

# 《人全基因组高通量测序数据质量评价方法》

## 标准征求意见稿编制说明

### 一、工作简况

1、任务来源：本标准由国家药品监督管理局提出，全国医用临床检验实验室和体外诊断系统标准化技术委员会归口。任务来源为《国家标准化管理委员会关于下达2023年第一批推荐性国家标准计划及相关标准外文版计划的通知》（国标委发〔2023〕10号），标准中文名称为《人全基因组高通量测序数据质量评价方法》，国家标准计划号：20230028-T-464，标准英文名称为《The Data Quality Evaluation Method of Human Whole Genome Sequencing》，外文版项目计划编号：W20233147。

2、工作过程：主要包括预研阶段、立项阶段、起草阶段、验证阶段、征求意见阶段、审查阶段。

#### 2.1 预研阶段

（1）自2020年10月至2021年5月，标准起草单位组织相关技术人员对标准的研究背景、研究进展及必要性进行了调研，查阅收集了国内外相关标准、指南及文献，了解了国内外相关技术动态，并且明确了工作思路和进程安排。

（2）2022年1月至4月，按照《中检院关于征集2023年度医用高通量测序标准化归口单位标准预立项提案的通知》（中检诊断函〔2022〕67号）要求，提交了项目建议书以及国家标准草案。于2022年9月，医用高通量测序标准化归口单位将该标准报送全国医用临床检验实

验室和体外诊断系统标准化技术委员会。

## 2.2 立项阶段

(1) 2022年9月1日，起草小组对全国医用临床检验实验室和体外诊断系统标准化技术委员会汇报了《人全基因组高通量测序数据质量评价方法》标准的起草工作，与会专家进行了讨论，提出了宝贵的意见和建议。主要意见包括评价方法描述的清晰度和完整性等。标准起草小组根据专家意见进行了修改和完善。

(2) 2023年3月21日，收到国家标准化管理委员会〔2023〕10号文件《国家标准化管理委员会关于下达2023年第一批推荐性国家标准计划及相关标准外文版计划的通知》(国标委发〔2023〕10号)，本标准纳入国家标准计划名单，标准中文名称为《人全基因组高通量测序数据质量评价方法》，国家标准计划号：20230028-T-464，标准英文名称为《The Data Quality Evaluation Method of Human Whole Genome Sequencing》，外文版项目计划编号：W20233147。

## 2.3 起草阶段：

(1) 2023年6月对标准进行了起草研制和验证方案初步设计，征求了8家企业及研究机构的修订建议，进行了草案的进一步修订。收集到主要问题归纳为以下几点：①部分定义描述有误或单词拼写问题，如：标签、有效测序深度、GC含量20x测序覆盖率表述等；②部分定义建议删除，如基因组唯一比对率；③部分指标难以达到的问题，如SV的准确性和灵敏度指标较高。

(2) 2023年6月20日，在北京召开了标准启动与研讨会，来自企

业、审评、检测机构、医院等单位的代表共计200余人参加了讨论。会上标准起草人对标准的起草过程予以说明并汇报了工作组讨论稿的主要内容。会后起草小组根据意见对工作组讨论稿进行完善。

#### 2.4 验证阶段：

(1) 2023年7月进行了验证方案的撰写、修订和确认。

(2) 2023年8月至2023年12月，对方案进行了验证。验证工作包括准备测试标准品、8家验证单位验证数据产出、HJ标准细胞系中国人参考序列/高置信变异集合更新、验证数据分析评估。

(3) 2023年12月至2024年1月，对变异集进行优化和进一步更新，并重新分析评估验证数据。

#### 2.5 征求意见阶段：

2024年1月根据验证评估结果，进行草案修订、英文草案撰写修订，形成了征求意见稿。

## 二、标准编制原则和确定标准主要内容的论据

### 1、标准制定的意义、原则

人全基因组高通量测序检测流程复杂，主要包括样本采集、DNA提取纯化、测序文库制备、文库上机测序、测序数据分析，完成每个环节的试剂耗材、仪器软件可来自国内外不同的供应商，操作方法也存在多样化和各种差异。虽然对全基因组测序应用于临床(如遗传病、新生儿危急重症中心等)，全球范围内已发布多个专家指南，指导临床工作者，包括患者选择、伦理考量、多学科证据、突变分级等，但缺乏对全基因组测序技术这一底层技术的规范和质控。全基因组测序

技术标准,目前国际标准和国内标准处于标准空缺或指标不明确的现状,使得不同使用机构(如医院、检验机构等)采用不一致的定义和质控标准(如测序深度、测序读长、GC含量等),质量标准差异极大。此外,由于目前鲜有标准品同时具备实物标准和配套人全基因组测序的标准数据集(SNP、Indel、SV等),使得不同实验室间无法进行变异检测结果一致性比较。为规范临床检测中心、医院中心实验室的人全基因组测序数据质量,迫切需要建立一套通用型的数据质量评价方法,以统一规范市场。

本标准编写格式按照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》的格式要求进行编写。制定的标准应能反映人全基因组高通量测序的质量评价要求和水平。制定标准时尽可能地做到简化、统一、协调和优化;既要考虑其先进性,也要考虑到实用性和可行性;既要符合国内外发展的需要,也要结合国内目前的实际状况。本标准制定原则符合相关法律、法规的要求。

2、本标准性能指标制定依据,对于有争议指标的处理及验证情况。

征求意见稿形成前,根据多方验证结果,与会专家及企业代表对标准工作组草案进行了全面讨论,其中针数据比对和变异检测的部分衡量指标,进行了全面的讨论。包括如下内容:

1) GC bias的衡量指标:考虑到GC含量与AT偏离率、GC偏离率、AT+GC偏离率这3个指标存在相关性,而GC含量更容易进行质量控制,

因此舍弃了AT偏离率、GC偏离率、AT+GC偏离率这3个指标，保留了GC含量的指标。

2) 覆盖均匀性的衡量指标：因高深度测序下，20x测序覆盖率和特殊区域测序覆盖率更能反应覆盖均匀性，故删除测序覆盖率的指标。

3) 变异检测的衡量指标：因SNP位点数受分析流程、过滤参数、标准变异集等因素影响差异较大，故建议不固定标准数值范围。

三、主要实验（或验证）的分析、综述报告、技术经济论证、预期的经济效果

1) GC 含量：通过验证数据可观察到在同一验证平台，存在标准品 GC 含量正常，但需要提取的细胞、血液 GC 含量偏高的问题，说明此指标可指导实验室更好质控 DNA 提取环节。另一方面，也存在在某一个平台 GC 含量普遍偏高的情况，说明平台可能存在一定系统 GC bias 问题。

2) 20x 测序覆盖率：在截取的测序深度  $\geq 40x$  的数据中，20x 测序覆盖率  $\geq 95\%$  的合格率为 66.7%。分析原因，主要是平台质量影响和提取 DNA 质量影响（GC 含量偏低或偏高）。

3) 变异检测准确率和灵敏度：在截取的测序深度  $\geq 40x$  的数据中，SNP 准确率、SNP 灵敏度、Indel 准确率、Indel 灵敏度、SV 准确率、SV 灵敏度的合格率分别为：75.8%、75.8%、48.5%、75.8%、84.8%、72.7%。不合格的原因主要是不同平台检测变异的质量有差异。

此外，还存在较多指标，可反应数据质量。由此说明，本标准制定的指标和参数可操作性强，指标设置科学合理，能够满足临床使用的要求。

四、采用国际标准和国外先进标准的程度，以及与国际、国外同类标准水平的对比情况，或与测试的国外样品、样机的有关数据对比情况。

目前国际上没有相关产品的官方标准，在制定过程中，参考了国内先进试剂企业的技术要求和说明书等。

五、与有关的现行法律、法规和强制性国家标准的关系。

本标准与有关的现行法律、法规和强制性国家标准不冲突。

六、重大分歧意见的处理经过和依据。

本行业标准制定过程中无重大分歧。

七、国家标准作为强制性国家标准或推荐性国家标准的建议。

建议本标准为推荐性国家标准。

建议中文版国家标准发布后 12 个月实施，外文版发布后实施日期与中文版保持一致。

八、贯彻行业标准的要求和措施建议（包括组织措施、技术措施、过渡办法等内容）

建议在本行业标准发布后实施前进行标准宣贯，宣贯对象是企业、各级医疗器械监管部门。

建议标准发布后一年内实施宣贯，宣贯对象是企业、医学实验室、省市药监局等相关部门。

九、废止现行有关标准的建议。

无。

十、其他应予说明的事项。

无其他说明。

标准起草工作组

2024年 01 月 18 日