**妊娠早期血浆蛋白组学分析揭示先天性心脏病的新型生物标志物**

翻译：梁丽俊 山西省儿童医院

审校：杨晓芳 首都医科大学附属北京安贞医院

**摘要**

先天性心脏病（CHD）的产前诊断主要依赖于孕中期进行的胎儿超声心动图检查，不同中心和不同医生对该检查的敏感性各不相同。因此，一种客观的早期诊断方法是很有必要的。在此，我们进行了一项病例对照研究，招募了 103 名有健康后代的孕妇和 104 名有先天性心脏病后代的病例，包括 VSD（42/104）、ASD（20/104）和其他先天性心脏病表型。在怀孕头三个月收集血浆并进行蛋白质组分析。主成分分析显示，对照组与CHD之间存在明显差异。在发生显著改变的蛋白质中，25个上调的蛋白质富集于氨基酸代谢、细胞外基质受体和肌动蛋白骨架调节，而49个下调的蛋白质富集于碳水化合物代谢、心肌收缩和心肌病。机器学习模型的曲线下面积达到了0.964，在识别CHD方面具有很高的准确性。这项研究为更好地识别先天性心脏病的病因提供了非常有价值的蛋白质组学资源，并为早期识别先天性心脏病开发了一种可靠的客观方法，有助于早期干预和更好的预测。

**关键词** 先天性心脏病；血浆；妊娠期蛋白组学

**引言**

先天性心脏病（CHD）是全球最常见的出生缺陷（van der Linde et al，2011；Zhang et al，2021a）。由于其发病率和致死率高，在孕早期及时发现先天性心脏病对预防和治疗至关重要（Botto，2000）。对先天性心脏病的诊断越早，预后越好（Sadeck Ldos等，1997年）。目前，即使胎儿超声心动图已成为筛查先天性心脏病的工具，但在常规产前检查中，心脏畸形仍被忽视，检出率从 6% 到 35% 不等（Sharland & Allan, 1992; Garne et al, 2001; Jaeggiet al, 2001）。此外，由于缺乏标准化，各中心的超声波检查结果也不尽相同（Friedberg等，2009年）。产前超声检查检测先天性心脏病的准确性受很多因素影响，如操作人员的经验、超声设备的质量、病变类型、部门政策和指南（Stumpflen等，199年；Fernandez等，1998年；Stoll等，1998年；Isaksen 等，1999年；Allan，2000年）。这也降低了欠发达地区的先心病检出率。因此，开发早期诊断先天性心脏病的新方法已成为预防和治疗出生缺陷的关键。

除了影像学方法（Wang等，2021年），遗传和生化方法也可用于先天性心脏病的早期检测。叶酸代谢基因中的各种单核苷酸多态性与先天性心脏病的风险有关（Zhao等，2012，2013，2014；Wang等，2014，2017）。然而，由于这些变异在人群中也常见，因此它们在预测后代患先天性心脏病风险方面的适用性并不令人满意。最近，人们发现母体生物标志物与胎儿在子宫内的先天性心脏病有关，包括游离β-人绒毛膜促性腺激素（β-hCG）和支链氨基酸水平的升高（Zhang等，2022年），以及妊娠前三个月妊娠相关血浆蛋白-A（PAPP-A）水平的降低（Michai-lidis & Economides，2001年；Souter 等，2002 年；Makrydimas 等，2003年；Jelliffe-Pawlowski 等，2008年）。此外，血液microRNAs和lncRNAs谱的改变也作为胎儿先天性心脏病产前诊断的新型生物标志物被探索（Zhu等，2013；Gu等，2016）。然而，目前在鉴定生物标志物方面取得的进展仍无法满足人们对改进生物标志物以早期诊断先天性心脏病的迫切需求。

无论冠心病的病因是遗传、营养还是环境，蛋白质都是心血管发展的分子机器。与疾病相关的蛋白质活性的改变可导致疾病的发生。另外，蛋白质的变化也可能是先天性心脏病的后果，因为胎儿心血管系统的结构和功能发育缺陷可能会反映在母体外周血的蛋白质组成中。不管它们是CHD的原因还是后果，在胎儿心脏发育的预期阶段，心血管细胞中的这些蛋白质变化可能会导致母体血液中蛋白质水平随之变化，这可以在孕早期检测到，并为CHD的早期诊断提供新的机会。

血液作为一种重要的研究来源，可能含有人类细胞中发现的任何类型的蛋白质（Chen等，2023b）。血液中的蛋白质不仅反映生理和病理状况，还包括疾病和治疗效果的生物标志物（Ku等，2023年）。本研究旨在对患有和未患有先天性心脏病胎儿的前三个月孕妇的血浆蛋白进行无偏见的蛋白质组学分析，以全面鉴定新型诊断生物标志物。一组可靠的蛋白质生物标志物而不是单一的标志物，还能开发出用于早期妊娠诊断的高度特异性检测方法，并可能为了解 CHD的发病机制提供新的视角。

**研究结果**

**母体血浆的蛋白质组学特征**

为了绘制妊娠早期先天性心脏病后代与健康后代之间血浆蛋白质组学的变化图，我们分析了两个独立的病例对照组，共招募了206人。其中第一组包括复旦大学附属妇产科医院的67例CHD后代患者和71例对照组，第二组包括中国福利会国际和平妇幼保健院的37例病例和32例对照组，共计103例对照组和104例病例组，最常见的表型为VSD（42/104）和ASD（20/104）（图1A，数据集EV1）。

实验工作流程如图1A所示。两组所有样本的 LC-MS/MS 分析均采用数据独立采集（DIA）方法，并对所有血浆蛋白组学数据进行了分析。应用这种稳健的工作流程，我们对每个血浆样本定量了平均2220个（1组）和 1926个（第2组）蛋白质（图1B；数据集EV2和EV3）。没有异常值，所有样本都可用于进一步分析（图1B）。此外，我们在第1组和第2组中分别鉴定出 8624 和7049个蛋白质（图1C和D）。随着样本数量的增加，蛋白质的数量逐渐趋于稳定，这表明蛋白质检测具有深度覆盖和良好的稳定性。与第1组相比，第 2组血浆样本中鉴定出的蛋白质数量较少，这可能是批次效应造成的。尽管如此，我们还是实现了单次高通量工作流程，从2μl血浆样本中实现了蛋白组学覆盖（图1E）。在DIA获得的数据中，有351个蛋白质的完整度达到100%，1023个蛋白质的完整度达到75%，1604个蛋白质的完整度达到 50%（图 1E）。在所有样本中，对照组和CHD组的定量蛋白质强度跨越8个数量级，排名前10的高丰度蛋白贡献了各自数据集中所有血浆蛋白质丰度的40%和39%（图1F）。此外，我们还通过分析整个测量范围内相关蛋白质的丰度来评估蛋白质组数据的可重复性。从207个样本中随机抽取20个样本（10个CHD 样本和10个对照样本）进行分析。结果显示，在健康对照组中，单个样品之间量化蛋白质信号的平均相关值为0.978，范围为 0.96-0.99。同样，CHD 组的平均相关值为 0.965，范围为 0.93-0.99。CHD组与健康对照组的平均相关值为0.939（图EV1A）。此外，为了说明每个病例中数据的良好重现性，我们选取了五种表征良好的血浆蛋白，发现在103个健康对照组和104个CHD病例中的定量结果具有很高的重现性（图EV1B）。



**图 1. 研究概况和母体血浆的蛋白质组学特征。**

A 研究组概况和蛋白组学原理示意图。母体血样在妊娠10-12周时采集。说明了每组的总人数和蛋白质组分析的基本程序。

B 蛋白质的定量采用1%的伪发现率 (FDR) 截止值。n = 71（第1组对照）、67（第1组 CHD后代）、32（第2组对照）和 37（第2组CHD 后代）作为生物重复。

C 第1组中已鉴定蛋白质的累计数量（左图显示对照组的结果，右图显示CHD组的结果）。 数据集中的蛋白质数量（Y 轴）与样本数量（X 轴）相对应。

D 第2组中已鉴定蛋白质的累计数量（左图为对照组的结果，右图为CHD组的结果）。 数据集中的蛋白质数量（Y轴）与样本数量（X轴）相对应。

E 数据完整性曲线。数据集中的蛋白质数量（Y轴）与量化蛋白质的最少样本数量（X 轴）的关系图。箭头表示数据完整性值为50、75 和100%。

F CHD 组（红色）和对照组（蓝色）的蛋白质丰度分布图。方框内标出了含量最高的前 10种蛋白质。

**检测母体血浆中与CHD相关的蛋白质组改变**

主成分分析（PCA）显示，两组中病例与对照之间存在明显差异，表明在妊娠早期，怀有 CHD 胎儿的孕妇与怀有健康胎儿的孕妇表现出不同的血浆蛋白组特征（图2A和C）。此外，我们还在两组孕妇中发现了与CHD相关的分子特征。第1组中发现对照组和CHD组之间有397个差异表达蛋白（DEPs）。其中，CHD组检测到184个显著上调蛋白和213个下调蛋白（Student's t 检验，*P*＜0.05，差异倍数＞2 或＜0.5）（图2B，数据集EV4）。在第2组中，对照组和CHD组之间发现了225个DEPs。其中，CHD组中有80个蛋白质显著上调，145个蛋白质显著下调（Student's t检验，*P* < 0.05，差异倍数> 2或< 0.5）（图2D，数据集EV5）。这些DEPs包括之前确定的几个血浆CHD标志物，如神经蛋白-2、ATP-柠檬酸合成酶（ACLY）、蛋白S100-A7（S100A7）、肌球蛋白调节轻链9（MYL9）、葡萄糖-6-磷酸1-羧酸（Glucose-6-phosphate 1-羧酸）、葡萄糖-6-磷酸 1-脱氢酶（G6PD）、丝氨酸羟甲基转移酶（SHMT1）、类内酰胺酶蛋白（COTL1）和纤溶酶原激活剂抑制剂1（SERPINE1）（Jain等，2004； Nembhard等，2017；Song等，2017；Ference等，2019；Xiong等，2019；Zhanget等，2019；Tan等，2020；Luo等，2022）。此外，我们还发现了其他发生显著变化的蛋白质，如脱氧核苷三磷酸三磷酸水解酶1（SAMHD1）和富含半胱氨酸的分泌蛋白（SPARC）。我们还用Western印迹法验证了蛋白质组学的蛋白质丰度结果，证实在怀有CHD胎儿的孕妇血浆样本中，G6PD和SHMT1的表达量减少，而MYL9的表达量增加（图 EV2）。

为了进一步分析母体血浆中变化蛋白的概况，我们利用基因组学（GO）技术对两组中显著的DEPs进行了注释，并确定了受胎儿CHD影响的生物学过程。在两组中，CHD组有264个蛋白明显上调，主要富集在氨基酸代谢、细胞外基质（ECM）受体、肌动蛋白骨架调节、Ras-MAPK信号通路和PI3K-Akt信号通路中。相反，358个明显下调的蛋白与碳水化合物代谢、心肌收缩力和心肌病密切相关（图2E）。此外，为了了解各种上调和下调蛋白的分子通路关系，对改变的蛋白进行了蛋白-蛋白相互作用（PPI）网络分析，揭示了各通路中的关键分子（图2F，表EV1和EV2）。这些结果表明，怀有CHD胎儿的孕妇血浆中表达了大量与胚胎器官发育相关的蛋白质，这与怀有健康胎儿的孕妇有显著差异，这些显著改变的蛋白质有望成为疾病的生物标志物。



**图 2：两组中 CHD组与对照组血浆蛋白质组的差异。**

A 第1组血浆样本蛋白质的主成分分析（PCA）。 对照组用红色表示，CHD组用蓝色表示。

B CHD/对照组样本的差异倍数与第1组的统计学意义之间的关系。红色表示上调的蛋白质，蓝色表示下调的蛋白质，灰色虚线以上的蛋白质具有统计学意义（P < 0.05）。

C 第2组血浆蛋白质的PCA。红色代表对照组，蓝色代表CHD组。

D CHD/对照组样本的差异倍数与第2组的统计学意义之间的关系。红色表示上调的蛋白质，蓝色表示下调的蛋白质，灰色虚线以上的蛋白质具有统计学意义（P < 0.05）。

E 上调或下调蛋白质的基因组学（GO）注释（P < 0.05）。

F 两组富集通路的蛋白质相互作用分析见 (E)。红色表示上调蛋白之间的相互作用，蓝色表示下调蛋白之间的相互作用。

本图的原始数据可在线获取。

**两组中CHD相关蛋白的复制**

8种已知的CHD相关蛋白的表达水平，包括热休克蛋白HSP 90-α（HSP90AA1）（Sevim Bayraket al, 2020； Xiao et al, 2020）、苹果酸脱氢酶（MDH2）（Liu et al, 2014）、MYL9（Xiong et al, 2019）、ACLY（Ference et al, 2019）、N-乙基马来酰亚胺敏感因子（NSF）（Mei et al, 2020）、a-肌球蛋白基因（TPM）（Hirono et al, 2020； Zhang等人，2020）、SERPINE1（Song等人，2017）和补体因子H相关2（CFHR2）（Zhang等人，2016）在两组中均发生了一致的改变。HSP90AA1、MDH2、MYL9和ACLY的蛋白表达水平在CHD组高于对照组。相比之下，NSF、TPM1、SERPINE1和CFHR2在CHD组的蛋白表达水平低于对照组（图3A）。这些结果表明，母体血浆表达的蛋白质与心脏发育有关。

此外，74种蛋白质在CHD组和对照组血浆中的水平变化在第1组和第2组之间是一致的。两组中共有25种蛋白质显著上调，49种蛋白质显著下调（图3B；表EV3）。我们进一步分析了这74个DEPs，发现25个上调的蛋白主要参与代谢、先天性免疫反应和细胞周期途径，而49个下调的蛋白主要参与葡萄糖代谢、脂质代谢和血管相互作用等过程，这些途径可能对胎儿心脏的发育至关重要（图3C和D）。



**图 3. 两组血浆蛋白质组的变化。**

A 以前报道的心脏病相关蛋白的表达水平。\*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05; *P*值来自非配对t检验。n = 71（第1组对照）、67（第1组CHD后代）、32（第2组对照）和 37（第 2 组CHD后代）作为生物重复。

B 上调和下调蛋白的维恩图。

C 两组中已识别蛋白质的表达水平和通路富集。热图显示了两组中25个上调蛋白（上图）和49个下调蛋白（下图）的表达水平、频率和 *P* 值。

D 两组差异表达蛋白的通路模式。

本图的原始数据可在线获取。

**构建与验证CHD蛋白共表达网络**

为了确定CHD病理的潜在驱动因素，我们在随后的分析中纳入了缺失值少于30%的蛋白，并利用加权基因共表达网络分析（WGCNA）选择了约2280个蛋白生成蛋白共表达网络。共表达网络由10个蛋白模块（M1-M10）组成，其大小从20个蛋白到492个蛋白不等（图4A；数据集EV6）。随后，对 12 种临床病理表型进行了分层聚类分析，将其分为以下三个表型群：第1组，包括法洛四联症（TOF）、主动脉瓣狭窄（AS）、持续性动脉导管未闭（PTA）、大动脉转位（TGA）和右室流出道梗阻（RVOTO）；第2组，包括三尖瓣反流（TR）和房间隔缺损（ASD）； 第3组包括肺动脉狭窄 (PS)、室间隔缺损 (VSD)、永存左上腔静脉 (PLSVC)、左室流出道梗阻 (LVOTO) 和房室间隔缺损 (AVSD)。应用GO富集分析来研究与CHD表型相关的蛋白质模块的生物学过程。我们观察到六个与临床表型簇显著相关的模块。簇1与蛋白质模块4和8显著相关，富含碳水化合物代谢、谷胱甘肽代谢和免疫反应。簇2与蛋白质模块9和10明显相关，并富含ECM受体反应、蛋白质转运和信号转导。簇3 与蛋白模块2和3有明显关联，主要富集于心肌生长、发育和肌动蛋白细胞骨架调控（图4B）。为了进一步研究不同集群中影响疾病发生的潜在因素，我们分析了通路的PPI网络（图4B；表EV4）。数据显示了各通路中的关键分子，如淀粉酶α2A（AMY2A）、免疫球蛋白λ样多肽 5（IGLL5）、MYL9、整合素亚单位α1（ITGA1）、RAS 协同基因家族成员 RAB6A 和 catenin beta 1（CTNNB1）（图4C）。这些枢纽基因可将CHD与对照组区分开来。

此外，我们还研究了10个蛋白质模块与18个临床指标之间的关系，其中只有3个模块（M2、M3和M6）与CHD病理表现之间存在强相关性。第3组（M2和M3）与血脂呈正相关，与甲状腺激素呈正相关。这些结果进一步表明，血脂和甲状腺激素水平的升高可能是导致CHD的危险因素。此外，M6水平与甲状腺激素水平呈负相关（*P*＜0.05）（图4D）。 此外，我们还比较了对照组和不同CHD组妊娠早期妇女血浆中七种脂质的浓度，发现CHD组中游离脂肪酸的水平显著下降，而其他六种脂质的水平则显著升高。而其他六种脂类，包括低密度脂蛋白（LDL）、甘油三酯、高密度脂蛋白、总胆固醇（TC）、载脂蛋白A（APOA）和载脂蛋白 B（APOB）在CHD组中则呈上升趋势（图4E）。这些结果与之前的发现一致，即血脂是CHD的主要风险因素（Pencina et al, 2019; Bogachkov et al, 2020; Hu et al, 2020; Kalaivani & Jaleel, 2020; Mehta et al, 2020; Dugani et al, 2021），表明血脂波动可能与CHD的发生有关。由于我们对WGCNA分析中的蛋白质模块2特别感兴趣，因此我们进一步研究了CHD 样本与对照样本之间的DEPs。在模块2中的26个蛋白质中，有两个蛋白质与血脂水平呈正相关（图4F），SELENBP1和MSN在CHD第3组中显著增加，其中SELENBP1增加了51% (P = 0.0114)，MSN 增加了 201% (P = 0.0162)。

**免疫景观中的CHD**

为了研究CHD发生后免疫微环境的变化，我们使用xCell分析确定了与免疫（CD4+ T细胞和NK细胞）或间质（脂肪细胞和内皮细胞）特征相关21种不同细胞类型，并通过特定细胞类型的存在或缺失对每个样本进行区分（图 4G）。对照组富集了内皮细胞、肌细胞和外周细胞，而CHD组富集了CD4+T 细胞、脂肪细胞和前体脂肪细胞（图4G）。这些结果表明，如图4I所示，CHD 的发生伴随着血脂、脂肪细胞和 CD4 +T细胞的增加，以及内皮细胞、肌细胞和外周细胞的减少，而这些细胞可能参与了CHD的发病机制。血脂升高会抑制心肌细胞的增殖和迁移，阻碍血管生成，并激活CD4+ T细胞，从而诱发冠CHD。活化的CD4+ T细胞可释放促炎症细胞因子，激活巨噬细胞和血管细胞，引起急性炎症。它们还可引起慢性炎症状态，进一步损害心脏发育（图 4I）。



**图4. CHD血浆蛋白质组与临床表型和指标的关系**

A 对207份血浆样本进行的加权基因共表达网络分析（WGCNA）显示，根据模块蛋白之间的相关性，CHD的12个病理特征可整合为三个簇。双色热图显示了正相关（红色）和负相关（蓝色）的强度。Pearson相关系数和*P*值由WGCNA软件包计算。\*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05。

B 每个簇中蛋白质的基因本体（GO）分析。

C 三种主要病理特征的蛋白-蛋白相互作用分析。

D 不同CHD蛋白网络的变化及其与临床指标的相关性分析。

E 对照组和CHD组血脂的表达水平。\*\*\**P* < 0.001；\*\**P* < 0.01；\**P* < 0.05；*P*值来自独立样本*t*检验。n = 103（对照组）、10（簇1）、23（簇2）和 71（簇3）个生物重复。

F 蛋白表达与临床指标（血脂）的相关性。\*\*\**P* < 0.001； \**P* < 0.05；*P* 值来自独立样本 *t* 检验。n = 103（对照组）、10（簇1）、23（簇2）和 71（簇3）个生物重复。

G 对照组和CHD组中特异性免疫细胞群类型的热图。

H 散点图说明对照组和 CHD组特定细胞类型的xCell分数。\*\*\**P* < 0.001；\*\**P* < 0.01；\**P* < 0.05；显示了独立样本 *t* 检验的 *P* 值。n = 103（对照组）、10（簇1）、23（簇2）和 71（簇3）个生物重复。

I 血脂与CHD的潜在发病机制。

本图的原始数据可在线获取。

**基于机器学习识别CHD 的生物标记物组合**

基于第1组的血浆蛋白质组学数据，采用机器学习方法识别在孕妇妊娠前三个月期间识别患有CHD胎儿的潜在生物标记物组合。潜在生物标志物的选择标准见材料与方法。为了优化训练模型的参数并评估模型的性能，我们进行了10倍交叉验证。机器学习模型是基于第1组80%的蛋白质组学数据建立的，选择的生物标记物组合包含以下九种蛋白：钙蛋白酶-5（CAPN5）、烯醇化磷酸酶 E1（ENOPH1）、组蛋白 H2A 1-C 型（H2AC6）、HSP90AA1、导入蛋白亚基 beta-1 （KPNB1）、MDH2、MYL9、radixin（RDX）和 SAMHD1。该模型在训练集中的曲线下面积（AUC）为0.964（95% CI= 0.862-0.968）（图5A）。第1组的剩余数据用于测试数据集，其AUC值为0.989（95% CI = 0.718-0.977）（图5D）。这些结果表明，基于母体血浆样本的这种生物标志物组合对诊断先天性心脏病具有临床意义。为了进一步验证该组合的性能，我们从第2 组收集了69份血浆样本进行组间验证（32份对照组样本和37份CHD 病例组样本），结果AUC值为0.963（95% CI = 0.878-0.997）（图5G）。此外，为了评估机器学习策略的可靠性，这些生物标记物组合的混淆矩阵和PCA结果显示，对照组和CHD组的分类准确率相对较高（图 5B、C、E、F、H和I）。利用独立生物标记物识别CHD的潜在临床意义见图 EV3。



**图 5. 利用机器学习开发预测CHD的生物标志物组合。**

A-C (A) 簇1训练数据集的接收者操作特征曲线（ROC），(B) 训练数据集中组合生物标志物的混淆矩阵，(C) 预测CHD和对照组结果的主成分分析图（PCA）。

D-F (D) 簇1测试数据集的ROC曲线，(E) 测试数据集中组合生物标记物的混淆矩阵，(F) 预测CHD和对照组结果的PCA图。

G-I (G) 簇2外部验证集的ROC曲线，(H) 外部验证集中组合生物标记物的混淆矩阵，(I) 预测CHD和对照组结果的PCA图。

**讨论**

由于近年来心脏介入治疗和手术策略的改进，对于新生儿来说可能危及生命的先天性心脏病的致命性似乎降低了。要更好地治疗先天性心脏病，尤其是危重病例，关键在于早期诊断和及时治疗，甚至可以在产前对主动脉瓣狭窄进行胎儿心脏介入治疗，从而避免左心发育不全综合征的发生。因此，对先天性心脏病的产前诊断不仅能让父母和医务人员为孩子出生后的治疗做好准备，还能进行早期干预，并可能获得更好的预后。在妊娠中期进行胎儿超声心动图的扫描（SAS）是目前公认的先天性心脏病产前诊断方法。在未经选择的人群中，使用SAS对CHD的检出率约为45%，不同类型的CHD检出率不同，例如，三尖瓣闭锁的检出率为90.8%，主动脉狭窄的检出率为22.3%（van Velzenet al，2018），并且受个人经验的影响。由于胎儿心脏发育在怀孕6周左右完成（Helle& Priest，2020年），因此需要能在心脏发育完全之前识别CHD 的客观指标，以进行准确的产前诊断和未来的产前干预。

一些血液学生物标志物，包括PAPP-A、游离b-HCG、钠尿肽和染色体微阵列分析，在早期发现CHD方面具有潜在价值（Miyoshiet al，2018；Wang et al，2018b；Alanen et al，2019；Gu et al，2019；Bahado-Singh et al，2020；Chen et al，2022）。在本研究中，我们从两家不同医院的两组孕妇中收集了怀有或没有先天性心脏病胎儿的初产孕妇的血浆样本。这项研究设计的目的是鉴定母体蛋白表达谱，以作为早期检测胎儿先天性心脏病的新型非侵入性生物标记物。在这里，我们在第1组中发现184个明显上调的蛋白和213个下调的蛋白，在第2组中发现了80个上调的蛋白和145个下调的蛋白，维恩图显示两组中均有25个上调的蛋白和49个下调的蛋白，这表明孕早期母体血浆蛋白谱有可能反映出CHD的发病情况。我们利用WGCNA识别10个蛋白模块，并利用层次聚类分析将CHD亚型分为三组，发现有6个模块与临床表型组显著相关，这表明每个组可能存在共同的发病机制。此外，为了识别潜在的生物标志物以准确鉴定胎儿先天性心脏病，我们开发了基于机器学习的管道，结果识别出七个生物标志物以及一组生物标志物组合，它们可以准确预测不同的先天性心脏病结局，而且灵敏度很高。

此外，我们的研究还发现了CHD组和对照组中参与新陈代谢的DEPs。这与代谢失调（如孕产妇糖尿病）与先天性心脏病风险相关的事实是一致的（Lisowski et al, 2010）。最近，母亲肥胖也被认为是CHD的一个风险因素（Persson等，2019年）。由于肥胖与糖尿病的表型重叠，肥胖是通过促进糖尿病而增加先天性心脏病风险，还是通过血脂异常和其他次生效应（包括高胰岛素血症、胰岛素抵抗和氧化应激）损害发育中的心脏，尚有待阐明（Helle & Priest，2020；Chen et al，2023a）。在这里，我们发现血脂（包括低密度脂蛋白和总胆固醇）在CHD组中上调，细胞类型免疫富集表明CHD组中脂肪细胞和前脂肪细胞富集。这表明血脂异常可能是导致CHD的原因之一。免疫CD4+ T细胞也在CHD组中富集。现有的关于CHD与免疫和炎症相关性的知识还很有限。有几种免疫特征与CHD相关（Singam-palli等，2021年），免疫/炎症状态的改变可能是CHD的原因（Swirski & Nahrendorf，2018年），也可能是CHD的结果，或伴随有共同病因的CHD。由于CD4+ T细胞可受脂肪细胞衍生脂质的调节，它们可能在肥胖者发生CHD的过程中发挥积极作用。

在这些研究中，首先，我们不能排除临床前疾病对蛋白质图谱结果的潜在影响。本研究招募的孕妇被证实没有明显的健康问题或严重的代谢紊乱。然而，可能有某些临床前病症未被发现，这些病症会因怀孕而加重，从而干扰蛋白质图谱结果。在未来的研究中，需要从怀孕前或怀孕后的同组人群中获取另一组蛋白质组学数据，以消除这种可能性。其次，本研究应谨慎应用于预测危重的CHD。在我们中心，大多数危重先天性心脏病病例都是在妊娠早期或中期通过超声心动图检查发现的，然后根据父母的意愿中止妊娠，因此没有机会进行遗传学研究或产后表型确认。由于我们招募的病例仅在出生后由新生儿科和儿科心脏病专家进行了完整的体格检查、超声心动图检查和遗传学检查，因此研究的病例主要是轻度VSD和ASD，没有足够的数据进行分层分析以评估疾病严重程度的特异性和敏感性，而这对临床治疗无疑是非常重要的。扩大样本量或专门招募危重先天性心脏病患者的进一步研究将有助于解决这一问题。此外，由于样本量有限，且普通人群中CHD的发病率较低，因此在得出结论时应慎重，并在未来谨慎使用这种微观方法。在应用于临床之前，需要进行更大规模的试验来验证结论并测试诊断面板。此外，在我们的研究中，与对照组相比，病例的血脂更高，因此参与脂质代谢的蛋白质与病例组的相关性可能反映了过度拟合。进一步研究在更大范围内比较病例组和对照组的脂质组成和参与脂质代谢的蛋白质，并揭示其潜在机制，将有助于了解脂质代谢在冠心病发生中的作用。

总之，我们发现并验证了包含九种蛋白质的生物标记物组合，它们可作为新型的非侵入性生物标记物用于检测胎儿先天性心脏病的孕产妇。只需在妊娠早期检测孕妇的2μl血浆，采用这种组学方法，指导有高风险的孕妇接受更深入的随访，并为产前干预做好准备，这无疑有助于有针对性地使用医疗资源。此外，这项工作还为研究界提供了非常宝贵的蛋白质组学资源，有助于更好地了解先天性心脏病的病因，确定一系列有价值的生物标志物，并为潜在的治疗策略提供线索。

**材料与方法**

**研究对象与伦理**

在本研究中，分析了来自两个独立病例对照组的孕早期（孕10-12周）母体血浆样本。本研究的设计和实施经复旦大学附属妇产科医院伦理委员会（伦理委员会）2015-17-C1和中国福利会国际和平妇幼保健院伦理委员会（伦理委员会）GKLW-2018-37批准和监督，符合《赫尔辛基宣言》规定的标准，所有实验均符合美国卫生与公众服务部贝尔蒙特报告。

如前所述（Zhang et al, 2022），第一组受试者于2018年1月至2019年12 月在中国上海复旦大学附属妇产科医院招募。该组包括67名后来被确诊为怀有CHD胎儿的孕妇和71名后代健康的对照组。第二组包括来自中国福利会国际和平妇幼保健院的37例病例和32例对照，这两组病例和对照是同时独立招募和分析的。对照组中的孕妇一般健康状况良好，并通过生命体征和生理指标的基线比较进行确认。有以下情况之一的孕妇被排除在外：临床或生化感染迹象、多胎妊娠、糖尿病或其他严重代谢紊乱。妊娠22 周时，首先通过检查畸形来确定先天性心脏病的表型，出生后再通过彩色超声心动图进行确认。本研究不包括孤立的动脉导管未闭、卵圆孔未闭、主动脉瓣双尖瓣、冠状动脉畸形以及主要与血管系统有关的先天性心脏病病例。所有患有遗传性综合征或已知染色体异常（如唐氏综合征、Holt-Oram 综合征、Alagille 综合征、DiGeorge 综合征、William 综合征和 Noonan 综合征）或一级亲属（父母、兄弟姐妹或子女）有心脏病家族史的患者均排除在外。所有参与者均为无血缘关系的汉族人。表1和数据集EV1 显示了怀有或没有先天性心脏病的孕妇的人口统计学特征，包括叶酸使用情况、母/父吸烟和饮酒情况、血液生化指标以及先天性心脏病后代的表型。



**表1 参与者的人口特征**

数据以均数±标准差表示。孕妇特征的 *P* 值来自独立样本 *t* 检验（两组具有相同的 SD 值）或经韦尔奇校正的独立样本 *t* 检验（两组不具有相同的 SD 值）。

独立样本 *t* 检验，并进行韦尔奇校正（两组的 SD 值不相等）；后代特征的 *P* 值通过卡方检验得出。VSD，室间隔缺损；ASD，房间隔缺损；PLSVC，持续性左上腔静脉；PS，肺动脉狭窄；LVOTO，左心室流出道梗阻；AVSD，房室隔缺损；PS，肺动脉狭窄；LVOTO，左心室流出道梗阻。AS：主动脉瓣狭窄；TR：三尖瓣反流；PTA：持续性动脉导管未闭；TOF：法洛氏四联症；TGA：大动脉转位、RVOTO：右室流出道梗阻；NA：不详。

**样品制备**

从血浆样本中提取蛋白质时，首先使用免疫耗竭试剂盒（赛默飞世尔）按照制造商的说明去除血浆中含量最高的14种蛋白质，然后在85°C下灭活10 min。用胰蛋白酶以1:25的酶与蛋白质量比在37°C下消化去除的血浆，然后提取肽并干燥（SpeedVac，Eppendorf）。

**液相色谱-串联质谱分析**

液相色谱-串联质谱（LC-MS/MS）分析在EASY-nLC 1200超高压系统上进行，该系统通过纳米电喷雾离子源（Thermo Fisher Scientific）与Orbitrap Fusion Lumos 质谱仪相连。

在280 bar的最大压力下，用14μl A溶液（0.1%甲酸水溶液）将5μl样品装入捕集柱（C18，100μm，2.5cm）。多肽在内径为150μm I.D.×15 cm的色谱柱（C18，1.9 μm，120A，Maisch GmbH）上分离，流动相B（乙腈和 0.1%甲酸）的梯度为15%-30%，流速为 600 nl/min，分离时间为75min。

DIA方法包括在60 K分辨率（自动增益控制 [AGC] 目标值为 4e5 或 50 ms）下从300-1,400 m/z的MS1扫描。在15 K分辨率下顺序采集30个DIA片段，最大注入时间的AGC目标值为5e4或22 ms。启用了 "在所有可用的可并行时间内注入离子"的设置。高能量碰撞诱导解离析出设置为归一化碰撞能 30%，MS2的默认电荷状态设置为3。

**质谱数据处理**

所有数据均使用Firmiana（Feng等，2017年）进行处理，DIA数据与 UniProt人类蛋白质数据库进行了检索 (更新于2019.12.17，20,406个条目），使用 FragPipe (v12.1) 和 MSFragger (2.2) (Kong et al, 2017)。使用胰蛋白酶蛋白水解裂解，允许最多两次漏切和最小肽长为7个氨基酸。搜索包括作为固定修饰的半胱氨酸氨甲基化和N-乙酰化和蛋氨酸氧化作为可变修饰。肽的前体质量精确度偏差为20 ppm，片段质量偏差为50 mmu。前体得分电荷设置为+2、+3和 +4。蛋白质和肽水平的伪发现率（FDR）均为1%。共有327作为参考谱库，并将DIA 结果合并到这些参考谱库中。使用 SpectraST 软件将DIA结果合并到这些参考库中。

使用 DIA-NN (v1.7.0）（Demichev等人，2020 年）分析与数据无关的采集数据。默认设置用于DIA-NN（前体FDR：5%，Log lev：1，质量精度：20 ppm、MS1精度：10 ppm，扫描窗口：30，隐性蛋白质组：基因，定量策略：稳健液相色谱（高精度））。鉴定出的确定的肽的定量为所有参照库中色谱碎片离子峰面积的平均值。片段离子峰面积的平均值进行量化。蛋白质定量采用无标记和基于强度的绝对定量（iBAQ）方法（Zhang等，2012年）。

**生物信息学分析**

**缺失值估算**

对于血浆蛋白质组数据，采用FOT乘以10进行定量，缺失值用最小值的1/10进行估算。

**统计分析**

数据统计分析采用GraphPad Prism 8软件和R-studio（版本1.3.1093）。缺失值输入和数据归一化后，独立样本采用*t*检验进行显著性分析，以确定CHD和健康对照组之间的差异表达蛋白。蛋白质强度与甲状腺激素、血脂和血糖的相关性采用*Pearson*相关系数进行分析。

**差异表达蛋白**

生物信息学分析使用 R-studio（版本 1.3.1093）进行。差异表达分析仅考虑NAs<50%（缺失值）的蛋白质。对每种蛋白质的相对丰度进行对数转换，得到最终的相对丰度值，并用*t* 检验来确定第1组和第2组病例和对照之间有显著变化的蛋白质。*P*<0.05、差异倍数>2 或<1/2 的显著DEPs分别标记为上调基因和下调基因。差异表达以曲线图表示。

**功能富集分析**

使用 DAVID v6.8 (http://david.ncifcrf.gov, anonline bioinformatics tool for gene function annotation) 和 Reac-tome (https://reactome.org) 富集分析软件分析病例样本和对照样本之间DEPs的功能富集（Huang da et al, 2009）。*P*<0.05的通路被认为具有统计学意义。使用String 11.0数据库和 Cytoscape软件对蛋白质相互作用（PPIs）进行可视化。

**加权基因共表达网络分析**

使用 R 软件包 WGCNA（Langfelder & Horvath，2008年）构建蛋白质共表达网络。我们在WGCNA中输入了208名患者中30% 以上存在的2280 个蛋白质。使用分块模块WGCNA函数计算蛋白质表达谱之间的*Spearman*相关系数，其设置如下：软阈值功率 β = 5（因为这是导致无标度R2拟合值为 0.85的最小阈值），最小模块大小= 20，合并切割高度= 0.3，从而使用双相关函数计算拓扑重叠（TOM）。此外，使用1-TOM（dissTOM）作为距离度量对基因进行分层聚类。每个模块由按比例模块表达谱的第一个主要分量来概括，称为模块特征基因。

**细胞类型富集分析**

使用x-Cell(http://xcell.ucsf.edu)工具（Aran等，2017年）推断了每种细胞类型的丰度，该工具从基因表达数据中对64种免疫和基质细胞类型进行了细胞类型富集分析，并生成了每个样本的免疫评分。

**潜在的CHD诊断生物标志物**

数据处理和机器学习使用 R-Studio（v1.3.1093）进行。我们将第一组数据的80%定义为训练集，将剩余数据的20%定义为测试集。使用随机森林（n\_trees = 1,000）选择特征。为避免过度拟合，组合中蛋白质类型的数量应大大少于样本数量。因此，选择并优化了少于10种蛋白质的生物标记物组合。此外，还将第2组的血浆样本设置为外部验证集。测试集和外部验证集用于评估候选诊断生物标志物。为了评估模型的灵敏度和特异性，采用了10倍交叉验证和10次重复验证。

**Western 印迹**

Western印迹检测按照之前的描述进行（Wang等，2018a；Zhang等，2021b）。简而言之，80μl血浆与20μl SDS上样缓冲液混合，然后按照标准免疫印迹程序进行。用于Western印迹分析的一抗如下：抗SHMT1（稀释 1:1,000，#80715S，CST）、抗G6PD（稀释 1:1,000，25413-1-AP，Proteintech）和抗 MYL9（稀释1:1,000，A8738，ABclonal）。用 Typhoon FLA 9500（GE Healthcare，LittleChalfont，UK）测量化学发光来检测蛋白质丰度。