论著·基础研究

基于单细胞RNA测序的结直肠癌预后预测模型的建立和验证

马燕如[1]，季林华[2]，童天颖[1]，严宇青[1]，沈超琴[1]，张昕雨[1]，曹颖颖[1]，洪　洁[1]，陈豪燕[1]
1.上海交通大学医学院附属仁济医院消化科，上海 200001；2.上海交通大学医学院附属仁济医院胃肠外科，上海 200001

［摘要］**目的**·基于单细胞RNA测序（single cell RNA sequence，scRNA-seq）技术构建结直肠癌预后预测模型。**方法**·利用GEO（Gene Expression Omnibus）数据库获取结直肠癌样本的scRNA-seq数据集，筛选与结直肠癌转移相关的差异基因作为预测模型的候选基因，运用套索回归算法（LASSO）、Logistic回归和Kaplan-Meier生存分析进一步在癌症基因组图谱（The Cancer Genome Atlas，TCGA）数据库中筛选及验证与结直肠癌预后相关的基因集，并建立结直肠癌预后预测模型。通过决策曲线分析和受试者工作特征（receiver operating characteristic，ROC）曲线评估预测模型在临床应用中的价值。**结果**·利用GEO数据库获取的scRNA-seq数据筛选出30个差异表达基因，进一步在TCGA数据库中利用LASSO回归得到9个关键基因，并以此对每例患者的关键基因表达进行评分。分别在训练集和验证集中对复发和未复发患者的评分进行比较，差异均有统计学意义（P<0.05）。采用Logistic回归分析将肿瘤原发灶分级（T stage）和是否发生远处转移（M stage）2个独立的临床变量纳入评分-临床变量整合模型。对评分-临床变量整合模型的实际预测价值进行评估，ROC曲线在训练集和验证集的曲线下面积分别为0.775和0.705。**结论**·基于scRNA-seq结果，构建了较为稳定的结直肠癌预后预测模型，可供临床评估患者预后参考。

［关键词］单细胞RNA测序；结直肠癌；预后；生物信息学
［DOI］10.3969/j.issn.1674-8115.2021.02.006　［中图分类号］R735.3；Q78　［文献标志码］A

Establishment and validation of prognostic prediction model of colorectal cancer based on single-cell RNA sequencing

MA Yan-ru[1], JI Lin-hua[2], TONG Tian-ying[1], YAN Yu-qing[1], SHEN Chao-qin[1], ZHANG Xin-yu[1], CAO Ying-ying[1], HONG Jie[1], CHEN Hao-yan[1]
1.Department of Gastroenterology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200001, China; 2.Department of Gastrointestinal Surgery, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200001, China

[Abstract] **Objective**·To establish a model for predicting the prognosis in patients with colorectal cancer (CRC) using single cell RNA sequencing (scRNA-seq). **Methods**·scRNA-seq data of patients with CRC from Gene Expression Omnibus (GEO) database was used to filter out candidate genes, which were related to metastatic CRC. The least absolute shrinkage and selection operator (LASSO) regression, Logistic regression and Kaplan-Meier analysis were used to select and evaluate the significance of the hub gene filtered out in The Cancer Genome Atlas (TCGA) database, and to develop the prognostic prediction model of CRC. Decision curve analysis and receiver operating characteristic (ROC) curve were used to assess the clinical use of the prediction model. **Results**·Thirty candidate genes were filtered out from the scRNA-seq data which was downloaded in GEO database, and then 9 hub genes were selected by LASSO regression in the TCGA database. The hub-gene expression was scored for each patient. The scores had significant difference between the groups with and without recurrence both in the training set and the validation set (P<0.05). In addition, Logistic regression analysis was carried out to incorporate the two independent clinical variables of primary tumor grade (T stage) and metastasis status (M stage) into the score-clinical variable integration model. Area under curve of the ROC curve in the training set and validation set were 0.775 and 0.705, respectively. **Conclusion**·A relatively stable model for predicting prognosis in CRC was constructed based on the results of scRNA-seq, which has certain guiding significance for treatment decision and prognostic prediction.

[Key words] single cell RNA sequencing (scRNA-seq); colorectal cancer; prognosis; bioinformatics

目前，结直肠癌（colorectal cancer，CRC）是世界上第三大常见的癌症，高居癌症相关死因第二位[1]。约25%的CRC患者存在转移，而约20%的患者在患病期间就出现远处转移[2-3]。转移性CRC的治疗对临床来说仍是一个挑战，实际治愈率较低[4]。因此，越来越多的研究开始尝试建立各种模型来完善CRC的临床分级和预后判断，为临床决策提供参考[5-7]。

RNA测序（RNA sequence，RNA-seq）技术已广泛应用于转录组分析，用以研究转录结构、剪接模式、基因表达差异等[8-9]。然而，普通RNA-seq方法忽视了组织

[基金项目] 上海市教育委员会高峰高原学科建设计划（20161309）；上海交通大学医学院高水平地方高校创新团队（SSMU-ZLCX20180200）。
[作者简介] 马燕如（1995—），女，硕士生；电子信箱：mayanru0213@163.com。
[通信作者] 陈豪燕，电子信箱：haoyanchen@shsmu.edu.cn。
[Funding Information] Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support (20161309); Innovative Research Team of High-Level Local Universities in Shanghai (SSMU-ZLCX20180200).
[Corresponding Author] CHEN Hao-yan, E-mail：haoyanchen@shsmu.edu.cn.

内存在的异质性，无法识别不同细胞群的内部特征。近年来，单细胞RNA测序（single cell RNA sequence，scRNA-seq）技术的发展为根据肿瘤内基因表达活性分析细胞类型和状态提供了新思路[10-11]，有助于肿瘤细胞的进一步分类。该技术已经成功应用于胰腺导管腺癌、乳腺癌和肺癌等[12-13]。此外，scRNA-seq在难治性肿瘤的研究中具有一定的应用价值，在监测循环肿瘤细胞、分析瘤内异质性和利用其敏感性评估复发性肿瘤预后等方面具有一定的优势[14]。Zhang等[15]利用scRNA-seq对患者来源的CRC样本进行了分析，发现发生和未发生肝转移的肿瘤组织在肿瘤微环境的信号转导、代谢、免疫调节等多个方面的基因表达上存在差异。

本研究利用GEO数据库的CRC患者scRNA-seq数据筛选在转移组织中差异表达的候选基因，并对候选基因进行大样本分析，以验证其在预测CRC预后中的意义，为进一步的个体化治疗提供参考。

1 材料与方法

1.1 数据获取

从美国国家生物技术信息中心的Gene Expression Omnnibus（GEO）数据库（https://www.ncbi.nlm.nih.gov/geo/）下载CRC scRNA-seq数据集GSE97693。原数据集对12例CRC患者的肿瘤原发灶、淋巴结转移灶和远处转移灶样本进行了单细胞测序，共190个单细胞数据通过了质量控制。本研究提取了其中淋巴结转移灶和相应的肿瘤原发灶数据以供后续分析。从癌症基因组图谱（The Cancer Genome Atlas，TCGA）数据库下载588例CRC样本转录组表达谱以及对应的临床信息，包括年龄、性别、肿瘤原发灶分级（T stage）、是否有淋巴结转移（N stage）、是否发生远处转移（M stage）、无进展生存期（progression free survival，PFS）和无进展生存状态共7个指标。

1.2 数据处理

提取下载的scRNA-seq数据，通过Scanpy包进行数据分析[16]，参考基因组为GRCh38。利用统一流形逼近与投影（uniform manifold approximation and projection，UMAP）算法对scRNA-seq数据进行降维分析，采用Louvain算法进行聚类分析。以标准化P=0.05为截止标准，再根据基因表达比值对数的绝对值（|log fold change|，|log FC|）由高到低排序，分别筛选出转移组和未转移组前15个差异表达基因，组成候选基因集，结果以热图的形式展现。

1.3 CRC患者中标记基因的筛选和鉴定

将来自TCGA数据库的588例CRC患者表达谱随机分为训练集和验证集（49.8%：50.2%）。提取训练集中293例患者标记基因的转录组图谱。在scRNA-seq获得的候选基因集中，使用glmnet包选取其中9个与预后相关的关键基因，以是否复发为因变量，通过Logistic回归分析的方法建立套索回归算法（the least absolute shrinkage and selection operator，LASSO）回归模型。根据建立的回归模型对每例患者的关键基因表达进行评分，评分=$-\log10\left[-\Sigma\left(\beta_i\times Exp_i\right)\right]$；其中$\beta_i$为系数，代表赋予的每个关键基因表达的加权。随后，分别在训练集和验证集中对复发和未复发患者的关键基因评分利用Wilcoxon检验进行比较；并通过Kaplan-Meier分析评估训练集、验证集和总TCGA集预后结局的差异。

1.4 预后预测模型的建立和评估

合并总TCGA集的评分和其他临床变量。采用Logistic回归分析对重要的临床变量进行评估。排除无统计学意义的变量后，利用广义线性模型（generalized linear model，GLM）建立完整的模型。采用具有曲线下面积（area under curve，AUC）的受试者工作特征（receiver operating characteristic，ROC）曲线评估列线图的实际预测价值。为了评估模型的临床使用价值，通过对总TCGA集中不同阈值概率下患者的净获益进行量化，采用决策曲线分析（decision curve analysis，DCA）确定列线图的临床有效性。

1.5 生物信息学分析和统计学分析

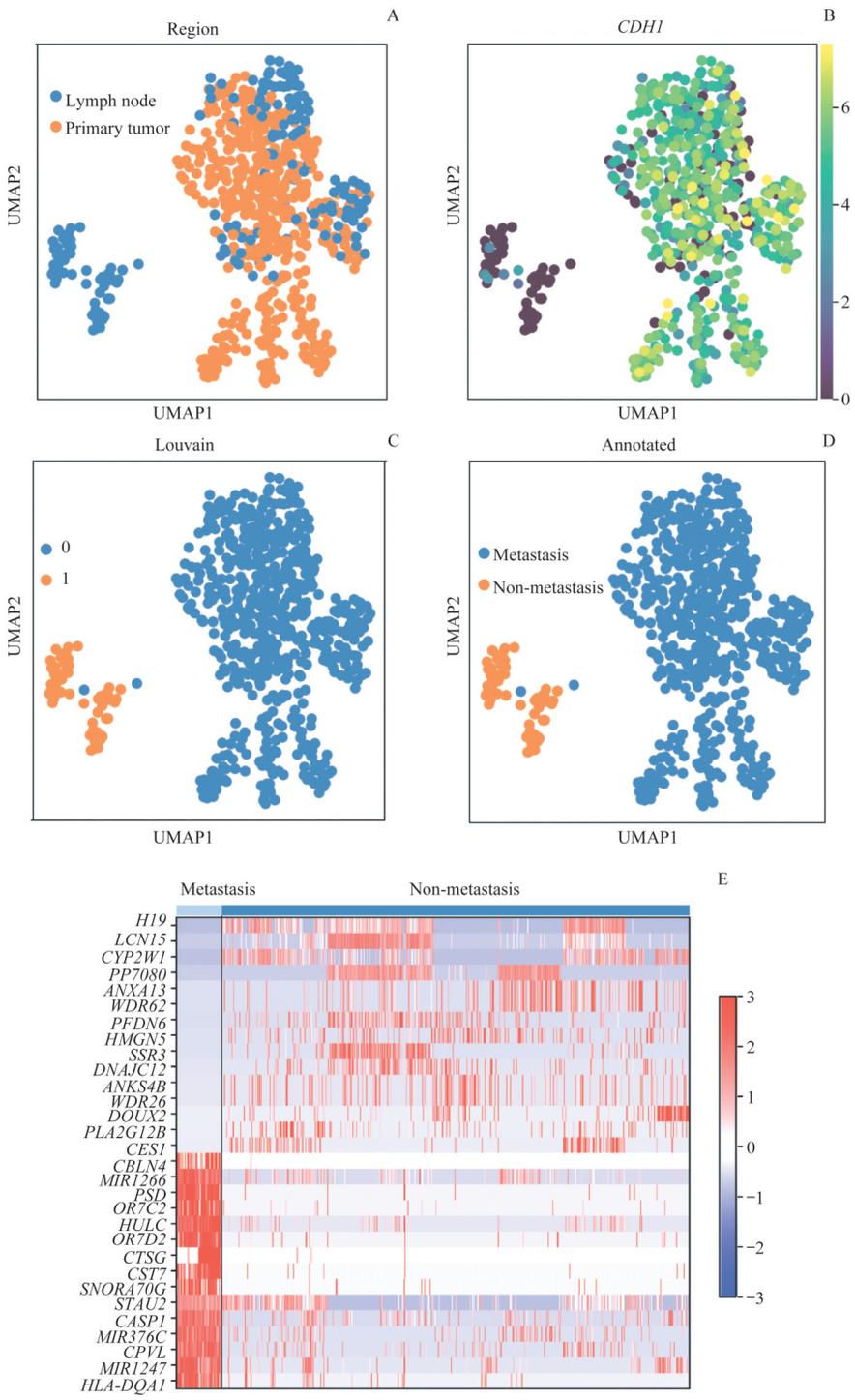
采用glmnet和survival包进行LASSO回归、Logistic回归和Kaplan-Meier分析。GLM和列线图用rms包建立。使用Wilcoxon秩和检验判断评分在复发和未复发患者中的差异是否具有统计学意义。所有统计分析均在R studio（3.6.1版）软件中进行，P<0.05表示差异有统计学意义。

2 结果

2.1 在单细胞水平筛选CRC预后相关基因

为从单细胞水平筛选CRC预后相关基因，从GEO数据库获取CRC样本scRNA-seq数据集，通过UMAP算法，对scRNA-seq数据进行降维分析，结果显示淋巴结转移灶与肿瘤原发病灶来源的细胞群分布具有较大差异（图1A）。进一步采用Louvain算法进行聚类分析，将scRNA-seq数据分为2个亚群，其中肿瘤转移相关基因钙黏蛋白1

（cadherin 1，*CDH1*）的表达在细胞群0和细胞群1中具有显著差异（图1B、C）。研究[17]表明，*CDH1*突变与细胞黏附减少而增殖、侵袭、转移增加相关；因此，本研究将细胞群0和1分别定义为未转移组和转移组，以进一步分析CRC转移相关基因（图1D）。分别在转移组和未转移组选择具有统计学意义（P<0.05）的前15个差异表达基因作为后续研究的候选基因（图1E）。



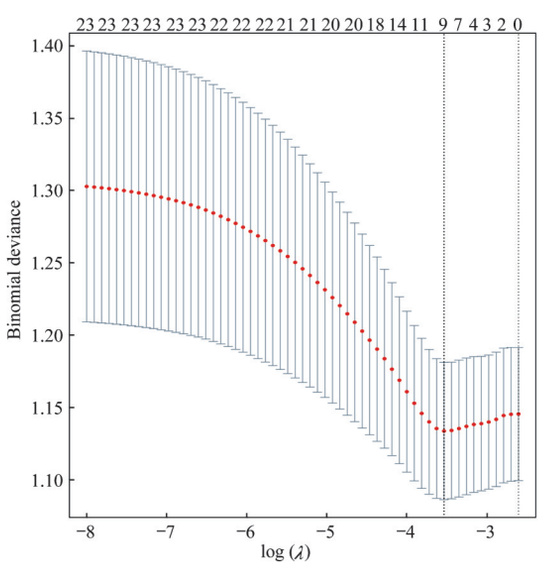
Note: A. The distribution of cell population in primary tumor and lymph node metastasis of CRC patients was analyzed by the UMAP algorithm. B/C. The distribution of *CDH1* in the cell population (B), and based on this, the cells were divided in to 0 and 1 groups (C). D. Based on the function of *CDH1*, the cells were divided into metastatic and non-metastatic groups. E. The heat map shows significant differences between the metastasis group and the non-metastasis group (P<0.05), and the darker the color is, the greater the difference is.

图1　单细胞水平筛选CRC预后相关基因

Fig 1　Screening genes associated with prognosis of CRC from single-cell level

2.2 对关键基因的评分和验证

在训练集人群中，采用LASSO回归从候选基因集选择了9个与预后相关的关键基因（图2），分别是膜联蛋白A13（annexin A13，*ANXA13*）、凋亡蛋白酶1（caspase 1，*CASP1*）、卵黄样羧肽酶（carboxypeptidase vitellogenic like，*CPVL*）、胱抑素7（cystatin-7，*CST7*）、长链非编码RNA *H19*、嗅觉受体家族7亚家族C成员2（olfactory receptor family 7 subfamily C member 2，*OR7C2*）、信号序列受体3（signal sequence receptor subunit 3，*SSR3*）、WD重复结构域26（WD repeat domain 26，*WDR26*）和*WDR62*。并根据建立的回归模型对每例患者的关键基因表达情况进行评分。CRC患者的年龄、性别、肿瘤原发灶分级、是否有淋巴结转移、是否发生远处转移、关键基因评分（score）在训练集和验证集中的差异均无统计学意义（表1）。



Note: The adjustment parameter selection in the LASSO model passes the minimum standard using 5-fold cross validation. The solid gray line represents the local likelihood. A dashed vertical line is drawn with the best value by using the minimum criteria and the 1 standard error of the minimum criteria (the 1−SE criteria). The optimal λ value of 0.029 06 with log (λ)=−3.538 4 was selected.

图2 采用LASSO回归选择关键基因

Fig 2 Texture feature selection using LASSO binary Logistic regression model

分别在训练集和验证集中对复发和未复发患者的评分进行比较，差异均有统计学意义（图3A、B，*P*值分别为0.000和0.002）。根据评分的中位数，分别在训练集、验证集和总TCGA集中将患者分为高分组和低分组，再进行Kaplan-Meier分析。结果表明，训练集中评分高的患者PFS结局明显较差（图3C，*P*=0.002），并在随访时间>1个月的验证集和总TCGA集中得到了一致验证（图3D、E，*P*值分别为0.017和0.000）。

表1 CRC患者的临床特征

Tab 1 Clinical characteristics of patients with CRC

Item	Training set (*n*=293)	Validation set (*n*=295)	*P* value①
Gender/*n* (%)			0.406
Male	160 (54.61)	145 (49.20)	
Female	133 (45.39)	150 (50.80)	
Age/year	66.86±12.77	65.43±12.96	0.178
T stage/*n* (%)			0.478
T1	9 (3.07)	13 (4.41)	
T2	56 (19.11)	45 (15.25)	
T3	198 (67.58)	201 (68.14)	
T4	30 (10.24)	36 (12.20)	
N stage/*n* (%)			0.592
N0②	173 (59.05)	165 (44.93)	
N1③	119 (40.61)	129 (43.73)	
NA④	1 (0.34)	1 (0.34)	
M stage/*n* (%)			0.342
M0	220 (75.09)	217 (73.56)	
M1	36 (12.29)	46 (15.59)	
NA④	37 (12.62)	32 (10.85)	
Score	−0.02±0.16	0.04±0.17	0.222

Note: ①The *P* value is obtained from univariate analysis between the training set and the validation set. ②N0 means no lymph node metastasis. ③N1 means lymph node metastasis. ④NA means missing data.

2.3 构建预后预测模型

将评分与其他临床变量相结合，构建完整的预测CRC患者预后的模型。排除单因素Logistic回归分析中无统计学意义的年龄和性别变量（表2，*P*值分别为0.406和0.779）以及在多变量Logistic回归模型中无统计学意义的淋巴结转移分级变量（N stage）。选择是否发生转移、肿瘤原发灶分级和关键基因评分3个独立变量作为最终的模型变量。利用GLM回归算法建立了包含这3个变量的列线图（图4A）。列线图预测预后结局的AUC分别达到0.775和0.705（图4B、C）。

表2 训练集中评分和临床候选预测变量的单因素Logistic回归分析

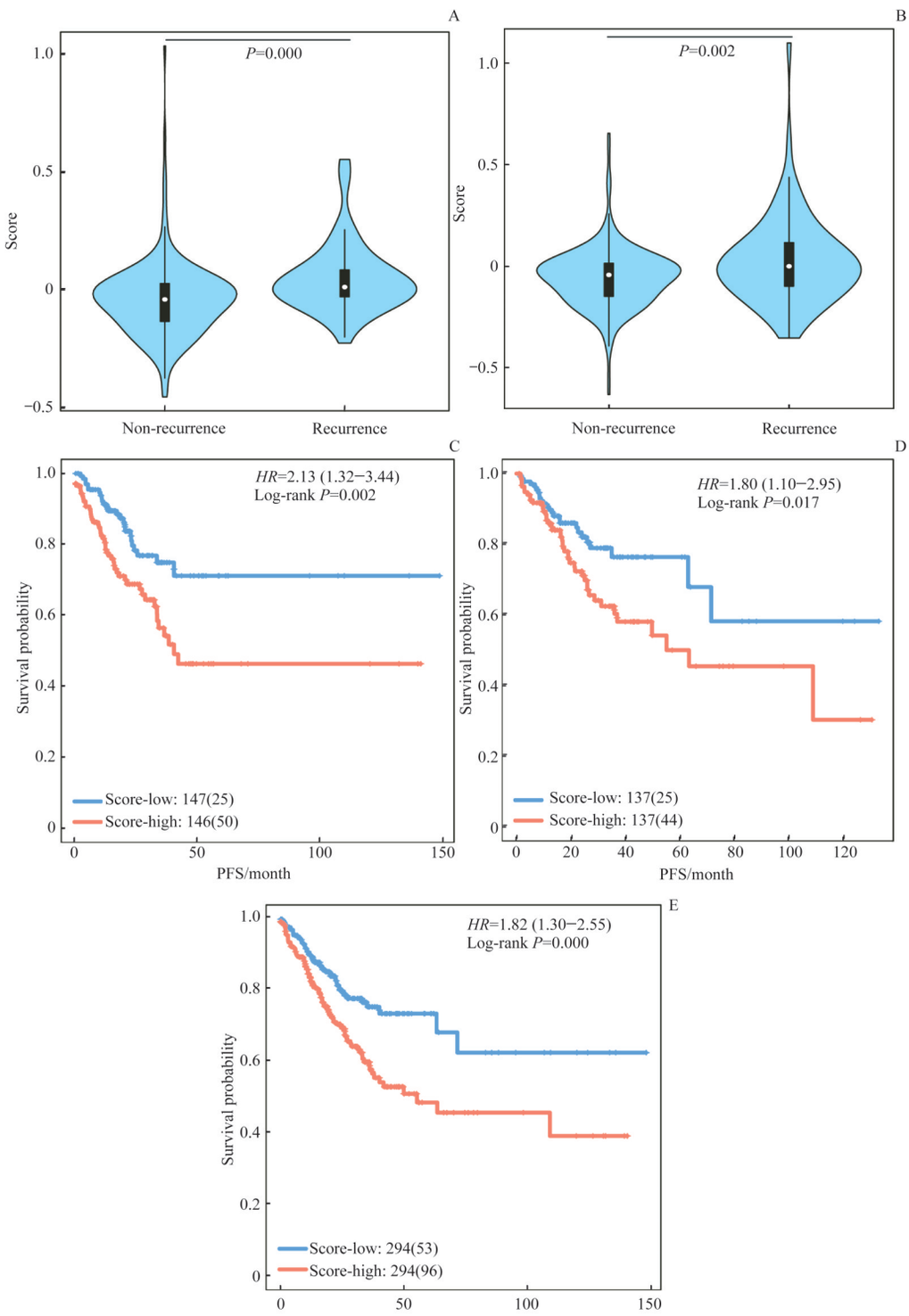
Tab 2 Univariate Logistic regression analysis of scores and clinical candidate predictors in training set

Variable	*OR*(95%*CI*)	*P* value
Age	1.01 (0.99–1.03)	0.406
Gender	0.93 (0.54–1.57)	0.779
T stage	2.82 (1.73–4.82)	0.000
N stage	2.15 (1.55–3.00)	0.000
M stage	6.47 (3.09–13.95)	0.000
Score	35.35 (6.27–241.99)	0.000

2.4 模型的临床应用价值

评分和评分-临床变量整合模型的DCA曲线如图5所

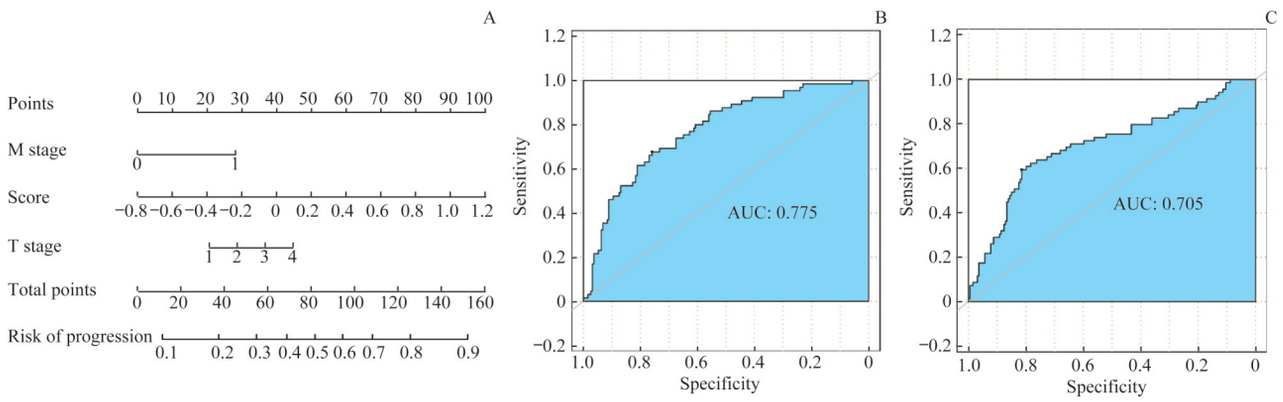
示。该曲线表明，如果复发风险阈值在4%~60%，使用评分-临床变量整合模型预测预后比治疗所有患者的方案或无治疗方案更有益；且在此范围内，使用整合模型预测的效果优于单纯使用评分。



Note：A/B. Violin plot of the score for prognosis in the training (A) and validation (B) set, respectively. C–E. Kaplan-Meier analysis shows that patients with high score have significantly poorer prognosis in the training set (C, *P*=0.002), validation set (D, *P*=0.017) and total TCGA set (E, *P*=0.000), respectively.

图3　关键基因评分验证

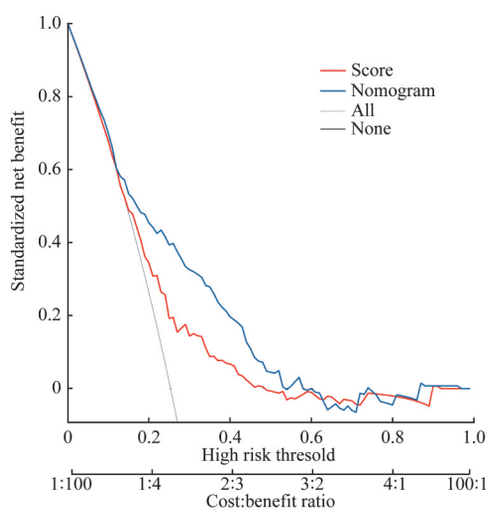
Fig 3　Hub gene score validation



Note：A. Nomogram for predicting prognosis. B/C. The ROC curves of the training set (B) and the validation set (C) are shown respectively.

图4　基于评分和部分临床数据预测复发的列线图

Fig 4　Nomogram for predicting prognosis based on the score and some clinical variables



Note：The *Y*-axis is the standardized net benefit, and the *X*-axis is the ratio of high-risk threshold probability and risk benefit. The net benefit is the proportion of patients who were true positive minus the proportion of patients who were false positive and then weighted according to the relative harm of continuing treatment *versus* the adverse consequences of unnecessary treatment. The figure shows a prognostic model of hub gene scores with (blue line) and without (red line) clinical variables, the solid gray line represents progression in all patients and the solid black line represents progression in none.

图5　列线图和关键基因评分的DCA

Fig 5　DCA of the nomogram and the hub gene scores

3　讨论

本研究通过GEO数据库获取CRC的scRNA-seq数据，从单细胞水平探究CRC预后相关的基因组特征。既往研究表明，*CDH1*基因编码产物钙黏蛋白（E-cadherin）是一种钙依赖的细胞间黏附分子和抑癌蛋白，通过细胞间黏附复合物在上皮细胞形成和分化中起关键作用[17-18]。*CDH1*的突变与肿瘤的浸润和转移能力相关，黏附复合物的破坏使细胞间黏附丧失，细胞活性增加[18]。基于此，本研究根据*CDH1*的表达将细胞群分为转移和未转移2个群。分别在转移组和未转移组选择具有统计学意义（P<0.05）的前15个差异表达基因作为后续研究的候选基因，利用LASSO回归算法进行筛选，确认了9个关键基因，其中部分基因在恶性肿瘤的进展中起重要作用。与未发生转移的CRC患者相比，*SSR3*在发生转移的CRC患者中显著升高，提示其可能与CRC的疾病进展相关[19]。在急性淋巴细胞白血病中，*CASP1*的过表达导致糖皮质激素受体分裂，糖皮质激素诱导的转录反应减弱，糖皮质激素耐药性增强[20]。*H19*是一个长链非编码RNA，其可以通过H19/S-腺苷高半胱氨酸水解酶/DNA甲基转移酶3B（H19/SAHH/DNMT3B）轴导致细胞自噬激活，从而引起乳腺癌患者对他莫昔芬治疗产生耐药[21]。*CST7*则在肿瘤逃逸和耐受、肝癌晚期复发、肝癌进展和某些肝癌亚类中显著富集[22]。这为后续的研究提供了一个方向。

根据建立的回归模型对每例患者的关键基因表达进行评分后，利用随机分成的验证集检验评分在临床的应用价值，发现复发组评分显著高于未复发组，这与生存分析结果是一致的。至于其是否具有潜在的治疗预测价值尚不清楚，这将是另一个有意义的研究方向。随后采用Logistic回归分析，排除无统计学意义的年龄、性别和淋巴结转移分级变量，选择是否发生转移、肿瘤原发灶分级和关键基因评分3个独立的变量作为最终的整合模型变量。ROC曲线显示整合模型对CRC患者的预后具有较高的预测价值，且临床也可以据此对患者的个体化治疗方案进行及时调整和完善。

本研究将GEO数据库获得的scRNA-seq数据分析与TCGA数据库中的大样本人群数据验证相结合。与CRC传统转录组测序分析相比[23-24]，传统RNA-seq可能会覆盖潜在的重要标记基因，而scRNA-seq在这方面具有一

定的优势。本研究仍存在以下局限性：首先，下载的TCGA数据大多来自美国或欧洲人群，而这是否适用于亚洲人种尚不清楚，因此，该研究结论需在中国人群中进一步验证；其次，虽然该整合模型在大量CRC人群中得到了很好的验证，但仍需补充基础实验来揭示其促进肿瘤发展的具体机制。

综上所述，本研究基于scRNA-seq技术筛选标记基因，并对标记基因进行大样本分析，对预测临床CRC患者预后、优化治疗方案具有一定的指导意义。

参·考·文·献

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020[J]. CA Cancer J Clin, 2020, 70(1): 7-30.

[2] Hackl C, Neumann P, Gerken M, et al. Treatment of colorectal liver metastases in Germany: a ten-year population-based analysis of 5 772 cases of primary colorectal adenocarcinoma[J]. BMC Cancer, 2014, 14: 810.

[3] Elferink MAG, Jong KP, Klaase JM, et al. Metachronous metastases from colorectal cancer: a population-based study in North-East Netherlands[J]. Int J Color Dis, 2015, 30(2): 205-212.

[4] Brenner H, Kloor M, Pox CP. Colorectal cancer[J]. Lancet, 2014, 383 (9927): 1490-1502.

[5] Guan X, Ma CX, Quan JC, et al. A clinical model to predict the risk of synchronous bone metastasis in newly diagnosed colorectal cancer: a population-based study[J]. BMC Cancer, 2019, 19(1): 704.

[6] Liu JN, Li CY, Xu J, et al. A patient-oriented clinical decision support system for CRC risk assessment and preventative care[J]. BMC Med Inform Decis Mak, 2018, 18(Suppl 5): 118.

[7] Boursi B, Mamtani R, Hwang WT, et al. A risk prediction model for sporadic CRC based on routine lab results[J]. Dig Dis Sci, 2016, 61(7): 2076-2086.

[8] Gao MM, Zhong A, Patel N, et al. High throughput RNA sequencing utility for diagnosis and prognosis in colon diseases[J]. World J Gastroenterol, 2017, 23(16): 2819-2825.

[9] Han YX, Gao SG, Muegge K, et al. Advanced applications of RNA sequencing and challenges[J]. Bioinform Biol Insights, 2015, 9(Suppl 1): 29-46.

[10] Xie T, Wang YZ, Deng N, et al. Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis[J]. Cell Rep, 2018, 22(13): 3625-3640.

[11] Olsen TK, Baryawno N. Introduction to single-cell RNA sequencing[J]. Curr Protoc Mol Biol, 2018, 122(1): e57.

[12] Peng JY, Sun BF, Chen CY, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma[J]. Cell Res, 2019, 29(9): 725-738.

[13] Chung W, Eum HH, Lee HO, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer[J]. Nat Commun, 2017, 8: 15081.

[14] Navin NE. The first five years of single-cell cancer genomics and beyond[J]. Genome Res, 2015, 25(10): 1499-1507.

[15] Zhang YB, Song JJ, Zhao ZW, et al. Single-cell transcriptome analysis reveals tumor immune microenvironment heterogenicity and granulocytes enrichment in colorectal cancer liver metastases[J]. Cancer Lett, 2020, 470: 84-94.

[16] Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis[J]. Genome Biol, 2018, 19(1): 15.

[17] Takeichi M. Cadherins in cancer: implications for invasion and metastasis[J]. Curr Opin Cell Biol, 1993, 5(5): 806-811.

[18] Christofori G, Semb H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene[J]. Trends Biochem Sci, 1999, 24(2): 73-76.

[19] Zhi JJ, Sun JW, Wang ZC, et al. Support vector machine classifier for prediction of the metastasis of colorectal cancer[J]. Int J Mol Med, 2018, 41(3): 1419-1426.

[20] Paugh SW, Bonten EJ, Savic D, et al. NALP3 inflammasome upregulation and CASP1 cleavage of the glucocorticoid receptor cause glucocorticoid resistance in leukemia cells[J]. Nat Genet, 2015, 47(6): 607-614.

[21] Wang J, Xie SD, Yang JJ, et al. The long noncoding RNA *H19* promotes tamoxifen resistance in breast cancer *via* autophagy[J]. J Hematol Oncol, 2019, 12: 81.

[22] Zhou X, Wang X, Huang K, et al. Investigation of the clinical significance and prospective molecular mechanisms of cystatin genes in patients with hepatitis B virus-related hepatocellular carcinoma[J]. Oncol Rep, 2019, 42(1): 189-201.

[23] Weren RD, Ligtenberg MJ, Kets CM, et al. A germline homozygous mutation in the base-excision repair gene *NTHL1* causes adenomatous polyposis and colorectal cancer[J]. Nat Genet, 2015, 47(6): 668-671.

[24] Lin SH, Raju GS, Huff C, et al. The somatic mutation landscape of premalignant colorectal adenoma[J]. Gut, 2018, 67(7): 1299-1305.

［收稿日期］2020-05-21　　［本文编辑］吴　洋