

RNA-Seq Analysis in Partek® Flow®

HANDS-ON TRAINING

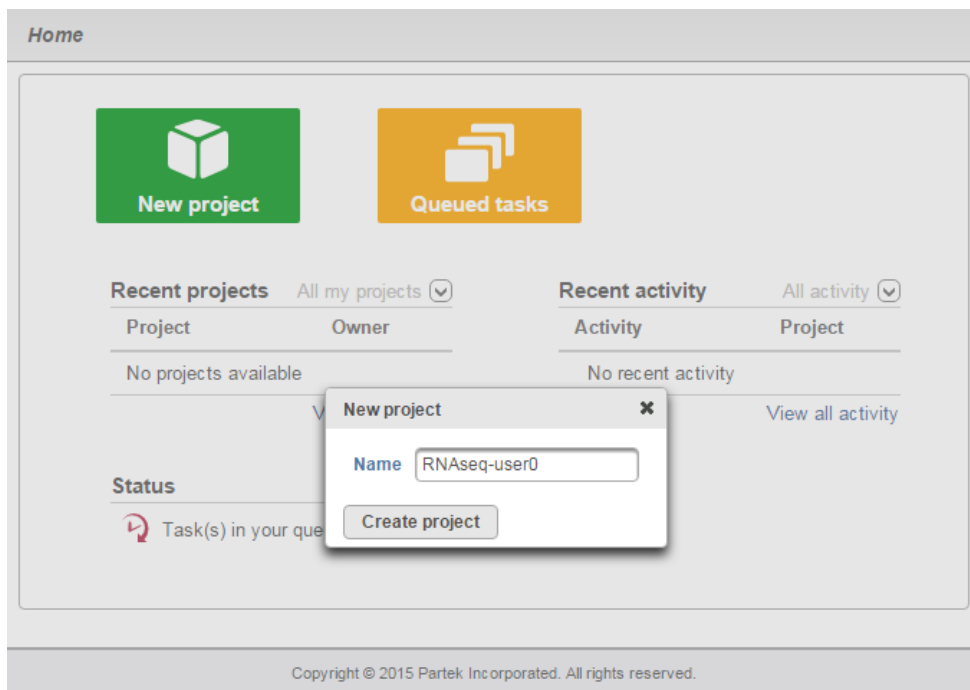
CCR, NCI
April 5th, 2016



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

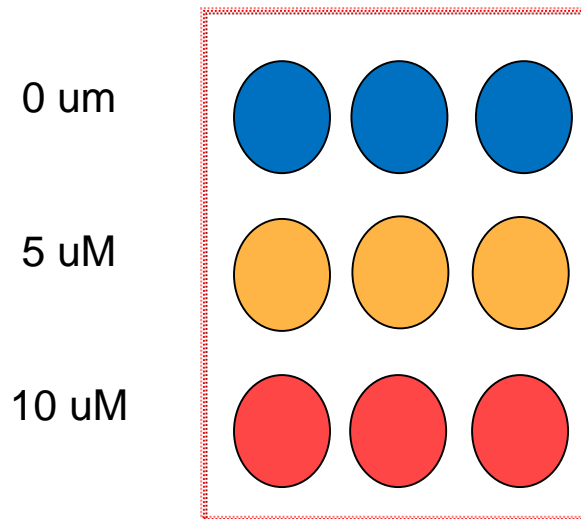


Notes: _____

Training Dataset

Data files in the project:

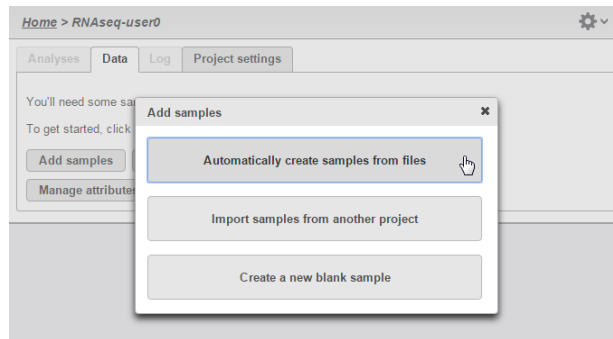
- HT29 colon cancer cells exposed to 5-aza drug with 3 different doses
 - 0 uM (Control)
 - 5 uM
 - 10 uM
- mRNA purified and sequenced using Illumina HiSeq
- Paired end reads
- Xu et al. 2013 BMC Bioinformatics (PMID: 23902433)



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)



Home > RNAseq-user0 > Data > Create samples from files

Select files from Partek Flow server My computer URL

Select files

Current directory
/home/flow/FlowData/RNA-seq

Server Computer

- home
 - flow
 - demo_db
 - FlowData
 - DNA-seq
 - library_files
 - ngs-training-files
 - Output
 - RNA-seq

18 files selected

<input type="checkbox"/>	Name	Size
<input type="checkbox"/>	sampleInfo.txt	161 B
<input checked="" type="checkbox"/>	SRR592573_1.fastq.gz	10.48 MB
<input checked="" type="checkbox"/>	SRR592573_2.fastq.gz	10.6 MB
<input checked="" type="checkbox"/>	SRR592574_1.fastq.gz	15.64 MB
<input checked="" type="checkbox"/>	SRR592574_2.fastq.gz	15.8 MB
<input checked="" type="checkbox"/>	SRR592575_1.fastq.gz	21.8 MB
<input checked="" type="checkbox"/>	SRR592575_2.fastq.gz	22.01 MB
<input checked="" type="checkbox"/>	SRR592576_1.fastq.gz	25.1 MB
<input checked="" type="checkbox"/>	SRR592576_2.fastq.gz	25.63 MB

Valid file types are bcf, idat, sff, txt, bz2, vcf, fasta, bgx, gz, cfasta, bpm, raw, fastq, qual, zip, sra, bam, sam, probe_tab, CEL, tar or cfastq files...

Copyright © 2015 Partek, Inc. All rights reserved.

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

The screenshot shows a web interface for RNAseq data management. At the top, there is a breadcrumb 'Home > RNAseq-user0' and a settings icon. Below this are tabs for 'Analyses', 'Data', 'Log', and 'Project settings'. The main content area features a table with sample information:

	Sample name	Attributes
		Treatment
1	SRR592573	0uM
2	SRR592574	0uM
3	SRR592575	0uM
4	SRR592576	5uM
5	SRR592577	5uM
6	SRR592578	5uM
7	SRR592579	10uM
8	SRR592580	10uM
9	SRR592581	10uM

Below the table are buttons for 'Show data files' and 'Download'. Further down, there are several action buttons: 'Add samples', 'Assign sample attributes from a file' (highlighted with a mouse cursor), 'Add a system-wide attribute column', and 'Manage attribute' (with a tooltip that reads 'Apply attributes by importing a file with information about your samples'). At the bottom, the project output directory is shown as '/home/flow/FlowData/Output/Project_RNAseq-user0 (254.18 GB free)'. A copyright notice at the very bottom reads 'Copyright © 2015 Partek Incorporated. All rights reserved.'

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

The screenshot shows the 'Analyses' tab in a software interface. At the top, there is a breadcrumb 'Home > RNAseq-user0' and a settings gear icon. Below this are tabs for 'Analyses', 'Data', 'Log', and 'Project settings'. The main workspace contains two nodes: a circular 'Unaligned reads' data node and a rectangular 'Pre-alignment QA/QC' task node. A tooltip points to the task node with the text: 'Assess the quality of raw reads to decide whether trimming or filtering is necessary before alignment.' On the right, a 'Context sensitive menu' is open, listing options: 'Unaligned reads', 'QA/QC', 'Pre-alignment QA/QC' (highlighted), 'ERCC (Bowtie)', 'Pre-alignment tools', and 'Aligners'. At the bottom of the menu is a 'Download data (352 MB)' button. At the bottom of the interface, there are links for 'Make a pipeline' and 'Import a pipeline', and a 'Project disk space' indicator with a green bar.

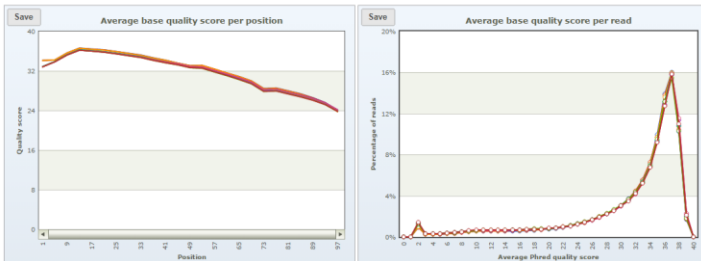
Notes: _____

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

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

Sample name	Total reads	Read length	Avg. read quality	% N	% GC
SRR592573	232,700	98.00	32.10	0%	53.74%
SRR592574	347,698	98.00	32.07	0%	53.61%
SRR592575	484,720	98.00	32.04	0%	53.25%
SRR592576	562,736	98.00	31.80	0%	52.95%
SRR592577	503,142	98.00	31.78	01%	52.02%
SRR592578	587,508	98.00	31.77	0%	52.89%
SRR592579	283,848	98.00	31.79	0%	51.96%
SRR592580	478,754	98.00	31.59	01%	53.06%
SRR592581	413,422	98.00	31.59	0%	51.98%



Home > RNAseq-user0 > Pre-alignment QA/QC report > SRR592573


Sample name SRR592573 Total reads 232,700 Read length 98.00
 Median read length 98.00 Average read quality 32.10 Quality score format Phred+33



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

Home > RNAseq-user0 > Trim bases

Trim based on


- Quality score 
- From 3' end 
- From 5' end 
- Both ends 


Quality trimming


End min quality level (Phred)

Trim from end

Advanced options

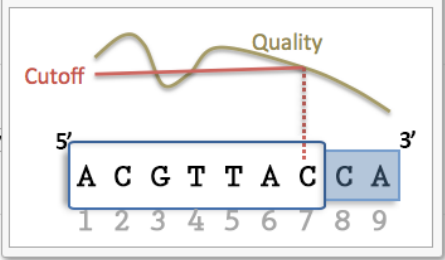
Min read length 

Max N  %

Quality encoding 

Copyright © 2015 Partek Incorporated. All rights reserved.

This mode scans the read from the 5' or 3' end (or both) for the first base at or above the specified Phred quality score. All bases previous to this position are trimmed (from the left if the 5' end, from the right if the 3' end). This mode is not available if none of your files contain quality scores.



The diagram illustrates the quality trimming process. A green curve labeled 'Quality' represents the Phred quality score across a sequence of bases. A horizontal red line labeled 'Cutoff' is drawn at a quality level of 20. A vertical dashed red line indicates the first base that meets or exceeds this cutoff, which is the 7th base (C). The bases to the right of this position (C and A) are highlighted in a blue box, indicating they are retained. The bases to the left (A, C, G, T, T, A) are not highlighted, indicating they are trimmed. The sequence is labeled with 5' and 3' ends, and the bases are numbered 1 through 9.

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**

The screenshot shows the Partek Flow web interface for configuring the STAR aligner. At the top, there is a navigation bar with the Partek Flow logo, a 'Queue' dropdown, a 'Projects' dropdown, and a user profile for 'admin'. Below the navigation bar is a breadcrumb trail: 'Home > RNAseq-user0 > STAR'. The main content area is titled 'Select STAR 2.4.1d index' and contains two dropdown menus: 'Genome build' set to 'Homo sapiens (human) - hg19_chr22' and 'Index' set to 'Whole genome'. Below this is the 'Alignment options' section, which includes a 'Generate unaligned reads' checkbox (unchecked) with an information icon. The 'Advanced options' section features an 'Option set' dropdown menu set to '-- Default --' and a 'Configure' link. At the bottom of the configuration area are two buttons: 'Back' and 'Finish'. A footer bar at the very bottom contains the copyright notice: 'Copyright © 2015 Partek Incorporated. All rights reserved.'

Notes:

Post-alignment QA/QC

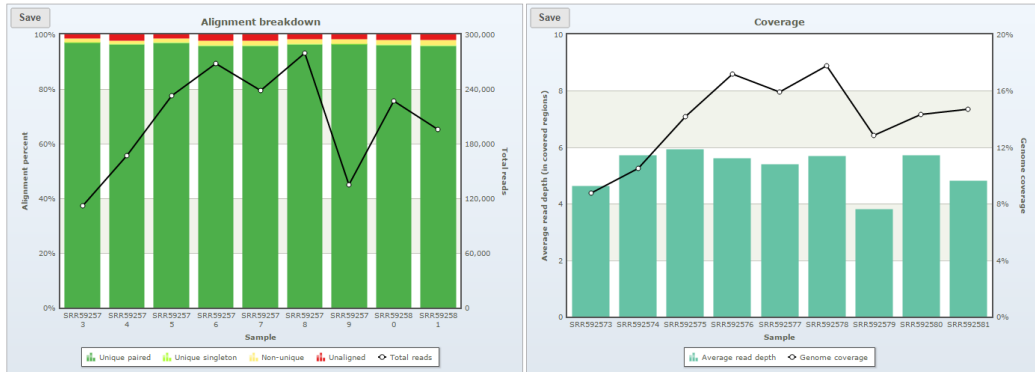
- Perform Post-alignment QA/QC to assess the quality of the alignment task
- Select **Aligned reads** data node
- Click **Post-alignment QA/QC** from the **QA/QC** section of the menu
- Use default settings and click **Finish**
- Click on a sample name to get QA/QC results for that sample name

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.



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**

Home > RNAseq-user0 > Quantify to transcriptome (Partek E/M)

Select annotation file

Genome build *Homo sapiens (human) - hg19_chr22*

Annotation model

Quantification options

Strict paired-end compatibility

Require junction reads to match introns

Strand specificity

Minimum read overlap with feature

Percent of read length

Number of bases

Report unexplained regions

Min reads for unexplained region

Include BAM files in output project file

Notes: _____

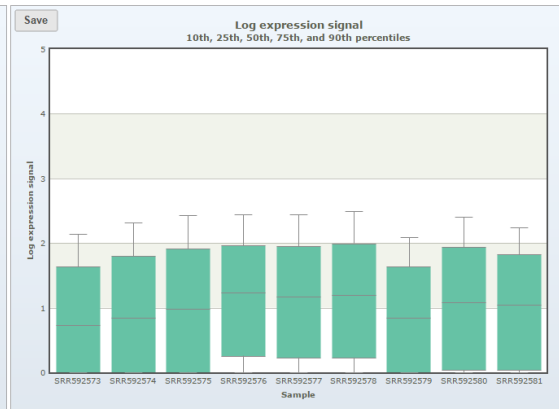
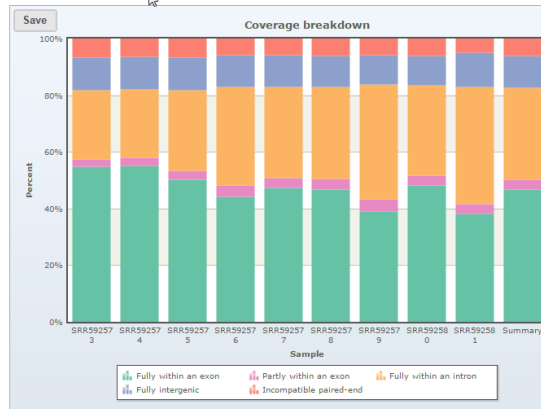
Viewing Quantification results

- 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

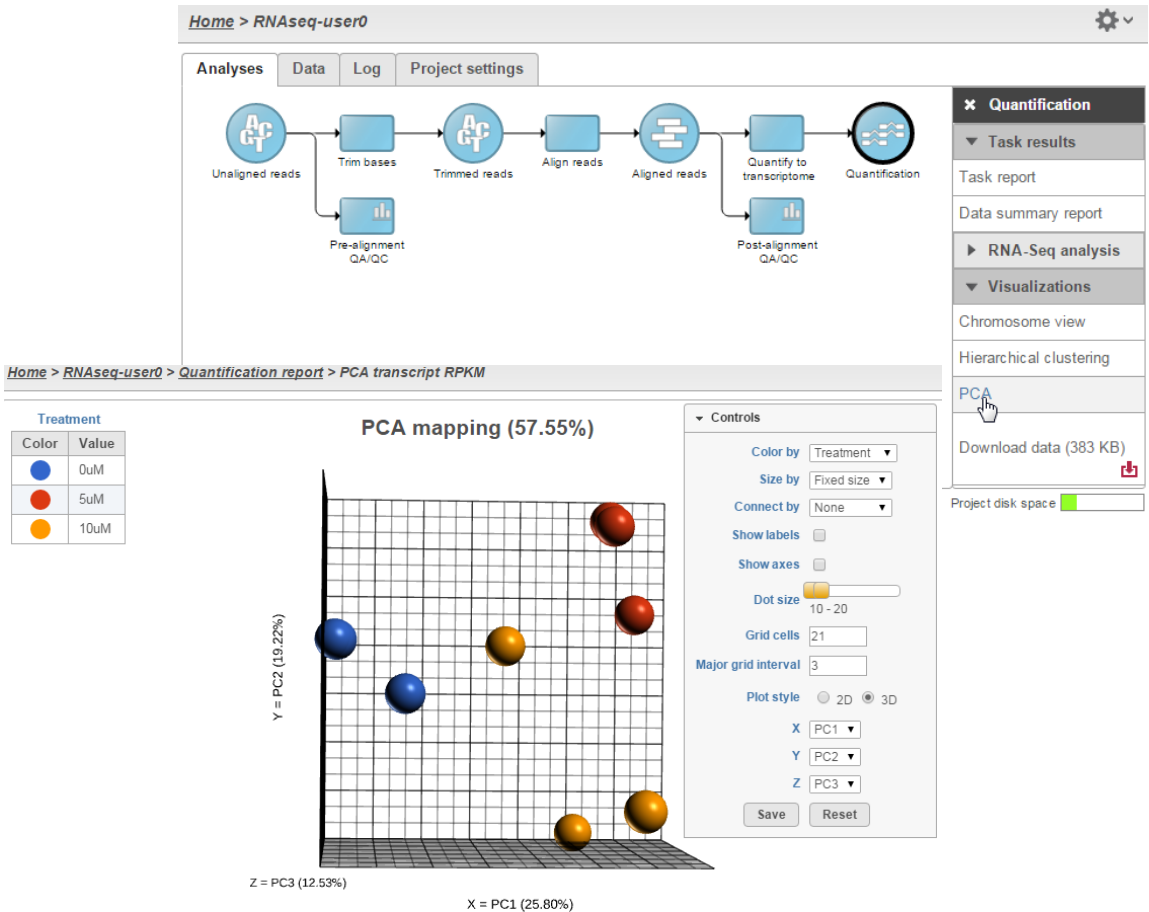
These samples were quantified to RefSeq
[Download gene-level read counts](#) | [Download transcript-level read counts](#)



Notes:

Principal Components Analysis

- The principal components analysis (PCA) scatter plot allows you to assess relatedness between samples and identify outliers
- This can only be 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**

The screenshot shows the RNA-Seq analysis workflow and the configuration for Differential Gene Expression Analysis (GSA).

Workflow Diagram: Unaligned reads → Trim bases → Trimmed reads → Align reads → Aligned reads → Quantify to transcriptome → Quantification. A side path shows Pre-alignment QA/QC and Post-alignment QA/QC.

Navigation: Home > RNAseq-user0 > Analyses > Data > Log > Project settings

Quantification Panel:

- Task results
- RNA-Seq analysis
 - Differential gene expression (GSA) (selected)
 - Differential gene expression (ANOVA)
- Visualizations
- Download data (383 KB)
- Project disk space

Comparison selector:

- Treatment 10uM: 0uM, 5uM, 10uM
- vs.
- Treatment 0uM: 0uM, 5uM, 10uM

Add comparison:

	Treatment	vs.	Treatment	
1	5uM	vs.	0uM	✘
2	10uM	vs.	0uM	✘

Analysis level: Run analysis on Gene level Transcript level

Advanced options: Option set -- Default -- [Configure](#)

[Back](#) [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”

Gene list

Results: 22

Filter

Gene ID

Total reads

P-value

FDR step up

All contrasts
 Per contrast

Less than or eq.

0 1

Ratio

Fold change

All contrasts
 Per contrast

from to

Exclude range

Low Expressed

Save filter Clear filter

Saved filters

(No saved filters available)

Optional columns




	View	Gene ID	Total reads	P-value	FDR step up	5uM vs 0uM			
						Ratio	Fold change	LSMean(5uM)	LSMean(0uM)
1		ADRBK2	2,127.00	7.12E-5	1.14E-3	0.31	-3.18	729.93	2,320.72
2		APOL3	88.00	1.21E-4	1.79E-3	11.39	11.39	74.42	6.53
3		CDC42EP1	5,195.00	6.72E-5	1.13E-3	0.36	-2.81	1,883.89	5,294.25
4		CDC45	442.00	1.38E-3	9.64E-3	5.14	5.14	341.87	66.49
5		CENPM	117.00	9.82E-4	7.88E-3	4.01	4.01	93.86	23.38
6		DERL3	190.00	3.35E-5	6.9E-4	7.29	7.29	190.37	26.13
7		EIF3L	4,421.00	4.2E-4	4.09E-3	0.47	-2.11	1,871.87	3,945.52
8		EMID1	20.00	3.88E-6	2.3E-4	168,331.72	168,331.72	16.83	1E-4
9		GCAT	435.00	1.45E-3	9.64E-3	2.91	2.91	291.37	100.15
10		GGT1	969.18	4.32E-5	7.86E-4	2.59	2.59	691.08	266.95
11		H1F0	19,473.00	2.02E-4	2.52E-3	0.42	-2.36	8,602.78	20,274.32
12		KLHDC7B	7,421.00	5.14E-7	1.18E-4	13.27	13.27	6,736.63	507.77
13		MCM5	616.00	9.34E-4	7.66E-3	2.08	2.08	320.89	154.32
14		PRODH	271.00	7.56E-6	2.54E-4	5.47	5.47	220.90	40.42
15		RPL3	8,308.95	2.44E-4	2.81E-3	0.36	-2.80	3,185.03	8,933.04
16		SELM	392.00	1.25E-6	1.18E-4	51.38	51.38	410.53	7.99
17		SEPT5	52.76	1.17E-6	1.18E-4	500,258.08	500,258.08	50.03	1E-4
18		SHANK3	135.00	3.96E-4	4.07E-3	10.40	10.40	129.77	12.48
19		SUSD2	228.00	2.88E-3	0.02	4.45	4.45	165.73	37.21
20		TBX1	91.00	1.16E-3	8.72E-3	3.04	3.04	70.31	23.10
21		TIMP3	402.00	6.34E-8	3E-5	4,840,729.20	4,840,729.20	484.07	1E-4
22		TUBA8	184.00	5.33E-3	0.03	2.75	2.75	124.57	45.26

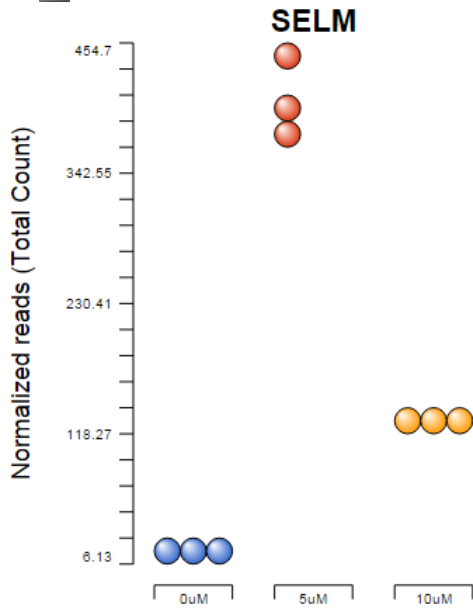
Rows per page (1 of 1)

[Generate 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



Legend

- 0uM
- 5uM
- 10uM

[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		

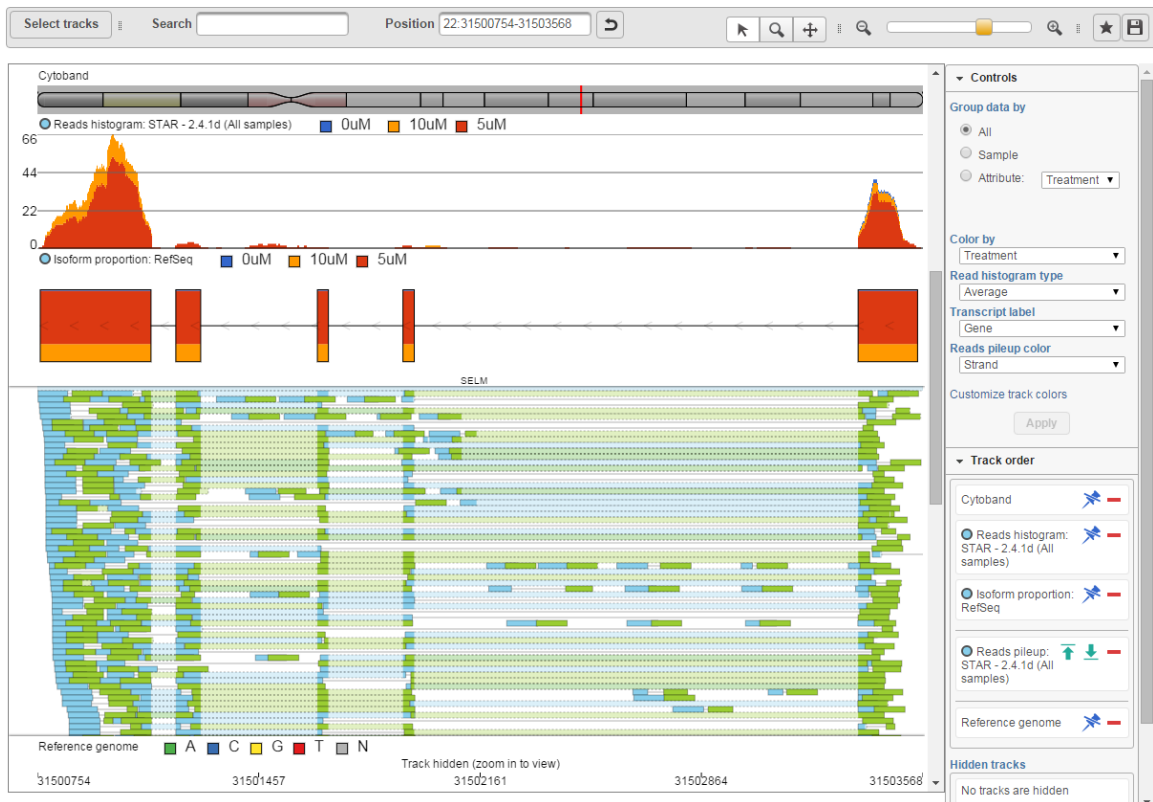
Model 1

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

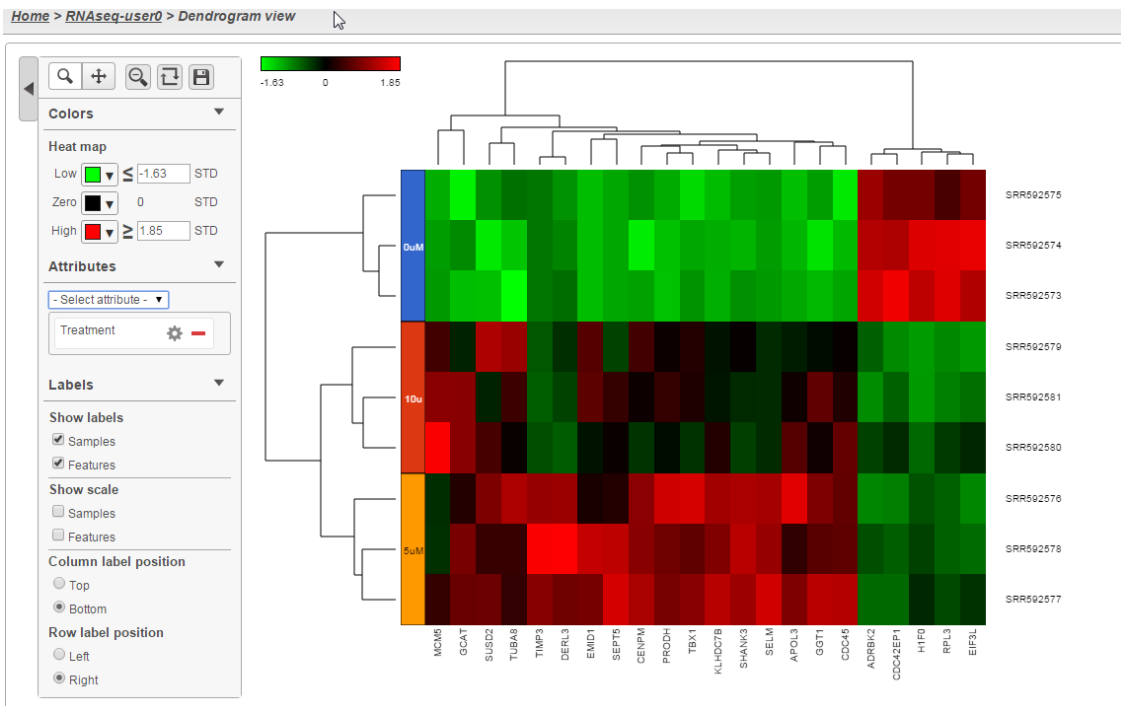
- **Select tracks** allows you to select different annotations or datasets to view together
- Sample grouping, color and transcript labeling can be edited in the **Controls** panel
- Search for any gene using the **Search** box
- Navigate to a genomic coordinate using the **Position** box
- Change and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

AmiGO 2 Home Search Tools & Resources Help Feedback About AmiGO 1.8 Quick search Search

alpha-amino acid metabolic process

Term information

Accession GO:1901605
 Name alpha-amino acid metabolic process
 Ontology biological_process
 Synonyms alpha-amino acid metabolism
 Definition The chemical reactions and pathways involving an alpha-amino acid. Source: GOC:TermGene
 Comment None
 History See term history for GO:1901605 at QuickGO
 Subset None
 Community (0) Add badge comments for this term on the CC0/3 wiki
 Related (0) Link to all genes and gene products annotated to alpha-amino acid metabolic process.
 (0) Link to all direct and indirect annotations to alpha-amino acid metabolic process.
 Feedback Contact the GO Helpdesk if you find mistakes or have concerns about the data you find here.

Annotations Graph Views Inferred Tree View Accessions and Children Mappings

Free-text filtering X Found entities Total: 27631, showing 1-10 Results count 10

Your search is pinned to these filters

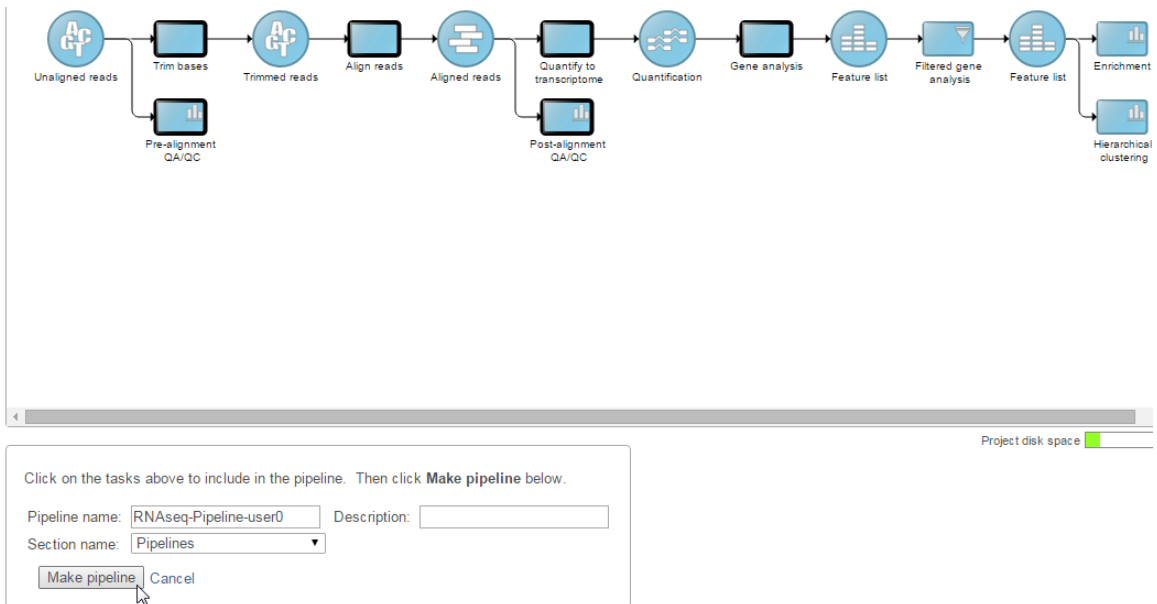
- document_category: annotation
- regulates_closure: GO:1901605
- No current user filters

Gene product	Gene product name	Qualifier	Direct annotation	Annotation extension	Assigned by	Taxon	Evidence	Evidence with	PANTHER family
0010_02731			gene biosynthetic process via aminoacyl: acid	GO_Central	34-00000	IBA	PANTHER:PTN000205224	phosphoserine transferase pf012215	
pr0-4			protein biosynthetic process	GO_Central	Neurospora	IBA	PANTHER:PTN000115402	glutamate semialdehyde dehydrogenase pf011063	
F70177	Adenosylhomocysteinase	S-	adenosylmethionine	GO_Central	Macaca mulatta	IBA	PANTHER:PTN000002719	adenosylhomocysteinase pf07	

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: _____

