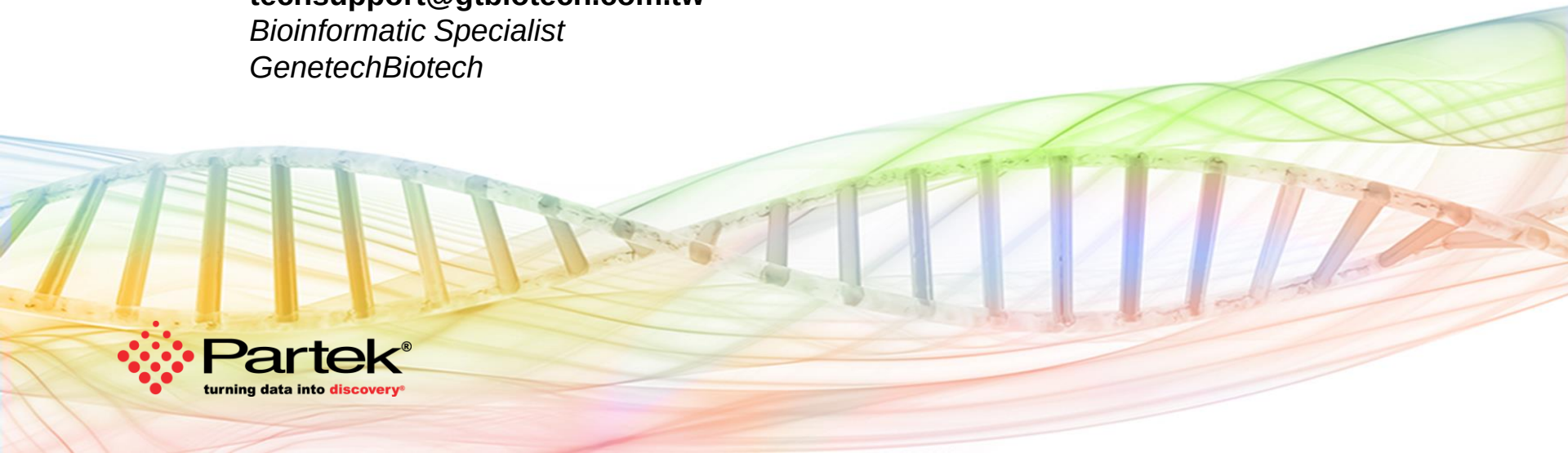
From Raw Data to Pathways: Easy Genomics Analysis with Partek Flow

Eric Chen
techsupport@gtbiotech.com.tw
Bioinformatic Specialist
GenetechBiotech



Partek®
turning data into discovery®

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

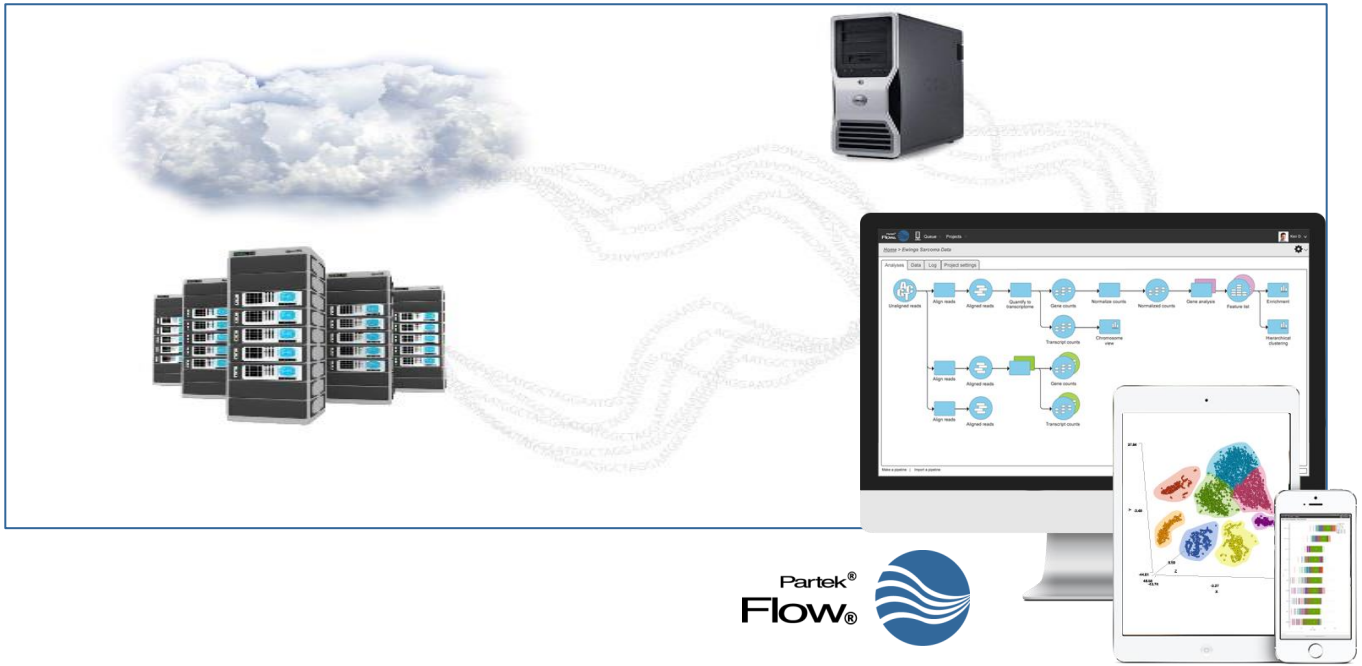
40

countries

Copyright © 2019 Partek Incorporated. All rights reserved.

Partek®

Partek Flow – Web Based Application

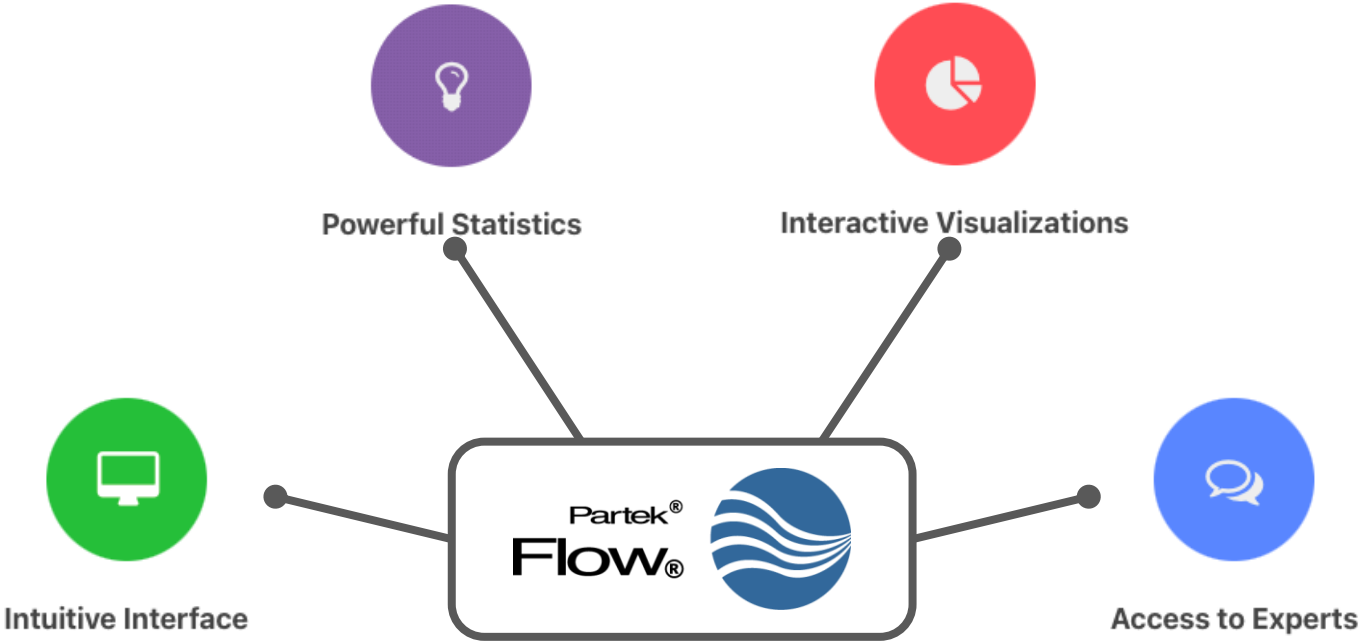


- Flow **server**
  - Desktop
  - Cluster
  - Cloud
- Flow **user**
  - Internet

Copyright © 2019 Partek Incorporated. All rights reserved.

Comprehensive Bioinformatics Software

User Friendly and Powerful Data Analysis Tool



Copyright © 2019 Partek Incorporated. All rights reserved.

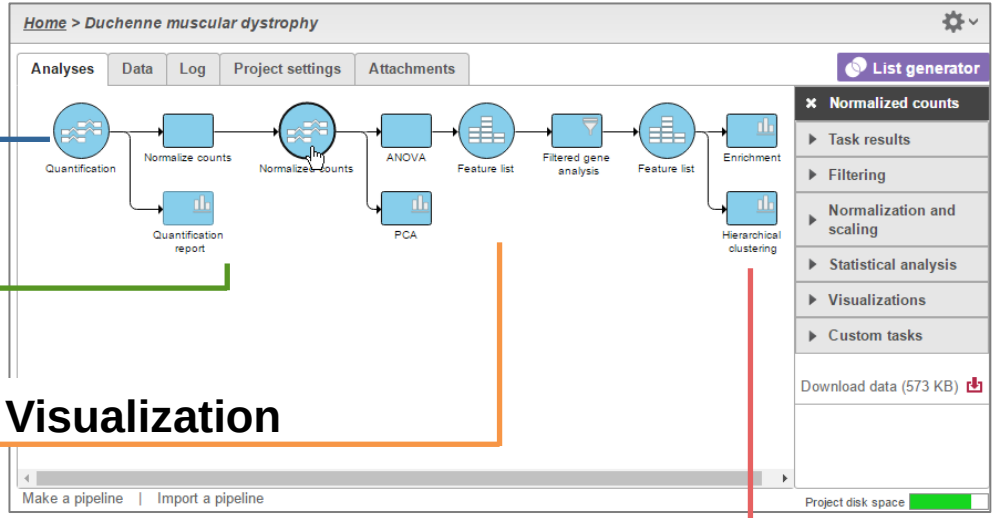
Visual Analysis Process

1 Import Data

2 QA/QC

3 Powerful Statistics & Visualization

4 Biological Interpretation



Copyright © 2019 Partek Incorporated. All rights reserved.

Publicly Available Statistical Algorithms and Tools

Alignment
Bowtie
BWA
Isaac
TopHat
TMAP
Bowtie
GSNAP
STAR
HISAT

Differential analysis
Limma
DESeq2
Poisson
Negative binomial
Non-parametric ANOVA

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
Heat map
Violin plot
Histograms
Chromosome view
2D & 3D Scatter Plot
Pie chart
UMAP

Variant calling
Samtools
LoFreq
CNVkit
FreeBayes
Strelka
GATK

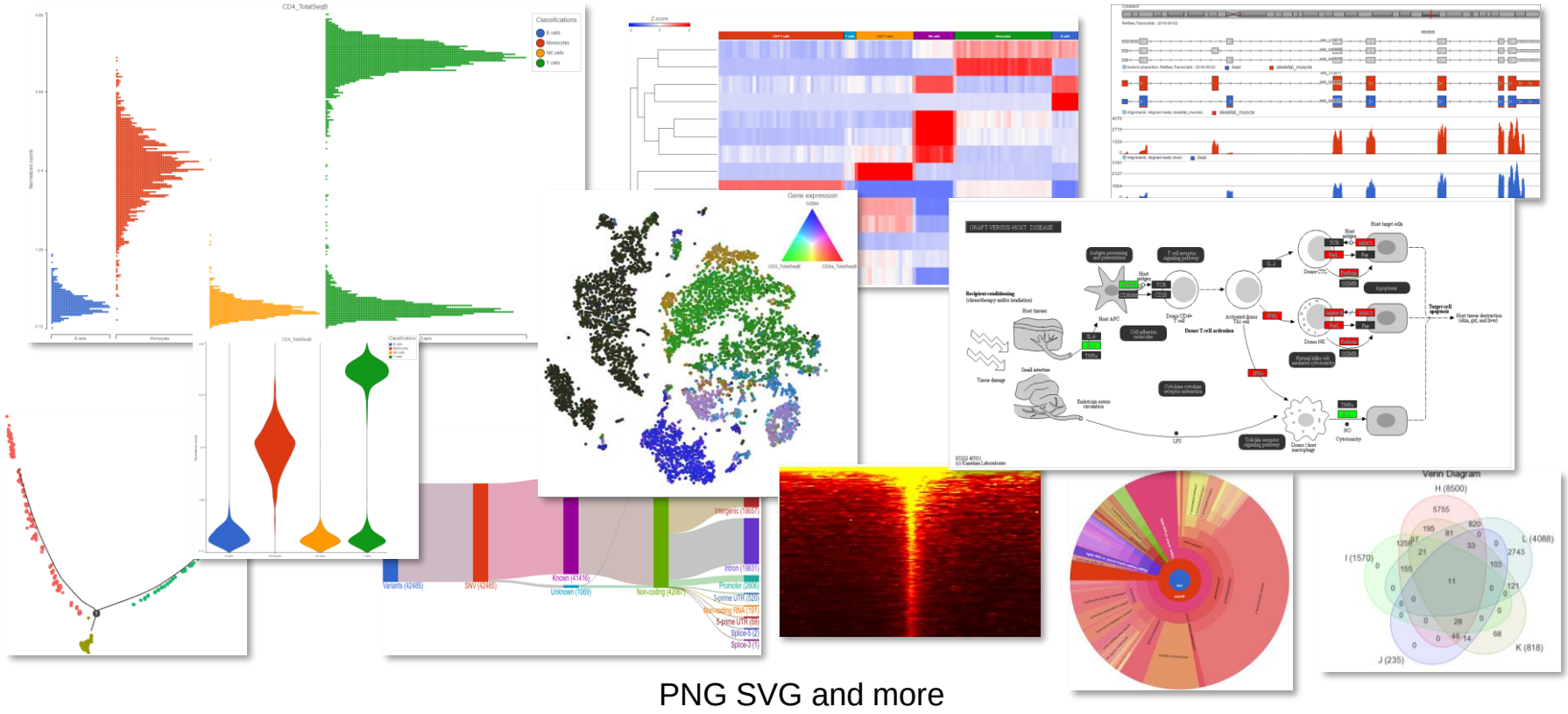
Variant annotation
SnpEff
dbSNP
VEP
Custom databases

Peak calling
MACS2
TSS plot
Motif detection

Quantification
Partek E/M
HTSeq
Cufflinks

Copyright © 2019 Partek Incorporated. All rights reserved.

Compelling and Publishable Visualizations

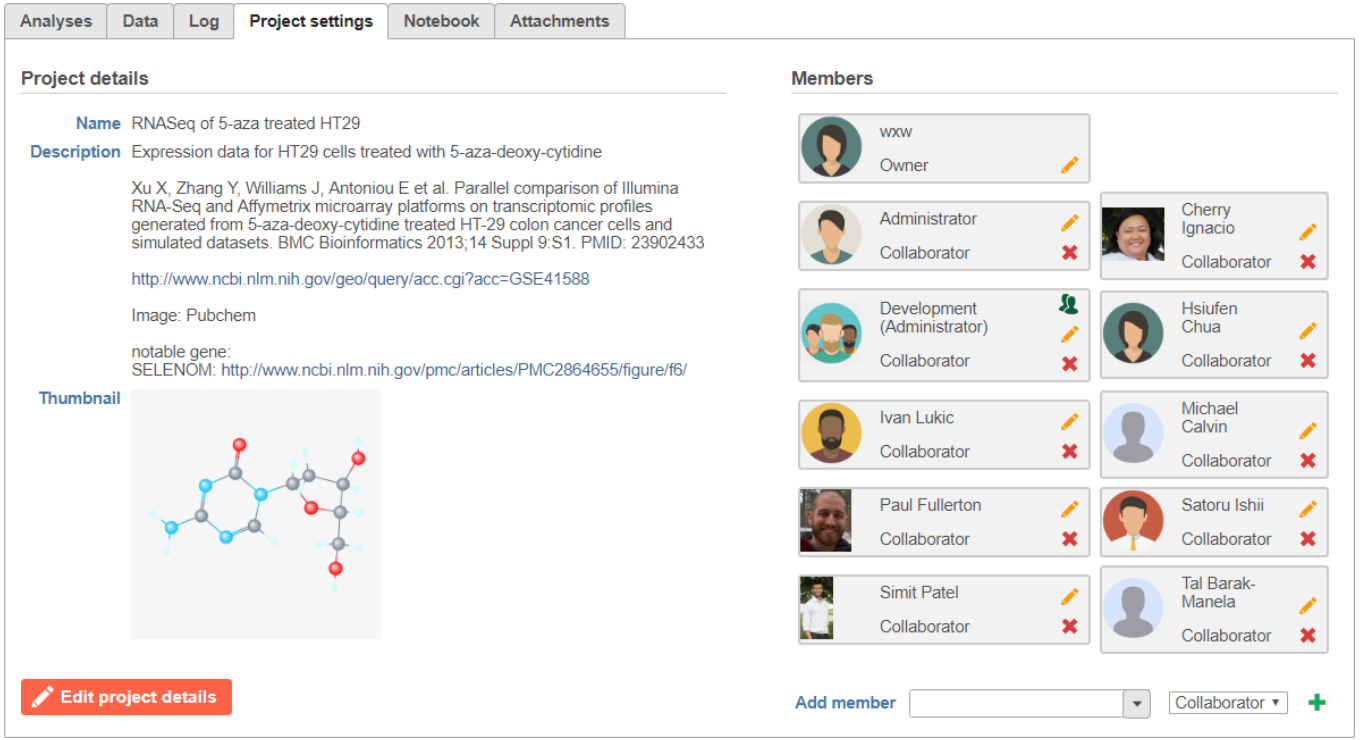


PNG SVG and more

Copyright © 2019 Partek Incorporated. All rights reserved.

Partek®

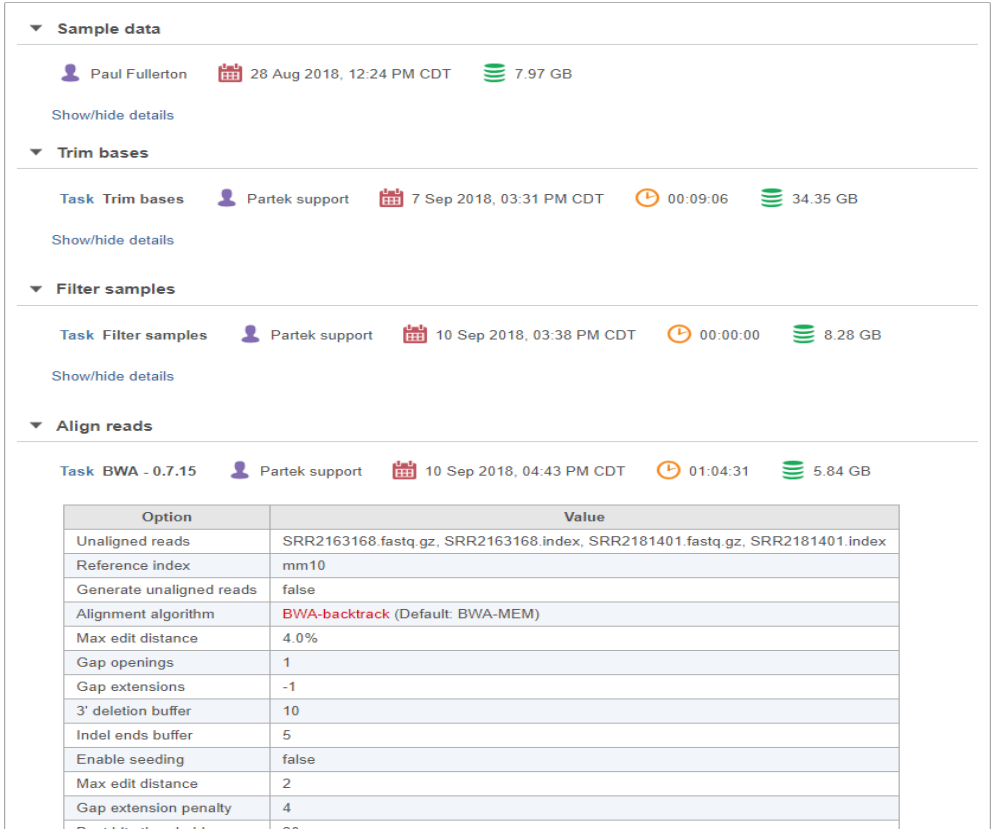
Collaborate and Share



Copyright © 2019 Partek Incorporated. All rights reserved.

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

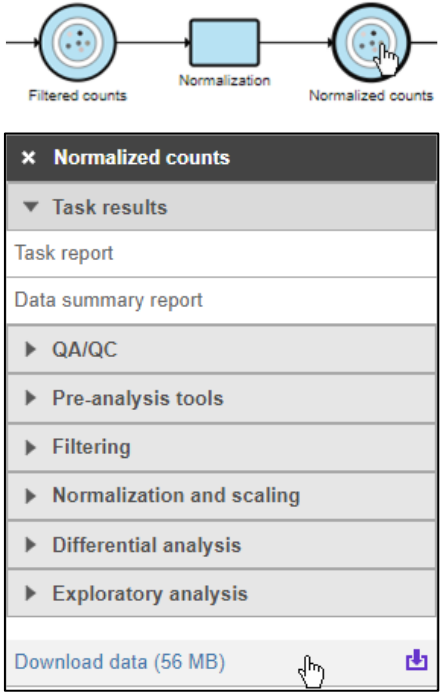
Task **BWA - 0.7.15** | Partek support | 10 Sep 2018, 04:43 PM CDT | 01:04:31 | 5.84 GB

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

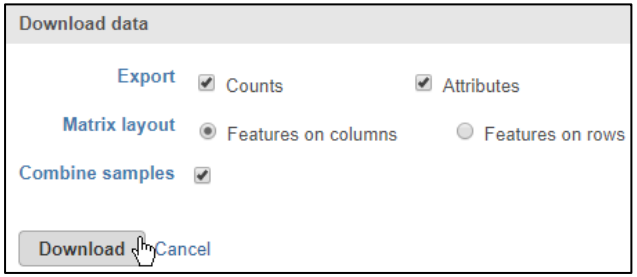
Copyright © 2019 Partek Incorporated. All rights reserved.

Export Data

Choose Any Data

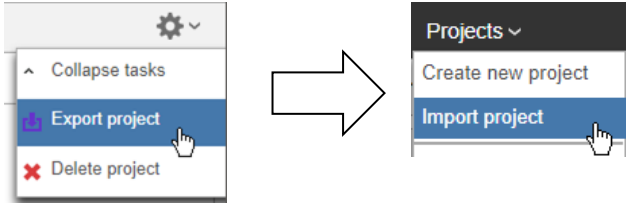


Download in Industry Standard Formats



FASTQ, BAM, TXT, and more

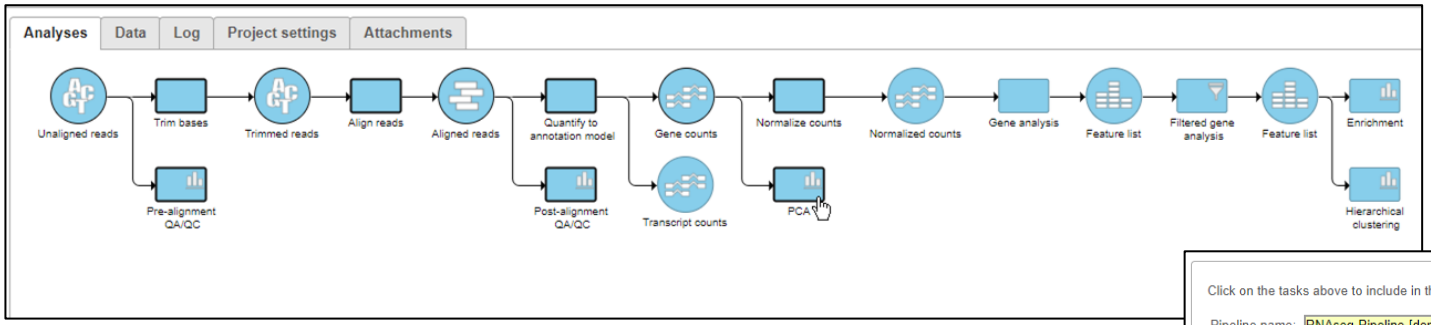
Export and Import Analysis Projects

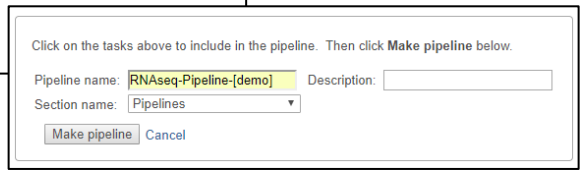


Copyright © 2019 Partek Incorporated. All rights reserved.

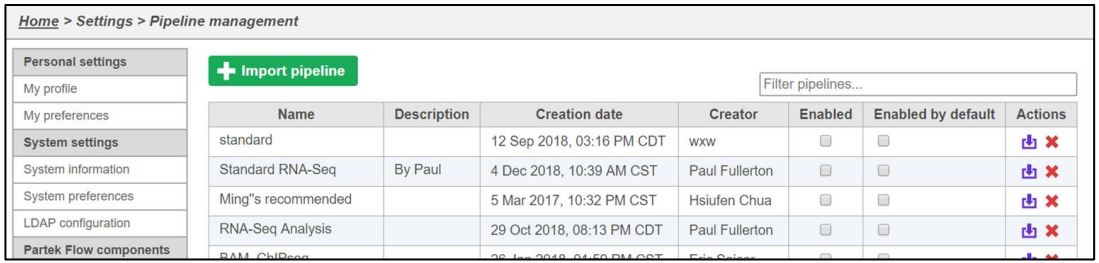
Build, Reuse, and Share Analysis Pipelines

Build Analysis Pipelines





Save, Share, and Manage



Copyright © 2019 Partek Incorporated. All rights reserved.

Compatible with All Major Genomics Formats and Assays

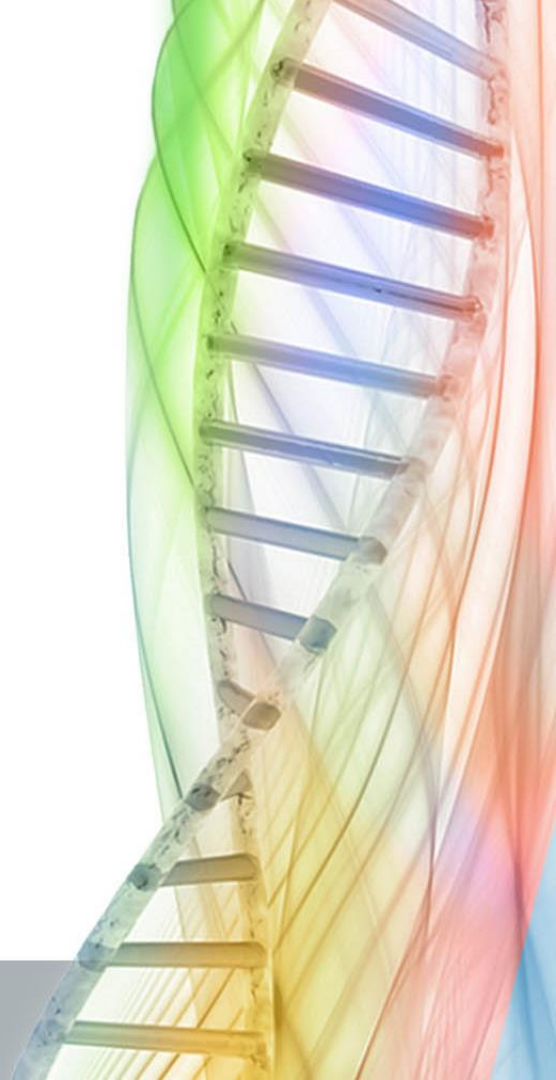
RNA | Noncoding RNA | SC RNA | DNA | ChIP | ATAC | Metagenomics



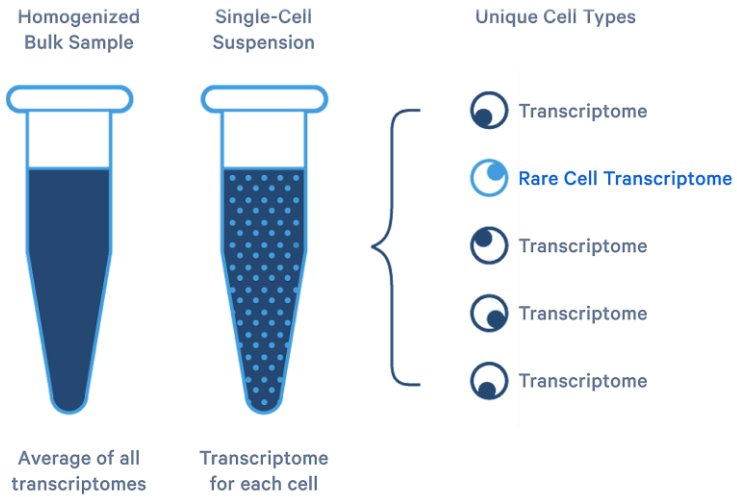
Microarray | Next Generation Sequencing | qPCR

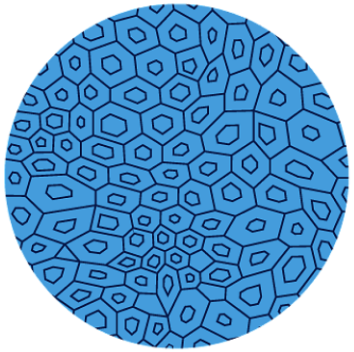
Copyright © 2019 Partek Incorporated. All rights reserved.

Single Cell Analysis

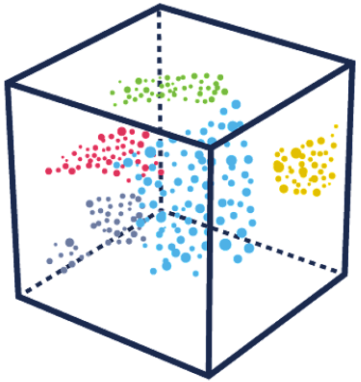


Introduction of Single-cell Analysis





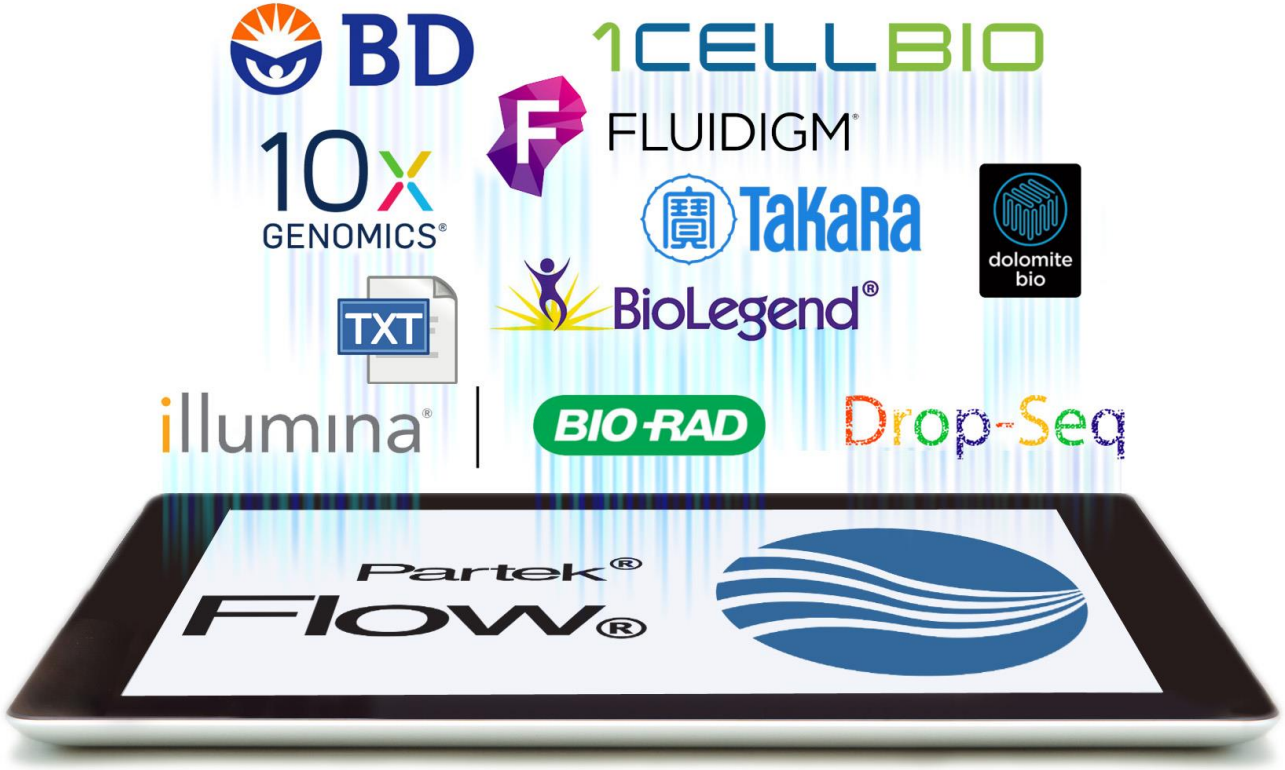
Tissue Specimen with a spatial relationship between cells.



Relationship between cells by similarity of gene expression.

https://www.10xgenomics.com/single-cell-technology

Supports All Major Single Cell Platforms



Copyright © 2019 Partek Incorporated. All rights reserved.

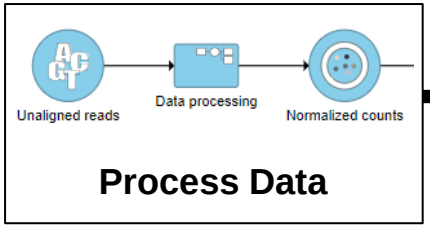
Support for Wide Variety of Single Cell Technologies

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

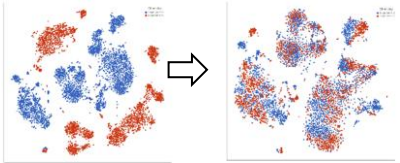
Copyright © 2019 Partek Incorporated. All rights reserved.

Data Processing and Analysis, All in One Place

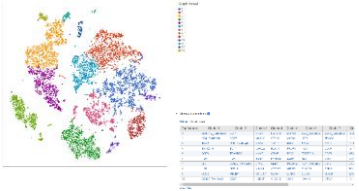
FASTQ, BAM, or Counts



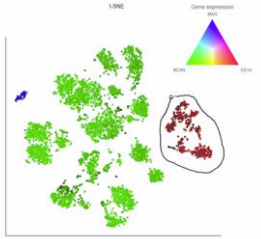
Batch correction



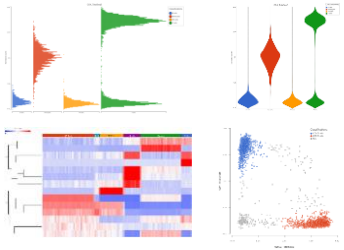
Clustering



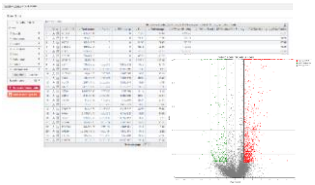
Interactive Visual Analysis



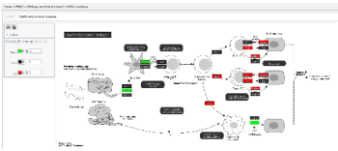
Data Visualization



Differential Analysis



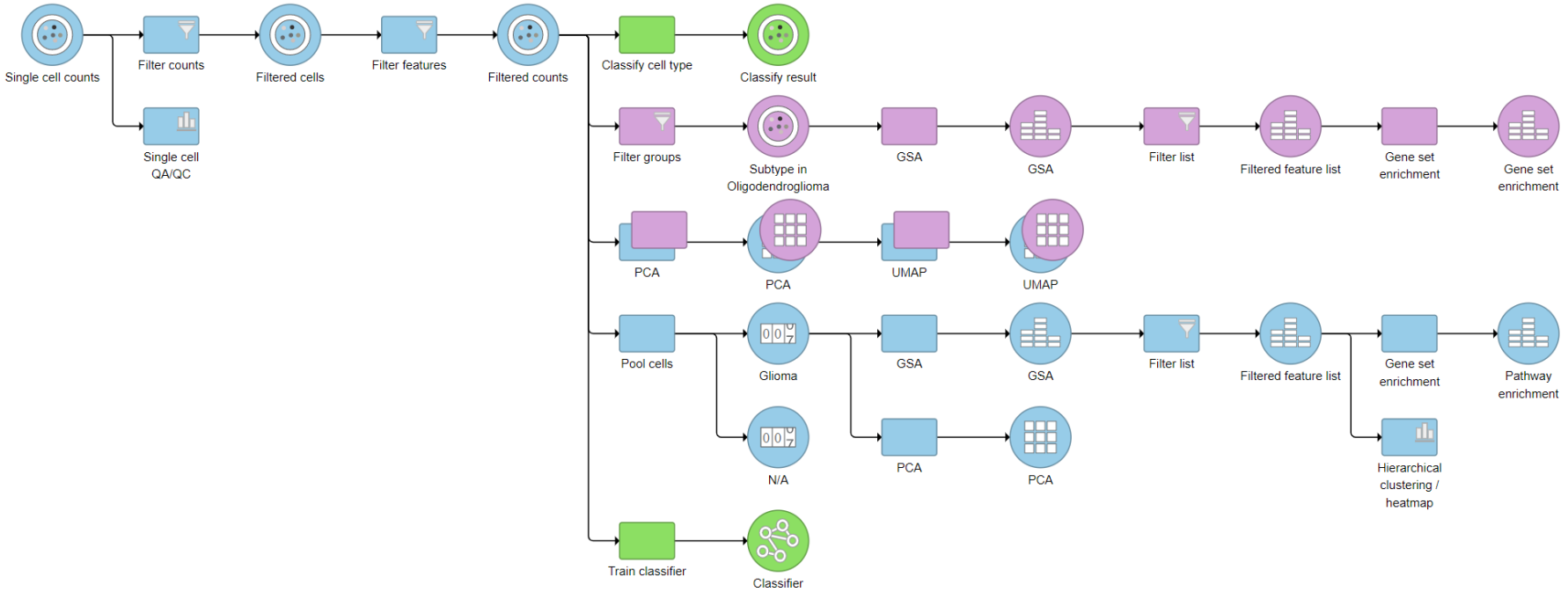
Pathway Analysis



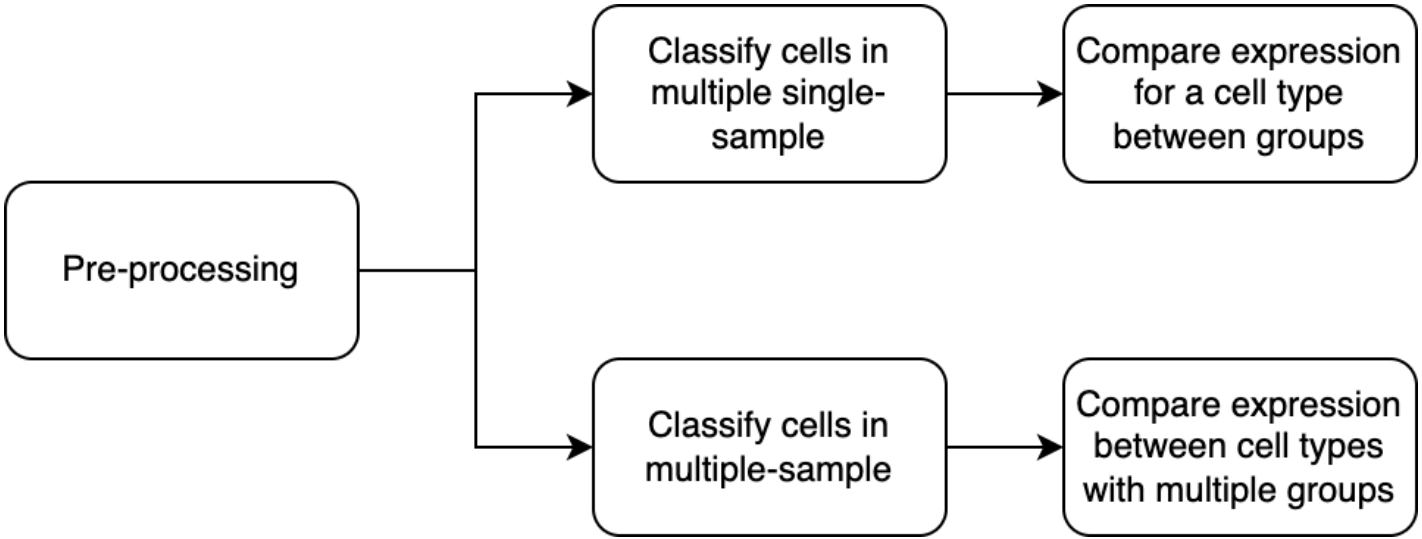
Copyright © 2019 Partek Incorporated. All rights reserved.

Partek®

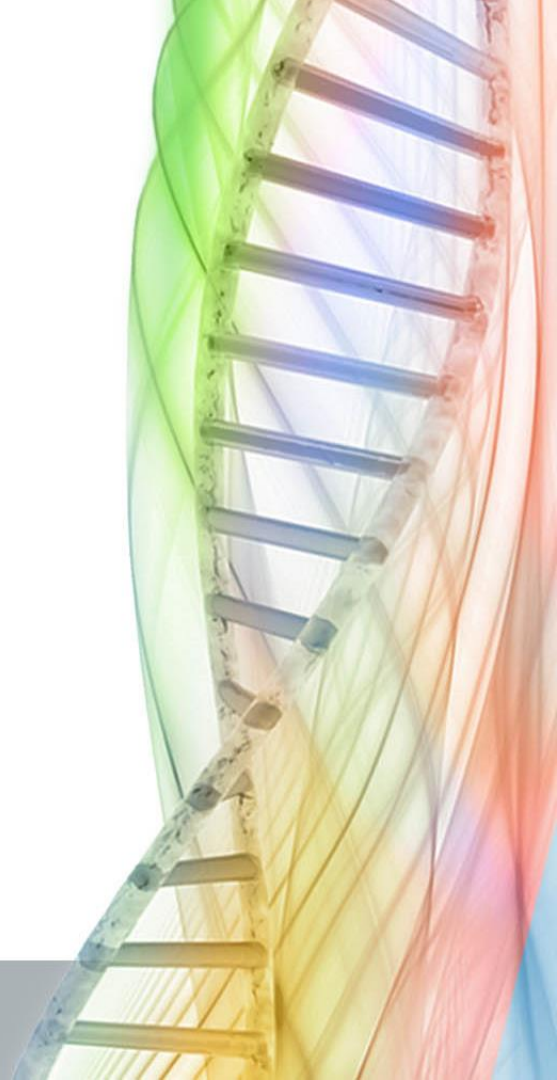
Single Cell RNA-Seq Analysis (Multiple Samples) pipeline



Overview of the whole Analysis

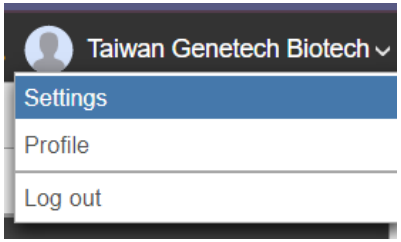


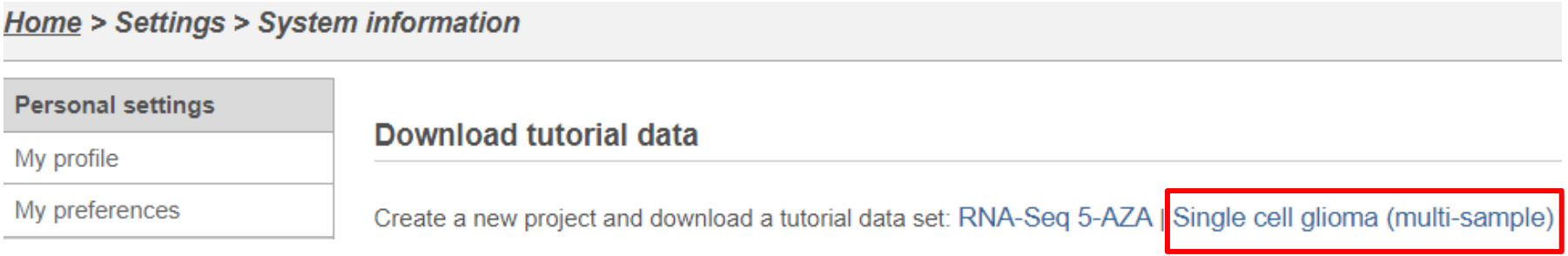
Live Demo



Creating a new project and importing the tutorial data set

1. Log in to **Partek Flow**
   a. https://taiwangtbiotech.partek.com/flow/login.xhtml
2. Click **Settings**

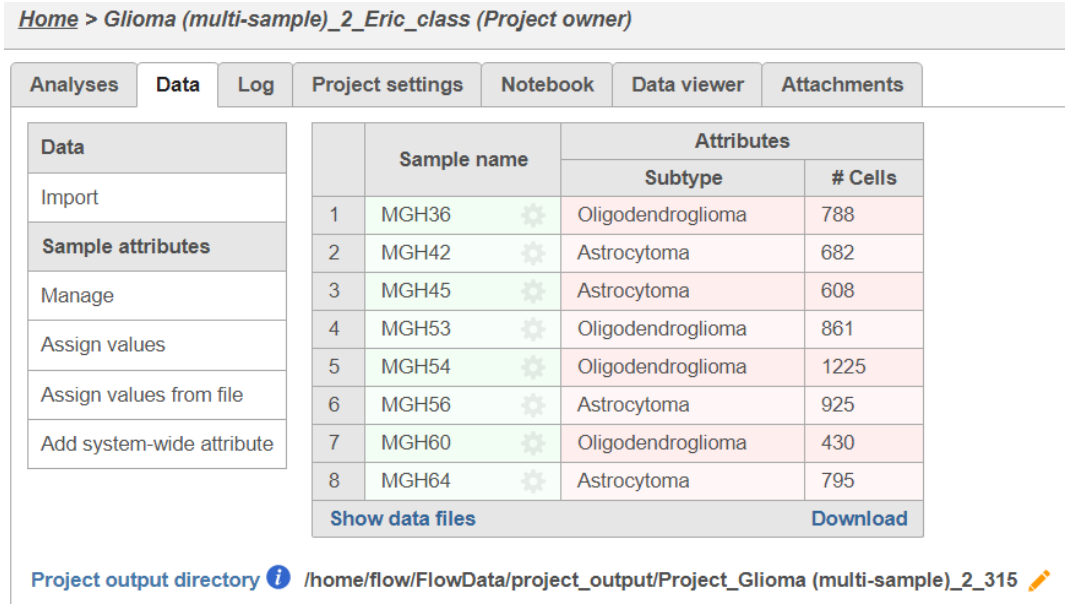




1. Click **Single cell glioma (multi-sample)**
2. The tutorial data set will be downloaded onto your Partek Flow server and a new project, **Glioma (multi-sample)**, will be created.

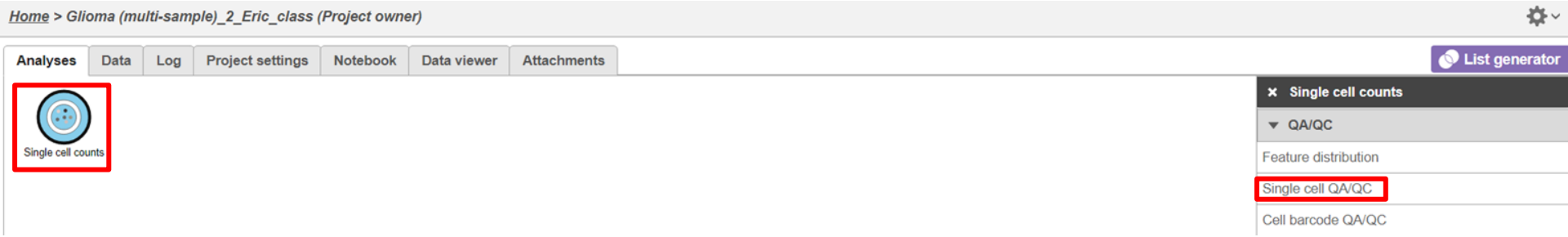
Sample data table

- Once the download completes, the sample table will appear in the Data tab, with one row per sample



Filtering cells in single cell RNA-Seq data

- Click **Analyses** to switch to the Analyses tab
- Click on the **Single cell counts** node
- Expand the **QA/QC** section of the task menu
- Click on **Single cell QA/QC**



Filtering cells in single cell RNA-Seq data

- Click the **Single cell QA/QC** node once it finishes running
- Click **Task report** on the task menu



Filtering cells in single cell RNA-Seq data

- Click on **Single cell counts** in Get Data on the left
- Click and drag the **Sample name** attribute onto the **Counts plot** and drop it onto the **X-axis**
- Repeat this for the **Detected genes** plot



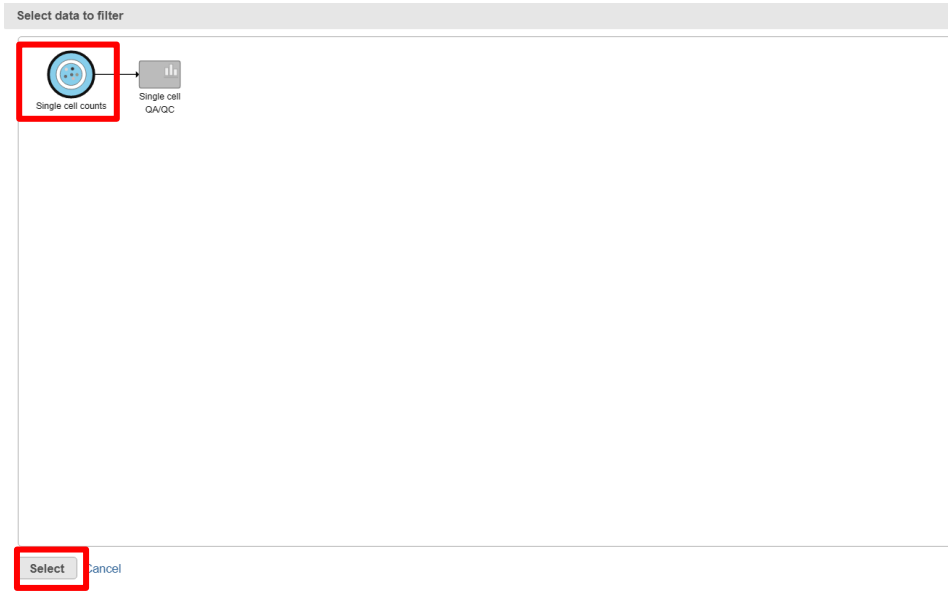
Filtering cells in single cell RNA-Seq data

- Set the Counts thresholds to **8000 and 20500**



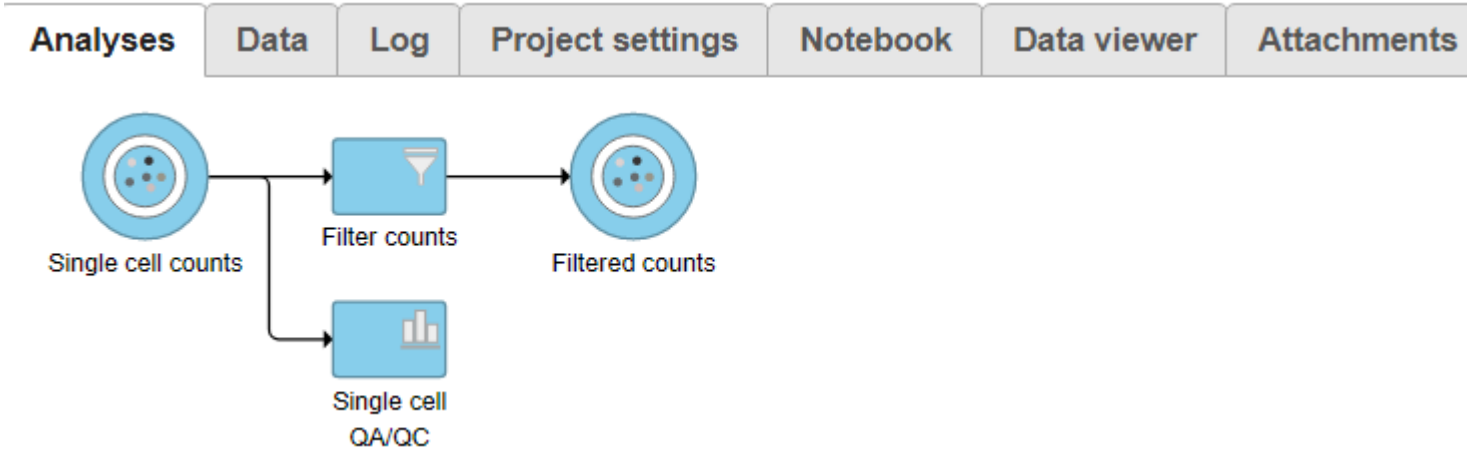
Filtering cells in single cell RNA-Seq data

- Click  in the Select & Filter window
- Click **Apply observation filter...**
- Click the **Single cell counts** data node in the pipeline preview
- Click **Select**



Filtering cells in single cell RNA-Seq data

- Click on the **Glioma (multi-sample)** project name at the top to go back to the Analyses tab
- Your browser may warn you that any unsaved changes to the data viewer session will be lost. Ignore this message and proceed to the Analyses tab



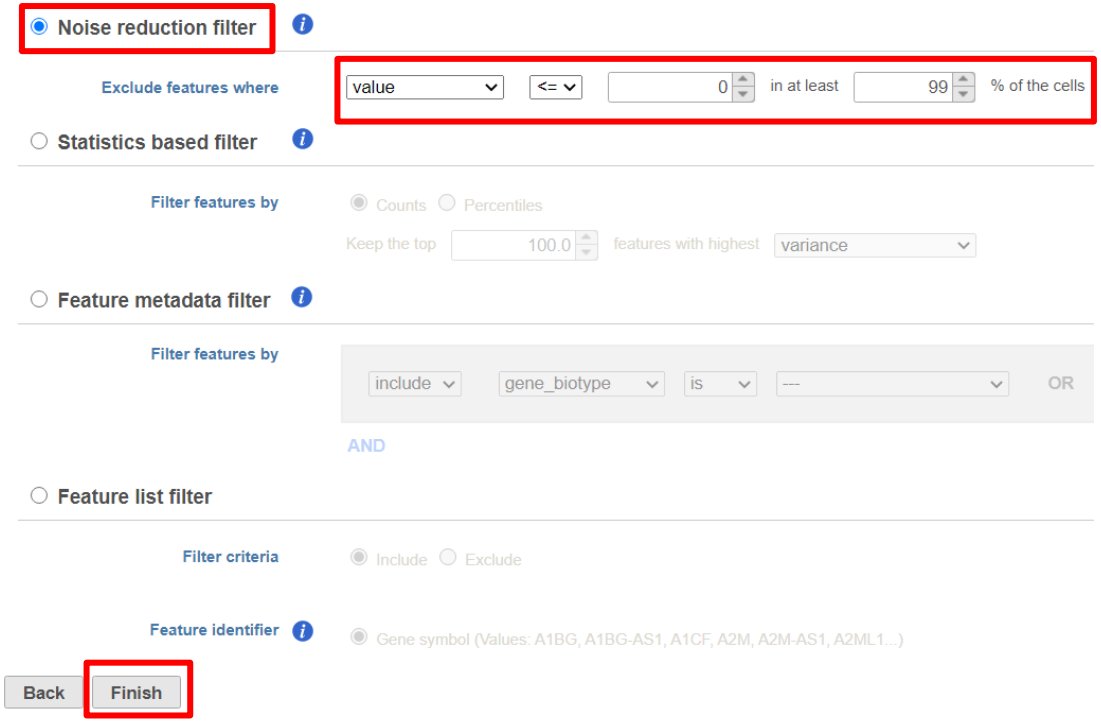
Filtering genes in single cell RNA-Seq data

- Click the **Filter counts** node produced by the Filter counts task
- Click **Filtering** in the task menu
- Click **Filter features**



Filtering genes in single cell RNA-Seq data

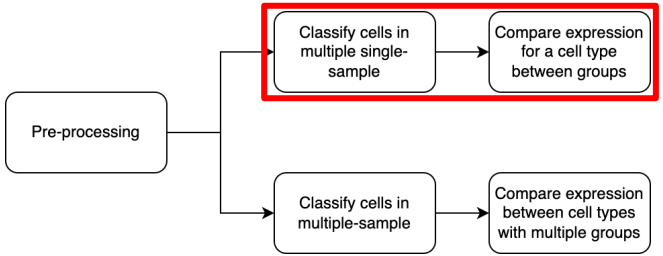
- Click the **Noise reduction filter** checkbox
- Set the Noise reduction filter to **Exclude features where value <= 0 in 99% of cells** using the drop-down menus and text boxes
- Click **Finish** to apply the filter



Normalizing single cell RNA-Seq data

- We are omitting normalization in this tutorial because the data has already been normalized.

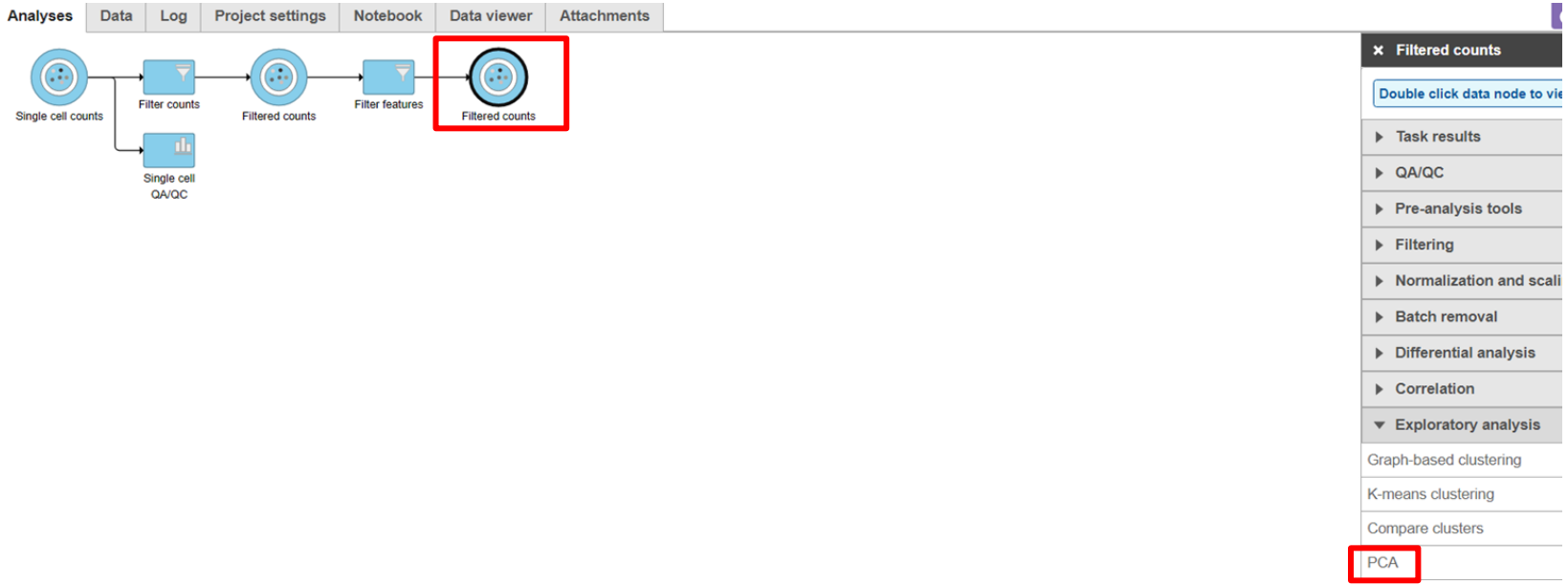
- The tutorial data set is taken from a published study and has already been normalized using TPM (Transcripts per million), which normalizes for the length of feature and total reads, and transformed as log2(TPM/10+1). This normalization and transformation scheme can be performed in Partek Flow, along with other commonly used RNA-Seq data normalization methods.

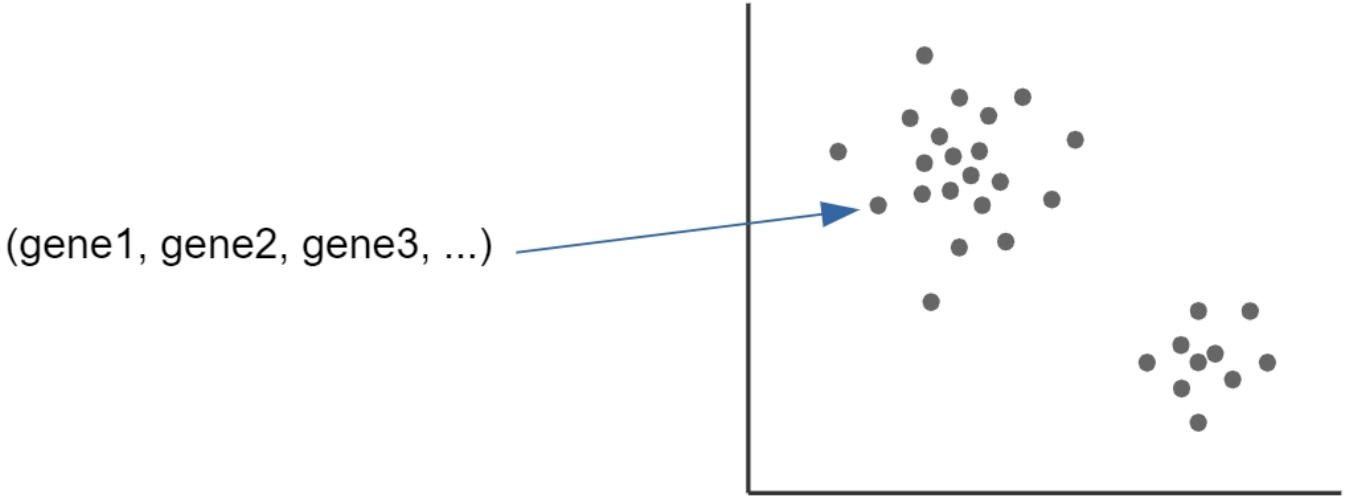


Compare expression for a cell type between groups

Multiple single-sample t-SNE plots

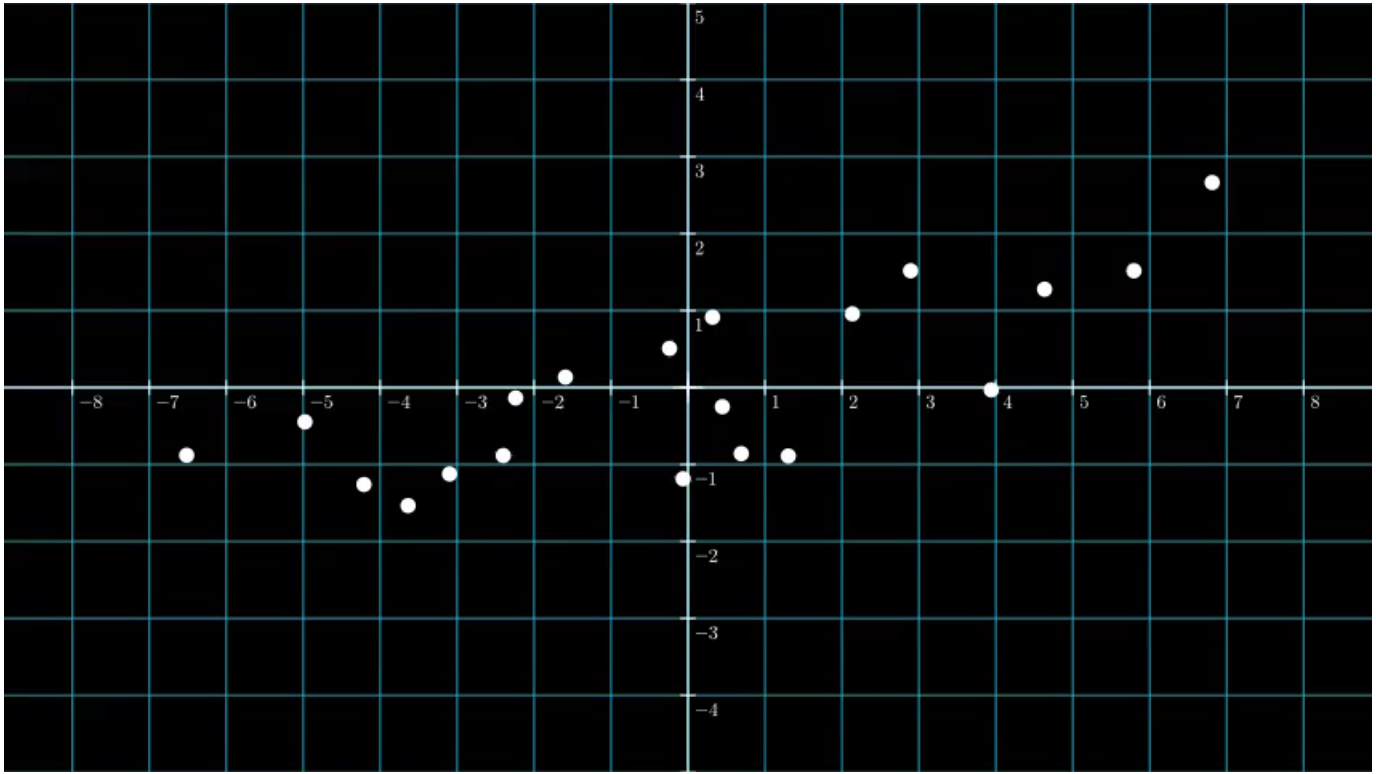
- Click the **Filtered counts** data node after the Filter features task
- Select **PCA** from the **Exploratory analysis** section of the task menu





(w, x, y, z)	(x', y')
(0, 0, 0, 0)	(0, 0)
(0, 0, 0, 1)	(0, 1)
(0, 0, 1, 0)	(1, 0)
(0, 0, 1, 1)	(1, 1)
(0, 1, 0, 0)	

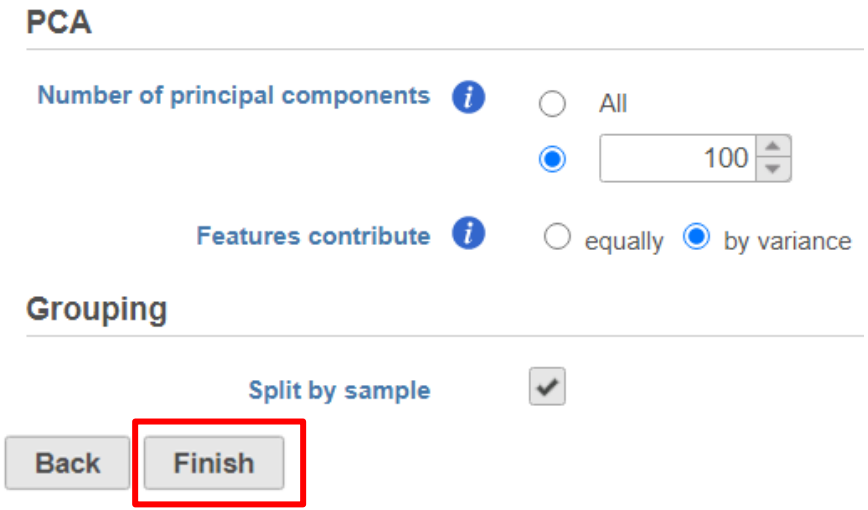
PCA



https://leemeng.tw/essence-of-principal-component-analysis.html

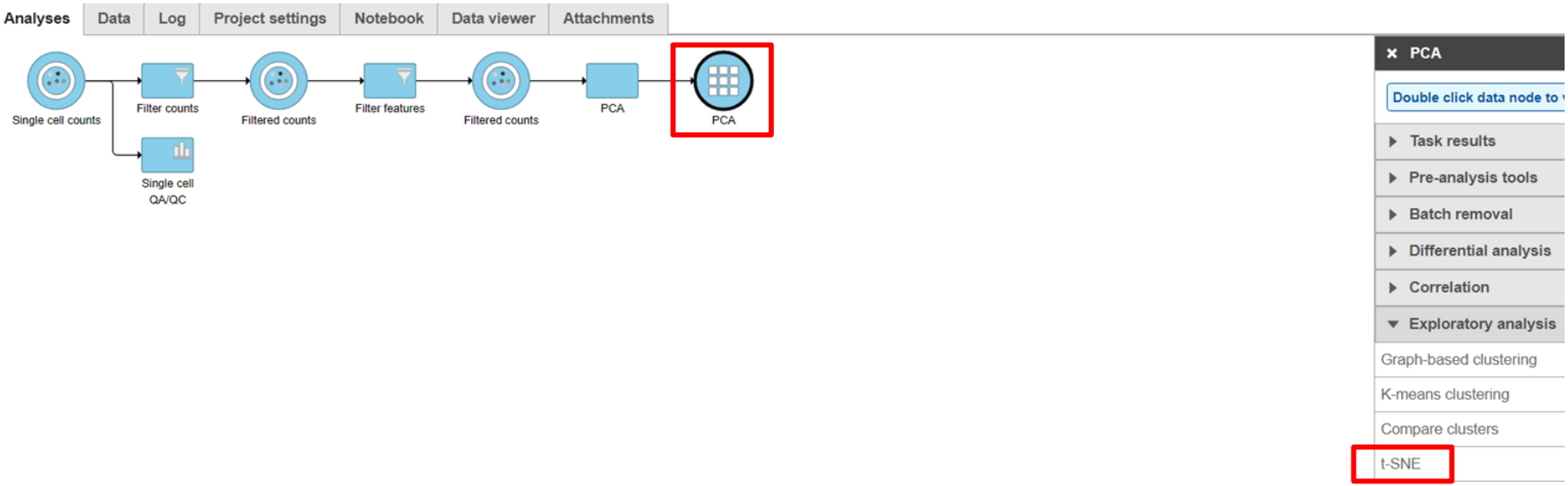
Multiple single-sample t-SNE plots

- Click **Finish** to run PCA with default settings

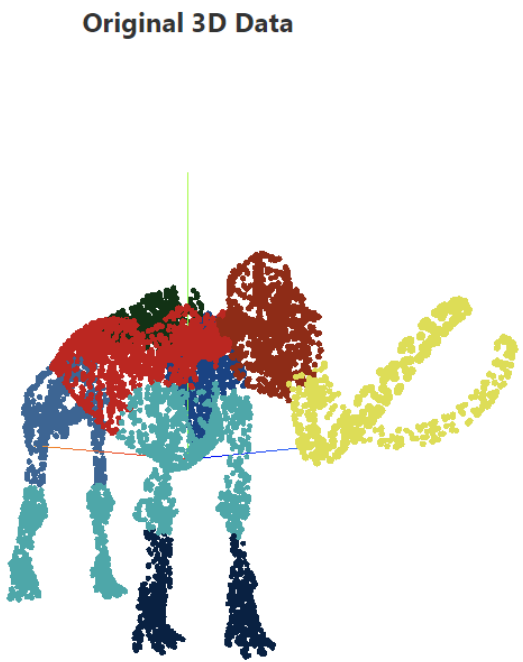


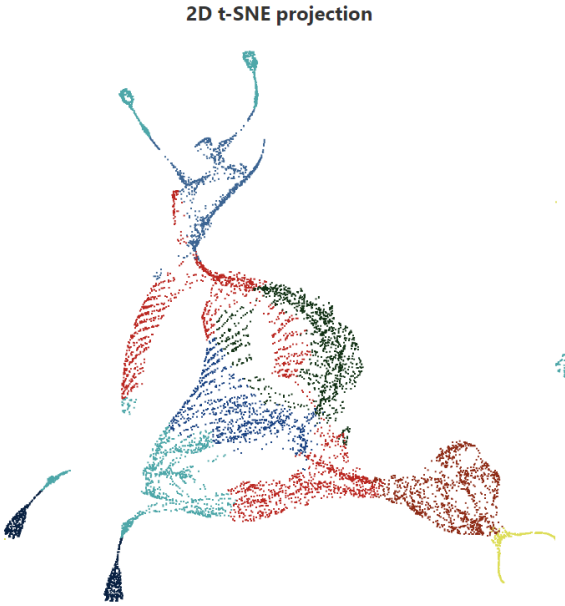
Multiple single-sample t-SNE plots

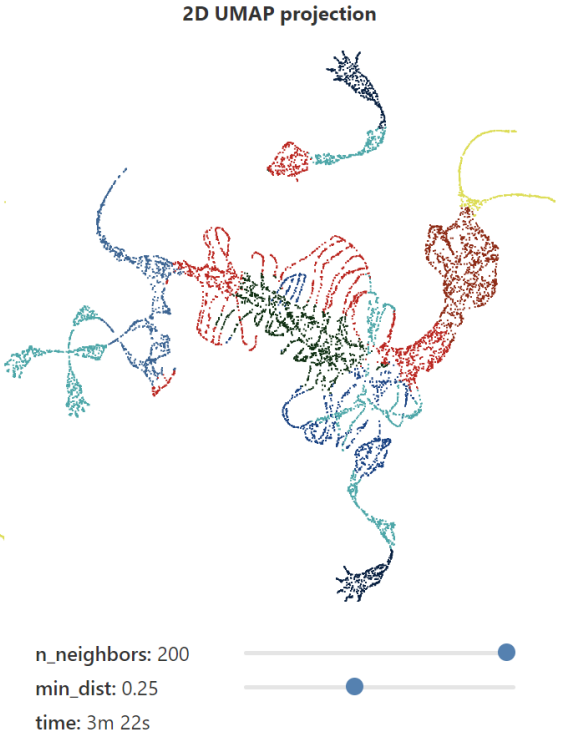
- Click the **PCA** data node
- Select **t-SNE** from the **Exploratory analysis** section of the task menu



UMAP & t-SNE



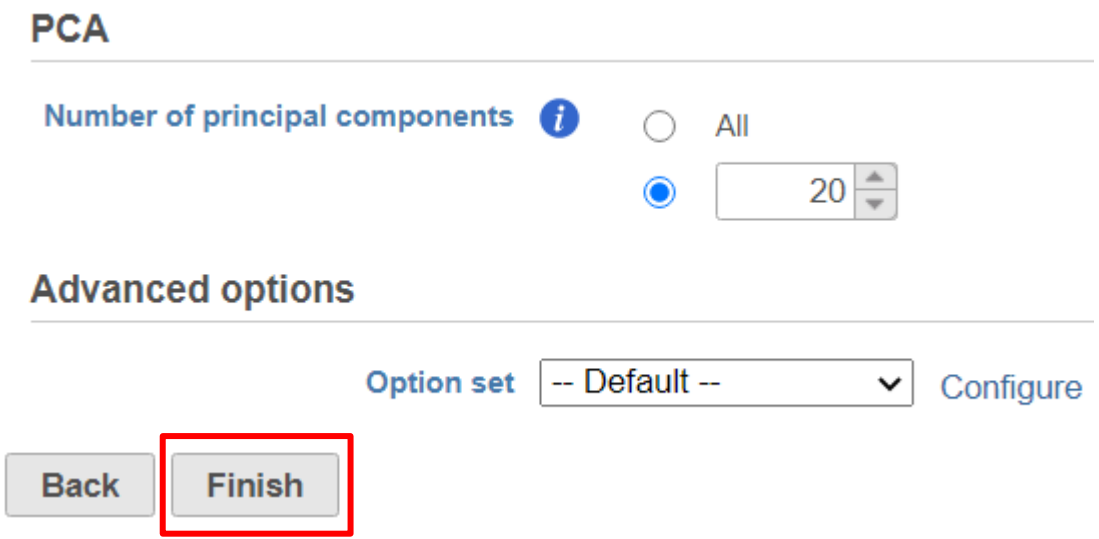




https://pair-code.github.io/understanding-umap/

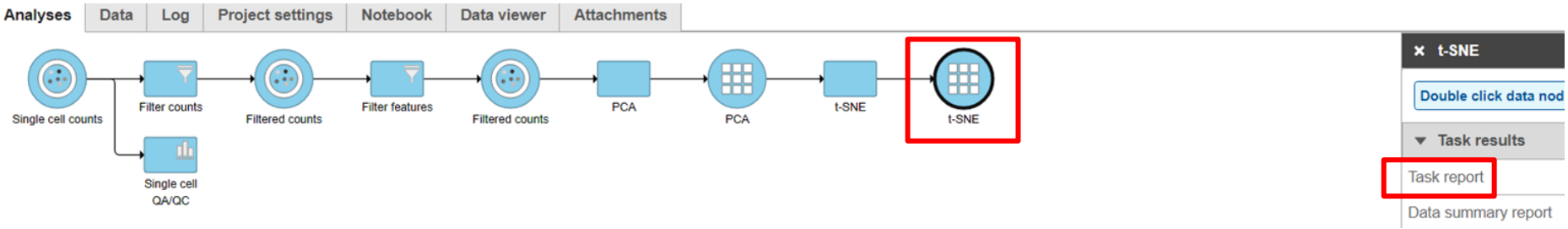
Multiple single-sample t-SNE plots

- Click Finish from the t-SNE dialog to run t-SNE with the default settings



Multiple single-sample t-SNE plots

- Click the **t-SNE** node
- Click **Task report** from the task menu or double click the t-SNE node



Multiple single-sample t-SNE plots

- Click **Axes** in the Configure on the left
- Select the ▶ icon below the sample name to go to the next sample
  - The t-SNE plot has switched to show the next sample, MGH42



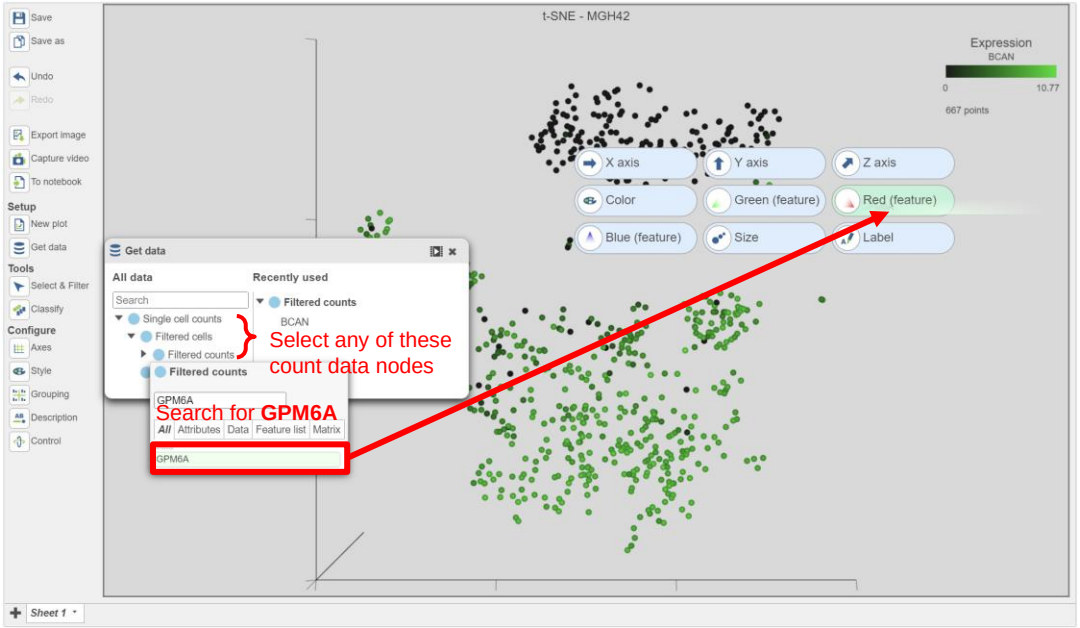
Multiple single-sample t-SNE plots

- Select any of the count data nodes from the Get Data window
- Search for the **BCAN** gene
- Click and drag the **BCAN** gene onto the plot and drop it over the **Green (feature)** option



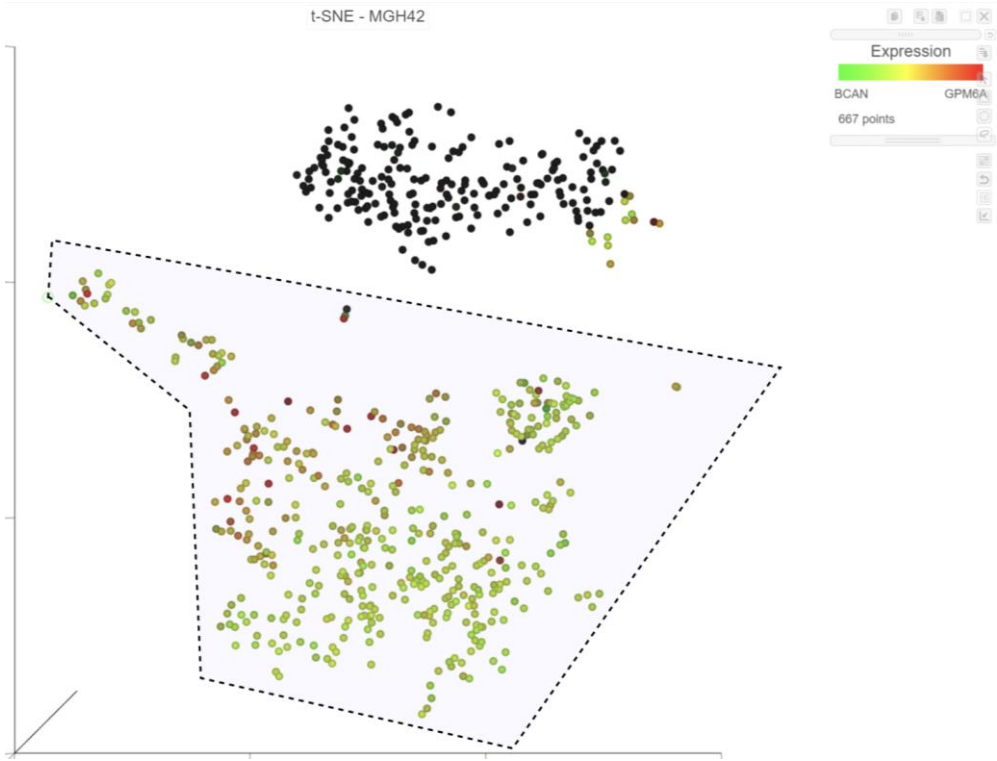
Multiple single-sample t-SNE plots

- Select any of the count data nodes from the Get Data window
- Search for the **GPM6A** gene
- Click and drag the **GPM6A** gene onto the plot and drop it over the **Red (feature)** option



Multiple single-sample t-SNE plots

- Switch to lasso mode by clicking in the top right of the plot
- Draw the lasso around the cluster of green, red, and yellow cells and click the circle to close the lasso



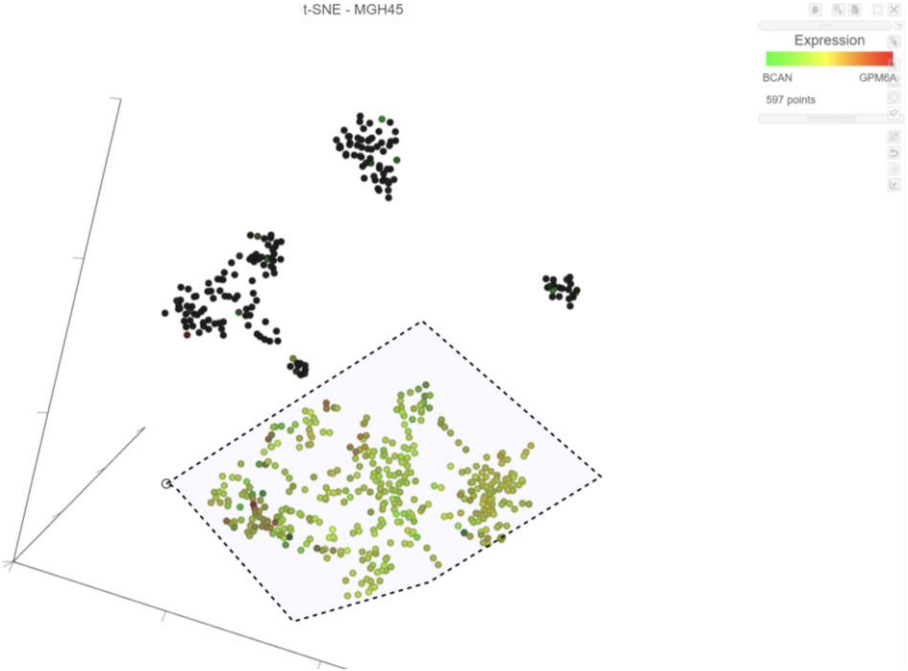
Multiple single-sample t-SNE plots

- Click **Classify selection** in the Classify window
- Name the classification **Glioma**
- Click **Save**



Multiple single-sample t-SNE plots

- Select the ▶ icon in Axes below the sample name to go to the next sample, **MGH45**
- Rotate the 3D t-SNE plot to get a better view of cells from the green, red, and yellow cluster
- Switch to lasso mode by selecting in the top right corner of the plot
- Draw the lasso around the cluster of colored cells and click the circle to close the lasso



Multiple single-sample t-SNE plots

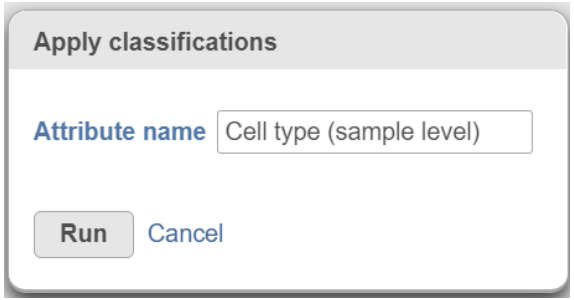
- Select **Classify selection** in the Classify window
- Type **Glioma** or select **Glioma** from the drop-down list
- Click **Save**



Repeat these steps for each of the 6 remaining samples. Remember to go back to the first sample (MGH36) to classify the glioma cells in that samples too.

Multiple single-sample t-SNE plots

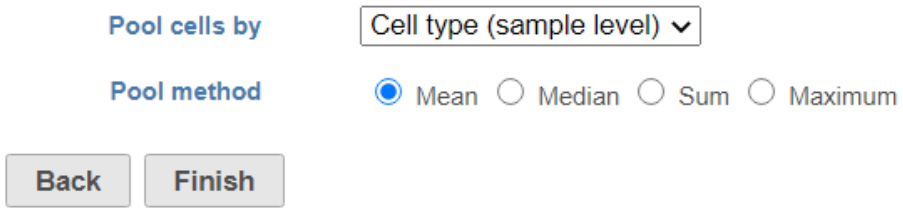
- Click **Apply classifications** in the Classify window
- Name the classification attribute **Cell type (sample level)**
- Click **Run**
- Click **OK** on the information box that says an attribute task has been created





Pool cells

- Click the **Filtered counts** data node
- Expand the **Pre-analysis tools** section of the task menu
- Click **Pool cells**
- Choose **Cell type (sample level)** from the Pool cells by drop-down list
- Keep the default pool method (**Mean**)
- Click **Finish**

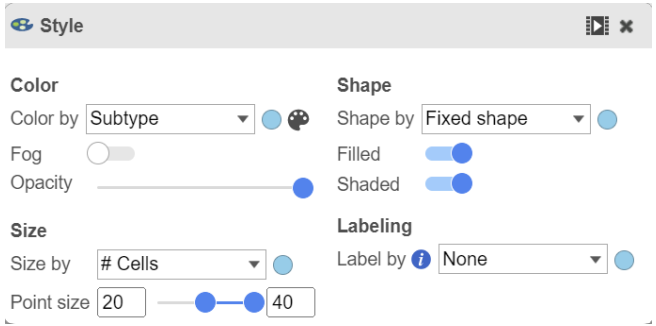


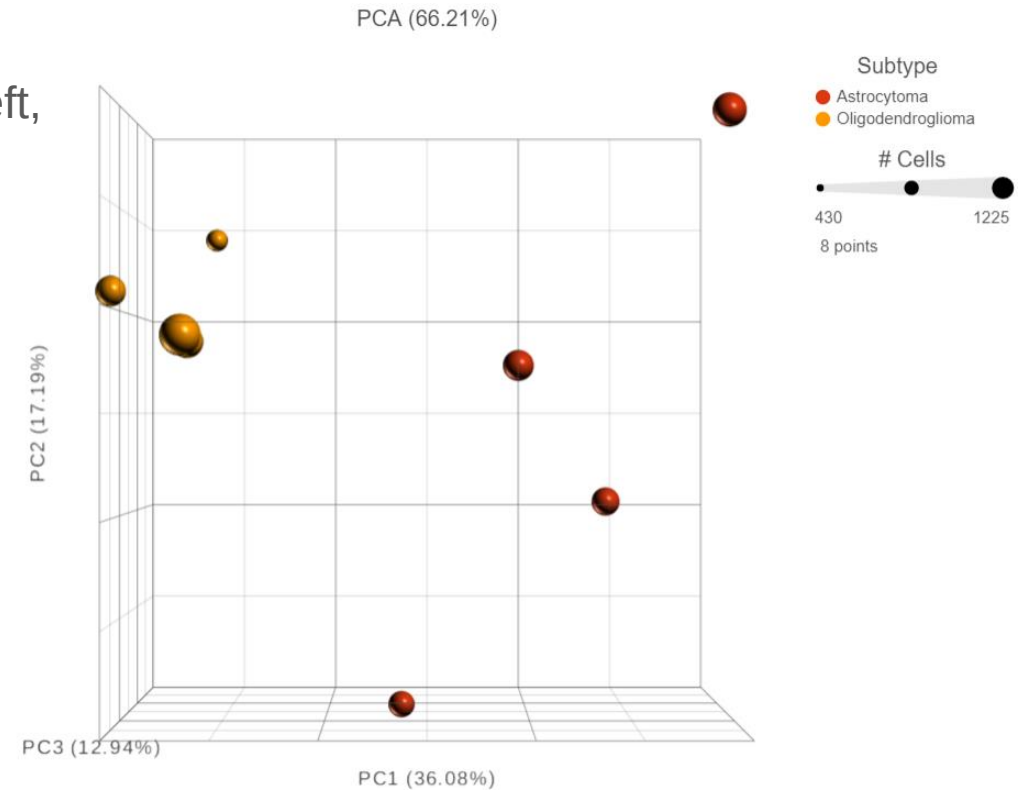
Explore data with PCA

- Click the **Glioma** data node
- Click the **Exploratory analysis** section of the task menu
- Click **PCA**
- Click **Finish** in the PCA dialog to run PCA with default settings
- Once the task is complete, double click the **PCA** node to open the PCA scatter plot

Explore data with PCA

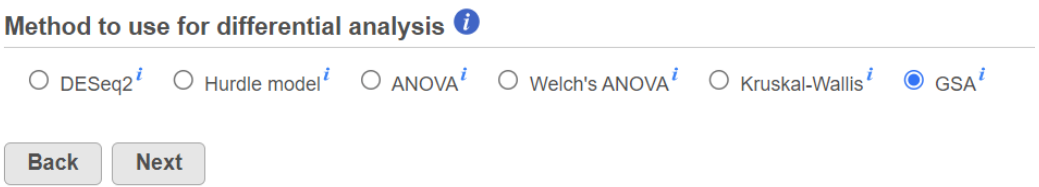
- In the **Configure** section on the left, click the **Style** and choose **Subtype** from the **Color by** drop-down list
- Besides, choose **# Cells** from the **Size by** drop-down list

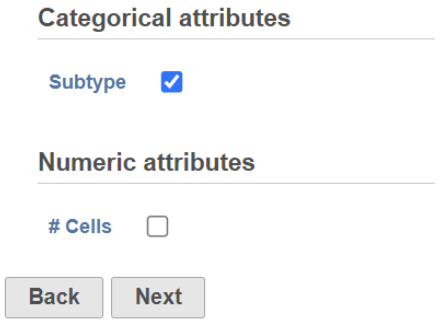




Identify differentially expressed genes

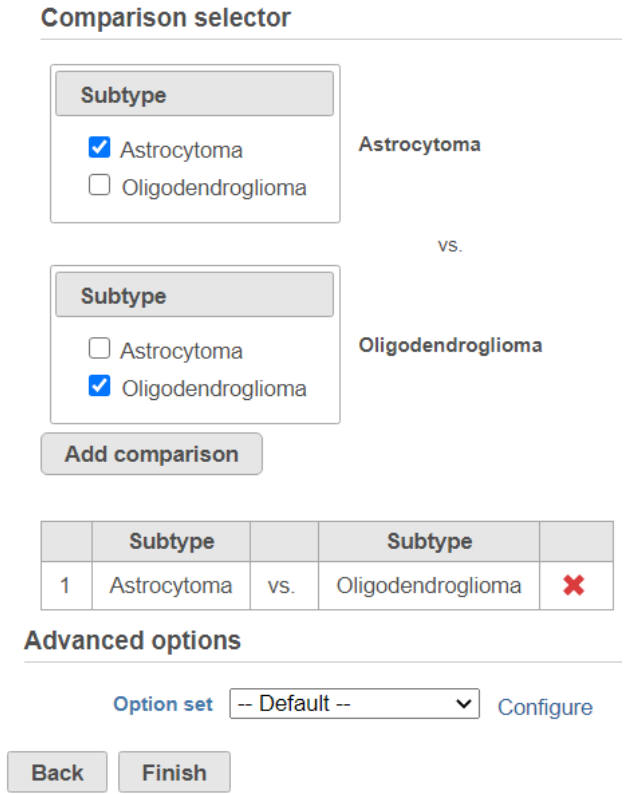
- Click **Glioma (multi-sample)** to return to the Analyses tab
- Click the **Glioma** data node
- Expand **Statistics** in the task menu
- Click **Differential Analysis** in the task menu
- Check the **GSA** attribute
- Click **Next**
- Check the **Subtype** attribute
- Click **Next**





Identify differentially expressed genes

- Click **Astrocytoma** in the top panel
- Click **Oligodendroglioma** in the bottom panel
- Click **Add comparison**
- Click **Finish** to run the GSA



Identify differentially expressed genes

- Double click the **GSA** data node
- Select  in the top right corner of the table to open the volcano plot for a comparison

Feature list

Results: 17242

Filter

- Feature ID
- P-value
- FDR step up
- Ratio
- Fold change
- LSMean
- Low expressed

Save filter | Clear filter

Saved filters

(No saved filters available)

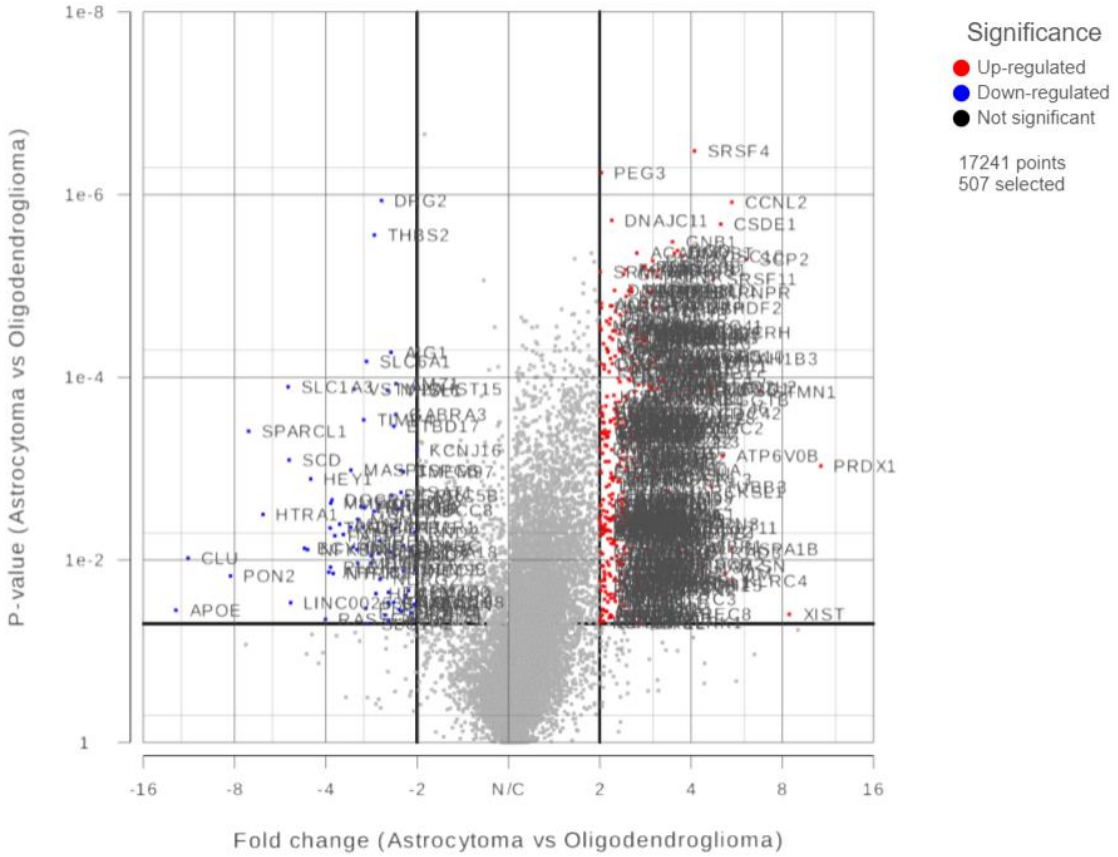
Generate filtered node

Save as managed list

Optional columns

			Astrocytoma vs Oligodendroglioma					
	View	Feature ID	P-value	FDR step up	Ratio	Fold change	LSMean(Astrocytoma)	LSMean(Oligodendroglioma)
1		SYT4	2.19E-7	2.87E-3	0.53	-1.89	1.03	1.95
2		SRSF4	3.33E-7	2.87E-3	4.11	4.11	7.53	1.83
3		PEG3	5.79E-7	3.33E-3	2.03	2.03	2.03	1.00
4		DRG2	1.16E-6	4.19E-3	0.38	-2.62	1.08	2.82
5		CCNL2	1.21E-6	4.19E-3	5.46	5.46	18.42	3.38
6		DNAJC11	1.92E-6	4.84E-3	2.19	2.19	3.59	1.64
7		CSDE1	2.1E-6	4.84E-3	5.01	5.01	22.64	4.52
8		THBS2	2.78E-6	4.84E-3	0.36	-2.77	1.06	2.93
9		GNB1	3.3E-6	4.84E-3	3.48	3.48	7.52	2.16
10		PGD	4.11E-6	4.84E-3	3.61	3.61	13.69	3.79
11		ACADM	4.35E-6	4.84E-3	2.65	2.65	4.36	1.65
12		GNL2	4.38E-6	4.84E-3	1.88	1.88	2.48	1.32
13		DDOST	4.42E-6	4.84E-3	3.52	3.52	18.51	5.25
14		EXOSC10	4.75E-6	4.84E-3	3.76	3.76	9.05	2.41
15		STIM2	5.05E-6	4.84E-3	1.66	1.66	1.85	1.12
16		SCP2	5.12E-6	4.84E-3	6.09	6.09	20.43	3.36

Identify differentially expressed genes



Identify differentially expressed genes

- Click the **GSA report** tab in your browser to return to the feature list
- Select  next to SYT4 to open its feature plot

Feature list

Results: 17242

Filter

- Feature ID
- P-value
- FDR step up
- Ratio
- Fold change
- LSMean
- Low expressed

Save filter | Clear filter

Saved filters

(No saved filters available)

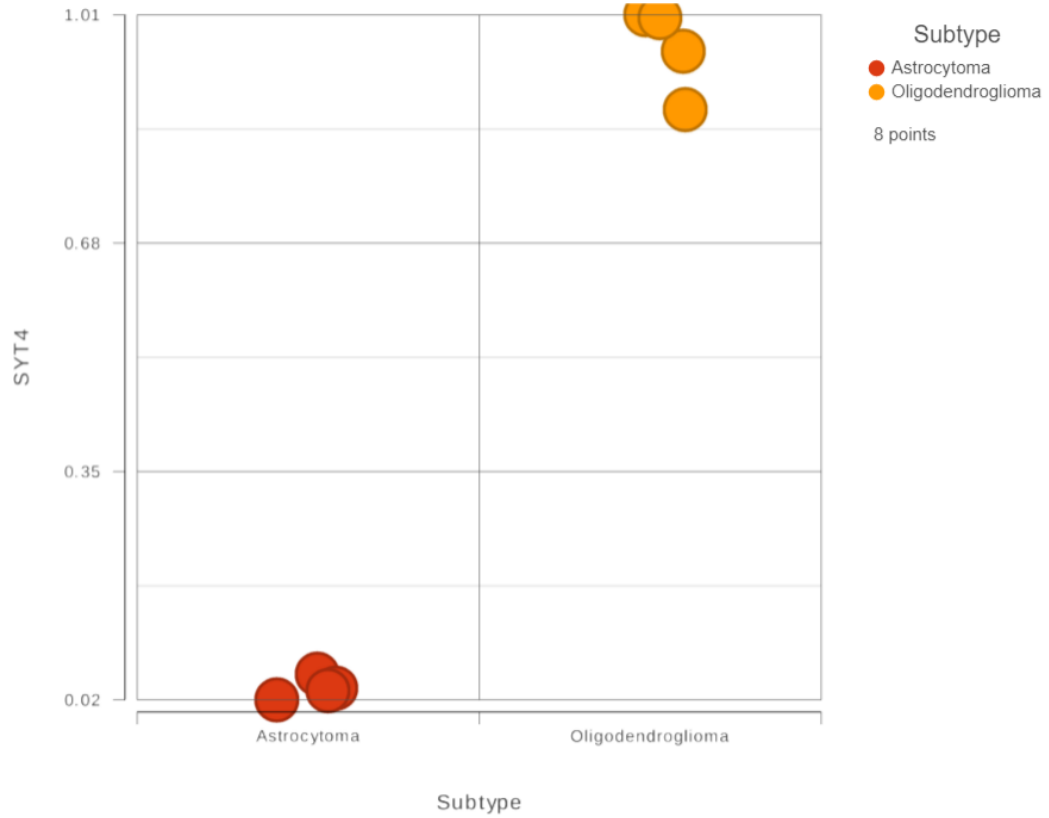
Generate filtered node

Save as managed list

Optional columns

			Astrocytoma vs Oligodendroglioma					
	View	Feature ID	P-value	FDR step up	Ratio	Fold change	LSMean(Astrocytoma)	LSMean(Oligodendroglioma)
1		SYT4	2.19E-7	2.87E-3	0.53	-1.89	1.03	1.95
2		SRSF4	3.33E-7	2.87E-3	4.11	4.11	7.53	1.83
3		PEG3	5.79E-7	3.33E-3	2.03	2.03	2.03	1.00
4		DRG2	1.16E-6	4.19E-3	0.38	-2.62	1.08	2.82
5		CCNL2	1.21E-6	4.19E-3	5.46	5.46	18.42	3.38
6		DNAJC11	1.92E-6	4.84E-3	2.19	2.19	3.59	1.64
7		CSDE1	2.1E-6	4.84E-3	5.01	5.01	22.64	4.52
8		THBS2	2.78E-6	4.84E-3	0.36	-2.77	1.06	2.93
9		GNB1	3.3E-6	4.84E-3	3.48	3.48	7.52	2.16
10		PGD	4.11E-6	4.84E-3	3.61	3.61	13.69	3.79
11		ACADM	4.35E-6	4.84E-3	2.65	2.65	4.36	1.65
12		GNL2	4.38E-6	4.84E-3	1.88	1.88	2.48	1.32
13		DDOST	4.42E-6	4.84E-3	3.52	3.52	18.51	5.25
14		EXOSC10	4.75E-6	4.84E-3	3.76	3.76	9.05	2.41
15		STIM2	5.05E-6	4.84E-3	1.66	1.66	1.85	1.12
16		SCP2	5.12E-6	4.84E-3	6.09	6.09	20.43	3.36

Identify differentially expressed genes



Identify differentially expressed genes

- Click **GSA report** tab to return to the results table
- The filter panel is used to filter results to identify significantly differentially regulated genes. Here, we will create a list of genes that have FDR step up < 0.05 and Fold change > 2 or < -2.
  - Click **FDR step up**
  - Set it to **0.05** with the drop-down menu set to **Less than or equal to**
  - Press **Enter**
  - Click **Fold change**
  - Set it to **-2 to 2** for the excluded range
  - Press **Enter**
  - Click 

to apply the filter

Identify differentially expressed genes

Results: 334

Filter

- Feature ID
- P-value
- FDR step up: Less than or equ 0.05; 0 — 1
- Ratio
- Fold change: From -2 to 2; Exclude range
- LSMean
- Low expressed

Save filter | Clear filter

Saved filters

(No saved filters available)

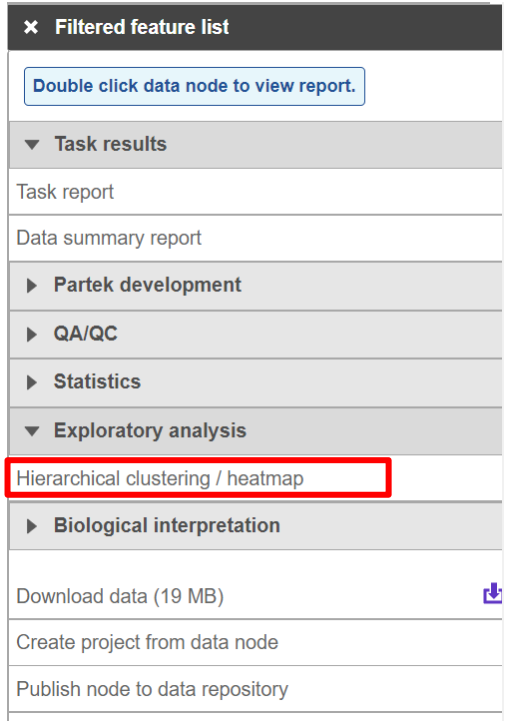
Generate filtered node

Optional columns

			Astrocytoma vs Oligodendroglioma					
	View	Feature ID	P-value	FDR step up	Ratio	Fold change	LSMean(Astrocytoma)	LSMean(Oligodendroglioma)
1		SRSF4	3.33E-7	2.87E-3	4.11	4.11	7.53	1.83
2		PEG3	5.79E-7	3.33E-3	2.03	2.03	2.03	1.00
3		DRG2	1.16E-6	4.19E-3	0.38	-2.62	1.08	2.82
4		CCNL2	1.21E-6	4.19E-3	5.46	5.46	18.42	3.38
5		DNAJC11	1.92E-6	4.84E-3	2.19	2.19	3.59	1.64
6		CSDE1	2.1E-6	4.84E-3	5.01	5.01	22.64	4.52
7		THBS2	2.78E-6	4.84E-3	0.36	-2.77	1.06	2.93
8		GNB1	3.3E-6	4.84E-3	3.48	3.48	7.52	2.16
9		PGD	4.11E-6	4.84E-3	3.61	3.61	13.69	3.79
10		ACADM	4.35E-6	4.84E-3	2.65	2.65	4.36	1.65
11		DDOST	4.42E-6	4.84E-3	3.52	3.52	18.51	5.25
12		EXOSC10	4.75E-6	4.84E-3	3.76	3.76	9.05	2.41
13		SCP2	5.12E-6	4.84E-3	6.09	6.09	20.43	3.36
14		RBBP4	5.33E-6	4.84E-3	2.99	2.99	12.66	4.23
15		TLE2	6.2E-6	4.84E-3	2.82	2.82	3.60	1.28
16		PRPF38B	6.22E-6	4.84E-3	2.78	2.78	6.24	2.24
17		FOXJ3	6.64E-6	4.84E-3	2.43	2.43	4.23	1.74
18		AKIRIN1	6.68E-6	4.84E-3	3.13	3.13	6.56	2.09
19		NRD1	6.77E-6	4.84E-3	2.91	2.91	6.29	2.16

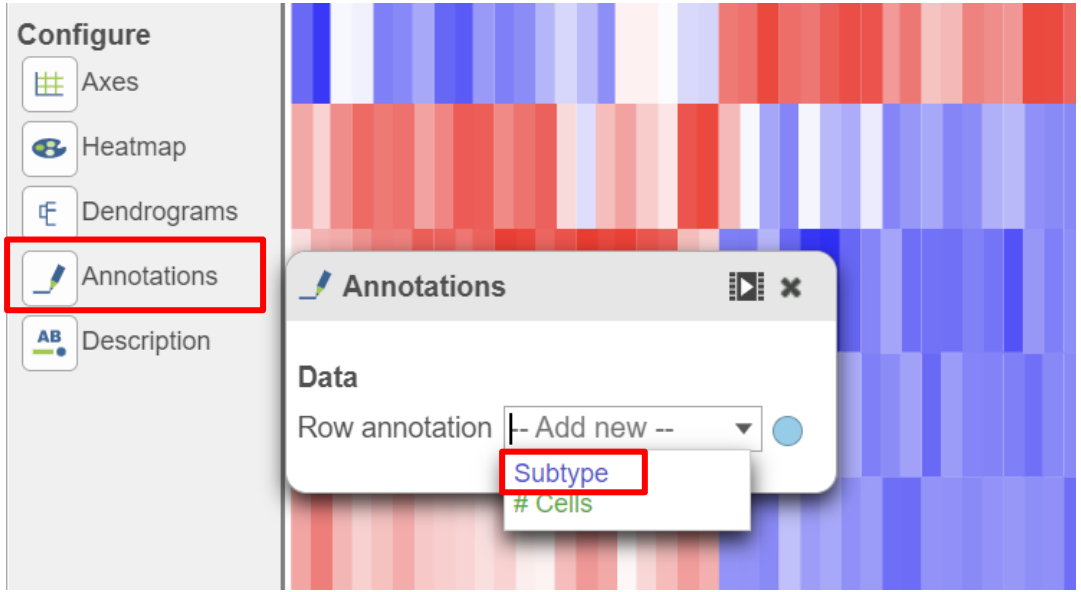
Analyze differentially expressed genes

- Click the **Filtered feature list** data node produced by the GSA task node
- Expand **Exploratory analysis** in the task menu
- Click **Hierarchical clustering / heat map**
- Click **Finish** to continue with default settings
- Double click the **Hierarchical clustering** task node to open the heat map



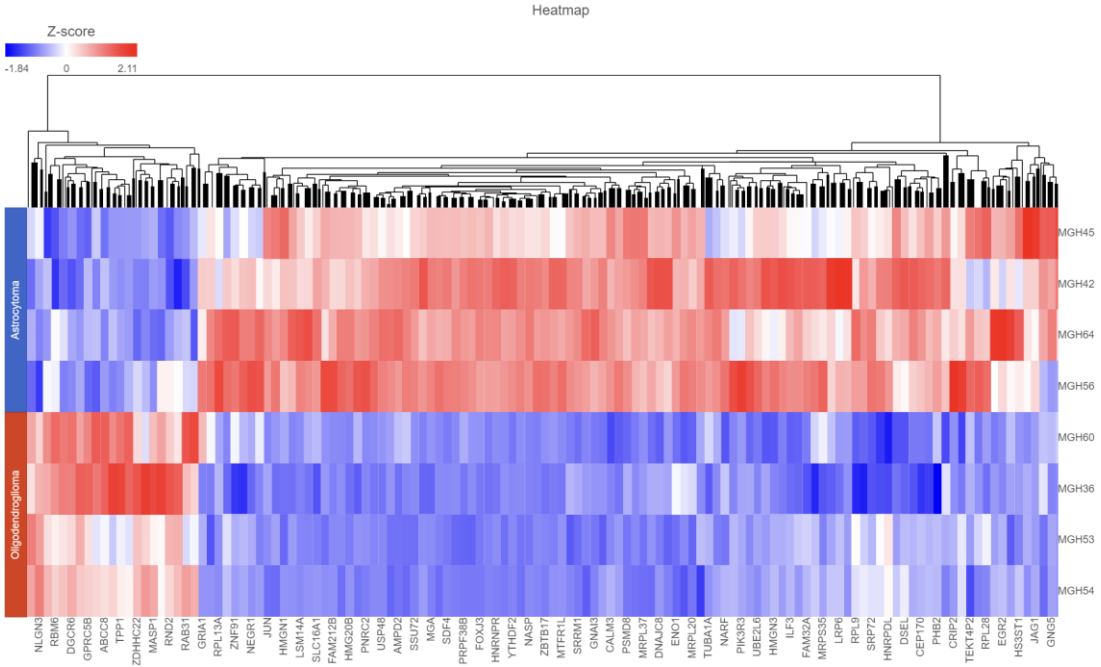
Analyze differentially expressed genes

- Click **Annotations** on the left
- Select **Subtype** from the **Row annotation** drop-down menu



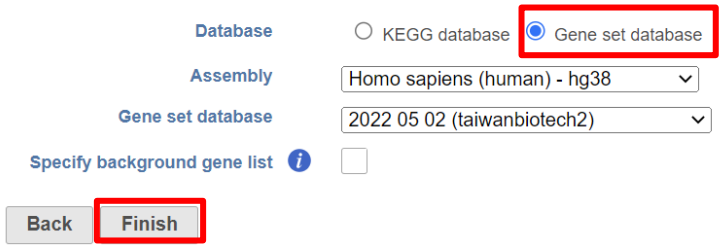
Analyze differentially expressed genes

- The heat map can be saved as a publication-quality image by clicking the save image icon  or sent to a page in the Notebook tab by clicking  .



Analyze differentially expressed genes

- Click **Glioma (multi-sample)** to return to the Analyses tab
- Click the **Filtered feature list** data node
- Expand **Biological interpretation** in the task menu
- Click **Gene set enrichment** analysis
- Click **Gene set database**
- Select **Homo sapiens (human) - hg38** from the Assembly drop-down menu
- Select a gene set database from the drop-down menu. If one isn't available, choose **Add gene set database** and download one
- Click **Finish**



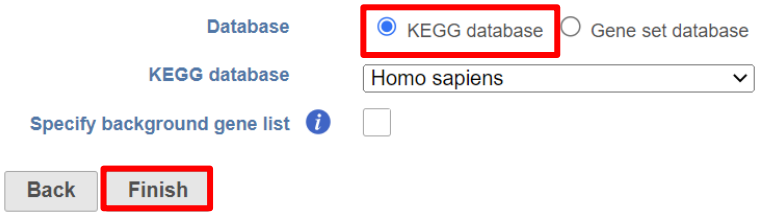
Analyze differentially expressed genes

- Double click the **Gene set enrichment** node to open the task report

Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list
GO:0003723	RNA binding	45.62	1.54E-20	83	1,473
GO:0005654	nucleoplasm	37.62	4.61E-17	125	3,336
GO:0032991	protein-containing complex	33.31	3.41E-15	150	4,672
GO:0003676	nucleic acid binding	32.14	1.1E-14	128	3,715
GO:0016071	mRNA metabolic process	32.12	1.13E-14	43	571
GO:0043229	intracellular organelle	30.32	6.82E-14	226	8,926
GO:0097159	organic cyclic compound binding	26.08	4.71E-12	160	5,583
GO:0006396	RNA processing	25.99	5.15E-12	44	722
GO:1902494	catalytic complex	25.88	5.77E-12	59	1,209
GO:1901363	heterocyclic compound binding	25.80	6.22E-12	158	5,500
GO:0043226	organelle	23.46	6.49E-11	236	10,026
GO:0008380	RNA splicing	22.63	1.48E-10	27	322
GO:0006397	mRNA processing	22.39	1.9E-10	29	376
GO:0006139	nucleobase-containing compound metabolic process	22.20	2.28E-10	87	2,428
GO:1990904	ribonucleoprotein complex	21.85	3.25E-10	37	609
GO:0005515	protein binding	21.76	3.54E-10	248	10,916
GO:1901360	organic cyclic compound metabolic process	21.43	4.93E-10	96	2,851
GO:0046483	heterocycle metabolic process	21.37	5.23E-10	90	2,596
GO:0044237	cellular metabolic process	21.19	6.3E-10	177	6,812
GO:0034641	cellular nitrogen compound metabolic process	21.18	6.31E-10	97	2,908
GO:0016070	RNA metabolic process	20.92	8.2E-10	61	1,452
GO:0043231	intracellular membrane-bounded organelle	19.77	2.59E-9	187	7,466

Analyze differentially expressed genes

- Click **Glioma (multi-sample)** to return to the Analyses tab
- Click the **Filtered feature list** data node
- Expand **Biological interpretation** in the task menu
- Click **Gene set enrichment** analysis
- Click **KEGG database**
- Verify that the species in the drop-down menu is **Homo sapiens**
- Click **Finish** to run enrichment analysis



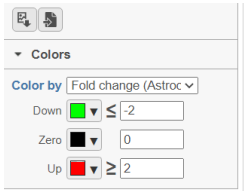
Analyze differentially expressed genes

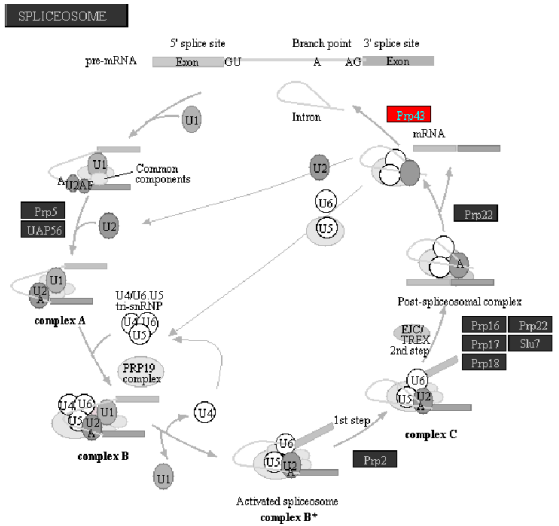
- Double click the **Pathway enrichment** node

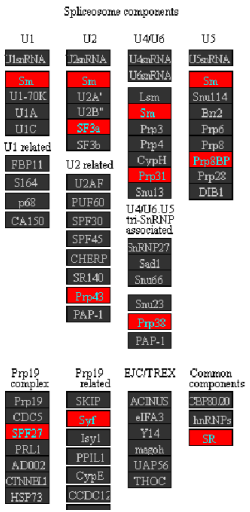
Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
path:hsa03040	Spliceosome	9.01	1.22E-4	11	125	
path:hsa05012	Parkinson disease	7.95	3.51E-4	10	119	
path:hsa04714	Thermogenesis	7.38	6.21E-4	13	205	
path:hsa03015	mRNA surveillance pathway	5.80	3.03E-3	7	85	
path:hsa00190	Oxidative phosphorylation	5.59	3.74E-3	8	113	
path:hsa03320	PPAR signaling pathway	5.30	4.98E-3	6	70	
path:hsa01212	Fatty acid metabolism	5.10	6.07E-3	5	51	
path:hsa03010	Ribosome	4.85	7.86E-3	8	129	
path:hsa00100	Steroid biosynthesis	4.84	7.87E-3	3	17	
path:hsa00072	Synthesis and degradation of ketone bodies	4.05	0.02	2	8	
path:hsa00280	Valine, leucine and isoleucine degradation	4.04	0.02	4	44	
path:hsa01040	Biosynthesis of unsaturated fatty acids	4.01	0.02	3	24	
path:hsa05016	Huntington disease	3.98	0.02	11	246	
path:hsa00270	Cysteine and methionine metabolism	3.97	0.02	4	45	
path:hsa05214	Glioma	3.91	0.02	5	70	
path:hsa03013	RNA transport	3.88	0.02	8	155	
path:hsa04721	Synaptic vesicle cycle	3.71	0.02	5	74	
path:hsa05203	Viral carcinogenesis	3.69	0.02	9	192	
path:hsa04146	Peroxisome	3.52	0.03	5	78	
path:hsa04932	Non-alcoholic fatty liver disease (NAFLD)	3.44	0.03	7	138	
path:hsa00640	Propanoate metabolism	3.40	0.03	3	31	
path:hsa01200	Carbon metabolism	3.34	0.04	6	111	

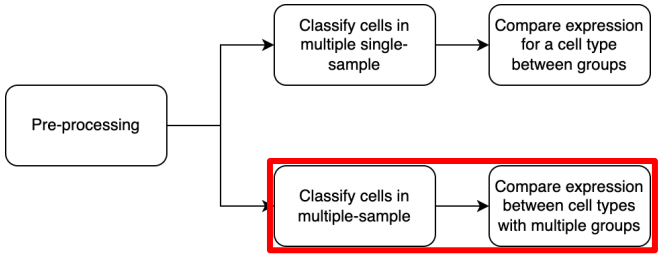
Analyze differentially expressed genes

- Click **path:hsa03040** (Spliceosome) to open its KEGG pathway map
- Here, we can see that several components of the spliceosome pathway are upregulated in astrocytoma vs. oligodendroglioma.





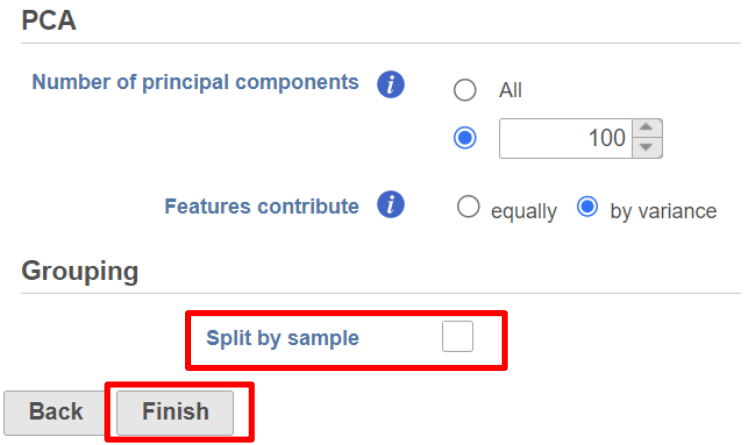




Compare expression between cell types with multiple groups

One multi-sample t-SNE plot

- For some data sets, cell types can be distinguished when all samples can be visualized together on one t-SNE plot. We will use a t-SNE plot of all samples to classify glioma, microglia, and oligodendrocyte cell types.
    - Click on the **Glioma (multi-sample)** project name at the top to go back to the Analyses tab
    - Click the **Filtered counts** data node after the Filter features task
    - Click **PCA** in the **Exploratory analysis** section of the task menu
    - Uncheck the **Split by sample** checkbox
    - Click **Finish**

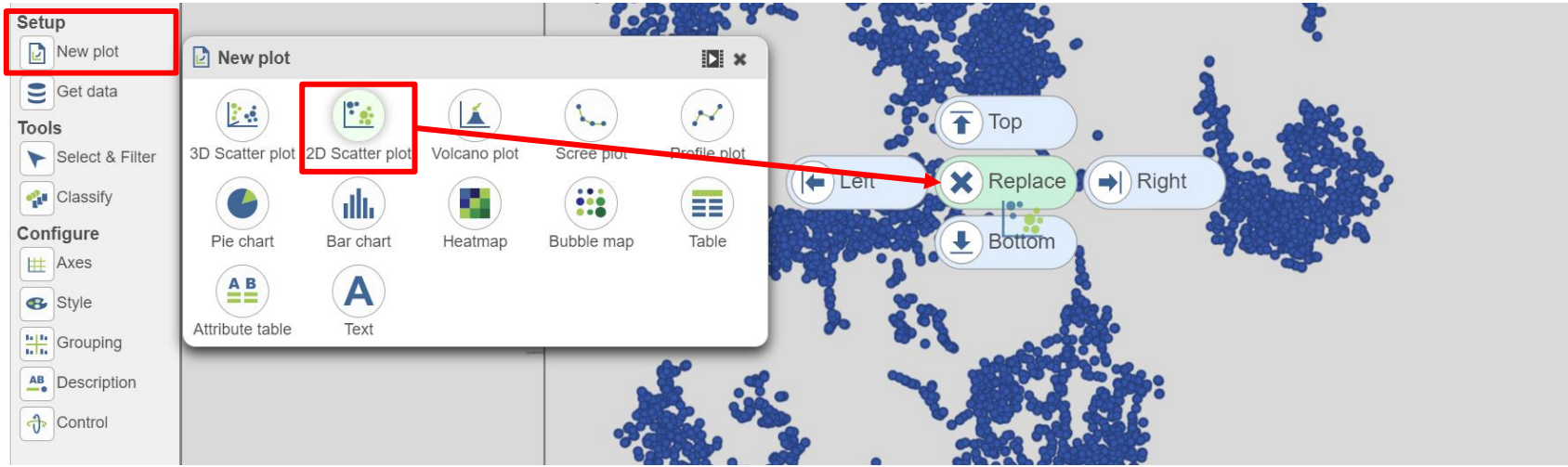


One multi-sample t-SNE plot

- Click the **new PCA** data node
- Select **t-SNE** from the **Exploratory analysis** section of the task menu
- Click **Finish** to run the t-SNE task with default settings
- Double-click the green **t-SNE** data node to open the t-SNE scatter plot

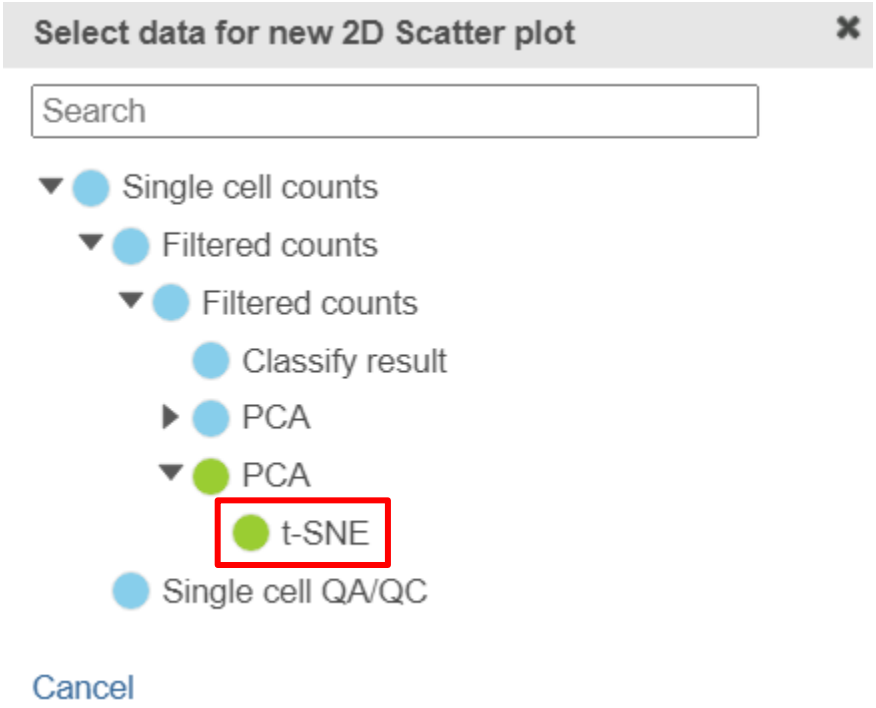
One multi-sample t-SNE plot

- Click and drag the **2D scatter plot** icon in the New Plot window onto the canvas and **replace** the 3D scatter plot



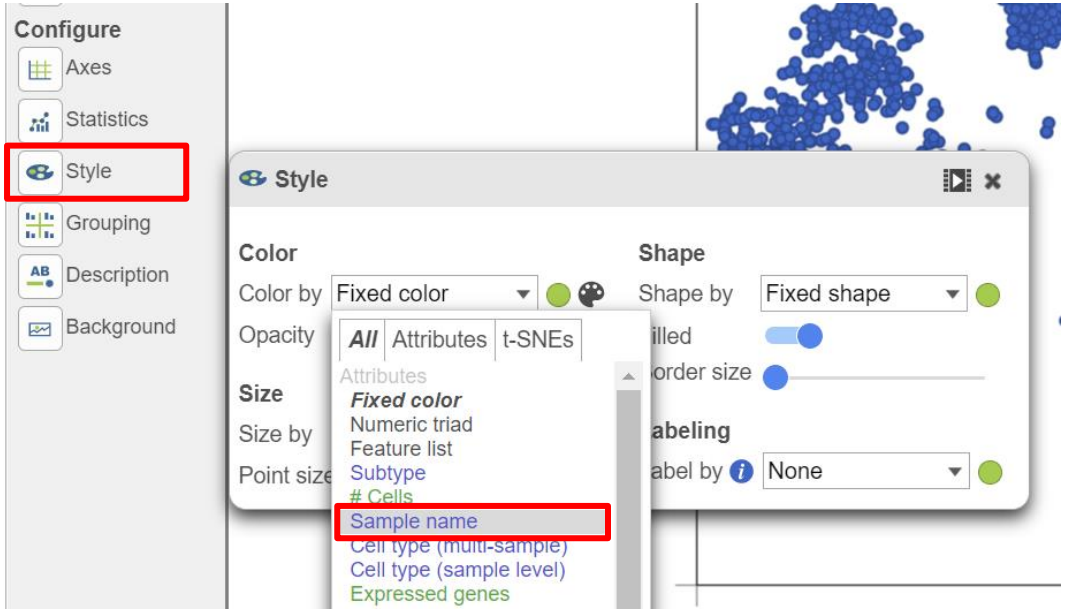
One multi-sample t-SNE plot

- Search for and select the **new t-SNE** data node



One multi-sample t-SNE plot

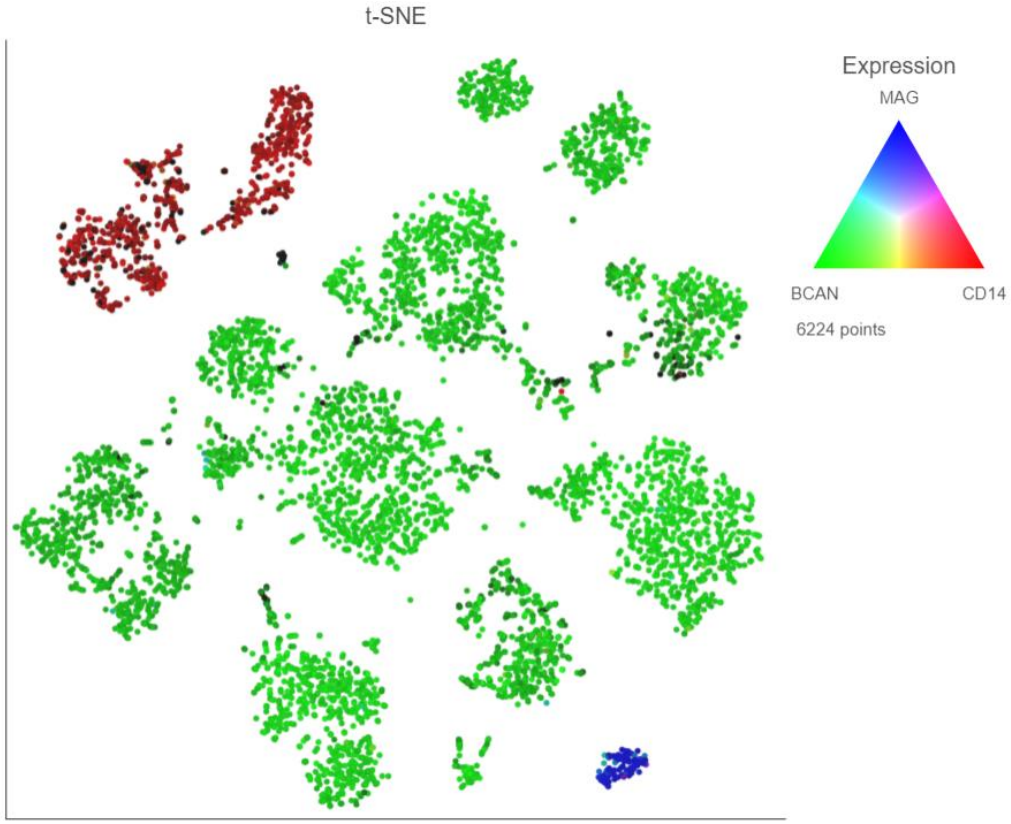
- In the Configure section, click **Style** and choose **Sample name** from the **Color by** drop-down list



One multi-sample t-SNE plot

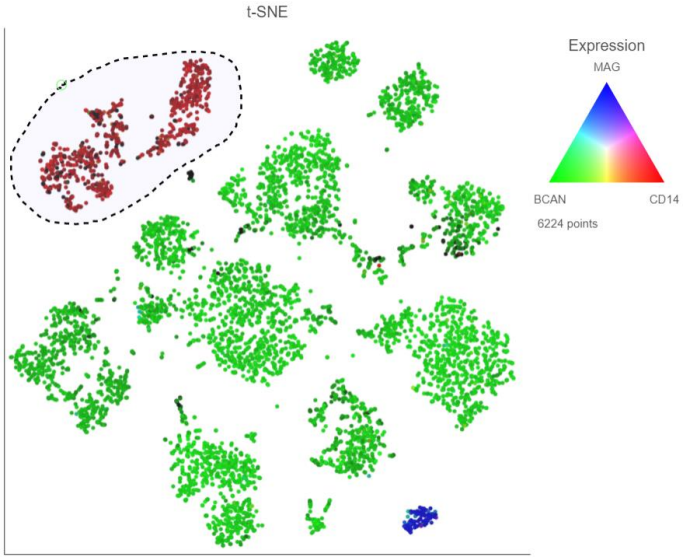
- Using marker genes, **BCAN (glioma), CD14 (microglia), and MAG (oligodendrocytes)**, we can assess whether these multi-sample clusters belong to our known cell types.
  - Select any of the count data nodes from the **Data** card on the left (Single cell counts, or any of the Filtered counts)
  - Search for the **BCAN** gene
    - Click and drag the BCAN gene onto the plot and drop it over the **Green (feature)** option
  - Search for the **CD14** gene
    - Click and drag the CD14 gene onto the plot and drop it over the **Red (feature)** option
  - Search for the **MAG** gene
    - Click and drag the MAG gene onto the plot and drop it over the **Blue (feature)** option

One multi-sample t-SNE plot



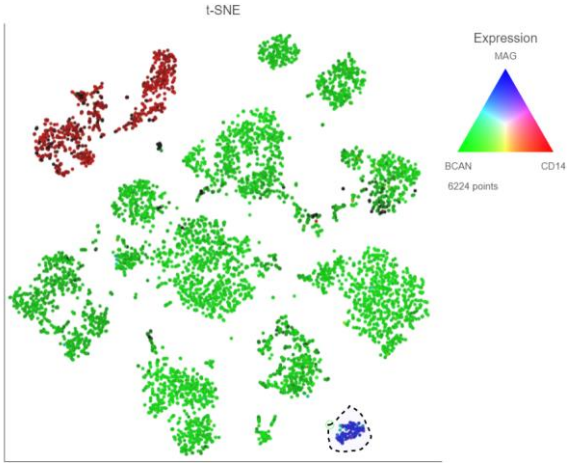
One multi-sample t-SNE plot

- The red cells are CD14 positive, indicating that they are the microglia from every sample.
  - Switch to lasso mode by clicking the  icon in the top right of the plot
  - Draw the lasso around the cluster of red cells and click the circle to close the lasso
  - Click **Classify selection** in Classify
  - Name the classification **Microglia**
  - Click **Save**



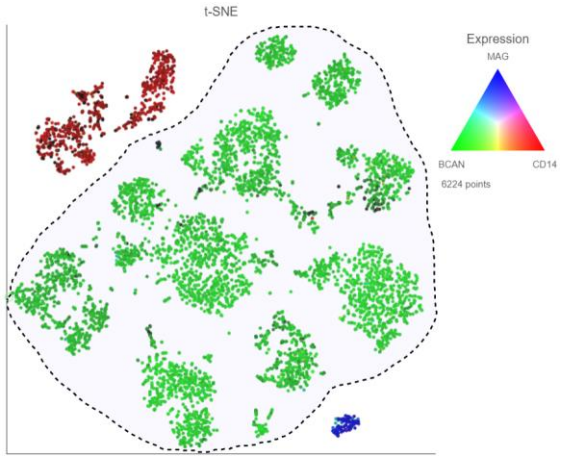
One multi-sample t-SNE plot

- The blue cells are MAG positive, indicating that they are the oligodendrocytes from every sample.
  - Switch to pointer mode by clicking  in the top right corner of the plot
  - Deselect the cells by clicking on any blank space on the plot
  - Switch to lasso mode again by clicking the  icon in the top right of the plot
  - Draw the lasso around the cluster of blue cells and click the circle to close the lasso
  - Click **Classify selection** in Classify
  - Name the classification **Oligodendrocytes**
  - Click **Save**



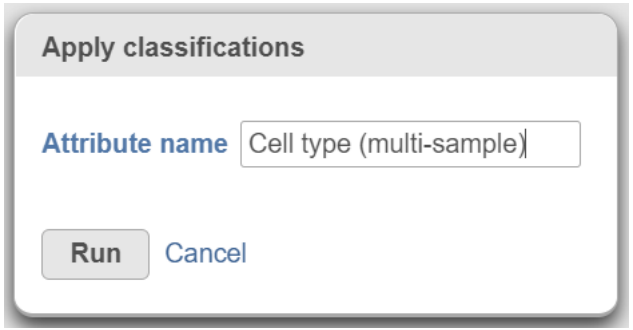
One multi-sample t-SNE plot

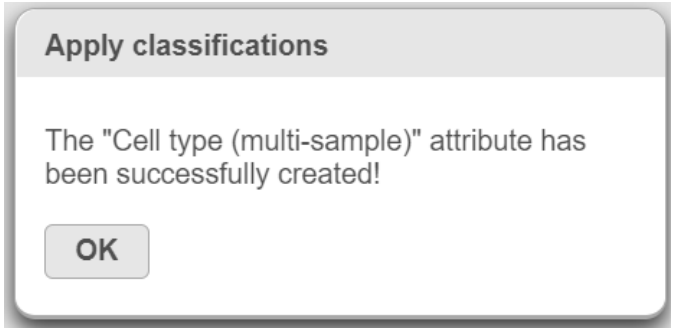
- Finally, we will classify the BCAN expressing cells on the plot as glioma cells from every sample.
  - Switch to pointer mode by clicking  in the top right corner of the plot
  - Deselect the cells by clicking on any blank space on the plot
  - Switch to lasso mode again by clicking the  icon in the top right of the plot
  - Draw the lasso around the cluster of green cells and click the circle to close the lasso
  - Click **Classify selection** in Classify
  - Name the classification **Glioma**
  - Click **Save**



One multi-sample t-SNE plot

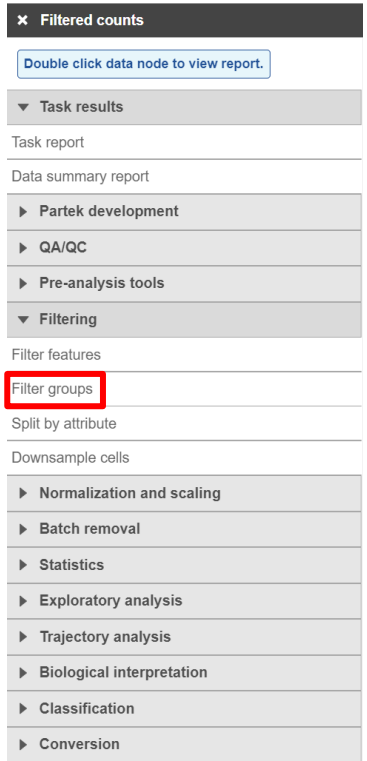
- Click **Apply classifications** in the Classify window
- Name the classification attribute **Cell type (multi-sample)**
- Click **Run**
- Click **OK** on the information box that says an attribute task has been created





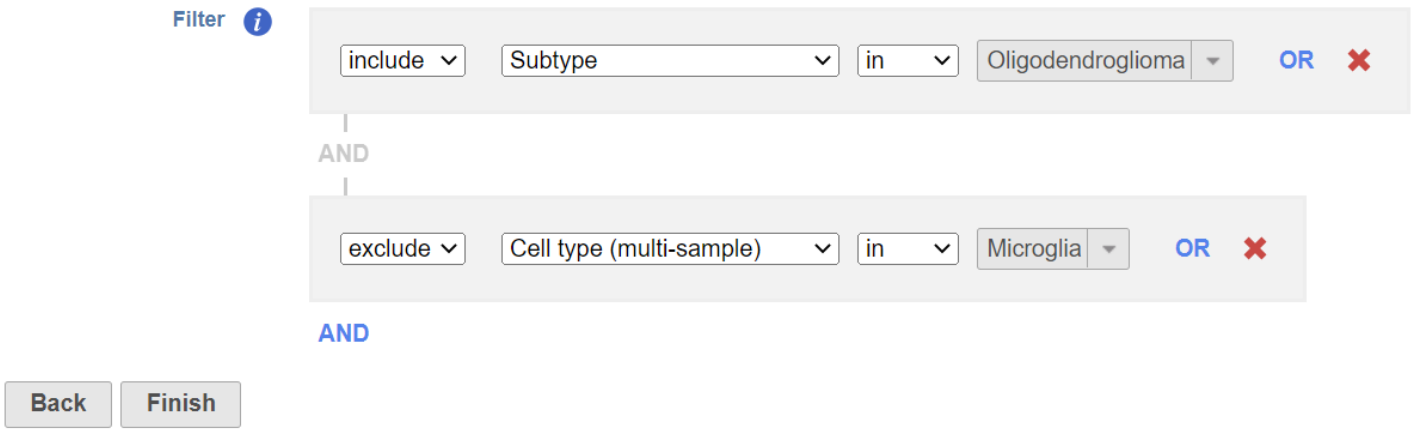
Filter groups

- To analyze only the oligodendroglioma subtype, we can filter the samples
- Click the **Filtered counts** data node
- Expand **Filtering** in the task menu
- Click **Filter groups**



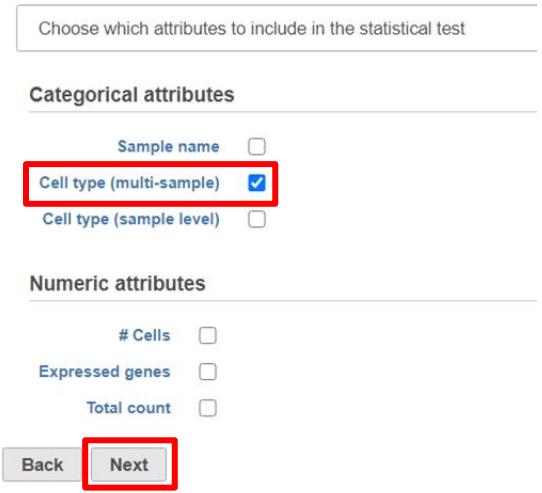
Filter groups

- Set the filter to **Include** samples where **Subtype in Oligodendroglioma**
- Click **AND**
- Set the second filter to **exclude Cell type (Multi-sample) in Microglia**
- Click **Finish** to apply the filter



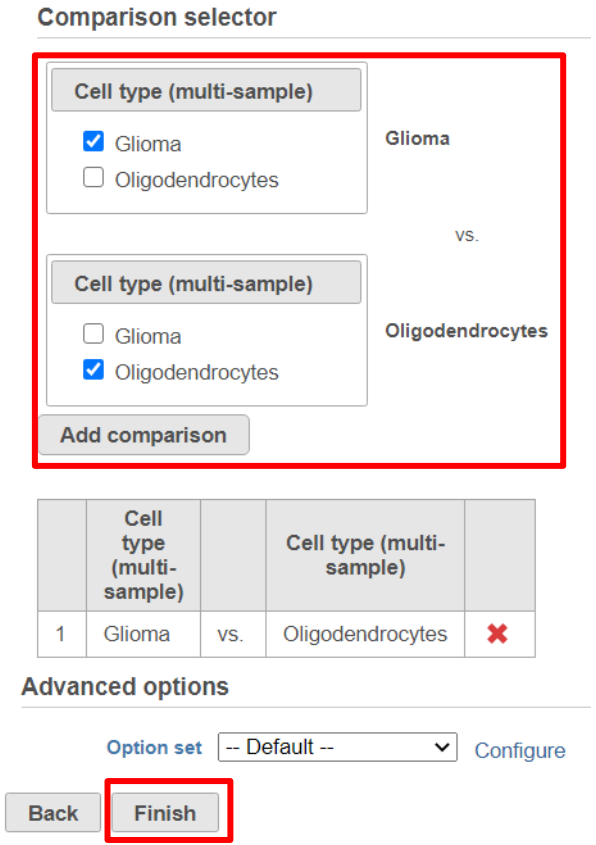
Identify differentially expressed genes

- Click the **new Filtered counts** data node produced by the Filter groups task node
- Expand **Statistics** in the task menu
- Click **Differential Analysis** in the task menu
- Check the **GSA** attribute
- Click **Next**
- Check the **Cell type (multi-sample)** attribute
- Click **Next**



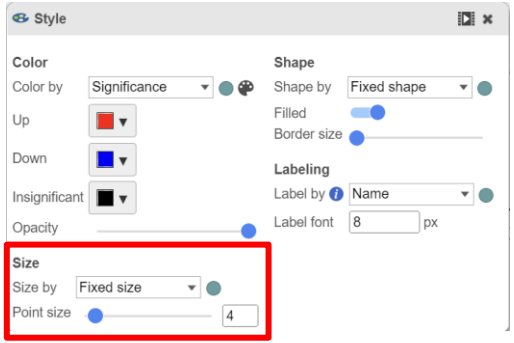
Identify differentially expressed genes

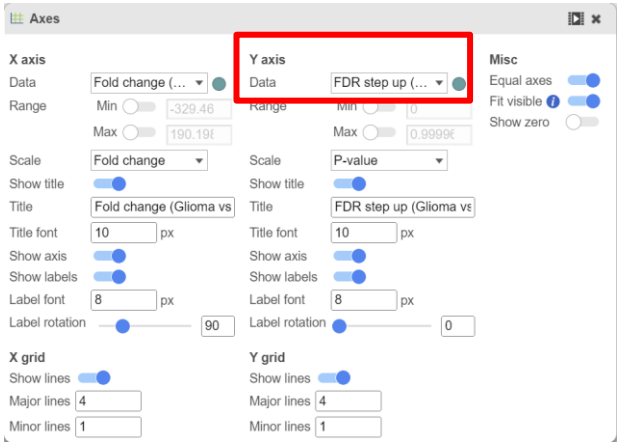
- Click **Glioma** in the top panel
- Click **Oligodendrocytes** in the bottom panel
- Click **Add comparison**
- Click **Finish** to run the GSA

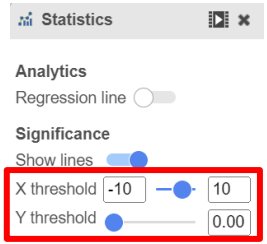


Identify differentially expressed genes

- Double-click the **pink GSA** data node to open the GSA report
- Click  to view the **Volcano plot**
- Click **Style** on the left and increase the **point size to 4** in Size
- Click **Axes** and change the **Y-axis** to **FDR step up (Glioma vs Oligodendrocytes)**
- Click **Statistics** and change **X threshold to -10 and 10** and the **Y threshold to 0.001**

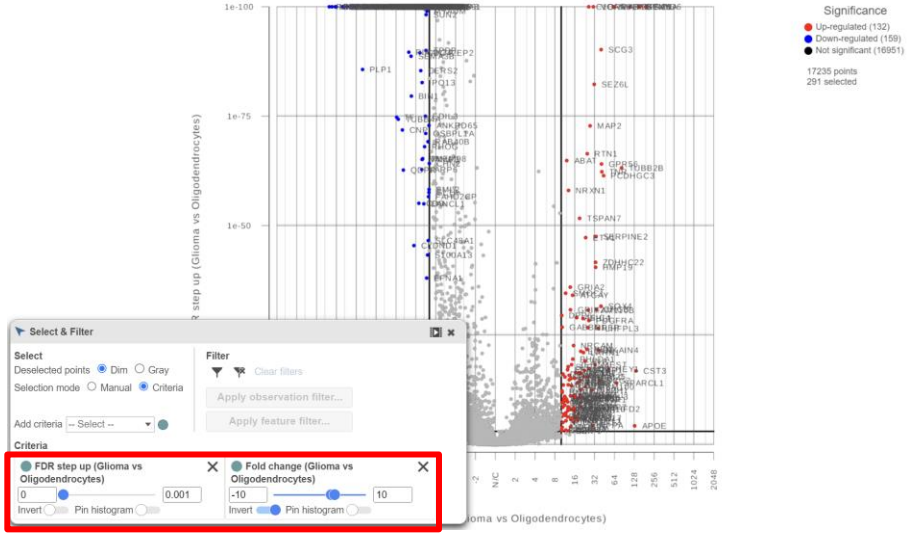






Identify differentially expressed genes

- Click **Select & Filter** on the right, set the **Fold change thresholds to -10 and 10**
- In the **Select & Filter** window, click ✖ to remove the **P-value (Glioma vs Oligodendrocytes)** selection rule. From the drop-down list, add **FDR step up (Glioma vs Oligodendrocytes)** as a selection rule and set the maximum to **0.001**



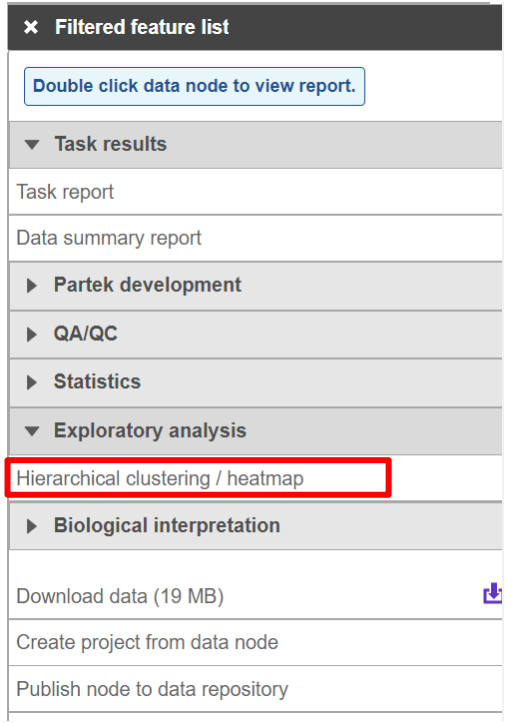
Identify differentially expressed genes

- Click **GSA report** tab in your web browser to return to the GSA report
- Click **FDR step up**
- Set the **FDR step up** filter to **Less than or equal to 0.001**
- Press **Enter**
- Click **Fold change**
- Set the **Fold change** filter to **From -10 to 10**
- Press **Enter**
- Click 

to apply the filter and generate a Filtered Feature list node

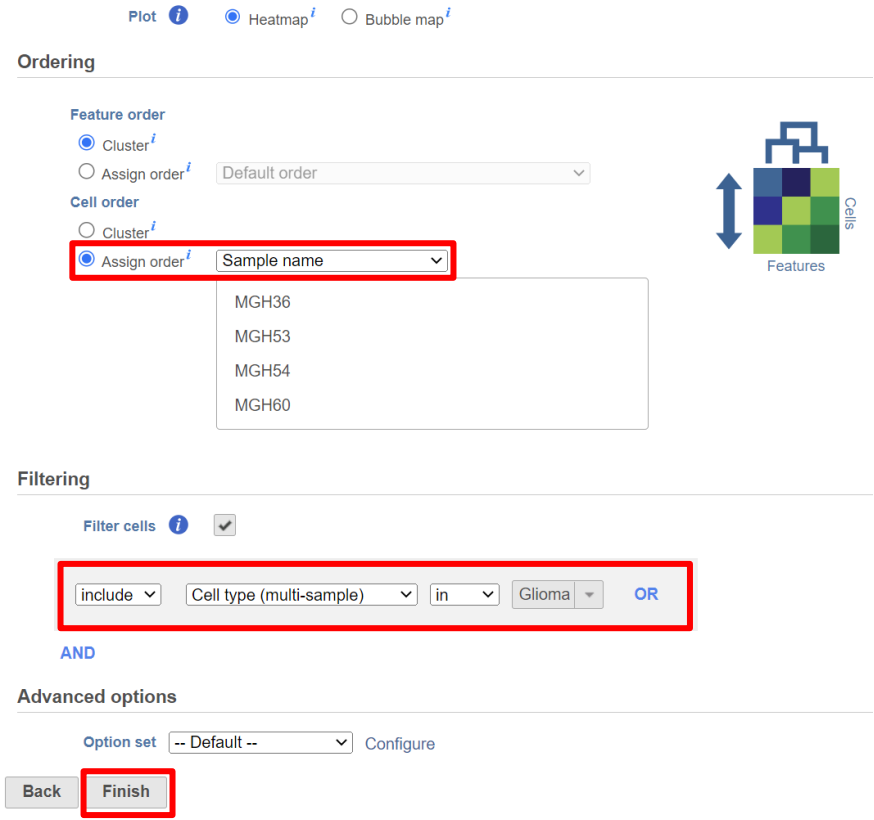
Exploring differentially expressed genes

- Click the **new Filtered feature list** produced by the GSA filter task
- Click **Exploratory analysis** in the task menu
- Click **Hierarchical clustering/heat map**



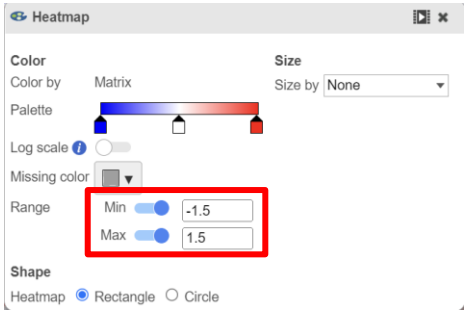
Exploring differentially expressed genes

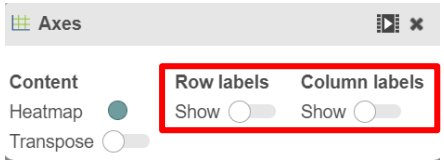
- Set the Assign order of **Cell order** as **Sample name**
- Click **Filter cells** and set the filter to **include Cell type (multi-sample) in Glioma**
- Click **Finish**

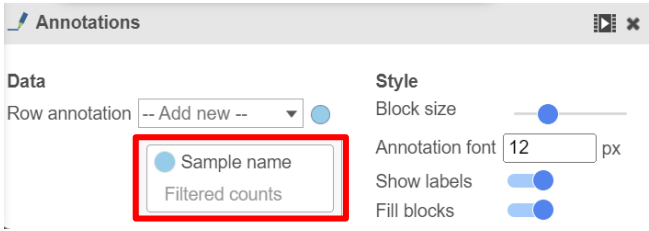


Exploring differentially expressed genes

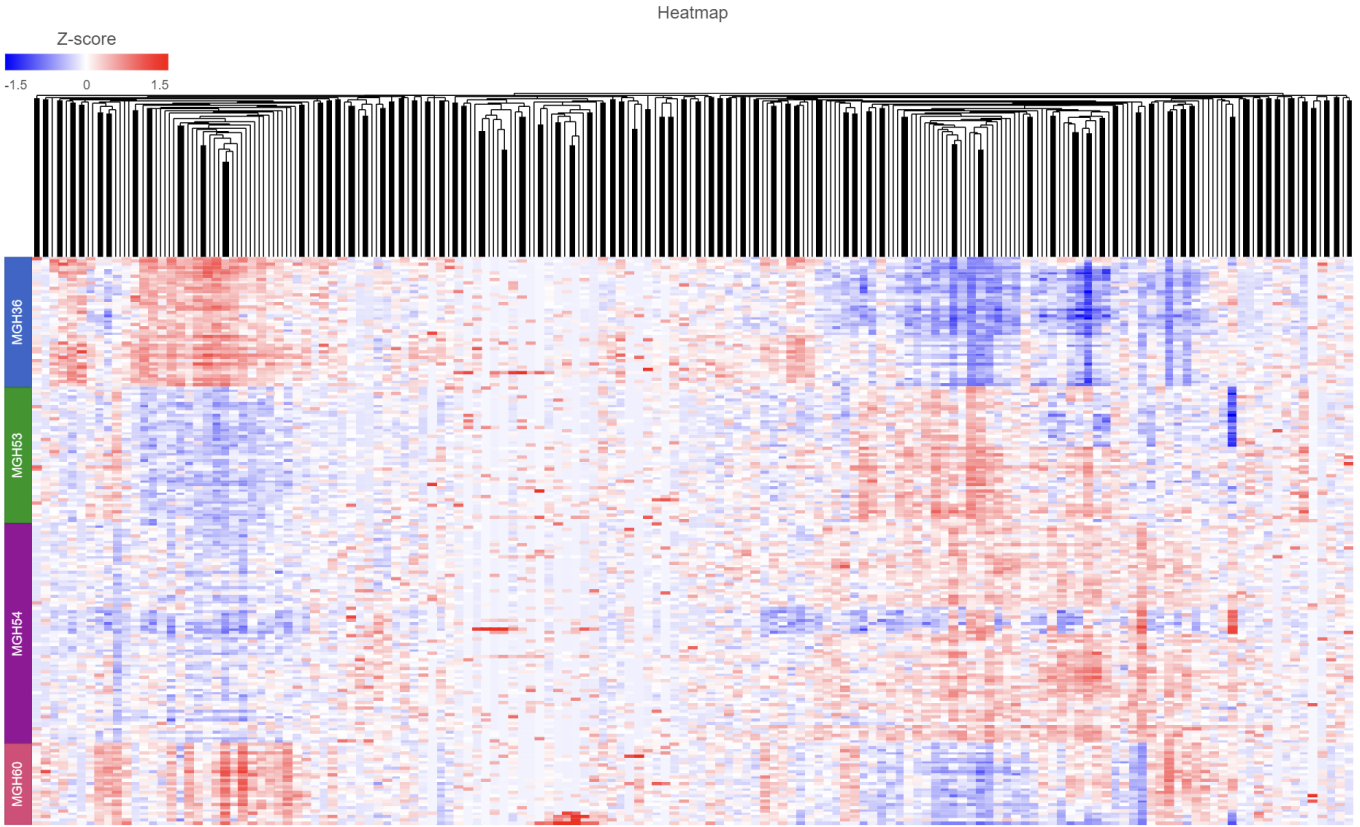
- Double click the **pink Hierarchical clustering** node to open the heat map
- Click **Heatmap** on the left
- Set **Min** to **-1.5** and **Max** to **1.5**
- Click **Axes** and deselect **Show** in both **Row labels** and **Column labels** to hide sample and feature names
- Click **Annotations** on the left and select **Filtered counts** data node
- Select **Sample name** from the **Row annotation** drop-down menu





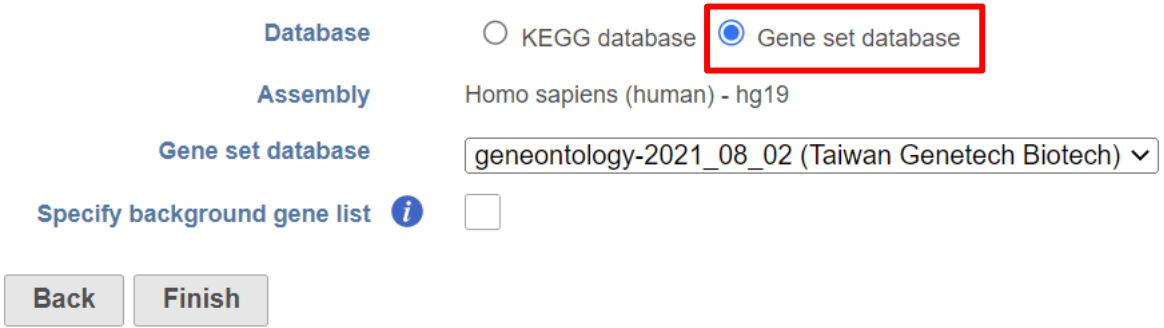


Exploring differentially expressed genes



Exploring differentially expressed genes

- Click **Glioma (multi-sample)** to return to the Analyses tab.
- Click the **new Filtered feature** list node
- Click **Biological interpretation** in the task menu
- Click **Gene set enrichment**
- Click **Gene set database**
- Select the **most recent Gene set database**
- Click **Finish**



Exploring differentially expressed genes

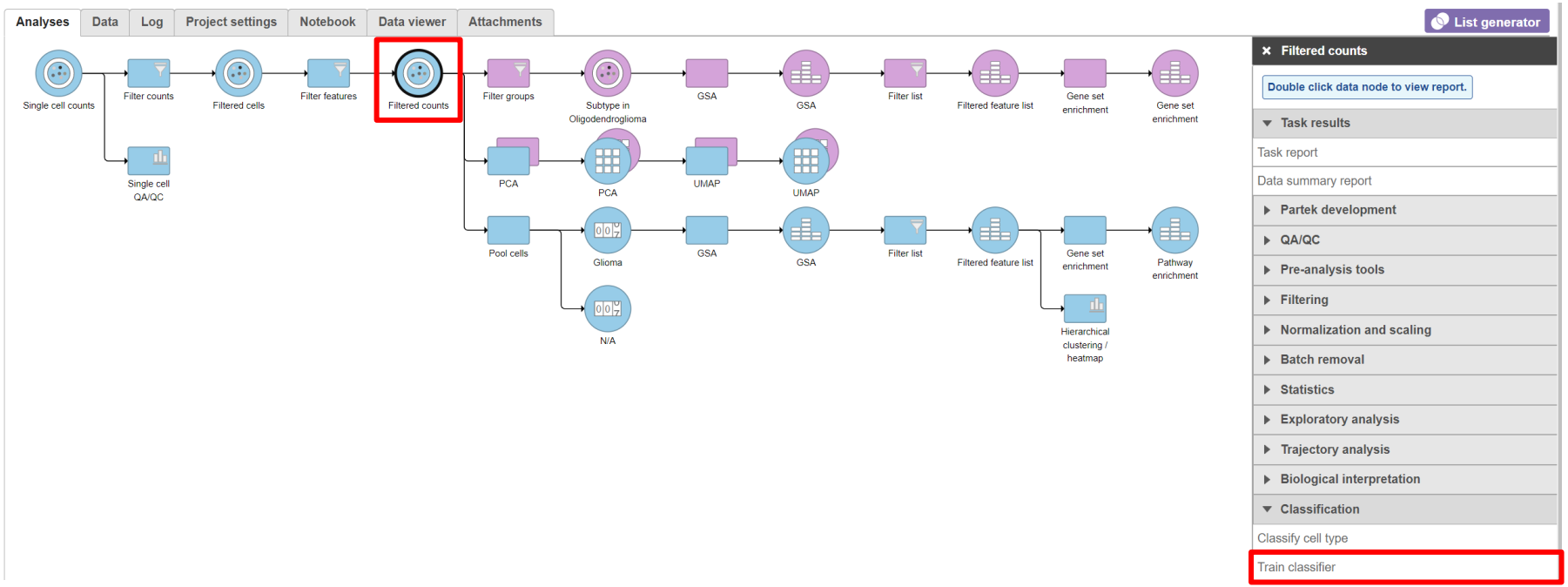
- Double-click the **new Gene set enrichment** task node to open the task report

Top GO terms in the enrichment report include "ensheathment of neurons" and "axon ensheathment" , which corresponds well with the role of oligodendrocytes in creating the myelin sheath that supports and protect axons in the central nervous system.

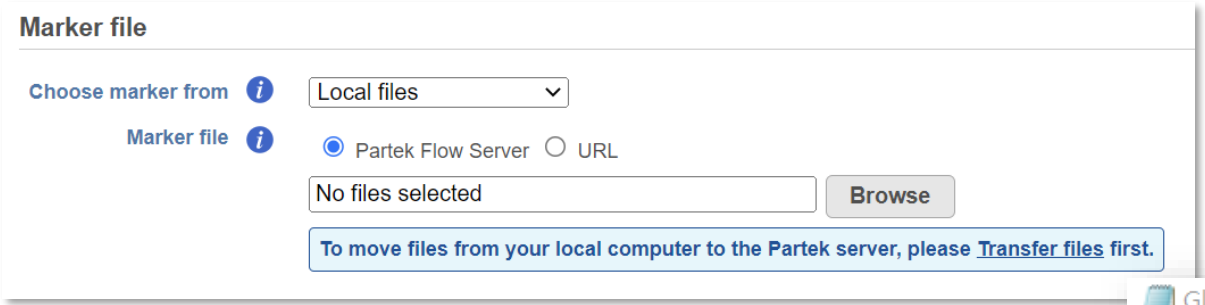
Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
GO:0048856	anatomical structure development	25.12	1.23E-11	94	2,973	
GO:0048731	system development	24.62	2.02E-11	38	661	
GO:0031344	regulation of cell projection organization	23.31	7.56E-11	35	595	
GO:0032502	developmental process	22.87	1.17E-10	124	4,660	
GO:0051960	regulation of nervous system development	22.73	1.35E-10	28	395	
GO:0120035	regulation of plasma membrane bounded cell projection organization	22.61	1.51E-10	34	579	
GO:0010977	negative regulation of neuron projection development	22.30	2.07E-10	16	114	
GO:0014013	regulation of gliogenesis	21.96	2.9E-10	14	82	
GO:0007417	central nervous system development	21.73	3.66E-10	16	119	
GO:0010975	regulation of neuron projection development	21.14	6.56E-10	27	396	
GO:0045685	regulation of glial cell differentiation	20.43	1.34E-9	12	61	
GO:0007155	cell adhesion	20.27	1.57E-9	40	848	
GO:0022610	biological adhesion	20.05	1.97E-9	40	855	
GO:0045202	synapse	20.03	2.01E-9	38	783	
GO:0014015	positive regulation of gliogenesis	19.68	2.84E-9	11	51	
GO:0050767	regulation of neurogenesis	18.77	7.07E-9	23	326	
GO:0051128	regulation of cellular component organization	18.08	1.4E-8	69	2,164	
GO:0031345	negative regulation of cell projection organization	18.00	1.52E-8	16	158	
GO:0032879	regulation of localization	17.85	1.78E-8	75	2,462	
GO:0007272	ensheathment of neurons	17.74	1.98E-8	11	63	
GO:0008366	axon ensheathment	17.74	1.98E-8	11	63	
GO:0009611	response to wounding	17.71	2.03E-8	17	185	
GO:0022010	central nervous system myelination	17.64	2.18E-8	6	7	
GO:0032291	axon ensheathment in central nervous system	17.64	2.18E-8	6	7	
GO:0061024	membrane organization	17.64	2.19E-8	29	537	

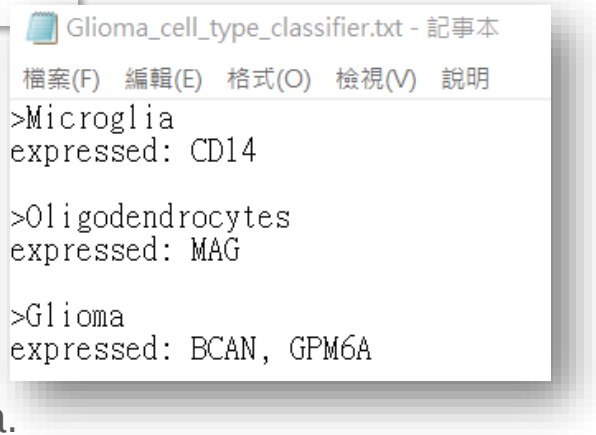
Rows per page 25 (1 of 204) Download

Garnett classifiers in Partek Flow



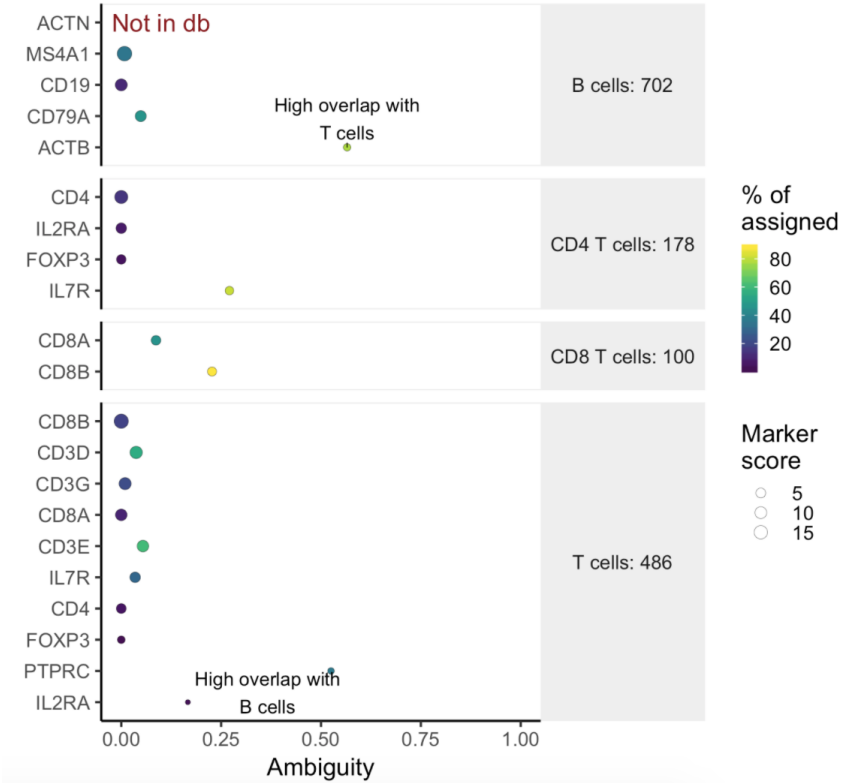
Garnett classifiers in Partek Flow





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

Garnett classifiers in Partek Flow



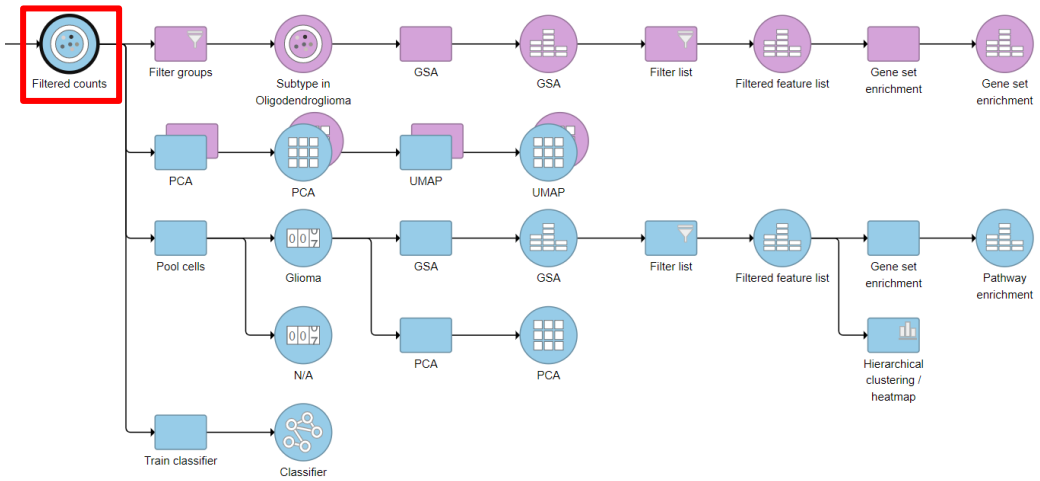
- Double click the **Classifier** data node
- Ambiguity scores are calculated for each of the markers which indicates how many cells receive ambiguous labels when this marker is included

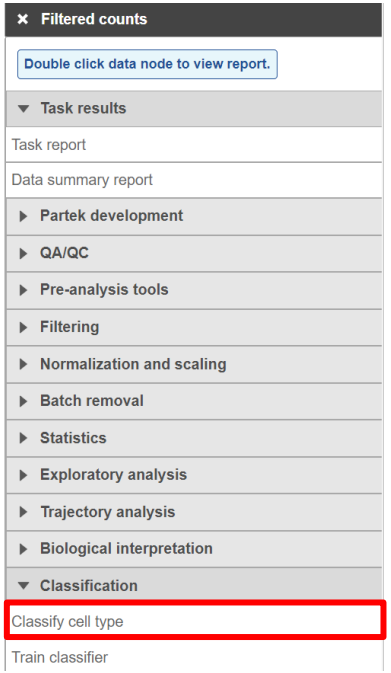
Garnett classifiers in Partek Flow

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

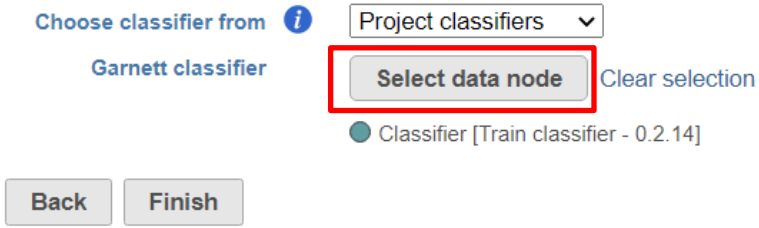
Feature	Glioma	Microglia	Oligodendrocytes	Unknown
(Intercept)	-39.80	9.48	14.21	16.11
BCAN	2.63	-1.00	-0.80	-0.83
GPM6A	2.43	-0.60	-0.96	-0.87
CD14	0.82	1.96	-1.48	-1.30
MAG	0.52	-0.50	2.71	-2.73

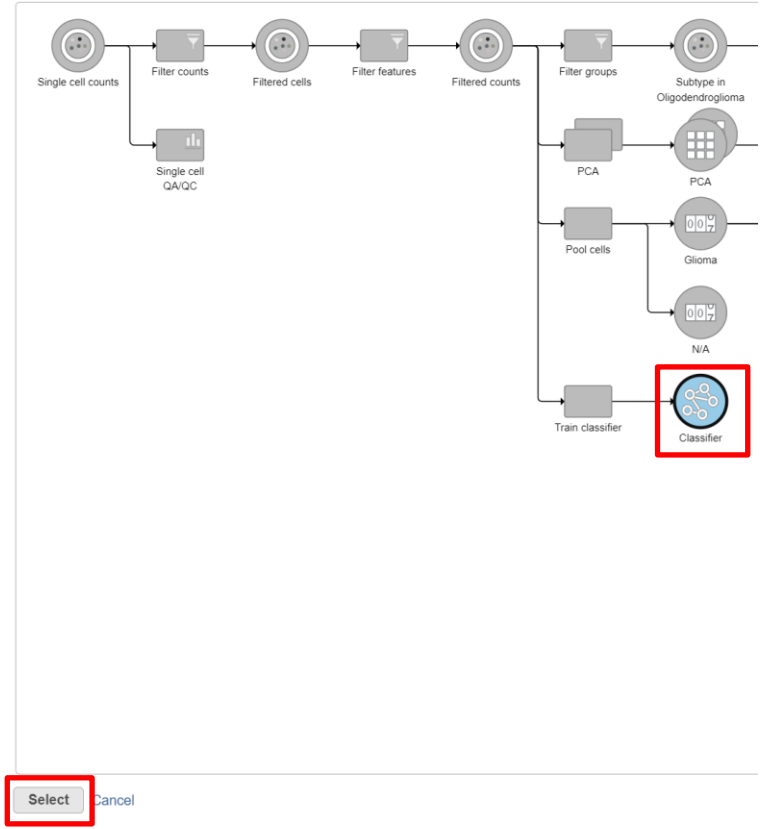
Garnett classifiers in Partek Flow





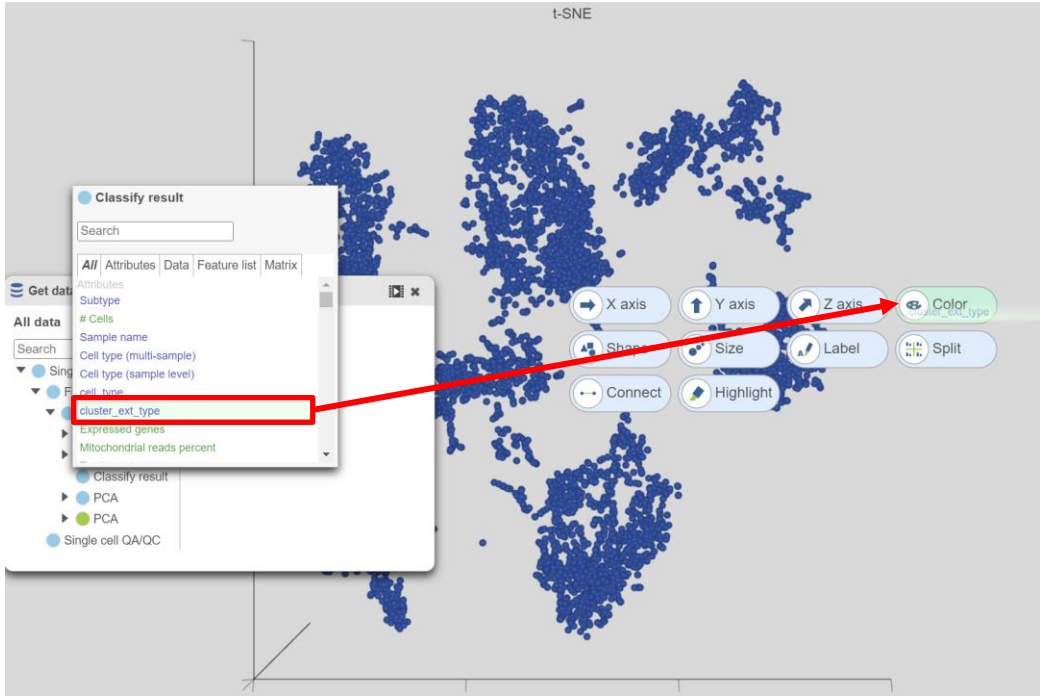
Garnett classifiers in Partek Flow



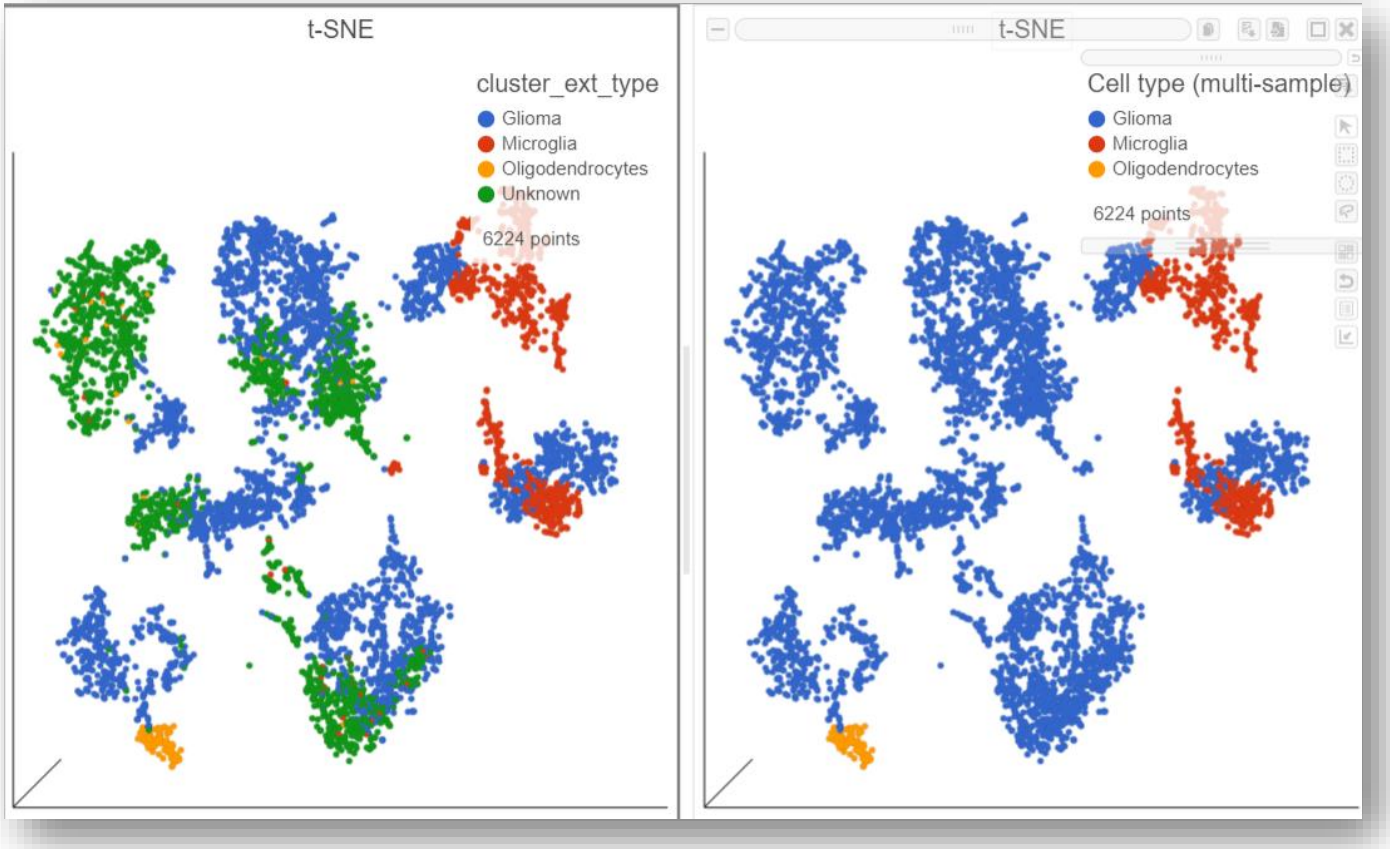


Garnett classifiers in Partek Flow

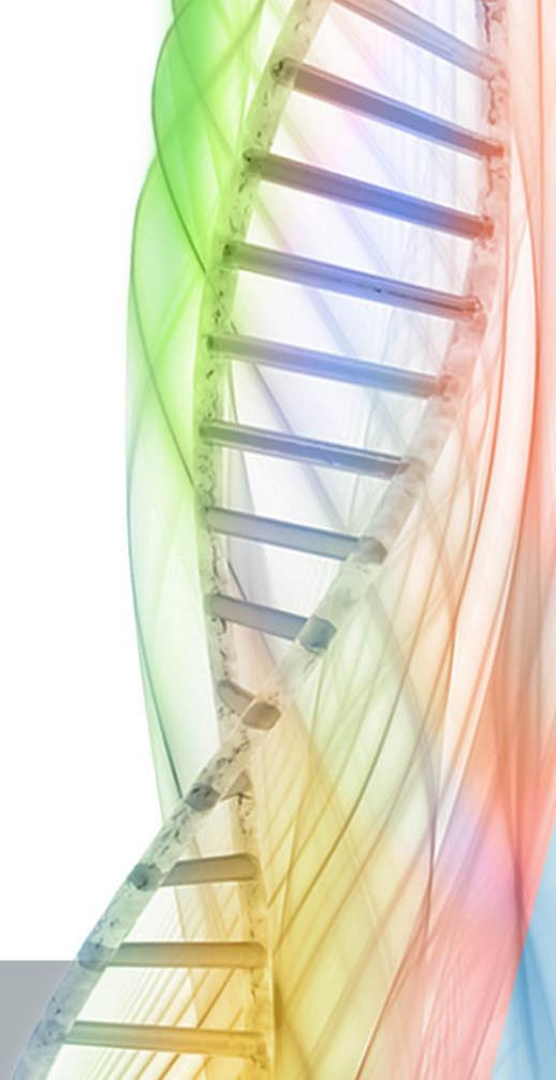
- **“cell_type"** is the cell type assignments directly from Garnett model.
- **"cluster_ext_type"** is the cell type that's determined by expanding cell type assignments to nearby cells using Louvain clustering.



Garnett classifiers in Partek Flow

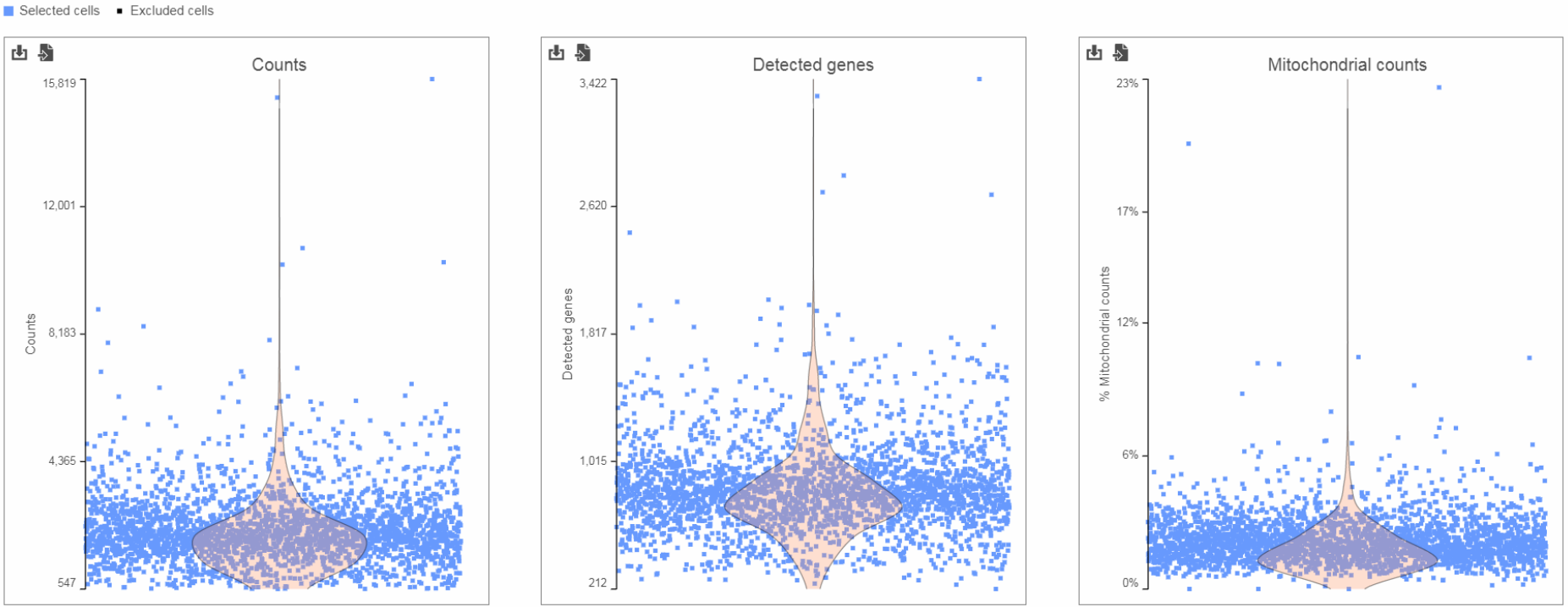


Appendix



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



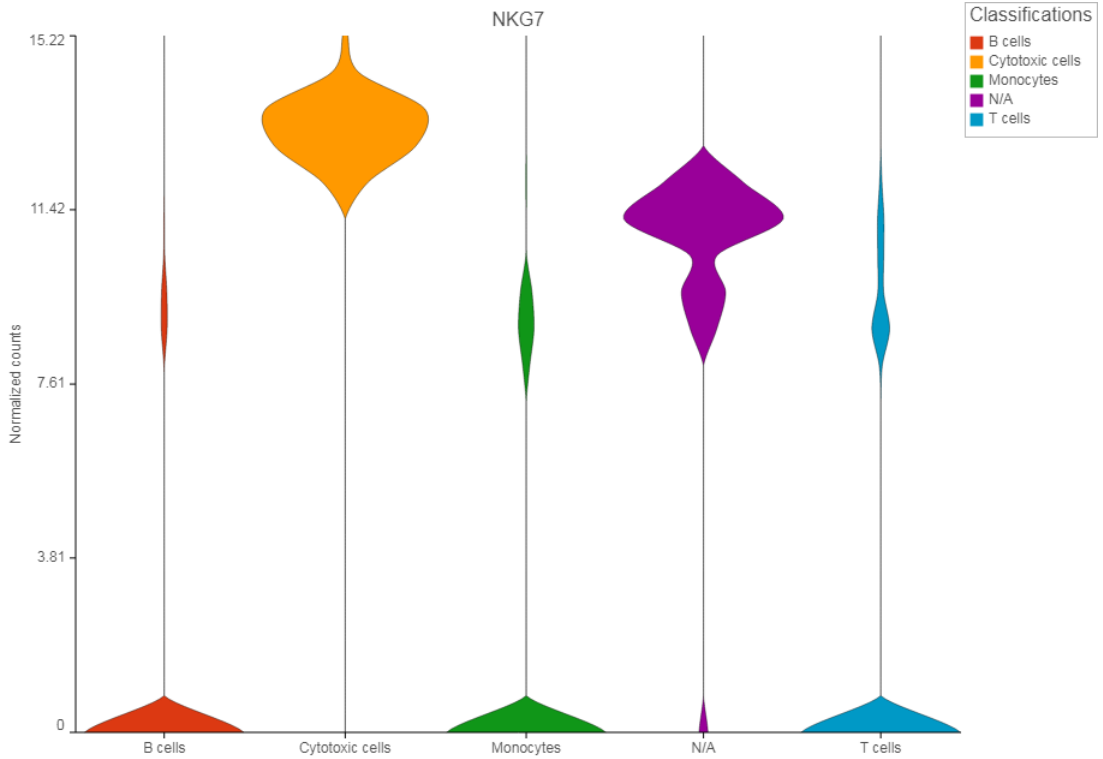
Copyright © 2019 Partek Incorporated. All rights reserved.

均泰生物科技有限公司
Genetech Biotech Co., Ltd

Partek®

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

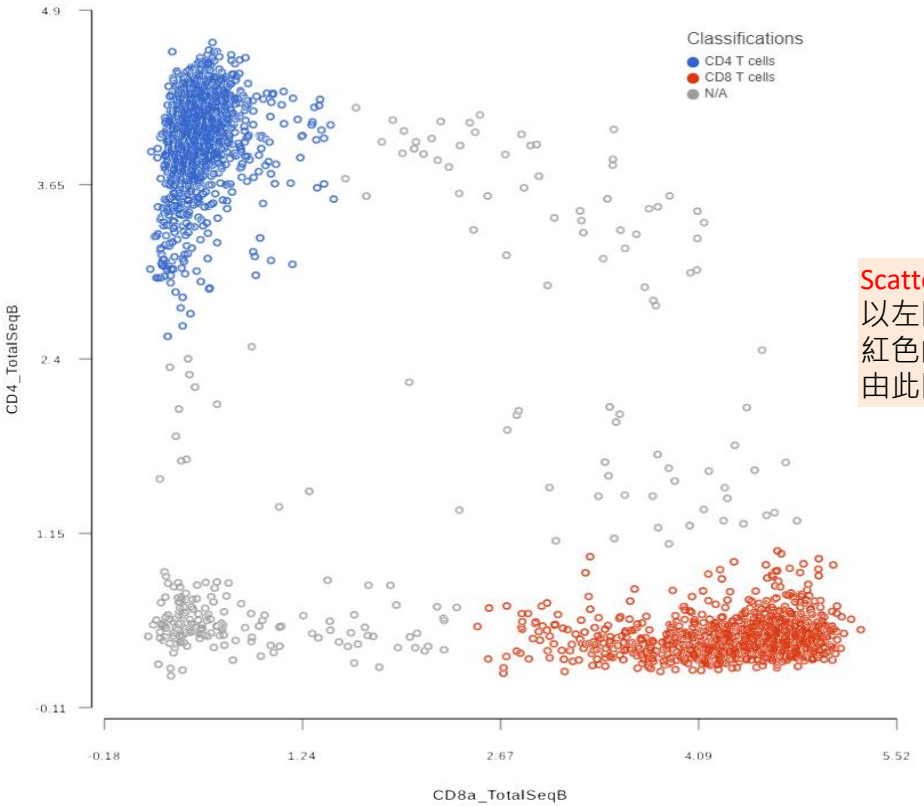


Copyright © 2019 Partek Incorporated. All rights reserved.

均泰生物科技有限公司
Genetech Biotech Co., Ltd

Partek®

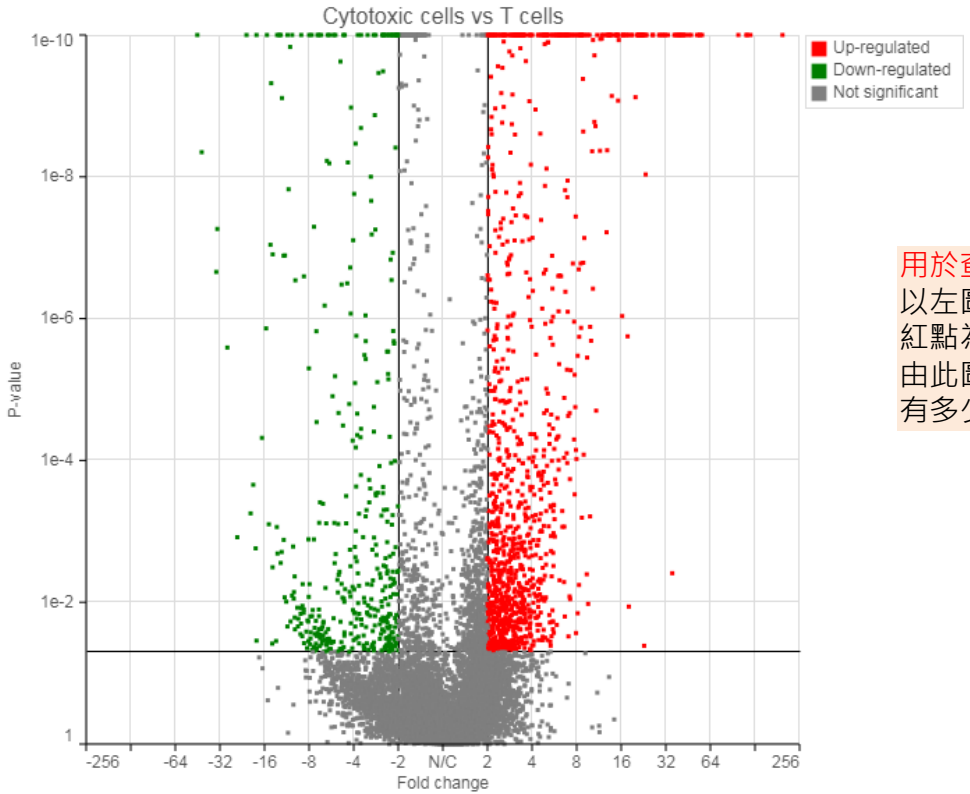
Scatter Plot



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

Copyright © 2019 Partek Incorporated. All rights reserved.

Volcano Plot

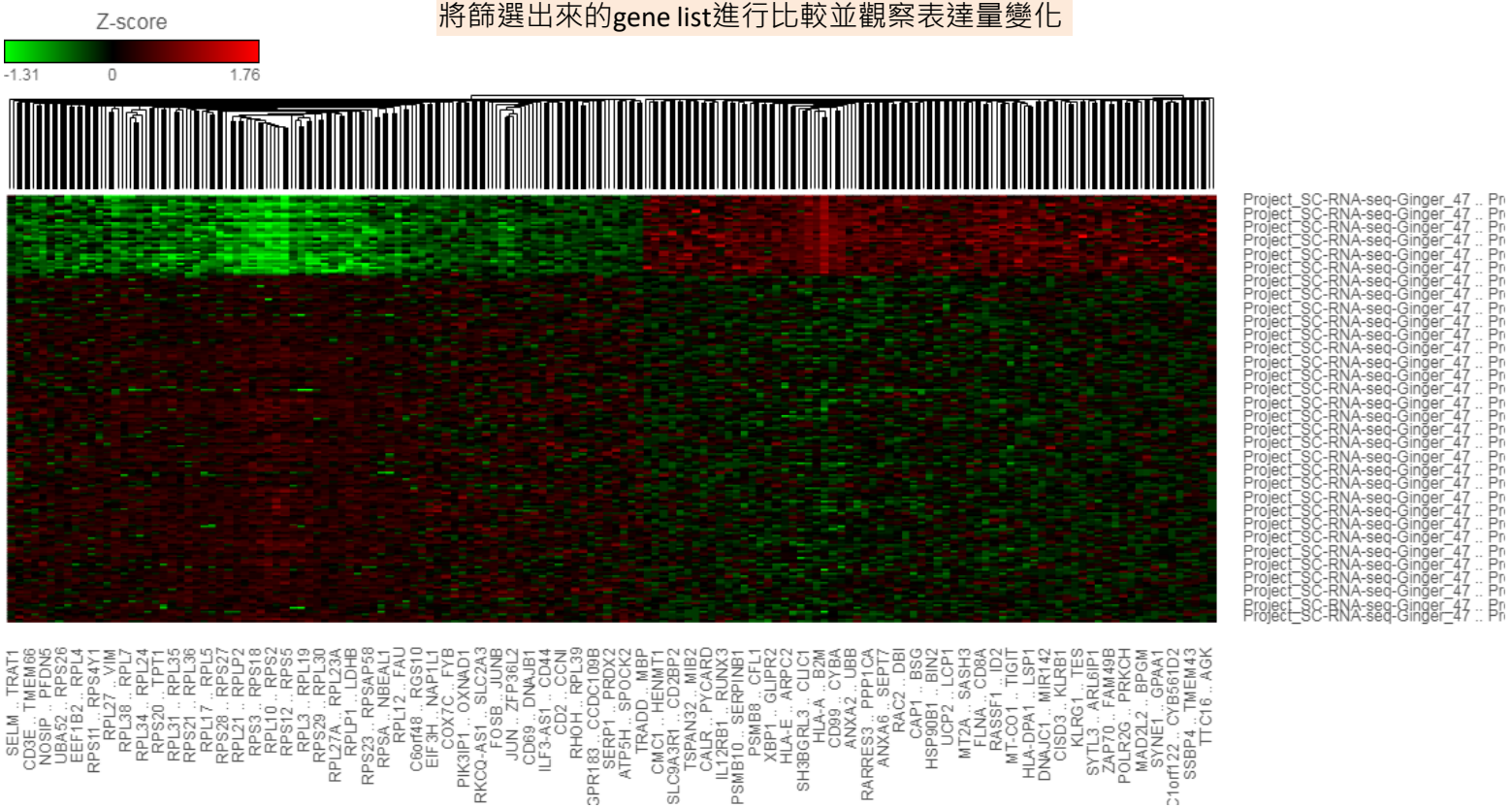


用於查看特定細胞群中高表達基因及低表達基因的數量
以左圖說明，X軸為Fold change，Y軸為P-value；
紅點為Up-regulated gene，綠點為Down-regulated gene
由此圖可看出cytotoxic cells 和 T-cells 這兩個種類的細胞群相比之後，
有多少up-regulated, down-regulated 及 un-change 的基因

Copyright © 2019 Partek Incorporated. All rights reserved.

Heat Map

將篩選出來的gene list進行比較並觀察表達量變化

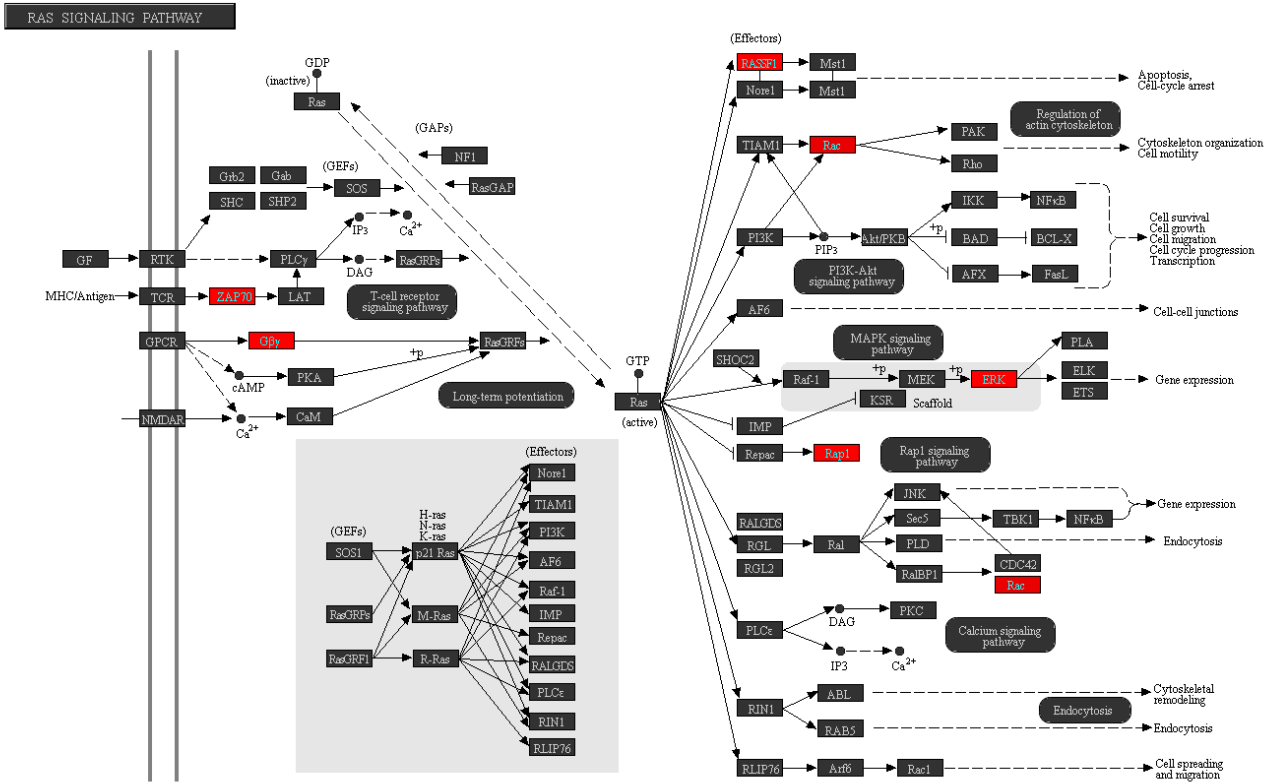


Copyright © 2019 Partek Incorporated. All rights reserved.

均泰生物科技有限公司 Genetech Biotech Co., Ltd

Partek®

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

Copyright © 2019 Partek Incorporated. All rights reserved.

均泰生物科技有限公司
Genetech Biotech Co., Ltd

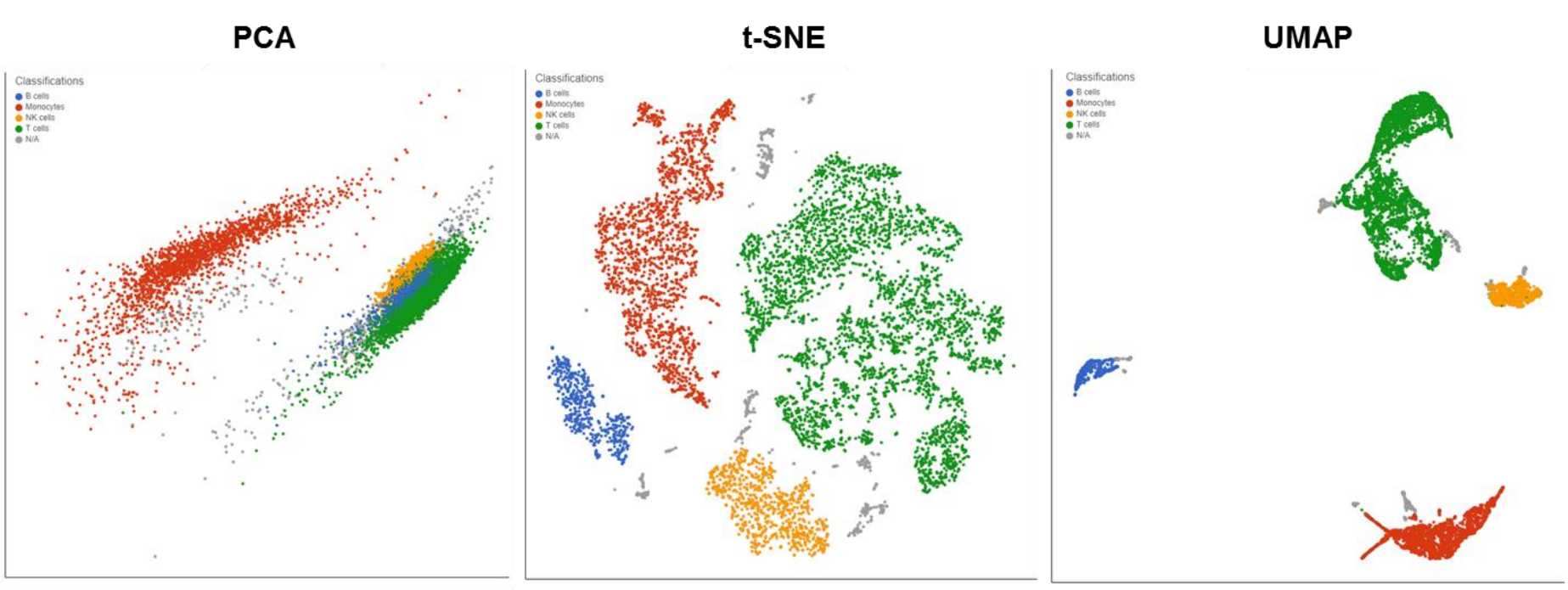
Partek®

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

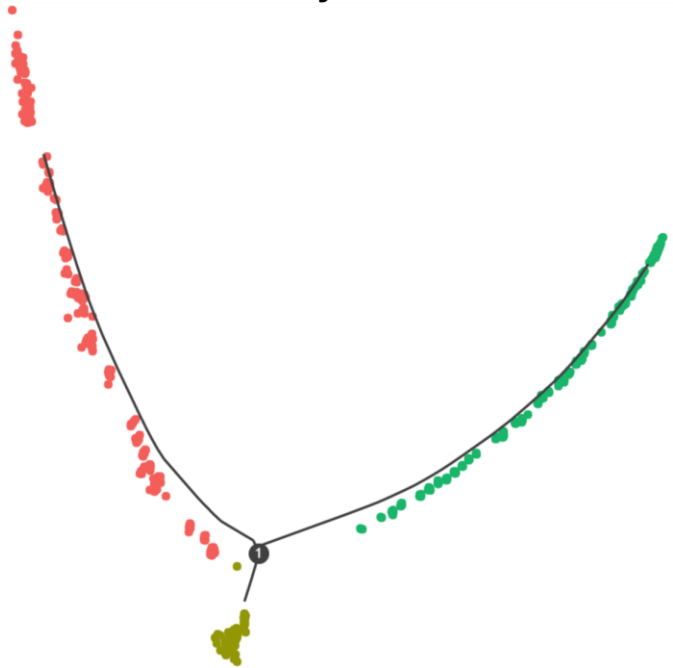
Partek Flow 提供2D及3D的呈現方式，讓使用者更有效了解樣品中不同細胞的相關性



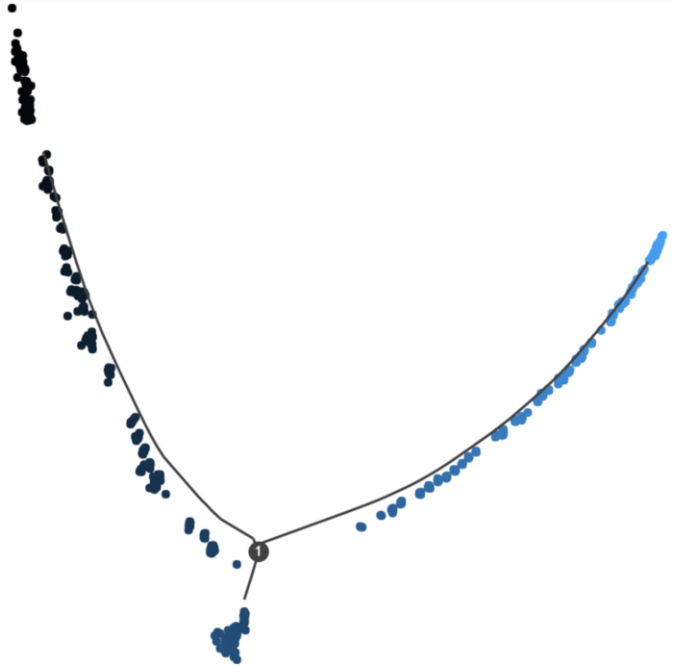
Run Trajectory Analysis with Monocle

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



Copyright © 2019 Partek Incorporated. All rights reserved.