Seurat Cheatsheet

Basic Concepts

Seurat is an R package designed for QC, analysis, and exploration of single-cell RNA-seq data.

- Seurat Object: The central data structure holding counts, metadata, and analysis results (reductions, neighbors, etc.).
- Assay: Represents different data types (e.g., "RNA", "SCT", "ADT").

Workflow Start: Object Creation and QC

```
library(Seurat)

# Create object
sobj ← CreateSeuratObject(counts =
counts_matrix, project = "sample1",
min.cells = 3, min.features = 200)

# QC Metrics Calculation (e.g.,
mitochondrial percentage)
sobj[["percent.mt"]] ←
PercentageFeatureSet(sobj, pattern =
"^MT-")

# Visualization of QC
VlnPlot(sobj, features =
c("nFeature_RNA", "nCount_RNA",
"percent.mt"), ncol = 3)

# Filtering
sobj ← subset(sobj, subset =
nFeature_RNA > 200 & nFeature_RNA < 2500
& percent.mt < 5)
```

Normalization and Feature Selection

```
# Standard Log Normalization
sobj ← NormalizeData(sobj,
normalization.method = "LogNormalize",
scale.factor = 10000)

# Variable Feature Selection (HVGs)
sobj ← FindVariableFeatures(sobj,
selection.method = "vst", nfeatures =
2000)

# Scaling (Standardizing)
sobj ← ScaleData(sobj, features =
rownames(sobj))
```

Dimensionality Reduction

```
# Principal Component Analysis
sobj ← RunPCA(sobj, features =
VariableFeatures(object = sobj))

# Visualization of PCA
VizDimLoadings(sobj, dims = 1:2,
reduction = "pca")
DimPlot(sobj, reduction = "pca")
DimHeatmap(sobj, dims = 1, cells = 500,
balanced = TRUE)

# Non-linear Reduction (UMAP/tSNE)
sobj ← RunUMAP(sobj, dims = 1:10)
DimPlot(sobj, reduction = "umap")
```

Clustering

```
# Find Neighbors and Clusters
sobj ← FindNeighbors(sobj, dims = 1:10)
sobj ← FindClusters(sobj, resolution =
0.5)

# View Clusters
table(Idents(sobj))
```

Differential Expression (Marker Identification)

```
# Find all markers for all clusters
all_markers ← FindAllMarkers(sobj,
only.pos = TRUE, min.pct = 0.25,
logfc.threshold = 0.25)

# Find markers for a specific cluster
cluster1_markers ← FindMarkers(sobj,
ident.1 = 1, min.pct = 0.25)
```

Advanced Visualization

- `FeaturePlot()`: Visualize gene expression on UMAP/tSNE.
- `VlnPlot()`: Violin plots of gene expression per cluster.
- `DotPlot()`: Dot plots showing percentage of cells express and avg expression.
- `DoHeatmap()`: Heatmap of top markers across cells/clusters.

SCTransform (Modern Normalization)

```
# Single command for normalization,
finding features, and scaling
sobj ← SCTransform(sobj,
vars.to.regress = "percent.mt", verbose
= FALSE)
```

Tips for scRNA-seq Analysis

Determining Dimensionality

Use `ElbowPlot(sobj)` or `JackStraw` (deprecated in newer versions, manual check preferred) to decide how many PCs to include for clustering and UMAP.

Cell Cycle Scoring

Calculate cell cycle phase scores using `CellCycleScoring()` and regress them out in `ScaleData` or `SCTransform` if they introduce unwanted bias.

Multi-sample Integration

Use standard Integration workflows (`FindIntegrationAnchors`, `IntegrateData`) or newer SCT-based methods to remove batch effects.