

整合分析构建基于增强子的非小细胞肺癌预后风险模型

张伟国^{1,2}, 陆秀宏², 黄钢², 靳明明², 程云章¹

1. 上海理工大学健康科学与工程学院, 上海 200093; 2. 上海健康医学院附属嘉定中心医院上海市分子影像学重点实验室, 上海 201318

【摘要】目的:通过整合DNA甲基组数据和转录组数据构建基于增强子的非小细胞肺癌预后风险模型。**方法:**使用加权基因共表达网络分析(WGCNA)从甲基化差异位点的增强子中识别非小细胞肺癌相关基因。并基于转录组数据通过LASSO-Cox回归算法构建并验证预后风险模型。**结果:**基于非小细胞肺癌的DNA甲基组数据分析获得了19784个差异甲基化位点,并对其分布模式进行了表征,其中包括6089个差异甲基化增强子位点。WGCNA从这6089个位点筛选出79个和非小细胞肺癌高度相关的增强子位点。基于转录组数据通过LASSO-Cox回归对79个增强子位点靶基因分析,构建10个基因的预后风险模型。在训练集和验证集中分析3年、5年和10年时间依赖的受试者工作特征曲线下面积(AUC)来评估预后风险模型。结果显示,训练集和验证集中的3年、5年和10年AUC均大于0.7。最后构建预测非小细胞肺癌患者3、5、10年生存情况的列线图。**结论:**本研究为理解增强子在非小细胞肺癌中的作用提供了新的见解,并具有通过指导个性化治疗决策来改善患者预后的潜力。

【关键字】非小细胞肺癌;增强子;甲基化;加权基因共表达网络分析;预后

【中图分类号】R318;R735.7

【文献标志码】A

【文章编号】1005-202X(2025)01-0112-10

Integrative analysis reveals enhancer-based prognostic risk prediction model for non-small cell lung cancer

ZHANG Weiguo^{1,2}, LU Xiuhong², HUANG Gang², JIN Mingming², CHENG Yunzhang¹

1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. Shanghai Key Laboratory of Molecular Imaging, Jiading District Central Hospital Affiliated to Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

Abstract: Objective To construct an enhancer-based prognostic risk prediction model for non-small cell lung cancer (NSCLC) by integrating DNA methylome data and transcriptome data. **Methods** The weighted gene co-expression network analysis (WGCNA) was used to identify NSCLC related genes from the differentially methylated positions (DMPs) of enhancers. Based on the transcriptome data, the prognostic risk prediction model was constructed using LASSO-Cox regression algorithm. **Results** Through the analysis on DNA methylome data of NSCLC, 19 784 DMPs were obtained and their distribution patterns were characterized, including 6 089 DMPs of enhancers. WGCNA analysis screened 79 highly correlated DMPs of enhancer with NSCLC from the 6 089 DMPs. After analyzing the target genes of 79 DMPs with LASSO-Cox regression based on the transcriptome data, 10 genes were used to construct a prognostic risk prediction model. The prognostic risk prediction model was evaluated by calculating the areas under the curve (AUC) of 3-, 5-, and 10-year time-dependent receiver operating characteristic (ROC) curves in training set and validation set; and the results showed that the 3-, 5-, and 10-year AUC in training set and validation set were all higher than 0.7. Finally, a nomogram was constructed to predict the 3-, 5-, and 10-year survival probabilities of NSCLC. **Conclusion** This study provides new insights into the role of enhancers in NSCLC and has the potential to improve the prognosis by guiding personalized treatment decisions.

Keywords: non-small cell lung cancer; enhancer; methylation; weighted gene co-expression network analysis; prognosis

【收稿日期】2024-09-20

【基金项目】国家自然科学基金(82127807);上海市分子影像学重点实验室建设项目(18DZ2260400)

【作者简介】张伟国,博士研究生,研究方向:生物医学,E-mail: zhangwg@sumhs.edu.cn

【通信作者】程云章,教授,研究方向:生物医学,E-mail: cyz2008@usst.edu.cn

前言

肺癌是中国最常见的癌症,2022年,中国有870 982例新发肺癌病例和766 898例肺癌死亡病例^[1]。非小细胞肺癌约占所有肺癌的80%~84%,包括腺癌(30%~50%)、鳞癌(20%~35%)和大细胞癌(9%)。小细胞癌约占所有肺癌16%~20%^[2]。其他非小细胞肺癌亚型,如肉瘤样癌、腺鳞癌等,为少见类型。非小细胞肺癌早期(I期、II期)诊断和治疗有望获得较好的预后,然而临床上检测到I期或II期只是少部分,此时肿瘤可以手术切除,大部分非小细胞肺癌患者在诊断时已为局部晚期或转移(III期或IV期),此时可能无法进行手术切除^[3]。

在哺乳动物中,DNA甲基化的主要形式是在胞嘧啶环的第5位碳原子上添加甲基^[4]。DNA甲基化可引起基因调控活性的永久性或短暂性改变^[5]。鉴于DNA甲基化既表现出可干预的修饰性,又表现出测量的稳定性^[6],已经被广泛用于研究癌症、治疗靶点和生物标志物^[7]。增强子是基因的顺式作用元件,可以位于基因的上游、下游或临近的基因内,与靶基因的距离从十万到百万个碱基不等,增强子通过与靶基因启动子相互作用、招募转录因子和共激活蛋

白来促进基因转录^[8]。增强子的异常在癌症的发生发展中起到重要作用^[9]。

越来越多的研究表明,由于癌症本身复杂性和基因间的复杂交互作用,越来越多的研究开始探索多基因的预后风险模型,仅仅在非小细胞肺癌方面就有基于基因表达的肿瘤转移风险模型^[10],基于lncRNA表达的预后风险模型^[11],针对肺腺癌有直接利用甲基化水平预测患者生存的模型^[12],增强子作为关键的顺式作用元件,尚未见基于增强子的非小细胞肺癌预后风险模型报道。本文基于甲基组和转录组数据,运用加权基因共表达网络分析(Weighted Gene Co-expression Network Analysis, WGCNA)和LASSO-Cox方法,构建基于增强子的非小细胞肺癌预后风险模型,以期临床决策提供指导。

1 材料与方法

本研究使用WGCNA方法从基于甲基组数据的甲基化差异位点的增强子中识别非小细胞肺癌相关基因。并基于转录组数据通过LASSO-Cox回归算法构建并验证预后风险模型,同时构建和检验了生存预测列线图,具体的技术路线和方法如图1所示。

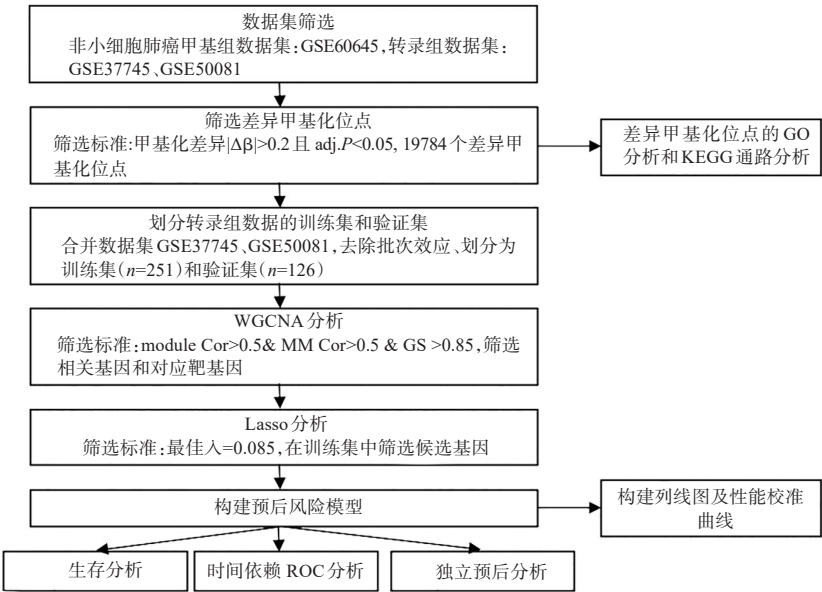


图1 本研究分析过程流程图
Figure 1 Flowchart of the analysis procedure in the study

1.1 数据资源

3个数据集 GSE60645^[13]、GSE37745^[14]和 GSE50081^[15]来自基因表达综合数据库(Gene Expression Omnibus, GEO)。其中甲基组数据集

GSE60645剔除3个小细胞肺癌样本数据,只留下非小细胞肺癌甲基组数据。此时的甲基组数据包含12例正常肺组织和121例肿瘤组织,肿瘤组织中腺癌83例、鳞癌23例、腺鳞癌1例、大细胞癌14例。转录组

数据集 GSE37745 和 GSE50081 基于同一芯片平台 (Affymetrix Human Genome U133 Plus 2.0 Array), 这两个数据集分别包括 106 例和 130 例腺癌, 66 例和 43 例鳞癌以及 24 例和 8 例大细胞癌病例, 转录组样本都具有完整临床信息和长期随访记录。

1.2 筛选差异甲基化位点

使用 R 包 'ChAMP' 对甲基组芯片数据进行筛选和注释, 筛选差异甲基化位点标准为甲基化差异 $|\text{beta}| > 0.2$ 且调整后 P 值 ($\text{adj.}P$) < 0.05 ^[16]。

1.3 基因本体论 (Gene Ontology, GO) 和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析

为了研究差异甲基化位点的功能, 利用 R 包 'clusterProfiler' 对差异甲基化位点进行 GO 分析和 KEGG 通路分析^[17]。

1.4 加权基因共表达网络的构建及相关模块的识别

对于差异甲基化位点中增强子位点, 使用 R 包 'wateRmelon' 进行归一化处理, 使用 R 包 'WGCNA' 进行加权基因共表达网络算法分析^[18]。筛选非小细胞肺癌相关模块的标准为模块相关性 (Module Membership, MM) > 0.5 且 $P < 0.01$ ^[19]。

1.5 鉴定非小细胞肺癌相关差异甲基化增强子位点靶基因

在 WGCNA 的候选模块中, 按照基因显著性值 (Gene Significance, GS) > 0.5 和 MM > 0.85 的标准筛选出非小细胞肺癌相关的增强子位点。此外, 每个非小细胞肺癌相关的增强子位点被注释到相应的基因^[20]。

1.6 特征基因的识别、验证及预后风险模型的构建

为了筛选关键基因并验证关键基因的预测能力, 对 GSE37745 和 GSE50081 的 377 例非小细胞肺癌病例转录组数据集进行合并, 使用 R 包 'sva' 去除批次效应, 使用 R 包 'caret' 将这 377 例非小细胞肺癌病例分为训练数据集和验证数据集。

在训练集中, 使用 R 包 'glmnet' 和 'survival' 进行 LASSO-Cox 回归, 为了选择最优 λ , 进行 10 折交叉验证^[21-22]。仅对单因素 Cox 回归分析中 $P < 0.05$ 的变量进行多因素 Cox 回归分析。

1.7 预后风险模型评价

在训练集和验证集中依次使用 R 包 'timeROC' 进行时间依赖的受试者工作特征曲线 (Receiver Operating Curves, ROC) 分析^[23], 然后使用 R 的 'pROC' 包进行独立的诊断 ROC 分析^[24], 使用 R 包 'survival' 进行生存分析, 并通过 R 包 'survminer' 和 R 包 'ggplot2' 进行可视化, 通过 R 包 'rms' 绘制列线图。

2 结果

2.1 非小细胞肺癌差异甲基化位点概况

以甲基化差异 $|\text{beta}| > 0.2$ 且 $\text{adj.}P < 0.05$ 为标准筛选得到 19 784 个差异甲基化位点, 19 784 个差异甲基化位点包括 54.16% 的 (10 716/19 784) 高甲基化位点和 45.84% 的 (9 068/19 784) 低甲基化位点。这些差异甲基化位点中增强子 (Enhancer: 6 089 个) 位点多于目前甲基化研究较多的基因体 (Body: 6 044 个) 和启动子 (5'UTR: 1 411 个, TSS200: 1 319 个, TSS1500: 2 388 个, 1stExon: 892 个, 共 6 010 个) 位点 (图 2a)。CpG 岛在基因不同部位分布差异明显 (图 2b、c)。差异甲基化位点和差异甲基化增强子位点在 (图 2d、e) 染色体上的分布存在明显差异, 6 089 个差异甲基化增强子位点中有 3 523 个高甲基化位点和 2 566 个低甲基化位点。

2.2 差异甲基化位点的 GO 和 KEGG 富集分析

GO 和 KEGG 富集分析探究差异甲基化位点的生物学功能。GO 分析表明, 差异甲基化位点与突触 ($\text{adj.}P < 0.001$)、发育 ($\text{adj.}P < 0.001$)、DNA 结合 ($\text{adj.}P < 0.001$) 等功能显著相关 (图 3a)。KEGG 通路分析表明, 差异甲基化位点主要参与 Rap1 信号通路 ($\text{adj.}P < 0.001$)、癌症蛋白聚糖通路 ($\text{adj.}P < 0.01$)、Hippo 信号通路 ($\text{adj.}P < 0.01$)、Wnt 信号通路 ($\text{adj.}P < 0.001$)、MAPK 信号通路 ($\text{adj.}P < 0.001$) 和多能干细胞调节信号通路 ($\text{adj.}P < 0.001$) (图 3b)。

2.3 非小细胞肺癌相关模块和基因的鉴定

构建 WGCNA 对差异甲基化增强子位点进行分析。为了符合无标度网络的分布, 根据 pickSoftThreshold 函数 (图 4a、b) 分析结果, 将软阈值设置为 $\beta = 6$ ($R^2 = 0.91$)。将相似模块 (相关性在 0.25 以上) 进行合并最终共识别出 15 个模块 (图 4c)。深绿色模块 ($\text{cor} = 0.77$, $P = 4\text{e-}27$) 和黑色模块 ($\text{cor} = -0.69$, $P = 3\text{e-}20$) 被筛选出来作为非小细胞肺癌相关模块 (图 4d、e)。

随后, 基于 GS > 0.5 和 MM > 0.85 筛选标准, 从深绿色模块和黑色模块中鉴定出 79 个相关的差异甲基化增强子位点 (表 1), 这 79 个增强子位点对应 65 个靶基因。

2.4 筛选关键基因并构建预后模型

合并来自 GSE37745 和 GSE50081 的 377 例非小细胞肺癌病例并去除批次效应。随后, 将 377 例非小细胞肺癌病例分为 251 例训练数据集和 126 例验证数据集。

为了构建风险模型, 在训练数据集中, LASSO 回归对筛选到的 79 个增强子位点和对应的靶基因进行进一步分析, 10 折交叉验证选择最小惩罚系数 λ 进而

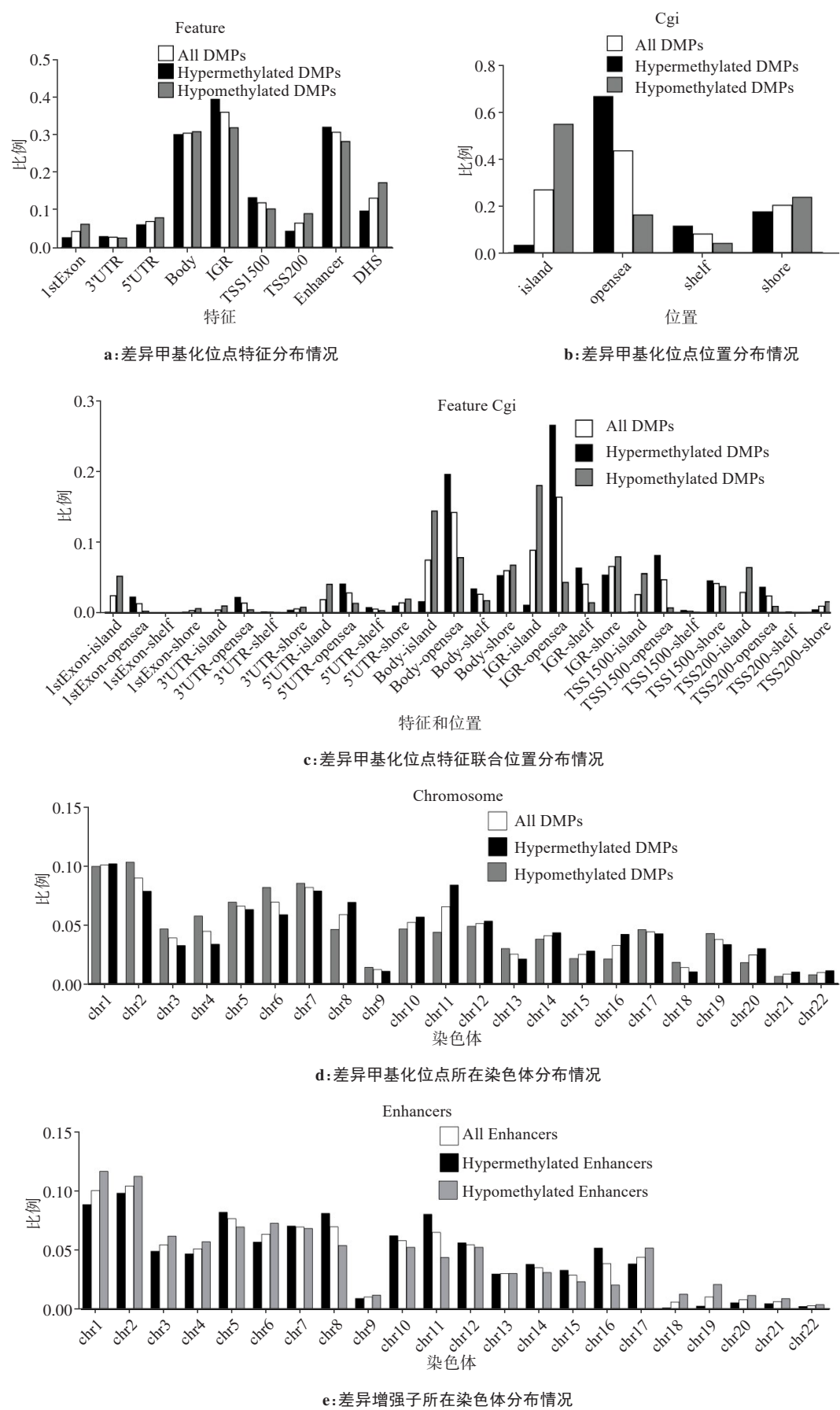
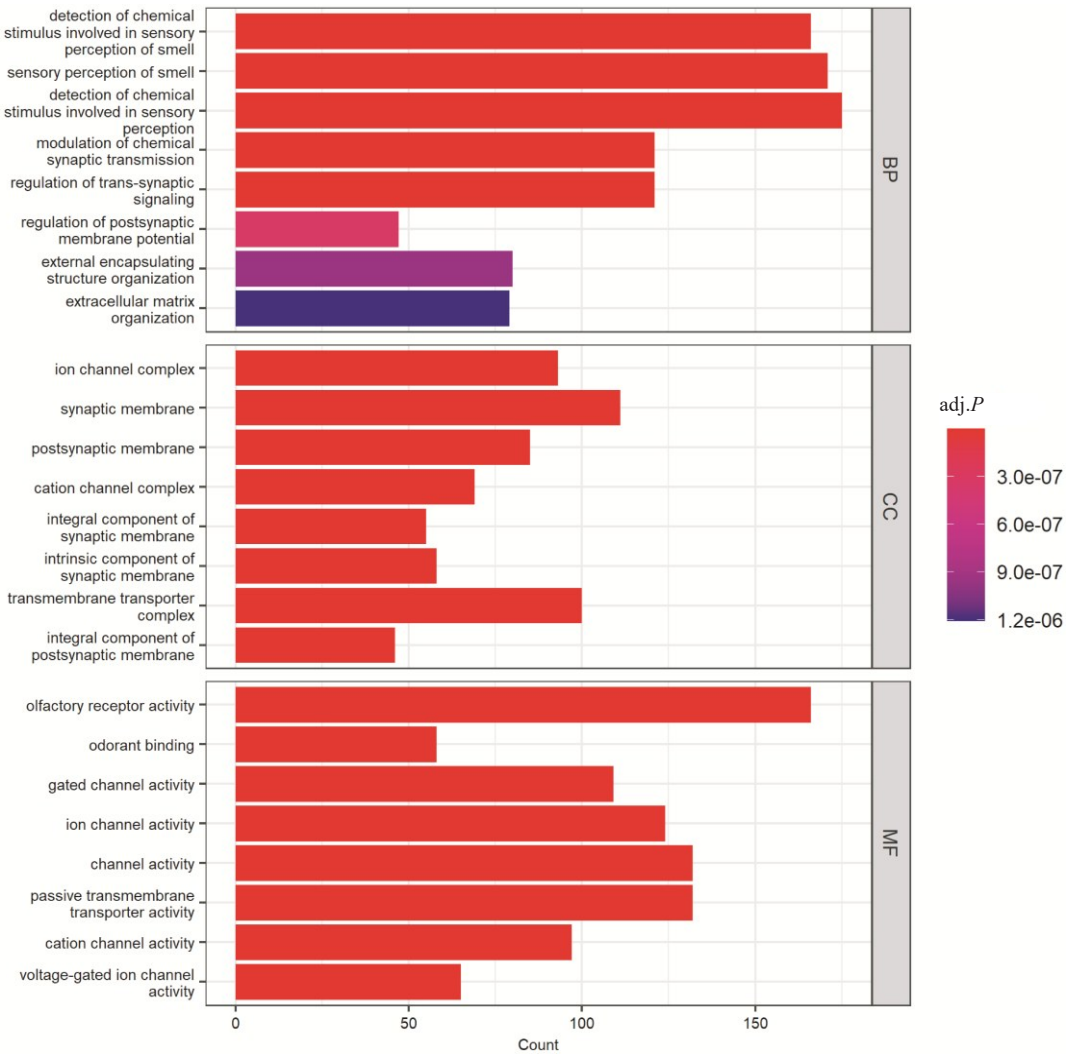
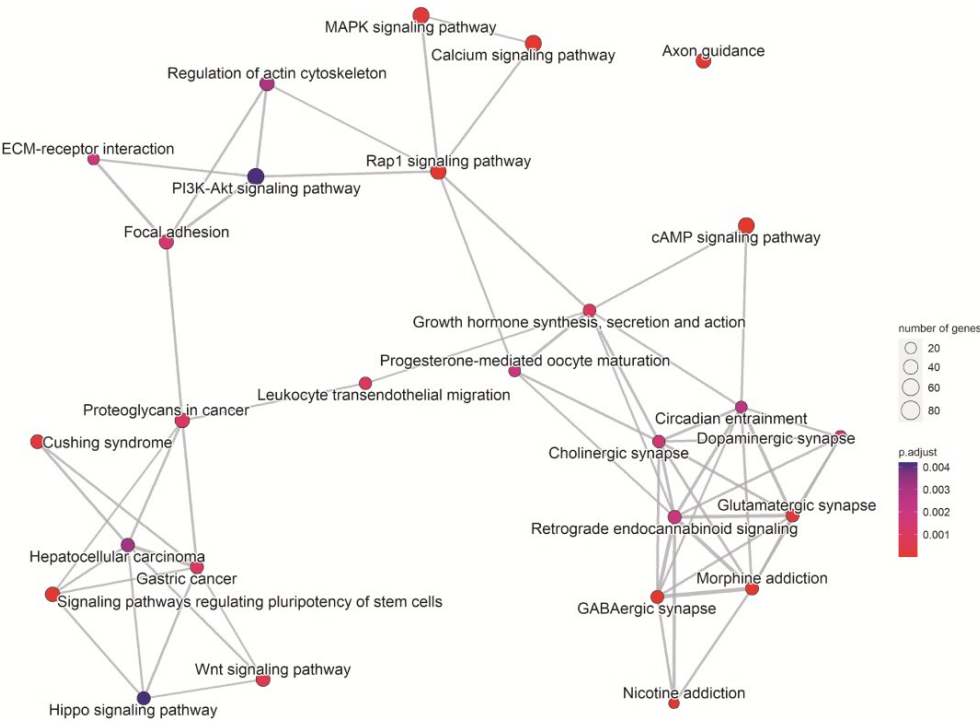


图2 非小细胞肺癌差异甲基化位点分布

Figure 2 Distributions of differentially methylated positions (DMPs) in non-small cell lung cancer (NSCLC)



a:GO 分析(BP:生物过程,CC:细胞组分,MF:分子功能)



b:KEGG 富集分析

图3 差异甲基化位点的GO和KEGG 富集分析

Figure 3 GO and KEGG enrichment analyses for the DMPs

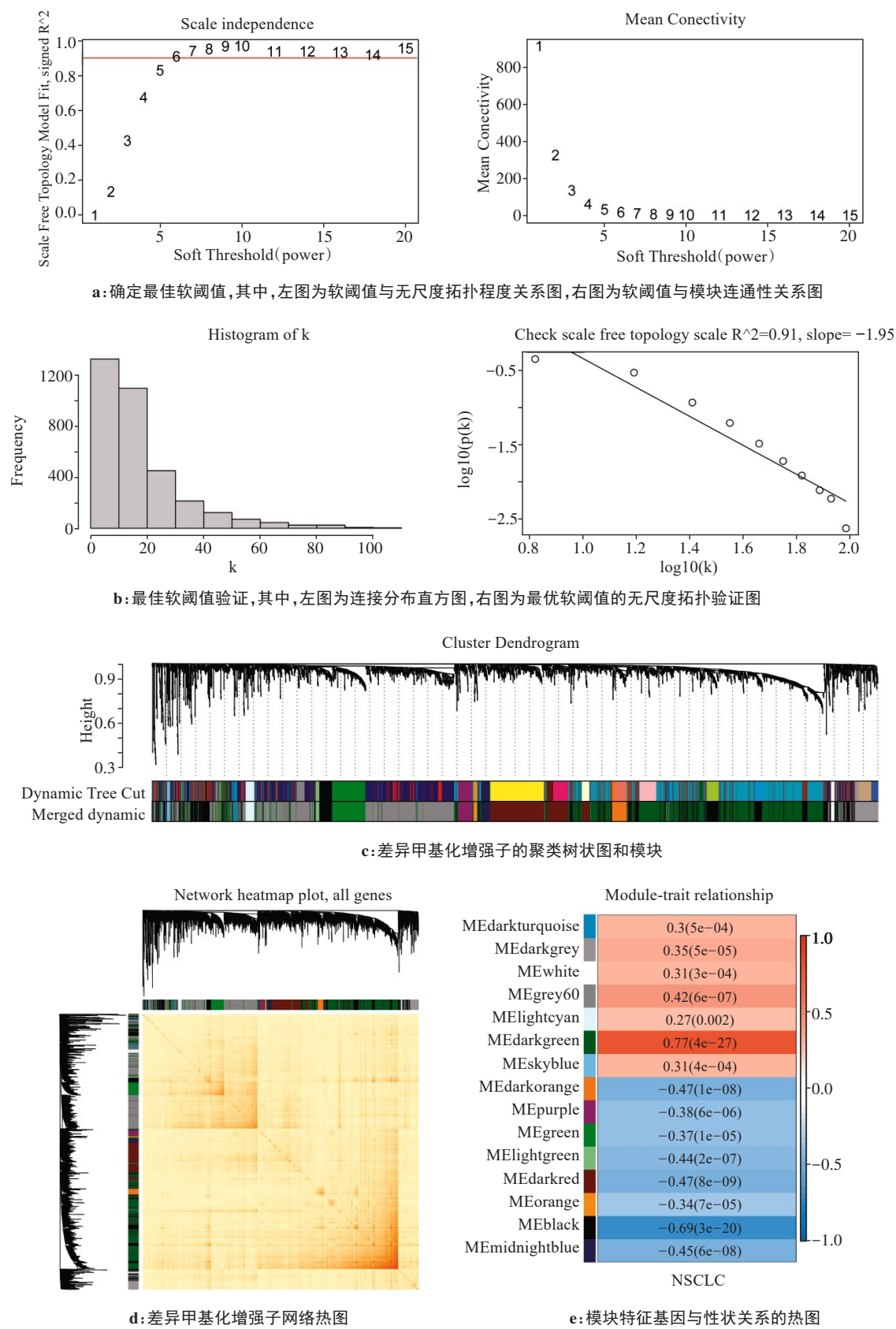


图4 WGCNA 分析鉴定非小细胞肺癌相关模块

Figure 4 WGCNA analysis for identifying NSCLC related modules

实现变量选择(图 5a), 最优 $\lambda=0.085$ 从而得到 14 个基因(图 5b)。进一步对这 14 个基因进行 Cox 回归分析, 10 个单因素 Cox 回归分析中 $P<0.05$ 的基因被用于构建风险模型(表 2)。基于多因素 Cox 回归分析

表 1 WGCNA 分析鉴定的非小细胞肺癌相关差异甲基化增强子位点
Table 1 NSCLC related DMPs identified by WGCNA analysis

cg26026748, cg11936643, cg13327911, cg12099279, cg07961632, cg05516390, cg10020892, cg19588399, cg13564459, cg02196651, cg27450010, cg25788793, cg06745704, cg11326574, cg04608356, cg12503394, cg17934775, cg10184289, cg23838005, cg03613662, cg19017553, cg05943655, cg16925459, cg27496339, cg18432877, cg11855693, cg03915012, cg02075410, cg08943107, cg13654884, cg13625187, cg13595913, cg14369455, cg02580005, cg17724627, cg17343374, cg11302791, cg14791193, cg01552919, cg03143486, cg17523415, cg10843210, cg11872373, cg25426203, cg02596427, cg09307985, cg14468280, cg19178151, cg15031753, cg05049448, cg08686960, cg07300534, cg16326402, cg09820244, cg22954906, cg02883503, cg20614671, cg15888569, cg18483459, cg06215107, cg06381455, cg03729251, cg26232945, cg07904475, cg01377200, cg04503570, cg17897352, cg10628699, cg06155414, cg23440816, cg19130973, cg08677210, cg11196788, cg22041712, cg15007228, cg19368031, cg03622664, cg23082657, cg09685096
--

的预后风险分值(Risk Score, RS)模型如下:
 $RS=9.6978-LRBA\times0.1525+DNAAF5\times0.3956+ERMP1\times0.1187-ITPKB\times0.3915-LNX1\times0.2544+$

$CMIP\times0.1989-IQSEC1\times0.8306-CCNP\times0.6923+PINX1\times0.2548+SLC2A9\times0.1108$ 。风险因子图显示该模型能较好地地区分高低风险人群(图 5c)。

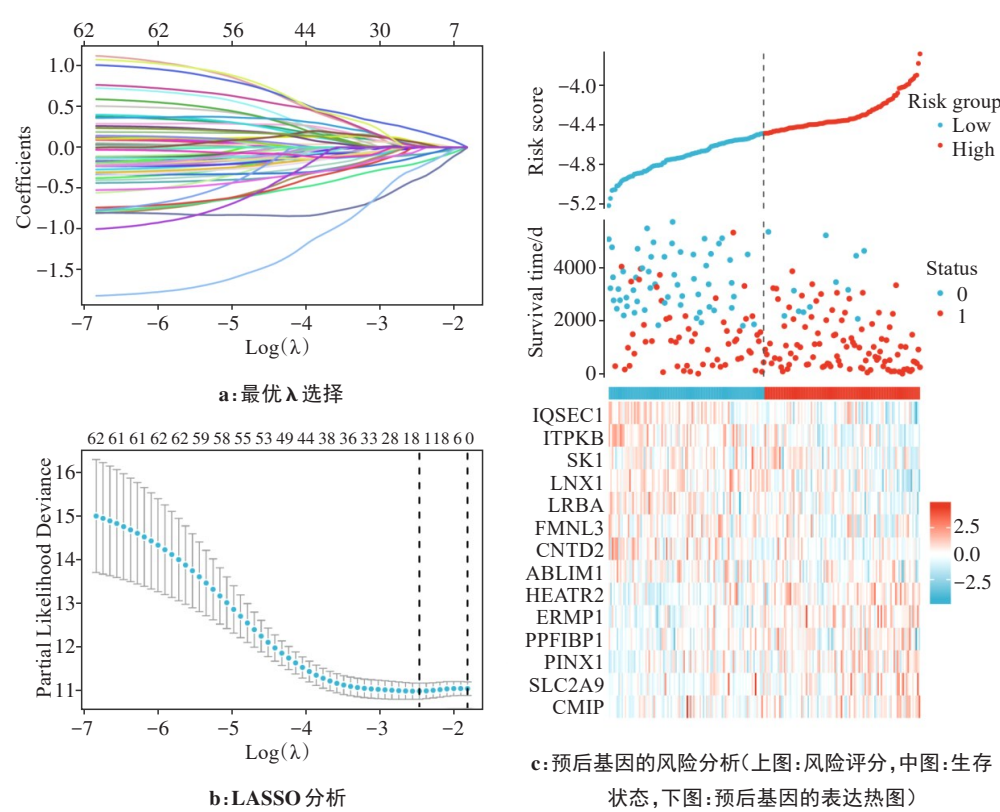


图 5 LASSO-Cox 回归构建预后风险模型

Figure 5 Prognostic risk prediction model constructed by LASSO-Cox regression

2.5 预后风险模型在训练集和验证集中的表现

在训练集中,总生存预后风险模型的风险比(Hazard Ratio, HR)为2.88(95%CI:2.04~4.07),见图 6a。基于RS的总生存预后风险模型(0.751)的ROC曲线下面积(Area Under Curve, AUC)优于基于年龄(0.646)、性别(0.524)、分期(0.611)、组织学(0.552)等

模型的表现(图 6b)。3年、5年和10年生存率预后风险模型的AUC分别为0.735、0.725和0.751(图 6c)。

在验证集中,总生存预后风险模型的HR为2.71(95%CI:1.64~4.48)(图 6d)。基于RS的总生存预后风险模型的AUC(0.704)优于基于年龄(0.685)、性别(0.527)、分期(0.627)、组织学(0.526)的表现(图 6e)。

表2 预后风险模型的10个基因详细信息
Table 2 Detailed information of 10 genes for prognostic risk prediction model construction

基因	探针	adj.P值	beta	染色体	特征	位置
LRBA	cg02580005	2.93E-24	0.303 04	4	Body	island
	cg06215107	5.24E-17	0.337 82	4	Body	island
	cg03729251	1.89E-16	0.205 98	4	Body	island
	cg07904475	2.06E-16	0.232 72	4	Body	island
	cg02075410	2.49E-30	0.26535	4	Body	island
DNAAF5	cg03143486	1.00E-22	0.30175	7	Body	shelf
ERMP1	cg13595913	6.30E-26	0.22466	9	3'UTR	opensea
ITPKB	cg04503570	2.47E-16	0.21477	1	Body	opensea
	cg04503570	2.47E-16	0.21477	1	Body	opensea
CMIP	cg07300534	2.39E-18	0.21555	16	Body	opensea
IQSEC1	cg19130973	3.56E-14	0.24468	3	Body	shore
CCNP	cg12099279	1.69E-23	0.221193	19	1stExon	island
PINX1	cg11302791	3.31E-23	0.28304	8	Body	opensea
SLC2A9	cg25788793	3.04E-18	0.273439	4	5'UTR	opensea
LNX1	cg02196651	3.13E-19	0.235191	4	TSS200	opensea

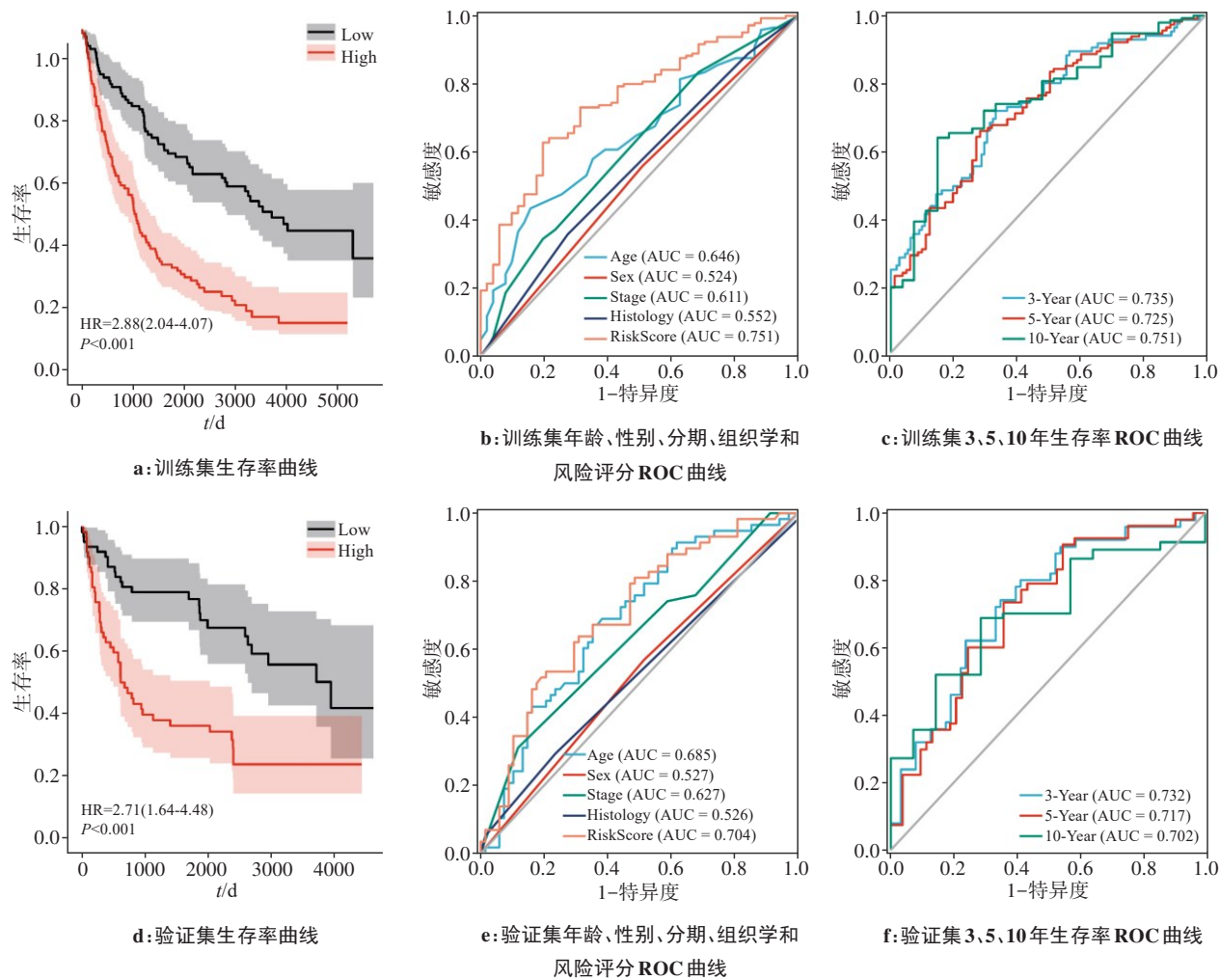


图6 预后风险模型验证
Figure 6 Validation of prognostic risk prediction model

3年、5年和10年生存率预后风险模型的AUC分别为0.732、0.717和0.702(图6f)。

基于377例非小细胞肺癌病例构建了包括年龄、

性别、分期和RS的列线图,以预测3年、5年和10年总生存情况(图7a)。校准曲线均接近参考理想预测模型线(图7b)。

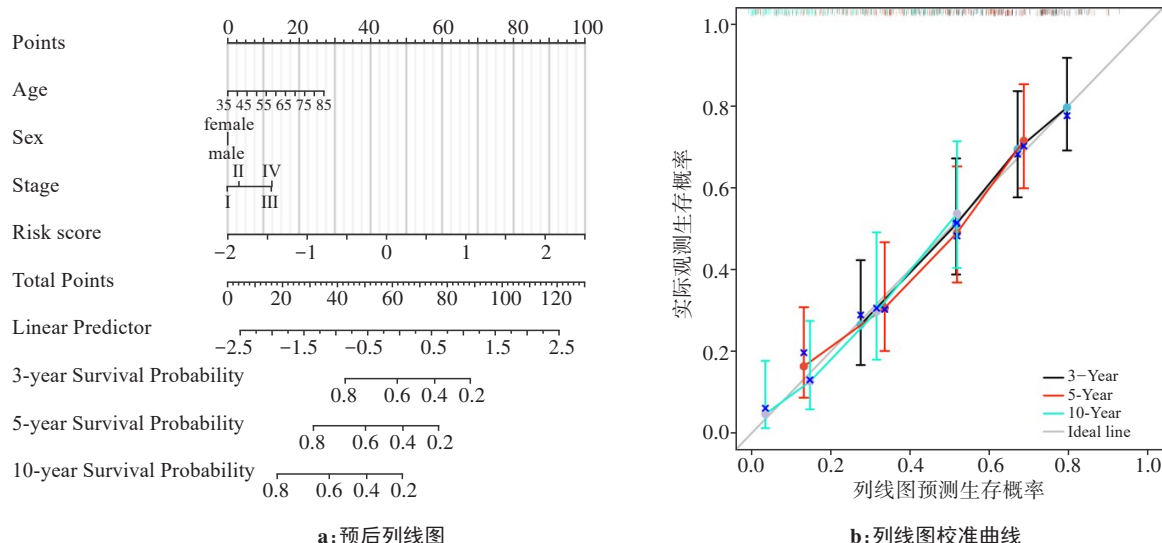


图7 列线图 and 校准曲线

Figure 7 Nomogram and calibration curves

3 讨论

目前,基因芯片和测序已广泛应用于非小细胞肺癌研究,但这些研究中专门针对非小细胞肺癌增强子作用的研究较少。笔者对非小细胞肺癌样本中的甲基组数据和转录数据进行整合分析。通过肿瘤和正常样本的比较,筛选到19784个差异甲基化位点,并分析了它们的全基因组分布。GO和KEGG富集分析显示,这些差异甲基化位点涉及Rap1信号通路、癌症中的蛋白聚糖通路、Hippo信号通路、Wnt信号通路、MAPK信号通路和调节干细胞多能性的信号通路。有研究显示调控Rap1信号通路可以抑制非小细胞肺癌的转移^[25]。靶向Wnt和Hippo信号之间的YAP1串话可以在非小细胞肺癌中产生治疗作用^[26]。激活Ras/MAPK通路会促进非小细胞肺癌的增殖和迁移^[27]。

作为一种动态的基因调控机制,DNA甲基化的证据大多局限于特定的例子^[28]。总的来说,DNA高甲基化抑制转录起始并进一步下调基因表达^[29]。然而,一些研究表明,DNA甲基化可能并不总是起抑制作用,体内和体外试验表明,高甲基化的增强子可以允许转录^[30],因此增强子在非小细胞肺癌中的作用远未被完全了解。

笔者接下来对19784个差异甲基化位点作进一步分析,筛选出其中的6089个增强子位点,通过WGCNA分析,笔者从6089个增强子位点筛选得到79个增强子位点。考虑到非小细胞肺癌的复杂性和非小细胞肺癌

相关基因的交互作用,本研究旨在整合多个基因和相应的系数,构建非小细胞肺癌的预后风险模型,而不是聚焦于单个基因。因此,对这些增强子位点对应的基因进行LASSO-Cox回归分析,构建了一个包含10个基因和相应系数的预后风险模型,该模型在训练集和验证集中均有较好表现。最后构建预测生存情况的列线图,校准曲线显示列线图有较好的预测性能。

实际上这10个基因中,文献报道LRBA基因调节内体运输,特别是配体激活受体的内吞作用^[31],LRBA与CTLA4基因共定位于内体囊泡,参与调节CTLA4的表达^[32]。DNAAF5基因促进肝细胞癌的发展^[33]。ERMP1基因的上调在人类癌症中很常见^[34],在乳腺癌细胞中发现ERMP1基因过表达^[35]。ITPKB基因在肺癌组织中表达下调^[36]。CMIP基因激活MAPK/ERK信号通路,进一步促进肺癌的增殖和转移^[37]。CMIP基因在胃癌中也是一个已知的癌基因^[38]。IQSEC1基因与许多致癌信号通路相互作用,如RAF、mTORC2和NFkB; IQSEC1基因的过表达与肿瘤的发生有关^[39]。CCNP在结肠癌和肺癌中过表达^[40-41]。在乳腺癌、肺癌和结直肠癌中,CCNP基因上调多能性(SOX2、OCT4、NANOG)标志物的表达,并促进球体形成和干性^[42]。在人类乳腺癌中,PINX1基因通过抑制NF-kB/MMP-9信号通路抑制侵袭和转移^[43]。PINX1基因也是非小细胞肺癌中潜在的预后因子^[44]。SLC2A9基因与尿酸浓度相关,

是肺腺癌的诊断标志物和治疗靶点之一^[45]。这10个基因与多个癌症的发生发展有密切联系,理论上作为非小细胞肺癌的预后风险模型构成因素是合理的,值得进一步的研究。

本研究存在一些局限性。首先,在该数据集中,非小细胞肺癌的亚型分布不平衡。其次,预后风险模型基于甲基组数据进行构建,并在转录组数据进行验证,但未在缺乏临床信息的甲基组数据进行验证。第三,本研究的风险模型是基于数据构建,分析结果有待实验验证。

综上所述,本研究采用整合DNA甲基组和mRNA转录组数据来构建基于增强子的非小细胞肺癌预后风险模型。预后风险模型对非小细胞肺癌的预后风险判断具有潜在的临床应用价值。

【参考文献】

- [1] Xia CF, Dong XS, 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] Bajbouj K, Al-Ali A, Ramakrishnan RK, et al. Histone modification in NSCLC: molecular mechanisms and therapeutic targets[J]. Int J Mol Sci, 2021, 22(21): 11701.
- [3] Osmani L, Askin F, Gabrielson E, et al. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy[J]. Semin Cancer Biol, 2018, 52 (Pt 1): 103-109.
- [4] Bestor TH. The DNA methyltransferases of mammals[J]. Hum Mol Genet, 2000, 9(16): 2395-2402.
- [5] Skvortsova K, Iovino N, Bogdanović O. Functions and mechanisms of epigenetic inheritance in animals[J]. Nat Rev Mol Cell Biol, 2018, 19(12): 774-790.
- [6] Joo JE, Dowty JG, Milne RL, et al. Heritable DNA methylation marks associated with susceptibility to breast cancer[J]. Nat Commun, 2018, 9(1): 867.
- [7] Feinberg AP, Tycko B. The history of cancer epigenetics[J]. Nat Rev Cancer, 2004, 4(2): 143-153.
- [8] Adhikary S, Roy S, Chacon J, et al. Implications of enhancer transcription and eRNAs in cancer[J]. Cancer Res, 2021, 81(16): 4174-4182.
- [9] Angeloni A, Bogdanovic O. Enhancer DNA methylation: implications for gene regulation[J]. Essays Biochem, 2019, 63(6): 707-715.
- [10] Ding H, Shi L, Chen Z, et al. Construction and evaluation of a prognostic risk model of tumor metastasis-related genes in patients with non-small cell lung cancer[J]. BMC Med Genomics, 2022, 15 (1): 187.
- [11] Zhang QZ, Liu XC, Chen ZN, et al. Novel GIRIncRNA signature for predicting the clinical outcome and therapeutic response in NSCLC[J]. Front Pharmacol, 2022, 13: 937531.
- [12] Wang XL, Zhou B, Xia YX, et al. A methylation-based nomogram for predicting survival in patients with lung adenocarcinoma[J]. BMC Cancer, 2021, 21(1): 801.
- [13] Karlsson A, Jönsson M, Lauss M, et al. Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome[J]. Clin Cancer Res, 2014, 20(23): 6127-6140.
- [14] Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation[J]. Clin Cancer Res, 2013, 19(1): 194-204.
- [15] Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients[J]. J Thorac Oncol, 2014, 9(1): 59-64.
- [16] Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k chip analysis methylation pipeline[J]. Bioinformatics, 2014, 30(3): 428-430.
- [17] Yu GC, Wang LG, Han YY, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. OMICS, 2012, 16(5): 284-287.
- [18] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. BMC Bioinformatics, 2008, 9: 559.
- [19] Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules[J]. BMC Syst Biol, 2007, 1: 54.
- [20] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis[J]. Stat Appl Genet Mol Biol, 2005, 4: Article17.
- [21] Tibshirani R. The lasso method for variable selection in the cox model[J]. Stat Med, 1997, 16(4): 385-395.
- [22] Engebretsen S, Bohlin J. Statistical predictions with glmnet[J]. Clin Epigenetics, 2019, 11(1): 123.
- [23] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks[J]. Stat Med, 2013, 32 (30): 5381-5397.
- [24] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves[J]. BMC Bioinformatics, 2011, 12: 77.
- [25] Kan J, Fu BQ, Zhou RS, et al. He-Chan Pian inhibits the metastasis of non-small cell lung cancer via the miR-205-5p-mediated regulation of the GREM1/Rap1 signaling pathway[J]. Phytomedicine, 2022, 94: 153821.
- [26] Mui CW, Chan WN, Chen BN, et al. Targeting YAP1/TAZ in non-small-cell lung carcinoma: From molecular mechanisms to precision medicine[J]. Int J Cancer, 2023, 152(4): 558-571.
- [27] Liu ZA, Zhao MN, Jiang XZ, et al. Upregulation of KLHL17 promotes the proliferation and migration of non-small cell lung cancer by activating the Ras/MAPK signaling pathway[J]. Lab Invest, 2022, 102 (12): 1389-1399.
- [28] Angeloni A, Bogdanovic O. Enhancer DNA methylation: implications for gene regulation[J]. Essays Biochem, 2019, 63(6): 707-715.
- [29] Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications[J]. Nat Rev Cancer, 2011, 11(10): 726-734.
- [30] Bestor TH. The DNA methyltransferases of mammals[J]. Hum Mol Genet, 2000, 9(16): 2395-2402.
- [31] Alangari A, Alsultan A, Adly N, et al. LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency[J]. J Allergy Clin Immunol, 2012, 130(2): 481-488.e2.
- [32] Charbonnier LM, Janssen E, Chou J, et al. Regulatory T-cell deficiency and immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like disorder caused by loss-of-function mutations in LRBA[J]. J Allergy Clin Immunol, 2015, 135(1): 217-227.
- [33] Liu YP, Wu Q, Sun TT, et al. DNAAF5 promotes hepatocellular carcinoma malignant progression by recruiting USP39 to improve PFKL protein stability[J]. Front Oncol, 2022, 12: 1032579.
- [34] Grandi A, Santi A, Campagnoli S, et al. ERMP1, a novel potential oncogene involved in UPR and oxidative stress defense, is highly expressed in human cancer[J]. Oncotarget, 2016, 7(39): 63596-63610.
- [35] Wu J, Liu S, Liu G, et al. Identification and functional analysis of 9p24 amplified genes in human breast cancer[J]. Oncogene, 2012, 31(3): 333-341.
- [36] Bäder S, Glaubke E, Grüb S, et al. Effect of the actin- and calcium-regulating activities of ITPKB on the metastatic potential of lung cancer cells[J]. Biochem J, 2018, 475(12): 2057-2071.
- [37] Yu XY, Wang M, Qian JJ. CMaf-Inducing protein promotes LUAD proliferation and metastasis by activating the MAPK/ERK pathway[J]. Evid Based Complement Alternat Med, 2022, 2022: 2501846.
- [38] Zhang JL, Huang J, Wang XY, et al. CMIP is oncogenic in human gastric cancer cells[J]. Mol Med Rep, 2017, 16(5): 7277-7286.
- [39] Nacke M, Sandilands E, Nikolatou K, et al. An ARF GTPase module promoting invasion and metastasis through regulating phosphoinositide metabolism[J]. Nat Commun, 2021, 12(1): 1623.
- [40] Sánchez-Botet A, Gasa L, Quandt E, et al. The atypical cyclin CNTD2 promotes colon cancer cell proliferation and migration[J]. Sci Rep, 2018, 8(1): 11797.
- [41] Gasa L, Sanchez-Botet A, Quandt E, et al. A systematic analysis of orphan cyclins reveals CNTD2 as a new oncogenic driver in lung cancer[J]. Sci Rep, 2017, 7(1): 10228.
- [42] Sánchez-Botet A, Quandt E, Masip N, et al. Atypical cyclin P regulates cancer cell stemness through activation of the WNT pathway[J]. Cell Oncol (Dordr), 2021, 44(6): 1273-1286.
- [43] Shi ML, Cao MH, Song J, et al. PinX1 inhibits the invasion and metastasis of human breast cancer via suppressing NF-κB/MMP-9 signaling pathway[J]. Mol Cancer, 2015, 14: 66.
- [44] Wang SG, Zhang H, Zhu JQ, et al. PinX1 Is a Potential prognostic factor for non-small-cell lung cancer and inhibits cell proliferation and migration[J]. Biomed Res Int, 2017, 2017: 7956437.
- [45] Zhang YL, Qin H, Bian J, et al. SLC2As as diagnostic markers and therapeutic targets in LUAD patients through bioinformatic analysis[J]. Front Pharmacol, 2022, 13: 1045179.

(编辑:薛泽玲)