KFO 309 VIRUS-INDUCED LUNG INJURY

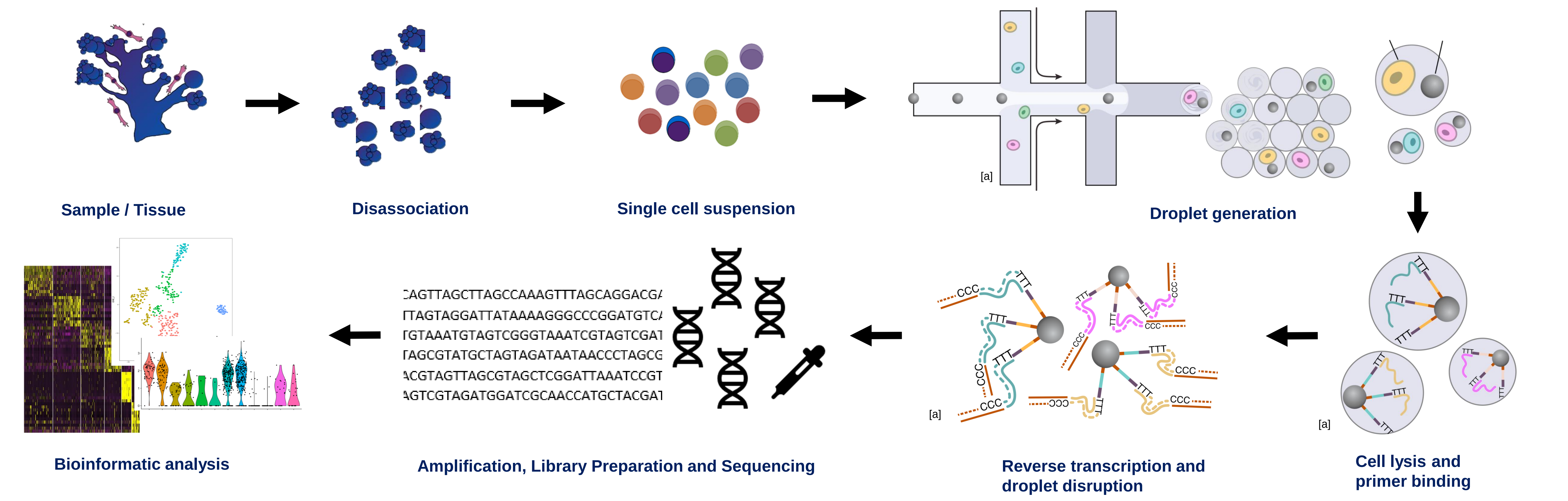
WASP - A versatile web-accessible single cell RNA-Seq processing platform

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Andreas Hoek[1], Katharina Maibach[1], Ebru Özmen[1], Torsten Hain[2], Susanne Herold[3], Alexander Goesmann[1]

[1]Bioinformatics and Systems Biology, Justus Liebig University Giessen, [2]Institute of Medical Microbiology, Justus Liebig University Giessen, [3]Department of Medicine II, Justus Liebig University Giessen

Processing of single cells


Sample / Tissue
Disassociation
Single cell suspension
[a]
Droplet generation
Bioinformatic analysis
CAGTTAGCTTAGCCAAAGTTTAGCAGGACGA
TTAGTAGGATTATAAAAGGGCCCGGATGTCA
TGTAAATGTAGTCGGGTAAATCGTAGTCGAT
TAGCGTATGCTAGTAGATAATAACCCTAGCG
ACGTAGTTAGCGTAGCTCGGATTAAATCCGT
AGTCGTAGATGGATCGCAACCATGCTACGAT
Amplification, Library Preparation and Sequencing
[a]
Reverse transcription and droplet disruption
[a]
Cell lysis and primer binding

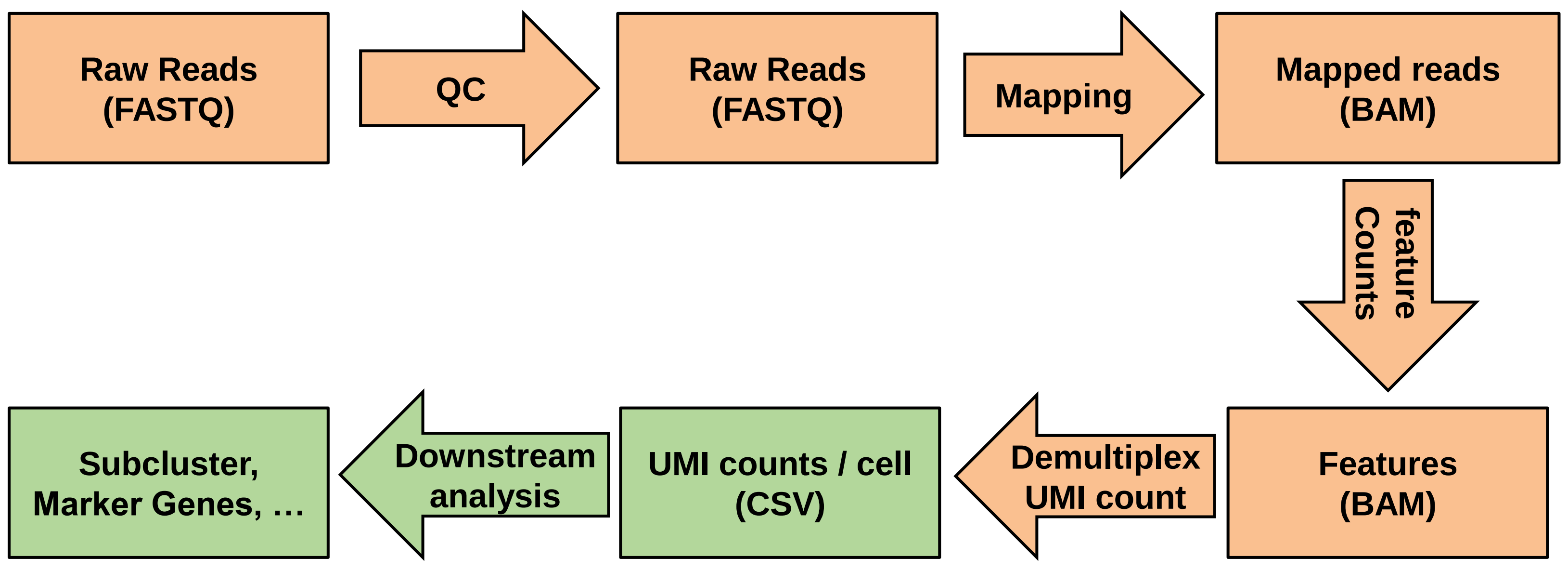
Single cell RNA sequencing:

- Single cell RNA sequencing (scRNA-Seq) allows analysis of samples, e.g. tissues at high resolutions
- In contrast to bulk RNA-Seq, scRNA-Seq does allow the detection of cellular heterogeneity, cell types and gene expression of specific cells that are masked in bulk RNA-Seq
- Protocols developed in the early stages (2009 Tang et. al) only allowed analysis of a few cells
- Newer protocols such as Drop-Seq lowered costs and increased throughput allowing analysis of up to thousands of cells per run

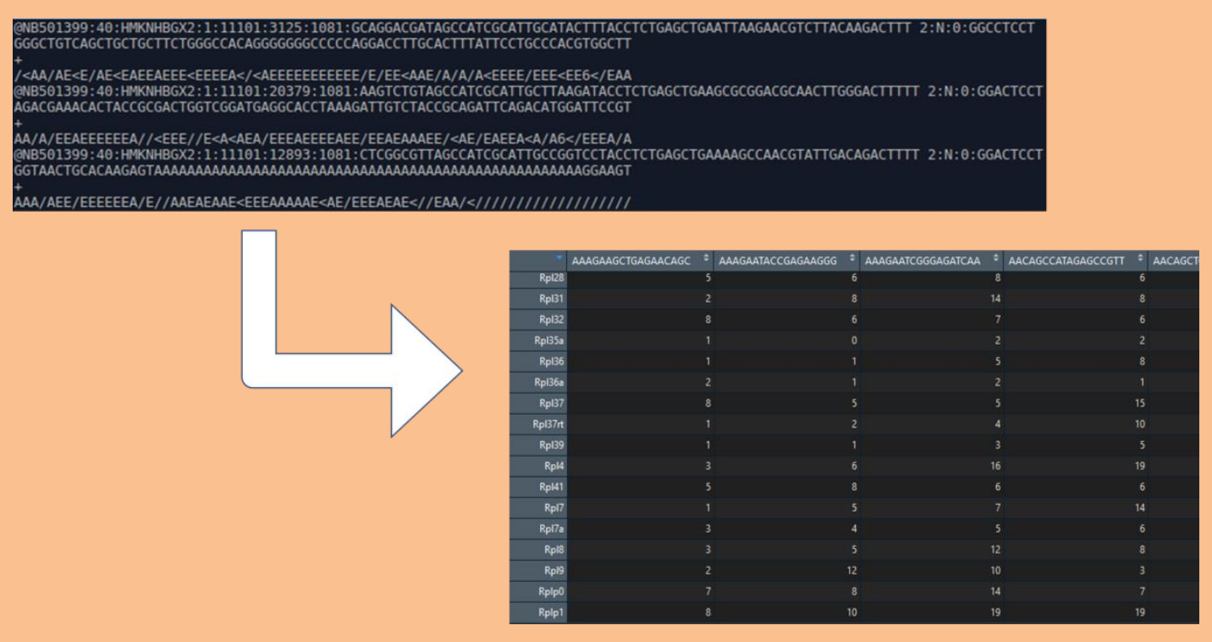
Bioinformatic analysis of single cell RNA sequencing data

Workflow

- Analysis starts with the FASTQ file(s) containing the raw reads obtained from sequencing and ends with detection of cellular clusters, differentially expressed genes and marker genes including visualizations
- This workflow can be separated into two major steps:
  - Pre-Processing – Processing of raw reads to a gene-expression matrix
  - Post-Processing – Processing of gene-expression matrix to cellular clusters, differential expression, subclusters, marker-genes, visualization of results

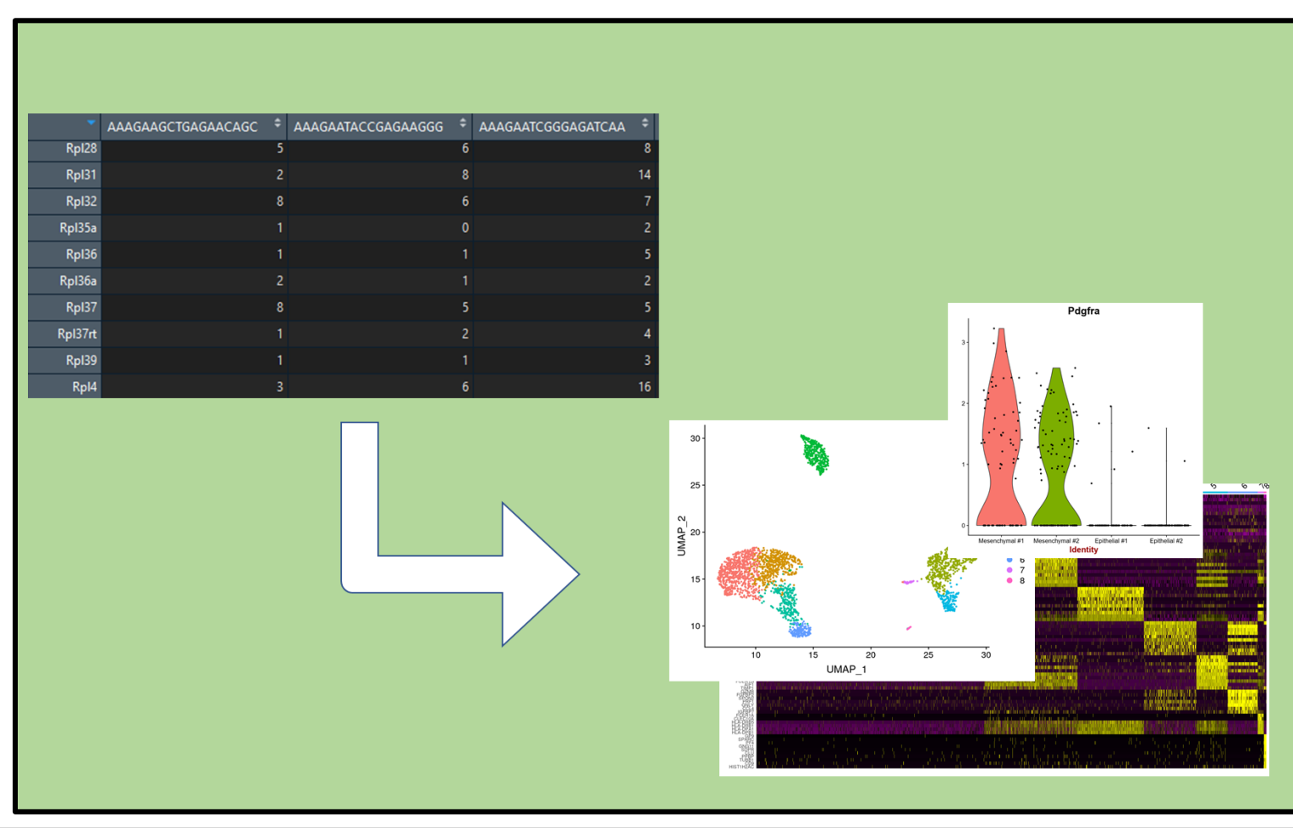


Pre-Processing:



- Quality control of reads using FASTQC
- Reads are mapped to reference genome using STAR
- Counting of reads mapping to genomic features with featureCounts
- Demultiplexing reads to their cell of origin
- Counting unique mRNA fragments (UMI) per genomic feature

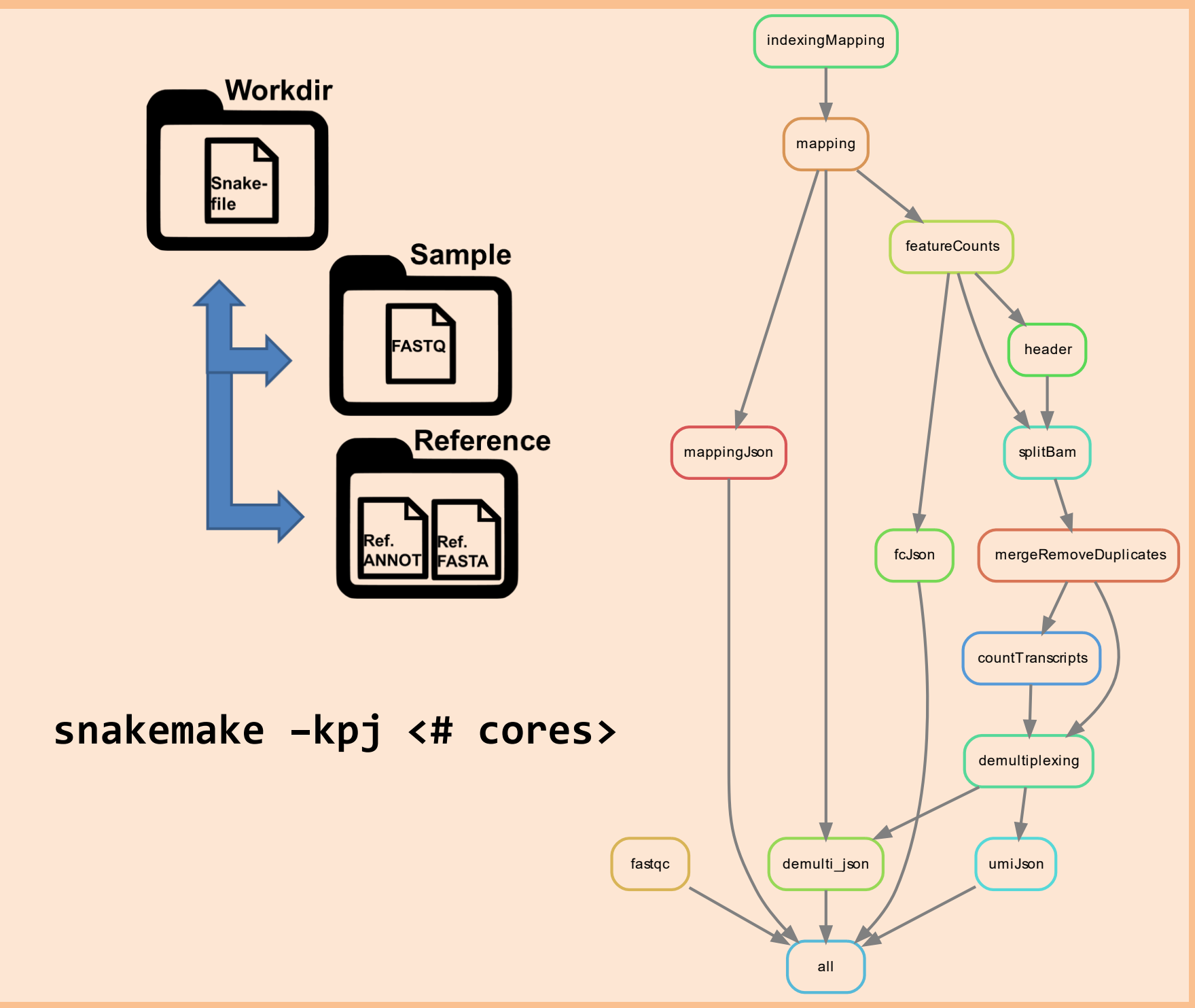
Post-Processing:



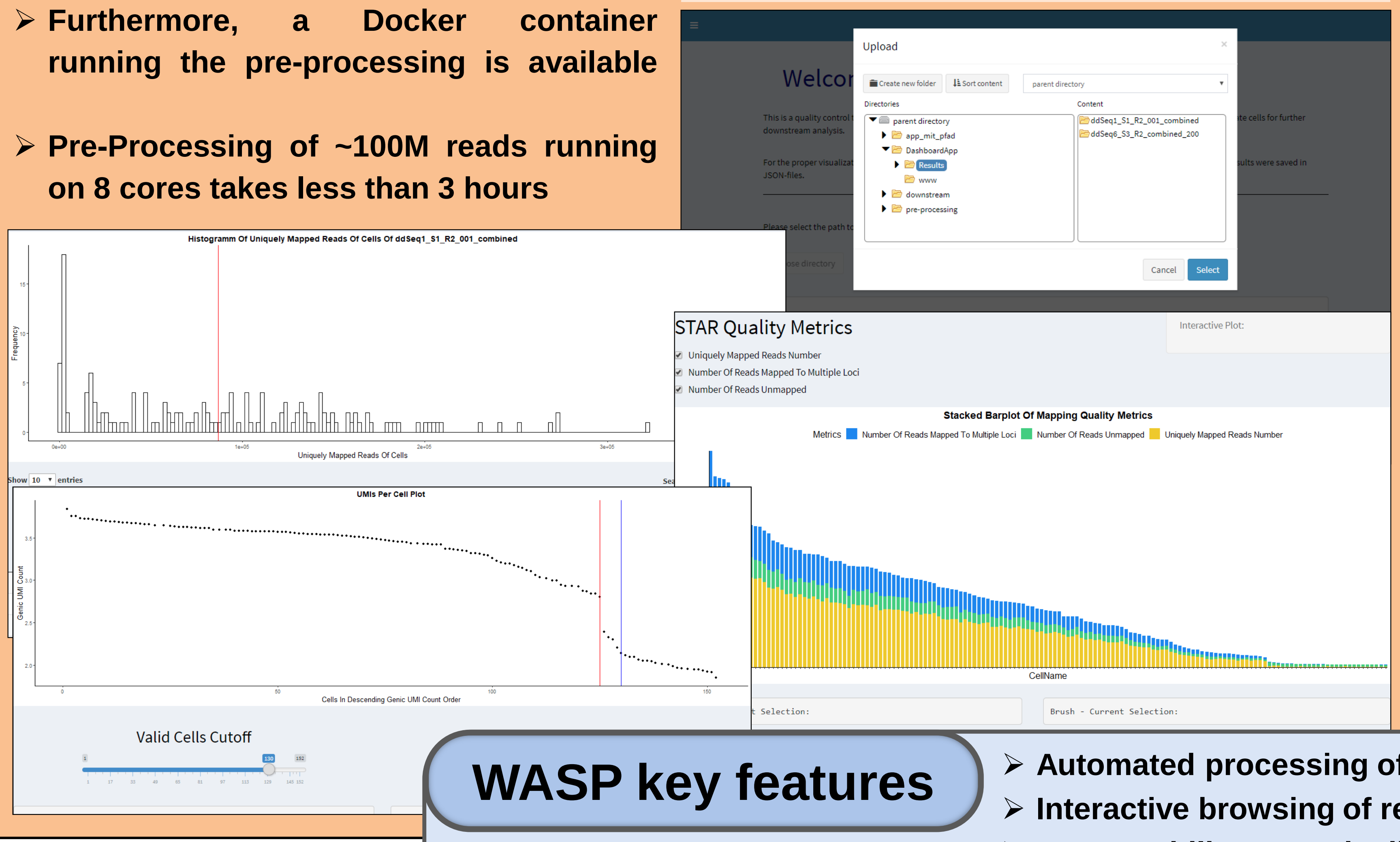
- Steps performed using R packages
- Quality control & normalization
- Detection of highly variable genes
- Dimension reduction (PCA, t-SNE, UMAP)
- Clustering of cells and characterization of clusters, detection of differential gene expression, …
- Visualization of results

Processing scRNA-seq data with WASP

- Pre-processing is performed using a snakemake workflow
- Users only have to provide reference genome, annotation and the FASTQ file(s)
- Metrics for each step are presented as an interactive R-Shiny web page
- After processing, users can select the number of cells to proceed
- As final result the pre-processing pipeline generates a gene expression matrix
- All dependencies (Tools and packages) can simply be installed via Conda
- Furthermore, a Docker container running the pre-processing is available
- Pre-Processing of ~100M reads running on 8 cores takes less than 3 hours



```
snakemake -kpj <# cores>
```

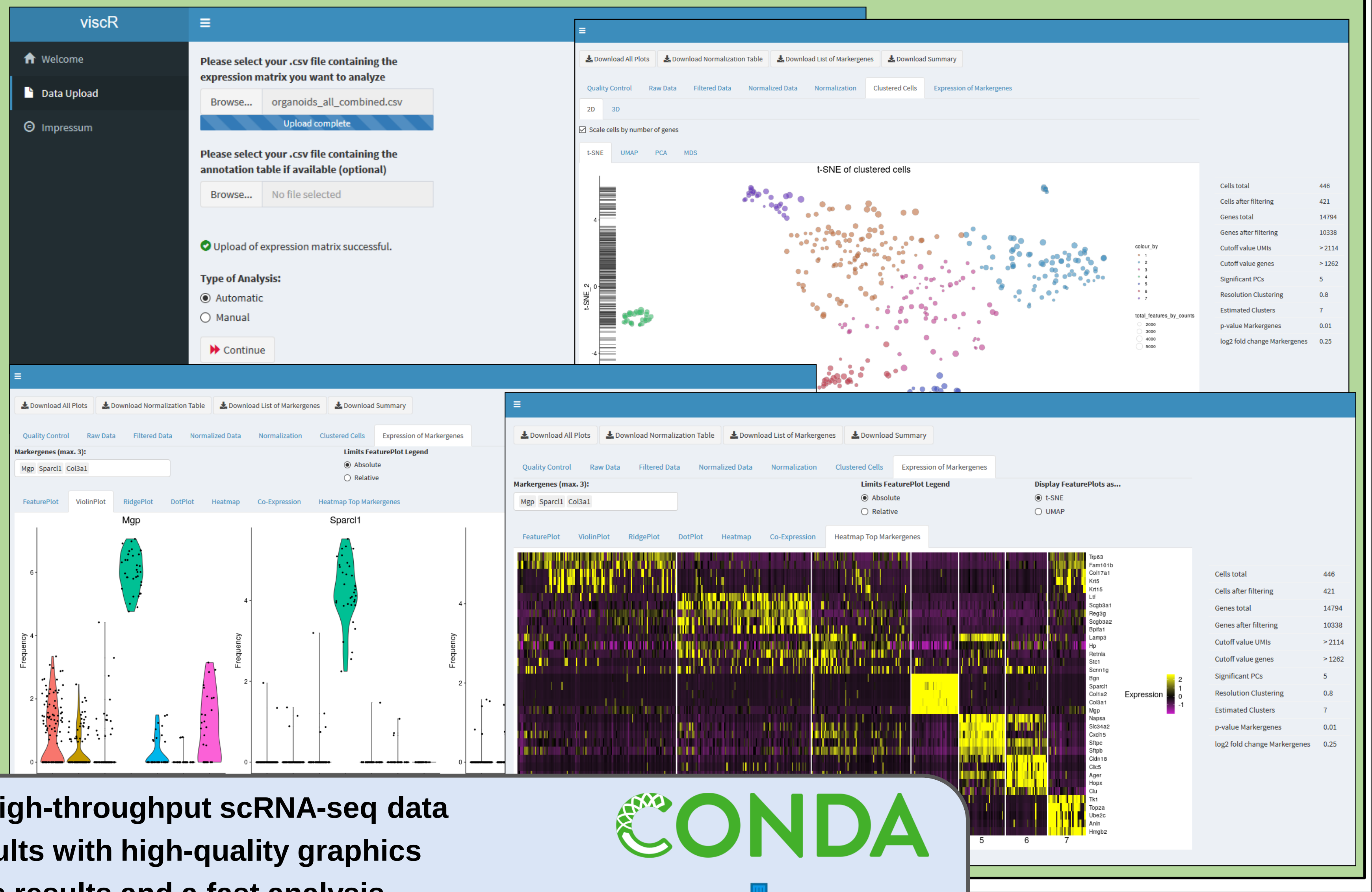

Upload
Create new folder
Sort content
parent directory
Directories
Content
parent directory
app_mit_pfad
DashboardApp
Results
www
downstream
pre-processing
ddSeq1_S1_R2_001_combined
ddSeq6_S3_R2_combined_200
Cancel
Select
Welcom

This is a quality control ... cells for further downstream analysis.

For the proper visualizat... sults were saved in JSON-files.

Histogramm Of Uniquely Mapped Reads Of Cells Of ddSeq1_S1_R2_001_combined
Frequency
Uniquely Mapped Reads Of Cells
Show 10 entries
UMIs Per Cell Plot
Genic UMI Count
Cells In Descending Genic UMI Count Order
Valid Cells Cutoff
STAR Quality Metrics
Interactive Plot:
Uniquely Mapped Reads Number
Number Of Reads Mapped To Multiple Loci
Number Of Reads Unmapped
Stacked Barplot Of Mapping Quality Metrics
Metrics
Number Of Reads Mapped To Multiple Loci
Number Of Reads Unmapped
Uniquely Mapped Reads Number
CellName
Brush - Current Selection:

- Post-processing is performed using a variety of designated R packages
- Visualizations of all analysis steps are presented as an interactive R-Shiny web page
- WASP provides an automated workflow for unexperienced users and a manual workflow
- The manual workflow allows the user to choose parameters for each step and directly shows the impact in R-Shiny
- All dependencies (Python, R and packages) can be installed for Windows and Linux
- Furthermore, a Docker container running the post-processing is available
- Post-Processing of 1,000 cells running on a standard PC/Laptop takes ~5min



Please select your .csv file containing the expression matrix you want to analyze

Browse... organoids_all_combined.csv

Upload complete

Please select your .csv file containing the annotation table if available (optional)

Browse... No file selected

Upload of expression matrix successful.

Type of Analysis:

● Automatic

○ Manual

Continue

☑ Scale cells by number of genes

Cells total	446
Cells after filtering	421
Genes total	14794
Genes after filtering	10338
Cutoff value UMIs	> 2114
Cutoff value genes	> 1262
Significant PCs	5
Resolution Clustering	0.8
Estimated Clusters	7
p-value Markergenes	0.01
log2 fold change Markergenes	0.25

Markergenes (max. 3): Mgp Sparcl1 Col3a1

Limits FeaturePlot Legend: ● Absolute ○ Relative

Display FeaturePlots as...: ● t-SNE ○ UMAP

WASP key features

- Automated processing of high-throughput scRNA-seq data
- Interactive browsing of results with high-quality graphics
- Easy usability, reproducible results and a fast analysis
- User-provided reference genome, annotation and FASTQ file(s)
- … or directly start with a gene expression matrix
- Local installation possible, Docker container planned

References

[a] Macosko EZ., *et al.*, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets, Cell 2015

Funding

DFG Deutsche Forschungsgemeinschaft de.NBI

DFG KFO 309 | de.NBI network BMBF Grant FKZ 031A53

Contact

www.computational.bio/software/wasp
andreas.hoek-2@computational.bio.uni-giessen.de