Introduction to single-cell RNA-seq analysis - Data sets and analyses

Stephane Ballereau

July 2021

Data set

Childhood acute lymphoblastic leukemia (cALL)

- Caron et al. 2020
- the most common pediatric cancer
- characterized by bone marrow lymphoid precursors
- that acquire genetic alterations,
- resulting in disrupted maturation and uncontrollable proliferation
- up to 85–90% of patients are cured
- others do not respond to treatment or relapse and die

Aim: characterise the heterogeneity of gene expression at the cell level, within and between patients

Samples

Four type of samples are considered:

- eight patients:
  - six B-ALL
    - four ‘t(12;21)’ or ‘ETV6-RUNX1’
    - two ‘High hyper diploid’ or ‘HHD’
  - two T-ALL (‘PRE-T’)
- three healthy pediatric controls
- eight healthy adult controls, publicly available

As the study aims at identifying cell populations, large numbers of cells were sequenced with the droplet-based 10X Chromium assay.

Analyses

We will follow several steps:

- sequencing quality check
- alignment of reads to the human genome (GRCh38) with 10X software cellranger
- quality control (cell calls, cells and genes filtering)
- UMI count normalisation
- data set integration (PBMMC and ETV6-RUNX1)
- feature selection and dimensionality reduction
- clustering
- marker gene identification
- differential expression and abundance between conditions
- trajectory analysis