

# 深圳市地方标准

## 基因身份证技术规程 编制说明

（送审稿）

起草工作组

2019 年 1 月 18 日

## 目 录

一、工作简况.....	3
1、任务来源.....	3
2、主要工作过程.....	3
二、标准编制原则和确定主要内容的论据及解决的主要问题 .....	3
1、标准制定的基本原则 .....	3
2、标准编制的主要内容及依据 .....	4
三、主要验证情况分析.....	4
(1) DNA 提取.....	5
(2) 文库建立和测序 .....	5
(3) 基因分型结果 .....	5
四、其他情况说明.....	7

# 地方标准《基因身份证技术规程》编制说明

## 一、工作简况

### 1、任务来源

本标准由深圳市发展和改革委员会归口。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负责标准的主要编写工作。

### 2、主要工作过程

(1) 2016 年 7 月至 2017 年 2 月，标准起草单位组织相关技术人员对《基因身份证技术规程》项目进行了预研，证明了项目的可行性。

(2) 2017 年 3 月，标准起草单位在收到深圳市市场监督管理局《深圳市市场监督管理局关于开展 2016 年深圳市技术标准文件制修订项目申报工作的通知》后，提交了项目建议书及项目汇总表。

(3) 2017 年 4 月，本标准正式立项。

(4) 2017 年 5 月-2018 年 7 月，标准起草单位对标准初稿及编制说明进行反复讨论修改，最终形成标准草案和编制说明。

(5) 2018 年 7 月至 8 月，标准编制组积极组织国内相关单位和专家对草案进行内部专家征求意见，标准起草小组根据专家意见历经多次面对面沟通和电话会议，对标准草稿进行了修改和完善，形成标准征求意见稿和编制说明。

### 3、标准编制的主要成员单位及其所做的工作

标准由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申报立项并负责起草制定，为此专门成立标准编制小组，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负责标准内容的撰写及修订。

。

## 二、标准编制原则和确定主要内容的论据及解决的主要问题

### 1、标准制定的基本原则

标准编制遵循“统一性、适用性、一致性、规范性”的原则，注重标准的可操作性，本标准严格按照《GB/T 1.1-2009 标准化工作导则 第 1 部分：标准的

结构和编写规则》的规定进行编写和表述。

本标准的编制原则是符合基因产业发展、应用现状，满足并兼容现阶段多种遗传标记，形成内容准确、技术先进、广泛适用的基因身份证规范。

## 2、标准编制的主要内容及依据

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、数据分析的技术要求及质量控制的要求。

标准基本构架包括：范围、范性引用文件、术语和定义、缩略语、基础条件、基因组遗传标记选取类型及要求，制作流程和附录。

### 3、编制过程中解决的主要问题（做出的贡献）

基因组识别以其难以替代的高可信度和高准确度在公共安全、信息安全、电子商务、电子政务、军事等众多领域都得到了广泛的应用。传统 DNA 识别技术基于美国的 CODIS 系统（含有 13 个 STR 等位基因位点），改良的方法只是将 STR 位点增加至几十个，并不能改变该系统存在等位基因突变率高、基因多态性差以及基因频率在各人群间不适用等天然缺陷，从而引起结果难以解释甚至结果误判，由此导致的错案不断增加。

本标准采用的基因组分型技术是从全基因组水平选取适用于中国人群的多态性位点，结合生物计算和人群大数据构成的生物特征识别系统，具有科学、精准、先进、实用等优势。本标准提供了一套基因身份证制作技术规程，规范基于新一代测序技术（NGS）的样本采集、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。

## 三、主要验证情况分析

本标准的相关要求都经主办单位组织进行了测试，以下为本项目的综述报告。

本实验室共收集 170 例个体样本，记录血型、HLA 分型（部分已有 Sanger 测序分型结果）、民族等信息，其中 3 例家系。按照标准流程制作基因身份证：

(1) DNA 提取。采用相应商业试剂盒提取基因组 DNA，OD260/OD280 值均在 1.7-1.9 之间，琼脂糖电泳检测 DNA 无降解或少量降解。

(2) 按照建库测序试剂盒步骤进行文库建立和测序，其中，筛选用于个体识别的 SNP 位点集，经千人基因组数据库中基因频率计算得到匹配概率  $P_m$  小于  $10^{-500}$ ，世界上 70 亿人没有两个人的分型一样。

$$P_m = \sum_{i=1}^n P_i^2$$

其中  $n$  为某一遗传标记的表型数目， $P_i$  为该群体第  $i$  个表型的频率。 $\sum_{i=1}^n P_i^2$  指人群中随机抽取两个无关个体在某一基因座上两者表型纯粹由于机会而一致的概率。

(3) 基因分型结果

表 3-1 ABO 血型验证情况

序号	样本编号	血清学表型	商业试剂盒验证	NGS 分型
1	TID-4	A	A01	001/A101
共计		55	6	55

检测同时有 ABO 血清学结果（体检或献血由医院检测提供）和 NGS 分型结果的共计 55 例，其中 4 例为购买标准品，血清学结果与本实验 NGS 分型结果一致，准确性高达 100%。

表 3-2 RH 血型验证情况

序号	样本编号	血清学表型	商业试剂盒验证	NGS 分型
1	TID-4	RH-Positive	阳性	DdCCee
共计		55	4	55

检测同时有 RH 血清学结果（体检或献血由医院检测提供）和 NGS 分型结果的共计 55 例，其中 4 例为购买标准品，血清学结果与本实验 NGS 分型结果一致，准确性高达 100%。

表 3-3 HLA 分型验证情况

序号	样本编号	Sanger 分型结果	NGS 分型
----	------	-------------	--------

1	TID-4	HLA-A: 02:07, 11:01 HLA-B: 44:03, 46:01 HLA-C: 01:03, 04:01 HLA-DQB1: 02:02, 03:01 HLA-DRB1: 07:01, 11:01	HLA-A: 02:07:01, 11:01:01 HLA-B: 44:03:01, 46:01:01 HLA-C: 01:03, 04:01:01 HLA-DQB1: 02:02, 03:01:01 HLA-DRB1: 07:01:01, 11:01:01
共计		16	16

同时有 Sanger 结果和本研究 NGS 分型结果的检测样本共 16 例，Sanger 结果展示 4 位精度，本研究大多数能做到 6 位精度。

表 3-4 祖源分析检测人群

人群	检测数目（人）	数据来源
AFR 非洲人	663	千人基因组
CAS 高加索人	21	本研究
HAN 汉族人	52	本研究

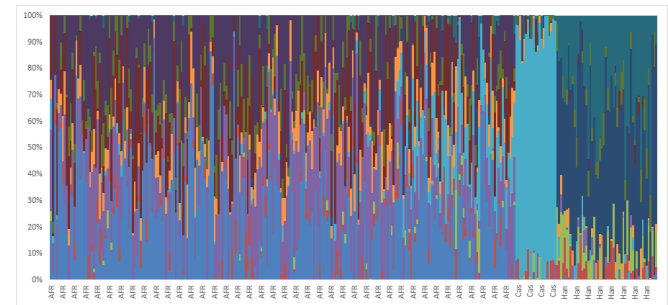


图 3-1 STRUCTURE 群体结构分析——板块

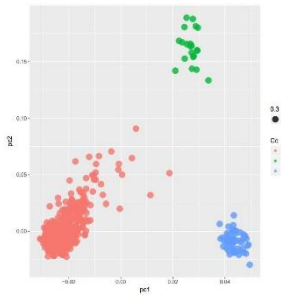


图 3-2 PCA 结果图

结果：本祖源分析相关 SNP 位点集可区分到大陆板块，非洲、欧洲和汉族。  
位点选取说明：

$\delta$  值，绝对等位基因频率差异，指来自两个不同的地区或民族的群体，某一位点上等位基因频率差的绝对值，其公式为：

$$\delta = \frac{\sum_{x=1}^k |p_x - q_x|}{2}$$

其中  $\delta \geq 0$ ，k 为该位点等位基因数， $p_x$  与  $q_x$  分别表示某等位基因 x 在群体 p 和群体 q 中的频率，据文献<sup>[1][2][3]</sup>，当  $\delta > 0.5$  时，是适合个体祖先地域或种族来源推断的理想遗传标记。

Fst 值，为族群间遗传分化系数，是指等位基因频率在群体间分布的差异性

情况，是群体遗传学参数与检测遗传结构的指标。测量的是平均杂合度在亚群间与总群体中的差异程度，用于祖先信息推断时， $F_{st}$  值越大说明某位点在不同人群之间的基因频率差别越大，适合用来区分人群<sup>[3][4]</sup>。

#### 四、知识产权情况说明

本标准暂未发现相关知识产权问题。

#### 五、产业化情况、推广应用论证和预期达到的经济效果

基因组识别技术的出现，使得 DNA 作为生物特征进行身份验证已经被公安、司法、法医等领域高度认可，广泛应用于个体识别、亲权鉴定、数据库建设等方面，具有唯一性、稳定性、稳定遗传性、不可复制性等特点。麦肯锡将高通量测序技术作为十年后拥有影响经济巨大潜力的颠覆性技术之一，预计到 2025 年，市场规模将达到 0.7-1.6 万亿美元，通过快速 DNA 诊断与靶向药物等，延长及改善 75% 的生命。高通量测序技术将在五年内占领法医（个体识别）市场的 50%。印度唯一身份认证机构（Unique Identification Authority of India, UIDAI）是历史上最大的人类生物识别项目，最终目标是为该国的 12 亿居民提供独特的基因身份证。

目前个体 DNA 身份检测主要采用 DNA 扩增和测序结合的方式，从方法学到仪器设备、试剂耗材，再到数据库建设都依赖和受制于美国，严重滞后于国家安全和法制建设体系迫切需求。近些年来，随着 NGS 技术的出现并快速发展，我国完成了从跟随到引领的跨越，华大基因成为全球最大的生物信息学测序中心和数据产出中心，围绕基因组识别，从核心技术到国际标准，从专用试剂到关键遗传分析仪器，从识别计算到智能数据系统，华大基因具有整套独立知识产权和发明专利。通过国际救援合作研究和实践，NGS 技术在司法实践中将发挥巨大的不可替代的作用。例如，2005 年参与泰国海啸的救援，并代表中国完成了泰国近 2000 份海啸遇难者遗体的 DNA 鉴定任务，为中国争得了荣誉；2005 年 7 月至 10 月，为青海省公安厅和司法厅免费检测了近两万人份的犯罪人群 DNA 数据；2007 年，为了帮助那些失散多年的人采用最准确最经济的办法寻亲，建立起了中国首个寻亲人员 DNA 数据库，截止到 2018 年，这个数据库已成功帮助 78 个家庭找到失散的亲人。到目前为止，据不完全统计，华大基因受理 DNA 识别案件超过 200,000 例，华大基因司法鉴定中心入选司法系统十大司法鉴定机构，多

次受到党和政府的表彰。

本标准采用的基因组分型技术是从全基因组水平选取适用于中国人群的多态性位点，结合生物计算和人群大数据构成的生物特征识别系统，具有科学、精准、先进、实用等优势。本标准提供了一套基因身份证制作技术规程，规范基于新一代测序技术（NGS）的样本采集、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，符合本标准产生的信息可存储于身份证、社保卡、医保卡等留有存储空间卡片。本标准若经批准实施，将有利于规范整个个体识别行业市场，确保各检测单位检测质量，从而提升准确性和安全性，促进我国个体身份鉴定产业健康稳定快速发展。

## **六、采用国际标准和国外先进标准情况**

本标准为国内外首个基因身份证技术规程标准，未采用国际标准。

## **七、与现行相关法律、法规、规章及相关标准的协调性**

本标准与我国现行其他标准不存在依赖关系或不协调的问题。

## **八、重大分歧意见的处理经过和依据**

本标准在制定和编制过程中没有重大分歧。

## **九、地方标准作为强制性地方标准或推荐性地方标准的建议**

建议本标准作为深圳市地方标准，指导性技术文件发布。

## **十、贯彻标准的要求和措施建议**

本标准的制定将使基因身份证制作有标准依据，建议标准尽快实施发布。

## **十一、替代或废止现行相关标准的建议**

本标准属于新制定的地方标准，不存在替代旧标准的问题。

## **十二、其它应予说明的事项**

无。

深圳市标准《基因身份证技术规程》

编制工作组

## **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、贾竞 (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