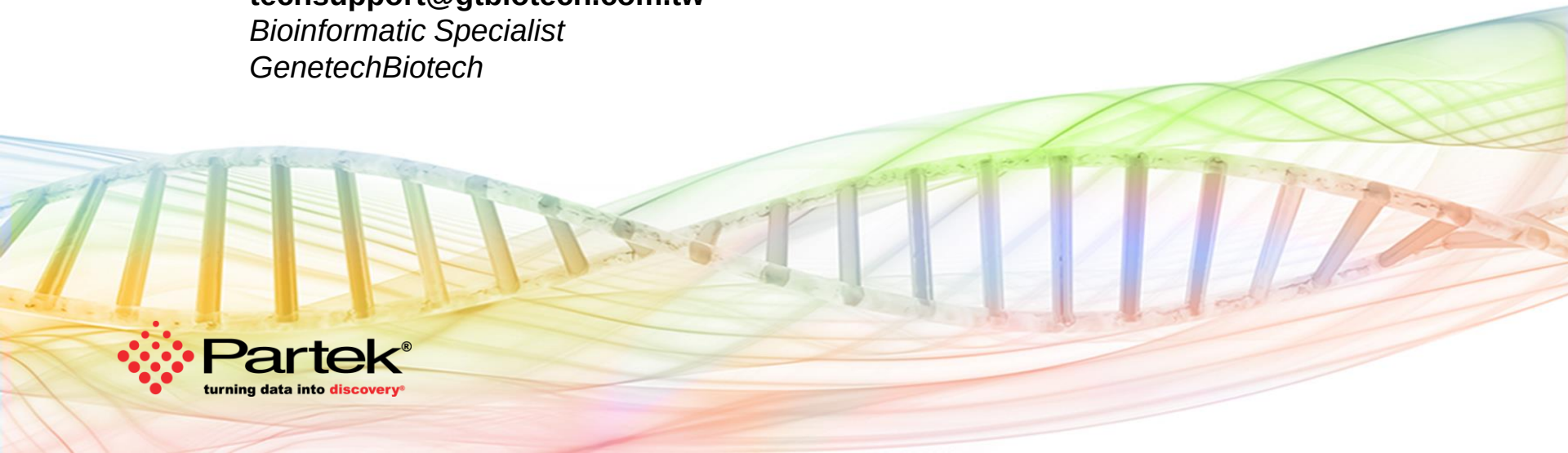
From Raw Data to Pathways: Easy Genomics Analysis with Partek Flow

Tanya Chiu
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

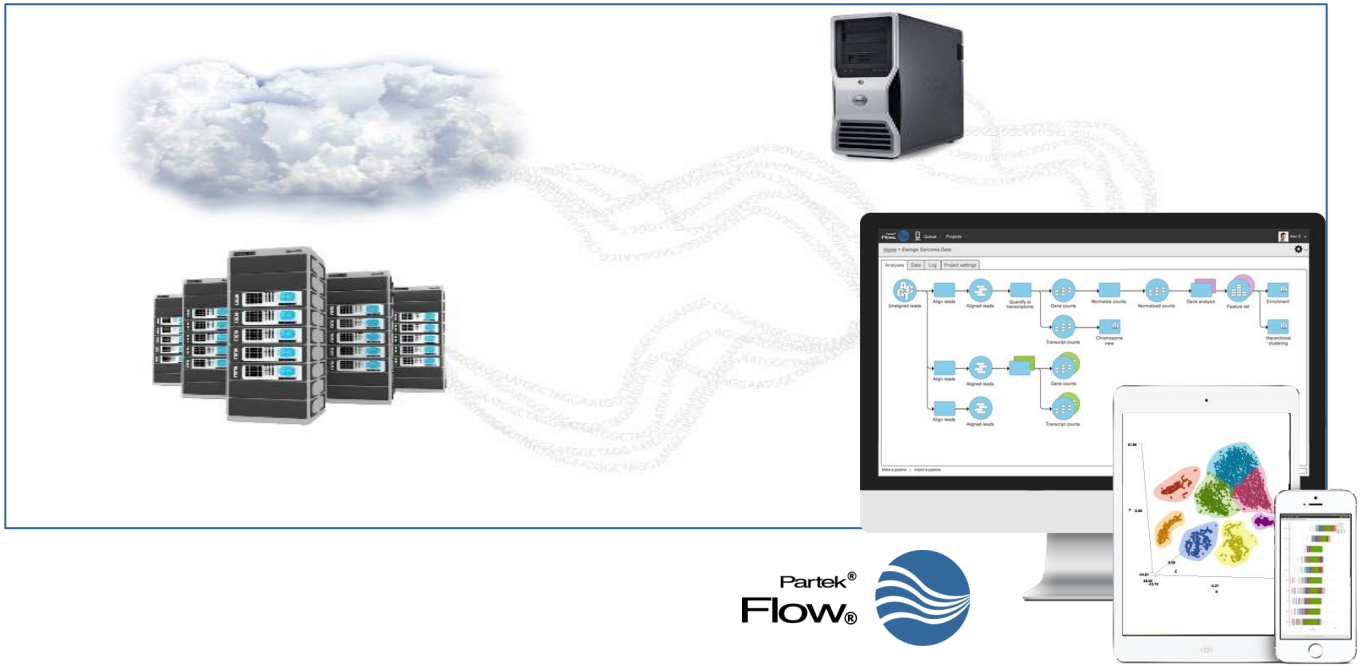
40

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.

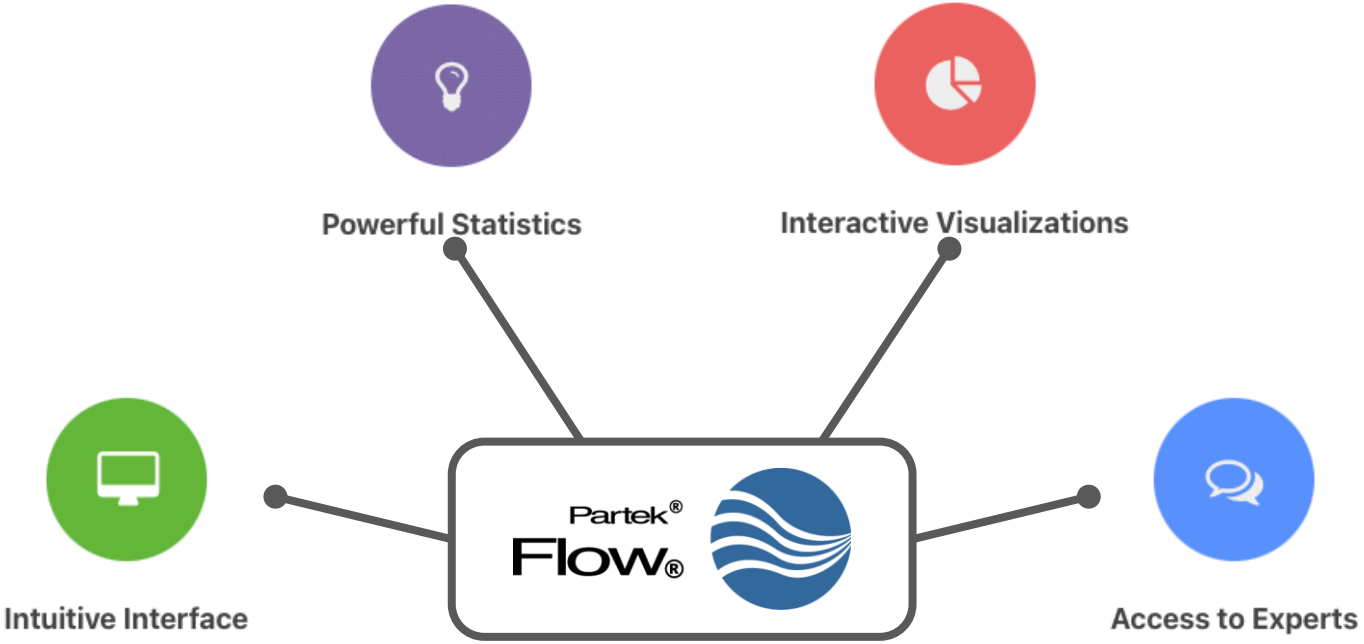
Hardware Requirements

- **Linux**: Ubuntu® 18.04, Redhat® 8, CentOS® 8 or later versions of these distributions
- 64-bit 2GHz quad-core processor
- 48GB of RAM
- > 2TB of storage available for data
- > 100GB on the root partition
- A broadband internet connection

Copyright © 2019 Partek Incorporated. All rights reserved.

Comprehensive Bioinformatics Software

User Friendly and Powerful Data Analysis Tool



Copyright © 2019 Partek Incorporated. All rights reserved.

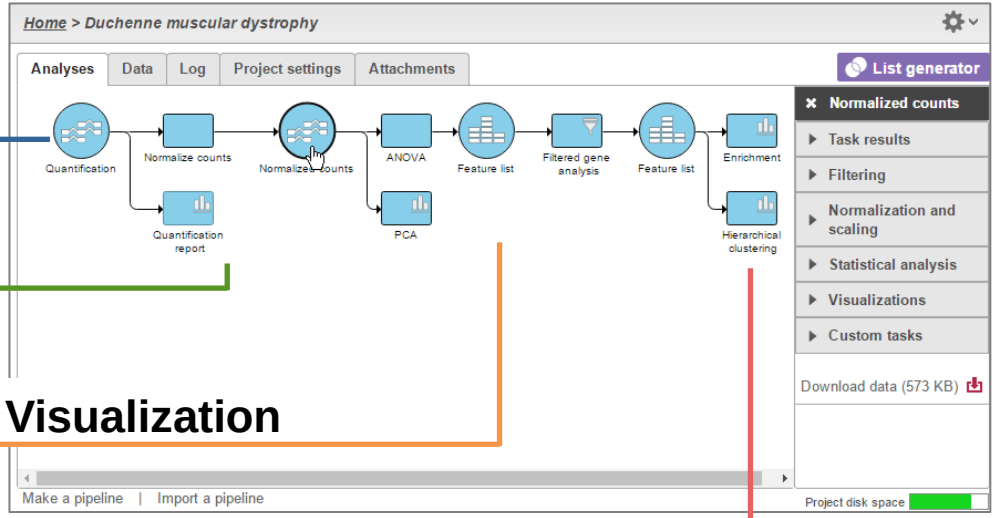
Visual Analysis Process

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

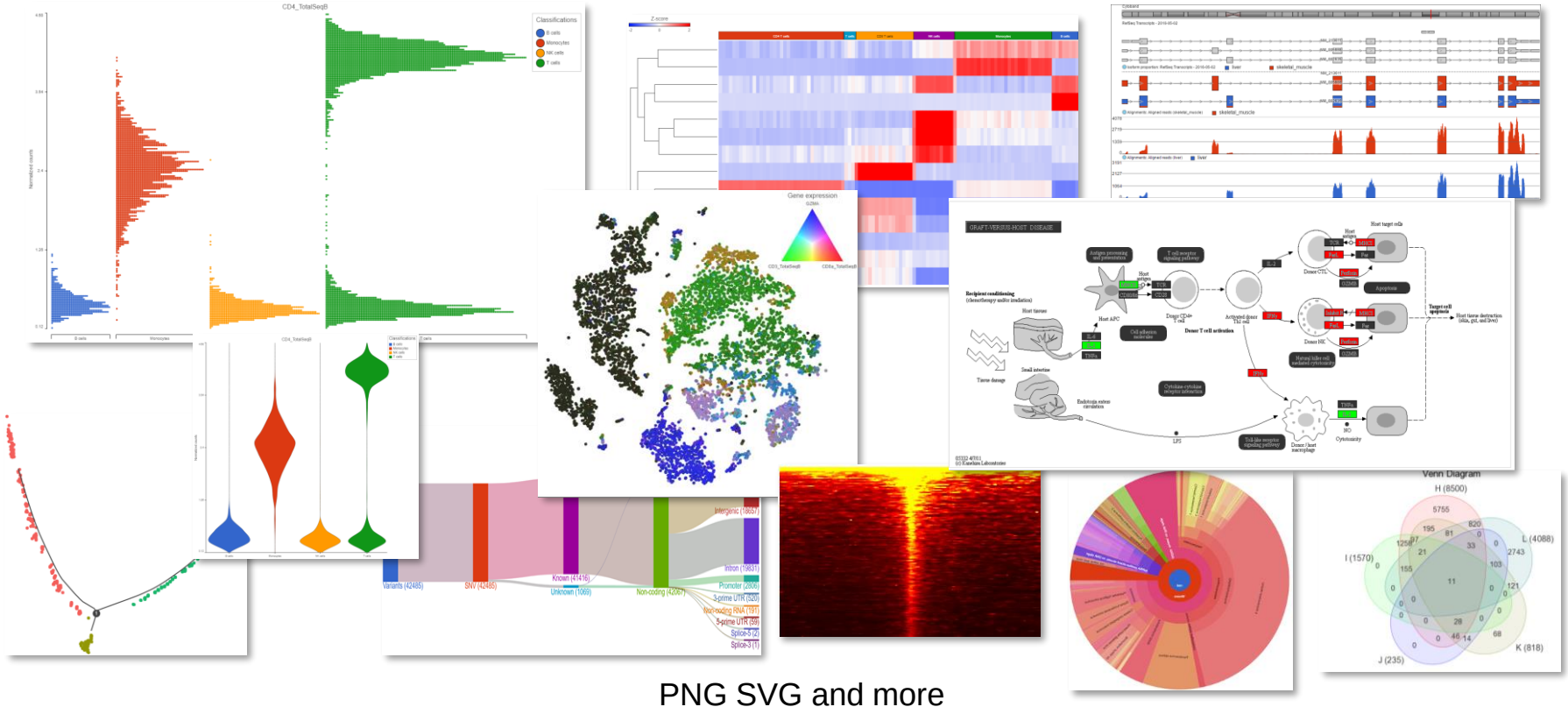
Variant annotation
SnpEff
dbSNP
VEP
Custom databases

Peak calling
MACS2
TSS plot
Motif detection

Quantification
Partek E/M
HTSeq
Cufflinks

Copyright © 2019 Partek Incorporated. All rights reserved.

Compelling and Publishable Visualizations

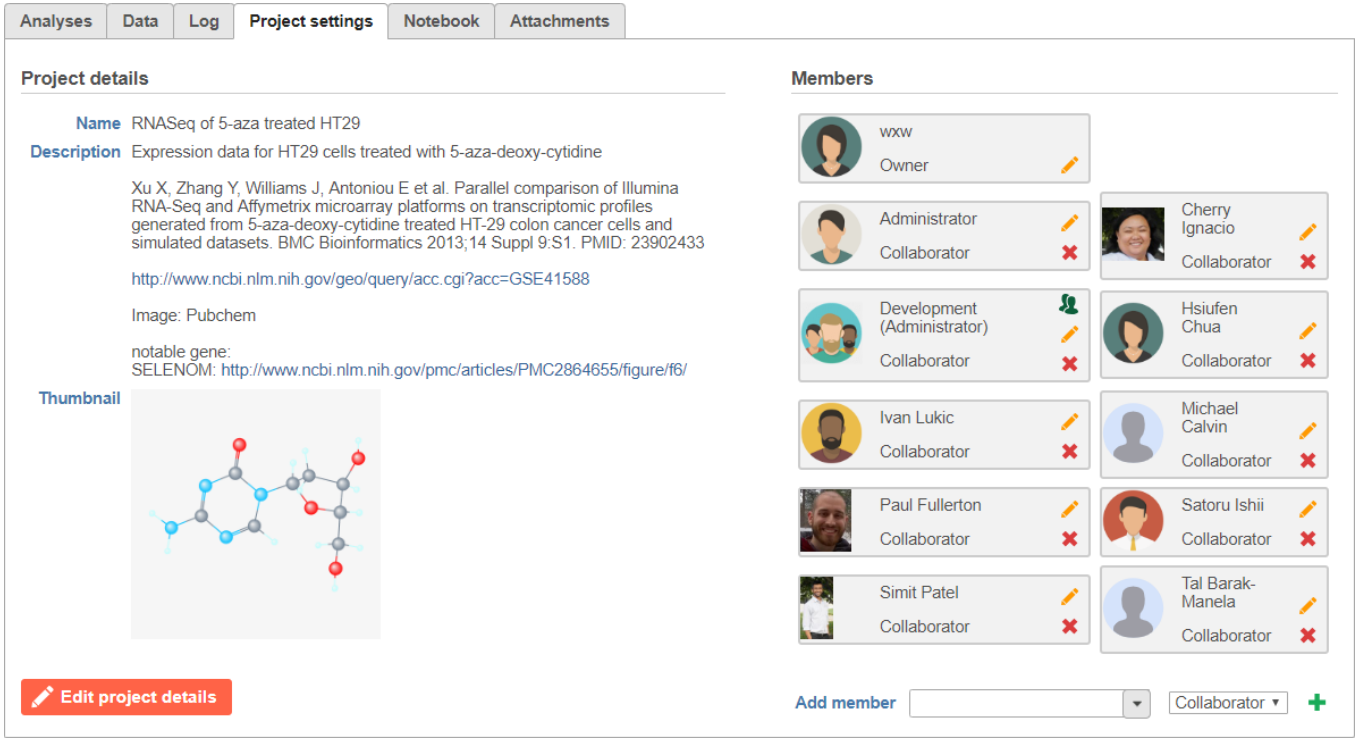


PNG SVG and more

Copyright © 2019 Partek Incorporated. All rights reserved.

Partek®

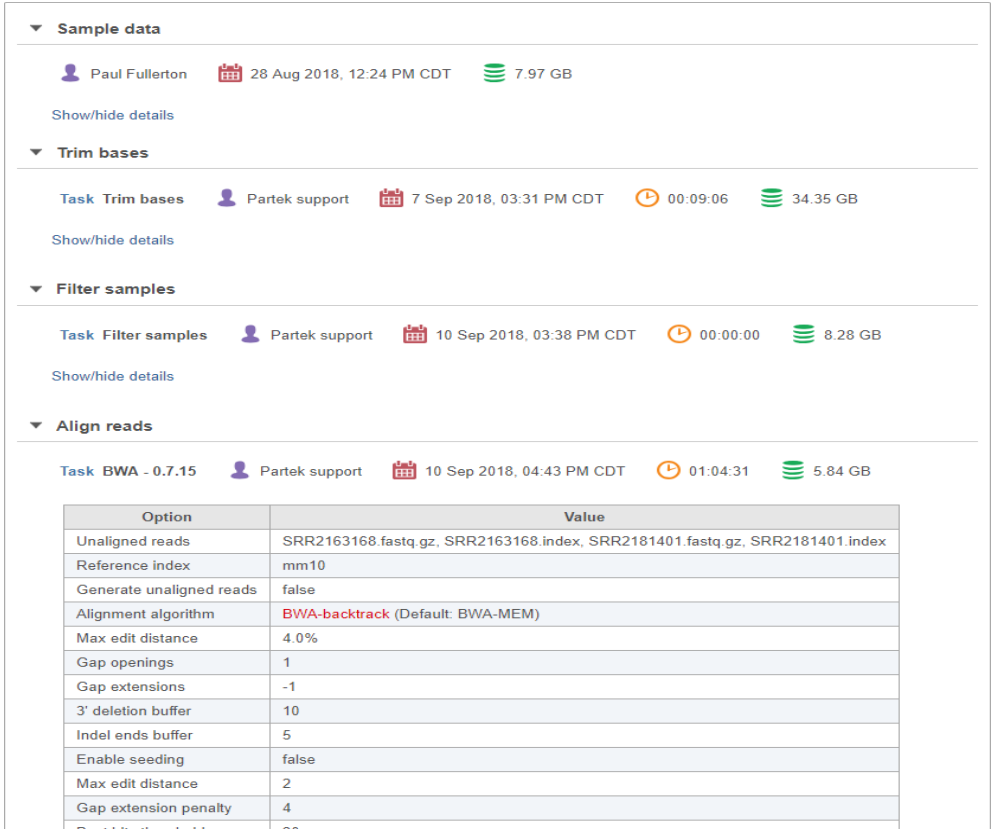
Collaborate and Share



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

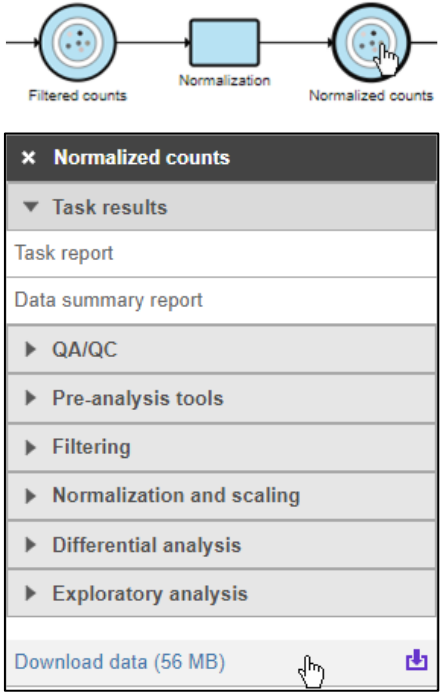
Task **BWA - 0.7.15** | Partek support | 10 Sep 2018, 04:43 PM CDT | 01:04:31 | 5.84 GB

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

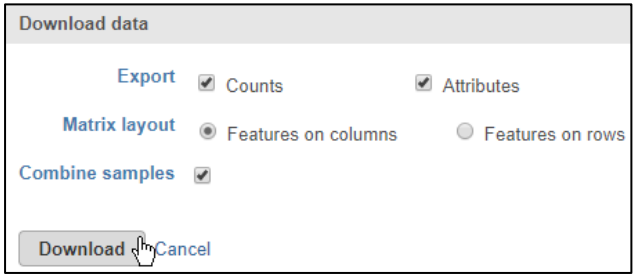
Copyright © 2019 Partek Incorporated. All rights reserved.

Export Data

Choose Any Data

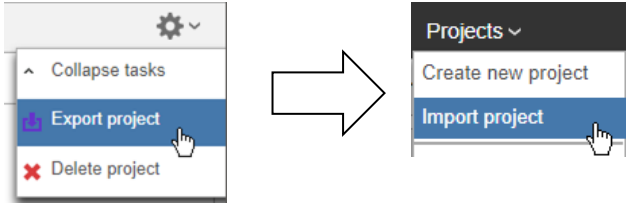


Download in Industry Standard Formats



FASTQ, BAM, TXT, and more

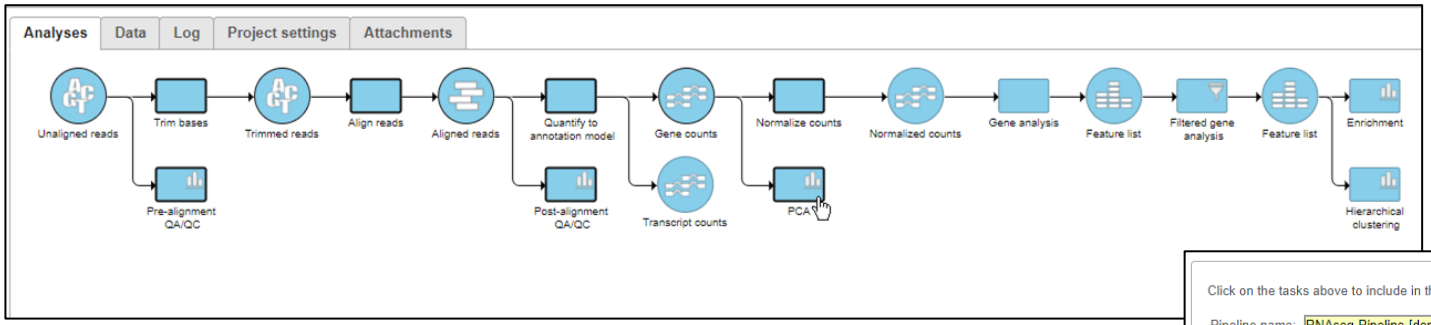
Export and Import Analysis Projects

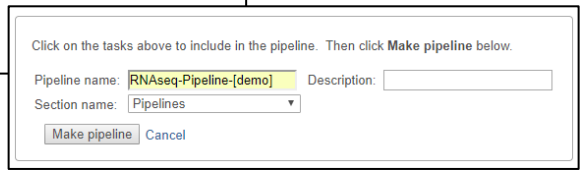


Copyright © 2019 Partek Incorporated. All rights reserved.

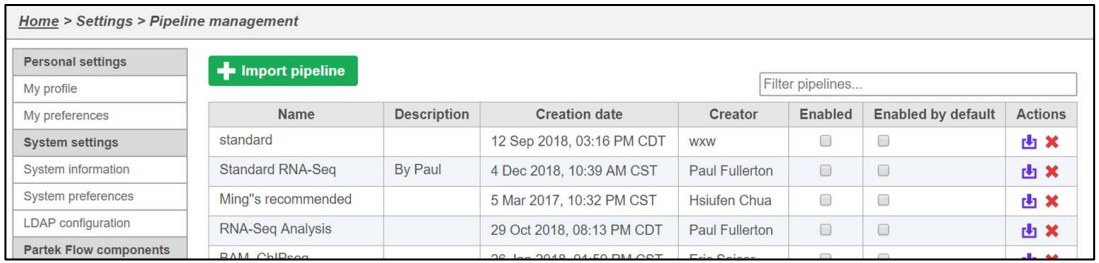
Build, Reuse, and Share Analysis Pipelines

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

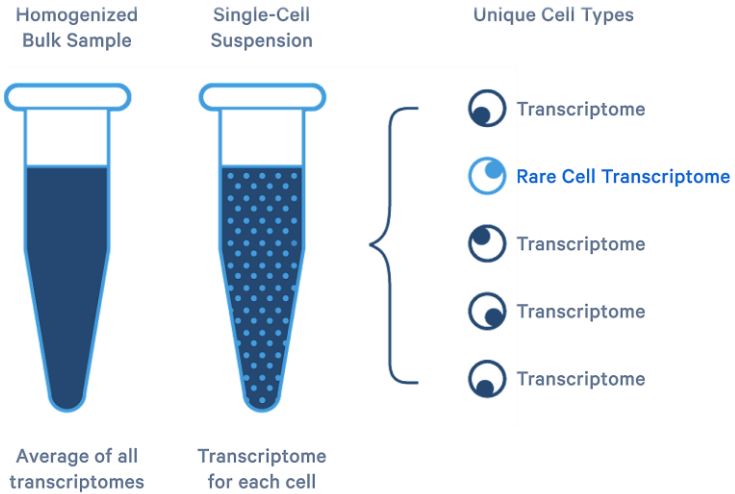


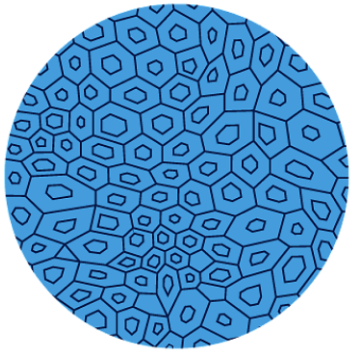
Microarray | Next Generation Sequencing | qPCR

Copyright © 2019 Partek Incorporated. All rights reserved.

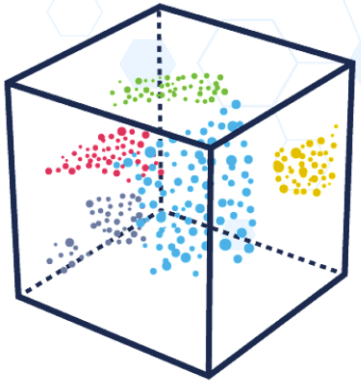
Partek Flow
scRNA-Seq Analysis

Introduction of Single-cell Analysis





Tissue Specimen with a spatial relationship between cells.

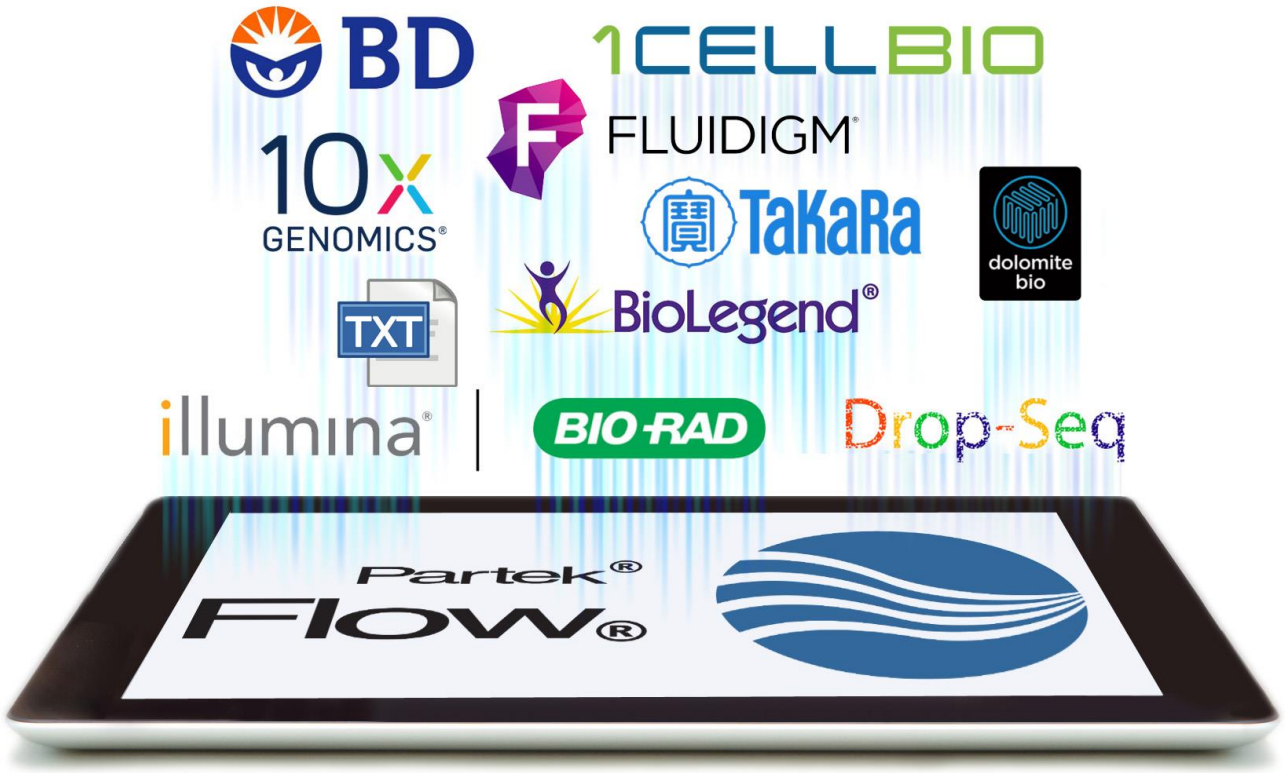


Relationship between cells by similarity of gene expression.

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

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

Supports All Major Single Cell Platforms

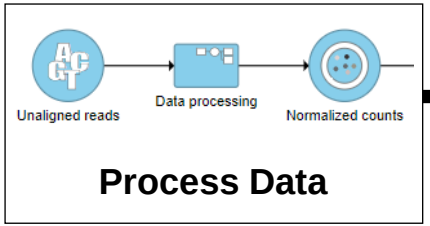


Support for Wide Variety of Single Cell Technologies

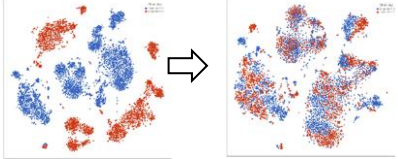
- ✓ Single Cell RNA-Seq
- ✓ Whole Transcriptome Single Cell RNA-Seq
- ✓ Gene & Protein Expression
- ✓ ECCITE-Seq
- ✓ Spatial Transcriptomics
- ✓ Trajectory analysis

Data Processing and Analysis, All in One Place

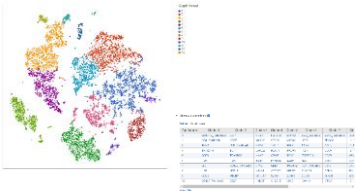
FASTQ, BAM, or Counts



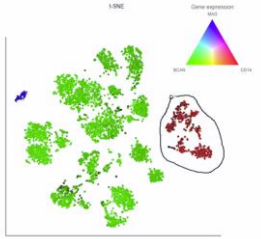
Batch correction



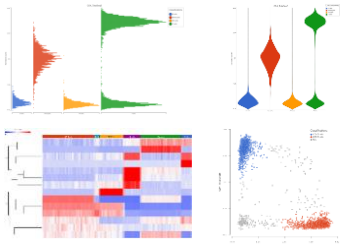
Clustering



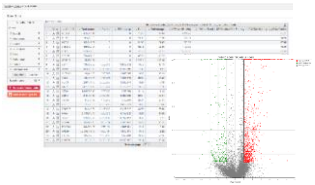
Interactive Visual Analysis



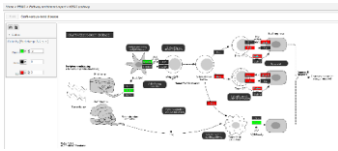
Data Visualization



Differential Analysis



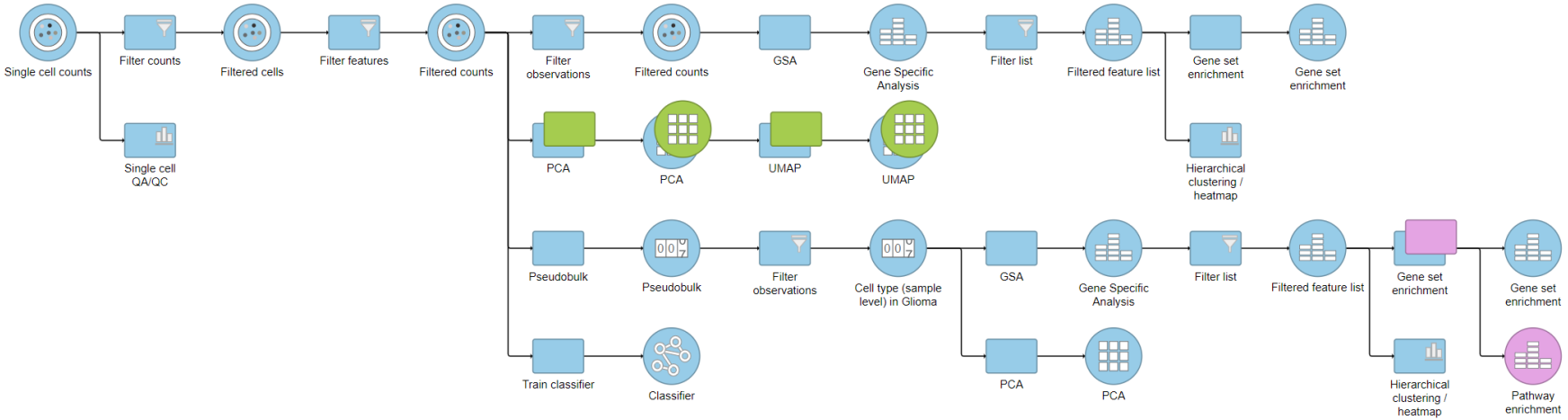
Pathway Analysis



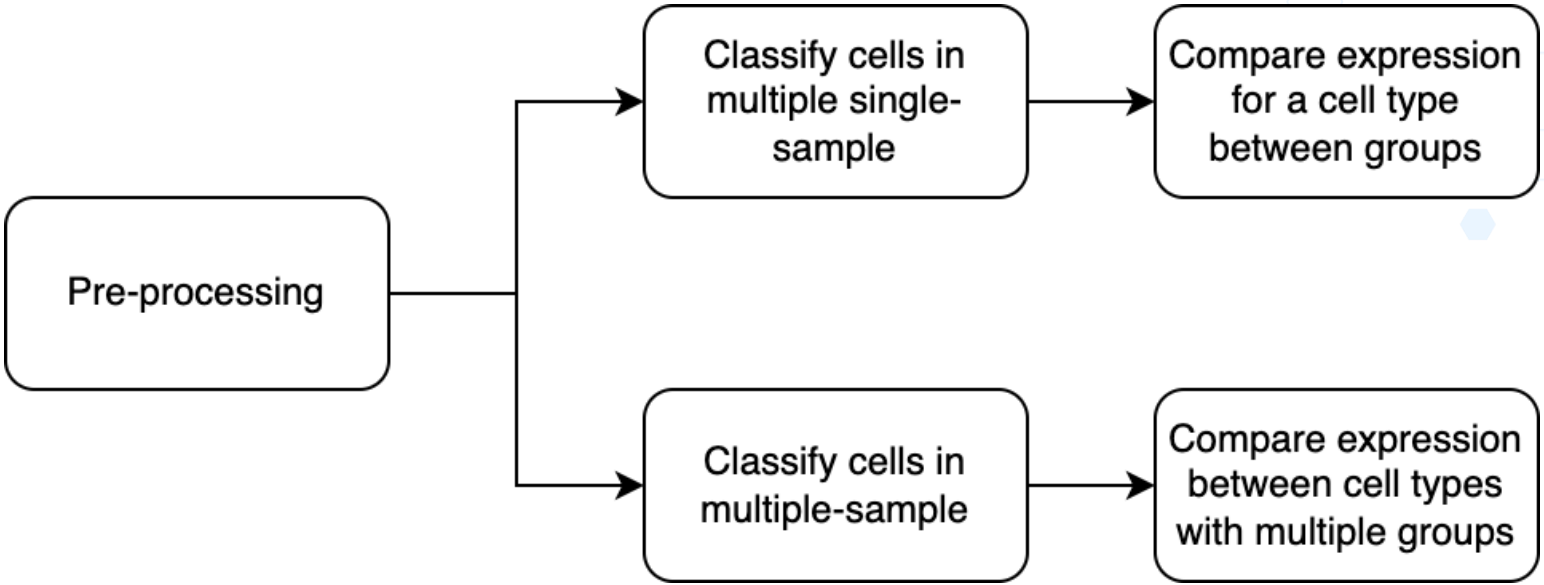
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Demo

Single Cell RNA-Seq Analysis Demo Pipeline

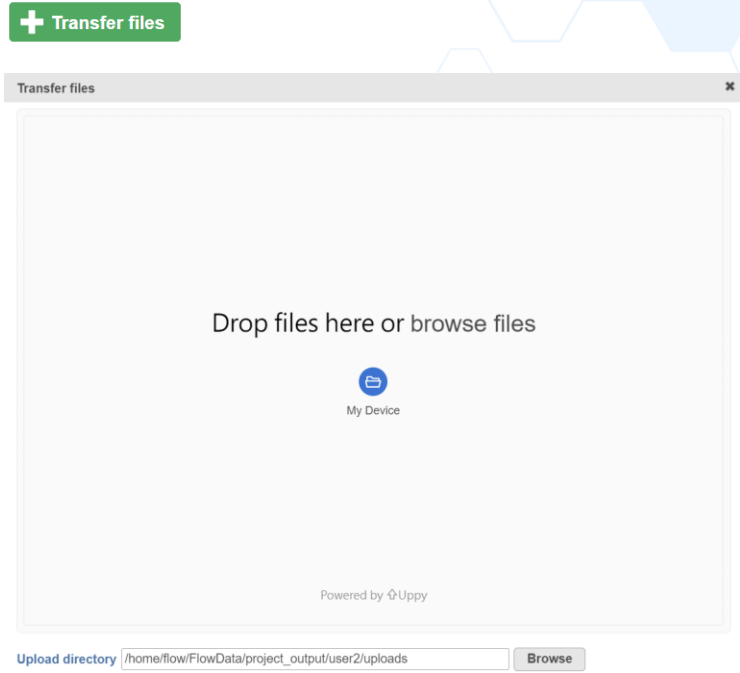


Overview of Demo Analysis

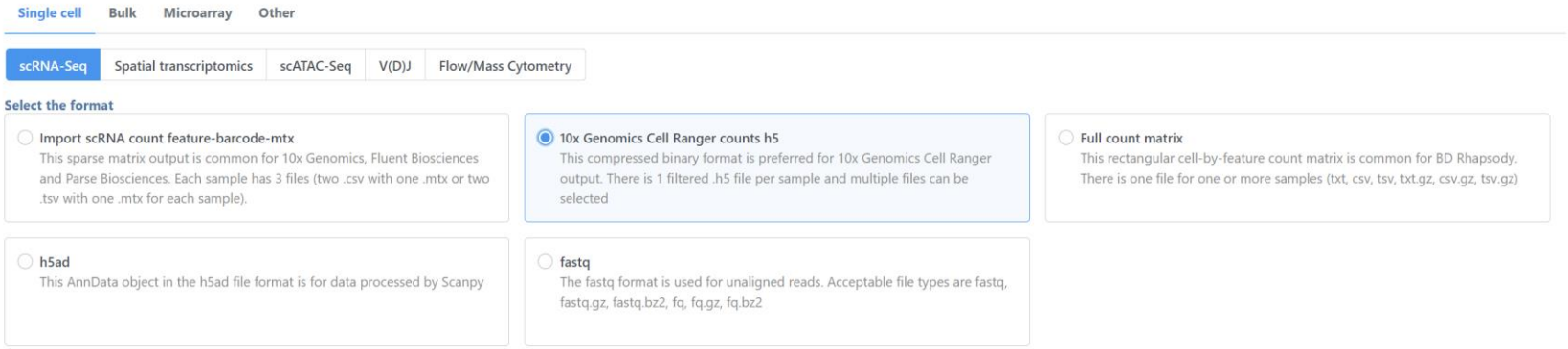


Transfer files

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



Import your own data

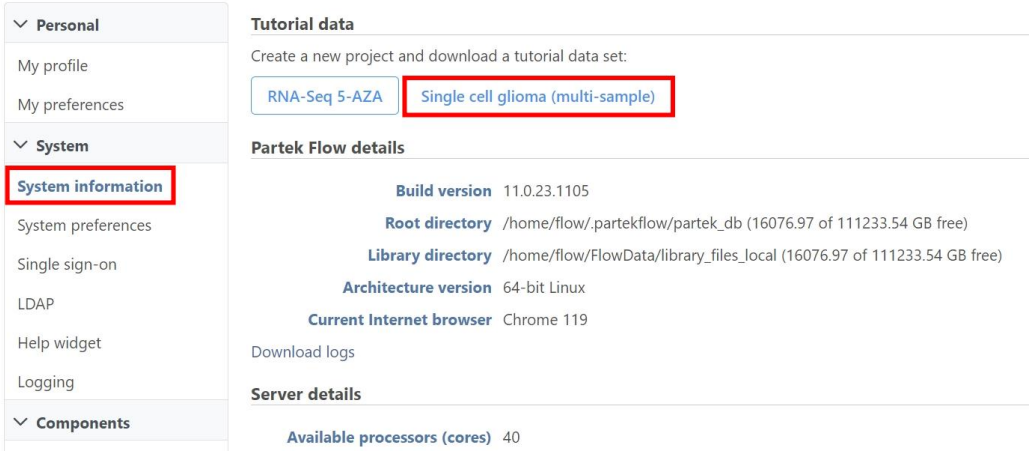


If you want to import your own data

- Click **+New project** and enter a project name
- Select the format
- Select all files and click **Next**

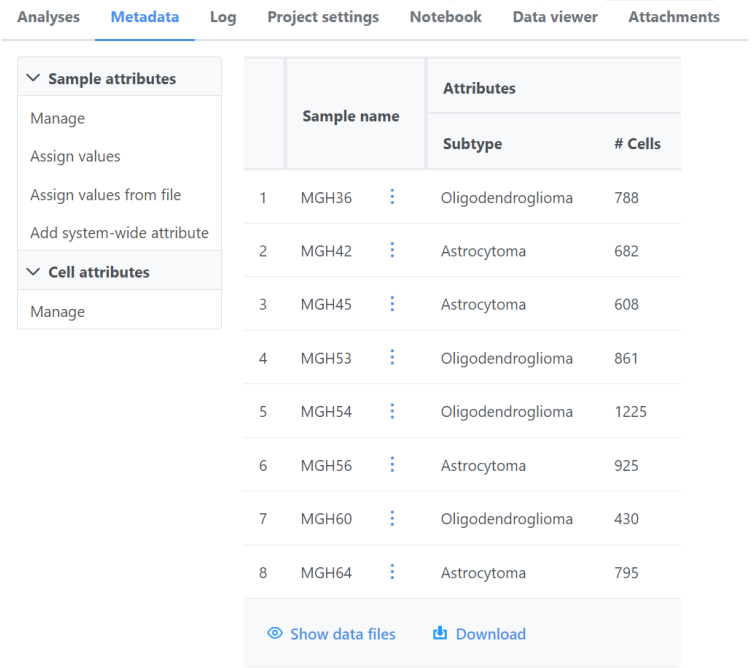
Import tutorial data

- Click your avatar in the upper right
- Click **Settings**
- On **System Information** page, click **Single cell glioma (multi-sample)** from Download tutorial data block



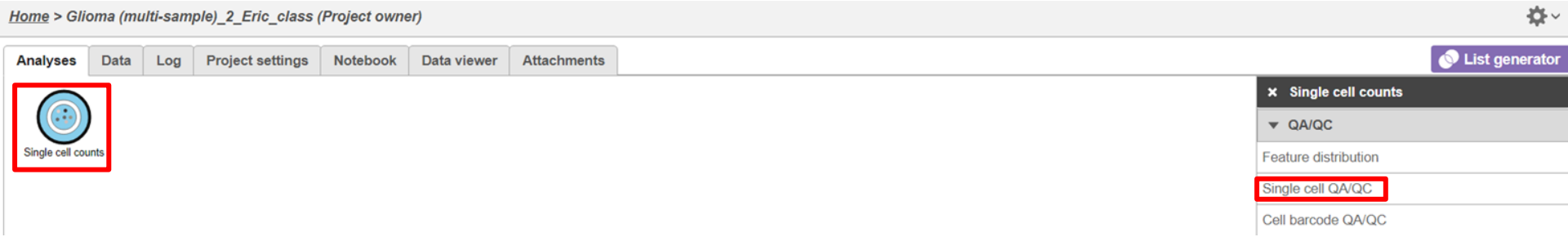
Sample data table

- 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



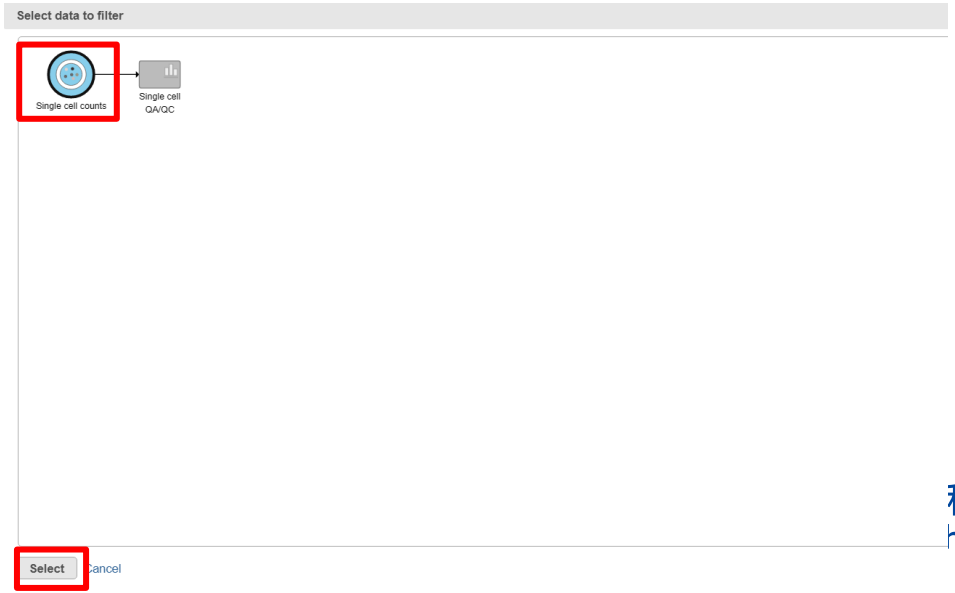
Filtering cells in single cell RNA-Seq data

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



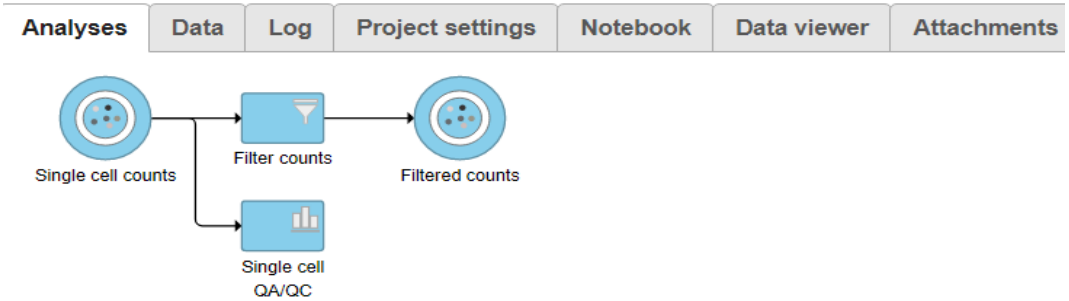
Filtering cells in single cell RNA-Seq data

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



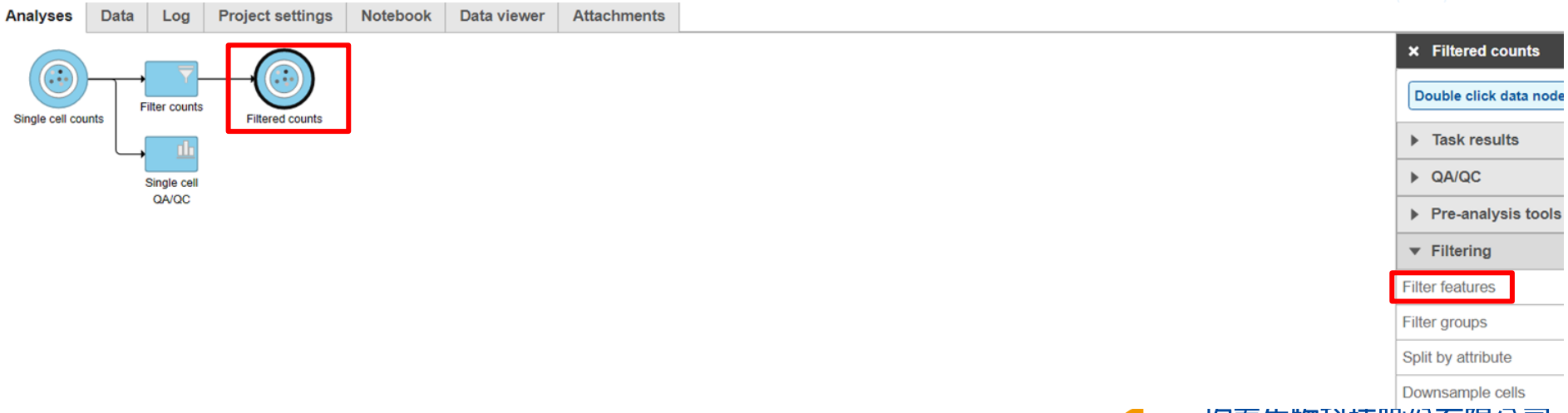
Filtering cells in single cell RNA-Seq data

- 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



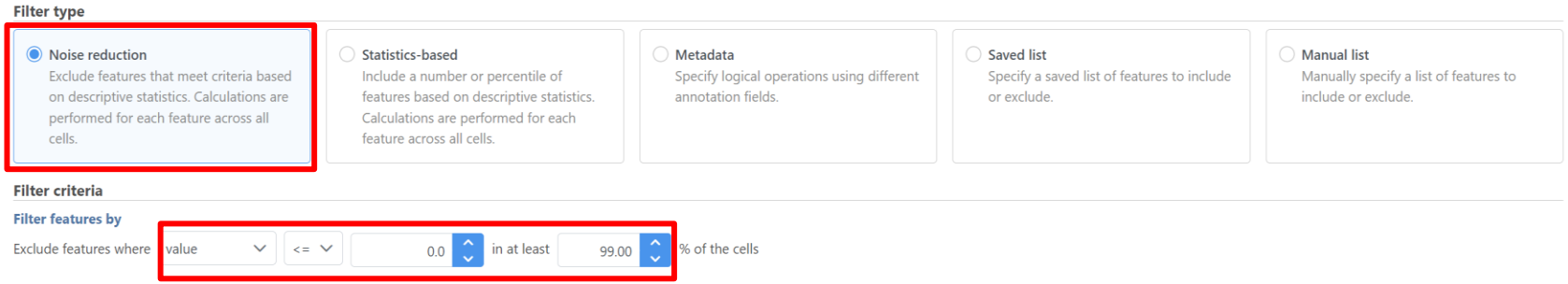
Filtering genes in single cell RNA-Seq data

- 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

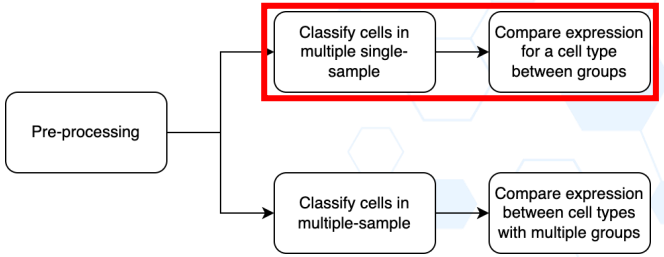
- Click the **Noise reduction** checkbox
- Set the Noise reduction filter to **Exclude features where value <= 0 in 99% of cells**
- Click **Finish** to apply the filter



Normalizing single cell RNA-Seq data

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

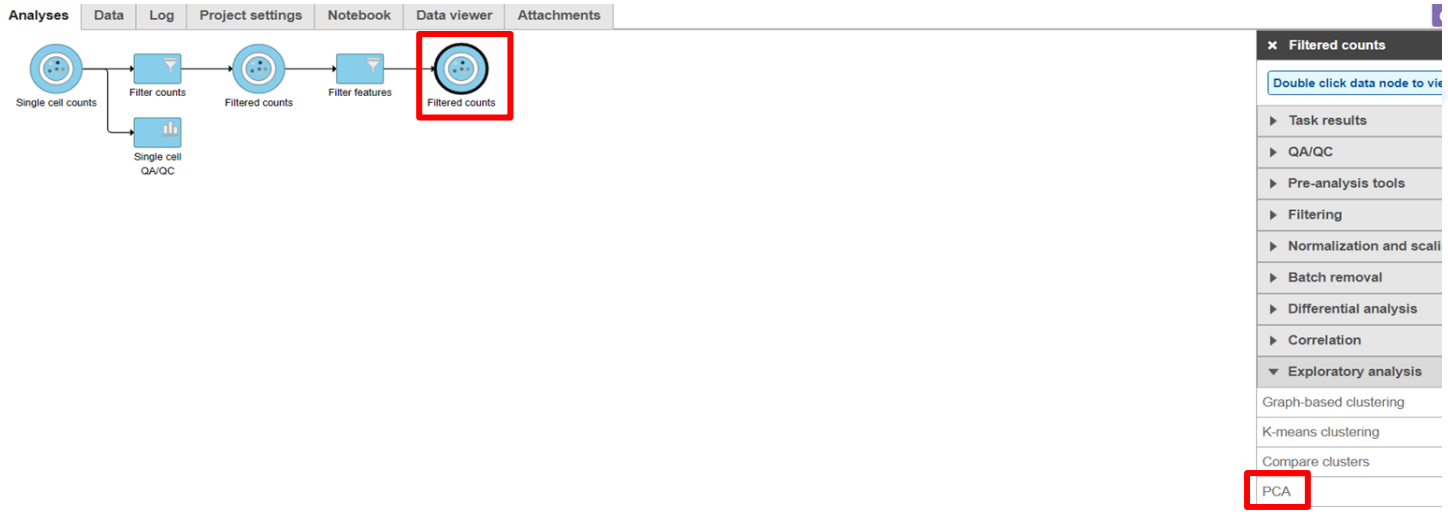
- The tutorial data set is taken from a published study and has already been normalized using TPM (Transcripts per million), which normalizes for the length of feature and total reads, and transformed as log2(TPM/10+1). This normalization and transformation scheme can be performed in Partek Flow, along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



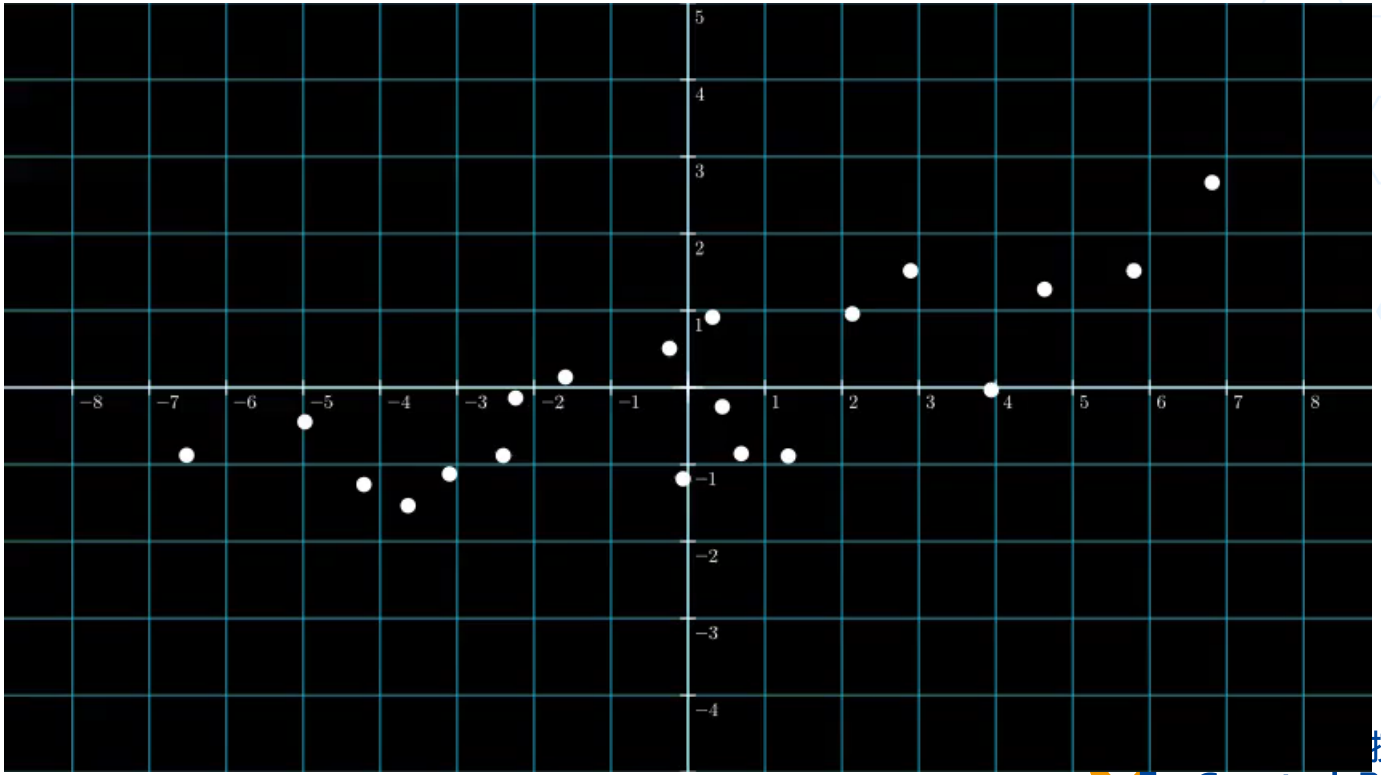
Compare expression for a cell type between groups

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



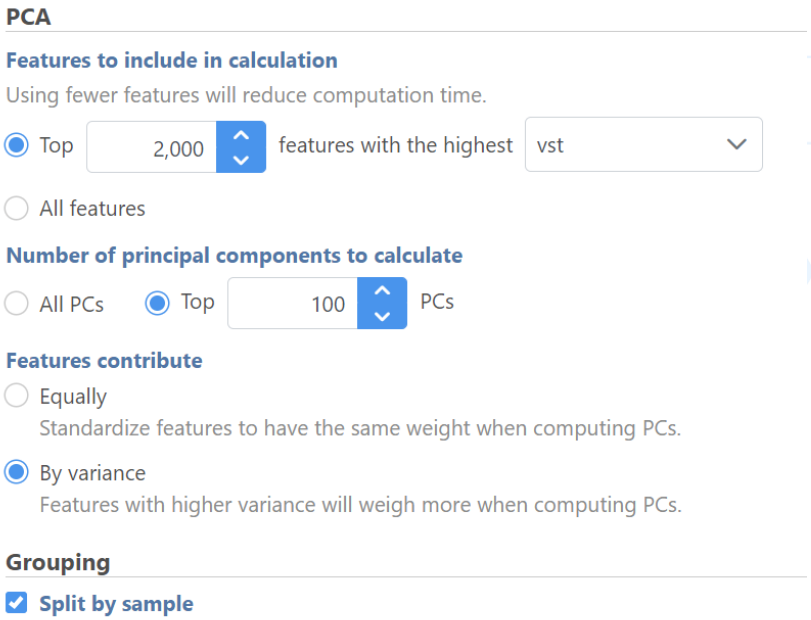
PCA



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

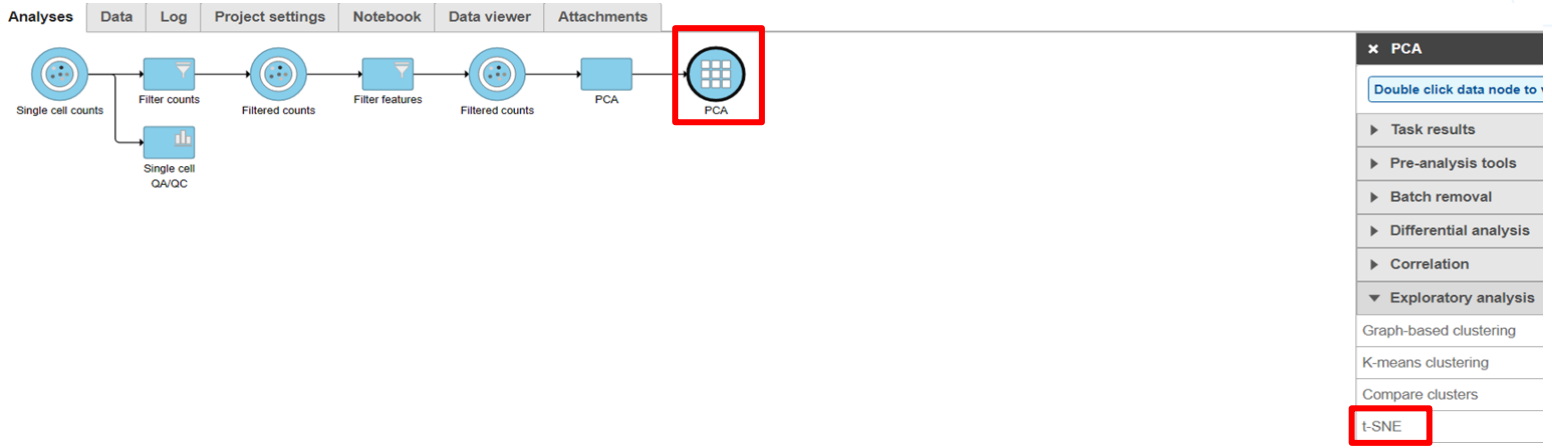
Multiple single-sample t-SNE plots

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

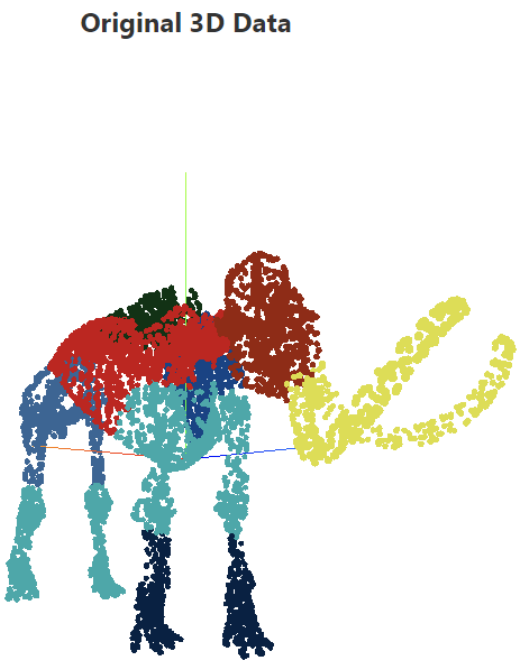


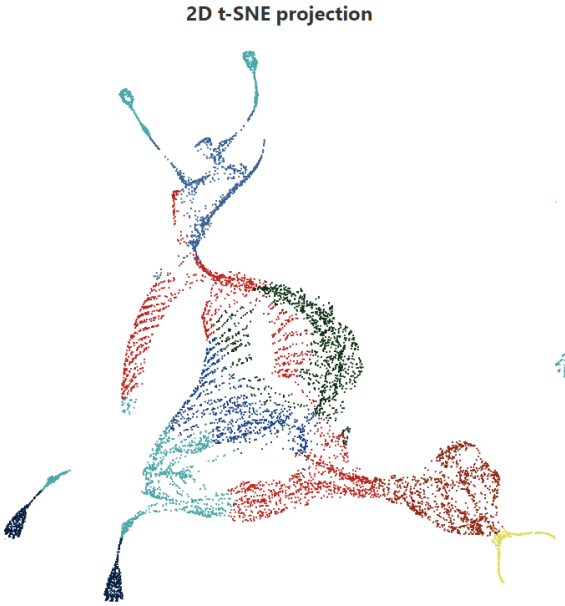
Multiple single-sample t-SNE plots

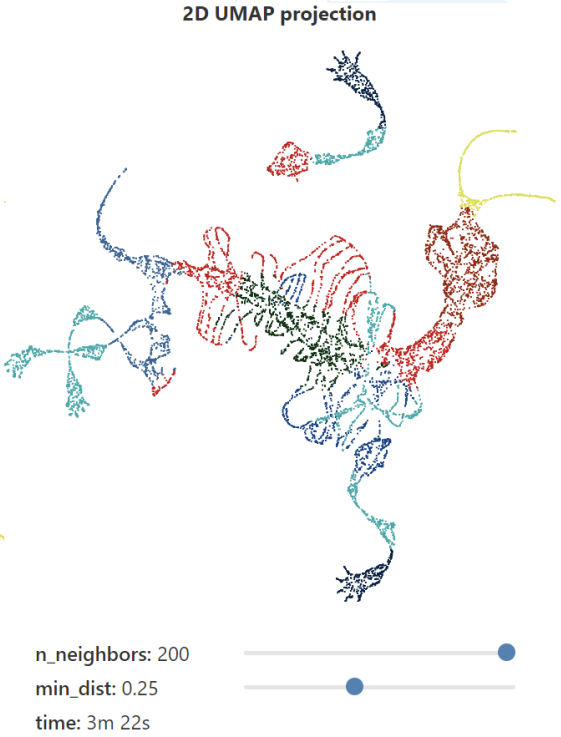
- Click the **PCA** data node
- Select **t-SNE** from the **Exploratory analysis** section of the task menu



UMAP & t-SNE



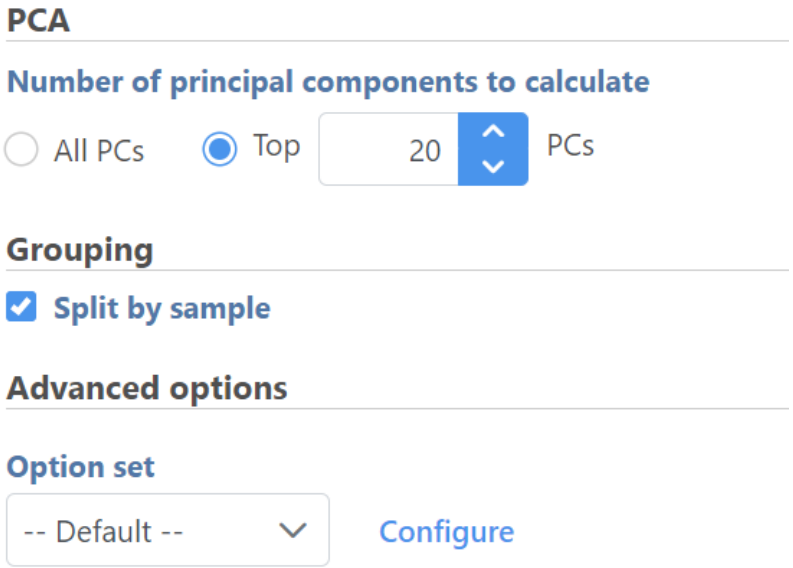




https://pair-code.github.io/understanding-umap/

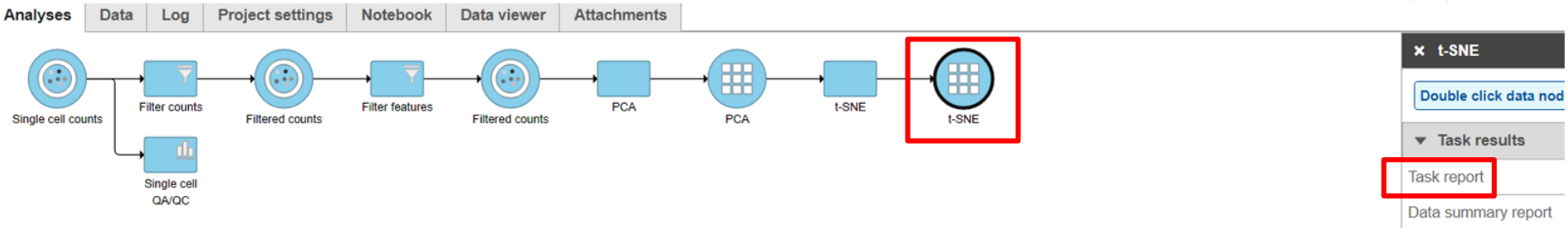
Multiple single-sample t-SNE plots

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



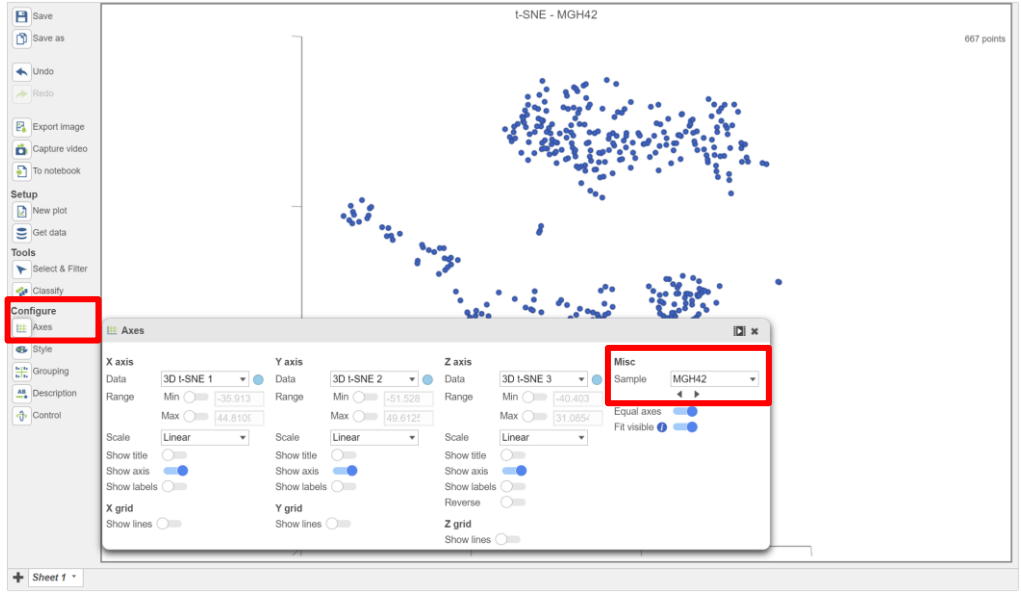
Multiple single-sample t-SNE plots

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



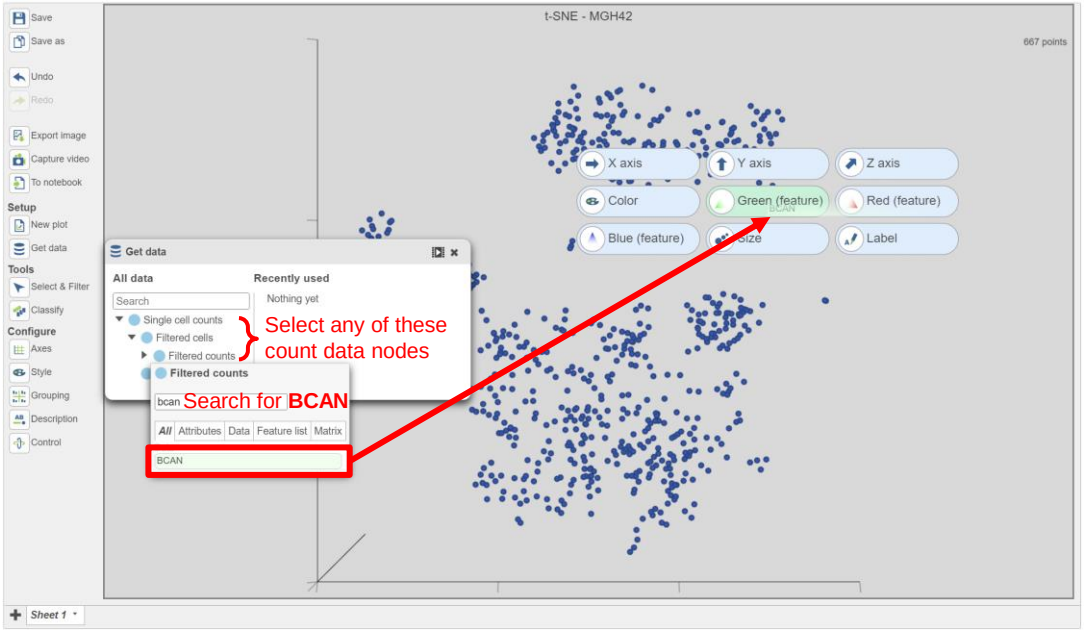
Multiple single-sample t-SNE plots

- 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

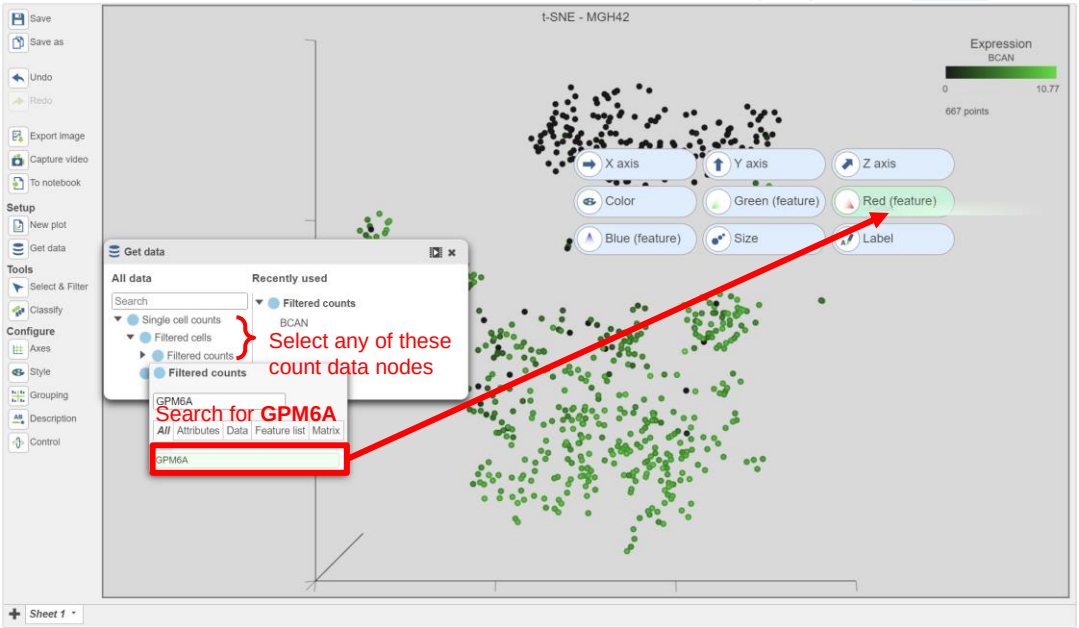
- Select any of the count data nodes from the Get Data window
- Search for the **BCAN** gene
- Click and drag the **BCAN** gene onto the plot and drop it over the **Green (feature)** option



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

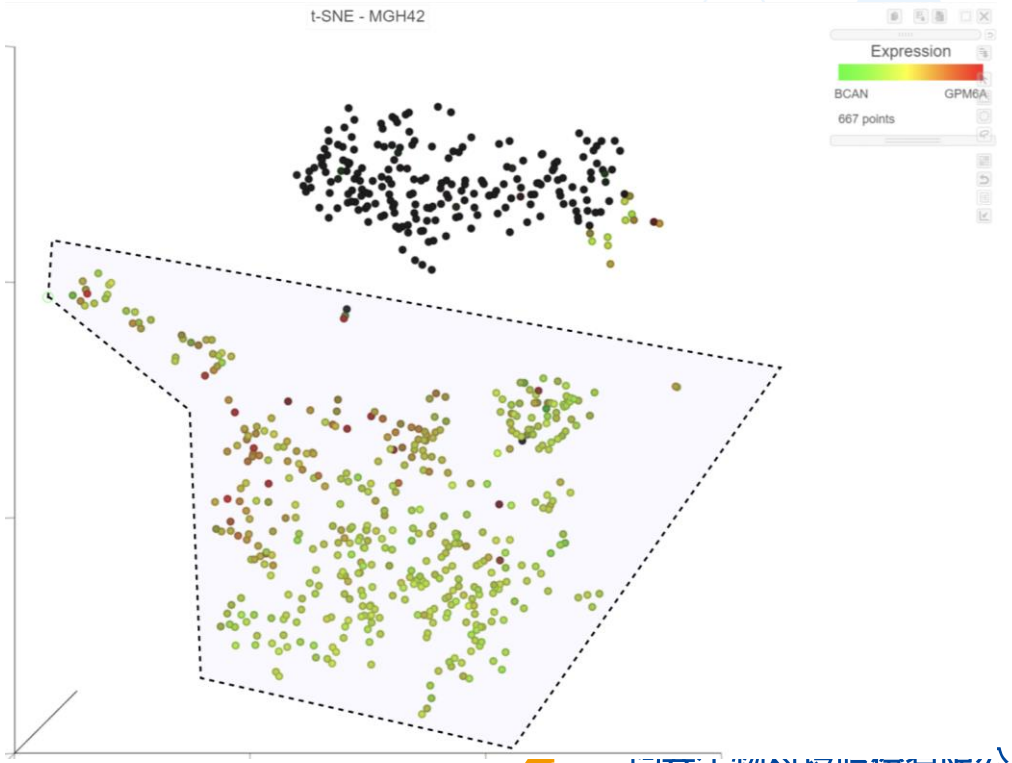
Multiple single-sample t-SNE plots

- 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



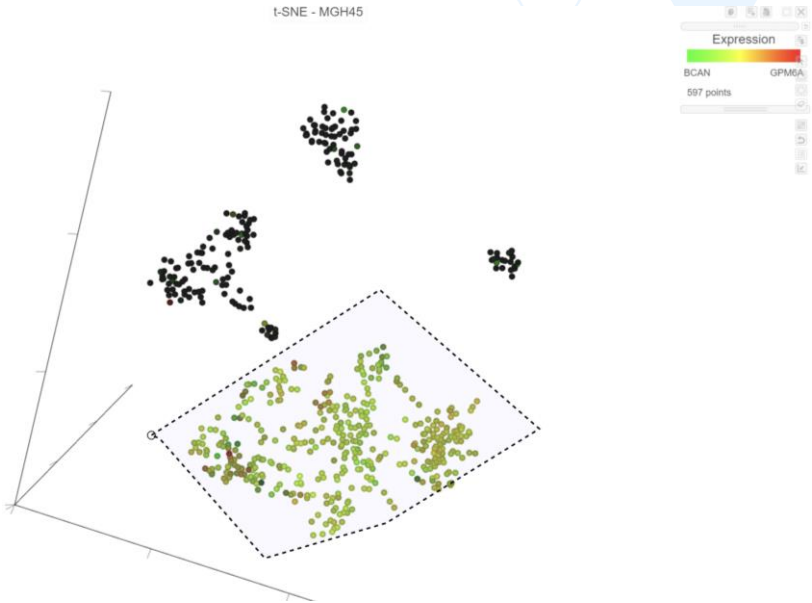
Multiple single-sample t-SNE plots

- 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

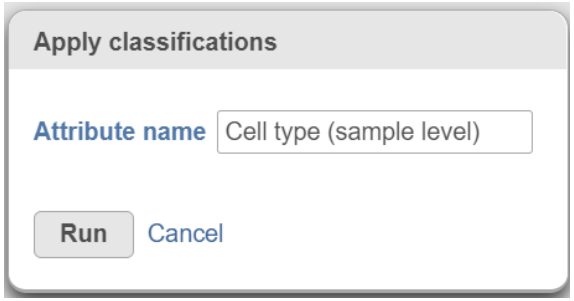
- Select **Classify selection** in the Classify window
- Type **Glioma** or select **Glioma** from the drop-down list
- Click **Save**



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

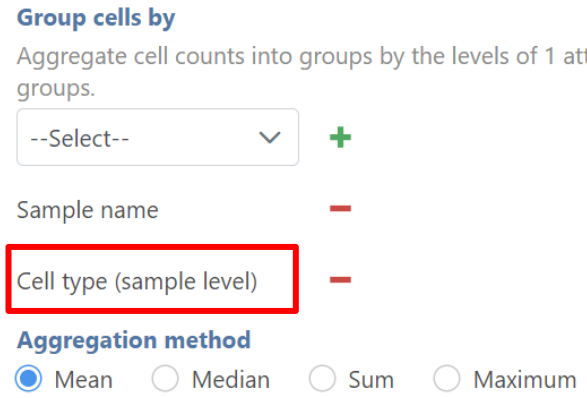
- Click **Apply classifications** in the Classify window
- Name the classification attribute **Cell type (sample level)**
- Click **Run**
- Click **OK** on the information box that says an attribute task has been created





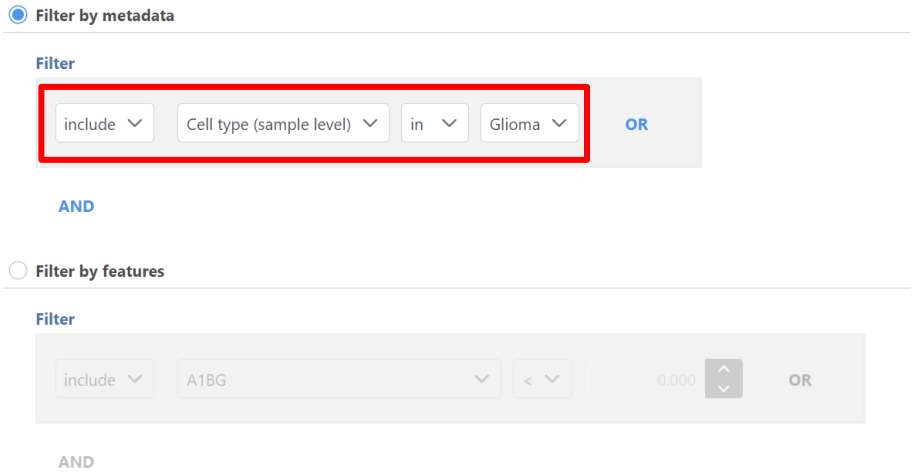
Pool cells (Pseudobulk)

- Click the **Filtered counts** data node
- Expand the **Pre-analysis tools** section of the task menu
- Click **Pseudobulk**
- Group cells by **Cell type (sample level)**
- Keep the default aggregation method (**Mean**)
- Click **Finish**



Pool cells (Pseudobulk)

- Click the **Pseudobulk** data node
- Expand the **Filtering** section of the task menu
- Choose **Filter pseudobulk**
- Set the filter to **include Cell type (sample level) in Glioma**
- Click **Finish**

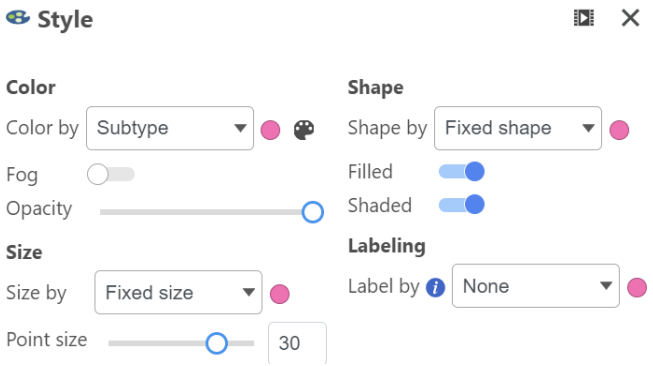


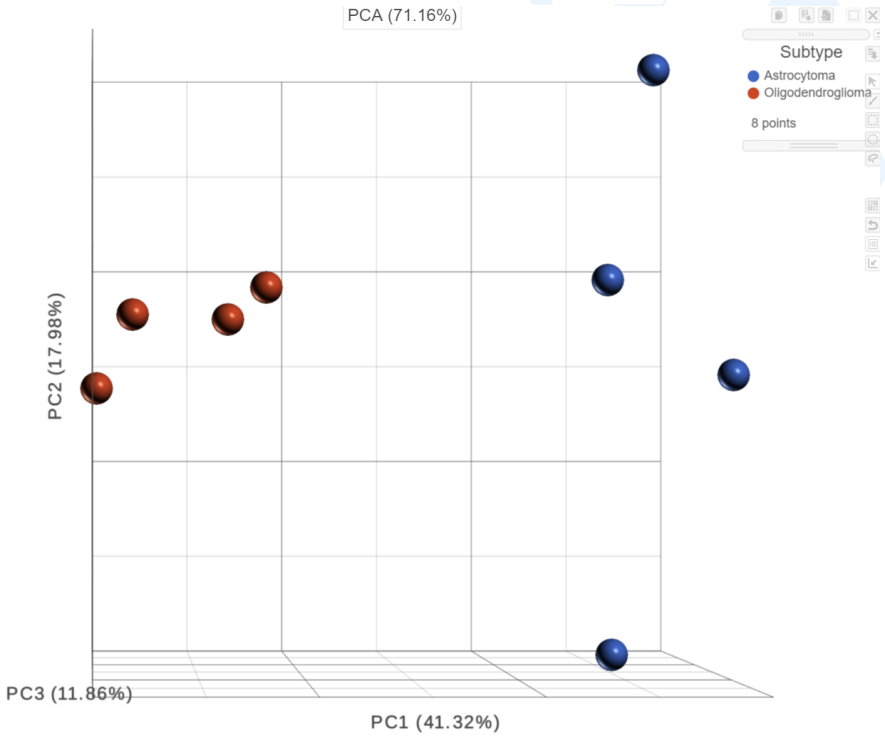
Explore data with PCA

- 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



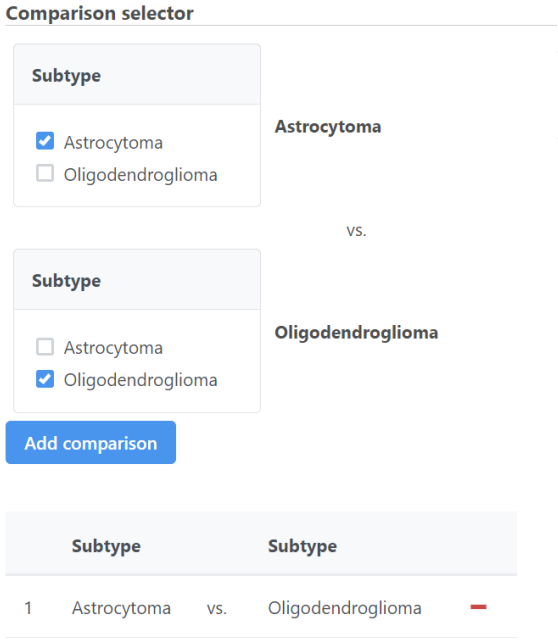


Identify differentially expressed genes

- 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
- Choose **Gene Specific Analysis**
- 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  to open the volcano plot for a comparison

Feature list

Results: 17196

Filter Clear all

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

Save filter

Saved filters

Generate filtered node

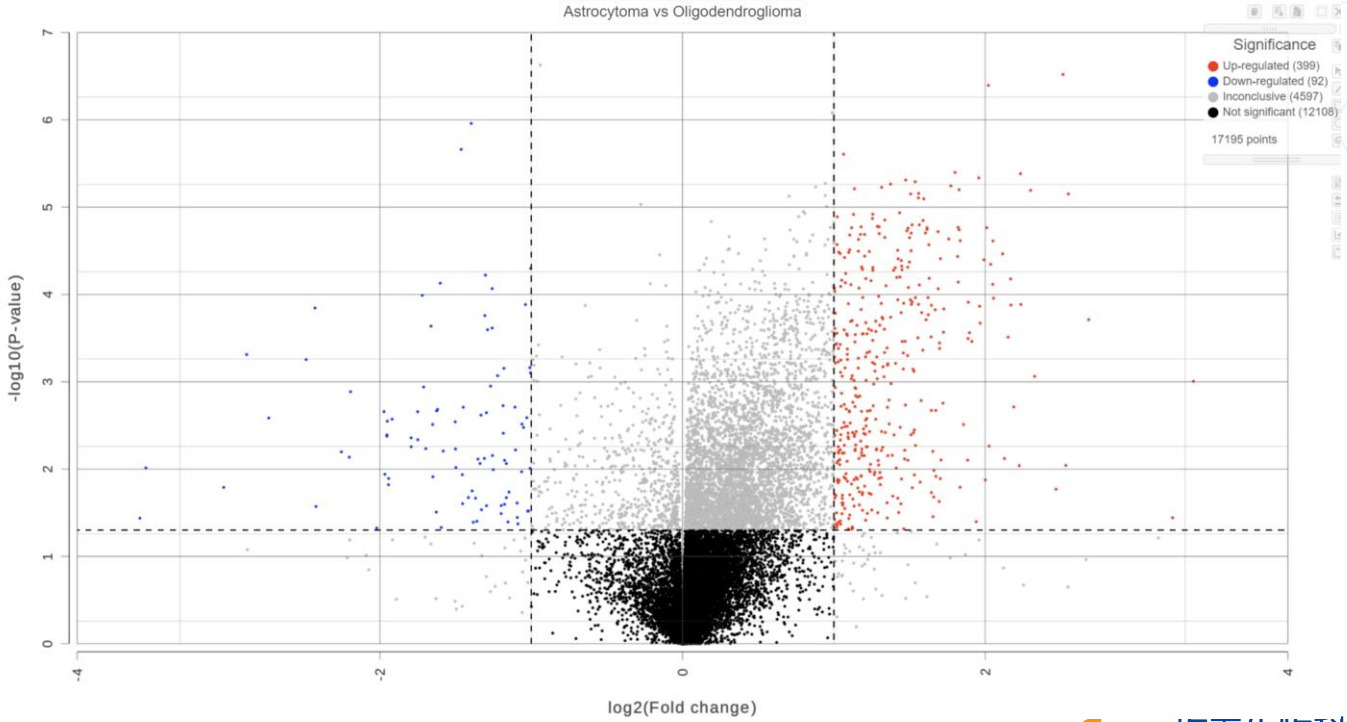
Save as managed list

Optional columns

Astrocytoma vs Oligodendroglioma

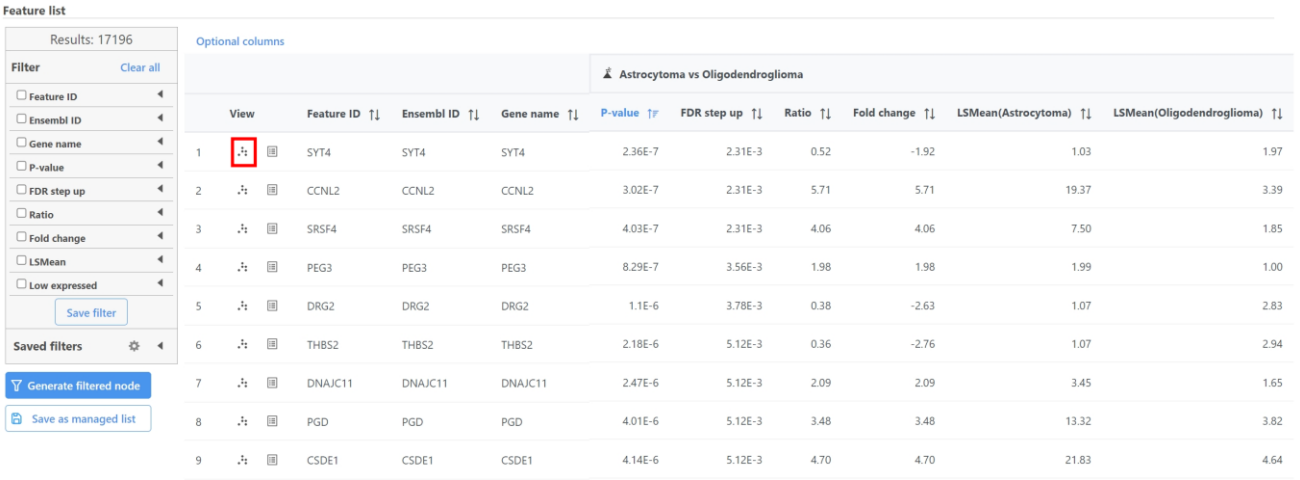
	View	Feature ID	Ensembl ID	Gene name	P-value	FDR step up	Ratio	Fold change	LSMean(Astrocytoma)	LSMean(Oligodendroglioma)
1		SYT4	SYT4	SYT4	2.36E-7	2.31E-3	0.52	-1.92	1.03	1.97
2		CCNL2	CCNL2	CCNL2	3.02E-7	2.31E-3	5.71	5.71	19.37	3.39
3		SRSF4	SRSF4	SRSF4	4.03E-7	2.31E-3	4.06	4.06	7.50	1.85
4		PEG3	PEG3	PEG3	8.29E-7	3.56E-3	1.98	1.98	1.99	1.00
5		DRG2	DRG2	DRG2	1.1E-6	3.78E-3	0.38	-2.63	1.07	2.83
6		THBS2	THBS2	THBS2	2.18E-6	5.12E-3	0.36	-2.76	1.07	2.94
7		DNAJC11	DNAJC11	DNAJC11	2.47E-6	5.12E-3	2.09	2.09	3.45	1.65
8		PGD	PGD	PGD	4.01E-6	5.12E-3	3.48	3.48	13.32	3.82
9		CSDE1	CSDE1	CSDE1	4.14E-6	5.12E-3	4.70	4.70	21.83	4.64

Identify differentially expressed genes

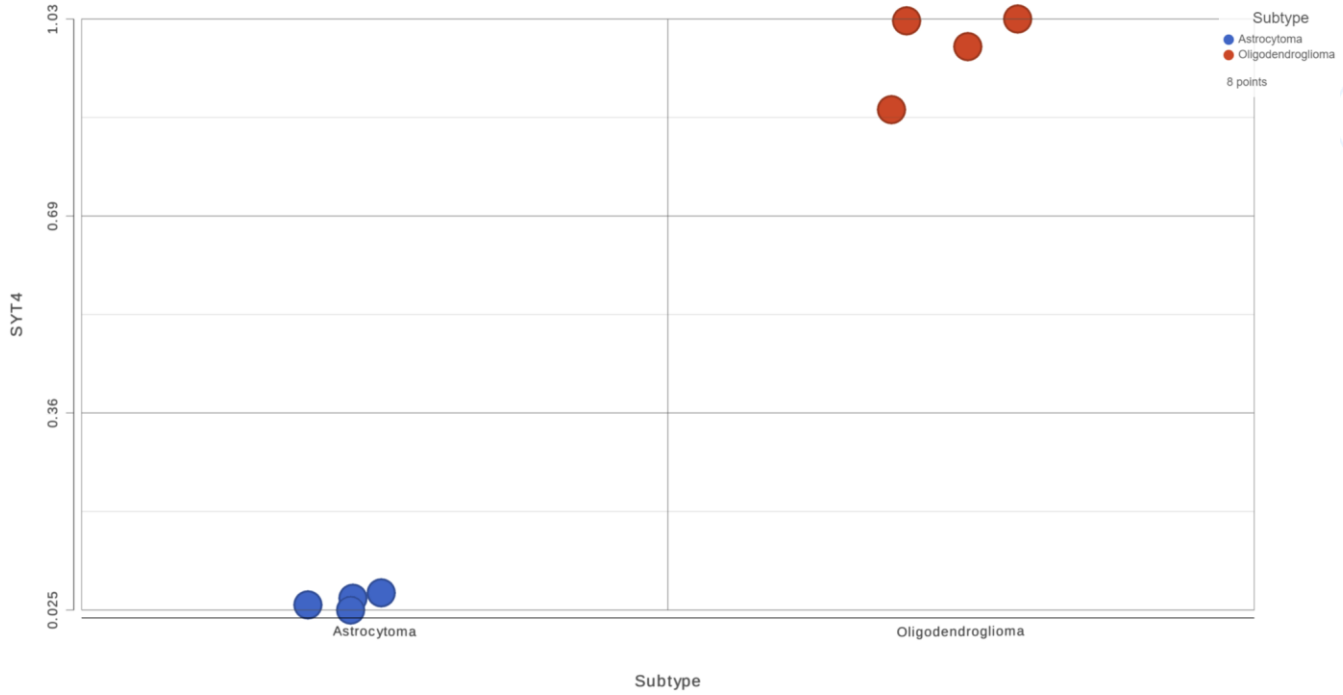


Identify differentially expressed genes

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



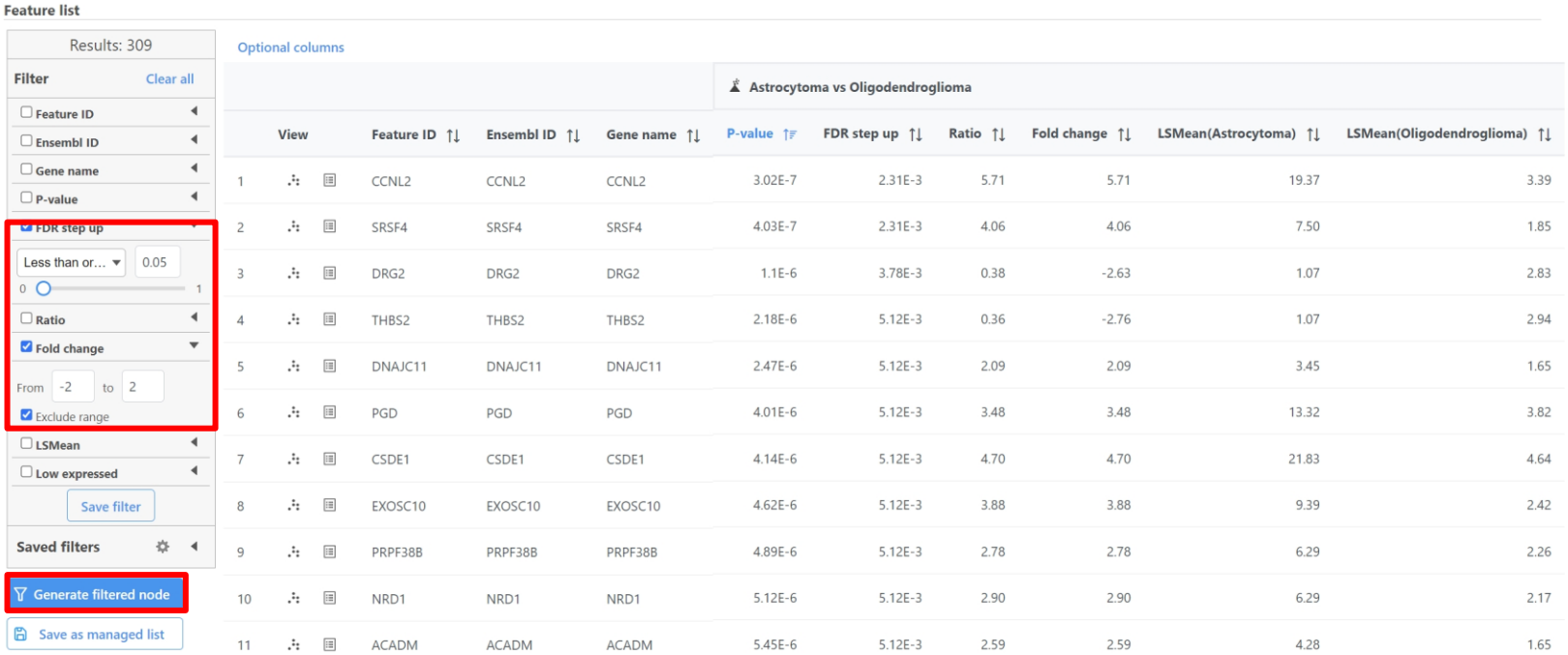
Identify differentially expressed genes



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 **Generate filtered node** to apply the filter

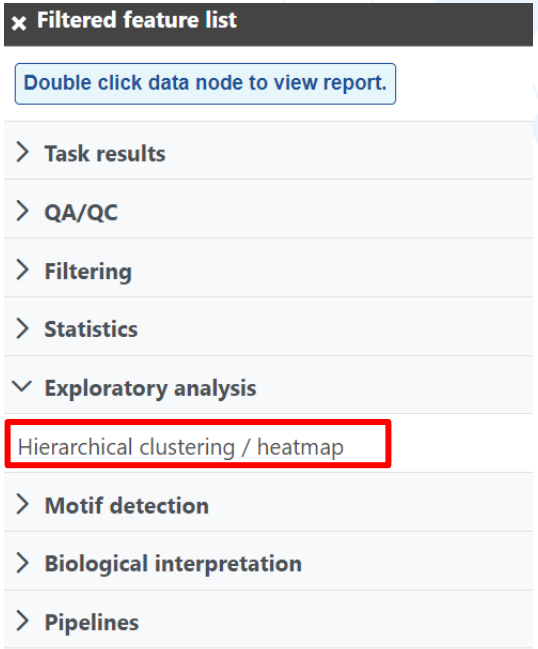
Identify differentially expressed genes



	View	Feature ID	Ensembl ID	Gene name	P-value	FDR step up	Ratio	Fold change	LSMean(Astrocytoma)	LSMean(Oligodendroglioma)
1		CCNL2	CCNL2	CCNL2	3.02E-7	2.31E-3	5.71	5.71	19.37	3.39
2		SRSF4	SRSF4	SRSF4	4.03E-7	2.31E-3	4.06	4.06	7.50	1.85
3		DRG2	DRG2	DRG2	1.1E-6	3.78E-3	0.38	-2.63	1.07	2.83
4		THBS2	THBS2	THBS2	2.18E-6	5.12E-3	0.36	-2.76	1.07	2.94
5		DNAJC11	DNAJC11	DNAJC11	2.47E-6	5.12E-3	2.09	2.09	3.45	1.65
6		PGD	PGD	PGD	4.01E-6	5.12E-3	3.48	3.48	13.32	3.82
7		CSDE1	CSDE1	CSDE1	4.14E-6	5.12E-3	4.70	4.70	21.83	4.64
8		EXOSC10	EXOSC10	EXOSC10	4.62E-6	5.12E-3	3.88	3.88	9.39	2.42
9		PRPF38B	PRPF38B	PRPF38B	4.89E-6	5.12E-3	2.78	2.78	6.29	2.26
10		NRD1	NRD1	NRD1	5.12E-6	5.12E-3	2.90	2.90	6.29	2.17
11		ACADM	ACADM	ACADM	5.45E-6	5.12E-3	2.59	2.59	4.28	1.65

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



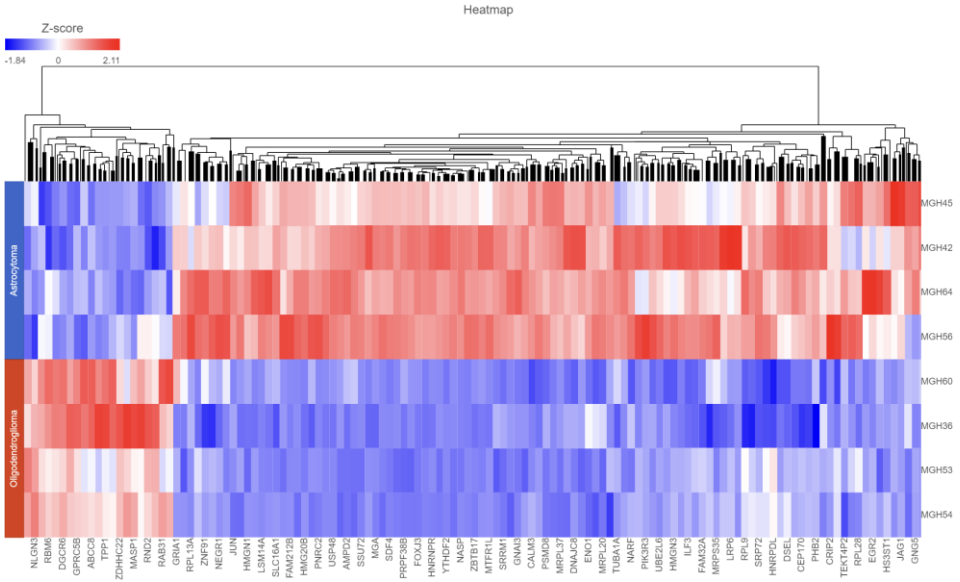
Analyze differentially expressed genes

- 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	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:0003723	RNA binding	molecular function	29.34	1.8E-13	3.8E-9	0.05	1,474	74	1,400	212	12,156
GO:0005654	nucleoplasm	cellular component	23.23	8.18E-11	5.89E-7	0.04	3,382	119	3,263	167	10,293
GO:0008380	RNA splicing	biological process	23.20	8.4E-11	5.89E-7	0.09	316	28	288	258	13,268
GO:0006396	RNA processing	biological process	21.96	2.9E-10	1.53E-6	0.06	749	44	705	242	12,851
GO:0032991	protein-containing complex	cellular component	21.42	4.98E-10	2.1E-6	0.03	4,908	152	4,756	134	8,800
GO:0003676	nucleic acid binding	molecular function	20.63	1.1E-9	3.86E-6	0.03	3,310	114	3,196	172	10,360
GO:0016071	mRNA metabolic process	biological process	20.01	2.04E-9	6.12E-6	0.07	515	34	481	252	13,075
GO:0006397	mRNA processing	biological process	19.42	3.7E-9	9.73E-6	0.08	373	28	345	258	13,211
GO:0043229	intracellular organelle	cellular component	18.36	1.06E-8	2.45E-5	0.03	8,542	221	8,321	65	5,235
GO:1990904	ribonucleoprotein complex	cellular component	18.27	1.16E-8	2.45E-5	0.06	665	38	627	248	12,929
GO:0016070	RNA metabolic process	biological process	16.19	9.31E-8	1.65E-4	0.04	1,238	54	1,184	232	12,372

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**
- Choose a database
- Click **Finish** to run enrichment analysis

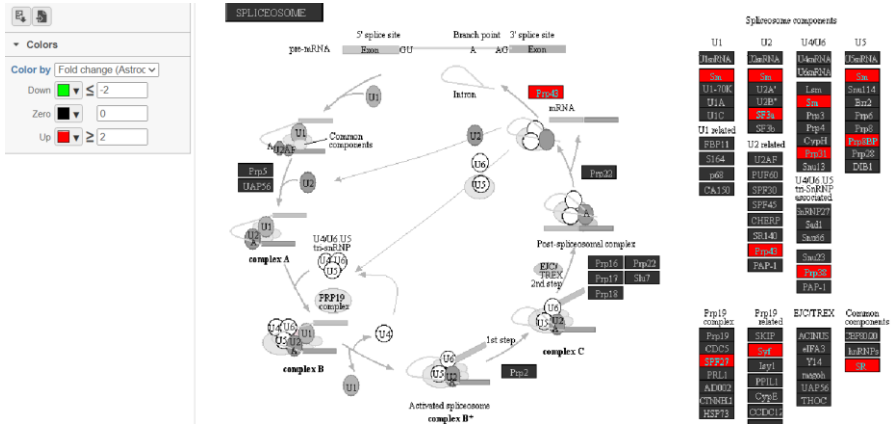
Analyze differentially expressed genes

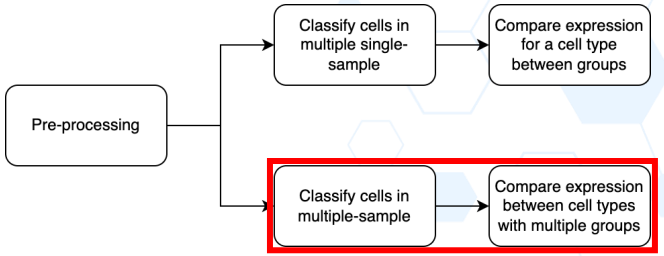
- Double click the **Pathway enrichment** node

Gene set	Description	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
path:hsa03040	Spliceosome	9.22	9.93E-5	0.04	0.09	132	12	120	152	6,305
path:hsa05012	Parkinson disease	8.39	2.27E-4	0.04	0.07	235	16	219	148	6,206
path:hsa03082	ATP-dependent chromatin remodeling	8.10	3.05E-4	0.04	0.09	107	10	97	154	6,328
path:hsa03320	PPAR signaling pathway	7.53	5.38E-4	0.05	0.12	58	7	51	157	6,374
path:hsa04714	Thermogenesis	6.50	1.5E-3	0.11	0.06	202	13	189	151	6,236
path:hsa00190	Oxidative phosphorylation	5.02	6.59E-3	0.39	0.07	112	8	104	156	6,321
path:hsa00100	Steroid biosynthesis	4.83	7.96E-3	0.40	0.18	17	3	14	161	6,411
path:hsa01212	Fatty acid metabolism	4.54	0.01	0.47	0.09	54	5	49	159	6,376
path:hsa05203	Viral carcinogenesis	4.37	0.01	0.50	0.06	177	10	167	154	6,258
path:hsa05016	Huntington disease	3.95	0.02	0.68	0.05	275	13	262	151	6,163
path:hsa03015	mRNA surveillance pathway	3.82	0.02	0.71	0.07	88	6	82	158	6,343

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.





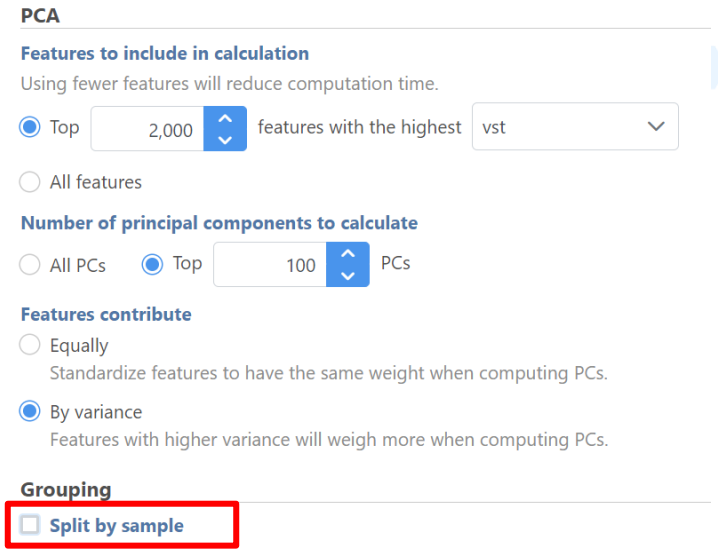
Compare expression between cell types with multiple groups

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.

One multi-sample t-SNE plot

- 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**

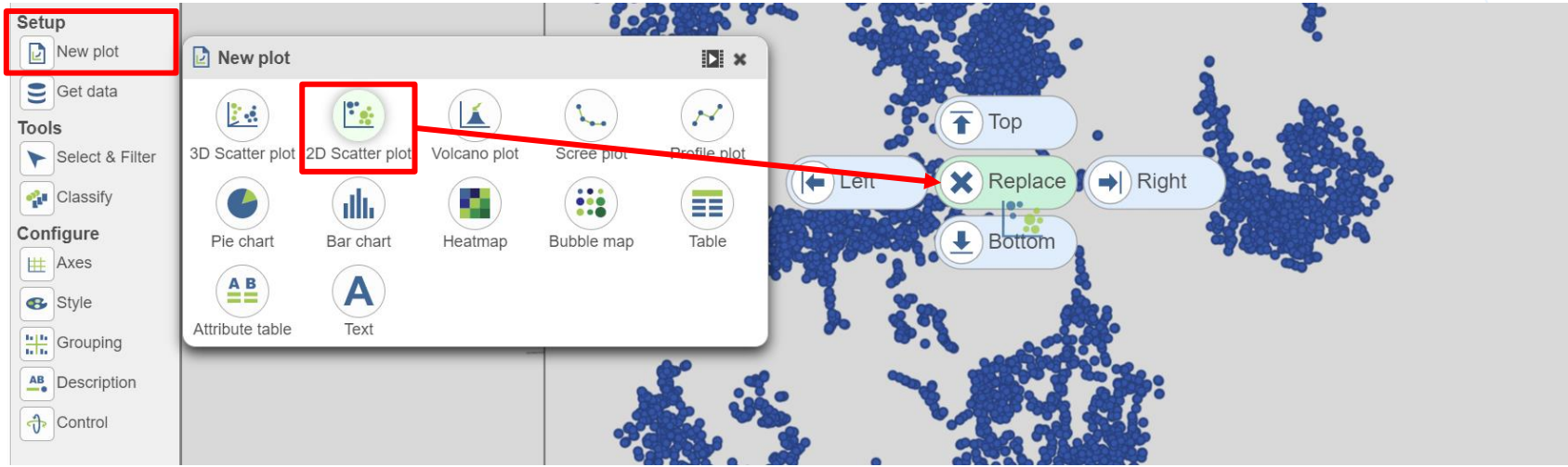


One multi-sample t-SNE plot

- 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

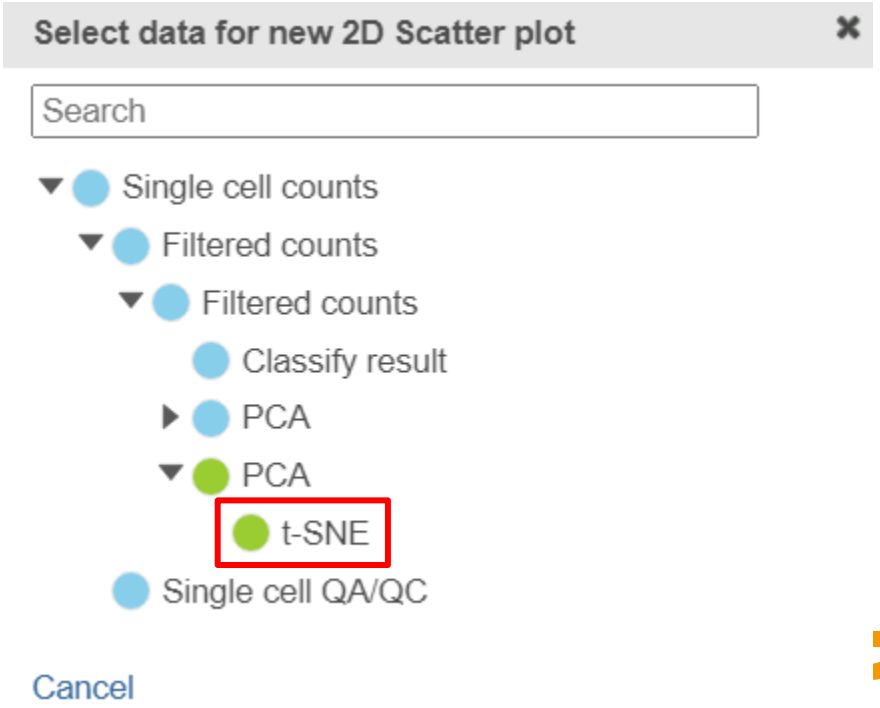
One multi-sample t-SNE plot

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



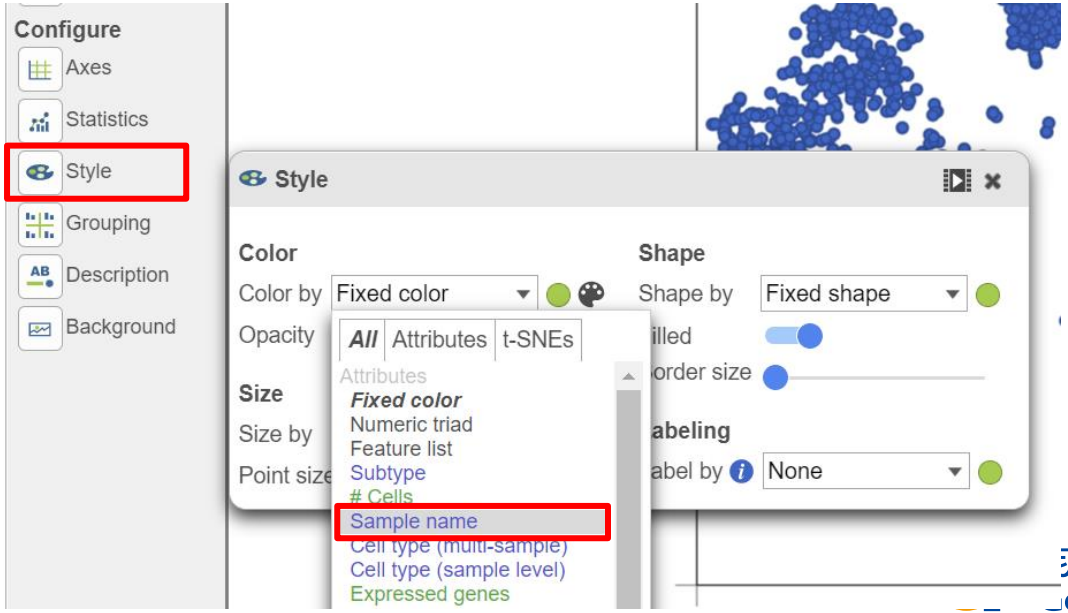
One multi-sample t-SNE plot

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



One multi-sample t-SNE plot

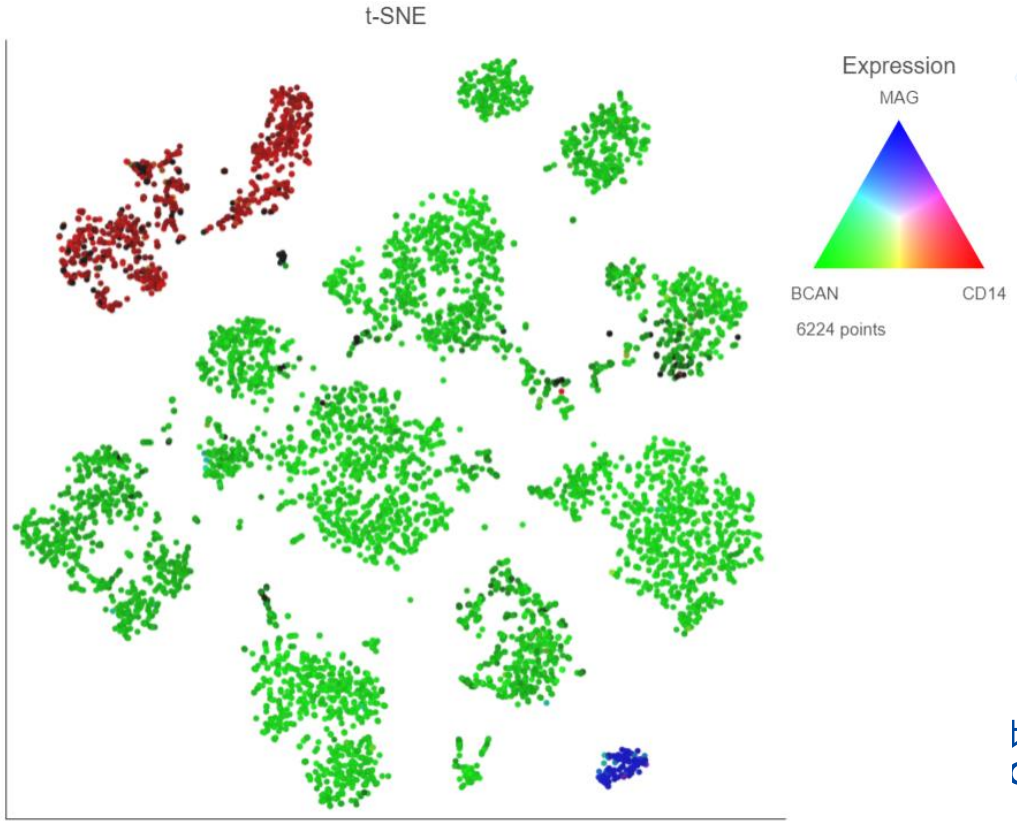
- In the Configure section, click **Style** and choose **Sample name** from the **Color by** drop-down list



One multi-sample t-SNE plot

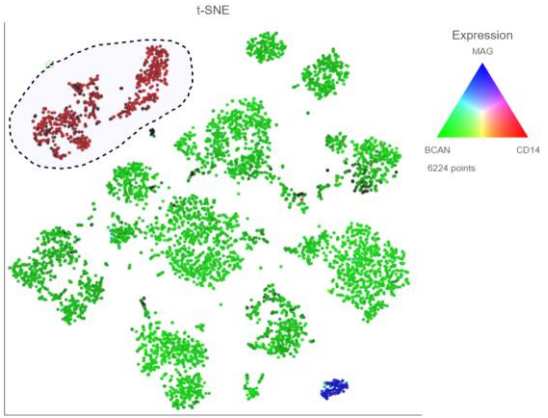
- Using marker genes, BCAN (glioma), CD14 (microglia), and MAG (oligodendrocytes), we can assess whether these multi-sample clusters belong to our known cell types.
  - Select any of the count data nodes from the **Data** card on the left (Single cell counts, or any of the Filtered counts)
  - Search 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



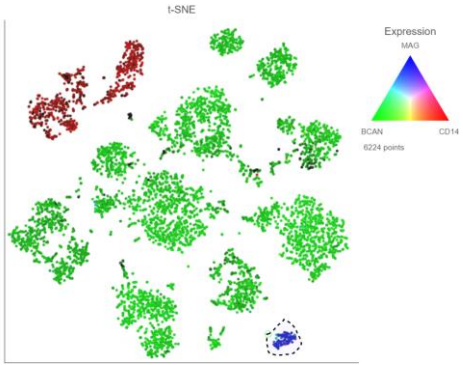
One multi-sample t-SNE plot

- 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**



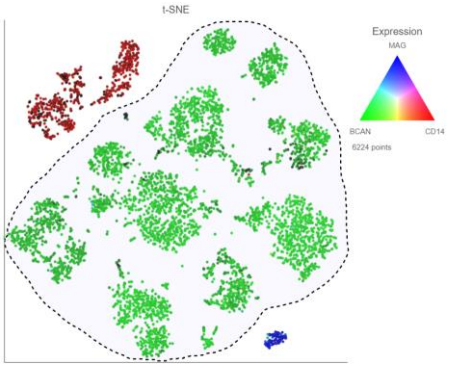
One multi-sample t-SNE plot

- 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**



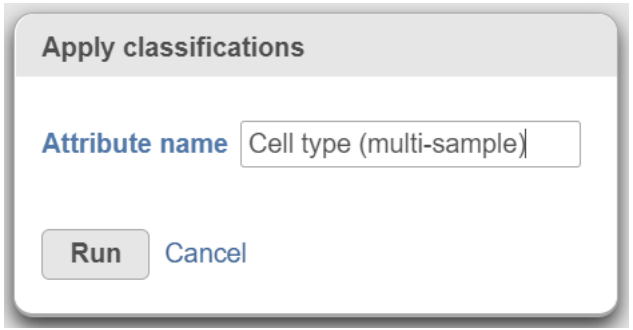
One multi-sample t-SNE plot

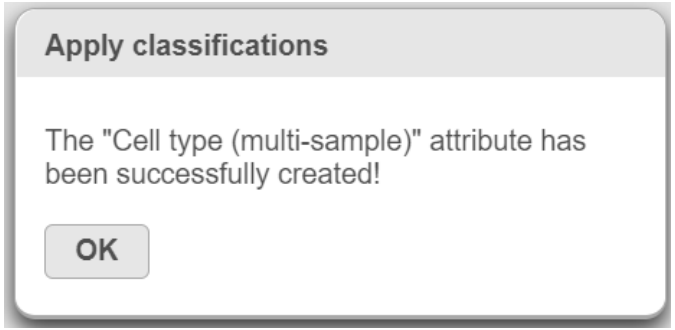
- Finally, we will classify the BCAN expressing cells on the plot as glioma cells from every sample.
  - Switch to pointer mode by clicking  in the top right corner of the plot
  - Deselect the cells by clicking on any blank space on the plot
  - Switch to lasso mode again by clicking the  icon in the 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

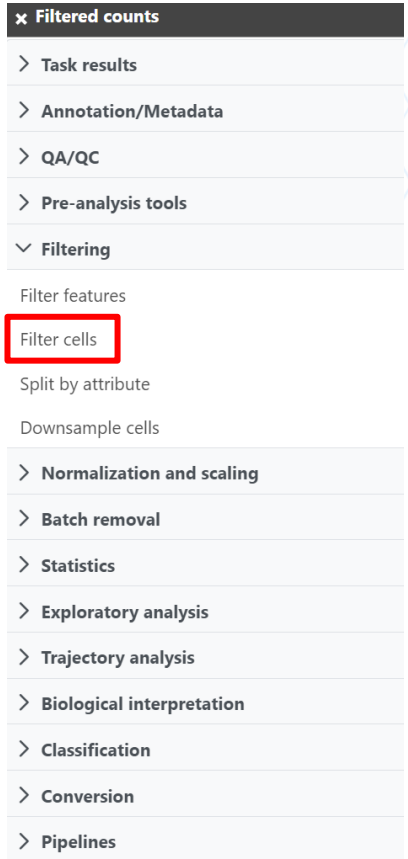
- Click **Apply classifications** in the Classify window
- Name the classification attribute **Cell type (multi-sample)**
- Click **Run**
- Click **OK** on the information box that says an attribute task has been created





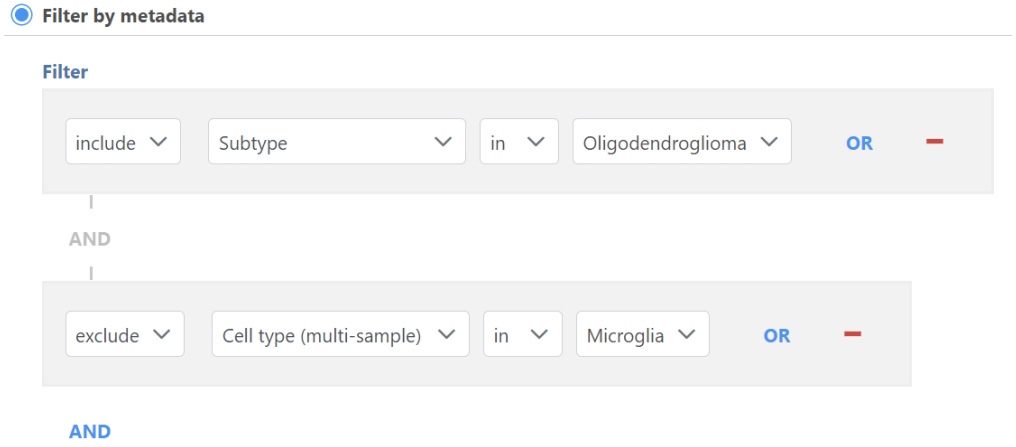
Filter groups

- 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 cells**



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



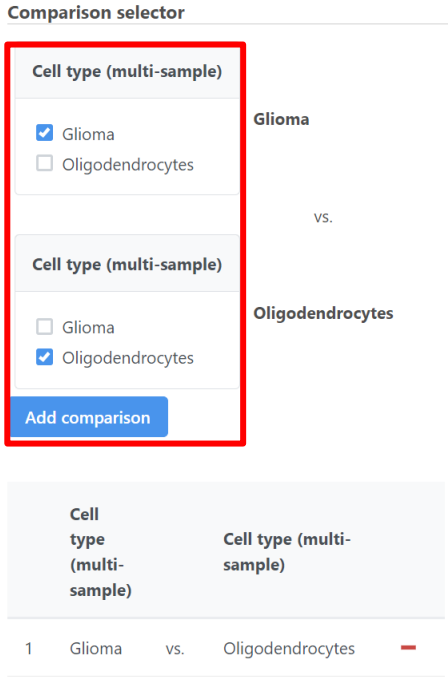
Identify differentially expressed genes

- 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
- Choose **Gene Specific Analysis**
- Click **Next**
- Check the **Cell type (multi-sample)** attribute
- Click **Next**



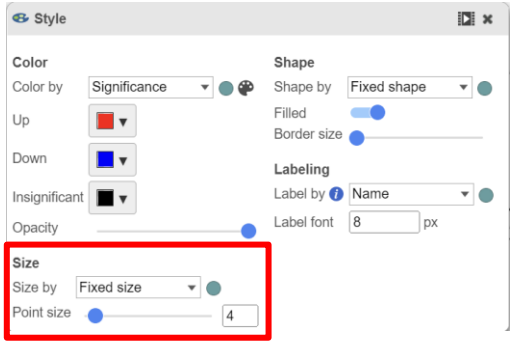
Identify differentially expressed genes

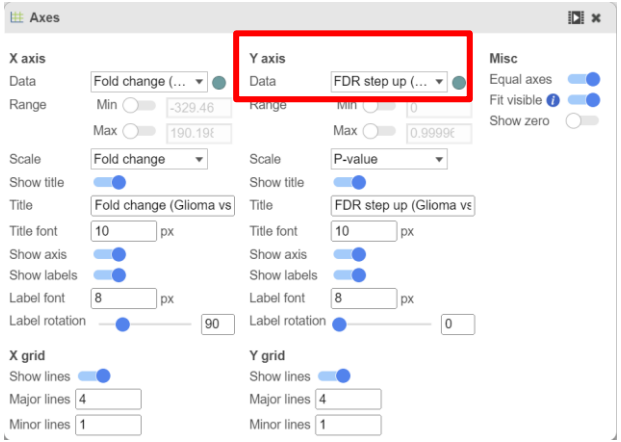
- Click **Glioma** in the top panel
- Click **Oligodendrocytes** in the bottom panel
- Click **Add comparison**
- Click **Finish** to run the GSA

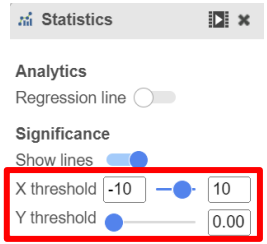


Identify differentially expressed genes

- 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**



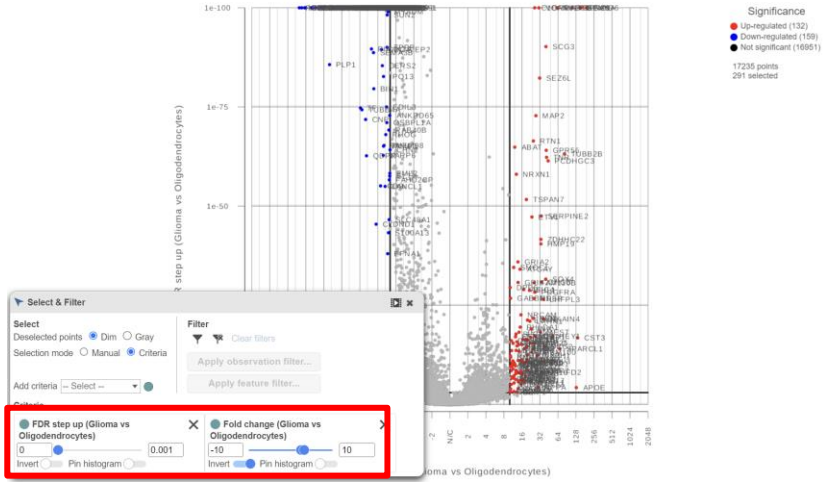




均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

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**

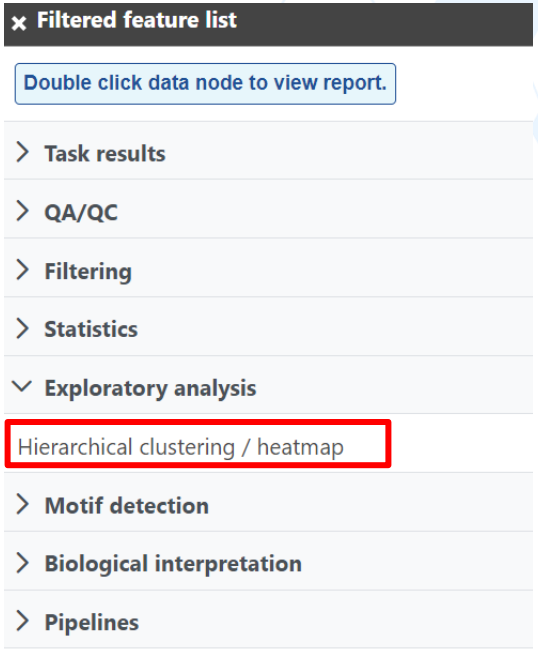


Identify differentially expressed genes

- 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 **Generate filtered node** to apply the filter and generate a Filtered Feature list node

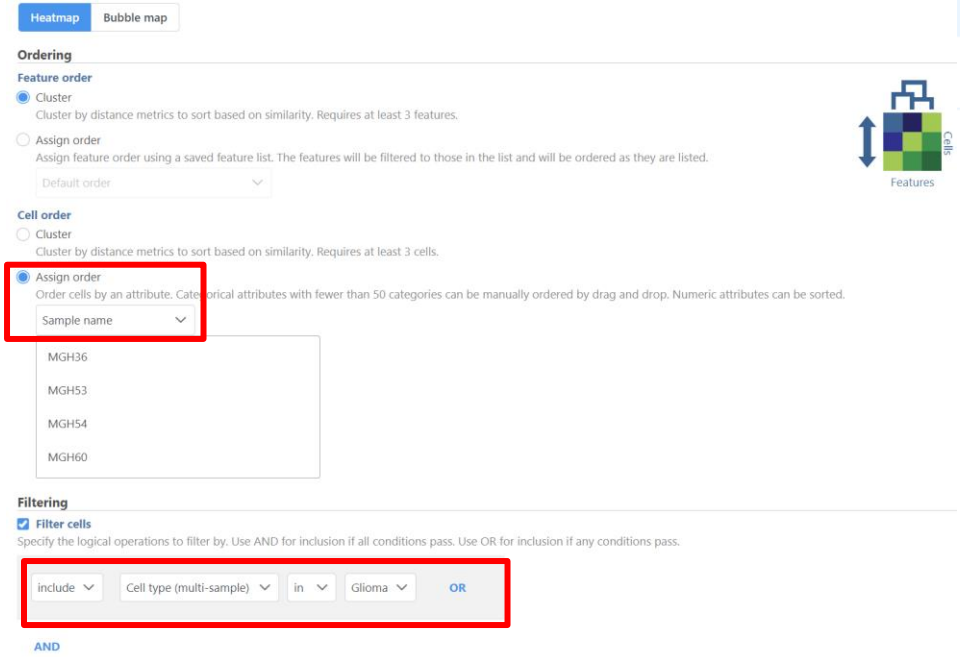
Exploring differentially expressed genes

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



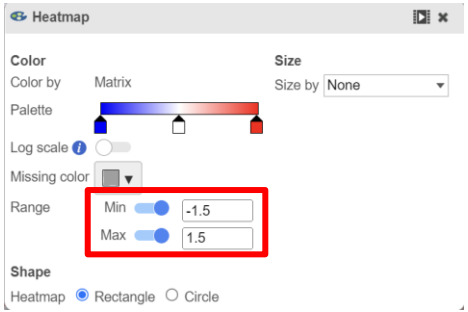
Exploring differentially expressed genes

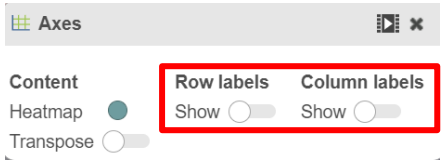
- Set the Assign order of **Cell order** as **Sample name**
- Click **Filter cells** and set the filter to **include Cell type (multi-sample) in Glioma**
- Click **Finish**

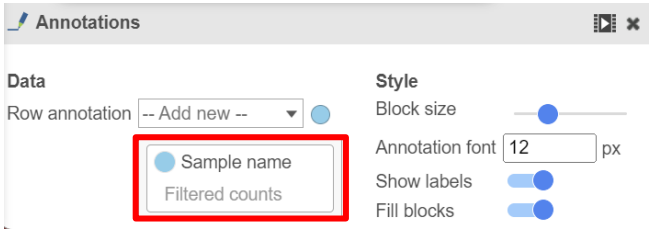


Exploring differentially expressed genes

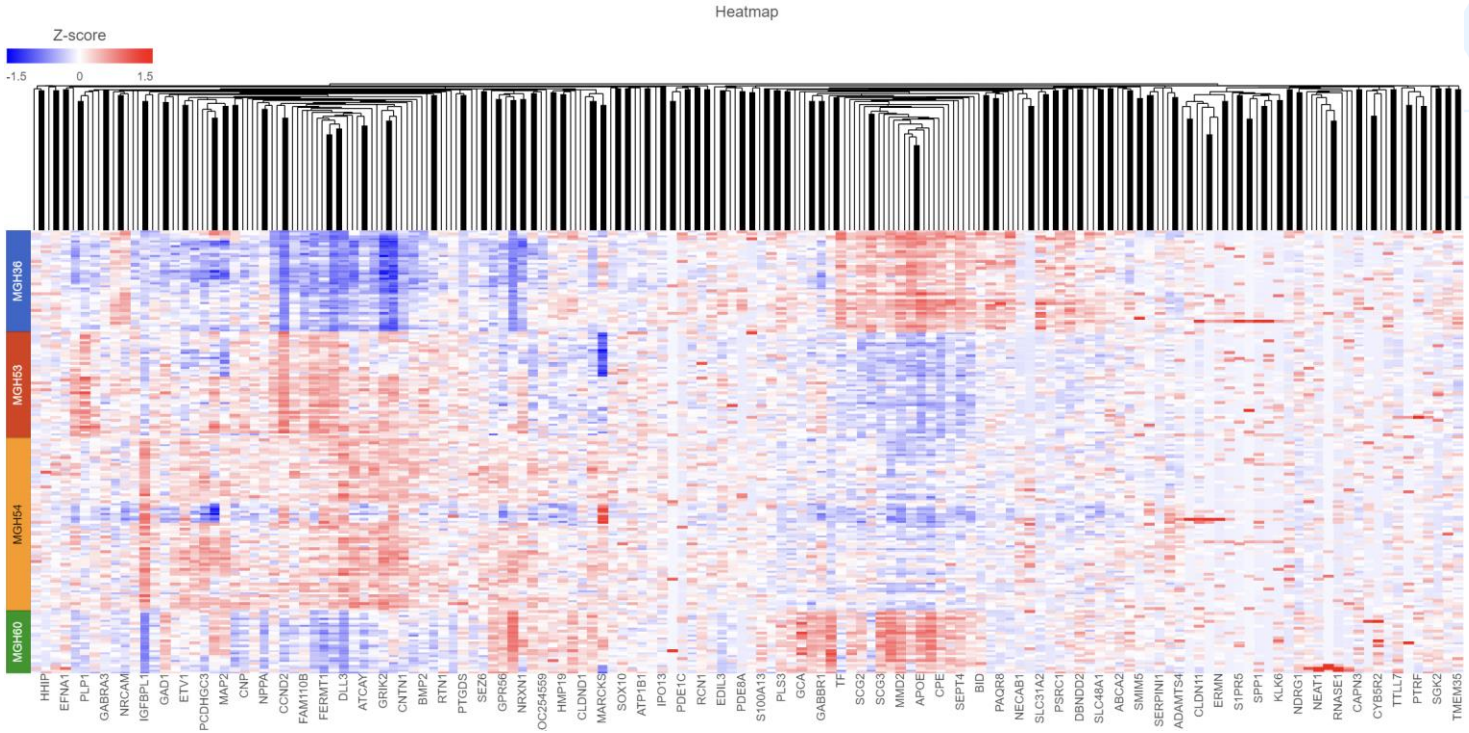
- Double click the **pink Hierarchical clustering** node to open the heat map
- Click **Heatmap** on the left
- Set **Min** to **-1.5** and **Max** to **1.5**
- Click **Axes** and deselect **Show** in both **Row labels** and **Column labels** to hide sample and feature names
- Click **Annotations** on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







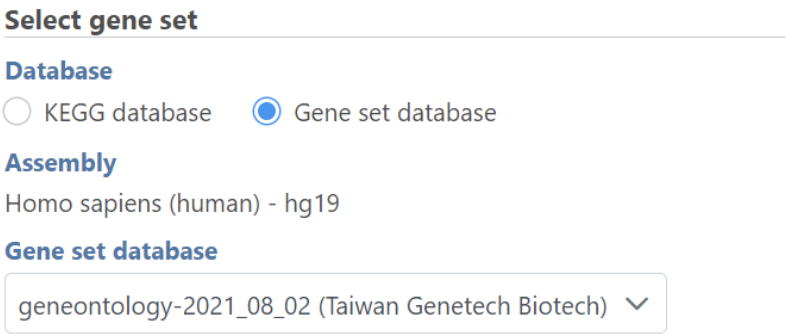
Exploring differentially expressed genes



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

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**



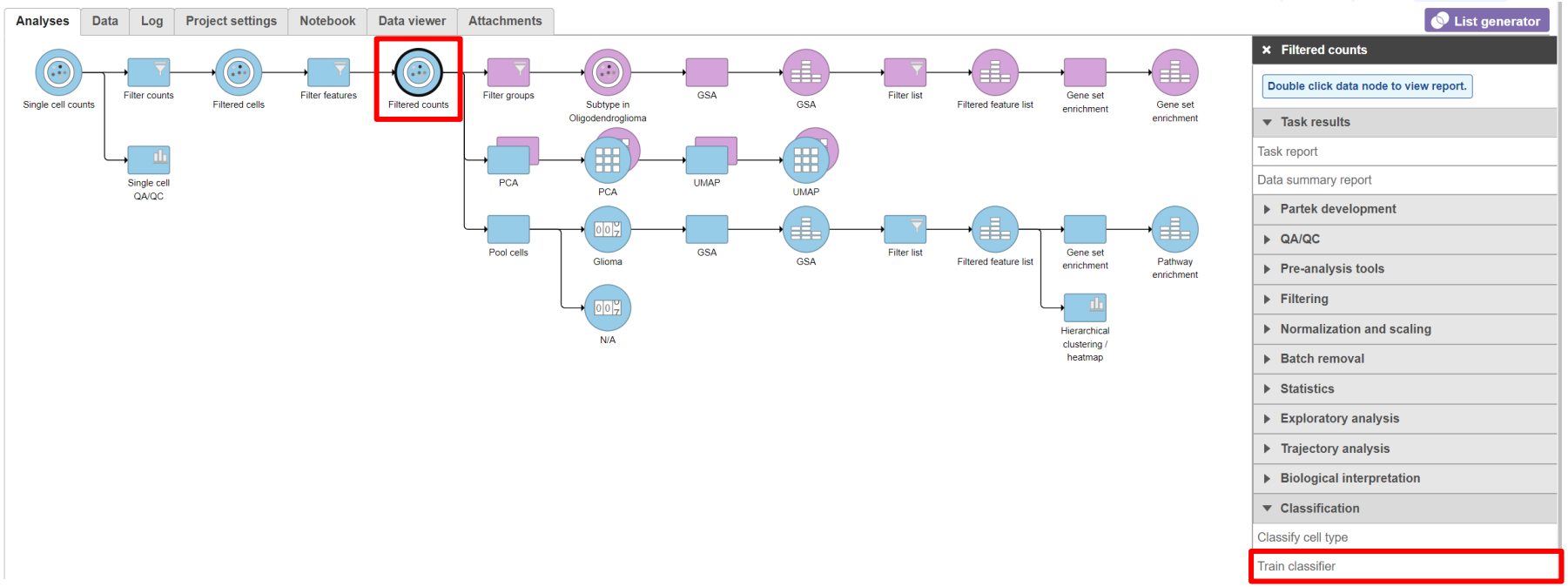
Exploring differentially expressed genes

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

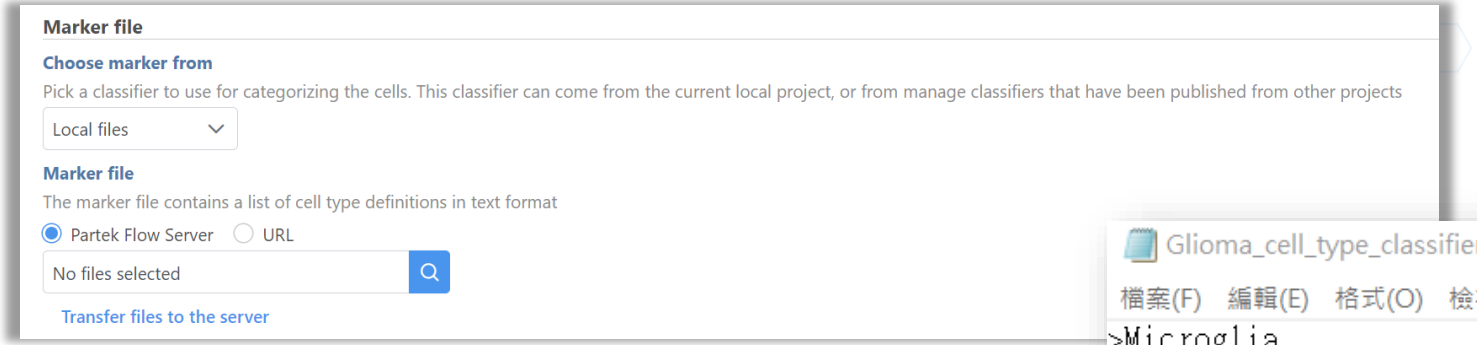
Top GO terms in the enrichment report include "ensheathment of neurons" and "axon ensheathment" , which corresponds well with the role of oligodendrocytes in creating the myelin sheath that supports and protect axons in the central nervous system.

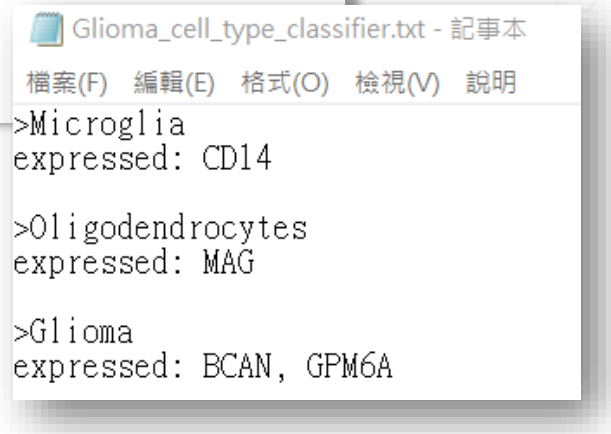
Gene set	Description	Type	Enrichment score	P-value	FDR step up	Rich factor	Genes in set	Genes in list	Genes not in list	Genes in list, not in set	Genes not in list, not in set
GO:0048731	system development	biological process	25.54	8.13E-12	1.33E-7	0.05	699	38	661	227	16,513
GO:0048856	anatomical structure development	biological process	25.14	1.21E-11	1.33E-7	0.03	3,067	92	2,975	173	14,199
GO:0032502	developmental process	biological process	23.60	5.63E-11	3.68E-7	0.03	4,784	122	4,662	143	12,512
GO:0051960	regulation of nervous system development	biological process	23.43	6.69E-11	3.68E-7	0.07	423	28	395	237	16,779
GO:0031344	regulation of cell projection organization	biological process	22.70	1.39E-10	6.11E-7	0.05	630	34	596	231	16,578
GO:0007417	central nervous system development	biological process	22.17	2.36E-10	8.66E-7	0.12	135	16	119	249	17,055
GO:0120035	regulation of plasma membrane bounded cell projection organization	biological process	21.98	2.84E-10	8.92E-7	0.05	613	33	580	232	16,594
GO:0010975	regulation of neuron projection development	biological process	21.81	3.37E-10	9.27E-7	0.06	423	27	396	238	16,778
GO:0051128	regulation of cellular component organization	biological process	21.26	5.87E-10	1.44E-6	0.03	2,233	71	2,162	194	15,012
GO:0045202	synapse	cellular component	20.88	8.54E-10	1.88E-6	0.05	821	38	783	227	16,391
GO:0045685	regulation of glial cell differentiation	biological process	20.77	9.55E-10	1.9E-6	0.16	73	12	61	253	17,113
GO:0007272	ensheathment of neurons	biological process	20.61	1.12E-9	1.9E-6	0.16	74	12	62	253	17,112
GO:0008366	axon ensheathment	biological process	20.61	1.12E-9	1.9E-6	0.16	74	12	62	253	17,112

Train Classifier



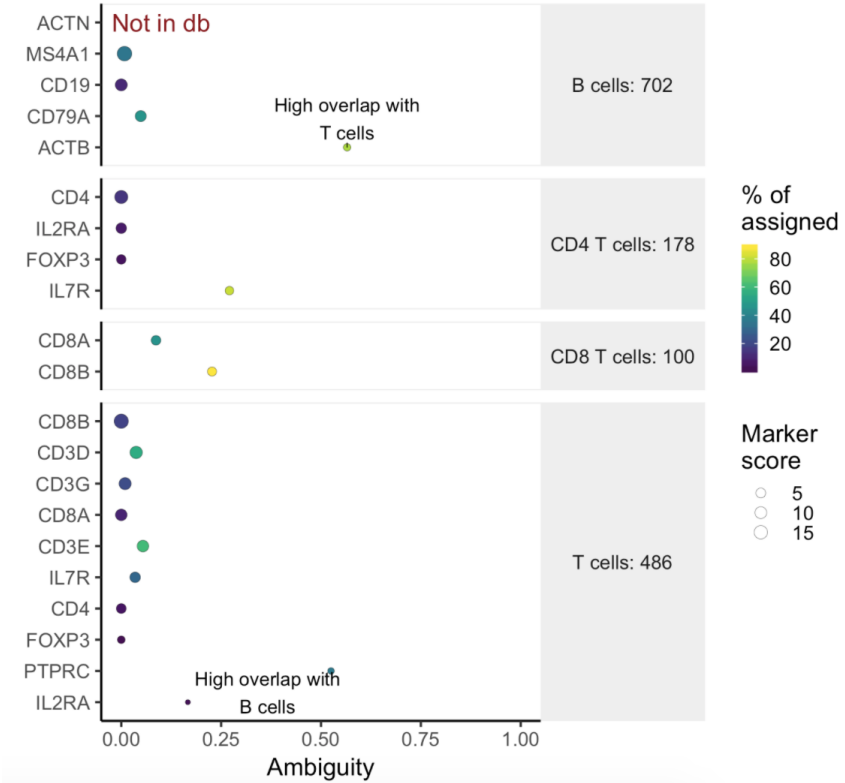
Train Classifier





There should be a space character after the colon and that there should 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

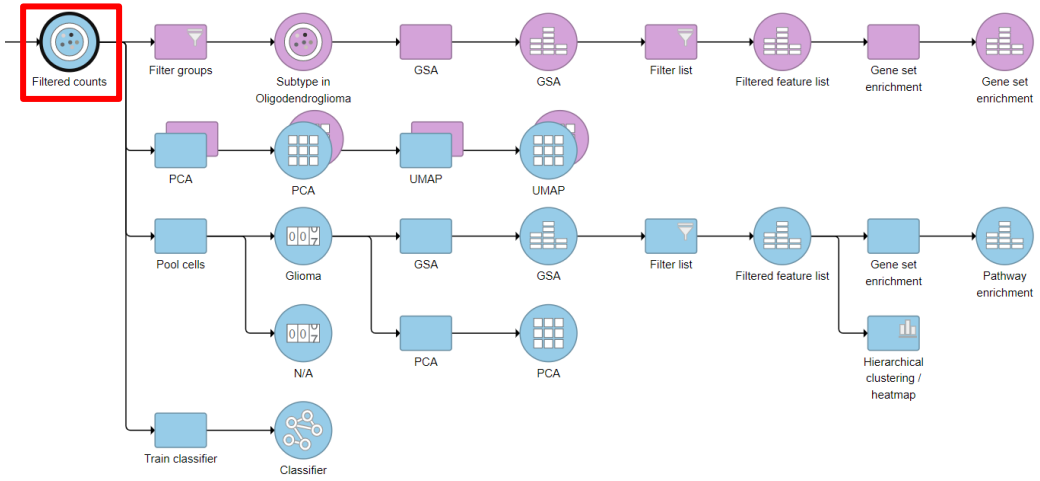
Features in classifier

Optional columns

Feature ↑↓	Glioma ↑↓	Microglia ↑↓	Oligodendrocytes ↑↓	Unknown ↑↓
(Intercept)	-39.27	9.59	13.30	16.38
BCAN	2.45	-0.90	-0.74	-0.81
GPM6A	2.56	-0.63	-0.97	-0.96
CD14	0.86	1.63	-1.24	-1.25
MAG	0.79	-0.53	4.15	-4.41

Download

Classify Cell Type





Classify Cell Type – Project classifiers

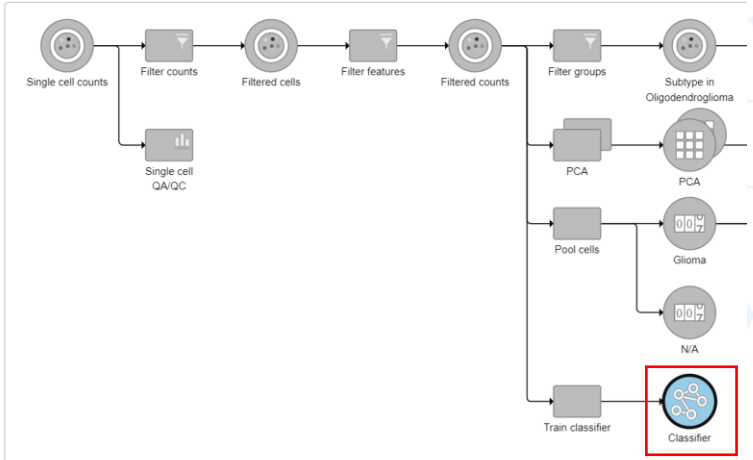
Choose classifier from Project classifiers

Garnett classifier

Select data node Clear selection

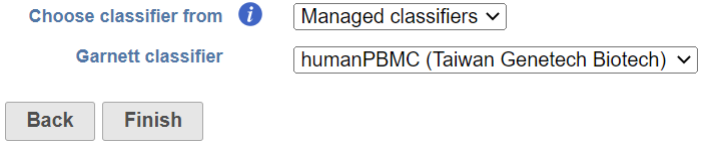
Classifier [Train classifier - 0.2.14]

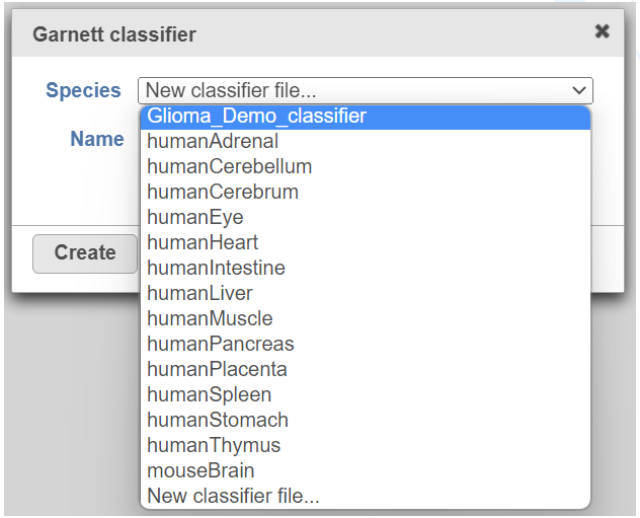
Back Finish



Select Cancel

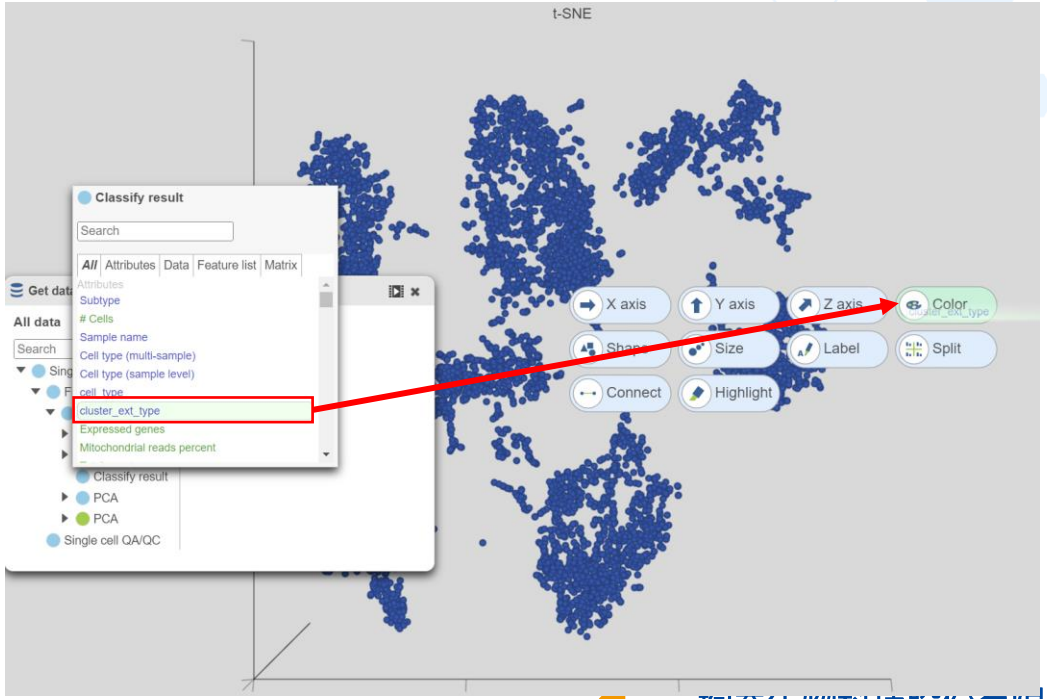
Classify Cell Type – Managed classifiers





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

