



# Identification of Epithelial-Mesenchymal Transition-Related lncRNAs that Associated With the Prognosis and Immune Microenvironment in Colorectal Cancer

## OPEN ACCESS

### Edited by:

Rui Cao,  
Capital Medical University, China

### Reviewed by:

Tingting Yan,  
Shanghai JiaoTong University, China  
J. Joshua Smith,  
Memorial Sloan Kettering Cancer  
Center, United States

### \*Correspondence:

Chengliang Zhao  
38221965@qq.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Protein and RNA Networks,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 26 November 2020

**Accepted:** 19 February 2021

**Published:** 01 April 2021

### Citation:

Liu C, Hu C, Li J, Jiang L and Zhao C  
(2021) Identification of Epithelial-  
Mesenchymal Transition-Related  
lncRNAs that Associated With the  
Prognosis and Immune  
Microenvironment in  
Colorectal Cancer.  
*Front. Mol. Biosci.* 8:633951.  
doi: 10.3389/fmolb.2021.633951

Chuan Liu<sup>1†</sup>, Chuan Hu<sup>2†</sup>, Jianyi Li<sup>2</sup>, Liqing Jiang<sup>1</sup> and Chengliang Zhao<sup>2\*</sup>

<sup>1</sup>Graduate College, China Medical University, Shenyang, China, <sup>2</sup>Department of Spinal Surgery, The Affiliated Hospital of Qingdao University, Qingdao, China

**Background:** The expression of long non-coding RNA (lncRNA) is associated with the epithelial-mesenchymal transition (EMT) in tumorigenicity, but the role of EMT-related lncRNA in colorectal cancer (CRC) remains unclear.

**Methods:** The clinical data and gene expression profile of CRC patients were obtained from The Cancer Genome Atlas database. Differential expression analysis, Cox regression model, and Kaplan-Meier analysis were used to study the relationship between EMT-related lncRNAs and the prognosis of CRC. Functional analysis and unsupervised clustering analysis were performed to explore the influence of certain lncRNAs on CRC. Finally, Cytoscape was used to construct mRNA-lncRNA networks.

**Results:** Two signatures incorporating six and ten EMT-related lncRNAs were constructed for predicting the overall survival (OS) and disease-free survival (DFS), respectively. Kaplan-Meier survival curves indicated that patients in the high-risk group had a poorer prognosis than those in the low-risk group. The results of the functional analysis suggested that the P53 and ECM-receptor pathways affect the prognosis of CRC, and AL591178.1 is a key prognostic EMT-related lncRNA, which is negatively related to immune cells, P53 pathway, and ECM-receptor pathway.

**Conclusion:** Six OS-related and ten DFS-related EMT-related lncRNAs were correlated with the prognosis of CRC by potentially affecting the immune microenvironment, and AL591178.1 plays a key role as a prognostic factor.

**Keywords:** epithelial-mesenchymal transition, lncRNA, colorectal cancer, immune cells, prognosis

## INTRODUCTION

Epithelial-mesenchymal transition (EMT) is a dynamic and reversible process that occurs in embryogenesis, organ development, wound healing, and fibrosis (Hay, 2005; Kalluri and Weinberg, 2009; Zeisberg and Neilson, 2009). It is characterized by the interaction between polarized epithelial cells and the basement membrane. Different processes make them show different mesenchymal cell phenotypes, including invasiveness migration, anti-apoptosis, and high levels of extracellular matrix (ECM) (Kalluri and Neilson, 2003). More importantly, there is evidence that the dysregulation of EMT plays an important role in the occurrence and progression of cancers, especially in the process of invasion, metastasis, and resistance of antineoplastic drug (Montemayor-Garcia et al., 2013; Zhao et al., 2013; Nieto et al., 2016). Recent studies have shown that the intermediate state of EMT is partial EMT (p-EMT), which is involved in the tumor progression and promotes the migration of tumor cells and the formation of circulating tumor cell clusters (Aiello et al., 2018; Saitoh, 2018). Besides, the process of EMT is regulated by various genes to affect the tumor progression (Huang et al., 2020; Xie et al., 2020). Thus, it is essential to further explore the function of the EMT-related genes in cancers.

With the emergence of next-generation sequencing technology, comprehensive genomic atlas based on tumor molecular characteristics have been received extensive research (Sanchez-Vega et al., 2018; Nacev et al., 2019). Different genomics, including mRNA, microRNA, long non-coding RNA (lncRNA), and other molecules, help to broaden our understanding of genomic characteristics in different tumors. A series of studies have proved that lncRNAs play an important role in the progression of tumors and can be used as robust predictors of the prognosis for cancer patients (Wu et al., 2020; Zhuang et al., 2020). Interestingly, the EMT mechanism of lncRNA is involved. For example, lncRNA ADAMTS9-AS1 is an EMT-related gene and is associated with lymph node invasion and prognosis of colorectal cancer (CRC) (Chen W. et al., 2020). Additionally, increased expression of lncRNA PTAR can promote EMT and metastasis in tumorigenicity of ovarian cancer cells (Liang et al., 2018). However, most research mainly focused on the impact of a single lncRNA on tumors, and it is necessary to conduct a comprehensive analysis of the potential role of EMT-related lncRNAs on cancer prognosis.

CRC is the most common gastrointestinal adenocarcinoma, with 1.8 million new cases occurring every year and causing approximately 900,000 deaths (Bray et al., 2018). In recent years, morbidity and mortality have been on the rise in many countries, especially in Asian (Chen et al., 2016). Although the improvement of early diagnosis, immunotherapy, and chemotherapy has significantly improved the diagnosis and treatment of CRC, the survival rate is still not ideal (Cunningham et al., 2010). Due to the molecular heterogeneity of CRC cells, it is common to get completely different clinical outcomes (Guinney et al., 2015). Therefore, it is imperative to explore the molecular mechanisms of the prognosis for CRC patients. In particular, EMT is a field associated with lncRNA

expression and cancer development, which may lead to a deeper understanding of this kind of tumor and provide new insights.

## METHODS

### Data Collection and Processing

We obtained the gene expression profile of CRC patients from the TCGA database (<https://cancergenome.nih.gov/>), and the corresponding clinical information were obtained from the cBioportal database (<https://www.cbioportal.org/>) (Cerami et al., 2012). According to the gene annotation in the GENCODE project (<https://www.encodegenes.org/>) (Derrien et al., 2012), a list of lncRNA was distinguished. Then the clinical data of CRC patients were processed for further analysis, including age, gender, tumor grade, AJCC TNM stage, follow-up time and follow up data. All these data are publicly available, thus no specific ethical approval and informed consent are required and the work is exempt.

### Identification of Differentially Expressed Epithelial-Mesenchymal Transition-Related lncRNA

The list of EMT-related genes (ERGs) was downloaded from the EMT gene database (<http://dbemt.bioinfo-minzhao.org/download.cgi>). Pearson correlation analysis was used to identify EMT-related lncRNAs (correlation coefficient  $|r| > 0.4$  and  $p < 0.05$ ). Then, differentially expressed analysis was performed by comparing the lncRNA expression of tumor samples with normal samples using the “limma” package in the R software (Ritchie et al., 2015), differentially expressed EMT-related lncRNAs would be confirmed when  $FDR < 0.05$  and  $|\log_{2}FC| > 1$ .

### Construction and Validation of Prognostic Signatures

In order to identify differentially expressed EMT-related lncRNAs associated with the overall survival (OS) and disease-free survival (DFS), 533 CRC patients with both OS and DFS information were included. Firstly, all CRC patients were randomly divided into a training set (70%) and a validation set (30%). The training set was used to perform survival analyses and establish prognostic signatures, while training set was used to validate corresponding signature. The univariate Cox analysis was then performed to select OS- and DFS-related differentially expressed EMT-related lncRNAs, and lncRNAs with  $p$  value  $< 0.05$  were listed as candidate prognostic factors. Next, multivariate Cox analysis was used to select the most appropriate OS-related or DFS-related differentially expressed EMT-related lncRNAs and construct two signatures. The corresponding risk scores of each CRC patients in both training and validation sets were calculated simultaneously. The formula is as follows:

$$\text{Risk Score} = \sum_{i=0}^n \beta_i * G_i$$

Here, ' $\beta_i$ ' is the estimated regression coefficient of the gene from the multivariate Cox proportional hazards regression analysis, and ' $G_i$ ' is the expression of the selected gene.

Then, the patients were divided into two risk groups according to the median of risk score. The Kaplan-Meier (K-M) survival analysis was used to analyze the different prognosis between the two groups. Besides, the receiver operating characteristic (ROC) curves at 1-, 3-, and 5-years were used to study the prediction efficiency of signatures. Both signatures were further validated by the validation set. To further understand the discrimination of EMT signatures in difference stage patients, the ROC curves in four stage groups were generated and AUC values at 1-, 3-, and 5-years were calculated.

### Construction of mRNA-lncRNA Network

In order to further understand the regulatory relationship between mRNA and lncRNA, we used the Pearson test to study the interaction between mRNA and lncRNA. When the correlation coefficient  $|r| > 0.4$  and  $p < 0.05$ , it is considered a significant correlation. Cytoscape (V3.7.2) was used to visualize the regulatory networks.

### Evaluation of Correlation With Immune Tumor Microenvironment

To further understand the correlation between EMT-related lncRNAs and immune features in CRC patients, the cluster of CRC cohort was performed by the unsupervised consensus approach by the "Consensus Cluster Plus" package. Because the overlapping lncRNAs between OS- and DFS-related differentially expressed EMT-related lncRNAs are the most highly conserved, we suggested they are the most likely to be associated with the progression of CRC. Therefore, the unsupervised consensus cluster analysis was performed based on the overlapping differentially expressed EMT-related lncRNAs. Meanwhile, the ESTIMATE algorithm was performed to quantify the tumor microenvironment, including immune score, stromal score, and ESTIMATE score (Yoshihara et al., 2013). The Single Sample Gene Set Enrichment Analysis (ssGSEA) algorithm was performed to quantify the fraction of 22 types of immune cells. The associations between clusters and prognosis, clinicopathological variables, and immune cells were also analyzed.

### Exploration of a Key Epithelial-Mesenchymal Transition-Related lncRNA in Colorectal Cancer

In our research, thirteen differentially expressed EMT-related lncRNAs were incorporated into the final signatures, including six to be incorporated into the OS signature and ten to be incorporated into the DFS model. Among them, there were three differentially expressed EMT-related lncRNAs, and they were included in both OS signature and DFS signature. Then, the survival curve of this three lncRNA was plotted and the

**TABLE 1** | Clinicopathologic characteristics of CRC patients.

Characteristics	Whole cohort (n = 533)
Gender	
Male	289
Female	244
Age	
< 65	230
≥ 65	303
Location	
Colon	389
Rectum	144
T stage	
T1-2	113
T3-4	419
Unknown	1
N stage	
N0	316
N1-2	215
Unknown	2
M stage	
M0	408
M1	62
Unknown	63
AJCC	
I-II	301
III-IV	217
Unknown	15

log-rank test was performed. The lncRNA which was most relevant to the prognosis of CRC patients was selected for further exploration.

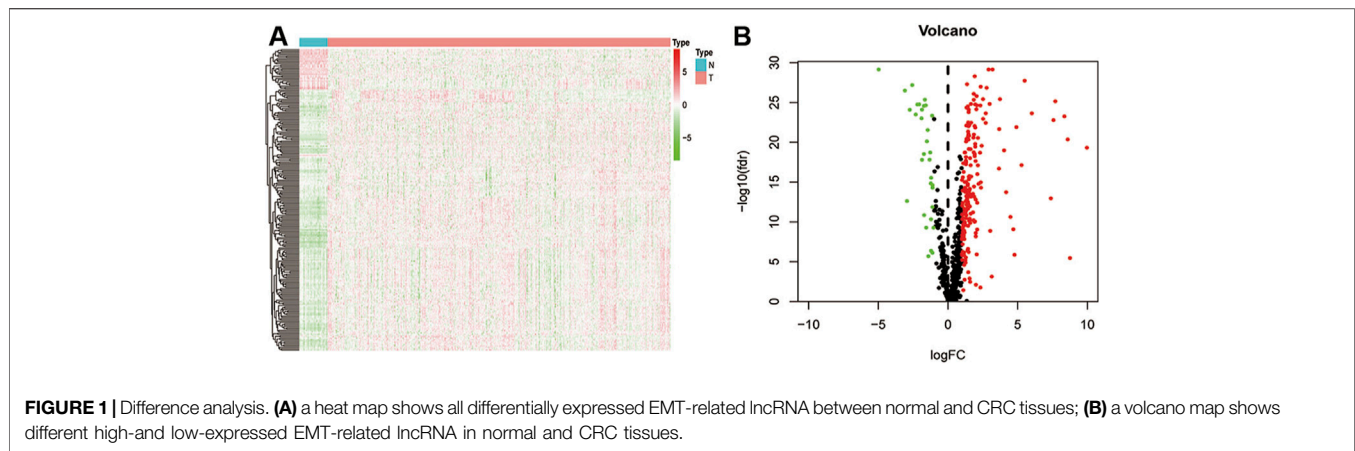
Gene set variation analysis (GSVA) is commonly used to estimate changes in the path and biological process activity in samples of expression datasets (Hänzelmann et al., 2013). To perform GSVA analysis, the gene sets of "c2.cp.kegg.v7.1.-symbols" were obtained from the MSigDB database. Then, we used univariate Cox regression to identify pathways related to prognosis. Pearson correlation analysis was applied to study its relationship with the immune microenvironment and prognosis-related pathways.

## RESULTS

### Overview of Differentially Expressed Epithelial-Mesenchymal Transition-Related lncRNA

Based on the criteria, 616 CRC and 51 adjacent normal samples were included in our research. A total of 533 CRC patients with complete follow-up information, including OS and DFS, were incorporated into the survival analyses. The characteristics and clinical data of these patients are shown in Table 1. As shown in the table, there were more patients in stage I-II and most CRC patients without distant metastases.

By intersecting the EMT gene database and the TCGA database, we obtained the expression of 1,000 EMT-related mRNAs in CRC cohort. Then, 650 lncRNAs were determined as EMT-related mRNAs. To identify the CRC-specific EMT-related lncRNAs, differences between 616 primary CRC tissues



**FIGURE 1** | Difference analysis. **(A)** a heat map shows all differentially expressed EMT-related lncRNA between normal and CRC tissues; **(B)** a volcano map shows different high- and low-expressed EMT-related lncRNA in normal and CRC tissues.

and 51 adjacent normal tissues were compared to identify the differentially expressed EMT-related lncRNAs. Totally, 231 EMT-related lncRNAs were determined, which include 32 downregulated and 199 upregulated in tumors (**Figure 1A** and **Supplementary Table S1**). The volcano map also showed that the majority of EMT-related lncRNAs were upregulated in tumor tissues, indicating that these lncRNAs contribute to the tumorigenesis of CRC (**Figure 1B**).

## Establishment and Validation of Prognostic Signatures

Firstly, a total of 533 CRC patients with complete follow-up information were randomly divided into training and validation sets. A total of 376 patients were incorporated into the training set and 157 patients were incorporated into the validation set. The univariate Cox proportional hazard model was conducted to determine the prognostic EMT-related lncRNAs in the training set. A total of 17 and 34 EMT-related lncRNAs were identified as OS- and DFS-related EMT-related lncRNAs, respectively (**Supplementary Tables S2, S3**). Subsequently, stepwise model selection using the Akaike information criterion (AIC) and multivariate Cox regression model was performed to construct two signatures, and six OS-related EMT-related lncRNAs and ten DFS-related EMT-related lncRNAs were incorporated into the final signatures (**Tables 2 and 3**). Risk scores were calculated based on the selected EMT-related lncRNAs in the final signatures (risk score of OS signature =  $AC127024.4 \times 0.027 + AL591178.1 \times 0.333 + FENDRR \times -0.411 + LINC01315 \times -0.234 + LINC01503 \times 0.359 + MMP25-AS1 \times 0.543$ ; risk score of DFS signature =  $AC021218.1 \times -0.066 + AC074117.1 \times -0.208 + AC104958.2 \times 0.062 + AL031985.3 \times -0.905 + AL033519.3 \times -0.351 + AL591178.1 \times 0.301 + BLACAT1 \times 0.116 + LINC01503 \times 0.245 + MMP25-AS1 \times 0.572 + AC005837.3 \times 0.307$ ).

According to the median risk score, 188 and 188 patients were grouped in the low- and high-risk groups, respectively (**Figures 2A,B**). The K-M survival curves and the log-rank tests showed that the OS and DFS of high-risk patients are significantly worse than that in the low-risk group. The AUC values of OS signature at 1-, 3-, and 5-years were 0.828, 0.789, and 0.730, respectively, while the

**TABLE 2** | OS-related EMT-related lncRNA for CRC patients.

lncRNA	Coef	HR	95% CI	p-value
AL591178.1	0.333	1.395	1.157–1.682	0.000479
LINC01503	0.359	1.432	1.169–1.754	0.00052
MMP25-AS1	0.543	1.722	1.235–2.340	0.001339
AC127024.4	0.027	1.027	1.009–1.045	0.002823
LINC01315	-0.234	0.791	0.669–0.936	0.006149
FENDRR	-0.411	0.663	0.434–1.012	0.056918

**TABLE 3** | DFS-related EMT-related lncRNA for CRC patients.

lncRNA	Coef	HR	95% CI	p-value
AL591178.1	0.301	1.351	1.166–1.565	6.26E-05
MMP25-AS1	0.572	1.772	1.333–2.356	8.27E-05
AL033519.3	-0.351	0.704	0.591–0.839	8.61E-05
AL031985.3	-0.905	0.404	0.251–0.652	0.0002
AC021218.1	-0.066	0.936	0.895–0.980	0.004304
AC104958.2	0.062	1.064	1.017–1.114	0.007646
AC005837.3	0.307	1.359	1.077–1.716	0.009871
LINC01503	0.245	1.278	1.037–1.575	0.021535
BLACAT1	0.116	1.124	1.004–1.258	0.043102
AC074117.1	-0.208	0.812	0.637–1.036	0.093718

corresponding AUC values of DFS signature were 0.765, 0.753, and 0.746. The above AUC values suggested the favorable discrimination of both signatures (**Figures 2A,B**). In the validation set, the K-M survival curves showed that the prognosis of the high-risk group is also significantly worse than that of the low-risk group and that all AUC values were higher than 0.550, indicating the stability of both signatures (**Figures 2C,D**).

To further understand the discrimination of signatures in subgroups, the ROC curves of both OS and DFS signatures for four stage patients are shown in **Supplementary Figure S1**. Overall, the discrimination of both signatures showed favorable performance in most subgroups. However, in stage I patients, the performance of both signatures showed unsatisfactory and the AUC values range from none to 0.873, which can be attributed to the better prognosis in stage I patients and stage I patients have a lower mortality and recurrence rate.

## Two mRNA-lncRNA Networks

To further detect the regulatory network between mRNA and lncRNA, we then constructed the mRNA-lncRNA co-expression network based on the correlation analysis. Six lncRNAs incorporated into the final OS signature and 97 mRNAs containing 105 relationships were obtained to generate a network (Figure 3A). Meanwhile, ten lncRNAs incorporated into the final DFS signature, and 95 mRNAs containing 96 relationships were selected to generate another network (Figure 3B). The in-depth analyses of three lncRNAs that were simultaneously incorporated into OS and DFS signatures revealed that HMGB1 was predicted as the target of AL591178.1, which laid a certain foundation for further research. For the other two EMT-related lncRNAs, 11 and 60 mRNAs were confirmed as the targets of MMP25-AS1 and LINC01503, respectively.

## The Features of Gene Set Variation Analysis and Immune Cell Level

To initially understand the potential mechanisms involved in the biological function of EMT-related lncRNA, the unsupervised clustering analysis was performed. According to the consensus matrix heatmap (Figure 4A), two clusters were identified (C1 ( $n = 205$ , 38.46%), C2 ( $n = 328$ , 61.54%)). The results of K-M analyses showed that the prognosis was significantly different between C1 and C2, including OS (Figure 4B) and DFS (Figure 4C).

In order to clarify the clinicopathological and immune characteristics of EMT-related lncRNA clusters, we firstly explored the relationship between clusters and clinical and

immune status, which is shown in a heatmap (Figure 4D). The distribution of T stage in CRC patients between the two clusters was not random, and there were eight prognosis-related lncRNAs expressed differently between C1 and C2 (Figure 4D). Among these lncRNAs, only LINC01503 was highly expressed in C2, while the others were opposite. Besides, the ESTIMATE score, immune score and stromal score were significantly different between C1 and C2. As shown in Figures 4E,F, the stromal score and immune score of C2 are significantly higher than that of C1. At the same time, by comparing the composition of 22 immune cells, we found that the level of most immune cells is significantly higher in C2, especially immune activated cells, including CD8<sup>+</sup> T cells, natural killer (NK) cells and dendritic cells (DCs) (Figure 4G). The above results indicated that the reason why C2 had a better prognosis than C1 may be related to the expression of different lncRNAs and immune mechanisms. In a word, C2 patients with favorable prognosis showed immune activation but C1 patients with unfavorable prognosis showed immune inactivation relative to immunosuppressive status.

## AL591178.1 is a Key Epithelial-Mesenchymal Transition-Related lncRNA in Colorectal Cancer

In order to find out the biomarker for CRC more accurately, we focused on three EMT-related lncRNAs that were simultaneously incorporated into OS and DFS signatures, including MMP25-AS1, AL591178.1, and LINC01503. The

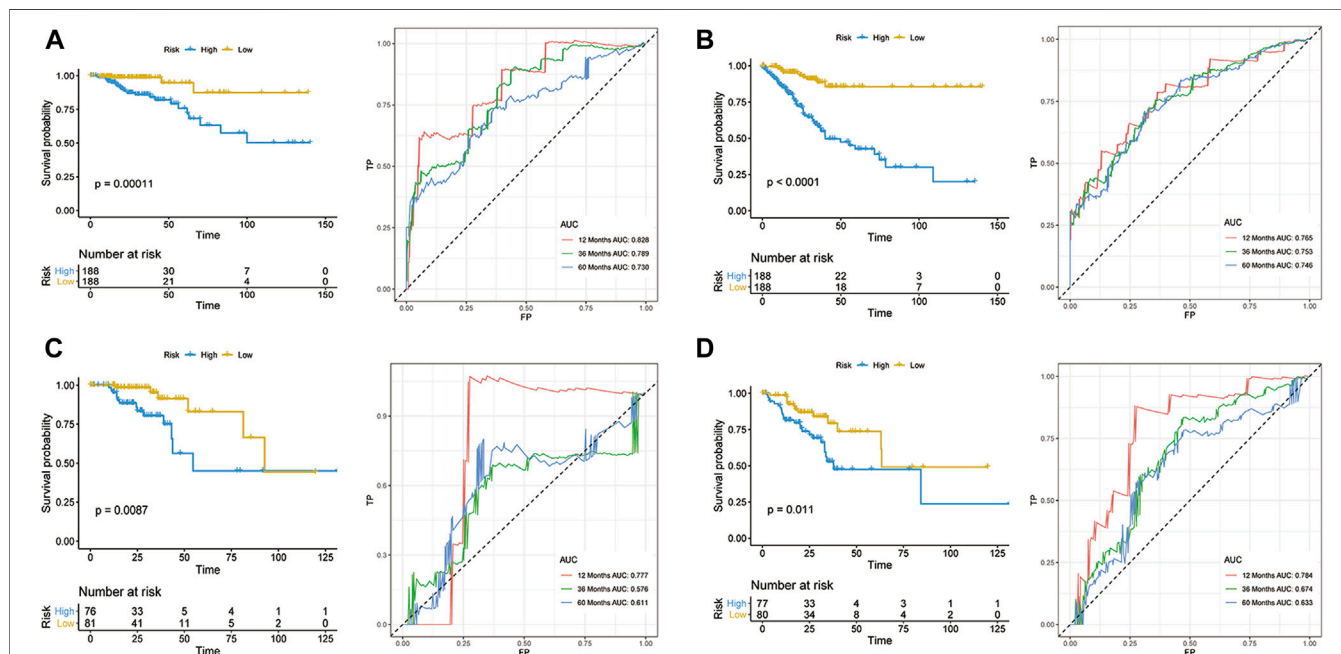
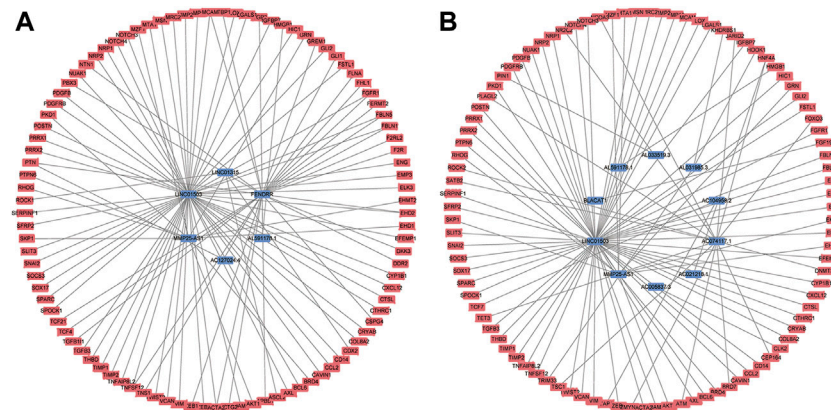


FIGURE 2 | Development of prognostic models for CRC. OS (A,C) and DFS (B,D) in the training set (A,B) and validation set (C,D).



**FIGURE 3** | The mRNA-lncRNA co-expression network based on the correlation analysis. **(A)** the final OS signature; **(B)** the final DFS signature.

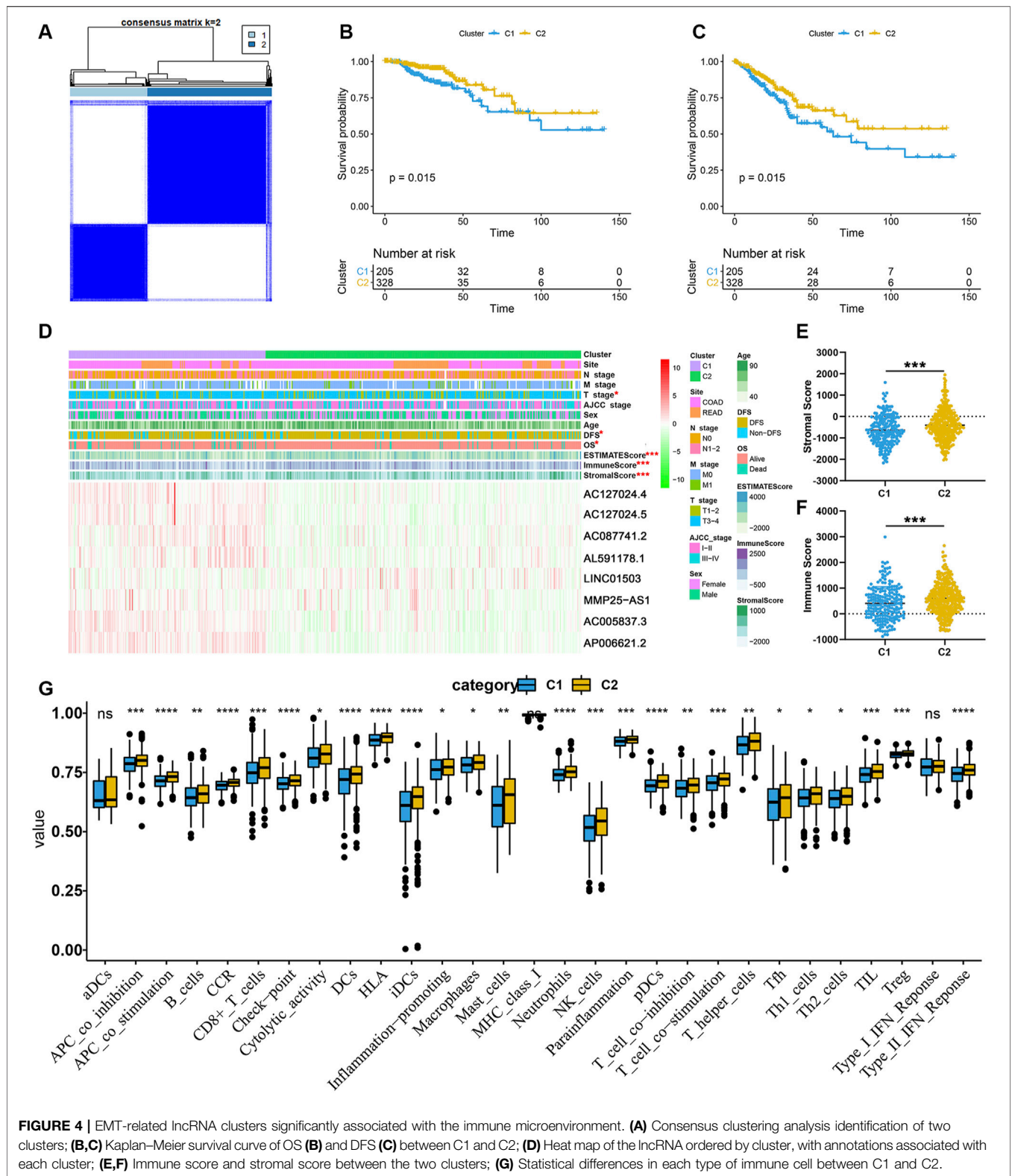
results showed that high expression of AL591178.1 and LINC01503 was associated with poor OS and DFS, but the differences for MMP25-AS1 were not associated with a significant difference ( $p = 0.15$ ) (Figures 5A–F). To further explore the biological function of the three lncRNAs, the univariate Cox analysis of GSVA score was performed. A total of 21 and 48 KEGG pathways were associated with OS and DFS, respectively (Supplementary Tables S4, S5). In these significant pathways, we found that 17 common pathways were related to both OS and DFS (Supplementary Table S6). Further correlation analysis indicated that there is a high degree of correlation between the above three lncRNAs and these 17 pathways. Interestingly, two well-known pathways were detected, including the P53 signaling pathway that was associated with the tumorigenesis and immune formation of CRC and the ECM-receptor interaction pathway that was significantly associated with the formation of the tumor microenvironment (Cho et al., 2020; Yin et al., 2020; Chen Y. et al., 2020; Zou et al., 2019). The correlation plots between lncRNAs and two vital pathways are shown in Figures 6A–F.

According to previous results of our study, EMT-related lncRNA may affect the prognosis of CRC patients through immune mechanisms (Figure 4). Therefore, the lncRNA AL591178.1 was selected, which is most associated with the prognosis of CRC patients (both  $p$  values  $< 0.0001$ ) and negatively associated with the GSVA scores of P53 signaling and ECM-receptor interaction pathways, to further study the association with the specific immune cells. As shown in the radar chart (Figure 7), the expression level is negatively correlated with the existence of immune cells, including CD4 T cells, NK cells, macrophages, and monocytes. In previous studies, we learned that the ECM-receptor interaction pathway hurts the prognosis of cancer patients (Zhang et al., 2020), but our study found that the ECM-receptor interaction pathway is negatively correlated with AL591178.1 (Figure 6D). It is speculated that AL591178.1 inhibits the tumor immune microenvironment, which may cover the prognosis of ECM's effect.

## DISCUSSION

Preliminary studies have shown that ERG and lncRNAs are involved in the progression of many cancers (Zou et al., 2011; Chen et al., 2013; Stavropoulou et al., 2016; Martini et al., 2017). In this study, we established two signatures to predict the OS and DFS of CRC patients, which can stratify CRC patients into significantly different risk groups to effectively predict their prognosis. Besides, the results of unsupervised consensus showed that there are higher immune score and a higher level of immune cells in C2 with a better prognosis compared with C1. In OS and DFS signatures, there are three common lncRNAs, and their expression is significantly related to prognosis. AL591178.1, as a prognostic risk factor, was found to be negatively correlated with the level of immune cells and ECM-receptor pathway. To the best of our knowledge, it is the first study to focus on the characteristics of CRC from the perspective of EMT-related lncRNAs, so as to provide a larger field of vision and choice for such patients' prediction and molecular targeting.

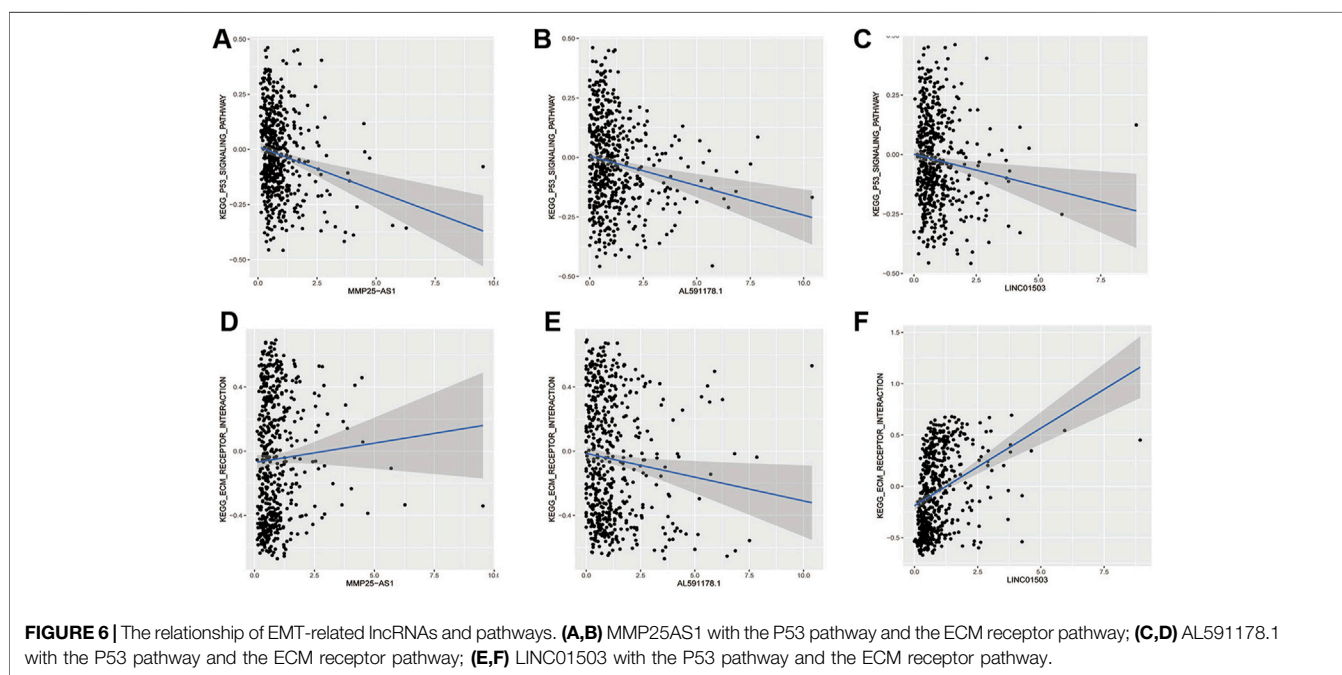
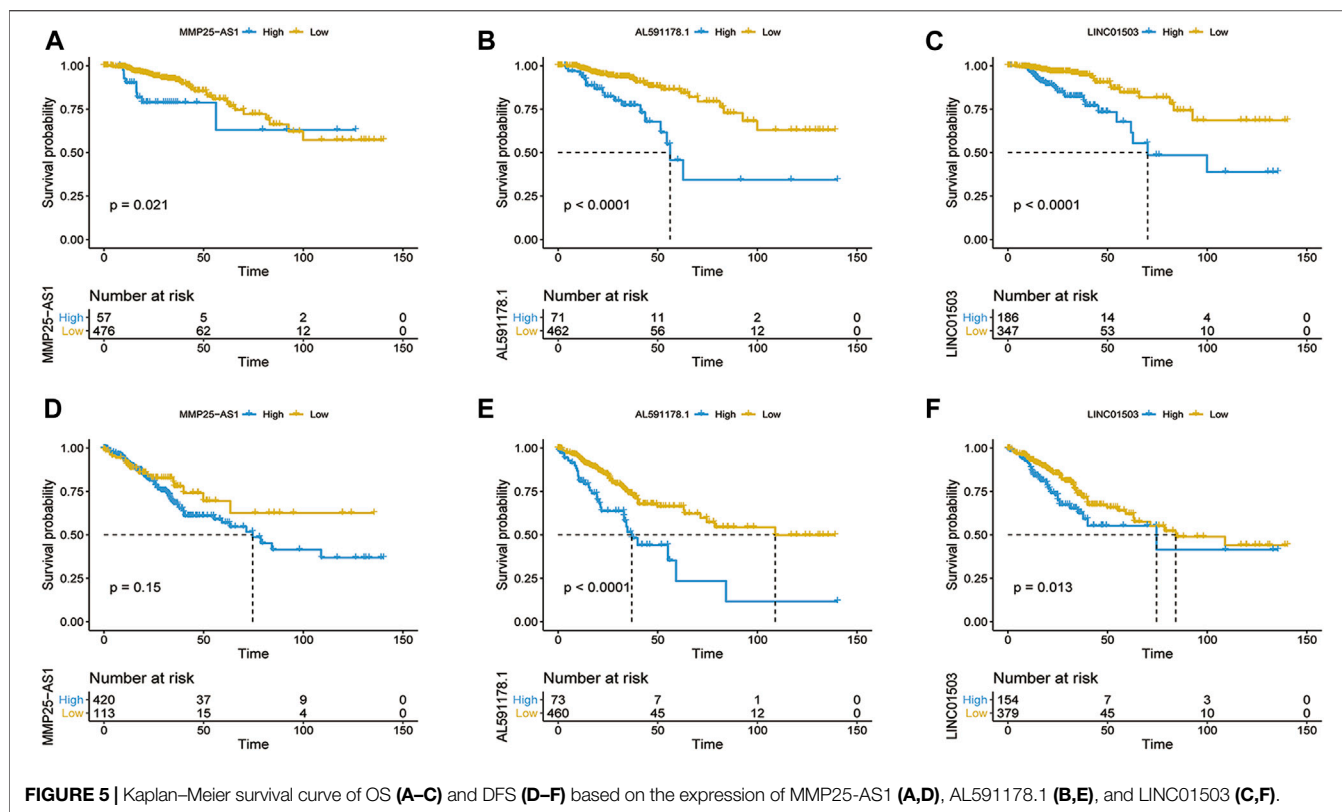
Normal tissues can precisely regulate the stability of genome expression and maintain their normal physiological functions. However, abnormal development of the genome is a typical feature of tumor cells, including lncRNA (Hanahan and Weinberg, 2011). In recent years, with the advancement of high-throughput sequencing technology, lncRNA, as an important genome for maintaining functions in the human body, has gradually gained attention. Our study used more common analysis methods to explore the prognostic biomarkers of CRC patients from a perspective of EMT-related lncRNA. And the results proved that our findings are accurate and reliable. For example, LINC01503 was confirmed to be correlated with the OS of CRC patients, and other studies have also found that it is highly expressed in different tumor cells, and is positively correlated with tumor cell proliferation and metastasis, including nasopharyngeal carcinoma (He et al., 2020), gastric cancer (Ding et al.,



**FIGURE 4 |** EMT-related lncRNA clusters significantly associated with the immune microenvironment. **(A)** Consensus clustering analysis identification of two clusters; **(B,C)** Kaplan-Meier survival curve of OS **(B)** and DFS **(C)** between C1 and C2; **(D)** Heat map of the lncRNA ordered by cluster, with annotations associated with each cluster; **(E,F)** Immune score and stromal score between the two clusters; **(G)** Statistical differences in each type of immune cell between C1 and C2.

2020; Ma et al., 2020), and CRC (Wei et al., 2020). Interestingly, in cholangiocarcinoma, LINC01503 can also promote the EMT process (Qu et al., 2019), confirming that this lncRNA is an EMT-related lncRNA. Our study found that

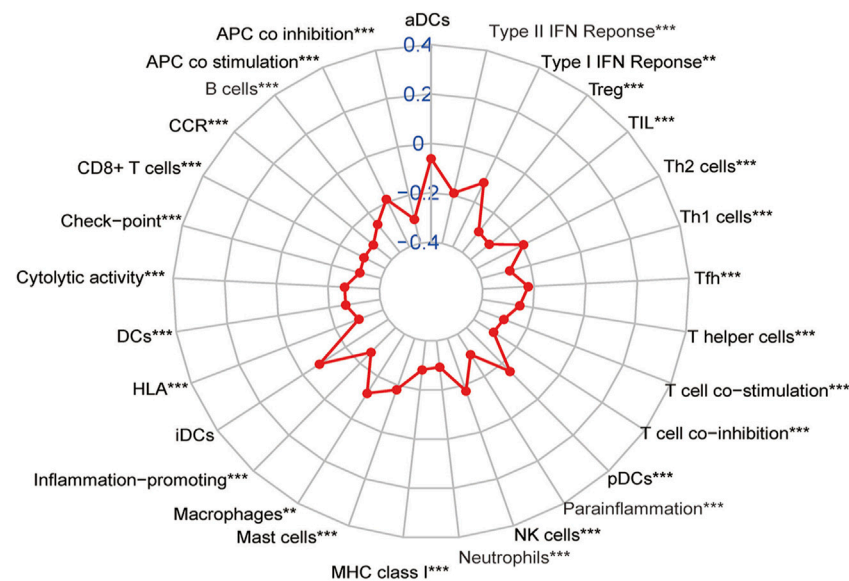
LINC01871 is a protective factor for the prognosis of CRC. He et al. (He and Wang, 2020) found that its expression level is closely related to CD8<sup>+</sup> T cell enrichment levels, PD-L1 expression levels, and immune cytolitic activity, which



means that LINC01871 may inhibit the growth of CRC cells by regulating the formation of immune microenvironment. Therefore, the above studies all showed the reliability of our

results. Besides, some of the lncRNAs, contained in the two signatures were not reported in previous studies, such as AC127024.4 and AL033519.3, which illustrates the





**FIGURE 7 |** Association between 22 types of immune cells and the expression of AL591178.1.

innovation of our study to a certain extent, and provides a theoretical basis for further research. In addition to lncRNAs, the corresponding targeting mRNA also participate immune regulation. Yang et al. (Yang et al., 2020) found that lipopolysaccharide induced the release of pro-inflammatory cytokines in a HMGB1-dependent manner to improve colon cancer progression. The discovery proved the immune role of the lncRNA and mRNA we identified.

In our results, there are differences in the levels of immune cells in C1 and C2, indicating that EMT-related lncRNA may affect the prognosis of patients through immune mechanisms. Previous studies have also confirmed it (Wang et al., 2019; Chen Q.-H. et al., 2020). The expression of lncRNA SNHG14 is up-regulated in diffuse large B cells, and its deletion will hinder the proliferation, migration, and EMT (Zhao et al., 2019). Mechanistically, SNHG14 up-regulates zinc finger E-box binding homeobox 1 (ZEB1), thereby activating PD-L1 to promote immune evasion of diffuse large B cells (Zhao et al., 2019). In order to further clarify the role of EMT-related lncRNA we found in CRC, we took the intersection of the two established signatures and found three lncRNAs related to both OS and DFS. Among them, we further focused on AL591178.1, which was negatively correlated with the infiltration level of most immune cells. The tumor immune microenvironment is an important regulatory factor of tumor development, such as CD4 T cells, CD8 T cells, and NK cells. Previous studies have found that they all have a positive impact on the prognosis of cancers (Ahlén Bergman et al., 2018; Wang et al., 2018; Wang D. et al., 2020). Therefore, we can speculate that AL591178.1 may affect the prognosis of CRC patients by regulating the abundance of immune cells. According to the GSVA function analysis and survival correlation analysis, we found that the two

pathways are significantly related to the expression of AL591178.1. The P53 pathway plays a role in inhibiting EMT and invasion of cancer and is beneficial to the prognosis of patients (Wu et al., 2019; Jackson-Weaver et al., 2020), so the results of its negative correlation with the expression of AL591178.1 are consistent with their conclusion. Interestingly, the ECM-receptor pathway has been considered to promote tumorigenesis in previous studies and is negatively correlated with the prognosis of cancer patients (Zhang et al., 2016; Wang J. et al., 2020), so we speculated that AL591178.1 may suppress the tumor immune microenvironment and suppress immune function. But the mechanism involved needs further research to clarify.

Although our research is somewhat innovative, some limitations require attention. First, no independent cohort of CRC patients was used to prove that the prognosis model presented here is reproducible. Second, this study is a retrospective study, which has certain biases, and the clinical data included is not rich enough. Third, although our research has found a potential way to interfere with the prognosis of CRC patients, further experiments are needed to verify and find the corresponding target.

## CONCLUSION

The results we obtained highlighted the prognostic value of EMT-related lncRNA in CRC patients and explored potential mechanisms and regulatory networks. In detail, AL591178.1 was found to be a risk factor for the prognosis of CRC patients, which is inversely proportional to the level of immune cell infiltration, P53 pathway, and ECM-receptor pathway. Thus, this study identified a novel potential target to provide treatment opportunities for CRC patients.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

CH and CZ conceived and designed the study. CL wrote the paper and revised the manuscript. JL performed the analysis.

## REFERENCES

- Ahlén Bergman, E., Hartana, C. A., Johansson, M., Linton, L. B., Berglund, S., Hyllienmark, M., et al. (2018). Increased CD4+ T cell lineage commitment determined by CpG methylation correlates with better prognosis in urinary bladder cancer patients. *Clin. Epigenet* 10 (1), 102. doi:10.1186/s13148-018-0536-6
- Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., et al. (2018). EMT subtype influences epithelial plasticity and mode of cell migration. *Dev. Cel.* 45 (6), 681–695. doi:10.1016/j.devcel.2018.05.027
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.cd-12-0095
- Chen, C. L., Mahalingam, D., Osmulski, P., Jadhav, R. R., Wang, C. M., Leach, R. J., et al. (2013). Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer. *Prostate* 73 (8), 813–826. doi:10.1002/pros.22625
- Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer statistics in China, 2015. *CA: a Cancer J. Clin.* 66 (2), 115–132. doi:10.3322/caac.21338
- Chen, Q.-H., Li, B., Liu, D.-G., Zhang, B., Yang, X., and Tu, Y.-L. (2020). lncRNA KCNQ1OT1 sponges miR-15a to promote immune evasion and malignant progression of prostate cancer via up-regulating PD-L1. *Cancer Cel. Int.* 20, 394. doi:10.1186/s12935-020-01481-8
- Chen, W., Tu, Q., Yu, L., Xu, Y., Yu, G., Jia, B., et al. (2020). lncRNA ADAMTS9-AS1, as prognostic marker, promotes cell proliferation and EMT in colorectal cancer. *Hum. Cel.* 33 (4), 1133–1141. doi:10.1007/s13577-020-00388-w
- Chen, Y., Huang, Y., Gao, X., Li, Y., Lin, J., Chen, L., et al. (2020). CCND1 amplification contributes to immunosuppression and is associated with a poor prognosis to immune checkpoint inhibitors in solid tumors. *Front. Immunol.* 11, 1620. doi:10.3389/fimmu.2020.01620
- Cho, Y.-H., Ro, E. J., Yoon, J.-S., Mizutani, T., Kang, D.-W., Park, J.-C., et al. (2020). 5-FU promotes stemness of colorectal cancer via p53-mediated WNT/ $\beta$ -catenin pathway activation. *Nat. Commun.* 11 (1), 5321. doi:10.1038/s41467-020-19173-2
- Cunningham, D., Atkin, W., Lenz, H.-J., Lynch, H. T., Minsky, B., Nordlinger, B., et al. (2010). Colorectal cancer. *Lancet* 375 (9719), 1030–1047. doi:10.1016/S0140-6736(10)60353-4
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22 (9), 1775–1789. doi:10.1101/gr.132159.111
- Ding, J., Shi, F., Xie, G., and Zhu, Y. (2020). Long non-coding RNA LINC01503 promotes gastric cancer cell proliferation and invasion by regulating wnt signaling. *Dig. Dis. Sci.* 66, 452. doi:10.1007/s10620-020-06215-4
- Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21 (11), 1350–1356. doi:10.1038/nm.3967
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

LJ reviewed and edited the manuscript. CZ gave some advice for the revised manuscript. All authors read and approved the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.633951/full#supplementary-material>.

- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7
- Hay, E. D. (2005). The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Dev. Dyn.* 233 (3), 706–720. doi:10.1002/dvdy.20345
- He, S.-W., Xu, C., Li, Y.-Q., Li, Y.-Q., Zhao, Y., Zhang, P.-P., et al. (2020). AR-induced long non-coding RNA LINC01503 facilitates proliferation and metastasis via the SFPQ-FOSL1 axis in nasopharyngeal carcinoma. *Oncogene* 39 (34), 5616–5632. doi:10.1038/s41388-020-01388-8
- He, Y., and Wang, X. (2020). Identification of molecular features correlating with tumor immunity in gastric cancer by multi-omics data analysis. *Ann. Transl. Med.* 8 (17), 1050. doi:10.21037/atm-20-922
- Huang, L., Chen, S., Fan, H., Ji, D., Chen, C., and Sheng, W. (2020). GINS2 promotes EMT in pancreatic cancer via specifically stimulating ERK/MAPK signaling. *Cancer Gene Ther.* [Epub ahead of print]. doi:10.1038/s41417-020-0206-7
- Jackson-Weaver, O., Ungvijanpunya, N., Yuan, Y., Qian, J., Gou, Y., Wu, J., et al. (2020). PRMT1-p53 pathway controls epicardial EMT and invasion. *Cell Rep.* 31 (10), 107739. doi:10.1016/j.celrep.2020.107739
- Kalluri, R., and Neilson, E. G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. *J. Clin. Invest.* 112 (12), 1776–1784. doi:10.1172/jci2053010.1172/jci200320530
- Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119 (6), 1420–1428. doi:10.1172/jci39104
- Liang, H., Yu, T., Han, Y., Jiang, H., Wang, C., You, T., et al. (2018). lncRNA PTAR promotes EMT and invasion-metastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression. *Mol. Cancer* 17 (1), 119. doi:10.1186/s12943-018-0870-5
- Ma, Z., Gao, X., Shuai, Y., Wu, X., Yan, Y., Xing, X., et al. (2020). EGFR-mediated linc01503 promotes cell cycle progression and tumorigenesis in gastric cancer. *Cel. Prolif.* 54, e12922. doi:10.1111/cpr.12922
- Martini, P., Paracchini, L., Caratti, G., Mello-Grand, M., Fruscio, R., Beltrame, L., et al. (2017). lncRNAs as novel indicators of patients' prognosis in stage I epithelial ovarian cancer: A retrospective and multicentric study. *Clin. Cancer Res.* 23 (9), 2356–2366. doi:10.1158/1078-0432.ccr-16-1402
- Montemayor-Garcia, C., Hardin, H., Guo, Z., Larrain, C., Buehler, D., Asioli, S., et al. (2013). The role of epithelial mesenchymal transition markers in thyroid carcinoma progression. *Endocr. Pathol.* 24 (4), 206–212. doi:10.1007/s12022-013-9272-9
- Nacev, B. A., Feng, L., Bagert, J. D., Lemiesz, A. E., Gao, J., Soshnev, A. A., et al. (2019). The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* 567 (7749), 473–478. doi:10.1038/s41586-019-1038-1
- Nieto, M. A., Huang, R. Y.-J., Jackson, R. A., and Thiery, J. P. (2016). EMT: 2016. *Cell* 166 (1), 21–45. doi:10.1016/j.cell.2016.06.028
- Qu, Y. K., Qu, X. S., Chen, G., Feng, Y., Teng, X. L., Liu, W. X., et al. (2019). LINC01503 promotes cell proliferation, invasion and EMT process in cholangio-carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 23 (15), 6445–6452. doi:10.26355/eurrev\_201908\_1852610.1111/jcmm.14062
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

- Saitoh, M. (2018). Involvement of partial EMT in cancer progression. *J. Biochem.* 164 (4), 257–264. doi:10.1093/jb/mvy047
- Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173 (2), 321–337. doi:10.1016/j.cell.2018.03.035
- Stavropoulou, V., Kaspar, S., Brault, L., Sanders, M. A., Juge, S., Moretini, S., et al. (2016). MLL-AF9 expression in hematopoietic stem cells drives a highly invasive AML expressing EMT-related genes linked to poor outcome. *Cancer Cell* 30 (1), 43–58. doi:10.1016/j.ccell.2016.05.011
- Wang, D., Yu, W., Lian, J., Wu, Q., Liu, S., Yang, L., et al. (2020). Th17 cells inhibit CD8+ T cell migration by systematically downregulating CXCR3 expression via IL-17A/STAT3 in advanced-stage colorectal cancer patients. *J. Hematol. Oncol.* 13 (1), 68. doi:10.1186/s13045-020-00897-z
- Wang, J., Zhang, Q., Li, S., Chen, Z., Tan, J., Yao, J., et al. (2020). Low molecular weight fucoidan alleviates diabetic nephropathy by binding fibronectin and inhibiting ECM-receptor interaction in human renal mesangial cells. *Int. J. Biol. macromolecules* 150, 304–314. doi:10.1016/j.ijbiomac.2020.02.087
- Wang, Q.-M., Lian, G.-Y., Song, Y., Huang, Y.-F., and Gong, Y. (2019). lncRNA MALAT1 promotes tumorigenesis and immune escape of diffuse large B cell lymphoma by sponging miR-195. *Life Sci.* 231, 116335. doi:10.1016/j.lfs.2019.03.040
- Wang, W.-T., Zhu, H.-Y., Wu, Y.-J., Xia, Y., Wu, J.-Z., Wu, W., et al. (2018). Elevated absolute NK cell counts in peripheral blood predict good prognosis in chronic lymphocytic leukemia. *J. Cancer Res. Clin. Oncol.* 144 (3), 449–457. doi:10.1007/s00432-017-2568-2
- Wei, J., Ge, X., Tang, Y., Qian, Y., Lu, W., Jiang, K., et al. (2020). An autophagy-related long noncoding RNA signature contributes to poor prognosis in colorectal cancer. *J. Oncol.* 2020, 4728947. doi:10.1155/2020/4728947
- Wu, Q., Ma, J., Wei, J., Meng, W., Wang, Y., and Shi, M. (2020). lncRNA SNHG11 promotes gastric cancer progression by activating Wnt/ $\beta$ -catenin pathway and oncogenic autophagy. *Mol. Ther.* doi:10.1016/j.ymthe.2020.10.011
- Wu, Z. H., Liu, C. C., Zhou, Y. Q., Hu, L. N., and Guo, W. J. (2019). OnclncRNA-626 promotes malignancy of gastric cancer via inactivated the p53 pathway through interacting with SRSF1. *Am. J. Cancer Res.* 9 (10), 2249–2263.
- Xie, Q., Tang, T., Pang, J., Xu, J., Yang, X., Wang, L., et al. (2020). LSD1 promotes bladder cancer progression by upregulating LEF1 and enhancing EMT. *Front. Oncol.* 10, 1234. doi:10.3389/fonc.2020.01234
- Yang, Y., Yang, L., Jiang, S., Yang, T., Lan, J., Lei, Y., et al. (2020). HMGB1 mediates lipopolysaccharide-induced inflammation via interacting with GPX4 in colon cancer cells. *Cancer Cel. Int* 20, 205. doi:10.1186/s12935-020-01289-6
- Yin, H., Xie, J., Jiang, P., Jiang, X., Duan, D., Qi, J., et al. (2020). KRASchAuranib selectively inhibits colorectal cancer with wild-type by modulation of ROS through activating the p53 signaling pathway. *Am. J. Cancer Res.* 10 (11), 3666–3685. doi:10.21203/rs.3.rs-46779/v1
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612
- Zeisberg, M., and Neilson, E. G. (2009). Biomarkers for epithelial-mesenchymal transitions. *J. Clin. Invest.* 119 (6), 1429–1437. doi:10.1172/jci36183
- Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m(6A) regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0
- Zhang, H.-j., Tao, J., Sheng, L., Hu, X., Rong, R. M., Xu, M., et al. (2016). RETRACTED: Twist2 promotes kidney cancer cell proliferation and invasion via regulating ITGA6 and CD44 expression in the ECM-Receptor-Interaction pathway. *Biomed. Pharmacother.* 81, 453–459. doi:10.1016/j.biopha.2016.02.042
- Zhao, L., Li, W., Zang, W., Liu, Z., Xu, X., Yu, H., et al. (2013). JMJD2B promotes epithelial-mesenchymal transition by cooperating with  $\beta$ -catenin and enhances gastric cancer metastasis. *Clin. Cancer Res.* 19 (23), 6419–6429. doi:10.1158/1078-0432.ccr-13-0254
- Zhao, L., Liu, Y., Zhang, J., Liu, Y., and Qi, Q. (2019). lncRNA SNHG14/miR-5590-3p/ZEB1 positive feedback loop promoted diffuse large B cell lymphoma progression and immune evasion through regulating PD-1/PD-L1 checkpoint. *Cel. Death Dis.* 10 (10), 731. doi:10.1038/s41419-019-1886-5
- Zhuang, L., Ding, W., Ding, W., Zhang, Q., Xu, X., and Xi, D. (2020). lncRNA ZNF667-AS1 (NR\_036521.1) inhibits the progression of colorectal cancer via regulating ANK2/JAK2 expression. *J. Cel. Physiol* 236, 2178. doi:10.1002/jcp.30004
- Zou, J., Luo, H., Zeng, Q., Dong, Z., Wu, D., and Liu, L. (2011). Protein kinase CK2 $\alpha$  is overexpressed in colorectal cancer and modulates cell proliferation and invasion via regulating EMT-related genes. *J. transl. Med.* 9, 97. doi:10.1186/1479-5876-9-97
- Zou, Y., Zheng, S., Deng, X., Yang, A., Xie, X., Tang, H., et al. (2019). The role of circular RNA CDR1as/ciRS-7 in regulating tumor microenvironment: a pan-cancer analysis. *Biomolecules* 9 (9), 429. doi:10.3390/biom9090429

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Liu, Hu, Li, Jiang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.