



**HAL**  
open science

## scRNA-seq with Nanopore sequencing: benchmark of approaches based on hybrid sequencing

Ali Hamraoui, Catherine Sénamaud-Beaufort, Audrey Onfroy, Stéphane Le Crom,  
Morgane Thomas-Chollier, Laurent Jourdren, Sophie Lemoine

### ► To cite this version:

Ali Hamraoui, Catherine Sénamaud-Beaufort, Audrey Onfroy, Stéphane Le Crom, Morgane Thomas-Chollier, et al.. scRNA-seq with Nanopore sequencing: benchmark of approaches based on hybrid sequencing. jobim 2022, Jul 2022, Rennes (Campus de Beaulieu), France. . <hal-05144874>

**HAL Id: hal-05144874**

**<https://hal.science/hal-05144874v1>**

Submitted on 4 Jul 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# scRNA-seq with Nanopore sequencing: benchmark of approaches based on hybrid sequencing

Ali Hamraoui<sup>1</sup>, Catherine Senamaud-Beaufort<sup>1</sup>, Audrey Onfroy<sup>3,4</sup>, Stéphane Le Crom<sup>1,2</sup>, Morgane Thomas-Chollier<sup>1</sup>, Laurent Jourden<sup>1</sup> and Sophie Lemoine<sup>1</sup>

1. GenomiqueENS, Institut de biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
2. Sorbonne Université, CNRS, Institut de Biologie Paris-Seine (IBPS), Laboratory of Computational and Quantitative Biology (LCQB), 75005 Paris, France
3. Institut de Biologie de l'École Normale Supérieure, UMR CNRS 8197, Inserm U1024, PSL Université, Paris, France
4. Institut Mondor de Recherche Biomédicale, Inserm U955 - Team 9, Créteil, France

## Introduction

Long-read sequencing has recently been incorporated into single-cell RNA sequencing studies, and compared to the standard Illumina short-read library that primarily captures abundance information. Nanopore sequencing has several advantages: it can produce reads of several kb, it provides more precise transcriptome mapping/assembly and can detect complex isoform variants. However, this technique is limited by its low sequencing depth and the inaccurate assignment of cell barcodes (BC) and unique molecular identifiers (UMIs), which are critical for elimination of PCR bias and artifacts that are inherent to the sequencing material preparation [1]. To deal with this issue, several bioinformatics approaches have been developed using hybrid sequencing to guide the allocation of Nanopore reads using Illumina data. We performed a benchmark analysis of the available Single-Cell long-read hybrid sequencing based pipelines: SiCeLoRe [1], Snuupy [2] and FLAMES [3].

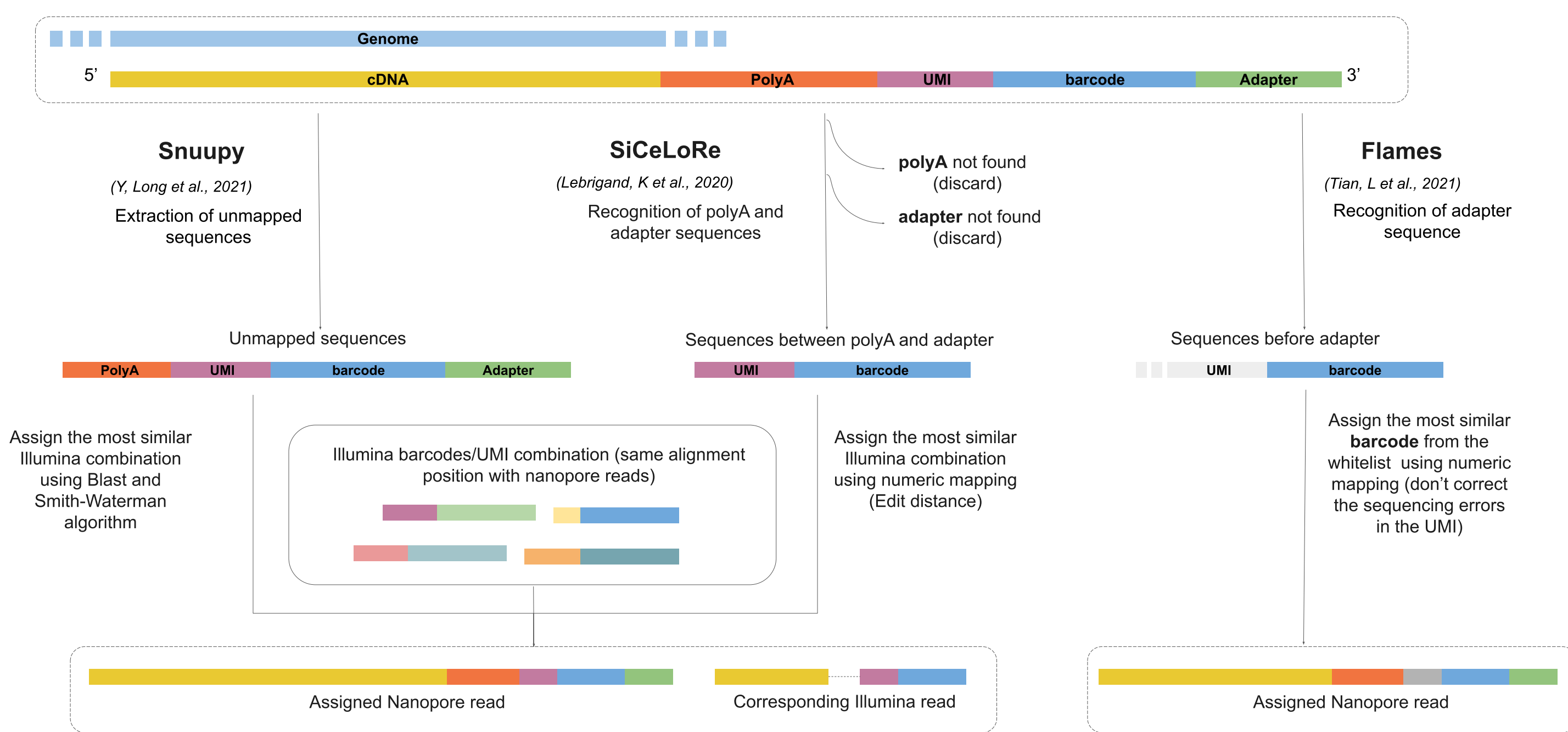
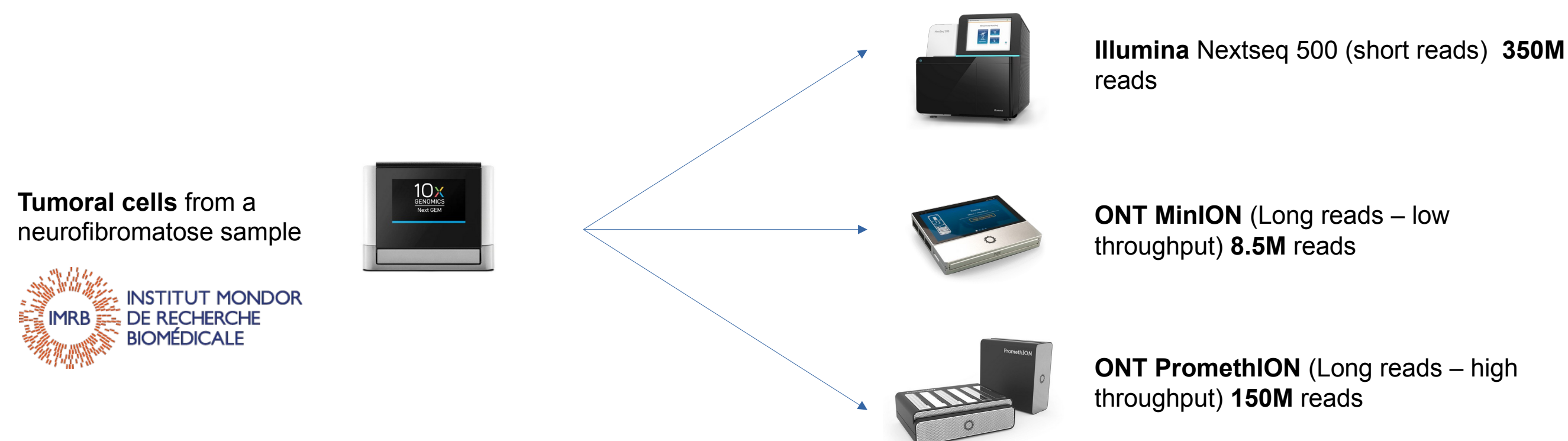


Figure 1: Available Single-Cell long-read hybrid sequencing based approaches: SiCeLoRe, Snuupy and FLAMES.

## Our dataset

We have the same biological sample from mice tumor sequenced on three different sequencing platforms:



## Barcodes and UMIs assignment

We compared the performance of the pipelines in term of correction and assignment of barcodes and UMIs. Our results show a high number of reads recovered by FLAMES and Snuupy but a good accurate assignment of SiCeLoRe on both MinION and PromethION data.

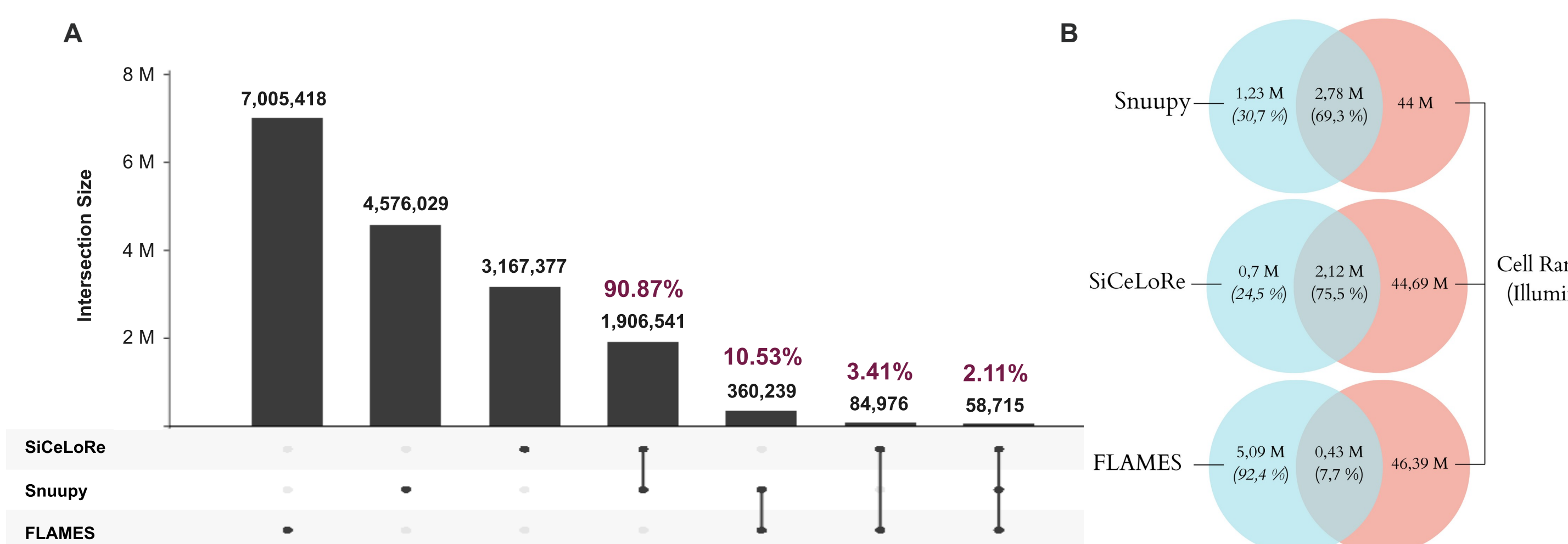


Figure 2: (A) Upset plot showing the overlap of readName-CellIBC-UMI between FLAMES, SiCeLoRe and Snuupy pipelines within MinION dataset (8.5M reads). (B) Venn diagrams showing the overlap of CellIBC-UMI between FLAMES, SiCeLoRe and Snuupy between the Illumina and the PromethION dataset (150M reads).

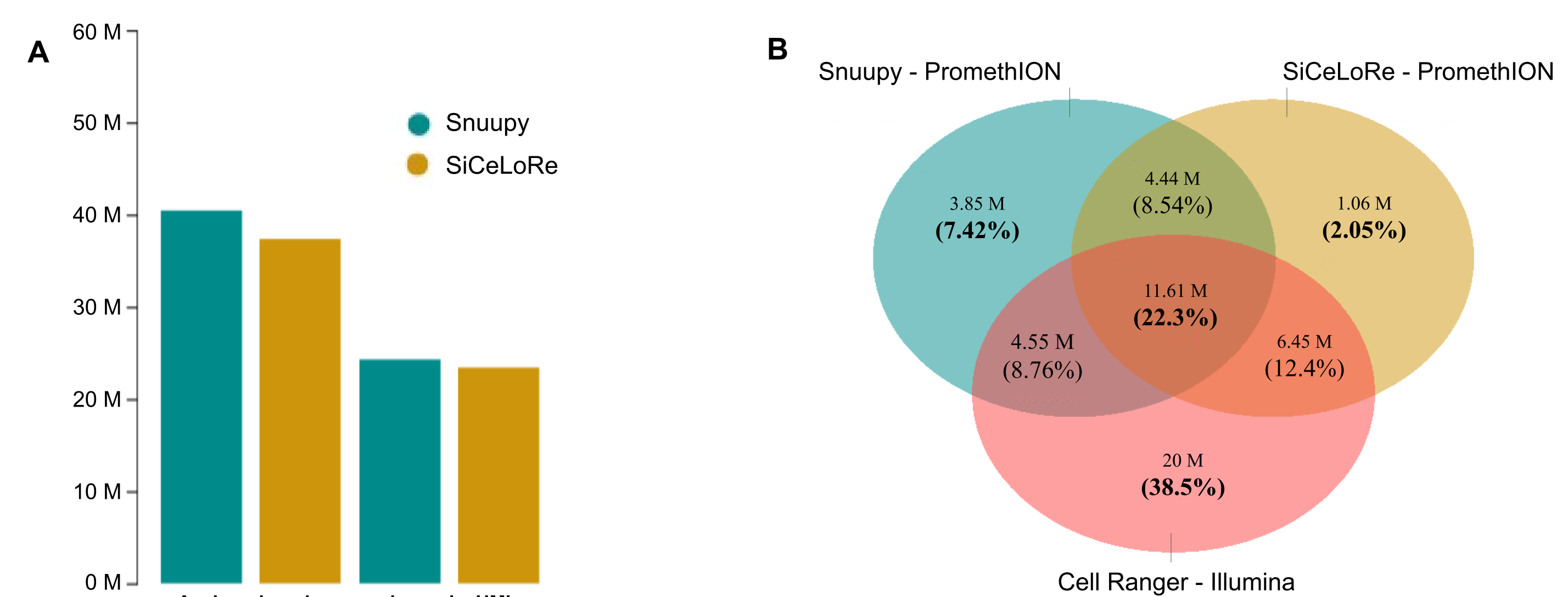


Figure 3: (A) Number of reads and CellIBC-UMI assigned (B) Venn diagrams showing the overlap of CellIBC-UMI between SiCeLoRe and Snuupy between the Illumina and the PromethION dataset (150M reads).

## Biological results

We then compared the biological results of SiCeLoRe and Snuupy vs Cell Ranger Illumina (short-read) at the gene level using the MinION and the PromethION dataset. Snuupy cannot perform on the PromethION dataset and the best results are produced with SiCeLoRe on the PromethION dataset.

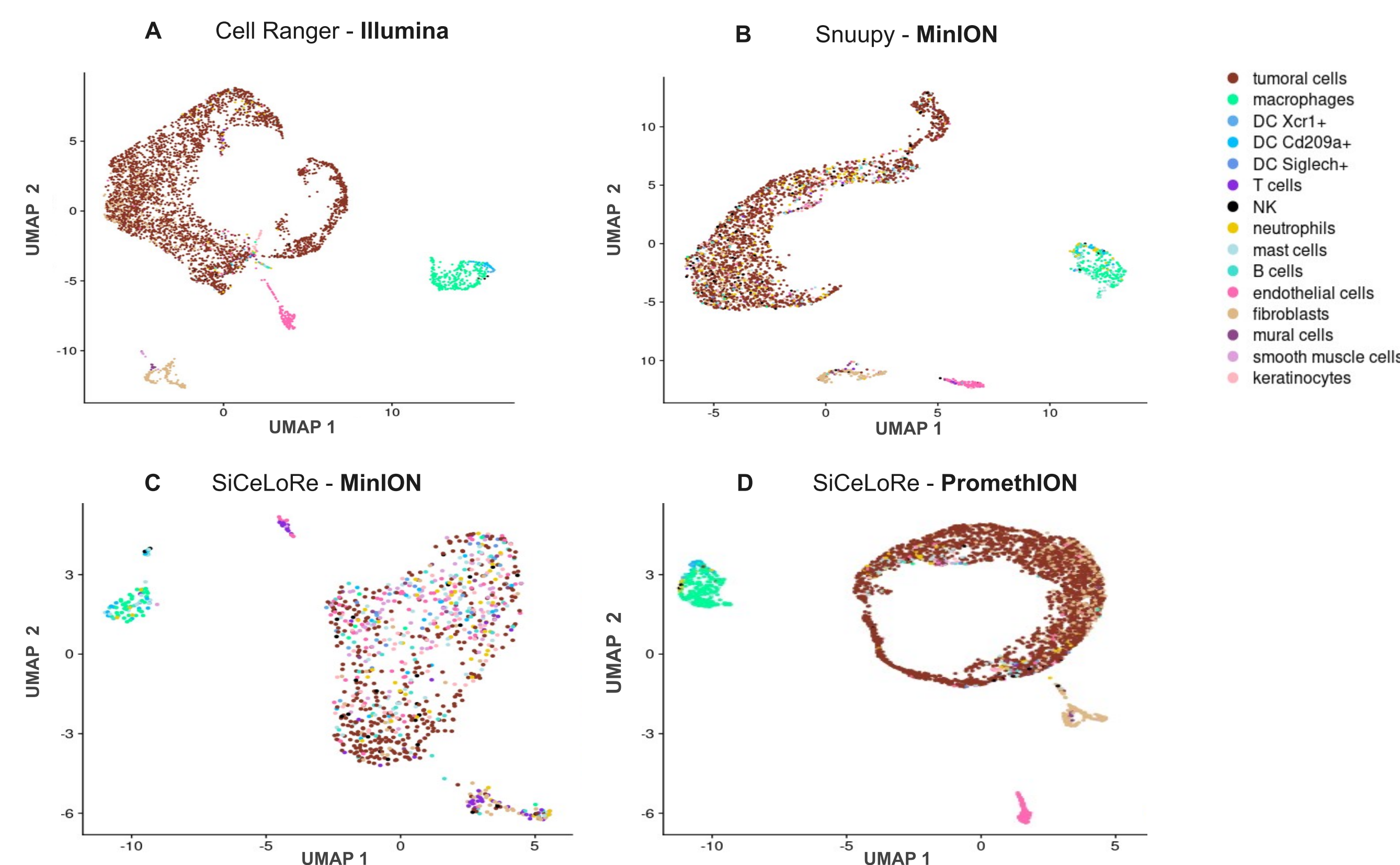


Figure 4: UMAP plots comparing the biological results of SiCeLoRe and Snuupy vs Cell Ranger Illumina (short-read) at the gene level using the MinION and the PromethION dataset.

## Isoform analysis

To compare SiCeLoRe isoform analysis module to FLAMES isoform detection module, we combined SiCeLoRe assignment part to the FLAMES isoform detection module. Then we compared the results of the full SiCeLoRe workflow to a hybrid SiCeLoRe-FLAMES workflow. SiCeLoRe outperform FLAMES in both quantity and diversity of isoforms detected.

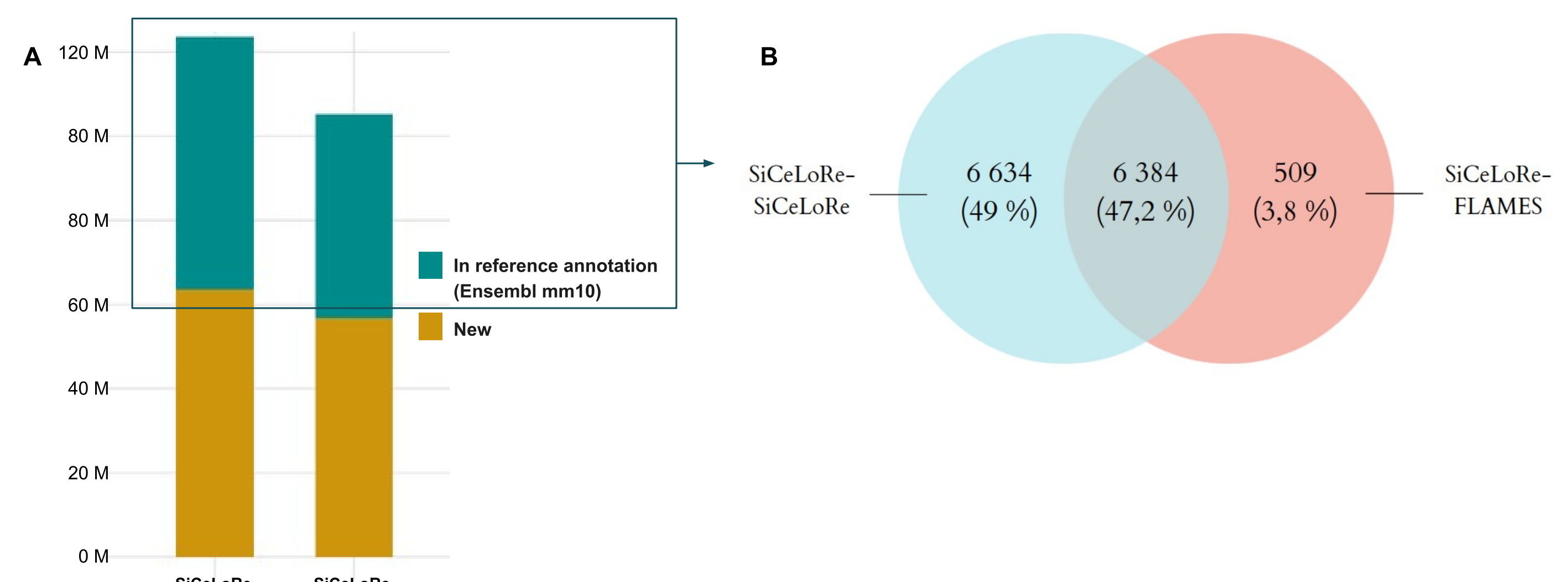


Figure 5: (A) Number of transcripts detected by each method. (B) Venn diagrams showing the degree of annotated isoform overlap between SiCeLoRe and FLAMES outputs.

Then we compared the biological results of SiCeLoRe and FLAMES at the isoform level of the Dcn gene.

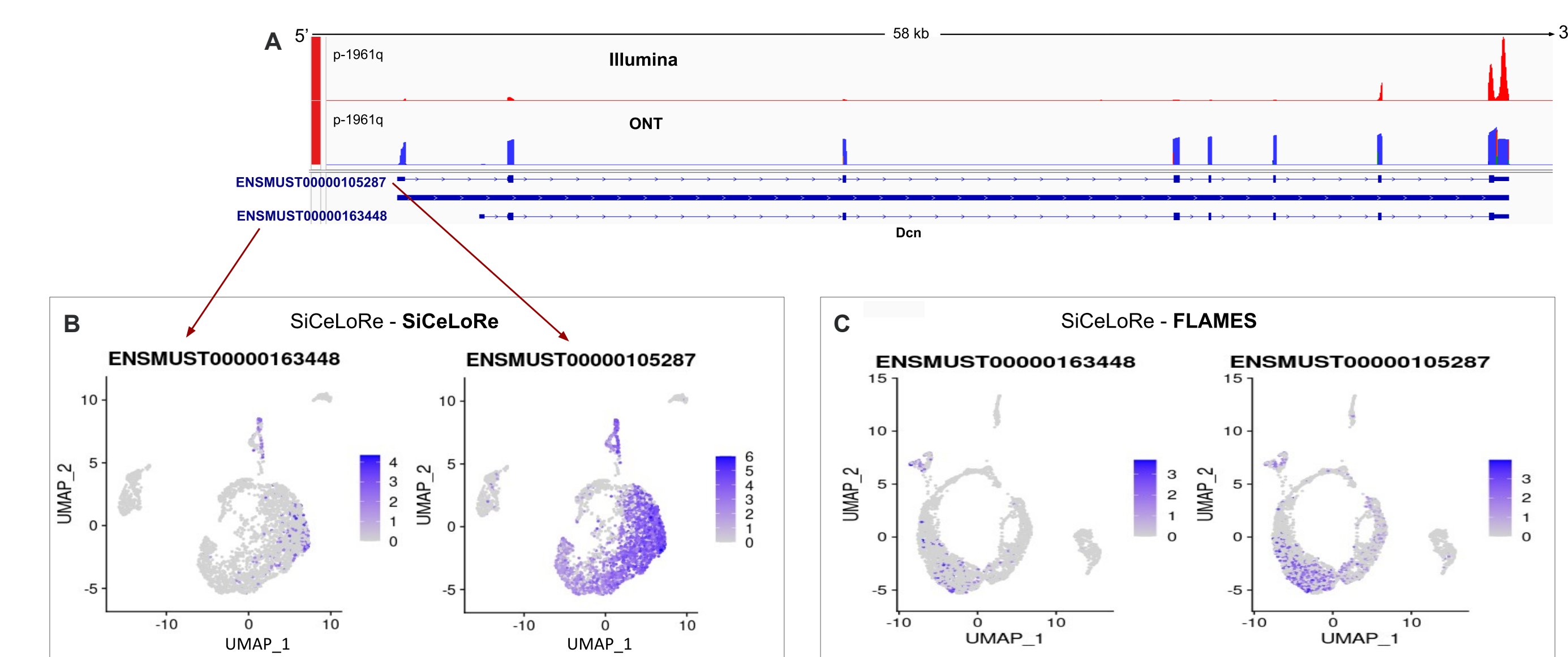


Figure 6: (A) Visualization of Dcn transcripts (IGV) (B and C) UMAP visualization of cells colored by the expression of two Dcn isoforms with differential expression across different cell types.

## Conclusion

- SiCeLoRe is an efficient method that provides a good compromise between the number of reads assigned, accuracy and computing cost.
- Snuupy is a good alternative to SiCeLoRe for small datasets, it recovers 15% more reads compared to SiCeLoRe in MinION and biological results are closer to the results generated on the Illumina data.
- SiCeLoRe can also detect and quantify isoforms correctly.

## References

- [1] Lebrigand, K., et al. High throughput error corrected Nanopore single cell transcriptome sequencing. Nat Commun 11, 4025 (2020)
- [2] Long, Y. et al. F1snRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. Genome Biol. 22, 66 (2021)
- [3] Tian, L. et al. Comprehensive characterization of single cell full-length isoforms in human and mouse with long-read sequencing. 2020.08.10.243543 (2020)