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 KE[1], Regunath K[1], Guzman A[2], Lee SM[3], Pfister NT[1], Akanni O[1], Kaufman LJ[2], Prives C[4], Prywes R[4].

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 ΔNp63α, the predominant p63 isoform in basal mammary epithelial cells. This regulation occurs at the transcriptional level, and a short region of the ΔNp63 promoter is sufficient for repression induced by H-RasV12. The suppression of ΔNp63α expression by these oncogenes concomitantly leads to an epithelial-to-mesenchymal transition (EMT). In addition, the depletion of ΔNp63α alone is sufficient to induce EMT. Both H-RasV12 expression and ΔNp63α 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 ΔNp63α.

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

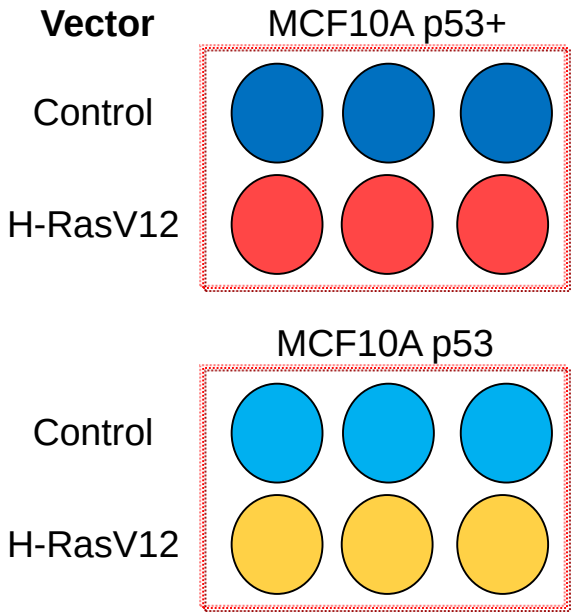
Data files in the project:

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

-Control vector

-H-RasV12 vector

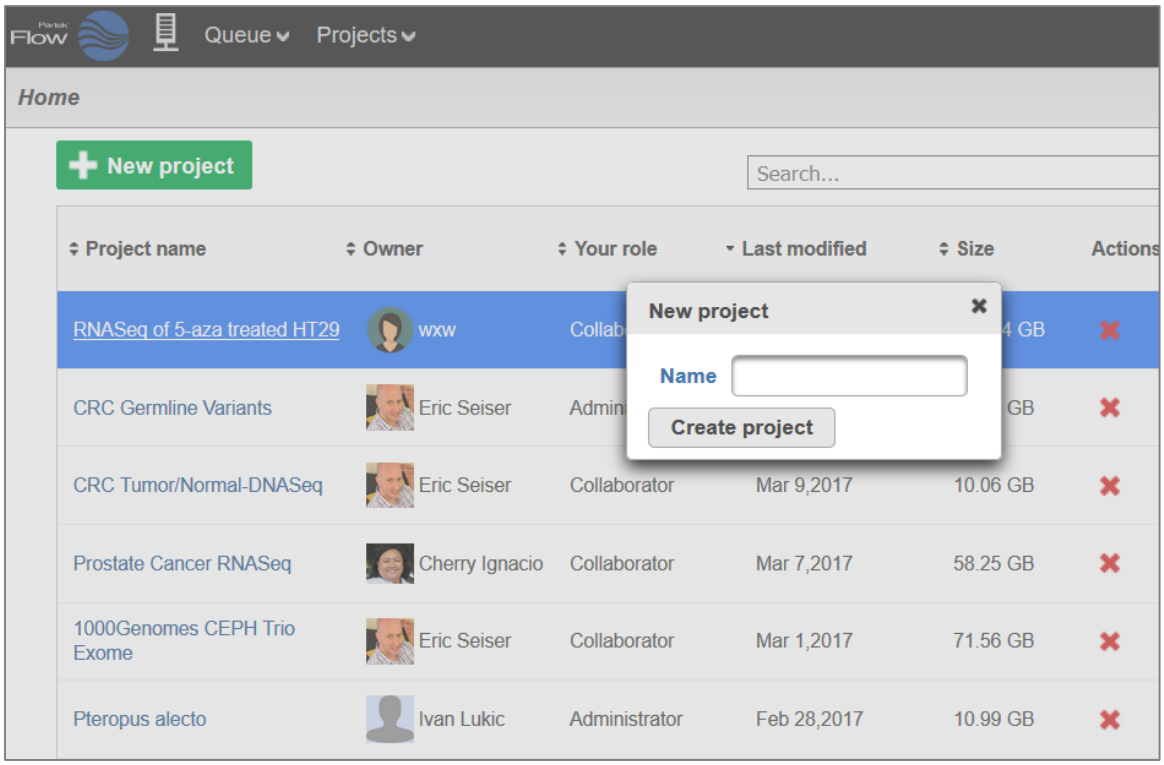
- mRNA purified and sequenced (paired-end reads) using the Illumina HiSeq 2000



Notes:_____

Login and Project Set-up

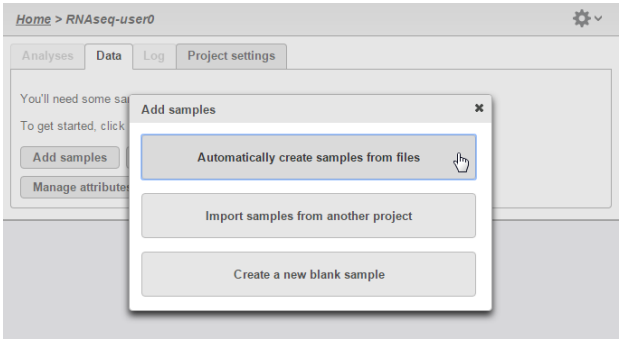
- Open Google Chrome/Firefox and enter: NCIdemo.partek.com
- Log in using the username and password given to you
- This will open to the Partek Flow homepage
- Click **New Project** and enter project name: RNAseq-[username]
- This will create a new project

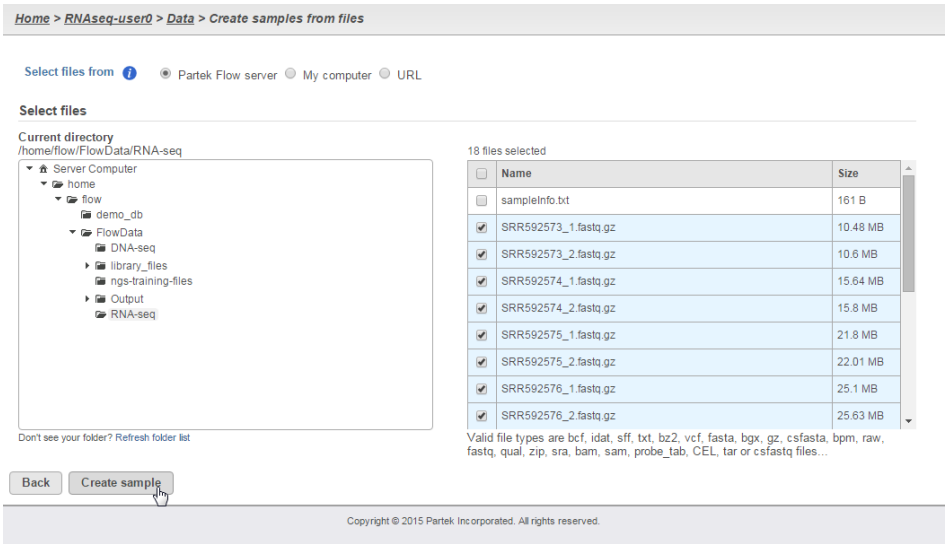


Notes:_____

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)

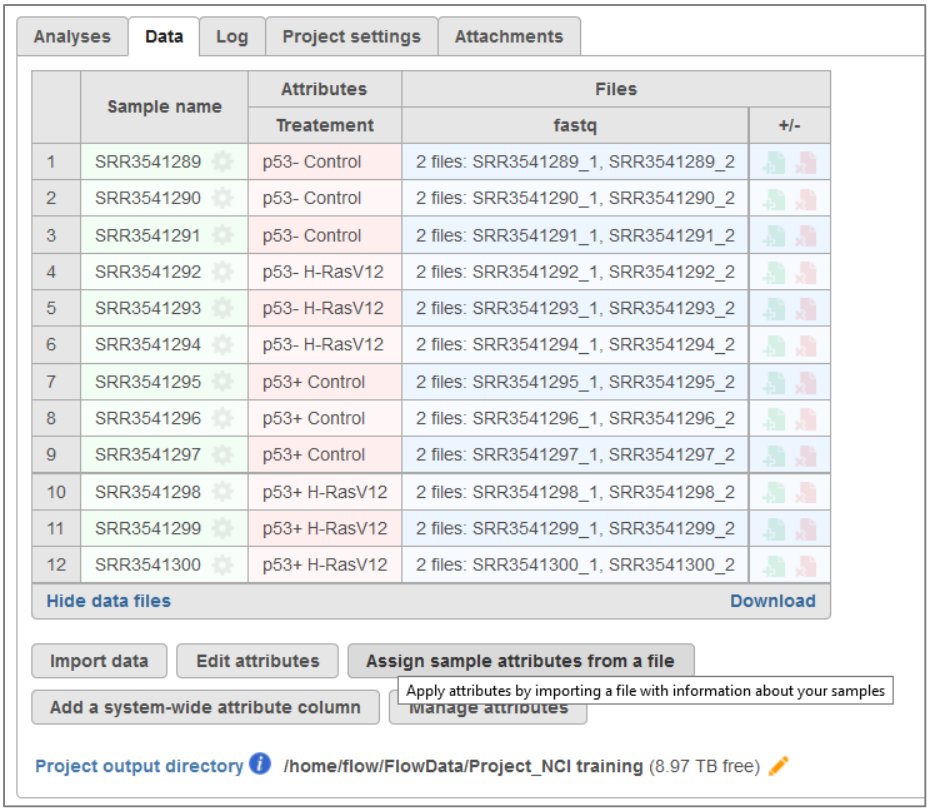




Notes:

Sample Attribute Assignment

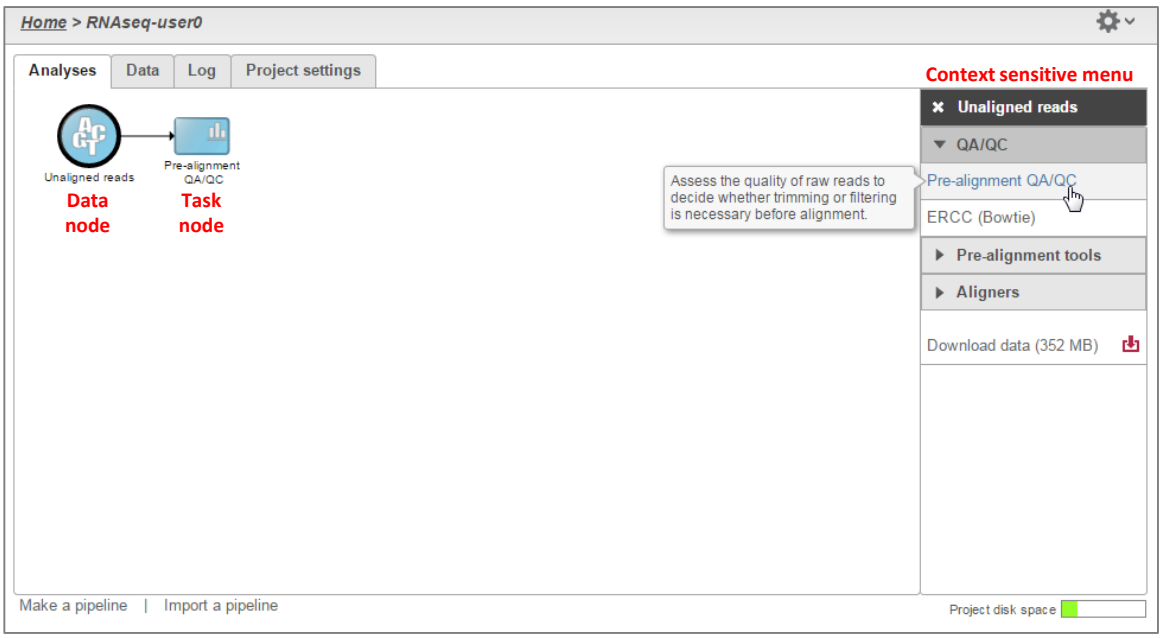
- Assign sample attributes using a tab-delimited text file
  - Contains table with ID in 1st column, followed by corresponding treatment groups
- Click **Assign sample attributes from a file**
- In the same folder, select sampleInfo.txt, click **Next**
- Click **Import**
- This will assign treatment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



Notes:_____

Analyses Tab Overview

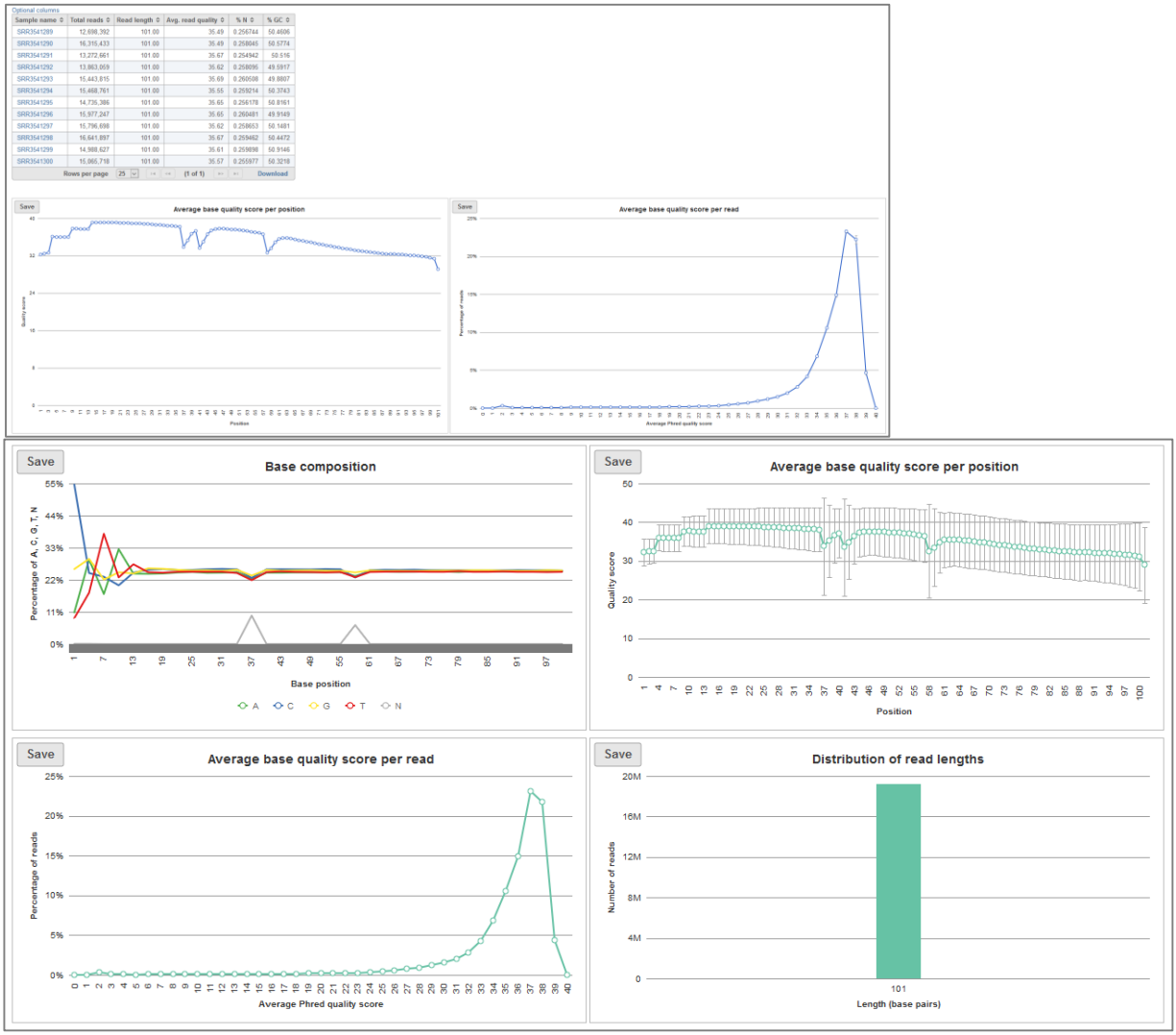
- Go to the **Analyses** tab
- Your first data node, the **Unaligned reads** node appears. All data nodes are circles
- Select the **Unaligned reads** data node and select **Pre-alignment QA/QC**
- Use the default settings and click **Finish**
- This will create a new task node in the **Analyses** tab.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

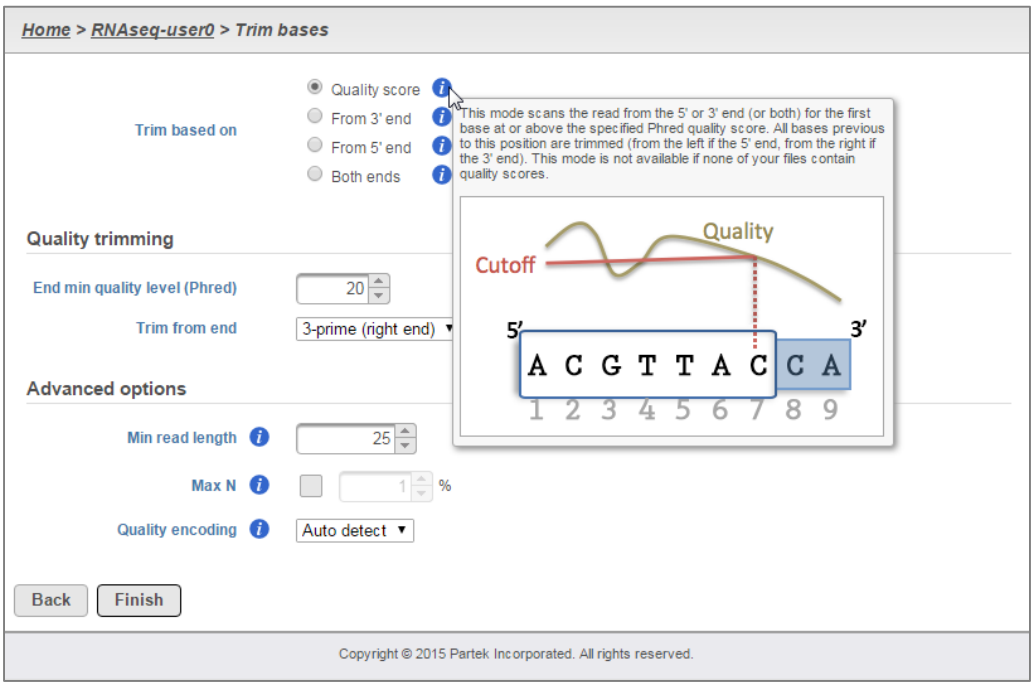


Notes:_____

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

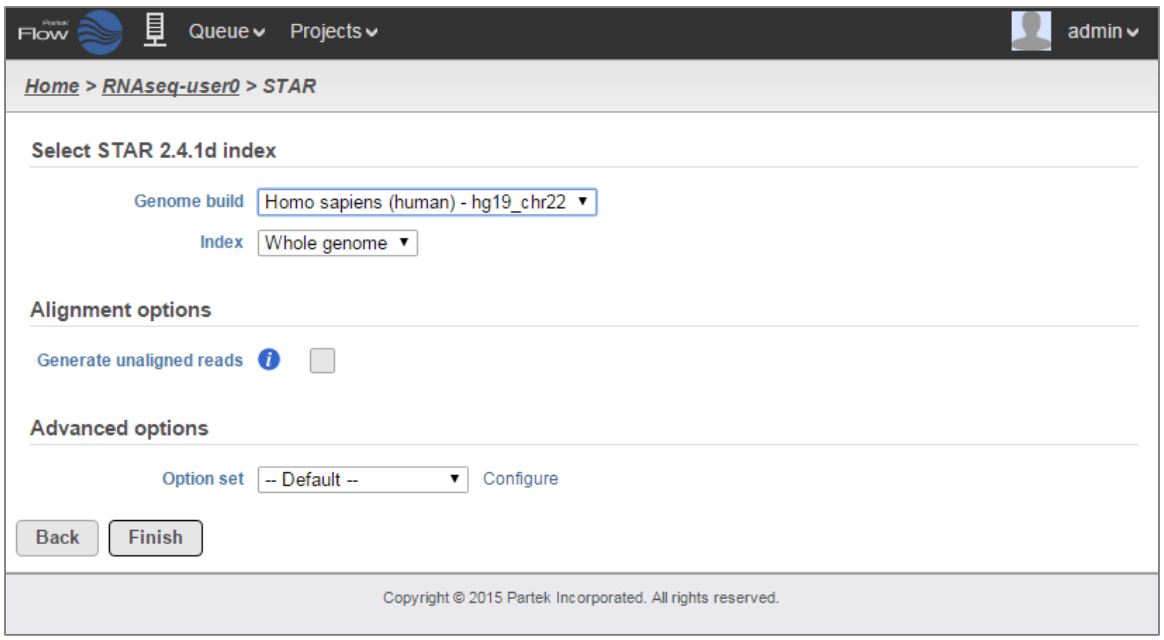


Notes:_____

Alignment

Using STAR aligner (can do junction reads)

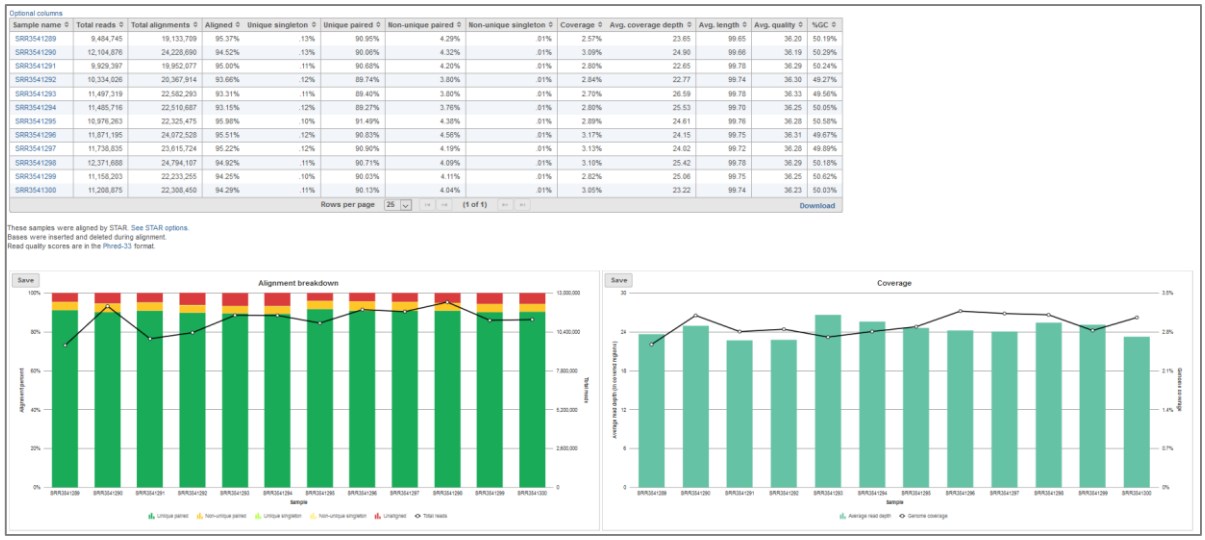
- Select the **Trimmed reads** data node
- Click **STAR** from the **Aligner** section of the menu
- Select STAR index:
  - Genome build: **Homo sapiens (human) - hg19_chr22**
  - Index: **Whole genome**
- Use the default options, click **Finish**



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



Notes:_____

_____

_____

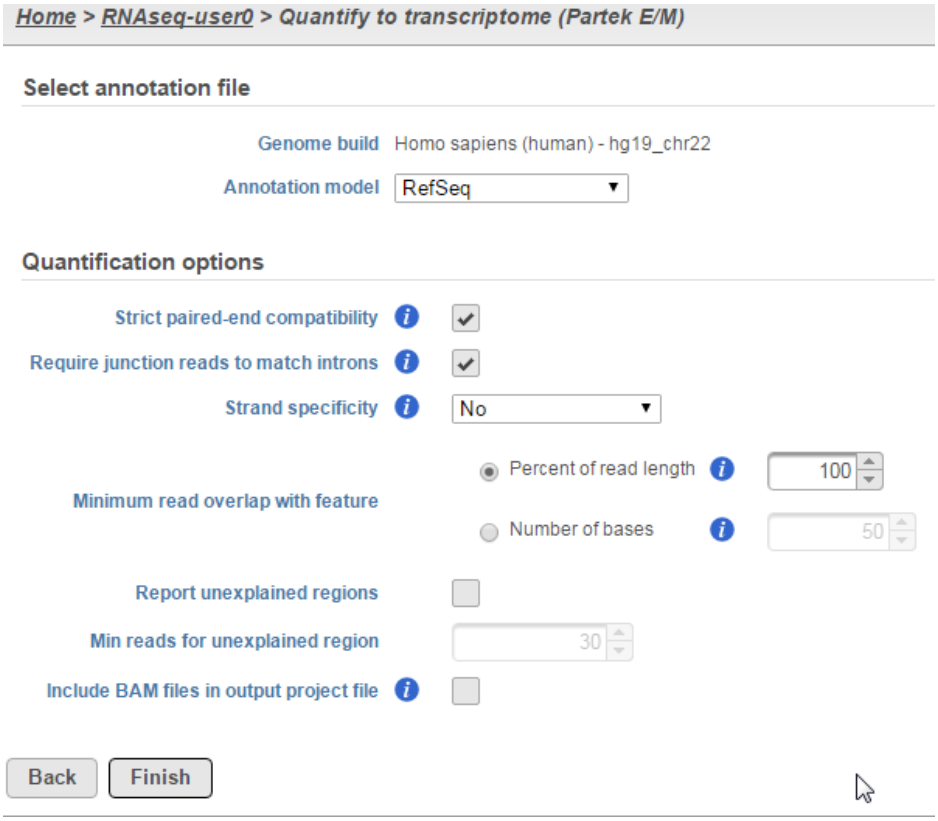
_____

_____

_____

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:_____

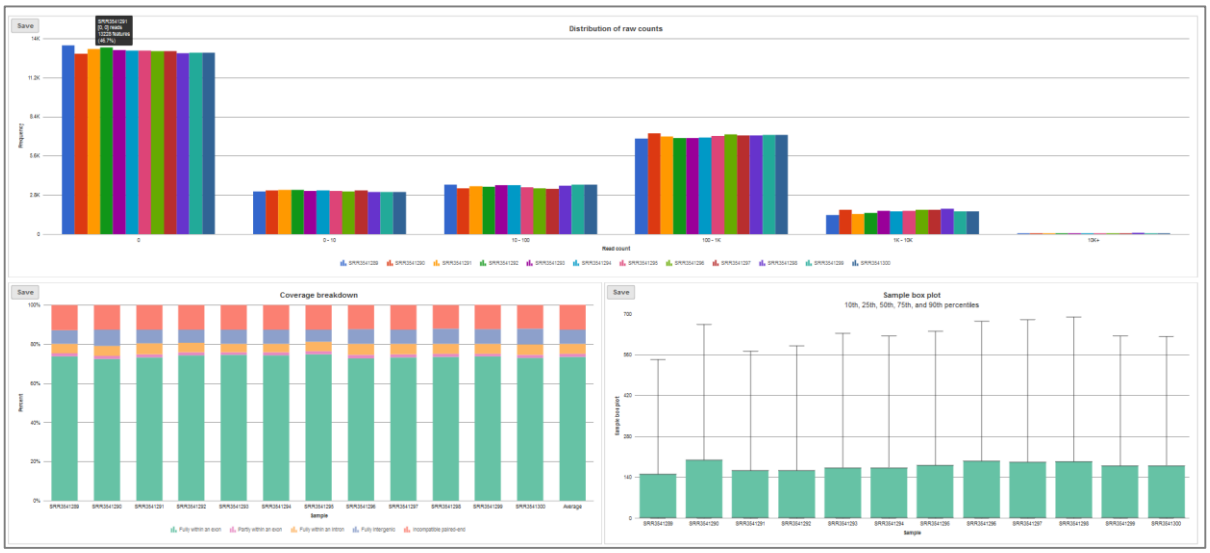
Viewing Quantification results

- To view the results of the quantification, select the **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

Optional columns

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	

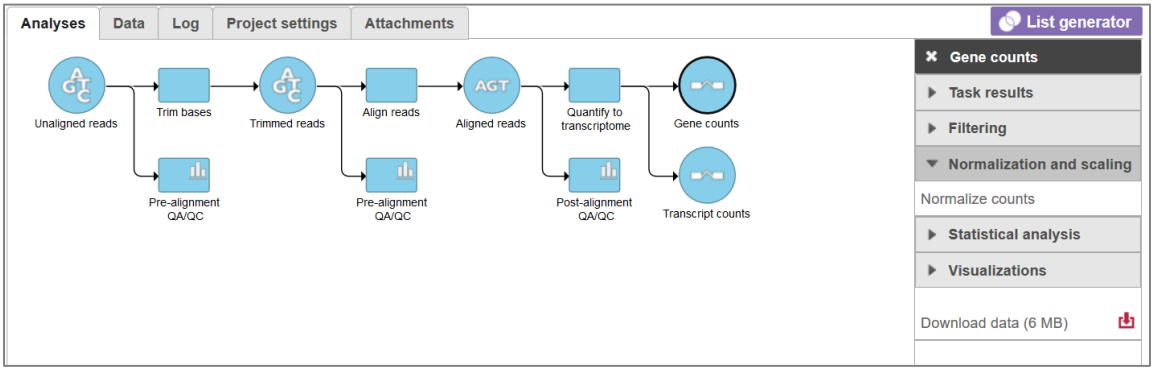
Rows per page 25 (1 of 1) Download

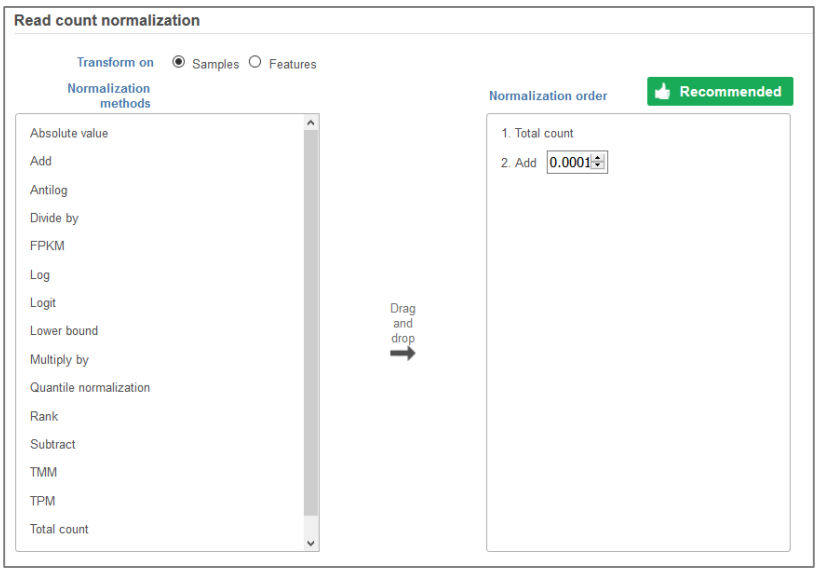


Notes:_____

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

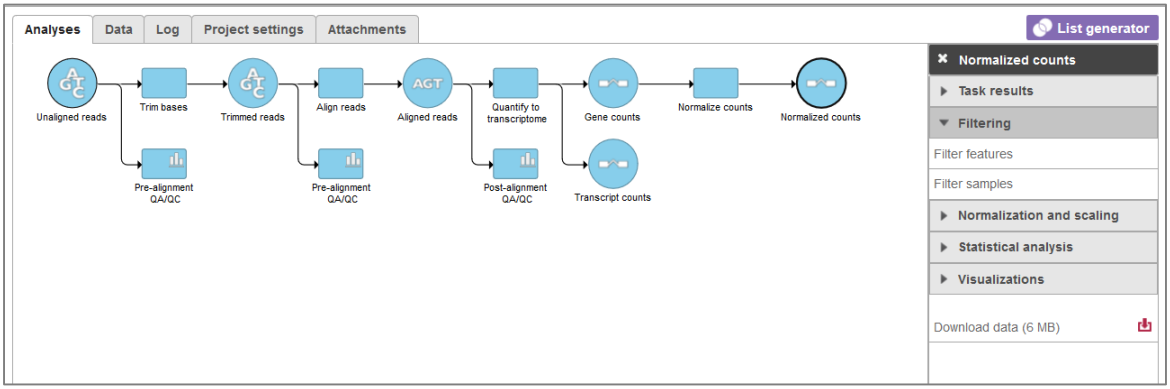


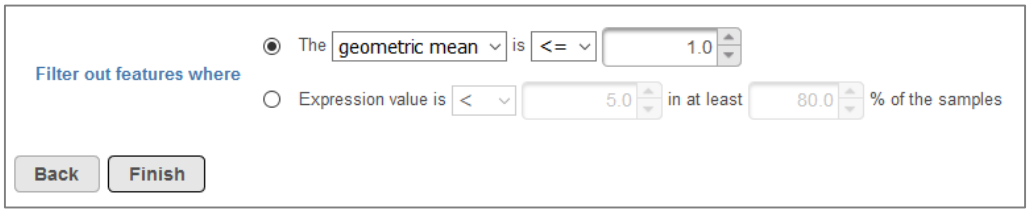


Notes:_____

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

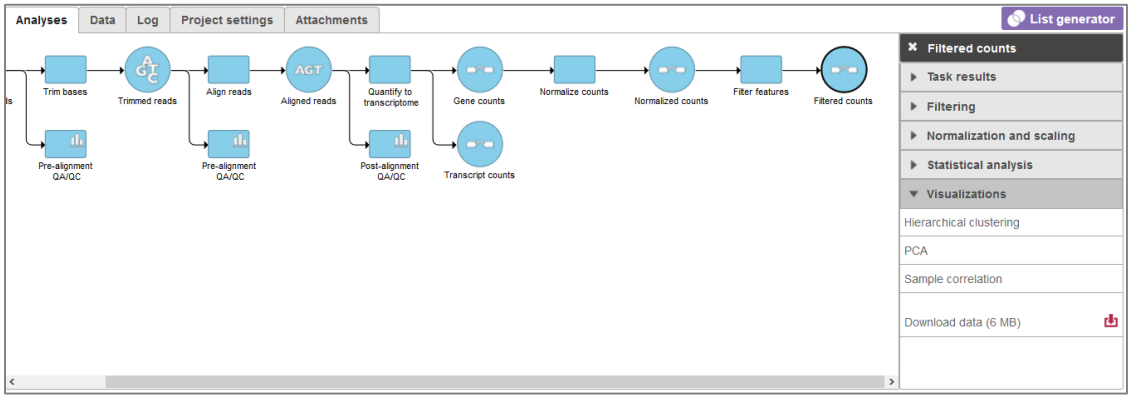


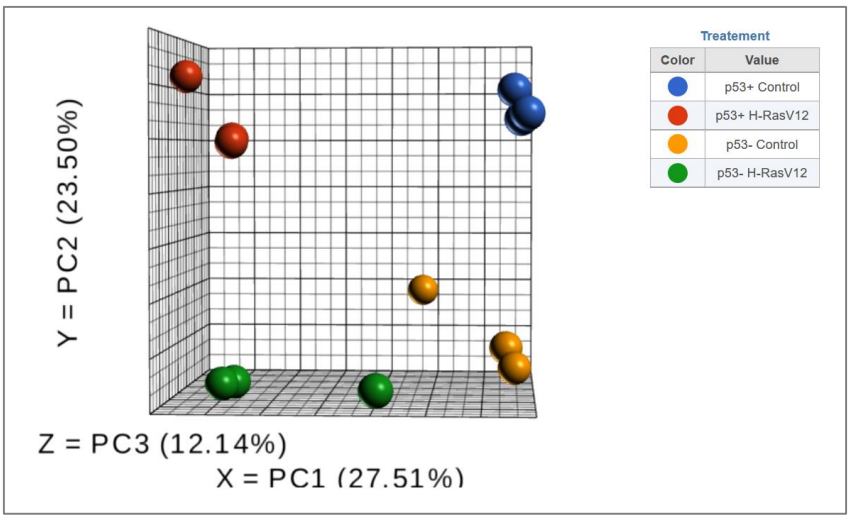


Notes:_____

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 **Filtered counts** data node and select **PCA** under the **Visualizations** portion of the menu

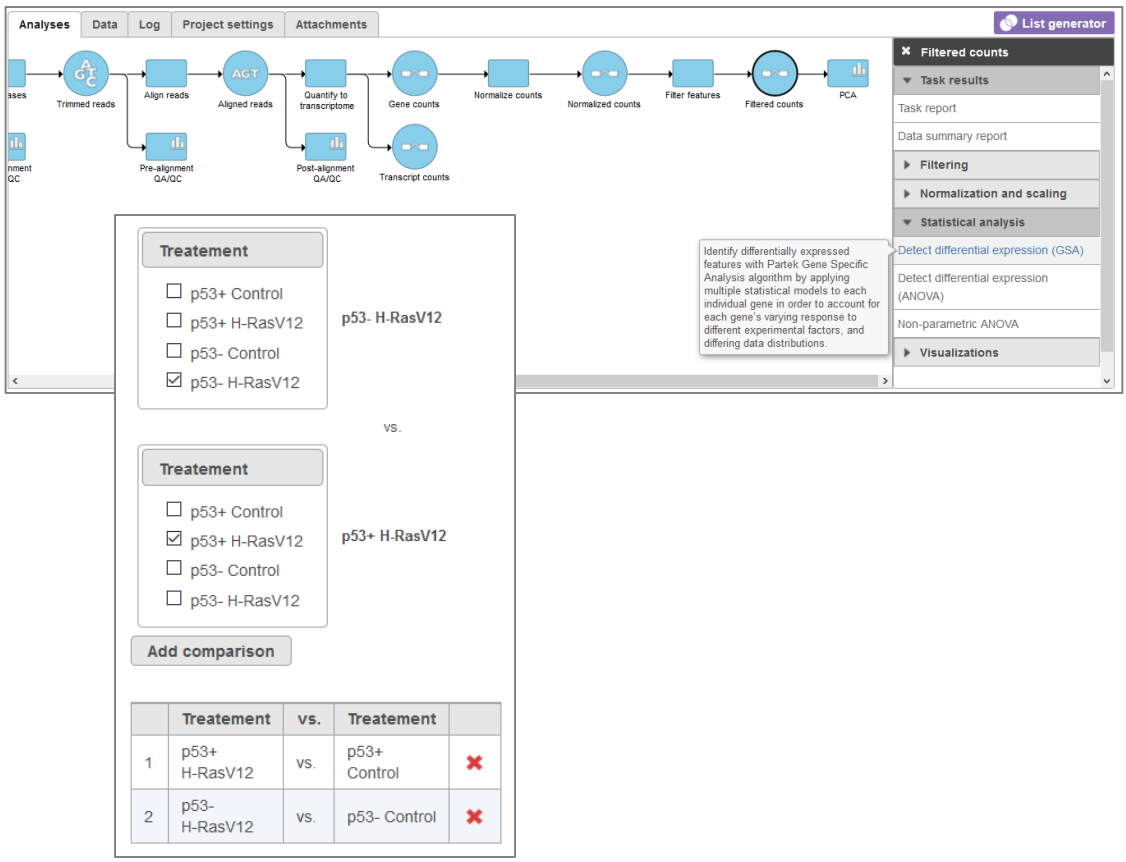




Notes:_____

Differential Expression Analysis

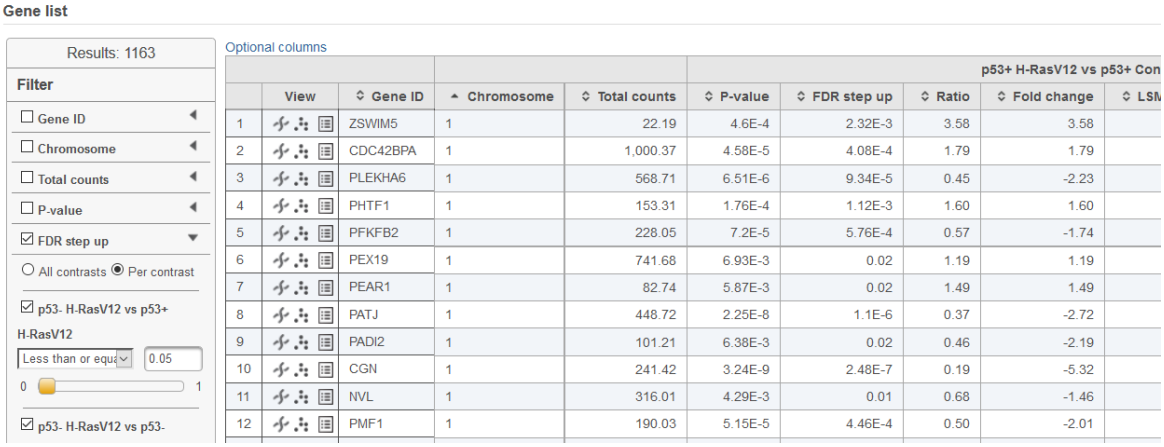
- Select the Quantification data node
- Click **Differential gene expression (GSA)** from the **Statistical analysis** section of the menu
- Select **Treatment** as the categorical attribute
- Select **p53+ H-RasV12 vs p53+ Control** and click **Add comparison**
- Select **p53- H-RasV12 vs p53- Control** and click **Add comparison**
- Select **p53- H-RasV12 vs p53+ H-RasV12**, click **Add comparison** and click **Finish**



Notes:_____

Creating a Filtered Gene List

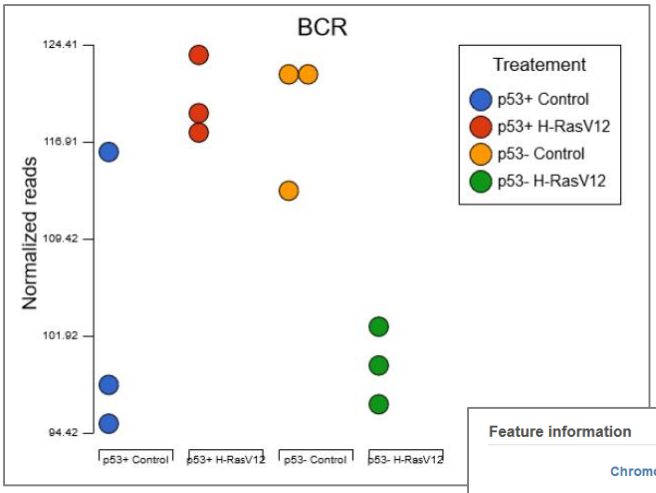
- Select **Feature List** data node and then click **Task report** in the toolbox
- Under the **Gene list** section, on the **Filter** panel select:
  - **FDR step up**, then select **Per contrast** for **p53+ H-RasV12 vs p53+ Control** and set it to Less than or equal to 0.05
- At the bottom of the 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



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	Gene ID	BCR
Start	23522552	Maximum counts	124.41
Stop	23660225	Average coverage	110.04
Strand	+	Total counts	1,327.62

Best model information

Best model	Treatement	Best AICc	97.36
Best model type	Lognormal with shrinkage	Best Akaike weight	1.00

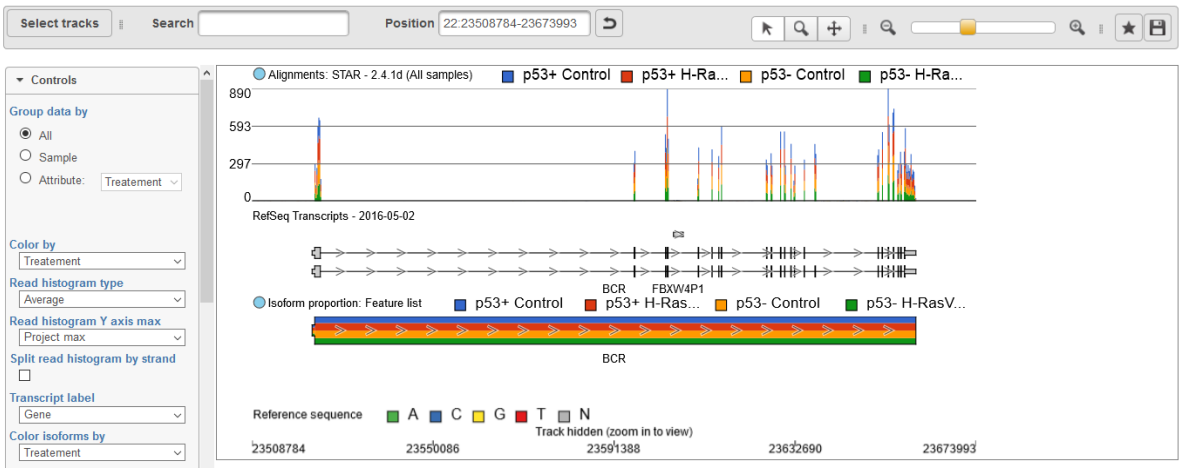
p53+ H-RasV12 vs p53+ Control

Multi-model estimate	0.17	LSMean(p53+ H-RasV12)	120.92
P-value (F)	0.02	LSMean(p53+ Control)	102.92
FDR step up	0.04	FC 95% lower limit	1.03
Ratio	1.17	FC 95% upper limit	1.34
Fold change	1.17		

Notes:_____

Chromosome Viewer

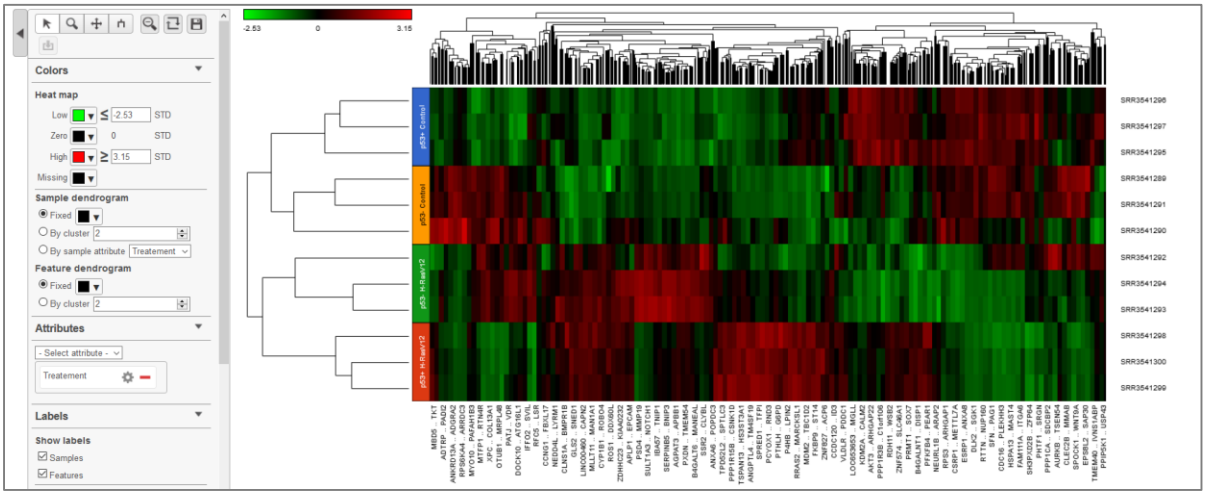
- **Select tracks** allows you to select different annotations or datasets to view together
- Sample grouping, color and transcript labeling can be edited in the **Controls** panel
- Search for any gene using the **Search** box
- Navigate to a genomic coordinate using the **Position** box
- Change and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

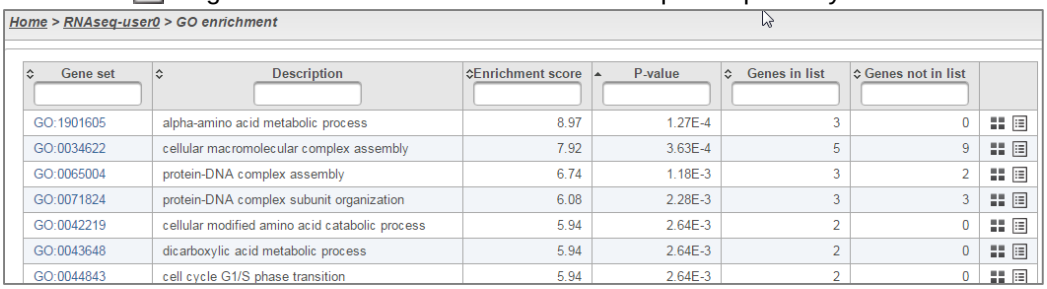
- Select any **Feature list** data node to perform clustering on that list of genes/transcripts
- For this training, select any **Feature list** produced after filtering
- Click **Hierarchical clustering** from the **Visualization** section of the menu
- Use default parameters and click **Finish**
-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



Notes:_____

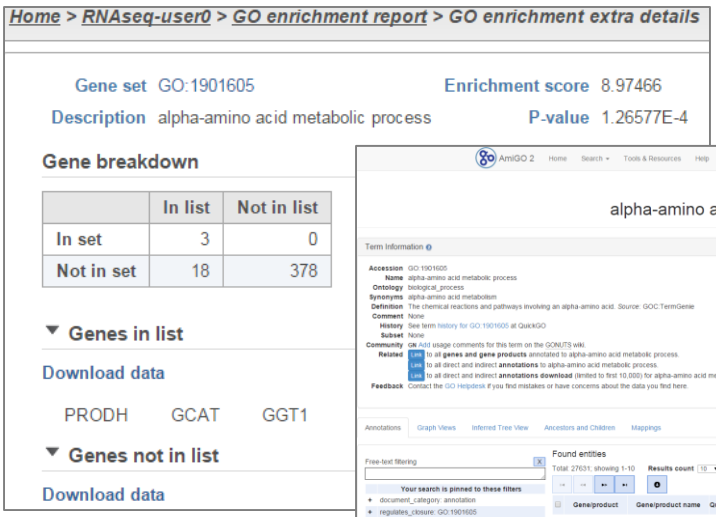
Enrichment Analysis

- 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 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

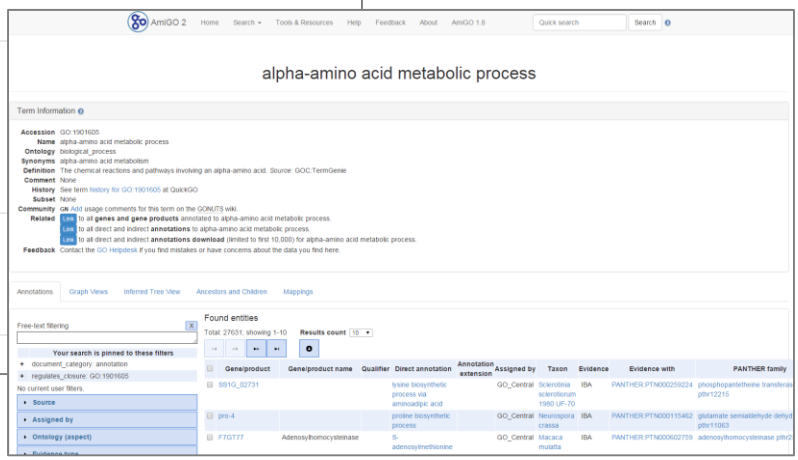
▼ Genes in list

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:TermGenie
Comment None
History See term history for GO:1901605 at QuickGO
Subset None
Community Add usage comments for this term on the GONUTS wiki.
Related Lists to all genes and gene products annotated to alpha-amino acid metabolic process.
Lists to all direct and indirect annotations to alpha-amino acid metabolic process.
Lists to all direct and indirect annotations download (limited to first 10,000) for 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 Ancestors and Children Mappings

Free-text filtering

Your search is pinned to these filters
document_category: annotation
regulates_closure: GO:1901605
No current user filters.
Source
Assigned by
Ontology (aspect)

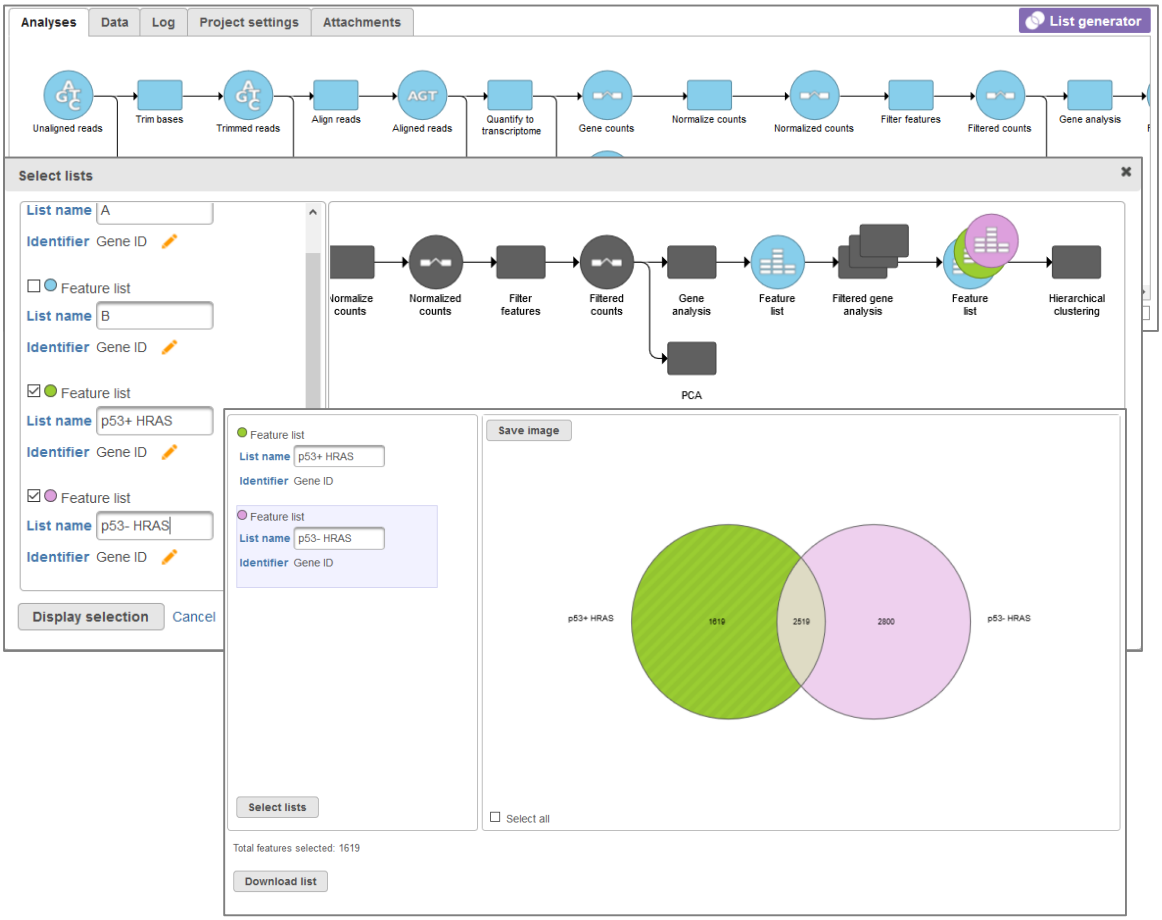
Found entities
Total: 27631; showing 1-10 Results count 10

Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Assigned by	Taxon	Evidence	Evidence with	PANTHER family
SS1G_02731			lysine biosynthetic process via aminoadipic acid		GO_Central	Sclerotinia sclerotiorum 1980 UF-70	IBA	PANTHER:PTN000259224	phosphopantetheine transferas pthr12215
pro-4			proline biosynthetic process		GO_Central	Neurospora crassa	IBA	PANTHER:PTN000115462	glutamate semialdehyde dehyd pthr11063
F7GT77	Adenosylhomocysteinase		S-adenosylmethionine		GO_Central	Macaca mulatta	IBA	PANTHER:PTN000602759	adenosylhomocysteinase pthr2

Notes:_____

Creating Gene Lists

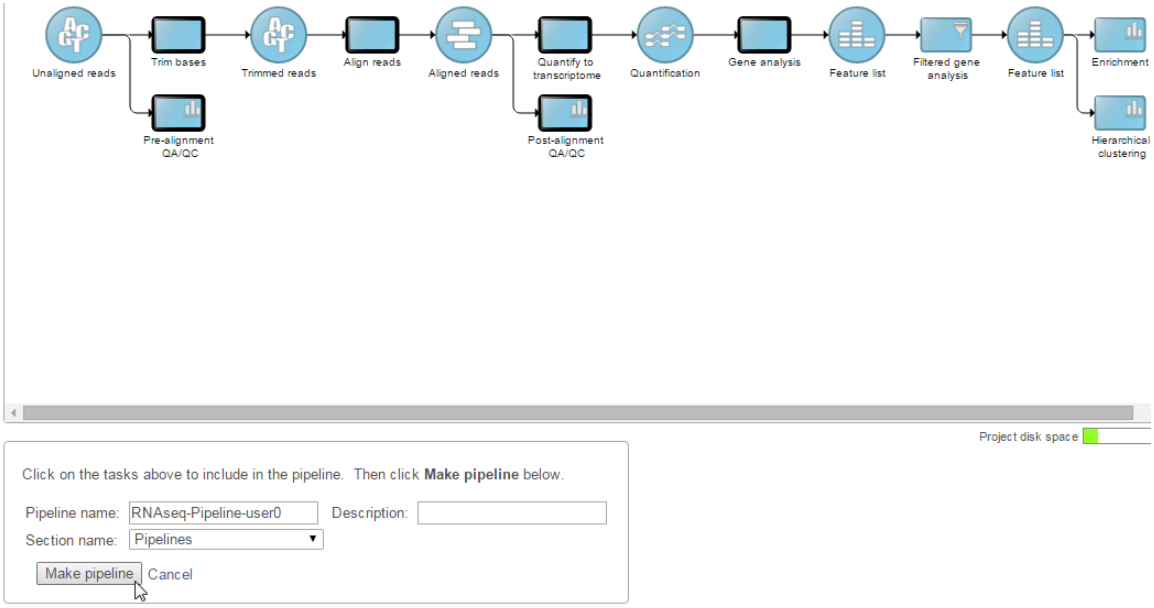
- In the Analysis tab, select **List generator** at the top right of the pipeline
- In the dialog, select 2 lists to compare (rename them for clarity), then click **Display selection**
- The venn diagram allows for selection of any region(s) and the associated list can be saved by clicking **Download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



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: