scientific reports

Check for updates

OPEN

Identification of progression related LncRNAs in colorectal cancer aggressiveness

Wei Xi[1], Xinxin Sun[2], Mingwei Wang[3], Xizi Wang[2], Kun Li[2], Runze Jiang[3], Xiaodong Jia[2] & Wenxiao Wang[4]✉

Colorectal cancer (CRC) progression involves complex molecular alterations, including the dysregulation of long non-coding RNAs (lncRNAs). In this study, we identified key progression-related lncRNAs in CRC by integrating transcriptomic data from TCGA and single-cell RNA sequencing (scRNA-seq). Differential expression analysis revealed numerous lncRNAs associated with CRC progression. To systematically prioritize these lncRNAs, we developed a scoring system incorporating multiple progression-related signatures, differential expression, and survival analysis. This approach identified 198 key lncRNAs, including both known (e.g., LINC01615) and novel candidates (e.g., AC007998.3). Experimental validation confirmed that LINC01615 was significantly upregulated in CRC tissues, whereas AC007998.3 was downregulated. Further analyses indicated that these lncRNAs influence CRC progression through cis-, trans-, and post-transcriptional regulation. Patients were classified into distinct molecular subgroups based on lncRNA expression, exhibiting significant differences in prognosis and immune microenvironment composition. The enrichment of progression-related lncRNAs among differentially expressed lncRNAs was statistically significant, reinforcing their functional relevance. Validation across independent datasets demonstrated the robustness of our findings. Our research provides novel insights into the molecular mechanisms underlying CRC progression and highlights the potential of progression-related lncRNAs as prognostic biomarkers and therapeutic targets.

Keywords Colorectal cancer, LncRNAs, Cancer progression, Tumor microenvironment, Immune regulation, Immune infiltration

Colorectal cancer (CRC) is the third typical cancer, and the second biggest cause of cancer-related deaths worldwide[1,2]. Despite advancements in screening, early detection, and multimodal treatment strategies, the prognosis for advanced-stage CRC remains poor[3]. This underscores the need for a deeper understanding of the genetic mechanisms behind CRC progression, enabling the creation of more effective, tailored therapy[4]. Over the last decade, long non-coding RNAs (lncRNAs) have emerged as significant regulators of gene expression in numerous cancers, including CRC[5]. Unlike protein-coding genes, lncRNAs are a diverse class of transcripts longer than 200 nucleotides that function primarily through transcriptional, post-transcriptional, and epigenetic mechanisms[6,7]. Their roles in tumor biology are increasingly recognized, influencing processes such as tumor growth, metastasis, drug resistance, and immune evasion[8,9].

The clinical and molecular heterogeneity of CRC poses significant challenges for personalized treatment[10,11]. This heterogeneity is shaped by both intrinsic factors, such as genetic and epigenetic alterations, and extrinsic factors, including interactions with the tumor microenvironment[12,13]. While prior research has identified key protein-coding genes involved in CRC progression, the contributions of non-coding RNAs, particularly lncRNAs, remain less understood. Emerging data reveals that lncRNAs play critical roles in coordinating multi-level regulatory networks in cancer, encompassing cis-regulation of nearby genes, trans-regulation of distal targets, and post-transcriptional interactions with microRNAs (miRNAs) and competing endogenous RNAs (ceRNAs)[14,15]. These multifaceted mechanisms allow lncRNAs to orchestrate complex gene expression programs, making them attractive targets for therapeutic intervention[16,17].

[1]Department of Oncology, Liaocheng People's Hospital, Liaocheng 252000, Shandong, China. [2]Joint Laboratory for Translational Medicine Research, Liaocheng People's Hospital, Liaocheng 252000, Shandong, China. [3]Traditional Chinese Medicine Innovation Research Institute, Shandong University of Traditional Chinese Medicine, Jinan 250035, Shandong, China. [4]Department of Gastrointestinal Surgery, Liaocheng People's Hospital, Liaocheng 252000, Shandong, China. ✉email: Doctorwang427@163.com

Despite these advances, several knowledge gaps remain. First, the systematic identification and prioritization of progression-related lncRNAs in CRC have been limited. Second, the functional roles of lncRNAs in shaping the tumor microenvironment, particularly their interactions with immune and stromal cells, require further investigation. Finally, the integration of multi-level regulatory mechanisms, including cis-, trans-, and post-transcriptional interactions, has not been comprehensively explored in the context of CRC. Addressing these gaps is crucial for understanding the broader biological impact of lncRNAs in CRC and their potential as diagnostic and therapeutic targets.

In this study, we identified and prioritized progression-related lncRNAs in CRC by using transcriptomic data from The Cancer Genome Atlas (TCGA). Furthermore, we explored the associations between prioritized lncRNAs and tumor microenvironment features, including immune and stromal cell infiltration, to uncover their contributions to CRC progression. By applying this framework, we identified key lncRNAs associated with distinct molecular subtypes of CRC, highlighting their roles in tumor invasion, immune modulation, and metastasis. We further systematically analyzed the roles of lncRNAs in cis-, trans-, and post-transcriptional regulation. This comprehensive approach provides a foundation for future studies aimed at validating these lncRNAs as biomarkers and therapeutic targets, with the ultimate goal of improving personalized treatment strategies for CRC.

Materials and methods

Data acquisition and preprocessing

The RNA-seq data for the present research were obtained from TCGA database (https://portal.gdc.cancer.gov/). Specifically, the datasets included colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) RNA-seq data, which were downloaded in the form of fragments per kilobase of transcript per million mapped reads (FPKM). The clinical data associated with the RNA-seq datasets, including survival information, tumor stage, and subtype classification, were also obtained from TCGA. To eliminate potential batch effects between COAD and READ datasets, we applied the 'ComBat' algorithm from the sva (v3.48.0) R package. The gene annotation file for identifying lncRNAs was obtained from GENCODE v38.

Framework of identifying progression-related LncRNAs

The framework involves three major steps[18]:

(I) Integration of lncRNA and mRNA expression profiles from CRC datasets. To comprehensively analyze the role of lncRNAs in CRC progression, we integrated transcriptomic data from TCGA, specifically COAD and READ datasets after batch removement. This step ensured that the integration of different datasets did not introduce technical biases.

(II) Calculation of relationships between lncRNAs and mRNAs through partial correlation correcting tumor purity and ranking mRNAs. Given that tumor purity can influence correlation analyses, we applied partial correlation analysis, adjusting for tumor purity using ESTIMATE algorithm. Partial correlation coefficients were computed to assess the direct association between lncRNAs and mRNAs while controlling for tumor purity. Significant correlations were defined based on false discovery rate (FDR) < 0.05 to account for multiple testing corrections. All mRNAs were ranked based on their correlation strength with progression-related lncRNAs.

(III) Functional enrichment using gene set enrichment analysis (GSEA) to link candidate lncRNAs with CRC progression-related gene signatures. To determine the functional roles of candidate lncRNAs, we performed GSEA to identify enriched biological pathways and gene signatures associated with CRC progression. The ranked list of mRNAs associated with progression-related lncRNAs was subjected to GSEA. Enrichment scores were calculated to identify biological processes and pathways significantly associated with lncRNA-regulated genes.

Identification of differentially expressed LncRNAs

The limma (v 3.56.2) R software was used to perform differential expression analysis between normal and tumorous tissues. LncRNAs with an FDR < 0.05 [false discovery rate; Benjamini-Hochberg (BH) method] and an absolute $\log_2$ fold-change > 1 were considered differentially expressed lncRNA.

Prioritization of progression-related LncRNAs

Progression-related lncRNAs were prioritized by integrating multiple features, including (i) their enrichment in cancer progression-related gene signatures, (ii) their association with survival outcomes (hazard ratio and p-value from Cox proportional hazards regression analysis), and (iii) their differential expression between normal and tumor tissues. In addition, low-expression lncRNA was excluded. The rules to follow are as follows: (i) eliminating any lncRNA with a 50 th percentile FPKM equal to 0; (ii) selecting only those lncRNAs whose 90 th percentile FPKM value is greater than 0.1[19].

Tumor microenvironment analysis

The tumor microenvironment was characterized by estimating the proportions of immune and stromal cells using the CIBERSORTx algorithms[20]. Associations between prioritized lncRNAs and specific cell types were evaluated using Spearman correlation (Spearman's $r > 0.3$, FDR < 0.05) and Fisher's exact tests. We uploaded gene expression data to the online site using TR4[21,22] and LM22 signature matrices. TR4 produces epithelial cells, endothelial cells, fibroblasts, and somatic immune cells with EPCAM, CD31, CD10, and CD45, respectively.

Stratification of patients based on prioritized progression-related LncRNAs

For patient stratification based on progression-related lncRNAs, we employed consensus clustering using the ConsensusClusterPlus (v1.64.0)[23] R package. This method categorizes patients into subgroups based on the expression patterns of the top 10 lncRNAs linked with CRC development. The clustering method was pam, and the distance parameter was pearson. Consensus clustering was performed with 1000 resampling iterations and the k-means algorithm, testing cluster numbers (k) from 2 to 6. The optimal number of clusters was identified through evaluation of the consensus matrix and silhouette scores, ensuring the stability and biological relevance of the patient subgroups.

Analysis of cis-regulatory interactions

Cis-regulatory interactions were identified by analyzing lncRNAs and mRNAs located within a 1 Mb genomic distance. Spearman correlation analysis was implemented to determine the co-expression correlations between lncRNAs and their neighboring genes. Significant interactions were defined as those with an absolute correlation coefficient > 0.3 and an FDR < 0.05.

Construction of trans-regulatory networks

Trans-regulatory networks were constructed based on significant correlations between lncRNAs and distal mRNAs. Spearman correlations were calculated across all samples, and significant interactions (absolute correlation coefficient > 0.3, FDR < 0.05) were included in the network.

Construction of CeRNA networks

Negative correlations (Spearman correlation < −0.3, FDR < 0.05) between lncRNAs and miRNAs, as well as miRNAs and mRNAs, were used to identify ceRNA interactions. Positive correlations (Spearman correlation > 0.3, FDR < 0.05) between lncRNAs and mRNAs were further integrated to construct the ceRNA network.

Sample collection and realtime fluorescence quantitative PCR (RT-qPCR)

Four samples of CRC tumor tissues and matched adjacent normal tissues were collected from March 12, 2025 to March 16, 2025 at Liaocheng People's Hospital. Samples were placed in sterilized centrifuge tubes and quickly transferred to −80 °C for storage. This study was approved by the medical ethics committee of Liaocheng People's Hospital (No: 2025063, 12 March 2025) and strictly followed the guidelines of the Declaration of Helsinki. All involved patients had signed informed consent for participating in this study.

Total RNA of CRC and adjacent normal tissues was extracted using the RNA extraction kit (QIAGEN, German). cDNA synthesis was then performed using HiScript III RT SuperMix kit (Vazyme, China). Finally, ChamQ Universal SYBR qPCR Master Mix kit (Vazyme, China) and primers of corresponding genes (as follows) were used for RT-qPCR reaction. The relative lncRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method, with GAPDH as the reference gene. Differences between relative expression values of different groups were tested using two-sided Wilcoxon's rank-sum test. The primer sequences were as follows:

LINC01615-F: 5' – CTGGCTCCGTCAGAGGATGGG – 3';
LINC01615-R: 5' – ACGCAGGCTTGTTGTTGGATGG – 3';
AC007998.3-F: 5' – CCAAGCAGCACAGGAACATAGTC – 3';
AC007998.3-R: 5' – GAAGGAAAGCGAGGGCACATC – 3';
LINC00460-F: 5' – CTTTCCCACGCAGTGGATGA – 3';
LINC00460-R: 5' – GAATGCGTCTTCTTTCCCACG – 3';
GAPDH-F: 5' – GGAGCGAGATCCCTCCAAAAT – 3';
GAPDH-R: 5' – GGCTGTTGTCATACTTCTCATGG – 3';

Comprehensive summary of tools

Category	Database/Tool/Package	Version
Organ and tissue model	Biorender	-
RNA-seq data acquisition	TCGA GDC portal; GEO	-
RNA-seq batch effect correction	sva (R package)	3.48.0
Partial correlation analysis	ppcor (R package)	1.1
Tumor purity	estimate (R package)	1.0.13
Differential expression	limma (R package)	3.56.2
Pathway enrichment	clusterProfiler; fgsea; enrichplot (R package)	4.8.1; 1.26.0; 1.20.0
Survival analysis	survival; survminer (R package)	3.6.4; 0.5.0
scRNA-seq data processing	Seurat; harmony (R package)	5.1.0; 1.2.0
Tumor microenvironment	CIBERSORTx	-
Patient stratification	ConsensusClusterPlus (R package)	1.64.0
Heat map plot	ComplexHeatmap (R package)	2.16.0
Network visualization	Cytoscape	3.9.1
LncRNA Subcellular localization	RNALocate; LncATLAS	-

Category	Database/Tool/Package	Version
Drug-target data acquisition	DGIdb	-
Data visualization	ggplot2; ggpubr (R package)	3.5.1; 0.6.0
Statistical computing environment	R	4.3.2

Results

Identification and validation of progression-related LncRNAs in colorectal cancer

Using an integrated approach, we identified and validated long non-coding RNAs (lncRNAs) associated with progression in colorectal cancer (CRC)[18]. As shown in Fig. 1A, the framework involves three major steps (Method): (I) integration of lncRNA and mRNA expression profiles from CRC datasets; (II) calculation of relationships between lncRNAs and mRNAs through partial correlation correcting tumor purity and ranking mRNAs; and (III) functional enrichment using gene set enrichment analysis (GSEA) to link candidate lncRNAs with CRC progression-related gene signatures[24] (Table S1).

To ensure a robust analysis, we integrated TCGA COAD and READ datasets (Table S2). Principal component analysis (PCA) demonstrates the successful integration of the datasets, indicating effective removal of batch effects (Fig. 1B). Differential expression analysis between normal and cancerous tissues identified numerous differentially expressed lncRNAs (Fig. 1C). Among these, many have been previously proven to be attributed with CRC progression, such as ELFN1-AS1[25] and SNHG25[26], further supporting their potential roles in driving CRC progression.

Further, we investigated the proportion of progression-related lncRNAs among all lncRNAs and differentially expressed lncRNAs. The results reveal an enrichment of progression-related lncRNAs within the differentially expressed lncRNAs, providing the identified relationships between progression-related lncRNAs

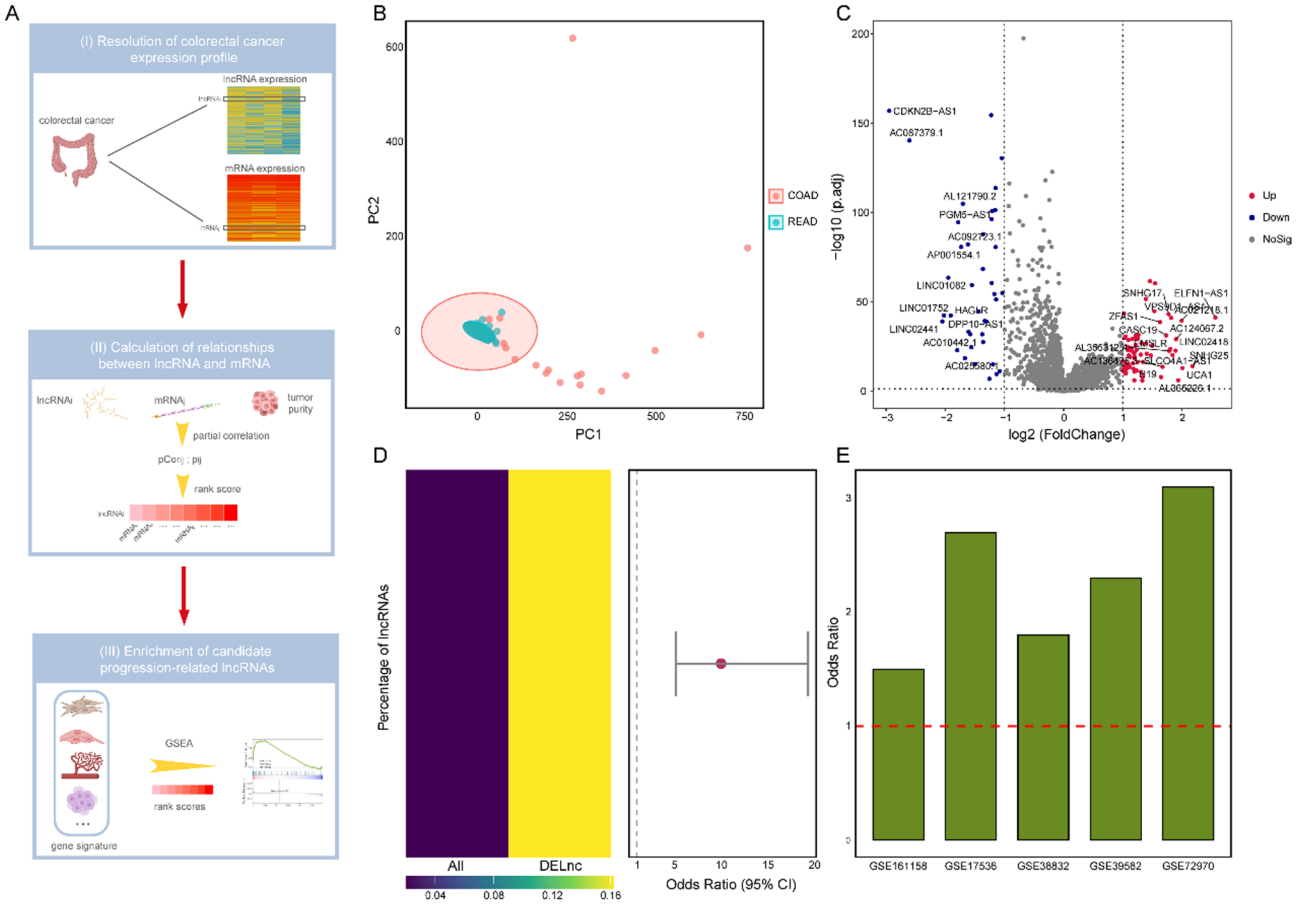

Fig. 1. Identification of progression-related lncRNAs in CRC. (**A**) Schematic representation of the analysis pipeline for identification progression-related lncRNAs. (**B**) PCA plot showing the segregation of COAD and READ patients based on lncRNA expression profiles. (**C**) Volcano plot illustrating differentially expressed lncRNAs between normal and cancer tissues. (**D**) Distribution and overlap between differentially expressed lncRNAs and progression-related lncRNAs. Two-sided Fisher's exact test. (**E**) Validation of progression-related lncRNAs using external datasets. Two-sided hypergeometric test.

and differentially expressed lncRNAs (two sides Fisher's exact test; Fig. 1D). To validate our findings, we applied the same framework to five additional CRC datasets (Fig. 1E). The hypergeometric test results indicate strong overlap between the progression-related lncRNAs identified in TCGA and those identified in these independent datasets, reinforcing the robustness and reliability of our results.

Prioritization and characterization of progression-related LncRNAs in colorectal cancer

To prioritize progression-related lncRNAs in CRC, we developed a scoring system based on ten key features: scores from seven progression-related signatures, differential expression $\log_2$(fold change) values, and survival hazard ratios (HR). For each lncRNA, we calculated a total score, averaged it, and then normalized it using a 0–1 scaling approach. Given that most lncRNAs are either antisense lncRNA or long intergenic non-coding RNA (lincRNA), we focused specifically on these two categories (Fig. 2A). Furthermore, we filtered out lncRNAs with low expression to ensure relevance, resulting in a ranked list of 198 prioritized lncRNAs sorted in descending order of standardized scores (Table S3).

Among them, LINC01615 is a progression-related lncRNA in CRC that has been confirmed by existing studies[27,28]. It was discovered that, in contrast to normal tissues, it was markedly elevated in cancer tissues (Fig. 2B). Additionally, our prioritization approach identified novel lncRNAs associated with CRC progression, such as AC007998.3, which showed significantly higher expression in CRC normal tissues compared to cancer tissues (Fig. 2C). Subsequently, we experimentally validated LINC01615 and AC007998.3. The results showed that LINC01615 was significantly upregulated in cancer tissues, consistent with our data (Fig. S1A). However, AC007998.3 did not show significant differential expression between cancer and normal tissues (Fig. S1B). Furthermore, we validated another lncRNA, LINC00460, which was significantly upregulated in cancer tissues both in the RNA-seq data (Fig. S1C) and in the experimental results (Fig. S1D). In addition to expression patterns, both LINC01615 and AC007998.3 were significantly enriched in progression-related signatures associated with CRC, including CAF (cancer-associated fibroblasts), angiogenesis, and matrix remodeling (Fig. 2D and E). These findings highlight their potential roles in driving CRC progression and underscore the effectiveness of our prioritization framework in identifying crucial progression-related lncRNAs.

Relationships between prioritized progression-related LncRNAs and colorectal cancer microenvironment cells

To comprehend how the tumor microenvironment and progression-related lncRNAs interact in colorectal cancer, we analyzed immune and non-immune cell infiltration, as well as a single-cell dataset, to uncover significant associations and functional implications. First, different patterns were found in the proportions

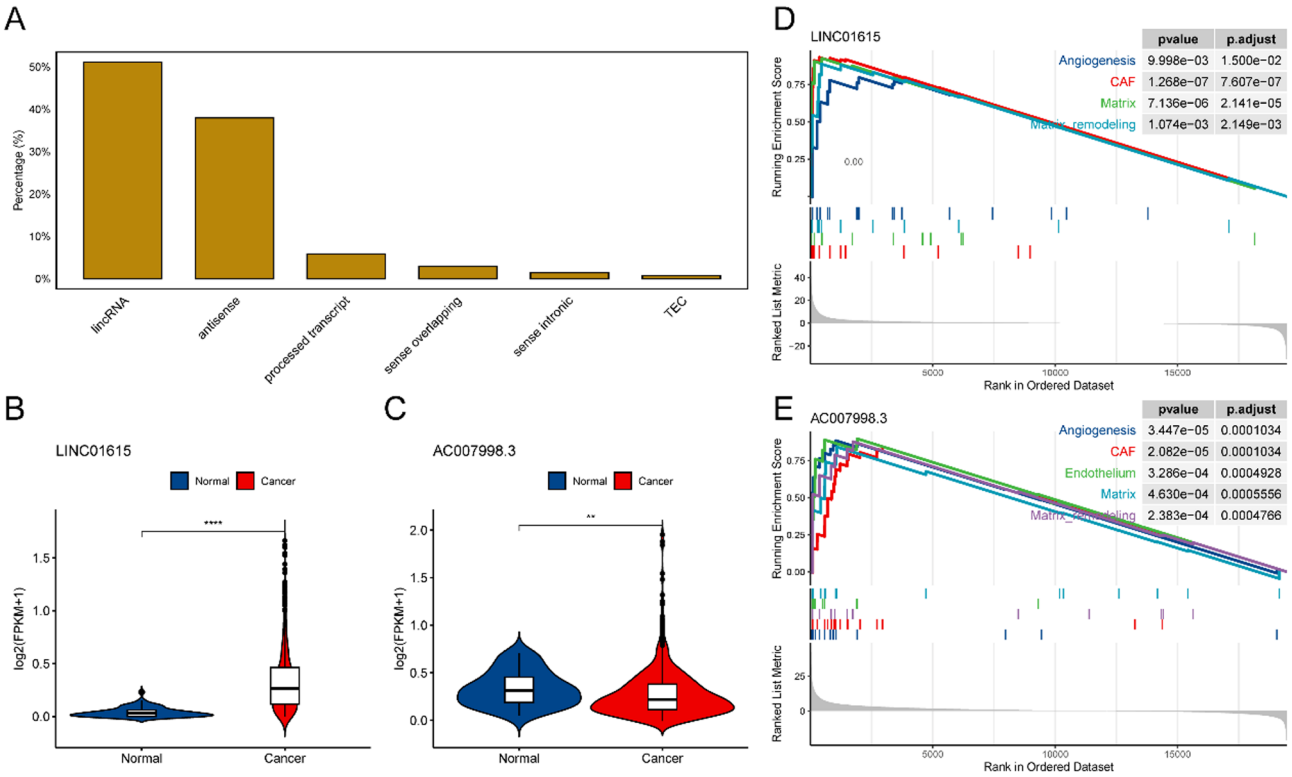

Fig. 2. Functional and expression characterization of prioritized lncRNAs in CRC. (**A**) Distribution of different RNA biotypes among progression-related RNAs. (**B**-**C**) Violin plots displaying the expression levels of LINC01615 (**B**) and AC007998.3 (**C**) in normal and cancer tissues. Two-sided Wilcoxon's rank-sum test. (**D**-**E**) GSEA results for LINC01615 (**D**) and AC007998.3 (**E**).

of immune cells and non-immune components that infiltrated CRC samples. Among immune cells, specific subtypes like naïve B cells and regulatory T cells (Tregs) showed prominent variations, while among non-immune cells, fibroblasts (CD10+), endothelial cells (CD31+), and epithelial cells (EPCAM+) were dominant (Fig. S2). These findings indicate a diverse microenvironment composition in CRC tumors, integrating both immune and stromal elements.

Next, we found that several cell types were significantly associated with prioritized lncRNAs. Progression-related lncRNAs demonstrated strong links with fibroblasts, epithelial cells, Tregs, and macrophages, suggesting their involvement in stromal and immune regulation within the CRC microenvironment (Fig. 3A). Single-cell analysis was initially performed to annotate the scRNA-seq data, identifying distinct clusters of immune and stromal cells. Marker gene analysis subsequently confirmed the identities of these clusters (Fig. 3B and C). Furthermore, single-cell enrichment analysis highlighted significant associations between prioritized lncRNAs and cell types like malignant cells (malignant epithelial cells), Tregs, CD8 + exhausted T cells (CD8 Tex), fibroblasts, and endothelial cells (Fig. 3D). The involvement of lncRNAs in tumor growth and stromal remodeling is shown by their abundance in fibroblasts and cancerous cells. Notably, Tregs and CD8 Tex showed distinct patterns of enrichment, with Tregs being more associated with immunosuppressive mechanisms, potentially aiding immune evasion, while CD8 Tex were linked to impaired anti-tumor immunity due to exhaustion. These results underscore the dual role of prioritized lncRNAs in modulating both pro-tumor and anti-tumor immune components of the CRC microenvironment.

To further describe the role of certain lncRNAs, LINC00839, ranked 24 th in our prioritized list, was used as a representative example. Most lncRNAs express themselves at a lower level than mRNA in single cells[29]. Its top 100 co-expressed mRNAs showed significant enrichment in tumor-promoting cell types, including malignant cells and fibroblasts, highlighting the likely involvement of LINC00839 in CRC progression[30] (Fig. 3E). These

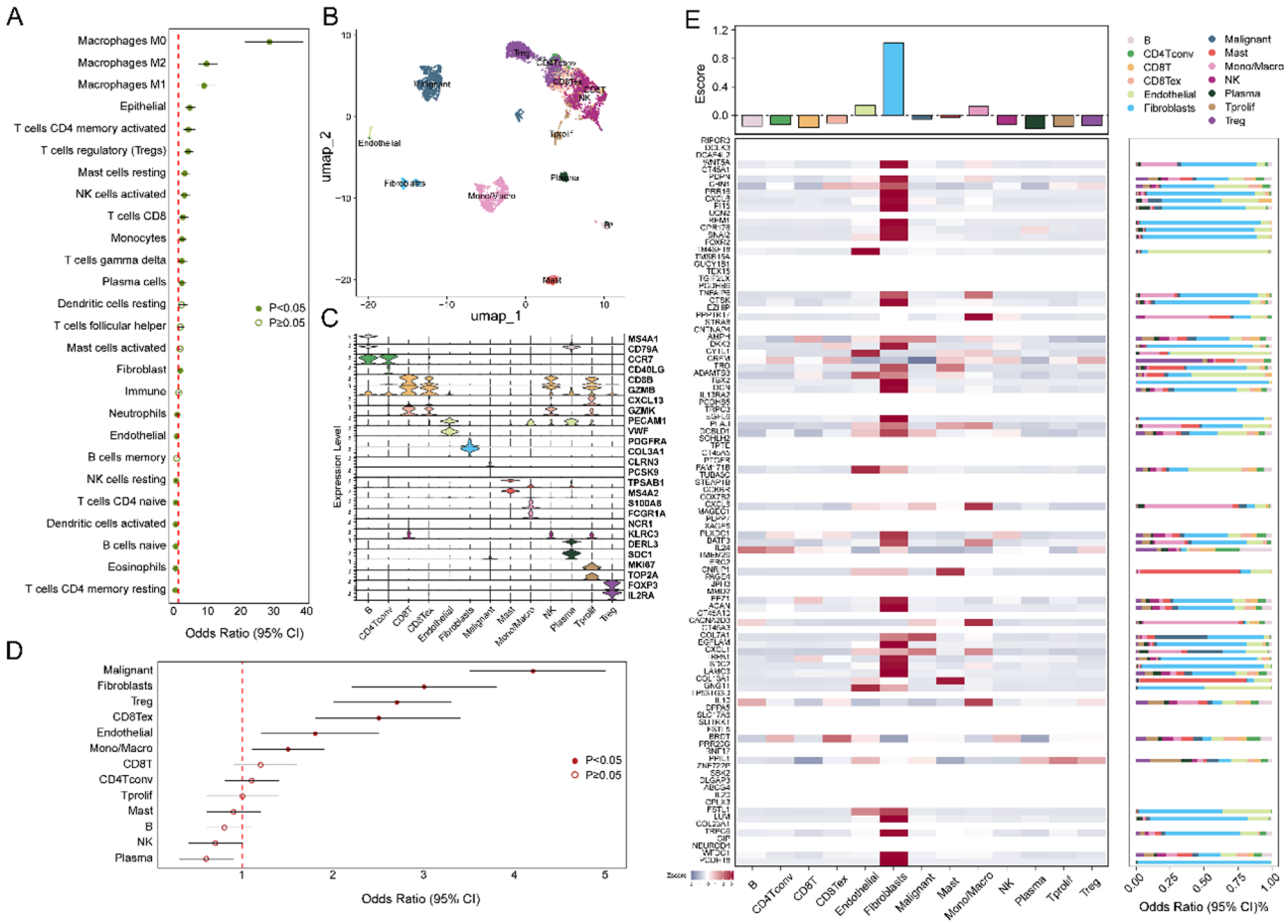

Fig. 3. Cellular landscape and key regulatory features in CRC. (**A**) Forest plot showing odds ratios and 95% confidence intervals between cell-related lncRNAs and prioritized progression-related lncRNAs in CRC. Two-sided Fisher's exact test. (**B**) UMAP plot representing scRNA-seq clustering in CRC. (**C**) Expression levels of representative marker genes across identified cell types. (**D**) Forest plot showing odds ratios and 95% confidence intervals for cell type enrichment in correlation with prioritized progression-related lncRNAs. Two-sided Fisher's exact test. (**E**) Heatmap showing the top 100 mRNAs significantly correlated with LINC00839 in single-cell data. The bar plot on the top displays the summed expression scores across cell types, and the bar plot on the right shows the distribution of odds ratios for cell type enrichment. Two-sided Fisher's exact test.

findings suggest that prioritized lncRNAs are critical mediators of tumor-stromal interactions, influencing the microenvironment to promote tumor growth and invasion.

Stratifying contribution of prioritized progression-related LncRNAs to colorectal cancer patients

Molecular subtypes are currently used to drive preclinical and clinical therapy development and treatment for many cancer types[31,32]. To investigate the clinical and molecular diversity of CRC and evaluate the potential of prioritized progression-related lncRNAs in patient stratification, we identified four distinct CRC subtypes. These subtypes exhibited substantial differences in survival, tumor characteristics, and molecular features. Specifically, we classified 638 CRC patients into four subtypes: S1 ($n = 189$), S2 ($n = 121$), S3 ($n = 132$), and S4 ($n = 196$). These subtypes showed significant survival differences (Fig. S3 and Fig. 4A), with S2 having the worst prognosis and S4 exhibiting the most favorable survival, highlighting the clinical and biological heterogeneity among them.

The distribution of tumor stages varied significantly among the subtypes. S2 was enriched in more advanced-stage patients, while S4 had a higher proportion of early-stage patients, reflecting the aggressiveness of S2 and the more indolent nature of S4 (Fig. 4B). Key cancer progression-related signature activity, including CAF, angiogenesis, EMT, and proliferation rate, showed significant differences across four subtypes (Fig. 4C). In addition, S3 exhibited the highest levels of angiogenesis, EMT, and matrix remodeling, indicating its strong association with invasive and metastatic potential. S2 showed elevated levels of CAF activity, matrix remodeling, and endothelial involvement, but moderate levels of angiogenesis and EMT, reflecting a stroma-driven tumor phenotype with an intermediate prognosis. S1 displayed relatively balanced levels across most signatures, with moderate activity in CAF and matrix remodeling but lower levels of angiogenesis and EMT, representing a less aggressive but metabolically active subtype. In contrast, S4 demonstrated lower levels of CAF activity, matrix remodeling, and endothelial involvement compared to S3 but higher levels of angiogenesis and EMT, suggesting a less stroma-driven but more invasive phenotype relative to S1. Notably, the proliferation rate was highest in S2 and S3, further emphasizing their aggressive tumor behavior and the unfavorable clinical outcomes associated with these subtypes.

The expression of the top 50 prioritized progression-related lncRNAs differed markedly among subtypes. Certain lncRNAs were clearly upregulated in S2, while others were shared by S1 and S3, suggesting that these lncRNAs play a part in the molecular heterogeneity of CRC (Fig. 4D). The composition of the tumor microenvironment also varied significantly among the four subtypes, highlighting distinct cellular contributions to their biological characteristics (Fig. 4E). S2 exhibited the higher proportions of fibroblasts and endothelial cells, aligning with its aggressive stroma-driven tumor phenotype. S3 was characterized by elevated levels of epithelial and immune cells, consistent with its strong association with invasive pathways such as EMT.

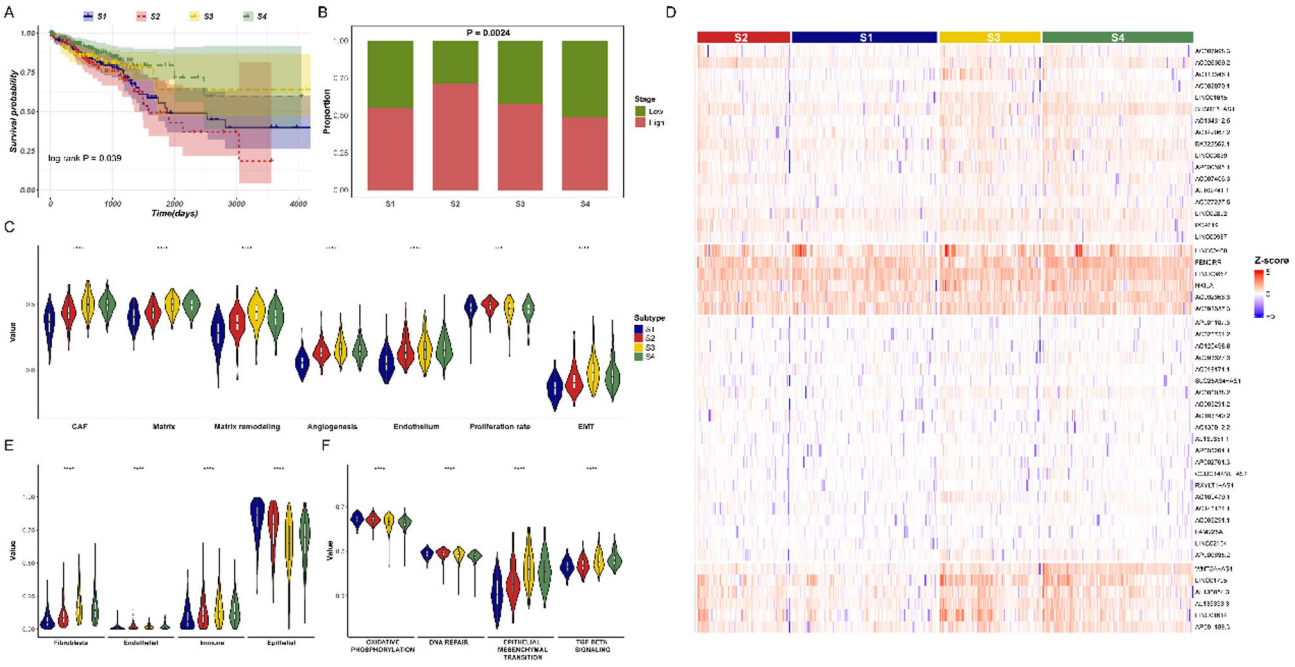

Fig. 4. Molecular and clinical characterization of CRC subtypes. (**A**) Kaplan-Meier survival curves of overall survival among the four CRC subtypes (S1-S4). Log-rank test. (**B**) Distribution of tumor stages across the four subtypes. Two-sided Fisher's exact test. (**C**) Cancer progression-related signature activity across the four subtypes. Two-sided Wilcoxon's rank-sum test. (**D**) Heatmap illustrating the expression profiles of the top 50 prioritized lncRNAs across the four subtypes. (**E**) Proportions of tumor microenvironment cell types across the four subtypes. Two-sided Wilcoxon's rank-sum test. (**F**) Activity of hallmark pathways across four subtypes. Two-sided Wilcoxon's rank-sum test.

S1 displayed intermediate levels of immune and stromal cells, suggesting a mixed tumor microenvironment profile. In contrast, S4 had a lower abundance of fibroblasts and endothelial cells and a relatively balanced microenvironment composition, reflecting its less invasive and more favorable prognosis.

Hallmark pathways such as oxidative phosphorylation, DNA repair, EMT, and TGF-beta signaling were significantly different among subtypes. S1 was characterized by higher activity in the oxidative phosphorylation pathway, while S3 showed enhanced EMT, consistent with its aggressive phenotype (Fig. 4F). S2 showed significant enrichment in DNA repair pathways, implying a heightened capability to manage and adapt to high levels of DNA damage and genomic instability. S4 had significant enrichment in TGF-beta signaling pathways, which are known to play a key role in regulating the tumor immune microenvironment. The four CRC subtypes were defined as follows: S1 (metabolically active subtype), characterized by enhanced oxidative phosphorylation and active energy metabolism; S2 (genomic instability subtype), driven by active DNA repair pathways and high genomic instability; S3 (invasive and metastatic subtype), marked by strong EMT activity and aggressive invasive behavior; and S4 (immune microenvironment-regulated subtype), associated with TGF-beta signaling and a modulated tumor immune microenvironment.

The identification of four distinct CRC subtypes (S1, S2, S3, and S4) highlights significant clinical and molecular heterogeneity, with each subtype exhibiting unique survival outcomes, tumor microenvironment compositions, and hallmark pathway activities. These findings provide a foundation for developing subtype-specific therapeutic strategies, addressing the metabolic, genomic, invasive, and immune-regulated characteristics of CRC.

LncRNA-mediated multi-level regulatory mechanisms in the S3 subtype of CRC

Understanding the complex regulatory roles of lncRNAs in CRC progression is critical for elucidating their biological functions and potential therapeutic implications. This study focuses on the S3 subtype of CRC patients, characterized by strong invasive and metastatic potential, aiming to unravel the cis-, trans-, and post-transcriptional interactions mediated by key lncRNAs and their regulatory networks. We first identified cis-regulatory interactions between the lncRNA LINC01614 and two neighboring genes, FN1 and IGFBP5, located on chromosome 2 (Fig. 5A). These interactions are within a 1 Mb genomic distance and exhibit a high degree of co-expression. The Spearman correlations for LINC01614 with FN1 and IGFBP5 are 0.72 and 0.59, respectively, highlighting potential direct regulatory effects of LINC01614 on its adjacent genes. LINC01614 is also significantly enriched in functions related to the malignant progression of cancer such as matrix remodeling and angiogenesis (Fig. 5B). These results suggest a potential cis-regulatory relationship that regulates malignant progression in patients with S3.

In addition, the trans-regulatory network of S3 subtype patients was constructed using significant correlations between lncRNAs and mRNAs (Fig. 5C and Table S4). The network comprises three lncRNAs (LINC01614, LINC01705, and AC113346.1), which target hundreds of downstream mRNAs. The central nodes in the network include lncRNAs and key CRC-associated mRNAs, forming highly connected hubs indicative of their pivotal roles in regulating cancer progression. The three trans-lncRNAs (LINC01614, LINC01705, and AC113346.1) were among the top-ranked progression-associated lncRNAs in the prioritized list. Specifically, their normalized ranks are 20 th, 12 th, and 5 th, respectively, emphasizing their strong association with aggressive tumor behavior in the S3 subtype (Fig. 5D). The mRNAs targeted in the trans-regulatory network are enriched in pathways associated with extracellular matrix organization, immune cell migration, and chemotaxis, reflecting the invasive and immune-modulating properties of the S3 subtype (Fig. 5E). Notably, processes such as neutrophil chemotaxis and collagen metabolism are significantly overrepresented, aligning with the metastatic phenotype of this group. These three potential trans-regulatory lncRNAs also had significantly expression perturbation between cancer and normal tissues (Fig. 5F).

Finally, the ceRNA network revealed interactions among lncRNAs, miRNAs, and mRNAs, with miRNAs acting as central mediators. To identify key miRNAs, we examined miRNA-lncRNA and miRNA-mRNA interactions based on Spearman correlation. This analysis highlighted MIR647 and MIR3189 as key mediators, connecting LINC01614, LINC01705, and AC113346.1 to downstream mRNAs such as SPP1, IL1RN, and HLA-DRB1 (Fig. 6A). SPP1 (osteopontin) is known to promote tumor invasion, metastasis, and immune modulation by facilitating extracellular matrix remodeling and enhancing immune cell recruitment[33]. IL1RN (interleukin-1 receptor antagonist) modulates inflammatory responses, potentially creating a tumor-promoting microenvironment[34]. Major Histocompatibility Complex (MHC) class II molecule HLA-DRB1 may have an impact on tumor immune evasion and immune surveillance[35]. These interactions underscore the role of lncRNAs in coordinating post-transcriptional regulation via miRNA sponging, highlighting their regulatory importance in CRC.

Given that lncRNAs must normally be found in the cytoplasm for ceRNA-mediated regulation[36], we assessed their subcellular distribution of LINC01614, LINC01705, and AC113346.1 (Fig. 6B). Functional enrichment study of target mRNAs linked with the ceRNA network showed various biological processes critical to the advancement of CRC, and the research demonstrated that LINC01614 and LINC01705 are localized in the cytoplasm, confirming their functions as important regulators in the ceRNA network (Fig. 6C). Key pathways included signal release, inflammatory response, and cellular response to interleukin-1, all of which play vital roles in shaping the tumor microenvironment and facilitating tumor invasion. Additionally, processes such as hormone transport and secretion suggest broader systemic influences of the ceRNA network on cancer progression. The drug-target network analysis identified a range of approved drugs targeting key mRNAs within the ceRNA network (Fig. 6D and Table S5). Notably, drugs like fenretinide[37] and bevacizumab[38], which are already used in CRC treatment, were mapped to targets such as SPP1 and HLA-DRB1. Interestingly, several other drugs targeting these mRNAs have yet to be explored for CRC treatment. For example, canakinumab, an IL-1β inhibitor[39], may hold potential as repurposed agents to target CRC pathways. These findings indicate that

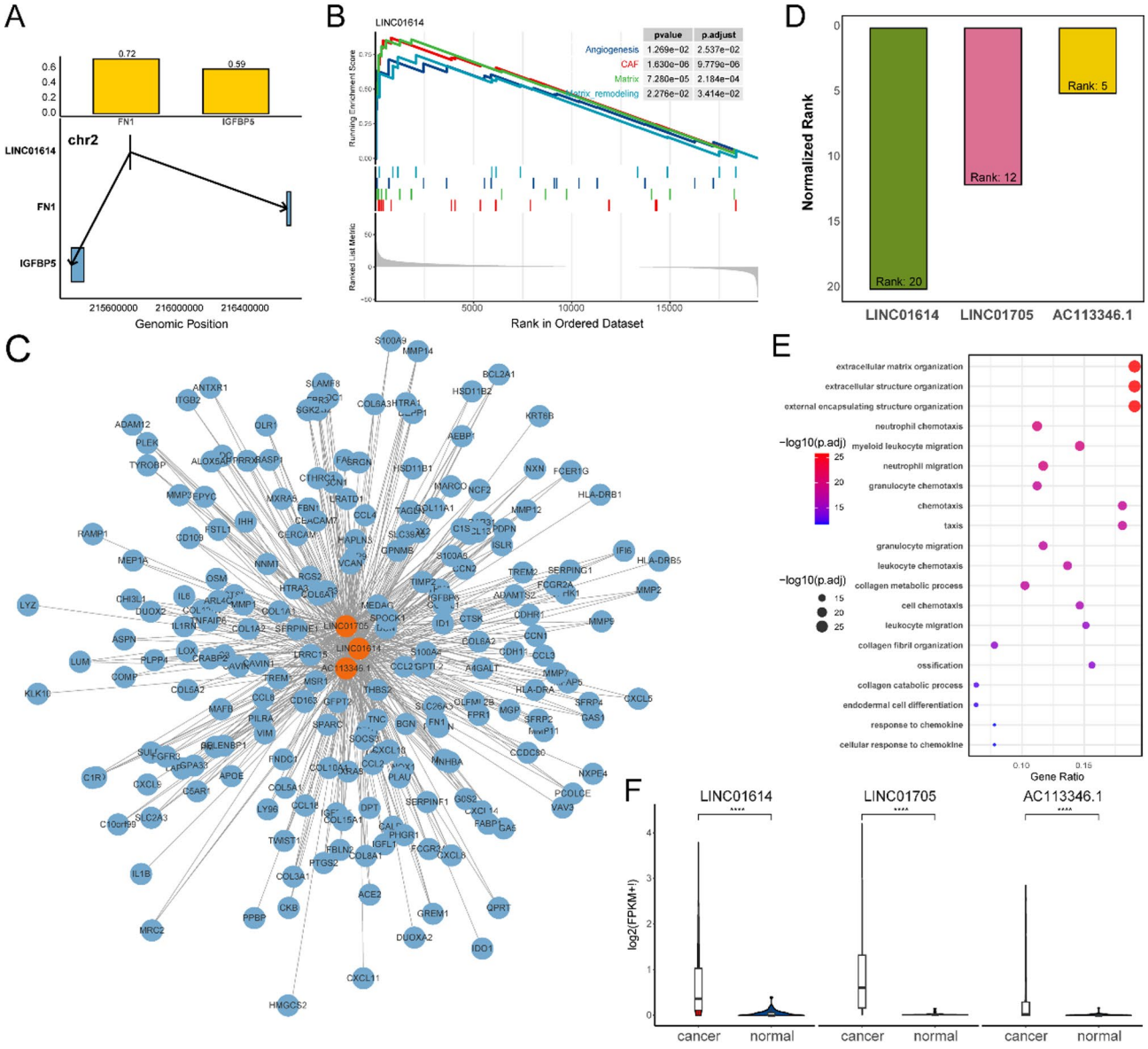

Fig. 5. Cis- and trnas- regulatory mechanisms mediated by lncRNAs in the S3 subtype of CRC patients. (**A**) Cis-regulatory interactions. (**B**) Function enrichment of LINC01614. (**C**) Trans-regulatory network. (**D**) Ranking of trans-lncRNAs in the prioritized progression-related list. (**E**) GO enrichment analysis of trans-regulated mRNAs. (**F**) Expression distribution of three trans-regulatory lncRNAs.

the ceRNA network not only gives mechanistic insights into CRC growth, but also acts as a significant resource for developing innovative treatment options.

Together, this study underscores the multifaceted roles of lncRNAs, such as LINC01614, LINC01705, and AC113346.1, in CRC. By integrating ceRNA interactions, subcellular localization, functional enrichment, and drug-target networks, we reveal how these lncRNAs orchestrate complex regulatory programs driving tumor invasion, immune modulation, and metastasis. These discoveries not only advance our understanding of CRC biology, but also provide the groundwork for prospective treatment strategies that target lncRNA-mediated regulatory networks.

Discussion

This study underscores the central roles of lncRNAs in colorectal cancer (CRC) by integrating their identification, characterization, immune relationships, regulatory mechanisms, and patient stratification. Through a systematic analysis, we highlight how lncRNAs contribute to CRC progression, immune modulation, and tumor heterogeneity, offering insights into their potential clinical applications.

The identification of progression-related lncRNAs is a critical step toward understanding their biological functions and clinical implications. By applying a robust computational framework, this study prioritized lncRNAs associated with CRC progression based on expression patterns, survival relevance, and functional

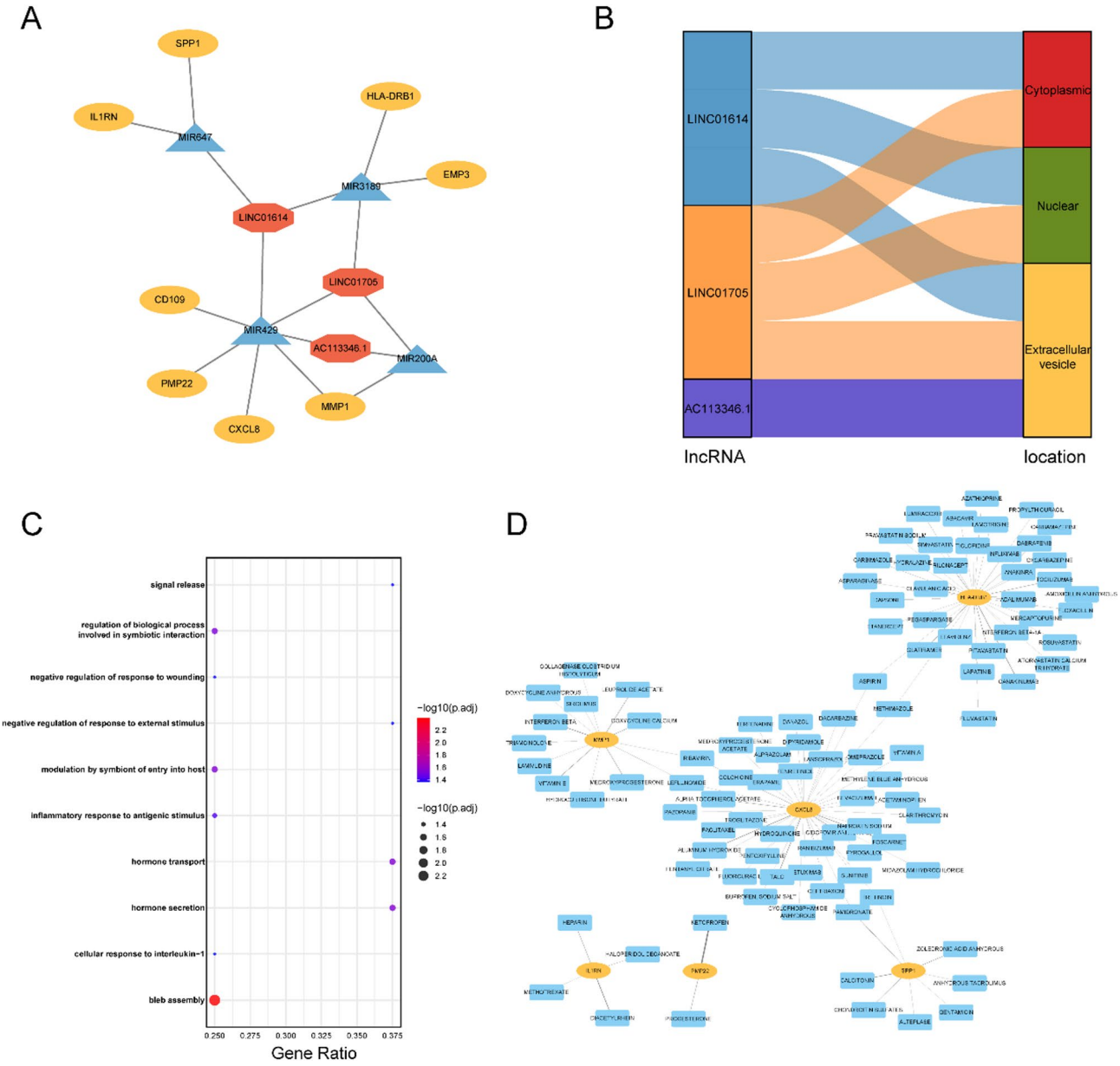

Fig. 6. CeRNA regulatory analysis of S3 subtype. (**A**) CeRNA network. Red octagons represent core lncRNAs, blue triangles represent miRNAs, and yellow ovals represent target genes (mRNAs). (**B**) RNA subcellular localization. (**C**) GO enrichment target genes. (**D**) Drug-target interaction network. Yellow ovals represent target genes (mRNAs) and blue rectangles represent corresponding approved drugs.

pathways. The inclusion of well-established lncRNAs, such as LINC01615, alongside novel candidates like AC007998.3, illustrates the capacity of this approach to capture both known and underexplored lncRNAs. These findings align with existing evidence suggesting that many lncRNAs are intricately linked to hallmark cancer pathways, such as angiogenesis and CAF activation[40,41]. However, while prioritization offers a starting point, it is imperative to validate the biological relevance of these lncRNAs experimentally.

The partnership between lncRNAs and the immunological microenvironment has garnered increasing attention, and our findings expand this field by linking prioritized lncRNAs to immune cell subtypes and stromal components. For instance, lncRNAs like LINC00839 demonstrated strong associations with fibroblasts and Tregs, suggesting their role in shaping a tumor-promoting microenvironment. This reflects a broader trend in cancer research, where non-coding RNAs are increasingly recognized as modulators of immune evasion and stromal remodeling[42,43]. Furthermore, single-cell RNA-seq data provide an unprecedented resolution to explore these interactions, revealing cell-specific enrichment patterns that deepen our understanding of lncRNA functionality in diverse tumor contexts. Despite these insights, the functional mechanisms underlying these associations remain speculative and warrant further investigation. In addition, only 11 out of the 198 prioritized lncRNAs identified from bulk RNA-seq were detected in the scRNA-seq dataset, due to the low expression

levels of lncRNAs and the dropout effects in single-cell sequencing. These lncRNAs exhibited weak and diffuse expression patterns, making it challenging to directly infer their cell-type specificity from scRNA-seq data (Fig. S4). Given these limitations, we adopted an alternative approach to infer the potential cellular context of these lncRNAs. While this approach provides valuable insights, it is important to acknowledge that indirect inference based on co-expressed protein-coding genes may not fully capture the functional roles of lncRNAs at the single-cell level. Future studies with deeper sequencing coverage, improved lncRNA-specific single-cell methodologies, or spatial transcriptomics may help to better characterize the cellular origins and biological functions of these lncRNAs.

The stratification of CRC patients into molecular subtypes (S1-S4) based on progression-related lncRNAs demonstrates the potential of lncRNAs as biomarkers for personalized medicine. Subtype-specific characteristics, such as the strong EMT activity in S3 or the immune-modulating properties of S4, underscore the heterogeneity of CRC and the importance of tailored therapeutic strategies. These findings parallel other cancer studies that emphasize the value of molecular subtypes in guiding treatment decisions and improving patient outcomes[44,45]. Notably, the integration of lncRNAs into subtype analysis bridges the gap between molecular biology and clinical oncology, positioning lncRNAs as both mechanistic drivers and clinical indicators[46]. However, transferring these findings into clinical practice necessitates confirmation in larger cohorts and functional investigations to establish their predictive and therapeutic relevance.

The regulatory versatility of lncRNAs was highlighted through their involvement in cis-, trans-, and post-transcriptional mechanisms. Cis-regulatory interactions, such as those observed for LINC01614 with neighboring genes FN1 and IGFBP5, emphasize the role of lncRNAs in modulating local transcriptional landscapes. Such interactions likely drive the invasive and metastatic behaviors characteristic of aggressive CRC subtypes. Beyond cis-regulation, trans-regulatory networks revealed the downstream impact of lncRNAs on pathways critical for stromal remodeling and immune cell migration, aligning with findings in other malignancies that position lncRNAs as central transcriptional regulators[47,48]. The ceRNA network added another layer of complexity, highlighting how lncRNAs interact with miRNAs to influence post-transcriptional gene regulation. While these mechanisms are well-supported by computational evidence, experimental validation is crucial to fully elucidate their biological significance.

Despite the extensive research offered in this paper, certain limitations must be focused on. Firstly, the regulatory networks and associations identified are primarily based on correlation and computational predictions, lacking direct experimental validation. Second, the work is based on bulk RNA-seq and single-cell data, which may not completely reflect the spatial heterogeneity and dynamic interactions found in the tumor microenvironment. Additionally, we selected three lncRNAs based on their functional relevance and statistical significance for experimental validation. However, due to the limited sample size and the scope of this initial investigation, the number of lncRNAs validated was small. We acknowledge that a more comprehensive validation with a larger sample size and additional lncRNAs will be crucial for confirming the findings. In future studies, we plan to extend our investigation to explore the molecular mechanisms underlying these lncRNAs in greater depth. This will involve further cellular and animal experiments, which will help to refine and validate our findings. Finally, while the findings are consistent across several datasets, more research in larger, independent cohorts and functional studies is required to establish the clinical and biological significance of the prioritized lncRNAs and regulatory mechanisms. Future research should focus on broadening these analyses to incorporate spatial transcriptomics and proteomics, which may give more details into the spatial organization and functional dynamics of lncRNAs within the tumor microenvironment[49].

Conclusion

This study underscores the significance of lncRNAs in CRC, highlighting their profound impact on tumor progression, microenvironment dynamics, and immune modulation. By systematically analyzing their roles in cis-, trans-, and post-transcriptional regulatory mechanisms, we identified key lncRNAs that drive aggressive phenotypes and molecular diversity in CRC. These findings advance our understanding of CRC biology and establish the groundwork for future research into lncRNA-targeted precision therapy medicine.

Data availability

The datasets analyzed in this study were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) for COAD and READ, and additional datasets, including GSE161158, GSE17536, GSE38832, GSE39582, GSE72970, and GSE146771 are available through the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). These datasets are also referenced in the Materials and Methods section as well as in the Supplementary Materials.

Received: 15 January 2025; Accepted: 12 May 2025
Published online: 18 May 2025

References

1. Morgan, E. et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* **72**, 338–344. https://doi.org/10.1136/gutjnl-2022-327736 (2023).
2. Patel, S. G. & Dominitz, J. A. Screening for colorectal cancer. *Ann. Intern. Med.* **177**, ITC49-ITC64. https://doi.org/10.7326/AITC202404160 (2024).
3. Chuang, J. P. et al. Comprehensive review of biomarkers for the treatment of locally advanced colon cancer. *Cells* **11**, 3744. https://doi.org/10.3390/cells11233744 (2022).
4. Shin, A. E., Giancotti, F. G. & Rustgi, A. K. Metastatic colorectal cancer: mechanisms and emerging therapeutics. *Trends Pharmacol. Sci.* **44**, 222–236. https://doi.org/10.1016/j.tips.2023.01.003 (2023).

5. Xu, M. D., Qi, P. & Du, X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod. Pathol.* **27**, 1310–1320. https://doi.org/10.1038/modpathol.2014.33 (2014).
6. Roberts, T. C., Morris, K. V. & Weinberg, M. S. Perspectives on the mechanism of transcriptional regulation by long non-coding RNAs. *Epigenetics* **9**, 13–20. https://doi.org/10.4161/epi.26700 (2014).
7. Zhang, X., Hong, R., Chen, W., Xu, M. & Wang, L. The role of long noncoding RNA in major human disease. *Bioorg. Chem.* **92**, 103214. https://doi.org/10.1016/j.bioorg.2019.103214 (2019).
8. Saleh, R. O. et al. lncRNA-microRNA axis in cancer drug resistance: particular focus on signaling pathways. *Med. Oncol.* **41**, 52. https://doi.org/10.1007/s12032-023-02263-8 (2024).
9. Entezari, M. et al. Long non-coding RNAs and Exosomal LncRNAs: potential functions in lung cancer progression, drug resistance and tumor microenvironment remodeling. *Biomed. Pharmacother.* **150**, 112963. https://doi.org/10.1016/j.biopha.2022.112963 (2022).
10. Punt, C. J., Koopman, M. & Vermeulen, L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat. Rev. Clin. Oncol.* **14**, 235–246. https://doi.org/10.1038/nrclinonc.2016.171 (2017).
11. Sinicrope, F. A., Okamoto, K., Kasi, P. M. & Kawakami, H. Molecular biomarkers in the personalized treatment of colorectal cancer. *Clin. Gastroenterol. Hepatol.* **14**, 651–658. https://doi.org/10.1016/j.cgh.2016.02.008 (2016).
12. Zhang, A., Miao, K., Sun, H. & Deng, C. X. Tumor heterogeneity reshapes the tumor microenvironment to influence drug resistance. *Int. J. Biol. Sci.* **18**, 3019–3033. https://doi.org/10.7150/ijbs.72534 (2022).
13. Hass, R., von der Ohe, J. & Ungefroren, H. Impact of the tumor microenvironment on tumor heterogeneity and consequences for cancer cell plasticity and stemness. *Cancers (Basel).* **12**, 3716. https://doi.org/10.3390/cancers12123716 (2020).
14. Zhang, P., Wu, W., Chen, Q. & Chen, M. Non-coding RNAs and their integrated networks. *J. Integr. Bioinform.* **16**, 20190027. https://doi.org/10.1515/jib-2019-0027 (2019).
15. Li, S. et al. Integrated modeling framework reveals co-regulation of transcription factors, MiRNAs and LncRNAs on cardiac developmental dynamics. *Stem Cell. Res. Ther.* **14**, 247. https://doi.org/10.1186/s13287-023-03442-0 (2023).
16. Winkle, M., El-Daly, S. M., Fabbri, M. & Calin, G. A. Noncoding RNA therapeutics - challenges and potential solutions. *Nat. Rev. Drug Discov.* **20**, 629–651. https://doi.org/10.1038/s41573-021-00219-z (2021).
17. Chi, Y., Wang, D., Wang, J., Yu, W. & Yang, J. Long Non-Coding RNA in the pathogenesis of cancers. *Cells* **8**, 1015. https://doi.org/10.3390/cells8091015 (2019).
18. Wang, M., Jiang, M., Xie, A., Zhang, N. & Xu, Y. Identification of CAF-related LncRNAs at the pan-cancer level represents a potential carcinogenic risk. *Hum. Mol. Genet.* **33**, 1064–1073. https://doi.org/10.1093/hmg/ddae042 (2024).
19. Yan, X. et al. Comprehensive genomic characterization of long Non-coding RNAs across human cancers. *Cancer Cell.* **28**, 529–540. https://doi.org/10.1016/j.ccell.2015.09.006 (2015).
20. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782. https://doi.org/10.1038/s41587-019-0114-2 (2019).
21. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5496.e5428. https://doi.org/10.1016/j.cell.2021.09.014 (2021).
22. Steen, C. B. et al. The landscape of tumor cell States and ecosystems in diffuse large B cell lymphoma. *Cancer Cell* **39**, 1422–1437. e1410. https://doi.org/10.1016/j.ccell.2021.08.011 (2021).
23. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170 (2010).
24. Bagaev, A. et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845–865. e847. https://doi.org/10.1016/j.ccell.2021.04.014 (2021).
25. Lei, R., Feng, L. & Hong, D. ELFN1-AS1 accelerates the proliferation and migration of colorectal cancer via regulation of miR-4644/TRIM44 axis. *Cancer Biomark.* **27**, 433–443. https://doi.org/10.3233/cbm-190559 (2020).
26. Gong, T. et al. SNHG25 promotes colorectal cancer metastasis by regulating MMP2. *Aging (Albany NY).* **15**, 10105–10116. https://doi.org/10.18632/aging.205060 (2023).
27. Xiao, Y. et al. Interaction between linc01615 and miR-491-5p regulates the survival and metastasis of colorectal cancer cells. *Transl Cancer Res.* **9**, 2638–2647. https://doi.org/10.21037/tcr.2020.03.03 (2020).
28. Hu, Z., Yang, C., Guo, S., Li, Y. & Li, Y. LINC01615 activates ZEB2 through competitively binding with miR-3653-3p to promote the carcinogenesis of colon cancer cells. *Cell. Cycle.* **21**, 228–246. https://doi.org/10.1080/15384101.2021.2015670 (2022).
29. Cabili, M. N. et al. Localization and abundance analysis of human LncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20. https://doi.org/10.1186/s13059-015-0586-4 (2015).
30. Liu, X. et al. LINC00839 promotes colorectal cancer progression by recruiting RUVBL1/Tip60 complexes to activate NRF1. *EMBO Rep.* **23**, e54128. https://doi.org/10.15252/embr.202154128 (2022).
31. Collisson, E. A., Bailey, P., Chang, D. K. & Biankin, A. V. Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 207–220. https://doi.org/10.1038/s41575-019-0109-y (2019).
32. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767. https://doi.org/10.1172/jci45014 (2011).
33. Zhong, H. et al. Machine learning predicts the oxidative stress subtypes provide an innovative insight into colorectal cancer. *Oxid. Med. Cell. Longev.* **2023**, 1737501. https://doi.org/10.1155/2023/1737501 (2023).
34. Reiner, A. P. et al. Polymorphisms of the IL1-receptor antagonist gene (IL1RN) are associated with multiple markers of systemic inflammation. *Arterioscler. Thromb. Vasc Biol.* **28**, 1407–1412. https://doi.org/10.1161/atvbaha.108.167437 (2008).
35. Jiang, N. et al. HLA and tumour immunology: immune escape, immunotherapy and immune-related adverse events. *J. Cancer Res. Clin. Oncol.* **149**, 737–747. https://doi.org/10.1007/s00432-022-04493-1 (2023).
36. Bai, Y. et al. Comprehensive analysis of a CeRNA network reveals potential prognostic cytoplasmic LncRNAs involved in HCC progression. *J. Cell. Physiol.* **234**, 18837–18848. https://doi.org/10.1002/jcp.28522 (2019).
37. Kouhara, J. et al. Fenretinide up-regulates DR5/TRAIL-R2 expression via the induction of the transcription factor CHOP and combined treatment with Fenretinide and TRAIL induces synergistic apoptosis in colon cancer cell lines. *Int. J. Oncol.* **30**, 679–687 (2007).
38. Prager, G. W. et al. Trifluridine-tipiracil and bevacizumab in refractory metastatic colorectal cancer. *N Engl. J. Med.* **388**, 1657–1667. https://doi.org/10.1056/NEJMoa2214963 (2023).
39. Dhimolea, E. Canakinumab. *MAbs* **2**, 3–13. https://doi.org/10.4161/mabs.2.1.10328 (2010).
40. Wang, J. et al. Regulatory roles of long noncoding RNAs implicated in cancer hallmarks. *Int. J. Cancer.* **146**, 906–916. https://doi.org/10.1002/ijc.32277 (2020).
41. Elimam, H. et al. Unraveling the influence of LncRNA in gastric cancer pathogenesis: a comprehensive review focus on signaling pathways interplay. *Med. Oncol.* **41**, 218. https://doi.org/10.1007/s12032-024-02455-w (2024).
42. Arora, L. & Pal, D. Remodeling of stromal cells and immune landscape in microenvironment during tumor progression. *Front. Oncol.* **11**, 596798. https://doi.org/10.3389/fonc.2021.596798 (2021).
43. Uddin, M. N. & Wang, X. The landscape of long non-coding RNAs in tumor stroma. *Life Sci.* **264**, 118725. https://doi.org/10.1016/j.lfs.2020.118725 (2021).
44. Wang, R. C. & Wang, Z. Precision medicine: disease subtyping and tailored treatment. *Cancers (Basel).* **15**, 3837. https://doi.org/10.3390/cancers15153837 (2023).

45. Prat, A. et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **24**(Suppl 2), 26–35. https://doi.org/10.1016/j.breast.2015.07.008 (2015).
46. Wang, Z. et al. Multi-omics integrative bioinformatics analyses reveal long non-coding RNA modulates genomic integrity via competing endogenous RNA mechanism and serves as novel biomarkers for overall survival in lung adenocarcinoma. *Front. Cell. Dev. Biol.* **9**, 691540. https://doi.org/10.3389/fcell.2021.691540 (2021).
47. Huarte, M. The emerging role of LncRNAs in cancer. *Nat. Med.* **21**, 1253–1261. https://doi.org/10.1038/nm.3981 (2015).
48. Nandwani, A., Rathore, S. & Datta, M. LncRNAs in cancer: regulatory and therapeutic implications. *Cancer Lett.* **501**, 162–171. https://doi.org/10.1016/j.canlet.2020.11.048 (2021).
49. Jin, Y. et al. Advances in Spatial transcriptomics and its applications in cancer research. *Mol. Cancer*. **23**, 129. https://doi.org/10.1186/s12943-024-02040-9 (2024).

Acknowledgements
This study benefited from the TCGA and GEO databases. We appreciate the data platforms and authors for sharing their data.

Author contributions
W.W. and W.X. conceived and designed the study. W.X., X.S., and M.W. wrote the manuscript. M.W., X.W., K.L., and R.J. processed and performed the bioinformatic analysis. W.W. and X.J. revised the manuscript. All authors contributed to the article and approved the submitted version.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Medical Ethics Committees of Liaocheng People's Hospital (No: 2025063, 12 March 2025). Written informed consent was provided by all participants.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-02096-7.

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.