

Single Cell RNA-Seq Data Analysis with Partek Flow software

Eric Chen

techsupport@gtbiotech.com.tw

Bioinformatic Specialist

GenetechBiotech

Who is Partek

Mission

To empower scientists to make scientific breakthroughs in human genetics, disease relationships, drug discoveries, diagnoses, and disease treatments.



Founded in

1993

for data mining and artificial
intelligence

Over

8,500

peer-reviewed citations

More than

40,000

researcher questions answered

Customers in over

40

countries

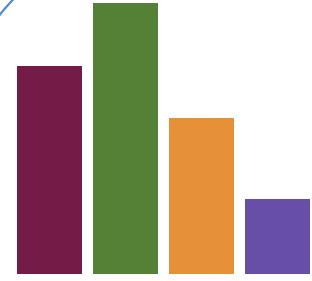
Partek Flow: Start-to-Finish Bioinformatics Solution



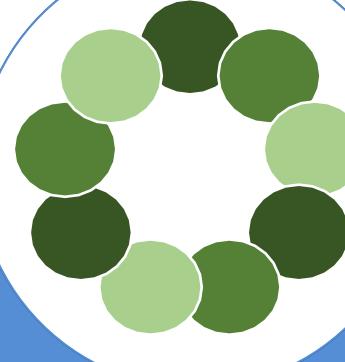
Interactive
Visualizations



Collaborative
Web
Environment



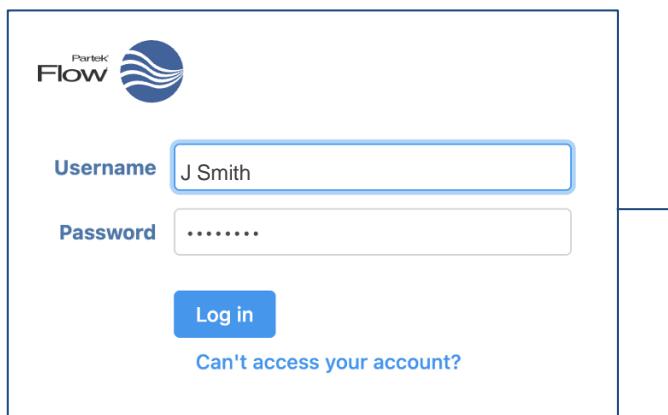
Powerful
Statistics



Comprehensive
Application
Support

User Friendly Analysis and Visualizations

Access from
Your Favorite
Browser



The login interface for Partek Flow. It features a logo at the top left, followed by fields for "Username" (containing "J Smith") and "Password" (containing a series of dots). Below these is a blue "Log in" button. At the bottom is a link "Can't access your account?".



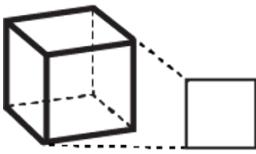
Comprehensive Statistics and Tools



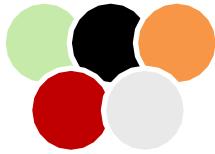
QA/QC



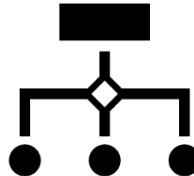
Normalization



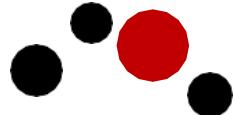
Dimension Reduction



Unsupervised Clustering



Automatic Cell Classification



Batch Removal



Differential Analysis



Cell Type
Abundance
Analysis

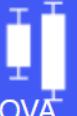


Trajectory Analysis



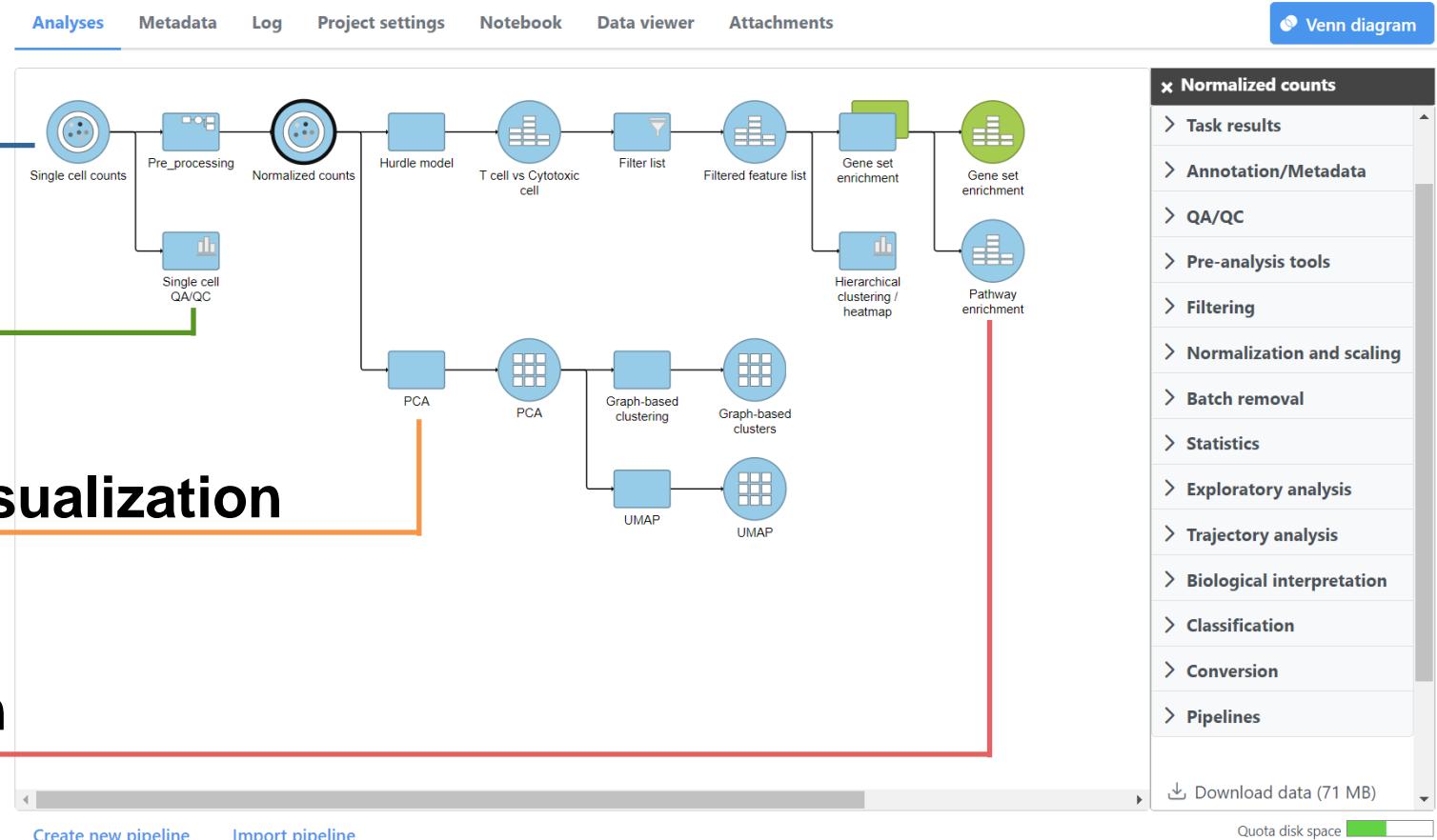
Biological Interpretation

Publicly Available Statistical Algorithms and Tools

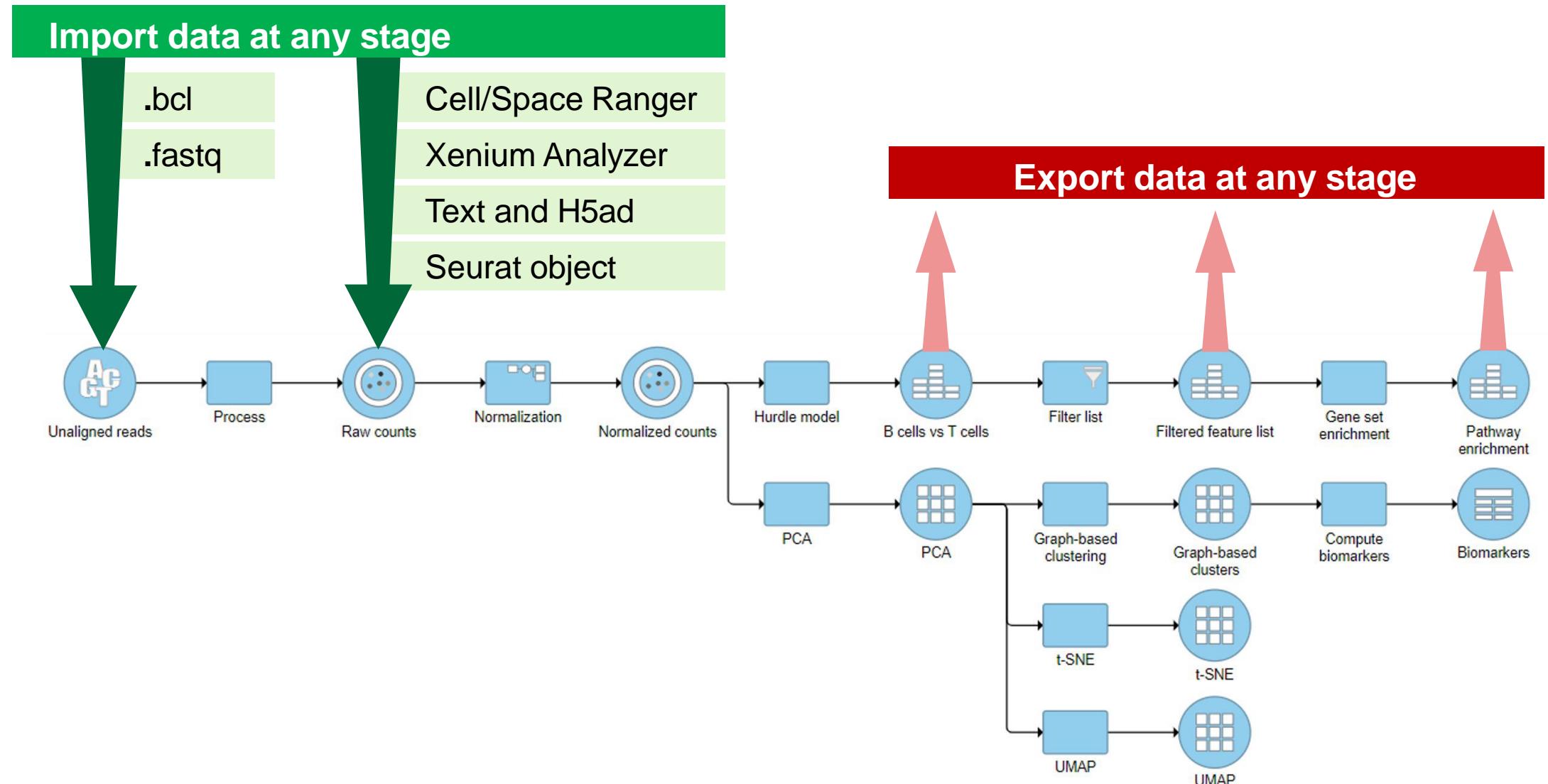
Alignment Bowtie BWA Isaac TopHat TMAP	
Differential analysis Limma DESeq2 Poisson	
Metagenomics Kraken Alpha and beta diversity Quantification at taxonomic levels Differential analysis at taxonomic levels	
QA/QC reports Pre-alignment Post-alignment ERCC spike-in Single cell quality	
Clustering Hierarchical K-means Graph-based	
Data exploration PCA t-SNE Dot plot Box plot Pathway Bar chart Bubble map	
Variant calling Samtools LoFreq CNVkit	
Variant annotation SnpEff dbSNP	
Peak calling MACS2 TSS plot	
Quantification Partek E/M HTSeq	

Visual Analysis Process

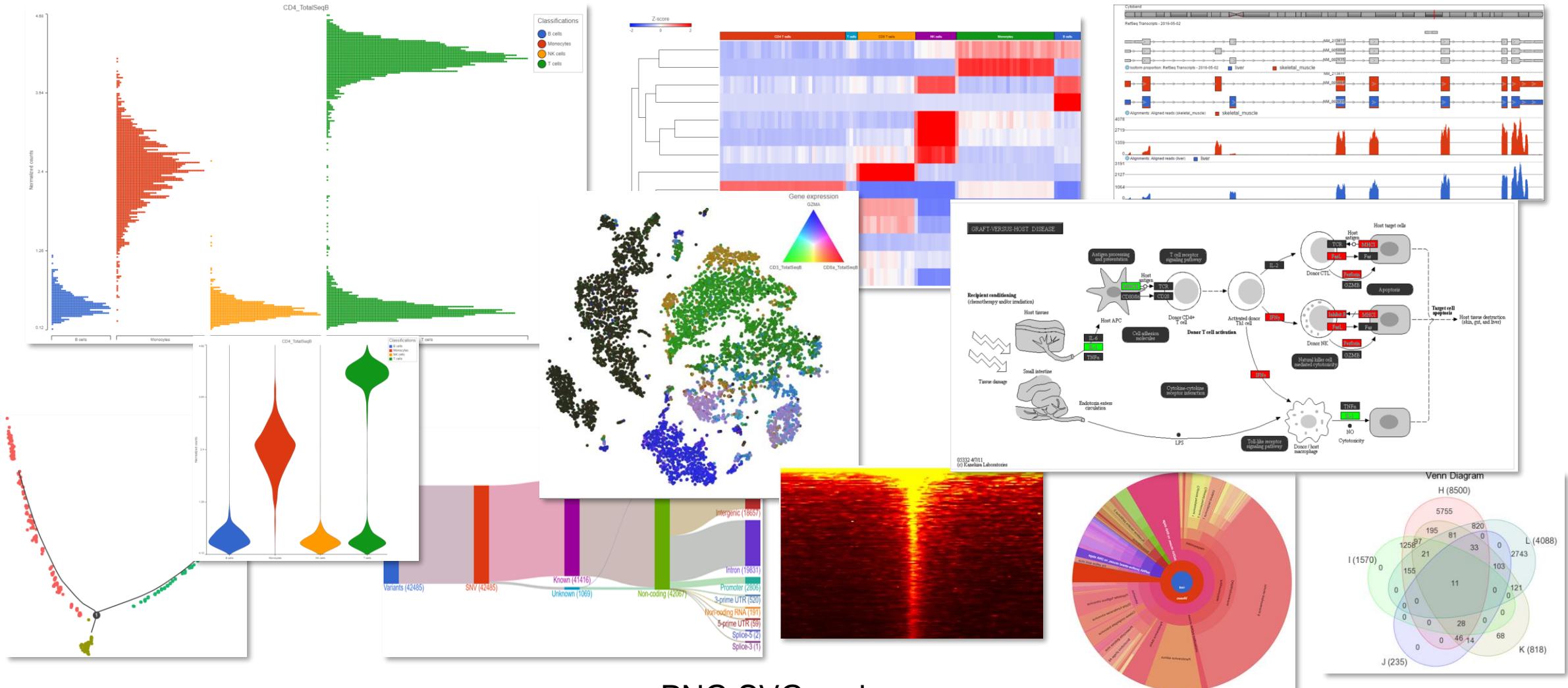
1 Import Data



Import and Export Data at Any Stage



Compelling and Publishable Visualizations



PNG SVG and more

Summary Report

- Who
- When
- What
- How long
- How much

▼ Sample data

👤 Paul Fullerton 📅 28 Aug 2018, 12:24 PM CDT ⚡ 7.97 GB

Show/hide details

▼ Trim bases

Task Trim bases 🤙 Partek support 📅 7 Sep 2018, 03:31 PM CDT ⚡ 00:09:06 ⚡ 34.35 GB

Show/hide details

▼ Filter samples

Task Filter samples 🤙 Partek support 📅 10 Sep 2018, 03:38 PM CDT ⚡ 00:00:00 ⚡ 8.28 GB

Show/hide details

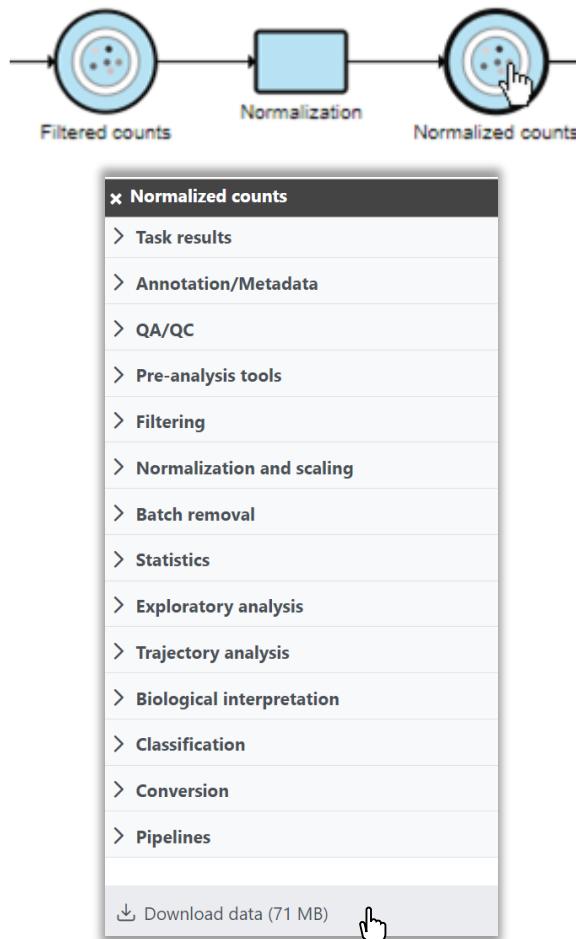
▼ Align reads

Task BWA - 0.7.15 🤙 Partek support 📅 10 Sep 2018, 04:43 PM CDT ⚡ 01:04:31 ⚡ 5.84 GB

Option	Value
Unaligned reads	SRR2163168.fastq.gz, SRR2163168.index, SRR2181401.fastq.gz, SRR2181401.index
Reference index	mm10
Generate unaligned reads	false
Alignment algorithm	BWA-backtrack (Default: BWA-MEM)
Max edit distance	4.0%
Gap openings	1
Gap extensions	-1
3' deletion buffer	10
Indel ends buffer	5
Enable seeding	false
Max edit distance	2
Gap extension penalty	4

Export Data

Choose Any Data



Download in Industry Standard Formats

Files will be available to download from task result

Export format

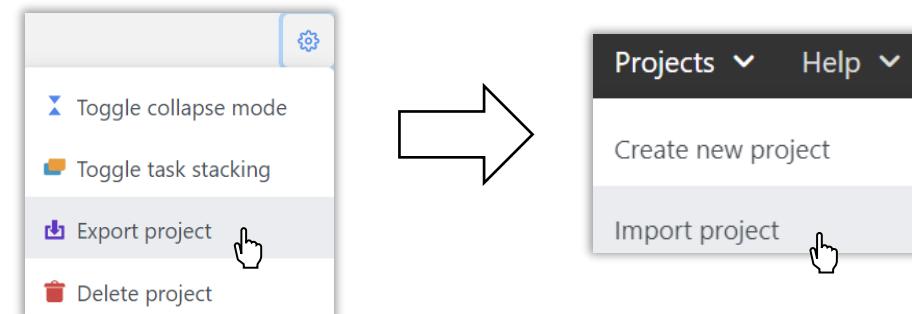
Features on columns (.txt)
 Features on rows (.txt)
 10X CellRanger HDF5 (.h5)

Include content

Annotations Counts

FASTQ, BAM, TXT, and more

Export and Import Analysis Projects



Build, Reuse, and Share Analysis Pipelines

Build Analysis Pipelines

The screenshot shows a pipeline builder interface. On the left, a flowchart illustrates the analysis workflow from mRNA input through various processing steps like Trim bases, STAR, and GSA, leading to final reports like PCA and Gene set enrichment. On the right, a configuration panel allows users to name their pipeline ('Pipeline name: RNA-seq') and choose its section ('Section name: Pipelines'). It also includes a 'Create pipeline' button and a 'Cancel' button.

Click on the tasks above to include in the pipeline. Then click **Create pipeline** below.

Pipeline name: RNA-seq Description:

Section name: Pipelines

Create pipeline **Cancel**

Save, Share, and Manage

The screenshot shows a pipeline management interface. A sidebar on the left includes links for Personal (My profile, My preferences) and System (System information, System preferences, Single sign-on, LDAP, Help widget, Logging). The main area displays a table of saved pipelines:

Name	Description	Creation date	Creator	Ignore	Actions
Agilent Gene Expression Pipeline		11 Dec 2023, 09:45 PM CST	[redacted]	<input type="checkbox"/>	Download pipeline
lncRNA Pipeline		11 Dec 2023, 09:45 PM CST	[redacted]	<input type="checkbox"/>	Share pipeline
Dolomite Bio Drop-Seq v2		11 Dec 2023, 09:45 PM CST	[redacted]	<input type="checkbox"/>	Delete pipeline
Exome germline variant detection		11 Dec 2023, 09:45 PM CST	[redacted]	<input type="checkbox"/>	

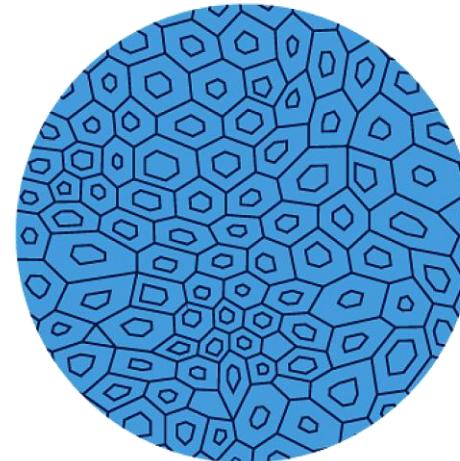
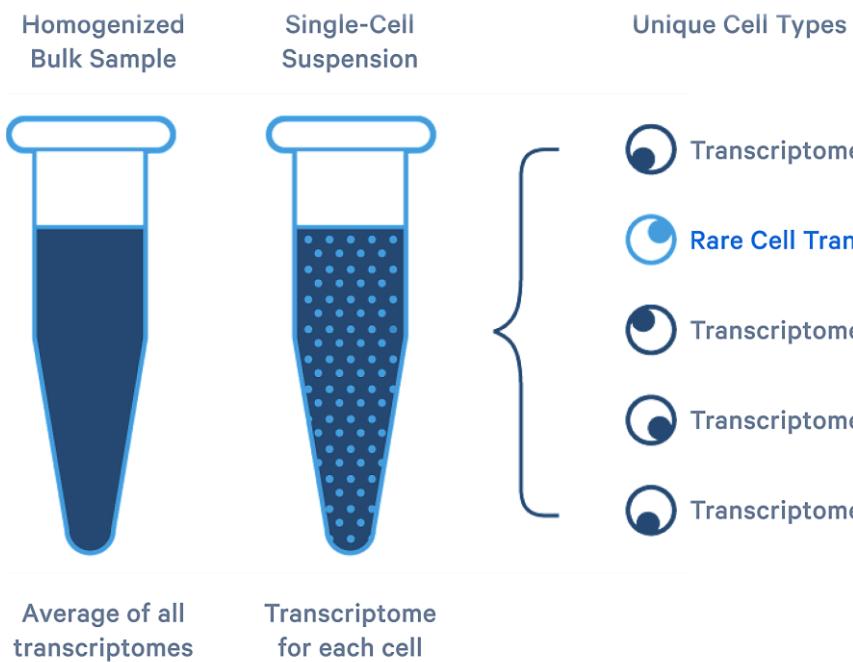
Compatible with All Major Genomics Formats and Assays



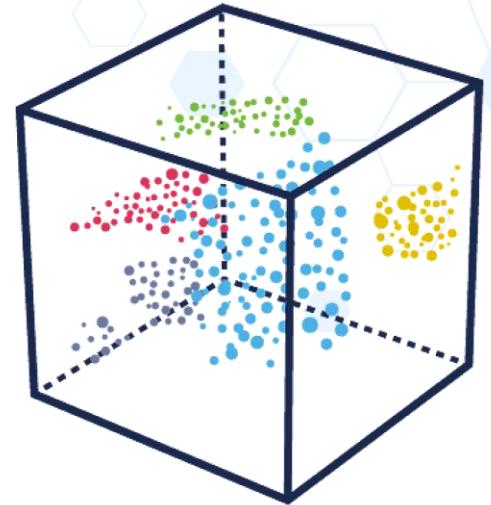
Available Toolkits

- RNA-Seq
 - DNA-Seq
 - Metagenomics
 - Microarray
 - ChIP-Seq
 - Single Cell

Introduction of Single-cell Analysis



Tissue Specimen with a spatial relationship between cells.

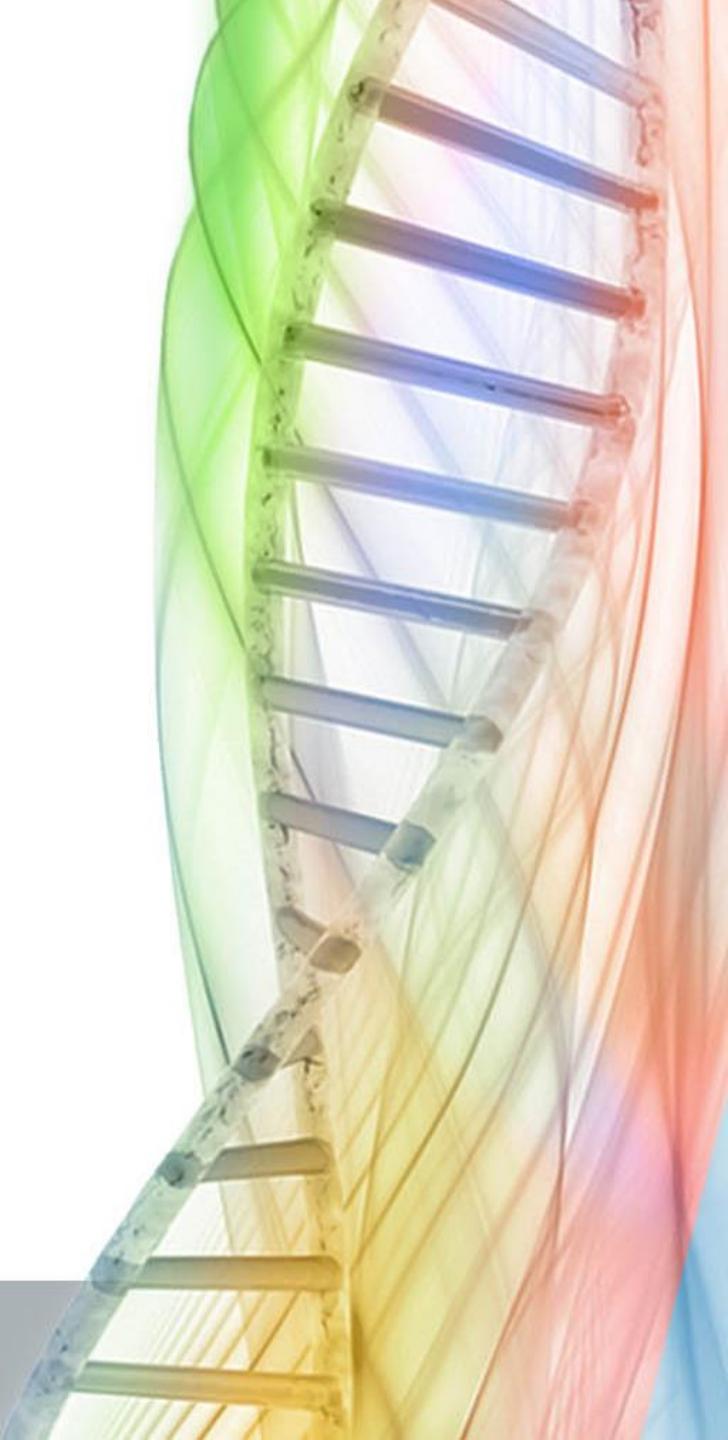


Relationship between cells by similarity of gene expression.



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Single Cell Analysis

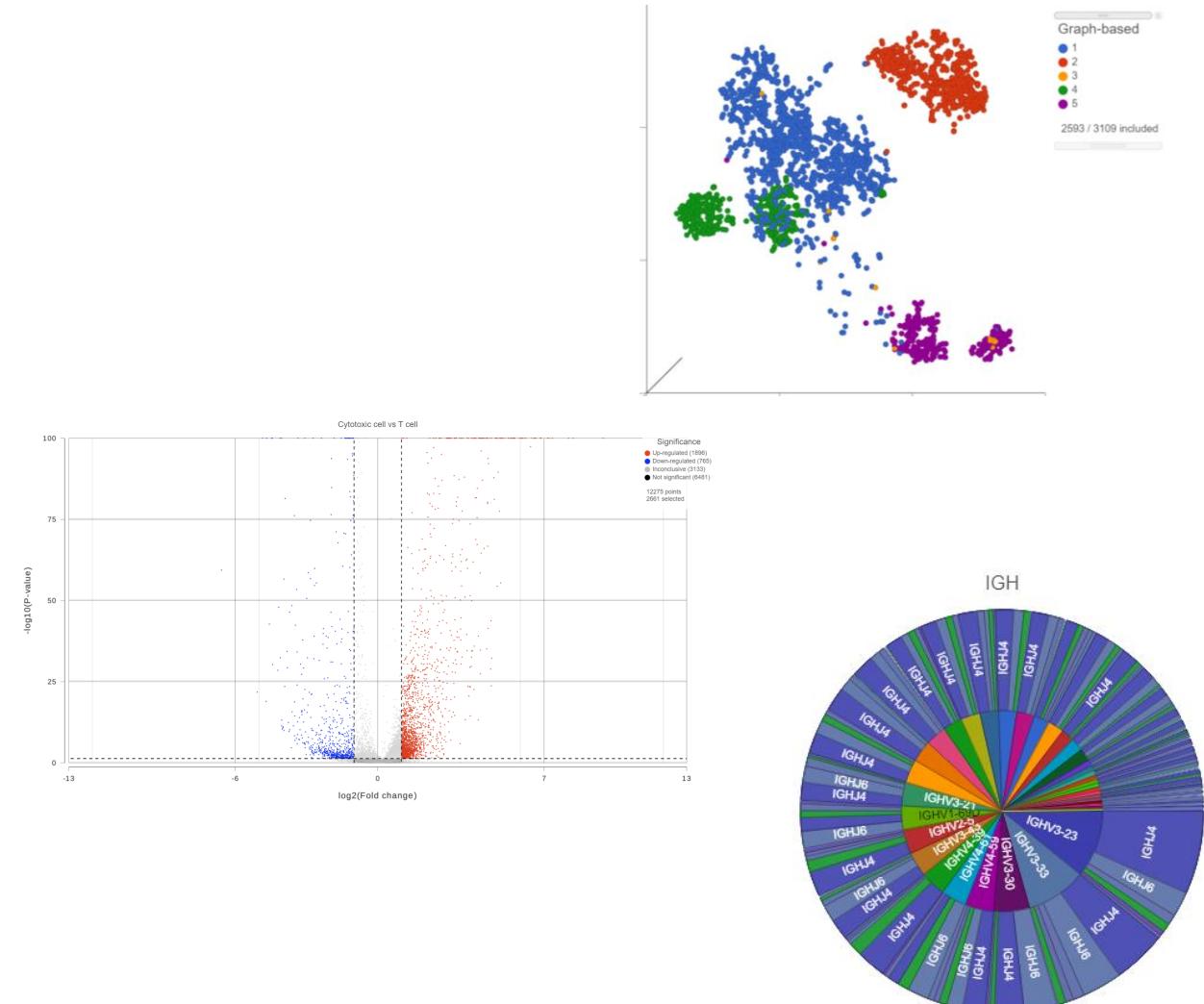


Supports All Major Single Cell Platforms

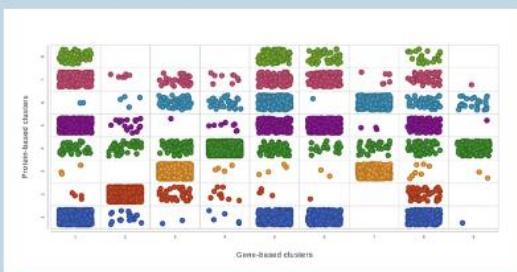


Support for Wide Variety of Single Cell Technologies

- ✓ Single Cell RNA-Seq
- ✓ Single Nucleus RNA-Seq
- ✓ CITE-Seq
- ✓ ECCITE-Seq
- ✓ Spatial Transcriptomics
- ✓ Feature Barcoding
- ✓ V(D)J

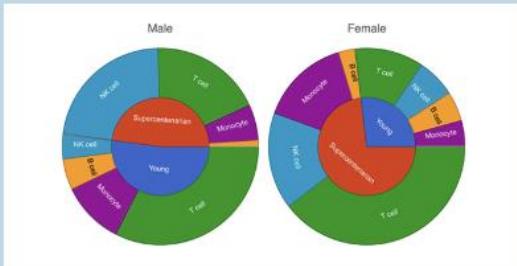


Interactive Visualizations in Partek Flow



Multomics Analysis

Easily integrate and visualize RNA-Seq and ATAC-Seq, or other assays



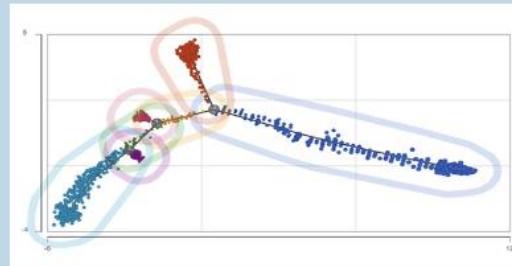
Cell Type Abundance

Determine cell type abundance using a variety of plot types



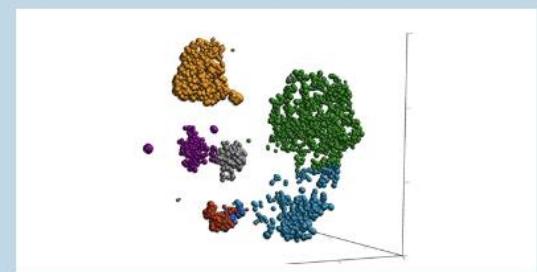
Spatial Transcriptomics

Overlay gene expression data to visualize spatial morphology



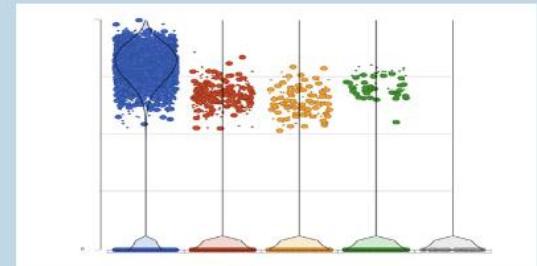
Trajectory Analysis

Analyze biological processes using trajectories



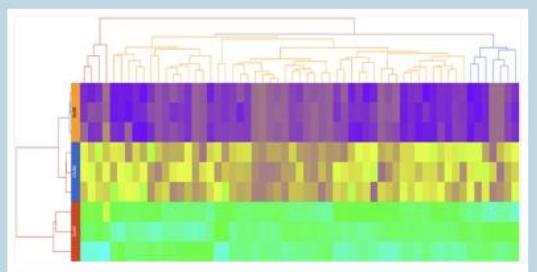
Classification of Cells

Classify cells by traditional methods or automatically



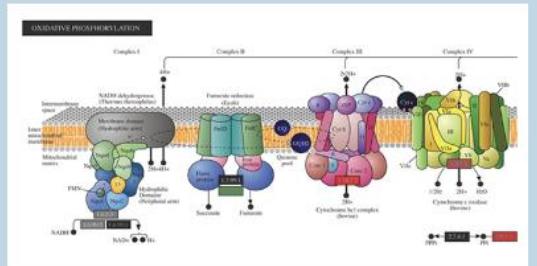
Differential Analysis

Detect gene expression by type using a dot/violin plot



Hierarchical Clustering

Customize heatmaps based on gene lists

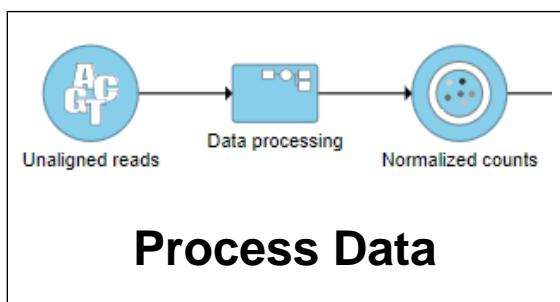


Biological Interpretation

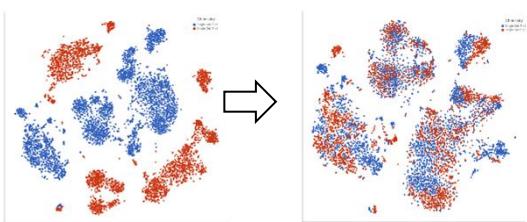
Discover meaningful biological insights using integrated pathways

Data Processing and Analysis, All in One Place

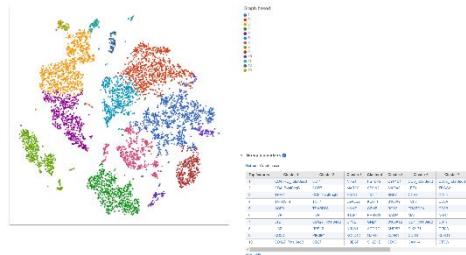
FASTQ, BAM, or Counts



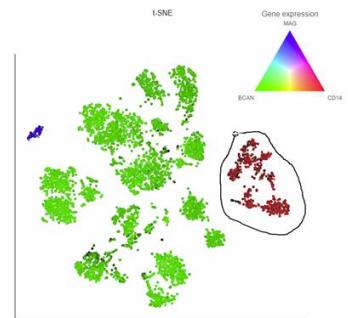
Batch correction



Clustering

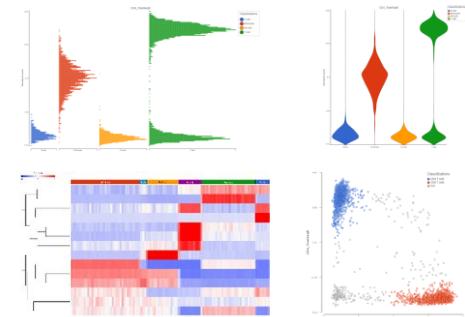


Interactive Visual Analysis

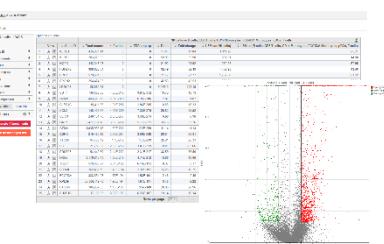


Data Visualization

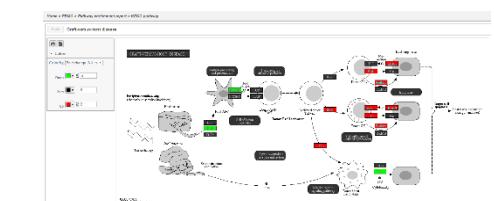
Data Visualization



Differential Analysis



Pathway Analysis



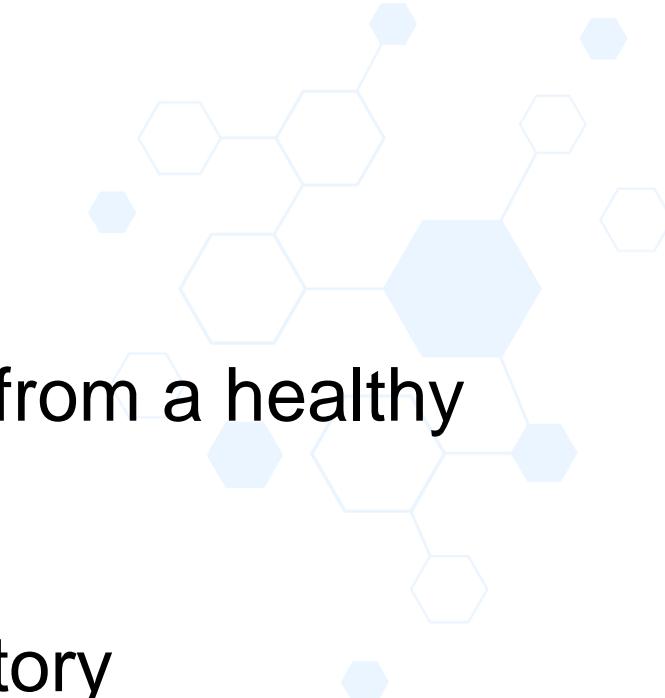
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Demo



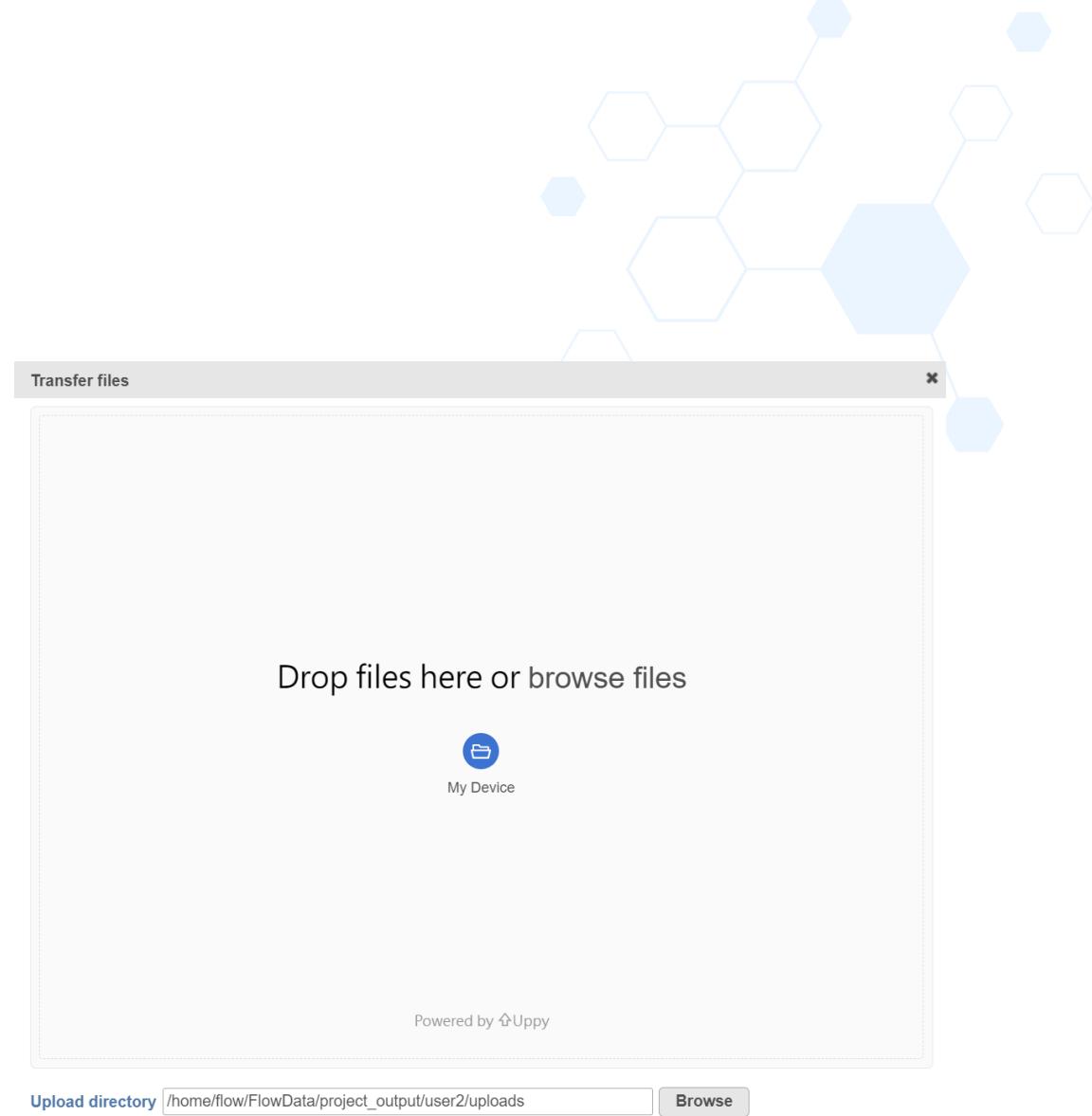
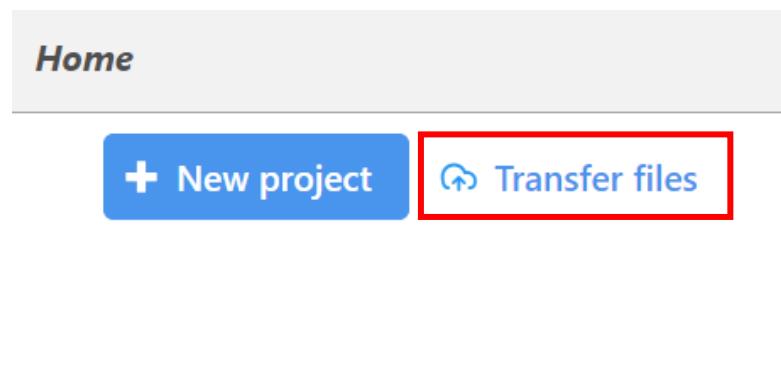
Experiment Description

- 5k peripheral blood mononuclear cells (PBMCs) from a healthy donor
 - Any peripheral blood cell having a round nucleus
- Downloaded from 10X Genomics' dataset repository
 - http://cf.10xgenomics.com/samples/cell-exp/3.0.2/5k_pbmc_v3/5k_pbmc_v3_filtered_feature_bc_matrix.h5
- Partek Flow supports file types: bcl, fastq, bam, h5, txt etc.
- Goal: **Identify different blood cell populations**



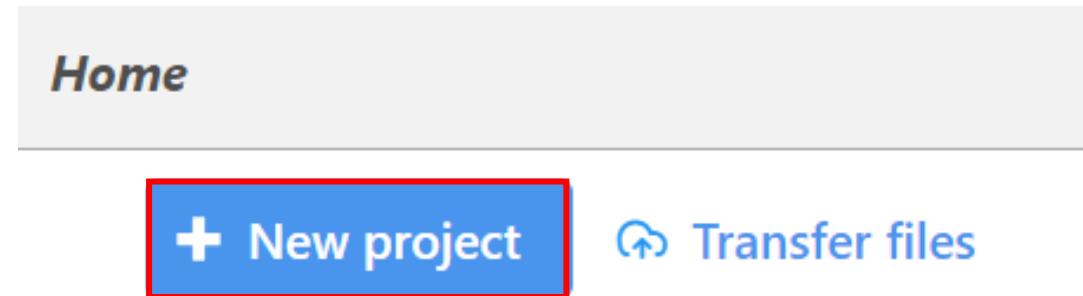
Transfer files

- To move files from your local computer to the Partek server, please **Transfer files** first



Create a new project

- Click **New project** from home page



Import your own data



Single cell Bulk Microarray Other

scRNA-Seq Spatial transcriptomics scATAC-Seq V(D)J Flow/Mass Cytometry

Select the format

Import scRNA count feature-barcode-mtx
This sparse matrix output is common for 10x Genomics, Fluent Biosciences and Parse Biosciences. Each sample has 3 files (two .csv with one .mtx or two .tsv with one .mtx for each sample).

10x Genomics Cell Ranger counts h5
This compressed binary format is preferred for 10x Genomics Cell Ranger output. There is 1 filtered .h5 file per sample and multiple files can be selected

Full count matrix
This rectangular cell-by-feature count matrix is common for BD Rhapsody. There is one file for one or more samples (txt, csv, tsv, txt.gz, csv.gz, tsv.gz)

h5ad
This AnnData object in the h5ad file format is for data processed by Scanpy

fastq
The fastq format is used for unaligned reads. Acceptable file types are fastq, fastq.gz, fastq.bz2, fq, fq.gz, fq.bz2

If you want to import your own data

- Select the format
- Select all files and click **Next**

Specify Annotation

- Set **Sample name** to 5k_pbmc
- Click the **Use annotation file** checkbox and set the annotation
 - Assembly: Homo sapiens (human) - hg38
 - Gene annotation: Ensembl transcripts release 110
- Click **Finish** to import sample



Sample names

Sample name	Files	Cells	Features
5k_pbmc	5k_pbmc_v3_filtered_feature_bc_matrix.h5	5025	33538

Feature annotation

Use annotation file

Select the file that has been used to generate the feature counts (e.g. gene or protein information).

Assembly

Homo sapiens (human) - hg38

Annotation model

Ensembl Transcripts release 110 (Taiwan Genetech Biotech)

Primary feature identifier

Feature name (Values: MIR1302-2HG, FAM138A, OR4F5, AL627309.1, AL627309....)

Feature ID (Values: ENSG00000243485, ENSG00000237613, ENSG00000186092,...)

Deduplication method

If the feature ID is not unique, the feature will be summarized by the selected method.

Mean Maximum Sum

Count value format

Raw count Normalized count with log base None

Report

All features Features with non-zero values across all samples

Cells with total read count at least

A low total read count threshold will result in a large number of cells which might take a long time to import

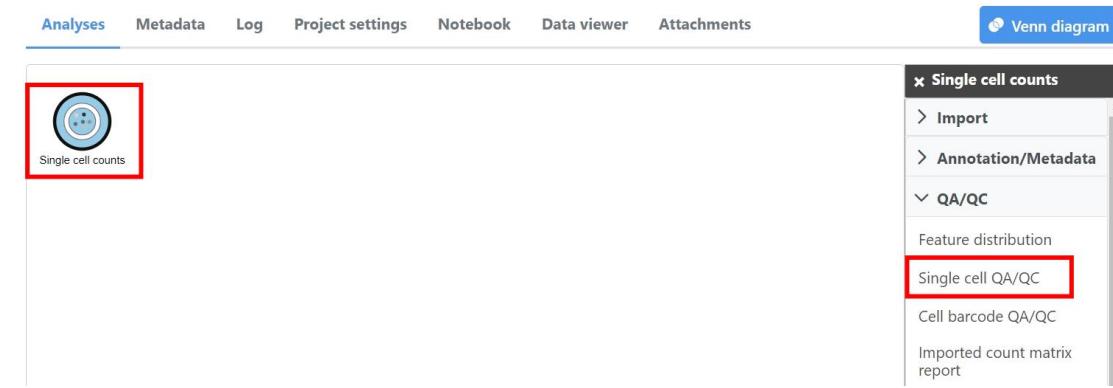
400

Back Finish

Single Cell QA/AC



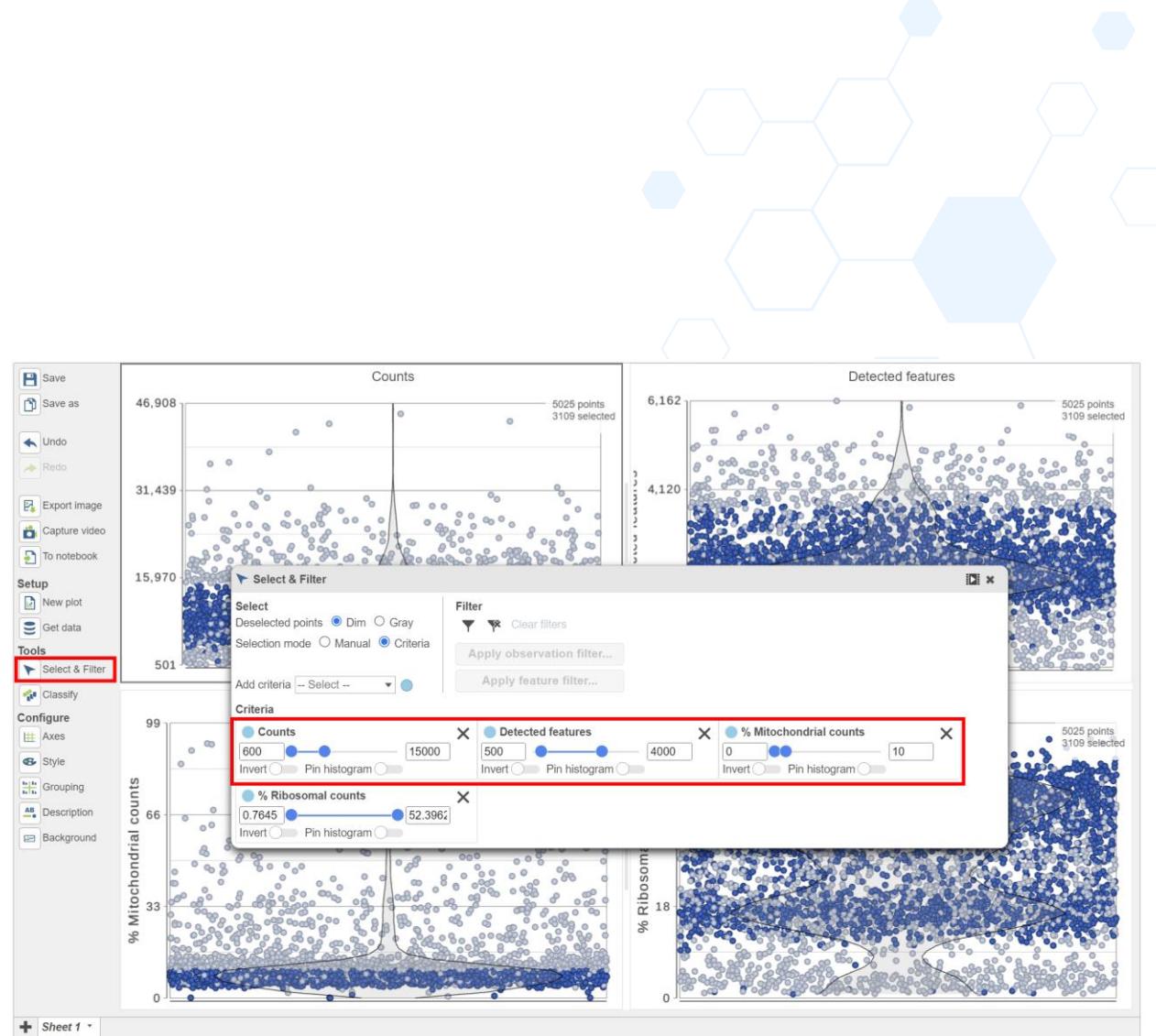
- Go to the **Analyses** tab
- The **Single cell counts** data node appears after the data imported
- Click the data node
- Select **Single Cell QA/QC** from the QA/QC section of the task menu



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Single Cell QA/AC

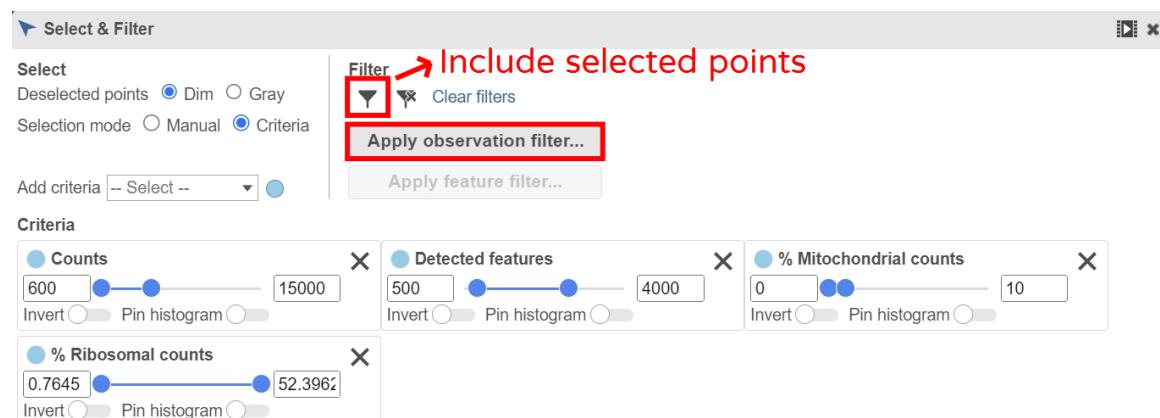
- Double click the **Single Cell QA/QC** task node to open the task report
- Use the **Select & Filter** card to set the Min and Max thresholds:
 - Counts: 600 – 15000
 - Detected features: 500 – 4000
 - Mitochondrial counts 0 – 10



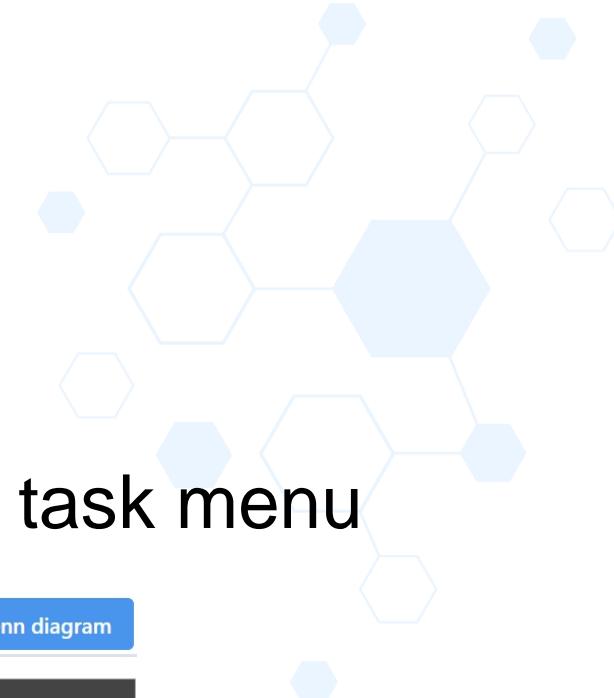
Single Cell QA/AC



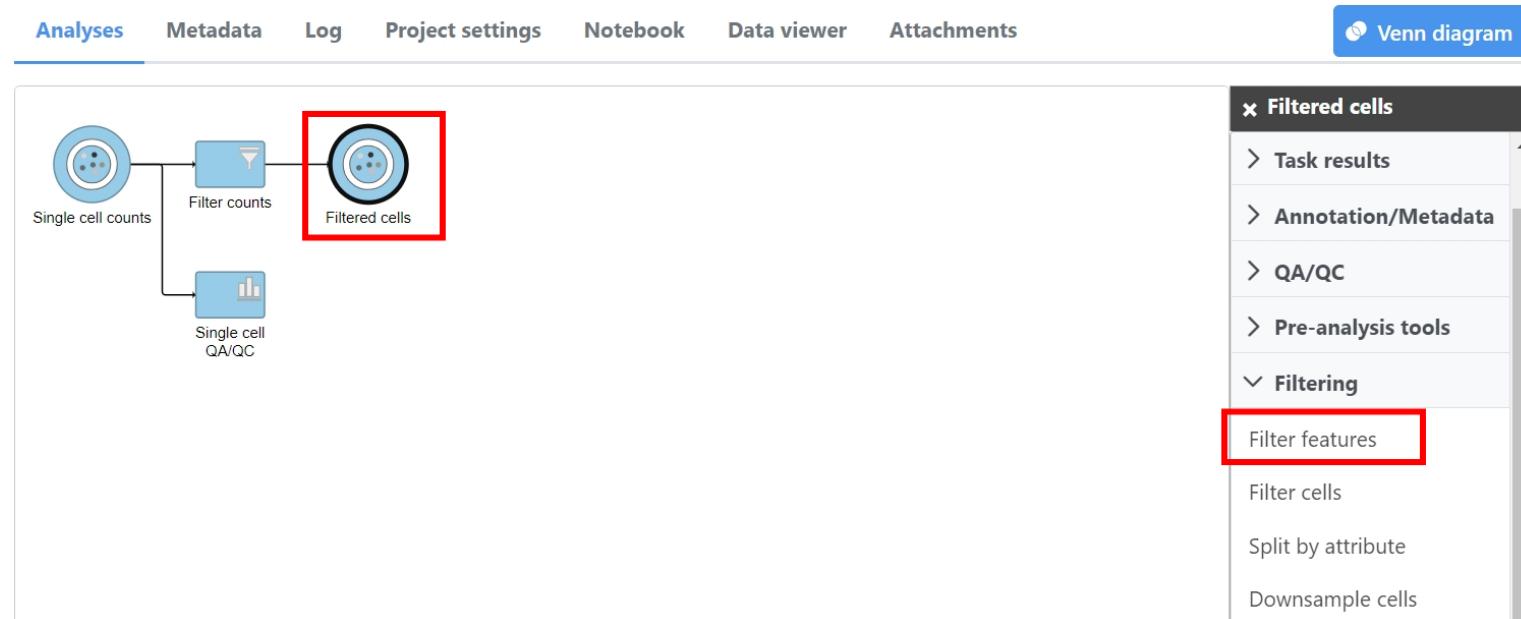
- Select **Include selected points** button
- Select **Apply observation filter...**
- Select the circular **Single cell counts** data node to filter
- Click **OK** on the message in the middle of the screen and click the project name to go back to the Analyses tab
 - This runs the Filter cells task and outputs a new Single cell data node



Applying a Noise reduction filter



- Click the **Filtered cells** data node
- Click **Filter features** in the **Filtering** section of the task menu



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Applying a Noise reduction filter



- Click the **Noise reduction filter** checkbox
- Create the following filter using the drop-downs and text boxes
 - Exclude features where value ≤ 0 in at least 99% of the cells
- Click **Finish** to apply the filter

Filter type

Noise reduction
Exclude features that meet criteria based on descriptive statistics. Calculations are performed for each feature across all cells.

Statistics-based
Include a number or percentile of features based on descriptive statistics. Calculations are performed for each feature across all cells.

Metadata
Specify logical operations using different annotation fields.

Saved list
Specify a saved list of features to include or exclude.

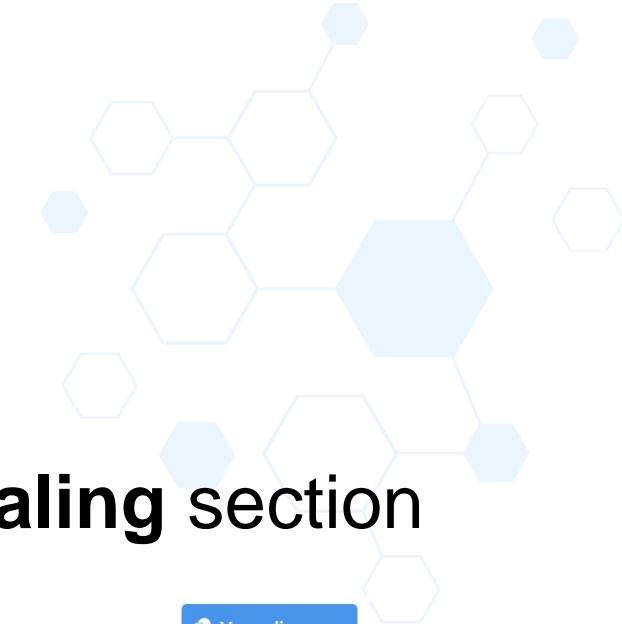
Manual list
Manually specify a list of features to include or exclude.

Filter criteria

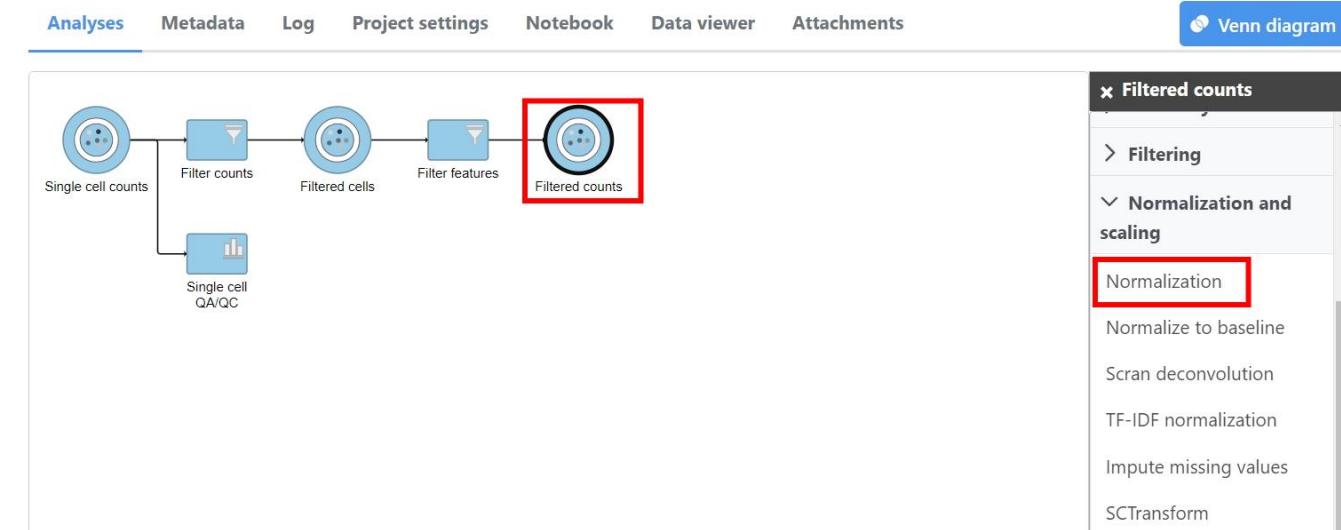
Filter features by

Exclude features where in at least % of the cells

Normalizing counts



- Click the **Filtered counts** node
- Click **Normalization** in the **Normalization and scaling** section of the task menu



Normalizing counts

- Click on the **Recommended** button
- Click **Finish** to run

Count normalization

Transform on

Cells Features

Available methods

- Absolute value
- Add
- Antilog
- Arcsinh
- CLR
- CPM (counts per million)
- Divide by
- Log
- Logit
- Lower bound
- Median ratio (DESeq2 only)

Drag and drop ➔

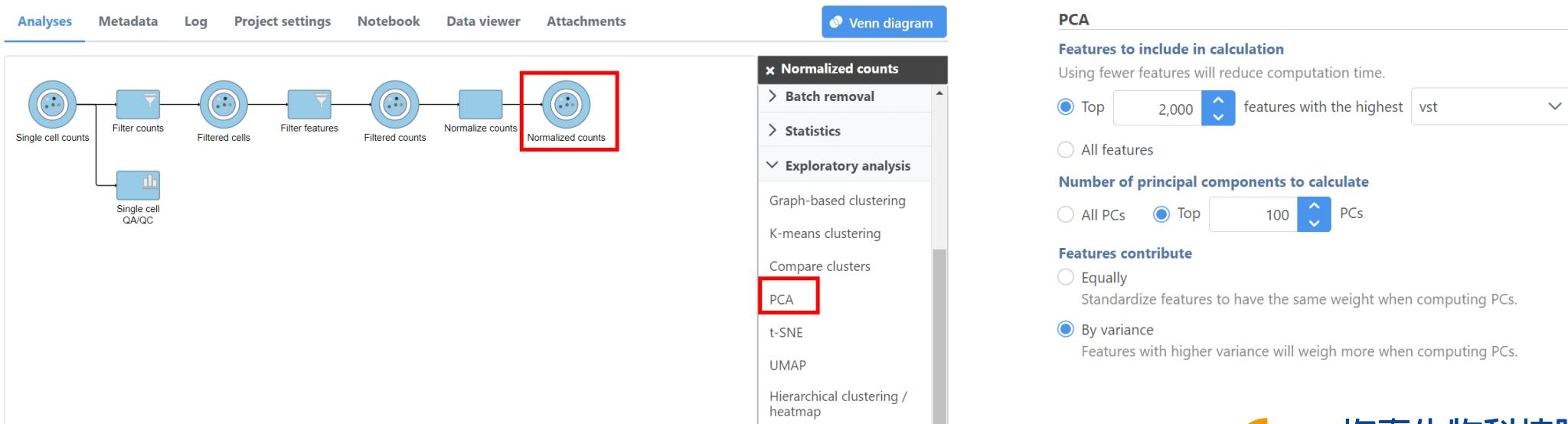
Selected methods

Use recommended

1. CPM (counts per million)
2. Add
3. Log

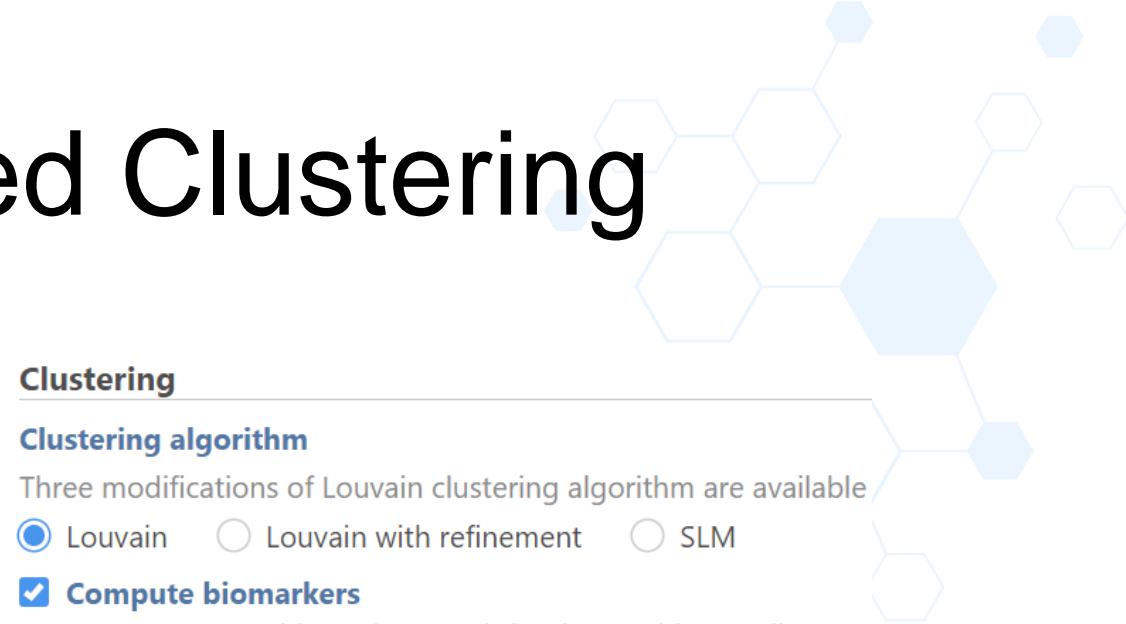
Performing Principal Components Analysis

- Click the **Normalized counts** data node
- Click **PCA** in the **Exploratory analysis** section
- Click **Finish** to run with default settings



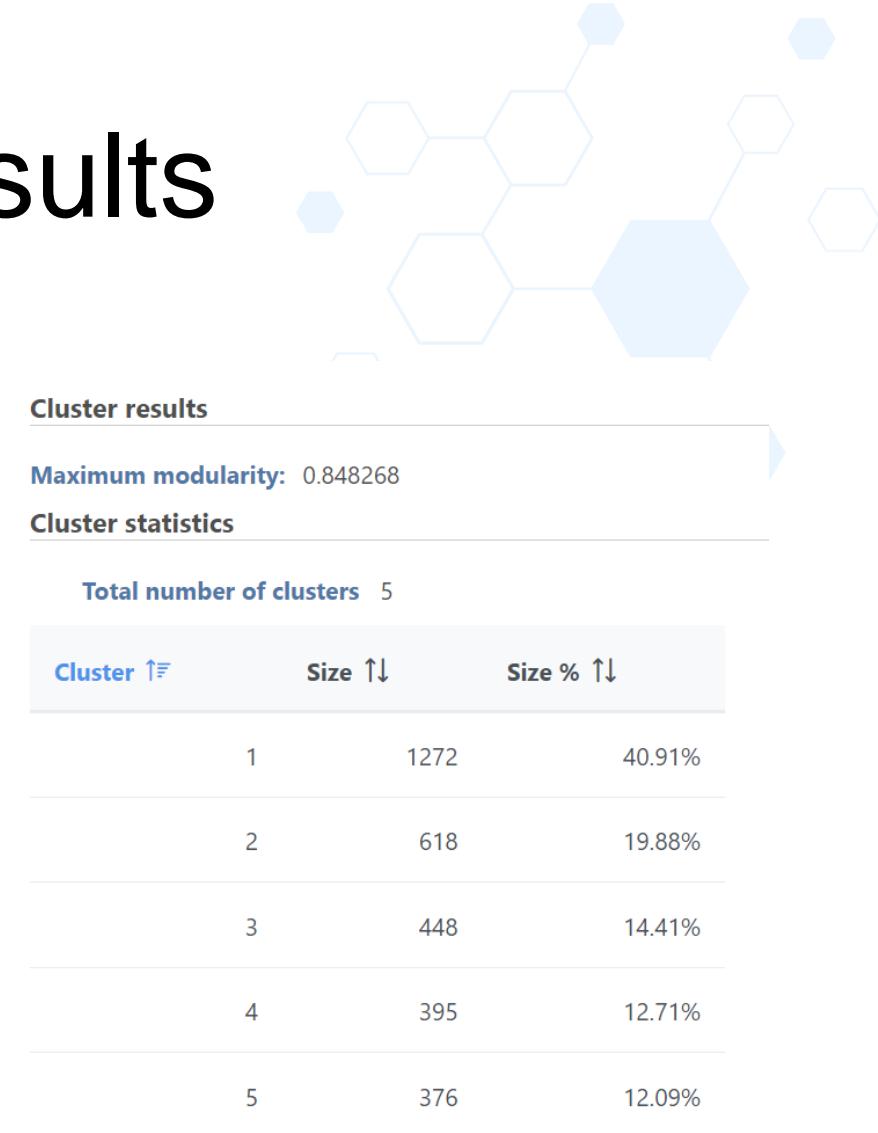
Performing Graph-based Clustering

- Click the **PCA** data node
- Click **Graph-based clustering** in the **Exploratory analysis** section of the task menu
- Click **Finish** to run with default settings



Graph-based Clustering Results

- Double-click the **Graph-based clusters** data node to open the Task report
- The *Maximum modularity* is a measure of the quality of the clustering result. Higher modularity (close to 1) indicates a better result
- The *Cluster statistics* shows the number of clusters, cluster size and the percentage of cells in each cluster



Biomarkers Results

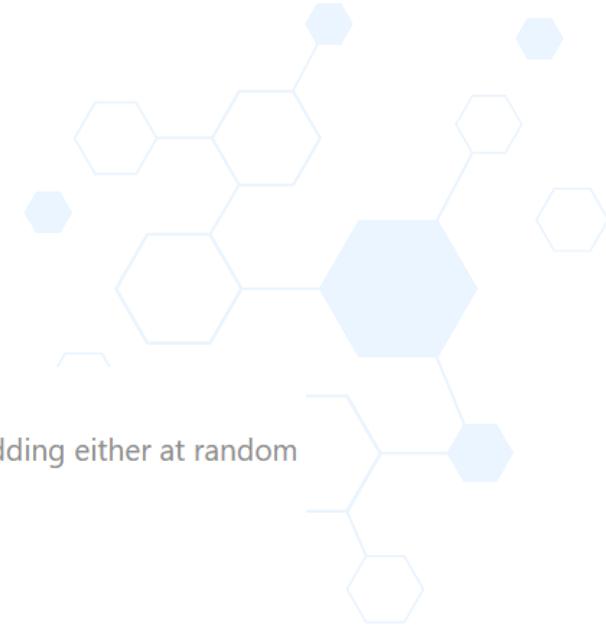
- Double-click the **Biomarkers** data node

Biomarkers for Graph-based					
Top features ↑	Cluster 1 ↑	Cluster 2 ↑↓	Cluster 3 ↑↓	Cluster 4 ↑↓	Cluster 5 ↑↓
1	TRABD2A	S100A8	TNFRSF4	IGKC	FGFBP2
2	LEF1	S100A9	LMNA	IGHM	GNLY
3	CCR7	S100A12	AQP3	IGHD	GZMH
4	TCF7	LYZ	IL32	TCL1A	NKG7
5	TPT1	FCN1	KLRB1	MS4A1	KLRD1
6	RPL35A	CD14	MAF	CD79A	ADGRG1
7	RPS15A	VCAN	IL7R	VPREB3	KLRF1
8	RPS27A	MNDA	NPDC1	JCHAIN	PRSS23
9	LRRN3	CSTA	SYNE2	SPIB	SPON2
10	CD3E	SERPINA1	NSG1	BANK1	PRF1



Perform UMAP

- Click the **Graph-based clusters** data node
- Click **UMAP** in the **Exploratory analysis** section
- Click **Finish** to run the UMAP task with default settings



Initialize output values
Initialize the low dimensional embedding either at random

PCA

Number of principal components to calculate

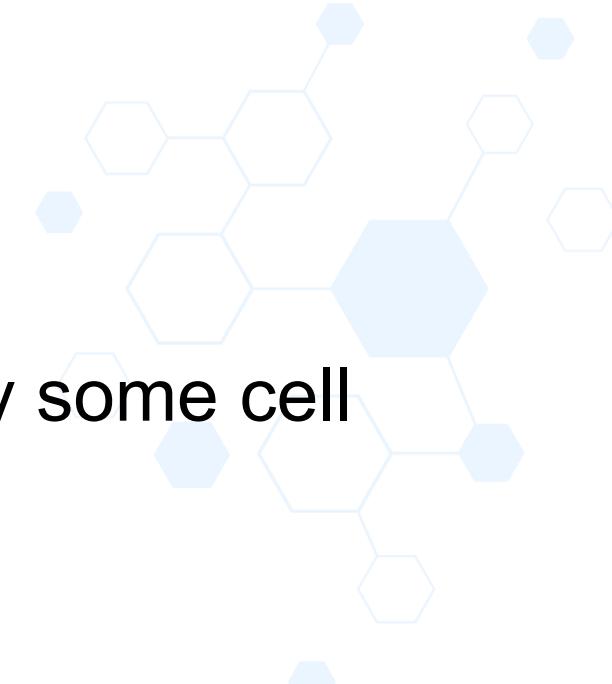
All PCs Top

Advanced options

Option set

-- Default -- [Configure](#)

Identifying Cell Types

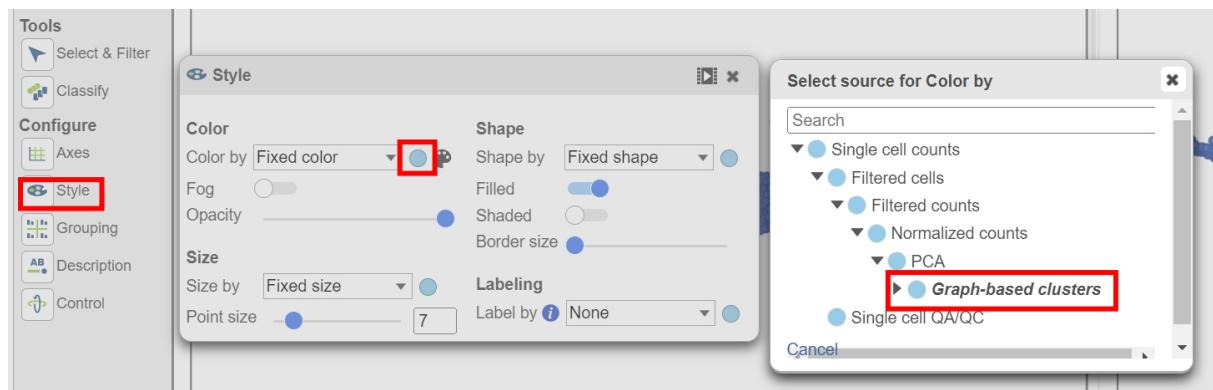


- We'll be using a combination of methods to identify some cell types commonly found in PBMCs. Namely:
 - Unbiased clustering (Graph-based)
 - Visualizing expression using
 - Canonical gene markers
 - Gene lists
 - Lassoing cell populations on the plot

Cell Type	Gene Markers
T-cells	CD3D, CD3E
Cytotoxic cells	NKG7, GNLY
B cells	CD79A, CD79B (list)
Monocytes	CD68, CD14

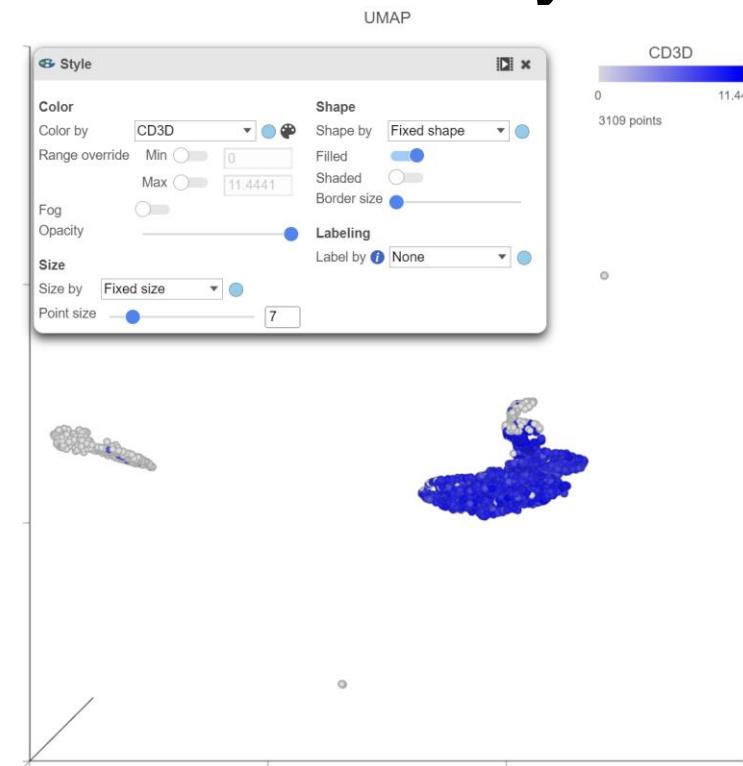
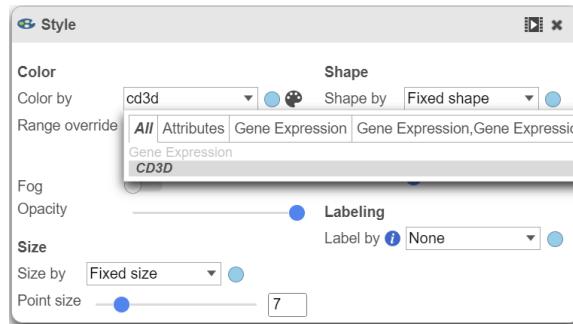
Classify T cells

- Duplicate the UMAP plot by clicking 
- Color one of the plots using Graph-based classification
 - Click **Style** and **Select source for Color by** as Graph-based clusters
 - Set **Color by** as **Graph-based**



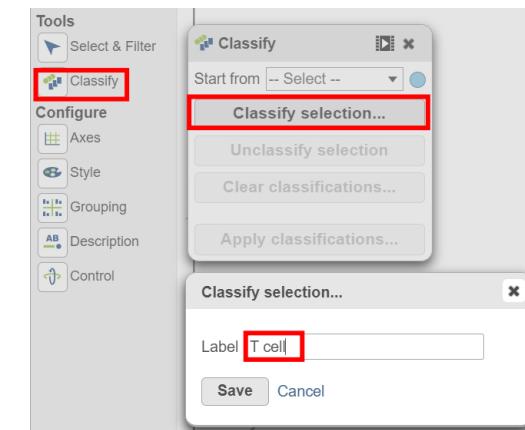
Classify T cells

- Click on the other UMAP plot
- Color the plot using a gene marker, CD3D
 - Click **Style** and **Select source for Color by** as Normalized counts
 - Enter **CD3D** in the box

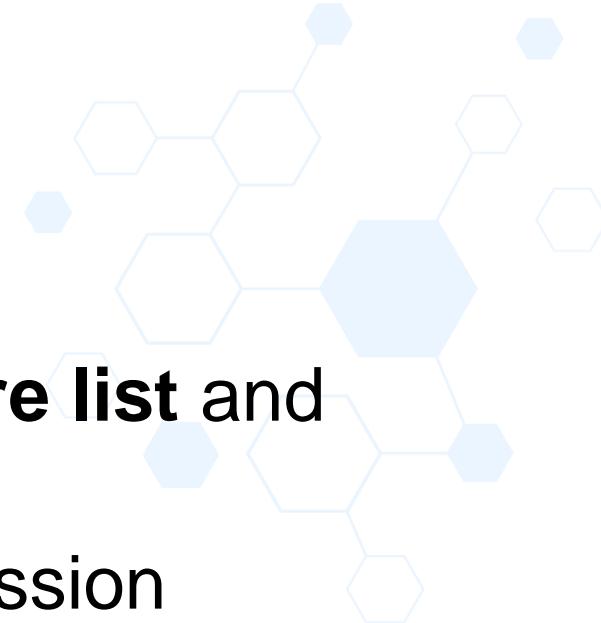


Classify T cells

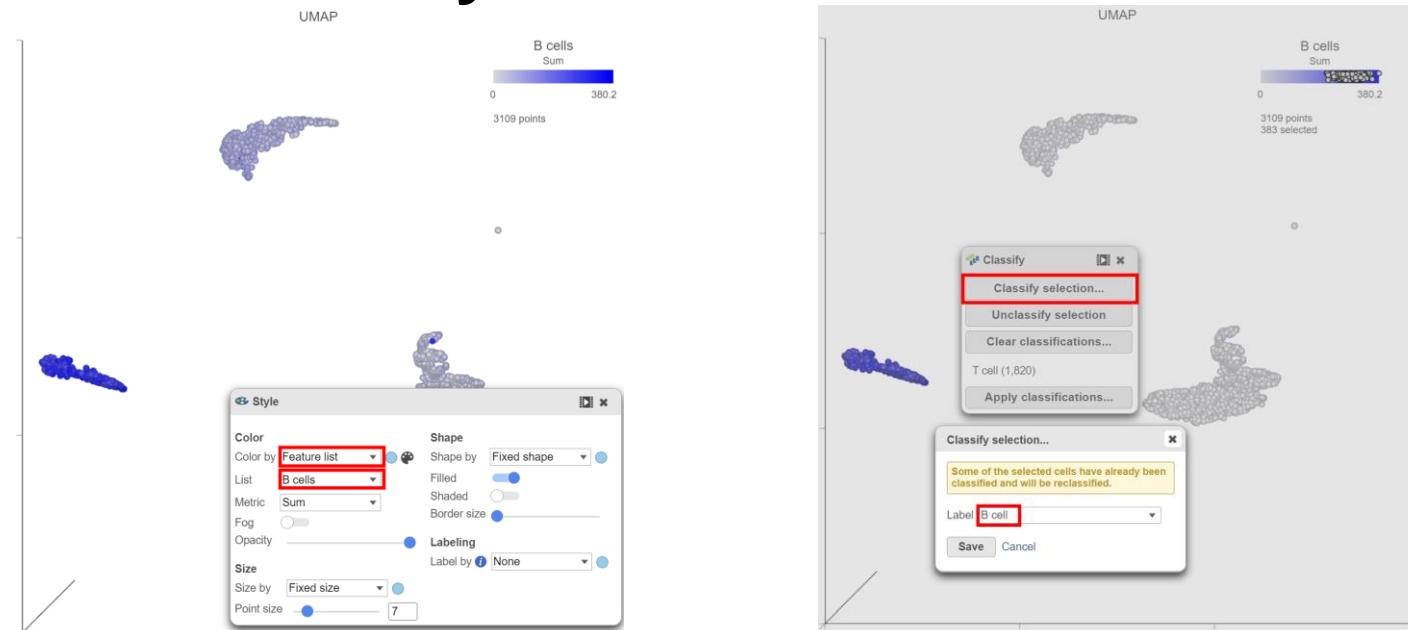
- Click **Select & Filter**
- Add criteria as **Graph-based** and choose 1 and 3
- Click **Classify** and **Classify selection...**
- Specify the name of selected cells as **T cell** and click **Save**



Classify B cells



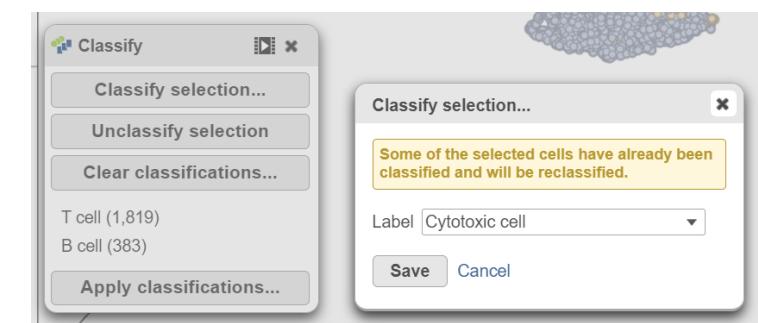
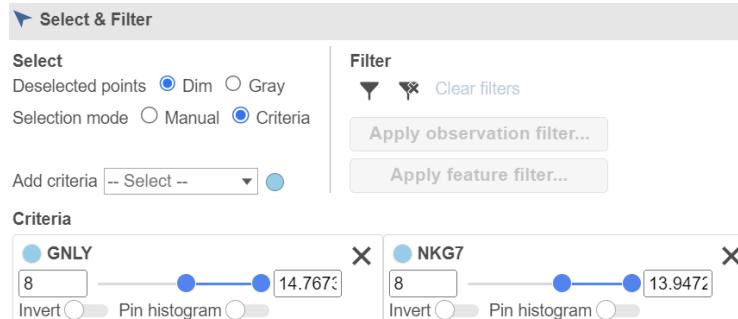
- Select the 2nd UMAP plot, choose Color by **Feature list** and select **B cells**
- Use lasso tool to select the cells with high expression
- Click on **Classify selection** to name selected cells as **B cell**



Classify Cytotoxic cells

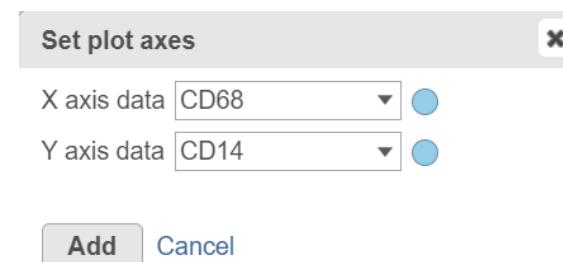
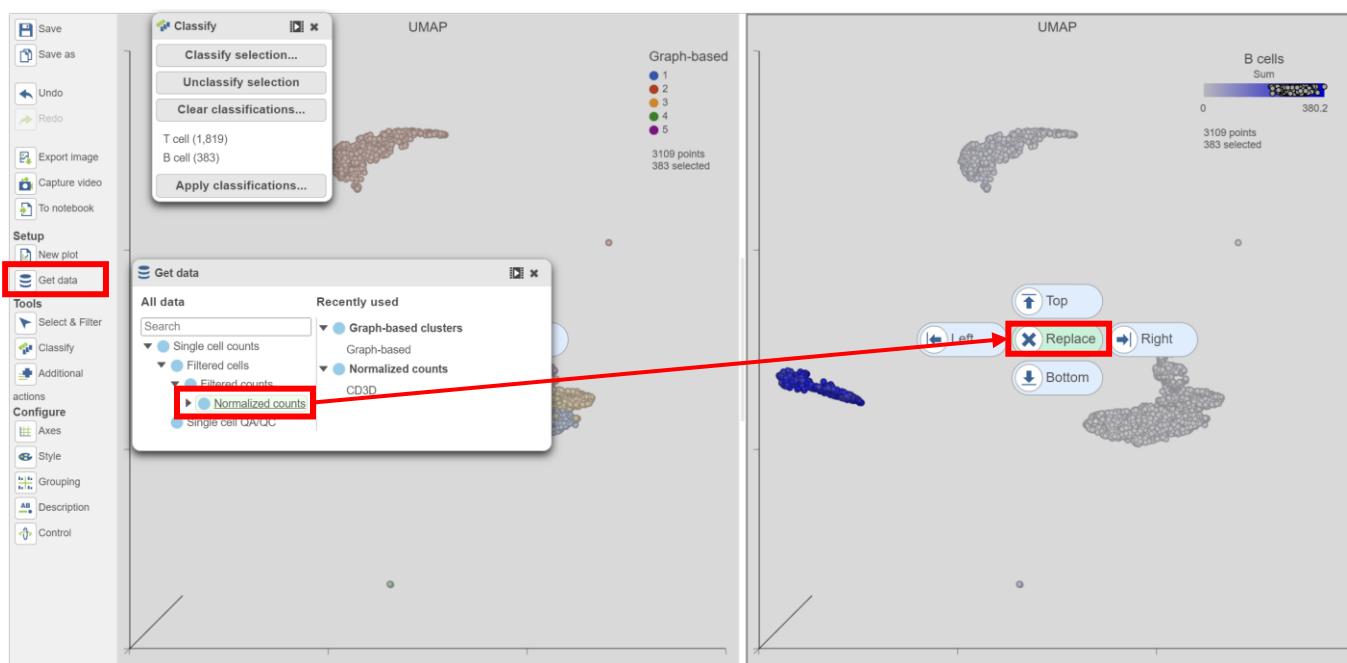


- Click **Select & Filter**
- Set **Select source for Color by** as Normalized counts
- Find the NKG7 and specify the min as 8
- Add GNLY and specify the min as 8
- Click **Classify selection** to name it as **Cytotoxic cell**
- Any number of genes can be used to build the rule



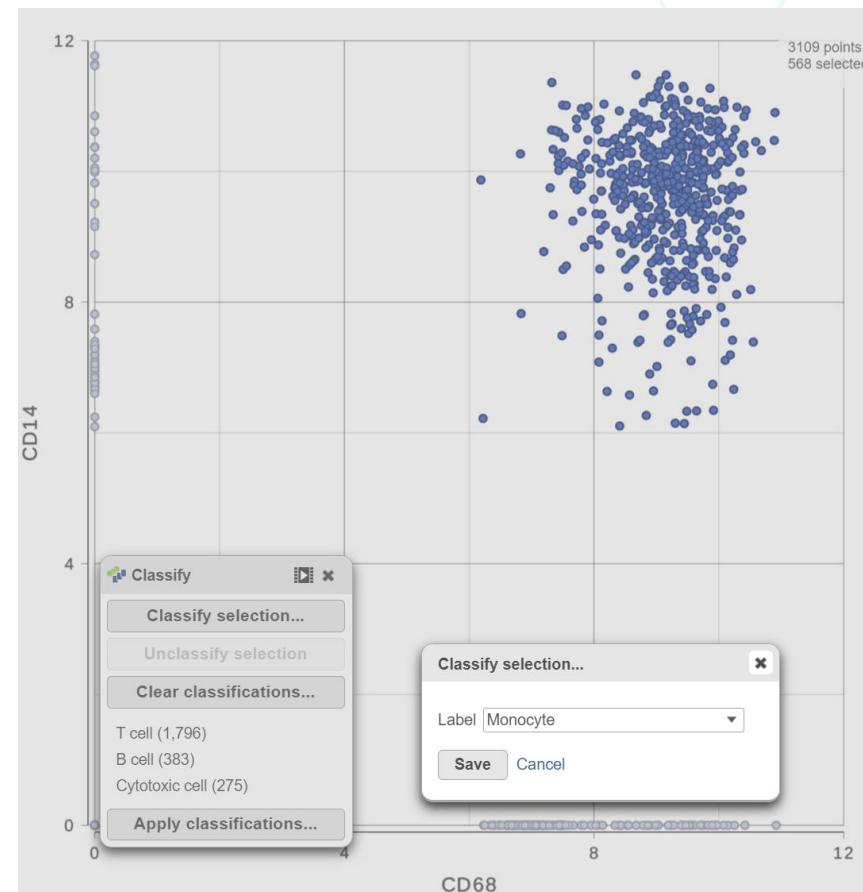
Classify Monocytes

- Click and drag the **Normalized counts** data node onto the canvas and replace the second UMAP, add a 2D scatter plot
- Set CD68 as X axis, and CD14 as Y axis



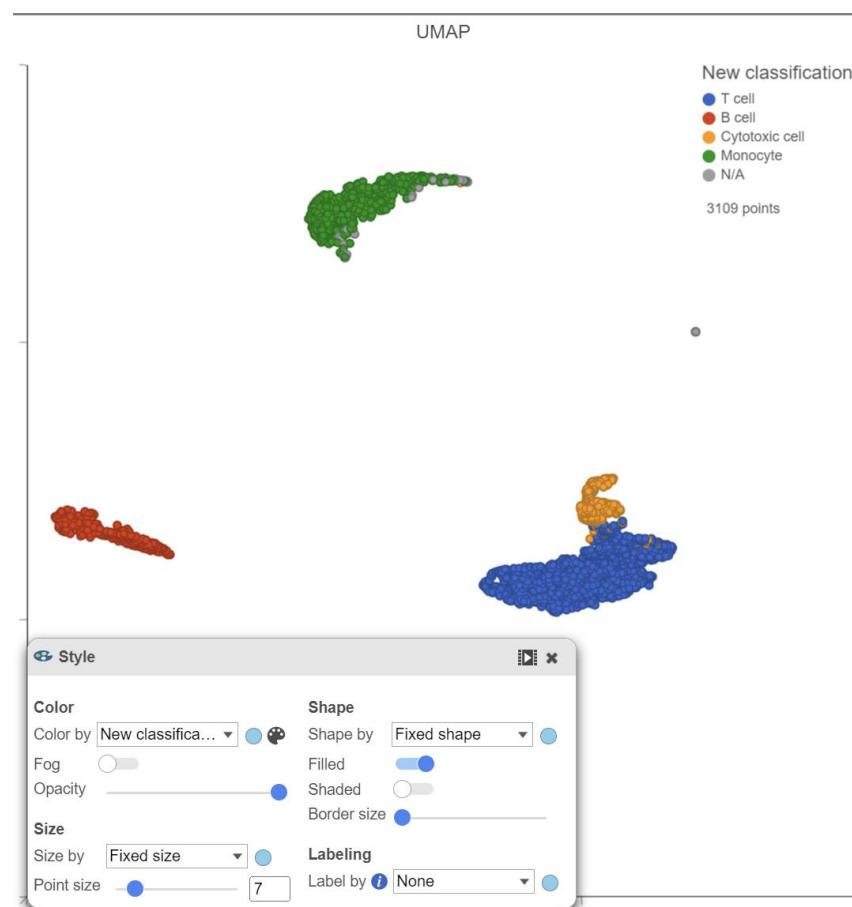
Classify Monocytes

- Use lasso tool to select cells with high expression on both genes (upper-right corner)
- Click **Classify selection**, name it as **Monocyte** and **Save**

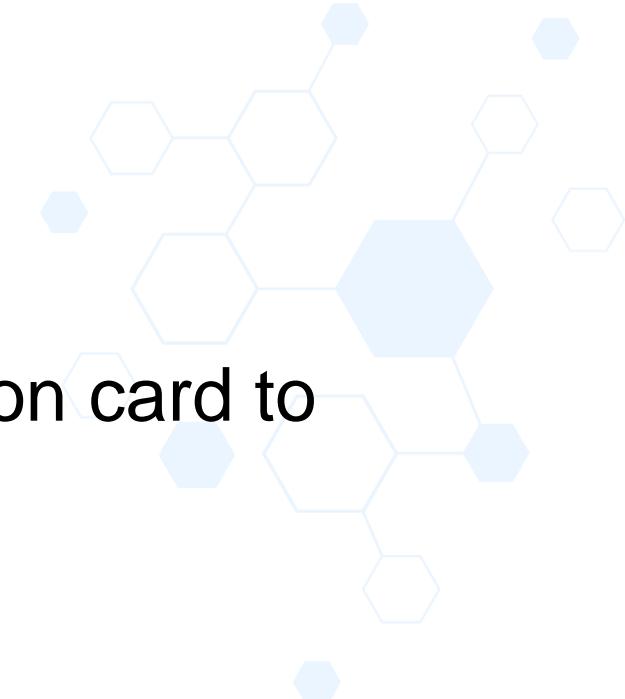


Viewing Classifications

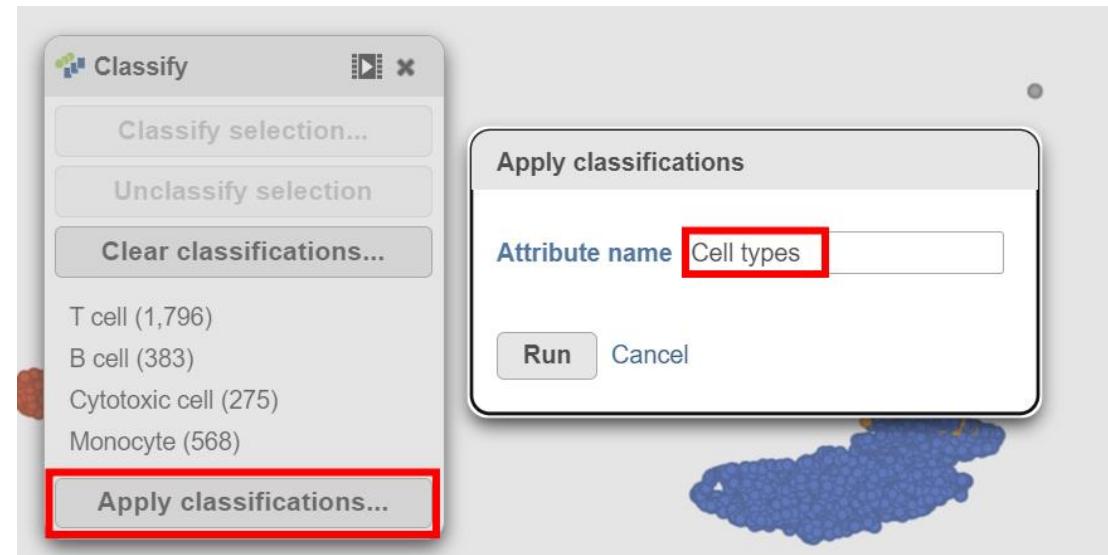
- Click on the UMAP plot, choose Color by **New classifications**



Viewing Classifications



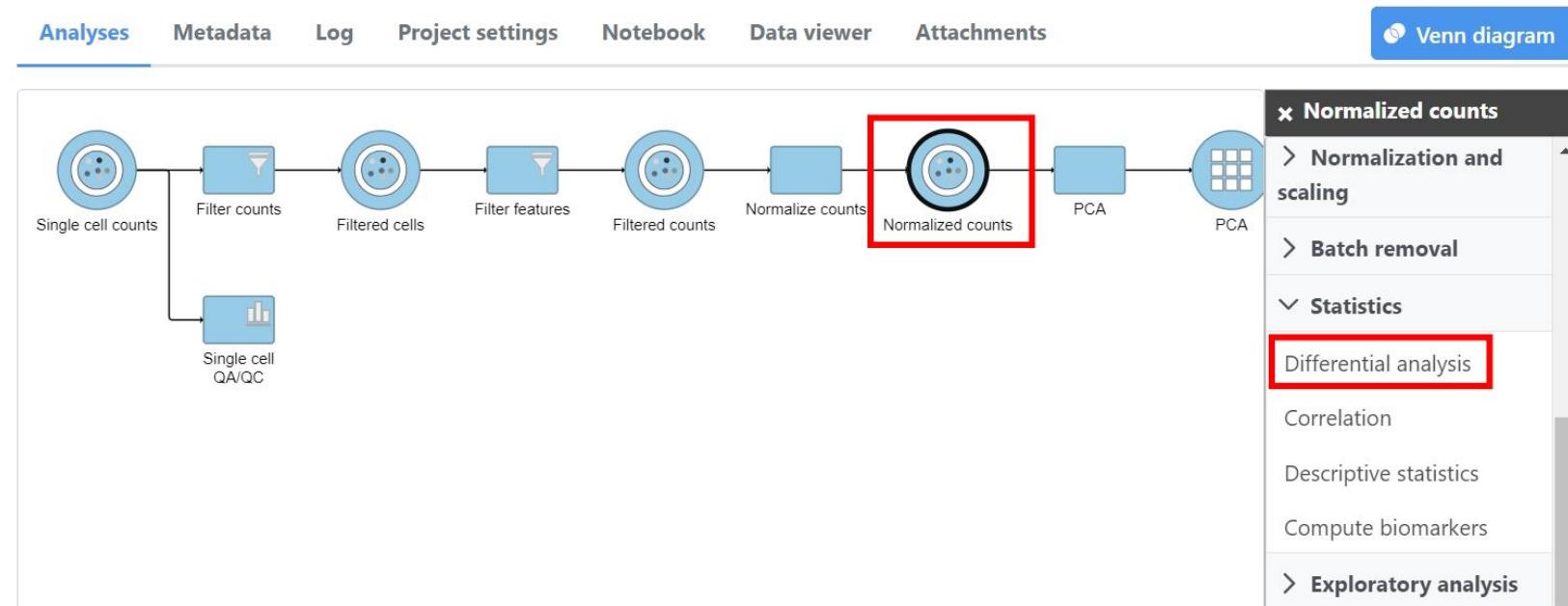
- Click **Apply classification...** button in Classification card to generate a new data node
- Name the new attribute **Cell types**
- Click **Run**



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Identifying Differentially Expressed Genes

- Click the **Normalized cells** data node
- Click **Differential analysis** in the **Statistics** section of the task menu



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Identifying Differentially Expressed Genes

- Choose **Hurdle** and click **Next**



Method to use for differential analysis i

DESeq2

Recommended for bulk RNA-Seq data with small sample size e.g. < 20 samples.

Hurdle model

Recommended for single cell RNA-Seq and CITE-Seq data.

Limma-trend

Recommended for continuous data with small sample size e.g. < 20 samples.

Limma-voom

Recommended for bulk RNA-Seq data with small sample size e.g. < 20 samples.

Kruskal-Wallis

Recommended for data that is not normally distributed and large sample size e.g. > 20 samples.

ANOVA

Recommended for continuous data including bulk and single cell expression data.

Welch's ANOVA

Recommended for continuous data including bulk and single cell expression data.

Gene Specific Analysis

Recommended for data with no replicates in any groups.



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Identifying Differentially Expressed Genes

- Choose **Cell types** and click **Next**
- Choose to compare Cytotoxic cell vs T cells, click **Add comparison**
- Click **Finish**

Select factor(s) for analysis

Categorical factors

Cell types

Numeric factors

Expressed genes Mitochondrial reads percent Ribosomal reads percent Total count

Add factors **Add interaction**

Selected factor(s)

Factor	Delete
Cell types	-

Define comparisons

Factor Cell types

B cell	>	Cytotoxic cell	Numerator
Cytotoxic cell	<	vs	Denominator
Monocyte			
T cell			
N/A			

Combine Pairwise

Add comparison

Comparisons

Comparison	Delete
Cytotoxic cell vs. T cell	-

Viewing GSA Results

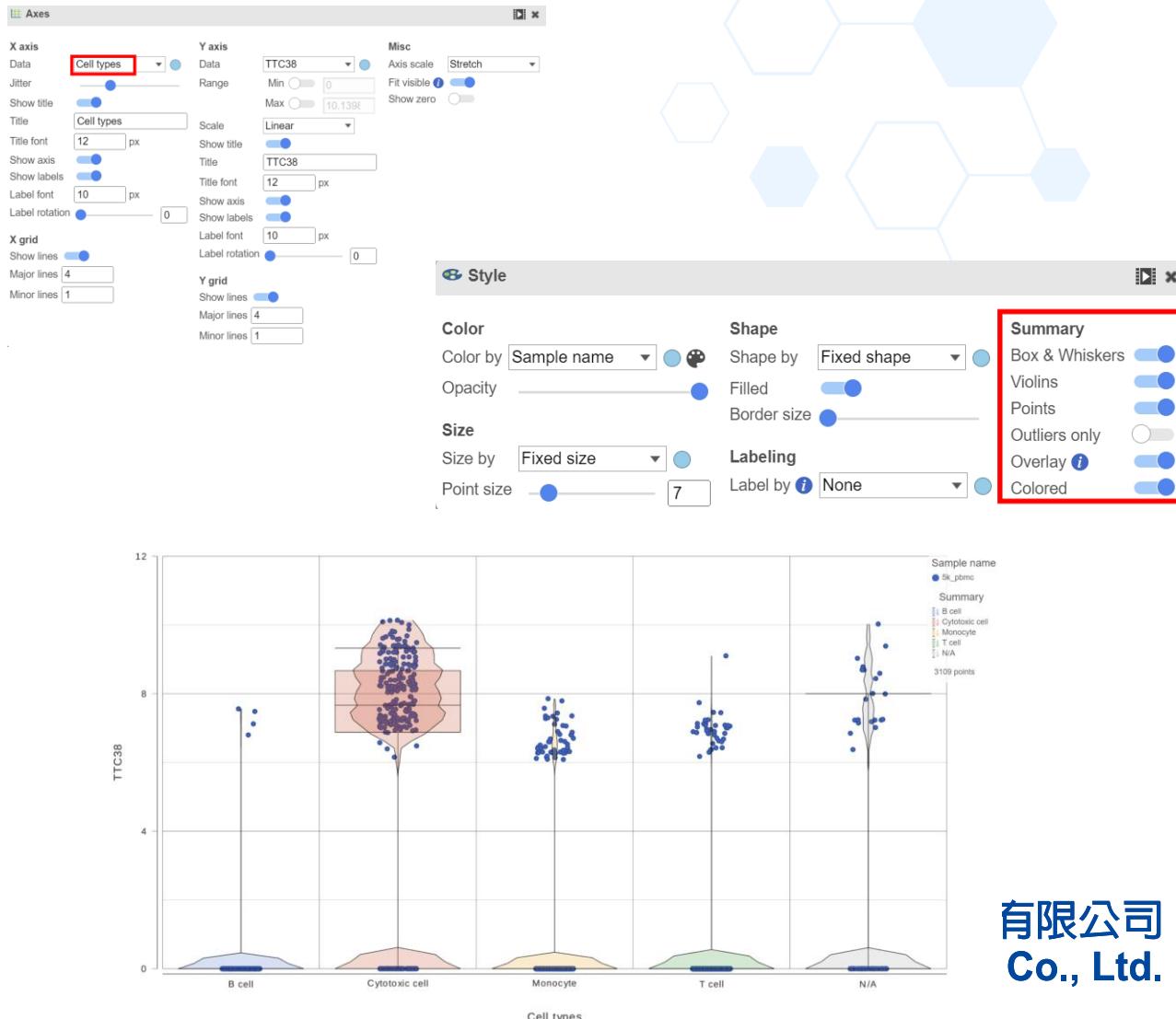


- Double click the **T cell vs Cytotoxic cell** data node
- Genes are listed starting with the lowest p-value

T cell vs Cytotoxic cell											
	View	Gene ID ↑↓	Gene name ↑↓	P-value ↑↓	FDR step up ↑↓	Ratio ↑↓	Fold change ↑↓	LSMean(T cell) ↑↓	LSMean(Cytotoxic cell) ↑↓	Pct(T cell) ↑↓	Pct(Cytotoxic cell) ↑↓
1	↶ ↷ ↴	PDGFD	PDGFD	0	0	0.38	-2.62	1.02	2.67	3.9E-3	0.20
2	↶ ↷ ↴	PRELID1	PRELID1	0	0	0.13	-7.69	18.78	144.34	0.57	0.87
3	↶ ↷ ↴	PREX1	PREX1	0	0	0.20	-4.90	2.23	10.91	0.16	0.45
4	↶ ↷ ↴	PRF1	PRF1	0	0	1.6E-3	-624.95	1.97	1,232.67	0.13	0.98
5	↶ ↷ ↴	ARHGEF3	ARHGEF3	0	0	0.26	-3.79	3.18	12.04	0.23	0.48
6	↶ ↷ ↴	ARHGDIB	ARHGDIB	0	0	0.71	-1.42	548.70	777.43	0.99	0.99
7	↶ ↷ ↴	ARHGDI	ARHGDI	0	0	0.27	-3.68	18.11	66.70	0.56	0.76
8	↶ ↷ ↴	PRKCA	PRKCA	0	0	6.14	6.14	14.83	2.42	0.53	0.17
9	↶ ↷ ↴	PRKCB	PRKCB	0	0	0.23	-4.33	9.98	43.17	0.45	0.68
10	↶ ↷ ↴	PRKCH	PRKCH	0	0	0.26	-3.82	12.67	48.35	0.50	0.71
11	↶ ↷ ↴	PRDX5	PRDX5	0	0	0.18	-5.56	15.57	86.61	0.54	0.80
12	↶ ↷ ↴	ERH	ERH	0	0	0.39	-2.55	18.59	47.48	0.58	0.72
13	↶ ↷ ↴	PRMT2	PRMT2	0	0	0.39	-2.54	37.56	95.47	0.68	0.80
14	↶ ↷ ↴	ARHGAP18	ARHGAP18	0	0	0.49	-2.05	1.14	2.34	0.03	0.17
15	↶ ↷ ↴	PRR5	PRR5	0	0	0.09	-10.81	2.59	28.01	0.19	0.62

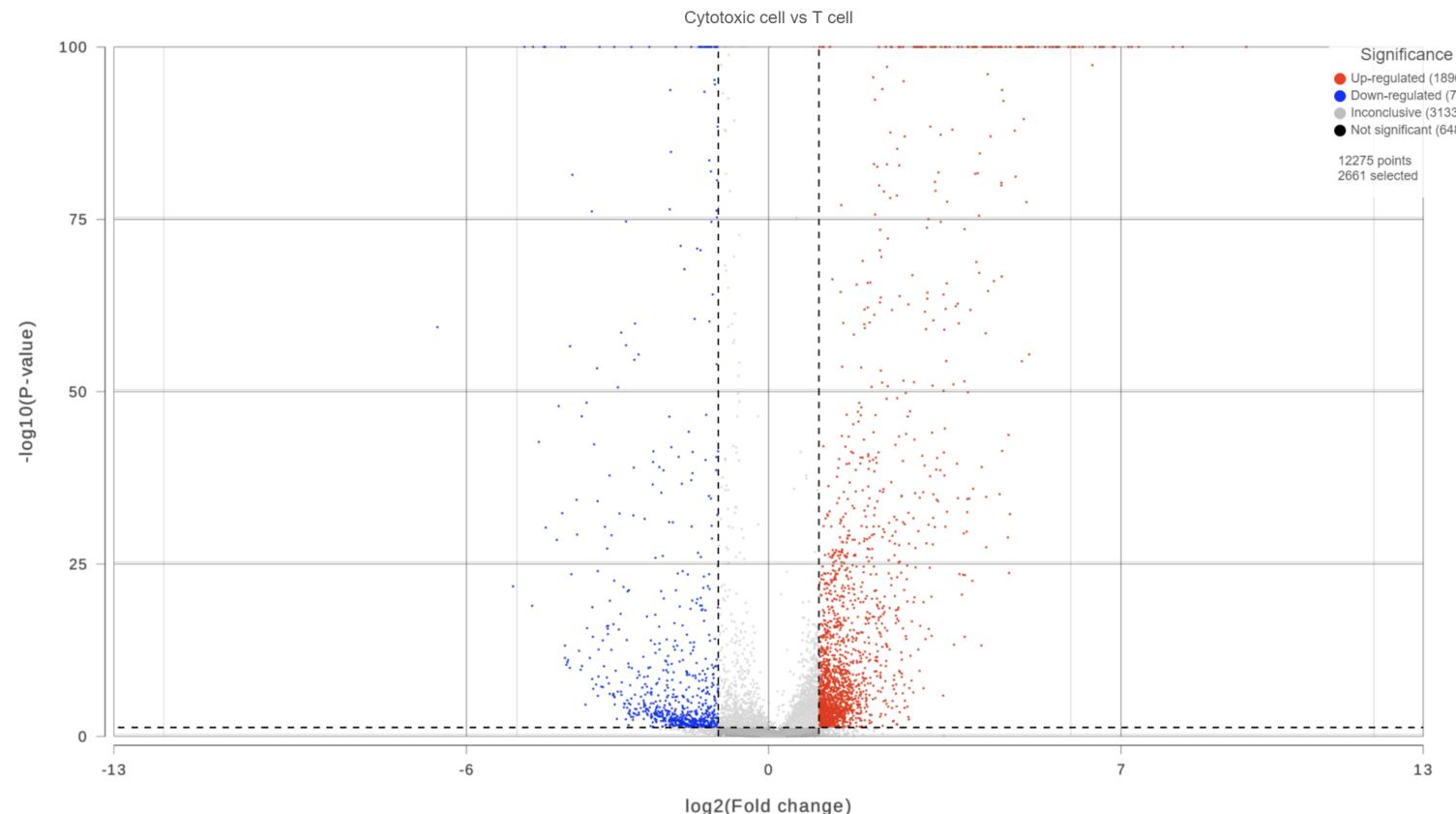
Viewing GSA Results

- Click the icon  next to a gene under View to open dot plot
- Set **Cell types** as X axis
- The plot can be added violins or box Whiskers in **Summary** session from **Style**



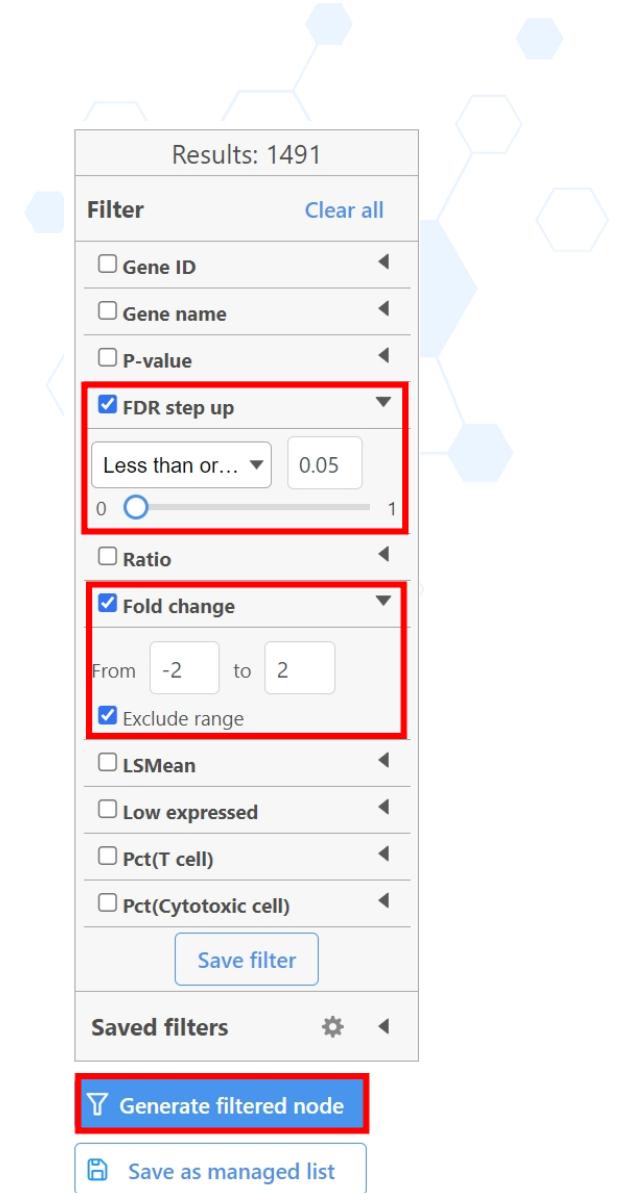
Viewing GSA Results

- Click the icon  to invoke volcano plot



Identify Significantly DEG

- Use the **Filter** on the left-hand side of the table
 - FDR step up: less than or equal to 0.05
 - Fold change: exclude range -2 to 2
- Click **Generate filtered node** to run the filter task

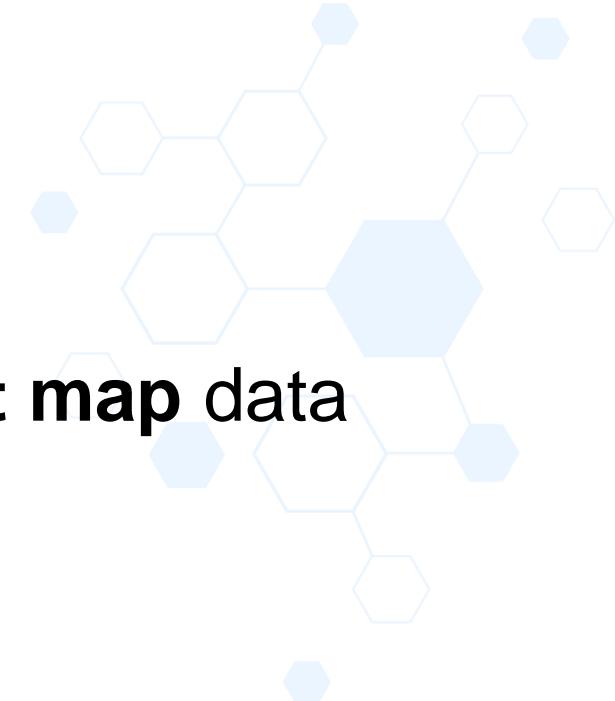


Configuring Hierarchical Clustering

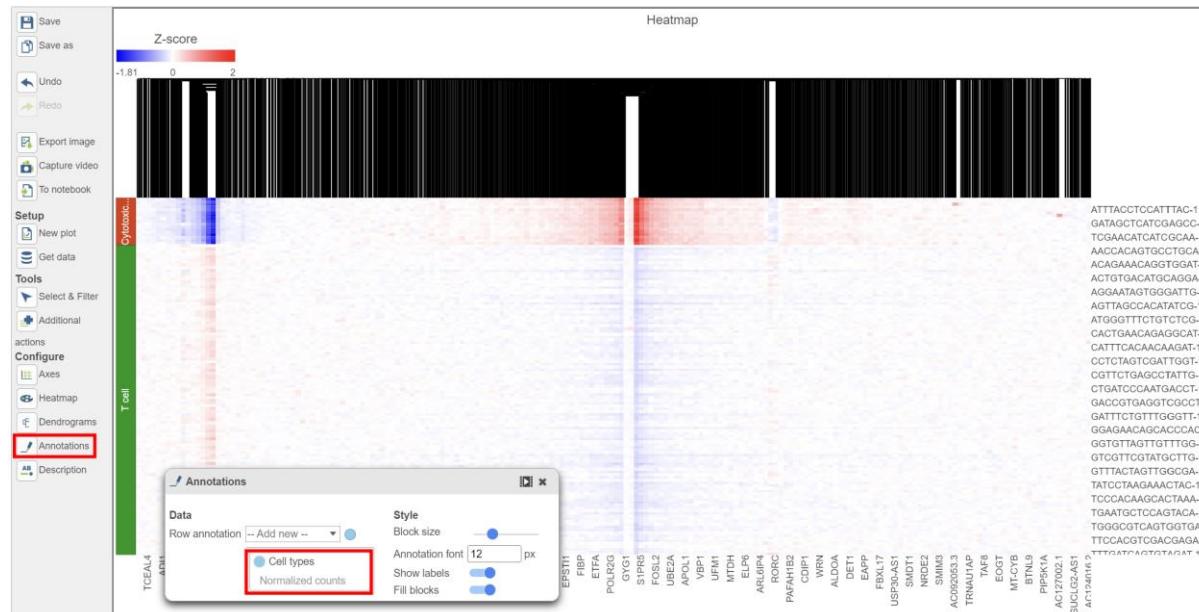
- Click the **Filtered feature list** data node
- Click **Hierarchical clustering / heat map** in the **Exploratory analysis** section of the task menu
- Check Cluster for Feature order
- Check **Filter cells** and set to *Include Cell types in T cells OR Include Cell types in Cytotoxic cells*

The screenshot shows the 'Hierarchical clustering' configuration window. At the top, there are three tabs: 'Plot' (with an info icon), 'Heatmap' (selected, with an info icon), and 'Bubble map' (with an info icon). Below the tabs is a section titled 'Ordering'. Under 'Feature order', the radio button 'Cluster' is selected (highlighted with a red box). Under 'Cell order', the radio button 'Assign order' is selected (highlighted with a red box), and the dropdown menu is set to 'Cell types' (also highlighted with a red box). A list of cell types is shown: B cell, Cytotoxic cell, Monocyte, T cell, and N/A. To the right of the ordering section is a diagram showing a grid of colored squares labeled 'Features' and a vertical bar labeled 'Cells' with double-headed arrows indicating their relationship. Below the ordering section is a 'Filtering' section. A large red box highlights the 'Filter cells' checkbox, which is checked. Inside this box are two filter criteria: 'include' followed by 'Cell types' dropdowns set to 'in' and 'Cytotoxic cell' (with an 'OR' option and a red 'X'), and another 'include' followed by 'Cell types' dropdowns set to 'in' and 'T cell' (with an 'OR' option and a red 'X'). Below the filtering section is an 'Advanced options' section with a dropdown menu set to 'Default' and a 'Configure' link. At the bottom are 'Back' and 'Finish' buttons.

Hierarchical Clustering Results



- Double-click on the **Hierarchical clustering / heat map** data node to view the result
- Use **Annotations** to annotate the cell types



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Biological Interpretation



- Click the Filtered feature list data node
- Click **Gene set enrichment** in the **Biological interpretation** section of the task menu
- Select **Gene set database** and choose the database
- Click **Finish**

Select gene set

Database

KEGG database Gene set database

Assembly

Homo sapiens (human) - hg38

Gene set database

2023 02 01 (Taiwan Genetech Biotech)

Biological Interpretation

- Double-click on the **Gene set enrichment** data node to view the report

Gene set ↑↓	Description ↑↓	Type ↑↓	Enrichment score ↑↓	P-value ↑↓	FDR step up ↑↓	Rich factor ↑↓	Genes in set ↑↓	Genes in list ↑↓	Genes not in list ↑↓	Genes in list, not in set ↑↓	Genes not in list, not in set ↑↓	i
GO:0070062	extracellular exosome	cellular component	121.88	1.17E-53	2.26E-49	0.28	1,310	369	941	1,057	8,376	■ ■
GO:0043230	extracellular organelle	cellular component	119.56	1.19E-52	5.75E-49	0.28	1,321	369	952	1,057	8,365	■ ■
GO:1903561	extracellular vesicle	cellular component	119.56	1.19E-52	5.75E-49	0.28	1,321	369	952	1,057	8,365	■ ■
GO:0065010	extracellular membrane-bounded organelle	cellular component	119.56	1.19E-52	5.75E-49	0.28	1,321	369	952	1,057	8,365	■ ■
GO:0031982	vesicle	cellular component	100.13	3.27E-44	1.26E-40	0.23	2,046	476	1,570	950	7,747	■ ■
GO:0002376	immune system process	biological process	84.06	3.1E-37	1E-33	0.26	1,199	313	886	1,113	8,431	■ ■
GO:0002682	regulation of immune system process	biological process	68.71	1.45E-30	4E-27	0.26	1,044	269	775	1,157	8,542	■ ■
GO:0030055	cell-substrate junction	cellular component	66.64	1.15E-29	2.67E-26	0.38	322	122	200	1,304	9,117	■ ■
GO:0005925	focal adhesion	cellular component	66.56	1.24E-29	2.67E-26	0.38	318	121	197	1,305	9,120	■ ■

Resolving complexity with spatial

Spatial



Single cell

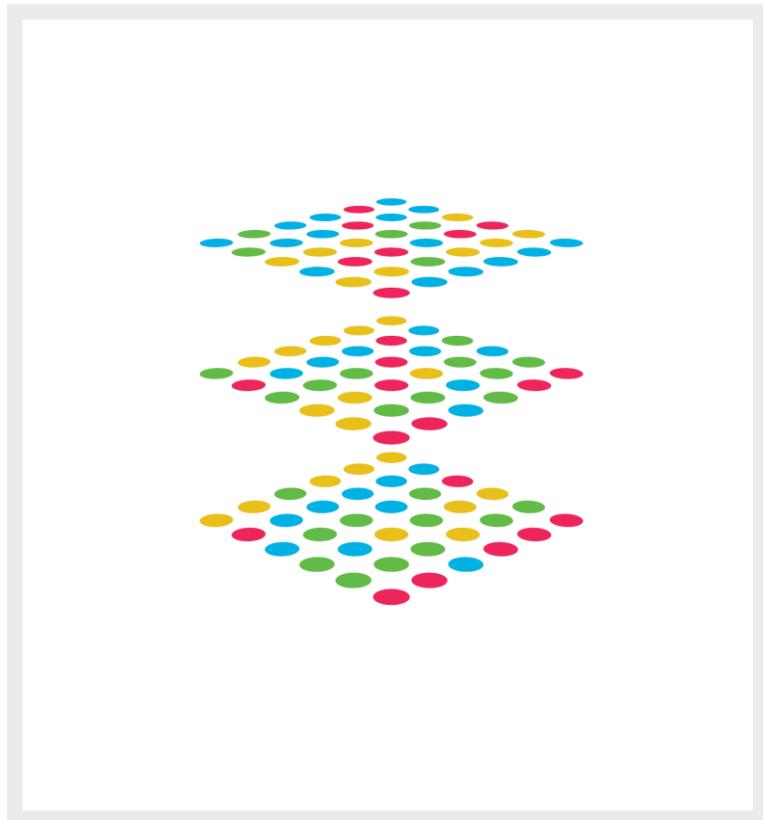


Bulk

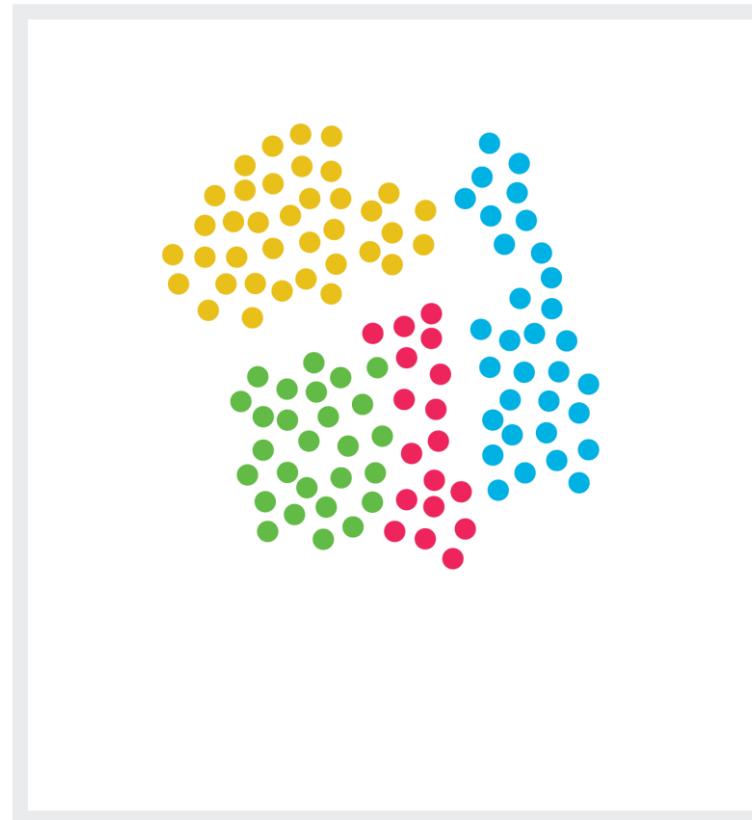


Resolving complexity with spatial

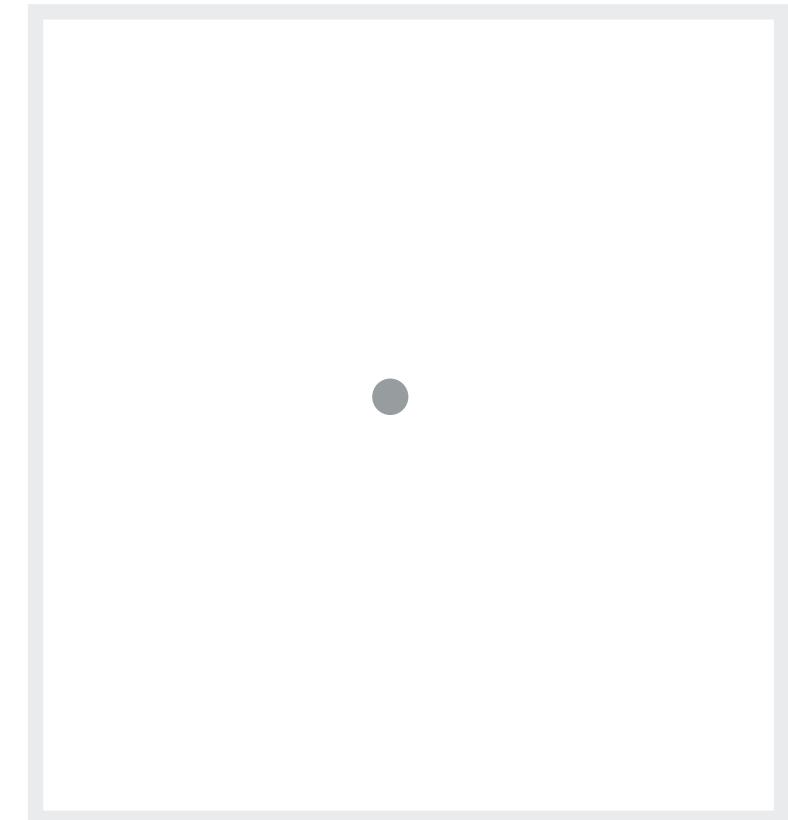
Spatial



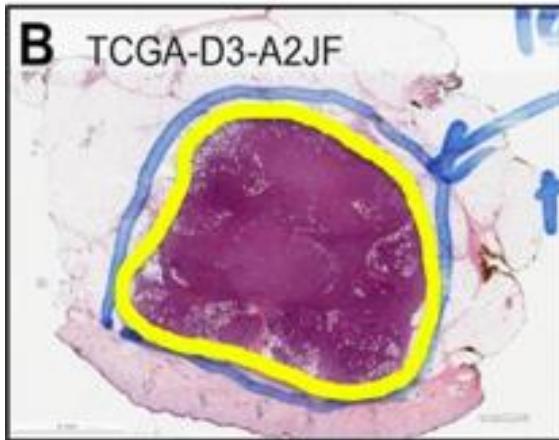
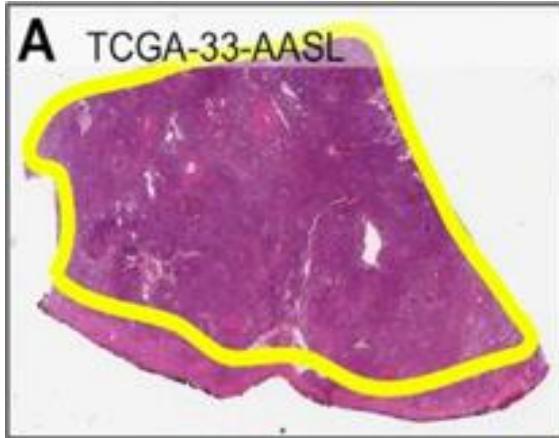
Single cell



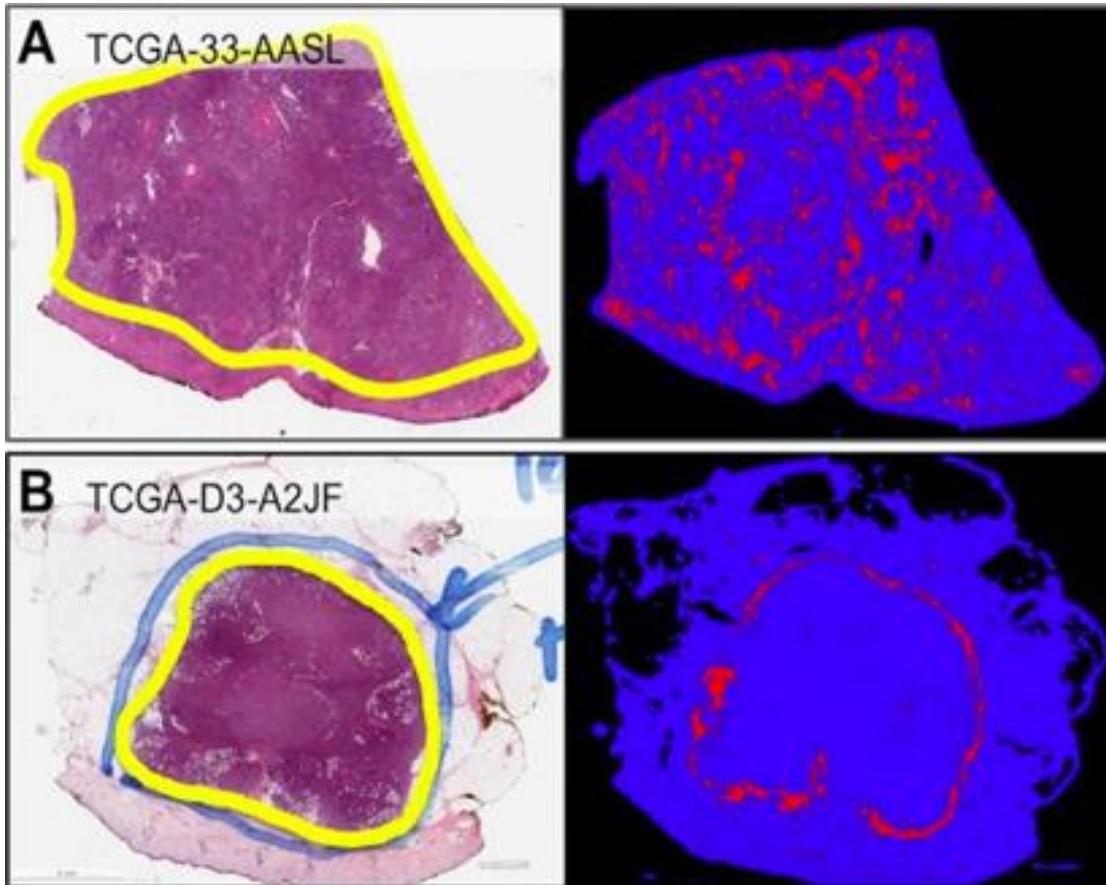
Bulk



Why spatial analysis



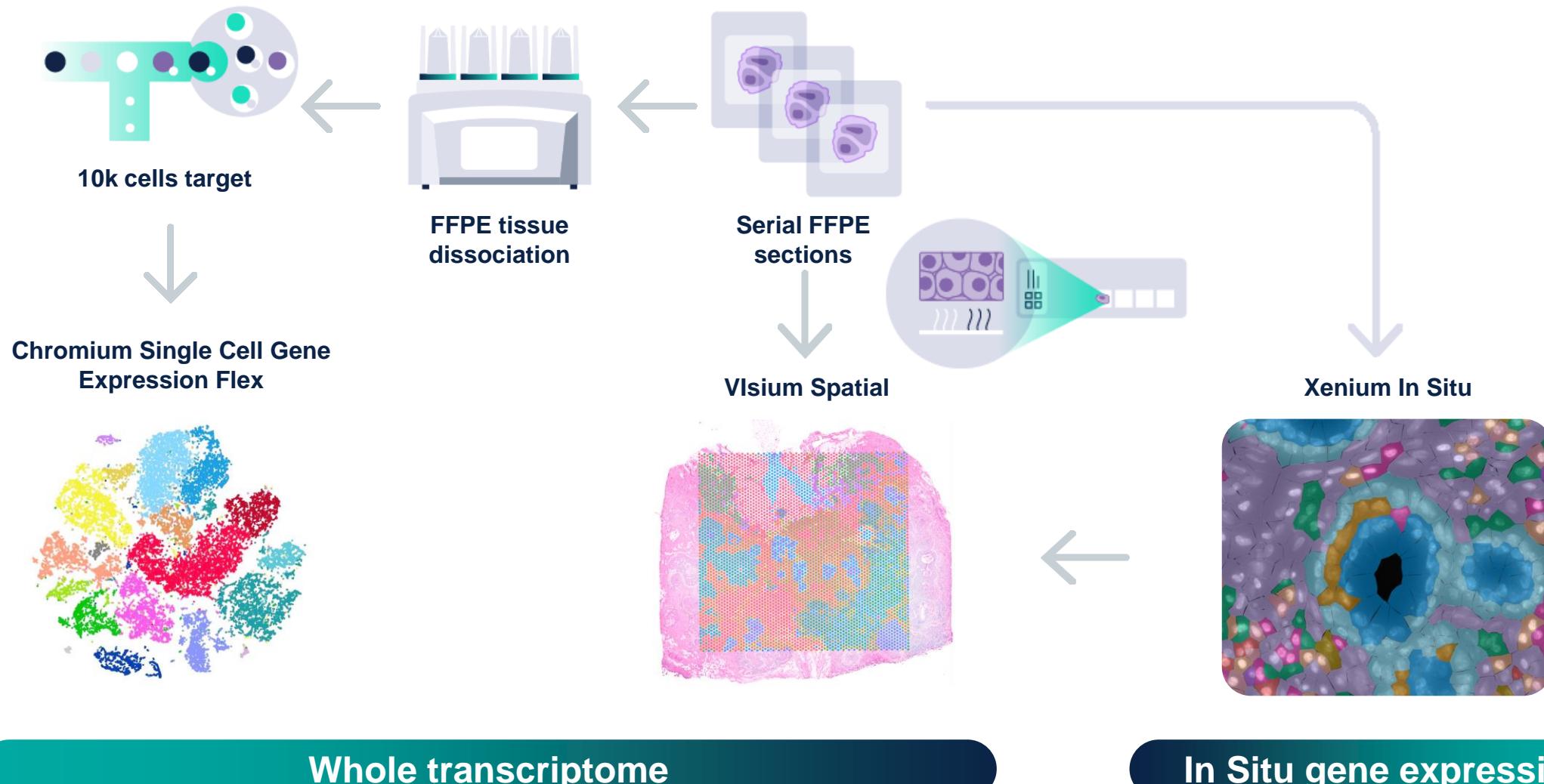
Why spatial analysis



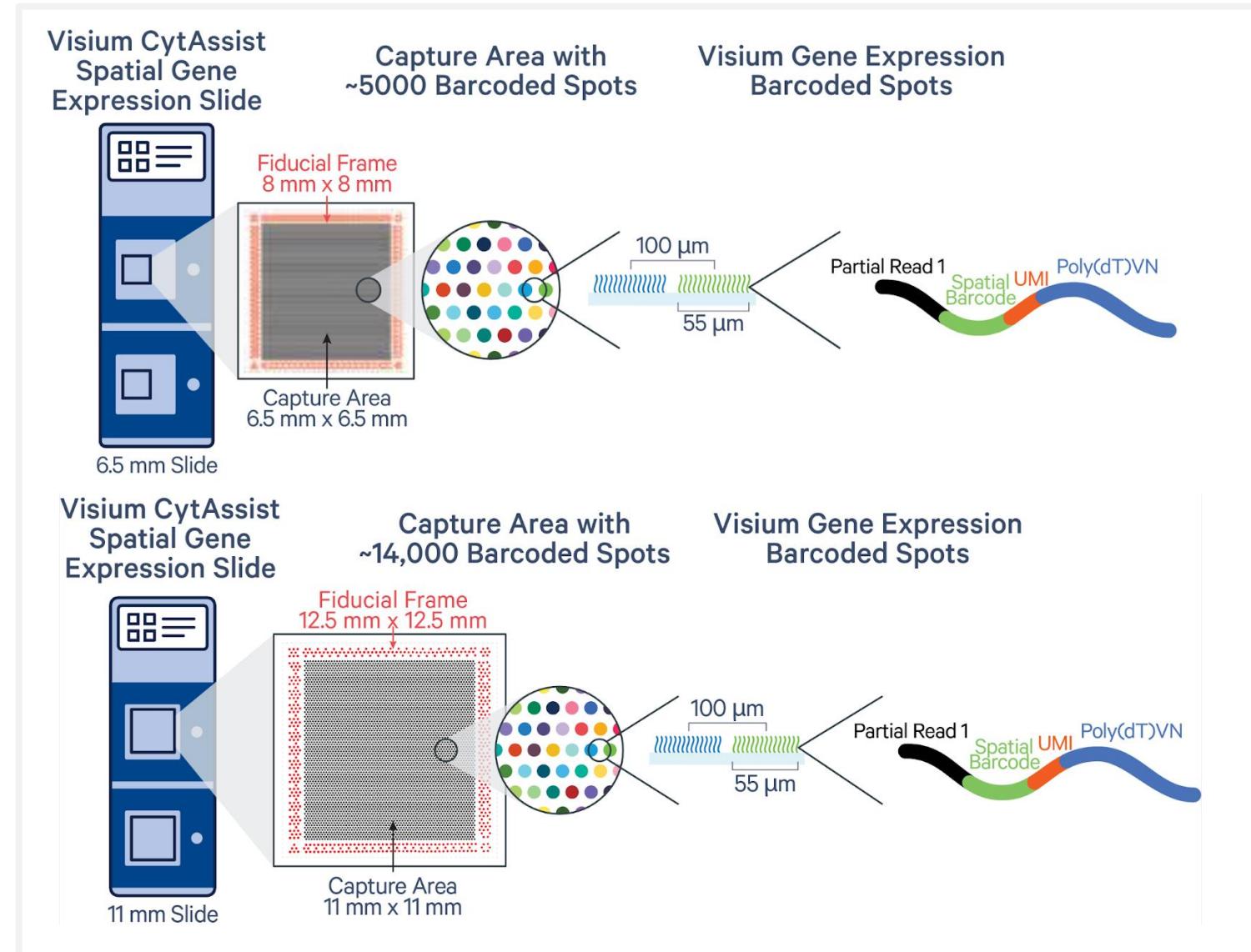
Lymphocytes infiltrating tumor

Lymphocytes stopped
at tumor boundary

Exploring Breast Cancer Biology with 10x Genomics

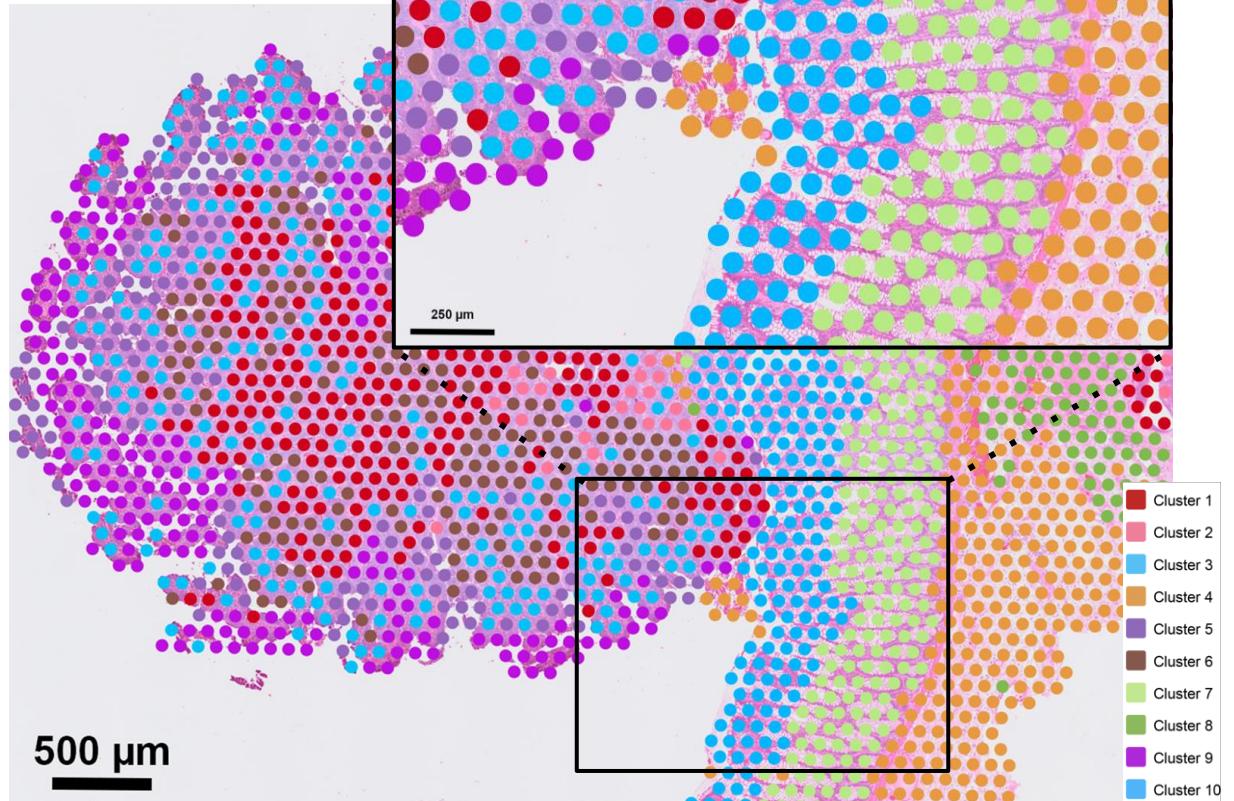


Visium CytAssist Gene Expression Slide Architecture

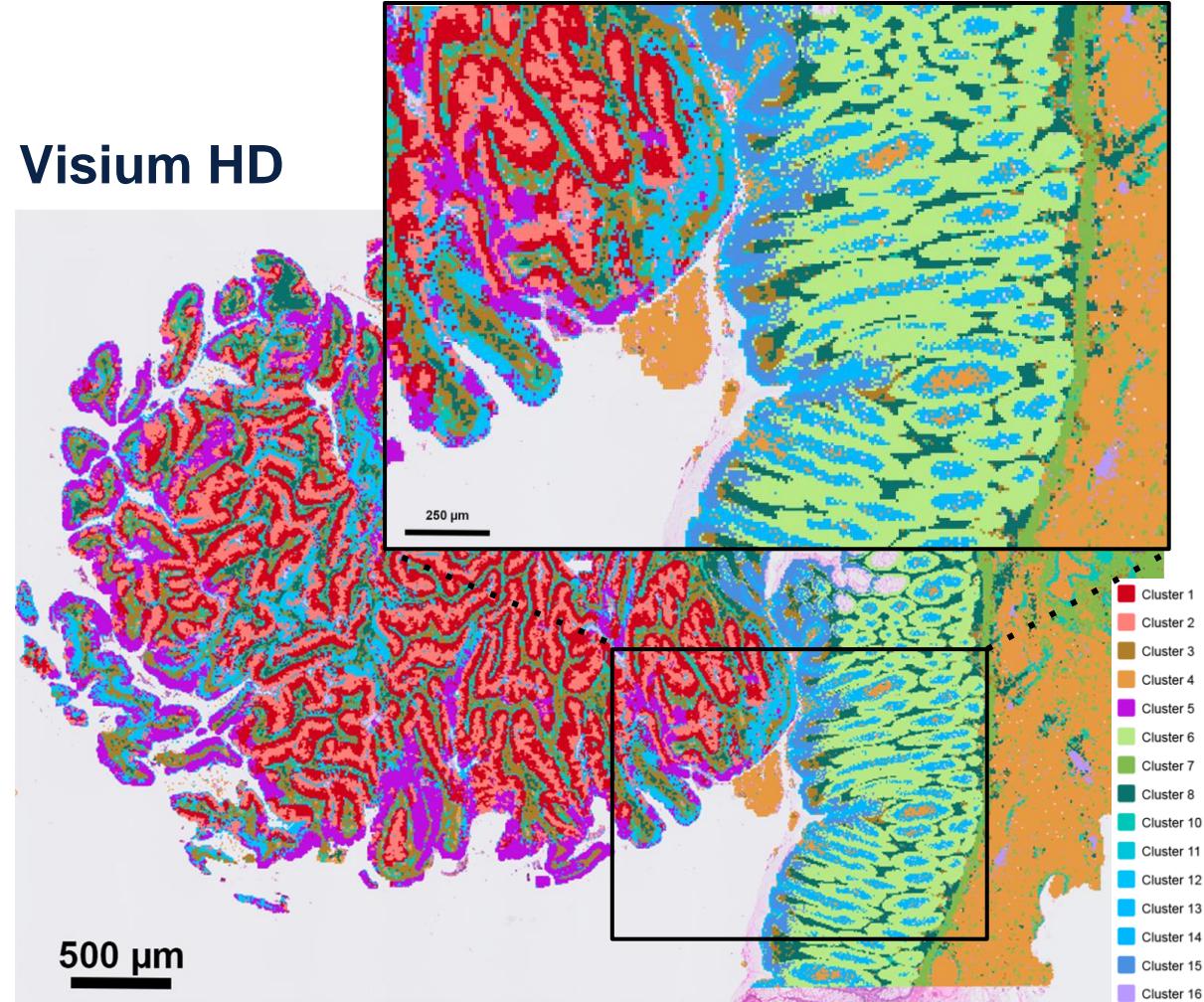


Introducing Visium HD

Visium



Visium HD



Spatial Analysis in Partek Flow



Import Data Based on Data Type



Single cell Bulk Microarray Other

scRNA-Seq Spatial scATAC-Seq V(D)J Flow/Mass Cytometry

Select the format

10x Genomics Visium Space Ranger output
10x Genomics Space Ranger output can be count matrix data as 1 filtered .h5 file per sample or sparse matrix files for each sample as 3 files (two .csv with one .mtx or two .tsv with one .mtx for each sample). The spatial output files should be in compressed format (.zip). The high resolution image can be uploaded and is optional.

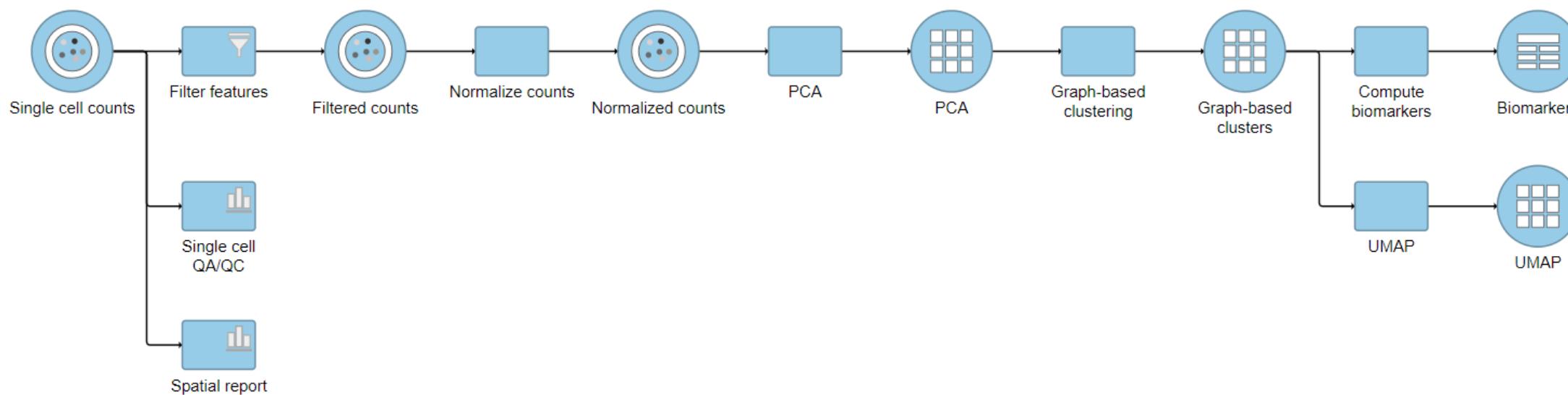
10x Genomics Xenium
10x Genomics Xenium data should include the unzipped Xenium Output Bundle with the preferred input image file (TIFF) for each sample.

NanoString CosMx
NanoString CosMx data should include 5 files (exprMat_file.csv, metadata_file.csv, polygons.csv, tx_file.csv, fov_positions_file.csv) and an image folder (CellComposite) per sample

10x Genomics Visium fastq
Unaligned fastq reads (fastq, fastq.gz, fastq.bz2, fq, fq.gz, fq.bz2) can be processed using the 10x Genomics Space Ranger task. Please follow a naming convention only containing letters, digits, underscores and dashes.

- Visium: Space Ranger output or Fastq

Demo Pipeline



Import Space Ranger Output

Samples and files
Count matrix files should be one hdf5 file or three feature-barcode matrix files (features.tsv, barcodes.tsv and matrix.mtx)

+ Add sample

Sample name	Cells	Features	Count matrix files	Spatial outputs	High resolution image (optional)	Action
Mouse Olfactory Bulb	1185	32285	Visium_Mouse_Olfactory_Bulb_filter...	Visium_Mouse_Olfactory_Bulb_spati...		-

Feature annotation

Use annotation file
Select the file that has been used to generate the feature counts (e.g. gene or protein information).

Assembly Mus musculus (mouse) - mm10 **Annotation model** Ensembl Transcripts release 102 (Taiwan Genetech Biotech)

Primary feature identifier

Feature name (Values: Xkr4, Gm1992, Gm19938, Gm37381, Rp1, Sox17, Gm3758..)
 Feature ID (Values: ENSMUSG0000051951, ENSMUSG0000089699, ENSMUSG000...)

Deduplication method
If the feature ID is not unique, the feature will be summarized by the selected method.

Mean Maximum Sum

Data format

Count value format

Raw counts Normalized counts with log base None

Filtering

Features to report

All features
 Features with non-zero values across all samples



Assign files to samples individually and choose the annotation model.



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Import Visium Fastq



If you only have Visium fastq,
Space Ranger is available in
Partek Flow!

- ▼ 10x Genomics
- STARsolo
- Cell Ranger - Gene Expression
- Cell Ranger - ATAC
- Space Ranger

Assay type

10X assay type
The selected data node must have fastq files

Spatial gene expression CytAssist gene expression

Reference assembly

Assembly Select genome, then select annotation index.	Index Ensembl Transcripts release 108 (Taiwan Genetech Biotech)
Homo sapiens (human) - hg38	

Image and barcode files

Sample files
For Spatial GEX, image files are single H&E brightfield images in TIFF or JPG format; For CytaAssist GEX, image files are CytaAssist instrument captured eosin stained Brightfield tissue image with fiducial frame in TIFF format. Probe set files are optional CSV files specifying the probe set used. Formalin-fixed paraffin-embedded (FFPE) image file requires probe set file.

Sample name	Image file	Browse image file	Probe set file	Browse probe set file
Sample 1	<input type="button" value="Q"/>	<input type="button" value="Q"/>		

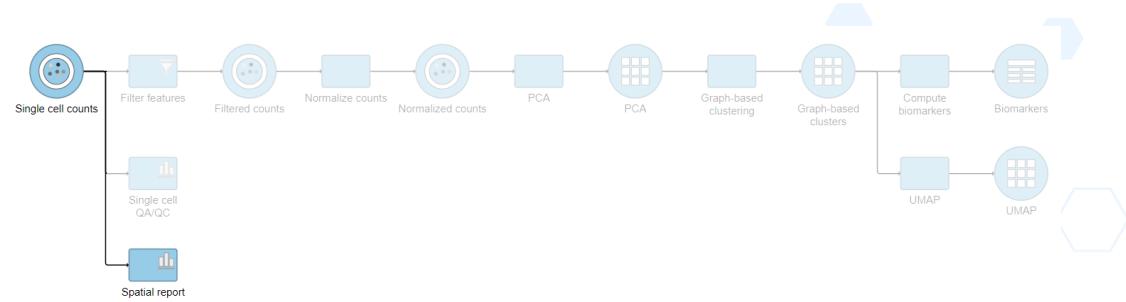
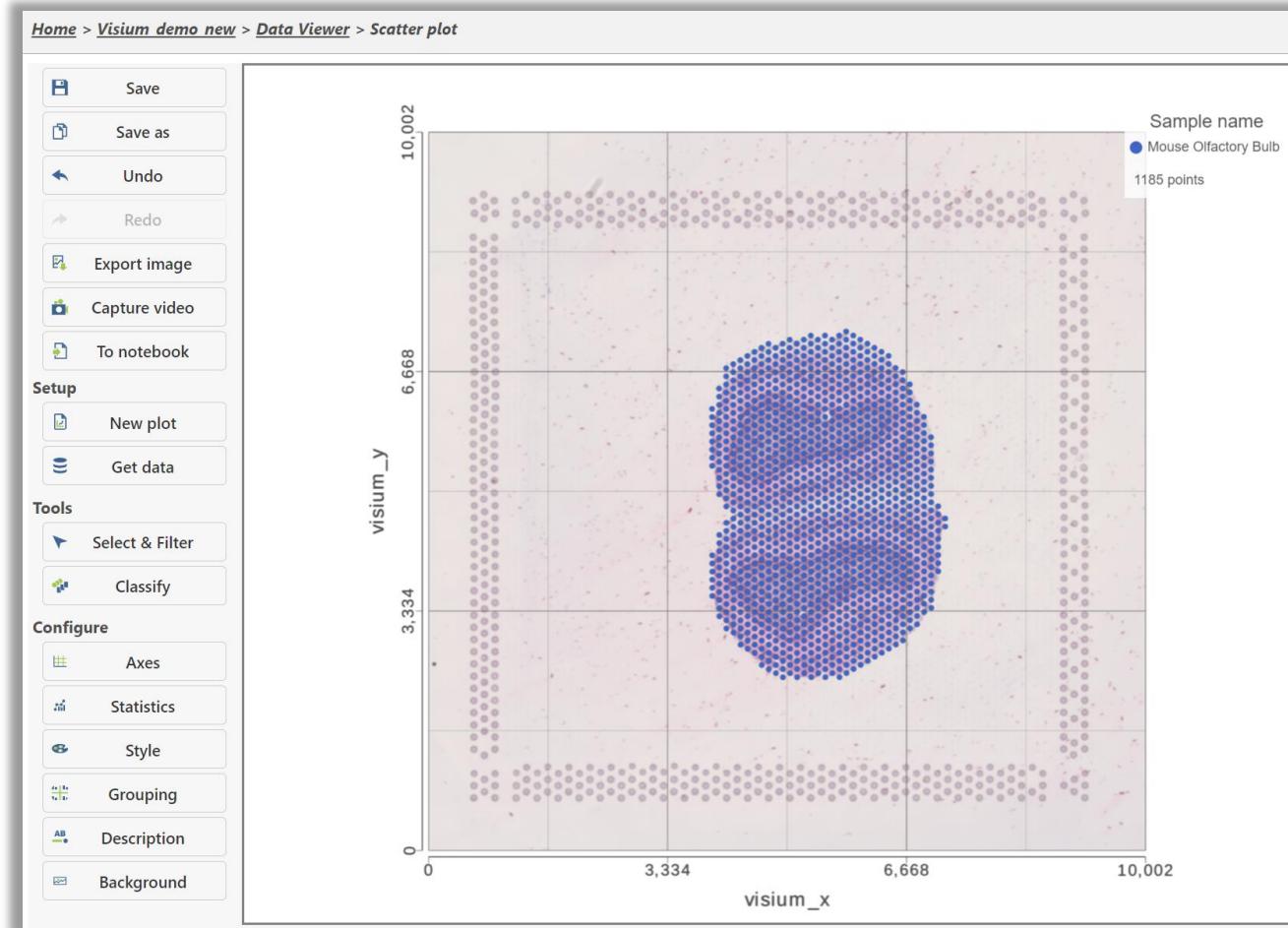
Advanced options

Use slide serial number file
Select a file specifying samples' slide and area information

Option set

-- Default --

Spatial Report

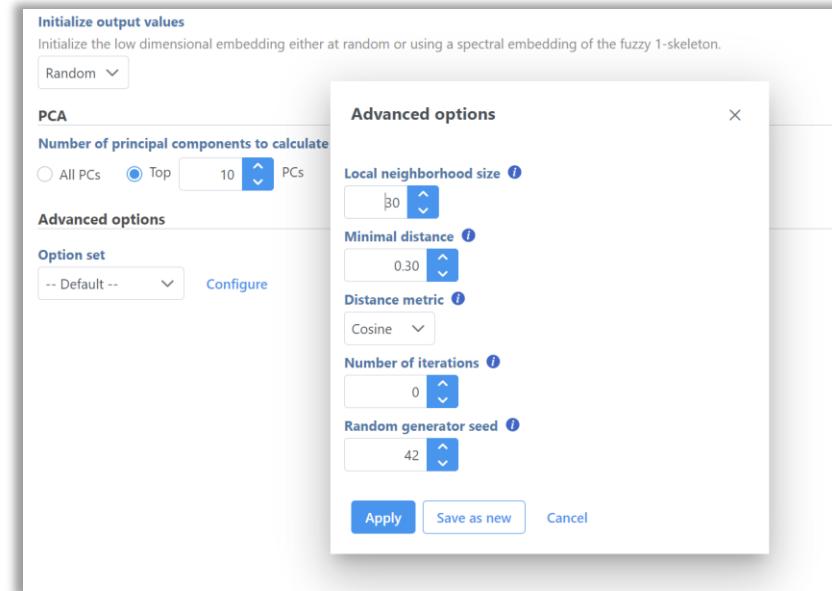


Spatial report would be generated automatically, which visualizing all spots (points) on the tissue image.

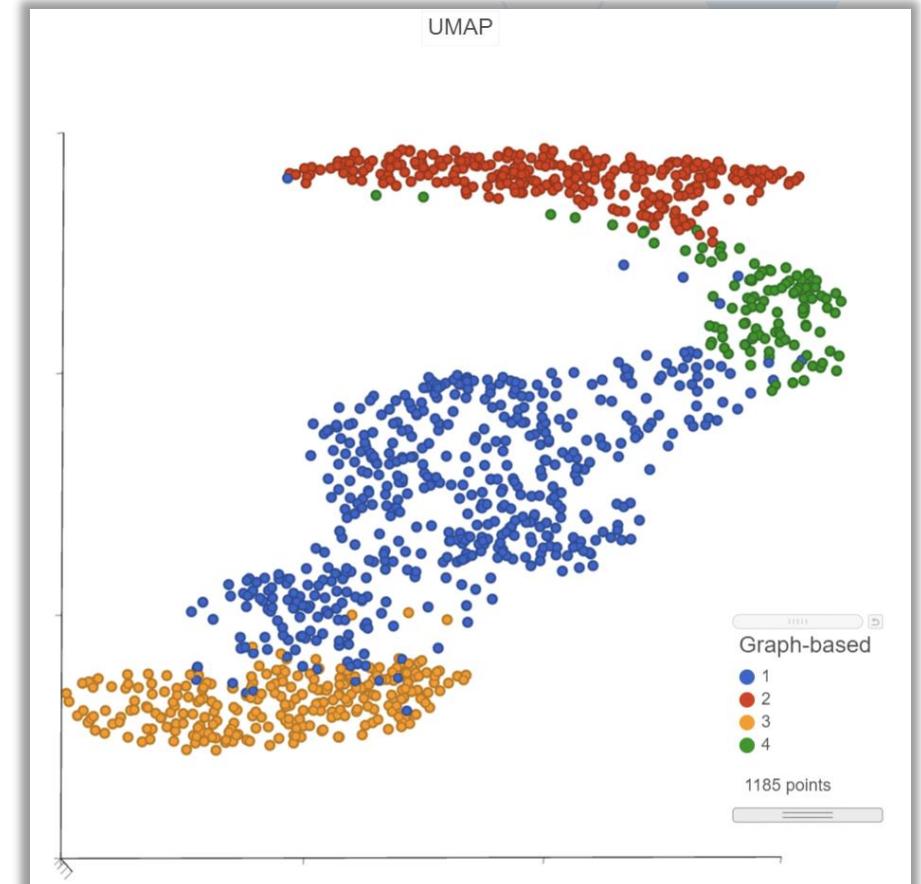
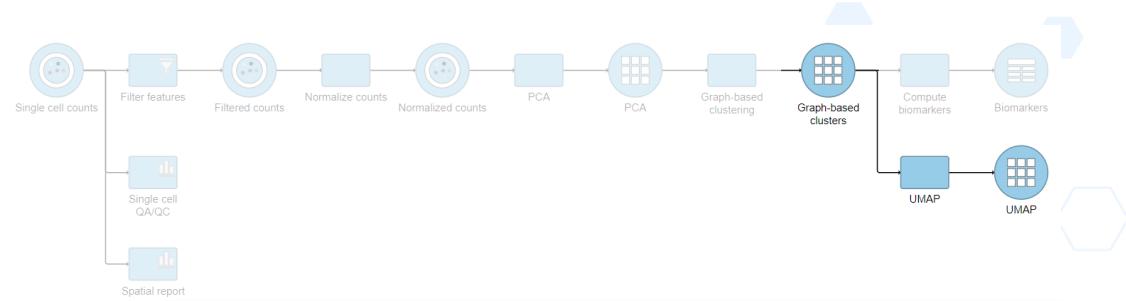


均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

UMAP



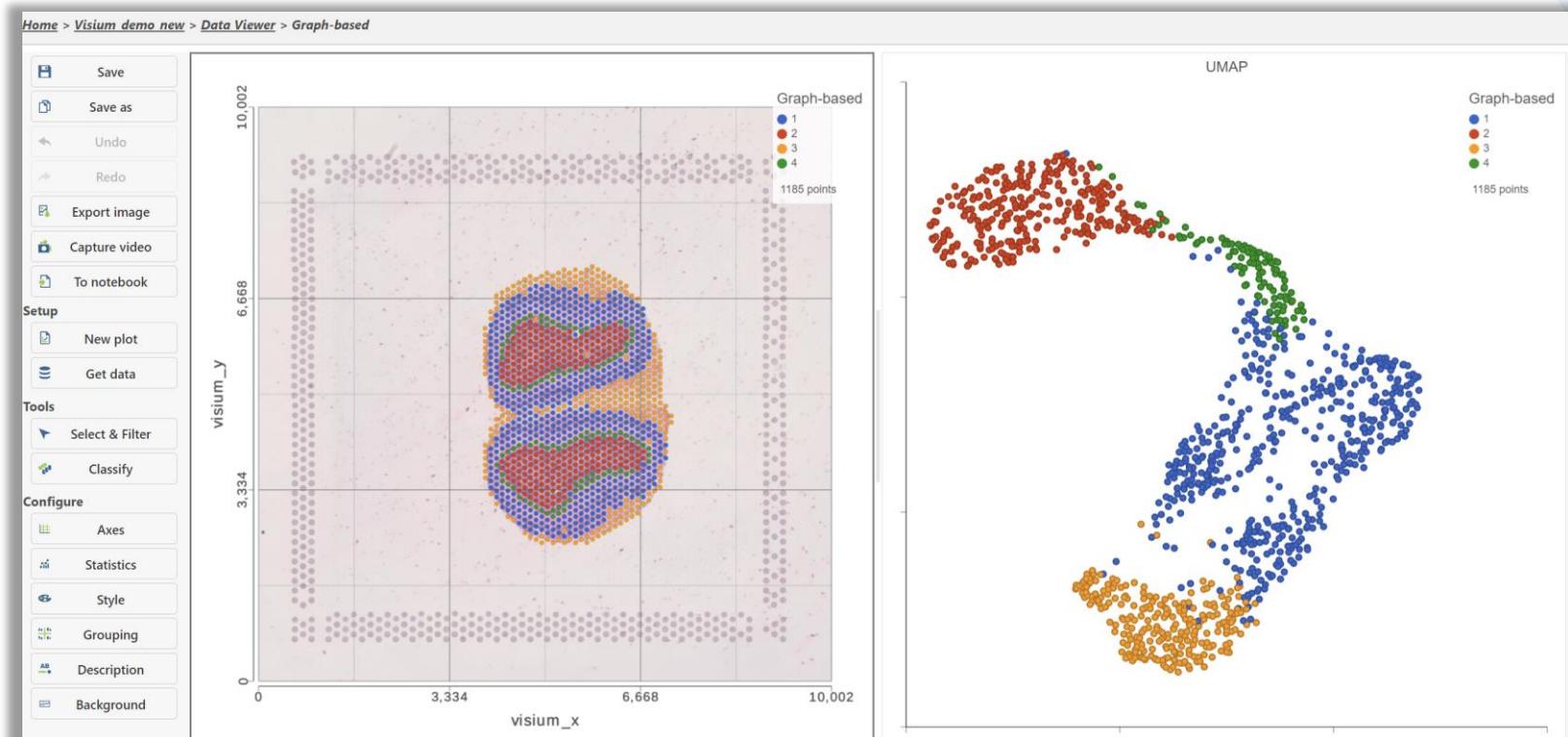
UMAP is particularly useful for visually identifying groups of similar samples or cells in large high-dimensional data sets.



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Visualization of Clustering Results

In data viewer, multiple plots can be shown at the same time.

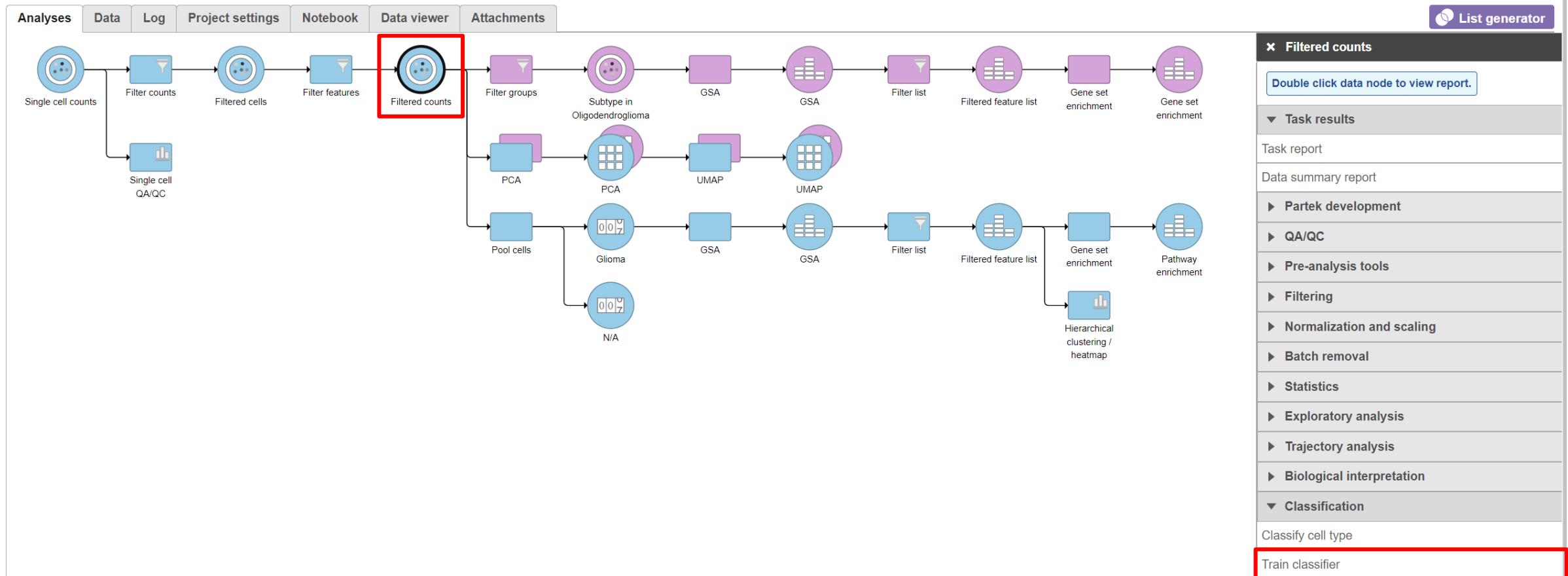


均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Appendix – Garnett Classifier



Train Classifier



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Train Classifier

Marker file

Choose marker from i Local files ▼

Marker file i Partek Flow Server URL

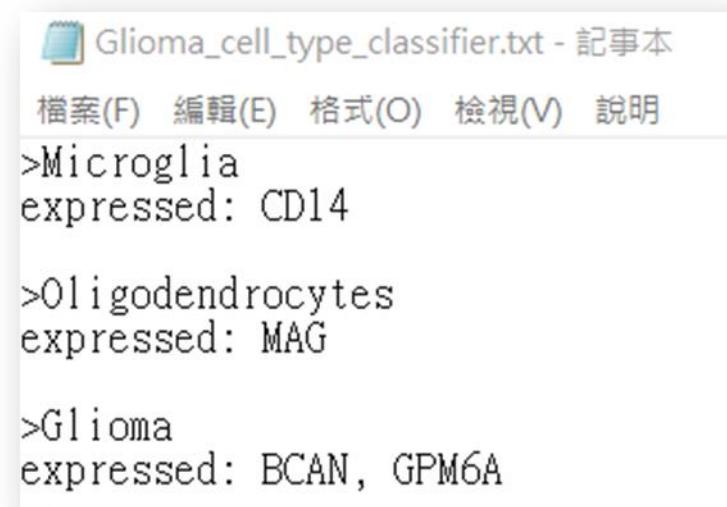
No files selected Browse

To move files from your local computer to the Partek server, please [Transfer files first.](#)



Constructing a marker file

- Each cell type definition starts with a '>' symbol and the cell type name.
- Definition lines start with a keyword and a ':' and entries are separated by a comma.
- There has to be a space character after the colon and that there has to be a space character after the comma.



The screenshot shows a Windows Notepad window titled "Glioma_cell_type_classifier.txt - 記事本". The menu bar includes "檔案(F)" (File), "編輯(E)" (Edit), "格式(O)" (Format), "檢視(V)" (View), and "說明" (Help). The content of the file is as follows:

```
>Microglia
expressed: CD14

>Oligodendrocytes
expressed: MAG

>Glioma
expressed: BCAN, GPM6A
```



均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Constructing a marker file

Recommended expression specifications

Format	Example
expressed: gene1, gene2	expressed: MYOD1, MYH3
not expressed: gene1, gene2	not expressed: PAX6, PAX3

Meta data specifications

Format	Example
subtype of: celltype	subtype of: T cells
custom meta data: attribute1, attribute2	tissue: spleen, thymus



Marker file example

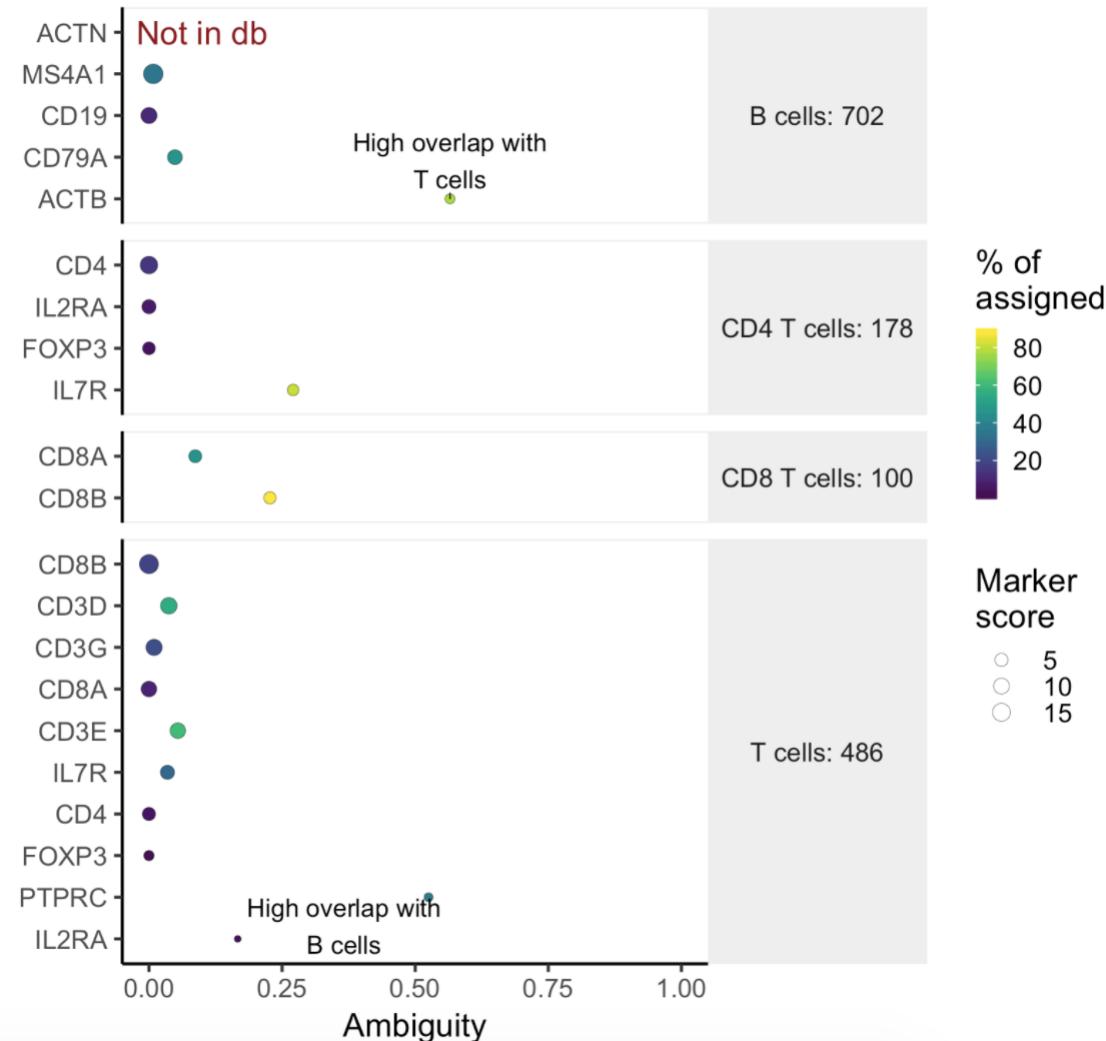
```
>B cells
expressed: CD19, MS4A1
expressed above: CD79A 10
references: https://www.abcam.com/primary-antibodies/b-cells-basic-immunophenotyping, 10.3109/07420528.2013.775654

>T cells
expressed: CD3D
sample: blood # A meta data specification

>Helper T cells
expressed: CD4
subtype of: T cells
references: https://www.ncbi.nlm.nih.gov/pubmed/?term=12000723
```



Train Classifier Results



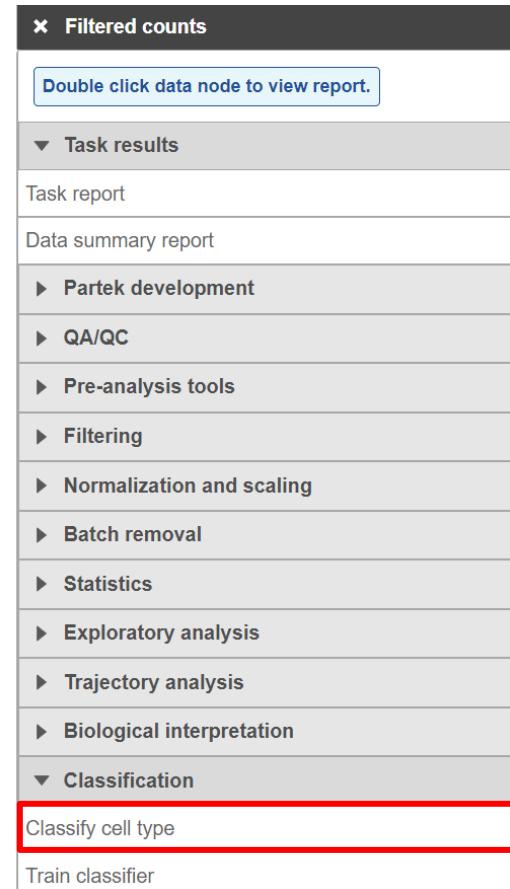
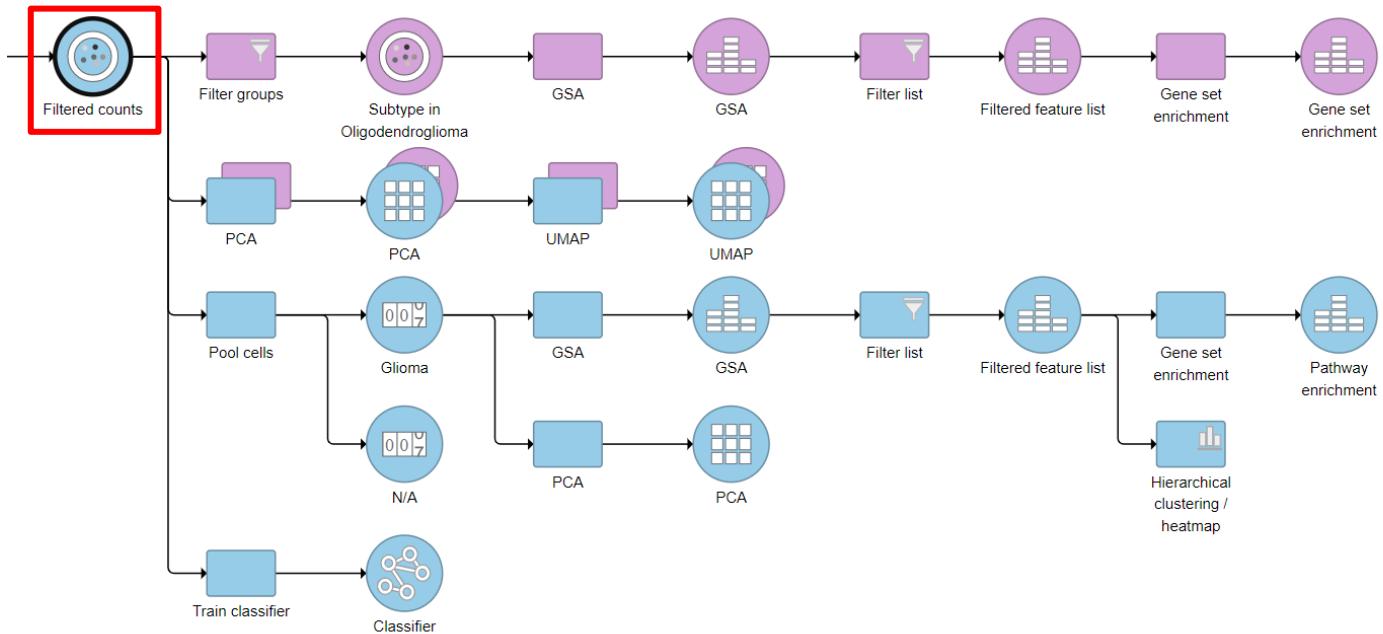
- Double click the **Classifier** data node
- Ambiguity scores are calculated for each of the markers which indicates how many cells receive ambiguous labels when this marker is included

Train Classifier Results

- The classification gene table may give a hint to which genes are chosen as the relevant genes for distinguishing between different cell types

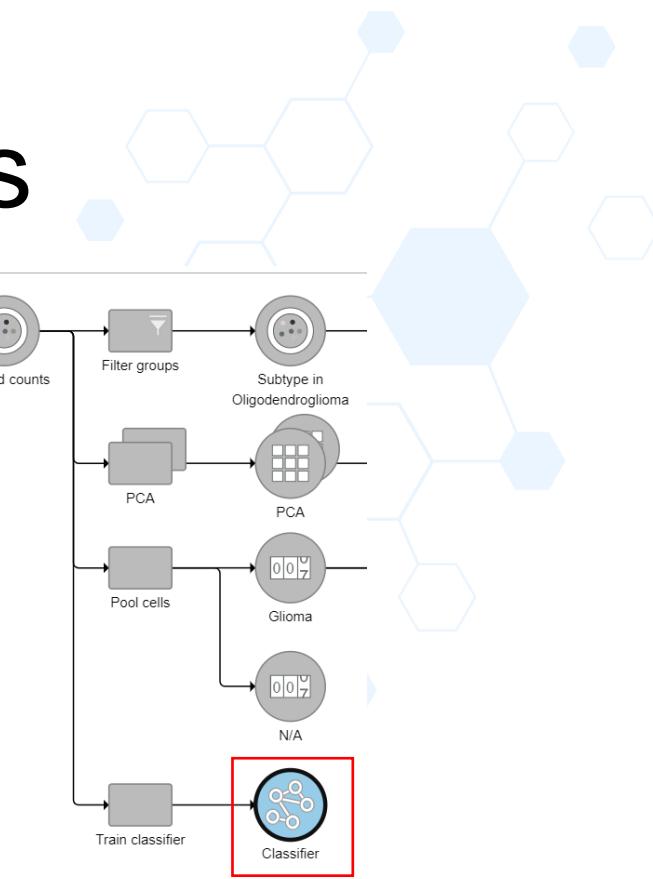
Feature	Glioma	Microglia	Oligodendrocytes	Unknown
(Intercept)	-39.80	9.48	14.21	16.11
BCAN	2.63	-1.00	-0.80	-0.83
GPM6A	2.43	-0.60	-0.96	-0.87
CD14	0.82	1.96	-1.48	-1.30
MAG	0.52	-0.50	2.71	-2.73

Classify Cell Type



均泰生物科技股份有限公司 Genetech Biotech Co., Ltd.

Classify Cell Type – Project classifiers



Choose classifier from i Project classifiers

Garnett classifier

Select data node Clear selection

● Classifier [Train classifier - 0.2.14]

Back Finish

Select Cancel

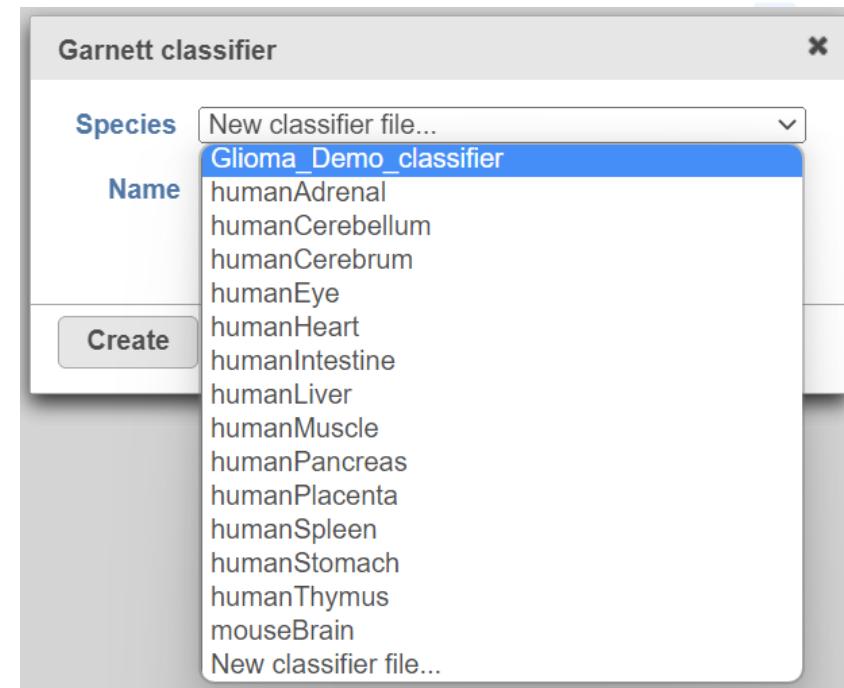
Classify Cell Type – Managed classifiers



Choose classifier from i Managed classifiers ▾

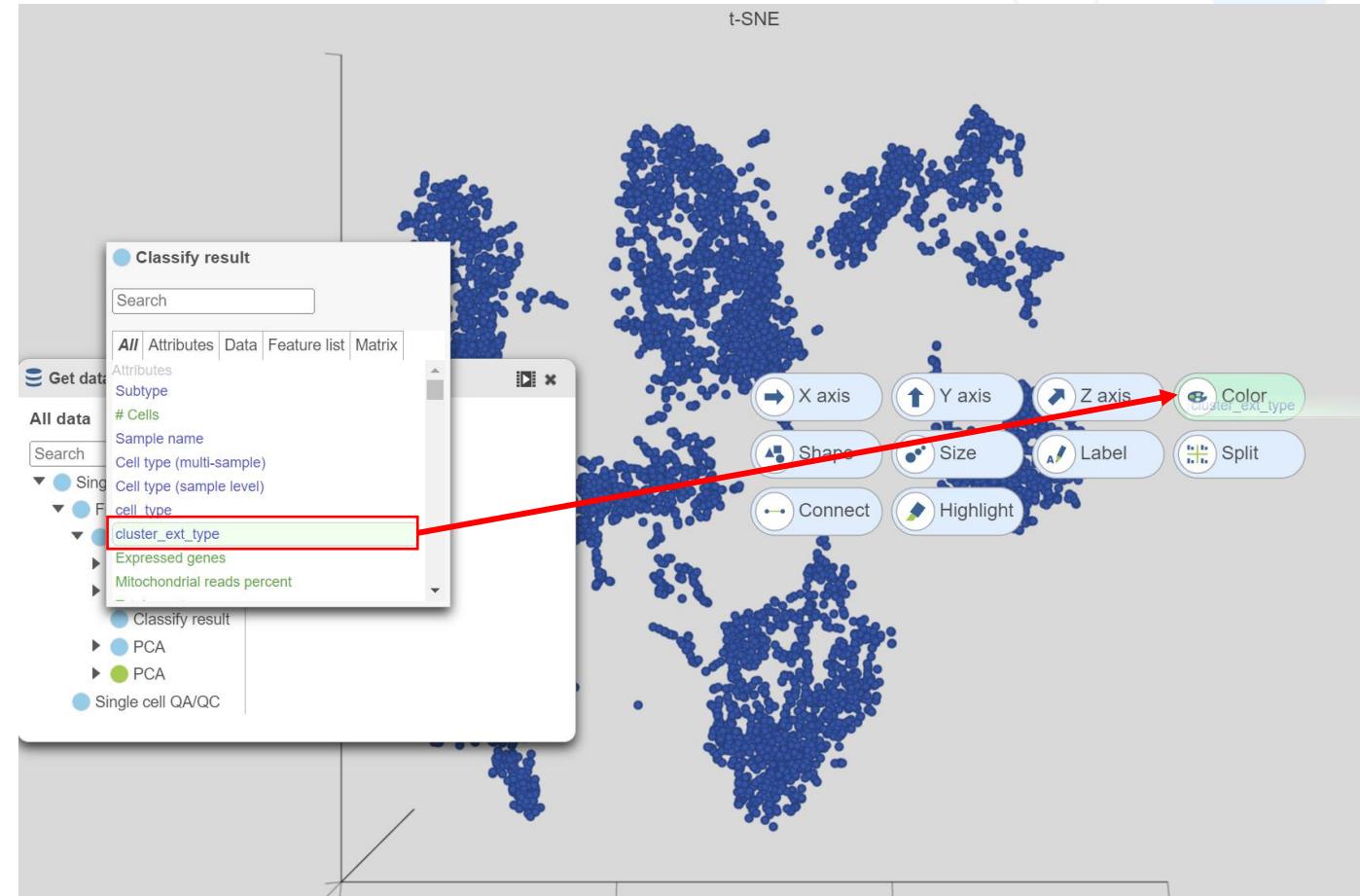
Garnett classifier humanPBMC (Taiwan Genetech Biotech) ▾

Back Finish

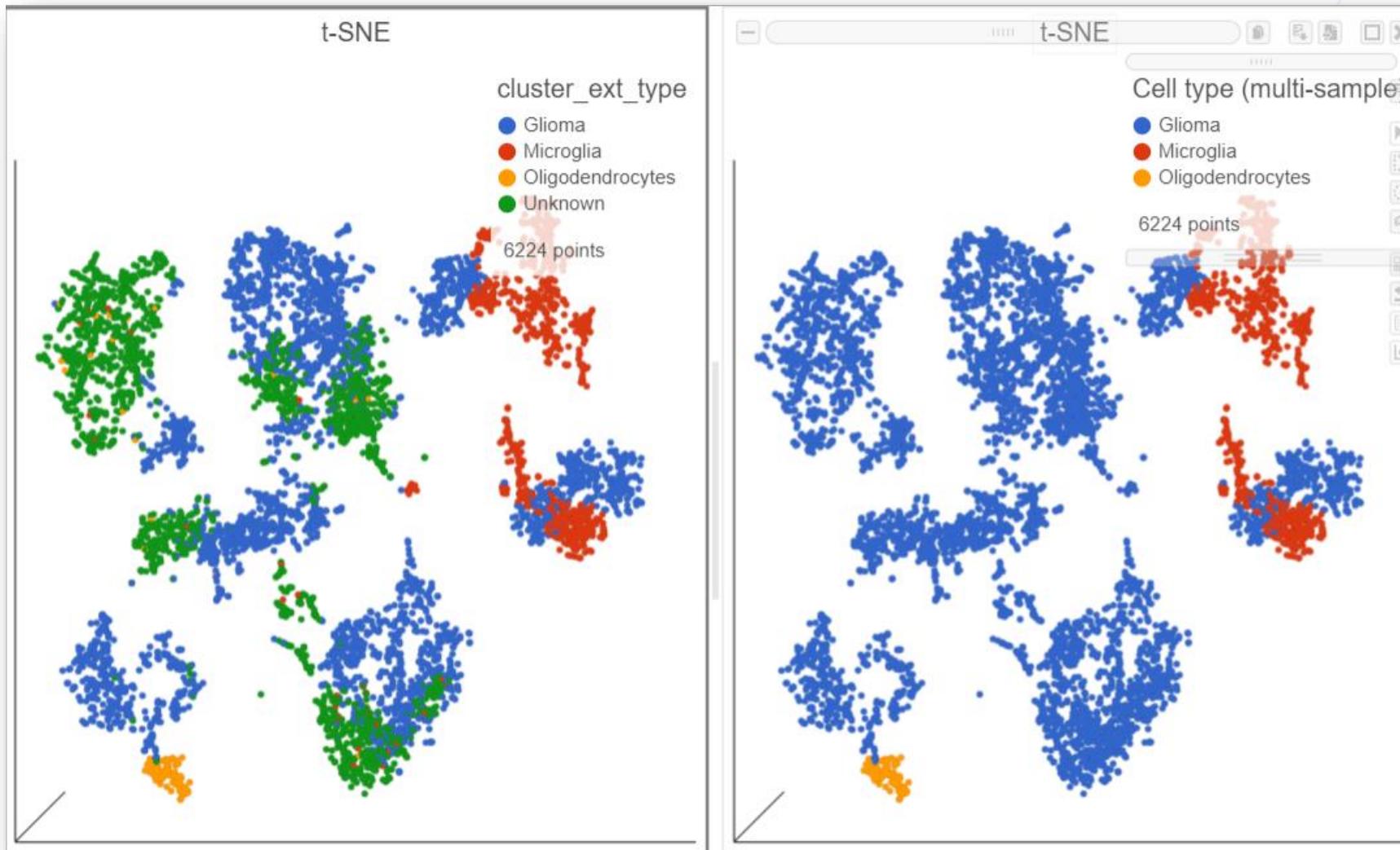


Classification Results

- “**cell_type**” is the cell type assignments directly from Garnett model.
- “**cluster_ext_type**” is the cell type that's determined by expanding cell type assignments to nearby cells using Louvain clustering.



Garnett Classifiers vs. Manual Classification



Plot Interpretation

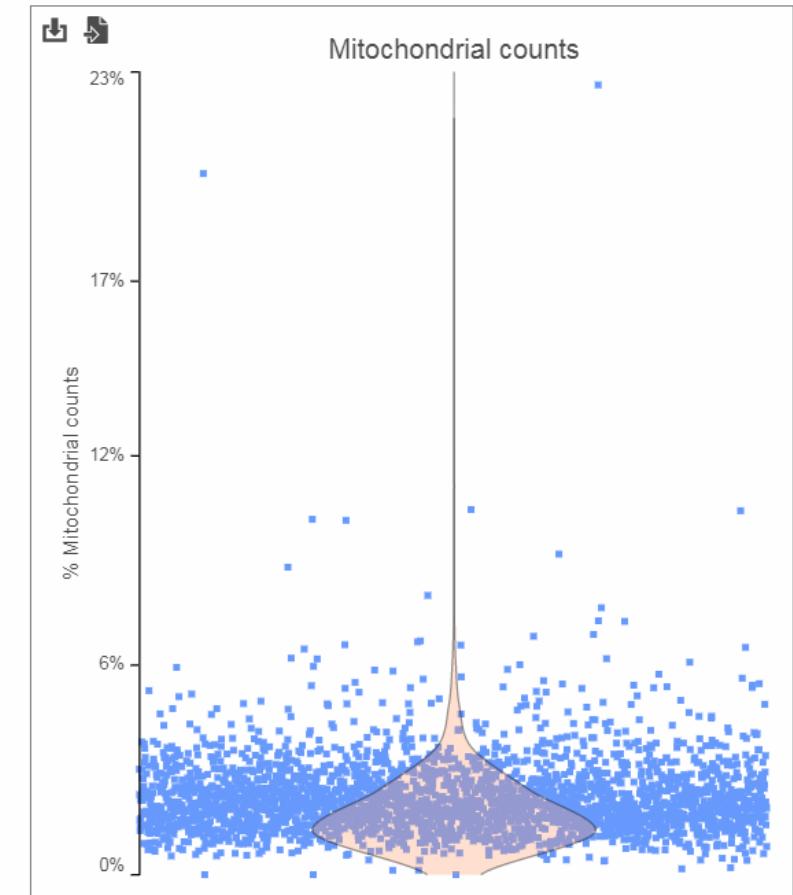
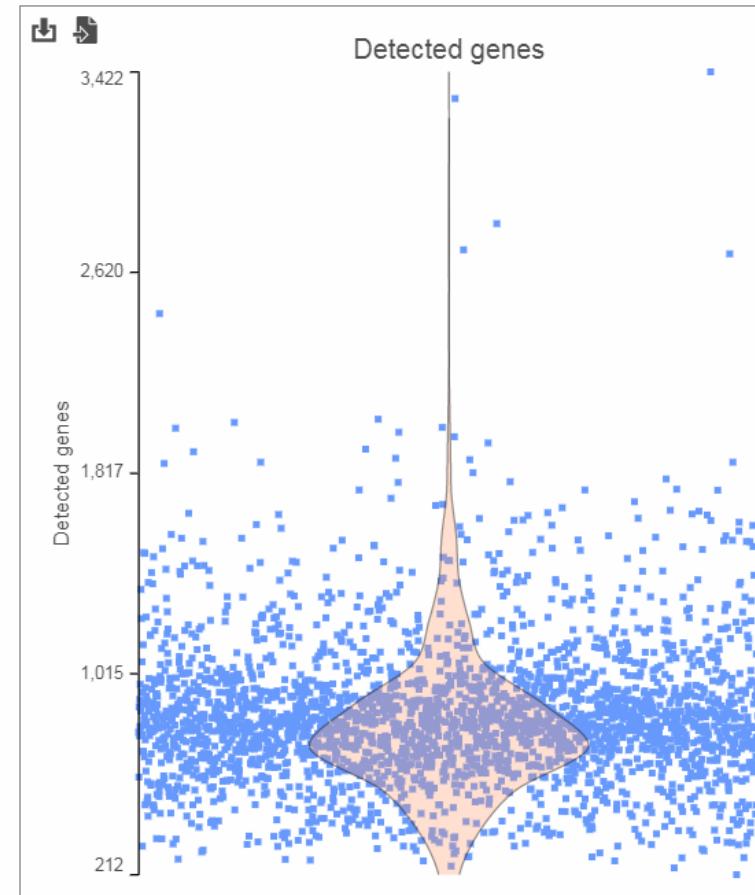
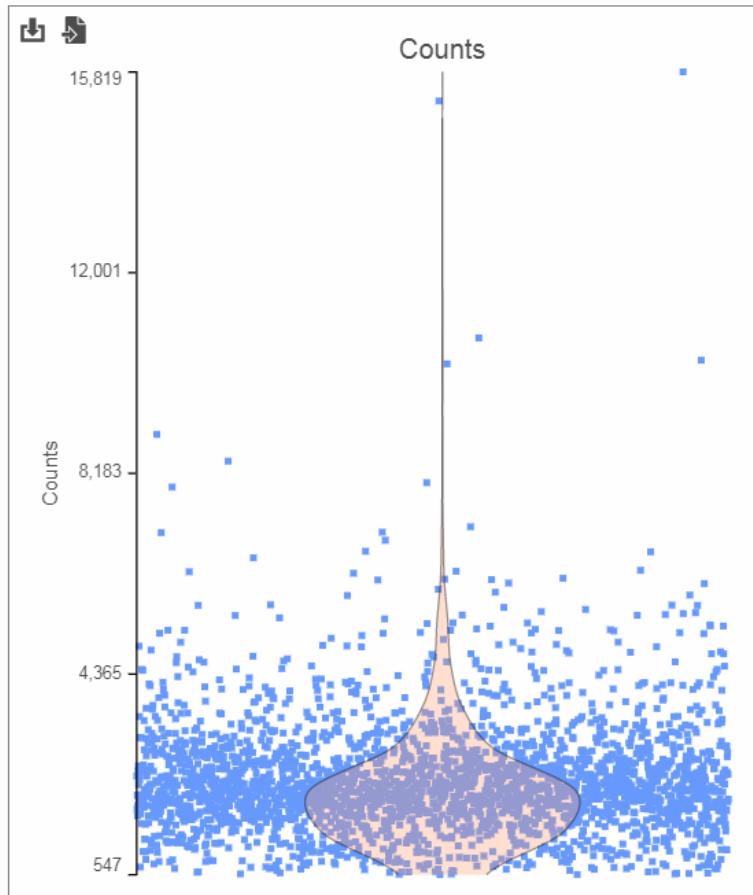


Single cell QA/QC report - Violin Plot



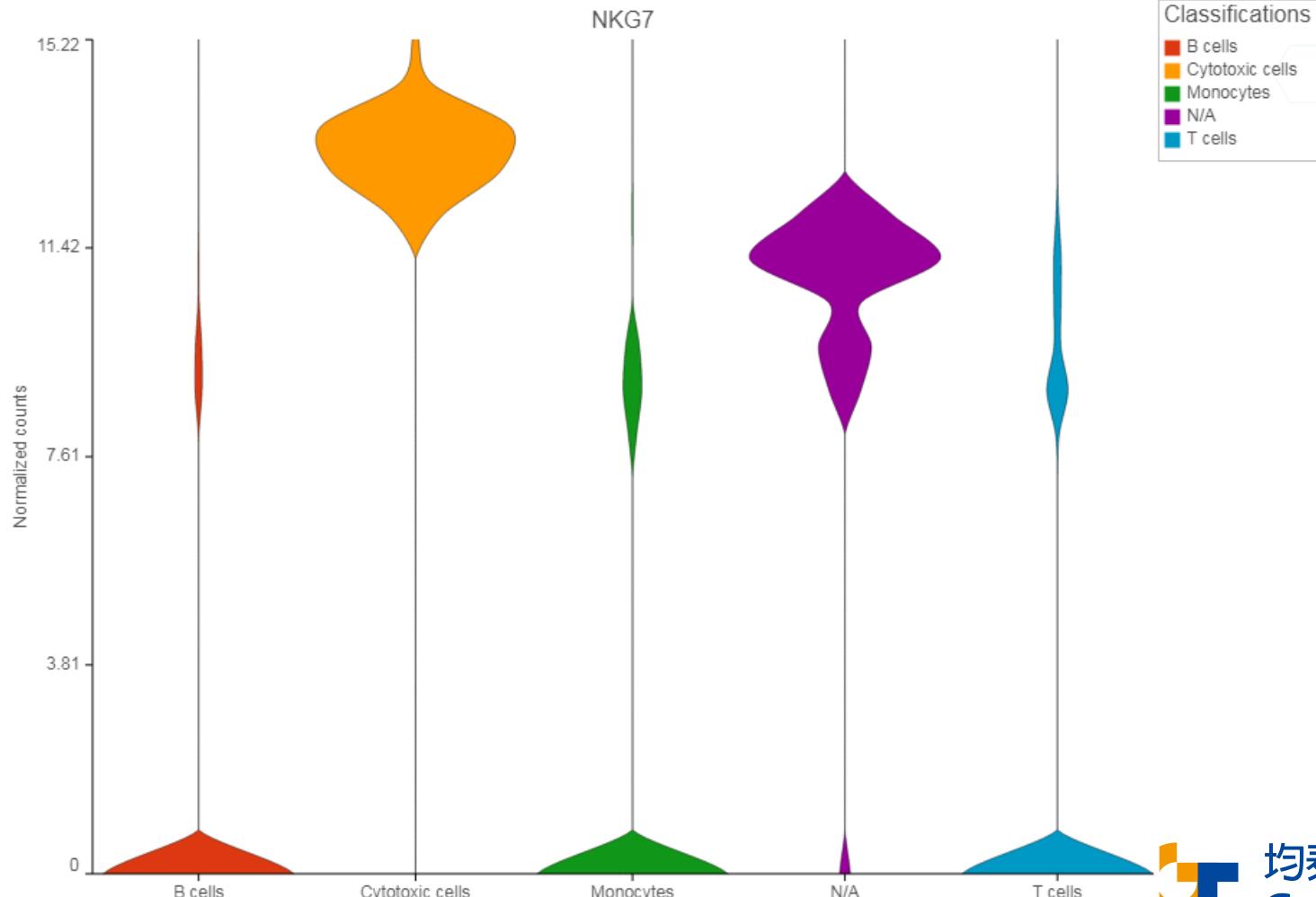
由左至右分別代表細胞中的read數量、基因數量以及Mitochondria gene表達量
X軸沒有意義，目的是為了避免有兩個以上的cells有相同的count重疊看不出來；Y軸代表total count；每個點代表一個細胞
Violin plot 越寬代表密度越大，可以由這張圖明顯看到cell集中於哪個數量區域，並進一步留下較有生物意義的細胞

Selected cells Excluded cells

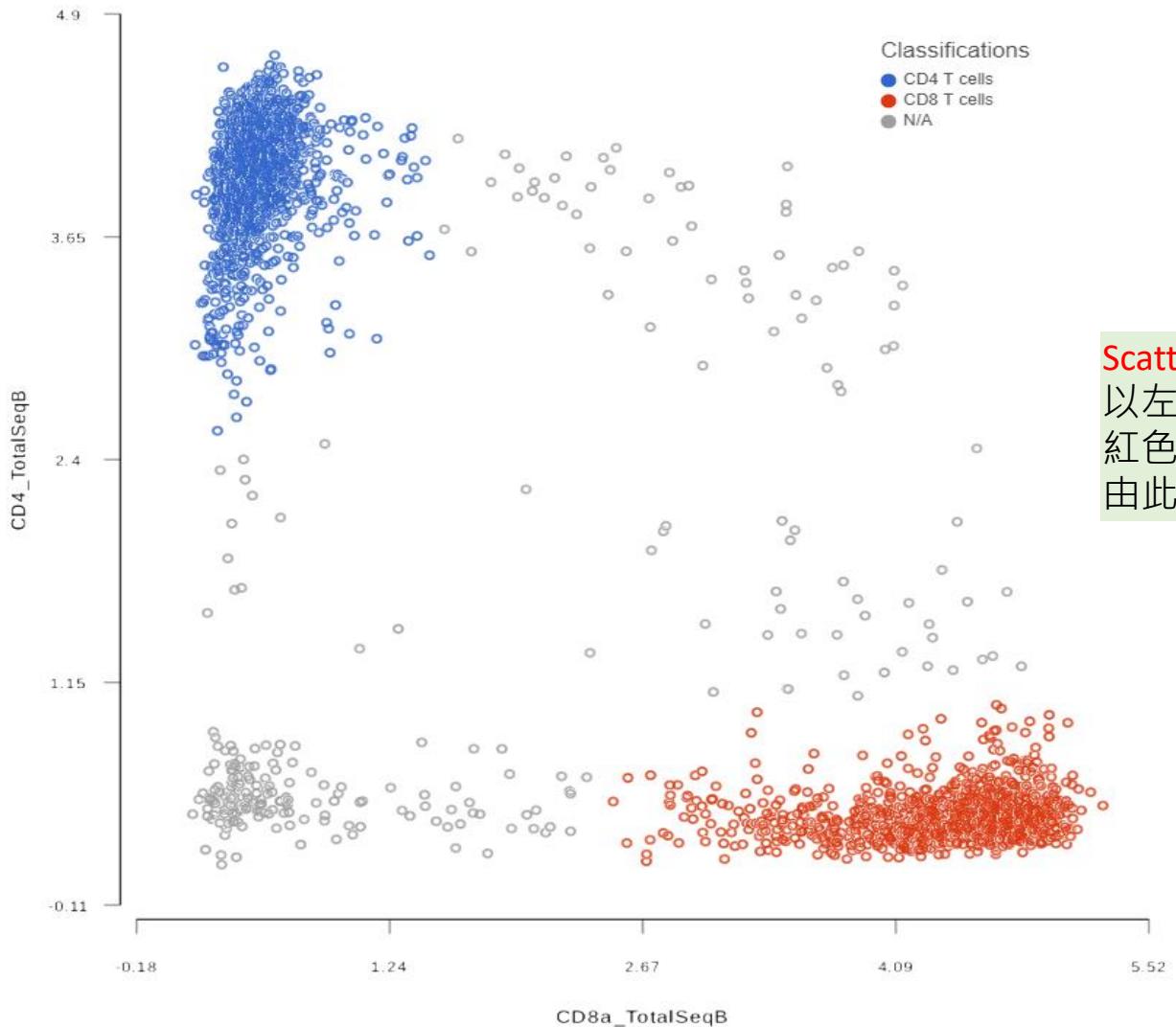


Feature Plot

X軸為不同的細胞類別，Y軸為Normalized後的 Read count數；客戶可自行將細胞分類，並透過Feature Plot了解特定基因在不同類別中的RNA表現量



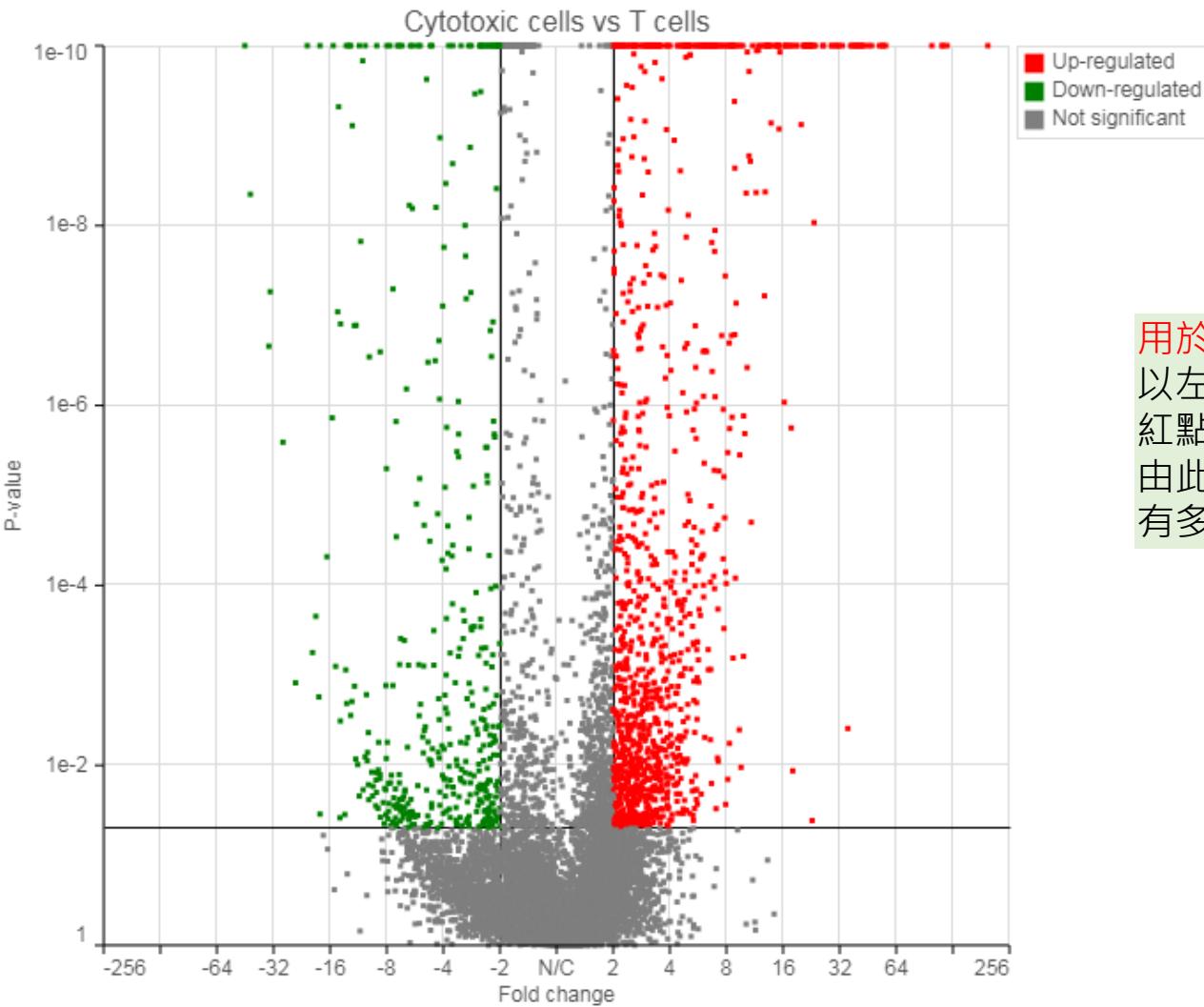
Scatter Plot



Scatter Plot可以看出不同Biomarker在不同種類的細胞是否具有相關性
以左圖說明，XY軸分別是CD8 及CD4兩種biomarker 表達量，
紅色的CD8 T-cell 群有高表達CD8及低表達CD4的特性，CD4 T-cell 群則反之；
由此圖可知這兩個Biomakers能有效分出藍色及紅色這兩個種類的細胞



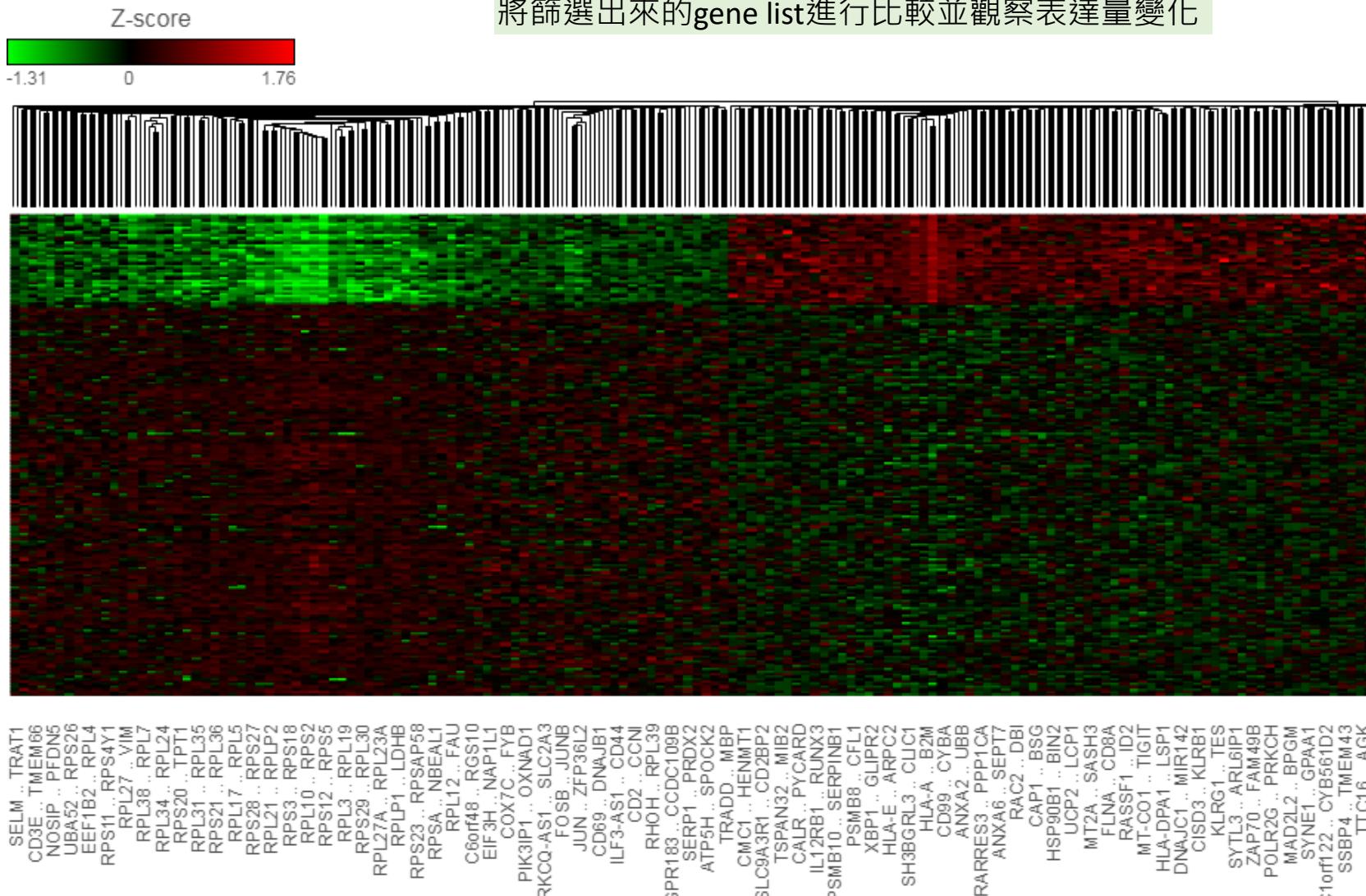
Volcano Plot



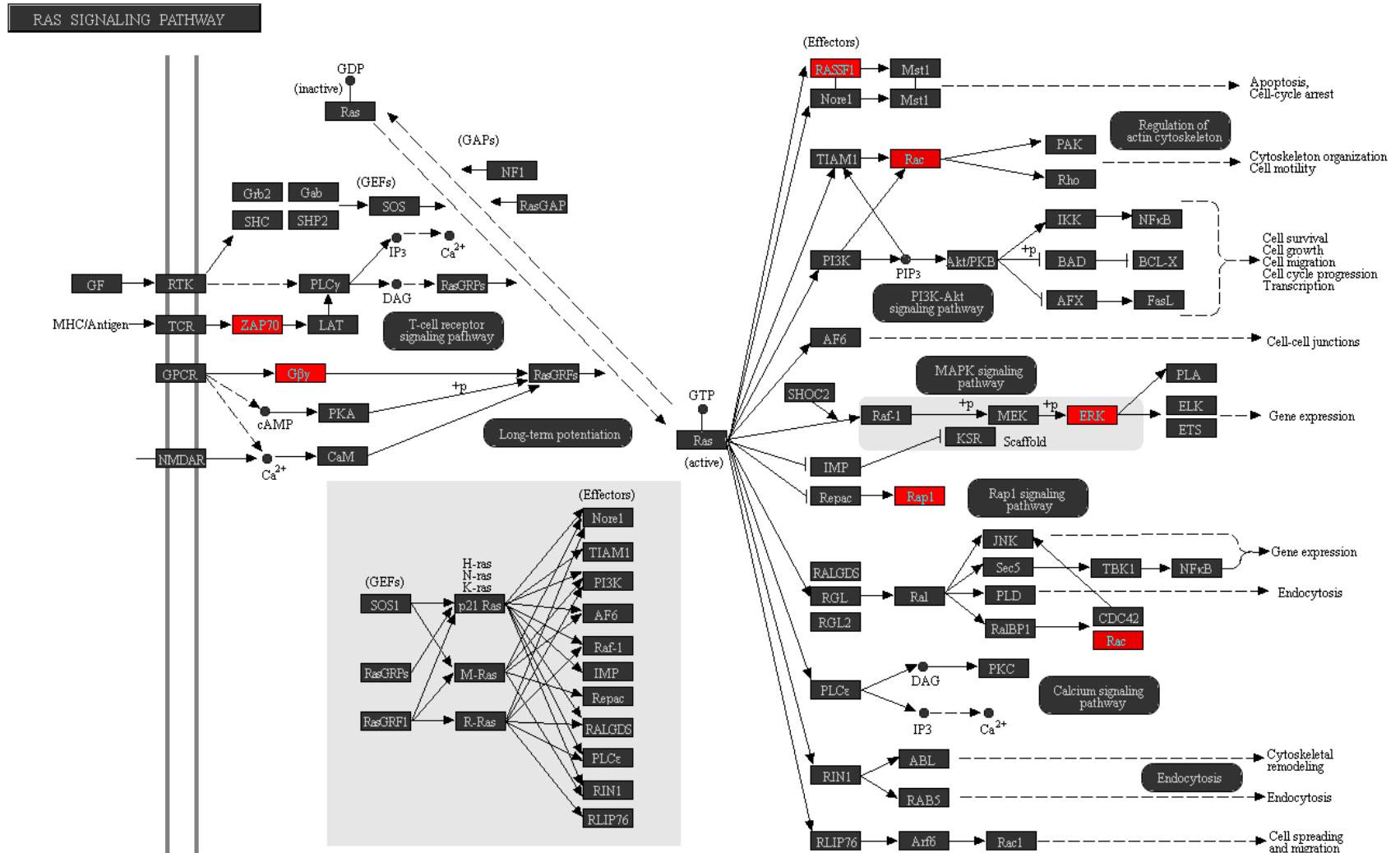
用於查看特定細胞群中高表達基因及低表達基因的數量
以左圖說明，X軸為Fold change，Y軸為P-value；
紅點為Up-regulated gene，綠點為Down-regulated gene
由此圖可看出cytotoxic cells 和 T-cells 這兩個種類的細胞群相比之後，
有多少up-regulated, down-regulated 及 un-change 的基因



Heatmap



KEGG Pathway result

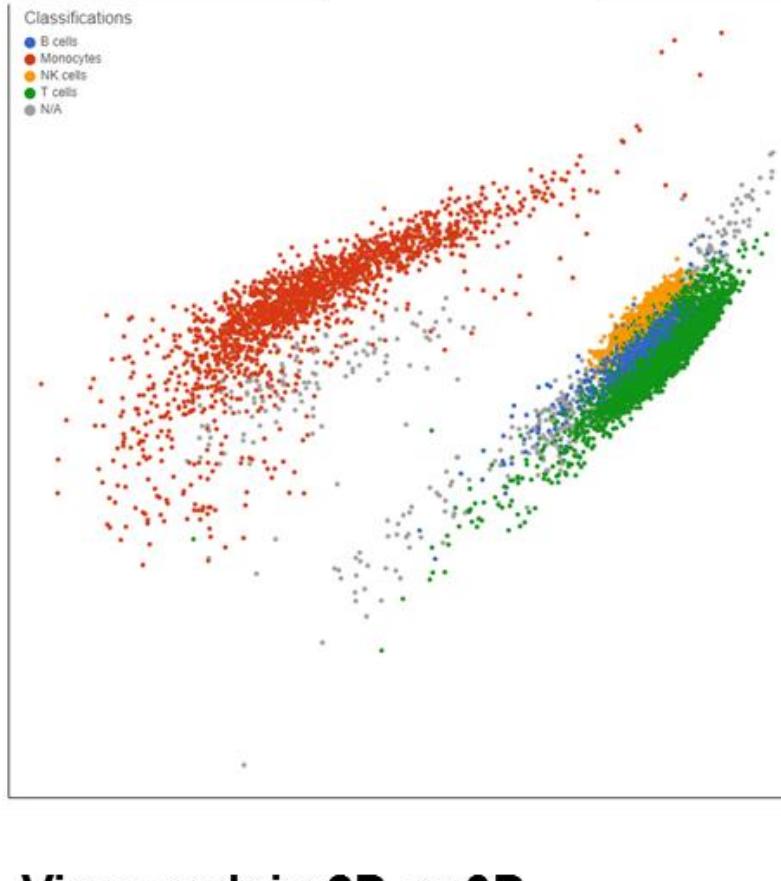


挑出感興趣的Gene list，
並分析這些基因是否影響特定pathway

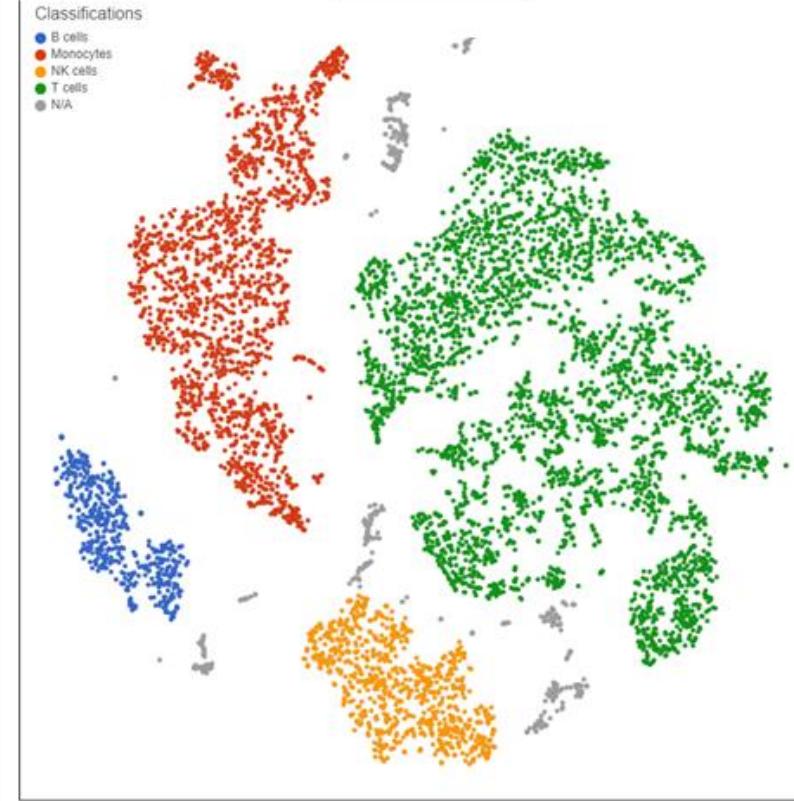
Dimensionality Reduction: PCA, t-SNE, UMAP

細胞分群後的圖表呈現，因每個細胞皆有上千、萬個基因，相等於上千、萬個維度，必須透過降維才能比較各個細胞間不同基因表達量的相關性
PCA, t-SNE, UMAP分別為三種不同的降維方法，是依照各細胞的基因表達量來分群，同一群的細胞所表達的基因越相似
Partek Flow 提供2D及3D的呈現方式，讓使用者更有效了解樣品中不同細胞的相關性

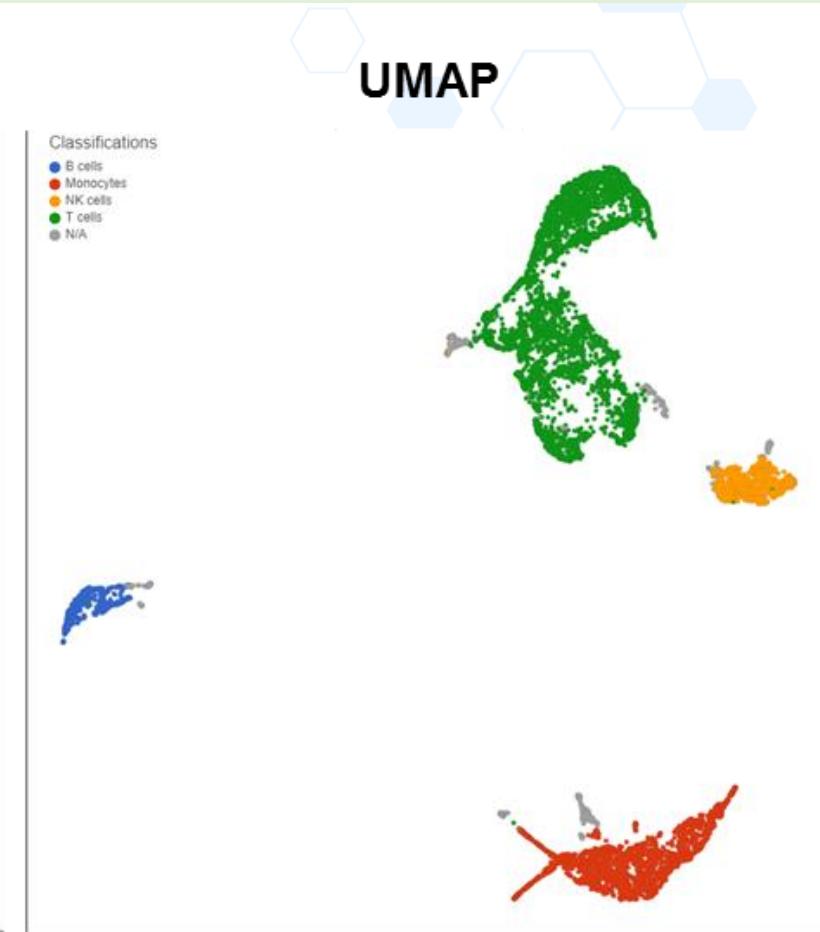
PCA



t-SNE



UMAP



View each in 2D or 3D



Genetech Biotech Co., Ltd.
94

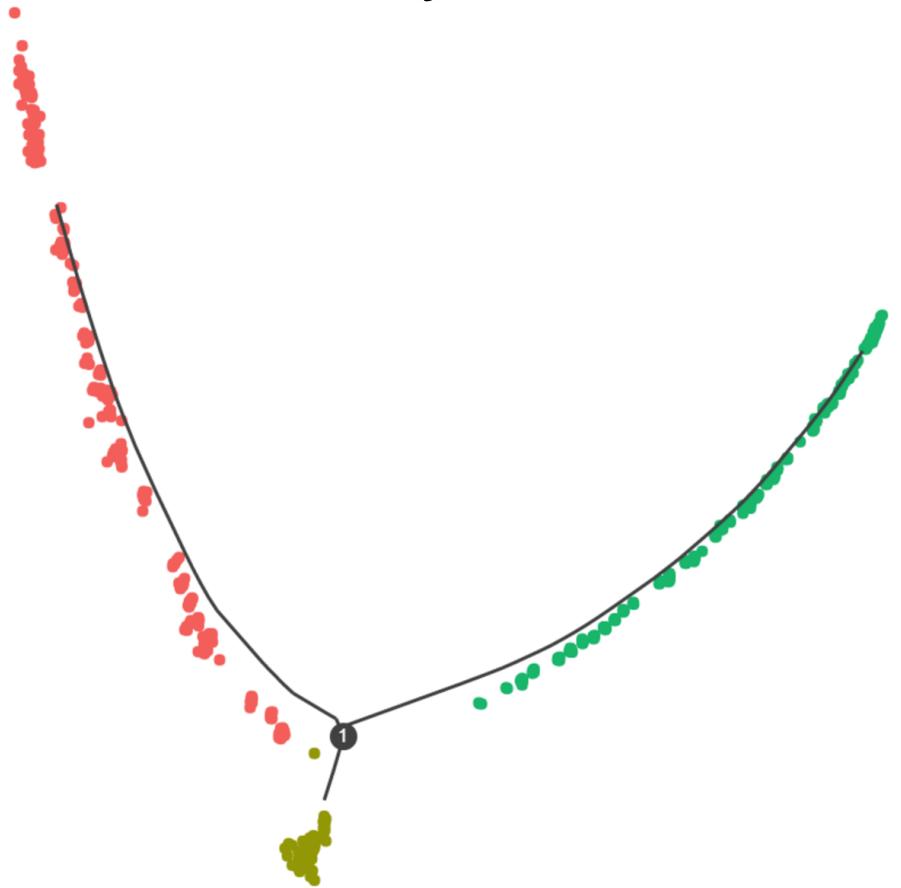
Run Trajectory Analysis with Monocle

透過Trajectory分析，將不同的細胞群依照基因的表達量來預測發育細胞的分化軌跡或細胞的演化過程

Identify States:根據表現量的分佈建構出細胞分化過程的樹狀結構

Calculate Pseudotime: 了解每個細胞在該樹狀結構中的位置，可進一步進行差異分析探索細胞分化過程的重要基因，常用於發育相關研究

Identify States



Calculate Pseudotime

