Novogene | Advancing Genomics, Improving Life

Technical Note: Nanopore Single Cell Full-Length Transcriptome

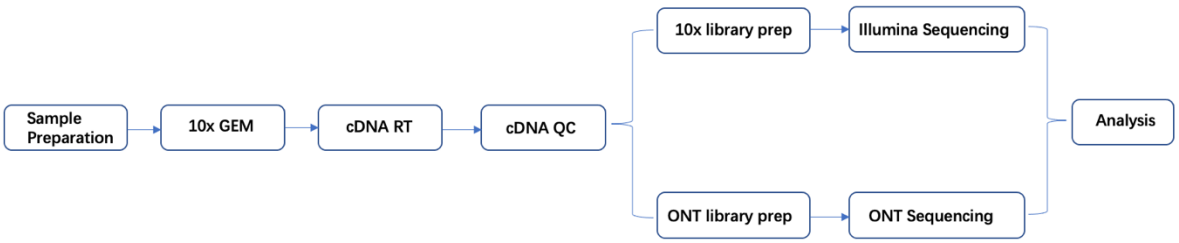
May 2023

Combining long-read sequencing with single cell assays enables the unambiguous identification of alternative splicing at single cell resolution. Traditional single cell assays have relied on short-read sequencing, which loses valuable information about transcript isoforms relevant to health, development, and disease.

This document covered the technical information for Nanopore Single Cell Full-length Transcriptome:

- Project workflow
- Sample requirements
- Data suggestion
- Validation data
- Analysis pipeline

1. Project Workflow



2. Sample Requirements

Sample Type	Sample Amount	Cell Viability	Concentration	Storage
Cells	At least 50, 000 cells	Cell viability should be >80% Cell size to be <40 μm	-	-
cDNA	At least 50ng	-	> 2 ng/ul	No more than 2 month under -20℃

3. Data Suggestion

Objective	Detection of structural variants of isoforms	Accurate gene quantification
Data (~5,000 cells)	ONT long read (1 Cell)	ONT long read (1 Cell) + Illumina short read (100Gb)

4. Validation Data

To construct single cell libraries, we utilized Human (mm10) and Mouse samples (hsa) and followed 10x Chromium single cell 3' protocol and the respective manufacturer's protocols illumina short read and Nanopore long read. The short-read sequencing was performed on an Illumina NovaSeq 6000 system using paired-end reads of 100Gb data. For long read sequencing, the library was sequenced on an individual Oxford Nanopore PromethION Cell.

We achieved a data output of more than 150M total reads for each library. The data quality assessment, including cell number and median gene number, yield a high level of concordances between short read and long read data. Furthermore, we demonstrated that the data output from a single Nanopore's PromethION cell can yield sufficient reads for isoform detection, structural variation detection and cell clustering for single cell analysis.

wf-single-cell	Reads	Cells	Genes (total)	Transcripts (total)
sample mm10	168,648,648	7,710	20,748	47,806
Sample hsa	154,825,486	4,917	29,534	83,627

Seurat		sampleC1_5(mm 10)	sample 1 *hsa
ONT gene	ncells	7,212	4,349
	median	3,195	4,083
	Median umi	7,522	13,129
ONT transript	ncells	7,142	4,687
	median	3,599	4,915
	Median umi	6,985	11,643
Illu gene	ncells	6,969	3,977
	median	2,056	1,979
	Median umi	6,312	6,605

Fig1. Data Output and Data Quality. The data output of the 2 samples was greater than 150M total reads, in line with the official recommendation of Nanopore.

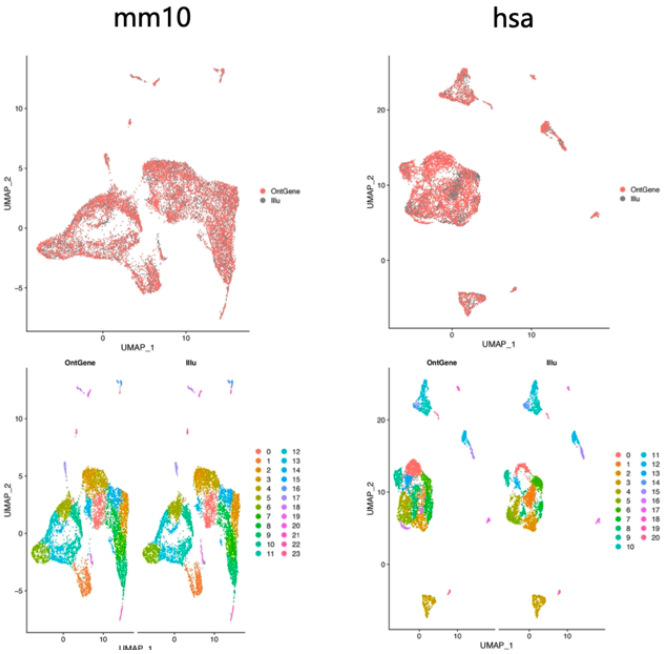


Fig2. UMAP plots show high consistency of the cell annotation grouping results in both short reads and long reads sequencing data.

5. Analysis Pipeline

10x single cell gene expression – Cell Ranger

Nanopore full-length transcriptome - Sockeye

Cell Ranger	Sockeye
Read processing	Orient
Align	Align
Corrected barcodes	Uncorrected barcodes
Filter UMIs	Barcode shortlist
Count UMIs	Corrected barcodes
Call cells	Gene assignments
Secondary analysis	Corrected UMIs
	Matrix
	UMAP