

基于公共基因表达数据库和临床样本队列构建 白内障预测模型

郭志强^{1*}, 张立友², 许利娟¹, 宫美娜¹, 韩笑¹

(1. 沧州市第四医院(南皮县人民医院), 中国河北 沧州 061500; 2. 沧州市眼科医院, 中国河北 沧州 061000)

摘要: 基于基因表达数据库(Gene Expression Omnibus, GEO)筛选差异基因, 建立白内障预测模型并对其进行评价。首先, 采用生物信息学方法从 GEO 中筛选出与白内障相关的芯片数据, 并采用 GEO2R 软件和 NetworkAnalyst 工具分析得到最显著的差异表达基因。然后, 依托我院健康管理中心白内障筛查队列, 采用 Cox 比例风险回归构建白内障发病风险预测模型, 绘制列线图, 通过 C 指数、校准曲线、受试者操作特征(receiver operating characteristic, ROC)曲线、决策曲线分析(decision curve analysis, DCA)评价模型的区分度、校准度、预测能力和获益情况。结果显示, 在 GSE5645、GSE193629 和 GSE161701 数据集中, 鱼精蛋白 1 (*protamine 1*, *PRM1*) 为高表达基因, 5羟色胺 2C 受体(*serotonin 2C receptor*, *HTR2C*)为低表达基因; 白内障组和非白内障组在年龄、体重、收缩压、对比敏感度(contrast sensitivity, CS)、客观散射指数(objective scatter index, OSI)、调制传递函数截止频率(modulation transfer function cut off, MTF cut off)、斯特列尔比(Strehl ratio, SR)、动态视力、*PRM1*、*HTR2C* 和 *CX46* 等指标的差异均有统计学意义($P < 0.05$); 预测模型最终纳入年龄、OSI、MTF cut off、*PRM1* 和 *HTR2C* 共 5 个变量($P < 0.05$), 建立的预测模型为 $\log[h(t)/h_0(t)] = 2.689 + 0.012 \times \text{年龄} + 1.320 \times \text{OSI} - 0.041 \times \text{MTF cut off} + 0.029 \times \text{PRM1} - 6.549 \times \text{HTR2C}$; 模型 C 指数为 0.875, 置信区间为 0.862~0.886; 模型的预测概率与实际概率接近; ROC 曲线下面积(area under the curve, AUC)为 0.904 (95% CI: 0.884~0.923), 灵敏度和特异度分别为 82.4%和 92.3%; 平均 AUC 为 0.911; 当模型的高风险阈值为 0.25~0.75 时, 净收益率 > 0 。本研究建立的白内障临床预测模型具有很好的区分度、校准度、预测能力、内部有效性和临床效益, 具备较高的临床应用价值。

关键词: 白内障; 临床预测模型; 列线图; 校准曲线; 受试者操作特征(ROC)曲线; 决策曲线分析(DCA)

中图分类号: Q81, R776.1

文献标志码: A

文章编号: 1007-7847(2023)05-0447-08

Model Construction for Cataract Prediction Based on Public Expression Database and Clinical Cohort

GUO Zhiqiang^{1*}, ZHANG Liyou², XU Lijuan¹, GONG Meina¹, HAN Xiao¹

(1. The Fourth Hospital of Cangzhou City (Nanpi County People's Hospital), Cangzhou 061500, Hebei, China; 2. Cangzhou Eye Hospital, Cangzhou 061000, Hebei, China)

Abstract: A cataract prediction model was established and evaluated by screening differential genes based on the Gene Expression Omnibus (GEO) database. Firstly, the microarray data related to cataract were screened from GEO by bioinformatics method and analyzed by GEO2R and NetworkAnalyst software to obtain the most significant differentially expressed genes. Then, based on the cataract screening cohort of the Health Management Center of our hospital, a cataract risk prediction model was constructed and the nomogram was drawn by using Cox proportional hazards regression. The degrees of differentiation and calibration, predictive ability and benefit of the model were evaluated through C-index, calibration curve, the receiver operating characteristic (ROC) curve and decision curve analysis (DCA). In GSE5645, GSE193629 and GSE161701 datasets, *protamine 1* (*PRM1*) is a high expression gene, and *serotonin 2C receptor* (*HTR2C*) a low expression

收稿日期: 2022-10-27; 修回日期: 2023-02-13; 网络首发日期: 2023-04-17

基金项目: 河北省 2020 年度医学科学研究课题计划(20200343)

作者简介: * 通信作者: 郭志强(1973—), 男, 河北南皮人, 副主任医师, 主要从事白内障、青光眼、眼底病研究, E-mail: yaosxx85@163.com。

gene. There were statistically significant differences ($P<0.05$) between the cataract and non-ataract groups in the age, body mass, systolic blood pressure, contrast sensitivity (CS), objective scatter index (OSI), modulation transfer function (MTF) cut off, Strehl ratio (SR), dynamic vision, *PRM1*, *HTR2C* and *CX46*. Five variables including age, OSI, MTF cut off, *PRM1* and *HTR2C* were finally included in the prediction model ($P<0.05$). The final prediction model was $\log[h(t)/h_0(t)]=2.689 2+0.012\times\text{age}+1.320\times\text{OSI}-0.041\times\text{MTF cut off}+0.029\times\text{PRM1}-6.549\times\text{HTR2C}$. The C-index of the model was 0.875, confidence interval (CI) was 0.862~0.886, and the predicted probability was close to the actual probability. The area under the curve (AUC) was 0.904 (95% CI: 0.884~0.923), and the sensitivity and specificity were 82.4% and 92.3%, respectively. The average AUC by the ten-fold crossover method was 0.911. The DCA diagram showed that, when the high risk threshold of the model was 0.25~0.75, the net return rate would be greater than 0. The cataract clinical prediction model established in this study proved to have good differentiation, calibration, prediction ability, internal effectiveness and clinical benefit, and would possess high clinical application value.

Key words: cataract; clinical prediction model; nomogram; calibration curve; receiver operating characteristic (ROC) curve; decision curve analysis (DCA)

(*Life Science Research*, 2023, 27(5): 447-454)

白内障是由多种原因导致晶状体代谢紊乱和变性,发生混浊,从而产生视觉障碍的一种疾病^[1]。白内障是首要的致盲因素,其所致盲人数占所有盲人总数的50%^[2]。白内障根据发生部位可以分为囊膜下、核性和皮质性,病理表现为囊膜弹性变差、悬韧带强度降低、易出现退行性改变,虽然其初期对远视力影响较小,但主观感觉有明显异常,包括暗视力、眩光、重影、强光下模糊等^[3]。临床常用视力表检查和视觉质量评估法筛查白内障,但是效果有限,难以查出早期白内障患者^[4]。近年来,在其他疾病中,不少临床预测模型均纳入了疾病相关基因进行建模,效果显著^[5]。但目前尚没有包含白内障相关基因的预测模型。本研究基于基因表达数据库(Gene Expression Omnibus, GEO)筛选白内障的差异表达基因,构建Cox比例风险预测模型并进行评价,为临床上筛查白内障提供参考。

1 资料与方法

1.1 差异表达基因筛选

采用生物信息学方法,从美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)的GEO数据库中筛选出与白内障相关的芯片数据GSE5645、GSE193629和GSE161701。其中,数据集GSE5645为表达谱芯片,平台号:GPL81 [MG_U74Av2] Affymetrix Murine Genome U74A Version 2 Array;数据集GSE193629为表达谱芯片,平台号:GPL21185 Agilent-072363 SurePrint G3 Human GE v3 8x60K Microarray

039494;数据集GSE161701为测序数据,平台号:GPL24676 Illumina NovaSeq 6000。以上数据集的获取时间为2022年08月09日。采用GEO2R软件和NetworkAnalyst工具进行分析,通过设置 $|\log_2(\text{FC})|$ (FC: fold change)和 P 值得到最显著的差异表达基因,其中 $|\log_2(\text{FC})|$ 越大,基因表达水平越高; P 值越小,基因在两组间的差异越显著。

1.2 一般资料

白内障筛查队列来源于沧州市第四医院(南皮县人民医院)健康管理中心。按照纳入标准和排除标准,共纳入2692例,其中,男性1468例,女性1224例。根据诊断标准,将所有入组人员分为结局发生白内障者(白内障组, $n=564$ 例)和结局未发生白内障者(非白内障组, $n=2128$ 例)。纳入标准:年龄 ≥ 18 岁;自愿接受且认真配合检查;屈光状态为球镜度数-3.00~3.00 D,柱镜度数-1.50~1.50 D;国际标准视力表单眼最佳远视力 ≥ 0.8 ;既往无眼外伤及眼部手术史;签署知情同意书;无重要信息缺失。排除标准:筛查前已确诊白内障者;合并高度屈光不正、角膜病、视神经病变、眼底病变等其他眼病;裂隙灯活体显微镜检查和直接检眼镜检查发现除晶状体混浊外其他异常;因白内障以外其他病因所致视力下降;不服从筛查及试验规定者。本研究已通过沧州市第四医院(南皮县人民医院)伦理委员会批准。

1.3 诊断标准

参照《中华眼科学》^[6]白内障的相关诊断:经裂隙灯活体显微镜在暗室内检查,自然瞳孔下晶

状体皮质混浊(以皮质混浊为主,核无或仅见轻微混浊,无后囊及囊下混浊),晶状体内可见空泡、水裂或板层分离。

1.4 调查方法

白内障筛查队列的数据包括一般检查指标[身高、体重、身体质量指数(body mass index, BMI)、舒张压、收缩压]、实验室检查指标[谷丙转氨酶(alanine transaminase, ALT)、血肌酐(creatinine, Cr)、空腹血糖]和问卷调查三部分内容。一般检查指标与实验室检查指标由医生和筛查员获得。一般情况(年龄)、生活习惯(吸烟、饮酒)和其他疾病(高血压、糖尿病、高脂血症)等信息通过问卷调查获得。随访以了解研究对象的患病情况。

1.5 视觉质量评估

采用以下指标进行视觉质量评估^[7]。1) 对比敏感度(contrast sensitivity, CS): 采用德国 OCU-LUS Binoptometer4 双目视力仪检测,使用对比度形式,通过不同对比度设置(80.0%、40.0%、20.0%、10.0%、5.0%、2.5%),让受试者识别视标,从而获取对比敏感度;2) 客观散射指数(objective scatter index, OSI): 使用 OQASTM II (欧卡斯)客观视觉质量分析系统(西班牙 Visiometrics 公司),经综合验光仪验光矫正后戴镜进行检查;3) 调制传递函数截止频率(modulation transfer function cut off, MTF cut off): 测量方法同 OSI;4) 斯特列尔比(Strehl ratio, SR): 测量方法同 OSI;5) 动态视力: 使用 TD-J905 动态视力检测仪(上海驼峰自动化技术有限公司)进行测量,在规定时间内辨别不同速度运动视标,测量受试者的动态视力。

1.6 基因表达水平的实时荧光定量 PCR 检测

采集受试者清晨空腹静脉血 5 mL 于 EDTA 抗凝管中,采用天根生化科技有限公司试剂盒(货号: DP419)提取总 RNA,使用紫外分光光度计(赛默飞世尔科技公司,型号: NanoDrop 2000)检测 RNA 浓度,将 RNA 反转录成 cDNA。采用实时荧光定量 PCR 仪(Bioneer 公司,型号: Exicycler 96)扩增 cDNA,具体程序如下: 94 °C 热启动 5 min; 94 °C 变性 10 s, 70 °C 延伸 50 s, 共 40 个循环; 40 °C 退火 50 s。采用 GAPDH 作为内参,通过 $2^{-\Delta\Delta C_t}$ 法计算 CX46 基因、鱼精蛋白 1 (protamine 1, PRM1) 和五羟色胺 2C 受体(serotonin 2C receptor, HTR2C) 的 mRNA 表达量。引物由上海生工生物工程股份有限公司提供,序列如下: PRM1 上游引物为 5'-AGAGCTGTGAGGACGATCAG-3', 下游引物为

5'-TGCAGTGGTCTATCTCACTTCG-3'; HTR2C 上游引物为 5'-AGTCATGTAGCTAGCTAGCTAA-3', 下游引物为 5'-ACTAACGTAGCCAGCTAGCTGA-3'; CX46 上游引物为 5'-GACTCATCACCATCA-ATGA-3', 下游引物为 5'-ACATTCTTTAGGCA-CGCCT-3'; GAPDH 上游引物为 5'-ATCCCATC-ACCATCTTCC-3', 下游引物为 5'-TGACCCTT-TGGCTCCCC-3'。

1.7 数据处理

GEO 数据库表达谱芯片采用 GEO2R 软件进行分析,临床数据采用 SPSS 软件和 R 4.0.4 软件进行分析。计量资料用 $(\bar{x} \pm s)$ 表示,采用 Kolmogorov-Smirnov 检验数据是否符合正态分布。若数据符合正态分布则进一步检验方差齐性,如果方差齐,采用 t 检验;如果方差不齐,采用校正 t 检验。若数据不符合正态分布,则采用非参数秩和检验。计数资料用例(%)表示,采用 χ^2 检验或 Fisher 精确概率检验。使用 Cox 回归分析筛选协变量并构建 Cox 比例风险回归预测模型。使用 R 软件包的 rms 包绘制列线图,计算 C-指数,评价模型的区分度,同时绘制校准曲线,评价模型的校准度。采用受试者操作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)评价模型的预测能力,并使用十折交叉验证法对模型进行内部验证。使用 R 软件的 mda 包绘制决策曲线分析(decision curve analysis, DCA)图,评价模型的获益情况。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 差异表达显著的基因

如图 1 所示, GSE5645 数据集共有 8 738 个基因; 满足 $|\log_2(FC)| > 2$ 且 $P < 0.01$ 的基因有 15 个, 其中, 高表达基因 1 个, 低表达基因 14 个; 满足 $|\log_2(FC)| > 3$ 且 $P < 0.01$ 的基因有 2 个, 其中, 高表达基因 1 个, 为 PRM1, 低表达基因 1 个, 为 HTR2C。在 GSE193629 数据集中, PRM1 的 $|\log_2(FC)| = 3.34$, $P = 0.002$; HTR2C 的 $|\log_2(FC)| = 4.16$, $P = 0.007$ 。在 GSE161701 数据集中, PRM1 的 $|\log_2(FC)| = 4.25$, $P = 0.003$; HTR2C 的 $|\log_2(FC)| = 3.61$, $P = 0.005$ 。以上结果表明, PRM1 和 HTR2C 在 3 个基因集中均存在差异表达。

2.2 队列基线特征描述

本研究使用的白内障筛查队列的平均随访时间为 (1.06 ± 0.37) 年, 随访时间中位数为 1.12 年。随

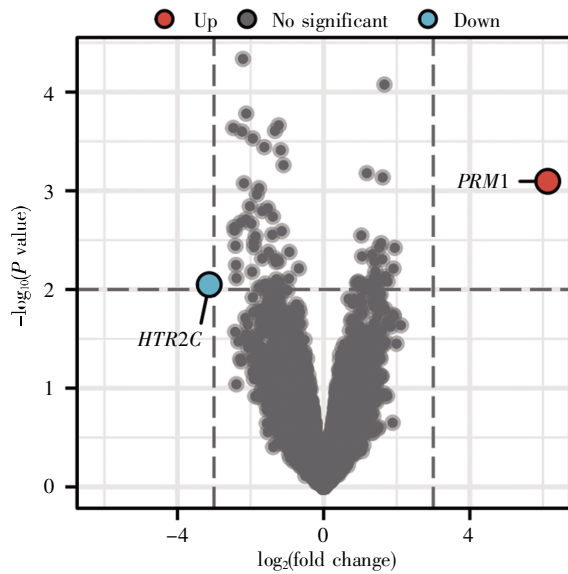


图1 GSE5645数据集中差异表达最显著的两个基因
Fig.1 The two most differentially expressed genes in the GSE5645 dataset

访期间共确诊新发白内障 564 例。从表 1 可知，白内障组和非白内障组在年龄、体重、收缩压、CS、

OSI、MTF cut off、SR 和动态视力等指标的差异均有统计学意义($P < 0.05$)。

2.3 基因表达水平比较

在结果 2.1 中，虽然差异表达最明显的 2 个基因不包括 CX46，但其表达量在两组间仍有明显的分离趋势，且该基因与结果 2.1 中的基因 PRM1 和 HTR2C 存在功能上的相关性，故一并纳入分析。从表 2 和图 2 可知，白内障组 PRM1 的表达水平明显高于非白内障组($t=22.934$; $P < 0.001$)，HTR2C 和 CX46 的表达水平明显低于非白内障组(t 分别为 -12.057 和 -20.773 ; $P < 0.001$)。

2.4 多因素 Cox 回归分析建立预测模型

将上述有统计学意义的指标纳入 Cox 比例风险回归模型，进行多因素分析和变量筛选，以构建白内障风险预测模型。最终纳入年龄、OSI、MTF cut off、PRM1 和 HTR2C 共 5 个变量($P < 0.05$, 表 3)。其中，年龄、OSI 和 PRM1 是白内障的危险因素[风险比(hazard ratio, HR) > 1]，MTF cut off 和 HTR2C 是白内障的保护因素($HR < 1$)。常数项为 2.689 2。

表 1 队列基线特征描述

Table 1 Description of the baseline characteristics of the cohort

General information	Cataract group (n=564)	Non-cataract group (n=2 128)	t/χ^2	P value
Age ($\bar{x} \pm s$, year)	65.67 \pm 7.72	58.31 \pm 7.68	15.237	<0.01
Male [n (%)]	298 (52.84)	1 129 (53.05)	0.337	0.516
Height ($\bar{x} \pm s$, cm)	162.83 \pm 7.95	162.92 \pm 8.01	0.752	0.483
Weight ($\bar{x} \pm s$, kg)	69.84 \pm 11.02	66.83 \pm 10.93	5.317	<0.01
BMI ($\bar{x} \pm s$)	25.91 \pm 3.12	25.82 \pm 3.10	1.948	0.183
Smoke [n (%)]	84 (14.89)	320 (15.04)	0.257	0.612
Alcohol [n (%)]	129 (22.87)	489 (22.98)	0.316	0.527
Hypertension [n (%)]	226 (40.07)	867 (40.74)	0.284	0.594
Systolic blood pressure ($\bar{x} \pm s$, mmHg)	139.68 \pm 21.52	134.23 \pm 19.82	7.321	<0.01
Diastolic blood pressure ($\bar{x} \pm s$, mmHg)	81.56 \pm 12.89	81.53 \pm 12.72	0.972	0.316
Hyperlipidemia [n (%)]	217 (38.48)	822 (38.63)	0.049	0.976
Diabetes [n (%)]	63 (11.17)	236 (11.09)	2.341	0.127
Fasting blood glucose ($\bar{x} \pm s$, mmol/L)	5.96 \pm 1.67	5.89 \pm 1.69	1.627	0.135
ALT ($\bar{x} \pm s$, U/L)	25.92 \pm 3.58	26.37 \pm 3.64	-1.867	0.154
Cr ($\bar{x} \pm s$, μ mol/L)	57.81 \pm 8.52	58.93 \pm 8.45	1.613	0.139
CS ($\bar{x} \pm s$)	1.32 \pm 0.12	1.63 \pm 0.15	-5.167	<0.01
OSI ($\bar{x} \pm s$)	1.78 \pm 0.16	0.65 \pm 0.12	9.348	<0.01
MTF cut off ($\bar{x} \pm s$, c/d)	24.62 \pm 3.18	37.89 \pm 3.36	-10.562	<0.01
SR ($\bar{x} \pm s$)	0.13 \pm 0.03	0.18 \pm 0.04	-7.264	<0.01
Dynamic vision ($\bar{x} \pm s$)	0.24 \pm 0.04	0.53 \pm 0.06	-13.561	<0.01

表 2 白内障组与非白内障组 PRM1、HTR2C 和 CX46 的表达水平比较

Table 2 Comparison of PRM1, HTR2C and CX46 expression levels between cataract and non-cataract groups

Group	n	PRM1	HTR2C	CX46
Non-cataract group	2 128	1.00 \pm 0.17	1.00 \pm 0.14	1.00 \pm 0.16
Cataract group	564	1.69 \pm 0.21	0.76 \pm 0.11	0.85 \pm 0.12
t		22.934	-12.057	-20.773
P		<0.001	<0.001	<0.001

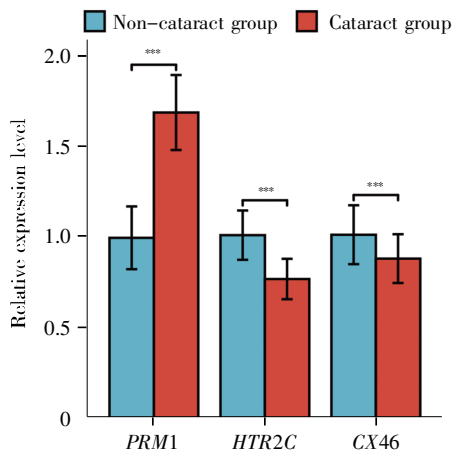


图 2 白内障组与非白内障组 *PRM1*、*HTR2C* 和 *CX46* 的表达水平比较
Fig.2 Comparison of *PRM1*, *HTR2C* and *CX46* expression levels between cataract and non-cataract groups
 ***: $P < 0.001$.

最终建立的预测模型为 $\log[h(t)/h_0(t)] = 2.689 \times 2 + 0.012 \times \text{年龄} + 1.320 \times \text{OSI} - 0.041 \times \text{MTF cut off} + 0.029 \times \text{PRM1} - 6.549 \times \text{HTR2C}$ 。

2.5 绘制列线图将模型可视化

对多因素分析筛选出的变量进行可视化, 构建列线图, 结果如图 3 所示。根据列线图, 可得出个体每个预测指标对应的分值(points), 并计算出个体所有指标的总分, 与总分对应的预测概率即为个体患白内障的概率。例如: 某人年龄 70 岁(12 分), OSI 为 1.4 (41 分), MTF cut off 为 20 (15 分), *PRM1*

表 3 多因素 Cox 回归分析结果

Table 3 Results of multivariate Cox regression analysis

Variable	β	HR	Confidence interval	P value
Age	0.012	1.012	1.001~1.022	0.033
OSI	1.320	3.742	2.465~5.680	<0.01
MTF cut off	-0.041	0.960	0.936~0.984	0.012
<i>PRM1</i>	0.029	1.030	1.003~1.059	0.039
<i>HTR2C</i>	-6.549	0.097	0.016~0.149	<0.01

水平为 1.4 (38 分), *HTR2C* 水平为 0.9 (36 分), 可算出其总分为 142 分, 对应其 1 年内患白内障的概率是 0.5。

2.6 C-指数和校准曲线评价模型的区分度与校准度

该模型的 C-指数为 0.875, 置信区间为 0.862~0.886, 说明模型有很好的区分度。校准曲线如图 4 所示, 横坐标是预测概率, 纵坐标是实际概率。由图 4 可知, 该模型的预测概率与实际概率线(对角线)非常接近, 说明模型具有很好的校准度。

2.7 ROC 曲线和十折交叉法评价模型的预测能力与内部有效性

ROC 曲线如图 5 所示, AUC 为 0.904 (95% CI: 0.884~0.923), 最佳切点为 0.456, 灵敏度和特异度分别为 82.4%和 92.3%, 阳性预测值为 0.926, 阴性预测值为 0.817, 约登指数为 0.747, 说明模型有很好的预测能力。十折交叉法验证结果显示, 平均 AUC 为 0.911, 说明模型具有较高的内部有效性。

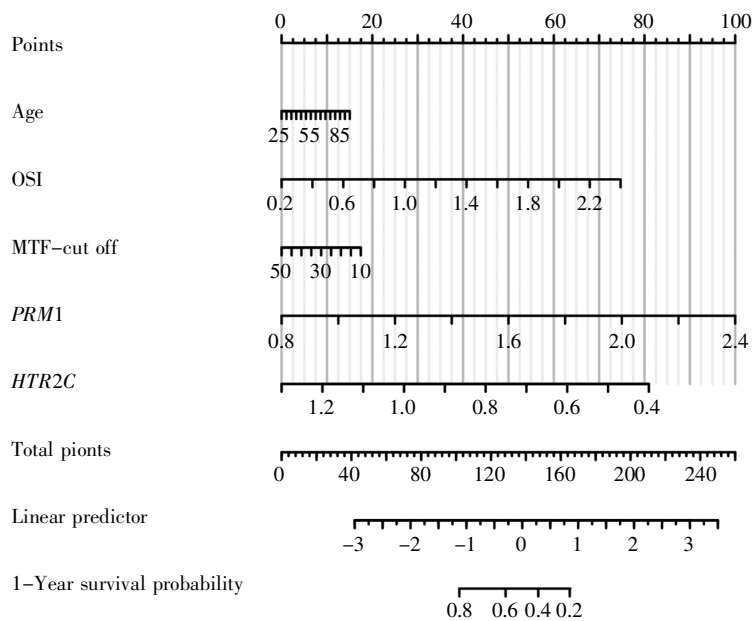


图 3 预测模型的列线图
Fig.3 Nomogram of the prediction model

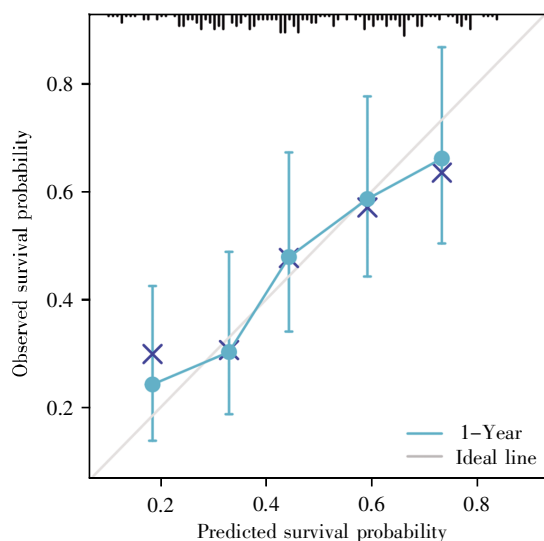


图4 评价模型校准度的校准曲线

Fig.4 Calibration curve for evaluating calibration degree of the model

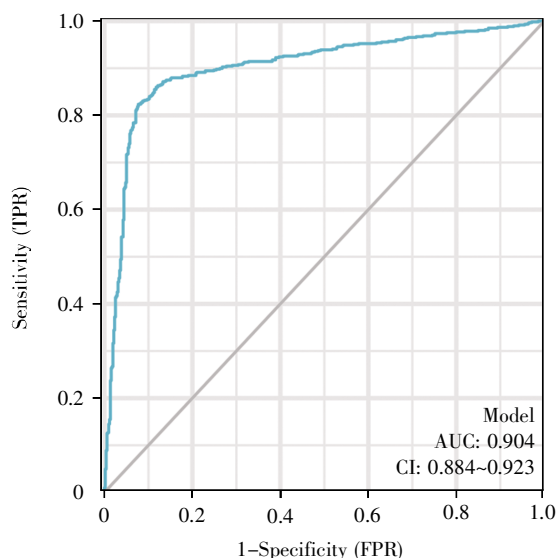


图5 评价模型预测能力的ROC曲线

Fig.5 ROC curve for evaluating predictive ability of the model

2.8 决策曲线评价模型的临床效益

如图6所示,当高风险阈值为0.25~0.75时,净收益率 >0 ,模型有临床意义;且高风险阈值取值越小,净收益率越大,说明该模型具有很好的临床效益。

3 讨论

白内障的发病机制比较复杂。有研究表明,年龄、性别、种族、代谢、遗传等多种因素均在白内障的发展中扮演重要角色^[8];紫外线、糖尿病、高血压、心血管疾病、饮酒、吸烟等因素也可诱发白

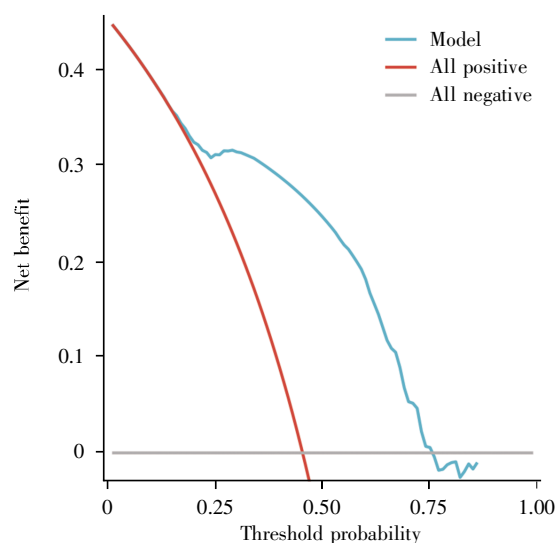


图6 评价模型临床效益的DCA图

Fig.6 DCA diagram for evaluating clinical benefit of the model

内障的形成^[9]。白内障患者的晶体混浊主要分布于周边,视轴区晶体尚处于透明状态,远视力暂未发生变化,视力表检查难以发现问题^[10]。目前,临床上主要采用超声乳化白内障摘除联合人工晶状体植入术治疗白内障,但是术后易出现眼内炎、黄斑水肿、后囊膜破裂、眼干燥症等并发症^[11-12]。所以,早期筛查与诊断仍是目前应对白内障的最佳策略。

预测模型的指标选择至关重要。本研究结果表明,从GEO数据库中筛选的 *PRM1* 和 *HTR2C* 这2个基因满足 $|\log_2(FC)| > 3$ 且 $P < 0.01$, 提示其在两组中的水平变化最为显著,适合参与建立预测模型。*PRM1* 位于人16号染色体,长度426 bp,可翻译为含有50个氨基酸的碱性核蛋白,该蛋白质具有3个结构域:左右部位均含有半胱氨酸残基,中间部位可结合DNA,精氨酸含量占70%,其能够保护细胞的基因组和表观遗传信息^[13]。细胞中的 *PRM1* 能通过阻断3'非翻译区与RNA的结合,抑制翻译,从而导致染色质浓缩,转录停止^[14]。有研究表明, *PRM1* 的激活会促进细胞增殖和转移,导致肺癌等疾病发生^[15]。*HTR2C* 分布于人体内多处器官,其中大脑含量最高;该基因位于X染色体的q24区域,其中759C/T (rs3813929)和697G/C (rs518147)这两处突变会作用于启动子,影响相应的受体表达量^[16]。有研究发现, *HTR2C* 的激活会导致癫痫等疾病^[17]。目前尚没有 *PRM1* 和 *HTR2C* 与白内障相关的报道,本研究通过GEO数据库分析和实时荧光定量PCR验证,首次表明了二者与

白内障发病相关。

OSI 常被用来评价人眼的散射指数, OSI 越高, 则视觉越差^[18]。有研究表明, 在早期白内障发生时, 视力无明显变化或仅有轻度下降, 此时 OSI 已有明显升高^[19]。这与本研究结果(表 1)一致, 表明在早期白内障患者的 OSI 即出现显著升高趋势。MTF cut off 代表了眼光学呈现和实物的对比度损失情况, MTF cut off 值越大, 成像越清晰, 视觉越好^[20]。有研究表明, MTF cut off 值越小, 视力损害越严重^[21]。本研究结果显示, 白内障组的 MTF cut off 值明显低于非白内障组(表 1), 这与之前的研究^[22]一致。此外, 年龄与白内障发生也有密切关系^[23]。本研究与之前的研究结果^[24]相似, 年龄每增加 1 岁, 发病风险提高 1.012 倍(HR=1.012)(表 3)。总的来讲, 本研究表明 OSI 和 MTF cut off 指标可以作为独立预后因素来预测早期白内障。而且, 年龄、OSI 和 *PRM1* 是白内障的危险因素(HR>1), MTF cut off 和 *HTR2C* 是白内障的保护因素(HR<1)。

目前, 关于白内障的预测模型多数是病例对照研究, 主要是回顾性分析已收集的临床资料, 预测模型多采用 logistic 回归模型。比如: 黄伟梅等^[25]通过建立模型对采用复明 2 号车行白内障手术后的感染风险进行了预测研究。该研究假定两组患病率是相同的, 这往往与真实情况有差异。而本研究采用的是前瞻性队列研究设计, 预测模型采用的是 Cox 比例风险回归模型, 证据等级高于回顾性分析, 而且通过列线图这种可视化的方式, 能直观计算出每一个研究对象的白内障发病风险。在模型评价方面, 大多数研究只采用 ROC 曲线和 AUC 评价模型的预测能力。比如: 杨惠春等^[26]在构建预测模型后, 用 ROC 曲线评价晶状体溶解性青光眼患者白内障术后高眼压风险。目前尚未见使用校准曲线和 DCA 曲线评价白内障预测模型的研究。而本研究采用校准曲线、C-指数、ROC 曲线和 DCA 曲线 4 种评价方法, 分别验证了模型的校准度、区分度、预测能力和获益情况。结果表明, 该模型具有很好的区分度、校准度、预测能力和临床价值, 为模型的临床应用提供了全面的证据。

本研究表明, 性别、吸烟、饮酒和糖尿病这 4 个因素在组间没有明显差异(表 1), 原因可能有以下三点。1) 报告偏倚: 在问卷调查中, 研究对象可能隐瞒吸烟、饮酒和糖尿病史; 2) 样本量偏小: 有的白内障研究纳入了超 8 000 例研究对象, 结果

提示存在组间差异^[27]; 3) 选择偏倚: 由于本研究是针对健康人及可能具有白内障患病危险因素的人群进行的筛查, 排除了已确诊的白内障患者和视力严重损害者, 所以可能导致上述 4 个因素在健康人和白内障初期患者中没有明显差异。另外, 身高、体重、收缩压、ALT、Cr、CS、SR 这些因素在单因素分析中有明显差异($P<0.05$), 但在多因素分析中无显著差异($P>0.05$), 原因可能有两点。1) 区分度较低: 某些因素虽然在两组间差异明显, 但建模时未能达到很好的区分度, 不能很好预测结局, 导致模型预测效率下降, 故被模型舍弃; 2) 多重共线性: 某些因素可以被其他因素通过线性关系解释, 导致其成为多余变量, 故被模型舍弃。

综上所述, 本研究建立的白内障临床预测模型具有很好的区分度、校准度、预测能力、内部有效性和临床效益, 能够为识别白内障高危人群、评估患病风险提供有力证据, 具备较高的临床应用价值。

参考文献(References):

- [1] XU J J, FU Q L, CHEN X J, *et al.* Advances in pharmacotherapy of cataracts[J]. *Annals of Translational Medicine*, 2020, 8(22): 1552.
- [2] GARAY-ARAMBURU G, BERGADO-MIJANGOS R, IRIZAR-AMILLETA R, *et al.* Risk factors for predicted refractive error after cataract surgery in clinical practice. Retrospective observational study[J]. *Archivos de la Sociedad Española de Oftalmología (English Edition)*, 2022, 97(3): 140-148.
- [3] 孙晓艳. 超声乳化白内障吸除术治疗皮质性白内障临床疗效评[J]. 影像研究与医学应用(SUN Xiaoyan. Clinical effect of phacoemulsification in the treatment of cortical cataract[J]. *Journal of Imaging Research and Medical Applications*), 2020, 4(6): 169-170.
- [4] 王诺, 贾松柏. 视觉质量分析在白内障患者中的临床应用[J]. 中南大学学报(医学版)(WANG Nuo, JIA Songbai. Clinical application of visual quality analysis in cataract patients[J]. *Journal of Central South University (Medical Science)*), 2019, 44(4): 461-468.
- [5] 盛超, 金芳和. 基于生物信息学构建卵巢浆液性腺癌预后风险模型及潜在中药预测分析[J]. 中医药临床杂志(SHENG Chao, JIN Fanghe. Construction of prognostic risk model for ovarian serous adenocarcinoma based on bioinformatics and predictive analysis of potential traditional Chinese medicine[J]. *Clinical Journal of Traditional Chinese Medicine*), 2022, 34(7): 1252-1259.
- [6] 李凤鸣, 谢立信. 中华眼科学[M]. 3 版. 北京: 人民卫生出版社(LI Fengming, XIE Lixin. *Chinese Ophthalmology*[M]. 3rd ed. Beijing: People's Medical Publishing House), 2014: 874-876.
- [7] ZHONG Y Y, ZHU Y N, WANG W, *et al.* Femtosecond laser-assisted cataract surgery versus conventional phacoemulsification: comparison of internal aberrations and visual quality[J]. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 2022, 260(3): 901-911.

- [8] KHAN S, RANA N, NASIR H, *et al.* Mutational analysis of *CRYAA* gene of cataract and investigating risk assessment factors responsible for eye diseases in district Buner, KPK, Pakistan[J]. Cellular and Molecular Biology, 2023, 69(9): 1-7.
- [9] 聂敏, 曲超. 皮质性及核性白内障影响因素分析[J]. 实用医院临床杂志(NIE Min, QU Chao. Analysis of influencing factors for the cortical and nuclear cataract[J]. Practical Journal of Clinical Medicine), 2019, 16(4): 1-5.
- [10] 蒋永祥, 卢奕. 现代视功能检查在微创白内障手术中的应用意义[J]. 中国眼耳鼻喉科杂志(JIANG Yongxiang, LU Yi. Application of modern visual performance measurements in minimally invasive cataract surgery[J]. Chinese Journal of Ophthalmology and Otorhinolaryngology), 2012, 12(3): 145-148.
- [11] 桑万春, 张小平. 白内障患者行超声乳化术中发生后囊膜破裂的相关危险因素分析[J]. 现代医学与健康研究电子杂志(SANG Wanchun, ZHANG Xiaoping. Analysis of risk factors for posterior capsule rupture in cataract patients undergoing phacoemulsification[J]. Modern Medicine and Health Research Electronic Journal), 2022, 6(16): 106-109.
- [12] 喻娟, 邓小丽, 傅小毅, 等. 16 例白内障术后非感染性眼内炎的临床分析[J]. 中华眼视光学与视觉科学杂志(YU Juan, DENG Xiaoli, FU Xiaoyi, *et al.* The clinical analysis of noninfectious endophthalmitis after cataract surgery[J]. Chinese Journal of Optometry Ophthalmology and Visual Science), 2022, 24(8): 622-627.
- [13] 杨丁全. 鱼精蛋白 1 在胃癌腺癌诊疗中的作用及促进肿瘤生长的机制研究[D]. 长春: 吉林大学(YANG Dingquan. The Role of Protamine 1 in the Diagnosis and Treatment of Gastric Adenocarcinoma and the Mechanism of Promoting Tumor Growth[D]. Changchun: Jilin University), 2021.
- [14] 任圣男. 鱼精蛋白 1 在结直肠癌诊断及治疗中的价值研究[D]. 长春: 吉林大学(REN Shengnan. Investigation of Protamine 1 in Colorectal Cancer Diagnosis and Treatment[D]. Changchun: Jilin University), 2021.
- [15] 李小丰. PRM1 在非小细胞肺癌中的表达及生物学作用研究[D]. 长春: 吉林大学(LI Xiaofeng. Study on the Expression and Biological Effects of PRM1 in Non-small Cell Lung Cancer[D]. Changchun: Jilin University), 2021.
- [16] 罗伏钢, 闫盼, 方悦, 等. 5-*HTR2C* 基因多态性对奥氮平治疗精神分裂症患者脂质代谢的影响[J]. 中国现代应用药学(LUO Fugang, YAN Pan, FANG Yue, *et al.* Effects of 5-*HTR2C* gene polymorphisms on glucolipid metabolism in schizophrenia patients treated with olanzapine[J]. Chinese Journal of Modern Applied Pharmacy), 2022, 39(10): 1347-1351.
- [17] 徐珑洋, 伍雪英, 林旭, 等. 5-羟色胺 2C 受体基因多态性与特发性癫痫关联研究[J]. 中风与神经疾病杂志(XU Longyang, WU Xueying, LIN Xu, *et al.* Association between 5-*HTR2C* gene polymorphism and idiopathic epilepsy[J]. Journal of Apoplexy and Nervous Diseases), 2018, 35(4): 324-327.
- [18] 徐彦. 不同人工晶状体植入治疗老年白内障患者的效果及对视觉质量的影响[J]. 医疗装备(XU Yan. The effect of different intraocular lens implantation on senile cataract patients and its influence on visual quality[J]. Medical Equipment), 2021, 34(8): 103-104.
- [19] GALLIOT F, PATEL S R, COCHENER B. Objective scatter index: working toward a new quantification of cataract?[J]. Journal of Refractive Surgery, 2016, 32(2): 96-102.
- [20] 郭磊, 梁先军, 张希乔, 等. 飞秒激光白内障手术联合PanOptix三焦点人工晶状体植入术的疗效[J]. 国际眼科杂志(GUO Lei, LIANG Xianjun, ZHANG Xiqiao, *et al.* Clinical effect of femtosecond laser-associated cataract surgery combined with PanOptix trifocal intraocular lens implantation[J]. International Eye Science), 2023, 23(2): 312-315.
- [21] 孙堂胜, 赵广愚, 郑两定, 等. 年龄相关性白内障患者三焦点与单焦点 IOL 植入术后早期视觉质量比较[J]. 国际眼科杂志(SUN Tangsheng, ZHAO Guangyu, ZHENG Liangding, *et al.* Comparison and analysis of early effect of tri-focal and monofocal intraocular lens implantation on the visual quality of patients with age-related cataract after surgery[J]. International Eye Science), 2021, 21(12): 2124-2129.
- [22] 刘敏锐, 彭稚喜, 吴峥峥, 等. 连续视程人工晶状体与双焦点、单焦点人工晶状体植入术后患者视力、视觉质量及满意度对比分析[J]. 眼科新进展(LIU Minrui, PENG Zhixi, WU Zhengzheng, *et al.* Comparative analysis of visual acuity, visual quality and satisfaction after extended depth, bifocal and monofocal intraocular lens implantation[J]. Recent Advances in Ophthalmology), 2023, 43(2): 127-130.
- [23] 杨斐, 鲍永珍. 年龄相关性白内障合并浅前房患者角膜屈光力与前房形态的相关性研究[J]. 中华眼科杂志(YANG Fei, BAO Yongzhen. Correlation between corneal refractive power and anterior chamber morphology in age-related cataract patients with shallow anterior chambers[J]. Chinese Journal of Ophthalmology), 2022, 58(5): 348-353.
- [24] 刘雯婷, 陈华洁, 沈念慈, 等. 图形翻转视觉诱发电位对轻度白内障术后视功能预测的价值[J]. 中国眼耳鼻喉科杂志(LIU Wenting, CHEN Huajie, SHEN Nianci, *et al.* Pattern reversal visual evoked potential examination in the prediction of the post-operative vision acuity of age-related cataract patients[J]. Chinese Journal of Ophthalmology and Otorhinolaryngology), 2020, 20(2): 98-100.
- [25] 黄伟梅, 林丹花, 郑玉生, 等. 复明 2 号车批量化白内障手术感染风险预测模型的建立及应用效果研究[J]. 智慧健康(HUANG Weimei, LIN Danhua, ZHENG Yusheng, *et al.* Establishment and application effect research of infection risk prediction model for batch cataract surgery of fuming No.2 vehicle[J]. Smart Healthcare), 2021, 7(12): 54-56.
- [26] 杨惠春, 宋学英, 齐绍文, 等. 晶状体溶解性青光眼患者白内障手术后高眼压的危险因素[J]. 国际眼科杂志(YANG Hui-chun, SONG Xueying, QI Shaowen, *et al.* Analysis of risk factors of high intraocular pressure after cataract surgery for phacolytic glaucoma[J]. International Eye Science), 2021, 21(12): 2170-2174.
- [27] 陆晓兰, 丁春波, 陈晴, 等. 老年白内障患者人工晶体植入术后眼内感染的危险因素分析[J]. 中华医院感染学杂志(LU Xiaolan, DING Chunbo, CHEN Qing, *et al.* Risk factors of intraocular infection after implantation of intraocular lens in elderly patients with cataract[J]. Chinese Journal of Nosocomiology), 2017, 27(7): 1594-1596, 1600.