

心肌梗死潜在生物标志物的生物信息学研究

钱 晗, 曹梦菲, 吕书梅, 袁 伟

(江苏大学附属医院 心血管内科, 江苏 镇江, 212001)

摘要: **目的** 基于外周血单个核细胞(PBMCs)寻找心肌梗死潜在的生物学标志物。**方法** 从基因表达综合(GEO)数据库中下载心肌梗死患者 PBMCs 的测序数据集 GSE59867, 基于加权基因共表达网络分析(WGCNA)和差异表达基因(DEGs)探讨心肌梗死过程中的生物学变化和共同基因, 并进行共同转录因子预测。通过 LASSO 回归分析确定关键基因, 绘制受试者工作特征(ROC)曲线, 评估关键基因的临床价值。通过基因集富集分析(GSEA)探讨与关键基因有关的生物学变化, 并基于心肌梗死数据集 GSE123342 验证关键基因的表达。**结果** 心肌梗死过程伴随着免疫炎症反应, 其中关键基因有 6 个, 即 *CD163*、*RNASE2*、*HP*、*FAM20A*、*MCEMP1* 和 *FAM198B*。ROC 曲线分析结果显示, 这些关键基因的表达对心肌梗死均有较高的诊断价值。在心肌梗死恢复过程中, 这些基因的表达呈现明显下降趋势。GSEA 结果显示, 这些关键基因涉及的生物学改变大多与糖脂代谢、活性氧、免疫炎症等有关。**结论** 基于 PBMCs 进行分析, 本研究发现 *CD163*、*RNASE2*、*HP*、*FAM20A*、*MCEMP1* 和 *FAM198B* 基因与心肌梗死有关, 这可能为心肌梗死的早期诊断和梗死恢复的监测、评估与管理提供新的见解。

关键词: 心肌梗死; 基因表达综合数据库; 外周血单个核细胞; 加权基因共表达网络分析; 差异表达基因; 生物信息学

中图分类号: R 542.2; Q 811.4 文献标志码: A 文章编号: 1672-2353(2023)03-021-08 DOI: 10.7619/jcmp.20223235

Bioinformatics study on potential biomarkers of myocardial infarction

QIAN Han, CAO Mengfei, LYU Shumei, YUAN Wei

(Vasculocardiology Department, Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu, 212001)

Abstract: **Objective** To find potential biomarkers of myocardial infarction from the perspective of peripheral blood mononuclear cells (PBMCs). **Methods** The sequencing dataset GSE59867 of peripheral blood mononuclear cells from patients with myocardial infarction was downloaded from the Gene Expression Omnibus (GEO) database. Weighted gene co-expression network (WGCNA) and differentially expressed genes (DEGs) were used to investigate the biological changes during process of myocardial infarction and to find common genes. Common transcription factors for these genes were predicted. Finally, hub genes were identified by LASSO regression analysis. Receiver operating characteristic curve (ROC) analysis was used to evaluate the clinical value of hub genes. Gene set enrichment analysis (GSEA) was used to explore the biological changes related to hub genes, and the expression of hub genes was verified based on myocardial infarction dataset GSE123342. **Results** Myocardial infarction was accompanied by an immune inflammatory response. Six novel hub genes were identified (*CD163*, *RNASE2*, *HP*, *FAM20A*, *MCEMP1*, and *FAM198B*). ROC analysis showed that the expression of these hub genes had good diagnostic for myocardial infarction. The expression of these genes also decreased significantly during recovery from infarction during recovering from myocardial infarction. GSEA results suggested that most of these hub genes were related to glucose and lipid metabolism, reactive oxygen species, immune inflammation and so on. **Conclusion** Based on the analysis of PBMCs, this study find that *CD163*, *RNASE2*, *HP*, *FAM20A*, *MCEMP1* and *FAM198B* genes are associated with myocardial infarction, which may provide new insights for the early diagnosis of myocardial infarction and the monitoring, evaluation and management of infarction recovery.

Key words: myocardial infarction; gene expression omnibus database; peripheral blood mononuclear cells; weighted gene co-expression network; differentially expressed gene; bioinformatics

在人口老龄化和代谢危险因素的双重压力下,中国居民的心血管疾病风险持续增加,心肌梗死患者的病死率亦逐渐上升^[1]。心肌梗死是指心肌缺血性坏死,通常是在冠状动脉病变的基础上发生冠状动脉血供减少或中断,心肌细胞因严重且持续的缺血缺氧而发生坏死^[2]。心肌细胞的死亡会引发一系列免疫炎症反应,这些免疫炎症反应介导的受损心肌修复在梗死后心室重构、心功能和患者结局等方面起着关键作用^[3]。急性心肌梗死的诊断主要依据临床表现、血清生物学标志物和心电图,其中临床常用的生物学标志物包括肌红蛋白、肌酸激酶-MB、心肌肌钙蛋白 I 和心肌肌钙蛋白 T^[4-5]。然而,这些标志物水平在心力衰竭、肾衰竭和甲状腺疾病患者中也同样升高,故结合多组学平台探寻更多特异性的生物学标志物非常必要^[6-7]。外周血单个核细胞(PBMCs)包括 T 细胞、B 细胞、单核细胞、树突状细胞和自然杀伤细胞,在免疫应答、免疫监测和免疫治疗中发挥着重要作用^[8]。本研究基于 PBMCs 分析心肌梗死和梗死恢复过程中潜在的关键基因和生物学改变,旨在为心肌梗死的诊断、评估、管理和预后分析提供新的见解。

1 材料和方法

1.1 数据集获取

从 GEO 数据库中下载关于心肌梗死 PBMCs 的转录组测序芯片数据集 GSE59867^[9-10],芯片注释平台为 GPL6244,芯片信息为 Affymetrix Human Gene 1.0 ST Array。该数据集收集了 111 例 ST 段抬高型心肌梗死(STEMI)患者入院、出院、心肌梗死后 1 个月、心肌梗死后 6 个月时的外周血样本数据。对照组为 46 例无心肌梗死病史的稳定型冠心病患者。基于芯片注释平台对数据集进行基因名转换,用于后续共表达网络的构建。

1.2 加权基因共表达网络分析(WGCNA)

通过 WGCNA R 包筛选与心肌梗死有关的易感基因模块^[11]。WGCNA 是一种常用的生物信息学方法,可以识别具有相似表达模式的基因模块,分析基因模块与样本表型之间的关系,绘制基因模块中的调控网络,识别关键模块和调控基因^[12-13]。首先,选取数据集中方差变异程度最高

的前 3 000 个基因构建共表达网络;其次,利用 R 函数 pickSoftThreshold 计算软阈值 $1 \sim 20$;然后,对这些基因进行层次聚类 and 动态树切割来划分模块,并合并相似的模块(阈值为 0.25);最后,计算模块与疾病之间的 Spearman 相关系数,进而确定易感基因模块。为了探讨所选模块的潜在生物学功能,通过 ClusterProfiler R 包对模块中的基因进行基因本体论(GO)分析和京都基因与基因组百科全书(KEGG)通路分析^[14]。GO 分析包括生物过程(BP)、分子功能(MF)和细胞成分(CC)。通过 Ggplot2 R 包进行结果展示。

1.3 基因差异分析和转录因子预测

使用 Limma R 包进行基因差异分析^[15]。差异表达基因(DEGs)的筛选标准为校正后 $P < 0.05$, $|\log_2(\text{倍数变化})| > 0.5$ 。基于 ClusterProfiler R 包对 DEGs 进行 GO 分析和 KEGG 通路分析, $P < 0.05$ 表示显著富集。对上述易感基因模块和 DEGs 进行交集分析,通过 ChEA3 数据库对交集基因进行转录因子预测。ChEA3 是一种转录因子富集分析工具,整合了多个数据库,如 ENCODE、CTEx、ARCHS4 等数据库^[16]。由于 ChEA3 数据库并未对预测结果可靠度进行严格划分,参考相关研究^[16-17]结论,本研究对平均排名分 ≤ 30 分的共同转录因子进行结果展示。基于 STRING 数据库构建交集基因的蛋白相互作用网络,并将交互评分 > 0.4 分的数据导入 Cytoscape 3.8.2 软件进行可视化处理^[18-19]。

1.4 关键基因的确定和受试者工作特征(ROC)曲线分析

LASSO 回归分析是常用的筛选变量的压缩估计方法,通过构造惩罚函数获得较为精炼的模型,使得其压缩一些回归系数,同时设定一些回归系数为 0,因此保留了子集收缩的优点,是一种处理具有复共线性数据的有偏估计,常被用于协变量筛选^[20]。基于 glmnet R 包对交集基因中存在蛋白相互作用的基因进行 LASSO 回归分析,进一步筛选关键基因^[21]。采用 10 倍交叉验证法选择惩罚项(λ),并选择最小二项偏差在 1 个标准误差内的最简单模型的 λ 值(λ_{1se})。通过 pROC R 包对关键基因进行 ROC 曲线分析,计算曲线下面积(AUC),评估关键基因对心肌梗死的

诊断价值^[22]。

1.5 基因富集分析(GSEA)和外部数据集验证

根据每个关键基因的表达量将数据集划分为高表达组和低表达组,通过 GSEA 探讨高表达组和低表达组之间潜在的生物学功能变化,将特征基因集作为预先定义的基因集。作为一种常见的生物信息学方法,GSEA 可评估一个预先定义的基因集的基因在与表型相关度排序的基因表中的分布趋势,从而判断其对表型的影响^[23]。从 GEO 数据库中选取另一个数据集 GSE123342,将其用于验证关键基因的表达。GSE123342 数据集包括急性心肌梗死(65 例)、梗死后 30 d(64 例)、梗死后 1 年(37 例)、稳定型冠心病(22 例)样本数据以及 4 例技术重复样本数据。

2 结果

2.1 易感基因模块的确定

根据无标度网络原理,选取 10 作为共表达网络的软阈值,见图 1A。通过层次聚类 and 动态树切割,将前 3 000 个基因划分为 15 个基因模块,见图 1B。剔除聚类失败的灰色模块,棕色模块的总体表达与心肌梗死的相关性最高($r=0.4$),且与出院至心肌梗死后 6 个月的演变过程呈负相关($r=-0.33$),见图 1C。棕色模块中的共表达基因伴随着心肌梗死的发生呈现高表达趋势,而在心肌梗死恢复过程中则呈现低表达趋势。由此提示,棕色模块可能是易感基因模块,其中包含 243 个基因,即 *ABHD5*、*ACER3*、*ACP3*、*ACSL1*、*ADAM17*、*ADAM9*、*ADM*、*AHR*、*ALCAM*、*ALDH2*、*ALPK1*、*ANKRD50*、*ANO5*、*AQP9*、*ARHGAP24*、*ARHGAP29*、*ARRDC4*、*ASPH*、*ATP6V0A1*、*ATP6V1A*、*B3GNT5*、*BACH1*、*BLVRB*、*BST1*、*C3AR1*、*C5*、*C9orf72*、*CALCRL*、*CAPG*、*CAPZA2*、*CARD6*、*CASP1*、*CASP5*、*CCDC88A*、*CCR2*、*CD14*、*CD163*、*CD1D*、*CD33*、*CD36*、*CD63*、*CLEC1A*、*CLEC4A*、*CLEC4D*、*CLEC4E*、*CLEC7A*、*CLMN*、*CNTLN*、*CPED1*、*CPM*、*CPNE8*、*CPVL*、*CR1*、*CRIL*、*CREG1*、*CRISPLD2*、*CS-TA*、*CTNNA1*、*CTSH*、*CYBRD1*、*CYP1B1*、*CYP1B1-AS1*、*DACH1*、*DDIAS*、*DOCK4*、*DOCK5*、*DRAM1*、*DSC2*、*DSE*、*DUSP6*、*EDNRB*、*ENTPD1*、*ERLIN1*、*EVI5*、*F5*、*FAM114A1*、*FAM13A*、*FAM151B*、*FAM20A*、*FAM198B*、*FAR2*、*FBN2*、*FBP1*、*FCGR2A*、*FGD4*、*FGD6*、*FLT3*、*FLVCR2*、*FMO5*、*FPR1*、*FPR2*、*FRRS1*、*FUCA1*、*FUCA2*、*GAPT*、*GAS2L3*、*GCA*、*GI-*

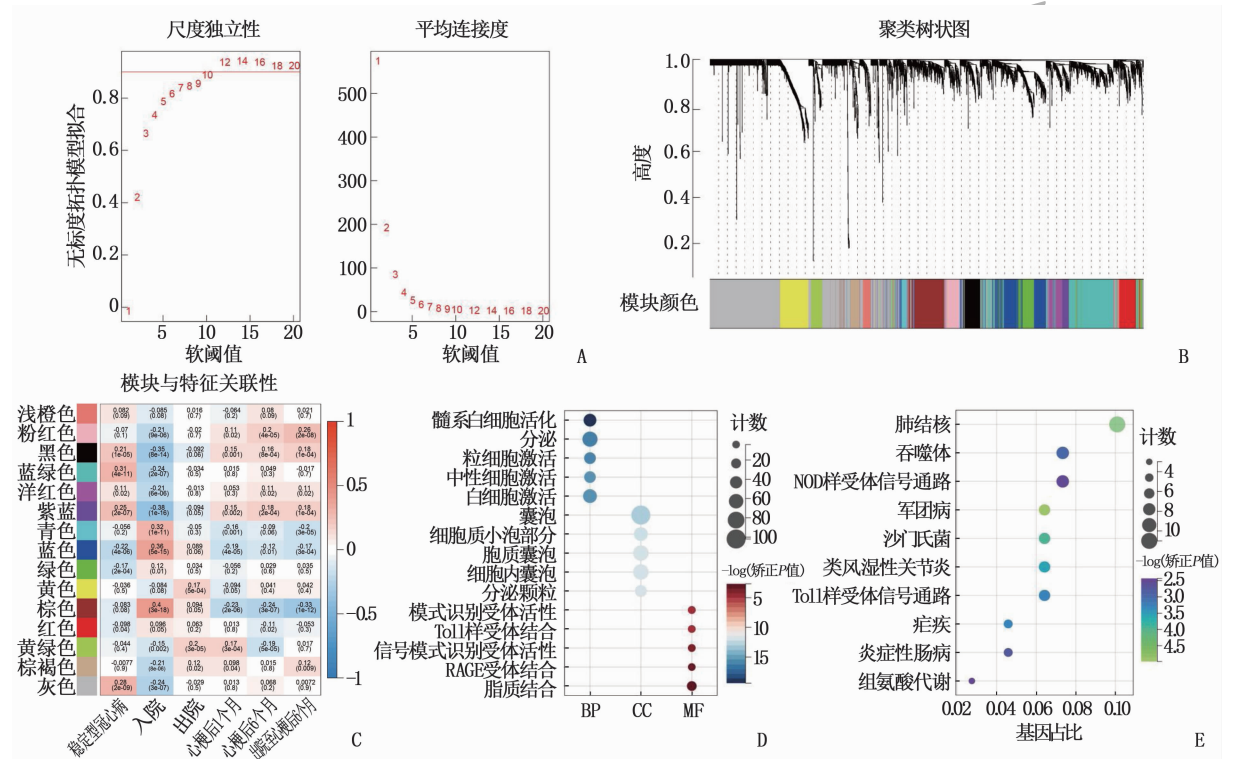
MAP8、*GLT1D1*、*GPAT3*、*GPR141*、*HAL*、*HGF*、*HMGB2*、*HNMT*、*HORMAD1*、*HP*、*HPSE*、*IDH1*、*IFN-GR1*、*IL13RA1*、*IL15*、*IL18*、*IMPA2*、*IRAK3*、*JAK2*、*KCNJ15*、*KCNJ2*、*KIF13A*、*KLHL8*、*KYNU*、*LHFPL2*、*LILRA5*、*LIN7A*、*LIPN*、*LMNB1*、*LPCAT2*、*LRMDA*、*LRRK2*、*LTA4H*、*LY86*、*LYVE1*、*MAP3K20*、*MCEMP1*、*ME1*、*METTL7A*、*METTL7B*、*MFSD1*、*MGST1*、*MGST2*、*MILR1*、*MNDA*、*MOSPD2*、*MS4A4A*、*MSRB1*、*MTARC1*、*MYOF*、*NAAA*、*NAIP*、*NETO2*、*NLN*、*NLRC4*、*NPL*、*OLFML2B*、*P2RX7*、*P2RY13*、*PGD*、*PIP4P2*、*PLA2G4A*、*PLBD1*、*PLD1*、*PLIN2*、*PLSCR1*、*PPARG*、*PPT1*、*PRRG4*、*PSTPIP2*、*PYGL*、*QPCT*、*RAB39A*、*RALB*、*RBP7*、*RGL1*、*RNASE2*、*RNASE6*、*RNF141*、*RNF217*、*RRAGD*、*S100A12*、*S100A9*、*S100Z*、*SAT2*、*SEMA3C*、*SEPTIN10*、*SERPINB1*、*SERPINB2*、*SESTD1*、*SGMS2*、*SH3PXD2B*、*SHTN1*、*SIPA1L2*、*SIRPD*、*SLC15A2*、*SLC1A3*、*SLC22A15*、*SLC22A4*、*SLC26A8*、*SLC31A1*、*SLC36A4*、*SLC7A7*、*SLCSA1*、*SLITRK4*、*SMAD1*、*SMPDL3A*、*SNX10*、*SOD2*、*SORT1*、*SPATA6*、*ST3GAL6*、*ST6GALNAC3*、*STEAP4*、*SULT1B1*、*TASL*、*TBC1D12*、*TCN2*、*TDRD9*、*TFEC*、*TGFB1*、*TLR1*、*TLR2*、*TLR4*、*TLR5*、*TLR6*、*TLR7*、*TLR8*、*TM6SF1*、*TMEM144*、*TMEM167A*、*TMTC2*、*TNFAIP6*、*TNFSF13B*、*TPST1*、*TSP0*、*UBE2D1*、*UGGT2*、*VNN2*、*VNN3*、*WASF1*、*WDFY3*、*WLS*、*ZC3H12C*、*ZFYVE16*、*ZNF438*。

棕色模块 GO 分析最显著的结果分别是骨髓白细胞激活(BP)、囊泡(CC)和模式识别受体活性(MF),见图 1D;KEGG 分析结果则主要为感染、免疫炎症等相关生物学过程的改变,包括 Toll 样受体信号通路等,见图 1E。

2.2 基因差异分析

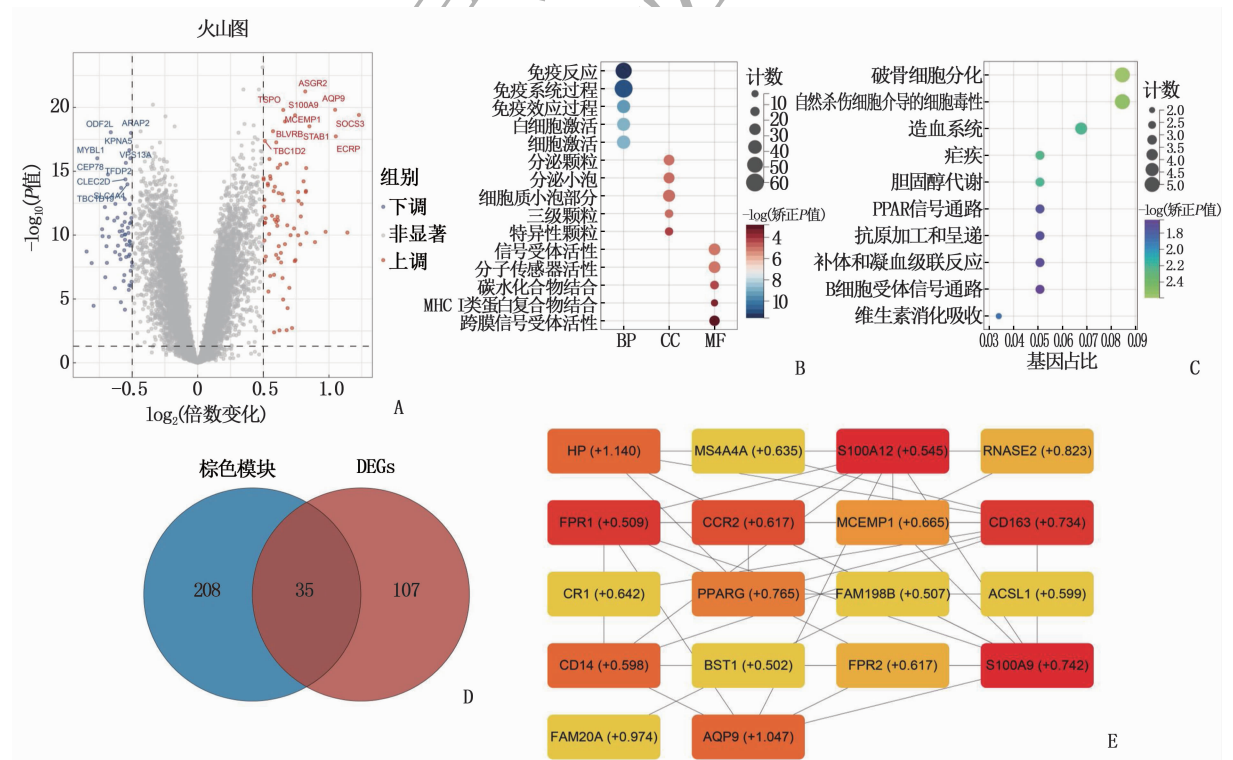
进一步对数据集进行基因差异分析,根据筛选标准共得到 142 个 DEGs(图 2A),其中上调的 DEGs 有 77 个,下调的 DEGs 有 65 个,见表 1。BP 分析结果主要包括免疫反应、免疫系统过程和免疫效应过程,CC 分析结果包括分泌颗粒、分泌小泡和细胞质小泡部分等,MF 分析结果主要包括信号受体活性、分子传感器活性和碳水化合物结合,见图 2B。KEGG 分析提示,这些 DEGs 主要涉及免疫、炎症和感染的相关生物学改变,见图 2C。DEGs 与上述棕色模块存在 35 个交集基因,见图 2D。基于 ChEA3 数据库对 35 个交集基因进行共同转录因子预测,其中平均排名分 ≤ 30 分的共同转录因

子分别是 CREB5、MTF1、NFE4、SPI1、ZNF467、NFE2、TFEC、MXD1、NR1H3、BORCS8MEF2B、NFIL3 和 CEBPE，见表 2。构建蛋白互作网络，其中有 18 个基因存在蛋白质相互作用，见图 2E。



A: 共表达网络的无标度拟合指数; B: 基因聚类树状图; C: 模块特征基因与临床特征相关性热图; D、E: GO 分析结果、KEGG 分析结果的气泡图展示。

图 1 易感基因模块的确定与功能注释



A: 差异基因火山图; B、C: 差异基因的 GO 分析、KEGG 分析结果; D: DEGs 与棕色模块的交集基因; E: 18 个交集基因的蛋白相互作用网络以及各基因的 $\log_2(\text{倍数变化})$ 改变量。

图 2 基因差异分析和蛋白互作网络

表 1 差异基因与交集基因

上调 DEGs	下调 DEGs	与棕色模块交集的 DEGs
ACSL1、ADM、AQP9、ASGR2、BLVRB、BST1、CCR1、CCR2、CD14、CD163、CES1、CES1P1、CRI、CYP1B1、CYP27A1、DSC2、DYSF、ECRP、EGRI、EGR2、FAM20A、FAM20C、FLVCR2、FMN1、FOLR3、FPR1、FPR2、FAM198B、GPR162、HLX、HP、IF-ITM3、IL1R2、KCNJ15、LILRA3、LILRA5、LILRB4、LRP1、MAFB、MCEMP1、MERTK、MIR21、MS4A4A、MTARC1、NFE2、NRG1、PADI2、PLB1、PLBD1、PPARG、QPCT、RNASE1、RNASE2、S100A12、S100A9、SASH1、SERPINB10、SERPINB2、SH3PXD2B、SIGLEC16、SIGLEC9、SLED1、SOC3、ST14、STAB1、STEAP4、TBC1D2、TCN2、TDRD9、TMEM150B、TMEM176A、TMEM176B、TNFAIP6、TREM1、TRIB1、TSPO、VSIG4	ARAP2、BIRC3、C12orf75、CD160、CD3G、CENPK、CEP78、CLC、CLEC2D、CPA3、DTHD1、EPB411A4、FAM169A、FCRL6、FGFBP2、GBP4、GNLY、HRH4、IL18RAP、JAKMIP2、KLRA1P、KLRC1、KLRC3、KLRC4-KLRK1、KLRD1、KLRF1、KLRG1、KPNAS、MIR29A、MS4A3、MYBL1、MYO6、ODF2L、PDGFD、PRSS23、RNU5A-1、RORA、SAMD3、SCARNA9、SH2D1A、SLC4A4、SLCO4C1、SNORA16B、SNORA20、SNO-RA75、SNORD20、SNORD28、SNORD30、SNORD45C、SNORD50A、SNORD54、SNORD58A、SNORD59A、SNORD59B、SNORD61、SNORD81、SNORD82、SYTL2、TARP、TBC1D19、TFDP2、TGFBF3、VPS13A、ZNF675、ZNF737	ACSL1、ADM、AQP9、BLVRB、BST1、CCR2、CD14、CD163、CRI、CYP1B1、DSC2、FAM20A、FLVCR2、FPR1、FPR2、FAM198B、HP、KCNJ15、LILRA5、MCEMP1、MS4A4A、MTARC1、PLBD1、PPARG、QPCT、RNASE2、S100A12、S100A9、SERPINB2、SH3PXD2B、STEAP4、TCN2、TDRD9、TNFAIP6、TSPO

表 2 共同转录因子预测结果(平均排名分≤30 分)

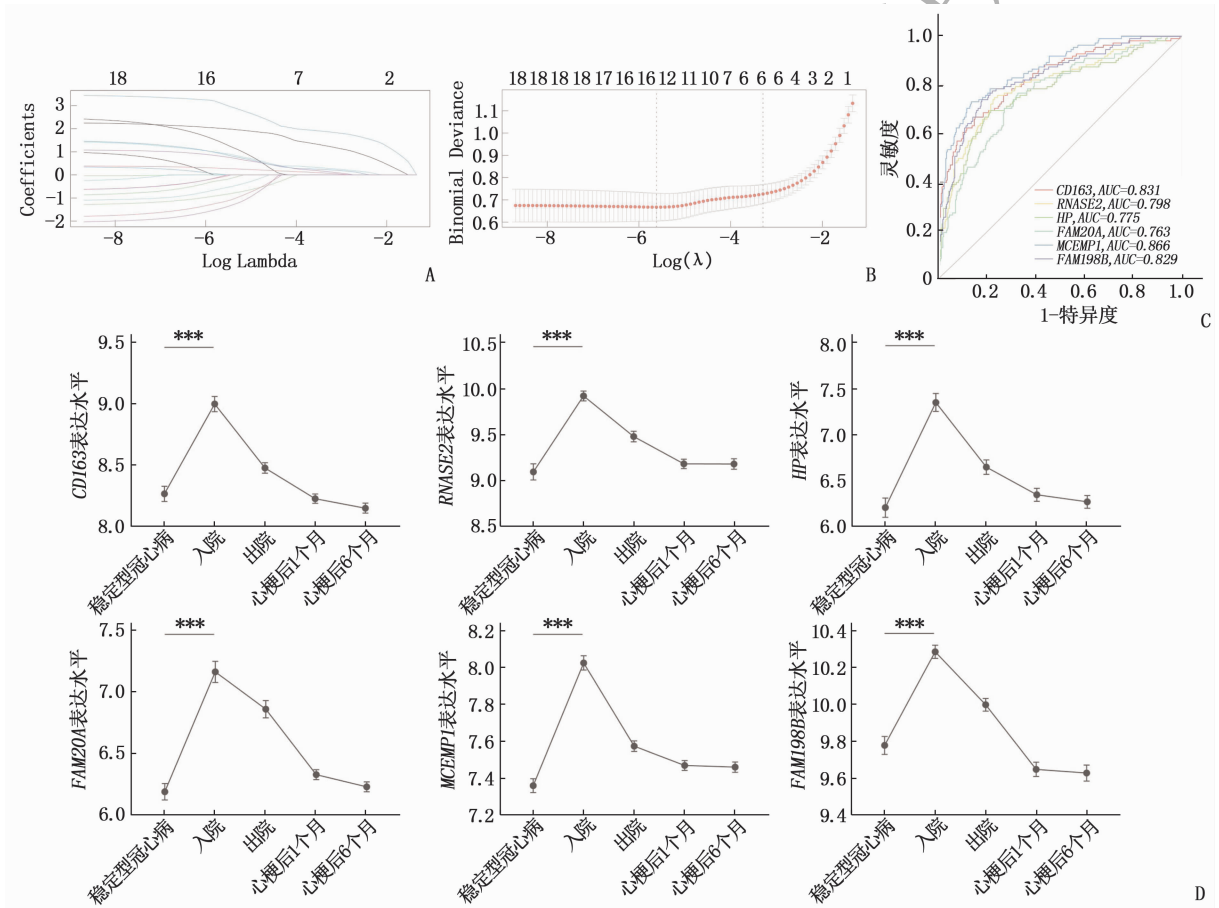
排名	转录因子	平均得分/分	转录因子数据库计算方法与得分	重叠基因
1	CREB5	7.667	ARCHS4 Coexpression, 5 分; Enrichr Queries, 14 分; GTEx Coexpression, 4 分	CRI、STEAP4、TNFAIP6、ACSL1、AQP9、FPR1、MCEMP1、ADM、FPR2、LILRA5、BST1、QPCT、PLBD1、CYP1B1、CD14、S100A9、CCR2
2	MTF1	8.500	ARCHS4 Coexpression, 6 分; GTEx Coexpression, 11 分	CRI、SERPINB2、TNFAIP6、ACSL1、AQP9、FPR1、FPR2、LILRA5、BST1、PLBD1、CYP1B1、FLVCR2、CD14、S100A9
3	NFE4	9.000	GTEx Coexpression, 9 分	BST1、CRI、AQP9、FPR1、PLBD1、MCEMP1、FPR2、LILRA5、CCR2
4	SPI1	9.667	Literature ChIP-seq, 5 分; ARCHS4 Coexpression, 11 分; ENCODE ChIP-seq, 8 分; Enrichr Queries, 11 分; ReMap ChIP-seq, 5 分; GTEx Coexpression, 18 分	CD163、CRI、STEAP4、SERPINB2、ACSL1、AQP9、FPR1、TDRD9、MCEMP1、FPR2、RNASE2、LILRA5、MS4A4A、BST1、QPCT、PLBD1、CYP1B1、CD14、S100A9、CCR2
5	ZNF467	19.333	ARCHS4 Coexpression, 18 分; Enrichr Queries, 30 分; GTEx Coexpression, 10 分	CRI、AQP9、FPR1、MCEMP1、ADM、FPR2、LILRA5、BST1、PLBD1、CYP1B1、BLVRB、CD14、S100A9、CCR2
6	NFE2	21.000	ARCHS4 Coexpression, 16 分; ENCODE ChIP-seq, 78 分; Enrichr Queries, 1 分; ReMap ChIP-seq, 8 分; GTEx Coexpression, 2 分	CRI、SERPINB2、AQP9、FPR1、MCEMP1、ADM、FPR2、RNASE2、LILRA5、BST1、TCN2、QPCT、PLBD1、S100A12、BLVRB、TSPO、CYP1B1、PPARG、CD14、S100A9、CCR2
7	TFEC	21.667	ARCHS4 Coexpression, 21 分; Enrichr Queries, 4 分; GTEx Coexpression, 40 分	CD163、CRI、SERPINB2、TNFAIP6、ACSL1、AQP9、FPR1、MS4A4A、BST1、TCN2、FLVCR2、PPARG、CD14、S100A9、CCR2
8	MXD1	23.000	ARCHS4 Coexpression, 4 分; Enrichr Queries, 9 分; GTEx Coexpression, 56 分	CRI、SERPINB2、STEAP4、ACSL1、AQP9、FPR1、ADM、FPR2、LILRA5、BST1、PLBD1、CYP1B1、CD14、S100A9
9	NR1H3	26.600	Literature ChIP-seq, 10 分; ARCHS4 Coexpression, 17 分; Enrichr Queries, 10 分; ReMap ChIP-seq, 36 分; GTEx Coexpression, 60 分	CD163、STEAP4、TCN2、ACSL1、FLVCR2、CYP1B1、PPARG、ADM、CD14、S100A9、MS4A4A
10	BORCS8MEF2B	29.000	ARCHS4 Coexpression, 29 分	AQP9、FPR1、TSPO、FPR2、S100A9、LILRA5
11	NFH3	30.000	ARCHS4 Coexpression, 1 分; Enrichr Queries, 21 分; GTEx Coexpression, 68 分	CRI、TNFAIP6、ACSL1、AQP9、FPR1、ADM、FPR2、LILRA5、MS4A4A、BST1、CYP1B1、PLBD1、CD14、S100A9
12	CEBPE	30.000	ARCHS4 Coexpression, 31 分; Enrichr Queries, 12 分; GTEx Coexpression, 47 分	BST1、FPR1、PLBD1、ADM、MCEMP1、FPR2、CD14、RNASE2、S100A9、CCR2、LILRA5

计算方法包括 ARCHS4 Coexpression、ENCODE ChIP-seq、Enrichr Queries、GTEx Coexpression、Literature ChIP-seq、ReMap ChIP-seq; 平均得分为各计算方法得分之和除以计算方法数量所得之商。

2.3 关键基因的确 定

对 18 个存在蛋白相互作用的基因进一步行 LASSO 回归分析,这些基因不同惩罚参数值所对应的系数见图 3A。本研究选取最小二项偏差在 1 个标准误差内的 λ 值 (lambda. 1se, 6 个基因), 该 λ 值提供更精简的模型,见图 3B。ROC 曲线分析显示, *CD163*、*RNASE2*、*HP*、*FAM20A*、*MCEMP1*

和 *FAM198B* 基因表达水平对心肌梗死的发生均具有良好的诊断价值, *AUC* 分别为 0.831、0.798、0.775、0.763、0.866 和 0.829, 见图 3C。与对照组稳定型冠心病患者相比,这些关键基因的表达水平在入院时心肌梗死患者中显著上升,并在出院时、心肌梗死后 1 个月、心肌梗死后 6 个月逐步下降,见图 3D。



A、B: 基于惩罚项的 LASSO 回归系数图和交叉验证图; C: 基于 ROC 分析评估关键基因表达在预测心肌梗死中的作用; D: 关键基因在不同时点的表达折线图(两者比较, ****P* < 0.001)。

图 3 LASSO 回归与 ROC 分析

2.4 GSEA 结果和验证

GSEA 结果提示,这些关键基因涉及的生物学改变主要与糖脂代谢、活性氧、免疫炎症等有关,见图 4。这些基因在外部数据集中大多也存在差异表达。基于外部验证数据集 GSE123342, 本研究同样发现,与稳定型冠心病相比, *HP*、*FAM198B*、*CD163*、*FAM20A*、*MCEMP1* 表达水平在急性心肌梗死发生时显著上升,而 *RNASE2* 仅表现出上升趋势,见图 5。

3 讨 论

本研究基于 PBMCs 转录组学的变化探讨心

肌梗死及梗死恢复过程中可能存在的生物学改变和潜在标志物, WGCNA 和 DEGs 分析结果显示,心肌梗死过程伴随着免疫炎症紊乱。基于 LASSO 分析,本研究从上述 2 种分析方法得到的共同基因中鉴定出 6 个关键基因,即 *CD163*、*RNASE2*、*HP*、*FAM20A*、*MCEMP1* 和 *FAM198B*。ROC 曲线分析结果提示,这些关键基因对心肌梗死的发生均具有较高的诊断价值。本研究还发现,在梗死恢复过程中,这些基因的表达呈现明显下降趋势。GSEA 分析结果表明,这些关键基因涉及的生物学改变主要与糖脂代谢、活性氧、免疫炎症等有关。

CD163 分子是 I 型膜蛋白,蛋白表达限于单

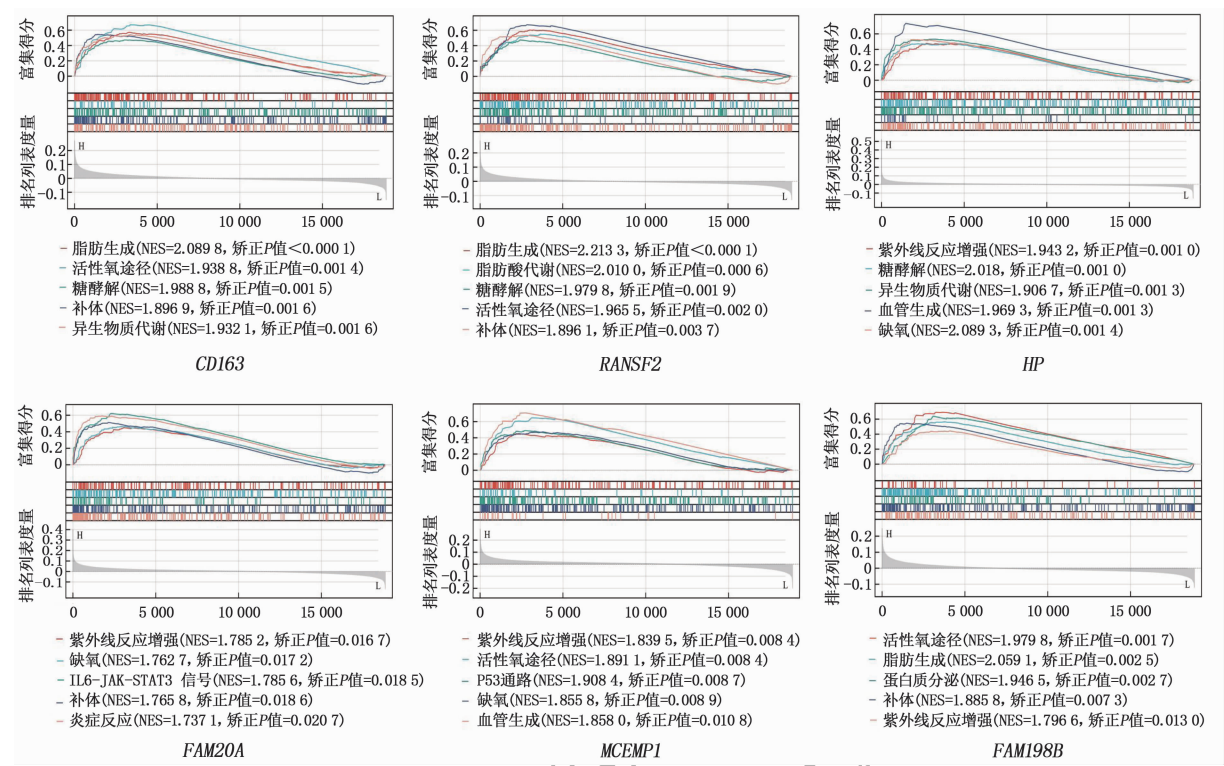
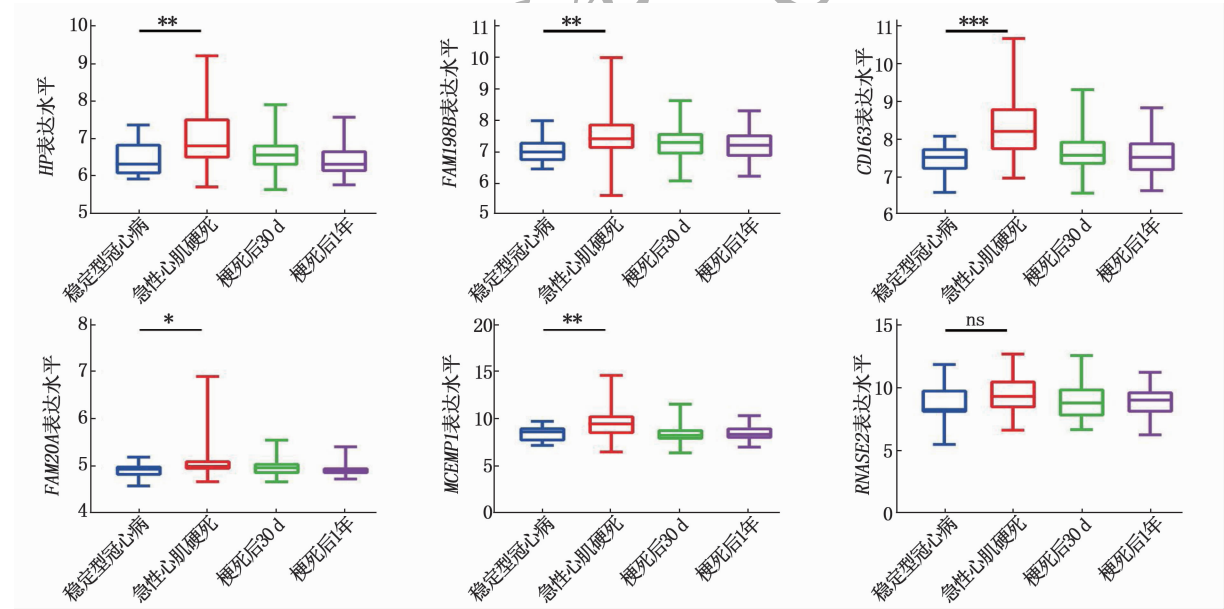


图4 关键基因的 GSEA 分析结果



两者比较, ns > 0.05, *P < 0.05, **P < 0.01, ***P < 0.001。

图5 外部数据集验证关键基因的表达情况

核细胞/巨噬细胞系, *CD163* 抗原特异性释放机制可能在炎症调节过程中起重要作用。在动脉粥样硬化过程中, *CD163*⁺ 巨噬细胞能促进血管生成和增加血管通透性,并伴随炎症反应;此外,破裂的冠状动脉斑块中 *CD163* 的表达增加了心肌梗死和冠心病发生风险^[24]。*RNASE2* 编码的蛋白质是非分泌型核糖核酸酶,属于胰核糖核酸酶家族^[25]。*FAM20A* 是激酶编码基因家族的成员,本

身不具有激酶活性,通过与家族成员 *FAM20C* 形成复合物,增强 *FAM20C* 的激酶活性,从而使分泌通路内的蛋白磷酸化^[26]。*MCEMP1* 又称 *C19ORF59*,可编码表达于肥大细胞、巨噬细胞等的跨膜蛋白^[27]。研究^[27]表明, *MCEMP1* 基因可能是卒中诊断和预后评估的新生物标志物。生物信息学研究^[28-29]提示, *RNASE2*、*FAM20A* 和 *MCEMP1* 可能是心肌梗死的关键基因,然而这些

基因在心肌梗死及梗死恢复过程中的具体作用机制有待进一步研究。*HP* 基因编码一种前蛋白,经处理后产生 α 链和 β 链,结合为四聚体产生触珠蛋白,其基因型与急性心肌梗死的发病风险密切相关,并且能决定心肌梗死面积^[30-31]。*FAM198B* (*GASK1B*)是目前未知功能的新基因,可能编码定位在高尔基体上的膜结合糖蛋白,被认为参与癌症的转移与进展^[32]。本研究分析结果显示,*FAM198B* 可能是心肌梗死和反映心肌梗死恢复的关键基因。

本研究尚存在一定局限性:首先,本研究基于 GEO 数据库中下载的心肌梗死患者 PBMCs 的测序数据集进行分析,还需要进一步开展分子生物学实验来验证这些基因在心肌梗死中的表达及其可能的生物学机制;其次,本研究仅分析了现有的样本数据,未来还需基于更大的样本规模和更详细的样本类型进一步深入研究;最后,由于缺乏临床资料,本研究未分析这些关键基因的表达与心肌梗死患者年龄、性别、射血分数等临床指标的相关性。

综上所述,本研究基于 PBMCs 确定了 6 个可能与心肌梗死密切相关的关键基因,即 *CD163*、*RNASE2*、*HP*、*FAM20A*、*MCEMP1* 和 *FAM198B*,为心肌梗死的诊断、评估、管理和预后分析提供了新的思路。

参考文献

- [1] ZHAO D, LIU J, WANG M, *et al.* Epidemiology of cardiovascular disease in China: current features and implications[J]. *Nat Rev Cardiol*, 2019, 16(4): 203-212.
- [2] 商广芸, 高燕飞, 曹洪波, 等. 急性心肌梗死的危险因素分析[J]. *实用临床医药杂志*, 2020, 24(12): 48-50.
- [3] WEIL B R, NEELAMEGHAM S. Selectins and immune cells in acute myocardial infarction and post-infarction ventricular remodeling: pathophysiology and novel treatments[J]. *Front Immunol*, 2019, 10: 300.
- [4] DE WINTER R J, KOSTER R W, STURK A, *et al.* Value of myoglobin, troponin T, and CK-MBmass in ruling out an acute myocardial infarction in the emergency room[J]. *Circulation*, 1995, 92(12): 3401-3407.
- [5] DE LEMOS J A, DRAZNER M H, OMLAND T, *et al.* Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population[J]. *JAMA*, 2010, 304(22): 2503-2512.
- [6] GE Y, WANG T J. Identifying novel biomarkers for cardiovascular disease risk prediction[J]. *J Intern Med*, 2012, 272(5): 430-439.
- [7] WANG Y, ZHANG X Y, DUAN M, *et al.* Identification of potential biomarkers associated with acute myocardial infarction by weighted gene coexpression network analysis[J]. *Oxid Med Cell Longev*, 2021, 2021: 5553811.
- [8] SEN P, KEMPPAINEN E, OREŠIĆ M. Perspectives on systems modeling of human peripheral blood mononuclear cells[J]. *Front Mol Biosci*, 2017, 4: 96.
- [9] BARRETT T, WILHITE S E, LEDOUX P, *et al.* NCBI GEO: archive for functional genomics data sets; update[J]. *Nucleic Acids Res*, 2013, 41(Database issue): D991-D995.
- [10] MACIEJAK A, KILISZEK M, MICHALAK M, *et al.* Gene expression profiling reveals potential prognostic biomarkers associated with the progression of heart failure[J]. *Genome Med*, 2015, 7(1): 26.
- [11] LANGFELDER P, HORVATH S. WGCNA: an R package for weighted correlation network analysis[J]. *BMC Bioinformatics*, 2008, 9: 559.
- [12] DAS S, MEHER P K, RAI A, *et al.* Statistical approaches for gene selection, hub gene identification and module interaction in gene co-expression network analysis: an application to aluminum stress in soybean (glycine max l.) [J]. *PLoS One*, 2017, 12(1): e0169605.
- [13] 朱宝华, 孙岩. 冠状动脉疾病差异基因富集和加权基因共表达网络分析[J]. *实用临床医药杂志*, 2021, 25(17): 15-21.
- [14] YU G C, WANG L G, HAN Y Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284-287.
- [15] RITCHIE M E, PHIPSON B, WU D, *et al.* Limma Powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [16] KEENAN A B, TORRE D, LACHMANN A, *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration[J]. *Nucleic Acids Res*, 2019, 47(W1): W212-W224.
- [17] FU Z J, XU Y Q, CHEN Y, *et al.* Construction of miRNA-mRNA-TF regulatory network for diagnosis of gastric cancer[J]. *Biomol Res Int*, 2021, 2021: 9121478.
- [18] SZKLARCZYK D, GABLE A L, NASTOU K C, *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. *Nucleic Acids Res*, 2021, 49(D1): D605-D612.
- [19] SHANNON P, MARKIEL A, OZIER O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Res*, 2003, 13(11): 2498-2504.
- [20] MCELIGOT A J, POYNOR V, SHARMA R, *et al.* Logistic LASSO regression for dietary intakes and breast cancer[J]. *Nutrients*, 2020, 12(9): 2652.

为中点,分为 6 等份(图 1),并将直径的 1/6 作为取样框的大小。对于血管条件良好且无明显颈动脉斑块的患者而言,血管内的血流主要为层流,改良环形近似法无明显优势($P>0.05$),反而增大了测量难度;对于存在动脉斑块的患者而言,改良环形近似法则可显著提高血流量测量的稳定性,原因可能是血管斑块导致血流为非层流,血流方向可能与管壁不平行,可能会出现涡流、返流和二次流等复杂流动形态^[11]。

本研究还存在许多不足之处:只分析了血流速度及血流量测量的稳定性,缺少与真实血流量的比较;样本量较小,需进一步扩大样本量,减少误差;在改良环形近似法的比较中,因缺乏软件未与环形近似法进行比较;在三角函数单双切面测量法中,运用沿椭圆形长轴 5 个等分点测得血流速度,然后运用三角函数得到垂直血管平面 5 个点的血流速度,其与真实平面 5 个点的测量速度可能并不相等。

综上所述,三角函数单切面测量法可测得较为稳定的血流速度,对于颈动脉斑块患者,改良环形近似法可获得较稳定的血流量,但评估稳定性有待进一步提高。

参考文献

[1] 周学华,李玉凤,沈霞. 围术期成人脑血流监测方法的研究进展[J]. 复旦学报:医学版, 2022, 49(4): 620-627.

[2] 沈莹莹,杜宜纲,朱磊,等. 基于向量速度的超声血流量计算方法探讨[J]. 中国医疗器械信息, 2019, 25(1): 25-28.

[3] 曹铁生,段云友,游国雄,等. 椎动脉二维超声显象及脉冲多普勒血流量定量[J]. 中国超声医学杂志, 1989, 5(4): 213-216.

[4] 李卫民. 硝酸甘油对脑动脉血流影响的超声检测[D]. 南京:东南大学, 2016.

[5] HOLBEK S, EWERTSEN C, BOUZARI H, *et al.* Ultrasonic 3-D vector flow method for quantitative in vivo peak velocity and flow rate estimation[J]. IEEE Trans Ultrason Ferroelectr-Freq Control, 2017, 64(3): 544-554.

[6] 陈友岱,宋孟龙,李依,等. 危重患者股动脉与颈总动脉血流量的关系[J]. 实用医院临床杂志, 2017, 14(5): 50-52.

[7] EICKE B M, TEGELER C H, DALLEY G, *et al.* Angle correction in transcranial Doppler sonography[J]. J Neuroimaging, 1994, 4(1): 29-33.

[8] GILL R W. Measurement of blood flow by ultrasound: accuracy and sources of error[J]. Ultrasound Med Biol, 1985, 11(4): 625-641.

[9] ZHAO S Z, XU X Y, HUGHES A D, *et al.* Blood flow and vessel mechanics in a physiologically realistic model of a human carotid arterial bifurcation[J]. J Biomech, 2000, 33(8): 975-984.

[10] HANSEN K L, UDESEN J, GRAN F, *et al.* Fast blood vector velocity imaging using ultrasound: in-vivo examples of complex blood flow in the vascular system[C]. IEEE Intern UltrasonSymp, 2008: 1068-1071.

[11] HANSEN K L, MØLLER-SØRENSEN H, KJAERGAARD J, *et al.* Analysis of systolic backflow and secondary helical blood flow in the ascending aorta using vector flow imaging[J]. Ultrasound Med Biol, 2016, 42(4): 899-908.

(本文编辑:周冬梅)

(上接第 28 面)

[21] ENGBRETSSEN S, BOHLIN J. Statistical predictions with glmnet[J]. Clin Epigenetics, 2019, 11(1): 123.

[22] 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.

[23] SUBRAMANIAN A, TAMAYO P, MOOTHA V K, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles[J]. Proc Natl Acad Sci USA, 2005, 102(43): 15545-15550.

[24] GUO L, AKAHORI H, HARARI E, *et al.* CD163⁺ macrophages promote angiogenesis and vascular permeability accompanied by inflammation in atherosclerosis[J]. J Clin Invest, 2018, 128(3): 1106-1124.

[25] WU T F, CHEN Y X, YANG L Y, *et al.* Ribonuclease A family member 2 promotes the malignant progression of glioma through the PI3K/Akt signaling pathway[J]. Front Oncol, 2022, 12: 921083.

[26] LI L L, SAIYIN W, ZHANG H, *et al.* FAM20A is essential for amelogenesis, but is dispensable for dentinogenesis[J]. J Mol Histol, 2019, 50(6): 581-591.

[27] RAMAN K, O'DONNELL M J, CZLONKOWSKA A, *et al.*

Peripheral blood *MCEMP1* gene expression as a biomarker for stroke prognosis[J]. Stroke, 2016, 47(3): 652-658.

[28] QIU L, LIU X. Identification of key genes involved in myocardial infarction[J]. Eur J Med Res, 2019, 24(1): 22.

[29] KILISZEK M, BURZYNSKA B, MICHALAK M, *et al.* Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction[J]. PLoS One, 2012, 7(11): e50054.

[30] GURUNG R L, YIAMUNAA M, LIU S, *et al.* Association of haptoglobin phenotype with incident acute myocardial infarction in Chinese patients with type 2 diabetes[J]. Cardiovasc Diabetol, 2019, 18(1): 65.

[31] BLUM S, ASAF R, GUETTA J, *et al.* Haptoglobin genotype determines myocardial infarct size in diabetic mice[J]. J Am Coll Cardiol, 2007, 49(1): 82-87.

[32] HSU C Y, CHANG G C, CHEN Y J, *et al.* FAM198B is associated with prolonged survival and inhibits metastasis in lung adenocarcinoma via blockage of ERK-mediated MMP-1 expression[J]. Clin Cancer Res, 2018, 24(4): 916-926.

(本文编辑:陆文娟)