|  |  |
| --- | --- |
| ICS  | 11.100 |
| CCS  |

|  |
| --- |
|  |

C 44 |

团体标准

T/SZAS XXXX—XXXX

单基因遗传病临床全外显子组测序性能验证及室内质控规程

Validation and internal quality control of clinicial whole exome sequencing

XXXX - XX - XX发布

XXXX - XX - XX实施

深圳市标准化协会  发布

目次

[前言 II](#_Toc175922391)

[1 范围 1](#_Toc175922392)

[2 规范性引用文件 1](#_Toc175922393)

[3 术语和定义 1](#_Toc175922394)

[4 检测性能要求 3](#_Toc175922395)

[5 性能确认要求 4](#_Toc175922396)

[6 室内质控要求 6](#_Toc175922397)

[附录A（资料性） 临床全外显子组测序实验相关建议 7](#_Toc175922398)

1. 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。

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。

本文件由北京协和医院提出。

本文件由深圳市标准化协会归口。

本文件起草单位：北京协和医院、浙江省人民医院、大连市妇女儿童医疗中心（集团）、吉林大学第一医院、中国医科大学附属盛京医院、重庆医科大学附属儿童医院、云南省第一人民医院、深圳华大基因股份有限公司、深圳华大医学检验实验室、安诺优达基因科技（北京）有限公司、北京贝瑞和康医学检验实验室有限公司。

本文件主要起草人：张抒扬、姜丹、唐少华、唐美芳、邱玲、徐秀华、刘睿智、刘雅萍、何蓉、何晓燕、吕涛、郭丹、吴平、李云、张嘉文、刘鹏、吴祺航、唐玉婧、肖红豆、邱娇、蒋婷、郑乔松、沈鑫曌。

单基因遗传病临床全外显子组测序性能验证及室内质控规程

* 1. 范围

本文件规范了临床全外显子组测序性能验证及室内质量控制（室内质控）的过程，涵盖检测性能和室内质控的要求、性能确认的过程、方法和评估要求等。

本文件旨在为检测服务提供商及提供相关检测的各类检测机构，在开展全外显子组测序对人类基因组进行单基因遗传病变异检测时的质量控制程序提供规范性建议，包含但不局限于运行前性能确认及运行中室内质控的流程及要求，覆盖范围包括但不限于实验操作、数据分析、结果判读以及报告出具过程。

适用于以大规模并行测序技术为技术路线的单基因遗传性疾病全外显子组测序，适用于该类型检测服务提供商及提供相关检测的各类检测机构。

* 1. 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/T 19703—2020/ISO 15194:2009 体外诊断医疗器械 生物源性样品中量的测量 有证参考物质及支持文件内容的要求

GB/T 35890—2018 高通量测序数据序列格式规范

GB/T 42751—2023  信息技术 生物特征识别 高通量测序基因分型系统规范

GA/T 1693—2020 法庭科学 DNA二代测序检验规范

T/CHIA 21.2—2021  组学样本处理与数据分析标准 第2部分：全外显子组测序数据分析

T/SZAS 13—2019 基因组学数据集

* 1. 术语和定义

下列术语和定义适用于本文件。

高通量测序 high-throughput sequencing

能够一次并行对大量核酸分子进行平行序列测定的技术，通常一次测序反应能产生不低于100M碱基对的测序数据。

[来源：GB/T 42751—2023，3.1]

外显子组 exome

个体基因组中能够编码产生蛋白质氨基酸序列的DNA区域的总和，在人类基因组中往往呈不连续的片段分布，占整个人类基因组的1～2%。

全外显子组测序 whole exome sequencing, WES

利用探针杂交富集外显子区域的DNA序列，然后通过高通量测序，主要识别和研究与疾病、种群进化相关的编码区及其调控区域相关遗传突变的技术手段。

[来源： T/CHIA 21.2—2021，3.3，有修改]

临床全外显子组测序 clinical whole exome sequencing; cWES

利用序列捕获技术将外显子区域捕获并富集后进行高通量基因测序，针对公共数据库（如OMIM数据库）中明确致病的基因，对获得的全外显子组数据进行单基因遗传病相关的变异校正及解读信息注释，从而获得与受检者相关的遗传病变异信息的检测方法。

性能确认 validation

性能确认是指对未经确认或自建的检验程序，以及明显修改过的检验程序，通过提供客观证据对特定的预期用途或应用要求已得到满足的认定。临床实验室应根据相应临床需要和目前技术水平制定检验程序分析性能标准（可接受的性能指标限值），作为确认结果解释的主要依据。

Barcode

一段特征性的脱氧核苷酸段片段，在多样本混合测序时，充当识别特定样本来源的唯一编号。

[来源：GA/T 1693—2020,3.10]

Q20

测序数据中，碱基识别质量值大于20的碱基占所有碱基的比例。

1. 碱基识别质量值为20时，表示碱基的正确率为99%以上。Q20≥95%，则表示测序数据中95%及以上的碱基正确率为99%以上。

[来源：T/SZAS 13—2019,3.1.12，有修改]

Q30

测序数据中，碱基识别质量值大于30的碱基占所有碱基的比例。

1. 碱基识别质量值为30时，表示碱基的正确率为99.9%以上。Q30≥85%，则表示测序数据中85%及以上的碱基正确率为99.9%以上。

[来源：T/SZAS 13—2019,3.1.13，有修改]

平均测序深度 average sequencing depth

测序得到的落在目标区域的碱基数与目标区域长度的比值。

覆盖度 coverage ratio

捕获芯片的覆盖度指捕获范围占参考序列总区域的百分比。而外显子组测序数据中的覆盖度则是指将测序序列比对到参考序列上时，所有被比对到的区域占捕获芯片范围总区域的百分比。

1. 本文件中如无特殊注明，覆盖度均指外显子组测序数据中的覆盖度。

单核苷酸变异 single nucleotide variation; SNV

在基因组水平，由单个核苷酸位点的变异（替代、插入或缺失）所引起的脱氧核糖核苷酸序列变异。

[来源：T/SZAS 13—2019,3.1.2，有修改]

插入缺失型变异 insertion and deletion; InDel

在基因组的某个位置上所发生的小片段序列的插入或者缺失，插入或缺失片段的长度在50 bp以下。

[来源：T/SZAS 13—2019,3.1.3]

拷贝数变异 copy number variation; CNV

大片段缺失和大片段重复又叫拷贝数变异，即连续较长的序列发生了缺失或者重复，与插入缺失型变异的区别在于变异的长度更长。

线粒体变异 mitochondrial mutation

线粒体DNA（mtDNA）发生的变异，这些变异可以导致线粒体功能障碍或疾病。

杂合性缺失 loss of heterozygosity; LOH

包含拷贝数减少的杂合性缺失（copy losses-LOH, CL-LOH）和拷贝数中性杂合性缺失（copy neutral-LOH，CN-LOH）两种情况。在一对同源染色体上，位于相同基因座的两个等位基因中的一个发生缺失，导致的杂合状态丧失，即CL-LOH；由同源重组等多种机制造成的，未发生拷贝数变化的等位基因失衡，则为CN-LOH。

1. 本文件中如无特殊注明，LOH均指CN-LOH。

短串联重复序列 short tandem repeats; STR

短串联重复序列是由2到6个核苷酸组成的序列，在基因组中串联重复出现。

染色体非整倍体 aneuploidy

在正常的染色体组中，丢失或增加了一条或几条完整的染色体，如单体（丢失单条染色体）和三体（额外增加一条染色体）。

灵敏度 sensitivity

灵敏度，又称查全率或召回率，它是针对真实结果而言的，是指实际为阳性的结果，被正确检测到的概率，其公式为：灵敏度=TP/(TP+FN)×100%。TP指与标准变异数据集基因变异位点一致的检测基因变异位点个数。FN指标准变异数据集中包含、但未被检出的基因变异位点个数，即漏检的个数。

精确率 precision

精确率，又称查准率，它是针对检出结果而言的，是在所有检出为阳性的结果中，实际为阳性的比例，其公式为：精确率=TP/(TP+FP)×100%。TP指与标准变异数据集基因变异位点一致的检测基因变异位点个数。FP指与标准变异数据集基因变异位点不一致的检测基因变异位点个数，即误报的结果。

参考品 reference material; RM

在一项或多项特性上具有足够的均匀性和稳定性，用于校准、给其他物质赋值或提供质量保证的物质。

[来源：GB/T 19703—2020/ISO 15194:2009，3.4，有修改]

测序片段 reads

高通量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—2018，3.1]

基因型质量值 genotype quality; GQ

Phred格式的质量值，表示在该位点该基因型存在的可能性，该值越高，则Genotype结果正确的可能性越大。

* 1. 检测性能要求
		1. 样本质量控制要求

各实验室应综合考虑检测范围及目的，确定应包含的样本类型，如外周血、羊水、流产组织及基因组DNA等。

各样本类型的运输流程，应综合考虑样本采集器储存条件及样本本身的特性，且应包含采样器无破裂、漏液等可能造成的交叉污染和样本降解、样本量不足等问题的质量要求，应包含受检者签署的知情同意书、样本编码及送检信息清晰完整的要求，确保无混样风险。

样本质量控制要求应综合考虑样本特性，进行充分测试，评估各样本类型的建议采样量，以及各类型样本在检测流程中的结果准确性及稳定性。建议采样量应为文库构建环节DNA投入量的3～4倍，各样本类型的建议采样量可参考附录A。

* + 1. DNA提取质量要求

提取试剂的批号、名称、外观应符合说明书标示，有效期应在规定范围内。

提取后的DNA质量要求如下：

1. DNA总量应为文库构建环节DNA投入量的2倍以上；
2. 电泳检测DNA主带明显，无明显降解；
3. DNA吸光值A260/280比值为1.7～1.9。

大批量自动化提取时应包含污染质控，即空白对照提取浓度＜1 ng/μL。

* + 1. 文库构建及杂交捕获质量要求

文库构建及杂交捕获试剂的批号、名称、外观应符合说明书标示，有效期应在规定范围内。杂交后的文库DNA总量应为测序上机环节投入量的2倍以上，文库的片段大小、体积、浓度等均需要符合相应平台的上机标准。其中，MGI测序平台建议文库总量≥600 ng；Illumina测序平台建议文库浓度≥1 nM且文库体积≥20 μL。文库片段大小建议在200 bp～500 bp。

室内质控参考品应在建库环节加入，与临床样本一同参与建库及后续的全部实验流程。大批量自动化提取时应包含污染质控（即空白对照），每板或每批次实验需包含空白对照、阳性质控品及阴性质控品各一个，空白对照提取浓度应＜1 ng/μL，阴阳性质控品质量要求同4.3.1。

* + 1. 测序质量要求

不同的测序平台有不同的合格阈值，不同平台可参考测序试剂盒或测序平台说明书，制定能保证分析成功率的测序数据合格指标，以保证测序质量和靶区覆盖深度。测序的Q30应不低于85%。

* + 1. 信息分析质量要求

测序下机后的数据建议关注以下关键指标：平均测序深度、20×覆盖度、Q30和污染比例，这些指标的实测值应不低于各平台设置的最低要求且非胎儿样本性别应与送检信息一致。其中，目标区域的平均测序深度应根据检测策略对基因变异判定所需的数据深度的需求而定。

其余指标的建议合格标准为：测序下机原始数据经过滤后，Q30应不低于85%，20×覆盖度应不低于97.5%，生信环节提示无污染且非胎儿样本性别与基本送检信息一致。

* + 1. 遗传解读质量要求

变异注释所选用的数据库应至少包括人群频率数据库、疾病数据库、表型数据库，以及预测结果数据库。所有数据库均需要定期更新，公共数据库应至少一年检查一次更新。（建议选用的数据库参见附录A。）

变异注释的信息需至少包括：变异基本信息（染色体坐标、合子状态、基因名称、转录本、HGVS命名等）、群体数据（该位点在正常人群和本地样本中的频率）、基因元件与功能影响、基因相关的疾病信息、疾病数据库收录情况、软件计算与预测结果等。

位点筛选主要考虑三要素：所涉及疾病的相关特异性表型与临床主诉吻合、符合疾病遗传模式，以及致病性达到报告筛选标准，解读参考信息参见附录A。

针对不同的变异类型，各实验室应在临床大样本量验证数据的基础上，对拟报告位点建立起低质量位点判断阈值，以界定经筛选后的拟报告变异位点，哪些应进行额外的金标准方法学验证实验补充（如Sanger、qPCR、MLPA等）。低质量位点阈值示例参见附录A。

* 1. 性能确认要求
		1. 性能确认时机

实验室建立起cWES检测程序后，检测程序常规应用前应先进行性能确认。在此基础上，正常收样的检测实验室建议每年至少进行一次性能确认。

任何可能严重影响cWES检测程序性能的情况发生前，如发生试剂升级、试剂替换、实验流程变更、生信流程变更、解读流程变更及重要技术操作人员变更等情况，以及在出现重大技术性差错、重大不符合、客户反复投诉等情况下，应在检测程序重新启用前进行性能确认。

* + 1. 性能确认样本要求
			1. 性能确认样本来源

应参照以下原则选择用于临床全外显子组测序性能确认的样本：

1. 性能确认样本可使用过往临床样本或参考品。选择的临床样本应有可信的参比方法提供的位点数据，即使用金标准方法（如：使用Sanger验证SNV变异）、行业公认方法或经确认性能符合要求满足临床预期用途的方法（如：通过ISO 15189认可实验室使用的相同检测方法）检测过；
2. 除了临床样本，性能确认还推荐使用位点准确可信的独立第三方参考品，即由与检测程序提供方或试剂盒制造商无关的第三方制造，独立开发，不受仪器或试剂制造商的影响的第三方参考物质。参考品应足够均匀和稳定，能证明测量系统处于统计控制下时，能提供期望的可信结果。
	* + 1. 性能确认样本种类

用于临床全外显子组测序性能确认的样本应符合以下要求：

1. 性能确认样本应包含参考序列参考品、特定位点阴性及阳性参考品或临床明确诊断阳性样本。参考序列参考品需使用有公开参考序列信息的通用参考品（如NA12878和炎黄一号），或经多个二代测序平台交叉验证过，有可信外显子组数据可供参考比对的参考品；
2. 阳性参考品或临床明确诊断阳性样本选取时，变异类型应覆盖制造商说明书声明的检测范围内的全部变异类型，如SNV、InDel、线粒体变异、杂合性缺失（LOH）、短串联重复序列（STR）、外显子CNV、CNV及染色体非整倍体等。如为检测范围内未包含的变异类型，在选取阳性参考品时无需包含；
3. 特定位点阴性指挑选临床已知致病或疑似致病位点至少100个，选取经多个二代测序平台交叉验证或金标准方法验证过，在挑选出的位点上无致病变异的参考品或临床样本。挑选的已知位点应尽可能覆盖所有染色体，且应覆盖涉及疾病的不同遗传模式。
4. 因cWES检测范围广，阳性参考品或阳性临床样本，同样可能检出其他致病或疑似致病位点，因此性能确认时仅关注经验证过的准确可信的阳性变异位点，对于其他检出位点，不予计入评估范围中。
5. 因cWES检测范围广，表型正常人群同样会携带多个致病或疑似致病的遗传病变异位点，因此cWES检测中并无传统意义上的“阴性”样本。此处的特定位点阴性参考品或临床样本，特指挑选出的位点经验证过均为阴性，性能确认时仅关注挑选出的位点，对于其他检出的致病或疑似致病位点，不予计入评估范围中。
	* 1. 性能确认评估要求
			1. 准确性评估

选取参考序列参考品至少1种、特定位点阴性参考品或临床样本至少5种、阳性参考品或临床样本至少每种变异类型各1种，且阳性参考品或临床样本总数不少于10种（其余要求参见5.2）。性能确认样本使用待确认的检测程序进行检测，准确性评估合格要求为：

1. 各实验环节均需达到检测程序的质控指标；
2. 参考序列参考品：SNV灵敏度≥98%，SNV精确率≥95%；InDel灵敏度≥84%，InDel精确率≥74%；
3. 阳性参考品的阳性变异全部检出，阳性变异为制造商说明书声明的检测范围内的变异类型，如SNV、InDel、线粒体变异、杂合性缺失（LOH）、短串联重复序列（STR）、外显子CNV、CNV及染色体非整倍体等；
4. 特定位点阴性参考品或临床样本，经验证过的位点上全部无阳性变异检出。
	* + 1. 重复性评估（批内稳定性）

选取参考序列参考品至少1种、阳性参考品或临床样本至少2种，批次内重复建库三次，合格要求为：

1. 各实验环节均应达到检测程序的质控指标；
2. 将所有SNV及InDel检出结果按图1汇总，计算一致性比值作为重复性评价参数，该参数不得低于90%。
	* + 1. 重现性评估（批间稳定性）

选取参考序列参考品至少1种、阳性参考品或临床样本至少2种，分三个批次进行重复建库，合格要求为：

1. 各实验环节均应达到检测程序的质控指标；
2. 将所有SNV及InDel检出结果按图1汇总，计算一致性比值作为重现性评价参数，该参数不得低于90%。



1. 一致性比值验证
2. R1、R2和R3分别为同一参考品的3个重复检出的阳性变异位点数。一致性比值=g/(a+b+c+d+e+f+g) ×100%，即交集数除以并集数。
	* + 1. 结果判定

准确性评估、重复性评估及重现性评估均合格，则该实验室的cWES检测流程及其程序可被认定为性能合格。

* + - 1. 不合格措施

如性能确认不合格，应排查实验及分析流程，修正检测程序后重新进行性能确认。

* 1. 室内质控要求
		1. 室内质控时机

正常开展临床全外显子组测序检测的实验室，实验室应建立室内质控体系，实时监控检测程序的稳定性，确保检测流程稳定、检测结果真实可信。室内质控推荐使用第三方参考品，并应在每一批临床样本检测中加入，应进行文库构建环节的监控，并且该室内质控品应与临床样本一同进行后续的全流程检测实验及分析。

* + 1. 室内质控品选择

室内质控品应至少包含一种参考序列参考品及一种阳性变异参考品。阳性变异参考品建议每半年至一年进行一次更换，尽量覆盖不同的变异类型，其余要求同5.2.2性能确认样本种类。

* + 1. 室内质控评估要求
			1. 结果判定

室内质控参考品需与临床样本一并进行文库构建及后续全部实验流程。合格要求为：

1. 各实验环节均应达到检测程序的质控指标；
2. 参考序列参考品：SNV灵敏度≥98%，SNV精确率≥95%；InDel灵敏度≥84%，InDel精确率≥74%；
3. 阳性参考品的阳性位点全部检出。
	* + 1. 不合格措施

如室内质控不合格，应排查实验及分析流程，修正检测程序后重新实验，该批次临床样本结果以室内质控合格批次的重测结果为准。

1.
2. （资料性）
临床全外显子组测序实验相关建议
	1. cWES建议采样量

以300 ng的建库投入量为例，建议采样量要求如下：外周血≥2 mL，如为婴幼儿取外周血或脐带血≥1 mL；流产组织取≥100 mg的胚胎、胎儿组织或绒毛；羊水取≥10 mL；基因组DNA总量≥1 µg；口腔拭子＞4支；干血斑＞3个，单个直径＞1 cm。

* 1. 变异注释数据库及注释信息选择

变异注释所选用的数据库应至少包括人群频率数据库、疾病数据库、表型数据库，以及预测结果数据库。染色体坐标建议使用hg19参考基因组，转录本建议使用MANE转录本。人群频率数据库应包括常见的公用数据库，如千人基因组、dbSNP、gnomAD等，建议包含区域细分频率库（如ExAC东亚数据库East Asian）以及实验室基于自身样本构建的本地频率数据库。疾病数据库，应至少包括ACMG指南推荐的ClinVar和HGMD。表型数据库建议使用OMIM。预测结果数据库建议使用dbNSFP，建议使用Revel值综合反映预测结果。

* 1. 致病性解读参考信息

致病性解读主要参考ACMG变异解读指南、ClinGen序列变异解读工作组和英国临床基因组科学学会等对该指南细化的规则，以及变异位点相关文献。

* 1. 低质量位点阈值

不同测序平台、不同测序深度下，低质量位点阈值存在差异，各实验室应在临床大样本量验证数据的基础上，对拟报告位点建立起低质量位点判断阈值，以下为低质量位点阈值的示例：

1. 基于MGI测序平台，在200×测序深度下的低质量位点建议阈值为：SNV位点测序深度＜40×或SNV变异率＜40%；InDel位点测序深度＜60×或InDel变异率＜45%；纯合子或半合子位点测序深度＜40×或变异率＜85%；
2. 基于Illumina测序平台，在60×测序深度下的低质量位点建议阈值为：位点突变率＜35%或突变深度＜10×且GQ值(genotype quality)＜20；纯合子或半合子位点突变率＜85%或突变深度＜10×且GQ值(genotype quality)＜20。

