



中华人民共和国国家标准

GB/TXXXXX—XXXX

单分子基因测序 第2部分：基于纳米孔测序技术的单分子测序仪性能评价方法

Single-molecule gene sequencing - Part 2: Methods for evaluation of sequencer performance

(点击此处添加与国际标准一致性程度的标识)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX-XX-XX 发布

XXXX-XX-XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

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。

本文件是GB/T ****-****《单分子基因测序》的第2部分：测序仪性能评价方法。

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。

本文件由国家药品监督管理局提出。

本文件由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（SAC/TC136）归口。

本文件起草单位：

本文件主要起草人：

引 言

GB/T ****-规定了单分子基因测序的术语，测序仪性能评价方法，样本处理规范及质量要求，数据质控要求、分析流程及格式，以及软件要求。拟由5部分组成：

—第1部分：术语，第1部分为第2部分、第3部分等后续部分的单分子基因测序技术实施和应用的理论基础。目的在于界定单分子基因测序领域的术语和定义，规定具有连续测序特征的单分子基因测序主要参数指标术语和技术相关术语；

—第2部分：测序仪性能评价方法。第2部分是第1部分在测序仪性能评价方面的展开和具体实施应用。目的在于规定单分子基因测序仪的性能评价要求及试验方法；

—第3部分：样本处理规范及质量要求。第3部分是第1部分在样本处理规范方面的展开和具体实施应用。目的在于界定适用于单分子基因测序的样本类型及样本处理技术要求，规定样本质量要求及评价方法；

—第4部分：数据质控要求、分析流程及格式。第4部分是第1部分在数据质控方面的展开和具体实施应用。目的在于界定单分子基因测序数据质量的术语和定义、规定数据质量要求及评价方法，描述测序信号至碱基序列的分析流程及测序序列层级目录格式和数据整体格式；

—第5部分：软件。第5部分是第1部分、第2部分、第4部分在数据分析方面的展开和具体实施应用。目的在于对单分子基因测序数据进行生物信息学分析的软件结构组成、通用要求、专用要求和检验方法，及测序软件的计算机配置要求、接口规范等。

单分子基因测序仪 第2部分 测序仪性能评价方法

1 范围

本文件规定了单分子基因测序仪质量评价的要求，描述了相应的试验方法。

本文件适用于基于单分子实时荧光测序法、单分子纳米孔链测序法、单分子纳米孔标签测序法，具有连续测序特征的单分子基因测序仪的质量评价。

本文件不适用于：

- 利用多个分立步骤进行非连续测序的单分子基因测序仪的质量评价；
- 利用共标签进行短序列连接的长片段测序的基因测序仪的质量评价。

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 4793.1 测量、控制和实验室用电气设备的安全要求 第1部分：通用要求
- GB 4793.9 测量、控制和实验室用电气设备的安全要求 第9部分：实验室用分析和其他目的自动和半自动设备的特殊要求
- GB/T 14710 医用电器环境要求及试验方法
- GB/T 18268.1 测量、控制和实验室用的电设备 电磁兼容性要求 第1部分：通用要求
- GB/T 18268.26 测量、控制和实验室用的电设备 电磁兼容性要求 第26部分：特殊要求 体外诊断（IVD）医疗设备
- GB/T 34065 分析仪器的安全要求
- GB/T 11606 分析仪器环境试验方法
- GB/T 191 包装储运图示标志
- GB/T 29791.3 体外诊断医疗器械 制造商提供的信息（标示） 第3部分：专业用体外诊断仪器
- YY 0648 测量、控制和实验室用电气设备的安全要求 第2-101部分：体外诊断（IVD）医用设备专用要求。
- GB XXXXXX 单分子基因测序 第1部分：术语

3 术语和定义

下列术语和定义适用于本文件。

3.1 单分子基因测序 single-molecule gene sequencing

在单分子水平上对核酸分子进行连续碱基序列测定。

注1：通常基于光学或电学等信号转换成碱基信息的单分子长片段测序方式，区别于依赖核酸模板扩增实现碱基序列测定的桑格测序、大规模平行高通量测序等以及基于共标签进行短序列连接的单分子测序方式。

注2：单分子基因测序中，根据不同的测序平台和建库方式，可实现万级或兆级连续碱基直接测量。

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

单个核酸分子被纳米孔捕获后，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分子的碱基序列及修饰信息的测序方式。

3.3 纳米孔 nanopore

具有贯穿通道的纳米量级的孔道，是实现碱基序列编码为电信号的转换器。

注：通常分为生物蛋白孔道与固态孔道。

3.4 测序通量 throughput of sequencing

单次运行可获得序列信息的片段数量或可测定的脱氧核糖核酸和核糖核酸(以碱基表示)数量。

注：通量指标需标注完整单位，以明确通量描述的是片段数或碱基数。

[来源：GB/T 30989—2014, 3.21]

3.5 测序准确度 accuracy of sequencing

单次运行获得的质量合格的序列片段，其原始序列或经处理后的单分子一致性序列与参考序列的一致程度。

测序准确度的计算方法可用式(1)表示：

$$Accuracy = Matches / (Matches + Substitutions + Insertions + Deletions) \dots\dots\dots (1)$$

式中：

Accuracy——测序准确度；

Matches——比对正确碱基数；

Substitutions——替换错误碱基数；

Insertions——插入错误碱基数；

Deletions——缺失错误碱基数。

注：用于单条序列的评价。

3.6 平均准确度 average accuracy

单次运行获得的质量合格的所有序列片段，其原始序列或经处理后的单分子一致性序列经与参考序列比对后，所有序列片段与参考序列的一致程度。

注：用于所有序列的整体统计评价。

3.7 中位数准确度 median accuracy

将单次运行获得的质量合格的序列片段按测序准确度由高至低进行排序和累积计数，当累积数量刚好达到或超过总片段数的一半时计数的最后一条片段的准确度。

3.8 众数准确度 modal accuracy

单次运行获得的质量合格的各序列片段的测序准确度所做直方图中的最高峰对应的准确度。

3.9 测序覆盖度 sequencing coverage

待测样本的核苷酸序列检测结果覆盖于参考序列上的比例。

3.10 测序深度 sequencing depth

待测样本中某个指定的核苷酸被检测的次数。

3.11 碱基识别 base calling

测序过程中从电信号、光信号或其他由于测序反应而产生的信号转换成碱基序列信息的过程。

注：在纳米孔测序反应中一般采用机器学习或深度学习算法模型。

3.12 测序片段 read

经测序平台产生的碱基序列片段，比如从单分子进入纳米孔识别开始，到整条单分子过孔与识别完成生产的序列片段。

3.13 测序读长 read length of sequencing

单次运行可读取的质量合格的序列片段长度，通常以碱基数表示。

注：单分子基因测序评价指标包含最长读长、平均读长、读长N50等。

[来源：YY/T 1723—2020，3.4]

3.14 参考序列 reference sequence

序列片段对应的物种基因组序列。

4 要求

4.1 外观

仪器外观符合如下要求：

- 仪器表面不应有明显的凹凸、划痕、脱漆、流痕，金属部位不应有锈蚀及机械损伤；
- 仪器开关、按键应灵活、可靠，紧固件应紧固、无松动；
- 标识、符号应清晰、端正。

4.2 性能指标

4.2.1 测序通量

4.2.1.1 单机测序通量

按照各测序仪器厂商提供的测序样品制备流程和测序操作进行待测序样品处理、质控和测序，待测序反应完成后，测序仪使用测序信号收集器（根据标记信号的不同选用相应不同的收集器，如图像信号收集器中的高精度 CMOS，电信号收集器中的半导体传感器）。对测序信号完成搜集。再通过信号分析软件，对采集得到的信号进行分析，将信号信息转变为核苷酸序列信息。

检测国家参考物质基因组 DNA，测序仪单机所有芯片一次运行的通量应符合制造商要求。

注：本文件采用经国家权威机构批准发布的正常人细胞系提取的长片段人基因组 DNA、以及其他种属不同片段大小的基因组 DNA 作为质量评价指标的实物标准。

注：针对直接 RNA 测序的性能评价另行要求。

4.2.1.2 单芯片测序通量

检测国家参考物质基因组 DNA，单张芯片单次运行的通量应符合制造商要求。

注：本文件采用经国家权威机构批准发布的正常人细胞系提取的长片段人基因组 DNA、以及其他种属不同片段大小的基因组 DNA 作为质量评价指标的实物标准。

4.2.1.3 单位时间测序通量

检测国家参考物质基因组 DNA，单张芯片单次运行过程中，在制造商规定的单位时间内通量应符合制造商要求。

4.2.2 测序读长

4.2.2.1 测序读长范围

单分子实时荧光测序原理的制造商应规定检测的测序读长范围。

注：单分子纳米孔链测序不适用。

4.2.2.2 测序读长

检测国家参考物质基因组 DNA，单张芯片单次运行过程中，获得的质量合格的序列片段，片段读长 N50、最长读长、中位数读长、平均读长应符合制造商要求。

4.2.3 测序准确度

4.2.3.1 单分子一致性测序准确度

检测国家参考物质基因组 DNA，单张芯片单次运行的单分子一致性测序准确度应符合制造商的要求。

4.2.3.2 单次测序准确度

检测国家参考物质基因组 DNA，单张芯片单次运行的准确度可选下列其中一项，应符合制造商的要求：

- a) 中位数准确度
- b) 众数准确度
- c) 平均准确度

4.2.3.3 一致性准确度

在制造商规定的测序时长、测序覆盖率和测序深度下，检测国家参考物质基因组 DNA，比对率应符合制造商的要求，测序一致性准确度应不低于 99.0%；

* 注1：对限定只用于人基因组DNA检测的，可只进行人基因组DNA参考品检测；

* 注2：对限定只用于病原微生物检测的，可只进行细菌和病毒DNA参考品检测；

注：针对直接RNA测序的性能评价另行要求。

4.2.3.4 变异检测准确度

在制造商规定的测序时长、测序覆盖率和测序深度下，检测国家参考物质基因组 DNA，比对率应符合制造商的要求，与单碱基变异 SNP 和插入缺失 Indel 及结构变异 SV 参考数据集比较，SNP 和 Indel 及 SV 的准确率和灵敏度应当符合制造商的要求；

注：针对直接 RNA 测序的性能评价另行要求。

4.2.4 重复性

进行三次重复，应均能满足 4.2.1~4.2.3 的要求。

4.3 安全性

一般分析用单分子基因测序仪应满足 GB/T 34065安全性的要求，医用临床诊断用单分子基因测序仪应满足GB4793.1和/和YY 0648的要求。

4.4 电磁兼容性

一般分析用单分子基因测序仪应满足GB/T 18268.1的要求，医用临床诊断用单分子基因测序仪还应满足GB/T 18268.26的要求。

4.5 环境适应性

一般分析用单分子基因测序仪应满足GB/T 11606的要求，医用临床诊断用单分子基因测序仪还应满足GB/T 14710的要求。

4.6 软件功能

单分子基因测序仪一般配套随机软件使用，通常可实现包括仪器运行查询、仪器控制、信号采集、实时信号处理、实时结果统计、离线信号处理等功能。可采用其他序列数据分析软件对制造商规定的下机数据文件进行分析比对，其他序列分析软件应经过评估确认其适用性。

5 试验方法

5.1 外观

用目测法对外观进行检查。结果应符合 4.1 的要求。

5.2 性能指标

5.2.1 测序通量

5.2.1.1 单机测序通量

按照 4.2.1.1 规定的检测方法，结果应符合 4.2.1.1 的要求。

5.2.1.2 单芯片测序通量

按照 4.2.1.2 规定的检测方法，结果应符合 4.2.1.2 的要求。

5.2.1.3 单位时间测序通量

按照 4.2.1.3 规定的检测方法，结果应符合 4.2.1.3 的要求。

5.2.2 测序读长

5.2.2.1 测序读长范围

按照 4.2.2.1 规定的检测方法，结果应符合 4.2.2.1 的要求。

5.2.2.2 测序读长 N50

按照 4.2.2.2 规定的检测方法，结果应符合 4.2.2.2 的要求。

5.2.3 测序准确度

5.2.3.1 单分子一致性测序准确度

在相同测序深度，以适宜的统计方法分析，结果应符合 4.2.3.1 的要求。

5.2.3.2 单次测序准确度

按照 4.2.3.2 规定的检测方法，单张芯片单次运行的准确度可选下列其中一项，结果应符合 4.2.3.2 的要求。

- a) 中位数准确度
- b) 众数准确度
- c) 平均准确度

5.2.3.3 一致性准确度

在制造商规定的测序覆盖率和测序深度下，通过与参考数据集进行比对，根据公式（1）计算一致性准确度，结果应符合 4.2.3.3 的要求；

$$AC = \left(1 - \frac{MS}{MB} \right) \times 100\% \dots\dots\dots (1)$$

式中：

AC——一致性准确度；

MS——错配碱基数；

MB——指定目标区域基因组碱基数；

其中，MS 包含置换、插入和删除。

5.2.3.4 变异检测准确度

在制造商规定的测序覆盖率和测序深度下，通过与 SNP、Indel、SV 参考数据集进行比对，统计测序结果和目标参考序列错配的碱基数，根据公式（2）和公式（3）计算 SNP 和 Indel 及 SV 变异检测准确率、灵敏度，结果应符合 5.2.3.4 的要求；

$$P = \frac{T_p}{T_p + F_p} \dots\dots\dots (2)$$

式中：

P—SNP 或 Indel 或 SV 变异检测准确率；

T_p—测序数据检出与标准品 SNP 或 Indel 或 SV 参考数据集一致的变异位点总数；

F_p—测序数据检出的标准品 SNP 或 Indel 或 SV 但不存在于参考数据集的变异位点总数。

$$R = \frac{T_p}{T_p + F_n} \dots\dots\dots (3)$$

式中：

R——SNP 或 Indel 或 SV 变异检测灵敏度；

T_p——测序数据检出与标准品 SNP 或 Indel 或 SV 参考数据集一致的变异位点总数；

F_n——测序数据未检出与标准品 SNP 或 Indel 或 SV 参考数据集的变异位点总数。

5.2.4 重复性

按照 4.2.1~4.2.3 规定的检验方法，依照操作流程进行三次重复，分别统计相应的技术指标，结果均应符合 4.2.1~4.2.3 的要求。

5.3 安全要求

一般分析用单分子基因测序仪应满足 GB/T 34065 的要求，医用临床诊断用单分子基因测序仪应满足 GB4793.1 和 YY 0648 的要求，其结果应符合 4.3 的要求。

5.4 电磁兼容

一般分析用单分子基因测序仪应满足 GB/T 18268.1 的要求，医用临床诊断用单分子基因测序仪还应满足 GB/T 18268.26 的要求，其结果应符合 4.4 的要求。

5.5 环境适应性

一般分析用单分子基因测序仪应满足 GB/T 11606 的要求，医用临床诊断用单分子基因测序仪还应满足 GB/T 14710 的要求，其结果应符合 4.5 的要求。

5.6 软件功能

通过运行测试用例，检查软件功能应符合 4.6 的要求。

6 包装、运输和贮存

6.1 包装标签

6.1.1 仪器应该按照 GB/T 191 的有关规定进行包装。

6.1.2 仪器包装箱内应附有合格证、装箱单和产品使用说明书。

6.1.3 仪器标签或铭牌应符合 GB/T 29791.3 的规定。

6.2 运输

测序仪应该按照制造商的要求运输。

6.3 贮存

测序仪应该按照制造商规定的条件贮存。

参 考 文 献

附录 C 单分子基因测序仪 性能评价用参考品

编号	参考品名称	片段长度	类别	来源	评价指标
1	人全基因组 DNA	> 30k bp	人类基因组	外周血 B 淋巴细胞	SNP, Indel 和 SV 的准确率和灵敏度、一致性准确度
2	HPV11 型基因组 DNA	13k bp	病毒类	HPV 病毒重组质粒	一致性准确度
3	金黄色葡萄球菌基因组 DNA	> 30k bp	低 GC 细菌类	金黄色葡萄球菌	一致性准确度
4	铜绿假单胞菌基因组 DNA	> 30k bp	高 GC 细菌类	铜绿假单胞菌	一致性准确度
5	大肠埃希菌基因组 DNA	> 30k bp	长片段样本	大肠埃希菌	测序通量, 测序读长, 单次测序准确率, 一致性准确度