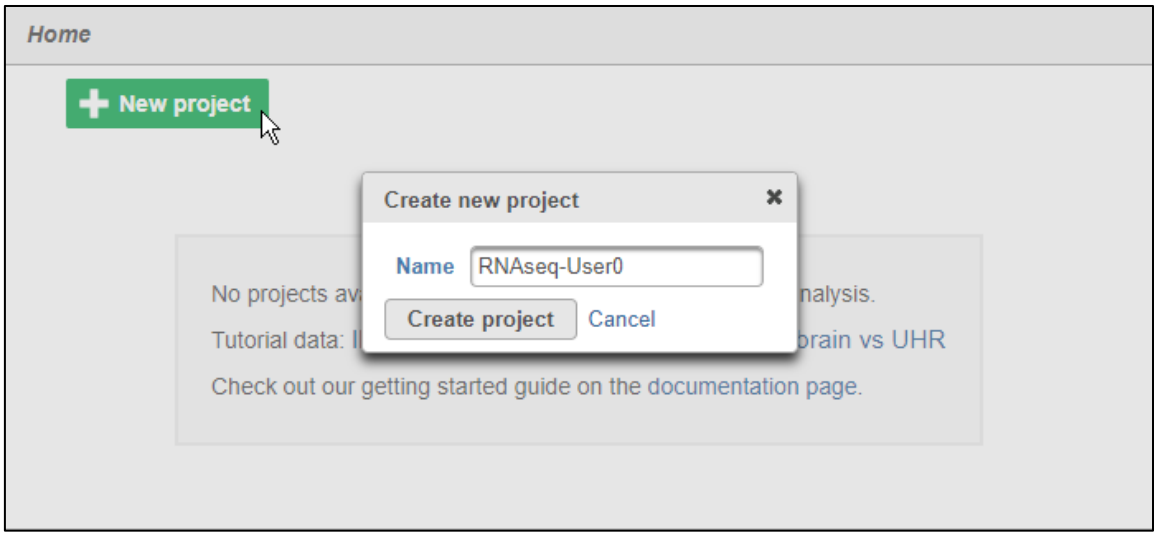
RNA-Seq Analysis in Partek® Flow®

HANDS-ON TRAINING

Partek Incorporated
support@partek.com

Login and Project Set-up

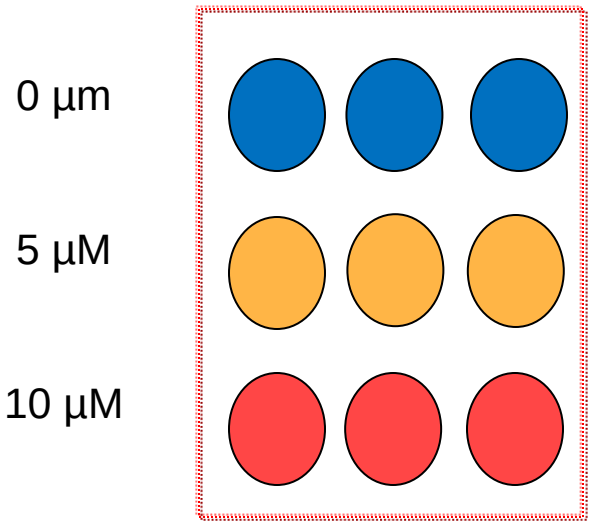
- Open Google Chrome and go to the server URL
- Log in using the username and password given to you
- This will open to the Partek Flow homepage
- Click **New Project** and enter project name: RNAseq-[username]
- This will create a new project



Notes:

Experiment Description

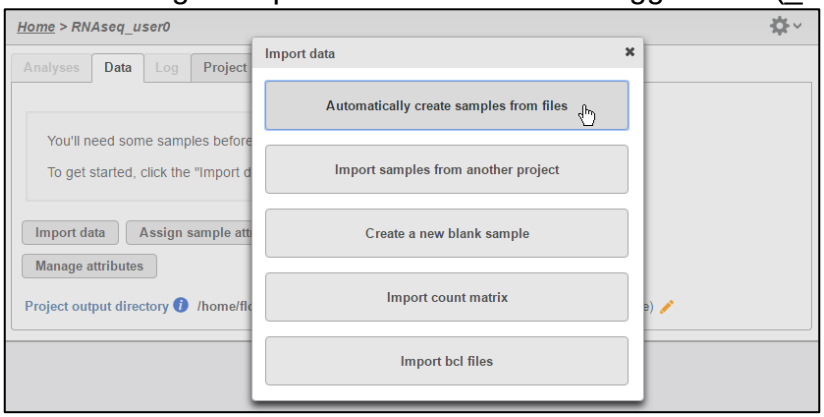
- HT29 colon cancer cells exposed to 5-aza drug with 3 different doses
    - 0 µM (Control)
    - 5 µM
    - 10 µM
- *Goal: Identify differentially expressed genes between different groups*
- mRNA purified and sequenced using Illumina HiSeq (Paired end reads)
- Xu et al. 2013 BMC Bioinformatics (PMID: 23902433)

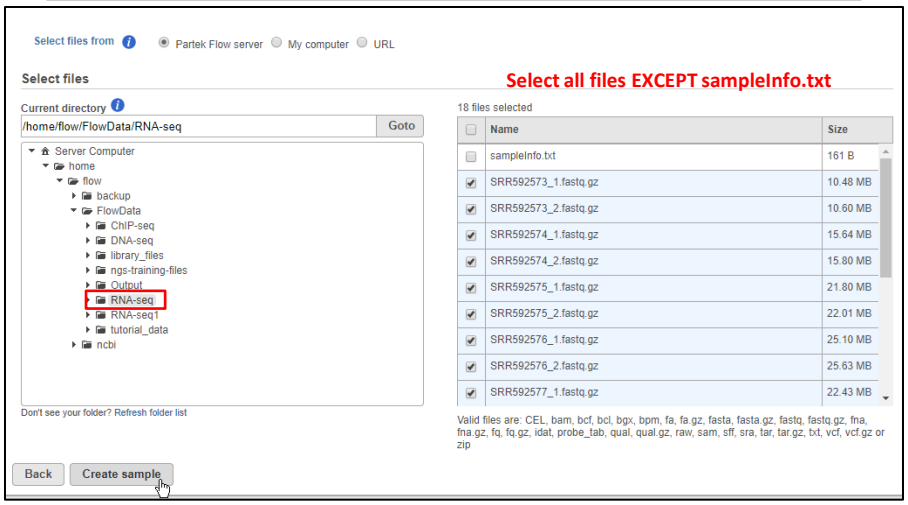


Notes:

Data Upload

- Creating a new project automatically opens up the **Data** tab
- To upload your data, click **Import data>Automatically create samples from files**
- Browse to the training RNA-Seq data
- Select *all 18 fastq.gz files* and click **Create sample**
  - Partek Flow recognizes paired-end read data if tagged with (_1 or _R1)

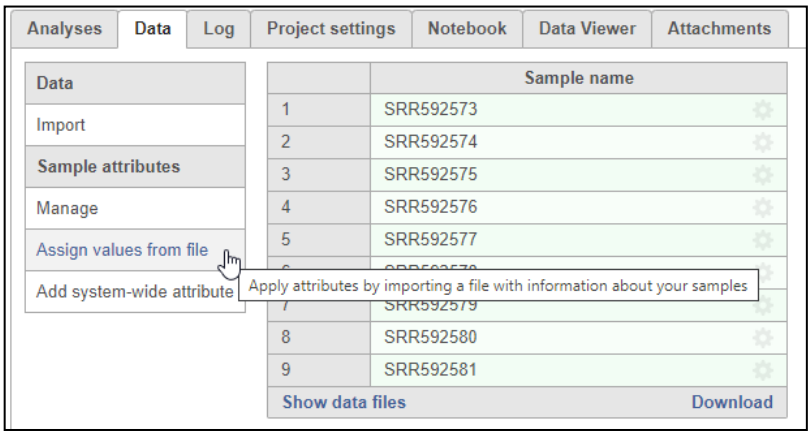




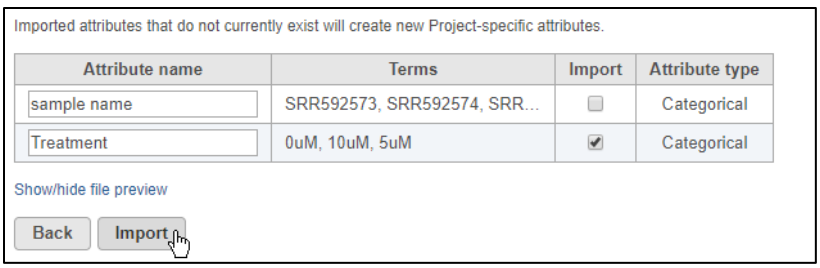
Notes:

Sample Attribute Assignment

- Assign sample attributes using a tab-delimited text file
  - Contains table with ID in 1st column, followed by corresponding treatment groups
- Click **Assign values from file**
- In the same folder, select *sampleInfo.txt*, click **Next**
- Click **Import**
- This will assign treatment groups to all samples



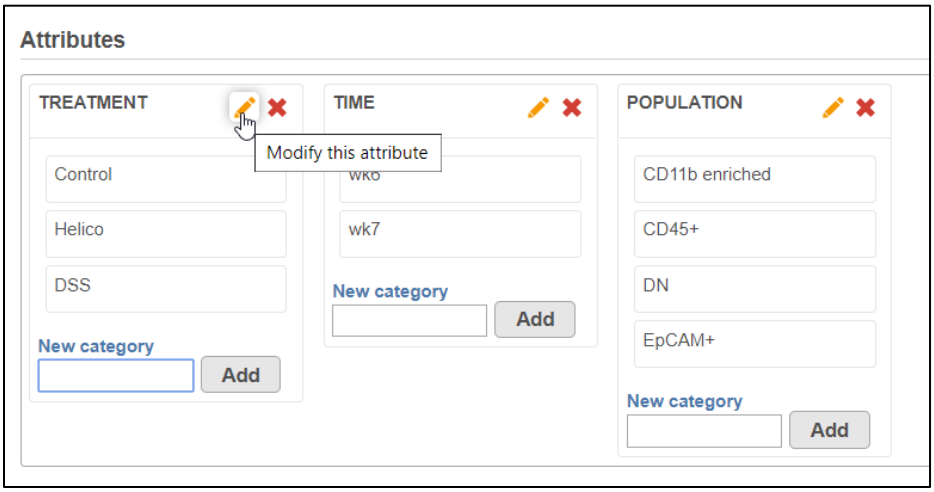
sample name	Treatment
SRR592573	0uM
SRR592574	0uM
SRR592575	0uM
SRR592576	5uM
SRR592577	5uM
SRR592578	5uM
SRR592579	10uM
SRR592580	10uM
SRR592581	10uM



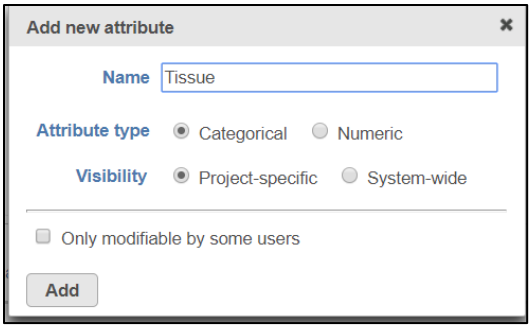
Notes:

Manage attributes

- Click and drag an attribute name to change the order the attributes on data tab and downstream display
- Click on to edit the name of the attribute or category
- Click on to delete the attribute or category
- Type a category name to add a new one within an attribute



- Click on Add new attribute to manually add another attribute

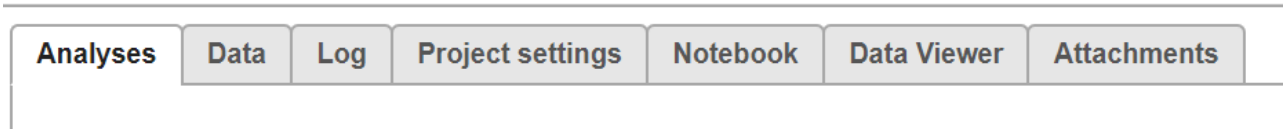


Note, the images here are just for your information, there is no need to edit the attributes for this tutorial

Notes:

Project Tabs

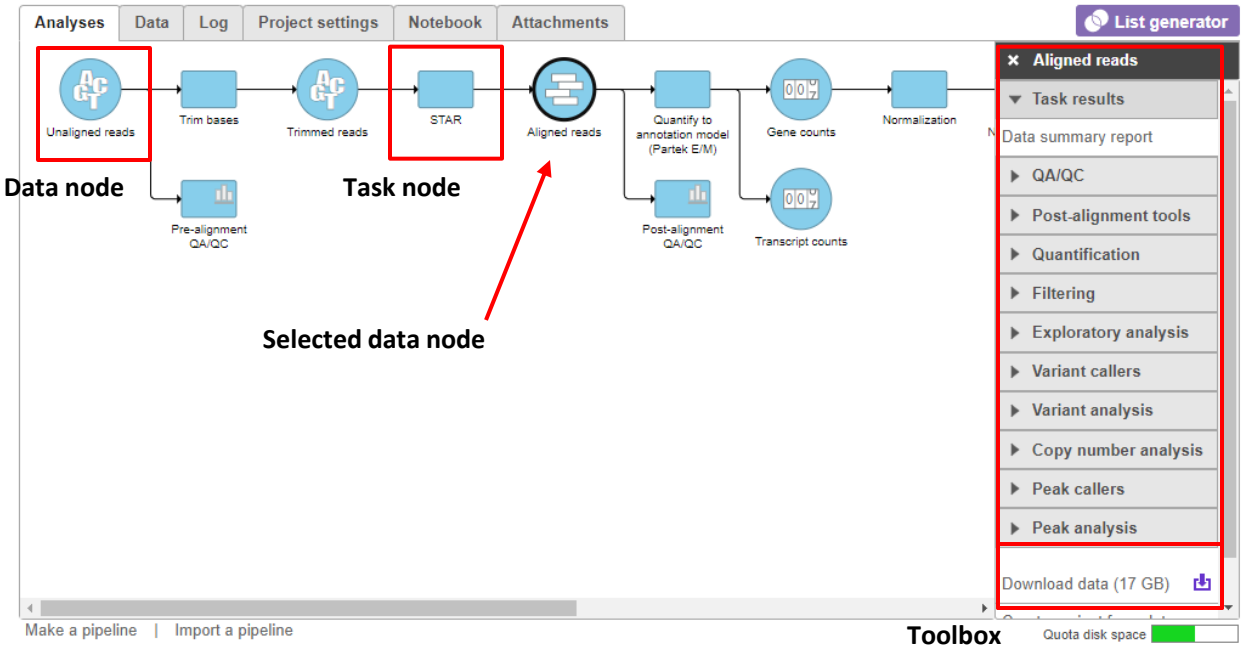
- Each project consits of seven tabs, which relate to different aspect of the project
  - Analysis
  - Data
  - Log
  - Project settings
  - Notebook
  - Data Viewer
  - Atachments



Notes: _____

Analysis Tab

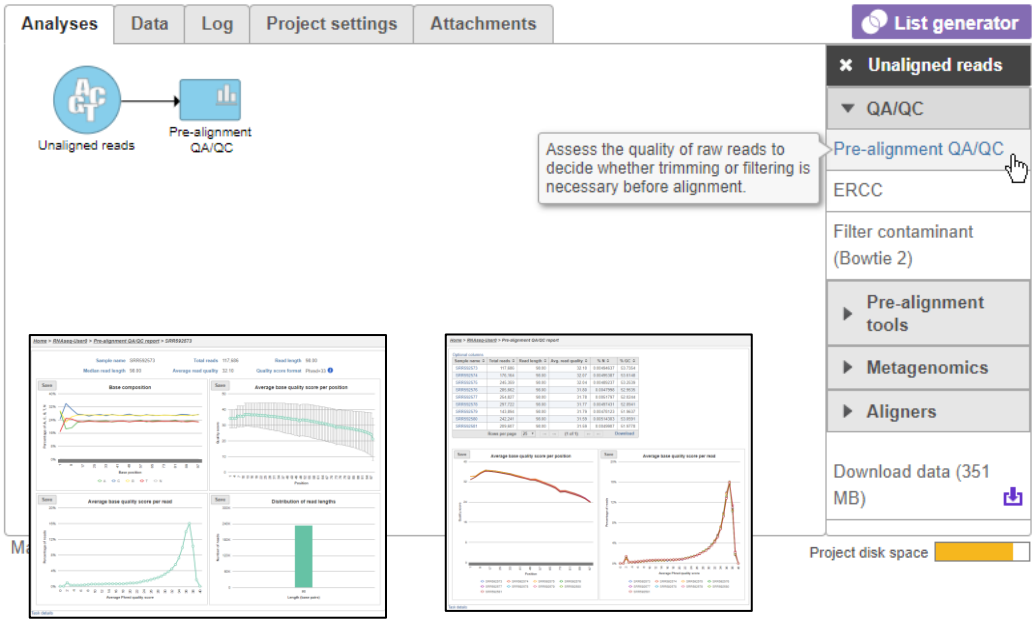
- Analysis tab is the main tab of a project. It provides an overview of the analysis and is populated by two types of symbols
  - **Data nodes** (circles): They contain data / analysis results. Actions are done on data nodes.
  - **Task node** (rectangles): These are the analysis steps performed on the data
- To launch an analysis task, select (left click) a data node. The toolbox appears on the right, showing only the tasks that can be performed on that particular node (context sensitivity). Seleced nodes are higlighted by a black circle
- Download data to export the data to local computer



Notes: _____

Pre-alignment QA/QC

- Select the **Unaligned reads** data node and select **Pre-alignment QA/QC**
- Use the default settings and click **Finish**
- Double-clicking on **the Pre-alignment QA/QC** node opens the task report
- Clicking each sample name also shows QA/QC results per sample



Quality score is -10log$_{10}$Prob

Phred Quality Score	Prob. of error	Base call accuracy
10	1/10	90%
20	1/100	99%
30	1/1000	99.9%
40	1/10000	99.99%

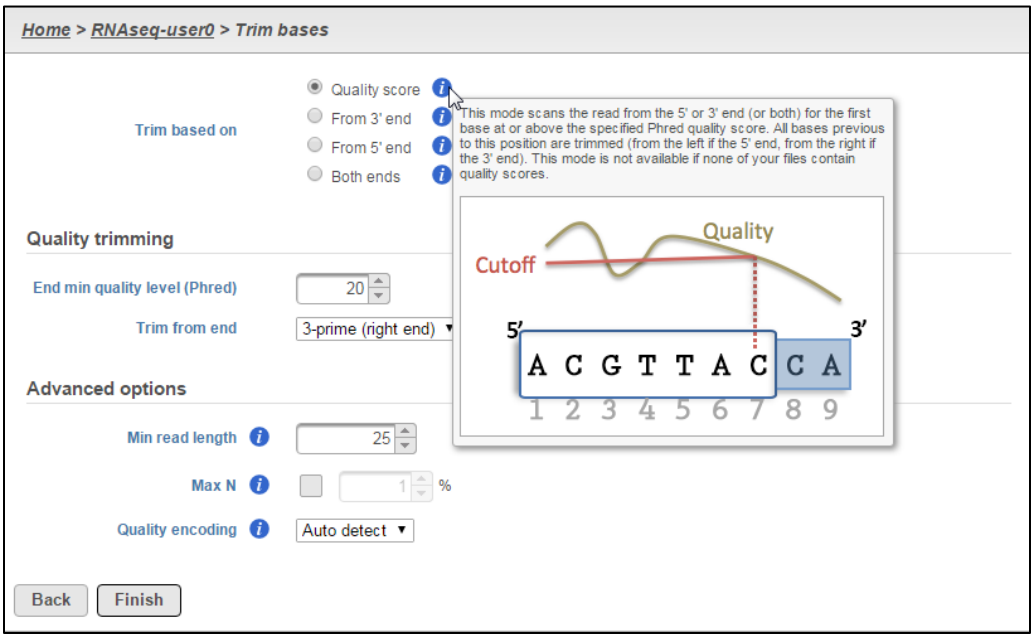
Notes:

Pre-analysis Tools: Trim Bases

Base trimming based on quality score

- Select **Unaligned reads** data node
- Click **Trim bases** from the **Pre-alignment tools** section in the toolbox
- Select **Trim based on: Quality score** with default settings and click **Finish**
- This will trim the reads at the 3’ end with a Phred quality score less than 20
- This produces your 1st new data node, the **Trimmed reads** data node

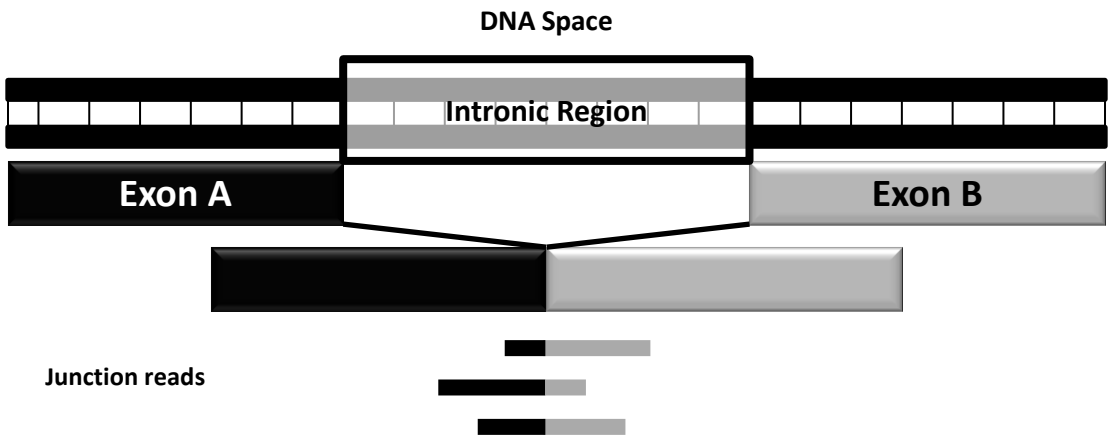
Tip: Hover over any ⓘ to get additional information about a specific option



Notes:

Aligning RNA-Seq Data

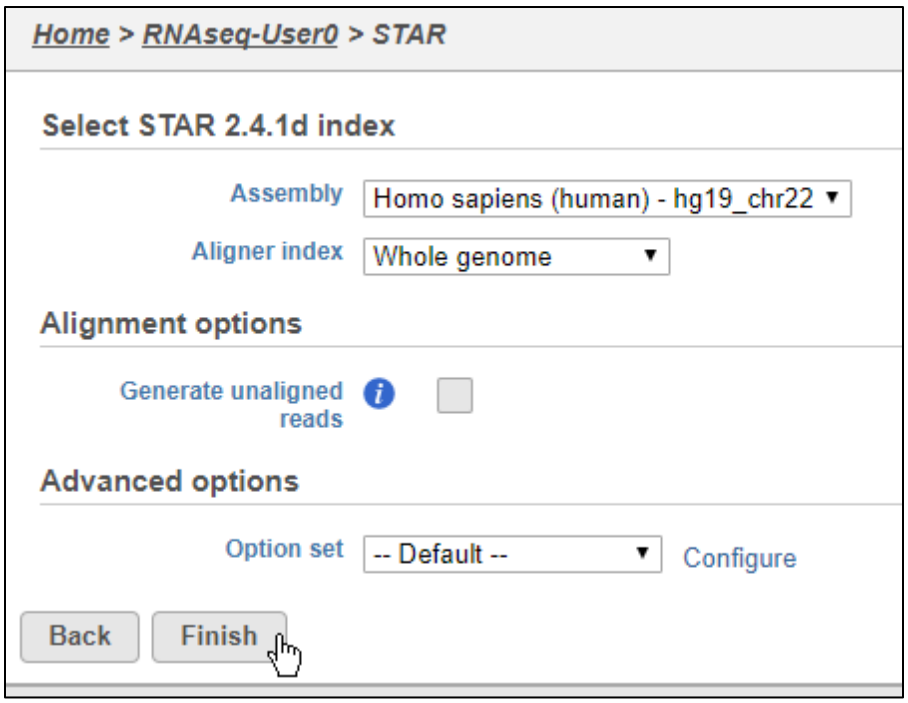
- RNA-Seq data must be aligned using an aligner that supports junction reads
- A junction read is one that spans two exons
- STAR is one of several aligners in Flow that you can use
  - Others include TopHat, TopHat2, HISAT and GSNAP



Notes:

Alignment

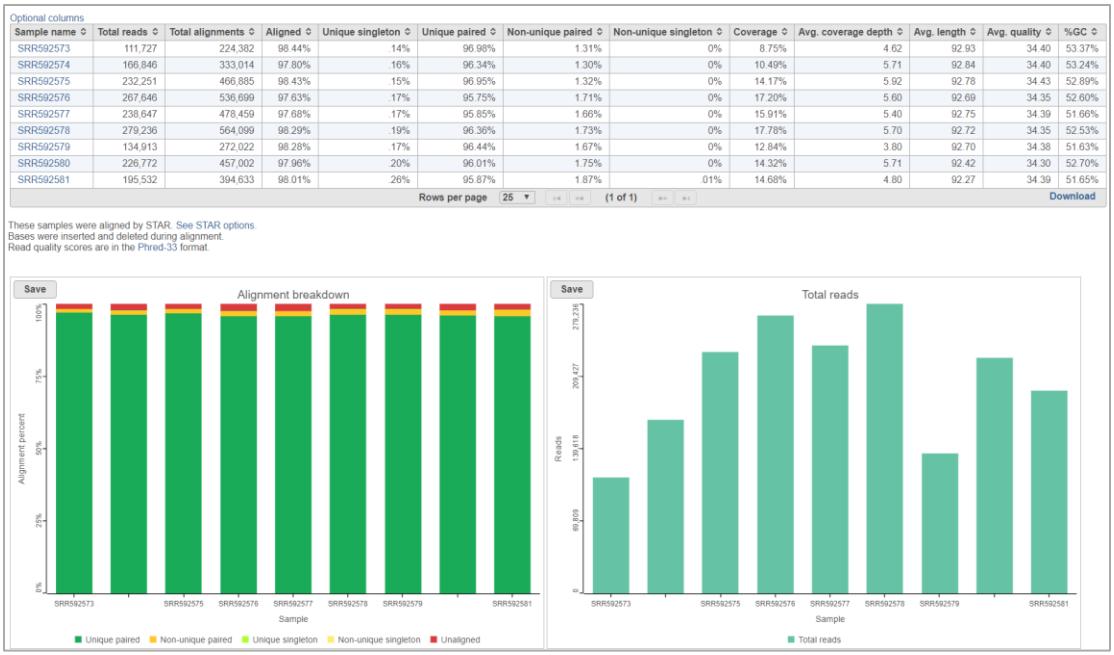
- Select the **Trimmed reads** data node
- Click **STAR** from the **Aligners** section of the menu
- Select STAR index:
  - Genome build: **Homo sapiens (human) - hg19_chr22**
  - Index: **Whole genome**
- Use the default options, click **Finish**



Notes:

Post-alignment QA/QC

- Perform Post-alignment QA/QC to assess the quality of the alignment task
- Select **Aligned reads** data node
- Click **Post-alignment QA/QC** from the **QA/QC** section of the menu
- Use default settings and click **Finish**
- Click on a sample name to get QA/QC results for that sample name



Optional columns

Sample name	Total reads	Total alignments	Aligned	Unique singleton	Unique paired	Non-unique paired	Non-unique singleton	Coverage	Avg. coverage depth	Avg. length	Avg. quality	%GC
SRR592573	111,727	224,382	98.44%	.14%	96.98%	1.31%	0%	8.75%	4.62	92.93	34.40	53.37%
SRR592574	166,846	333,014	97.80%	.16%	96.34%	1.30%	0%	10.49%	5.71	92.84	34.40	53.24%
SRR592575	232,251	466,885	98.43%	.15%	96.95%	1.32%	0%	14.17%	5.92	92.78	34.43	52.89%
SRR592576	267,646	536,699	97.63%	.17%	95.75%	1.71%	0%	17.20%	5.60	92.69	34.35	52.60%
SRR592577	238,647	478,459	97.68%	.17%	95.85%	1.66%	0%	15.91%	5.40	92.75	34.39	51.66%
SRR592578	279,236	564,099	98.29%	.19%	96.36%	1.73%	0%	17.78%	5.70	92.72	34.35	52.53%
SRR592579	134,913	272,022	98.28%	.17%	96.44%	1.67%	0%	12.84%	3.80	92.70	34.38	51.63%
SRR592580	226,772	457,002	97.96%	.20%	96.01%	1.75%	0%	14.32%	5.71	92.42	34.30	52.70%
SRR592581	195,532	394,633	98.01%	.26%	95.87%	1.87%	.01%	14.68%	4.80	92.27	34.39	51.65%

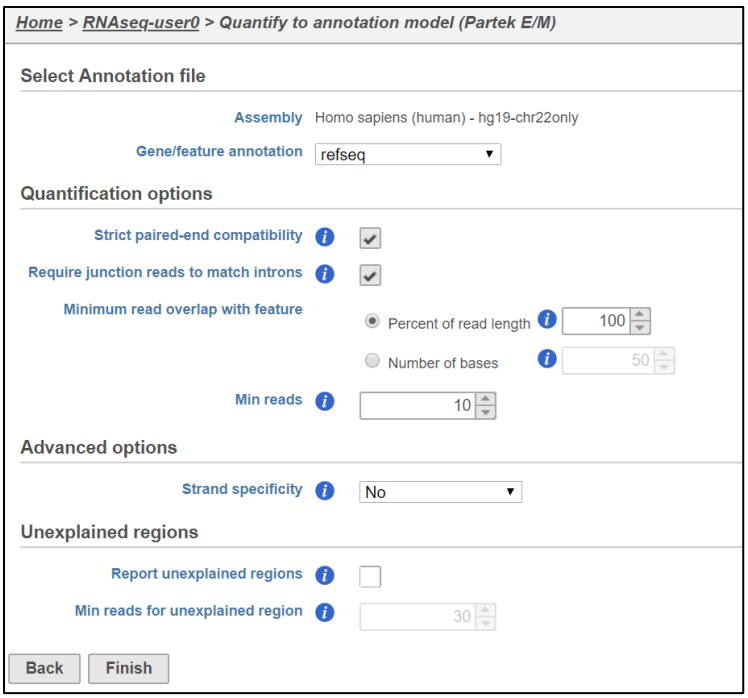
Rows per page 25 (1 of 1) Download

These samples were aligned by STAR. See STAR options.
Bases were inserted and deleted during alignment.
Read quality scores are in the Phred-33 format.

Notes: _____

Quantification to Annotation Model

- Mapping aligned reads to a database of known transcripts
  - This method can be used with any gene or feature annotation
- Select **Aligned reads** data node
- Click **Quantify to annotation model (Partek E/M)** from the **Quantification** section of the menu
- Select **RefSeq** as the Annotation model and click **Finish**
  - By default, features with total number of reads less than 10 will be filtered out



Notes:

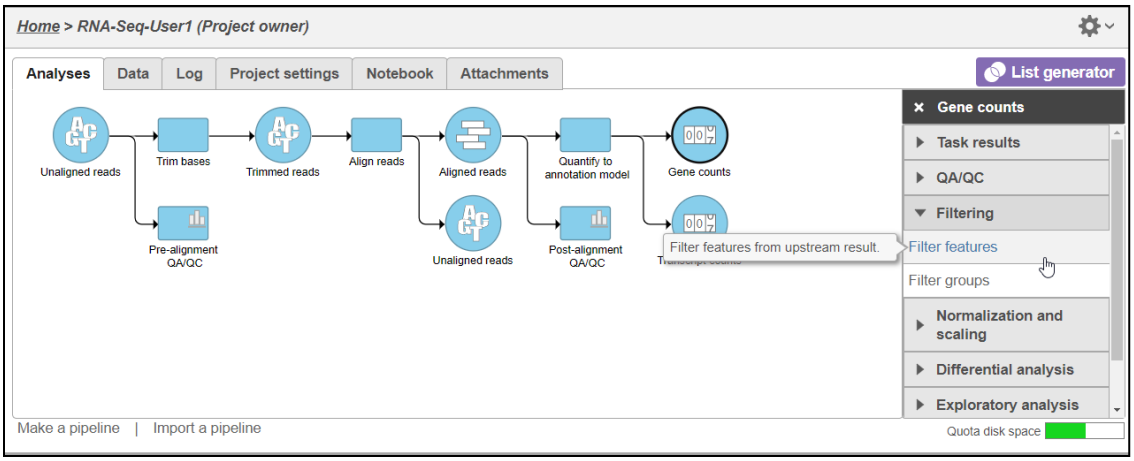
Viewing Quantification Results

- Since the RefSeq annotation has both *gene*- and *transcript-level* information, this task will generate 2 data nodes:
  - Gene counts
  - Transcript counts
- To view the results, double-click the **Gene counts** data node
- Data at this level can be downloaded as text file containing count matrix

Notes: _____

Filter Features

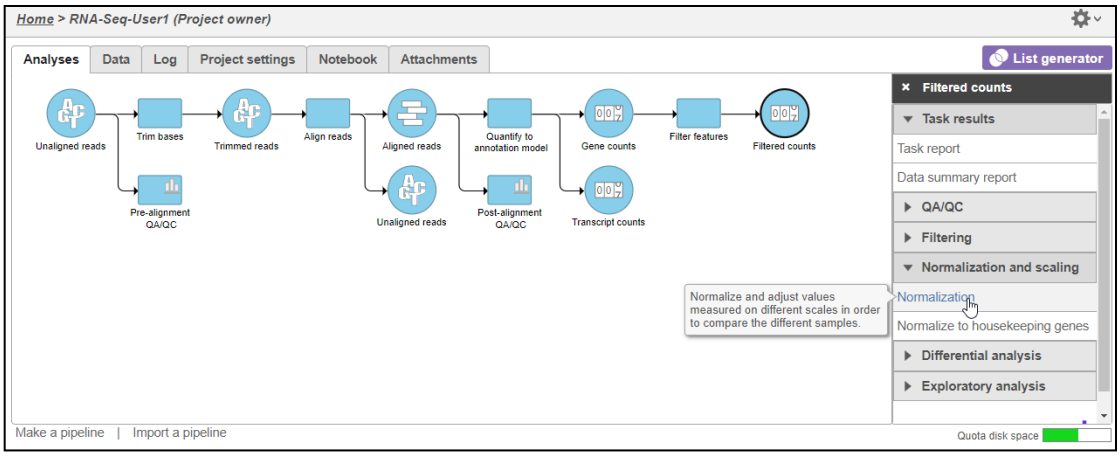
- Low expression genes maybe indistinguishable from noise, will decrease the sensitivity of DEG detection
- To filter out low expression at the gene level, select the **Gene counts** data node and click **Filter features** under the **Filtering** portion of the menu
- Choose **Filter exclude features if Maximum<=10**
- Click **Finish**

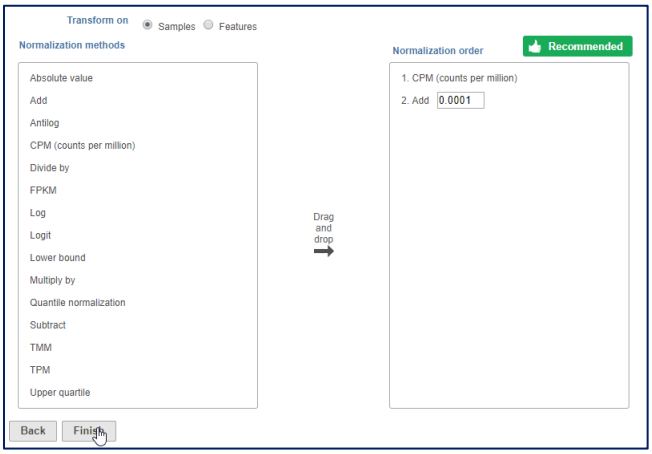


Notes: _____

Normalize Counts

- Data must be normalized before differential expression analysis
- Select the filtered gene count node and click **Normalize counts** under the **Normalization and scaling** portion of the menu
- Click the **Recommended** button and select **Finish**

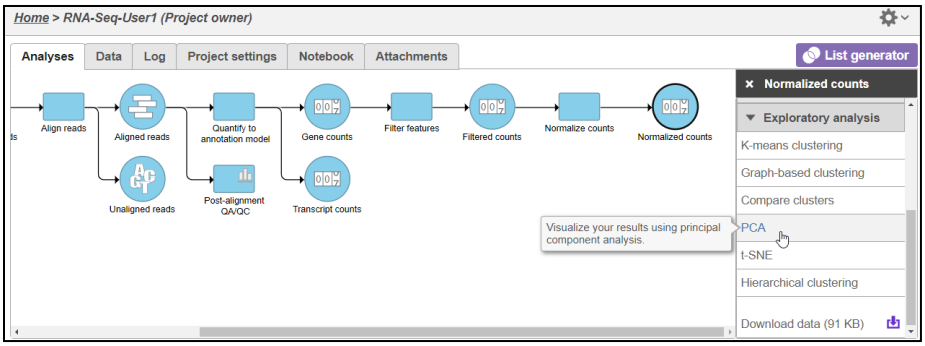


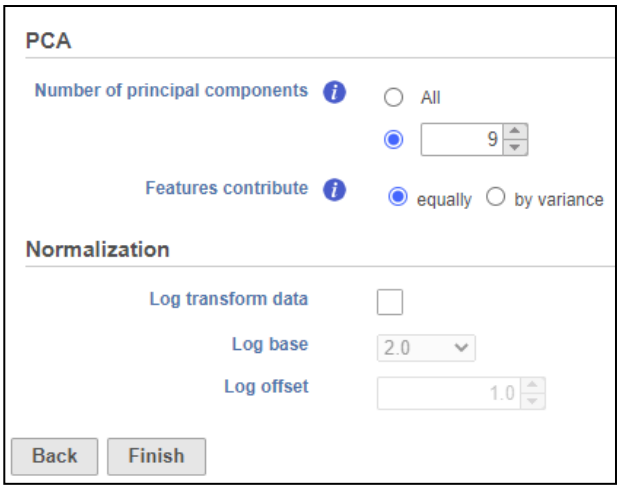


Notes:

Principal Components Analysis

- The principal components analysis (PCA) scatter plot allows you to assess relatedness between samples and identify outliers
- This can only be performed on quantified data
- To create the PCA plot, select the **Normalized counts** data node**,** click **PCA** under the **Exploratory analysis** portion of the menu, use default settings and select **Finish**

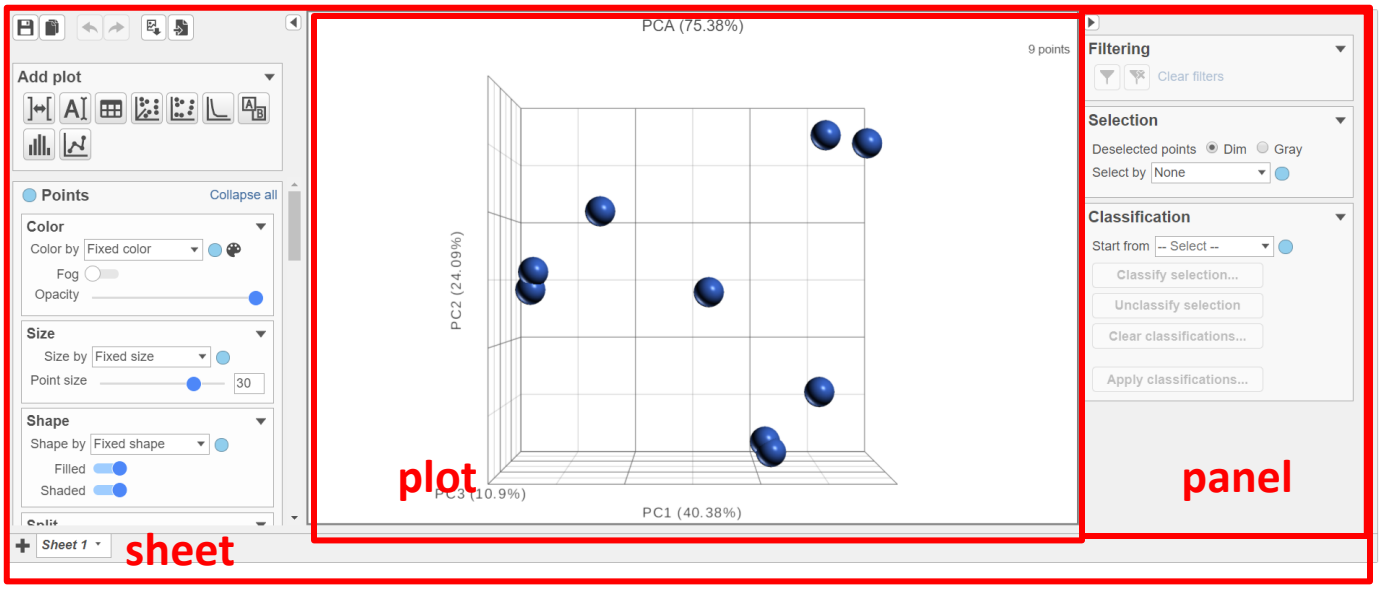




Notes: _____

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
- *Sheet:* one or more linked plots with shared controls
- *Data Viewer session*: a collection of one or more sheets



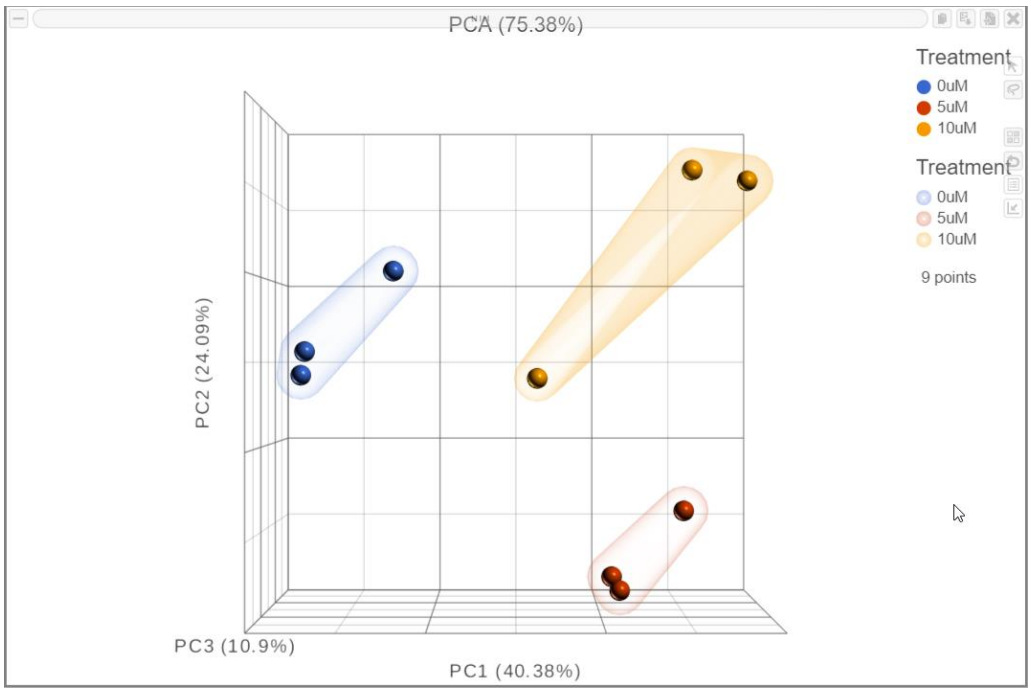
- *Panel*
  - Configuration (left)
  - Selection (right)

Notes: _____

PCA scatterplot configuration

- Collapse the right panel by clicking
- Click on PCA scatter plot to select it
- On the configuration panel on the left to configure the plot
  - Color by Treatment
  - Change the size of the points
  - Highlight by Treatment

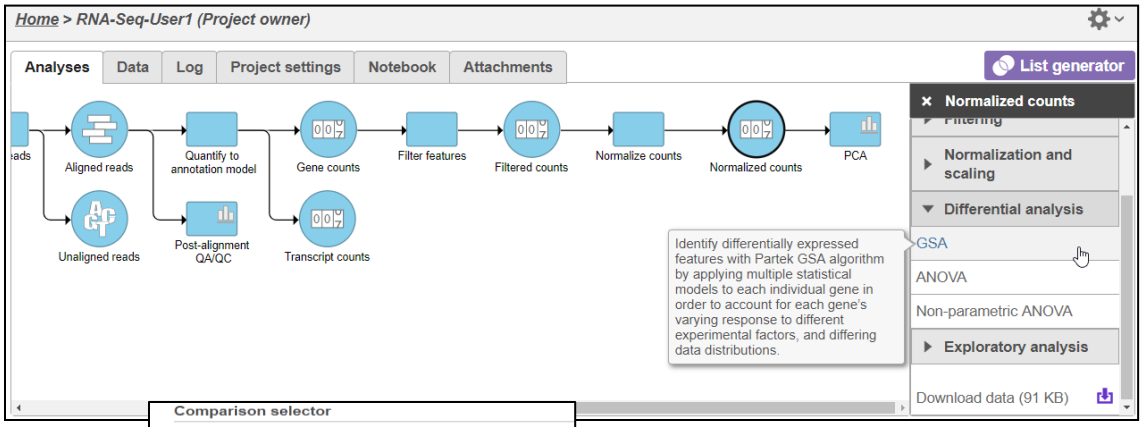
- Click on Save button to save this session

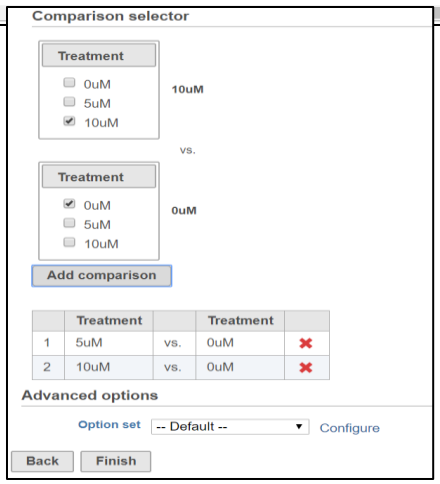


Notes:

Differential Expression Analysis

- Select the **Normalized counts** data node
- Click **GSA** from the **Differential analysis** section of the menu
- Select **Treatment** as an attribute to include in statistical test and click **Next**
- Setup the following comparisons and click the *Add comparison* button
  - **5uM vs 0uM**
  - **10uM vs 0uM**
- Click **Finish**

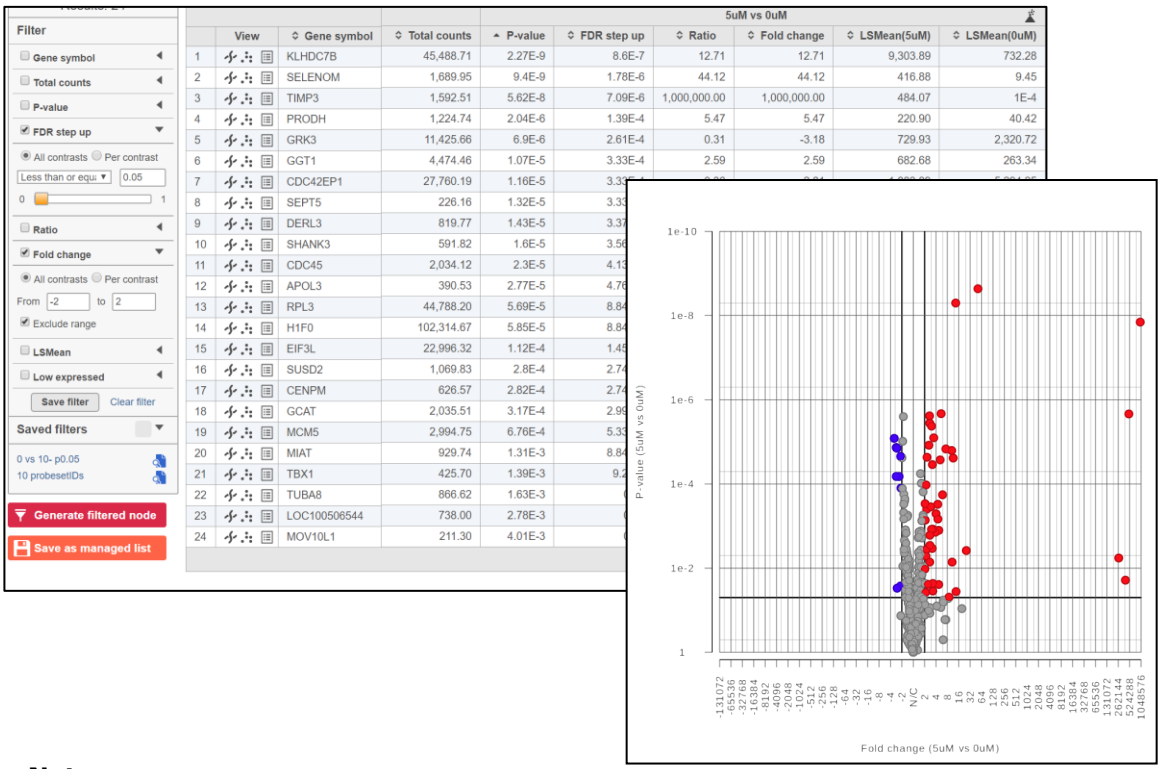




Notes:

Creating a Filtered Gene List

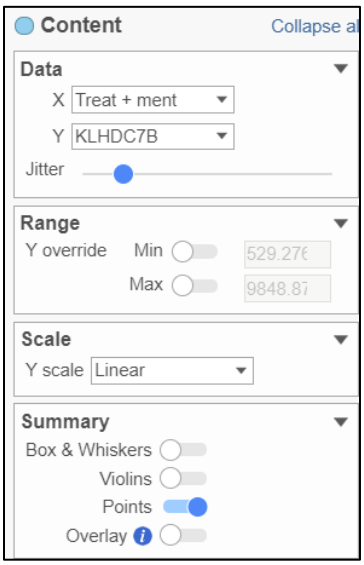
- Select **GSA** data node and then click **Task report** in the toolbox
- To get a sense of how to filter list, view the *Volcano plot* by clicking
- Under the **Gene list** section, on the **Filter** panel select:
  - **FDR step up**, then select **All contrasts** and set it to Less than or equal to 0.05
  - **Fold-change**, then select **All contrasts** and set it to From **-2** to **2**, with **Exclude range** selected
- At the bottom of the table, click **Generate filtered node**

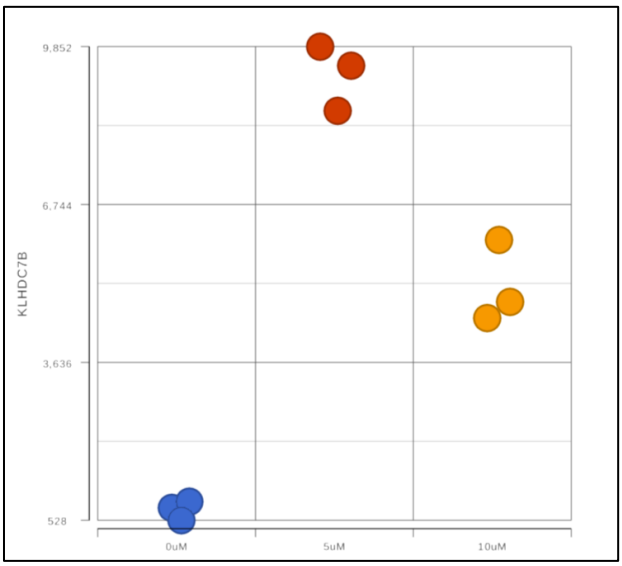


Notes:

Viewing Gene/Transcript Level Results

- Select **Filtered feature List** data node and then click **Task report** in the toolbox
- On the table, under the **View** column, select
  - to view the Dot plot
    - Jitter controls the horizontal positioning of the points
    - Box & Whiskers and Violin plots can be added
  - to see the region in Chromosome View
  - to see additional information about the statistical results

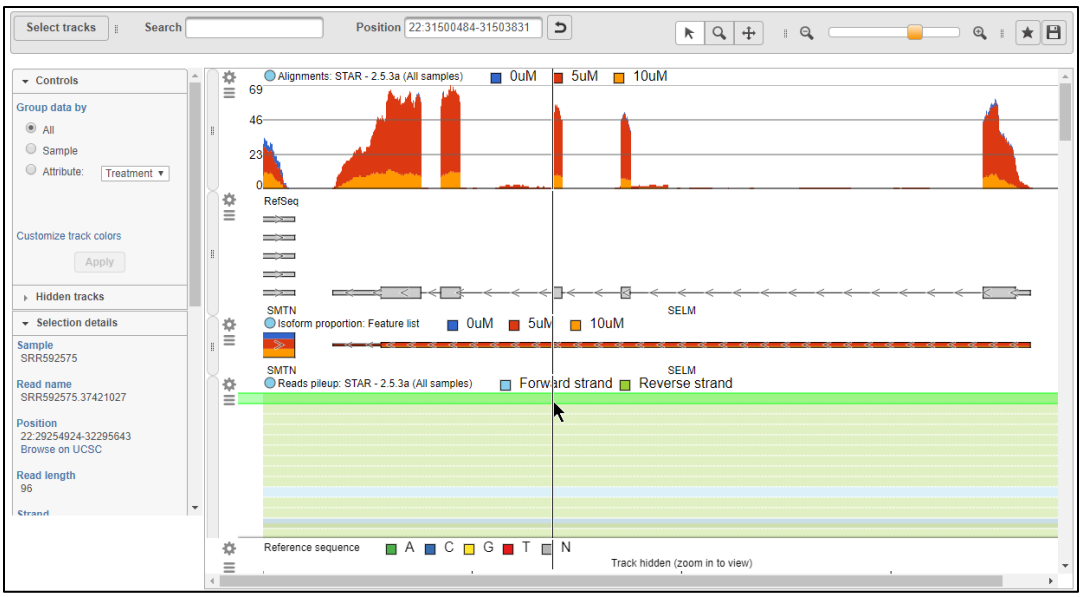




Notes: _____

Chromosome Viewer

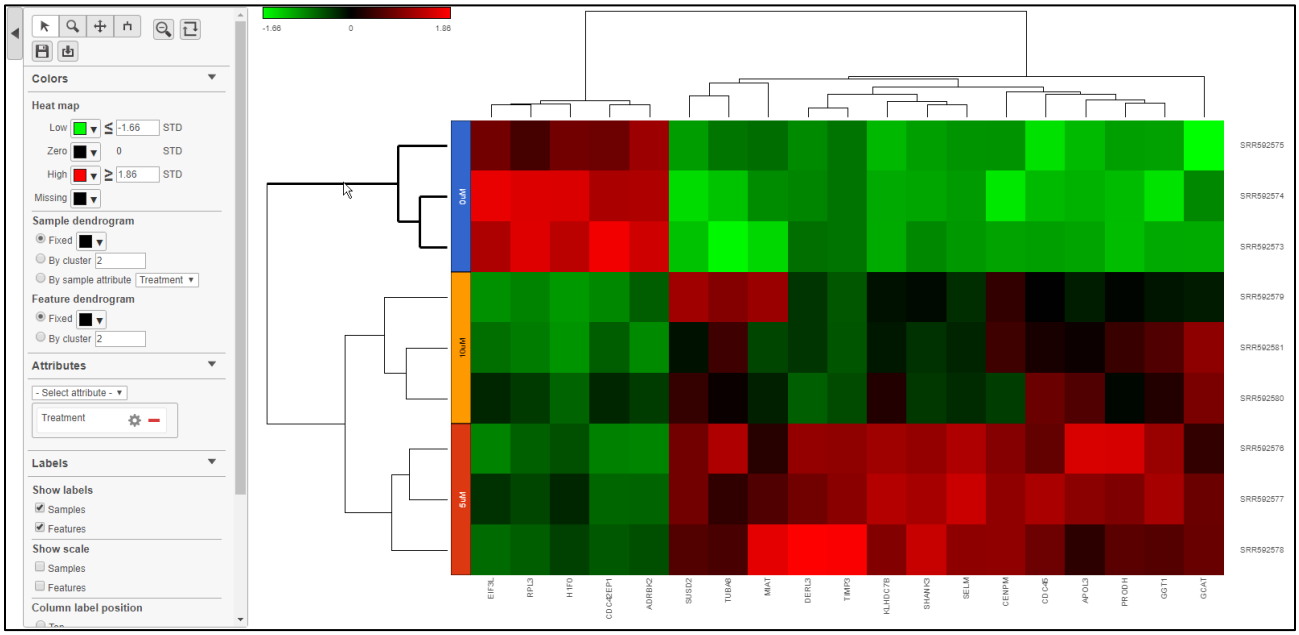
- **Select tracks** allows you to select different annotations or datasets to view together
- Sample grouping, color and transcript labeling can be edited in the **Controls** panel
- Search for any gene using the **Search** box
- Navigate to a genomic coordinate using the **Position** box
- Change and pin any displayed tracks using **Track order**
- Select any read in the reads pileup track to display additional information about the read



Notes:

Hierarchical Clustering

- Select any **Feature list** or **counts** data node to perform clustering on that list of genes/transcripts
- For this training, select the **Filtered feature list** produced after filtering
- Click **Hierarchical clustering** / **heatmap** from the **Exploratory analysis** section of the menu
- Click **Finish** to run hierarchical clustering with default settings
- Select the **Hierarchical clustering** task node and click on **Task Report**

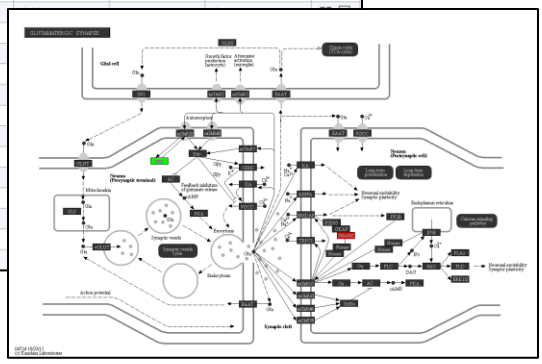


Notes: _____

Enrichment Analysis

- Perform gene set enrichment analysis using filtered list of genes
- Select **Filtered** f**eature List** data node resulting from the Filtered gene analysis task
- Select **Pathway enrichment** from the **Biological interpretation** section of the menu
- Make sure the species is Homo sapiens and click **Finish**
- Select the **Pathway enrichment** task node and click on **Task Report**
- Select  to get additional information about each specific pathway
- Click on pathway ID to see the gene network,
- Genes displayed in the network:
  - Rectangle represents gene in the network
  - colored genes are in the significant gene list, color is showing up/down regulation of the comparison
  - Black color represent genes are not in the significant list
  - Click on a gene to get detailed information about the gene at KEGG website

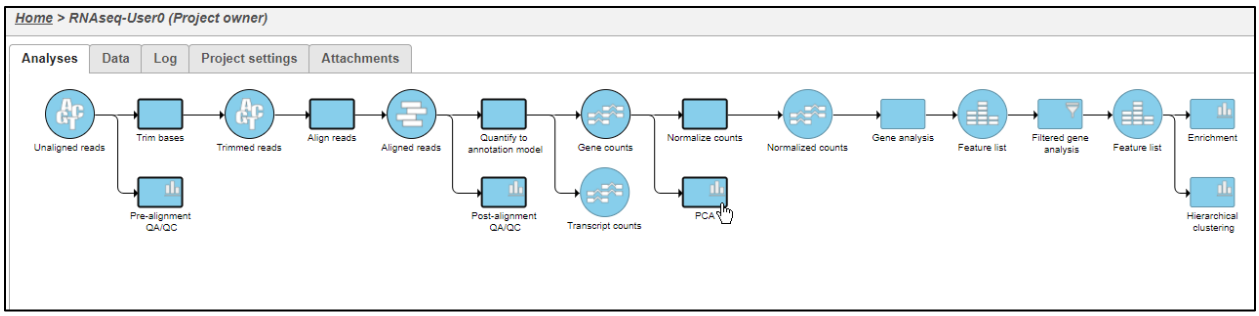
Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list
path:hsa04724	Glutamatergic synapse	4.19	0.02	2	1
path:hsa00260	Glycine, serine and threonine metabolism	2.57	0.08	1	0
path:hsa00330	Arginine and proline metabolism	2.57	0.08	1	0
path:hsa00430	Taurine and hypotaurine metabolism	2.57			
path:hsa03010	Ribosome	2.57			
path:hsa04740	Olfactory transduction	2.57			
path:hsa05032	Morphine addiction	2.57			
path:hsa05130	Pathogenic Escherichia coli infection	2.57			
path:hsa00590	Arachidonic acid metabolism	1.91			
path:hsa04145	Phagosome	1.91			
path:hsa04340	Hedgehog signaling pathway	1.91			
path:hsa04530	Tight junction	1.54			
path:hsa04540	Gap junction	1.54			
path:hsa00480	Glutathione metabolism	1.29			
path:hsa04062	Chemokine signaling pathway	1.29			
path:hsa04144	Endocytosis	1.29			

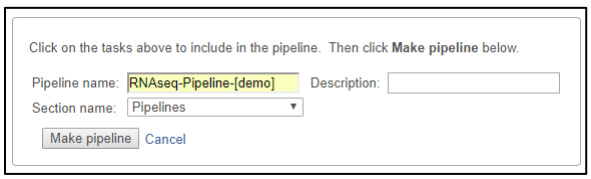


Notes:

Creating Pipelines

- Pipelines allows you to repeat the same set of tasks on different datasets
- On the Analyses tab, click **Make a pipeline** at the lower-left of the page
- Name the pipeline as **RNAseq-Pipeline-[username]**
- Select **Section name: Pipelines** then select the task nodes (rectangles) to include in the pipeline
- Click **Make pipeline** to create the pipeline

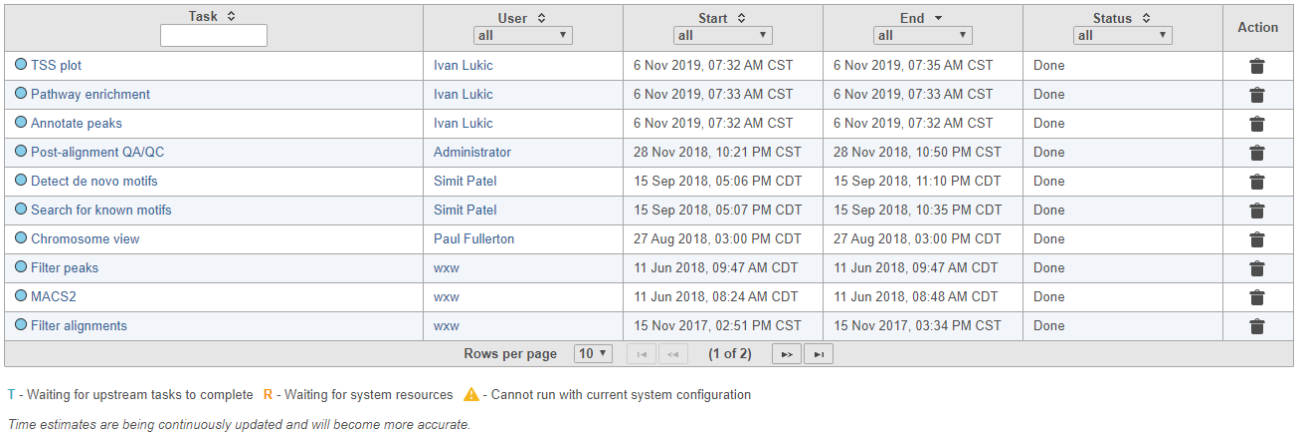




Notes:

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.

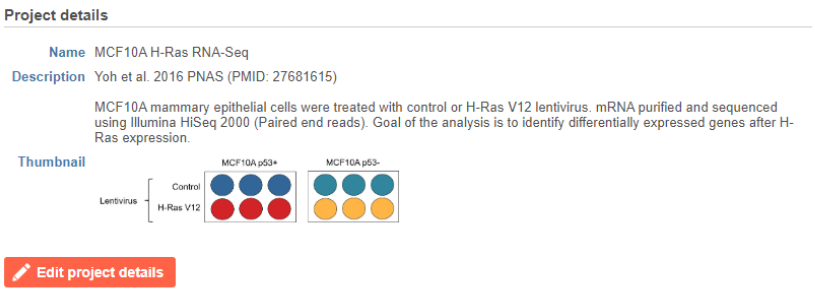
Notes:

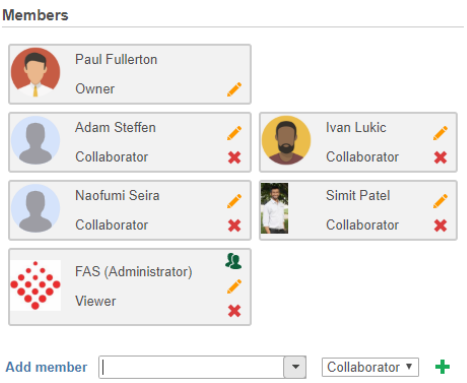
Project Settings Tab

- 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

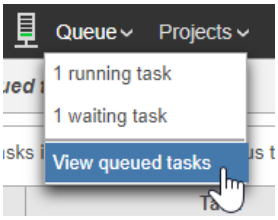




Notes: _____

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

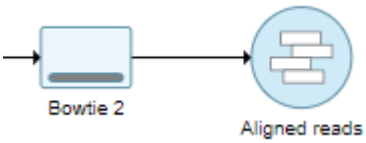

There are 2 tasks in the queue. (Anonymous tasks are not being displayed)

Status	Task	Project	User	Submitted	End	Workers	Cancel
	STAR	MCF10A H-Ras RNA-Seq	Ivan Lukic	6 Nov 2019, 01:12 PM CST	7 Nov 2019, 12:52 AM CST	iontorrent	
Waiting R	Quantify to annotation model (Partek E/M)	Ampliseq 21k Brain	Ivan Lukic	6 Nov 2019, 01:13 PM CST	Unknown		

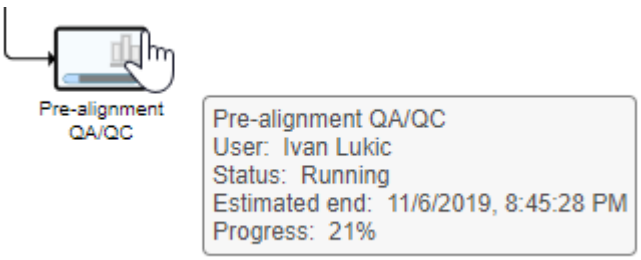
T - Waiting for upstream tasks to complete R - Waiting for system resources ⚠ - Cannot run with current system configuration

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

- 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

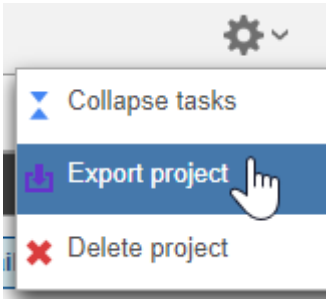


- Mousing over a task in progress provides basic info on the task



Other GUI Features: 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 

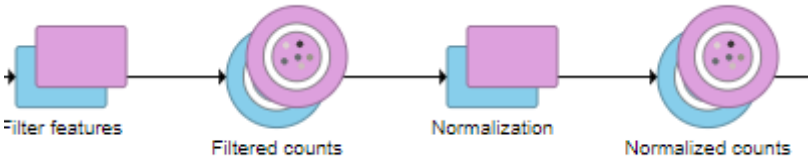
- To imported an exported project, use **Projects > Import project**

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

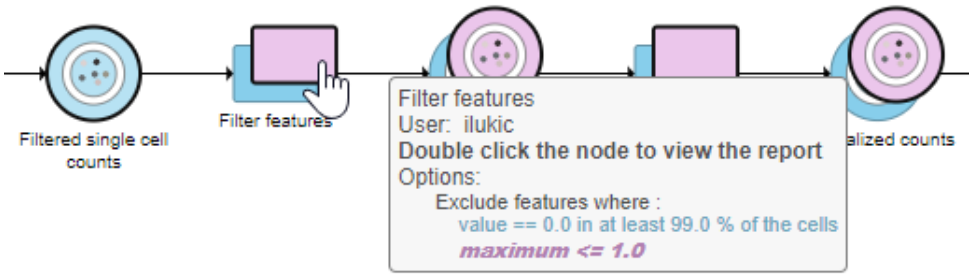
Notes:

Other GUI Features: Layers

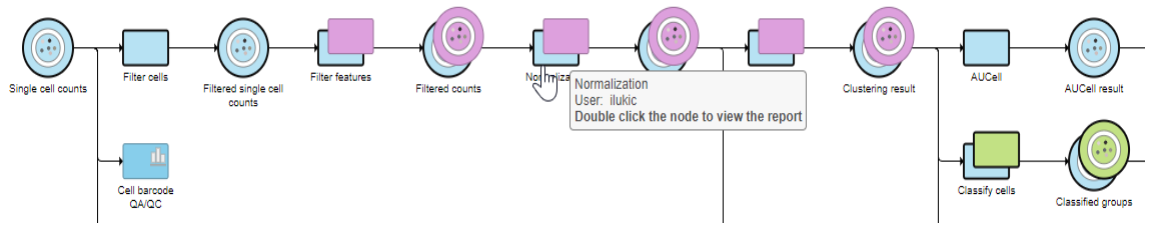
- Identical tasks ran with different task options are represented using layers of different colour. The image below shows two layers (blue and pink)



- 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

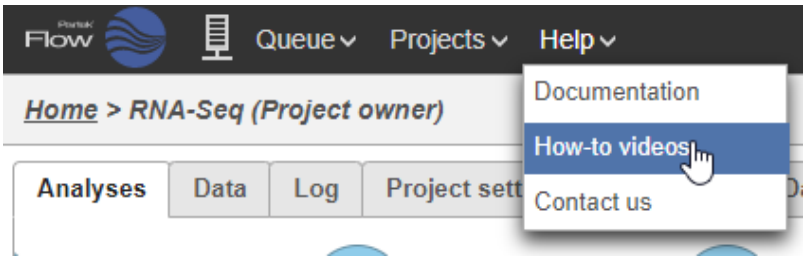


Notes: _____

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: