

From Raw Data to Pathways: Easy Genomics Analysis with Partek Flow

Tanya Chiu
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



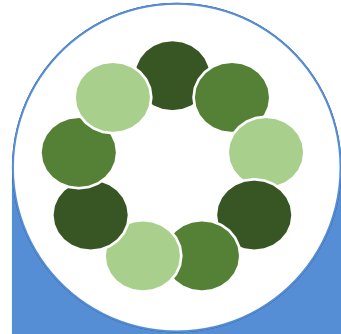
**Collaborative
Web
Environment**



**Powerful
Statistics**



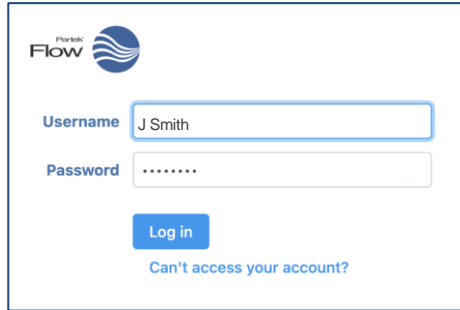
**Interactive
Visualizations**



**Comprehensive
Application
Support**

User Friendly Analysis and Visualizations

Access from
Your Favorite
Browser



Partek
Flow

Username

Password

[Log in](#)

[Can't access your account?](#)



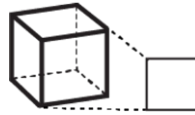
Comprehensive Statistics and Tools



QA/QC



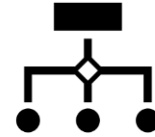
Normalization



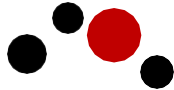
Dimension
Reduction



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	Bowtie
BWA	GSNAP
Isaac	STAR
TopHat	HISAT
TMAP	



QA/QC reports

Pre-alignment
Post-alignment
ERCC spike-in
Single cell quality



Variant calling

Samtools	FreeBayes
LoFreq	Strelka
CNVkit	GATK



Differential analysis

Limma	Negative binomial
DESeq2	Non-parametric ANOVA
Poisson	



Clustering

Hierarchical
K-means
Graph-based



Variant annotation

SnEff	VEP
dbSNP	Custom databases



Metagenomics

Kraken
Alpha and beta diversity
Quantification at taxonomic levels
Differential analysis at taxonomic levels



Data exploration

PCA	Heat map
t-SNE	Violin plot
Dot plot	Histograms
Box plot	Chromosome view
Pathway	2D & 3D Scatter Plot
Bar chart	Pie chart
Bubble map	UMAP



Peak calling

MACS2	Motif detection
TSS plot	

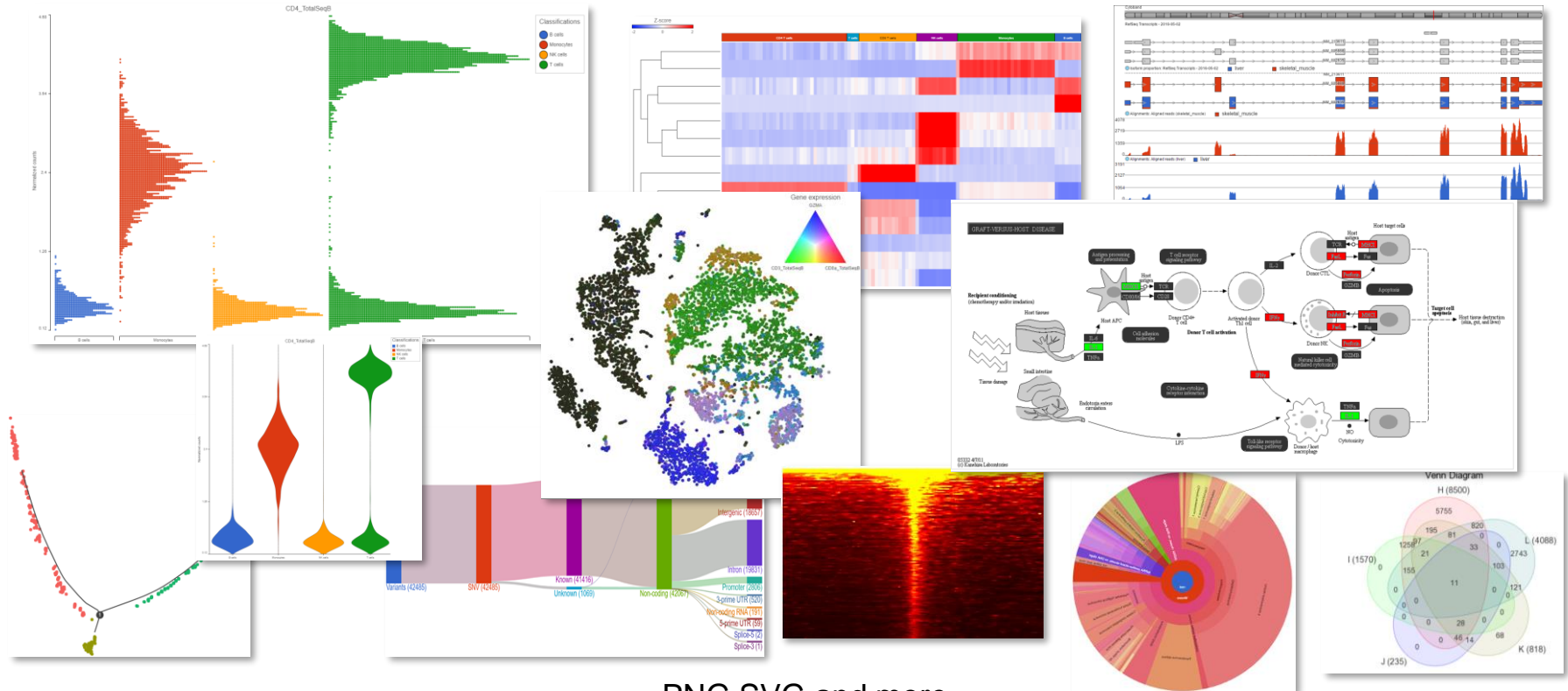


Quantification

Partek E/M	Cufflinks
HTSeq	



Compelling and Publishable Visualizations



PNG SVG and more

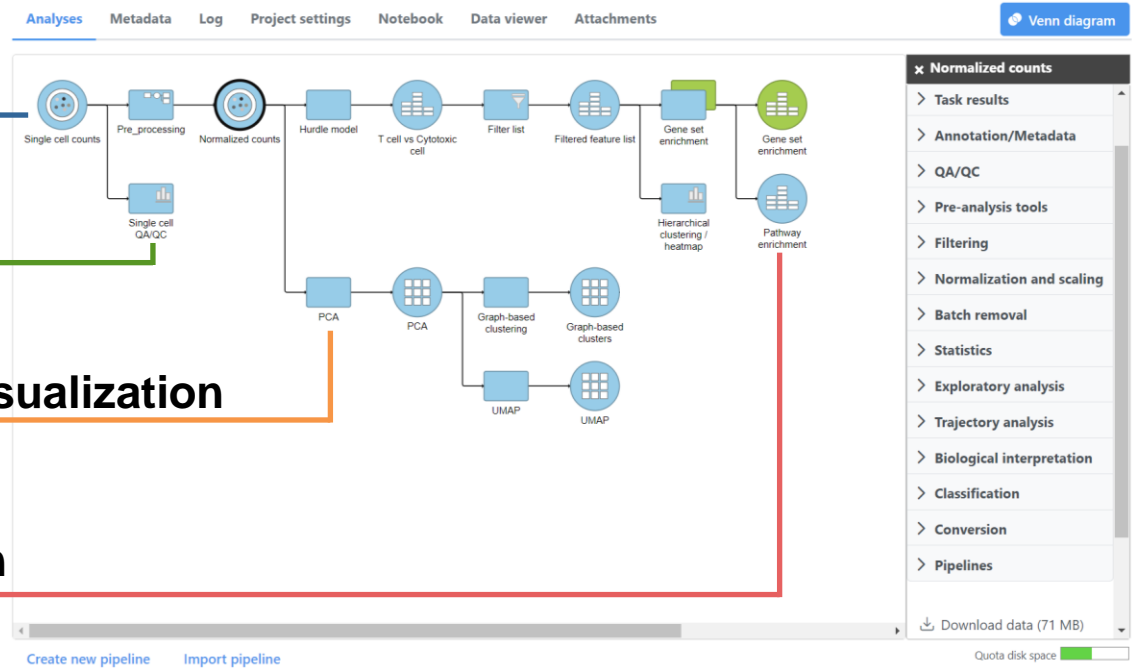
Visual Analysis Process

1 Import Data

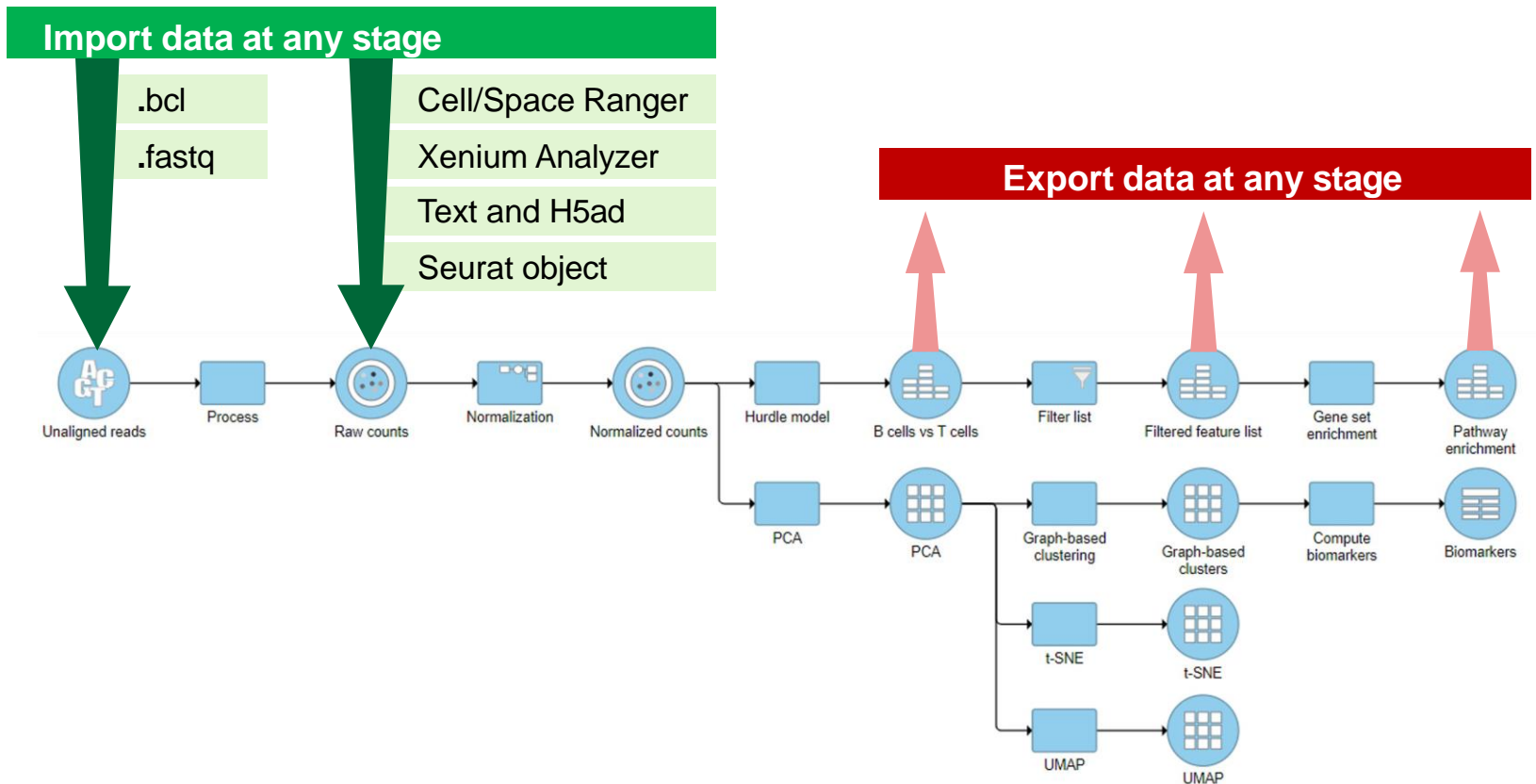
2 QA/QC

3 Powerful Statistics & Visualization

4 Biological Interpretation

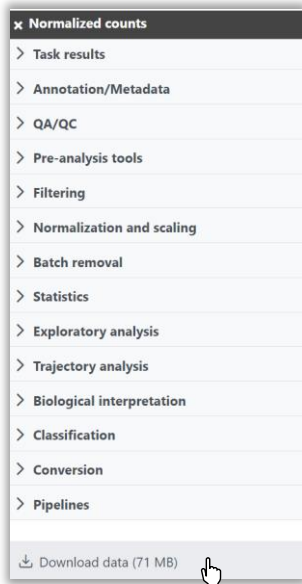
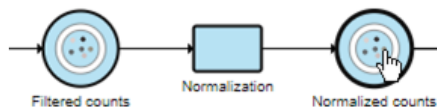


Import and Export Data at Any Stage

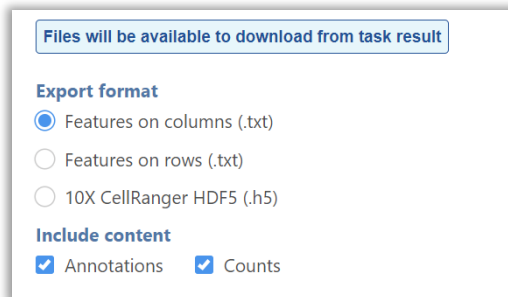


Export Data

Choose Any Data

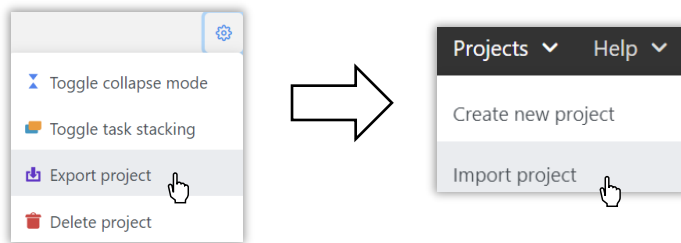


Download in Industry Standard Formats



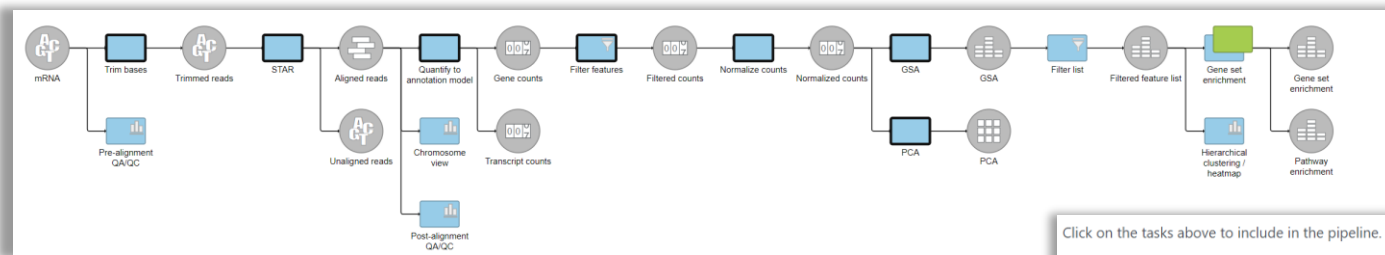
FASTQ, BAM, TXT, and more

Export and Import Analysis Projects



Build, Reuse, and Share Analysis Pipelines

Build Analysis Pipelines



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

Pipeline name: Description:

Section name:

Save, Share, and Manage

Home > Settings > Pipelines

Personal

- My profile
- My preferences

System




- System information
- System preferences
- Single sign-on
- LDAP
- Help widget
- Logging

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

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

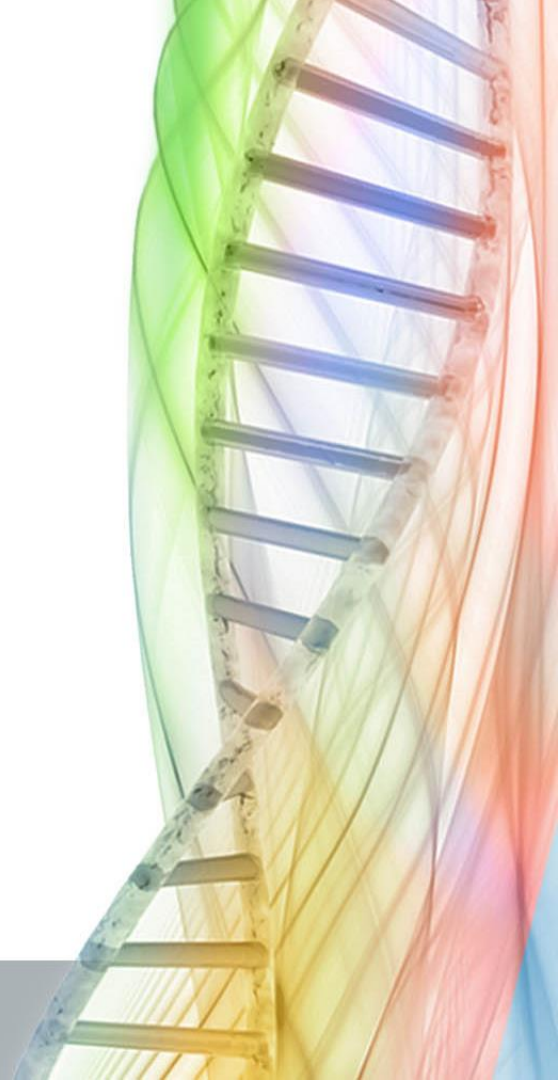
Compatible with All Major Genomics Formats and Assays



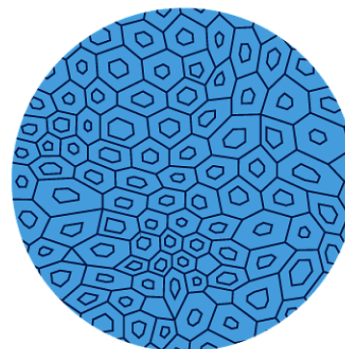
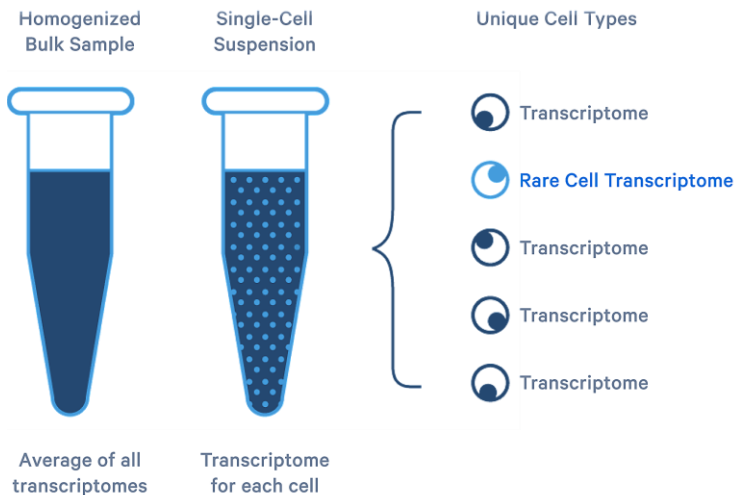
Available Toolkits

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

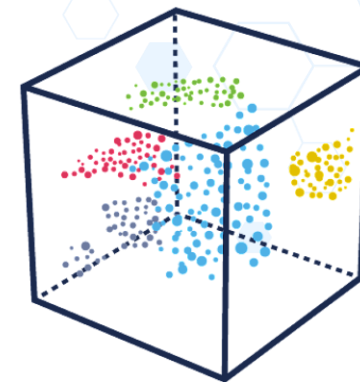
Single Cell Analysis



Introduction of Single-cell Analysis



Tissue Specimen with a spatial relationship between cells.



Relationship between cells by similarity of gene expression.

Supports All Major Single Cell Platforms



1CELLBIO

10X
GENOMICS®



FLUIDIGM®



TakaRa



BioLegend®

illumina®

BIO-RAD

Drop-Seq



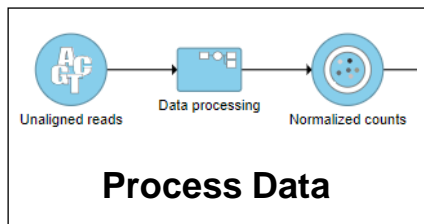
Support for Wide Variety of Single Cell Technologies

- ✓ Single Cell RNA-Seq
- ✓ Whole Transcriptome Single Cell RNA-Seq
- ✓ Gene & Protein Expression
- ✓ ECCITE-Seq
- ✓ Spatial Transcriptomics
- ✓ Trajectory analysis

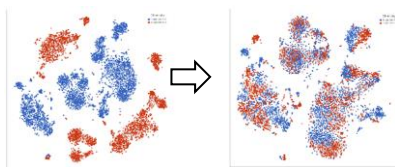


Data Processing and Analysis, All in One Place

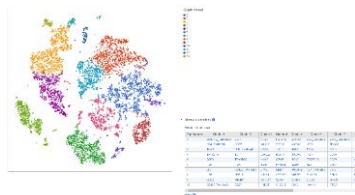
FASTQ, BAM, or Counts



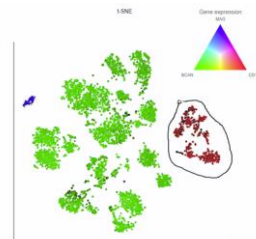
Batch correction



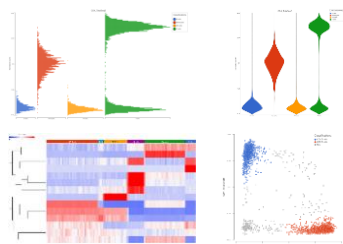
Clustering



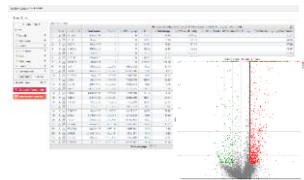
Interactive Visual Analysis



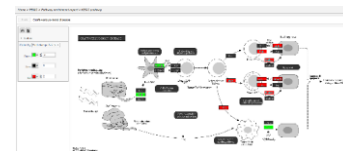
Data Visualization



Differential Analysis



Pathway Analysis



Demo



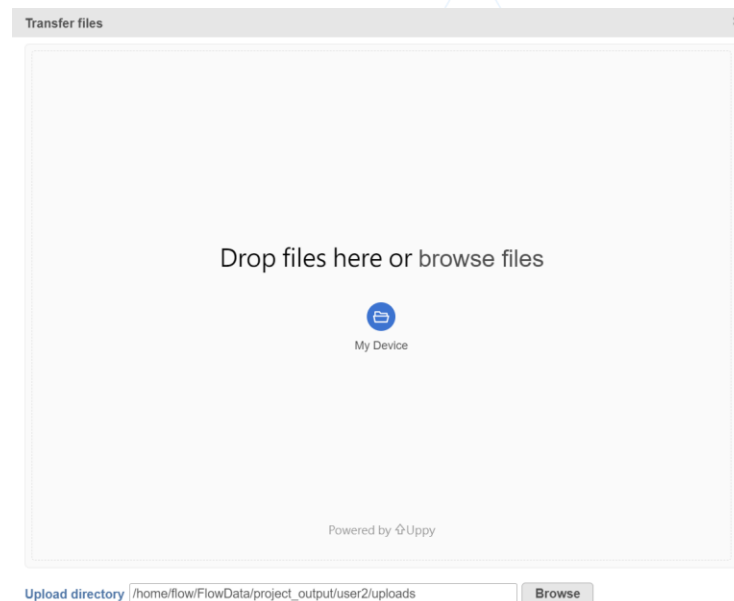
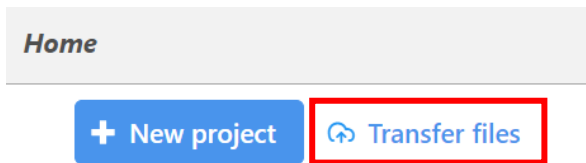
Experiment Description

- 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**



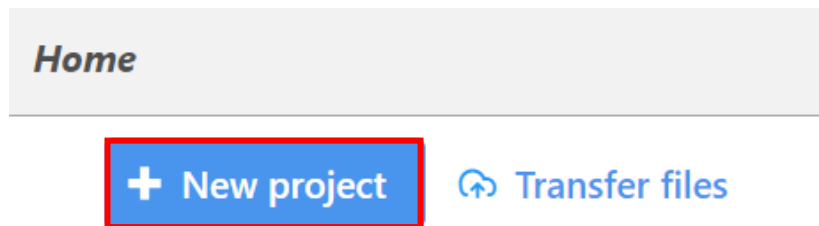
Transfer files

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

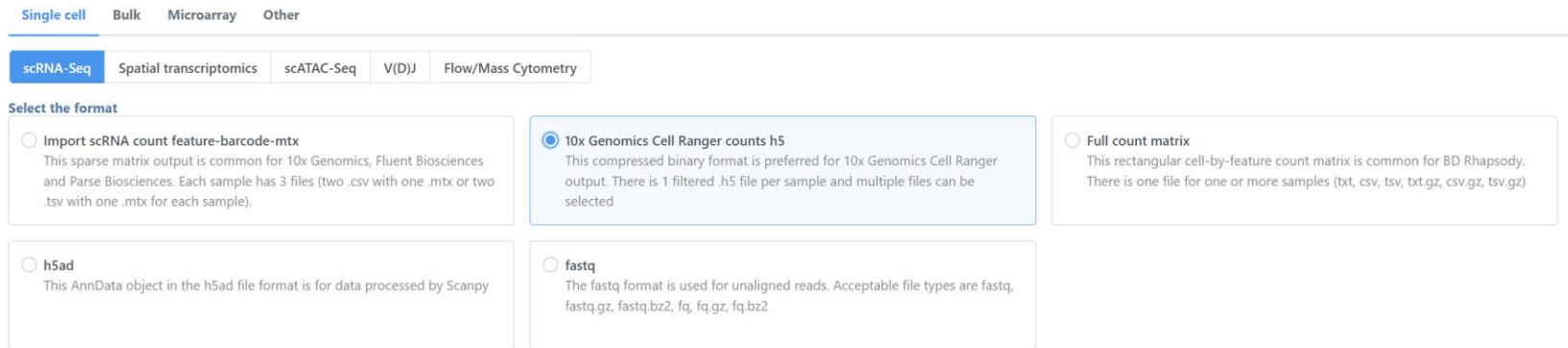


Create a new project

- Click **New project** from home page



Import your own data



The screenshot shows a web interface for importing data. At the top, there are tabs for 'Single cell', 'Bulk', 'Microarray', and 'Other'. Below these, there are sub-tabs for 'scRNA-Seq', 'Spatial transcriptomics', 'scATAC-Seq', 'V(D)J', and 'Flow/Mass Cytometry'. The 'scRNA-Seq' tab is active. Underneath, there is a section titled 'Select the format' with five radio button options:

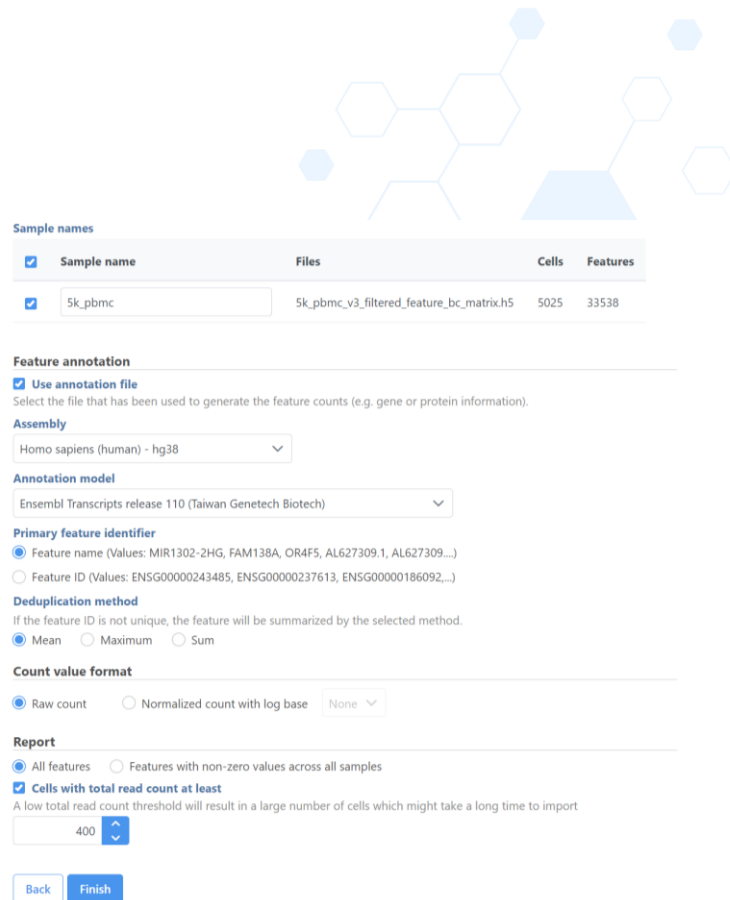
- Import scRNA count feature-barcode-mtx
This sparse matrix output is common for 10x Genomics, Fluent Biosciences and Parse Biosciences. Each sample has 3 files (two .csv with one .mtx or two .tsv with one .mtx for each sample).
- 10x Genomics Cell Ranger counts h5
This compressed binary format is preferred 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 checkbox and set the annotation
 - Assembly: Homo sapiens (human) - hg38
 - Gene annotation: Ensembl transcripts release 110
- Click **Finish** to import sample



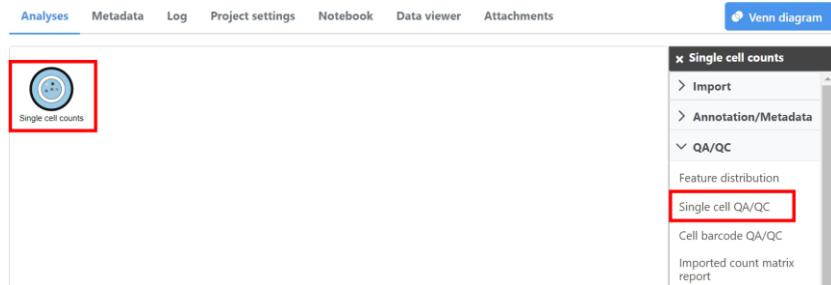
The screenshot displays the 'Specify Annotation' step in a web application. At the top right, there is a decorative graphic of interconnected hexagons. Below it, a table lists sample information:

Sample names	Files	Cells	Features
<input checked="" type="checkbox"/> Sample name			
<input checked="" type="checkbox"/> 5k_pbmc	5k_pbmc_v3_filtered_feature_bc_matrix.h5	5025	33538

Below the table, the 'Feature annotation' section is active, with the 'Use annotation file' checkbox checked. A note states: 'Select the file that has been used to generate the feature counts (e.g. gene or protein information)'. The 'Assembly' dropdown is set to 'Homo sapiens (human) - hg38'. The 'Annotation model' dropdown is set to 'Ensembl Transcripts release 110 (Taiwan Genetech Biotech)'. Under 'Primary feature identifier', the 'Feature name' radio button is selected. Under 'Deduplication method', the 'Mean' radio button is selected. Under 'Count value format', the 'Raw count' radio button is selected. The 'Report' section has 'All features' selected. The 'Cells with total read count at least' dropdown is set to 400. At the bottom, there are 'Back' and 'Finish' buttons.

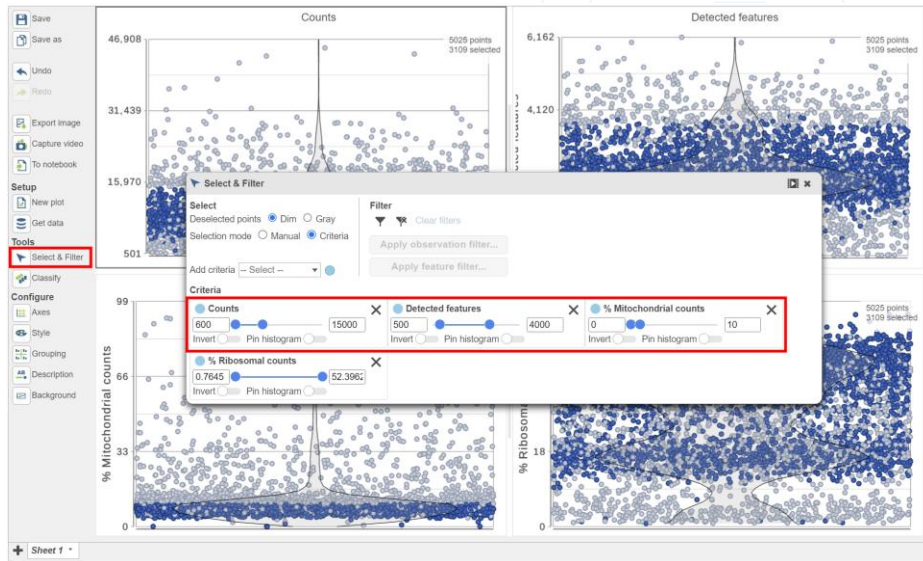
Single Cell QA/AC

- 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



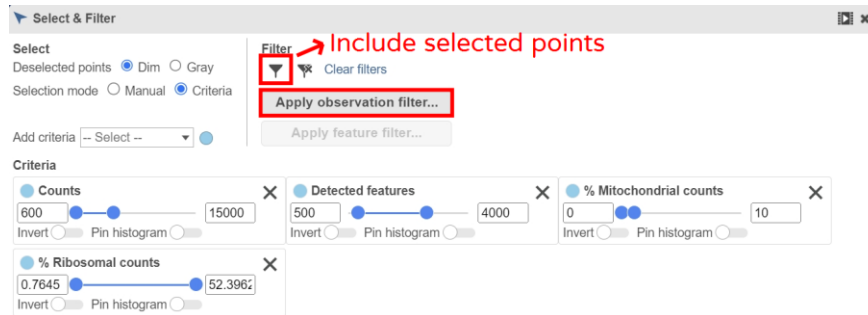
Single Cell QA/AC

- 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

The screenshot displays a software interface with a top navigation bar containing tabs for 'Analyses', 'Metadata', 'Log', 'Project settings', 'Notebook', 'Data viewer', and 'Attachments'. A 'Venn diagram' button is located on the right side of the navigation bar. The main workspace shows a workflow diagram with three nodes: 'Single cell counts', 'Filter counts', and 'Filtered cells'. The 'Filtered cells' node is highlighted with a red box. Below the 'Filter counts' node is a 'Single cell QA/QC' node. On the right side, a task menu is open for the 'Filtered cells' node. The menu is titled 'Filtered cells' and contains several expandable sections: 'Task results', 'Annotation/Metadata', 'QA/QC', 'Pre-analysis tools', and 'Filtering'. The 'Filtering' section is expanded, and the 'Filter features' option is highlighted with a red box. Other options in the 'Filtering' section include 'Filter cells', 'Split by attribute', and 'Downsample cells'.



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

<input checked="" type="radio"/> Noise reduction Exclude features that meet criteria based on descriptive statistics. Calculations are performed for each feature across all cells.	<input type="radio"/> Statistics-based Include a number or percentile of features based on descriptive statistics. Calculations are performed for each feature across all cells.	<input type="radio"/> Metadata Specify logical operations using different annotation fields.	<input type="radio"/> Saved list Specify a saved list of features to include or exclude.	<input type="radio"/> Manual list Manually specify a list of features to include or exclude.
---	--	--	--	--

Filter criteria

Filter features by

Exclude features where value 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

The screenshot displays a workflow interface with a top navigation bar containing 'Analyses', 'Metadata', 'Log', 'Project settings', 'Notebook', 'Data viewer', and 'Attachments'. A 'Venn diagram' button is located in the top right. The main workflow area shows a sequence of nodes: 'Single cell counts' (circular icon), 'Filter counts' (square icon), 'Filtered cells' (circular icon), 'Filter features' (square icon), and 'Filtered counts' (circular icon). The 'Filtered counts' node is highlighted with a red box. Below the 'Single cell counts' node is a 'Single cell QA/QC' node. A dropdown menu is open for the 'Filtered counts' node, showing a 'Filtering' section and an expanded 'Normalization and scaling' section. Within 'Normalization and scaling', the 'Normalization' option is highlighted with a red box. Other options in the 'Normalization and scaling' section include 'Normalize to baseline', 'Scran deconvolution', 'TF-IDF normalization', 'Impute missing values', and 'SCTransform'.



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



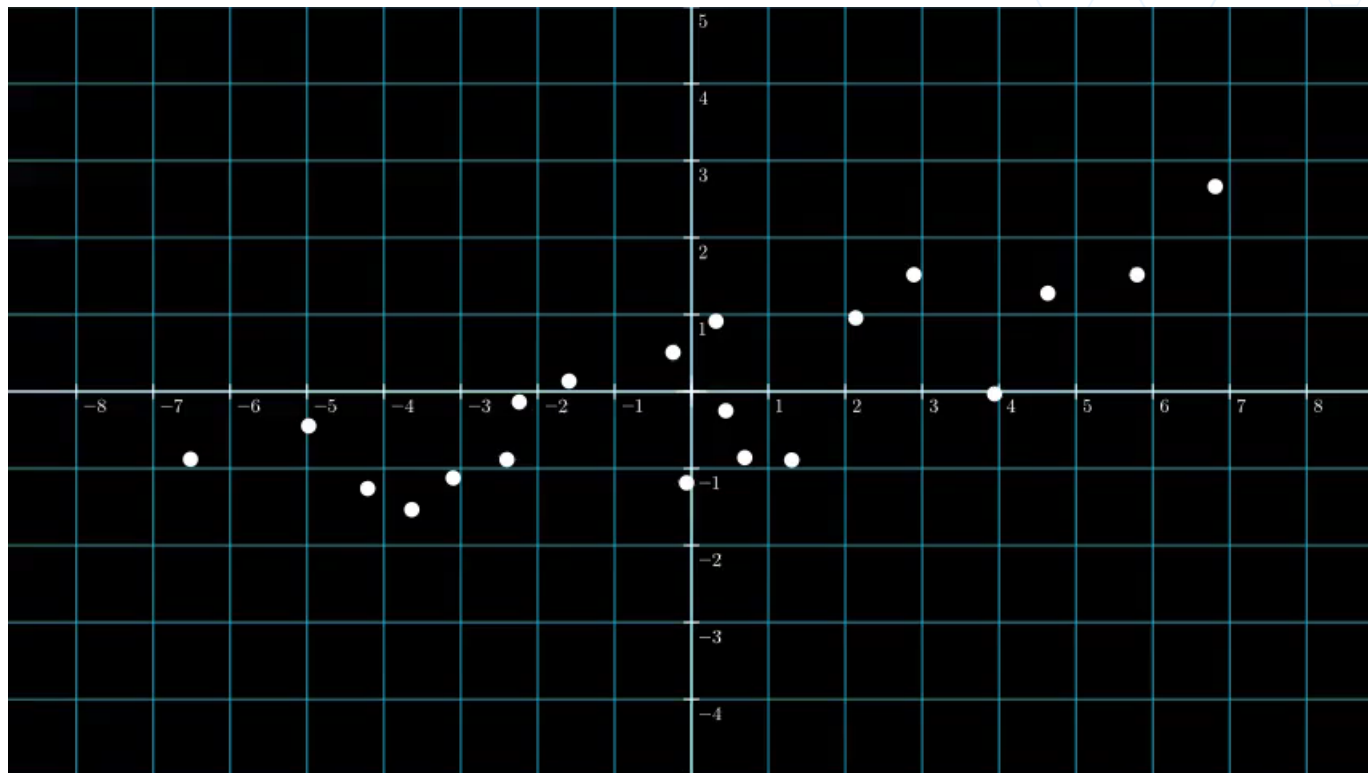
Performing Principal Components Analysis

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

The screenshot displays a software interface for data analysis. At the top, there are navigation tabs: **Analyses**, **Metadata**, **Log**, **Project settings**, **Notebook**, **Data viewer**, and **Attachments**. A **Venn diagram** button is visible on the right. The main workflow consists of several nodes: **Single cell counts**, **Filter counts**, **Filtered cells**, **Filter features**, **Filtered counts**, **Normalize counts**, and **Normalized counts**. The **Normalized counts** node is highlighted with a red box. Below the workflow, a sidebar menu is open, showing the **Normalized counts** section. Under the **Exploratory analysis** sub-section, the **PCA** option is highlighted with a red box. To the right of the workflow, the **PCA** configuration panel is shown. It includes the following settings:

- Features to include in calculation:** **Top** (selected), 2,000 features with the highest vst.
- Number of principal components to calculate:** **Top** (selected), 100 PCs.
- Features contribute:** **By variance** (selected).

PCA



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

Clustering

Clustering algorithm

Three modifications of Louvain clustering algorithm are available

Louvain Louvain with refinement SLM

Compute biomarkers

Queue a "Compute biomarkers" task for the resulting attribute, w

PCA

Number of principal components to calculate

All PCs Top PCs

Advanced options

Option set

[Configure](#)



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



Cluster results

Maximum modularity: 0.848268

Cluster statistics

Total number of clusters 5

Cluster ↑	Size ↑↓	Size % ↑↓
1	1272	40.91%
2	618	19.88%
3	448	14.41%
4	395	12.71%
5	376	12.09%



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

Random ▾

PCA

Number of principal components to calculate

All PCs Top PCs

Advanced options

Option set

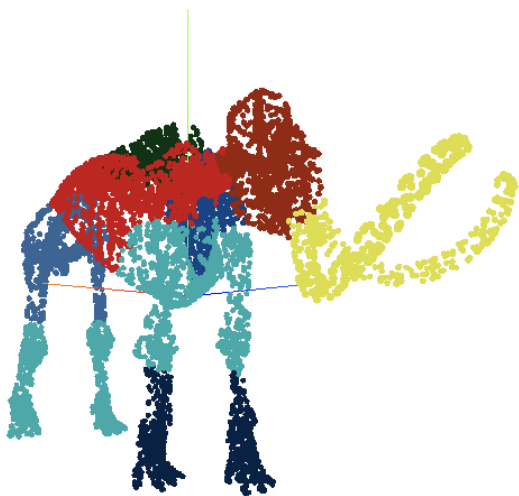
-- Default -- ▾

[Configure](#)



UMAP & t-SNE

Original 3D Data

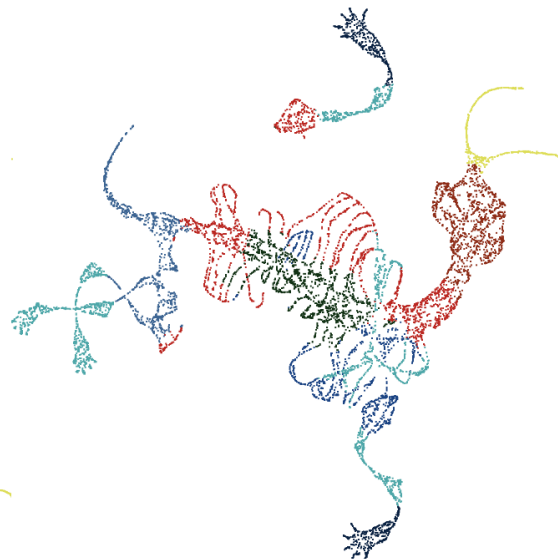


2D t-SNE projection



perplexity: 2000
time: 2h 5m

2D UMAP projection



n_neighbors: 200
min_dist: 0.25
time: 3m 22s




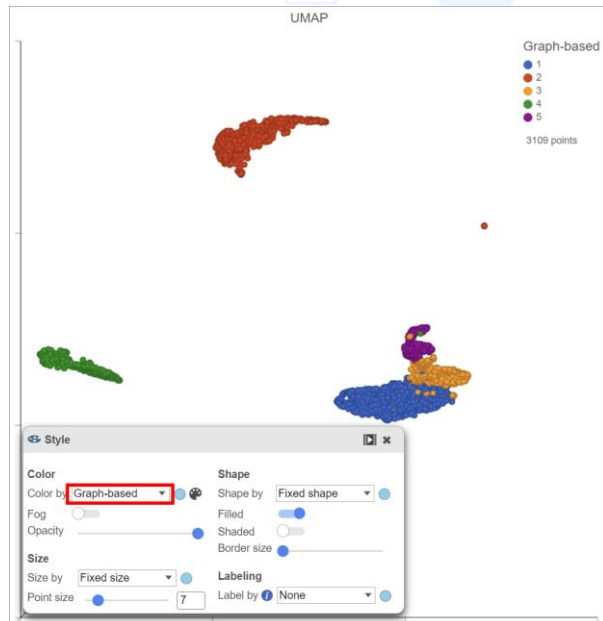
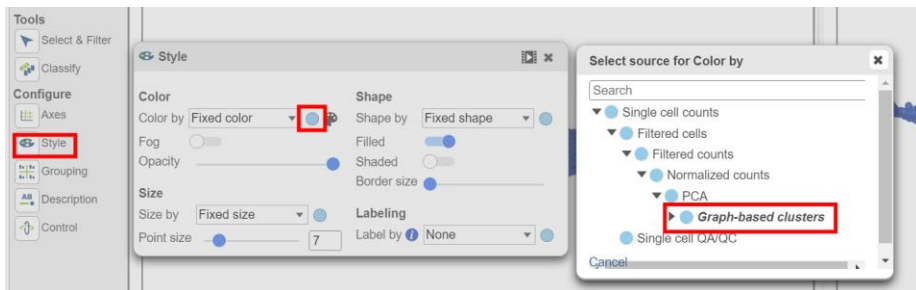
Identifying Cell Types

- 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

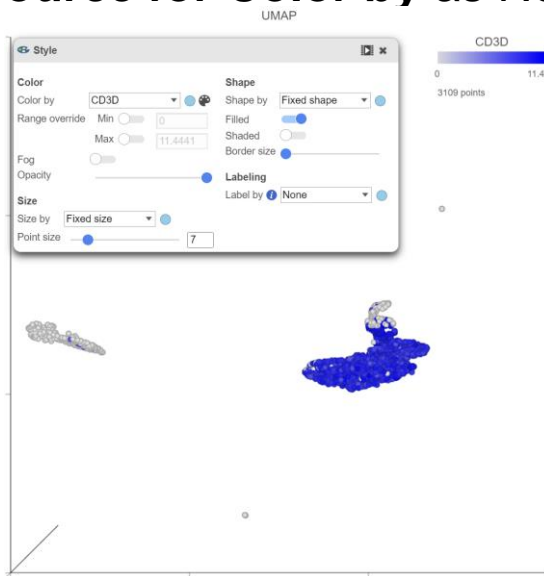
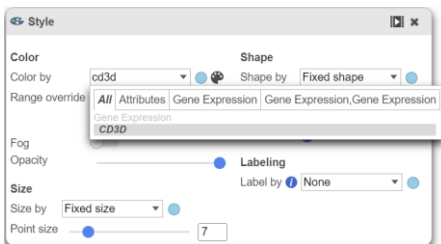
Classify T cells

- 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**



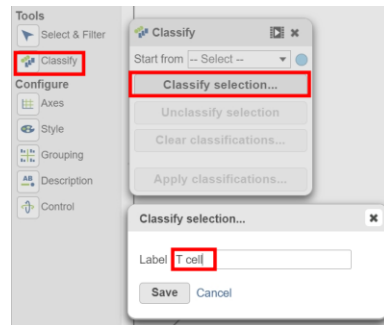
Classify T cells

- 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




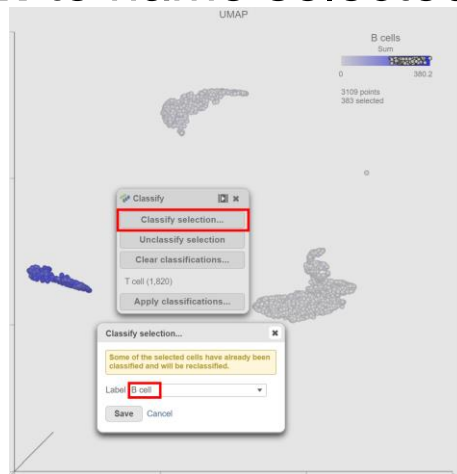
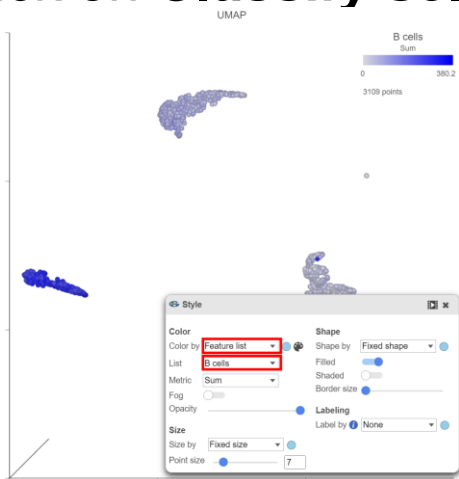
Classify T cells

- 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**



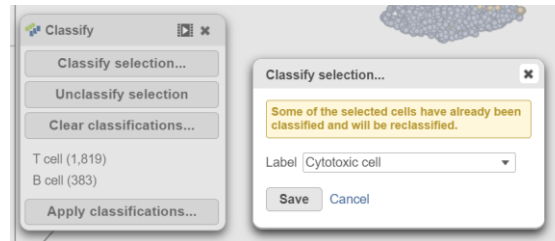
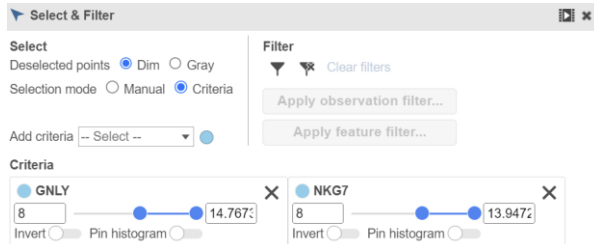
Classify B cells

- 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



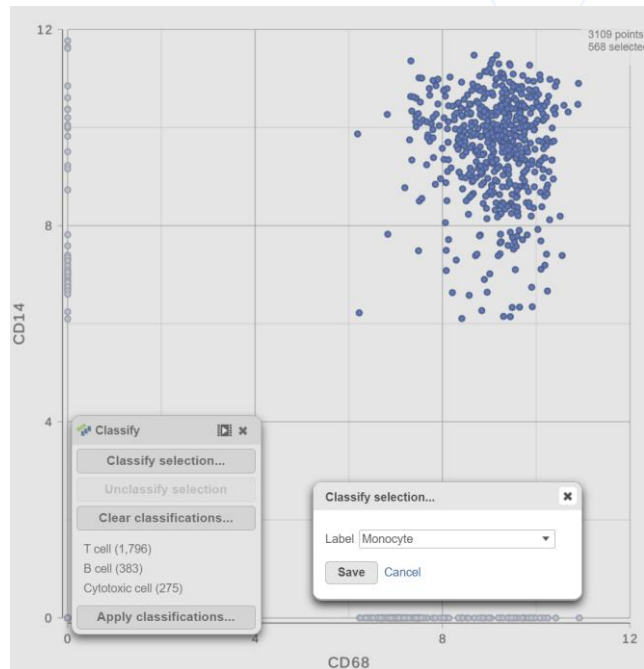
Set plot axes ✕

X axis data

Y axis data

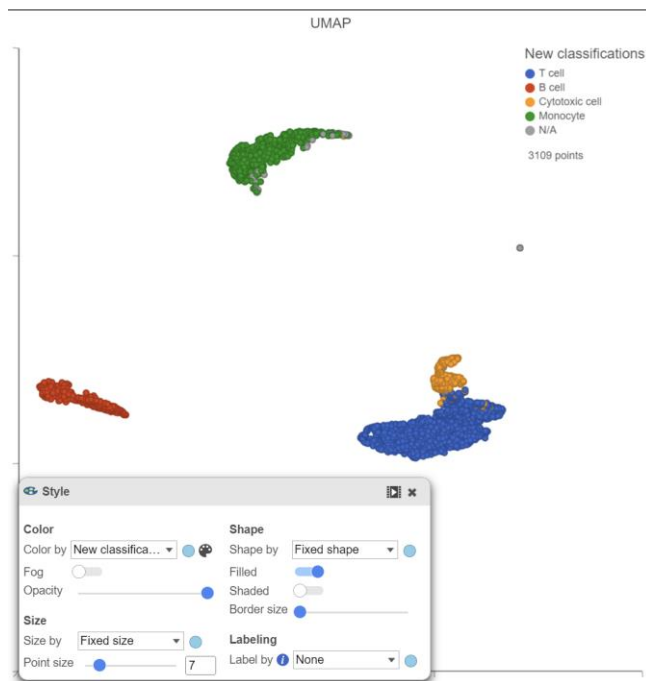
Classify Monocytes

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



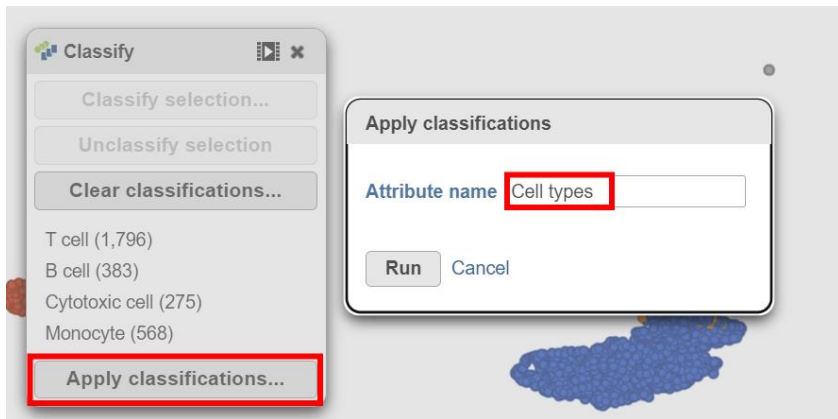
Viewing Classifications

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



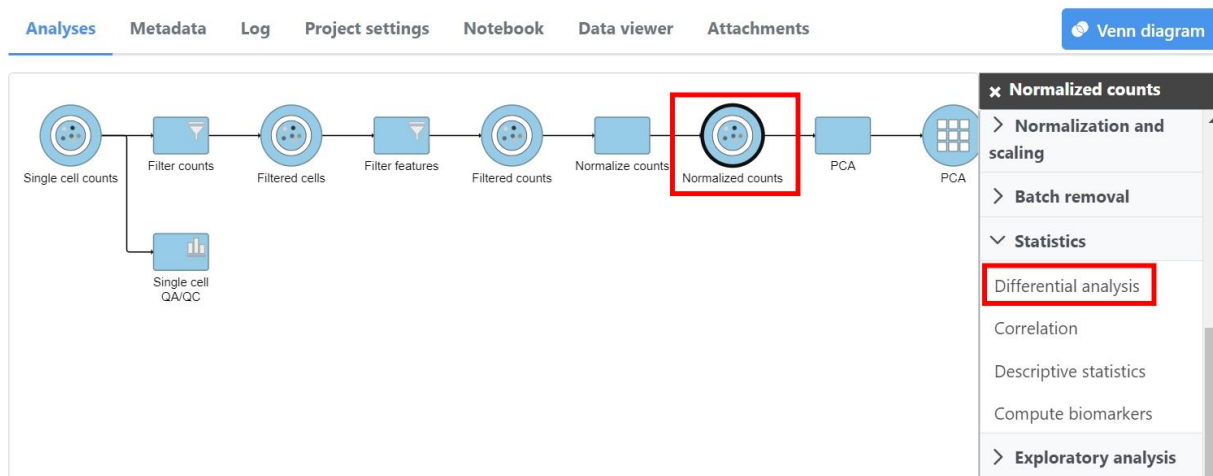
Viewing Classifications

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



Identifying Differentially Expressed Genes

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



Identifying Differentially Expressed Genes

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

Method to use for differential analysis

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.

ANOVA

Recommended for continuous data including bulk and single cell expression 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.

Welch's ANOVA

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

Kruskal-Wallis

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

Gene Specific Analysis

Recommended for data with no replicates in any groups.

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

Combine  Pairwise 

Add comparison

Comparisons

Comparison	Delete
Cytotoxic cell vs. T cell	—




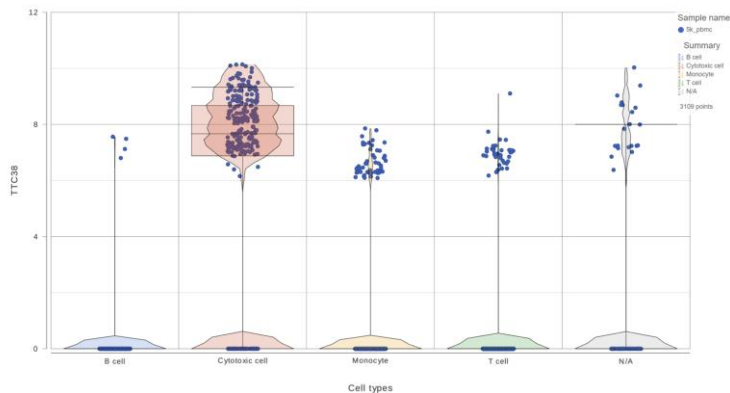
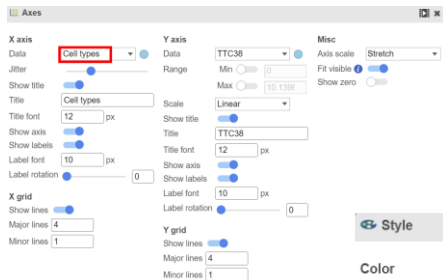
Viewing Hurdle Results

- 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	ARHGDI8	ARHGDI8	0	0	0.71	-1.42	548.70	777.43	0.99	0.99		
7	ARHGDI9	ARHGDI9	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	PRRS	PRRS	0	0	0.09	-10.81	2.59	28.01	0.19	0.62		

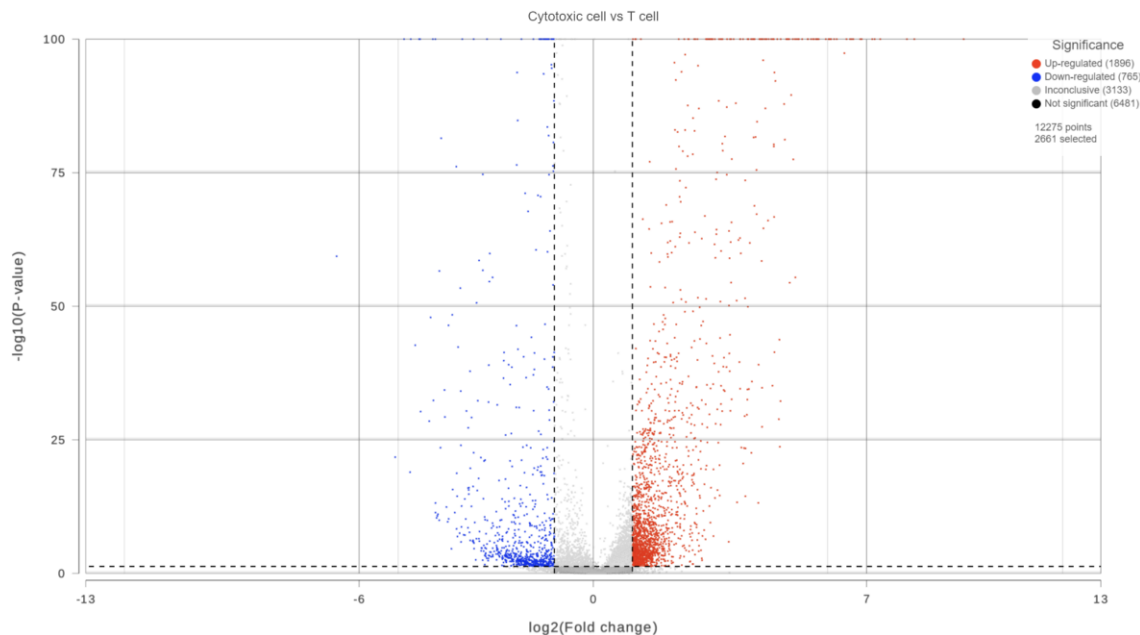
Viewing Hurdle Results

- 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 Hurdle 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

Results: 1491

Filter [Clear all](#)

Gene ID ◀

Gene name ◀

P-value ◀

FDR step up ◻

Less than or... 0.05

0 1

Ratio ◀

Fold change ◻

From -2 to 2

Exclude range

LSMean ◀

Low expressed ◀

Pct(T cell) ◀

Pct(Cytotoxic cell) ◀

[Save filter](#)

Saved filters ⚙️ ◀

[Generate filtered node](#)

[Save as managed list](#)



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*

Plot **i** Heatmap ^f Bubble map ^f

Ordering

Feature order
 Cluster ^f
 Assign order ^f Default order

Cell order
 Cluster ^f
 Assign order ^f Cell types

B cell
Cytotoxic cell
Monocyte
T cell
N/A

Filtering

Filter cells **i**

include Cell types in Cytotoxic cell OR

include Cell types in T cell OR

AND

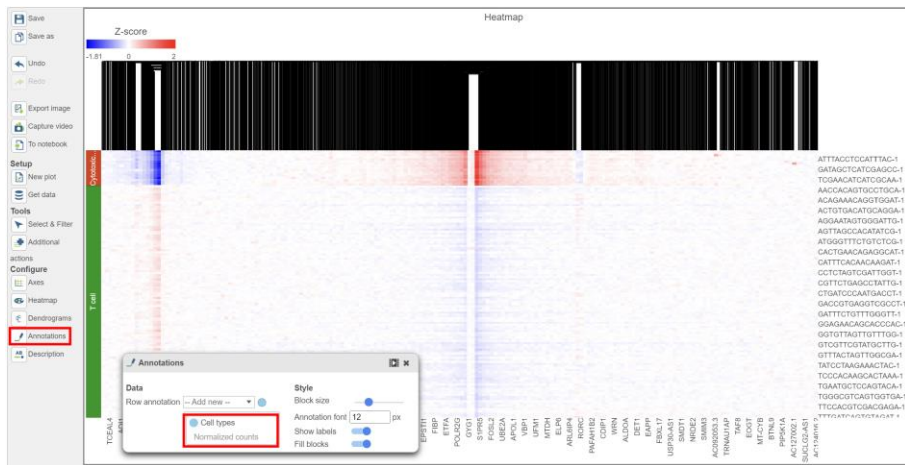
Advanced options

Option set -- Default -- Configure

Back Finish

Hierarchical Clustering Results

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



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) ▼

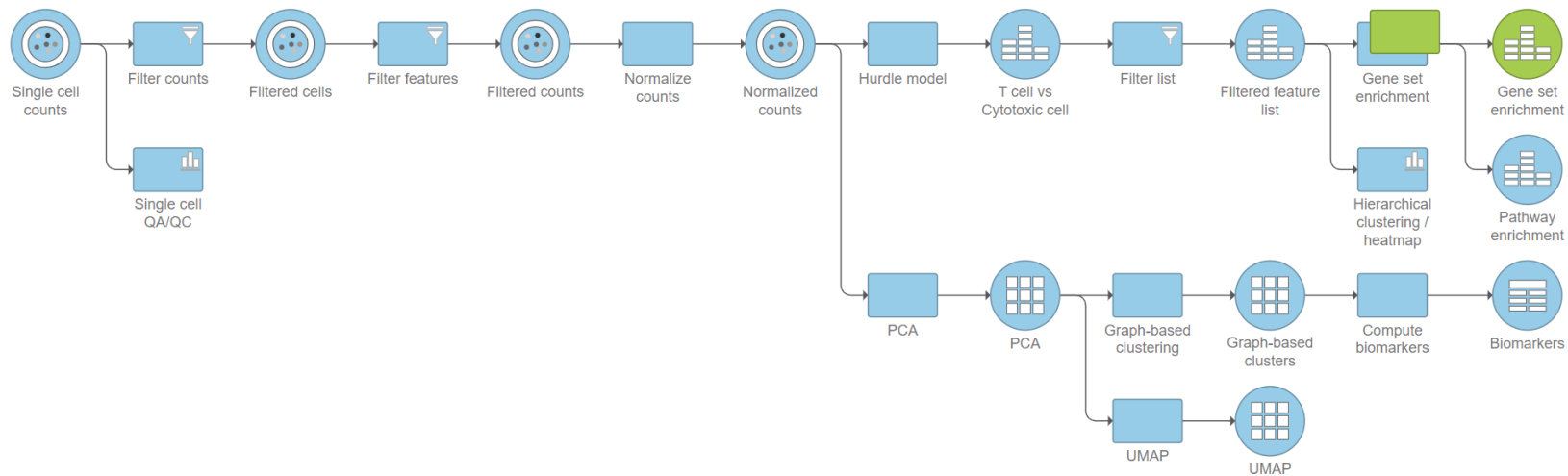


Biological Interpretation

- 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 ↑↓	
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	☰ ☒

Pipeline Overview



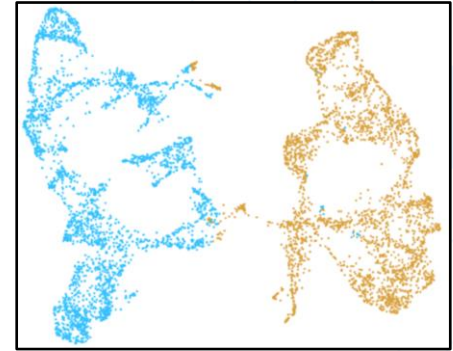
Appendix – Batch Removal



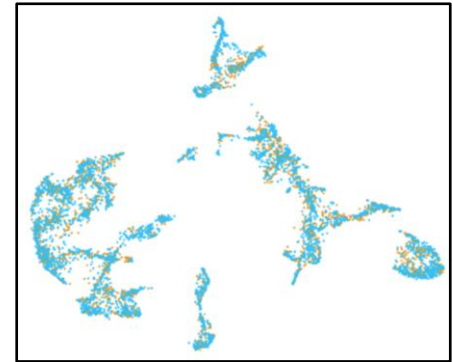
Purpose

- When a project contains multiple libraries, the data might contain variabilities due to technical differences (e.g. sequencing machine, library prep kit etc.) in addition to biological differences (like treatment, genotype etc.)
- Batch removal is essential to remove the noise and discover biological variations.

Batch effect



Batch effect correction



Assign batch to each sample

- Go to the **Metadata** tab
- Click Manage
- **Add new attribute** and enter a name
- Add categories



Analyses **Metadata** Log Project settings Notebook Data viewer Attachments

▼ Sample attributes

- Manage
- Assign values
- Assign values from file
- Add system-wide attribute

▼ Cell attributes

- Manage

Sample name	Attributes
	# Cells
1 Mouse_Brain_Anterior	2823
2 Mouse_Brain_Posterior	3289

[Show data files](#) [Download](#)

Sample Attributes

No sample attributes have been added to the project.

[Add new attribute](#) [Add system-wide attribute](#) [Back to metadata tab](#)

Add new attribute

Name

Attribute type Categorical Numeric

Visibility Project-specific System-wide

Only modifiable by some users

[Add](#) [Cancel](#)

batch

1

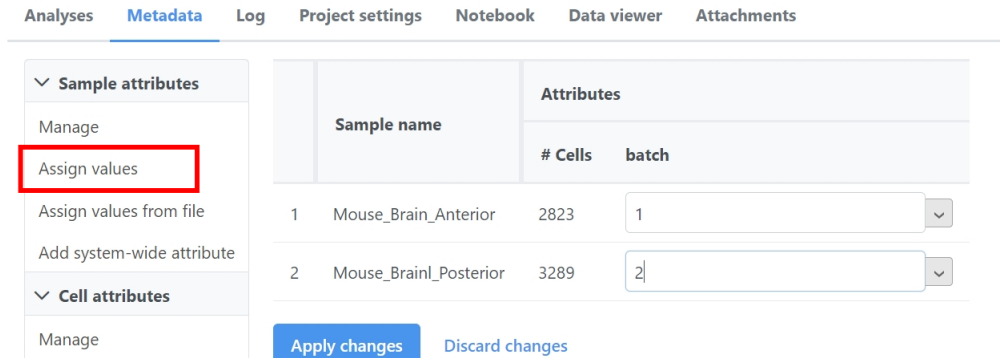
New category

2



Assign batch to each sample

- Back to **Metadata** tab
- Click **Assign values** and choose the batch of each sample
- **Apply changes**



Analyses **Metadata** Log Project settings Notebook Data viewer Attachments

▼ Sample attributes

- Manage
- Assign values**
- Assign values from file
- Add system-wide attribute

▼ Cell attributes

- Manage

	Sample name	Attributes	
		# Cells	batch
1	Mouse_Brain_Anterior	2823	1
2	Mouse_BrainI_Posterior	3289	2

Apply changes Discard changes



Performing Batch Removal

- Click the **Normalized counts** data node
- Click **Seurat3 integration** in the **Batch removal** section
- Select the attribute name for integration
- Click **Finish**
- A new data node will be created

Select a factor for integration

(--factor)

batch

Advanced options

Option set

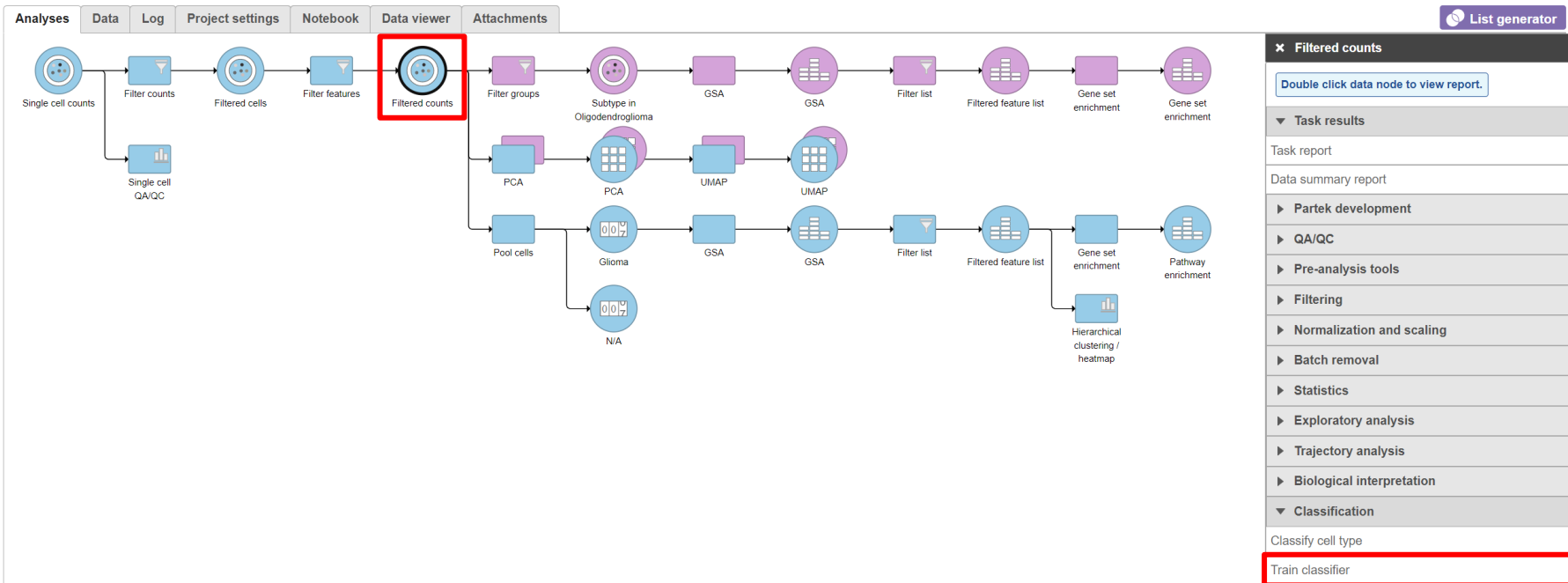
-- Default --

[Configure](#)



Appendix – Garnett Classifier


Train Classifier




Train Classifier



Marker file

Choose marker from  Local files

Marker file  Partek Flow Server URL

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

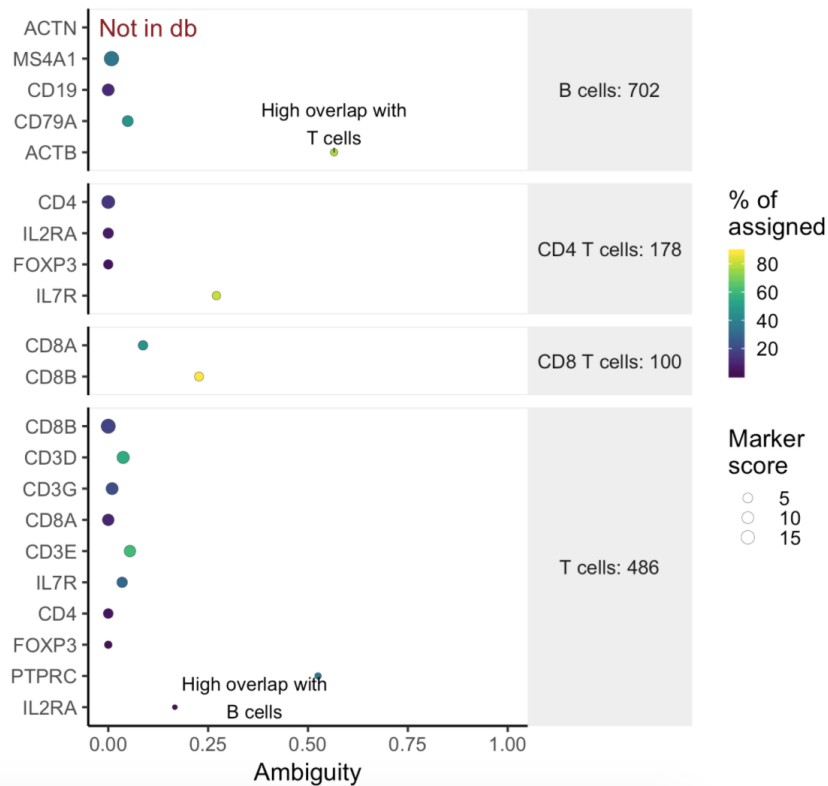
```
Glioma_cell_type_classifier.txt - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明
>Microglia
expressed: CD14

>Oligodendrocytes
expressed: MAG

>Glioma
expressed: BCAN, GPM6A
```

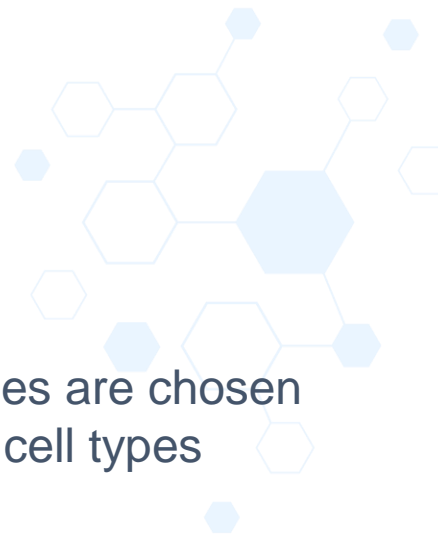
There has to be a space character after the colon and that there has to be a space character after the comma.

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

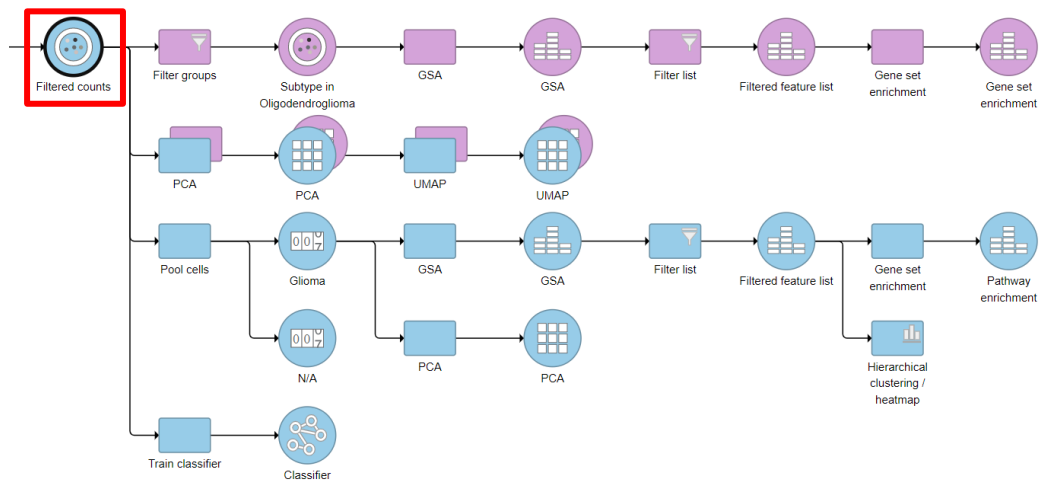
Train Classifier Results



- 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



× Filtered counts

Double click data node to view report.

▼ Task results

Task report

Data summary report

► Partek development

► QA/QC

► Pre-analysis tools

► Filtering

► Normalization and scaling

► Batch removal

► Statistics

► Exploratory analysis

► Trajectory analysis

► Biological interpretation

▼ Classification


Classify cell type

Train classifier

Classify Cell Type – Project classifiers


Choose classifier from 

Garnett classifier

Project classifiers 

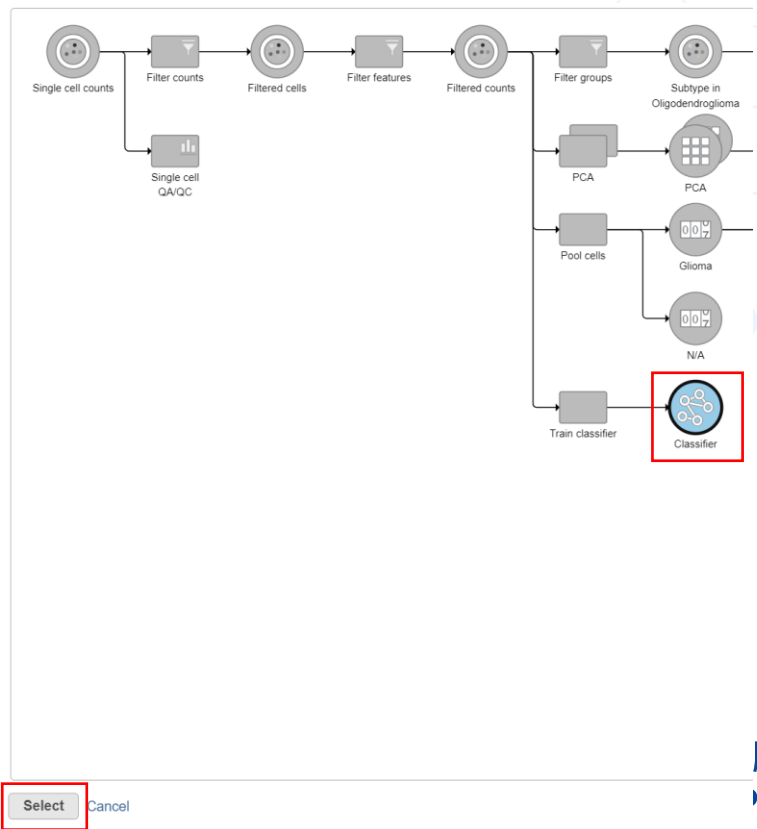
Select data node

Clear selection


 Classifier [Train classifier - 0.2.14]

Back


Finish



Classify Cell Type – Managed classifiers

Choose classifier from 

Garnett classifier

Garnett classifier 

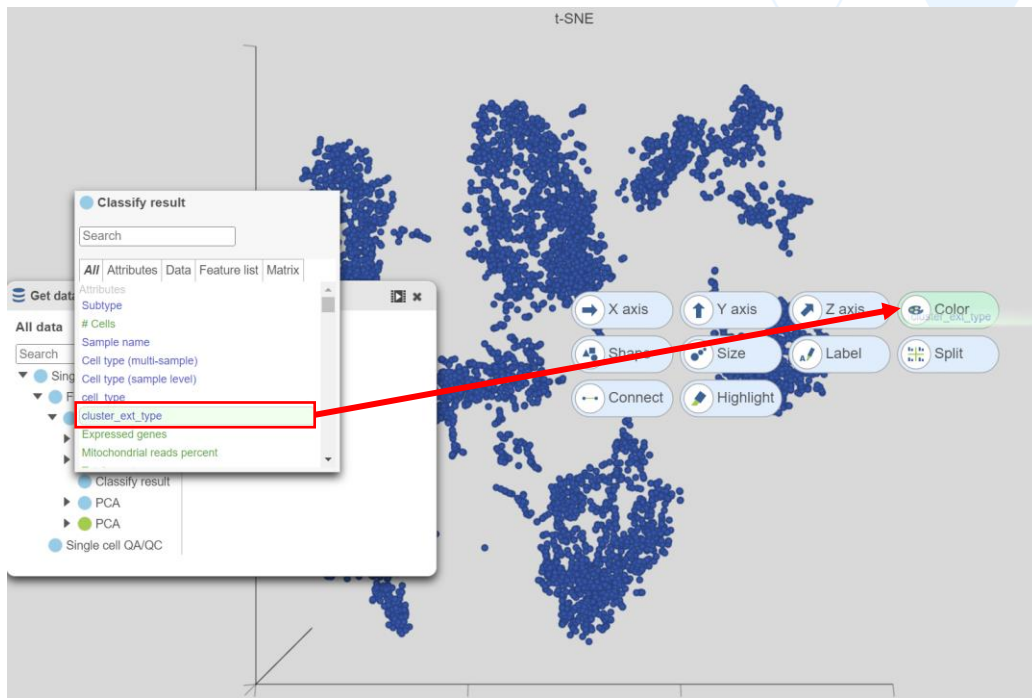
Species

Name

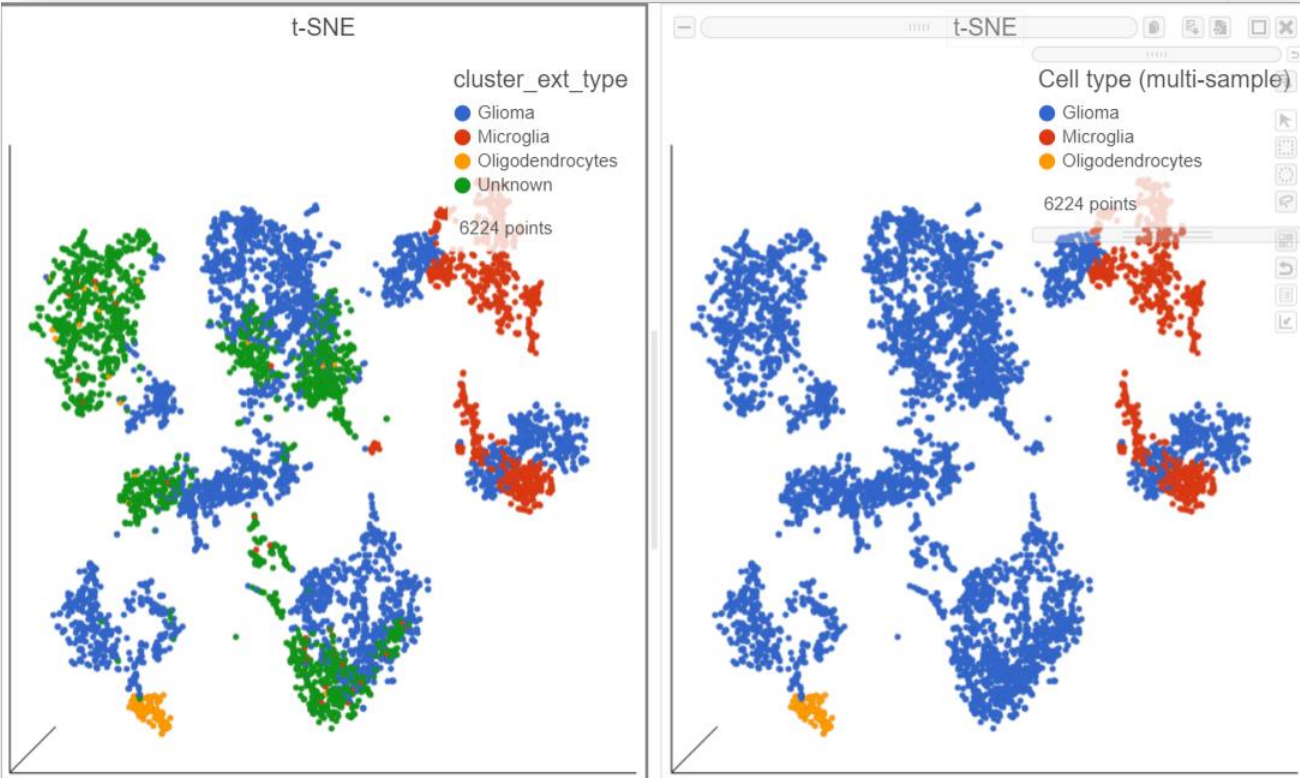
- Glioma_Demo_classifier**
- humanAdrenal
- humanCerebellum
- humanCerebrum
- humanEye
- humanHeart
- humanIntestine
- humanLiver
- humanMuscle
- humanPancreas
- humanPlacenta
- humanSpleen
- humanStomach
- humanThymus
- mouseBrain
- New classifier file...

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



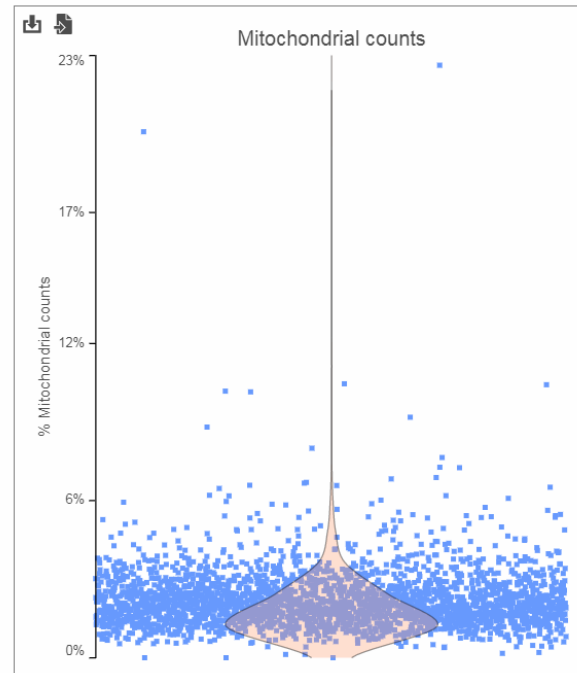
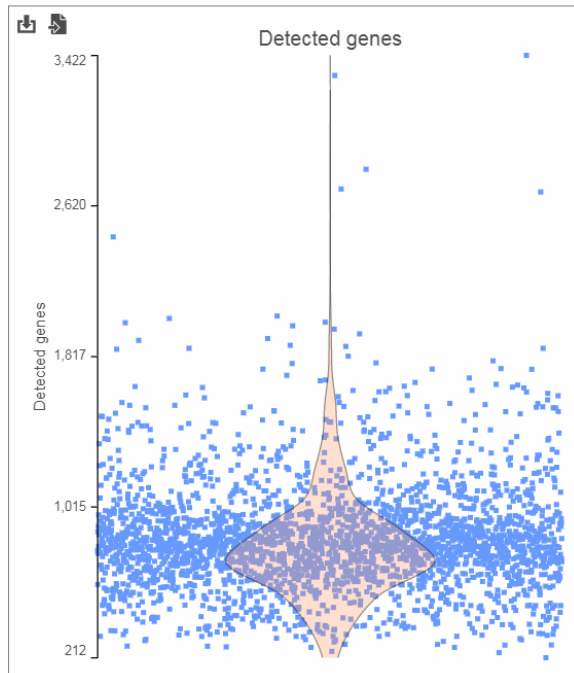
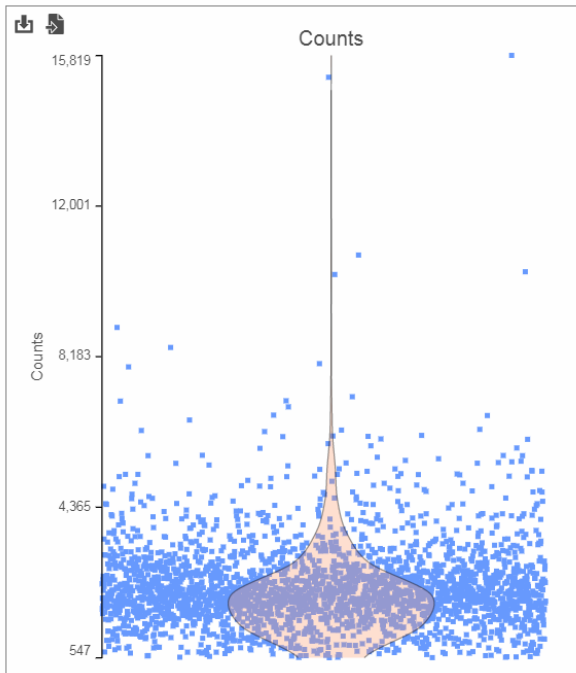
Plot Interpretation



Single cell QA/QC report - Violin Plot

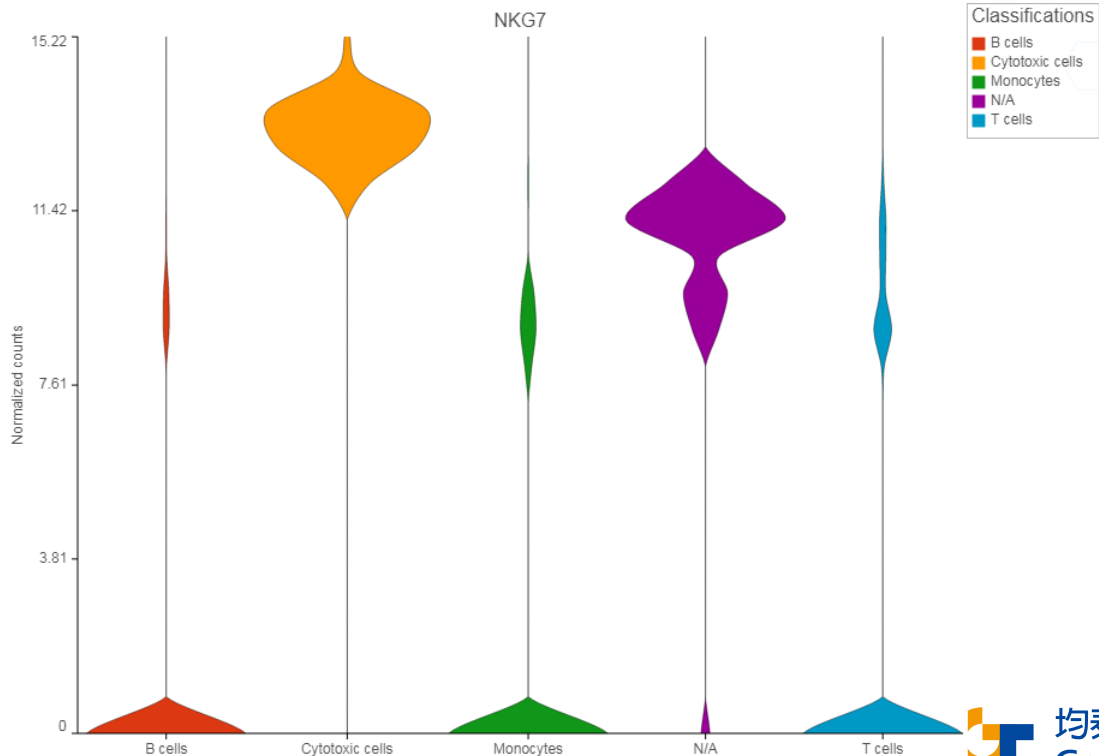
由左至右分別代表細胞中的read數量、基因數量以及Mitochondria gene表達量
x軸沒有意義，目的是為了避免有兩個以上的cells有相同的count重疊看不出來；y軸代表total count；每個點代表一個細胞
Violin plot 越寬代表密度越大，可以由這張圖明顯看到cell集中於哪個數量區域，並進一步留下較有生物意義的細胞

■ Selected cells ■ Excluded cells

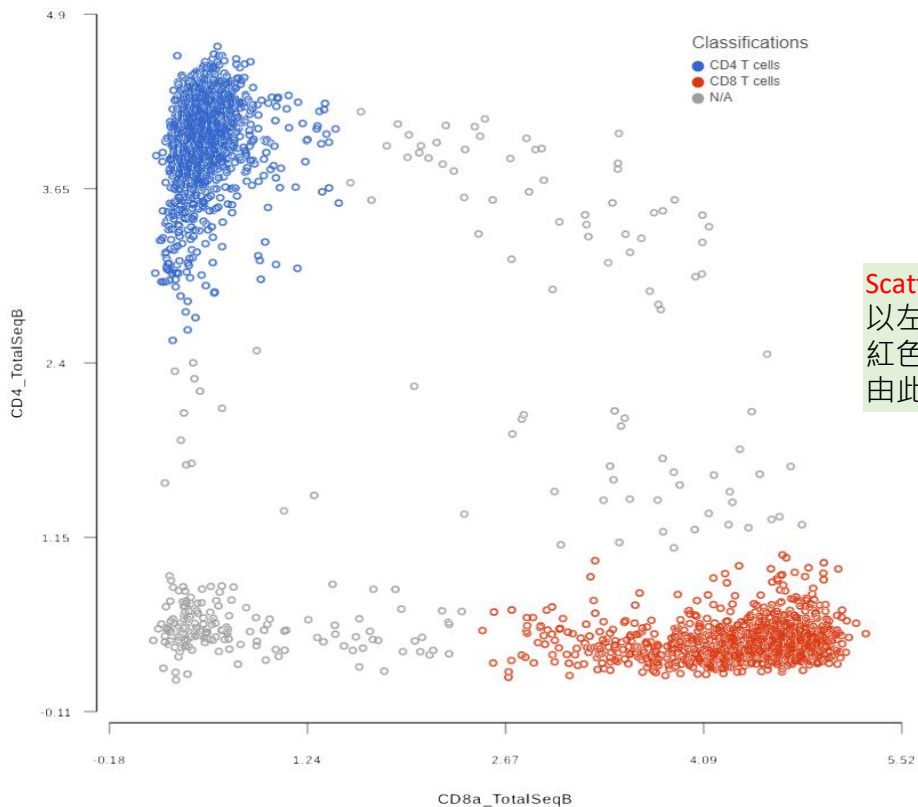


Feature Plot

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



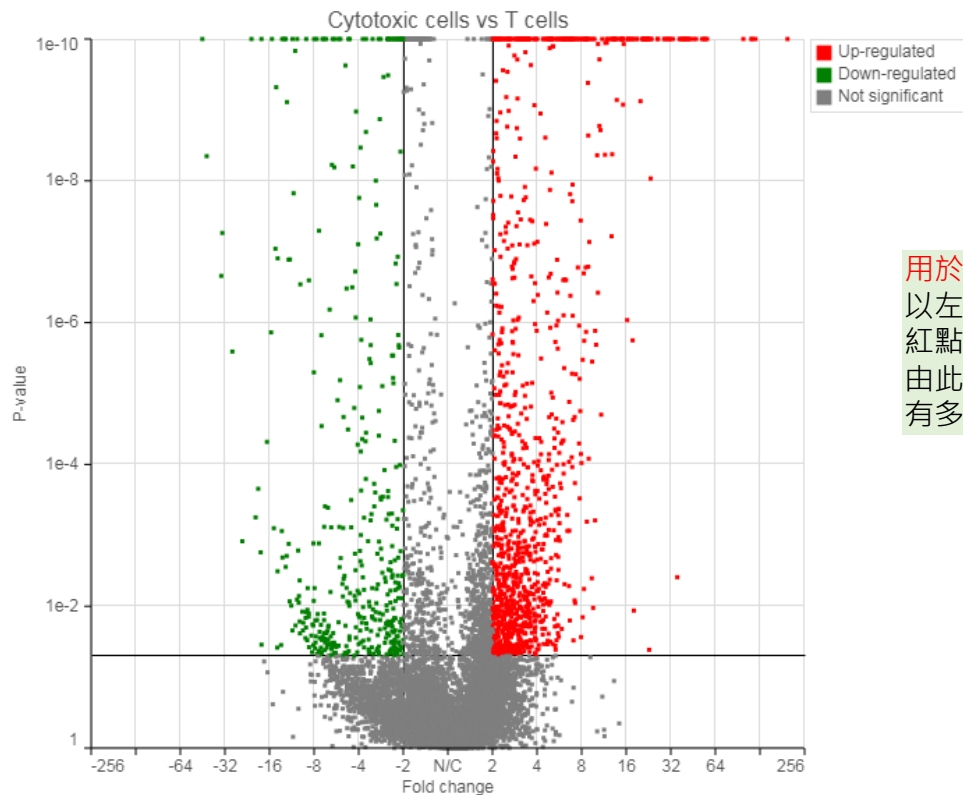
Scatter Plot



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Biomarkers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

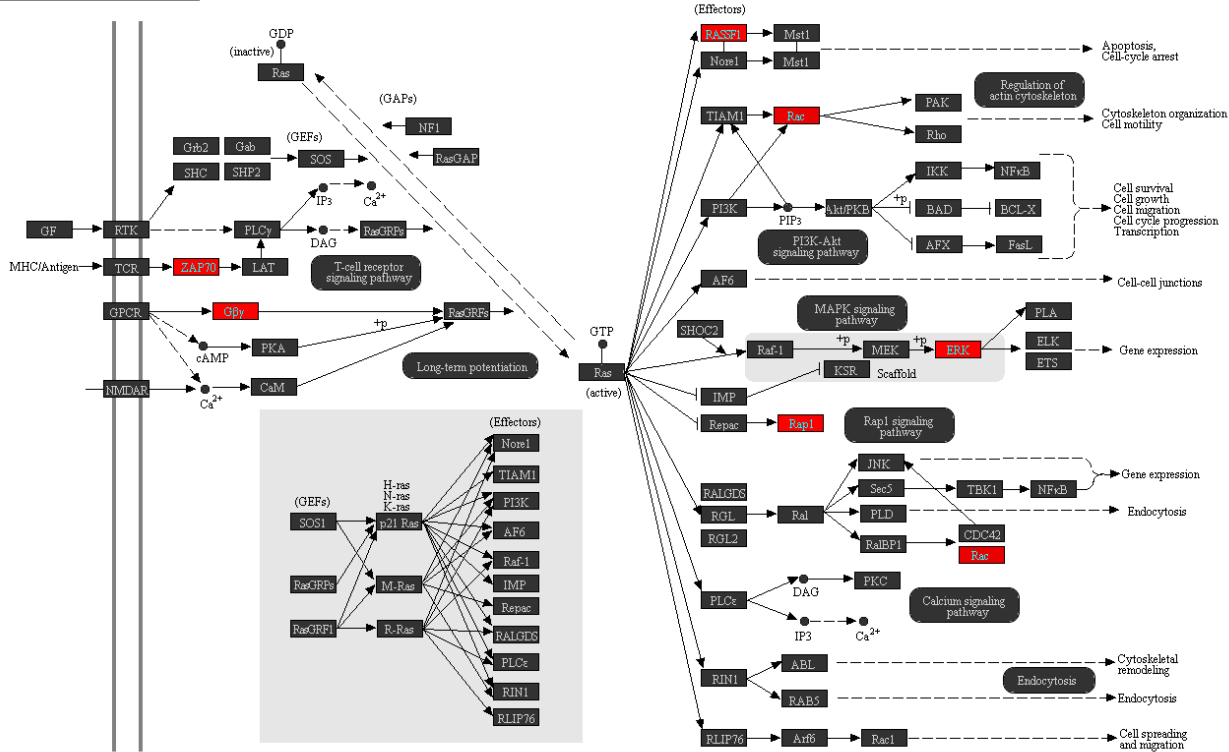
以左圖說明，X軸為Fold change，Y軸為P-value；

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

KEGG Pathway result

RAS SIGNALING PATHWAY



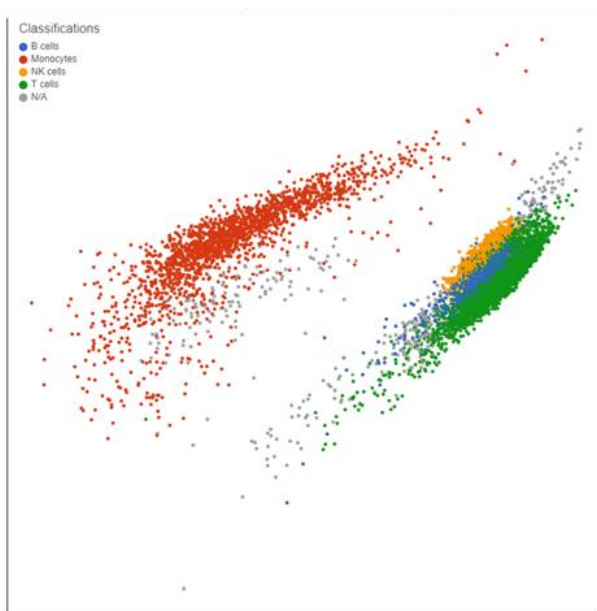
挑出感興趣的Gene list，並分析這些基因是否影響特定pathway

Dimensionality Reduction: PCA, t-SNE, UMAP

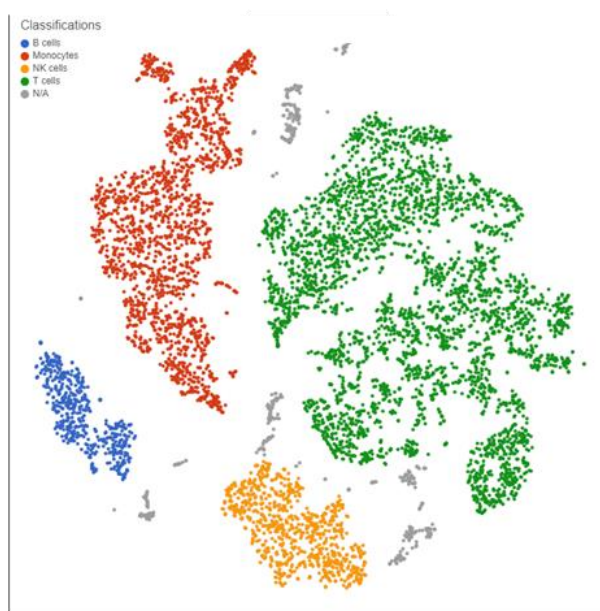
細胞分群後的圖表呈現，因每個細胞皆有上千、萬個基因，相等於上千、萬個維度，必須透過降維才能比較各個細胞間不同基因表達量的相關性

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

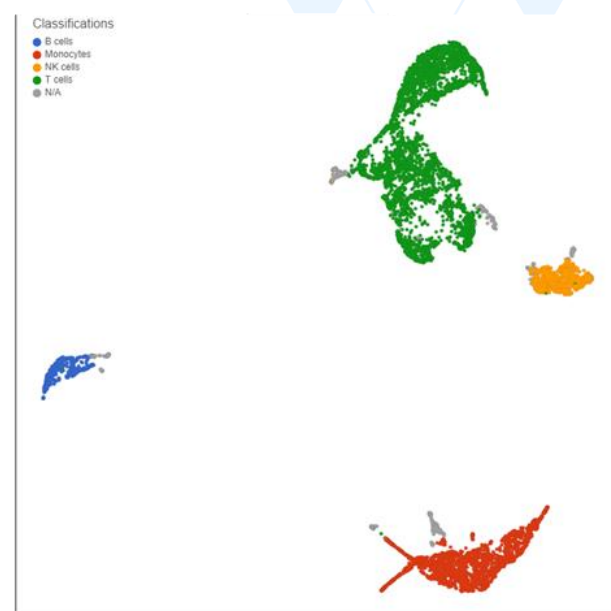
PCA



t-SNE



UMAP



View each in 2D or 3D

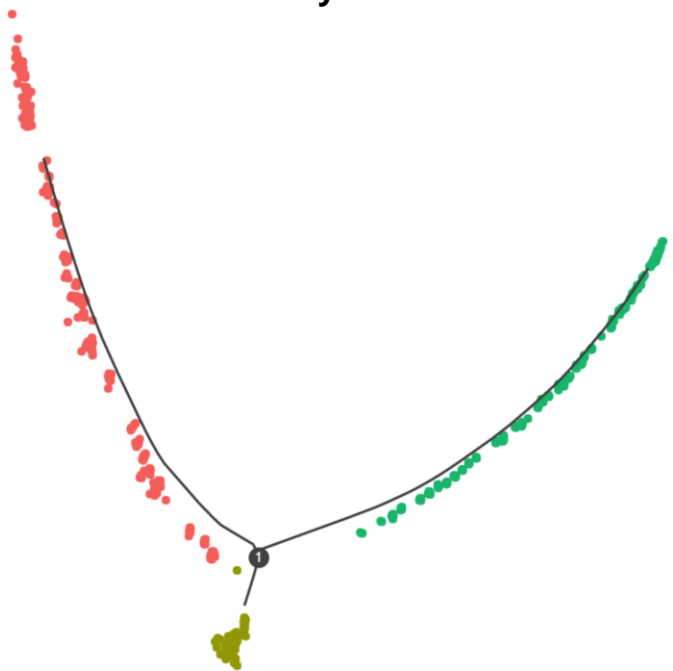
Run Trajectory Analysis with Monocle

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

