Research Article

Single-Cell RNA Sequencing Reveals the Role of Epithelial Cell Marker Genes in Predicting the Prognosis of Colorectal Cancer Patients

Kai-yu Shen [ID],[1] **Bin-yu Chen** [ID],[1] **and Wen-cang Gao** [ID][2]

[1]*The Second Clinical Medical College of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China*
[2]*Department of Oncology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China*

Correspondence should be addressed to Wen-cang Gao; 20164580@zcmu.edu.cn

Single-cell RNA sequencing (scRNA-seq) is increasingly used in studies on gastrointestinal cancers. This study investigated the prognostic value of epithelial cell-associated biomarkers in colorectal cancer (CRC) using scRNA-seq data. We downloaded and analysed scRNA-seq data from four CRC samples from the Gene Expression Omnibus (GEO), and we identified marker genes of malignant epithelial cells (MECs) using CRC transcriptome and clinical data downloaded from The Cancer Genome Atlas (TCGA) and GEO as training and validation cohorts, respectively. In the TCGA training cohort, weighted gene correlation network analysis, univariate Cox proportional hazard model (Cox) analysis, and least absolute shrinkage and selection operator regression analysis were performed on the marker genes of MEC subsets to identify a signature of nine prognostic MEC-related genes (MECRGs) and calculate a risk score based on the signature. CRC patients were divided into high- and low-risk groups according to the median risk score. We found that the MECRG risk score was significantly correlated with the clinical features and overall survival of CRC patients, and that CRC patients in the high-risk group showed a significantly shorter survival time. The univariate and multivariate Cox regression analyses showed that the MECRG risk score can serve as an independent prognostic factor for CRC patients. Gene set enrichment analysis revealed that the MECRG signature genes are involved in fatty acid metabolism, p53 signalling, and other pathways. To increase the clinical application value, we constructed a MECRG nomogram by combining the MECRG risk score with other independent prognostic factors. The validity of the nomogram is based on receiver operating characteristics and calibration curves. The MECRG signature and nomogram models were well validated in the GEO dataset. In conclusion, we established an epithelial cell marker gene-based risk assessment model based on scRNA-seq analysis of CRC samples for predicting the prognosis of CRC patients.

1. Introduction

Colorectal cancer (CRC) is among the most common malignant tumours of the digestive system worldwide, and its high incidence and mortality are second only to those of lung cancer and breast cancer [1, 2]. Unhealthy lifestyle habits, such as smoking, drinking, and consuming a high-fat diet, have led to an increase in CRC incidence [3], with 1.5% annual increase reported in people aged 30 to 39 years from 2007 to 2016 [4]. At present, tumour node metastasis (TNM) staging, histopathology, and completion of surgical resection are mainly evaluated to determine CRC prognosis, and molecular markers are widely used for CRC diagnosis and treatment. Various medical treatments, including surgery, postoperative adjuvant chemoradiotherapy, and molecular targeted therapy, have been used to treat CRC [5]. CRC symptoms are generally minor in early disease stages, and therefore, patients are often at an advanced stage of the disease by the time they seek medical attention. Furthermore, the optimal treatment timing can easily be missed

in cases of advanced, metastatic, and recurrent CRC, and the effects of conventional treatment are usually inconsistent, resulting in variable prognoses.

As a malignant tumour of epithelial origin, the development and progression of CRC are closely related to epithelial tissue [6, 7]. Normal epithelial cells have antitumor activities and are able to eliminate oncogenic transformed cells by regulating cytoskeletal proteins [8]. According to Lv et al., epithelial cells secrete periostin, which inhibits the growth of gastric cancer cells by stabilising P53 and E-cadherin proteins [9]. Epithelial cells exhibit apical-basal polarity and cell–cell adhesion. Correct regulation of polarity is essential to inhibit tumour growth [10]. Royer and Lu suggested that the malignant transformation of epithelial cells in the presence of oncogene activation is generally closely associated with the loss of cell polarity and disorganisation and that the disruption of epithelial cell polarity promotes epithelial-mesenchymal transition (EMT), which is a key step in the invasion of the surrounding stroma by epithelial tumour cells [11]. In addition, under hypoxic conditions, pulmonary epithelial cells can downregulate connexin (CX)26 and CX43 via the P53 signalling pathway to promote lung carcinogenesis [12]. Epithelial cells are both the structural basis of most organ tissues in the body and the source of most human tumours [13]. With the progressive research on epithelial cells, epithelial cell-related biomarkers have become a research hotspot in recent years. Keratins are widely used as epithelial cell biomarkers in the pathological diagnosis of tumours and in predicting survival prognosis. In CRC patients, decreased expression of keratin (K)8 and K20 is closely related to EMT and suggestive of poor survival and high tumour invasion [14]. Increased serum levels of a cleaved K18 fragment produced by apoptotic epithelial cells suggest a high risk of CRC recurrence within 3 years [15]. However, because of the uncontrollable pathological types, stages, and metastases of CRC [16], using conventional biomarkers to predict prognosis does not achieve adequate results. Therefore, to explore more epithelial cell-related biomarkers with clinical application potential, we used high-resolution omics tools to perform a more accurate analysis of CRC epithelial cells.

Single-cell RNA sequencing (scRNA-seq) allows the construction of a gene-regulatory network at the cellular level [17]. Analysis of the genome, transcriptome, or epigenome of single cells individually or simultaneously enables the detection of gene expression profiles and tracking of cell development at the single-cell level [18]. The RNA-seq technology has been widely used to evaluate the tumour immune microenvironment [19, 20]; however, few studies have applied this method to CRC epithelial cells to predict CRC prognosis.

In this study, we identified marker genes of CRC epithelial cell subsets using scRNA-seq analysis, determined the prognostic significance of nine malignant epithelial cell- (MEC-) related genes (MECRGs) using data from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), and integrated the MECRG-based risk score and clinical traits to construct a nomogram for the prediction of CRC patient prognosis.

2. Materials and Methods

2.1. Data Availability. ScRNA-seq data of four CRC samples were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (GSE161277). Relevant clinical information related to the CRC samples is provided in Supplementary Table S1. The dataset comprises four CRC samples, four adenoma samples, three normal tissue samples, one paracancer sample, and one blood sample [21]. Tissue samples other than the CRC samples were not considered in this study. Transcriptome data of 473 CRC samples and 41 nontumour tissues were obtained from the TCGA database (https://portal.gdc.cancer.gov/), including 435 CRC samples with matched clinical data. The extracted clinical information included sex, age, and stage. TCGA-CRC samples with complete clinical information were used to construct the training set used to develop a prediction score. A total of 232 CRC samples with complete clinical information obtained from the GEO database (GSE17538) were used as the external validation set.

2.2. scRNA-seq Data Processing. The scRNA-seq data of the four CRC samples were processed using the R language (v4.1.0; https://www.r-project.org/) and the "Seurat" package [22] (https://cran.r-project.org/web/packages/Seurat/index.html). The "harmony" package was used to remove batch effects from the scRNA-seq data of the four CRC samples. First, we determined the percentage of mitochondrial genes in each cell using the "PercentageFeatureSet" function with the parameter set to pattern = "^MT-". Subsequently, using the "subset" function, genes expressed in <10 cells and cells expressing <200 genes were eliminated. Further, we excluded noncells and cell aggregates. Cell samples with a mitochondrial gene proportion < 15% were included in the subsequent analysis and logarithmically normalized. Principal component analysis (PCA) was used for unsupervised clustering, and the "JackStraw" function was used to determine and visualise the number of principal components. We used nonlinear dimension reduction (t-distributed stochastic neighbour embedding; t-SNE) clustering and the "FindAllMarkers" function (parameters: minimum% = 0.3, log function threshold = 0.25) to identify marker genes based on between-cluster differences. The cell clusters were then annotated based on reported cell-specific marker genes [23–29].

2.3. Analysis of Chromosome Copy Number Variation (CNV) and Screening of MEC Marker Genes. To distinguish epithelial cell malignancy, we identified chromosome CNV in each sample using the "infercnv" package [30] (https://github.com/broadinstitute/inferCNV; default parameters). Using immune cells as a reference [31], we calculated CNV scores for epithelial cell subsets and defined the epithelial cell subset showing a median CNV score > 1,000 as the MEC subset, which was used for subsequent analysis. Differences in CNV scores between epithelial cell clusters were compared using the Kruskal–Wallis test. We then identified marker genes in the MEC subset based on a $|\log 2(fold\ change)| > 0.5$ and a false discovery rate (FDR) < 0.01.

2.4. Analysis of Differential Gene Enrichment of Epithelial Cell Subsets. The potential biological mechanisms of the MEC marker genes were determined using the "cluster-Profiler" (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) and "org.Hs.eg.db" (https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html) packages. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed, using a q value < 0.05 to determine statistically significant enrichment.

2.5. MEC Marker Gene Analysis Using Weighted Gene Correlation Network Analysis (WGCNA). WGCNA can be used to find modules of correlated genes and identify disease-related biomarkers. We used the WGCNA R package [32] (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/) to identify MEC marker genes related to CRC. We generated a similarity matrix between MEC marker genes using the Pearson correlation analysis and then calculated an adjacency matrix and constructed a topological overlap matrix. Next, we plotted a tree diagram of modules and clustered closely related MEC marker genes within the modules using hierarchical clustering. The MEC marker genes in the final correlation module ($P < 0.05$) were used for subsequent analysis.

2.6. Construction and Validation of the MECRG Signature. The univariate Cox analysis was used to find MECRGs significantly related to overall survival (OS) in the TCGA-CRC cohort, and a forest map was plotted. MECRGs were identified using least absolute shrinkage and selection operator (LASSO) analysis, using 10-fold cross-validation and 1,000 iterations to determine the minimum value of the penalty parameter (λ) and construct a MECRG signature. The regression coefficients of the nine MECRGs thus identified were calculated using the "predict" function. The following formula was used to calculate the risk score:

$$\text{MECRG risk score} = \sum X, G * \text{coef } G, \qquad (1)$$

where "coef G" is the regression coefficient and "$X G$" describes the expression levels of the MECRGs. Patients from the TCGA-CRC training cohort were divided into high- and low-risk groups according to the median value of the MECRG risk score. A Kaplan–Meier survival curve was used to explore differences in survival and prognosis between the two groups. We then used a receiver operating characteristic (ROC) curve to evaluate the predictive value of the MECRG signature. Finally, the signature was validated using the GSE17538 GEO dataset.

2.7. Analysis of the Prognostic Accuracy of the MECRG Risk Score. Combining the MECRG risk score with clinical features, we used univariate and multivariate Cox analyses to assess whether the risk score could serve as an independent prognostic factor. Using the training and validation sets, we then performed a survival analysis of the MECRG signature for patients of different clinical subgroups. Additionally, we identified the relationships between MECRG risk groups

and clinical traits (including sex, age, and TNM stage) and generated a heat map.

2.8. Gene Set Enrichment Analysis (GSEA) of the MECRG Signature. Patients in the TCGA-CRC training cohort were grouped according to the median value of the risk score and we used GSEA to evaluate the functions of and signalling pathways associated with MECRGs in the high- and low-risk groups, using $P < 0.05$ as a threshold. The "AUCell" package (https://www.rdocumentation.org/packages/msigdbr/versions/7.4.1) was used to present the results of enrichment analysis according to cell subsets.

2.9. Construction of a MECRG Nomogram. Using the training set, we integrated the three independent prognostic factors of age, stage, and MECRG risk group to plot a MECRG nomogram capable of predicting the 1-, 3-, and 5-year OS of CRC patients for clinical application. ROC and calibration curves were used to evaluate the predictive value of the MECRG nomogram, and feasibility was confirmed by external validation using the GSE17538 GEO dataset.

2.10. Statistical Analysis. The "survival" (https://cran.r-project.org/web/packages/survival/index.html) and "survminer" (https://cran.r-project.org/web/packages/survminer/index.html) packages were used to construct Kaplan–Meier survival curves. The log-rank test was used to determine significant differences in survival between the high- and low-risk groups according to the training and validation datasets. LASSO regression was performed using the "glmnet" package (https://cran.r-project.org/web/packages/glmnet/index.html), and the "timeROC" package (https://cran.r-project.org/web/packages/timeROC/index.html) was used to generate the ROC curve to evaluate model accuracy. The univariate and multivariate Cox regression analyses were used to determine the independent predictors of OS outcomes in CRC patients. The "rms" package (https://cran.r-project.org/web/packages/rms/) was used to construct the nomogram model. Comparisons between two groups were made using the Wilcoxon rank sum test, and comparisons between multiple groups were made using the Kruskal–Wallis test. Correlation analysis was conducted using the Pearson method. $P < 0.05$ was considered significant.

3. Results

3.1. CRC scRNA-seq Data Analysis. We obtained 15,465 cells from the four CRC samples. After applying the screening criteria, 8,798 high-quality cell samples were obtained. Gene numbers (nFeature_RNA), sequencing depth (nCount_RNA), and mitochondrial gene percentage (percent.mt) are shown in Supplementary Figure S1. The Pearson correlation coefficient between gene count and sequencing depth was 0.92 (Supplementary Figure S2), suggesting a positive correlation. PCA used to classify the high-quality cells identified 40 principal components (Supplementary Figure S3), and t-SNE of the first 11 principal components ($P < 0.05$) (Supplementary Figure S4) allowed visualisation of the high-dimensional CRC scRNA-seq data and the distribution of the cell subsets, as well as classification of

the cells into 18 subclasses (Figures 1(a) and 1(b)). Among the 18 subclasses, clusters 0, 4, 7, 9, 10, and 11 were identified as epithelial cell subtypes 1, 2, 3, 4, 5, and 6, respectively, based on the presence of epithelial cell marker genes (*EPCAM*, *KRT19*, and *CDH1*), clusters 2, 12, 13, 15, and 17 were identified as B cell subsets (*CD79A*, *MS4A1*, and *CD19*), clusters 1, 3, 6, and 8 were identified as T cell subsets (*CD3D*, *CD8A*, and *CD3E*), clusters 5 and 16 were identified as macrophage subsets (*CD14*, CD68, and CD163), and cluster 14 was identified as an endothelial cell subset (*IL3RA*, *SERPINF1*, and *CCDC88A*). For the purpose of selecting epithelial cell subsets with high malignancy, we did not merge the epithelial cell subsets. Based on the above annotation effects, we summarised the final cell subpopulation annotations in Figure 2(a). In addition, we calculated the proportions of the five cell types in the four CRC samples, which revealed that the epithelial cell subpopulation accounted for a relatively large fraction (Figure 2(b)). The five most strongly expressed marker genes in each cell subset are shown in Figure 2(c).

3.2. Evaluation of Chromosome CNV in the Epithelial Cell Subsets.

We next determined the chromosome CNV in each sample based on the transcriptome data to evaluate the degree of malignancy in the epithelial cell subsets. We observed low CNV in the immune cell subsets (macrophages, B cells, and T cells) in control samples, whereas high CNV was observed in epithelial cells. Chromosome amplification mainly occurred in chromosomes 7, 8, 9, 12, 13, 16, 19, 20, and 21, and deletions were most prevalent in chromosomes 4, 6, 8, 12, 14, 15, 17, 18, 19, and 22 (Figure 3(a)). The MEC subpopulation was screened using a median CNV score > 1,000 as the threshold. Notably, the CNV scores for epithelial cell subtypes 1 through 5 were more significant than epithelial cell subtype 6 (Figure 3(b) and Supplementary Table S2), suggesting a higher degree of malignancy of CRC lesions associated with these cell subsets. Therefore, we defined these as MEC subsets, and, using $|\log 2(fold\ change)| > 0.5$ and FDR < 0.01 as thresholds, we identified 1,259 marker genes (Supplementary Table S3), and we speculated that 1259 MEC marker genes function differently in CRC than in normal colonic epithelial cells and therefore require further analysis.

3.3. GO and KEGG Enrichment Analyses of MEC Subset Marker Genes.

We performed GO function (Figure 4(a)) and KEGG pathway (Figure 4(b)) enrichment analyses of the 1,259 marker genes in the five MEC subsets to determine their possible biological functions. We found that the differentially expressed genes in the MEC subsets were mainly involved in processes associated with ATP metabolism, focal adhesion, formation of cell–substrate junctions, cadherin binding, formation of adherens junctions, reactive oxygen species (ROS), and thyroid cancer. Notably, cell–substrate junction formation, calmodulin binding, and adhesion junctions are closely related to the characteristics of epithelial cells themselves, ROS is strongly associated with tumor progression, and as thyroid cancer and CRC are both epithelial-derived malignancies, we hypothesised that common epithelial cell-related genes might be involved in the development of both. These results tentatively suggest that marker genes of the MEC subpopulation are involved in the occurrence and progression of CRC mainly through the above-mentioned biological mechanisms.

3.4. WGCNA of MEC Marker Genes.

Using the TCGA-CRC cohort, we performed WGCNA of the expression profiles of the identified 1259 marker genes. We clustered the MEC genes into modules associated with clinical traits ("tumour" and "normal") based on a soft threshold of $\beta = 7$. As shown in Figure 5, to prevent meaningful MEC marker genes from being missed, with the exception of the turquoise module ($P > 0.05$), all genes in the remaining modules were significantly associated with clinical characteristics ("tumour" and "normal") ($P < 0.05$) and were used in subsequent analyses. To build a clinical prediction model by linking to the clinical characteristics of CRC patients, we initially identified 787 genes significantly associated with clinical characteristics ("tumour" and "normal") and further identified MEC marker genes associated with survival prognosis.

3.5. Establishment and Validation of a MECRG Signature for Predicting CRC Patient Survival.

The univariate Cox analysis (Figure 6(a)) and LASSO regression analysis (Supplementary Figure S7) of the TCGA-CRC training cohort identified 47 MEC marker genes capable of predicting OS, the regression coefficients of the nine MECRGs were calculated using the "predict" function, and nine MECRGs were identified based on the minimum λ ($\lambda = 0.01970$): galectin 2 (*LGALS2*), glycerophosphodiester phosphodiesterase 1 (*GDE1*), monocyte chemoattractant protein 1 (*MPC1*), bone marrow stromal cell antigen 2 (*BST2*), tropomyosin 2 (*TPM2*), PRELI domain-containing 2 (*PRELID2*), G protein subunit gamma 5 (*GNG5*), calcyphosin (*CAPS*), and calcium voltage-gated channel subunit alpha 1D (*CACNA1D*). Validation using the GSE17538 data identified *BST2*, *TPM2*, and *CAPS* as risk genes (hazard ratio (HR) > 1) and *LGALS2*, *GDE1*, *MPC1*, *GNG5*, *PRELID2*, and *CACNA1D* as protective genes (HR < 1). The MECRG risk score was defined as follows: (−0.13614 ∗ *expression* of *LGALS*2) + (−0.60881 ∗ *expression* of *GDE*1) + (−0.49609 ∗ *expression* of *MPC*1) + (−0.20530 ∗ *expression* of *GNG*5) + (0.05613 ∗ *expression* of *BST*2) + (0.12894 ∗ *expression* of *TPM*2) + (−0.16624 ∗ *expression* of *PRELID*2) + (0.48635 ∗ *expression* of *CAPS*) + (−0.83941 ∗ *expression* of *CACNA1D*). The final risk scores for all CRC samples in the training and validation cohorts are shown in Supplementary Tables S4 and S5. The MECRGs were predominantly distributed in the epithelial cell subsets (Figure 6(b)).

Next, we divided the patients into high- and low-risk groups according to the median MECRG risk score. The Kaplan–Meier survival analysis demonstrated that in the TCGA-CRC training and GSE17538 validation cohorts, the survival time of the high-risk group was shorter than that of the low-risk group ($P < 0.001$) (Figures 6(c) and 6(d)). In both cohorts, an increase in the risk score was
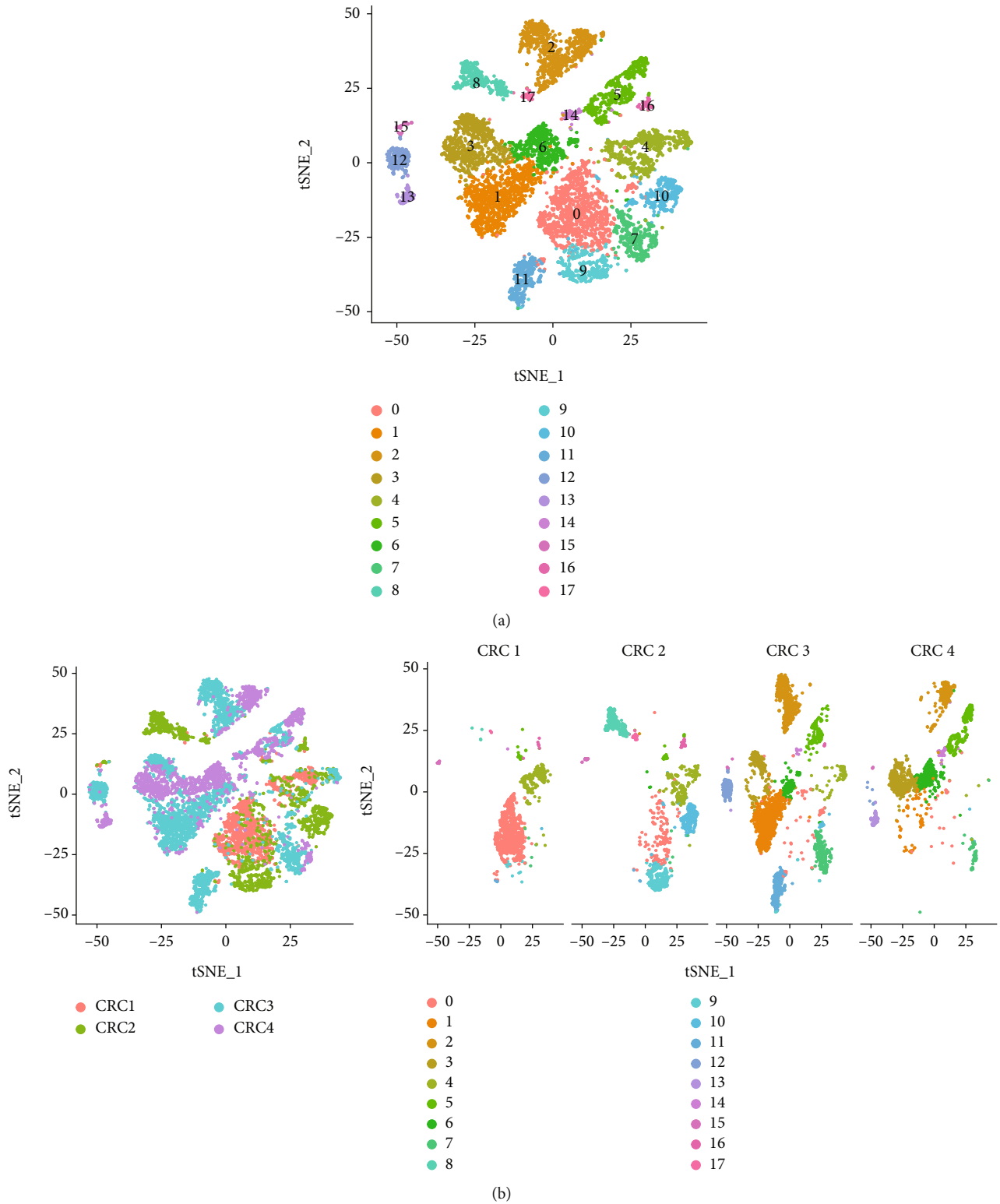
(a)

(b)

Figure 1: Characterisation of scRNA-seq from 15,465 cells. (a) The cells were classified into 18 subsets using the t-SNE algorithm. (b) Distribution ratios of the cell subsets in the four CRC samples.

accompanied by an increase in patient mortality (Figures 6(e) and 6(f)). The area under the ROC curve (AUC) at 1, 3, and 5 years was 0.746, 0.726, and 0.710, respectively, in the training cohort, and 0.618, 0.668, and 0.658, respectively, in the validation cohort (Figures 6(g) and 6(h)). These data indicated that the MECRG risk score shows good sensitivity and specificity for predicting the prognosis of CRC patients.
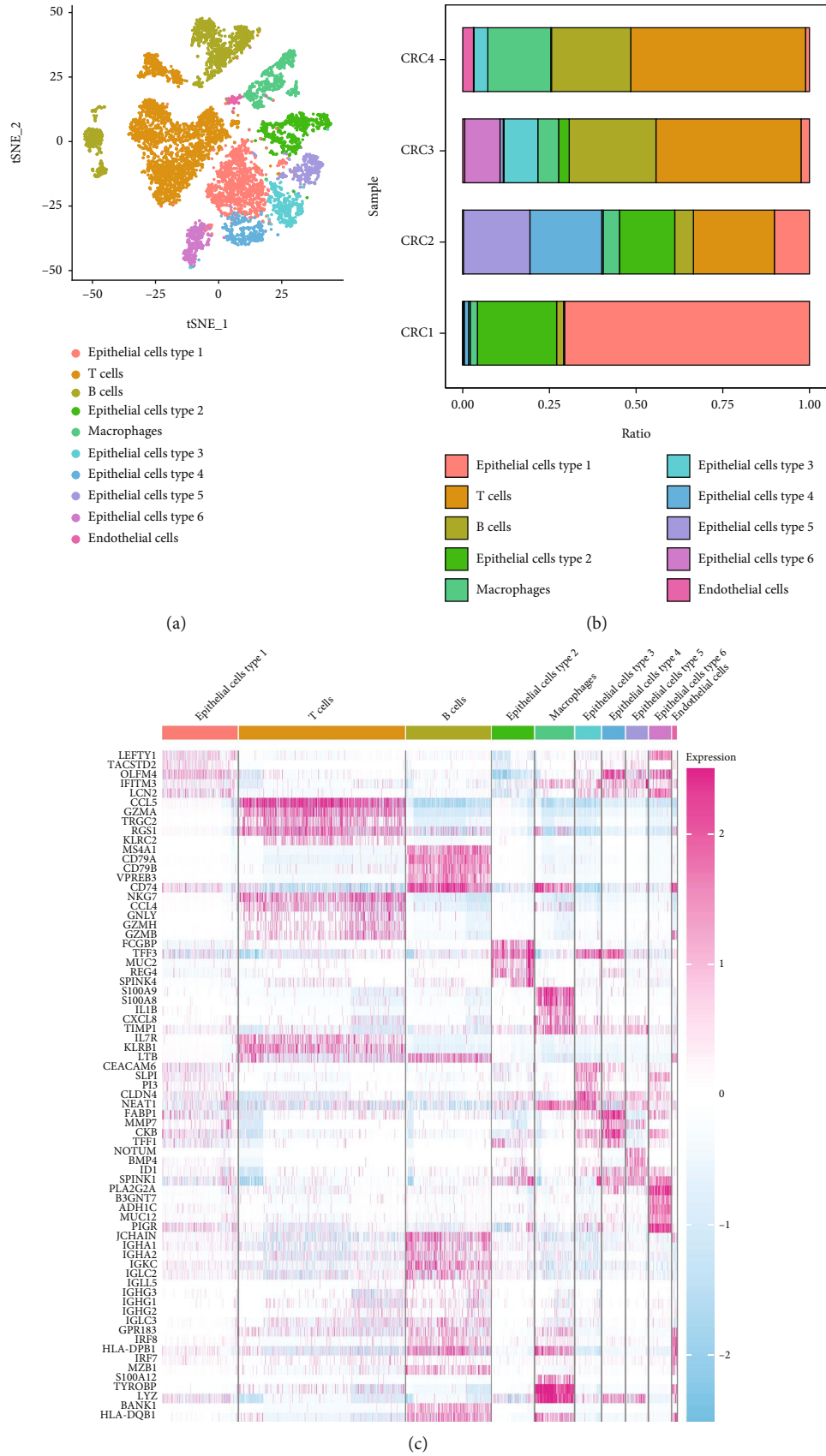
(a)

(b)

(c)

FIGURE 2: Characteristics of the cell subsets. (a) Annotated cell subsets. (b) Proportions of the various types of cells in the four CRC samples. (c) Heat map of the five most differentially expressed marker genes in each cell subset.
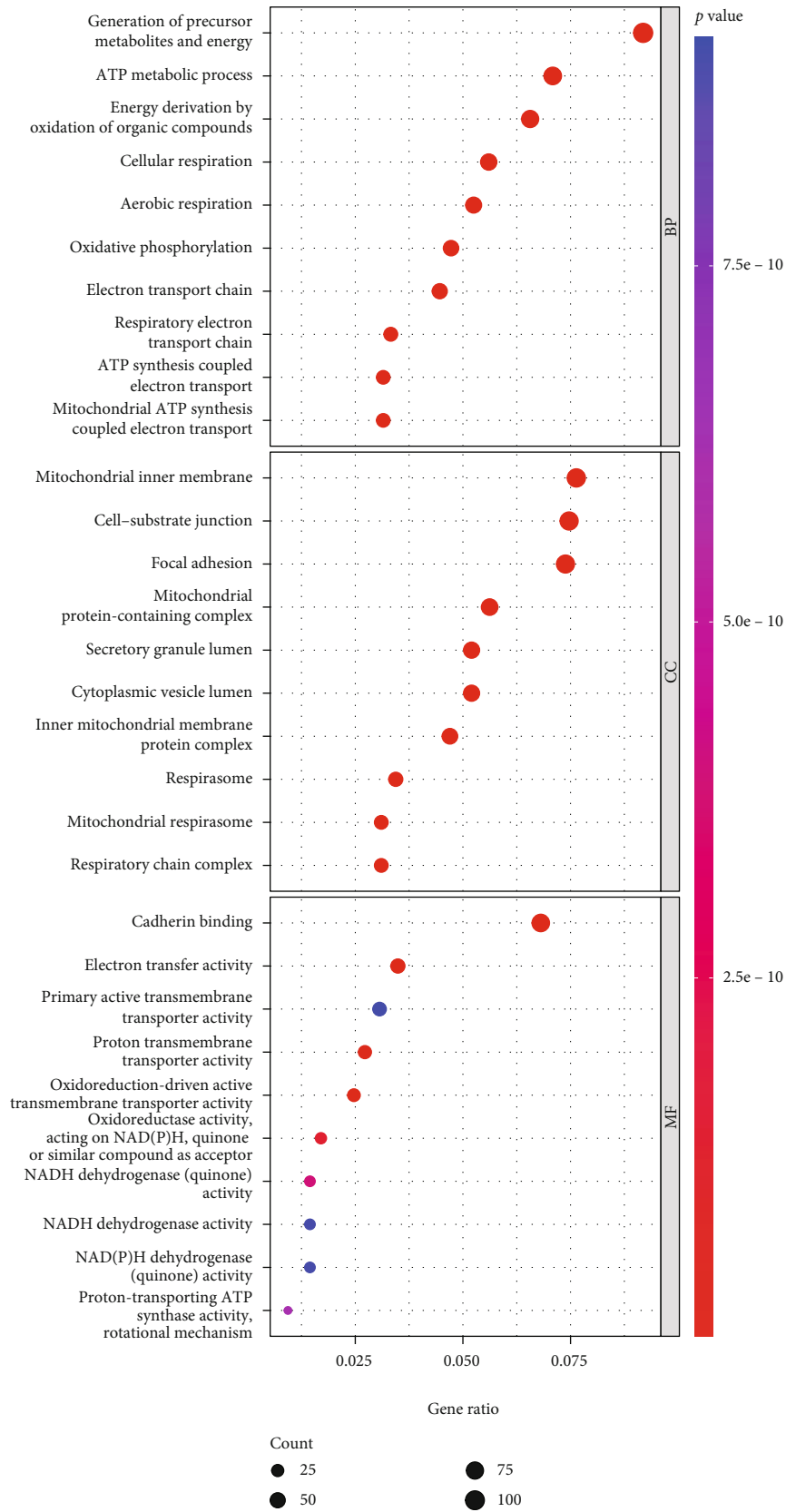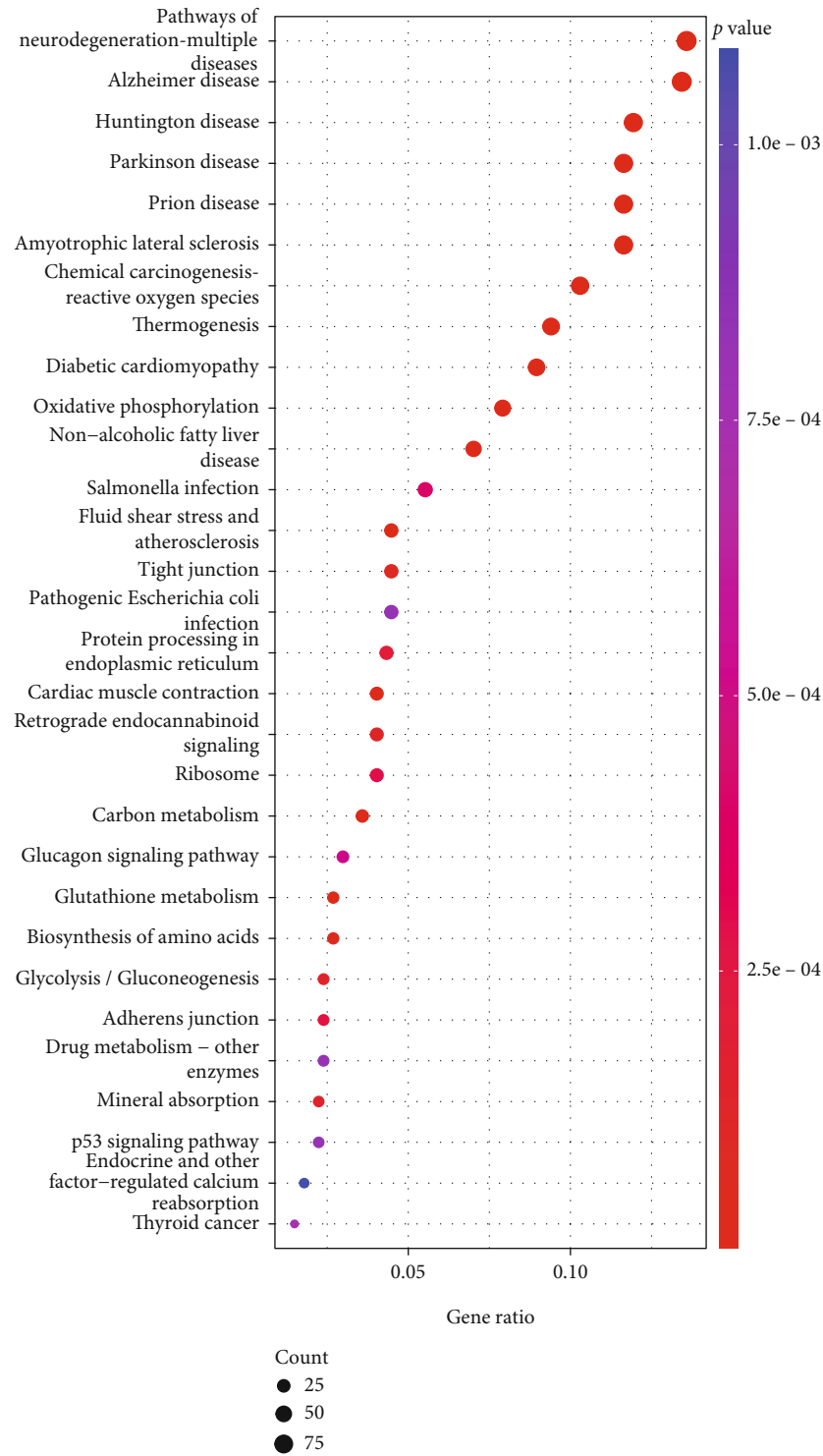
(a)

(b)

Figure 3: CNV analysis of epithelial cells from CRC patients. (a) Heat map of CNV in epithelial cells from the four CRC samples. (b) CNV score distribution among different epithelial cell subtypes.

(a)

Figure 4: Continued.

FIGURE 4: Functional enrichment analysis of marker genes in the MEC subsets. Results of GO function (a) and KEGG pathway (b) enrichment analyses.

3.6. Independent Prognostic Value of the MECRG Risk Score. To determine whether the MECRG risk score can predict prognosis independently of traditional clinical features, such as age, sex, and TNM stage, we performed univariate and multivariate Cox regression analyses. In the training cohort,

age (HR = 1.029 and 1.040, 95% confidence interval (CI): 1.010–1.048 and 1.021–1.060, respectively; $P = 0.002$ and $P < 0.001$), TNM stage (HR = 2.068 and 2.106, 95% CI: 1.628–2.627 and 1.642–2.701, respectively; $P < 0.001$), and the MECRG risk score (HR = 1.294 and 1.234, 95% CI:
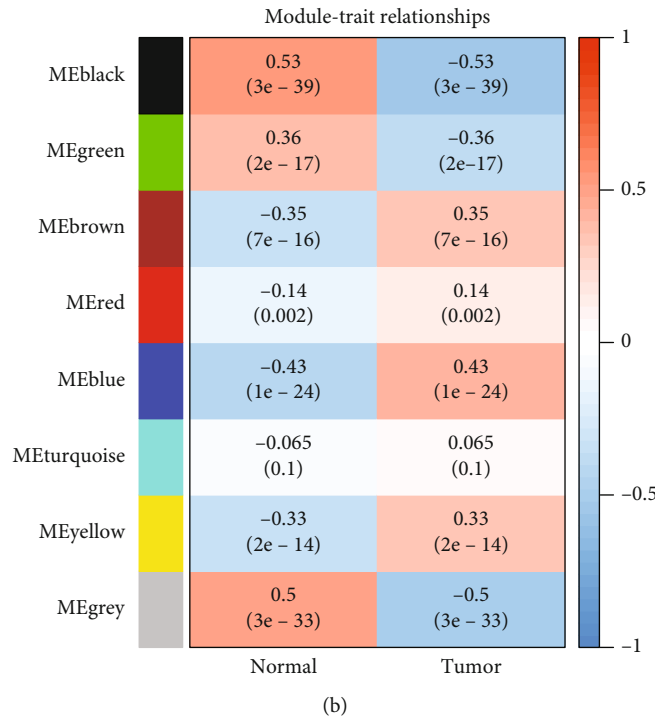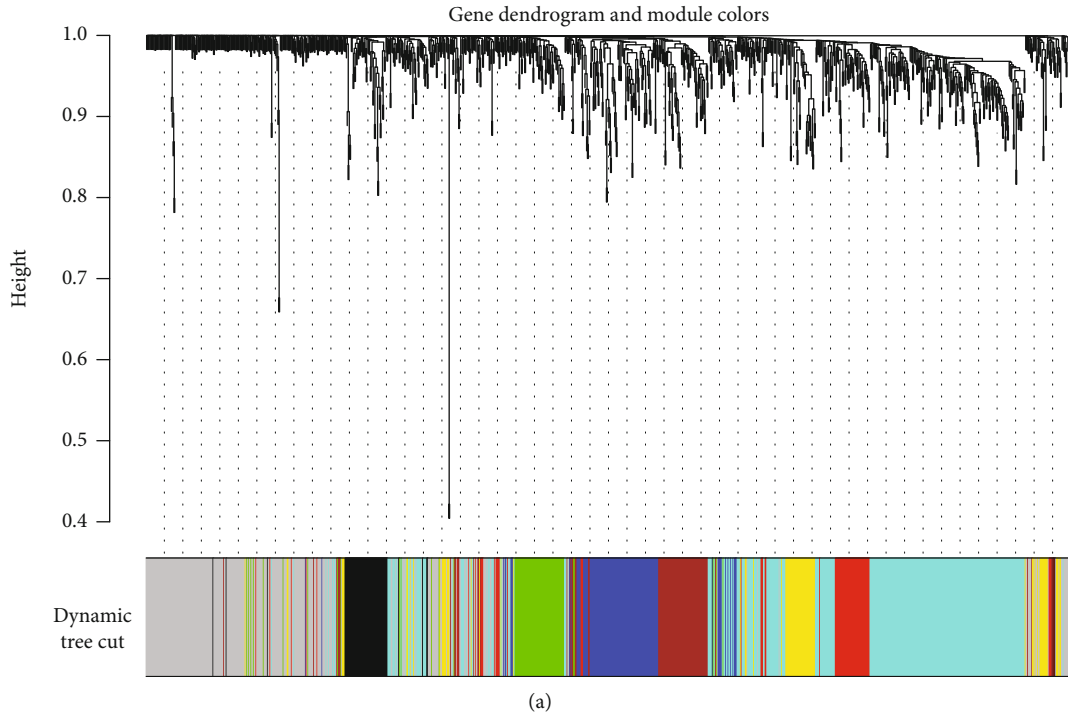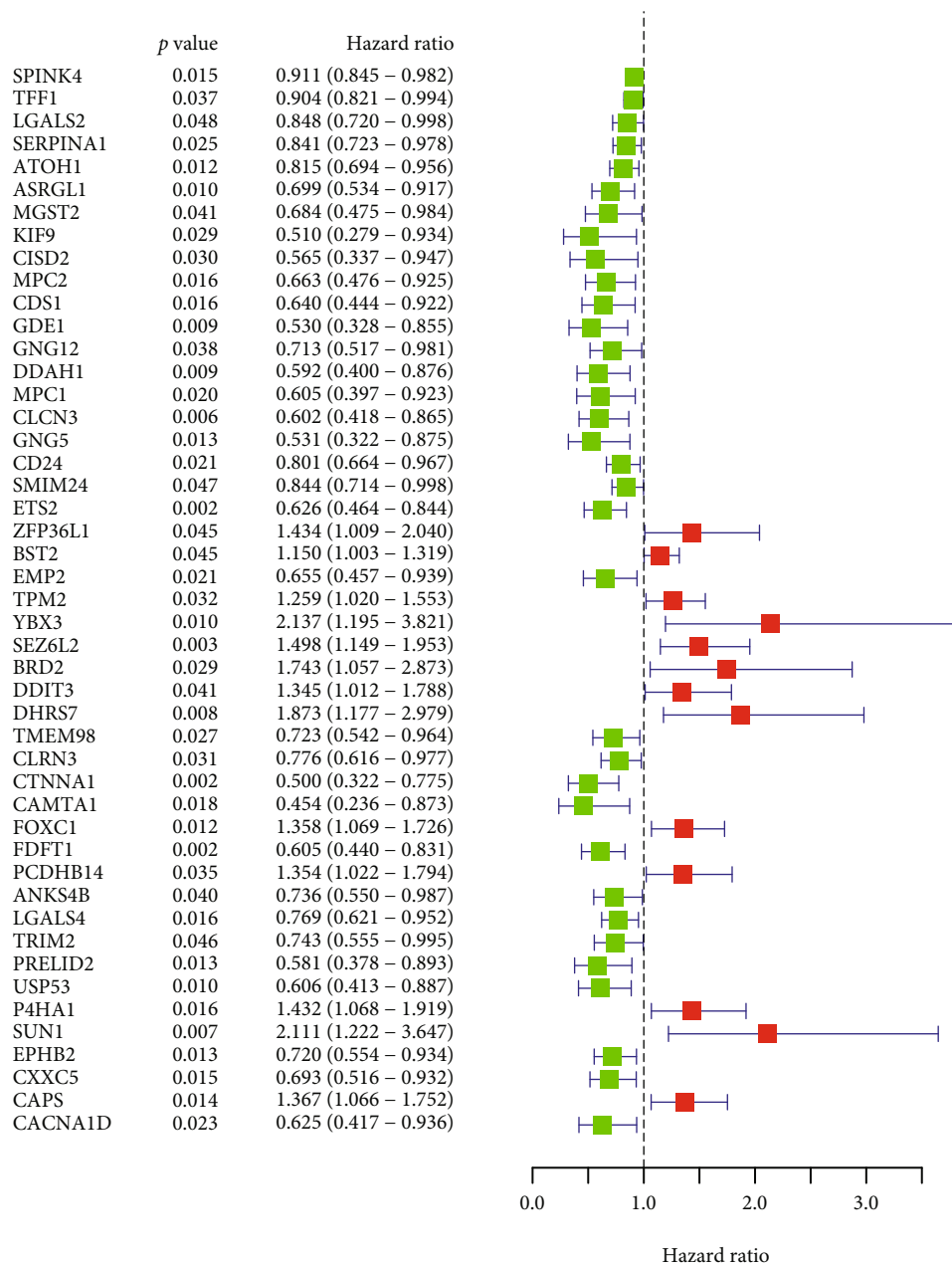
(a)

(b)

Figure 5: MEC marker genes associated with CRC. (a) Dendrogram of MEC subset marker genes obtained by WGCNA according to colour. (b) Correlations between characteristic genes of different modules and CRC.

1.200–1.396 and 1.139–1.336, respectively; $P < 0.001$) were independent predictors of OS. Analysis of the GSE17538 validation cohort confirmed that TNM stage (HR = 2.712 and 3.027, 95% CI: 2.077–3.541 and 2.323–3.945, respectively; $P < 0.001$) and the MECRG risk score (HR = 1.138 and 1.298, 95% CI: 1.005–1.289 and 1.127–1.496, respectively; $P = 0.042$ and $P < 0.001$) were independent predictors of OS (Figure 7). The above results indicated that the MECRG
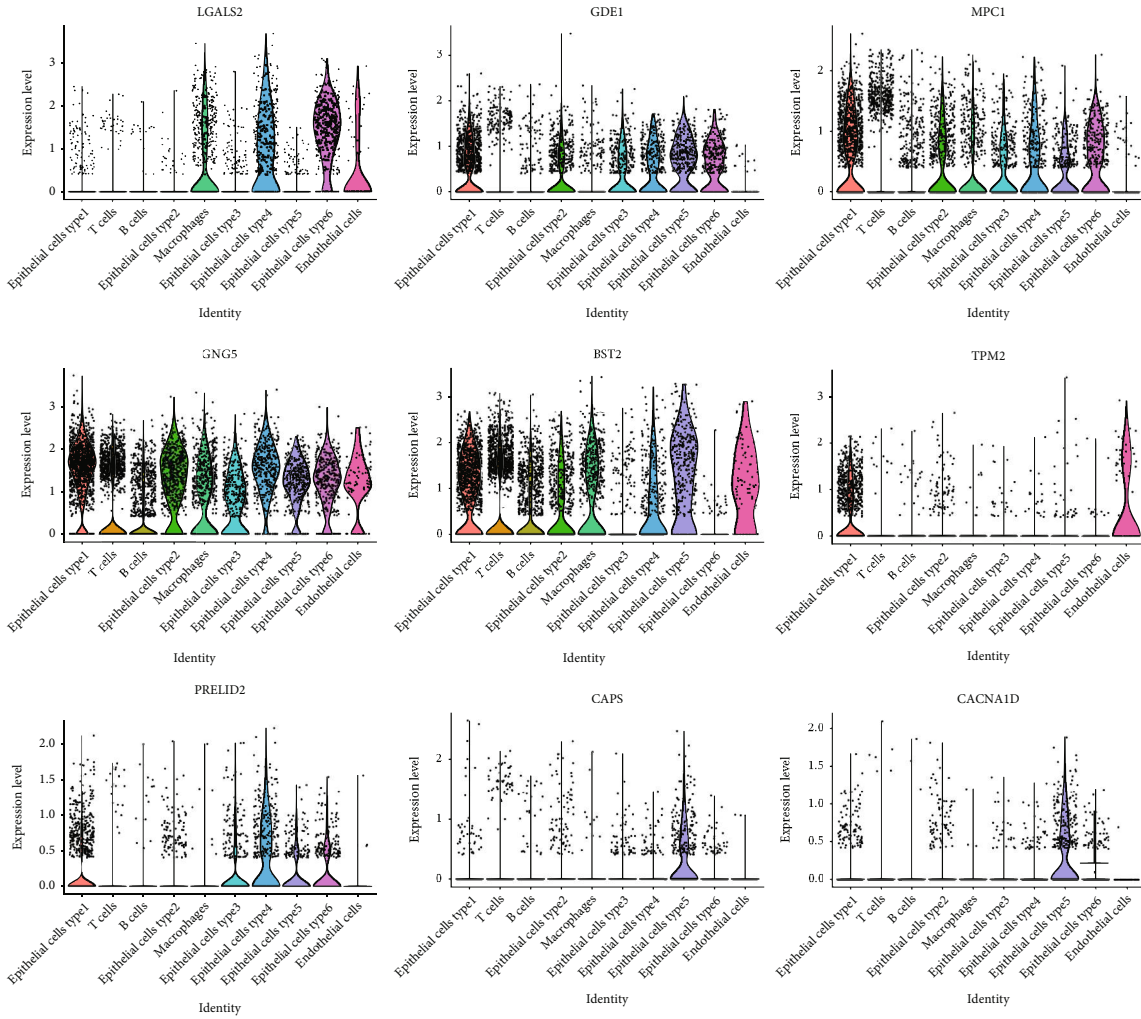
risk score was an independent prognostic factor in both the training and validation sets.

3.7. Correlation between MECRG Risk Groups and Clinical Features. We evaluated the correlation between clinical traits and risk groups among the 435 patients in the TCGA-CRC training cohort and the 232 CRC patients in the GSE17538 validation cohort (with complete clinical information). The

	p value	Hazard ratio
SPINK4	0.015	0.911 (0.845 − 0.982)
TFF1	0.037	0.904 (0.821 − 0.994)
LGALS2	0.048	0.848 (0.720 − 0.998)
SERPINA1	0.025	0.841 (0.723 − 0.978)
ATOH1	0.012	0.815 (0.694 − 0.956)
ASRGL1	0.010	0.699 (0.534 − 0.917)
MGST2	0.041	0.684 (0.475 − 0.984)
KIF9	0.029	0.510 (0.279 − 0.934)
CISD2	0.030	0.565 (0.337 − 0.947)
MPC2	0.016	0.663 (0.476 − 0.925)
CDS1	0.016	0.640 (0.444 − 0.922)
GDE1	0.009	0.530 (0.328 − 0.855)
GNG12	0.038	0.713 (0.517 − 0.981)
DDAH1	0.009	0.592 (0.400 − 0.876)
MPC1	0.020	0.605 (0.397 − 0.923)
CLCN3	0.006	0.602 (0.418 − 0.865)
GNG5	0.013	0.531 (0.322 − 0.875)
CD24	0.021	0.801 (0.664 − 0.967)
SMIM24	0.047	0.844 (0.714 − 0.998)
ETS2	0.002	0.626 (0.464 − 0.844)
ZFP36L1	0.045	1.434 (1.009 − 2.040)
BST2	0.045	1.150 (1.003 − 1.319)
EMP2	0.021	0.655 (0.457 − 0.939)
TPM2	0.032	1.259 (1.020 − 1.553)
YBX3	0.010	2.137 (1.195 − 3.821)
SEZ6L2	0.003	1.498 (1.149 − 1.953)
BRD2	0.029	1.743 (1.057 − 2.873)
DDIT3	0.041	1.345 (1.012 − 1.788)
DHRS7	0.008	1.873 (1.177 − 2.979)
TMEM98	0.027	0.723 (0.542 − 0.964)
CLRN3	0.031	0.776 (0.616 − 0.977)
CTNNA1	0.002	0.500 (0.322 − 0.775)
CAMTA1	0.018	0.454 (0.236 − 0.873)
FOXC1	0.012	1.358 (1.069 − 1.726)
FDFT1	0.002	0.605 (0.440 − 0.831)
PCDHB14	0.035	1.354 (1.022 − 1.794)
ANKS4B	0.040	0.736 (0.550 − 0.987)
LGALS4	0.016	0.769 (0.621 − 0.952)
TRIM2	0.046	0.743 (0.555 − 0.995)
PRELID2	0.013	0.581 (0.378 − 0.893)
USP53	0.010	0.606 (0.413 − 0.887)
P4HA1	0.016	1.432 (1.068 − 1.919)
SUN1	0.007	2.111 (1.222 − 3.647)
EPHB2	0.013	0.720 (0.554 − 0.934)
CXXC5	0.015	0.693 (0.516 − 0.932)
CAPS	0.014	1.367 (1.066 − 1.752)
CACNA1D	0.023	0.625 (0.417 − 0.936)

Hazard ratio

(a)

Figure 6: Continued.
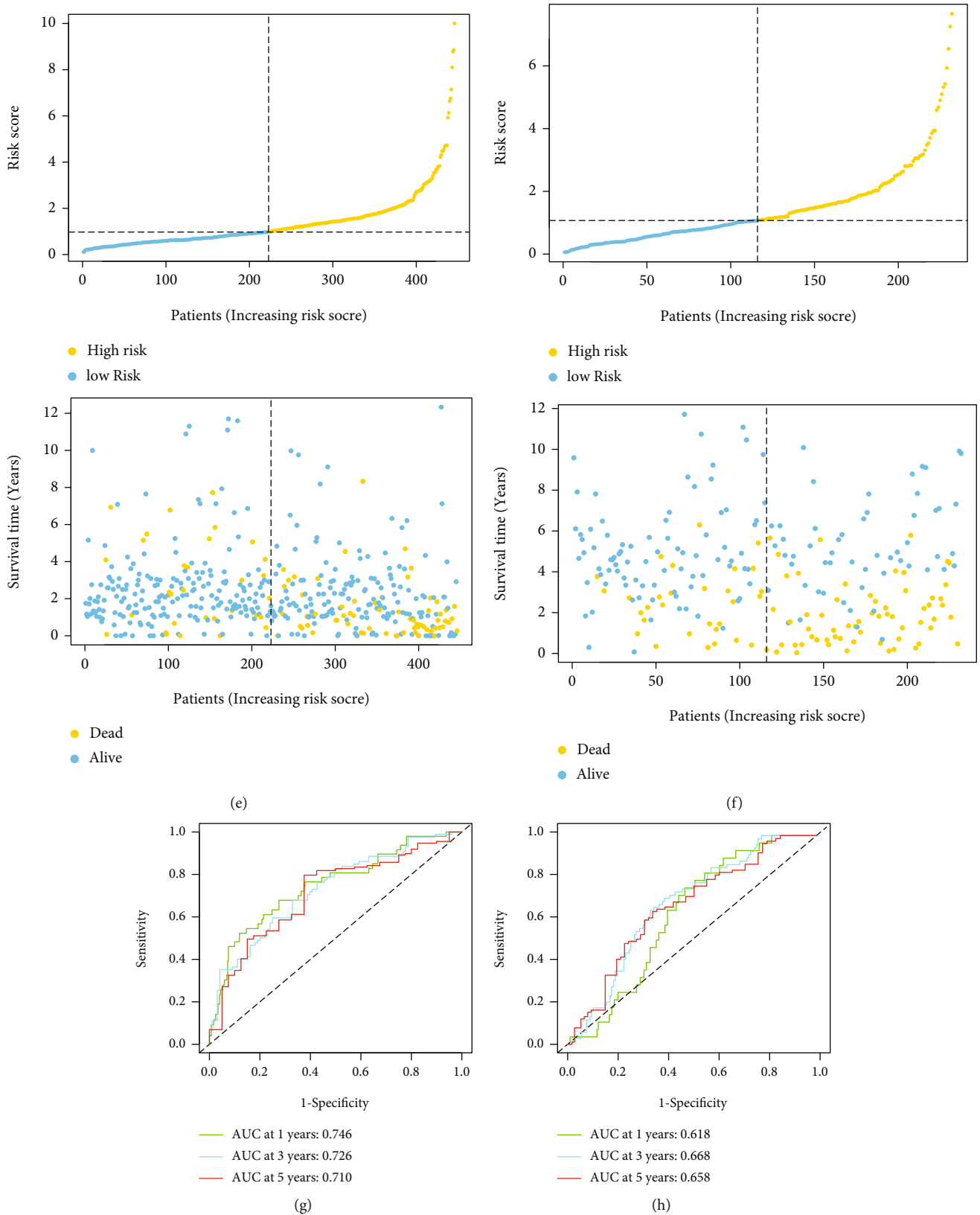
(b)

Figure 6: Continued.
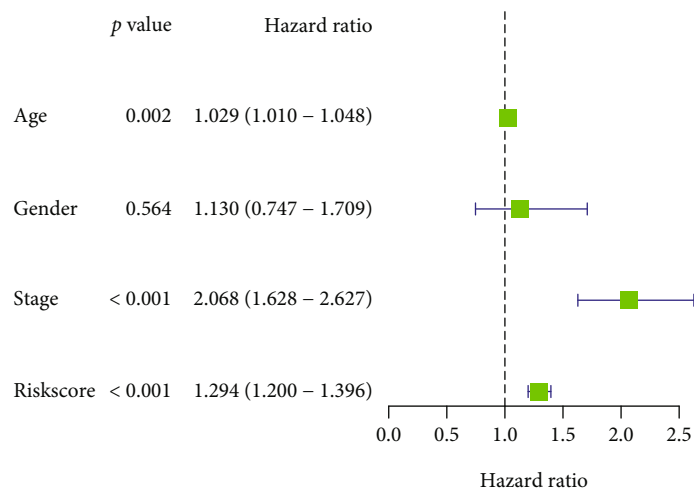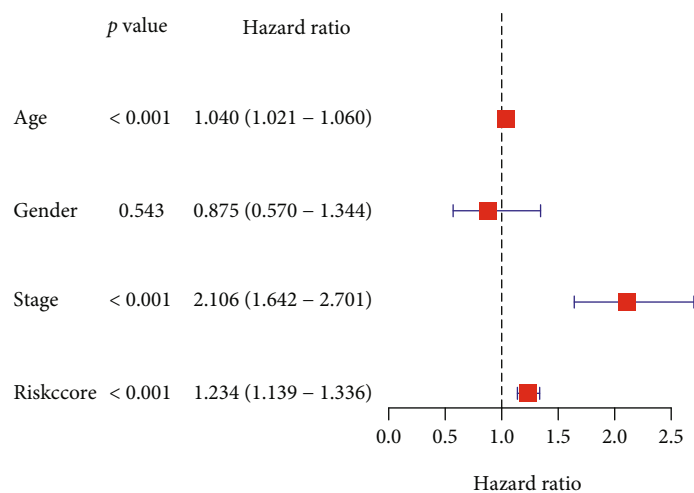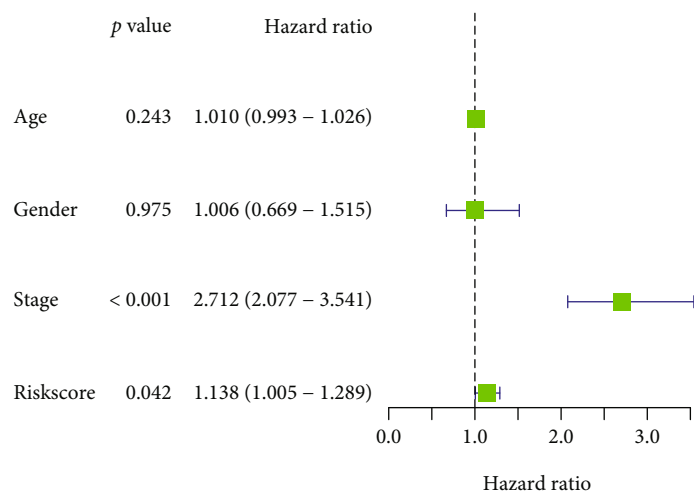
(c)

(d)
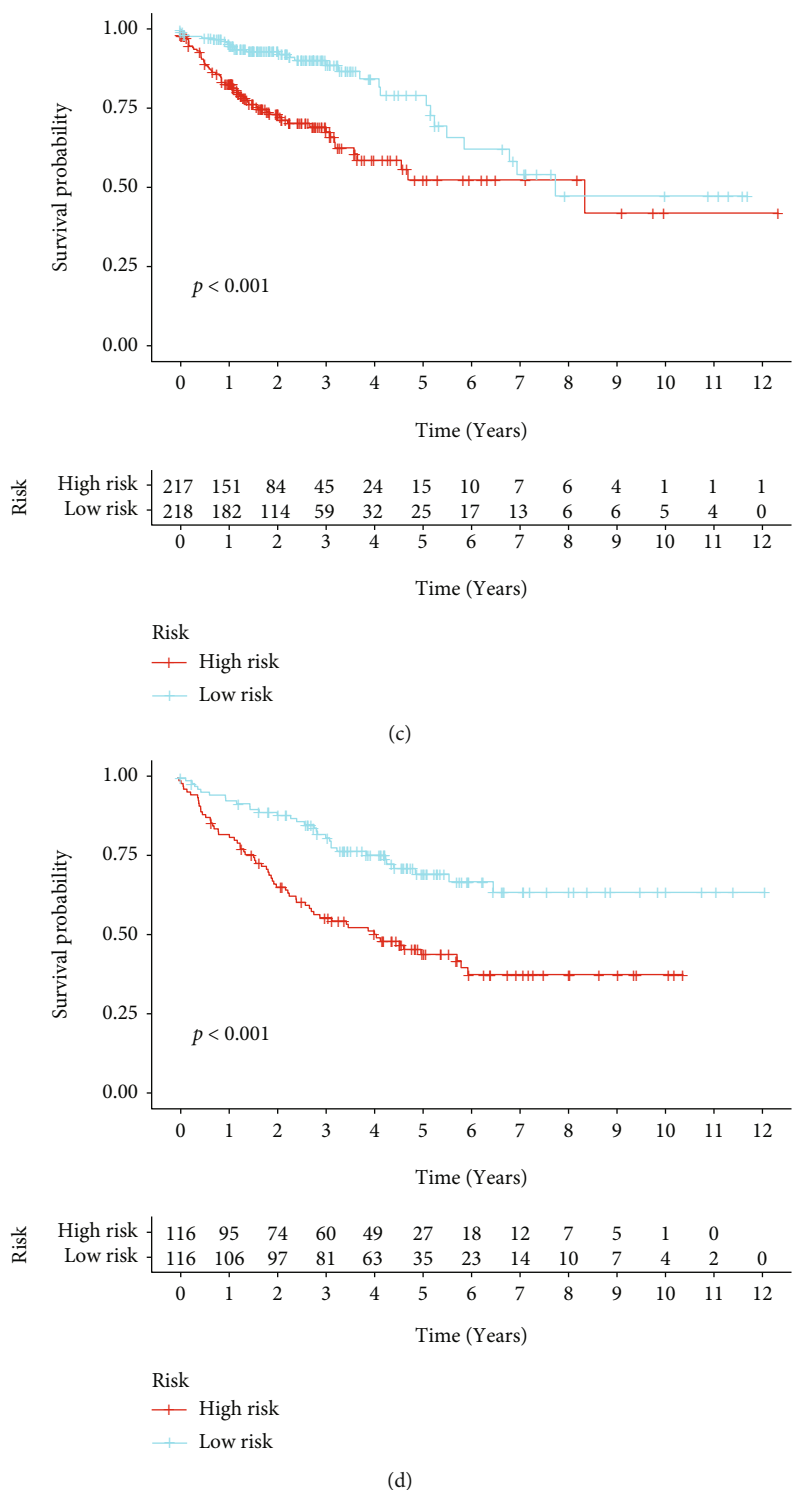
Figure 6: Continued.

Figure 6: Identification and validation of MECRGs in the TCGA-CRC training and GSE17538 validation cohorts. (a) Univariate Cox analysis was used to screen MECRGs with prognostic significance. (b) Distribution of MECRGs in cell subsets. (c and d) Kaplan–Meier survival curve showing the prognostic value of the MECRG signature in the training cohort (c) and the validation cohort (d). (e and f) Distribution of the MECRG risk scores and survival status of CRC patients in the training cohort (e) and validation cohort (f). ROC curve representing the efficiency of the MECRG signature in predicting 1-, 3-, and 5-year OS in CRC patients in the training cohort (g) and validation cohort (h).
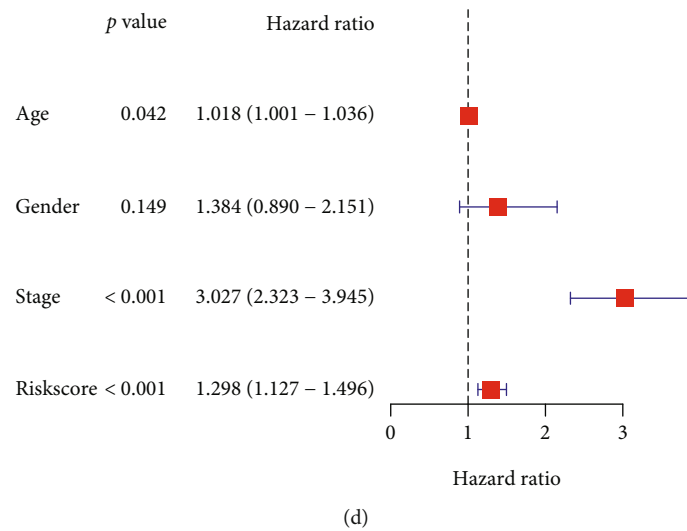
	p value	Hazard ratio
Age	0.002	1.029 (1.010 − 1.048)
Gender	0.564	1.130 (0.747 − 1.709)
Stage	< 0.001	2.068 (1.628 − 2.627)
Riskscore	< 0.001	1.294 (1.200 − 1.396)

Hazard ratio

(a)

	p value	Hazard ratio
Age	< 0.001	1.040 (1.021 − 1.060)
Gender	0.543	0.875 (0.570 − 1.344)
Stage	< 0.001	2.106 (1.642 − 2.701)
Riskccore	< 0.001	1.234 (1.139 − 1.336)

Hazard ratio

(b)

	p value	Hazard ratio
Age	0.243	1.010 (0.993 − 1.026)
Gender	0.975	1.006 (0.669 − 1.515)
Stage	< 0.001	2.712 (2.077 − 3.541)
Riskscore	0.042	1.138 (1.005 − 1.289)

Hazard ratio

(c)

Figure 7: Continued.

	p value	Hazard ratio	
Age	0.042	1.018 (1.001 − 1.036)	
Gender	0.149	1.384 (0.890 − 2.151)	
Stage	< 0.001	3.027 (2.323 − 3.945)	
Riskscore	< 0.001	1.298 (1.127 − 1.496)	

Hazard ratio

(d)

FIGURE 7: Independent prognostic value of the MECRG signature. Univariate and multivariate Cox analyses of the MECRG risk score in the TCGA-CRC training cohort (a and b) and GSE17538 validation cohort (c and d).

patients were grouped according to sex (male or female), age (≤65 or >65 years), and TNM stage (I–II or III–IV), and analyses were performed using the log-rank test and Kaplan–Meier analysis. We found that among female and male patients ($P < 0.05$), patients aged >65 years and ≤65 years ($P < 0.05$), and patients at TNM stage III–IV ($P \leq 0.001$), those in the high-risk group had a shorter survival time than those in the low-risk group (Figures 8(a) and 8(b)). A heat map revealed significant differences in tumour staging between the high- and low-risk groups (TCGA-CRC: $P = 0.024$; GSE17538: $P = 0.02$) (Figures 8(c) and 8(d)). These results suggested that the MECRG risk score can significantly affect the survival prognosis of CRC patients across gender, age, and stage.

3.8. Functional Enrichment Analysis of the MECRG Signature. We used GSEA to evaluate the functions of and signalling pathways associated with the genes in the high- and low-risk groups (Figure 9(a)). The results showed that the high-risk group was significantly enriched in functions related to the positive regulation of migration involved in sprouting angiogenesis, the phosphoinositide 3-kinase (PI3K)-AKT-mammalian target of rapamycin (mTOR) signalling pathway, basal cell carcinoma, extracellular matrix (ECM) receptor interaction, and angiogenesis, whereas the low-risk group was significantly enriched in apoptosis and the p53 signalling pathway. Additionally, enrichment analysis of the epithelial cell subsets indicated significant enrichment of activities involving collagen-containing ECM, fatty acid metabolism, basal cell carcinoma, PI3K-AKT signalling, and p53 signalling (Figure 9(b)). We found that the above signalling pathways are more likely to mediate the process of tumor progression in CRC epithelial cells than in other cell subpopulations.

3.9. Nomogram for Predicting the Prognosis of CRC Patients. To establish a practical method for predicting the probability of CRC patient survival, we constructed a MECRG nomo-

gram using the TCGA-CRC training cohort to predict OS at 1, 3, and 5 years (Figure 10(a)). The predictors included three clinical features (age, sex, and TNM stage) and the risk score. ROC curve analysis of nomogram reliability revealed AUC values at 1, 3, and 5 years of 0.787, 0.806, and 0.777, respectively (Figure 10(b)). Analysis of the GSE17538 validation cohort yielded AUC values of 0.843, 0.812, and 0.827, respectively (Figure 10(c)). The calibration curve showed that the predicted survival rate agreed well with the actual survival rate (Figures 10(d) and 10(e)). Our results demonstrated that the MECRG nomogram constructed in combination with the clinical characteristics of CRC can provide good prediction.

4. Discussion

Epithelial adenocarcinoma represents the most prevalent type of CRC and arises from benign colon adenoma [33]. Studies have shown that in epithelial malignant tumours, unstable adhesion between epithelial cells strongly correlates with an increased invasiveness of tumour cells. Recently, numerous mechanisms underlying cell–cell junctions regulated by E-cadherin expression in epithelial cells have been discovered [34–36]. Apical-basal polarity is the main characteristic of epithelial cells. Epithelial cell-associated polarity proteins are associated with the origin and poor prognosis of colorectal tumours, hepatocellular carcinoma, and endometrial cancer [37–39]. Therapeutic targets based on epithelial cell apical-basal polarity complexes have been reported. For example, partitioning-defective 6 (Par6) is expected to be a therapeutic target for breast cancer [37], and atypical protein kinase C (aPKC) has been suggested as a possible therapeutic target for gastric cancer [40]. Based on the above findings, we believed that the discovery of CRC epithelial cell-related biomarkers would facilitate the development of new therapeutic and predictive targets for CRC prognosis. ScRNA-seq technology allows the sequencing of RNA from individual cells to systematically track the dynamic changes
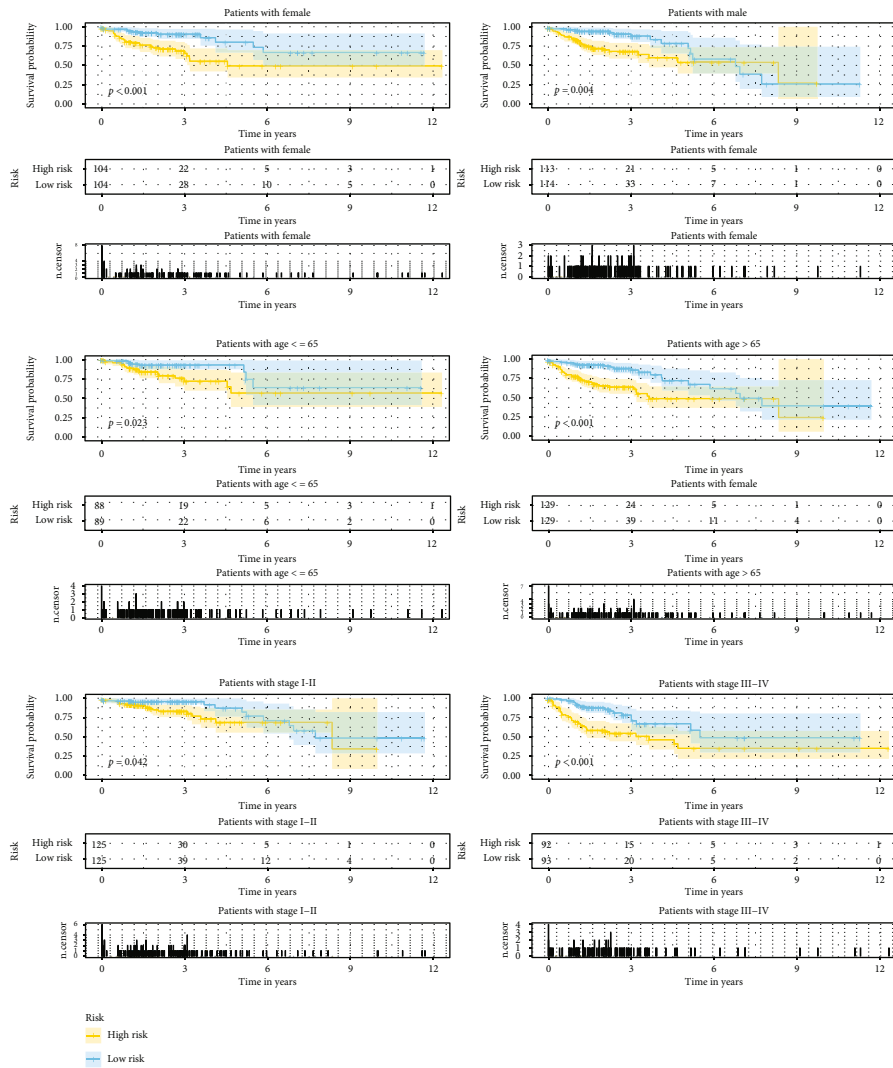
(a)

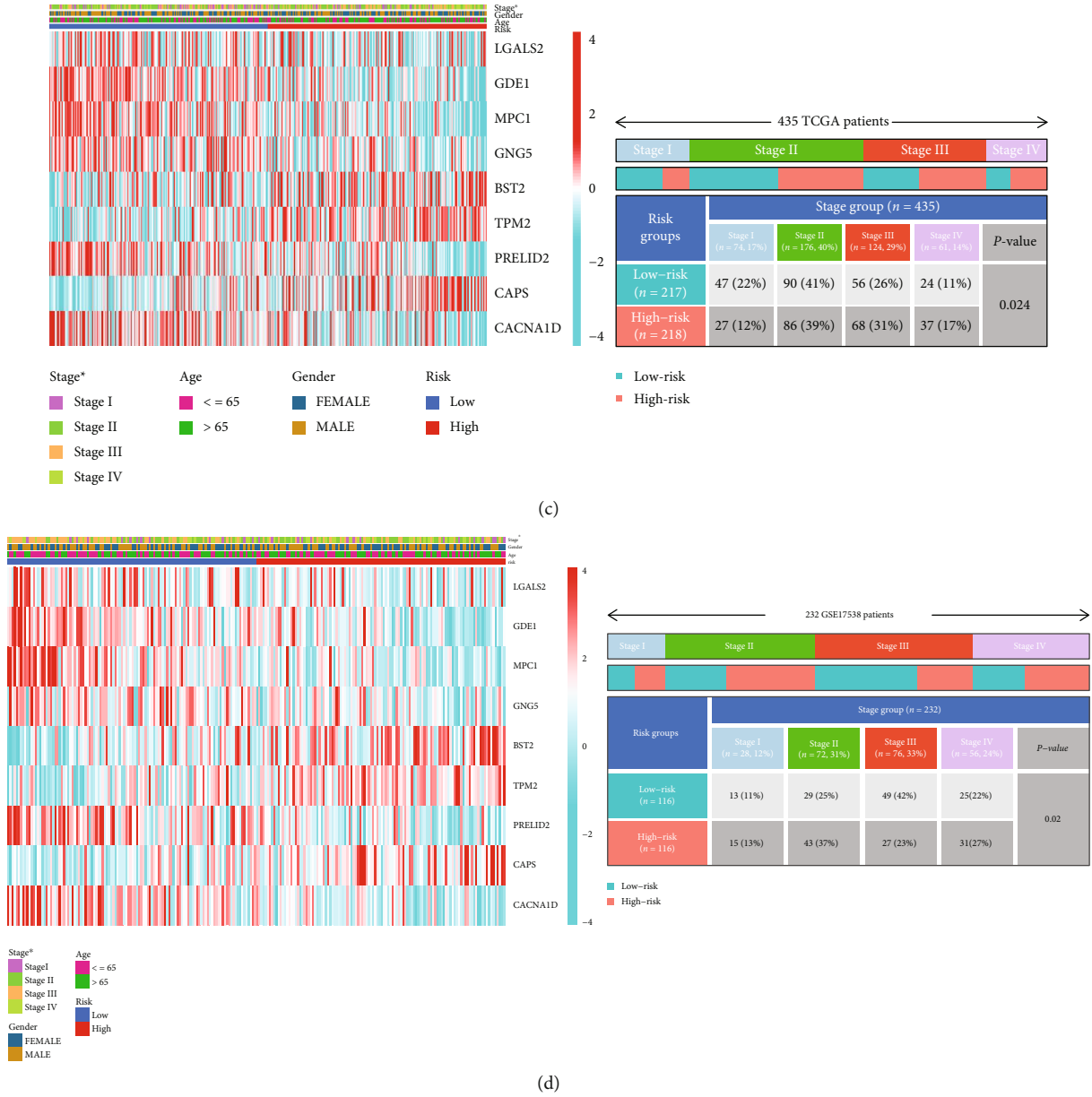FIGURE 8: Continued.
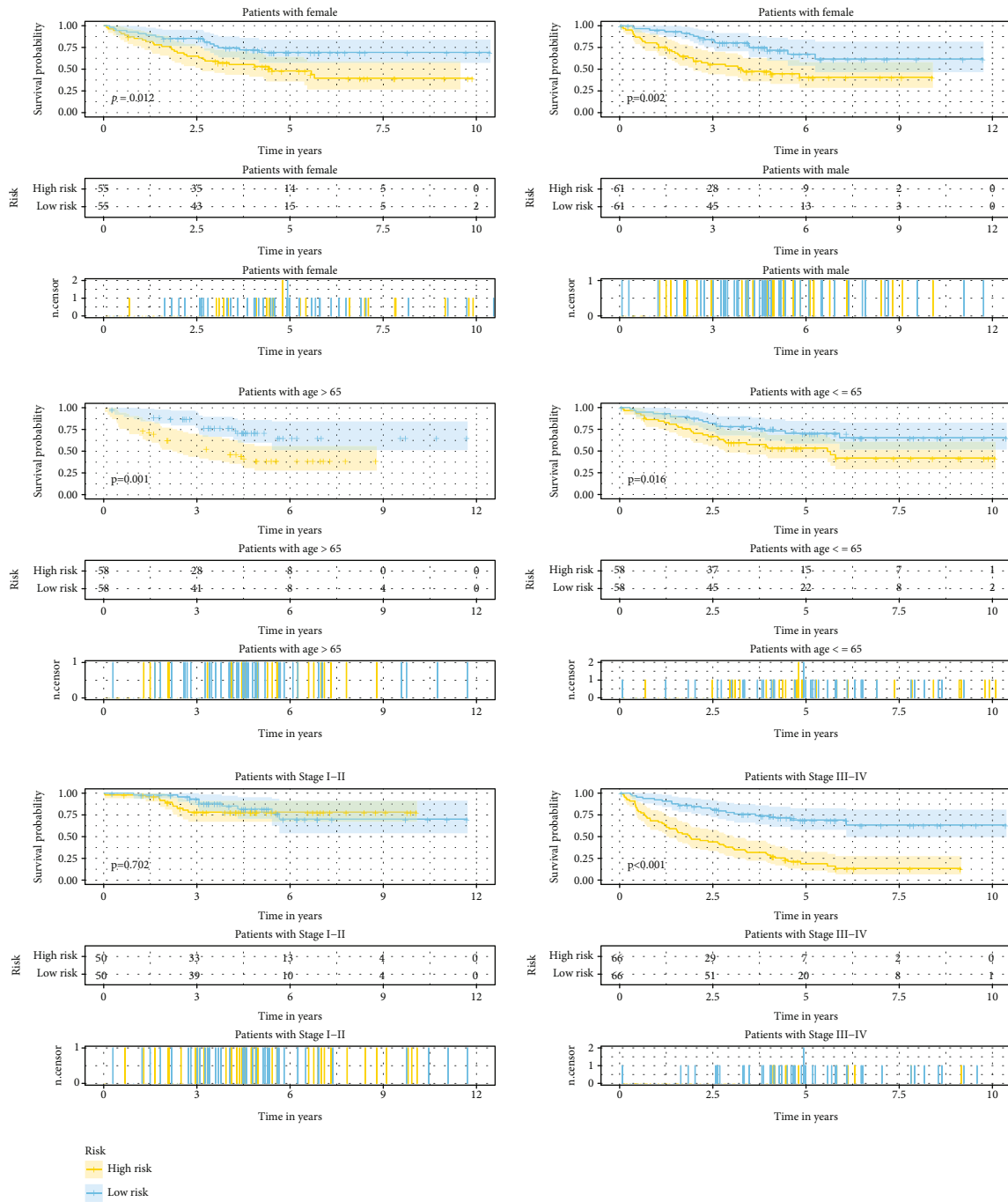
(b)

Figure 8: Continued.

(c)

(d)

FIGURE 8: Analysis of the relationships between the MECRG signature and clinical features using the TCGA-CRC training and GSE17538 validation cohorts. (a) Survival analysis of the MECRG signature in clinical features based on the training cohort. (b) Survival analysis of the MECRG signature in clinical features based on the validation cohort. (c and d) Heat maps showing the correlation between MECRG risk grouping and TNM stage in the training cohort (c) and validation cohort (d). ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

in individual cells and deepen the understanding of cellular states and gene expression regulation in pathological disease processes [41]. scRNA-seq analysis provides analytical detail on the cellular level [42]. The technology has been recently applied to study epithelial cells in serous epithelial ovarian cancer and endometrial cancer [43, 44]. In nontumour research, scRNA-seq has been used for gene expression profiling of breast epithelial cells during their development [45]. Xu et al. revealed the biological mechanisms implicated in the involvement of lung epithelial cells in idiopathic fibrosis using scRNA-seq [46]. Additionally, scRNA-seq technology has received increasing attention for predicting the prognosis of cancer patients. For example, Wang et al. constructed a

model based on 10 biomarkers of pancreatic ductal epithelial cells to predict the prognosis of patients with pancreatic adenocarcinoma [47], Zheng et al. [48] screened nine fibroblast marker genes from scRNA-seq CRC data as potential prognostic markers, and Li et al. [49] identified seven macrophage marker genes from breast cancer scRNA-seq data as promising diagnostic and prognostic biomarkers.

In our study, we constructed a CRC prognosis model by analysing scRNA-seq data from CRC epithelial cells. Specifically, we analysed scRNA-seq data of 15,465 cells from four CRC patients, followed by PCA for unsupervised clustering and the identification of MEC subsets based on chromosome CNV analysis. We then used WGCNA, univariate Cox, and
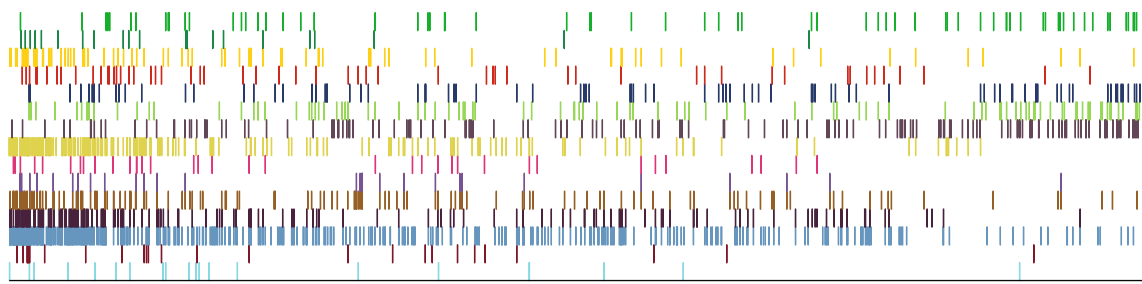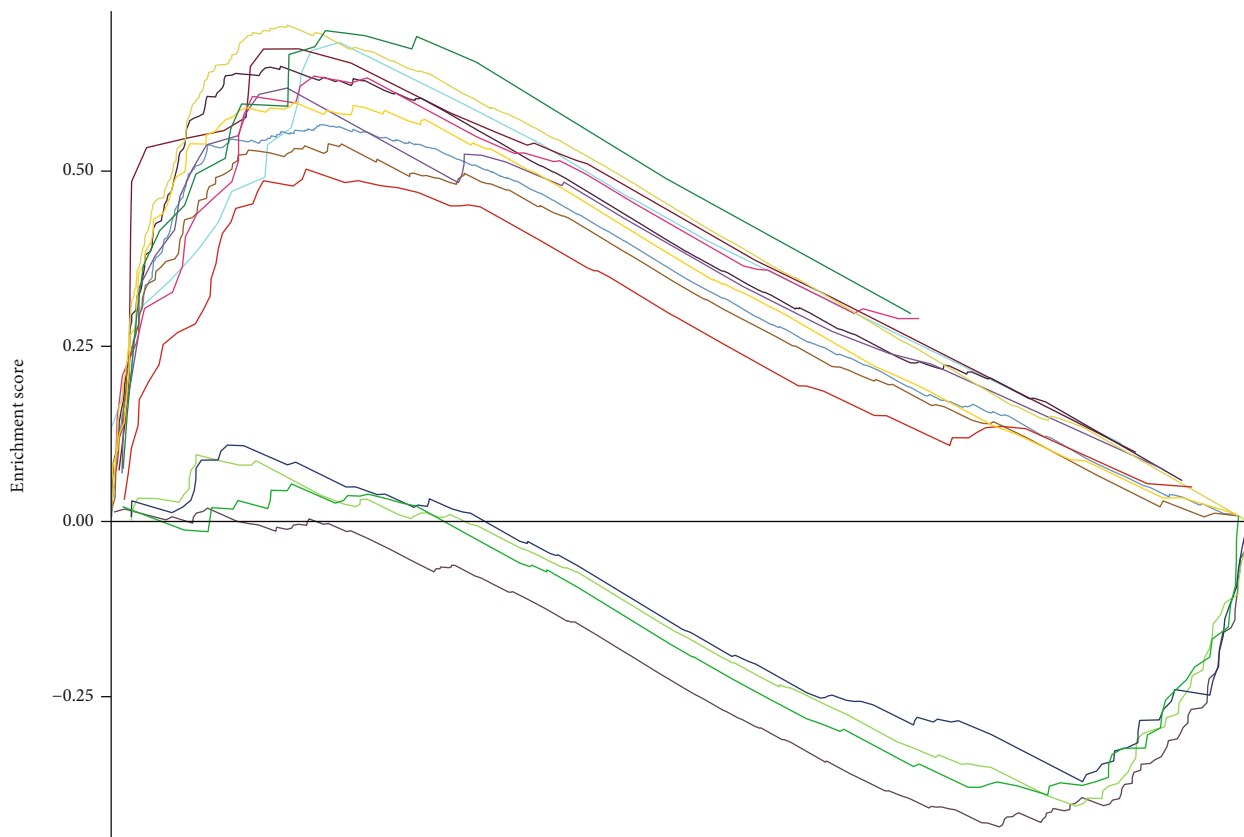
GOBP_POSITIVE_REGULATION_OF_CELL_MIGRATION_INVOLVED_IN_SPROUTING_ANGIOGENESIS
GOBP_REGULATION_OF_VASCULAR_ENDOTHELIAL_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY
GOCC_COLLAGEN_CONTAINING_EXTRACELLULAR_MATRIX
GOMF_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT
GOMF_GROWTH_FACTOR_BINDING
GOMF_INSULIN_LIKE_GROWTH_FACTOR_BINDING
HALLMARK_ANGIOGENESIS
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_PI3K_AKT_MTOR_SIGNALING
KEGG_APOPTOSIS
KEGG_BASAL_CELL_CARCINOMA
KEGG_ECM_RECEPTOR_INTERACTION
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE
KEGG_P53_SIGNALING_PATHWAY

(a)

Figure 9: Continued.
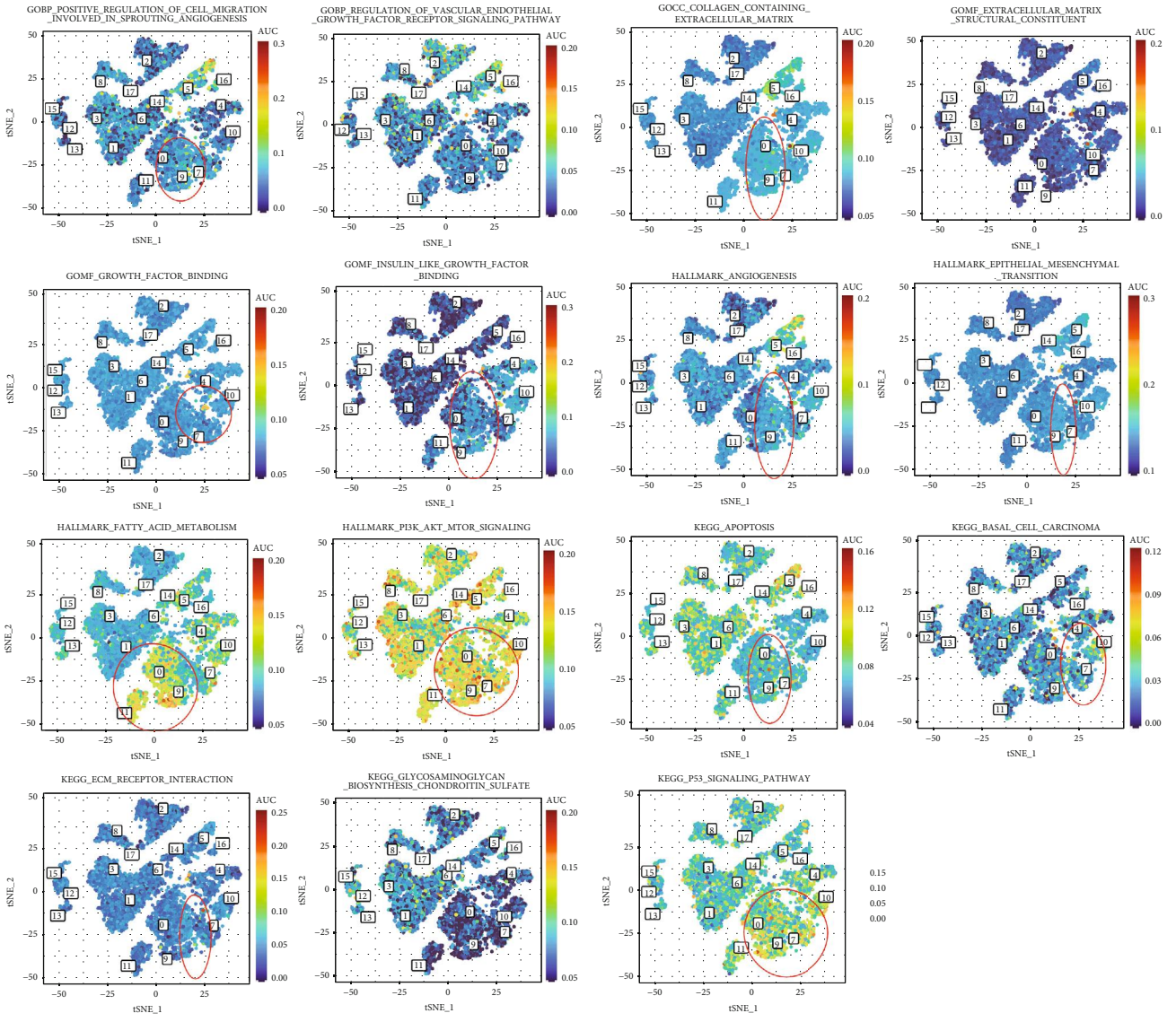
(b)

FIGURE 9: Functional enrichment analysis of the MECRG signature and distribution of enriched pathways in the cell subsets. (a) GSEA of the MECRG high- and low-risk groups. (b) Pathways enriched in the cell subsets of the MECRG risk groups.

LASSO regression analyses to construct a MECRG signature of nine genes related to CRC prognosis. We validated the MECRG signature using both training and validation cohorts. The Kaplan–Meier analysis confirmed a significantly shortened OS for CRC patients with a high MECRG risk score. Furthermore, the univariate and multivariate Cox regression analyses showed that the MECRG risk score may be an independent predictor of OS. Based on the MECRG risk score coupled with the clinical characteristics of sex, age, and TNM stage (III–IV), low-risk patients showed significantly longer survival times than high-risk patients. It is worth noting that there were significant differences between the MECRG signature-based risk groups at different CRC stages. Furthermore, the MECRG nomogram showed excellent prediction in both the training and validation sets, suggesting that it may efficiently predict CRC patient survival in the clinical setting.

Elevated expression of *LGALS2* reportedly inhibits the development of CRC and lymph node metastasis of gastric cancer [50, 51]. *GDE1* expression is significantly reduced in drug-resistant ovarian cancer samples [52]. Consistent with these findings, our results suggested that high expression of *LGALS2* and *GDE1* in CRC patients implies a good prognosis. However, the relationship between *GDE1* and CRC requires further study. Lysine demethylase 5A (KDM5A) regulates *MPC1* expression, and the KDM5A–MPC1 axis is involved in regulating the mesenchymal characteristics of cancer cells during EMT [53]. Schell et al. [54] found that loss of *MPC1* expression enhances the Warburg effect and promotes the proliferation of CRC cells. Deletion of *MPC1* is related to a poor prognosis in glioblastoma [55]. Furthermore, *BST2* activates the nuclear factor-κB-Snail-Raf kinase inhibitor protein axis to promote tumour invasion and metastasis via EMT [56, 57]. *BST2* overexpression correlates with poor prognosis in CRC, stomach cancer, and

(a)

AUC at 1 years: 0.787
AUC at 3 years: 0.806
AUC at 5 years: 0.777

(b)

AUC at 1 years: 0.843
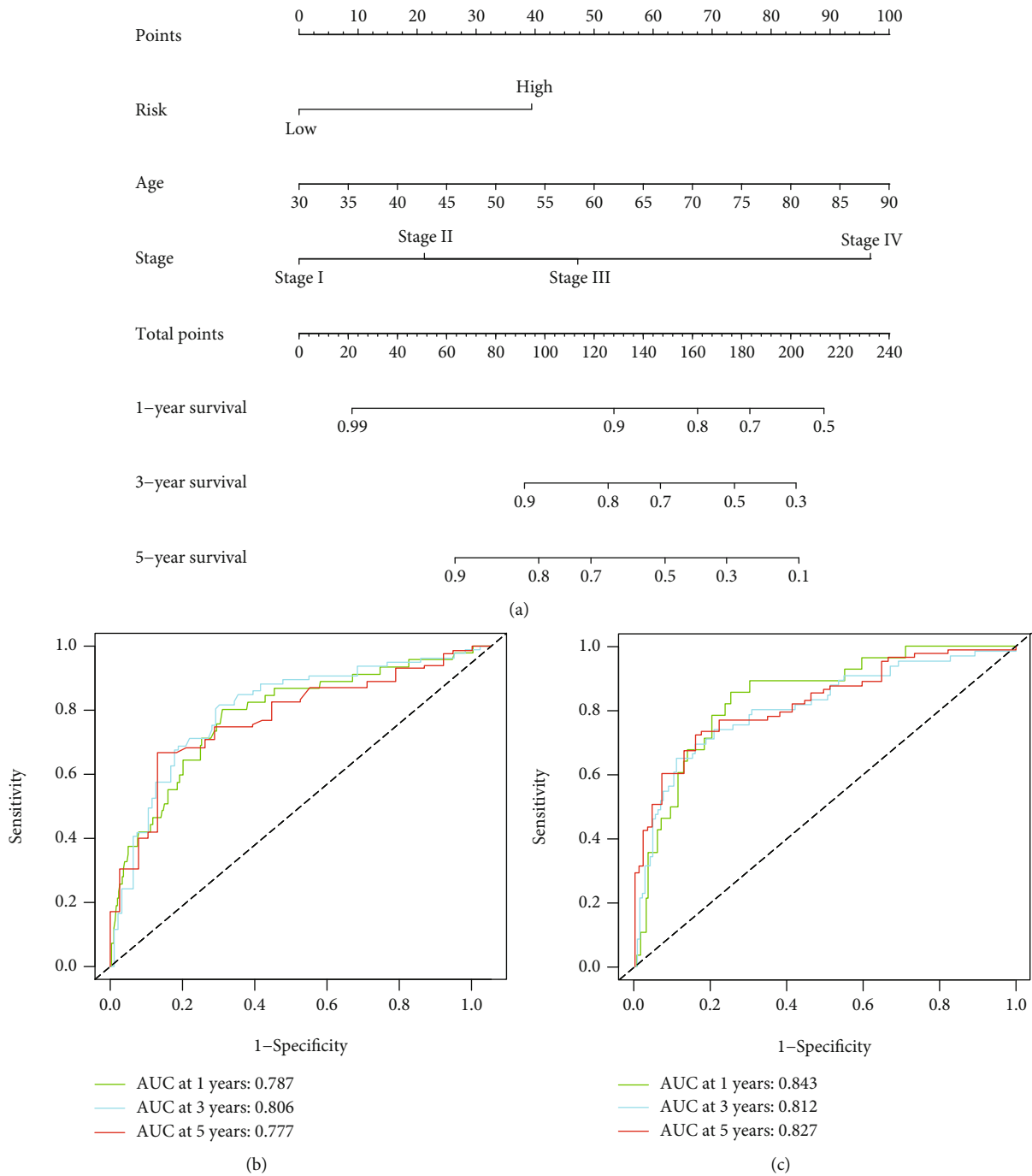AUC at 3 years: 0.812
AUC at 5 years: 0.827
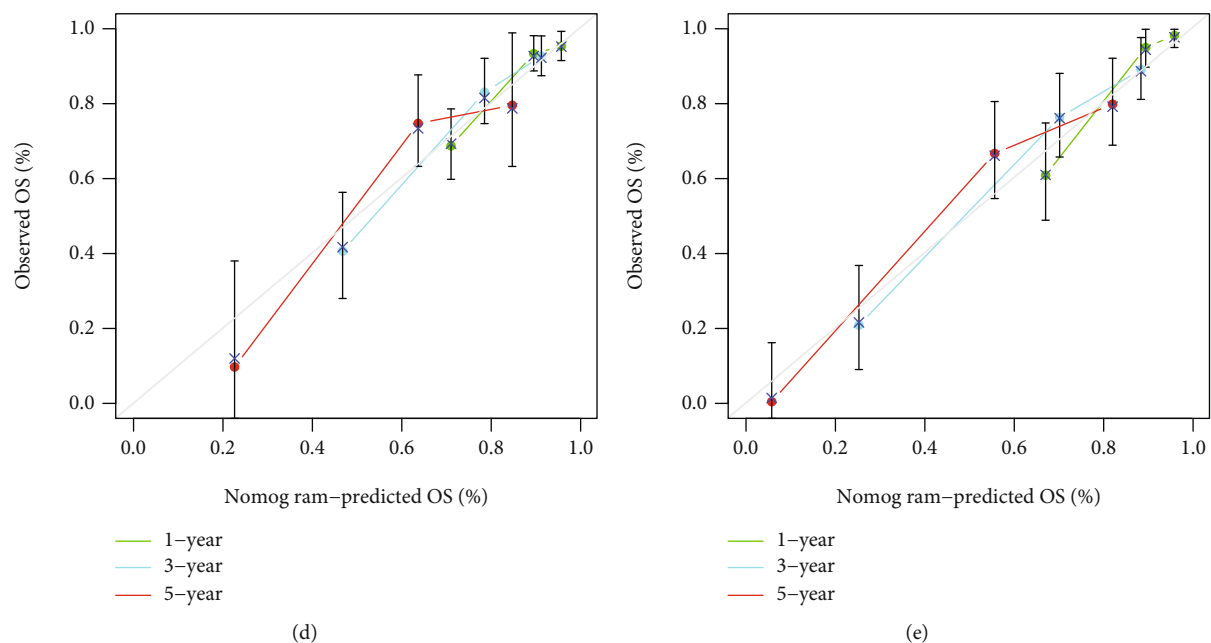
(c)

FIGURE 10: Continued.

FIGURE 10: Construction of a nomogram based on the MECRG signature. ROC curve analysis of the nomogram (a) using the TCGA-CRC training cohort (b) and the GSE17538 validation cohort (c). (d and e) Calibration curves for the nomogram using the training cohort (d) and the validation cohort (e).

oesophageal cancer [58]. *BST2* expression is specifically upregulated in oral squamous cell carcinoma and is responsible for drug resistance [59]. *BST2* is more highly expressed in breast cancer cells derived from patients presenting bone metastasis than in human primary breast cancer cells [60]. Furthermore, Mukai et al. [58] showed that *BST2* knockout *in vitro* inhibited the proliferation of gastric cancer cells. Together, these findings suggest that *MPC1* is a protective gene and *BST2* a risk gene in CRC, which is consistent with the results of the present study. *CAPS* is a calcium-binding protein related to cell proliferation and differentiation signals [61] and associated with poor prognosis in CRC and gliomas and drug resistance in breast cancer [62–64], which is consistent with our results. Abnormal expression of *TPM2* is involved in actin cytoskeleton remodelling during EMT of lens epithelial cells [65]. Shibata et al. [66] showed that downregulation of *TPM2* expression decreased EMT in injured mouse lens epithelium, resulting in delayed lens wound healing. Additionally, *TPM2* is upregulated in ovarian cancer, liver cancer, and breast cancer [67–69]. Zhou et al. [70] found that elevated *TPM2* expression in CRC patients was predictive of poor prognosis, which is in line with the findings of the present study. However, Ma et al. [71] showed that *TPM2* expression was downregulated in CRC; therefore, the precise mechanism of *TPM2* in CRC requires clarification. Patients with glioma exhibiting elevated *GNG5* expression have a shorter survival time [72], and patients with head and neck squamous cell carcinoma and elevated *PRELID2* expression have a poor prognosis [73]. Tan et al. [74] reported that excessive aldosterone secretion from aldosteronoma is related to a *CACNA1D* mutation. We identified *GNG5*, *PRELID2*, and *CACNA1D* as protective genes in CRC patients; however, further basic experimental studies of these three genes in CRC are needed. Additionally, we showed that *MPC1*, *BST2*, and *TPM2* are closely

related to EMT, suggesting that these molecules are potentially important EMT-related therapeutic targets.

GSEA suggested that genes related to collagen-containing ECM, fatty acid metabolism, PI3K-AKT signalling, p53 signalling, EMT, and other related pathways were enriched in epithelial cell subsets. Among these, EMT-related pathways were more significantly enriched in epithelial cell clusters 7, 9, and 10. Activation of the PI3K-AKT signalling pathway is a key feature of the EMT programme during tumour progression [75]. The expression of genes involved in fatty acid synthesis is upregulated in CRC epithelial cells, where the accumulation of polyunsaturated fatty acids contributes to the development of CRC [76]. Targeting fatty acid synthesis genes may become a new strategy for the treatment of CRC in the future. ECM remodelling can affect the signalling in the tumour microenvironment [77]. The P53 signalling pathway is a common tumour suppressor pathway [78], and we speculate that epithelial cells may promote CRC progression by interfering with this pathway. Thus, in the CRC microenvironment, epithelial cells may mediate the development of CRC through collagen-containing ECM, fatty acid metabolism, PI3K-AKT signalling, p53 signalling, EMT, and other pathological mechanisms.

This was a retrospective study that used scRNA-seq data and bulk data from public databases to construct a model for predicting survival in CRC patients. However, this study had some limitations. First, the mechanisms of *GDE1*, *PRELID2*, *GNG5*, and *CACNA1D* in CRC have not been clarified; therefore, the data suggesting their prognostic value need to be validated. Second, we did not evaluate tumour size, metastasis, surgery, postoperative chemoradiotherapy, and other prognostic factors in this study. This may have affected the predictive accuracy of the model. In future studies, we will include additional data to increase the accuracy of the model.

5. Conclusion

We evaluated epithelial cell marker genes with prognostic significance in CRC using scRNA-seq data and bulk data and generated a MECRG signature and risk score, which was confirmed to show independent prognostic efficacy for CRC patients. A nomogram based on the MECRG signature along with specific clinical features demonstrated accurate prediction of CRC patient survival, suggesting its potential utility for clinical application.

Data Availability

The datasets generated during and/or analysed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

Supplementary Materials

Supplementary 1. Figure S1: quality control of the scRNA-seq data from the four CRC samples. Figure S2: gene numbers in the four CRC samples are positively correlated with sequencing depth. Figure S3: PCA classification of the cells based on the scRNA-seq data. Figure S4: eleven statistically significant principal components as identified by PCA. Figure S5: dot diagram showing the expression of specific marker genes in the cell subsets. Figure S6: distribution of specific marker genes in the four cell subsets. Figure S7: LASSO regression was used to screen MECRGs with prognostic significance.

Supplementary 2. Table S1: clinical characteristics of the four CRC patients. Table S2: CNV scores of epithelial cell subclusters. Table S3: MEC subclusters of marker genes. Table S4: MECRG signature using the TCGA-CRC training cohort. Table S5: MECRG signature using the GSE17538 validation cohort.

References

[1] Y. N. Peng, M. L. Huang, and C. H. Kao, "Prevalence of depression and anxiety in colorectal cancer patients: a literature review," *International Journal Environmental Research and Public Health*, vol. 16, no. 3, p. 411, 2019.

[2] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[3] F. Islami, A. Goding Sauer, K. D. Miller et al., "Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 31–54, 2018.

[4] K. D. Miller, M. Fidler-Benaoudia, T. H. Keegan, H. S. Hipp, A. Jemal, and R. L. Siegel, "Cancer statistics for adolescents and young adults, 2020," *CA: a Cancer Journal for Clinicians*, vol. 70, no. 6, pp. 443–459, 2020.

[5] S. B. Edge and C. C. Compton, "The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.

[6] A. Jemal, R. Siegel, E. Ward et al., "Cancer statistics, 2008," *CA: a Cancer Journal for Clinicians*, vol. 58, no. 2, pp. 71–96, 2008.

[7] V. Karantza, "Keratins in health and cancer: more than mere epithelial cell markers," *Oncogene*, vol. 30, no. 2, pp. 127–138, 2011.

[8] N. Tanimura and Y. Fujita, "Epithelial defense against cancer (EDAC)," *Seminars in Cancer Biology*, vol. 63, pp. 44–48, 2020.

[9] H. Lv, R. Liu, J. Fu et al., "Epithelial cell-derived periostin functions as a tumor suppressor in gastric cancer through stabilizing p53 and E-cadherin proteins via the Rb/E2F1/p14ARF/Mdm2 signaling pathway," *Cell Cycle*, vol. 13, no. 18, pp. 2962–2974, 2014.

[10] L. M. McCaffrey and I. G. Macara, "Epithelial organization, cell polarity and tumorigenesis," *Trends in Cell Biology*, vol. 21, no. 12, pp. 727–735, 2011.

[11] C. Royer and X. Lu, "Epithelial cell polarity: a major gatekeeper against cancer?," *Cell Death & Differentiation*, vol. 18, no. 9, pp. 1470–1477, 2011.

[12] S. G. Zeng, X. Lin, J.-C. Liu, and J. Zhou, "Hypoxia-induced internalization of connexin 26 and connexin 43 in pulmonary epithelial cells is involved in the occurrence of non-small cell lung cancer via the P53/MDM2 signaling pathway," *International Journal of Oncology*, vol. 55, no. 4, pp. 845–859, 2019.

[13] J. Varga and F. R. Greten, "Cell plasticity in epithelial homeostasis and tumorigenesis," *Nature Cell Biology*, vol. 19, no. 10, pp. 1133–1141, 2017.

[14] T. Knösel, V. Emde, K. Schlüns, P. M. Schlag, M. Dietel, and I. Petersen, "Cytokeratin profiles identify diagnostic signatures in colorectal cancer using multiplex analysis of tissue microarrays," *Cell Oncology*, vol. 28, no. 4, pp. 167–175, 2006.

[15] C. Ausch, V. Buxhofer-Ausch, U. Olszewski et al., "Caspase-cleaved cytokeratin 18 fragment (M30) as marker of postoperative residual tumor load in colon cancer patients," *European Journal of Surgical Oncology*, vol. 35, no. 11, pp. 1164–1168, 2009.

[16] C. K. Cheong, K. R. Y. Nistala, C. H. Ng et al., "Neoadjuvant therapy in locally advanced colon cancer: a meta-analysis and systematic review," *Journal of Gastrointestinal Oncology*, vol. 11, no. 5, pp. 847–857, 2020.

[17] D. Lambrechts, E. Wauters, B. Boeckx et al., "Phenotype molding of stromal cells in the lung tumor microenvironment," *Nature Medicine*, vol. 24, no. 8, pp. 1277–1289, 2018.

[18] F. Tang, C. Barbacioru, Y. Wang et al., "mRNA-Seq whole-transcriptome analysis of a single cell," *Nature Methods*, vol. 6, no. 5, pp. 377–382, 2009.

[19] E. Azizi, A. J. Carr, G. Plitas et al., "Single-cell map of diverse immune phenotypes in the breast tumor microenvironment," *Cell*, vol. 174, no. 5, pp. 1293–1308.e36, 2018.

[20] Q. Zhang, Y. He, N. Luo et al., "Landscape and dynamics of single immune cells in hepatocellular carcinoma," *Cell*, vol. 179, no. 4, pp. 829–845.e20, 2019.

[21] X. Zheng, J. Song, C. Yu et al., "Single-cell transcriptomic profiling unravels the adenoma-initiation role of protein tyrosine kinases during colorectal tumorigenesis," *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, 2022.

[22] S. Mangiola, M. A. Doyle, and A. T. Papenfuss, "Interfacing Seurat with the R tidy universe," *Bioinformatics*, vol. 37, no. 22, pp. 4100–4107, 2021.

[23] Z. Chen, H. Zhang, Y. Bai et al., "Single cell transcriptomic analysis identifies novel vascular smooth muscle subsets under high hydrostatic pressure," *Science China Life Sciences*, vol. 64, no. 10, pp. 1677–1690, 2021.

[24] J. H. Pan, H. Zhou, L. Cooper et al., "LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers," *Frontiers in Immunology*, vol. 10, p. 6, 2019.

[25] G. Lombardo, M. Gili, C. Grange et al., "IL-3R-alpha blockade inhibits tumor endothelial cell-derived extracellular vesicle (EV)-mediated vessel formation by targeting the β-catenin pathway," *Oncogene*, vol. 37, no. 9, pp. 1175–1191, 2018.

[26] H. Ichimiya, K. Maeda, A. Enomoto, L. Weng, M. Takahashi, and T. Murohara, "Girdin/GIV regulates transendothelial permeability by controlling VE-cadherin trafficking through the small GTPase, R-Ras," *Biochemical and Biophysical Research Communications*, vol. 461, no. 2, pp. 260–267, 2015.

[27] O. Gires, M. Pan, H. Schinke, M. Canis, and P. A. Baeuerle, "Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years?," *Cancer and Metastasis Reviews*, vol. 39, no. 3, pp. 969–987, 2020.

[28] G. Corso, J. Figueiredo, S. P. De Angelis et al., "E-cadherin deregulation in breast cancer," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 11, pp. 5930–5936, 2020.

[29] L. Fang, G. Yu, W. Yu, G. Chen, and B. Ye, "The correlation of WDR76 expression with survival outcomes and immune infiltrates in lung adenocarcinoma," *PeerJ*, vol. 9, article e12277, 2021.

[30] A. P. Patel, I. Tirosh, J. J. Trombetta et al., "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma," *Science*, vol. 344, no. 6190, pp. 1396–1401, 2014.

[31] Q. Xu, S. Chen, Y. Hu, and W. Huang, "Single-cell RNA transcriptome reveals the intra-tumoral heterogeneity and regulators underlying tumor progression in metastatic pancreatic ductal adenocarcinoma," *Cell Death Discovery*, vol. 7, no. 1, p. 331, 2021.

[32] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.

[33] A. S. Takhar, O. Eremin, and S. A. Watson, "The role of gastrin in colorectal carcinogenesis," *The Surgeon*, vol. 2, no. 5, pp. 251–257, 2004.

[34] D. M. Bryant and J. L. Stow, "The ins and outs of E-cadherin trafficking," *Trends in Cell Biology*, vol. 14, no. 8, pp. 427–434, 2004.

[35] A. P. Kowalczyk and A. B. Reynolds, "Protecting your tail: regulation of cadherin degradation by p120-catenin," *Current Opinion in Cell Biology*, vol. 16, no. 5, pp. 522–527, 2004.

[36] G. Swaminathan and C. A. Cartwright, "Rack1 promotes epithelial cell-cell adhesion by regulating E-cadherin endocytosis," *Oncogene*, vol. 31, no. 3, pp. 376–389, 2012.

[37] Y. Kamei, K. Kito, T. Takeuchi et al., "Human scribble accumulates in colorectal neoplasia in association with an altered distribution of β-catenin," *Human Pathology*, vol. 38, no. 8, pp. 1273–1281, 2007.

[38] Z. Ouyang, W. Zhan, and L. Dan, "hScrib, a human homolog of Drosophila neoplastic tumor suppressor, is involved in the progress of endometrial cancer," *Oncology Research*, vol. 18, no. 11, pp. 593–599, 2009.

[39] S. Wan, A.-S. Meyer, S. M. E. Weiler et al., "Cytoplasmic localization of the cell polarity factor scribble supports liver tumor formation and tumor cell invasiveness," *Hepatology*, vol. 67, no. 5, pp. 1842–1856, 2018.

[40] R. Takagawa, K. Akimoto, Y. Ichikawa et al., "High expression of atypical protein kinase C lambda/iota in gastric cancer as a prognostic factor for recurrence," *Annals of Surgical Oncology*, vol. 17, no. 1, pp. 81–88, 2010.

[41] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann, "The technology and biology of single-cell RNA sequencing," *Molecular Cell*, vol. 58, no. 4, pp. 610–620, 2015.

[42] X. Ren, B. Kang, and Z. Zhang, "Understanding tumor ecosystems by single-cell sequencing: promises and limitations," *Genome Biology*, vol. 19, no. 1, p. 211, 2018.

[43] A. J. Shih, A. Menzin, J. Whyte et al., "Identification of grade and origin specific cell populations in serous epithelial ovarian cancer by single cell RNA-seq," *PLoS One*, vol. 13, no. 11, article e0206785, 2018.

[44] Y.-E. Guo, Y. Li, B. Cai et al., "Phenotyping of immune and endometrial epithelial cells in endometrial carcinomas revealed by single-cell RNA sequencing," *Aging*, vol. 13, no. 5, pp. 6565–6591, 2021.

[45] K. Bach, S. Pensa, M. Grzelak et al., "Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing," *Nature Communications*, vol. 8, no. 1, p. 2128, 2017.

[46] Y. Xu, T. Mizuno, A. Sridharan et al., "Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis," *JCI Insight*, vol. 1, no. 20, article e90558, 2016.

[47] X. Wang, X. Dou, X. Ren et al., "A ductal-cell-related risk model integrating single-cell and bulk sequencing data predicts the prognosis of patients with pancreatic adenocarcinoma," *Frontiers in Genetics*, vol. 12, article 763636, 2021.

[48] H. Zheng, H. Liu, Y. Ge, and X. Wang, "Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer," *Cancer Cell International*, vol. 21, no. 1, 2021.

[49] Y. Li, X. Zhao, Q. Liu, and Y. Liu, "Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis," *Annals of Medicine*, vol. 53, no. 1, pp. 1019–1031, 2021.

[50] H. Li, L. Zhao, Y. S. Lau, C. Zhang, and R. Han, "Genome-wide CRISPR screen identifies LGALS2 as an oxidative stress-responsive gene with an inhibitory function on colon tumor growth," *Oncogene*, vol. 40, no. 1, pp. 177–188, 2021.

[51] J.-H. Jung, H.-J. Kim, J. Yeom et al., "Lowered expression of galectin-2 is associated with lymph node metastasis in gastric cancer," *Journal of Gastroenterology*, vol. 47, no. 1, pp. 37–48, 2012.

[52] J. Lu, Y. Li, Y. A. Li et al., "In vivo detection of dysregulated choline metabolism in paclitaxel-resistant ovarian cancers with proton magnetic resonance spectroscopy," *Journal of Translational Medicine*, vol. 20, no. 1, p. 92, 2022.

[53] J. H. You, J. Lee, and J. L. Roh, "Mitochondrial pyruvate carrier 1 regulates ferroptosis in drug-tolerant persister head and neck cancer cells via epithelial-mesenchymal transition," *Cancer Letters*, vol. 507, pp. 40–54, 2021.

[54] J. C. Schell, K. A. Olson, L. Jiang et al., "A role for the mitochondrial pyruvate carrier as a repressor of the Warburg effect and colon cancer cell growth," *Molecular Cell*, vol. 56, no. 3, pp. 400–413, 2014.

[55] Y. Chai, C. Wang, W. Liu, Y. Fan, and Y. Zhang, "*MPC1* deletion is associated with poor prognosis and temozolomide resistance in glioblastoma," *Journal of Neurooncology*, vol. 144, no. 2, pp. 293–301, 2019.

[56] A. Matsuda, Y. Suzuki, G. Honda et al., "Large-scale identification and characterization of human genes that activate NF-κB and MAPK signaling pathways," *Oncogene*, vol. 22, no. 21, pp. 3307–3318, 2003.

[57] K. Wu and B. Bonavida, "The activated NF-κB-Snail-RKIP circuitry in cancer regulates both the metastatic cascade and resistance to apoptosis by cytotoxic drugs," *Critical Reviews of Immunology*, vol. 29, no. 3, pp. 241–254, 2009.

[58] S. Mukai, N. Oue, T. Oshima et al., "Overexpression of transmembrane protein BST2 is associated with poor survival of patients with esophageal, gastric, or colorectal cancer," *Annals of Surgical Oncology*, vol. 24, no. 2, pp. 594–602, 2017.

[59] H. Jin, L. Zhang, S. Wang, and L. Qian, "BST2 promotes growth and induces gefitinib resistance in oral squamous cell carcinoma via regulating the EGFR pathway," *Archives of Medicinal Science*, vol. 17, no. 6, pp. 1772–1782, 2021.

[60] D. Cai, J. Cao, Z. Li et al., "Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis," *BMC Cancer*, vol. 9, no. 1, p. 102, 2009.

[61] R. Lecocq, F. Lamy, C. Erneux, and J. E. Dumont, "Rapid purification and identification of calcyphosine, a Ca^{2+}-binding protein phosphorylated by protein kinase A," *Biochemistry Journal*, vol. 306, no. 1, pp. 147–151, 1995.

[62] W. Shao, Q. Wang, F. Wang, Y. Jiang, M. Xu, and J. Xu, "Abnormal expression of calcyphosine is associated with poor prognosis and cell biology function in colorectal cancer," *Onco Targets Therapy*, vol. 9, pp. 477–487, 2016.

[63] Z. Zhu, J. Wang, J. Tan et al., "Calcyphosine promotes the proliferation of glioma cells and serves as a potential therapeutic target," *Journal of Pathology*, vol. 255, no. 4, pp. 374–386, 2021.

[64] H. J. Johansson, B. C. Sanchez, J. Forshed et al., "Proteomics profiling identify CAPS as a potential predictive marker of tamoxifen resistance in estrogen receptor positive breast cancer," *Clinical Proteomics*, vol. 12, no. 1, p. 8, 2015.

[65] E. Kubo, N. Hasanova, N. Fatma, H. Sasaki, and D. P. Singh, "Elevated tropomyosin expression is associated with epithelial-mesenchymal transition of lens epithelial cells," *Journal of Cellular and Molecular Medicine*, vol. 17, no. 1, pp. 212–221, 2013.

[66] T. Shibata, S. Shibata, Y. Ishigaki et al., "Tropomyosin 2 heterozygous knockout in mice using CRISPR-cas9 system displays the inhibition of injury-induced epithelial-mesenchymal transition, and lens opacity," *Mechanisms of Ageing and Development*, vol. 171, pp. 24–30, 2018.

[67] D. Q. Li, L. Wang, F. Fei et al., "Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry," *Proteomics*, vol. 6, no. 11, pp. 3352–3368, 2006.

[68] Z. Tian, J. Zhao, and Y. Wang, "The prognostic value of TPM1–4 in hepatocellular carcinoma," *Cancer Medicine*, vol. 11, no. 2, pp. 433–446, 2022.

[69] H.-Y. Tang, L. A. Beer, J. L. Tanyi, R. Zhang, Q. Liu, and D. W. Speicher, "Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer," *Journal of Proteomics*, vol. 89, pp. 165–178, 2013.

[70] Y. Zhou, S. Bian, X. Zhou et al., "Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer," *Cancer Cell*, vol. 38, no. 6, pp. 818–828.e5, 2020.

[71] Y. Ma, T. Xiao, Q. Xu, X. Shao, and H. Wang, "iTRAQ-based quantitative analysis of cancer-derived secretory proteome reveals TPM2 as a potential diagnostic biomarker of colorectal cancer," *Frontiers in Medicine*, vol. 10, no. 3, pp. 278–285, 2016.

[72] W. Zhang, Z. Liu, B. Liu et al., "*GNG5* is a novel oncogene associated with cell migration, proliferation, and poor prognosis in glioma," *Cancer Cell International*, vol. 21, no. 1, 2021.

[73] G. Liu, X. Zeng, B. Wu, J. Zhao, and Y. Pan, "RNA-seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with radiotherapy response of nasopharyngeal carcinoma and prognosis of head and neck cancer," *Cancer Biology and Therapy*, vol. 21, no. 2, pp. 139–146, 2020.

[74] G. C. Tan, G. Negro, A. Pinggera et al., "Aldosterone-producing adenomas: histopathology-genotype correlation and identification of a novel *CACNA1D* mutation," *Hypertension*, vol. 70, no. 1, pp. 129–136, 2017.

[75] L. Larue and A. Bellacosa, "Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol $3'$ kinase/AKT pathways," *Oncogene*, vol. 24, no. 50, pp. 7443–7454, 2005.

[76] J. Hofmanová, J. Slavík, M. Ciganek et al., "Complex alterations of fatty acid metabolism and phospholipidome uncovered in isolated colon cancer epithelial cells," *International Journal of Molecular Science*, vol. 22, no. 13, p. 6650, 2021.

[77] V. Mohan, A. Das, and I. Sagi, "Emerging roles of ECM remodeling processes in cancer," *Seminars in Cancer Biology*, vol. 62, pp. 192–200, 2020.

[78] L. J. Hernández Borrero and W. S. El-Deiry, "Tumor suppressor p53: biology, signaling pathways, and therapeutic targeting," *Biochimica Biophysica Acta Reviews on Cancer*, vol. 1876, no. 1, article 188556, 2021.