



Photo credits: ©



*Leveraging OmicSoft Array Suite to Share Single Cell RNA-Seq Data with Colleagues & Collaborators*

Michael R. Dufault

09.18.2019

# *Agenda*

---

**1. Introduction**

**2. Issue: Need to quickly & easily share single cell data**

**3. Solution: Array Studio Suite (Array Studio & Array Viewer)**

**4. Example**

# Introduction

---

- Single cell RNA-Seq studies have quickly become a significant proportion of the RNASeq analysis performed by the Bioinformatics department.
- A majority of the studies are collaborations between the Bioinformatics department and various therapeutic areas.
- Due to the iterative nature of scRNA-Seq analysis, sharing data with collaborators can be an important step in the analytical process.
- Our current data analysis pipelines use a mix of commercial applications and open source/internal packages which were not designed with sharing in mind.

... How can the data be shared and communicated efficiently?

# *Desired Attributes*

---

- We needed a mechanism to provide access to the single cell data with collaborators in real-time
- Platform has to be accessible to everyone with the company
- Ideally, it would be in an user friendly environment to the bench scientists

# *OmicSoft Array Suite*

---

## Array Studio + Array Viewer

- Corporate License
  - Infrastructure in place
  - Company-wide accessibility
- Environment that both our bioinformaticians & biologists have been exposed to
  - Little/no learning curve (compared to other analytical tools)
- Omicsoft has a staff that is willing to help develop the tools and views needed to share single cell data

# Array Viewer

---

- **What is it?**

- Array Viewer is OmicSoft's lightweight software for viewing shared projects and data without having to install Array Studio.

- **Why use it?**

- It's an interactive application that provides researchers the ability to share predefined visualizations enabling colleagues & collaborators to view and interrogate the data in real-time.

# Default Collection of Views to Share Single Cell Data

## Metadata

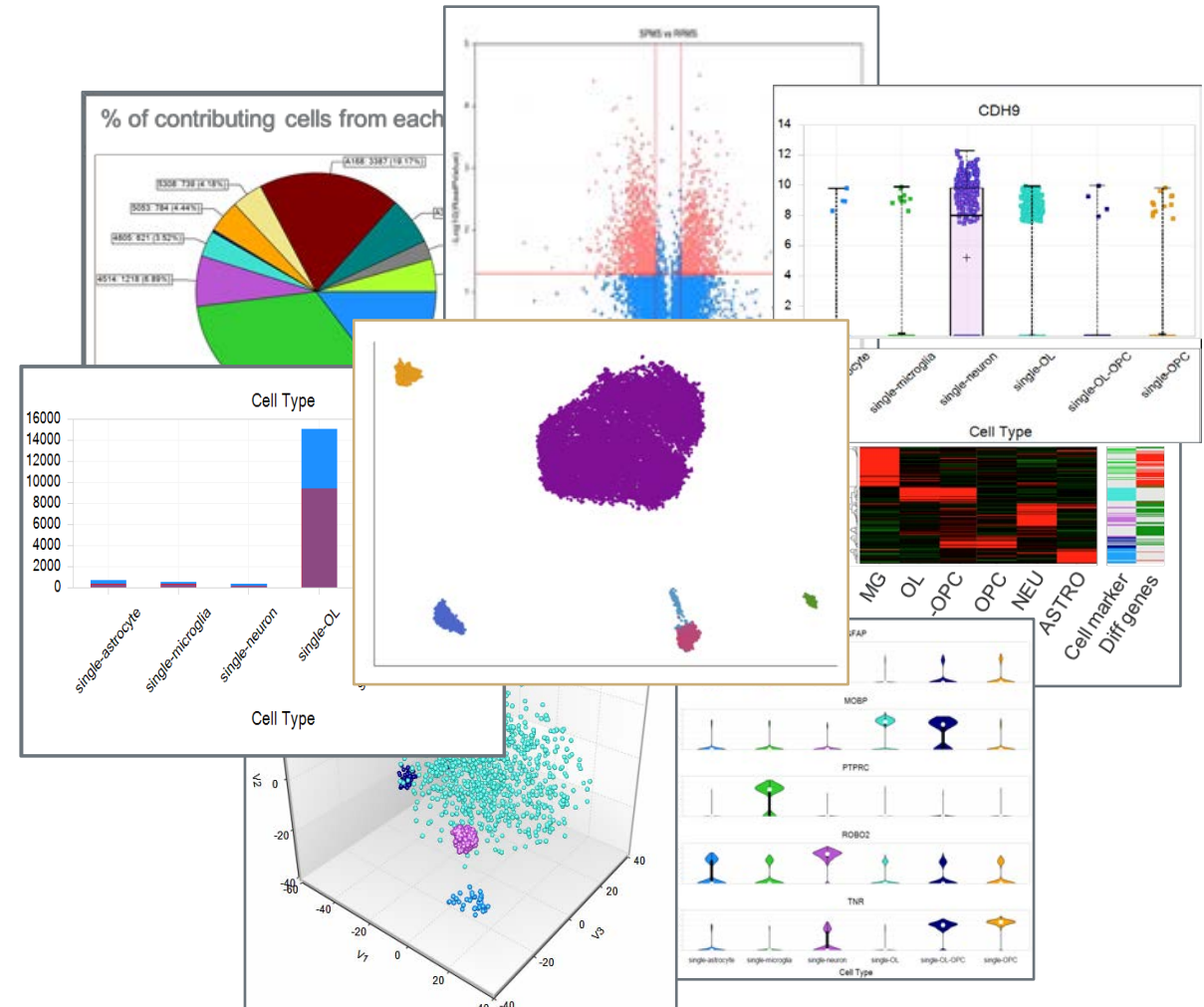
- Sample level
- Cell level

## Expression matrix

- Bar graph
- Violin Plot

## UMAP/tSNE

- Sample level
- Cell level



# *Example: snRNA-Seq Data from MS Study*

---

- **Goal:** Identify novel mechanisms underlying MS progression and enable the discovery of cell-specific targets for progressive MS.
- **Experimental Approach:** Compare SPMS vs. Control

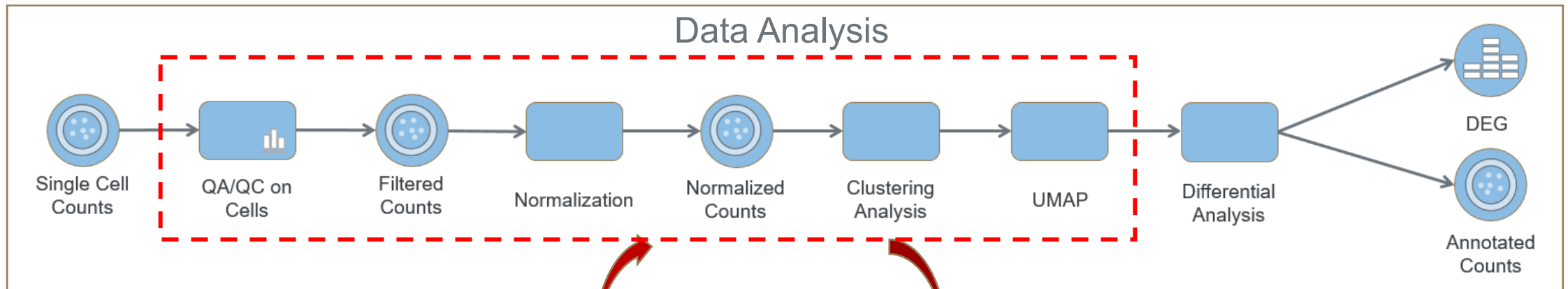
Group	Samples
Control	6
SPMS	5

- **Important steps**
  - Identification of the different cell populations
  - Determine the expression profile for each cell



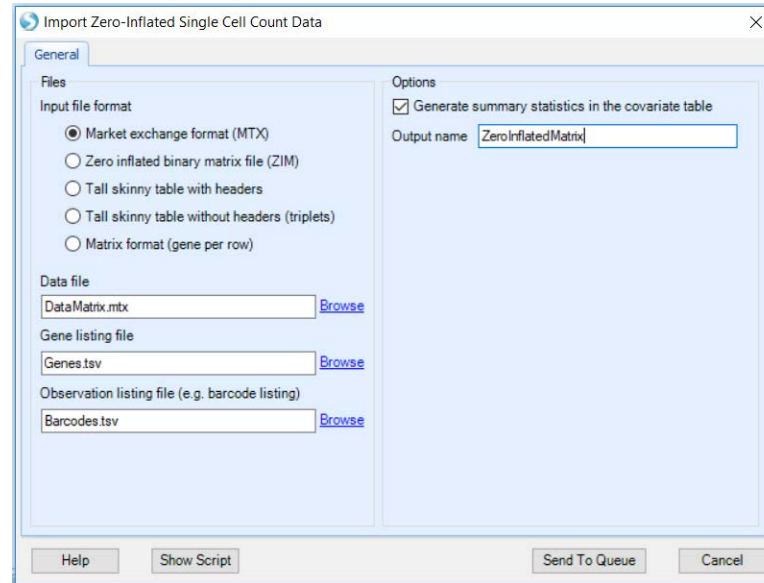
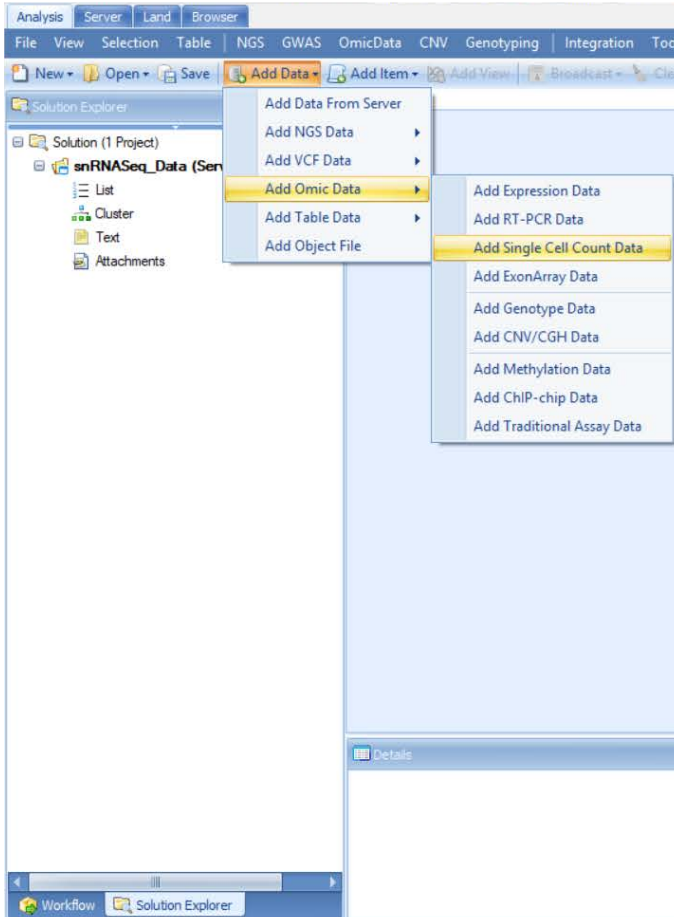
# snRNA-Seq Workflow

## Sample/library prep & sequencing



Share Data  
Array Studio + Array Viewer

# Import Zero Inflated Matrix



The screenshot shows a data table with 21896 rows and 23382 columns. The table has the following structure:

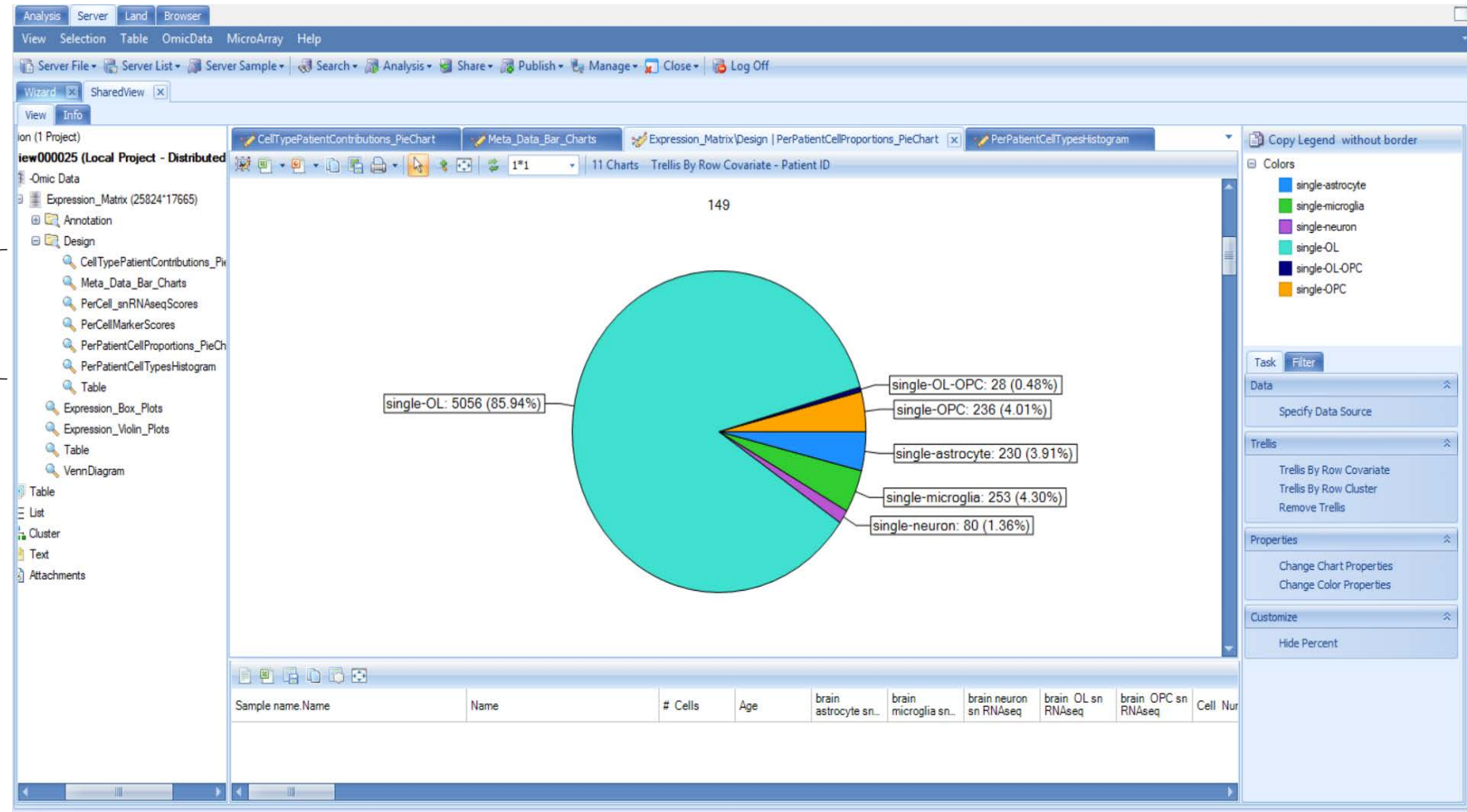
ID	CGAACATTC TCACATT-1 SN1 P04 NN	CGAATGTAG AGTCTGG-1 SN1 P04 NN	CGAATGTAG CCGTCGT-1 SN1 P04 NN	CGAATGTAG CGATAGC-1 SN1 P04 NN	CGAATGTCA TCACGTA-1 SN1 P04 NN
AADACL2-AS1	0	0	0	0	0
AAMDC	0	0	0	0	0
AAMP	0	0	0	0	0
AASDHPPT	2.32008000...	0	0	0	0
ABCB1	0	0	0	0	0
ABCC1	0	1.89889000...	0	0	0
ABCC4	0	0	0	0	0
ABCC5	2.32008000...	0	0	0	0
ABCD3	0	0	0	0	2.70764000...
ABHD3	0	0	0	3.01878000...	0
ABHD18	2.32008000...	0	2.72846000...	0	2.70764000...
ABL1	2.32008000...	0	0	0	0
ABR	0	0	0	3.92691000...	0

# Metadata Views

Derived from the ZIM design table

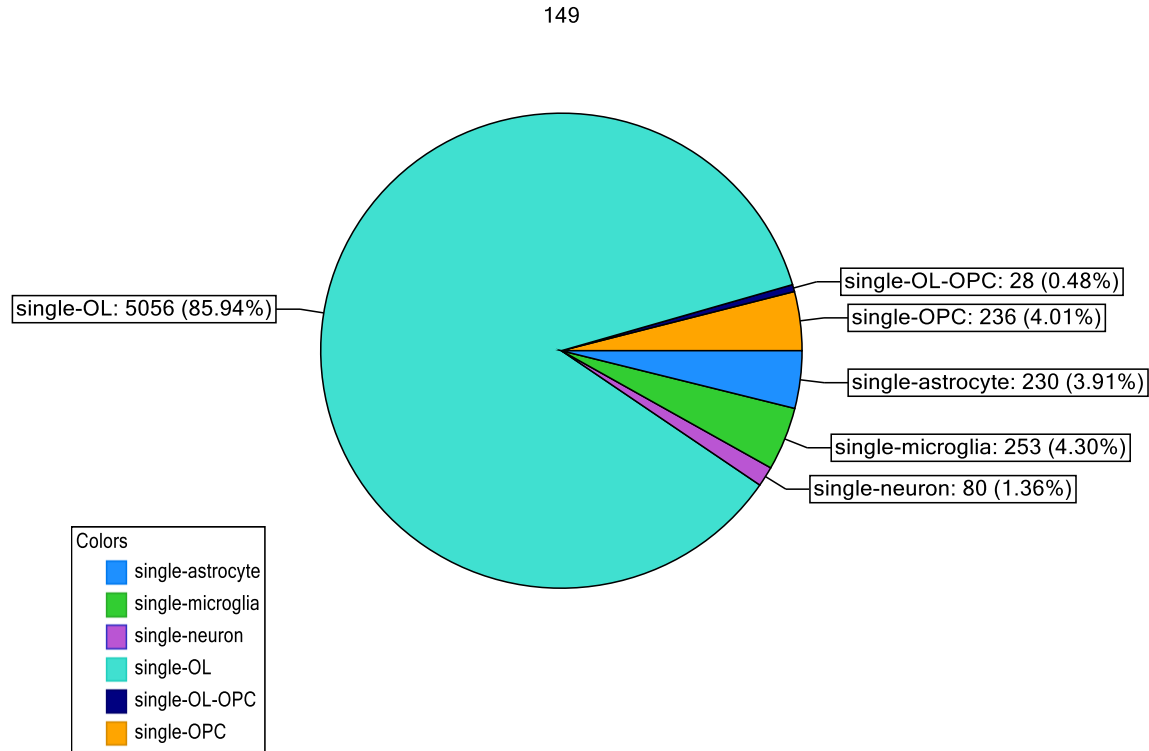
## Metadata

- Sample level
- Cell level

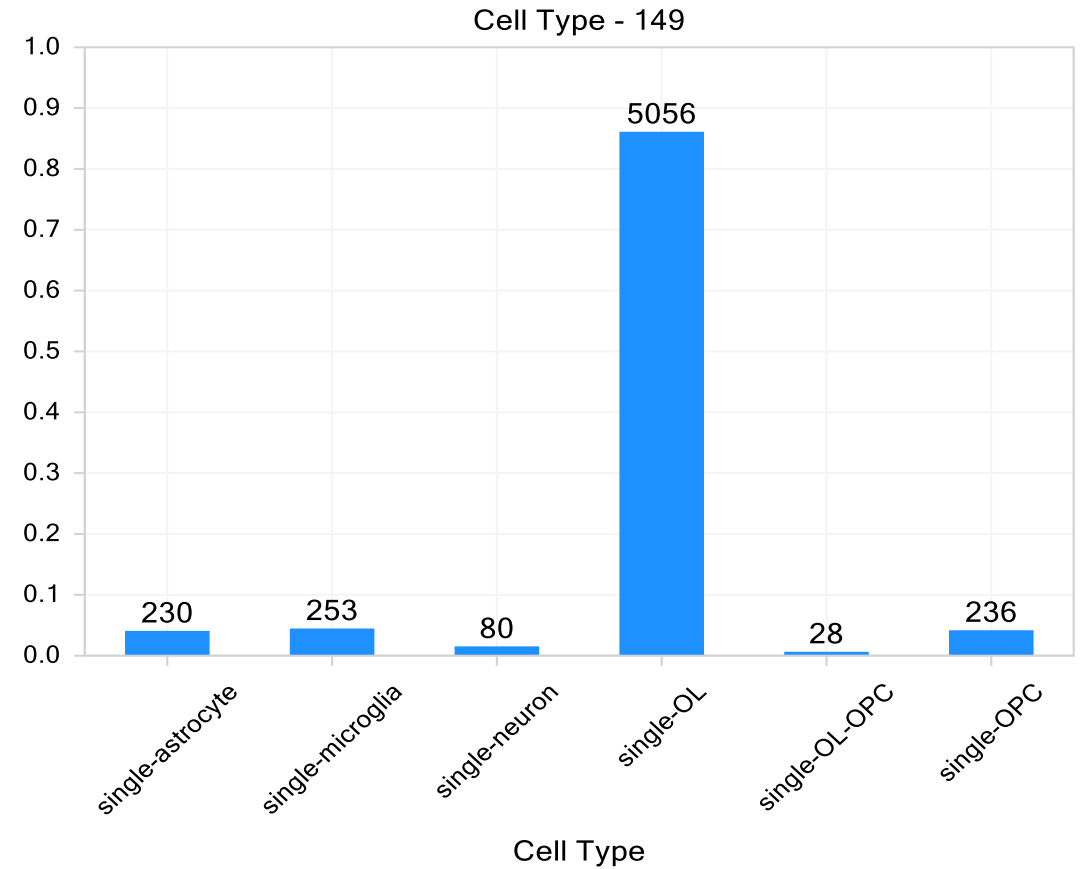


# Per Patient Cell Information

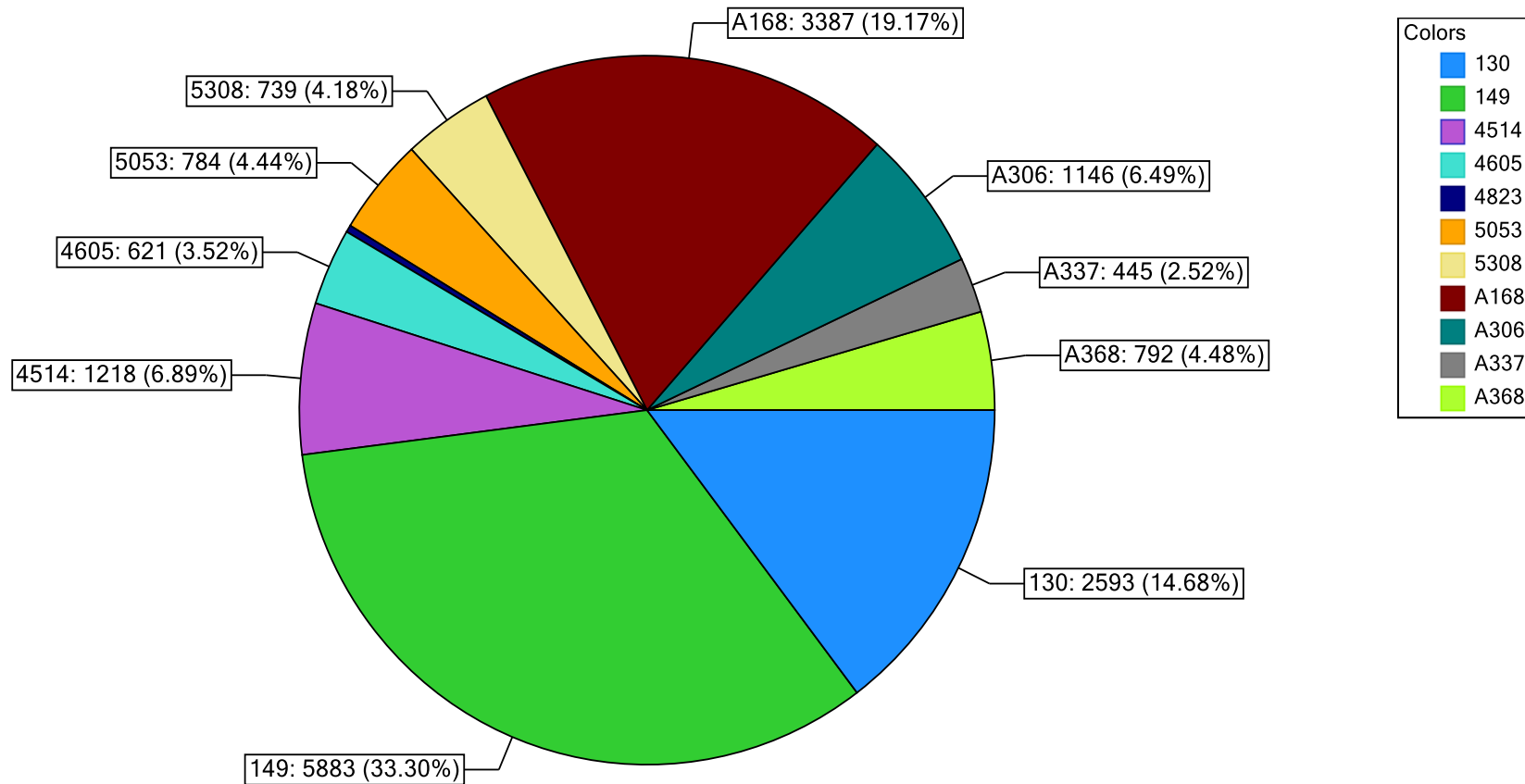
Percentage



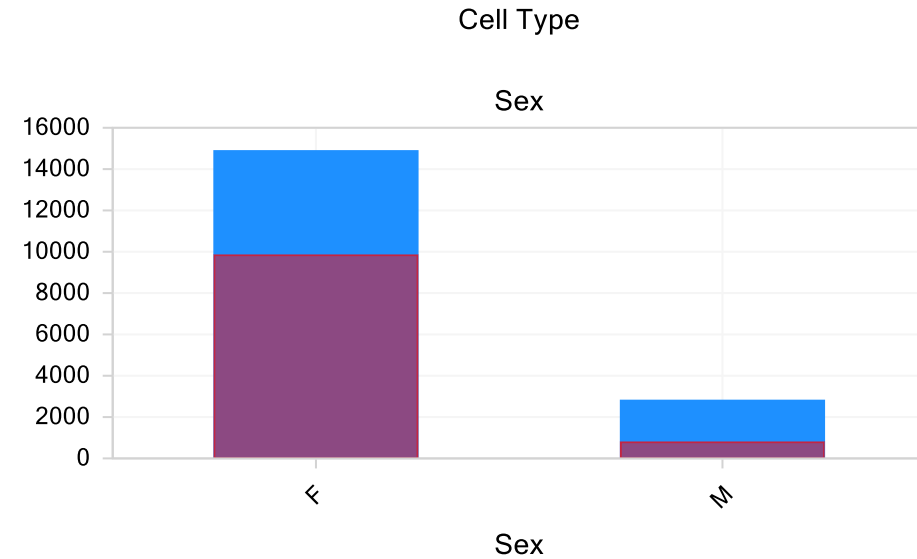
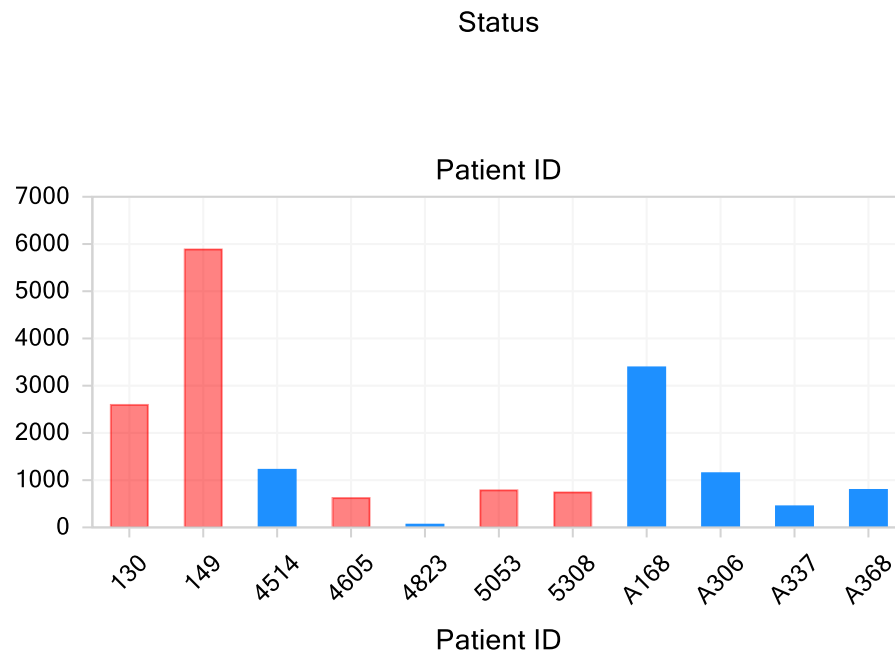
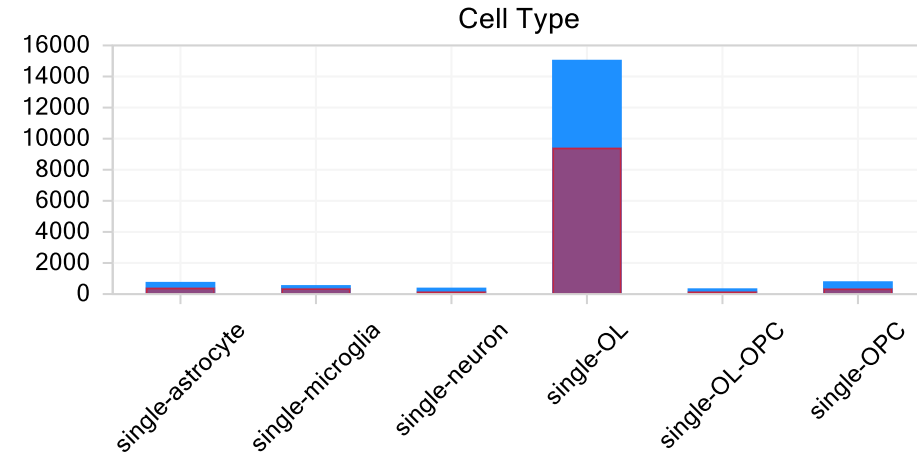
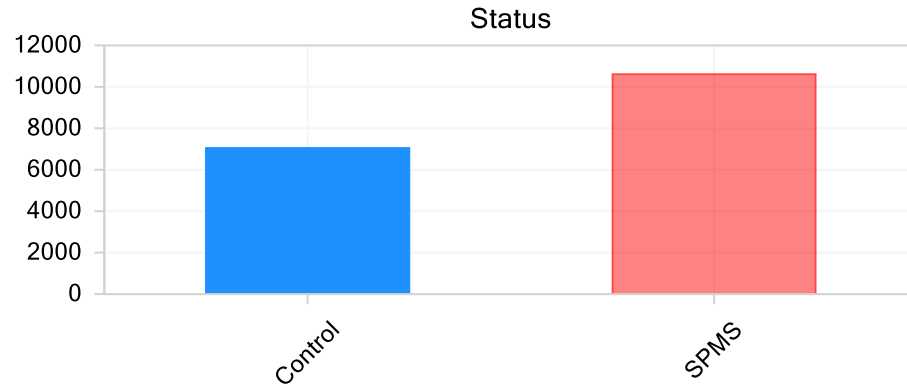
Histogram



# Percentage of Contributing Cells from each Patient



# Metadata Views are Dynamically Linked



# Gene Expression Views

## Navigation panel

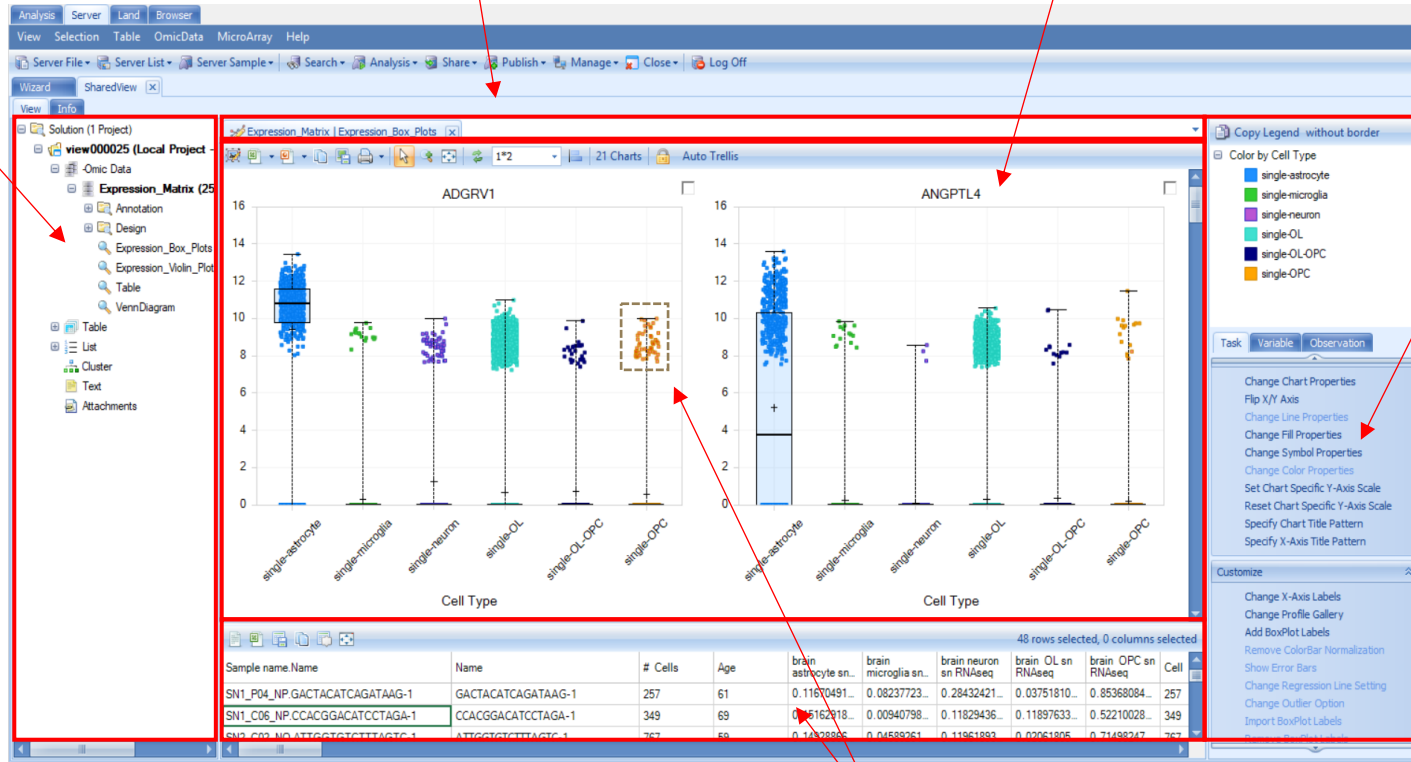
- Expression Data (ZIM)
- Box Plot
- Violin Plot

## Result view tab

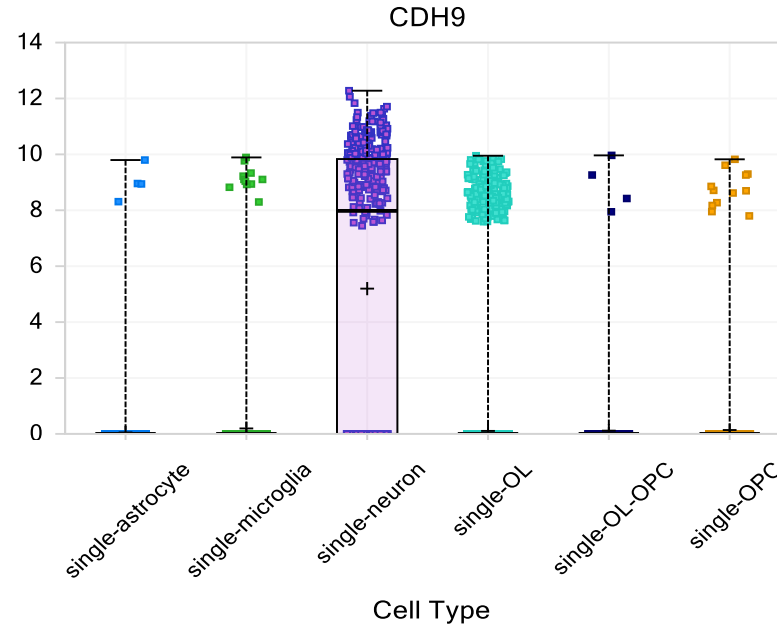
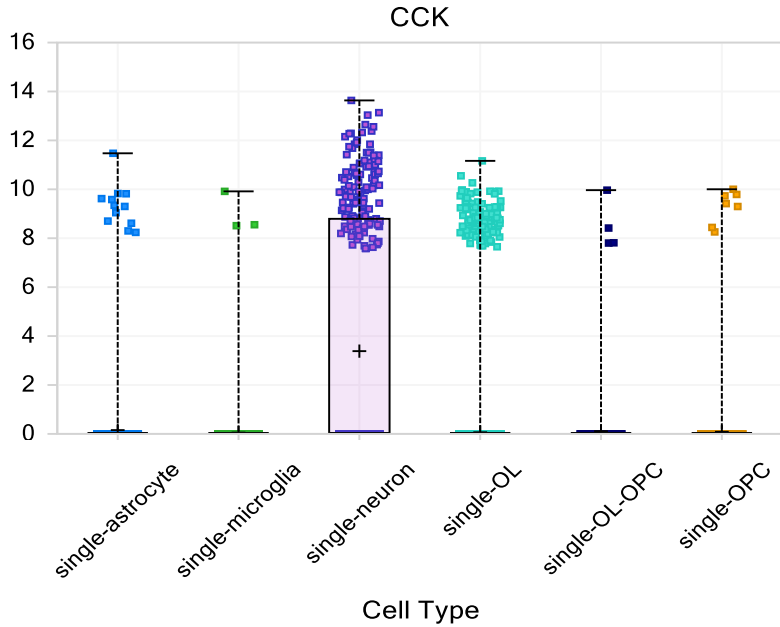
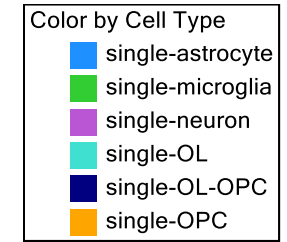
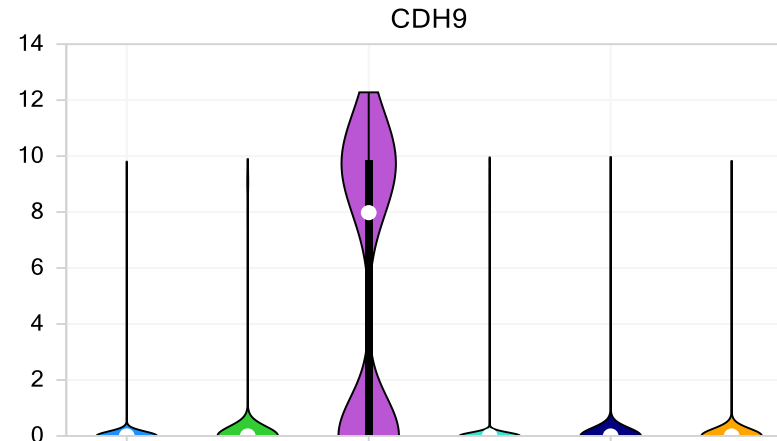
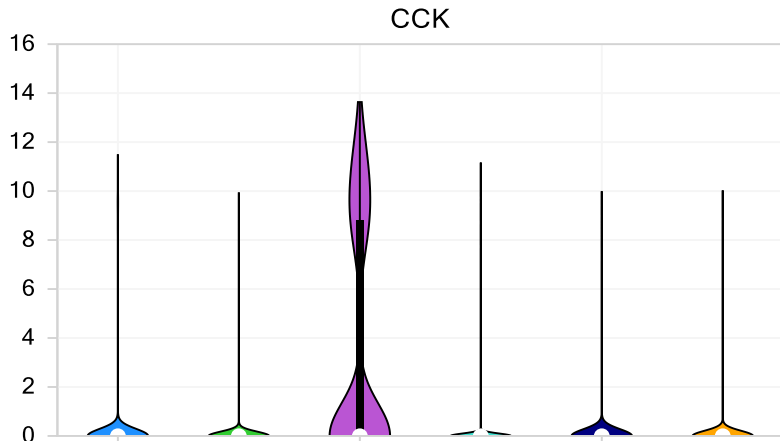
## Result view panel

## Customization panel

- Graphic visual output
- Data subsetting/filtering

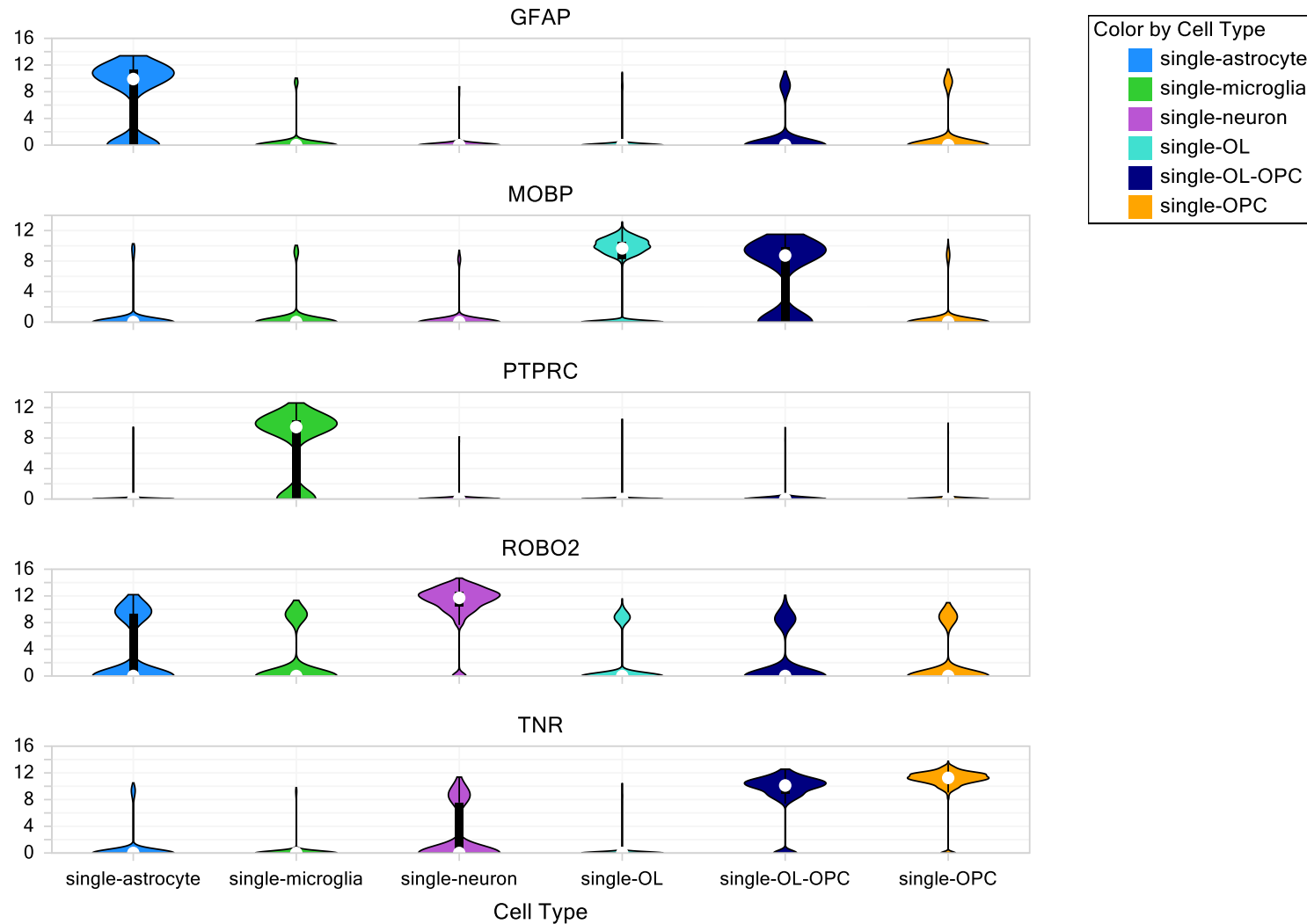


# Expression of Known Neural Markers Genes





# Visualization of Markers Genes in Different Cell Types

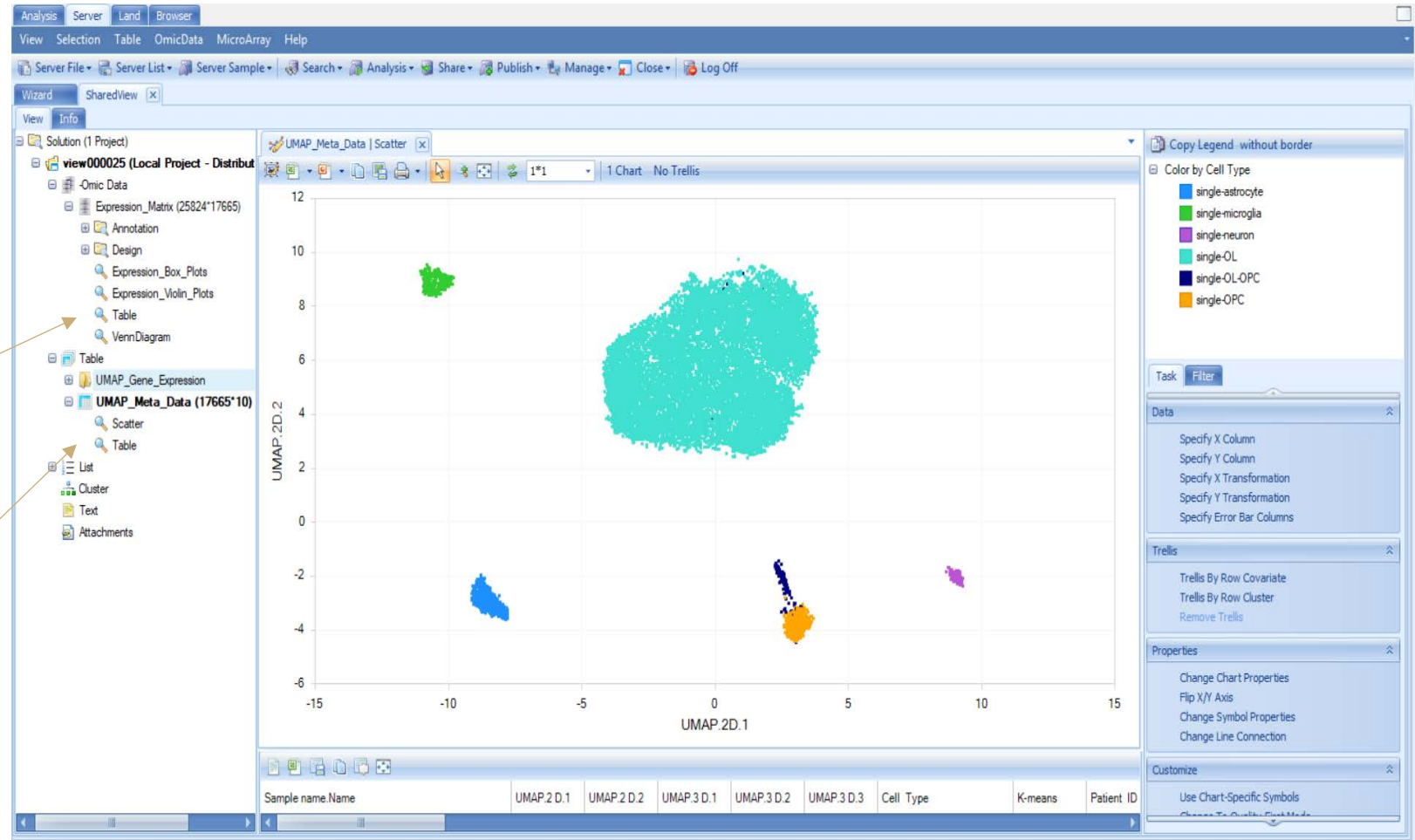


# UMAP Projections

- UMAP Coordinates
- Metadata
- Expression data

Overlay of gene expression comes from the ZIM

Metadata for overlay must be included as a column in the UMAP table



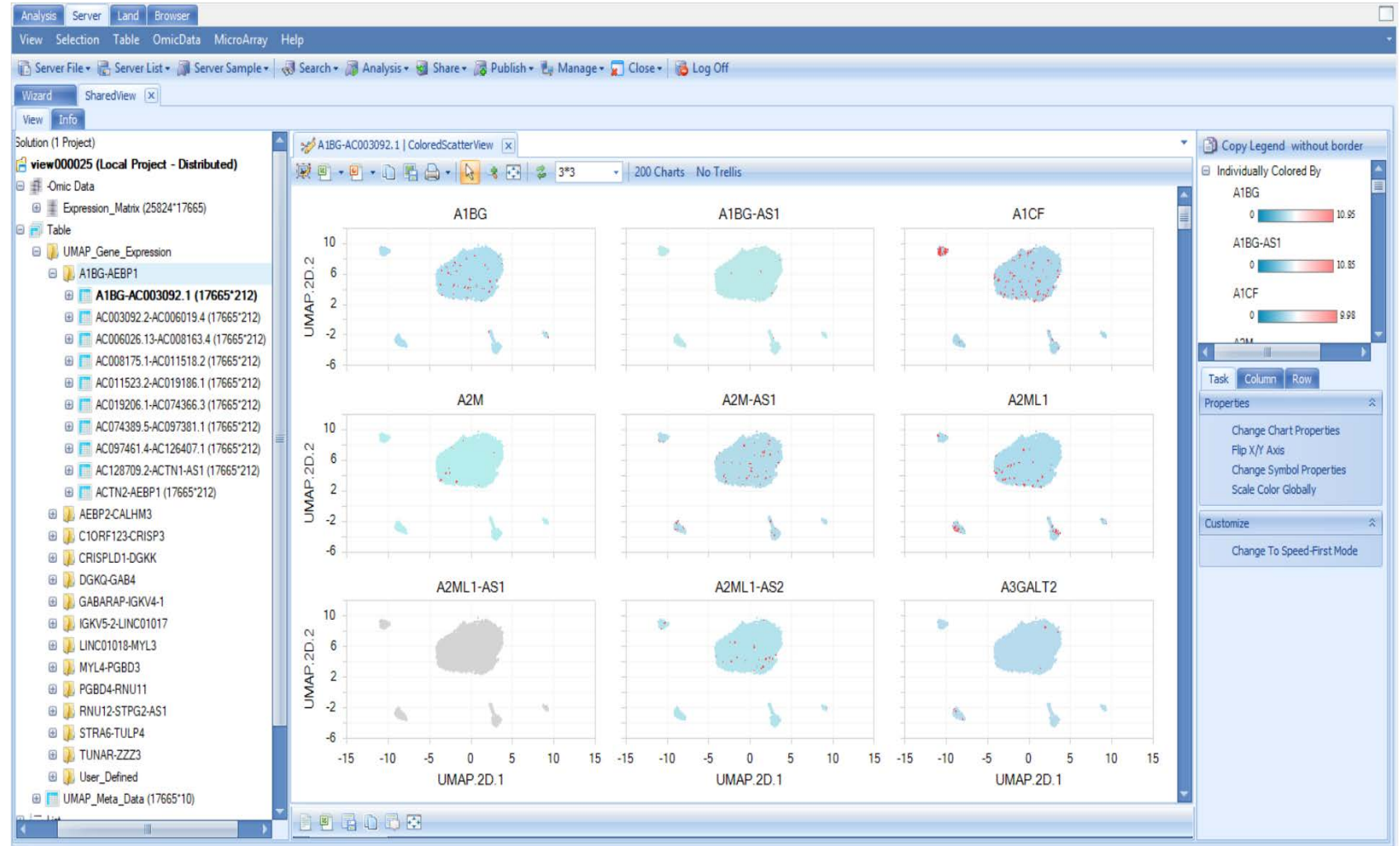
# Gene Expression Overlaid onto UMAP Projections

Overlaying genes onto the projection helps to elucidate the expression profile of the each cell type.

Opening the UMAP visualizations is quick compared to R-Shiny and other applications.

Each overlay is limited to 200 genes.

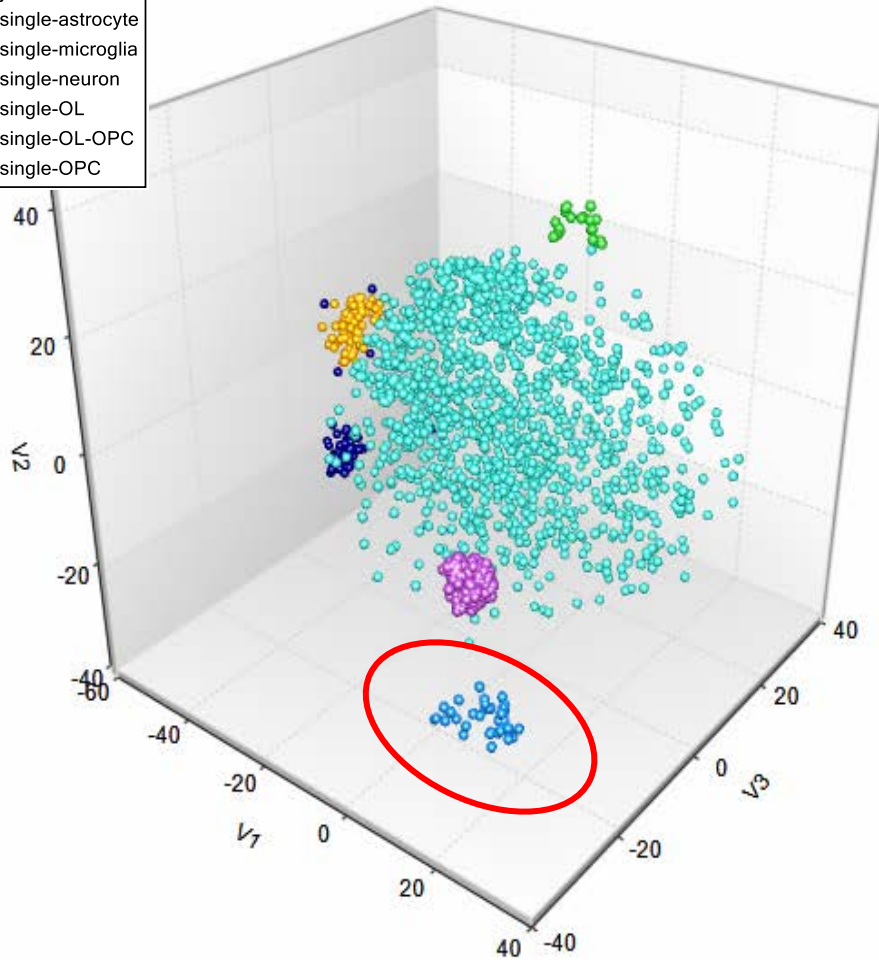
There is no search function, so an organizational strategy is helpful.



# t-SNE Projections

Color by Classifications

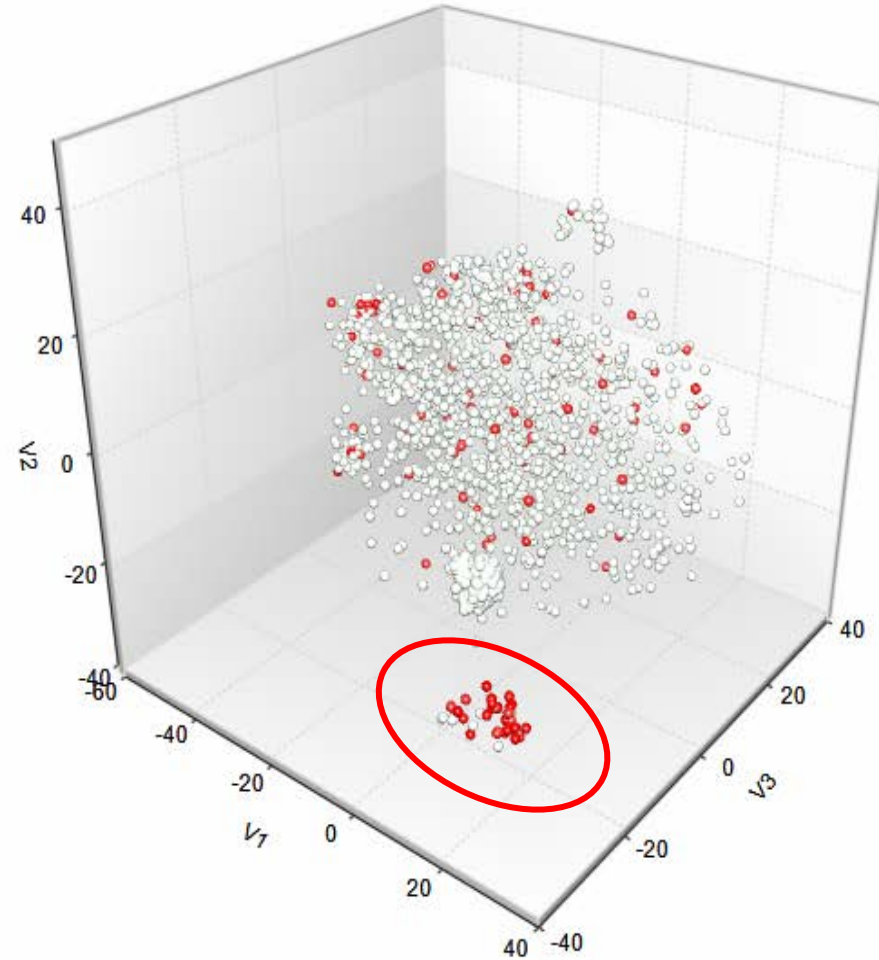
- single-astrocyte
- single-microglia
- single-neuron
- single-OL
- single-OL-OPC
- single-OPC



Colored by Cell Type

Color by GFAP

0 13.39



Colored by Gene Expression

# Summary

---

- OmicSoft's Array Suite enables the Bioinformatics department to share single cell data with collaborators at any stage, in real-time.
- Corporate License
  - Infrastructure in place
  - Company-wide accessibility
- We partnered with Omicsoft to develop single cell tools and views within Array Suite that are currently being used in the analysis of our MS study.
- Omicsoft continues to address some of the sticking points in order to make the process more efficient.

# Acknowledgements

---

## Sanofi

---

### Translational Sciences

Kathy Klinger

### Bioinformatics

Mindy Zhang  
Yi-Chien Chang  
Deepak Rajpal

### Genomics

Yinyin Huang  
Stephen Madden

### Rare and Neurological Diseases Research

Jonathan Proto  
Dimitry Ofengeim

## Qiagen/OmicSoft

---

Nirav Amin  
Larry Brescio  
Andrew Olsen  
Joseph Pearson  
Elina Tserlin

THANK YOU

