

转录组学公共数据获取

1. 公共数据库资源



2. 数据检索与下载



3. 数据格式与质量



4. 注意事项

- 1.1 NCBI GEO
- 1.2 EMBL-EBI ArrayExpress
- 1.3 Genome sequence archive

- 2.1 使用疾病名称、基因名称或实验类型等关键词检索
- 2.2 通过fastq-dump或prefetch命令可以从NCBI sequence read archive下载指定Accession号的fastq文件

- 3.1 使用fastqc等工具对数据质量进行初步评估：读长的碱基质量分布、序列复杂度、接头污染
- 3.2 对数据进行归一化或过滤

- 4.1 数据权限与引用
- 4.2 实验设计与元数据
- 4.3 数据整合与可比性

常见问题解析

转录组学数据获取的常见问题集中于三环节：1. RNA质量问题，如RIN值 <7 （因样本冻存/处理延迟致降解），需重新采集并全程控RNase；若含基因组DNA污染，需用DNase I消化后重提。2. 文库构建问题，接头二聚体过多（试剂比例失衡），需优化接头 / 模板浓度；PCR扩增偏差（循环数 >15 ），应减少循环并验证均一性。3. 测序问题，Q30 $<80\%$ （测序仪状态异常），需维护仪器并重测；数据量不足（文库浓度低），需调整浓度后补测，确保覆盖度满足后续分析需求。

参考文献

- [1] Song Y, Song B, Huang D, et al. Multimodal zero-shot learning of previously unseen epitranscriptomes from RNA-seq data [J]. *Brief Bioinform*, 2025, 26(4).
- [2] Wang J, Sun Z, Wang G, Miao Y. HPTAS: An Alignment-Free Haplotype Phasing Algorithm Focused on Allele-Specific Studies Using Transcriptome Data [J]. *Int J Mol Sci*, 2025, 26(12).