

DOI: 10.20135/j.issn.1006-8147.2024.03.0205

论著

基于多组学 *Lasso* 回归分析构建肝癌预测模型

罗焱瑞, 赵倩

(天津医科大学基础医学院细胞生物学系, 天津 300070)

摘要 目的: 通过生物信息学方法整合肝癌转录组和甲基化组数据, 利用 *Lasso* 回归分析筛选肝癌特异性标志物, 并构建肿瘤预测模型。方法: 在 GEO (gene expression omnibus) 中下载 GSE70091、GSE77314 数据集, 共 53 例肝癌患者的转录组测序 (RNA-seq) 和全基因组甲基化测序 (WGBS) 数据, 分别将肝癌与癌旁对照间的转录组、甲基化组数据进行差异分析, 并对差异表达基因 (DEG) 和差异甲基化基因 (DMG) 进行整合, 以筛选出候选肝癌标志基因。对候选肝癌标志基因进行 GO (gene ontology) 和 Reactome 通路富集分析, 使用 *Lasso* 回归分析筛选标志基因, 构建肝癌预测模型, 并在其他队列中进行性能验证。结果: 共筛选出 288 个 DEG ($|\log_2\text{FC}| > 1, P_{\text{adj}} < 0.05$) 和 28 528 个 DMG ($P < 0.05$), 通过 DEG 和 DMG 的交叉分析找到 51 个高甲基化下调 (Hyper-Down) 基因和 111 个低甲基化上调 (Hypo-Up) 基因。GO 和 Reactome 通路富集分析显示, Hypo-Up 基因主要富集在细胞有丝分裂通路 ($FDR < 0.05, P < 0.05$), Hyper-Down 基因主要与转录激活的功能相关 ($FDR < 0.05, P < 0.05$)。使用 *Lasso* 回归分析筛选出 11 个具有非零系数的基因并构建肝癌预测模型。最后在 GSE77314、外部验证队列 (TCGA-LIHC) 中验证出模型曲线下面积 (AUC) 分别为 1、0.998。结论: 通过整合肝癌多组学数据, 使用 *Lasso* 回归分析筛选出 11 个基因标志物, 并构建肝癌预测模型。

关键词 肝癌; 差异表达基因; 差异甲基化基因; *Lasso* 回归分析

中图分类号 R735.7

文献标志码 A

文章编号 1006-8147(2024)03-0205-08

A liver cancer prediction model based on multi-omics *Lasso* regression analysis

LUO Yanrui, ZHAO Qian

(Department of Cell Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China)

Abstract **Objective:** To integrate hepatocellular carcinoma transcriptomic and methylationomic data by bioinformatics methods, screen hepatocellular carcinoma-specific markers and construct tumor prediction models using *Lasso* regression analysis. **Methods:** The transcriptome sequencing (RNA-seq) and whole genome methylation sequencing (WGBS) data of a total of 53 hepatocellular carcinoma patients from the GSE70091 and GSE77314 datasets were downloaded from the GEO (gene expression omnibus) database, and the transcriptome and methylation data between hepatocellular carcinomas and adjacent normal tissues were analyzed for differences, respectively. Differentially expressed genes (DEGs) and differentially methylated genes (DMGs) were integrated to screen out hepatocellular carcinoma candidate marker genes. The GO (gene ontology) and Reactome pathway enrichment analyses were performed on the candidate marker genes, and *Lasso* regression analysis was used to screen the marker genes and construct the liver cancer prediction model, and the performance was validated in other cohorts. **Results:** Totally, 288 DEGs ($|\log_2\text{FC}| > 1, P_{\text{adj}} < 0.05$) and 28 528 DMGs ($P < 0.05$) were screened out, and 51 Hyper-Down genes and 111 Hypo-Up genes were identified by cross-analysis of DEG and DMG. GO and Reactome results showed that Hypo-Up genes were enriched in the cellular mitotic pathway ($FDR < 0.05, P < 0.05$) and Hyper-Down genes were mainly associated with transcriptional activation functions ($FDR < 0.05, P < 0.05$). A liver cancer prediction model was created by screening 11 genes with non-zero coefficients using *Lasso* regression analysis. Finally, in GSE77314 and TCGA-LIHC cohorts, the model's area under curve (AUC) was validated as 1 and 0.998, respectively. **Conclusion:** A total of 11 gene markers and construct a hepatocellular carcinoma prediction model by are screened integrating the multi-omics data of hepatocellular carcinoma and using *Lasso* regression analysis.

Key words liver cancer; differentially expressed genes; differentially methylated genes; *Lasso* regression analysis

肝癌是全球范围内最常见的致命恶性肿瘤之一, 肝癌分别位居我国恶性肿瘤发病率、死亡率的第四、第二^[1]。对高危人群进行肝癌筛查有助于早期发现、早期诊断和早期治疗, 对于改善肝癌患者的

预后至关重要^[2]。

甲胎蛋白 (alpha-fetoprotein, AFP) 是目前应用最广泛的肝癌肿瘤标志物, 有研究表明, 以 20 ng/mL 为临界值, AFP 的灵敏度为 63%, 特异性为 88.7%, 曲线下面积 (AUC) 为 0.829, 其敏感性和特异性仍待提高^[3-6]。此外, 现有的标志物往往都是基于基因

作者简介 罗焱瑞 (1994-), 男, 硕士, 研究方向: 计算基因组学; 通信作者: 赵倩, E-mail: zhaoq@tmu.edu.cn。

组、转录组或者蛋白质组等单个组学的数据筛选而来,很少有研究利用多个组学的数据整合分析来筛选标志物,这些标志物不能从系统生物学角度体现其特异性。

本研究拟利用配对样本的全基因组甲基化(whole genome bisulfite sequencing, WGBS)和转录组测序(RNA-sequencing, RNA-seq)数据来联合筛选受甲基化调控的差异表达基因(differentially expressed gene, DEG),搭建了全新的分析平台,并使用Lasso回归分析最终选出基因标志物,构建新的肝癌预测模型,为肝癌的早期筛查提供新思路。

1 材料与方法

1.1 数据来源 从GEO(gene expression omnibus)收集来自GSE70091和GSE77314共53例肝癌患者的高通量测序数据^[7-8],由于这两个数据集包含患者肿瘤、正常组织的RNA-seq数据,且GSE70091包含了WGBS的数据,因此能够同时从甲基化修饰和转录调控反映患者的分子特征,有助于肿瘤标志物的筛选。本研究从GSE77314中随机筛选出50%的样本作为训练集,其余50%的样本作为内部验证集,同时从TCGA(the cancer genome atlas)数据库下载外部验证队列(TCGA-LIHC)的转录组数据用于模型外部验证^[9]。

1.2 方法 从发现队列(GSE70091)的RNA-seq、WGBS数据中筛选DEG和差异甲基化基因(differentially methylated gene, DMG)进行整合,对整合筛选基因(DEG & DMG)进行GO(gene ontology)与Reactome富集分析;并在GSE77314的RNA-seq训练数据集(Train)中使用Lasso回归分析,进一步筛选肝癌特异性标志基因,同时构建预测模型;最后在GSE77314的RNA-seq验证数据集(Test)和TCGA-LIHC中进行验证(图1)。

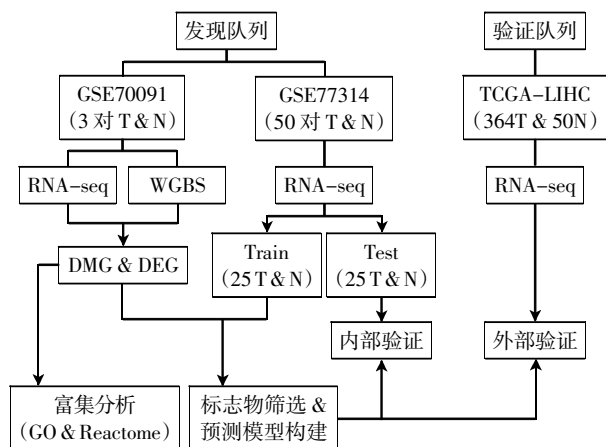


图1 研究流程图

Fig.1 The flowchart of the study

1.2.1 RNA-seq 数据处理 利用trim_galore(版本0.6.6)软件从原始RNA-seq数据中去除接头和低质量的序列,使用hisat2(版本2.2.1)将序列与人类hg38参考基因组进行比对^[10],最后使用stringtie(版本2.1.5)进行基因表达定量^[11]。

1.2.2 DEG 的获取 使用R软件(版本4.1.2)中的DESeq2软件包(版本1.38.3),在肝癌与癌旁的基因表达数据间(RNA-seq),以 $|\log_2(\text{FC, fold change})| > 1$ 和Bonferroni校正 $P < 0.05$ 为标准,筛选DEGs。

1.2.3 富集分析 使用R软件的clusterProfiler(版本4.8.2)包进行GO功能富集和Reactome通路富集分析,以实现检测与关键差异基因相关的生物功能和潜在的信号通路途径^[12-13]。GO和Reactome分析采用 $FDR < 0.05$ 为筛选标准。

1.2.4 WGBS 数据处理 使用trim_galore从原始WGBS测序数据中去除接头和低质量序列,然后使用BisMark(版本0.23.0)将序列与人类参考基因组(hg38)进行比对,生成BAM(binary alignment map)文件,同时使用bismark_methylation_extractor模块提取每个CpG位点的甲基化状态^[14]。每个CpG位点甲基化水平的Beta值由甲基化和未甲基化信号的比率计算。用基因转录起始位点(transcription start site, TSS)上下游1 kb(TSS \pm 1k)区间的CpG平均甲基化值来表示基因的甲基化水平。基因组按照起始位置划分为200 kb的连续区间,用来评估基因组局部位置的平均甲基化水平。

1.2.5 WGBS 差异甲基化区间获取 使用DSS软件包(版本2.47.1)从肝癌、癌旁的BAM文件中寻找差异甲基化CpG位点,并将相邻的差异甲基化CpG位点合并得到差异甲基化区域(differentially methylated regions, DMRs)^[15]。两组间的平均甲基化差值的绝对值 $\Delta > 0.1$, $P < 0.05$ 为差异有统计学意义。最后使用R软件的ChIPseeker(版本1.30.3)、annotatr包(版本1.20.0)和GeneHancer数据库(版本4.7)分别对DMR所在基因区间、CpG岛和增强子进行注释,并获取DMGs。DMG是根据其启动子甲基化状态来定义的,如果一个基因的启动子区域存在DMR,则该基因被视为DMG^[16-18]。

1.2.6 基因标志物筛选 整合基因的甲基化和表达状态,进一步确定差异基因,根据DMGs和DEGs之间的交集,将基因划分为高甲基化上调(Hyper-Up)、高甲基化下调(Hyper-Down)、低甲基化上调(Hypo-Up)和低甲基化下调(Hypo-Down)基因。启动子甲基化导致肿瘤抑制基因失活,而启动子低甲基化导致肿瘤基因的表达激活,这是肿瘤发生的重

要机制,因此挑选 Hyper-Down 和 Hypo-Up 基因为候选基因标志物。

1.2.7 Lasso 回归分析 Lasso 回归可以自动识别并舍弃无关或多余的变量,这一特点使得 Lasso 特别适用于特征选择^[19]。本研究使用 Lasso 回归方法,从候选基因标志物中筛选出标志基因,并构建 Lasso 回归模型用于样本预测。由于 Lasso 依赖于调节参数 λ ,本研究使用交叉验证的方法挑选出最优调节参数 λ 。

对于一般线性回归模型:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

其中, y 被称为因变量, x_1, x_2, \cdots, x_p 为自变量。在对模型的系数进行估计时, Lasso 的基本思想是将 L1 惩罚项施加在最小二乘法上,从而达到变量选择的目的,并且去除模型中不显著的变量,其表达式如下:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^d x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^d |\beta_j| \right\}$$

其中 $\hat{\beta}$ 为自变量的系数估计值, d 为回归模型中自变量的个数。 λ 为 Lasso 方法的调节参数,它起到对变量系数的压缩程度进行控制的作用,不同的 λ 可以得到不同的模型。

1.2.8 模型效能评估 应用受试者工作特征(ROC)曲线对模型进行评价。其中 AUC 值反映了模型预测的真实性,当 AUC>0.5 时,模型具有实际应用价值, AUC 值越接近 1.0, 预测模型的真实性越高。模型外部验证队列使用 TCGA-LIHC^[20]。

1.3 统计学处理 TCGA-LIHC 队列肿瘤和正常组

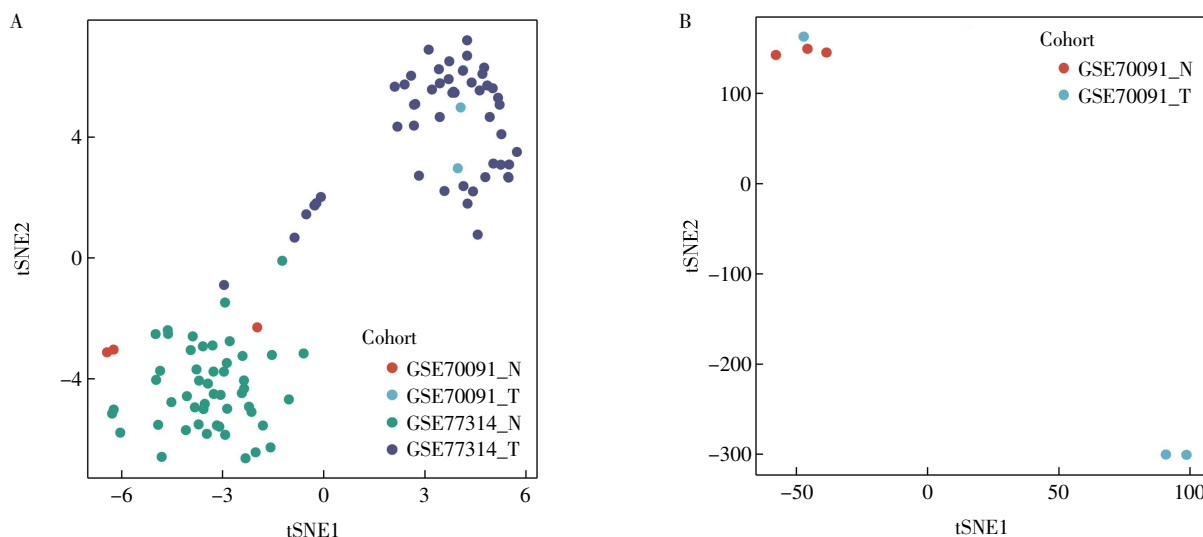
织表达水平比较采用秩和检验方法检验。样本相关性采用 t -SNE 方法进行分析。 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 样本相关性及差异表达分析 分别将 GSE-70091、GSE77314 数据集的肝癌和癌旁对照的 RNA-seq 数据进行合并,并分别对 RNA-seq 和 WGBS 数据进行相关性分析(图 2)。将肝癌和癌旁的基因进行差异表达分析,共筛选出 DEG 288 个,包含上调基因 212 个,下调差异基因 76 个(图 3A),并对 DEG 进行聚类分析和绘制热图(图 3B)。

2.2 差异甲基化区域筛选 差异甲基化分析共筛选出 65 734 个高甲基化(hyper-methylated)的区域(图 4B), 921 306 个低甲基化(hypo-methylated)的区域(图 4C)。通过注释分析,差异甲基化区域被注释到 28 528 个基因的 TSS +/-1k 上,其中高甲基化的基因 13 710 个,低甲基化基因 14 818 个。这些差异甲基化区域主要存在于基因间区和基因体上,少部分在增强子、基因启动子和 CpG 岛上(图 4D)。其中高甲基化的区域主要发生在基因的启动子和启动子附近的 CpG 岛上(约 42.9%,图 4B),而低甲基化的区域主要发生在基因间区和基因体上(约 80.4%,图 4C)。相对于正常组织,肿瘤在全基因组范围内呈现 DNA 低甲基化的改变(图 4A)。

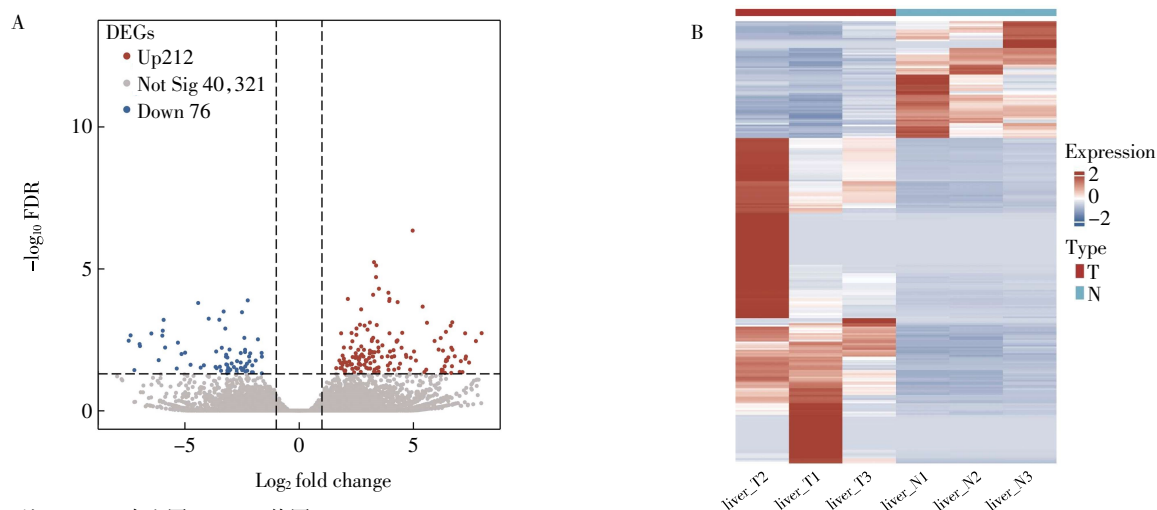
2.3 DMGs 和 DEGs 的整合分析 通过 DMGs 和 DEGs 的交叉分析(图 5A),发现 92 个 Hyper-Up 基因、51 个 Hyper-Down 基因、111 个 Hypo-Up 基因和 32 个 Hypo-Down 基因(图 5B)。



注:A:基于转录组数据比较样本相关性;B:基于甲基化组数据比较样本相关性

图2 在转录组和甲基化组中比较样本间的相关性

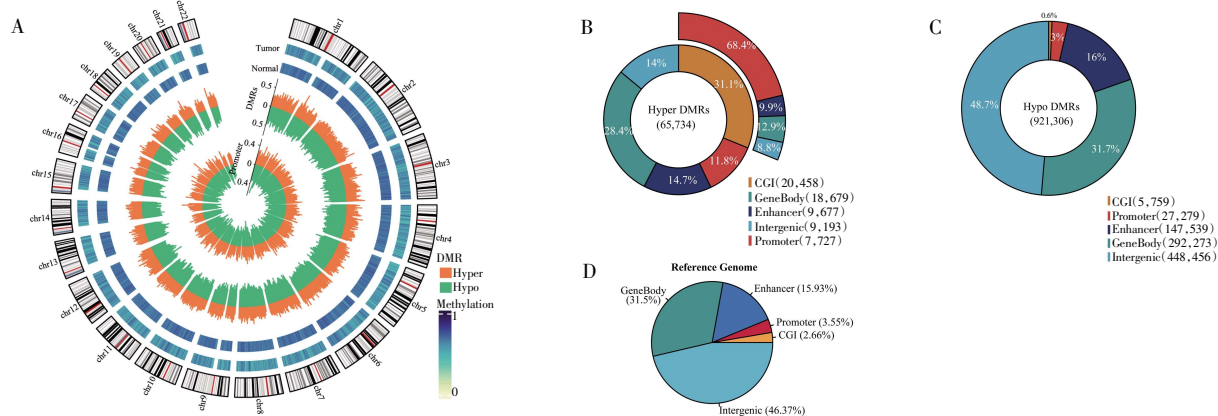
Fig.2 Correlations comparison of samples in the transcriptome and methylationome



注:A:DEG 火山图;B:DEG 热图

图 3 差异基因表达分析

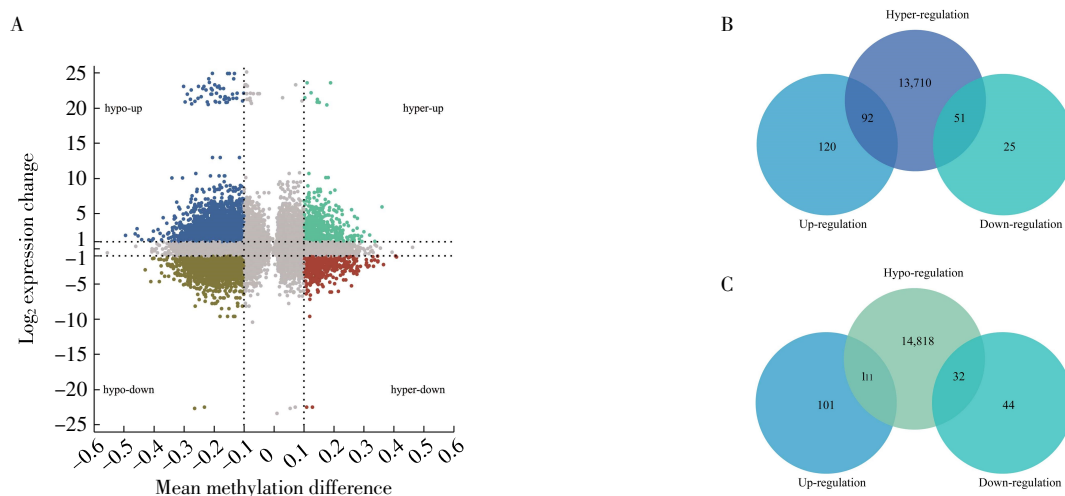
Fig.3 Differential gene expression analysis



注:A:差异甲基化环状图,从外往内依次为染色体、肿瘤平均甲基化水平、正常组织平均甲基化水平、DMR 区间内肿瘤与正常组织的甲基化水平差值(橙色为肿瘤高甲基化,绿色为肿瘤低甲基化)、注释到启动子上的 DMR 区间内肿瘤与正常组织的甲基化水平差值(橙色为肿瘤高甲基化,绿色为肿瘤低甲基化);B:高甲基化 DMR 的注释环状图,内圈为各部分在基因组中占比,外圈为 CGI(CpG island)上的 DMR 在启动子、增强子、基因体和基因间区的比例;C:低甲基化 DMR 的注释环状图;D:不同基因结构位置的 DMR 占比

图 4 差异甲基化区域筛选及注释

Fig.4 Differential methylation region detection and annotation



注:A:DMR 的平均甲基化差值与 \log_2 表达变化的散点图,每个点代表一个 DMR-Gene 对;B:韦恩图展示高甲基化基因与 DEGs 之间的交集;C:韦恩图展示低甲基化基因与 DEGs 之间交集

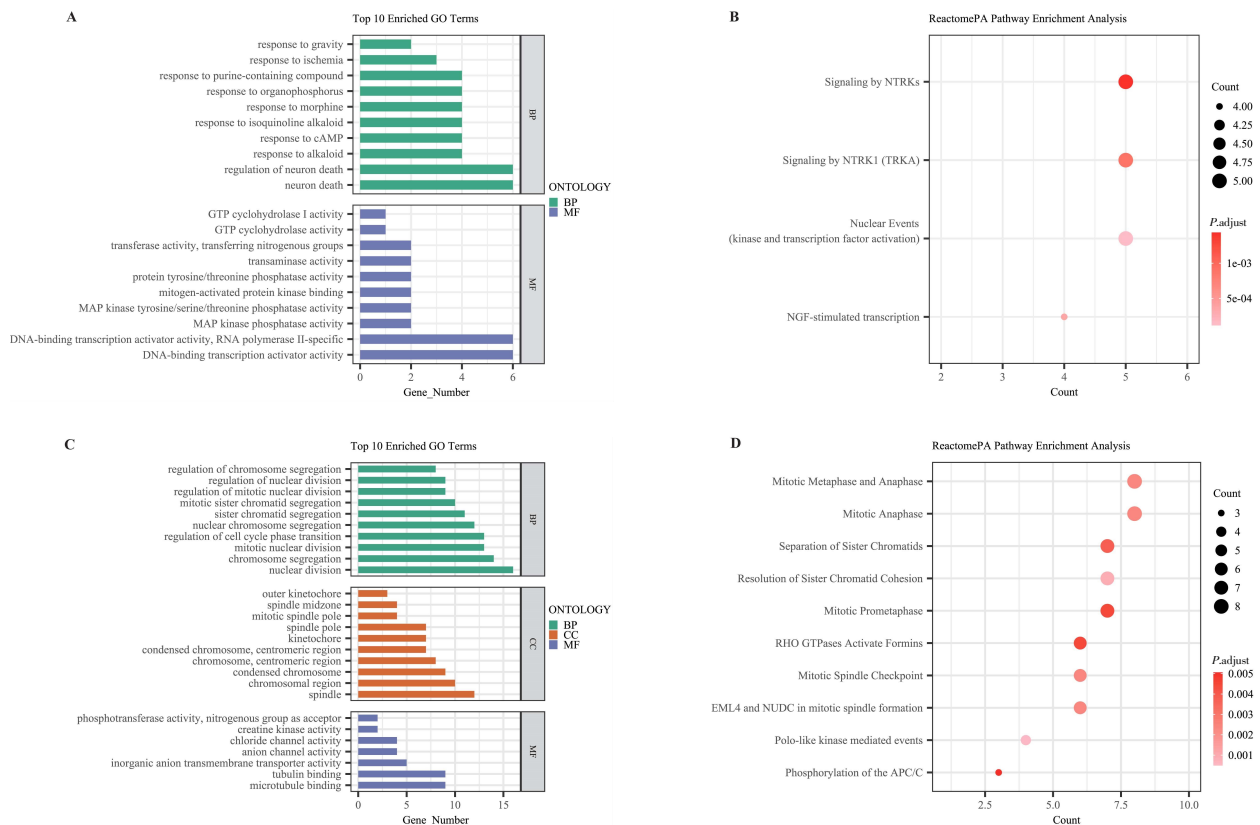
图 5 DMG 和 DEG 整合

Fig.5 Integration of DMG and DEG

2.4 Hyper-Down 和 Hypo-Up 基因富集分析 将差异甲基化和差异表达共同筛选的基因(DMG & DEG)进行 GO 富集分析,结果显示 Hyper-Down 的基因主要富集在对 cAMP 的反应、神经元死亡调控等分子功能(BP)上及 RNA 聚合酶 II 特异性 DNA 结合型转录激活剂活性、DNA 结合型转录激活剂活性等分子功能(MF)上(图 6A)。Hypo-Up 基因主要富集在有丝分裂核分裂、染色体分离的调控等生物学过程(BP)上,细胞组分(CC)主要富集在凝聚的染色体着丝粒区域、有丝分裂纺锤极等组分上,分子功能(MF)主要富集在离子通道活性、肌酸激酶活性等功能上(图 6C)。Reactome 通路富集分析发现 Hyper-Down 的基因主要富集在 NTRK1 (TRKA) 的信号传递、刺激的 NGF 转录等通路上(图 6B);Hypo-Up 基因主要富集在 Polo 样激酶介导的事件、有

丝分裂期等通路上(图 6D)。

2.5 Lasso 回归分析及模型构建 使用 Hyper-Down 和 Hypo-Up 基因作为自变量,通过 Lasso 算法在不同水平 λ 分别对自变量的系数进行压缩,经过不断调整优化模型,当 $\lambda = 0.2$ 时,筛选出 11 个具有非零系数的基因(图 7A),即 *AADAT*、*ARPP21*、*ASPM*、*DUSP1*、*E2F8*、*ESM1*、*FBXO43*、*FOS*、*GADD45B*、*SCN4A* 和 *TERT*, 其 Lasso 回归系数分别为 -1.268 、 -0.097 、 0.407 、 -0.018 、 1.644 、 0.676 、 0.258 、 -0.086 、 -0.551 、 2.520 和 0.597 。对模型中特征基因的重要性值进行了排序,前 5 位最重要的特征基因包括 *ESM1*、*ASPM*、*AADAT*、*E2F8* 和 *TERT*(图 7B)。GSE7734 和 TCGA-LIHC 两个队列中,这 11 个基因的表达量在肿瘤、正常组织间都存在显著差异(图 7C、7D)。



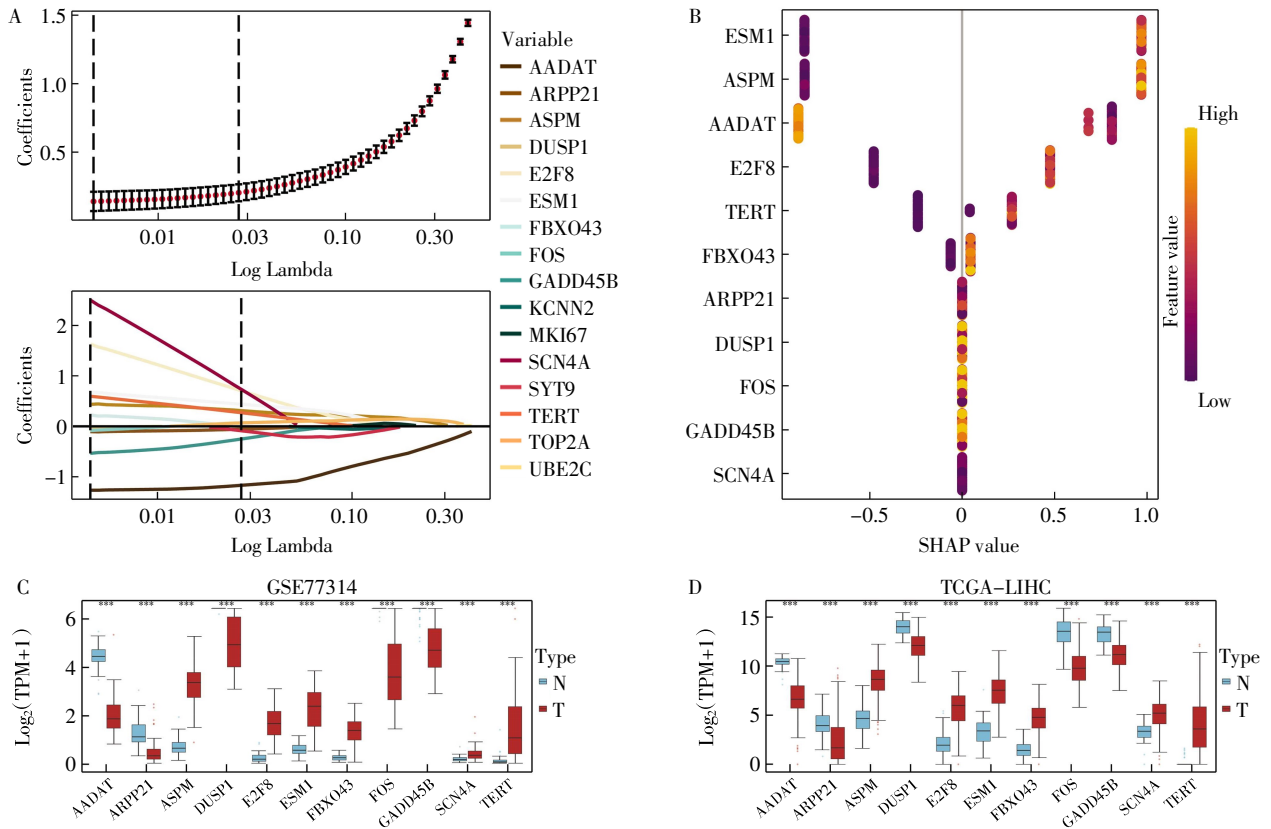
注:A:Hyper-Down 基因的 GO 注释;B:Hyper-Down 基因的 KEGG 注释;C:Hypo-Up 基因的 GO 注释;D:Hypo-Up 基因的 KEGG 注释

图 6 DMG & DEG 富集分析

Fig.6 Enrichment analysis of DMG and DEG

2.6 模型验证 使用 GSE77314_Test 和 TCGA-LIHC 队列对模型的性能进行了验证,在验证集中,模型能有效区分肿瘤和正常组织($P < 0.001$,图 8A)。其 AUC 在两个队列中分别为 1 和 0.998(图 8B)、灵敏

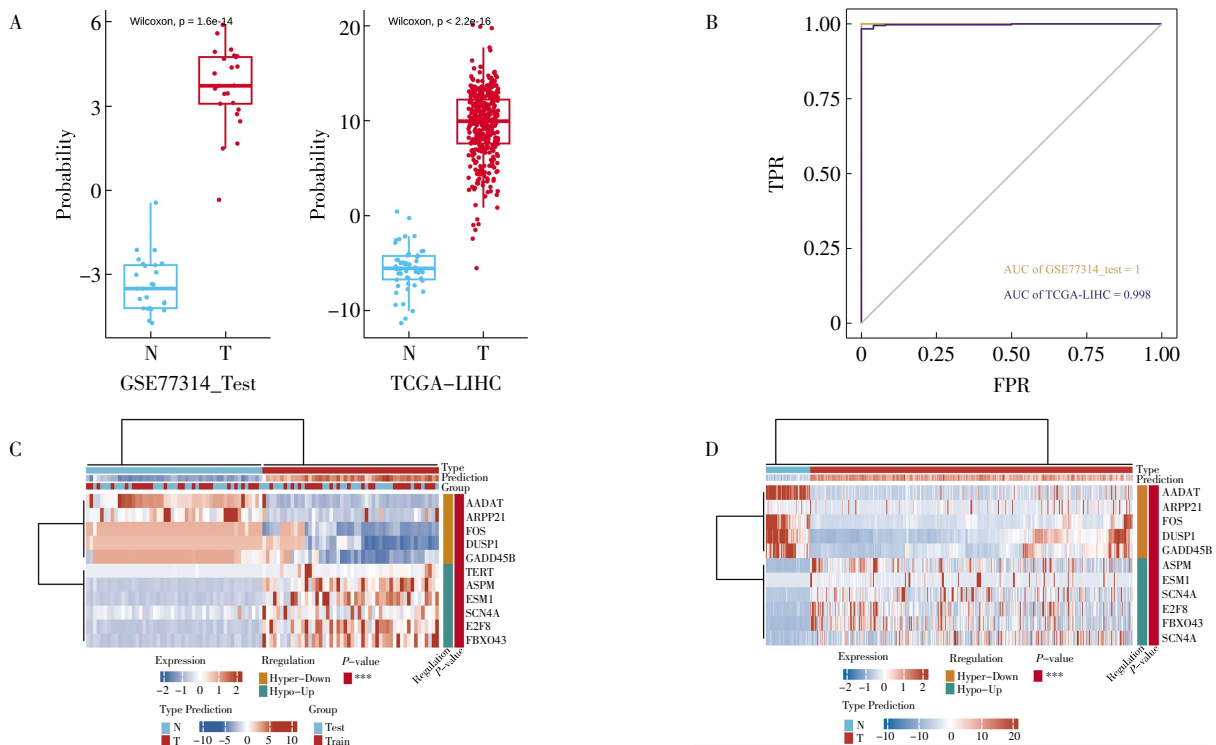
度分别为 1 和 0.978、特异性为 1 和 0.991。表达热图也显示出 11 个标志基因的表达水平均存在显著差异($P < 0.05$,图 8C、8D),其中 5 个 Hyper-Down、6 个 Hypo-Up 基因。



注:A:Lasso 回归模型中的最佳惩罚系数 λ 的变化过程和 Lasso 模型回归系数轨迹;B: 筛选变量的重要性排序;C: 使用秩和检验对 GSE77314 队列中 T 和 N 之间的表达量进行比较;D:使用秩和检验对 TCGA-LIHC 队列中 T 和 N 之间的表达量进行比较;*** $P < 0.001$

图 7 利用 Lasso 构建肝癌预测模型

Fig.7 Liver cancer prediction model constructed by Lasso



注:A:箱线图展示 Lasso 模型的预测效率;B:曲线下面积(AUC)分析,描述 Lasso 模型的预测效率;C:热图展示 GSE77314 队列样本的表达图谱, P -value 表示 T 和 N 之间基因表达量的秩和检验结果;D:热图展示 TCGA-LIHC 队列样本的表达图谱, P -value 表示 T 和 N 之间的秩和检验结果;*** P -value < 0.001

图 8 使用其他队列验证预测模型

Fig.8 Validation of the prediction model by using additional cohorts

3 讨论

肝癌是全球最常见、最致命的肿瘤之一,研究者一直在致力于探索和改进肝癌的筛查、诊断和治疗策略,以改善患者的预后。因此,筛选监测肝癌的生物标志物、构建有效预测模型是非常重要的研究内容。目前缺乏用于监测和早期诊断肝癌的生物标志物,现有血清生物标志物的截断点不同,特异性也不尽相同^[3,21],且这些研究往往都是基于单个组学数据来筛选标志物,因此这些标志物不能够从系统生物学角度体现其特异性。

本研究从整合肝癌及其配对正常组织的 WGBS 和 RNA-seq 的数据、基因表达调控修饰和转录水平角度来筛选肝癌特异性的标志基因。大量研究表明,抑癌基因启动子的甲基化会导致抑癌基因失活以及致癌基因的异常低甲基化,导致癌基因表达激活,启动子异常甲基化是肿瘤发生、发展的一个重要机制^[22-24]。因此在下游标志物的鉴定过程中,将重点放在了51个 Hyper-Down 和111 个 Hypo-Up 基因上。

对 Hyper-Down 基因的 GO 分析显示,这些基因与 RNA 聚合酶 II 转录激活功能相关。转录调节因子通过与顺式调节区域内特定双链 DNA 序列的选择性非共价结合,调节基因组的转录。肿瘤中 RNA 聚合酶 II 对基因转录调控的变化,是几乎所有复杂人类疾病的根源,这种变化在肿瘤中得到了最广泛的记录,肿瘤特有的基因表达特征是确定新型靶向疗法的重要方式^[25]。Hypo-Up 基因 GO 主要富集在有丝分裂、染色体分离的调控等功能上,表明肝癌细胞的细胞周期和细胞增殖异常,这一结果与许多肿瘤的特征相似^[26]。Reactome 通路富集结果显示 Hypo-up 基因主要富集在 Polo 样激酶介导的有丝分裂期通路上。这些基因在有丝分裂的启动、维持和完成过程中发挥着重要作用,而基因的异常低甲基化可导致基因的异常高表达,可能会促进肿瘤转化,并推动肿瘤进展进程,在多种人类肿瘤中均发现 PLK1、FOX M1 等基因的过表达,并且与肿瘤的不良预后有关^[27-28]。

为了进一步从 Hyper-Down 和 Hypo-Up 的基因中筛选出与肝癌相关的标志基因,本研究利用 Lasso 回归分析从 162 个候选基因中筛选出 11 个肝癌特异性的标志基因,并构建预测模型。为了验证筛选的 11 个标志基因是否能够区分肝癌和正常组织,同时为了验证构建的肝癌预测模型的性能,本研究使用 GSE77314_Test 和 TCGA-LIHC 两个队列的 RNA-seq 数据进行验证,发现这 11 个基因的表达量在肿瘤和正常组织之间都呈现显著差异,并

且 AUC 值分别为 1 和 0.998,证明预测模型有着较为优异的预测性能。因此,后续工作将会基于筛选的 11 个标志基因,在肝癌患者外周血 ctDNA 的检测研究中做进一步验证和筛选,同时进一步优化模型以提升预测性能,期望在肝癌诊断领域能获取更有价值的结果。

本研究存在一定的局限性,由于同时具有 WGBS、RNA-seq 数据的肿瘤和配对正常组织的样本较少,无法细分肝癌中不同亚型和发现不同亚型之间的异质性特征。未来的研究可以通过扩大样本量来解决这些问题。

综上所述,本研究整合了肝癌患者的甲基化组和转录组数据,利用 Lasso 回归分析筛选出 11 个肝癌基因标志物,并构建了新的肝癌预测模型,在 GSE77314 和 TCGA-LIHC 队列中验证出模型有着优异的预测性能。这些基因标志物以及构建的预测模型可能会为肝癌的早期筛查提供一定的提示。

参考文献:

- [1] XIA C,DONG X,LI H,et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants[J]. Chin Med J (Engl), 2022, 135(5):584-590.
- [2] LOUD J T,MURPHY J. Cancer screening and early detection in the 21(st) century[J]. Semin Oncol Nurs, 2017, 33(2): 121-128.
- [3] LEE C W, TSAI H I, LEE W C, et al. Normal alpha-fetoprotein hepatocellular carcinoma: are they really normal[J]. J Clin Med, 2019, 8(10):1736-1751.
- [4] BOUATTOUT M, MEHTA N, HE A R, et al. Systemic treatment for advanced hepatocellular carcinoma[J]. Liver Cancer, 2019, 8(5): 341-358.
- [5] BRUIX J, GORES G J, MAZZAFERRO V. Hepatocellular carcinoma: clinical frontiers and perspectives[J]. Gut, 2014, 63(5):844-855.
- [6] WORLAND T, HARRISON B, DELMENICO L, et al. Hepatocellular carcinoma screening utilising serum alpha-fetoprotein measurement and abdominal ultrasound is more effective than ultrasound alone in patients with non-viral cirrhosis[J]. J Gastrointest Cancer, 2018, 49(4): 476-480.
- [7] LI X, LIU Y, SALZ T, et al. Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver[J]. Genome Res, 2016, 26(12): 1730-1741.
- [8] LIU G, HOU G, LI L, et al. Potential diagnostic and prognostic marker dimethylglycine dehydrogenase (DMGDH) suppresses hepatocellular carcinoma metastasis *in vitro* and *in vivo*[J]. Oncotarget, 2016, 7(22): 32607-32616.
- [9] LIU J, LICHTENBERG T, HOADLEY K A, et al. An integrated tga pan-cancer clinical data resource to drive high-quality survival outcome analytics[J]. Cell, 2018, 173(2):400-416.
- [10] KIM D, PAGGI J M, PARK C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype[J]. Nat Biotechnol, 2019, 37(8): 907-915.

(下转第 254 页)

- 析[D]. 山东大学, 2021.
- [5] CAPUTO M, UCCIERO A, MELE C, et al. Use of administrative health databases to estimate incidence and prevalence of acromegaly in Piedmont Region, Italy[J]. *J Endocrinol Invest*, 2019, 42(4): 397–402.
- [6] ESPOSITO D, RAGNARSSON O, JOHANSSON G, et al. Prolonged diagnostic delay in acromegaly is associated with increased morbidity and mortality[J]. *Eur J Endocrinol*, 2020, 182(6): 523–531.
- [7] BIAGIO C, EMANUELE G, CATERINA P, et al. Psychological complications in patients with acromegaly: relationships with sex, arthropathy, and quality of life[J]. *Endocrine*, 2022, 77(3): 510–518.
- [8] DAL J S, BENEDIKTE G, ANDERSEN M, et al. Sex differences in acromegaly at diagnosis: a nationwide cohort study and meta-analysis of the literature[J]. *Clin Endocrinol*, 2021, 94(4): 625–635.
- [9] 盛志峰. 肢端肥大症药物治疗进展[J]. *临床内科杂志*, 2022, 39(8): 515–516.
- [10] 李冉, 段炼, 潘慧, 等. 奥曲肽敏感试验预测长效生长抑素受体配体治疗垂体 GH 腺瘤的疗效 [J]. *中华内分泌代谢杂志*, 2022, 38(4): 288–293.
- [11] URAI S, YAMAMOTO M, YAMAMOTO N, et al. Newer parameters of the octreotide test in patients with acromegaly[J]. *Pituitary*. 2023, Published online.
- [12] CHIN S, KU C, KIM B, et al. Medical treatment with somatostatin analogues in acromegaly: position statement[J]. *Endocrinol Metab*, 2019, 34(1): 53–62.
- [13] FLESERIU M B, BEVERLY M K, FREDA P U, et al. A pituitary society update to acromegaly management guidelines[J]. *Pituitary*, 2021, 24(1): 1–13.
- [14] 刘洪宇, 徐剑, 梁丹, 等. 长效生长抑素治疗对肢端肥大症术后患者糖脂代谢的影响[J]. *中国医师进修杂志*, 2021, 44(12): 1111–1114.
- [15] 李一琳, 郭晓鹏, 幸兵. 垂体生长激素腺瘤对生长抑素类似物的耐药机制研究进展[J]. *基础医学与临床*, 2015, 35(6): 846–850.
- [16] GOLA M, BONADONNA S, MAZZIOTTI G, et al. Resistance to somatostatin analogs in acromegaly: an evolving concept [J]. *J Endocrinol Invest*, 2014, 29(1): 86–93.
- [17] POTORAC I, PETROSSIANS P, DALY A, et al. T2-weighted MRI signal predicts hormone and tumor responses to somatostatin analogs in acromegaly[J]. *Endocr-Relat Cancer*, 2016, 23(11): 871–881.
- [18] WANG M, SHEN M, HE W, et al. The value of an acute octreotide suppression test in predicting short-term efficacy of somatostatin analogues in acromegaly[J]. *Endocr J*, 2016, 63(9): 819–834.

(2023-12-11 收稿)

(上接第 211 页)

- [11] KOVAKA S, ZIMIN A V, PERTEA G M, et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2 [J]. *Genome Biol*, 2019, 20(1): 278.
- [12] WU T, HU E, XU S, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data[J]. *Innovation (Camb)*, 2021, 2(3): 100141.
- [13] GILLESPIE M, JASSAL B, STEPHAN R, et al. The reactome pathway knowledge base 2022[J]. *Nucleic Acids Res*, 2022, 50(D1): 687–692.
- [14] LEE H. Analysis of bisulfite sequencing data using bismark and dmrcaller to identify differentially methylated regions[J]. *Methods Mol Biol*, 2022, 2443: 451–463.
- [15] WU H, XU T, FENG H, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates[J]. *Nucleic Acids Res*, 2015, 43(21): e141.
- [16] FISHILEVICH S, NUDEL R, RAPPAPORT N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards[J]. *Database (Oxford)*, 2017, 2017.
- [17] YU G, WANG L G, HE Q Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization[J]. *Bioinformatics*, 2015, 31(14): 2382–2383.
- [18] CAVALCANTE R G, SARTOR M A. Annotatr: genomic regions in context[J]. *Bioinformatics*, 2017, 33(15): 2381–2383.
- [19] LI Y, LU F, YIN Y. Applying logistic LASSO regression for the diagnosis of atypical Crohn's disease[J]. *Sci Rep*, 2022, 12(1): 11340.
- [20] HOADLEY K A, YAU C, HINOUE T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer[J]. *Cell*, 2018, 173(2): 291–304.
- [21] FORNER A, REIG M, BRUIX J. Hepatocellular carcinoma[J]. *Lancet*, 2018, 391(10127): 1301–1314.
- [22] LEE S, HWANG K S, LEE H J, et al. Aberrant CpG island hypermethylation of multiple genes in colorectal neoplasia[J]. *Lab Invest*, 2004, 84(7): 884–893.
- [23] LICCHESI J D, WESTRA W H, HOOKER C M, et al. Promoter hypermethylation of hallmark cancer genes in atypical adenomatous hyperplasia of the lung[J]. *Clin Cancer Res*, 2008, 14(9): 2570–2578.
- [24] LAKSHMINARASIMHAN R, LIANG G. The role of DNA methylation in cancer[J]. *Adv Exp Med Biol*, 2016, 945: 151–172.
- [25] WOUTERS J, KALENDER ATAK Z, AERTS S. Decoding transcriptional states in cancer[J]. *Curr Opin Genet Dev*, 2017, 43: 82–92.
- [26] LEAL-ESTEBAN L C, FAJAS L. Cell cycle regulators in cancer cell metabolism[J]. *Biochim Biophys Acta Mol Basis Dis*, 2020, 1866(5): 165715.
- [27] LIU Z, SUN Q, WANG X. PLK1, a potential target for cancer therapy[J]. *Transl Oncol*, 2017, 10(1): 22–32.
- [28] ZIEGLER Y, LAWS M J, SANABRIA GUILLEN V, et al. Suppression of FOXM1 activities and breast cancer growth *in vitro* and *in vivo* by a new class of compounds[J]. *NPJ Breast Cancer*, 2019, 5: 45.

(2023-11-07 收稿)