

RNA-Seq Analysis in Partek[®] Flow[®] and Genomics Suite[®]

HANDS-ON TRAINING

NCI BTEP Workshop

March 21st, 2017



Eric Seiser, PhD

Field Application Scientist

Partek Incorporated

support@partek.com

Training Dataset

Proc Natl Acad Sci U S A. 2016 Oct 11;113(41):E6107-E6116. Epub 2016 Sep 28.

Repression of p63 and induction of EMT by mutant Ras in mammary epithelial cells.

Yoh KE¹, Regunath K¹, Guzman A², Lee SM³, Pfister NT¹, Akanni O¹, Kaufman LJ², Prives C⁴, Prywes R⁴.

Author information

Abstract

The p53-related transcription factor p63 is required for maintenance of epithelial cell differentiation. We found that activated forms of the Harvey Rat Sarcoma Virus GTPase (H-RAS) and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA) oncogenes strongly repress expression of Δ Np63 α , the predominant p63 isoform in basal mammary epithelial cells. This regulation occurs at the transcriptional level, and a short region of the Δ Np63 promoter is sufficient for repression induced by H-RasV12. The suppression of Δ Np63 α expression by these oncogenes concomitantly leads to an epithelial-to-mesenchymal transition (EMT). In addition, the depletion of Δ Np63 α alone is sufficient to induce EMT. Both H-RasV12 expression and Δ Np63 α depletion induce individual cell invasion in a 3D collagen gel in vitro system, thereby demonstrating how Ras can drive the mammary epithelial cell state toward greater invasive ability. Together, these results suggest a pathway by which RAS and PIK3CA oncogenes induce EMT through regulation of Δ Np63 α .

KEYWORDS: H-Ras; breast cancer; epithelial mesenchymal transition; p63; transcriptional repression

PMID: 27681615 PMCID: PMC5068336 [Available on 2017-04-11] DOI: 10.1073/pnas.1613417113

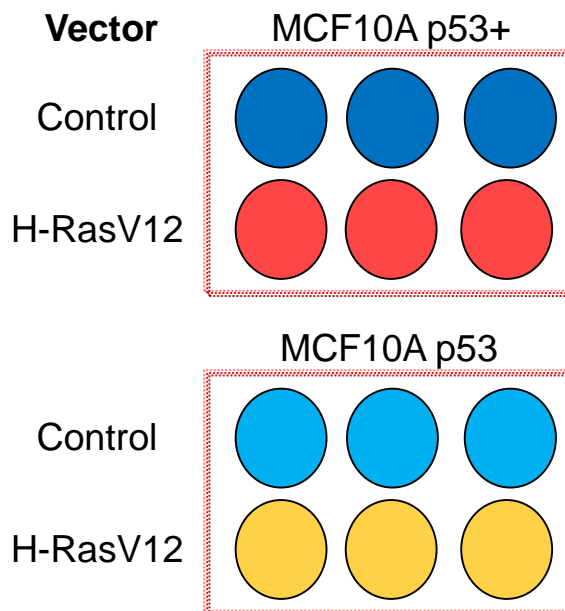
Data files in the project:

- MCF10A mammary epithelial cells, with or without p53 deletion, were transfected with lentiviral vectors:

-Control vector

-H-RasV12 vector

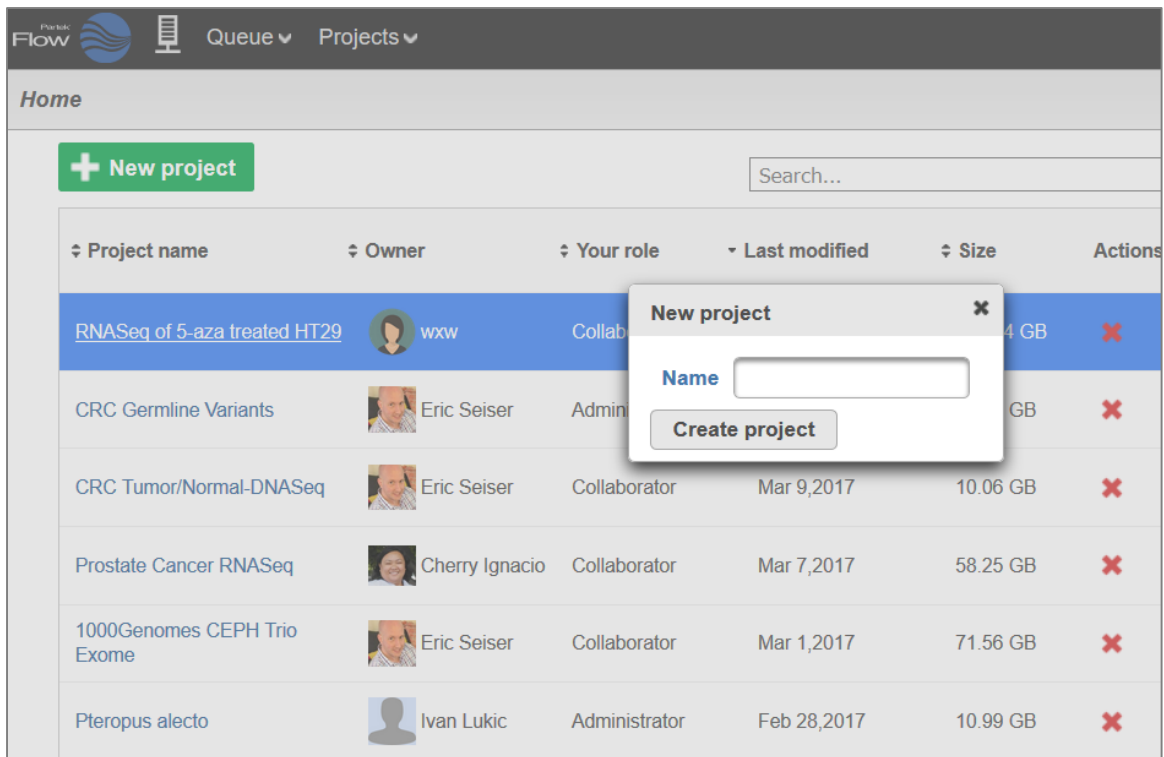
- mRNA purified and sequenced (paired-end reads) using the Illumina HiSeq 2000



Notes: _____

Login and Project Set-up

- Open Google Chrome/Firefox and enter: NCIdemo.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



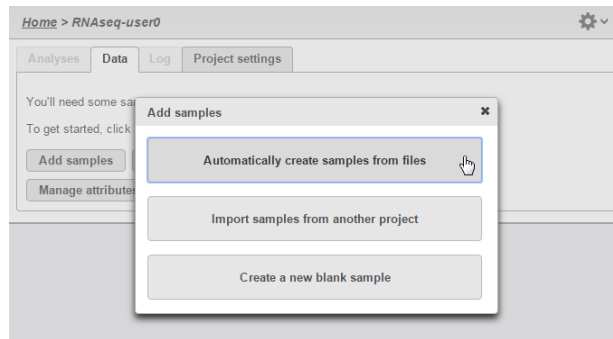
The screenshot shows the Partek Flow web interface. At the top, there is a navigation bar with the Partek Flow logo, a 'Queue' dropdown, and a 'Projects' dropdown. Below the navigation bar, the page title is 'Home'. A green button labeled '+ New project' is visible on the left. A search bar is located on the right. The main content area displays a table of projects with columns for Project name, Owner, Your role, Last modified, Size, and Actions. A modal dialog box titled 'New project' is open, featuring a text input field for 'Name' and a 'Create project' button. The table contains the following data:

Project name	Owner	Your role	Last modified	Size	Actions
RNASeq of 5-aza treated HT29	wxw	Collab		4 GB	✖
CRC Germline Variants	Eric Seiser	Admin		GB	✖
CRC Tumor/Normal-DNASeq	Eric Seiser	Collaborator	Mar 9,2017	10.06 GB	✖
Prostate Cancer RNASeq	Cherry Ignacio	Collaborator	Mar 7,2017	58.25 GB	✖
1000Genomes CEPH Trio Exome	Eric Seiser	Collaborator	Mar 1,2017	71.56 GB	✖
Pteropus alecto	Ivan Lukic	Administrator	Feb 28,2017	10.99 GB	✖

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-seq1
- Select all 24 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, csfasta, bpm, raw, fastq, qual, zip, sra, bam, sam, probe_tab, CEL, tar or csfastq 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

Analyses
Data
Log
Project settings
Attachments

	Sample name	Attributes	Files	
		Treatment	fastq	+/-
1	SRR3541289	p53- Control	2 files: SRR3541289_1, SRR3541289_2	
2	SRR3541290	p53- Control	2 files: SRR3541290_1, SRR3541290_2	
3	SRR3541291	p53- Control	2 files: SRR3541291_1, SRR3541291_2	
4	SRR3541292	p53- H-RasV12	2 files: SRR3541292_1, SRR3541292_2	
5	SRR3541293	p53- H-RasV12	2 files: SRR3541293_1, SRR3541293_2	
6	SRR3541294	p53- H-RasV12	2 files: SRR3541294_1, SRR3541294_2	
7	SRR3541295	p53+ Control	2 files: SRR3541295_1, SRR3541295_2	
8	SRR3541296	p53+ Control	2 files: SRR3541296_1, SRR3541296_2	
9	SRR3541297	p53+ Control	2 files: SRR3541297_1, SRR3541297_2	
10	SRR3541298	p53+ H-RasV12	2 files: SRR3541298_1, SRR3541298_2	
11	SRR3541299	p53+ H-RasV12	2 files: SRR3541299_1, SRR3541299_2	
12	SRR3541300	p53+ H-RasV12	2 files: SRR3541300_1, SRR3541300_2	

Hide data files
Download

Import data
Edit attributes
Assign sample attributes from a file

Add a system-wide attribute column

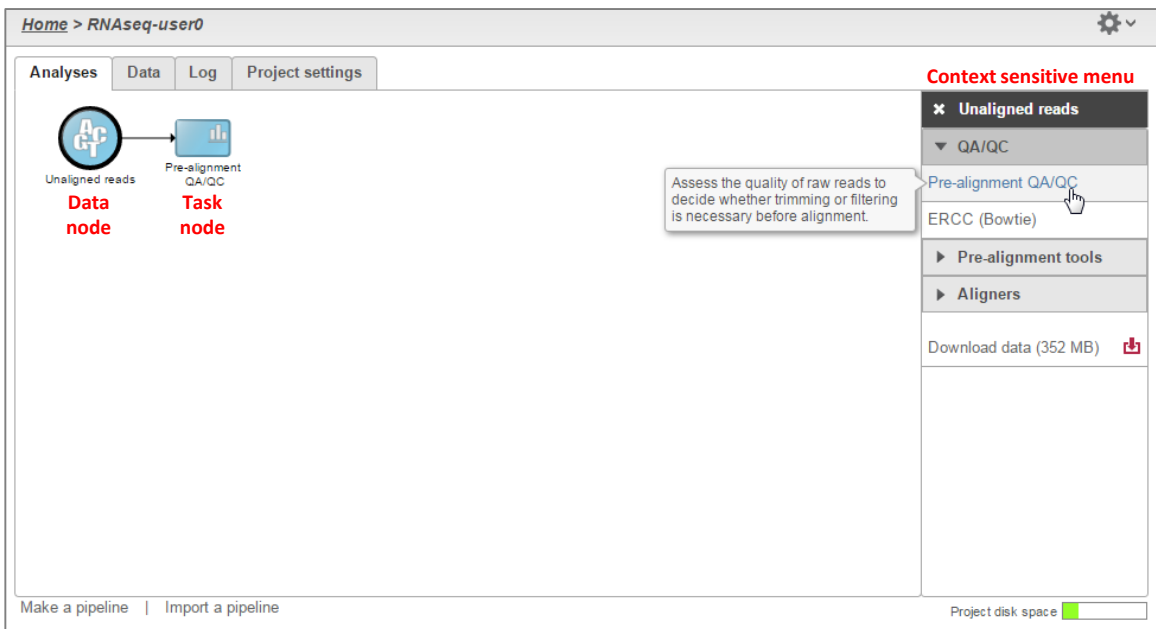
Apply attributes by importing a file with information about your samples
[manage attributes](#)

Project output directory /home/flow/FlowData/Project_NCI training (8.97 TB free)

Notes:

Analyses Tab Overview

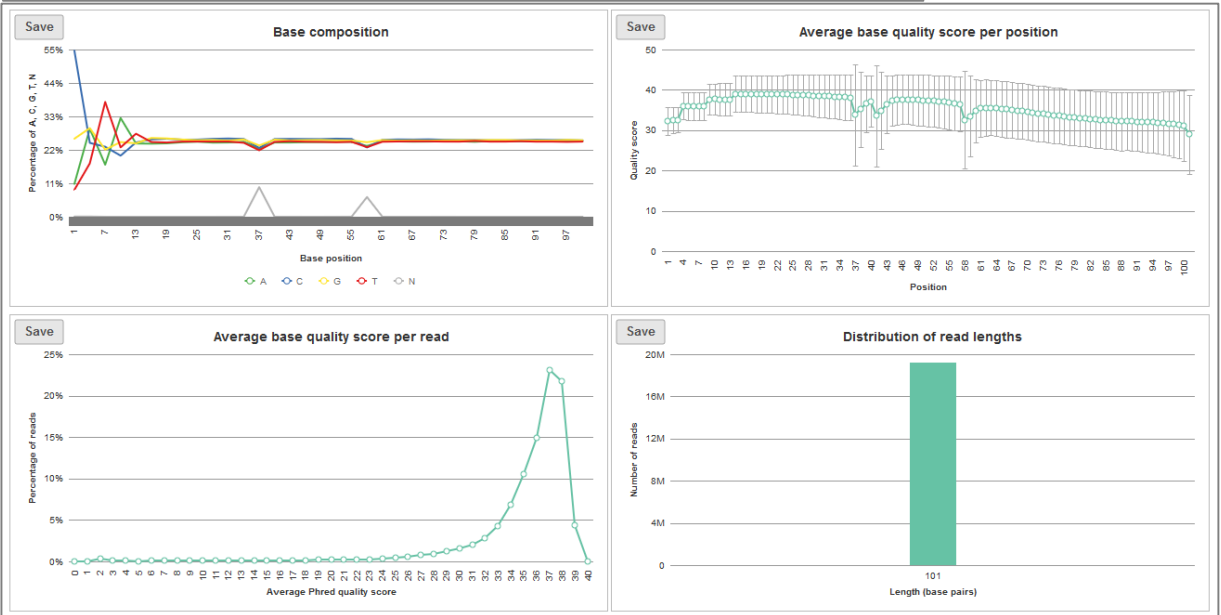
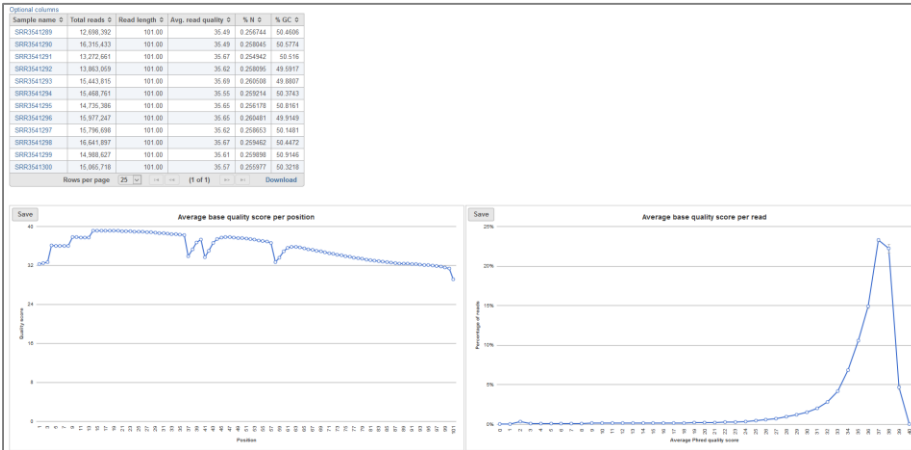
- Go to the **Analyses** tab
- Your first data node, the **Unaligned reads** node appears. All data nodes are circles
- Select the **Unaligned reads** data node and select **Pre-alignment QA/QC**
- Use the default settings and click **Finish**
- This will create a new task node in the **Analyses** tab.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

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.

Tooltip: 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.

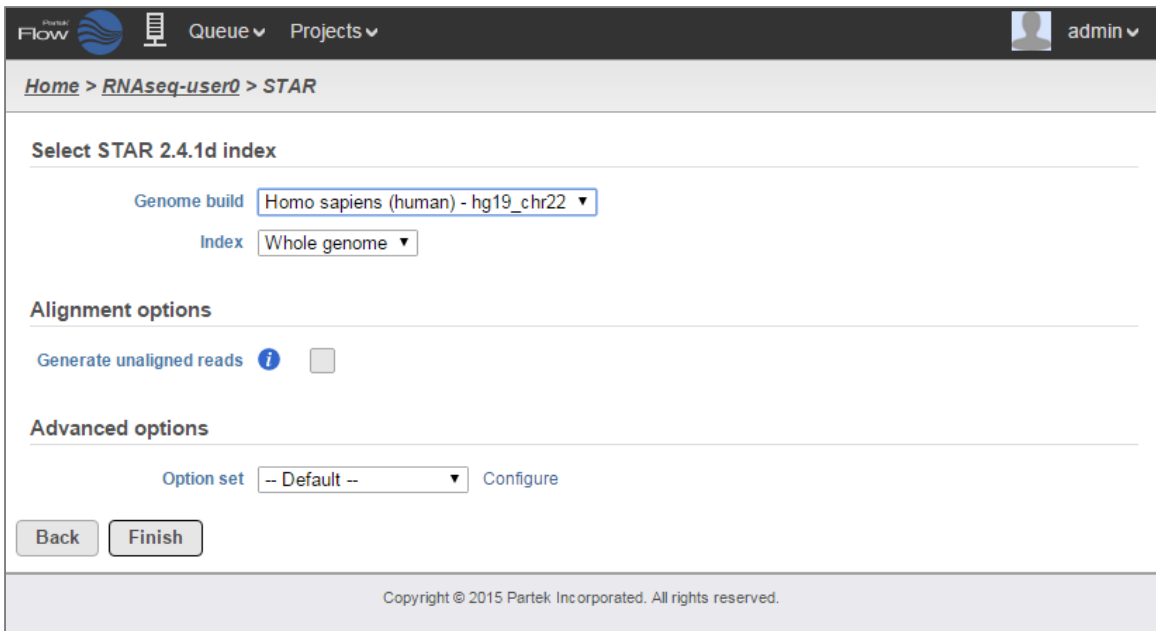
Diagram: Quality curve above read sequence: 5' A C G T T A C C A 3'. Cutoff line at quality 20. Trim point at base 7 (C).



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**



Flow  Queue ▾ Projects ▾  admin ▾


Home > RNAseq-user0 > STAR

Select STAR 2.4.1d index

Genome build

Index

Alignment options

Generate unaligned reads 

Advanced options

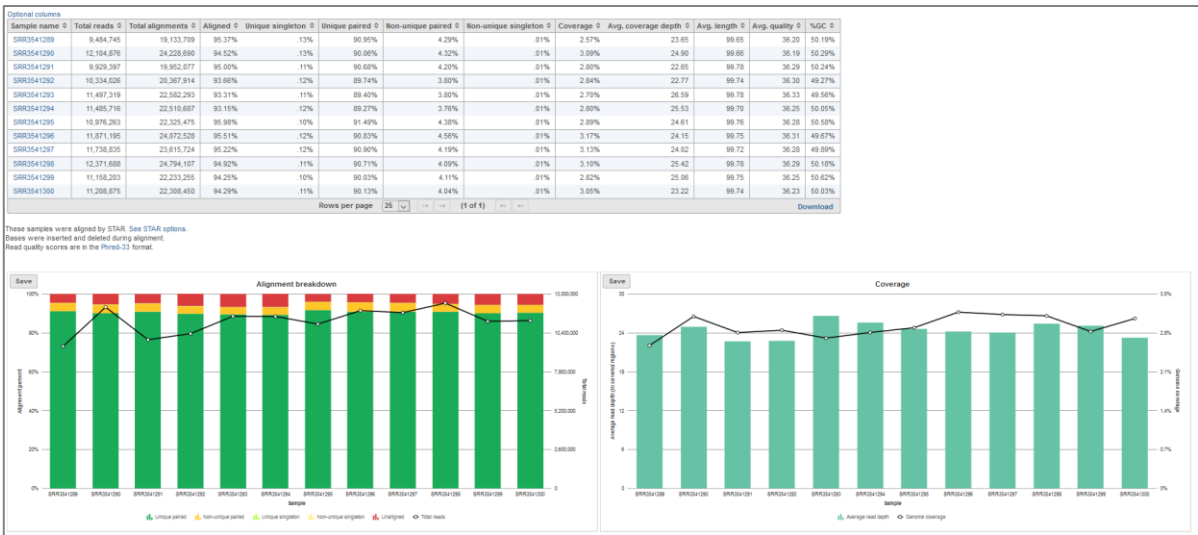
Option set [Configure](#)

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



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

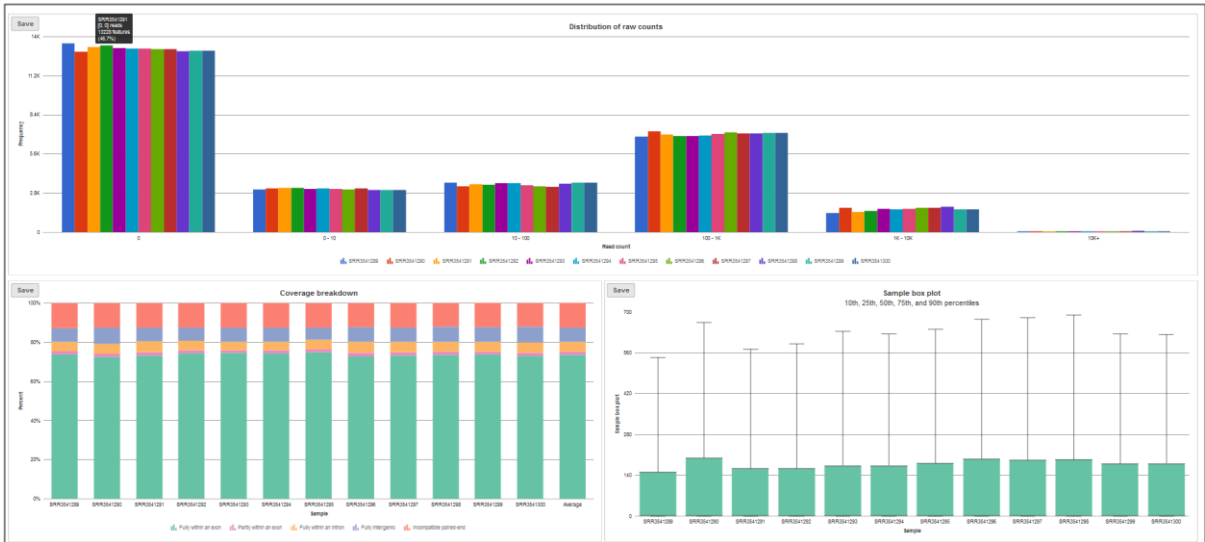
Notes: _____

Viewing Quantification results

- To view the results of the quantification, select the **Gene counts** data node
- Click **Task report** on the menu
- Click **Download** in the context sensitive menu to download a .txt file of the read counts

Sample name	Total reads	Fully within an exon	Partly within an exon	Fully within an intron	Fully intergenic	Incompatible paired-end	Compatible junctions	Total junctions	View
SRR3541289	9,055,372.00	73.72%	1.62%	4.77%	6.90%	12.99%	3,598,952.00	4,466,521.00	
SRR3541290	11,456,351.00	72.33%	1.62%	5.13%	8.35%	12.57%	4,387,016.00	5,440,403.00	
SRR3541291	9,443,178.00	73.18%	1.64%	5.46%	7.03%	12.70%	3,630,548.00	4,542,733.00	
SRR3541292	9,690,019.00	74.12%	1.57%	4.89%	6.59%	12.83%	3,784,962.00	4,691,222.00	
SRR3541293	10,741,194.00	74.41%	1.49%	4.22%	7.26%	12.62%	4,181,182.00	5,168,187.00	
SRR3541294	10,712,578.00	74.15%	1.53%	4.45%	7.13%	12.74%	4,200,474.00	5,195,114.00	
SRR3541295	10,547,744.00	74.64%	1.62%	4.92%	6.06%	12.76%	4,142,194.00	5,128,654.00	
SRR3541296	11,353,730.00	72.72%	1.65%	5.61%	7.53%	12.49%	4,326,763.00	5,373,354.00	
SRR3541297	11,191,984.00	73.17%	1.63%	5.23%	7.23%	12.73%	4,353,389.00	5,391,138.00	
SRR3541298	11,759,093.00	73.41%	1.55%	5.17%	7.63%	12.24%	4,460,098.00	5,543,317.00	
SRR3541299	10,529,292.00	73.60%	1.52%	5.04%	7.44%	12.41%	3,924,308.00	4,927,868.00	
SRR3541300	10,582,254.00	72.93%	1.57%	5.28%	7.93%	12.29%	3,983,667.00	4,955,208.00	
Average	10,588,565.75	73.51%	1.58%	5.02%	7.28%	12.60%	4,081,129.42	5,068,643.25	

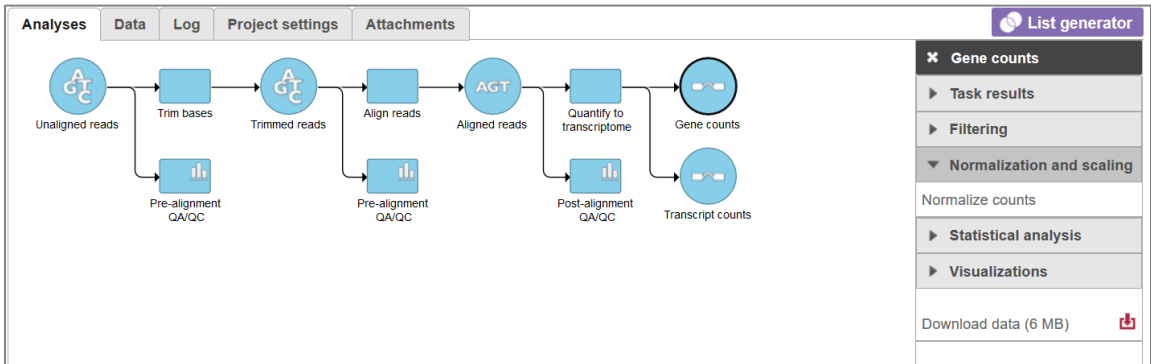
Rows per page: 25 (1 of 1) [Download](#)



Notes:

Normalization

- To adjust for technical variation in the data, click on the **Gene Counts** data node and go to Normalize counts under the Normalization and scaling tab
- In the dialog, click **Recommended** and then click **Finish**
- This will create a **Normalized counts** data node



Read count normalization

Transform on Samples Features

Normalization methods

- Absolute value
- Add
- Antilog
- Divide by
- FPKM
- Log
- Logit
- Lower bound
- Multiply by
- Quantile normalization
- Rank
- Subtract
- TMM
- TPM
- Total count

Drag and drop →

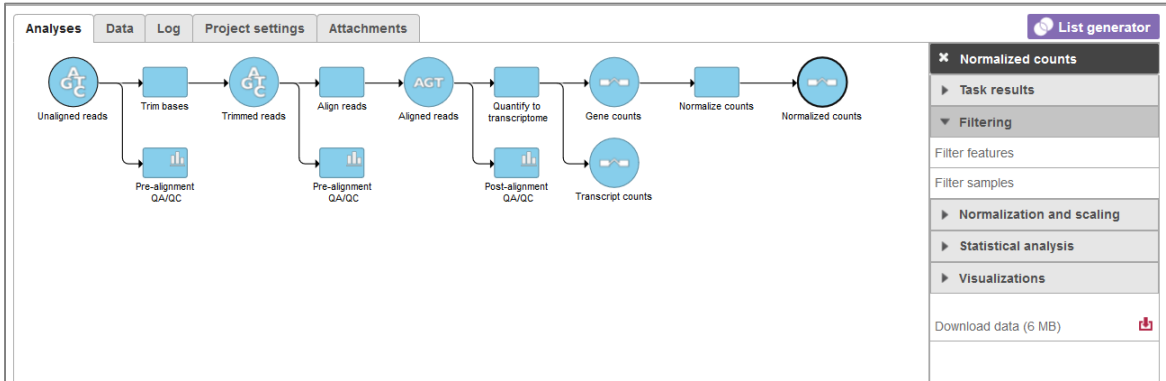
Normalization order Recommended

1. Total count
2. Add

Notes: _____

Filtering Low Expression

- To remove genes that have low expression across samples, click on the **Normalized counts** data node and select **Filter features** under the Filtering tab in the menu
- In the dialog, ensure that *Filter out features where* is set to **geometric mean** and **<=1.0**
- Clicking Finish will create a **Filtered counts** data node



Filter out features where

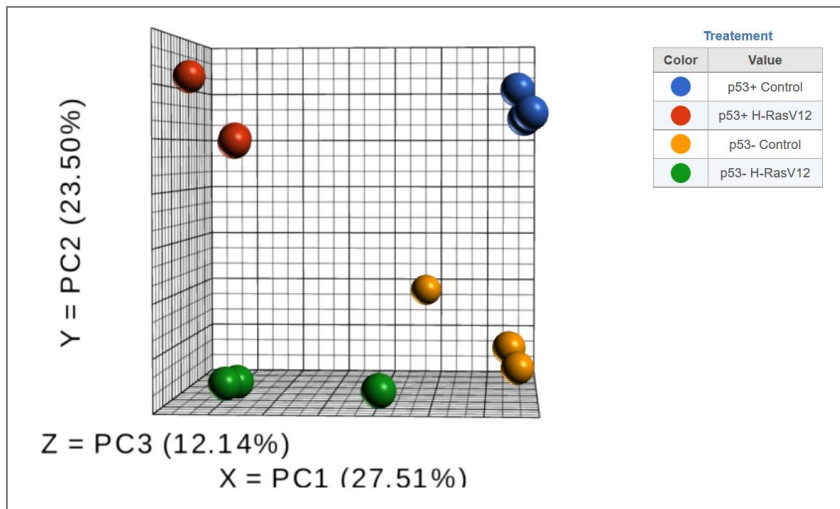
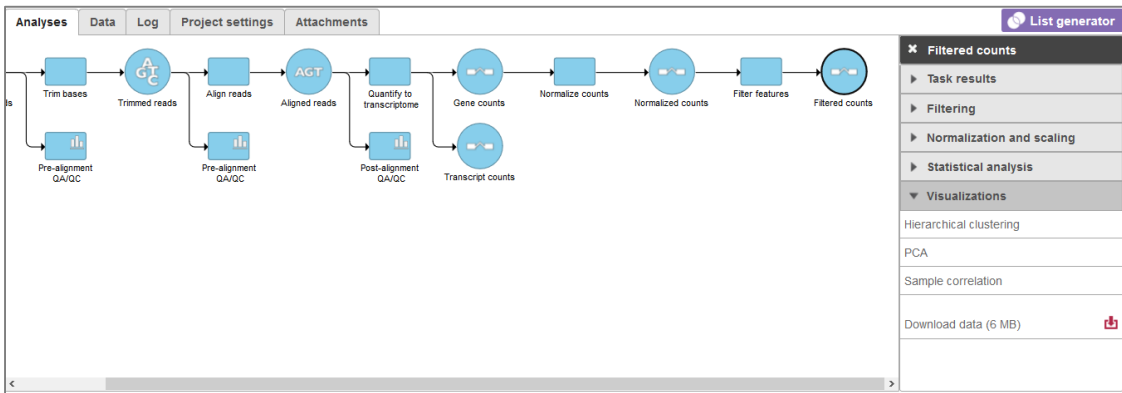
The **geometric mean** is **<=**

Expression value is **<** in at least % of the samples

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 **Filtered counts** 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 **Statistical analysis** section of the menu
- Select **Treatment** as the categorical attribute
- Select **p53+ H-RasV12 vs p53+ Control** and click **Add comparison**
- Select **p53- H-RasV12 vs p53- Control** and click **Add comparison**
- Select **p53- H-RasV12 vs p53+ H-RasV12**, click **Add comparison** and click **Finish**

Identify differentially expressed features with Partek Gene Specific Analysis algorithm by applying multiple statistical models to each individual gene in order to account for each gene's varying response to different experimental factors, and differing data distributions.

Treatment

p53+ Control

p53+ H-RasV12

p53- Control

p53- H-RasV12

p53- H-RasV12

vs.

Treatment

p53+ Control

p53+ H-RasV12

p53- Control

p53- H-RasV12

p53+ H-RasV12

Add comparison

	Treatment	vs.	Treatment	
1	p53+ H-RasV12	vs.	p53+ Control	✘
2	p53- H-RasV12	vs.	p53- Control	✘

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 **Per contrast** for **p53+ H-RasV12 vs p53+ Control** and set it to Less than or equal to 0.05
- At the bottom of the table, click to **Generate filtered Node**
- Repeat this two more times for the other two contrasts performed to create 3 total filtered **Feature list** nodes

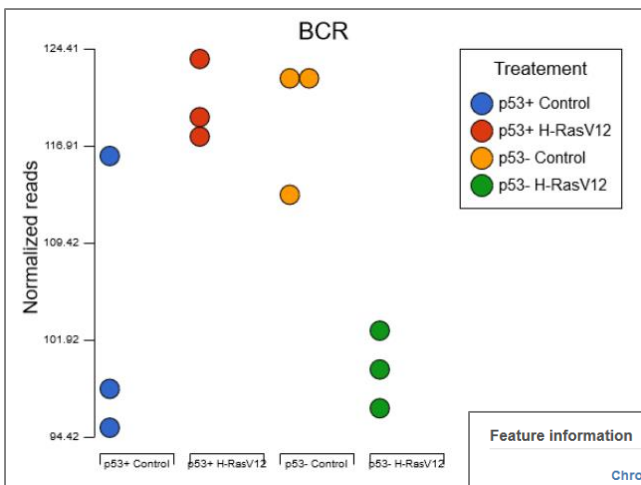
Gene list

Results: 1163		Optional columns								
Filter		View	Gene ID	Chromosome	Total counts	P-value	FDR step up	Ratio	Fold change	LSN
<input type="checkbox"/> Gene ID	◀	1	ZSWIM5	1	22.19	4.6E-4	2.32E-3	3.58	3.58	
<input type="checkbox"/> Chromosome	◀	2	CDC42BPA	1	1,000.37	4.58E-5	4.08E-4	1.79	1.79	
<input type="checkbox"/> Total counts	◀	3	PLEKHA6	1	568.71	6.51E-6	9.34E-5	0.45	-2.23	
<input type="checkbox"/> P-value	◀	4	PHTF1	1	153.31	1.76E-4	1.12E-3	1.60	1.60	
<input checked="" type="checkbox"/> FDR step up	▼	5	PFKFB2	1	228.05	7.2E-5	5.76E-4	0.57	-1.74	
<input type="radio"/> All contrasts	<input checked="" type="radio"/> Per contrast	6	PEX19	1	741.68	6.93E-3	0.02	1.19	1.19	
<input checked="" type="checkbox"/> p53- H-RasV12 vs p53+		7	PEAR1	1	82.74	5.87E-3	0.02	1.49	1.49	
H-RasV12		8	PATJ	1	448.72	2.25E-8	1.1E-6	0.37	-2.72	
Less than or equal to	0.05	9	PADI2	1	101.21	6.38E-3	0.02	0.46	-2.19	
<input type="range"/>	0 1	10	CGN	1	241.42	3.24E-9	2.48E-7	0.19	-5.32	
<input checked="" type="checkbox"/> p53- H-RasV12 vs p53-		11	NVL	1	316.01	4.29E-3	0.01	0.68	-1.46	
		12	PMF1	1	190.03	5.15E-5	4.46E-4	0.50	-2.01	

Notes:

Viewing Gene/Transcript Level Results

- Select **Feature List** data node and then click **Task report** in the menu
- 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

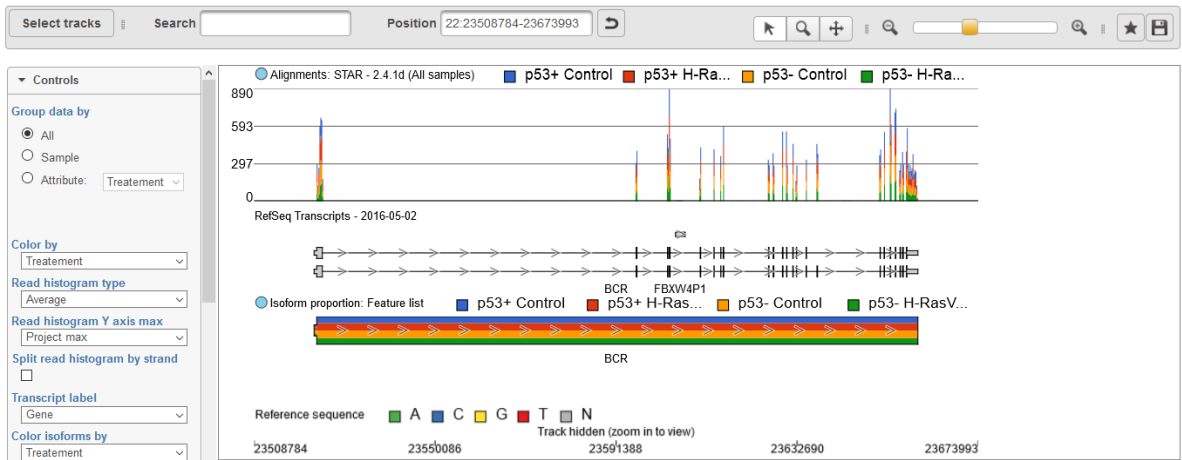


Feature information	
Chromosome	22
Start	23522552
Stop	23660225
Strand	+
Gene ID	BCR
Maximum counts	124.41
Average coverage	110.04
Total counts	1,327.62
Best model information	
Best model	Treatment
Best model type	Lognormal with shrinkage
Best AICc	97.36
Best Akaike weight	1.00
p53+ H-RasV12 vs p53+ Control	
Multi-model estimate	0.17
P-value (F)	0.02
FDR step up	0.04
Ratio	1.17
Fold change	1.17
LSMean(p53+ H-RasV12)	120.92
LSMean(p53+ Control)	102.92
FC 95% lower limit	1.03
FC 95% upper limit	1.34

Notes: _____

Chromosome Viewer

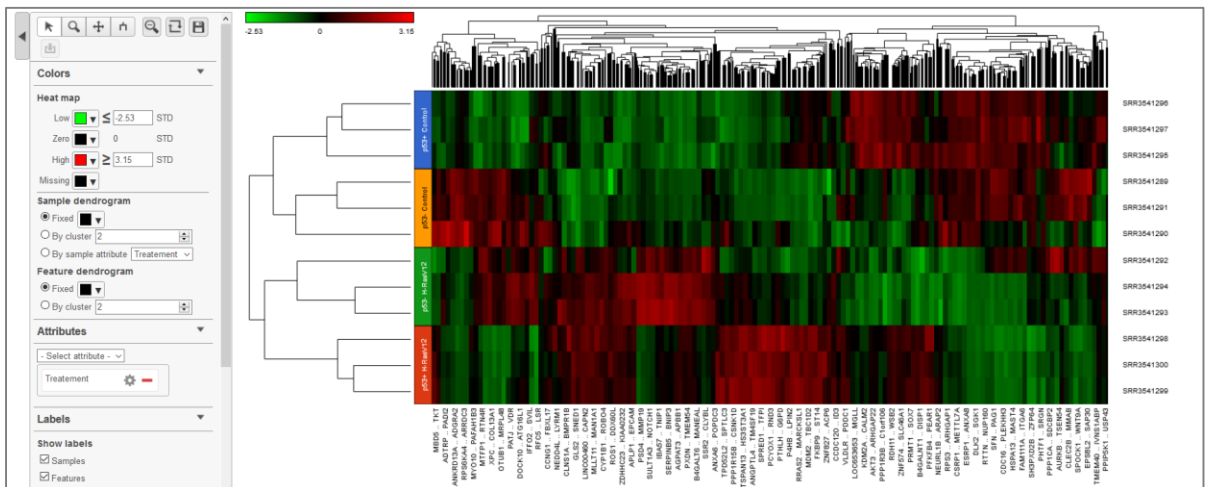
- **Select tracks** allows you to select different annotations or datasets to view together
- Sample grouping, color and transcript labeling can be edited in the **Controls** panel
- Search for any gene using the **Search** box
- Navigate to a genomic coordinate using the **Position** box
- Change and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

Description alpha-amino acid metabolic process

Enrichment score 8.97466

P-value 1.26577E-4

Gene breakdown

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

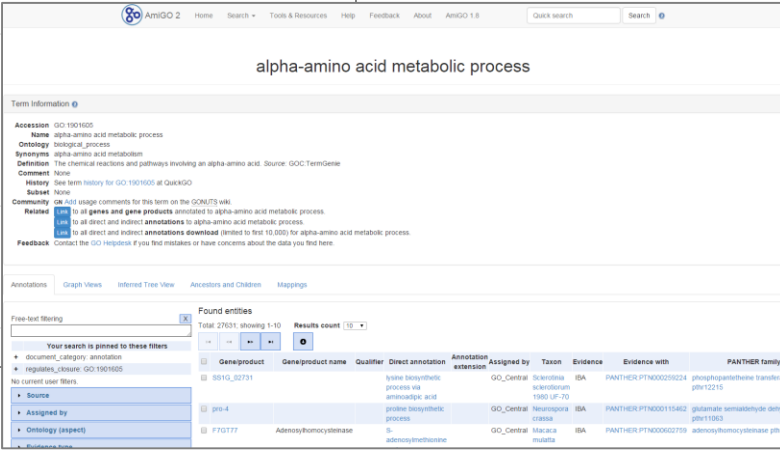
▼ **Genes in list**

Download data

PRODH GCAT GGT1

▼ **Genes not in list**

Download data



The screenshot shows the AmiGO 2 interface for the GO term 'alpha-amino acid metabolic process'. It includes sections for Term Information (Accession, Name, Synonyms, Definition, Comment, History, Subset, Community), Annotations (Graph Views, Inferred Tree View, Accessions and Children, Mappings), and a table of Found entities. The Found entities table lists genes like SBH1_G2731, ppi-4, and F7G177 with their respective annotations and evidence.

Notes:

Creating Gene Lists

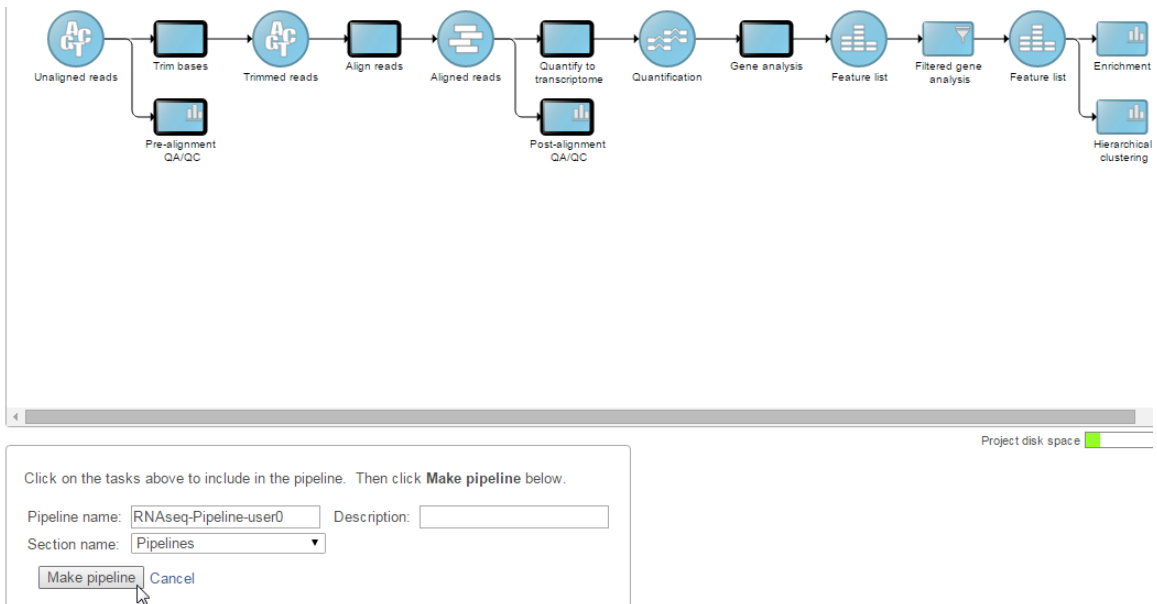
- In the Analysis tab, select **List generator** at the top right of the pipeline
- In the dialog, select 2 lists to compare (rename them for clarity), then click **Display selection**
- The venn diagram allows for selection of any region(s) and the associated list can be saved by clicking **Download list**

The screenshot displays the 'List generator' dialog box within a software pipeline. The pipeline steps are: Unaligned reads, Trim bases, Trimmed reads, Align reads, Aligned reads, Quantify to transcriptome, Gene counts, Normalize counts, Normalized counts, Filter features, Filtered counts, and Gene analysis. The 'List generator' dialog is open, showing a list of feature lists on the left and a Venn diagram on the right. The Venn diagram compares two feature lists: 'p53+ HRAS' (green circle) and 'p53- HRAS' (purple circle). The numbers in the Venn diagram are: 1619 for p53+ HRAS only, 2510 for the intersection, and 2800 for p53- HRAS only. The dialog includes fields for 'List name' and 'Identifier' for each list, a 'Display selection' button, and a 'Download list' button. A 'Save image' button is also present above the Venn diagram.

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: _____

