团体标准
《RNA定量测序技术规程》
编制说明书

2023年5月

《RNA定量测序技术规程》编制说明

一、工作简况

1、任务来源

本标准基于卫健委/科技部国家重点研发计划《医学生命组学数据质量控制关键技术研发与应用》（2018YFC0910200，2018年立项，2021年结题）和广东省重点领域研发计划《多组学整合技术研发及标准化组学数据质量控制技术的推广应用》（2019B020226001，2019年立项，2022年结题）两个项目的研究成果，本标准由深圳承启生物科技有限公司于2023年3月提出申报，根据中国国际科技促进会标准化工作委员会[2023]中科促标字第227号文件：关于开展《RNA定量测序技术规程》团体标准立项通知，正式确立立项，本项目计划编号为：C12023072。

2、目的和意义

随着精准医学的发展，转录组（RNA）测序在科研和医学检验中的应用越来越多，可望成为数千亿美元的大市场。但由于RNA容易降解，各种实验和分析流程众多，导致行业内一直存在保存运输条件严苛、结果不准确、不稳健的问题，同一个样品测序两次都可以出现较大的偏差，严重制约了行业的发展。另一方面，测得的海量数据必须用算法进行运算才可得到各基因表达量的信息，但目前行业内算法众多，基本都是国外的算法，准确性低，运算速度慢，现有研究已充分显示算法本身所造成的问题甚至大过于实验，造成了结果往往难以被验证和被重复。国外的算法无法高效地在全国产计算平台上完成计算尤其是在海关检验检疫、公共卫生事件应急、生物战防治等领域，产生了较为严重的安全问题。

目前国内外均无相关的标准流程，各转录组测序单位处于各自为战的状态。流程不通用，结果不可比。

本标准对样品采集、处理、运输、质控和检测以及数据产生、质量评估和数据分析全

流程进行标准化和规范化。该标准的特点是：流程标准化，可应对一定程度的RNA降解而保持结果的稳定性，对样品保存运输的要求低；核心试剂、仪器、软件分析算法和计算平台可做到100%国产化，具有完整自主知识产权，完全可应对贸易和技术封锁。

3、起草单位

本标准负责起草单位：深圳承启生物科技有限公司。

本标准参加起草单位：武汉承启医学检验实验室有限公司、暨南大学、华南理工大学、赛哲生物。

本标准主要起草人：张　弓、陈　洋、余　卓、王　鑫、卢少华、金静洁、杜红丽、卢　红、陈　杰。

参加单位工作分配：深圳承启生物科技有限公司负责标准方案的制定及网页发布；深圳承启生物科技有限公司、暨南大学、华南理工大学和赛哲生物参与了标准方案的实验测试和质量评估；暨南大学进行了标准起草和编制说明编写，其他参与单位对标准各个版本的全部内容提出编写和修改意见。

4、起草过程

4.1 起草阶段

2018年12月至2019年5月，标准起草单位组织相关技术人员对标准项目进行了预研，课题组成员广泛收集了国内外RNA定量测序技术的应用现状及技术发展趋势相关的标准、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；

2019年6月至2022年3月，标准起草单位对标准草案进行实验验证和技术论证并确立了RNA定量测序技术从采样、处理、运输、质控和检测以及数据产生、质量评估和数据分析全流程，形成了《RNA定量测序技术规程》的初稿，并于2023年3月以网页形式向社会发布（https://translatome.net/Resources/omicsStandard/index.html）；

2022年8月，国家卫生健康委医药卫生科技发展研究中心组织专家对本标准流程初稿

进行评审，参会专家做“基本完成主要相关考核指标”的评价，建议进一步加强相关数据采集与检测标准的认定，同时提出应加强应用示范的建议；

2022年9月至11月，编制组针对收集的意见，对标准进行了补充和完善。包括：1）增加了核酸提取和文库的质量检测标准；2）增加了使用MHCC97H标准数据集进行数据质量控制的内容；3）在武汉承启医学检验有限公司、广州瑞惠生物科技有限公司、深圳市人民医院临床医学研究中心和佛山普利盛生物科技有限公司等单位宣传和推广本标准，充分交流并进一步收集相关单位的反馈意见；

2022年12月，向广东省科技创新监测研究中心组织的专家组进行总结报告，专家组考察了标准实施场地，审阅了相关资料并进行了咨询。2023年4月，研究中心经讨论形成意见，意见内容如下：在算法、标准制定等方面取得了一定进展，部分研究成果已经成为行业应用的参考标准，受到国内外同行关注，项目开发技术已在多家机构推广应用，取得较好的评价和一定的经济效益；

本标准由深圳承启生物科技有限公司提出，由暨南大学主编，其他单位共同参与起草，于2023年3月提交团体标准《RNA定量测序技术规程》的立项申请书，经中国国际科技促进会标准化工作委员会及相关专家技术审核，于2023年3月30日正式发布立项公告，并于同日在全国团体标准平台发布立项公示。

2023年4月，中国国际科技促进会标准化工作委员会完成了对本标准内容的审核并提出对格式重排和规范书写的意见，编制组进一步参考标准书写规范，修改和完善了标准文本内容，形成了《RNA定量测序技术规程》团体标准征求意见稿和团体标准编制说明。

4.2 征求意见阶段

2023年5月，本标准由中国国际科促会标准化工作委员会，在全国团体标准信息平台面向社会进行公开征求意见。同时由编制组向相关单位进行定向征求意见。

4.3 审查阶段

4.4 报批阶段

二、标准编制原则、主要内容及其确定依据，修订标准时，还包括修订前后技术内容的对比

1、标准的编写原则

本标准的制定工作遵循“先进性、适用性、可操作性、实用性、规范性”的原则，依据 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》和 GB/T 20001.4-2015《标准编写规则 第 4 部分：试验方法标准》给出的规则编写。

2、标准主要内容

(1) 动植物组织、培养细胞、全血和细菌/真菌样本量标准；

(2) RNA样本运输标准流程；

(3) 总RNA核酸样本质量检测标准；

(4) 转录组建库样本质量检测标准；

(5) RNA核酸样品采集标准流程；

(6) RNA建库及测序标准化实验流程；

(7) 转录组数据分析标准。

3、制定本标准的基础和依据

本起草工作组或单位在该领域的既往工作情况和成绩：

（1）建立了超高精度比对 FANSe 系列算法

张弓教授研发的 FANSe 算法避开了对错配难以准确处理的 BWT 原理，采用数学上更稳健的 seed-hash 原理，将一个 read 切分成尽可能不重叠的 seed，只要有一个 seed 内部不含错配，即可保证找到其最优匹配位置，因而在数学上十分稳健，其错误率有数学保证[1]，绝大多数情况下可控制在 10-9 以下，可完美容忍 7～12%的错配率而保证得到数学上

的最优解[2]。实际测试表明，在基因组突变检测[3]、转录组基因表达检测[1]、新剪切变体检测[4]、非模式物种分析[5-7]等应用上，FANSe 算法以绝对优势击败 BWA，Bowtie 等算法，几乎完全符合实验验证。如此高的准确度，使高度精密和稳健的组学分析成为可能。30 个生物学样品中高丰度内参基因测序定量误差仅为 0.0053% [8]，全转录组定量生物学重复相关性 R2 高达 0.98，这是以往技术重复都不容易达到的高相关性[9]。另一方面，如此高的精度也可以大幅度降低测序通量的需求。使用 FANSe 算法，可以只使用一般转录组测序通量的百分之一，便可可靠定量细胞内单拷贝 mRNA，而如此低通量下所鉴定到的基因差异表达倍数，都可以被 qRT-PCR 所验证[10, 11]。这使转录组测序成本大幅度下降。

（2）分析速度刷新世界纪录的第三代 FANSe 算法

FANSe3 代算法，采用了全新的数据结构和并行策略，大幅度提升了匹配速度与 indel 处理效率，并针对 Intel Xeon E5 之后的 CPU 环形总线/Mesh 总线架构进行优化，在人全基因组测序数据的比对过程中比 BWA 快 7.5 倍以上，且能更充分利用 CPU 超线程的能力。FANSe3 在家用电脑单机上达成半小时 mapping 完人全基因组、半分钟人外显子组的惊人速度，同时精度不打折扣。至于转录组和翻译组测序，张弓教授团队已在 2017 年展示了平均 1 秒多分析一个转录组测序数据集的云平台，可同时并行处理分析海量数据。深圳承启生物科技有限公司已将 FANSe3 部署在云平台上，为公众提供准确快速的测序数据处理。在云平台的支持下，FANSe 的云平台版可以单机 5 分多钟完成 30x 人全基因组的分析（输出 SNV 列表），受邀在央视节目上进行现场演示。用户不必自己学习复杂的生物信息学知识，也不需要自己购买维护高性能计算集群，通过网络即可享受立等可取的测序分析。这将彻底改变目前业界分析成本高昂的现状。

（3）国际领先：FANSe 算法通过科技成果评价

2021 年 7 月 13 日，经科技部批准成立的第三方专业科技成果评价机构中科合创（北京）科技成果评价中心，组织专家以视频连线方式召开了科技成果评价会，对承启张弓教

授团队完成的“全自主高性能高稳健的核酸测序快速比对算法及云分析系统”进行科技评价。经过专家评审，FANSe 系列算法同时达到了极高计算速度与准确性的双重目标，具有超强的容错能力及远超传统算法的实验可验证性，在解决“精准医学不精准”的世界性难题上做出了重要的贡献。作为可商用的基因数据底层比对核心技术，其比对精度比欧美算法高出两个数量级，其云平台的投入应用得到了用户的一致好评，体现出良好的商业效益和社会效益。FANSe 系列算法完全自主研发，保障了中国的科研和精准医学的独立自主性。评价专家组一致认为，该项目具有很强的科学性、创新性和实用性，研究成果总体上达到国际领先水平。这是国家权威机构对 FANSe 的“官方认证”。

（4）FANSe3 文章成为 Phenomics 杂志 2022 年最受欢迎文章

2021 年 2 月 22 日，《表型组学》（Phenomics）期刊在线发表了暨南大学张弓和金静洁团队以及深圳承启生物科技有限公司合作题为《The Ultrafast and Accurate Mapping Algorithm FANSe3: Mapping a Human Whole Genome Sequencing Dataset Within 30 Minutes》的研究论文。该研究通过全新的数据结构和并行策略，开发了 FANSe3 代算法，该算法在高达 12%的错配率情况下可保证给出数学上的最优解，提供准确而快速的测序数据处理。文章发表一周年后，《Phenomics》期刊于 2023 年 2 月 8 日认证 FANSe3 文章为《Phenomics》杂志 2022 年最受欢迎文章。

（5）FANSe 可在纯国产计算机上高效分析测序数据

2022 年 5 月，FANSe 算法在纯国产平台上运行成功，CPU 是兆芯开先 KX-U6780A，固态硬盘来自长江存储，内存芯片来自长鑫。与之对比的是一台双 Intel E5V3 (共 28 核 56 线程)的服务器，都部署在承启基因测序分析云平台上。在常见的转录组、翻译组、细菌菌群宏基因组等应用中，纯国产计算机的性能仅比服务器慢 1%～12%，使用中几乎感觉不到其性能差异。由于 FANSe 的超高算法效能，这是纯国产计算设备在大规模测序数据处理上第一次在实际应用性能上几乎追平国际主流服务器，而其他算法由于并非完全自主开发，

较难针对国产 CPU 进行针对性的优化，远无法达到同等性能。此前，承启生物和华大智造联合发布了国产大规模测序的全流程，将测序仪、试剂、分析算法等全面实现国产化，不依赖国外知识产权，有力地捍卫了中国精准医学的自主权。如今，承启再进一步，凭借专门优化的 FANSe 算法，将纯国产计算设备的应用性能提高到比肩国际主流的水平，标志着中国的基因测序的安全性和自主可控达到一个全新的高度，不仅能支撑诸多精准医学应用，更能在检验检疫、流行病控制、生物战防治、法医鉴定等直接关系到国家安全的方面大显身手。

以下文献，对本标准的编制具有重要的参考作用。

1. Zhang, G., et al., *FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads.* Nucleic acids research, 2012. **40**(11): p. e83-e83.
2. Xiao, C.-L., et al., *FANSe2: a robust and cost-efficient alignment tool for quantitative next-generation sequencing applications.* PloS one, 2014. **9**(4): p. e94250.
3. Wu, X., et al., *Iterative genome correction largely improves proteomic analysis of nonmodel organisms.* Journal of proteome research, 2014. **13**(6): p. 2724-2734.
4. Mai, Z., et al., *Low-cost, low-bias and low-input RNA-seq with high experimental verifiability based on semiconductor sequencing.* Scientific Reports, 2017. **7**(1): p. 1053.
5. Luo, B., et al., *Revealing crosstalk of plant and fungi in the symbiotic roots of sewage-cleaning Eichhornia crassipes using direct de novo metatranscriptomic analysis.* Scientific reports, 2015. **5**: p. 15407.
6. Cao, X., et al., *Autoactivation of Translation Causes the Bloom of Prorocentrum donghaiense in Harmful Algal Blooms.* J Proteome Res, 2021. **20**(6): p. 3179-3187.
7. Zhang, M., et al., *Increased photosystem II translation efficiency as an important photoprotective mechanism in an Arabidopsis thaliana ecotype (Tibet-0) adapted to high light environments.* Environmental and Experimental Botany, 2021. **183**: p. 104350.
8. Liao, X.H., et al., *Enhancing co-translational folding of heterologous protein by deleting non-essential ribosomal proteins in Pichia pastoris.* Biotechnology for Biofuels, 2019. **12**.
9. Zhang, G., Y. Zhang, and J. Jin, *The Ultrafast and Accurate Mapping Algorithm FANSe3: Mapping a Human Whole-Genome Sequencing Dataset Within 30 Minutes.* Phenomics, 2021. **1**(1): p. 22-30.
10. Li, S.Y., et al., *In vitro biomimetic platforms featuring a perfusion system and 3D spheroid culture promote the construction of tissue-engineered corneal endothelial layers.* Scientific Reports, 2017. **7**.
11. Hsu, C.-L., et al., *Integrated genomic analyses in PDX model reveal a cyclin-dependent kinase inhibitor Palbociclib as a novel candidate drug for nasopharyngeal carcinoma.* Journal of Experimental & Clinical Cancer Research, 2018. **37**(1): p. 1-15.

4、实验内容

对 RNA 定量测序技术进行调研评估，搜集相关资料及研究结果，结合文献及众多科研项目，以及深圳承启生物科技有限公司内部标准化流程及评估标准，梳理出了具有代表性的关键步骤和通用流程。因此本标准规定了基于高通量测序进行转录组检测的主要内容，主要实验内容包括了：样品采集、保存和运输、RNA 提取处理和质控、RNA 建库及测序以及数据产生、质量评估和分析。

5、实际应用效果

通常高丰度 mRNA 的定量重现性（技术重复）以 CV 值（变异系数，等于标准差除以平均值）进行统计，通常为 10%。2019 年 2 月 12 日发表的文章中，华南理工大学和暨南大学的团队利用本标准流程，将转录组中高丰度 mRNA 的定量重复性做到了 CV=53 ppm，即 0.0053%。在 30 个菌株中稳定表达一个基因，进行 RNA 转录组测序。尽管各样本的测序通量相差 100 倍之多，但这个高丰度基因的表达量却惊人的一致，CV 仅为 53 ppm。在更为困难的 RNC-seq（翻译组测序）上，30 个样本的生物学重复，该高丰度 mRNA 的翻译定量 CV 也做到了 150 ppm，即 0.015%。该应用说明《RNA 测序定量技术》标准流程已经达到误差极小、重现性极高的 RNA 测序定量水平。相关研究成果《Enhancing co-translational folding of heterologous protein by deleting non-essential ribosomal proteins in Pichia pastoris》已经发表在《Biotechnology for Biofuels》杂志。

三、试验验证的分析、综述报告，技术经济论证，预期的经济效益、社会效益和生态效益

1、主要技术指标

（1）总 RNA 核酸样本完整度及纯度指标

确定起始材料总 RNA 的完整性是基因表达分析中的一个关键步骤，评估 RNA 完整性是

获得有意义基因表达数据的第一个关键步骤。RNA 的质量和数量是确保基因表达分析和其他基于 RNA 的下游应用准确性的重要因素。

总 RNA 样本的完整性应符合 GB/T40664 的要求。通过凝胶电泳分析总 RNA 的 rRNA 的 28S/18S（真核样品）和 23S/16S（原核样品）rRNA 比值进行判断，rRNA 条带亮度比值接近 2:1，rRNA 条带明亮完整，说明无明显降解。样本总 RNA 浓度应大于 1 ng/μL。RNA 样本纯度：OD260/280=1.8～2.0。

（2）RNA 建库前投入量指标

常见的动物、植物、真菌等物种，包括人、鼠、水稻、拟南芥、酵母等。推荐使用的投入总 RNA 量为 10 ng～1 μg。对于如水稻、拟南芥等植物 mRNA 丰度较低的物种，建议提高投入总 RNA 量至 1 μg～2.5 μg。当样本轻微降解可适当增加投入量，总投入量不超过 2.5 μg。

（3）PCR 产物片段大小及文库浓度指标

若插入片段为 250 bp 则 PCR 产物主片段在 330 bp 左右，插入片段为 150 bp 则 PCR 产物主片段在 230 bp 左右，条带颜色均匀。使用 Qubit® dsDNA HS Assay Kit 测定文库浓度。要求最终 PCR 产物的摩尔产量≥1 pmoL，不同片段大小的 PCR 产物对应的产量请参考表 1。

表 1 不同片段大小 PCR 产物 1 pmoL 对应产量

插入片段主片段大小（bp）	PCR 产物主片段大小（bp）	1 pmol 对应产量（ng）
150	230	152
250	330	218

（4）转录组测序数据质控指标

用于高通量测序的核酸类样本应根据 GB/T 3557（高通量基因测序结果评价要求）中 4.1 和 4.2 章节中的碱基测序准确率的要求进行质量控制。

（5）测序仪质控指标

使用 MHCC97H 细胞来源的 RNA 标准品进行转录组测序，对测序仪稳定性进行评估：测序数据使用 FANSe3 将 reads 比对到 NCBI refseq 标准参考转录组，FANSe3 设置参数为-E5 -I1，当 mapping 率＞70%，Q30＞80%，说明仪器和试剂状态良好，可正常进行转录组高通量测序。

2、主要试验和验证数据分析

按照本标准的方法，我们组织实施了关键性能指标的试验项目验证，实施验证的内容为对不同来源细胞株采集提取 RNA 后，我们设计了不同 RNA 富集方法、不同建库试剂盒、不同操作人、不同时空和不同测序平台上进行转录组高通量测序检测，并进行对比分析，来论证本标准流程的稳健性。

（1）RNA 完整性检测

培养 MHCC97H、HCCLM6、HCCLM3、A549 和 Hela 细胞株，连续传代培养 8～12 代，收集每一代中细胞样品。从每个样品中提取总 RNA，并通过琼脂糖凝胶电泳检查 RNA 质量确保它们没有降解。琼脂糖凝胶电泳检测结果显示细胞总 RNA 的核糖体大亚基 28S rRNA 和核糖体小亚基 18S rRNA 条带明亮完整，比值接近 2:1，说明 RNA 样品完整性好，未发生降解（图 1A-E）。

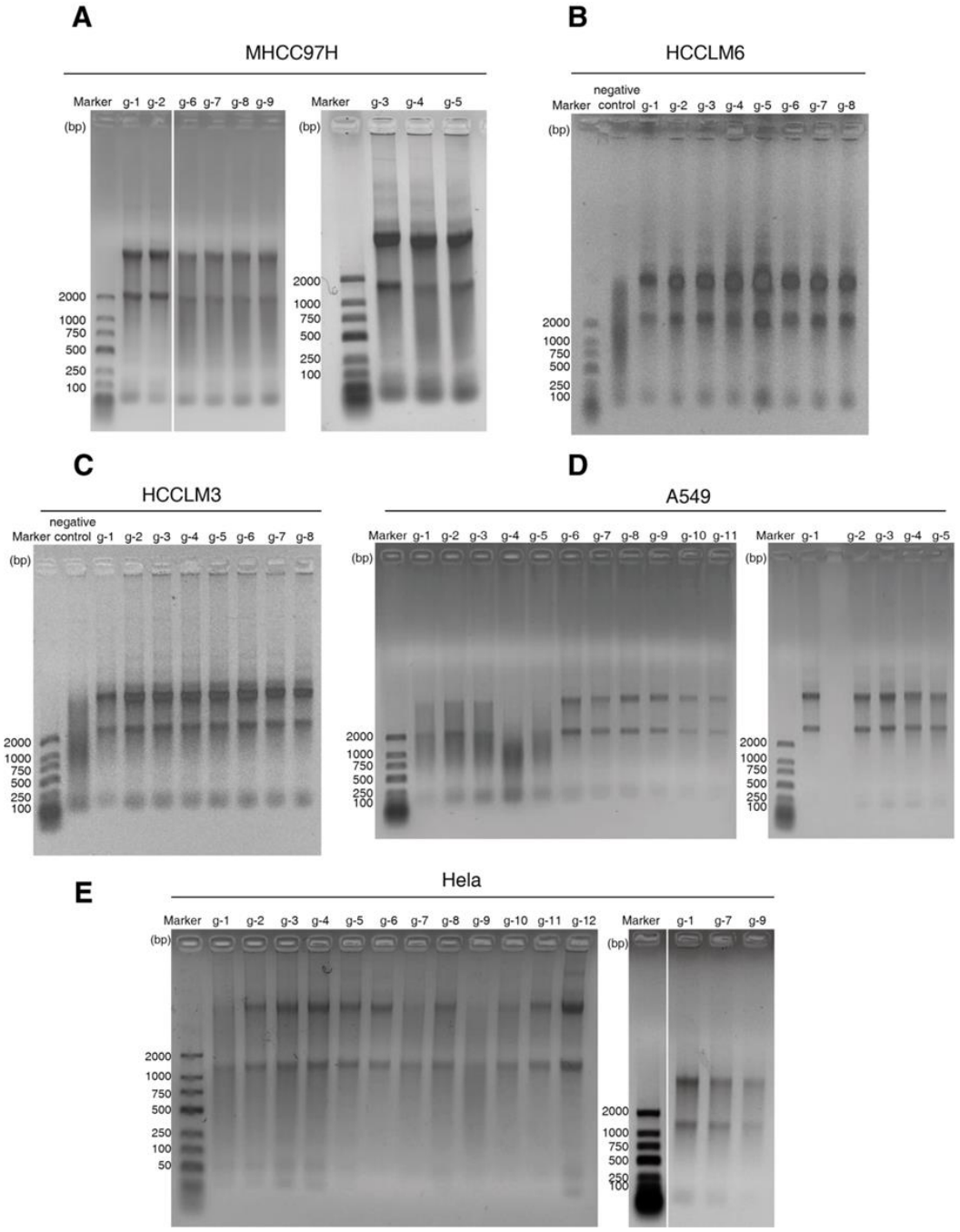


图1 五株实验细胞株RNA完整性质控检测

（2）转录组高通量测序数据质量检测

我们对转录组测序数据进行了质量检测，并生成了一系列质控指标（表 2）。测序数据集的原始数据和比对（mapping）数据整体质量令人满意：（1）原始测序数据的平均 reads 数>20M（表 2）；（2）原始测序数据的平均比对率为 73%（表 2）；（3）原始测序数据的平均 Q30（测序质量）为 89%（表 1）。值得注意的是，即使使用不同 RNA 提取方法、

不同品牌文库构建方法、不同降解程度的 RNA 样品和不同测序仪得到的测序数据质量也达到质量控制要求，说明实验流程和分析流程具有极强的稳定性。

表 2 转录组高通量测序数据质量检测

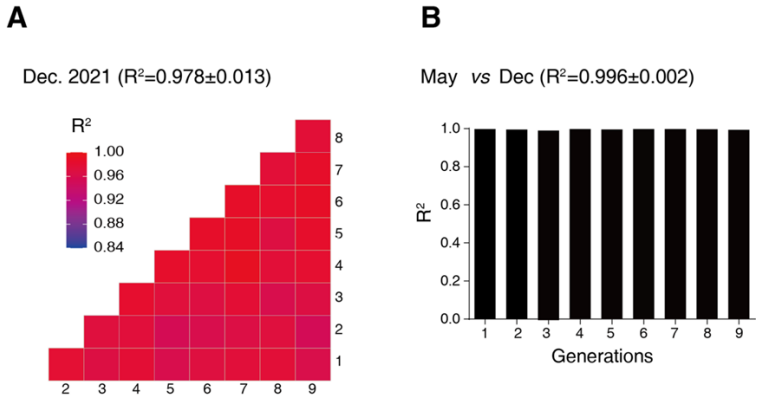
测序类型	样品类型	细胞代次	RNA提取方法	文库构建方法	测序数据量	比对率（%）	Q30（%）	测序平台及策略
未降解RNA的转录组测序	MHCC97H	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	42,050,104	78.63	88.18	MGI2000，PE150
		g-2			48,009,291	70.83	88.53	
		g-3			45,082,691	78.98	89.23	
		g-4			47,067,086	80.84	88.13	
		g-5			43,325,933	78.14	88.62	
		g-6			44,213,387	79.19	89.02	
		g-7			49,744,186	80.74	88.56	
		g-8			44,588,420	74.03	89.13	
		g-9			40,661,384	79.44	88.90	
	HCCLM3	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	22,452,336	82.68	94.70	MGI2000，PE150
		g-2			22,640,103	78.03	94.53	
		g-3			29,579,372	80.37	94.44	
		g-4			23,254,924	83.31	94.81	
		g-5			20,630,337	80.01	95.07	
		g-6			25,180,963	80.01	94.66	
		g-7			19,948,077	80.27	94.77	
		g-8			13,929,561	82.41	95.55	
	HCCLM6	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	20,700,126	86.93	94.56	MGI2000，PE150
		g-2			17,062,837	77.97	95.59	
		g-3			18,371,716	86.04	95.54	
		g-4			17,470,697	82.01	95.77	
		g-5			17,436,057	79.76	95.76	
		g-6			15,483,406	81.69	95.51	
		g-7			17,425,879	84.03	95.57	
		g-8			14,459,889	77.97	95.63	
	Hela	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	19,375,938	81.50	87.70	MGI2000，PE150
		g-2			41,795,149	82.28	90.64	
		g-3			19,817,541	82.20	88.13	
		g-4			34,359,124	82.17	86.87	
		g-5			12,396,219	81.83	84.97	
		g-6			17,183,738	81.55	90.45	
		g-7			17,056,778	82.64	85.35	
		g-8			20,734,377	80.80	89.54	
		g-9			15,792,131	83.44	81.60	
		g-10			14,744,576	82.35	87.41	
		g-11			17,604,495	83.79	88.14	

		g-12			43,790,435	77.64	82.46	
	A549	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	15,406,504	79.14	91.21	MGI2000, PE150
		g-2			14,871,959	81.77	88.48	
		g-3			19,157,100	82.31	85.61	
		g-4			15,340,056	81.90	87.55	
		g-5			18,131,540	81.91	89.96	
		g-6			10,146,087	77.97	80.83	
		g-7			11,822,422	78.96	83.92	
		g-8			12,922,652	78.31	82.88	
		g-9			14,187,863	78.45	86.11	
		g-10			18,343,988	80.90	89.68	
		g-11			15,785,336	82.73	83.68	
	MHCC97H-SAGENE Co. Ltd	g-1	Trizol法+mRNA磁珠捕获	VAHTS® Universal V6 RNA-seq Library Prep Kit for Illumina	12,922,652	84.11	92.15	Illumina NovaSeq-6000, PE150
		g-3			24,401,139	86.75	93.11	
		g-4			28,618,943	85.58	95.90	
		g-9			23,961,476	85.05	94.81	
	MHCC97H-Chi-biotech Co. Ltd	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	17,798,917	79.60	91.73	MGI2000, PE150
		g-3			16,996,034	78.96	89.95	
		g-4			15,919,486	78.69	91.17	
		g-9			18,151,614	78.69	91.49	
	MHCC97H-Dec. 2021	g-1	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	22,803,892	80.05	94.31	MGI2000, PE150
		g-2			20,045,340	73.58	94.92	
		g-3			17,677,312	72.23	95.61	
		g-4			22,868,070	82.10	95.02	
		g-5			20,301,562	80.51	95.22	
		g-6			23,712,158	80.74	95.30	
		g-7			27,098,829	82.04	94.86	
		g-8			24,297,942	76.24	95.27	
		g-9			29,961,926	80.20	94.73	
	MHCC97H-Ribominus	g-1	Trizol法+rRNA depletion法	MGIEasy mRNA文库制备试剂盒	22,257,405	48.98	94.55	MGI2000, PE150
		g-2			16,828,685	49.57	94.65	
		g-3			18,333,770	49.67	94.12	
		g-4			20,087,087	50.33	94.14	
		g-5			14,390,179	50.79	94.31	
		g-6			21,278,562	51.03	94.64	
		g-7			20,652,039	49.32	94.33	
		g-8			16,899,331	51.15	94.88	
		g-9			11,637,544	52.98	94.18	
降解RNA的转录组测序	MHCC97H-Ribominus_Deg	g-2	Trizol法+rRNA depletion法	MGIEasy mRNA文库制备试剂盒	15,456,998	56.18	94.00	MGI2000, PE150
		g-4			20,681,705	54.53	93.68	
		g-6			13,951,465	53.64	93.74	
		g-7			20,921,601	52.19	93.76	
		g-8			16,393,850	51.05	94.11	
		g-9			17,962,464	52.75	93.94	

	MHCC97H-PolyA+mRNA_Deg	g-2	Trizol法+mRNA磁珠捕获	MGIEasy mRNA文库制备试剂盒	15,137,022	77.18	88.50	MGI2000, PE150
		g-4			10,778,271	77.04	88.10	
		g-6			9,835,709	78.44	86.56	
		g-7			15,157,228	79.08	87.76	
		g-8			10,493,760	76.81	87.37	
		g-9			10,635,207	76.26	85.99	
	MHCC97H-RNaseA-Riboninus_Deg	NC	Trizol法+rRNA depletion法	MGIEasy mRNA文库制备试剂盒	24,198,150	54.35	84.46	MGI2000, PE150
		30s			27,395,623	39.61	83.86	
		1min			16,787,772	35.09	77.63	
		2min			24,095,943	35.24	75.43	
		5min			31,702,917	35.28	72.97	
		1h			29,052,190	54.60	84.36	

（3）不同时间批次下转录组稳定性检测

一个可靠的标准操作程序，产生的高通量组学数据可在不同技术人员，不同时间批次，不同实验室、不同平台以及不同数据分析方法下，始终产生相似的结果。我们分别在2021年5月和12月对两批 MHCC97H 细胞系进行了传代培养。对其转录组测序数据进行定量分析，发现每一批样本的基因表达情况是相似的，平均Pearson相关性R2=0.978±0.013，（图2A）。两个批次之间同一代次的细胞样品转录组基因表达定量分析结果均展现出较高的Pearson相关性R2= 0.996±0.002（图2B）。



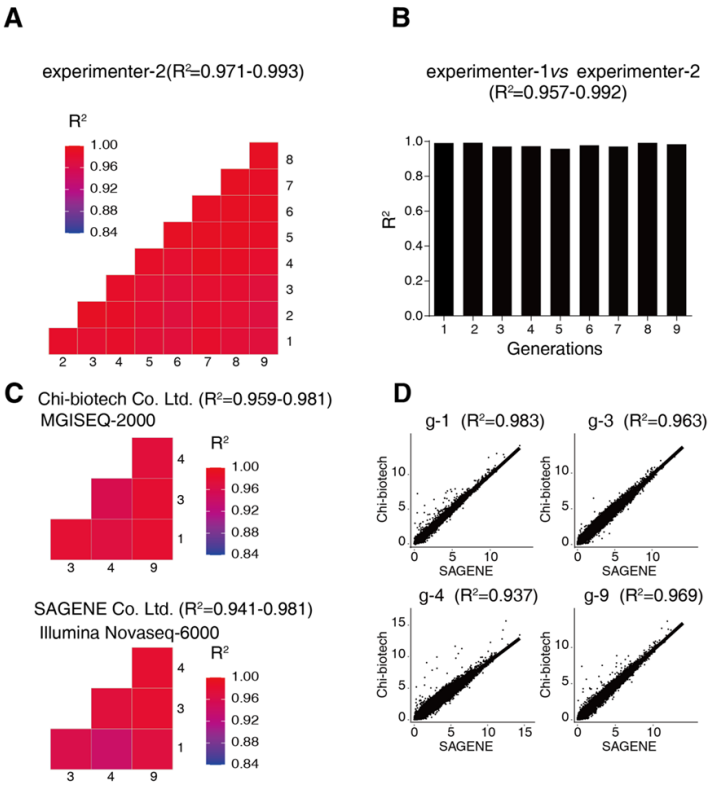
（A）2021年11月不同细胞代次的MHCC97H细胞系的Pearson相关性分析；（B）2021年12月和5月两批次传代MHCC97H细胞系同代相关性比较。

图2 不同时间批次下转录组定量稳定性比较

（4）转录组测序流程稳定性综合测评

我们测试了不同的实验员和实验室对 MHCC97H 转录组稳定性的影响。相同实验室的另

一位实验员对 MHCC97H 进行细胞培养，然后收集细胞 RNA，使用诺唯赞的 VAHTS® Universal V6 RNA-seq Library Prep Kit for Illumina 的建库试剂盒构建转录组文库。结果（图 3A～B）与前实验者（图 2A）几乎相同。此外，我们向 1000 多公里以外的两家商业测序服务商随机发送了 4 个代次的 RNA 样本。武汉承启医学检验实验室（Chi-Biotech Co. Ltd.）使用 MGISEQ-2000 测序仪，广州赛哲生物（Sagene Co. Ltd.）使用 Illumina NovaSeq-6000 测序仪。同样的，两家商业实验室也使用我们的标准操作流程构建转录组文库，并根据制造商的说明操作测序仪。Pearson 相关性 R2 分别达到 0.959～0.981 和 0.941～0.981（图 3C），两个商业化实验室的 4 个跨代次样品的基因表达相关性相似（图 3D）。这些结果表明在本标准流程的应用下， 在不同的实验室、不同操作员和不同测序平台的操作和检测下都有很好的稳定性以及数据可重现性，说明流程的稳定和可靠。

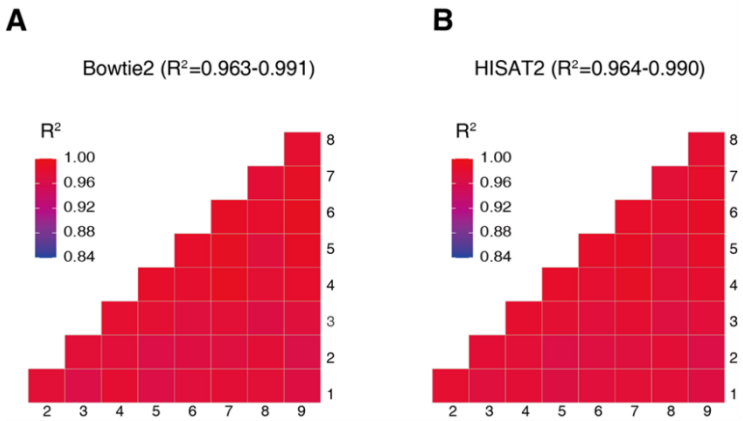


（A）另一名实验员构建的MHCC97H转录组文库；（B）两个实验者测试下的MHCC97H转录组基因表达定量相关性分析；（C）随机选择MHCC97H细胞4个代次的RNA样品（第1、3、4和9代）分别分发给两个商业测序服务公司，对每个实验室内转录组定量分别进行Pearson相关性分析；（D）MHCC97H每代基因表达定量跨实验室Pearson相关性分析。

图3 不同实验员和不同平台MHCC97H转录组定量稳定性比较

（5）不同分析算法转录组定量稳定性检测分析

我们研究了比对和定量算法对 MHCC97H 转录组基因表达稳定性的影响。本研究使用的算法是暨南大学张弓教授开发的 FANSe3 算法对基因表达进行定量，在 MHCC97H 连续多代次的转录组数据集中定量相关性达到 R2=0.966～0.995（图 2A）。除此之外，本研究对目前转录组测序数据分析广泛使用的主流比对和定量算法 Bowtie2 + featureCounts 和 HISAT2 + featureCounts 进行了测试。相对于 FANSe 算法，这些算法显示出略低的相关性，Bowtie2+ featureCount 的 Pearson 相关性为 R2 = 0.963～0.991，HISAT2 + featureCounts 的 Pearson 相关性为 R2 = 0.964～0.990（图 4A～B），表明这些算法不如 FANSe 稳健。



（A）使用Bowtie2 + featureCounts算法分析原始测序数据集；（B）HISAT2 + featureCounts 算法分析原始测序数据集。

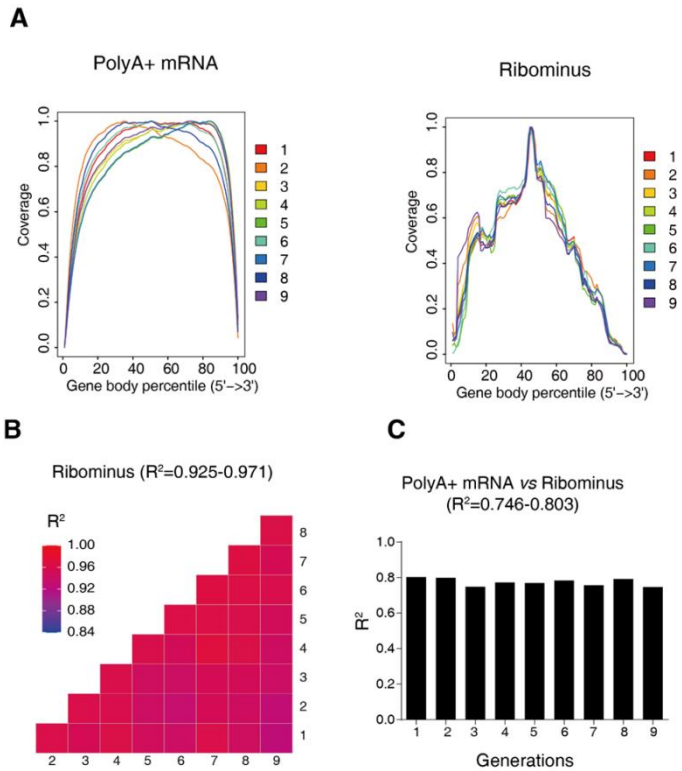
图4 分析算法对转录组一致性的影响

（6）不同建库方式转录组稳定性检测分析

不同的转录组建库方法会对测序数据产生影响。例如，在比较短插入片段法和长插入片段法构建转录组文库的研究中，显示出两种方法的在数据产生率、数据质量上均不同，使用长插入片段法构建的转录组文库在生物信息学上发现更多具有研究性的问题。

我们测试了使用不同的 mRNA 富集策略的建库方式对 MHCC97H 转录组稳定性的影响。我们的标准方案选择大多数研究中都广泛使用的 oligo-dT 来捕获 polyA+ mRNA（成熟 mRNA）。另一种对比策略为 rRNA 去除策略（也称为“Ribominus”或“Ribo-Zero”），它通过探针杂交去除 rRNA，然后进行磁珠提取或 RNaseH 消化。使用 rRNA 去除策略，Pearson

R2=0.925～0.971（图 5B），远低于 polyA+ mRNA 方法。此外，使用 oligo-dT 来富集 polyA+ mRNA 策略，读取的 reads 在基因组上的覆盖率比 rRNA 去除策略均匀得多（图 5A）。这些结果表明 polyA+ mRNA 的方法比 rRNA 去除策略更稳健，应尽可能优先考虑。正如预期，两种策略之间的基因表达相关性不高，R2 仅为 0.746～0.803（图 5C），建议如果使用两种不同策略产生的数据不应该混合进行分析。其次在设计课题时，仔细根据自己的实验目的选择合适的实验方法，并且同一个项目的实验流程保持一致也是获得可靠结果的重要因素。



（A）polyA+ mRNA-seq建库策略中reads在基因组上的分布情况（左），Ribominus建库策略中reads在基因组上的分布情况（右）；（B）使用rRNA去除策略（Ribominus）构建RNA-seq库进行转录组定量相关性分析；（C）PolyA + mRNA和Ribominus策略每代之间基因表达的相关性。

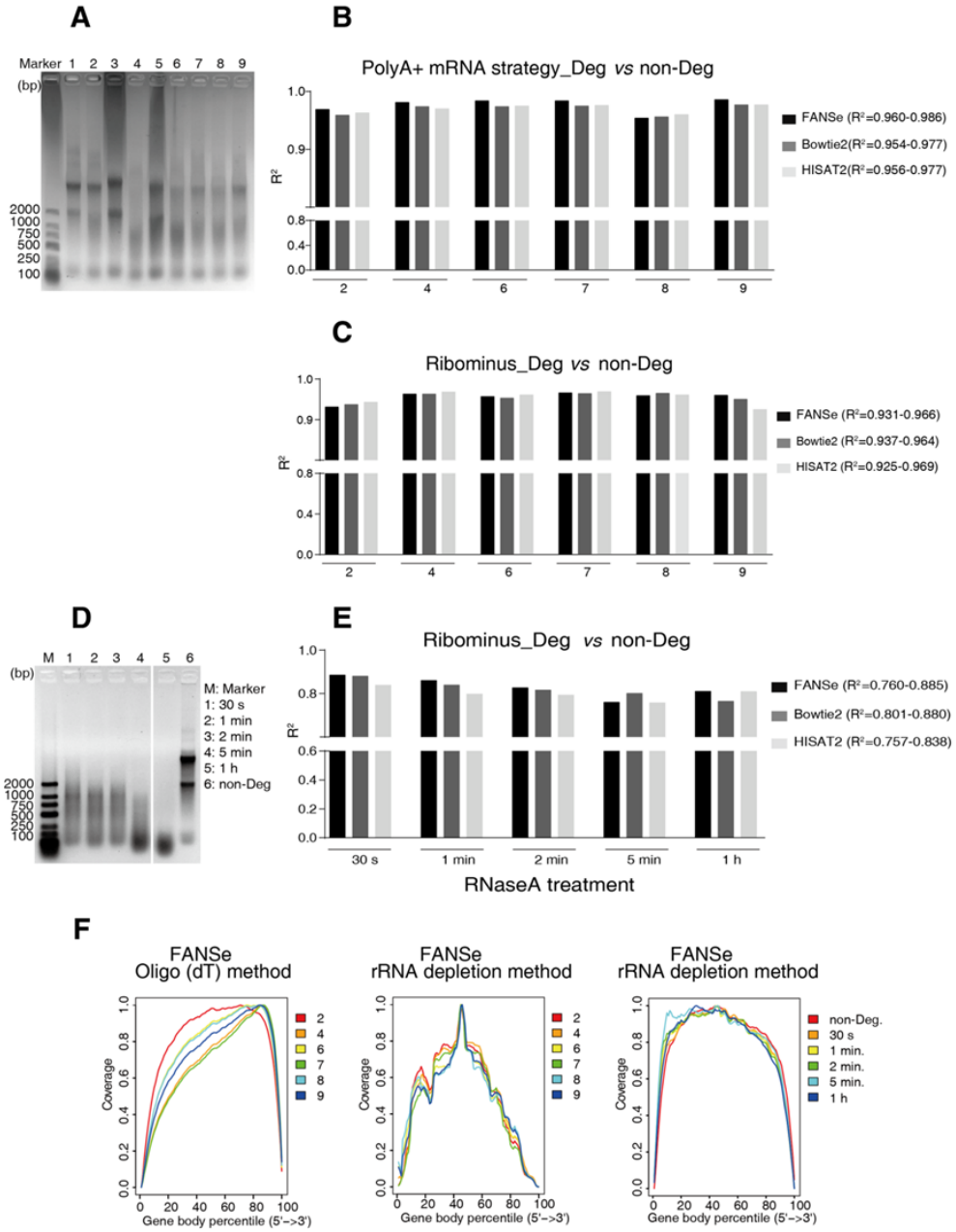
图5 使用不同建库方法测试转录组测序重复性

（7）降解和未降解 RNA 样品转录组建库方法比较分析

RNA 容易受到无处不在的 RNase 酶和环境变化（例如反复冻融）的影响。因此，在 RNA 样品的储存和运输过程中，可能会造成不可避免的轻微或不同程度的降解。我们测试不同降解情况下的 RNA 样本，以研究本标准流程是否适用于降解样品的转录组定量研究。首

先，我们创建了一个模拟环境即暴露在空气中导致 RNA 自然降解的场景：RNA 样本在室温下长时间暴露在空气中，这样环境中的 RNases 可能会进入 EP 管中并降解核糖核酸。然后将样品反复冻融。琼脂糖凝胶电泳显示 MHCC97H 的 1 到 9 代次总 RNA 有不同程度的降解（图 6A）。使用 oligo-dT 方法富集 mRNA，如预期所料，mRNA 因降解而断裂产生较明显的 3’端偏好性（图 6F 左侧）。意料之外的是，原本只适用于非降解样品的本标准操作规程，也可以直接用于降解样品的整个操作处理，并且无需在实验和分析中进行任何更改，同时可产生与非降解样品高度可比的基因表达定量值（Pearson R2>0.95，图 6B)。rRNA 去除策略避免了 3’端偏好性（图 6F 中间），并且也显示出与未降解样品有较高的一致性（R2> 0.92，图 10C），但仍低于 oligo-dT 方法。

大多数环境 RNase 是核酸外切酶，从 3'端开始切割 mRNA。然而，核酸内切酶可能将 mRNA 降解为更小的片段。我们将 RNase A 添加到未降解的 MHCC97H 总 RNA 中，并孵育 30 秒至 1 小时，以创建一系列不同降解程度的内切核酸酶降解的样品（图 6D）。由于核酸内切酶切割，使用 oligo-dT 策略不能成功构建测序文库，但使用 rRNA 去除的策略按照降解样本的建库说明流程构建文库则可成功构建文库。reads 在基因组上的的覆盖率比核酸外切酶产生的降解样本要均匀得多（图 6F，右侧）。然而，与未降解的对应物相比，核酸内切酶降解的样品显示出相当低的一致性。采用 FANSe3 算法，R2=0.760～0.885；而使用 Bowtie2/HISAT2 + featureCounts，相关性更低 R2=0.757～0.880（图 6E）。



（A）MHCC97H总RNA反复冻融+暴露在空气中后不同降解程度的琼脂糖凝胶电泳图；（B）降解及其相对应的未降解RNA样品使用polyA+mRNA的建库策略构建转录组；分别使用FANSe3、Bowtie2和HISAT2分析原始测序数据集；（C）降解及其相对应的未降解RNA样品使用Ribominus的建库策略构建转录组；分别使用FANSe3、Bowtie2和HISAT2 分析原始测序数据集；（D）设置30秒到1小时的时间梯度，通过添加RNase A人工降解MHCC97H 总RNA；（E）对RNaseA降解的RNA使用Ribominus策略建库，测试转录组稳定性；（F）测序reads在基因组上的分布情况（左：使用polyA+mRNA建库策略处理自然降解的RNA样本；中：使用Ribominus建库策略处理自然降解的RNA样本；右：RnaseA酶处理的RNA样本）。

图6 降解的RNA样本的转录组定量比较

3、真实性验证

根据标准要求，标准起草小组对关键指标参数进行了实验验证，实验验证样品由标准起草单位提供，将验证样品分别委托至武汉承启医学检验实验室（Chi-Biotech Co. Ltd）

和广州赛哲生物（Sagene Co. Ltd.）两家公司，两家公司均使用本文件规定的检测指标进行实验和测序，结果见表1、图3C和图3D。从以上结果可知，数据质量和相关指标均符合本文件技术要求，两家验证单位再重复性条件下获得的两次独立测定结果与起草单位测试结果一致，满足该标准要求。说明该标准流程稳定可靠。

4、经济与社会效益

（1）经济效益

将该标准流程应用于RNA测序定量分析，在近三年已为相关企业实现收入2000多万人民币。随着基因测序产业的不断扩大，预计将进一步提升相关收入，随着技术的不断成熟，预计在2023年实现收入1000万元。

（2）社会效益

生命数据分析已成为生物医药产业的重心与瓶颈，谁解决了此问题，谁就掌握了产业龙头。各种不同类型的、复杂的RNA样本对高通量测序全流程提出挑战，而不规范的检测方法极易造成数据的混乱及结果矛盾，严重影响了这一先进技术在各领域的推广应用。本标准的建立和普及能够使RNA测序定量技术流程做到规范化，进而产出可靠的数据结果。本标准中实验试剂、仪器和分析算法均采用具有自主知识产权的国产产品，完全可应对贸易和技术封锁。值得一提的是，分析流程中应用的FANSe高通量mapping算法是目前实验已验证假阳性、假阴性最少的基因二代测序数据分析快速mapping算法。作为可商用的基因数据比对核心技术，其精度比欧美算法高出两个数量级。另外FANSe系列算法还可以解决非模式生物二代测序应用所面临的参考库不准确问题。由于基于FANSe的分析结果的高度实验可验证性，FANSe以其精准的计算为基因组、转录组和蛋白质组学研究提供可靠的结果，很好地解决了当前大规模测序结果不能被验证、重复的危机。这尤其对需要极高准确度、稳健性和速度的医学临床应用提供了良好的支撑。FANSe算法对生命数据的有效挖掘与应用，将极大提升各领域的核心研发能力，且蕴含着巨大的创新机会，也是孵化相关领

域优秀创业企业的沃土。

四、与国际、国外同类标准技术内容的对比情况，或者与测试的国外样品、样机的有关数据对比情况

本标准没有采用国际标准，本标准在制定过程中未查到同类国际标准。

本标准未采用国内标准，本标准在制定过程中未查到针对性的国家标准、行业标准和地方标准。

本标准的总体技术水平属于国内外领先水平。

五、以国际标准为基础的起草情况，以及是否合规引用或者采用国际国外标准，并说明未采用国际标准的原因

无。

六、与有关法律、行政法规及相关标准的关系

1. 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》要求进行编写。

2. 参照相关法律、法规和规定，在编制过程中着重考虑了科学性、适用性和可操作性。

3.规范性引用文件包括：

GB/T 35537-2017　高通量基因测序结果评价要求

GB/T 40664-2021　用于高通量测序的核酸类样本质量控制通用要求

T/CHIA 21.4-2021　组学样本处理与数据分析标准 第4部分：转录组文库构建

七、重大分歧意见的处理经过和依据

无。

八、涉及专利的有关说明

无。

九、实施标准的要求，以及组织措施、技术措施、过渡期和实施日期的建议

等措施建议

建议本标准在批准发布6个月后实施。

本标准发布后，应向有转录组测序需求的各企、事业单位宣传和推荐执行本标准。

十、其他应当说明的事项。

无。