

血清蛋白质组学和机器学习揭示的青少年结核病新诊断标志物

陈禹^{1,3#}, 许鸿翔^{2#}, 田瑶¹, 何谦³, 赵晓云⁴, 张国斌⁵, 谢建平^{2*}

1 沈阳市第十人民医院(沈阳市胸科医院)结核病科, 辽宁 沈阳 110044

2 西南大学 生命科学学院现代生物医药研究所, 重庆 400715

3 中国医科大学 公共卫生学院, 辽宁 沈阳 110044

4 沈阳药科大学 生命科学与生物制药学院, 辽宁 沈阳 110016

5 沈阳市疾病预防控制中心, 辽宁 沈阳 110623

陈禹, 许鸿翔, 田瑶, 何谦, 赵晓云, 张国斌, 谢建平. 血清蛋白质组学和机器学习揭示的青少年结核病新诊断标志物[J]. 生物工程学报, 2025, 41(4): 1478-1489.

CHEN Yu, XU Hongxiang, TIAN Yao, HE Qian, ZHAO Xiaoyun, ZHANG Guobin, XIE Jianping. Serum proteomics and machine learning unveil new diagnostic biomarkers for tuberculosis in adolescents and young adults[J]. Chinese Journal of Biotechnology, 2025, 41(4): 1478-1489.

摘 要: 青少年(adolescents and young adults, AYA)是结核病的主要患病群体之一。为了研究针对这个群体的独特性诊断标志物, 本研究招募了 81 名 AYA 受试者, 通过定量蛋白质组学的方法, 绘制了 AYA 结核病患者高质量的血清蛋白质图谱。血清蛋白组数据表明, 活动性肺结核(active tuberculosis, ATB)患者中血红蛋白和载脂蛋白相对丰度显著降低。通路富集分析表明, ATB 组下调蛋白显著富集于抗氧化、细胞脱毒相关通路, 表明存在广泛的氧化应激损伤。通过随机森林(random forest, RF)和极致梯度提升(extreme gradient boosting, XGBoost)联合评估蛋白重要性, 获得了一组区分 ATB 和非 ATB 的候选蛋白标志物。基于特征递归消除的支持向量机算法, 发现载脂蛋白 A1 (apolipoprotein A-I, APOA1)、血红蛋白亚基 α (hemoglobin subunit alpha-1, HBA1)、血红蛋白亚基 β (hemoglobin subunit beta, HBB)这 3 个蛋白的组合在诊断 ATB 过程中具有最高的准确性和敏感性。临床生化常规检测的血红蛋白(hemoglobin, HGB)和白蛋白(albumin, ALB)含量可以作为评估 APOA1、HBB 蛋白表达变化的血液生化指标。本研究建立了 AYA 结核病患者的血清蛋白组全貌, 获得了该类群结核病的新诊断标志物。

关键词: 结核病; 血清蛋白组学; 机器学习; 诊断标志物; 载脂蛋白

资助项目: 2021 沈阳市科学技术计划(21-173-9-72); 2022 年辽宁省科学技术计划面上项目(2022-MS-432)

This work was supported by the 2021 Shenyang Science and Technology Plan (21-173-9-72) and the 2022 Liaoning Provincial Science and Technology Plan General Project (2022-MS-432).

#These authors contributed equally to this work.

*Corresponding author. E-mail: georgex@swu.edu.cn

Received: 2024-05-27; Accepted: 2024-07-03; Published online: 2024-07-04

Serum proteomics and machine learning unveil new diagnostic biomarkers for tuberculosis in adolescents and young adults

CHEN Yu^{1,3#}, XU Hongxiang^{2#}, TIAN Yao¹, HE Qian³, ZHAO Xiaoyun⁴, ZHANG Guobin⁵, XIE Jianping^{2*}

1 Department of Tuberculosis, China Shenyang the Tenth People's Hospital (China Shenyang Chest Hospital), Shenyang 110044, Liaoning, China

2 Institute of Modern Biomedical Research, School of Life Sciences, Southwest University, Chongqing 400715, China

3 School of Public Health, China Medical University, Shenyang 110044, Liaoning, China

4 School of Life Science and Biopharmaceutics, Shenyang Pharmaceutical University, Shenyang 110016, Liaoning, China

5 Shenyang Disease Control Center, Shenyang 110623, Liaoning, China

Abstract: Adolescents and young adults (AYAs) are one of the major populations susceptible to tuberculosis. However, little is known about the unique characteristics and diagnostic biomarkers of tuberculosis in this population. In this study, 81 AYAs were recruited, and the high-quality serum proteome of the AYAs with tuberculosis was profiled by quantitative proteomics. The data of serum proteomics indicated that the relative abundance of hemoglobin and apolipoprotein was significantly reduced in the patients with active tuberculosis (ATB). The pathway enrichment analysis showed that the downregulated proteins in the ATB group were mainly involved in the antioxidant and cell detoxification pathways, indicating extensive oxidative stress damage. Random forest (RF) and extreme gradient boosting (XGBoost) were employed to evaluate protein importance, which yielded a set of candidate proteins that can distinguish between ATB and non-ATB. The analysis with the support vector machine algorithm (recursive feature elimination) suggested that the combination of apolipoprotein A-I (APOA1), hemoglobin subunit beta (HBB), and hemoglobin subunit alpha-1 (HBA1) had the highest accuracy and sensitivity in diagnosing ATB. Meanwhile, the levels of hemoglobin (HGB) and albumin (ALB) can be used as blood biochemical indicators to evaluate changes in the protein levels of APOA1 and HBB. This study established the serum proteome landscape of AYAs with tuberculosis and identified new biomarkers for the diagnosis of tuberculosis in this population.

Keywords: tuberculosis; serum proteomics; machine learning; diagnostic biomarkers; apolipoprotein

结核分枝杆菌(*Mycobacterium tuberculosis*, Mtb)感染导致的结核病是全球公共卫生重大威胁, 每年新发结核病患者超过 1 000 万^[1]。在感染 Mtb 的人群中, 约 5%–15% 最终会发展成活动性肺结核(active tuberculosis, ATB)^[1]。结核病监测通常将人群简单分为“儿童”(0–14 岁)和“成人”(≥15 岁), 而忽视了青少年(adolescents and young

adults, AYA)这一特殊群体(定义为 10–24 岁)^[2]。最近一项研究表明, 全球每年新增约 180 万 AYA 结核病患者, 占新发结核病总患者总数的 17%^[3]。自 1990 年以来, AYA 结核病发病率有所下降, 但其中耐药结核病的发病率却在增加^[4]。尽管 AYA 的结核病风险正处于上升时期, 但 AYA 在结核病控制工作中尚未被视为独特人群^[2]。然

而, AYA 群体社会接触较为密集, 这意味着该群体的结核病高效诊断对于结核病防控至关重要^[5]。Mtb 培养是结核病诊断的金标准, 但耗时较长, 需要 4-8 周^[6]。 γ -干扰素释放试验(interferon- γ release assay, IGRA)可用于辅助诊断 Mtb 感染, 但不能区分活动性结核病(active tuberculosis, ATB)和潜伏性结核感染(latent tuberculosis infection, LTBI)^[7]。传统的诊断方法复杂、耗时, 不利于结核病的快速诊断。因此, 寻找结核病的快速诊断标志物, 建立新的结核病诊断模型尤为迫切。

关于结核病诊断新方法的研究非常多。粪便定量 PCR 可以提高成人和青少年结核病的检出率^[8]; 另外由于外周血易于收集, 近年来筛选和识别血浆或者血清中的标志物越来越受到研究人员重视^[9-10]。结核病患者外周血细胞的基因转录水平变化, 特别是特定基因转录水平的变化可用于结核病诊断^[11-13]。然而, 由于转录后修饰、翻译和翻译后修饰等机制的影响, 基因转录水平的变化并不总能真实反映最终人体蛋白质水平的变化。Mtb 感染与特定蛋白质的表达水平、蛋白质-蛋白质或蛋白质-核酸的相互作用有关^[3]。目前蛋白质组学技术已经广泛用于癌症诊断标志物的研发^[14]。例如, 部分血清蛋白被发现与特定疾病高度相关^[15]; 而经过筛选的血清蛋白集合能够提前多年预测癌症风险^[16]。对于结核病患者, 特定血清蛋白同样可作为潜在的诊断标志物^[17]。然而, 人群遗传背景、环境、Mtb 流行菌株的基因组差异等均可能导致血清蛋白组的差异^[17]。

本研究使用无标记定量蛋白组学技术, 建立了我国 AYA 结核病患者血清蛋白质组图谱, 并系统比较了其同龄 LTBI 或健康对照(healthy control, HC)的血清蛋白质组特征; 还采用多种机器学习算法, 挖掘并鉴定了我国 AYA

结核病诊断的新型标志物, 以期未来结核病高效诊断提供新方法, 并为 AYA 结核病的深入研究奠定基础。

1 材料与方法

1.1 样本采集

使用人类蛋白质组组织(Human Proteome Organization, HUPO)建议的标准操作程序处理样本。对于每个参与者, 空腹采集全血 2 mL。对于 ATB 患者, 在治疗前收集血样。全血于室温放置 2 h 后, 4 °C、3 000×g 离心 20 min, 并收集上清。加入乙二胺四乙酸(ethylene diamine tetraacetic acid, EDTA)后, 将血清样品分配到 1 mL 低蛋白吸附管中, 并储存在-80 °C超低温冰箱中备用。本实验由沈阳市第十人民医院医学伦理委员会审查批准(批准号: KYXM-2021-003-1)。

1.2 血清蛋白组样品制备

通过加入预冷的 6 倍体积丙酮沉淀血清样品中的总蛋白, 并在-20 °C下孵育过夜。离心后, 将干燥的蛋白质沉淀重悬于 0.1 mol/L 四乙基氯化铵(tetraethylammonium bromide, TEAB)缓冲溶液中。通过二辛可酸测定法测量蛋白质浓度。通过 SDS-PAGE 和银染法确认蛋白质完整性。使用 5 mmol/L 二硫苏糖醇(dithiothreitol, DTT)在 32 °C下进行 30 min 的二硫键还原, 然后在室温下用 10 mmol/L 碘乙酸酯在黑暗中烷基化 45 min。然后, 将血清蛋白质用 1.5 μ g 胰蛋白酶在 37 °C下消化过夜。消化的肽段随后用 10%三氟乙酸酸化并使用 C18 离心柱进行脱盐。干燥脱盐后的肽段用缓冲液(1%甲酸和 1%乙腈)溶解。

1.3 蛋白组质谱法定量

液相色谱分离后, 在 Orbitrap Fusion Lumos 质谱仪(ThermoFisher Scientific)上使用数据非依赖采集(data independent acquisition, DIA)模式分析肽混合物。将样品加载到毛细管柱, 并

用非线性梯度(600 nL/min)洗脱 150 min。肽在高压(2.2 kV)下电离。质谱仪调查扫描的分辨率为 120 000, 目标值设定为 400 000 个离子, 扫描的质量范围为 350–1 500 *m/z*。使用 Spectronaut 软件(v15.7)对 Uniprot 数据库中的人类蛋白序列文件进行搜索, 使用默认参数, 设置错误发现率(false discovery rate, FDR)<0.01。

1.4 蛋白通路富集分析

使用 limma 包^[18], 以差值倍数>1.5 或<0.67 和错误发现率 FDR<0.05 筛选显著差异表达的蛋白。使用 ClusterProfiler 包进行通路富集分析^[19], 以 FDR<0.05 筛选显著富集的通路。使用 STRINGdb 包对差异蛋白构建蛋白互作网络^[20]。

1.5 机器学习鉴定蛋白诊断标志物

对于数据预处理, 采用分位数法对蛋白质组数据集进行归一化, 并去除缺失率超过 20% 的蛋白质。使用均值填补法填补缺失值。利用 mlr3verse 包(<https://github.com/mlr-org/mlr3verse>), 通过 XGBoost 和 RF 机器学习算法筛选区分组间差异的重要血清蛋白。对于极致梯度提升(extreme gradient boosting, XGBoost)使用默认参数提取重要变量; 对于随机森林(random forest, RF), 保留 Mean Decrease Gini>0.5 的变量。将 2 种机器学习算法筛选得到的变量取并集后作为候选蛋白。使用基于特征递归消除的支持向量机(support vector machine feature based recursive feature elimination, SVM-RFE)算法, 并进行 5 折交叉验证, 以获取诊断效果最优的蛋白组合。此外, 使用 2 个外部队列的结核病血清蛋白组数据验证 3 个蛋白的表达变化, 并构建受试者工作曲线(receiver operating characteristic, ROC)^[10,17]。

1.6 单细胞转录组中 APOA1 表达量分析

从 NCBI 下载结核病患者外周血单细胞转录组数据集^[21]。利用 Seurat 包整合并标准化处理基因表达数据^[22]。使用 Celltypist 包自动化注释

细胞亚群^[23], 并进行人工校正。

1.7 统计分析及可视化

所有统计分析及可视化过程均在 R studio 中使用 R 语言(v4.3.1)进行。使用 Benjamini-Hochberg 法校正 *P* 值。符合参数模型的数据以平均值±标准差(mean±SD)表示, 符合非参数模型的数据以中位数(范围)表示。连续变量采用 *t* 检验或 Mann-Whitney 检验。

2 结果与分析

2.1 人口统计学及血液生化指标

共收集了 81 例受试者的血清蛋白组数据, 包括 30 例 HC、21 例 LTBI 和 30 例 ATB。对于 LTBI 和 ATB 组, 获取了人口统计学信息、血液指标以及免疫细胞亚群的比例(表 1)。所有样本年龄在 13–24 岁之间。LTBI 组中男性占比 61.9%; ATB 组中男性占比 30%。绝大多数血液生化指标在 LTBI 组和 ATB 组中无显著差异, 但 HGB 和 ALB 的浓度在 ATB 组中相较于 LTBI 组显著降低(表 1 中加粗的部分, *P*<0.05)。

2.2 结核病患者血清蛋白组学特征

血清蛋白质组共鉴定并定量了 1 252 种蛋白质(图 1A)。主成分分析结果表明, LTBI 组与 HC 组具有较为相似的血清蛋白质组特征, 而 ATB 组的血清蛋白质组特征与两者差异较大(图 1B)。差异分析表明, 与 HC 组相比, ATB 组有 35 个显著上调表达的蛋白质和 63 个显著下调表达的蛋白质, 其中 A2J1M4 (rheumatoid factor RF-ET7)、H0Y300 (haptoglobin)、CRP 等蛋白质上调表达倍数最高, 而 ECI1、HBG2、F5 等蛋白质下调幅度最多; 与 LTBI 组相比, ATB 组有 33 个显著上调的蛋白质和 57 个显著下调的蛋白质。其中 CRP、KCRM、PSPB 等蛋白质上调表达倍数最高, 而 HBG2、F5、GALA 等蛋白质下调最多(图 1C)。

表 1 受试者人口统计学及血液生化指标

Table 1 Demographics and blood biochemical indicators of participants

Item	LTBI	ATB	FDR
Participants (n)	21	30	—
Age [Median, (min–max)]	18, 15–23	18, 13–24	NS
Male (n, %)	13 (61.90%)	9 (30.00%)	0.048
Blood sugar (mmol/L)	5.18±0.52	5.20±0.58	NS
White blood cells (×10 ⁹ /L)	9.25±12.59	7.17±2.23	NS
Neutrophils (%)	60.19±11.24	65.24±11.28	NS
Lymphocytes (×10 ⁶ /L)	2 057.86±508.14	1 894.53±700.35	NS
Lymphocytes (%)	26.67±9.19	23.32±9.90	NS
Hemoglobin (g/L)	139.48±19.01	121.87±15.52	0.026
Platelets (×10 ⁹ /L)	278.00±79.32	303.00±66.36	NS
Glutathione aminotransferase (U/L)	18.19±3.91	19.10±7.29	NS
Alanine aminotransferase (U/L)	15.10±9.49	15.33±9.33	NS
Albumin (g/L)	44.32±3.95	40.79±4.28	0.043
Total bilirubin (μmol/L)	9.09±5.20	7.82±3.19	NS
Direct bilirubin (μmol/L)	3.98±2.35	3.69±2.05	NS
T cells, CD3 ⁺ /CD45 ⁺ (×10 ⁶ /L)	1 413.62±434.95	1 345.33±520.09	NS
T cells (%)	68.14±7.8	70.93±8.44	NS
Helper T cells (×10 ⁶ /L)	777.33±211.75	710.17±347.77	NS
Helper T cells (%)	37.76±5.69	37.33±9.07	NS
Inhibitory T cells (×10 ⁶ /L)	553.57±189.61	557.17±225.85	NS
Inhibitory T cells (%)	26.76±5.23	29.47±7.14	NS
B cells (×10 ⁶ /L)	338.81±176.57	322.37±243.36	NS
B cells (%)	16.57±7.59	16.40±7.48	NS
Natural killer cells (×10 ⁶ /L)	251.05±157.93	151.23±72.47	NS
Natural killer cells (%)	12.29±7.32	8.87±5.07	NS

NS: No significant difference.

2.3 差异蛋白质互作网络及差异通路分析

基于蛋白差异分析得到的显著差异蛋白，构建了基于 STRING 数据库的蛋白互作网络(图 2)。结果显示，在 ATB 组相较于 HC 组或 LTBI 组显著下调的蛋白质中，均包含大量载脂蛋白(APOA1、APOA2、APOA4、APOC3)或血红蛋白家族(HBB、HBD、HBG1、AHSP)相关蛋白，且具有较高的互作得分。GO 富集分析表明，与 HC 组或 LTBI 组相比，在生物过程(biological process, BP)(图 3A)和分子功能(molecular function, MF)方面

(图 3B)，ATB 组的抗氧化、解毒、脂质代谢相关通路显著下调，而内肽酶活性的抑制通路显著上调；在细胞成分(cellular component, CC)方面(图 3C)，ATB 组表现为胞外基质、血红蛋白复合体等显著下调，而囊泡腔相关成分显著上调。

2.4 机器学习筛选 TB 血清蛋白质诊断标志物

基于 XGBoost 和 RF 评估了不同血清蛋白质区分 ATB 组和非 ATB 组的重要性，最终保留了 APOA1、HBA1、A0A193CHR5 (10E8 light chain variable region)、A0A024QZT9 [NAD(P)H

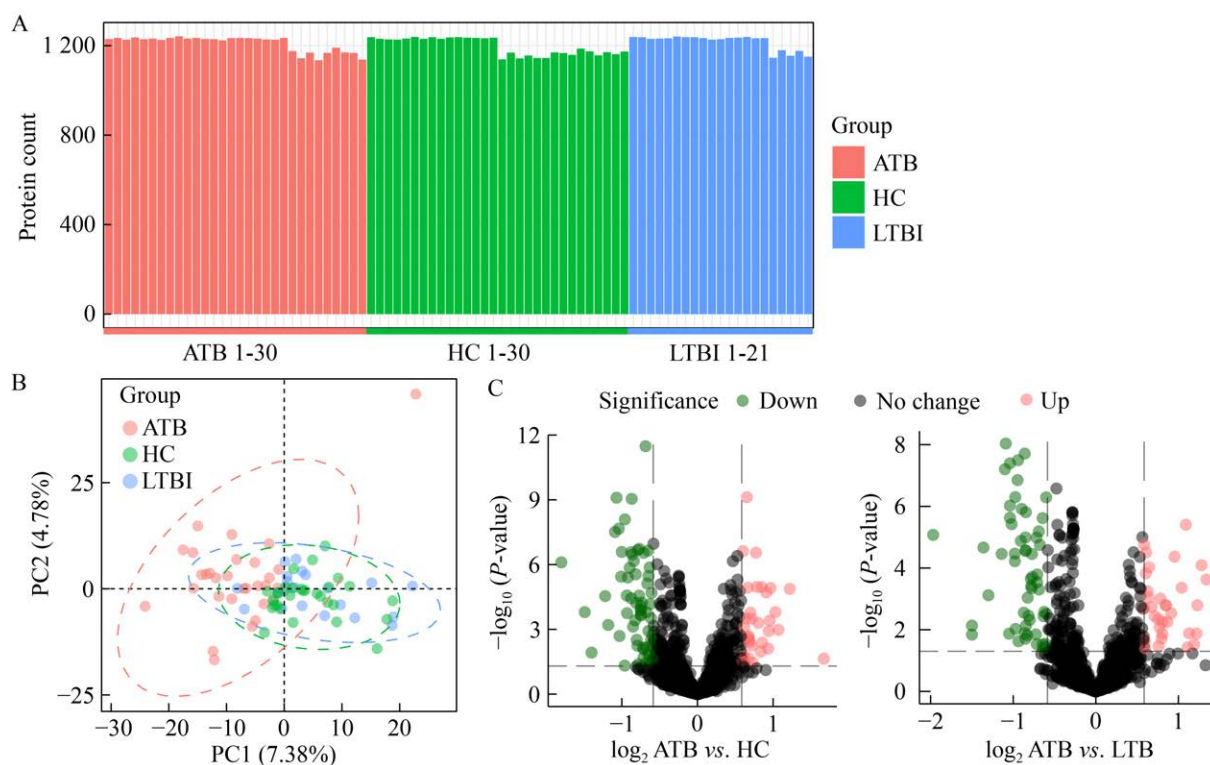


图1 ATB组与LTBI组和HC组具有不同的血清蛋白质组学特征 A: 每组样本鉴定到的蛋白质数量; B: 血清蛋白质组学主成分分析; C: 差异蛋白质火山图。

Figure 1 Serum proteomic characteristics of ATB were different from LTBI or HC. A: The number of proteins identified in each group; B: Principal component analysis of serum proteome; C: Volcano plot of differential protein.

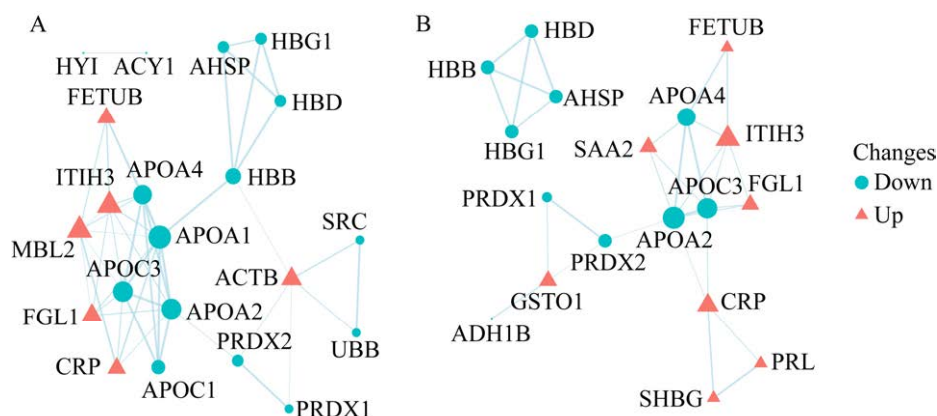


图2 差异蛋白质互作网络 A: ATB组相较于HC组的显著差异蛋白质; B: ATB组相较于LTBI组的显著差异蛋白质。节点大小由每个节点与其他节点的连接数量决定。

Figure 2 Differential protein interactions network. A: Significant difference in proteins between the ATB and the HC group; B: Significant difference in proteins between the ATB group and the LTBI group. The size of node is determined by the number of connections between the node and others.

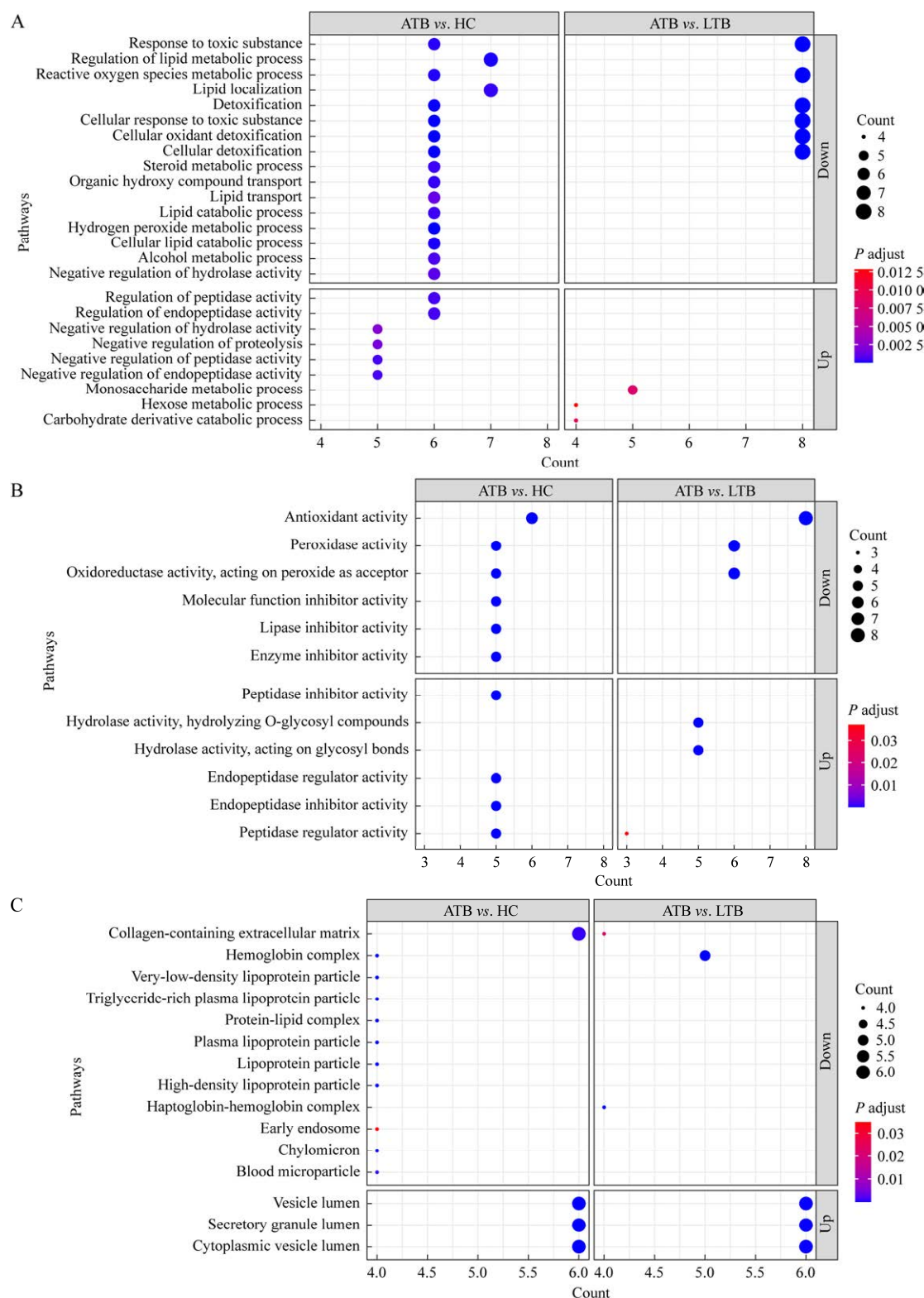


图3 差异蛋白质 GO 富集分析 基于生物过程(A)、分子功能(B)和细胞成分(C)的显著差异通路。

Figure 3 Differential protein GO enrichment analysis. Significant enriched pathways based on biological processes (A), molecular functions (B), and cellular components (C).

dehydrogenase]、HBB、NCAM1、MUC18、CRAC1 等 8 个候选蛋白质用于对 ATB 诊断标志物的筛选。通过 SVM-RFE 方法计算得出,使用 APOA1、HBB、HBA1 的蛋白质组合具有最高的曲线下面积(area under curve, AUC) (95.42%) (图 4A)。此外差异分析表明,与 HC 或 LTBI 相比,ATB 血清蛋白质组中 APOA1、HBB 和 HBA1 的蛋白质丰度均显著降低(图 4B)。对于单个蛋白质, APOA1 诊断的特异性为 0.833, 敏感度为 0.902; HBB 的特异性为 0.767, 敏感度为 0.902; HBA1 的特异性为 0.833, 敏感度为 0.667 (图 4C)。Spearman 相关性分析表明, APOA1、HBB 的蛋白质丰度与 ALB 和 HGB 显著正相关(图 4D)。外部数据也表明,这 3 个蛋白质的相对丰度在结核病患者中明显下调(图 4E, 3 例 ATB、3 例 LTBI、3 例 HC); 而利用外部队列的 APOA1 和血红蛋白表达量预测结核病患者结果表明,这种蛋白质组合一定程度上能预测 AYA-TB, 具有相对较好的特异性(0.746)和敏感度(0.778) (图 4F, 37 例 ATB、104 例 HC)。然而,在外周血单个核细胞(peripheral blood mononuclear cell, PBMC)的单细胞转录组中,无论是来自结核病患者还是健康对照的样本,均只有不超过 3%的细胞表达 APOA1 基因,表明 PBMC 并非 APOA1 蛋白的主要来源(图 5)。

3 讨论与结论

AYA 群体处于结核病患病概率的上升期^[2],而常规的结核病诊断方法无法满足 AYA 结核病快速诊断的需求,因此,亟须开发新的结核病快速诊断方法。本研究的结果显示,在我国 AYA 人群中,ATB 组与 LTBI 或 HC 组的血清蛋白组明显不同,主要表现为 ATB 中抗氧化和解毒能力的下降。这可能是由于宿主在抵抗结核分枝杆菌过程中会产生大量活性氧,而当其含量超

过宿主的抗氧化能力时可能导致氧化应激,并可以通过组织损伤和细胞死亡来放大该效应^[24]。本研究中的 CRP 是 ATB 组中表达量上调最多的蛋白之一。CRP 即 C 反应蛋白,是炎症和组织损伤的非特异性标志物,在结核病患者中通常会增加^[25]。在南非一项青少年结核病研究中,CRP 被认为可作为结核病诊断的蛋白之一,但其在部分队列中敏感性较低^[17]。此外还发现,基于 APOA1、HBB、HBA1 这 3 个蛋白在血清中的相对表达量即能够实现 ATB 的精准诊断。相较于 HC 或 LTBI 组,ATB 组中这 3 个蛋白的表达量均显著下调,且这一结果在外部队列数据中得到了验证。APOA1 是高密度脂蛋白的主要载体蛋白,能够与其他脂质和胆固醇结合,形成高密度脂蛋白颗粒。APOA 蛋白的下调可能会削弱宿主对 Mtb 的免疫反应;而利福平或异烟肼治疗后, APOA1 的表达量显著上升^[26]。此外,蛋白互作网络也显示大量载脂蛋白共同下调,这可能表明 ATB 组脂质转运能力的广泛受损。载脂蛋白被认为与结核病易感性有关,例如敲除载脂蛋白 E 基因的小鼠对结核分枝杆菌的敏感性增加,感染后细菌数量更多、肉芽肿增大^[26]。先前也有研究表明, APOA1 是结核病诊断的潜在标志物^[10],这提示本研究结果可能可以拓展到非 AYA 结核病患者。然而, APOA1 蛋白在血清中的表达量与其在 PBMC 中的转录水平不一致。在 PBMC 的单细胞转录组中, ATB 中 APOA1 的转录水平相较于 LTBI 与 HC 无显著差异。APOA1 主要在肝脏和小肠合成,血液中 APOA1 转录水平较低,提示 APOA1 可能无法作为 ATB 的血液转录标志物,但可作为蛋白质标志物来诊断结核病。

活动性肺结核患者的血红蛋白水平通常会降低^[27]。本研究中, HGB 和 ALB 在 ATB 组显著下调,这可能是结核分枝杆菌利用血红素作

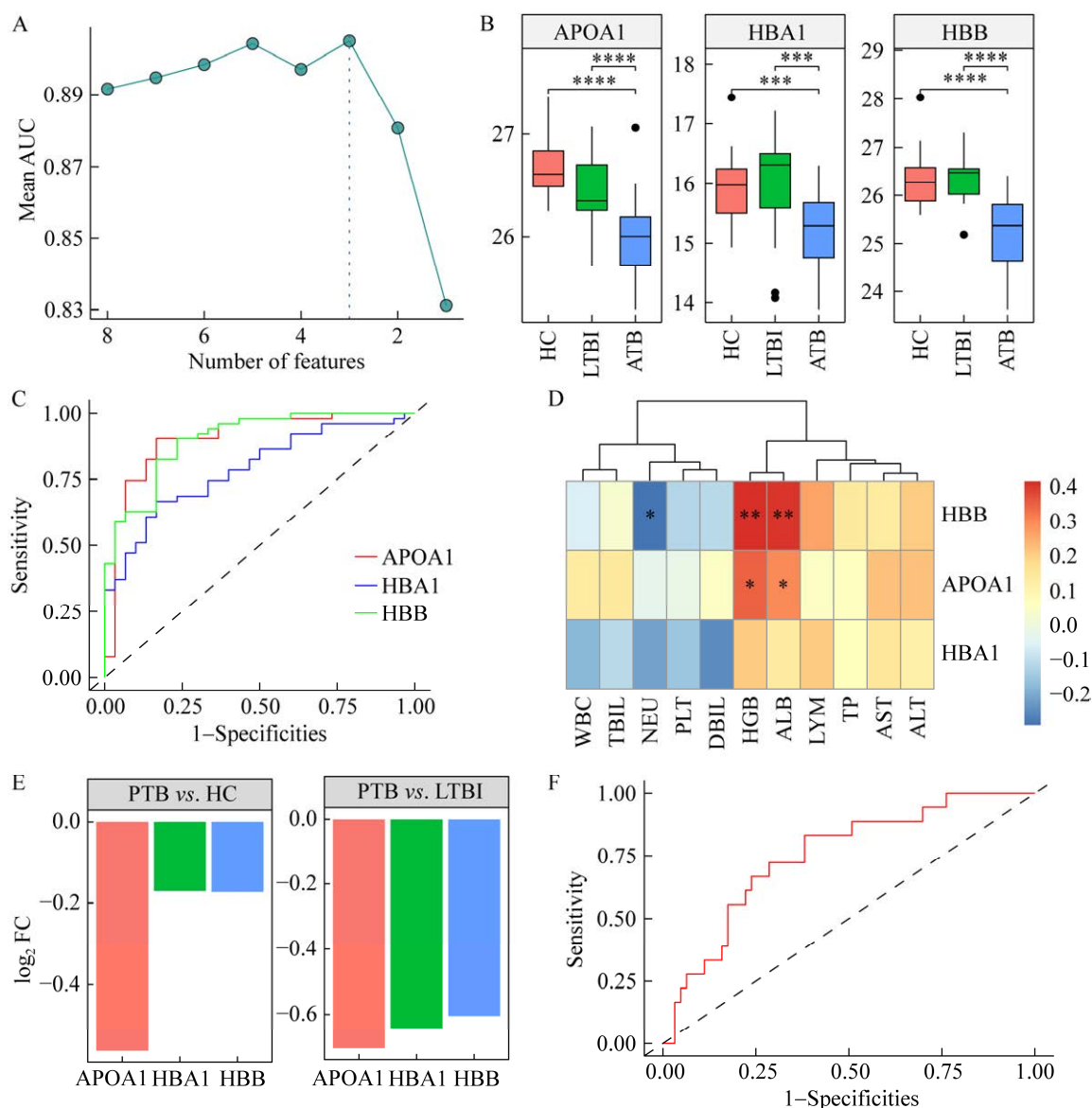


图 4 机器学习鉴定结核病血清蛋白质诊断标志物 A: 基于 SVM-RFE 算法鉴定结核病血清蛋白质诊断标志物最优蛋白质组合。B: 选定蛋白质的相对丰度在 ATB 组中显著下调。***: $P < 0.001$; ****: $P < 0.0001$ 。C: 基于 3 个蛋白质预测 AYA-ATB 的受试者工作曲线。D: 选定蛋白质与临床血液指标的 Spearman 相关性分析。*: $P < 0.05$; **: $P < 0.01$ 。E: 外部数据证实 3 个血清蛋白质在结核病患者的血清蛋白质中下调表达。F: 基于 3 个蛋白质预测的外部结核病患者队列的受试者工作曲线。

Figure 4 Machine learning identification of tuberculosis serum protein diagnostic biomarkers. A: The optimal combination of serum protein diagnostic markers for tuberculosis was identified based on SVM-RFE algorithm. B: The relative abundance of selected proteins was significantly downregulated in the ATB group. ***: $P < 0.001$; ****: $P < 0.0001$. C: ROC curve analysis for predicting AYA-ATB based on three proteins. D: Spearman correlation analysis between selected proteins and clinical blood indicators. *: $P < 0.05$; **: $P < 0.01$. E: External data confirmed that the selected protein was down regulated in serum protein of tuberculosis patients. F: ROC curve analysis for predicting tuberculosis in an external cohort of tuberculosis patients based on the three proteins.

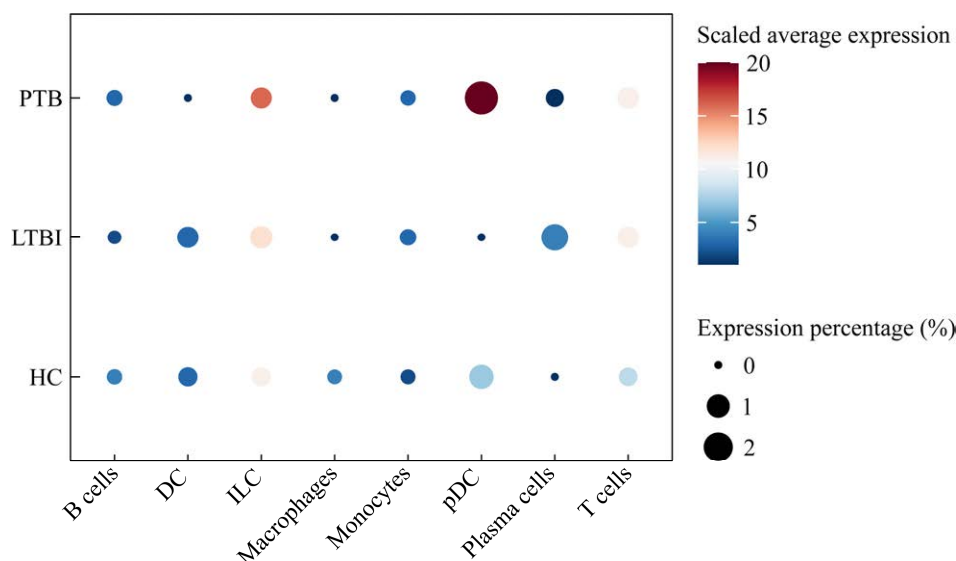


图5 结核病患者 PBMC 中的单细胞转录组中 APOA1 表达量极低

Figure 5 The mRNA level of APOA1 in the single cell transcriptome of PBMC from tuberculosis patients is extremely low.

为铁的来源以满足自身生长需求^[28]。同时,促炎因子如 IL-6、TNF- α 和 IFN- γ 的过度释放也会导致促红细胞生成素形成减少、骨髓反应受损和铁代谢改变,进而导致缺铁性贫血^[29-30]。两个诊断标志物 HBB 和 HBA1 均属于血红蛋白亚基,且 HBB 的蛋白相对丰度与临床指标 HGB 和 ALB 显著相关,提示 ATB 组血液运输氧能力的下降。因此,临床上可以考虑在结核病诊断过程中将这 2 个血液生化指标纳入诊断依据中,并与血清蛋白诊断标志物联用,以提高结核病诊断的敏感性和精度。需要指出的是,本研究中蛋白组学为相对丰度。因此,未来需要利用质谱法对血浆中 3 个蛋白的特异靶标肽段进行绝对定量,并确定 ATB 诊断的阈值。此外,本研究未比较 3 个蛋白在其他微生物感染或炎症性疾病过程中的表达水平,因此对其诊断 ATB 的特异性仍需进一步评估。

综上所述,本研究基于 AYA 群体的血清蛋白组,系统比较了 AYA 结核病患者与健康人的血清蛋白丰度差异,并通过机器学习方法挖掘

得到 AYA 结核病患者血清蛋白的诊断标志物。然而,本研究仍然存在一些局限性:(1) 本研究仅获取了 ATB 和 LTBI 组的血液生化指标,而未获取 HC 组的血液生化指标。(2) 本研究利用血清蛋白组学未能找到良好区分 LTBI 和 HC 的诊断标志物。(3) 本研究的受试人群只局限在单一地区。未来进行多中心的临床验证将有助于加深对 AYA 结核病的认识,以及这几个标志物在不同遗传背景人群中的诊断价值。

REFERENCES

- [1] GETAHUN H, MATTEELLI A, CHAISSON RE, RAVIGLIONE M. Latent *Mycobacterium tuberculosis* infection[J]. The New England Journal of Medicine, 2015, 372(22): 2127-2135.
- [2] SNOW KJ, CRUZ AT, SEDDON JA, FERRAND RA, CHIANG SS, HUGHES JA, KAMPMANN B, GRAHAM SM, DODD PJ, HOUBEN RM, DENHOLM JT, SAWYER SM, KRANZER K. Adolescent tuberculosis[J]. The Lancet Child & Adolescent Health, 2020, 4(1): 68-79.
- [3] SNOW KJ, SISMANIDIS C, DENHOLM J, SAWYER SM, GRAHAM SM. The incidence of tuberculosis among adolescents and young adults: a global estimate[J]. The European Respiratory Journal, 2018, 51(2): 1702352.
- [4] SHANG WJ, CAO GY, JING WZ, LIU J, LIANG WN,

- LIU M. Global burden of tuberculosis in adolescents and young adults: 1990–2019[J]. *Pediatrics*, 2024, 153(4): e2023063910.
- [5] LAYCOCK KM, ENANE LA, STEENHOFF AP. Tuberculosis in adolescents and young adults: emerging data on TB transmission and prevention among vulnerable young people[J]. *Tropical Medicine and Infectious Disease*, 2021, 6(3): 148.
- [6] BRODIE D, SCHLUGER NW. The diagnosis of tuberculosis[J]. *Clinics in Chest Medicine*, 2005, 26(2): 247-271.
- [7] SESTER M, SOTGIU G, LANGE C, GIEHL C, GIRARDI E, MIGLIORI GB, BOSSINK A, DHEDA K, DIEHL R, DOMINGUEZ J, LIPMAN M, NEMETH J, RAVN P, WINKLER S, HUITRIC E, SANDGREN A, MANISSERO D. Interferon- γ release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis[J]. *The European Respiratory Journal*, 2011, 37(1): 100-111.
- [8] KAY A, VASILIU A, CARRATALA-CASTRO L, MTAIFYA B, REYES JEM, MAPHALALA N, MUNGUAMBE S, MULENGWA D, NESS T, SAAVEDRA B, BACHA J, MAPHALALA G, MEJIA R, MTETWA G, ACACIO S, MANJATE P, MAMBUQUE E, SHIBA N, KOTA N, ZIYANE M, et al. Performance of a stool-based quantitative PCR assay for the diagnosis of tuberculosis in adolescents and adults: a multinational, prospective diagnostic accuracy study[J]. *The Lancet Microbe*, 2024, 5(5): e433-e441.
- [9] SUN HS, PAN LP, JIA HY, ZHANG ZG, GAO MQ, HUANG ML, WANG JH, SUN Q, WEI RR, DU BP, XING AY, ZHANG ZD. Label-free quantitative proteomics identifies novel plasma biomarkers for distinguishing pulmonary tuberculosis and latent infection[J]. *Frontiers in Microbiology*, 2018, 9: 1267.
- [10] MATEOS J, ESTÉVEZ O, GONZÁLEZ-FERNÁNDEZ Á, ANIBARRO L, PALLARÉS Á, RELJIC R, MUSSÁ T, GOMES-MAUEIA C, NGUILICHANE A, GALLARDO JM, MEDINA I, CARRERA M. Serum proteomics of active tuberculosis patients and contacts reveals unique processes activated during *Mycobacterium tuberculosis* infection[J]. *Scientific Reports*, 2020, 10: 3844.
- [11] GONG Z, GU YZ, XIONG KL, NIU JX, ZHENG RJ, SU B, FAN L, XIE JP. The evaluation and validation of blood-derived novel biomarkers for precise and rapid diagnosis of tuberculosis in areas with high-TB burden[J]. *Frontiers in Microbiology*, 2021, 12: 650567.
- [12] KAFOROU M, WRIGHT VJ, ONI T, FRENCH N, ANDERSON ST, BANGANI N, BANWELL CM, BRENT AJ, CRAMPIN AC, DOCKRELL HM, ELEY B, HEYDERMAN RS, HIBBERD ML, KERN F, LANGFORD PR, LING L, MENDELSON M, OTTENHOFF TH, ZGAMBO F, WILKINSON RJ, et al. Detection of tuberculosis in HIV-infected and-uninfected African adults using whole blood RNA expression signatures: a case-control study[J]. *PLoS Medicine*, 2013, 10(10): e1001538.
- [13] SIVAKUMARAN D, RITZ C, GJØEN JE, VAZ M, SELVAM S, OTTENHOFF THM, DOHERTY TM, JENUM S, GREWAL HMS. Host blood RNA transcript and protein signatures for sputum-independent diagnostics of tuberculosis in adults[J]. *Frontiers in Immunology*, 2021, 11: 626049.
- [14] GÁMEZ-POZO A, BÉRGES-SORIA J, AREVALILLO JM, NANNI P, LÓPEZ-VACAS R, NAVARRO H, GROSSMANN J, CASTANEDA CA, MAIN P, DÍAZ-ALMIRÓN M, ESPINOSA E, CIRUELOS E, FRESNO VARA JÁ. Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications[J]. *Cancer Research*, 2015, 75(11): 2243-2253.
- [15] EMILSSON V, ILKOV M, LAMB JR, FINKEL N, GUDMUNDSSON EF, PITTS R, HOOVER H, GUDMUNDSDOTTIR V, HORMAN SR, ASPELUND T, SHU L, TRIFONOV V, SIGURDSSON S, MANOLESCU A, ZHU J, OLAFSSON Ö, JAKOBSDOTTIR J, LESLEY SA, TO J, ZHANG J, et al. Co-regulatory networks of human serum proteins link genetics to disease[J]. *Science*, 2018, 361(6404): 769-773.
- [16] PAPIER K, ATKINS JR, TONG TYN, GAITSKELL K, DESAI T, OGAMBA CF, PARSAEIAN M, REEVES GK, MILLS IG, KEY TJ, SMITH-BYRNE K, TRAVIS RC. Identifying proteomic risk factors for cancer using prospective and exome analyses of 1463 circulating proteins and risk of 19 cancers in the UK Biobank[J]. *Nature Communications*, 2024, 15: 4010.
- [17] PENN-NICHOLSON A, HRAHA T, THOMPSON EG, STERLING D, MBANDI SK, WALL KM, FISHER M, SULIMAN S, SHANKAR S, HANEKOM WA, JANJIC N, HATHERILL M, KAUFMANN SHE, SUTHERLAND J, WALZL G, de GROOTE MA, OCHSNER U, ZAK DE, SCRIBA TJ, AND GC6-74 COHORT STUDY GROUPS ACS. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: a prospective cohort study[J]. *PLoS Medicine*, 2019, 16(4): e1002781.
- [18] RITCHIE ME, PHIPSON B, WU D, HU YF, LAW CW, SHI W, SMYTH GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Research*, 2015, 43(7): e47.
- [19] YU GC, WANG LG, HAN YY, HE QY. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *Omics: a Journal of Integrative Biology*, 2012, 16(5): 284-287.
- [20] SZKLARCZYK D, GABLE AL, NASTOU KC, LYON D, KIRSCH R, PYYSALO S, DONCHEVA NT, LEGEAY M, FANG T, BORK P, JENSEN LJ, von MERING C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. *Nucleic Acids Research*, 2021, 49(D1): D605-D612.
- [21] CAI Y, DAI YC, WANG YJ, YANG QQ, GUO JB, WEI CL, CHEN WX, HUANG HP, ZHU JL, ZHANG C, ZHENG WD, WEN ZH, LIU HY, ZHANG MX, XING SJ, JIN Q, FENG CG, CHEN XC. Single-cell

- transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis[J]. *EBioMedicine*, 2020, 53: 102686.
- [22] HAO YH, HAO S, ANDERSEN-NISSEN E, MAUCK WM 3rd, ZHENG SW, BUTLER A, LEE MJ, WILK AJ, DARBY C, ZAGER M, HOFFMAN P, STOECKIUS M, PAPALEXI E, MIMITOU EP, JAIN J, SRIVASTAVA A, STUART T, FLEMING LM, YEUNG B, ROGERS AJ, et al. Integrated analysis of multimodal single-cell data[J]. *Cell*, 2021, 184(13): 3573-3587.e29.
- [23] CONDE CD, XU C, JARVIS LB, RAINBOW DB, WELLS SB, GOMES T, HOWLETT SK, SUCHANEK O, POLANSKI K, KING HW, MAMANOVA L, HUANG N, SZABO PA, RICHARDSON L, BOLT L, FASOULI ES, MAHBUBANI KT, PRETE M, TUCK L, RICHOS N, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans[J]. *Science*, 2022, 376(6594): eabl5197.
- [24] YANG S, OUYANG J, LU YQ, HARYPURSAT V, CHEN YK. A dual role of heme oxygenase-1 in tuberculosis[J]. *Frontiers in Immunology*, 2022, 13: 842858.
- [25] SCHIFF HF, WALKER NF, UGARTE-GIL C, TEBRUEGGE M, MANOUSOPOULOU A, GARBIS SD, MANSOUR S, WONG PHM, ROCKETT G, PIAZZA P, NIRANJAN M, VALLEJO AF, WOELK CH, WILKINSON RJ, TEZERA LB, GARAY-BAQUERO D, ELKINGTON P. Integrated plasma proteomics identifies tuberculosis-specific diagnostic biomarkers[J]. *JCI Insight*, 2024, 9(8): e173273.
- [26] MARTENS GW, ARIKAN MC, LEE J, REN FC, VALLERSKOG T, KORNFELD H. Hypercholesterolemia impairs immunity to tuberculosis[J]. *Infection and Immunity*, 2008, 76(8): 3464-3472.
- [27] GIL-SANTANA L, CRUZ LAB, ARRIAGA MB, MIRANDA PFC, FUKUTANI KF, SILVEIRA-MATTOS PS, SILVA EC, OLIVEIRA MG, MESQUITA EDD, RAUWERDINK A, COBELENS F, OLIVEIRA MM, KRITSKI A, ANDRADE BB. Tuberculosis-associated anemia is linked to a distinct inflammatory profile that persists after initiation of antitubercular therapy[J]. *Scientific Reports*, 2019, 9: 1381.
- [28] MITRA A, KO YH, CINGOLANI G, NIEDERWEIS M. Heme and hemoglobin utilization by *Mycobacterium tuberculosis*[J]. *Nature Communications*, 2019, 10: 4260.
- [29] SINGH PP, GOYAL A. Interleukin-6: a potent biomarker of mycobacterial infection[J]. *SpringerPlus*, 2013, 2: 686.
- [30] FARHADIAN M, VEISI S, FARHADIAN N, ZAMANIAN MH. Hematological parameters in newly diagnosed TB patients: a systematic review and meta-analysis[J]. *Tuberculosis*, 2024, 144: 102430.