



中华人民共和国公共安全行业标准

GA/T 1693—2020

法庭科学 DNA 二代测序检验规范

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

2020-10-30 发布

2021-01-01 实施

中华人民共和国公安部 发布

目 次

前言 I

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 总则 3

6 原理 4

7 检验器材和试剂 4

8 检验流程 4

9 数据分析 5

10 遗传参数计算..... 7

附录 A（资料性附录） 法庭科学 DNA 二代测序检验记录表 8

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

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这些专利的责任。

本标准由全国刑事技术标准化技术委员会法医检验分技术委员会(SAC/TC 179/SC 6)提出并归口。

本标准起草单位:公安部物证鉴定中心、四川大学、广东省公安厅。

本标准主要起草人:叶健、季安全、王乐、康克莱、侯一平、张驰、李海燕、刘开会、张广峰。

法庭科学 DNA 二代测序检验规范

1 范围

本标准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术语和定义、原理、检验器材和试剂、检验流程,以及数据分析、遗传参数计算应遵守的基本要求。

本标准适用于针对各类法医学物证检材的 STR、SNP、InDel 等遗传标记以及线粒体 DNA 进行二代测序检验分析,为法庭科学实践提供遗传学信息,适用于所有从事法庭科学检验的 DNA 实验室和产品制造商。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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—2019 检测和校准实验室能力的通用要求

GA/T 383—2014 法庭科学 DNA 实验室检验规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

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。

3.2

二代测序 second generation sequencing

区别于传统 Sanger(双脱氧法)测序,能够一次并行对大量核酸分子进行序列测定的技术,通常一次测序反应能产生不低于 100 Mb 的测序数据。

3.3

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测序,有效降低测序成本,提高测序深度,更为有效地研究特定区域的遗传变异。

3.4

桥式 PCR bridge polymerase chain reaction

文库两端的 DNA 序列与芯片表面固定的引物序列特异性结合,以文库 DNA 为模板,进行桥形扩增。通过不断变性和扩增的循环过程使得每个 DNA 片段在各自的位置上复制集中成束,生成 DNA 簇。每个 DNA 簇包含成千上万单克隆扩增子,以每个 DNA 单链为模板,逐个合成 DNA 互补链,将碱基信号强度放大,进而达到测序所需信号的强度。

3.5

乳化 PCR emulsion polymerase chain reaction

将用于 PCR 反应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 反应空间,如果 PCR 反应体系中加入的核酸模板数量合适,每个独立反应空间中将只有一个模板分子。在形成的独立反应空间中进行相同模板的平行扩增,即可得到大量相同的扩增结果,实现基因扩增。

3.6

DNA 纳米球 DNA nanoball

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 个~3 个数量级最终产生的产物。

3.7

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。

[GB/T 30989—2014,定义 3.20]

3.8

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,如基因组文库、互补 DNA 文库、噬菌体展示肽文库等。

[GB/T 30989—2014,定义 3.5]

3.9

接头 adapter

用于标记和固定待测序片段的一小段已知序列的 DNA 片段。

3.10

标签或条码 index;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。

3.11

测序通量 throughput of gene sequencing

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。

[GB/T 30989—2014,定义 3.21]

3.12

测序读长 read length of gene sequencing

测序能测得的最长 DNA 片段的长度,通常以碱基数表示。

[GB/T 30989—2014,定义 3.24]

3.13

测序深度 depth of sequencing

待测样本中某个指定的核苷酸被检测的次数。

[GB/T 30989—2014,定义 3.31]

注:对于 STR 的二代测序,指待测样本中某条指定的序列被检测的次数。

3.14

测序准确率 accuracy of gene sequencing

对已知序列的基因进行测序,获得的正确碱基数占可识别的碱基数比例。

[GB/T 30989—2014,定义 3.25]

3.15

碱基识别正确率 accuracy of base calling

单次测序正确碱基数占碱基总数的比例。

[GB/T 30989—2014, 定义 3.27]

3.16

碱基识别错误率 inaccuracy of base calling

单次测序错误碱基数占碱基总数的比例,与**碱基识别正确率**(3.15)相对。

[GB/T 30989—2014, 定义 3.28]

3.17

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:

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

式中:

Q —— 碱基识别质量;

P —— 碱基识别错误率。

[GB/T 30989—2014, 定义 3.29]

3.18

等位基因覆盖度比 allele coverage ratio

基因座分型为杂合子时,两个等位基因的测序深度的较低值与较高值的比值($0 < ACR \leq 1$),常用于评估基因座内测序信号的均衡性。

3.19

有效序列 effective read

测序结果中,所有能比对到参考基因组的目标区域序列数占有所有测序序列数的比例。

4 缩略语

下列缩略语适用于本文件。

ACR:等位基因覆盖度比(Alelle Coverage Ratio)

DNA:脱氧核糖核酸(Deoxyribonucleic Acid)

DNB:DNA 纳米球(DNA Nanoball)

EPCR:乳化 PCR(Emulsion Polymerase Chain Reaction)

InDel:插入缺失多态性(Insertion Deletion Polymorphism)

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)

5 总则

5.1 鉴定人应具有法医物证鉴定执业资格,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。

5.2 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。

5.3 实验室的环境与设施应符合 GB/T 27025—2019 的规定。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,各区间不能直通。所有实验操作应在规定区域进行。

6 原理

将生物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,将待测样品放大收集成库,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理,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。

7 检验器材和试剂

7.1 通则

7.1.1 对于采用的检验器材和试剂等产品需按照 GB/T 27025—2019 的要求经方法确认后方可使用于日常检案。

7.1.2 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需进行质量评价验证,包括但不限于准确性、灵敏度、稳定性、一致性、均衡性、种属特异性及遗传标记的群体遗传学参数调查。

7.2 器材

主要检验器材包括:

- a) 二代测序仪,基本参数要求包括:碱基识别质量大于 20, Q 值过滤之后的准确率大于或等于 99.9%;
- b) PCR 仪;
- c) 荧光定量 PCR 仪;
- d) 离心机;
- e) 低温冰箱;
- f) 移液器。

7.3 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) PCR 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、高效扩增酶及缓冲液;
- 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、高效测序反应酶及缓冲液。

8 检验流程

8.1 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、文库浓度控制、测序原始数据质量控制、测序结果质量评价等。

8.2 DNA 提取与纯化

按照 GA/T 383—2014 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 浓度进行定量并记录。

8.3 目标序列扩增

将适量的 DNA 加至引物混合液和 PCR 反应的预混体系中进行 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以相应试剂说明书为准。实验需设置阳性对照和阴性对照。

8.4 文库构建

主要步骤包括添加接头、文库纯化、文库质量控制和标准化文库。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验文库浓度,可选择采用生物分析仪检测文库片段大小;
- 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。

8.5 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。材料和方法以模板制备试剂说明书为准。

8.6 测序

测序反应即加入底物核苷酸,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。材料和方法以二代测序试剂说明书为准,其他注意事项包括:

- a) 实验室应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读长以保证测序结果的准确性和完整性;
- c) 根据不同测序平台的特性和检测样本量确定合适的测序覆盖度和深度以保证测序结果的准确性,并进行充分验证。

9 数据分析

9.1 通则

9.1.1 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。

9.1.2 测序原始数据质量控制

9.1.2.1 原始数据质量评估

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。

9.1.2.2 数据过滤

针对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):

- 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。

9.1.3 数据比对分析

采用二代测序平台推荐的商业化分析软件或者其他适合的分析软件,对过滤后的测序数据中的候选法庭科学遗传标记进行分型。统计测序通量、测序深度、遗传标记的等位基因覆盖度比、有效序列比例等测序结果质量评价指标。

9.2 STR 分型

9.2.1 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。

9.2.2 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。

9.2.3 测序读长要足够覆盖 STR 扩增子长度,不可通过序列拼接的方式进行 STR 分型,未测通扩增子的数据要舍去。

9.2.4 STR 分型可针对不同基因座设定合适的测序深度作为分析阈值。

9.3 SNP 分型

9.3.1 以 GRCh38 作为比对参考序列。

9.3.2 SNP 位点的测序深度需大于或等于 100。

9.3.3 采用 dbSNP 中 ID 号码(rs#)的方式进行 SNP 位点命名。

9.4 InDel 分型

9.4.1 以 GRCh38 作为比对参考序列。

9.4.2 InDel 位点的测序深度需大于或等于 100。

9.4.3 插入等位基因以序列信息标示,缺失等位基因用“—”进行标示。

9.5 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,数据测序深度需大于或等于 100。

10 遗传参数计算

10.1 将 STR 作为序列多态性遗传标记进行法庭科学参数和亲权指数计算,即长度相同而序列不同的 STR 等位基因是独立的不同等位基因,所采用的人群频率数据应是序列多态 STR 等位基因频率统计数据。

10.2 在进行法庭科学参数和亲权指数计算过程中,应充分考虑所检遗传标记间连锁关系的影响。

附录 A

(资料性附录)

法庭科学 DNA 二代测序检验记录表

法庭科学 DNA 二代测序检验记录表见表 A.1~表 A.6。

表 A.1 扩增模板 DNA 定量结果登记表

×××× DNA 实验室

控制编号：

扩增模板 DNA 定量结果登记表

操作人：_____

扩增模板 DNA 定量时间：自__年__月__日__时__分至__年__月__日__时__分(24 小时制)

定量仪器：☐Qubit 荧光计 ☐实时荧光定量 PCR 仪 ☐其他_____

仪器编号：_____

定量试剂：☐试剂一 ☐试剂二 ☐其他_____

试剂 LOT 号：_____

样本编号	浓度	样本编号	浓度	样本编号	浓度

浓度单位默认为 ng/ μ L，若使用其他浓度单位需填写在表格中。

第__页/共__页

表 A.2 扩增作业单

×××× DNA 实验室

控制编号：

DNA 检验记录——扩增作业单

操作人：_____

扩增时间：自__年__月__日__时__分至__年__月__日__时__分(24 小时制)

扩增试剂：☐试剂一 ☐试剂二 ☐试剂三 ☐试剂四 ☐其他：_____

试剂 LOT 号：_____

扩增仪器编号：_____ 扩增体系：____ μL 循环数：____

序号	位置	样本 编号	模板	序号	位置	样本 编号	模板	序号	位置	样本 编号	模板	序号	位置	样本 编号	模板
1	A1			25	A4			49	A7			73	A10		
2	B1			26	B4			50	B7			74	B10		
3	C1			27	C4			51	C7			75	C10		
4	D1			28	D4			52	D7			76	D10		
5	E1			29	E4			53	E7			77	E10		
6	F1			30	F4			54	F7			78	F10		
7	G1			31	G4			55	G7			79	G10		
8	H1			32	H4			56	H7			80	H10		
9	A2			33	A5			57	A8			81	A11		
10	B2			34	B5			58	B8			82	B11		
11	C2			35	C5			59	C8			83	C11		
12	D2			36	D5			60	D8			84	D11		
13	E2			37	E5			61	E8			85	E11		
14	F2			38	F5			62	F8			86	F11		
15	G2			39	G5			63	G8			87	G11		
16	H2			40	H5			64	H8			88	H11		
17	A3			41	A6			65	A9			89	A12		
18	B3			42	B6			66	B9			90	B12		
19	C3			43	C6			67	C9			91	C12		
20	D3			44	D6			68	D9			92	D12		
21	E3			45	E6			69	E9			93	E12		
22	F3			46	F6			70	F9			94	F12		
23	G3			47	G6			71	G9			95	G12		
24	H3			48	H6			72	H9			96	H12		

模板一栏单位默认为 ng,若使用其他单位需填写在表格中。

第__页/共__页

表 A.3 文库构建作业单

×××× DNA 实验室

控制编号：

DNA 检验记录——文库构建作业单(测序平台)

操作人：

建库时间：自 年 月 日 时 分 至 年 月 日 时 分(24 小时制)

建库方法：☐手动 ☐ 仪器编号：

建库试剂：☐建库试剂一 ☐建库试剂二 ☐其他

试剂 LOT 号：

标签试剂：☐标签试剂一 ☐其他

试剂 LOT 号：

纯化试剂：☐纯化试剂一 ☐其他

试剂 LOT 号：

样本编号	标签号	样本编号	标签号	样本编号	标签号

第 页/共 页

表 A.4 文库定量结果登记表

×××× DNA 实验室

控制编号：

文库定量结果登记表

操作人：

文库定量时间：自 年 月 日 时 分 至 年 月 日 时 分 (24 小时制)

定量仪器：☐Qubit 荧光计 ☐实时荧光定量 PCR 仪 ☐其他

仪器编号：

定量试剂：☐试剂一 ☐试剂二 ☐其他

试剂 LOT 号：

样本编号	浓度	样本编号	浓度	样本编号	浓度

浓度单位默认为 ng/μL,若使用其他浓度单位需填写在表格中。

第 页/共 页

表 A.5 测序作业单

×××× DNA 实验室

控制编号：

DNA 检验记录——测序作业单*（__测序平台）

操作人：_____

模板制备时间：自__年__月__日__时__分至__年__月__日__时__分（24 小时制）

测序芯片：☐芯片一 ☐芯片二 ☐芯片三 ☐芯片四 ☐芯片五 芯片编号：_____

模板制备方法：☐手动 ☐工作站_____ 仪器编号：_____

模版制备试剂：☐试剂一 试剂 LOT 号：_____

☐试剂二 试剂 LOT 号：_____

测序时间：自__年__月__日__时__分至__年__月__日__时__分（24 小时制）

测序仪器：☐仪器一 ☐仪器二 ☐仪器三 仪器编号：_____

测序试剂：☐试剂一 试剂 LOT 号：_____

☐试剂二 试剂 LOT 号：_____

☐试剂三 试剂 LOT 号：_____

测序参数：①DNA 文库上样浓度：_____ ②Flows：_____ ③测序读长_____

样本编号	标签号	样本编号	标签号	样本编号	标签号

第__页/共__页

* 若模版制备由测序仪器内部完成,则不填相应内容。

表 A.6 DNA 分型记录单

×××× DNA 实验室

控制编号：

DNA 分型记录——序列多态 STR 分型研判记录单

样本编号：_____

未获得 STR 分型：_____；

获得混合 STR 分型：_____；

获得单一 STR 分型：_____；

研判依据：_____

基因座	分型	序列信息

注 1：“研判依据”栏内填写分型所依据的原始测序结果报告，以文件名表示。

注 2：本记录所示混合 STR 分型数据系检验人员的初步研判，最终分型以原始测序结果报告为准。

签名/日期：_____

第__页/共__页

表 A.6 (续)

×××× DNA 实验室

控制编号：

DNA 分型记录——SNP 分型* 研判记录单

样本编号：_____

未获得 SNP 分型：_____；

无法明确判读 SNP 分型：_____；

获得单一 SNP 分型：_____；

研判依据：_____

位点序号	分型	备注	位点序号	分型	备注	位点序号	分型	备注

注 1：“研判依据”栏内填写分型所依据的原始测序结果报告，以文件名表示。

注 2：本记录所示无法明确判读 SNP 分型数据系检验人员的初步研判，最终分型以原始测序结果报告为准。

签名/日期：_____

第__页/共__页

* InDel 分型研判记录单可参照该表格式。

中华人民共和国公共安全
行 业 标 准
法庭科学 DNA 二代测序检验规范
GA/T 1693—2020

*

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)
北京市西城区三里河北街16号(100045)

网址: www.spc.org.cn

服务热线: 400-168-0010

2021年4月第一版

*

书号: 155066 · 2-35884

版权专有 侵权必究



GA/T 1693-2020



码上扫一扫 正版服务到