DOI: 10.14188/j.1671-8852.2024.0161　　**文章编号** 1671-8852(2025)10-1282-11

基于B细胞相关基因构建心力衰竭多基因诊断列线图

凌　懿[1]　梁晶晶[1]　潘治豪[1]　王扬淦[1,2]
武汉大学中南医院　[1]心血管内科，[2]综合医疗科　湖北　武汉　430071

摘要　目的：运用生物信息学和分子生物学方法探究B细胞相关基因在心力衰竭(HF)中的诊断价值。方法：应用CD19免疫荧光检测HF模型小鼠心脏B细胞浸润水平。对26例HF患者心脏组织单细胞测序结果分析确定B细胞相关基因。从GEO数据库下载6个HF相关microarray测序数据集及2个HF相关RNA-seq测序数据集，分别将其合并作为训练集和验证集。使用差异基因分析、PPI网络分析、逻辑回归、BORUTA算法、LASSO回归及SVM-RFE算法进行特征选择。利用多因素逻辑回归构建诊断列线图，并通过ROC、Calibration以及DCA分析其可信度。通过多重免疫荧光验证结论。结果：与对照组相比，HF组心脏B细胞浸润水平上升($P<0.05$)。单细胞测序分析发现125个B细胞相关基因，其中86个在组间差异表达。通过多种算法分析，诊断列线图包括CCL5、CD74、ITGB2、KLRD1和SELL 5个基因。该列线图在训练集及验证集中均被证实为一个可靠的HF诊断工具。免疫荧光显示，HF组小鼠心肌组织具有更高水平$CD19^+$ $KLRD1^+$ B细胞浸润($P<0.05$)。结论：基于B细胞相关基因构建了诊断HF的5基因列线图，为阐明HF的发病机制提供了一个潜在的工具和可能的切入点。

关键词　B细胞；心力衰竭；诊断标志物；列线图；KLRD1；生物信息学
中图分类号　R541.6
文献标识码　A　　**开放科学(资源服务)标识码(OSID)：**

Constructing a multigene diagnostic nomogram for heart failure based on B cell-related genes

LING Yi[1], LIANG Jingjing[1], PAN Zhihao[1], WANG Yanggan[1,2]
[1]Dept. of Cardiovascular Medicine, [2]Dept. of Integrative Medicine, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei, China

Abstract

Objective: To comprehensively apply bioinformatics and molecular biology methods to explore the diagnostic value of B cell-related genes in heart failure (HF). **Methods:** CD19 immunofluorescence staining was used to detect the B cell infiltration level in HF model mice′s hearts. Single-cell RNA sequencing (scRNA-seq) data from the heart tissues of 26 HF patients were analyzed to determine B cell-related genes. Six microarray and two RNA-seq data sets related to HF were downloaded from the Gene Expression Omnibus (GEO) database and merged as training and validation sets, respectively. Differential gene analysis, protein-protein interaction (PPI) network analysis, logistic regression, Boruta algorithm, LASSO regression, and support vector machine recursive feature elimination (SVM-RFE) algorithms were used for feature selection. A diagnostic nomogram was constructed using multiple-factor logistic regression, and its reliability was assessed through receiver operating char-

课题来源：国家自然科学基金资助项目(编号：8207021743)
作者简介：凌懿，女，1998—，医学硕士，主要从事心力衰竭及心律失常相关研究，E-mail: 1083736834@qq.com
通信作者：王扬淦，男，1963—，医学博士，主任医师，教授，博士生导师，主要从事心律失常、电生理及心力衰竭等方面的研究，E-mail: wb000813@whu.edu.cn

acteristic (ROC), calibration, and decision curve analysis (DCA). The results were validated by multiplex immunofluorescence. The results were validated by multiplex immunofluorescence. **Results:** Compared with that in the control group, the level of B cell infiltration in the heart of HF patients was increased ($P<0.05$). Single-cell RNA sequencing analysis identified 125 B cell-related genes, of which 86 were differentially expressed between groups. Analysis using various algorithms revealed that the diagnostic nomogram included five genes: CCL5, CD74, ITGB2, KLRD1, and SELL. The nomogram was confirmed to be a reliable HF diagnostic tool in the training and validation sets. Immunofluorescence showed that HF model mice had higher $CD19^+$ $KLRD1^+$ B cell infiltration levels in myocardial tissue ($P<0.05$). **Conclusion:** A 5-gene diagnostic nomogram for HF was constructed based on B cell-related genes, providing a potential tool and a possible entry point for elucidating the pathogenesis of HF.

Key Words　B Cells; Heart Failure; Diagnostic Markers; Nomogram; KLRD1; Bioinformatics

心力衰竭(heart failure, HF)是一种以呼吸困难、踝关节肿胀等为主要症状,并伴有颈静脉怒张、肺部啰音等体征的临床综合征[1]。由于人口老龄化,全球心力衰竭总体发病率在逐渐增加[2]。早期和及时的诊断对心力衰竭患者的治疗和预后至关重要。目前对于心力衰竭的诊断除了症状和体征外,还需结合生物标志物辅助诊断,例如B型利钠肽(BNP)和N末端B型利钠肽前体(NT-proBNP),这些生物标志物在HF诊断和风险分层中的作用已得到广泛证明,是人群筛查的新兴工具[3]。但BNP或NT-proBNP对于HF诊断的特异性不高,有许多其他原因也能引起其升高,如心房颤动或肾病等[4]。并且心力衰竭临床症状是多种多样的,在临床实践中很容易漏诊或误诊,因此我们仍然需要寻找能够辅助诊断心力衰竭更有价值的生物标志物。

越来越多的证据表明,B淋巴细胞与心脏密切相关。B淋巴细胞存在于健康和受损的心肌内,作为必需的免疫调节剂,B淋巴细胞在维持免疫稳态和预防自身免疫性疾病方面起着至关重要的作用[5]。心力衰竭的特征之一是免疫反应的产生。导致心力衰竭的诸多病因如缺血或血流动力学过载引起的心肌损伤可导致心脏固有免疫系统和适应性免疫系统的激活[6]。固有免疫系统是针对病原体或组织损伤的全局非特异性防御,其诱导的炎症反应可上调一系列细胞保护反应,为心脏提供短期适应增加的应激(生理性炎症)。而适应性免疫系统是固有免疫系统激活导致促炎细胞因子和趋化因子的表达增加,进而通过激活B细胞和T细胞所介导的高度特异性应答。然而,炎症反应可能变得失调,导致慢性炎症,从而导致左室功能障碍和左室重塑。在诱导压力超负荷时,B细胞亚群在疾病的早期和晚期都被激活并进行扩增(或从外周招募)[7]。可见B淋巴细胞在心力衰竭中表现活跃。有研究[8]表明B细胞不仅可作为开发新型免疫调节治疗心力衰竭的潜在靶点,其相关生物标志物对于诊断心力衰竭的潜在价值也值得探索。

1　材料与方法

1.1　实验材料

1.1.1　*实验动物*　本研究中所使用10只SPF级雄性C57BL/6小鼠,购买于湖北省预防医学科学院-湖北省实验动物研究中心,小鼠为6～8周龄,体质量22～25 g。实验设计经武汉大学动物伦理委员会批准,动物使用过程中的所有操作符合美国国立卫生研究院出版的《实验动物护理和使用指南》。

1.1.2　*数据集*　从基因表达数据库(Gene Expression Omnibus, GEO, https://ncbi.nlm.nih.gov/geo/)下载人类HF数据集GSE145154,GSE16499、GSE26887、GSE42955、GSE57338、GSE76701和GSE79962(表1)。GSEA分析的参考数据集下载于MSigDB数据库(https://www.gsea-msigdb.org/gsea/msigdb/index.jsp)。

1.1.3　*公共数据库*　STRING数据库(https://cn.string-db.org/)用于构建蛋白质-蛋白质相互作用(Protein-Protein Interaction, PPI)网络。利用Metascape数据库(https://metascape.org/gp/index.html#/main/step1)进行功能富集分析。

1.2　实验方法

1.2.1　*小鼠压力负荷型心力衰竭模型及假手术模型的构建*　利用随机数字表法,将10只小鼠按照1∶1随机划分为HF模型组(HF)和假手术组(Sham)。

表1 本研究所使用的GEO数据集详细信息

GEO编号	实验类型	分类	样本量(对照组/心力衰竭组)	地区	参考文献(PMID)	种属	组织来源
GSE145154	单细胞RNA测序	训练集	4/26	中国	34601654	人类	心脏
GSE16499	微阵列测序	训练集	15/15	美国	20124440	人类	心脏
GSE26887	微阵列测序	训练集	5/19	意大利	22427379	人类	心脏
GSE42955	微阵列测序	训练集	5/24	西班牙	24339868	人类	心脏
GSE57338	微阵列测序	训练集	136/177	美国	25528681	人类	心脏
GSE76701	微阵列测序	训练集	4/4	美国	26756417	人类	心脏
GSE79962	微阵列测序	训练集	11/20	美国	28067713	人类	心脏
GSE46224	RNA测序	验证集	8/16	美国	24429688	人类	心脏
GSE116250	RNA测序	验证集	14/50	美国	32868781	人类	心脏

HF模型组采用主动脉弓缩窄术(transverse aortic constriction,TAC)建立心力衰竭模型,假手术组动物,操作一致,但不做主动脉结扎,作为对照。

1.2.2 小鼠心脏超声 在造模手术两周后使用小动物超声心动图仪(Visual Sonics Vevo 2100)进行心脏超声检测验证造模是否成功。用1.0%异氟烷麻醉小鼠用于M模式成像,测量左室收缩末期内径(LVESd)、左室舒张末期内径(LVEDd)、左室前壁舒张末期厚度(LVAWT)、左室后壁舒张末期厚度(LVPWT),取3次心动周期的平均值作为最终测量结果。在这些参数测量的基础上,计算左室射血分数(LVEF)。LVEF<50%认为造模成功。

1.2.3 Masson染色 心脏组织经4%多聚甲醛固定72 h,脱水浸蜡包埋切片,放入二甲苯及乙醇中梯度脱蜡。切片放入铁苏木素染液10 min,分化后返蓝5 min,丽春红品红染色10 min,乙醇脱水,透明处理后中性树胶封片。镜下胶原纤维呈天蓝色至深蓝色,正常心肌细胞呈红色。

1.2.4 HE染色 石蜡切片先用苏木素染15 min,再用伊红染3 min,经乙醇脱水和透明处理后封片。光镜下观察小鼠心肌组织形态,胞核呈蓝色,胞质呈深浅不同的粉红色至桃红色,胶原纤维呈淡粉红色,弹力纤维呈亮粉红色,红细胞呈橘红色。

1.2.5 免疫荧光分析 切片脱蜡,抗原修复后加双氧水孵育。BSA封闭后,滴加1:500稀释的CD19抗体(GB12061-100,Servicebio),4 ℃过夜。PBS洗,加二抗,室温孵育50 min。PBS洗,加TSA避光孵育10 min。柠檬酸修复液处理10 min去除抗体。BSA封闭后,滴加1:200稀释的KLRD1抗体(DF6773,Affinity),4 ℃过夜。二抗孵育,重复上述步骤。DAPI染色,荧光淬灭处理,封片,采集图像。

1.2.6 B细胞相关基因确定 利用R软件的Seurat软件包处理GSE145154数据集中单细胞数据。使用主成分分析(PCA)及t分布随机邻域嵌入(t-SNE)维度聚类分析对细胞进行二维可视化处理。通过Seurat包的"Find All Markers"功能识别的细胞标记基因和过滤阈值设置如下:$|\log_2FC|>1$且错误发现率(False Discovery Rate,FDR)<0.05。最后,利用R中的SingleR包进行细胞类型注释。最后得到的B细胞Marker基因被认为是B细胞相关基因纳入进一步的分析[9-11]。

1.2.7 数据集处理 为了减少数据集的批次效应,利用"sva"包进行矫正,并用主成分分析(PCA)验证效果。将处理后的数据集合并为"Meta-Microarray"训练集(sham=176,HF=259),并经过$\log_2$转换和批次矫正。从GEO数据库收集了两个RNA-seq的HF队列(表1),用同样方法去除批次效应,合并成"Meta-RNAseq"验证集(sham=22,HF=66)。

1.2.8 差异基因分析 利用R软件的"limma"包以FDR<0.05作为筛选标准,将上述得到的B细胞相关基因筛选能得到与心力衰竭相关的差异表达基因。

1.2.9 PPI网络构建 将可信度设置为0.4,构建PPI网络,利用Cytoscape软件(版本:3.8.0)对该网络进行进一步可视化。使用"Boruta"包进行BORUTA算法分析,"glmnet"R包进行LASSO回归分析及"caret"R包通过支持向量机递归特征消除(SVM-RFE)对初步筛选出的差异基因进一步分析得到对心力衰竭具有更高诊断价值的基因。

1.2.10 功能富集分析和基因集富集分析(GSEA) 将获得的差异基因输入Metascape数据库进行基因功能富集分析,使用MSigDB数据库的GOBP数据集,收集B细胞相关的基因集,作为GSEA的参考数据集。通过GSEA软件(4.3.2版)进行分析,认为Nominal $P<0.05$且FDR<0.25的通路具有显著意义。

1.2.11　预测模型构建和验证　使用“rms”包对筛选出的基因进行单因素逻辑回归，得到相应的概率（*OR*）和回归系数（β）。根据以下公式：风险评分＝常数＋$\sum_{i=1}^{n}\beta_i \times \mathrm{EXP}(\mathrm{gene})_i$，从而预测 HF 的发生风险。我们用列线图可视化 HF 预测模型，为每个基因分配相应分数并累加得到总分，以此确定 HF 的风险。通过绘制标定曲线以评估列线图预测准确性，并根据预测风险与实际结果的一致性对模型的标定进行描述。使用“rms”包建立 Calibration 校准图，横轴表示模型预测概率，纵轴表示实际概率，用中位数加均值的 error bar 形式表征，并绘制了一条斜率为 1 的理想曲线作为参照，使用 R 软件的“pROC”包分别在训练集和验证集中绘制受试者工作特征（ROC）曲线，计算曲线下面积（*AUC*）及其对应的 95% 置信区间（*CI*）。该 R 包也被用于德龙检验（Delong-test）以比较两个 *AUC* 值的大小。此外，我们使用“rmda”包绘制决策曲线（DCA）来计算诊断特征的净收益并评估诊断性能。

1.2.12　统计学分析　本研究统计分析基于 R 软件（版本 3.6.3）。计量资料以均数±标准差表示，首先进行正态性和方差齐性检验，检验符合正态分布且方差齐，组间采用配对或非配对 *t* 检验，多组间比较采用单因素方差分析或重复测量数据的方差分析。在本研究中，若无特别说明，以 $P<0.05$ 为差异具有统计学意义。

2　结果

2.1　HF 模型小鼠心肌组织中 B 细胞浸润水平升高　2 周后，相对于假手术组小鼠，心脏超声发现 HF 组小鼠心肌肥厚，心腔扩大，LVEF 平均为 38.3%（$P<0.05$，图 1A）。Masson 染色显示 HF 组心肌细胞纤维化明显，心腔增大明显（图 1B）。HE 染色结果显示 HF 组心肌细胞排列紊乱，心肌纤维化及心肌细胞横截面积显著增加（图 1C）。B 细胞最具特征的标志物是 CD19[8]，利用针对 CD19 特异性抗体进行免疫荧光染色发现，与假手术组相比，HF 组小鼠心脏切片中 $CD19^+$ B 细胞浸润明显（$P<0.05$，图 1D），提示 B 细胞可能在 HF 发生过程中发挥重要作用。

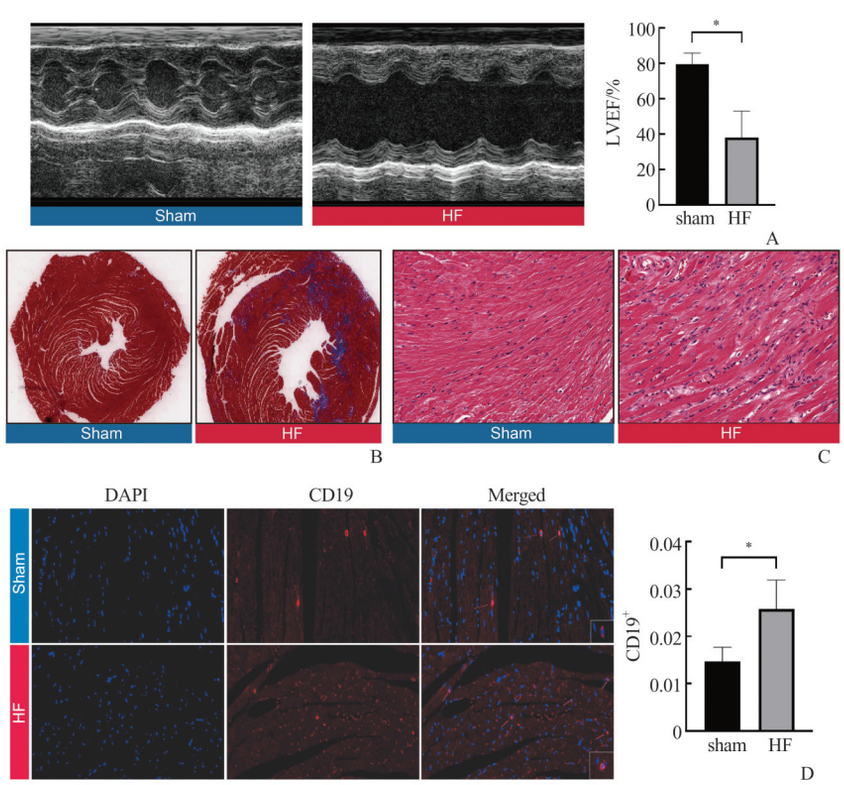


图 1　HF 模型小鼠心肌组织中 B 细胞浸润水平上升

A. HF 组小鼠 LVEF 显著下降且均低于 50%；B. HF 组小鼠心肌组织中胶原纤维（蓝色为纤维化）含量显著增加；C. HF 组小鼠心肌组织中心肌细胞排列紊乱（蓝色为细胞核），心肌细胞扭曲、肥大，细胞间隙增加，血管周围细胞外间质增多；D. 相对于假手术组小鼠，HF 组小鼠心肌组织中 $CD19^+$ B 细胞（红色）浸润水平显著上升；*$P<0.05$

2.2 B细胞相关基因的鉴定和训练集(Meta-Microarray)的处理 利用从GEO数据库下载的GSE145154数据集提供的26个人类HF状态下心脏组织的单细胞测序数据。通过t-SNE降维分析(图2A),识别出多种细胞群,包括NK细胞、T细胞等,并确定了125个B细胞Marker基因(表2)。并用"sva"包对其他6个数据集进行了批次效应矫正,PCA分析显示,与矫正前(图2B)相比,矫正后批次效应显著降低(图2C)。最终将以上数据集整合成"Meta-Microarray"训练集。

表2 从心力衰竭患者心脏组织单细胞测序数据中确定125个B细胞Marker基因

基因名称
LAPTM5, CCL5, CTSC, CD74, IFITM3, IFI6, IFITM2, FYN, GSN, LEPROTL1, TIMP1, NKG7, HLA-DRA, CEBPB, ITGB2, GLUL, HLA-DMB, MEF2C, ITM2B, KLRD1, GNG7, CEBPD, VIM, ORAI2, HVCN1, S100A10, IRF8, MYL6, SRGN, CD63, SH3BGRL3, ITM2A, SWAP70, S100A11, PFN1, SELL, AAK1, TNFAIP3, MGST3, FCER2, BANK1, FCRLA, ID2, CD79A, HCST, ARL4C, CD3E, CD79B, CD24, IGFBP7, GNLY, CD3G, ANXA2, IL32, LCK, HLA-DOB, CD83, HERPUD1, VPREB3, BIRC3, CALM1, PRF1, CD9, RALGPS2, SPARCL1, CD55, BCL11A, TNFRSF13C, BST2, ACTB, GIMAP7, TCL1A, RHOH, MYL12A, TPM1, LY9, CTSW, CTSD, ATP5J, ANXA5, TYROBP, CD7, CCL4, CCR7, POU2F2, STOM, SNX2, S100A6, PTMS, CD37, HSPB1, LGALS1, GZMB, GZMA, USMG5, NDUFA4, CMC1, ZFP36, IFITM1, GIMAP4, GZMH, MS4A1, LGALS3, CD2, RHOC, UQCR11, FYB, LTB, EZR, MYL2, SYNE2, ANXA1, RHOB, CD247, CST7, ITGB1, RARRES3, GZMM, CD22, RGS1, ADAM28, DSTN, CST3, RPS11, CD3D

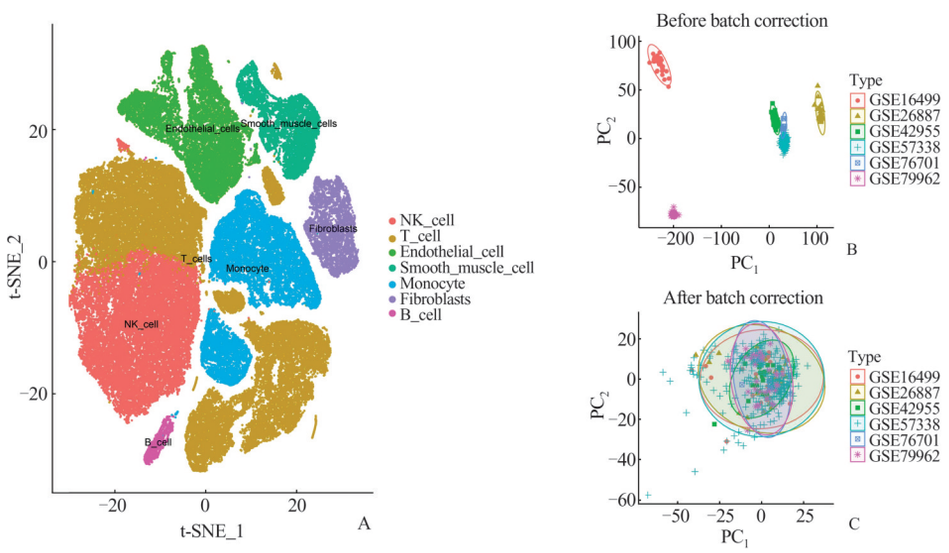


图2 B细胞相关基因的鉴定和训练集(Meta-Microarray)的处理

A.单细胞测序分析显示在人类心力衰竭心脏中不同的细胞亚群;B、C.主成分分析揭示矫正前(B)和矫正后(C)的不同GEO数据集的组间差异

2.3 筛选差异表达的B细胞相关基因 使用R软件的"limma"包分析上述得到的125个B细胞相关基因,以FDR<0.05筛选出86个心力衰竭相关的差异表达基因(图3A)。将这些基因上传STRING数据库,可信度设置为0.4,构建PPI网络(图3B),利用Cytoscape软件(3.8.0版)使其可视化,cytohubba插件分析基因的重要性,算法选择为"Degree",展示Degree最高的前50个基因(图3C)。接着将86个差异表达基因上传至Metascape数据库进行功能富集分析(图3D),发现这些基因主要与细胞活化和免疫反应正向调节相关。

2.4 进一步筛选对HF具有高诊断价值的基因 利用R软件的"rms"包对86个差异基因进行单因素逻辑回归分析,以$P<0.001$作为筛选标准选出与心力衰竭更相关的基因($n=58$)(图4A)。通过"Boruta"包进行BORUTA算法分析,将基因分为3类:"Confirmed"表示确定可以作为HF的诊断基因($n=49$, 56.98%);"Tentative"表示不能确定($n=6$, 6.98%);"Rejected"表示不可以作为HF的诊断基因($n=31$, 36.04%),用饼图表示最后的计算结果(图4B)。接下来把确定的49个基因利用"glmnet"R包进行LASSO回归分析,确定了19个基因具有高度诊断价值(图4C)。再利用"caret"R包通过支持向量机递归特征消除(SVM-RFE)的方

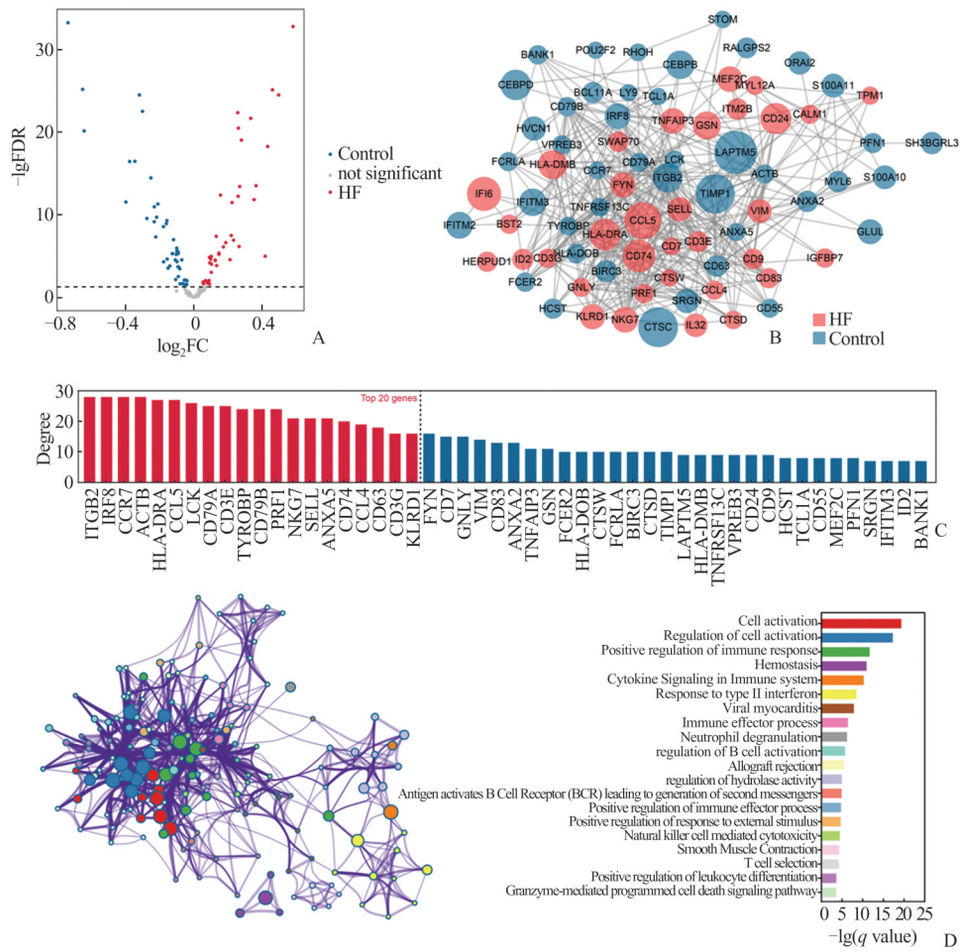


图 3　筛选差异表达的 B 细胞相关基因

A. 与心力衰竭相关的 86 个 B 细胞相关基因；B. 差异基因的 PPI 网络；C. 在 PPI 网络中 Degree 最高的前 50 个基因；D. 86 个差异表达基因的功能富集分析

法对差异基因做进一步特征筛选（图 4D），识别出 16 个基因具有较高的诊断效能。使用 Venn 图综合以上 4 种算法，并结合 PPI 网络 Degree 排名前 20 的基因，最终确定了 5 个基因具有高度诊断价值，分别是：趋化因子 C-C 配体 5（C-C Motif Chemokine Ligand 5，CCL5），CD74（CD74 Molecule，白细胞分化抗原 74），ITGB2（Integrin Subunit Beta 2，整合素 β_2），KLRD1（Killer Cell Lectin Like Receptor D1，杀伤细胞凝集素样受体 D1）和 SELL（Selectin L，L-选择素）（图 4E）。

2.5　CCL5、CD74、ITGB2、KLRD1 和 SELL 对 HF 的诊断效能　接下来对上述所得 5 个基因对 HF 的诊断效能进行了分析。分析结果显示，在训练集“Meta - Microarray”中，CCL5 的 AUC=0.818（95%CI：0.775～0.857，图 5A）；CD74 的 AUC=0.719（95%CI：0.666～0.765，图 5B）；ITGB2 的 AUC=0.777（95%CI：0.733～0.821，图 5C）；KLRD1 的 AUC=0.768（95%CI：0.717～0.812，图 5D）；SELL 的 AUC=0.674（95%CI：0.621～0.725，图 5E）。而在验证集“Meta-RNAseq”中，CCL5 的 AUC=0.780（95%CI：0.669～0.881，图 5F）；CD74 的 AUC=0.879（95%CI：0.795～0.944，图 5G）；ITGB2 的 AUC=0.607（95%CI：0.485～0.735，图 5H）；KLRD1 的 AUC=0.811（95%CI：0.702～0.902，图 5I）；SELL 的 AUC=0.730（95%CI：0.608～0.839，图 5J）。上述结果进一步验证了我们特征选择的可靠性。

2.6　构建和验证基于上述 5 个基因的诊断列线图　为进一步提高预测效能，基于以上 5 个基因的表达水平，我们利用多因素逻辑回归构建了诊断模型，并利用该模型计算了每个样本的风险评分。风险评分具体计算如下：风险评分＝−10.059＋4.028×EXP（CCL5）＋3.378×EXP（CD74）−7.441×EXP（ITGB2）＋1.303×EXP（KLRD1）

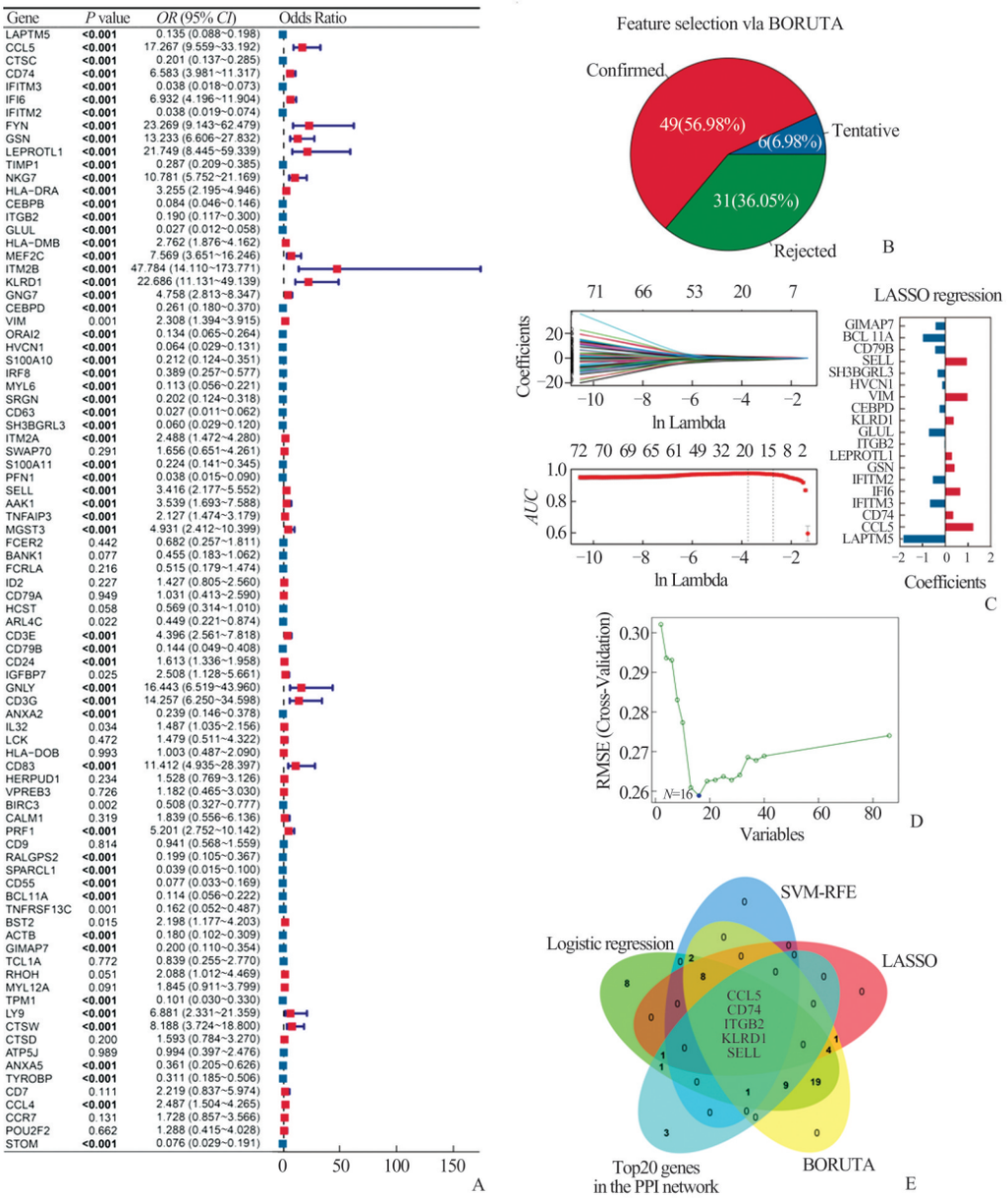


图4 基于机器学习算法筛选对HF有高度诊断价值的基因

A. 单因素逻辑回归分析;B. Boruta算法分析;C. LASSO回归分析;D. 基于支持向量机递归特征消除(SVM-RFE)筛选生物标志物;E. Venn图显示了以上四种算法以及PPI网络分析获得的基因的交集

+1.925×EXP(SELL)。从该公式各基因的回归系数分析除ITGB2可能对HF的发生具有抑制作用,其余基因的高表达可能促进HF的发生。为了使预测模型更加直观,利用R软件的"rms"包构建了HF的诊断列线图(图6A)并建立了Calibration校准图(图6B),发现实际曲线非常接近于理想曲线,表明校准度较好,即模型预测结果与实际结果的偏差较小,模型效果较好。我们使用R软件的"pROC"包分别在"Meta-Microarray"训练集和"Meta-RNAseq"验证集中绘制ROC曲线(图6C和图6D),训练集中*AUC*为0.951(95%*CI*:0.928~0.971),验证集中*AUC*为0.928(95%*CI*:0.845~0.983),表明特征性生物标志物具有较高的诊断能力。另外,我们利用R软件的"rmda"包绘制训练集(图6E)和验证集(图6F)的DCA决策曲线,结果显示在两组中该列线图获益均较高,数据较可靠。

2.7 比较诊断列线图与BNP对HF的诊断效能

BNP是目前应用最广泛的HF诊断标志物之一,接下来我们比较了本研究所构建诊断列线图和BNP对HF的诊断效能。德龙检验结果显示,无论是在

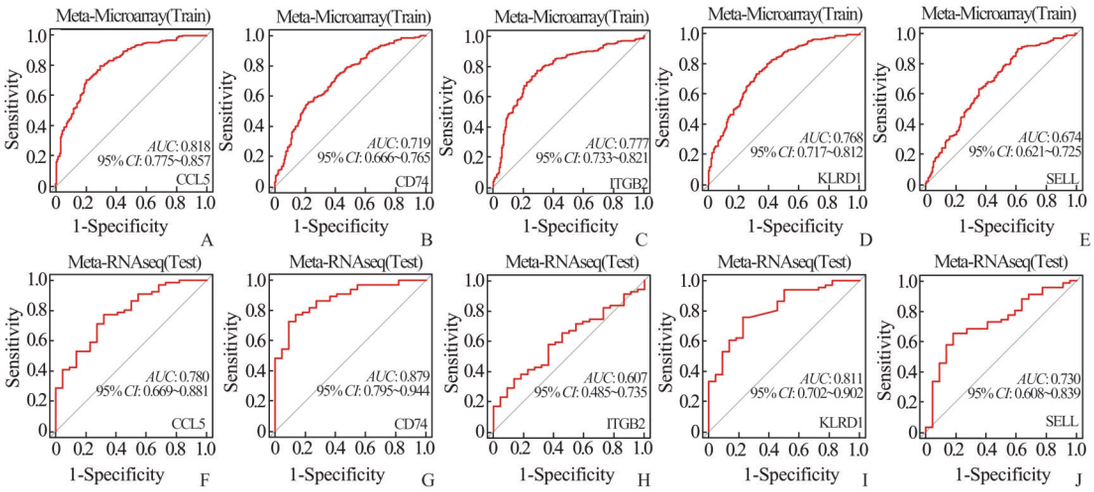


图 5　所筛选基因对 HF 的诊断效能

A～E. 在训练集中探究 CCL5(A)、CD74(B)、ITGB2(C)、KLRD1(D)和 SELL(E)对 HF 的诊断效能;F～J. 在验证集中探究 CCL5(F)、CD74(G)、ITGB2(H)、KLRD1(I)和 SELL(J)对 HF 的诊断效能

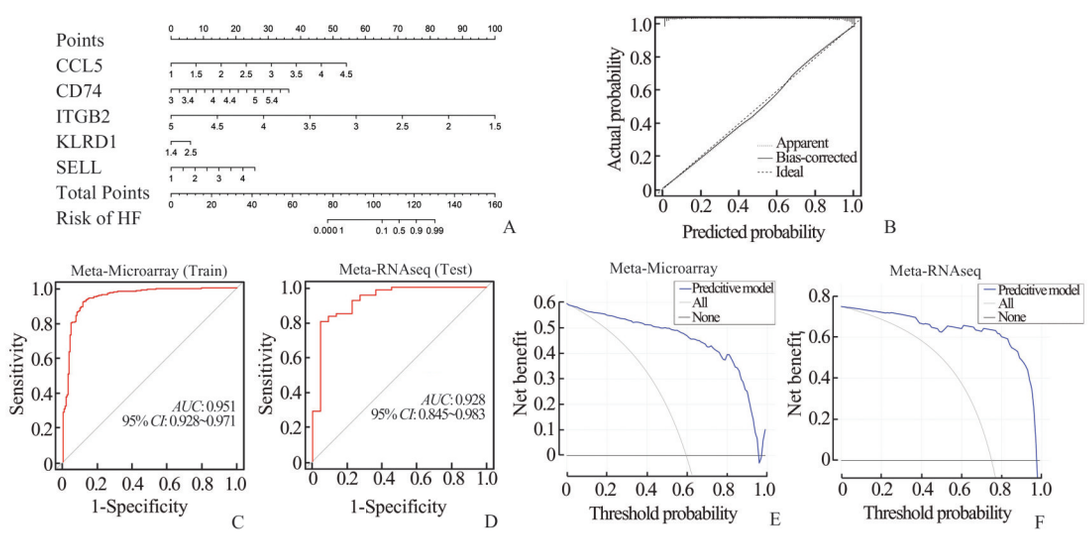


图 6　构建和验证基于上述 5 个基因的诊断列线图

A. 列线图用于预测 HF 的发生;B. Calibration 校准图;C. 训练集的 ROC 曲线;D. 验证集的 ROC 曲线;E. 训练集的 DCA 决策曲线;F. 验证集的 DCA 决策曲线

训练集还是验证集中,本研究所构建的诊断列线图诊断效能均优于 BNP($P<0.05$,表 3)。

2.8　诊断列线图对 HF 不同亚型的诊断效能　鉴于 HF 具有多种不同的亚型,接下来我们探究了诊断列线图对缺血性心肌病心力衰竭和扩张型心肌病心力衰竭的诊断效能。在“Meta-Microarray”数据集中,明确诊断为缺血性心肌病心力衰竭样本为 105 例,诊断为扩张型心肌病心力衰竭样本为 91 例,分别占到所有 HF 样本的 40.541% 和 35.135%。利用上述构建的列线图预测缺血性心肌病心力衰竭的发生概率,*AUC* 值为 0.954(95%*CI*:0.927～0.977)(图 7A);而利用诊断列线图预测扩张型心肌病心力衰竭的发生率结果也较为理想,*AUC* 值为 0.959(95%*CI*:0.928～0.983)(图 7B)。但必须要

表 3　在训练集和验证集中比较列线图和 BNP 对 HF 的诊断效能

队列	列线图		脑钠肽(BNP)		德龙检验	
	AUC	95%*CI*	*AUC*	95%*CI*	*Z*	*P*
Meta-Microarray	0.951	0.928～0.971	0.881	0.852～0.911	3.689	<0.001
Meta-RNAseq	0.928	0.845～0.983	0.780	0.656～0.883	2.089 6	0.037

说明的是,该诊断列线图并不能很好区分扩张型心肌病心力衰竭患者和缺血性心肌病心力衰竭患者,*AUC*为0.498(95%*CI*:0.416~0.578)(图7C)。

2.9 CCL5、CD74、ITGB2、KLRD1和SELL在B细胞中的潜在功能分析 使用单基因GSEA策略进一步分析CCL5、CD74、ITGB2、KLRD1和SELL在B细胞中的潜在功能。以CCL5为例,提取GSE145154数据集中HF状态下心脏组织的单细胞测序数据中的B细胞样本(n=914)。以CCL5表达水平中位值为截断值,将样本划分为高表达和低表达两组。从GOBP数据集收集B细胞相关的基因集,作为GSEA的参考数据集进行GSEA分析,选择与CCL5表达水平相关性最高的前10个B细胞相关功能进行展示,即NES绝对值最高的前10个通路。可以观察到CCL5表达水平与B细胞分化、激活、增殖等生物学功能显著正相关(图8A)。同样方式,依次分析了CD74、ITGB2、KLRD1和SELL与B细胞功能的相关性。分析结果显示,CCL5表达水平与B细胞分化最相关(NES=1.581,Nominal P<0.05,FDR<0.25,图8A);CD74表达水平与B细胞激活最相关(NES=1.926,Nominal P<0.001,FDR<0.001,图8B);ITGB2表达水平与B细胞激活最相关(NES=1.692,Nominal P<0.001,FDR<0.01,图8C);KLRD1表达水平与B细胞分化最相关(NES=1.534,Nominal P<0.01,FDR<0.25,图8D);而SELL与B细胞受体(BCR)信号通路激活最相关(NES=1.735,Nominal P<0.001,FDR<0.001,图8E)。总之,CCL5、CD74、ITGB2、KLRD1和SELL与多种B细胞功能显著相关,可能在B细胞中发挥关键功能。

2.10 $CD19^+$ $KLRD1^+$ B细胞在HF模型组小鼠心肌组织中高浸润 在确定的5个基因中,除KLRD1外,CCL5、CD74、ITGB2、SELL均已被报道过与心力衰竭的关系,并有可能作为诊断的潜在生物标志物[11-14]。因此我们选择KLRD1进行进一步实验验证。使用特异性KLRD1和CD19蛋白抗体,在假手术组和HF组小鼠心肌组织中进行免疫荧光染色分析。分析结果表明,假手术组小鼠心肌组织中$CD19^+$ $KLRD1^+$ B细胞浸润比例为33%±5%,而HF组小鼠的$CD19^+$ $KLRD1^+$ B细胞浸润比例为80%±5%。相对于假手术组小鼠,HF模型组小鼠$CD19^+$ $KLRD1^+$细胞浸润比例显著上升(P<0.05,图9)。

3 讨论

本研究通过构建心力衰竭小鼠模型,应用CD19免疫组织化学染色检测小鼠心脏中B细胞浸润水平,发现与假手术组相比,HF小鼠心脏B细胞浸润水平上升(P<0.05)。基于对数据集GSE145154中提供的26个人类HF状态下心脏组织的单细胞测序数据分析,确认了在HF心脏中有B淋巴细胞的存在,并利用R软件Securt包分析确定了125个B细胞相关基因。使用R软件的"limma"包分析后,以FDR<0.05作为筛选标准,筛选得到了86个与HF相关的差异表达基因。从GEO数据库中下载6个HF患者microarray测序数据集,并将其合并作为训练集(对照组=176,HF=259)。同时下载2个HF患者RNA-seq测序数据集,将其合并作为验证集(对照组=22,HF=66)。将86个差异基因上传到STRING数据库,构建了蛋白质-蛋白质相互作用(PPI)网络,利用Cytoscape软件对该网络进行进一步可视化,算法选择为"Degree",筛选出了Degree最高的前50个基因。同时将86个差异基因上传到Metascape数据库进行基因功能富集

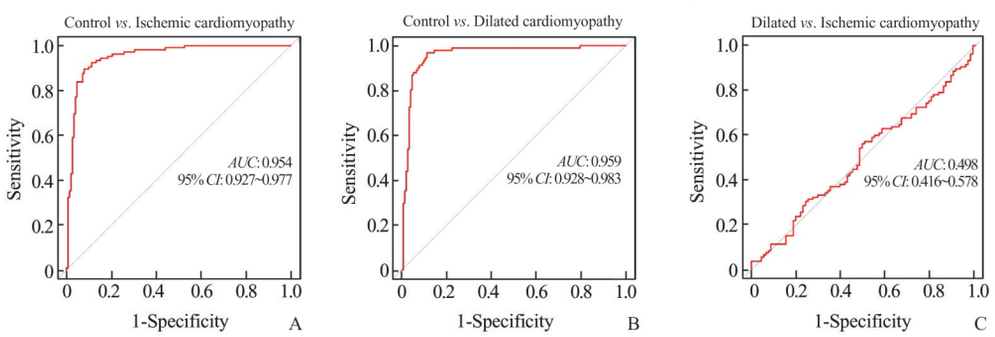


图7 诊断列线图对缺血性心肌病心力衰竭和扩张型心肌病心力衰竭的诊断效能

A.区分健康人群和缺血性心肌病心力衰竭患者诊断效能;B.区分健康人群和扩张型心肌病心力衰竭患者诊断效能;C.区分两型心肌病心力衰竭患者诊断效能

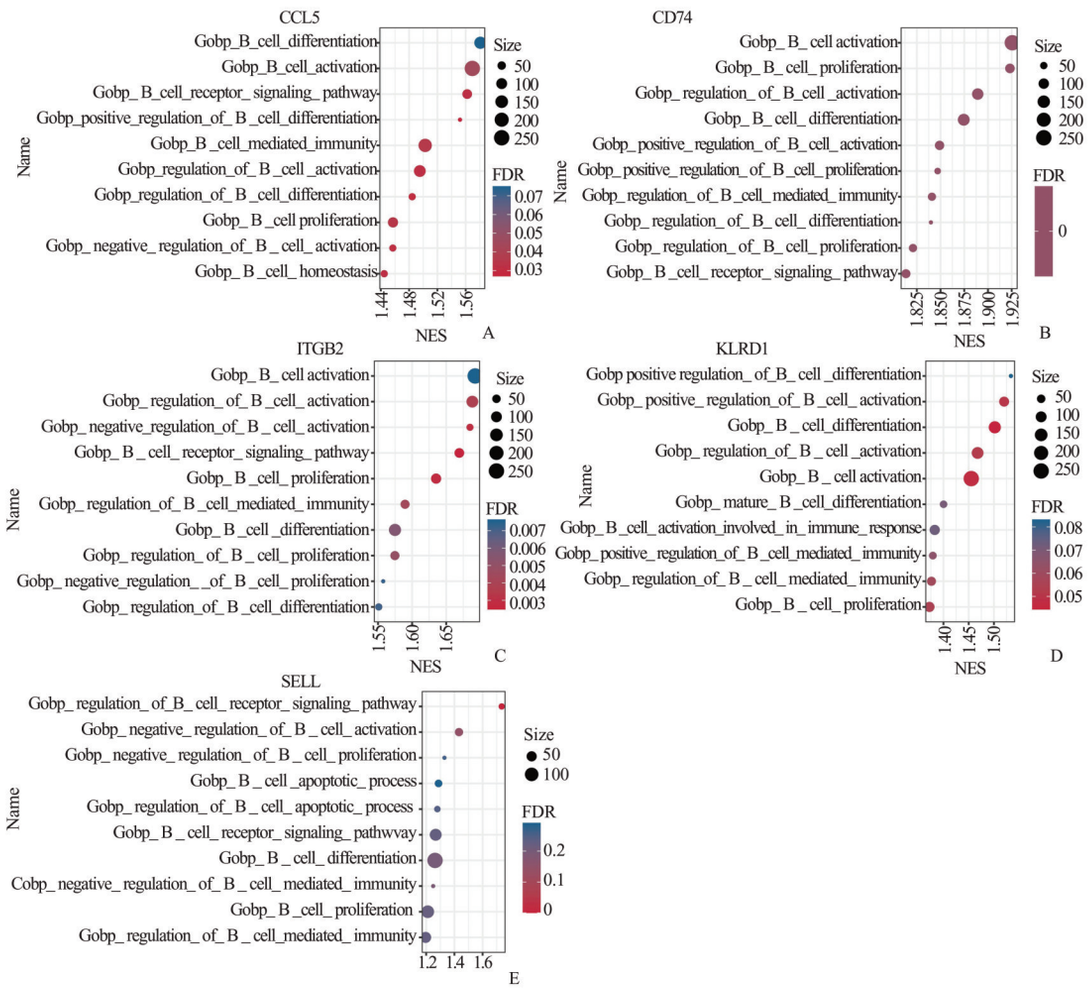


图 8　CCL5、CD74、ITGB2、KLRD1 和 SELL 表达水平与 B 细胞功能的相关性

A. CCL5 与 B 细胞分化最相关；B. CD74 与 B 细胞激活最相关；C. ITGB2 与 B 细胞激活最相关；D. KLRD1 与 B 细胞分化最相关；E. SELL 与 B 细胞受体信号通路（BCR）激活最相关

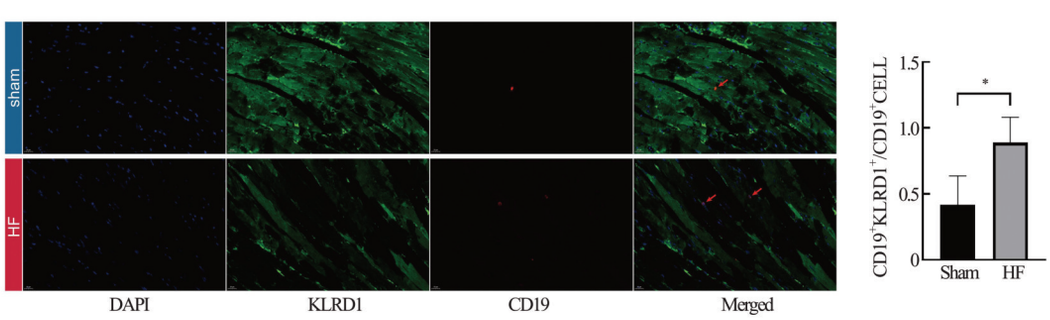


图 9　HF 组小鼠心肌组织中 $CD19^+$（红色）$KLRD1^+$（绿色）B 细胞浸润水平显著上升

$^*P<0.05$

分析，发现差异基因的功能大多数与 B 细胞活化以及免疫反应的正向调节等相关。使用单因素逻辑回归、BORUTA 算法、LASSO 回归以及 SVM-RFE 算法，结合 PPI 网络分析确定具有高度诊断价值的 B 细胞相关基因有 5 个分别为：CCL5、CD74、ITGB2、KLRD1 和 SELL。利用多因素逻辑回归构建诊断列线图，该列线图是一个可靠的 HF 诊断工具，并通过 ROC、Calibration 以及 DCA 分析明确其诊断价值。通过小鼠心肌组织 CD19 和 KLRD1 多重免疫荧光染色验证结论 KLRD1 与 B 细胞以及心力衰竭的相关性，结果显示，相对于对照组小鼠心肌组织，HF 组小鼠心肌组织具有更高水平的

$CD19^+$ $KLRD1^+$。

目前心力衰竭(HF)仍然是一种不断上升的全球流行病。基于基因表达数据库挖掘相关差异表达基因和信号通路的生物信息学方法对于寻找心力衰竭诊断和治疗的生物标志物至关重要。多种心脏疾病、遗传性缺陷和全身性疾病都可能导致心衰,但心衰的诊断性生物标志物和治疗靶点仍然有限[15]。现在临床上常用于诊断心力衰竭的生物标志物是BNP和NT-proBNP,但因其特异性不高,仍然需要寻找更精确的诊断性生物标志物或方法。我们发现在HF小鼠心脏中B细胞浸润水平上升($P<0.05$),提示在HF发生时会有免疫反应的产生。B淋巴细胞在正常组织中主要有三个功能:抗体产生、抗原加工或呈递以及细胞因子的产生或调节其他白细胞的功能。但B细胞在受损心肌中会被激活,活化的B细胞分泌细胞因子和趋化因子,会促进炎症产生并损害心脏功能,并可能直接参与导致不良心脏重塑的途径[16]。因此我们认为,在HF患者中,与B细胞相关的基因可能对HF诊断有帮助。

本研究通过多种算法对B细胞相关基因进行筛选,最终发现了5个与心力衰竭有关的差异基因,分别是CCL5、CD74、ITGB2、KLRD1和SELL。为了更好地支持临床决策,我们利用多因素逻辑回归构建诊断列线图,该列线图是一个可靠的HF诊断工具,并通过Calibration、ROC以及DCA分析明确其诊断价值,结果表明预测结果与实际观察结果之间具有良好的一致性。此外,我们还发现该诊断列线图对HF的两个重要亚型:缺血性心肌病心力衰竭和扩张型心肌病心力衰竭,具有更高的诊断能力。进一步分析表明,相比于传统HF诊断标志物BNP,该诊断列线图表现出更强的预测效能。综上所述,我们基于B细胞相关基因构建了一个5基因诊断列线图,为HF患者的诊断提供了一个潜在的临床工具。

因CCL5、CD74、ITGB2以及SELL均已被报道过与心力衰竭的关系。本研究着重研究KLRD1。KLRD1位于6号染色体上发现的NK复合体(NKC)中的Ly49基因簇和NKR-P1基因簇之间,编码NK细胞受体CD94,CD94与NKG2家族成员形成异二聚体,并识别靶细胞上的HLA-E。最近一项Meta分析[17]显示,KLRD1表达被确定为预测流感易感性的反向生物标志物,另外KLRD1也被报道与心肌梗死有关[18-20],提示KLRD1可能是心肌梗死进展和复发诊断和预后生物标志物。我们利用免疫荧光的方法探究KLRD1与心力衰竭及与B细胞的关系,结果发现相对于对照组小鼠心肌组织,HF组小鼠心肌组织具有更高水平的$CD19^+$ $KLRD1^+$ B细胞浸润($P<0.05$),提示KLRD1或能作为一个潜在的HF的诊断性分子生物标志物,同时也为从B细胞角度探究HF发病机制提供了可能的切入点。

尽管我们的研究结果具有潜在的临床意义,但仍存在一些局限性。首先从数据库中提取的信息是有限的,仍缺乏多中心大样本的临床数据。其次,本研究实验结果有限,仍需要更多的实验数据进一步证实。最后,需要进一步研究这个基因作为潜在治疗靶点的价值。

总之,我们的研究结果也很好地为之后的研究指明了方向,即未来可以在生物信息学的指导之下,利用多基因诊断列线图为心力衰竭的诊断提供临床工具,并可进一步设计相关实验研究KLRD1对心力衰竭的影响,我们还可以通过二代测序和心肌活检对心力衰竭患者心肌细胞内KLRD1的表达程度和预后水平进行比对,进而探索KLRD1和心力衰竭患者预后的关系。

参考文献

[1] POURAFKARI L, TAJLIL A, NADER N D. Biomarkers in diagnosing and treatment of acute heart failure [J]. Biomark Med, 2019, 13(14): 1235-1249.

[2] CONRAD N, JUDGE A, TRAN J, et al. Temporal trends and patterns in heart failure incidence: A population-based study of 4 million individuals [J]. Lancet, 2018, 391(10120): 572-580.

[3] CASTIGLIONE V, AIMO A, VERGARO G, et al. Biomarkers for the diagnosis and management of heart failure [J]. Heart Fail Rev, 2022, 27(2): 625-643.

[4] MUELLER C, MCDONALD K, DE BOER R A, et al. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations [J]. Eur J Heart Fail, 2019, 21(6): 715-731.

[5] TANG Q, CEN Z, LU J, et al. The abnormal distribution of peripheral B1 cells and transition B cells in patients with idiopathic dilated cardiomyopathy: A pilot study [J]. BMC Cardiovasc Disord, 2022, 22(1): 78.

[6] ADAMO L, ROCHA-RESENDE C, PRABHU S D, et al. Reappraising the role of inflammation in heart failure [J]. Nat Rev Cardiol, 2020, 17(5): 269-285.

(下转第1306页)