RNA-Seq Analysis in Partek® Flow®

HANDS-ON TRAINING

CCR, NCI

April 5th, 2016

Eric Seiser, PhD

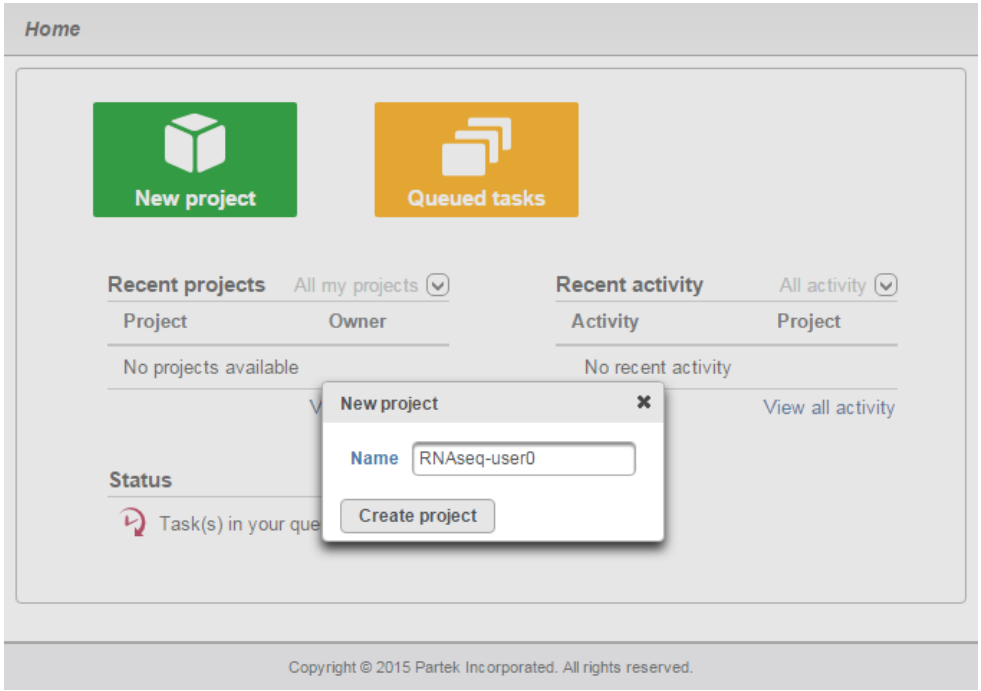
Field Application Scientist

Partek Incorporated

support@partek.com

Login and Project Set-up

- Open Google Chrome and enter: demo1.partek.com
- 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

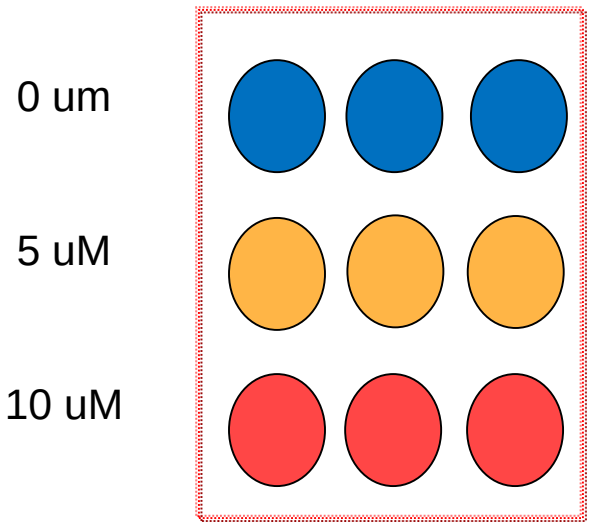


Notes:_____

Training Dataset

Data files in the project:

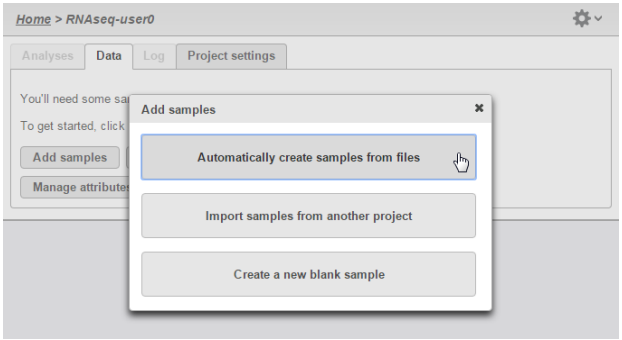
- HT29 colon cancer cells exposed to 5-aza drug with 3 different doses
    - 0 uM (Control)
    - 5 uM
    - 10 uM
- mRNA purified and sequenced using Illumina HiSeq
- Paired end reads
- Xu et al. 2013 BMC Bioinformatics (PMID: 23902433)

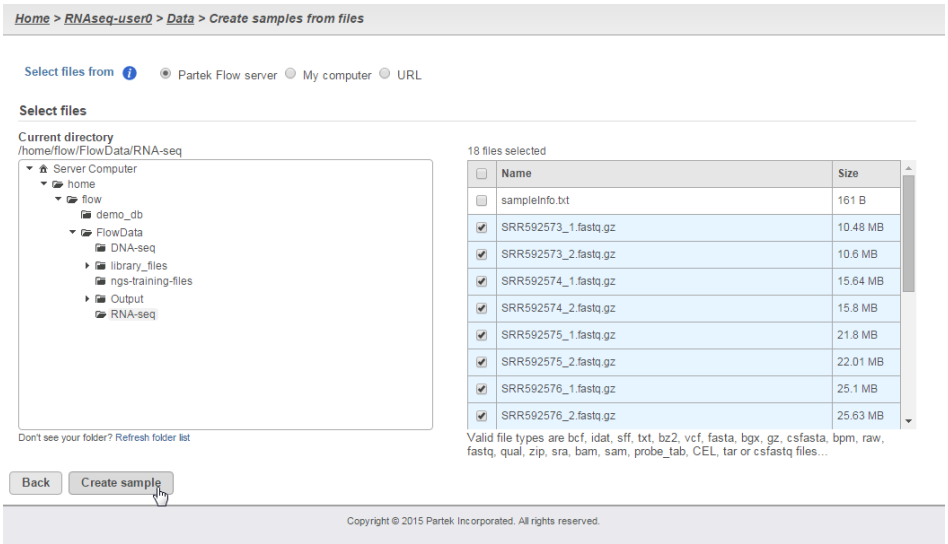


Notes:

Data Upload

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

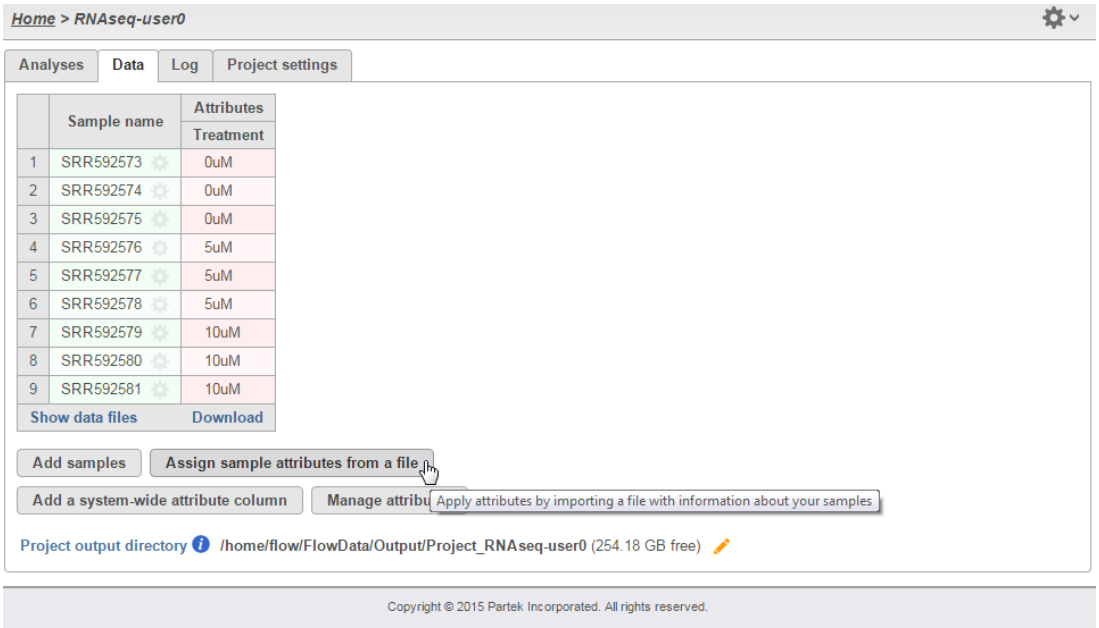




Notes:

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 sample attributes from a file**
- In the same folder, select sampleInfo.txt, click **Next**
- Click **Import**
- This will assign treatment groups to all samples



Notes:_____

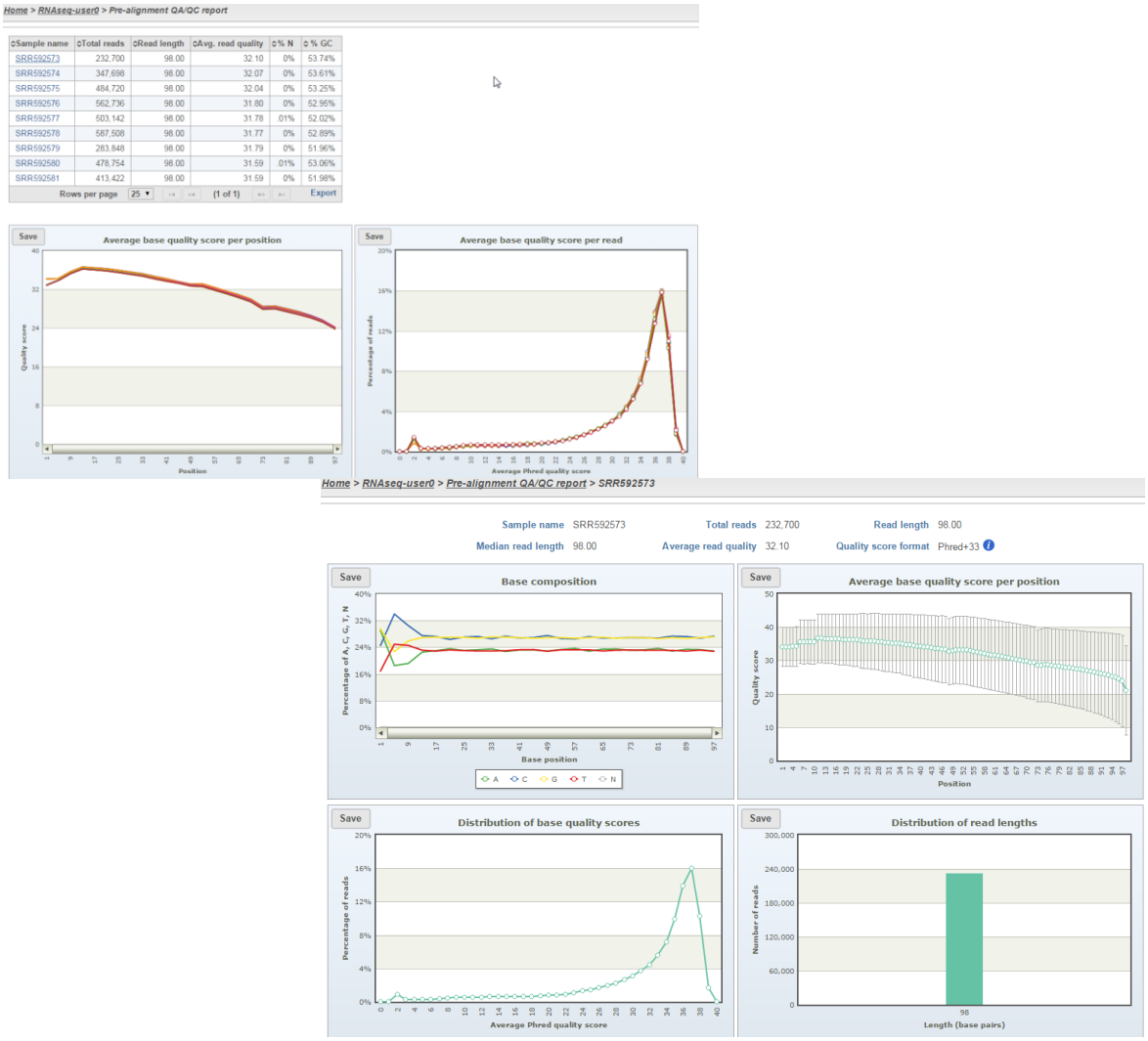
Analyses Tab Overview

- Go to the **Analyses** tab
- Your first data node, the **Unaligned reads** node appears. All data nodes are circles
- Select the **Unaligned reads** data node and select **Pre-alignment QA/QC**
- Use the default settings and click **Finish**
- This will create a new task node in the **Analyses** tab. All task nodes are rectangles
- 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

Notes:

Pre-alignment QA/QC

- Double-clicking on **the Pre-alignment QA/QC** node opens the task report
- Double-clicking each sample name also shows QA/QC results for each sample

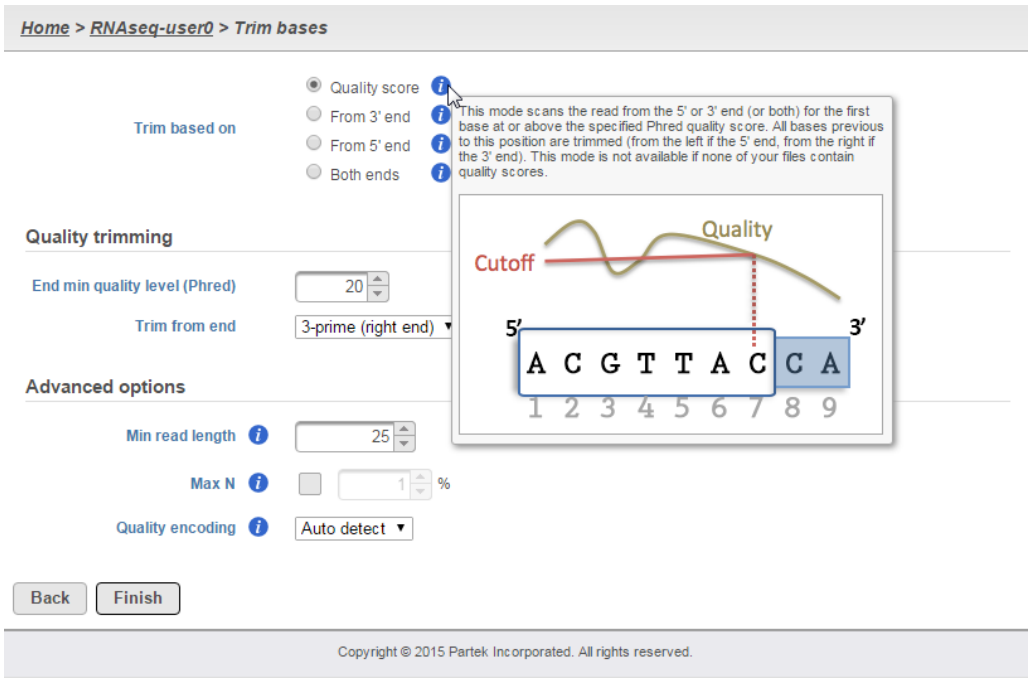


Notes:_____

Pre-analysis Tools: Trim bases

Base trimming based on quality score

- Select **Unaligned reads** data node
- Click **Trim bases** from the **Pre-analysis 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 parameter

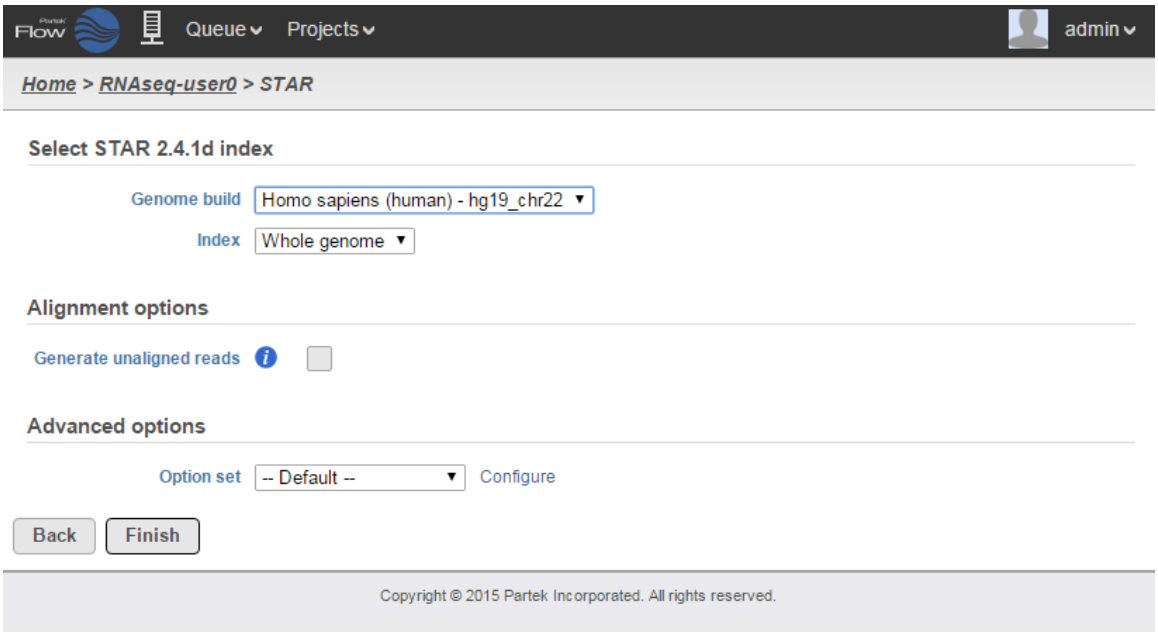


Notes:_____

Alignment

Using STAR aligner (can do junction reads)

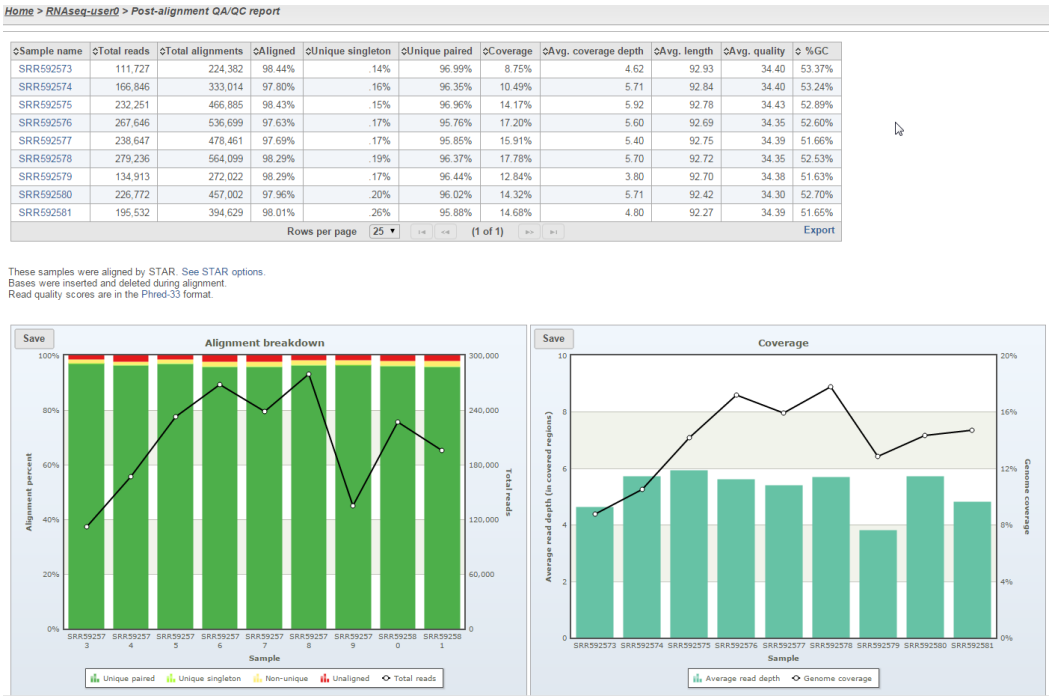
- Select the **Trimmed reads** data node
- Click **STAR** from the **Aligner** section of the menu
- Select STAR index:
  - Genome build: **Homo sapiens (human) - hg19_chr22**
  - Index: **Whole genome**
- Use the default options, click **Finish**



Notes:_____

Post-alignment QA/QC

- Perform Post-alignment QA/QC to asses 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



Home > RNAseq-user0 > Post-alignment QA/QC report

Sample name	Total reads	Total alignments	Aligned	Unique singleton	Unique paired	Coverage	Avg. coverage depth	Avg. length	Avg. quality	%GC
SRR592573	111,727	224,382	98.44%	.14%	96.99%	8.75%	4.62	92.93	34.40	53.37%
SRR592574	166,846	333,014	97.80%	.16%	96.35%	10.49%	5.71	92.84	34.40	53.24%
SRR592575	232,251	466,885	98.43%	.15%	96.96%	14.17%	5.92	92.78	34.43	52.89%
SRR592576	267,646	536,699	97.63%	.17%	95.76%	17.20%	5.60	92.69	34.35	52.60%
SRR592577	238,647	478,461	97.69%	.17%	95.85%	15.91%	5.40	92.75	34.39	51.66%
SRR592578	279,236	564,099	98.29%	.19%	96.37%	17.78%	5.70	92.72	34.35	52.53%
SRR592579	134,913	272,022	98.29%	.17%	96.44%	12.84%	3.80	92.70	34.38	51.63%
SRR592580	226,772	457,002	97.96%	.20%	96.02%	14.32%	5.71	92.42	34.30	52.70%
SRR592581	195,532	394,629	98.01%	.26%	95.88%	14.68%	4.80	92.27	34.39	51.65%

Rows per page 25 (1 of 1) Export

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.

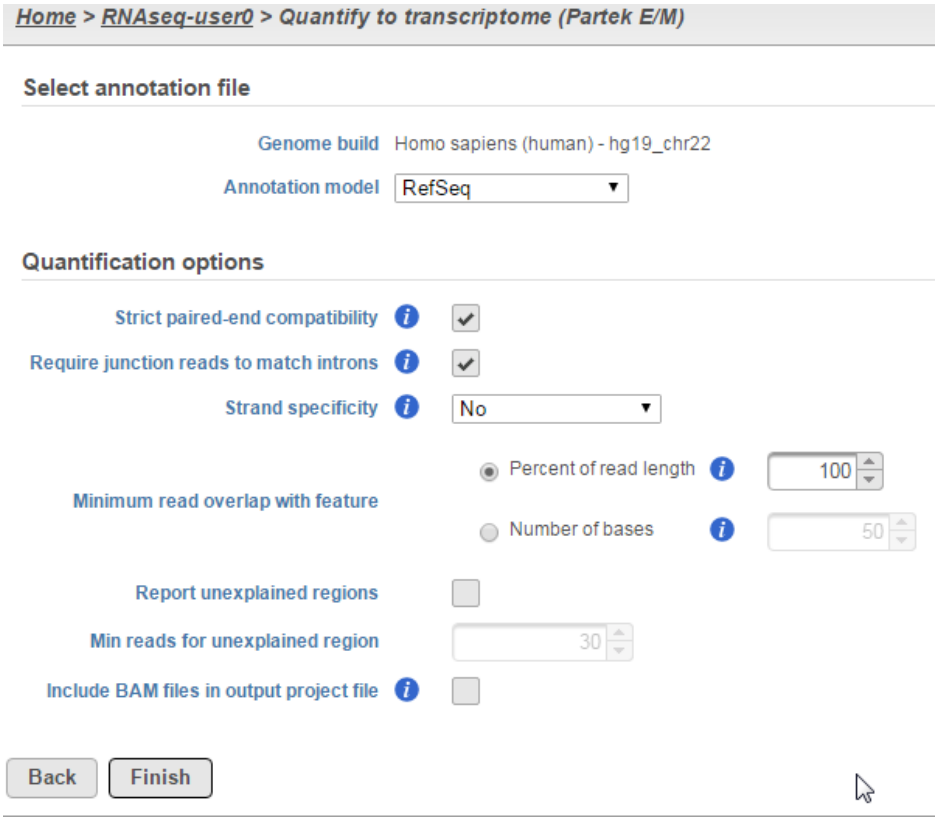
Save

Save

Notes:_____

Quantification to Transcriptome

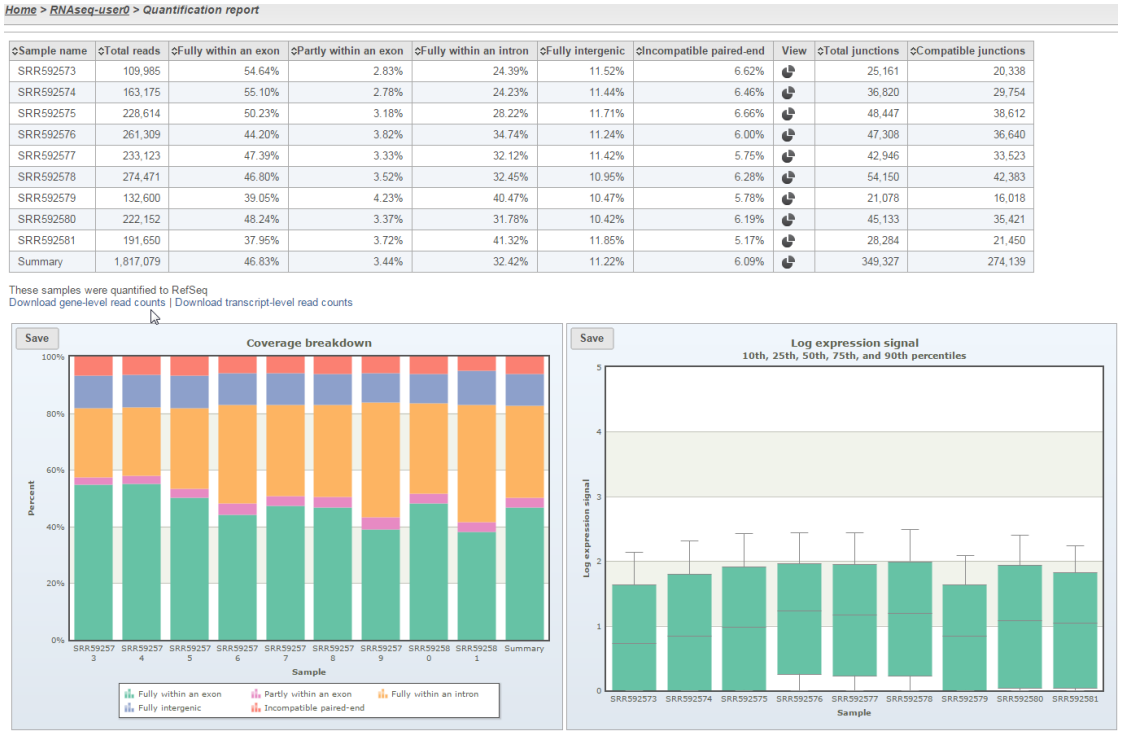
- Mapping aligned reads to transcriptome database
- Select **Aligned reads** data node
- Click **Quantify to transcriptome (E/M)** from the **RNA-Seq Analysis** section of the menu
- Select **Refseq** as the Annotation model and click **Finish**



Notes:_____

Viewing Quantification results

- To view the results of the quantification, select the **Quantification** data node
- Click **Task report** on the menu
- Click **Download gene-level read counts**, to download a .txt file of the read counts



Home > RNAseq-user0 > Quantification report

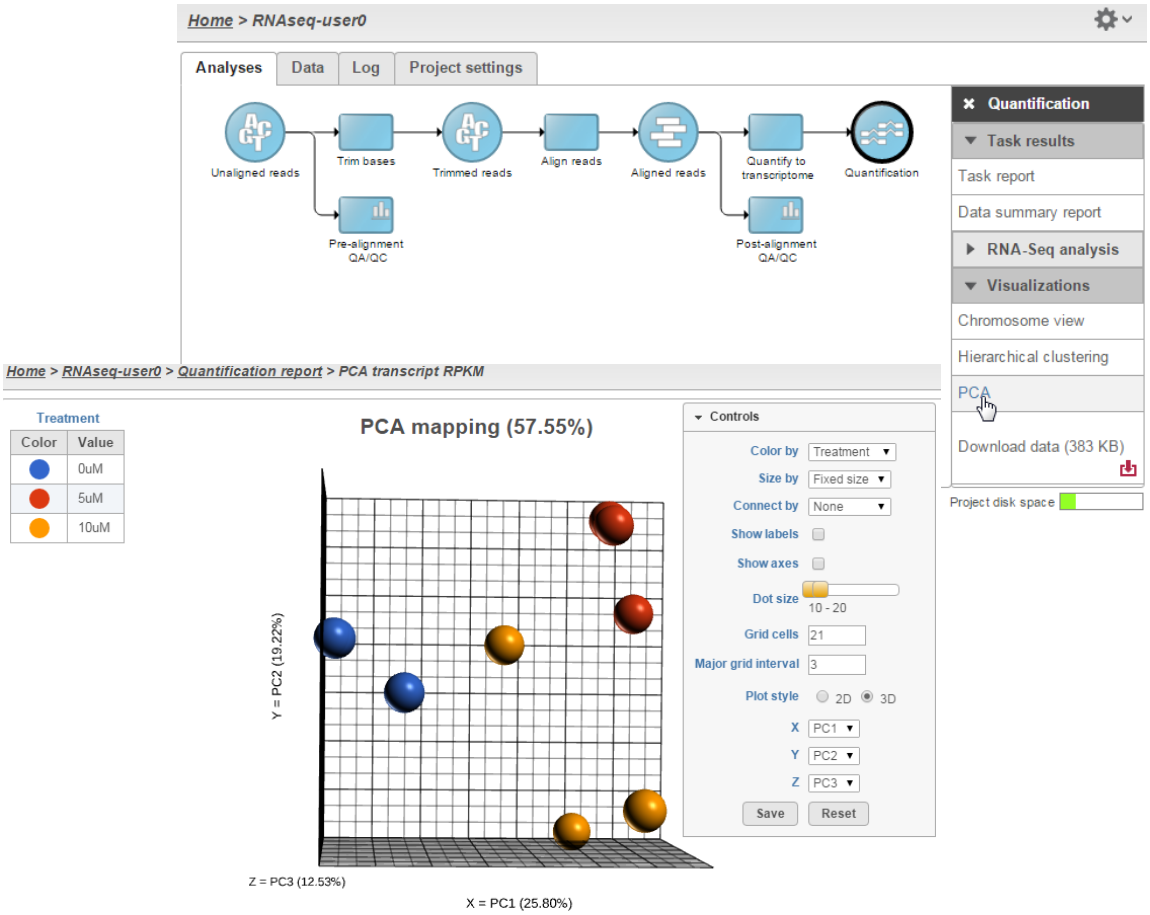
Sample name	Total reads	Fully within an exon	Partly within an exon	Fully within an intron	Fully intergenic	Incompatible paired-end	View	Total junctions	Compatible junctions
SRR592573	109,985	54.64%	2.83%	24.39%	11.52%	6.62%		25,161	20,338
SRR592574	163,175	55.10%	2.78%	24.23%	11.44%	6.46%		36,820	29,754
SRR592575	228,614	50.23%	3.18%	28.22%	11.71%	6.66%		48,447	38,612
SRR592576	261,309	44.20%	3.82%	34.74%	11.24%	6.00%		47,308	36,640
SRR592577	233,123	47.39%	3.33%	32.12%	11.42%	5.75%		42,946	33,523
SRR592578	274,471	46.80%	3.52%	32.45%	10.95%	6.28%		54,150	42,383
SRR592579	132,600	39.05%	4.23%	40.47%	10.47%	5.78%		21,078	16,018
SRR592580	222,152	48.24%	3.37%	31.78%	10.42%	6.19%		45,133	35,421
SRR592581	191,650	37.95%	3.72%	41.32%	11.85%	5.17%		28,284	21,450
Summary	1,817,079	46.83%	3.44%	32.42%	11.22%	6.09%		349,327	274,139

These samples were quantified to RefSeq
Download gene-level read counts | Download transcript-level read counts

Notes:

Principal Components Analysis

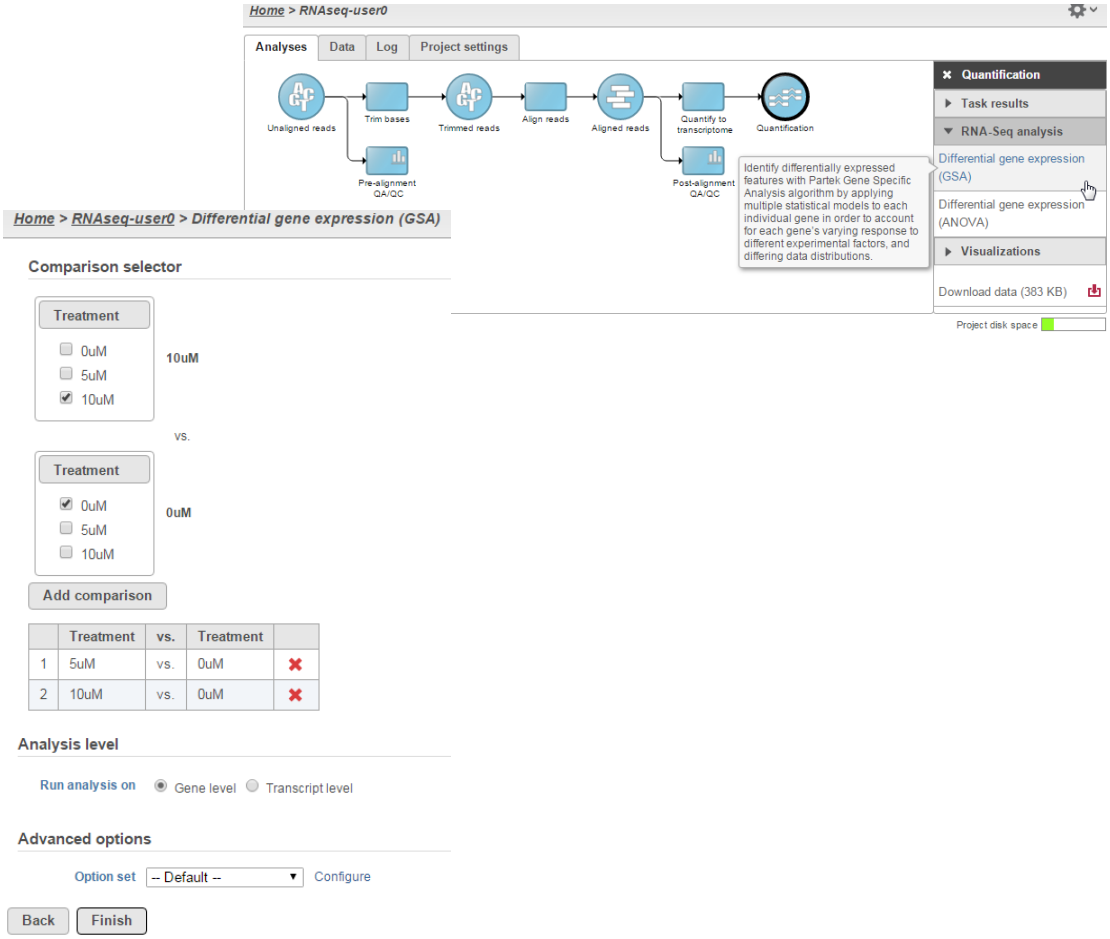
- The principal components analysis (PCA) scatter plot allows you to assess relatedness between samples and identify outliers
- This can only performed on quantified data
- To create the PCA plot, select the **Quantification** data node and select **PCA** under the **Visualizations** portion of the menu



Notes:_____

Differential Expression Analysis

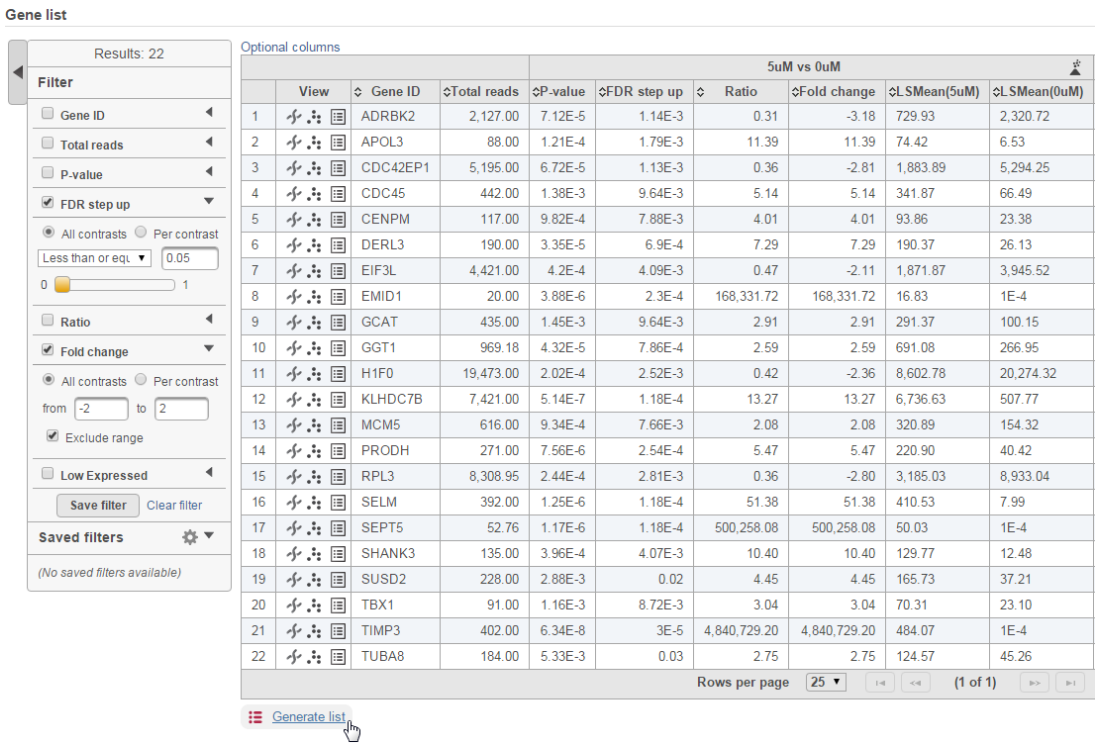
- Select the Quantification data node
- Click **Differential gene expression (GSA)** from the **RNA-Seq Analysis** section of the menu
- Select **5uM vs 0uM** and click **Add comparison**
- Select **10uM vs 0uM** and click **Add comparison**
- Select **Run analysis on Gene-level** and click **Finish**



Notes:

Creating a Filtered Gene List

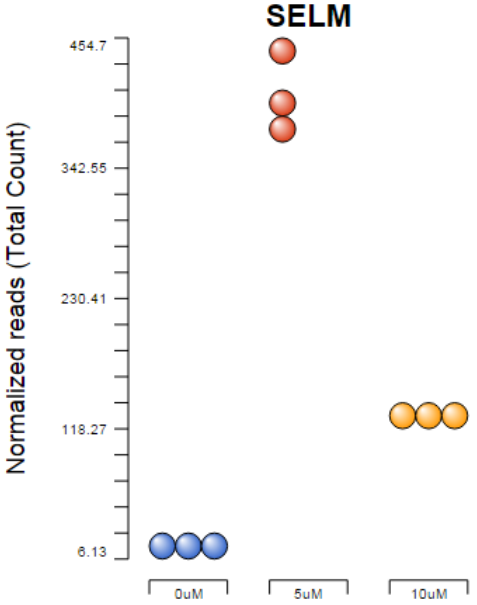
- Select **Feature List** data node and then click **Task report** in the toolbox
- Under the **Gene list** section, on the **Filter** panel select:
  - **FDR step up**, then select **All contrasts** and set it to Less than or equal to 0.05
  - **Fold-change**, then select **All contrasts** and set it to From -2 to 2, then check **Exclude range**
- At the bottom of the table, click to generate a new “Feature List”



Notes:_____

Viewing Gene/Transcript Level Results

- Select **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
  - to see the region in Chromosome View
  - to see additional information about the statistical results



Home > RNAseq-user0 > Gene-specific analysis report > Gene-specific analysis extra details

Feature information

Chromosome	22	Maximum reads	106.00
Start	31500763	Total reads	392.00
Stop	31503552	Total reads, normalized	1,662.03
Strand	-	Normalization method	Total Count
Gene ID	SELM		

Best model information

Best model	Treatment	Best AICc	80.88
Best model type	Negative binomial	Best Akaike weight	0.96

5uM vs 0uM

Multi-model estimate	3.94	Fold change	51.38
P-value (F)	1.25E-6	LSMean(5uM)	410.53
FDR step up	1.18E-4	LSMean(0uM)	7.99
Ratio	51.38		

10uM vs 0uM

Multi-model estimate	2.83	Fold change	16.96
P-value (F)	9.91E-6	LSMean(10uM)	135.48
FDR step up	6.7E-4	LSMean(0uM)	7.99
Ratio	16.96		

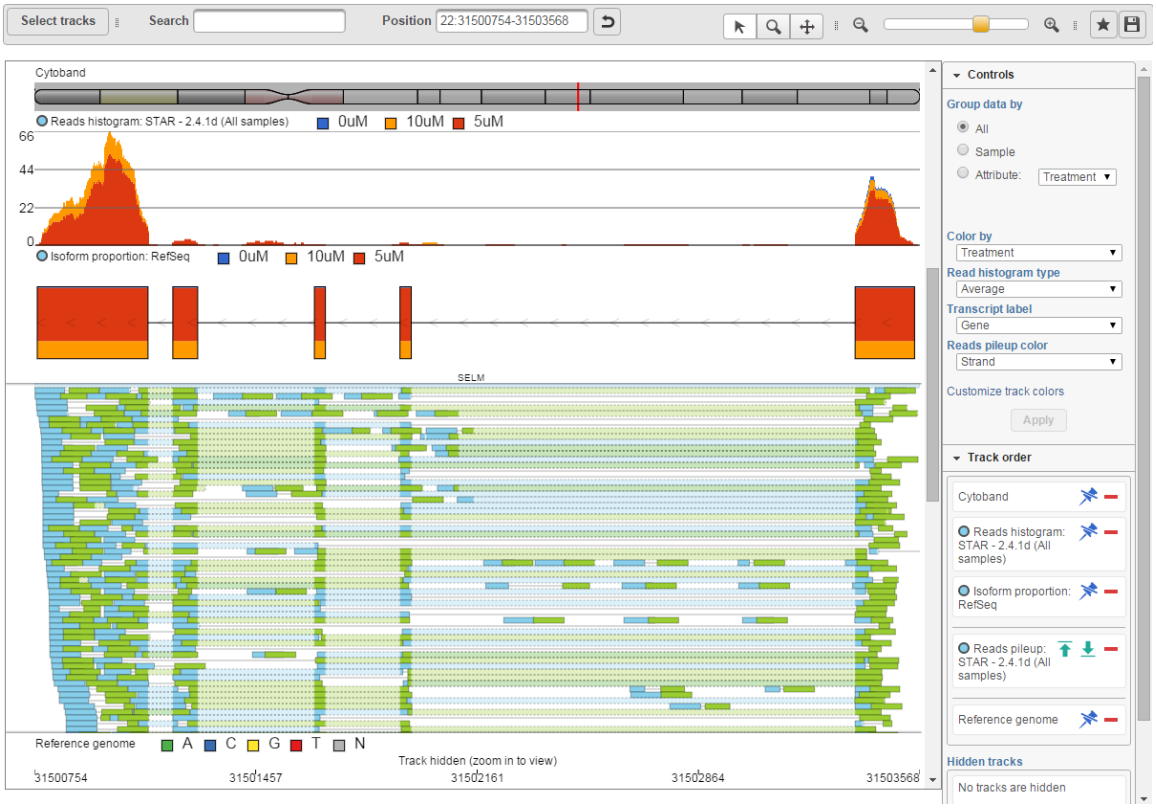
Model 1

Model	Treatment	Lognormal with shrinkage Akaike weight	0.03
Normal AICc	96.16	Negative binomial AICc	80.88
Normal Akaike weight	4.65E-4	Negative binomial Akaike weight	0.96
Lognormal with shrinkage AICc	87.54		

Notes:

Chromosome Viewer

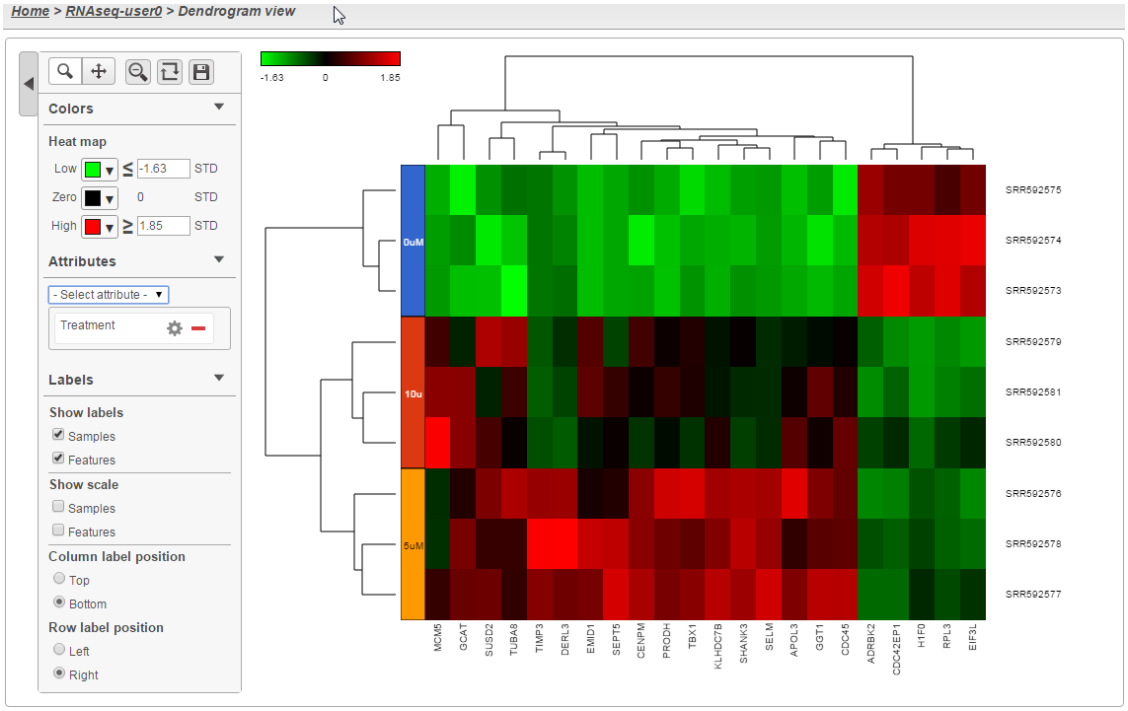
- **Select tracks** allows you to select different annotations or datasets to view together
- Sample grouping, color and transcript labeling can be edited in the **Controls** panel
- Search for any gene using the **Search** box
- Navigate to a genomic coordinate using the **Position** box
- Change and 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

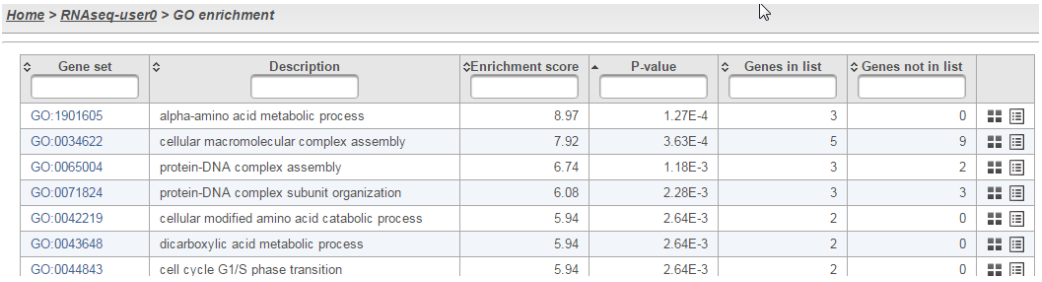
- Select any **Feature list** data node to perform clustering on that list of genes/transcripts
- For this training, select the **Feature list** produced after filtering
- Click **Hierarchical clustering** from the **RNA-Seq Analysis** section of the menu
- Use default parameters and click **Finish**
- Select the **Hierarchical clustering** task node and click on **Task Report**



Notes:_____

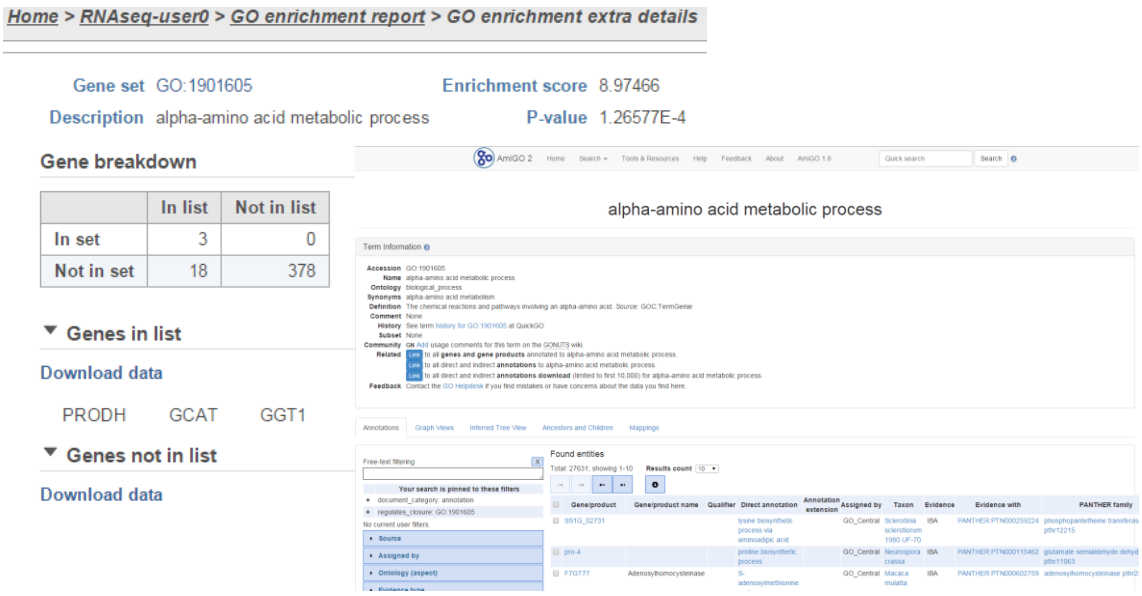
Enrichment Analysis

- Perform gene set enrichment analysis using filtered list of genes
- Select **Feature List** data node resulting from the Filtered gene analysis task
- Select **Enrichment analysis** from the **Biological interpretation** section of the menu
- Select **GO** (Gene Ontology) as Gene set annotation and then click **Finish**
- Select the **Enrichment** task node and click on **Task Report**
- Select  to get additional information about each specific pathway



Home > RNAseq-user0 > GO enrichment

Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
GO:1901605	alpha-amino acid metabolic process	8.97	1.27E-4	3	0	
GO:0034622	cellular macromolecular complex assembly	7.92	3.63E-4	5	9	
GO:0065004	protein-DNA complex assembly	6.74	1.18E-3	3	2	
GO:0071824	protein-DNA complex subunit organization	6.08	2.28E-3	3	3	
GO:0042219	cellular modified amino acid catabolic process	5.94	2.64E-3	2	0	
GO:0043648	dicarboxylic acid metabolic process	5.94	2.64E-3	2	0	
GO:0044843	cell cycle G1/S phase transition	5.94	2.64E-3	2	0	



Home > RNAseq-user0 > GO enrichment report > GO enrichment extra details

Gene set GO:1901605 Enrichment score 8.97466

Description alpha-amino acid metabolic process P-value 1.26577E-4

Gene breakdown

	In list	Not in list
In set	3	0
Not in set	18	378

Genes in list

Download data

PRODH GCAT GGT1

Genes not in list

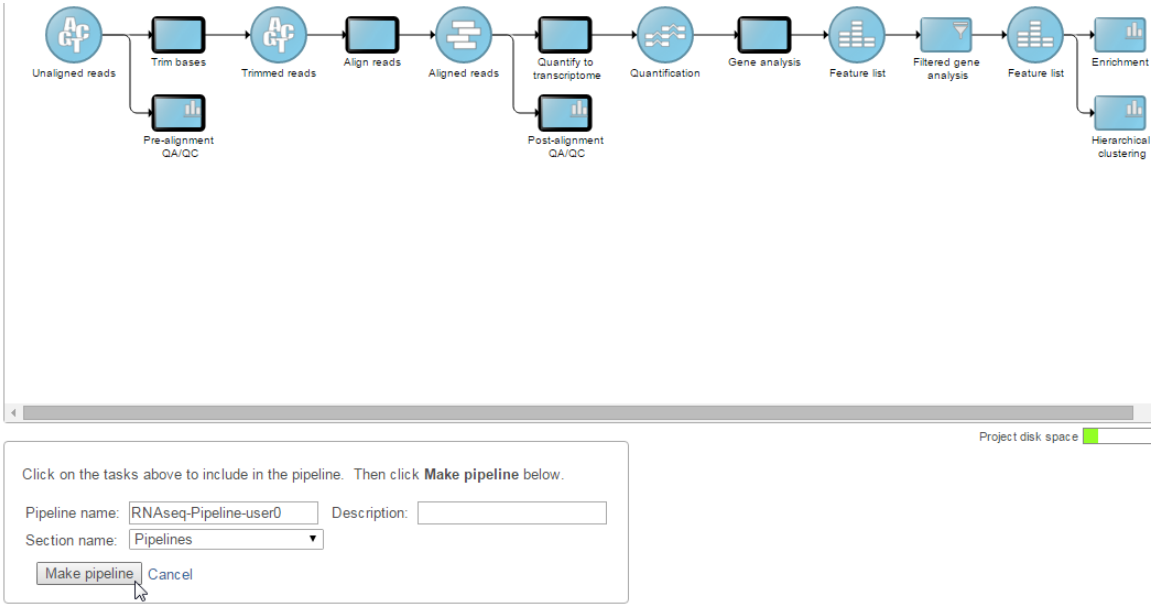
Download data

alpha-amino acid metabolic process

Notes:_____

Creating Pipelines

- Creating pipelines allows you to repeat a series of tasks on different projects
- 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:

Further Training

Self-learning

- Check out http://www.partek.com/resources-partek-flow for resources
- Recorded webinars available on Partek Incorporated's YouTube page

Regional Technical Support

- Email: support@partek.com
- Phone: +1-314-878-2329

Notes: