

团体标准

T/GBA xxx—xxxx

大人群全基因组测序 湿实验技术规范

Technical specifications for wet lab population-scale whole genome
sequencing

(征求意见稿)

xxxx-xx-xx 发布

xxxx-xx-xx 实施

粤港澳大湾区标准创新联盟 发布

目 次

前 言	II
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 缩略语	7
5 总体要求	8
5.1 总目标及技术方案	8
5.2 湿实验设计原则	8
5.3 环境与设施基本要求	9
5.4 实验室生物安全要求	9
5.5 质量体系与能力保障要求	9
6 技术要求	10
6.1 样本接收与分装存储	10
6.2 核酸提取	10
6.3 文库构建	11
6.4 上机测序	11
6.5 数据交付	12
附 录 A （资料性） 大人群全基因组测序主体技术方案	13
附 录 B （资料性）	14
附 录 C （资料性） 基因组 DNA 质量分级标准	15
附 录 D （资料性） 基因组 DNA 存储技术要求	18
附 录 E （资料性） 人全基因组文库质量分级标准	19
附 录 F （资料性） 人全基因组测序原始数据质量分级标准	22
参 考 文 献	25

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》的规定起草。

本文件由深圳华大智造科技股份有限公司提出。

本文件由粤港澳大湾区标准创新联盟归口。

本文件授权粤港澳大湾区标准创新联盟组织伙伴和所有成员单位使用，联盟组织伙伴需等同采用转化为自身团体标准，并在全国团体标准信息平台上公开标准基本信息。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

大人群全基因组测序湿实验技术规范

1 范围

本文件规定了大人群全基因组测序湿实验的样本接收与分装存储、核酸提取、文库构建、上机测序、数据交付等环节的技术要求。

本文件适用于采用高通量基因测序技术、单分子基因测序技术或上述技术的组合开展的大人群全基因组测序湿实验。

本文件不适用于以下情形：

——样本采集和干实验环节；

——以临床诊断为直接目的的全基因组测序检测。

本文件适用的人类样本类型为血液、唾液等核酸质量和产量较高的样本；不适用于血浆、尿液、福尔马林固定石蜡包埋组织等核酸量少、质量受限的样本，以及粪便、皮肤拭子等非检测人类全基因组序列的样本。

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—2008 实验室 生物安全通用要求

GB/T 30989—2014 高通量基因测序技术规程

GB/T 39766—2021 人类生物样本库管理规范

GB/T 40171—2021 磁珠法 DNA 提取纯化试剂盒检测通则

GB/T 40664—2021 用于高通量基因测序的核酸类样本质量控制通用要求

GB/T 40974—2021 核酸样本质量评价方法

GB/T 44467—2024/ISO 4307: 2021 分子体外诊断检验 唾液检验前过程的规范 提取人类 DNA

GB/T 45214—2025 人全基因组高通量基因测序数据质量评价方法

GB/T 46302—2025 医用全自动样本处理系统

GB/T 37864—2019/ISO 20387: 2018 生物样本库质量和能力通用要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

大人群全基因组测序 population-scale whole genome sequencing

采用高通量基因测序技术或单分子基因测序技术或高通量与单分子联合测序技术，在限定时间内完成不少于1000个人的全基因组测序。

注：根据限定时间内完成人全基因组测序样本个数，可细分为千级（1,000至9,999个）、万级（10,000至99,999个）、

十万级（100,000至999,999个）、百万级（1,000,000至9,999,999个）、千万级（10,000,000至99,999,999个）等不同等级。通常，限定时间与需完成样本个数呈正相关，与实验室配置核心设备通量呈负相关。

3.2

湿实验 wet lab

在符合生物安全与质量控制要求的受控环境下，以人的血液、唾液等生物样品为研究对象，利用试剂、耗材及测序相关仪器设备，通过手工或自动化操作完成样本处理与存储、核酸提取、文库构建及上机测序等步骤，从而获得下机原始测序数据及质量控制信息的实验活动。

3.3

干实验 dry lab

在计算与信息技术环境下，以湿实验交付的下机原始测序数据及质控信息为研究对象，利用计算机硬件、生物信息学软件、算法及数据库等资源，通过数据质控、序列比对、变异识别、功能注释、统计分析或人工智能预测等手段，完成数据处理与生物学信息解析的实验活动。

3.4

遗传基线 genetic baseline

指个体或群体在基因组层面的基础遗传特征，通常通过全基因组测序（WGS）或芯片技术获取单核苷酸多态性（SNP）、插入缺失变异（InDel）、串联重复序列变异（TR）、拷贝数变异（CNV）、结构变异（SV）等遗传变异信息。

3.5

表型关联 phenotype association

指特定遗传变异信息与可观测生物学特征之间的统计学或功能性联系。

3.6

应用转化 application translation

将遗传基线与表型关联的研究成果转化为临床诊断、治疗或预防策略的过程。

3.7

自建湿实验中心 internal wet lab/bench

指机构内部建立的湿实验专用的实验场地。

3.8

第三方服务公司 external service provider

提供实验检测、数据处理等服务的外部机构。

3.9

核心设备 core equipment

实验流程中不可或缺的关键仪器。

3.10

标准样品 reference material

具有一种或多种规定特性足够均匀且稳定的材料，已被确定其符合测量过程的预期用途。

3.11

真实样本 authentic specimen

直接从生物体（如人体、动物、植物或环境）采集的、未经人工修饰或标准化的原始实验物质。

3.12

过程记录参数 process tracking parameters

实验、生产过程中关键环节的各类信息记录，包括涉及人、机、料、法、环方面的日志、型号、编号、货号、批号、数值、状态等。

3.13

数字化管理系统 digital management system

用于管理组织数字化资产、流程和数据的软件系统。

3.14

全血 whole blood

从人体采集并保存在采血袋或采血管内的血液混合物，包含所有血细胞（红细胞、白细胞、血小板）和血浆成分。

3.15

白膜层 buffy coat

全血经离心后，位于红细胞和血浆之间的含有白细胞和血小板的薄层。

3.16

自动化存储系统 automated storage and retrieval system

实现生物样本自动存取、信息记录和环境监控的仪器集成系统。

3.17

自动化样本处理系统 automated sample processing system

自动完成生物样本稀释、分装、加样等前处理工序的仪器集成系统。

3.18

通量 throughput

单位时间内产生的数据量，或检测、处理的样本数量。

3.19

提取前处理方法 pre-extraction treatment method

在正式开始核酸提取流程前，针对特定待提取样本类型、样本保存条件或干扰物质状态，所采用的物理或化学预处理步骤，包括分离、切割、研磨、打孔、浓缩等。

3.20

提取方法 extraction method

用于从经过（或无需）前处理的生物样本（或其特定组分）中分离、富集并纯化目标核酸（DNA或RNA）的具体技术路线或操作原理，包括酚氯仿有机抽提、硅胶膜或二氧化硅颗粒吸附（固相可逆吸附法，磁珠或离心柱）、盐析沉淀法、阴离子交换法等。

3.21

自动化程度 degree of automation

在进行湿实验操作时，自动化设备干预和操作所需比例的等级。根据应用场景需求分为无自动化（全程人工）、半自动化（部分步骤自动化）、全自动化（流程完全由自动化平台执行）。

3.22

可行性 **feasibility**

在给定资源和约束条件下，完成特定目标或项目的可能性。

3.23

重复性 **repeatability**

在相同测量条件下，对同一被测量物进行连续多次测量所得结果之间的一致性。

3.24

稳定性 **stability**

在特定条件下，标准样品保持其规定特性值不变的能力。

3.25

单分子实时荧光测序 **single-molecule real-time fluorescent sequencing**

单个核酸被固定在零模波导孔底部的单个聚合酶捕获后，聚合酶延伸引物链添加新碱基，并产生连续脉冲荧光信号，经碱基识别获得核酸分子的碱基序列与修饰信息的测序方式。

3.26

单分子纳米孔链测序 **single-molecule nanopore strand sequencing**

单个核酸被纳米孔捕获后，碱基单元在马达蛋白或其他方式的控制下逐一通过纳米孔，并产生连续的电信号变化，该时序电信号经碱基识别获得核酸的碱基序列及修饰信息的测序方式。

3.27

测序片段质量值 **quality of read**

衡量单条测序片段中各个碱基被正确识别的平均概率，通常以数字值直接表示。

3.28

原始下机数据 **raw sequencing data**

测序仪直接输出的未经过滤、纠错的信号文件或未经比对过滤的原始碱基序列文件和质量分数。

注：1. 高通量基因测序平台通常以BCL格式和/或FASTQ格式存储。

2. 单分子基因测序平台存储格式包括“.fast5”、“.pod5”、“.bax.h5”、“.ccf5”、“.fastq”、“.bam”等格式。

3.29

下机质控报告 **sequencing quality control report**

测序运行结束后，由测序平台软件生成的运行质量统计报告。

3.30

MD5 校验 **message-digest algorithm 5**

基于MD5算法生成的散列值，用于校验数据传输或存储的完整性。

4 缩略语

下列缩略语适用于本文件。

A_{260/230}: 260 nm 与 230 nm 的吸光度比值 (Absorbance Ratio at 260 nm to 230 nm)

A_{260/280}: 260 nm 与 280 nm 的吸光度比值 (Absorbance Ratio at 260 nm to 280 nm)

bp: 碱基对 (base pair)

CNV: 拷贝数变异 (Copy Number Variation)

CAPA: 不符合项纠正措施 (Corrective Action and Preventive Action)

DNA: 脱氧核糖核酸 (Deoxyribonucleic Acid)

DIN: DNA 完整性质控数 (DNA Integrity Number)

EDTA: 乙二胺四乙酸 (Ethylene Diamine Tetraacetic Acid)

GWAS: 全基因组关联研究 (Genome-Wide Association Study)

GQN: 基因组质量数 (Genomic Quality Number)

GC: 鸟嘌呤和胞嘧啶 (Guanine and Cytosine)

HLA: 人类白细胞抗原 (Human Leukocyte Antigen)

InDel: 插入缺失变异 (Insertion and Deletion)

kb: 千碱基对 (kilo basepair)

LRS: 长读长测序 (即单分子基因测序) (Long Read Sequencing)

MPS: 大规模并行测序 (即高通量基因测序) (Massively Parallel Sequencing)

mM: 毫摩尔每升 (mmol/L) (millimolar)

N50: 组装长度分布的第 50 百分位数 (50th percentile of the assembly length distribution)

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

PCR-free: 不需要聚合酶链式反应 (Polymerase Chain Reaction-free)

PFGE: 脉冲场凝胶电泳 (Pulsed-Field Gel Electrophoresis)

PE: 双端 (Paired-End)

pH: 氢离子浓度指数 (Potential of Hydrogen)

QMS: 质量管理体系 (Quality Management System)

qPCR: 荧光定量聚合酶链式反应 (Quantitative Polymerase Chain Reaction)

Q20: 质量值 20 (Quality Score 20)

Q30: 质量值 30 (Quality Score 30)

RNA: 核糖核酸 (Ribonucleic Acid)

RNase: 核糖核酸酶 (Ribonuclease)

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

SV: 结构变异 (Structural Variation)

SOP: 标准化操作程序 (Standard Operating Procedure)

Tris-HCl: 三羟甲基氨基甲烷盐酸盐 (Tris (hydroxymethyl) aminomethane Hydrochloride)

TR: 串联重复序列 (Tandem Repeats)

UPS: 不间断电源 (Uninterruptible Power Supply)

WGS: 全基因组测序 (Whole Genome Sequencing)

5 总体要求

5.1 总目标及技术方案

大人群全基因组测序旨在解析设定范围内人群遗传变异特征，建立“遗传基线—表型关联—应用转化”的全流程研究体系。该体系包括样本与表型采集、湿实验产出测序数据、干实验完成数据分析三大环节，本文件主要规范湿实验环节的技术要求。

应根据项目应用方向、经费预算、目标样本个数等，选择明确的湿实验主体技术方案，可参照附录 A 执行。

5.2 湿实验设计原则

5.2.1 为确保湿实验交付数据的时效性，应根据项目完成时间要求、总体经费预算、能力建设要求等确定湿实验开展中心的个数和产出测序数据的湿实验机构。湿实验开展中心的数量包括 1 个或多个，湿实验机构包括自建湿实验中心、第三方服务公司或两者结合。

5.2.2 为确保湿实验数据质量的可靠性，应提前做好项目全生命周期的整体规划和预运行方案验证，宜固化样本分装与存储、核酸提取、文库构建、上机测序等湿实验环节的实施方案和质量控制标准、核心试剂耗材和核心设备，具体实施可部分参考 GB/T 30989—2014 相关内容。应在项目正式开展前，使用标准样品和少量真实样本进行湿实验产出测序数据质量可用性的评估。

5.2.3 为确保湿实验数据质量的一致性，宜减少湿实验开展的中心个数及实施过程的变量。湿实验在多个中心开展时，应使用相同批次的标准样品（和真实样本）进行多中心间湿实验质量一致性的监控。变更湿实验方法、核心试剂或核心设备前，应进行预实验以评估变更对湿实验成功率的影响和干实验分析结果的影响。变更前后的数据对比宜使用不低于 30 个样本（含标准样品和真实样本）。必要时，应通过算法进行系统性偏差校正。若通过校正仍不能消除系统性偏差，应终止统一的方案进行湿实验数据的产出与交付。

5.2.4 为确保湿实验数据的稳定性，应对湿实验各环节设定明确的质量控制要求。更换试剂批次、仪器设备时，宜加入不低于1个标准样品，进行不同批次、不同设备的湿实验数据质量稳定性监控。在容易出现样本管或样本板摆位错误和/或气溶胶污染的实验环节，宜设置空白对照以监控实验过程污染。宜建立样本个体识别 SNP 位点检测体系（如质谱检测、测序检测等）来纠正湿实验过程中可能出现的样本信息混淆的问题。

5.2.5 为确保湿实验全流程数据的可追溯性，应标准化湿实验的过程记录参数，定期备份记录数据或将记录数据录入实验室数字化管理系统。

5.3 环境与设施基本要求

5.3.1 湿实验区域应按功能分区，包括样本库区、核酸提取区、文库构建区、测序准备区、上机测序区等，宜实行单向流设计，避免交叉污染。文库构建涉及 PCR 扩增时，应进行 PCR 扩增前区、后区的分区设计。

5.3.2 各分区的基础设施（包括实验室面积、承重、电路、水路、网络等）和环境条件（包括实验室温度、湿度、洁净度等）应符合对应实验室功能的要求。实验室空间或设备有限时，在可避免实验室交叉污染的前提下，可合并基础设施要求相同的功能区（如样本前处理和核酸提取区合并、核酸提取区与文库构建前区合并等）。

5.3.3 全流程宜配备不间断电源（UPS），保障关键设备（包括超低温冰箱、基因测序仪等）在市电中断时维持不少于20分钟正常运行，并配置温湿度与断电实时监控报警系统。

5.4 实验室生物安全要求

湿实验场所应根据所操作人类生物样本的来源、类型、前处理状态和潜在生物危害程度开展风险评估，并确定相应的生物安全防护要求。实验室的设计、建设与运行应参考 GB 19489—2008 的相关规定。实验人员应根据生物安全风险评估结果和操作要求配备个人防护装备，并严格遵守实验室生物安全操作规程。涉及病原微生物相关操作的，应在相应生物安全等级的实验室中开展。

湿实验全过程的样本信息与测序数据应严格去标识化处理，宜采用唯一项目编码与唯一样本编号双层索引机制。测序数据的传输应采取加密措施，数据存储应具备访问权限控制与防篡改机制，防止数据泄露、丢失或未授权访问。涉及人类遗传资源的信息采集、保藏、对外提供及合作研究等活动，应符合《中华人民共和国人类遗传资源管理条例》的相关规定。

实验人员上岗前应接受生物安全专项培训，内容应包括实验室生物安全法规、生物危害识别与个人防护、废弃物处置及应急处置措施等。宜定期组织生物安全复训与应急演练。

5.5 质量体系与能力保障要求

5.5.1 湿实验实施单位应建立覆盖全流程的质量管理体系（QMS），至少应包含以下内容：

——人员培训与上岗考核制度；

——设备校准与维护规程；

- 试剂耗材验收与有效期管理；
- 环境监控记录；
- 不符合项纠正措施（CAPA）流程。

5.5.2 关键岗位（包括样本接收、核酸提取、文库构建、测序上机、质量检测等）操作人员应经标准化操作程序（SOP）培训与实操考核合格后持证上岗。

6 技术要求

6.1 样本接收与分装存储

大人群全基因组测序的常见样本类型有血液（包括全血、血细胞、白膜层）、唾液、口腔拭子，针对特殊人群（如：肿瘤队列、生育队列等）还有组织（包括新鲜组织、冰冻组织）、羊水、血斑等类型。若技术方案包含单分子基因测序，首选样本类型为血液。

大人群队列生物样本库的建设通常是自建湿实验中心的重要模块，其建设与管理规范可参考 GB/T 39766—2021、GB/T 39766—2021，相关质量及能力通用要求可参考 GB/T 37864—2019/ISO 20387:2018。

接收样本后，应根据样本类型、应用需求、保存需求等，选择样本的处置方式，包括：

- a) 立即进行样本存储；
- b) 或先分离后再进行样本存储；
- c) 或先进行样本分装后再进行样本存储；
- d) 或先进行样本分离并分装后再进行样本存储。

样本保存温度应符合样本类型需求，可参照附录 B 执行。若存储样本达到 100,000 管及以上和/或存储时间超过 2 年，宜采用自动化存储系统。若单次分装样本管数超过 96 管，宜采用自动化样本处理系统进行分装操作，选型可参考 GB/T 46302—2025。

分装样本管与原始样本管应分别带有唯一识别的管标签。管标签可采用手工打印粘贴方式或保管预印条码方式，宜根据分装体积、存储温度要求以及分装和存储设备自动化程度等匹配适用的样本分装管和管标签类型。存储样本管对应的样本信息及存储位置数据应准确无误，宜通过人工扫描（若适用）和双人复核、自动化设备自动扫描等方式实现管标签与样本信息及存储位置的关联记录，并应设计防差错措施避免样本架、样本管或样本板摆放位置错误。

6.2 核酸提取

应根据测序平台对基因组 DNA 质量的要求、待提取样本类型、提取前样本保存条件、提取通量要求等，选择匹配的前处理方法（若适用）、提取方法及试剂（盒）与不同自动化程度的核心设备。当湿实验主体技术方案为单分子基因测序与高通量基因测序相结合时，应以单分子基因测序平台对基因组 DNA 完整性的要求来选择提取方法，并确保提取核酸的总量至少同时满足一次单分子基因测序

标准建库与一次高通量基因测序标准建库的需要。半自动化或自动化提取宜采用磁珠法，试剂盒选型可参考 GB/T 40171—2021。宜使用标准样品或真实样本进行提取模块方案的可行性、重复性与稳定性验证，并确保提取核酸的测序数据质量符合应用需求。

应建立核酸质量控制的固定检测方法并制定质量分级标准，可参照附录 C 执行，具体实施方法可参考 GB/T 40664—2021、GB/T 40974—2021、GB/T 44467—2024/ISO 4307: 2021 和 GB/T 45214—2025 的相关内容。当样本量大且样本采集及核酸提取方案固化时，可选择基因组 DNA 完整性与纯度抽检、浓度全检的核酸质控策略。

当采用 96 孔板进行核酸提取时，宜设置至少 1 个无样本的空白孔，以监控整板文库是否存在操作污染和/或位置错乱。宜设置至少 1 个以人细胞系标准样品为待提取样本的标准孔，以监控该板核酸提取质量稳定性。

提取的基因组 DNA 宜溶解于含有 EDTA 的 Tris 缓冲液中（如：含 10 mM Tris-HCl (pH 8.0) 和 0.1~1 mM EDTA 的缓冲液），并避免反复冻融。当需要建立基因组 DNA 样本库时，宜对基因组 DNA 进行分装保存，宜按照单次使用量进行分装。

应根据预期存储时间和应用场景，选择基因组 DNA 存储温度与存储设备，可参照附录 D 执行。

6.3 文库构建

应根据应用方向和测序平台对全基因组测序文库质量的要求、待检测样本可提取基因组 DNA 的总量范围与建库通量要求等，选择匹配的建库方法及试剂（盒）与不同自动化程度的核心设备。在保证质量的前提下，宜采用步骤少的建库方法（如：PCR-free 建库法等）。宜使用标准样品和/或真实样本进行建库模块方案可行性、重复性与稳定性验证，并确保文库构建的测序数据质量符合应用的需求。对于需要混合文库上机的情况，应建立质控方法来测定湿实验全流程的样本标签串扰率，并确保其符合特定应用要求。

应建立文库质量控制的固定检测方法并制定质量分级标准，可参照附录 E 执行。当样本量大，且基因组 DNA 质量稳定、文库构建条件固化时，可选择文库片段抽检，浓度全检的文库质控策略。

当采用 96 孔板进行文库构建时，宜设置 1 个无核酸样本的空白孔，以监控整板文库是否存在操作污染和/或位置错乱。宜设置 1 个以人基因组 DNA 标准样品为待建库样本的标准孔，以监控该板文库质量。

纯化的文库宜溶解于匹配后续上机测序要求的溶液中，并避免反复冻融。通常，文库宜置于-20℃冰箱暂存，待数据交付且确认数据质量无问题后，可根据项目需求决定继续存储或进行销毁。

6.4 上机测序

应根据湿实验主体技术方案、预算、干实验数据质量要求等，选择匹配的测序平台及仪器型号、测序试剂盒及试剂版本。宜采用标准样品构建的文库进行测序模块方案可行性、重复性与稳定性验证，并确保测序数据质量符合应用需求。对于需要混合文库上机的情况，应根据单张测序芯片或单个

测序通道测序原始碱基数的平均产出值、单文库需产出的测序原始碱基数、所选定混合测序方案的标签拆分率均一性等，计算单张芯片可混合上机文库的数量。

应建立测序原始数据质量分级标准，可参照附录 F 执行。当预算允许时，宜按照一定频率将标准样品构建的文库与待检样本构建的文库混合上机，以监控全生命周期的数据质量稳定性。

6.5 数据交付

应根据需求建立下机数据交付标准、交付数据内容和数据交付方式。交付标准应至少包括单文库测序原始碱基数、单文库碱基识别质量百分比或测序片段质量值。交付数据应至少包括原始下机数据和下机质控报告。数据交付方式包括线下物理介质交付、线上网络传输交付和本地化部署交付。

当存在数据传输或拷贝操作时，应建立验证数据完整性验证方法（如：MD5 校验）。交付的数据应暂存本地存储，待需求方确认收到完整数据后再进行本地删除。

附 录 A

(资料性)

大人群全基因组测序主体技术方案

表A.1 提供了大人群全基因组测序主体技术方案及对应主要应用方向、推荐样本个数的分类参考。

表A.1 大人群全基因组测序主体技术方案分类

序号	主体技术方案		主要应用方向	推荐样本个数
	测序技术类型	有效测序深度		
1	高通量基因测序	超低深度 0.1 - 1×	群体等位基因频率估计、群体结构检测、选择信号扫描；适用于经费有限但需最大化样本量的 GWAS 预筛	≥10,000
2	高通量基因测序	低深度 1 - 10×	多样化人群遗传基线建立、群体特异性变异发现、遗传多样性评估；缺乏参考面板的欠代表人群研究。	≥1,000
3	高通量基因测序	中等深度 10 - 20×	常见+低频变异关联分析（GWAS/Burden test）、罕见单例变异关联研究、HLA 分型、人群参考面板构建	≥1,000
4	高通量基因测序	标准 / 高深度 30 - 50×	高精度 SNV、InDel、TR 全面检测、罕见变异精细鉴定、临床级 WGS（遗传病诊断）、癌症体细胞突变分析	≥500
5	单分子基因测序	中等深度 8 - 15×	群体规模结构变异（SV）、TR 系统鉴定与基因分型、转座元件活性分析、串联重复序列变异图谱构建	≥100
6	单分子基因测序	高深度 20 - 30×	二倍体基因组从头组装、复杂嵌套 SV、TR 精细解析、单倍型定相、罕见病致病 SV 鉴定	≥10
7	高通量+单分子联合测序	MPS 30× + LRS 8 - 15×	综合性人群遗传变异目录构建(SNV+SV+TR 全谱系)；先以 MPS 大规模覆盖，再以 LRS 补充 SV 和复杂区域	≥500 (MPS 全部样本) + ≥100 (LRS 代表性子集)
8	高通量+单分子联合测序	MPS 4 - 7× + LRS 10 - 20×	经费受限下的群体 SV+常见变异联合分析；MPS 低深度覆盖群体频率，LRS 聚焦 SV 发现	≥1,000 (MPS) + ≥50 (LRS)
9	高通量+单分子联合测序	MPS 30× + LRS 30×	高精度参考级人群基因组资源构建（如人群特异性泛基因组图谱）；兼具碱基级准确性和结构完整性	≥10

附 录 B
(资料性)

大人群全基因组测序不同样本类型保存技术要求

表B.1 提供了大人群全基因组测序样本的运输存储条件、适配测序平台和应用场景的分类参考标准。

表B.1 大人群全基因组测序样本类型及存储条件要求分类

序号	样本类型	运输及短期存储	长期存储	是否适配单分子基因测序平台	是否高通量测序平台	主要应用 大人群队列 类型
1	全血	干冰（仅运输） 4 ° C（≤ 72 h） 3.常温（含核酸稳定剂的采血管）	-80 ° C	是（需用EDTA管保存管4 ° C运输）	是	自然人群、遗传病人群、肿瘤人群、母婴人群
2	血细胞、白膜层	干冰（仅运输） 4 ° C（≤ 72 h）	-80 ° C （样本库常用全血分离后分装）	是	是	自然人群、遗传病人群、肿瘤人群
3	唾液、口腔拭子	干冰（仅运输） 4 ° C（≤ 72 h） 常温（含核酸稳定剂的唾液采集管）	常温/4 ° C （用稳定剂，1年~5年）； -20 ° C/-80 ° C（5年以上）	是（需验证可行性）	是	自然人群、遗传病人群
4	组织（新鲜、冰冻）	干冰（仅运输） 4 ° C（≤ 24 h）	-80 ° C	是（需验证可行性）	是	肿瘤疾病人群、遗传病人群（涉及体细胞嵌合体或特定组织特异性遗传代谢病时）
5	羊水	干冰（仅运输） 4 ° C（≤ 24 h）	-80 ° C	否	是	母婴人群
6	干血斑	常温（干燥）	常温（干燥）	否	是	母婴人群及拓展自然人群

注：短期存储时间指不超过 4 周，长期存储指超过短期存储时间。

附 录 C

(资料性)

基因组 DNA 质量分级标准

表C.1 提供了基因组DNA从浓度及总量、纯度、完整性进行质量分级的参考标准。

表C.1 基因组DNA质量分级标准

序号	质量等级	浓度及总量	纯度	完整性	使用建议及对下游可能影响
1	优质/合格 (High Quality/Pass)	荧光染料法定量： 提取浓度$\geq$标准建库最低浓度阈值，且提取总量$\geq$2次标准建库起始量阈值	1. 目测：DNA 样本透明澄清，无明显黏稠状，无明显色素或杂质。 2. 紫外分光光度计检测法： A260/280：1.7~1.9，A260/230：2.0~2.2 3. 电泳法：无明显RNA、蛋白条带污染。	高通量基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳：DIN或GQN：≥ 7.0 片段主峰：> 20 kb（无显著拖尾） 2. 1%~2%琼脂糖凝胶电泳：基因组条带集中在 23 kb 及以上，无明显拖尾或弥散。 单分子基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳：片段主峰或N50（若可测）：≥ 50 kb DIN或GQN：≥ 8.0（若可测） 2. 脉冲场凝胶电泳（Pulsed-Field Gel Electrophoresis, PFGE）：基因组条带集中在 50 kb 及以上，无明显< 20 kb 拖尾。	风险极低，可使用。 对文库构建质量无明显干扰，建库成功率高（建库模块无异常时）。

表C.1 基因组DNA质量分级标准（续）

序号	质量等级	浓度及总量	纯度	完整性	使用建议及对下游可能影响
2	中等/有风险 (Moderate Quality/ Risk)	荧光染料法定量： 1. 提取浓度$\geq$标准建库最低浓度阈值，且2次标准建库起始量阈值$>$提取总量$\geq$1次标准建库起始量阈值 2. 标准建库最低浓度阈值$>$提取浓度$\geq$风险建库最低浓度阈值，且提取总量$\geq$1次风险建库起始量阈值	1. 目测：DNA 样本轻微浑浊、淡黄色、轻微黏稠。 2. 紫外分光光度计检测法： A260/280:1.6~1.7 或 1.9~2.0, A260/230: 1.5~2.0 3. 电泳法：有轻微或明显 RNA（需加入 RNase 消化后重新定量）条带污染、有轻微蛋白条带污染。	高通量基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳： DIN 或 GQN: 5.0~6.9 片段主峰: 5 kb~20 kb 2. 1%~2%琼脂糖凝胶电泳： 基因组条带$>$50%集中在 23 kb，可见轻微弥散或部分降解条带。 单分子基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳： 20 kb $\leq$ 片段主峰或 N50（若可测）$<$50 kb DIN 或 GQN : 5.0~7.9 2. PFGE： 基因组条带$>$ 50%集中在 50 kb，可见明显低分子量（20 kb~50 kb）条带弥散。	风险中度，高通量基因测序需谨慎使用，通常不建议用于单分子基因测序。 对文库构建质量可能有干扰，存在建库失败风险。

表C.1 基因组DNA质量分级标准（续）

序号	质量等级	浓度及总量	纯度	完整性	使用建议及对下游可能影响
3	不合格/失败 (Low Quality/Fail)	荧光染料法定量： 提取浓度<风险建库最低浓度阈值，和/或提取总量<1次风险建库起始量阈值	1. 目测：DNA 样本明显浑浊/沉淀、深色（褐/黑/绿）、高度黏稠/拉丝、分层。 2. 紫外分光光度计检测法： A260/280: <1.6 或>2.0 A260/230: <1.5 3. 电泳法：有严重RNA、蛋白条带污染。	高通量基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳： DIN 或 GQN: <5.0 片段主峰: < 5 kb 2. 1%~2%琼脂糖凝胶电泳： 条带严重弥散状分布，无明显 23 kb 主带。 单分子基因测序平台（以下检测方法二选一）： 1. 微流控毛细管电泳： 片段主峰或 N50（若可测）<20kb DIN 或 GQN : <5.0 2. PFGE： 条带严重弥散状分布，无明显 50 kb 主带。	风险高，不可用。宜重新提取或重新采样。 会明显影响文库构建质量和导致建库失败。

附 录 D

(资料性)

基因组 DNA 存储技术要求

表D.1 提供了基因组DNA存储温度的技术要求参考标准。

表D.1 基因组DNA存储温度技术要求

序号	基因组DNA存储时间	用于高通量基因测序	用于单分子基因测序
1	短期 (≤1个月)。	4 ℃冰箱。	4 ℃冰箱。
2	中期 (1个月~3个月)。	-20 ℃冰箱。	-80 ℃冰箱。
3	长期 (≥3个月)。	-80 ℃冰箱或自动化存储系统。	-80 ℃冰箱或自动化存储系统或 -196 ℃液氮气相存储系统。

附 录 E
(资料性)
人全基因组文库质量分级标准

表E.1 提供了人基因组文库从片段分布、纯度、浓度及总量进行质量分级的参考标准。

表E.1 人全基因组文库质量分级标准

序号	质量等级	片段分布	纯度	浓度及总量	使用建议及对下游可能影响
1	优质/合格 (High Quality/ Pass)	高通量基因测序平台（推荐 PE150 测序）： 微流控毛细管电泳： 1. 主峰大小：选定测序平台品牌后，进行建库测序预实验，选定测序生信比对分析插入片段主峰在 300 bp~450 bp 范围的条件对应的湿实验中间产物质控主峰大小。 2. 片段分布：目的产物主峰单一、尖锐、呈对称的主峰（近似高斯分布），插入片段主峰在 200 bp~800 bp 范围内的片段占比不低于 90%。无杂峰或仅有极微量（< 5%）杂峰。 单分子基因测序平台： 高灵敏度脉冲场毛细管电泳或微流控毛细管电泳： 1. 主峰大小：主峰位于使用测序平台的预期范围（如单分子实时荧光测序 15 kb~25 kb，单分子纳米孔链测序 10 kb~35 kb）。 2. 片段分布：目的产物主峰单一、尖锐、对称，无杂峰；大片段拖尾少。DIN 或 GQN 不小于设定阈值（若可测）。短片段残留的阈值下限占比<5%。无明显引物或接头残留峰。	1. 目测：DNA 文库透明澄清，无明显磁珠残留。 2. 紫外分光光度计检测法（如适用）：A260/280：1.8~2.0，A260/230：≥2.0	荧光染料法定量或 qPCR 定量： 文库浓度 ≥ 测序平台标准上机文库浓度合格阈值，且文库总量 ≥ 2 次标准上机投入文库总量阈值	风险极低，可使用。 对测序质量无明显干扰，测序成功率高（测序模块无异常时）。

表 E.1 人全基因组文库质量分级标准（续）

序号	质量等级	片段分布	纯度	浓度及总量	使用建议 及对下游可能影响
2	中等/有风险 (Moderate Quality/ Risk)	高通量基因测序平台（推荐 PE150 测序）： 微流控毛细管电泳： 1. 主峰大小：选定测序平台品牌后，进行建库测序预实验，选定测序生信比对分析插入片段主峰在 250 bp~300 bp 或 450 bp~550 bp 范围的条件对应的湿实验中间产物主峰大小。 2. 片段分布：目的产物主峰明显但有轻微双峰和/或轻微拖尾，和/或插入片段主峰在 200 bp~800 bp 范围内的片段占比不低于 80%。有较明显（5%~15%）杂峰。 单分子基因测序平台： 高灵敏度脉冲场毛细管电泳或微流控毛细管电泳： 1. 主峰大小：主峰略低于使用测序平台的预期范围（如单分子实时荧光测序 10 kb~15 kb，单分子纳米孔链测序 5 kb~15 kb）。 2. 片段分布：目的产物主峰分布较宽，均一性一般。有轻微降解迹象，但仍有主峰。DIN 或 GQN 小于设定阈值的差值不大于 3（若可测）。短片段残留的阈值下限占比在 5%~30% 范围内。无明显引物或接头残留峰。	1. 目测：DNA 文库透明澄清，有轻微磁珠残留。 2. 紫外分光光度计检测法（如适用）：A260/280：1.6~1.8 A260/230:1.7~2.0	荧光染料法定量或 qPCR 定量： 1. 文库浓度 $\geq$ 测序平台标准上机文库合格浓度阈值，且 2 次标准上机文库投入总量阈值 $>$ 文库总量 $\geq$ 1 次标准上机投入文库总量阈值 2. 标准上机合格浓度阈值 $>$ 文库浓度 $\geq$ 风险上机最低文库浓度阈值，且文库总量 $\geq$ 1 次风险上机投入文库总量阈值	风险中度，需谨慎使用，需根据文库质量情况选择额外纯化和/或风险上机。 对测序质量可能有干扰，存在测序失败风险。

表E.1 人全基因组文库质量分级标准（续）

序号	质量等级	片段分布	纯度	浓度及总量	使用建议及下游可能影响
3	不合格/失败 (Low Quality/ Fail)	高通量基因测序平台（推荐 PE150 测序）： 微流控毛细管电泳： 1. 主峰大小：选定测序平台品牌后，进行建库测序预实验，选定测序生信比对分析插入片段主峰在小于 250 bp 或大于 550 bp 范围的条件对应的湿实验中间产物主峰大小。 2. 片段分布：无明显主峰，有明显双峰、多峰，和/或条带严重弥散，和/或插入片段主峰在 200 bp~800 bp 范围内的片段占比低于 80%。有明显（>15%）杂峰。 单分子基因测序平台： 高灵敏度脉冲场毛细管电泳或微流控毛细管电泳： 1. 主峰大小：主峰严重低于使用测序平台的预期范围（如单分子实时荧光测序 5 kb~10 kb，单分子纳米孔链测序 5 kb~10 kb）。 3. 片段分布：目的产物主峰严重弥散，几乎无长片段。DIN 或 GQN 小于设定阈值的差值大于 3（若可测）。短片段残留的阈值下限占比大于 30%，和/或有明显接头残留峰。	1. 目测：DNA 文库有混浊或肉眼可见杂质，有明显磁珠残留。 2. 紫外分光光度计检测法（如适用）：A260/280：<1.6 或 >2.0 A260/230：<1.7	荧光染料法定量或 qPCR 定量： 文库浓度 < 风险上机最低文库浓度阈值，和/或文库总量 < 1 次风险上机投入文库总量阈值	风险高，不可用。宜重新建库。 会明显影响测序质量和导致测序失败。
注：微流控毛细管电泳质控高通量基因测序平台湿实验中间产物主峰时，通常 PCR-free 文库推荐质控打断纯化后主峰或连接产物取 1 μL~2 μL PCR 纯化后主峰，PCR 文库推荐质控 PCR 纯化后主峰。					

附录 F
(资料性)

人全基因组测序原始数据质量分级标准

表F.1 提供了人基因组测序原始数据质量分级的参考标准。

表F.1 人全基因组测序原始数据质量分级标准

序号	质量等级	单张测序芯片或单条测序通道			单个测序文库			使用建议及对下游可能影响
		原始测序碱基数 (Gb)	碱基识别质量百分比 (%Q20 或 %Q30)	标签拆分率 (%) (若适用)	原始测序碱基数 (Gb)	碱基识别质量百分比 (%Q20 或 %Q30) 或测序片段质量值	GC 含量 (%) (若适用)	
1	优质/合格 (High Quality/Pass)	≥ 预期通量的 100%	1. 高通量基因测序: Q30 ≥ 85% 2. 单分子基因测序: 1) 单分子实时荧光测序: Q20 ≥ 90%; 2) 单分子纳米孔链测序: 测序片段质量值不小于测序平台规定的合格标准值。	拆分率 ≥ 90%	≥ 预期产量的 100%	1. 高通量基因测序: Q30 ≥ 85% 2. 单分子基因测序: 1) 单分子实时荧光测序: Q20 ≥ 90%; 2) 单分子纳米孔链测序: 测序片段质量值不小于测序平台规定的合格标准值。	43% ≥ GC 含量 ≥ 39%	风险极低, 可使用。生信分析碱基错配率低, 在全基因组不同 GC 区域的覆盖较均一。

表F.1 人全基因组测序原始数据质量分级标准（续）

序号	质量等级	单张测序芯片或单条测序通道			单个测序文库			使用建议及 对下游可能影响
		原始测序 碱基数 (Gb)	碱基识别 质量百分比 (%Q20 或 %Q30)	标签拆 分率 (%) (若适用)	原始测序 碱基数 (Gb)	碱基识别 质量百分比 (%Q20 或 %Q30) 或 测序片段 质量值	GC 含量 (%) (若适用)	
2	中等/有 风险 (Moderate Quality/ Risk)	预期通量的 80%~100%	1. 高通量 基因测序: 85%>Q30 ≥80% 2. 单分子 基因测序: 1) 单分子 实时荧光 测序: 90% >Q20≥ 80%; 2) 单分子 纳米孔链 测序: 测 序平台规 定的合格 标准值> 测序片段 质量值≥ 测序平台 规定的合 格标准值 的 95%	90%>拆 分率≥ 85%	预期产量的 80%~100%	1. 高通量 基因测序: 85%>Q30 ≥80% 2. 单分子 基因测序: 1) 单分子 实时荧光 测序: 90% >Q20≥ 80%; 2) 单分子 纳米孔链 测序: 测 序平台规 定的合格 标准值> 测序片段 质量值≥ 测序平台 规定的合 格标准值 的 95%	44%≥GC 含量≥ 43%或 39%≥GC 含量≥ 38%	风险中 度, 需 谨慎使 用或需 补测数 据或需 排查原 因。 生信分 析碱基 错配率 较高, 在全基 因组不 同 GC 区域的 覆盖较 差或存 在非人 类基因 组 DNA 污染。

表 F.1 人全基因组测序原始数据质量分级标准（续）

序号	质量等级	单张测序芯片或单条测序通道			单个测序文库			使用建议及 对下游可能影响
		原始测序 碱基数 (Gb)	碱基识别 质量百分 比 (%Q20 或 %Q30)	标签拆 分率 (%) (若适 用)	原始测序 碱基数 (Gb)	碱基识别 质量百分 比 (%Q20 或 %Q30) 或 测序片段 质量值	GC 含量 (%) (若适 用)	
3	不合格/ 失败 (Low Quality/ Fail)	<预期通 量的 80%	1. 高通量 基因测 序： Q30<80% 2. 单分子 基因测 序： 1) 单分子 实时荧光 测序：Q20 <80%； 2) 单分子 纳米孔链 测序：测 序片段质 量值<测 序平台规 定的合格 标准值的 95%	拆分率 <85%	<预期产 量的 80%	1. 高通量 基因测 序： Q30<80% 2. 单分子 基因测 序： 1) 单分子 实时荧光 测序：Q20 <80%； 2) 单分子 纳米孔链 测序：测 序片段质 量值<测 序平台规 定的合格 标准值的 95%	GC 含量 >44%或 GC 含量 <38%	风险 高，不 可用。宜重新 测序，并排查 原因（若是 唾液、口腔拭 子样 本，可根据实 际情况判断或 反馈客户是否 交付数据）。 生信分 析碱基 错配率 高，在全基 因组不同 GC 区域 的覆盖 差或存在非 人类基因 组 DNA 严重污 染。

参 考 文 献

- [1] GB/T 15000.2—2019 标准样品工作导则 第2部分：常用术语及定义
- [2] GB/T46942.1—2025 单分子基因测序 第1部分：术语
- [3] T/CPMA 001—2018 大型人群队列研究数据处理技术规范
- [4] T/CHIA 21.1—2021 组学样本处理与数据分析标准 第1部分：全基因组测序数据分析
- [5] T/LTIA 30—2025 大规模人群全基因组高通量基因测序数据质量评估方法
- [6] T/CITS 275—2025 临床下一代测序自动化与常规化技术要求
- [7] T/SZAS 103—2026 高通量基因测序试剂、耗材质检评级规则

