

中华人民共和国司法行政行业标准

SF/T 0070—2020

染色体遗传标记高通量测序与法医学应用
规范

Specification for high-throughput sequencing of chromosomal genetic markers and
its forensic applications

2020 - 05 - 29 发布

2020 - 05 - 29 实施

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 缩略语	4
5 总体要求	4
6 检验程序	5
7 结果分析	7
8 结果应用	8
参考文献	9

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。

本标准由司法鉴定科学研究院提出。

本标准由司法部信息中心归口。

本标准起草单位：司法鉴定科学研究院、四川大学、中山大学、复旦大学、河北医科大学。

本标准主要起草人：李成涛、张素华、侯一平、孙宏钰、谢建辉、李淑瑾。

染色体遗传标记高通量测序与法医学应用规范

1 范围

本标准规定了染色体遗传标记高通量测序的总体要求、检验程序、结果分析和结果应用。

本标准适用于人体血液（斑）、口腔拭子（唾液斑）、精液（斑）、带毛囊毛发、羊水和组织块等检材的染色体遗传标记高通量测序与法医学应用。

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 6682 分析实验室用水规格和试验方法

GB/T 27025 检测和校准实验室能力的通用要求

GB/T 37223—2018 亲权鉴定技术规范

SF/Z JD0105011—2018 法医学STR基因座命名规范

SF/Z JD0105012—2018 个体识别技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

基因座 locus

基因在染色体上所占的位置。

3.2

等位基因 allele

位于同源染色体的相同位置上有差异的DNA片段互称为等位基因。

3.3

基因型 genotype

个体在一个或者多个基因座上等位基因的组合。

3.4

杂合子 heterozygote

二倍体中同源染色体同一位点上的两个等位基因不相同的基因型个体。

3.5

纯合子 homozygote

二倍体中同源染色体同一位点上的两个等位基因相同的基因型个体。

3.6

微变异等位基因/中间等位基因 microvariant allele/ intermediate allele

STR基因座上表现为含有不完整重复单位的等位基因。

3.7

文库 library

通过生物来源、人工合成或克隆技术等获取的重建分子群。

注：如基因组文库、互补DNA文库、噬菌体展示文库等。

3.8

短串联重复序列 short tandem repeat, STR

人类基因组中由2~6个碱基对作为核心单位，串联重复所形成的一类多态性遗传标记。

3.9

单核苷酸多态性 single nucleotide polymorphism, SNP

人类基因组中同一位置由单个核苷酸的变异所形成的一类多态性遗传标记。

3.10

插入缺失多态性 insertion deletion polymorphism, InDel

人类基因组中插入或缺失不同长度DNA片段所形成的一类多态性遗传标记。

3.11

引物 primer

在DNA复制过程中，能与模板链互补并作为复制延伸的起始点、具有一定长度和碱基顺序的寡核苷酸链。

3.12

DNA聚合酶 DNA polymerase

以DNA单链为模板，基于碱基互补配对原则，催化底物脱氧核糖核苷三磷酸分子生成双链DNA时所需要的酶。

3.13

聚合酶链反应 polymerase chain reaction, PCR

模板DNA先经高温变性为两条单链，在适宜的温度下，成对的引物分别与两条模板DNA单链上的一段互补序列发生退火，随后在DNA聚合酶作用下，以四种dNTP（dATP、dGTP、dTTP、dCTP）为底物，催化磷酸二酯键形成，使引物得以延伸，如此反复变性、退火和延伸这一循环，使位于两段靶序列之间的DNA片段呈几何倍数扩增。

3.14

乳液聚合酶链反应 emulsion polymerase chain reaction, ePCR

乳化 PCR

按照一定的混合比例，将用于PCR反应的水相体系与油相体系混合，产生大量油包水的微乳液滴，每个微乳液滴形成独立PCR反应空间；如果PCR反应体系中加入的核酸模板数量合适，每个微乳液滴内将只含有一个模板分子；然后，每个微乳液滴进行平行扩增，即可同时得到大量模板分子的扩增产物。

3.15

桥式聚合酶链反应 bridge polymerase chain reaction

桥式PCR bridge PCR

连接在芯片上的DNA单链的3'末端序列与邻近的连接在芯片上的寡核苷酸链以碱基互补的形式进行结合；然后，在PCR反应过程中，以该寡核苷酸链作为引物进行扩增生成新的单链。如此反复该过程，可以获得含有单分子的簇。

3.16

高通量测序 high-throughput sequencing

大规模平行测序 massively parallel sequencing, MPS

下一代测序 next generation sequencing, NGS

二代测序 second generation sequencing, SGS

能同时对大量DNA分子进行序列测定的技术。

注：通常一次测序反应能产生不低于100Mb的测序数据。

3.17

接头 adapter

一段人工合成的长度较短的能与平末端或者黏性末端匹配（连接）的DNA片段。

3.18

碱基识别 base calling

测序过程中荧光信号或其它测序反应产生的信号转换为碱基序列信息的过程。

3.19

测序通量 sequencing throughput

单次测序可获得DNA序列信息的数据量。

注：通常以碱基数表示。

3.20

测序读长 read length of gene sequencing

测序能测得的DNA片段的长度。

注：以碱基数表示。

3.21

测序深度 sequencing depth

待测遗传标记进行高通量测序获得的总碱基数与遗传标记大小的比值。

注：通常以次数（×）表示。

3.22

碱基识别正确率 accuracy of base calling

单次测序正确的碱基数占测序获得的总碱基数比例。

3.23

碱基识别错误率 inaccuracy of base calling

单次测序错误的碱基数占测序获得的总碱基数比例。

3.24

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公式（1）表示：

$$Q = -10 \log P \dots\dots\dots (1)$$

式中：

Q——碱基识别质量；

P——碱基识别错误率。

3.25

等位基因测序深度比 allele coverage (count) ratio, ACR

位于常染色体和X染色体上的基因座为杂合子时，两个等位基因的测序深度的百分比值。

注：常用来评估遗传标记经高通量测序后杂合子的两个等位基因被检测的均衡性。

4 缩略语

下列缩略语适用于本文件。

CPI 累积亲权指数 (Combined Parentage Index)

DNA 脱氧核糖核酸 (Deoxyribonucleic Acid)

LR 似然率 (Likelihood Ratio)

PI 亲权指数 (Parentage Index)

5 总体要求

5.1 人员

鉴定人应具有法医物证鉴定执业资格，熟悉并掌握基于高通量测序进行染色体遗传标记检测的原理和方法，并能使用分析软件对测序结果进行正确的分析与评价。

5.2 检测步骤

使用高通量测序仪进行染色体遗传标记检测的步骤如下：

- a) 采样；
- b) DNA 提取和定量；
- c) 构建测序文库（遗传标记的靶向富集、连接接头、文库扩增与文库质控）；
- d) 高通量测序（选择合适的文库上机浓度、测序模板的制备与测序）；
- e) 结果分析与结果应用。

测序活动完毕后，应将各个环节的记录进行分类归档。

注：本标准中染色体遗传标记指分布于人类基因组常染色体和性染色体上的STR、SNP以及InDel遗传标记。

5.3 实验室设施与环境

实验室的设施与环境除应符合GB/T 27025的规定外，还包括：

- a) 实验室宜根据处理步骤设置分隔独立的工作区域，并有明显的标识，各区间不能直通；
- b) 各区之间如果是紧密相连，应安装物品传递舱；
- c) 所有实验操作宜在规定区域进行。

6 检验程序

6.1 基本要求

6.1.1 实验室应对采用的高通量测序仪撰写作业指导书，并组织实验人员进行培训与考核。

6.1.2 实验室应对采用的商品化或自行开发的染色体遗传标记高通量检测试剂盒进行法医学验证，包括但不限于准确性、灵敏度、稳定性、种属特异性验证及遗传标记的群体遗传学参数调查。实验室也可采用已公开发表的文献或出版物中提供的法医学验证结果。

6.1.3 实验室应根据采用的商品化或自行开发的染色体遗传标记高通量检测试剂盒及相应的高通量测序仪进行内部方法验证，并制定相应的作业指导书，内容包括但不限于：

- a) 实验方法：包括但不限于 DNA 提取与定量方法、测序文库构建方法、文库定量与质控方法、文库均一化方法等；其中，实验对照样本包括但不限于试剂空白对照、阳性对照及阴性对照；
- b) 高通量测序结果评估方法：确定测序信息评估参数阈值，应评估的参数包括但不限于单次测序反应的总测序数、单个样本的总测序数、正向链和反向链的测序均衡性等；
- c) 染色体遗传标记测序结果评估方法：对染色体遗传标记高通量测序结果建立合理的分析阈值，分析阈值应设置为测序信息可以与背景噪音信号可靠地进行区分的最小阈值。其中，不同的遗传标记测序效能可能有较大差异，需要对过高或过低测序信号的染色体遗传标记建立单独的分析方法。

6.1.4 实验室应使用商品化或自行开发的软件对染色体遗传标记高通量测序结果进行分析，分析软件应提供用户使用手册，详细描述分析方法。

示例：针对 STR 遗传标记，分析方法描述中应提供 STR 基因座核心重复序列的起始位置、终止位置和重复序列结构等。

6.1.5 实验室应对所采用的软件进行准确性与科学性的验证，确保高通量测序产生的染色体遗传标记分型结果与基于其它类型技术产生的分型结果具有一致性。

6.2 采样

样本的采集、包装及保存的要求包括：

- a) 样本可为人体的血液（斑）或口腔拭子（唾液斑），也可为人体其它生物学样本；

示例：精液（斑）、带毛囊毛发、羊水和组织块等。

- b) 样本应分别包装，进行唯一性标识并注明采样日期；
- c) 样本采集后应干燥、冷藏或冷冻保存。

6.3 DNA 提取和定量

6.3.1 DNA 提取

DNA提取要求包括：

- a) DNA 提取应主要采用物理、化学或生物酶的方法，释放样本中的基因组 DNA，进一步对其进行纯化，充分去除样本中的蛋白质、脂类等 PCR 反应抑制物以及提取过程中加入的试剂；
- b) 宜采用相关商品化 DNA 提取试剂盒进行待测样本 DNA 的提取。

6.3.2 DNA 定量

DNA定量要求包括：

- a) 提取后 DNA 的质量和浓度应能满足下游实验要求；
- b) 宜采用与检材状况相适应的定量方法进行定量。

6.4 测序文库构建

6.4.1 构建步骤

测序文库构建的步骤分为遗传标记的靶向富集和文库制备，其中文库制备应包括以下步骤：

- a) 连接接头；
- b) 文库扩增；
- c) 文库质控。

建库过程中应设置阴性对照与阳性对照。

6.4.2 遗传标记的靶向富集

本标准中高通量测序对象是染色体遗传标记。目前主要采用PCR技术靶向富集包含目标检测遗传标记的DNA片段。具体的PCR体系及反应条件应参照所采用的试剂盒说明书。

6.4.3 文库制备

6.4.3.1 连接接头

应采用合适的方法对6.4.2中靶向富集的PCR产物两端加上适用于相应高通量测序仪的接头(含有与扩增及测序引物互补的序列)，制备成两端已知的特定接头和中间未知的DNA片段组成的DNA分子的集合，即待测文库。

6.4.3.2 文库扩增

文库扩增的主要要求如下：

- a) 宜以 6.4.3.1 中待测文库中的 DNA 序列为模板进行单分子扩增，获得用于高通量测序的测序文库；
- b) 单分子扩增的方法宜采用桥式 PCR 或乳液 PCR；
- c) 扩增试剂盒内至少应包含扩增引物、高效扩增酶缓冲液；
- d) 单分子扩增产物应能满足所选用高通量测序仪厂家的质量要求。

6.4.3.3 文库质控

可采用实时荧光定量PCR或微流体技术检验文库浓度与片段大小。质检标准要求包括：

- a) 文库终浓度应 $\geq 2\text{nmol/L}$ ，体积 $\geq 10\mu\text{L}$ ；
- b) 文库片段大小应与预期相符，且没有游离的接头序列和引物二聚体。

6.5 高通量测序

6.5.1 选择合适的文库上机浓度

文库上机浓度的选择应注意：

- a) 实验人员应根据所采用的染色体遗传标记高通量检测试剂盒和高通量测序仪建立合适的文库上机浓度。浓度过低，会导致仪器测序能力的浪费；浓度过高，会超过测序仪器的分辨率，从而降低测序数据的质量甚至导致测序失败；
- b) 对于不同类型甚至同一类型的不同单机，即使实验人员选择相同的上机浓度，成簇密度也可能存在差异。实验人员应根据实践经验来修正某一台特定仪器的上机浓度，不宜照搬其它类型高通量测序仪的上机浓度数值。

6.5.2 测序模板制备

测序模板制备的步骤及要求如下：

- a) 文库变性；
- b) 将变性后的文库稀释至合适的上机浓度。

6.5.3 测序

测序要求包括：

- a) 除另有规定外，试剂为生化试剂。实验用水应符合 GB/T 6682 中一级水的要求；
- b) 用于对测序文库进行高通量测序的试剂盒应至少包含测序引物、高效测序反应液、缓冲液；
- c) 按要求准备好测序试剂和测序芯片，文库及试剂装载操作应按照高通量测序仪系统界面提示进行。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本 DNA 模板进行结合，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；
- d) 测序完毕，应进行仪器清洗。

7 结果分析

7.1 高通量测序信号的分析

高通量测序仪输出序列信息后，统计测序通量和测序读长，并对测序过程中加入的阳性对照进行分析，将测序得到的原始序列比对到人类参考基因组序列（Hg19或GRCh38），统计相关运行参数。运行参数要求包括但不限于：

- a) 测序通量 $\geq 100\text{Mb}$ ；
- b) 碱基识别正确率 $\geq 99\%$ ；
- c) 碱基识别质量 > 20 。

7.2 基于高通量测序的 STR 基因座的分型结果

对经高通量测序后STR基因座的分型结果分析包括但不限于以下步骤：

- a) 采用软件对 STR 基因座的高通量测序信息进行基因型分型分析，其中分型结果的命名应符合 SF/Z JD0105011—2018 的要求；
- b) 对于微变异等位基因或者未在实验室检见及文献中报道的等位基因应仔细核对序列信息，宜进行再次测序及不同测序方法的验证。同时，实验人员应考虑软件是否对该等位基因进行了正确的命名；
- c) 当 STR 基因座的测序深度 $<100\times$ 或等位基因测序深度比（ACR） $<60\%$ 时，基因座的分型结果应谨慎采纳。

7.3 基于高通量测序的 SNP 位点的分型结果

对经高通量测序后 SNP 位点的分型结果分析包括但不限于：

- a) 采用软件对 SNP 位点的高通量测序信息进行基因型分型分析，其中分型结果的命名应指出：
 - 1) 位点名称及染色体编号；
 - 2) 等位基因名称（等位基因以 A、C、G、T 表示，等位基因间以“/”分隔）；
 - 3) 参考基因组版本号。
- b) 当 SNP 位点的测序深度 $<50\times$ 或等位基因测序深度比（ACR） $<60\%$ 时，位点的分型结果应谨慎采纳。

7.4 基于高通量测序的 InDel 位点的分型结果

对经高通量测序后 InDel 位点的分型结果分析包括但不限于：

- a) 采用软件对 InDel 位点的高通量测序信息进行基因型分型分析，其中分型结果的命名宜指出：
 - 1) 位点名称及染色体编号；
 - 2) 等位基因名称（插入等位基因以序列信息标示，缺失等位基因用“-”进行标示，等位基因间以“/”分隔）；
 - 3) 参考基因组版本号。
- b) 当 InDel 位点的测序深度 $<100\times$ 或等位基因测序深度比（ACR） $<60\%$ 时，位点的分型结果应谨慎采纳。

8 结果应用

8.1 应用于亲权鉴定或个体识别

应按照 GB/T 37223—2018 中的规定对染色体遗传标记高通量测序分型结果进行 PI 值及 CPI 值的计算，并将其结果应用于二联体亲权鉴定或三联体亲权鉴定。

应按照 SF/Z JD0105012—2018 中的规定对染色体遗传标记高通量测序分型结果进行 LR 值的计算，并将其结果应用于个体识别。

8.2 其它用途

可根据已公开发表的文献或出版物中相关方法的描述将染色体遗传标记高通量测序信息及分型结果应用于复杂亲缘关系鉴定或法医 DNA 数据库建设等领域。

参 考 文 献

[1] Parson W, Ballard D, Budowle B, et al. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements [J]. *Forensic Sci Int Genet*, 2016, 22: 54-63.

[2] Scientific Working Group on DNA Analysis Methods Addendum to "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" to Address Next Generation Sequencing.

[3] Sims D, Sudbery I, Ilott NE, et al. Sequencing depth and coverage: key considerations in genomic analyses [J]. *Nat Rev Genet*. 2014, 15(2):121-32.

[4] Churchill JD, Novroski NMM, King JL, et al. Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System [J]. *Forensic Sci Int Genet*. 2017, 30:81-92.
