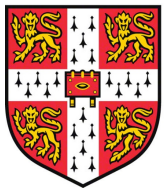


UNIVERSITY OF CAMBRIDGE

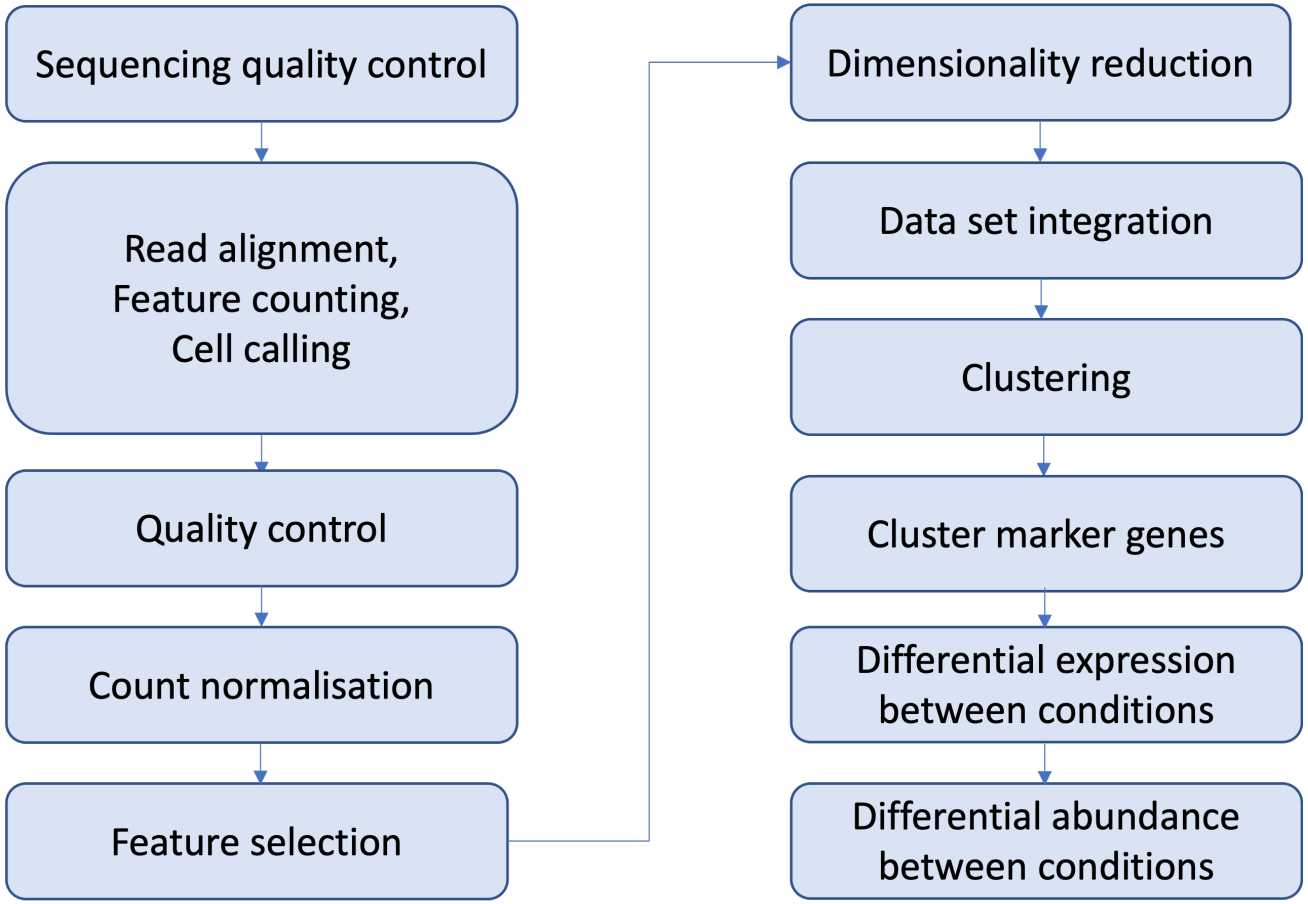
CANCER RESEARCH UK | CAMBRIDGE CENTRE

single-cell RNA-seq Normalisation

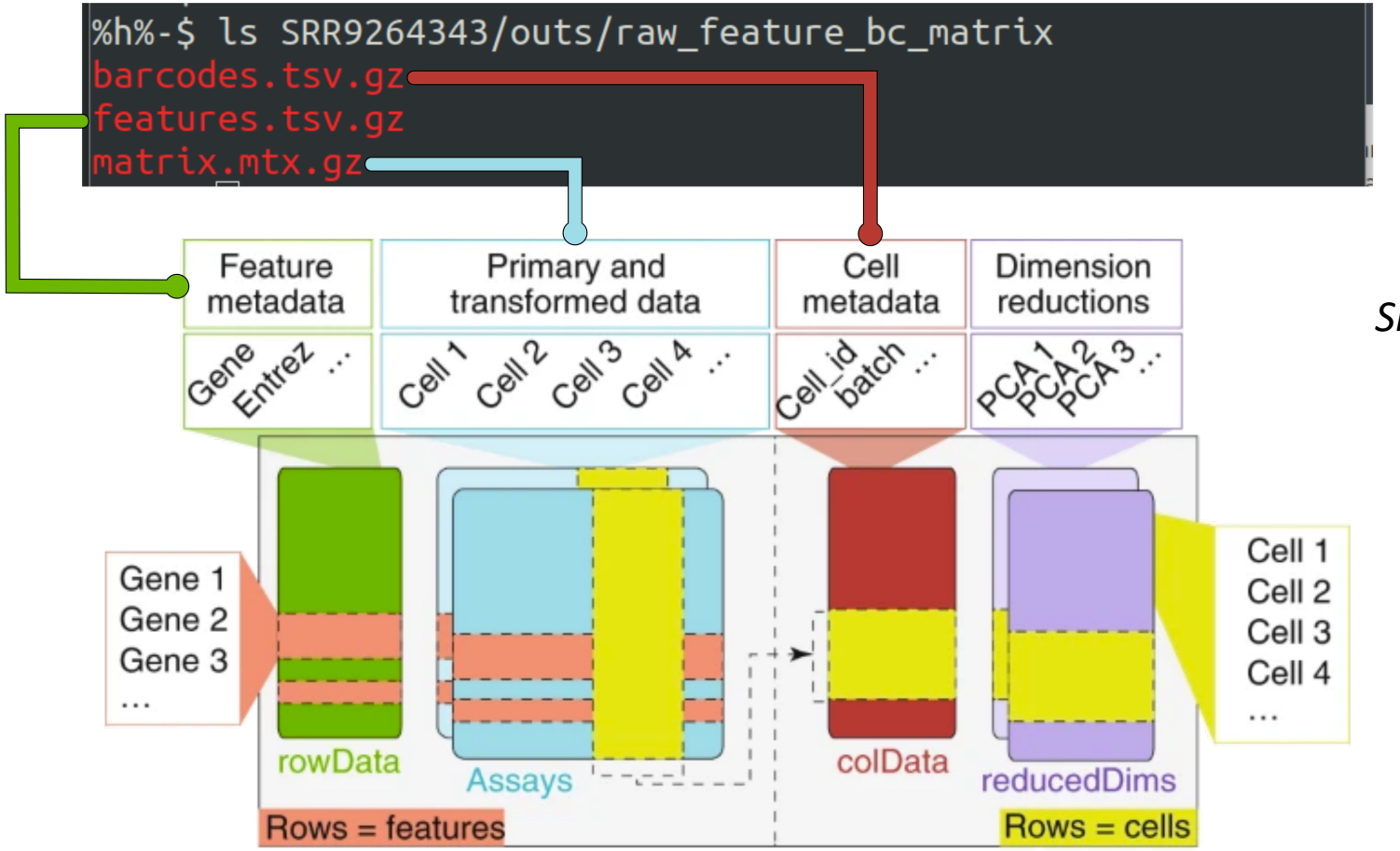
Kamal Kishore

03/05/2022

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

Quality Control (Recap)

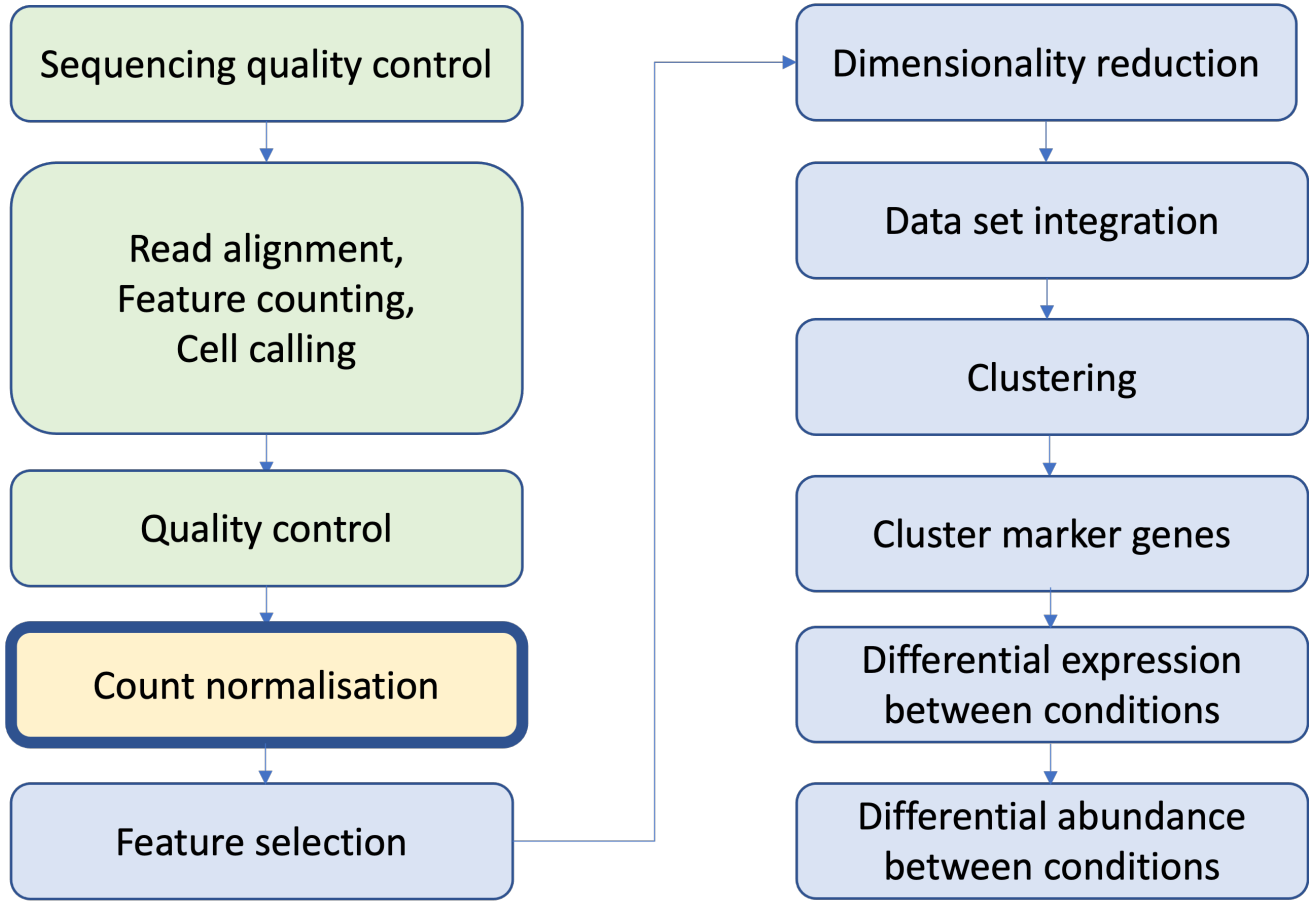


SingleCellExperiment Object

Quality Control (Recap)

- Remove undetected genes
- Empty = no cell in droplet : low gene count
- Doublet/multiplet = more than one cell in droplet : large gene count
- Broken/dead cell in droplet : high proportion of UMIs mapping to mitochondrial genes

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

Why Normalize?

Systematic differences in sequencing coverage between libraries occur because of:

- ✓low input material
- ✓differences in mRNA capture
- ✓differences in PCR amplification

Scaling Normalization

✓ In scaling normalization, the “normalization factor” is an estimate of the library size relative to the other cells.

✓ Steps usually include:

- Computation of a cell-specific 'scaling' or 'size' factor
- division of all counts for the cell by that factor to remove that bias.

✓ Assumption: any cell specific bias will affect genes the same way.

Library Size Normalization

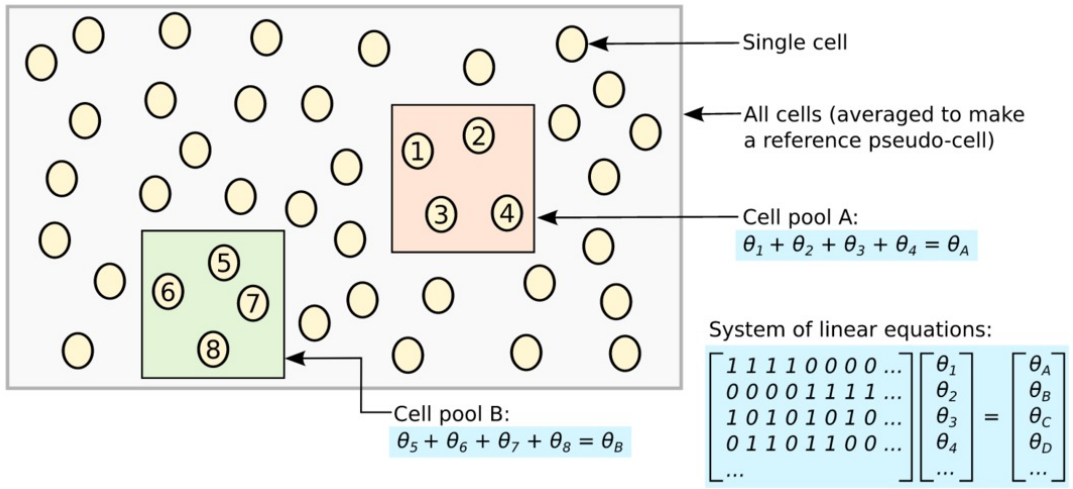
- ✓ The library size factor is proportional to the library size such that the average size factor across cell is one.
- ✓Divide gene's UMI count in a cell by total number of UMIs in the cell.
- ✓Multiply the ratio by a scale factor (10,000 by default).
- ✓Transform the result by taking natural log.

Deconvolution

Deconvolution strategy [Lun et al 2016]

The deconvolution method consists of several key steps:

- Defining a pool of cells
- Summing expression values across all cells in the pool
- Normalizing the cell pool against an average reference, using the summed expression values
- Repeating this for many different pools of cells to construct a linear system
- Deconvolving the pool-based size factors to their cell-based counterparts (Fig. 3)



Steps :

- *compute Scaling Factor*
- *apply scaling Factor*