses to chemotherapy. Therefore, a better understanding of the importance of the clinical classifications of the anatomic subtypes of colon cancer is needed.

Methods: We collected colon cancer patients' transcriptome data, clinical information, and somatic mutation data from the Cancer Genome Atlas (TCGA) database portal. The transcriptome data were taken from 390 colon cancer patients (172 LCC samples and 218 RCC samples); the somatic mutation data included 142 LCC samples and 187 RCC samples. We compared the expression and prognostic differences of LCC and RCC by conducting a multi-omics analysis of each using the clinical characteristics, immune microenvironment, transcriptomic differences, and mutation differences. The prognostic signatures was validated using the internal testing set, complete set, and external testing set (GSE39582). We also verified the independent prognostic value of the signature.

Results: The results of our clinical characteristic analysis showed that RCC had a significantly worse prognosis than LCC. The analysis of the immune microenvironment showed that immune infiltration was more common in RCC than LCC. The results of differential gene analysis showed that there were 360 differentially expressed genes, with 142 upregulated genes in LCC and 218 upregulated genes in RCC. The mutation frequency of RCC was generally higher than that of LCC. BRAF and KRAS gene mutations were the dominant genes mutations in RCC, and they had a strong mutual exclusion with APC, while APC gene mutation was the dominant gene mutation in LCC. This suggests that the molecular mechanisms of RCC and LCC differed. The 4-mRNA and 6-mRNA in the prognostic signatures of LCC and RCC, respectively, were highly predictive and may be used as independent prognostic factors.

Conclusion: The clinical classification of the anatomic subtypes of colon cancer is of great significance for early diagnosis and prognostic risk assessment. Our study provides directions for individualized treatment of left and right colon cancer.

Submitted 3 December 2020
Accepted 20 April 2021
Published 12 May 2021

Corresponding author
Qunguang Jiang, fbron.student@sina.com

Academic editor
Elaine Dunlop

Additional Information and Declarations can be found on page 19

DOI 10.7717/peerj.11433

Subjects Bioinformatics, Gastroenterology and Hepatology, Oncology, Computational Science
Keywords Left-side colon cancer, Right-side colon cancer, Mutation, Gene expression, Prognosis, Immune microenvironment

INTRODUCTION

Colon cancer is one of the most common cancers in the world and it is the second leading cause of cancer-related deaths in the United States (*Siegel et al., 2020*). The location of the tumor itself has not received much attention due to the belief that accurately locating the tumor would not affect patient survival.

However, in the past decade the differences between LCC and RCC have received more attention (*Mik et al., 2017*). The embryonic origin may help explain the genesis of this disease (*Bufill, 1990*). RCC is known to originate from the midgut, which includes the cecum, ascending colon, and hepatic flexure. In contrast, LCC originates from the hindgut, which includes the splenic flexure, descending colon, and sigmoid colon.

LCC and RCC have received increased attention because of clear differences in their prognosis, outcomes, and clinical response to chemotherapy. It has been reported that LCC is associated with a better prognosis compared with RCC (*Kalantzis et al., 2020*). A recent systematic review noted that many studies have identified differences in their epidemiology, clinical presentation, pathology, and genetic mutations through anatomical subsites (*Imperial et al., 2018*).

Most of the studies indicated that patients with RCC showed lower survival rates compared with LCC (*Nakagawa-Senda et al., 2019*). However, the data are still controversial. Weiss et al. demonstrated that when analysis was adjusted for multiple variables, including patient, disease, comorbidity, and treatment, there was no overall difference in the 5-year mortality between LCC and RCC (*Weiss et al., 2011*).

Additional studies on LCC and RCC are needed. We performed a multi-omics analysis of LCC and RCC using clinical characteristics, the immune microenvironment, transcriptomic differences, and mutation differences to determine the importance of classifying these anatomic colon cancer subtypes.

METHOD AND DATA

Data collection and preprocessing

First, we downloaded colon cancer patients' transcriptome data, clinical information, and somatic mutation data from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov). The transcriptome data were comprised of 390 colon cancer patients (172 LCC samples and 218 RCC samples), and the somatic mutation data were comprised of 329 colon cancer patients (142 LCC samples and 187 RCC samples). According to Dwertmann et al., the LCC consists of the descending colon, sigmoid colon, and splenic flexure of colon and the RCC consists of the ascending colon, cecum, and hepatic flexure of colon (*Hsu et al., 2019*). We used genecode.v22.annotation (https://www.gencodegenes.org/) to comment on the transcriptional data downloaded from TCGA database.

Clinical analysis in LCC and RCC

We used R to classify the data used to analyze the differences between LCC and RCC in terms of age, gender, pT, pN, pM, pStage and survival. Pearson's chi-square ($\chi 2$) test was used to calculate differences in clinical characteristics between LCC and RCC. We compared the overall survival rates of LCC and RCC in different clinical subtypes using the survival package in R.

Immune microenvironment in LCC and RCC

We obtained immune-related gene sets with 29 immune cell types and immune-related functions from previous studies (*Bedognetti et al., 2015*; *Ren et al., 2020*; *Liu et al., 2018*; *Maibach et al., 2020*; *Aran et al., 2016*) to explore the differences between LCC and RCC in the immune microenvironment. We used the single sample gene set enrichment analysis (ssGSEA) algorithm to obtain the scores of 29 immune cell types and immune-related functions with the 'GSVA' package in R (*Hänzelmann, Castelo & Guinney, 2013*). We visualized the results using the pheatmap package in R (*Galili et al., 2018*). We used the estimate package in R to analyze the differences between LCC and RCC in the immune microenvironment to calculate the immune score, stromal score, ESTIMATE score, and tumor purity. Then we compared the differences between the two groups using the Mann–Whitney *U* test. We also compared the expression levels of the human leukocyte antigen (HLA) gene family and immune check-point genes, and the abundance of immune cell infiltration in LCC and RCC. We obtained the immune cell infiltration data using CIBERSORT (*Chen et al., 2018*).

Screening differential genes in LCC and RCC

We used the |log2 fold change(LogFC)|>1 and the false discovery rate (FDR) <0.05 with the Wilcox test to identify the differences in expression of mRNAs in LCC and RCC. The results were visualized using heatmap and volcano diagrams. Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to determine the enrichment of the differential genes through the Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) (*Yang et al., 2019*). The top 20 biological processes (BP) of GO enrichment analysis (*Chen et al., 2017*) were depicted in a circle diagram, and the top 15 KEGG pathways (*Altermann & Klaenhammer, 2005*) were depicted in a bubble diagram.

Screening prognostic mRNAs in LCC and RCC by univariate COX analysis

We included different genes in our study. We used the survival package in R with $P < 0.005$ to identify the prognostic mRNAs in LCC and RCC, respectively, using univariate COX (*Tibshirani, 2009*). According to univariate COX analysis, there were 22 genes associated with the prognosis of RCC, with a potential collinear relationship among them. We used the LASSO regression algorithm with a penalty term to delete genes with multicollinearity for additional analysis.

Construction and verification of the prognostic signature and validation of prognostic models

Prognostic genes related to LCC and RCC were included in our study. We set up a random number seed in order to divided LCC patients from TCGA into a training set and an internal testing set with a 1:1 ratio and established a 4-mRNA LCC prognosis model using multivariate COX regression analysis with a noose penalty (*Grant, Hickey & Head, 2019*; *Zhou et al., 2019*). We used the same method to establish a 6-mRNA RCC prognosis signature (*Marisa et al., 2013*; *Zhang et al., 2016*; *Zhang et al., 2020*). The samples was divided into two groups using the median risk score. We judged the efficacy of the model by plotting the Kaplan–Meier (KM) curve and receiver operating characteristic (ROC) curve (*Obuchowski & Bullen, 2018*). The GSE39582 data set was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582) (*Barrett et al., 2013*) and was used as an external validation set. We validated the model by plotting the Kaplan–Meier (KM) curve. The GSE39582 data set included 566 colon cancer samples (342 LCC samples and 224 RCC samples) and their corresponding survival information in accordance with the GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) (Table S1). We performed an independent prognostic analysis of the risk score in the total TCGA set to further verify the model's efficacy. The risk score was calculated as (*Cho et al., 2019*):

$$Riskcore = \sum_{i=1}^{N}(Expi \times Coef) \tag{1}$$

with N representing the number of signature genes, Expi representing the gene expression levels, and Coef representing the estimated regression coefficient value from the Cox proportional hazards analysis.

Single gene mutation analysis in LCC and RCC

On the JAVA8 platform, we analyzed the number of variants and the length of exons for each sample using Perl scripts to calculate mutation frequency (*Tabibzadeh et al., 2020*). Samples were divided into two groups according to the location of colon cancer and the Mann–Whitney test (*McGee, 2018*) was used to compare the tumor mutation burden (TMB) difference between two groups. We used the maftools package (*Mayakonda et al., 2018*) for visualization and performed Fisher's exact test in pairs between the top 25 mutated genes to analyze the mutational exclusion and co-occurrence. We also used oncoplot in R to visualize the top 30 mutated genes from the 142 LCC samples and 187 RCC samples to produce waterfall plots. Then we used the ggplot2 and boxplot packages to visualize the classification and frequency of mutation types, frequency of variant types, frequency of SNV classes, the tumor mutation burden in specific samples, and the top 10 mutated genes in LCC and RCC. The top 10 mutated genes in LCC were: *APC*, *TP53*, *TTN*, *KRAS*, *MUC16*, *SYNE1*, *FAT4*, *RYR2*, *PIK3CA*, and *OBSCN*. The top 10 mutation genes in RCC were: *TTN*, *APC*, *MUC16*, *SYNE1*, *TP53*, *KRAS*, *FAT4*, *PIK3CA*, *PCLO*, and *ZFHX4*.

Table 1 Clinical features for the COAD patients in the LCC and RCC in TCGA.

Parameters	LCC patients (n = 172)	RCC patients (n = 218)	χ^2	P value
Age, y			7.814	0.005
≤65	85	76		
>65	87	142		
Gender			0.035	0.852
Male	89	116		
Female	83	102		
pT			0.263	0.608
T1-2	37	41		
T3-4	135	176		
unknow	0	1		
pN			3.022	0.082
N0	93	138		
N1-2	79	80		
pM			2.099	0.147
M0	126	162		
M1	31	25		
unknow	15	31		
pStage			2.934	0.087
Stage I–II	89	130		
Stage III–IV	81	81		
unknow	2	7		
Survival			5.122	0.024
Alive	144	163		
Dead	26	55		
unknow	2	0		

Note:
LCC: Left-side colon cancer; RCC: Right-side colon cancer; TCGA: The Cancer Genome Atlas; χ^2: Chi-square value.

RESULTS

Differences in clinical characteristics between LCC and RCC

The LCC and RCC data in the TCGA database and the results of the chi-square test on clinical characteristics are shown in Table 1. We classified the data by stage, T, N, M, and age after separating the data by LCC and RCC. We used the Kaplan–Meier (KM) curve of over survival (OS) to compare the survival differences of different clinical characteristics between the two groups. The results indicated that RCC had a worse prognosis than LCC, which was also seen in stages III-IV, T3-4, and N1-2 (Figs. 1A–1D). The survival rate of RCC was worse than that of LCC (Figs. 1E–1F) although there was no statistical difference between the M1 and age > 65 subgroups.

Immune microenvironment landscape between LCC and RCC

The ssGSEA algorithm showed that 29 types of immune cells and their functions were enriched in each sample. We then obtained the immune score, stromal score, ESTIMATE

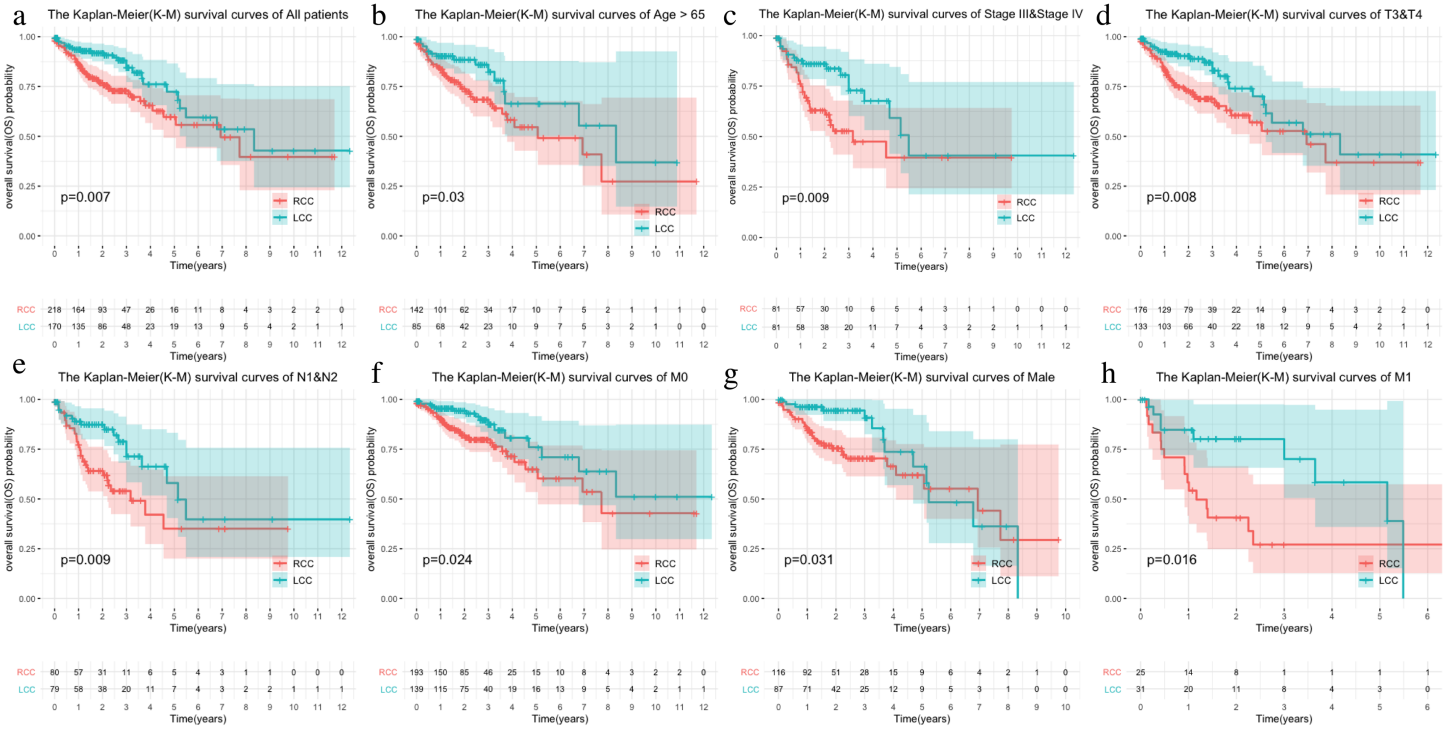

Figure 1 Comparison of survival rates of LCC and RCC in different clinical subtypes. Survival analysis of different clinical characteristics including (A) all patients, (B) Age > 65, (C) Stage III & Stage IV, (D) T3&T4, (E) N1&N2, (F) M0, (G) Male, (H) M1.

Full-size DOI: 10.7717/peerj.11433/fig-1

score, and tumor purity. The heatmap indicated that the RCC had a higher immune invasion than the LCC (Fig. 2A). Comparing the two groups' scores, we found that only the immune scores were significantly different (Fig. 2B). We confirmed that RCC had a higher immune infiltration than LCC by further comparing the expression levels of the *HLA* gene family and immune checkpoint-related genes and the abundance of immune cell infiltration (Figs. 2C–2D). Previous studies have shown that the changes in *HLA* class I genes in colon cancer are closely related to RCC, suggesting microsatellite instability (MSI). In addition, the high expression of *PD-L1* also occurs more frequently in RCC, indicating MSI (*Kikuchi et al., 2019*). Our results support the conclusion that RCC has more immune infiltration and is highly correlated with MSI. Therefore, this result suggests that right-side colon cancer was significantly more reactive than left-side colon cancer in immune response, which might provide new treatments for colon cancer.

Differential gene analysis between LCC and RCC

The Wilcox test was used to extract differential mRNAs to obtain 360 differential genes, which included 218 up-regulated genes in RCC and 142 up-regulated genes in LCC (Figs. 3A–3B). All of the differential genes are shown in Table S2. All of the differentially expressed genes were enriched by the biological processes of GO and KEGG pathways in the DAVID database (Tables S3 and S4). The top 20 biological processes of GO

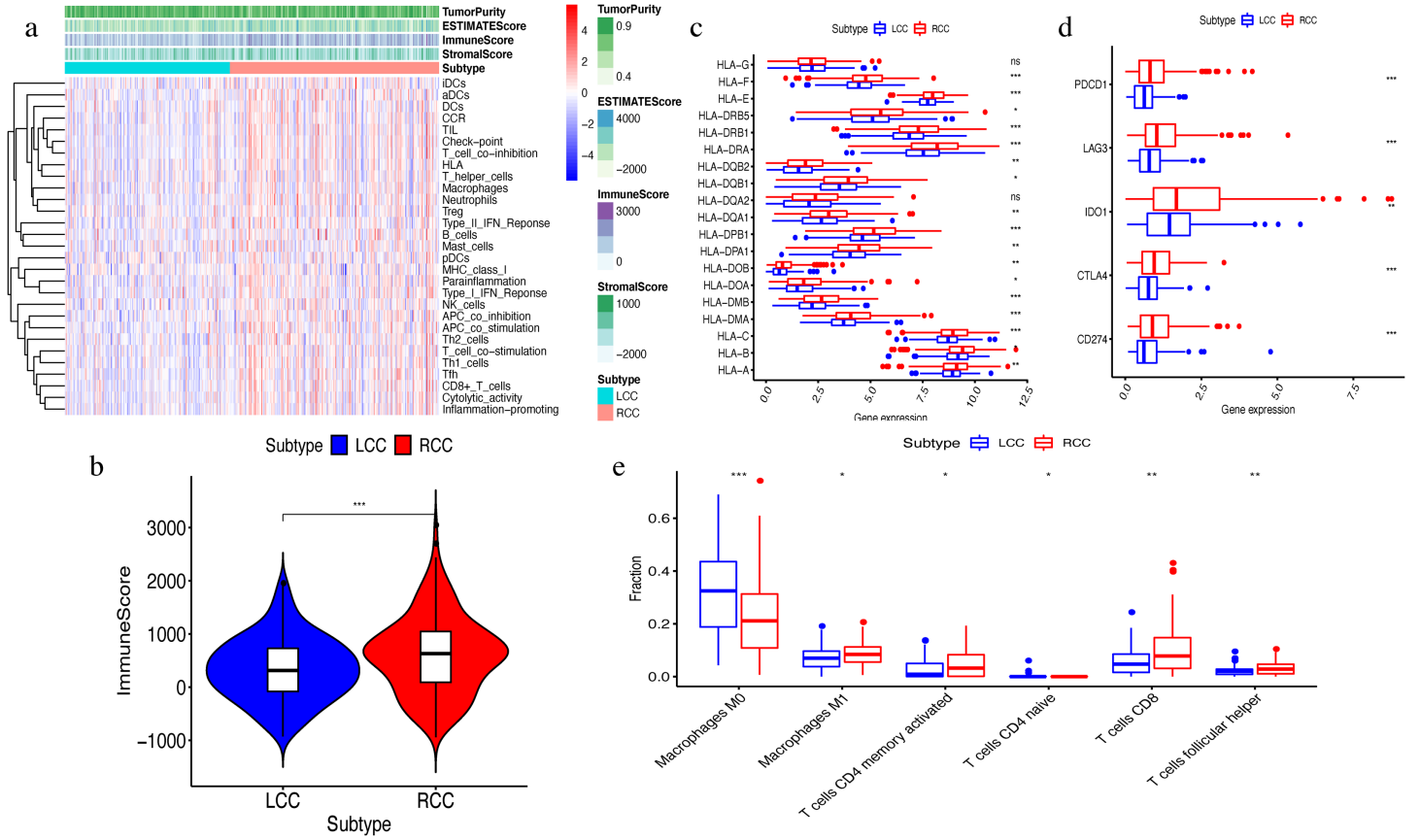

Figure 2 Exploration and validation the differences of immune microenvironment between LCC and RCC. Through ssGSEA, 29 immune-related gene sets were enriched, including immune cells and immune processes. (A) The heat map is also included the tumor purity, ESTIMATE score, immune score and stromal score. (B) Variance analysis of the immune score between LCC and RCC. (C) The expression levels of HLA gene family in samples from LCC and RCC. (D) The expression levels of immune checkpoint genes (PDCD1, LAG3, IDO-1, CTLA4, CD274) in samples from LCC and RCC. (E) The abundance of six types of infiltrating immune cells in samples from LCC and RCC. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Full-size DOI: 10.7717/peerj.11433/fig-2

enrichment analysis were graphed in a circle diagram, while the top 15 KEGG pathways were displayed in a bubble diagram (Figs. 3C–3D). The top three biological pathways were 'associative learning', 'arachidonic acid secretion', and 'anterior/posterior pattern specification'. The differentially expressed genes were significantly enriched in the steroid hormone biosynthesis pathway.

Univariate COX screening of prognostic genes in LCC and RCC

We screened the genes related to the prognosis of LCC and RCC using univariate Cox analysis in the LCC and RCC patients with $P < 0.005$. We obtained 9 genes related to prognosis in LCC and 22 genes related to prognosis in RCC (Tables 2–3). In order to avoid model overfitting, we performed LASSO regression analysis with the penalty term on RCC to solve the multicollinearity problem again by dimension reduction, and finally obtained 12 genes related to prognosis in RCC (Fig. 5B).

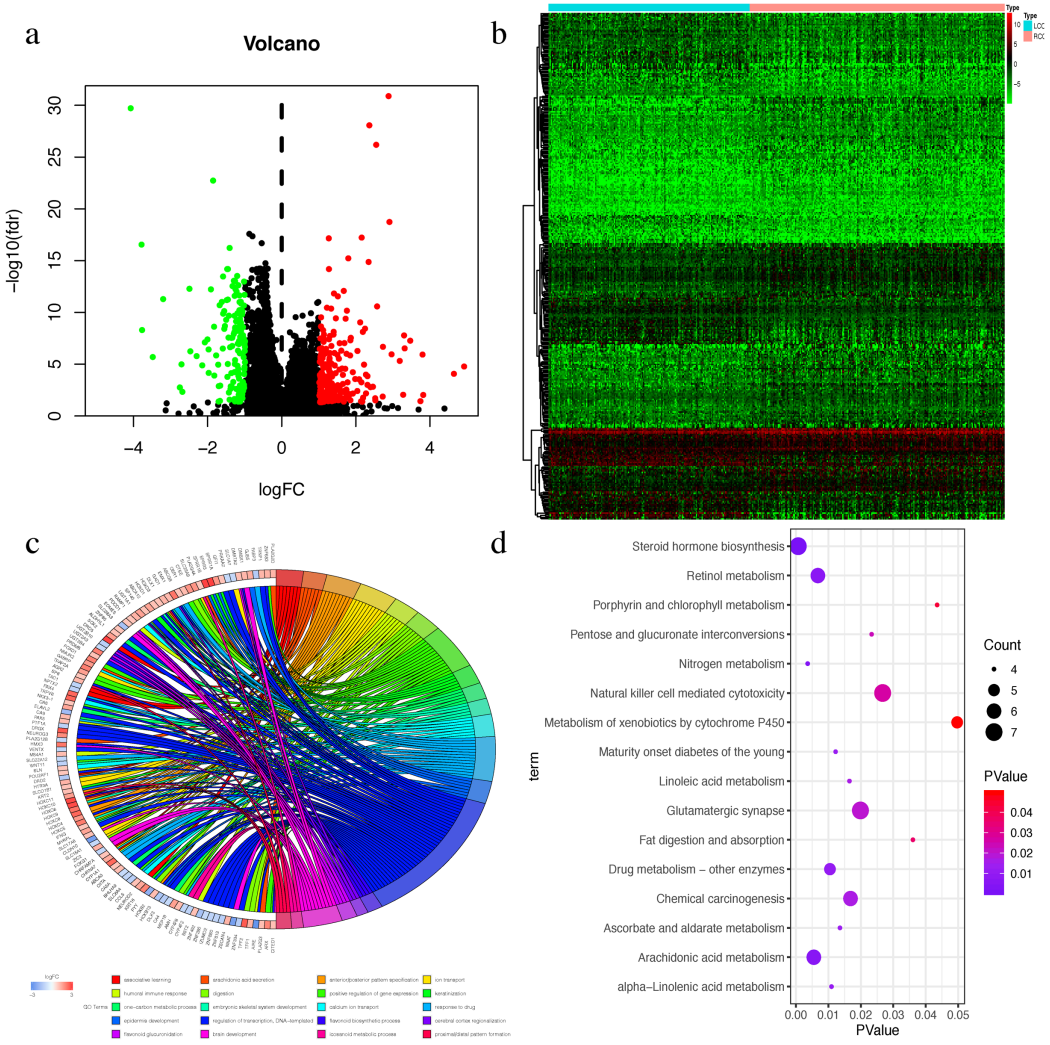

Figure 3 The Differential expressed mRNAs in LCC and RCC. (A) Volcano plot for differential expressed mRNAs in LCC and RCC. Green dots represent up-regulated genes in LCC, while red dots represent up-regulated genes in RCC. (B) Heatmap of differential expressed mRNAs between LCC and RCC. (C) Circle diagram demonstrated the top 20 biological processes of GO enrichment analysis. (D) Bubble diagram demonstrated the top 15 KEGG pathways. Full-size DOI: 10.7717/peerj.11433/fig-3

Construction of prognosis signature in LCC and RCC

TCGA LCC patients were divided into a training set and an internal testing set at a 1:1 ratio. Multivariate COX regression analysis with noose penalty was then used to establish a 4-mRNA LCC prognosis signature and a 6-mRNA RCC prognosis signature.

The 4-mRNA LCC prognosis signature and risk score were calculated as: *C1orf105**0.458+*FAM132B**1.703+*TNNT1**0.130+*RSPO4**0.268 (Table 4). The median risk score (0.622) in the training set was used to assign patients to the high risk or low risk group. Patients with a high risk score had significantly worse survival rates than those with low-risk scores (P = 0.046, Fig. 4A). Furthermore, the AUC of the risk score for 1-year, 2-year, 3-year, and 5-year OS were 0.751, 0.810, 0.860, and 0.904, respectively

Table 2 The prognostic mRNAs by univariable Cox analysis in LCC.

mRNA	HR	95% CI		*P* value
		Low	High	
C1orf105	1.412	1.154	1.729	0.001
OSR1	10.074	2.908	34.898	<0.001
FAM132B	3.310	1.597	6.858	0.001
WNT7A	1.750	1.212	2.525	0.003
FDCSP	1.083	1.026	1.144	0.004
SMTNL2	1.317	1.109	1.564	0.002
FCER2	1.446	1.144	1.828	0.002
TNNT1	1.114	1.050	1.182	<0.001
RSPO4	1.230	1.074	1.410	0.003

Note:
LCC: Light-side colon cancer; mRNA: messenger RNA; CI: confidence interval; HR: hazard ratio.

Table 3 The prognostic mRNAs by univariable Cox analysis in RCC.

mRNA	HR	95% CI		*P* value
		Low	High	
LMX1A	3.964	1.712	9.176	0.001
COLGALT2	1.165	1.070	1.270	0.000
SNCB	2.748	1.537	4.912	0.001
OFCC1	48.830	6.787	351.331	0.000
FABP7	4.518	1.896	10.770	0.001
PAX4	1.130	1.045	1.223	0.002
KLRG2	1.253	1.124	1.398	<0.001
PAX5	1.347	1.126	1.612	0.001
PCDH8	166.438	7.157	3,870.700	0.001
HS6ST3	4.836	1.677	13.953	0.004
SYNGR3	1.116	1.045	1.191	0.001
CHST6	1.266	1.111	1.443	<0.001
SLC22A31	1.635	1.324	2.020	<0.001
NEUROD2	1.877	1.299	2.712	0.001
TCAP	1.791	1.250	2.566	0.001
GREB1L	2.482	1.369	4.502	0.003
FCER2	1.118	1.042	1.198	0.002
SLC7A10	2.237	1.370	3.653	0.001
APLP1	1.087	1.029	1.147	0.003
RSPO4	1.479	1.216	1.799	<0.001
INSM1	1.038	1.012	1.064	0.004
CCDC160	1.436	1.200	1.718	<0.001

Note:
RCC: Right-side colon cancer; mRNA: messenger RNA; CI: confidence interval; HR: hazard ratio.

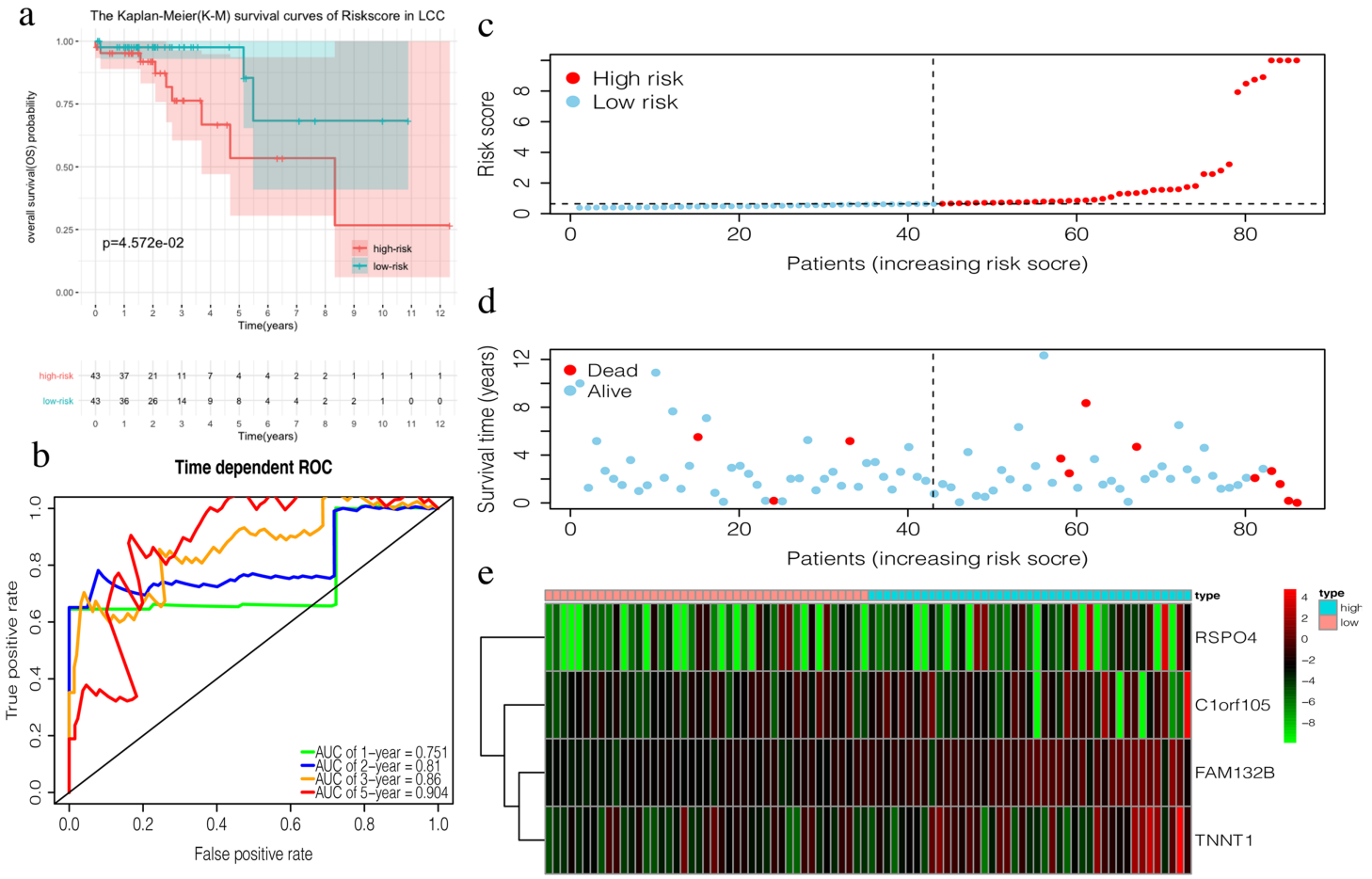

Figure 4 Construction of the prognostic model in the training group of LCC. (A) The Kaplan–Meier (K–M) survival curves in the training set, (B) Time-dependent ROC curves in the training set at 1-year, 2-year, 3-year and 5-year. (C) The survival status of LCC patients in the training group. Green dots represent the patient is still alive, while red dots represent the patient has dead. (D) Risk scores of LCC patients in the training group. Green dots represent the patient assigned to the low risk group, while red dots represent the patient assigned to the high risk group. (E) mRNAs expression levels of four mRNA LCC prognosis signature in the training group. Full-size DOI: 10.7717/peerj.11433/fig-4

(Fig. 4B). The survival status, risk scores, and gene expression data of LCC patients in the training group are shown in Figs. 4C–4E. *RSPO4*, *FAM132B*, and *TNNT1* were highly expressed in the high-risk group, while *C1orf105* was not well-expressed in the high-risk group.

The risk score of the 6-mRNA RCC prognosis signature was calculated as: *OFCC1**4.834 +*KLRG2**0.195+*PAX5**0.461+*SYNGR3**0.096+*SLC22A31**1.232+*CCDC160**0.368 (Table 5). The median risk score (0.689) in the training set was used to assign patients to the high risk or low risk group. Patients with a high-risk score was had a significantly worse survival rate than those with a low-risk score (0.012, Fig. 5A). Furthermore, the AUC of the risk score for 1-year, 2-year, 3-year, and 5-year OS were 0.776, 0.714, 0.670, and 0.792, respectively (Fig. 5C). The survival status, risk scores, and gene expression data of RCC patients in the training group are shown in Figs. 5D–5F. All six genes were highly-expressed in the high-risk group.

Table 4 Multivariate Cox regression modeling in LCC.

id	coef	HR	95% CI		*P* value
			Low	High	
C1orf105	0.458	1.581	1.134	2.205	0.007
FAM132B	1.703	5.492	1.390	21.693	0.015
TNNT1	0.130	1.139	1.026	1.265	0.015
RSPO4	0.268	1.307	1.087	1.572	0.004

Note:
LCC: Light-side colon cancer; CI: confidence interval; HR: hazard ratio.

Table 5 Multivariate Cox regression modeling in RCC.

id	coef	HR	95% CI		*P* value
			Low	High	
OFCC1	4.834	125.723	8.492	1861.210	<0.001
KLRG2	0.195	1.215	1.011	1.461	0.038
PAX5	0.461	1.586	1.201	2.095	0.001
SYNGR3	0.096	1.101	1.007	1.204	0.035
SLC22A31	1.232	3.428	1.599	7.350	0.002
CCDC160	0.368	1.444	1.213	1.720	<0.001

Note:
RCC: Right-side colon cancer; CI: confidence interval; HR: hazard ratio.

Validation of the prognosis signature in LCC and RCC

The prognostic accuracy of the prognosis signature was validated in three independent cohorts, including the testing set, the total TCGA data set, and the GSE39582 data set.

The OS in the high-risk group was significantly worse than that of the low-risk group in the testing set in the 4-mRNA LCC prognosis signature ($P = 0.016$, Fig. 6A). The predicted 1-year, 2-year, 3-year, and 5-year OS was 0.731, 0.760, 0.779, and 0.700, respectively (Fig. 6B). The total TCGA set also validated the prognostic accuracy of the signature ($P = 0.001$, Fig. 6C), with respective AUCs of 0.732, 0.776, 0.820, and 0.793 for 1-year, 2-year, 3-year, and 5-year OS (Fig. 6D).

The OS of the high-risk group was significantly worse than that of the low-risk group in the testing set for the 6-mRNA RCC prognosis signature ($P = 0.042$, Fig. 7A). The predicted 1-year, 2-year, 3-year, and 5-year OS was 0.770, 0.754, 0.689, and 0.646, respectively (Fig. 7B). The total TCGA set validated the prognostic accuracy of the signature ($P = 0.002$, Fig. 7C), with respective AUCs of 0.760, 0.718, 0.663, and 0.718 for 1-year, 2-year, 3-year, and 5-year OS (Fig. 7D).

The GSE39582 data set showed the same conclusion in the 4-mRNA LCC prognosis signature ($P = 0.185$) and the 6-mRNA RCC prognosis signature ($P = 0.25$) ($P = 0.018$, Fig. 6E; $P = 0.025$, Fig. 7E). The survival status, risk scores and gene expression data of LCC and RCC patients in the testing set and total TCGA set are shown in Figs. S2 and S3.

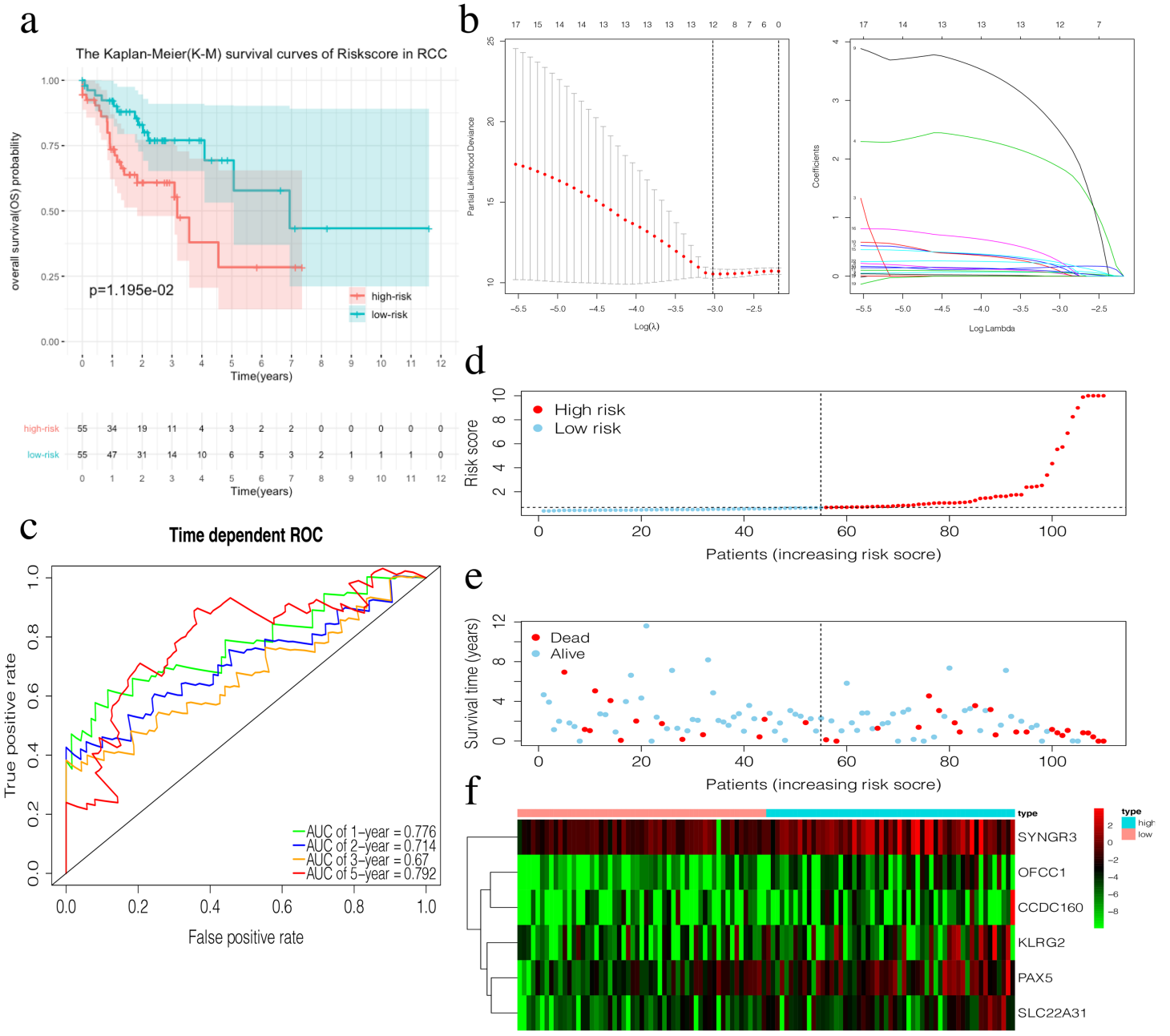

Figure 5 Construction of the prognostic model in the training group of RCC. (A) The Kaplan–Meier (K–M) survival curves in the training set. (B) Optimal parameters for Lasso Regression Analysis. (C) Time-dependent ROC curves in the training set at 1-year, 2-year, 3-year and 5-year. (D) The survival status of RCC patients in the training group. Green dots represent the patient is still alive, while red dots represent the patient has dead. (E) Risk scores of RCC patients in the training group. Green dots represent the patient assigned to the low risk group, while red dots represent the patient assigned to the high risk group. (F) mRNAs expression levels of six mRNA LCC prognosis signature in the training group.

Full-size DOI: 10.7717/peerj.11433/fig-5

The prognosis signature confers additional prognostic power for LCC and RCC patients

Clinical characteristics, including the pStage ($P < 0.001$), pN ($P < 0.001$), pM ($P = 0.004$), and the risk score ($P < 0.001$), were closely associated with patient survival in LCC

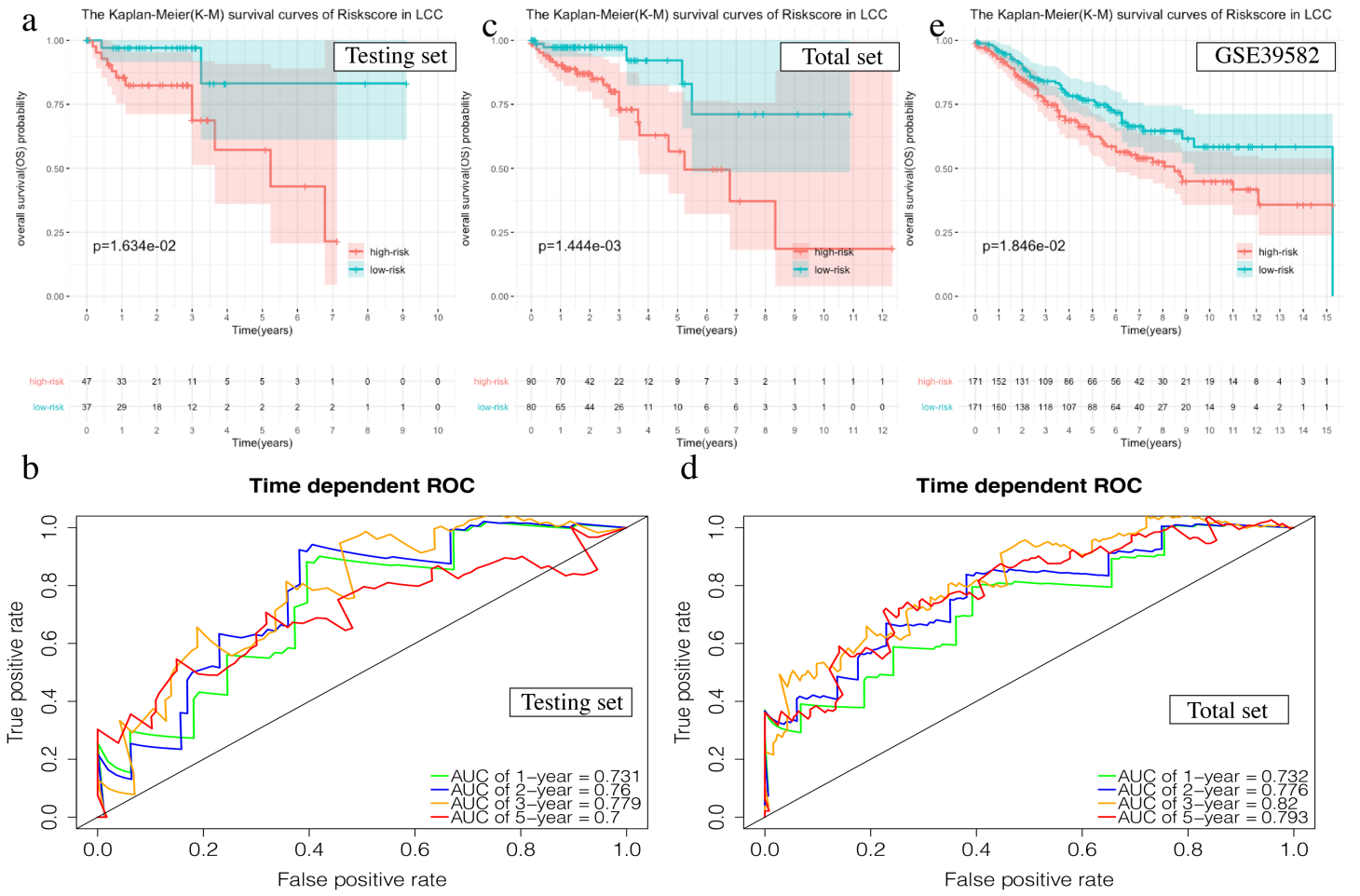

Figure 6 Validation of the prognostic signature of LCC. (A, C, E) The Kaplan–Meier (K–M) survival curves in the testing set, the total set and GSE39582. (B, D) Time-dependent ROC curves in the testing set and the total set at 1-year, 2-year, 3-year and 5-year.

Full-size DOI: 10.7717/peerj.11433/fig-6

(Fig. 8A). The pStage ($P < 0.001$), pT ($P < 0.001$), pN ($P < 0.001$), pM ($P < 0.001$), age ($P = 0.013$), and risk score were closely associated with patient survival in RCC (Fig. 8B). Multivariate Cox regression analysis further showed that the our signature is an independent prognostic indicator for OS in LCC and RCC (Figs. 8C–8D, Tables S5–S6).

Single gene mutation landscape in LCC and RCC

The most obvious mutations, including missense mutations, were: deletion, nonsense mutation, splice site, insertion, translation start site, and nonstop mutation. The missense mutation was the most obvious. We also found that single nucleotide polymorphisms (SNP) were more frequent than insertions or deletion and the most common single nucleotide variant (SNV) was C > T (*Li et al., 2011*). The number of altered bases in each sample was counted and the mutation types were plotted in a box plot. The 10 most prevalent mutated genes in LCC and RCC were shown with ranked percentages (Figs. S4 and S5). The Mann–Whitney test was used to compare the TMB of LCC and RCC, and the results showed that the RCC had a higher TMB (Fig. 9A). The mutation information

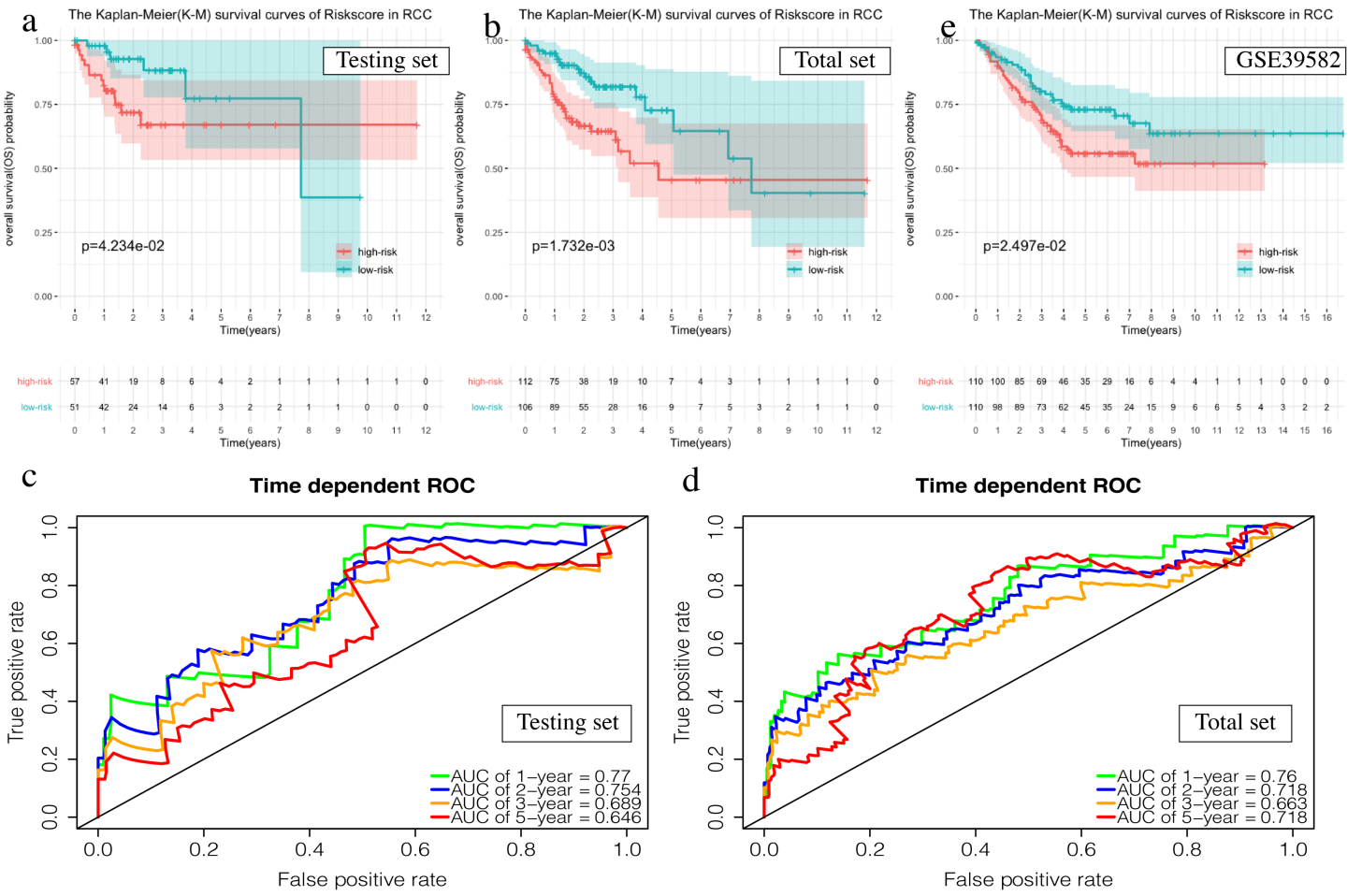

Figure 7 Validation of the prognostic signature of RCC. (A, C, E) The Kaplan–Meier (K-M) survival curves in the testing set, the total set and GSE39582. (B, D) Time-dependent ROC curves in the testing set and the total set at 1-year, 2-year, 3-year and 5-year.

Full-size DOI: 10.7717/peerj.11433/fig-7

of each sample in LCC and RCC was graphed in a waterfall plot (Figs. 9B–9C), which showed that the mutation frequency of RCC was generally higher than that of LCC. *APC*, *TP53*, *TTN*, and *KRAS* mutations were present in both LCC and RCC (*Cappell, 2008*). We found that *BRAF* mutations were more pronounced in RCC, and *APC* mutations were significantly higher in LCC. The higher immune infiltration and the higher *BRAF* mutation in RCC suggested that RCC is closely related to MSI. The study of Lochhead P et al. showed that BRAF mutations in colorectal cancer were linked to MSI through the methylation of CIMP and MLH1 promoter methylation (*Lochhead et al., 2013*). This is consistent with results from previous research (*Popescu et al., 2021*). The high *APC* mutation in LCC suggests that it may be related to the inactivation of the Wnt pathway (*Faux et al., 2021*). LINC02418 has been shown to be a tumor driver in colon cancer (*Tian et al., 2020*), and whether there is an inherent relationship between LINC02418 and the mutated gene has not been investigated. The association between the mutated gene and LINC02418 may become the direction of future research.

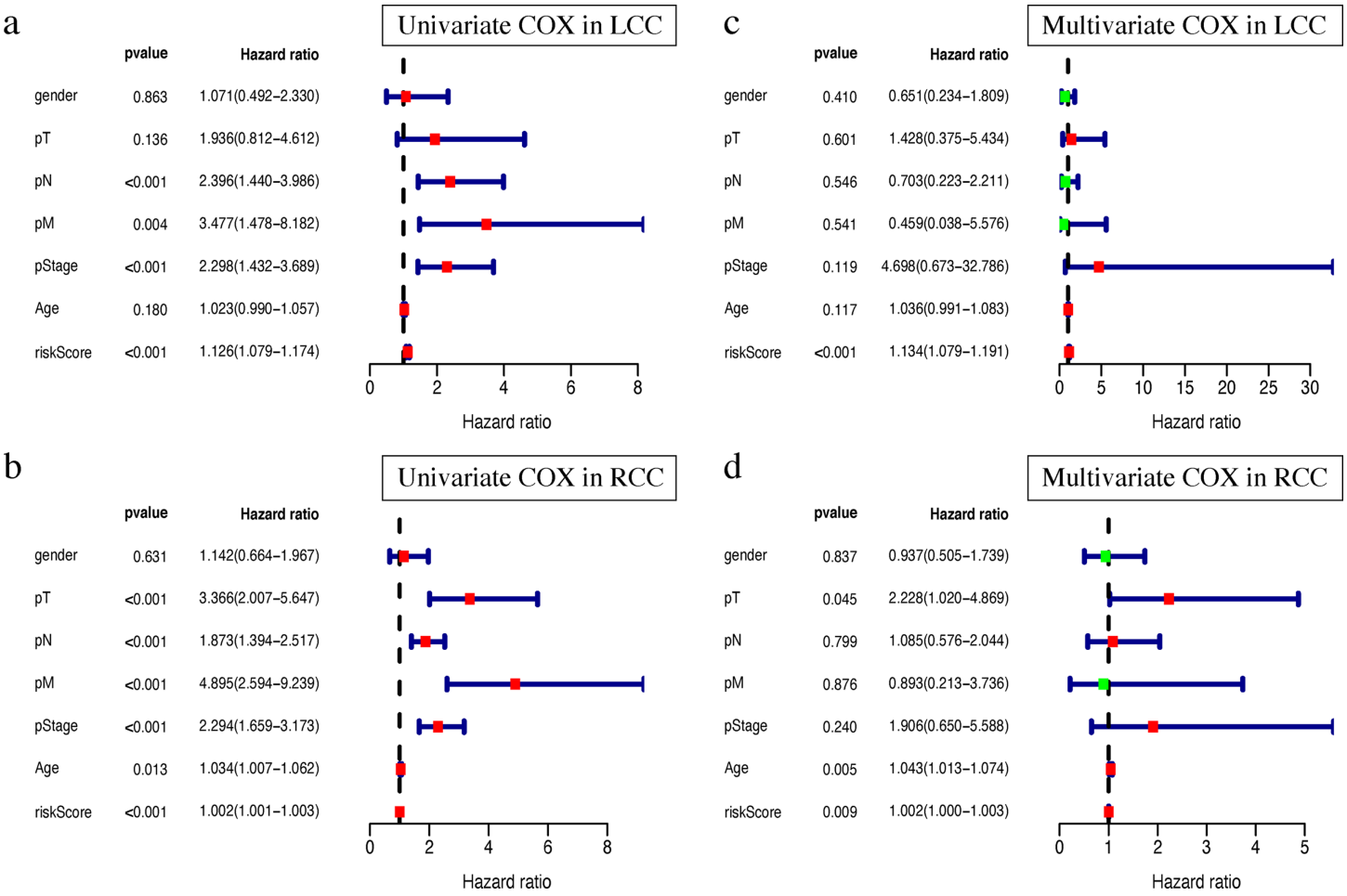

Figure 8 Independent prognostic analysis of two prognostic signatures. (A) Univariate COX analysis of LCC prognostic signatures and clinical characteristics. (B) Univariate COX analysis of RCC prognostic signatures and clinical characteristics. (C) Multivariate COX analysis of LCC prognostic signatures and clinical characteristics. (D) Multivariate COX analysis of RCC prognostic signatures and clinical characteristics.
Full-size DOI: 10.7717/peerj.11433/fig-8

The correlation analysis of top 25 mutated genes in LCC and RCC was conducted using the maftools package. Molecular interactions were more frequent in RCC than in LCC (Figs. 9D –9E). In RCC, the co-occurrence of *APC* and *KRAS* and the mutually exclusive relationship of *BRAF* with *APC* and *KRAS* further indicated a potential relationship with CIN and CIMP (*Issa, 2008*).The different molecular mechanisms of RCC and LCC suggest that they may require different therapeutic approaches and prognoses.

DISCUSSION

Colon cancer is one of the most common malignant tumors of the digestive system. Colon cancer can be defined as a left-sided or right-sided cancer according to the primary location of the tumor. The primary site of left-side colon cancer includes the splenic flexure, descending colon, and sigmoid colon. The right-side colon cancer includes the cecum, ascending colon, and hepatic flexure. The literature shows that prognosis of the left-side colon cancer is better than that of the right-side (*Klose et al., 2020*), and the

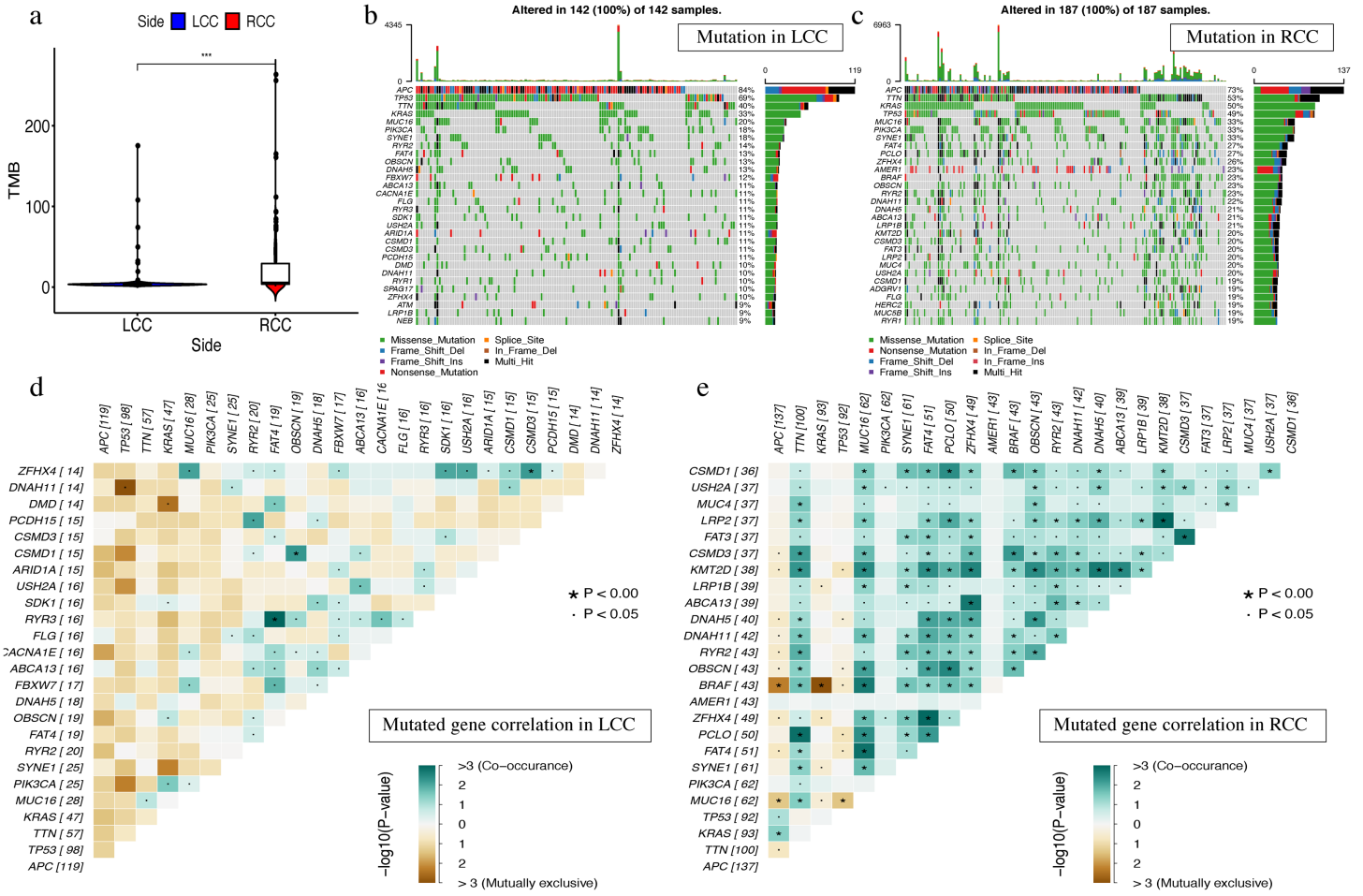

Figure 9 The landscape of single gene mutation in LCC and RCC. (A) The TMB of samples from two immune subgroups (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$) (B) Waterfall plot displayed the top 30 frequently mutated genes in LCC. (C) Waterfall plot displayed the top 30 frequently mutated genes in RCC. (D) The coincident and exclusive associations across mutated genes in LCC. (E) The coincident and exclusive associations across mutated genes in RCC.
Full-size DOI: 10.7717/peerj.11433/fig-9

survival rate is higher. Therefore, we need a better understanding of the classification of clinical subtypes of colon cancer.

We used the ssGSEA algorithm to obtain scores for 29 immune cell types and immune-related functions. We observed a high level of immune function in our analysis of RCC. This is manifested in *PD-L1* and *HLA* class I genes as well as high immune cell infiltration. Previous studies have shown that the changes of *HLA* class I genes and the high expression of *PD-L1* are both closely related to RCC (*Kanno et al., 2020*). We also observed that the immune score was significantly different between the two groups; the RCC presented with a higher immune score. This is consistent with previous reports (*Pentheroudakis et al., 2015*). The same is true for *HLA* family genes and immune checkpoint related gene expression and the abundance of immune cell infiltrations. Our results suggest that the RCC has more immune infiltration than the LCC, while RCC has a worse prognosis than LCC. TMB analysis showed that RCC with high immunity had a higher mutational burden than LCC, suggesting the existence of immune-evasive

mutations and immune escape. Our comprehensive analysis found that the level of immune cell infiltration should not be the only determinant of prognosis.

The genetic factors of colorectal cancer include chromosome instability (CIN) and microsatellite instability (MSI) (*Kaiser, Meckbach & Jacob, 2014*). DNA mismatch repair (MMR) gene mutations or modifications may lead to a lack of MMR proteins, referred to as "microsatellite instability" (MSI), which can detect the rise or decline in the number of repeat sequences in tumor tissues and is caused by a repetitive sequence of insertions or deletions in the DNA. We found that *APC* mutations were significantly present in LCC, and *BRAF* mutations were significantly present in RCC. *APC* is a multifunctional gene, whose mutation is often associated with chromosome instability (CIN) (*Hoevenaar et al., 2020*), and plays an important role in the regulation of the Wnt signaling pathway. *APC* regulates the Wnt pathway by controlling the formation of β-catenin/Tcf, a nuclear complex that initiates Wnt target gene transcription (*Raji, Sasikumar & Jacob, 2018*). Both CIMP and *BRAF* mutations are closely related to RCC. Moreover, CIMP is often associated with an increased risk of malignant transformation, and *BRAF* mutations are suggestive of MSI. A number of studies have suggested that the occurrence of RCC is closely related to MSI. However, that RCC tends to have a poor prognosis despite its high MSI is contrary to previous studies (*Laghi et al., 2020*) where a high MSI indicates a good prognosis. This may suggest that *BRAF* negatively impacts the occurrence of RCC. The high correlation between RCC and *BRAF* suggests that there may be other prognostic pathways in the occurrence of RCC that are worth exploring. Based on these results, it is reasonable to assume that the prognostic efficacy of MMR is weak. It is reasonable to believe that the combination of immunotherapy and the analysis of the related signaling pathways will have important significance in the future cancer therapy.

From the above results and the clinicopathological analysis of the right hemicolon (*Kalantzis et al., 2020*), LCC and RCC may be tumors of different properties and have different carcinogenic mechanisms. We constructed a 4-mRNA LCC prognostic signature and a 6-mRNA RCC prognostic signature. Among the key genes, *C1ORF105* was associated with a larger inter-adventitial common carotid artery diameter (ICCAD) (*Harrison et al., 2013*), and *FAM132b* can be increased by induction of erythrogenesis (*Gurieva et al., 2017*), suggesting that they may play an important role in the occurrence and development of cancer. The overexpression of *TNNT1* may play a role in the development of diffuse midline gliomas (DMGs) (*Vitanza et al., 2020*). *RspO4* can activate the Wnt/β-catenin signaling pathway and promote the progression of esophageal squamous cell carcinoma (*Chai et al., 2020*). Genetic variation in *KLRG2* may influence the aggressiveness of prostate cancer (*Liu et al., 2011*). *MiR-1254* targets *PAX5* to reduce *HIPPO* signal, thereby promoting the proliferation, migration, and invasion of HCC cells (*Lu et al., 2021*). *SLC22A31* was differentially expressed between LCC and RCC in a study based on sequencing data (*Liang et al., 2018*). Currently, there is no reported association between *OFCC1*, *Syngr3*, *CCDC160* and cancer. Additional studies on the mechanism of action of these key genes in LCC and RCC are needed. The model has been

verified by an internal testing set, a complete set, and an external testing and has been validated as an independent prognostic indicator. Prognostic signatures were established for left-sided and right-sided colon cancers, and have been validated internally and externally. These signatures provide the basis for individualized treatment of left and right-sided colon cancers.

RCC has a more pronounced mutation landscape than LCC according to previous studies (*Jensen, Villanueva & Loaiza-Bonilla, 2018*). We found that the expression of mutated genes in LCC was more positively correlated, and the results were more significant than that of RCC. These results suggest that the classification of clinical subtypes of colon cancer may be of great significance for the determination of clinical diagnosis and treatment in the future.

CONCLUSIONS

We observed significant differences in the clinical characteristics, immune microenvironment, transcriptomic differences, and single gene mutation differences in the multi-omics analysis of LCC and RCC, suggesting that the difference in gene expression can be analyzed and divided into different clinical subtypes to help the early clinical diagnosis and prognosis of colon cancer. Our results may provide individualized treatment options and better prognostic evaluation for patients with left-side or right-side colon cancer. The 4-mRNA LCC prognostic signature and 6-mRNA RCC prognostic signature may provide a basis for personalized treatment of colon cancer. Further clinical testing is required to validate our results.

ABBREVIATIONS

LCC	left-side colon cancer
RCC	right-side colon cancer
TCGA	The Cancer Genome Atlas
ssGSEA	single sample gene set enrichment analysis
HLA	human leukocyte antigen
DAVID	the Database for Annotation, Visualization and Integrated Discovery
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
K–M	Kaplan–Meier
ROC	receiver operating characteristic
GEO	the Gene Expression Omnibus
OS	Overall survival
SNP	single nucleotide poly-morphism
SNV	single nucleotide variants

ACKNOWLEDGEMENTS

The authors thank all the researchers who supported the The Cancer Genome Atlas and the Gene Expression Omnibus Research Network.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by a grant from National Natural Science Foundation of China (nfsc:81560397 and 81660403). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China (nfsc): 81560397 and 81660403.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yanyi Huang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jinzhong Duanmu conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Yushu Liu conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mengyun Yan conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Taiyuan Li conceived and designed the experiments, authored or reviewed drafts of the paper, statistical analyses, and approved the final draft.
- Qunguang Jiang conceived and designed the experiments, authored or reviewed drafts of the paper, statistical analyses, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
The raw data is available at TCGA-COAD and GSE39582.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11433#supplemental-information.

REFERENCES

Altermann E, Klaenhammer TR. 2005. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* **6(1)**:60 DOI 10.1186/1471-2164-6-60.

Aran D, Lasry A, Zinger A, Biton M, Pikarsky E, Hellman A, Butte AJ, Ben-Neriah Y. 2016. Widespread parainflammation in human cancer. *Genome Biology* **17(1)**:145 DOI 10.1186/s13059-016-0995-z.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Research* **41(Database)**:D991–D995 DOI 10.1093/nar/gks1193.

Bedognetti D, Hendrickx W, Marincola FM, Miller LD. 2015. Prognostic and predictive immune gene signatures in breast cancer. *Current Opinion in Oncology* **27(6)**:433–444 DOI 10.1097/CCO.0000000000000234.

Bufill JA. 1990. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Annals of Internal Medicine* **113(10)**:779–788 DOI 10.7326/0003-4819-113-10-779.

Cappell MS. 2008. Pathophysiology, clinical presentation, and management of colon cancer. *Gastroenterology Clinics of North America* **37(1)**:1–24 DOI 10.1016/j.gtc.2007.12.002.

Chai T, Shen Z, Zhang Z, Chen S, Gao L, Zhang P, Lin W, Kang M, Lin J. 2020. LGR6 is a potential diagnostic and prognostic marker for esophageal squamous cell carcinoma. *Journal of Clinical Laboratory Analysis* **34(4)**:e23121 DOI 10.1002/jcla.23121.

Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. 2018. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology* **1711**:243–259 DOI 10.1007/978-1-4939-7493-1_12.

Chen L, Zhang YH, Wang S, Zhang Y, Huang T, Cai YD. 2017. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLOS ONE* **12(9)**:e0184129 DOI 10.1371/journal.pone.0184129.

Cho YA, Lee J, Oh JH, Chang HJ, Sohn DK, Shin A, Kim J. 2019. Combined lifestyle factors and risk of colorectal cancer. *Cancer Research and Treatment* **51(3)**:1033–1040 DOI 10.4143/crt.2018.447.

Faux MC, King LE, Kane SR, Love C, Sieber OM, Burgess AW. 2021. APC regulation of ESRP1 and p120-catenin isoforms in colorectal cancer cells. *Molecular Biology of The Cell* **32(2)**:120–130 DOI 10.1091/mbc.E20-05-0321.

Galili T, O'Callaghan A, Sidi J, Sievert C. 2018. Heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34(9)**:1600–1602 DOI 10.1093/bioinformatics/btx657.

Grant SW, Hickey GL, Head SJ. 2019. Statistical primer: multivariable regression considerations and pitfalls. *European Journal of Cardio-Thoracic Surgery* **55(2)**:179–185 DOI 10.1093/ejcts/ezy403.

Gurieva I, Frýdlová J, Rychtarčíková Z, Vokurka M, Truksa J, Krijt J. 2017. Erythropoietin administration increases splenic erythroferrone protein content and liver TMPRSS6 protein content in rats. *Blood Cells, Molecules and Diseases* **64**:1–7 DOI 10.1016/j.bcmd.2017.02.007.

Harrison SC, Zabaneh D, Asselbergs FW, Drenos F, Jones GT, Shah S, Gertow K, Sennblad B, Strawbridge RJ, Gigante B, Holewijn S, De Graaf J, Vermeulen S, Folkersen L, van Rij AM, Baldassarre D, Veglia F, Talmud PJ, Deanfield JE, Agu O, Kivimaki M, Kumari M, Bown MJ, Nyyssönen K, Rauramaa R, Smit AJ, Franco-Cereceda A, Giral P, Mannarino E, Silveira A, Syvänen AC, de Borst GJ, van der Graaf Y, de Faire U, Baas AF, Blankensteijn JD, Wareham NJ, Fowkes G, Tzoulaki I, Price JF, Tremoli E, Hingorani AD, Eriksson P, Hamsten A, Humphries SE. 2013. A gene-centric study of common carotid artery remodelling. *Atherosclerosis* **226(2)**:440–446 DOI 10.1016/j.atherosclerosis.2012.11.002.

Hoevenaar WHM, Janssen A, Quirindongo AI, Ma H, Klaasen SJ, Teixeira A, van Gerwen B, Lansu N, Morsink FHM, Offerhaus GJA, Medema RH, Kops GJPL, Jelluma N. 2020. Degree

and site of chromosomal instability define its oncogenic potential. *Nature Communications* **11(1)**:643 DOI 10.1038/s41467-020-15279-9.

Hsu YL, Lin CC, Jiang JK, Lin HH, Lan YT, Wang HS, Yang SH, Chen WS, Lin TC, Lin JK, Lin PC, Chang SC. 2019. Clinicopathological and molecular differences in colorectal cancer according to location. *The International Journal of Biological Markers* **34(1)**:47–53 DOI 10.1177/1724600818807164.

Hänzelmann S, Castelo R, Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**:7 DOI 10.1186/1471-2105-14-7.

Imperial R, Ahmed Z, Toor OM, Erdoğan C, Khaliq A, Case P, Case J, Kennedy K, Cummings LS, Melton N, Raza S, Diri B, Mohammad R, El-Rayes B, Pluard T, Hussain A, Subramanian J, Masood A. 2018. Comparative proteogenomic analysis of right-sided colon cancer, left-sided colon cancer and rectal cancer reveals distinct mutational profiles. *Molecular Cancer* **17(1)**:177 DOI 10.1186/s12943-018-0923-9.

Issa JP. 2008. Colon cancer: it's CIN or CIMP. *Clinical Cancer Research* **14(19)**:5939–5940 DOI 10.1158/1078-0432.CCR-08-1596.

Jensen CE, Villanueva JY, Loaiza-Bonilla A. 2018. Differences in overall survival and mutation prevalence between right- and left-sided colorectal adenocarcinoma. *Journal of Gastrointestinal Oncology* **9(5)**:778–784 DOI 10.21037/jgo.2018.06.10.

Kaiser JC, Meckbach R, Jacob P. 2014. Genomic instability and radiation risk in molecular pathways to colon cancer. *PLOS ONE* **9(10)**:e111024 DOI 10.1371/journal.pone.0111024.

Kalantzis I, Nonni A, Pavlakis K, Delicha EM, Miltiadou K, Kosmas C, Ziras N, Gkoumas K, Gakiopoulou H. 2020. Clinicopathological differences and correlations between right and left colon cancer. *World Journal of Clinical Cases* **8(8)**:1424–1443 DOI 10.12998/wjcc.v8.i8.1424.

Kalantzis I, Nonni A, Pavlakis K, Delicha EM, Miltiadou K, Kosmas C, Ziras N, Gkoumas K, Gakiopoulou H. 2020. Clinicopathological differences and correlations between right and left colon cancer. *World Journal of Clinical Cases* **8(8)**:1424–1443 DOI 10.12998/wjcc.v8.i8.1424.

Kanno H, Miyoshi H, Yoshida N, Sudo T, Nakashima K, Takeuchi M, Nomura Y, Seto M, Hisaka T, Tanaka H, Okuda K, Akagi Y, Ohshima K. 2020. Differences in the immunosurveillance pattern associated with DNA mismatch repair status between right-sided and left-sided colorectal cancer. *Cancer Science* **111(8)**:3032–3044 DOI 10.1111/cas.14495.

Kikuchi T, Mimura K, Okayama H, Nakayama Y, Saito K, Yamada L, Endo E, Sakamoto W, Fujita S, Endo H, Saito M, Momma T, Saze Z, Ohki S, Kono K. 2019. A subset of patients with MSS/MSI-low-colorectal cancer showed increased CD8(+) TILs together with up-regulated IFN-γ. *Oncology Letters* **18(6)**:5977–5985 DOI 10.3892/ol.2019.10953.

Klose J, Kloor M, Warschkow R, Antony P, Liesenfeld LF, Büchler MW, Schneider M, Tarantino I. 2020. Does side really matter? Survival analysis among patients with right- versus left-sided colon cancer: a propensity score-adjusted analysis. *Annals of Surgical Oncology* **28(5)**:2768–2778 DOI 10.1245/s10434-020-09116-y.

Laghi L, Negri F, Gaiani F, Cavalleri T, Grizzi F, De' Angelis GL, Malesci A. 2020. Prognostic and predictive cross-roads of microsatellite instability and immune response to colon cancer. *International Journal of Molecular Sciences* **21(24)**:9680 DOI 10.3390/ijms21249680.

Li Y, Tang J, Pan Z, Xiao P, Zhou D, Jin L, Pan M, Lu Z. 2011. Single nucleotide polymorphism genotyping and point mutation detection by ligation on microarrays. *Journal of Nanoscience and Nanotechnology* **11(2)**:994–1003 DOI 10.1166/jnn.2011.3056.

Liang L, Zeng JH, Qin XG, Chen JQ, Luo DZ, Chen G. 2018. Distinguishable Prognostic signatures of left- and right-sided colon cancer: a study based on sequencing data. *Cellular Physiology and Biochemistry* **48(2)**:475–490 DOI 10.1159/000491778.

Liu X, Cheng I, Plummer SJ, Suarez BK, Casey G, Catalona WJ, Witte JS. 2011. Fine-mapping of prostate cancer aggressiveness loci on chromosome 7q22-35. *Prostate* **71(7)**:682–689 DOI 10.1002/pros.21284.

Liu Z, Li M, Jiang Z, Wang X. 2018. A comprehensive immunologic portrait of triple-negative breast cancer. *Translational Oncology* **11(2)**:311–329 DOI 10.1016/j.tranon.2018.01.011.

Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, Qian ZR, Morikawa T, Shen J, Meyerhardt JA, Fuchs CS, Ogino S. 2013. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *JNCI: Journal of the National Cancer Institute* **105(15)**:1151–1156 DOI 10.1093/jnci/djt173.

Lu X, Yang C, Hu Y, Xu J, Shi C, Rao J, Yu W, Cheng F. 2021. Upregulation of miR-1254 promotes hepatocellular carcinoma cell proliferation, migration, and invasion via inactivation of the Hippo-YAP signaling pathway by decreasing PAX5. *Journal of Cancer* **12(3)**:771–789 DOI 10.7150/jca.49680.

Maibach F, Sadozai H, Seyed Jafari SM, Hunger RE, Schenk M. 2020. Tumor-infiltrating lymphocytes and their prognostic value in cutaneous melanoma. *Frontiers in Immunology* **11**:2105 DOI 10.3389/fimmu.2020.02105.

Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou JF, Benchimol D, Berger A, Lagarde A, Pencreach E, Piard F, Elias D, Parc Y, Olschwang S, Milano G, Laurent-Puig P, Boige V. 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLOS Medicine* **10(5)**:e1001453 DOI 10.1371/journal.pmed.1001453.

Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Research* **28(11)**:1747–1756 DOI 10.1101/gr.239244.118.

McGee M. 2018. Case for omitting tied observations in the two-sample t-test and the Wilcoxon-Mann-Whitney Test. *PLOS ONE* **13(7)**:e0200837 DOI 10.1371/journal.pone.0200837.

Mik M, Berut M, Dziki L, Trzcinski R, Dziki A. 2017. Right- and left-sided colon cancer: clinical and pathological differences of the disease entity in one organ. *Archives of Medical Science* **13(1)**:157–162 DOI 10.5114/aoms.2016.58596.

Nakagawa-Senda H, Hori M, Matsuda T, Ito H. 2019. Prognostic impact of tumor location in colon cancer: the Monitoring of Cancer Incidence in Japan (MCIJ) project. *BMC Cancer* **19(1)**:431 DOI 10.1186/s12885-019-5644-y.

Obuchowski NA, Bullen JA. 2018. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Physics in Medicine and Biology* **63(7)**:07TR01 DOI 10.1088/1361-6560/aab4b1.

Pentheroudakis G, Raptou G, Kotoula V, Wirtz RM, Vrettou E, Karavasilis V, Gourgioti G, Gakou C, Syrigos KN, Bournakis E, Rallis G, Varthalitis I, Galani E, Lazaridis G, Papaxoinis G, Pectasides D, Aravantinos G, Makatsoris T, Kalogeras KT, Fountzilas G. 2015. Immune response gene expression in colorectal cancer carries distinct prognostic implications according to tissue, stage and site: a prospective retrospective translational study in the context of a hellenic cooperative oncology group randomised trial. *PLOS ONE* **10(5)**:e0124612 DOI 10.1371/journal.pone.0124612.

Popescu RC, Tocia C, Brînzan C, Cozaru GC, Deacu M, Dumitru A, Leopa N, Mitroi AF, Nicolau A, Dumitru E. 2021. Molecular profiling of the colon cancer in South-Eastern Romania: results from the MERCUR study. *Medicine (Baltimore)* **100(1)**:e24062 DOI 10.1097/MD.0000000000024062.

Raji RJ, Sasikumar R, Jacob E. 2018. Multiple roles of Adenomatous Polyposis Coli gene in Wnt signalling: a computational model. *Biosystems* **172(13)**:26–36 DOI 10.1016/j.biosystems.2018.08.001.

Ren Y, Cherukuri Y, Wickland DP, Sarangi V, Tian S, Carter JM, Mansfield AS, Block MS, Sherman ME, Knutson KL, Lin Y, Asmann YW. 2020. HLA class-I and class-II restricted neoantigen loads predict overall survival in breast cancer. *Oncoimmunology* **9(1)**:1744947 DOI 10.1080/2162402X.2020.1744947.

Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A. 2020. Colorectal cancer statistics. *CA: A Cancer Journal for Clinicians* **70(3)**:145–164 DOI 10.3322/caac.21601.

Tabibzadeh A, Tameshkel FS, Moradi Y, Soltani S, Moradi-Lakeh M, Ashrafi GH, Motamed N, Zamani F, Motevalian SA, Panahi M, Esghaei M, Ajdarkosh H, Mousavi-Jarrahi A, Niya MHK. 2020. Signal transduction pathway mutations in gastrointestinal (GI) cancers: a systematic review and meta-analysis. *Scientific Reports* **10(1)**:18713 DOI 10.1038/s41598-020-73770-1.

Tian J, Cui P, Li Y, Yao X, Wu X, Wang Z, Li C. 2020. LINC02418 promotes colon cancer progression by suppressing apoptosis via interaction with miR-34b-5p/BCL2 axis. *Cancer Cell International* **20(1)**:460 DOI 10.1186/s12935-020-01530-2.

Tibshirani RJ. 2009. Univariate shrinkage in the cox model for high dimensional data. *Statistical Applications in Genetics and Molecular Biology* **8(1)**:1–18 DOI 10.2202/1544-6115.1438.

Vitanza NA, Biery MC, Myers C, Ferguson E, Zheng Y, Girard EJ, Przystal JM, Park G, Noll A, Pakiam F, Winter CA, Morris SM, Sarthy J, Cole BL, Leary SES, Crane C, Lieberman NAP, Mueller S, Nazarian J, Gottardo R, Brusniak MY, Mhyre AJ, Olson JM. 2020. Optimal therapeutic targeting by HDAC inhibition in biopsy-derived treatment-naïve diffuse midline glioma models. *Neuro Oncology* **23(3)**:376–386 DOI 10.1093/neuonc/noaa249.

Weiss JM, Pfau PR, O'Connor ES, King J, LoConte N, Kennedy G, Smith MA. 2011. Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results—Medicare data. *Journal of Clinical Oncology* **29(33)**:4401–4409 DOI 10.1200/JCO.2011.36.4414.

Yang Q, Wang S, Dai E, Zhou S, Liu D, Liu H, Meng Q, Jiang B, Jiang W. 2019. Pathway enrichment analysis approach based on topological structure and updated annotation of pathway. *Briefings in Bioinformatics* **20(1)**:168–177 DOI 10.1093/bib/bbx091.

Zhang Z, He G, Feng Q, Zheng P, Lv Y, Mao Y, Xu Y, Xu J. 2020. Analysis of tumor microenvironment-related key mRNAs and construction of a prognosis signature in colon cancer. *Clinical and Translational Medicine* **10(2)**:e104 DOI 10.1002/ctm2.104.

Zhang CD, Wang JN, Sui BQ, Zeng YJ, Chen JQ, Dai DQ. 2016. Prognostic and predictive model for stage II colon cancer patients with nonemergent surgery: who should receive adjuvant chemotherapy? *Medicine* **95(1)**:e2190 DOI 10.1097/MD.0000000000002190.

Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, Bin J, Liao Y, Liao W. 2019. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I–III colon cancer. *Cancer Immunology, Immunotherapy* **68(3)**:433–442 DOI 10.1007/s00262-018-2289-7.