

DB4403

深 圳 市 地 方 标 准

DB4403/T X—XXXX

基因身份证技术规程

Technical Regulation of DNA ID

(送审稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

深圳市市场监督管理局 发布

前 言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由深圳华大法医科技有限公司提出。

本标准由深圳市发展和改革委员会归口。

本标准的附录A为资料性附录，附录B为资料性附录。

本标准起草单位：深圳华大法医科技有限公司、深圳市华大司法技术协同创新研究院。

本标准起草人：李生斌、张喆、常辽、刘文嘉、王轶男、沈悦生。

本标准为首次制定。

基因身份证技术规程

1 范围

本标准规定了基因身份证制作的技术流程，包括基于新一代测序技术（NGS）的样本采集、基因组位点选择、实验流程、数据分析、数据存储的技术要求及质量控制的要求。

本标准适用于基因身份证制作。

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，仅所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 19489-2004 实验室生物安全通用要求

GB 50346-2004 生物安全实验室建筑技术规范

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

GA 461-2004 居民身份证制证用数字相片技术标准

GA 1012-2012 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

基因身份证又称 DNA 身份证，它由个人基础资料和基因组信息组成。基因身份证利用个体基因组多态性原理，甄别个体、种族、血型等，为识别个人身份提供证据。

3.2 ABO 血型基因 ABO gene

ABO血型基因座位于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 gene typing

从基因序列水平对ABO血型进行分型，其结果精确到NCBI数据库级别血清学水平后的亚型。ABO血型基因分型能准确定位到亚型，可解决临床上受检者由于药物治疗、血液病等原因导致的抗原、抗体减弱或缺失所导致的正反定型结果不一致的疑难血型样本。

3.4 Rh 血型基因 Rh gene

DB4403/T X—XXXX

Rh血型基因座位于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 gene typing

从基因序列水平对Rh血型进行分型，可区分免疫原性最强的RHD基因全缺失、部分D、弱D、DEL型，其结果能精确定位到亚型。

3.6 HLA 高分辨分型 HLA high resolution typing

即从基因序列水平对 HLA 进行分型的技术，其结果精确到血清学水平后的亚型。HLA 高分辨分型与过去的低分辨分型相比配型速度更快，配型更精确，使得移植排斥反应更小，手术成功率和术后存活率更高。

3.7 种族溯源 ancestry traceability

通过筛选不同地域和民族中基因与基因型频率存在明显群体差异的一组SNP位点，用于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的成分组成。

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

是通过测序技术获得个人基因组上多态性位点的分型信息，比较这些分型数据与任意参考族群的相似性，用以计算祖源成分。

3.9 探针捕获 probe capture

选择靶基因区域，设计探针长度为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。分为“固相”杂交芯片和磁珠的“液相”杂交系统。

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。
[GB/T 29859-2013, 2.6.5]

4 缩略语

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）

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 基础条件

5.1 实验室条件

普通实验室应根据GB 50346-2004和GB 19489-2004要求，参考生物安全实验室建筑平面、装修和结构的技术要求，给水排水、气体供应、配电、自动控制和消防设施配置的原则等。基于二代测序技术的基因身份证制作实验室应符合基因扩增检验实验室的要求，对PCR实验室的设置和仪器设备进行分区和仪器专用明确标记，标本来样处理、测定中的核酸提取扩增和产物分析等。

5.2 设施设备条件

5.2.1 主要仪器

- a) 高通量测序仪；
- b) 聚合酶链式反应仪；
- c) 掌上离心机；
- d) 微量移液管；
- e) 恒温温育仪。

5.2.2 主要试剂

- a) DNA提取试剂盒；
- b) 文库制备试剂盒；
- c) 高通量测序试剂套装。

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

筛选遗传标记进行个体识别时，针对目前应用较为广泛的STR和SNP遗传标记，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；
- b) 已有种属特异性、灵敏度、稳定性研究；
- c) 遗传方式符合相应的遗传规律；
- d) 遗传标记的分型不受年龄、疾病及其他因素影响，终生不变；
- e) 体细胞稳定性，同一个体的不同组织有相同的分型；
- f) 具有遗传多态性，符合哈迪-温伯格(Hardy-Weinberg, H-W)法则。STR遗传标记的 $H>0.5$ 、 $PIC>0.5$ ；SNP遗传标记的 $H>0.4$ 、 $MAF>0.4$ ；
- g) 有可供使用的并公开发表的群体遗传学数据；
- h) 经家系调查，至少观察500次减数分裂后，遗传标记的突变率 $<0.2\%$ ；
- i) STR侧翼50 bp范围内无超过5 bp的InDel以保证捕获效率；

DB4403/T X—XXXX

- j) 位点间互不连锁 ($r^2 < 0.01$);
- k) 检验方法操作简单, 重复性好, 结果明确可靠。

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需满足以下条件:

- 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:

- a) 所选位点构建的体系符合H-W平衡法则 ($P > 0.001$);
- b) 人群特异性位点, 即在三大人群中, 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 $MAF > 0.01$ ^[3], $\delta > 0.5$ ^{[1][2][3]} (绝对等位基因频率差异);
- c) $F_{st} > 0.3$ ^{[3][4]}。

7 制作流程

基因身份证的制作有5个步骤, 具体流程见图1所示:

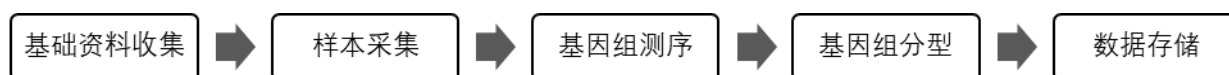


图1 基因身份证制作流程

7.1 基础资料收集

基因身份证基础资料包括: 本人身份证号码、性别、民族、出生年月日、彩色正面免冠数字化照片及指纹, 并签署知情同意书。

7.2 样本采集、运输及保存

7.2.1 样本类型

外周血、口腔拭子等。

7.2.2 采集、运输及保存方法

参见SZDB/Z 124-2015第4.2 受检者资料和样本的采集、运输与保存。

7.3 基因组测序

7.3.1 DNA 提取

按照商业试剂盒中相应检材的处理步骤进行提取操作。

7.3.2 标签文库建立与测序

构建基于芯片捕获的标签文库，依据不同测序平台按照其参考步骤构建文库及上机测序，建库过程中设置阴性对照与阳性对照。测序仪基本要求参见GB/T 30989-2014第7.3 测序仪器设备。

7.4 基因组分型

将测序数据进行过滤比对，并对比对结果进行筛选统计，从而判断各个位点的基因型，具体过程见图2。



图2 基因组分型流程图

7.4.1 过滤

对原始测序数据进行过滤，去除接头污染、扩增重复与低质量序列（碱基识别质量<30）。

7.4.2 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。

7.4.3 分型

通过对比对文件的统计分析，判断各位点的基因型。其中，提取祖源相关SNP位点分型结果，进行主成分分析（PCA）及群体结构分析得到结论。

7.5 质量控制

7.5.1 DNA 提取

采取荧光定量仪或酶标仪检测，对提取的DNA进行质控，应符合DNA浓度 $\geq 20\text{ng}/\mu\text{L}$ 、DNA总质量 $\geq 1\mu\text{g}$ 、OD 260/280值在1.7-1.9之间的要求

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。对于微量生物学检材，若提取DNA不足10 ng，建议采用全基因组扩增商业试剂盒，扩增后再进行文库建立。

7.5.2 文库构建

采用荧光定量仪对PCR后DNA进行定量，采用毛细管电泳仪检测文库的片段大小及产量，DNA总量可根据相应文库构建试剂、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。

7.5.3 芯片捕获

纯化后定量检测，建议DNA浓度 $>2.5\text{ng}/\mu\text{L}$ 。

7.5.4 上机测序前 DNA 样品的质量控制

上机测序前定量检测DNA浓度，建议根据高通量测序设备操作说明及项目要求进行定量。

7.5.5 测序数据

DB4403/T X—XXXX

下机的测序数据质量要求Q30%>80，SplitRate%>90。

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

7.5.6 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 ¹ (X)	最低的碱基支持数 ² (X)
5000-7000	300
2500	250
1000	50
500	40
200	10
100	5
50	5
注1：表示被测基因组上单个碱基被测到的平均次数。	
注2：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。	

7.5.7 异常记录

分析流程中，任何偏离数据标准分析流程的地方都要详细记录，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。分析流程中出现的一些漏洞，纰漏也需要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。异常记录文档需要与相关案例分析报告对应起来，而实验室主管也需要将这些与数据分析SOP差异的地方与对应的实验人员进行交流。

分析一些特殊样本或基因组区间的时候，相对于SOP而言，整个分析流程中所涉及到的任何修改变动，对应的解释说明都需要体现在最终的结果报告中。

7.6 结果记录

基因身份证结果记录应有以下内容。

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、	国际IMGT/HLA数	HLA-A: 11:01:01, 24:02:01

	DQB1、DRB1	数据库	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

7.7 数据储存

基因身份证数据需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
(资料性附录)
DNA 位点

本附录提供了符合遗传标记筛选条件的基因座及遗传信息,本指导性规范建议可在以下遗传标记中挑选。

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. 1.5 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. 1.6 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
(资料性附录)
基于第二代测序技术的基因身份证信息采集单

送检信息

送检人姓名_____ 送检人联系方式_____

送检人身份证号码_____ 送检人家庭住址_____

彩色照片黏贴处

送检人个人信息登记

姓名	性别 1、 2、	民族	A B O 血型	R h 血 型	H L A 分 型	既往史 1、 疫治疗 2、 无输血 3、 无器官移植	送检样本类型 1、 周血 2、 腔拭子 3、 血 4、 NA 5、 他（须注明）	入库 条 码 标 记
张三	男	汉	A	阳 性	— —	无	2	

注：
ABO、Rh血型及HLA分型不知则“——”；
彩色照片采集按照《GA 461-2004 居民身份证制证用数字相片技术标准》；
指纹采集按照《GA 1012-2012 居民身份证指纹采集和比对技术规范》；
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）_____

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
-