



中华人民共和国国家标准

GB/T XXXXX—XXXX

肿瘤微卫星不稳定性高通量测序分析技术 规程

Technical Guideline for High-Throughput Sequencing Analysis of Tumor
Microsatellite Instability

(点击此处添加与国际标准一致性程度的标识)

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 要求 5

5 分析软件 9

附录 A..... 11

参考文献 12

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。

本文件由国家药品监督管理局提出。

本文件由全国医用高通量测序标准化技术委员会（SAC/TU 007）归口。

本文件起草单位：

本文件主要起草人：

肿瘤微卫星不稳定性高通量测序分析技术规程

1 范围

本文件描述了基于高通量测序的肿瘤微卫星不稳定性分析方法及评价方法。

本文件适用于拟采用高通量测序方法开展肿瘤微卫星不稳定性检测的医学实验室，在开展临床检测前检测流程建立及方法学性能确认过程中使用。也可供体外诊断开发商、制造商，以及相关研究和诊断机构参考使用。

本文件不包括基于PCR和免疫组化蛋白的微卫星不稳定性检测方法。

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 29859-2013 生物信息学术语
- GB/T 30989-2014 高通量基因测序技术规程
- GB/T 35537-2017 高通量基因测序结果评价要求
- GB/T 35890-2018 高通量测序数据系列格式规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

临床测序样本 clinical sequencing sample

为支持疾病、障碍或异常身体状态或其症状的评估、诊断、治疗、缓解或预防而收集或获取的组织、体液、食物或其他物质的样本

3.2

高通量测序 high-throughput sequencing or massively parallel sequencing

能一次并行几十万到几十亿条核酸分子序列测定和一般读长较短等为标志的测序技术。

注1：改写GB/T 35890-2018, 定义3.1。

注2：又叫大规模平行测序。

3.3

人全基因组测序 human whole genome sequencing

对人类不同个体或群体进行全基因组测序。

注：包括人的23对染色体核酸序列、线粒体核酸序列和人样本中含有的微生物序列，但生信分析不包括人样本中

含有的微生物序列。

3.4

人全外显子测序 human whole exome sequencing

对人类不同个体或群体进行全外显子组测序。

3.5

目标区域测序 target region sequencing

根据需要对人类的特定基因或者特定区域进行测序，例如大panel测序。

3.6

微卫星 microsatellite

是一种DNA重复序列，又称简单重复序列或者短串联重复序列，由串联重复单元组成，重复单元长度一般为1-6bp，重复次数一般超过5次。

3.7

模体 motif

微卫星的重复单元序列。

3.8

错配修复系统 mismatch repair system

识别DNA复制或者重组过程中的插入、删除、错配并修复的系统，也可以修复其他形式的DNA损伤。

3.9

微卫星不稳定性 microsatellite instability

由DNA错配修复（Mismatch repair，MMR）系统受损引起的遗传超突变性的分子表型。

3.10

参考基因组 human reference genome

已公开发表的某物种的供参考的全基因组序列，一般为FASTA格式。

3.11

人基因组参考序列版本 human genome reference sequence version

由不同基因组研究团体发布的人参考基因组序列。

3.12

变异 variant

某一测序样本中基因组上某一位置与参考基因组或者其他对照样本的序列差异。

3.13

单碱基突变 single nucleotide variant

DNA上单个碱基位点的替换变异。

3.14

删除 deletion

相对参考基因组发生序列删除的变异。

3.15

插入 insertion

相对参考基因组发生序列插入的变异。

3.16

Indel

相对参考基因组发生序列插入或者删除的变异，一般为小于50bp的变异。

3.17

Read

测序产生的序列片段。

3.18

GC含量 GC content

测序片段碱基中鸟嘌呤和胞嘧啶的数量占所有嘌呤碱基（腺嘌呤和鸟嘌呤）和嘧啶碱基（胸腺嘧啶和胞嘧啶）的百分比。

3.19

数据过滤 data filtering

根据不同的数据过滤参数，对原始测序片段（reads）进行过滤，去除未达到过滤参数要求的测序片段。

注：通常会过滤掉设定参数下的包含低质量，N碱基和接头污染的测序片段。

3.20

序列比对 sequencing alignment

比较两个或两个以上核苷酸或氨基酸序列间相似性的过程。

3.21

有效测序深度 sequencing depth

即一个基因组测序样本经过数据过滤和序列比对之后得到的某一区域或位置的平均测序深度，指某一个区域所有覆盖的测序read的平均值。

3.22

重复测序片段 duplicate reads

测序片段比对到参考序列的位置、方向及碱基序列均一致的重复测序片段（reads）。

3.23

20X 测序覆盖率 coverage rate of sequencing at least 20x

待测样本某一区域中的核苷酸检测结果覆盖参考序列20X及以上深度的百分比。

注1：20X测序覆盖率=20X深度及以上覆盖区域长度÷参考序列总长度*100%

注2：参考序列总长度不包括N碱基。

4 要求

4.1 样本要求

测序样本类型为人基因组DNA。模式可以为肿瘤、肿瘤-正常对照两种。正常对照可以为全血、血液白膜层、唾液、羊水、口腔拭子、癌旁组织。

4.2 实验要求

采样及检测过程所涉及的实验操作应按GB19489、GB/T 30989、SN/T 1193 规定执行。

4.3 测序数据质量要求

测序数据质量应符合 GB/T 37872-2019 和 GB/T 35537-2017 的要求。

4.4 分析软件命名与版本

按《医用软件通用名称命名指导原则》要求进行分析软件命名和版本设定。

按产品简称-应用领域-设计序号顺序命名产品型号。

4.5 分析软件结构组成

微卫星分析软件一般由输入模块、处理模块和输出模块组成。结合自身软件实际特点，对各模块下的子功能模块进行明确说明。分析软件每一个版本应该包含完整的使用说明，包括适用范围、输入、输出、参数、异常处理等。分析软件还应包含用于测试的示例。

5 分析软件

微卫星分析软件包含数据预处理，序列比对，微卫星有效位点检测，微卫星位点稳定性判别，微卫星检测报告生成等内容。

5.1 数据预处理

对测序下机数据进行处理，根据不同测序方案，测序平台，建库方法对测序 read 进行过滤获得高质量测序 read。如果混样测序，需要对数据进行拆分。输出处理后的格式一般为 fastq 格式。

5.2 序列比对

将待测样本分别比对到参考基因组上并根据测序方案进行一定的处理，包括去重，比对质量分析等。如果有 UMI 分析标签，则需对同一个测序分子进行合并矫正。对目标区域或者全基因组范围内的有效测序深度进行评估。如果无法达到设定值，则此分析不可靠。

5.3 微卫星有效位点检测

根据参考基因组上微卫星位点结合测序范围设定检测的微卫星位点，根据序列比对结果确定有效微卫星位点。有效微卫星位点的有效测序覆盖度一般应不低于设定值。

5.4 微卫星位点状态判别

针对不同的样本组合对微卫星有效位点的稳定性进行判别。对于每个位点，应该包括其测序深度，微卫星突变状态等信息。

5.5 微卫星检测报告

根据所检测微卫星位点的状态对样本微卫星状态进行判别，并给出相关报告。一般应该包括样本基本信息，样本组合模式，有效微卫星位点，参考基因组信息等。

参 考 文 献
