



## OPEN ACCESS

## EDITED BY

Weiren Luo,  
The Second Affiliated Hospital of  
Southern University of Science and  
Technology, China

## REVIEWED BY

Calin Cainap,  
Iuliu Hațieganu University of Medicine  
and Pharmacy, Romania  
Qian Xu,  
China Medical University, China

## \*CORRESPONDENCE

Mingwei Liu  
liumingwei@cqmu.edu.cn

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Gastroenterology,  
a section of the journal  
Frontiers in Medicine

RECEIVED 23 February 2022

ACCEPTED 09 September 2022

PUBLISHED 27 September 2022

## CITATION

Long F, Tian L, Chai Z, Li J, Tang Y and  
Liu M (2022) Identification of  
stage-associated exosome miRNAs in  
colorectal cancer by improved robust  
and corroborative approach  
embedded miRNA-target network.  
*Front. Med.* 9:881788.  
doi: 10.3389/fmed.2022.881788

## COPYRIGHT

© 2022 Long, Tian, Chai, Li, Tang and  
Liu. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Identification of stage-associated exosome miRNAs in colorectal cancer by improved robust and corroborative approach embedded miRNA-target network

Fei Long<sup>†</sup>, Luyao Tian<sup>†</sup>, Zixuan Chai, Jing Li, Ying Tang and Mingwei Liu\*

Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China

**Background:** Colorectal cancer (CRC) is a common gastrointestinal tumor with high morbidity and mortality. At the molecular level, patients at different stages present considerable heterogeneity. Although the miRNA in exosome is an effective biomarker to reveal tumor progression, studies based on stage-associated exosome miRNA regulatory network analysis still lacking. This study aims to identify CRC stage-associated exosome miRNAs and reveal their potential function in tumor progression.

**Methods:** In this study, serum and cellular exosome miRNA expression microarrays associated with CRC were downloaded from GEO database. Stage-common (SC) and stage-specific (SS) differentially expressed miRNAs were extracted and their targets were identified based on 11 databases. Furthermore, miRNA SC and SS regulatory function networks were built based on the CRC phenotypic relevance of miRNA targets, and the corresponding transcription factors were identified. Concurrently, the potential stage-associated miRNAs were identified by receiver-operating characteristic (ROC) curve analysis, survival analysis, drug response analysis, ceRNA analysis, pathway analysis and a comprehensive investigation of 159 publications.

**Results:** Ten candidate stage-associated miRNAs were identified, with three SC (miR-146a-5p, miR-22-3p, miR-23b-3p) and seven SS (I: miR-301a-3p, miR-548i; IIIA: miR-23a-3p; IV: miR-194-3p, miR-33a-3p, miR-485-3p, miR-194-5p) miRNAs. Additionally, their targets were enriched in several vital cancer-associated pathways such as TGF-beta, p53, and hippo signaling pathways. Moreover, five key hotspot target genes (*CCNA2*, *MAPK1*, *PTPRD*, *MET*, and *CDKN1A*) were demonstrated to associated with better overall survival in CRC patients. Finally, miR-23b-3p, miR-301a-3p and miR-194-3p were validated being the most stably expressed stage-associated miRNAs in CRC serum exosomes, cell exosomes and tissues.

**Conclusions:** These CRC stage-associated exosome miRNAs aid to further mechanism research of tumor progression and provide support for better clinical management in patients with different stages.

KEYWORDS

colorectal cancer, exosome, miRNA, stage, network analysis

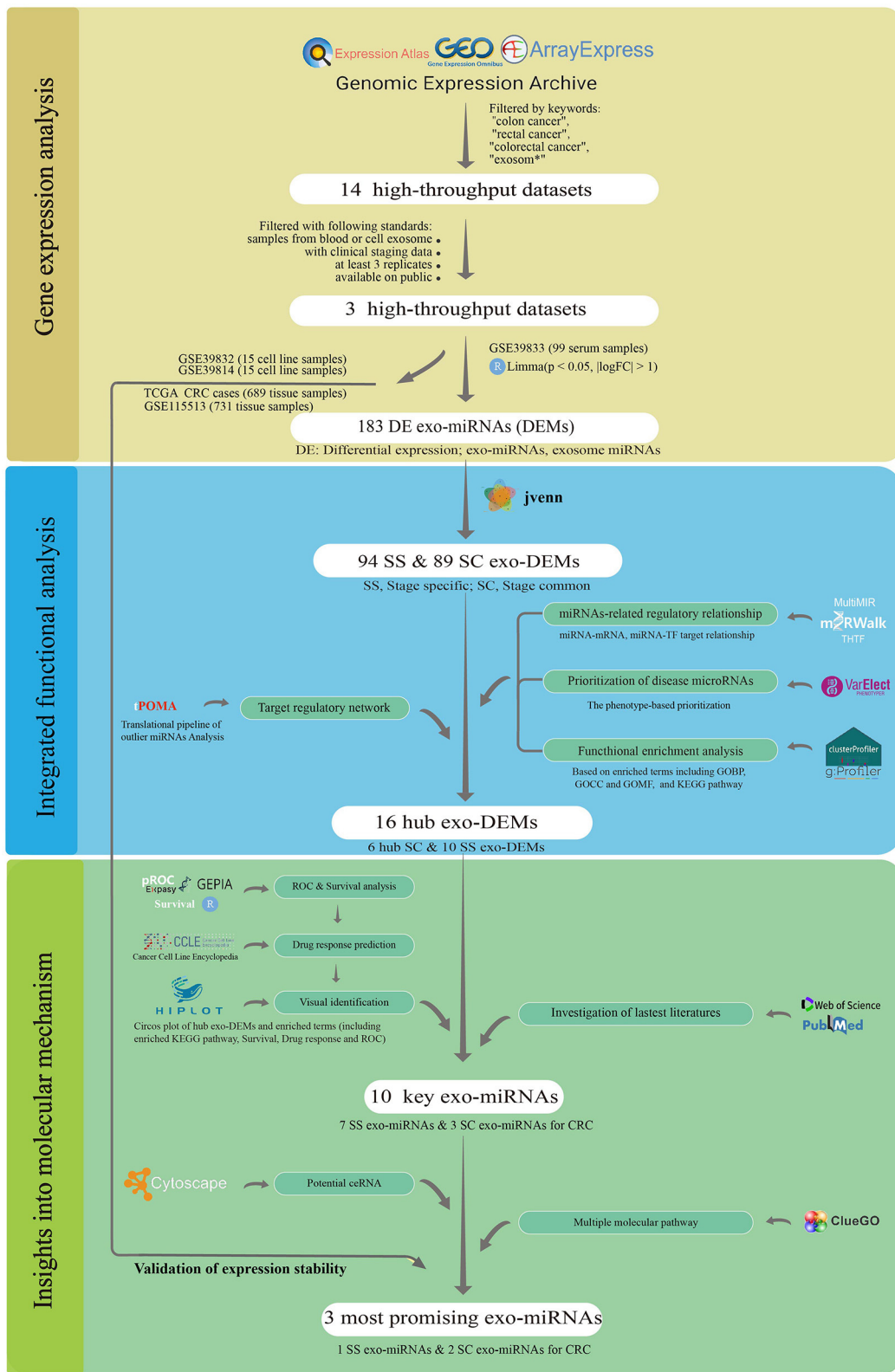
## Introduction

Colorectal cancer (CRC) is an aggressive and fatal malignancy that develops from normal intestinal cells, with up to 1.85 million new cases diagnosed each year worldwide. The mortality rate is as high as 45.9% (850,000 cases) (1), which makes it the world's third most common and second-deadliest cancer (2). CRC is characterized by insidious and rapid progression, metastasis, and treatment difficulty in the middle and late stages. Previous studies have shown that high inter-patient and intra-tumor heterogeneity promotes CRC metastasis and patient deterioration (3). Exosomes have been discovered to be a powerful screening and diagnostic tool for highly heterogeneous malignancies, including CRC, due to their widespread presence in blood, urine, cerebrospinal fluid, and other body fluids, with the outstanding advantages of richer molecules, better representation, higher stability, and greater clinical applicability (4–8). Exosome miRNAs (exo-miRNAs), the cancer genome dark matter, maybe suitable biomarkers for cancer screening, stage monitoring and precise treatment (9, 10). Furthermore, exosome miRNAs have a dramatic impact on metastasis, progression, prognosis and regulation of biological networks in CRC patients.

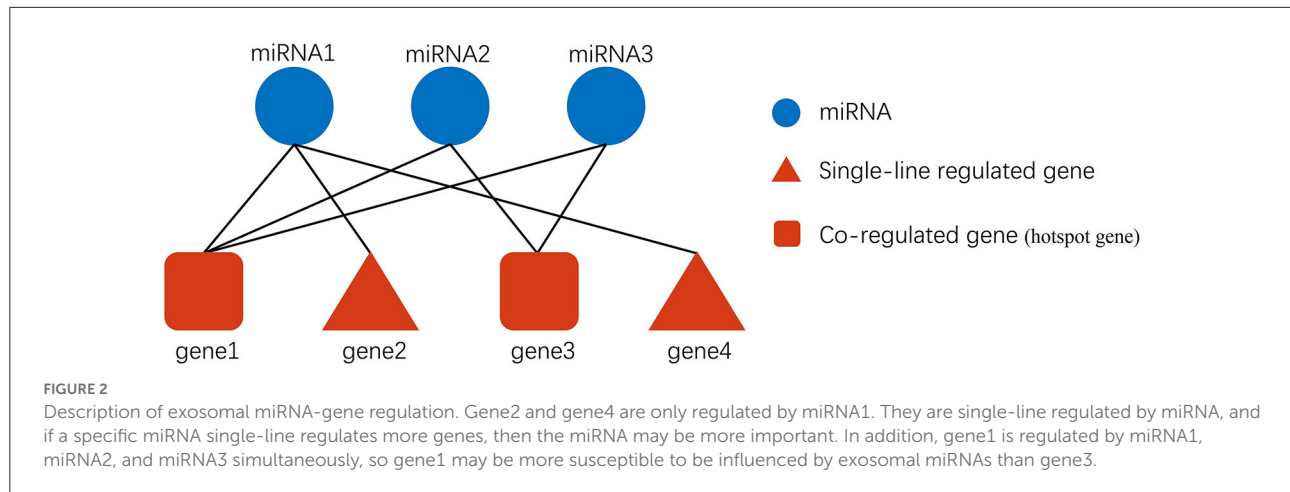
For instance, Matsumura et al. (11) discovered that serum exosome miR-17-92a cluster expression level was correlated with CRC progression and miR-19a with the prognostic significance of expression might be a prognostic biomarker for recurrence in CRC patients. The miR-21 was reported to be highly expressed in various CRC cell exosomes, plasma exosomes and tissues, and correlated with disease staging, liver metastasis and prognosis, which could be used as a reference marker for gastrointestinal cancers (12). Nearly a hundred exosome miRNAs, such as miR-23a, miR-92a, miR-1246, and miR-486, have been identified as promising biomarkers for CRC (13–17). However, the correlation of exosome miRNAs with CRC staging and how it affects CRC progression is less understood. Identification of stage-associated exosome miRNA biomarkers is critical to facilitate further mechanistic studies and dissect tumor progression.

It is encouraging that untargeted omics data and text-derived knowledge have improved biomarker identification. The omics-driven and network-based approaches, composed of biological and computational processes, can lead to the more precise discovery of potential novel biomarkers due to corroborative evidence integrated with interacting molecular entities and incorporating measurements of molecule expression and biological function (18–21). The miRNA targeting topology networks are frequently used to extract key miRNAs from noisy backgrounds for further translational studies (22, 23). In contrast to this approach, Chen et al. (24) increased our understanding of miRNA biology by developing a single-line regulatory miRNA vulnerability network and identifying the corresponding biomarkers. However, single network analysis is insufficient, and the integration of diverse data (e.g., disease relevance of target molecules, ability to regulate transcription factors, and biological pathways) can aid in identifying the stage-associated biomarkers.

Here, we systematically presented an enhanced analysis with a robust and corroborative workflow to identify stage-associated exo-miRNAs biomarkers. In this study, the miRNA single-line regulation vulnerability network construction pipeline was proposed and combined with the marker robust analysis strategy proposed in our previous work (25). As illustrated in Figure 1, three CRC serum and cell exosome datasets were first collected from four public databases. Subsequently, stage-common (SC) and stage-specific (SS) miRNA targeting topology networks for CRC were built, and their distinct biological functions were elucidated. Ten key stage-associated miRNA were further selected from different networks and assessed by receiver operating characteristic (ROC) curve analysis. Additionally, survival analysis and drug response analysis were performed for them at the tissue and cellular levels. Moreover, the competing endogenous RNA (ceRNA) networks were constructed, and the targeting pathways were explored. Finally, three miRNAs were validated presenting stably expression in serum exosomes, cellular exosomes, and tumor tissues. The multi-level and multi-information integration approach enables a robust analysis for biomarker discovery.



**FIGURE 1**  
 The workflow of this study.



## Materials and methods

### Collection and filtering of exosome data sets

To obtain complete datasets of exosome miRNA (exo-miRNA) related gene expression in CRC, four public databases were retrieved using the keywords “colon cancer,” “rectal cancer,” “colorectal cancer,” “exosome,” “exosomes,” and “exosomal.” The databases Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) (26), ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>), ExpressionAtlas (<https://www.ebi.ac.uk/gxa>), and Genomic Expression Archive (GEA, <https://www.ddbj.nig.ac.jp/gea>) (Supplementary Table 1). Afterward, the following criteria were used to filter the data: (i) the data objects were miRNA expression profiles in blood or cellular exosomes. (ii) The blood exosome samples contained clinical staging information. (iii) At least three sample replicates. (iv) They were completely free and downloadable. Finally, three CRC exosome datasets were obtained: one associated with serum exosomes (GSE39833), which included 20 patients with stage I, 20 patients with stage II, 20 patients with stage IIIA, 16 patients with stage IIIB, 12 patients with stage IV, and 11 healthy individuals; the other two were associated with CRC cell line exosomes (GSE39832 and GSE39814), which included four fetal bovine serum-treated controls, three normal colon cells (FHC) and 15 CRC cells (HCT116, HT29, RKO, SW48, SW480; three of each).

### Transcriptome data pre-processing and differentially expressed exo-miRNA identification

Background-corrected and normalized Agilent microarray expression matrixes from the datasets above were downloaded,

negative values were removed, and interpolation was performed using the caret R package’s preProcess function (27). All values were log<sub>2</sub>-transformed. The limma R package (28) was used to analyze differentially expressed miRNAs in serum exosomes from different stages of CRC. Differential significance was determined with  $p < 0.05$  and  $|\log_{2}FC| > 1$  (Supplementary Figure 1), and differential intersection screening and definition were performed by Jvenn online tool (29). MiRBase 22.1 (<http://www.mirbase.org/index.shtml>) was used to transform the names of all miRNAs obtained. Heatmap data were normalized using the scale function, clustered using fully linked hierarchical clustering based on Euclidean distances (if necessary), and visualized by the pheatmap R package (30).

To determine exo-miRNA target gene expression and miRNA expression in tissues, the GDC-TCGA (31) colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) log<sub>2</sub>(count+1)-transformed mRNA expression matrix, log<sub>2</sub>(RPM+1)-transformed miRNA expression matrix and the corresponding clinical data were downloaded from UCSC-ZENA (<https://xenabrowser.net/datapages/>) (32). COAD included 41 normal samples and 471 samples from colon cancer patients and READ included 10 normal samples and 167 samples from rectal cancer patients. Moreover, dataset GSE115513 (33) was used to validate the miRNA expression in the tissue level, which included 649 normal intestinal mucosa samples and 731 CRC tumor tissue samples (including 231 in stage I, 181 in stage II, 219 in stage III, and 100 in stage IV). The normalized expression matrix was downloaded from GEO and log<sub>2</sub> transformed. Key miRNAs were selected and compared among different stages, the Wilcoxon test was used to evaluate the differences between two stages and the Kruskal-Wallis test was applied to estimate the differences among all stages. Gene expression and drug sensitivity data of 22 CRC cell lines were downloaded from the CCLE database (34).



## Prediction and screening of targeted genes from differentially expressed exo-miRNAs

To obtain the targeted genes from differentially expressed exo-miRNAs, 11 databases were searched based on the multiMiR R package (35) and miRWalk 2.0 (36), including three experimental validation databases [miRecords 4.0 (37), mirTarBase 7.0 (38) and Tarbase 8.0 (39)] and eight prediction databases [DIANA-microT-CDS 5.0 (40), EIMMo 5.0 (41), MicroCosm 5.0 (<http://www.ebi.ac.uk/enright-srv/microcosm/cgi-bin/targets/v5/download.pl>), miRanda (42), miRDB 6.0 (43), PicTar 2.0 (44), PITA 6.0 (45) and TargetScan 7.2 (46)]. The top 20% of data with high confidence based on *p*-value and included in at least 4 (5,127 miRNA-gene pairs) of the 8 (133,738 miRNA-gene pairs) prediction databases were taken. miRanda, PITA 6.0, and TargetScan 7.2 were specifically chosen to include only predictions with conserved loci. Furthermore, data from at least 2 (2,825 miRNA-gene pairs) of the 3 (46,455 miRNA-gene pairs) experimental validation databases were chosen. Finally, 7,306 miRNA-gene pairs were obtained from the 11 databases. The accuracy of prediction and the richness of miRNA targeting prediction results were ensured by combining experimental validation databases and prediction databases.

The VarElect online tool was used to perform phenotypic correlation analysis based on the keywords “Colon Cancer” OR “Rectal Cancer” OR “Colorectal Cancer” to further screen for target genes associated with CRC (47). Phenotype scores greater than five, and only direct phenotypic associations were considered.

## Exosomal miRNA targeting analysis process establishment

The Pipeline of Outlier miRNA Analysis (POMA) is an analytical process developed by Chen et al. (24) for identifying key miRNAs. A translational POMA (tPOMA) in exo-miRNA was created to identify key exo-miRNAs in CRC (Figure 2). tPOMA is constructed based on two key assumptions: (i) The number of targeted CRC-related genes/TFs can reflect the regulatory activity of CRC exo-miRNAs. (ii) If only one exo-miRNA target a CRC-related gene/TF, that exo-miRNA is more likely to show regulatory specificity to that gene/TF. Thus, the greater the number of CRC-related genes/TFs regulated by an exo-miRNA in a single line, the more important that miRNA is likely to be. Furthermore, we suppose many CRC exo-miRNAs regulate a CRC-related gene. In that case, the gene is more likely to be repressed (miRNAs typically repress gene expression), and such a

gene was referred to as a hotspot gene regulated by exo-miRNAs.

Based on this assumption, a tPOMA-based screening was performed in four steps on all obtained SC exo-miRNAs: (i) First, a global targeting network of SC exo-miRNAs was built using the previously mentioned 11 databases. (ii) The miRNA single-line regulatory network was further constructed, in which genes only regulated by one exo-miRNA (other exo-miRNAs do not regulate the gene). (iii) To generate a hotspot gene network, single-line regulated genes were removed from the global miRNA-gene network (genes are regulated by at least two exo-miRNAs). (iv) Using the VarElect tool, CRC phenotype-related genes were selected to build the CRC-related single-line regulated gene network and hotspot gene network (47). (v) The TF was confirmed using the THTF database (<http://humantfs.cabr.utoronto.ca>) (48). Finally, the network of exo-miRNAs targeting genes was mapped using Cytoscape (3.7.1) (49).

The online tool miRNet (50) developed by Chang et al. was used to construct a target network of SC exo-miRNAs (target molecules including mRNAs, lncRNAs, circRNAs, pseudogenes, and sncRNAs) and ranked the miRNAs by the degree to demonstrate the reliability of our tPOMA model. The top 10 miRNAs were compared to the top 10 miRNAs predicted by our tPOMA model (Supplementary Figure 2), and there is a high degree of overlap (6 out of 10 miRNAs were consistent).

The following three steps were used to build the SS exo-miRNA subnetworks: (i) The global network of SS exo-miRNAs and target genes was built using the results of the previously mentioned 11 databases. (ii) Genes only present at a specific stage were extracted by Jvenn. (iii) SS exo-miRNAs and target gene sub-networks were further constructed by Cytoscape.

## Enrichment analysis of biological functions and pathways

To better understand the role of these targeted genes in the development of CRC, enrichment analysis of Gene Ontology (GO) (51) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (52) was performed by the clusterProfiler R package (53), and further ranking enrichment analysis was performed using the g: Profiler online tool (<http://biit.cs.ut.ee/gprofiler/gost>) (54). According to the importance of targets, the top 50 genes with the highest relevance to CRC were extracted (FDR < 0.05) to identify the more important biological entries and relevant pathways. The GOplot R package (55) was used to plot and visualize enrichment string maps.

## ROC and survival analysis of key exo-miRNAs

To obtain exo-miRNAs with high AUC, the pROC R package (56) was used for ROC analysis (the ROC curves of all SS and SC exo-miRNAs in [Supplementary Figure 3](#)). For combined-multivariate ROC analysis, logistic regression models were first developed using the glmnet R package (57), and further assessed by pROC.

Furthermore, to determine whether these exo-miRNAs and their target hotspot genes affect CRC patients' overall survival (OS), the survival R package (58) was used to conduct a survival analysis of key SC and SS exo-miRNAs. The TCGA survival data included both COAD and READ data, with means calculated for high and low expression groups, and log-rank  $p < 0.05$  was considered a significant survival impact. GEPIA (59) was used to assess targeted hotspot genes for expression and survival analysis [367 CRC samples and 667 normal samples from TCGA (31) and GTEx (60) were included,  $|\logFC| > 0.5$ ,  $p < 0.05$ , survival analysis grouped by median expression].

## Response analysis of the EGFR-targeted drug for CRC cell lines

Twenty-two cell lines were grouped according to their median expression in the CCLE database (34). The sensitivity of two grouped cell lines to the EGFR-targeting drug lapatinib (as measured by IC50) was compared using a two-tailed  $t$ -test for each targeted EGFR tyrosine kinase inhibitor resistance pathway genes.

## ceRNA network construction and targeting pathway analysis of key exo-miRNAs

The following four steps were used to identify the key CRC exo-miRNA-associated ceRNA networks: (i) The lncRNAs positively associated with miRNA-targeted mRNA were extracted from cbioportal online tool (61) (640 COAD samples from TCGA; Spearman analysis,  $p < 0.05$ ). (ii) The experimentally validated miRNA-lncRNA pairs were determined in the lncBase v2 database (62). (iii) The corresponding ceRNA was identified when lncRNAs and mRNAs linked to the same miRNA. (iv) The corresponding networks were built using Cytoscape (49). Targeted pathway analysis of 10 key miRNAs was accomplished using Cytoscape's ClueGO plugin (63), with relevant KEGG pathways selected at  $p < 0.05$ .

## Results

### Identification of differentially expressed serum exo-miRNAs in different stages of CRC patients

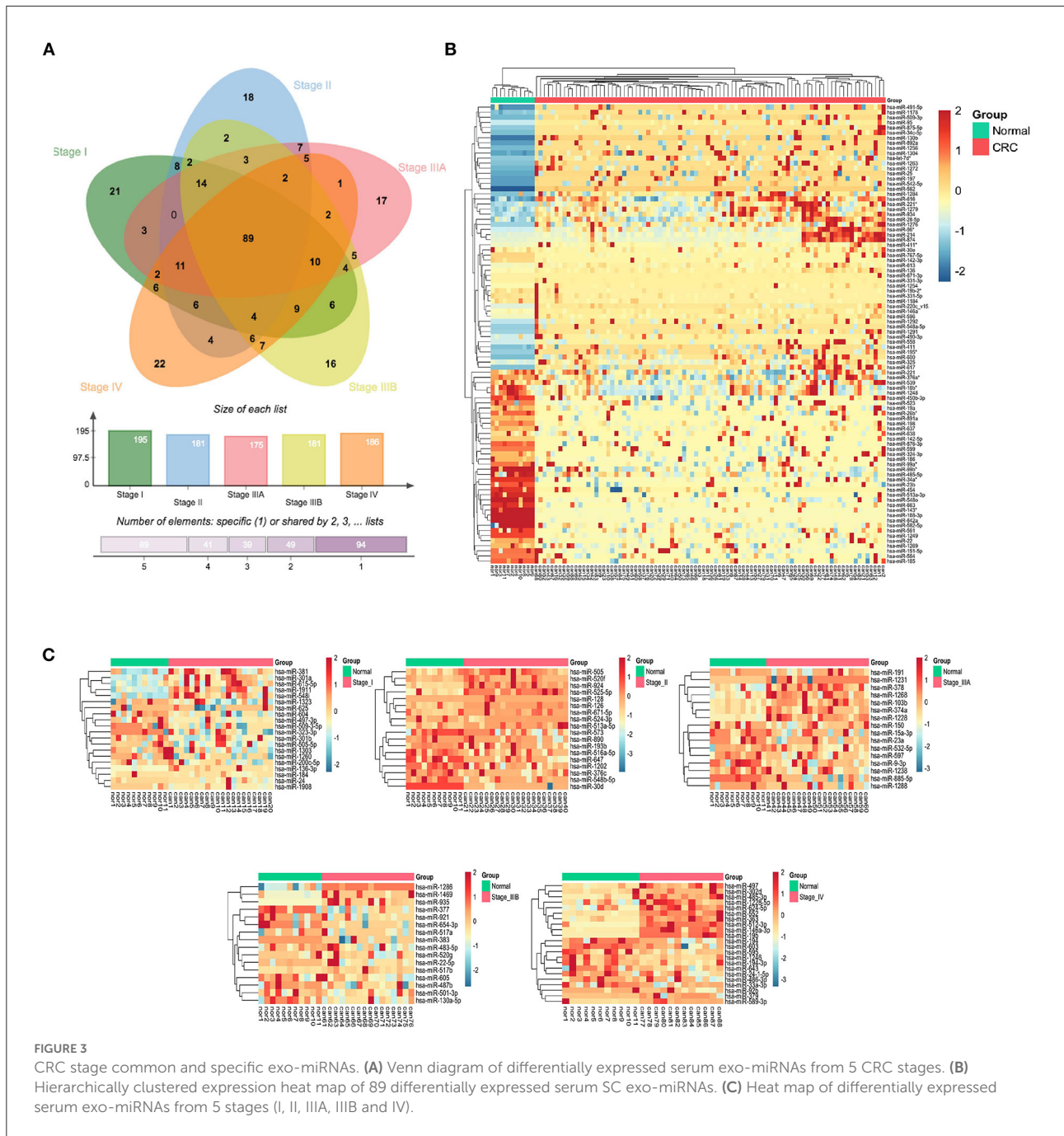
Differential expression analysis of serum exo-miRNAs in different stages of CRC patients yielded 89 SC exo-miRNAs and 94 SS exo-miRNAs ( $p < 0.05$ ,  $|\logFC| > 1$ ) ([Figure 3A](#)). Using hierarchical clustering of heatmaps, 89 SC exo-miRNAs were discovered to effectively distinguish CRC patients from healthy individuals ([Figure 3B](#)). The expression of SS exo-miRNAs was also investigated ([Figure 3C](#)), with stage I and IIIB being poorly differentiated, but stages II, IIIA, and IV being better differentiated, with stage IV showing the most significant difference.

### SC miRNA-target network construction of exo-miRNAs

An exosomal miRNA-gene targeting network for the 89 SC exo-miRNAs was constructed ([Figure 4A](#)), which included 89 exo-miRNAs and their 7,306 predicted target genes from 11 databases (see Materials and methods section for detailed steps). Second, by establishing the tPOMA model (detailed steps are included in the Materials and methods section), a single-line regulatory network of exo-miRNAs was obtained in CRC ([Figure 4B](#)). Then, only genes with a high degree of relevance to CRC were selected from the overall miRNA-gene network and the miRNA single-line regulatory network (for detailed steps, see Materials and methods section). Concurrently, single-line regulated genes were excluded from the overall exosomal miRNA-gene network. Finally, a hotspot gene network was constructed including 56 exo-miRNAs and targeted 173 genes associated with CRC (at least two exo-miRNAs regulated each gene) ([Figure 4C](#); [Supplementary Table 2](#)), as well as a miRNA single-line regulatory network constructed including 43 exo-miRNAs and targeted 186 genes associated with CRC (a single exo-miRNA regulated each gene) ([Figure 4D](#); [Supplementary Table 3](#)).

### Exploring the biological function of genes with different targeting burden

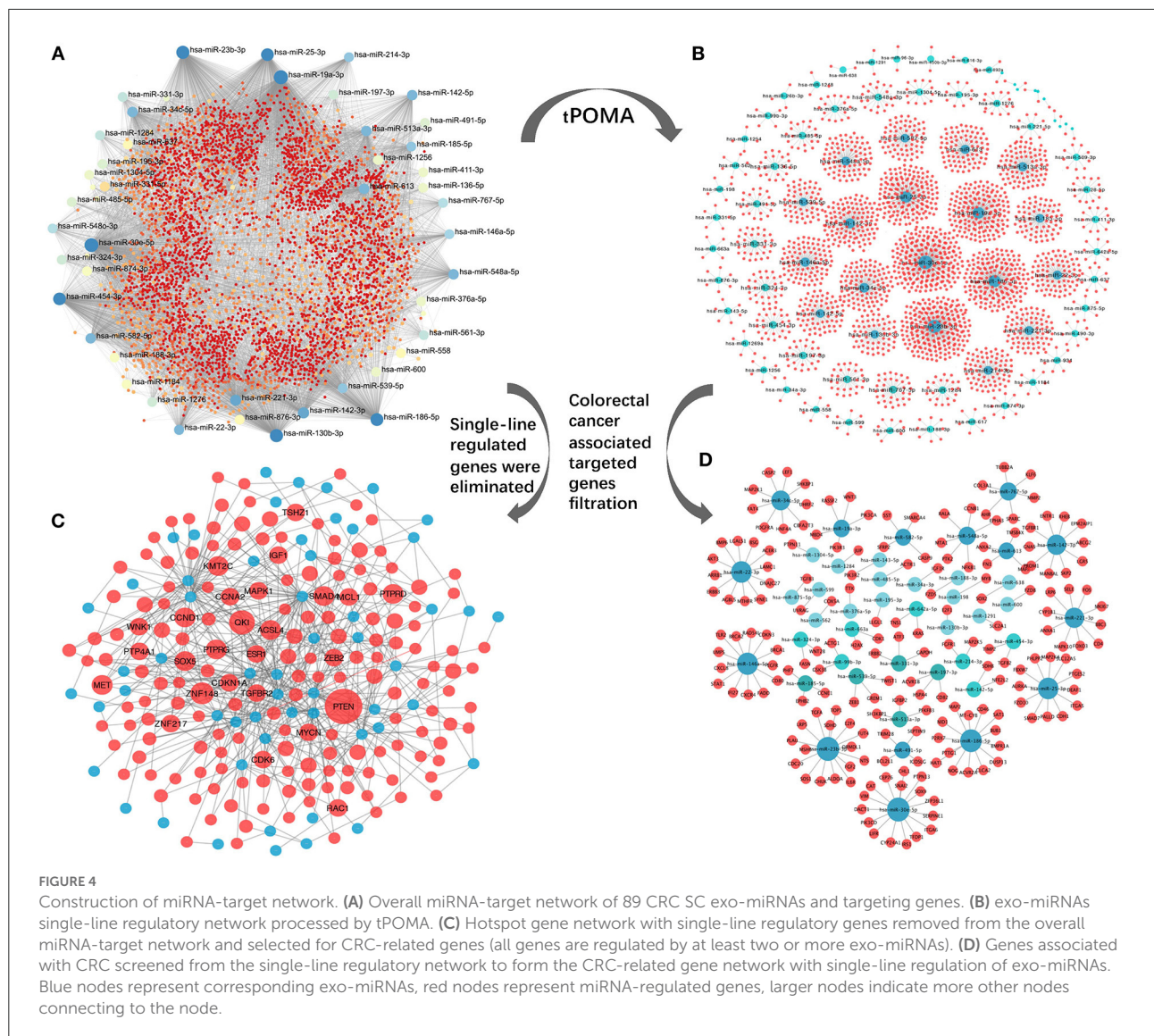
To confirm that screened exo-miRNA-targeted genes are indeed related to the biological functions of exosomes and to compare the biological functions of exo-miRNA-regulated genes with varying targeting burdens, the corresponding enrichment analysis of biological functions was performed.



First, GO and KEGG enrichment analyses were performed for 173 hotspot genes and 186 single-line genes (Supplementary Figure 4A), respectively. It was found that these genes were enriched in signaling and regulation (e.g., signaling receptor binding, identical protein binding, enzyme binding, growth factor binding, and kinase activity); cell membrane complex structures (e.g., membrane-bounded organelle, membrane region, intracellular organelle, cell surface and receptor complex); regulation of substance secretion,

transport and cellular response (e.g., positive regulation of metabolic process, positive regulation of cellular process, cellular response to chemical stimulus, positive regulation of biological process and cell population proliferation) and various biological pathways associated with cancer (e.g., colorectal cancer, gastric cancer, breast cancer, PI3K-AKT signaling pathway and proteoglycans in cancer). The findings show that these genes are important for secretory transport and exosome regulation.

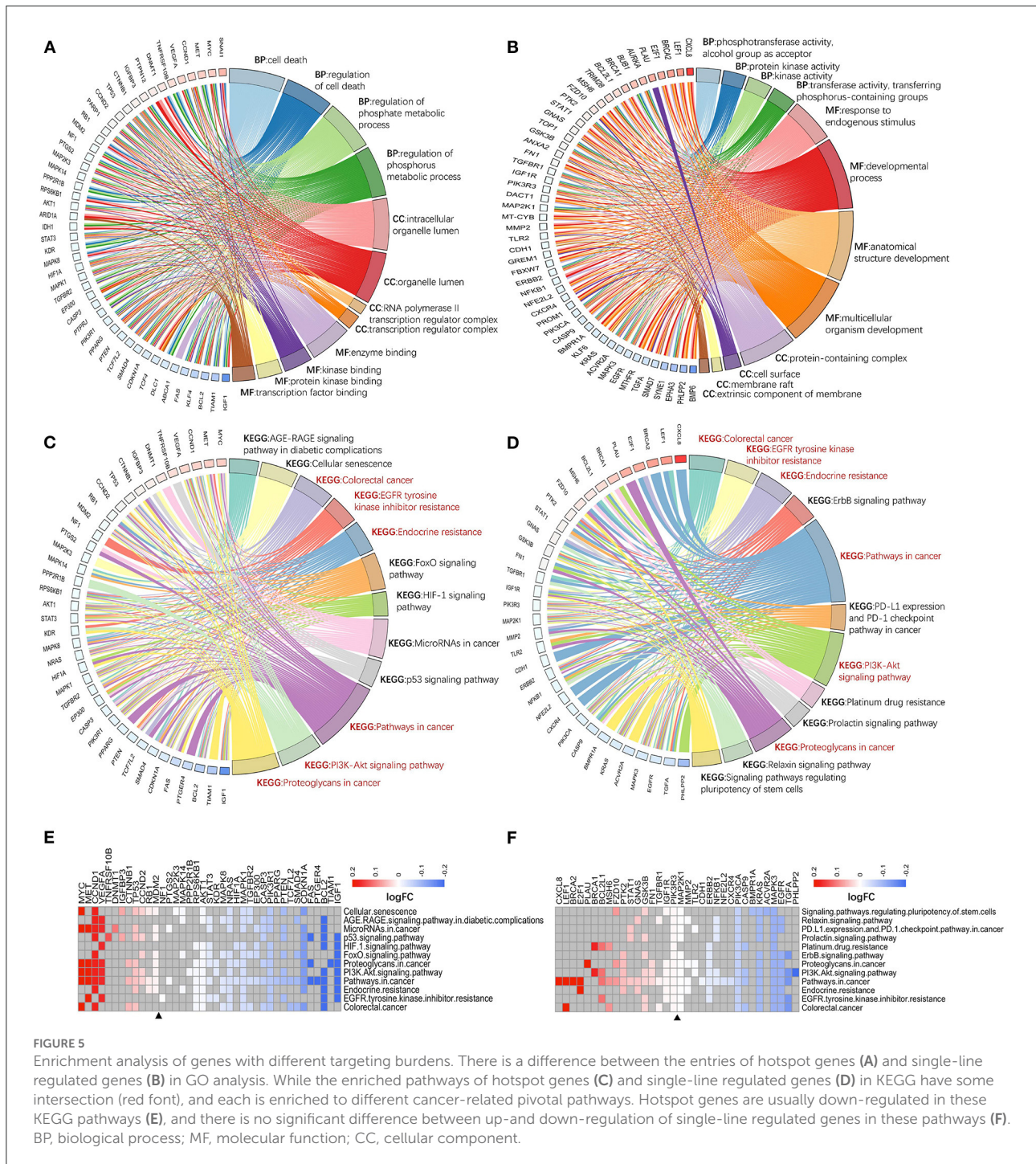




Second, two classes of genes was ranked based on CRC-related scores calculated by VarElect, and their biological functions were further explored using ranking enrichment analysis. Some differences in GO entries between hotspot genes (Figure 5A; Supplementary Table 4) and single-line regulated genes (Figure 5B; Supplementary Table 4) were discovered, as the former is more related to the regulation of cell death and phosphorylation metabolism in biological processes, with molecular functions focused on the binding of enzymes, proteins, and transcription factors, with the majority of these molecules localized in organelle membranes. The latter is primarily involved in activating enzymes in biological processes, molecular functions in cell development regulation and are primarily localized on the cell surface and membrane-related components. Both types of molecules are enriched in several cancer-associated pathways, including colorectal cancer,

EGFR tyrosine kinase inhibitor resistance, endocrine resistance, PI3K/AKT signaling pathway, and cancer proteoglycans. However, they were also enriched in various pathways that are important for cancer development, immune checkpoint, drug tolerance, and pluripotent differentiation of stem cells (e.g., P53 signaling pathway, HIF-1 signaling pathway, FOXO signaling pathway, specific to hotspot genes; PD-L1 expression and PD-1 checkpoint pathway in cancer, platinum drug resistance, signaling pathways regulating pluripotency of stem cells, specific to single-line regulated genes) (Figures 5C,D; Supplementary Table 4).

Finally, the expression of genes from cancer-associated pathways was investigated. The hotspot genes being more down-regulated in these pathways (Figure 5E), which may be related to their negative regulation by multiple exo-miRNAs. In contrast, genes regulated by a single exo-miRNA were almost



**FIGURE 5** Enrichment analysis of genes with different targeting burdens. There is a difference between the entries of hotspot genes (A) and single-line regulated genes (B) in GO analysis. While the enriched pathways of hotspot genes (C) and single-line regulated genes (D) in KEGG have some intersection (red font), and each is enriched to different cancer-related pivotal pathways. Hotspot genes are usually down-regulated in these KEGG pathways (E), and there is no significant difference between up- and down-regulation of single-line regulated genes in these pathways (F). BP, biological process; MF, molecular function; CC, cellular component.

identical in terms of up- and down-regulation in these pathways (Figure 5F).

Through traditional and further ranking enrichment analysis, distinct regulatory patterns for two distinct types of genes with varying targeting burdens were identified in colorectal carcinogenesis and development, paving the way for future research into more precise regulation of exo-miRNAs.

## Identification of three key SC exo-miRNAs and five survival-related hotspot genes

According to the tPOMA model construction hypothesis, miRNAs targeting more CRC-related genes or transcription factors (TF) in the single-line regulatory network of



**TABLE 1** Important miRNAs and hotspot genes screened based on miRNA-target network (top 26).

miRNA	NCGSM-TF	NCGSM	NGSM	NGM	Gene	Indegree
miR-30e-5p	3	16	237	573	PTEN	13
miR-221-3p	2	10	75	251	QKI	8
miR-34c-5p	2	9	101	212	ZNF148	7
miR-767-5p	2	6	33	100	KMT2C	6
miR-539-5p	2	3	55	152	MET	6
miR-23b-3p	1	16	224	489	SMAD4	6
miR-146a-5p	1	13	87	126	CCND1	6
miR-25-3p	1	12	163	418	PTPRD	6
miR-548a-5p	1	6	78	210	PTPRG	6
miR-19a-3p	1	5	166	619	SOX5	6
miR-331-3p	1	3	30	73	ACSL4	6
miR-142-5p	1	2	66	204	TSHZ1	5
miR-186-5p	0	14	181	437	RAC1	5
miR-22-3p	0	12	78	182	WNK1	5
miR-142-3p	0	10	117	268	MAPK1	5
miR-613	0	4	82	206	CCNA2	5
miR-582-5p	0	4	64	195	ESR1	5
miR-491-5p	0	4	19	44	IGF1	5
miR-513a-3p	0	3	83	208	PTP4A1	5
miR-185-5p	0	3	68	152	TGFBR2	5
miR-197-3p	0	3	23	54	ZEB2	5
miR-214-3p	0	2	71	159	ZNF217	5
miR-324-3p	0	2	27	52	MYCN	5
miR-454-3p	0	2	26	415	CDKN1A	5
miR-642a-5p	0	2	8	17	CDK6	5
miR-663a	0	2	4	4	MCL1	5

NCGSM-TF, the number of CRC-related TF regulated in a single line by exo-miRNA; NCGSM, the number of CRC-related genes regulated in a single line by exo-miRNA; NGSM, the number of genes regulated in a single line by exo-miRNA; NGM, the number of genes regulated by exo-miRNAs; INDEGREE, the number of exo-miRNAs that regulate this gene.

exo-miRNAs may be more important. Additionally, in the hotspot gene network, CRC-related genes/TF targeted by a greater number of exo-miRNAs may be more susceptible to repressed. The corresponding exo-miRNAs and hotspot genes were arranged in descending order according to NCGSM-TF and INDEGREE (Table 1). To narrow down the screening of key miRNAs/genes and identify molecules of greater importance. NCGSM-TF and NCGSM were combined to thoroughly examine the top 15 miRNAs and hotspot genes with INDEGREE  $\geq$  5 (26 genes). Five of these miRNAs (miR-23b-3p, miR-186-5p, miR-22-3p, miR-142-3p and miR-221-3p) had significantly lower expression (Figure 6A) and high AUC (AUC > 85%, Figure 6B) in serum exosomes of CRC patients compared with healthy individuals, while miR-146a-5p demonstrated higher expression.

Interestingly, most miRNAs in CRC patients' serum exosomes were down-regulated (data not shown), which is consistent with the general down-regulation of original miRNA levels in cancer patients (64) and suggests that most miRNAs in exosomes may act as cancer suppressors. Ten of 26 hotspot genes screened were upregulated in tissue expression in CRC patients, and seven were downregulated ( $|\log_{2}FC| > 0.5$ ,  $p < 0.05$ ), and nine were below threshold (Supplementary Figure 5). Furthermore, five genes (*CCNA2*, *MAPK1*, *PTPRD*, *MET*, and *CDKN1A*) were significantly associated with CRC patients' better survival (Figure 6C). There is reason to believe that these genes with a high miRNA suppressive burden in CRC patients act as protective molecules in the compensatory increase after tumor occurrence and positively correlate with patient survival. Finally, the exo-miRNA regulatory network of five survival-related genes was constructed (Figure 6D). Three miRNAs were identified (miR-146a-5p, miR-22-3p, and miR-23b-3p) with the strong single-line regulatory ability and the ability to regulate hotspot genes in conjunction with other exo-miRNAs.

## Construction of miRNA-target network and subnetworks for SS exo-miRNAs

First, target genes and screened CRC-related genes were identified for all SS exo-miRNAs obtained based on 11 databases, and an SS miRNA-target network was constructed (Figure 7A). Subsequently, the CRC-related genes of each stage were identified by Venn diagram (Figure 7B), and finally, five SS CRC miRNA-target sub-networks were obtained (stage IV, Figure 7C; stage I, Figure 7D; stage II, Figure 7E; stage IIIA, Figure 7F; stage IIIB, Figure 7G). Additionally, GO and KEGG enrichment analyses were performed to explore the role of different SS subnetworks in CRC development.

According to the results of GO analysis, stage I focuses on cellular development and metabolic processes (e.g., cellular development process, cell differentiation, and positive regulation of metabolic process) and is found on the cell surface, endomembrane system, vesicle, and intracellular vesicle (Supplementary Figure 4B). Stages II and IIIA have increased signal transduction and response (e.g., regulation of cell communication, signaling, signal transduction, and chemical response). They are associated with transcriptional regulation and cellular junctions (e.g., transferase complex, chromosome, DNA polymerase complex, and cell-cell junction). The main focus of stage IV is on metabolic and phosphorylation-related processes (e.g., positive regulation of the cellular metabolic process, phosphorus metabolic process, and phosphorylation regulation), which are closely related to cell adhesion and transcriptional regulation (e.g., cell junction, anchoring junction, focal adhesion, and transcription regulator complex). All stages are linked to molecular binding and enzyme

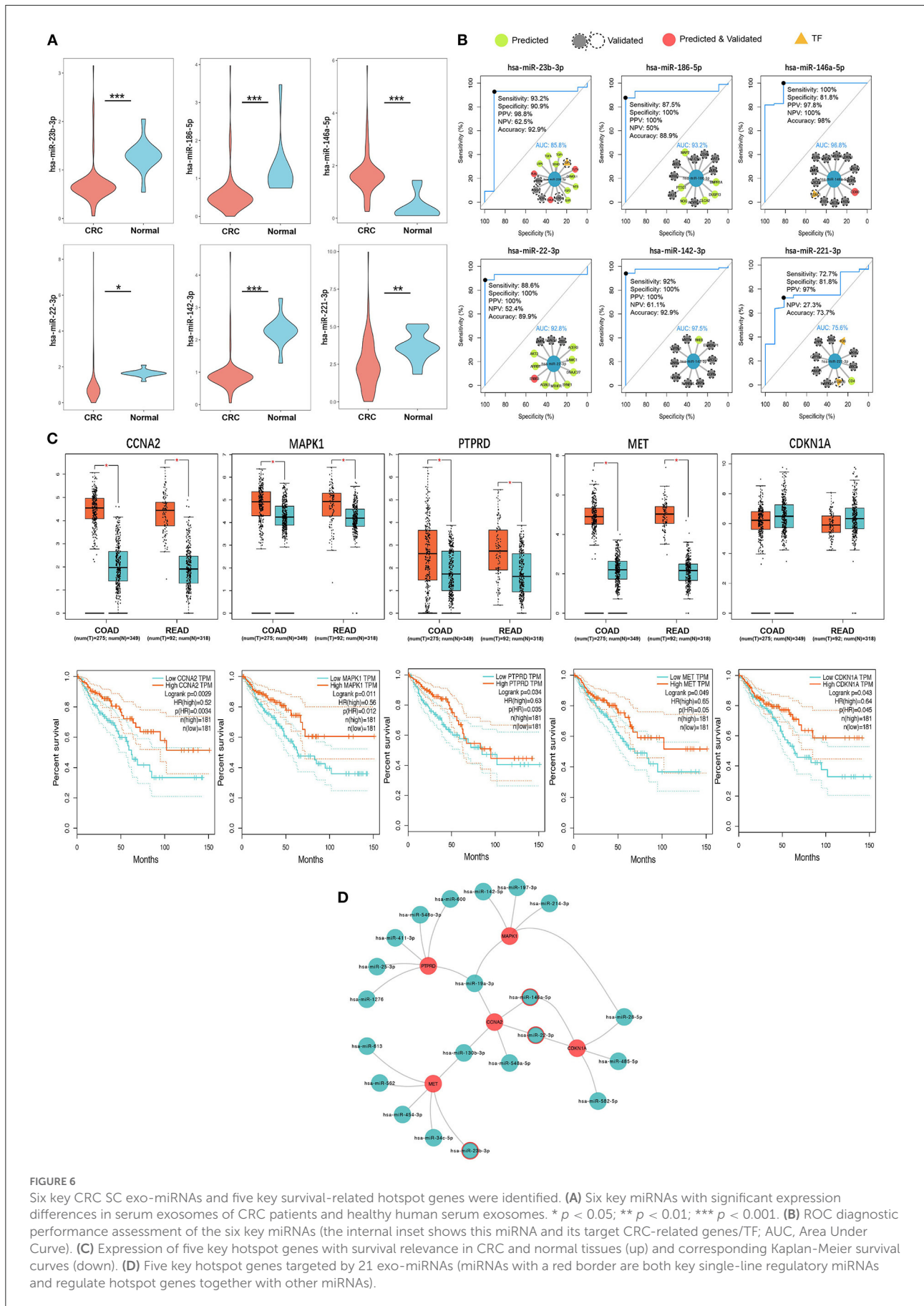
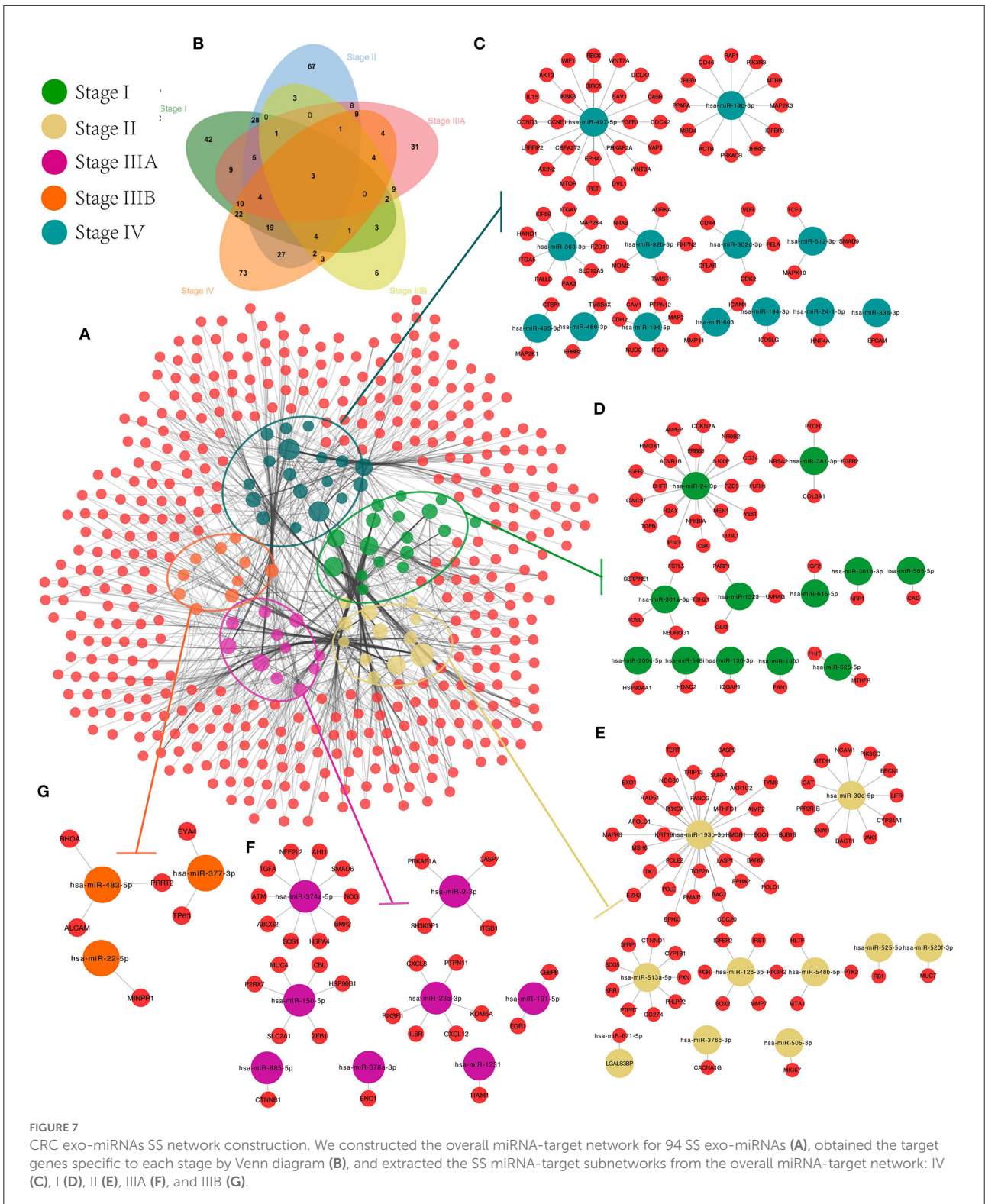


FIGURE 6

Six key CRC SC exo-miRNAs and five key survival-related hotspot genes were identified. (A) Six key miRNAs with significant expression differences in serum exosomes of CRC patients and healthy human serum exosomes. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . (B) ROC diagnostic performance assessment of the six key miRNAs (the internal inset shows this miRNA and its target CRC-related genes/TF; AUC, Area Under Curve). (C) Expression of five key hotspot genes with survival relevance in CRC and normal tissues (up) and corresponding Kaplan-Meier survival curves (down). (D) Five key hotspot genes targeted by 21 exo-miRNAs (miRNAs with a red border are both key single-line regulatory miRNAs and regulate hotspot genes together with other miRNAs).



activation in molecular function. Furthermore, genes in different SS subnetworks showed some pathway similarities and differences.

The effects of exo-miRNAs-regulated genes on CRC progression and changes in biological processes were observed by exploring the biological functions of genes

TABLE 2 Expression changes of key miRNAs in different datasets.

Type	miRNA	Serum exosome	Cell exosome	Tissue	Tissue
		GSE39833	GSE39832 GSE39814	TCGA	GSE115513
SC	miR-146a-5p	Up	Up/down	Up	Up/down
	miR-22-3p	Down	Up/down	Up	Up/down
	miR-23b-3p	Down	Down	Down	Down
SS-I	miR-301a-3p	Up	Up	Up	NA
	miR-548i	Up	Up/down	NA	NA
SS-III A	miR-23a-3p	Down	Up/down	Up	Up
SS-IV	miR-194-3p	Down	Down	Down	Down
	miR-33a-3p	Down	Up	Down	NA
	miR-485-3p	Up	Up/down	Down	NA
	miR-194-5p	Down	Up/down	Up	Down

in different subnetworks, which aids in understanding the regulation of exo-miRNAs in different stages of CRC patients.

## Identification of seven key SS exo-miRNAs

The expression and ROC curve diagnostic performance of all SS exo-miRNAs were evaluated. There were two I-specific miRNAs (miR-301a-3p, miR-548i), two IIIA-specific miRNAs (miR-23a-3p, miR-532-5p), and six IV-specific miRNAs (miR-1246, miR-194-3p, miR-33a-3p, miR-485-3p, miR-194-5p and miR-379-5p) present high Stage-AUC (Stage-AUC > 65%, it represent the performance of miRNA can distinguish one stage from other stages).

Their stage diagnostic performance (Stage-AUC) was 65–75%, and their overall diagnostic performance (the miRNA can distinguish CRC patients from healthy people) was 62–85% (Figure 8; Supplementary Table 5). Furthermore, seven miRNAs found in previously constructed SS subnetworks were selected as key SS exo-miRNAs, with two for stage I (miR-301a-3p and miR-548i), one for stage IIIA (miR-23a-3p), and four for stage IV (miR-194-3p, 33a-3p, miR-485-3p and miR-194-5p). Given the low stage differentiation properties of these SS exo-miRNAs, a combined diagnostic analysis of two I-specific exo-miRNAs and four IV-specific exo-miRNAs was performed, with the stage I combined diagnostic model achieving 72% AUC (Figure 9A) and the stage IV combined diagnostic model achieving 93% AUC (Figure 9B).

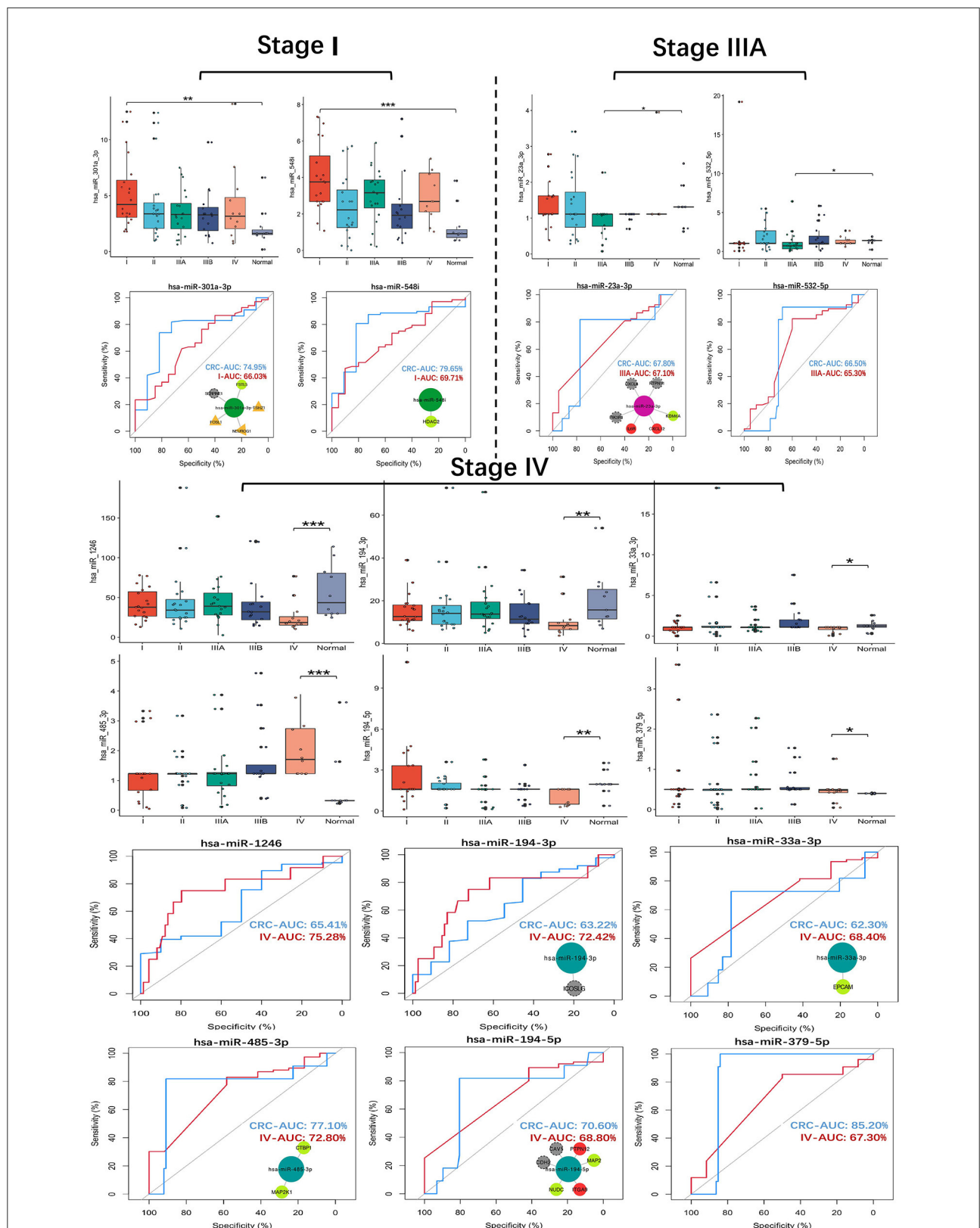
## Survival and expression validation of SC/SS miRNA

A survival analysis based on CRC data in TCGA was performed to further understand the impact of these SC/SS exo-miRNAs on the survival of CRC patients. It was discovered that low expression of miR-1246 was associated with worse overall survival (OS) of patients (Figure 9C), while low expression of miR-379 and miR-485 predicted worse survival after 3 years (Figure 9D). Additionally, the expression of key exo-miRNAs in colon cancer tissues and healthy tissues were explored in TCGA (Figure 9E), in which the miR-23b-3p, miR-146a-5p, miR-301a-3p, and miR-194-3p present the same expression changes compared with CRC serum exosomes. The miRNA expression was further validated in GSE115513, and most of the miRNA expression alterations were similar to that of the TCGA cohort, except for miR-194-5p (SS-IV) had the opposite expression (Supplementary Figure 6). Among them, both tissue and exosome expression of miR-194-3p (SS-IV) were significantly decreased in stage IV, suggesting its possible role as a potential tumor suppressor miRNA for exosomal transport. Finally, the expression of key exo-miRNAs in the exosomes of five different CRC cell lines was investigated (Supplementary Figure 7). The expression of exo-miRNAs in CRC cell line exosomes was largely consistent with that in serum exosomes (e.g., miR-22-3p, miR-23b-3p, miR-301a-3p, miR-194-3p, miR-485-3p). In summary, miR-23b-3p, miR-301a-3p and miR-194-3p were validated being the most stably expressed stage-associated miRNAs in CRC serum exosomes, cell exosomes and tissues (Table 2).

## Identification of exo-miRNAs and their target genes that respond to the EGFR-targeting drug

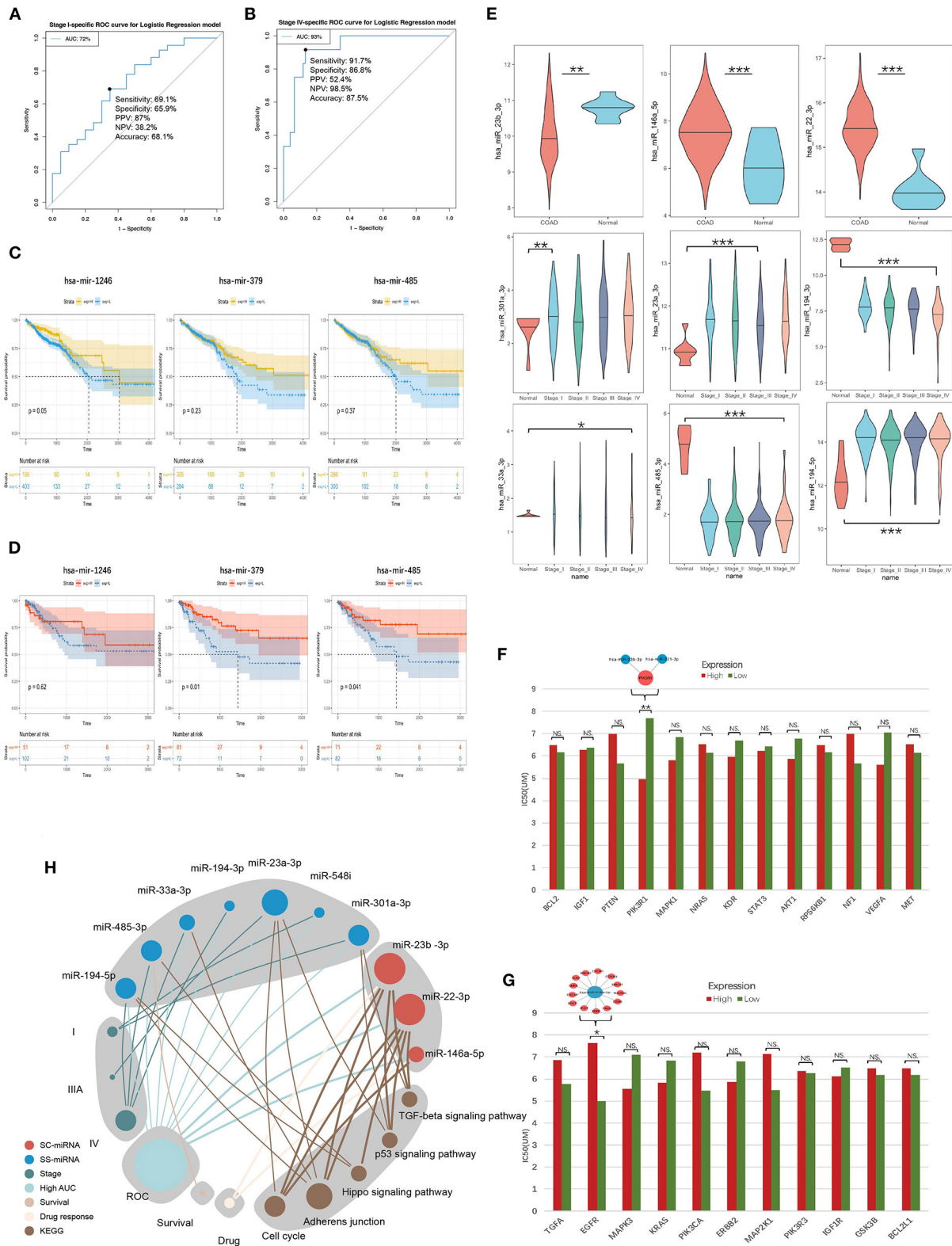
The EGFR tyrosine kinase inhibitor tolerance pathway was enriched in hotspot genes and single-line regulated genes. The response of exo-miRNAs regulatory genes in this pathway to the EGFR-targeted drug lapatinib was analyzed to further identify exo-miRNAs regulatory genes with tolerance response to EGFR tyrosine kinase inhibitors (of the three EGFR targeting drugs in the CCLE database, only lapatinib responded to these genes). Low PIK3R1 expression among the hotspot genes in the CRC Cell significantly increased resistance to lapatinib (Figure 9F), and the previously identified key SC exo-miRNAs: miR-23b-3p and miR-221-3p were able to regulate this gene directly. In contrast, only high EGFR expression among single-line regulated genes in the CRC cell increased resistance to lapatinib (Figure 9G), and EGFR was directly regulated by miR-146a-5p.





**FIGURE 8**  
 SS exo-miRNAs expression assessment and ROC curve analysis. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . CRC-AUC: area under the curve of this miRNA in CRC patients and healthy human serum exosomes; Stage-AUC: area under the curve of this miRNA in this stage versus other stages; ROC curves with network insets are key miRNAs and their regulated SS CRC-related genes present in the SS subnetworks.





**FIGURE 9** Survival analysis and drug response analysis. **(A)** ROC curves include two stage I-specific exo-miRNAs (miR-301a-3p, miR-548i) for combined differentiation of stage I and other stages of CRC patients. **(B)** ROC curves include four stage IV-specific exo-miRNAs (miR-194-3p, miR-33a-3p, miR-485-3p and miR-194-5p) were used to jointly distinguish stage IV CRC patients from other stages. **(C,D)** Kaplan-Meier curve demonstrated *(Continued)*

**FIGURE 9 (Continued)**

mir-1246 was associated with overall survival, mir-379 and mir-485 were associated with 3-years later survival of CRC patients. **(E)** Tissue expression assessment of three key SC exo-miRNAs and six SS exo-miRNAs in CRC (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ). **(F)** EGFR tyrosine kinase inhibitor resistance analysis of hotspot genes in the EGFR tyrosine kinase inhibitor tolerance pathway (the higher the IC50 value, the stronger the resistance; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ). **(G)** Resistance analysis of single-line regulated genes in the EGFR tyrosine kinase inhibitor tolerance pathway (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ). **(H)** Summary of molecular features of 10 key exo-miRNAs.

Finally, the potential pathways, ROC diagnostic capabilities, survival, and drug response of 10 key stage-associated exo-miRNAs were summarized to demonstrate the potential importance of each miRNA (Figure 9H). MiR-22-3p, miR-23a-3p and miR-23b-3p are the top3 important stage-associated exo-miRNAs, and they display close correlation with several vital cancer-associated pathways such as TGF-beta signaling, p53 signaling, and hippo signaling pathways.

## Exploring the potential ceRNA regulatory functions of 10 key exo-miRNAs

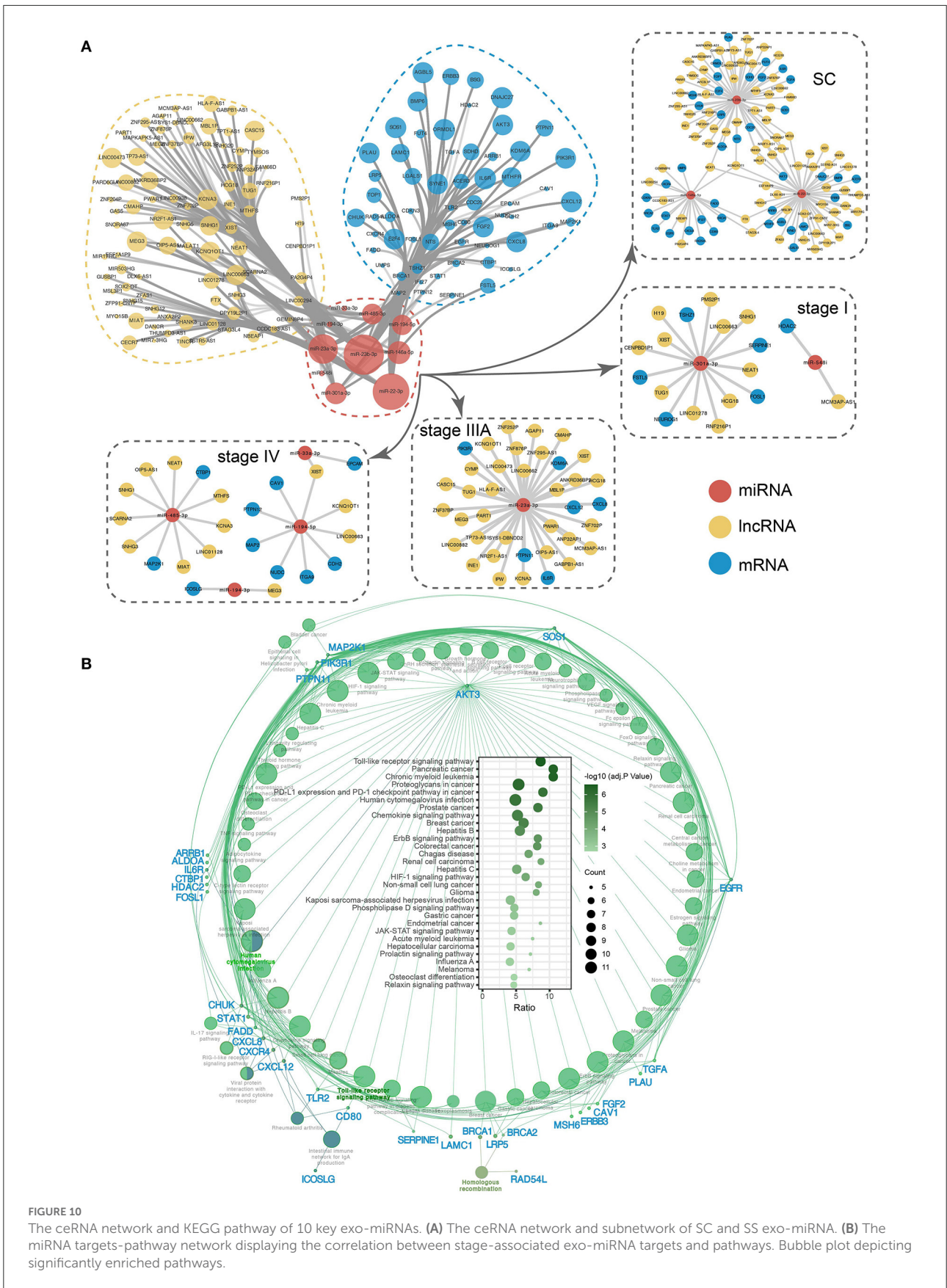
In addition to targeting the corresponding mRNAs, inhibiting their expression, and influencing downstream functions, miRNAs can also interact with the corresponding lncRNAs, and a competitive relationship between lncRNAs and mRNAs that bind to the same miRNA, known as competing endogenous RNA (ceRNA) regulation, can form. To explore the potential ceRNA regulation of stage-associated exo-miRNAs, the ceRNA network was constructed (Figure 10A; Supplementary Table 6). NEAT1 and KCNQ1OT1 are lncRNAs associated with all three ceRNA regulatory modules in the SC exo-miRNA ceRNA network. NEAT1 and KCNQ1OT1 have been reported to be upregulated in a variety of cancers and promote tumorigenesis by altering the expression levels of sponged miRNAs (65, 66). Additionally, the corresponding ceRNA regulatory networks in stages I, IIIA, and IV were identified, which have corresponding roles in regulating the various stages. To better understand the impact of these miRNAs on downstream biological pathways, their miRNA targeting pathways were analyzed using the ClueGO plugin in Cytoscape (63). There were 59 significantly enriched pathways (Supplementary Table 7), among which Toll-like receptor signaling pathway is the most significant (adj.  $p = 2.60e-07$ , adjusted by Bonferroni step down) (Figure 10B). Additionally, *AKT3* was identified as the gene that is most strongly associated with all pathways and is involved in a variety of cancer-related processes. *AKT3* is a member of the protein encoding gene of AKT serine-Threonine protein kinase. AKT is usually overexpressed in CRC and serves as a key target of targeted therapy based on PI3K/AKT signaling pathway (67, 68).

## Discussion

Identification of stage-associated exo-miRNAs biomarker is critical for better understanding the unique developmental progression of CRC at different stages, developing optimal stage monitoring protocols, and determining precise tumor treatment strategies. However, traditional miRNA markers are frequently discovered based on their altered expression and discriminatory ability between experimental and control groups, ignoring their regulatory ability and biological function. Combining the differential expression alteration and *in vivo* targeting or the regulatory ability of miRNA target molecules with their function and disease relevance can aid in identifying more reliable miRNA biomarkers.

The miRNA targeting network considers miRNAs' targeting and regulatory abilities and effectively identify disease-related miRNAs. However, a single miRNA can frequently regulate the expression of a large number of genes simultaneously, and the diversity of regulatory genes frequently complicates our understanding of the miRNA's regulatory function. The POMA model developed by Chen et al. (24) simplifies miRNA regulation by identifying single-line regulatory molecules and vulnerable nodes in the miRNA regulatory network. These genes, which are only regulated by one miRNA, avoid regulatory interference from other miRNAs, effectively reflecting that miRNA's regulatory specificity. This concept was applied to screening key exo-miRNAs in this study, resulting in developing a screening strategy for the tPOMA. Finally, 10 stage-associated exo-miRNAs were identified by tPOMA and further stage subnetwork analysis, ROC analysis, survival analysis and drug response analysis.

Three key SC exo-miRNAs (miR-146a-5p, miR-22-3p, and miR-23b-3p), which can significantly discriminate CRC patients and healthy individuals (AUC = 96.8, 92.8, and 85.8%). miR-146a-5p expression is significantly increased in CRC tissues and promotes CRC cell migration (69), which likely explains the significant increase in miR-146a-5p expression in CRC serum exosomes. Similarly, the significant increase of miR-146a-5p in exosomes may promote CRC cell migration, which needs to be further investigated. And miR-146a-5p also induces tumor formation by inhibiting NUMB protein, resulting in the asymmetric segregation of CRC stem cells (70). The miR-22-3p was found to be significantly upregulated in CRC tissues (71), but it was downregulated in serum and cell exosomes from our



**FIGURE 10**  
The ceRNA network and KEGG pathway of 10 key exo-miRNAs. (A) The ceRNA network and subnetwork of SC and SS exo-miRNA. (B) The miRNA targets-pathway network displaying the correlation between stage-associated exo-miRNA targets and pathways. Bubble plot depicting significantly enriched pathways.



results. Sastre et al. (72) transfected miR-22-3p mimics into CRC cells (HCT116), which were found to inhibit cell proliferation and induce cell death significantly. However, miR-23b-3p was downregulated in CRC tissues, serum exosomes, and cell exosomes, which has been identified as a key regulatory molecule in some cancers (73–77). Ostensfeld et al. (78) discovered that miR-23b-3p was elevated in extracellular vesicles of CRC blood samples and decreased after the tumor was surgically removed. It appears to be different from our finding of reduced miR-23b-3p in CRC patients' serum exosomes, which may be caused by the tumor heterogeneity.

Seven key SS exo-miRNAs include two key I-specific exo-miRNAs (miR-301a-3p and miR-548i), one key IIIA-specific exo-miRNA (miR-23a-3p) and four key IV-specific exo-miRNAs (miR-194-3p, miR-33a-3p, miR-485-3p, and miR-194-5p). The miR-301a-3p is significantly upregulated in CRC tissues (79, 80), and Zhang et al. (80) discovered that overexpression of miR-301a-3p promoted CRC cell proliferation, migration, invasion, as well as the growth of xenograft tumors and liver metastasis *in vivo*. However, miR-548i was found to be significantly downregulated in early colon adenocarcinoma tissue when compared to precancerous and colonic intraepithelial neoplasia tissue (81). Additionally, miR-301a-3p and miR-548i were both upregulated in CRC patients' serum exosomes and were most abundant in stage I patients, suggesting a role for early detection of CRC. Along with miR-376c-3p, miR-23a-3p has been successfully used as a serum marker for colorectal cancer (CRC) to predict 3-year OS in 70% of patients (82). Notably, it was discovered that miR-485 in CRC-IV exosomes could predict patients' better survival after 3 years at the level of CRC tissues, which may aid in the prognosis of advanced CRC patients. In CRC-IV exosomes, miR-194-3p and miR-194-5p showed intriguing results. miR-194-3p expression was significantly downregulated in CRC tissues, whereas miR-194-5p expression was significantly upregulated, but both were significantly downregulated in exosomes. According to Zhang et al. (83), CRC cells' migration and invasion ability transfected with miR-194 mimics appeared to be reduced. Still, it is unknown whether CRC cells can enhance their migration and invasion ability by decreasing miR-194 family secretion in exosomes.

Finally, it was discovered that a drug-resistance pathway (EGFR tyrosine kinase inhibitor resistance) was enriched for both types of targeted genes and that CRC cell lines with low expression of *PIK3R1* (targeted by exosome miR-23b-3p/miR-221-3p) exhibited resistance to the EGFR-targeted drug lapatinib. Additionally, CRC cell lines with high *EGFR* (targeted by exosome miR-146a-5p) also demonstrated resistance to lapatinib. Briefly, the reduced expression of SC miR-23b-3p and elevated expression of miR-146a-5p in serum exosomes of CRC patients (Figure 6A) may help to enhance the sensitivity to EGFR-targeted drugs in all patients without considering stage. Concurrently, SS-IIIA miR-23a-3p could

also target *PIK3R1* (Figure 7F), the expression of miR-23a-3p was specifically reduced in serum exosome of stage IIIA CRC patients, which may help to enhance the response to EGFR-targeted drugs in stage IIIA patients. The regulation of some CRC-related genes by these exo-miRNAs exhibited distinct responses to drug treatment, suggesting that exo-miRNAs may represent hope for drug-resistant therapy in some CRC patients. Moreover, the genes targeted by stage-associated exosomal miRNA are indeed involved in many main processes of colorectal carcinogenesis and important cancer-related pathways such as: angiogenesis, immune regulation, cellular metabolism (Supplementary Table 4), and TGF-beta, p53, hippo signaling pathways (Figure 9H). In detail, we found that miR-23a-3p (SS-IIIA) could target *CXCL8* (Figure 7F), which has been reported in several studies to play a key role in angiogenesis in CRC (84), and the specific reduction of miR-23a-3p in exosomes from stage IIIA patients may upregulate *CXCL8* expression and thus promote angiogenesis. In addition, inflammatory response is one of the typical features of CRC, and *SERPINE1* is one of the targets of miR-301a-3p (SS-I) (Figure 7D), which has been reported to play a key role in shaping the immune microenvironment and activation of immune-related pathways in CRC (85).

In general, these stage-associated exo-miRNAs have been studied in tissue/cell from CRC or other cancers, but they have been reported in exosomes relatively infrequently or not at all. Despite some sample type differences, a corresponding comparison and side-by-side validation were performed using cellular exosomes and tissues data. A literature mining-based confirmatory survey of the 10 CRC exosome-associated miRNAs identified as markers was also performed further to corroborate our study. Finally, 159 studies had been investigated and the pro-/anti-tumor effects of 10 stage-associated exo-miRNAs were revealed (Supplementary Table 8). It is worth noting that there are still some limitations of the present study. First, the main focus of our study was on the functional impact of these stage-related exo-miRNAs on CRC, and the limited amount of data does not support their use as valid diagnostic indicators. Second, although we obtained three exo-miRNAs stably expressed in cellular exosomes, serum exosomes and tissues, their pro-/anti-tumor effects need further experimental validation.

## Conclusion

Three SC and seven SS exo-miRNAs were identified by stage-related miRNA-target network analysis and assessed by ROC analysis. Three exo-miRNAs were validated being stably expressed in serum exosomes, cell exosomes and tissues. These stage-associated exo-miRNAs are expected to be new biomarkers for CRC patients' staging surveillance and further mechanism research.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

## Author contributions

FL and LT designed the ideas for this paper. ZC, YT, and JL contributed to data collation and data analysis. FL analyzed and interpreted the data. All authors contributed to the writing of the manuscript. All authors contributed to this article and approved the submitted version.

## Funding

This work was funded by the Science Innovation Program of College of Laboratory Medicine, Chongqing Medical University (CX201704).

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA*. (2021) 325:669–85. doi: 10.1001/jama.2021.0106
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492
3. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in colorectal cancer: a challenge for personalized medicine? *Int J Mol Sci*. (2018) 19:3733. doi: 10.3390/ijms19123733

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.881788/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

PCA quality control and volcano plot of exo-miRNAs expression in five CRC stages versus healthy individuals.

### SUPPLEMENTARY FIGURE S2

Reliability verification of tPOMA model.

### SUPPLEMENTARY FIGURE S3

ROC and expression of all stage-specific exo-miRNAs.

### SUPPLEMENTARY FIGURE S4

Enrichment analysis of stage-common and stage-specific exo-miRNAs target genes.

### SUPPLEMENTARY FIGURE S5

Expression of 26 hotspot genes in COAD and READ tissues.

### SUPPLEMENTARY FIGURE S6

Tissue expression validation of three key SC exo-miRNAs and seven SS exo-miRNAs in GSE115513.

### SUPPLEMENTARY FIGURE S7

Expression of 10 key exo-miRNAs in exosomes of five CRC cell lines.

### SUPPLEMENTARY TABLE 1

CRC exosome datasets collection results.

### SUPPLEMENTARY TABLE 2

Hotspot gene network and CRC-related scores.

### SUPPLEMENTARY TABLE 3

Single-line regulated genes and CRC-related score.

### SUPPLEMENTARY TABLE 4

Results of GO and KEGG enrichment analysis for hotspot genes and single-line regulated genes.

### SUPPLEMENTARY TABLE 5

Sensitivity, specificity, PPV, NPV and accuracy of SS exo-miRNAs.

### SUPPLEMENTARY TABLE 6

ceRNA.

### SUPPLEMENTARY TABLE 7

Results of ClueGO enrichment analysis for 10 key exo-miRNAs.

### SUPPLEMENTARY TABLE 8

Literature verification of 10 key exo-miRNAs.

4. van Niel G, Porto-Carreiro I, Simoes S, Raposo G. Exosomes: a common pathway for a specialized function. *J Biochem*. (2006) 140:13–21. doi: 10.1093/jb/mvj128
5. Mori MA, Ludwig RG, Garcia-Martin R, Brandão BB, Kahn CR. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. *Cell Metab*. (2019) 30:656–73. doi: 10.1016/j.cmet.2019.07.011
6. Vader P, Breakefield XO, Wood MJ. Extracellular vesicles: emerging targets for cancer therapy. *Trends Mol Med*. (2014) 20:385–93. doi: 10.1016/j.molmed.2014.03.002
7. Thind A, Wilson C. Exosomal miRNAs as cancer biomarkers and therapeutic targets. *J Extracell Vesic*. (2016) 5:31292. doi: 10.3402/jev.v5.31292



8. Zhou B, Xu K, Zheng X, Chen T, Wang J, Song Y, et al. Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct Target Ther.* (2020) 5:144. doi: 10.1038/s41392-020-00258-9
9. Wu J, Hu S, Zhang L, Xin J, Sun C, Wang L, et al. Tumor circulome in the liquid biopsies for cancer diagnosis and prognosis. *Theranostics.* (2020) 10:4544–56. doi: 10.7150/tno.40532
10. Drula R, Ott LE, Berindan-Neagoe I, Pantel K, Calin GA. MicroRNAs from liquid biopsy derived extracellular vesicles: recent advances in detection and characterization methods. *Cancer (Basel).* (2020) 12:2009. doi: 10.3390/cancers12082009
11. Matsumura T, Sugimachi K, Iinuma H, Takahashi Y, Kurashige J, Sawada G, et al. Exosomal microRNA in serum is a novel biomarker of recurrence in human colorectal cancer. *Br J Cancer.* (2015) 113:275–81. doi: 10.1038/bjc.2015.201
12. Vautrot V, Chanteloup G, Elmallah M, Cordonnier M, Aubin F, Garrido C, et al. Exosomal miRNA: small molecules, big impact in colorectal cancer. *J Oncol.* (2019) 2019:8585276. doi: 10.1155/2019/8585276
13. Mousavi S, Moallem R, Hassanian SM, Sadehghzade M, Mardani R, Ferns GA, et al. Tumor-derived exosomes: potential biomarkers and therapeutic target in the treatment of colorectal cancer. *J Cell Physiol.* (2019) 234:12422–32. doi: 10.1002/jcp.28080
14. Rapado-González Ó, Álvarez-Castro A, López-López R, Iglesias-Canle J, Suárez-Cunquero MM, Muínelo-Romay L. Circulating microRNAs as promising biomarkers in colorectal cancer. *Cancer (Basel).* (2019) 11:898. doi: 10.3390/cancers11070898
15. Desmond BJ, Dennett ER, Danielson KM. Circulating extracellular vesicle microRNA as diagnostic biomarkers in early colorectal cancer—a review. *Cancer.* (2019) 12:52. doi: 10.3390/cancers12010052
16. Baassiri A, Nassar F, Mukherji D, Shamseddine A, Nasr R, Temraz S. Exosomal non coding RNA in LIQUID biopsies as a promising biomarker for colorectal cancer. *Int J Mol Sci.* (2020) 21:1398. doi: 10.3390/ijms21041398
17. Alves Dos Santos K, Clemente Dos Santos IC, Santos Silva C, Gomes Ribeiro H, de Farias Domingos I, Nogueira Silbiger V. Circulating exosomal miRNAs as biomarkers for the diagnosis and prognosis of colorectal cancer. *Int J Mol Sci.* (2020) 22:346. doi: 10.3390/ijms22010346
18. Ghosh D, Poisson L. “Omics” data and levels of evidence for biomarker discovery. *Genomics.* (2009) 93:13–6. doi: 10.1016/j.ygeno.2008.07.006
19. Oulas A, Minadakis G, Zachariou M, Sokratous K, Bourdakou M, Spyrou G. Systems bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches. *Brief Bioinform.* (2019) 20:806–24. doi: 10.1093/bib/bbx151
20. Sheng K, Kang L, Pridham K, Dunkenberg L, Sheng Z, Varghese R. An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. *Sci Rep.* (2020) 10:21246. doi: 10.1038/s41598-020-78126-3
21. Deyati A, Younesi E, Hofmann-Apitius M, Novac N. Challenges and opportunities for oncology biomarker discovery. *Drug Discov Today.* (2013) 18:614–24. doi: 10.1016/j.drudis.2012.12.011
22. Hu B, Ma X, Fu P, Sun Q, Tang W, Sun H, et al. miRNA-mRNA regulatory network and factors associated with prediction of prognosis in hepatocellular carcinoma. *Genom Proteom Bioinform.* (2021) 19:913. doi: 10.1016/j.gpb.2021.03.001
23. Zaka M, Sutton CW, Peng Y, Konur S. Model-based integration analysis revealed presence of novel prognostic miRNA targets and important cancer driver genes in triple-negative breast cancers. *Cancer (Basel).* (2020) 12:632. doi: 10.3390/cancers12030632
24. Chen J, Zhang D, Zhang W, Tang Y, Yan W, Guo L, et al. Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis. *J Transl Med.* (2013) 11:169. doi: 10.1186/1479-5876-11-169
25. Wang X, Chai Z, Pan G, Hao Y, Li B, Ye T, et al. ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer. *Brief Bioinform.* (2021) 22:bbaa088. doi: 10.1093/bib/bbaa088
26. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol.* (2016) 1418:93–110. doi: 10.1007/978-1-4939-3578-9\_5
27. Kuhn MJ. Building predictive models in R using the caret package. *J Stat Softw.* (2008) 28:1–26. doi: 10.18637/jss.v028.i05
28. Smyth GK. *Limma: Linear Models for Microarray Data.* *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* Berlin: Springer (2005). p. 397–420. doi: 10.1007/0-387-29362-0\_23
29. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp CJ. jvenn: an interactive Venn diagram viewer. *BMC Bioinform.* (2014) 15:1–7. doi: 10.1186/1471-2105-15-293
30. Kolde R, Kolde MR. Package ‘pheatmap’. *R Package.* (2015) 1:790.
31. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkol.* (2015) 19:A68. doi: 10.5114/wo.2014.47136
32. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8
33. Slattery ML, Herrick JS, Pellatt DF, Stevens JR, Mullany LE, Wolff E, et al. MicroRNA profiles in colorectal carcinomas, adenomas and normal colonic mucosa: variations in miRNA expression and disease progression. *Carcinogenesis.* (2016) 37:245–61. doi: 10.1093/carcin/bgv249
34. Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER 3rd, Kalocsay M, et al. Quantitative proteomics of the cancer cell line encyclopedia. *Cell.* (2020) 180:387–402 e16. doi: 10.1016/j.cell.2019.12.023
35. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* (2014) 42:e133–e. doi: 10.1093/nar/gku631
36. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nature Methods.* (2015) 12:697. doi: 10.1038/nmeth.3485
37. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* (2009) 37(suppl\_1):D105–10. doi: 10.1093/nar/gkn851
38. Chou C-H, Chang N-W, Shrestha S, Hsu S-D, Lin Y-L, Lee W-H, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucl Acids Res.* (2016) 44:D239–D47. doi: 10.1093/nar/gkv1258
39. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* (2018) 46:D239–D45. doi: 10.1093/nar/gkx1141
40. Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG. Functional microRNA targets in protein coding sequences. *Bioinformatics.* (2012) 28:771–6. doi: 10.1093/bioinformatics/bts043
41. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinform.* (2007) 8:69. doi: 10.1186/1471-2105-8-69
42. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* (2010) 11:R90. doi: 10.1186/gb-2010-11-8-r90
43. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* (2020) 48:D127–D31. doi: 10.1093/nar/gkz757
44. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, et al. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* (2012) 40:D180–D6. doi: 10.1093/nar/gkr1007
45. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet.* (2007) 39:1278–84. doi: 10.1038/ng2135
46. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* (2015) 4:7554. doi: 10.7554/eLife.05005
47. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genom.* (2016) 17:195–206. doi: 10.1186/s12864-016-2722-2
48. Lambert S, Jolma A, Campitelli L, Das P, Yin Y, Albu M, et al. The human transcription factors. *Cell.* (2018) 172:650–65. doi: 10.1016/j.cell.2018.01.029
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303
50. Chang L, Zhou G, Soufan O, Xia J. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res.* (2020) 48:W244–51. doi: 10.1093/nar/gkaa467
51. Consortium GO. Gene ontology consortium: going forward. *Nucleic Acids Res.* (2015) 43:D1049–56. doi: 10.1093/nar/gku1179

52. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* (2000) 28:27–30. doi: 10.1093/nar/28.1.27
53. Yu G, Wang L-G, Han Y, He Q-YJ. clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
54. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Bioinformatics.* (2019) 47(W1):W191–8. doi: 10.1093/nar/gkz369
55. Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* (2015) 31:2912–4. doi: 10.1093/bioinformatics/btv300
56. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* (2011) 12:77. doi: 10.1186/1471-2105-12-77
57. Friedman J, Hastie T, Tibshirani R. glmnet: lasso and elastic-net regularized generalized linear models. *R package version* (2009).
58. Therneau TM, Lumley T. Package 'survival'. *R Top Doc.* (2014) 2:3.
59. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* (2017) 45:W98–102. doi: 10.1093/nar/gkx247
60. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet.* (2013) 45:580–5. doi: 10.1038/ng.2653
61. Gao J, Aksoy B, Dogrusoz U, Dresdner G, Gross B, Sumer S, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* (2013) 6:pl1. doi: 10.1126/scisignal.2004088
62. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* (2016) 44:D231–D8. doi: 10.1093/nar/gkv1270
63. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* (2009) 25:1091–3. doi: 10.1093/bioinformatics/btp101
64. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature.* (2005) 435:834–8. doi: 10.1038/nature03702
65. Cagle P, Qi Q, Niture S, Kumar D. KCNQ1OT1: an oncogenic long noncoding RNA. *Biomolecules.* (2021) 11:1602. doi: 10.3390/biom11111602
66. Li K, Yao T, Zhang Y, Li W, Wang Z. NEAT1 as a competing endogenous RNA in tumorigenesis of various cancers: role, mechanism and therapeutic potential. *Int J Biol Sci.* (2021) 17:3428. doi: 10.7150/ijbs.62728
67. Yarla NS, Madka V, Pathuri G, Rao CV. Molecular targets in precision chemoprevention of colorectal cancer: an update from pre-clinical to clinical trials. *Int J Mol Sci.* (2020) 21:9609. doi: 10.3390/ijms21249609
68. Bahrami A, Khazaei M, Hasanzadeh M, ShahidSales S, Joudi Mashhad M, Farazestanian M, et al. Therapeutic potential of targeting PI3K/AKT pathway in treatment of colorectal cancer: rational and progress. *J Cell Biochem.* (2018) 119:2460–9. doi: 10.1002/jcb.25950
69. Lu D, Yao Q, Zhan C, Le-Meng Z, Liu H, Cai Y, et al. MicroRNA-146a promote cell migration and invasion in human colorectal cancer via carboxypeptidase M/src-FAK pathway. *Oncotarget.* (2017) 8:22674–84. doi: 10.18632/oncotarget.15158
70. Hwang W-L, Yang M-H. Numb is involved in the non-random segregation of subcellular vesicles in colorectal cancer stem cells. *Cell Cycle.* (2016) 15:2697–703. doi: 10.1080/15384101.2016.1218101
71. Schee K, Lorenz S, Worren MM, Gunther CC, Holden M, Hovig E, et al. Deep sequencing the microRNA transcriptome in colorectal cancer. *PLoS ONE.* (2013) 8:e66165. doi: 10.1371/journal.pone.0066165
72. Sastre D, Baiocchi J, de Souza Lima IM, Canto de Souza F, Corveloni AC, Thome CH, et al. Focused screening reveals functional effects of microRNAs differentially expressed in colorectal cancer. *BMC Cancer.* (2019) 19:1239. doi: 10.1186/s12885-019-6468-5
73. Taha M, Mitwally N, Soliman AS, Yousef E. Potential diagnostic and prognostic utility of miR-141, miR-181b1, and miR-23b in breast cancer. *Int J Mol Sci.* (2020) 21:8589. doi: 10.3390/ijms21228589
74. Hayashi M, Yamada S, Kurimoto K, Tanabe H, Hirabayashi S, Sonohara F, et al. miR-23b-3p plays an oncogenic role in hepatocellular carcinoma. *Ann Surg Oncol.* (2021) 28:3416–26. doi: 10.1245/s10434-020-09283-y
75. Zhang Y, Chen D, Zhang G, Wu X, Zhou L, Lin Y, et al. MicroRNA-23b-3p promotes pancreatic cancer cell tumorigenesis and metastasis via the JAK/PI3K and Akt/NF-kappaB signaling pathways. *Oncol Lett.* (2020) 20:160. doi: 10.3892/ol.2020.12021
76. Campos-Viguri GE, Peralta-Zaragoza O, Jimenez-Wences H, Longinos-Gonzalez AE, Castanon-Sanchez CA, Ramirez-Carrillo M, et al. MiR-23b-3p reduces the proliferation, migration and invasion of cervical cancer cell lines via the reduction of c-Met expression. *Sci Rep.* (2020) 10:3256. doi: 10.1038/s41598-020-60143-x
77. Bahnassy AA, Salem SE, El-Sayed M, Khorshid O, Abdellateif MS, Youssef AS, et al. MiRNAs as molecular biomarkers in stage II Egyptian colorectal cancer patients. *Exp Mol Pathol.* (2018) 105:260–71. doi: 10.1016/j.yexmp.2018.09.002
78. Ostenfeld MS, Jensen SG, Jeppesen DK, Christensen L-L, Thorsen SB, Stenvang J, et al. miRNA profiling of circulating EpCAM+extracellular vesicles: promising biomarkers of colorectal cancer. *J Extracell Vesic.* (2016) 5:31488. doi: 10.3402/jev.v5.31488
79. Kara M, Yumrutas O, Ozcan O, Celik OI, Bozgeyik E, Bozgeyik I, et al. Differential expressions of cancer-associated genes and their regulatory miRNAs in colorectal carcinoma. *Gene.* (2015) 567:81–6. doi: 10.1016/j.gene.2015.04.065
80. Zhang L, Zhang Y, Zhu H, Sun X, Wang X, Wu P, et al. Overexpression of miR-301a-3p promotes colorectal cancer cell proliferation and metastasis by targeting deleted in liver cancer-1 and runt-related transcription factor 3. *J Cell Biochem.* (2019) 120:6078–89. doi: 10.1002/jcb.27894
81. Liu J, Liu F, Li X, Song X, Zhou L, Jie J. Screening key genes and miRNAs in early-stage colon adenocarcinoma by RNA-sequencing. *Tumour Biol.* (2017) 39:1010428317714899. doi: 10.1177/1010428317714899
82. Vychytilova-Faltejskova P, Radova L, Sachlova M, Kosarova Z, Slaba K, Fabian P, et al. Serum-based microRNA signatures in early diagnosis and prognosis prediction of colon cancer. *Carcinogenesis.* (2016) 37:941–50. doi: 10.1093/carcin/bgw078
83. Zhang Q, Wei T, Shim K, Wright K, Xu K, Palka-Hamblin HL, et al. Atypical role of sprouty in colorectal cancer: sprouty repression inhibits epithelial-mesenchymal transition. *Oncogene.* (2016) 35:3151–62. doi: 10.1038/ncr.2015.365
84. Liu Q, Li A, Tian Y, Wu JD, Liu Y, Li T, et al. The CXCL8-CXCR1/2 pathways in cancer. *Cytokine Growth Factor Rev.* (2016) 31:61–71. doi: 10.1016/j.cytogfr.2016.08.002
85. Wang S, Pang L, Liu Z, Meng X. SERPINE1 associated with remodeling of the tumor microenvironment in colon cancer progression: a novel therapeutic target. *BMC Cancer.* (2021) 21:1–13. doi: 10.1186/s12885-021-08536-7