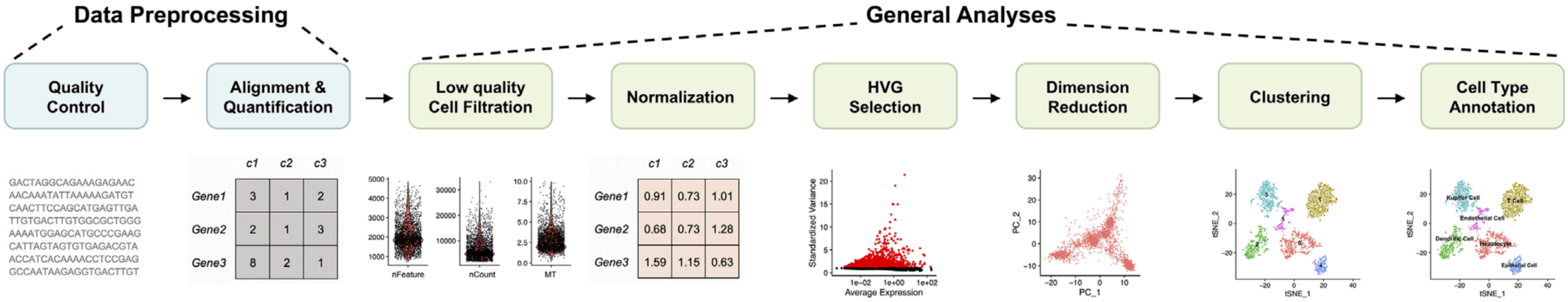
SINGLE CELL RNA-seq WORKSHOP

TREX x BioHPC

Week 2

Faraz Ahmed - 02/19/24

Analysis Overview:



Data Preprocessing

General Analyses

Quality Control → Alignment & Quantification → Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

GACTAGGCAGAAAGAGAAC
AACAAATATTAAAAAGATGT
CAACTTCCAGCATGAGTTGA
TTGTGACTTGTGGCGCTGGG
AAAATGGAGCATGCCCGAAG
CATTAGTAGTGTGAGACGTA
ACCATCACAAAACCTCCGAG
GCCAATAAGAGGTGACTTGT

	c1	c2	c3
Gene1	3	1	2
Gene2	2	1	3
Gene3	8	2	1

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

DOI: 10.1002/ctm2.694

REVIEW

CLINICAL AND TRANSLATIONAL MEDICINE Open Access WILEY

Single-cell RNA sequencing technologies and applications: A brief overview

Dragomirka Jovic[1,2] | Xue Liang[1,2,3] | Hua Zeng[4] | Lin Lin[5,6] | Fengping Xu[1,2] | Yonglun Luo[1,2,5,6]

Cellranger 10x Suite

Data Preprocessing
General Analyses
Quality Control
Alignment & Quantification
Low quality Cell Filtration
Normalization
HVG Selection
Dim Red
5000
4000
3000
2000
1000
nFeature
30000
20000
10000
nCount
10.0
7.5
5.0
2.5
0.0
MT
Standardized Variance
20
15
10
5
0
1e−02
1e+00
1e+02
Average Expression
PC_2
30
20
10
0
−10
−20
−1

GACTAGGCAGAAAGAGAAC
AACAAATATTAAAAAGATGT
CAACTTCCAGCATGAGTTGA
TTGTGACTTGTGGCGCTGGG
AAAATGGAGCATGCCCGAAG
CATTAGTAGTGTGAGACGTA
ACCATCACAAAACCTCCGAG
GCCAATAAGAGGTGACTTGT

	c1	c2	c3
Gene1	3	1	2
Gene2	2	1	3
Gene3	8	2	1

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

DOI: 10.1002/ctm2.694

CLINICAL AND TRANSLATIONAL MEDICINE WILEY

REVIEW

Single-cell RNA sequencing technologies and applications: A brief overview

Dragomirka Jovic[1,2] | Xue Liang[1,2,3] | Hua Zeng[4] | Lin Lin[5,6] | Fengping Xu[1,2] | Yonglun Luo[1,2,5,6]

General Analyses
Low quality Cell Filtration
Normalization
HVG Selection
Dimension Reduction
Clustering
Cell Type Annotation
5000
4000
3000
2000
1000
nFeature
30000
20000
10000
nCount
10.0
7.5
5.0
2.5
0.0
MT
Standardized Variance
20
15
10
5
0
1e−02
1e+00
1e+02
Average Expression
PC_2
30
20
10
0
−10
−20
−10
0
10
PC_1
tSNE_2
20
0
−20
−20
0
20
40
tSNE_1
3
1
5
2
0
4
Kupffer Cell
T Cell
Endothelial Cell
Dendritic Cell
Hepatocyte
Epithelial Cell

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

DOI: 10.1002/ctm2.694

CLINICAL AND TRANSLATIONAL MEDICINE WILEY

REVIEW

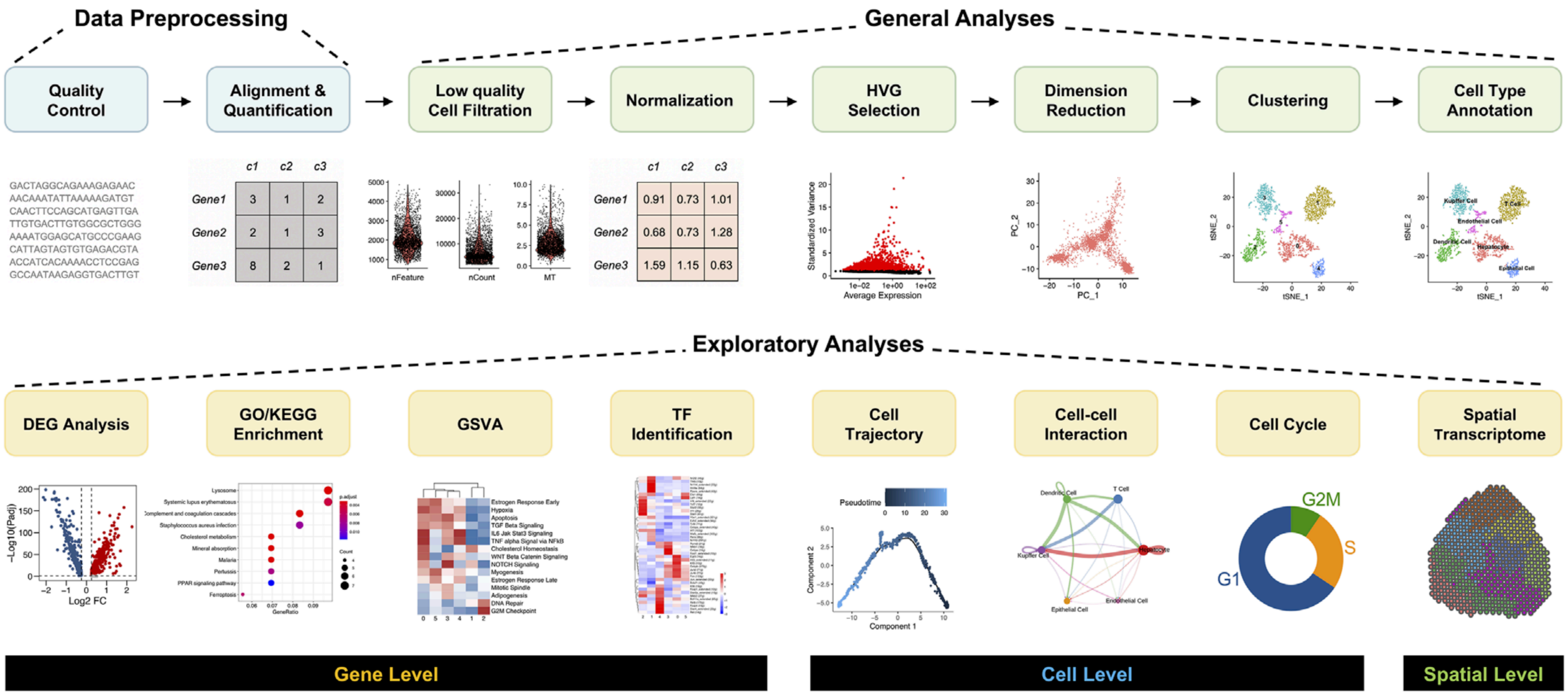
Single-cell RNA sequencing technologies and applications: A brief overview

Dragomirka Jovic[1,2] | Xue Liang[1,2,3] | Hua Zeng[4] | Lin Lin[5,6] | Fengping Xu[1,2] | Yonglun Luo[1,2,5,6]

Seurat

Scanpy

Bioconductor/OSCA



Data Preprocessing

Quality Control → Alignment & Quantification

General Analyses

→ Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

GACTAGGCAGAAAGAGAAC
AACAAATATTAAAAAGATGT
CAACTTCCAGCATGAGTTGA
TTGTGACTTGTGGCGCTGGG
AAAATGGAGCATGCCCGAAG
CATTAGTAGTGTGAGACGTA
ACCATCACAAAACCTCCGAG
GCCAATAAGAGGTGACTTGT

	c1	c2	c3
Gene1	3	1	2
Gene2	2	1	3
Gene3	8	2	1

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Exploratory Analyses

DEG Analysis | GO/KEGG Enrichment | GSVA | TF Identification | Cell Trajectory | Cell-cell Interaction | Cell Cycle | Spatial Transcriptome

Gene Level | Cell Level | Spatial Level

DOI: 10.1002/ctm2.694

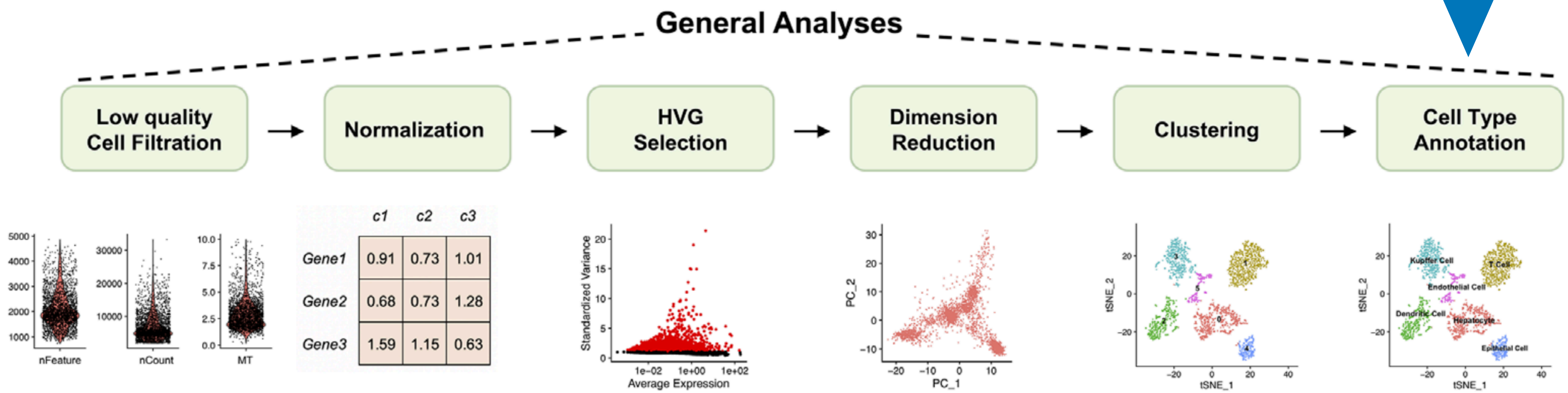
REVIEW

CLINICAL AND TRANSLATIONAL MEDICINE WILEY

Single-cell RNA sequencing technologies and applications: A brief overview

Dragomirka Jovic[1,2] | Xue Liang[1,2,3] | Hua Zeng[4] | Lin Lin[5,6] | Fengping Xu[1,2] | Yonglun Luo[1,2,5,6]

Analysis Overview:



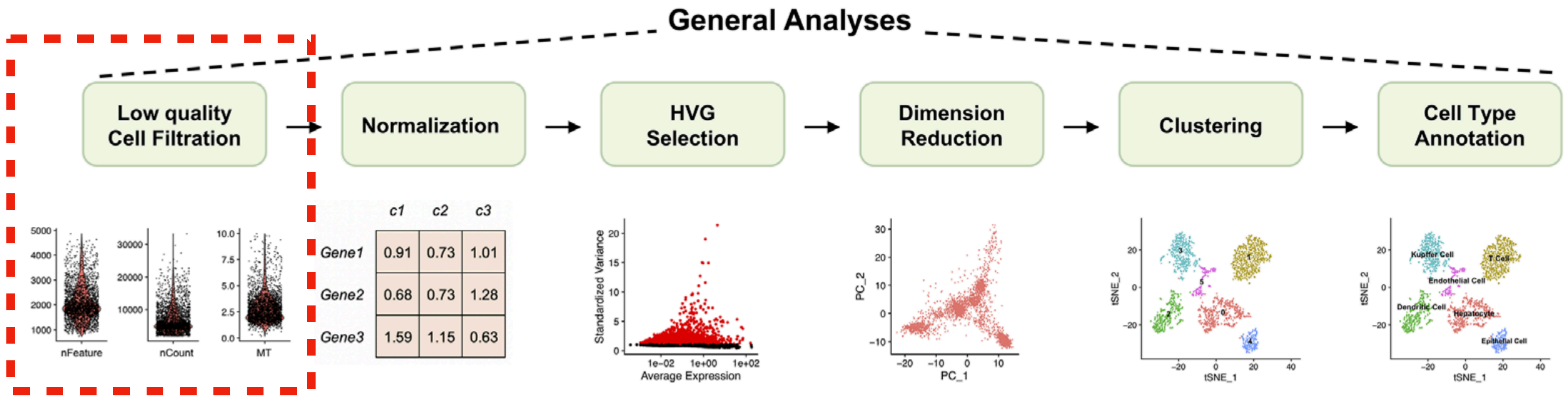
General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Doublet Filtering
Ambient RNA Removal
CellCycle Regression
Normalization Method
Integration

Low Quality Cell Filtration:



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

- Filter Data to only include cells of high quality

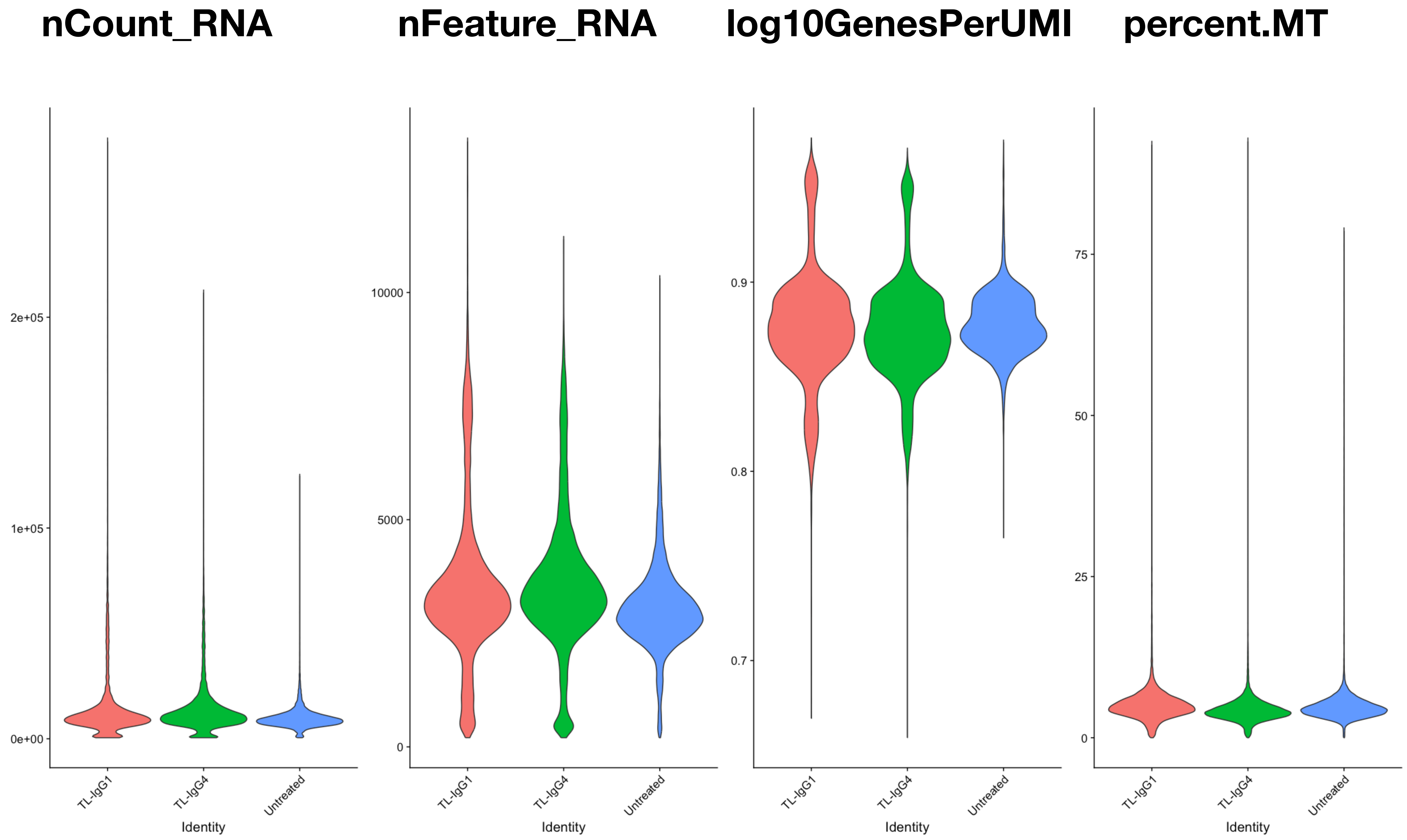
Challenges:

- Delineating cells from poor quality from less complex cells
- Choosing appropriate thresholds

Recommendations:

- Have a good idea of your expectations:
  - Do we expect low complexity cells? Same cell types? PBMCS?
  - Do we expect cells to have high MT reads?

Low Quality Cell Filtration:



nCount_RNA

nFeature_RNA

log10GenesPerUMI

percent.MT

nCount_RNA:

Number of UMI's detected Per Cell

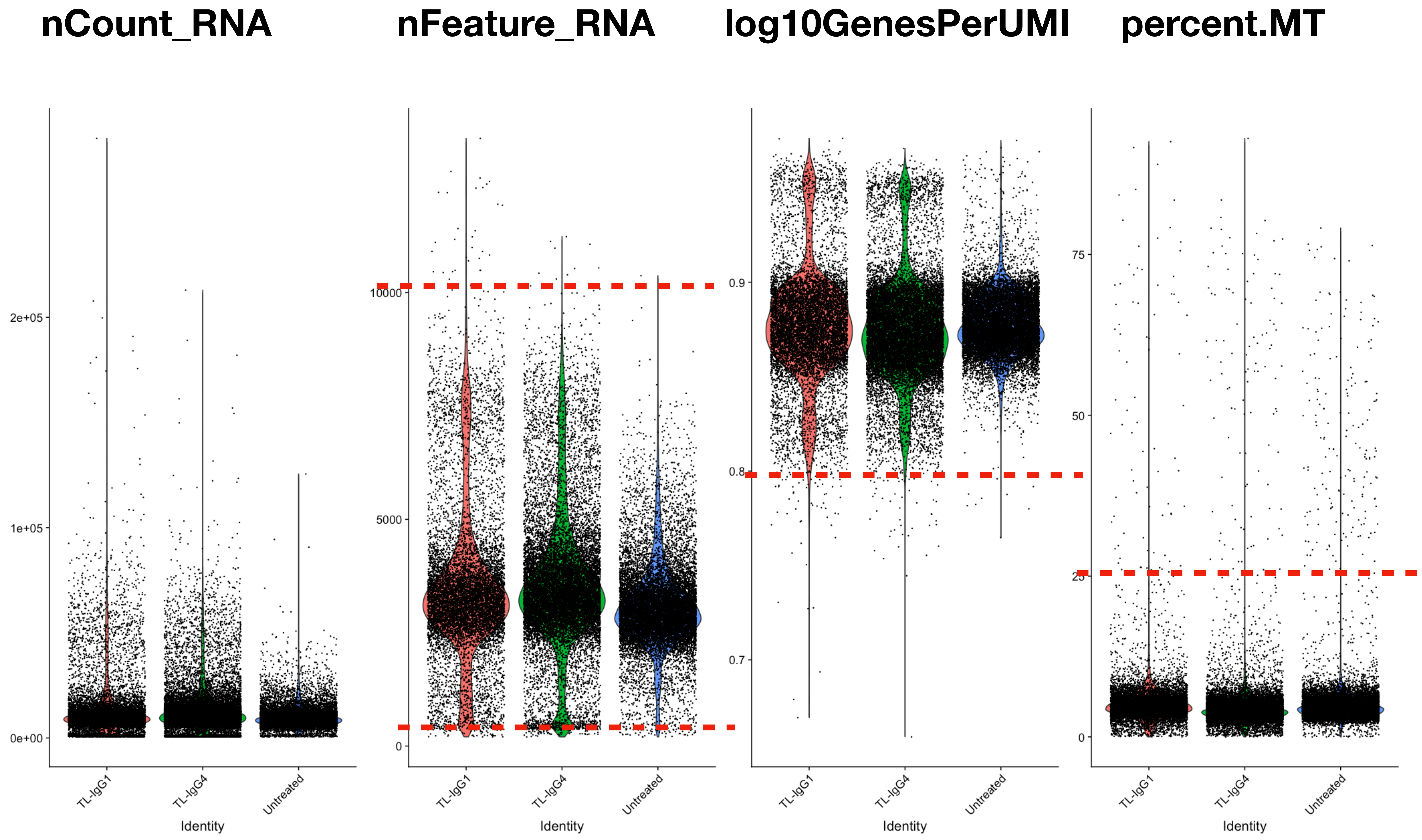
nFeature_RNA:

Number of features/genes detected per cell

percent.MT:

Proportion of Mitochondiral Reads Per cell

Low Quality Cell Filtration:



nCount_RNA

nFeature_RNA

log10GenesPerUMI

percent.MT

nCount_RNA:

Number of UMI's detected Per Cell

nFeature_RNA:

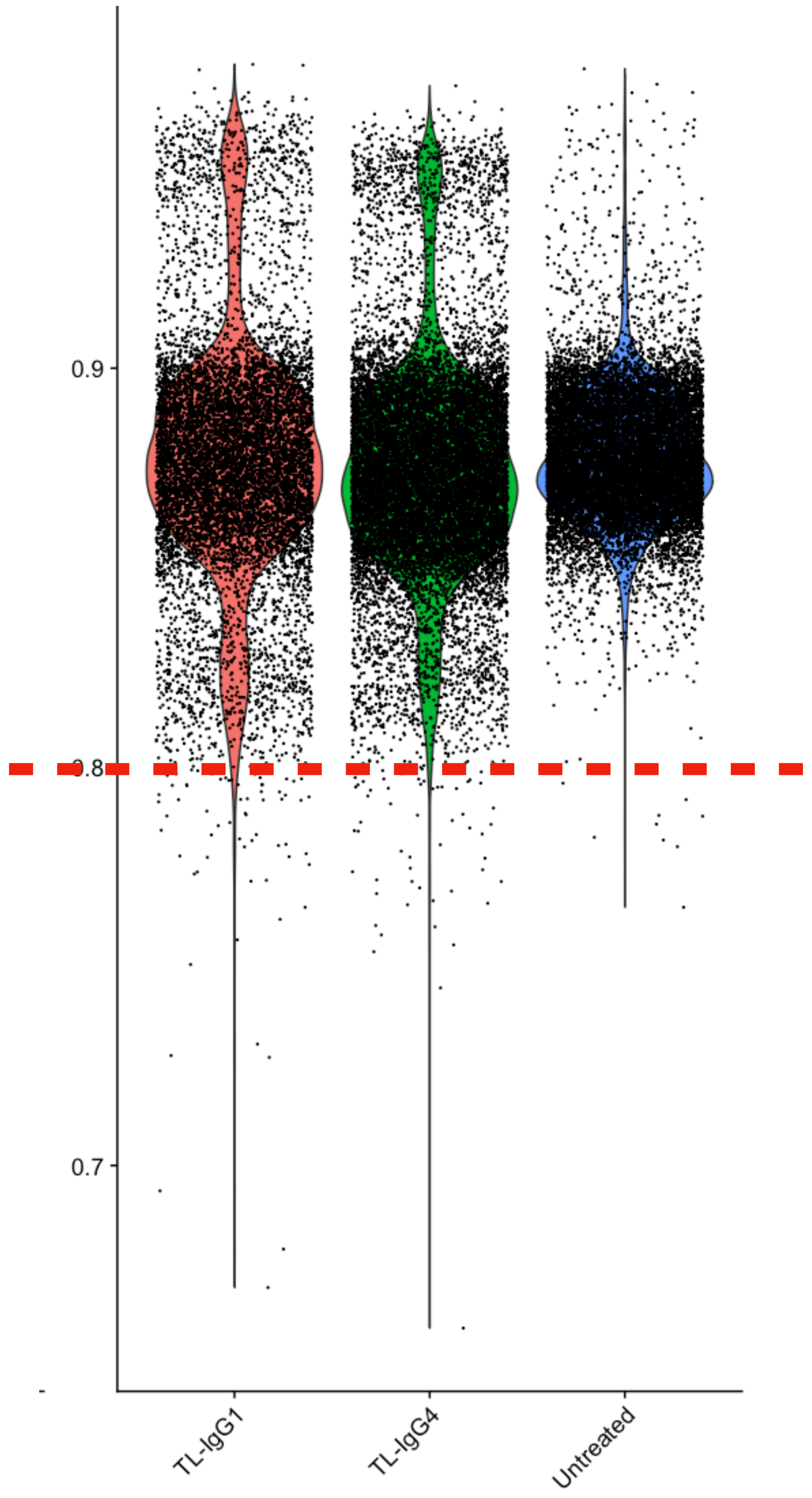
Number of features/genes detected per cell

percent.MT:

Proportion of Mitochondiral Reads Per cell

Low Quality Cell Filtration:

log10GenesPerUMI

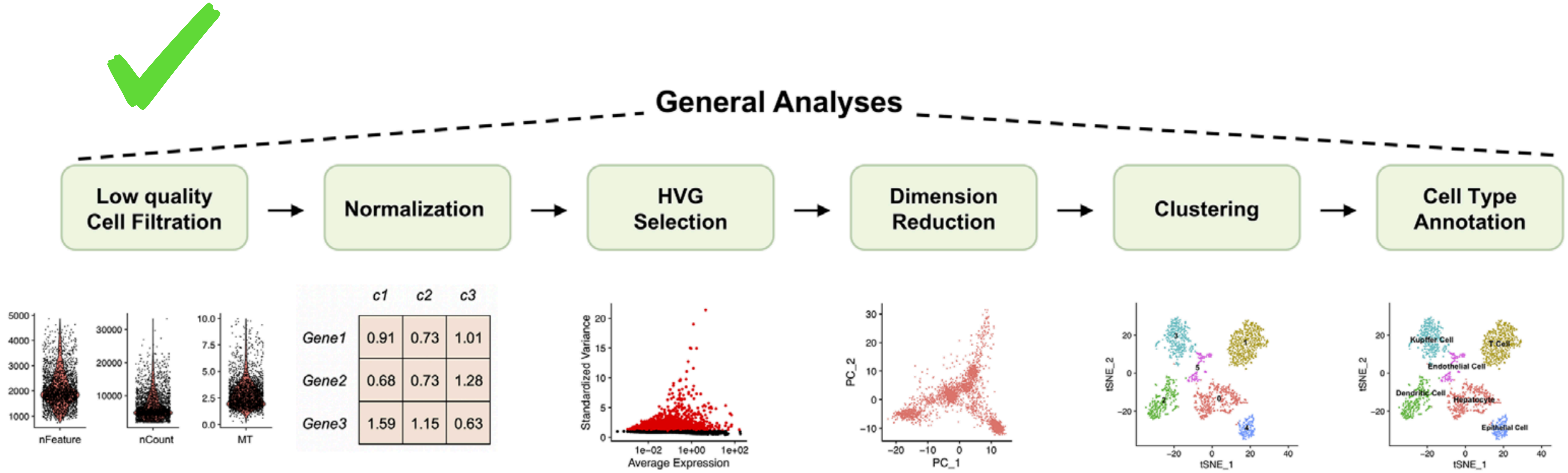


log10GenesPerUMI

- **Also referred to as Novelty Score**
- **Provides insights for RNA complexity Per Cell**

log10(nFeature_RNA) / log10(nCount_RNA)

Analysis Overview:

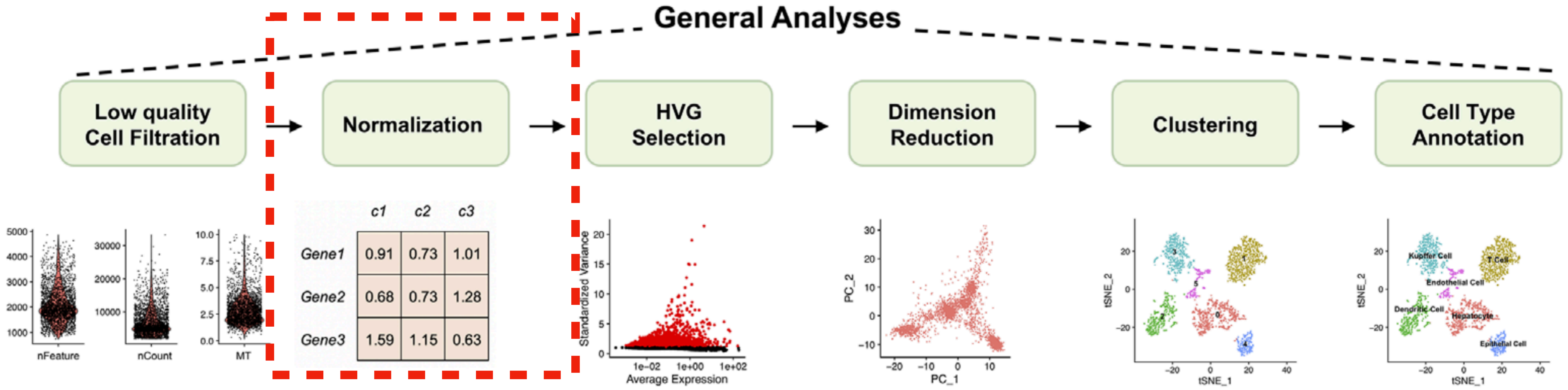


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Normalization



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

- Normalize UMI counts to account for differences in sequencing depth and over-dispersed count values

Challenges:

- Removing unwanted variation so we do not drive downstream clustering by artifacts

Recommendations:

- Regress out number of UMIs, mitochondrial contamination, cell cycle (if needed and appropriate)

Normalization

Various methods have been developed specifically for scRNA-seq normalization

In Seurat we can either use *LogNormalize* method or *SCTransform* method

In general, normalization is a two Step Process

- **Scaling**
- **Simple Transformation *OR* Complex Transformation**

Normalization

LogNormalize:

Scaling —> (Divide Counts for each *Gene* / Total Counts in a Given *Cell*) * scale.factor (default: 10,000)

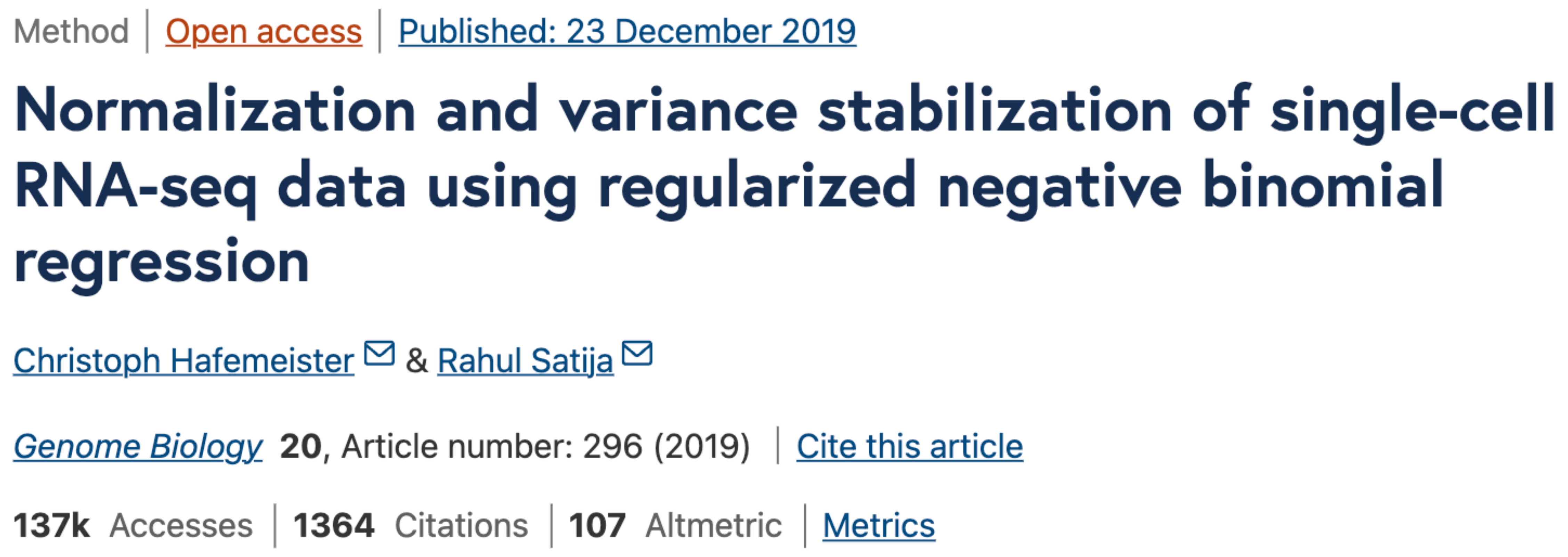
Transformation —> Log Transformation (same for each gene, hence simple transformation)

SCTransform:

Scaling —> Multiplies each measurement by a gene-specific weight

Transformation —> Pearson Residuals from regularized negative binomial regression

More evidence == more weight; Genes that are expressed in only a small fraction of cells will be favored (useful for finding rare cell populations)



Method | Open access | Published: 23 December 2019

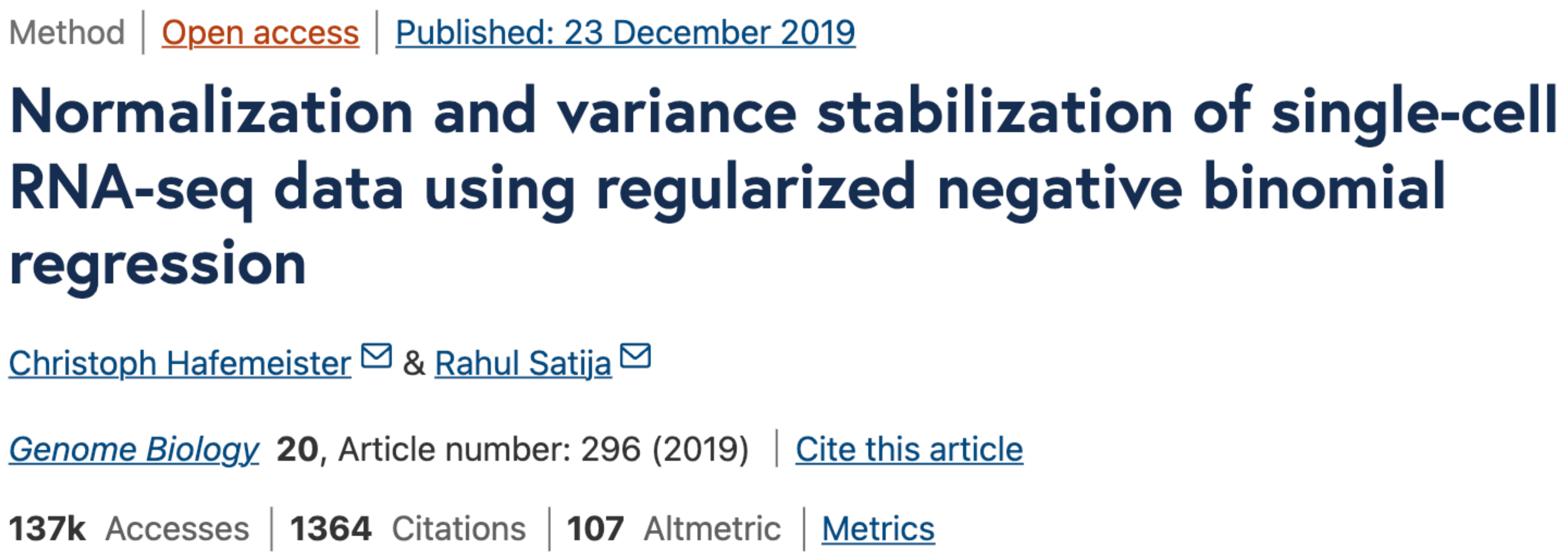
Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression

Christoph Hafemeister & Rahul Satija

Genome Biology **20**, Article number: 296 (2019) | Cite this article

137k Accesses | **1364** Citations | **107** Altmetric | Metrics

Normalization



Method | Open access | Published: 23 December 2019

Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression

Christoph Hafemeister & Rahul Satija

Genome Biology **20**, Article number: 296 (2019) | Cite this article

137k Accesses | **1364** Citations | **107** Altmetric | Metrics

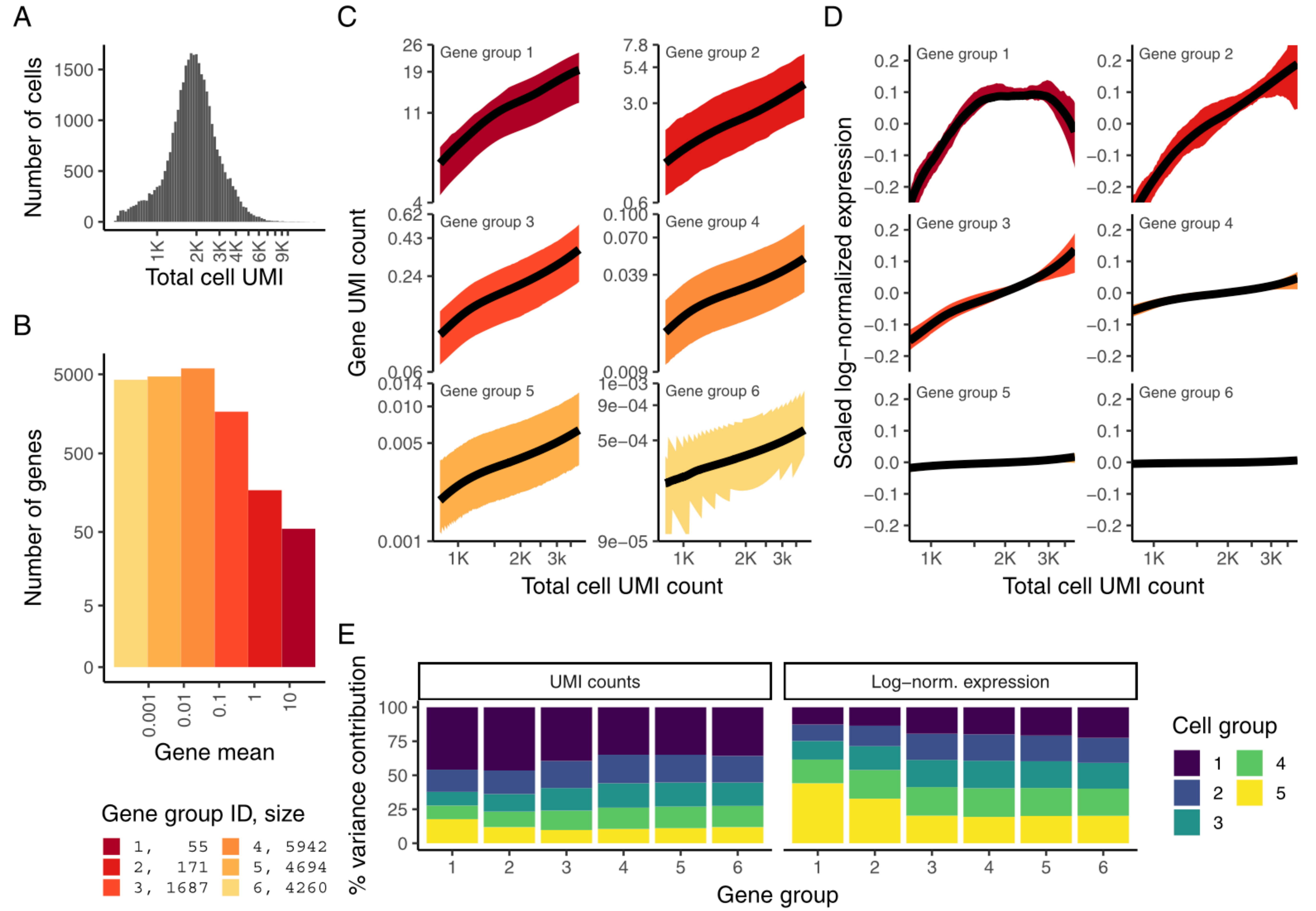


Fig. 1 33,148 PBMC dataset from 10X Genomics. **a** Distribution of total UMI counts / cell ("sequencing depth"). **b** We placed genes into six groups, based on their average expression in the dataset. **c** For each gene group, we examined the average relationship between observed counts and cell sequencing depth. We fit a smooth line for each gene individually and combined results based on the groupings in **b**. Black line shows mean, colored region indicates interquartile range. **d** Same as in **c**, but showing scaled log-normalized values instead of UMI counts. Values were scaled

Normalization

Method | Open access | Published: 23 December 2019

Normalization and variance stak
RNA-seq data using regularized
regression

Christoph Hafemeister & Rahul Satija

Genome Biology 20, Article number: 296 (2019) | Cite this

137k Accesses | 1364 Citations | 107 Altmetric | Metrics

Takeaway:

Cells with low total UMI counts show disproportionately high variances' — post LogNormalization

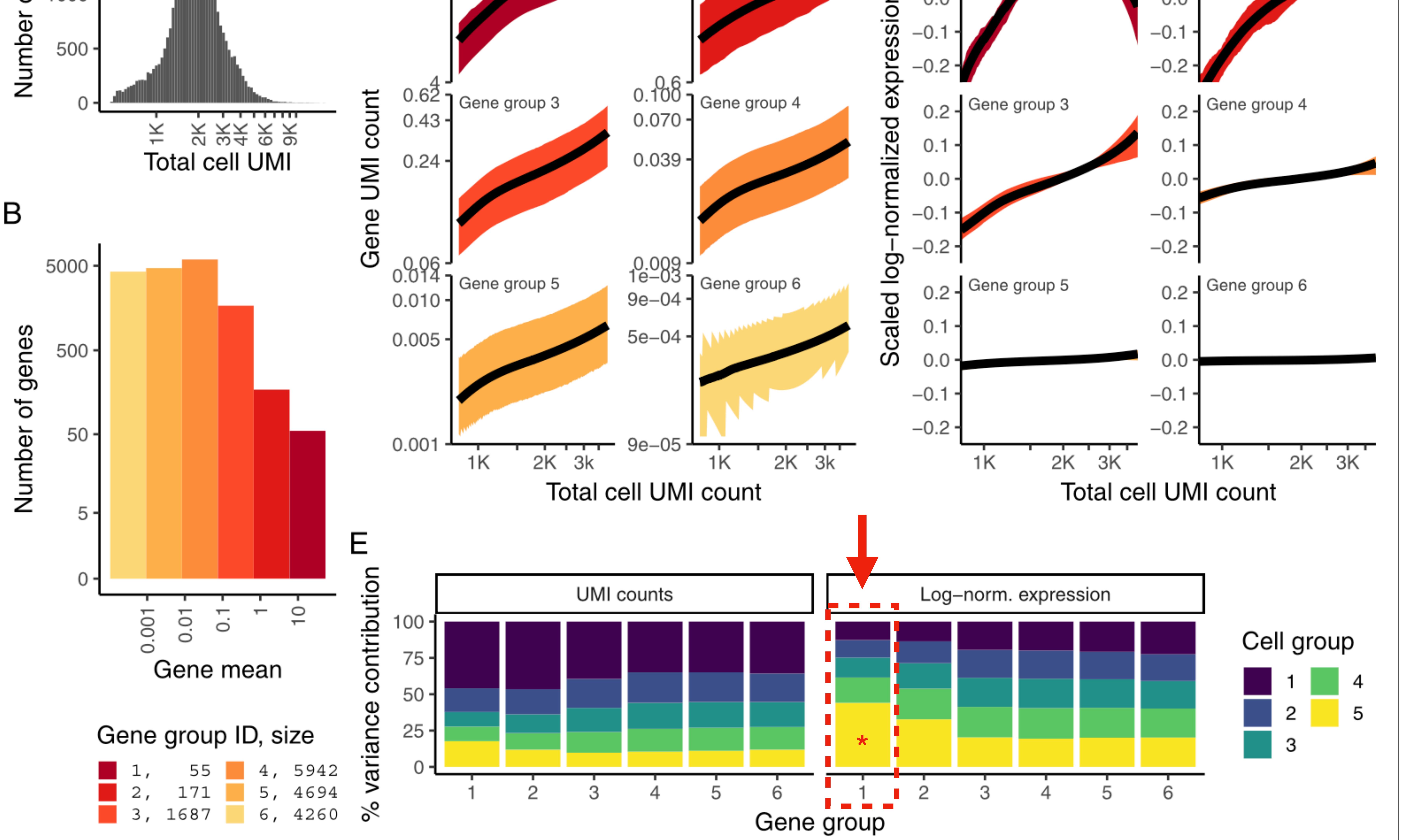
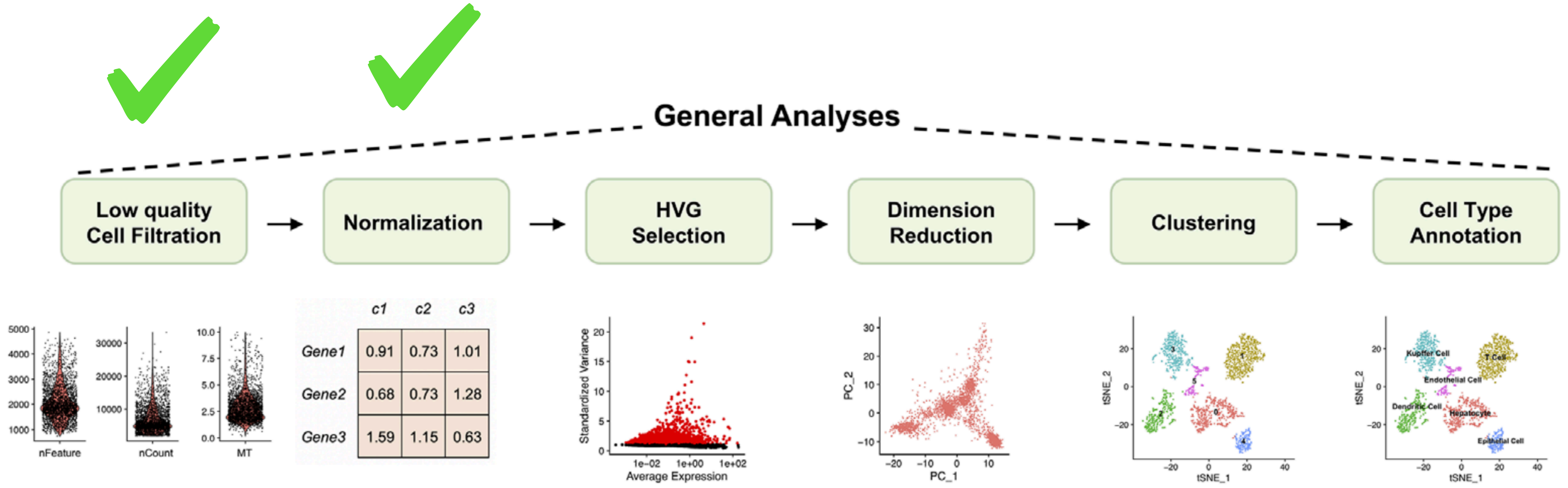


Fig. 1 33,148 PBMC dataset from 10X Genomics. **a** Distribution of total UMI counts / cell ("sequencing depth"). **b** We placed genes into six groups, based on their average expression in the dataset. **c** For each gene group, we examined the average relationship between observed counts and cell sequencing depth. We fit a smooth line for each gene individually and combined results based on the groupings in **b**. Black line shows mean, colored region indicates interquartile range. **d** Same as in **c**, but showing scaled log-normalized values instead of UMI counts. Values were scaled

Analysis Overview:

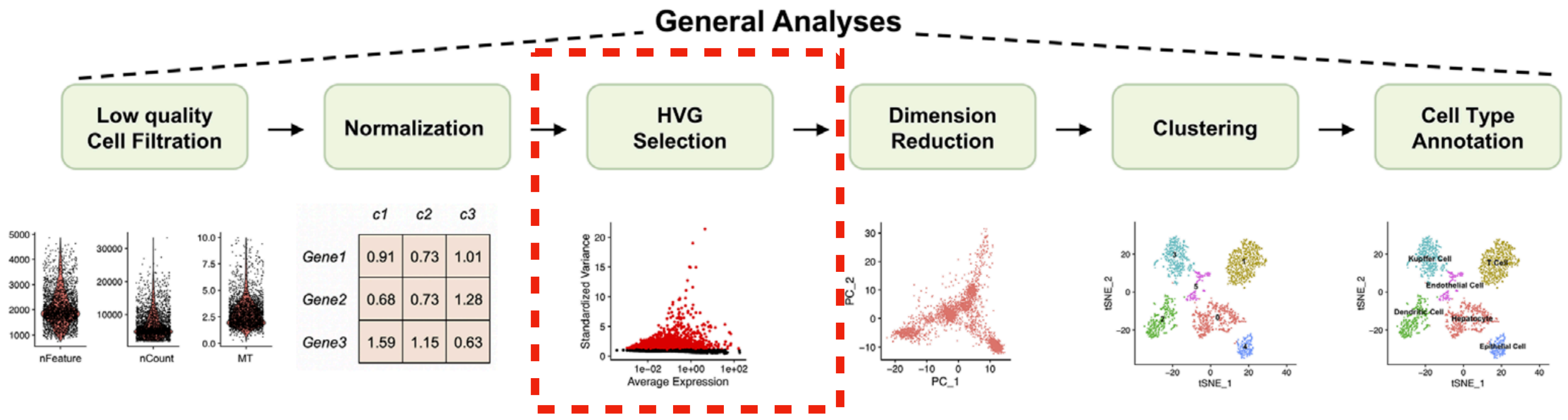


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	*c2*	*c3*
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

High Variable Gene Selection



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

These genes will be used for Clustering

- Find Most Interesting Features in an Unsupervised Manner
- Optimize Signal:Noise ratio

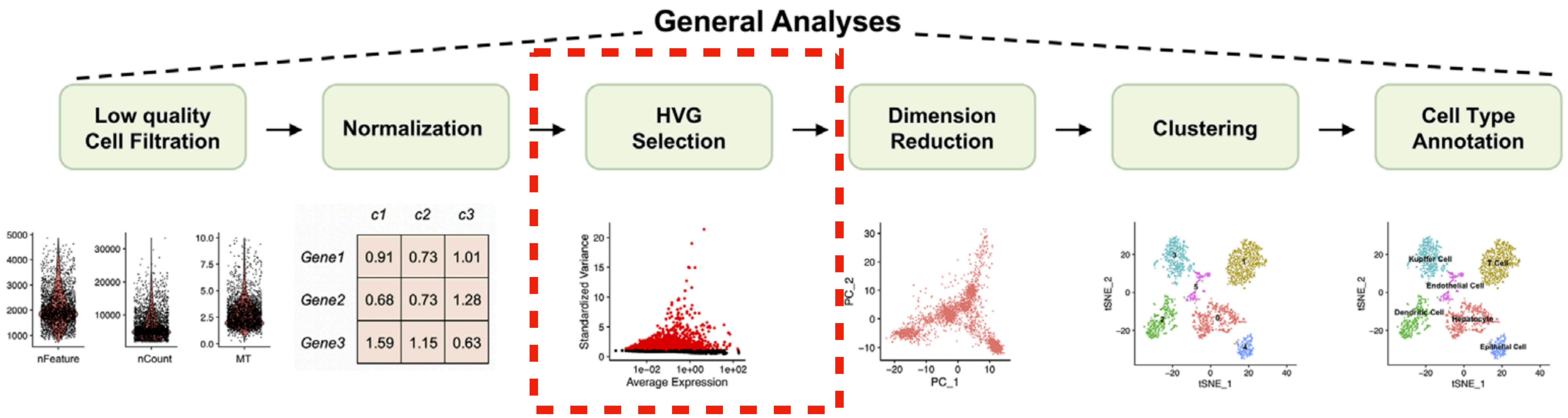
Challenges:

- The high sparsity and zero-inflation of the scRNA-seq data makes it difficult to distinguish true variability from technical noise or dropout events
- The heterogeneity and complexity of the cell populations

Recommendations:

- Validate HVG lists …

HVG Selection *



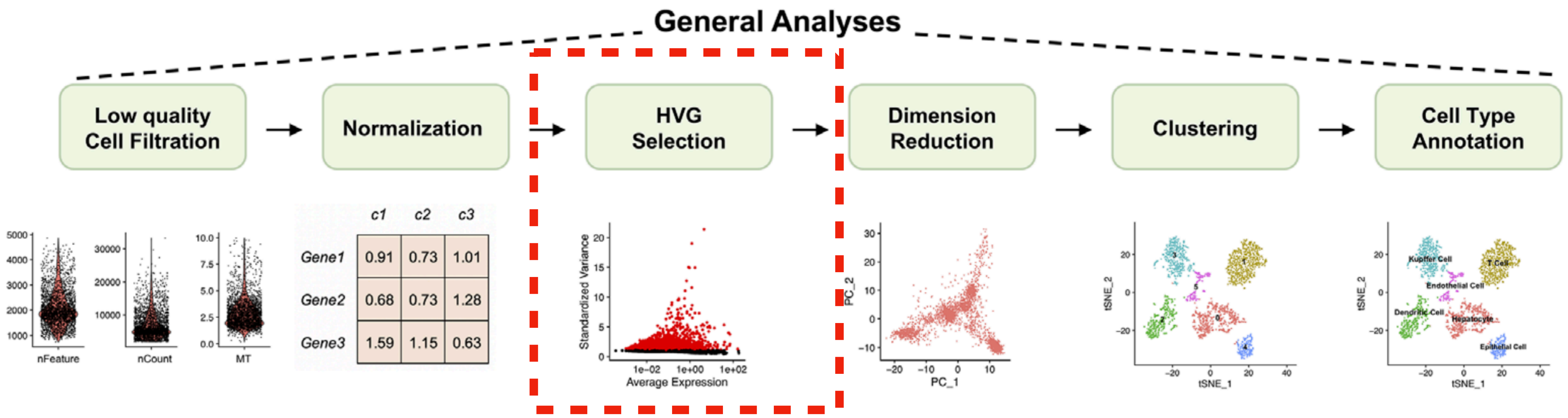
	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

In Seurat, there are a few ways to find High Variable Genes

1. **VST (Variance Stabilized Transformation) Method (Default)**
2. **MVP (Mean Variance Plot) Method**
3. **Dispersion Method**

* https://satijalab.org/seurat/reference/findvariablefeatures

HVG Selection *



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

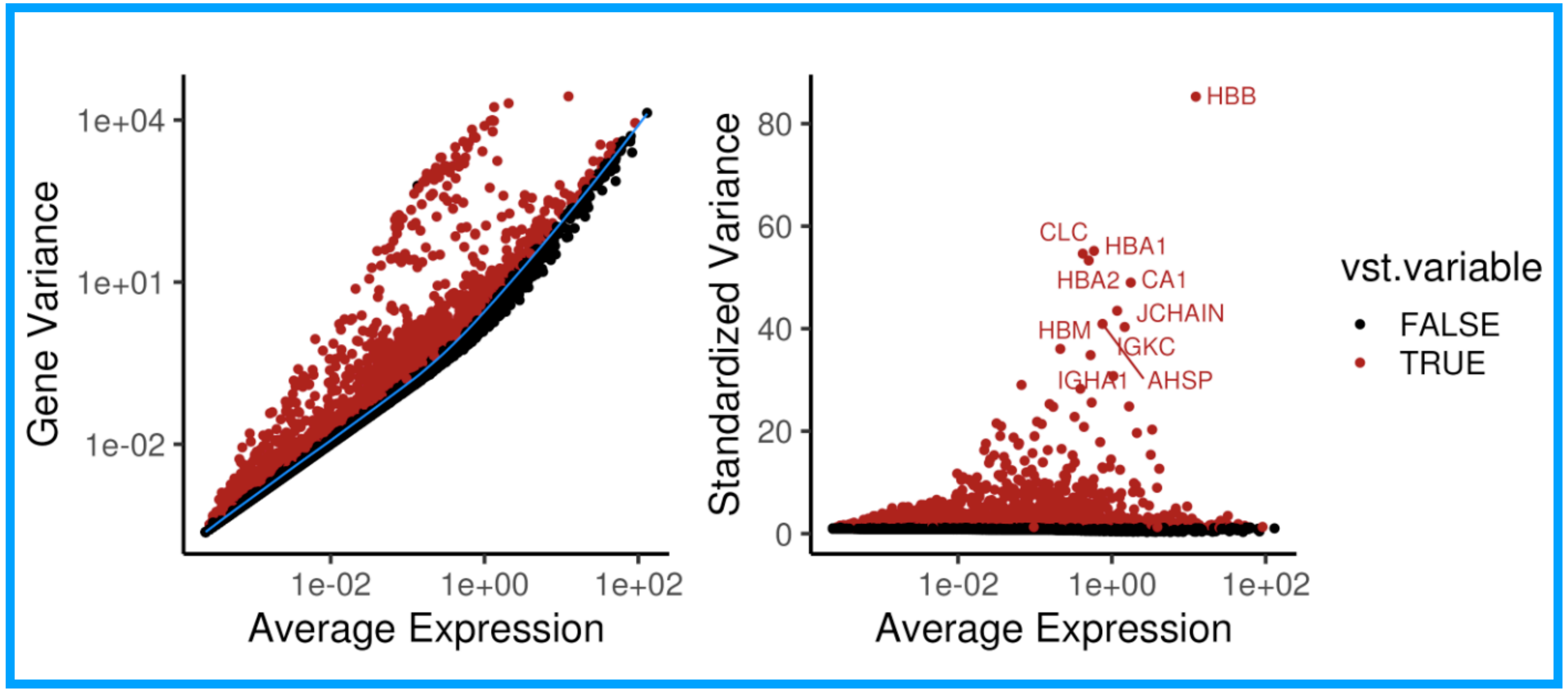
VST Method (Default):

First, it fits a line to the relationship of log(variance) and log(mean) using local polynomial regression (loess). This line represents the expected variance for a given mean expression level.

Then, it standardizes the feature values using the observed mean and expected variance (given by the fitted line). This removes the effect of the mean expression on the variance, and makes the features more comparable.

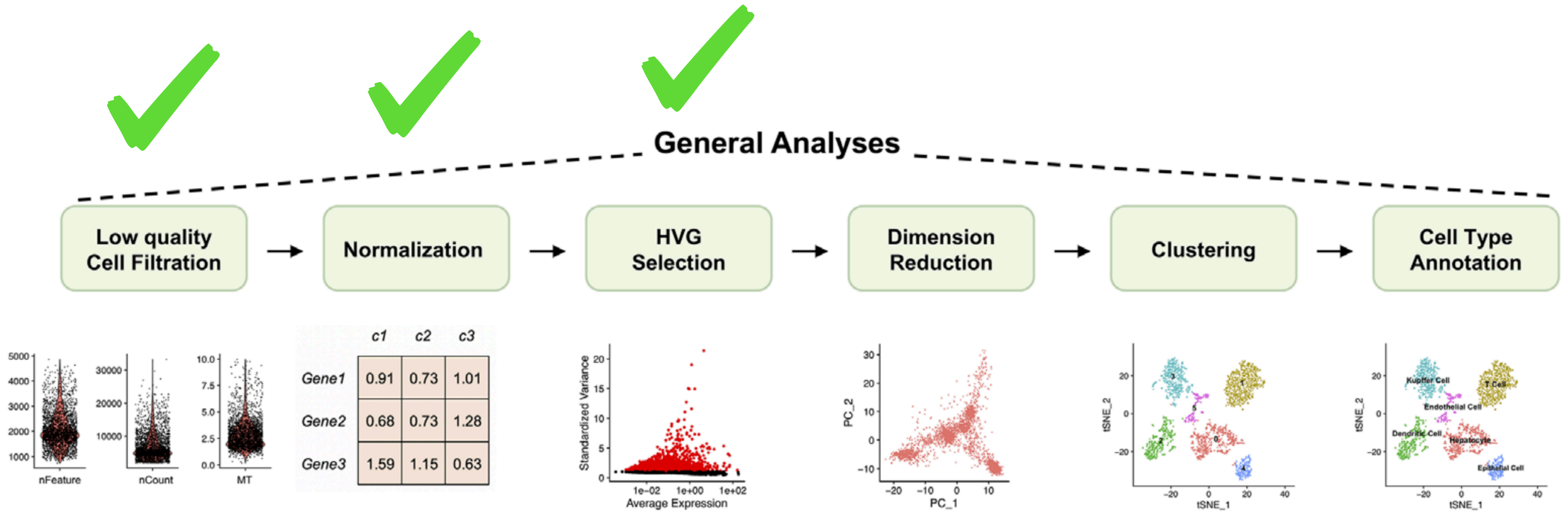
Next, it calculates the variance of the standardized values, after clipping them to a maximum value. This value is set to the square root of the number of cells by default, but can be changed by the user. Clipping helps to reduce the influence of outliers or extreme values on the variance calculation.

Finally, it selects the features with the highest variance among the standardized values, as these are the most variable features across cells.



These genes essentially drive the Clustering Analysis

Analysis Overview:

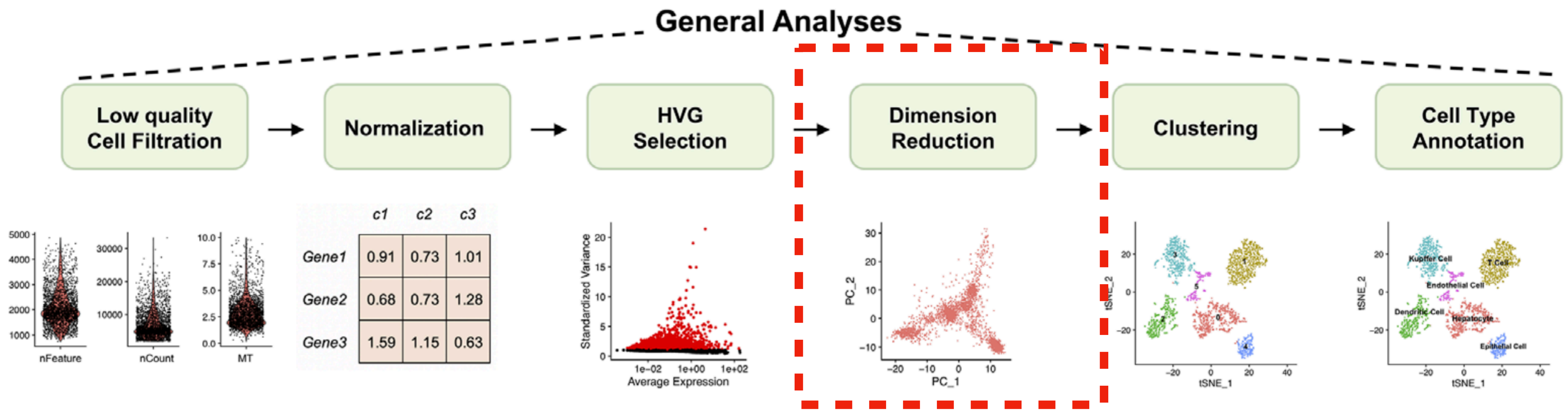


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Dimension Reduction (PCA)



	c1	*c2*	*c3*
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

- Use HVG’s to perform dimensionality reduction

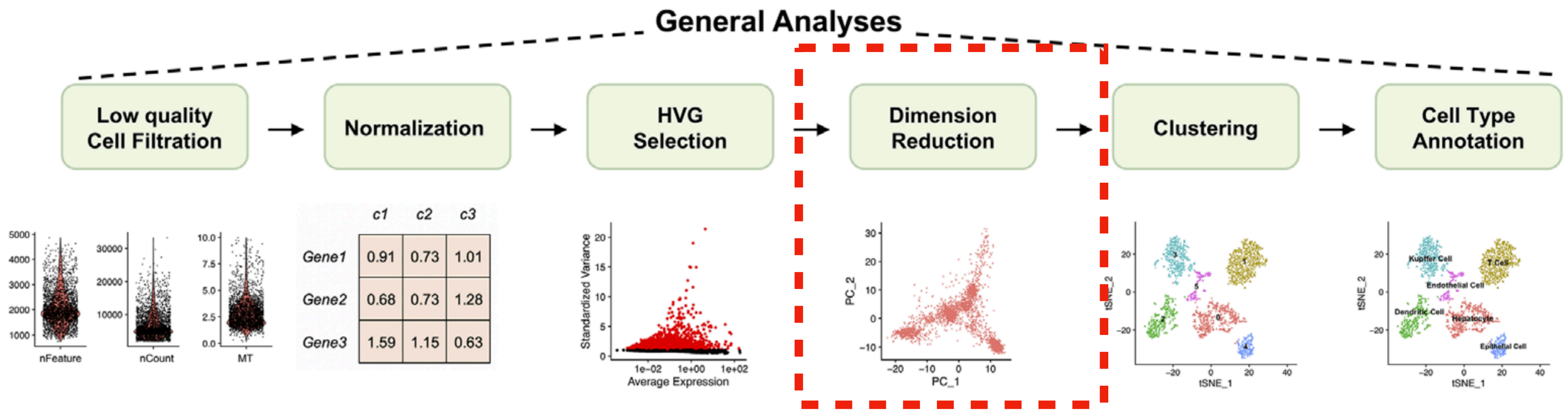
Challenges:

- Can be affected by batch effects + other unwanted sources of variation
- Separating technical variation from true biological variation

Recommendations:

- Batch Correction / Integration

Dimension Reduction (PCA)

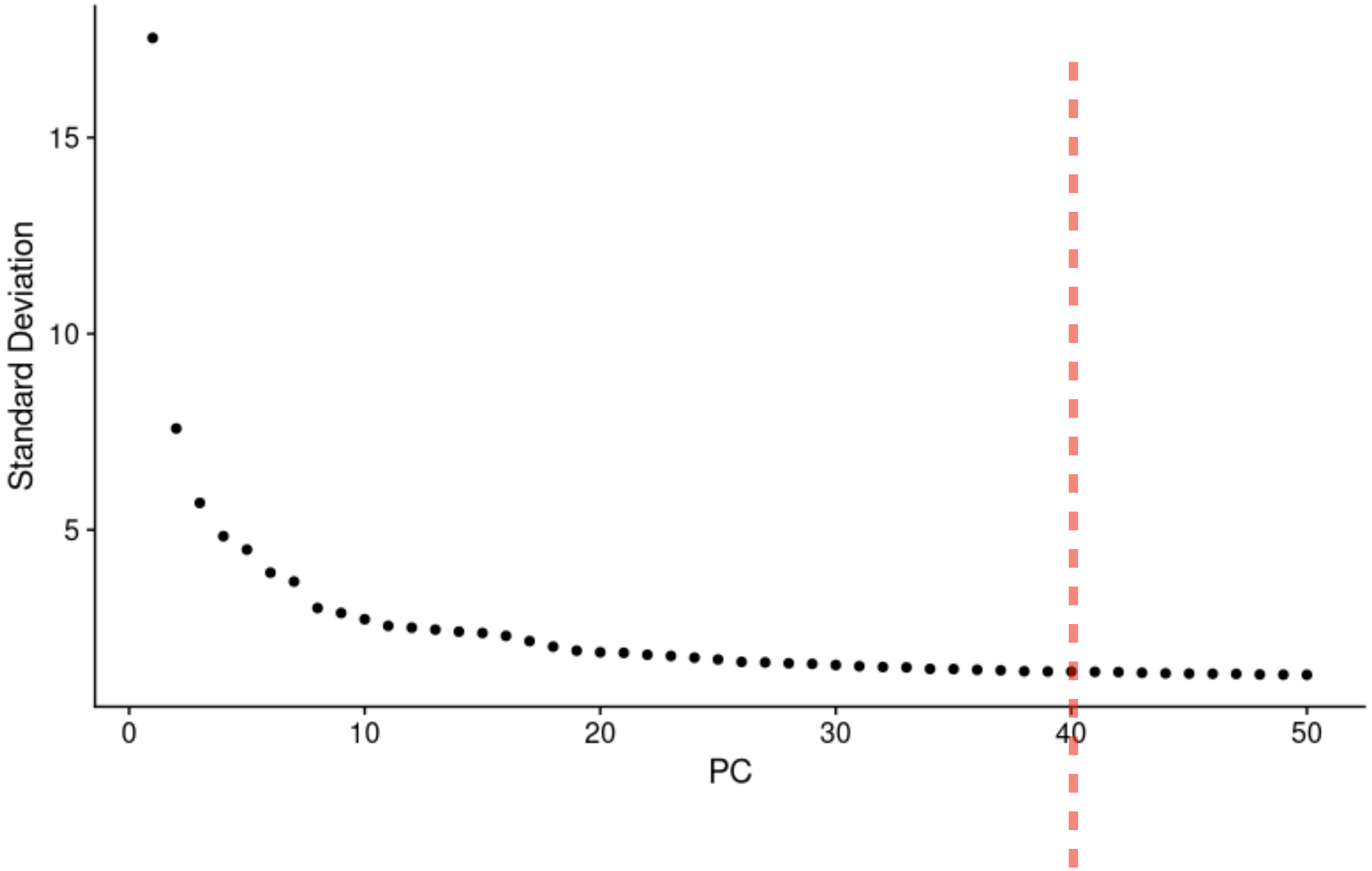


General Analyses

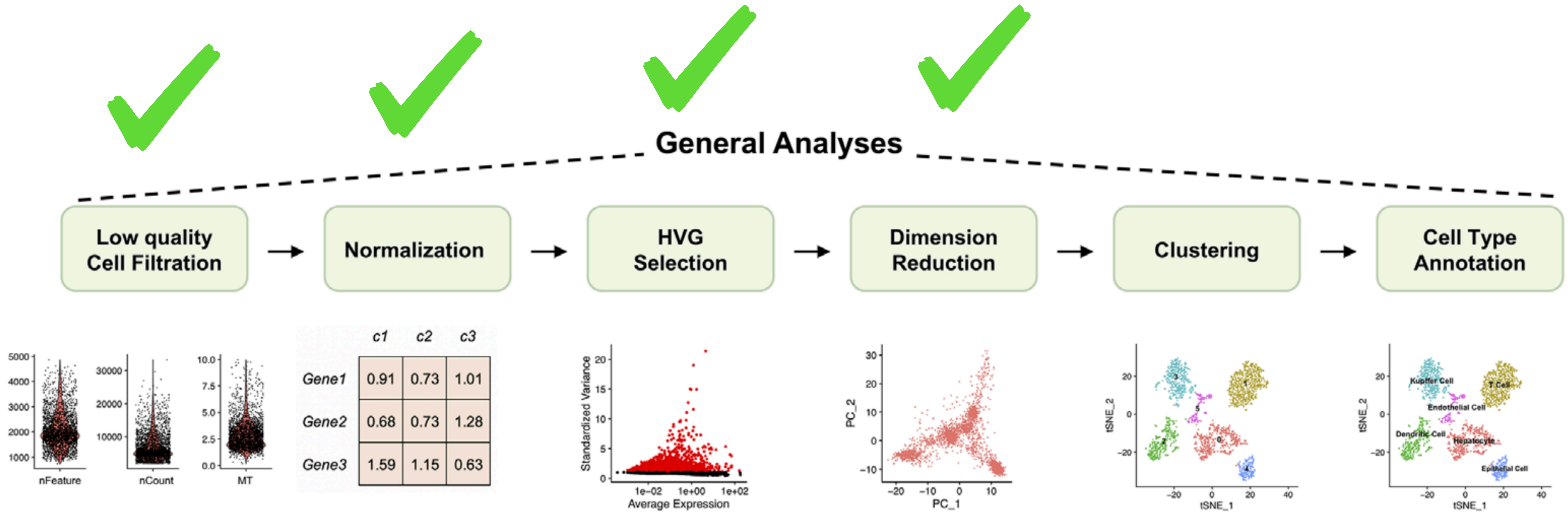
Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

ElbowPlot



Analysis Overview:

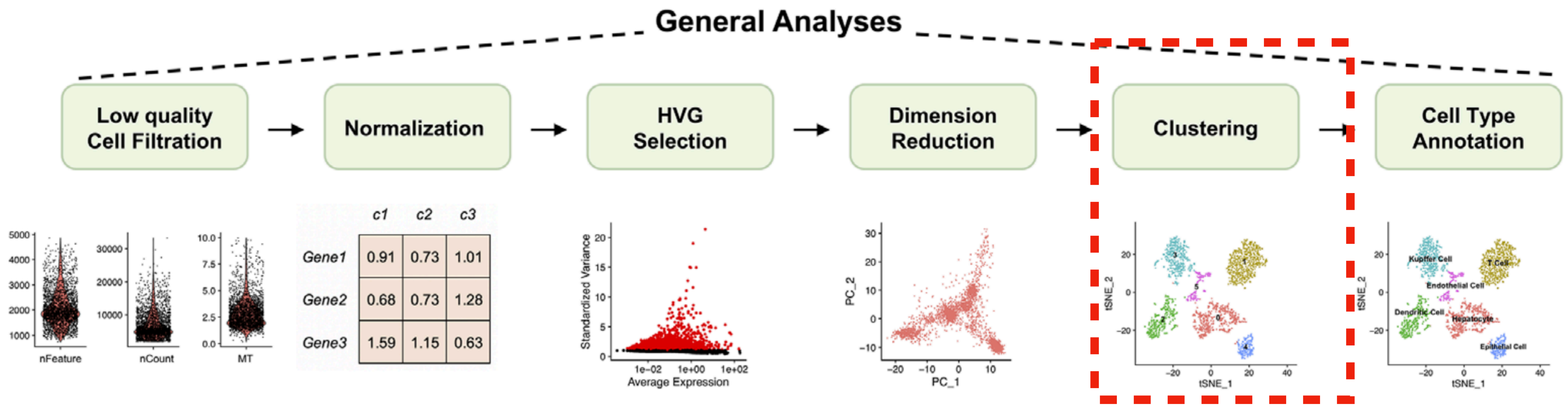


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Clustering



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

- **Generate cell type-specific clusters**
- **Determine whether clusters represent true cell types or cluster due to biological or technical variation, such as clusters of cells in the S phase of the cell cycle, clusters of specific batches, or cells with high mitochondrial content.**

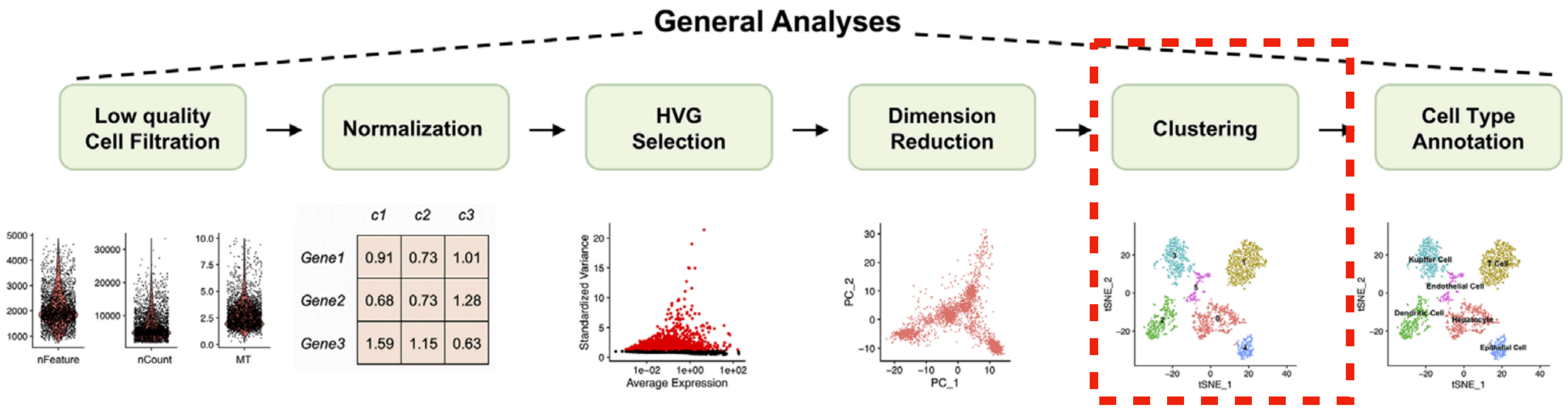
Challenges:

- **Identifying poor quality clusters that may be due to uninteresting biological or technical variation**
- **Iterative Process, revise QC thresholds**

Recommendations:

- **Expectations?**
- **Try different resolutions**

Clustering



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

There are three main approaches:

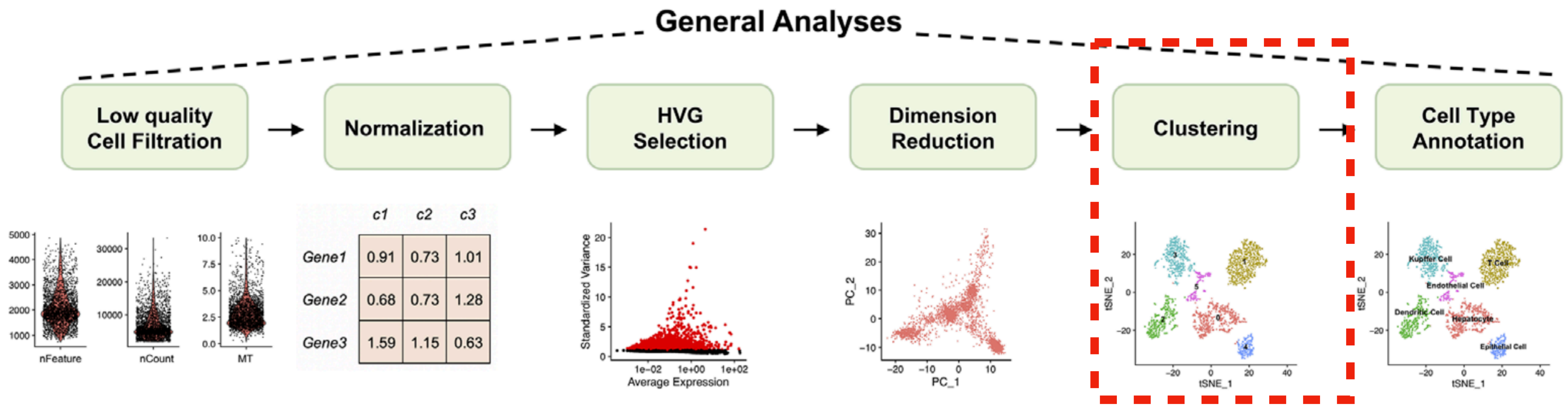
1. **Hierarchical Clustering:** These methods build a tree-like structure of clusters, where each node represents a cluster and the distance between nodes reflects the similarity between clusters

2. **K-means Clustering:** These methods partition the cells into a predefined number of clusters, such that the within-cluster variation is minimized and the between-cluster variation is maximized

3. **Graph-Based Clustering:** These methods construct a graph where each node represents a cell and each edge represents the similarity or distance between two cells. Then, they apply graph partitioning algorithms to find clusters of densely connected nodes.

https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/clustering-and-cell-annotation.html

Clustering

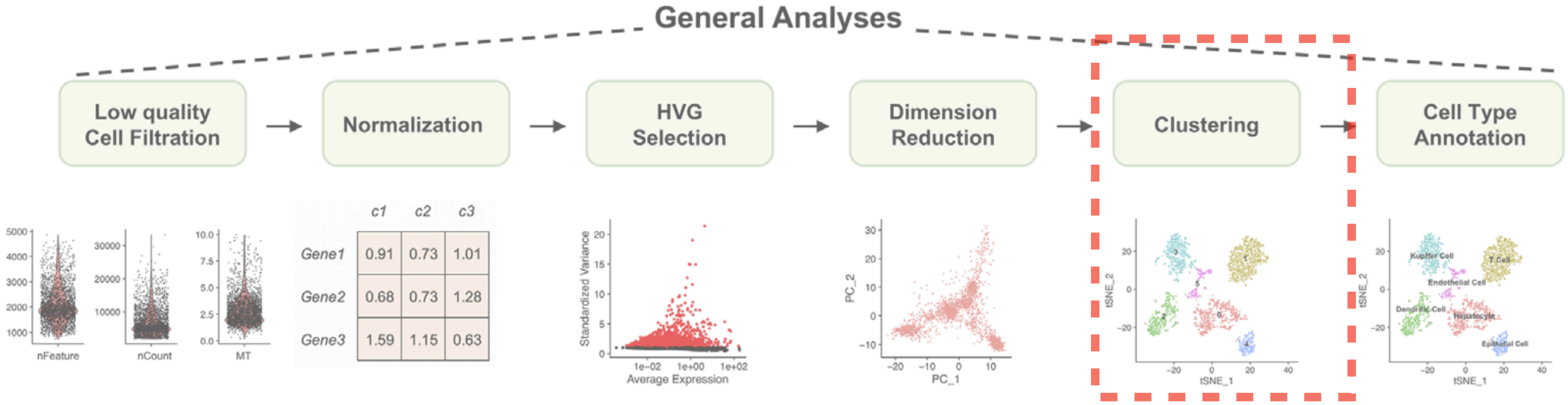


Seurat uses Graph-Based Clustering:

The default clustering algorithm in Seurat is the *Louvain* algorithm which is a fast and scalable method for finding communities in large networks.

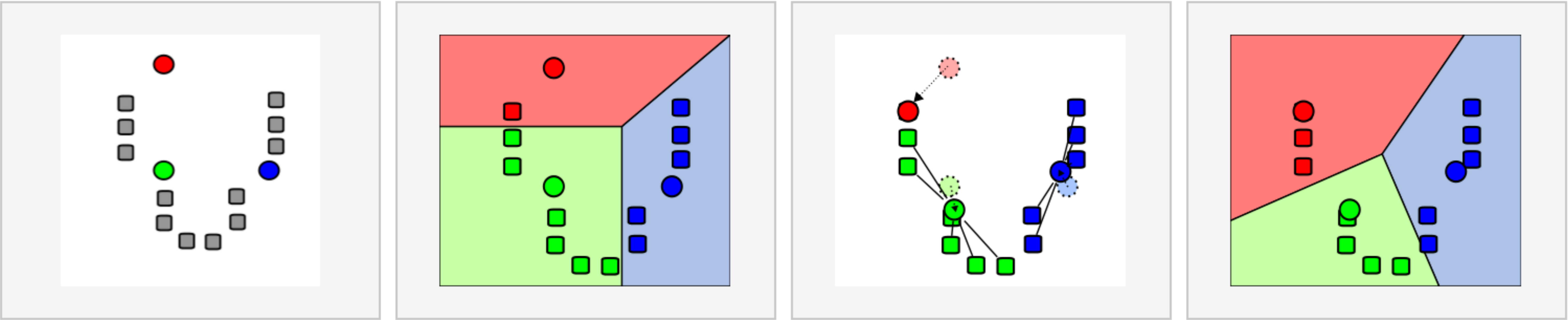
https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/clustering-and-cell-annotation.html

Clustering



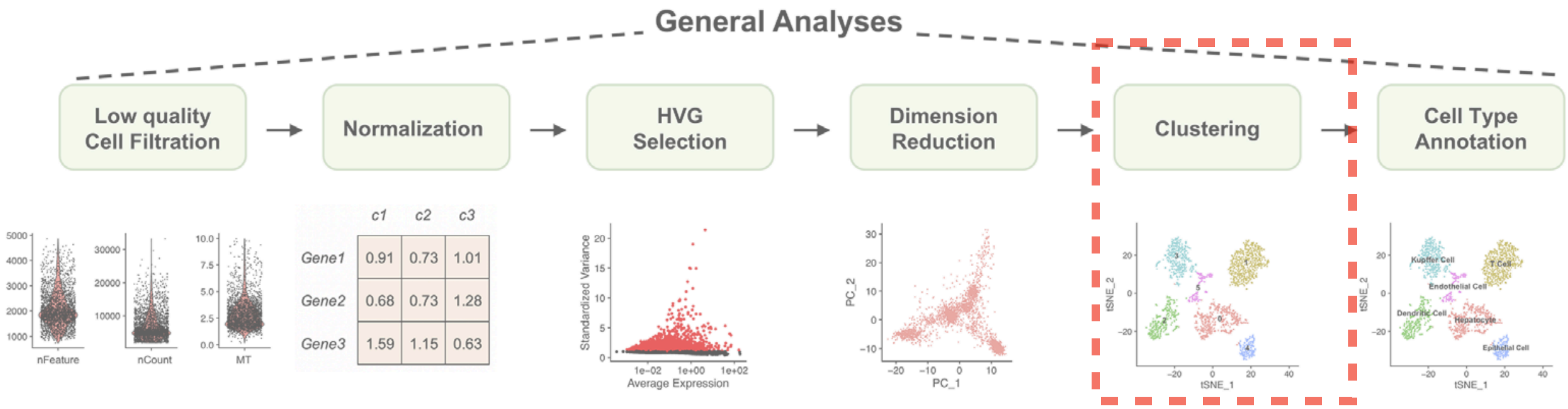
First Step is to construct a KNN graph **(Uses Euclidean Distances in PCA space)**

Second Step is to apply the Louvain Algorithm to find communities



https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/clustering-and-cell-annotation.html

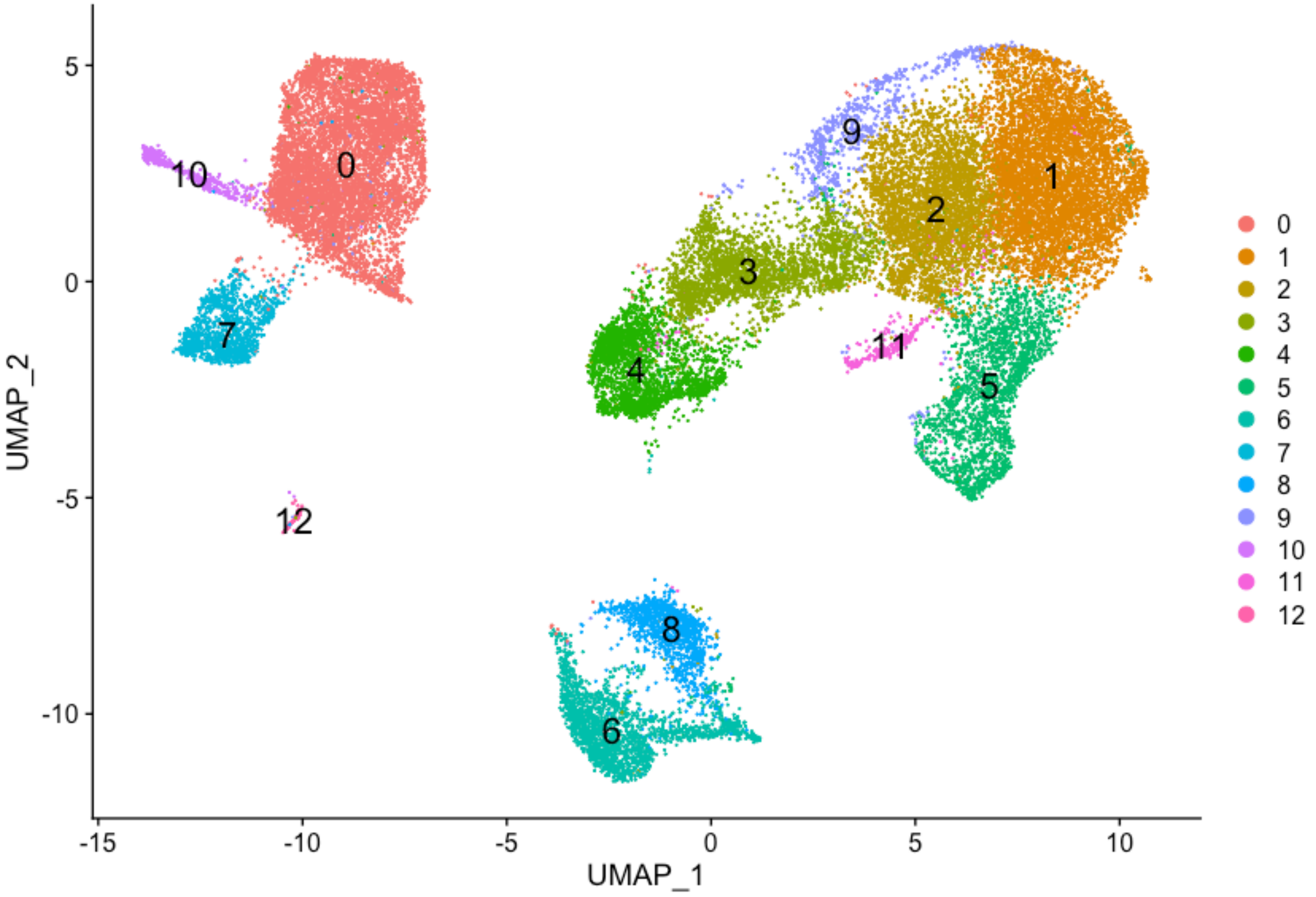
Clustering


General Analyses

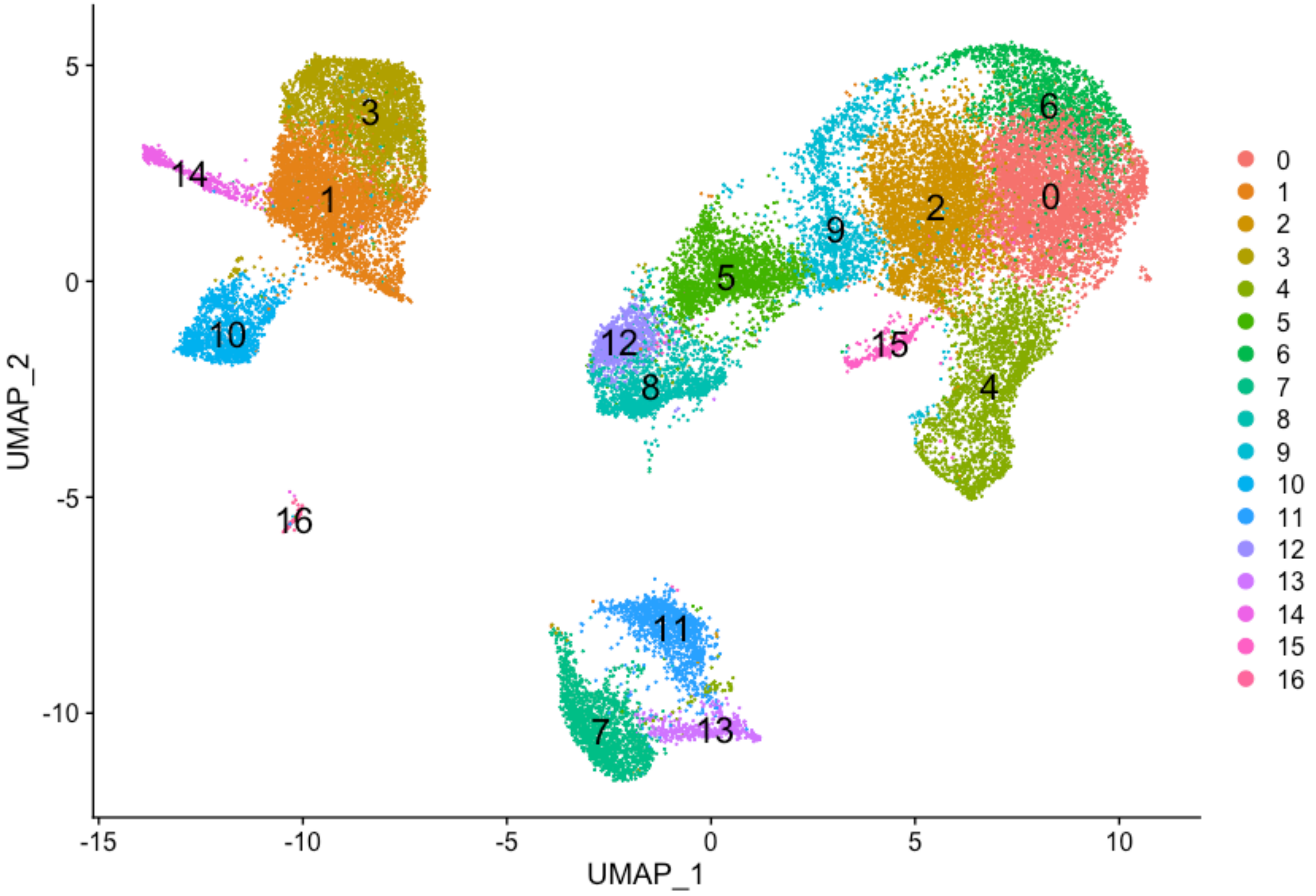
Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

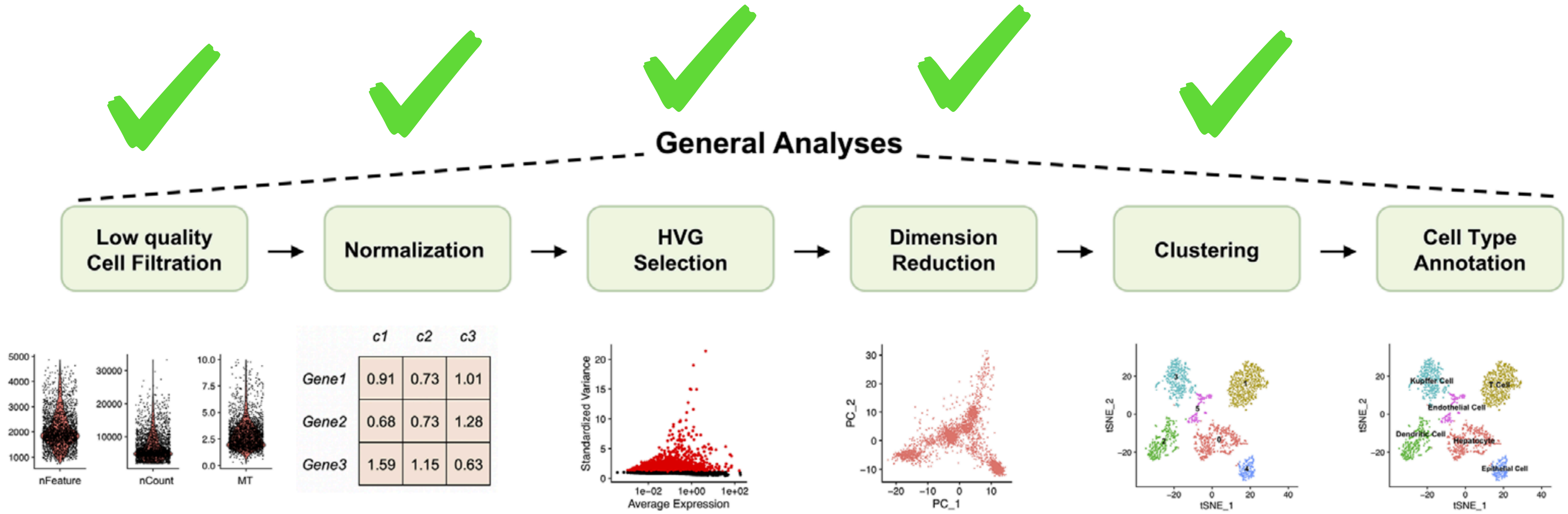
Resolution = 0.4



Resolution = 0.8



Analysis Overview:

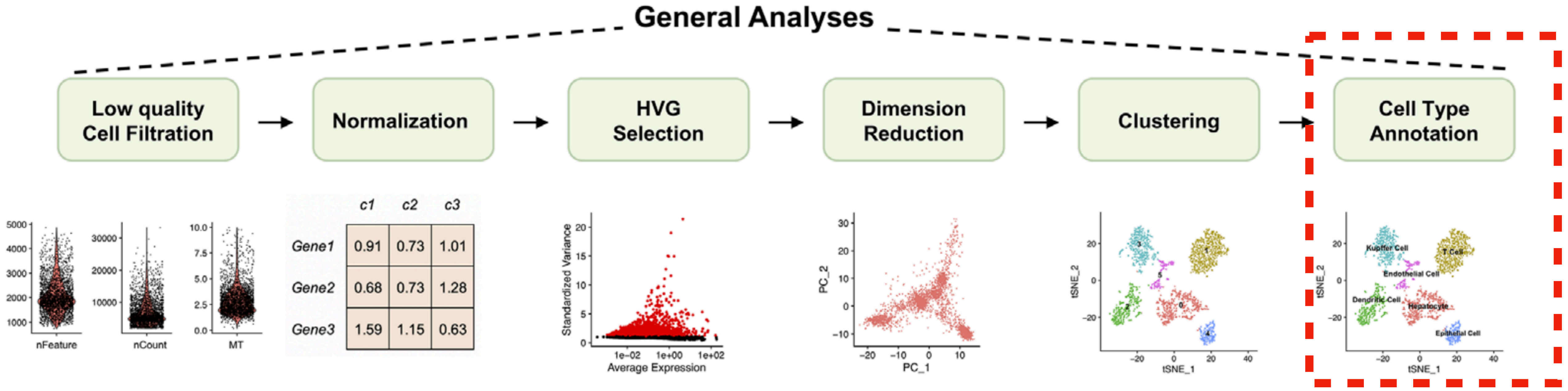


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	*c2*	*c3*
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Analysis Overview:



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Goals:

- Determine gene markers for each cluster
- Identify cell type of each cluster using markers

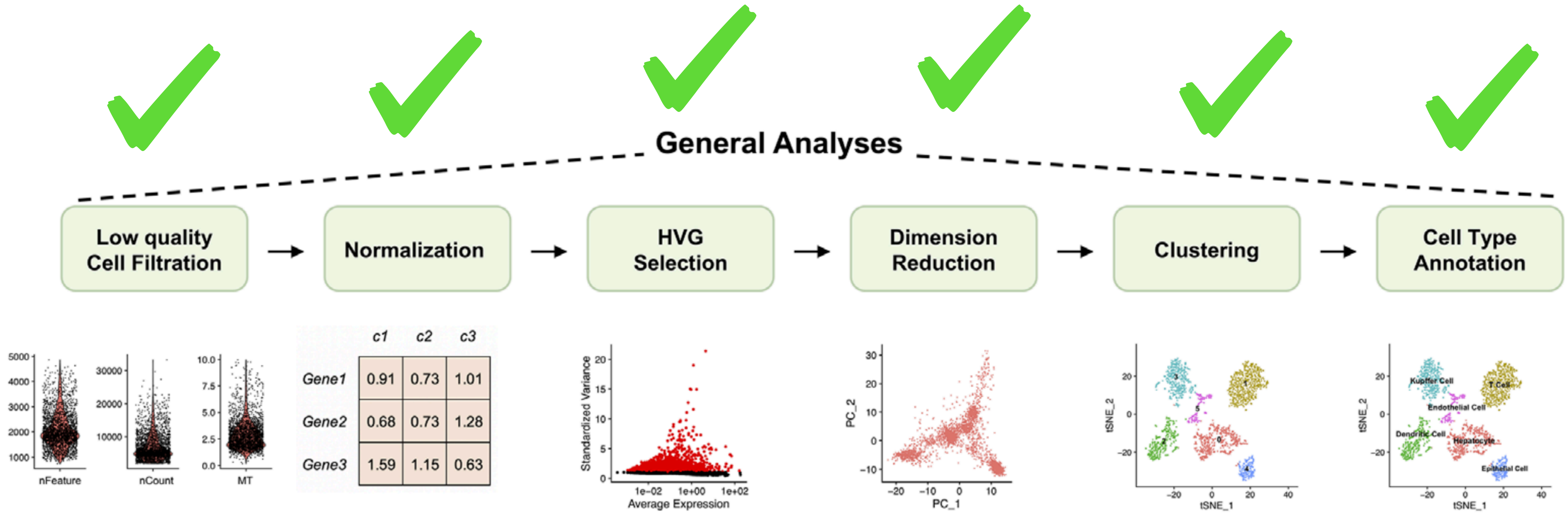
Challenges:

- Highly dependent on the quality of clusters

Recommendations:

- Top Markers are trustworthy (inflated p-values, each cell is a replicate)
- Identify Conserved Markers between conditions for each cluster
- Identify markers that are differentially expressed between specific clusters

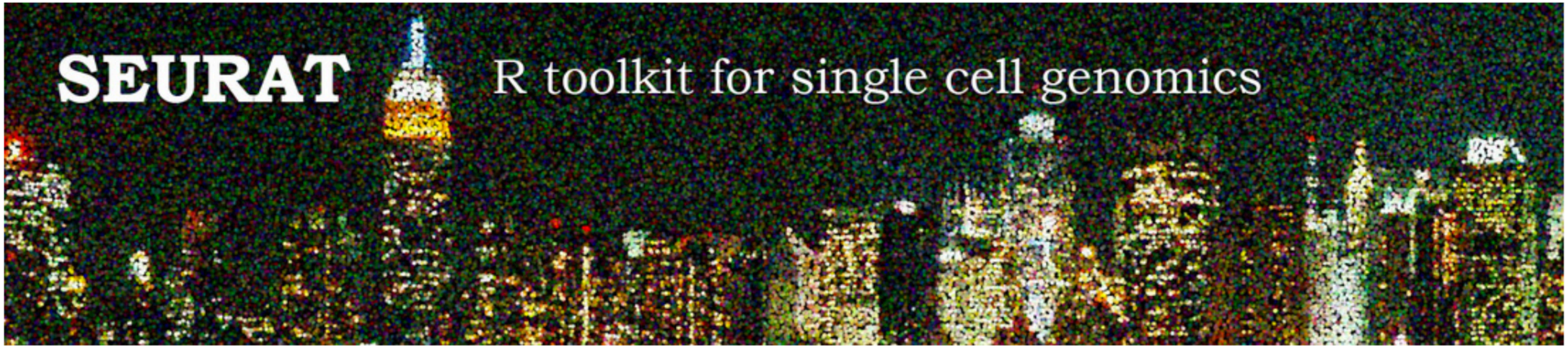
Analysis Overview:

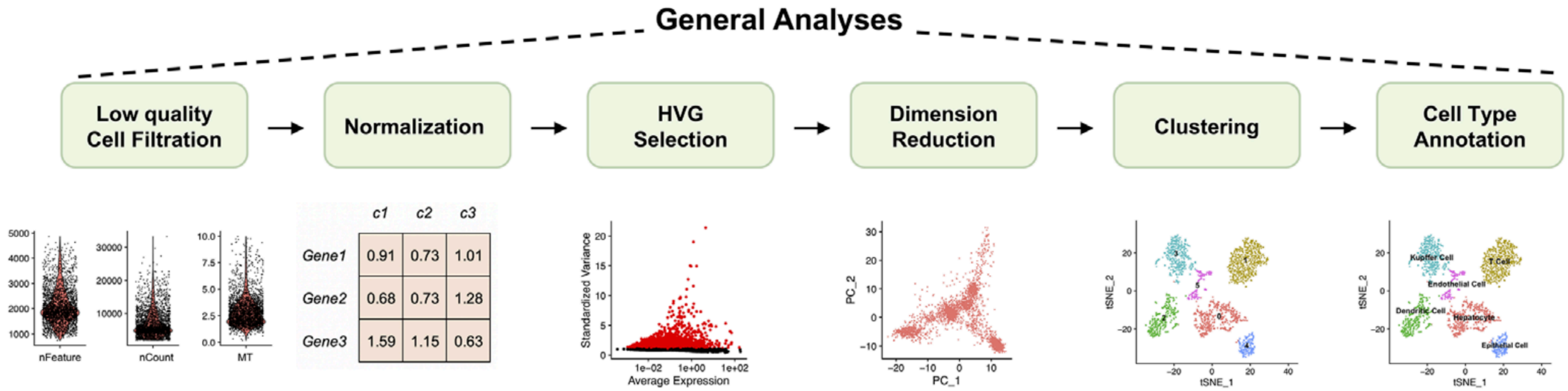


General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

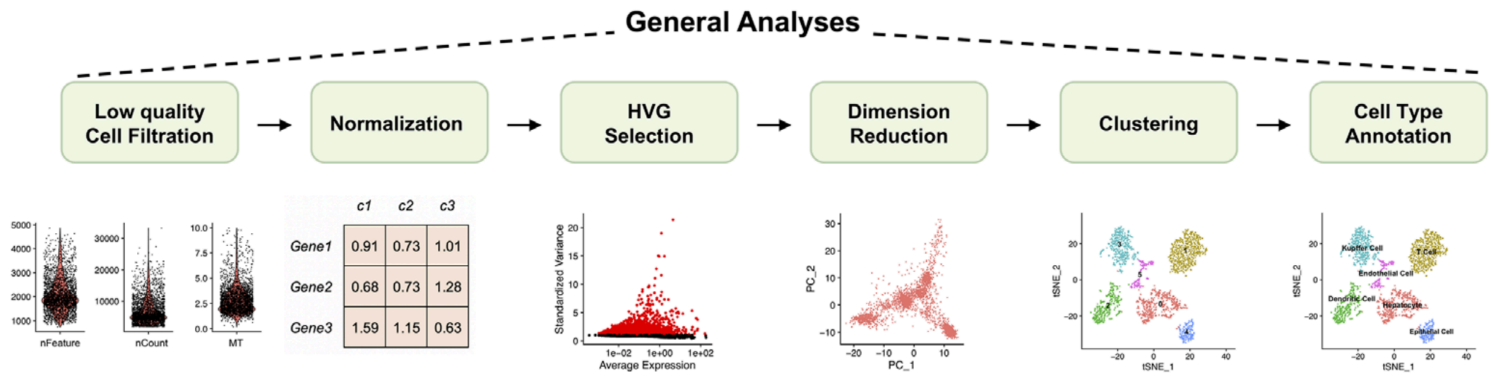
	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63





	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

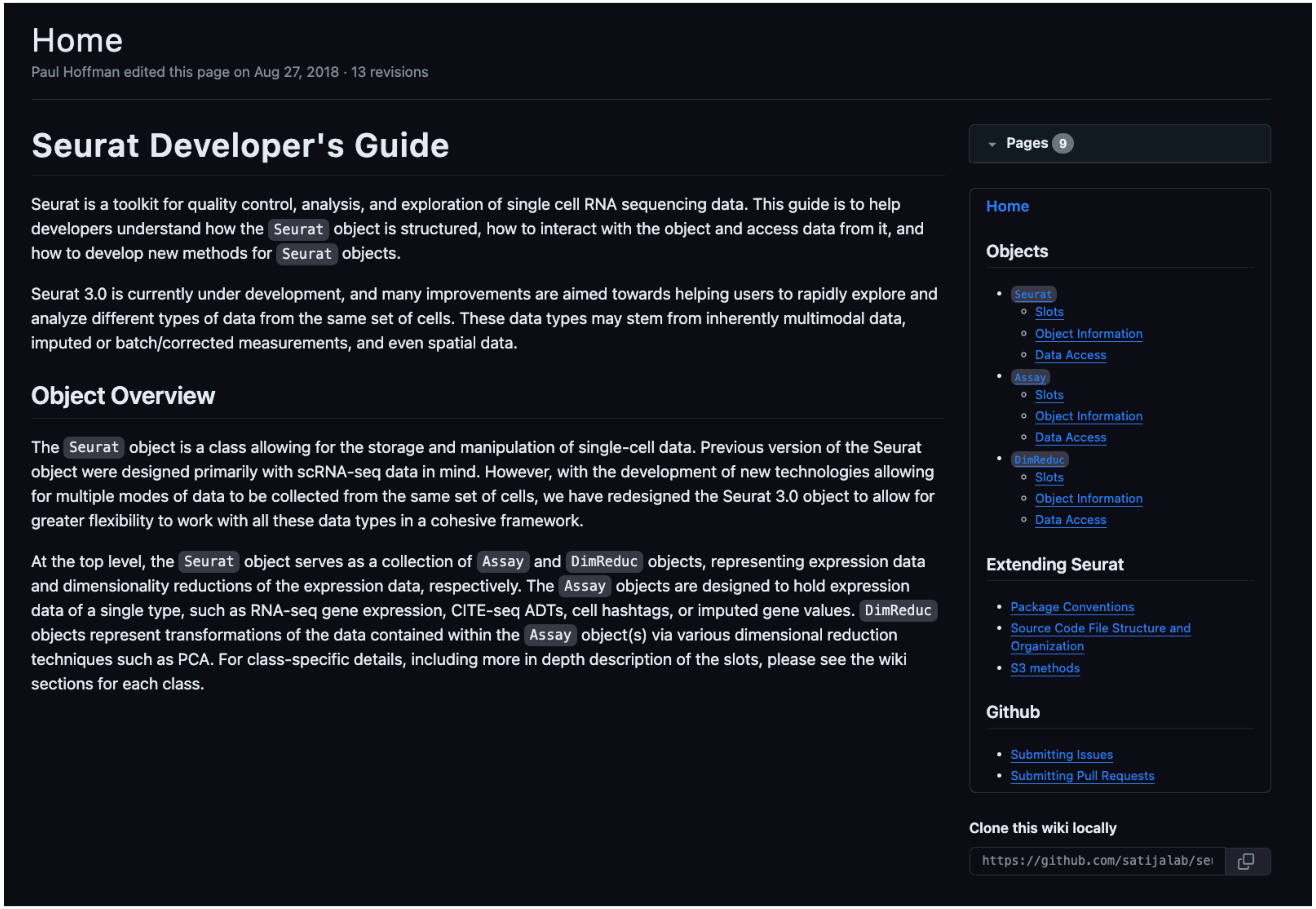
https://github.com/satijalab/seurat/wiki



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63



Home

Paul Hoffman edited this page on Aug 27, 2018 · 13 revisions

Seurat Developer's Guide

Seurat is a toolkit for quality control, analysis, and exploration of single cell RNA sequencing data. This guide is to help developers understand how the `Seurat` object is structured, how to interact with the object and access data from it, and how to develop new methods for `Seurat` objects.

Seurat 3.0 is currently under development, and many improvements are aimed towards helping users to rapidly explore and analyze different types of data from the same set of cells. These data types may stem from inherently multimodal data, imputed or batch/corrected measurements, and even spatial data.

Object Overview

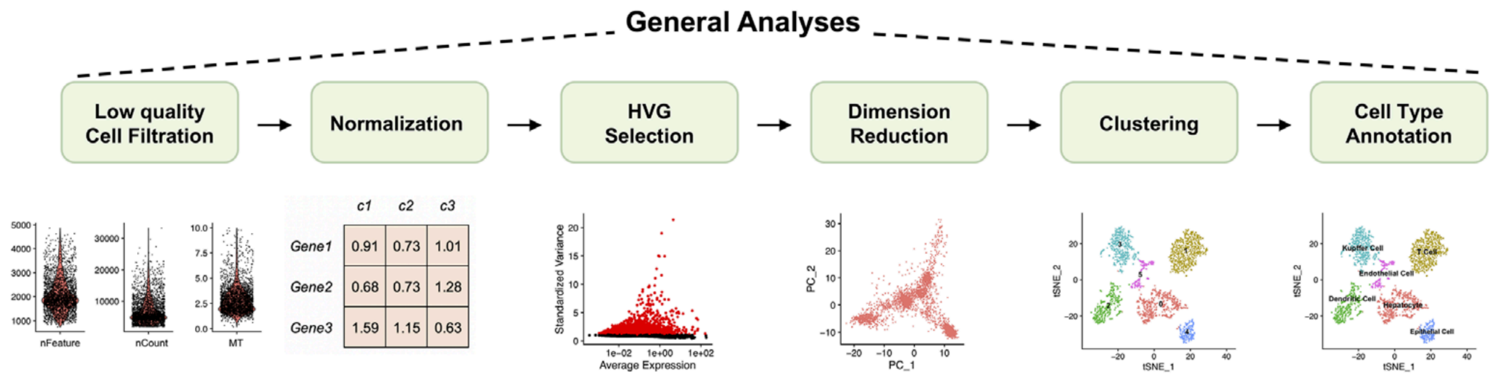
The `Seurat` object is a class allowing for the storage and manipulation of single-cell data. Previous version of the Seurat object were designed primarily with scRNA-seq data in mind. However, with the development of new technologies allowing for multiple modes of data to be collected from the same set of cells, we have redesigned the Seurat 3.0 object to allow for greater flexibility to work with all these data types in a cohesive framework.

At the top level, the `Seurat` object serves as a collection of `Assay` and `DimReduc` objects, representing expression data and dimensionality reductions of the expression data, respectively. The `Assay` objects are designed to hold expression data of a single type, such as RNA-seq gene expression, CITE-seq ADTs, cell hashtags, or imputed gene values. `DimReduc` objects represent transformations of the data contained within the `Assay` object(s) via various dimensional reduction techniques such as PCA. For class-specific details, including more in depth description of the slots, please see the wiki sections for each class.

Clone this wiki locally

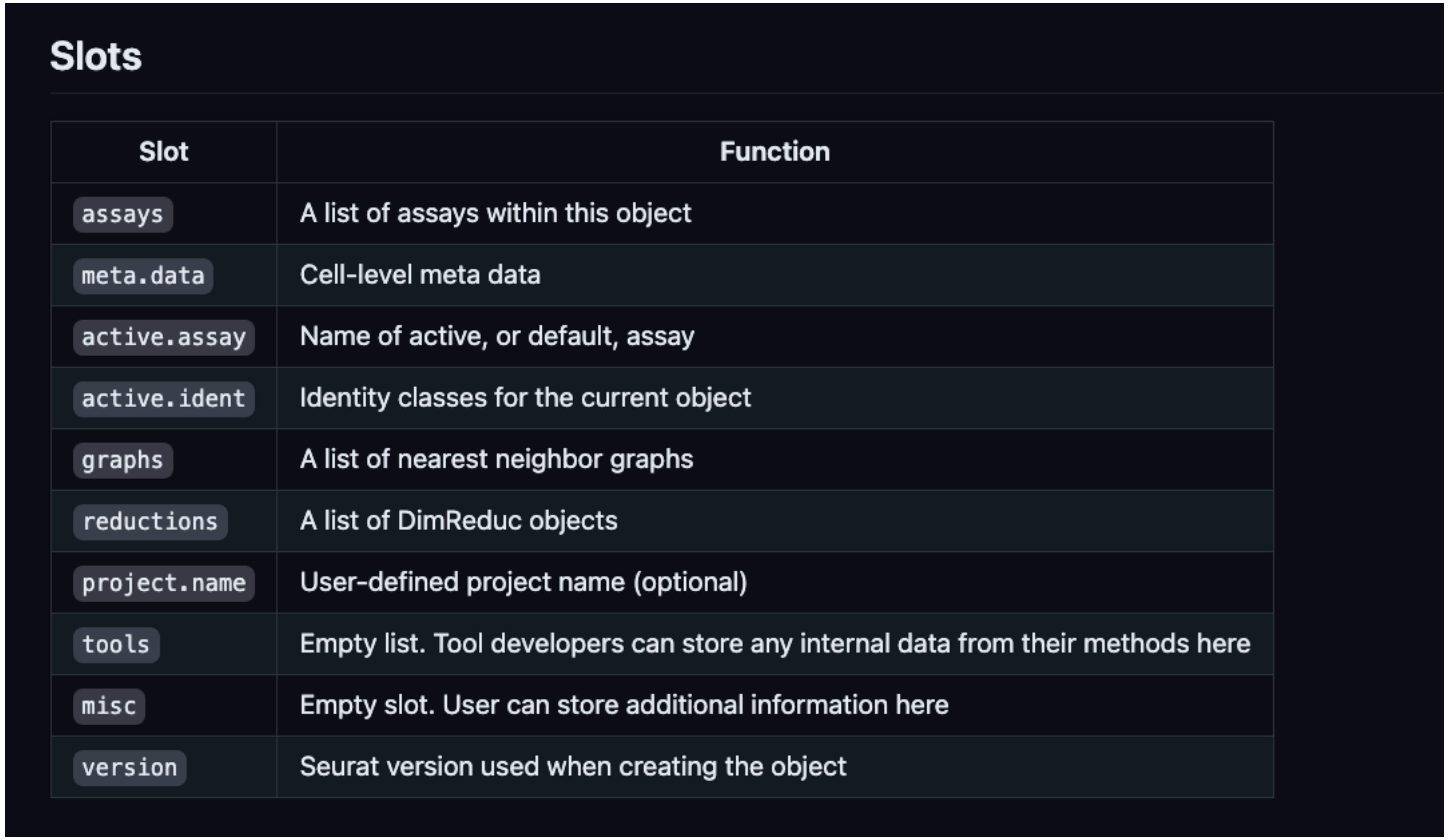
`https://github.com/satijalab/se`

https://github.com/satijalab/seurat/wiki



Seurat Object is a representation of single-cell expression data for R

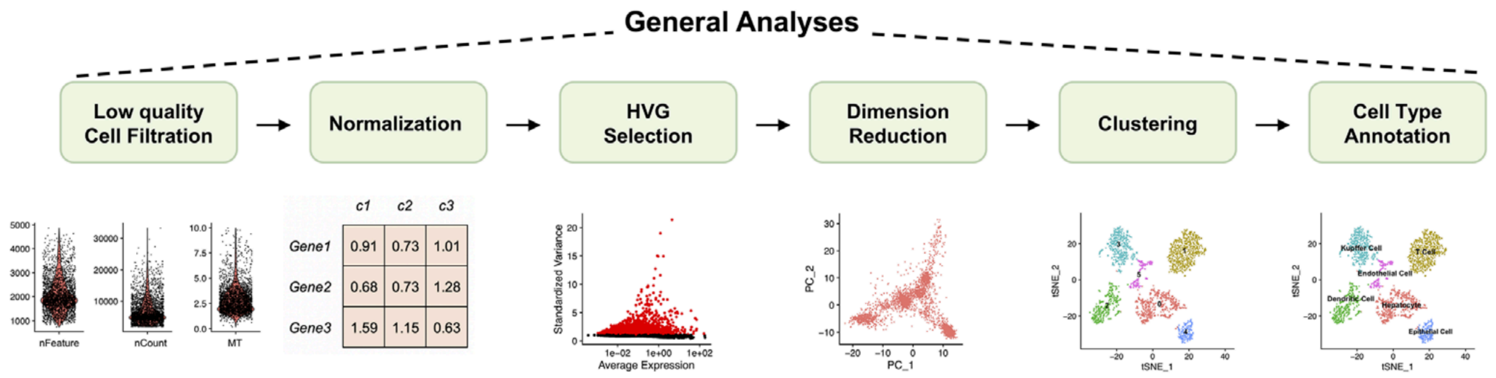
Collection of Expression Data *(Assay)* + Dimensionality Reductions *(DimReduc)*



Slots

Slot	Function
`assays`	A list of assays within this object
`meta.data`	Cell-level meta data
`active.assay`	Name of active, or default, assay
`active.ident`	Identity classes for the current object
`graphs`	A list of nearest neighbor graphs
`reductions`	A list of DimReduc objects
`project.name`	User-defined project name (optional)
`tools`	Empty list. Tool developers can store any internal data from their methods here
`misc`	Empty slot. User can store additional information here
`version`	Seurat version used when creating the object

https://github.com/satijalab/seurat/wiki



Assay’s

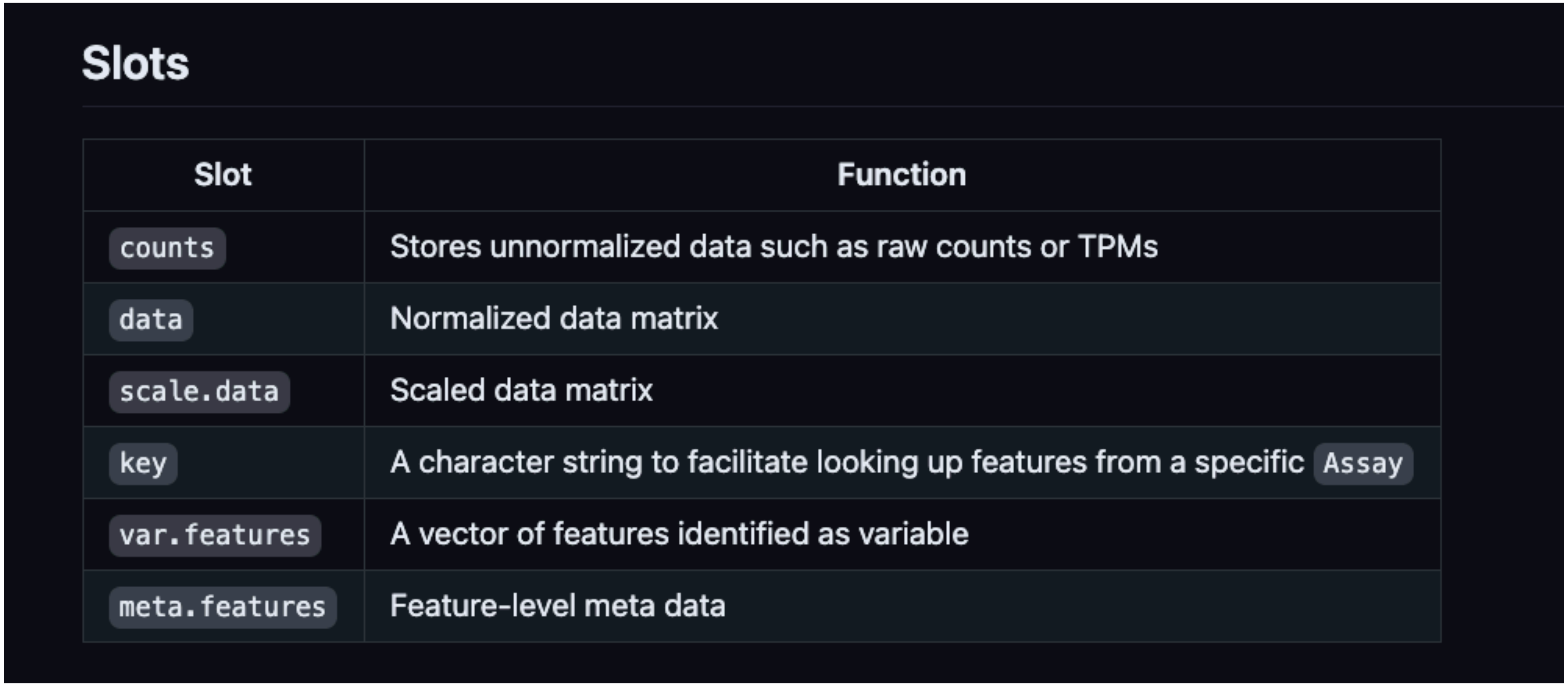
For a typical scRNA-seq experiments, a Seurat object will have a single Assay (“RNA").

Complex scRNA-seq experiments = Multiples Assay’s

scRNA

CITE seq

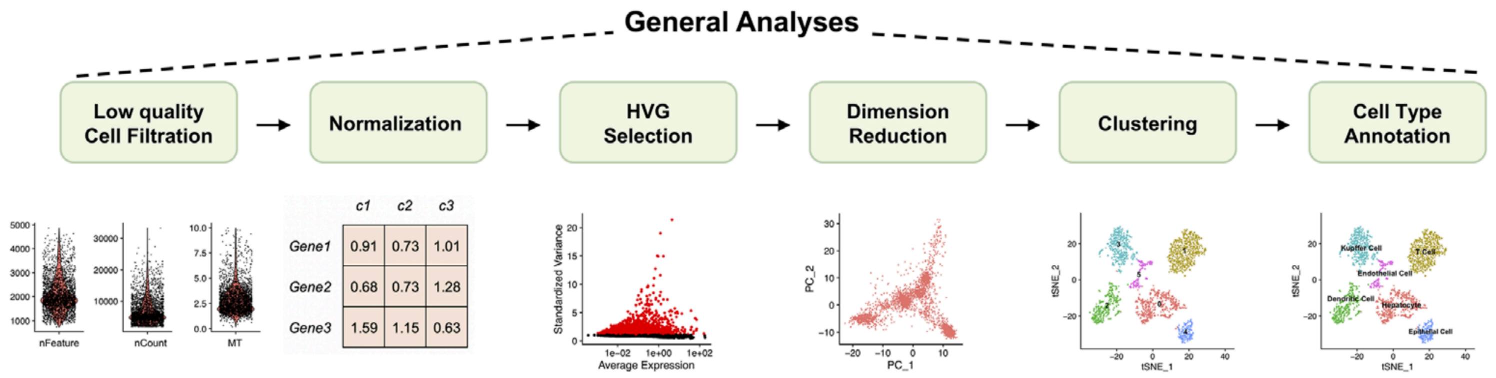
Spatial



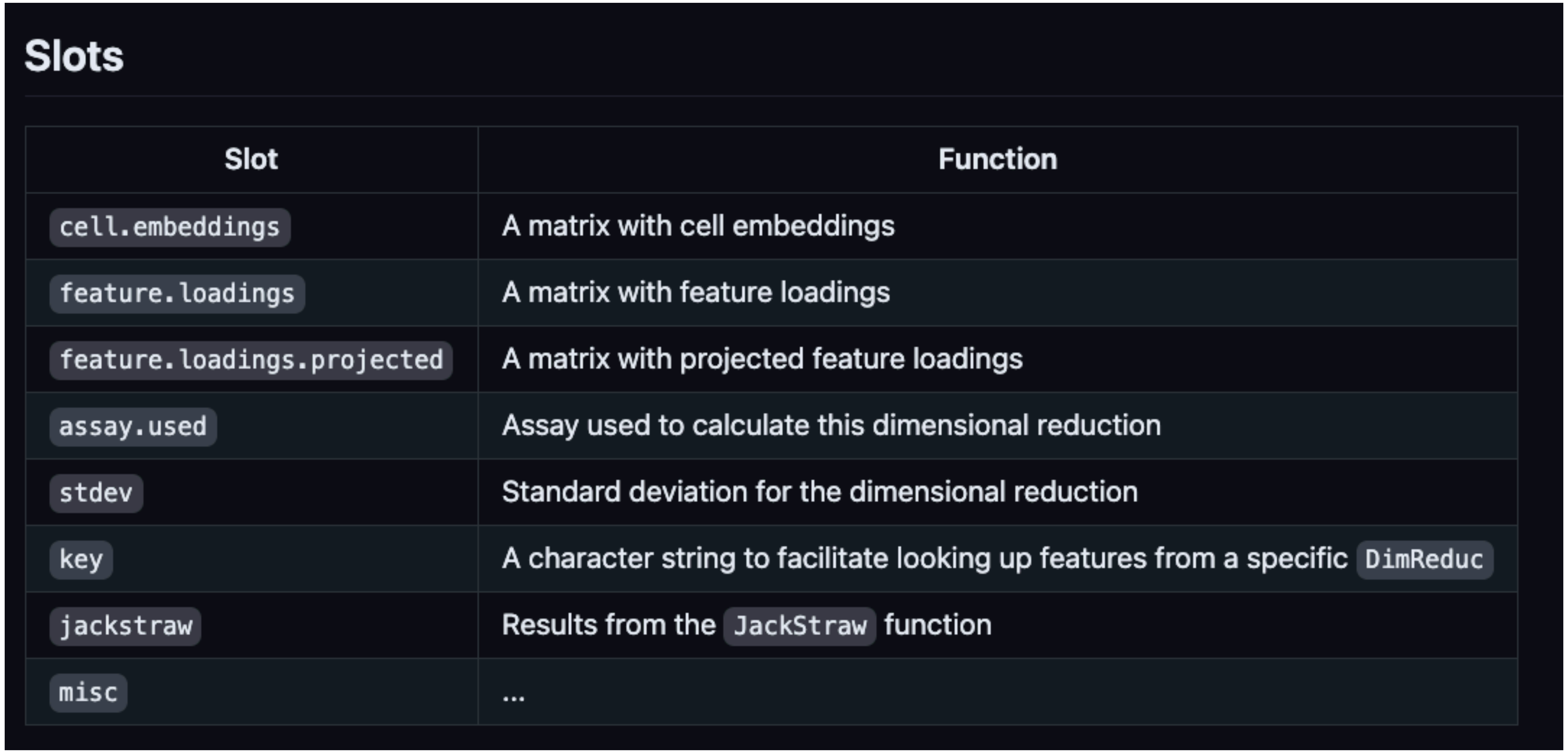
Slots

Slot	Function
`counts`	Stores unnormalized data such as raw counts or TPMs
`data`	Normalized data matrix
`scale.data`	Scaled data matrix
`key`	A character string to facilitate looking up features from a specific `Assay`
`var.features`	A vector of features identified as variable
`meta.features`	Feature-level meta data

https://github.com/satijalab/seurat/wiki



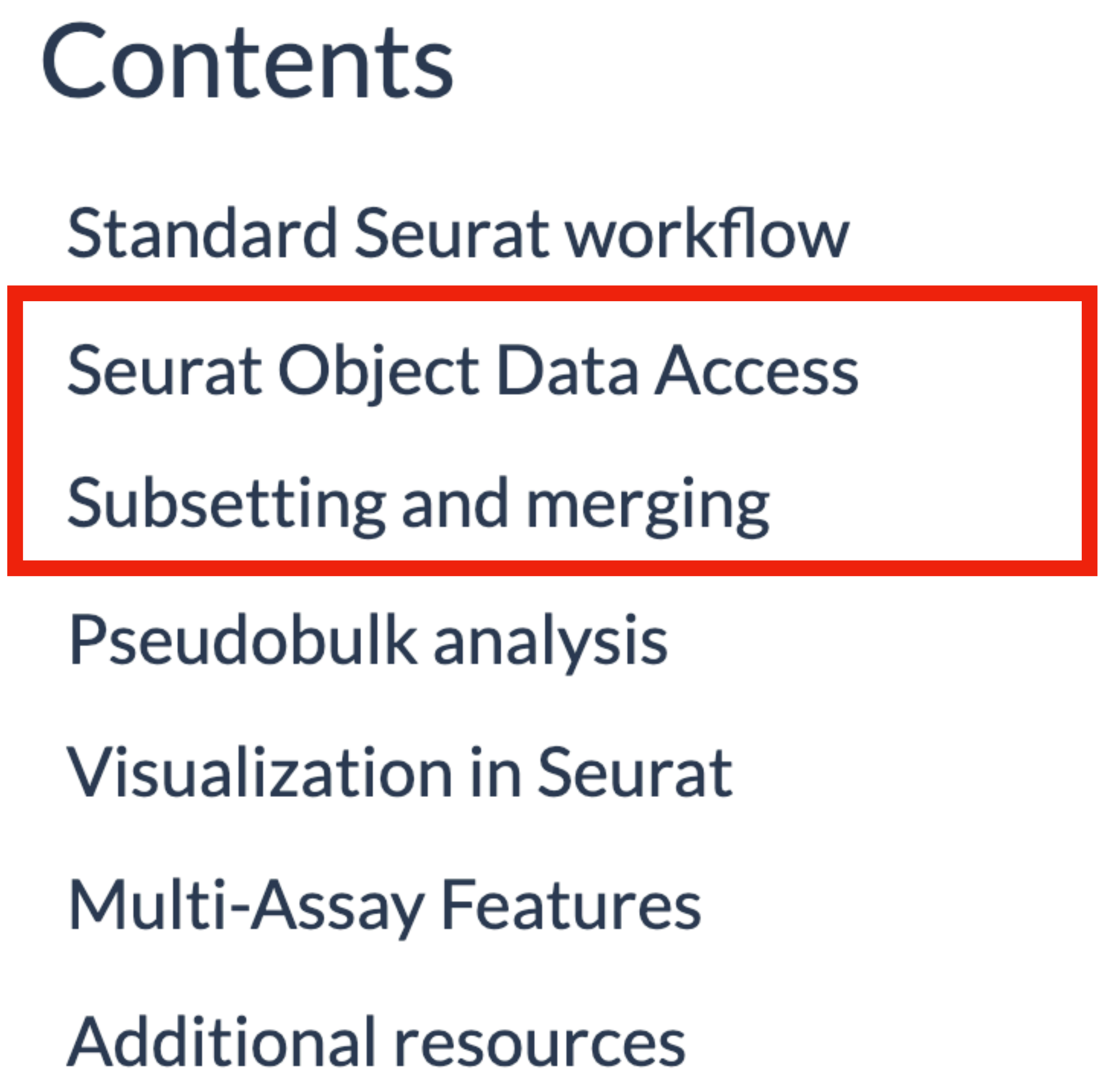
DimReduc object represents a dimensional reduction taken upon the Seurat object.



Slots

Slot	Function
`cell.embeddings`	A matrix with cell embeddings
`feature.loadings`	A matrix with feature loadings
`feature.loadings.projected`	A matrix with projected feature loadings
`assay.used`	Assay used to calculate this dimensional reduction
`stdev`	Standard deviation for the dimensional reduction
`key`	A character string to facilitate looking up features from a specific `DimReduc`
`jackstraw`	Results from the `JackStraw` function
`misc`	...

https://satijalab.org/seurat/articles/essential_commands



Contents

Standard Seurat workflow

Seurat Object Data Access

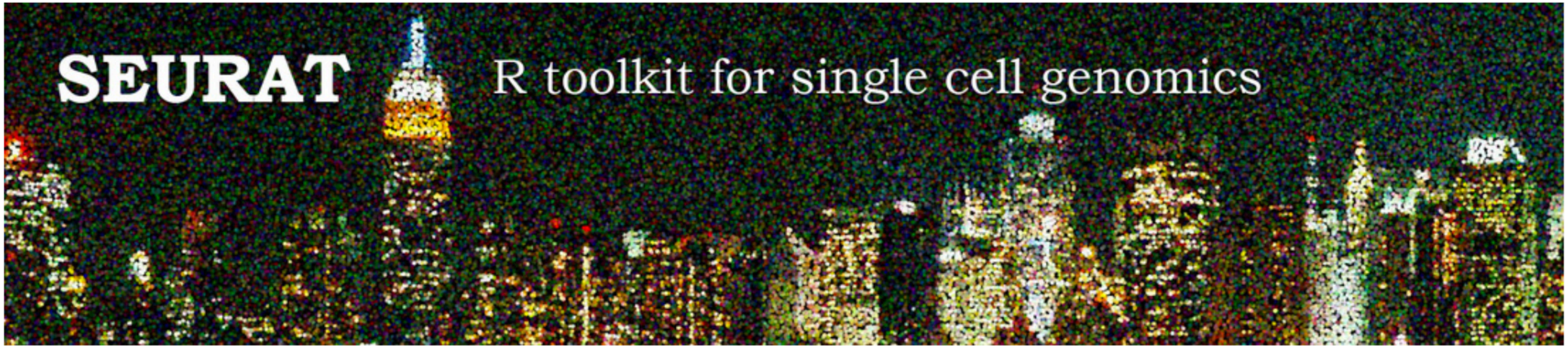
Subsetting and merging

Pseudobulk analysis

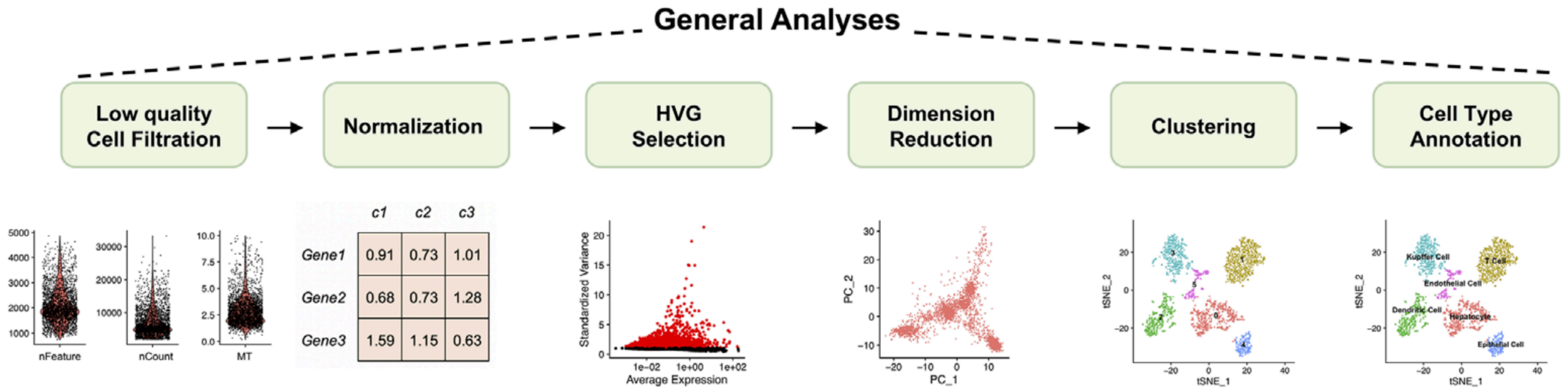
Visualization in Seurat

Multi-Assay Features

Additional resources



SEURAT R toolkit for single cell genomics



General Analyses

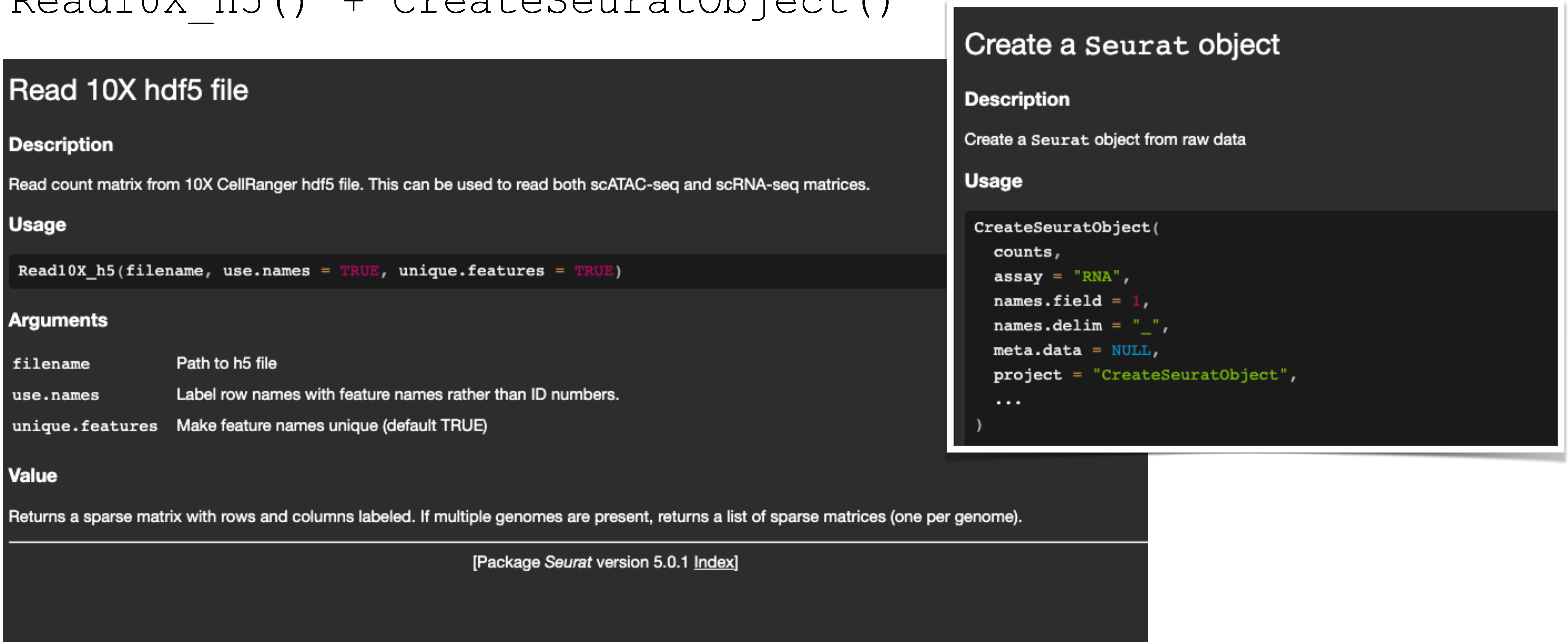
Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	*c2*	*c3*
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Creation of Seurat Object:

Seurat has a handful of functions that can directly import cellranger outputs

`Read10X_h5() + CreateSeuratObject()`



Read 10X hdf5 file

Description

Read count matrix from 10X CellRanger hdf5 file. This can be used to read both scATAC-seq and scRNA-seq matrices.

Usage

```
Read10X_h5(filename, use.names = TRUE, unique.features = TRUE)
```

Arguments

`filename`	Path to h5 file
`use.names`	Label row names with feature names rather than ID numbers.
`unique.features`	Make feature names unique (default TRUE)

Value

Returns a sparse matrix with rows and columns labeled. If multiple genomes are present, returns a list of sparse matrices (one per genome).

[Package *Seurat* version 5.0.1 Index]

Create a `Seurat` object

Description

Create a `Seurat` object from raw data

Usage

```
CreateSeuratObject(
  counts,
  assay = "RNA",
  names.field = 1,
  names.delim = "_",
  meta.data = NULL,
  project = "CreateSeuratObject",
  ...
)
```

Creation of Seurat Object:

In a *multisample* experiment:

Create a seurat object for each sample

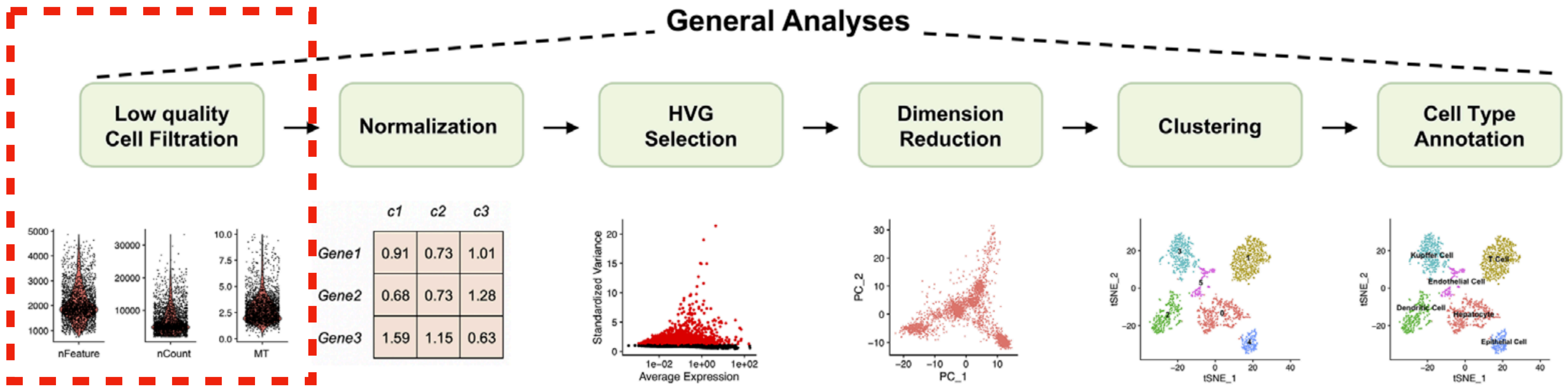
Merge all seurat objects together using the merge function

```
sobj <- merge(x = sobj.list[[1]], y = sobj.list[2:length(sobj.list)], merge.data=TRUE)
```

Main Function

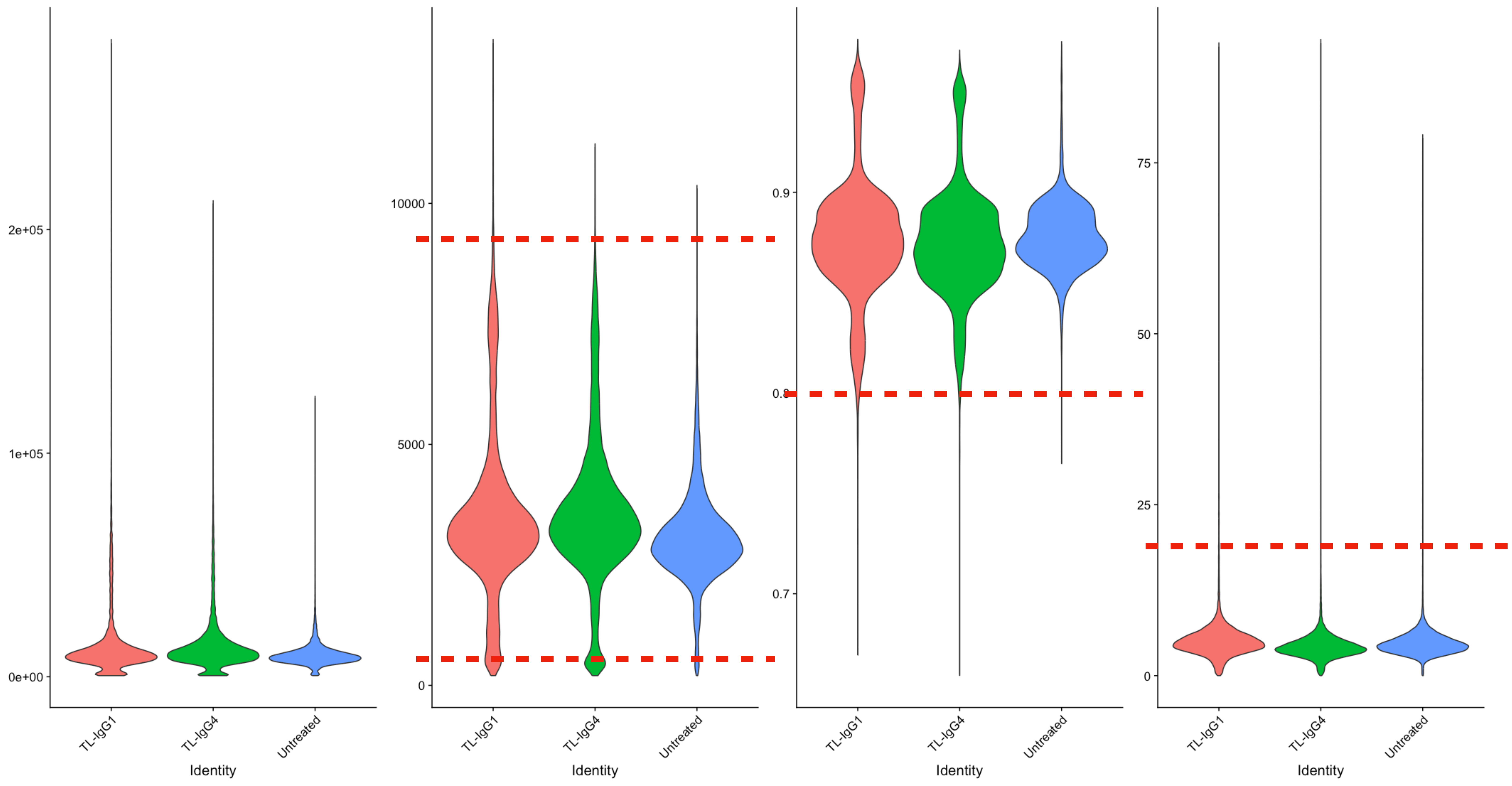
An R list object containing all Seurat Objects

Low Quality Cell Filtration:



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

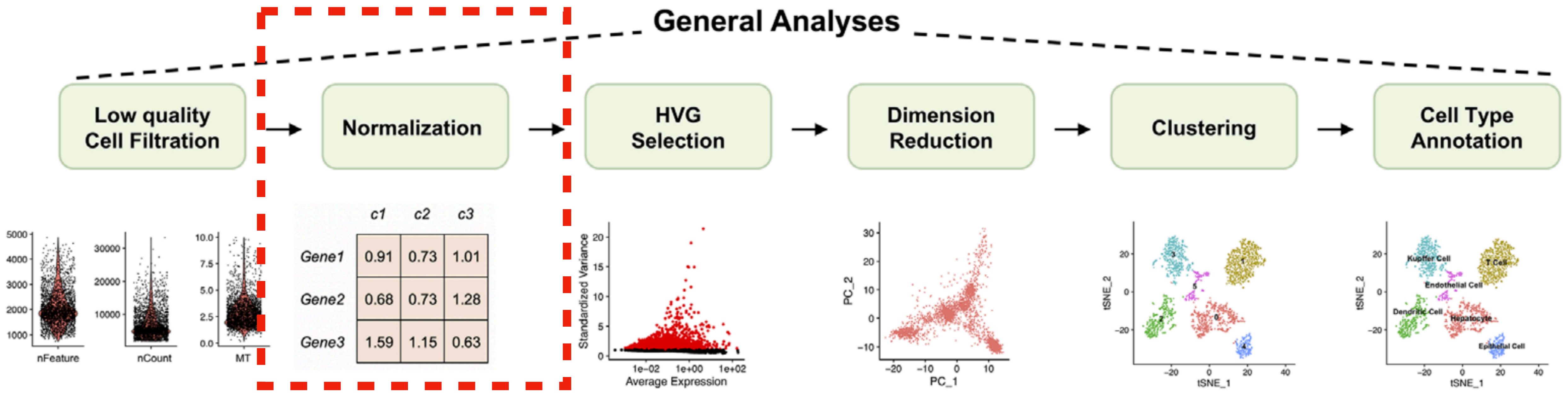
nCount_RNA nFeature_RNA log10GenesPerUMI percent.MT



Merged Seurat Object

```
sobj.filtered <- subset(sobj,
subset = nFeature_RNA > 1000 &
nFeature_RNA < 9000 &
percent.mt < 20 &
log10GenesPerUMI > 0.80)
```

Normalization



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

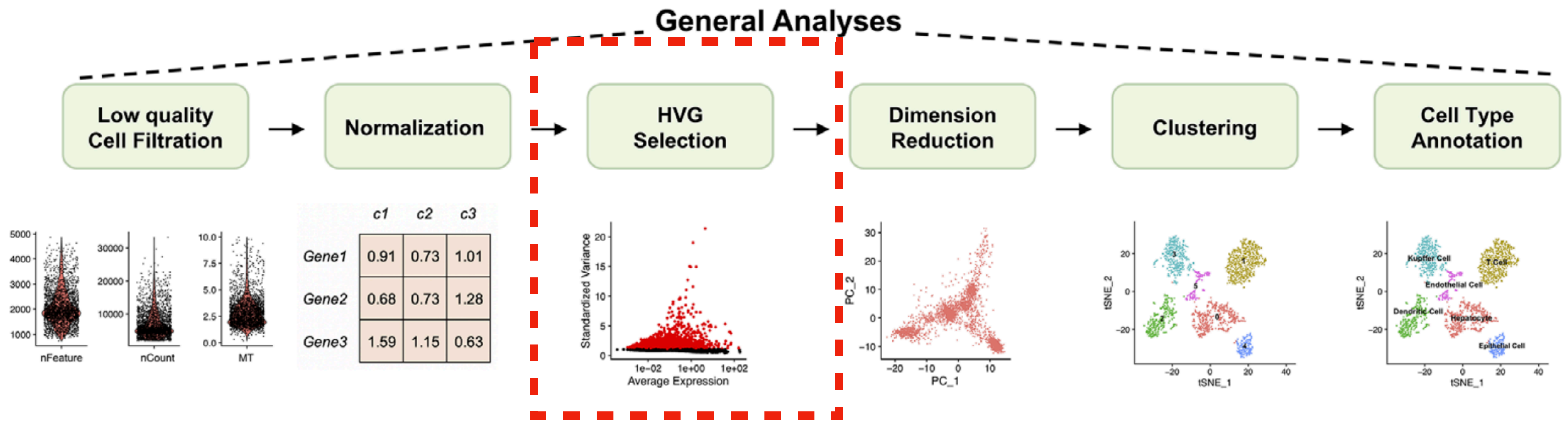
	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Main Function

```
sobj.filtered <- NormalizeData(sobj.filtered, normalization.method = "LogNormalize", verbose = T)
```

Normalization Method

HVG Selection



General Analyses

Low quality Cell Filtration → Normalization → HVG Selection → Dimension Reduction → Clustering → Cell Type Annotation

	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

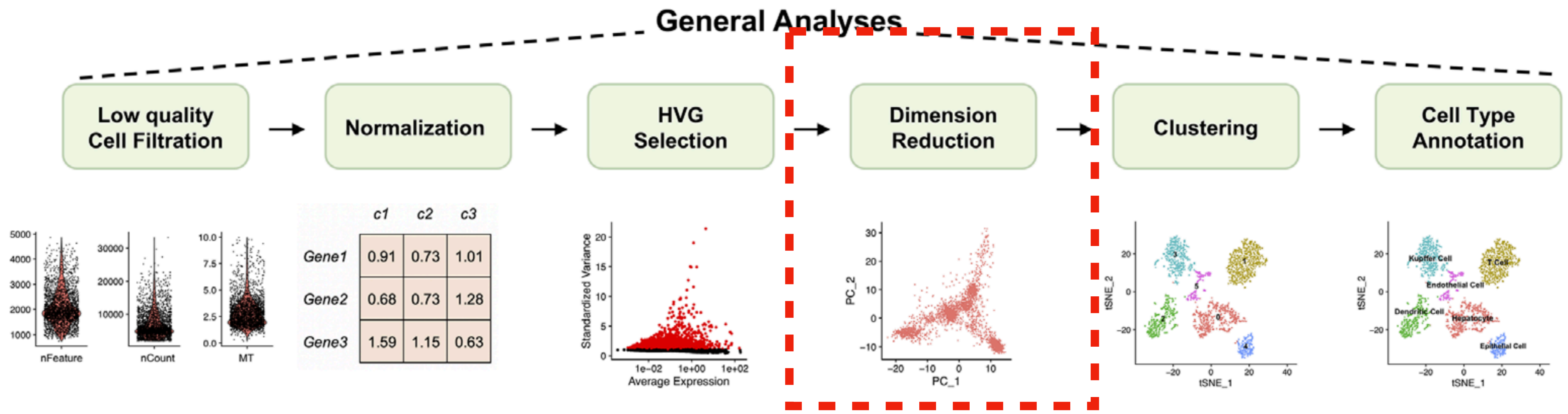
Main Function

Number of top variable features selected

```
sobj.filtered <- FindVariableFeatures(sobj.filtered, nfeatures = 3000,
selection.method = "vst", verbose = T)
```

Selection Method

Dimension Reduction



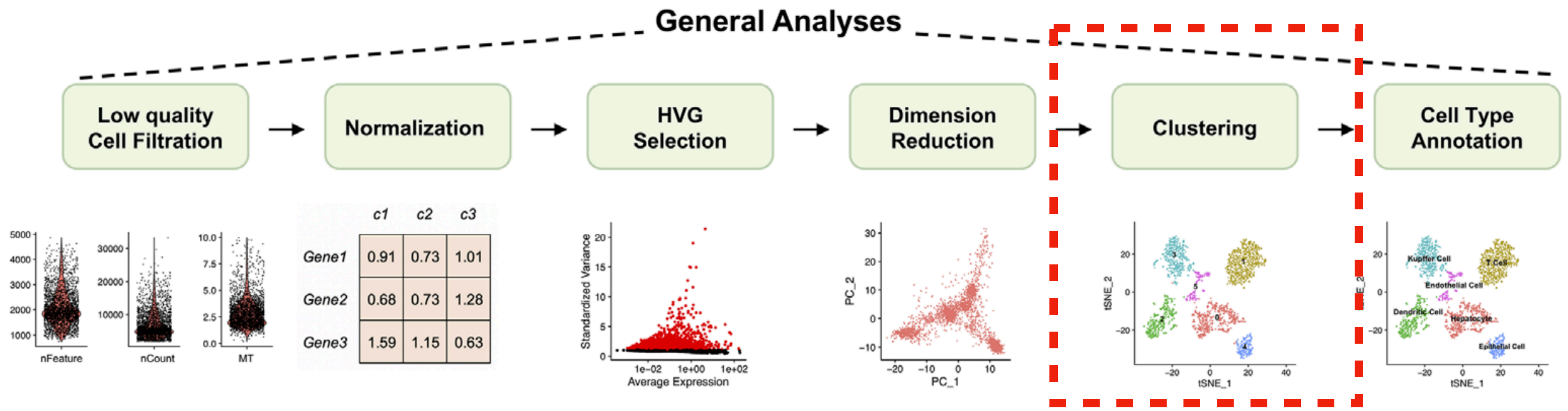
Main Function

```
sobj.filtered <- RunPCA(sobj.filtered, npcs = 50, verbose = T)
```

Number of PC’s to compute

```
ElbowPlot(sobj.filtered, ndims = 50, reduction = "pca")
```

Clustering



	c1	c2	c3
Gene1	0.91	0.73	1.01
Gene2	0.68	0.73	1.28
Gene3	1.59	1.15	0.63

Main Function

```
sobj.filtered <- FindNeighbors(sobj.filtered, dims = 1:50, reduction = "pca")
```

PCA Dimensions

```
sobj.filtered <- FindClusters(sobj.filtered, resolution = 0.4, verbose = T)
```

Main Function

Sets the granularity

```
sobj.filtered <- RunUMAP(sobj.filtered, dims = 1:50, reduction = "pca")
```

Main Function