

## 论著·基础研究

## 基于差异表达基因组合构建高度微卫星不稳定结直肠癌转移预测模型

徐莹, 褚以忞, 杨大明, 李吉, 张海芹, 彭海霞

上海交通大学医学院附属同仁医院内窥镜室, 上海 200336

**[摘要]** **目的**·从转录组层面探究影响高度微卫星不稳定 (microsatellite instability-high, MSI-H) 结直肠癌转移的潜在关键基因及基因表达特征, 并构建基因转移预测模型。 **方法**·从癌症基因组图谱数据库中收集 MSI-H 结直肠癌患者转录组数据, 根据转移信息分为转移组 (21 例) 和无转移组 (42 例), 分析 2 组间差异表达基因 (differentially expressed gene, DEG), 以基因本体数据库 (Gene Ontology, GO)、基因集富集分析 (Gene Set Enrichment Analysis, GSEA) 对 DEG 进行注释、聚类及信号通路富集; 使用 STRING、Cytoscape 软件筛选枢纽基因 (hub 基因); 选取 DEG 绘制列线图, 使用 Bootstrap 方法进行交叉验证; 分析列线图中每个基因对 MSI-H 结直肠癌无进展生存期 (progression-free survival, PFS) 的影响。 **结果**·转移组和无转移组间共得到 245 个 DEGs, 其中转移组较无转移组表达上调基因 204 个, 下调基因 41 个。GO 分析发现: DEG 在生物过程、分子功能上主要富集于离子穿膜转运、氯离子穿膜转运及氯离子通道活性; 在细胞组分中, 富集于细胞外部分、细胞外空间等。GSEA 结果显示: 上调基因富集于神经活性物质配体-受体相互作用和代谢信号通路。通过 Cytoscape 筛选出上调基因蛋白质互作网络中的前 10 位的 hub 基因。根据 DEG 中调整后  $P$  值最小且与肿瘤发生发展关联性高的前 10 个基因构建的转移预测模型有一定的预测效能, 其中训练集曲线下面积 (area under curve, AUC) = 0.975, 验证集 AUC=0.920; 模型中 *AC078993.1*、*IGLJ2* (immunoglobulin lambda joining 2) 的表达水平与 MSI-H 结直肠癌 PFS 呈明显负相关 ( $P=0.011$ ,  $P=0.005$ )。 **结论**·在 MSI-H 结直肠癌中, 离子通道变化及细胞外环境变化可能对肿瘤转移有重要影响, 神经活性物质配体-受体相互作用、代谢信号通路可能是对转移较重要的信号通路; 初步构建了 MSI-H 结直肠癌基因转移预测模型, 可为后续相关临床研究提供参考。

**[关键词]** 结直肠癌; 高度微卫星不稳定; 转移; 差异表达基因; 生物信息学; 列线图**[DOI]** 10.3969/j.issn.1674-8115.2021.09.010 **[中图分类号]** R735.34 **[文献标志码]** A

## Construction of a metastasis prediction model of microsatellite instability-high colorectal cancer based on differentially expressed gene assembly

XU Ying, CHU Yi-min, YANG Da-ming, LI Ji, ZHANG Hai-qin, PENG Hai-xia

Endoscopy Center, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200336, China

**[Abstract]** **Objective**·To explore the potential key genes and the gene expression characteristics of microsatellite instability-high (MSI-H) colorectal cancer (CRC) with metastasis at the transcriptome level, and establish a metastasis prediction gene model. **Methods**·The transcriptome data of MSI-H CRC patients was obtained from The Cancer Genome Atlas database. The patients were divided into metastatic group (21 patients) and non-metastatic group (42 patients). The differentially expressed genes (DEGs) between the two groups were analyzed by Gene Ontology (GO) and Gene Set Enrichment Analysis (GSEA) to annotate, and cluster DEGs and enrich the signaling pathways. STRING and Cytoscape were used to select the hub genes. Nomogram was drawn based on the selected DEGs. The cross validation of the model was performed by Bootstrap method. Survival analysis was done to explore the influences of each gene in the nomogram on progression-free survival (PFS) of MSI-H CRC. **Results**·A total of 245 DEGs were obtained from the metastatic group and non-metastatic group, among which 204 genes were up-regulated and 41 genes were down-regulated. GO analysis showed that DEGs were mainly clustered in ion transmembrane transport, chloride transmembrane transport and chloride channel activity in terms of biological process and molecular function. In terms of cellular component, DEGs were mainly clustered in extracellular region and extracellular space. GSEA showed that the neuroactive ligand-receptor interaction and metabolic pathways were enriched in the up-regulated genes. The top 10 hub genes in the protein-protein interaction network of the up-regulated genes were screened by Cytoscape. The metastasis prediction gene model, which was set up based on the top 10 DEGs with the lowest adjusted  $P$  value and high physiological relevance to tumor, had certain predictive efficiency [area under curve (AUC)

**[基金项目]** 上海市科学技术委员会基金 (18ZR1434900); 上海市长宁区科学技术委员会一般项目 (CNKW2018Y02); 上海交通大学医工 (理) 交叉基金资助 (ZH2018QNB24); 上海市第六人民医院医疗集团科研基金 (ly202003)。**[作者简介]** 徐莹 (1987—), 女, 主治医师, 博士生; 电子信箱: xy3459@shtrhospital.com。**[通信作者]** 彭海霞, 电子信箱: phx1101@shtrhospital.com。**[Funding Information]** Fund of Shanghai Municipal Science and Technology Committee (18ZR1434900); General Program of Science and Technology Commission of Changning District of Shanghai (CNKW2018Y02); Interdisciplinary Program of Shanghai Jiao Tong University (ZH2018QNB24); Research Fund of Shanghai Sixth People's Hospital Medical Group (ly202003)。**[Corresponding Author]** PENG Hai-xia, E-mail: phx1101@shtrhospital.com。

=0.975 for training, AUC=0.920 for validation]. The expression levels of *AC078993.1* and *IGLJ2* (immunoglobulin lambda joining 2) were significantly negatively correlated with PFS of MSI-H CRC ( $P=0.011$ ,  $P=0.005$ ). **Conclusion** The changes in ion channels and extracellular environment may have important impacts on metastasis of MSI-H CRC. Neuroactive ligand-receptor interaction and metabolic pathways may be two important signaling pathways for metastasis of MSI-H CRC. A metastasis prediction gene model is established, which can provide reference for the follow-up related clinical researches.

**[Key words]** colorectal cancer (CRC); microsatellite instability-high (MSI-H); metastasis; differentially expressed gene (DEG); bioinformatics; nomogram

结直肠癌是世界范围内发病率第四的恶性肿瘤, 其死亡率位列第二<sup>[1]</sup>。在我国, 结直肠癌位于所有恶性肿瘤发病率的第三位, 死亡率排名第五<sup>[2]</sup>。虽然伴随早期肿瘤发现的增多及肿瘤筛查的推广, 结直肠癌患者的5年生存率有所提高, 但是仍有25%的患者确诊时为IV期, 并且有25%~50%的患者确诊时为早期而后发展为转移性疾病<sup>[3-6]</sup>。转移性结肠癌患者的预后较差。一项基于美国人群的统计数据显示, 其5年中位生存率仅有12.5%<sup>[4]</sup>。因此, 寻找预判转移的生物学标志物有着重要的意义。

微卫星指的是广泛分布于原核和真核生物基因组中的短的串联重复序列, 约占人类基因的10%, 核心序列长度为1~6 bp; 微卫星不稳定 (microsatellite instability, MSI) 指基因组中短串联重复序列次数的增加或者减少<sup>[7-8]</sup>。MSI状态的检测可以通过PCR扩增特定微卫星标志位点、定向二代测序 (next-generation sequencing, NGS) 或免疫组化检测DNA错配修复 (mismatch repair, MMR) 蛋白表达来确定, 并可根据PCR扩增出的标志位点不稳定数或定向NGS中的不定位点累积得分将MSI分为高度微卫星不稳定 (microsatellite instability-high, MSI-H) 和低度微卫星不稳定 (microsatellite instability-low, MSI-L)<sup>[9-11]</sup>。MSI为结直肠癌的特殊分子表型, 在结直肠癌中占10%~20%<sup>[7-8]</sup>。既往研究<sup>[12-15]</sup>显示, MSI结直肠癌淋巴结转移和远处转移的发生率低于微卫星稳定 (microsatellite stable, MSS) 肿瘤。但是在转移性结直肠癌中, MSI-H患者的预后较差<sup>[16]</sup>。并且在早期结直肠癌中, MSI-H的比例高达20%, 而晚期转移性结直肠癌中, 该比例则显著下降 (4%~5%)<sup>[17-18]</sup>, 这提示在MSI-H结直肠癌中存在一部分高转移潜能的肿瘤。

本研究拟从转录组层面进一步探索影响MSI-H结直肠癌高转移潜能的因素。从癌症基因组图谱 (The Cancer Genome Atlas, TCGA) 数据库中筛选有无转移的MSI-H结直肠癌患者中的差异表达基因 (differentially expressed gene, DEG), 选取其中的关键基因构建转移预测列线图, 以帮助结直肠癌临床治疗及随访策略的制定。

## 1 材料与方法

### 1.1 数据收集

根据以下标准, 从TCGA数据库中筛选符合要求的患者: ①确诊结直肠癌。②分子分型为MSI-H (通过NGS的全外显子组测序数据得到)<sup>[19]</sup>。③患者标注有转移信息。

### 1.2 DEG筛选与功能富集分析

使用R语言edgeR软件包<sup>[20]</sup>对收集到的转移组及无转移组中DEGs进行分析。将差异阈值设定为误报率 (false detective rate, FDR) <0.05、 $\log_2$ 差异倍数 (fold change, FC) >1, 其中FDR为P值的多重校正值,  $\log_2$ FC为基因表达变化的数值和方向。

使用DAVID数据库 (the Database for Annotation, Visualization and Integrated Discovery) 对DEGs进行基因本体数据库 (Gene Ontology, GO) 注释与聚类, 筛选条件定为 $P<0.05$ 、基因数 $\geq 5$ 。

使用在线分析工具WEB-based GENE SeT Analysis Toolkit, 参照京都基因和基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 和Reactome信号通路数据库, 对DEGs进行基因集富集分析 (Gene Set Enrichment Analysis, GSEA), 分析DEGs涉及的信号通路。

使用STRING在线工具, 构建上调基因 (转移组较无转移组表达升高基因) 的蛋白质互作 (protein-protein interaction, PPI) 网络, 通过Cytoscape软件<sup>[21]</sup>筛选出PPI网络中高连接度等级前10位的枢纽基因 (hub基因)。

### 1.3 转移预测模型构建

选取DEGs中调整后P值 (adjusted P value,  $P_{adj}$ ) 最小, 且根据既往报道与肿瘤发生发展关联性高的前10个基因, 通过R语言的rms包对DEGs构建Logistic回归模型, 并通过R语言的“Boot”函数<sup>[22]</sup>, 使用Bootstrap方法, 将收集的63个肿瘤样本每次随机抽取25个, 进行15次检验, 对构建模型进行交叉验证。使用一致性指数

(concordance index, C-index)、受试者工作特征曲线(receiver operating characteristic curve, ROC 曲线)对模型预测效能进行评价。使用R语言的rms包绘制可视化列线图。

### 1.4 转移预测模型中各基因对 MSI-H 结直肠癌无进展生存期影响分析

结合收集到的TCGA数据库中的临床信息,使用Log-rank 检验并将基因中位表达值作为临界值,采用R语言survminer包行生存分析,分析列线图中每个基因表达水平对 MSI-H 结直肠癌无进展生存期(progression-free survival, PFS)的影响。

### 1.5 统计学方法

转移预测模型构建使用 Logistic 回归模型。模型预测效能评价使用 C-index、ROC 曲线。其中 C-index>0.7 则证明预测模型有可靠性;ROC 曲线的曲线下面积(area under curve, AUC)用来预测准确性,当 0.5<AUC<1 时提示优于随机猜测,模型有预测价值。

## 2 结果

### 2.1 MSI-H 结直肠癌转移与无转移组间 DEGs 筛选

基于筛选标准,一共从TCGA数据库中纳入63例患者。纳入患者的基本信息见表1。根据标注的转移信息将患者分为转移组(21例)及无转移组(42例),转移组为存在淋巴结转移或/和远端转移,无转移组为不存在淋巴结转移和远端转移。对2组进行转录组分析,共获得245个DEGs,其中转移组较无转移组表达升高的有204个,转移组较无转移组表达降低的有41个(图1)。上调及下调前10位DEG的详细信息见表2。

### 2.2 功能注释与富集分析

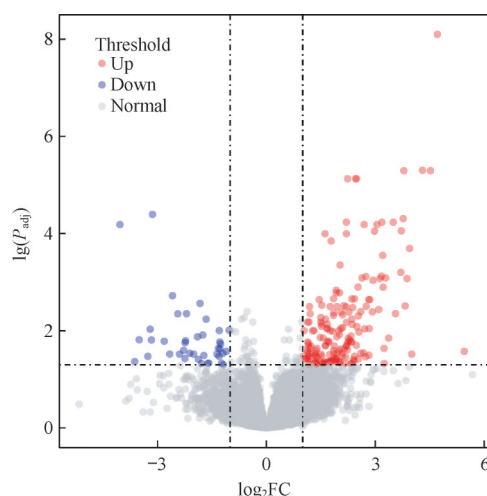
对DEGs进行GO注释及富集分析,将筛选条件定为 $P<0.05$ ,基因数 $\geq 5$ ,在上调基因中得到8项生物过程(biological process, BP)、12项细胞组分(cellular component, CC)、3项分子功能(molecular function, MF)。上调基因的BP主要为分泌、离子穿膜转运、神经肽信号通路等,CC主要为细胞外部分、质膜的组成部分、细胞外空间等,MF主要为激素活性、氯离子通道活性和生长因子活性。在下调基因中得到1项CC,为细胞外部分(图2)。

分别使用KEGG和Reactome信号通路数据库对

表1 63例纳入患者基本信息

Tab 1 Basic information of 63 included patients

Characteristic	Metastasis group (n=21)	Non-metastasis group (n=42)
Gender/n(%)		
Female	12 (57.1)	22 (52.4)
Male	9 (42.9)	20 (47.6)
Race/n(%)		
Unknown	9 (42.9)	7 (16.7)
Asian	0 (0)	1 (2.4)
Black or African American	2 (9.5)	7 (16.7)
White	10 (47.6)	27 (64.3)
Survival status/n(%)		
Alive	17 (81.0)	41 (97.6)
Dead	4 (19.0)	1 (2.4)



**Note:** Each dot in the graph represents a specific gene or transcript, with red dots representing significantly up-regulated genes, blue dots representing significantly down-regulated genes, and gray dots representing non-significantly different genes.

图1 MSI-H 结直肠癌转移组和无转移组间 DEGs 火山图

Fig 1 Volcano plot of DEGs between metastatic group and non-metastatic group of MSI-H colorectal cancer

DEGs进行信号通路的GSEA,并选取上调及下调基因富集前10位的信号通路;通过对2个数据库富集信号通路取交集发现,上调基因中神经活性物质配体-受体相互作用、代谢信号通路在2个数据库中都得到了富集(图3)。

使用STRING工具构建了上调基因PPI网络(图4A),并且通过Cytoscape软件筛选出其中高连接度等级前10位的hub基因,分别为胰高血糖素(glucagon, GCG)、生长激素抑制素(somatostatin, SST)、神经降压素(neurotensin, NTS)、 $\alpha 2$ -HS糖蛋白( $\alpha 2$ -HS glycoprotein, AHSB)、载脂蛋白B(apolipoprotein B, APOB)、嗜铬粒蛋白B(chromogranin B, CHGB)、突触素(synaptophysin, SYP)、胰岛素样生长因子结合蛋

表2 转移组和无转移组间上调及下调前10位 DEGs

Tab 2 Top 10 up-regulated and down-regulated DEGs between metastasis and non-metastasis group

DEG	Full name of gene	FDR	log <sub>2</sub> FC
Up-regulated			
<i>CA1</i>	Carbonic anhydrase 1	$5.85 \times 10^{-10}$	5.012
<i>IGLJ2</i>	Immunoglobulin lambda joining 2	$8.21 \times 10^{-6}$	4.686
<i>MS4A12</i>	Membrane spanning 4-domains A12	$3.68 \times 10^{-6}$	4.622
<i>SST</i>	Somatostatin	$9.04 \times 10^{-7}$	4.547
<i>GCG</i>	Glucagon	$1.58 \times 10^{-3}$	3.987
<i>SLC26A3</i>	Solute carrier family 26 member 3	$2.87 \times 10^{-5}$	3.960
<i>IGKV3OR2-5</i>	Immunoglobulin $\kappa$ variable 3 or 2-5 (pseudogene)	$4.42 \times 10^{-3}$	3.901
<i>OGDHL</i>	Oxoglutarate dehydrogenase 1	$1.53 \times 10^{-5}$	3.878
<i>AQP8</i>	Aquaporin 8	$4.66 \times 10^{-4}$	3.828
<i>HTR3C</i>	5-hydroxytryptamine receptor 3C	$6.22 \times 10^{-3}$	3.813
Down-regulated			
<i>GP2</i>	Glycoprotein 2	$8.44 \times 10^{-5}$	-4.070
<i>UICLM</i>	Up-regulated in colorectal cancer liver metastasis	$3.59 \times 10^{-2}$	-3.663
<i>FGL1</i>	Fibrinogen like 1	$1.08 \times 10^{-2}$	-3.245
<i>GPRC6A</i>	G protein-coupled receptor class C group 6 member A	$1.91 \times 10^{-2}$	-3.231
<i>DEFA6</i>	Defensin $\alpha$ 6	$3.69 \times 10^{-2}$	-3.231
<i>ACTL8</i>	Actin like 8	$6.77 \times 10^{-5}$	-3.161
<i>HSD3B1</i>	Hydroxy- $\delta$ -5-steroid dehydrogenase, 3 $\beta$ - and steroid $\delta$ -isomerase 1	$2.12 \times 10^{-2}$	-2.789
<i>PADI3</i>	Peptidyl arginine deiminase 3	$2.72 \times 10^{-2}$	-2.509
<i>MRLN</i>	Myoregulin	$3.02 \times 10^{-2}$	-2.477
<i>SLC9A4</i>	Solute carrier family 9 member A4	$2.03 \times 10^{-2}$	-2.344

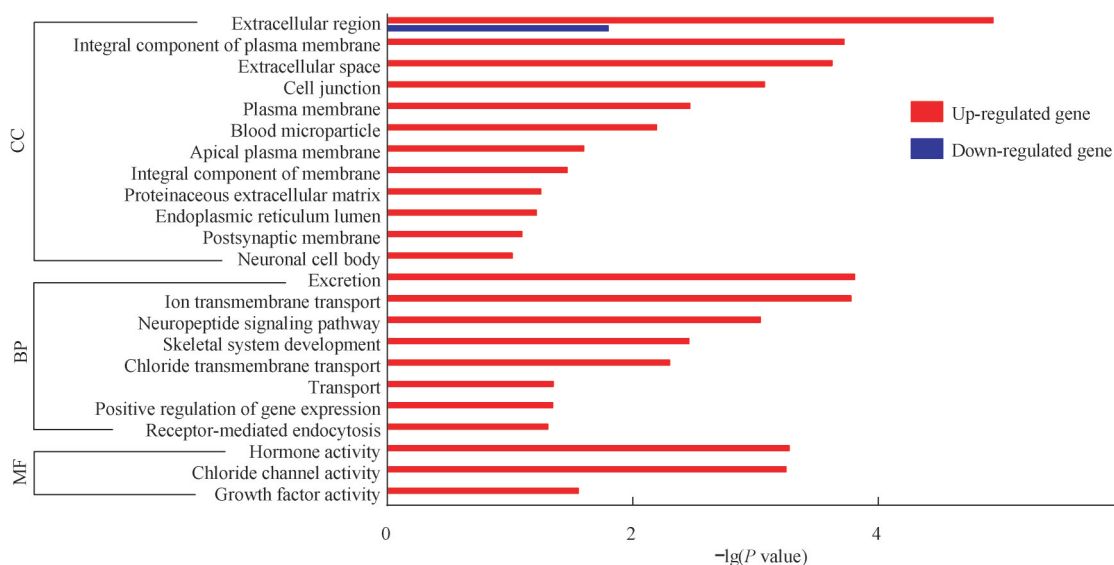


图2 GO分析上调与下调基因的BP、CC与MF

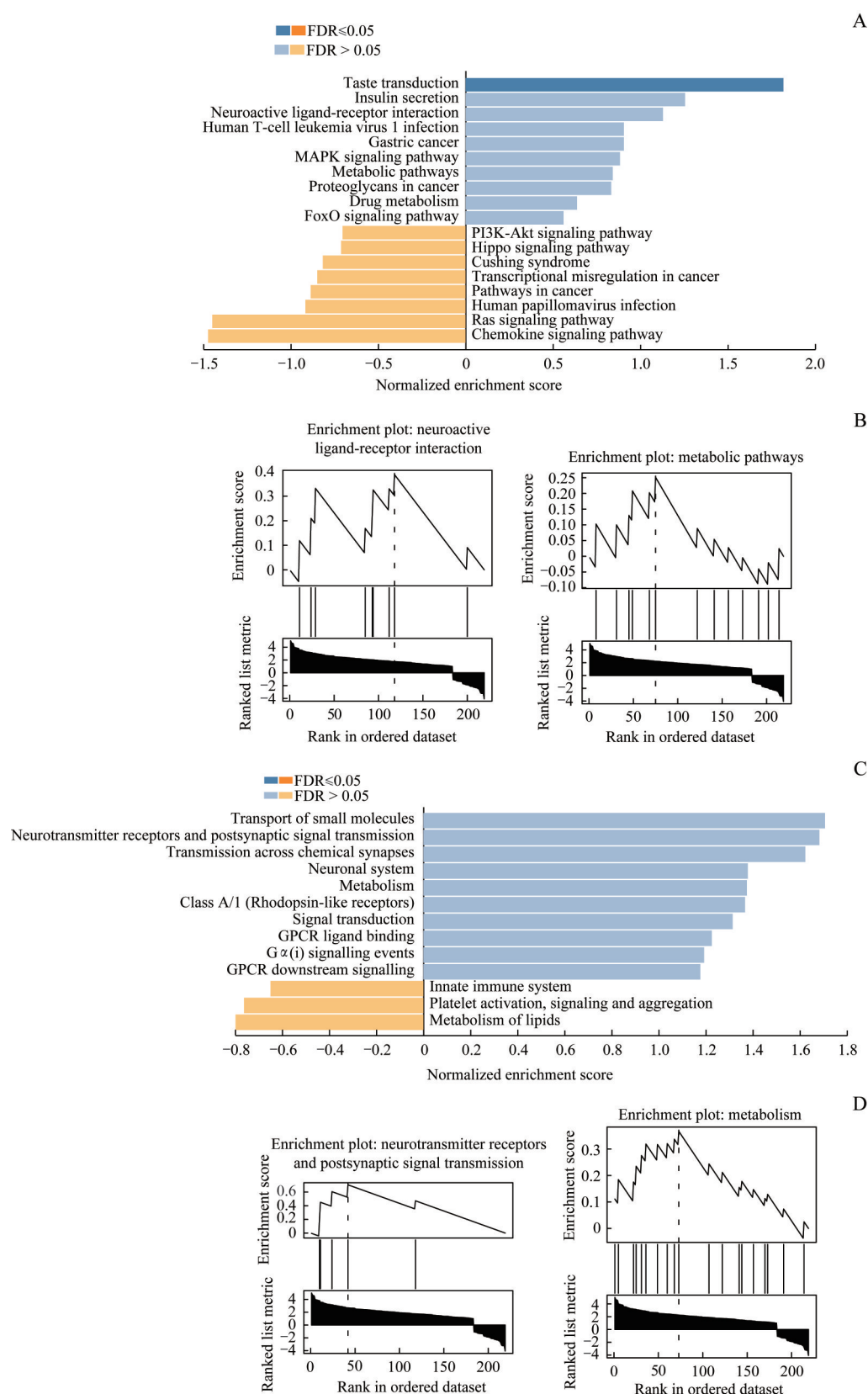
Fig 2 Relative BP, CC and MF of up-regulated and down-regulated genes

白3 (insulin like growth factor binding protein 3, *IGFBP3*)、精氨酸加压素受体2 (arginine vasopressin receptor 2, *AVPR2*)、分泌粒蛋白3 (secretogranin III, *SCG3*) (图4B)。

### 2.3 MSI-H结直肠癌基因转移预测模型的建立

将训练集 AUC=0.975, 验证集 AUC=0.920, C-index=0.832 (95%CI 0.798~0.866) 的预测模型使用 R 语言的 rms 包绘制可视化列线图, 即 MSI-H 结直肠癌基因转移预

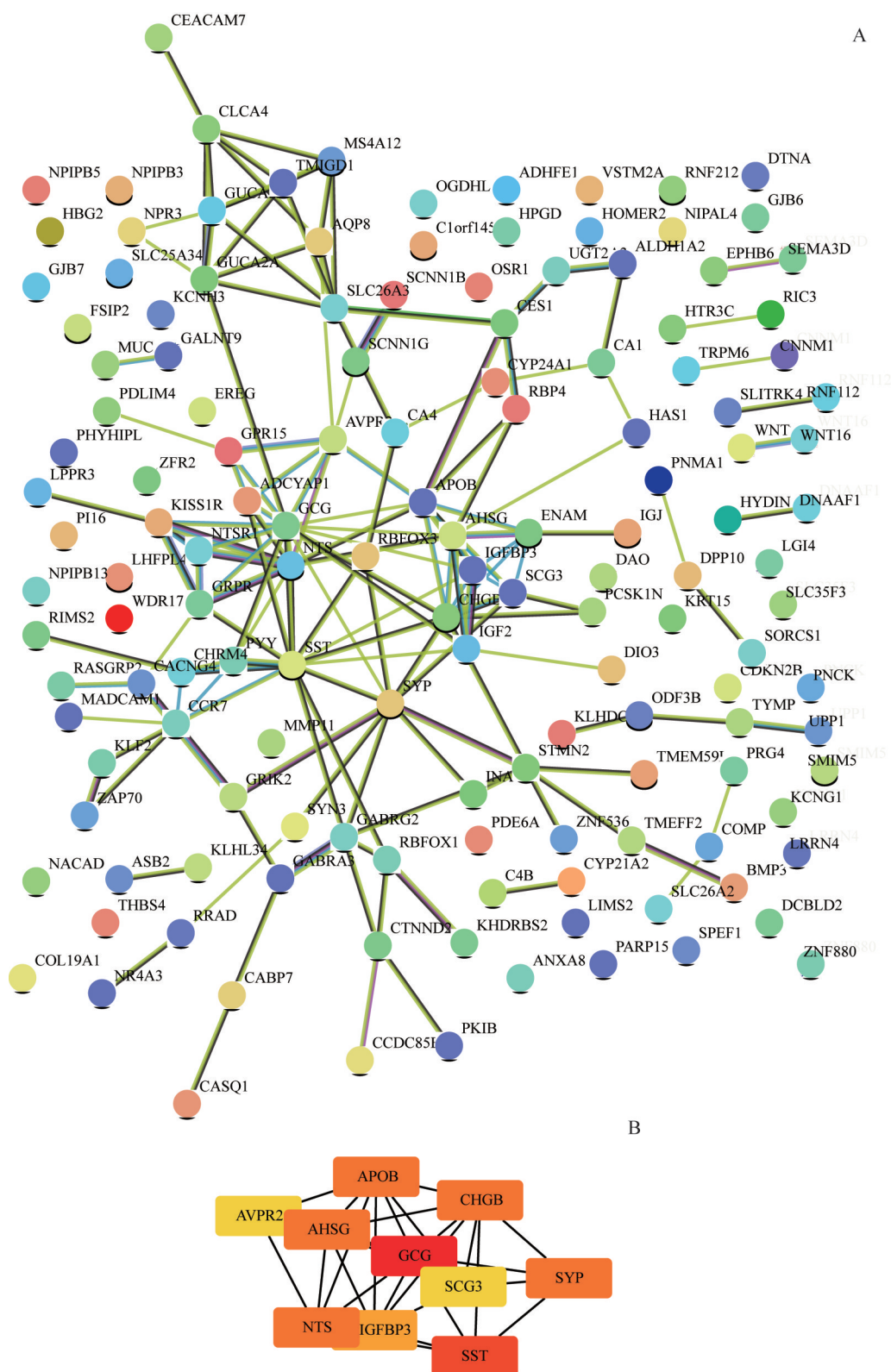




**Note:** A. The respective top 10 pathways of up-regulated genes and down-regulated genes were enriched by GSEA in KEGG database. B. Neuroactive ligand-receptor interaction and metabolic pathways were the same two pathways enriched in KEGG database and Reactome database. C. The respective top 10 pathways of up-regulated genes and down-regulated genes were enriched by GSEA in Reactome database. D. Neurotransmitter receptors and postsynaptic signal transmission and metabolism were the same two pathways enriched in Reactome database and KEGG database.

图3 DEG信号通路的GSEA

Fig 3 GSEA of DEG pathway



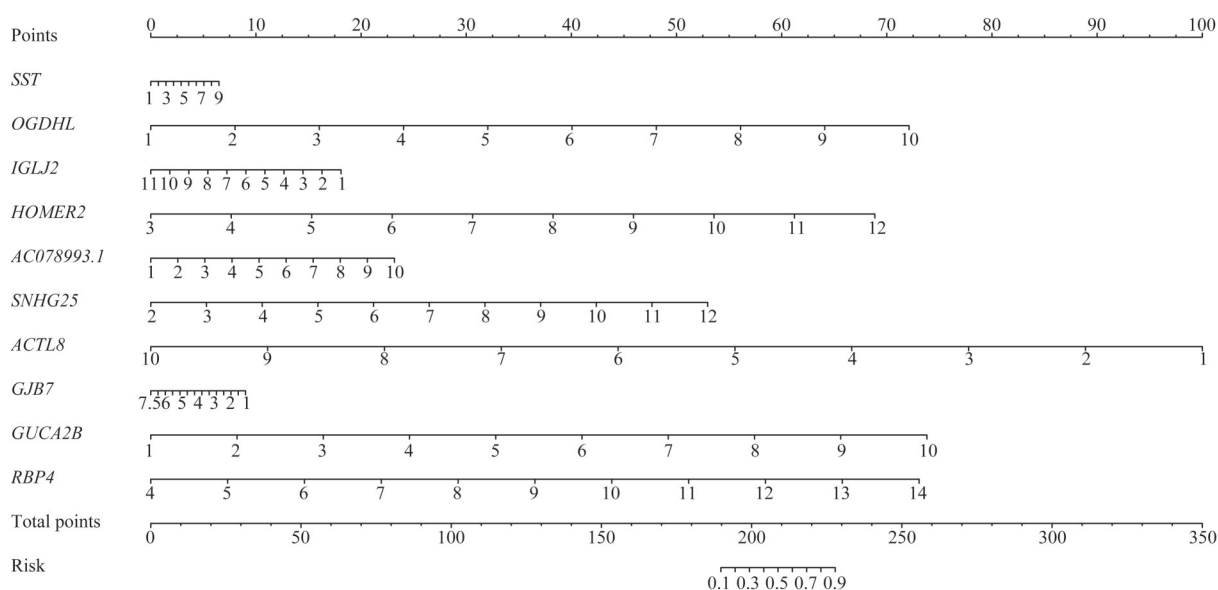
**Note:** A. The PPI network of the up-regulated genes was constructed by STRING. B. The top 10 hub genes were analysed by Cytoscape.

图4 上调基因PPI网络及等级前10位的hub基因

Fig 4 PPI network of the up-regulated genes and the top 10 hub genes

测列线图模型 (图5)。其中, 肌动蛋白8 (actin like 8, *ACTL8*)、鸟苷酸环化酶激活剂2B (guanylate cyclase activator 2B, *GUCA2B*)、L 氧化戊二酸脱氢酶

(oxoglutarate dehydrogenase L, *OGDHL*) 及视黄醇结合蛋白4 (retinol binding protein 4, *RBP4*) 对结直肠癌转移有较大影响。



**Note:** The value of  $\log_2|FC|$  of the gene corresponds to the relative value on the gene scale, and then corresponds to the position on “point” scale to get the relative score. The sum of all the scores corresponds to the relative value on “total points” scale, and then corresponds to the position on “risk” scale to get the relative metastatic risk value. *HOMER2*—homer scaffold protein 2; *SNHG25*—small nucleolar RNA host gene 25; *GJB7*—gap junction protein beta 7.

图5 MSI-H 结直肠癌转移风险列线图模型

Fig 5 Nomogram model of MSI-H colorectal cancer metastatic risk

## 2.4 转移预测模型中各基因对PFS的影响

将列线图中各基因表达水平对MSI-H结直肠癌PFS的影响进行生存分析(图6),发现:*AC078993.1*和*IGLJ2*的表达水平与MSI-H结直肠癌PFS呈明显负相关( $P=0.011$ ,  $P=0.005$ ),二者均为上调基因;其他基因的表达水平对MSI-H结直肠癌PFS影响无统计学意义(均 $P>0.05$ )。

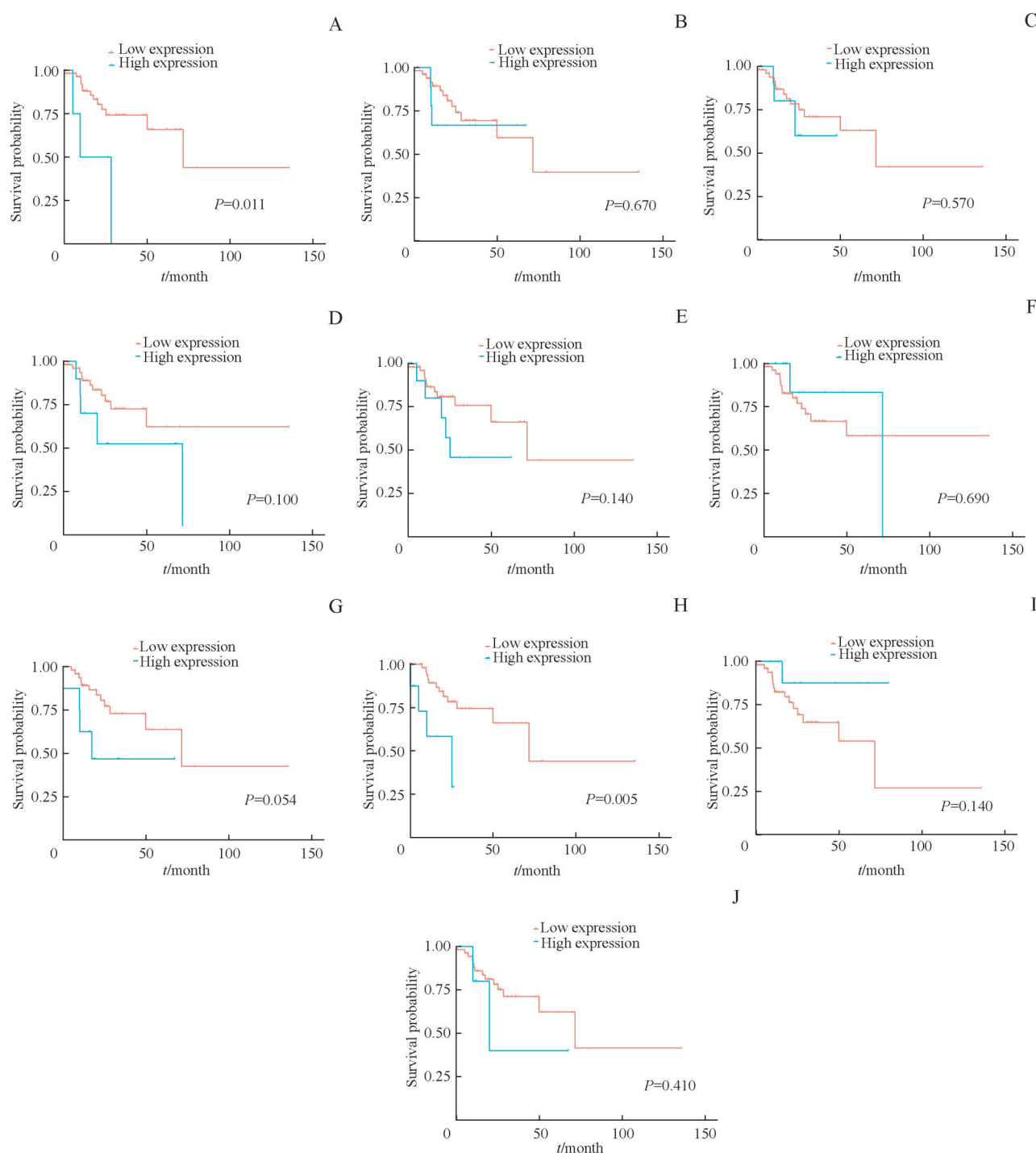
## 3 讨论

基因组不稳定性的产生是结直肠癌发展过程中一个重要特征,而MSI是造成基因组不稳定性的的重要途径之一<sup>[23]</sup>。MSI常常反映由于DNA错配修复缺陷(deficient mismatch repair, dMMR)导致的DNA复制错误<sup>[7-8]</sup>。在Ⅱ期和Ⅲ期结直肠癌中,MSI分别约占20%及12%,Ⅳ期中MSI占4%~5%<sup>[17-18]</sup>。虽然在Ⅱ期的散发性结直肠癌中,MSI为有益于预后的标志物<sup>[24]</sup>,但是在转移性结直肠癌中dMMR-MSI-H预后较差。这些现象提示,在MSI-H结直肠癌中可能存在一类高转移潜能亚型肿瘤,而目前尚无针对影响MSI-H结直肠癌转移转录组层面的研究。本研究希望从转录组层面探索影响MSI-H结直肠癌转移、产生高转移潜能的可能因素。

首先从TCGA数据库中筛选出63例MSI-H结直肠癌患者的转录组数据。通过生物信息学方法分析了MSI-H结直肠癌有转移和无转移患者间的DEG,共获得245个

DEGs。在转移组上调的前10位基因中,4次跨膜结构域A12(membrane spanning 4-domains A12, MS4A12)是结肠特异性的钙池调控钙离子内流通道蛋白,在结肠癌细胞中MS4A12蛋白表达降低会减弱癌细胞的增殖、运动性和趋化侵袭<sup>[25]</sup>。转移组下调的前10位基因中,*FGLI*编码的蛋白为纤维蛋白原样蛋白1(fibrinogen like 1, FGL1),属于纤维蛋白原家族成员。*FGLI*在丝氨酸/苏氨酸激酶11(serine/threonine kinase 11, *LKB1*)突变的肺腺癌中表达缺失会诱导上皮间质转化和血管形成<sup>[26]</sup>。在肝细胞癌中,*FGLI*的表达缺失与肝细胞癌的低分化表型有相关性,且*FGLI*通过丝氨酸-苏氨酸激酶依赖机制在肝细胞癌中起到抑癌作用<sup>[27]</sup>。肽基精氨酸脱亚胺酶3(peptidyl arginine deiminase 3, *PADI3*)编码的蛋白为肽基精氨酸脱亚胺酶家族成员,*PADI3*蛋白在结肠癌组织中低表达,并且*PADI3*可通过热休克蛋白90(heat shock protein, Hsp90)/细胞周期蛋白依赖性激酶调节亚基1(cyclin kinase subunit 1, *CKS1*)途径或沉默信息调节因子2(silent information regulator2, *Sirt2*)/蛋白激酶B(protein kinase B, *AKT*)/p21途径发挥抑癌作用<sup>[28-29]</sup>。基于*MS4A12*、*FGLI*和*PADI3*既往在各种肿瘤中的功能报道,三者可能为MSI-H结直肠癌转移过程中较关键的基因,其具体作用机制值得进一步探索。

进一步通过GO分析,在BP及MF富集中发现MSI-H结直肠癌转移组离子穿膜转运、氯离子穿膜转运及氯离子通道活性表现活跃。近年来,肿瘤组织中离子通道的



**Note:** The influence of the expression level of *AC078993.1* (A), *GJB7* (B), *HOMER2* (C), *OGDHL* (D), *SNHG25* (E), *ACTL8* (F), *GUCA2B* (G), *IGLJ2* (H), *RBP4* (I) and *SST* (J) on PFS in MSI-H CRC.

图6 生存分析列线图中10个基因对MSI-H结直肠癌PFS影响

Fig 6 Survival analysis of the 10 genes in Nomogram model on PFS of MSI-H colorectal cancer

改变越发受到重视。离子通道改变会导致细胞内离子稳态失调, 从而影响细胞体积调节、细胞膜电位变化、细胞的机械传导及肿瘤微环境等, 而这些改变都会对肿瘤细胞的增殖、凋亡、迁移、血管生成造成重要影响<sup>[30]</sup>。例如, 在肝细胞癌中, 氯离子选择性通道CIC-3通过调节

氯离子流调控细胞形态和体积从而促进细胞迁移<sup>[30]</sup>; 在胶质瘤细胞突起中, 氯离子通道CIC-3与基质金属蛋白酶2聚集, 共同调节胶质瘤细胞的迁移和侵袭<sup>[31]</sup>等。本研究分析得到的基因富集结果提示, 离子通道变化对MSI-H结直肠癌的转移的影响值得关注。



而在CC富集中发现, MSI-H结直肠癌转移组中细胞外部分、细胞外空间表现活跃。而肿瘤微环境是肿瘤细胞外的重要部分, 主要包括血管、与肿瘤相关的成纤维细胞、免疫细胞和免疫抑制细胞、信号分子、骨髓源性细胞、细胞外基质等<sup>[32]</sup>。我们得到的在MSI-H结直肠癌转移组中细胞外成分的富集结果可能提示了肿瘤微环境对MSI-H结直肠癌转移的发生和促进存在重要影响。

对DEGs信号通路的富集分析发现, 代谢通路在上调基因中得到富集。既往研究<sup>[33]</sup>也提示细胞代谢失调为肿瘤重要标志之一, 这说明在MSI-H结直肠癌转移过程中代谢通路改变也有着重要的意义。而既往报道神经系统异常, 包括神经递质、神经营养因子和其受体的异常, 在结肠癌肝转移中有重要作用<sup>[34]</sup>。我们上调基因中也富集到神经活性物质配体-受体相互作用通路, 说明该方向在MSI-H结直肠癌的转移研究中也值得关注。通过PPI网络及Cytoscape软件筛选出的hub基因中, NTS被报道其蛋白的表达水平与结肠癌的预后呈负相关, 并且NTS可以促进多种结肠癌细胞的生长<sup>[35]</sup>; 结肠癌组织中SYP蛋白表达水平与肿瘤预后呈负相关<sup>[36]</sup>; *IGFBP3*的基因水平与结肠癌的预后呈负相关<sup>[37]</sup>, 且IGFBP3可以促进肺腺癌的脑转移<sup>[38]</sup>。这些都提示发生转移的MSI-H结直肠癌相较于无转移肿瘤, 在转录组水平体现出更强的转移潜能。

通过对筛选出的DEGs进行列线图模型构建, 得到了预测MSI-H结直肠癌转移的基因列线图。该列线图有较好的区分度和一定的预测效能。既往有研究<sup>[39]</sup>从转录组层面分析结直肠癌的基因标签, 也有研究<sup>[40]</sup>提出一组可以用于预测结肠癌转移及复发的基因标签。但是运用这类基因标签进行评分, 其方式较复杂, 需要专业繁琐的统计学方法计算。本研究主要针对结直肠癌中的特定分子亚型——MSI-H结直肠癌, 特异性地构建了预测MSI-

H结直肠癌转移风险的基因表达预测模型。该模型较直观, 计算MSI-H结直肠癌发生转移的风险预测值较方便。随着基因测序技术的不断发展及在临床应用中的推广, 对于结肠癌术后行肿瘤组织转录组测序的患者, 可应用此类预测模型评估其发生转移的风险, 以协助术后治疗及随访方案的制定。将列线图中各基因表达水平对MSI-H结直肠癌PFS影响情况进行生存分析, 得到*AC078993.1*和*IGLJ2*与PFS呈负相关, 且两者均为MSI-H转移性结肠癌中上调基因。结合两者的表达水平及对PFS的影响, 提示其可能在MSI-H结直肠癌发生转移的过程中较重要。但其在肿瘤中均未有相关报道, 在之后进一步的研究中值得重点关注。

本列线图模型仅使用了单个数据库中的数据进行内部验证, 使结果有一定的局限性。而由于MSI结直肠癌在总结直肠癌中占10%~20%<sup>[7-8]</sup>, 且MSI结直肠癌转移发生率较MSS结直肠癌相对低等特点<sup>[12-15]</sup>, 在各数据库及数据集中可得到的发生转移的MSI结直肠癌病例数相对较少, 这使得之后的研究需进一步收集临床病例和其他数据库、数据集中的相关数据, 以加强验证及优化该列线图模型。并且本研究使用的数据无法区分MSI结直肠癌中林奇综合征患者, 之后的研究也需结合其他数据库信息, 区分出林奇综合征及错配修复蛋白甲基化导致的MSI-H结直肠, 进行更细致的探讨。

综上, 本研究从转录组层面探索了影响MSI-H结直肠癌产生高转移潜能的因素。通过对TCGA数据库中相关数据的生物信息学分析, 获得对MSI-H结直肠癌发生转移有影响的潜在关键基因; 并且构建了有一定预测效能的MSI-H结直肠癌转移基因预测模型。由于本研究局限于单一数据库的数据, 未来还需通过扩大数据的收集范围、收集临床样本及生物学实验等多个方面验证及完善本研究的结果。

## 参 · 考 · 文 · 献

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6): 394-424.
- [2] 郑荣寿, 孙可欣, 张思维, 等. 2015年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2019, 41(1): 19-28.
- [3] Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes[J]. Gut, 2015, 64(10): 1637-1649.
- [4] Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017[J]. CA Cancer J Clin, 2017, 67(3): 177-193.
- [5] Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates[J]. Cancer, 2010, 116(3): 544-573.
- [6] Sargent D, Sobrero A, Grothey A, et al. Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20, 898 patients on 18 randomized trials[J]. J Clin Oncol, 2009, 27(6): 872-877.
- [7] Copija A, Waniczek D, Witkoś A, et al. Clinical significance and prognostic relevance of microsatellite instability in sporadic colorectal cancer patients[J]. Int J Mol Sci, 2017, 18(1): E107.
- [8] Vilar E, Gruber SB. Microsatellite instability in colorectal cancer-the stable evidence[J]. Nat Rev Clin Oncol, 2010, 7(3): 153-162.
- [9] Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer[J]. Cancer Res, 1998, 58(22): 5248-5257.



- [10] Cohen R, Svrcek M, Dreyer C, et al. New therapeutic opportunities based on DNA mismatch repair and BRAF status in metastatic colorectal cancer[J]. *Curr Oncol Rep*, 2016, 18(3): 18.
- [11] Latham A, Srinivasan P, Kemel Y, et al. Microsatellite instability is associated with the presence of lynch syndrome pan-cancer[J]. *J Clin Oncol*, 2019, 37(4): 286-295.
- [12] Søreide K, Nedrebø BS, Søreide JA, et al. Lymph node harvest in colon cancer: influence of microsatellite instability and proximal tumor location[J]. *World J Surg*, 2009, 33(12): 2695-2703.
- [13] Buckowitz A, Knaebel HP, Benner A, et al. Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases[J]. *Br J Cancer*, 2005, 92(9): 1746-1753.
- [14] Malesci A, Laghi L, Bianchi P, et al. Reduced likelihood of metastases in patients with microsatellite-unstable colorectal cancer[J]. *Clin Cancer Res*, 2007, 13(13): 3831-3839.
- [15] Kim CG, Ahn JB, Jung M, et al. Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers[J]. *Br J Cancer*, 2016, 115(1): 25-33.
- [16] Venderbosch S, Nagtegaal ID, Maughan TS, et al. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies[J]. *Clin Cancer Res*, 2014, 20(20): 5322-5330.
- [17] Kawakami H, Zaanen A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer[J]. *Curr Treat Options Oncol*, 2015, 16(7): 30.
- [18] Colle R, Cohen R, Cochereau D, et al. Immunotherapy and patients treated for cancer with microsatellite instability[J]. *Bull Cancer*, 2017, 104(1): 42-51.
- [19] Liu Y, Sethi NS, Hinoue T, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas[J]. *Cancer Cell*, 2018, 33(4): 721-735. e8.
- [20] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*, 2010, 26(1): 139-140.
- [21] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Res*, 2003, 13(11): 2498-2504.
- [22] Wang SD, Yang L, Ci B, et al. Development and validation of a nomogram prognostic model for SCLC patients[J]. *J Thorac Oncol*, 2018, 13(9): 1338-1348.
- [23] Markowitz SD, Bertagnolli MM. Molecular origins of cancer: molecular basis of colorectal cancer[J]. *N Engl J Med*, 2009, 361(25): 2449-2460.
- [24] Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis[J]. *J Clin Oncol*, 2005, 23(3): 609-618.
- [25] Koslowski M, Türeci O, Huber C, et al. Selective activation of tumor growth-promoting  $\text{Ca}^{2+}$  channel MS4A12 in colon cancer by caudal type homeobox transcription factor CDX2[J]. *Mol Cancer*, 2009, 8: 77.
- [26] Bie FL, Wang GH, Qu X, et al. Loss of FGL1 induces epithelial-mesenchymal transition and angiogenesis in *LKB1* mutant lung adenocarcinoma[J]. *Int J Oncol*, 2019, 55(3): 697-707.
- [27] Nayeb-Hashemi H, Desai A, Demchev V, et al. Targeted disruption of fibrinogen like protein-1 accelerates hepatocellular carcinoma development[J]. *Biochem Biophys Res Commun*, 2015, 465(2): 167-173.
- [28] Chai ZB, Wang L, Zheng YB, et al. PADI3 plays an antitumor role via the Hsp90/CKS1 pathway in colon cancer[J]. *Cancer Cell Int*, 2019, 19: 277.
- [29] Chang XT, Chai ZB, Zou JR, et al. *PADI3* induces cell cycle arrest via the Sirt2/AKT/p21 pathway and acts as a tumor suppressor gene in colon cancer[J]. *Cancer Biol Med*, 2019, 16(4): 729-742.
- [30] Prevarskaya N, Skryma R, Shuba Y. Ion channels in cancer: are cancer hallmarks oncochannelopathies?[J]. *Physiol Rev*, 2018, 98(2): 559-621.
- [31] Lui VC, Lung SS, Pu JK, et al. Invasion of human glioma cells is regulated by multiple chloride channels including  $\text{ClC-3}$ [J]. *Anticancer Res*, 2010, 30(11): 4515-4524.
- [32] Siveen KS, Raza A, Ahmed EI, et al. The role of extracellular vesicles as modulators of the tumor microenvironment, metastasis and drug resistance in colorectal cancer[J]. *Cancers (Basel)*, 2019, 11(6): E746.
- [33] la Vecchia S, Sebastián C. Metabolic pathways regulating colorectal cancer initiation and progression[J]. *Semin Cell Dev Biol*, 2020, 98: 63-70.
- [34] Kasprzak A, Adamek A. The neuropeptide system and colorectal cancer liver metastases: mechanisms and management[J]. *Int J Mol Sci*, 2020, 21(10): 3494.
- [35] Qiu SY, Nikolaou S, Zhu J, et al. Characterisation of the expression of neurotensin and its receptors in human colorectal cancer and its clinical implications[J]. *Biomolecules*, 2020, 10(8): 1145.
- [36] Liu Y, He JJ, Xu JH, et al. Neuroendocrine differentiation is predictive of poor survival in patients with stage II colorectal cancer[J]. *Oncol Lett*, 2017, 13(4): 2230-2236.
- [37] Yamamoto N, Oshima T, Yoshihara K, et al. Clinicopathological significance and impact on outcomes of the gene expression levels of *IGF-1*, *IGF-2* and *IGF-1R*, *IGFBP-3* in patients with colorectal cancer: overexpression of the *IGFBP-3* gene is an effective predictor of outcomes in patients with colorectal cancer[J]. *Oncol Lett*, 2017, 13(5): 3958-3966.
- [38] Yang LS, Li JY, Fu SZ, et al. Up-regulation of insulin-like growth factor binding protein-3 is associated with brain metastasis in lung adenocarcinoma[J]. *Mol Cells*, 2019, 42(4): 321-332.
- [39] Loboda A, Nebozhyn MV, Watters JW, et al. EMT is the dominant program in human colon cancer[J]. *BMC Med Genomics*, 2011, 4: 9.
- [40] Schell MJ, Yang ML, Missiaglia E, et al. A composite gene expression signature optimizes prediction of colorectal cancer metastasis and outcome[J]. *Clin Cancer Res*, 2016, 22(3): 734-745.

[收稿日期] 2021-03-10

[本文编辑] 崔黎明