UNCHAINING THE FULL POTENTIAL OF PUBLIC SINGLE-CELL RNA-SEQ DATA

Natalie Russi[1], Frank Staubli[1], Jana Sponarova[1], Pavel Honsa[2], Stefan Bleuler[1], and Philip Zimmermann[1]
1. NEBION AG, Hohlstrasse 515, 8048 Zurich, Switzerland 2. NEBION AG, odstepny zavod, Rimska 526/20, Prague, 12000, Czech Republic

Presenter and Contact Person:
natalie.russi@immunai.com

Selected Study

- GEO accession: GSE145926
- Liao M et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. **Nat Med.** 2020 Jun;26(6):842-844

Aggregation

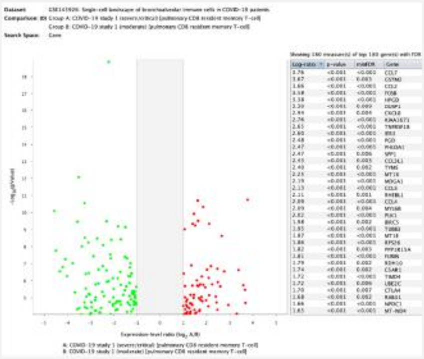
- After QC, 58'534 high-quality cells were retained
- Subsequent aggregation into 291 cell aggregates (common cell type/state/sample status)
- 54 cell-type level comparisons pre-defined by NEBION curators

Analysis Overview

- Differential expression analysis of exhausted pulmonary CD8$^+$ resident memory T-cells of patients with severe vs. moderate COVID-19
- Comparison with public bulk tissue data
- Clustering of aggregates for top regulated genes
- Gene Set Enrichment Analysis

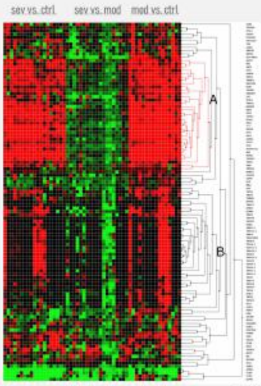
Identification of Cell-Type Specific Responses to COVID-19 using GENEVESTIGATOR®

Differential Expression



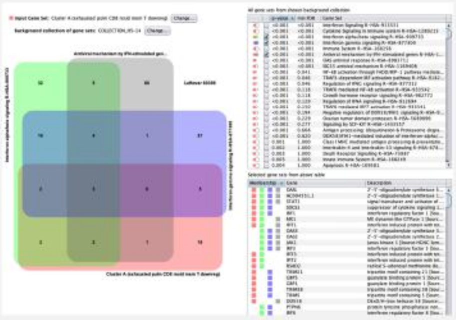
FDR of 0.01 and log-2 ratio of 1 (2-fold change)

Clustering of Aggregates



Cluster A: genes upregulated in almost all cell types in response to COVID-19
Cluster B: genes upregulated in CD4 and CD8 T-cells in response to COVID-19

Gene Set Enrichment Analysis



Processing and Curation Pipeline

- Data processing:
  - Read mapping
  - Doublet removal
  - Stringent quality control
  - Batch effect removal
- Clustering and cell-type identification:
  - UMAP visualization
  - Cluster description and evaluation
  - Sub-clustering of all clusters containing distinct cell types or cell states
- Biocuration of study and sample characteristics:
  - Cluster-level and sample-level annotation, using proprietary anatomy and cell type ontology with controlled vocabularies
  - Harmonization of cell-type annotations
  - Peer review

raw scRNA-Seq data → read mapping and expression quantification → quality control and doublets removal → normalization → batch effect removal and clustering → marker genes and cluster identification → cell-type annotation → metadata annotation using controlled vocabularies → peer-review process → data delivery

Figure 1. Pipeline – Overview of the single-cell RNA-Seq processing and curation pipeline at NEBION.

Anatomical Part	cluster positive biomarker	cluster positive biomarker	cluster positive biomarker	cluster positive biomarker	samples
cortical astrocyte	Foxj1	Myb	Fhl1	—	2
cortical astrocyte	Gfap	—	—	—	52
cortical astrocyte	Mfge8	—	—	—	25
cortical astrocyte	—	—	—	—	54
dentate gyrus astrocyte	Gfap	Hes5	Sox9	Aqp4	326
hippocampal astrocyte	Foxj1	Myb	Fhl1	—	18
hippocampal astrocyte	Gfap	—	—	—	16
hippocampal astrocyte	Mfge8	—	—	—	36
hippocampal astrocyte	—	—	—	—	11
hypothalamus astrocyte	Fabp7	Ntsr2	—	—	267
striatum astrocyte	Aldoc	Gja1	Slc1a2	—	63
striatum astrocyte	Gjb6	Sox9	—	—	107
subventricular astrocyte	—	—	—	—	34

Figure 2. Cluster description – Cell-level annotation of the most significantly and differentially expressed genes in each cell cluster. Marker genes are derived from literature and directly evaluated in GENEVESTIGATOR® before assignment to samples.

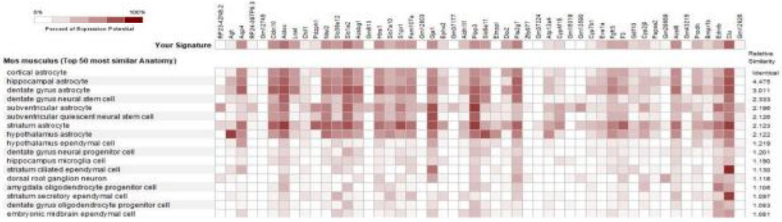


Figure 3. Cluster evaluation – each newly introduced cluster is compared to the current content in GENEVESTIGATOR®. Using the *Signature* tool, it is possible to evaluate the correct cluster assignment.

Results and Conclusions

- NEBION's data processing, cell-type identification, aggregation, and pre-defined comparisons allow fast identification of COVID-19 specific cell-type responses.
- Comparison to thousands of curated public studies shows inflammatory responses similar to a handful of other diseases.
- Gene set enrichment analysis shows strong up-regulation of human anti-viral mechanisms in diseased patients.