SC1: interactive web-based single cell RNA-seq analysis

ACM-BCB 2020 conference
Sep. 21

Marmar Moussa

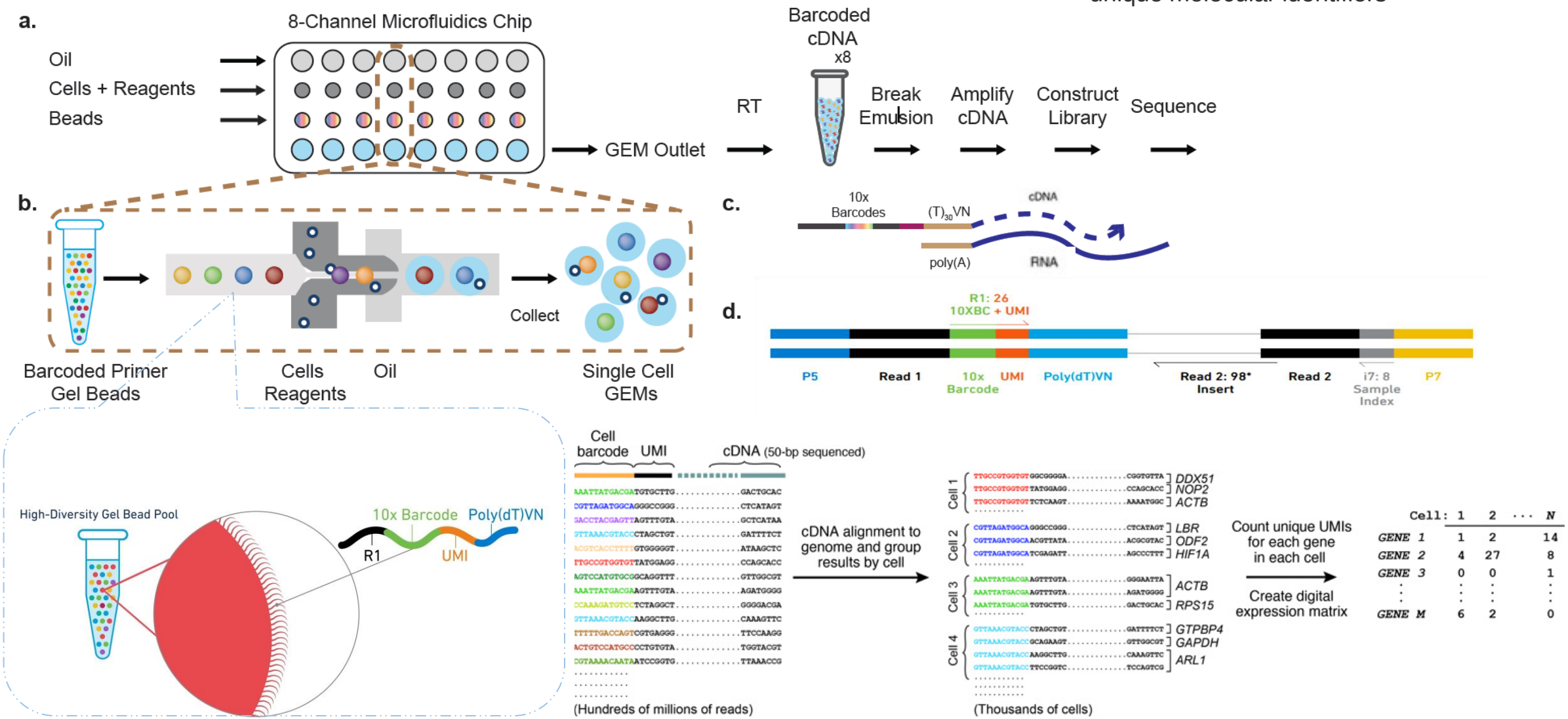
Carole and Ray Neag Comprehensive Cancer Center and
Center for Immunotherapy of Cancer and Infectious Diseases
UConn Health

Computer Science & Engineering Department
University of Connecticut

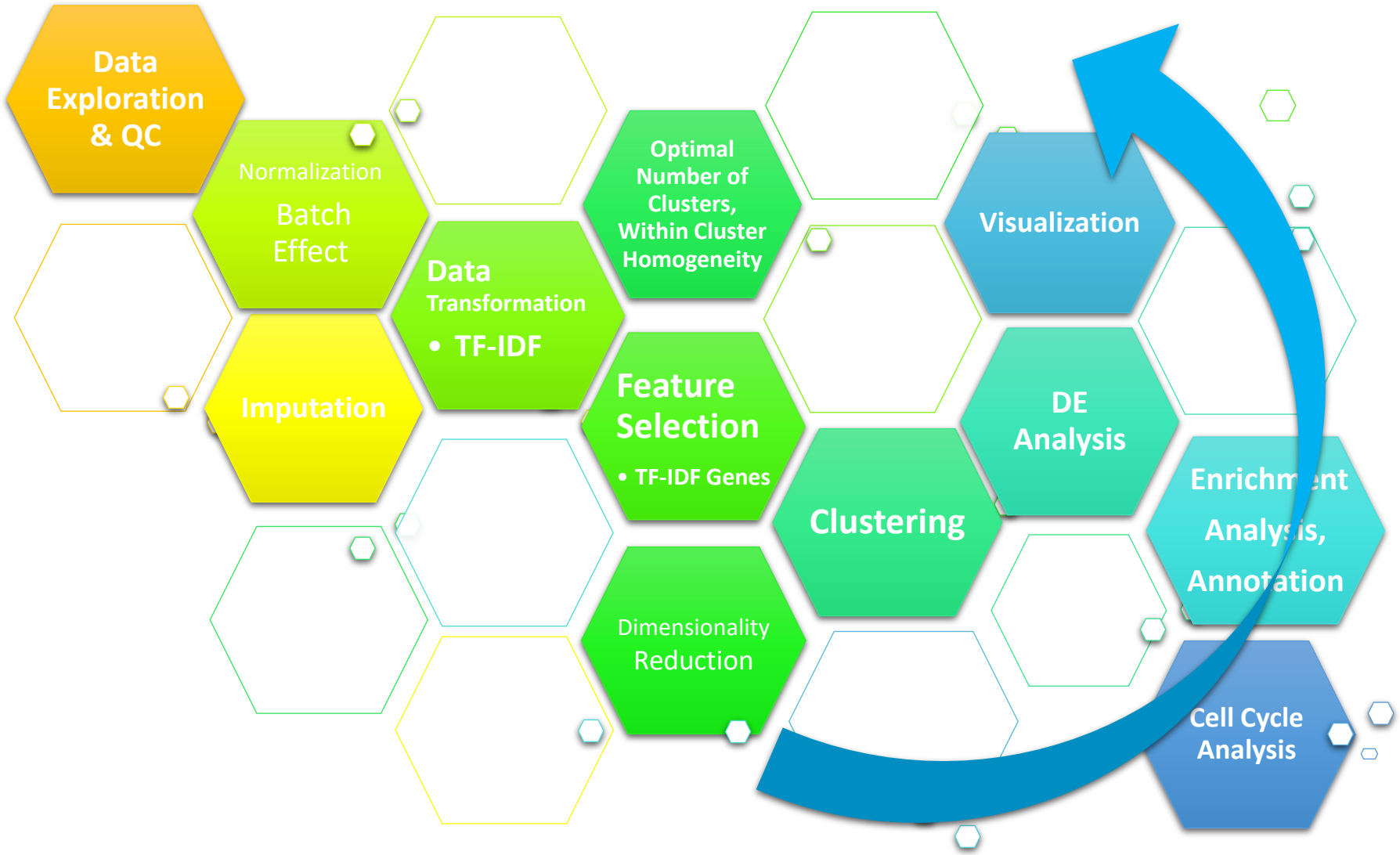
UCONN
HEALTH

UMI* Sequencing (10X Genomics)

*unique molecular identifiers



scRNA-Seq Analysis Workflow

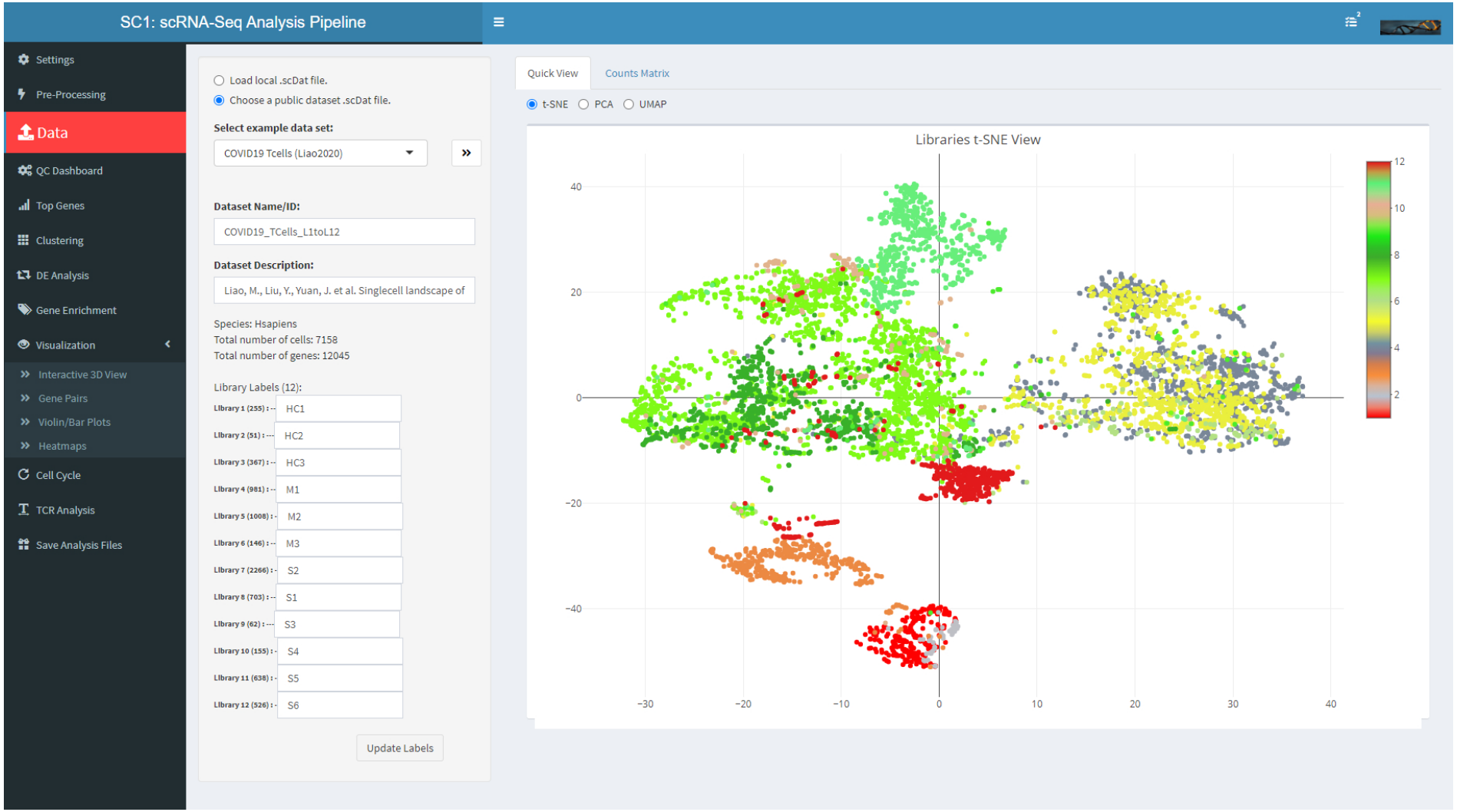


SC1 Tool

- An interactive web-based pipeline.
  - Only few tools without considerable coding efforts.
- Provide support for more detailed analysis & new methods:
  - LSImpute
  - TF-IDF based gene selection & clustering
  - Cell Cycle Analysis

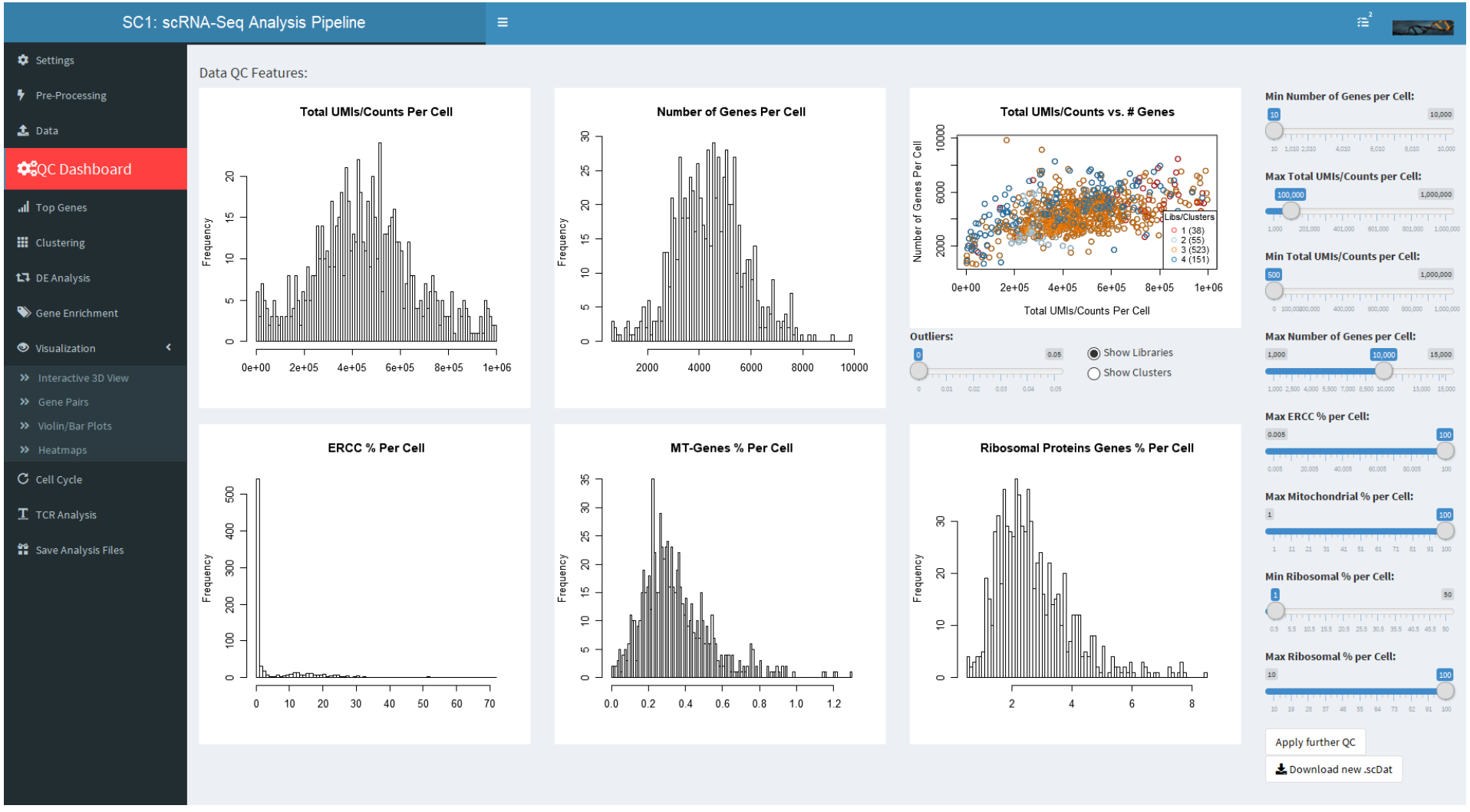
- SC1 public server: https://sc1.engr.uconn.edu/
- Mirror public server: https://cd8.engr.uconn.edu/
- Development server: https://sc1-dev.engr.uconn.edu/
- Older versions: https://sc1-prod.engr.uconn.edu/
- Support: https://sc1.zulip.engr.uconn.edu/

Data at a glance
(Example Dataset: COVID19 Tcells from Liao et al 2020)



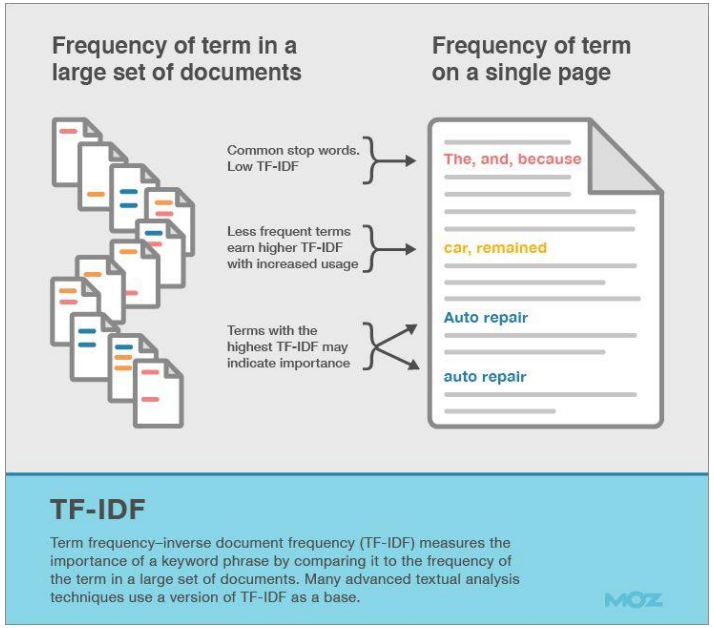
SC1 QC Dashboard

(Example Dataset: COVID19 Tcells from Liao et al 2020)



Gene Selection & Clustering Single Cell RNA-Seq Data using TF-IDF based Methods

- *Term frequency:* How frequently a term occurs in a document?
- *Inverse document frequency:* How uncommon the term is in the document collection?



- For scRNA-Seq data:
  - For gene i in cell j with count f_{ij} :

$$TF_{ij} = f_{ij} / \max_k f_{kj}$$

  - If gene i is detected in n_i out of N cells:

$$IDF_i = \log_2(N/n_i)$$

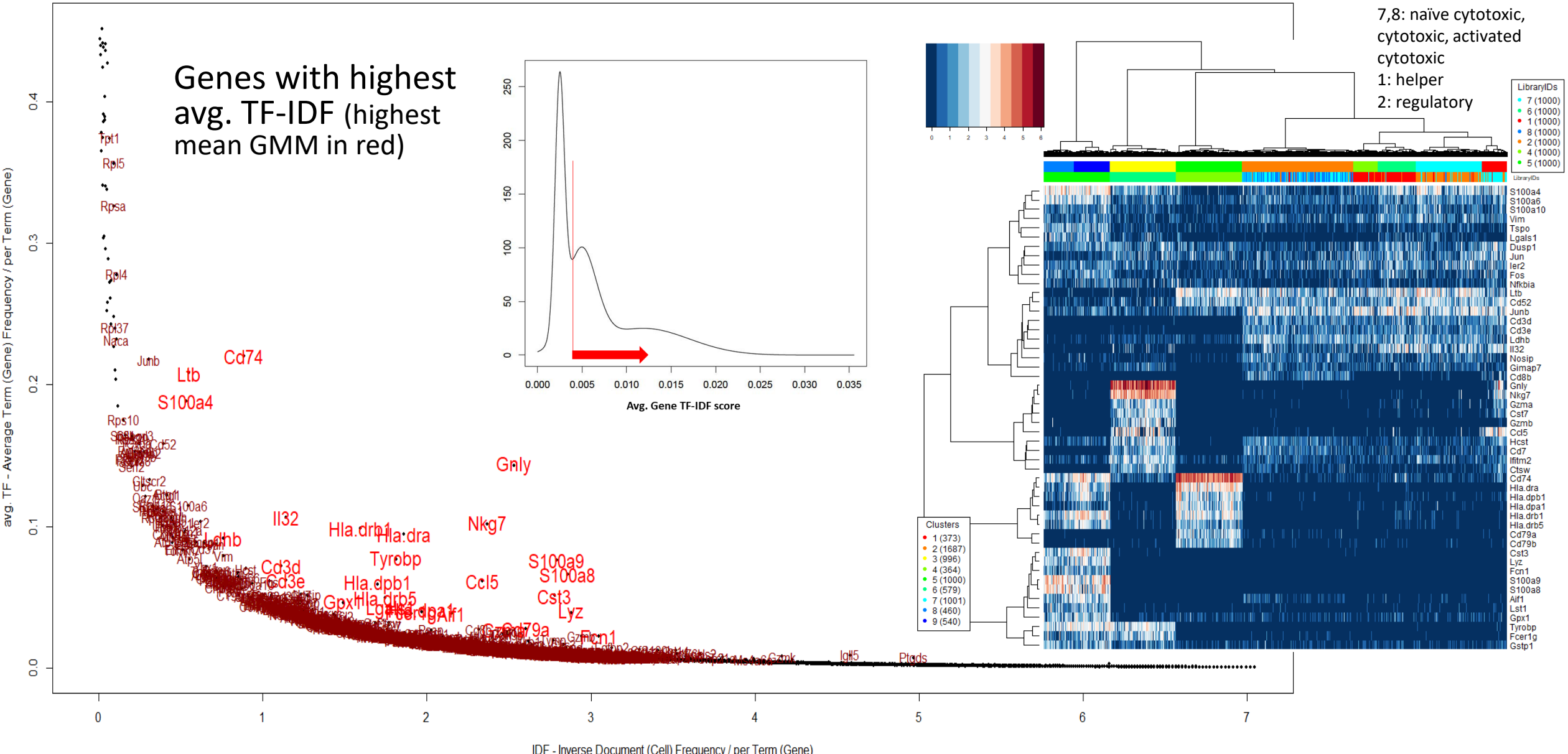
  - TF-IDF score:

$$TF_{ij} \times IDF_i$$

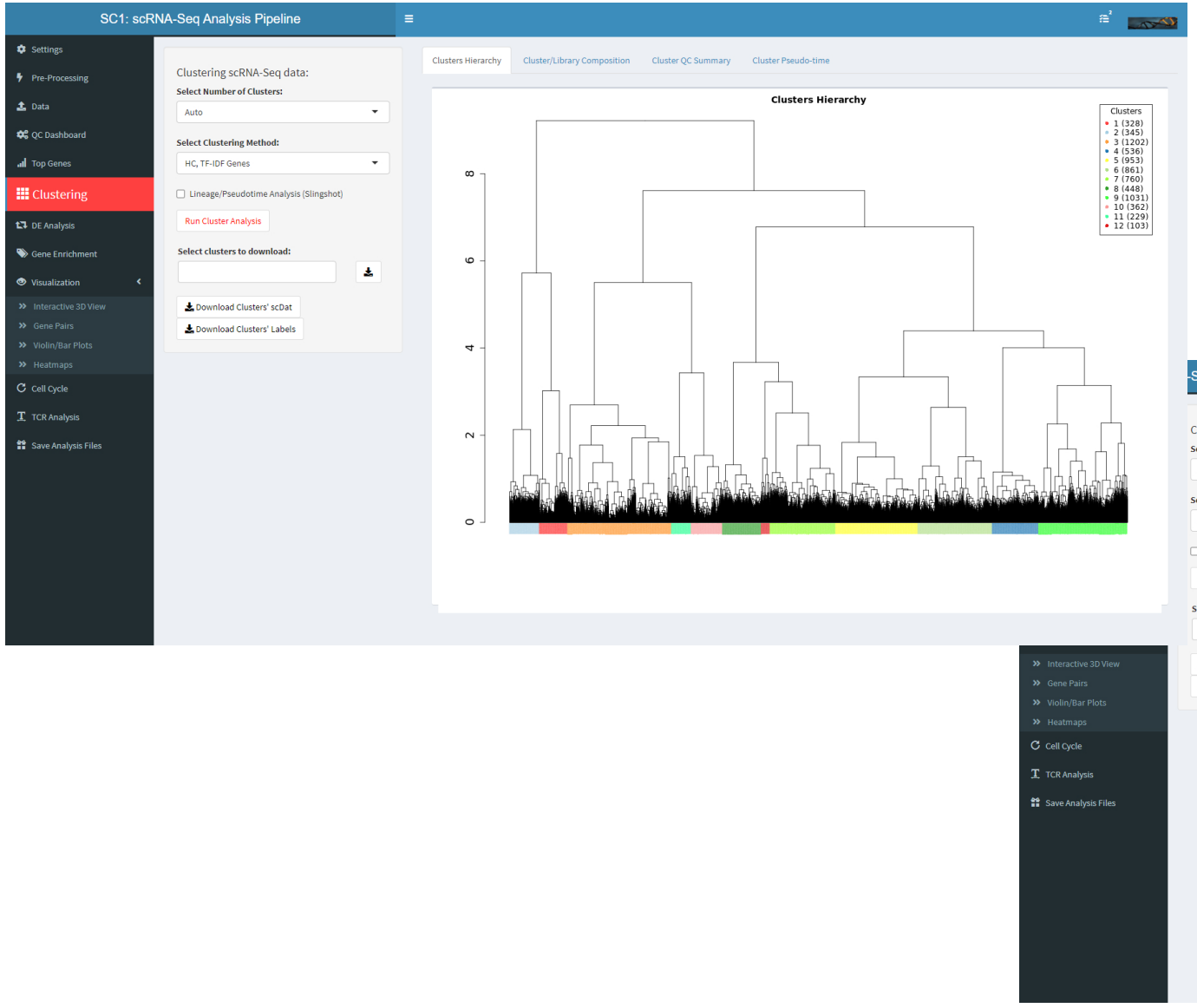
	ISG15	TNFRSF18	TNFRSF4	SDF4	UBE2J2	PUSL1	CPSF3L	GLTPD1	AURKAIP1	
GGAGCGCTTCCAAG-2	0	0.522461	0	0	0	0.8119	0	0	0	...
CAGATCGACCTCAC-1	0	0.180852	0	0	0	0	0	0	0.152962	
CGCTAAGAACCTGA-2	0	0	0	0	0	0	0	0.13172	0	
CCCATGTGGGGCAA-1	0	0	0.218698	0.162713	0.177152	0	0	0	0.116971	
GCCCAACTGGAGTG-1	0	0	0	0	0	0.315812	0	0	0	
AGTAATACGTTCAG-2	0	0	0	0.125732	0	0	0	0	0	
	...									

TF-IDF based Feature selection Example Dataset: PBMCs (10x)

5: Monocytes,
6: Natural Killer
4: B cells
7,8: naïve cytotoxic, cytotoxic, activated cytotoxic
1: helper
2: regulatory



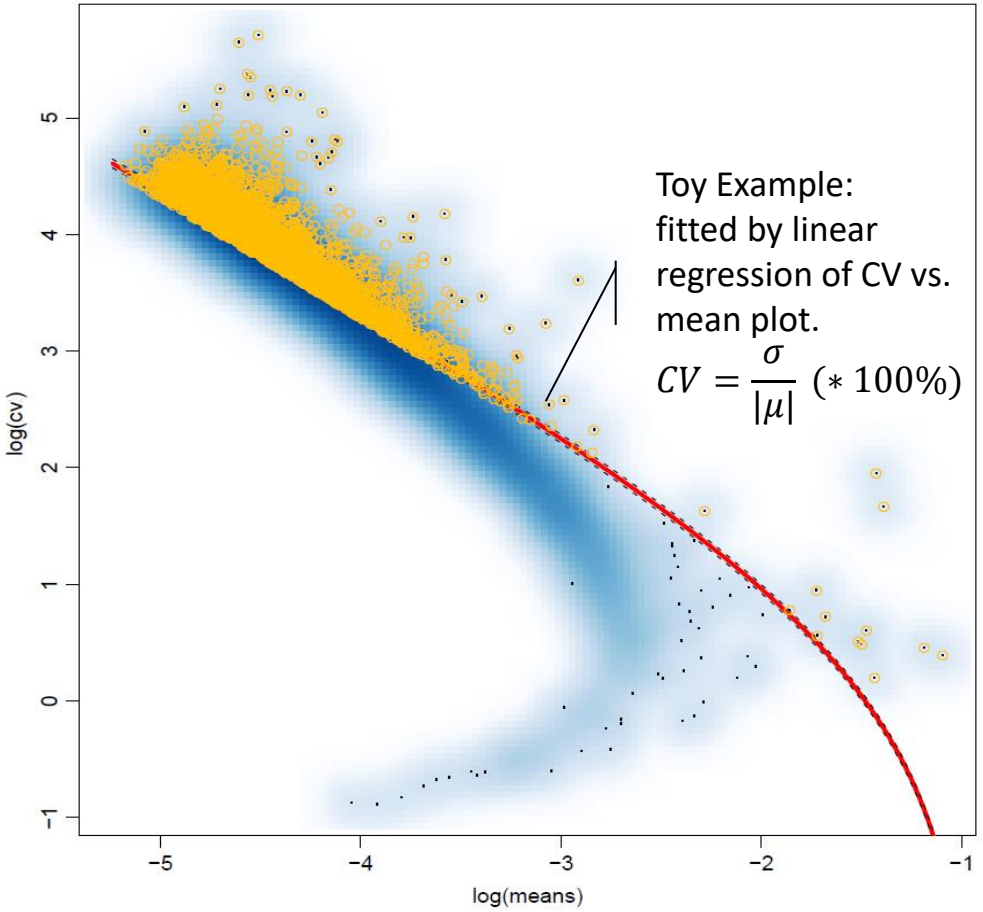
Clustering



(Example Dataset: COVID19 Tcells from Liao et al 2020)

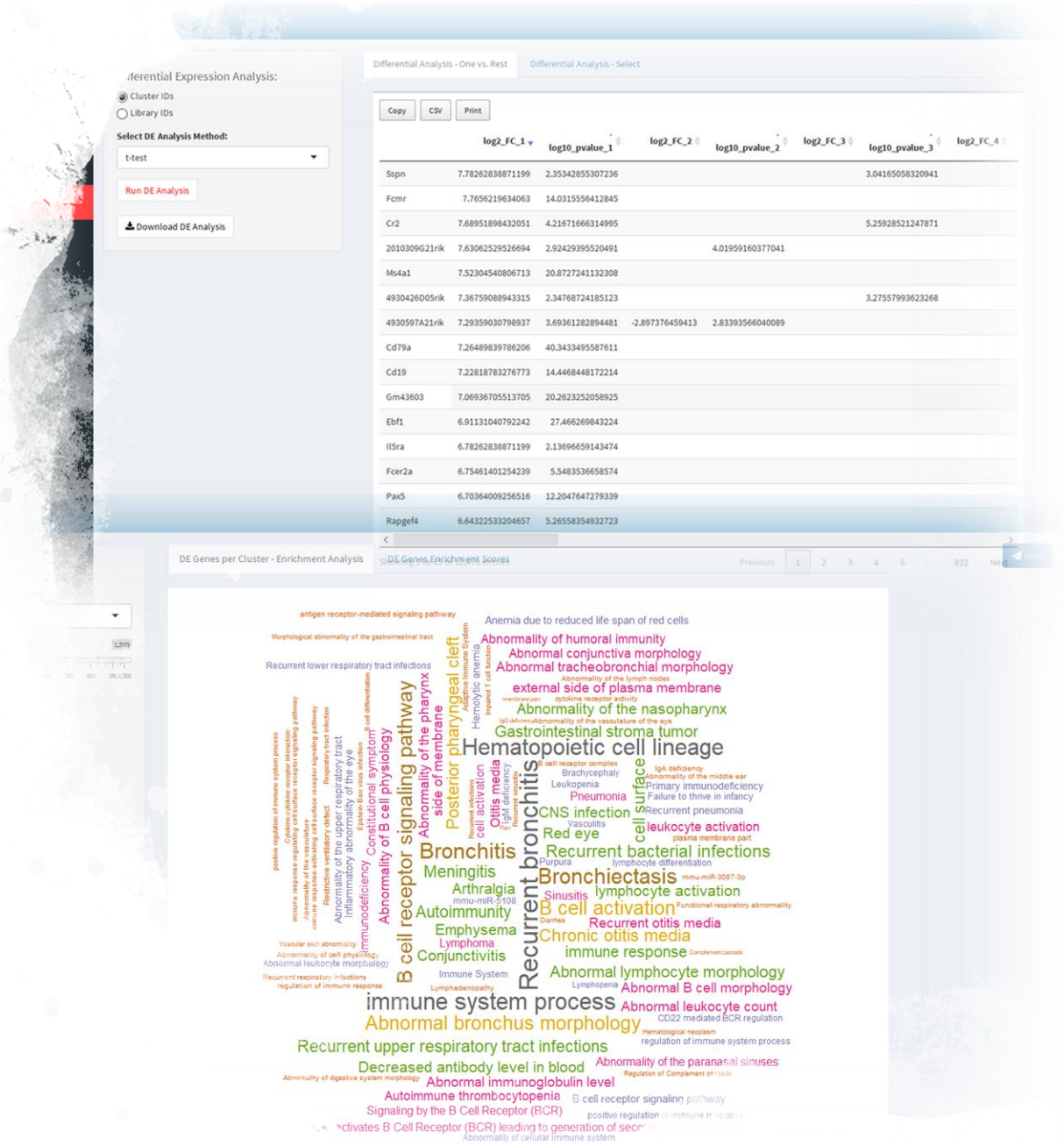
Other gene selection approaches

- genes with highest variability (Var) in values:
  - Variability decided by the relationship between the coefficient of variation (CV) and average expression levels.
  - CV (Dispersion) : ratio of the standard deviation to the mean. $CV = \frac{\sigma}{|\mu|}\ (* 100\%)$
  - Useful in comparison between data sets with different units or widely different means.
  - We pick the genes above the fitted line (fitted by linear regression) of CV vs. mean plot.



DE Analysis

- One Cluster / Library vs. Rest
- Simple t-test→ Fold Change, p-value
- Gene Enrichment Analysis using gProfiler2:
  - functional profiling of gene lists using various kinds of biological evidence

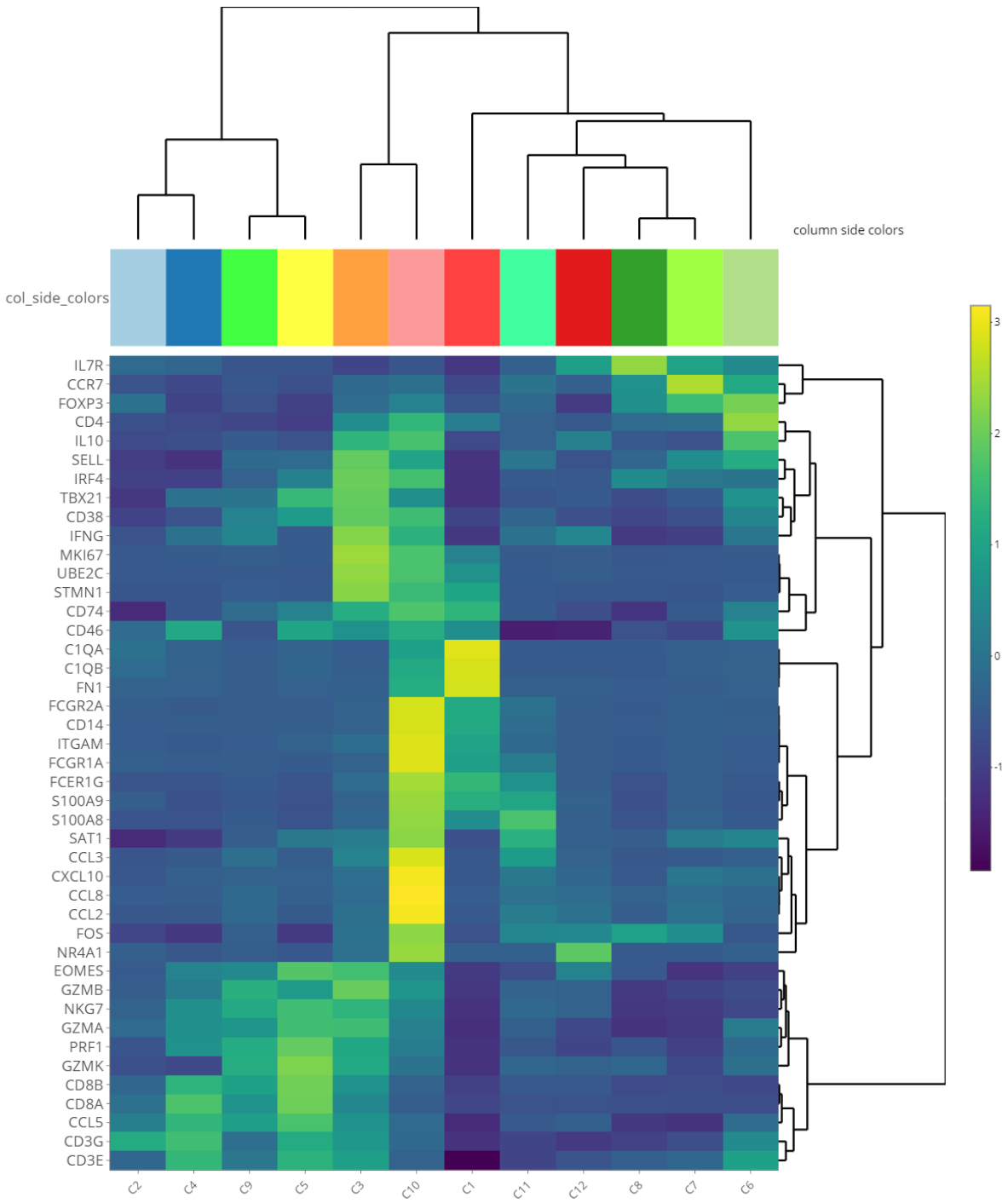


Visualization

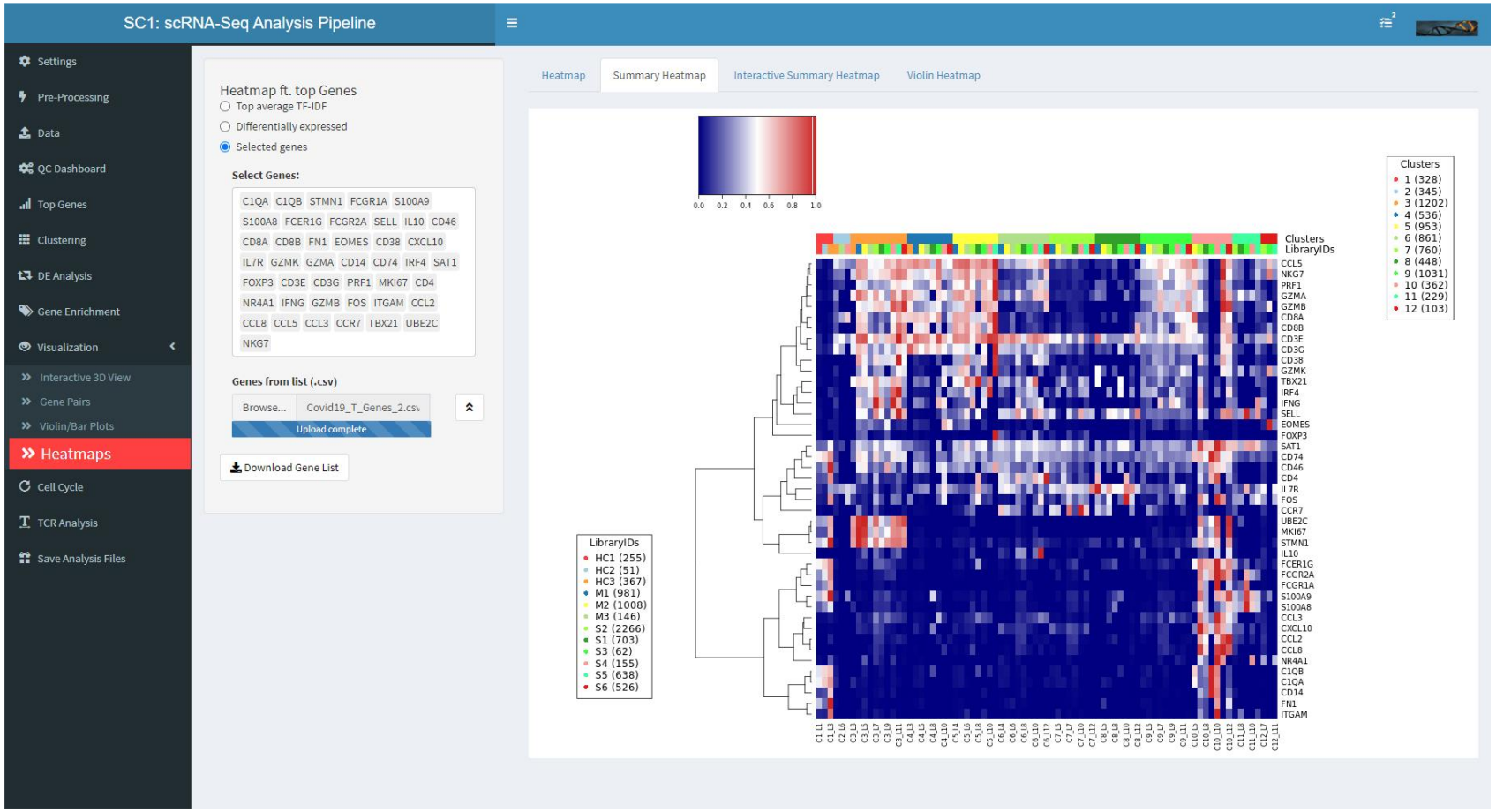
Visualization Packages used in SC1:

Heatmap3, scrattch.vis, plotly, ggplot2, Slingshot, …

(Example Dataset: COVID19 Tcells from Liao et al 2020)

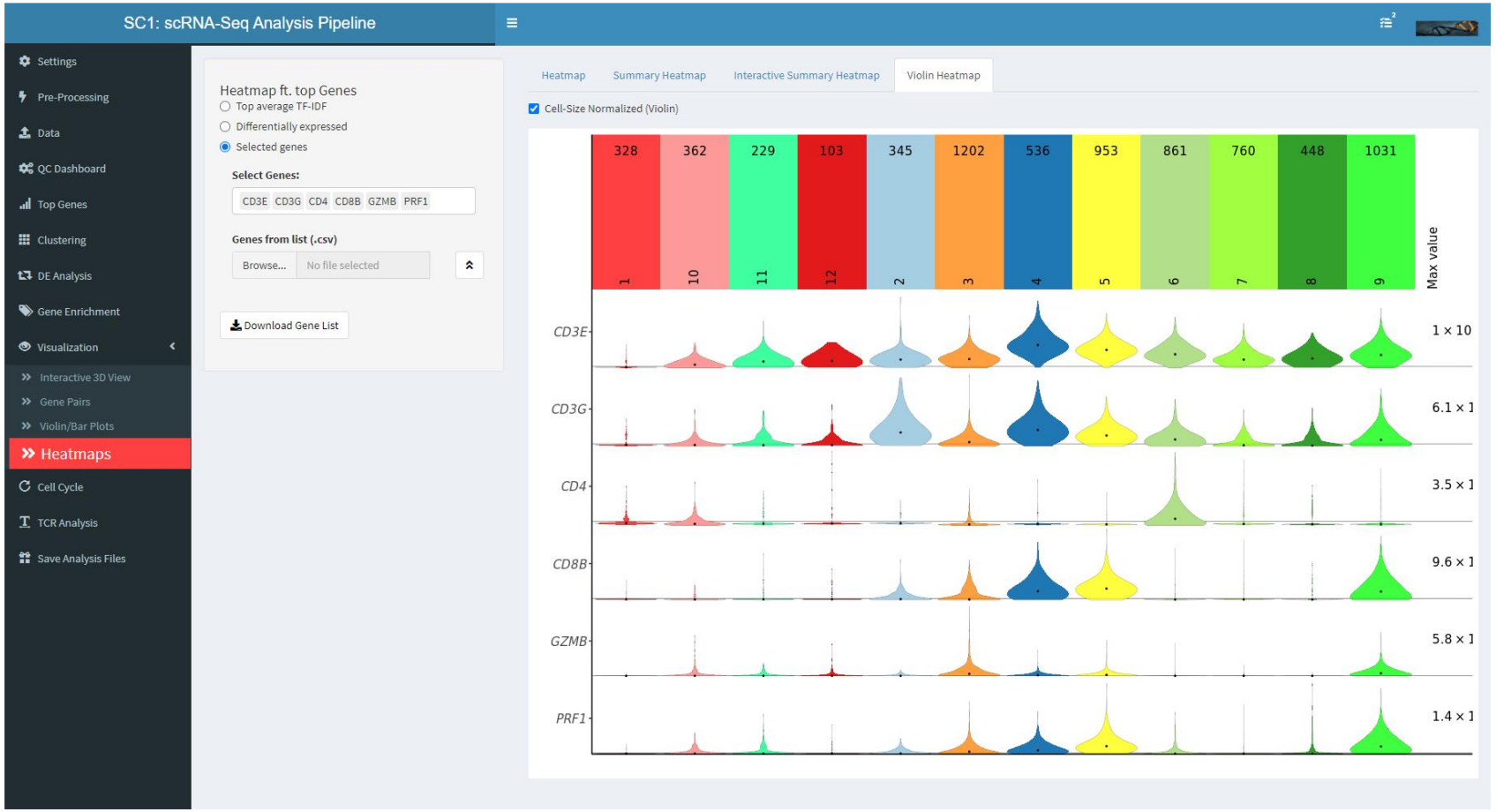


More Visualization Examples



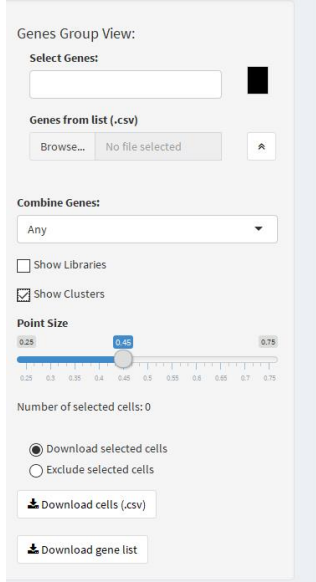
(Example Dataset: COVID19 Tcells from Liao et al 2020)

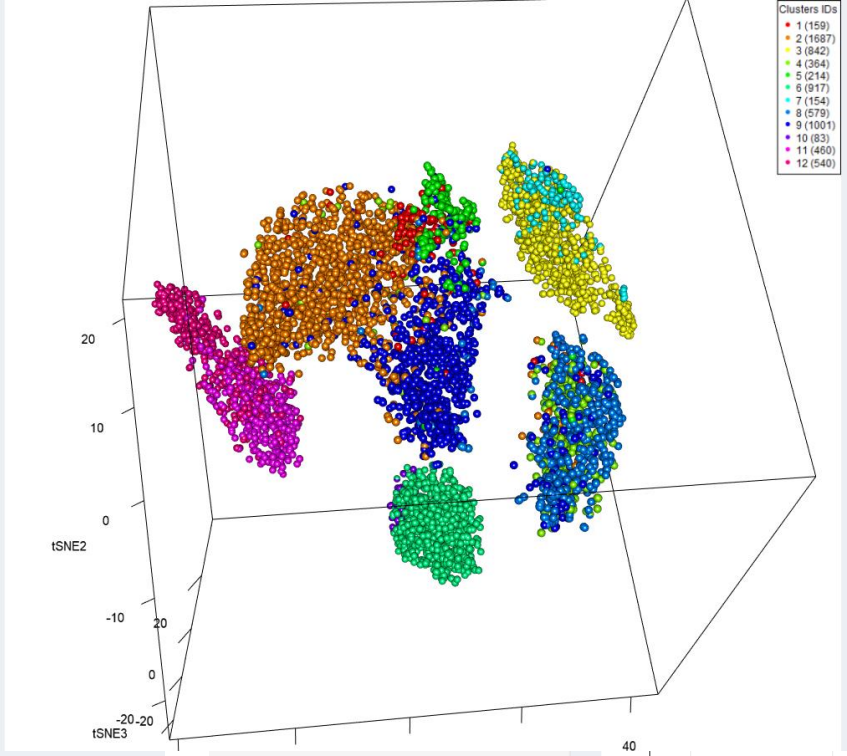
More Visualization Examples

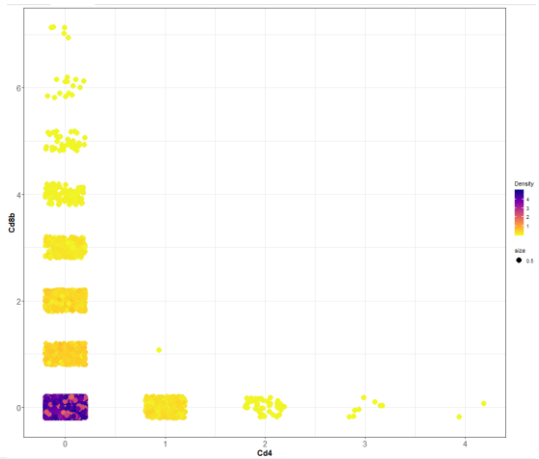


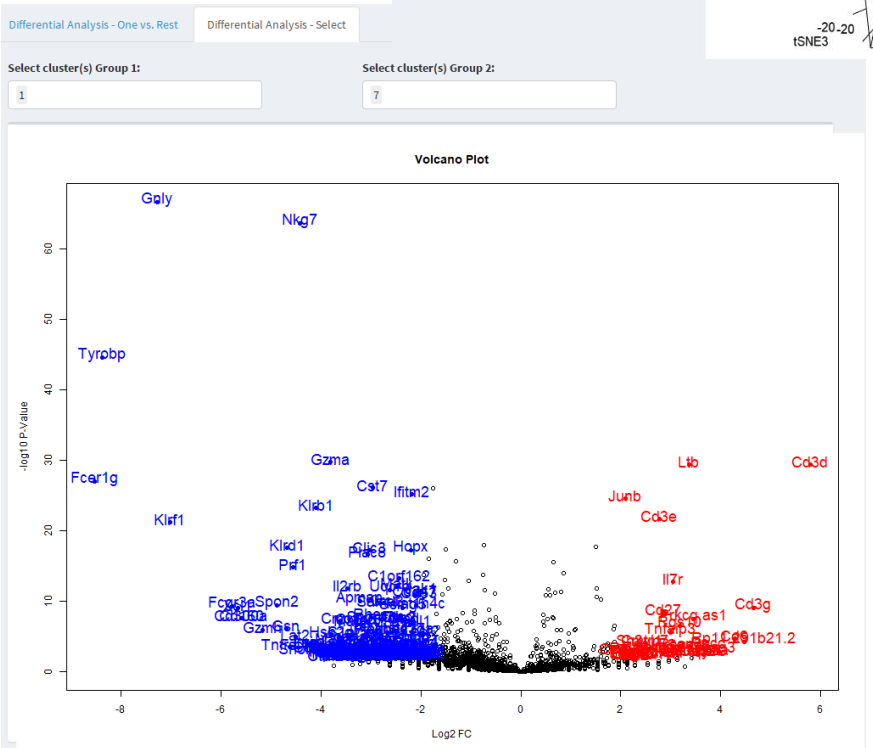
(Example Dataset: COVID19 Tcells from Liao et al 2020)

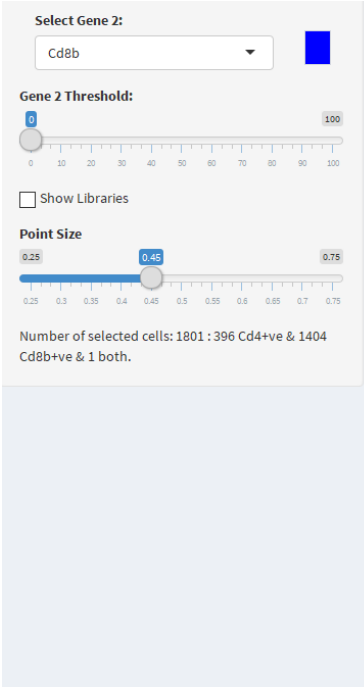
Visualization

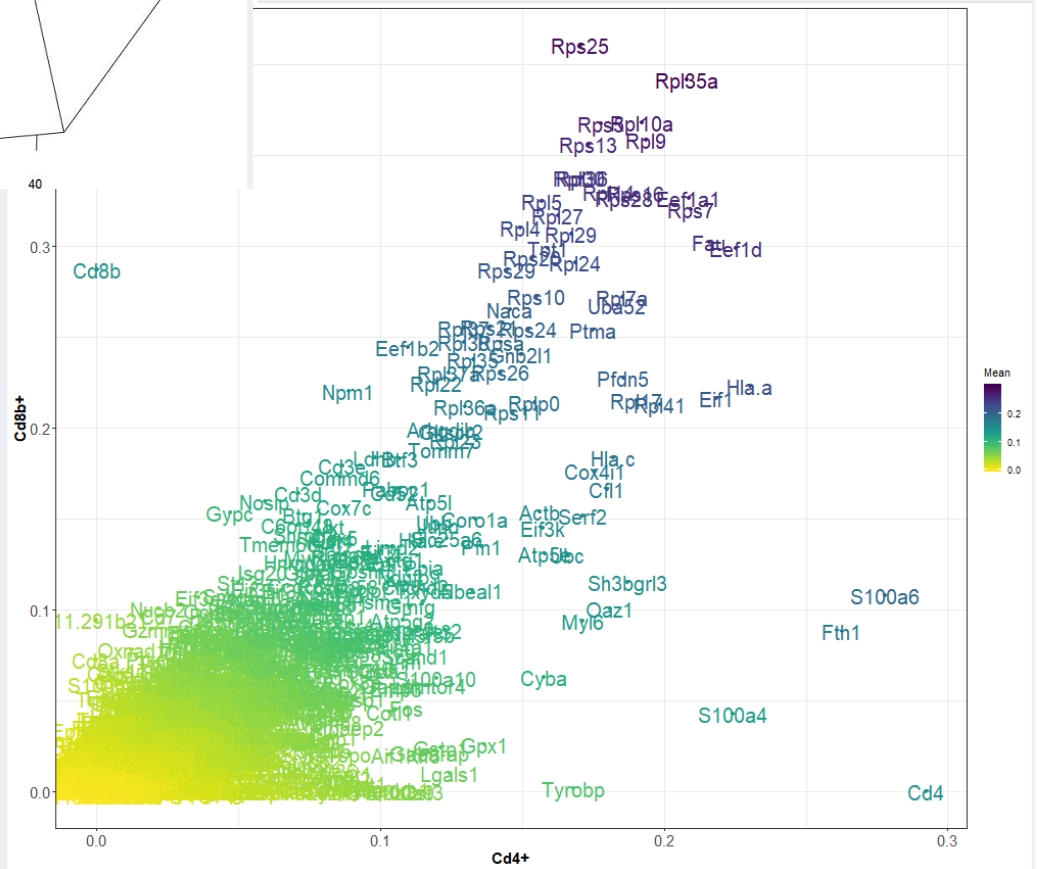




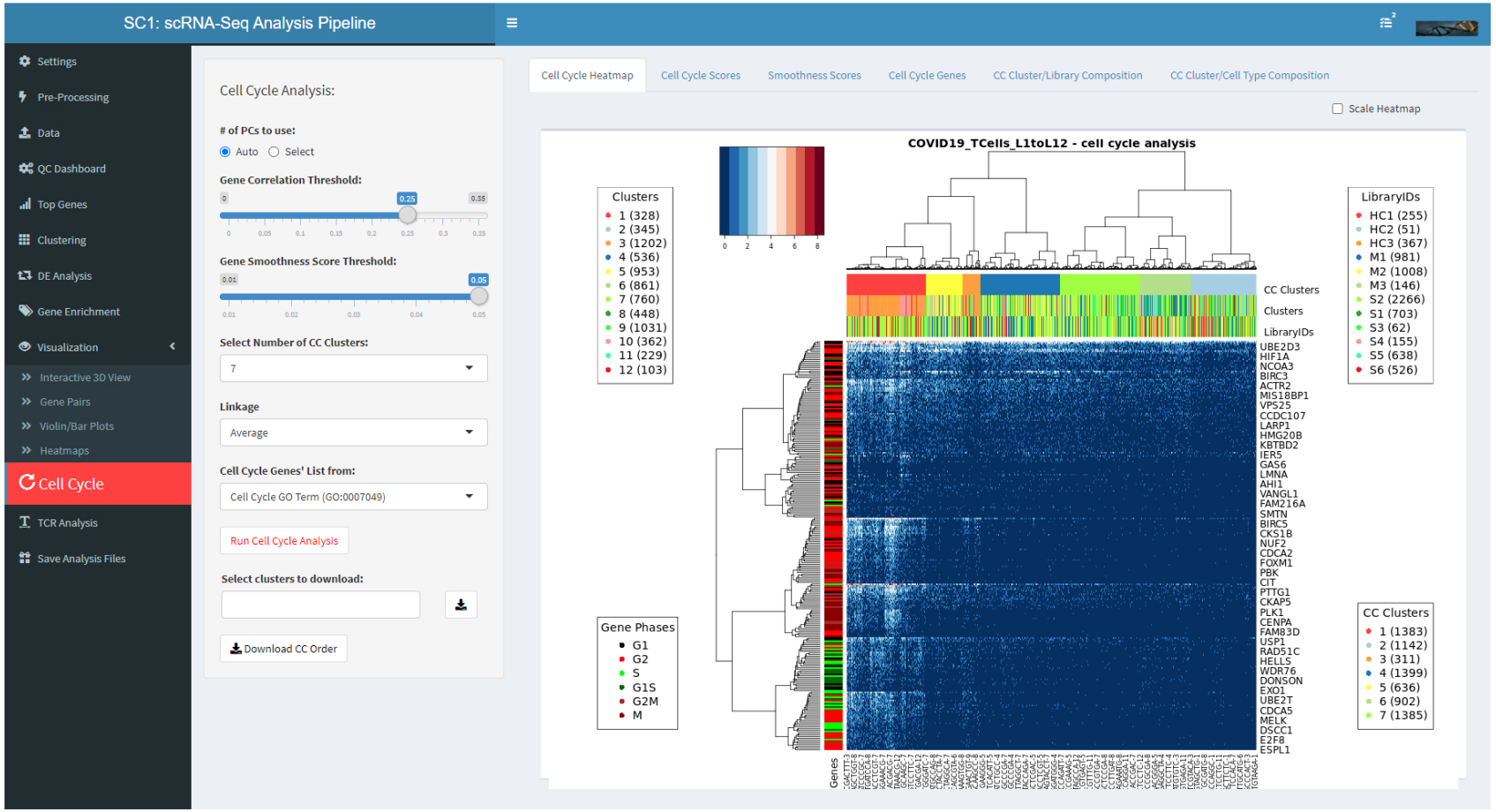






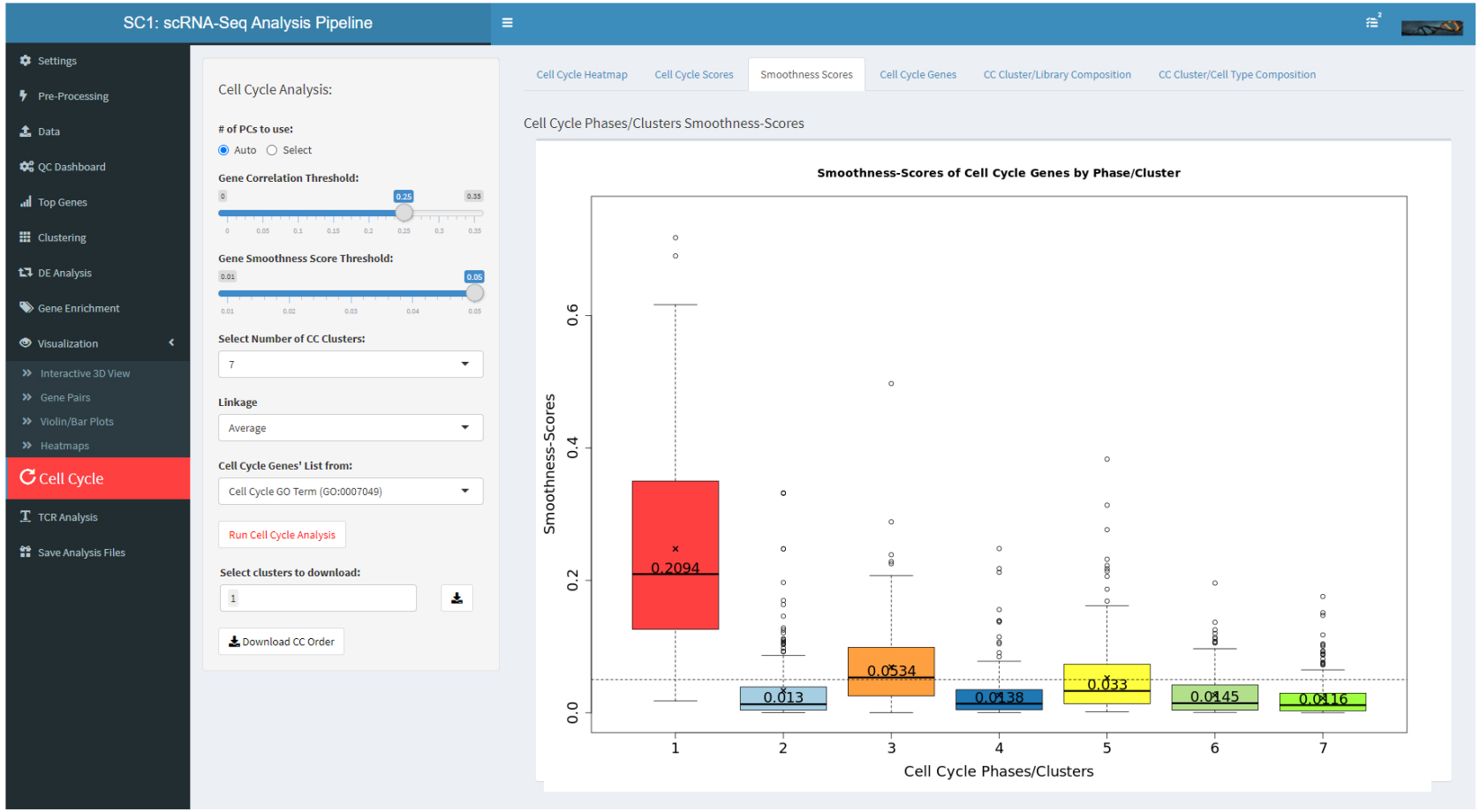


Cell Cycle Analysis



(Example Dataset: COVID19 Tcells from Liao et al 2020)

Cell Cycle Analysis



(Example Dataset: COVID19 Tcells from Liao et al 2020)

Cell Cycle Analysis

(Example Dataset: COVID19 Tcells from Liao et al 2020)

References

Moussa, M., Măndoiu, I. Single cell RNA-seq data clustering using TF-IDF based methods. *BMC Genomics* **19,** 569 (2018). https://doi.org/10.1186/s12864-018-4922-4

Moussa, M., & Măndoiu, I. I. (2020, December). SC1: A Tool for Interactive Web-Based Single Cell RNA-Seq Data Analysis. In *Bioinformatics Research and Applications* (pp. 389-397). Springer, Cham

Nevin, J. T., Moussa, M., Corwin, W. L., Mandoiu, I. I., & Srivastava, P. K. (2020). Sympathetic nervous tone limits the development of myeloid-derived suppressor cells. *Science Immunology*, *5*(51).

Example Datasets - References

- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., ... & Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, *19*(1), 477.
- Fletcher, R. B., Das, D., Gadye, L., Street, K. N., Baudhuin, A., Wagner, A., ... & Purdom, E. (2017). Deconstructing olfactory stem cell trajectories at single-cell resolution. *Cell stem cell*, *20*(6), 817-830.
- Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., ... & Liu, L. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nature medicine*, 1-3.
- https://support.10xgenomics.com/single-cell-gene-expression/datasets
- https://cf.10xgenomics.com/samples/cell-exp/4.0.0/Targeted_NGSC3_DI_PBMC_Immunology/Targeted_NGSC3_DI_PBMC_Immunology_filtered_feature_bc_matrix.tar.gz

Thank You!

- Questions?