

DB4403

深 圳 市 地 方 标 准

DB4403/T 126—2020

基因身份证技术规程

Technical Regulation of DNA ID

2020-11-23 发布

2020-12-01 实施

深圳市市场监督管理局 发布

目 次

前言	2
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 缩略语	5
5 基本要求	5
6 基因组遗传标记选取类型及要求	6
7 制作流程	7
附录 A（资料性） DNA 位点	10
附录 B（资料性） 基因身份证信息采集单	12
参考文献	13

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》的规则起草。

本文件由深圳市发展和改革委员会归口。

本文件起草单位：深圳华大法医科技有限公司、深圳华大基因股份有限公司，深圳华大基因科技有限公司。

本文件主要起草人：沈悦生、王秋娟、阳明霞、郑飞雪、胡雪松、吴昊、叶明芝、姜华艳、王洪琦、尹焯。

基因身份证技术规程

1 范围

本文件规定了基因身份证制作的基本要求、基因组遗传标记选取类型及要求、制作流程等内容。本文件适用于深圳市范围内基因身份证制作。

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 19489 实验室生物安全通用要求
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 37870—2019 个体鉴定的高通量测序方法
GB 50346 生物安全实验室建筑技术规范
GA/T 461—2019 居民身份证制证用数字相片技术要求
GA/T 1012—2019 居民身份证指纹采集和比对技术规范
SZDB/Z 124—2015 基于二代测序技术的HLA高分辨分型检测标准

3 术语和定义

GB/T 30989—2014界定的以及下列术语和定义适用于本文件。

3.1

基因身份证 DNA ID

DNA 身份证 DNA ID

由个人基本信息标识符和部分基因组信息组成。可甄别个体、种族、血型等，并为识别个人身份提供证据。

3.2

ABO 血型基因 ABO gene

位于9号染色体长臂，由7个外显子和6个内含子组成，基因全长约18 kb，其中7个外显子全长约1065 bp，大小从28 bp到688 bp不等。大多数临床表型及基因分型亚型多态性位于6号外显子。

3.3

ABO 血型基因分型 ABO genotyping

DB4403/T 126—2020

从基因序列水平对ABO血型基因进行分型，其结果精确到血清学水平后的亚型。

3.4

Rh 血型基因 Rh gene

位于1号染色体短臂，包括RHD、RHCE和小膜蛋白1基因（SMP1），RHD和RHCE基因3'端相对，相隔约30000 bp并且高度同源。

3.5

Rh 血型基因分型 Rh genotyping

从基因序列水平对Rh血型基因进行分型，可区分免疫原性最强的RHD基因全缺失、部分D、弱D、DEL型，其结果精确到血清学水平后的亚型。

3.6

HLA 高分辨分型 HLA high resolution genotyping

从基因序列水平对HLA基因进行分型，其结果精确到血清学水平后的亚型。

3.7

种族溯源 ancestry trace

通过人群基因与基因型频率存在明显群体差异的一组SNP位点，分析某一人群的遗传成分构成，或推断某一个体的群体来源及其成分构成。

3.8

祖先信息遗传标记 ancestry informative markers, AIMS

在不同人群中等位基因存在较大差异的遗传位点，也称人群特异标记。

3.9

探针捕获 probe capture

通过设计长度一般为90~120 nt的捕获探针群，在杂交系统中特异性结合经过酶切或超声打碎的DNA片段的过程。

3.10

主成分分析 principal component analysis, PCA

将分散在一组变量上的信息，集中到某几个综合指标（主成分）上的一种统计分析方法。

[来源：GB/T 29859—2013，定义2.6.5]

3.11

DNA 分型 DNA genotyping

利用生物学检测方法测定个体DNA序列，并将其与其他个体的DNA序列或参考DNA序列进行比对，以确定该个体与其他个体的遗传组成（基因型）差异的过程。

3.12

测序片段 reads

高通量测序平台产生的含有碱基序列和质量值的序列片段。

[来源：GB/T 35890—2018，定义3.1]

3.13

序列比对 sequence alignment

比较两个或两个以上核酸序列间的相似性的过程。

[来源：GB/T 29859—2013，定义2.2.1]

3.14

参考序列 reference genome sequence

测序片段对应的物种基因组序列。

[来源：GB/T 35890—2018，定义3.11]

4 缩略语

下列缩略语适用于本文件。

AIMs: 祖先信息遗传标记 (Ancestry Informative Markers)

AISNPs: 祖源SNP (Ancestry Informative SNPs)

bp: 碱基对 (Base Pair)

DNA: 脱氧核糖核酸 (Deoxyribonucleic Acid)

Fst: 族群间遗传分化指数 (Fixation index between subpopulation and total population)

InDel: 插入缺失 (Insertion Deletion)

IISNPs: 个体识别SNP (Individual Identification SNPs)

MAF: 次等位基因频率 (Minor Allele Frequency)

nt: 碱基 (Nucleotide)

PIC: 多态信息量 (Polymorphism Information Content)

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

STR: 短串联重复序列 (Short Tandem Repeat)

SOP: 标准作业程序 (Standard Operating Procedure)

5 基本要求

5.1 实验室要求

实验室应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。实验室建设其他要求应符合GB 19489和GB 50346的规定。

5.2 设施设备要求

5.2.1 主要仪器

试验过程中主要包括以下仪器：

- a) 高通量测序仪：应满足测序通量 ≥ 100 Mb，碱基识别质量 >20 ，碱基识别正确率 $\geq 99.9\%$ ；
- b) 聚合酶链式反应仪：应满足温度设置范围为 $0\text{ }^{\circ}\text{C}\sim 99\text{ }^{\circ}\text{C}$ ，最大循环数为99；
- c) 高速离心机：应满足最高转速 ≥ 16000 r/min；
- d) 微量移液管；
- e) 涡旋震荡仪；
- f) 冰箱：应满足温度调节范围为 $-20\text{ }^{\circ}\text{C}\sim 4\text{ }^{\circ}\text{C}$ 。

5.2.2 主要试剂

试验过程中主要包括以下试剂：

- a) DNA 提取试剂盒；
- b) PCR 扩增试剂：基因身份证的引物混合液及 PCR 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；
- c) 文库制备试剂盒：核酸样本的基因文库构建试剂包括 DNA 聚合酶、PCR 缓冲液；
- d) 高通量测序试剂：包括测序引物、高效测序反应酶、缓冲液。

6 基因组遗传标记选取类型及要求

6.1 遗传标记类型

遗传标记类型见表1。

表1 遗传标记类型

	个体识别	急救医学	种族溯源
遗传标记类型	IISNP、STR	ABO、Rh血型、HLA分型	AISNPs

6.2 遗传标记选取要求

6.2.1 个体识别

进行个体识别时，针对目前应用较为广泛的STR和SNP遗传标记，遗传标记选取应符合以下要求：

- a) 基因座定义和具有的特征已有文献报道，已有种属特异性、灵敏度、稳定性研究，有可供使用并公开发表的群体遗传学数据；
- b) 遗传方式符合孟德尔遗传规律；
- c) 遗传标记的分型不受年龄、疾病及其他因素影响，终生不变；
- d) 体细胞稳定性，同一个体的不同组织有相同的分型；
- e) 具有遗传多态性，符合哈迪—温伯格（Hardy—Weinberg，H—W）平衡法则，STR 遗传标记的 $H>0.5$ 、 $PIC>0.5$ ，SNP 遗传标记的 $H>0.4$ 、 $MAF>0.4$ ；
- f) 通过家系调查，至少观察 500 次减数分裂后，遗传标记的突变率 $<0.2\%$ ；
- g) STR 侧翼 50 bp 范围内无超过 5 bp 的 InDel，以保证捕获效率；
- h) 位点间互不连锁 ($r^2<0.01$)。

6.2.2 ABO、Rh 血型，HLA 分型

根据目前数据库收录表型与基因序列差异，用于ABO、Rh血型亚型分型和HLA五位点四位数字高分辨分型应满足以下条件：

- a) 探针捕获方法探针设计覆盖 ABO 基因全区域；
- b) 探针捕获方法探针设计覆盖 RHD 和 RHCE 基因全区域；

- c) HLA—A、B、C、DQB1、DRB1 五位点四位数字高分辨型别。探针捕获方法探针设计覆盖 HLA—A、B、C、DQB1、DRB1 五位点全基因区域。

6.2.3 种族溯源

一组SNP位点用于个体种族来源推断，应符合以下要求：

- 所选位点构建的体系符合哈迪—温伯格 (Hardy—Weinberg, H—W) 平衡法则 ($P > 0.001$)；
- 人群特异性位点，即某个位点在一个人群中具有多态性，而在另外的人群中不具备多态性特征，或者在一个人群中某等位基因较其他人群更多的出现。选择以 $MAF > 0.01$, $\delta > 0.5$ (绝对等位基因频率差异)；
- $F_{st} > 0.3$ 。

注：符合遗传标记选取要求的基因座及遗传信息参见附录A。

7 制作流程

7.1 概述

基因身份证的制作分为5个步骤，具体流程见图1。

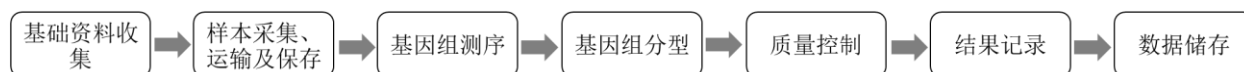


图1 基因身份证制作流程

7.2 基础资料收集

基因身份证基础资料应包括：本人姓名、性别、民族、身份证号、联系方式、家庭住址、既往史、彩色正面免冠数字化照片及知情同意书。基因身份证信息采集单参见附录B。

7.3 样本采集、运输及保存

7.3.1 样本类型

外周血、唾液、口腔脱落细胞等。

7.3.2 采集、运输及保存方法

参见SZDB/Z 124—2015中4.2 受检者资料和样本的采集、运输与保存的要求。

7.4 基因组测序

7.4.1 DNA 提取

根据样本类型，按照商业试剂盒中相应的操作说明进行提取操作。

7.4.2 标签文库建立与测序

构建基于芯片捕获的标签文库，并依据不同测序平台的参考步骤测序。建库过程中应设置阴性对照与阳性对照。测序仪要求参见GB/T 30989—2014，7.3测序仪器设备。

7.5 基因组分型

7.5.1 概述

将测序数据进行过滤、序列比对，并对比对结果进行筛选统计，判断各个位点的基因型，具体流程见图2。



图2 基因组分型流程图

7.5.2 过滤

去除测序数据中有接头污染、扩增重复与低质量（碱基识别质量值 <30 ）的序列。

7.5.3 序列比对

将过滤后的序列与参考序列进行比对，确定相对位置关系，生成比对文件。

7.5.4 基因组分型

通过统计分析比对文件，判断各位点的基因型。其中，祖源相关SNP位点分型结果可用于进行主成分分析（PCA）及群体结构分析。

7.6 质量控制

7.6.1 DNA 提取

7.6.1.1 采用荧光定量仪或酶标仪，对提取的DNA进行质控。合格的DNA提取产物应符合DNA浓度 ≥ 20 ng/ μ L、DNA总质量 ≥ 1 μ g、OD 260/280值在1.7~1.9之间的要求。

7.6.1.2 采用电泳方法检测DNA有无降解，电泳目标片段应条带清晰，不弥散。对于微量生物学样本，若提取DNA不足10 ng，宜采用全基因组扩增商业试剂盒，扩增后再进行文库构建。

7.6.2 文库构建

采用荧光定量仪对DNA提取产物进行定量，文库构建起始的DNA总量可根据相应文库构建试剂、高通量测序设备操作说明进行定量质控；片段大小应根据不同测序设备读长设定。

7.6.3 芯片捕获

纯化后定量检测，DNA浓度应 >2.5 ng/ μ L。

7.6.4 测序前DNA样品质量控制

测序前定量检测DNA浓度，宜根据高通量测序设备操作说明及项目要求进行定量。

7.6.5 测序数据

测序数据质量应满足Q30 $\%>80$ ，SplitRate $\%>90$ 。

注：Q30%，错误率在0.1%碱基质量值为30的序列比例。SplitRate%，拆分率，数据中成功拆除标签的序列占总数据的比例。

7.6.6 分型结果

每个位点最低支持reads数目应符合表2的要求，若不符合则判定该位点分型失败，若样本中90%以上位点分型失败则判定该样本检测失败（对照品除外）。

表2 参数数值对照表

测序reads支持数目 ^a (X)	最低的碱基支持数 ^b (X)
5000~7000	300
2500	250
1000	50
500	40
200	10
100	5
50	5
^a 表示被测基因组上单个碱基被测到的平均次数。	
^b 最低的碱基支持数为在相应的测序 reads 支持数目时，所需要的最少碱基支持数。	

7.6.7 异常记录

7.6.7.1 分析流程中，任何偏离数据标准分析流程之处均应详细记录，包括软件程序包、脚本、版本号、数据库、命令行或参数的任何变动。分析流程中出现的漏洞、纰漏应详细记录在异常记录文档中，包括问题的描述、问题发生原因的调查、修正措施、相关的交流及直接负责人的批示。异常记录文档应与相关案例分析报告对应。

7.6.7.2 分析特殊样本或基因组区间时，流程中任何与 SOP 不同的变动、相应的解释说明均应在最终的结果报告中体现。

7.7 结果记录

基因身份证结果记录应包括表3的内容。

表3 结果记录

内容	生物标记	结果分析	结果记录示例
个体识别SNP或STR位点集	SNP或STR	记录分型结果	rs1229984: T/T; D21S11: 29,30
ABO血型	ABO血型基因	NCBI数据库亚型	O04/B101
RH血型	RHD和RHCE	NCBI数据库RHD亚型	RHD 1227G > A /d
HLA	HLA—A、B、C、DQB1、DRB1	国际IMGT/HLA数据库	HLA—A: 11:01:01, 24:02:01 HLA—B: 35:01:01, 54:01:01 HLA—C: 1:02:01, 4:01:01 HLA—DQB1: 3:01:01, 3:03:02 HLA—DRB1: 9:01:02, 11:01:01
祖源分析	SNP	记录分型结果及群体结构分析	90%东亚, 10%欧洲

7.8 数据储存

基因身份证数据应存储于有安全性保障的数据库或存储于拥有不少于102400个字节内存的芯片卡。

附录 A

(资料性)

DNA 位点

A.1 个体识别

A.1.1 常染色体STR 位点

D21S11、D18S51、D5S818、D7S820、D13S317、D16S539、FGA、D8S1179、D3S1358、TPOX、Penta E、Penta D、D2S1338、D19S433、D12S391、D6S1043、D1S1656、D2S441、D3S1744、D3S3045、D4S2366、D5S2500、D6S477、D7S1517、D7S3048、D8S1132、D10S1248、D10S1435、D10S2325、D11S2368、D13S325、D14S608、D15S659、D17S1290、D18S535、D19S253、D21S2055、D22—GATA198B05。

A.1.2 X染色体STR 位点

DXS6789、DXS6795、DXS6803、DXS6809、DXS7132、DXS7133、DXS7423、DXS8377、DXS8378、DXS9895、DXS9898、DXS10101、DXS10134、DXS10135、DXS10074、GATA172D05、HPRTB。

A.1.3 Y染色体STR 位点

DYS456、DYS389I、DYS390、DYS389II、DYS458、DYS19、DYS385 a/b、DYS393、DYS391、DYS439、DYS635、DYS392、Y GATA H4、DYS437、DYS438、DYS448。

A.1.4 常染色体SNP位点

A.1.4.1 二等位SNP 位点

rs740910、rs1490413、rs1335873、rs1979255、rs1493232、rs2040411、rs1528460、rs717302、rs251934、rs8037429、rs891700、rs901398、rs873196、rs964681、rs737681、rs1463729、rs1360288、rs1382387、rs1413212、rs2056277、rs2107612、rs1015250、rs1005533、rs729172、rs10495407、rs1357617、rs719366、rs1031825、rs733164、rs938283、rs2111980、rs1886510、rs914165、rs354439、rs763869、rs2076848、rs1024116、rs1355366、rs735155、rs1454361、rs727811、rs917118、rs2831700、rs907100、rs1029047、rs2046361、rs722098、rs876724、rs2016276、rs826472、rs2830795、rs1028528。

A.1.4.2 三等位SNP 位点

rs1630312、rs3091244、rs2069945、rs6001030、rs140676、rs356167、rs941454、rs10045、rs3743842、rs2298556、rs3816662、rs2307223、rs10811897、rs17287498、rs385780、rs11141033、rs4540055、rs3812847、rs2032582、rs2278786。

A.1.5 X染色体SNP 位点

rs2056688、rs2128519、rs1534285、rs763056、rs1373592、rs993010、rs1557054、rs1243792、rs925178、rs1207480、rs1936313、rs1977719、rs1372687、rs1857602、rs985425、rs933315、rs2190288、rs1991961、rs1931662、rs149910、rs1573704、rs1340718、rs1930674、rs1339597、rs1981452。

A.1.6 Y染色体SNP 位点

rs11096433、M145、rs9306845、rs9786479、rs17276358、rs2075640、M134、M88、M95、rs16980426、rs17323322、M122、rs13447354、M89、rs9786707、M15、rs16980711、M9、rs17316592、rs17276345。

A.1.7 线粒体SNP 位点

709、1719、1736、3010、3394、3970、4216、4883、5147、5417、5460、6392、6455、8584、8701、9090、10397、10398、11914、12705、13708、13928、14318、14783、15487、16519。

注：在上述SNP基础上可增加文献报道且验证的其他SNP，以提高检测的系统效能

A.2 种族溯源

在祖源分析时，区分东亚、欧洲、非洲常用SNP基因座如下：

rs2814778、rs3737576、rs7554936、rs10497191、rs1834619、rs1876482、rs260690、rs3827760、rs6754311、rs798443、rs12498138、rs1919550、rs1229984、rs3811801、rs4833103、rs7657799、rs7722456、rs870347、rs16891982、rs192655、rs3823159、rs917115、rs1462906、rs1871534、rs2196051、rs6990312、rs3814134、rs4918664、rs1079597、rs174570、rs2238151、rs671、rs1572018、rs2166624、rs7326934、rs7997709、rs9522149、rs200354、rs12439433、rs1426654、rs1800414、rs735480、rs12913832、rs459920、rs11652805、rs17642714、rs2593595、rs4411548、rs4471745、rs2042762、rs3916235、rs4891825、rs7226659、rs7251928、rs310644、rs2024566。

注：在上述SNP基础上可增加文献报道且验证的其他SNP，以提高系统的区分能力，比如区分中国的少数民族人群。

附 录 B
(资料性)
基因身份证信息采集单

B.1 基因身份证信息采集单见图 B.1。

基因身份证信息采集单									
<p>送检信息</p> <p>送检人姓名_____ 送检人联系方式_____</p> <p>送检人身份证号码_____ 送检人家庭住址_____</p>								<div style="border: 1px solid black; width: 100%; height: 100%; text-align: center; padding: 5px;">彩色照片黏贴处</div>	
<p>送检人个人信息登记</p>									
姓名	性别	民族	身份证号	ABO 血型	Rh血 型	HLA 分型	既往史 1. 有无免疫治疗; 2. 有无异体输血; 3. 有无器官移植	送检样本类型 1. 外周血; 2. 口腔拭子; 3. 血痕; 4. DNA; 5. 其他(须注明)	入 库 条 码 标 记

注1: ABO、Rh 血型及 HLA 分型不知则填“—”;

注2: 彩色照片采集按照《GA/T 461—2019 居民身份证制证用数字相片技术要求》;

注3: 指纹采集按照《GA/T 1012—2019 居民身份证指纹采集和比对技术规范》;

注4: 样本采集参考《SZDB/Z 124—2015 基于二代测序技术的 HLA 高分辨分型检测标准》。

检测机构人员填写

样本接收日期_____

接收人_____

备注(对不符合接收标准的样本描述)_____

图B.1 基因身份证信息采集单

参 考 文 献

- [1] Dean, M., et al. (1994). "Polymorphic admixture typing in human ethnic populations." *American Journal of Human Genetics* 55(4): 788—808.
- [2] Edwards, A., et al. (1992). "Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups." *Genomics* 12(2): 241—253.
- [3] 贾竟 (2014). 生物物证供者种族来源推断的SNP体系研究, 重庆医科大学.
- [4] Holsinger, K. E. and B. S. Weir (2009). "Genetics in geographically structured populations: defining, estimating and interpreting FST." *Nature Reviews Genetics* 10(9): 639
-