

# 基于外周血单个核细胞 ATM 基因甲基化的联合模型在胰腺癌早期诊断中的价值

祝永才<sup>1</sup>, 谢艳<sup>2</sup>, 齐秋晨<sup>1</sup>, 李培龙<sup>1</sup>, 王传新<sup>1</sup>, 杜鲁涛<sup>2</sup>

(1.山东大学第二医院检验医学中心,山东 济南 250033;2.山东大学齐鲁医院检验医学中心,山东 济南 250012)

**摘要:**目的 利用随机森林算法建立并验证基于共济失调毛细血管扩张突变(ataxia telangiectasia mutated, ATM)基因甲基化和临床检验指标的胰腺癌联合诊断模型,为胰腺癌的早期诊断提供新途径。方法 回顾性收集118例胰腺癌患者(胰腺癌组)和73例健康对照者(健康对照组)的外周血单个核细胞(peripheral blood mononuclear cells, PBMCs)标本以及临床检验指标检测结果,分为发现集、训练集和验证集。在发现集中,利用935K甲基化芯片对两组的PBMCs样本进行差异甲基化位点(differentially methylated points, DMPs)分析。在训练集和验证集样本中,采用焦磷酸测序对候选DMP进行验证。在训练集中,利用随机森林算法进行变量筛选并构建联合诊断模型,并采用受试者工作特征(receiver operating characteristic, ROC)曲线在验证集中评价模型在胰腺癌各期以及在糖类抗原19-9(carbohydrate antigen 19-9, CA19-9)阴性患者中的诊断效能。结果 甲基化芯片结果显示,以 $|\Delta\beta| \geq 0.1$ ,且 $P < 0.01$ 的筛选条件,获得132个DMPs,其中ATM基因DMP的甲基化水平在胰腺癌组的PBMCs中显著高表达,且能够区分胰腺癌组和健康对照组( $AUC = 0.871, P < 0.001$ )。焦磷酸测序结果进一步表明,胰腺癌组PBMCs的ATM基因DMP呈高甲基化水平。基于随机森林算法的变量筛选获得3个变量ATM基因DMP、CA19-9和白蛋白(albumin, ALB),在训练集和验证集中,CA19-9表达水平在胰腺癌组中显著升高( $P < 0.001$ ),ALB表达水平明显降低( $P < 0.001$ )。以上述3个变量水平为特征,训练集中通过随机森林构建的联合诊断模型("AmCA")对胰腺癌诊断效能的曲线下面积(area under the curve, AUC)为0.992(95%CI:0.952~1.000)。验证集中模型对胰腺癌诊断效能的AUC为0.982(95%CI:0.895~1.000),优于单独CA19-9的AUC值[0.840(95%CI:0.705~0.930)];模型对早期胰腺癌诊断效能良好,I期和II期胰腺癌诊断效能的AUC分别为1.000(95%CI:0.863~1.000)和0.979(95%CI:0.840~1.000)。模型对CA19-9阴性胰腺癌患者诊断效能的AUC为0.751(95%CI:0.639~0.843),敏感度为52.2%,特异度为98.1%。结论 ATM基因在胰腺癌患者的PBMCs中呈高甲基化水平,以ATM基因高甲基化、高CA19-9和低ALB水平为特征,借助随机森林算法构建的联合诊断模型在胰腺癌的早期诊断中具有重要临床价值,并能够弥补常规标志物CA19-9诊断性能不足。

**关键词:**胰腺癌;外周血单个核细胞;DNA甲基化;早期诊断;分子标志物

中图分类号:R735.9

文献标志码:A

## Value of a combined model based on ATM gene methylation of peripheral blood mononuclear cells in the early diagnosis of pancreatic cancer

ZHU Yongcai<sup>1</sup>, XIE Yan<sup>2</sup>, QI Qiuchen<sup>1</sup>, LI Peilong<sup>1</sup>, WANG Chuanxin<sup>1</sup>, DU Lutao<sup>2</sup>

(1. Department of Clinical Laboratory, The Second Hospital of Shandong University, Jinan 250033, Shandong, China;

2. Department of Clinical Laboratory, Qilu Hospital of Shandong University, Jinan 250012, Shandong, China)

**Abstract: Objective** To provide a new approach for the early diagnosis of pancreatic cancer by establishing and validating a combined diagnostic model for pancreatic cancer based on the methylation of the ataxia telangiectasia mutated (ATM) gene and clinical test indicators using the random forest algorithm. **Methods** Retrospectively, 118 specimens

of peripheral blood mononuclear cells (PBMCs) from pancreatic cancer patients (pancreatic cancer group) and 73 specimens from healthy controls (healthy control group) were collected and their clinical test results were recorded. They were divided into discovery set, training set and validation set. In the discovery set, 935K methylation chip was used to analyze the differentially methylated points (DMPs) of PBMCs samples from both groups. In the training set and validation set samples, pyrosequencing was used to validate the candidate DMPs. In the training set, random forest algorithm was used for variable selection and to construct a combined diagnostic model. The diagnostic efficacy of the model in pancreatic cancer at different stages and in patients with negative carbohydrate antigen 19-9 (CA19-9) was evaluated in the validation set using the receiver operating characteristic (ROC) curve. **Results** The results of the DNA methylation beadchip indicated that, with the screening condition of  $|\Delta\beta| \geq 0.1$  and  $P < 0.01$ , 132 differentially methylated points (DMPs) were obtained. Among them, the methylation level of the ATM gene DMPs in PBMCs of the pancreatic cancer group was significantly higher than that of the healthy control group, and it could distinguish the pancreatic cancer group from the healthy control group (AUC=0.871,  $P < 0.001$ ). The pyrosequencing results further indicated that the ATM gene DMPs in PBMCs of the pancreatic cancer group was in a high-methylation level. Based on the variable selection of the random forest algorithm, three variables, ATM gene DMP, CA19-9, and albumin (ALB), were obtained. In the training set and validation set, the expression level of CA19-9 was significantly increased in the pancreatic cancer group ( $P < 0.001$ ), and the expression level of ALB was significantly decreased ( $P < 0.001$ ). Using the expression levels of the above three variables as features, the combined diagnostic model ("AmCA") constructed by random forest in the training set had an AUC of 0.992 (95%CI: 0.952-1.000) for pancreatic cancer diagnosis. The AUC of the model in the validation set was 0.982 (95%CI: 0.895-1.000), which was superior to the AUC value of CA19-9 alone (0.840, 95%CI: 0.705-0.930); the model had good diagnostic efficacy for early-stage pancreatic cancer, with AUC values of 1.000 (95%CI: 0.863-1.000) and 0.979 (95%CI: 0.840-1.000) for stage I and stage II pancreatic cancer, respectively. The AUC of the model for diagnosing CA19-9-negative pancreatic cancer patients was 0.751 (95%CI: 0.639-0.843), with a sensitivity of 52.2% and a specificity of 98.1%. **Conclusion** The ATM gene is hypermethylated in PBMCs of pancreatic cancer patients. Characterized by ATM hypermethylation, high CA19-9 and low ALB levels, the combined diagnostic model built with the random forest algorithm holds great clinical value for early pancreatic cancer diagnosis, and can make up for the diagnostic deficiency of the conventional biomarker CA19-9.

**Key words:** Pancreatic cancer; Peripheral blood mononuclear cells; DNA methylation; Early diagnosis; Molecular marker

近年来,国内外胰腺癌发病率和死亡率均呈明显上升趋势。其恶性程度较高,进展迅速,且早期症状不典型,大部分患者在首次就诊时已进展至中晚期<sup>[1]</sup>,导致胰腺癌患者预后差,五年生存率不足10%,远低于其他常见恶性肿瘤。因此提高胰腺癌的早期诊断率是改善胰腺癌患者整体预后的关键。然而胰腺癌的早期诊断是世界性难题,其早诊率仅为5%左右<sup>[2]</sup>。影像检查、以糖类抗原19-9(carbohydrate antigen 19-9, CA19-9)为代表的血清学标志物检查、“金标准”病理学检查等是现有的临床诊断方法,但是受到敏感度和准确度较低以及病理检查的有创性等因素影响,在早期胰腺癌的诊断方面未能获得理想效果<sup>[3-6]</sup>。因此,需寻找新型无创诊断标志物,推动实现胰腺癌的早期发现。

在现代肿瘤诊疗领域,肿瘤DNA甲基化标志物在癌症早期诊断中扮演着重要角色。DNA甲基化是指DNA甲基转移酶(DNA methyltransferases, DNMTs)将甲基加到DNA CpG序列中胞嘧啶的5'碳位,形成5-甲基胞嘧啶的过程,是一种DNA的

共价修饰形式。DNA甲基化异常可发生在肿瘤形成的早期阶段,通过检测与肿瘤发生发展相关的甲基化标志物如差异甲基化位点(differentially methylated points, DMPs),可以辅助癌症早期诊断<sup>[7]</sup>。有研究表明,外周血单个核细胞(peripheral blood mononuclear cells, PBMCs)中存在与肿瘤相关的特异性甲基化变化,为癌症早期诊断生物标志物研究提供了一个新方向<sup>[8]</sup>,但在胰腺癌中的诊断价值有待进一步探索。

随着机器学习大数据分析技术的发展,临床检验指标在肿瘤诊断中的作用日益凸显。本研究旨在利用随机森林算法联合PBMCs的DNA甲基化位点即共济失调毛细血管扩张突变(ataxia telangiectasia mutated, ATM)基因DMP甲基化水平、肝功能指标和CA19-9等指标开发一种胰腺癌早期无创辅助诊断模型,为胰腺癌筛查和分子诊断标志物研发提供实验依据和新思路,从而有助于提高胰腺癌的早期诊断率。

## 1 资料与方法

### 1.1 临床资料

#### 1.1.1 研究对象

选取2024年5月30日至10月30日于山东大学齐鲁医院确诊的118例胰腺癌患者临床资料,进行回顾性病例对照分析,其中男74例,女44例,38~75岁,平均58岁。随机选取14例胰腺癌样本作为发现集的胰腺癌组,其他样本以7:3的比例随机划分为训练集胰腺癌组( $n=72$ )和验证集胰腺癌组( $n=32$ );胰腺癌患者I期30例,II期36例,III期22例,IV期30例。CA19-9阴性的胰腺癌患者是指其患有胰腺癌但血清CA19-9水平低于37 U/mL。本研究CA19-9阴性的胰腺癌患者23例。同时选取山东大学齐鲁医院健康管理中心健康对照者73例,其中男48例,女25例,40~70岁,平均57岁,随机选取20例与发现集胰腺癌组性别、年龄相匹配的样本作为发现集健康对照组,其他样本以7:3的比例随机划分为训练集健康对照组( $n=37$ )和验证集健康对照组( $n=16$ )。本研究通过山东大学齐鲁医院伦理委员会批准许可(KYLL-202308-003),所有研究对象知情同意。

#### 1.1.2 纳入标准

胰腺癌组:①均为初诊患者,且无相关手术、放射和化疗治疗史;②经病理切除或者穿刺组织细胞学确认为胰腺癌且明确TNM分期。胰腺癌诊断依据国家卫生健康委办公厅《胰腺癌诊疗指南(2022年版)》<sup>[1]</sup>,TNM分期参考国际抗癌联盟/美国癌症联合会发布的第8版胰腺癌TNM分期标准<sup>[9]</sup>。健康对照组:①无肿瘤病史;②影像学CT检查等无明显异常;③无心、肝、肾等重大疾病。

#### 1.1.3 排除标准

胰腺癌组:①合并其他来源的恶性肿瘤;②患有自身免疫疾病;③有传染病或为携带者;④外周血标本的血液量过少。健康对照组:①有肿瘤病史;②患有自身免疫疾病;③有传染病或为携带者;④外周血标本的血液量过少。

#### 1.1.4 主要试剂与仪器

Histopaque-1077人淋巴细胞分离液购自美国Sigma公司,磷酸盐缓冲盐溶液(phosphate buffered saline, PBS)购自中国上海索莱宝科技有限公司,血液/细胞/组织基因组DNA提取试剂盒购自中国北京天根生化科技有限公司, Methylation-Gold™ Kit试剂盒购自美国Zymo Research公司,PCR扩增引物和焦磷酸测序引物购自中国苏州金唯智生物科技

有限公司,SYBR Green Pro Taq HS预混型qPCR试剂盒购自中国湖南艾科瑞生物工程有限公司。

Nanodrop 2000超微量分光光度计(美国Thermo公司),Illumina Infinium® Human Methylation 935 BeadChip测序平台(美国Illumina公司),PyroMark Q48焦磷酸测序仪(德国Qiagen公司),实时荧光定量PCR仪(美国Bio-Rad公司),台式高速离心机(中国上海索莱宝科技有限公司),-80℃超低温冰箱(美国Thermo公司)。

### 1.2 方法

#### 1.2.1 胰腺癌组和健康对照组的临床资料

收集两组的一般资料信息,包括性别、年龄、吸烟史、饮酒史和既往史等;临床检验指标包括CA19-9、丙氨酸氨基转移酶(alanine transaminase, ALT)、天门冬氨酸氨基转移酶(aspartate transaminase, AST)、 $\gamma$ -谷丙酰基转移酶(gamma-glutamyl transferase, GGT)、碱性磷酸酶(alkaline phosphatase, AKP)、总胆红素(total bilirubin, TBIL)、白蛋白(albumin, ALB)和血小板计数(platelet count, PLT)等。

#### 1.2.2 PBMCs的分离

使用Histopaque-1077人淋巴细胞分离液,严格按照说明书提取。步骤为:在15 mL离心管中加入约4 mL的分离液,用吸管将约2~4 mL的外周血沿管壁缓慢加入离心管,呈现出初步的分层状态,400 $\times$ g离心30 min。离心后,得到中间的白膜分层,用吸管小心吸取白膜层至新的15 mL离心管,加入10 mL PBS,洗涤1次,250 $\times$ g离心10 min,弃上清,底部沉淀即为PBMCs,放置于-80℃冰箱备用。

#### 1.2.3 DNA提取与转化

使用血液/细胞/组织基因组DNA提取试剂盒,严格按照说明书提取DNA,提取后用Nanodrop 2000超微量分光光度计测定浓度。使用Methylation-Gold™ Kit试剂盒,严格按照说明书对DNA进行亚硫酸氢盐转化,转化后的DNA放入-80℃冰箱备用。

#### 1.2.4 935K甲基化芯片检测

胰腺癌组(14例)和健康对照组(20例)PBMCs的DNA甲基化水平采用Illumina Infinium® Human Methylation 935 BeadChip进行全基因组(覆盖基因的不同区域,包括CpG岛、启动子、编码区、开放染色质和增强子等约935 000个CpG位点)甲基化测序,使用R包中的ChAMP包对测序数据进行分析,确定高甲基化和低甲基化差异位点。 $\beta$ 值表示DNA的甲基化水平, $\Delta\beta$ 表示胰腺癌组的平均甲基化水平减去健康对照组的平均甲基化水平。两组间的差异甲基化位点被定义为: $|\Delta\beta| \geq 0.1$ ,且 $P < 0.01$

的甲基化位点。

### 1.2.5 引物设计

由 PyroMark Assay Design 2.0 软件依据 935K 芯片中候选目标基因甲基化位点(即 ATM 基因 DMP)的“forward sequence”碱基序列进行 PCR 扩增引物(包括正向引物和反向引物)和焦磷酸测序引物的设计,生成的碱基序列为:正向引物 GATAT-AGTGAAAGATTTATTTAATGGTGTT;反向引物(5' 修饰)CACCAATATTTTCTCACAAAAAAAT-ATCT;测序引物 TGAAAGATTTATTTAATGGT-GTTA。

### 1.2.6 焦磷酸测序

取亚硫酸氢盐转化后的 DNA 标本,在冰上融化 SYBR Green Pro Taq HS 预混型 qPCR 试剂盒。按照如下条件配制每个八连排孔的反应体系:6.5  $\mu\text{L}$  SYBR Green Pro Taq HS 预混型 qPCR 试剂盒试剂、1.3  $\mu\text{L}$  转化后的 DNA 模板、0.26  $\mu\text{L}$  正向引物(10 mmol)、0.26  $\mu\text{L}$  反向引物(10 mmol)和 4.68  $\mu\text{L}$  无菌无酶水。

随后将配置好的体系(13  $\mu\text{L}$ )及亚硫酸氢盐转化后的 DNA 模板(1.3  $\mu\text{L}$ )加入 PCR 反应的八连排管孔中,盖紧盖子。将用于 PCR 反应的八连排放置于实时荧光定量 PCR 仪中,对转化后的 DNA 进行 PCR 扩增,反应程序:95  $^{\circ}\text{C}$  预变性 3 min,1 次;95  $^{\circ}\text{C}$  变性 15 s,35 次;55  $^{\circ}\text{C}$  退火 15 s,35 次;72  $^{\circ}\text{C}$  延伸 15 s,35 次;72  $^{\circ}\text{C}$  终末延伸 1 min,1 次。

向样本盘中的各孔加入 10  $\mu\text{L}$  扩增后产物和 3  $\mu\text{L}$  磁珠,使用 PyroMark Q48 焦磷酸测序仪检测目标基因位点的甲基化水平。

### 1.2.7 特征筛选

通过随机森林算法完全随机化分组,以 7:3 的比例划分为训练集和验证集,在训练集利用随机森林构建本研究 11 个变量的全变量模型并训练模型,调整超参数,建立包含全部特征的预测模型,得到各个特征的变量重要性;根据变量重要性,逐个纳入变量到随机森林模型中,并计算纳入不同数量变量时的模型评价指标;选择随着

变量增加,而模型性能提高不明显时的模型所包含的最少数量特征,作为后续模型构建的特征变量。

### 1.2.8 模型构建

利用随机森林算法在训练集上进行重抽样成五份,用于构建超参数网格搜索,寻找模型最优的超参数组合;将超参数应用于训练集进行模型的训练,其截止值(Cut-off 值)设定为使约登指数值达到最大时的模型预测值;固定模型参数应用于验证集来验证模型的性能。

## 1.3 统计学处理

采用 GraphPad Prism 8.3.0 和 SPSS 26.0 统计学软件。采用 Shapiro-Wilk 方法检验计量资料是否符合正态分布。符合正态分布的计量资料以  $\bar{x}\pm s$  表示,两组间比较采用两独立样本  $t$  检验;偏态分布的计量资料以  $M(P_{25}, P_{75})$  表示,两组间比较采用 Mann-Whitney  $U$  检验。分类变量用  $n(\%)$  表示,两组间比较采用  $\chi^2$  检验。随机森林机器学习算法在 R 4.4.1 上进行,采用随机抽样,其中 70% 的数据集作为训练集,30% 作为验证集。采用 MedCalc 20.0.22 软件绘制受试者工作特征(receiver operating characteristic, ROC)曲线,曲线下面积(area under the curve, AUC)用于分析联合模型对胰腺癌患者的辅助诊断价值,采用 DeLong 检验比较两组 AUC 值间的差异。使用 R 包的京都基因与基因组百科全书(kyoto encyclopedia of genes and genomes, KEGG)进行 KEGG 富集分析,富集分析使用 Fisher 精确检验。检验水准  $\alpha=0.05$ (双侧)。

## 2 结果

### 2.1 临床资料比较

胰腺癌组和健康对照组性别、年龄差异无统计学意义( $P>0.05$ );胰腺癌组血清学肿瘤标志物 CA19-9 表达水平显著高于健康对照组( $P<0.05$ );其他部分指标健康对照组缺失未统计差异,见表 1。

表 1 胰腺癌组与健康对照组的临床资料比较

Table 1 Comparison of clinical data between pancreatic cancer group and healthy control group

| 项目          | 胰腺癌组              |                   |                   |                   | 健康对照组             |                   |                   |                  | $t/Z/\chi^2$ | $P$     |
|-------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|--------------|---------|
|             | 发现集<br>( $n=14$ ) | 训练集<br>( $n=72$ ) | 验证集<br>( $n=32$ ) | 总计<br>( $n=118$ ) | 发现集<br>( $n=20$ ) | 训练集<br>( $n=37$ ) | 验证集<br>( $n=16$ ) | 总计<br>( $n=73$ ) |              |         |
| 年龄          | 62.1 $\pm$ 6.9    | 57.0 $\pm$ 8.0    | 58.8 $\pm$ 7.0    | 58.1 $\pm$ 7.8    | 61.1 $\pm$ 6.7    | 55.3 $\pm$ 5.2    | 54.7 $\pm$ 6.1    | 56.7 $\pm$ 6.1   | 1.340        | 0.181 9 |
| 性别/ $n(\%)$ |                   |                   |                   |                   |                   |                   |                   |                  | 0.181        | 0.671 0 |
| 男           | 7(50.0)           | 47(65.3)          | 20(62.5)          | 74(62.7)          | 10(50.0)          | 28(75.7)          | 10(62.5)          | 48(65.8)         |              |         |
| 女           | 7(50.0)           | 25(34.7)          | 12(37.5)          | 44(37.3)          | 10(50.0)          | 9(24.3)           | 6(37.5)           | 25(34.2)         |              |         |

续表

| 项目                 | 胰腺癌组                        |                           |                           |                           | 健康对照组             |                       |                   |                       | $t/Z/\chi^2$ | $P$    |
|--------------------|-----------------------------|---------------------------|---------------------------|---------------------------|-------------------|-----------------------|-------------------|-----------------------|--------------|--------|
|                    | 发现集<br>( $n=14$ )           | 训练集<br>( $n=72$ )         | 验证集<br>( $n=32$ )         | 总计<br>( $n=118$ )         | 发现集<br>( $n=20$ ) | 训练集<br>( $n=37$ )     | 验证集<br>( $n=16$ ) | 总计<br>( $n=73$ )      |              |        |
| CA19-9<br>/(U/mL)  | 229.5<br>(40.4,<br>6 070.5) | 209.5<br>(43.7,<br>800.0) | 159.0<br>(25.4,<br>608.8) | 186.5<br>(43.3,<br>800.0) | 13.0±6.6          | 9.8<br>(7.6,<br>14.7) | 10.0±4.4          | 9.8<br>(7.5,<br>15.4) | 8.709        | <0.001 |
| 有吸烟史/<br>$n$ (%)   | 4(28.6)                     | 29(40.3)                  | 8(25.0)                   | 41(34.7)                  | —                 | —                     | —                 | —                     |              |        |
| 有饮酒史/<br>$n$ (%)   | 6(42.9)                     | 25(34.7)                  | 9(28.1)                   | 40(33.9)                  | —                 | —                     | —                 | —                     |              |        |
| TNM 分期/<br>$n$ (%) |                             |                           |                           |                           |                   |                       |                   |                       |              |        |
| I                  | 2(14.3)                     | 19(26.4)                  | 9(28.1)                   | 30(25.4)                  | —                 | —                     | —                 | —                     |              |        |
| II                 | 5(35.7)                     | 19(26.4)                  | 12(37.5)                  | 36(30.5)                  | —                 | —                     | —                 | —                     |              |        |
| III                | 2(14.3)                     | 15(20.8)                  | 5(15.6)                   | 22(18.6)                  | —                 | —                     | —                 | —                     |              |        |
| IV                 | 5(35.7)                     | 19(26.4)                  | 6(18.8)                   | 30(25.4)                  | —                 | —                     | —                 | —                     |              |        |

## 2.2 935K 甲基化芯片分析差异甲基化位点及其所在基因的通路富集分析

935K 甲基化芯片结果显示,在 $|\Delta\beta| \geq 0.1$ ,且 $P < 0.01$ 的筛选条件下,共计发现 132 个差异甲基化位点,其中高甲基化位点 28 个,低甲基化位点 104 个(图 1A)。转录起始位点(transcription start site, TSS)上下游附近的 CpG 位点在调控基因表达中

发挥重要作用,分析两组间 TSS 上下游区域 CpG 位点甲基化差异情况,显示胰腺癌组 TSS 上下游附近的 CpG 位点整体甲基化水平明显低于健康对照组(图 1B)。KEGG 分析显示,差异甲基化位点所在的基因主要与癌症中的基因表达调控、细胞周期和信号转导、脂质代谢等通路相关(图 1C)。

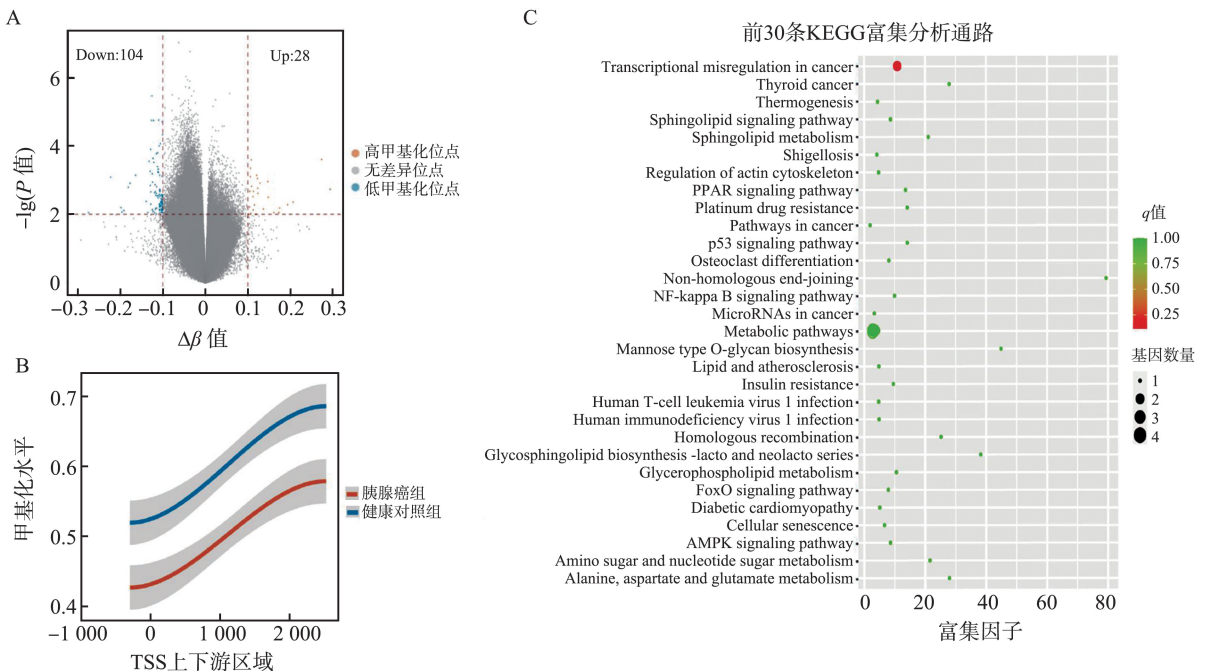


图 1 935K 甲基化芯片发现差异位点与 KEGG 通路富集分析

A: 935K 芯片差异甲基化位点火山图; B: TSS 区域附近的差异 CpG 位点甲基化水平; C: 差异甲基化位点 KEGG 富集分析的前 30 条通路图。

Figure 1 935K methylation microarray found differential sites and KEGG pathway enrichment analysis

A: 935K chip differential methylation site volcano map; B: The methylation level of CpG sites in the vicinity of the TSS region; C: Top 30 pathway maps for KEGG enrichment analysis of differential methylation sites.

### 2.3 甲基化标志物的筛选及验证

共筛选到 5 个高甲基化基因 DMPs, 将 5 个 DMPs 的 AUC 值依次排序(图 2A), 显示两性蛋白(amphiphysin, AMPH)基因和 ATM 基因 DMPs 的 AUC 分别为 0.882 和 0.871, 能较好的区分胰腺癌患者和健康对照者。溶质载体家族 29 成员 3(solute carrier family 29 member 3, SLC29A3)基因、鸟苷酸结合蛋白 5(guanylate binding protein 5, GBP5)基因和突变 S 同源物 3(muts homolog 3, MSH3)基因 DMPs 的 AUC 值分别为 0.807、0.760 和 0.760, 明显低于 AMPH 和 ATM 基因 DMPs, 因

此未纳入这 3 个基因的 DMPs 进一步验证。针对上述诊断效能较好的 2 个 DMPs, 采用 PyroMark Assay Design 2.0 软件对候选甲基化位点进行引物序列的设计。但是 AMPH 基因 DMP 的反向引物序列无法生成, 该基因 DMP 无法进行后续的焦磷酸测序验证; ATM 基因 DMP 在引物序列设计过程中表现良好且未提示异常信息, 将该位点进行后续的验证。将 ATM 基因 DMP 在训练集和验证集样本中进行焦磷酸测序验证, 结果显示, ATM 基因 DMP 的甲基化水平在胰腺癌患者中显著高表达( $P < 0.001$ ), 趋势与 935K 芯片发现集一致(图 2B、C)。

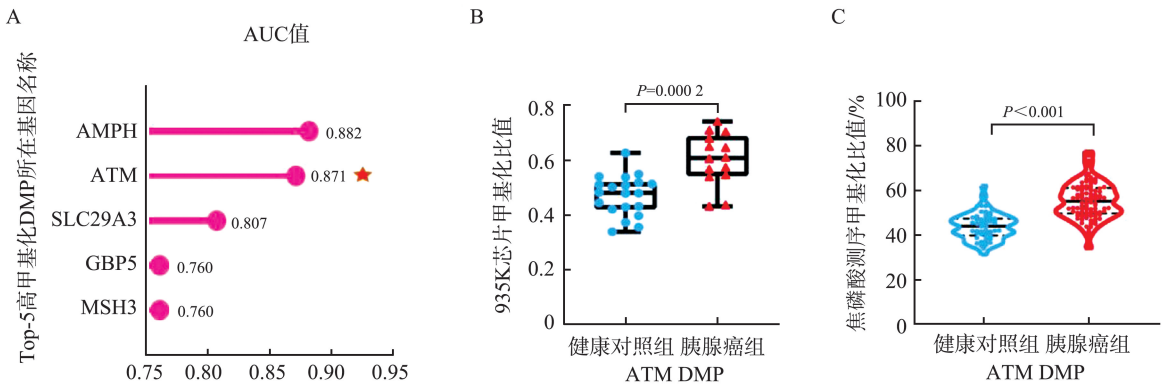


图 2 甲基化位点 ATM 基因 DMP 的筛选及验证

A: 发现集筛选到 5 个基因 DMPs 的 AUC 值排序棒棒糖图; B: ATM 基因 DMP 在 935K 芯片中两组的甲基化水平比较; C: ATM 基因 DMP 在训练集和验证集中的表达。

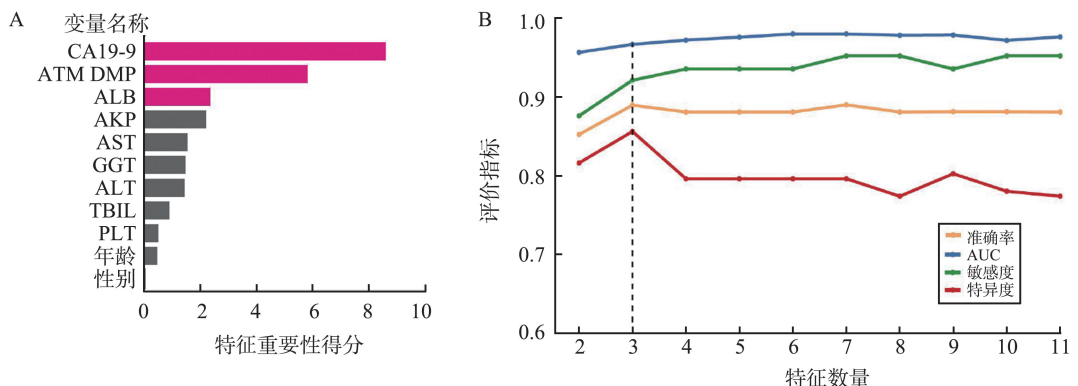
Figure 2 Screening and validation of ATM gene DMP at methylation site

A: The sorted bar chart of AUC values for the 5 DMPs genes identified through concentration screening; B: Comparison of methylation levels of ATM gene DMP in the two groups on 935K chip; C: The expression of ATM gene DMP was verified in both the training set and the validation set.

### 2.4 随机森林模型变量特征的筛选及纳入

将胰腺癌组和健康对照组临床资料进行统计, 可用于变量筛选的变量共计 11 个: ATM 基因 DMP 甲基化水平、CA19-9、性别、年龄、ALT、AST、GGT、AKP、TBIL、ALB 和 PLT。通过随机森林算法, 计算每个变量特征的重要性, 然后按照特征重要性排序(图 3A), 依次纳入模型, 通过观察模型在训练集的性能表现, 如准确度、AUC、敏感度和特异度等指标, 经综合判断, 最终筛选获得 ATM 基因 DMP 甲

基化、CA19-9 和 ALB 水平 3 个变量(图 3B)。在训练集和验证集样本中, 与健康对照组相比, 胰腺癌组 ATM 基因 DMP 甲基化和 CA19-9 水平均升高 [ATM 基因 DMP 为  $55.5 \pm 8.4$  vs.  $44.0 \pm 6.0$ ,  $t = 9.946$ ,  $P < 0.001$ ; CA19-9 为  $186.5 (43.7, 749.8)$  vs.  $9.7 (7.2, 13.9)$ ,  $Z = 7.656$ ,  $P < 0.001$ ]; ALB 水平降低 [ $42.7 (39.9, 45.2)$  vs.  $46.0 \pm 2.1$ ,  $Z = -6.178$ ,  $P < 0.001$ ], 差异均有统计学意义(图 2C、3C、D)。



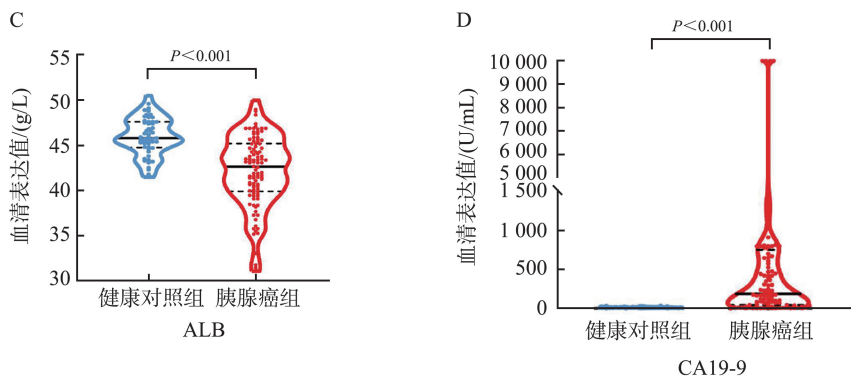


图3 基于随机森林的变量特征的筛选及纳入

A: 11个特征随机森林算法重要性得分; B: 模型逐步纳入前 Top-N 个指标后的性能比较; C: 两组在训练集和验证集中 ALB 血清表达值; D: 两组在训练集和验证集中 CA19-9 血清表达值。

Figure 3 Screening and inclusion of variable features based on random forest

A: Importance scores of 11 features in the random forest algorithm; B: Performance comparison after gradually incorporating the Top-N indicators into the model; C: The ALB serum expression values of the two groups in the training set and validation set; D: The CA19-9 serum expression values in the training set and validation set of the two groups.

## 2.5 基于随机森林的联合诊断模型构建及其诊断效能的评价

以 ATM 基因 DMP 甲基化、CA19-9 和 ALB 水平为特征,借助随机森林算法构建联合诊断模型(“AmCA”),训练集模型对胰腺癌诊断效能的 AUC 值为 0.992(95%CI:0.952~1.000);单独 ATM 基因 DMP 的 AUC 值为 0.861(95%CI:0.782~0.920),单独 CA19-9 的 AUC 值为 0.893(95%CI:0.819~0.944),单独 ALB 的 AUC 值为 0.778(95%CI:0.689~0.852),可见“AmCA”模型对胰腺癌的诊断效能优于单独 CA19-9( $Z=3.079, P=0.0021$ ,见图 4A)。训练集模型的 Cut-off 值设定为约登指数值达到最大时的模型预测值,则“AmCA”模型的 Cut-off

值为 0.705。此时“AmCA”模型在训练集的敏感度为 91.7%(95%CI:82.7%~96.9%),特异度为 100%(95%CI:90.5%~100.0%)。

验证集“AmCA”模型 AUC 值为 0.982(95%CI:0.895~1.000),敏感度为 84.4%(95%CI:67.2%~94.7%),特异度为 93.8%(95%CI:69.8%~99.8%);单独 ATM 基因 DMP AUC 值为 0.916(95%CI:0.799~0.976);单独 CA19-9 的 AUC 值为 0.840(95%CI:0.705~0.930),单独 ALB 的 AUC 值为 0.861(95%CI:0.731~0.944),可见“AmCA”模型对胰腺癌的诊断效能优于单独的传统血清学标志物 CA19-9( $Z=2.515, P=0.0119$ ),见图 4B)。

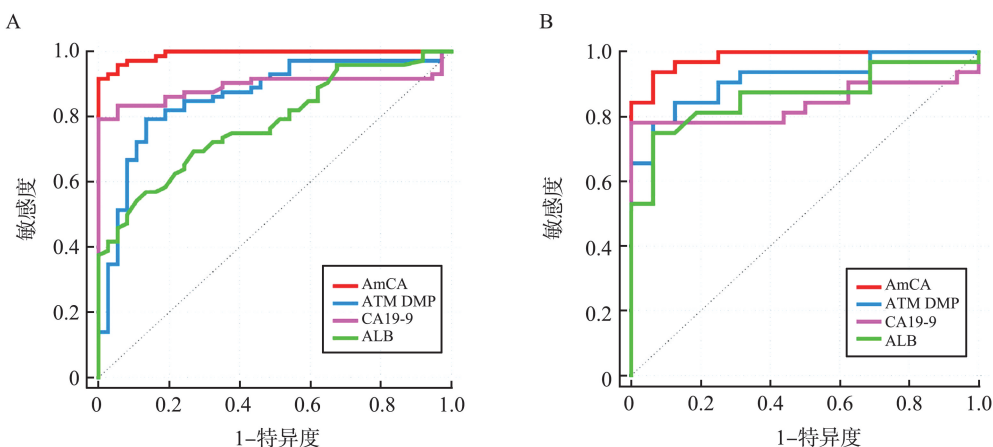


图4 “AmCA”模型与单个指标在训练集(A)和验证集(B)上的 ROC 曲线

Figure 4 ROC curves of the joint model and individual indicators on the training set (A) and the validation set (B)

验证集“AmCA”模型对 I 期胰腺癌的诊断效能 AUC 为 1.000(95%CI:0.863~1.000),敏感度为 100%(95%CI:66.4%~100.0%),特异度为 93.8%

(95%CI:69.8%~99.8%),见图 5A;对 II 期胰腺癌的诊断效能 AUC 为 0.979(95%CI:0.840~1.000),敏感度为 91.7%(95%CI:61.5%~99.8%),特异度

为 93.8% (95% CI: 69.8% ~ 99.8%), 见图 5B; 对 III 和 IV 期胰腺癌的诊断效能 AUC 为 0.972 (95% CI: 0.824 ~ 1.000), 敏感度为 63.6% (95% CI: 30.8% ~

89.1%), 特异度为 93.8% (95% CI: 69.8% ~ 99.8%), 见图 5C。“AmCA”模型在验证集各期的诊断 AUC 均高于单独 CA19-9 的 AUC。

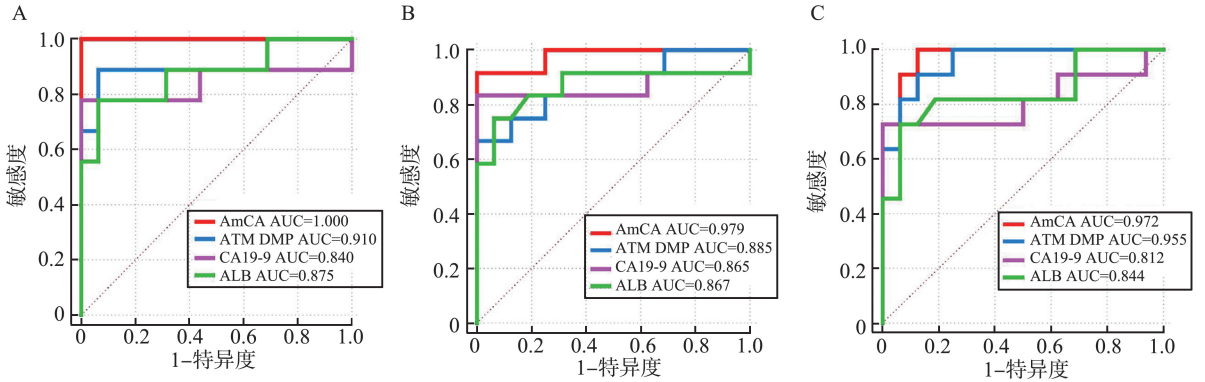


图 5 验证集中“AmCA”模型与单个变量对胰腺癌各期的诊断 ROC 曲线

A: I 期; B: II 期; C: III 和 IV 期。

Figure 5 The ROC curves of the combined model and a single variable for each stage of pancreatic cancer of validation set  
A: Stage I ; B: Stage II ; C: Stage III and IV.

### 2.6 “AmCA”模型对 CA19-9 阴性胰腺癌患者的诊断效能

当“AmCA”模型预测值的 Cut-off 值设定为 0.705 时,“AmCA”模型对区分健康对照者和

CA19-9 阴性胰腺癌患者的 AUC 为 0.751 (95% CI: 0.639 ~ 0.843)。“AmCA”模型对 CA19-9 阴性胰腺癌患者的诊断敏感度为 52.2% (95% CI: 30.6% ~ 73.2%), 特异度为 98.1% (95% CI: 89.9% ~ 100.0%)。见图 6。

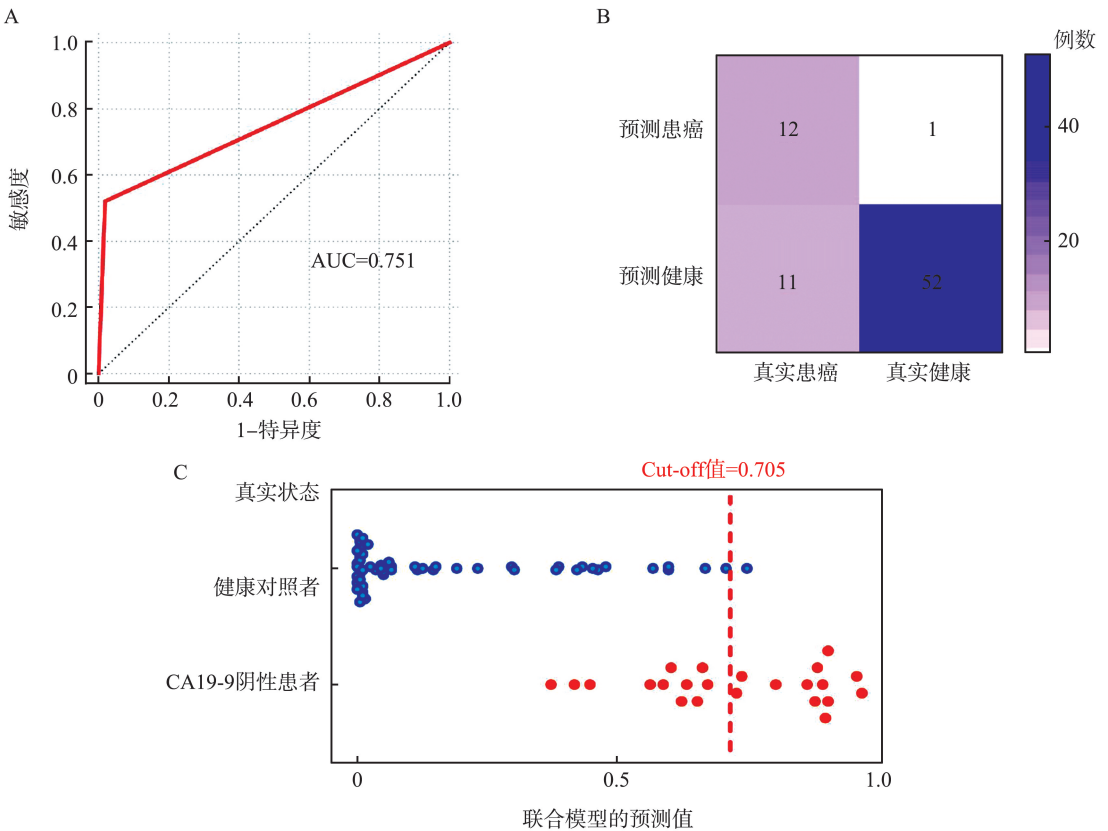


图 6 “AmCA”模型对 CA19-9 阴性胰腺癌患者的诊断效能

A: “AmCA”模型对 CA19-9 阴性患者检测的 ROC 曲线图; B: “AmCA”模型诊断 CA19-9 阴性患者的混淆矩阵图; C: “AmCA”模型对 CA19-9 阴性患者的诊断散点图。

Figure 6 Diagnostic efficacy of "AmCA" combined model in patients with CA19-9 negative pancreatic cancer  
A: The ROC curve of "AmCA" combined model for CA19-9 negative patients; B: Confusion matrix diagram of "AmCA" combined model for diagnosis of CA19-9 negative patients; C: Diagnostic scatterplot of "AmCA" combined model for CA19-9 negative patients.

### 3 讨论

DNA 甲基化是一种重要的表观遗传修饰方式,可以作为肿瘤诊断的生物标志物。近年来,大量的胰腺癌组织和细胞系研究解析了胰腺癌的异常 DNA 甲基化模式,可作为胰腺癌的诊断或预后工具。有研究报道,含钾通道四聚化结构域蛋白 8 (potassium channel tetramerization domain containing 8, KCTD8) 和高亮氨酸重复跨膜神经蛋白 1 (leucine-rich repeat transmembrane neuronal protein 1, LRRTM1) 基因的启动子区域高甲基化状态影响基因发挥抑癌功能,并且表明基因的异常甲基化状态可作为胰腺癌诊断生物标志物<sup>[10-11]</sup>。目前对血液中胰腺癌甲基化生物标志物的研究主要聚焦 ctDNA 或循环游离 DNA (cell-free DNA, cfDNA)。有研究检测 4 个 ctDNA 抑癌基因甲基化联合 CA19-9 诊断胰腺癌的敏感度为 100%, AUC 值为 0.96, 但特异度仅为 77.81%<sup>[12]</sup>。Wu 等<sup>[13]</sup> 基于血浆 ctDNA 开发的靶向甲基化测序分类器诊断胰腺癌的 AUC 值为 0.91, 联合 CA19-9 后 AUC 值提升至 0.94。Ben-Ami 等<sup>[14]</sup> 将 9 个 cfDNA 甲基化标志物联合金属蛋白酶抑制因子 1 (metallopeptidase inhibitor 1, TIMP1)、CA19-9, 在验证集中将单独 CA19-9 的 AUC 值从 0.82 提升至 0.86。以上研究表明, ctDNA/cfDNA 甲基化用于胰腺癌诊断具备一定的应用前景, 然而 ctDNA/cfDNA 在血浆中的丰度较低、半衰期短以及检测成本较高, 严重制约甲基化标志物的临床应用<sup>[15-16]</sup>。

PBMCs 可以反映癌症进展期间的系统免疫反应<sup>[17]</sup>, 作为癌症早期诊断的潜在工具具有多种优势。首先, 与 ctDNA 或 cfDNA 相比, 外周血中的 PBMCs 含量高且提取过程简单<sup>[8]</sup>; 其次, PBMCs 可以反映全身免疫反应以及癌症发生和发展的过程, 使得其用于发现并鉴定出早期癌症检测和个性化癌症管理的新型免疫相关生物标志物成为可能<sup>[18]</sup>。在描述 PBMCs 的免疫学特征时, 多个研究都对特定癌症中的 PBMCs 多样性进行描述<sup>[19-20]</sup>。有研究报道, 胃癌、肺癌、结直肠癌和乳腺癌患者全血中 PBMCs 的 DNA 甲基化水平用于早期诊断具有较好的检测性能<sup>[21-24]</sup>。本研究利用 935K 甲基化芯片发现, 健康对照者与胰腺癌患者 ATM 基因 DMP 的甲基化水平存在显著差异, 通过焦磷酸测序在更大的队列中进行了差异的验证, 明确了该位点在胰腺癌患者 PBMCs 中显著高表达, 对区分胰腺癌有很好

的潜力。基于随机森林算法筛选到 ATM 基因 DMP 甲基化水平、CA19-9 和 ALB3 个变量, 成功构建了联合诊断“AmCA”模型, 在胰腺癌诊断中具有较高的诊断性能 (训练集 AUC = 0.992, 验证集 AUC = 0.982), 其在早期阶段胰腺癌诊断中也表现出较高的诊断效能 (验证集 I 期 AUC = 1.000, II 期 AUC = 0.979), 优于单一的 CA19-9 或甲基化标志物; 同时, 该联合模型在 CA19-9 阴性的胰腺癌患者中也具有一定的诊断性能 (AUC = 0.751), 可以作为现有肿瘤标志物 CA19-9 的重要补充, 其具备作为胰腺癌早筛和早诊工具的潜力。

有研究报道, ATM 是一种重要的 DNA 损伤修复基因, 属于磷酸肌醇 3 激酶相关激酶 (phosphatidylinositol-3-kinase-related kinases, PIKK) 家族, 主要通过磷酸化一系列下游蛋白参与 DNA 修复和维持基因组稳定性<sup>[25]</sup>。研究表明, 约 5% ~ 10% 的胰腺癌患者存在 ATM 基因突变<sup>[26]</sup>。ATM 基因的突变可导致细胞 DNA 修复能力减弱, 与胰腺癌的发生风险增加密切相关<sup>[27-28]</sup>, 影响胰腺癌患者对 DNA 损伤剂 (如铂类药物) 和靶向药 (PARP 抑制剂) 的敏感性<sup>[25]</sup>。ATM 基因除了在胰腺癌细胞中有所改变外, 本研究结果发现, ATM 基因外显子区域的 DMP 在胰腺癌患者的 PBMCs 中呈现出高甲基化状态, 表明 ATM 基因甲基化在胰腺癌患者的免疫细胞群中发生了重要的改变, 并可作为胰腺癌诊断的分子标志物。也有研究表明, 乳腺癌患者的外周血白细胞 ATM 基因启动子区位点的甲基化水平明显升高<sup>[29]</sup>, 支持了恶性肿瘤发生进展中免疫细胞 ATM 基因甲基化模式发生改变, 与本研究结果相同。已有研究报道, ATM 在抗肿瘤免疫监视中具有重要的作用, 在一项对小鼠 B 细胞淋巴瘤的研究中发现, ATM 激活可上调肿瘤细胞表面 CD155 等免疫刺激分子的表达, 增强 NK 细胞和 CD4<sup>+</sup>/CD8<sup>+</sup> T 细胞的抗肿瘤免疫作用, 表明 ATM 基因与机体免疫细胞之间存在相互作用<sup>[30]</sup>。胰腺癌患者 PBMCs 的 ATM 基因高甲基化可能在癌症免疫监视和免疫逃逸中发挥作用, 但具体的调控作用及其机制尚需进一步深入研究。

有研究表明, 机器学习算法联合临床检验指标明显提高了联合诊断模型对癌症的检测能力。例如, 一项研究通过机器学习联合磁共振成像 (magnetic resonance imaging, MRI) 和经常被忽略的潜在相关临床参数 (如前列腺特异性抗原、前列腺体积和年龄等), 提高了模型对于前列腺癌的诊断准确性, 多模式 (结合了机器学习和临床参数) 的联合诊

断模型对早期诊断检测效能优于单一的临床指标和 MRI 机器学习算法<sup>[31]</sup>。在另一项研究中,应用机器学习算法成功鉴定出了 9 个关键基因用于诊断转移性结直肠癌,精确率-召回率曲线下面积平均提高了 21.16%,显著提高了预测模型的性能。由此可见,先进的特征选择方法,结合机器学习,在解决医学诊断中的关键作用<sup>[32]</sup>。CA19-9 和 ALB 作为常规检验指标,已有报道将其用于联合建模诊断早期胰腺癌<sup>[33]</sup>。在本研究中,基于随机森林算法依据变量的特征重要性评分,由高到低依次纳入变量,此时算法给出在联合模型达到较高诊断效能(之后若继续纳入变量但模型的准确率和 AUC 等性能指标提升不再明显)时特征数量最简洁的 3 个变量(ATM 基因 DMP、CA19-9 和 ALB),从而避免了潜在的模型过拟合问题,并最终利用这 3 个变量成功构建了具有良好的胰腺癌早期诊断潜力的联合诊断模型,优于单一传统血清标志物 CA19-9 的诊断性能。

CA19-9 作为临床上常用的血液诊断标志物,由于其在 Lewis 抗原阴性基因型的患者中无法发挥作用,从而造成假阴性,限制了其在胰腺癌早期诊断中的应用<sup>[6]</sup>。有报道表明,Lewis 抗原阴性的胰腺癌患者拥有更差的预后和更高的转移率<sup>[34]</sup>。因此,尽早识别出此类胰腺癌患者具有重要意义。本研究中构建的联合诊断“AmCA”模型对 CA19-9 阴性胰腺癌患者具有一定的诊断效能,在本研究设定的阈值下能够在特异度达到 98.1% 的情况下,诊断出 52.2% 的 CA19-9 阴性胰腺癌患者,提高了对于阴性患者的诊断效能,可以一定程度上弥补现有传统肿瘤标志物 CA19-9 的不足,具备作为辅助胰腺癌早期诊断工具的潜能。CA19-9 作为“AmCA”模型中具有重要贡献度的变量之一,针对 CA19-9 阴性患者的亚组分析显得尤为重要,当 CA19-9 的诊断效能受限时,联合诊断模型仍然具备 52.2% 的诊断敏感度,表明其较传统肿瘤标志物 CA19-9 的诊断性能有一定的提升作用,同时这也指导了我们未来优化联合模型的方向:通过纳入更多的变量特征,包括甲基化标志物(如 SLC29A3 基因 DMP)、免疫相关检测指标(如细胞因子、免疫球蛋白)及胰腺功能相关检查指标(如脂肪酶、淀粉酶)等,以提高针对 CA19-9 阴性胰腺癌患者的诊断敏感度和 AUC 值,并计划在更大的队列中进一步验证。

本研究将通过多中心研究,针对胰腺癌高危人群<sup>[35]</sup>(包括胰腺癌家族史、慢性胰腺炎、胰腺囊性肿瘤及新发糖尿病等)开展更大规模的前瞻性研究验证。将通过不断优化联合诊断模型,进一步明确其

在高危人群中的筛查性能及其临床诊断价值。

综上所述,本研究开发了一个基于外周血 PBMCs 的 DNA 甲基化 ATM 基因 DMP 与 CA19-9 和 ALB 水平的联合诊断模型“AmCA”,该模型可以显著提高胰腺癌早期阶段的诊断效能,一定程度上弥补了现有常用的血清学标志物 CA19-9 的不足,有助于临床推广使用。

## 参考文献:

- [1] 国家卫生健康委办公厅. 胰腺癌诊疗指南(2022 年版)[J]. 临床肝胆病杂志, 2022, 38(5): 1006-1015.
- [2] 中华医学会肿瘤学分会早诊早治学组. 中华医学会肿瘤学分会胰腺癌早诊早治专家共识[J]. 临床肝胆病杂志, 2020, 36(12): 2675-2680.  
Early Diagnosis and Treatment Group, the Oncology Committee of Chinese Medical Association. Expert consensus of Oncology Committee of Chinese Medical Association in early diagnosis and treatment of pancreatic cancer [J]. Journal of Clinical Hepatology, 2020, 36(12): 2675-2680.
- [3] 中国医师协会临床精准医疗专业委员会, 中国抗癌协会肿瘤胰腺病学专业委员会. 早期胰腺癌分子诊断专家共识(2023 年版)[J]. 临床肝胆病杂志, 2024, 40(3): 473-477.  
Professional Committee on Clinical Precision Medicine, Chinese Medical Doctor Association, The Society of Pancreas Cancer, Chinese Anti-Cancer Association. Expert consensus on the molecular diagnosis of early-stage pancreatic cancer (2023 edition) [J]. Journal of Clinical Hepatology, 2024, 40(3): 473-477.
- [4] Zhang L, Sanagapalli S, Stoita A. Challenges in diagnosis of pancreatic cancer [J]. World J Gastroenterol, 2018, 24(19): 2047-2060.
- [5] Müller MF, Meyenberger C, Bertschinger P, et al. Pancreatic tumors: evaluation with endoscopic US, CT, and MR imaging [J]. Radiology, 1994, 190(3): 745-751.
- [6] Bestari MB, Joewono IR, Syam AF. A quest for survival: a review of the early biomarkers of pancreatic cancer and the most effective approaches at present [J]. Biomolecules, 2024, 14(3): 364.
- [7] 中国抗癌协会肿瘤标志专业委员会. 肿瘤 DNA 甲基化标志物检测及临床应用专家共识(2024 版)[J]. 中国癌症防治杂志, 2024, 16(2): 129-143.
- [8] Li YZ, Fan ZY, Meng YF, et al. Blood-based DNA methylation signatures in cancer: a systematic review [J]. Biochim Biophys Acta BBA Mol Basis Dis, 2023, 1869(1): 166583.
- [9] Amin MB, Greene FL, Edge SB, et al. The eighth edition AJCC cancer staging manual: continuing to build a

- bridge from a population-based to a more “personalized” approach to cancer staging[J]. *CA Cancer J Clin*, 2017, 67(2): 93-99.
- [10] 黄攀丝. KCTD8 基因在胰腺癌中的表观遗传调控及功能机制研究[D]. 新乡: 新乡医学院, 2023.
- [11] 张帆, 张美英, 高媛媛, 等. LRRTM1 基因甲基化是胰腺癌的潜在诊断和预后标志物[J]. *胃肠病学和肝病学杂志*, 2024, 33(5): 585-589.
- ZHANG Fan, ZHANG Meiyang, GAO Aiai, et al. LRRTM1 gene methylation is a potential diagnostic and prognostic marker of pancreatic cancer[J]. *Chinese Journal of Gastroenterology and Hepatology*, 2024, 33(5): 585-589.
- [12] 辛万鹏. 二代测序技术检测 ctDNA 抑癌基因甲基化在胰腺癌诊治中的应用价值[D]. 南昌: 南昌大学, 2022.
- [13] Wu H, Guo S, Liu X, et al. Noninvasive detection of pancreatic ductal adenocarcinoma using the methylation signature of circulating tumour DNA[J]. *BMC Med*, 2022, 20(1): 458.
- [14] Ben-Ami R, Wang QL, Zhang J, et al. Protein biomarkers and alternatively methylated cell-free DNA detect early stage pancreatic cancer[J]. *Gut*, 2024, 73(4): 639-648.
- [15] Anagnostou V, Velculescu VE. Pushing the boundaries of liquid biopsies for early precision intervention[J]. *Cancer Discov*, 2024, 14(4): 615-619.
- [16] 王金, 蒋慧, 郑建明. 循环肿瘤 DNA 在胰腺癌诊疗中的应用及前景[J]. *中华胰腺病杂志*, 2022, 22(2): 152-155.
- [17] Zhang YH, Zhou XR, Zhong Y, et al. Pan-cancer scRNA-seq analysis reveals immunological and diagnostic significance of the peripheral blood mononuclear cells[J]. *Hum Mol Genet*, 2024, 33(4): 342-354.
- [18] Celešnik H, Potočnik U. Peripheral blood transcriptome in breast cancer patients as a source of less invasive immune biomarkers for personalized medicine, and implications for triple negative breast cancer[J]. *Cancers (Basel)*, 2022, 14(3): 591.
- [19] Xue J, Yang S, Zhang SS, et al. Deciphering the multifaceted immune landscape of unresectable primary liver cancer to predict immunotherapy response[J]. *Adv Sci (Weinh)*, 2024, 11(47): e2309631.
- [20] Ran X, Zheng J, Chen L, et al. Single-cell transcriptomics reveals the heterogeneity of the immune landscape of IDH-wild-type high-grade gliomas[J]. *Cancer Immunol Res*, 2024, 12(2): 232-246.
- [21] 王景. 基于 PBMC DNA 甲基化的多位点联合诊断模型在胃癌早期诊断中的应用研究[D]. 济南: 山东大学, 2022.
- [22] 刘士标, 张淑君, 李培龙, 等. cg20657709 位点甲基化对肺腺癌早期诊断的初步探讨[J]. *山东大学学报(医学版)*, 2023, 61(4): 18-25.
- LIU Shibiao, ZHANG Shujun, LI Peilong, et al. The cg20657709 site methylation in the early detection of lung adenocarcinoma[J]. *Journal of Shandong University (Health Science)*, 2023, 61(4): 18-25.
- [23] Xie Y, Li P, Sun D, et al. DNA methylation-based testing in peripheral blood mononuclear cells enables accurate and early detection of colorectal cancer[J]. *Cancer Res*, 2023, 83(21): 3636-3649.
- [24] Wang T, Li P, Qi Q, et al. A multiplex blood-based assay targeting DNA methylation in PBMCs enables early detection of breast cancer[J]. *Nat Commun*, 2023, 14(1): 4724.
- [25] Armstrong SA, Schultz CW, Azimi-Sadjadi A, et al. ATM dysfunction in pancreatic adenocarcinoma and associated therapeutic implications[J]. *Mol Cancer Ther*, 2019, 18(11): 1899-1908.
- [26] Yin L, Wei J, Lu Z, et al. Prevalence of germline sequence variations among patients with pancreatic cancer in China[J]. *JAMA Netw Open*, 2022, 5(2): e2148721.
- [27] Hsu FC, Roberts NJ, Childs E, et al. Risk of pancreatic cancer among individuals with pathogenic variants in the ATM gene[J]. *JAMA Oncol*, 2021, 7(11): 1664-1668.
- [28] Park W, O'Connor CA, Bandlamudi C, et al. Clinico-genomic characterization of ATM and HRD in pancreas cancer: application for practice[J]. *Clin Cancer Res*, 2022, 28(21): 4782-4792.
- [29] Begam N, Jamil K, Raju SG. Promoter hypermethylation of the ATM gene as a novel biomarker for breast cancer[J]. *Asian Pac J Cancer Prev*, 2017, 18(11): 3003-3009.
- [30] Tang ML, Gasser S. ATM activation mediates anticancer immunosurveillance by natural killer and T cells[J]. *Oncoimmunology*, 2013, 2(6): e24438.
- [31] Roest C, Yakar D, Renner Sitar DI, et al. Multimodal AI combining clinical and imaging inputs improves prostate cancer detection[J]. *Invest Radiol*, 2024, 59(12): 854-860.
- [32] Yang H, Liu J, Yang N, et al. Enhancing metastatic colorectal cancer prediction through advanced feature selection and machine learning techniques[J]. *Int Immunopharmacol*, 2024, 142(Pt A): 113033. doi: 10.1016/j.intimp.2024.113033
- [33] Zhang P, Zou M, Wen X, et al. Development of serum parameters panels for the early detection of pancreatic cancer[J]. *Int J Cancer*, 2014, 134(11): 2646-2655.
- [34] Liu C, Deng S, Jin K, et al. Lewis antigen-negative pancreatic cancer: an aggressive subgroup[J]. *Int J Oncol*, 2020, 56(4): 900-908.
- [35] 杨金收, 许瑞源, 王程程, 等. 胰腺癌的早期筛查与诊断策略[J]. *癌症*, 2023, 42(9): 453-472.