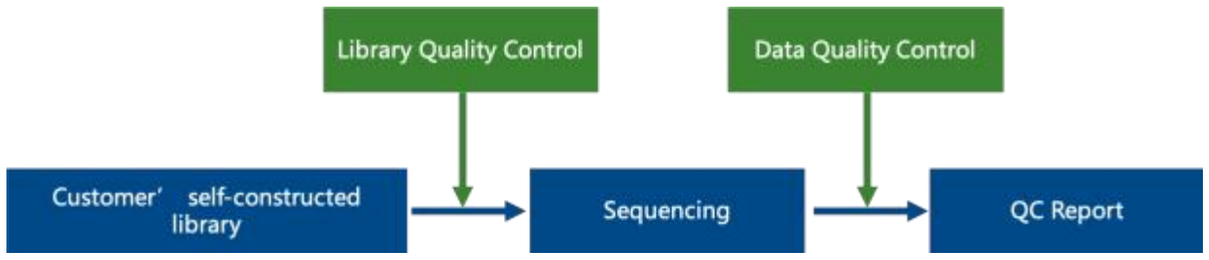
10X Single Cell CellRanger Only Analysis Method

A. Library Preparation and Sequencing

From the RNA samples to the final data, each step (including sample testing , library preparation and sequencing) will influence the data quality. The quality of data would have direct impacts on the analysis results . To guarantee the reliability of the data, Quality Control will be performed on each step of the procedure . For customer' self-constructed libraries , the quality detection will be performed before sequencing to ensure producing high quality of data.



1 Library Quality Control

There are mainly three methods in QC for library quality control:

(1) Qubit 2 .0 : tests the library concentration preliminarily.

(2) Agilent 2100 : tests the insert size .

(3) Q- PCR: quantifies the library effective concentration precisely.

2 Sequencing

The qualified libraries are fed into Illumina sequencers after pooling according to its effective concentration and expected data volume .

B. Data Analysis

1. 10x Sequencing data processing and quality control

Once the FASTQ files for each samples is generated , the data analysis begins . Reads from scRNA-seq were processed with Cell Ranger software (10x Genomics) with the default parameters for each sample separately. Briefly, N nucleotides (nt) of Read1s or Read2s were aligned against the reference genome with STAR. Barcodes and UMIs were filtered and corrected . PCR duplicates were marked using the barcode , UMI and gene ID. Only confidently mapped , non- PCR duplicates with valid barcodes and UMIs were used to generate a gene-barcode matrix for further analysis .

The filtered gene expression matrices containing only cellular barcodes were generated in data processing step . Then the Seurat R package was futher utilized to achieve quality control.

Alignment to Reference Genome

Cell Ranger uses an aligner called STAR, which peforms splicing-aware alignment of reads to the genome . It then uses the transcript annotation GTF to bucket the reads into exonic, intronic, and intergenic, and by whether the reads align (confidently) to the genome . A read is exonic if at least 50% of it intersects an exon , intronic if it is non-exonic and intersects an intron , and intergenic otherwise .

The exonic reads were further aligned to the existing transcriptome with annotation as well. A read that is compatible with the exons of an annotated transcript, and aligned to the same strand , is considered mapped to the transcriptome . Among mapped reads , these uniquely mapped reads are the only ones considered for UMI counting .

Cell Calling and UMI Counting

Before counting UMIs , Cell Ranger attempts to correct for sequencing errors in the UMI sequences . The UMI of the less supported read group is corrected to the UMI with higher support. Cell Ranger again groups the reads by barcode , UMI (possibly corrected) , and gene annotation . If two or more groups of reads have the same barcode and UMI, but different gene annotations , the gene annotation with the most supporting reads is kept for UMI counting , and the other read groups are discarded .

The cell-calling algorithm based on the EmptyDrops method in Cell Ranger determines cell-associated barcodes based on their UMI count or by their RNA profiles . The algorithm has two key steps:

1 . It uses a cutoff based on total UMI counts of each barcode to identify cells . This step identifies the primary mode of high RNA content cells .

2 . Then the algorithm uses the RNA profile of each remaining barcode to determine if it is an “ empty" or a cell containing partition . This second step captures low RNA content cells whose total UMI counts may be similar to empty GEMs .

Reference

1 . Z heng , G. X. Y., Terry, J. M., Belgrader , P., Ryvkin , P., Bent, Z .W. , & Wilson , R., et al. (2017) . Massively parallel digital transcriptional profiling of single cells . Nature Communications , 8 , 14049 .

2 . Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, III WMM, Hao Y, Stoeckius M, Smibert P, Satija R. (2019) . Comprehensive Integration of Single-Cell Data. Cell, 177 , 1888-1902 .

3 . Ayyaz , A., Kumar , S., Sangiorgi, B. et al. (2019) . Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. Nature , 569 , 121– 125 .

4 . Peng , Y. R., Shekhar , K., Yan , W., Herrmann , D., Sappington , A., & Bryman , G. S., et al. (2019) . Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. Cell, 176 , 1222-1237 .e22 .

5 . Kim, D. W., Yao , Z ., Graybuck, L. T., Kim, T. K., & Anderson , D. J,. (2019) . Multimodal analysis of cell types in a hypothalamic node controlling social behavior . Cell, 179(3) , 713-728 .e17 .

6 . Amezquita, R. A., Lun , A.T.L., Becht, E. et al. (2020) . Orchestrating single-cell analysis with Bioconductor . Nature Methods , 17 , 137– 145 .

7 . Geng , Chen , Baitang , Ning , Tieliu , & Shi. (2019) . Single-cell rna-seq technologies and related computational data analysis . Frontiers in Genetics , 10 , 317 .

8 . Dobin , A., Davis , C. A., Schlesinger , F., et al. (2013) . STAR: ultrafast universal RNA-seq aligner . Bioinformatics , 29(1): 15-21 .

9 . Becht, E., McInnes , L., Healy, J. et al. (2019) . Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnol 37 , 38–44 .

10 . Van der Maaten , L., (2014) . Accelerating t-SNE using Tree- Based Algorithms . Journal of Machine Learning Research , 15 , 3221-3245 .

11 . Butler , A., Hoffman , P., Smibert, P., Papalexi, E., & Satija, R. (2018) . Integrating single-cell transcriptomic data across different conditions , technologies , and species . Nature Biotechnology, 36(5) , 411–420 .

12 . Yu , G., Wang , L. G., Han , Y., & He , Q.Y. . (2012) . Clusterprofiler: an r package for comparing biological themes among gene clusters . Omics A Journal of Integrative Biology, 16(5) , 284 -287 .

13 . Kanehisa M, Goto S. (2000) . KEGG: kyoto encyclopedia of genes and genomes . Nucleic acids research , 28(1): 27-30 .

Yu , G., & He , Q. Y. . (2016) . Reactomepa: an r/bioconductor package for reactome pathway analysis and visualization . Molecular BioSystems , 12(2) , 477-479 .

14 . Schlesner , Matthias , Eils , Roland , Gu , & Z uguang . (2016) . Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics , 32 , 2847– 2849 .