Identification of Mirna Signature and Key Genes in Colorectal Cancer Lymph Node Metastasis

Xi Wang
Second Affiliated Hospital of Soochow University
Guangyu Gao
Second Affiliated Hospital of Soochow University
Zhengrong Chen
Second Affiliated Hospital of Soochow University
Zhihao Chen
Second Affiliated Hospital of Soochow University
Mingxiao Han
Second Affiliated Hospital of Soochow University
Xiaolu Xie
Second Affiliated Hospital of Soochow University
Qiyuan Jin
Second Affiliated Hospital of Soochow University
Hong Du
Second Affiliated Hospital of Soochow University
Zhifei Cao
Second Affiliated Hospital of Soochow University
Haifang Zhang (haifangzhang@sina.com)
The Second Affiliated Hospital of Soochow University https://orcid.org/0000-0001-6434-3561

Primary research

Keywords: MicroRNA, colorectal cancer, lymph node metastasis, prognostic signature, HS3ST2

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-385507/v1

Version of Record: A version of this preprint was published at Cancer Cell International on July 7th, 2021. See the published version at https://doi.org/10.1186/s12935-021-02058-9.

Abstract

Background: miRNAs and mRNAs can serve as biomarkers for the diagnosis, prognosis and therapy of colorectal cancer (CRC), whose metastasis to lymph node is closely related to the poor prognosis. The current study aimed to identify the novel gene signatures in the lymph node metastasis of CRC.

Methods: GSE56350, GSE70574 and GSE95109 were downloaded from the Gene Expression Omnibus (GEO) database and 569 colorectal cancer statistics were also downloaded from the The Cancer Genome Atlas (TCGA) database. Differentially expressed miRNAs (DE-miRNAs) were calculated by using R software. Besides, gene ontology and Enriched pathway analysis of target mRNAs were analyzed by using FunRich. Furthermore, the mRNA-miRNA network was constructed using Cytoscape software. Gene expression level was verified by GEO datasets and forty paired lymph node non-metastasis CRC tissues and lymph node metastatic CRC tissues obtained from patients with CRC using quantitative real-time PCR (qPCR) .

Results: In total, five DE-miRNAs were selected, and 34 mRNAs were identified after filtering. Moreover, 2 key miRNAs and one gene were identified including hsa-miR-99a, has-miR-100 and heparan sulfate-glucosamine 3-sulfotransferase 2 (HS3ST2). The GEO datasets analysis and qPCR results showed the expression of key miRNA and genes were consistent with that in the bioinformatic analysis. A novel miRNA-mRNA network, hsa-miR-99a-HS3ST2-has-miR-100 was found in lymph node metastasis of CRC after expression analysis, prognostic prediction and experiments confirmation.

Conclusions: In summary, the potential miRNAs and genes were found and a novel miRNA-mRNA network was established in CRC lymph node metastasis by systematic bioinformatic analysis and experiments validation, which may be used as potential biomarkers in the development of lymph node metastatic CRC.

Background

Colorectal cancer (CRC) is a serious health threat worldwide. Compared with the early stage, the response and overall survival of patients with advanced CRC are still very poor. The 5-year survival rate of patients with advanced CRC is reduced from 50–10% [1]. Surgical tumor resection is still the cornerstone of the treatment of local advanced CRC. There is no curable treatment for metastatic tumors that cannot be surgically removed, as well as those with poor chemotherapy and radiotherapy effects [2]. At present, AJCC's TNM staging system has limited value in predicting recurrence [3, 4]. Moreover, lymph node metastasis is not only the main type of metastasis in patients with advanced CRC, but also one of the most essential prognostic risk factors [5]. Therefore, it is urgent to find out the key factors influencing the lymph node metastasis of CRC, so as to promote the prognosis evaluation and individualized treatment.

MicroRNAs (miRNAs) are small, noncoding RNAs, having the function of regulating after gene transcription [6]. According to previous studies, miRNAs can regulate many target genes or one type of miRNAs can be regulated by many genes [7]. In particular, Sin T. K et al found that some miRNAs can

improve the therapeutic effect by improving the drug sensitivity of cancer cells [8]. Ma X et al found that miR-374a, miR-92a, and miR-106a increase drug resistance and promote growth and metastasis of lung cancer [9]. Kania EE et al also reported that miR-9-3p and miR-9-5p decrease DNA topoisomerase IIα expression levels in acquired resistance to etoposide and may act as biomarkers of responsiveness to TOP2-targeted therapy [10]. However, the mechanisms of miRNAs in the adenoma transformation to adenocarcinoma remain unknown.

In recent years, accumulating studies have well documented that bioinformatics analysis have provided a deeper understanding of the aberrant genetic pathways in the development, progression, and metastasis of various human cancers including breast cancer [11], lung cancer [12], liver cancer [13], CRC [14], and so on. However, systematic analysis of mRNAs and miRNAs in CRC lymph node metastasis is still not enough. In this study, we conducted systematic bioinformatic analysis and identified two miRNAs (hsa-miR-99a and hsa-miR-100) and one gene (heparan sulfate-glucosamine 3-sulfotransferase 2, HS3ST2), and established a novel mRNA-miRNA regulatory network in lymph node metastasis of CRC, which may be used in the early diagnosis and therapy of metastatic CRC.

Materials And Methods

Microarray data

The Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) database is a public functional genomics data set to help users download statistics and import gene expression. In our research, Gene expression profile data (GSE56350, GSE70574 and GSE95109) were obtained from GEO.

GSE56350, including 8 primary CRC tissues derived from stage II–III CRC patients with (n = 20) or without (n = 15) lymph node metastasis. MicroRNA expression profiling analysis of these samples was performed on Agilent-021827 Human miRNA Microarray [miRNA_107_Sep09_2_105]. Dataset GSE70574, 16 T1-stage CRCs (7 lymph node-positive and 9 lymph node- negative tumors), were processed by Agilent-031181 Unrestricted_Human_miRNA_V16.0_Microarray 030840 (Feature Number version). GSE95109, 13 patients in lymph node negative subgroup and 9 patients in lymph node positive subgroup. The mRNA profiles of all 22 patients were analyzed by Agilent microarray technology to explore the different expression between lymph node negative subgroup and lymph node positive subgroup.

Differently expressed miRNAs and mRNAs analysis

R software was used to compare two groups of tissues. Besides, $|\log2FC| \geq 1$ and $P<0.05$ were used as a cut-off criterion and a significant statistical difference would be considered if the statistics met our standards [15].

Functional and pathway enrichment analysis

FunRich (http://www.funrich.org) is a publicly accessible software with the ability to identify the enriched transcription factors and perform Gene Ontology (GO) functional analysis of upload differentially expressed miRNAs (DE-miRNAs). In this study, transcription factors enrichment analysis was carried out and 10 transcription factors that may regulate DE-miRNAs were identified. Besides, we also used this software to obtain target genes of DE-miRNAs. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed by using Cytoscape software. ClueGO is a Cytoscape plug-in that visualizes the non-redundant biological terms for large clusters of genes in a functionally grouped network. The network graph of ClueGO is created based on kappa statistics. Each node in the graph represents a term. The connection between nodes reflects the correlation between terms, while the color of nodes reflects the enrichment and classification of the node (which functional group it belongs to).

PPI network and clustered subnetworks construction

The exploration of protein interactions helps to reveal the underlying pathological mechanism of CRC. In this study, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) to construct a protein–protein interaction network.

Prediction of miRNA target genes and miRNA-mRNA regulatory network

A previous study suggested that the function of miRNAs lies in the regulation of target genes. Therefore, the prediction of target genes is particularly important. It can indirectly understand the biological function and enrichment pathway of miRNAs. The miRNA enrichment function in FunRich was used to achieve miRNA targeting prediction. Besides, GSE95109 was downloaded from GEO database. By combining the results from FunRich and difference analysis of GSE95109, screened genes were identified and the miRNA-mRNA regulatory network was built by using Cytoscape software.

Construction of a prognostic signature model

To find out the influence of differential expression of miRNAs on the prognosis of CRC patients, the univariate and multivariate Cox proportional risk regression analysis was carried out for different expression of miRNAs, the miRNAs related to CRC prognosis were selected, and the risk linear model based on The Cancer Genome Atlas (TCGA) data set was established. 569 CRC patients downloaded from TCGA were randomly divided into training groups and test groups. We built a predictive feature model in the training group and tested its validity in the test group. First, the training group was analyzed by univariate Cox regression analysis to select the prognoses related DEGs. In the further functional analysis and development of potential risk characteristics, the least absolute shrinkage and selection operator (LASSO) method was used to regress the high-dimensional prediction factors [16, 17]. The "glmnet" packet in R was used to calculate the coefficient and partial likelihood deviation [18]. Through multivariate Cox regression analysis, we further studied these miRNAs to identify significant targets and build a risk linear model. To better understand the relationship between the selected miRNAs and the prognosis of CRC patients, we constructed a risk prediction model. By using “survival ROC” package, we

calculated AUC (Area Under Curve) of 3 years and 5 years dependent ROC (Receiver Operating Characteristic) curve to assess the predictive power of identified miRNAs.

Verification of miRNA expression and mRNA expression

Twenty lymph node non-metastasis CRC tissues and twenty lymph node metastatic CRC tissues were obtained from patients with CRC. We obtained the approval from the ethics committee of Second Affiliated Hospital of Soochow University. Total RNA was extracted from the tissues using Trizol reagent (Mesgen Biotech Co., Shanghai, China). The reverse transcription was performed according to the manufcturer's protocol. Real-time PCR was performed on the QuantStudio 5 Real-Time PCR System (Thermo Fisher) using qPCR SYBR Green master mix (Vazyme Biotech Co., Nanjing, China). The primers are showed in Table S1. The expression levels of miRNAs were normalized to U6 (internal standard control) and calculated using the $2^{-\Delta\Delta Ct}$ method. All experiments were performed in triplicate.

Hematoxylin and eosin (H&E) staining and analysis.

Fresh colorectal carcinoma and lymph node tissues were fixed in 10% formalin and embedded in paraffin before sectioning and staining. Tissue sections 4 μm thick were deparaffinized in xylene and rehydrated in ethanol series. H&E staining was performed according to standard protocols.

Statistical analysis

Lymph node negative test and lymph node positive test were performed to evaluate the statistical significance between the two groups. All data analysis was performed using R software (version 3.6.6) and GraphPadPrism6 package (GraphPad Software, Inc., La Jolla, California, USA). Correlation between miRNA and its possible targeted mRNA among individual samples was assessed. A P value of <0.05 was viewed as statistically significant. The correlation analysis between the RT-qPCR results and RNA-seq results was calculated in Excel 2013 (Microsoft Corporation, Redmond, WA, USA) with the function of CORR.

Results

Identification of the miRNAs between lymph node non-metastasis tissues and lymph node metastasis tissues

A total of three GEO datasets were selected and downloaded, and R software was used to research the gene expression profiles from the GEO datasets GSE56350, GSE70574 and GSE95109. In total, 47 DE-miRNAs (13 upregulated and 34 downregulated) and 30 DE-miRNAs (13 upregulated and 17 downregulated) were obtained from datasets GSE56350 and GSE70574, respectively, and 34 DEGs (29 upregulated and 5 downregulated, Table 1) were identified from dataset GSE95109 according to the cut-off criteria ($P<0.05$ and $|\log2FC|\geq1$). The identified DE-miRNAs and DEGs were shown in heat map as well as the volcano plot (Figure 1A-1F). Furthermore, a total of 5 common DE-miRNAs were screened from

GSE56350 and GSE70574 datasets, including 5 downregulated miRNAs (Figure 2). The detailed information of the 5 DE-miRNAs was listed in Table 2.

Screening of potential transcription factors and enrichment analysis

To identify the shared transcriptional factors signatures of DE-miRNAs, FunRich was used. As shown in Figure 3A, the top ten enriched transcription factors were SP1, TEAD1, TCF3, SOX1, HNF4A, TFAP4, KLF7, NHLH1, HENMT1, and RREB1. To know more about the functions of these miRNAs, we uploaded these data into FunRich to perform GO analysis. The result showed that DE-miRNAs were most enriched in the cytoplasm, nuleoside, transcription factor activity, and RNA binding (Figure 3B). KEGG pathway analysis of these target genes was performed by using Cytoscape and ClueGO. These selected mRNAs were mainly enriched in 5 pathways including Fatty acid elongation, MAPK signaling pathway, Autophagy, Signaling pathways regulating pluripotency of stem cells, and Th17 cell differentiation (Figure 3C). Besides, the construction of a protein-protein interaction (PPI) network was shown in Figure 4. This networks included 598 nodes and 1004 edges, under the conditions that the comprehensive Gt score > 0.7 and unconnected points were removed.

Construction of miRNA-mRNA regulatory network

We predicted the potential target genes of screened DE-miRNAs using FunRich software, and 598 potential target genes were obtained. The following Venn Diagram analysis of target mRNA and differentially expressed genes (DEGs) of GSE 95109 were performed and one gene was identified (Figure 5A). To show the composition and relationship of target genes more intuitively, a complete network of target genes was constructed by using Cytoscape. Finally, two essential miRNA-mRNA regulatory networks were identified which implied the crucial effect of lymph node metastasis CRC (Figure 5B).

Construction of prognostic risk model and predictability assessment

To identify the best prognostic miRNAs, the LASSO Cox regression algorithm was applied for 20 prognostic-related miRNAs. 9 miRNAs were selected to build the risk signature based on the minimum standard (Figure 6A). Then, multivariate Cox proportional risk regression analysis was carried out for 9 candidate prognostic miRNAs to assess their independent prognostic values. According to Cox model, 7 candidate miRNAs (hsa-miR-125a-5p, hsa-miR-377, hsa-miR-100, hsa-miR-455-3p, hsa-miR-126, hsa-miR-199a, and hsa-miR-99a) were selected as independent significant prognostic factors. Seven prognostic miRNAs were then combined to build a model to predict patient outcomes. The AUC of 3 years survival for the 6-miRNA signature achieved 0.809 and the AUC of 5 years survival achieved 0.981, which proved that the model has good performance in predicting the survival risk of CRC patients (Figure 6B). According to this risk model, patients were divided into high and low risk groups. The results show that this model can well predict the clinical outcomes of patients. We also analyzed the risk score, survival status and distribution of 7 miRNAs expressions in each patient (Figure 6C, D).

Validation of the expression of DE-miRNAs and DEGs

In order to further identify the key genes in CRC lymph node metastasis, we assessed the expression of hsa-miR-99a, hsa-miR-100, and HS3ST2 using GEO database. Two different GEO datasets (GSE108153 and GSE126093) both showed that hsa-miR-99a and hsa-miR-100 were down-regulated in CRC tissues compared with the normal tissues (Figure 7A-7D, P<0.001), while HS3ST2 were up-regulated in CRC tissues compared with the normal tissues using two datasets GSE146587 and GSE110224 (Figure 7E-7F, P<0.001, P<0.01). Moreover, the expression of hsa-miR-99a, hsa-miR-100, and HS3ST2 was also validated by qPCR using forty CRC tissues with lymph node metastasis and forty CRC tissues without lymph node metastasis (Figure 8A-8D). As shown in Figure 8E-8G, the expression of hsa-miR-99a and hsa-miR-100 decreased significantly in CRC tissues with lymph node metastasis compared with CRC tissues without lymph node metastasis (P<0.01, P<0.05), while the levels of HS3ST2 greatly increased in CRC tissues with lymph node metastasis (P<0.05). Generally, our data indicated that has-miR-100, hsa-miR-99a, and HS3ST2 could be the candidate biomarkers for CRC lymph node metastasis.

Discussion

CRC can occur through the progression of adenomas, which is the result of epithelial cell genetic and epigenetic events. Some microarray articles have identified gene expression profiles in adenomas and cancers [19–21]. In this study, 6 cancer related gene expression patterns were identified, suggesting that these 27 miRNAs play a role in promoting the development of CRC. Function annotation indicated that these 27 miRNAs were primarily related to cytoplasm, nuleoside, transcription factor activity, and RNA binding, which is consistent with the recognition that transcription factor activity and RNA binding are the main causes of tumor occurrence and development [22, 23]. The cytoplasm in cancer cells is essentially different from that in normal cells [24]. KEGG pathway analysis indicated that these identified 27 genes were mainly enriched in 5 pathways including fatty acid elongation, MAPK signaling pathway, autophagy, signaling pathways regulating pluripotency of stem cells, and Th17 cell differentiation. MAPK signaling pathway participates in a diverse array of important cellular processes, including the survival, proliferation, differentiation, and activation of different cell types [25, 26]. It was reported that miR-146a acts as an important molecular brake which blocks the autocrine IL-6- and IL-21-induced Th17 differentiation pathways in autoreactive CD4 T cells, highlighting its potential as a therapeutic target for treating autoimmune diseases [27]. In addition, the autophagy can promote cancer through suppressing p53 and preventing energy crisis and cell death [28]. However, these pathwaysr have not been reported in the pathogenesis of CRC in lymph node metastasis of CRC yet.

miRNA-mRNA regulatory network was built Based on FunRich and Cytoscape. 5 DE-miRNAs (hsa-miR-100, hsa-miR-375, hsa-miR-125b, hsa-miR-143 and hsa-miR-99a) and 1 potential DEG were identified by combining two screening results. It was previously reported that down-regulation of hsa-miR-100 in non-small cell lung cancer (NSCLC) altered the expression of GRP78 and spliced XBP1 level Within UPR Pathway. Hsa-miR-100 may affect endoplasmic reticulum stress and lung cancer and be used as a diagnostic biomarker of activated unfolded protein response in NSCLC [29]. In addition, low expression of miR-99a significantly predicts poor prognosis in head and neck squamous cell carcinoma and regulates cancer cell migration and invasion [30]. Xu X *et al*. also demonstrated that miR-99a inhibited the

migratory and invasive abilities by regulating the expression of insulin-like growth factor 1 receptor and the miR-99a/IGF1R axis may provide novel insight into the pathogenesis of gastric cancer [31]. Hsa-miR-125b, as a tumor suppressor, can contribute to prostate tumorigenesis by modulations in PI3K/AKT and MAPK/ERK signaling pathways. At the same time, these key pathways also influence prostate cancer progression [32]. Hsa-miR-125b also can inhibit the development of bladder cancer by inhibiting SIRT7 and MALAT1 [33]. Besides, hsa-miR-125b was found to play essential role in progression of OSCC, as well as the target genes and transcription factors of hsa-miR-125b [34]. In this study, survival analysis indicated that overexpression of hsa-miR-125b were related to worse overall survival in patients with CRC by using KM-plot software.

Through the combination of the GEO and TCGA analysis results, 2 miRNAs (has-miR-100 and hsa-miR-99a) were especially focused and We found that has-miR-100 and hsa-miR-99a have different expression in lymph node non-metastasis tissues and lymph node metastasis tissues by qPCR. Intrestingly, has-miR-100 and hsa-miR-99a all target HS3ST2. HS3ST2, an enzyme mediating 3-O-sulfation of heparan sulfate, presents in all cell types and tissues and functionally interacts with growth factors, tyrosine kinase receptors, matrix metalloproteinases and extracellular matrix proteins to modulate cell adhesion, proliferation and motility [35, 36]. In breast cancer, CRC, lung cancer, cervical cancer, pancreatic cancer and recurrent prostate cancer, HS3ST2 is silenced due to hypermethylation, suggesting it may play an important role [37]. Previous study confirmed that abnormal methylation level of HS3ST2 is important in endometrial hyperplasia and carcinogenesis [38]. The prognostic significance of HS3ST2 mRNA expression in several cancer types has been evaluated [39, 40]. HS3ST2 protein expression could be used as a favorable prognostic tissue biomarker in patients with primary advanced-stage lung cancer [40]. For gastric cancer, statistical analyses using a chi-squared test showed that there is a statistically significant difference in methylation level of HS3ST2 between gastric cancer and uncancerous patients and it may act as novel cancer-related molecular mechanisms in detection of new treatment strategies [41]. In this study, we found that HS3ST2 is upregulated in the lymph node metastatic CRC tissues compared to CRC. Moreover, a novel miRNA-mRNA network, hsa-miR-99a-HS3ST2-has-miR-100, was firstly found in lymph node metastasis of CRC.

Our findings proved that many differentially expressed mRNAs and miRNAs involved in the lymph node metastasis of CRC by certain signaling pathways and had prognostic value. Since all our data were mostly achieved from the GEO database and TCGA by bioinformatics tools, as well as the limited number of relevant samples, more data analysis, and clinical experiments should be further performed for developing these potential biomarkers for predicting the recurrence of CRC.

Conclusion

Our study concluded certain mechanisms of the development of CRC. Plenty of differentially expressed mRNAs and miRNAs were identified between lymph node non-metastasis tissues and lymph node metastasis tissues by using bioinformatics methods. Also, has-miR-100, hsa-miR-99a and HS3ST2 were identified as potential biomarkers for predicting the recurrence of CRC.

Abbreviations

GEO: Gene Expression Omnibus; KEGG: Kyoto Encyclopedia of Genes and Genomes; miRNA: microRNAs; TCGA: The Cancer Genome Atlas; ROC: receiver operating characteristic; AUC: area under the curve; qPCR: quantitative real-time PCR; DE-miRNAs: differentially expressed miRNAs; DEGs: differentially expressed genes; CRC: colorectal cancer; HS3ST2: heparan sulfate-glucosamine 3-sulfotransferase 2.

Declarations

We obtained the approval from the ethics committee of Second Affiliated Hospital of Soochow University.

Author's contributions

Conception and design: HF Zhang, ZF Cao, X Wang

Development of methodology: X Wang, GY Gao, ZR Chen

Acquisition of data: X Wang, GY Gao, ZR Chen

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X Wang, GY Gao, ZH Chen, MX Han, H Du

Writing, review and/or revision of the manuscript: X Wang, HF Zhang, ZF Cao

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): MX Han, XL Xie, QY Jin

Study supervision: HF Zhang, ZF Cao

Acknowledgements

This study was supported by the Natural Science Foundation of Jiangsu province, China (BK20181173), Gusu health youth talents program of Suzhou (GSWS2019039), Jiangsu youth medical talents program (QNRC-866, 867), Discipline Construction Program of The Second Affiliated Hospital of Soochow University (XKTJ-TD202001), Innovation and Entrepreneurship Training Program for College Students in Jiangsu Province (202010285125Y), the Medical Research Programs of Health Commission Foundation of Jiangsu Province (H2019071), and the Science and Technology Program of Suzhou (SLT201934, SYS2020023, SS202056).

Competing Interests

The authors have declared that no competing interest exists.

Consent for publication

This study has not been published previously, and the authors agree the consent for publication.

Availability of data and materials.

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Steele, S. R.; Chang, G. J.; Hendren, S.; Weiser, M.; Irani, J.; Buie, W. D.; Rafferty, J. F., Practice guideline for the surveillance of patients after curative treatment of colon and rectal cancer. *Diseases of the Colon & Rectum* **2015**, 58 (8), 713-725, doi:10.1097/DCR.0000000000000410.
2. Vogel, J. D.; Eskicioglu, C.; Weiser, M. R.; Feingold, D. L.; Steele, S. R., The American Society of Colon and Rectal Surgeons clinical practice guidelines for the treatment of colon cancer. *Diseases of the Colon & Rectum* **2017**, 60 (10), 999-1017, doi:10.1097/DCR.0000000000000926.
3. Balch, C.; Ramapuram, J. B.; Tiwari, A. K., The epigenomics of embryonic pathway signaling in colorectal cancer. *Frontiers in pharmacology* **2017**, 8, 267, doi: 10.3389/fphar.2017.00267.
4. Dienstmann, R.; Mason, M.; Sinicrope, F.; Phipps, A.; Tejpar, S.; Nesbakken, A.; Danielsen, S. A.; Sveen, A.; Buchanan, D. D.; Clendenning, M., Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Annals of Oncology* **2017**, 28 (5), 1023-1031, doi:10.1093/annonc/mdx052.
5. Bosch, S.; Teerenstra, S.; De Wilt, J.; Cunningham, C.; Nagtegaal, I., Predicting lymph node metastasis in pT1 colorectal cancer: a systematic review of risk factors providing rationale for therapy decisions. *Endoscopy* **2013**, 45 (10), 827-841, doi: 10.1055/s-0033-1344238.
6. Akao, Y.; Nakagawa, Y.; Naoe, T., let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biological & Pharmaceutical Bulletin* **2006**, 29 (5), 903. 10.1248/bpb.29.903.
7. Ren, P.; Gong, F.; Zhang, Y.; Jiang, J.; Zhang, H., MicroRNA-92a promotes growth, metastasis, and chemoresistance in non-small cell lung cancer cells by targeting PTEN. *Tumor Biology* **2016**, 37 (3), 3215-3225, doi: 10.1007/s13277-015-4150-3.
8. Sin, T. K.; Wang, F.; Meng, F.; Wong, S.; Cho, W.; Siu, P. M.; Chan, L. W.; Yung, B. Y., Implications of microRNAs in the treatment of gefitinib-resistant non-small cell lung cancer. *International journal of molecular sciences* **2016**, 17 (2), 237, doi:10.3390/ijms17020237.
9. Ma, X.; Liang, A. L.; Liu, Y. J., Research progress on the relationship between lung cancer drug-resistance and microRNAs. *Journal of Cancer* **2019**, 10 (27), 6865-6875, doi:10.7150/jca.31952.
10. Kania, E. E.; Carvajal-Moreno, J.; Hernandez, V. A.; English, A.; Papa, J. L.; Shkolnikov, N.; Ozer, H. G.; Yilmaz, A. S.; Yalowich, J. C.; hsa-miR-9-3p and hsa-miR-9-5p as post-transcriptional modulators of DNA topoisomerase IIα in human leukemia K562 cells with acquired resistance to etoposide. *Molecular pharmacology* **2020**, 97 (3), 159-170, doi:10.1124/mol.119.118315.

11. Fridrichova, I.; Zmetakova, I., MicroRNAs Contribute to Breast Cancer Invasiveness. *Cells* **2019**, 8 (11), 1361, doi:10.3390/cells.8111361 .
12. Ma, Q.; Xu, Y.; Liao, H.; Cai, Y.; Xu, L.; Xiao, D.; Liu, C.; Pu, W.; Zhong, X.; Guo, X., Identification and validation of key genes associated with non-small-cell lung cancer. *Journal of Cellular Physiology* **2019**, 234 (12), doi: 10.1002/jcp.28839.
13. Zeng, L.; Fan, X.; Wang, X.; Deng, H.; Zhang, K.; Zhang, X.; He, S.; Li, N.; Han, Q.; Liu, Z., Bioinformatics Analysis based on Multiple Databases Identifies Hub Genes Associated with Hepatocellular Carcinoma. *Current Genomics* **2019**, 20 (5), doi:10.2174/1389202920666191011092410.
14. Liu, X.; Liu, L.; Dong, Z.; Li, J.; Yu, Y.; Chen, X.; Ren, F.; Cui, G.; Sun, R., Expression patterns and prognostic value of m6A-related genes in colorectal cancer. *American Journal of Translational Research* **2019**, 11 (7), 3972.
15. Li, L.; Wang, G.; Li, N.; Yu, H.; Si, J.; Wang, J., Identification of key genes and pathways associated with obesity in children. *Experimental & Therapeutic Medicine* **2017**, 14 (2), 1065-1073, doi:10.3892/etm.2017.4597.
16. Ikeda, N.; Nakajima, Y.; Tokuhara, T.; Hattori, N.; Sho, M.; Kanehiro, H.; Miyake, M., Clinical significance of aminopeptidase N/CD13 expression in human pancreatic carcinoma. *Clinical Cancer Research* **2003**, 9 (4), 1503-1508, doi:10.1093/carcin/bgg048.
17. Veenman, C.J.; Tax, D. M. J., LESS: a model-based classifier for sparse subspaces. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **2005**, 27 (9), 1496-500, doi:10.1109/TPAMI.2005.182.
18. Pollock, B. E.; Storlie, C. B.; Link, M. J.; Stafford, S. L.; Garces, Y. I.; Foote, R. L., Comparative analysis of arteriovenous malformation grading scales in predicting outcomes after stereotactic radiosurgery. *Journal of neurosurgery* **2017**, 126 (3), 852-858, doi: 10.3171/2015.11.JNS151300.
19. Kitahara, O.; Furukawa, Y.; Tanaka, T.; Kihara, C.; Ono, K.; Yanagawa, R.; Nita, M. E.; Takagi, T.; Nakamura, Y.; Tsunoda, T., Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer research* **2001**, 61 (9), 3544-3549, doi:10.1046/j.1523-5394.2001.009003155.x.
20. Lechner, S.; Müller-Ladner, U.; Renke, B.; Schölmerich, J.; Rüschoff, J.; Kullmann, F., Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. *Gut* **2003**, 52 (8), 1148-1153, doi: 10.1136/gut.52.8.1148.
21. Notterman, D. A.; Alon, U. A.; Sierk, A. J.; Levine, A. J., Transcriptional Gene Expression Profiles of Colorectal Adenoma, Adenocarcinoma, and Normal Tissue Examined by Oligonucleotide Arrays. *Cancer Research* **2001**, 61 (7), 3124-3130, doi: 10.1097/00002820-200104000-00012.
22. Perez, R.; Wu, N.; Klipfel, A. A.; Beart, R. W., A better cell cycle target for gene therapy of colorectal cancer: cyclin G. *Journal of gastrointestinal surgery* **2003**, 7 (7), 884-889, doi: 10.1016/j.gassur.2003.08.001.
23. Tominaga, O.; Nita, M. E.; Nagawa, H.; Fujii, S.; Tsuruo, T.; Muto, T., Expressions of cell cycle regulators in human colorectal cancer cell lines. *Japanese journal of cancer research* **1997**, 88 (9), 855-860, doi:

10.1111/j.1349-7006.1997.tb00461.x.

24. Djamgoz, M. B. A.; Coombes, R. C.; Schwab, A., Ion transport and cancer: From initiation to metastasis. *Philosophical Transactions of The Royal Society B Biological Sciences* **2014**, 369 (1638), 20130092, doi: 10.1098/rstb.2013.0092.
25. Sun, Y.; Liu, W. Z.; Liu, T.; Feng, X.; Yang, N.; Zhou, H. F., Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *Journal of Receptors and Signal Transduction* **2015**, 35 (6), 600-604, doi: 10.3109/10799893.2015.1030412.
26. Chen, S. X.; Zhao, F.; Huang, X. J., MAPK signaling pathway and erectile dysfunction. *Zhonghua nan ke xue= National Journal of Andrology* **2018**, 24 (5), 442-446.
27. Li, B.; Wang, X.; Choi, I. Y.; Wang, Y. C.; Liu, S.; Pham, A. T.; Moon, H.; Smith, D. J.; Rao, D. S.; Boldin, M. P., miR-146a modulates autoreactive Th17 cell differentiation and regulates organ-specific autoimmunity. *The Journal of clinical investigation* **2017**, 127 (10), 3702-3716, doi:10.1172/JCI94012.
28. Amaravadi, R.; Kimmelman, A. C.; White, E., Recent insights into the function of autophagy in cancer. *Genes & Development* **2016**, 30 (17), 1913-1930, doi: 10.1101/gad.287524.116.
29. Ahmadi, A.; Khansarinejad, B.; Hosseinkhani, S.; Ghanei, M.; Mowla, S. J., miR-199a-5p and miR-495 target GRP78 within UPR pathway of lung cancer. *Gene* **2017**, 620, 15-22, doi:10.1016/j.gene.2017.03.032.
30. Okada, R.; Koshizuka, K.; Yamada, Y.; Moriya, S.; Kikkawa, N.; Kinoshita, T.; Hanazawa, T.; Seki, N., Regulation of oncogenic targets by miR-99a-3p (passenger strand of miR-99a-duplex) in head and neck squamous cell carcinoma. *Cells* **2019**, 8 (12), 1535, doi:10.3390/cells8121535.
31. Xu, X.; Guo, A.; Pan, Q.; Chang, A.; Zhao, C., MiR-99a suppresses cell migration and invasion by regulating IGF1R in gastric cancer. *Eur Rev Med Pharmacol Sci* **2019**, 23 (17), 7375-7382, doi:10.26355/eurrev_201909_18845.
32. Budd, W. T.; Seashols-Williams, S. J.; Clark, G. C.; Danielle, W.; Valerie, C.; Emanuel, P.; Dragoescu, E. A.; Katherine, O.; Dual Action of miR-125b As a Tumor Suppressor and OncomiR-22 Promotes Prostate Cancer Tumorigenesis. *PLoS ONE* **2015**, 10 (11), e0142373-, doi: 10.1371/journal.pone.0142373.
33. Han, Y.; Liu, Y.; Zhang, H.; Wang, T.; Diao, R.; Jiang, Z.; Gui, Y.; Cai, Z., Hsa-miR-125b suppresses bladder cancer development by down-regulating oncogene SIRT7 and oncogenic long non-coding RNA MALAT1. *FEBS letters* **2013**, 587 (23), 3875-3882, doi: 10.1016/j.febslet.2013.10.023.
34. Yan, Z. Y.; Luo, Z. Q.; Zhang, L. J.; Li, J.; Liu, J. Q., Integrated Analysis and MicroRNA Expression Profiling Identified Seven miRNAs Associated With Progression of Oral Squamous Cell Carcinoma. J*ournal of Cellular Physiology* **2016**, doi: 10.1002/jcp.25728.
35. Afratis, N.; Gialeli, C.; Nikitovic, D.; Tsegenidis, T.; Karousou, E.; Theocharis, A. D.; Pavão, M. S.; Tzanakakis, G. N.; Karamanos, N. K., Glycosaminoglycans: key players in cancer cell biology and treatment. *The FEBS journal* **2012**, 279 (7), 1177-1197, doi:10.1111/j.1742-4658.2012.08529.x.
36. Vijaya Kumar, A.; Salem Gassar, E.; Spillmann, D.; Stock, C.; Sen, Y. P.; Zhang, T.; Van Kuppevelt, T. H.; Hülsewig, C.; Koszlowski, E. O.; Pavao, M. S., HS3ST2 modulates breast cancer cell invasiveness via

MAP kinase-and Tcf4 (Tcf7l2)-dependent regulation of protease and cadherin expression. *International journal of cancer* **2014**, 135 (11), 2579-2592, doi: 10.1002/ijc.28921.

37. Danková, Z.; Braný, D.; Dvorská, D.; Ňachajová, M.; Fiolka, R.; Grendár, M.; Hatok, J.; Kubatka, P.; Holubeková, V.; Halašová, E., Methylation status of KLF4 and HS3ST2 genes as predictors of endometrial cancer and hyperplastic endometrial lesions. *International journal of molecular medicine* **2018**, 42 (6), 3318-3328, doi:10.3892/ijmm.2018.3872.
38. Zuo, Q.; Zheng, W.; Zhang, J.; Pan, Z.; Liu, Y.; Long, H.; Fan, P.; Guo, C.; Li, F.; Shao, R., Methylation in the promoters of HS3ST2 and CCNA1 genes is associated with cervical cancer in Uygur women in Xinjiang. *The International journal of biological markers* **2014**, 29 (4), 354-362, doi:10.5301/jbm.5000107.
39. Hellec, C.; Delos, M.; Carpentier, M.; Denys, A.; Allain, F., The heparan sulfate 3-O-sulfotransferases (HS3ST) 2, 3B and 4 enhance proliferation and survival in breast cancer MDA-MB-231 cells. *PloS one* **2018**, 13 (3), e0194676, doi: 10.1371/journal.pone.0194676.
40. Hwang, J. A.; Kim, Y.; Hong, S. H.; Lee, J.; Cho, Y. G.; Han, J. Y.; Kim, Y. H.; Han, J.; Shim, Y. M.; Lee, Y. S., Epigenetic inactivation of heparan sulfate (glucosamine) 3-O-sulfotransferase 2 in lung cancer and its role in tumorigenesis. *PLoS One* **2013**, 8 (11), e79634, doi: 10.1002/ccd.1810270106.
41. Eyvazi, S.; Khamaneh, A. M.; Tarhriz, V.; Bandehpour, M.; Hejazi, M. S.; Sadat, A. T. E.; Sepehri, B., CpG islands methylation analysis of CDH11, EphA5, and HS3ST2 genes in gastric adenocarcinoma patients. *Journal of gastrointestinal cancer* **2020**, 51(2), 579-583, doi: 10.1007/s12029-019-00290-1.

Tables

Table 1. Identification of DEGs associated with lymph node metastasis of CRC.

ID	logFC	AveExpr	t	P.Value	adj.P.Val
MMP3	-3.383380607	2.743291273	8.441456328	1.62E-08	0.000255365
MMP1	-2.604327371	3.872452102	7.849649847	5.74E-08	0.000451091
CCL11	-2.545017495	4.141617407	7.216135246	2.33E-07	0.001165698
WNT5A	-2.285951543	6.469308505	-7.110177522	2.97E-07	0.001165698
PCDH18	-1.008646403	3.110155251	6.808152934	5.92E-07	0.001861911
SLAMF6	1.012364852	3.136288745	6.447210686	1.37E-06	0.00359875
HLA-DOA	1.124220113	3.019812643	6.185459043	2.55E-06	0.00573321
PVRIG	1.150652615	4.417614191	6.095269942	3.17E-06	0.006222161
TRIM22	1.169702283	4.947332919	5.9191503	4.83E-06	0.008443899
CCR7	1.219016804	2.731750916	5.803821202	6.39E-06	0.00994535
CD3E	1.281657782	5.313206737	5.75769866	7.14E-06	0.00994535
VPREB3	1.293250555	3.746465497	5.732721287	7.59E-06	0.00994535
CCDC102B	1.409808703	4.144533696	5.589740914	1.08E-05	0.012154586
LTB	1.440290668	6.533277848	5.587143365	1.08E-05	0.012154586
SASH3	1.480906594	4.121667145	-5.415593938	1.65E-05	0.017272876
C16orf54	1.498350606	5.088107616	5.3829589	1.79E-05	0.017546428
LILRA4	1.503138459	4.132907382	5.323556627	2.07E-05	0.019115674
CD69	1.531925947	3.32492286	5.279049319	2.31E-05	0.020148371
C7	1.556824402	4.306873776	5.191476469	2.86E-05	0.022719827
MS4A1	1.632143731	4.377301176	-5.119251869	3.43E-05	0.025641473
CD52	1.678266234	2.85855908	4.96443047	5.03E-05	0.035009824
CD22	1.681388143	3.364439797	4.941125861	5.33E-05	0.035009824
PCOLCE2	1.821419427	6.325012753	4.940084735	5.34E-05	0.035009824
HLA-DOB	1.856948346	3.984899196	4.907577183	5.80E-05	0.035619973
IGHM	1.942704339	3.267700289	4.893964811	5.99E-05	0.035619973
SELL	2.068994543	4.119056992	4.86316037	6.47E-05	0.03634284
TRAT1	2.101207727	4.919100006	-4.765556311	8.26E-05	0.041118009
BANK1	2.131469636	3.586958156	4.747757343	8.63E-05	0.041118009

GZMK	2.140117625	3.461884833	4.67721575	0.00010293	0.046348051
FCMR	2.187861765	4.307929701	-4.675107745	0.000103474	0.046348051
HS3ST2	2.281337525	5.157835286	4.656340675	0.000108439	0.046348051
CCL19	2.919633409	3.986878293	4.642180581	0.000112343	0.046477613
CR2	3.758336449	7.863578434	4.561609892	0.000137398	0.049439631
FDCSP	4.088409432	5.323765484	4.560403115	0.000137813	0.049439631

Table 2. Identification of DE-miRNAs associated with lymph node metastasis of CRC.

ID	logFC	AveExpr	t	P.Value	adj.P.Val
hsa-miR-99a	-3.2998	5.878924	-9.07224	1.47E-10	1.08E-07
hsa-miR-100	-2.35603	6.687555	-4.37844	0.00011	0.003868
hsa-miR-125b	-2.95247	5.767905	-4.71345	4.12E-05	0.002706
hsa-miR-143	-3.34466	5.496491	-5.96943	9.83E-07	0.000242
hsa-miR-375	-1.95009	7.292945	-3.1063	0.003834	0.049641

Figures

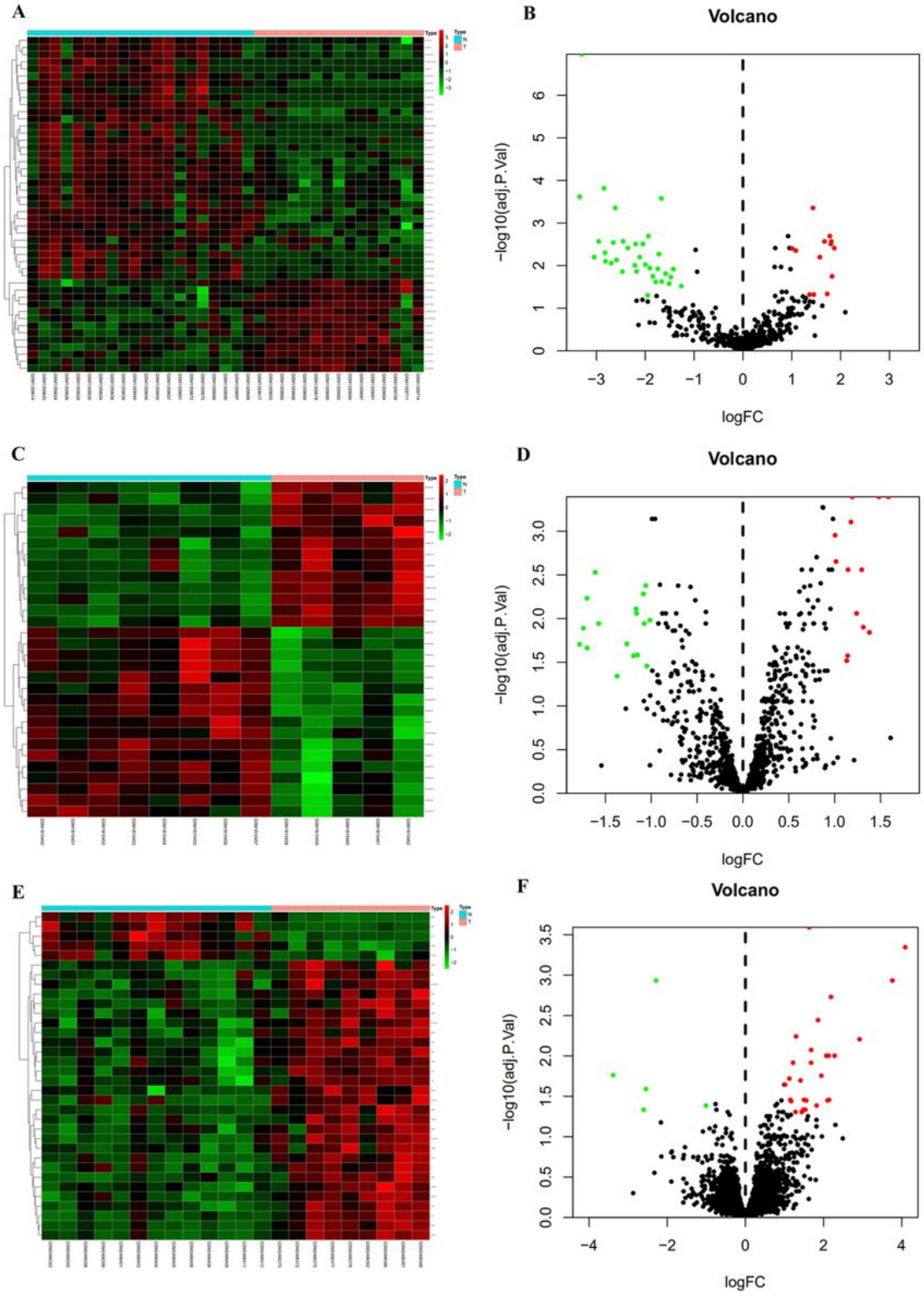

Figure 1

Volcano and heat maps of GSE56350, GSE70754 and GSE95109. (A) Unsupervised clustering analysis of differentially-expressed miRNAs (DE-miRNAs) in GSE56350. (B) Volcano plots of miRNAs in GSE56350. (C) Unsupervised clustering analysis of DE-miRNAs in GSE70754. (D) Volcano plots of miRNAs in GSE70754. (E) Unsupervised clustering analysis of the DE-mRNAs in GSE95109. (F) Volcano plots of mRNAs in GSE95109. A, C, E: Red dots indicate significantly up-regulated miRNAs or mRNAs, green dots

indicate significantly down-regulated miRNAs or mRNAs. B, D, F: Red dots indicate up-regulated DE-miRNAs or DE-mRNAs, green dots indicate down-regulated DE-miRNAs or DE-mRNAs, black dots indicate non-differentially expressed miRNAs or mRNAs.

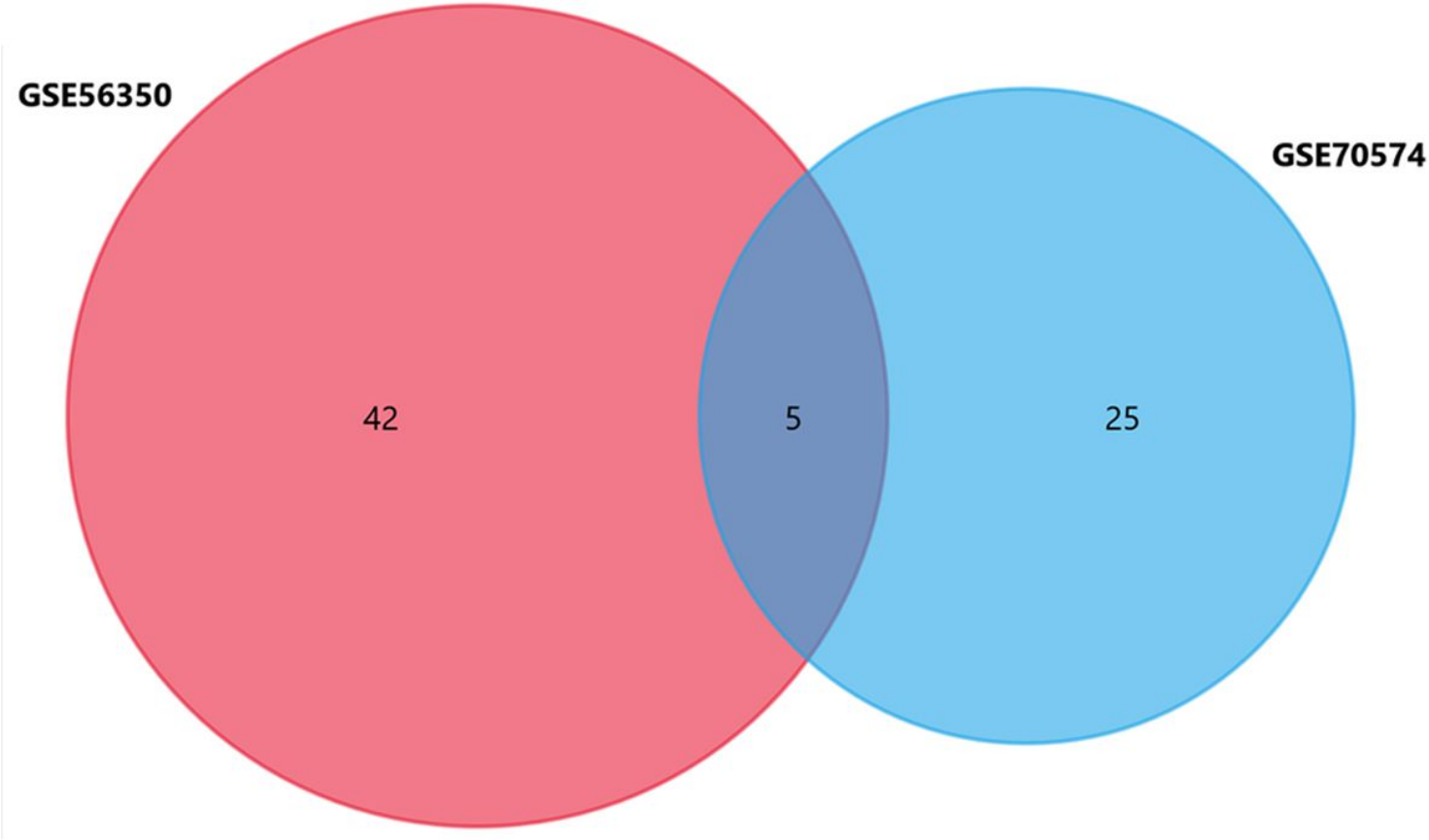

Figure 2

Venn Diagram of GSE56350 and GSE70574.

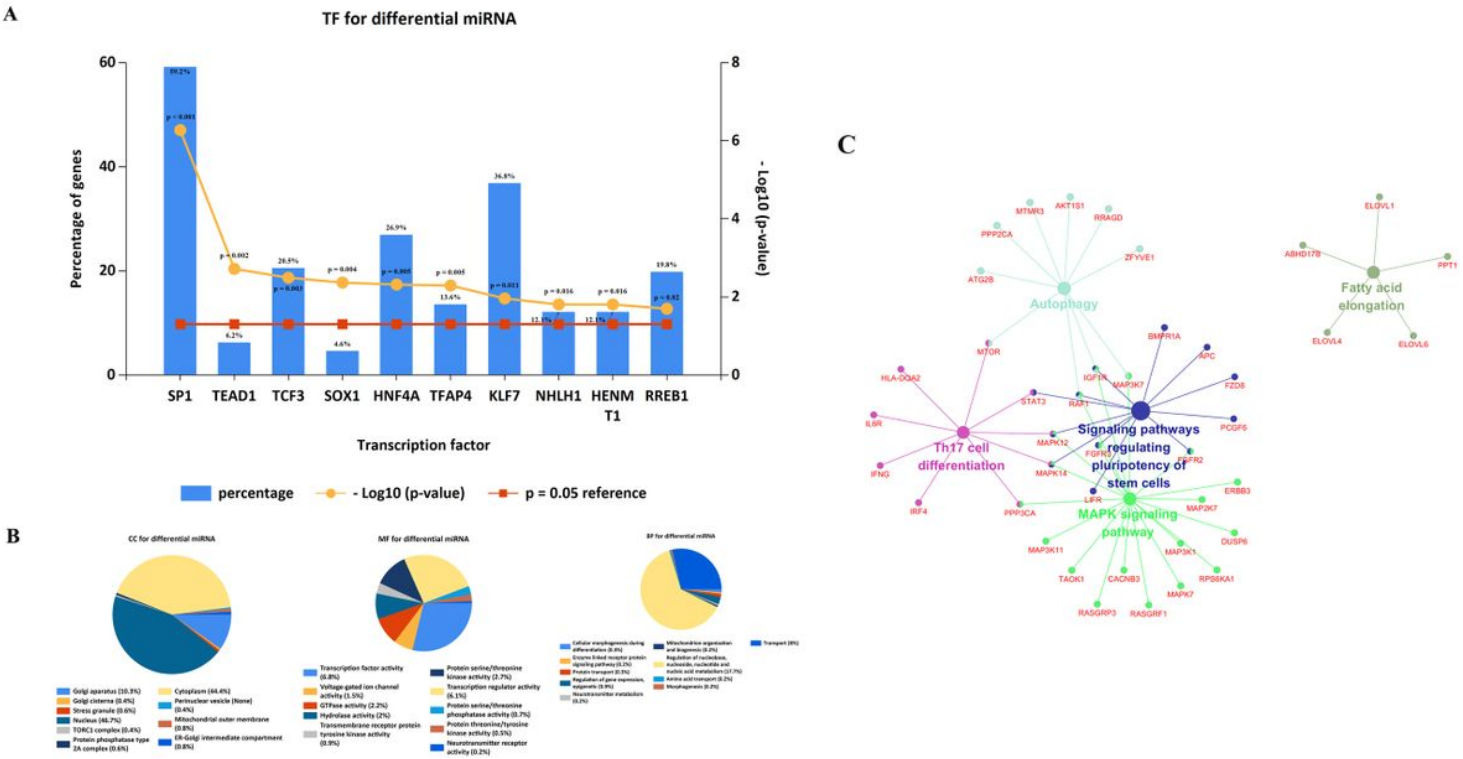

Figure 3

Screening of potential transcription factors and target genes of DE-miRNAs. (A) Identification of the potential transcription factors of DE-miRNAs by FunRich software. (B) The top 10 of biological process, cellular component, and molecular function of the target genes of DE-miRNAs. (C) KEGG pathway enriched by potential target mRNAs of DE-miRNAs.

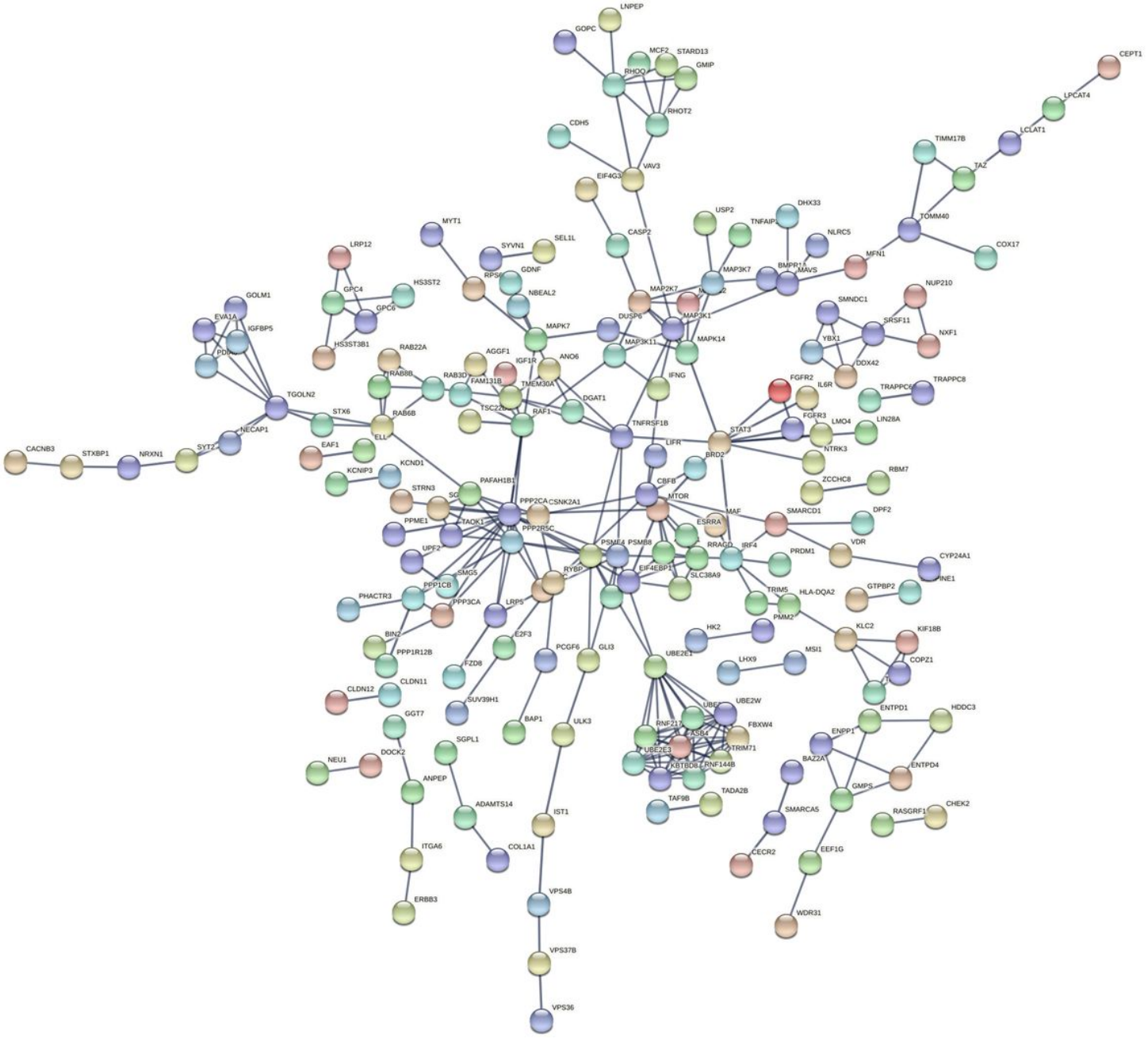

Figure 4

The PPI network of the target genes of the identified DE-miRNAs.

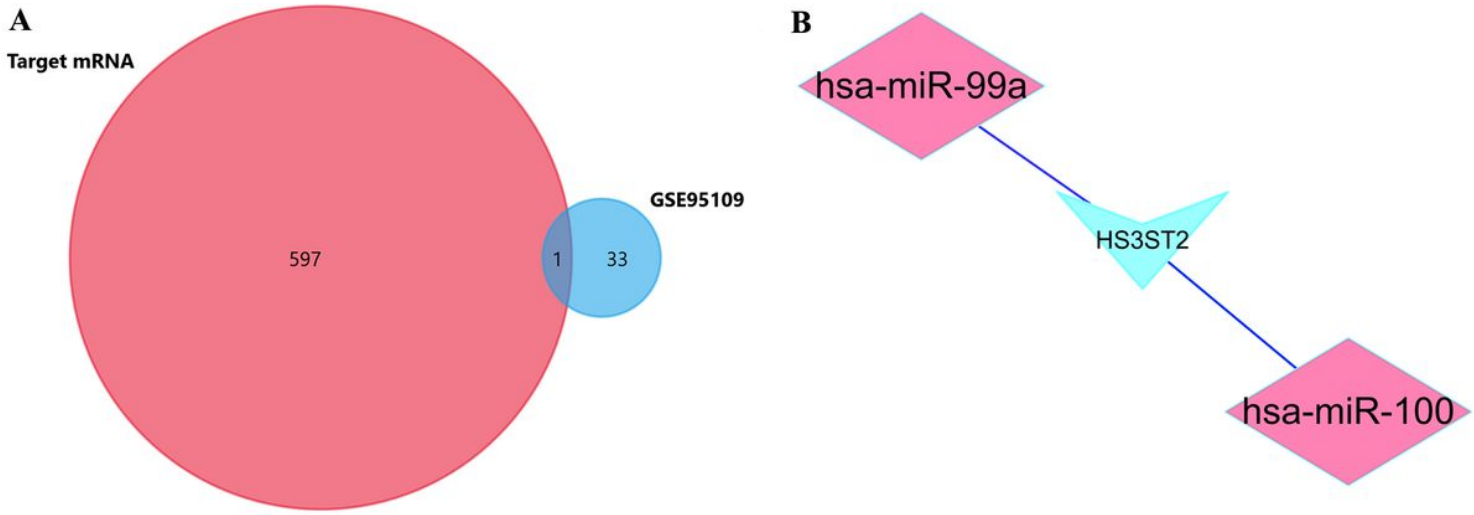

Figure 5

The miRNA-mRNA network of CRC lymph node metastasis. (A) Venn Diagram of target mRNAs of DE-miRNAs and DE-mRNAs of GSE95109. (B) The miRNA-mRNA regulatory network in the lymph node metastasis of CRC.

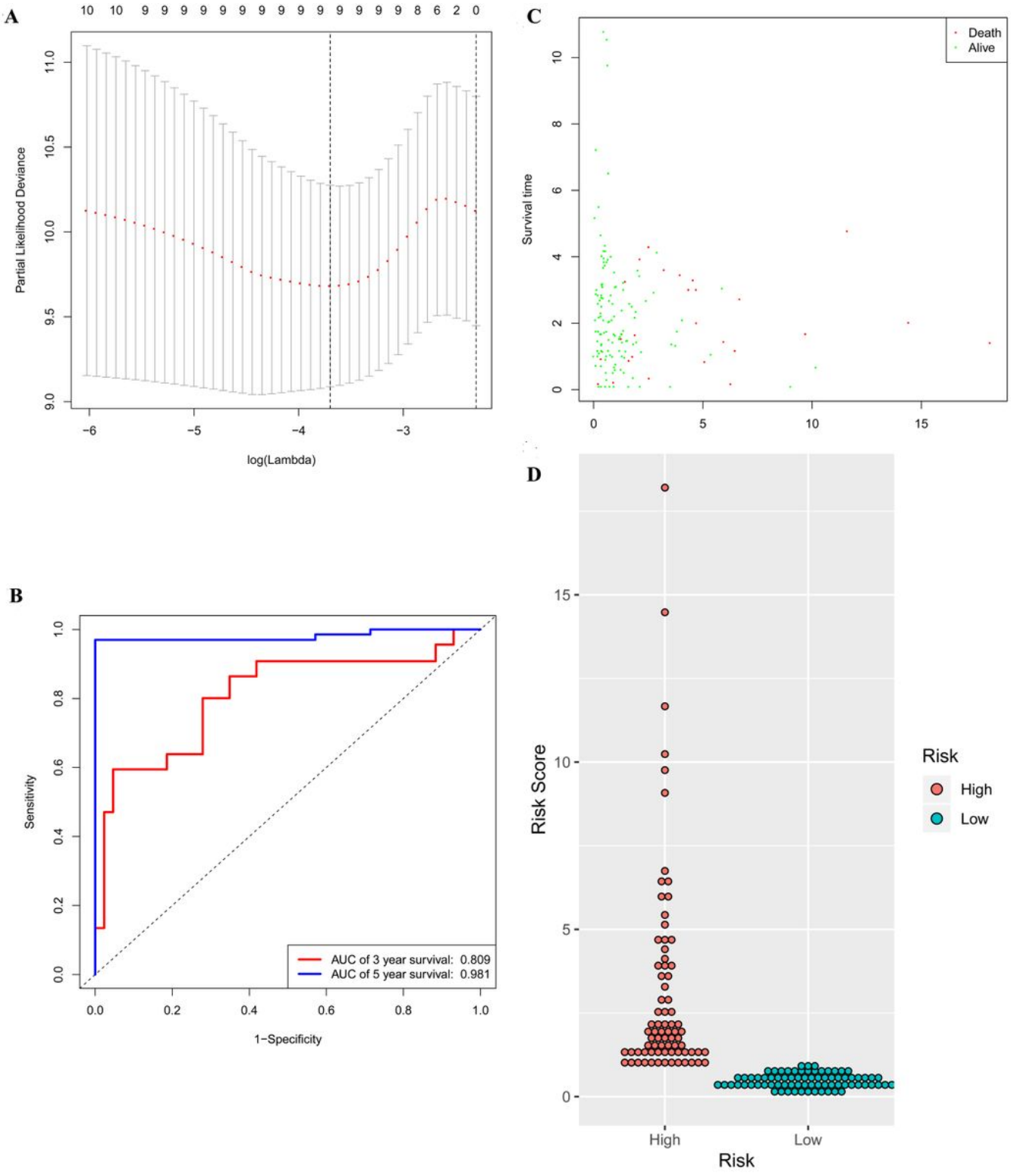

Figure 6

Identification of the signature significantly associated with the survival of patients with CRC in the training group. (A) LASSO Cox regression algorithm was used to reduce the scope. (B) Time-dependent ROC curves analysis. (C, D) Risk score distribution and survival status for patients in high- and low-risk groups by the signature. LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

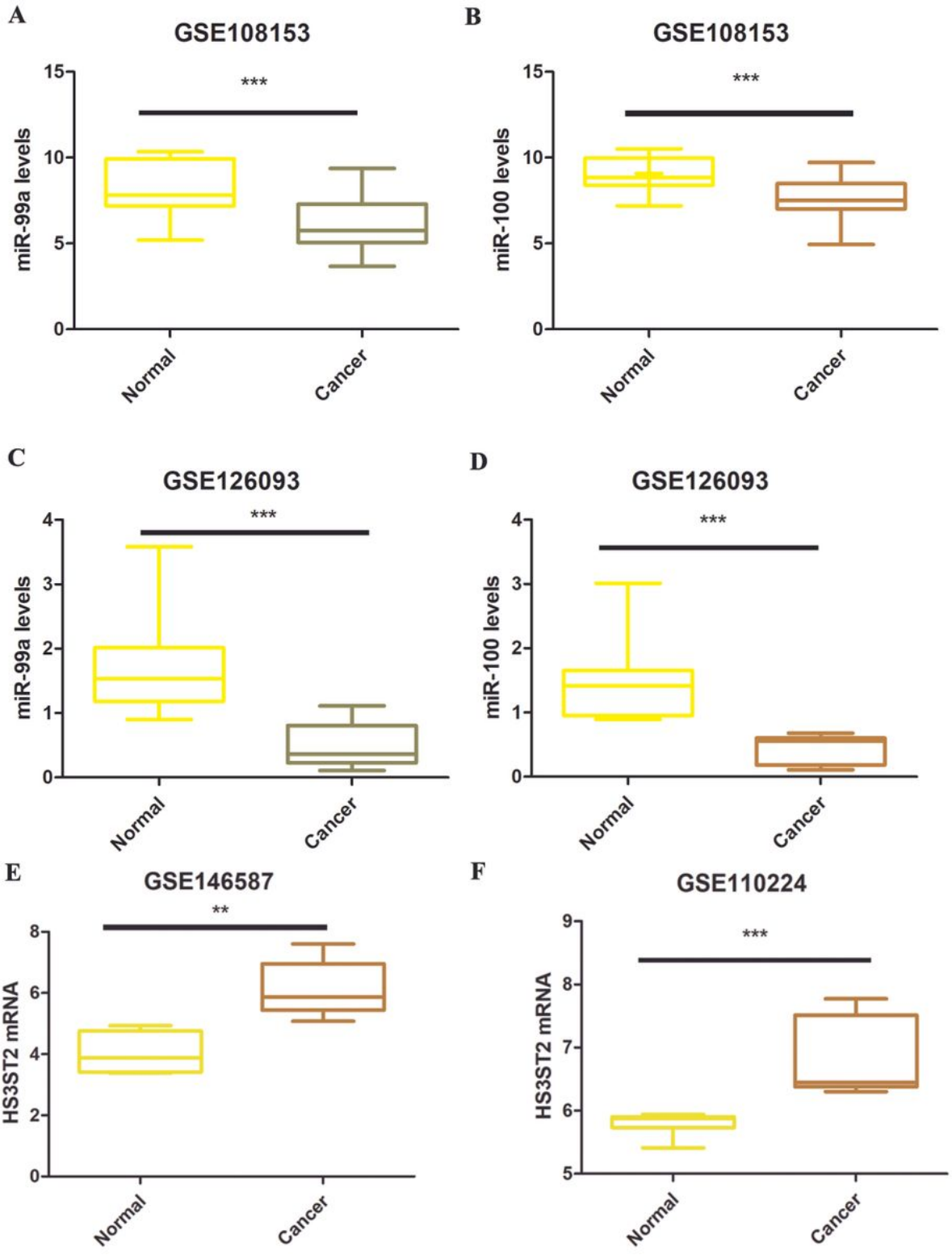

Figure 7

The differential expression of hsa-miR-99a and hsa-miR-100 in CRC tissues and their corresponding normal-appearing tissues. (A-D) Validation of hsa-miR-99a and hsa-miR-100 in GEO datasets GSE108153 and GSE126093, respectively . (E-F) Validation of HS3ST2 in GEO datasets GSE146587 and GSE110224, respectively. ** $P < 0.01$, *** $P < 0.001$

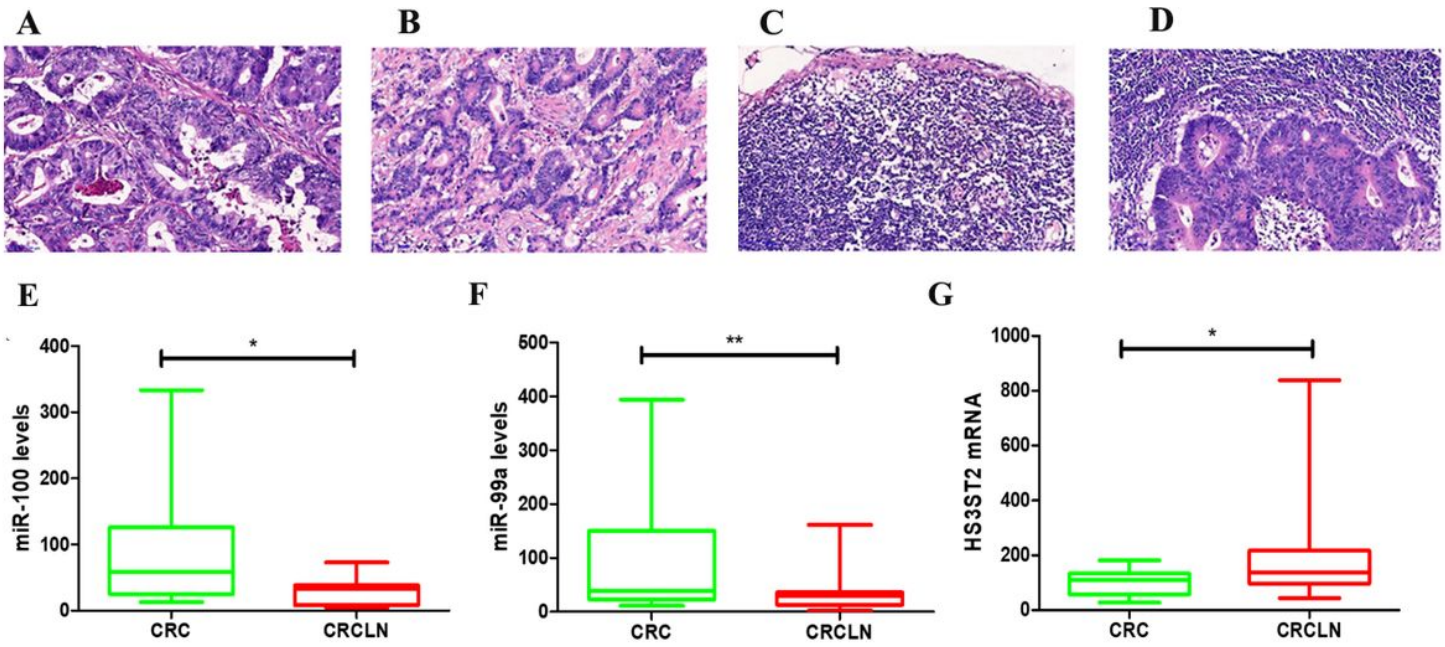

Figure 8

The expression levels of hsa-miR-100, hsa-miR-99a and HS3ST2 in CRC tissues and lymph node metastatic CRC tissues. (A) Hematoxylin and eosin (H&E) staining of CRC tissues without lymph node metastasis. (B) H&E staining of CRC tissues in patients with lymph node metastatic CRC. (C) H&E staining of lymph nodes in patients with CRC. (D) H&E staining of metastatic lymph nodes in patients with CRC. (E-G) Validation of hsa-miR-99a, hsa-miR-100 and HS3ST2 in GEO datasets GSE146587 and GSE110224, respectively.* $P < 0.05$,** $P < 0.01$

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tableS1.docx