

Intro to SeqGeq v1.6

Timothy Quinn Crawford, PhD

Field Application Scientist, BDB Informatics/FlowJo

Timothy.Crawford@BD.com

Ian.Taylor@BD.com

What is SeqGeq?



- Single cell gene expression analysis software for the bench scientist
 - Load a CSV expression matrix
 - Gate or cluster on individual genes, gene sets or synthetic parameters to define populations
 - Pivot the graph to compare populations and define gene sets of differentially expressed gene products
 - Integrate with other programming environments (Plugins)
 - Identify biological processes and define novel populations

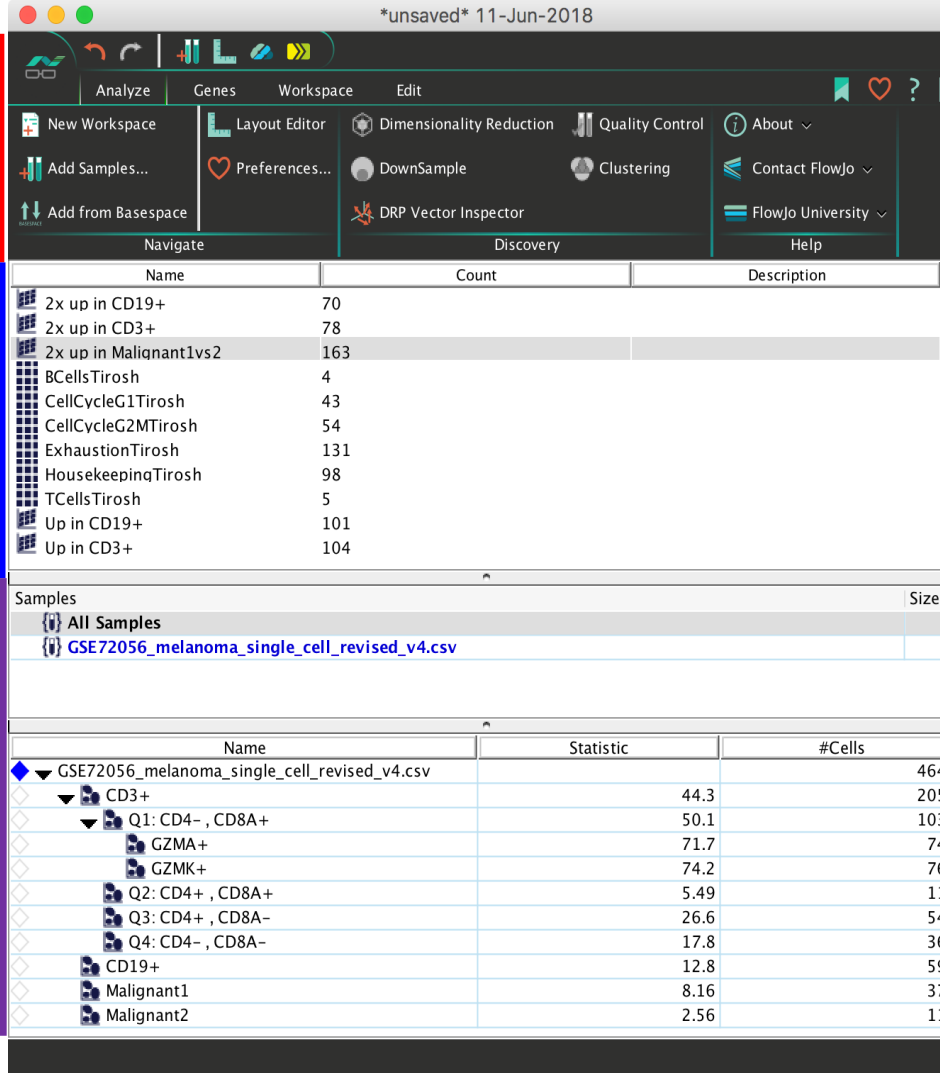
Why use SeqGeq?



- Requires no programming knowledge
- Allows bench scientist to explore their own scRNA Seq data w/o the bioinformatician
- Accepts CSV expression matrices from all single cell acquisition platforms
- If you already know FlowJo, the SeqGeq learning curve is minimal

The SeqGeq Workspace

- Organize your data and initiate actions
 - Ribbon, tabs, and bands
 - Gene Sets
 - Samples and sample analysis
- Double Click on a Sample to open a Graph Window and start your analysis



The screenshot displays the SeqGeq Workspace interface. At the top, a ribbon menu includes tabs for 'Analyze', 'Genes', 'Workspace', and 'Edit'. Below the ribbon, there are several toolbars: 'Navigate' (with 'New Workspace', 'Add Samples...', 'Add from Basespace'), 'Discovery' (with 'Layout Editor', 'Preferences...', 'Dimensionality Reduction', 'DownSample', 'DRP Vector Inspector', 'Clustering'), and 'Help' (with 'About', 'Contact FlowJo', 'FlowJo University').

The main workspace is divided into three sections:

- Gene Sets Table:** A table with columns 'Name', 'Count', and 'Description'. It lists various cell populations and their counts.
- Samples Table:** A table with columns 'Name' and 'Size'. It shows the loaded sample 'GSE72056_melanoma_single_cell_revised_v4.csv' with a size of 1.
- Cell Population Table:** A detailed table with columns 'Name', 'Statistic', and '#Cells'. It shows the hierarchical structure of the sample, starting from the total population and branching into various cell populations.

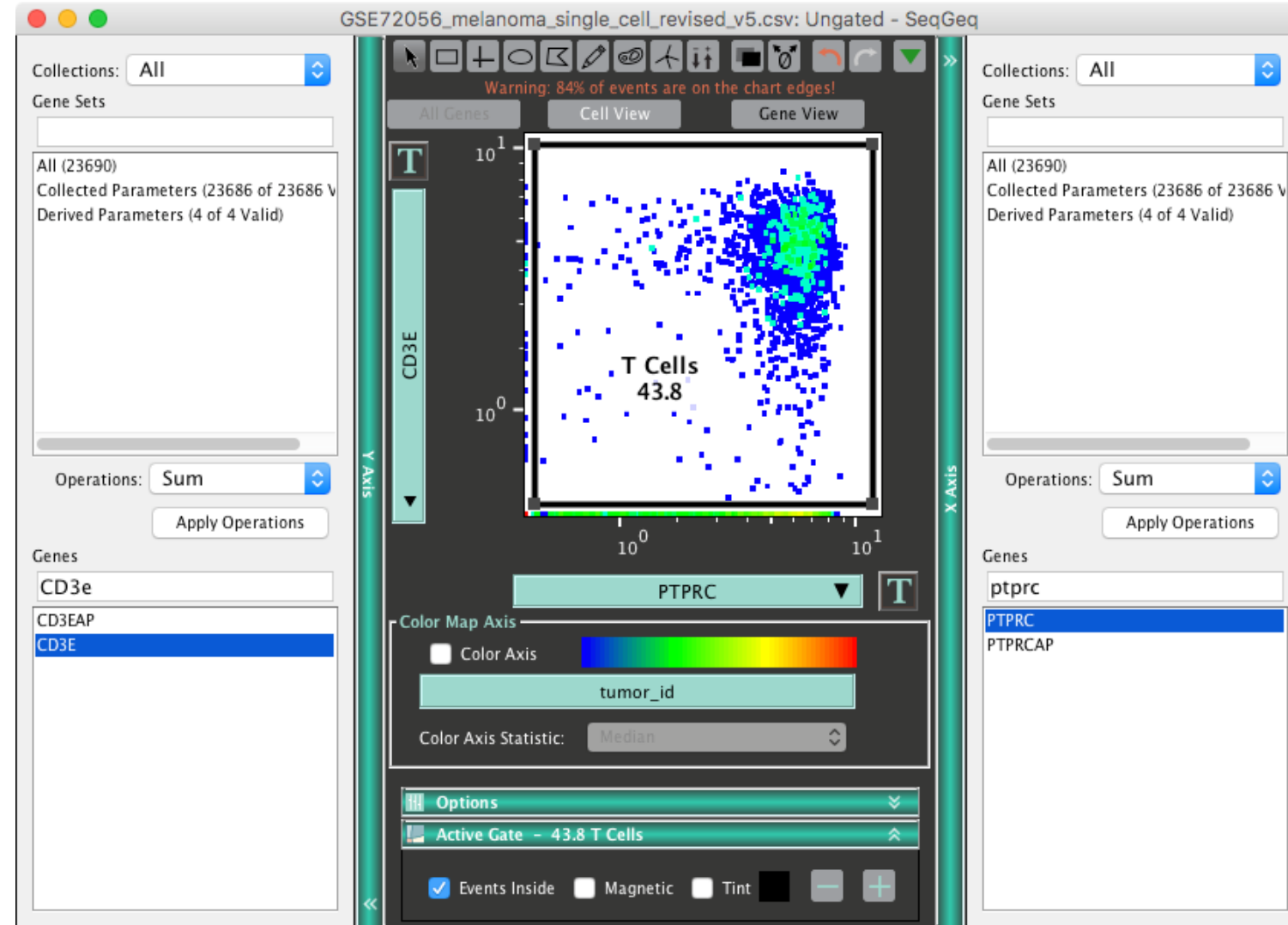
Name	Count	Description
2x up in CD19+	70	
2x up in CD3+	78	
2x up in Malignant1vs2	163	
BCellsTirosh	4	
CellCycleG1Tirosh	43	
CellCycleG2MTirosh	54	
ExhaustionTirosh	131	
HousekeepingTirosh	98	
TCellsTirosh	5	
Up in CD19+	101	
Up in CD3+	104	

Name	Size
All Samples	1
GSE72056_melanoma_single_cell_revised_v4.csv	1

Name	Statistic	#Cells
GSE72056_melanoma_single_cell_revised_v4.csv		4645
CD3+	44.3	2059
Q1: CD4-, CD8A+	50.1	1032
GZMA+	71.7	740
GZMK+	74.2	766
Q2: CD4+, CD8A+	5.49	113
Q3: CD4+, CD8A-	26.6	548
Q4: CD4-, CD8A-	17.8	366
CD19+	12.8	595
Malignant1	8.16	379
Malignant2	2.56	119

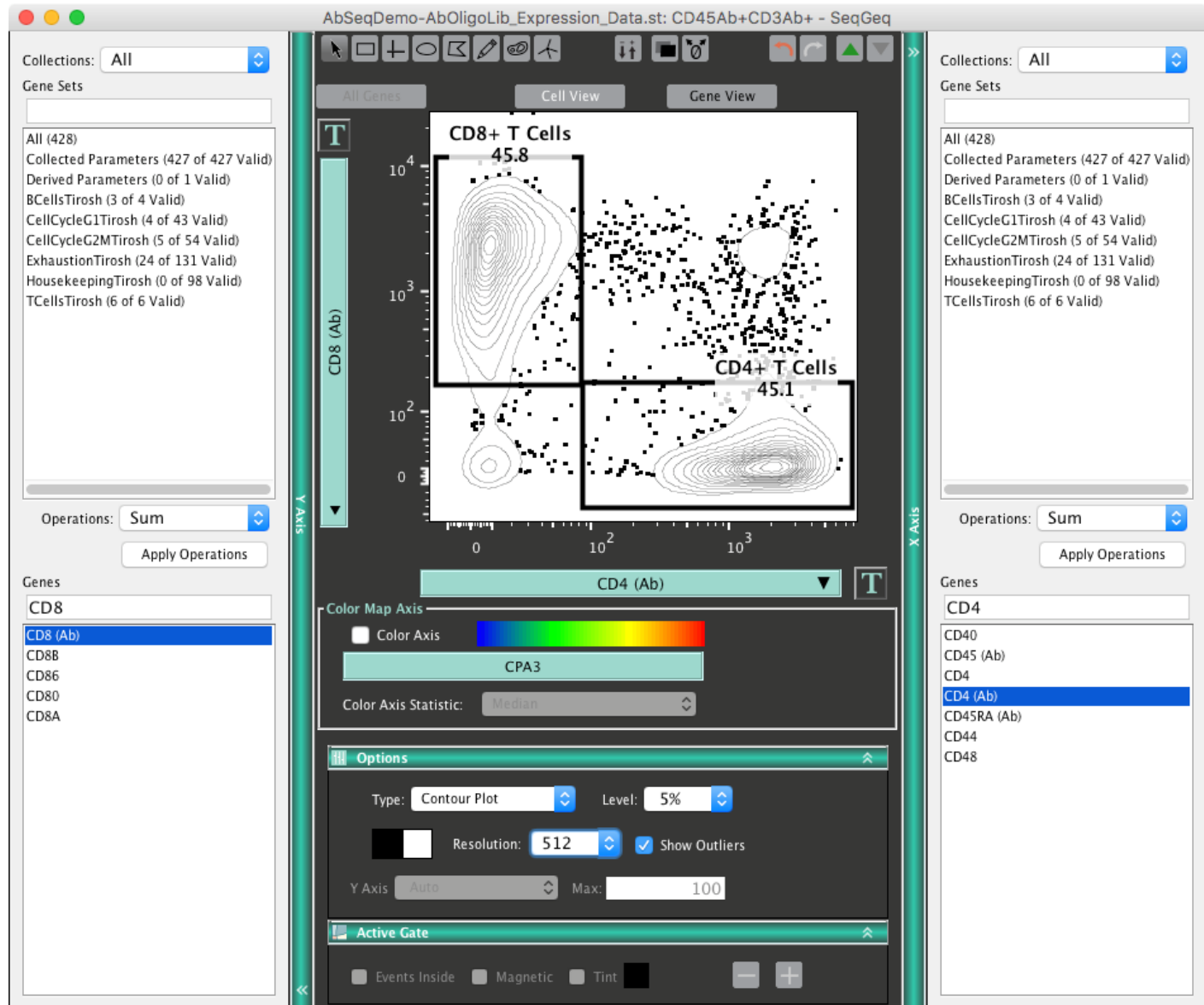
The SeqGeq Graph Window

- Searchable parameter selection wings
- Manual gating tools
 - Rectangle
 - Ellipse
 - Polygon
 - Pencil
 - Contour
 - Quadrant



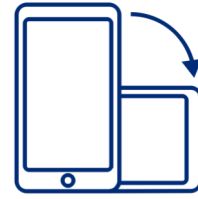
The SeqGeq Graph Window

- Transformation (T Button)
- Graph Plot View Options
- 3rd Parameter Color Map Axis
- Active Gate Options
- Cell View vs. Gene View



Cell View

cell	gene ₁	gene ₂	
cell id ₁	0	6	cell pop ₁
cell id ₂	2	0	
cell id ₃	2	4	cell pop ₂
cell id ₄	6	3	



Gene View

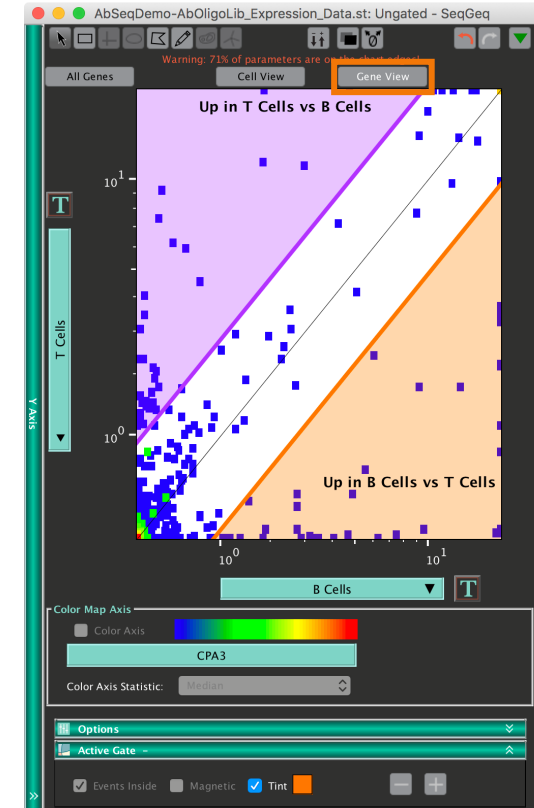
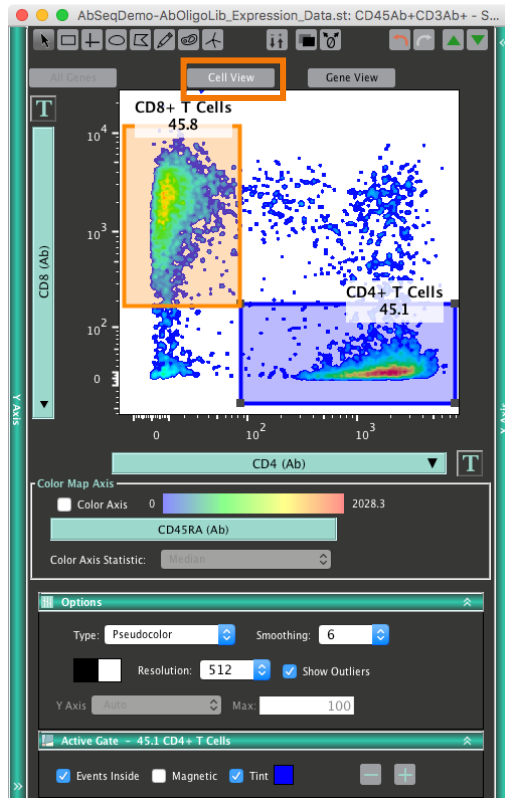
gene	cell pop ₁	cell pop ₂
gene ₁	{0,2}	{2,4}
gene ₂	{6,0}	{6,3}



- **Gene View** → Each dot is one or more measured parameters, comparing their expression between the two populations selected on the X and Y axes
- On the diagonal → Equally expressed
- Off the diagonal → Differentially expressed

- **Cell View** → Each dot is one or more measured events
- Just like a normal bivariate dot plot in flow cytometry

- Toggle between Cell and Gene view using buttons at the top of the graph plot



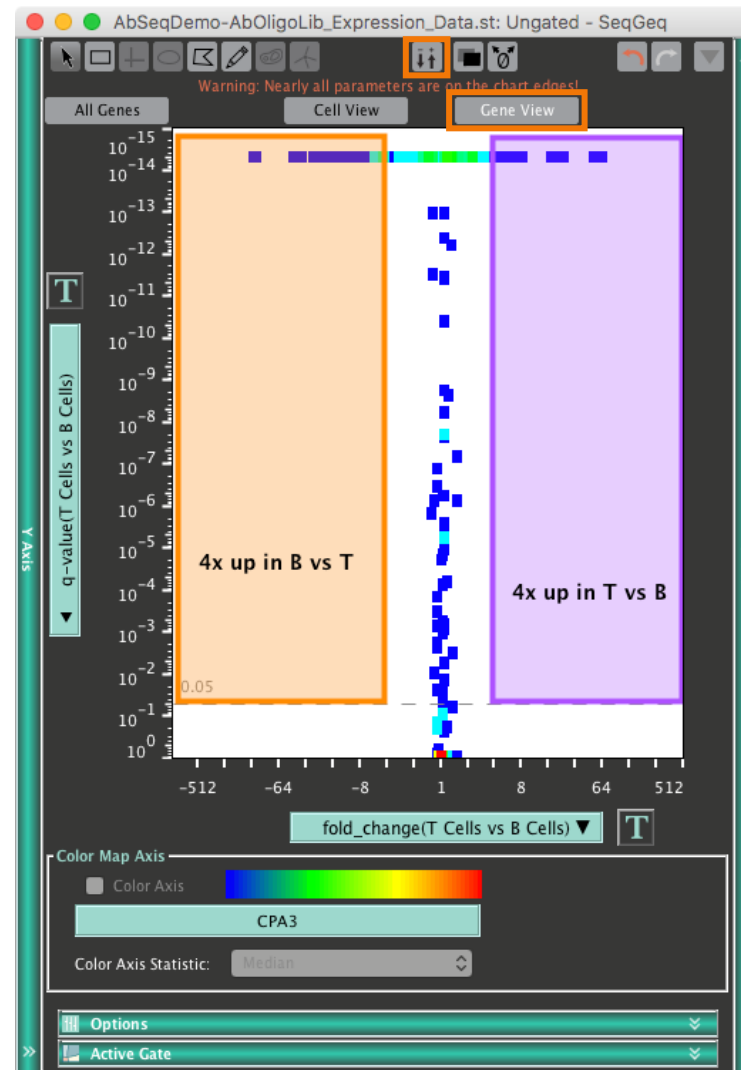
For Research Use Only.

For research use only. Not for use in any diagnostic or therapeutic procedure.

Volcano Plot

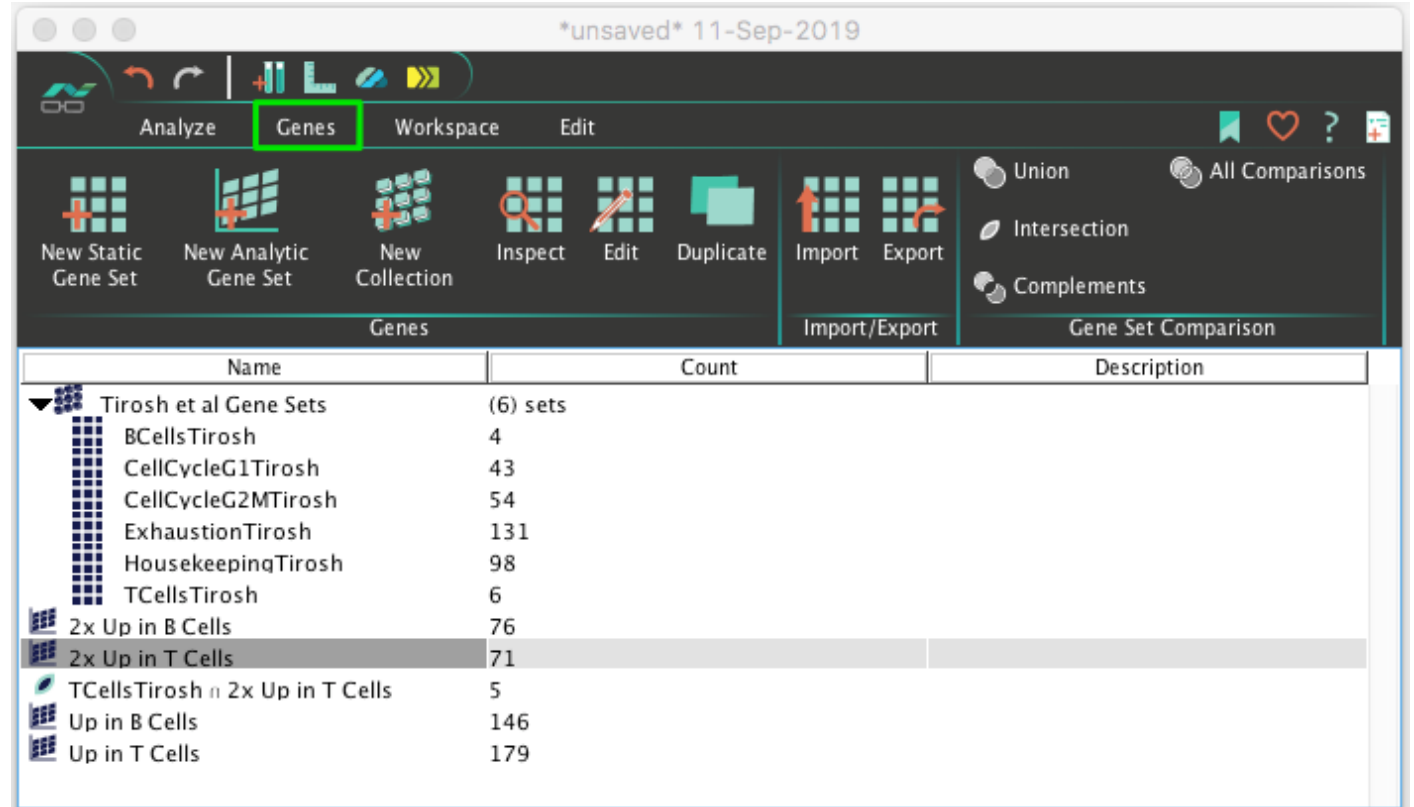
Derived Observations → Volcano Plot → Gate to define Differentially Expressed (DE) Genes

- Derived observations are per parameter computed values
- Calculation of fold change and p-value statistics allow for **statistical threshold gating** and identification of genes that are significantly differentially expressed between populations
- Gate in Gene View → Gene Sets!
- Filter on existing gene sets, or view all genes in the experiment



Gene Sets

- Two types of Gene Sets
 - **Analytic** – Created by a gate in gene view
 - **Static** – User defined. Create your own
- Can be compared (Intersection, Union, Complements)
- Can be exported and shared



The screenshot shows the 'Genes' tab in a software interface. The top toolbar includes buttons for 'New Static Gene Set', 'New Analytic Gene Set', 'New Collection', 'Inspect', 'Edit', 'Duplicate', 'Import', and 'Export'. On the right, there are options for 'Union', 'Intersection', 'Complements', and 'All Comparisons'. The main table lists gene sets with their names, counts, and descriptions.

Name	Count	Description
▼ Tirosh et al Gene Sets	(6) sets	
BCellsTirosh	4	
CellCycleG1Tirosh	43	
CellCycleG2MTirosh	54	
ExhaustionTirosh	131	
HousekeepingTirosh	98	
TCellsTirosh	6	
2x Up in B Cells	76	
2x Up in T Cells	71	
TCellsTirosh ∩ 2x Up in T Cells	5	
Up in B Cells	146	
Up in T Cells	179	

Gene Sets – Create and Edit

- Analytic Gene Set Inspector

Gene Set Inspector

Name: 2x up in Pop2 vs Pop1

Description:

Created: 1/18/19 12:14 PM

Modified: 1/24/19 1:50 PM

Gene Count: 21

Gene Set Type: ANALYTIC

Attributes: X Parameter: fold_change(Pop2 vs Pop1)
Y Parameter: q-value(Pop2 vs Pop1)
Parent set: All Genes

Parameters

Parameter	fold_chang...	q-value(Po...	Pop2	Pop1
CD56 (Ab)	6.55769	4.27e-15	116.096	16.8563
KLRC3	1.95903	4.27e-15	0.974038	0.00766283
Tim3 (Ab)	12.4451	4.27e-15	66.5183	4.42529
IL2RB	2.40962	4.27e-15	3.81923	1
TRDC	4.60569	4.27e-15	3.6851	0.0172414
PRF1	2.01234	4.27e-15	8.15577	3.54981
CTSW	2.03852	4.27e-15	4.88125	1.88506
IgG (Ab)	4.03933	4.27e-15	32.9784	7.41188
KLRK1	2.66567	4.27e-15	2.60529	0.35249
CD16 (Ab)	50.7729	4.27e-15	845.409	15.6705
GZMB	2.02158	4.27e-15	5.37067	2.15134
CD38 (Ab)	7.34518	4.27e-15	62.3346	7.6226
FCER1G	2.53138	4.27e-15	1.62837	0.0383142
CD7 (Ab)	7.47502	4.27e-15	230.525	29.9732
FCGR3A	5.51007	4.27e-15	6.4101	0.344828
ITGAX	2.17777	4.27e-15	1.26538	0.0402299

Buttons: ? Edit Gene Set Done

- Static Gene Set Editor

Static Gene Set Editor

Name: T Cell Lineage Genes

Description:

Collection: None

Collections: All

Gene Sets:

All (23697)
T Cell Lineage Genes (5 of 5 Valid)
T Cell Lineage Genes n Up in CD5+ B
Up in B Cells (63 of 63 Valid)
Up in CD5+ B Cells (45 of 45 Valid)
Up in Pop1 (59 of 59 Valid)
Up in T Cells (50 of 50 Valid)

Buttons: Add Selected > Add All >> Remove Selected x Remove All xx

Genes:

CD3E

Included Genes: (0 of 0)

CD4
PTPRC
CD5
CD8A
CD3E

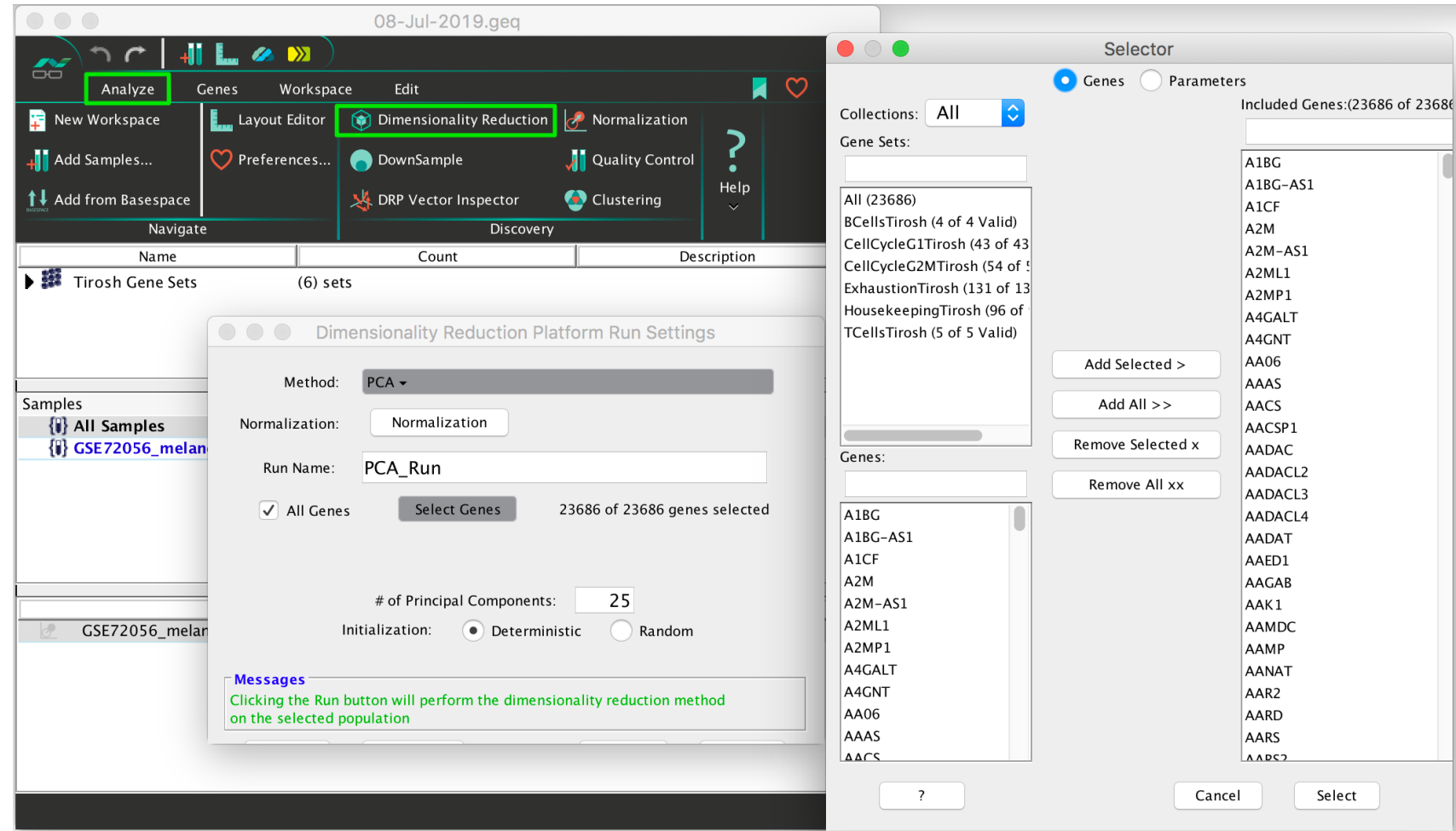
Buttons: ? Cancel Save

Dimensionality Reduction Platform

- Select Method and Genes, then click Run → New Analytic Parameters Created

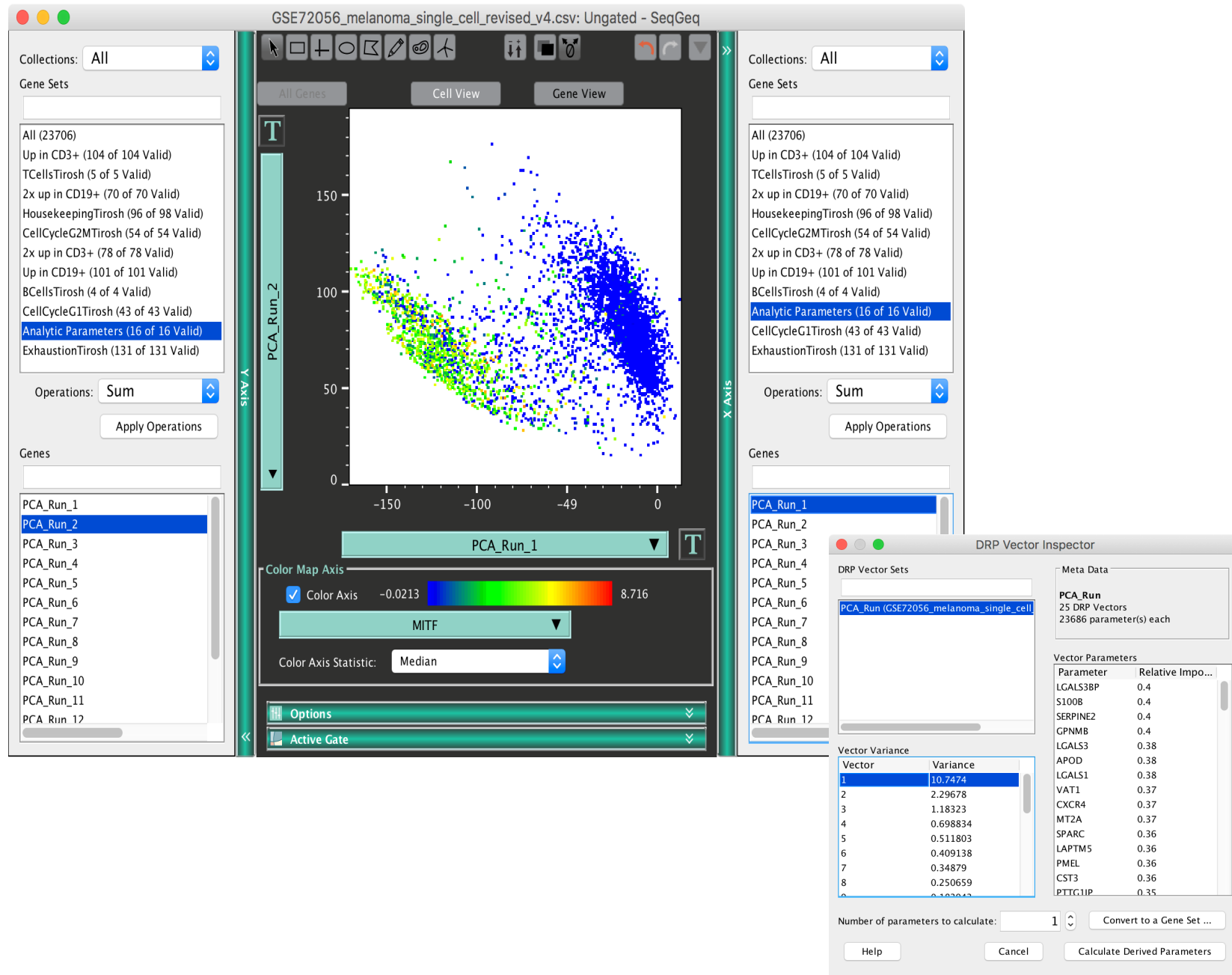
➤ PCA, LDA and tSNE options

- Suggestion



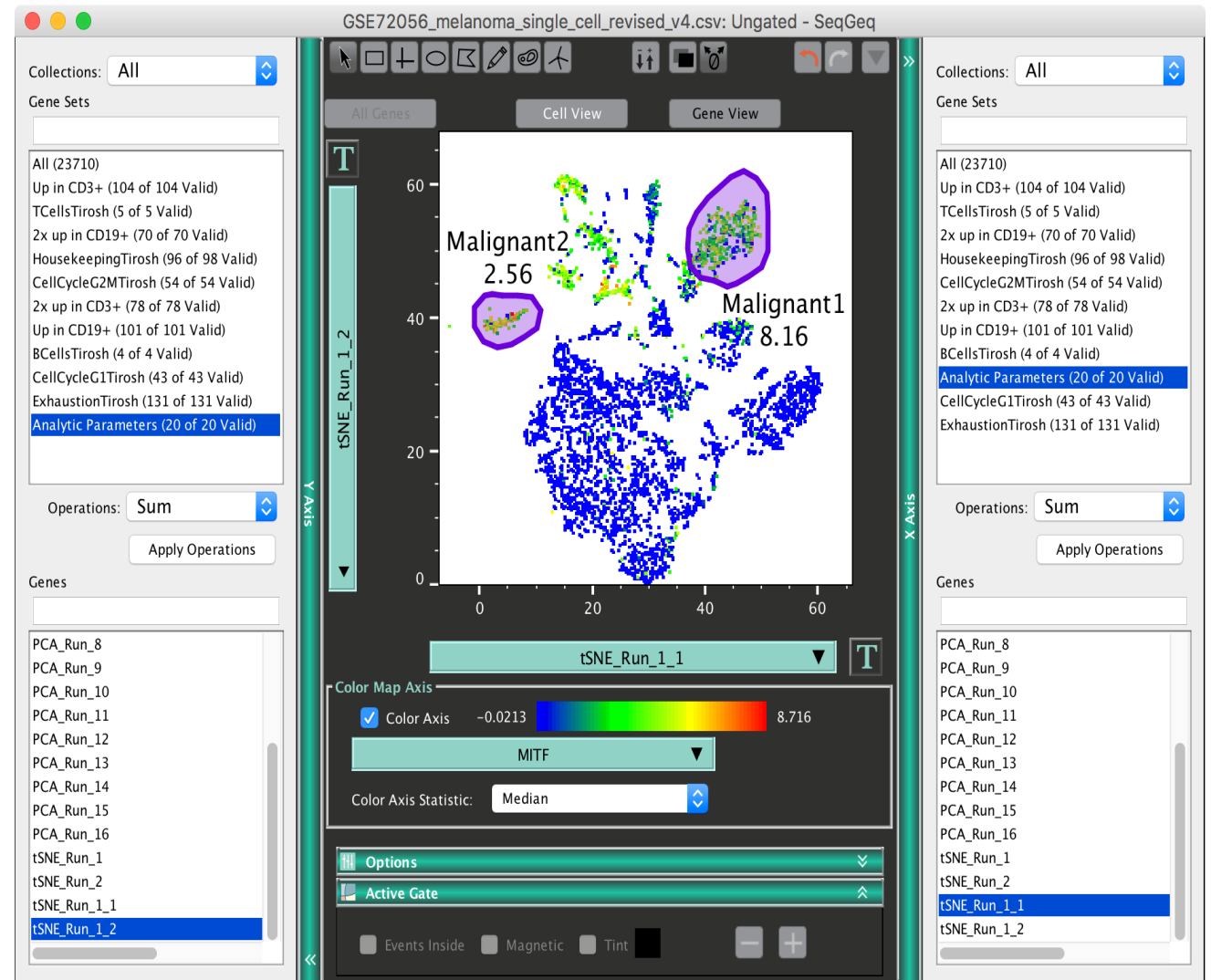
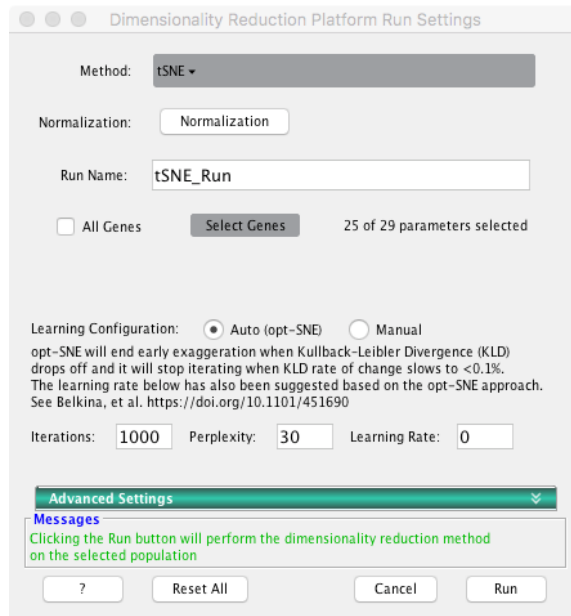
PCA Result

- Select Analytic Parameters to view
- Color map a 3rd parameter to explore expression of specific markers
- Gate on event clusters to define populations
- Vector Inspector → Gene sets of transcripts contributing to variance along the PCA axis



tSNE of PCA Parameters

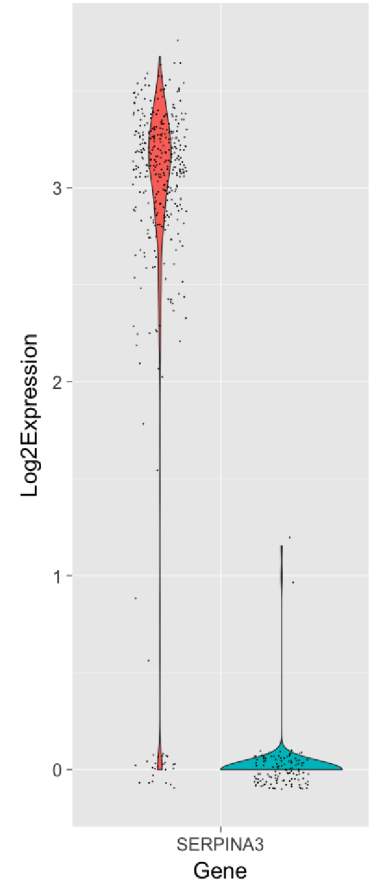
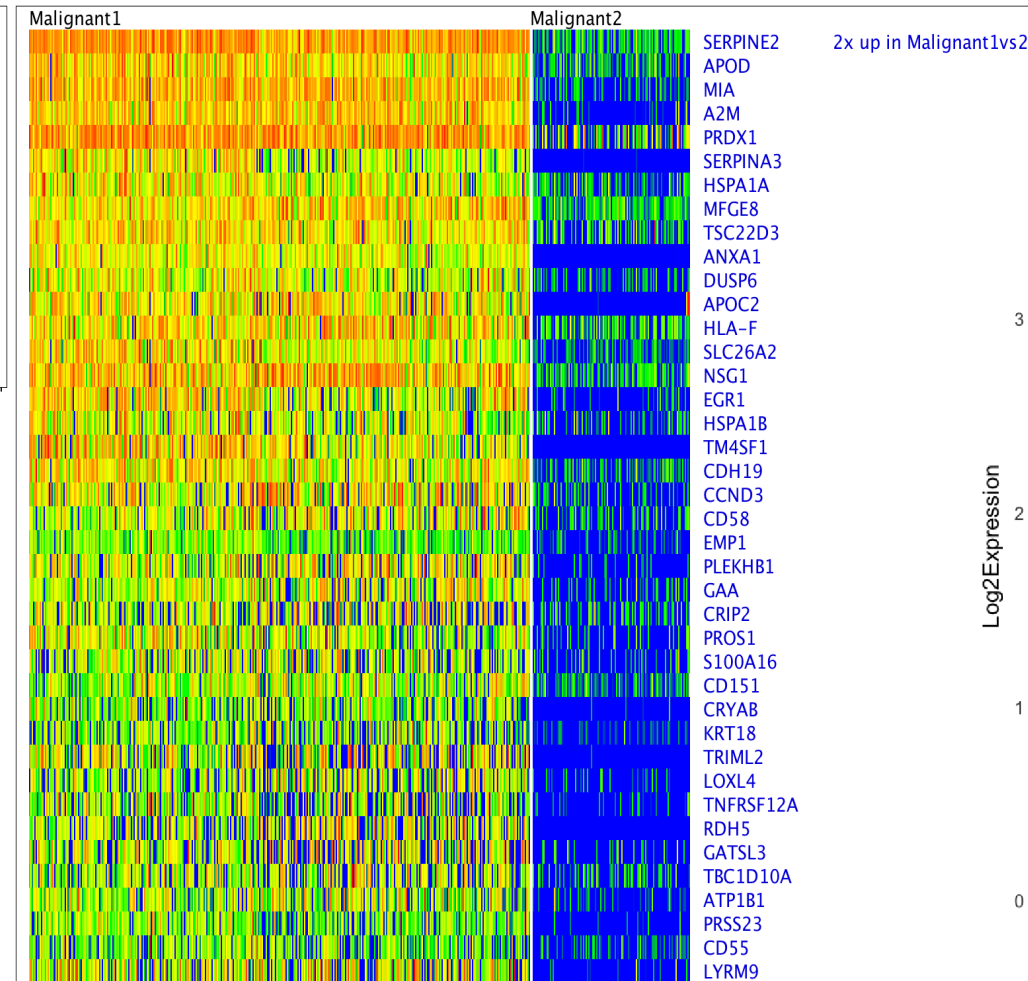
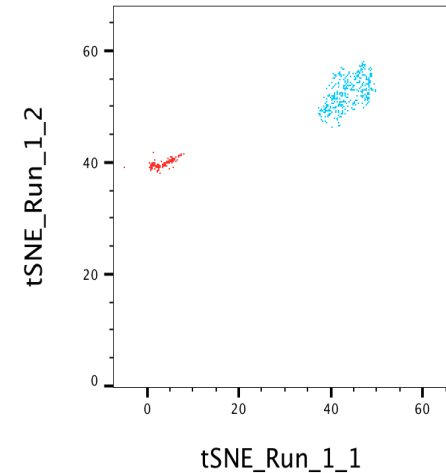
- Running tSNE using the PCA parameters → multiple distinct MITF+ populations
- Gate on event clusters to define populations
- Pivot into Gene View to determine DE genes



DE Gene Visualizations

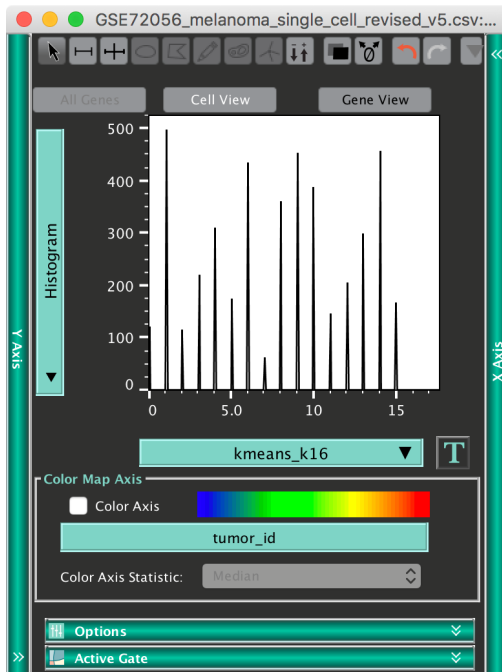
- Drag and drop populations to Layout Editor
- Right Click and select Make HeatMap to visualize expression of DE genes
- ViolinBox plugin → visualize expression of selected genes

	Sample Name	Subset Name	Count
■	GSE72056_melanoma_single_cell_revised_v4.csv	Malignant1	379
■	GSE72056_melanoma_single_cell_revised_v4.csv	Malignant2	119



Native Clustering Options

- Select Method and Genes, then click Run → Cluster Populations created
 - K Nearest Neighbors
 - K Means
- AutoGate Categorical



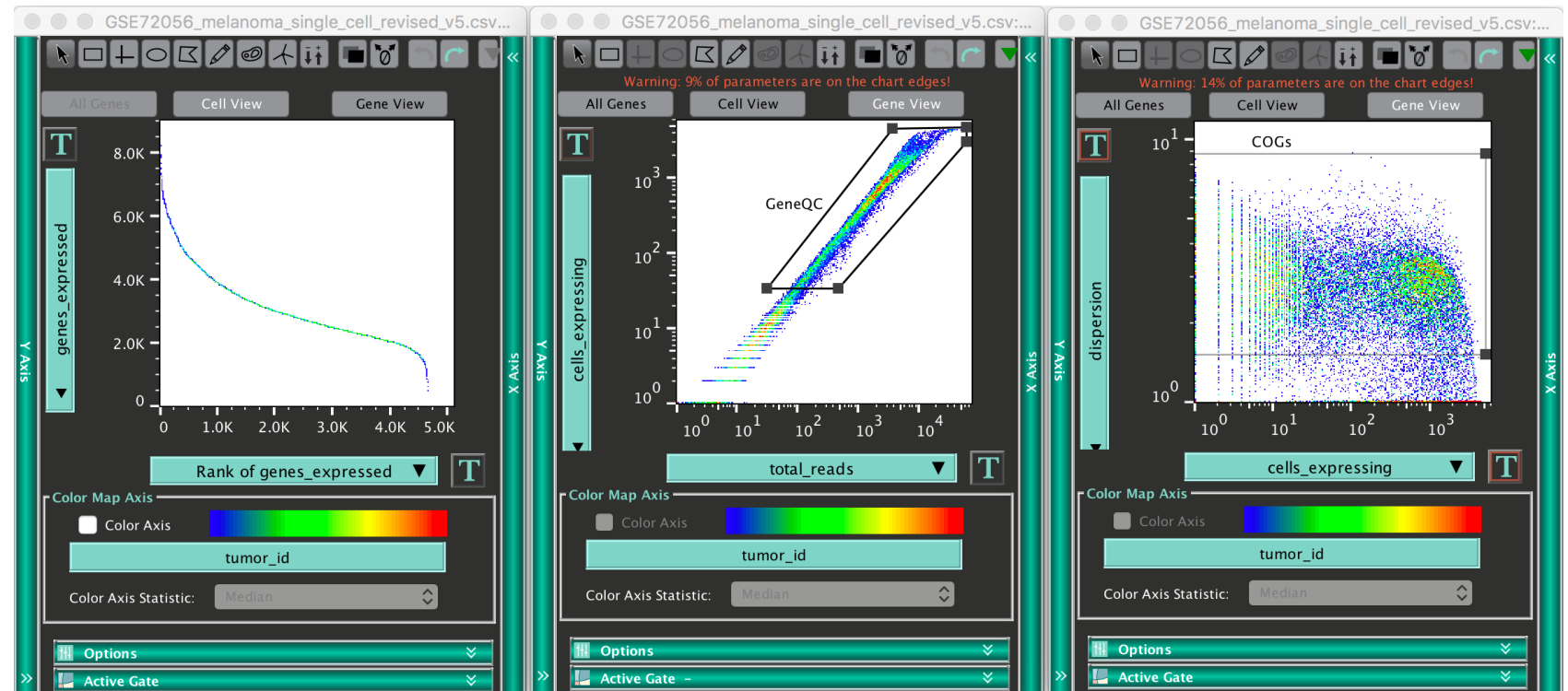
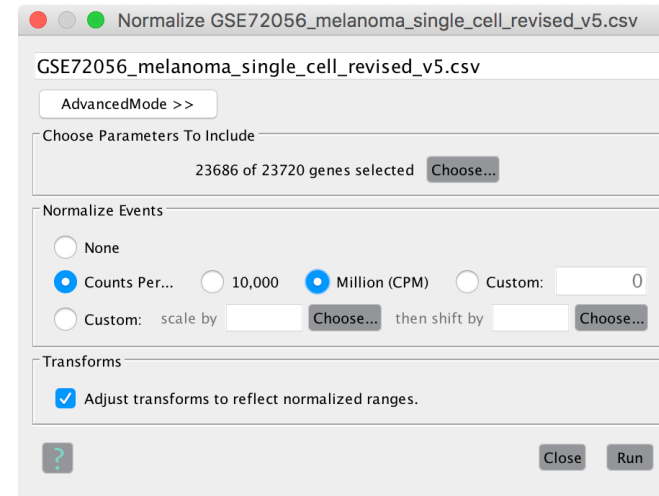
The screenshot shows the FlowJo software interface. The 'Analyze' menu is highlighted in green. The 'Clustering' option is also highlighted in green. A dialog box titled 'Cluster/Classify GSE72056_melanoma_single_cell...' is open, showing the 'KMeans Cluster' algorithm selected. The 'Run Name' is 'kmeans', the 'Population' is 'GSE72056_melanoma_single_cell_revised_v5.csv', and the 'Parameters' are '23686 of 23686 genes selected'. The 'K' value is set to 16. The 'Close' and 'Run' buttons are visible at the bottom of the dialog box.

Name	Count	Description
Tiresh Gene Sets	(6) sets	
BCellsTiresh	4	
CellCycleG1Tiresh	43	
CellCycleG2MTiresh	54	
ExhaustionTiresh	131	
HousekeepingTiresh	98	
TCellsTiresh	5	

Name	Statistic	#Cells
GSE72056_melanoma_single_cell_revised_v5.csv		4645

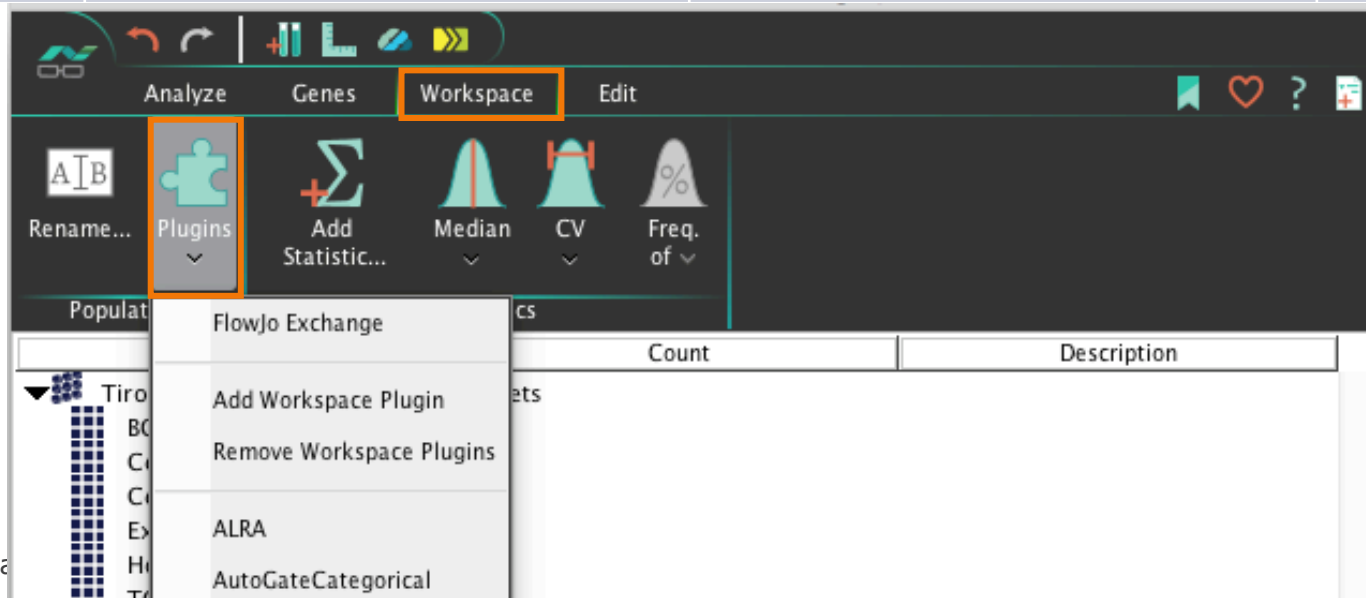
Additional Features

- Normalization tools
- Quality Control Plots
- Enrichment tests against gene set libraries
- Plugins to run your favorite pipeline or algorithm

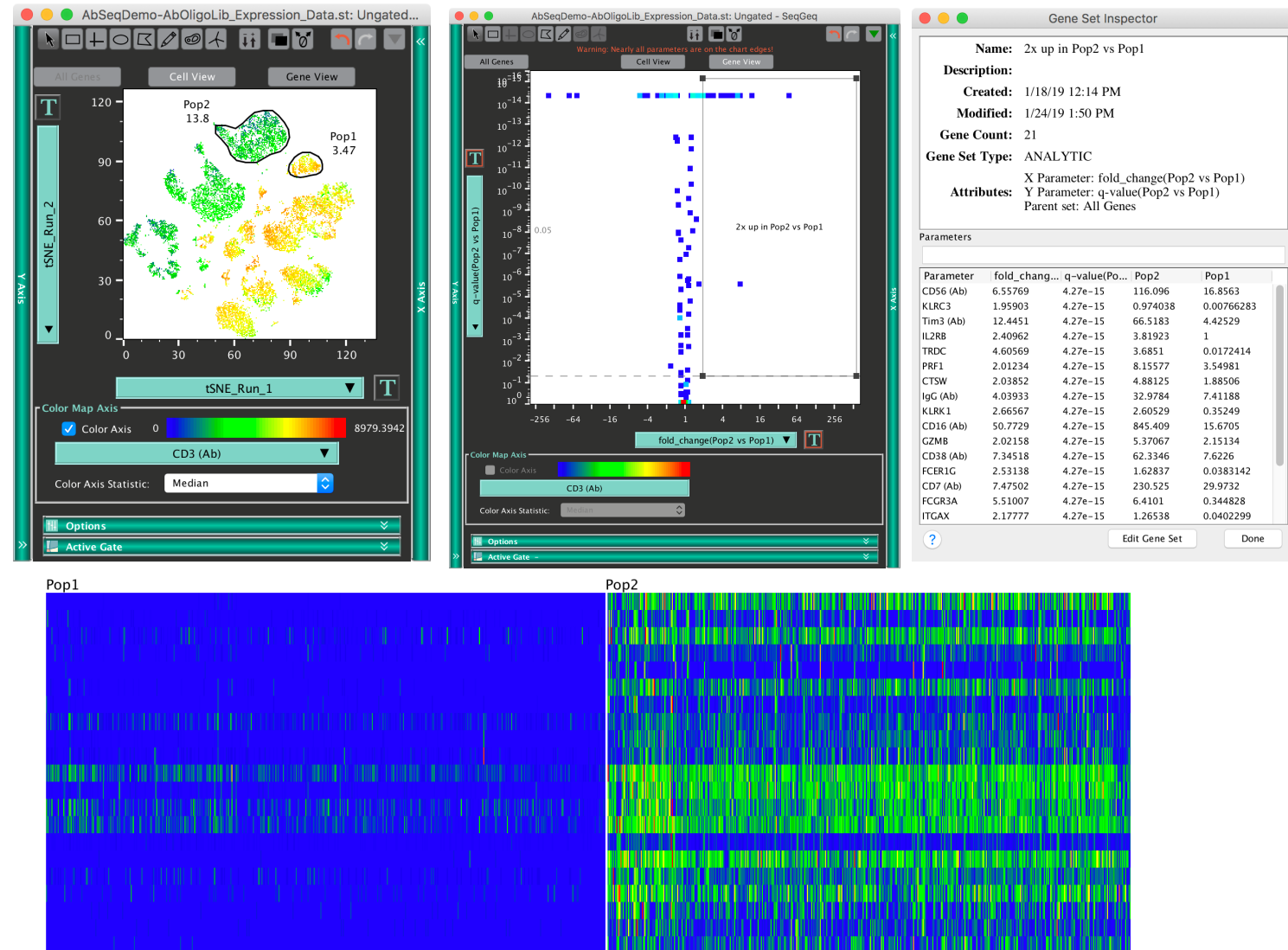
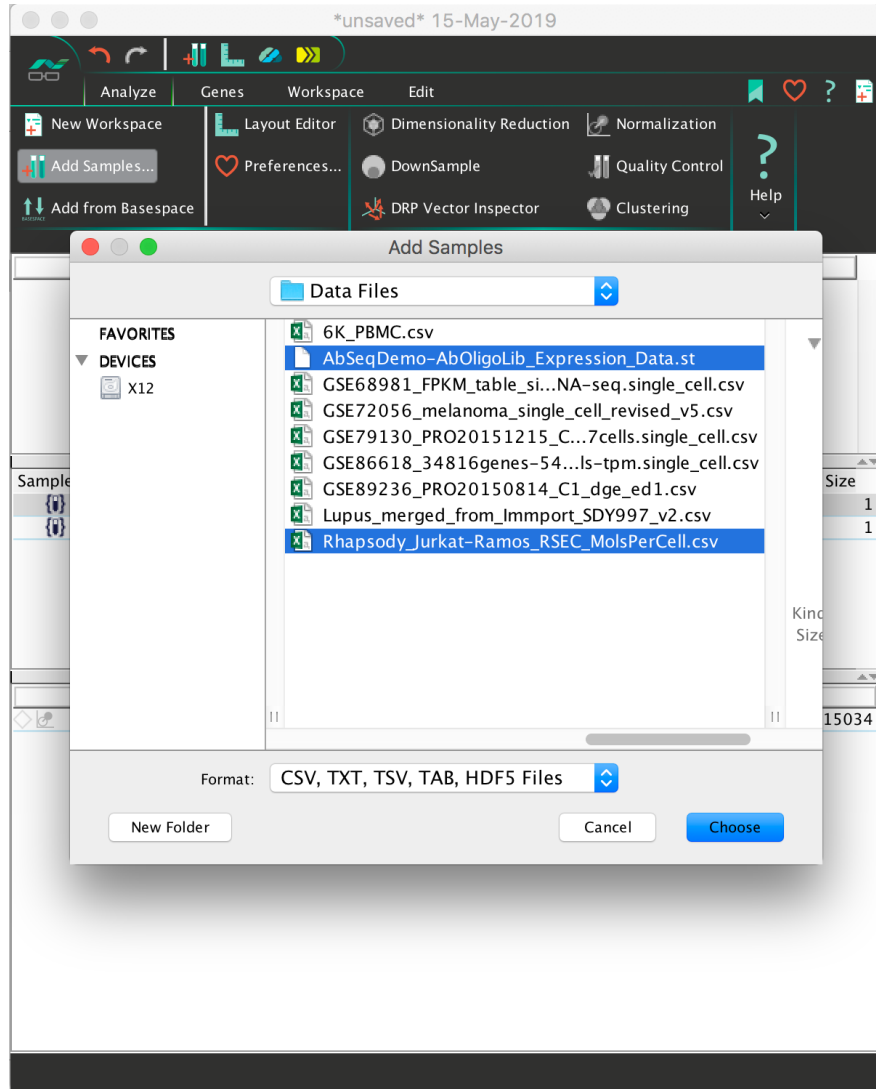


Plugins

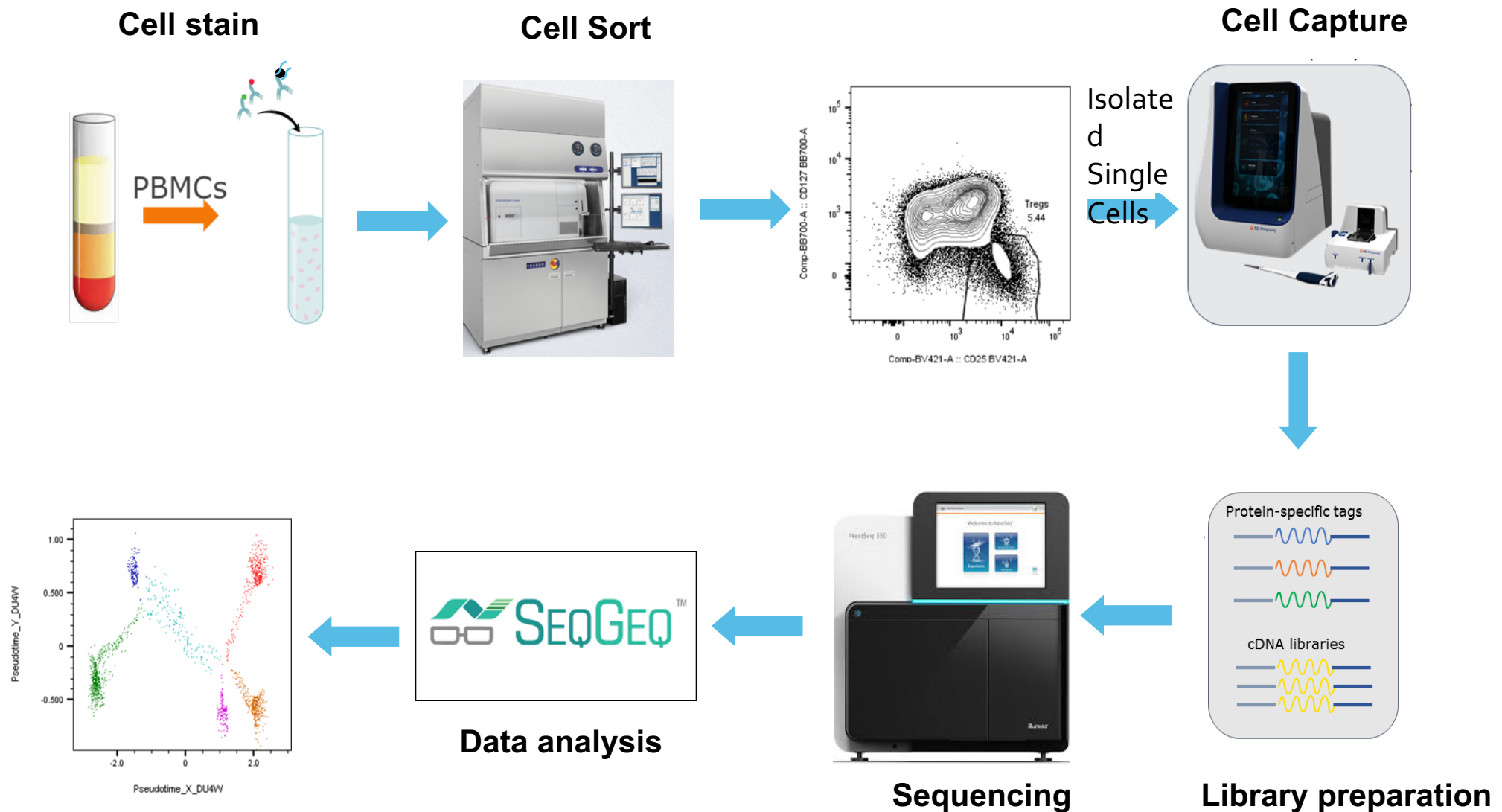
Pre-processing	Visualization	Clustering	Interpretation +Dig Deeper
AutoGateCategorical	UMAP	FlowSOM	VDJ Explorer
scImpute	TriMAP	Phenograph	
Lex	Monacle	X-Shift	
K-finder	ViolinBox	ClassyDL	
Seurat	iCellR	SC3	



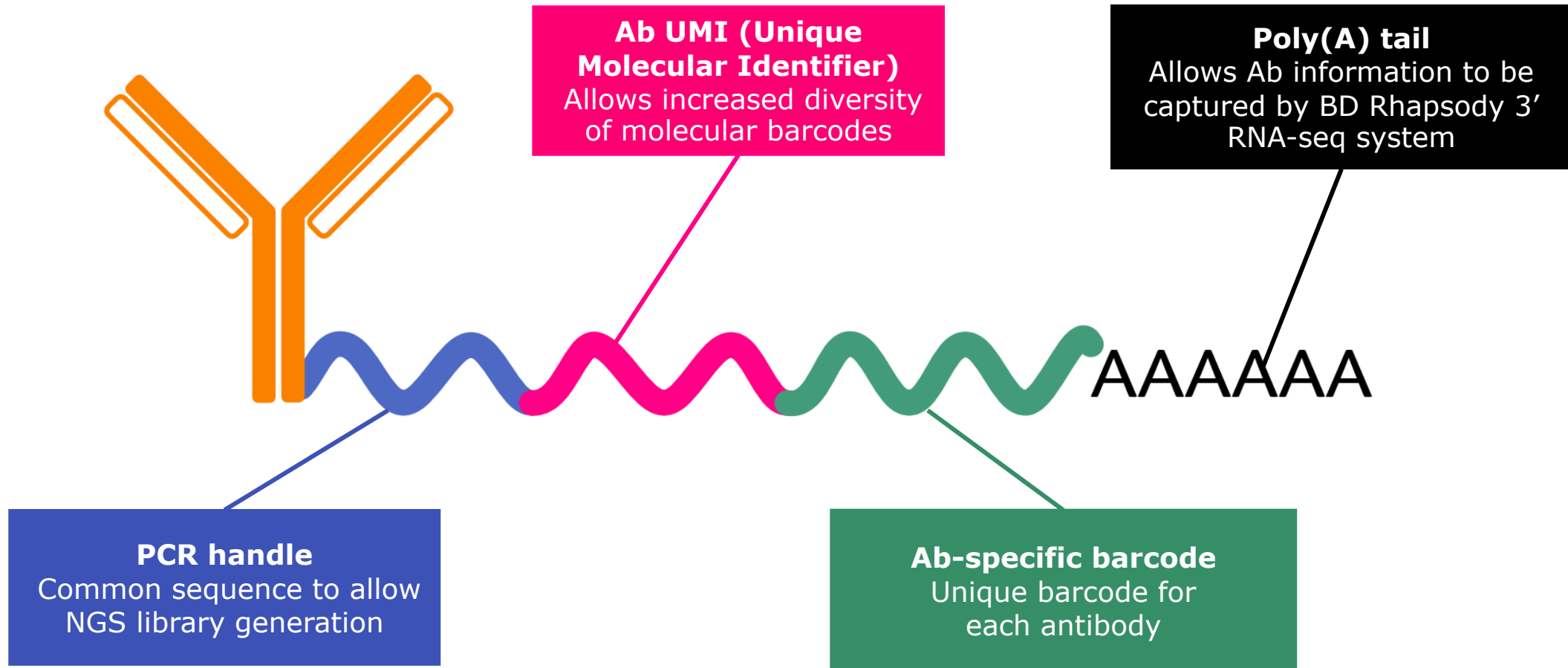
Demo Data Included with Download!



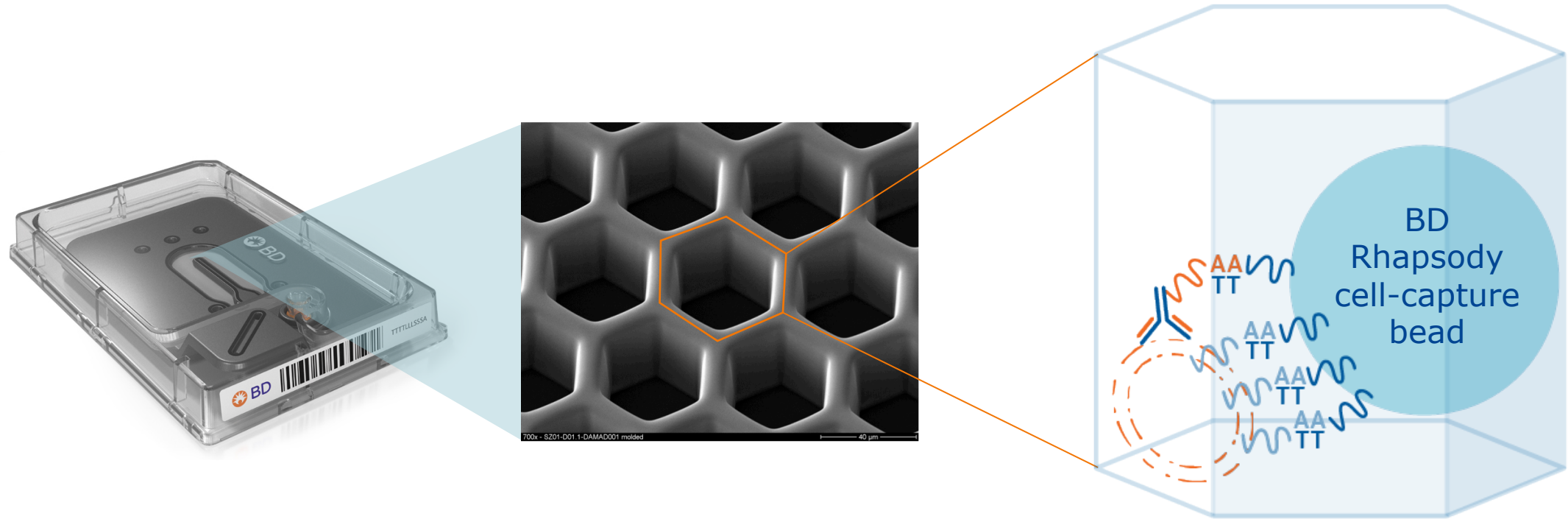
AbSeq Experimental Workflow



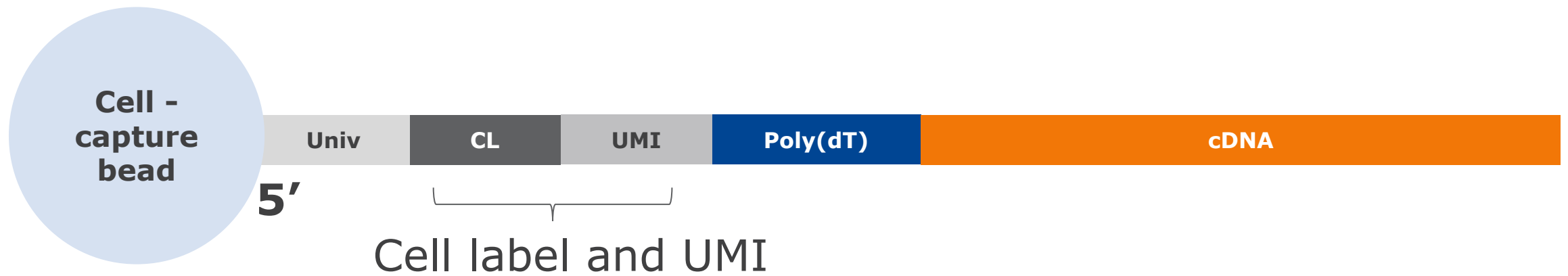
Poly(A) Ab-oligos allow antibody detection and readout using sequencing



Ab-oligos are captured with RNA



cDNA synthesis after single-cell capture encodes transcriptome (and Ab-oligo) information on beads



BD Rhapsody Bioinformatics



FastQ

.csv File

Insight

```
--->gzip -cd L2I_S1_L001_R1_001.fastq.gz
@M00805:5:000000000-A0VLL:1:1101:16473:
NTTGTCACTCAGCTGAAGATGAAATAGGATGTAATCAGAC
TTGGAAC TAGGT CAGCTGAAGATCCTGTGAGCGAAGTTC
+
#55<<?BBDBDDDDDDFFFFFH HHHHHFHHAFHHHHHHH
AFHHHHHHHHFGHDDHFBFHDFFHFFHHHHFFA=@BEE
@M00805:5:000000000-A0VLL:1:1101:15023:
NAGAAATCACAGACATACAAAGCAGTCTGTGCTCTTAGG
AGCATCTGCCGTCACATTGTTCTGCACACACCGTCCTTG
+
#55???BBDDDDDDDDGGGGGGIIIIIIIIIIIIIIIIII
HHHHHHHHHHHHHHHHHHHHHHHHHHHHGGFGE GGGGGGGG
@M00805:5:000000000-A0VLL:1:1101:14046:
NTTTCGTGGAAGTGGGTTACCTGACAGTGTGCACGCCCC
ACATGAGTGCCTCTCTTCAGAGCTGTCTGCTTTTCTG
```

Cell_Index	ABCB1	ABCG2_ALT1	ACTA2	ACTR3B
52803	0	0	1	1
718930	0	0	3	0
160403	0	0	59	0
720136	0	0	1	2
727124	0	0	1	0
815380	0	0	70	0
257971	0	0	83	0
46714	0	0	12	0
825447	0	0	0	1
443121	0	0	0	1
493478	0	0	4	0
315769	0	0	1	1
543894	0	0	3	1
580637	0	0	56	0
398887	0	0	42	0

SevenBridges

Projects

Data

Public Apps

Public projects

Developer

Staff

Dashboard

Files

Apps

Tasks

Quickstart

Copy project

Interactive Analysis

Settings

Notes

Description

Welcome to your new project!

Projects are the core building blocks of the Seven Bridges Platform. Each project corresponds to a distinct scientific investigation, serving as a container for its data, analysis pipelines, and results. Projects are shared only by designated project members.

Within your project, you can:

- Start exploring the public pipelines straight away
- Install your tools and create workflows
- Upload your own private data
- Collaborate securely with other researchers

After reviewing the information above, you can continue to use this space for adding notes about your project such as its aims, experimental context, and any other ideas that you'd like to share with your project members as everyone will see the same content. You can also use [markdown](#) here to add formatting to your notes.

To start adding your description, click **Add Description** below.

Remember that details of each pipeline execution you run on the Seven Bridges Platform are logged on the dedicated task page.

Good luck with your research! If you get stuck, take a look at the [Knowledge Center](#).

The Seven Bridges Team

Add description

Members

Email notifications

franklin

Witte, Cress, Executive, Admin

Don't work alone. The best research happens in teams.

Invite new members

Share your tools, data, and ideas with collaborators

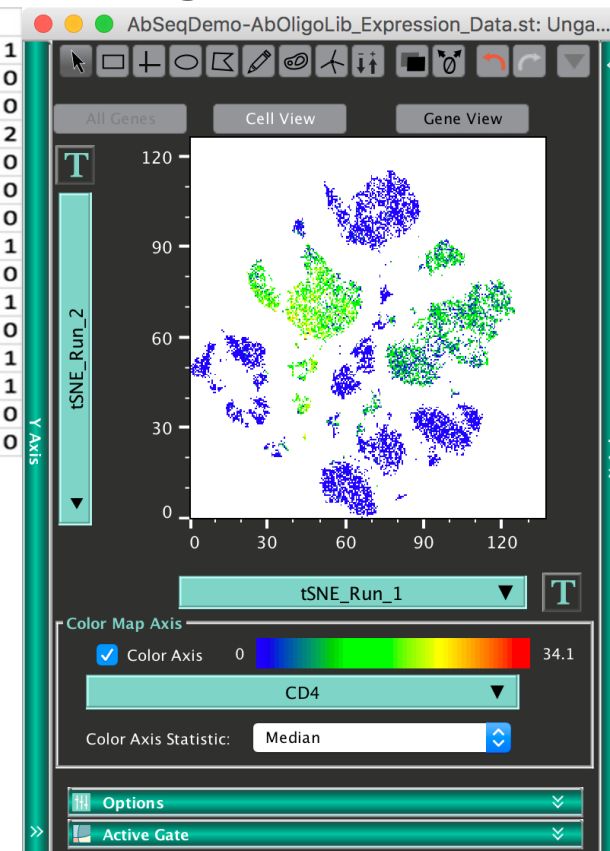
Analyses

Search

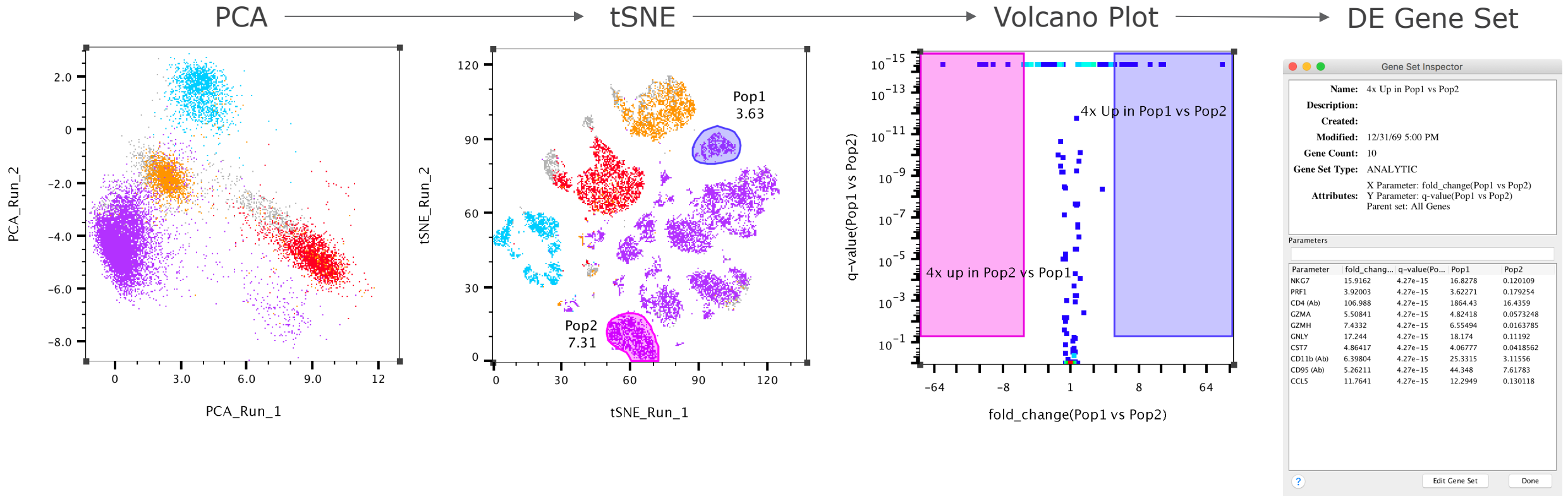
Tasks

Data Cruncher

Your executions will appear here. Before you start.

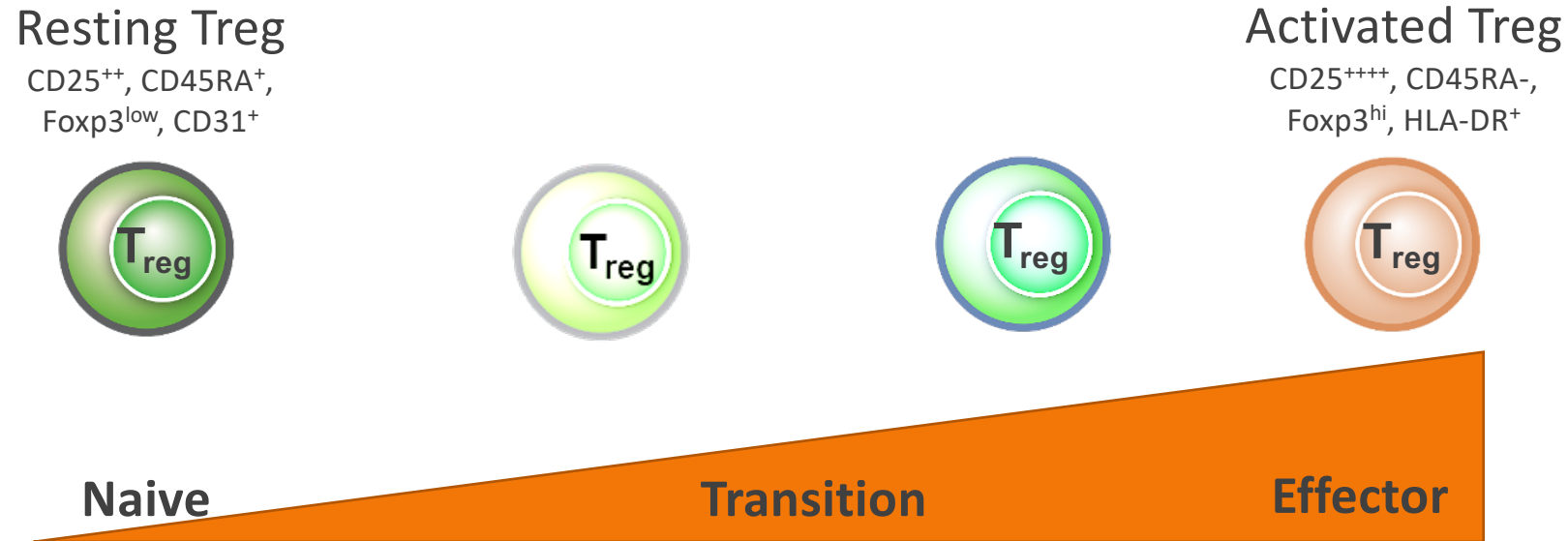


SeqGeq → Protein + RNA Analysis Tools



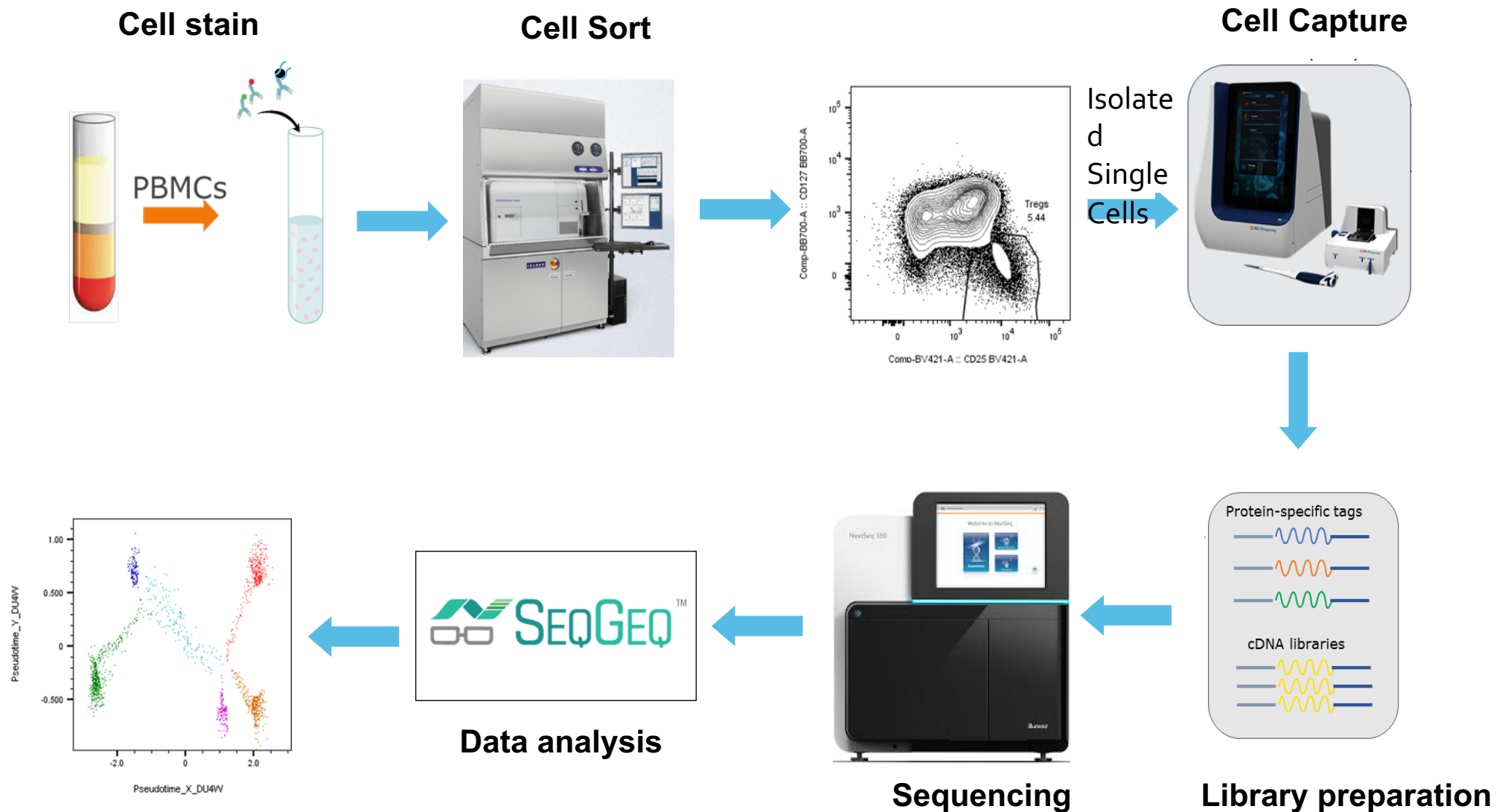
Simple, point and click FlowJo-like interface

Resolution of Treg heterogeneity through single cell multiomic analyses tools

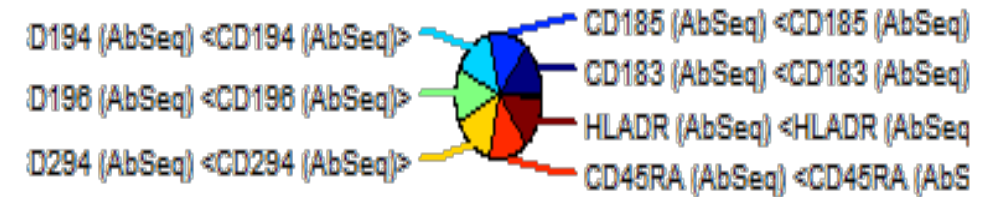
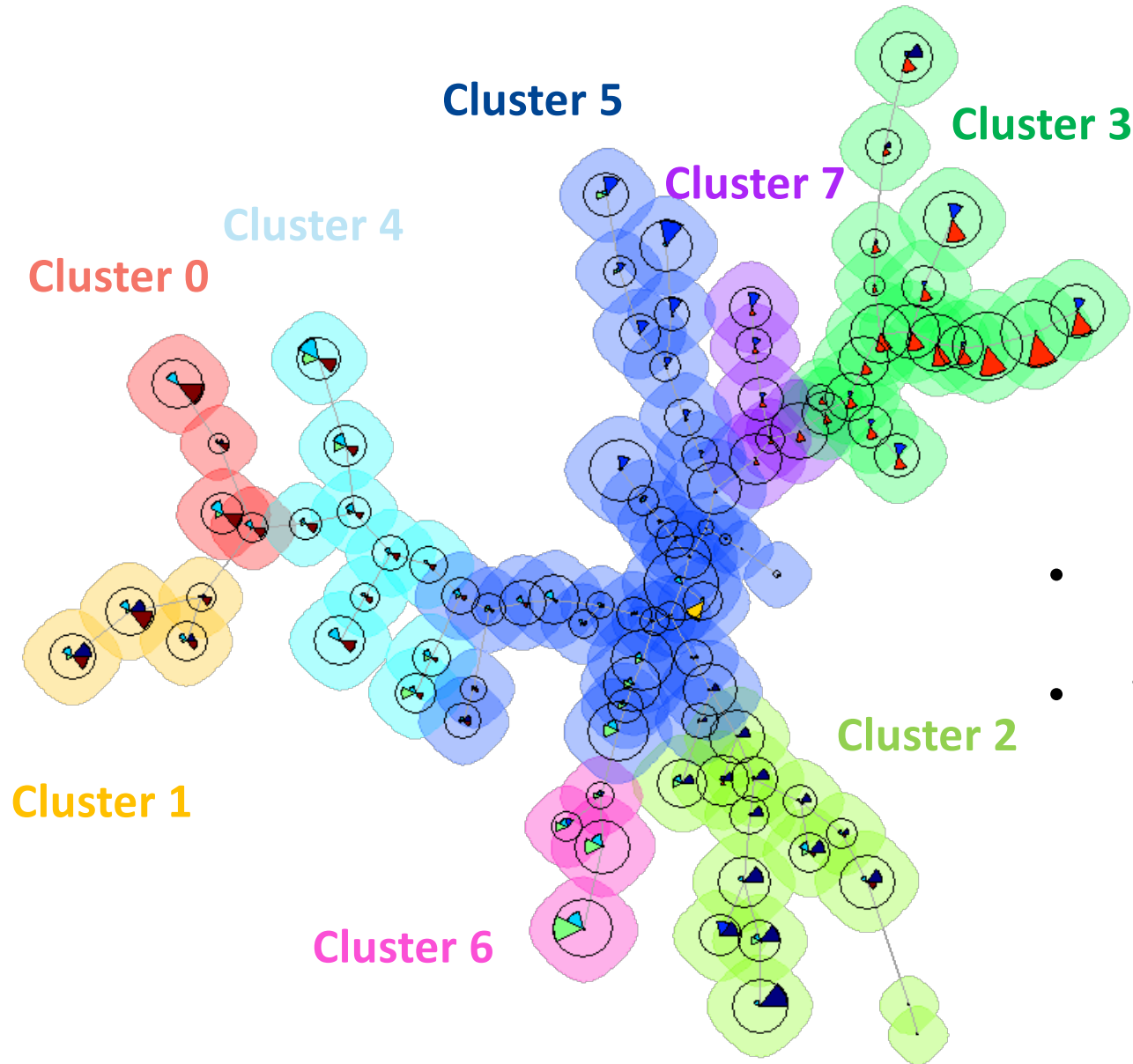


- HLA-DR and CD45RA expression defines 3 main T-reg subsets (naïve, memory, memory/activated)
- Distribution of these subsets can be altered in disease
- These markers are expressed over a continuum (secondary antigens)

AbSeq Experimental Workflow

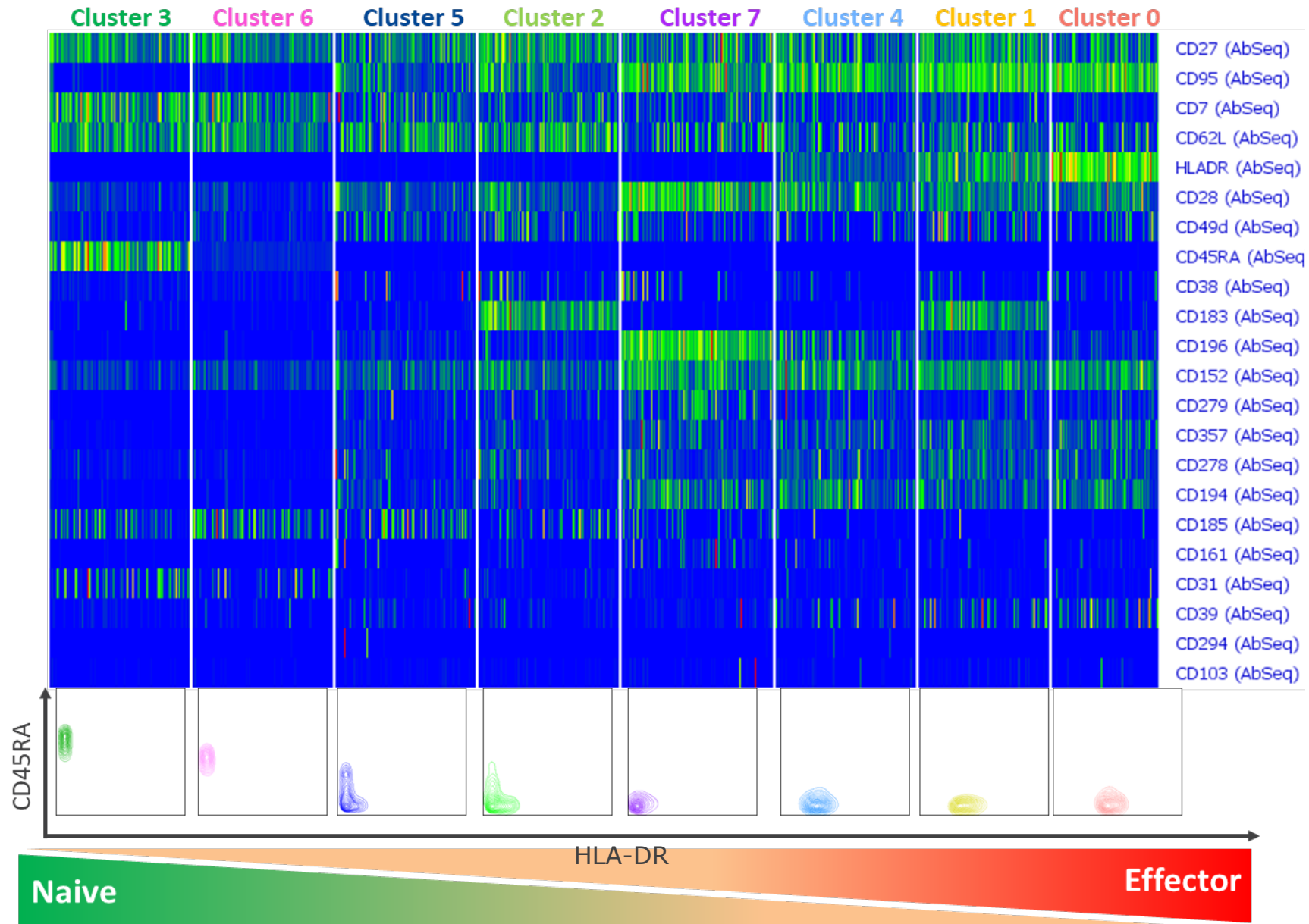


FlowSOM Clustering

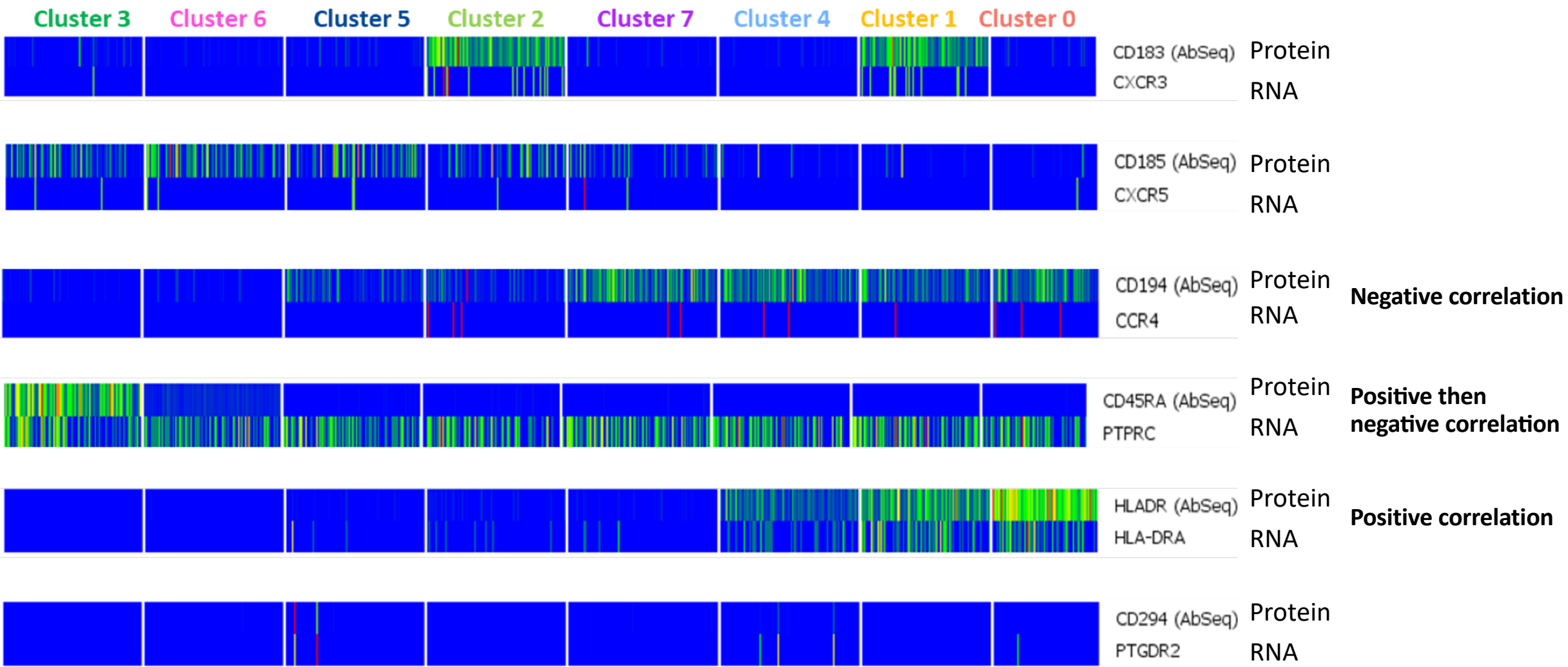


- Eight distinct subsets of Tregs were identified using AbSeq protein readouts
- The subsets are defined based on differential co-expression of CD45RA, HLA-DR, CD183, CD185, CD194, CD196 and CD294

Deeper dive into Treg clusters using AbSeq



Correlation between gene and protein expression at the single-cell level



Need access to or Help with SeqGeq?

- Get a free 60 day Trial
<https://www.flowjo.com/solutions/seqgeq/free-trial>
- SeqGeq Quotes, Technical Support and Questions
SeqGeq@bd.com

Manuals and help documents
docs.flowjo.com/seqgeq

Technical Support
flowjosupport@BD.com

BD Rhapsody – AbSeq Info or any other questions
Timothy.Crawford@BD.com



Thank You!

Questions?