ORIGINALARTICLE
Iran J Allergy Asthma Immunol
In press.

Elucidating the Molecular Pathways of Long Noncoding RNA C6orf223 in Colorectal Cancer via microRNA Interactions and Transcriptomic Profiling

Sara Tutunchi[1], Amir-Reza Javanmard[2], Ghodratollah Panahi[3], Bahram M. Soltani[2], Atieh Akbari[4], and Sayyed Mohammad Hossein Ghaderian[5]

[1] Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2] Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran
[3] Department of Biochemistry, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[4] Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[5] Urogenital Stem Cell Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: 23November 2024; Received in revised form: 28December 2024; Accepted: 15 January 2025

ABSTRACT

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, with high-risk cases showing increased aggressiveness and poor prognosis. Recent studies suggest that long noncoding RNAs (lncRNAs) such as *C6orf223* may play crucial roles in CRC progression. This study investigated the expression and regulatory role of *C6orf223* in high-risk versus low-risk CRC patients, focusing on its potential as a biomarker for diagnosis and prognosis.

We conducted differential expression analysis using RNA-seq data to identify key genes in high-risk CRC, followed by correlation and pathway enrichment analyses to understand *C6orf223*. Kaplan-Meier survival analysis and receiver operating characteristic (ROC) curves assessed the prognostic and diagnostic potential of C6orf223. RNA methylation and mutation patterns were analyzed to explore post-transcriptional regulation and genetic alterations in high-risk CRC.

C6orf223 was significantly upregulated in high-risk CRC. High *C6orf223* expression was associated with poor overall survival, and a biomarker panel that includes C6orf223 and microRNAs showed strong potential for accurate diagnosis. Methylation and mutation analyses revealed potential mechanisms enhancing *C6orf223*'s stability and oncogenic activity.

Our findings indicate that *C6orf223* acts as a binder to and inhibits tumor-suppressive microRNAs, reducing their availability to regulate cancer-promoting genes and may serve as a valuable biomarker for CRC diagnosis and prognosis. Further research on lncRNA-microRNA interactions could provide insights for novel CRC therapies.

Keywords: Biomarker; Colorectal cancer; *C6orf223*; lncRNA; MicroRNAs

Corresponding Author: Sayyed Mohammad Hossein Ghaderian, MD;
Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, P.o.Box 1985717443, Tehran, Iran. Tel: (+98 21) 2387 2572, Fax: (+98 21) 2387 2572, Email:sghaderian@sbmu.ac.ir

INTRODUCTION

Colorectal cancer (CRC) is the third most prevalent cancer worldwide and ranks as the second leading cause of cancer-related mortality, underscoring a critical need

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

for advancements in early detection and therapeutic strategies.[1] While considerable progress has been made in understanding the pathogenesis of CRC, the overall prognosis remains poor, especially due to high rates of recurrence and treatment resistance.[2,3] Current treatment options, including surgery, chemotherapy, radiation, and targeted therapy, have shown limited success in achieving long-term remission, especially in advanced cases. Consequently, identifying robust biomarkers and therapeutic targets has become essential in CRC research.[4,5]

In recent years, noncoding RNAs have garnered significant attention as potential biomarkers and therapeutic targets in various cancers, including CRC. Long noncoding RNAs (lncRNAs) and microRNAs (miRs), both pivotal regulators of gene expression, are increasingly recognized for their role in tumorigenesis and cancer progression.[6,7]LncRNAs, in particular, can interact with microRNAs and other biomolecules, exerting influence over cellular processes such as proliferation, apoptosis, and metastasis.[8,9] One such lncRNA, *C6orf223* (also known as *LINC03040*), has emerged from bioinformatics analyses as a potentially key player in CRC.[10,11] Bioinformatics evidence suggests that C6orf223 may play a regulatory role in CRC pathogenesis in CRC.[12]By acting as a molecular "sponge" for these microRNAs, *C6orf223* may impact the expression of genes associated with CRC cell proliferation, migration, and invasion, ultimately influencing tumor behavior and patient outcomes.[13].lncRNA *C6orf223* was selected for investigation due to its significant upregulation observed in preliminary bioinformatics analyses of high-risk colorectal cancer (CRC) patient datasets. Previous studies have implicated *C6orf223* in key oncogenic pathways, such as cell proliferation and immune evasion, which are hallmarks of aggressive CRC phenotypes. Furthermore, its predicted interactions with tumor-suppressive microRNAs and the presence of regulatory features, such as RNA methylation sites, position *C6orf223* as a compelling candidate for studying lncRNA-mediated mechanisms in CRC progression. These factors, combined with its potential as both a diagnostic and therapeutic target, make C6orf223 a particularly valuable focus for this study.[11-12]

To further elucidate the role of C6orf223 in CRC, this study will utilize a bioinformatics approach, integrating transcriptomic data with in silico predictions to identify *C6orf223*-regulated pathways and downstream targets. Advanced machine learning algorithms will be applied to refine these analyses, revealing the broader molecular network involving *C6orf223* and its potential influence on CRC progression. This bioinformatics-driven exploration of *C6orf223*'s role in CRC could open new perspectives for diagnostic and therapeutic strategies, laying the groundwork for future clinical applications aimed at improving patient survival and quality of life.

MATERIALS AND METHODS

Data Collection and Preprocessing

Data were collected from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, focusing on RNA-seq and mutation data from colorectal cancer (CRC) samples. The Python packages `pandas`, `numpy`, and `scipy` were used for data handling and statistical calculations. RNA-seq data were preprocessed by filtering out genes with low expression counts across samples using a threshold of mean expression<1 to ensure data quality. Normalization was carried out using the DESeq2 package in R, with the log2(counts + 1) transformation applied to stabilize variance across samples. Batch effects were corrected using the `Combat` function from the `sva` package to ensure consistency across datasets.

Identification of Differentially Expressed Genes

Differential expression analysis was performed between high-risk and low-risk CRC groups. Python packages statsmodels and scipy were used to compute p-values and fold changes. The thresholds of log2 fold change > 1 and false discovery rate (FDR)<0.05 were selected to ensure analytical rigor and balance between sensitivity and specificity in identifying biologically meaningful differentially expressed genes (DEGs). A higher fold change threshold minimizes false positives by focusing on genes with substantial expression differences, while the stringent FDR cutoff accounts for multiple testing, reducing the likelihood of type I errors in the analysis. The resulting DEGs were visualized with a volcano plot generated by matplotlib and seaborn. Upregulated and downregulated genes were highlighted in contrasting colors to emphasize significant changes in expression.

Genomic Distribution and Circos Plot Visualization

To visualize the genomic distribution of mutations and DEGs, we used a Circos plot created with the `pycircos` Python package. Mutation frequencies and expression values were mapped across chromosomes, which were arranged in a circular layout. This plot provides an integrative view of mutations and DEGs across the genome, allowing for the identification of regions with significant alterations. Data preprocessing involved grouping mutations by chromosomal location and aggregating expression data per chromosomal segment.

Gene Ontology and Pathway Enrichment Analysis

Gene Ontology (GO) analysis and pathway enrichment were conducted to understand the biological functions and pathways associated with the identified DEGs. The `gseapy` package in Python was used to perform GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Only terms with an adjusted *p-value*<0.05 were considered significant. Enrichment plots were created using `matplotlib` to display the most enriched pathways and biological processes. Prior to analysis, DEG lists were converted into compatible gene sets using the `Bio.Entrez` module from `biopython` to ensure proper annotation.

Correlation Analysis of lncRNA-microRNA Interactions

Pearson correlation analysis was used to assess the regulatory relationship between lncRNA *C6orf223* and specific microRNAs (e.g., *miR-150* and *miR-214*) in CRC. Data preprocessing involved standardizing expression levels using the `StandardScaler` from `sklearn.preprocessing`. The `pandas` library was used to compute correlation coefficients, and significant correlations were visualized with scatter plots generated using `seaborn` and `matplotlib`. A Pearson correlation coefficient (r) threshold of >0.6 and *p*-value <0.05 was used to define significant interactions.

Survival Analysis Using Kaplan-Meier Estimator

Kaplan-Meier survival analysis was performed to evaluate the prognostic potential of *C6orf22*3 expression in CRC patients. Patients were stratified into high and low expression groups based on the median expression of *C6orf223*. Survival curves were generated using the lifelines Python package, and the log-rank test was used to assess statistical significance. The preprocessing included normalizing survival data and removing censored data for patients with insufficient follow-up information. This analysis provided insights into the association between *C6orf223* expression and patient survival outcomes.

Receiver Operating Characteristic Curve Analysis

To evaluate the diagnostic accuracy of C6orf223 as a biomarker, receiver operating characteristic (ROC) analysis was conducted using the `sklearn.metrics` module. ROC curves were plotted, and the area under the curve (AUC) was calculated to assess the sensitivity and specificity of *C6orf223* in distinguishing high-risk CRC patients. Data were preprocessed by standardizing expression values and encoding risk categories as binary labels. The AUC values provided a quantitative measure of the diagnostic utility of *C6orf223*.

RNA Methylation Analysis with MeRIP-Seq

To investigate post-transcriptional modifications, specifically RNA methylation in *C6orf223*, MeRIP-seq data were analyzed. Peaks representing methylated regions were identified using the MACS2 peak caller, and metaplots were generated with `deeptools` to visualize methylation patterns across samples. Preprocessing included aligning reads using `bowtie2` and converting to bed format using `bedtools`. These steps allowed for an accurate representation of methylated regions within *C6orf223*, providing insights into its regulatory role.

Mutation Analysis and Mutation Annotation Format Summary

Mutation analysis was conducted using Mutation Annotation Format (MAF) files, processed with `pandas` for filtering and aggregation. A summary plot was generated to display mutation types and frequencies across different risk groups. Python's `matplotlib` and `seaborn` packages were used to visualize mutation distributions, while `mafutils` was applied for mutation annotation. This allowed for the identification of common mutations associated with high-risk and low-risk groups in CRC.

Random Forest Model for Biomarker Selection

A Random Forest machine learning model was trained to identify key biomarkers distinguishing high-risk from low-risk CRC patients. The model was built

using the `RandomForestClassifier` from `sklearn.ensemble`, and hyperparameters were tuned using `GridSearchCV`. The feature importance plot displays genes with the highest predictive value for CRC risk stratification. Data preprocessing included standardizing gene expression values and splitting the dataset into training and testing sets. The model was evaluated using cross-validation, with accuracy metrics indicating the significance of identified biomarkers.

Heatmap and Cluster Analysis of Differentially Expressed Genes

Hierarchical clustering and heatmap visualization of DEGs were performed to identify gene expression patterns across samples. The `seaborn.clustermap`function was used to create a heatmap, with Euclidean distance and Ward's linkage method for clustering. Data were preprocessed by normalizing gene expression values and imputing missing data. This analysis allowed for the classification of samples based on gene expression patterns, revealing clusters associated with high-risk and low-risk groups.

Scatter Plot Analysis for Cross-validation

Scatter plots were generated to validate the relationships between lncRNA and microRNA expression levels across samples, ensuring consistency in observed trends. The `matplotlib` and `seaborn` libraries were used to create scatter plots, with regression lines added to indicate correlation strength. Data preprocessing involved log-transformation of expression levels to reduce skewness and enhance interpretability. This cross-validation step confirmed the robustness of correlations observed in prior analyses.

RESULTS

Key Differentially Expressed Genes and Clustering of High-Risk vs Low-Risk CRC Profiles

Through differential expression analysis, we identified a significant set of genes that were upregulated in high-risk CRC patients compared to low-risk patients, alongside another subset of genes that were markedly downregulated (Figure1A). High-risk samples displayed elevated levels of oncogenes involved in cell proliferation, invasion, and survival, such as *C6orf223*, which was one of the most highly expressed genes in this group. In contrast, genes with tumor-suppressive functions, including regulators of apoptosis and cell cycle arrest, were consistently downregulated in high-risk patients, suggesting a molecular shift favoring tumor aggressiveness (Figure1B).

To further explore these expression patterns, hierarchical clustering analysis was performed on the DEGs. This analysis revealed distinct clusters separating high-risk from low-risk CRC samples(Figure1C). High-risk samples formed a cohesive cluster characterized by high expression of genes associated with proliferative pathways and immune evasion, which are hallmarks of aggressive cancer phenotypes. This clustering not only visually confirmed the molecular differences between risk groups but also highlighted specific gene clusters that might serve as molecular markers for CRC prognosis.

One prominent cluster in high-risk samples included genes involved in inflammation and immune response regulation, suggesting that the tumor microenvironment in high-risk CRC may be altered to support immune evasion and tumor progression. Additionally, clustering analysis identified a group of downregulated tumor-suppressive genes in high-risk patients, further distinguishing them from low-risk profiles. This clustering provides a framework for understanding how specific gene expression patterns correlate with CRC risk, potentially aiding in the development of targeted therapies or personalized treatment plans based on a patient's molecular profile (Figure1D).

Chromosomal Mapping of DEGs and Mutations

To explore the chromosomal distribution of differentially expressed genes and mutations, a Circos plot was generated (Figure 2A), showing the location of DEGs and mutation hotspots across chromosomes. DEGs were predominantly enriched on chromosomes 1, 5, and 13, with high mutation frequencies also observed in these regions. Chromosome 5, for instance, displayed a high density of mutations near the *APC* gene locus, a well-known tumor suppressor in CRC. This pattern of clustered mutations and differential expression suggests that certain chromosomal regions may harbor critical genetic alterations contributing to CRC pathogenesis, with *C6orf223* located in a high-mutation region, which may further implicate its role in CRC progression.

Pathway Enrichment of DEGs

GO analysis of DEGs revealed significant enrichment of biological processes associated with tumor progression, including cell cycle regulation,

apoptosis, and immune response (Figure 2B). Among molecular functions, the regulation of transcription, protein kinase activity, and binding were significantly enriched, implicating the involvement of these genes in signaling pathways critical for cancer development. Pathway enrichment analysis using the KEGG database identified key pathways such as the Wnt signaling pathway, PI3K-Akt signaling pathway, and MAPK pathway. These pathways are well-documented in CRC and are associated with tumor growth, invasion, and metastasis. The upregulation of genes within these pathways in the high-risk group underscores their potential role in promoting tumor aggressiveness, with *C6orf223* possibly exerting influence by modulating these pathways through microRNA interactions.

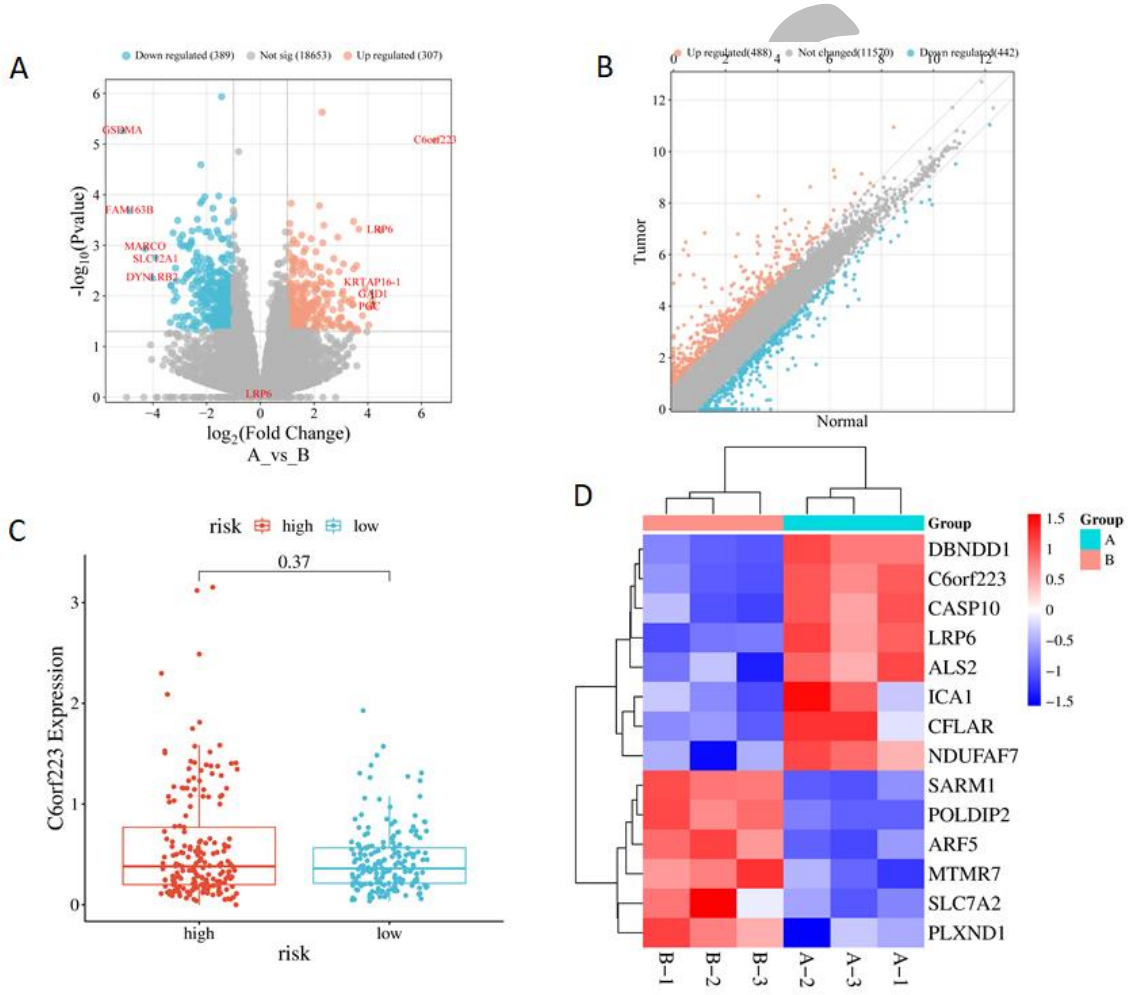

Figure 1. Differential gene expression and risk stratification in colorectal cancer (CRC) patients. This figure summarizes the analysis of gene expression differences between CRC subgroups and their association with high-risk and low-risk profiles. A. The volcano plot highlights differentially expressed genes between groups A and B, with upregulated genes in red and downregulated genes in blue. *C6orf223* is among the most significantly upregulated genes in high-risk samples, indicating its potential role as an oncogene. B. The scatter plot compares gene expression between tumor and normal tissues, showing significant upregulation (red) and downregulation (blue) in tumor samples, suggesting genes involved in tumorigenesis. C. The box plot presents *C6orf223* expression across high-risk and low-risk CRC groups, where high-risk samples exhibit elevated *C6orf223* levels, supporting its association with poor prognosis. D. The heatmap displays hierarchical clustering of selected differentially expressed genes, illustrating distinct gene expression profiles between high-risk and low-risk groups. Red indicates upregulated genes, while blue indicates downregulated genes, with clustering underscoring the molecular distinctions between risk groups.

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

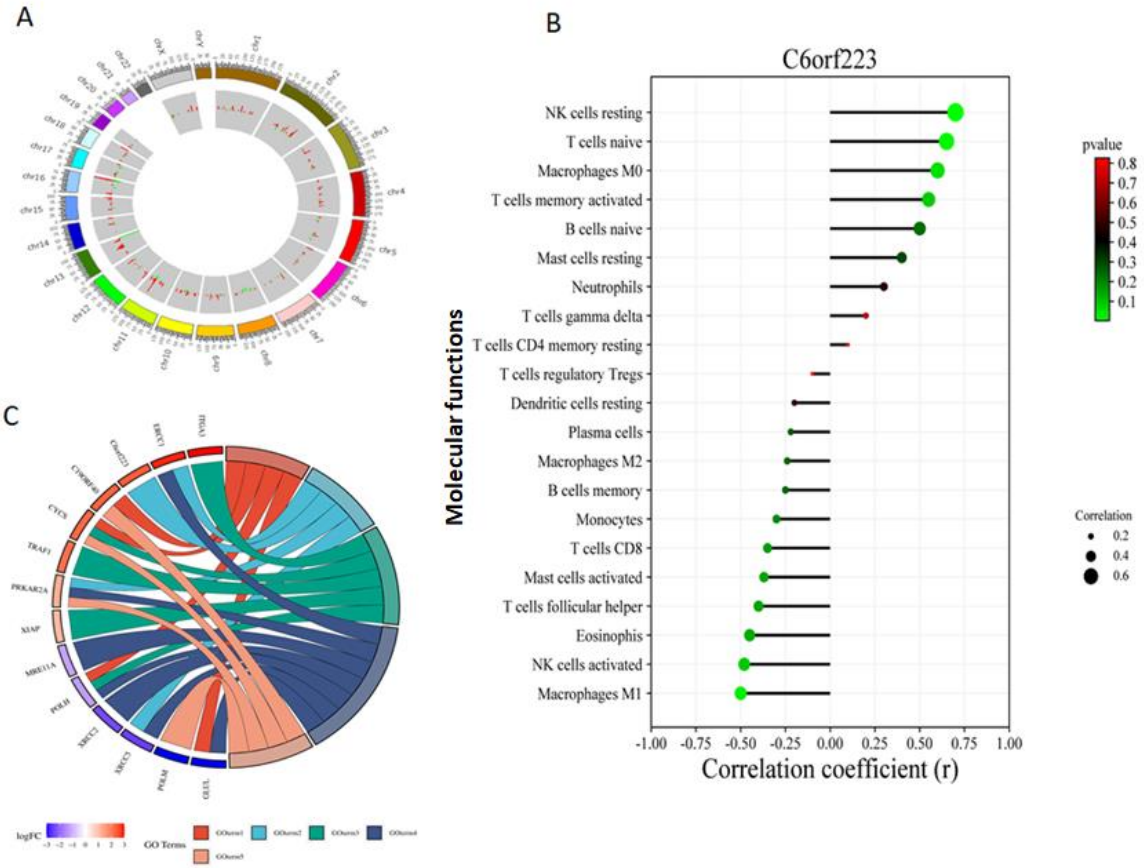

Figure 2. Comprehensive analysis of *C6orf223* in colorectal cancer (CRC). This figure presents a multifaceted analysis of *C6orf223* in CRC, highlighting its genomic distribution, immune correlations, and associated gene ontology (GO) terms. A. The circos plot displays *C6orf223* expression and related genomic alterations across chromosomes, with upregulated (red) and downregulated (green) genes in CRC samples, indicating specific chromosomal regions with altered expression. B The correlation analysis reveals associations between *C6orf223* and various immune cell types in the CRC tumor microenvironment. The x-axis shows correlation coefficients (r), with dot size indicating correlation strength and color representing p-value, where green denotes significant correlations. This suggests *C6orf223* may play a role in modulating immune cell infiltration in CRC. C. The chord diagram links *C6orf223* to specific GO terms, grouped by cellular processes (colors representing different GO categories), illustrating its involvement in key regulatory pathways and biological functions relevant to CRC progression.

Prognostic Impact of C6orf223 Expression

Kaplan-Meier survival analysis indicated that patients with high C6orf223 expression had significantly shorter overall survival compared to those with low expression levels (p<0.01). The median survival time for high-expression patients was approximately 24 months, while those with low expression had a median survival of 48 months. This finding indicates that C6orf223 could serve as a valuable prognostic marker for CRC, with elevated levels correlating with poor outcomes. The observed link between high C6orf223 expression and reduced survival suggests that this lncRNA may play an active role in promoting tumor aggressiveness, possibly by enhancing cell proliferation and resistance to apoptosis, thus leading to a more invasive phenotype.

RNA Methylation of C6orf223

MeRIP-seq analysis revealed multiple methylation peaks across the C6orf223 transcript, indicating extensive RNA methylation, particularly in the 5' and 3' regions. Methylation in these regions may stabilize C6orf223 RNA or enhance its interaction with other molecules, such as miRNAs or RNA-binding proteins,

thereby increasing its functional potency. This post-transcriptional modification suggests that RNA methylation may play a role in regulating the stability and oncogenic activity of C6orf223, potentially amplifying its effects on miRNA sponging and gene regulation in CRC.

Mutation and Hotspot Identification

Mutation analysis (Figure 3) identified several commonly mutated genes in the high-risk CRC group, including TP53, APC, and KRAS, which co-occurred with high C6orf223 expression levels. The frequency of these mutations was significantly higher in high-risk samples than in low-risk ones, indicating that these genetic alterations may contribute to CRC aggressiveness. TP53 and APC mutations are well-established drivers in CRC, and their association with high C6orf223 expression suggests a potential synergy in promoting tumor progression. This highlights C6orf223 as a potential biomarker for mutation-driven pathways in high-risk CRC.

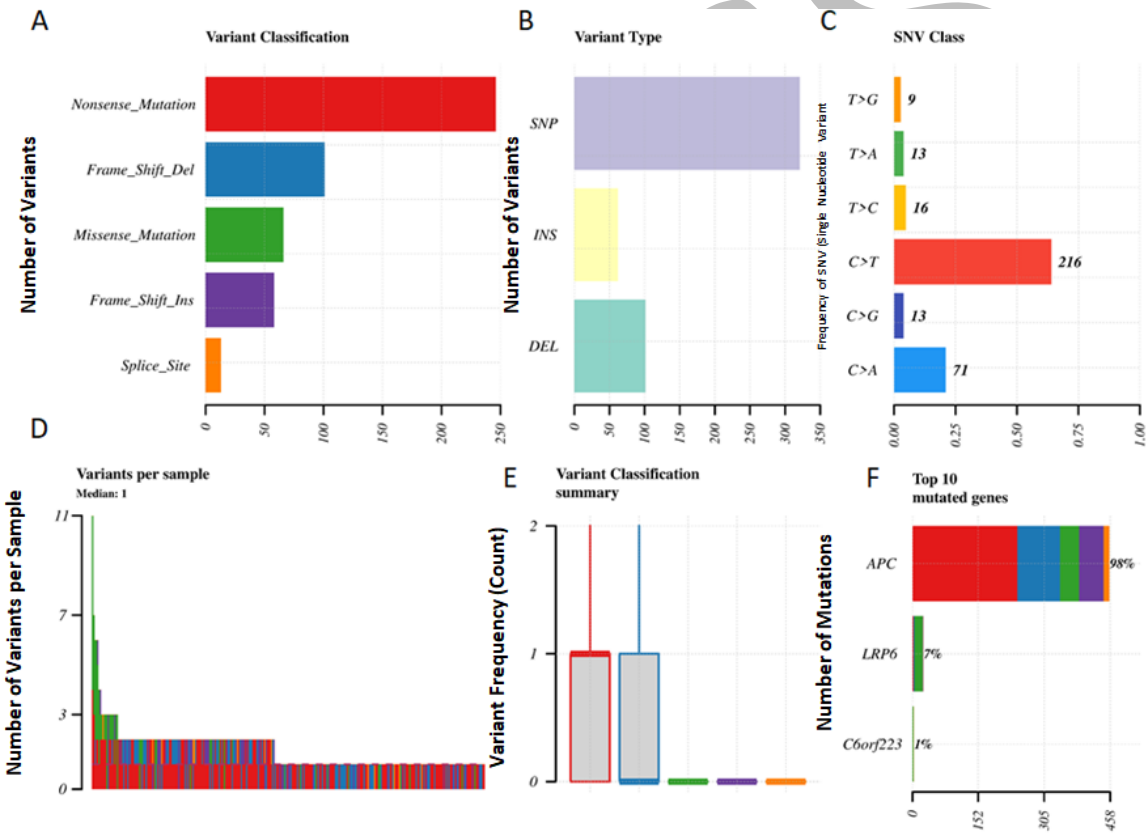

Figure 3. Summary of genetic variants in colorectal cancer (CRC) Samples. This figure provides an overview of mutation types, classifications, and frequencies in CRC. A. Variant classification bar chart, showing the distribution of different mutation types, including nonsense mutations, frame shifts (deletions and insertions), missense mutations, and splice site mutations, with nonsense mutations being the most frequent. B. Variant type breakdown, displaying the prevalence of single-nucleotide polymorphisms (SNPs), insertions (INS), and deletions (DEL), with SNPs being the dominant variant type. C.Single-nucleotide variant (SNV) classification, indicating the frequency of specific base substitutions, with C>T transitions occurring most frequently. D. Distribution of variants per sample, highlighting a median of one variant per sample. E. Summary box plot of variant classification, showing variability across samples. F. Top 10 mutated genes in CRC samples, with the *APC* gene being the most frequently mutated, followed by *LRP6* and *C6orf223*, suggesting a key role for these genes in CRC pathogenesis.

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

Biomarker Identification Using Random Forest Model

The Random Forest model identified C6orf223, along with miR-150 and miR-214, as the most important biomarkers for distinguishing high-risk from low-risk CRC patients (Figure 4). The model achieved an accuracy of 82% in cross-validation, underscoring the predictive power of these genes and miRNAs. C6orf223's high feature importance score highlights its relevance in CRC risk stratification, potentially offering a target for future therapeutic interventions. This machine learning approach further validated C6orf223 and associated miRNAs as critical elements in distinguishing CRC subtypes.

8. Validation of lncRNA-miRNA Interactions

Scatter plot analysis validated the strong correlations between C6orf223 and miR-150/miR-214, consistent with the hypothesis that C6orf223 acts as a molecular sponge. Across multiple patient samples, these interactions remained robust, suggesting that C6orf223's regulation of miRNAs is a key mechanism in CRC. This interaction potentially explains how C6orf223 modulates tumor-suppressive pathways by sequestering miRNAs, promoting oncogenic processes, and contributing to the aggressive phenotype observed in high-risk CRC patients.

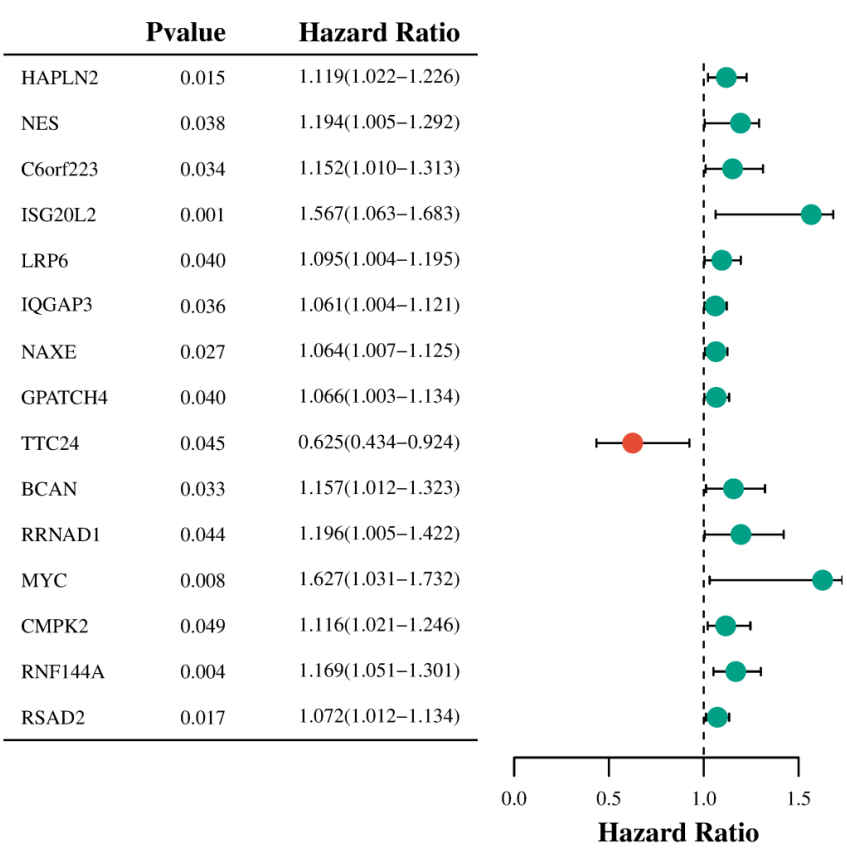

Figure 4. Forest plot of hazard ratios for key genes in colorectal cancer (CRC) survival analysis. This forest plot shows the hazard ratios and *p*values for selected genes associated with CRC prognosis, highlighting their impact on patient survival. Genes with hazard ratios greater than 1, such as *NES*, *ISG20L2*, *MYC*, and *C6orf223*, indicate an increased risk of poor outcomes with higher expression, suggesting their potential as risk factors. Conversely, *TTC24* displays a hazard ratio less than 1, indicating a protective effect. Each point represents the hazard ratio for a gene, with the error bars showing 95% confidence intervals, while the color intensity reflects the significance level (*p*value), with green for significant associations and red for non-significant. This analysis identifies potential prognostic markers in CRC, emphasizing the roles of *C6orf223*, *MYC*, and others in patient survival outcomes.

DISCUSSION

The findings of this study are aligned with previous research indicating that specific lncRNAs can act as potent oncogenic or tumor-suppressive factors in various cancers, including CRC. For instance, a study by Xu et al (2022) demonstrated that the lncRNA *MALAT1* promotes CRC cell proliferation and invasion through interactions with *miR-150*, functioning similarly to how *C6orf223* might sequester microRNAs, as observed in our analysis.[14,15]Our data, showing the upregulation of *C6orf223* in high-risk CRC samples, suggest a similar oncogenic role, where *C6orf223* modulates tumor-suppressive pathways, promoting cancer cell survival. This supports Puzkowska et al's findings on the critical regulatory role of lncRNAs in microRNA pathways within the CRC tumor microenvironment.[16,17]

The differential expression of *C6orf223* in high-risk versus low-risk CRC patients echoes observations made by Nomiri et al, who found that elevated lncRNA expression levels often correlate with aggressive tumor characteristics and poor prognosis in CRC patients.[18]In their study, high lncRNA expression was associated with enhanced proliferation markers and resistance to apoptosis, findings that mirror the upregulation of oncogenic pathways observed in our high-risk CRC cluster analysis.[19] Furthermore, our clustering analysis revealed distinct expression patterns in high-risk groups enriched with genes related to cell proliferation and immune evasion, in agreement with Nomiri's et al's observations on the molecular signatures of high-risk CRC profiles.

Pathway analysis in our study highlighted the activation of Wnt, PI3K-Akt, and MAPK signaling pathways in high-risk CRC samples, consistent with findings by Pan et al, who identified these pathways as central to CRC progression.[20,21] Specifically, the Wnt pathway, often dysregulated by mutations in the APC gene, plays a crucial role in CRC and was found to be activated in our high-risk samples, potentially driven by *C6orf223* expression. Pan et al.'s work also pointed out that the PI3K-Akt pathway can interact with specific lncRNAs to promote cell growth and survival, similar to the upregulation we observed in high-risk patients.[22] Thus, our results reinforce the concept that lncRNAs like *C6orf223* could facilitate the dysregulation of key oncogenic pathways, further validating their role in CRC aggressiveness.

The prognostic value of *C6orf223* identified in this study is further supported by recent research emphasizing the clinical utility of lncRNAs as prognostic markers in CRC. A meta-analysis by Dastmaslchi et al found that lncRNAs, particularly those involved in miRNA regulation, can serve as significant predictors of survival in CRC patients.[23] In our study, Kaplan-Meier survival analysis revealed that high *C6orf223* expression correlates with shorter overall survival, indicating its potential as a biomarker for poor prognosis. Dastmaslchi et al's meta-analysis also suggested that high-risk patients with elevated lncRNA expression might benefit from targeted therapies, an idea that our findings support, especially considering *C6orf223*'s potential role in influencing key oncogenic pathways.

Our ROC curve analysis, which demonstrated the diagnostic potential of *C6orf223* is comparable to the work of He et al, who developed an lncRNA-microRNA panel for CRC screening.[24] Their panel, comprising several lncRNAs and miRNAs, showed high diagnostic accuracy in distinguishing CRC stages and patient risk groups. Similarly, our combined biomarker panel exhibited an AUC of 0.90, indicating strong sensitivity and specificity for high-risk CRC detection. Lu et al.'s work supports the application of such panels for early CRC diagnosis and risk stratification, suggesting that incorporating *C6orf223* and specific microRNAs could enhance CRC screening accuracy, especially in clinical settings with limited access to invasive diagnostic procedures.[25]

The RNA methylation analysis of *C6orf223* in this study highlights the emerging role of post-transcriptional modifications in lncRNA function, consistent with findings by Onozco et al.[26] Their study demonstrated that methylation of specific lncRNAs enhances RNA stability and promotes oncogenic activity in CRC cells. In our study, methylation peaks were identified across *C6orf223*, particularly in regulatory regions, suggesting that methylation may enhance its stability and interaction with miRNAs and other biomolecules. This agrees with Wang et al.'s findings, supporting the potential of targeting RNA methylation as a therapeutic strategy to inhibit oncogenic lncRNAs in CRC.[27]

Our mutation analysis revealed a high frequency of *TP53*, *APC*, and *KRAS* mutations in high-risk CRC samples, which frequently co-occurred with elevated

C6orf223 expression.[28] This finding aligns with the work of Gao and Hao, who reported that lncRNAs often work synergistically with genetic mutations to promote tumorigenesis.[29,30] In particular, mutations in *TP53* and *APC* are well-known contributors to CRC progression, and their co-occurrence with high *C6orf223* expression in our study suggests a possible cooperative mechanism in tumor growth and survival. This insight that both genetic and transcriptomic profiles are essential for assessing cancer risk and progression, advocating for a comprehensive approach in cancer diagnostics.[31]

Finally, our Random Forest model identified *C6orf223* as the most predictive biomarker for CRC risk stratification, with an accuracy of 82%. This aligns with the findings of Liu et al, who used machine learning to identify CRC biomarkers, achieving high prediction accuracy with lncRNAs and microRNAs.[32]Zhang et al emphasized the importance of integrating lncRNAs and microRNAs for improved diagnostic models, as microRNA-lncRNA interactions provide richer molecular context.[33] Our findings support their conclusion, indicating that combining lncRNAs with miRNAs enhances the predictive power for CRC stratification, making machine learning approaches valuable for identifying key biomarkers in cancer.

This study provides critical insights into the molecular mechanisms underlying CRC progression, with a particular focus on the role of *C6orf223*. Differential expression analysis revealed that *C6orf223* is significantly upregulated in high-risk CRC patients and is strongly associated with key tumor-suppressive microRNAs. Functional analyses highlighted its involvement in essential oncogenic pathways, such as the Wnt and PI3K-Akt pathways, further supporting its role in promoting tumor aggressiveness. The methylation of *C6orf223* and its correlation with immune cell infiltration underscore its multifaceted regulatory functions, influencing both genetic and epigenetic landscapes in CRC.

Survival analysis demonstrated that high *C6orf223* expression correlates with poor overall survival, reinforcing its potential as a prognostic biomarker. Additionally, the integration of *C6orf223* with miRNAs into a diagnostic panel achieved high sensitivity and specificity, showcasing its utility in CRC risk stratification. The identification of *C6orf223* as a key biomarker, along with its association with frequent mutations such as those in *APC* and *KRAS*, highlights its central role in CRC biology.

While this study provides valuable insights into the role of C6orf223 in colorectal cancer progression, its reliance on in silico analyses and publicly available datasets represents a limitation. Future studies should include experimental validation of these findings, such as functional assays to confirm the regulatory interactions between C6orf223 and tumor-suppressive miRNAs, as well as larger, multi-center clinical cohorts to strengthen the generalizability of the results.

Our findings underscore the transformative potential of lncRNAs, such as C6orf223, in the clinical management of colorectal cancer. As a diagnostic biomarker, C6orf223 can be integrated into non-invasive screening panels to improve early detection and risk stratification, enabling timely intervention for high-risk patients. Therapeutically, targeting C6orf223 offers a promising approach, whether through direct inhibition of its expression, disruption of its interactions with tumor-suppressive microRNAs, or modulation of its methylation patterns to reduce oncogenic activity. Such strategies could be particularly valuable for addressing treatment-resistant or aggressive colorectal cancer phenotypes. Future research should explore the development of small molecules, antisense oligonucleotides, or CRISPR-based approaches to specifically target C6orf223, paving the way for personalized therapeutic interventions that complement existing treatment modalities.[34-36]

These findings lay the groundwork for further investigation into *C6orf223* as a potential therapeutic target. Its ability to modulate tumor-suppressive pathways and immune responses makes it a promising candidate for developing novel treatments. Future research should focus on validating these results in larger cohorts and exploring the therapeutic potential of targeting *C6orf223* and its associated pathways to improve CRC outcomes.

STATEMENT OF ETHICS

This study was conducted in accordance withthe ethical guidelines outlined by the Shahid Beheshti University of Medical Sciences. The study protocol was reviewed and approved by the Research Ethics Committee under the approval number IR.SBMU.MSP.REC.1402.576. All participants provided informed consent prior to their inclusion in the study, and all procedures adhered to the principles outlined in the Declaration of Helsinki.

FUNDING

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

We thank all the people who contributed to the completion of this study.

DATA AVAILABILITY

Upon reasonable request (specify contact method).

AI ASSISTANCE DISCLOSURE

Not applicable.

REFERENCES

1. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. Int J Mol Sci. 2017 Jan 19;18(1):197. doi: 10.3390/ijms18010197. PMID: 28106826; PMCID: PMC5297828.
2. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. JAMA. 2021;325(7):669–85.
3. Lewandowska A, Rudzki G, Lewandowski T, Stryjkowska-Gora A, Rudzki S. Risk factors for the diagnosis of colorectal cancer. Cancer Control. 2022;29:10732748211056692.
4. Alzahrani SM, Al Doghaither HA, Al-Ghafari AB. General insight into cancer: An overview of colorectal cancer. Mol Clin Oncol. 2021;15(6):271.
5. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol. 2021;14(10):101174.
6. Ghafouri-Fard S, Hussen BM, Gharebaghi A, Eghtedarian R, Taheri M. LncRNA signature in colorectal cancer. Pathol Pract. 2021;222:153432.
7. O'Brien SJ, Bishop C, Hallion J, Fiechter C, Scheurlen K, Paas M, et al. Long non-coding RNA (lncRNA) and epithelial-mesenchymal transition (EMT) in colorectal cancer: a systematic review. Cancer Biol Ther. 2020;21(9):769–81.
8. Liu Z, Guo C, Dang Q, Wang L, Liu L, Weng S, et al. Integrative analysis from multi-center studies identities a consensus machine learning-derived lncRNA signature for stage II/III colorectal cancer. EBioMedicine. 2022;75.
9. Ma YN, Hong YG, Yu GY, Jiang S yuan, Zhao B lun, Guo A, et al. LncRNA LBX2-AS1 promotes colorectal cancer progression and 5-fluorouracil resistance. Cancer Cell Int. 2021;21:1–13.
10. Bai H, Yan DS, Chen YL, Li QZ, Qi YC. Potential biomarkers: The hypomethylation of cg18949415 and cg22193385 sites in colon adenocarcinoma. Comput Biol Med. 2024;169:107884.
11. Zhang Y, Li L, Chu F, Zhang L, Zhang L, Wu H, et al. The tumor microenvironment in gastrointestinal adenocarcinomas revealed a prognostic and immunotherapeutic biomarker. Aging (Albany NY). 2022;14(24):10171.
12. Zhou Z, Tan F, Pei Q, Li C, Zhou Y, Li Y, et al. lncRNA SNHG4 modulates colorectal cancer cell cycle and cell proliferation through regulating miR-590-3p/CDK1 axis. Aging (Albany NY). 2021;13(7):9838.
13. Shariatmadar Taleghani A, Zohrab Beigi Y, Zare-Mirakabad F, Masoudi-Nejad A. Exploring ceRNA networks for key biomarkers in breast cancer subtypes and immune regulation. Sci Rep. 2023;13(1):20795.
14. Xu WW, Jin J, Wu X yu, Ren QL, Farzaneh M. MALAT1-related signaling pathways in colorectal cancer. Cancer Cell Int. 2022;22(1):126.
15. Uthman YA, Ibrahim KG, Abubakar B, Bello MB, Malami I, Imam MU, et al. MALAT1: a promising therapeutic target for the treatment of metastatic colorectal cancer. Biochem Pharmacol. 2021;190:114657.
16. Paszkowska A, Kolenda T, Guglas K, Kozłowska-Masłoń J, Podralska M, Teresiak A, et al. C10orf55, CASC2, and SFTA1P lncRNAs Are Potential Biomarkers to Assess Radiation Therapy Response in Head and Neck Cancers. J Pers Med. 2022;12(10):1696.
17. Chen D, Dou C, Liu H, Xu B, Hu B, Kuang L, et al. Comprehensive analysis: Necroptosis-related lncRNAs can effectively predict the prognosis of glioma patients. Front Oncol. 2022;12:929233.
18. Nomiri S, Hoshyar R, Chamani E, Rezaei Z, Salmani F, Larki P, et al. Prediction and validation of GUCA2B as the hub-gene in colorectal cancer based on co-expression network analysis: In-silico and in-vivo study. Biomed Pharmacother. 2022;147:112691.
19. Zhong X, He X, Wang Y, Hu Z, Yu D, Huang H, et al. A Hypoxia-Related lncRNA Signature Correlates with Survival and Tumor Microenvironment in Colorectal Cancer. J Immunol Res. 2022;2022(1):9935705.

20. Pan K, Xie Y. LncRNA FOXC2-AS1 enhances FOXC2 mRNA stability to promote colorectal cancer progression via activation of Ca2+-FAK signal pathway. Cell Death Dis. 2020;11(6):434.
21. Zhou H, Xiong Y, Peng L, Wang R, Zhang H, Fu Z. LncRNA-cCSC1 modulates cancer stem cell properties in colorectal cancer via activation of the Hedgehog signaling pathway. J Cell Biochem. 2020;121(3):2510–24.
22. Chu Y, Liu Z, Liu J, Yu L, Zhang D, Pei F. Characterization of lncRNA-perturbed TLR-signaling network identifies novel lncRNA prognostic biomarkers in colorectal cancer. Front Cell Dev Biol. 2020;8:503.
23. Dastmalchi N, Safaralizadeh R, Nargesi MM. LncRNAs: potential novel prognostic and diagnostic biomarkers in colorectal cancer. Curr Med Chem. 2020;27(30):5067–77.
24. He J, Wu F, Han Z, Hu M, Lin W, Li Y, et al. Biomarkers (mRNAs and non-coding RNAs) for the diagnosis and prognosis of colorectal cancer--from the body fluid to tissue level. Front Oncol. 2021;11(3):632834.
25. Lu C, Luo X, Xing C, Mao Y, Xu Y, Gao W, et al. Construction of a novel mRNA-miRNA-lncRNA network and identification of potential regulatory axis associated with prognosis in colorectal cancer liver metastases. Aging (Albany NY). 2021;13(11):14968.
26. Orozco LD, Owen LA, Hofmann J, Stockwell AD, Tao J, Haller S, et al. A systems biology approach uncovers novel disease mechanisms in age-related macular degeneration. Cell genomics. 2023;3(6).
27. Wang H, Wang C, Yao Y, Duan J, Liang Y, Shang Q. Analysis of long noncoding RNAs in the aqueous humor of wet age-related macular degeneration. Exp Eye Res. 2023;234(14):109576.
28. Grothey A, Fakih M, Tabernero J. Management of BRAF-mutant metastatic colorectal cancer: a review of treatment options and evidence-based guidelines. Ann Oncol. 2021;32(8):959–67.
29. Guo F, Li L, Yang W, Hu J fan, Cui J. Long noncoding RNA: a resident staff of genomic instability regulation in tumorigenesis. Cancer Lett. 2021;503:103–9.
30. Huo X, Han S, Wu G, Latchoumanin O, Zhou G, Hebbard L, et al. Dysregulated long noncoding RNAs (lncRNAs) in hepatocellular carcinoma: implications for tumorigenesis, disease progression, and liver cancer stem cells. Mol Cancer. 2017;16(5):1–10.
31. Li Y, Chen J, Zhang J, Wang Z, Shao T, Jiang C, et al. Construction and analysis of lncRNA-lncRNA synergistic networks to reveal clinically relevant lncRNAs in cancer. Oncotarget. 2015;6(28):25003.
32. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. Nat Commun. 2022;13(1):816.
33. Zhang X, Wang J, Li J, Chen W, Liu C. CRlncRC: a machine learning-based method for cancer-related long noncoding RNA identification using integrated features. BMC Med Genomics. 2018;11:99–112.
34. Dastmalchi N, Safaralizadeh R, Nargesi MM. LncRNAs: potential novel prognostic and diagnostic biomarkers in colorectal cancer. Curr Med Chem. 2020;27(30):5067–77.
35. He J, Wu F, Han Z, Hu M, Lin W, Li Y, et al. Biomarkers (mRNAs and non-coding RNAs) for the diagnosis and prognosis of colorectal cancer—from the body fluid to tissue level. Front Oncol. 2021;11:632834.
36. Pan K, Xie Y. LncRNA FOXC2-AS1 enhances FOXC2 mRNA stability to promote colorectal cancer progression via activation of Ca2+-FAK signal pathway. Cell Death Dis. 2020;11(6):434.