###### 泌尿生殖系统癌症早筛标志物创新科研项目

**项目申请书**

**项 目 名 称：基于机器学习的血浆代谢组学筛选泌尿生殖系统肿瘤生物标志物的病例对照研究**

**申请医院/科室：中山大学肿瘤防治中心/泌尿外科**

**课题负责人：陈东**

**通 讯 地址：广州市黄埔区中新广州知识城开阳五路1号**

**联 系 电话：**

**电 子 邮箱：**

**申 报 日期： 2024.12.30**

**项目起止时间：2025.1-2027.6**

**申请总金额： 10万元**

此表由北京生命绿洲公益服务中心“泌尿生殖系统癌症早筛标志物创新科研项目” 办公室编制，解释权归项目办公室。

**一、申请医院基本情况：**

|  |  |  |  |
| --- | --- | --- | --- |
| **申请医院** | **中山大学肿瘤防治中心** | **医院级别** | **三甲** |
| **申请科室** | **泌尿外科** | **科室负责人** | **陈东** |
| **职务** | **主任** | **联系方式** |  |
| **邮箱** |  | | |

**二、课题基本情况：**

|  |  |
| --- | --- |
| **课题名称** | 基于机器学习的血浆代谢组学筛选泌尿生殖系统肿瘤生物标志物的病例对照研究 |
| **课题摘要** | **项目主要研究目的：** 1)识别特异性生物标志物 —— 系统性地分析泌尿生殖系统肿瘤患者与健康对照的血浆代谢组，识别出与肿瘤发生密切相关的差异代谢物，这些代谢物有望作为潜在的生物标志物用于临床实践。2)构建预测模型 —— 应用机器学习方法根据已知病例和对照的血浆代谢数据训练和验证分类模型，以期达到高效、准确地识别各类泌尿生殖系统肿瘤患者的目的。3)评估临床应用价值 —— 评估所发现的生物标志物组合及其对应的预测模型在早期诊断、疾病分期、预后判断、治疗反应监测等方面的临床效用，包括但不限于灵敏度、特异性、阳性预测值、阴性预测值、受试者工作特征曲线（ROC曲线）下的面积（AUC）等评价指标。4)探索生物学机制 —— 结合已知的代谢通路知识，探讨鉴别出的生物标志物在肿瘤发生发展中的可能作用机制，为理解泌尿生殖系统肿瘤的病理生理过程提供新见解，并可能为未来治疗策略的制定提供理论依据。研究设计：本研究采用病例对照研究设计。通过临床标准分为病例组、对照组和干扰组。1）病例组：该部分研究对象主要包括已确诊为泌尿生殖系统三大肿瘤（前列腺癌、膀胱癌、肾癌）的患者。研究人员会在他们确诊时或治疗前采集血浆样本，以便分析与肿瘤相关的血浆代谢组变化。2）对照组：对照组则由健康志愿者组成，他们同样提供血浆样本作为比较的基础。对照组的选取是为了排除非肿瘤因素导致的代谢变化，从而确保鉴别出来的代谢标志物具有较高的特异性和灵敏度。通过对比分析病例组和对照组的血浆代谢谱，研究者运用机器学习方法来筛选和验证那些与肿瘤发生、发展密切相关的潜在生物标志物。3）干扰组（良性疾病组）：患有与研究的泌尿生殖系统肿瘤相似症状或可能导致代谢改变的良性疾病患者，如前列腺增生、肾结石、膀胱炎等，旨在开发高度特异性的肿瘤生物标志物，能够有效区分肿瘤与可能的混淆因素（如良性疾病）。诊断标准：经组织病理学检查确诊为前列腺癌/膀胱癌/肾癌主要纳入标准： a. 年龄≥18岁且＜75岁。  b. 患者愿意参与研究并签署知情同意书。  c. 泌尿三癌组：经病理学检查确诊为前列腺癌/膀胱癌/肾癌；  d. 对照组：通过详细的医疗史询问、体格检查、必要的实验室检查（如全血计数、生化分析等）和影像学筛查（如胸部X线或CT），确保参与者无其他慢性疾病或重大健康问题。  e. 干扰组（良性疾病组）：患有与泌尿生殖系统肿瘤症状相似的良性疾病，如良性前列腺增生、肾结石、尿路感染、妇科良性肿瘤等。 |

**三、项目专家组成员：**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 姓名 | 职称 | 学位 | 身份证号 | 单位名称/科室 | 电话 | 承担任务 |
| 1 | 陈东 |  |  |  | 中山大学肿瘤防治中心/泌尿外科 |  |  |
| 2 | 马冰磊 |  |  |  | 中山大学肿瘤防治中心/泌尿外科 |  |  |
| 3 | 龙星博 |  |  |  | 中山大学肿瘤防治中心/泌尿外科 |  |  |
| 4 |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |

**四、项目资金预算表 金额单位：万元人民币**

|  |  |  |
| --- | --- | --- |
| **科 目** | **经费** | **备注（计算依据和说明）** |
| **一、直接费用** | **/** | **/** |
| 设备费 | 1.5 | 购置离心机等分离血浆必备设备 |
| 材料费 | 1.2 | 采血管、冻存管、离心管、吸头、冻存盒等材料费用 |
| 测试化验加工费 | 3 | 1200例受试者血浆非靶向代谢组学检测费用。 |
| 出版/文献/信息传播/知识产权事务 | 1 | 发表文章出版费用 |
| 劳务费 | 1.5 | 需要研究生参与本项目，负责样本处理、样本检测等 |
| 税费 | 0.8 | 税费按照8%计算 |
| 管理费 | 1 | 管理费用按照10%计算 |
|  | / | / |
|  | / | / |
|  | / | / |
| **二、间接费用**  （不超过直接费用扣除设备购置费后的20%核定） | **/** | **/** |
| **总计** | **￥：10（人民币大写：拾万元整 ）** |  |
|  |  |  |
|  |  |  |
|  |  | |

**五、报告正文（中文）**

（内容包括：立项依据、研究内容、研究目标、研究方法、统计方法等）

|  |
| --- |
| **一、立项依据**：  癌症是全球重大公共卫生问题 。泌尿生殖系统癌症一般是指男性和女性泌尿系统和生殖系统男性器官。前列腺癌、膀胱癌和肾癌（简称泌尿三癌）是泌尿生殖系统癌症中最常见的组织类型。在2020年的癌症统计数据中，前列腺癌是男性中第二常见的癌症，并且是全球男性癌症相关死亡原因的第五位，与此同时，膀胱癌和肾癌分别位列全球癌症发病率的第10位和第14位 。预计未来十年，这三种癌症的发病率和死亡率都将上升。泌尿三癌在早期都可能表现为非特异性泌尿系统症状，如血尿、尿频、尿急、尿痛、排尿困难等，这些症状也常见于许多良性泌尿系统疾病，如尿路感染、结石、前列腺增生等。因此，仅凭症状难以准确区分癌症与其他良性疾病。  目前用于泌尿系统肿瘤早期筛查的手段有限且有效性尚不理想。传统的筛查工具如尿液分析、血液生化指标、超声波检查等在检测微小肿瘤或早期病变时敏感性不足。例如，虽然前列腺特异性抗原(PSA)测试可用于前列腺癌筛查，但其假阳性和假阴性率限制了其在早期诊断上的应用。尽管近年来液体活检、代谢组学、基因组学等新技术正在不断发展，但仍需进一步研究以验证其在泌尿系统肿瘤早期诊断中的实际应用价值，而且这些技术的标准化和临床转化也需要时间。  代谢组学作为系统生物学的重要分支，其核心在于全面、系统地分析生物体在不同生理或病理状态下所有代谢产物的动态变化 。在揭示疾病状态下，尤其是癌症发生和发展过程中，代谢组学展现出了独特的优越性 。首先，代谢组学能够捕捉到由基因表达变化经由蛋白质翻译和修饰，最终反映在代谢物水平上的细微变化，这些变化可能是疾病早期的表现，而在基因或蛋白质层面可能尚未出现明显异常。通过对比健康与疾病状态下的代谢谱，科学家们可以发现与疾病进程紧密相关的代谢物生物标志物。其次，代谢组学数据能够直观反映出细胞代谢途径的异常和重塑，揭示癌症发生的代谢机制，有助于从源头理解疾病的发生与发展，并为后续的药物靶点发现和治疗策略提供线索。在癌症早期诊断的应用前景上，代谢组学研究具有极大的潜力。已经有研究表明，通过血清、尿液或组织等样本的代谢组学分析，能够在临床症状出现之前就识别出与癌症相关的代谢异常。Daniele 等 使用液相色谱-四级杆飞行时间质谱 (LC-Q-TOF/MS），对 23 名乳腺癌（ BC） 患者和 35 名口腔健康妇女唾液进行非靶向代谢组学分析，使用t检验、卡方检验等单因素统计方法，对比 METLIN数据库，鉴定出乳腺癌组中有 31 种化合物上调（P<0.05）。其中，患者与健康人群相比，发现 7 种寡肽和 6 种甘油磷脂表达上调，表现出明显的代谢差异，说明唾液代谢物有望区分乳腺癌患者和健康人群，适用于早期诊断。Miyamoto 等用气相色谱-飞行质谱技术分析了非小细胞肺癌和其他肺癌患者的血液样本，标记了437种代谢物，发现非小细胞肺癌患者的麦芽糖、 棕榈酸、甘油、乙醇胺、谷氨酸和乳酸的水平升高，而色氨酸、组氨酸和赖氨酸的水平降低，由此可对二者进行区别。综上所述，代谢组学为揭示疾病状态下机体的代谢变化提供了有力工具，同时在癌症早期诊断方面展现出广阔的应用前景，随着技术的不断提升和大数据分析方法的引入，代谢组学在癌症研究中的地位将进一步加强。  机器学习作为一门重要的人工智能分支，其在处理和分析海量数据，尤其是在模式识别和关联分析方面的卓越能力，已被广泛应用于生物医学研究，尤其是在生物标志物的发现与验证方面。在生物医学研究中，机器学习技术能够高效处理包括基因组学、转录组学、蛋白质组学和代谢组学在内的多组学数据，通过构建和训练复杂的模型，挖掘隐藏在大量生物数据背后的深层次规律和关联性。这种能力使得科研人员能够在纷繁复杂的生物数据海洋中，识别出与疾病发生、发展、预后及治疗响应等密切相关的生物标志物。Roberto等 运用机器学习和多组学数据的综合分析相结合借助 Clarivate CBDD 系统生物学工具，确定了一组对肾上腺皮质癌（ACC）疾病具有高预后价值的生物标志物。另有一项针对食管癌（EC）的研究，开发一种基于风险因素的EC风险预测模型，利用机器学习方法对EC高风险人群进行分层，并在社区层面获得有效的预防目的 。鉴于此，机器学习技术在生物医学领域的应用趋势日益显著，越来越多的科研人员正借助其强大的数据处理能力和模式识别优势，加速生物标志物的发现与验证进程，推动精准医疗和个体化治疗的发展。  正是由于上述种种原因，泌尿系统恶性肿瘤的早期诊断成为了医学界关注的重要课题，研究人员不断努力探索新的生物标志物和检测技术，期望通过更加敏感和特异的早期筛查手段，有效提高肿瘤的早期发现率和治愈率。诸如基于机器学习的血浆代谢组学研究就是一个前沿方向，试图通过分析患者血液样本中的代谢物变化来识别早期肿瘤特征。  **二、研究内容**： 2.1研究设计：本研究采用病例对照研究设计。通过临床标准分为病例组、对照组和干扰组。1）病例组：该部分研究对象主要包括已确诊为泌尿生殖系统三大肿瘤（前列腺癌、膀胱癌、肾癌）的患者。研究人员会在他们确诊时或治疗前采集血浆样本，以便分析与肿瘤相关的血浆代谢组变化。2）对照组：对照组则由健康志愿者组成，他们同样提供血浆样本作为比较的基础。对照组的选取是为了排除非肿瘤因素导致的代谢变化，从而确保鉴别出来的代谢标志物具有较高的特异性和灵敏度。通过对比分析病例组和对照组的血浆代谢谱，研究者运用机器学习方法来筛选和验证那些与肿瘤发生、发展密切相关的潜在生物标志物。3）干扰组（良性疾病组）：患有与研究的泌尿生殖系统肿瘤相似症状或可能导致代谢改变的良性疾病患者，如前列腺增生、肾结石、膀胱炎等，旨在开发高度特异性的肿瘤生物标志物，能够有效区分肿瘤与可能的混淆因素（如良性疾病）。2.2样本量估算基于代谢组学进行的临床研究，充足的样本量对于统计学意义的满足以及实验结果的质量至关重要。代谢组学对于临床的样本建议每组最低50个生物学重复，但要尽可能排除生物学变量（其他疾病干扰或者是性别、年龄、BMI指数等）、过程变量（样本采集、处理、检测环节的人员和仪器操作）以及其他因素变量（地域、季节和气候等）的影响。因此，本研究拟纳入的样本量具体如下：病例组:900例，包含前列腺癌病例组300例，膀胱癌病例组300例，肾癌病例组300例； 干扰组：300例； 对照组：100例。2.3纳入与排除标准：1）病例组纳入标准：a. 年龄≥18岁且＜75岁。b. 经病理学检查确诊为前列腺癌/膀胱癌/肾癌。c. 有足够的临床资料，包括病理报告、影像学检查结果等。d. 具有稳定的全身状况，如心率、血压、血氧饱和度等生命体征正常，血常规、生化、凝血功能等实验室检查结果在可接受范围内。e. 患者愿意参与研究并签署知情同意书。2）病例组排除标准：a. 受试者除前列腺癌/膀胱癌/肾癌外，患有严重影响血浆代谢谱的重大疾病，如严重的肝脏或肾脏疾病、糖尿病未控制、严重感染、艾滋病、自身免疫性疾病或其他并发恶性肿瘤等。b. 接受过针对前列腺癌/膀胱癌/肾癌的任何治疗，包括化疗、放疗、靶向治疗等。c. 伴有其他恶性肿瘤，除非是已经治愈超过5年的皮肤基底细胞癌或鳞状细胞癌。e. 有精神疾病史、药物滥用史、药物依赖史。f. 近3个月内接受过手术者。g. 近3个月内献血或大量出血（＞450mL），或接受输血或者使用血制品者。h. 妊娠试验阳性、哺乳期妇女。i. 研究者认为的其他原因不适合入组的受试者。3）健康对照组纳入标准：a. 年龄≥18岁且＜75岁。b. 通过详细的医疗史询问、体格检查、必要的实验室检查（如全血计数、生化分析等）和影像学筛查（如胸部X线或CT），确保参与者无其他慢性疾病或重大健康问题。c. 患者愿意参与研究并签署知情同意书。d. 无严重的内科疾病或精神疾病，身体状况适合进行代谢组学研究。4）健康对照组排除标准：a. 存在任何类型恶性肿瘤的病史。b. 存在严重感染、营养不良、重度贫血等严重影响代谢组学检测结果的状况。c. 患有影响免疫功能或代谢的慢性疾病，如艾滋病、自身免疫性疾病、糖尿病未得到有效控制等。e. 有精神疾病史、药物滥用史、药物依赖史。f. 近3个月内接受过手术者。g. 近3个月内献血或大量出血（＞450mL），或接受输血或者使用血制品者。h. 妊娠试验阳性、哺乳期妇女。i. 研究者认为的其他原因不适合入组的受试者。5）良性疾病干扰组纳入标准a. 年龄≥18岁且＜75岁。b. 患有与泌尿生殖系统肿瘤症状相似的良性疾病，如良性前列腺增生、肾结石、尿路感染、妇科良性肿瘤等。c. 有足够的临床资料，包括病理报告、影像学检查结果等。d. 具有稳定的全身状况，如心率、血压、血氧饱和度等生命体征正常，血常规、生化、凝血功能等实验室检查结果在可接受范围内。e. 患者愿意参与研究并签署知情同意书。6）良性疾病干扰组排除标准a. 经充分评估受试者患有恶性肿瘤，尤其是泌尿生殖系统恶性肿瘤。b. 存在严重影响代谢或免疫功能的其他重大疾病，如终末期肾病、肝硬化、自身免疫疾病、糖尿病未控制等。d. 有精神疾病史、药物滥用史、药物依赖史。e. 近3个月内接受过手术者。f. 近3个月内献血或大量出血（＞450mL），或接受输血或者使用血制品者。g. 妊娠试验阳性、哺乳期妇女。h. 研究者认为的其他原因不适合入组的受试者。 **2.4研究流程**  1）受试者筛选与知情同意  研究者根据研究方案纳入和排除标准，研究者对初步符合筛选条件的受试者进行知情同意。受试者需要在样本采集前签署经伦理批准的知情同意书，研究者需要告知参加者参加本研究是自愿的、而且可以在研究的任何阶段退出而不受偏见。即使受试者决定不参加本研究项目，他的医疗待遇与权益不会受到任何影响。  2）受试者入组  根据受试者的年龄和性别，参考检验结果，如血液学报告、影像学结果等，经过研究医生的综合判断，认为符合研究方案纳入排除标准，即确定分组。  3）临床信息采集  参试者确定分组后，即可采集研究用的临床信息。信息采集需遵守知情同意的原则，保护受试者隐私。采集的数据包括但不限于：人口统计学、生活方式、疾病史、生化结果、分子水平检测、影像学检查结果、病理学检查结果、癌症分期、合并症情况、用药情况、随访数据等。  4）代谢组学样本采集  血浆样本：采样前一天嘱咐患者清淡饮食，晚8点后禁食，9点后禁水。次日清晨使用EDTA抗凝管采空腹全血3-5 mL，1管。样本低温（0-4 ℃）运输运送至院内实验室，自离体至离心、分装、冻存处理完毕，全程不超过4小时；如果是常温条件下，运输时间应小于30分钟，且样本从离体至离心分装直至冻存，全程不超过1小时。  5）样本分装与保存  a.血浆分离：EDTA抗凝管离心条件：1600 g，12 min，4 ℃。离心后将血浆分装至EP管（120 μL/支，2支）和冻存管（约1mL/支，1支）。液氮罐或-80 ℃冰箱储存。  b.样本剔除标准：血液样本发生溶血；样本离体后常温放置超过30分钟；4℃保存下，离体至冻存超过4小时；血浆分离获得样品总体积＜60μL。  c.样本转运：处理后的样本将转运至国内代谢组学实验室进行检测。转运过程采用干冰运输。  6）样本提取与检测  a.样本提取：经解冻的样本，涡旋30 s混匀，移取样本50 μL到对应编号离心管中，加入700 μL含内标提取液（甲基叔丁基醚：甲醇=3:1，v/v），涡旋混匀，冰浴超声15 min；加入350 μL水：甲醇（3:1，v/v），涡旋混匀，12700 rpm 离心5 min；将350 μL上层有机相转移到对应编号离心管中，后于真空浓缩仪（Speed-Vac）旋干。干燥后的样本加入200 μL乙腈：异丙醇（7:3，v/v）复溶，移取上清液用于LC-MS脂质代谢物分析。将400 μL下层水相转移到对应编号离心管中，加入1100 uL甲醇，涡旋混匀，4 ℃下以12700 rpm离心10 min，转移1000 uL上清液到对应编号Polar管中，并将350 uL下层液体转移到对应编号GC管中，后于真空浓缩仪（Speed-Vac）旋干。干燥后的Polar管样本加入200 μL水复溶，移取上清液用于LC-MS极性代谢物检测。干燥的GC管样本经甲氧铵盐吡啶溶液肟化后，加入N-甲基-N-（三甲基硅烷基）三氟乙酰胺（MSTFA）进行衍生化，用于GC-MS极性代谢物衍生化检测。提取流程图见图1。    图1：样本提取流程  b.样本上机检测：LC-MS脂质代谢物分析：配有ACQUITY UPLC BEH C8色谱柱（1.7um ,100mm × 2.1mm，Waters）的ACQUITY UPLC I-Class液相系统（Waters），流动相为含有10 mmol乙酸铵、0.1%乙酸水（流动相A）和乙腈：异丙醇（7:3，v/v) 含有10 mmol乙酸铵、0.1%乙酸（流动相B）；Orbitrap高分辨质谱（Thermo Fisher）配有电喷雾电离源（ESI），分别采用正负两种电离模式采集数据，一级全扫描（Full scan）分辨率为70,000 FWHM (@200 m/z) ，二级扫描采用ddMS2模式，分辨率35,000 FWHM (@200 m/z) 。  c.LC-MS极性代谢物分析： 配有ACQUITY UPLC HSS T3色谱柱（1.8um ,100mm × 2.1mm，Waters）的ACQUITY UPLC I-Class液相系统（Waters），流动相为含0.1%甲酸水（流动相A）和含0.1%甲酸乙腈（流动相B）；Orbitrap高分辨质谱（Thermo Fisher）配有电喷雾电离源（ESI），分别采用正负两种电离模式采集数据，一级全扫描（Full scan）分辨率为70,000 FWHM (@200 m/z) ，二级扫描采用ddMS2模式，分辨率35,000 FWHM (@200m/z) 。  d.GC-MS极性代谢物衍生化分析：样本经衍生化后，注入到配有Rxi®-5Sil MS色谱柱（Restek 30 m \* 0.25 mm \* 0.25 μm）的气相色谱系统（7890B，Agilent）和飞行时间质谱系统（PEGASUS-BT，Leco）检测 。采用高纯度氦气用作载气，流速为1 mL/min。气相色谱升温程序为：初始温度50 ℃，保持2 min，15 ℃/min连续升温至330 ℃，保持2 min。进样口主要参数为：280 ℃，分流进样（10∶1）。质谱主要参数为：电子电离源（EI），温度250 ℃，电子能量70 eV；扫描范围50~500 amu，检测电压1.2 kV e.质量控制：LC-MS和GC-MS检测样本均采用随机化序列。此外，实验所使用的样本合并为生物学质量控制样品，称为样本池（QC）。生物学质量控制样品的预处理与研究样品的预处理平行且相同。将代谢物混合标准品和质量控制样品均匀地插入到每组运行序列中，以监控分析的稳定性。图示, 示意图  描述已自动生成 研究路线示意图  **三、研究目标：** 1）识别特异性生物标志物：系统性地分析泌尿生殖系统肿瘤患者与健康对照的血浆代谢组，识别出与肿瘤发生密切相关的差异代谢物，这些代谢物有望作为潜在的生物标志物用于临床实践。2）构建预测模型：应用机器学习方法根据已知病例和对照的血浆代谢数据训练和验证分类模型，以期达到高效、准确地识别各类泌尿生殖系统肿瘤患者的目的。3）评估临床应用价值：评估所发现的生物标志物组合及其对应的预测模型在早期诊断、疾病分期、预后判断、治疗反应监测等方面的临床效用，包括但不限于灵敏度、特异性、阳性预测值、阴性预测值、受试者工作特征曲线（ROC曲线）下的面积（AUC）等评价指标。4）探索生物学机制：结合已知的代谢通路知识，探讨鉴别出的生物标志物在肿瘤发生发展中的可能作用机制，为理解泌尿生殖系统肿瘤的病理生理过程提供新见解，并可能为未来治疗策略的制定提供理论依据。综上所述，这项研究旨在利用先进的代谢组学技术和机器学习方法，系统性地揭示泌尿生殖系统肿瘤的代谢特征，开发出具有临床实用价值的生物标志物组合和预测模型，为改善此类肿瘤的早期诊断、精准分类和个体化治疗提供科学依据。 **四、统计方法**： 1）质谱数据处理：使用专业应用程序从质谱下机数据中提取有效峰信号，然后应用基线校正去除噪音，保留原始信号峰值；原始数据转换为中心离散数据；然后，对样本中的每个质谱峰对应色谱中的保留时间进行校正，并将样本间差值在误差范围内峰定义为同一峰，由此得到矩阵数据集。由于在实验过程中存在无法避免的系统性波动（如仪器波动、人员操作产生的波动等），因此，为了使数据分布更反应样本间的真实差异，便于后续分析，进行了缺失填充，均一化，log转化等处理。2）血浆代谢组学数据：代谢物鉴定：通过比对样本中质谱峰与库中化合物的保留时间，荷质比，同位素等特征，对检测物进行定性，化合物涵盖了脂质，氨基酸，糖类，有机酸，核苷酸及其衍生物等。代谢物差异分析：对比泌尿系统癌患者与健康对照组间特定代谢物的浓度变化，计算 fold change、t-test 、OPLS-DA或 Mann-Whitney U test 等统计量，确定显著差异代谢物（通常是 p-value < 0.05 并有生物学意义的 fold change）。3）机器学习模型构建与性能评估：模型选择：如支持向量机（SVM）、随机森林（RF）、逻辑回归（LR）、深度学习模型等。特征选择：基于变量重要性（如 VIP 值、特征重要性得分）、递归特征消除（RFE）、LASSO 等方法筛选关键代谢物标志物。模型性能评价指标：如准确率、敏感性、特异性、受试者工作特征曲线（ROC 曲线）下的面积（AUC）、混淆矩阵、Kappa系数等。4）生物标志物候选物评估：生物标志物组合：分析单个或多个代谢物联合作为标志物的表现，评估其在区分泌尿系统癌与健康对照中的协同效应。稳健性检验：通过 bootstrapping、交叉验证、独立验证等方式验证所选标志物组合的稳定性。5）生物标志物的临床应用价值：诊断效能：计算基于标志物的诊断模型的敏感性、特异性、阳性预测值（PPV）、阴性预测值（NPV）等。 |

**六、研究进度与考核指标**

|  |  |  |
| --- | --- | --- |
| **时间（年、月）** | **研究内容** | **考核指标** |
| 2025.1 -2026.1 | 筛选受试者，采集受试者样本，采集临床信息 | 完成方案既定受试者数量采集 |
| 2026.1 -2026.7 | 血浆非靶向代谢组学检测 | 完成已采集血浆的非靶向代谢组学检测 |
| 2026.7 -2026.12 | 统计分析 | 完成方案既定研究目的统计分析 |
| 2026.12 -2027.6 | 成果转化 | 发表高质量论文，申报发明专利 |

**七、管理条款**

1. 研究医院必须按要求向主办方提交中期进展报告及项目报告，逾期不报，将停拨经费，主办方对报告进行书面审核或汇报审核。

2. 研究医院在课题执行中不可调整研究内容。

**八、申请者承诺**

申请者承诺：

我保证申请书内容的真实性。如果获得资助，我将履行项目负责人职责，严格遵守北京生命绿洲公益服务中心“泌尿生殖系统癌症早筛标志物创新科研项目”的相关规定，按照本申请书所述，切实保证研究工作时间，认真开展工作，按时报送有关材料。若填报失实和违反规定，本人将承担外全部责任。

**申请者（签字）：**

**日期：**

**医院（盖章）：**

**日期：**