



中华人民共和国国家标准

GB/T 37870—2019

个体鉴定的高通量测序方法

Individual identification by high-throughput sequencing method

2019-08-30 发布

2019-08-30 实施

国家市场监督管理总局
中国国家标准化管理委员会 发布

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由国家标准物质研究中心提出并归口。

本标准起草单位：深圳华大生命科学研究院(原深圳华大基因研究院)、中国计量科学研究院、深圳华大智造科技有限公司、深圳华大基因科技有限公司、深圳基因产学研资联盟。

本标准主要起草人：谢一帆、王晶、谢伟伟、陈芳、蒋慧、杜佳婷、牛春艳、李陶莎、刘心、程奇、李倩一、李岱怡、谢强。

引 言

个体鉴定在恐怖事件、重大灾害、交通事故的急救医学和侦查破案中有较大的适用性。本标准是结合核酸生物技术和高通量测序技术以检测人体样本的 STR 位点和 SNP 位点进行个体鉴定,对于维护社会稳定有极其重要的意义。

个体鉴定的高通量测序方法

1 范围

本标准规定了个体鉴定高通量测序方法的术语和定义、原理、试验条件、仪器与设备、试剂、试验步骤和结果分析。

本标准适用于人体的血液及血痕、毛发、唾液斑等样本的个体鉴定。

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 29859—2013 生物信息学术语

GB/T 30989—2014 高通量基因测序技术规程

3 术语和定义

GB/T 29859—2013 及 GB/T 30989—2014 界定的以及下列术语和定义适用于本文件。为了便于使用,以下重复列出了 GB/T 29859—2013 及 GB/T 30989—2014 中的某些术语和定义。

3.1

测序 sequencing

测定氨基酸或核苷酸序列的过程。

[GB/T 29859—2013,定义 2.4.13]

3.2

高通量基因测序 high-throughput gene 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平行序列测定的技术,通常一次测序反应能产生不低于 100 Mb 的测序数据。

[GB/T 30989—2014,定义 3.19]

4 缩略语

下列缩略语适用于本文件。

DNA:脱氧核糖核酸(Deoxyribonucleic Acid)

dNTP:脱氧核糖核苷三磷酸(deoxy-ribonucleotide triphosphate)

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

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

STR:短串联重复序列(Short Tandem Repeat)

5 原理

采集样本后,结合核酸生物技术和高通量测序技术,检测特异的分子遗传标记位点(STR 位点和

SNP 位点),通过比较测试样本与待检者的分型结果是否均相同,鉴定测试样本是否来源于待检者。

若测试样本与待检者的分子遗传标记位点分型均相同,则判断为匹配。

若测试样本与待检者有两个及以上位点分型结果不同,在排除等位基因丢失等前提下,判断为不匹配。

若测试样本与待检者只有一个位点分型结果不同,不能排除测试样本来源于待检者,需要增加更多特异的分子遗传标记位点进行检测。增加检测更多位点后,测试样本与待检者有两个及以上位点分型结果不同,判断为不匹配。

测试样本与待检者匹配的原因有两种:一是测试样本可能来源于待检者;二是测试样本可能来源于人群中的随机个体,仅仅是分子遗传标记位点分型碰巧相同。因此,需要通过计算匹配概率和似然率来进行个体鉴定。

匹配概率是人群中的随机个体纯粹由于机会与待检者分型结果一致的概率,以此估计一个理论上的随机个体碰巧匹配的可能性。似然率在数值上等于样本来源于待检者的概率与样本来源于随机个体的概率比值。

6 试验条件

实验室应做到防止污染,根据功能划分为试剂储存和准备区、样本制备区、扩增区、扩增产物分析区和其他区域。实验室人员应具备良好的分子生物学专业技术操作能力。

7 仪器与设备

7.1 PCR 仪:温度设置范围为 0 °C~99 °C,最大循环数为 99。

7.2 高通量测序仪:测序通量 ≥ 100 Mb,碱基识别质量 > 20 ,碱基识别正确率 $\geq 99.9\%$ 。

7.3 离心机:实验室通用高速离心机,最高转速不低于 16 000 r/min。

7.4 涡旋振荡仪。

7.5 冰箱:温度调节范围为 -20 °C ~ 4 °C。

8 试剂

8.1 试验用水

一级水,电导率(25 °C) ≤ 0.01 mS/m,吸光度(254 nm,1 cm 光程) ≤ 0.001 ,可溶性硅(以 SiO₂ 计)含量 ≤ 0.01 mg/L。

8.2 DNA 提取试剂

8.2.1 Chelex-100 试剂

Chelex-100、TNE 缓冲液、蛋白酶 K、十二烷基硫酸钠(SDS)、二硫苏糖醇(DTT)。

注:Chelex 为聚苯乙烯二乙烯基苯树脂。

8.2.2 试剂盒

基因组 DNA 提取试剂盒。

8.3 PCR 扩增试剂

个体鉴定的引物混合液及 PCR 反应的预混体系。PCR 反应的预混体系包括 DNA 聚合酶、PCR

缓冲液、镁离子、dNTP、PCR 稳定剂和增强剂。

8.4 建库试剂

核酸样品的基因文库构建试剂包含 DNA 聚合酶、PCR 缓冲液。

8.5 平行扩增试剂

构建好的基因文库平行扩增试剂包含扩增引物、高效扩增酶缓冲液。

8.6 高通量测序试剂

高通量测序试剂包含测序引物、高效测序反应酶、缓冲液。

9 试验步骤

9.1 选择分子遗传标记位点

选取高多态性的 STR 位点和杂合度大于 0.3 的 SNP 位点,作为候选的分子遗传标记位点,并进行引物设计与合成(参见附录 A)。

当多个 SNP 位点同时使用时,或 SNP 位点与 STR 位点同时使用时,需有证据表明 SNP 位点两两之间是相互独立的,或 SNP 位点与 STR 位点两两之间是相互独立的。当不能证明相互独立时按多个位点组合成的单倍型进行匹配概率和似然率的计算。

9.2 DNA 提取

9.2.1 总则

采用 Chelex 法或试剂盒法进行 DNA 提取。提取后的 DNA 浓度应大于 1 ng/ μ L,应置于 $-20\text{ }^{\circ}\text{C}$ 或 $-20\text{ }^{\circ}\text{C}$ 以下条件保存。

9.2.2 Chelex 法

当样本为血液和血痕、毛发、唾液斑时可采用 Chelex 法。

- a) 血液和血痕:取少量血液或血痕加入适量纯水,剧烈振荡,室温下放置 30 min 以上。13 000 r/min 离心 3 min 后去上清液,沉淀中加入 200 μ L 5% Chelex-100 溶液(使用前要充分振摇,使 Chelex-100 颗粒悬浮)和 10 μ L 5 mg/mL 的蛋白酶 K,在振荡器上反复振荡,放入 56 $^{\circ}\text{C}$ 保温 30 min 以上。取出后振荡,100 $^{\circ}\text{C}$ 保温 8 min,13 000 r/min 离心 3 min 后,上清液用于 PCR 扩增或置 4 $^{\circ}\text{C}$ 冰箱内低温保存备用。
- b) 带毛囊的毛发:剪取毛根部分 5 mm~10 mm,无水乙醇、水、无水乙醇各冲洗一次,晾干后加入 30 μ L 5% Chelex-100 溶液和 2 μ L 5 mg/mL 的蛋白酶 K,56 $^{\circ}\text{C}$ 消化至完全溶解。取出后振荡,100 $^{\circ}\text{C}$ 保温 8 min,13 000 r/min 离心 3 min 后,上清液用于 PCR 扩增或置 4 $^{\circ}\text{C}$ 冰箱内低温保存备用。
- c) 唾液斑:剪取约 0.5 cm^2 的唾液斑加入适量纯水,剧烈振荡,在室温下放置 15 min。13 000 r/min 离心 3 min 后去上清液,沉淀中加入 200 μ L 5% Chelex-100 溶液(使用前要充分振摇,使 Chelex-100 颗粒悬浮)和 10 μ L 5 mg/mL 的蛋白酶 K,在振荡器上反复振荡,放入 56 $^{\circ}\text{C}$ 保温 30 min 以上。取出后振荡,100 $^{\circ}\text{C}$ 保温 8 min,13 000 r/min 离心 3 min 后,上清液用于 PCR 扩增或置 4 $^{\circ}\text{C}$ 冰箱内低温保存备用。

9.2.3 试剂盒法

当样本为血液和唾液时,可采用试剂盒法。按照 DNA 提取试剂盒说明书进行 DNA 提取。

9.3 文库制备与 DNA 测序

将候选分子遗传标记位点的引物混合液和 PCR 反应的预混体系加至适量的 DNA 中进行 PCR 扩增。扩增完成后选择合适的磁珠对 PCR 产物进行纯化。纯化后的 PCR 产物浓度应符合高通量测序仪的文库构建要求。取纯化后的 PCR 产物按照高通量测序仪说明书进行文库构建和文库质检。当文库的片段长度分布符合预期,且文库浓度符合高通量测序仪测序要求时,文库质检合格。最后将多个质检合格的文库按比例混合后,选择合适的测序读长进行高通量测序。

9.4 位点分型

对高通量测序数据进行过滤,去除带有高通量测序接头和测序质量值较低的测序序列。根据测序数据的文件格式和测序读长选择比对软件并设定比对参数,将过滤后的测序数据比对到人类参考基因组上。取出比对到人类参考基因组唯一位置的测序序列,并保留其比对信息。选择适用于高通量测序数据的 STR 分型软件和 SNP 分型软件,对候选的分子遗传标记位点进行分型。

STR 的分型结果包含样本编号、STR 位点及基因型,基因型以重复单元的个数表示,未得到分型或无法明确判定分型的记为 NA(缺失值)。

SNP 的分型结果包含样本编号、SNP 位点和基因型,基因型以碱基 A、C、G、T 表示,未得到分型或无法明确判定分型的记为 NA。

10 结果分析

按式(1)计算匹配概率,从理论上估计测试样本来自群体中的随机个体却因碰巧而与待检者发生匹配的可能性。以频率估计概率,群体中发现这种基因型的频率即是随机个体碰巧匹配概率。随机匹配概率越小,测试样本来自群体中随机个体的可能性就越小,说明测试样本与待检者匹配可能不是随机发生的,测试样本更可能来自待检者。

匹配概率的计算:

$$\Pr(E | H_d) = 1 \times P(X) \dots\dots\dots(1)$$

式中:

$\Pr(E | H_d)$ ——在 H_d 条件下获得这种基因型的概率;

E ——获得这种基因型的期望值;

H_d ——测试样本来自群体中的随机个体;

$P(X)$ ——群体中这种基因型的频率。

按式(2)计算似然率,判断在某种基因型组合下,测试样本更可能来自待检者,还是更可能来自人群中的随机个体。当似然率在数值上超过 1 时,从统计学上认为测试样本更可能来自待检者。反之,如果小于 1,测试样本更可能来自群体中的随机个体。当似然率在数值上超过全球人口总数时,认为测试样本来自待检者。因为从概率上估计几乎不可能在世界上找到具有同样基因型组合的另一个人。

似然率的计算:

$$LR = \Pr(E | H_p) / \Pr(E | H_d) \dots\dots\dots(2)$$

式中:

LR ——似然率;

$\Pr(E | H_p)$ ——在 H_p 条件下获得某种基因型组合的概率;

- E —— 获得某种基因型组合的期望值；
- H_p —— 测试样本来自待检者；
- $\Pr(E|H_d)$ —— 在 H_d 条件下获得某种基因型组合的概率；
- H_d —— 测试样本来自群体中的随机个体。

附 录 A
(资料性附录)
分子遗传标记位点

A.1 单核苷酸多态性位点

单核苷酸多态性位点参见表 A.1。

表 A.1 单核苷酸多态性位点

SNP 位点			
rs2235907	rs1657695	rs3805392	rs7429010
rs8124995	rs2292564	rs1657741	rs4478233
rs228104	rs2013162	rs6909306	rs2172651
rs5749426	rs1997660	rs62431284	rs325238
rs4076086	rs14134	rs562381	rs7715674
rs1106201	rs26821	rs3094868	rs1337823
rs3756050	rs7690296	rs929310	rs574202
rs11123823	rs5745448	rs1512612	rs7741536
rs2274212	rs11239930	rs1355634	rs4719491
rs3829868	rs10801520	rs3817211	rs1343469
rs2276967	rs3899750	rs6499422	rs472728
rs9821880	rs11714239	rs2356027	rs1698647
rs1049500	rs2293195		

A.2 短串联重复序列位点

短串联重复序列位点参见表 A.2。

表 A.2 短串联重复序列位点

STP 位点			
CSF1PO	D7S820	D18S51	TH01
D2S1338	D8S1179	D19S433	TPOX
D3S1358	D13S317	D21S11	vWA
D5S818	D16S539	FGA	AMEL

中 华 人 民 共 和 国
国 家 标 准
个 体 鉴 定 的 高 通 量 测 序 方 法

GB/T 37870—2019

*

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

2019年7月第一版

*

书号: 155066 · 1-62712

版权专有 侵权必究



GB/T 37870-2019