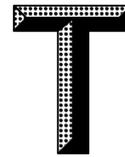


ICS 11.100  
CCS C 10



团 体 标 准

T/CI 1022—2025

# 结直肠癌早筛中多组学液体活检 技术应用指南

Technical application guidelines for multi-omics-based liquid biopsy in  
early screening of colorectal cancer

2025-05-21 发布

2025-05-21 实施

中国国际科技促进会 发布  
中国标准出版社 出版



## 目 次

前言 .....	Ⅲ
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	2
5 样本采集与处理 .....	2
6 多组学高通量测序 .....	4
7 多组学特征集合获取 .....	5
8 肠癌多组学早筛模型构建 .....	7



## 前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由浙江大学医学院附属邵逸夫医院提出。

本文件由中国国际科技促进会归口。

本文件起草单位：浙江大学医学院附属邵逸夫医院、杭州圣庭医疗科技有限公司、中国科学院合肥物质科学研究院健康与医学技术研究所、浙江大学医学院附属第二医院、上海锐序生物科技有限公司、浙江省肿瘤医院、浙江大学医学院附属第一医院、四川大学华西医院、中山大学附属第六医院、温州医科大学附属第一医院、通辽市人民医院。

本文件主要起草人：戴胜、王云飞、谷红仓、丁克峰、李军、王倩、朱玉萍、盛勤松、孟文建、蔡泽荣、李绍堂、张国友。



# 结直肠癌早筛中多组学液体活检 技术应用指南

## 1 范围

本文件提供了基于液体活检的多组学技术(包括片段组学和代谢组学)在结直肠癌早期筛查中的技术要求、操作规范、质量控制等方面的指导。

本文件适用于医疗机构、研究机构和企业开展基于多组学(包括片段组学和代谢组学)液体活检技术的结直肠癌早期筛查服务及相关产品开发。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 38576 人类血液样本采集与处理

GB/T 38736 人类生物样本保藏伦理要求

GB/T 40352.1 人类组织样本采集与处理 第1部分:手术切除组织

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**液体活检 liquid biopsy**

利用人体体液作为标本来源检测获取肿瘤相关信息的技术。

### 3.2

**游离核酸片段组学 circulating nucleic acid fragmentomics**

研究生物体内 cfDNA 片段的断裂位点分布模式及其与个体病理状态关联的研究学科,是基因组学的研究分支。

### 3.3

**代谢组学 metabolomics**

研究生物体内代谢物变化与个体生理病理关联的研究学科,是系统生物学的分支学科。

### 3.4

**CpG 岛 CpG island**

基因组中长度为 300 bp~3 000 bp 的富含 CpG 二核苷酸的 DNA 序列区域。

### 3.5

**RRBS 测序 reduced representation bisulfite sequencing**

通过限制性内切酶富集 CpG 密集区并进行亚硫酸盐处理的靶向甲基化测序技术。

### 3.6

#### 末端基序特征 **end motifs features**

cfDNA 分子双链断裂末端碱基的序列偏好性特征。

### 3.7

#### 断点基序特征 **breakpoint motifs features**

cfDNA 双链断裂位点上下游核苷酸序列的组成规律特征。

## 4 缩略语

下列缩略语适用于本文件。

cfDNA: 循环游离 DNA (cell free DNA)

RRBS: 简并代表性亚硫酸氢盐测序技术 (reduced representation bisulfite sequencing)

## 5 样本采集与处理

### 5.1 样本采集

#### 5.1.1 血液样本的采集

5.1.1.1 人类血液样本的采集应符合 GB/T 38576 和 GB/T 38736 的要求。具体血液样本的采集要求如下:

- a) 采集容器: 使用抗凝采血管 (EDTA 采血管或游离 DNA 采血管) 作为采集容器, 每管采集量为 10 mL;
- b) 采集环境: 样本采集应在清洁、无菌的环境中进行, 避免污染;
- c) 采集时间: 尽量在早晨空腹状态下采集血液样本, 以减少饮食对检测结果的影响;
- d) 采集标签: 所有样本应贴上清晰的标签, 标明患者信息、样本类型、采集日期和时间。

5.1.1.2 血液样本的采集流程如下:

- a) 采血: 使用一次性采血针和采血管, 按照标准操作程序进行采血;
- b) 混匀: 采集后立即轻轻颠倒采血管 8 次~10 次, 使血液与抗凝剂充分混匀;
- c) 记录: 记录采血时间和患者的基本信息。

5.1.1.3 血液样本的运输与储存要求如下:

- a) 运输条件: 样本应在 2℃~8℃ 条件下运输, 避免温度过高或过低;
- b) 储存条件: 采集后的血液样本应在 4℃ 冰箱中短期保存 (不超过 24 h), 长期保存需冷冻 (-80℃)。

#### 5.1.2 组织样本的采集

5.1.2.1 人类组织手术切除样本的采集应符合 GB/T 40352.1 和 GB/T 38736 的要求。组织样本的采集要求如下:

- a) 组织类型和样本量: 采集的组织类型可包括正常组织、肿瘤组织和癌旁组织等, 每份组织样本大小为 1 cm<sup>3</sup>~2 cm<sup>3</sup>;
- b) 采样环境: 样本采集应在无菌环境下进行, 避免污染;
- c) 采样标签: 所有样本应贴上清晰的标签, 标明患者信息、样本类型、采集日期和时间。

5.1.2.2 组织样本的采集流程如下:

- a) 取样: 使用无菌手术器械采集组织样本, 避免污染;
- b) 固定: 将采集的组织样本立即放入含有 RNA 保护剂 (如 RNeasy lysis buffer) 的无菌管中;

- c) 记录:记录取样时间和患者的基本信息。
- 5.1.2.3 组织样本的运输与储存要求如下:
- 运输条件:样本应在 $2^{\circ}\text{C}\sim 8^{\circ}\text{C}$ 条件下运输,避免温度过高或过低;
  - 储存条件:采集后的组织样本应在 $4^{\circ}\text{C}$ 冰箱中短期保存(不超过24 h),长期保存需冷冻( $-80^{\circ}\text{C}$ )。
- 5.2 样本处理
- 5.2.1 血液样本处理
- 5.2.1.1 血浆分离的流程如下。
- 离心:将采集的血液样本先后进行两步离心,首先在 $4^{\circ}\text{C}$ 环境下 $1\ 600\ g$ 低速离心10 min,分离出血浆层;其次将离心转移后的血浆层在 $4^{\circ}\text{C}$ 环境下继续 $16\ 000\ g$ 高速离心10 min,再次分离出血浆层。
  - 转移:将分离出的血浆小心转移到新的无菌管中,避免离心后下层白细胞和血小板的污染。
- 5.2.1.2 血液样本的DNA提取流程如下。
- 试剂盒选择:使用商业化的DNA提取试剂盒。按照试剂盒说明书进行DNA提取;
  - 质量控制:使用Qubit、NanoDrop分光光度计及毛细管电泳仪检测DNA的浓度、纯度和完整性。
- 5.2.1.3 血液样本的代谢物提取流程如下。
- 试剂选择:使用甲醇/水混合溶液提取血浆中的代谢物。
  - 提取步骤:取适量血浆样本于2 mL离心管中,浓缩干燥;加入 $500\ \mu\text{L}$ 甲醇溶液到干燥后的样本管中,涡旋1 min; $12\ 000\ \text{r}/\text{min}$   $4^{\circ}\text{C}$ 离心10 min,取全部上清液,转移至新的2 mL离心管中,浓缩干燥;加入 $150\ \mu\text{L}$  80% 甲醇水配制的2-氯-L-苯丙氨酸( $4\ \text{mg}/\text{L}$ )溶液复溶样品,取上清液过 $0.22\ \mu\text{m}$ 膜过滤,过滤液加入检测瓶中,用于后续检测。
  - 质量控制:取所有实验样本等体积混合,作为质控样本,用于监测仪器稳定性和数据重复性。加入同位素标记内标物(如2-氯-L-苯丙氨酸),用于后续数据归一化。
- 5.2.1.4 从血液样本中提取出的代谢物的保存要求如下:
- 短期保存:提取的DNA和代谢物应在 $4^{\circ}\text{C}$ 冰箱中短期保存(不超过24 h);
  - 长期保存:提取的DNA和代谢物应在 $-80^{\circ}\text{C}$ 条件下长期保存。
- 5.2.2 组织样本处理
- 5.2.2.1 组织样本的DNA提取流程如下。
- 试剂盒选择:使用商业化的DNA提取试剂盒。按照试剂盒说明书进行DNA提取。
  - 质量控制:使用Qubit和琼脂糖凝胶电泳检测DNA的浓度和纯度。
- 5.2.2.2 组织样本的代谢物提取流程如下。
- 试剂选择:使用甲醇/水混合溶液提取组织中的代谢物。
  - 提取步骤:取适量组织样本于2 mL离心管中,浓缩干燥;加入 $500\ \mu\text{L}$ 甲醇溶液到干燥后的样本管中,涡旋1 min; $12\ 000\ \text{r}/\text{min}$   $4^{\circ}\text{C}$ 离心10 min,取全部上清液,转移至新的2 mL离心管中,浓缩干燥;加入 $150\ \mu\text{L}$  80% 甲醇水配制的2-氯-L-苯丙氨酸( $4\ \text{mg}/\text{L}$ )溶液复溶样品,取上清液过 $0.22\ \mu\text{m}$ 膜过滤,过滤液加入检测瓶中,用于后续检测。
  - 质量控制:取所有实验样本等体积混合,作为质控样本,用于监测仪器稳定性和数据重复性。加入同位素标记内标物(如2-氯-L-苯丙氨酸),用于后续数据归一化。
- 5.2.2.3 从组织样本中提取出的代谢物的保存要求如下:
- 短期保存:提取的DNA和代谢物应在 $4^{\circ}\text{C}$ 冰箱中短期保存(不超过24 h);
  - 长期保存:提取的DNA和代谢物应在 $-80^{\circ}\text{C}$ 条件下长期保存。

## 6 多组学高通量测序

### 6.1 RRBS甲基化测序

#### 6.1.1 DNA质量控制流程

DNA质量控制的具体流程如下：

- a) 浓度检测:使用Qubit荧光计检测DNA的浓度,DNA的浓度控制在50 ng/ $\mu$ L~200 ng/ $\mu$ L之间;
- b) 纯度检测:使用NanoDrop分光光度计检测DNA的纯度,A 260/A 280比值应在1.8~2.0之间;
- c) 完整性检测:使用毛细管电泳分析仪检测DNA的完整性,确保没有明显的降解,对于游离DNA需确保没有大片段基因组污染。

#### 6.1.2 DNA片段化流程

DNA片段化的具体流程如下：

- a) 酶切:使用MspI酶对DNA进行切割,产生40 bp~220 bp的片段;
- b) 片段大小选择:使用凝胶电泳或磁珠法选择合适大小的片段。

#### 6.1.3 文库构建流程

##### 6.1.3.1 末端修复的具体流程如下：

- a) 添加A碱基:使用Klenow DNA聚合酶对DNA片段进行末端修复,并添加A碱基;
- b) 连接接头:将带有测序接头的双链DNA连接到末端修复后的DNA片段上。

6.1.3.2 使用亚硫酸氢盐转化试剂盒,按照试剂盒说明书进行DNA转化操作,将未甲基化的胞嘧啶转化为尿嘧啶,而甲基化的胞嘧啶则保持不变。

##### 6.1.3.3 PCR扩增流程如下：

- a) 引物设计:设计带有测序引物的PCR引物;
- b) 扩增条件:使用热循环仪进行PCR扩增,保证扩增效率和特异性。

##### 6.1.3.4 文库纯化流程如下：

- a) 磁珠纯化:使用磁珠对PCR产物进行纯化,去除未结合的接头和引物;
- b) 文库质量控制:使用毛细管电泳分析仪检测文库的质量和大小分布。

#### 6.1.4 测序方法要求

测序方法要求如下：

- a) 平台:使用Illumina HiSeq或NovaSeq平台进行高通量测序;
- b) 测序模式:选择PE 150(双端150 bp)测序模式,确保足够的测序深度。

#### 6.1.5 测序数据产出要求

数据产出要求如下：

- a) 测序深度:确保目标检测区域的平均CpG位点测序深度达到10 X以上;
- b) 数据格式:原始测序数据以FASTQ格式保存,便于后续分析。

### 6.2 代谢组学测序

#### 6.2.1 色谱条件如下。

- a) 仪器配置:使用高效液相色谱系统,使用C18反相色谱柱(2.1 mm×100 mm, 1.8 μm 粒径),流速为0.3 mL/min,柱温为40℃,进样量为2 μL。
  - b) 正离子模式:流动相为0.1%甲酸乙腈(B2)和0.1%甲酸水(A2),梯度洗脱程序为:0 min~1 min, 8%B2; 1 min~8 min, 8%~98%B2; 8 min~10 min, 98%B2; 10 min~10.1 min, 98%~8%B2; 10.1 min~12 min, 8%B2。
  - c) 负离子模式:流动相为乙腈(B3)和5 mmol 甲酸铵水(A3),梯度洗脱程序为:0 min~1 min, 8%B3; 1 min~8 min, 8%~98%B3; 8 min~10 min, 98%B3; 10 min~10.1 min, 98%~8%B3; 10.1 min~12 min, 8%B3。
- 6.2.2 质谱条件如下:
- a) 仪器配置:使用电喷雾离子源(ESI)高分辨率质谱仪,正负离子模式分别采集数据;
  - b) 离子化条件:正离子喷雾电压为3.50 kV,负离子喷雾电压为-2.50 kV,鞘气40 arb,辅助气10 arb;
  - c) 扫描参数:毛细管温度325℃,以分辨率70 000进行一级全扫描,一级离子扫描范围m/z 100~1 000,并采用高能碰撞解离(HCD)进行二级裂解,碰撞能量为30 eV,二级分辨率为17 500,采集信号前10的离子进行碎裂,同时采用动态排除去除无必要的MS/MS信息。
- 6.2.3 数据以MZXML或MZML格式保存(如通过proteowizard软件包中msconvert工具将原始质谱下机文件转换为MZXML文件格式),便于后续分析。

## 7 多组学特征集合获取

### 7.1 甲基化特征集合获取

#### 7.1.1 数据质控流程

数据质控流程如下:

- a) 质量评估:使用FastQC软件对原始测序数据进行质量评估,去除低质量的读段(reads);
- b) 接头去除:使用Trimmomatic或Cutadapt工具去除接头序列。

#### 7.1.2 数据比对流程

比对流程如下。

- a) 比对方法:使用比对工具(如Bismark)将处理后的reads比对到人类参考基因组(hg19)。
- b) 比对质量评估:
  - 1) 比对率:计算比对到参考基因组的读段比例,确保比对率达到60%以上;
  - 2) 唯一比对率:计算唯一比对到参考基因组的读段比例,确保唯一比对率达到60%以上;
  - 3) 未比对率:计算未比对到参考基因组的读段比例,确保未比对率低于10%。

#### 7.1.3 肿瘤特异性 CpG 岛区域鉴定

##### 7.1.3.1 甲基化水平计算方法如下:

将参考基因组按CpG岛注释进行区域划分,基于单倍型解析策略计算各区域甲基化特征。采用五类标准化算法量化单倍型特异性甲基化水平:

- a) MM(mean methylation);
- b) MHL(Methylated haplotype load);
- c) Entropy(Shannon's information entropy);
- d) PDR(Partial Discordance Rate);
- e) CHALM(Cellular Heterogeneity-Adjusted clonal Methylation)。

### 7.1.3.2 差异甲基化分析方法如下。

- a) 采用 Wilcoxon 秩和检验进行肿瘤与正常组织甲基化水平比较,检验条件包括:
  - 1) 独立分析不同单倍型背景;
  - 2) 以 CpG 岛为最小分析单元。
- b) 显著性判定标准:
  - 1) 统计显著性:Benjamini-Hochberg 校正后  $P$  值  $< 0.01$ ;
  - 2) 生物学效应值:  $|\Delta\beta| \geq 0.1$ ;
  - 3) 技术质量控制:要求差异甲基化区域在组内各样本的测序深度均  $\geq 10 \times$ , 差异甲基化区域在组内所有样本覆盖度  $\geq 50\%$ 。

## 7.2 片段组学特征集合获取

### 7.2.1 数据质控

同 7.1.1 规定的血浆 cfDNA 甲基化测序数据质控流程,获取结直肠癌患者组与健康对照组标准化测序数据。

### 7.2.2 末端基序特征获取

同 7.1.2 规定的比对流程,将质控后的读段比对到人类参考基因组上,获取各读段数据在参考基因组上的坐标。提取读段数据 3' 端在参考基因组上的  $p$  个碱基作为碱基片段集合,计算各碱基片段在总片段中的占比,作为末端基序特征。 $p$  代表一个介于 4~10 之间的任意整数。

### 7.2.3 断点基序特征获取

同 7.1.2 规定的比对流程,将质控后的读段比对到人类参考基因组上,提取各读段数据 3' 端在参考基因组上的上下游各  $q$  个碱基的序列,作为碱基片段集合,计算各碱基片段在总片段中的占比,作为断点基序特征。 $q$  代表一个介于 2~5 之间的任意整数。

## 7.3 代谢组学特征集合获取

### 7.3.1 数据质控

#### 7.3.1.1 质量评估

对原始质谱数据进行峰检测、峰过滤和峰对齐处理(例如使用 R XCMS 软件包,参数设置为  $bw=2$ ,  $ppm=15$ ,  $peakwidth=c(5, 30)$ ,  $mzwid=0.015$ ,  $mzdiff=0.01$ ,  $method="centWave"$ ),去除低质量的数据点,得到代谢物定量列表。

#### 7.3.1.2 数据预处理

基于质量控制(QC)样本,使用支持向量回归(SVR)方法进行校正,通过质谱连续扫描进行数据采集,生成基峰色谱图(Base Peak Chromatogram, BPC),以识别潜在的特征峰,基于质量控制(QC)样本,计算这些特征峰的相对标准偏差(RSD),即变异系数(CV),并筛选出 CV 小于 30% 的代谢物。对不符合要求的特征峰应予以删除,仅保留高质量、高稳定性的数据,进行后续分析。

### 7.3.2 代谢物注释

采用公共数据库(如 HMDB、massbank、LipidMaps、mzcloud、KEGG 等谱图数据库)进行物质鉴定,

参数设置为  $\text{ppm} < 30 \text{ ppm}$ 。根据一级质谱中母离子的质荷比( $m/z$ )确定代谢物的分子量,通过质量数偏差( $\text{ppm}$ )以及加合离子等信息进行分子式预测,与数据库进行匹配,实现代谢物的一级鉴定;定量列表中的二级谱图的代谢物与数据库中每个代谢物的碎片离子等信息进行匹配,实现代谢物的二级鉴定。

### 7.3.3 肿瘤特异性代谢物鉴定

#### 7.3.3.1 统计方法

对样本数据进行多维分析,包括 Student  $t$  检验、主成分分析(PCA)、偏最小二乘判别分析(PLS-DA)和正交偏最小二乘判别分析(OPLS-DA),以实现数据的降维处理,得到一级差异代谢物。以一级差异分析结果中统计检验预设的  $P$  值和变量投影重要度(VIP)阈值进行筛选,得到二级差异代谢物。这些分析可利用 R 软件包 Ropls 等工具进行。为验证模型的可靠性,采用置换检验(Permutation Test)进行过拟合检验。

#### 7.3.3.2 阈值设定

通过统计检验计算  $P$  值、OPLS-DA 分析计算变量投影重要度(VIP)以及 fold change 计算组间差异倍数,这些指标共同评估代谢物含量对样本分类判别的影响强度和解释能力。筛选标志代谢物的标准设定为  $P$  值小于 0.05 且 VIP 值大于 1,认为这些代谢物分子具有统计学显著差异。

## 8 肠癌多组学早筛模型构建

整合 7.1.3、7.2.2、7.2.3、7.3.3 获得的甲基化特征、末端基序特征、断点基序特征以及代谢组学特征。将上述提取的特征数据输入到一个集成机器学习分类模型,该模型由两层构成:第一层集成了四种不同的机器学习算法,每种算法针对训练集样本进行独立训练,以识别患癌概率。第二层则通过逻辑回归模型整合第一层的输出,以生成最终的分类决策。模型涉及的算法简要描述及设定的参数如下。

- a) 逻辑回归模型:逻辑回归是一种广泛使用的监督学习模型,适用于处理分类问题。通过建立代价函数,并通过优化方法迭代求解出最优的模型参数,以其计算效率和直观性著称。该方法适用于二分类问题,速度快且易于理解。
- b) 支持向量机(support vector machines, SVM):支持向量机是一种用于分类和回归分析的强大监督学习模型。通过将数据映射到高维空间,并找到一个最优的超平面来最大化不同类别之间的间隔。训练模型时,设定了如下参数:kernel="linear", scale=T, probability=TRUE, type="C-classification"。
- c) 随机森林(random forest, RF):随机森林是一种集成学习方法,通过构建多个决策树并对它们的预测结果进行投票或平均来提高预测准确性。训练模型时,设定了如下参数:method="rf", prox=TRUE, ntree=500, metric="Accuracy"。
- d) XGBoost:XGBoost是一种高效的梯度提升决策树框架,通过优化算法提高机器学习器的性能,特别适用于处理大规模数据集。训练模型时,设定了如下参数:nrounds=1000, params=list(booster="gbtree", objective="binary:logistic", eval\_metric="logloss")。
- e) CatBoost:CatBoost是一种先进的梯度提升决策树算法,特别设计以处理类别型特征并减少模型偏差,通过迭代优化提高预测精度,特别适用于处理类别型特征。训练模型时,设定了如下参数:params=list(loss\_function='Logloss', iterations=1000)。





中国国际科技促进会  
团体标准  
结直肠癌早筛中多组学液体活检  
技术应用指南

T/CI 1022—2025

\*

中国标准出版社出版发行  
北京市朝阳区和平里西街甲2号(100029)

网址 [www.spc.net.cn](http://www.spc.net.cn)

总编室:(010)68533533 发行中心:(010)51780238

读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷  
各地新华书店经销

\*

开本 880×1230 1/16 印张 1 字数 17 千字  
2025年8月第1版 2025年8月第1次印刷

\*

书号:155066·5-16291 定价 38.00 元

如有印装差错 由本社发行中心调换  
版权专有 侵权必究

举报电话:(010)68510107



T/CI 1022-2025