

UNIVERSITY OF
CAMBRIDGE



CANCER
RESEARCH
UK

Cambridge
Institute

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

September 2022

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

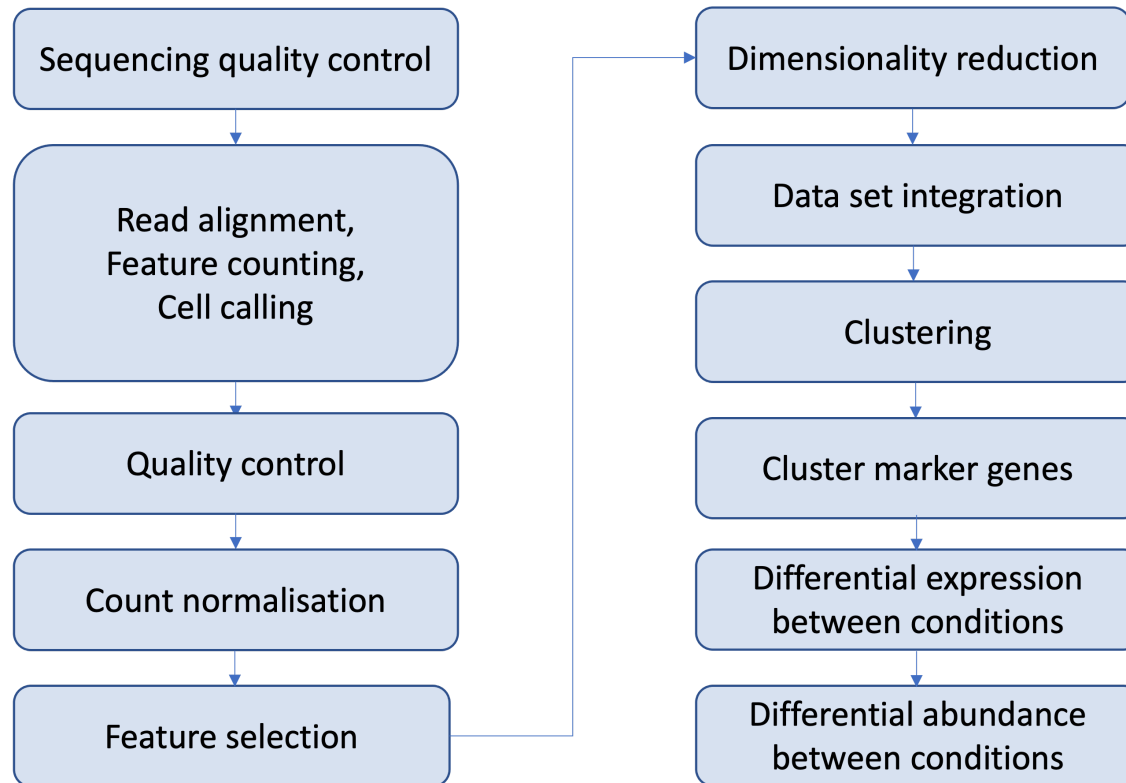
As the study aims at identifying cell populations, large numbers of cells were sequenced with the droplet-based 10x Chromium assay.

Analyses

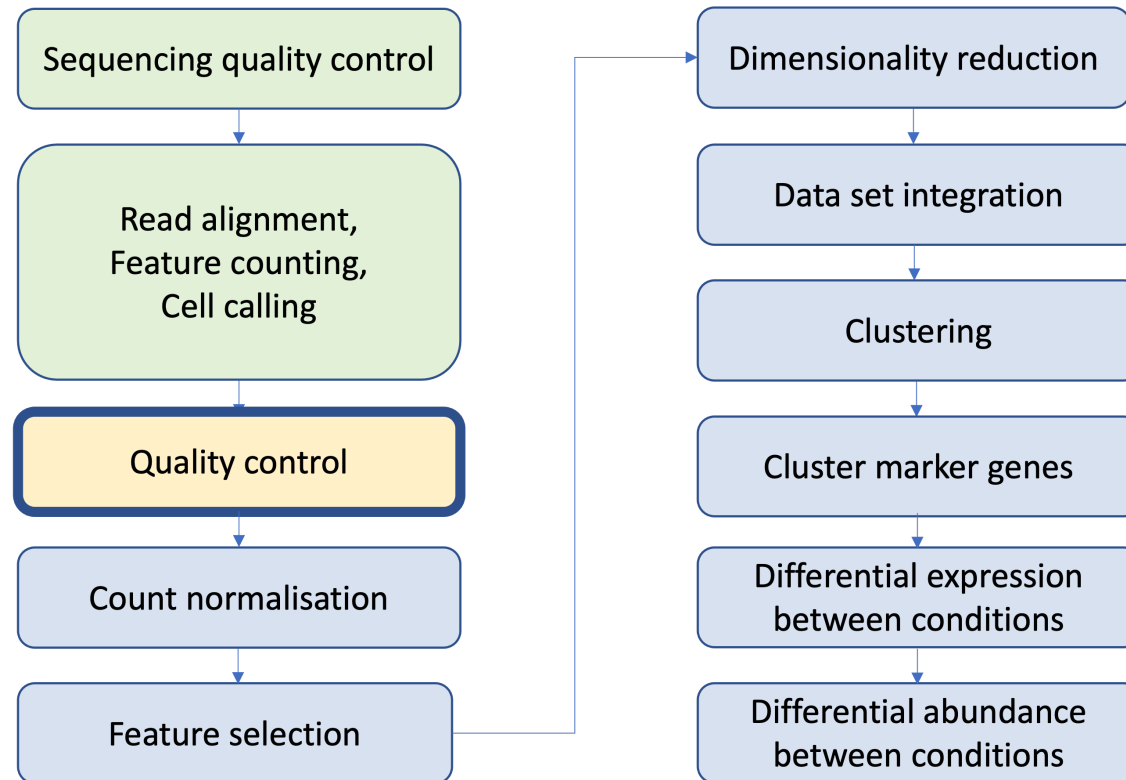
We will follow several steps:

- 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

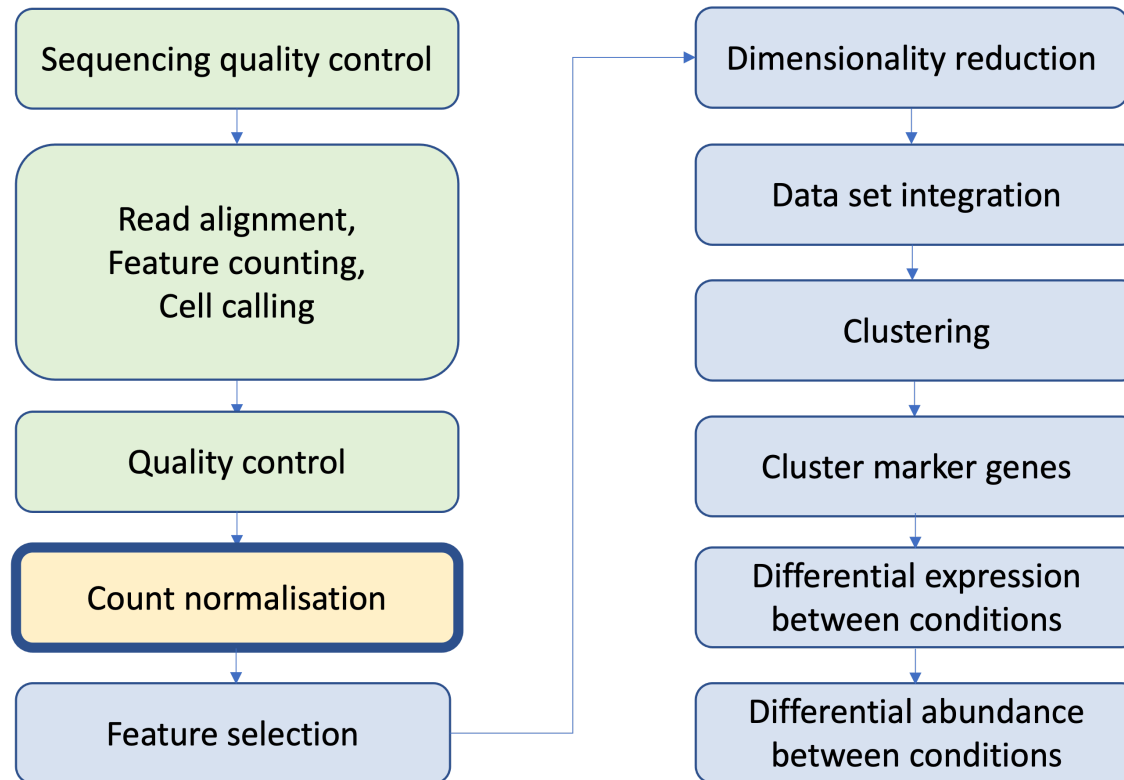
Workflow



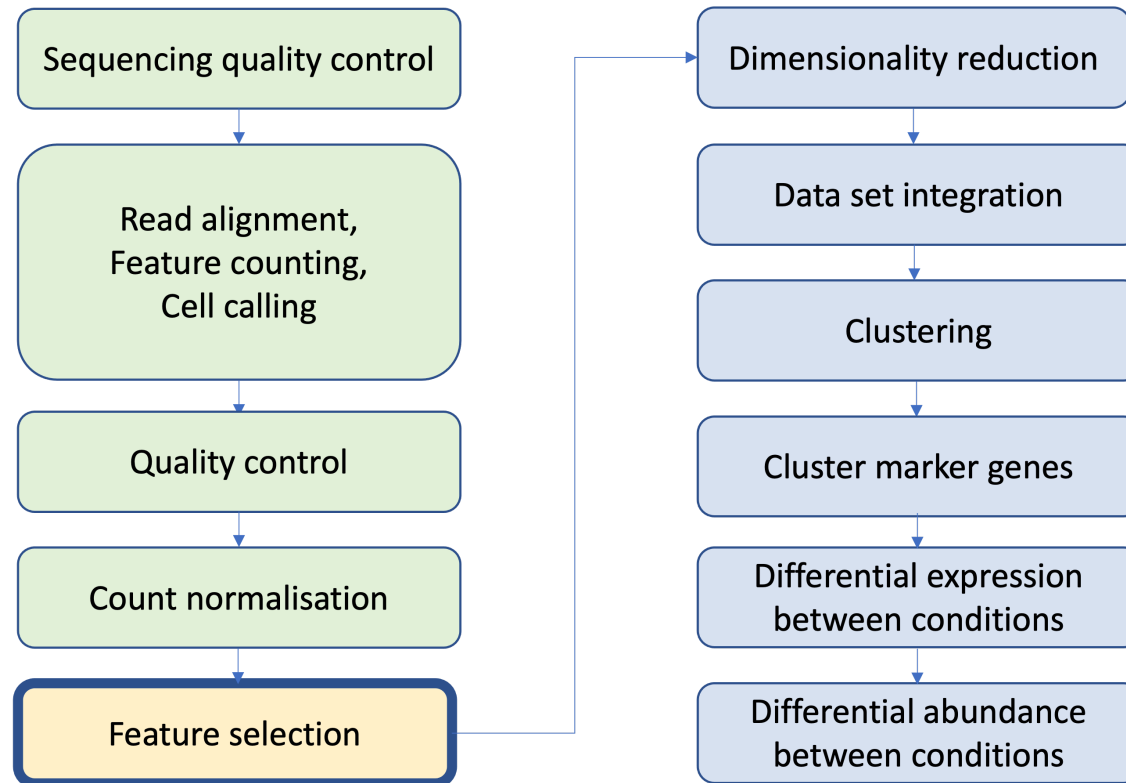
Workflow



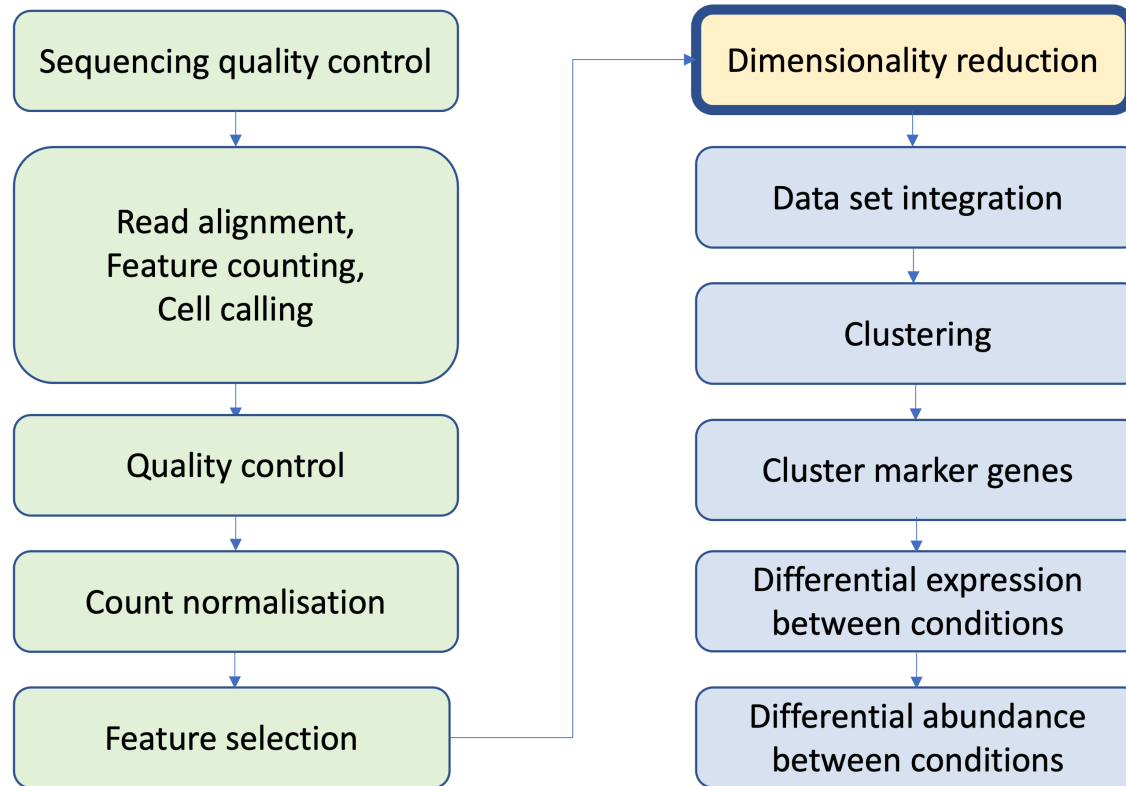
Workflow



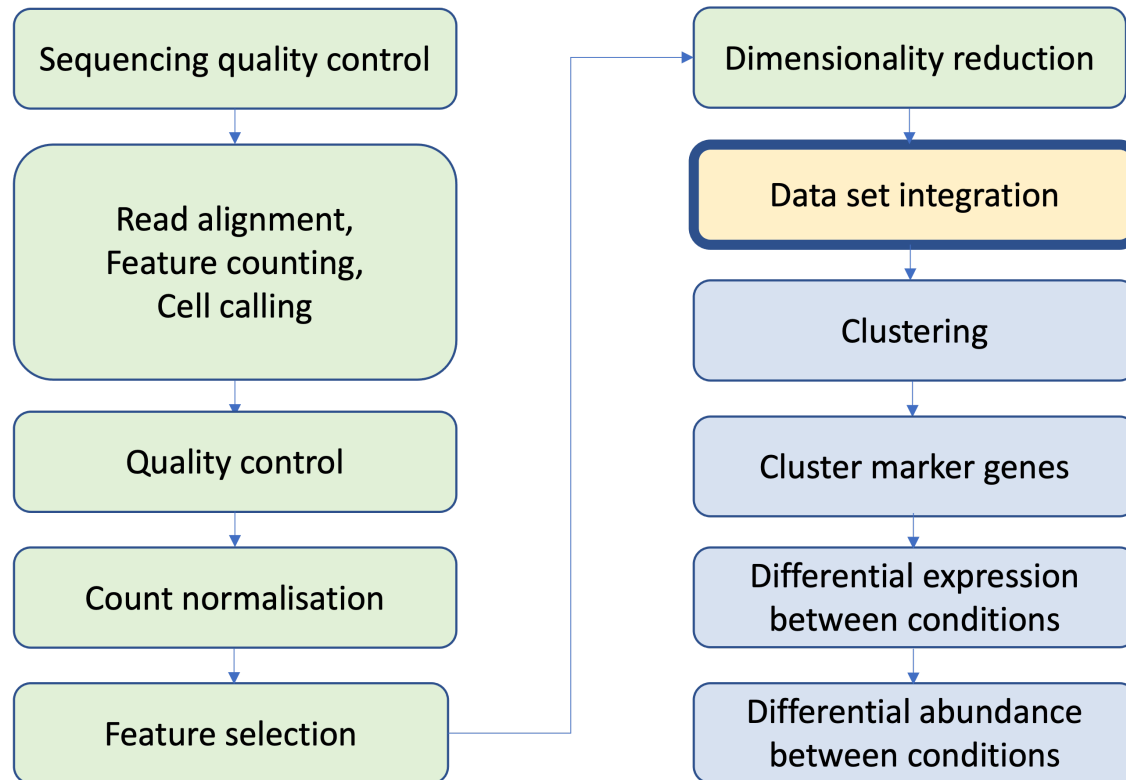
Workflow



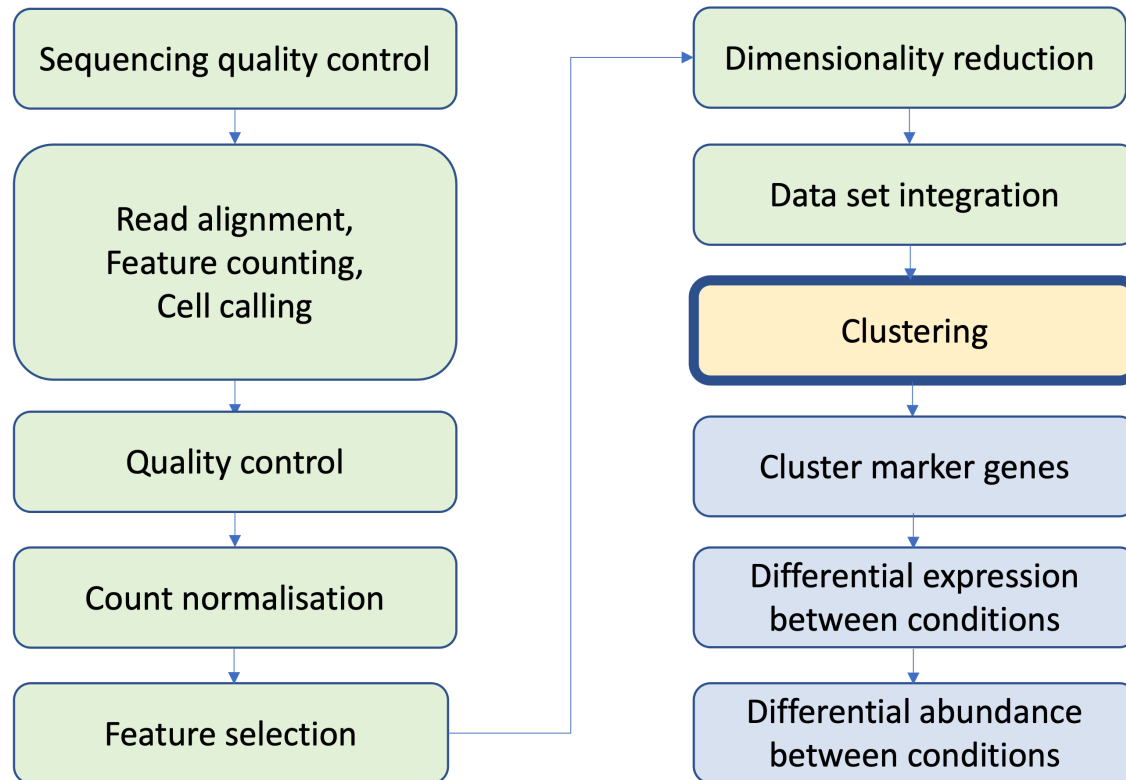
Workflow



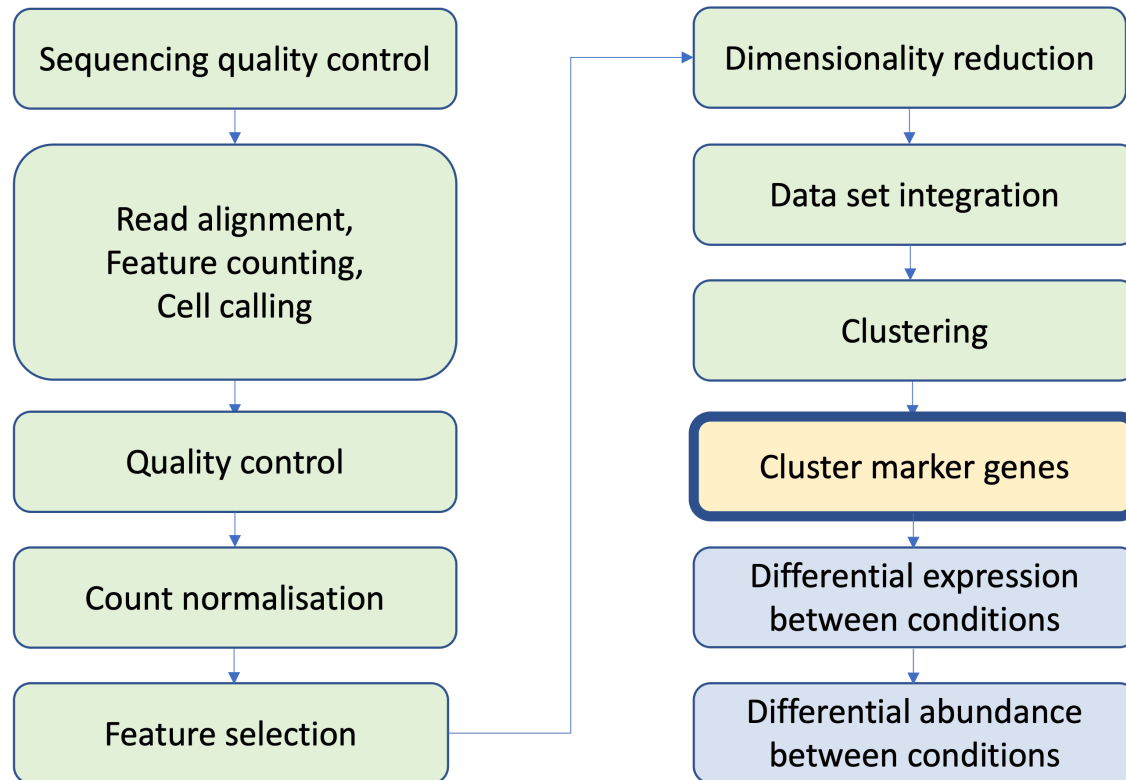
Workflow



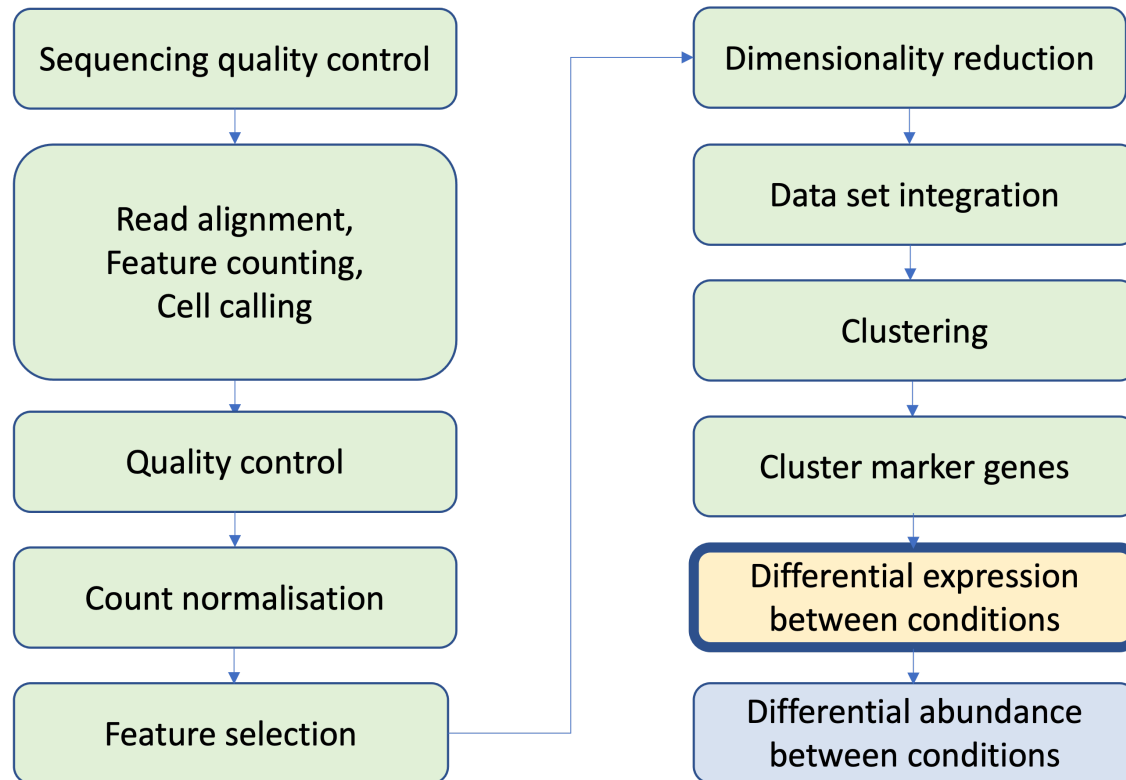
Workflow



Workflow



Workflow



Workflow

