

Single-cell transcriptomics helps to unlock our understanding of the subtleties of cellular diversity

A combination of single-cell approaches with full-length cDNA sequencing offers the potential to provide a level of detail to transcriptomic studies that is not available from bulk analyses

Contact: publications@nanoporetech.com More information at: www.nanoporetech.com and publications.nanoporetech.com





Fig. 1 Single-cell a) isolation b) reverse transcription and enrichment of full-length cDNAs

Single-cell transcriptomics can reveal differences in expression patterns of cells

Differences in the transcriptomic behaviour of individual cells are not visible in bulk analyses of heterogeneous cell populations. If the transcripts from each cell are given a specific label before analysis then it is possible to compare the expression levels of single cells. One way of achieving this is to encapsulate single cells in droplets along with a bead coated with reverse transcription (RT) primers. The primers surrounding any one bead contain the same cell-barcode sequence. Cell lysis and RT occur within the droplets. In this way, all cDNAs derived from the same cell are given the same cell barcode (Fig. 1a). Following RT and strand switching, all cDNAs can be pooled and amplified, before attachment of sequencing adapters (Fig. 1b).





Fig. 2 Enriching for full-length cDNAs in 10X single cell libraries by biotin capture

Maximising the proportion of full-length cDNAs in 10X single-cell libraries

PCR artefacts are frequently produced during amplification of the barcoded single-cell cDNAs, which limits the proportion of full-length transcript reads. The major 10X PCR artefact consists of a truncated cDNA flanked by copies of the strand-switching oligo (Fig. 2a). These strands can be depleted by biotin capture, giving a far higher proportion of full-length reads per run: ~90% (Fig. 2b). Typical yields from a PromethION[™] flow cell are shown. It is currently typical to generate around 100 million full-length reads per flow cell (Fig. 2c). By removing this major library artefact, we also see the correlation between our single-cell expression levels and those from a matching Illumina dataset increase substantially (Fig. 2d).



Fig. 3 Data analysis a) Sockeye bioinformatics workflow b) knee plot c) agreement with short reads

The Sockeye bioinformatics pipeline enables ONT-only cell-barcode identification

Following data generation, putative cell barcodes from the highest quality reads are clustered using the *Sockeye* pipeline, to identify the cell barcodes present in the sample (Fig. 3a). The resulting barcode clusters are visualised in a knee plot (Fig. 3b) to identify which cell barcodes have sufficient read support. Once these true barcodes are identified, cell barcodes are assigned to all reads based on edit-distance criteria. Barcode-assigned reads are then aligned and annotated using the appropriate references. The results can then be visualised through cell clustering based on gene expression, or full-length transcript consensus sequences can be generated to look at isoforms, alternative splicing and genotyping. An upset plot (Fig. 3c) shows that the majority of Illumina- and ONT-called barcodes are in agreement.

Fig. 4 Barcode assignments a) 'barnyard' experiment b) platform comparison c and d) UMAP plots

Demonstrating 10X single-cell profiling of a mixture of human and mouse cells

We analysed a sample prepared from a mixture of human and mouse cells, assigning barcodes to reads before aligning to a combined reference. We counted reads aligned to each gene, and produced a count of UMIs per gene per barcode. For each cell we then tabulated the number of UMIs mapping to mouse vs human genes. The vast majority of UMIs for each cell barcode belong to either human or mouse genes (Fig. 4a). Next we compared ONT and Illumina libraries of the same sample. We see excellent concordance between the two technologies (Fig. 4b). We then plotted transcripts per gene per barcode in both the Illumina (Fig. 4c) and ONT (Fig. 4d) datasets using *UMAP*. As expected, in each case the mouse and human cells are clearly separated based on observed patterns of single cell gene expression.

Oxford Nanopore Technologies, the Wheel icon, EPI2ME, Flongle, GridION, Metrichor, MinION, MinKNOW, PromethION, SmidgION, Ubik, and VoITRAX are registered trademarks of Oxford Nanopore Technologies plc in various countries. All other brands and names contained are the property of their respective owners. © 2022 Oxford Nanopore Technologies plc. All rights reserved. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition. PO_1110(EN)_V2_05May2022