ORIGINAL RESEARCH

A Six-microRNA Signature Nomogram for Preoperative Prediction of Tumor Deposits in Colorectal Cancer

Shihan Xiao[1,2]
Jianping Guo[3,4]
Wuming Zhang[1,2]
Xianqin Hu[1,2]
Ran Wang[1,2]
Zhikang Chen[1,2,5],*
Chen Lai[1,2,5],*

[1]Department of General Surgery, Xiangya Hospital, Central South University, Changsha, Hunan, People's Republic of China; [2]International Joint Research Center of Minimally Invasive Endoscopic Technology Equipment & Standardization, Xiangya Hospital, Central South University, Changsha, Hunan Province, People's Republic of China; [3]Department of Gastrointestinal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, People's Republic of China; [4]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, the Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, People's Republic of China; [5]Hunan Key Laboratory of Precise Diagnosis and Treatment of Gastrointestinal Tumor, Xiangya Hospital Central South University, Changsha, Hunan Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zhikang Chen; Chen Lai
Department of General Surgery, Xiangya Hospital, Central South University, 87th Xiangya Road, Kaifu District, Changsha, Hunan, People's Republic of China
Tel +86-13875982443
Tel +86-13875982443
Email chenzhikang1974@126.com; chenl_csu@126.com

Purpose: Tumor deposits (TDs) are acknowledged negative prognostic factors in colorectal cancer (CRC), and their pathogenesis remains a puzzle. This study aimed to construct and validate a nomogram available for preoperative TDs prediction in CRC patients.

Patients and Methods: Patients from the Surveillance, Epidemiology, and End Results (SEER) and the cancer genome atlas (TCGA) databases were randomly divided into training and validation sets according to the sample size ratio of 7:3. Univariate logistic regression was performed for identifying differentially expressed microRNAs between TDs and non-TDs. Nomograms for TDs prediction were developed from the multivariate logistic regression model with least absolute shrinkage and selection operator and were validated internally in terms of accuracy, calibration, and clinical utility. Based on the target genes, pathways tightly associated with TDs were selected using enrichment analysis.

Results: Six clinicopathologic factors and expressions of six microRNAs (miR-614, miR-1197, miR-4770, miR-3136, miR-3173, and miR-4636) differed significantly between TDs and non-TDs CRC patients from the SEER and TCGA training sets. We compared potential prediction discrimination between two nomograms: a clinicopathologic nomogram and a six-microRNA signature nomogram. The six-microRNA signature nomogram revealed better accuracy than the clinicopathologic one for TDs prediction (AUC values of 0.96 and 0.93 in the validation cohort). The calibration plots and decision curve analysis demonstrated that the six-microRNA signature nomogram had better validity and a greater prognostic benefit versus the clinicopathologic one for TDs prediction. Calcium signaling pathways were closely associated with roles of the six microRNAs in TDs of CRC patients.

Conclusion: The six-microRNA signature nomogram can be used as an efficient tool for preoperative TDs prediction in CRC patients.

Keywords: tumor deposits, colorectal cancer, microRNA, prediction, nomogram

Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer and the second leading cause of cancer death worldwide in 2020, causing an estimated 1.9 million new cases and 935,000 deaths every year.[1] Radical surgery and chemoradiotherapy remain valid options for CRC patients. Notably, site-specific factors have been constantly introduced into the tumor node metastasis (TNM) staging system of the American Joint Committee on Cancer (AJCC) for better prognostic prediction and guidance of adjuvant therapies.[2]

The earliest identification of site-specific factors includes tumor deposits (TDs), characterized by discrete tumor nodules (without residual lymphoid tissue) present

in peri-colorectal fatty tissue and along the lymphatic drainage pathways.[3,4] TDs was first observed in rectal carcinoma by Gabriel in 1935.[5] Until recently, TDs was detected in many neoplasms other than colorectal carcinomas, like gastric, biliary, and pancreatic carcinomas.[6] But growing evidence supports the presence of TDs to be an independent prognosis factor in CRC, regardless of lymph node metastasis, leaving behind much debate on the genesis and roles of TDs therein.[7–10] Limited by the current appraisal of TDs that is available as a postoperative pathological examination only,[11,12] we are almost nowhere near an accurate TDs assessment before the operation. An easy and fast preoperative prediction using gene panels and CRC hallmarks may help tackle this issue.

Genomic and transcriptomic technology upgrades provide universal access to gene screening, including microRNAs (miRNAs), for diagnostic biomarker identification for cancers. miRNAs are a category of small endogenous noncoding RNA molecules frequently detectable in human blood.[13,14] They regulate cancer-related gene expression post-transcriptionally and are reported to serve as potential noninvasive indicators for the diagnosis and prognosis of malignancies.[15–19] But it is unknown why miRNA signatures for TDs prediction have not been reported in preexisting CRC studies.

In the article, we aimed to identify clinicopathologic and miRNA biomarkers and compare their efficacy in preoperative TDs prediction. We extracted baseline data on clinicopathologic information and miRNA sequencing of CRC patients from the Surveillance, Epidemiology, and End Results (SEER) and the cancer genome atlas (TCGA) databases. Based on the selected clinicopathologic factors and miRNAs tightly associated with TDs, we developed two candidate nomograms and validated their discrimination power in differentiating TDs from non-TDs. Besides, biofunctions and pathways involved in the pathogenesis of TDs were discussed.

Patients and Methods
Data Preparing and Preprocessing
We identified 137,879 patients who were diagnosed with colorectal carcinomas and underwent radical surgery between 2010 and 2015 from the SEER database of the US National Cancer Institute (http://seer.cancer.gov/), a publicly accessible database allowing access to 18-regional or statewide cancer statistics.[20] We excluded 7579 patients with no follow-up information, 1022 with unknown race, 77,197 with no detailed clinicopathologic documentation (eg, TDs, tumor size, TNM stage, tumor grade, positive lymph node count, number of regional nodes examined, circumferential resection margin [CRM], and perineural invasion [PI]), and 19,821 with missing data regarding serum carcinoembryonic antigen (CEA) and chemoradiotherapy data from the analysis. And this process resulted in a SEER cohort of 32,260 patients. In addition, we also downloaded data on 118 CRC patients with exact TDs status and other information from TCGA (https://portal.gdc. cancer.gov), another authoritative database with clinicopathologic information and RNA expression profiles. Finally, each eligible cohort was randomly split into a training and a validation set (ratio 7:3) (Figure 1). Comparisons of baseline characteristics between the two sets were performed using the Chi-square test and Fisher's exact test.

Variable Selection Method
Univariate logistic regression analysis was employed to analyze differences in patient demography, clinical characteristics, and histopathological features between TDs versus non-TDs cases from SEER. MiRNAs and genes differentially expressed between TDs and non-TDs CRC patients from TCGA were filtered based on the cutoff criteria of the p-value < 0.05 and $|\log_2$ fold change (FC)$| > 1$ using DESeq2 package. The expression profiles and distribution of differentially expressed miRNAs (DEMis) and genes (DEGs) were visualized into heatmaps and volcano plots, respectively.

The least absolute shrinkage and selection operator (LASSO) method is commonly used to overcome multicollinearity in high-dimensional regression.[21] Here, we performed a LASSO regression modeling to select DEMis and TDs-related clinicopathologic variables for predictive nomogram building. Specifically, we assumed the candidate miRNA signature and clinicopathologic variables were associated with TDs. The risk score was calculated based on the expression level of miRNAs and their regression coefficients as follows: risk score = $(0.1269 \times$ miR-614) + $(0.8450 \times$ miR-1197) + $(1.3584 \times$ miR-4770) + $(-2.1202 \times$ miR-3136) + $(-1.1673 \times$ miR-3173) + $(-0.2194 \times$ miR-4636).

Classification Models Fitting and Validation
Relevant data from the training cohort were subjected to multivariate logistic regression, upon which two nomograms (ie, a six-miRNA signature nomogram and a clinicopathologic nomogram) for TDs prediction in CRC were built using rms package. Their accuracy and discriminative ability were evaluated in training and
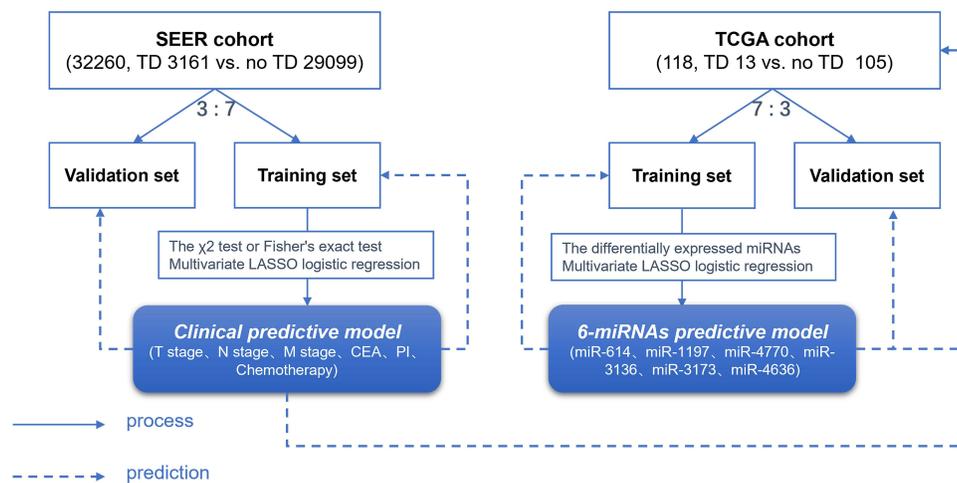
Figure 1 Flowchart of variables selection and models construction.

testing sets by the area under the receiver operating characteristic (ROC) curve (AUC) and calibration curves. Decision curve analysis (DCA) was conducted to evaluate the clinical application value of these models in the TCGA cohort using rmda package.

MiRNA-Targeted Gene Prediction and Functional Enrichment Analysis

We predicted target genes associated with the signature using TargetScan, miRDB, and miRTarBase algorithms, and those present in at least 2 of the 3 databases were screened out. The genes additionally interacting with TDs-related DEGs were selected for the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis. We mapped networks between the miRNA signature and overlapped genes using Cytoscape software version 3.7.2. All statistical analyses were conducted using R version 3.5.2 software (http://www.r-project.org/). In all tests, p-value < 0.05 (two-tailed) was considered statistically significant.

Results
Characteristics of the Study Population
Finally, 32,260 patients from SEER and 118 TCGA patients were enrolled. The incidence of TDs was similar in both cohorts, of 9.8% in the SEER cohort and 11% in the TCGA cohort. Subsequently, each cohort was randomly divided into a training and validation set according to the sample size ratio of 7:3. Clinicopathological characteristics of the training and validation sets are summarized in Table 1, demonstrating no significant difference

between both sets regardless of data source (p-value > 0.05 for all variables).

Clinicopathologic Risk Factors for TDs Prediction

The univariate logistic regression of the SEER training cohort initially identified the following candidates associated with TDs, including age, tumor size and location, pathologic T stage, lymph node metastasis (LNM), distant metastasis (DM), number of regional nodes examined, serum CEA, PI, CRM, and chemoradiotherapy (all p-value < 0.05; Table 2). Six optimal TDs-risk predictors for a multivariate logistic regression model with LASSO were confirmed (Supplementary Figure 1), incorporating higher T stage, positive lymph node count, DM, elevated CEA, PI, and neoadjuvant chemotherapy (Table 2). Therefore, a clinicopathologic diagnostic nomogram for preoperative TDs anticipation was developed for subsequent discrimination power comparisons.

Construction of TDs-Related miRNA Signature

We first used volcano plot analysis to identify TDs-related miRNAs (FC > 1.0 and p-value < 0.05) and recognized 38 DEMis and 1283 DEGs between TDs and non-TDs using the TCGA training set. Among them, 13 DEMis and 885 DEGs were upregulated, and 25 DEMis and 398 DEGs were downregulated. Their expression profiles were visualized by heatmaps, as shown in Figure 2. Then, predictive feature selection was carried out with LASSO, and six DEMis with non-zero coefficients were identified by logistic regression

Table 1 Characteristics of Training Sets and Validation Sets of the SEER Cohort and TCGA Cohort

Variables		SEER Cohort			TCGA Cohort		
		Training Set (N=22681)	Testing Set (N=9579)	P value	Training Set (N=88)	Testing Set (N=30)	P value
TDs	Negative	20,434 (90.1%)	8665 (90.5%)	0.323	77 (87.5%)	28 (93.3%)	0.587
	Positive	2247 (9.9%)	914 (9.5%)		11 (12.5%)	2 (6.7%)	
Age	≤60	8069 (35.6%)	3393 (35.4%)	0.791	22 (25.0%)	12 (40.0%)	0.117
	>60	14,612 (64.4%)	6186 (64.6%)		66 (75.0%)	18 (60.0%)	
Gender	Male	11,721 (51.7%)	4918 (51.3%)	0.589	49 (55.7%)	14 (46.7%)	0.52
	Female	10,960 (48.3%)	4661 (48.7%)		39 (44.3%)	16 (53.3%)	
Race	White	17,648 (77.8%)	7463 (77.9%)	0.777	71 (80.7%)	29 (96.7%)	0.0705
	Black	2610 (11.5%)	1116 (11.7%)		17 (19.3%)	1 (3.3%)	
	Other	2423 (10.7%)	1000 (10.4%)		–	–	
Location	Colon	18,772 (82.8%)	7897 (82.4%)	0.492	63 (71.6%)	23 (76.7%)	0.762
	Rectum	3909 (17.2%)	1682 (17.6%)		25 (28.4%)	7 (23.3%)	
Tumor size	<2cm	1917 (8.5%)	809 (8.4%)	1	–	–	
	≥ 2cm	20,764 (91.5%)	8770 (91.6%)		–	–	
Grade	I/II	18652 (82.2%)	7982 (83.3%)	0.019	–	–	
	III/IV	4029 (17.8%)	1597 (16.7%)		–	–	
T stage	T1	1689 (7.4%)	751 (7.8%)	0.619	1 (1.1%)	3 (10.0%)	0.113
	T2	3490 (15.4%)	1464 (15.3%)		17 (19.3%)	7 (23.3%)	
	T3	13,762 (60.7%)	5768 (60.2%)		57 (64.8%)	17 (56.7%)	
	T4	3740 (16.5%)	1596 (16.7%)		13 (14.8%)	3 (10.0%)	
LNM	0	13,290 (58.6%)	5650 (59.0%)	0.613	49 (55.7%)	17 (56.7%)	0.483
	1–3	5870 (25.9%)	2495 (26.0%)		24 (27.3%)	11 (36.7%)	
	4–6	1929 (8.5%)	775 (8.1%)		5 (5.7%)	1 (3.3%)	
	≥7	1592 (7.0%)	659 (6.9%)		10 (11.4%)	1 (3.3%)	
DM	No	19,804 (87.3%)	8427 (88.0%)	0.106	74 (84.1%)	27 (90.0%)	0.621
	Yes	2877 (12.7%)	1152 (12.0%)		14 (15.9%)	3 (10.0%)	
Examined	≤15	8440 (37.2%)	3533 (36.9%)	0.585	33 (37.5%)	7 (23.3%)	0.233
	>15	14,241 (62.8%)	6046 (63.1%)		55 (62.5%)	23 (76.7%)	
CEA	Negative	13,380 (59.0%)	5731 (59.8%)	0.166	59 (67.0%)	23 (76.7%)	0.448
	Positive	9301 (41.0%)	3848 (40.2%)		29 (33.0%)	7 (23.3%)	
CRM	Negative	19,500 (86.0%)	8246 (86.1%)	0.81	–	–	
	Positive	3181 (14.0%)	1333 (13.9%)		–	–	

(*Continued*)

Table 1 (Continued).

Variables		SEER Cohort			TCGA Cohort		
		Training Set (N=22681)	Testing Set (N=9579)	P value	Training Set (N=88)	Testing Set (N=30)	P value
PI	Negative	19,925 (87.8%)	8371 (87.4%)	0.258	58 (65.9%)	21 (70.0%)	0.852
	Positive	2756 (12.2%)	1208 (12.6%)		30 (34.1%)	9 (30.0%)	
Radiotherapy	No	19,778 (87.2%)	8283 (86.5%)	0.0779	–	–	
	Yes	2903 (12.8%)	1296 (13.5%)		–	–	
Chemotherapy	No	12,291 (54.2%)	5218 (54.5%)	0.65	50 (56.8%)	17 (56.7%)	1
	Yes	10,390 (45.8%)	4361 (45.5%)		38 (43.2%)	13 (43.3%)	

Abbreviations: TDs, tumor deposits; LNM, lymph node metastasis; DM, distant metastasis; Examined, examined lymph node number; CEA, carcinoembryonic antigen; CRM, circumferential resection margin; PI, perineural invasion.

(Supplementary Figure 1), incorporating miR-614, miR-1197, miR-4770, miR-3136, miR-3173, and miR-4636. Finally, a multivariable logistic regression model was developed to generate a six-miRNA-based classifier for predicting TDs status based on these DEMis. The signature-based predictive risk score could be calculated by multiplying their regression coefficients as follows: risk score = $(0.1269 \times miR\text{-}614) + (0.8450 \times miR\text{-}1197) + (1.3584 \times miR\text{-}4770) + (-2.1202 \times miR\text{-}3136) + (-1.1673 \times miR\text{-}3173) + (-0.2194 \times miR\text{-}4636)$.

Construction and Validation of a Six-miRNA Nomogram for TDs Prediction

The nomograms for performance comparisons were constructed using data from the validation cohort. As for clinicopathologic nomogram building, the total risk score was used and described by the sum of risk scores for all categories selected through the multivariate LASSO logistic regression analysis of the SEER training cohort. A total score represented the probability of TDs risk of the patient. Similarly, the miRNA signature was constructed using miRNA sequencing data from the TCGA training cohort (Supplementary Figure 2). ROC curves for analysis of their prediction performances showed that the AUCs of the clinicopathologic nomogram in preoperative TDs anticipation were 0.79 and 0.77 in the SEER training and SEER testing cohorts, respectively (Figure 3A). The AUCs for the six-miRNA signature nomogram were 0.96 and 0.96 in the TCGA training and TCGA testing cohorts (Figure 3B). Then, we further applied the clinicopathological model and the six-miRNA model to

the TCGA cohort, and the AUCs was 0.93 and 0.96 (Figure 3C), respectively. Moreover, the calibration curves of both models exhibited high consistency between the observed and predicted results (Supplementary Figure 2). ROC curves and calibration plots for the accuracy and validity among two models demonstrated the best agreement between predicted and observed probability in the six-miRNA signature nomogram (Figure 3C and D and Supplementary Figure 2). In order to further verify this result, we have made the DCA curves of two models, and the results show that if the threshold probability was lower than 55%, the six-miRNA nomogram could add more net prognosis benefits to more accurate anticipation than the clinicopathologic nomograms (Figure 3E). So, the six-miRNA signature showed a better prediction.

Identification of Target Genes Associated with the Six-miRNA Signature and Biological Pathways Involved

To unveil the roles of the six DEMis screened for the signature, we predicted their target genes from the above miRNA-related databases, and those reported in at least 2 of the 3 databases and also interacting with TDs-related DEGs were recognized as target genes. The Venn diagrams yielded two overlapping genes associated with miR-614, four with miR-1197, ten with miR-4770, 16 with miR-3136, 23 with miR-3173, and one with miR-4636. They were identified through the overlap between 1505 DEGs and 1373 DEMis target genes (Figure 4). The regulatory networks revealing associations between target genes and each DEMi were visualized by Cytoscape, as shown in Figure 5.

Table 2 Univariate and Multivariate Logistic Regression Analysis of TDs-Related Clinicopathologic Factors in Training Set

Variables		Univariate Logistic Regression Analysis		Multivariate Logistic Regression Analysis	
		Odd Ratio	**P value**	**Odd Ratio**	**P value**
Age	<50	Reference	<0.001	–	–
	50–70	0.70		–	
	≥70	0.53		–	
Gender	Male	Reference	0.569	–	–
	Female	0.97		–	
Race	White	Reference	0.618	–	–
	Black	1.02		–	
	Other	0.94		–	
Location	Colon	Reference	0.006	–	–
	Rectum	1.17		–	
Tumor size	<2cm	Reference	<0.001	–	–
	≥ 2cm	2.47		–	
Grade	I/II	Reference	<0.001	–	–
	III/IV	1.70		–	
T stage	T1	Reference	<0.001	Reference	<0.001
	T2	2.31		2.04	
	T3	9.10		4.23	
	T4	21.99		6.80	
LNM	0	Reference	<0.001	Reference	<0.001
	1–3	3.78		2.09	
	4–6	5.91		2.53	
	≥7	7.71		2.84	
DM	No	Reference	<0.001	Reference	<0.001
	Yes	4.39		1.83	
Examined	≤15	Reference	<0.001	–	–
	>15	0.77		–	
CEA	Negative	Reference	<0.001	Reference	<0.001
	Positive	2.12		1.22	
CRM	Negative	Reference	<0.001	–	–
	Positive	2.00		–	
PI	Negative	Reference	<0.001	Reference	<0.001
	Positive	4.17		2.09	
Radiotherapy	No	Reference	<0.001	–	–
	Yes	1.28		–	
Chemotherapy	No	Reference	<0.001	Reference	<0.001
	Yes	4.03		1.92	

Abbreviations: TDs, tumor deposits; LNM, lymph node metastasis; DM, distant metastasis; Examined, examined lymph node number; CEA, carcinoembryonic antigen; CRM, circumferential resection margin; PI, perineural invasion.

We performed enrichment analysis for target genes to understand potential functions and pathways related to the six DEMis. The results showed that regulation of membrane potential, transmembrane transporter complex, and substrate-specific channel activity were the top enriched terms regarding biological processes (BP),
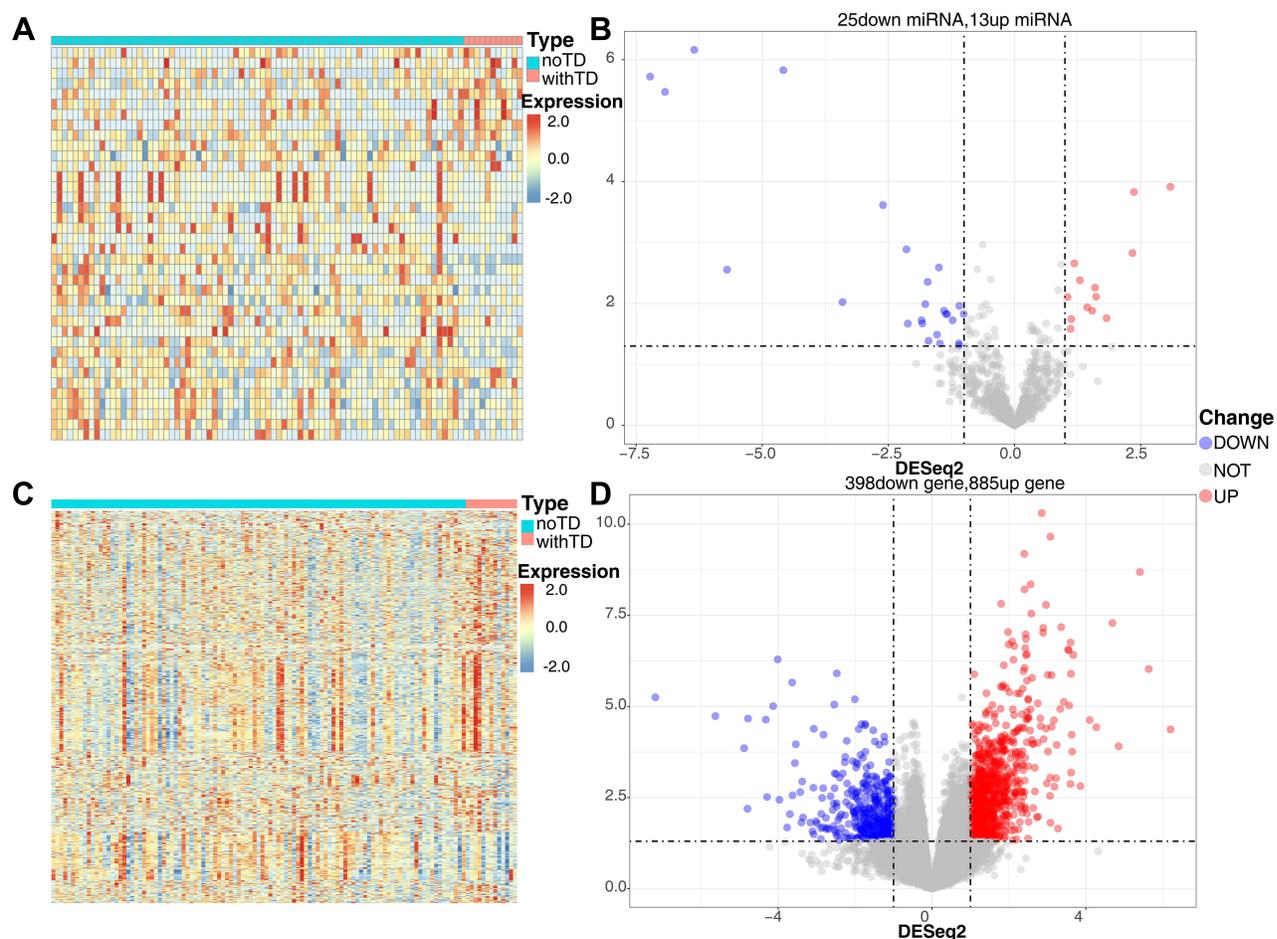
Figure 2 Differentially expressed of miRNAs and genes. Heatmap (**A**) and volcano plot (**B**) visualized the differentially expressed miRNAs between CRC patients with and without TDs. Heatmap (**C**) and volcano plot (**D**) visualized the differentially expressed genes between CRC patients with and without TDs. The blue patches and dots represented the downregulated miRNA and genes; the red patches and dots represented the upregulated miRNA and genes, and the gray dots represented the miRNA and genes without differential expression.

cellular components (CC), and molecular functions (MF) (Figure 5). The KEGG analysis revealed that the calcium signaling pathway were robustly enriched in the target genes (Figure 5).

Discussion

TDs have been well accepted as an important negative prognostic factor in CRC and were the first to be introduced to the 5th edition of the AJCC/TNM staging system for CRC.[22] Frequent modification of the definition of TDs in the latter staging systems over the years also highlighted the importance of TDs in cancer prognosis and management.[4,23] However, its origin and possible mechanism in CRC remain controversial. Restrained by current sampling and examination techniques, we have no access now to preoperative prediction of TDs status in cancer patients.

Specific biomarkers that allow a fast and readily evaluation of TDs status may help tackle this issue. This study produced a clinicopathologic nomogram (composed of six risk factors: T stage, LNM, DM, CEA, PI, and neoadjuvant chemotherapy) and a six-miRNA signature nomogram (composed of six miRNA). By comparing the AUCs, DCA curves and calibration plots between the two models, we successfully prove that the six-miRNA signature nomogram had better validity and a greater accuracy versus the clinicopathologic one for TDs prediction.

For all the efforts to upgrade the knowledge about prognostic values of TDs,[24–27] we still fall far short of predictors for TDs, particularly those targeting its origin. Some researchers consider they are potentially positive lymph nodes that are no longer identifiable due to total substitution by tumor metastasis.[28,29] Some believe that TDs should be regarded as a systemic disease rather than
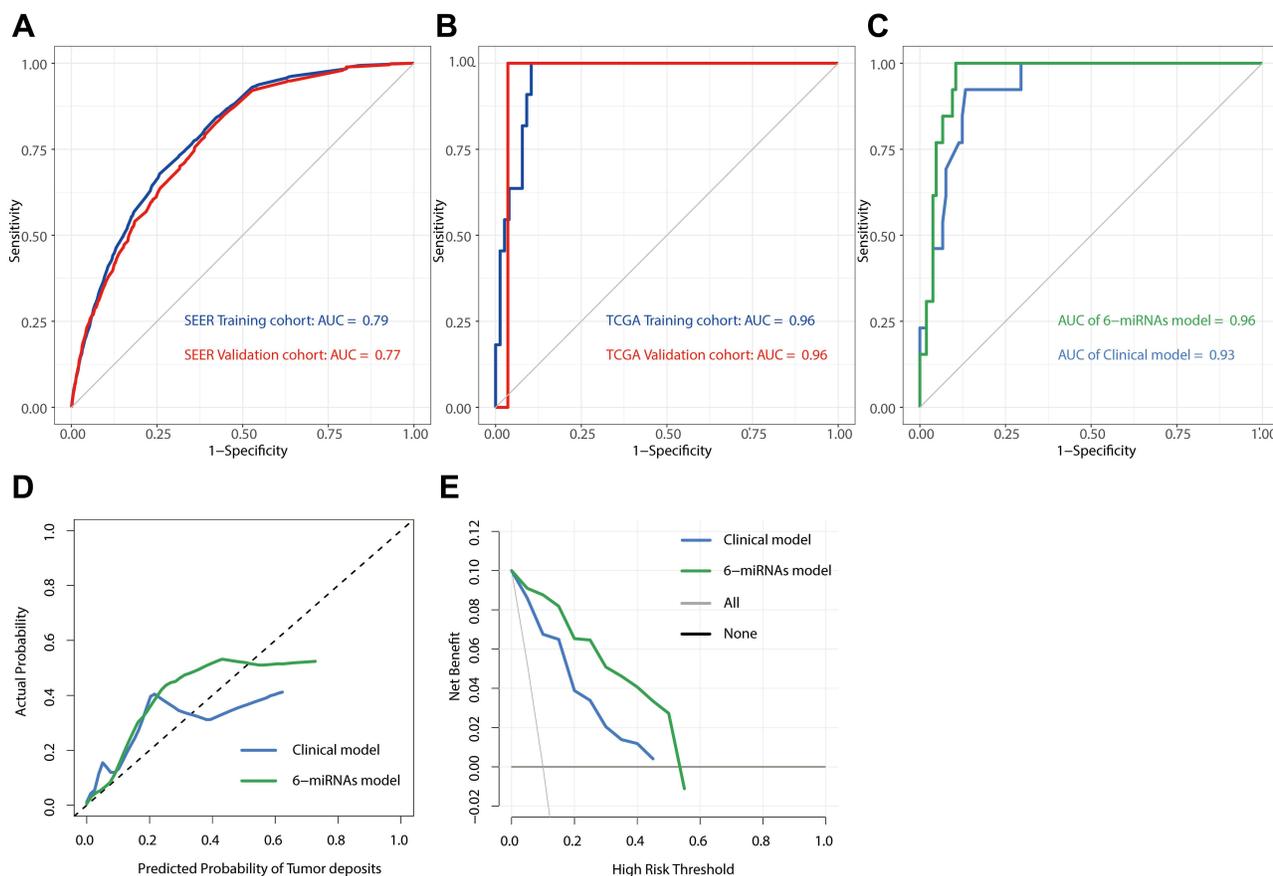
Figure 3 Comparison of nomograms for TDs prediction.The receiver operating characteristic (ROC) curve showed the area under curve (AUC) of clinicopathologic nomogram in SEER training and validation cohorts (**A**), the AUC of six-miRNA signature nomogram in TCGA training and validation cohorts (**B**), and the AUC of clinicopathologic nomogram and six-miRNA nomogram in TCGA cohort (**C**). The calibration plot compared the validity among the two models TCGA cohort (**D**). And the decision curve compared the clinical application among the two models in TCGA cohort (**E**).

a local disease as it represents a unique metastasis mode within or along vessels, nerves, or lymphatic channels.[28,30] Nevertheless, although the origin of TDs remains unclear, the 7th and 8th editions of the TNM/AJCC system give an explicit classification for TDs: isolated tumor foci without histologic evidence of a residual lymph nodule identified in the pericolic or perirectal adipose tissue away from the primary tumors.[3,23] And the 8th edition added a clarification that if any vascular or neural structure was recognizable, the nodes should be classified into lymphovascular invasion (LVI) or perineural invasion (PI) correspondingly. These indicate that TDs identification relies on histopathological techniques, such as H&E staining.[3,12] Unfortunately, current imaging tools and clinical parameters fail to provide access to preoperative TDs recognition. Despite some techniques like MRI texture analysis or phenotype determinations that have been introduced to roughly assess the TDs status, an accurate

prediction shaping personalized TDs features remains an arduous task.[31–34]

A predictive nomogram is currently an efficient tool to improve preoperative TDs prediction and investigate related mechanisms. Our nomogram incorporated six-gene risk score system (miR-614, miR-1197, miR-4770, miR-3136, miR-3173, and miR-4636) and exhibited satisfactory performances in high-risk group identification. The six clinicopathologic risk factors for TDs (higher T stage, positive lymph node count, DM, pathologically elevated CEA, PI, and neoadjuvant chemotherapy) are consistent with previous findings, although the AJCC/TNM staging system has definitely differentiated TDs from tumor extension, LNM, and DM.[30,35–37] And some studies even support the inclusion of TDs in a T, N, or M category.[23,27,38] The relationship between TDs and PI and neoadjuvant chemotherapy can be explained by diverse origins of TDs: tumor cells growing within or along nerves channels, the fragmentation of advanced tumors after neoadjuvant chemotherapy or tumors incompletely regressed after
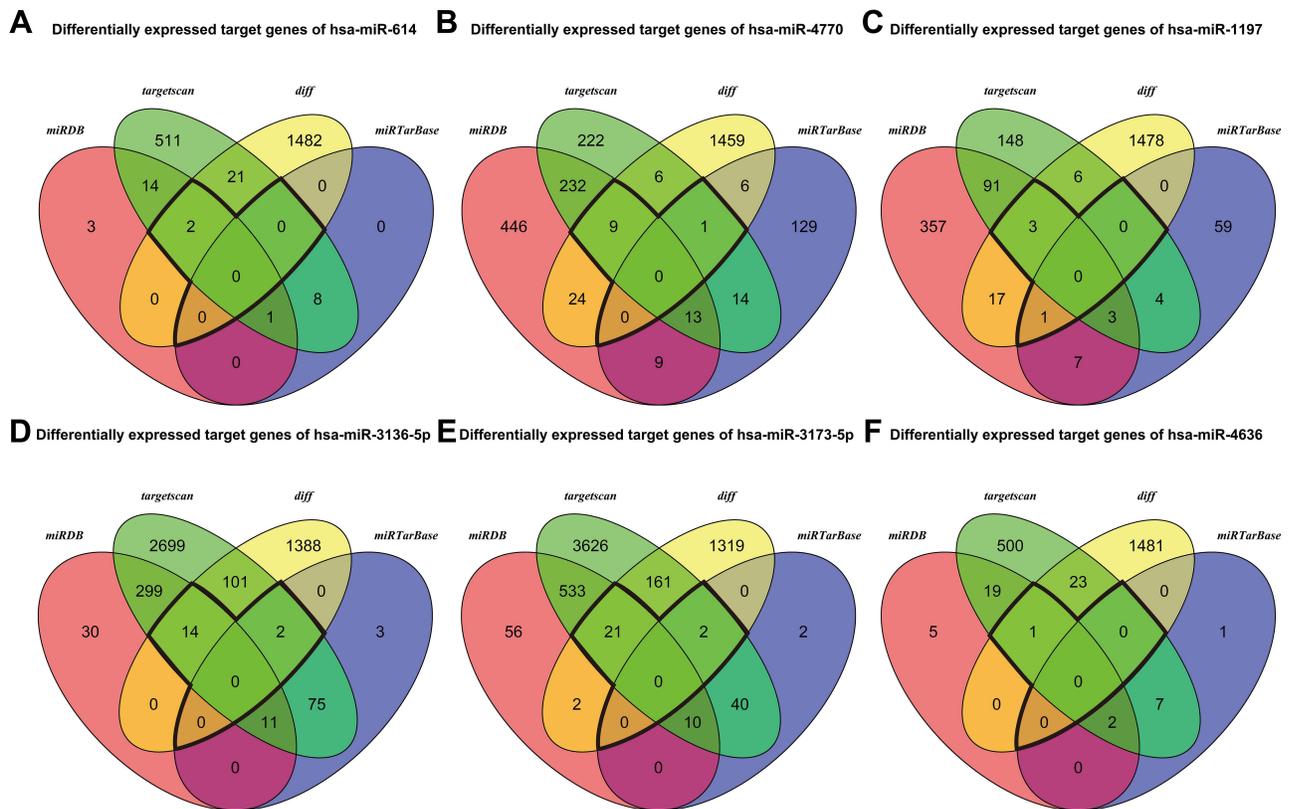
A Differentially expressed target genes of hsa-miR-614 **B** Differentially expressed target genes of hsa-miR-4770 **C** Differentially expressed target genes of hsa-miR-1197

D Differentially expressed target genes of hsa-miR-3136-5p **E** Differentially expressed target genes of hsa-miR-3173-5p **F** Differentially expressed target genes of hsa-miR-4636

Figure 4 The Venn diagram of the target genes of the six TDs-related DEMis. The Venn diagrams displayed two overlapping genes associated with miR-614 (**A**), ten with miR-4770 (**B**), four with miR-1197 (**C**), 16 with miR-3136 (**D**), 23 with miR-3173 (**E**) and one with miR-4636 (**F**), which were identified from the overlap between 1505 DEGs and 1373 DEMis target genes.

preoperative chemotherapy.[39,40] The association between TDs risk and a higher CEA level needs verifying in future studies.

Six TDs-related miRNAs were also identified in our study, including three upregulated (miR-614, miR-1197, and miR-4770) and three downregulated miRNAs (miR-3136, miR-3173, and miR-4636). Preexisting studies have shown that four of these miRNAs (miR-614, miR-1197, miR-3173, and miR-4636) can be used as prognostic factors in other cancers (eg, ovarian cancer, non-small cell lung cancer, B-cell acute lymphoblastic leukaemia, cervical cancer, gastric cancer, and pancreatic ductal adenocarcinoma).[41–47] MiR-614 can promote proliferation and inhibit apoptosis of ovarian cancer cells via suppressing PPP2R2A expression.[41] MiR-1197 has been recognized as a negative prognostic indicator in lung cancer and pancreatic cancer and may serve as a therapeutic target.[42–44] MiR-3173 downregulation has been found in B-cell acute lymphoblastic leukemia, which facilitated cell invasion.[45] MiR-4636 inhibits the proliferation, migration, and invasion of gastric cancer cells and serves as a favorable prognostic biomarker for cervical cancer survival.[46,47] The subsequent functional enrichment analysis of differentially

expressed target genes interacting with these miRNAs included calcium signaling pathways. And the GO annotation results indicated that these target genes were primarily involved in passive transmembrane transporter activity and channel activity, adding the evidence that TDs pathogenesis may be associated with calcium signaling pathways. Most studies agree that the dysregulated calcium signaling pathways play a critical role in CRC recurrence, metastasis, and prognosis.[48,49] An increase of intracellular calcium could fuel tumor proliferation, while cell apoptosis also resulted from sustained calcium increases in cell.[50] Such dual-function pathways in CRC open a new gateway into underlying molecular mechanisms related to TDs pathogenesis.

As we have understood both the clinicopathologic and the six-miRNA signature nomograms exhibit high agreement between the predicted and the actual probability of TDs status, we further estimated the discrimination validity of these two nomograms by comparing the AUCs of TCGA validation cohort. As expected, the six-miRNA signature nomogram showed better prediction with higher accuracy in preoperative TDs status assessment versus the clinicopathologic nomogram
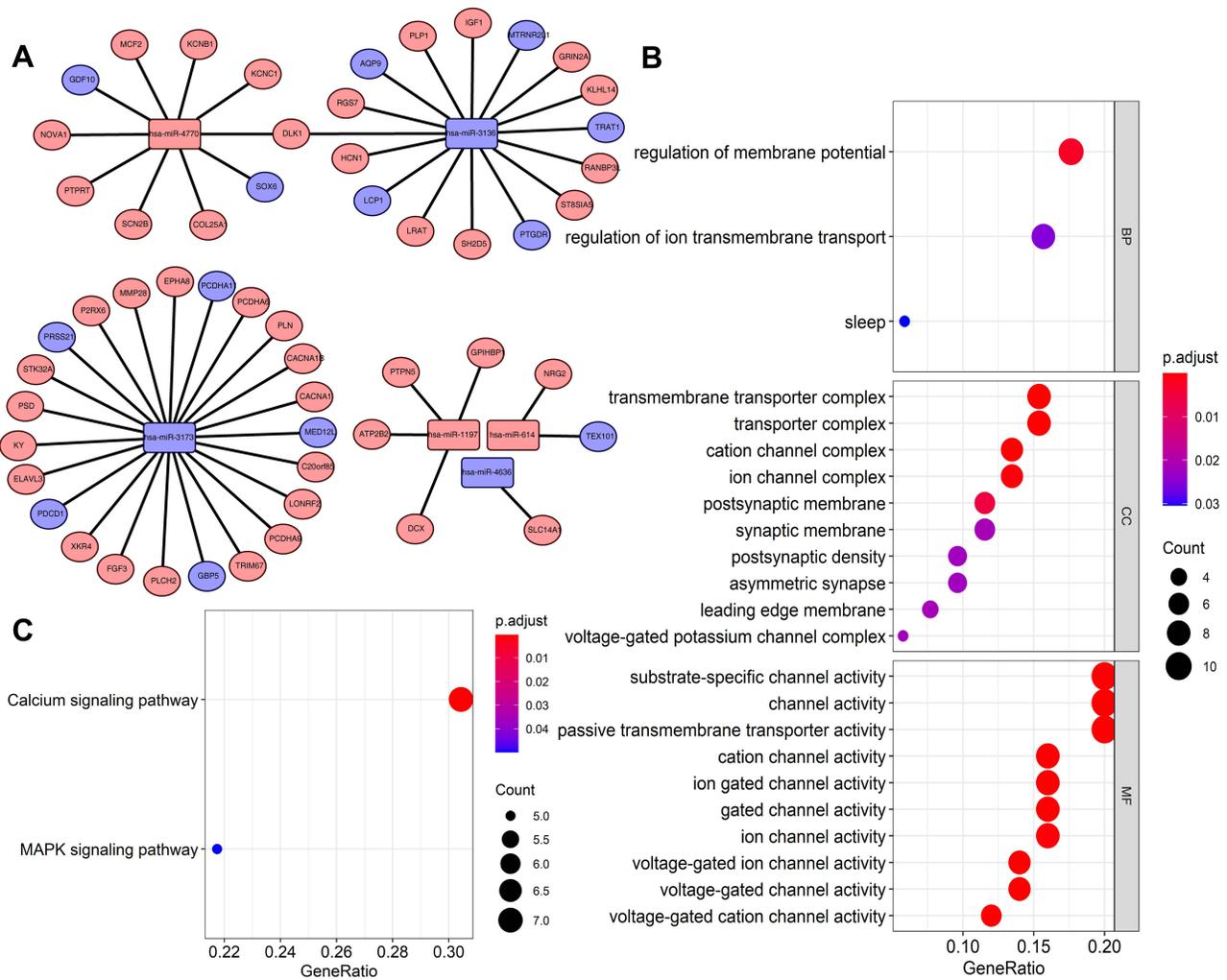
Figure 5 Regulatory networks between the six miRNA and their target genes (**A**), and functional enrichment analysis using the Gene Ontology (GO, (**B**) and Kyoto Encyclopedia of Genes and Genomes (KEGG, (**C**) pathways enrichment analysis.

(AUC values of 0.96 and 0.93, respectively). DCA analysis for clinical application further demonstrated that if the threshold probability was lower than 55%, the six-miRNA signature nomogram could offer a higher net prognostic benefit than the clinicopathologic nomograms. Therefore, we confirm that our six-miRNA signature nomogram has a favorable and reliable prediction efficacy for TDs status prediction and encourage its popularization in CRC diagnosis as an accurate, personalized decision-making tool.

The originality of our work is the unique nomogram for preoperative TDs prediction in CRC, which was developed by screening risk factors from hallmark miRNA sets of suitable patients and validated using data from two authoritative databases. Its application value was supported by DCA analysis. Enrichment analysis focused on the function of calcium signaling pathways, upon which future studies can investigate TDs genesis in various cancer types. Besides, we acknowledge that

this work has several limitations. Our conclusion should be carefully validated in larger-sample clinical trials as the current bioinformatics analysis lacks external validation and can only provide limited evidence based on the relatively insufficient sample size. Multicenter studies can offer more sources of independent cohorts to produce stronger results in the prediction efficacy and application value of this nomogram. Moreover, animal studies are needed for elucidating experimental evidence on the special roles of calcium signaling in CRC with TDs.

Conclusion

In conclusion, our study offers an nomogram composed of six-miRNA signature, which plays an excellent role in preoperative TDs prediction in CRC patients from authoritative databases. This predictive nomogram may serve as an efficient decision-making tool for practitioners in both TDs

assessment and personalized CRC treatment. And future studies are required to elucidate the roles of the six miRNAs and calcium signaling pathways in the pathogenesis of TDs.

Abbreviations

AJCC, American Joint Committee on Cancer; AUC, area under the receiver operating characteristics curve; BP, biological processes; CC, cellular components; CEA, carcinoembryonic antigen; CRC, colorectal cancer; CRM, circumferential resection margin; DCA, decision curve analysis; DEGs, differentially expressed genes; DEMis, differentially expressed miRNAs; DM, distant metastasis; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; HCC, hepatocellular carcinoma, LASSO, least absolute shrinkage and selection operator; LNM, lymph node metastasis; MF, molecular functions; miRNA, microRNA; PI, perineural invasion; ROC, receiver operating characteristics; SEER, Surveillance, Epidemiology, and End Results; TCGA, the cancer genome atlas; TDs, tumor deposits; TNM, tumor node metastasis.

Data Sharing Statement

The datasets generated and analysed during the current study are available in the Surveillance, Epidemiology, and End Results (SEER) and the cancer genome atlas (TCGA) databases (http://seer.cancer.gov/ and https://portal.gdc.cancer.gov).

Ethics Approval and Informed Consent

This study has been reviewed by the Medical Ethics Committee of Xiangya Hospital of Centre South University and has been granted ethical waiver because TCGA and SEER database belong to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open source data, so there are no ethical issues and other conflicts of interest.

Acknowledgments

Author' Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Chen VW, Hsieh MC, Charlton ME, et al. Analysis of stage and clinical/prognostic factors for colon and rectal cancer from SEER registries: AJCC and collaborative stage data collection system. *Cancer*. 2014;120 Suppl 23:3793–3806. doi:10.1002/cncr.29056
3. Weiser MR. AJCC 8th Edition: colorectal Cancer. *Ann Surg Oncol*. 2018;25(6):1454–1455. doi:10.1245/s10434-018-6462-1
4. Jin M, Roth R, Rock JB, et al. The impact of tumor deposits on colonic adenocarcinoma AJCC TNM staging and outcome. *Am J Surg Pathol*. 2015;39(1):109–115. doi:10.1097/pas.0000000000000320
5. Ersen A, Unlu MS, Akman T, et al. Tumor deposits in gastric carcinomas. *Pathol Res Pract*. 2014;210(9):565–570. doi:10.1016/j.prp.2014.03.006
6. Puppa G, Ueno H, Kayahara M, et al. Tumor deposits are encountered in advanced colorectal cancer and other adenocarcinomas: an expanded classification with implications for colorectal cancer staging system including a unifying concept of in-transit metastases. *Mod Pathol*. 2009;22(3):410–415. doi:10.1038/modpathol.2008.198
7. Landau MA, Zhu B, Akwuole FN, et al. Histopathological predictors of recurrence in Stage III colon cancer: reappraisal of tumor deposits and tumor budding using AJCC8 criteria. *Int J Surg Pathol*. 2019;27(2):147–158. doi:10.1177/1066896918787275
8. Nagtegaal ID, Knijn N, Hugen N, et al. Tumor deposits in colorectal cancer: improving the value of modern staging-a systematic review and meta-analysis. *J Clin Oncol*. 2017;35(10):1119–1127. doi:10.1200/jco.2016.68.9091
9. Pei JP, Zhang CD, Liang Y, et al. A modified pathological N stage including status of tumor deposits in colorectal cancer with nodal metastasis. *Front Oncol*. 2020;10:548692. doi:10.3389/fonc.2020.548692

10. Zheng P, Chen Q, Li J, et al. Prognostic significance of tumor deposits in patients with Stage III colon cancer: a nomogram study. *J Surg Res*. 2020;245:475–482. doi:10.1016/j.jss.2019.07.099

11. Peparini N. Beyond T, N and M: the impact of tumor deposits on the staging and treatment of colorectal and gastric carcinoma. *Surg Oncol*. 2018;27(2):129–137. doi:10.1016/j.suronc.2018.02.007

12. Jin M, Frankel WL. Lymph node metastasis in colorectal cancer. *Surg Oncol Clin N Am*. 2018;27(2):401–412. doi:10.1016/j.soc.2017.11.011

13. Romano G, Veneziano D, Acunzo M, et al. Small non-coding RNA and cancer. *Carcinogenesis*. 2017;38(5):485–491. doi:10.1093/carcin/bgx026

14. Iorio MV, Croce CM. microRNA involvement in human cancer. *Carcinogenesis*. 2012;33(6):1126–1133. doi:10.1093/carcin/bgs140

15. Dawood S. Novel biomarkers of metastatic cancer. *Expert Rev Mol Diagn*. 2010;10(5):581–590. doi:10.1586/erm.10.35

16. Lv Y, Huang Z, Lin Y, et al. MiRNA expression patterns are associated with tumor mutational burden in lung adenocarcinoma. *Oncoimmunology*. 2019;8(10):e1629260. doi:10.1080/2162402x.2019.1629260

17. Nishiwada S, Sho M, Banwait JK, et al. A MicroRNA signature identifies pancreatic ductal adenocarcinoma patients at risk for lymph node metastases. *Gastroenterology*. 2020;159(2):562–574. doi:10.1053/j.gastro.2020.04.057

18. Qu A, Yang Y, Zhang X, et al. Development of a preoperative prediction nomogram for lymph node metastasis in colorectal cancer based on a novel serum miRNA signature and CT scans. *EBioMedicine*. 2018;37:125–133. doi:10.1016/j.ebiom.2018.09.052

19. Wang M, Li R, Zou X, et al. A miRNA-clinicopathological nomogram for the prediction of central lymph node metastasis in papillary thyroid carcinoma-analysis from TCGA database. *Medicine (Baltimore)*. 2020;99(35):e21996. doi:10.1097/md.0000000000021996

20. Wingo PA, Jamison PM, Hiatt RA, et al. Building the infrastructure for nationwide cancer surveillance and control–a comparison between the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program (United States). *Cancer Causes Control*. 2003;14(2):175–193. doi:10.1023/a:1023002322935

21. Zhou ZR, Wang WW, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann Transl Med*. 2019;7(23):796. doi:10.21037/atm.2019.08.63

22. Sobin LH, Fleming ID. TNM Classification of Malignant Tumors, fifth edition (1997). Union Internationale Contre le Cancer and the American Joint Committee on Cancer. *Cancer*. 1997;80 (9):1803–1804. doi:10.1002/(sici)1097-0142(19971101)80:9<1803::aid-cncr16>3.0.co;2-9

23. Tong LL, Gao P, Wang ZN, et al. Is the seventh edition of the UICC/AJCC TNM staging system reasonable for patients with tumor deposits in colorectal cancer? *Ann Surg*. 2012;255(2):208–213. doi:10.1097/SLA.0b013e31821ad8a2

24. Chavali LB, Llanos AAM, Yun JP, et al. Radiotherapy for patients with resected tumor deposit-positive colorectal cancer: a surveillance, epidemiology, and end results-based population study. *Arch Pathol Lab Med*. 2018;142(6):721–729. doi:10.5858/arpa.2017-0099-OA

25. Delattre JF, Cohen R, Henriques J, et al. Prognostic value of tumor deposits for disease-free survival in patients with Stage III colon cancer: a post hoc analysis of the IDEA France Phase III trial (PRODIGE-GERCOR). *J Clin Oncol*. 2020;38(15):1702–1710. doi:10.1200/jco.19.01960

26. Li ZH, Xie PY, Zhang DF, et al. Nomogram for predicting disease-free survival among a multicenter cohort of Chinese patients with locally advanced rectal cancer. *Cancer Manag Res*. 2019;11:2471–2483. doi:10.2147/cmar.S196614

27. Wang Y, Zhang J, Zhou M, et al. Poor prognostic and staging value of tumor deposit in locally advanced rectal cancer with neoadjuvant chemoradiotherapy. *Cancer Med*. 2019;8(4):1508–1520. doi:10.1002/cam4.2034

28. Nagtegaal ID, Quirke P. Colorectal tumour deposits in the mesorectum and pericolon; a critical review. *Histopathology*. 2007;51 (2):141–149. doi:10.1111/j.1365-2559.2007.02720.x

29. Song YX, Gao P, Wang ZN, et al. Can the tumor deposits be counted as metastatic lymph nodes in the UICC TNM staging system for colorectal cancer? *PLoS One*. 2012;7(3):e34087. doi:10.1371/journal.pone.0034087

30. Yagi R, Shimada Y, Kameyama H, et al. Clinical significance of extramural tumor deposits in the lateral pelvic lymph node area in low rectal cancer: a retrospective study at two institutions. *Ann Surg Oncol*. 2016;23(Suppl 4):552–558. doi:10.1245/s10434-016-5379-9

31. Atre ID, Eurboonyanun K, Noda Y, et al. Utility of texture analysis on T2-weighted MR for differentiating tumor deposits from mesorectal nodes in rectal cancer patients, in a retrospective cohort. *Abdom Radiol (NY)*. 2021;46(2):459–468. doi:10.1007/s00261-020-02653-w

32. Fassan M, Vianello L, Sacchi D, et al. Assessment of intratumor immune-microenvironment in colorectal cancers with extranodal extension of nodal metastases. *Cancer Cell Int*. 2018;18:131. doi:10.1186/s12935-018-0634-8

33. Kim H, Shin S, Kim Y, et al. The clinicopathologic significance of extranodal tumor extension in locally advanced (pT3) colorectal adenocarcinoma and its association with the loss of E-cadherin expression. *Int J Clin Exp Pathol*. 2019;12(9):3417–3425.

34. Zhang M, Hu W, Hu K, et al. Association of KRAS mutation with tumor deposit status and overall survival of colorectal cancer. *Cancer Causes Control*. 2020;31(7):683–689. doi:10.1007/s10552-020-01313-0

35. Al Sahaf O, Myers E, Jawad M, et al. The prognostic significance of extramural deposits and extracapsular lymph node invasion in colon cancer. *Dis Colon Rectum*. 2011;54(8):982–988. doi:10.1097/DCR.0b013e31821c4944

36. Yamano T, Semba S, Noda M, et al. Prognostic significance of classified extramural tumor deposits and extracapsular lymph node invasion in T3-4 colorectal cancer: a retrospective single-center study. *BMC Cancer*. 2015;15:859. doi:10.1186/s12885-015-1885-6

37. Shimada Y, Takii Y. Clinical impact of mesorectal extranodal cancer tissue in rectal cancer: detailed pathological assessment using whole-mount sections. *Dis Colon Rectum*. 2010;53(5):771–778. doi:10.1007/DCR.0b013e3181cf7fd8

38. Wang S, Guan X, Ma M, et al. Reconsidering the prognostic significance of tumour deposit count in the TNM staging system for colorectal cancer. *Sci Rep*. 2020;10(1):89. doi:10.1038/s41598-019-57041-2

39. Quirke P, Williams GT, Ectors N, et al. The future of the TNM staging system in colorectal cancer: time for a debate? *Lancet Oncol*. 2007;8(7):651–657. doi:10.1016/s1470-2045(07)70205-x

40. Greene FL. Tumor deposits in colorectal cancer: a moving target. *Ann Surg*. 2012;255(2):214–215. doi:10.1097/SLA.0b013e3182430eaa

41. Zhang J, Gao D, Zhang H. Upregulation of miR-614 promotes proliferation and inhibits apoptosis in ovarian cancer by suppressing PPP2R2A expression. *Mol Med Rep*. 2018;17(5):6285–6292. doi:10.3892/mmr.2018.8714

42. Sun B, Hua J, Cui H, et al. MicroRNA-1197 downregulation inhibits proliferation and migration in human non- small cell lung cancer cells by upregulating HOXC11. *Biomed Pharmacother*. 2019;117:109041. doi:10.1016/j.biopha.2019.109041

43. Yang S, Yu F, Lin M, et al. Single-nucleotide polymorphism rs17548629 in RIPK1 gene may be associated with lung cancer in a young and middle-aged Han Chinese population. *Cancer Cell Int*. 2020;20:143. doi:10.1186/s12935-020-01215-w

44. You L, Wang J, Zhang F, et al. Potential four-miRNA signature associated with T stage and prognosis of patients with pancreatic ductal adenocarcinoma identified by co-expression analysis. *Mol Med Rep*. 2019;19(1):441–451. doi:10.3892/mmr.2018.9663

45. Tian L, Cao J, Ji Q, et al. The downregulation of miR-3173 in B-cell acute lymphoblastic leukaemia promotes cell invasion via PTK2. *Biochem Biophys Res Commun*. 2017;494(3–4):569–574. doi:10.1016/j.bbrc.2017.10.013

46. Yin S, Yang M, Li X, et al. Peripheral blood circulating microRNA-4636/-143 for the prognosis of cervical cancer. *J Cell Biochem*. 2020;121(1):596–608. doi:10.1002/jcb.29305

47. Tang J, Hu Y, Zhang C, et al. miR-4636 inhibits tumor cell proliferation, migration and invasion, and serves as a candidate clinical biomarker for gastric cancer. *Oncol Lett*. 2021;21(1):33. doi:10.3892/ol.2020.12294

48. Yang X, Lou J, Shan W, et al. Pathogenic roles of altered calcium channels and transporters in colon tumorogenesis. *Life Sci*. 2019;239:116909. doi:10.1016/j.lfs.2019.116909

49. Lee JM, Davis FM, Roberts-Thomson SJ, et al. Ion channels and transporters in cancer. 4. Remodeling of Ca(2+) signaling in tumor-igenesis: role of Ca(2+) transport. *Am J Physiol Cell Physiol*. 2011;301(5):C969–76. doi:10.1152/ajpcell.00136.2011

50. Wang W, Yu S, Huang S, et al. A complex role for calcium signaling in colorectal cancer development and progression. *Mol Cancer Res*. 2019;17(11):2145–2153. doi:10.1158/1541-7786.Mcr-19-0429