



浙江大學
ZHEJIANG UNIVERSITY

基于转录组测序数据的 T细胞受体重建方法评估

汇报人：吴佳欣

导师：刘琬璐 研究员

日期：2025.5.14



目录

01

研究背景及现状

02

研究目标与内容

03

研究结果

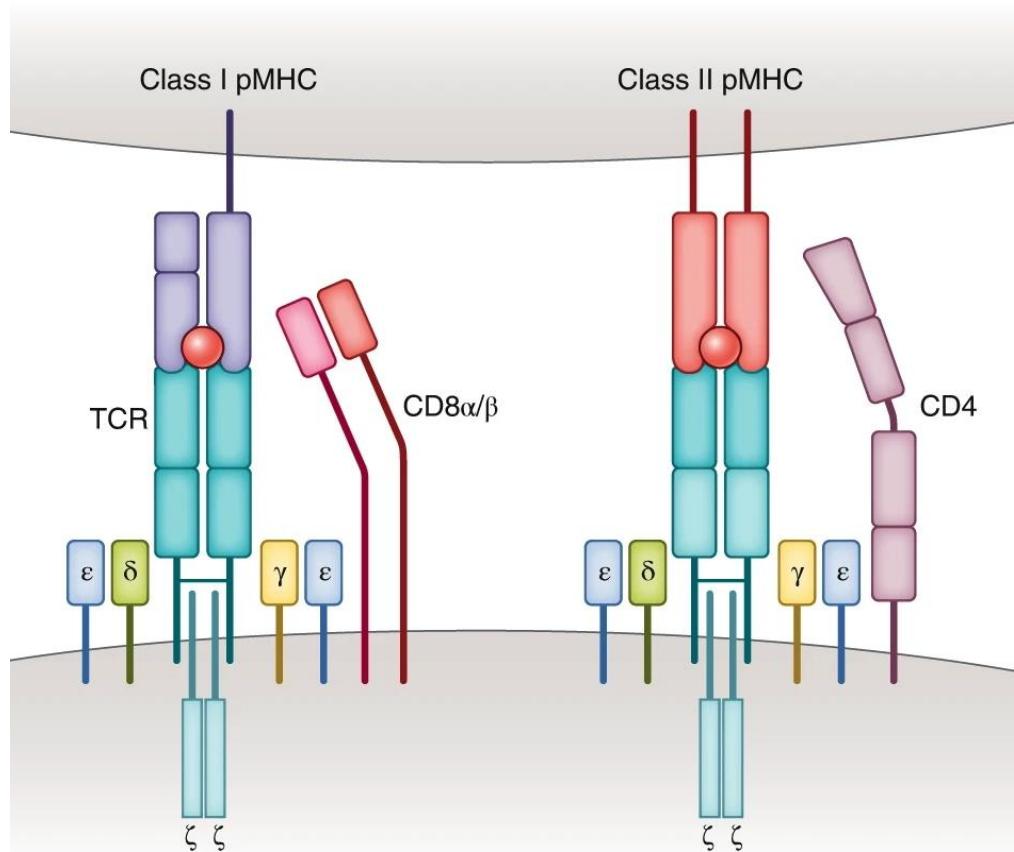
04

总结与展望

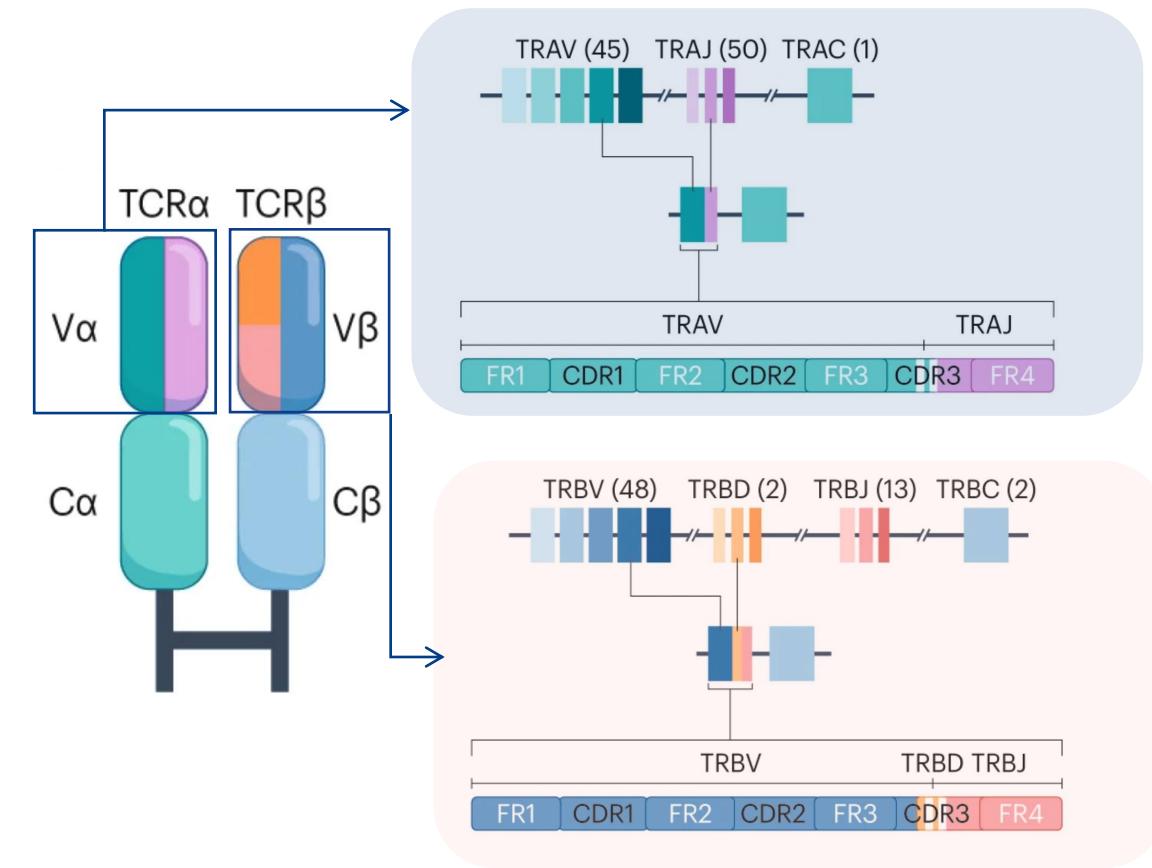
05

致谢

- T细胞受体(T Cell Receptor, TCR)是T细胞表面的特异性受体，负责识别由MHC所呈递的抗原；
- TCR多样性是适应症免疫精准应对多种病原体的基础，来源于V(D)J重排、轻链重链配对等机制。

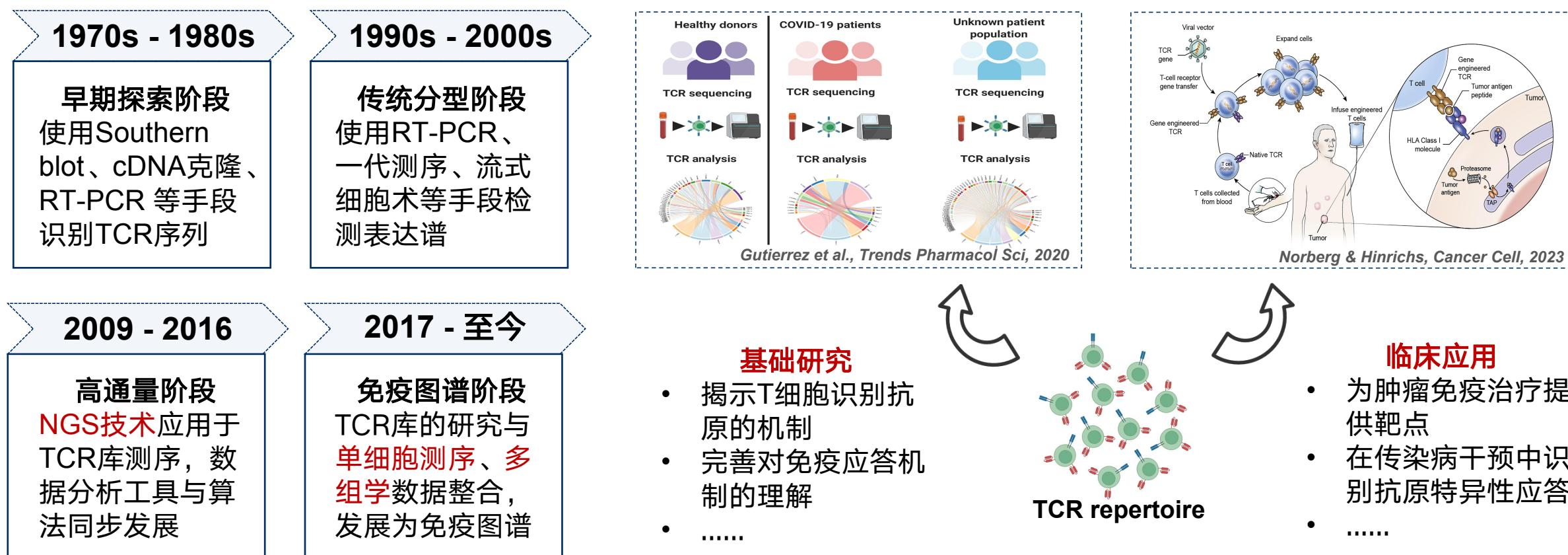


Joglekar & Li, *Nat. Methods*, 2021.



Mhanna, Vanessa, et al., *Nat Methods*, 2021.

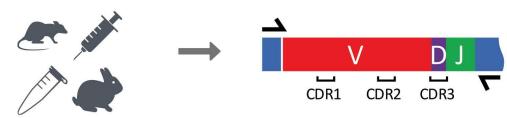
- TCR库反映了个体T细胞对潜在抗原的识别能力，体现适应性免疫应答的多样性和特异性；
- TCR库在基础研究和临床应用层面都有非常高的科学价值。



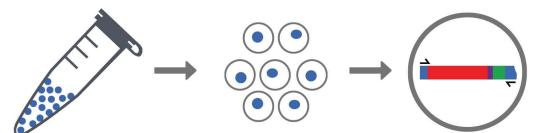
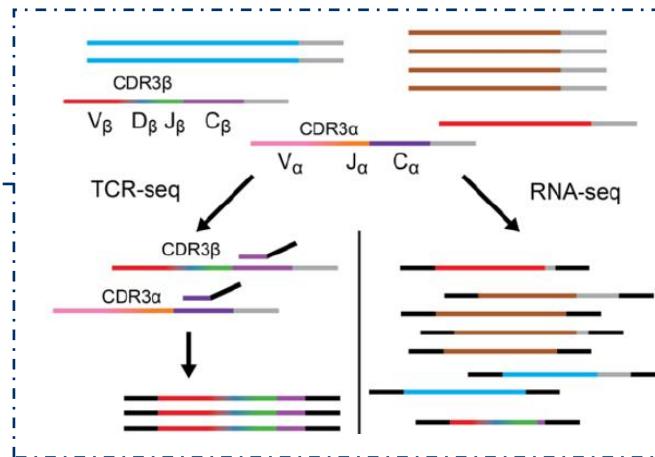
(sc)TCR-seq V.S. (sc)RNA-seq

- RNA-seq和scRNA-seq是对全转录组进行测序，能间接获取TCR相关的转录信息；
- TCR-seq和scTCR-seq针对TCR序列进行扩增和测序，能直接获取TCR信息。

Immuno-Proiling For bulk repertoire analysis

RNA, blood, cells,
or tissueV(D)J amplification of each
BCR/TCR chain

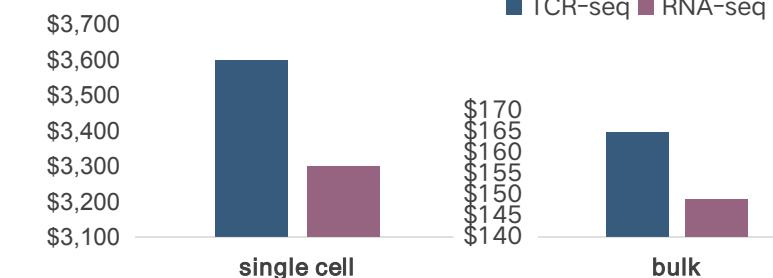
Single-Cell Immuno-Proiling For physical pairing of chains

Cells in
solutionSingle-cell
partitioningV(D)J amplification
and barcoding<https://www.genewiz.com>

bulk TCR-seq
技术成熟、通量高、
成本低，无法获取
TCR双链配对信息

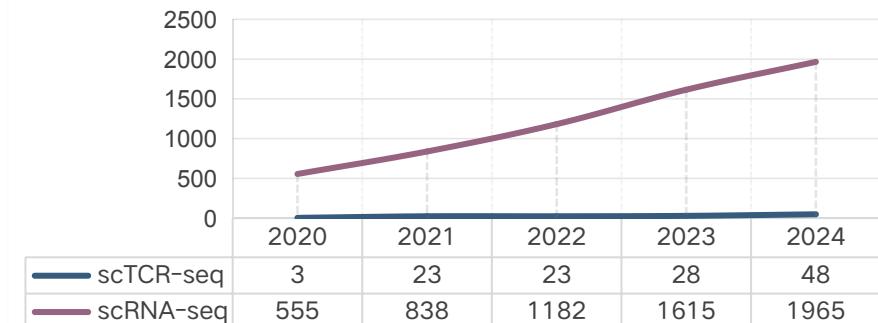
scTCR-seq
实验复杂、成本高，
目前发表公开数据
有限

(sc)TCR-seq和(sc)RNA-seq的建库成本（每个样本）



数据来源：<https://www.uth.edu/cgc/price.htm>
<https://www.advancedgenomics.pitt.edu/pricing>

近五年scTCR-seq和scRNA-seq的公开数据集数量

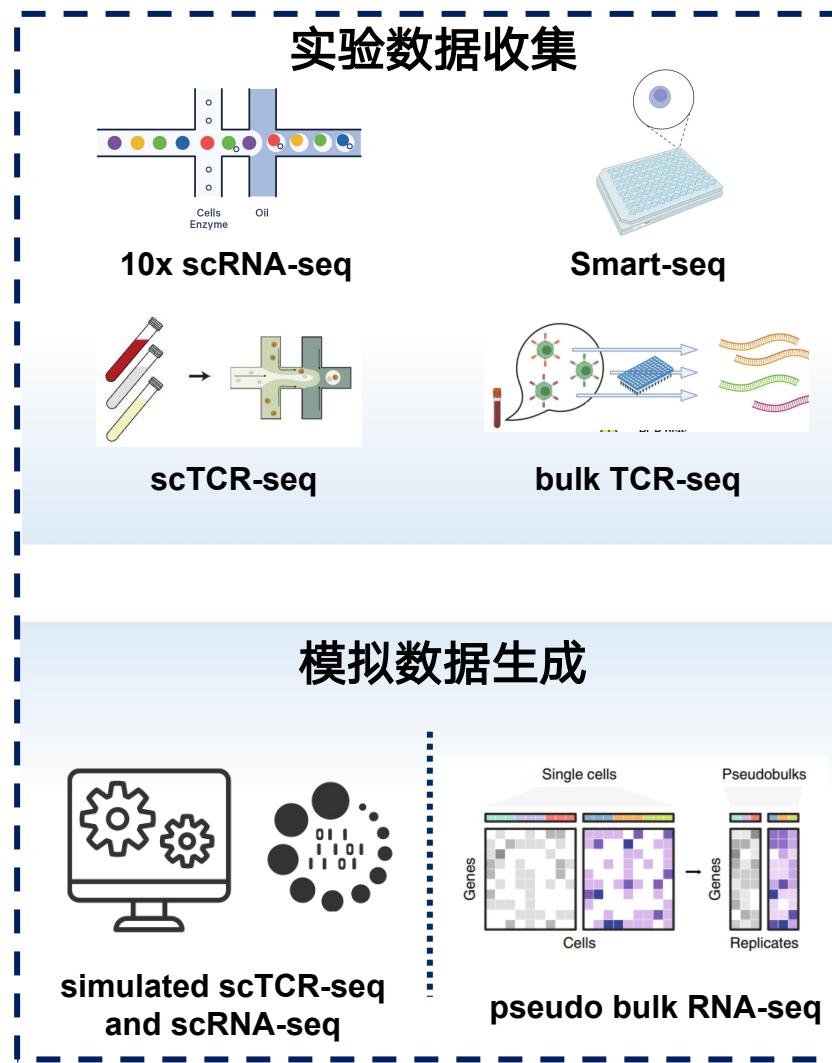


数据来源：<https://www.ncbi.nlm.nih.gov/gds>

(sc)RNA-seq具有成本低、公共数据资源丰富的优势；从其公开数据中重建TCR，可进一步发掘其数据潜力

软件名称	发表时间	支持细胞类型	1 确定候选序列	2 获得contigs	3 序列注释	TCR重建步骤
MiXCR	2015	✓ T细胞 ✓ B细胞	✓ k-mer哈希索引	⚠ overlap拼接 (非de novo)	✓ IMGT注释	
TraCeR	2016	✓ T细胞 ✓ B细胞	✗ 全量输入后期筛选	✓ Trinity(de novo)	✓ IMGT或自定义库注释	
VDJer	2016	✓ T细胞 ✗ B细胞	✓ k-mer哈希索引	⚠ V-D-J拼接 (非de novo)	⚠ 自定义库注释	
BASIC	2017	✓ T细胞 ✓ B细胞	✓ 利用已知V/J seeds	⚠ k-mer图遍历 (非de novo)	✓ IMGT注释	
BALDR	2018	✗ T细胞 ✓ B细胞	✗ 全量输入后期筛选	✓ Trinity(de novo)	✓ IMGT注释	
BraCeR	2019	✗ T细胞 ✓ B细胞	✗ 全量输入后期筛选	✓ Trinity(de novo)	⚠ 自定义库注释	
VDJPuzzle	2019	✓ T细胞 ✓ B细胞	✓ 比对至参考库	⚠ overlap拼接 (非de novo)	⚠ 自定义库注释	
ImRep	2020	✓ T细胞 ✓ B细胞	✓ k-mer哈希索引	✗ 无contig拼接	✓ IMGT注释	
CATT	2020	✓ T细胞 ✓ B细胞	✓ de Bruijn图识别	✓ de Bruijn图组装 (de novo)	✓ IMGT注释	
TRUST4	2021	✓ T细胞 ✓ B细胞	✓ pseudoalignment	✓ 类kallisto算法 (de novo)	✓ IMGT或自定义库注释	
DeRR	未发表	✓ T细胞 ✓ B细胞	✓ anchor reads识别	⚠ overlap拼接 (非de novo)	⚠ 自定义库注释	

如何系统评估不同软件在不同的数据类型上的表现差异?



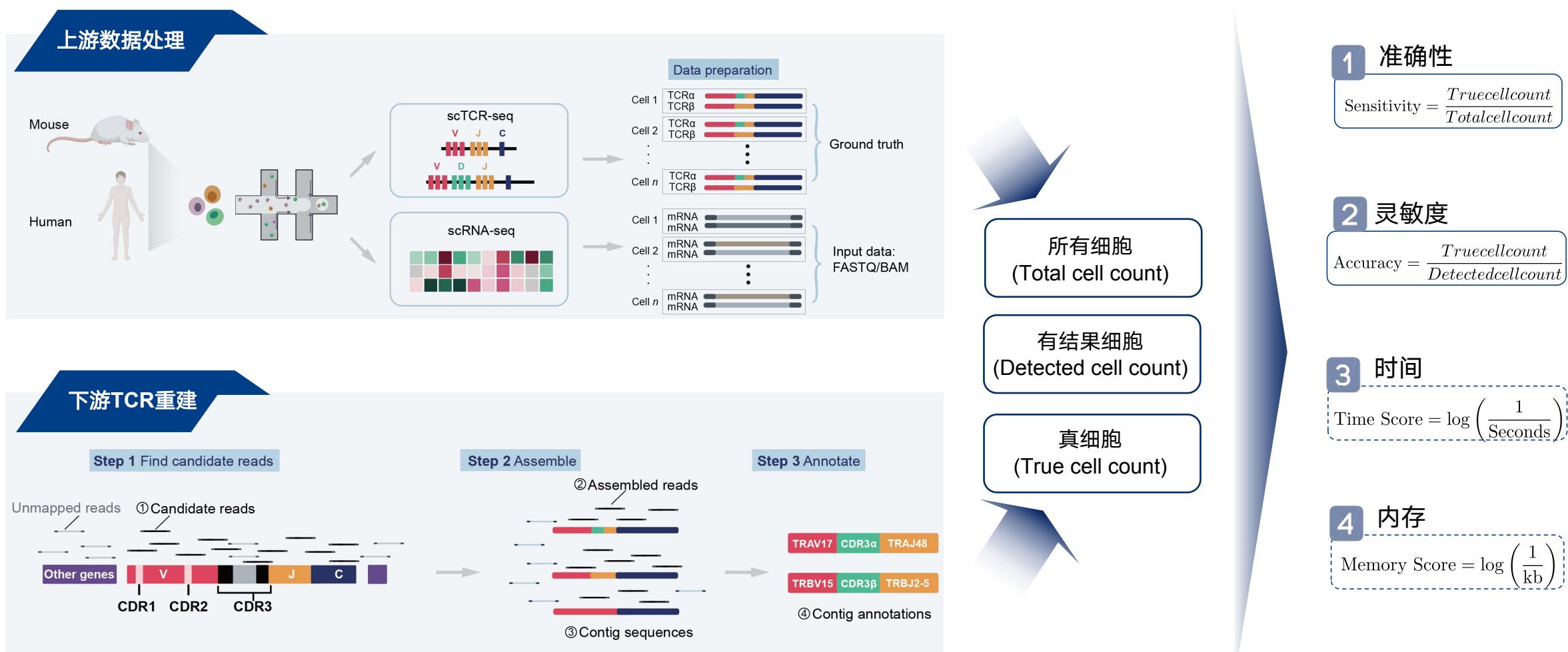
MiXCR
TraCeR
BASIC
ImRep
CATT
TRUST4
DeRR



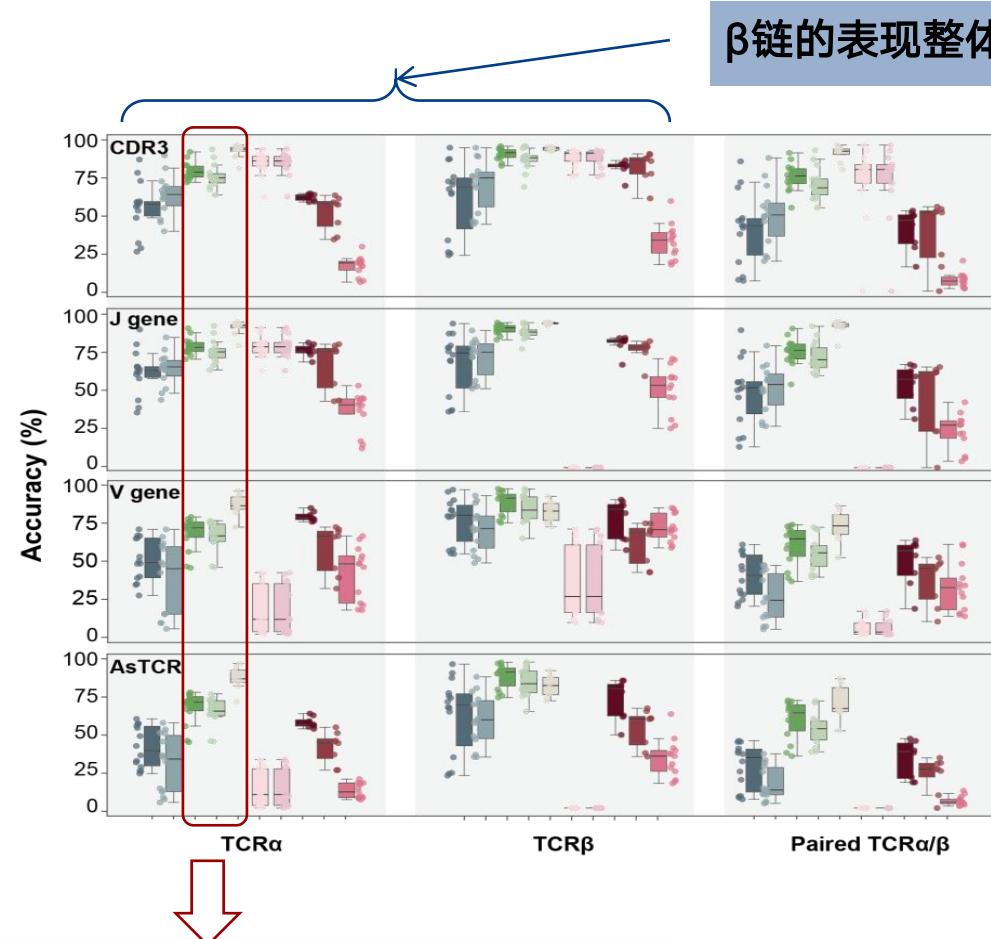
获得不同软件重建
TCR的综合评分

分析影响TCR重建
结果的主要原因

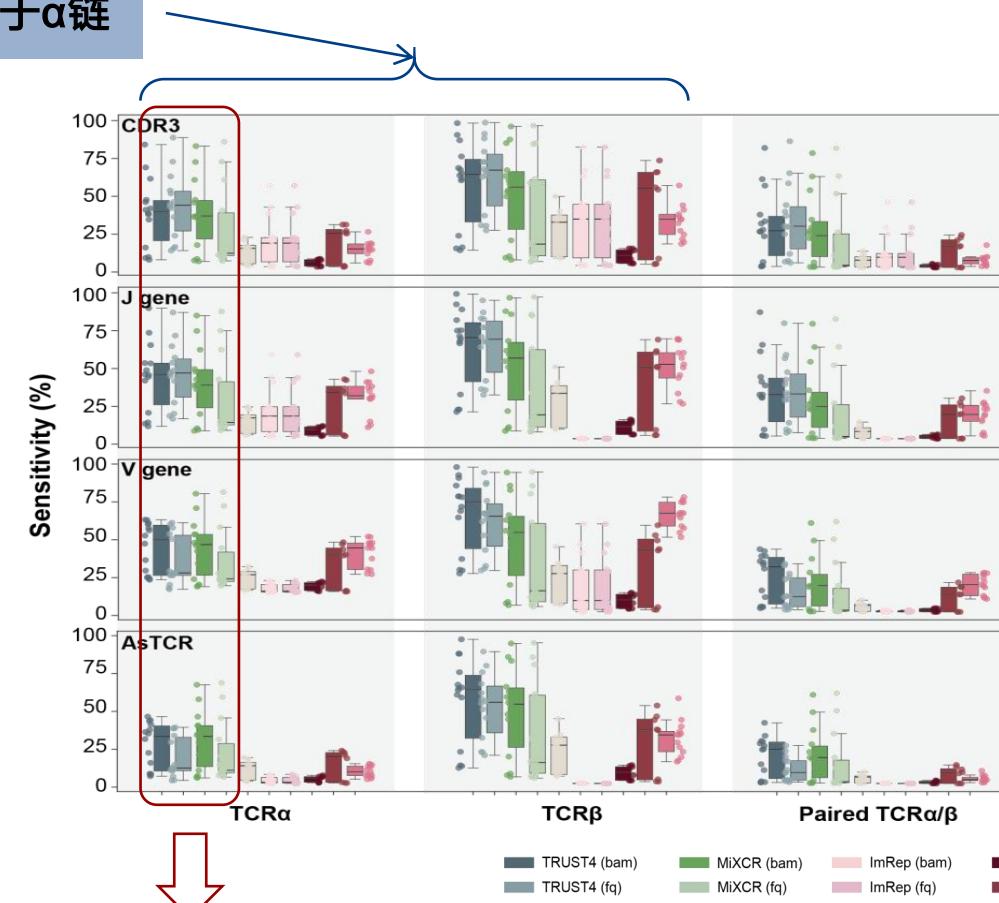
评估过程与关键评估维度的计算方法



■ 2.1 使用真实10x scRNA-seq数据评估不同方法的准确性和灵敏度



DeRR和MiXCR显示出相对较高的准确性



TRUST4和MiXCR表现出明显更高的灵敏度

不同的输入格式会影响重建效果，但是bam格式与fastq格式的影响在不同的软件中并不一致

■ 2.2 使用真实Smart-seq数据评估不同方法的准确性和灵敏度

Smart-seq数据集缺乏相应的scTCR-seq配对数据

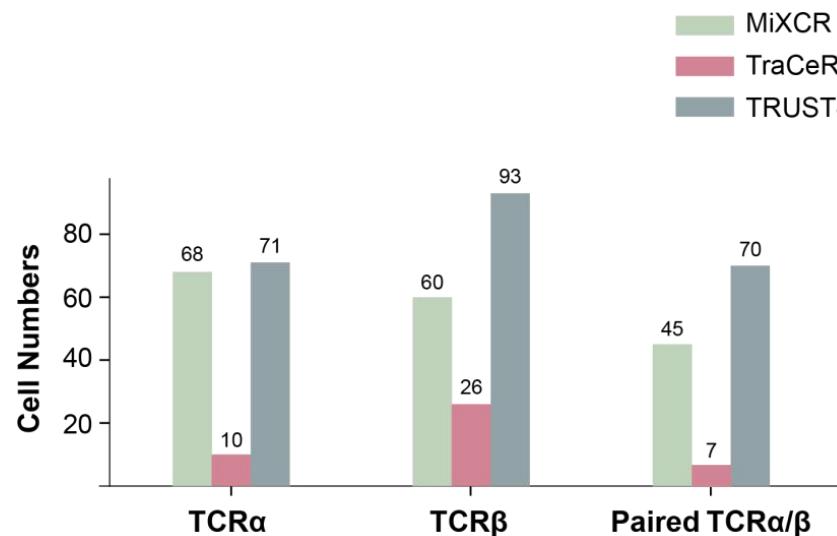


无法直接获得参考真值

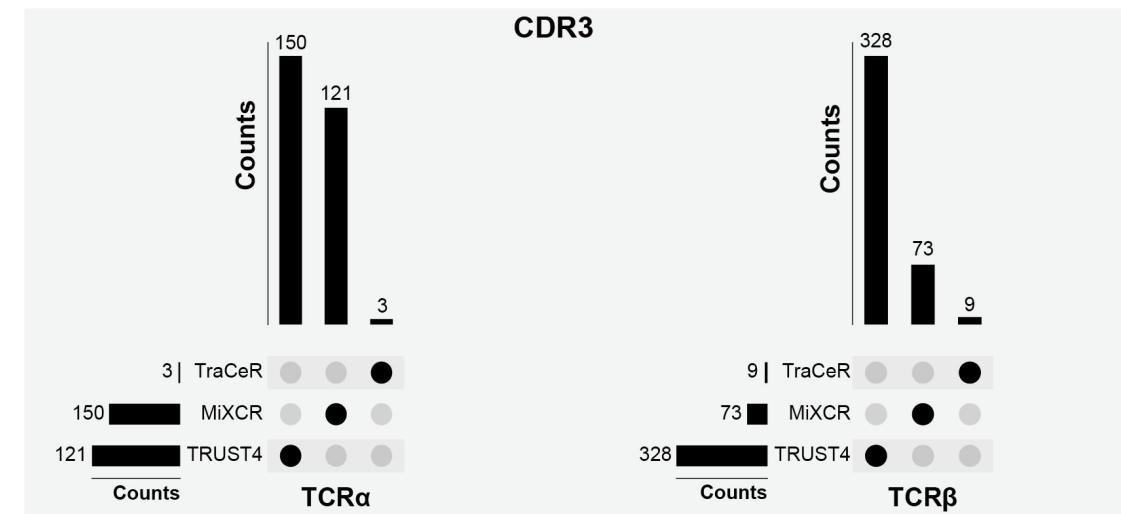


用数量直接评估TCR重建效果

不同软件成功构建TCR的细胞数量

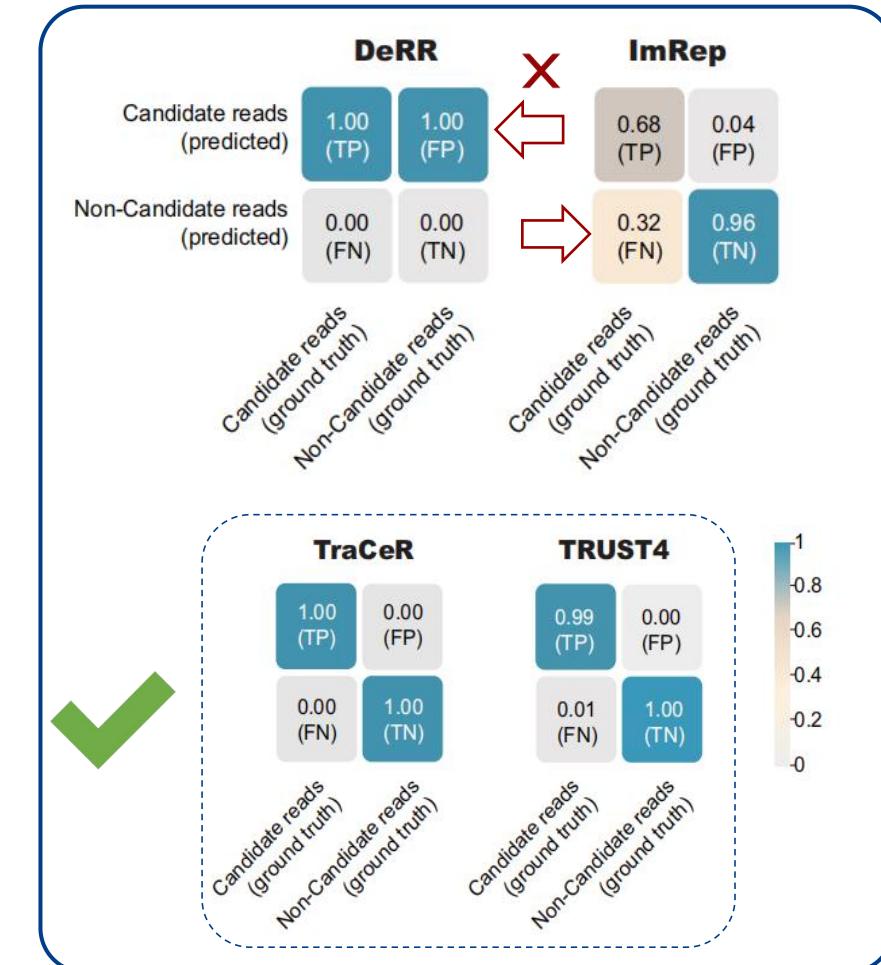
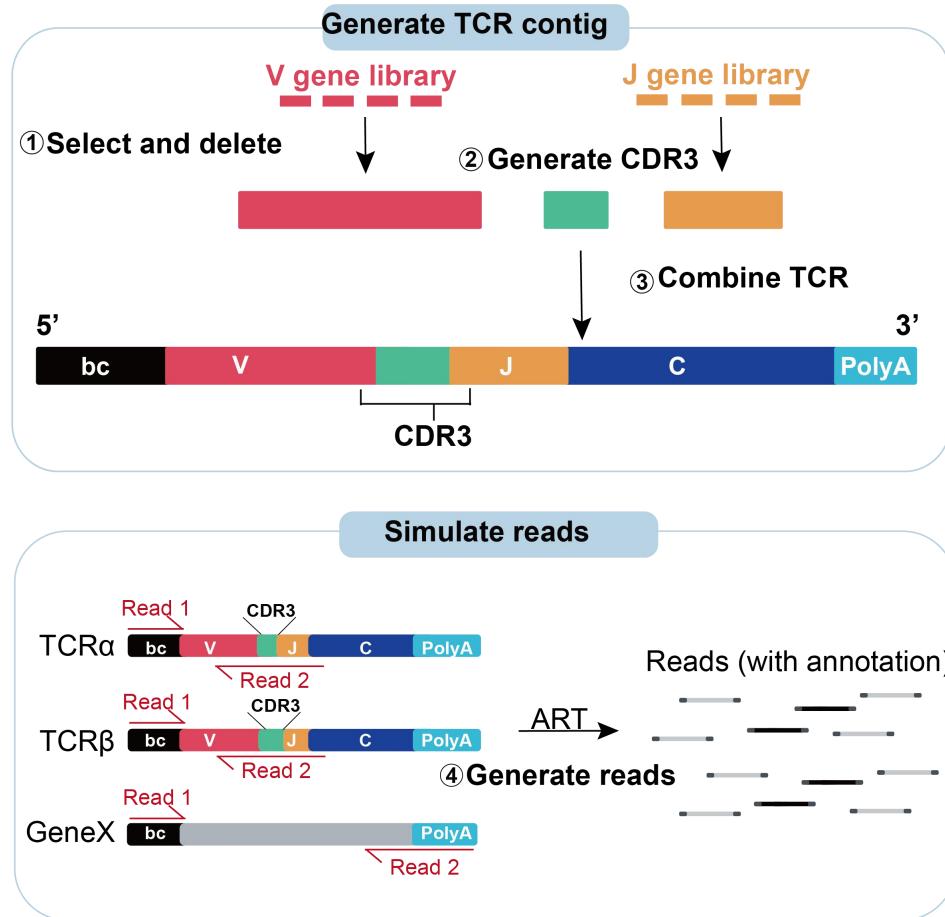


重建成功的TCR克隆数量



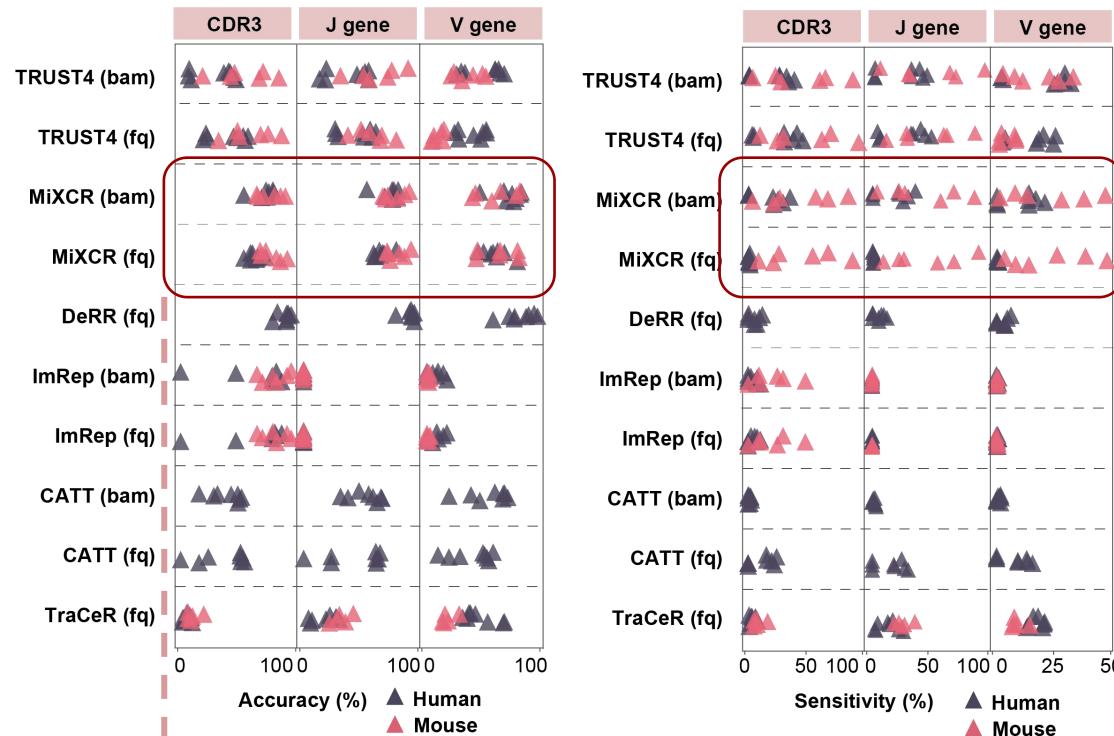
在成功构建TCR的细胞数量和成功构建TCR克隆数量两个维度上，TRUST4的重建结果均为最佳

■ 3. 开发YASIM-scTCR模拟数据生成工具并使用模拟的scTCR-seq数据评估不同方法



■ 4.1 不同方法在不同物种来源和测序策略下的性能分析

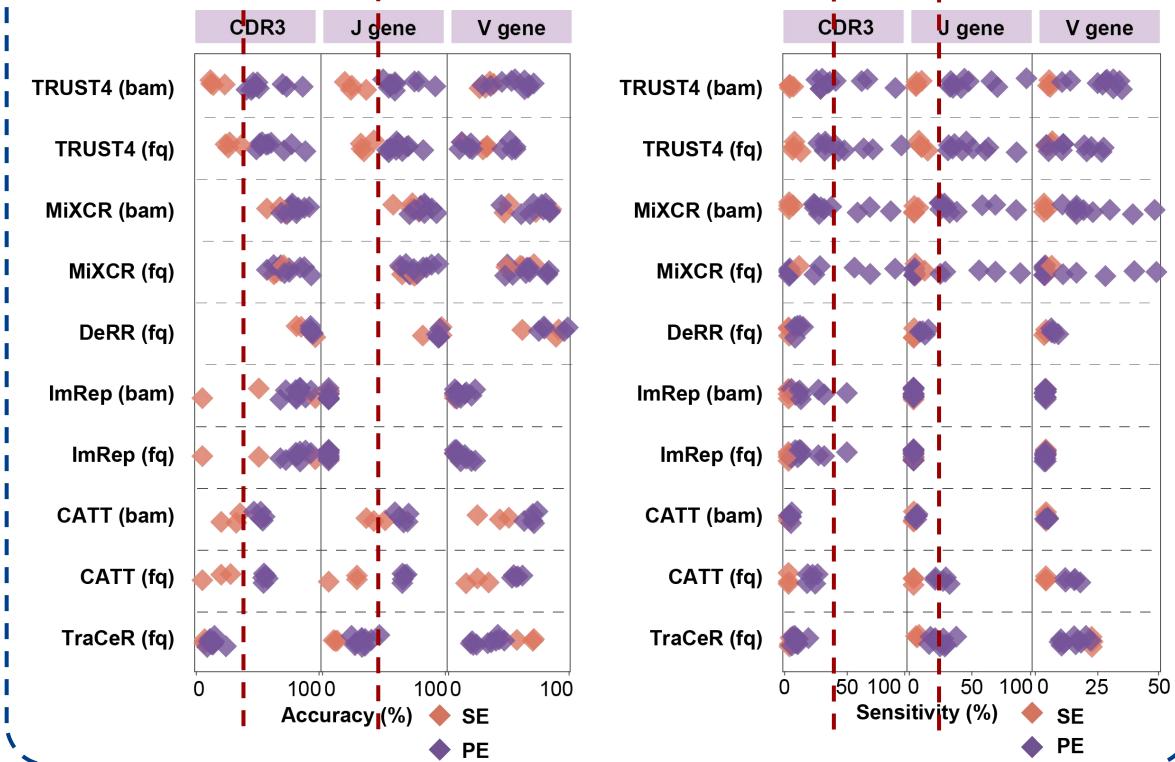
----- 不同物种的TCR重建准确性与灵敏度 -----



大多数方法在两个物种上均展现出相似的准确性与灵敏度

在物种内部观察到了一定的性能异质性

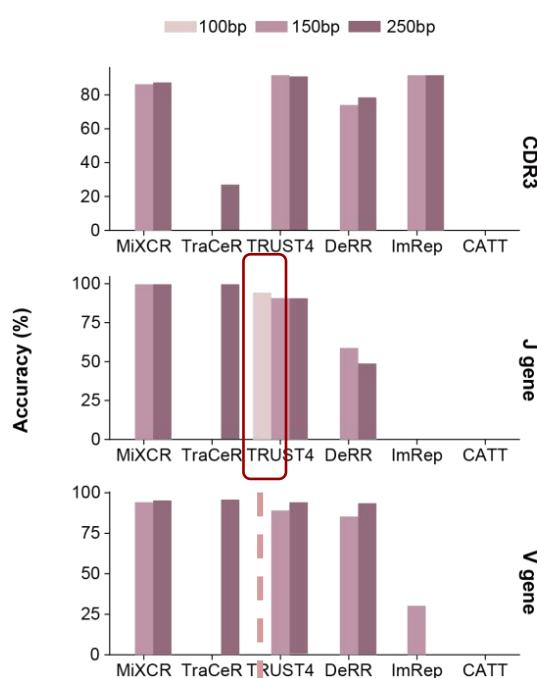
----- 不同文库构建策略的TCR重建准确性与灵敏度 -----



在大多数方法中，与单端测序数据相比，双端测序数据表现出更高的准确性和灵敏度

■ 4.2 不同方法在不同读长和测序深度下的性能分析

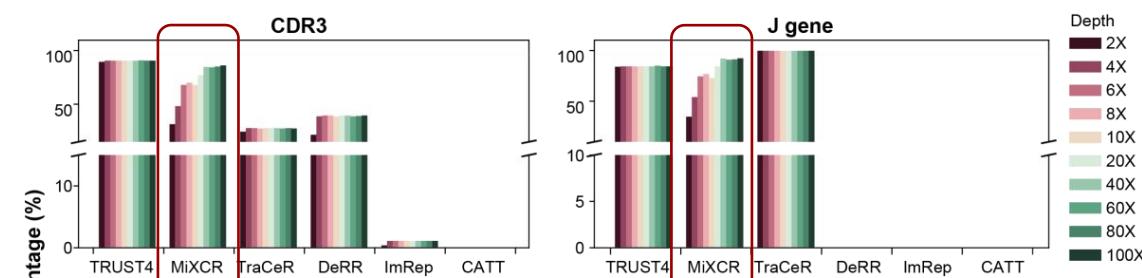
不同读长下的TCR重建准确性与灵敏度



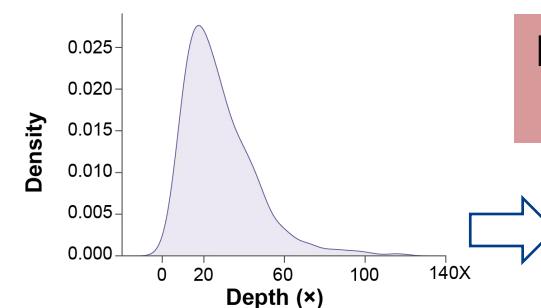
100 bp的读长在大部分软件中无法正常重建TCR

150 bp以上在大部分软件中不会影响重建TCR的结果

不同测序深度下的TCR重建准确性与灵敏度

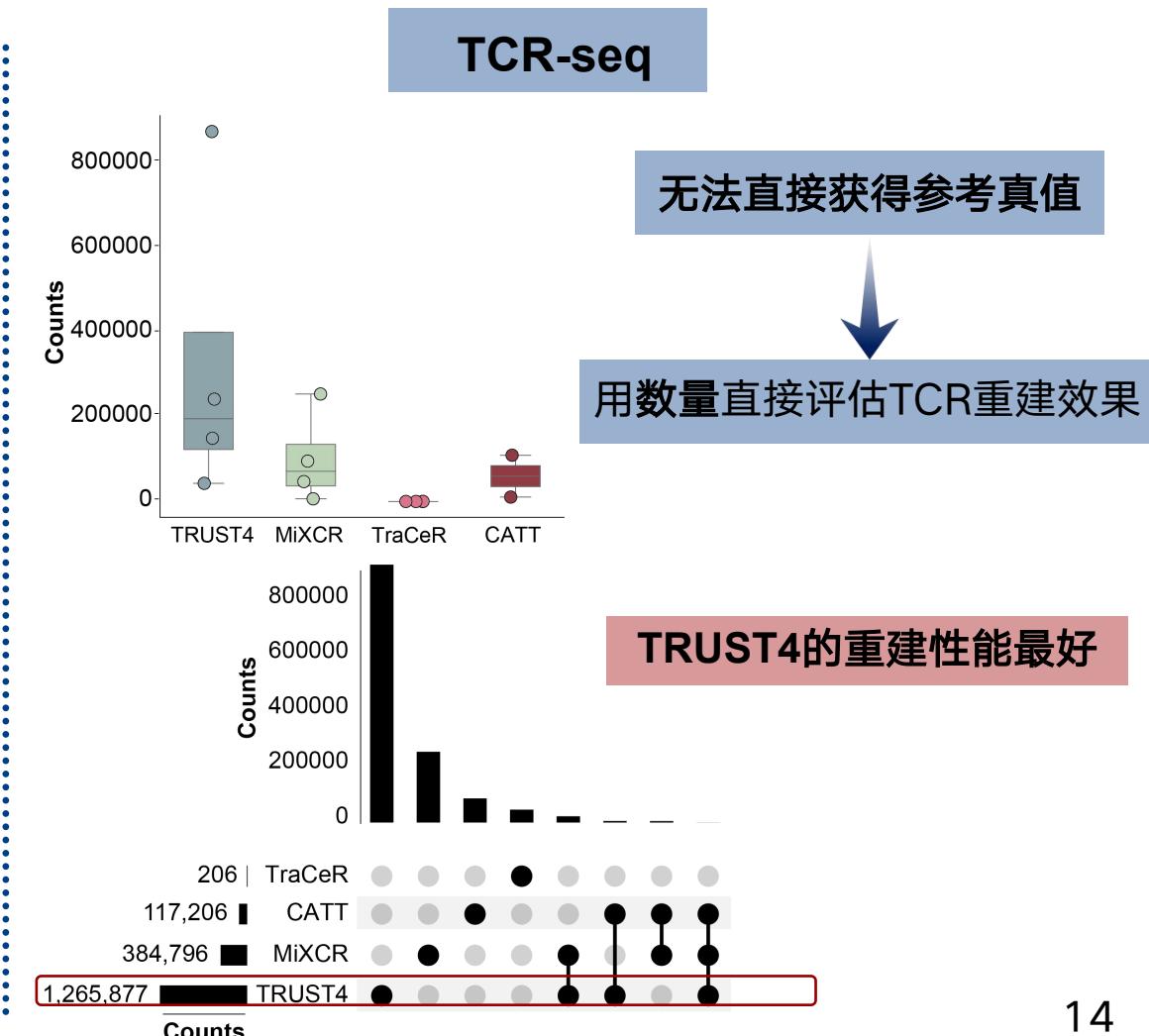
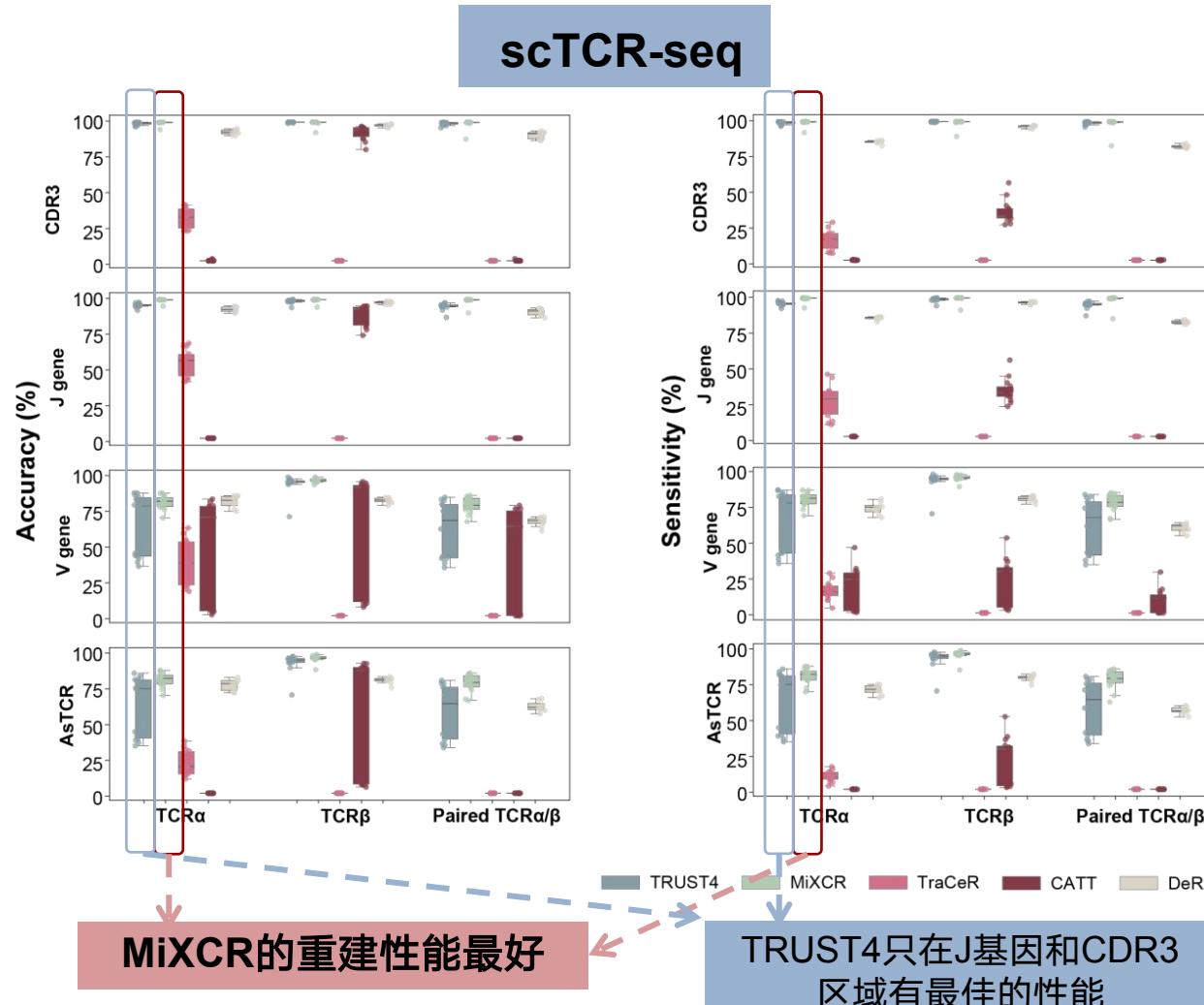


MiXCR需要大约100×测序深度才能实现最佳重建结果



真实测序数据中的深度统计

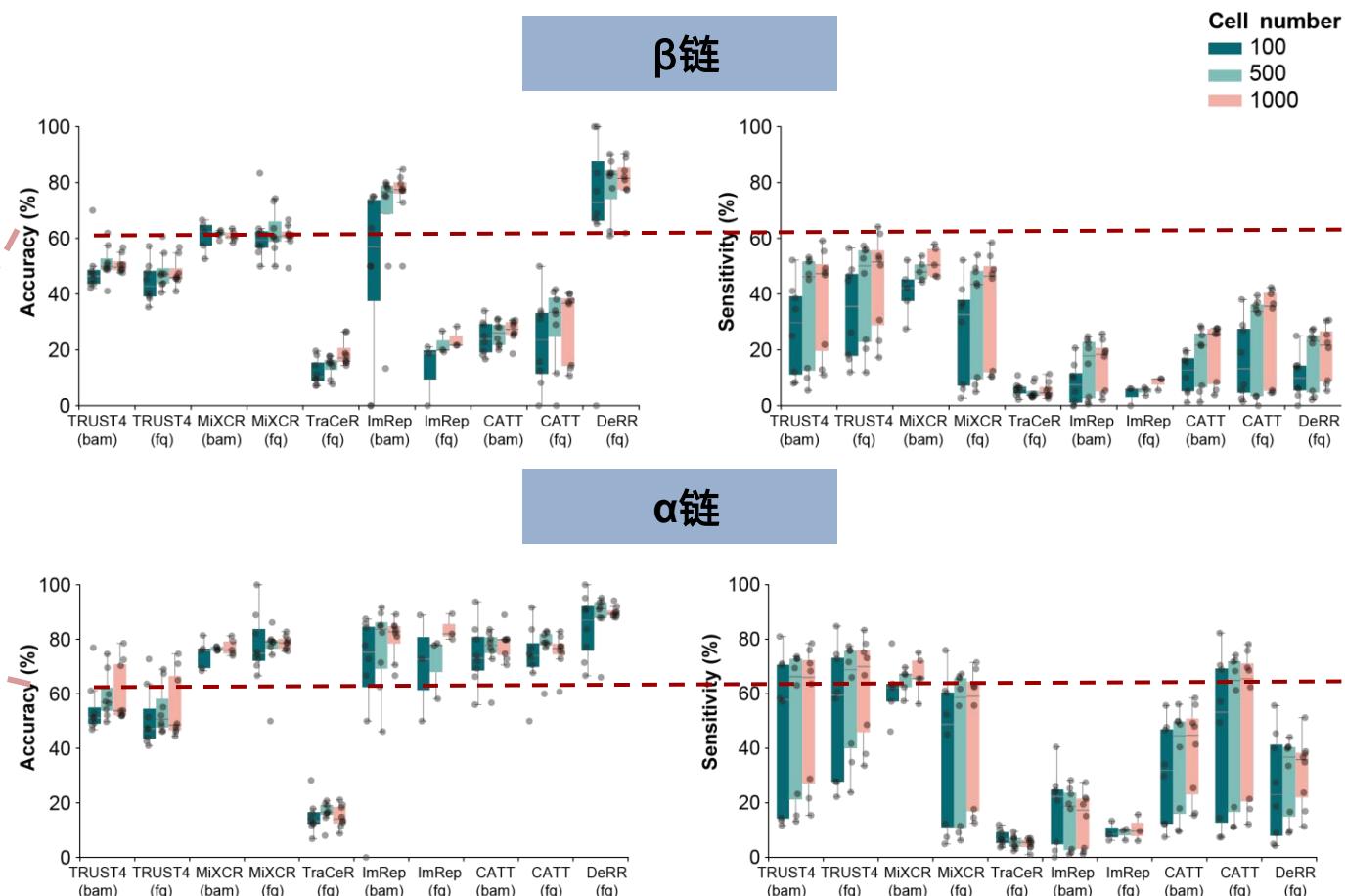
■ 5. 使用真实(sc)TCR-seq数据评估不同方法的性能



■ 6. 使用伪批量RNA-seq数据评估不同方法的性能

随着细胞数量的增加，大多数方法的性能都会提高

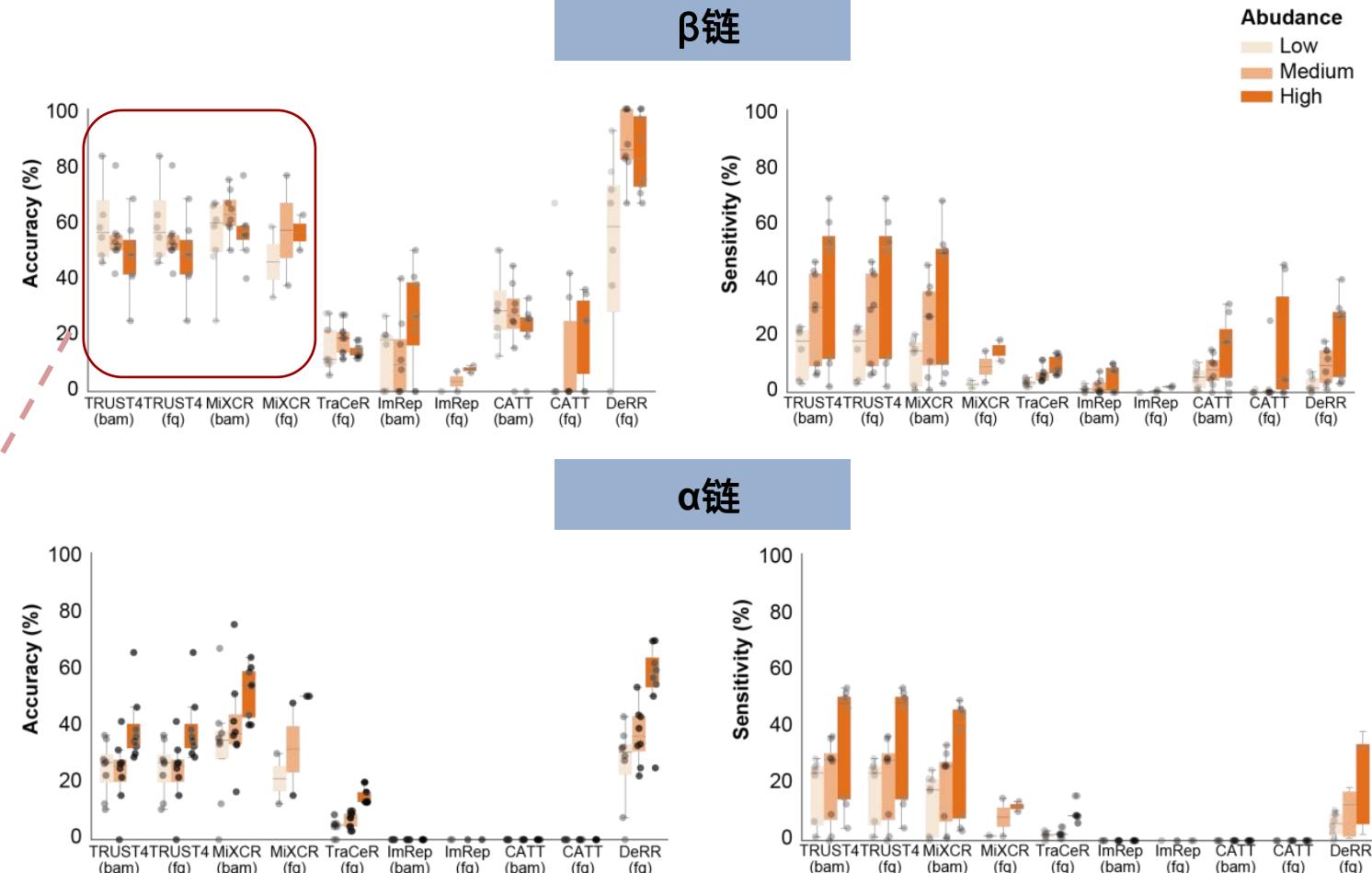
β 链的重建性能明显优于 α 链



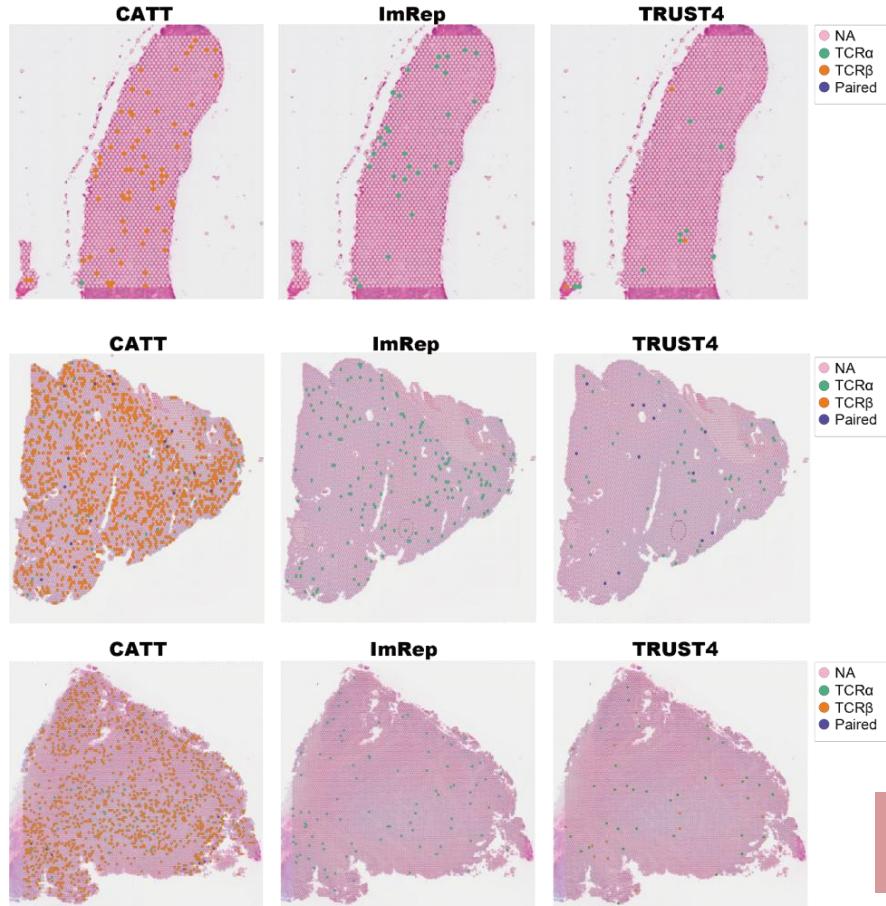
■ 6. 使用伪批量RNA-seq数据评估不同方法的性能

随着细胞丰度的增加，大多数方法的性能都会提高

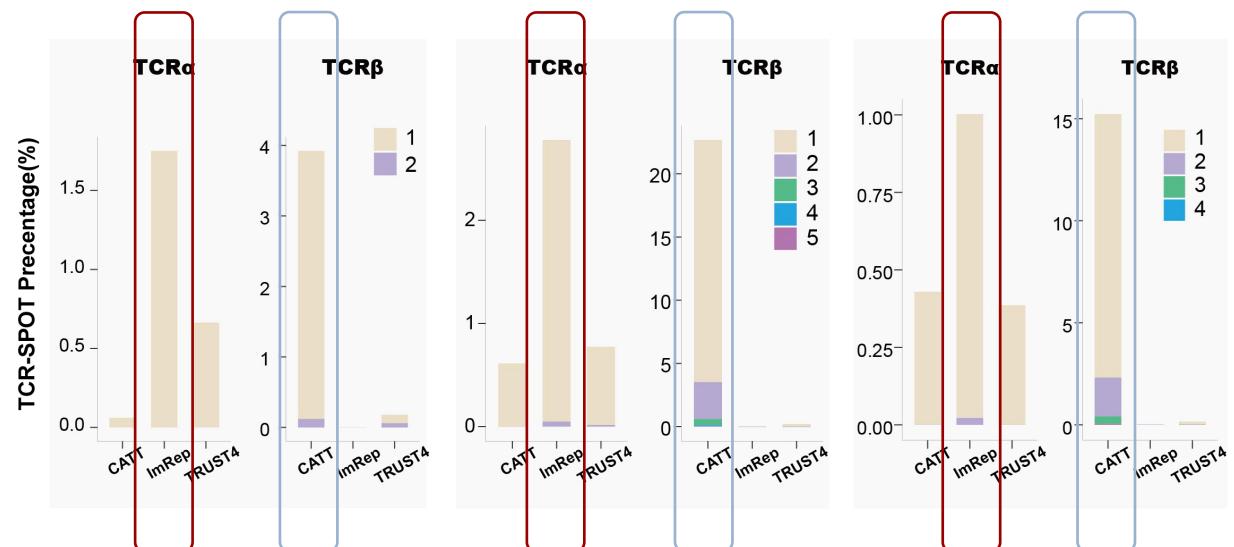
TRUST4和MiXCR的性能受TCR丰度影响相对较小



■ 7. 使用空间转录组数据评估不同方法的性能



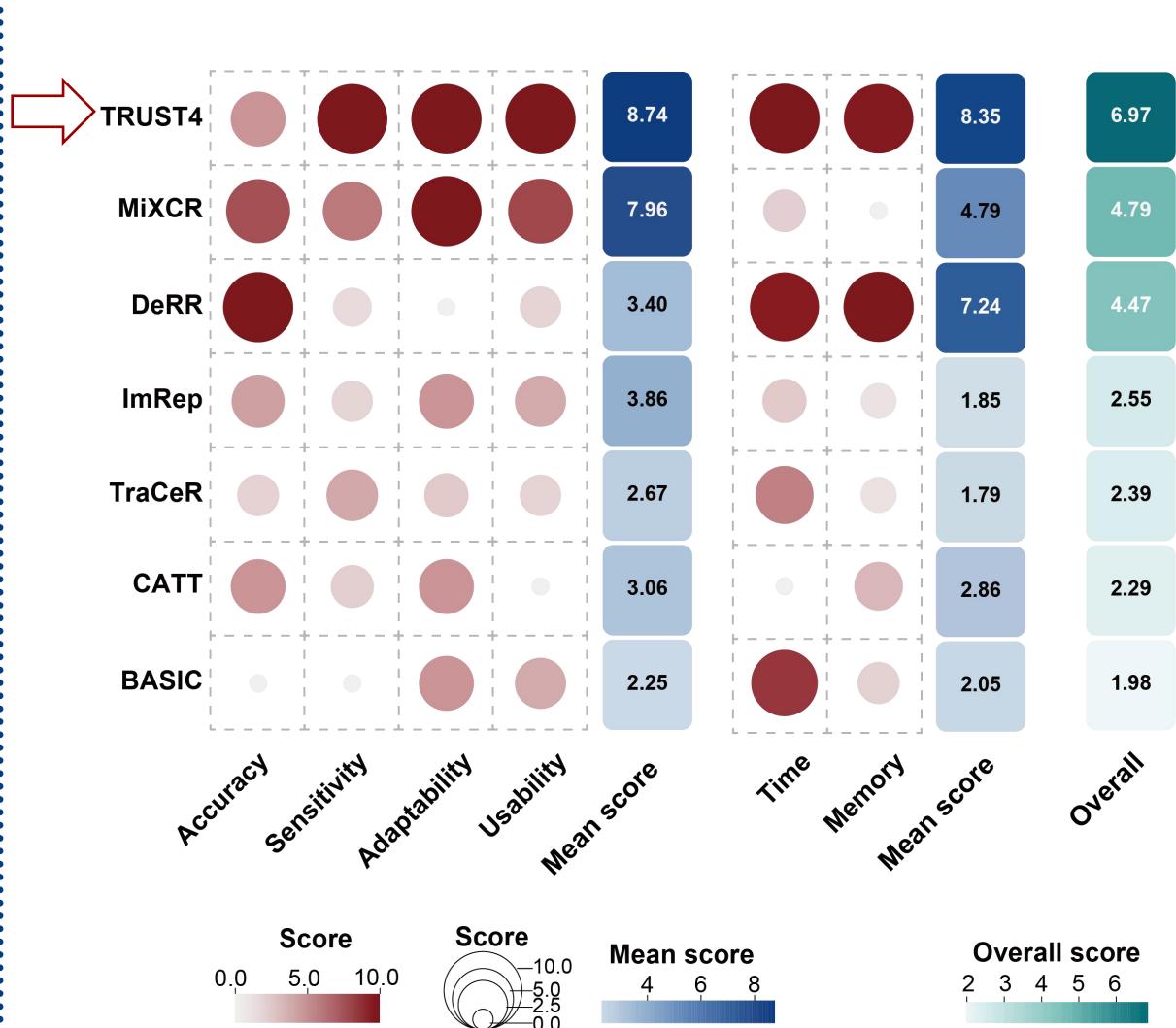
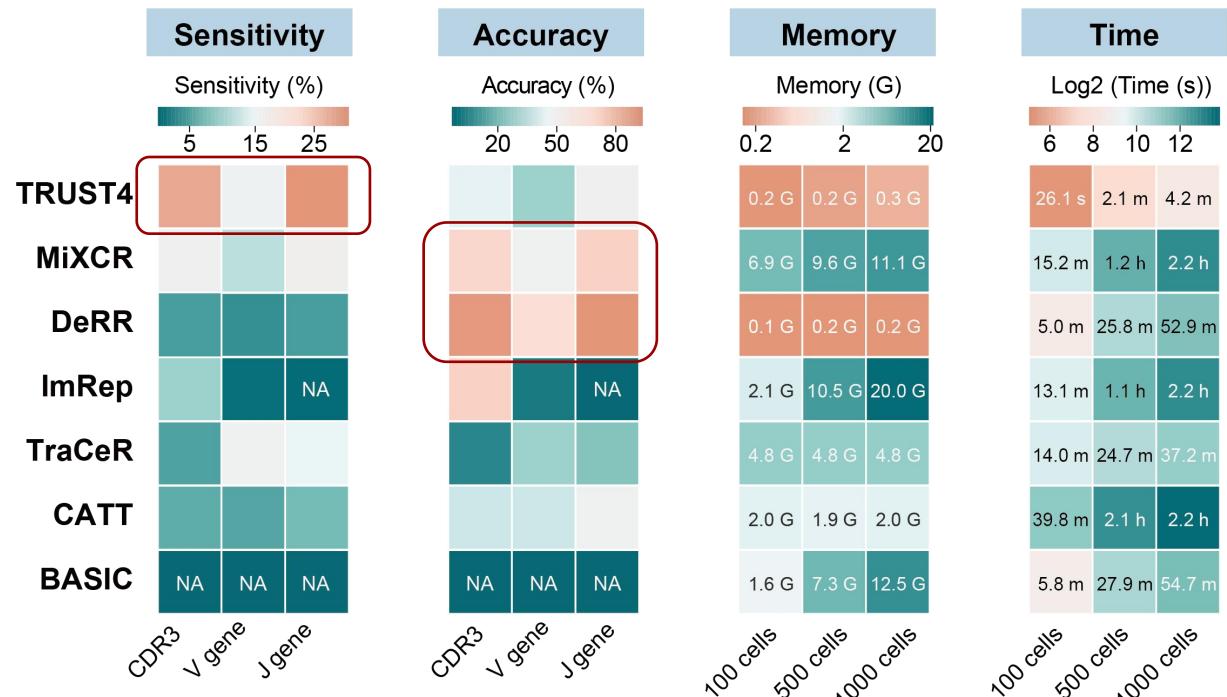
只有CATT、ImRep、TRUST4可以从空间转录组数据中成功重建出TCR



ImRep能够重建最多的TCR α链

CATT能够重建最多的TCR β链

■ 8. 准确性、灵敏度、适应性、可用性、时间消耗和内存使用情况的综合打分



本研究的主要内容与结论

不同TCR重建方法的优缺点

- TRUST4综合评分最高，MiXCR次之；
- DeRR在准确性上表现卓越；
- TRUST4的灵敏度最高。

影响TCR重建性能的关键因素

- 双端测序的TCR重建质量明显优于单端测序；
- 测序读长需要超过150 bp。

总结

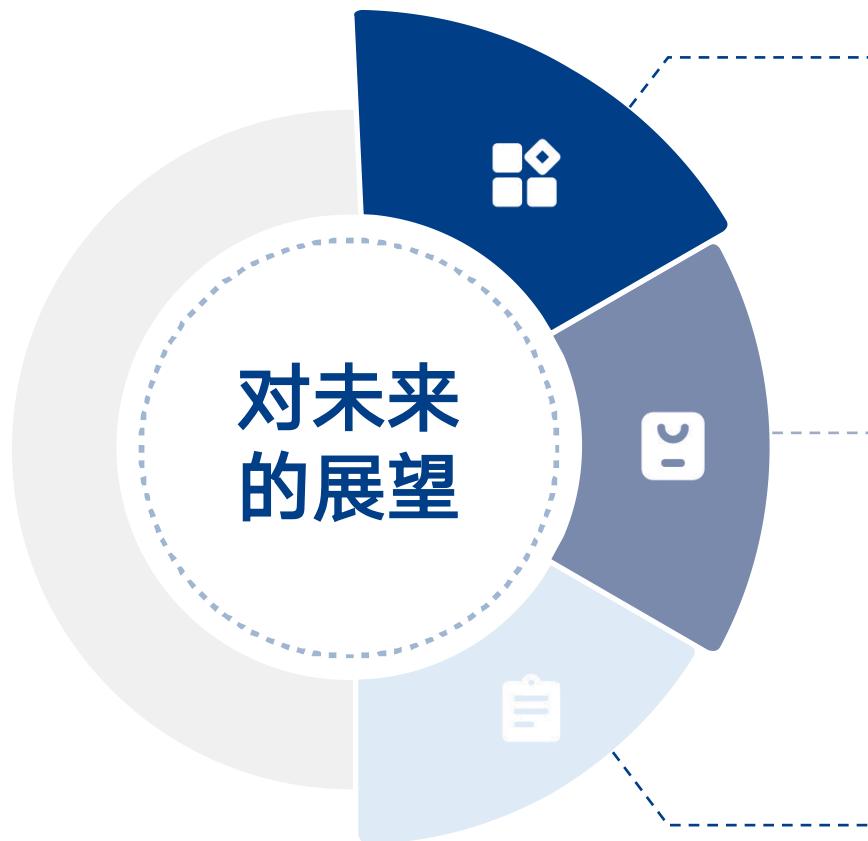
本研究的创新点与局限性

创新点

- 对多种TCR重建方法进行了系统性的评估，并提出了一套全面的性能衡量体系；
- 开发了模拟数据生成软件YASIM-scTCR。

局限性

- 未对TCR克隆扩增等更多生物学维度对重建性能的影响进行充分评估；
- Smart-seq部分的评估存在数据的缺乏。



算法优化方向可进一步深化

本研究所揭示的影响重建性能的关键因素，未来可作为算法改进和模型优化的重要参考，推动更高准确性与稳定性的 TCR 重建工具开发

应用范围亟待拓展

当前 TCR 重建方法主要集中于经典 T 细胞群体，其在非经典免疫细胞类型中的适用性尚未充分验证，未来应重视对这些细胞类型中的 TCR 表达特征的挖掘与适配性改良

多组学整合将成为重要发展方向

将 TCR 重建方法与表观组学、空间转录组学、TCR 结构功能分析等多维数据深度融合，有望提升重建结果的生物学解释力与应用广度，为精准免疫研究提供更全面的工具支持

感谢刘琬璐老师在课题的开展过程中给予的指导与建议；感谢田若楠、俞哲健、薛子为在课题中的卓越贡献；感谢阮登峰博士、高兵对课题提出的宝贵建议，感谢课题组的全体成员在科研生活中的慷慨帮助。





浙江大学
ZHEJIANG UNIVERSITY

敬请各位老师批评指正

谢 谢

