

中 华 人 民 共 和 国 国 家 标 准

GB/T 43636—2024

法庭科学 DNA 二代测序检验规范

Forensic sciences—Specifications for next generation
sequencing-based DNA examination

2024-03-15发布

2024-10-01 实施

国家市场监督管理总局
国家标准化管理委员会 发 布

目 次

前言 Ⅲ

1 范围 1

2 规范性引用文件 1

3 术语和定义、缩略语 1

4 总体要求 3

5 原理 3

6 器材和试剂 4

7 检验流程 4

8 数据分析 5

9 遗传参数计算 8

附录A（资料性）法庭科学 DNA 二代测序检验记录表 9

参考文献 17

前 言

本文件按照GB/T1.1—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。

本文件由中华人民共和国公安部提出。

本文件由全国刑事技术标准化技术委员会(SAC/TC 179)归口。

本文件起草单位：公安部鉴定中心、四川大学、中国医学科学院阜外医院、广东省公安厅、中央军委政法委员会侦查技术中心、中国政法大学、北京生物医药研究所。

本文件主要起草人：叶健、季安全、王乐、康克莱、侯一平、周洲、张驰、李海燕、刘开会、张广峰、汤鹏、袁丽、戴文申、赵杰

法庭科学 DNA 二代测序检验规范

1 范围

本文件规定了利用二代测序技术进行法庭科学人类 DNA 遗传标记靶向测序检验的总体要求以及检验器材和试剂、检验流程、数据分析、遗传参数计算的基本要求。

本文件适用于从事法庭科学人类遗传标记检验的实验室和产品制造商针对人类生物检材与样本的 STR、SNP、InDel 等遗传标记以及线粒体 DNA 进行二代测序检验分析。

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 27025 检测和校准实验室能力的通用要求

GB/T 43633 法庭科学 DNA 实验室建设规范

GB/T 43635 法庭科学 DNA 实验室检验规范

3 术语和定义、缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

DNA 测序 DNA sequencing

对 DNA 分子的核苷酸排列顺序的测定，即测定组成核酸分子的腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T) 的排列顺序。

3.1.2

二代测序 next generation sequencing; NGS

区别于传统 Sanger (双脱氧链终止法) 测序，能够一次并行对大量核酸分子进行序列测定的技术。

3.1.3

靶向测序 targeted sequencing

针对特定基因集或基因组区域内已知和新型变异进行的测序。

3.1.4

桥式 PCR bridge polymerase chain reaction

单链文库 DNA 一端序列与芯片表面固定的寡核苷酸特异性结合，当连接片段的另一端弯曲，与芯片表面固定的另一条互补寡核苷酸“形成桥”，以文库DNA 为模板进行扩增的反应。

3.1.5

乳化 PCR emulsion polymerase chain reaction

将用于 PCR 扩增的水相体系与油相体系混合，形成大量油包水的微乳液独立 PCR 扩增空间，在其中扩增得到大量相同来源扩增产物的反应。

3.1.6

DNA纳米球 DNA nanoball;DNB

在二代测序的模板扩增过程中，先将 DNA 进行片段化处理，加接头序列和环化形成单链环状 DNA，然后通过滚环扩增将单链环状 DNA 扩增2~3个数量级最终产生的产物。

3.1.7

单次测序 single run sequencing

测序仪器运行一个测序流程，如一次单向测序或一次双向测序的行为。

[来源：GB/T 30989—2014, 3. 20]

3.1.8

文库 library

通过生物来源的、人工合成的或克隆技术等所得到的一个重建分子群。

注：例如基因组文库、互补 DNA文库、噬菌体展示肽文库等。

[来源：GB/T30989—2014, 3. 5, 有修改]

3.1.9

接头 adapte

用于标记和固定待测序片段的一小段已知序列的DNA 片段。

3.1.10

标签 index

条码 barcode

一段特征性的脱氧核苷酸短片段，在多样本混合检测时，充当识别特定样本来源的唯一标志。

3.1.11

测序通量 sequencing throughput

单次测序可获得序列信息的基因片段数量或可测定的脱氧核糖核酸和核糖核酸(以碱基表示)数量。

[来源：GB/T 30989—2014, 3. 21]

3.1.12

测序读长 read length of sequencing

测序能测得的最长 DNA 片段的长度。

注：通常以碱基数表示。

[来源：GB/T30989—2014, 3. 24, 有修改]

3.1.13

测序深度 depth of sequencing

待测样本中某个指定的核苷酸或某条指定的序列被检测的次数。

[来源： GB/T 30989—2014, 3.31, 有修改]

3.1.14

碱基识别正确率 accuracy of base calling

单次测序正确碱基数占可识别碱基总数的比例。

[来源： GB/T 30989—2014, 3.27, 有修改]

3.1.15

碱基识别错误率 inaccuracy of base calling

单次测序错误碱基数占可识别碱基总数的比例，与碱基识别正确率(3.1.14)相对。

[来源： GB/T 30989—2014, 3.28, 有修改]

3.1.16

碱基识别质量 **quality of base calling**

衡量碱基正确识别的概率。通常以数字值直接表示。

碱基识别质量与碱基识别错误率之间的关系可用式(1)表示：

$$Q = -10 \cdot \lg P \quad \dots\dots\dots (1)$$

式中：

Q——碱基识别质量；

P——碱基识别错误率。

[来源：GB/T 30989—2014, 3.29, 有修改]

3.1.17

分析阈值 **analytical threshold**

能将检测到的信号与基线噪声进行可靠区分的最低测序深度(占比)。

3.2 缩略语

下列缩略语适用于本文件。

DNA: 脱氧核糖核酸(Deoxyribonucleic Acid)

DNB: DNA 纳米球(DNA Nanoball)

InDel: 插入或缺失(Insertion Deletion)

NGS: 二代测序(Next Generation Sequencing)

PCR: 聚合酶链式反应(Polymerase Chain Reaction)

SNP: 单核苷酸多态性(Single Nucleotide Polymorphism)

STR: 短串联重复序列(Short Tandem Repeat)

4 总体要求

4.1 实验室建设应符合GB/T43633 (法庭科学 DNA 实验室建设规范)的要求。

4.2 实验人员应熟悉并掌握基于二代测序进行法庭科学遗传标记检验的原理和方法,并能正确使用分析软件对测序结果进行分析与评价。

4.3 实验室应有完备的文书体系,准确、清晰、完整地记录实验流程和结果,相关格式参见附录A。原始记录、报告等应归档留存,保证其具有可追溯性。

4.4 实验室的环境与设施应符合GB/T 27025的要求。实验室应设置文库构建区、核酸定量区。为了有效防止 DNA 污染,实验室应设置分隔独立的工作区域进行PCR 扩增前操作(试剂配制、样本制备等)和后续 PCR 扩增、测序分析等检验操作,并标识区分不同工作区。检验流程应单向流动,必要时实现人员的单向流动。所有实验操作应在规定区域进行。

5 原理

将生物检材或样本制备成待测样品，通过桥式PCR、乳化 PCR或 DNA 纳米球扩增等 PCR扩增技术，将待测样品量放大，并收集成文库，然后进行平行循环测序。利用边合成边测序的策略，加入一种或多种带标记或不带标记的底物核苷酸，在聚合酶的作用下，按照碱基互补配对原则，进行 DNA 链的延伸，延伸一轮测取一轮底物与模板结合后发出的信号，包括荧光信号、电信号或半导体芯片检测的离子信号等，收集该信号并确定所测位点的碱基信息。当上一轮测序完成后，重复测序过程，即实现大规模并行基因测序。

6 器材和试剂

6.1 通则

6.1.1 对于采用的检验器材和试剂等产品应按照GB/T 27025的要求经方法确认后方可使用。

6.1.2 对于采用的商业化或自行开发的法庭科学遗传标记二代测序检测试剂应进行质量评价验证，包括但不限于准确性、灵敏度、稳定性、均衡性、重复性、检材适应性、种属特异性等。

6.2 器材

主要检验器材包括：

- a) 二代测序仪：碱基识别质量大于20, 碱基识别质量过滤之后的测序准确率大于或等于99.9%, 测序通量适合相应法庭科学遗传标记检验；
- b) PCR 仪；
- c) 荧光定量 PCR 仪；
- d) 离心机；
- e) 低温冰箱；
- f) 移液器。

6.3 试剂

主要检验试剂包括

- a) DNA 提取试剂：用于释放、分离获得各类生物样本中的所有 DNA 并进行纯化；
- b) PCK 复合扩增试剂：用于大规模平行复合扩增针对 STR、SNP、InDel 等遗传标记或线粒体 DNA 的各目标位点靶序列，试剂盒内至少应包含：扩增引物、扩增酶及缓冲液；
- c) 文库构建试剂：用于核酸样品的测序文库构建，试剂盒内至少应包含：接头、DNA 连接酶或 DNA 聚合酶、缓冲液；
- d) 测序试剂：用于对测序文库进行二代测序，试剂盒内至少应包含：测序引物、测序反应酶及缓冲液。

7 检验流程

7.1 通则

检验流程应符合相应试剂说明书和仪器操作规范要求，并进行实验室内部分方法验证，制定相应的作业指导书，内容至少包括：

- a) 检验步骤：至少包括 DNA 提取与定量、测序文库构建、测序文库纯化与定量、测序检验等；

- b) 质量控制：至少包括 DNA 质量和浓度控制、文库片段长度和浓度控制、测序原始数据质量控制、测序结果质量评价等。

7.2 DNA 提取与纯化

按照GB/T43635（法庭科学 DNA 实验室检验规范）的方法开展实验，并对 DNA 提取的质量和浓度进行质量控制。宜采用荧光定量方法对 DNA 浓度进行定量，应对符合文库构建的质量要求指标进行方法验证。如果核酸样品质量不符合要求，应重新提取或增加样品量再进行相应操作。

7.3 目标序列扩增

将适量的 DNA 加至引物混合液和 PCR 预混体系中进行 PCR 扩增，扩增后选择合适的方式对 PCR 产物进行纯化。扩增反应体系及循环参数参照相应试剂说明书操作。实验应设置阳性对照和阴性对照。

7.4 文库构建

主要步骤包括添加接头、文库纯化、文库质量控制和文库均一化。材料和方法以文库构建试剂说明书为准，其他注意事项包括：

- a) 清晰标记添加的标签接头，防止标签之间造成污染；
- b) 标签接头连接后采用磁珠法或相关文库纯化试剂对扩增产物进行纯化处理；
- c) 文库质量控制宜采用荧光定量方法检测文库浓度，可选择采用生物分析仪检测文库片段大小，实验室应采用方法验证确定符合测序要求的文库质量要求指标；
- d) 符合质量要求的文库进行均一化和混合，均一化浓度以测序试剂和测序仪器要求为准。

7.5 测序模板制备

测序文库均一化、混合后，宜采用桥式 PCR、乳化PCR 或 DNA 纳米球扩增等方法进行信号放大处理，使信号强度满足测序识别要求。材料和方法以测序模板制备试剂说明书为准。

7.6 测序

测序即加入底物核苷酸，在测序酶的作用下使底物核苷酸与测序文库中的待测样品进行结合，同时收集该过程中产生的信号，包括但不限于荧光信号、电信号或离子信号。材料和方法参照相应二代测序试剂说明书，其他注意事项包括：

- a) 根据采用的检测试剂和二代测序平台摸索确定合适的文库上机浓度；
- b) 根据检测的法庭科学遗传标记特性选择合适的测序读长以保证测序结果准确和完整；
- c) 根据不同测序平台的特性和检测样本量选择合适测序通量的测序试剂、确定合适的测序深度并进行充分验证，保证测序结果准确。

8 数据分析

8.1 通则

8.1.1 软件的分析流程开放

应选用公开数据分析流程、算法和数据分析参数的生物信息学软件，进行法庭科学遗传标记二代测序数据分析，确保二代测序实验结果准确、可溯源。

8.1.2 测序原始数据质量控制

8.1.2.1 原始数据质量评估

通过测序质量值量化评估原始数据质量，宜评估的指标包括碱基质量、序列质量、序列长度、N(未测到或不确定碱基)的比例、接头含量等。

8.1.2.2 数据过滤

根据评估结果过滤去除原始数据中的接头、低质量或未检出的碱基及序列，宜过滤的内容包括(以

下过滤参数应根据检测不同的法庭科学遗传标记或使用不同的二代测序平台进行调整，并经过实验室验证）：

- a) 检测接头序列并进行剪切；
- b) 去除含一定数目N碱基的序列；
- c) 设定低质量碱基的判定阈值，去除含一定比例以上低质量碱基的序列；
- d) 设定滑动窗口，去除序列平均碱基值低于阈值的序列；
- e) 去除剪切后低于一定长度的序列；
- f) 其他经过验证的剪切去除序列方式。

8.1.3 数据比对分析

采用适合的软件分析数据质量控制后的测序数据，对目标法庭科学遗传标记进行分型并统计测序深度。

8.2 STR分型

8.2.1 二代测序 STR 数据分析软件应支持将STR 作为序列多态遗传标记进行分型，即长度相同而序列不同的 STR 等位基因可被识别为不同的等位基因。序列多态分型结果命名应与现有长度多态命名兼容，并具备序列唯一性。

8.2.2 以 GRCh38 作为比对参考序列，以其正向序列作为 STR 序列比对依据。

8.2.3 测序读长应足够覆盖 STR 扩增子长度，不应通过序列拼接的方式进行STR 分型，应舍去未测通扩增子的数据。

8.3 SNP分型

8.3.1 以 GRCh38 作为比对参考序列。

8.3.2 采用dbSNP 数据库中ID 号码(rs#) 的方式进行 SNP 位点命名。

8.4 InDel分型

8.4.1 以 GRCh38 作为比对参考序列。

8.4.2 插入等位基因以序列信息标示，缺失等位基因用“-”进行标示。

8.5 线粒体DNA 测序

以人类线粒体 DNA 参考序列(NCBI Reference Sequence:NC_012920.1)作为比对参考序列，进行序列比对、拼接和定位，标出与之不同的碱基作为特征点。

8.6 测序深度阈值设定与使用

8.6.1 各基因座/位点设定合适的测序深度(占比)作为分型研判的分析阈值。若序列的测序深度占比(在该基因座/位点总测序深度中的比例)低于分析阈值时研判为噪声序列可被过滤；若序列的测序深度占比高于或等于分析阈值时应进一步研判序列的类型。

8.6.2 实验室应根据自行检测数据设定分析阈值中的测序深度占比。阈值设定方法参照表1、表2和图1的示例。针对已知分型的单一来源样本，统计某位点的噪声序列最高测序深度占该位点总测序深度的比例达1% (位点总测序深度为5000, 噪声序列最高测序深度为50, 见表1)。检测不少于50例样本，获得每例样本某位点的噪声序列最高测序深度占该位点总测序深度的比例 (见图1), 在图1的示例中可设定测序深度占比1.5%作为该位点的阈值比例。对于 STR 基因座，应先确定等位基因和影子峰，再统计噪声序列的最高测序深度占比，在表2的示例样本中噪声序列的最高测序深度占比1.8%。

表 1 已知分型样本的某SNP 位点的序列情况

序列	已知类型	SNP分型	测序深度	测序深度占比
TGAATGCCATCCGTATCACCTGTTGAAGG TTCACC	等位基因	T	4940	4940/5000=98.8%
TGAATGCCATCCGTATCACCTGTAGAAGG TTCACC	噪声序列	A	50	50/5000=1%
TGAATGCCATCCGTATCACCTGTCGAAGG TTCACC	噪声序列	C	10	10/5000=0.2%

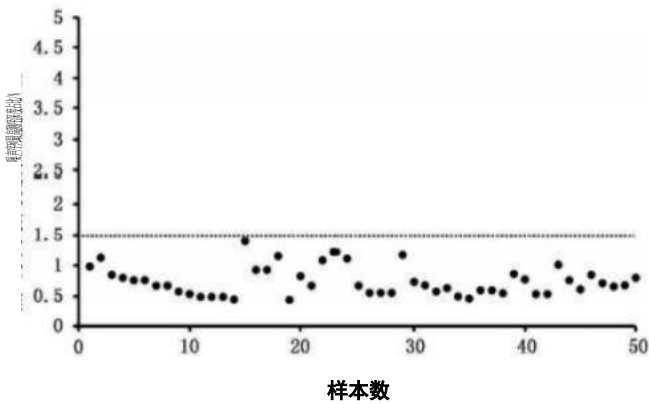


图 1 各样本中某位点的噪声序列最高测序深度占比情况

表 2 已知分型样本的某 STR 基因座的序列情况

长度	序列	已知类型	测序深度	测序深度占比
12	[ATCT]12	等位基因	2500	2500/5000=50%
10	[ATCT]10	等位基因	2100	2100/5000=42%
11	[ATCT]i	影子峰	150	150/5000=3%
9	[ATCT]。	影子峰	130	130/5000=2.6%
11	[ATCT]sAACT[ATCT]s	噪声序列	90	90/5000=1.8%
12	AACT[ATCT]	噪声序列	25	25/5000=0.5%
10	AACA[ATCT],	噪声序列	5	5/5000=0.1%

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/478123143135006101>