Single Cell RNA-Seq Analysis in Partek® Flow®

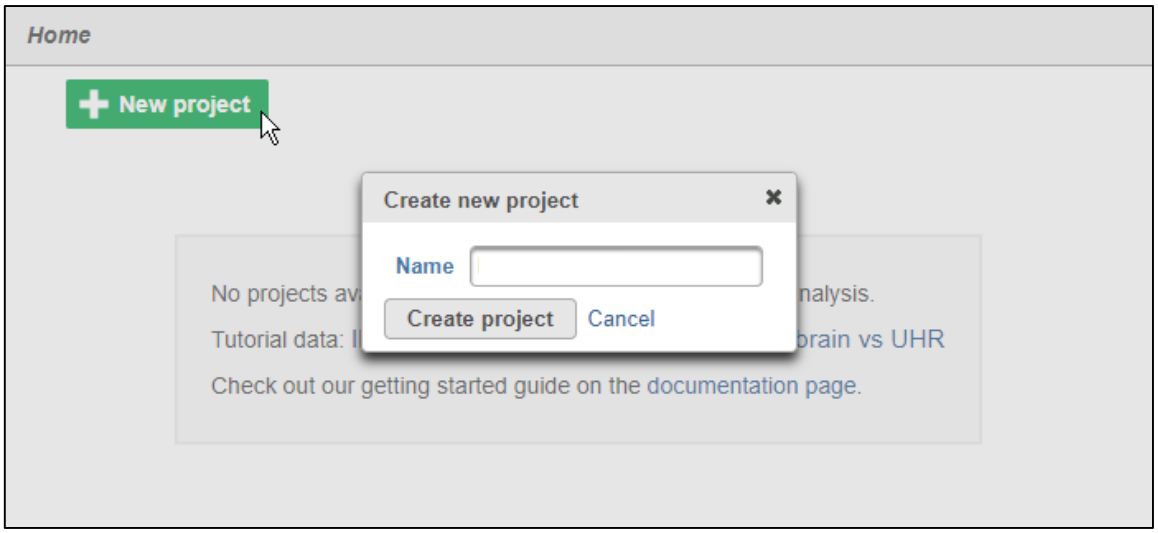
HANDS-ON TRAINING

National Institutes of Health
December 2019

Xiaowen Wang
Field Application Specialist
Partek Incorporated
support@partek.com

Login and Project Set-up

- Open your preferred web browser (Chrome, Firefox, etc. would work fine)
- Go to the server URL given by your instructor
- Log in using the username and password given to you
- This will open to the Partek Flow homepage
- Click **New Project** and enter project name: SC-RNAseq-[username]
- This will create a new project



Notes:

Experiment Description

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

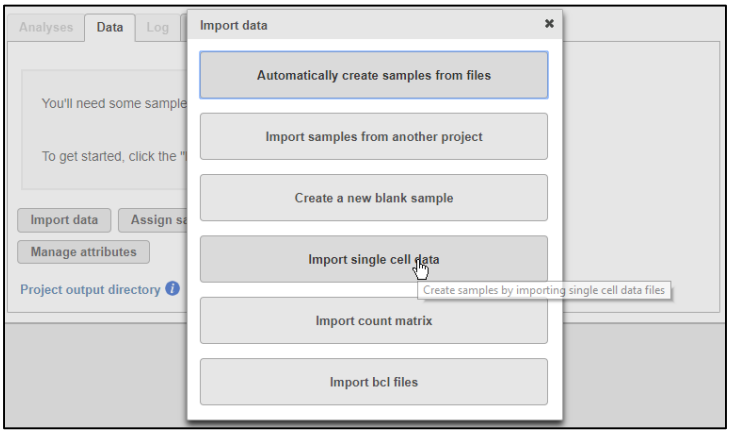
- *Goal for today: Identify different blood cell populations*

Input Files	Size	md5sum
FASTQs	27.25 GB	40274800f18c380b9bd7c25e3ed43419
Output Files format details →		
Genome-aligned BAM	21.69 GB	cd02b972a841487b09782f3b1724e42e
Genome-aligned BAM index	10.00 MB	9fbb1a8593a9421e1b5320aea76e7157
Per-molecule read information	245.20 MB	f592a76eba137b1ad07b845628a60112
Feature / cell matrix HDF5 (filtered)	17.26 MB	4dcd7861f61219bd5325190e4c5f6798
Feature / cell matrix (filtered)	41.12 MB	f741a636ede503cf65491320ed3ec719
Feature / cell matrix HDF5 (raw)	146.07 MB	7c5d7164c8e8a36fa3f6e334fb0281cd
Feature / cell matrix (raw)	85.41 MB	6616dcd7c6c5275b66ff7d6b806002c7
Clustering analysis	25.87 MB	a96940bfa361d8b52953047bb1afca79
Summary CSV	683 bytes	6178a3305959996189a48f3d5d73bb82
Summary HTML	3.77 MB	7c386237483575edccd1362e14ff7027
Loupe Cell Browser file	62.08 MB	370d0bc8472a161054ac5d6209e4de56

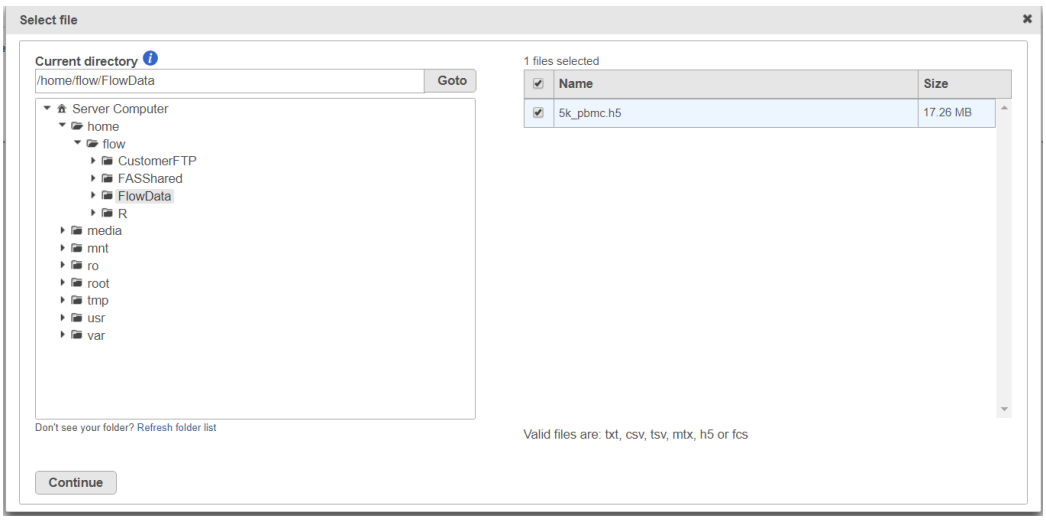
Notes:

Importing Single Cell Data

- Creating a new project automatically opens up the **Data** tab
- To import the data, click **Import data**, then click **Import single cell data**



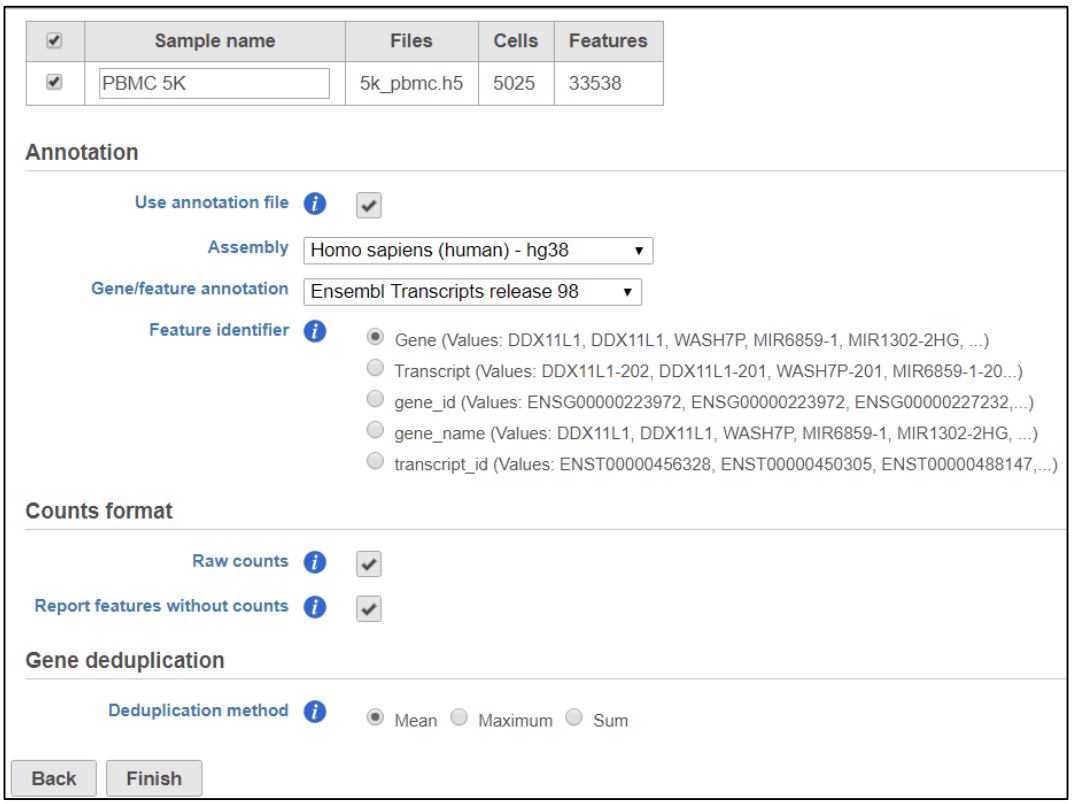
- Browse to select *5K-pbmc.h5,* click **Continue**, then click **Next**



Notes:_____

Specify Annotation

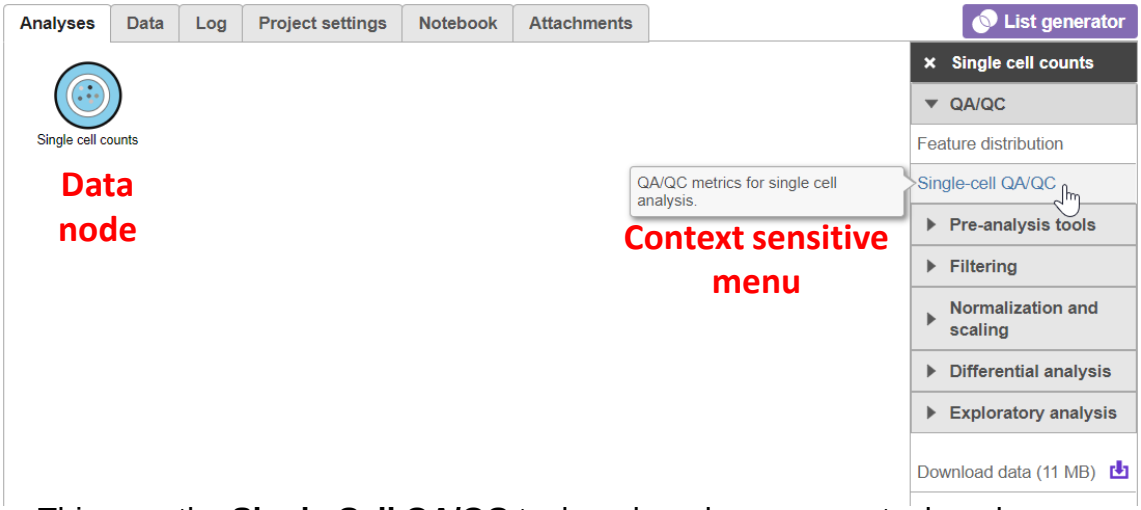
- Click the **Use annotation file** checkbox and set the annotation
  - Assembly: **Homo sapiens (human) - hg38**
  - Gene annotation: **Ensembl transcripts release 98**
- Set *Sample name* to ***PBMC 5K***
- Click **Finish** to import sample. This will create your first data node



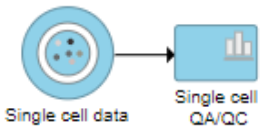
Notes:_____

Analyses Tab Overview and Running a Task

- Go to the **Analyses** tab
- Your first data node, the **Single cell data** node appears
  - *All data nodes are circles*
- Click the data node
- Clicking any node will bring up a **Context sensitive menu** on the right. Only the tasks that can be performed on that node will appear in this menu
- Select **Single Cell QA/QC** from the **QA/QC** section of the task menu



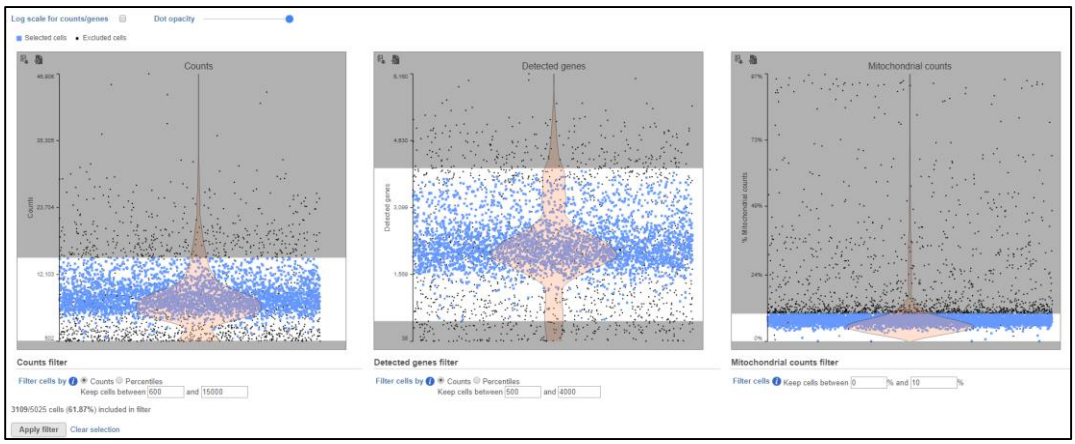
- This runs the **Single Cell QA/QC** task and produces a new task node
  - *All task nodes are rectangles*



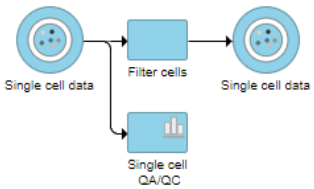
Notes:_____

Single Cell QA/AC

- Double click the Single Cell QA/QC task node to open the task report
- Single Cell QA/QC shows the most popular QC metrics used in the SC genomics community: the number of read counts per cell, detected genes per cell, and % of mitochondrial reads per cell in three violin plots
- Set the follow parameters for *Min* and *Max*
  - Total reads: **600 -- 15000**
  - Expressed genes: **500 -- 4000**
  - Mitochondrial reads **0 -- 10**
- Click **Apply filter**



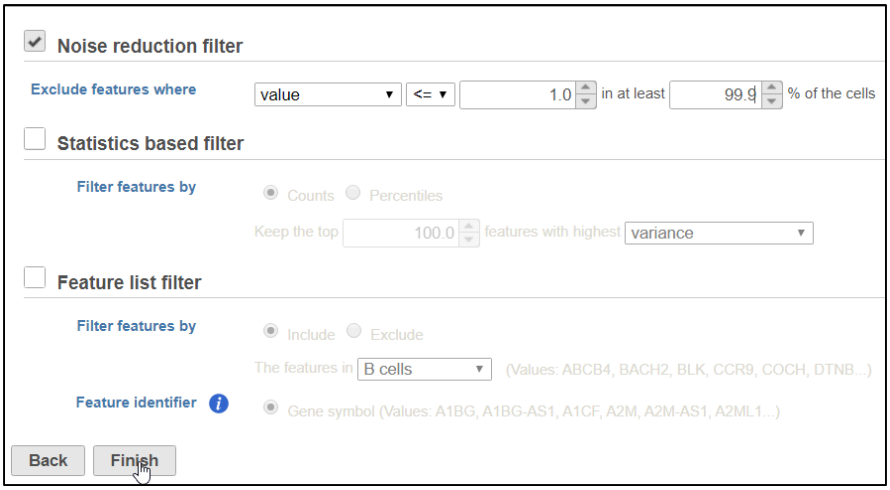
- This runs the **Filter cells** task and outputs a new **Single cell data** node



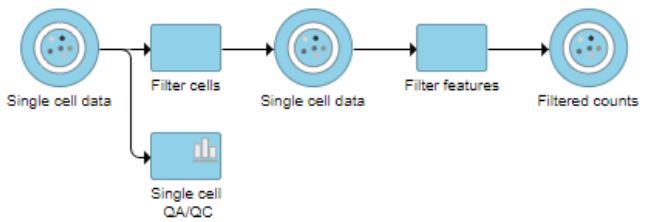
Notes:

Applying a Noise reduction filter

- Click the filtered **Single cell data** node
- Click **Filter features** in the **Filtering** section of the task menu
- This opens the **Filter features** task dialog
- Click the **Noise reduction filter** checkbox
- Create the following filter using the drop-downs and text boxes
  - Exclude features where **value <= 1** in at least **99.9**% of the cells
- Click **Finish** to apply the filer



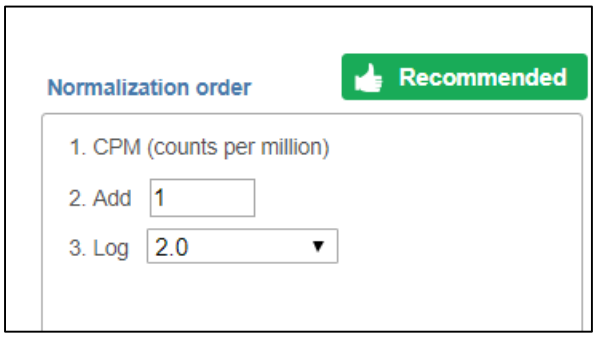
- The **Filter features** task creates a new **Filtered counts** data node



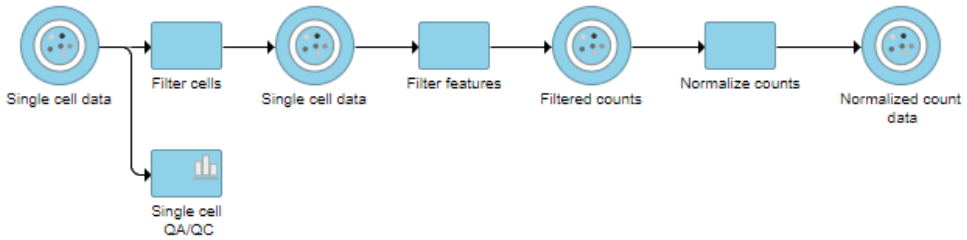
Notes:_____

Normalizing counts

- Click the **Filtered counts** node
- Click **Normalization** in the **Normalization and scaling** section of the task menu
- Click on the **Recommended** button
  - **CPM**
  - **Add 1**
  - **Log2**



- Click **Finish** to run the **Normalize counts** task



Notes:_____

Identifying Cell Types

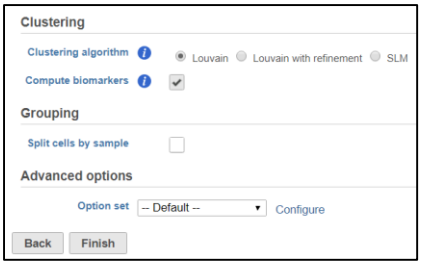
- We'll be using a combination of methods to identify some cell types commonly found in PBMCs. Namely:
  - Unbiased clustering (Graph-based)
  - Visualizing expression using
    - Canonical gene markers
    - Gene lists
  - Lassoing cell populations on the plot

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

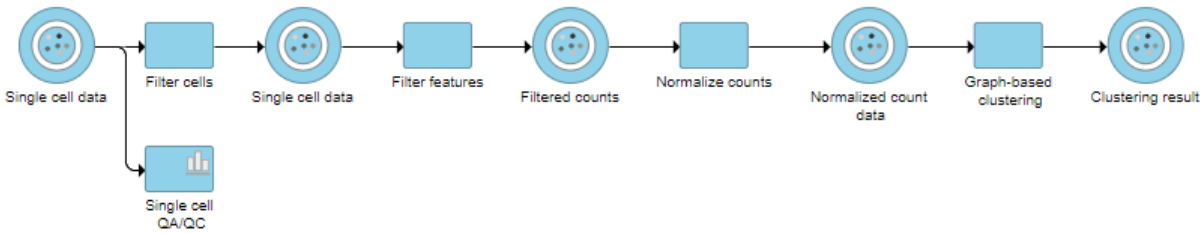
Notes:

Performing graph-based clustering

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



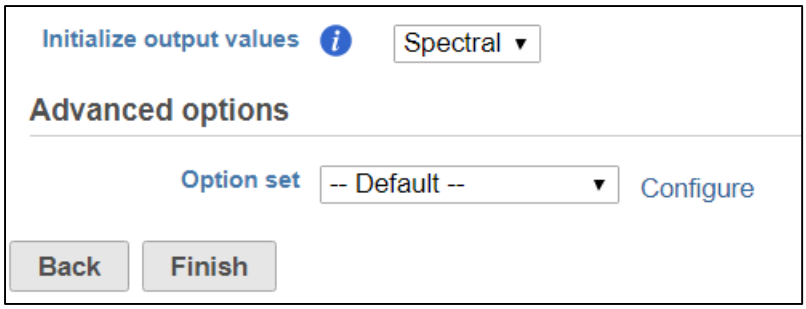
- **Graph-based clustering** produces a **Clustering result** data node

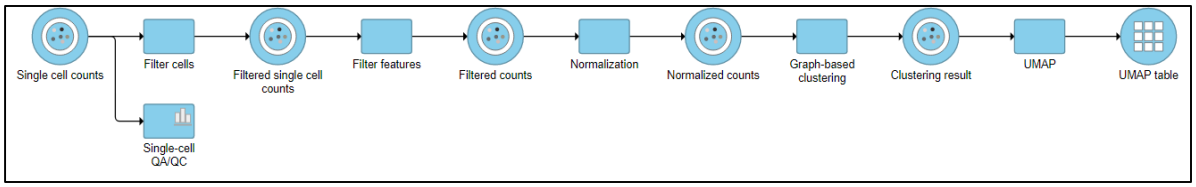


Notes:

Perform UMAP

- Click the **Clustering result** data node
- Click **UMAP** in the **Exploratory analysis** section of the task menu
- Click **Finish** to run the UMAP task with default settings
- A **UMAP table** node is produced, it contains the UMAP coordinates of all the cells



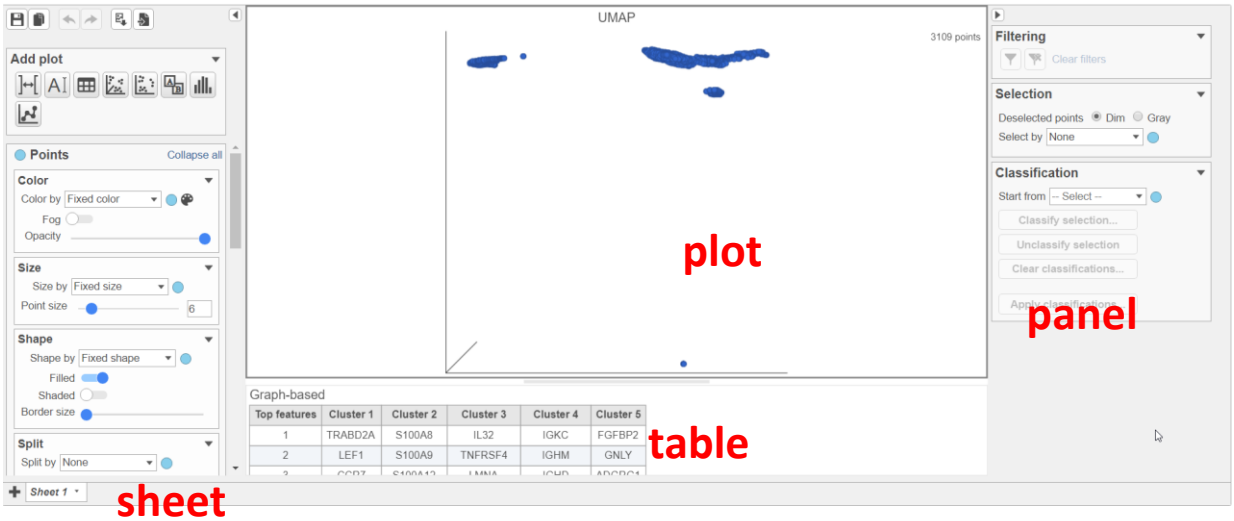


- Double click on UMAP table to open the scatter plot in *Data Viewer*
- We will use the interactive UMAP plot to view the clustering results and classify cells

Notes:_____

Data Viewer Components

- *Data Viewer* is an interactive data visualisation tool that enables you to use combine different pieces of data from one project
- *Plot:* an individual visualisation within the Data Viewer
- *Panel*
  - Configuration (left)
  - Selection (right)
- *Sheet:* one or more linked plots with shared controls
- *Data Viewer session*: a collection of one or more sheets



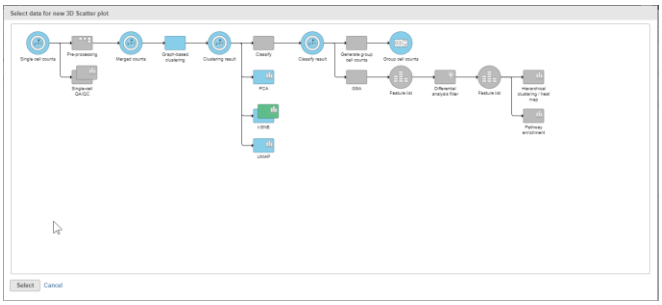
Notes:_____

Plot Types

- Use the *Add plot* card to add content to your Data Viewer session

- The buttons are (from left)
  - *Blank Space*. Adds an empty plot (no data). You may want to use this option to play with the layout
  - *Text*. Inserts a text box. Useful for captions and notes
  - Table. Inserts a table. E.g. table of biomarkers per cluster.
  - *3D Scatter Plot*. Inserts a 3D (axes) scatterplot. This is a plot showing tree data features at the same time.
  - *2D Scatter Plot*. Inserts a 2D scatter plot. This is a plot showing two data features at the same time
  - Attribute summary table. Insert a table which shows attributes
  - Histogram: histogram of a variable
  - Profile: bar or line chart on one or multiple variables

- Select a button to open the *Select data* dialog. Use it to point to the data which you would like to add to the Data Viewer. Available nodes are coloured



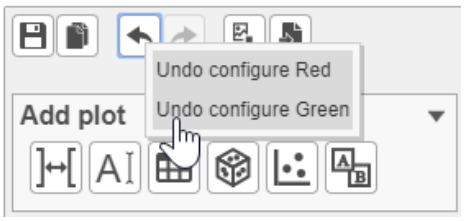
Notes:

Genaral Data Viewer Controls

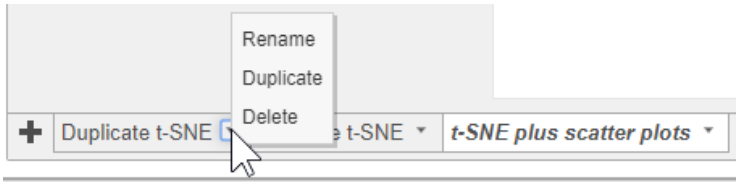
- General control buttons are in the upper left corner

- *Save*. Saves the session to the Data Viewer tab.
- *Save As*... Saves the session under a new name
- *Undo*
- *Redo*
- *Save Image*. Saves the entire canvas as an image to the local computer. Supported formats are .png and .svg
- *Send to Notebook Page*. Sends the entire canvas to the Notebook.

- *Undo* and *Redo* tools support multiple steps. Right-click on a button to see them



- Sheets are shown in the lower right corner. To add a blank sheet select the plus icon. For sheet options, click on the arrowhead



Notes:

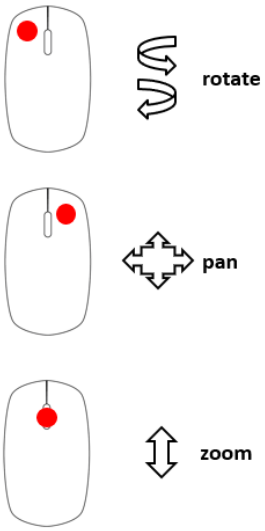
Customizing the Data Viewer Appearance

- Click on arrow icon to collapse or expand 

- Plots are separated by grippers. Mouse over a gripper to see options
  - Click and drag: resize the plot
  - Double-click: resets the plot to the center

  - To move a plot to a new position, use the handle at the top

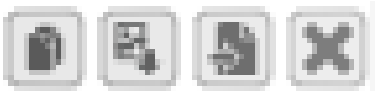
- Mouse operation on a plot



Notes:

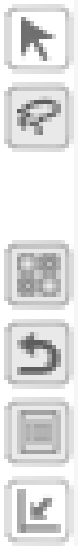
Plot Tools

- Plot tools are located in the upper right corner and appear upon a mouseover



- Horizontal series
  - *Duplicate plot*. Makes an exact replica
  - *Save Image*. Saves the selected plot only as an image to the local computer. Supported formats are .png and .svg
  - *Send to Notebook Page*. Sends the selected plot to the Notebook
  - *Remove plot*. Removes the selected plot from the canvas. Data Viewer session remains open

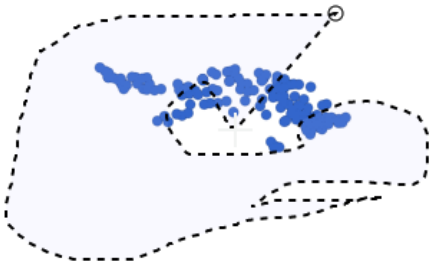
- Vertical series
  - *Pointer mode*. Click to select data points
  - *Lasso mode.* Draws a lasso to select data points
  - *Invert Selection*. Unselects the currently selected cells, and selects the others
  - *Reset View*. Resets plot rotation and zoom
  - *Toggle Legend*. Turns legend on and off
  - *Toggle Axes Autoscale.* Scales the axes with respect to the visible data points (e.g. after some data points have been filtered out)



Notes:_____

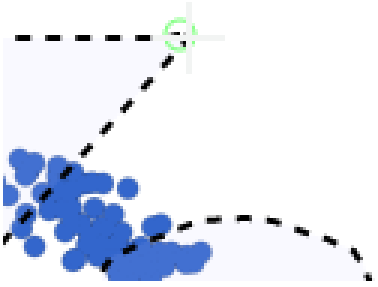
Lasso Mode

- Click and hold to make a curvy boundary



- Click and lift to make a polygon boundary

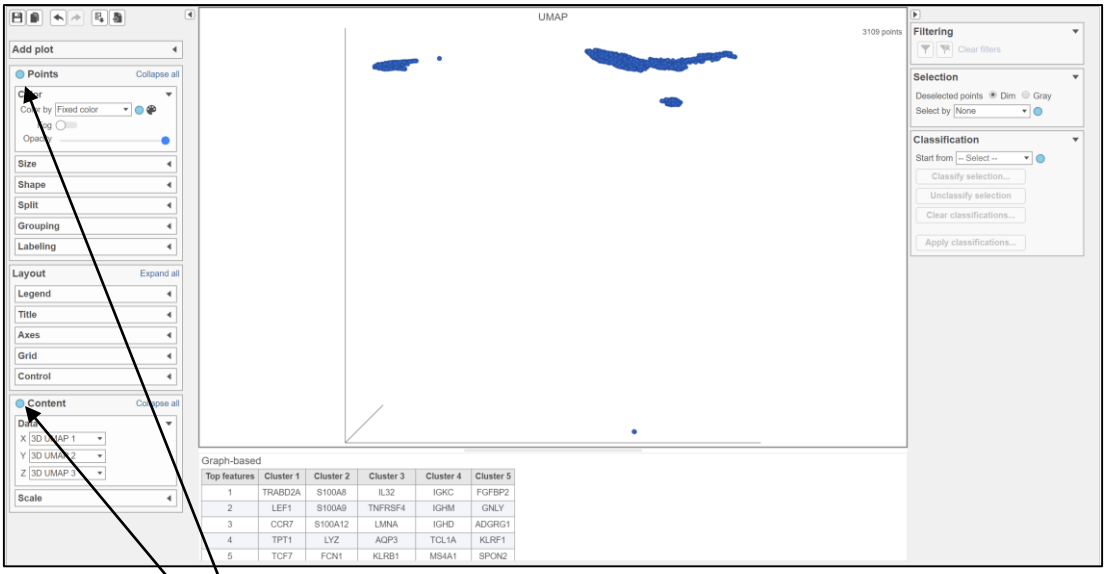
- Click on green circle (origin) or double click ot close hte gate



Notes:

UMAP Scatter Plot

- When double click on the UMAP table, it opens 3D scatterplot of the cells:
  - Axes: UMAP scores from *UMAP table* data node
  - Point rendering information is from the *Clustering result* data node
  - Biomarker table: from *Clustering result* data node



Click on the circle to change the data source

Notes:

Cell Rendering

Color

- Cell attributes and features can be used to color cells
- Click on *Color by* to choose Graph-based classification attribute
- Drag a marker gene from the table to plot to color the cells by gene expression
- Select one gene, two genes, three genes at a time to color cells
- Select a list of genes (if there are more than 3 genes) to color cells

Size

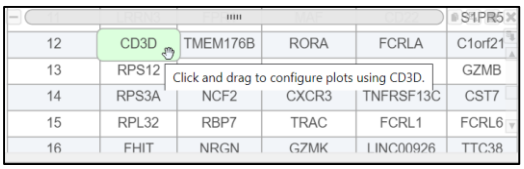
- Only cell attributes can be used to size cells

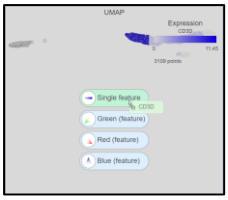
Shape, Split, Grouping

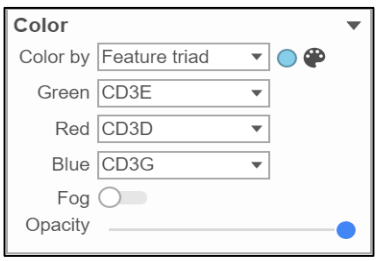
- Only categorical attributes can be used to shape cells

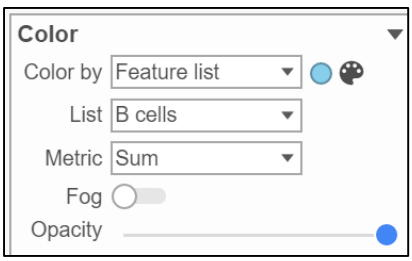
Labeling

- Only show on selected points within 100 points in the data source





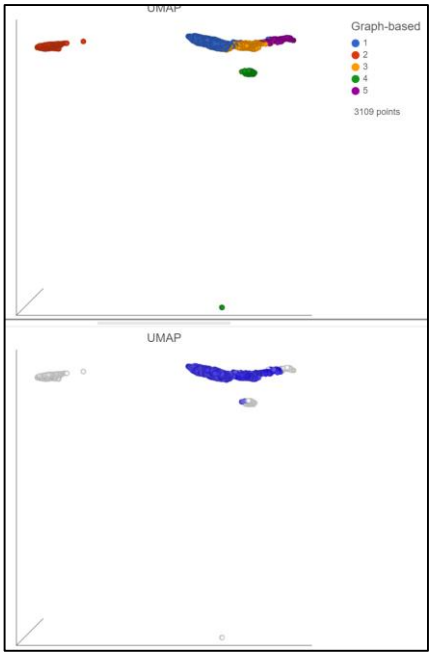


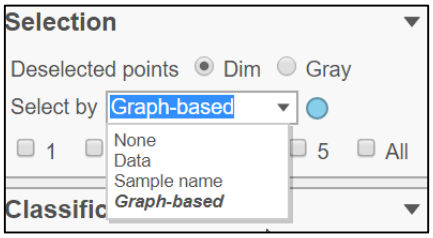


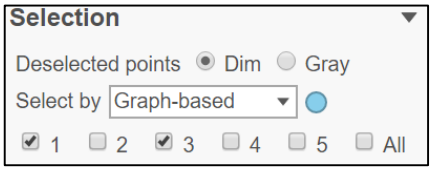
Notes:

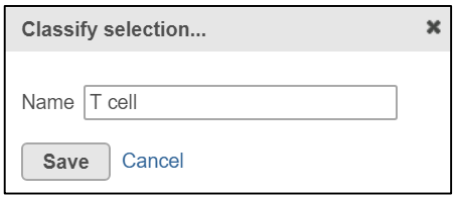
Classify T cells

- Color the cells using graph-based classification
- Duplicate the UMAP plot by clicking 
- On the 2nd plot, use CD3D to color the cells, cells having high expression on CD3D genes are mainly in cluster 1 and 3 in graph base classification.
- Select **Graph-based** from drop-down list in Selection
- Choose **1** and **3** and click **Classify selection** button in **Classification**, specify the name of selected cells as ***T cell*** and click **Save**





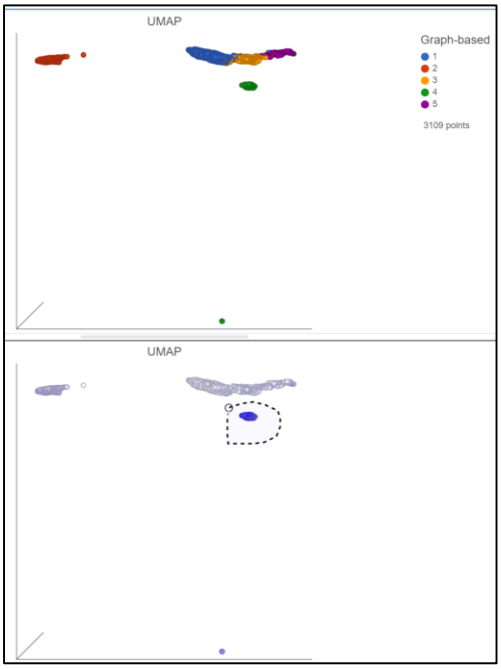


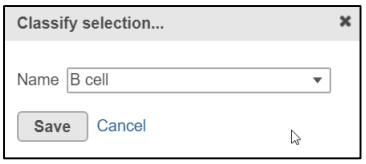


Notes:_____

Classify B cells

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



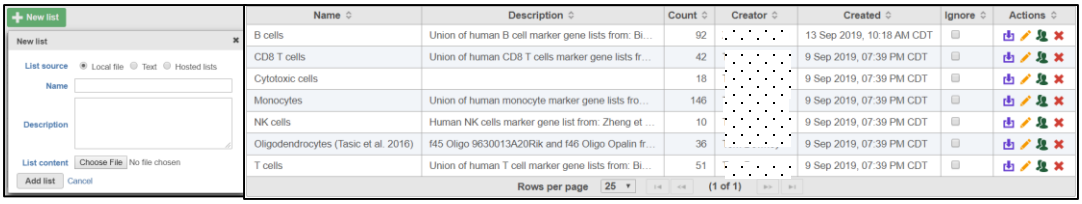


Notes:

List Management

The List management page allows users to create, view, and edit lists. List can be used to filter data, configure plots etc.

- Click on **username >Settings**, choose **List management** in Partek Flow components section on the left panel
- Choose New List to create a new list
  - List should have a unique for each user
  - Description section is optional
  - Local file --- List content can be generated from a text file which contains only one column with a list of names
  - Text – list of names can be typed in directly, one name per line
  - Hosted lists – Partek hosted some genes list generated from published paper
- List actions:
  - List can be downloaded as a text file
  - List name and description can be edited, list content cannot be edited
  - List can be deleted
  - List can be shared: lists downloaded from Partek hosted will be shared with everybody
  - Shared list can be Ignored by the user

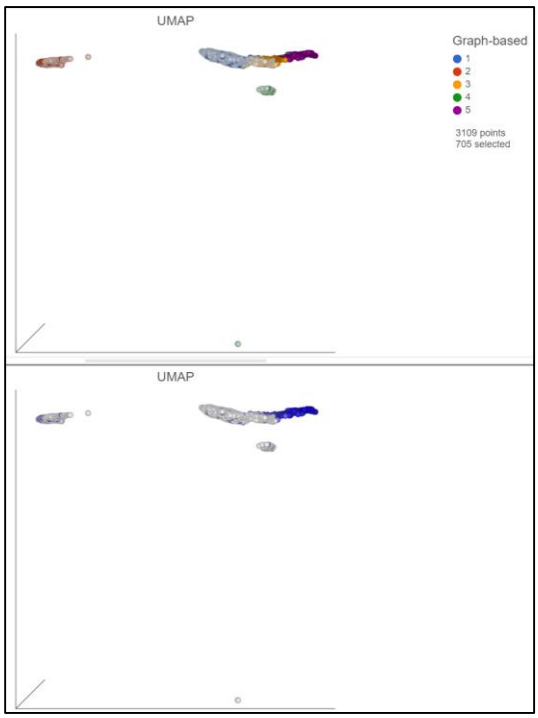


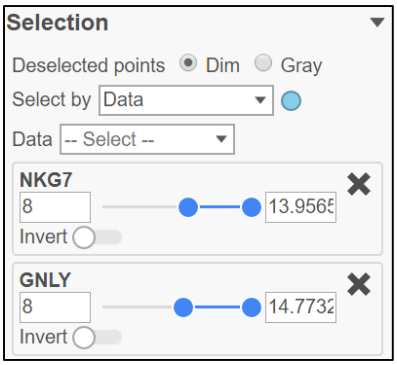
Notes:

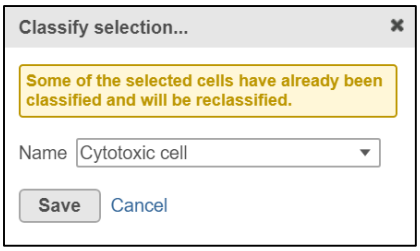
Classify Cytotoxic cells

- Select the 2nd UMAP plot, choose **NKG7** to color the cells from the biomarker table
- Choose **Data** from Select by drop-down list, and select **NKG7**
- Specify the *min* as **8** to select cells whose NGK7>=8
- Add GNLY from the drop-down list and specify GNLY>=8
- Click **Classify selection** to name it as **Cytotoxic cell**, click **Save**
- Any number of genes can be used to build the rule

Note: if cells were classified previously, the new class label will overwrite the previous class label



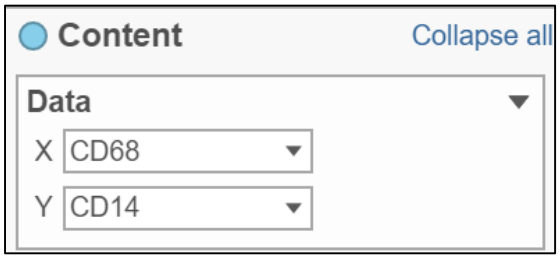


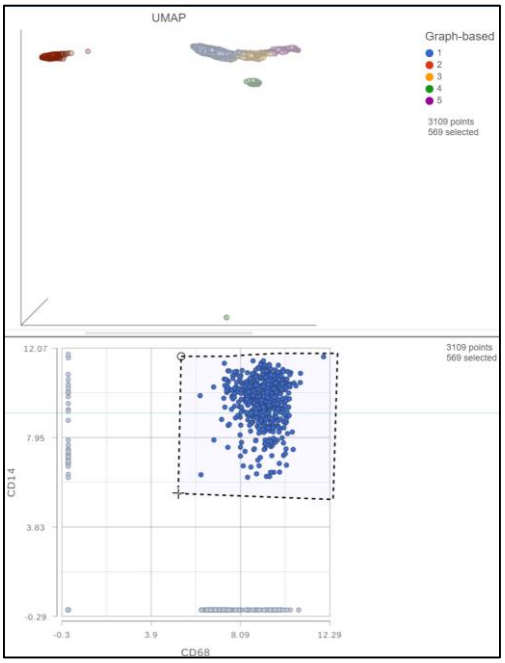


Notes:_____

Classify Monocytes

- Close the 2nd UMP plot by clicking on
- Add a 2D plot from **Normalized counts** data node
- In the Content card, specify marker genes of monocytes on the axes:
  - X-axis: **CD68**
  - Y-axis: **CD14**
- Use lasso tool to select cells with high expression on both genes (upper-right corner)
- Click **Classify selection**, name it as **Monocyte** and **Save**

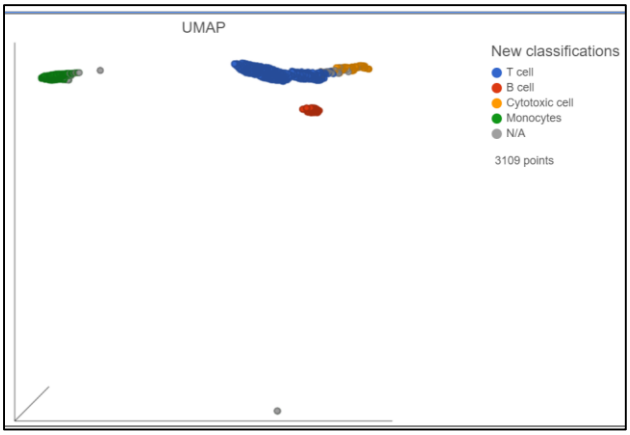


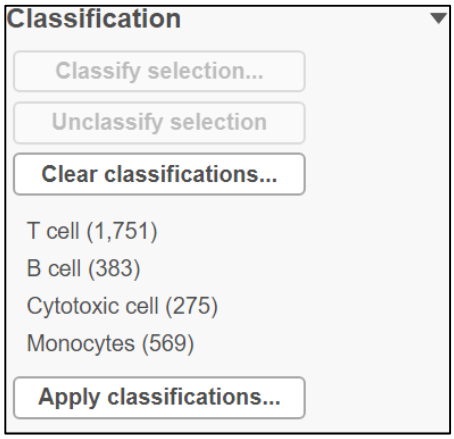


Notes:

Viewing classifications

- Click on the UMAP plot, choose Color by **New classification**
- Click **Apply classification…** button in *Classification* card to generate a new data node
- Select *Clustering result* data node as input data

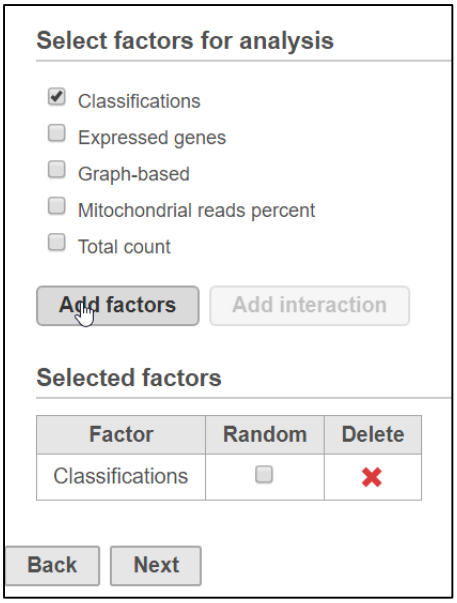


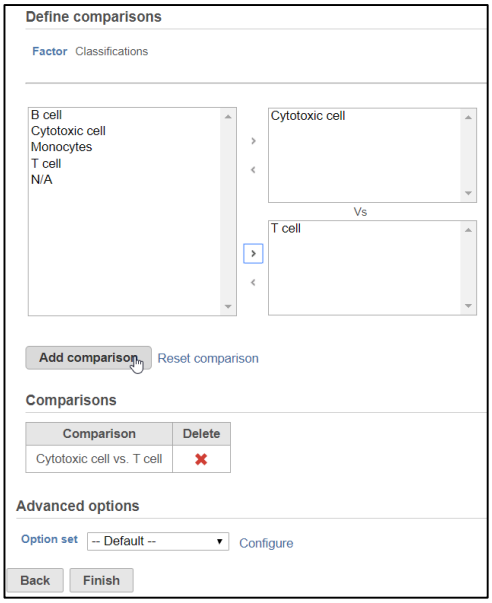


Notes:

Identifying differentially expressed genes

- Now that we have classified cells into cell types, we can compare expression between cell types
- Here, we will compare **Cytotoxic cells** and **T cells** to identify genes that are differentially expressed between these cell types
- Click the **Classified groups** node produced by the **Classify cells** task
- Click **ANOVA** in the **Differential analysis** section of the task menu
- Choose **Classifications** and click Add factors
- Click **Next**
- Choose to compare Cytotoxic cell vs T cell, click **Add comparison**
- **Click Finish**

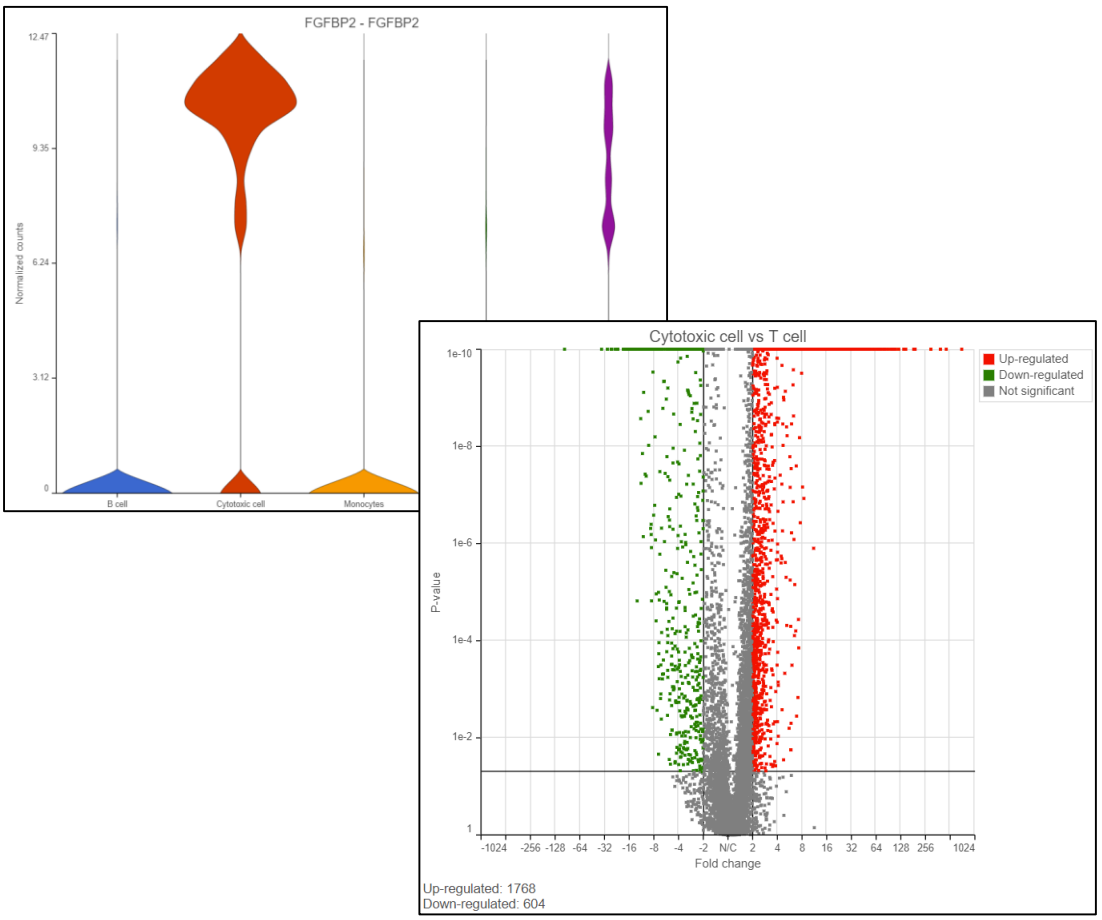




Notes:

Viewing results

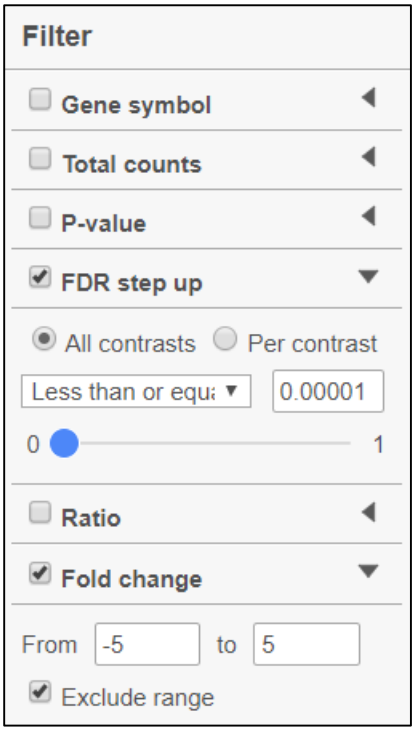
- Double click the **Feature Lists** data node to open the ANOVA report
- The **Gene list** table in the ANOVA report lists every gene that was
- *Genes are listed starting with the lowest p-value*
- Click the icon next to a gene under **View** to open a violin plot
- Click the icon to invoke volcano plot



Notes:

Filtering results

- To identify significantly differentially expressed genes, we can use the **Filter** on the left-hand side of the table
- Set **FDR step up** to **1e-5** and **Fold change** to **-5** to **5**
- The number of genes in the table changes with the filter applied



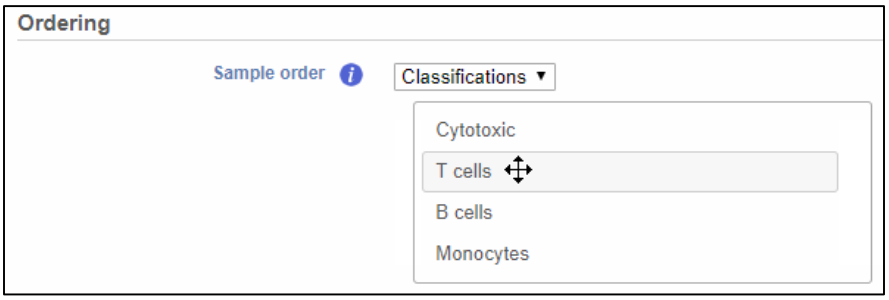
- Click  to run the **Differential analysis filter** task

Notes:

Configuring Hierarchical clustering

- To visualize the differentially expressed genes on our filtered list, we will create a hierarchical clustering heat map
- Click the **Feature list** node generated by **Filter list**
- Click **Hierarchical clustering** in the **Exploratory analysis** section of the task menu
- Uncheck **Cluster samples**
- Check **Filtering** and set to **Include Classification is T cells** OR **Include Classification is Cytotoxic cells**

- Under Ordering select **Classifications** from the **Sample order** drop-down menu to order cells by their classification

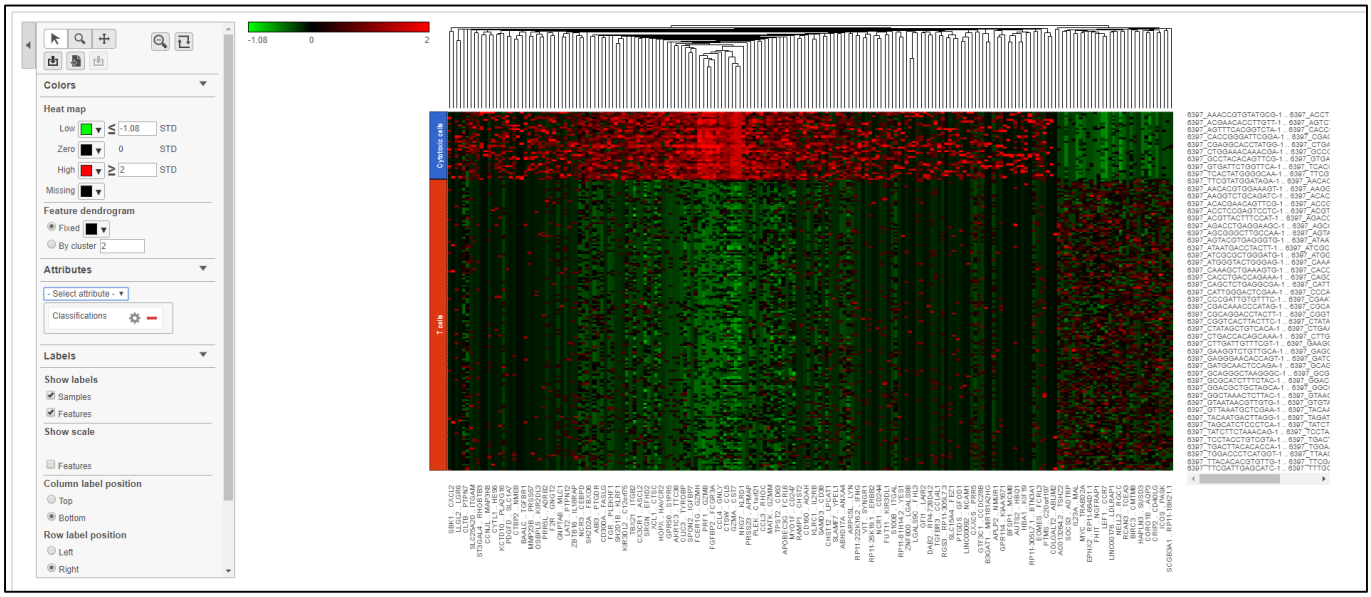


- Click **Finish** to run **Hierarchical clustering**

Notes:

Hierarchical clustering heat map

- Double-click the **Hierarchical clustering** node to open the heat map
- Set the **High** value to 2 to balance the colors
- Select **Classifications** from the **Attributes** drop-down menu to label cells with their classification



- Click the save image button to download the heat map as a publication-quality image

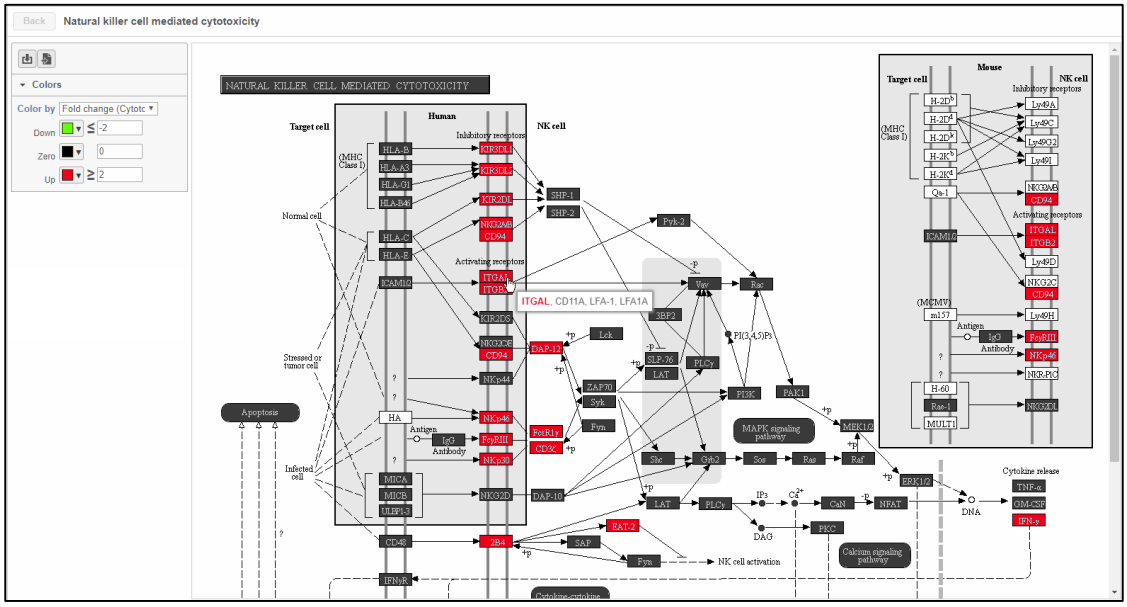
Notes:

Biological interpretation

- We can use **Biological interpretation** tools to learn more about the differentially expressed genes between **Cytotoxic cells** and **T cells**
- Click the **Feature list** node generated after filtering
- Click **Pathway analysis** in the **Biological interpretation** section of the task menu
- Click **Finish** to run enrichment analysis
- Double-click on the **Pathway enrichment** task node to view the report

Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
path:hsa04650	Natural killer cell mediated cytotoxicity	34.85	7.29E-16	19	108	
path:hsa05332	Graft-versus-host disease	21.15	6.52E-10	9	29	
path:hsa04060	Cytokine-cytokine receptor interaction	13.64	1.19E-6	15	256	

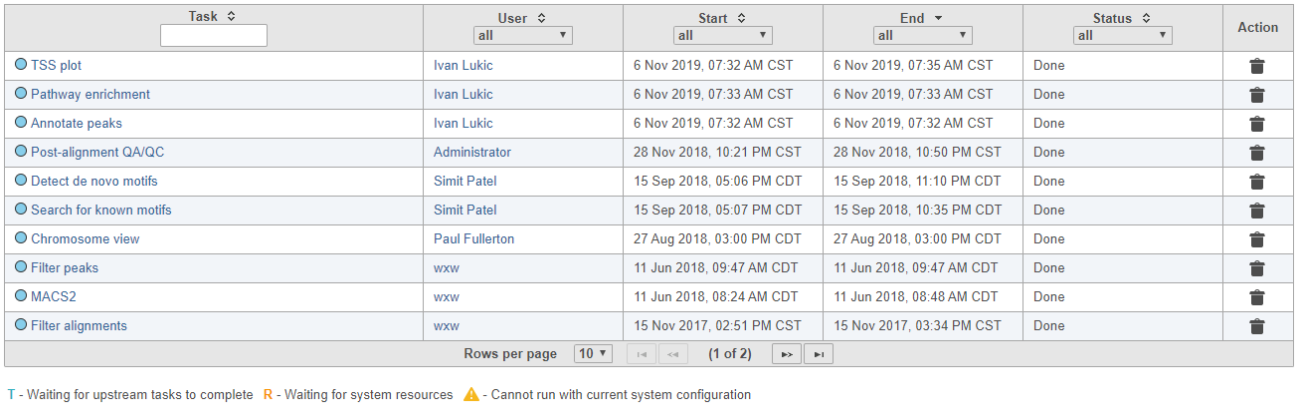
- The links on the table open to KEGG pathway maps overlaid with your differential gene expression results



Notes:_____

Log Tab

- Log tab lists all the tasks (current and past) within a project. Task names are links to Task details page for each task
- Task in progress can be stopped by using the stop button in the *Actions* column. Completed tasks can be deleted by selecting the **bin** icon



Task	User	Start	End	Status	Action
	all	all	all	all	
TSS plot	Ivan Lukic	6 Nov 2019, 07:32 AM CST	6 Nov 2019, 07:35 AM CST	Done	
Pathway enrichment	Ivan Lukic	6 Nov 2019, 07:33 AM CST	6 Nov 2019, 07:33 AM CST	Done	
Annotate peaks	Ivan Lukic	6 Nov 2019, 07:32 AM CST	6 Nov 2019, 07:32 AM CST	Done	
Post-alignment QA/QC	Administrator	28 Nov 2018, 10:21 PM CST	28 Nov 2018, 10:50 PM CST	Done	
Detect de novo motifs	Simit Patel	15 Sep 2018, 05:06 PM CDT	15 Sep 2018, 11:10 PM CDT	Done	
Search for known motifs	Simit Patel	15 Sep 2018, 05:07 PM CDT	15 Sep 2018, 10:35 PM CDT	Done	
Chromosome view	Paul Fullerton	27 Aug 2018, 03:00 PM CDT	27 Aug 2018, 03:00 PM CDT	Done	
Filter peaks	wxw	11 Jun 2018, 09:47 AM CDT	11 Jun 2018, 09:47 AM CDT	Done	
MACS2	wxw	11 Jun 2018, 08:24 AM CDT	11 Jun 2018, 08:48 AM CDT	Done	
Filter alignments	wxw	15 Nov 2017, 02:51 PM CST	15 Nov 2017, 03:34 PM CST	Done	

Rows per page 10 (1 of 2)

T - Waiting for upstream tasks to complete R - Waiting for system resources ⚠ - Cannot run with current system configuration

Time estimates are being continuously updated and will become more accurate.

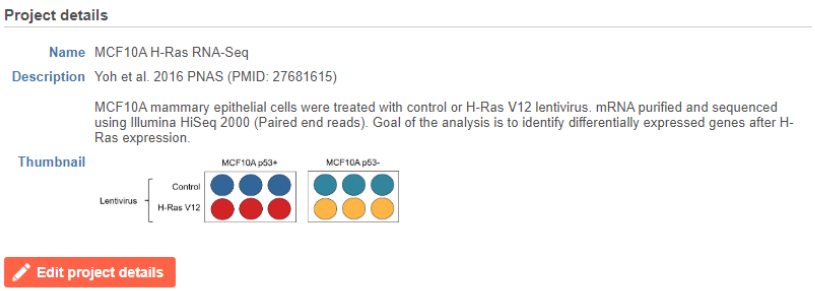
Notes:

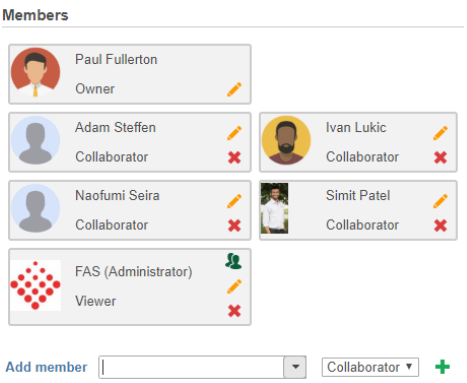
Project Settings Tab

- Projects settings tab is composed of two parts and contains general information on the project

- Project details
  - Optional metadata can be added to a project (**Edit project details**), such as a *Description* and a *Thumbnail* (the thumbnail appears on the home page)
  - A project can also be renamed by using the **Edit project details** and chaning the *Name* field

- Members
  - List of existing Partek Flow users which have access to the project
  - To remove a user from the project, click on the red **X** icon
  - To add a user to the project, click on the **Add member** drop down list

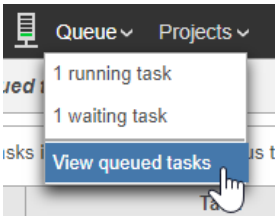




Notes:_____

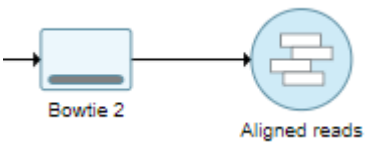
Other GUI Features: Tasks in Progress

- To monitor Partek Flow queue go to **Queue > View queued tasks**

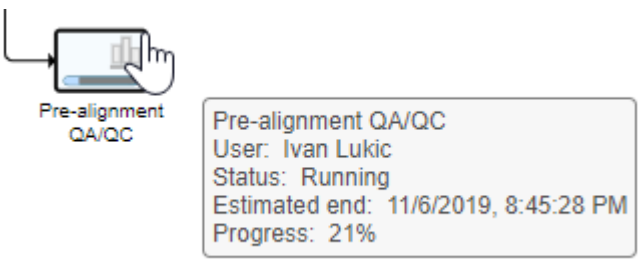


- The *Queue* shows tasks from all the projects, while the *Log* tab of a project shows only tasks launched from that project

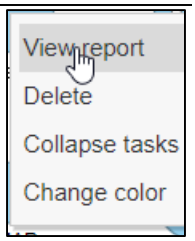
- A task in progress is shown as translucent and has a progress bar at the bottom. Once the task completes, the color changes. To move forward with analysis you do not need to wait on the task completion; you can work with data nodes of tasks in progress



- Mousing over a task in progress provides basic info on the task



Other GUI Features: Task Options

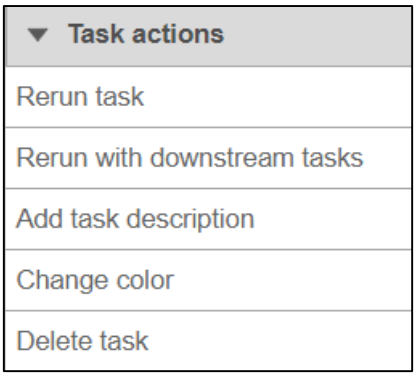


Right click on a task (rectangle) to:

- View report: get the task detail page
- Delete: delete all downstream pipeline
- Collapse task: choose another task to collapse the pipeline in between
- Change color: change the color of the selected task or all the downstream pipeline

Left click on a task to choose options from the menu:

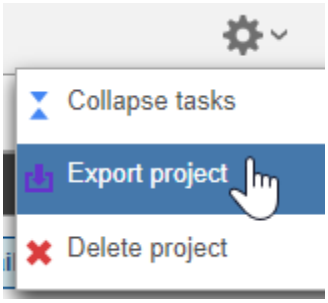
- Rerun task: you can change the parameters to re-run
- Rerun with downstream tasks: you can change the parameters of the selected task but keep the downstream tasks the same as previous one
- Add task description: once added text to a task, when mouse over the task, the text will display



Notes:_____

Other GUI Features: Project Operations

- An entire project (including all the data, the pipeline, and the annotation file) can be exported from Partek Flow using the **Export project** tool from the project



- Alternatively, a project can be exported from the *Home* screen 

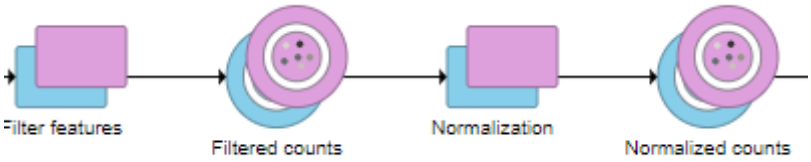
- To imported an exported project, use **Projects > Import project**

- To delete a project, use the **Delete project** tool within the project (see above) or the delete button on the *Home* screen

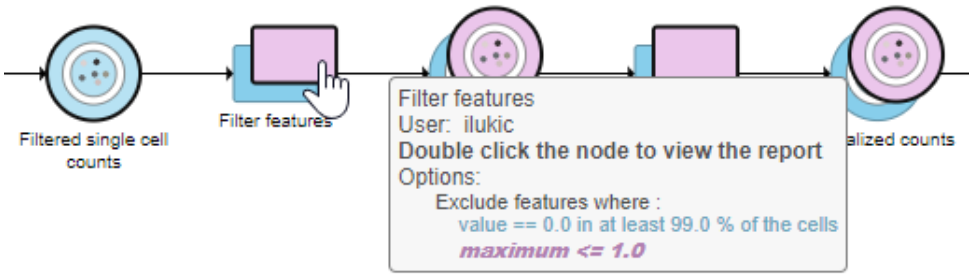
Notes:

Other GUI Features: Layers

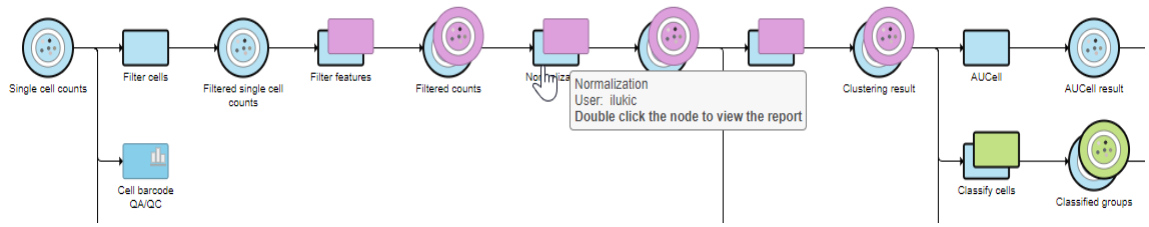
- Identical tasks ran with different task options are represented using layers of different colour. The image below shows two layers (blue and pink)



- To quickly tell a difference between the layers, mouse over the first task in a new layer. The figure below shows that *Filtered single cell counts* were further filtered (*Filter features*) using two different criteria. Baloon indicates the difference in filter settings between the blue and the pink layer



- Hovering over different parts of workflow bolds the workflow

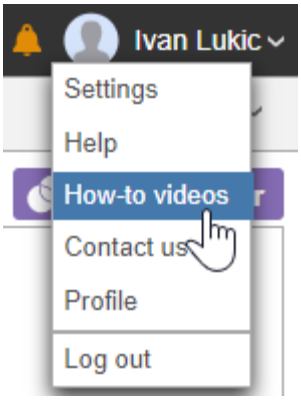


Notes:_____

Getting Help

Self-learning

- Partek Flow documentation https://documentation.partek.com/display/FLOWDOC/Partek+Flow+Documentation
- Step by step tutorials + practice data sets https://documentation.partek.com/display/FLOWDOC/Tutorials
- Recorded webinars https://documentation.partek.com/display/FLOWDOC/Webinars
- Partek blog page https://www.partek.com/blog/
- Tips and tricks on Partek Flow are regulary tweeted https://twitter.com/Partek_Inc
- *How-to videos* are accessable from the **Settings** menu



Technical Support

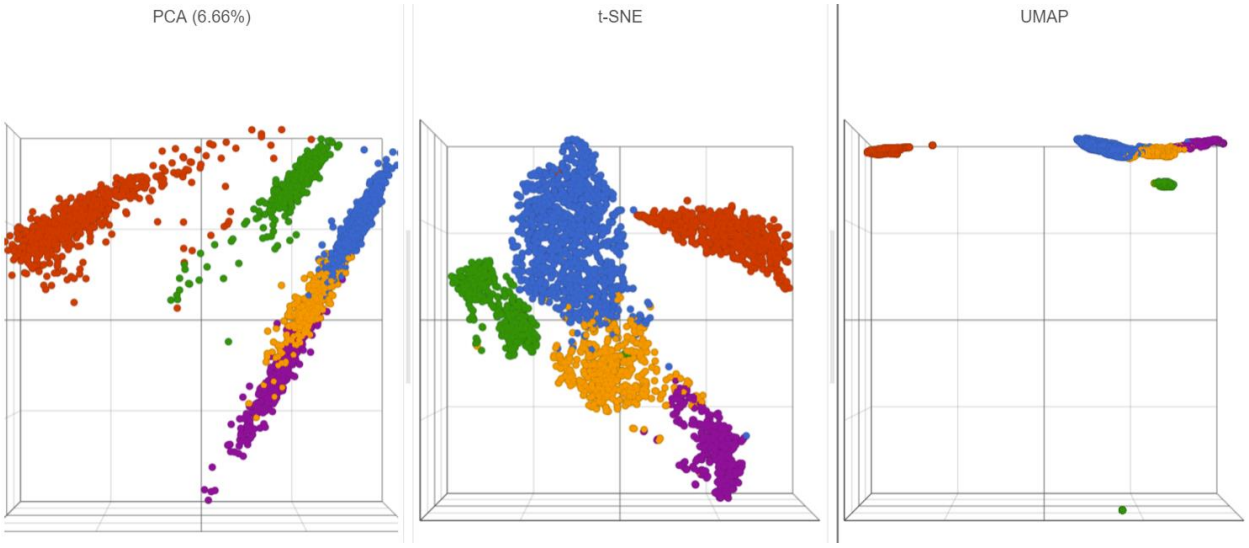
- Open a support ticket at partek.com/support
- Phone: +1-314-884-6172

Notes:_____

Exercise on Your Own

Compare different dimension reduction methods on the same data

- Run the PCA, UMAP, tSNE to generate report table
- Create a new session in Dataviewer
- Add 3D scatterplot from on each report table
- Select data source to render the cells

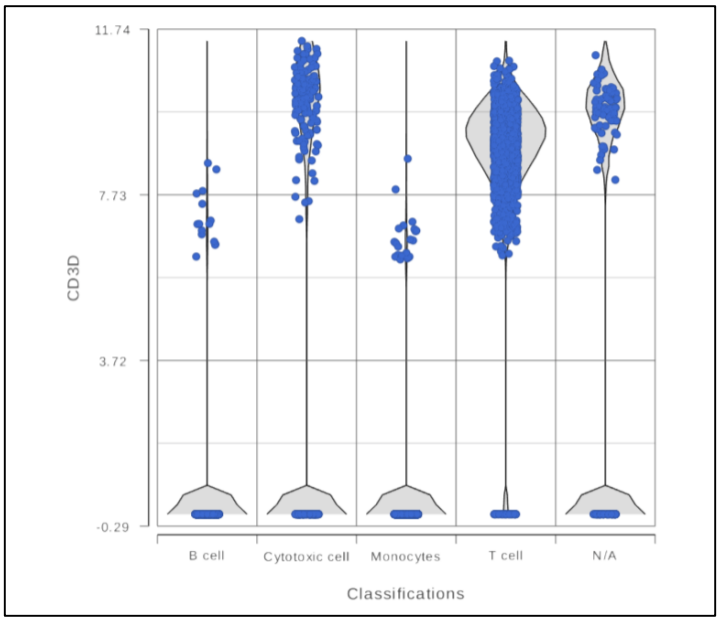


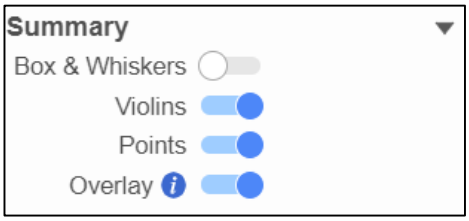
Notes:_____

Exercise on Your Own

Create a dot/violin/box plot on one gene

- Create a new session in Dataviewer
- Add a 2D scatterplot from the classify result data node
- X-axis represents classification, Y-axis represents a gene expression
- Turn on/off different plots in summary



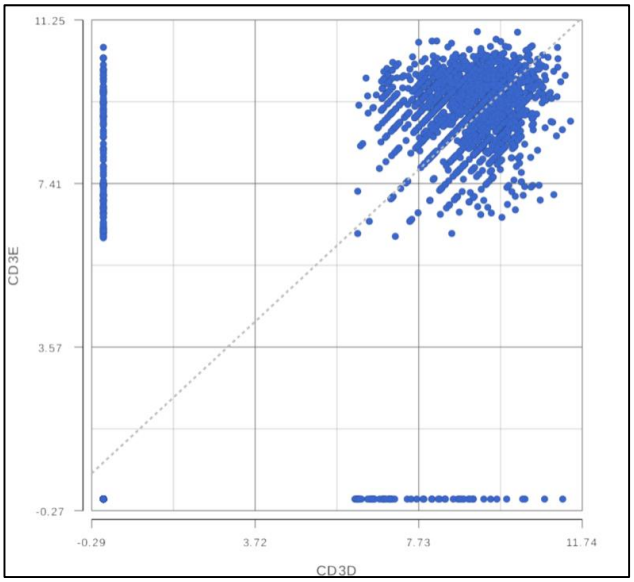


Notes:_____

Exercise on Your Own

Create correlation plot on two genes

- Create a new session in Dataviewer
- Add a 2D scatterplot from the classify result data node
- Select two genes to put on axes
- Turn on regression line

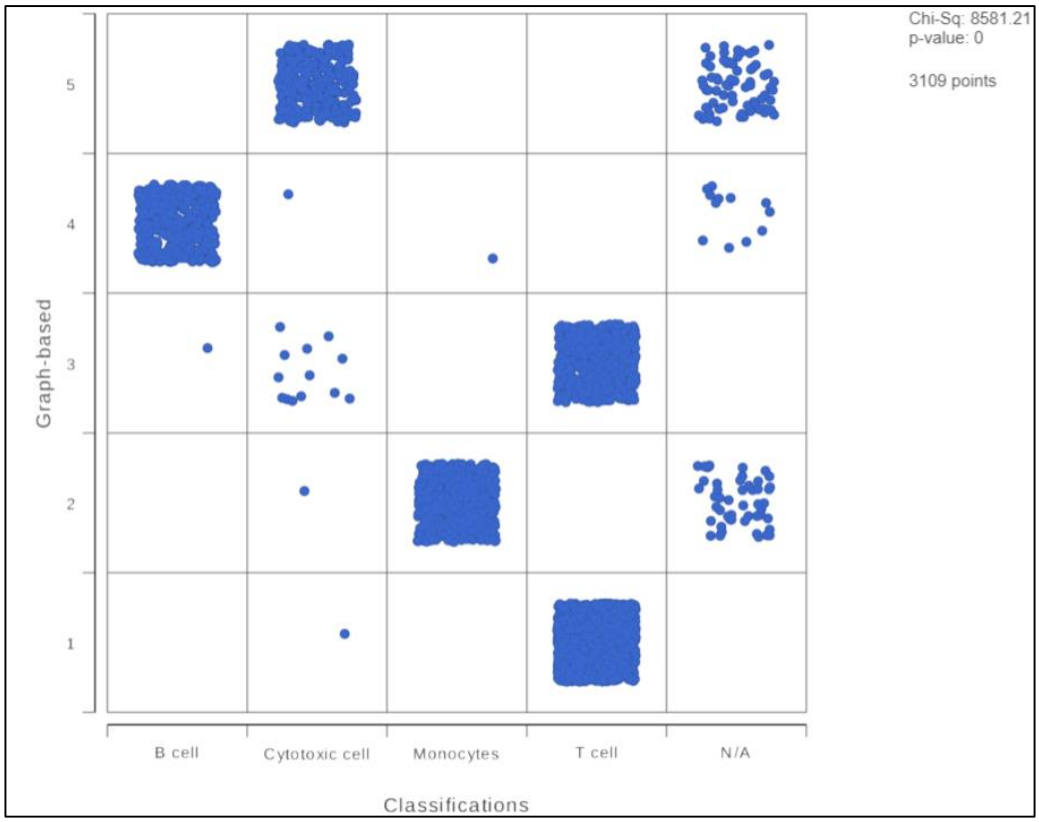


Notes:

Exercise on Your Own

Generate contingency table on two categorical variables

- Create a new session in Dataviewer
- Add a 2D scatterplot from the classify result data node
- X-axis represents classification, Y-axis represents graph-base clusters
- Chi-sq and p-value of the comparison is generated on the legend



Notes: