Original Article

Six mutator-derived lncRNA signature of genome instability for predicting the clinical outcome of colon cancer

Shujia Chen[1#], Xiaofei Li[2#], Jiachen Zhang[1], Li Li[1], Xueqiu Wang[1], Yinghui Zhu[2], Lianyi Guo[2], Jiwei Wang[3]

[1]Department of Gastroenterology, Panjin Central Hospital Affiliated to Jinzhou Medical University, Panjin, China; [2]Department of Gastroenterology, the First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China; [3]Department of Gastrointestinal Surgery, Xuzhou Central Hospital, Xuzhou, China

Contributions: (I) Conception and design: All authors; (II) Administrative support: J Wang; (III) Provision of study materials or patients: L Guo; (IV) Collection and assembly of data: S Chen; (V) Data analysis and interpretation: J Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jiwei Wang. Department of Gastrointestinal Surgery, Xuzhou Central Hospital, Xuzhou 221009, China. Email: 17310923623m@sina.cn; Lianyi Guo. Department of Gastroenterology, the First Affiliated Hospital of Jinzhou Medical University, Jinzhou 121001, China. Email: angel_gly@163.com.

Background: Colon adenocarcinoma (COAD) is one of the most common malignancies worldwide. Genomic instability is one of the hallmarks of colon cancer and is associated with prognosis. Nevertheless, the impact of genome instability-associated long non-coding RNAs (lncRNAs) along with their clinical significance in cancers has remained mostly unexplored.

Methods: In this study, a mutator hypothesis-derived computational frame integrating the somatic mutation profiles and lncRNA expression profiles in a tumor genome was developed, which enabled the identification of 137 novel genomic instability-associated lncRNAs in colon cancer. Subsequently, a genome instability-derived lncRNA signature (GILncSig) segregated the patients into low- and high-risk groups with prominent differences in outcomes.

Results: Combined with the overall survival data, we established 6 six lncRNA-based signature to predict prognosis, which were LINC00896, AC007996.1, NKILA, AP003555.2, MIRLET7BHG, and AC009237.14. We found that the expression level of PD-L1 (CD274) and somatic mutations in the high-risk group were higher than those in the low-risk group. This suggests that high-risk patients may be sensitive to immunotherapy. We further found that the prognosis of patients in the high-risk group was significantly lower than that of patients in the low-risk group, and that patients' prognosis was likely to be worse as the patient's risk score increased.

Conclusions: In conclusion, this study explores the role of lncRNAs in genomic instability and cancer prognosis and provides a new idea for the prognostic prediction of colon cancer.

Keywords: Genome instability; mutator phenotype; long non-coding RNAs (lncRNAs)

Submitted Jul 23, 2021. Accepted for publication Sep 07, 2021.
doi: 10.21037/jgo-21-494
View this article at: https://dx.doi.org/10.21037/jgo-21-494

Introduction

Worldwide, despite advances in the diagnosis and treatment of colon adenocarcinoma (COAD), its morbidity and mortality rates remain high (1). The TNM staging system remains the world's gold standard for selecting cancer treatment or predicting prognosis. Nonetheless, despite identical clinical characteristics, the prognosis amongst patients can differ significantly due to the high levels of heterogeneity in colon cancer (2). From the progress, development, and response to treatment, colon cancer has shown high complexity in terms of clinical and molecular heterogeneity (3). Hence, to assess the clinical outcomes of patients with colon cancer more accurately, the urgent identification of novel biomarkers is imperative.

One of the hallmarks of cancer has been reported to be genomic instability. There are many forms of genomic instability. Most cancers have a form called chromosomal instability (CIN), which refers to the high rate of change in the structure and number of chromosomes in cancer cells over time compared to normal cells. Other forms of genomic instability have also been described, including microsatellite instability (MSI; also known as MIN), and forms of genomic instability characterized by an increased frequency of base pair mutations (4). Genomic instability has been linked with the progression and survival of colon cancer, and is an important prognostic factor (5). In colorectal cancer (CRC), the cancer was induced by mutated genes The evolution of CRC is caused by a variety of genetic and epigenetic abnormalities, including defective DNA mismatch repair (DMMR) and mutations in Kirsten-RAS (KRAS) and BRAF proto-oncogenes (6,7). This suggests the potential of molecular signatures of genomic instability to predict prognosis. For example, a 12-gene genomic instability signature was identified through an analysis of the gene expression profiles of breast cancer specimens by Habermann *et al.* (8). A meta-analysis of the expression of miRNAs showed their association with MSI in colon cancer tissues (9). Many studies have also shown that upregulated immune checkpoints (e.g., CTLA 4, PD-1, and/or PD-L1/CD274) have been found in highly mutated tumors with DMMR or high MSI (MSI-H), which may benefit from immunotherapy (10,11).

Transcripts that do not indicate any potential for protein coding and are longer than 200 nt have been broadly classified as long non-coding RNAs (lncRNAs) (12). Growing *in vivo* and *in vitro* evidence over the years have indicated the significant role played by lncRNAs across various biological processes. In particular, aberrant lncRNA expression has been shown to play a role in metastasis, tumor progression, and cell proliferation (13). Despite lncRNAs being abnormally expressed in numerous cancers, the functions of these lncRNAs have not been determined. In the maintenance of genomic instability, the crucial role of lncRNAs has been revealed from emerging evidence. A study showed that the non-coding RNA NORAD regulates genomic stability by chelating pumilio proteins (14). Another study showed that RNA exosomes control super-enhancer activity by regulating lncRNA transcription (15). In colon cancer, it has been shown that lncRNA CCAT2 induces CIN through BOP1-AURKb signaling, leading to poor prognosis (16). However, so far, lncRNAs associated with genomic instability and their clinical significance have not been mapped and explored.

To evaluate the potential of the lncRNA signature being considered as an indicator of genomic stability in colon cancer, a mutator hypothesis-derived computational frame integrating lncRNA expression and somatic mutation profiles in a tumor genome was developed in this study. This study explored the role of lncRNAs in genomic instability and cancer prognosis in colon cancer which was the main difference from the article you mentioned. We further explored the expression of immune checkpoint between high-risk group and low-risk group in this article. We used the CIBERSORT algorithm to estimate the proportions of 22 immune cell types in colon cancer samples to further investigate the relationship between the genome instability-derived lncRNA signature (GILncSig) and immune cell infiltration.

We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/jgo-21-494).

Methods

Data download

The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) provided the information of the somatic mutations, the fragments per kilobase of exon model per million mapped fragments (FPKM) type of RNA-seq expression data, and the clinical features of patients with COAD. For further study, a total of 446 paired samples with common clinicopathological features, somatic mutation information, survival information, and mRNA and lncRNA expression profiles were obtained.

Table 1 Clinicopathological information of the three COAD patient cohorts in this study

Covariates	Type	Total	Testing	Training	P value
Age	≤65	183	93	90	0.786
	>65	263	129	134	
Gender	Female	212	104	108	0.8459
	Male	234	118	116	
Stage	Stage I–II	250	126	124	0.7071
	Stage III–IV	185	89	96	
	Unknown	11	7	4	
T	T1–2	86	47	39	0.3878
	T3–4	359	175	184	
	Unknown	1	0	1	
M	M0	329	161	168	1
	M1	61	30	31	
	Unknown	56	31	25	
N	N0	265	134	131	0.7585
	N1–2	181	88	93	

COAD, colon adenocarcinoma.

All the colon cancer patients were divided into two sets, namely the test and training sets. To identify the prognostic lncRNA signature and build a prognostic risk model, a total of 224 patients from TCGA were placed in the training set, while 222 patients were used to validate the performance of the prognostic risk model. Furthermore, from TCGA database, the corresponding lncRNA expression data and the somatic mutation information of patients were also downloaded. *Table 1* provides a brief summary of the pathological and clinical characteristics. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of genome instability-associated lncRNAs

We developed a mutator hypothesis-derived computational framework by integrating somatic mutation profiles and lncRNA expression profiles in tumor genomes to identify lncRNAs associated with genomic instability using the following steps: (I) the cumulative number of somatic mutations in each patient was calculated; (II) the cumulative number of somatic mutations among patients was ranked in decreasing order; (III) the bottom 25% of patients were defined as the genomically stable (GS)-like group, while the top 25% were defined as the genomically unstable (GU)-like group; (IV) the expression profiles of the lncRNAs between the GS and GU groups were compared; (V) the genome instability-associated lncRNAs were identified as differentially expressed lncRNAs [absolute value of the fold change greater than 2 and their false discovery rate (FDR) adjusted P value being less than 0.05].

Enrichment analysis

We identified mRNAs pairwise expressed with lncRNAs associated with genomic instability. The top 10 mRNAs associated with each lncRNA were selected. On this basis, a co-expression network was constructed. The 'clusterProfiler' R package and the 'ggplot2' R package were used for Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (17,18).

Estimation of immune cell infiltration

We used the CIBERSORT algorithm to estimate the proportions of 22 immune cell types in colon cancer samples from gene expression data to further investigate

the relationship between the GILncSig and immune cell infiltration (19). We removed samples with P>0.05 and retained samples with P<0.05 for further analysis. We used the Wilcoxon rank-sum test to identify if there was a significant difference in the proportion of immune cells between the low- and high-risk groups. In addition, the Kaplan-Meier method was used to evaluate the relationship between immune cell infiltration and patient prognosis.

Statistical analysis

To evaluate the overall survival and the association with the expression levels of genome instability-associated lncRNAs, multivariate and univariate Cox proportional hazard regression analyses were conducted. A GILncSig for prognostic prediction was developed as described below, based on the coefficients from the multivariate regression analysis and the levels of expression of the prognostic genome instability-associated lncRNAs:

$$GILncSig(patient) = \sum_{i=1}^{n} coef(IncRNAi) \times expr(IncRNAi) \quad [1]$$

where GILncSig (patient) represented the prognostic risk score for the colon cancer patient, lncRNAi indicated the ith prognostic lncRNA, and expr (lncRNAi) was the expression level of lncRNAi for the patient. The coef (lncRNAi) was representative of the contribution of the lncRNAi to the prognostic risk scores that were obtained from the multivariate Cox analysis regression coefficient. To segregate patients into the low-risk group with a low GILncSig or the high-risk group with a high GILncSig, the median scores of patients in the training set were used as the risk cutoff.

To assess the difference in survival between the low- and high-risk groups with a 5% significance level, the log-rank test was applied. To determine the survival rate, the Kaplan-Meier method was applied. For assessing the independence of the GILncSig from the other key clinical factors, stratified and multivariate Cox regression analyses were conducted. With the Cox analysis, the 95% confidence interval (CI) and hazard ratio (HR) were calculated. The time-dependent receiver operating characteristic (ROC) curve was used to evaluate the performance of the GILncSig. R version 4.0.3 was used to perform all statistical analyses.

Results

Identification of genomic instability-related lncRNAs in colon cancer patients

We calculated and sorted the cumulative number of somatic mutations per patient to identify the lncRNAs associated with genomic instability. Based on the results of somatic mutations in decreasing order, the top 25% (n=112) of patients were divided into the GU-like group and the bottom 25% (n=101) of patients were divided into the GS-like group. In order to identify the lncRNAs with significant differences, the lncRNA expression profiles of the 112 patients from the GU-like group were compared with the 101 patients in the GS-like group. A total of 137 lncRNAs were found to be significantly differentially expressed, with an absolute value of fold change greater than 2 and FDR-adjusted P value less than 0.05. In total, 81 lncRNAs were upregulated and 56 downregulated, as detailed in Table S1. *Figure 1A* shows the top 40 lncRNAs that were significantly different. Using the set of 137 differentially expressed lncRNAs, the samples from TCGA cohort were divided into GS-like and GU-like groups based on the expression levels of the differentially expressed lncRNAs (*Figure 1B*). Between the two groups, the somatic mutation pattern was found to be significantly different. The GU-like group contained higher cumulative somatic mutations compared to the GS-like group (P=0.014; *Figure 1C*). Further study showed higher PD-L1 (CD274) expression in the GU group, consistent with another previous study ($P=1.4\times10^{-8}$; *Figure 1D*) (11).

Enrichment analysis

We measured the expression correlation between the 137 differentially expressed lncRNAs and protein coding genes, and constructed a lncRNA-mRNA co-expression network (*Figure 2A*). The mRNAs co-expressed with lncRNAs were enriched by GO and KEGG analysis. Based on GO enrichment analysis, the mRNAs in this co-expression network were mainly enriched in gland development, organelle subcompartment, and steroid binding (*Figure 2B*). According to KEGG pathway enrichment analysis, the main enrichment pathways were the chemokine signaling pathway, proteoglycans in cancer, and the NOD-like receptor signaling pathway (*Figure 2C*). These pathways are

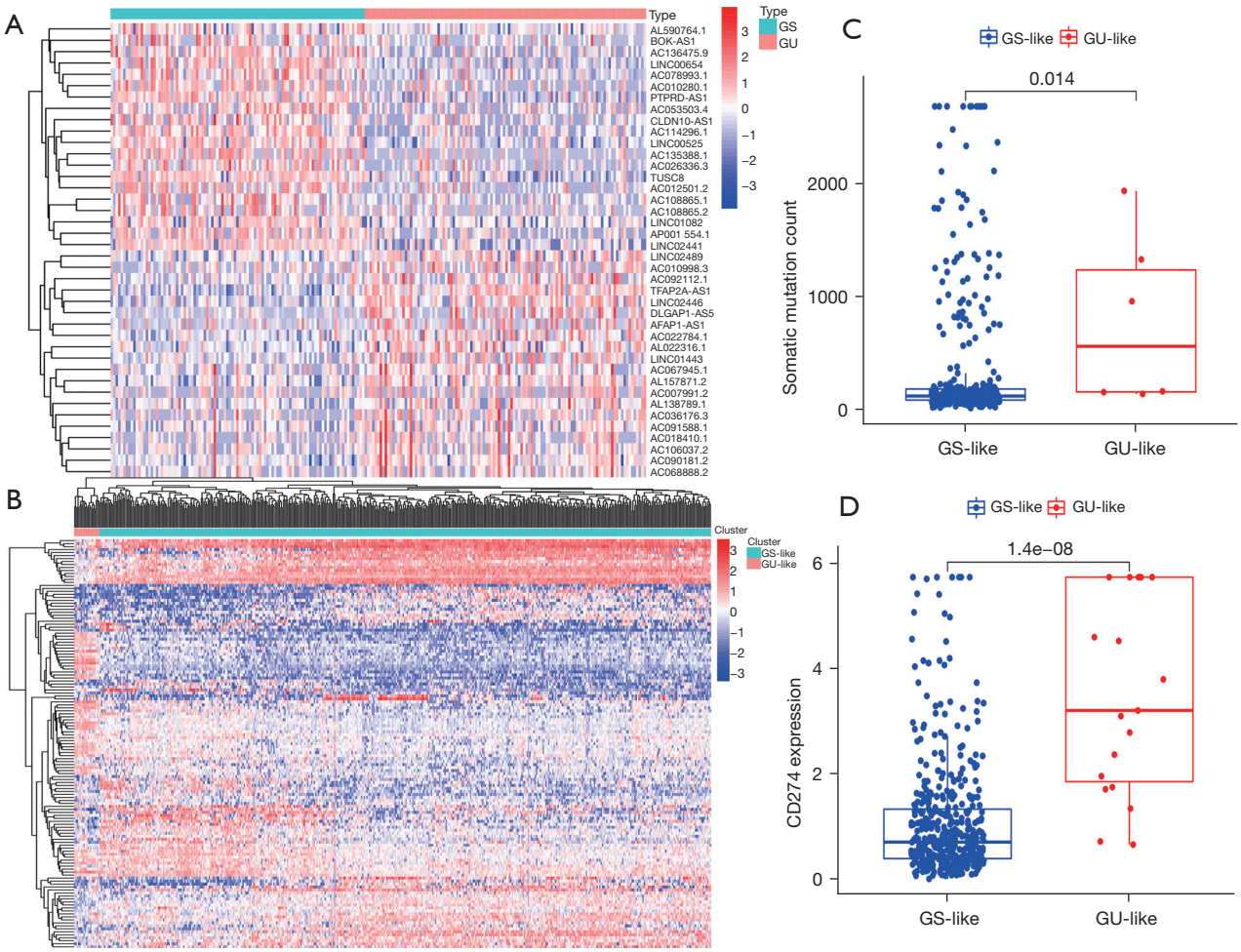

Figure 1 Identification of genomic instability-related lncRNAs in patients with colon cancer. (A) Heatmap showing significant changes in the top 40 lncRNAs between the GU-like group and the GS-like group. (B) Based on the expression patterns of 137 candidate lncRNAs associated with genomic instability, 446 colon cancer patients were clustered. (C) Boxplots of somatic mutations in the GU-like group and GS-like group. (D) Boxplots of CD274 expression levels in the GU-like group and GS-like group. LncRNAs, long non-coding RNAs; GU, genomically unstable; GS, genomically stable.

closely related to the occurrence and progression of cancer.

Development of the GILncSig

Colon cancer patients from TCGA were randomly divided into the training set and test set to further investigate the prognostic roles of these candidate genome instability-associated lncRNAs, as detailed in *Table 1*. It was observed that 12 genome instability-associated lncRNAs were significantly associated with the prognosis of colon cancer patients in the training set (all P<0.05), and the prognostic forest plot of these 12 lncRNAs is shown in *Figure 3*. Furthermore, the multivariate Cox proportional hazards regression analysis was conducted between the 12 candidate lncRNAs to single out the lncRNAs with independent prognostic value. Finally, 6 of the 12 candidate lncRNAs (LINC00896, AC007996.1, NKILA, AP003555.2, MIRLET7BHG, and AC009237.14) were identified as independent prognostic lncRNAs as their P values in the multivariate Cox analysis were less than 0.05 (*Table 2*). Subsequently, based on the expression levels of the six independent prognostic genomic instability-associated lncRNAs and the coefficients of the multivariate Cox analysis, a GILncSig was developed to determine the

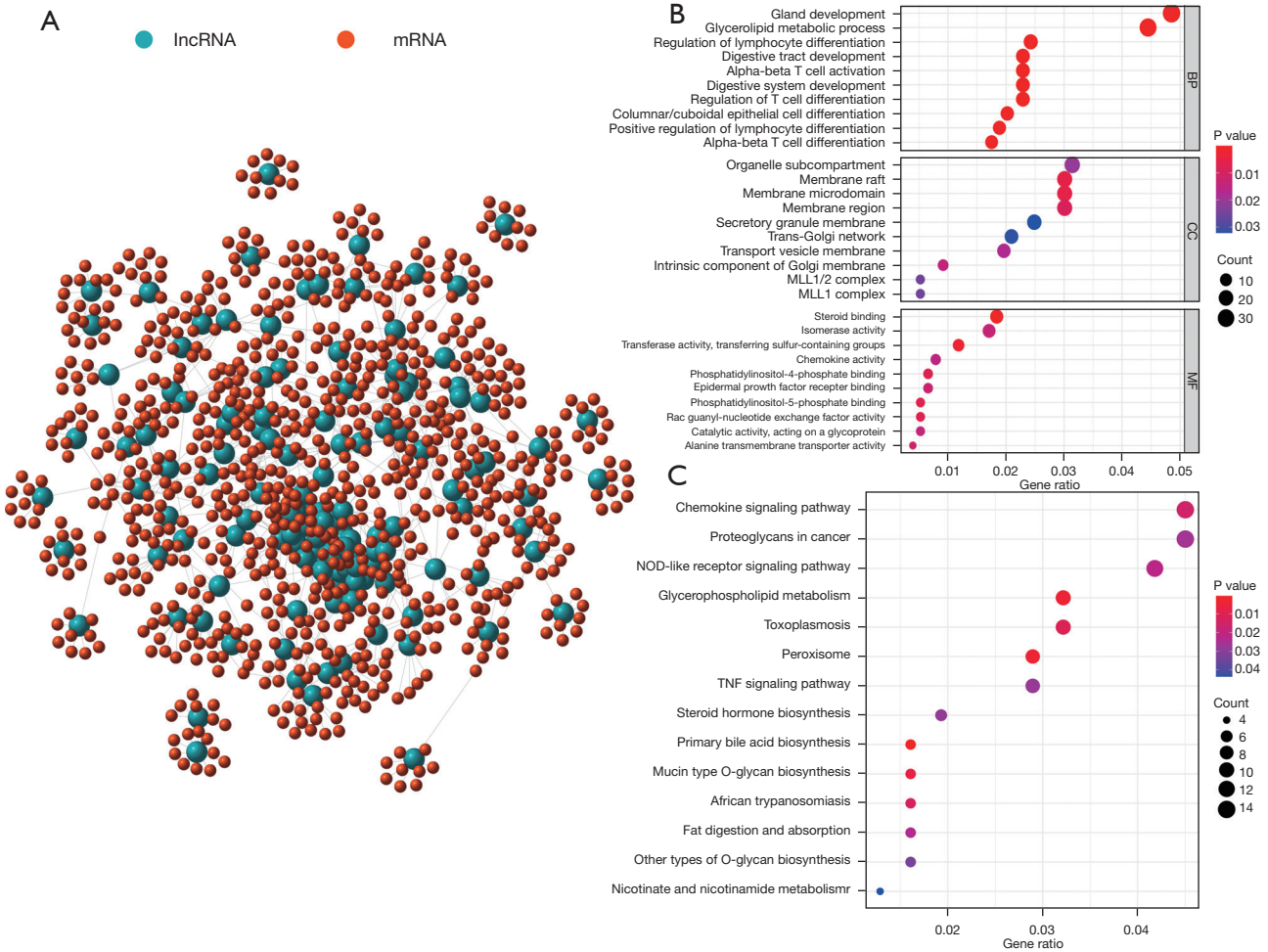

Figure 2 Co-expression network and enrichment analysis of differential lncRNAs. (A) Co-expression network of genomic instability-related lncRNAs and mRNAs. The blue circles represent lncRNAs and red circles represent mRNAs. (B) GO enrichment analysis and (C) KEGG pathway analysis of mRNAs associated with lncRNAs. LncRNAs, long non-coding RNAs; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

prognostic risk of patients with colon cancer. GILncSig score = (0.244155 × expression level of LINC00896) + (0.626462 × expression level of AC007996.1) + (0.295227 × expression level of NKILA) + (0.209866 × expression level of AP003555.2) + (0.351340 × expression level of MIRLET7BHG) + (0.196452 × expression level of AC009237.14). In the GILncSig score, the coefficients of all lncRNAs were positive, and their high expression levels were associated with poorer survival. The expression levels of lncRNAs in the GILncSig in the high- and low-risk groups are shown in *Figure 4*. Results showed that all 6 lncRNAs in the GILncSig were upregulated in the high-risk group, both in the training group (*Figure 4A*) and in TCGA patient groups (*Figure 4C*). However, only four lncRNAs (AC007996.1, NKILA, AP003555.2, and AC009237.14) were upregulated in the test group (*Figure 4B*).

Risk scores for patients in both the training set and the test set were obtained, and these patients were then classified into different prognostic groups using the median risk score as a threshold. We then performed Kaplan-Meier analyses of patients in the high- and low-risk groups, and the results showed that survival outcomes in the low-risk group were significantly better than those in the high-risk group in TCGA groups (training set, P<0.001, *Figure 5A*; test set, P=0.005, *Figure 5B*; TCGA sets, P<0.001, *Figure 5C*). Time-dependent ROC curve analysis of the GILncSig

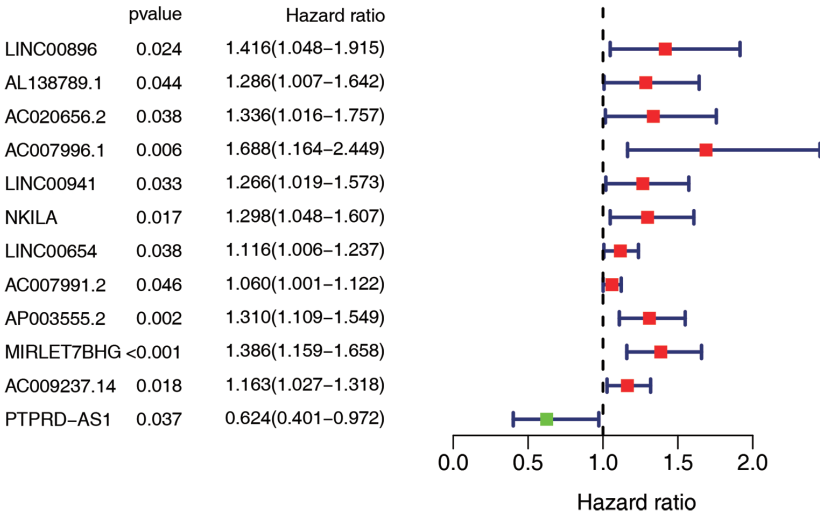

Figure 3 The prognostic forest plot of 12 genome instability-associated lncRNAs in the training set. LncRNAs, long non-coding RNAs.

Table 2 LncRNAs associated with the prognosis of colon cancer patients obtained after multivariate Cox analysis

ID	Coef	HR	HR.95L	HR.95H	P value
LINC00896	0.2441547	1.276541	0.960602	1.696392	0.092392
AC007996.1	0.6264617	1.870978	1.234731	2.835079	0.003133
NKILA	0.2952265	1.343430	1.080575	1.670227	0.007871
AP003555.2	0.2098663	1.233513	1.002866	1.517206	0.046915
MIRLET7BHG	0.3513404	1.420971	1.164792	1.733492	0.000532
AC009237.14	0.1964520	1.217076	1.068190	1.386715	0.003169

LncRNAs, long non-coding RNAs; coef, coefficient; HR, hazard ratio.

showed that the areas under the curve (AUCs) of the training set, testing set and TCGA groups were 0.691, 0.661, and 0.675, respectively (*Figure 5D-5F*).

The number of somatic mutations and the expression level of PD-L1 in each group

We then explored the number of somatic mutations and the differences in PD-L1 expression levels between the high- and low-risk patients. We first developed a set of risk figures for the three datasets, including a heat map of lncRNA expression and the distribution of patient risk scores. As is shown in *Figure 6A*, the expression levels of all lncRNAs in the training set increased with the increase in GILncSig score. These results were further verified in the test set and the TCGA set (*Figure 6B,6C*). As shown in *Figure 6D-6F*, the count of somatic mutations in patients in the high-risk group was higher than that of patients in the low-risk group (P=0.088, training set; P=0.066, test set; P=0.013, total set). The expression level of PD-L1 (CD274) was significantly higher in the high-risk group than in the low-risk group (P=0.01, training set; *Figure 6G*). However, in the test set, the expression of PD-L1 (CD274) was not significantly different among the high- and low-risk groups (P=0.15, *Figure 6H*). The expression level of PD-L1 (CD274) was significantly higher in the high-risk group than in the low-risk group (TCGA set; *Figure 6I*). We also did mutations of the molecule CD274 across different groups, but we found that it was very stable, with only two mutations in 400 patient samples (Table S2). Our results show that PD-L1 expression is higher in the high-risk group, suggesting that these patients may benefit from immunotherapy.

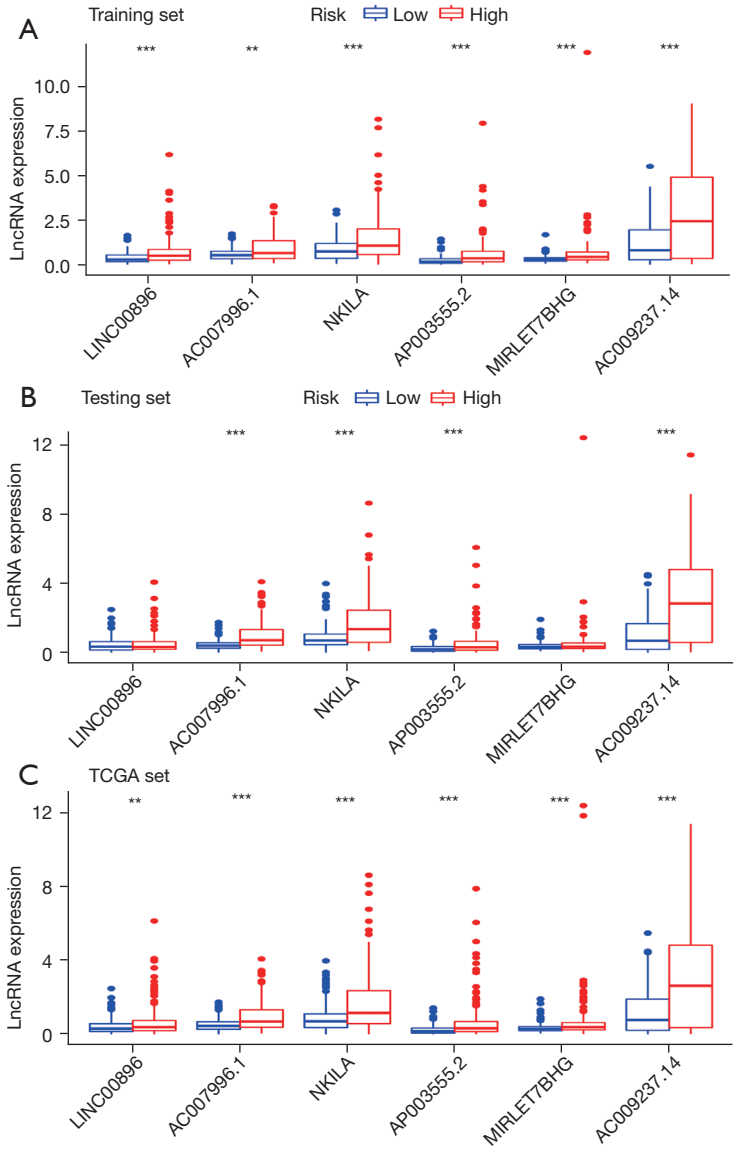

Figure 4 The expression levels of all six lncRNAs of the GILncSig in high- and low-risk groups from the training (A), test (B), and TCGA (C) sets. **, P<0.01; ***, P<0.001. LncRNAs, long non-coding RNAs; GILncSig, genome instability-derived lncRNA signature; TCGA, The Cancer Genome Atlas.

GILncSig and clinical factors

We used the chi-square test to investigate the relationship between risk and clinical characteristics. As shown in *Table 3*, the stages, T stage, and N stage were significantly correlated with the risk score in the training group and the total set. The higher a patient's risk score, the higher their stage and the higher their N stage. All the clinical characteristics showed no significant differences between the high- and low-risk groups in the test set.

The landscape of immune infiltration in COAD

To further investigate the relationship between risk score and immune cell infiltration, we used the CIBERSORT algorithm to estimate the proportions of 22 immune cell types in the COAD cohort from gene expression data. The Wilcoxon rank-sum test was used to explore whether there were differences between the 22 kinds of immune cells in different groups. As shown in *Figure 7A*, T follicular helper cells (P=0.018), resting NK cells (P=0.029), and M1

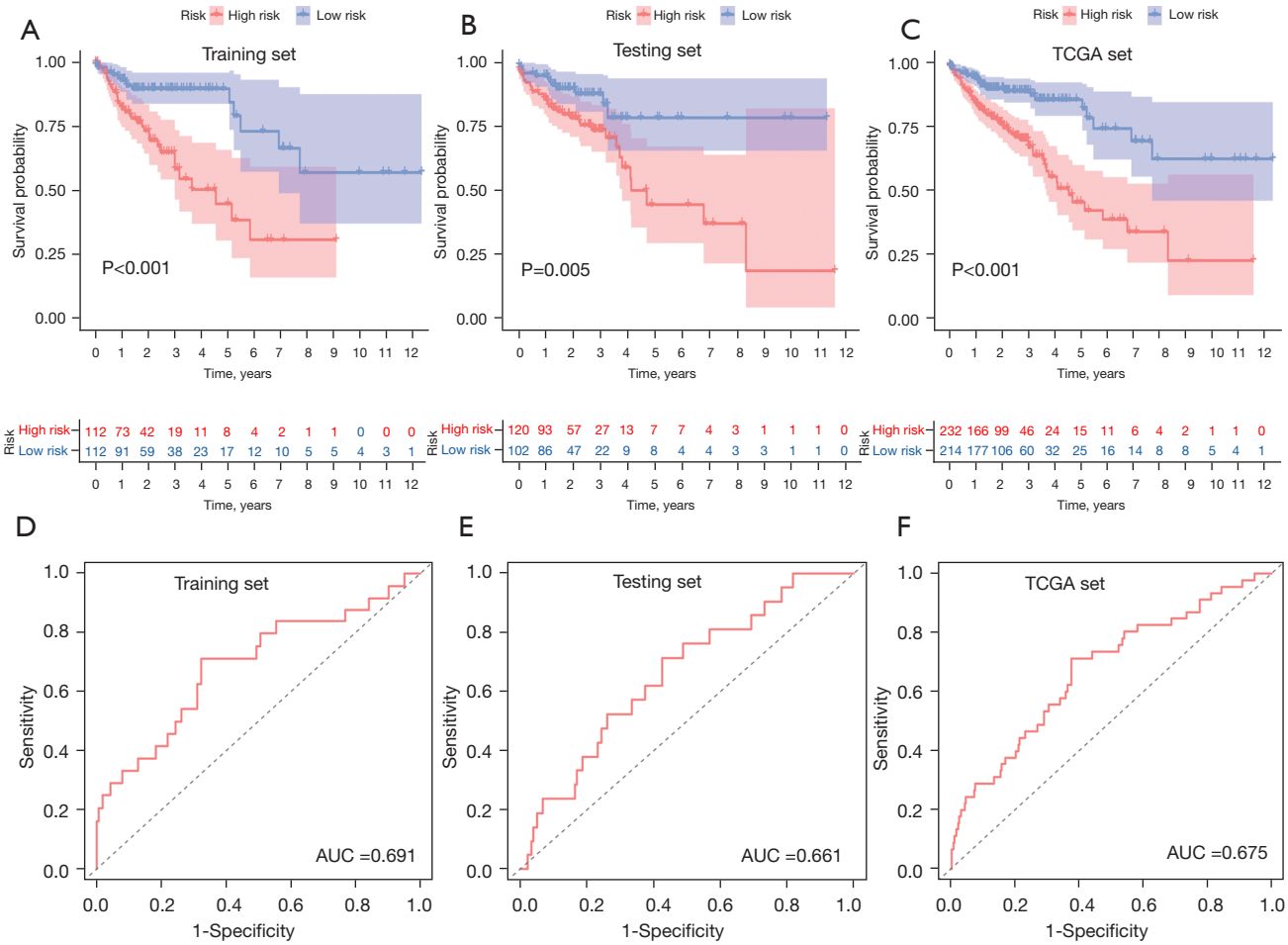

Figure 5 Kaplan-Meier curves and ROC curves of patients in the high- and low-risk groups. Kaplan-Meier curves of overall survival in low- and high-risk patients from the training (A), test (B), and TCGA (C) sets. Time-dependent ROC curve analysis of the GILncSig in the training (D), test (E), and TCGA (F) sets. ROC, receiver operating characteristic; GILncSig, genome instability-derived lncRNA signature; TCGA, The Cancer Genome Atlas.

macrophages (P=0.026) varied significantly between high- and low-risk score patients. In addition, Kaplan-Meier analysis showed that a lower proportion of regulatory T cells (Tregs) were associated with better overall survival (P=0.018, *Figure 7B*). However, a higher proportion of resting mast cells were associated with better overall survival (P=0.008, *Figure 7C*).

Discussion

The initiation, development, and treatment of colon cancer have been the subject of investigation for the past several years (20-22). Patients continue to be classified according to different therapeutic groups based on their pathological features, while the most important prognostic factors, such as the traditional histopathological features of tumor size, grade, and stage, also continue to be used (23,24). Nonetheless, due to the various limitations of traditional clinicopathological features, the clinical outcome of patients with colon cancer remains highly heterogeneous (25). Genomic instability is not only a common feature of most cancers, but is also considered to be one of the factors affecting the prognosis of colon cancer (26). In colon cancer progression and recurrence, genomic instability has a crucial and dominating role, thereby suggesting the important diagnostic and prognostic implications indicated by the pattern and degree of genomic instability (27,28). Nevertheless, quantitative measures of the degree

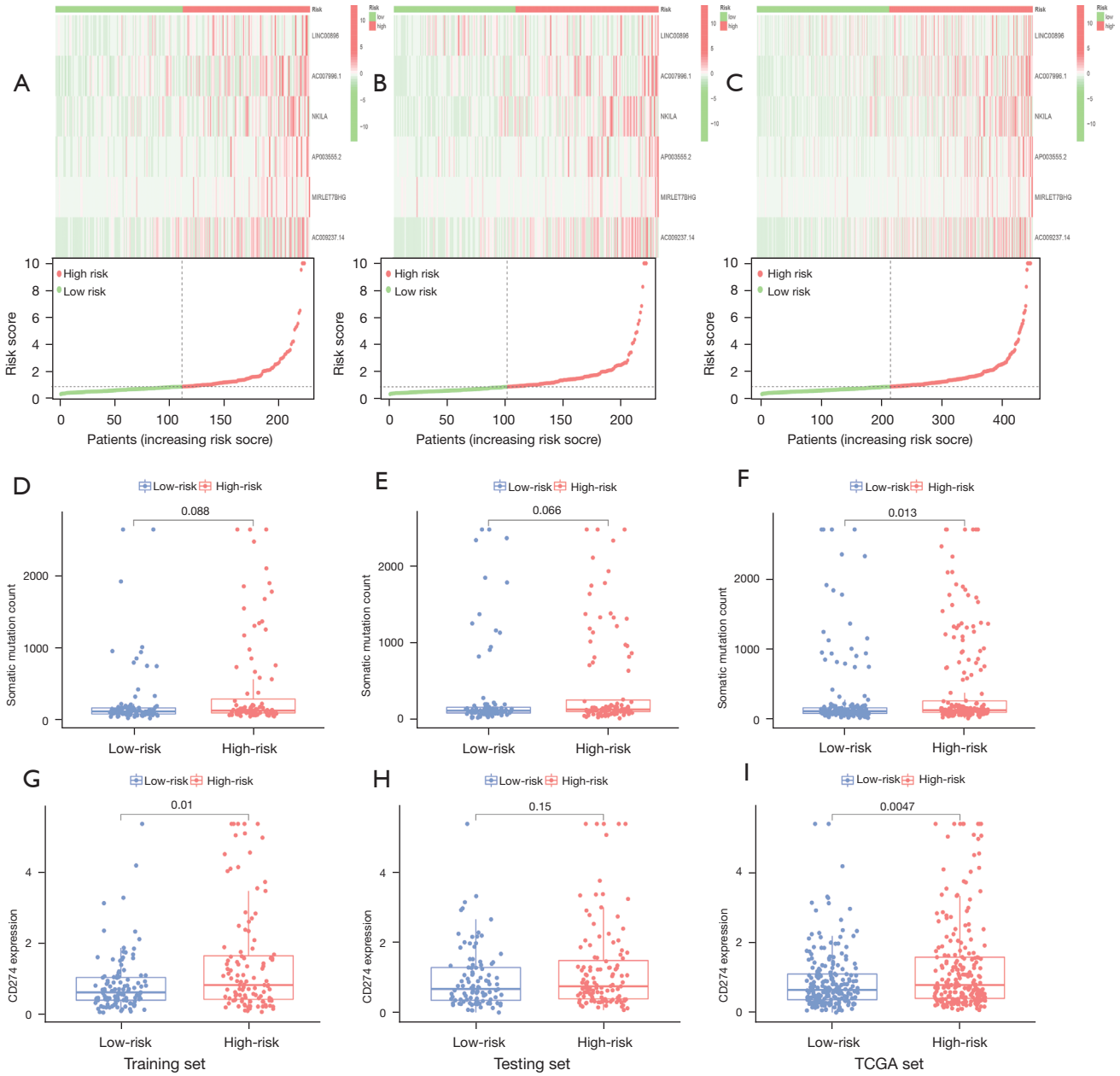

Figure 6 The number of somatic mutations and the expression level of CD274 in each group. LncRNA expression patterns with increasing GILncSig score in the training (A), test (B), and TCGA (C) sets. The distribution of somatic mutations in the training (D), test (E), and TCGA (F) sets. The expression levels of CD274 in the high- and low-risk groups from the training (G), test (H), and TCGA (I) sets. LncRNAs, long non-coding RNAs; GILncSig, genome instability-derived lncRNA signature; TCGA, The Cancer Genome Atlas.

of genomic instability have remained a challenge. For predicting genomic instability, concerted efforts are being made to develop gene or miRNA signatures and identify miRNAs associated with genomic instability (29).

LncRNAs, a novel class of non-coding RNAs, have recently gained significance as important components of tumors. The dysregulated expression of lncRNAs in cancer is related to disease progression, and lncRNAs have the potential to be used as prognostic markers for patients (30,31). With recent advances in the understanding of the

Table 3 Correlations between the risk scores of mutator-derived lncRNAs and the clinicopathological characteristics in the training set, test set, and total set

Parameters	Training set (n=224)				Testing set (n=222)				Total set (n=446)			
	HR	LR	χ^2	P	HR	LR	χ^2	P	HR	LR	χ^2	P
Age (y)												
≤65	42	48	0.4643	0.4956	52	41	0.1127	0.7371	94	89	0.0178	0.8938
>65	70	64			68	61			138	125		
Gender												
Male	57	59			66	52			123	111		
Female	55	53	0.0179	0.8936	54	50	0.2145	0.6432	109	103	0.0218	0.8826
Stage												
Stage I–II	52	72	7.4875	0.0062	64	62	0.9353	0.3335	116	134	7.3455	0.0067
Stage III–IV	59	37			52	37			111	74		
Unknown	4	1							5	6		
T												
T1–2	18	21	0.1035	0.7476	19	28	3.79	0.0516	37	49	2.9457	0.0861
T3–4	93	91			101	74			194	165		
Unknown	1	0							1	0		
M												
M0	79	89	1.5986	0.2061	86	75	0.6441	0.4222	165	164	2.5732	0.1087
M1	19	12			19	11			38	23		
Unknown	14	11			15	16			29	27		
N												
N0	53	78	10.5905	0.0011	68	66	1.1722	0.2789	121	144	9.956	0.0016
N1–2	59	34			52	36			111	70		

LncRNAs, long non-coding RNAs; HR, high risk; LR, low risk.

functional mechanisms of lncRNAs, lncRNAs are also critical to genome stability (32). The systematic exploration of their clinical significance in cancers and the genome-wide identification of genome instability-associated lncRNAs remain in their infancy despite certain efforts being made. Hence, to identify genome instability-associated lncRNAs combining lncRNA expression and the tumor mutator phenotype, a computational frame was developed. The results showed that 137 novel genome instability-associated lncRNAs were identified after the lncRNA expression profiles were combined with the somatic mutation profiles of colon cancer. We then constructed a lncRNA-mRNA co-expression network analysis of 128 genes that co-expressed lncRNAs associated with genomic instability. GO enrichment analysis showed that the mRNAs in this co-expression network were mainly enriched in gland development, organelle subcompartment, and steroid binding. According to KEGG pathway enrichment analysis, the main enrichment pathways were the chemokine signaling pathway, proteoglycans in cancer, and the NOD-like receptor signaling pathway, which are closely related to the occurrence and progression of cancer.

Colon cancer patients from TCGA were randomly divided into the training set and test set. Through multivariate Cox analysis, we observed that six genomic instability-related lncRNAs (LINC00896, AC007996.1, NKILA, AP003555.2, MIRLET7BHG, and AC009237.14) were significantly associated with the prognosis of patients in the training set.

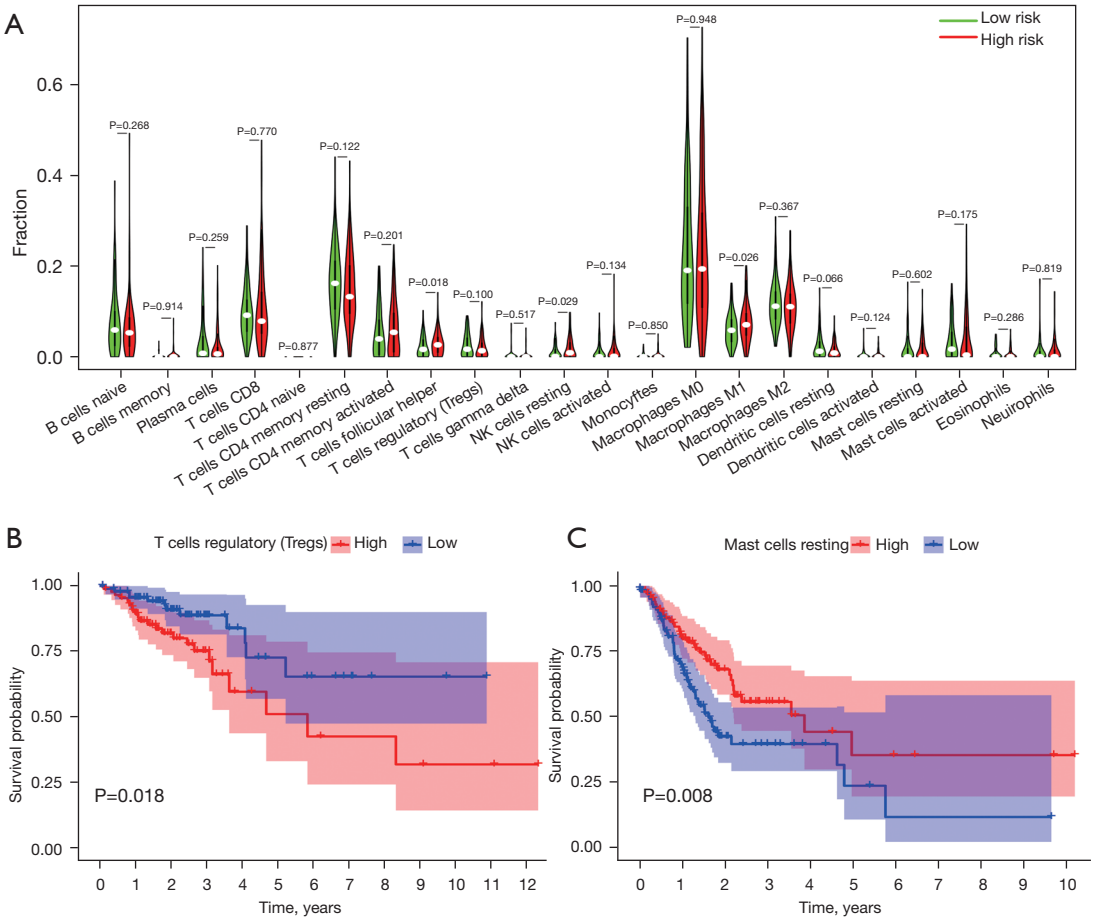

Figure 7 The landscape of immune infiltration in COAD. (A) The comparison of the fractions of immune cells between the high- and low-risk groups. Kaplan-Meier survival analysis of overall survival between patients with high and low levels of infiltrating Tregs (B) and resting mast cells (C). COAD, colon adenocarcinoma; Tregs, regulatory T cells.

A GILncSig was developed to determine the prognostic risk of patients with colon cancer: GILncSig score = (0.244155 × expression level of LINC00896) + (0.626462 × expression level of AC007996.1) + (0.295227 × expression level of NKILA) + (0.209866 × expression level of AP003555.2) + (0.351340 × expression level of MIRLET7BHG) + (0.196452 × expression level of AC009237.14). The results showed that all six lncRNAs were associated with poor prognosis, which has rarely been reported in the literature. NKILA is a lncRNA that interacts with NF-κB. A study showed that NKILA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death in breast cancer (33). Another study also showed that lncRNA AP003555.2 was associated with poor prognosis in patients with colon cancer (34). Furthermore, one study suggested that MIRLET7BHG was associated with polycystic ovary syndrome in women (35). Another study showed that autophagy-related lncRNA AC009237.14 was associated with poor prognosis in colon cancer patients (36). We further investigated the association between the GILncSig and clinical features. The results showed that higher risk scores were associated with more advanced stage and N stage. In addition, there was an obvious survival difference between the high- and low-risk groups according to our model, that is, the high-risk group had a shorter survival time and worse prognosis, and the AUC in TCGA groups was greater than 0.65, indicating that the prediction results of our model were relatively accurate.

Maintaining genomic stability to a large extent, nucleotide excision repair can specifically prevent mutations induced by environmental carcinogens. Moreover, some patients with MSI-H or DMMR CRC appear to be prone to persistent clinical responses to checkpoint inhibitors,

providing a new treatment option for patients with advanced disease (37). We then examined the number of somatic mutations and the expression of PD-L1 (CD274) between high- and low-risk patients. The results showed that the number of somatic mutations in the high-risk group was higher than that in the low-risk group, and the expression level of PD-L1 (CD274) in the high-risk group was significantly higher than that in the low-risk group. This suggests that our GILncSig is beneficial in differentiating patients who are sensitive to immunotherapy.

We further investigated the relationship between risk score and immune cell infiltration, and showed that there were more T follicular helper cells, resting NK cells, and M1 macrophages in the high-risk group. Besides, a low proportion of Tregs is related to a better prognosis. In a previous study, Tregs from cancer patients inhibited the mechanism of traditional T cell migration and thus affected patient outcomes (38). Mast cells have been shown to promote the development of colon cancer and could be a potential therapeutic target (39). All of these studies confirm our conclusions.

Although our study provides important insights for better assessing genomic instability and prognosis in colon cancer patients, it still has some limitations that require further investigation. We need more independent datasets such as Gene Expression Omnibus (GEO) datasets to validate the GILncSig to ensure its robustness and repeatability. In addition, we need to conduct further experimental validation to understand the regulatory mechanisms of the GILncSig in maintaining genomic instability.

Conclusions

In conclusion, examining the role of lncRNAs in genome instability through a mutator hypothesis-derived computational frame for identifying genome instability-associated lncRNAs was proposed by this study, providing a critical approach and resource for further studies. The levels of somatic mutations and PD-L1 (CD274) expression in the high-risk group were higher than those in the low-risk group, which indicated that high-risk groups may be sensitive to immunotherapy. The GILncSig can predict the outcome of patients with COAD and provides a new therapeutic direction.

Acknowledgments

For contributors to the TRIPOD statement, see https://www.annals.org.

Funding: This work was supported by the National Natural Science Foundation of China (No: 81970358), and Natural Science Foundation of Liaoning Province (1561449995778).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/jgo-21-494

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/jgo-21-494). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Institutional ethical approval and informed consent were waived.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Dienstmann R, Mason MJ, Sinicrope FA, et al. Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. Ann Oncol 2017;28:1023-31.
3. Hutchins G, Southward K, Handley K, et al. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. J Clin Oncol 2011;29:1261-70.
4. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic

instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol 2010;11:220-8.
5. Samowitz WS, Sweeney C, Herrick J, et al. Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. Cancer Res 2005;65:6063-9.
6. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113-30.
7. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135:1079-99.
8. Habermann JK, Doering J, Hautaniemi S, et al. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. Int J Cancer 2009;124:1552-64.
9. Mjelle R, Sjursen W, Thommesen L, et al. Small RNA expression from viruses, bacteria and human miRNAs in colon cancer tissue and its association with microsatellite instability and tumor location. BMC Cancer 2019;19:161.
10. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372:2509-20.
11. Llosa NJ, Cruise M, Tam A, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov 2015;5:43-51.
12. Sun M, Kraus WL. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. Endocr Rev 2015;36:25-64.
13. Sun M, Gadad SS, Kim DS, et al. Discovery, annotation, and functional analysis of long noncoding RNAs controlling cell-cycle gene expression and proliferation in breast cancer cells. Mol Cell 2015;59:698-711.
14. Lee S, Kopp F, Chang TC, et al. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. Cell 2016;164:69-80.
15. Pefanis E, Wang J, Rothschild G, et al. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell 2015;161:774-89.
16. Chen B, Dragomir MP, Fabris L, et al. The long noncoding RNA CCAT2 induces chromosomal instability through BOP1-AURKB signaling. Gastroenterology 2020;159:2146-2162.e33.
17. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.
18. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27-30.
19. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.
20. Dienstmann R, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. J Clin Oncol 2015;33:1787-96.
21. Gavin PG, Colangelo LH, Fumagalli D, et al. Mutation profiling and microsatellite instability in stage II and III colon cancer: an assessment of their prognostic and oxaliplatin predictive value. Clin Cancer Res 2012;18:6531-41.
22. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014;383:1490-502.
23. Klaver CEL, Wisselink DD, Punt CJA, et al. Adjuvant hyperthermic intraperitoneal chemotherapy in patients with locally advanced colon cancer (COLOPEC): a multicentre, open-label, randomised trial. Lancet Gastroenterol Hepatol 2019;4:761-70.
24. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 2018;391:2128-39.
25. Boyne DJ, Cuthbert CA, O'Sullivan DE, et al. Association between adjuvant chemotherapy duration and survival among patients with stage II and III colon cancer: a systematic review and meta-analysis. JAMA Netw Open 2019;2:e194154.
26. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology 2010;138:2059-72.
27. Chen WS, Chen JY, Liu JM, et al. Microsatellite instability in sporadic-colon-cancer patients with and without liver metastases. Int J Cancer 1997;74:470-4.
28. Westra JL, Schaapveld M, Hollema H, et al. Determination of TP53 mutation is more relevant than microsatellite instability status for the prediction of disease-free survival in adjuvant-treated stage III colon cancer patients. J Clin Oncol 2005;23:5635-43.
29. Mettu RK, Wan YW, Habermann JK, et al. A 12-gene genomic instability signature predicts clinical outcomes in multiple cancer types. Int J Biol Markers 2010;25:219-28.
30. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010;464:1071-6.
31. Wu Y, Zhang L, Zhang L, et al. Long non-coding RNA HOTAIR promotes tumor cell invasion and metastasis by recruiting EZH2 and repressing E-cadherin in oral

squamous cell carcinoma. Int J Oncol 2015;46:2586-94.
32. Hu WL, Jin L, Xu A, et al. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. Nat Cell Biol 2018;20:492-502.
33. Huang D, Chen J, Yang L, et al. NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. Nat Immunol 2018;19:1112-25.
34. Liu Y, Liu B, Jin G, et al. An integrated three-long non-coding RNA signature predicts prognosis in colorectal cancer patients. Front Oncol 2019;9:1269.
35. Butler AE, Hayat S, Dargham SR, et al. Alterations in long noncoding RNAs in women with and without polycystic ovarian syndrome. Clin Endocrinol (Oxf) 2019;91:793-7.
36. Cheng L, Han T, Zhang Z, et al. Identification and validation of six autophagy-related long non-coding RNAs as prognostic signature in colorectal cancer. Int J Med Sci 2021;18:88-98.
37. Oliveira AF, Bretes L, Furtado I. Review of PD-1/PD-L1 inhibitors in metastatic dMMR/MSI-H colorectal cancer. Front Oncol 2019;9:396.
38. Sundström P, Stenstad H, Langenes V, et al. Regulatory T cells from colon cancer patients inhibit effector T-cell migration through an adenosine-dependent mechanism. Cancer Immunol Res 2016;4:183-93.
39. Wang S, Li L, Shi R, et al. Mast cell targeted chimeric toxin can be developed as an adjunctive therapy in colon cancer treatment. Toxins (Basel) 2016;8:71.

(English Language Editor: C. Betlzar)

Cite this article as: Chen S, Li X, Zhang J, Li L, Wang X, Zhu Y, Guo L, Wang J. Six mutator-derived lncRNA signature of genome instability for predicting the clinical outcome of colon cancer. J Gastrointest Oncol 2021;12(5):2157-2171. doi: 10.21037/jgo-21-494

Table S1 Eighty-one lncRNAs were upregulated and 56 downregulated

Lnc	conMean	treatMean	LogFC	P value	FDR
SATB2-AS1	3.263644	1.331876	−1.29302	2.31E−12	2.27E−10
AL139022.1	0.215413	0.605613	1.491288	0.001622	0.005777
AC083837.1	0.274535	0.565583	1.042751	1.74E−06	1.78E−05
LINC00896	0.315604	0.833356	1.400815	1.51E−06	1.57E−05
LINC02418	4.469751	2.116022	−1.07884	2.13E−08	3.85E−07
AP000892.3	1.373846	0.56943	−1.27063	0.000281	0.001288
AC106037.2	0.185673	0.908019	2.289958	0.003639	0.011274
AL590764.1	1.0803	0.376926	−1.51908	0.000367	0.00165
AC007878.1	0.693002	1.953194	1.494904	0.007257	0.019408
AC108134.3	2.597882	1.03223	−1.33157	2.27E−12	2.27E−10
AP000439.2	6.933847	3.130735	−1.14715	1.09E−12	1.32E−10
AFAP1-AS1	1.227195	4.844596	1.981012	0.020043	0.043939
AL136115.2	0.467383	1.443113	1.626507	0.000972	0.003778
AP003071.4	0.545501	0.269064	−1.01963	5.14E−05	0.000326
CLDN10-AS1	1.170945	0.131933	−3.1498	1.03E−05	7.82E−05
EP300-AS1	0.602137	1.265507	1.071551	1.43E−11	7.76E−10
AC254629.1	4.695431	1.966961	−1.25529	6.14E−12	4.03E−10
AC114296.1	0.900505	0.22409	−2.00665	4.80E−11	2.22E−09
DLG3-AS1	0.427362	1.024622	1.261559	0.007271	0.019408
AC090181.2	0.947178	3.473434	1.874655	6.52E−10	2.18E−08
AC135388.1	0.771644	0.141235	−2.44984	2.40E−11	1.18E−09
AC100791.2	0.287139	0.667802	1.217671	0.000543	0.002284
AC026336.3	1.010562	0.261749	−1.9489	8.08E−09	1.77E−07
AC106876.1	7.297159	3.569219	−1.03173	2.28E−11	1.16E−09
LINC02041	0.624243	1.944109	1.638929	3.93E−10	1.41E−08
AL138789.1	0.181172	0.708085	1.966559	5.26E−05	0.000332
AC092756.1	0.274205	0.691437	1.334341	0.002458	0.008214
AL133330.1	0.534984	1.175185	1.135321	1.72E−05	0.000124
AC092112.1	0.113814	1.117944	3.296094	0.000429	0.001866
AC020656.2	0.342168	0.905786	1.404466	0.00023	0.001113
AC112484.3	0.243391	0.748677	1.621067	2.84E−05	0.000191
AC104695.3	1.476418	3.7837	1.357696	0.009679	0.024573
LINC00525	0.831191	0.254094	−1.70982	5.44E−15	2.14E−12
AL022316.1	0.438675	1.846152	2.073297	4.22E−08	6.57E−07
AC253536.3	0.358536	1.117436	1.640001	0.009721	0.024601
AC078993.1	4.208335	0.946461	−2.15263	1.45E−09	4.04E−08
AC005256.1	0.524732	1.590905	1.600195	5.87E−07	6.79E−06
AC136475.9	0.762446	0.1352	−2.49554	1.46E−09	4.04E−08
AC027290.2	0.480894	1.158222	1.268121	0.003076	0.00988
AC064801.1	0.199167	0.59223	1.572181	1.71E−07	2.25E−06
AP005271.1	0.777641	0.376353	−1.04702	6.76E−07	7.59E−06
AP005899.1	0.428537	1.282659	1.581647	2.98E−08	4.89E−07
AC130456.3	0.650838	1.716592	1.399176	4.89E−07	5.78E−06
AL391056.1	1.555264	0.688809	−1.17498	4.14E−10	1.45E−08
AC067945.1	0.265649	1.016358	1.935814	0.01529	0.035497
AC012317.1	0.441329	1.311318	1.571092	2.03E−05	0.000141

Table S1 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S1 (*continued*)

Lnc	conMean	treatMean	LogFC	P value	FDR
AC114760.2	0.344139	0.840834	1.288832	0.000852	0.003369
AC010998.3	0.23085	0.992921	2.10472	0.003529	0.011022
AC027607.1	0.364713	0.794035	1.122441	0.000347	0.001574
AL121761.1	0.768722	1.796768	1.22487	0.003922	0.012056
AC017074.1	2.783165	1.252842	–1.15152	4.85E–13	7.63E–11
AC132192.1	0.263042	0.579271	1.138946	0.005287	0.015259
AC007996.1	0.549296	1.337934	1.284351	2.03E–14	5.31E–12
AC036176.3	0.193443	0.776446	2.004976	0.000258	0.001214
LINC00941	0.48666	1.2478	1.3584	5.74E–07	6.69E–06
AC124067.4	10.37331	5.139954	–1.01305	6.33E–09	1.43E–07
AGAP1-IT1	0.786685	1.951518	1.310739	7.15E–07	7.98E–06
AC108865.2	1.791604	0.184711	–3.27791	0.000216	0.001056
AC032044.1	0.341677	0.698489	1.031604	6.10E–05	0.000365
NPTN-IT1	0.430873	1.098409	1.350079	0.014101	0.033227
AC108865.1	4.99933	0.406014	–3.62213	6.70E–06	5.52E–05
AP000785.1	0.648716	0.254914	–1.34758	1.39E–06	1.48E–05
AL121839.2	0.518638	1.077427	1.054791	2.24E–06	2.19E–05
LINC01082	1.768679	0.617166	–1.51894	1.34E–10	5.26E–09
LHFPL3-AS2	2.65626	1.028778	–1.36847	0.000146	0.000765
AC141930.1	0.201944	0.583749	1.53139	3.83E–08	6.03E–07
AC053503.4	2.498457	0.505568	–2.30506	5.15E–06	4.58E–05
AL606834.1	0.597492	1.474691	1.303422	5.20E–08	7.88E–07
AL162582.1	2.073973	0.799075	–1.37599	4.33E–06	3.88E–05
AC093585.1	0.688411	0.26148	–1.39657	2.83E–08	4.74E–07
AL390198.1	3.44889	1.38262	–1.31873	1.96E–08	3.66E–07
AL022313.2	1.007331	0.463489	–1.11993	1.64E–05	0.000119
AP002498.1	1.261239	3.622285	1.522058	0.0171	0.038615
AC091588.1	0.307998	1.290111	2.066502	0.00037	0.00166
AC012501.2	0.949959	0.3193	–1.57295	1.29E–09	3.83E–08
AC108058.1	0.555257	1.118173	1.009917	0.00494	0.014535
AP001554.1	0.694776	0.190462	–1.86704	4.89E–09	1.13E–07
AL121895.2	0.749682	0.358633	–1.06377	3.08E–07	3.73E–06
NKILA	1.696497	0.789099	–1.10428	5.95E–12	4.03E–10
AC005392.2	2.012855	4.656967	1.210147	0.006566	0.017819
AL139384.1	0.810717	0.400662	–1.01681	1.57E–09	4.19E–08
LINC00654	1.521138	0.396254	–1.94065	1.16E–14	3.66E–12
AC023825.2	0.228911	0.609865	1.413703	0.000307	0.001398
AP002761.4	1.827599	0.905832	–1.01263	9.01E–10	2.84E–08
AP000251.1	0.267671	0.597842	1.159301	0.01395	0.033019
AC009996.1	0.347304	0.793285	1.191642	0.000505	0.002164
LINC01443	0.131698	0.699876	2.409865	2.44E–05	0.000165
DLGAP1-AS5	0.054601	1.413953	4.694674	2.32E–05	0.00016
LINC01315	4.889959	2.419758	–1.01496	1.76E–08	3.38E–07
AC022784.1	0.207966	0.986104	2.245395	6.86E–08	1.02E–06
LOXL1-AS1	0.520056	1.226924	1.238308	1.44E–09	4.04E–08
AC115522.1	0.420704	0.928506	1.142107	2.80E–07	3.41E–06

Table S1 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S1 (*continued*)

Lnc	conMean	treatMean	LogFC	P value	FDR
AP001429.1	0.623855	1.521805	1.286501	0.002591	0.008584
AL109615.3	0.578406	1.169276	1.015462	2.94E–08	4.87E–07
AP001107.4	0.321578	0.720211	1.163249	0.000836	0.003322
AC007991.2	0.583263	2.098948	1.847448	0.011025	0.027231
AC073487.1	0.450909	1.002731	1.153026	5.75E–05	0.000351
AC004832.5	0.307127	0.679864	1.146411	0.00012	0.000652
AL121829.2	0.749064	0.316149	–1.24449	9.47E–06	7.38E–05
AL024508.1	0.447147	1.056642	1.240666	0.00018	0.000898
AP003555.2	0.412051	0.826696	1.004536	0.001451	0.005252
TUSC8	3.830048	1.277904	–1.58358	2.90E–15	1.52E–12
AC091057.4	0.488576	0.991232	1.020639	1.99E–07	2.51E–06
AC005911.1	0.412353	0.977438	1.245125	1.09E–10	4.51E–09
AC026801.2	0.69079	0.318617	–1.11643	2.85E–10	1.07E–08
AP001625.2	0.581906	1.203421	1.048285	0.000176	0.000885
TFAP2A-AS1	0.28822	1.161046	2.010183	8.80E–18	1.39E–14
AC068888.2	0.184258	0.721032	1.968338	0.007665	0.020126
XXYLT1-AS2	0.252436	0.611525	1.276494	9.77E–06	7.54E–05
BOK-AS1	1.629157	0.556553	–1.54954	0.000123	0.000664
LINC02195	0.424969	1.108173	1.382753	1.34E–09	3.91E–08
AC009812.1	1.351057	0.573863	–1.23531	2.10E–12	2.27E–10
AC009041.2	0.713702	1.585362	1.151419	0.000153	0.000792
AC023043.4	0.741086	1.553118	1.067454	2.49E–14	5.60E–12
FENDRR	2.206017	1.029552	–1.09943	6.74E–12	4.24E–10
AL157871.2	0.286166	1.020079	1.833757	1.98E–07	2.51E–06
AC018410.1	0.159333	0.688157	2.11069	0.006224	0.017215
HAR1B	0.373786	0.843845	1.174768	1.81E–05	0.000128
MIRLET7BHG	0.41134	0.947051	1.203111	6.09E–07	6.94E–06
LINC00543	3.583235	1.745928	–1.03727	2.66E–12	2.47E–10
SMIM2-AS1	1.689804	0.828729	–1.02788	4.36E–09	1.02E–07
LINC02441	3.840394	1.251533	–1.61756	2.18E–13	3.81E–11
AC025423.1	0.240071	0.610941	1.347573	0.000125	0.000671
AC009237.14	2.184394	1.033768	–1.07932	1.14E–11	6.42E–10
AL031651.2	0.573795	0.285993	–1.00456	0.021189	0.046003
AC009318.4	0.274172	0.631737	1.204246	0.006267	0.017271
LINC02446	0.24564	1.33461	2.4418	5.19E–12	3.71E–10
RHPN1-AS1	1.079381	0.434489	–1.31281	9.39E–10	2.84E–08
BOLA3-AS1	0.757076	0.330146	–1.19733	2.16E–10	8.27E–09
AC010280.1	0.91508	0.297898	–1.61908	1.34E–08	2.74E–07
OSER1-DT	5.511887	2.314471	–1.25186	4.58E–17	3.60E–14
PTPRD-AS1	0.97144	0.352669	–1.46181	1.10E–11	6.42E–10
DPP10-AS1	0.996428	0.465868	–1.09684	6.57E–06	5.50E–05
RARA-AS1	3.902149	1.506859	–1.37272	1.69E–08	3.29E–07
AL162724.2	0.506196	1.128372	1.156475	0.006072	0.016975
LINC02489	0.483755	1.608278	1.733169	1.43E–06	1.51E–05
FEZF1-AS1	0.905064	2.286439	1.337011	7.23E–10	2.37E–08

LncRNAs, long non-coding RNAs; FDR, false discovery rate.

https://dx.doi.org/10.21037/jgo-21-494

Table S2 Mutation analysis of the molecule CD274 across different groups

Gene	CD274
TCGA-AA-3984	Mutation
TCGA-AA-3684	Mutation
TCGA-AA-A02R	Wild
TCGA-5M-AAT5	Wild
TCGA-AA-3488	Wild
TCGA-AA-3844	Wild
TCGA-AA-3678	Wild
TCGA-CM-4746	Wild
TCGA-CM-5862	Wild
TCGA-DM-A280	Wild
TCGA-AA-3667	Wild
TCGA-AY-A71X	Wild
TCGA-D5-6537	Wild
TCGA-CM-6680	Wild
TCGA-G4-6317	Wild
TCGA-QG-A5YV	Wild
TCGA-AA-A00N	Wild
TCGA-D5-6531	Wild
TCGA-D5-6539	Wild
TCGA-G4-6298	Wild
TCGA-CK-6747	Wild
TCGA-A6-2677	Wild
TCGA-F4-6461	Wild
TCGA-AA-3848	Wild
TCGA-AA-3979	Wild
TCGA-AA-3866	Wild
TCGA-G4-6293	Wild
TCGA-AA-A01I	Wild
TCGA-AZ-6601	Wild
TCGA-AA-A01Z	Wild
TCGA-AZ-6598	Wild
TCGA-A6-6782	Wild
TCGA-AA-3697	Wild
TCGA-D5-6540	Wild
TCGA-AY-5543	Wild
TCGA-A6-2674	Wild
TCGA-CK-4948	Wild
TCGA-AU-6004	Wild
TCGA-AA-3673	Wild
TCGA-AA-A01X	Wild
TCGA-CM-6172	Wild
TCGA-CA-6719	Wild
TCGA-CM-6162	Wild
TCGA-AD-6895	Wild
TCGA-DM-A1D4	Wild
TCGA-G4-6309	Wild
TCGA-AY-4070	Wild
TCGA-DM-A28A	Wild
TCGA-A6-2682	Wild
TCGA-A6-2675	Wild
TCGA-AA-3989	Wild
TCGA-G4-6306	Wild
TCGA-AA-3663	Wild
TCGA-AA-3833	Wild
TCGA-AD-6964	Wild
TCGA-A6-6780	Wild
TCGA-D5-6922	Wild
TCGA-G4-6311	Wild
TCGA-WS-AB45	Wild
TCGA-AA-3713	Wild
TCGA-A6-A565	Wild
TCGA-AA-3860	Wild
TCGA-AA-3712	Wild
TCGA-G4-6626	Wild
TCGA-CM-6674	Wild
TCGA-AZ-5403	Wild
TCGA-T9-A92H	Wild
TCGA-AD-6548	Wild
TCGA-AA-3877	Wild
TCGA-AY-A69D	Wild
TCGA-QG-A5YX	Wild
TCGA-AA-A03J	Wild
TCGA-AD-6965	Wild
TCGA-AA-3685	Wild
TCGA-CK-4952	Wild
TCGA-D5-6930	Wild
TCGA-CM-4744	Wild
TCGA-D5-6529	Wild
TCGA-DM-A28H	Wild
TCGA-AA-3502	Wild
TCGA-AA-A02E	Wild
TCGA-A6-3810	Wild
TCGA-CM-4752	Wild
TCGA-AA-3971	Wild
TCGA-AA-3666	Wild
TCGA-A6-2685	Wild
TCGA-G4-6297	Wild
TCGA-CK-5912	Wild
TCGA-RU-A8FL	Wild
TCGA-CK-5913	Wild
TCGA-G4-6299	Wild

Table S2 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S2 (*continued*)

Gene	CD274
TCGA-CA-5254	Wild
TCGA-F4-6570	Wild
TCGA-DM-A1D7	Wild
TCGA-AZ-6599	Wild
TCGA-DM-A28C	Wild
TCGA-AA-3941	Wild
TCGA-AA-3818	Wild
TCGA-A6-2679	Wild
TCGA-CK-6748	Wild
TCGA-5M-AATA	Wild
TCGA-D5-5538	Wild
TCGA-4N-A93T	Wild
TCGA-AA-3496	Wild
TCGA-AA-A01V	Wild
TCGA-AA-3854	Wild
TCGA-AD-6901	Wild
TCGA-AA-A004	Wild
TCGA-AZ-4313	Wild
TCGA-AZ-6603	Wild
TCGA-CM-6677	Wild
TCGA-CA-6717	Wild
TCGA-G4-6321	Wild
TCGA-AA-3841	Wild
TCGA-D5-5541	Wild
TCGA-G4-6294	Wild
TCGA-QL-A97D	Wild
TCGA-D5-6926	Wild
TCGA-A6-5667	Wild
TCGA-AA-3811	Wild
TCGA-A6-5661	Wild
TCGA-AA-3845	Wild
TCGA-AZ-6605	Wild
TCGA-A6-2686	Wild
TCGA-AA-A02H	Wild
TCGA-AA-A02W	Wild
TCGA-AZ-6606	Wild
TCGA-AA-3852	Wild
TCGA-AA-3947	Wild
TCGA-AA-A02Y	Wild
TCGA-AA-A010	Wild
TCGA-AA-A02F	Wild
TCGA-AA-3492	Wild
TCGA-F4-6855	Wild
TCGA-F4-6808	Wild
TCGA-AA-3850	Wild
TCGA-CA-6716	Wild

Table S2 (*continued*)

Table S2 (*continued*)

Gene	CD274
TCGA-CK-4950	Wild
TCGA-G4-6628	Wild
TCGA-NH-A5IV	Wild
TCGA-AA-3664	Wild
TCGA-AZ-6608	Wild
TCGA-G4-6322	Wild
TCGA-CM-6163	Wild
TCGA-F4-6805	Wild
TCGA-CM-6171	Wild
TCGA-AZ-6607	Wild
TCGA-AA-3494	Wild
TCGA-A6-6651	Wild
TCGA-AA-3693	Wild
TCGA-AA-A02K	Wild
TCGA-CA-5255	Wild
TCGA-AA-A01P	Wild
TCGA-AA-3855	Wild
TCGA-AA-3861	Wild
TCGA-AA-3977	Wild
TCGA-A6-A567	Wild
TCGA-AM-5820	Wild
TCGA-A6-6781	Wild
TCGA-AA-3858	Wild
TCGA-AA-A022	Wild
TCGA-CM-4743	Wild
TCGA-AY-6197	Wild
TCGA-AA-3955	Wild
TCGA-AA-3939	Wild
TCGA-CM-6169	Wild
TCGA-AA-3966	Wild
TCGA-D5-6929	Wild
TCGA-CM-5341	Wild
TCGA-A6-6653	Wild
TCGA-A6-5656	Wild
TCGA-D5-5539	Wild
TCGA-AD-5900	Wild
TCGA-AA-3821	Wild
TCGA-F4-6806	Wild
TCGA-A6-2680	Wild
TCGA-D5-6536	Wild
TCGA-DM-A0X9	Wild
TCGA-DM-A28E	Wild
TCGA-G4-6586	Wild
TCGA-AA-3968	Wild
TCGA-AZ-4323	Wild
TCGA-AA-A024	Wild

Table S2 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S2 (*continued*)

Gene	CD274
TCGA-F4-6569	Wild
TCGA-AZ-4681	Wild
TCGA-G4-6588	Wild
TCGA-CA-6715	Wild
TCGA-DM-A28G	Wild
TCGA-F4-6704	Wild
TCGA-AA-3864	Wild
TCGA-F4-6459	Wild
TCGA-CM-6161	Wild
TCGA-A6-5660	Wild
TCGA-A6-2671	Wild
TCGA-AA-3715	Wild
TCGA-AA-3872	Wild
TCGA-D5-6534	Wild
TCGA-A6-3807	Wild
TCGA-AZ-6600	Wild
TCGA-DM-A28M	Wild
TCGA-A6-2672	Wild
TCGA-AA-3510	Wild
TCGA-AA-3662	Wild
TCGA-F4-6703	Wild
TCGA-AA-3819	Wild
TCGA-CK-5914	Wild
TCGA-AA-A01S	Wild
TCGA-NH-A6GA	Wild
TCGA-NH-A6GC	Wild
TCGA-AD-6889	Wild
TCGA-CM-6676	Wild
TCGA-D5-7000	Wild
TCGA-AZ-4315	Wild
TCGA-A6-5659	Wild
TCGA-AA-3812	Wild
TCGA-AZ-4616	Wild
TCGA-5M-AAT4	Wild
TCGA-AA-3495	Wild
TCGA-AA-3696	Wild
TCGA-AA-3506	Wild
TCGA-A6-6137	Wild
TCGA-QG-A5YW	Wild
TCGA-AA-3675	Wild
TCGA-DM-A1DB	Wild
TCGA-3L-AA1B	Wild
TCGA-CM-6165	Wild
TCGA-AA-3851	Wild
TCGA-AA-A01T	Wild
TCGA-AD-A5EK	Wild

Table S2 (*continued*)

Table S2 (*continued*)

Gene	CD274
TCGA-NH-A50U	Wild
TCGA-AA-3930	Wild
TCGA-AA-3982	Wild
TCGA-CA-5796	Wild
TCGA-AA-A017	Wild
TCGA-NH-A6GB	Wild
TCGA-AY-6386	Wild
TCGA-D5-6927	Wild
TCGA-AZ-4614	Wild
TCGA-CK-5915	Wild
TCGA-CM-6678	Wild
TCGA-CM-5863	Wild
TCGA-A6-6648	Wild
TCGA-F4-6856	Wild
TCGA-CM-6167	Wild
TCGA-AD-A5EJ	Wild
TCGA-G4-6304	Wild
TCGA-5M-AATE	Wild
TCGA-AA-3846	Wild
TCGA-CA-5256	Wild
TCGA-AA-3976	Wild
TCGA-CM-5861	Wild
TCGA-AY-4071	Wild
TCGA-G4-6323	Wild
TCGA-AA-3660	Wild
TCGA-D5-6920	Wild
TCGA-CM-6679	Wild
TCGA-CA-5797	Wild
TCGA-CM-4750	Wild
TCGA-NH-A50T	Wild
TCGA-AA-3949	Wild
TCGA-AA-3956	Wild
TCGA-F4-6854	Wild
TCGA-CM-5868	Wild
TCGA-DM-A1HB	Wild
TCGA-DM-A1D8	Wild
TCGA-AA-3952	Wild
TCGA-CM-6170	Wild
TCGA-D5-6538	Wild
TCGA-CK-5916	Wild
TCGA-AZ-4684	Wild
TCGA-G4-6320	Wild
TCGA-DM-A288	Wild
TCGA-D5-6530	Wild
TCGA-AA-3972	Wild
TCGA-A6-6654	Wild

Table S2 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S2 (*continued*)

Gene	CD274
TCGA-AA-3814	Wild
TCGA-AA-3688	Wild
TCGA-DM-A1HA	Wild
TCGA-AA-3692	Wild
TCGA-A6-2684	Wild
TCGA-AA-3870	Wild
TCGA-AA-A01C	Wild
TCGA-G4-6303	Wild
TCGA-D5-6532	Wild
TCGA-AA-3679	Wild
TCGA-AA-3862	Wild
TCGA-AA-3856	Wild
TCGA-G4-6295	Wild
TCGA-SS-A7HO	Wild
TCGA-DM-A1D6	Wild
TCGA-D5-5540	Wild
TCGA-NH-A50V	Wild
TCGA-AD-6963	Wild
TCGA-CM-6166	Wild
TCGA-CM-5348	Wild
TCGA-AA-3680	Wild
TCGA-F4-6463	Wild
TCGA-AA-3530	Wild
TCGA-D5-5537	Wild
TCGA-AA-3875	Wild
TCGA-AA-3511	Wild
TCGA-A6-6142	Wild
TCGA-D5-6932	Wild
TCGA-QG-A5Z1	Wild
TCGA-AY-A54L	Wild
TCGA-NH-A8F7	Wild
TCGA-AZ-5407	Wild
TCGA-AA-3695	Wild
TCGA-AA-3842	Wild
TCGA-AA-3973	Wild
TCGA-AD-6890	Wild
TCGA-DM-A28F	Wild
TCGA-AA-3681	Wild
TCGA-QG-A5Z2	Wild
TCGA-CM-4751	Wild
TCGA-AA-3994	Wild
TCGA-AA-3950	Wild
TCGA-A6-5666	Wild
TCGA-A6-4107	Wild
TCGA-AA-3815	Wild
TCGA-CM-5349	Wild

Table S2 (*continued*)

Table S2 (*continued*)

Gene	CD274
TCGA-DM-A0XD	Wild
TCGA-D5-6931	Wild
TCGA-DM-A282	Wild
TCGA-A6-5664	Wild
TCGA-CM-4747	Wild
TCGA-CM-6168	Wild
TCGA-A6-3809	Wild
TCGA-D5-6898	Wild
TCGA-AA-3975	Wild
TCGA-A6-6650	Wild
TCGA-AD-6888	Wild
TCGA-AA-3655	Wild
TCGA-D5-6541	Wild
TCGA-AA-3509	Wild
TCGA-CK-6751	Wild
TCGA-A6-6652	Wild
TCGA-A6-A566	Wild
TCGA-DM-A285	Wild
TCGA-F4-6460	Wild
TCGA-D5-6928	Wild
TCGA-AA-3831	Wild
TCGA-AA-3867	Wild
TCGA-D5-6924	Wild
TCGA-G4-6315	Wild
TCGA-A6-4105	Wild
TCGA-G4-6627	Wild
TCGA-A6-5665	Wild
TCGA-G4-6302	Wild
TCGA-A6-2681	Wild
TCGA-AA-3672	Wild
TCGA-AA-A02O	Wild
TCGA-AA-3980	Wild
TCGA-A6-3808	Wild
TCGA-CK-6746	Wild
TCGA-NH-A8F8	Wild
TCGA-AA-3986	Wild
TCGA-DM-A28K	Wild
TCGA-5M-AAT6	Wild
TCGA-AY-A8YK	Wild
TCGA-AA-A01R	Wild
TCGA-CA-6718	Wild
TCGA-A6-6649	Wild
TCGA-A6-A56B	Wild
TCGA-G4-6307	Wild
TCGA-G4-6310	Wild
TCGA-AU-3779	Wild

Table S2 (*continued*)

https://dx.doi.org/10.21037/jgo-21-494

Table S2 (*continued*)

Gene	CD274
TCGA-AA-3837	Wild
TCGA-F4-6809	Wild
TCGA-AZ-4615	Wild
TCGA-CM-5864	Wild
TCGA-DM-A1D0	Wild
TCGA-AA-3967	Wild
TCGA-CK-4947	Wild
TCGA-AA-A03F	Wild
TCGA-CM-4748	Wild
TCGA-CM-6675	Wild
TCGA-G4-6625	Wild
TCGA-D5-6533	Wild
TCGA-A6-5657	Wild
TCGA-AM-5821	Wild
TCGA-AA-3869	Wild
TCGA-A6-6138	Wild
TCGA-CK-4951	Wild
TCGA-AA-3489	Wild
TCGA-CM-5344	Wild
TCGA-D5-6535	Wild
TCGA-4T-AA8H	Wild
TCGA-DM-A1DA	Wild
TCGA-AZ-4682	Wild
TCGA-DM-A0XF	Wild
TCGA-CM-6164	Wild
TCGA-AD-6899	Wild
TCGA-AY-6196	Wild
TCGA-F4-6807	Wild
TCGA-DM-A1D9	Wild
TCGA-CM-5860	Wild
TCGA-G4-6314	Wild
TCGA-AZ-4308	Wild

TCGA, The Cancer Genome Atlas.

https://dx.doi.org/10.21037/jgo-21-494