Single Cell RNA-Seq Analysis in Partek® Flow®

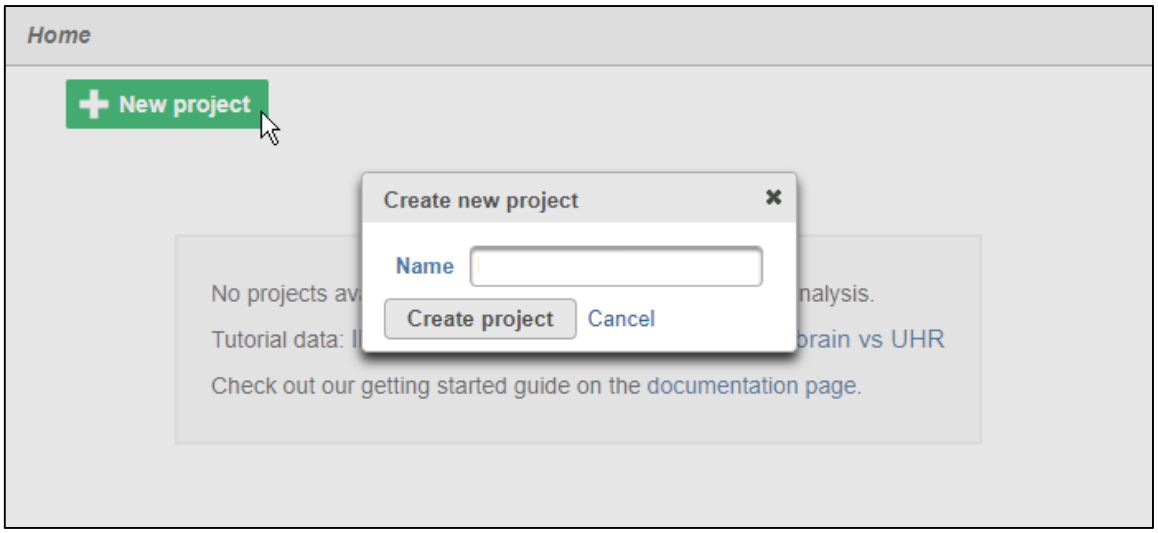
HANDS-ON TRAINING

National Institutes of Health
Aug 2019

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

Login and Project Set-up

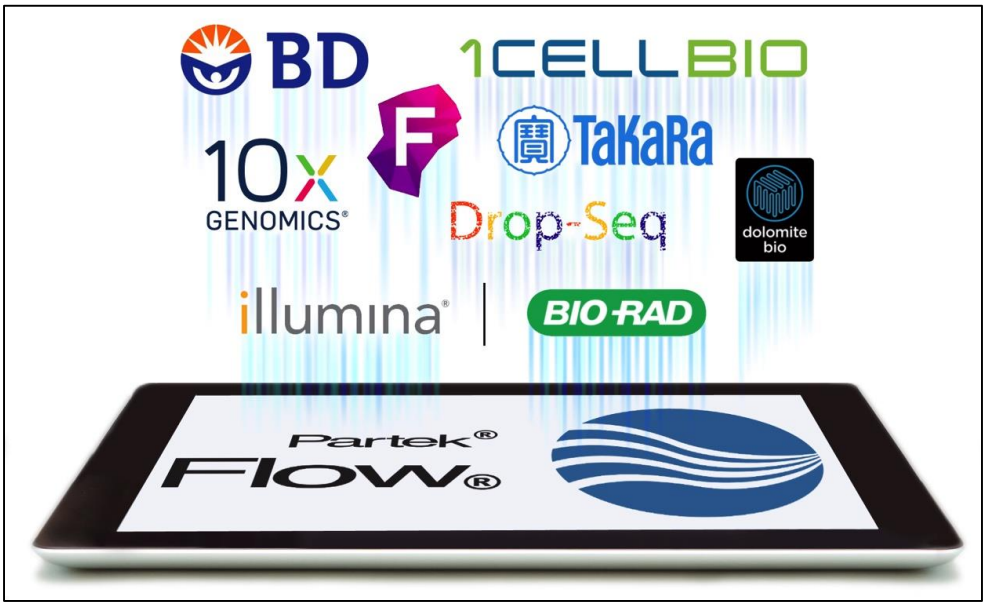
- Open your preferred web browser (Chrome, Firefox, etc. would work fine)
- Go to the server URL given by your instructor
- Log in using the username and password given to you
- This will open to the Partek Flow homepage
- Click **New Project** and enter project name: SC-RNAseq-[username]
- This will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



Notes:

Experiment Description

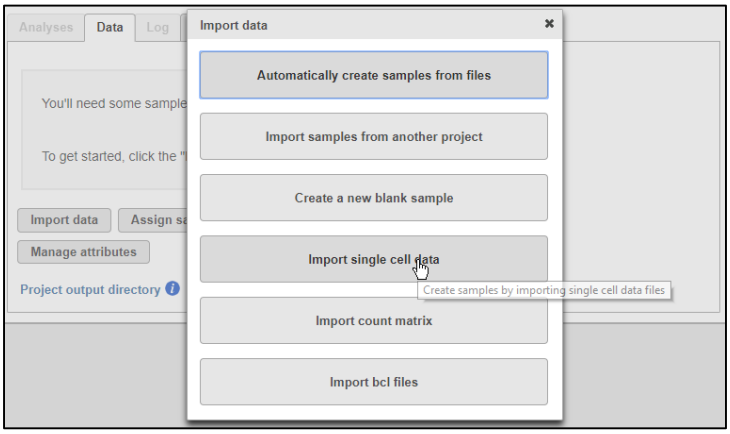
- 3k peripheral blood mononuclear cells (PBMCs) from a healthy donor
  - Any peripheral blood cell having a round nucleus
- Downloaded from 10X Genomics' dataset repository
  - https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k
- Today, will be importing the filtered gene/cell matrix from this dataset
- *Goal for today: Identify different blood cell populations*
- Partek Flow is versatile, supporting a wide variety of starting file types
- Partek Flow also supports a wide variety of single cell analysis platforms



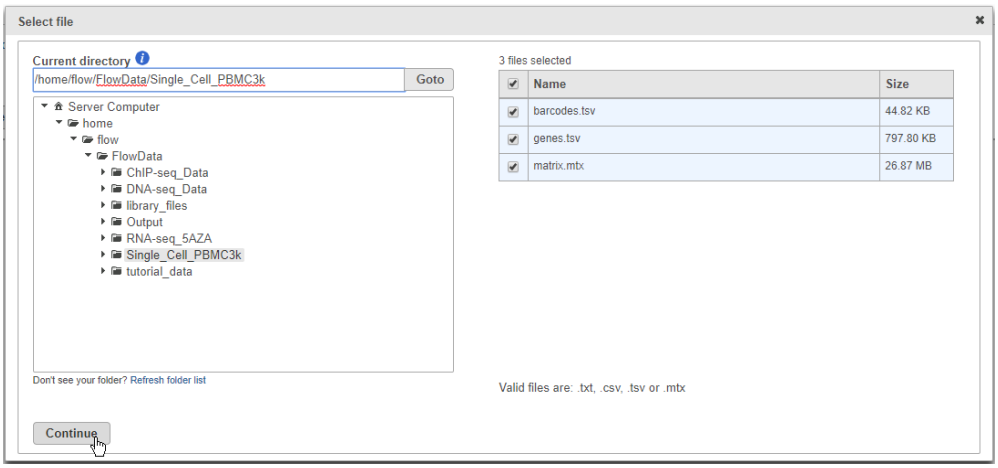
Notes:

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**



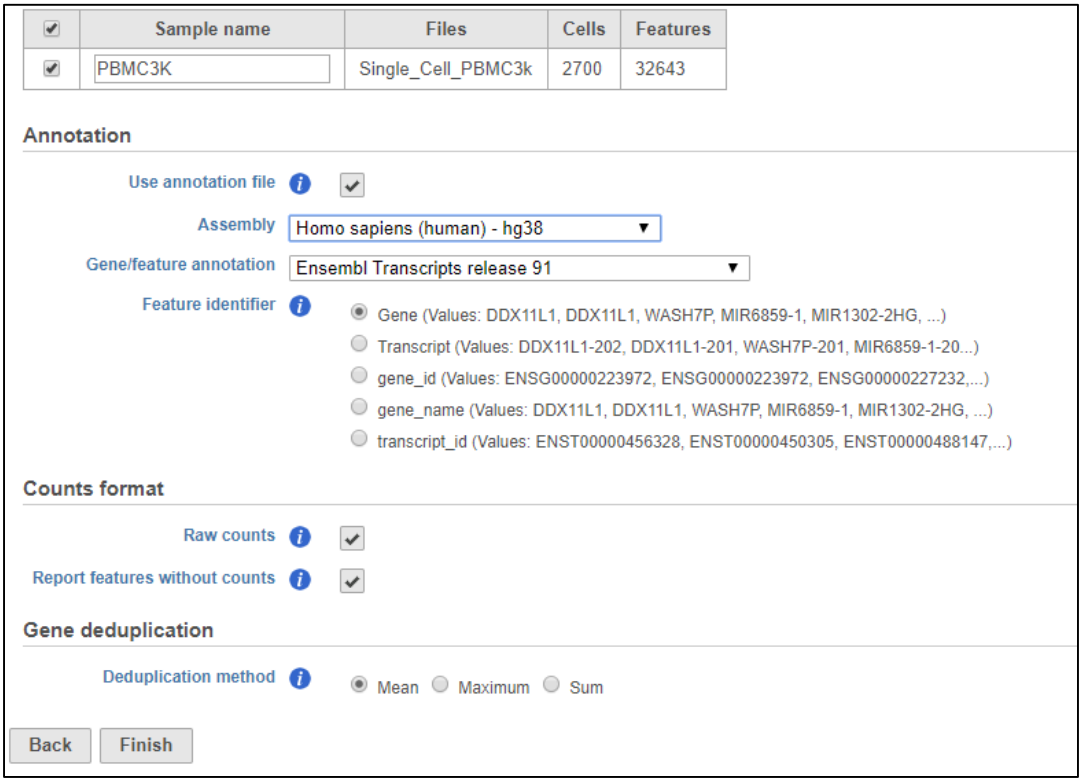
- Browse to */home/flow/FlowData/SingleCell_PBMC3k*
- Select <u>all 3 files</u> *(2 tsv and 1 mtx),* click **Continue**, then click **Next**
- **Note: Flow also support .h5 output from CellRanger**



Notes:_____

Specify Metadata

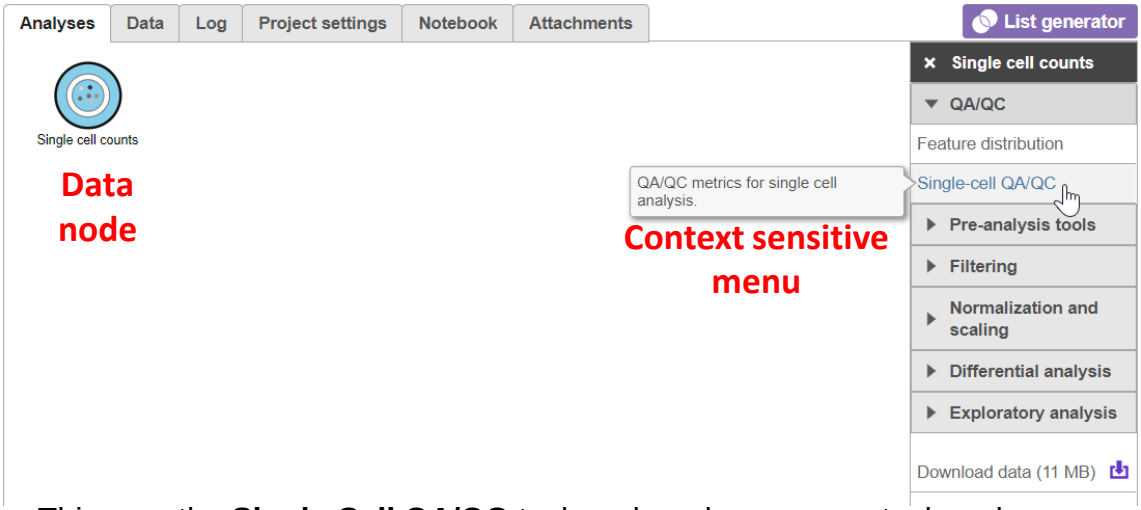
- Click the **Use annotation file** checkbox and set the annotation
  - Assembly: *Homo sapiens (human) - hg38*
  - Gene annotation: *Ensembl transcripts release 91*
- Set **Sample name** to *PBMC 2.7K*
- Click **Finish** to import sample. This will create your first data node



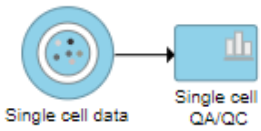
Notes:_____

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*



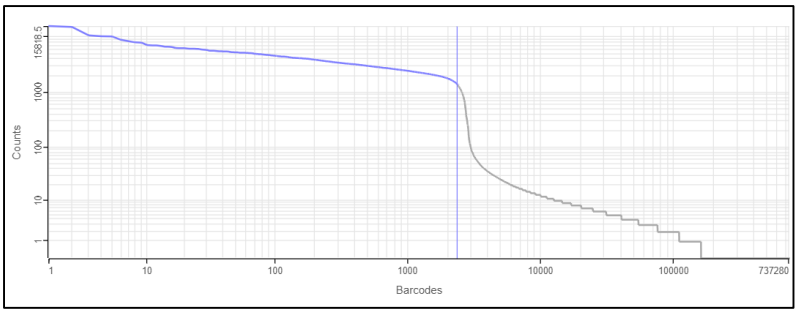
Task node

Notes:_____

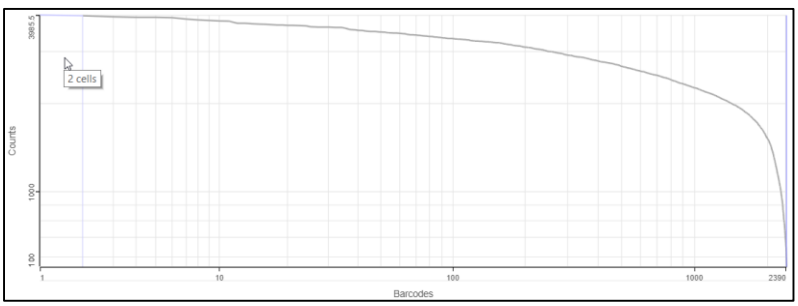
Cell barcode QA/QC

- Check if droplets (barcode) actually contain cells
- Filter out droplets (cell barcode) don't contain cells
- Knee plot: total read count for all the cell, X-axis represent barcodes in deceasing order on total counts

- **Need to be filtered**



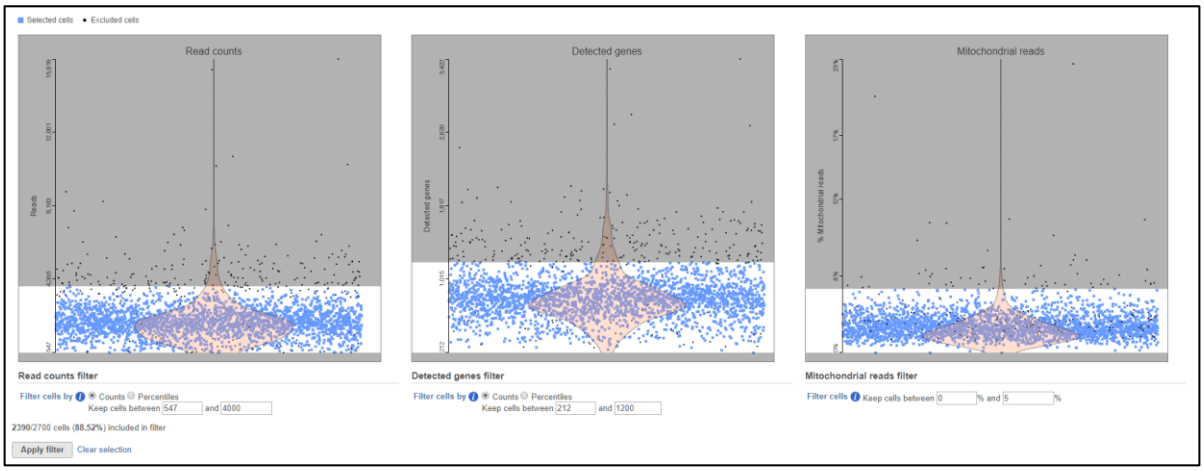
- **Good**



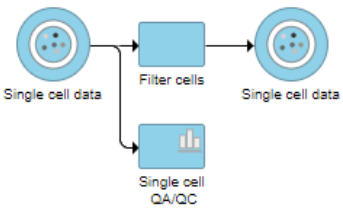
Notes:

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 **Read counts filter** to a max of **4000** reads, the **Detected genes filter** to a max of **1200** genes and the **Mitochondrial reads filter** to a max of **5%**
- Click **Apply filter**



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

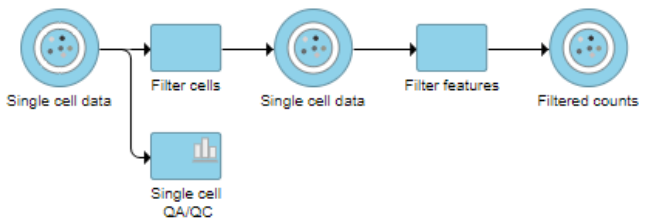


Notes:

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 == 0** in at least **99.9**% of the cells
- Click **Finish** to apply the filer

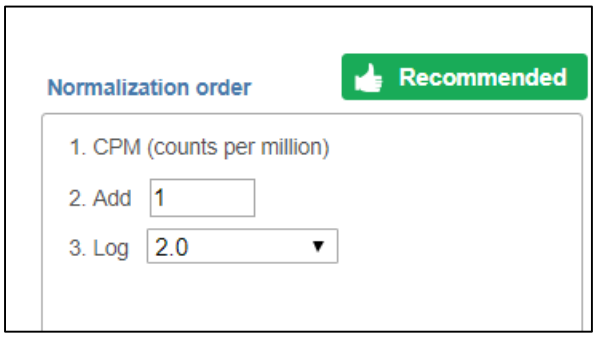
- The **Filter features** task creates a new **Filtered counts** data node



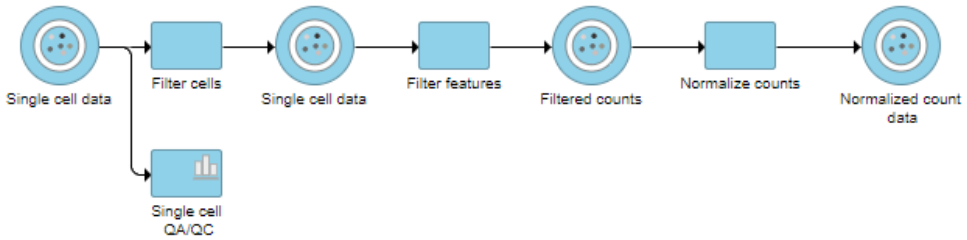
Notes:

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

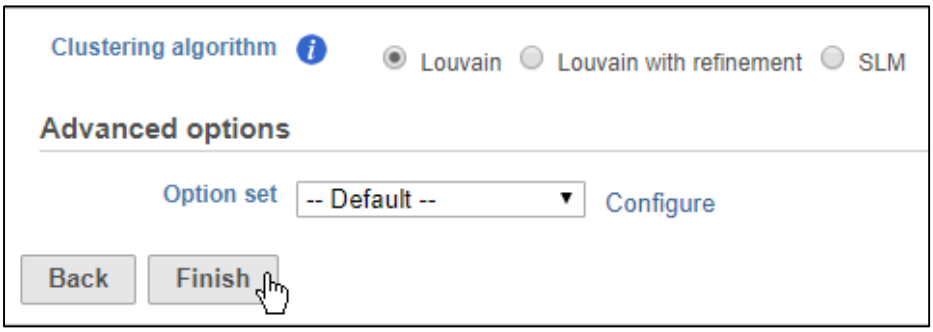
- We'll be using a combination of methods to identify some cell types commonly found in PBMCs. Namely:
  - Unbiased clustering (Graph-based)
  - Visualizing expression using
    - Canonical gene markers
    - Gene lists
  - Lassoing cell populations on the t-SNE plot

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

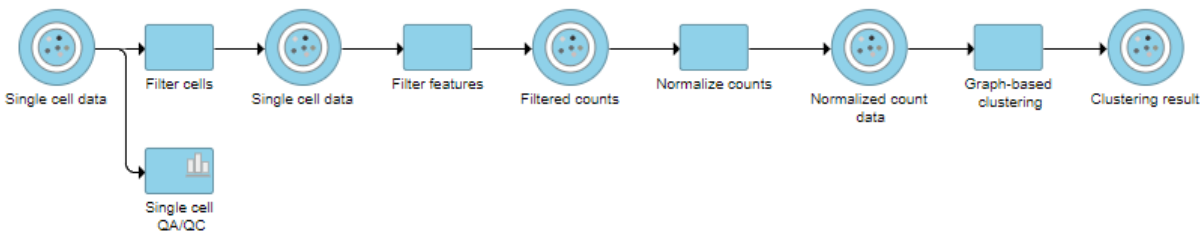
Notes:

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



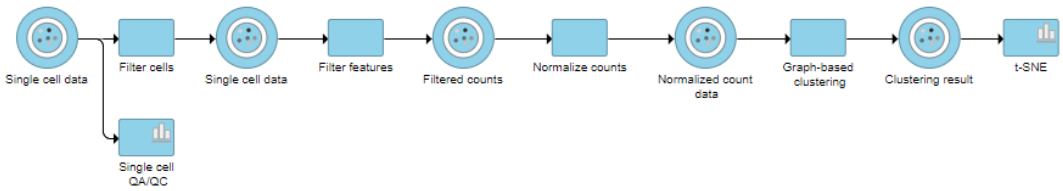
- **Graph-based clustering** produces a **Clustering result** data node



Notes:

Invoking the t-SNE plot

- Click the **Clustering result** data node
- Click **t-SNE** in the **Exploratory analysis** section of the task menu
- Click **Finish** to run the t-SNE task with default settings
- A **t-SNE** node is produced, double click it to open the t-SNE plot

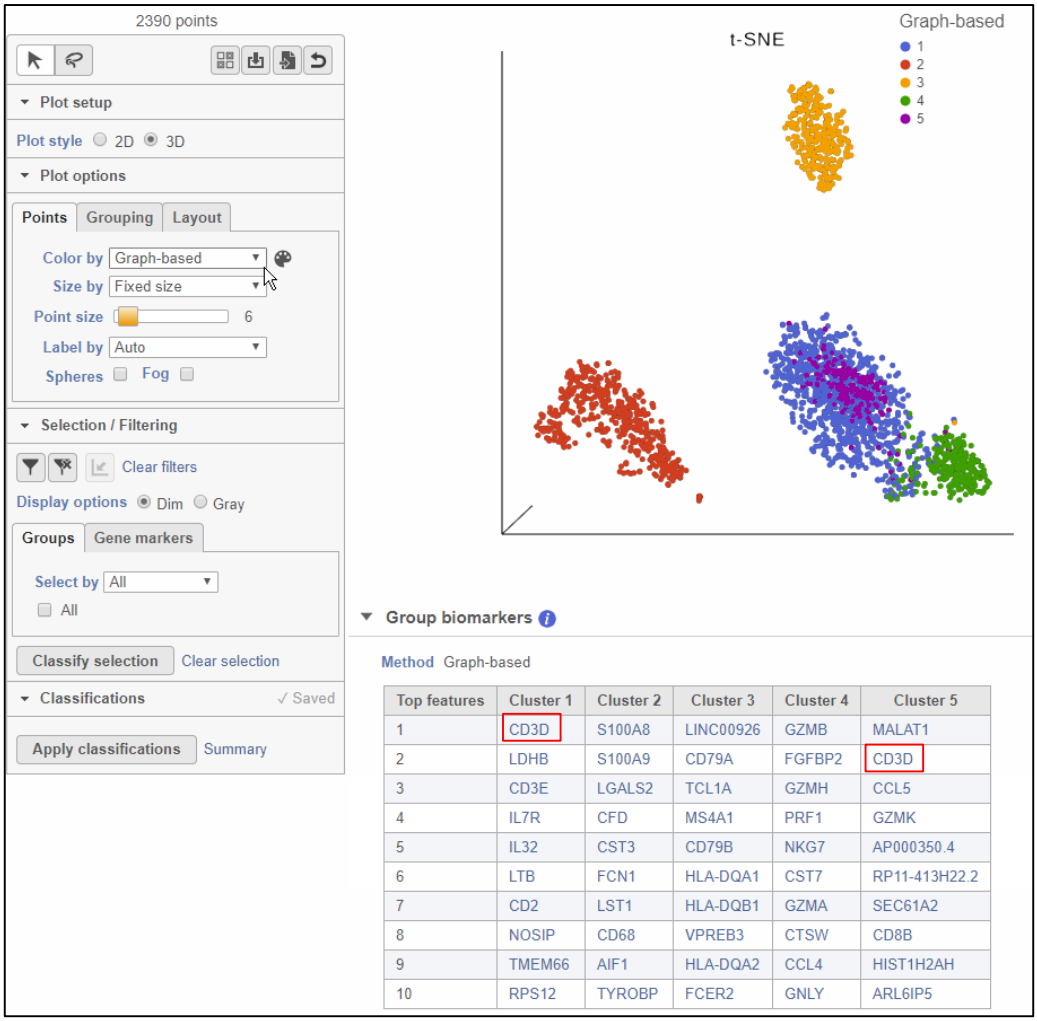


- We will use the interactive t-SNE plot to view the clustering results and classify our cells
  - *t-distributed stochastic neighbor embedding (t-SNE) is a popular technique for visualizing high-dimensional data*
  - *t-SNE draws cells that are similar to each other across the high-dimensional RNA-Seq data, where each gene is a dimension, close together on the plot*
  - *t-SNE uses principal components analysis to determine which cells are similar to each other*

Notes:

Identifying Cell Types

- Select **Graph-based** from the Color by drop-down menu
  - *You can see the data has been clustered into 5 clusters*
- The **Group biomarkers** table lists genes that distinguish each cluster
- **CD3D**, a T cell marker gene, is listed as a biomarker for Clusters 1 and 5



Group biomarkers

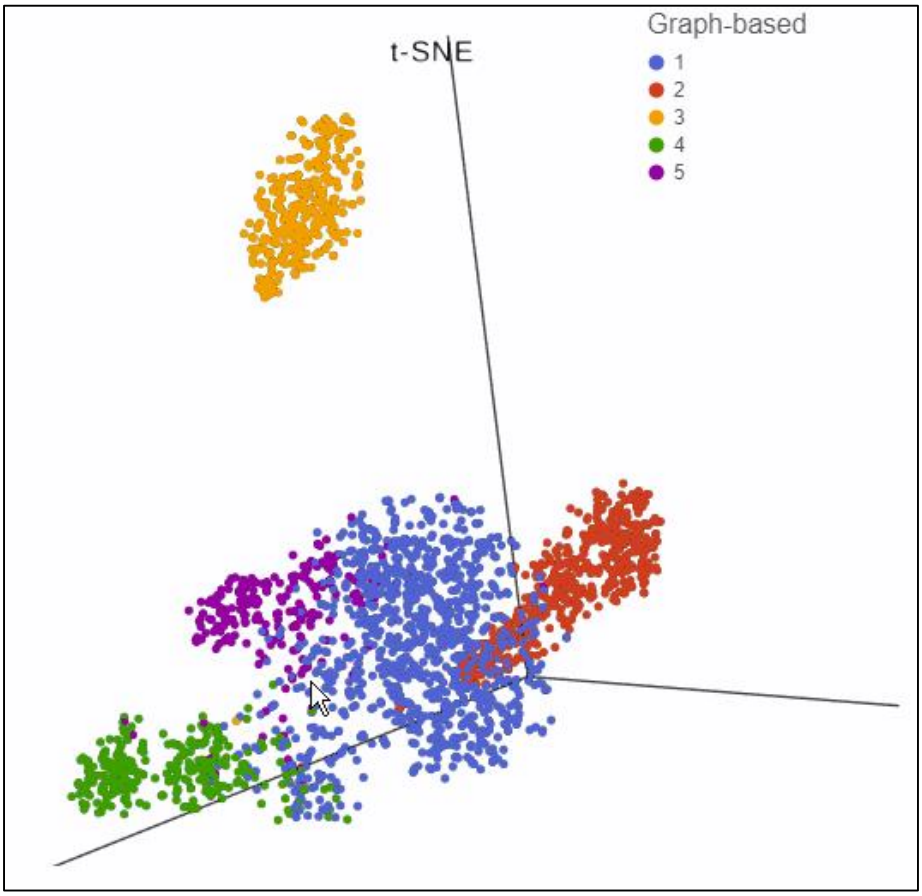
Method Graph-based

Top features	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	CD3D	S100A8	LINC00926	GZMB	MALAT1
2	LDHB	S100A9	CD79A	FGFBP2	CD3D
3	CD3E	LGALS2	TCL1A	GZMH	CCL5
4	IL7R	CFD	MS4A1	PRF1	GZMK
5	IL32	CST3	CD79B	NKG7	AP000350.4
6	LTB	FCN1	HLA-DQA1	CST7	RP11-413H22.2
7	CD2	LST1	HLA-DQB1	GZMA	SEC61A2
8	NOSIP	CD68	VPREB3	CTSW	CD8B
9	TMEM66	AIF1	HLA-DQA2	CCL4	HIST1H2AH
10	RPS12	TYROBP	FCER2	GNLY	ARL6IP5

Notes:

Rotating, panning, and zooming

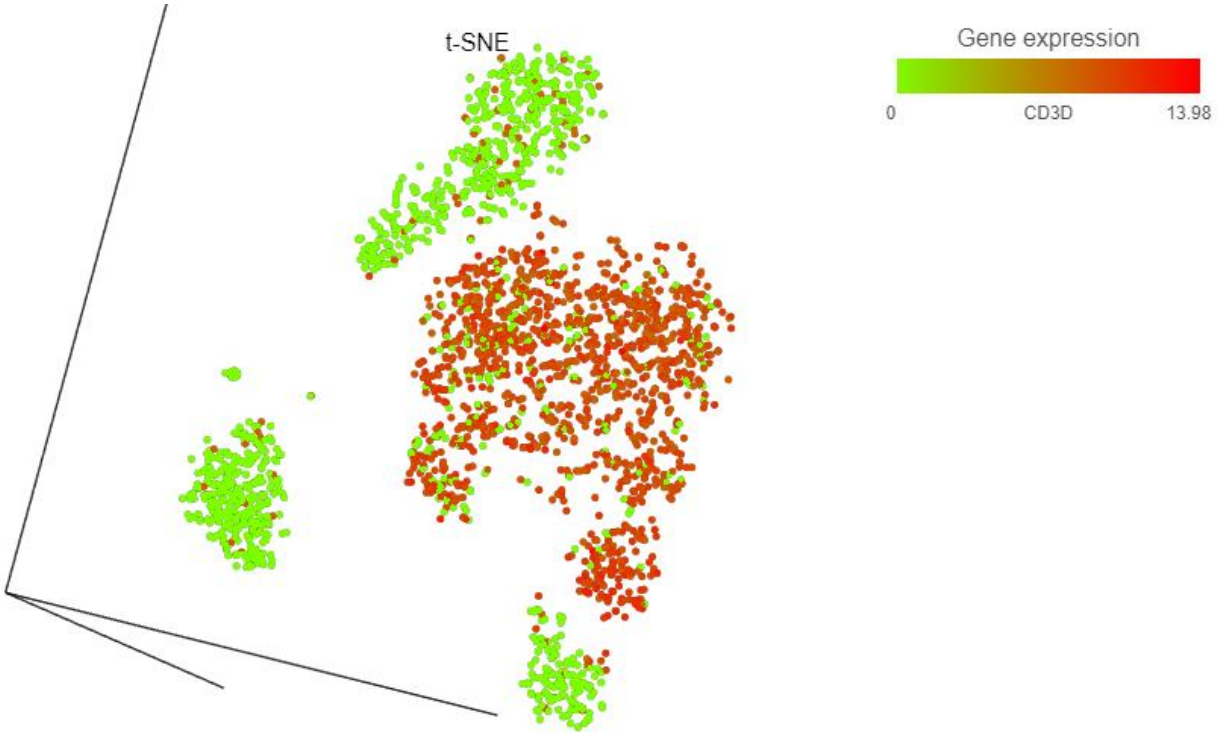
- To get a better view of the cluster, we can change our view
  - Rotate the plot by left-clicking and dragging
  - Zoom using the mouse wheel
  - Pan by right-clicking and dragging
  - Move the legend by left-clicking and dragging it



Notes:

Coloring cells by marker genes

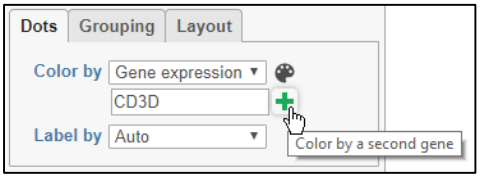
- Click **CD3D** on the *Group biomarkers* table to color by **CD3D** expression
- Cells are now colored by their expression values for **CD3D** from green (0) to red (max)



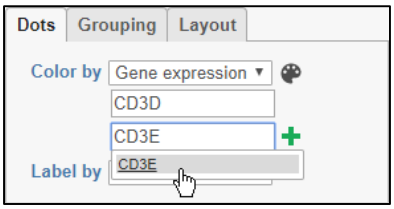
Notes:

Coloring by a second marker gene

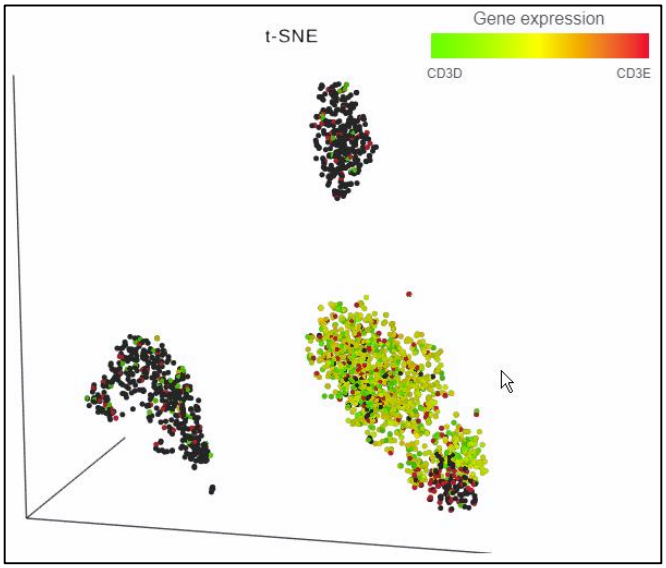
- Click the green plus to color by a second gene



- Type **CD3E** in the second text box and select **CD3E** from the list



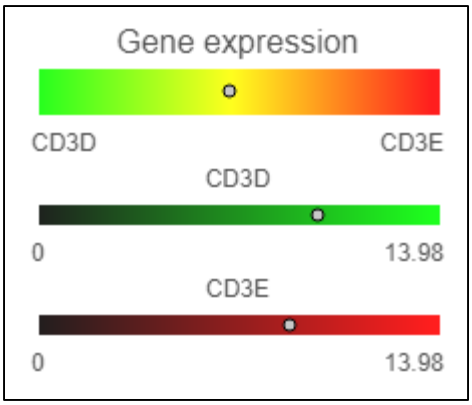
- The plot is now colored by CD3D (green) and CD3E (red) with cells that express both genes colored yellow



Notes:_____

Viewing expression values for individual cells

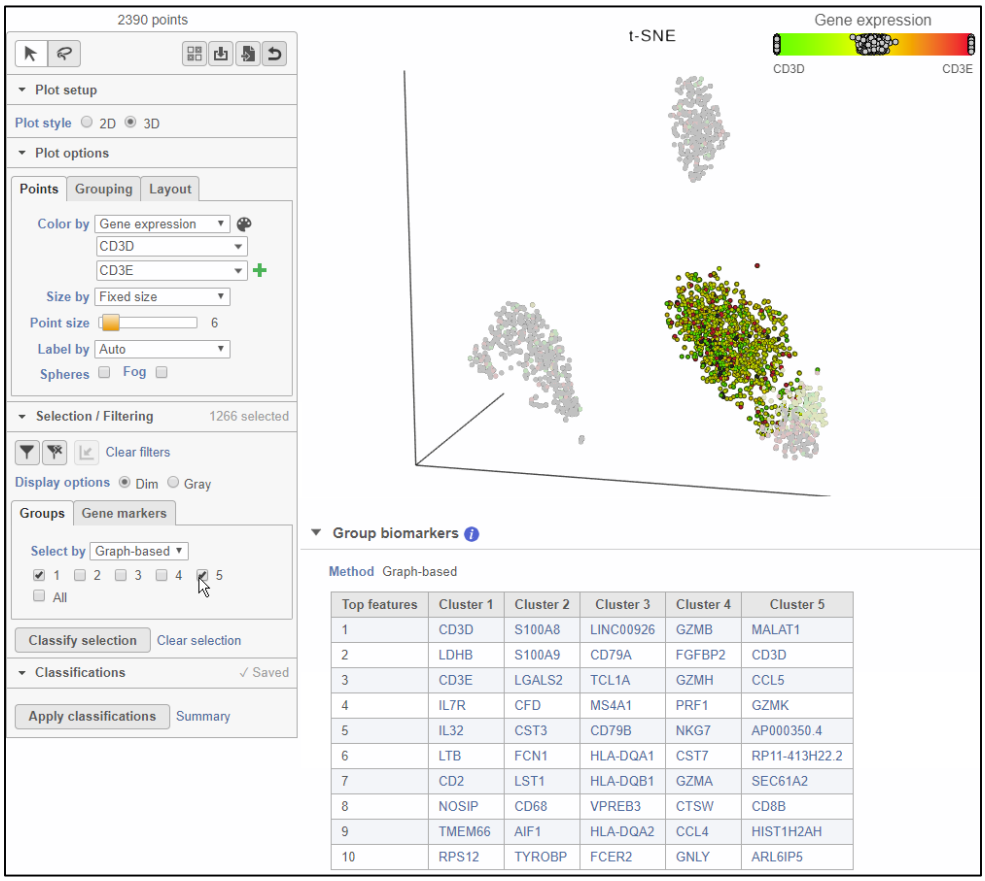
- Click a yellow cell on the plot
- The expression values of that cell are listed in the legend
  - Each gene is assigned a color channel (RGB)
  - Cells that express multiple genes have mixed color
- Yellow cells express both CD3D and CD3E



Notes:

Selecting by Cluster

- Cells in Clusters 1 and 5 express T cell marker genes, we want to classify these cells as T cells
- Choose **Graph Based** from the **Select by** drop-down menu
- Click the boxes for **1** and **5** to select cells in those clusters
  - *Selected cells on the plot are shown in bold*
  - *The distribution of expression values for selected cells is shown on the legend*

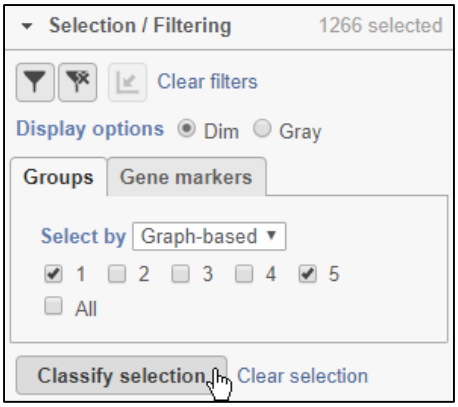


Top features	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	CD3D	S100A8	LINC00926	GZMB	MALAT1
2	LDHB	S100A9	CD79A	FGFBP2	CD3D
3	CD3E	LGALS2	TCL1A	GZMH	CCL5
4	IL7R	CFD	MS4A1	PRF1	GZMK
5	IL32	CST3	CD79B	NKG7	AP000350.4
6	LTB	FCN1	HLA-DQA1	CST7	RP11-413H22.2
7	CD2	LST1	HLA-DQB1	GZMA	SEC61A2
8	NOSIP	CD68	VPREB3	CTSW	CD8B
9	TMEM66	AIF1	HLA-DQA2	CCL4	HIST1H2AH
10	RPS12	TYROBP	FCER2	GNLY	ARL6IP5

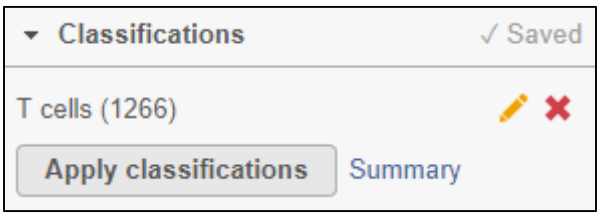
Notes:

Classifying selected cells

- Click **Classify selection**



- Name the classification **T cells**
- Click **Save**
- T cells is added to the **Classifications** section of the menu
  - *The number of T cells is listed in parentheses*

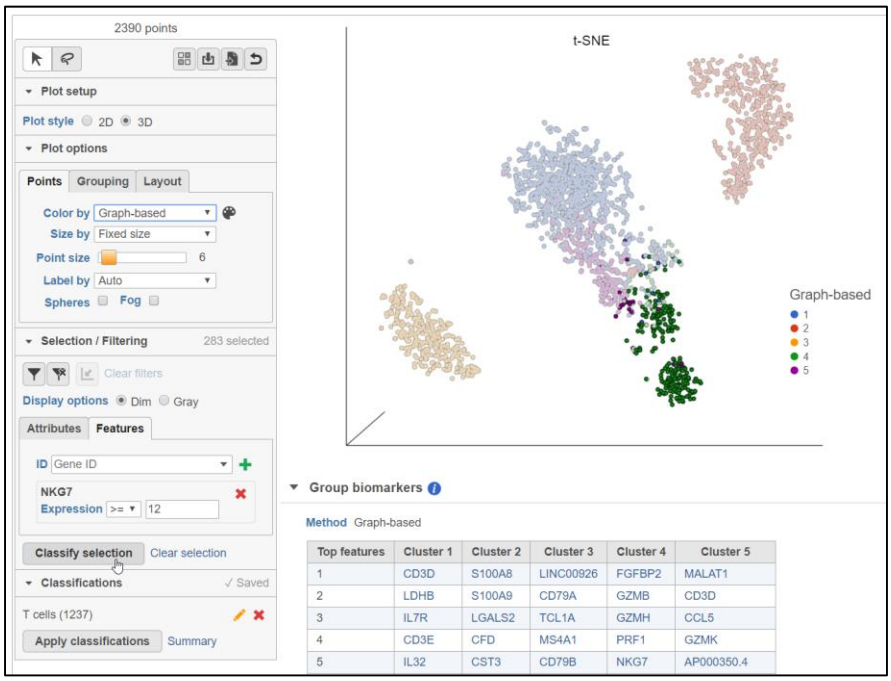


- Clear the selection by clicking a blank space on the plot
- Select **Graph-based** from the **Color by** drop-down menu

Notes:_____

Classifying cytotoxic cells

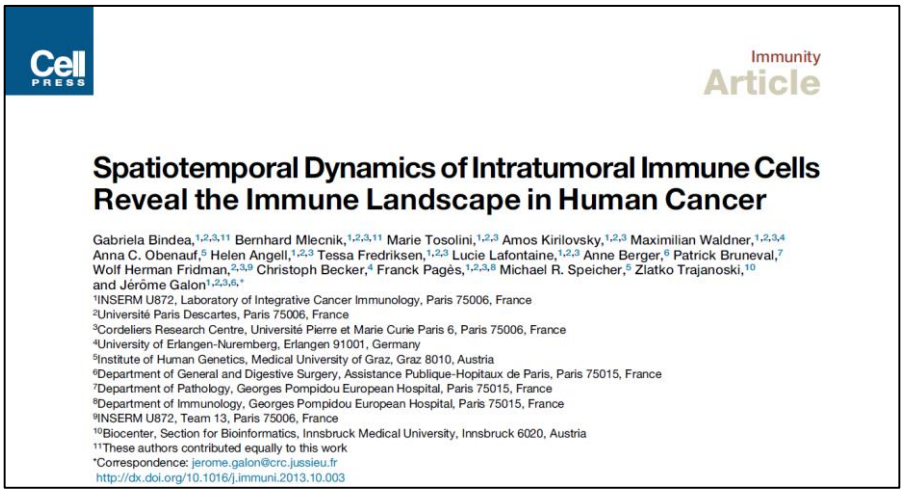
- In the *Selection/Filte*r section, choose *Features* tab, type in NKG7 click ✚
- Type expression >=12 to select the cells
- Click Classify selection and name the classification **Cytotoxic cells**
- Save the classification
- Click the plot to clear the selection



Notes:

Coloring by a gene list

- Cluster 3 lists the B-cell marker genes CD79A and CD79B as biomarkers
- To further verify that these are B cells cells, we can use a published list of 92 marker genes for B cells



Cell PRESS

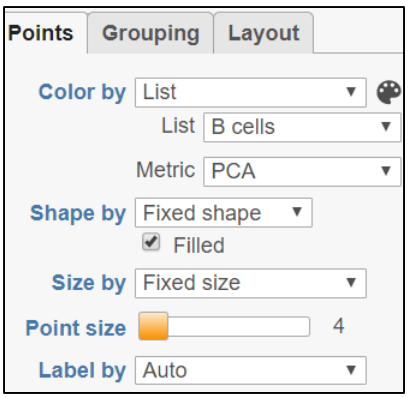
Immunity
Article

Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer

Gabriela Bindea,[1,2,3,11] Bernhard Mlecnik,[1,2,3,11] Marie Tosolini,[1,2,3] Amos Kirilovsky,[1,2,3] Maximilian Waldner,[1,2,3,4] Anna C. Obenauf,[5] Helen Angell,[1,2,3] Tessa Fredriksen,[1,2,3] Lucie Lafontaine,[1,2,3] Anne Berger,[6] Patrick Bruneval,[7] Wolf Herman Fridman,[2,3,9] Christoph Becker,[4] Franck Pagès,[1,2,3,8] Michael R. Speicher,[5] Zlatko Trajanoski,[10] and Jérôme Galon[1,2,3,6,*]

[1]INSERM U872, Laboratory of Integrative Cancer Immunology, Paris 75006, France
[2]Université Paris Descartes, Paris 75006, France
[3]Cordeliers Research Centre, Université Pierre et Marie Curie Paris 6, Paris 75006, France
[4]University of Erlangen-Nuremberg, Erlangen 91001, Germany
[5]Institute of Human Genetics, Medical University of Graz, Graz 8010, Austria
[6]Department of General and Digestive Surgery, Assistance Publique-Hopitaux de Paris, Paris 75015, France
[7]Department of Pathology, Georges Pompidou European Hospital, Paris 75015, France
[8]Department of Immunology, Georges Pompidou European Hospital, Paris 75015, France
[9]INSERM U872, Team 13, Paris 75006, France
[10]Biocenter, Section for Bioinformatics, Innsbruck Medical University, Innsbruck 6020, Austria
[11]These authors contributed equally to this work
*Correspondence: jerome.galon@crc.jussieu.fr
http://dx.doi.org/10.1016/j.immuni.2013.10.003

- Select **List** from the **Color by** drop-down menu

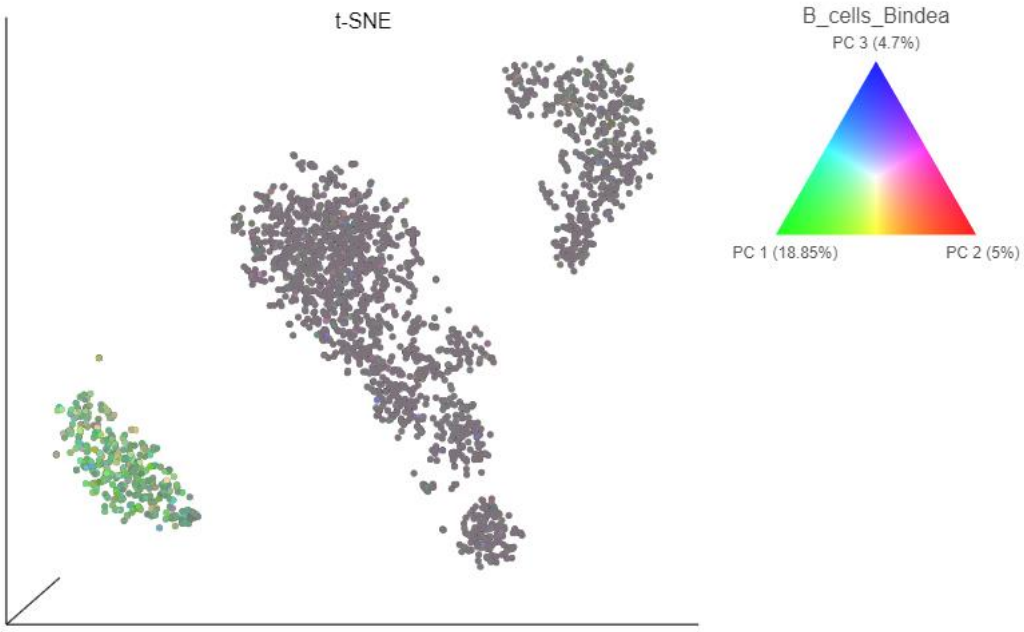


- Select **B cells**
- Choose the color metric

Notes:

Coloring by a gene list

- Coloring by a list performs principal components analysis on the gene list to identify cells that are distinguished by their expression of genes on the list
- The color of each cell is determined by its value for the first three PCs (PC1 green, PC2 red, PC3 blue)
- The cells from Cluster 3 are colored green and are distinguishable based on their expression of 92 B cell marker genes

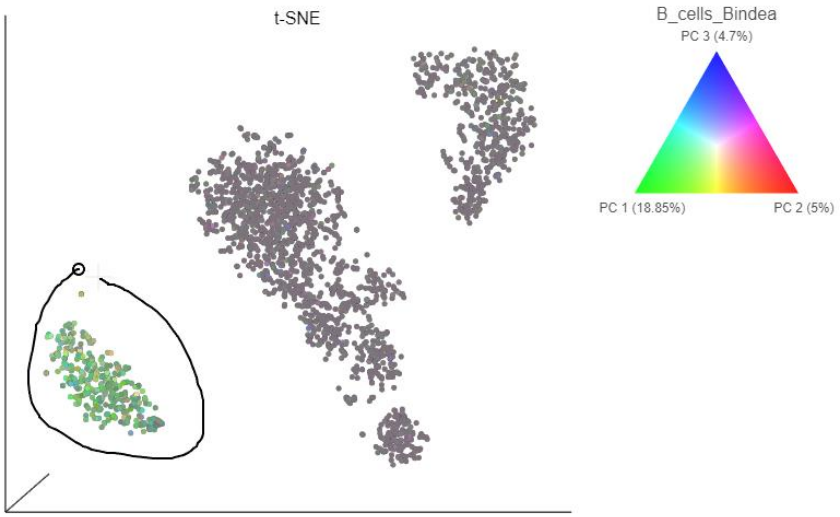


- Choose Sum from the Metric drop-down list

Notes:_____

Selecting cells using the 3D lasso tool

- Click the lasso icon to activate the **3D lasso tool** 
- Click and hold to draw a lasso around the cluster of green cells
- Click the starting circle to close the lasso and select the cluster

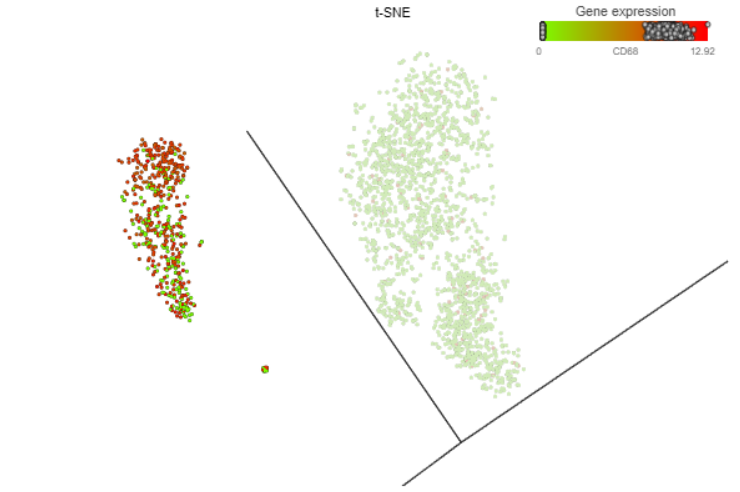


- Click **Classify selection** and name this group **B cells** (note that some cells may have already been classified)

Notes:_____

Identifying monocytes

- Clear the selection
- Select **Graph Based** from the **Color by** drop-down menu
- Pan and zoom to focus on Cluster 2
- A biomarker for Cluster 2 is a monocyte marker gene, **CD68**, click on this gene in the *Group biomarkers* table



Group biomarkers

Method Graph-based

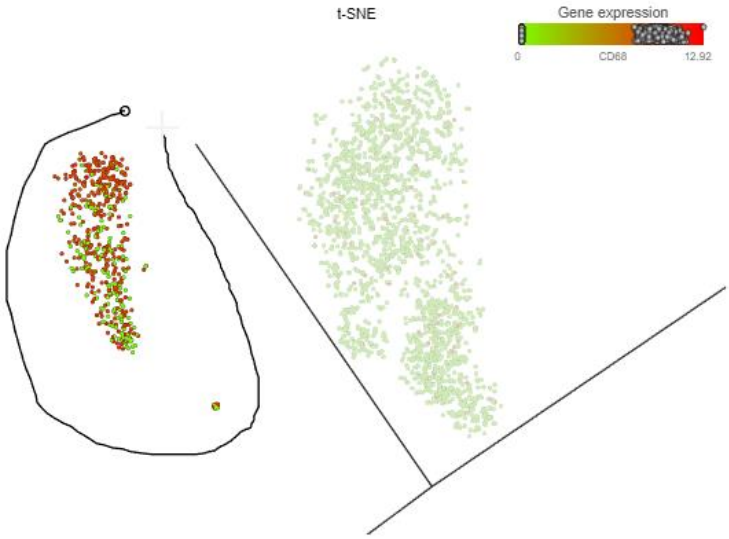
Top features	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	CD3D	S100A8	LINC00926	FGFBP2	MALAT1
2	LDHB	S100A9	CD79A	GZMB	CD3D
3	IL7R	LGALS2	TCL1A	GZMH	CCL5
4	CD3E	CFD	MS4A1	PRF1	GZMK
5	IL32	CST3	CD79B	NKG7	AP000350.4
6	LTB	FCN1	HLA-DQA1	CST7	RP11-413H22.2
7	CD2	LST1	HLA-DQB1	GZMA	SEC61A2
8	NOSIP	CD68	VPREB3	CTSW	CD8B
9	TMEM66	AIF1	HLA-DQA2	CCL4	HIST1H2AH
10	RPS12	TYROBP	FCER2	GNLY	ARL6IP5

Download

Notes:

Classifying monocytes

- Click the lasso icon to activate the **3D lasso tool**
- Click and hold to draw a lasso around the red cells
- Click the starting circle to close the lasso and select the cluster

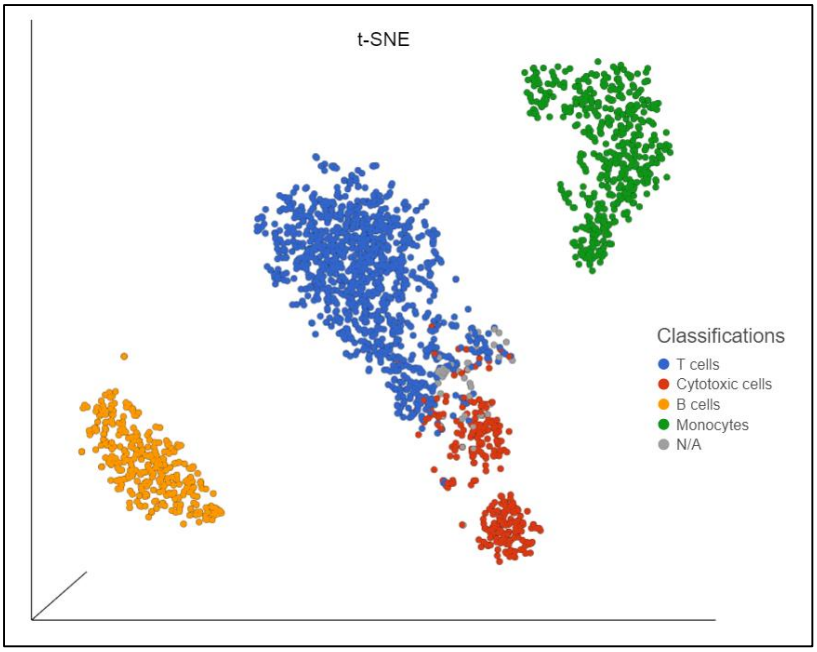


- Click **Classify selection** and name this group **Monocytes**
- Click the plot to clear the selection

Notes:_____

Viewing classifications

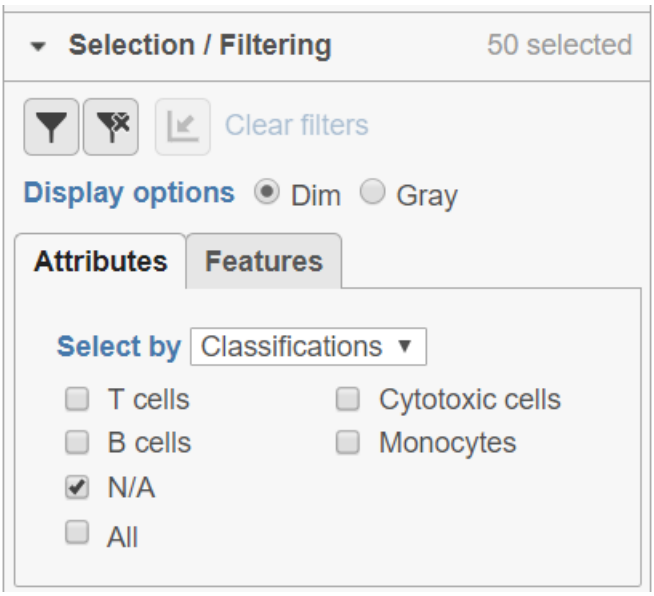
- Reset the view
- Select **Classifications** from the **Color by** drop-down menu
- N/A means the cell doesn't have be classified in the classifications attribute



Notes:

Select and filter

- Next we will classify the N/A cells
- In *Selection /Filtering* section *Attributes* tab, choose **Classifications** from the drop-down
- Select **N/A** to highlight the cells in this group
- Click on  to filter only include those cells to view
- Clear filters display all the cells
- Since N/A cells are close to Cytotoxic cells, we will classify them as cytotoxic cells.
- Click **Summary** in the **Classifications** section of the menu to view the **Classifications Summary** table
- Click **Apply classifications** to run the **Classify cells** task



Notes:_____

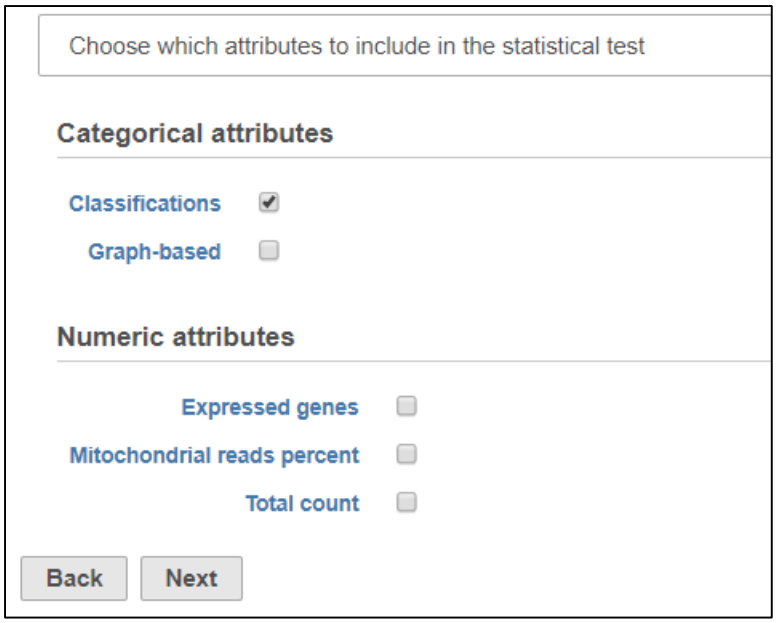
Exporting visualizations and notebook

- All visualizations in Partek Flow can be saved as publication-quality images
  - To save the t-SNE plot, click the **Save image** button 
- You can also export an image to a **Project Notebook**. The notebook is always associated with a specific project, so specific notes related to the analysis stays with the data
- To send the same image to the notebook, click the **Send to notebook page** button 
- This will prompt you to specify a notebook page to send the image too. Create a new page and call it *Cell classification* and click the **Send** button
- Click the *Cell classification* page link to navigate to the notebook page
- The page has helpful features for writing notes about specific observations, attaching relevant images and can be exported as a pdf file

Notes:

Identifying differentially expressed genes

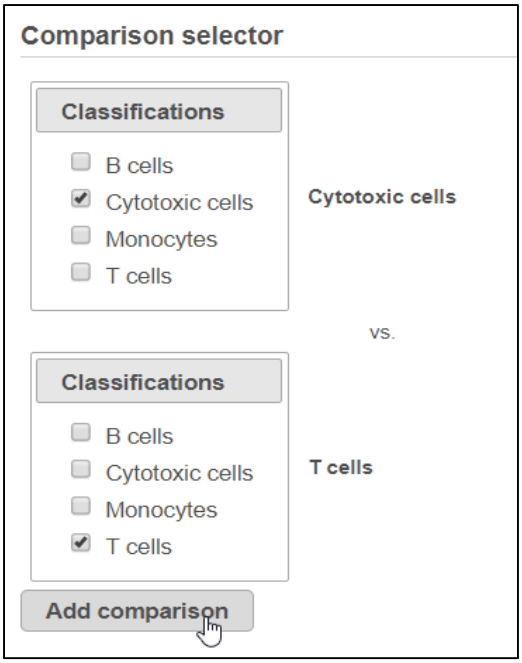
- Now that we have classified cells into cell types, we can compare expression between cell types
- Here, we will compare **Cytotoxic cells** and **T cells** to identify genes that are differentially expressed between these cell types
- Click the **Classified groups** node produced by the **Classify cells** task
- Click **GSA** in the **Differential analysis** section of the task menu
- Click **Classifications** to include it in the GSA model
  - *Adding a factor to the GSA model means that its effects will be considered in the statistical test*
- Click **Next**



Notes:

Adding a comparison

- Differential expression analysis lets us compare groups. Here, we want to compare **Cytotoxic cells** to **T cells**
- Click **Cytotoxic cells** in the top panel
- Click **T cells** in the bottom panel
  - *The top panel is the numerator and the bottom panel is the denominator for fold-change calculations*

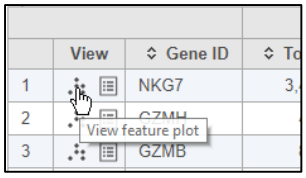


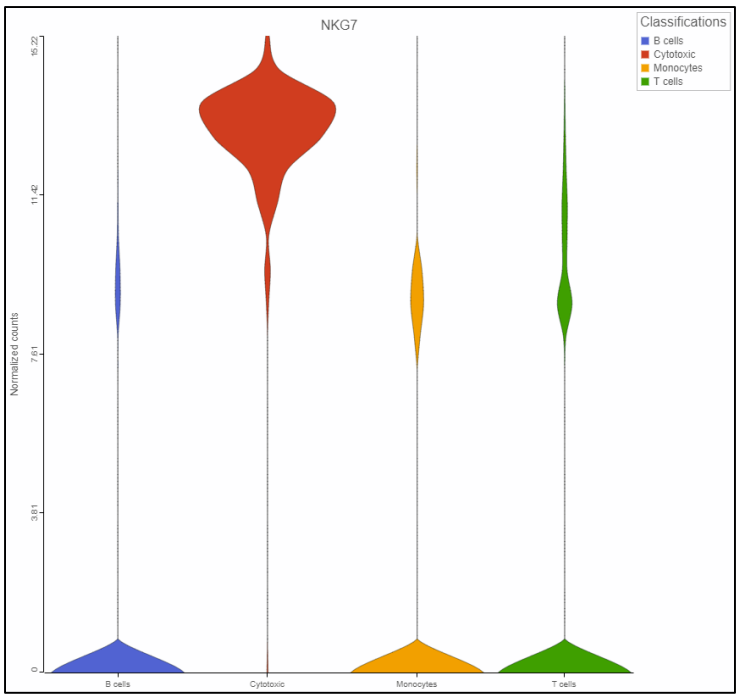
- Click **Add comparison**
- Click **Finish** to run the statistical test
  - Running the GSA task produces a **Feature list** data node

Notes:

Viewing results

- Double click the **Feature Lists** data node to open the ANOVA report
- The **Gene list** table in the GSA report lists every gene that was considered by the GSA
  - *Genes are listed starting with the lowest p-value*
- Click the **dots** icon next to **NKG7** under **View** to open a violin plot



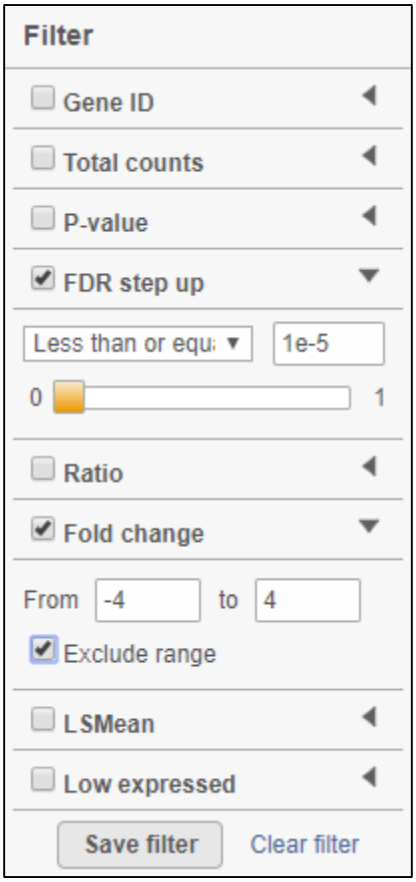


- Click **GSA report** to return to the gene list table

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 **-4** to **4**
- The number of genes in the table changes with the filter applied



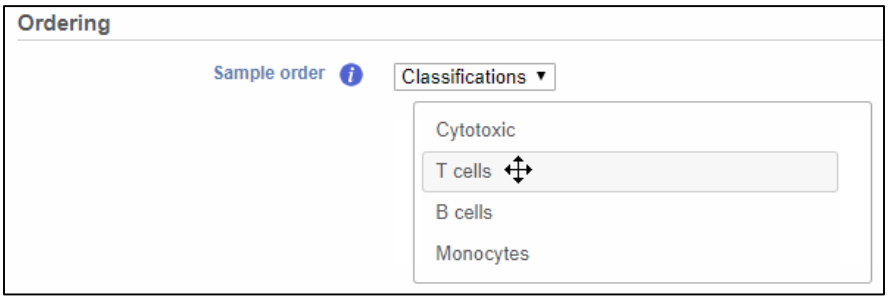
- Click Generate filtered node to run the **Differential analysis filter** task

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

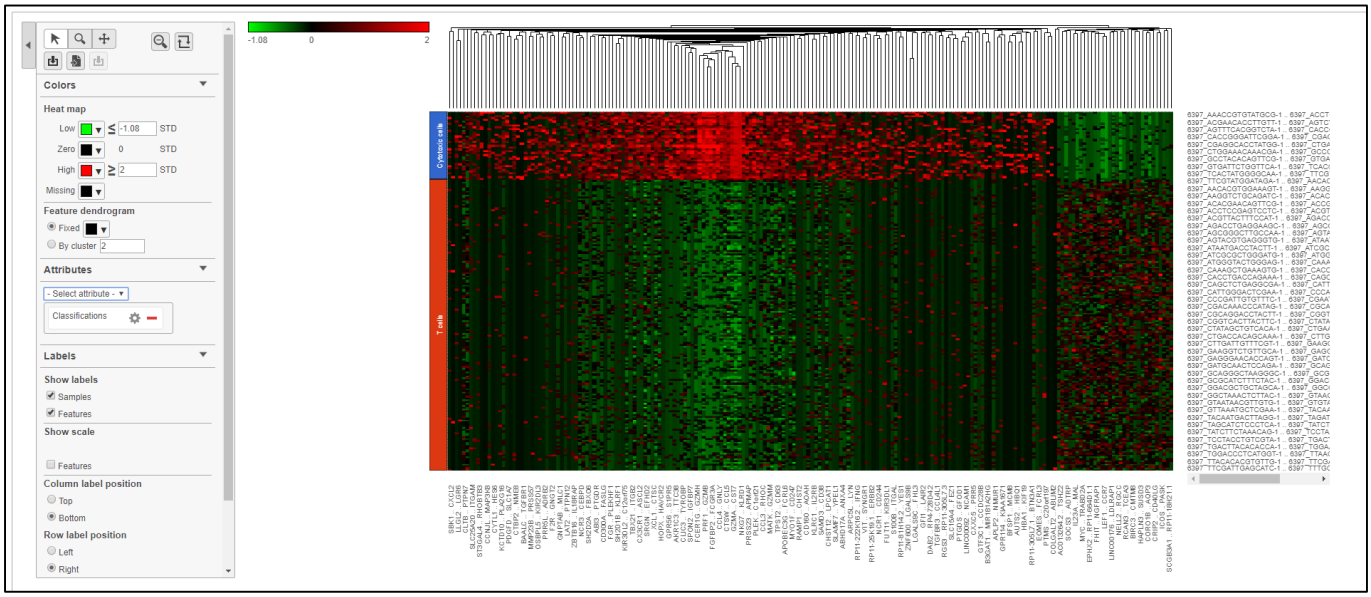


- Click **Finish** to run **Hierarchical clustering**

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

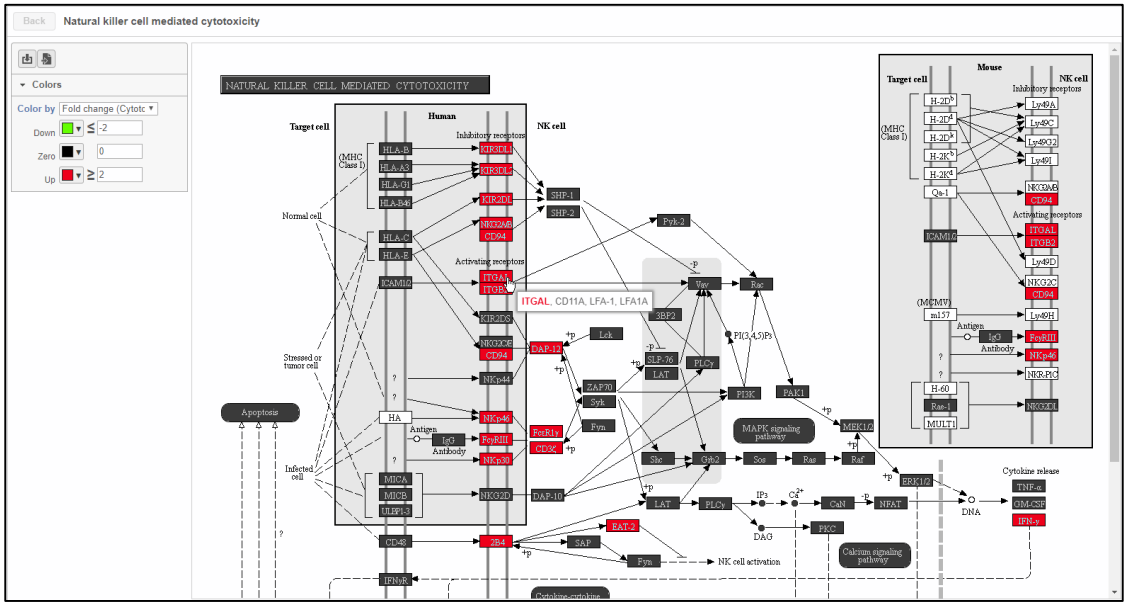
Notes:

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:_____

Further Training

Self-learning

- Check out http://www.partek.com/flow-resources for documentation and additional resources
- Recorded webinars available on http://www.partek.com/webinars
- Use the t-SNE to identify additional cell populations in the PBMC 2.7K data. A few suggestions:
  - CD14+ Monocytes
  - CGR3A+ Monocytes
  - CD8A+ Cytotoxic T cells
  - NK cells
- Ready to analyze a multi-sample dataset? Try our Glioma multi-sample tutorial

Regional Technical Support

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

Notes: