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

November 2025

Data sets

- 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
  - **cells**: Bone Marrow Mononuclear cells (BMMCs)

Samples

Five types of sample are considered:

- B-ALL patients:
  - **'ETV6-RUNX1'**, or 't(12;21)', four patients
  - **'HHD'**, or 'High hyper diploid', two patients
- T-ALL patients
  - **'PRE-T'**, two patients
- Healthy controls,
  - **Healthy pediatric controls**, three individuals

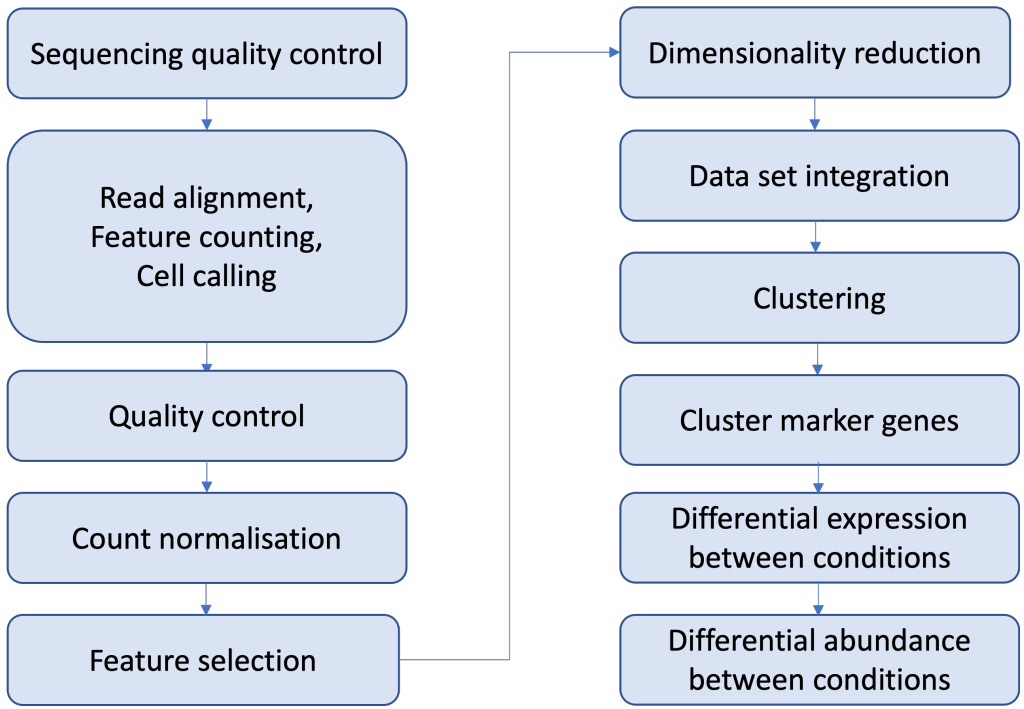
As the study aims at identifying cell populations, large numbers of cells were sequenced with the droplet-based 10x Chromium assay.

Analyses

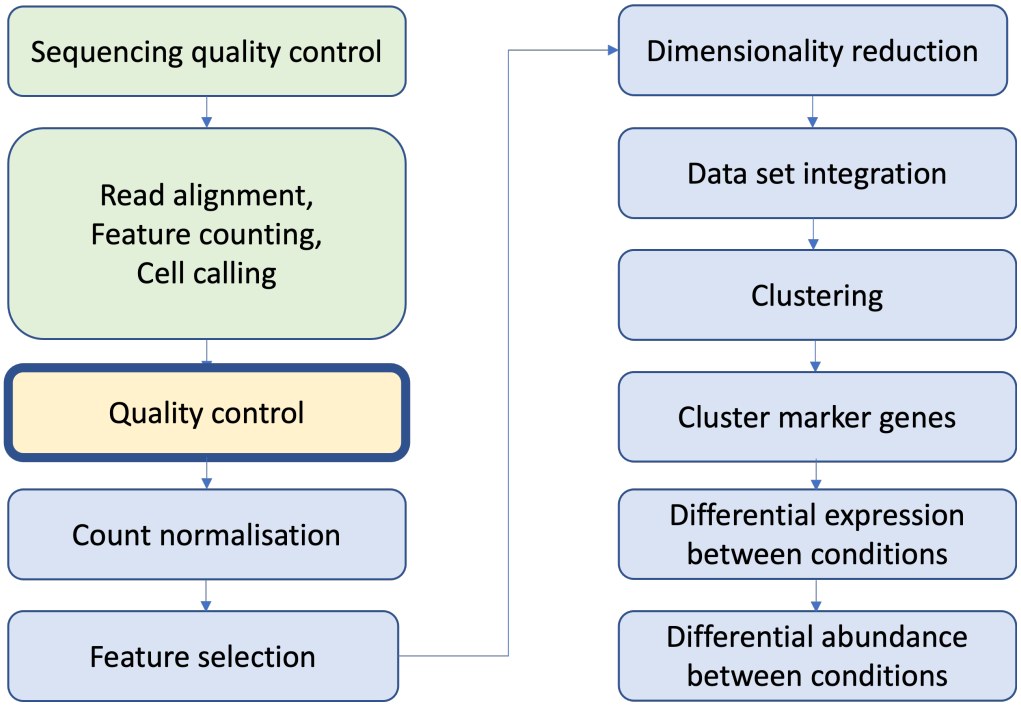
We will follow several steps:

- Sequencing quality check
- Alignment of reads to the human genome with 10x software cellranger
- Quality control (filter poor quality cells and remove uninformative genes)
- UMI count normalisation
- Feature selection and dimensionality reduction
- Data set integration
- Clustering
- Identification of cluster marker genes
- Differential expression and abundance between conditions

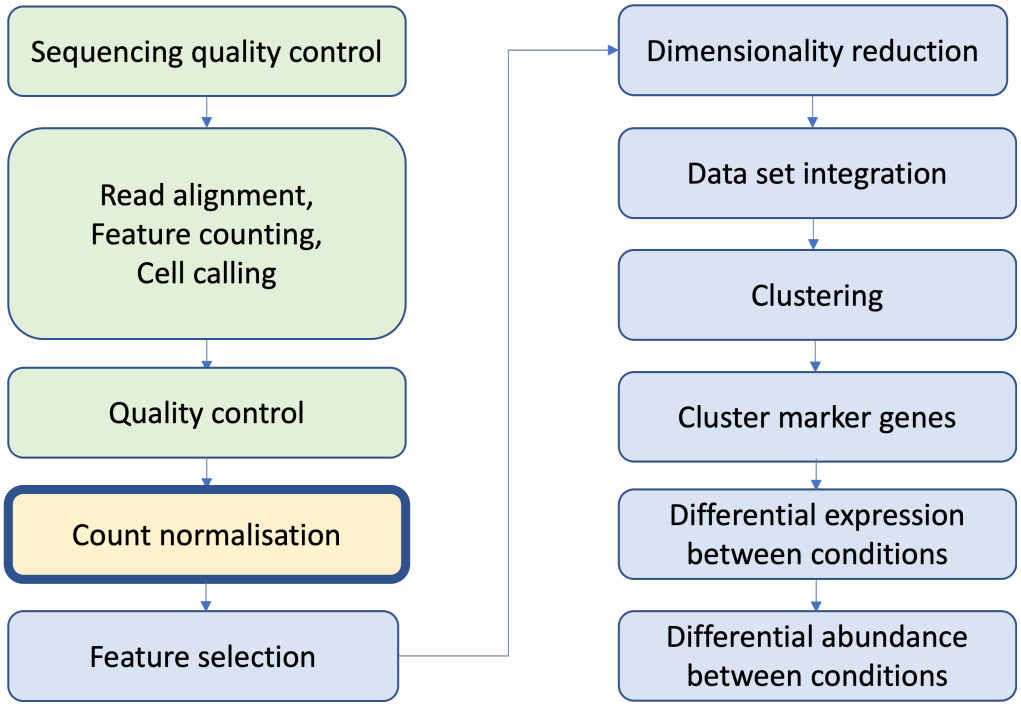
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

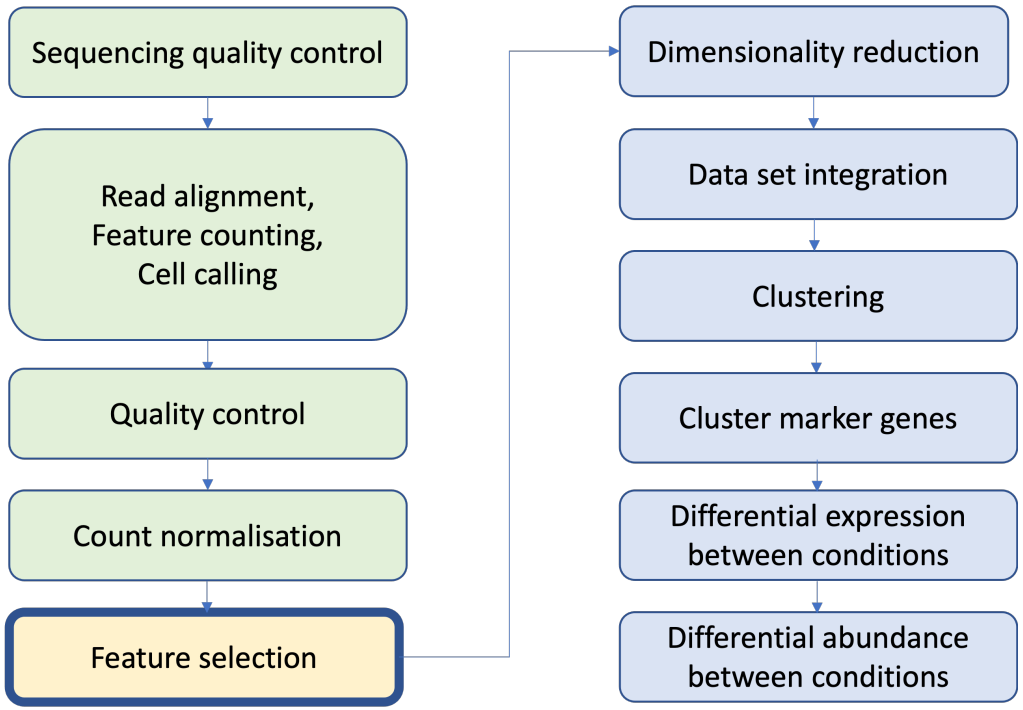
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

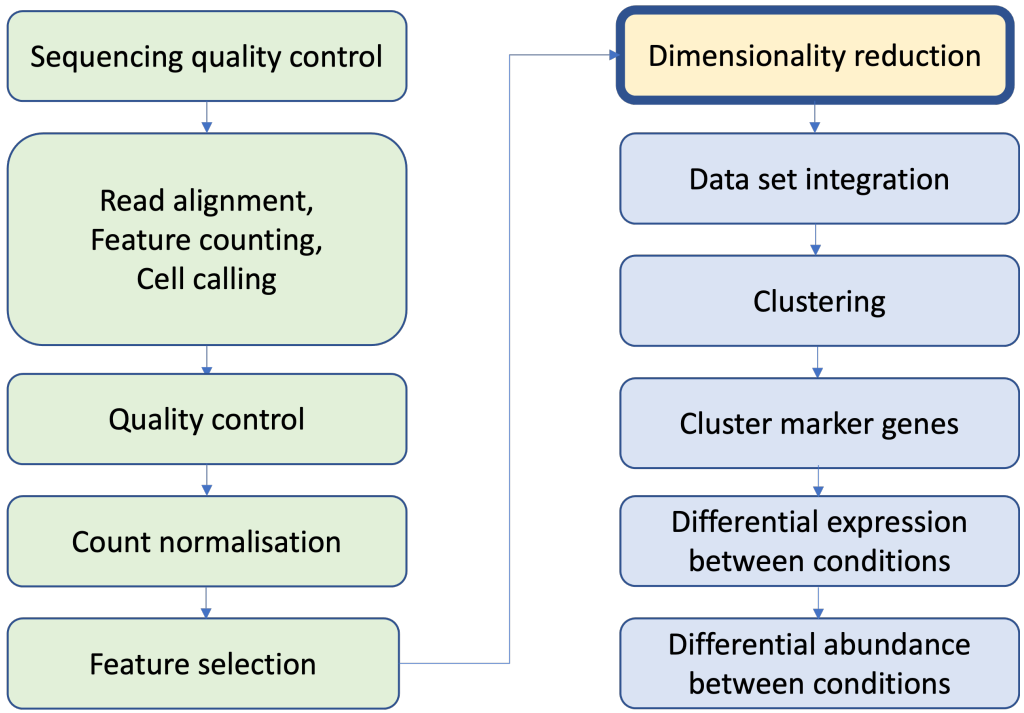
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

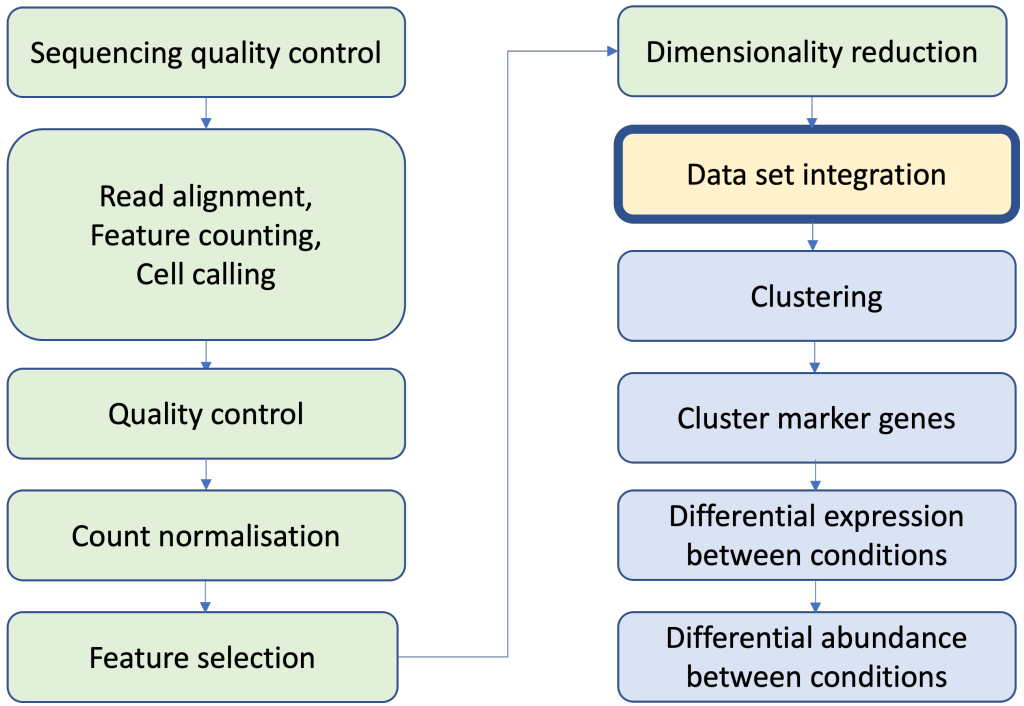
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

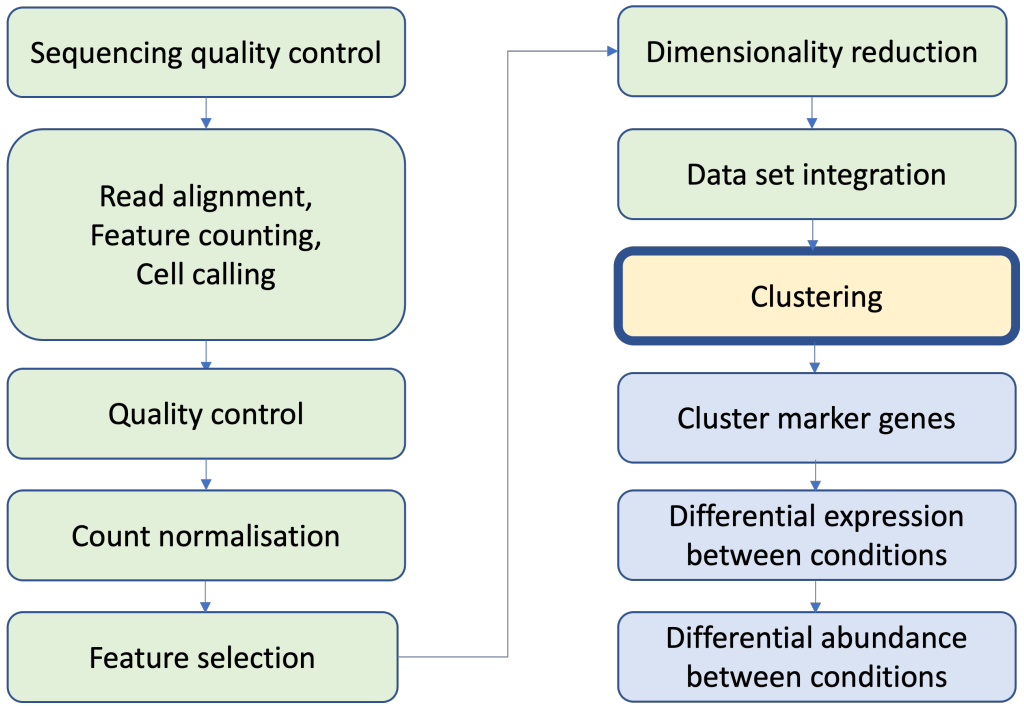
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

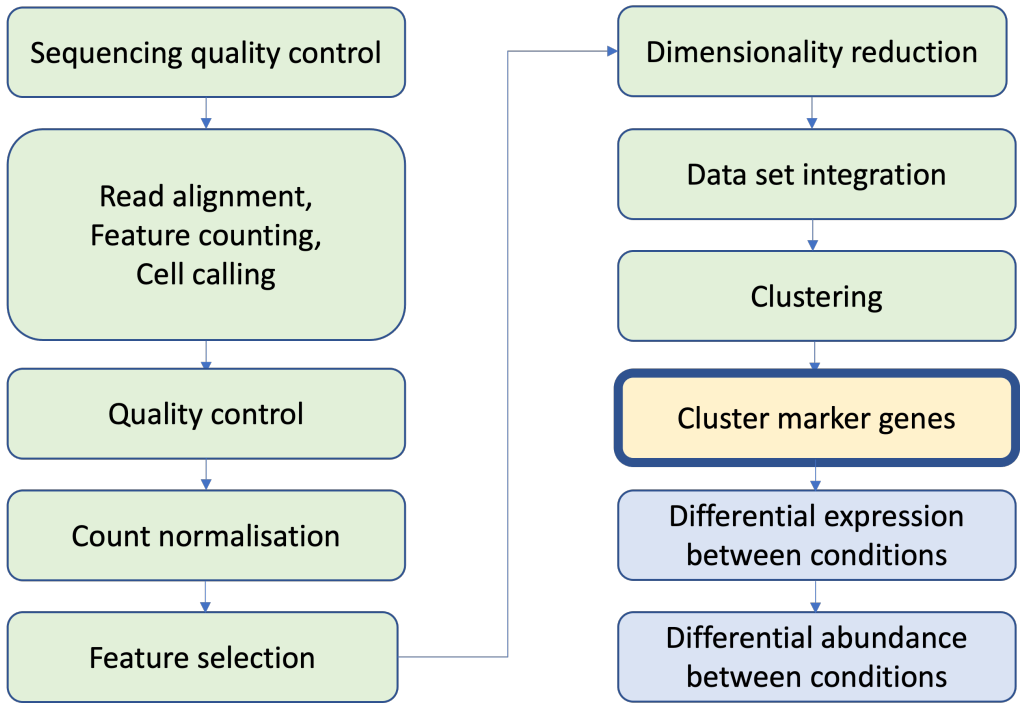
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

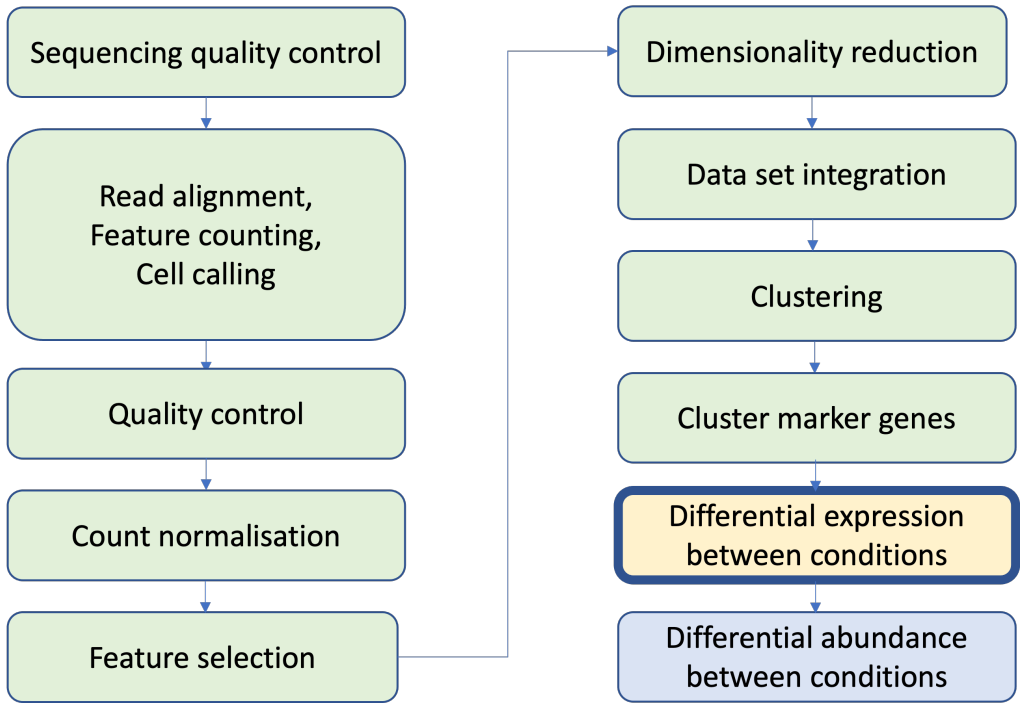
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

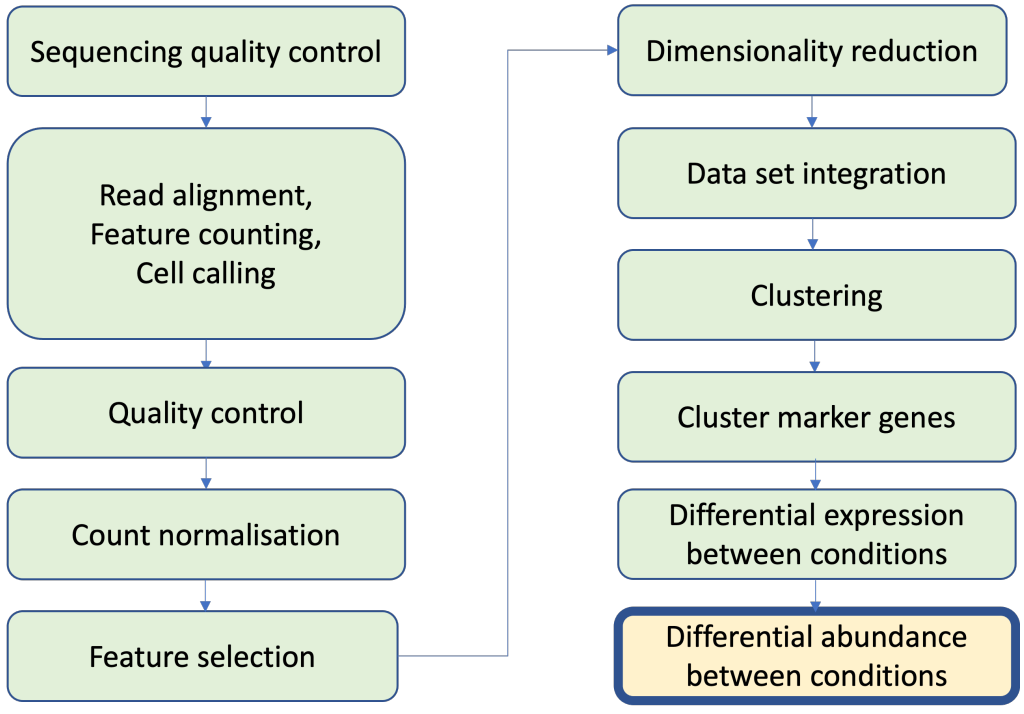
Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions

Workflow


Sequencing quality control
Read alignment,
Feature counting,
Cell calling
Quality control
Count normalisation
Feature selection
Dimensionality reduction
Data set integration
Clustering
Cluster marker genes
Differential expression
between conditions
Differential abundance
between conditions