

## scRNA Results

The pipeline is based on the standard Seurat (version 4) workflow and consists of a number of stages. Below is a list of the files (mostly plots/pdfs) and a short description of their contents.

Details on the Seurat package: <https://satijalab.org/seurat/>

### Stage I - Prefilter

- **seuratQC\_analysis\_01\_QC.pdf**: Plots showing a number of metrics to assess the quality of cells and help in the selection for threshold for cell filtering. The key metrics plotted are:
  - nFeature\_RNA: the number of genes detected in a given cell
  - nCounts\_RNA: the number of unique molecules found in each cell
  - percent.mt: the number of reads that map to mitochondrial genes.
- **seuratQC\_analysis\_02\_CellCycle.pdf**: A plot of the first two PCA coordinates colored by the cell cycle phase of the cells which is used to assess the extent of the cell cycle signal in the dataset.

### Stage II - Post filtering assessment

These plots assess how well the cell level filtering was done and also how well the Cell Cycle correction (via regression at the SCTransform step) worked. Not at this step the data is normalized and scaled using the SCTransform method.

- **seuratQC\_analysis\_11\_PostFilterQCTbls.pdf**: Table of the number and percentage of cells filtered out or retained:
- **seuratQC\_analysis\_12\_PostIntegrateCC.pdf**: Plot of top two PCA coordinates (same as stage I) but now after normalization and scaling with the cell cycle score regression done.
- **seuratQC\_analysis\_13\_VariableFeatures.pdf**: The final plot show the genes with highest variability. Should be inspected to determine if there are any potential artifacts that should be filtered out (such as an ribosomal or mitochondrial genes)

### Linear reduction, clustering and projection

At this stage the PCA transformation is done on the scaled data using the top (typically 2000) more variable features (genes) determined in the previous step. The top N PCA coordinates (typically 50) are then used to both cluster the data and also to compute the 2D projection (UMAP). The clustering is done at a number of different resolutions.

- **seuratQC\_analysis\_21\_PCADimMetric.pdf**: Variance as a function of pca coordinate
- **seuratQC\_analysis\_22\_UMAP\_20\_.pdf**: UMAP's colored by:
  - Cluster number at various resolutions
  - Samples: look for possible sample batch effects
  - Cell Cycle Phase: another assessment of how well cell cycle effects were removed
- **seuratQC\_analysis\_23\_ClusterChart\_20\_.pdf**: Fraction of samples per cluster, cluster per samples, cell cycle phase per clusters. QC for possible batch effects and to help decide the optimal cluster number.

### Cluster specific genes:

For a given cluster resolution; either chosen by default or selected after review of the cluster QC plots we find the genes that are differentially expressed between the clusters. We show the distribution of these marker genes expressions amount the cluster. Also included is a table of the marker genes.

- **seuratQC\_analysis\_31\_ClusterMarkersDot\_SCT\_snn\_res.0.2\_FDR\_0.05\_logFC\_1\_.pdf**:

- `seuratQC_analysis_32_ClusterMarkers_SCT_snn_res.0.2_FDR_0.05_logFC_1.pdf`:
- `tblClusterMarkers_SCT_snn_res.0.2_FDR_0.05_logFC_1.xlsx`: