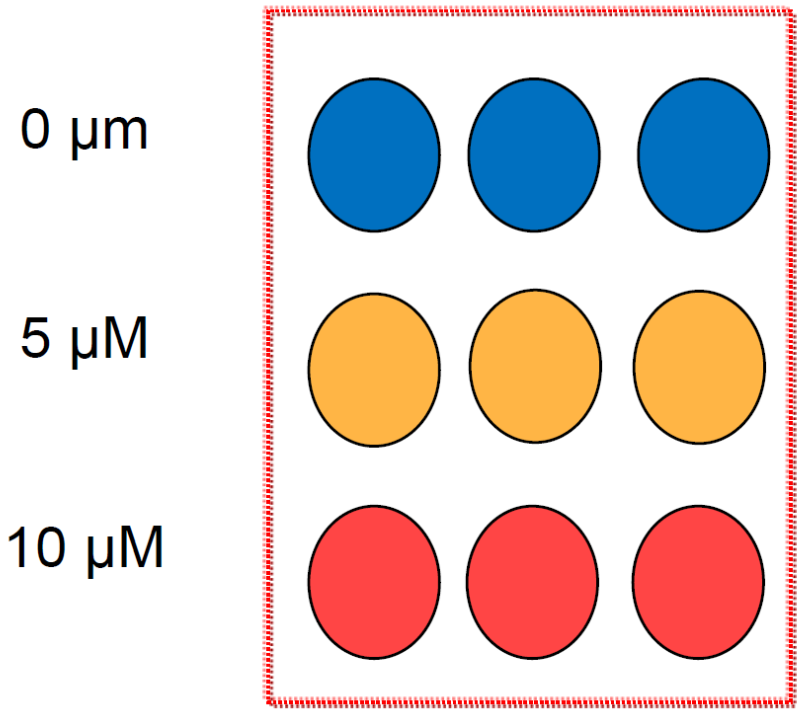
Partek Flow
RNA-Seq Analysis

Eric

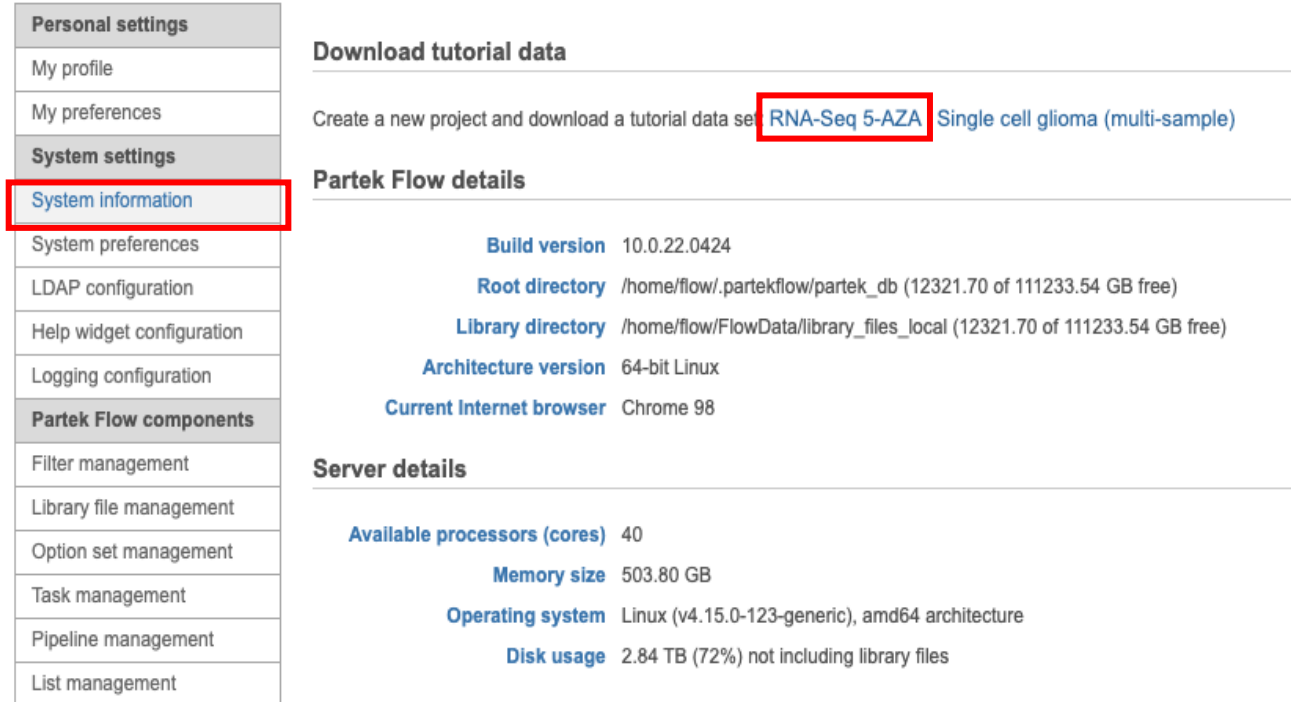
Experiment Description

- 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)

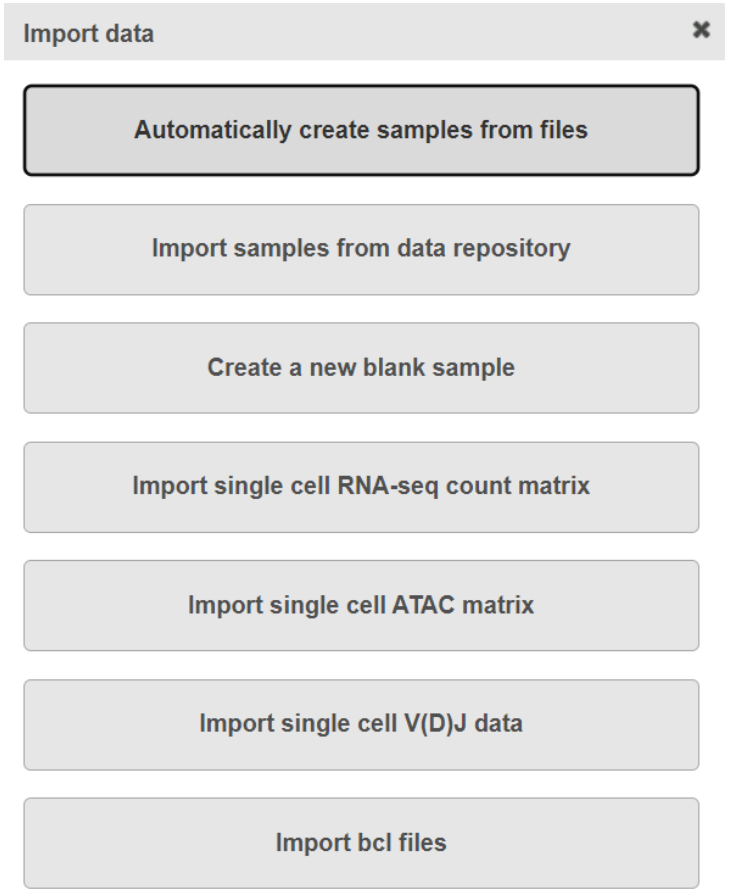


Import Tutorial Data

- Click your avatar in the upper right
- Click **Settings**
- On **System Information** page, click **RNA-Seq 5-AZA** from Download tutorial data block



Import Your Own Data



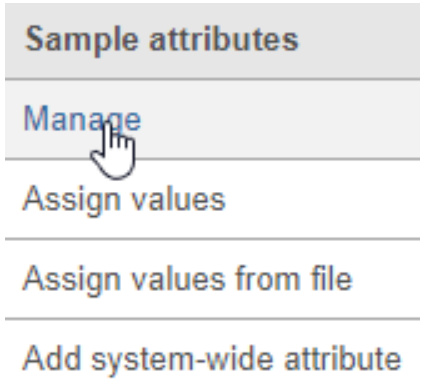
If you want to upload your own data

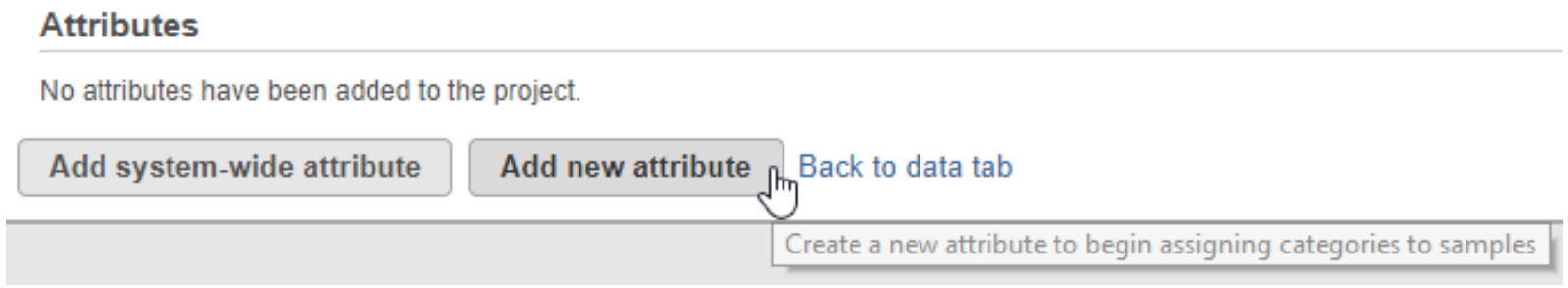
- Click **+New project** and enter a project name
- Import Fastq data by **Automatically create samples from files**
- Select all fastq.gz files and click **Create sample**
  - Partek Flow recognizes paired-end read data if tagged with (_1 or _R1)

Sample Attribute Assignment

Once the download completes, the sample table will appear in the **Data** tab.

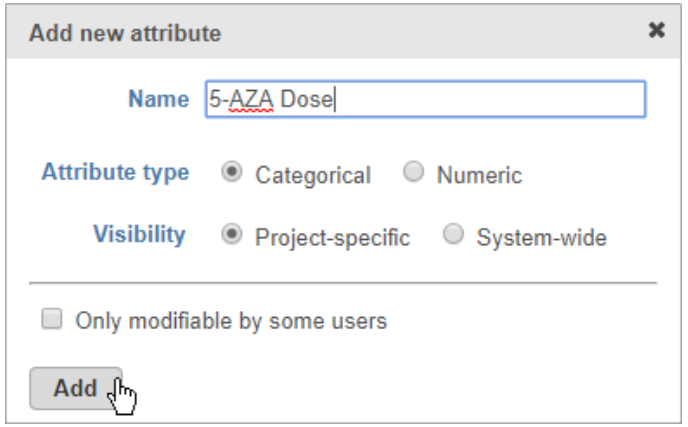
- Click **Data** tab
- Click **Manage** under Sample attributes
- Click **Add new attribute**

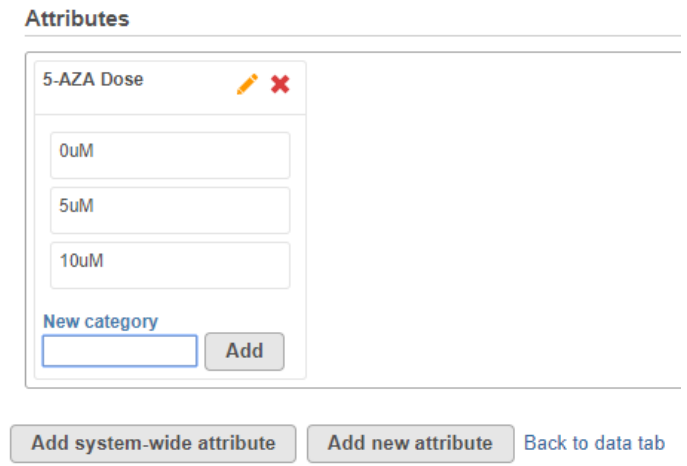




Sample Attribute Assignment

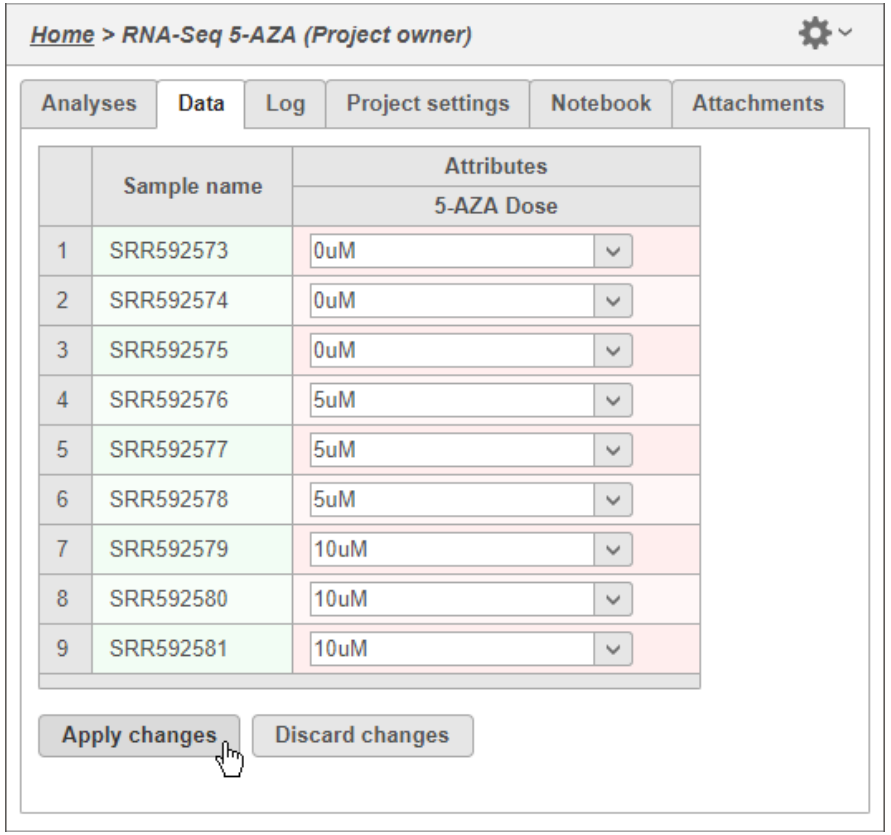
- Name the attribute **5-AZA Dose**
- Click **Categorical** and **Project-specific**
- Click **Add**
- Name the first New category **0uM**
- Click **Add**
- Repeat for two additional categories, **5uM** and **10uM**





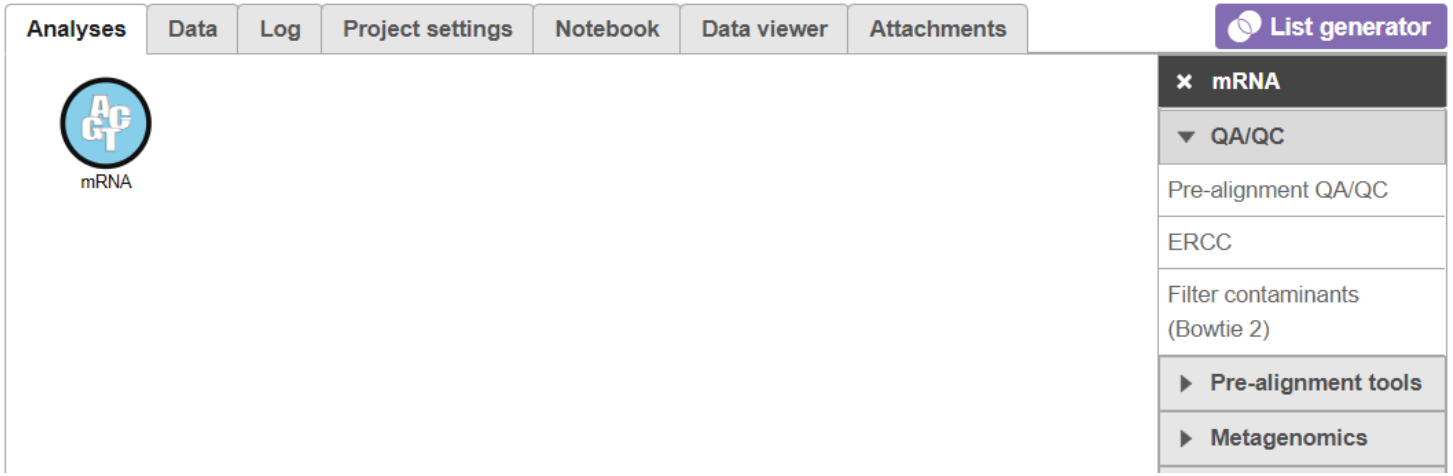
Sample Attribute Assignment

- Click **Back to data tab**
- Select **Assign values** under Sample attributes
- Edit the attribute for each sample with the drop-down menu
  - The first three samples should be 0uM, the next three samples should be 5uM, and the final three samples should be 10uM
- Click **Apply changes**



Pre-alignment QA/QC

- Click **Analyses** tab
- Click **mRNA** data node
- Click **Pre-alignment QA/QC** in the QA/QC section of the task menu
- Click **Finish** to run the task with default settings



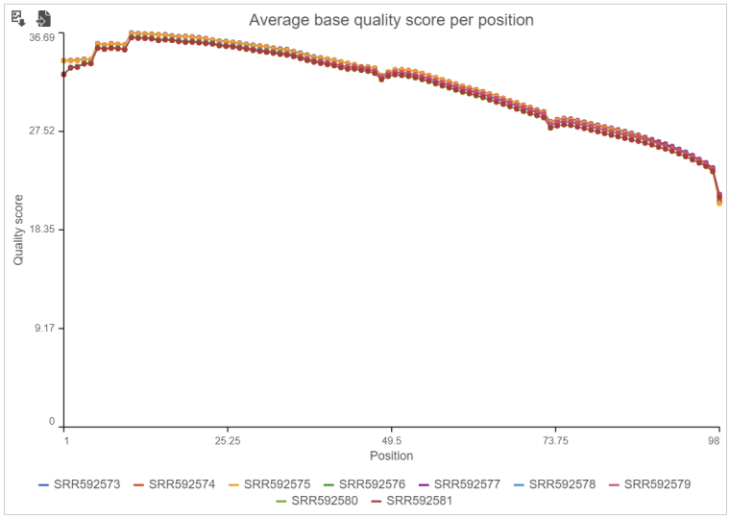
Pre-alignment QA/QC

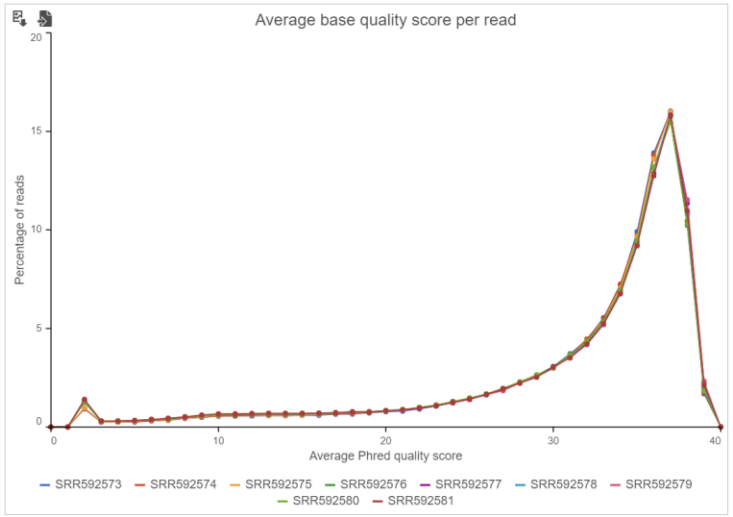
- Double click the **Pre-alignment QA/QC** node

Optional columns

Sample name	Total reads	Read length	Avg. read quality	% N	% GC
SRR592573	116,350	98.00	32.10	0%	53.74%
SRR592574	173,849	98.00	32.07	0%	53.61%
SRR592575	242,360	98.00	32.04	0%	53.25%
SRR592576	281,368	98.00	31.80	0%	52.95%
SRR592577	251,571	98.00	31.78	0.01%	52.02%
SRR592578	293,754	98.00	31.77	0%	52.89%
SRR592579	141,924	98.00	31.79	0%	51.96%
SRR592580	239,377	98.00	31.59	0.01%	53.06%
SRR592581	206,711	98.00	31.59	0%	51.98%

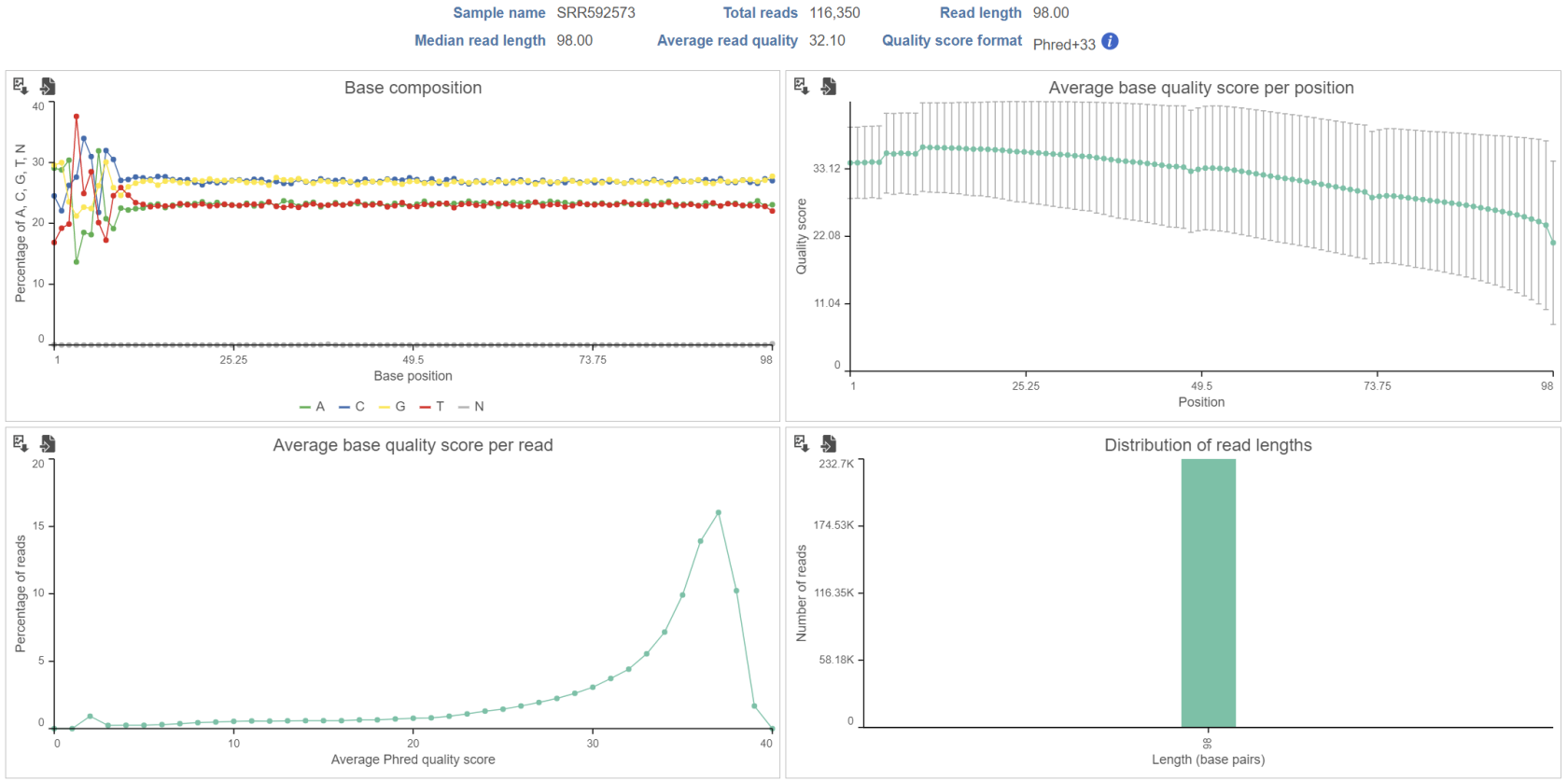
Rows per page 25 (1 of 1) Download





Pre-alignment QA/QC

- Click a sample name in the table to open a sample-level report



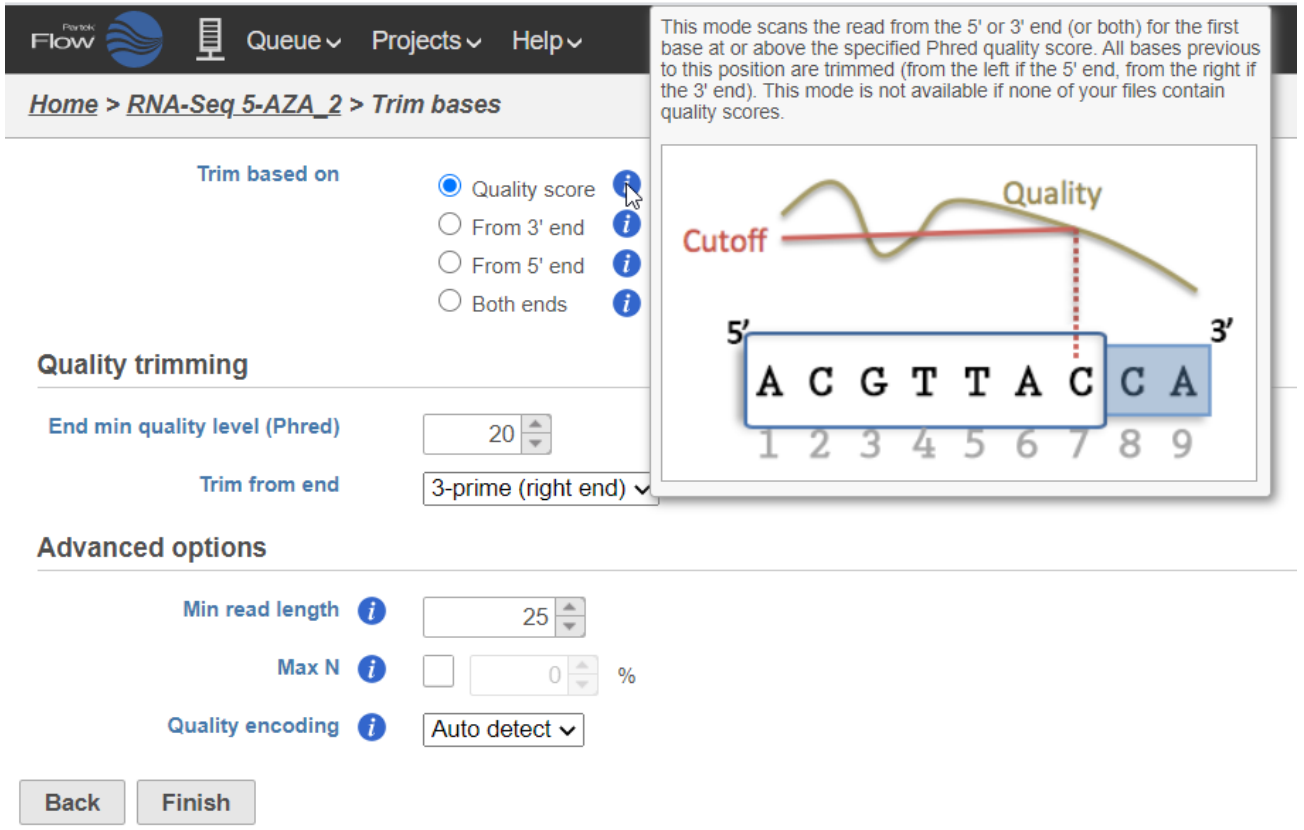
Trim Bases

- Click the project name to return to the Analyses tab
- Click **mRNA** data node
- Click **Pre-alignment tools** in the task menu
- Click **Trim bases**



Trim Bases

- Click **Finish** to run with default settings



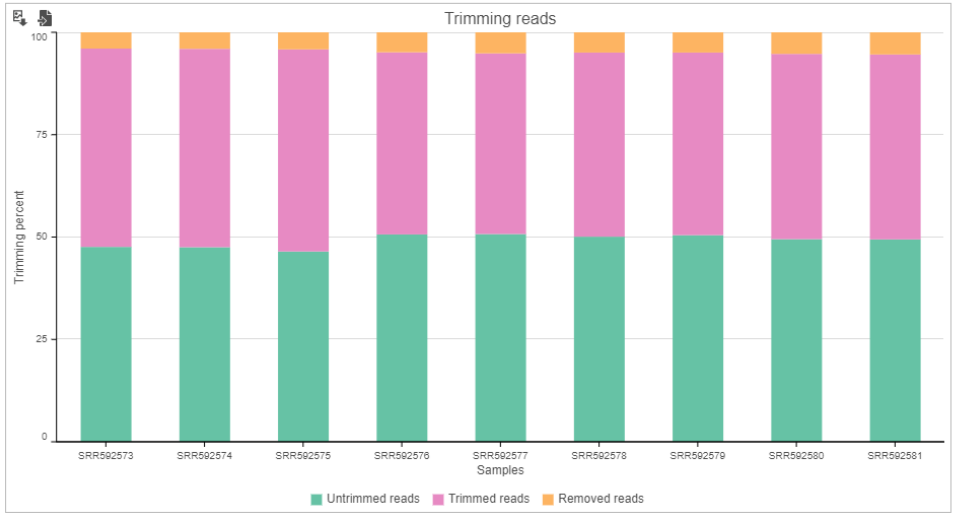
Trim Bases

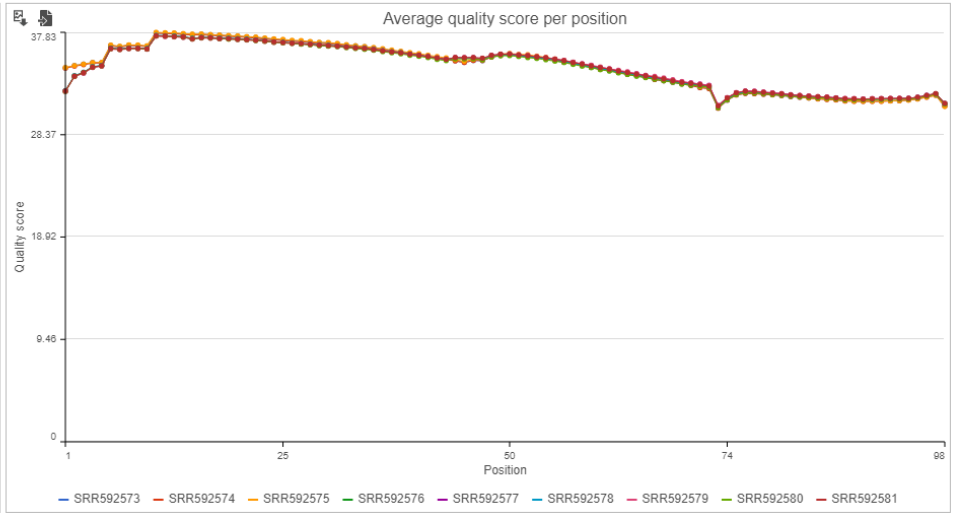
- Double-click **Trimmed reads** data node to open the task report

Optional columns

Sample name	Total reads	Reads trimmed	% Reads trimmed	Reads removed	% Reads removed	Average bases trimmed	Pre-trim quality	Post-trim quality
SRR592573	116,350	56,346	48.43%	4,623	3.97%	20.33	32.19	34.51
SRR592574	173,849	84,226	48.45%	7,003	4.03%	20.63	32.09	34.48
SRR592575	242,360	119,611	49.35%	10,109	4.17%	20.56	32.20	34.57
SRR592576	281,368	125,164	44.48%	13,722	4.88%	22.71	31.73	34.41
SRR592577	251,571	111,013	44.13%	12,924	5.14%	22.78	31.67	34.43
SRR592578	293,754	132,092	44.97%	14,518	4.94%	22.61	31.73	34.41
SRR592579	141,924	63,291	44.59%	7,011	4.94%	22.90	31.71	34.42
SRR592580	239,377	108,306	45.24%	12,605	5.27%	23.55	31.49	34.35
SRR592581	206,711	93,354	45.16%	11,179	5.41%	24.12	31.50	34.44

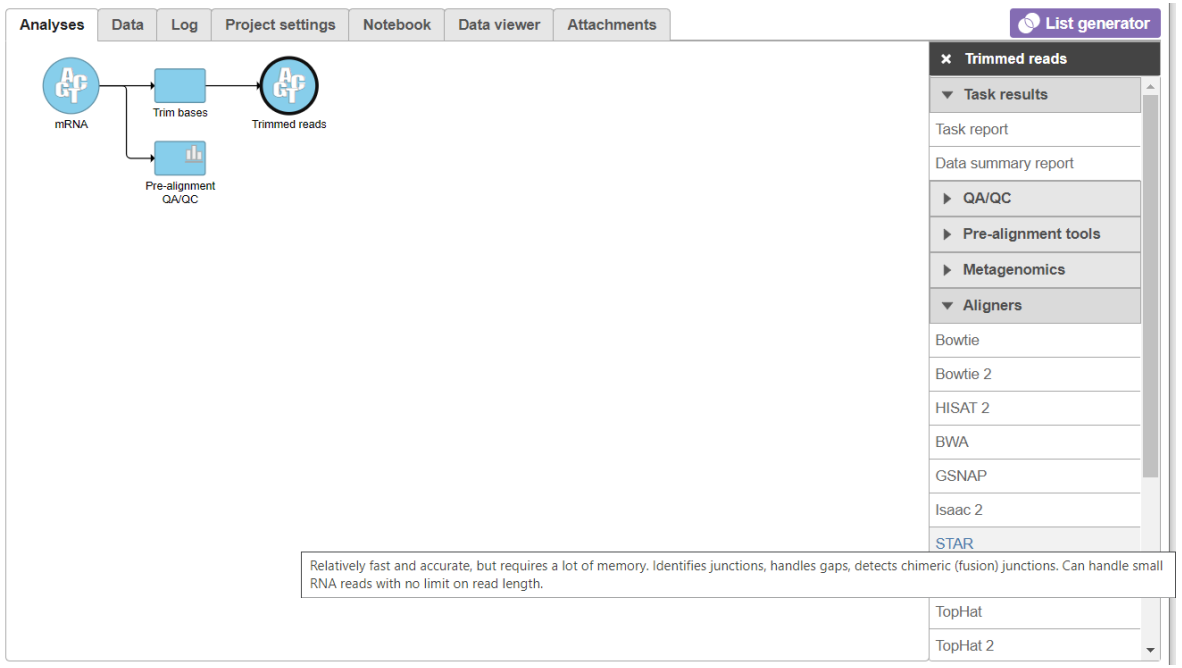
Rows per page 25 (1 of 1) Download





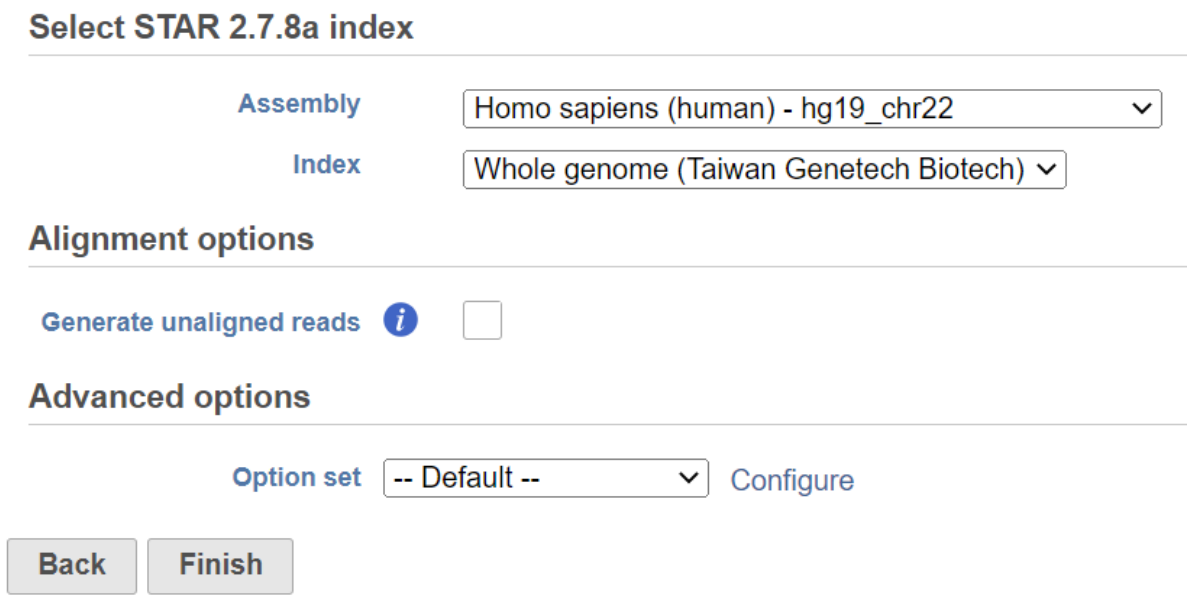
Alignment

- Click the project name to return to the Analyses tab
- Click **Trimmed reads**
- Click **Aligners** in the task menu to display available aligners
- Click **STAR**



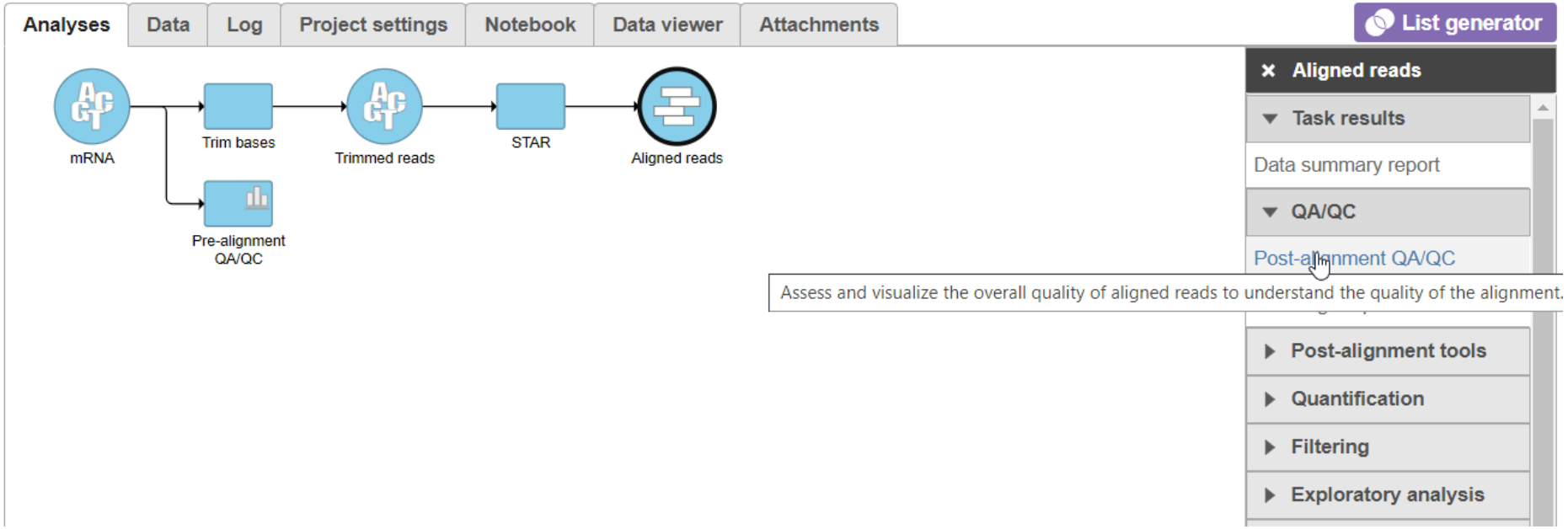
Alignment

- Choose **Homo sapiens - hg19_chr22** for Assembly and **Whole genome** for Index
- Click **Finish**



Post-alignment QA/QC

- Click **Aligned reads** data node
- Click **QA/QC** in the task menu
- Click **Post-alignment QA/QC**



Post-alignment QA/QC

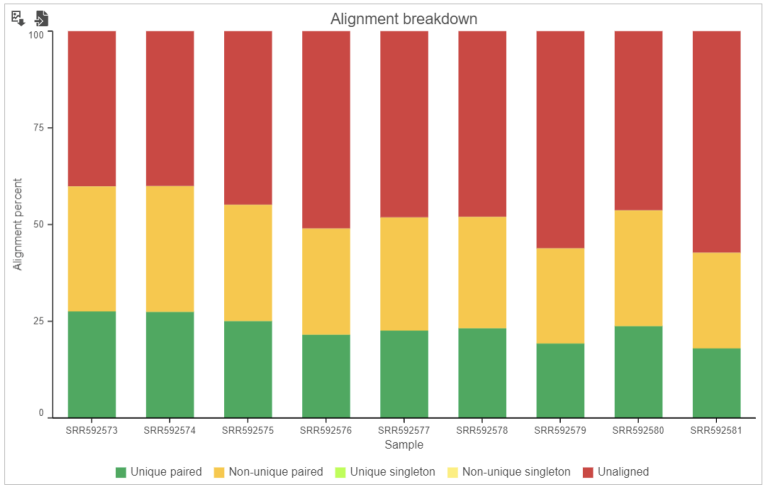
- Double-click the **Post-alignment QA/QC** task node to view the task report

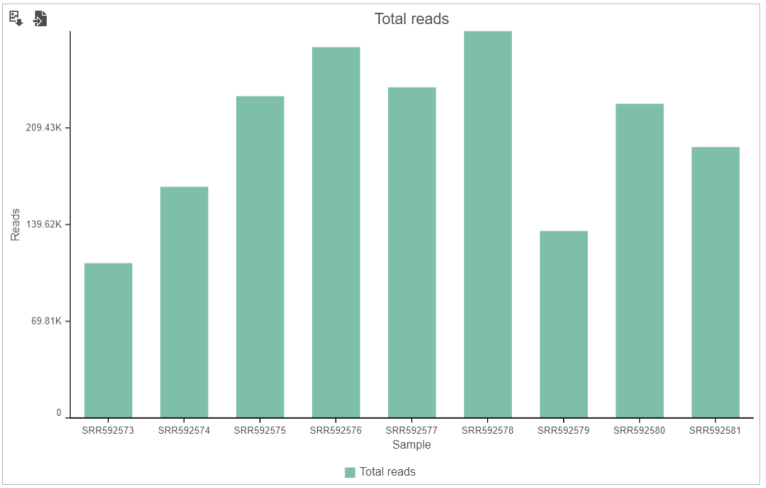
Optional columns

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	293,787	59.88%	0.02%	27.53%	32.29%	0.03%	1.84%	28.60	92.65	34.23	55.47%
SRR592574	166,846	438,638	59.96%	0.03%	27.42%	32.47%	0.04%	1.95%	40.31	92.61	34.26	55.25%
SRR592575	232,251	563,991	55.13%	0.03%	25.05%	30.02%	0.03%	2.07%	48.83	92.56	34.24	55.31%
SRR592576	267,646	596,983	49.00%	0.02%	21.53%	27.41%	0.03%	2.26%	47.09	92.49	34.13	55.50%
SRR592577	238,647	560,590	51.87%	0.03%	22.57%	29.24%	0.03%	2.21%	45.42	92.46	34.16	54.39%
SRR592578	279,236	654,863	52.01%	0.03%	23.19%	28.75%	0.04%	2.28%	51.34	92.48	34.15	55.12%
SRR592579	134,913	266,265	43.87%	0.03%	19.24%	24.57%	0.03%	1.96%	24.22	92.32	34.13	54.84%
SRR592580	226,772	547,676	53.69%	0.03%	23.73%	29.89%	0.04%	2.13%	45.94	92.33	34.09	55.41%
SRR592581	195,532	391,114	42.74%	0.03%	17.97%	24.70%	0.04%	2.03%	34.23	92.10	34.11	55.39%

Rows per page 25 (1 of 1) Download

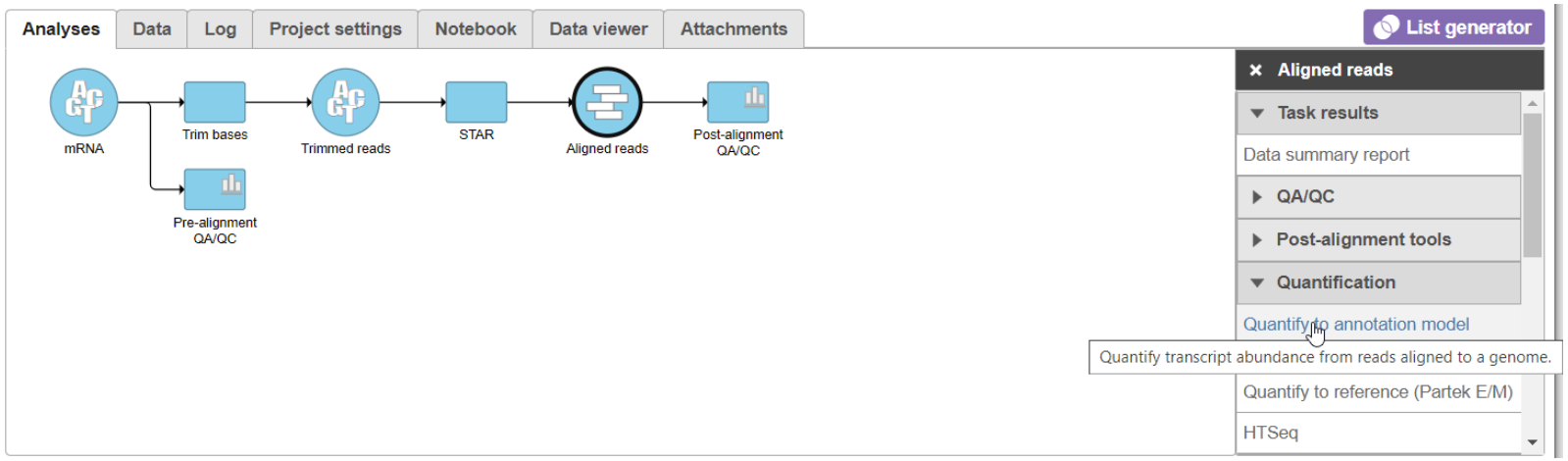
These samples were aligned by STAR. See STAR options.
Bases were inserted and deleted during alignment.
Read quality scores are in the Phred-33 format.





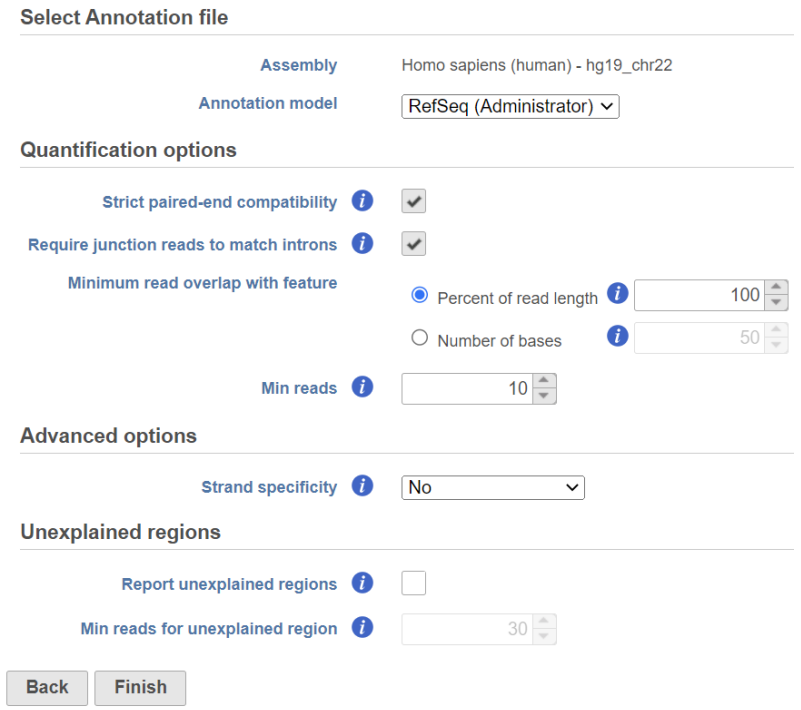
Quantification

- Click the project name to return to the Analyses tab
- Click **Aligned reads** data node
- Click **Quantification** in the task menu
- Click **Quantify to an annotation model (Partek E/M)**



Quantification

- Choose the RefSeq for Annotation model
  - You may need to download it first, via Library File Management
- Click **Finish**



Quantification

- Double-click **Gene counts** data node to view the task report

Summary of reads quantified to hg19_chr22 - RefSeq

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
SRR592573	66,904.00	99.76%	0.02%	0%	0%	0.22%	27,215.00	27,324.00	
SRR592574	100,040.00	99.72%	0.03%	0%	0%	0.25%	39,862.00	40,049.00	
SRR592575	128,042.00	99.76%	0.01%	0%	0%	0.23%	52,017.00	52,222.00	
SRR592576	131,137.00	99.66%	0.04%	0%	0%	0.30%	50,390.00	50,729.00	
SRR592577	123,783.00	99.67%	0.03%	0%	0%	0.30%	45,718.00	46,011.00	
SRR592578	145,219.00	99.64%	0.03%	0%	0%	0.33%	57,484.00	57,861.00	
SRR592579	59,185.00	99.63%	0.03%	0%	0%	0.34%	22,558.00	22,724.00	
SRR592580	121,755.00	99.67%	0.03%	0%	0%	0.30%	48,227.00	48,504.00	
SRR592581	83,569.00	99.63%	0.02%	0%	0%	0.35%	29,834.00	30,043.00	
Average	106,626.00	99.68%	0.03%	0%	0%	0.29%	41,478.33	41,718.56	

Rows per page 25 (1 of 1) Download

Feature distribution (9 samples; 476 genes)

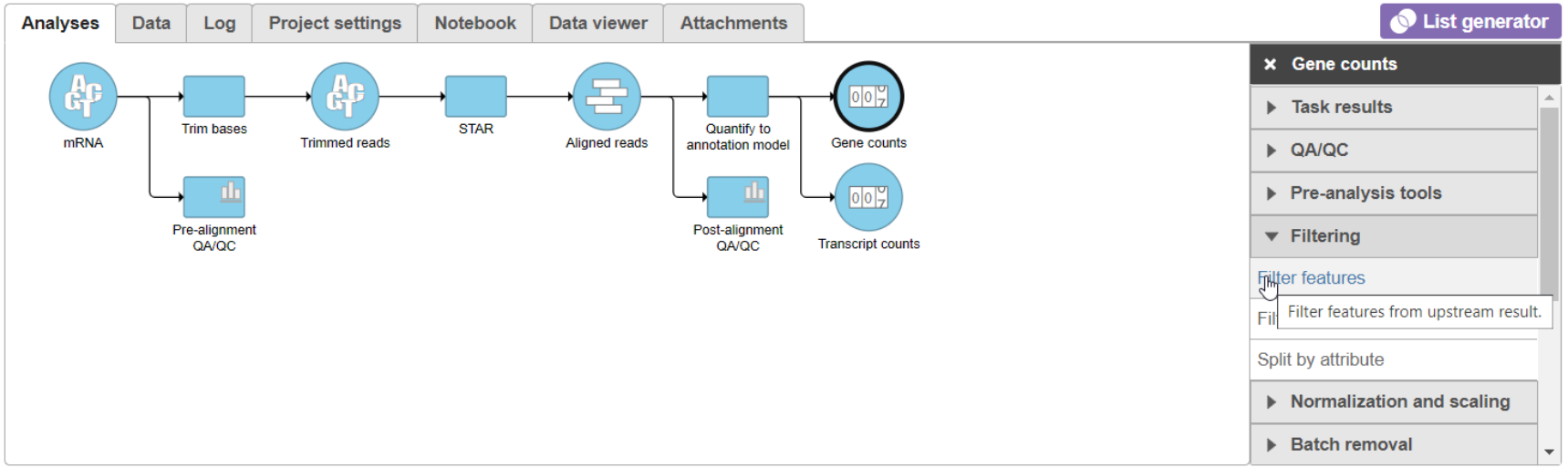
Optional columns

Sample name	Min	2nd min	Max	Mean	Median	Q1	Q3
SRR592573	0	3.35E-37	3,859.00	140.20	34.00	4.00	141.00
SRR592574	0	1.08E-31	5,940.00	209.55	49.00	5.03	212.00
SRR592575	0	3.12E-27	6,643.00	268.33	65.10	8.00	274.00
SRR592576	0	0.25	5,546.00	274.44	94.25	17.00	309.94
SRR592577	0	1.63E-19	4,521.00	259.09	84.00	14.30	288.54
SRR592578	0	3.29E-17	6,007.00	303.88	96.94	16.93	339.00
SRR592579	0	5.29E-29	2,985.00	123.85	37.50	5.92	136.50
SRR592580	0	3.73E-35	5,604.00	254.90	80.46	9.00	268.50
SRR592581	0	7.63E-19	3,818.00	174.85	67.50	9.00	210.00
All samples	0	3.35E-37	6,643.00	223.23	64.00	9.00	234.00

Rows per page 25 (1 of 1) Download

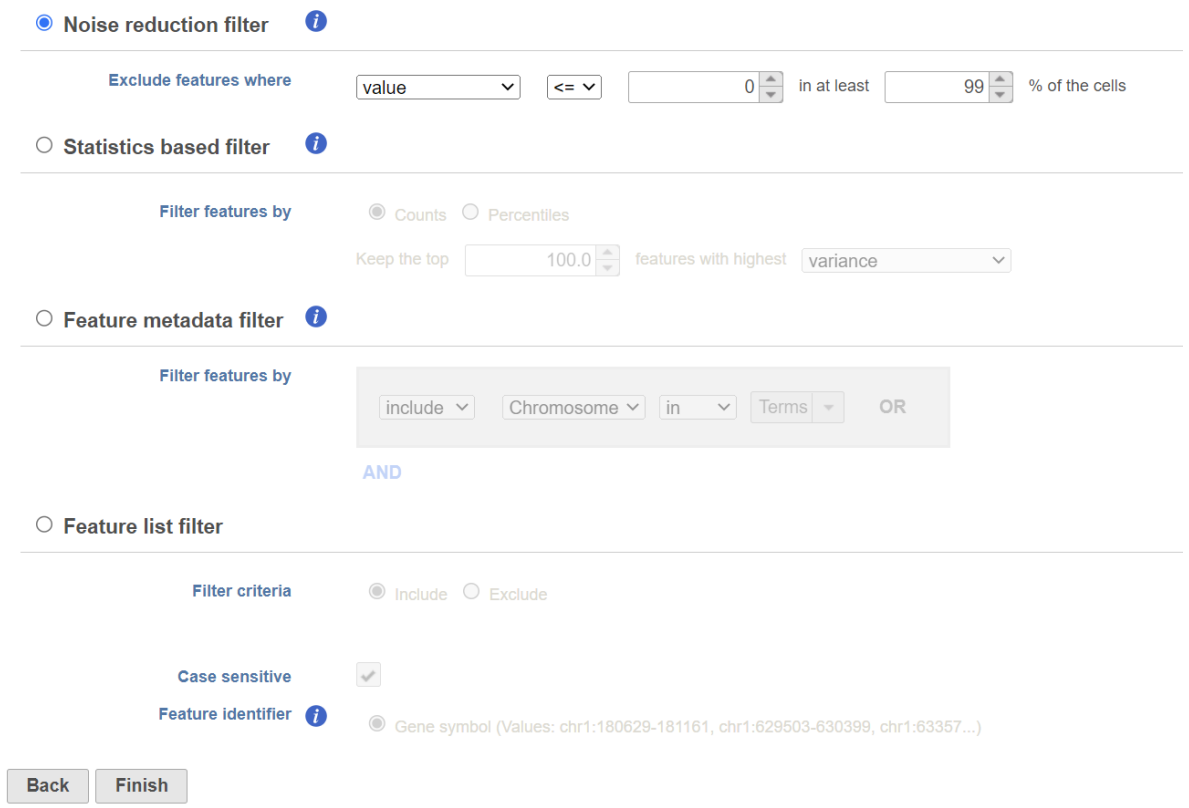
Filtering Features

- Click the project name to return to the Analyses tab
- Click **Gene counts** node
- Click **Filtering** in the task menu
- Click **Filter features**



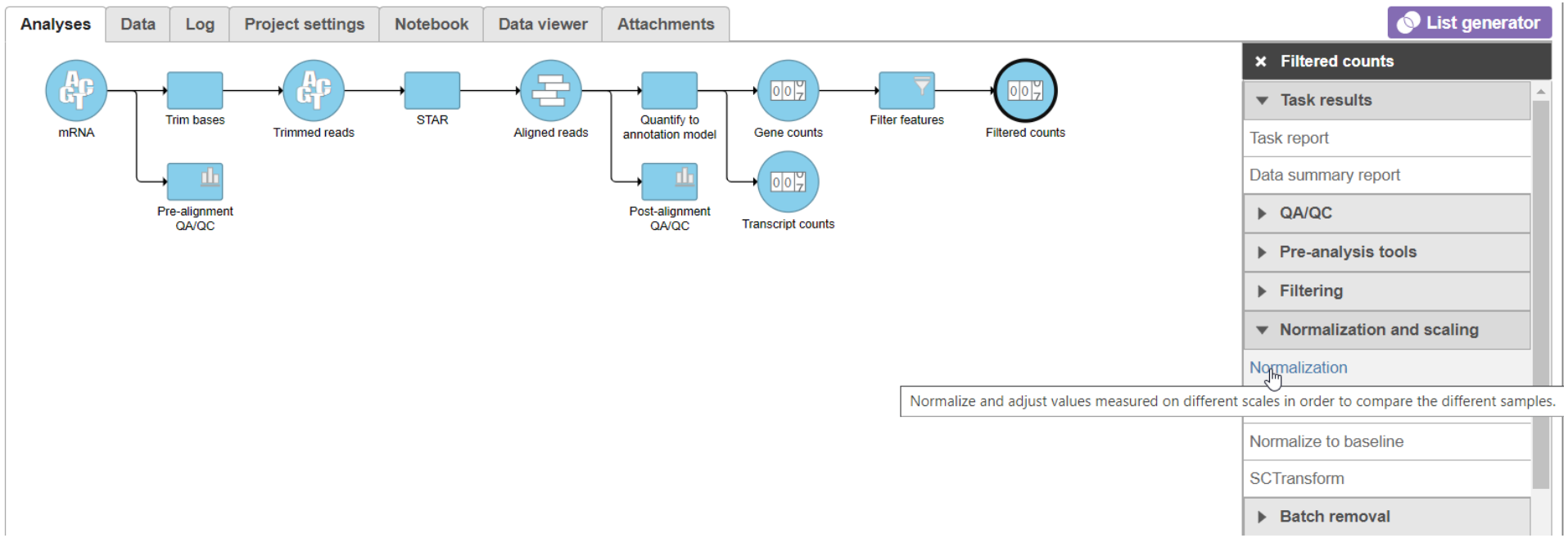
Filter Features

- Click the **Noise reduction filter** checkbox
- Set the filter to **Exclude features where value <= 0 in 99% of cells** using the drop-down menus and text boxes
- Click **Finish** to apply the filter



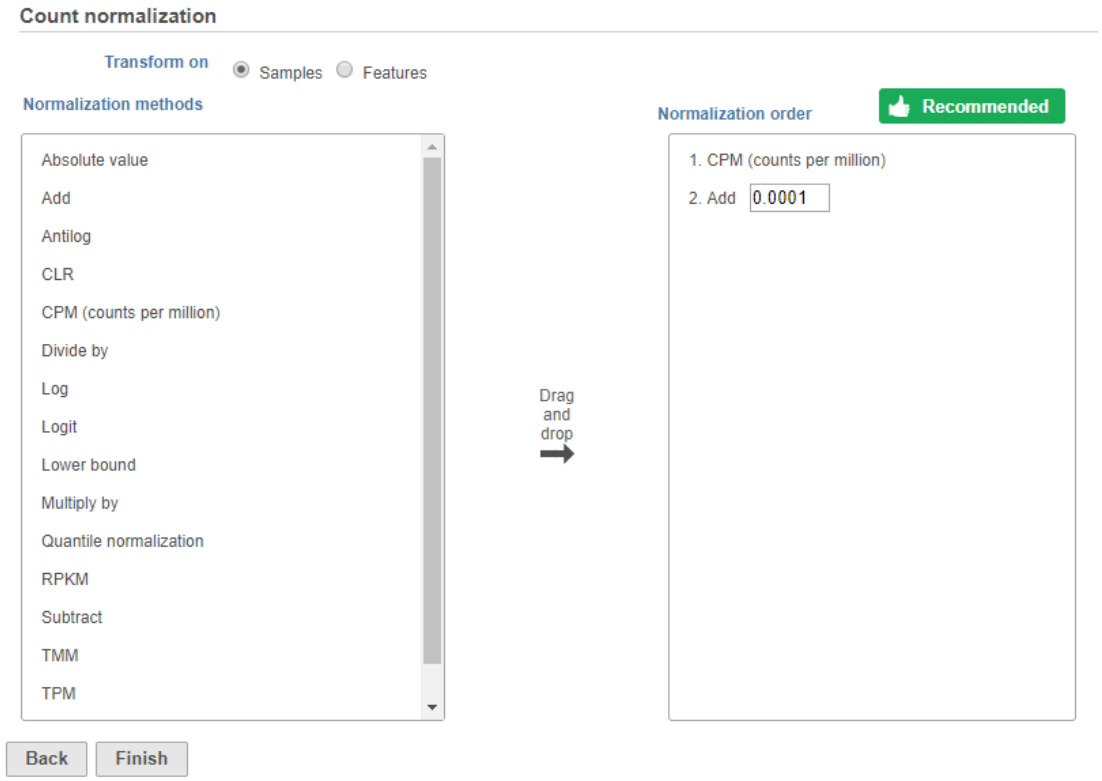
Normalization

- Click **Filtered counts** data node
- Click **Normalization and scaling** in the task menu
- Click **Normalization**



Normalization

- Select **Recommended** if no preferred normalization methods
- Click **Finish**



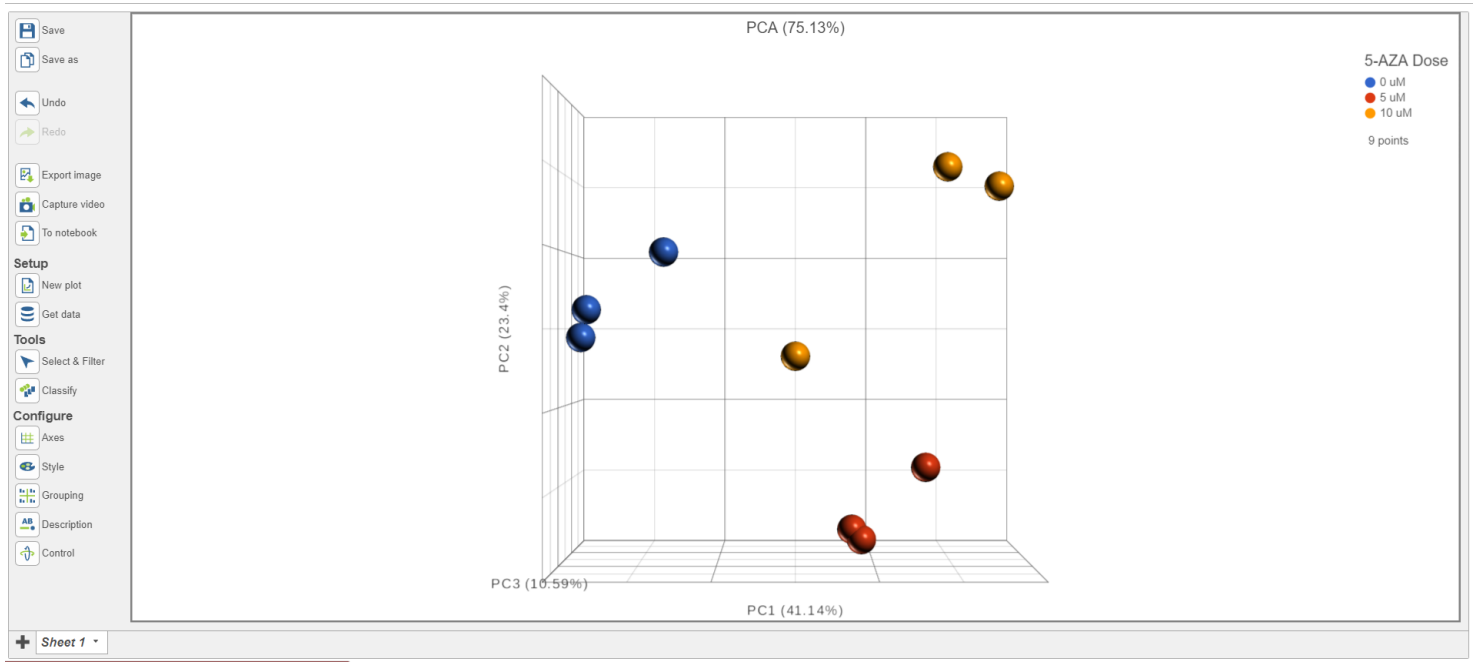
Principal Components Analysis (PCA)

The PCA scatter plot allows to visualize similarities and differences between the samples in a data set.

- Click **Normalized counts** data node
- Click **Exploratory analysis** in the task menu
- Click **PCA**
- Click **Finish** to run with the default options

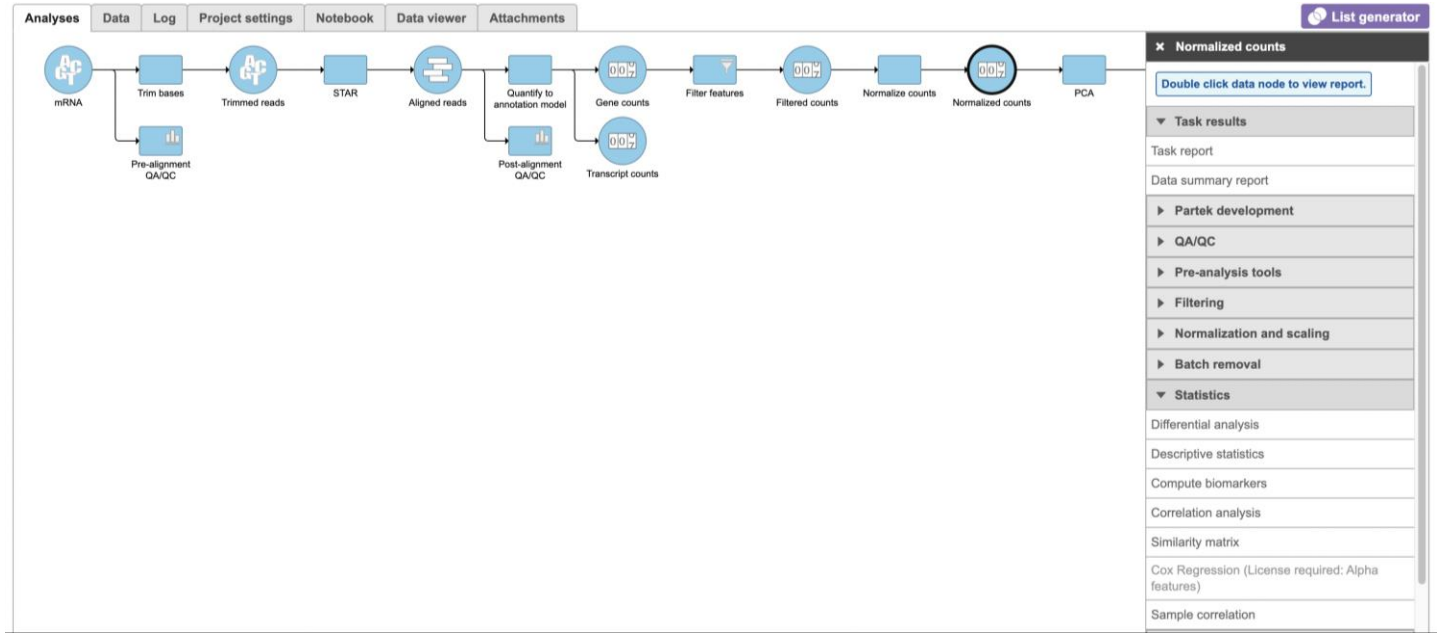
PCA

- Double click **PCA** data node to open the PCA scatter plot
- Click **Style** under Configure
- Set the **Color by** drop-down to **5-AZA Dose**



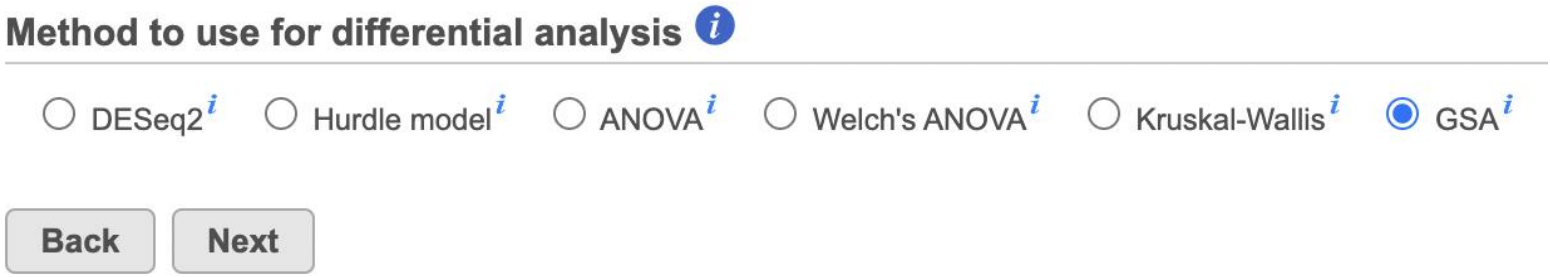
Differential Analysis

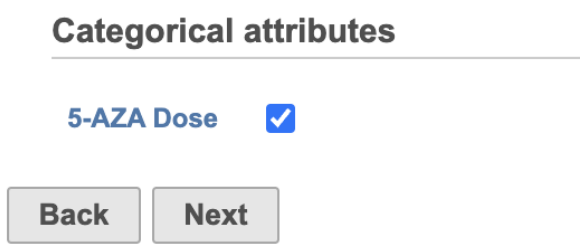
- Click **Normalized counts** node
- Click **Statistics** in the task menu
- Click **Differential analysis** in the task menu



Differential Analysis

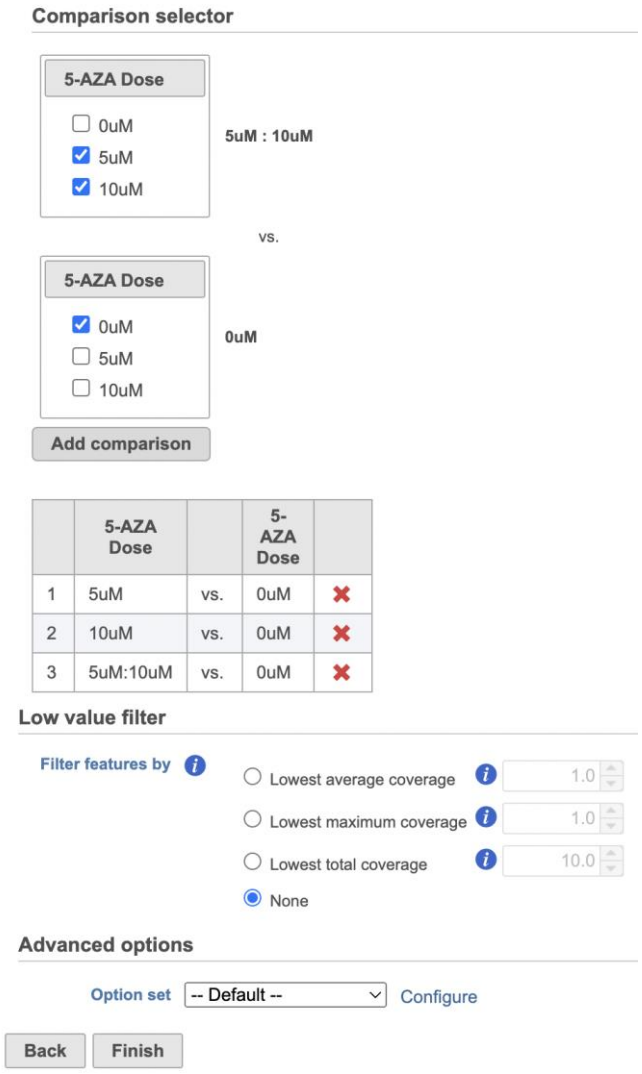
- Select the appropriate differential analysis method
- In this tutorial we are going to use **GSA**
- Click **Next**
- Select the attribute **5-AZA Dose**
- Click **Next**





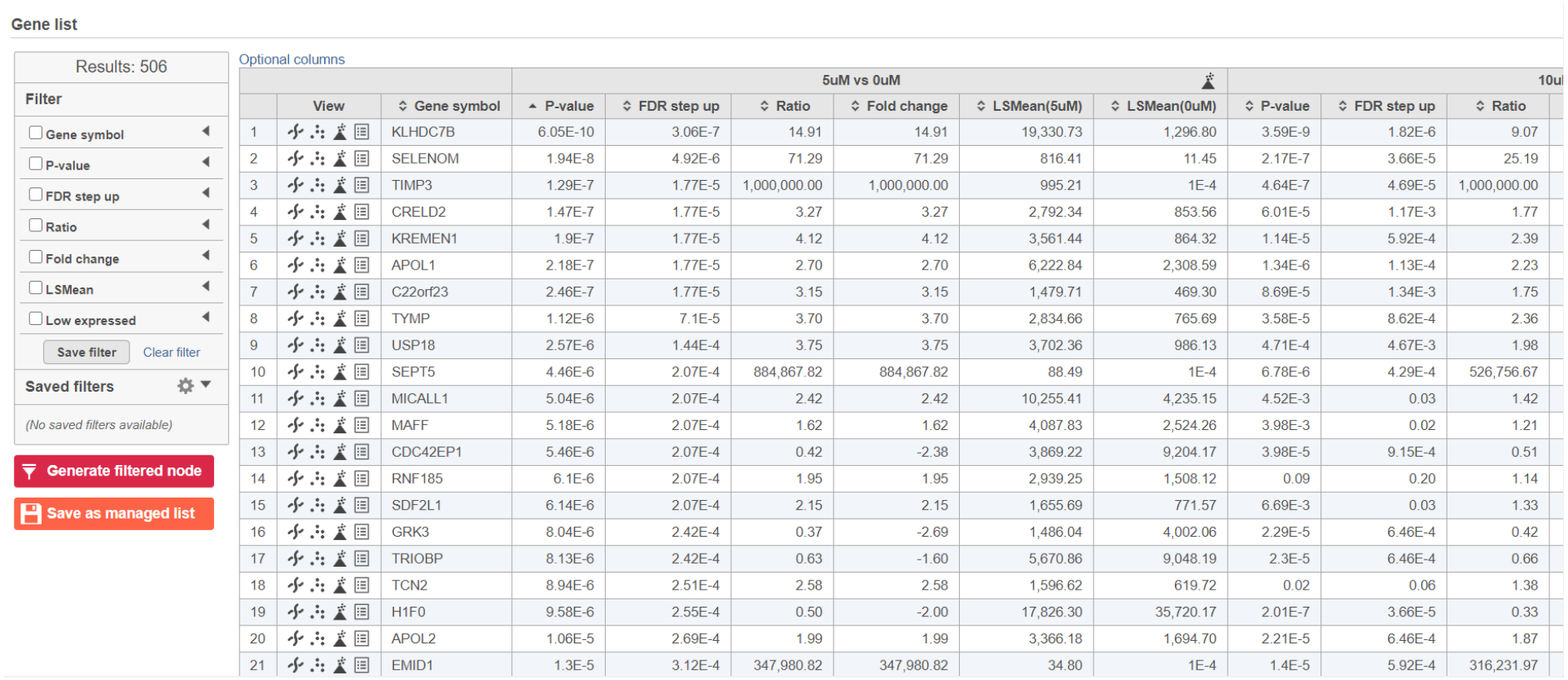
Differential Analysis

- Select **5µM** in the upper box
- Select **0µM** in the lower box
- Click **Add comparison**
- Repeat to create comparisons for **10µM vs. 0µM** and **5µM:10µM vs. 0µM**
- Click **Finish**



Differential Analysis

- Double click **GSA** data node to open the task report



Gene list

Results: 506

Filter

- Gene symbol
- P-value
- FDR step up
- Ratio
- Fold change
- LSMean
- Low expressed

Save filter | Clear filter

Saved filters

(No saved filters available)

Generate filtered node

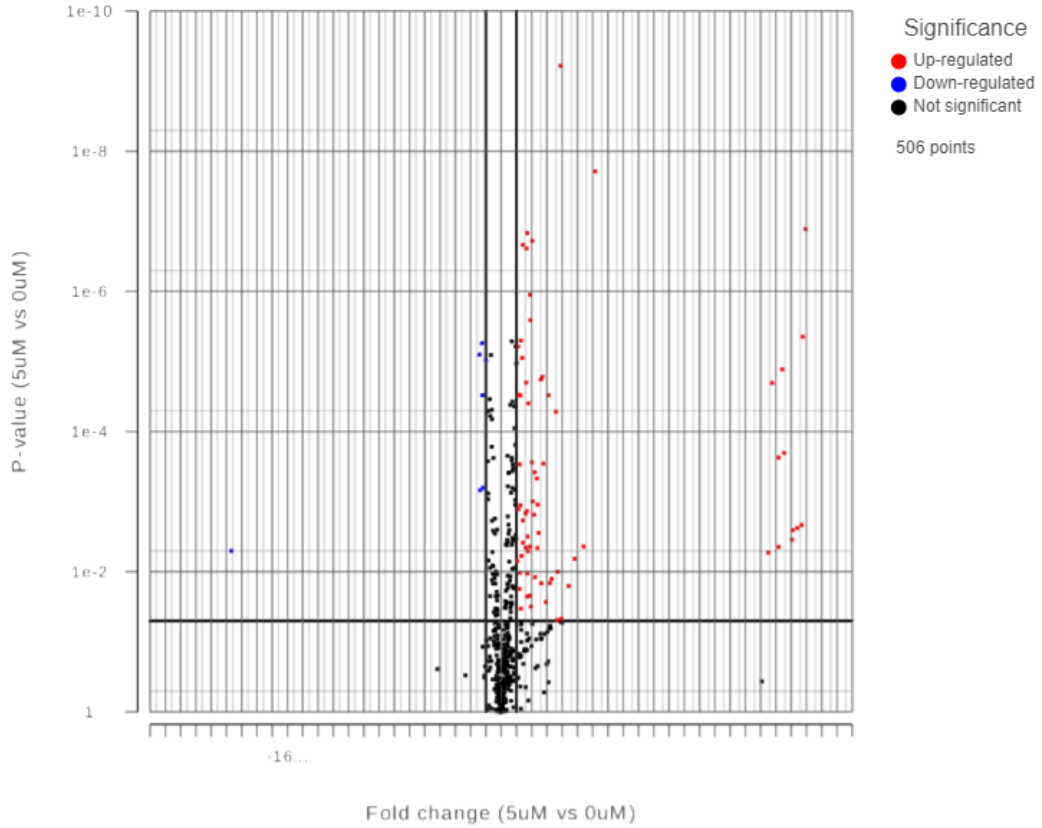
Save as managed list

Optional columns

			5uM vs 0uM						10u		
	View	Gene symbol	P-value	FDR step up	Ratio	Fold change	LSMean(5uM)	LSMean(0uM)	P-value	FDR step up	Ratio
1		KLHDC7B	6.05E-10	3.06E-7	14.91	14.91	19,330.73	1,296.80	3.59E-9	1.82E-6	9.07
2		SELENOM	1.94E-8	4.92E-6	71.29	71.29	816.41	11.45	2.17E-7	3.66E-5	25.19
3		TIMP3	1.29E-7	1.77E-5	1,000,000.00	1,000,000.00	995.21	1E-4	4.64E-7	4.69E-5	1,000,000.00
4		CRELD2	1.47E-7	1.77E-5	3.27	3.27	2,792.34	853.56	6.01E-5	1.17E-3	1.77
5		KREMEN1	1.9E-7	1.77E-5	4.12	4.12	3,561.44	864.32	1.14E-5	5.92E-4	2.39
6		APOL1	2.18E-7	1.77E-5	2.70	2.70	6,222.84	2,308.59	1.34E-6	1.13E-4	2.23
7		C22orf23	2.46E-7	1.77E-5	3.15	3.15	1,479.71	469.30	8.69E-5	1.34E-3	1.75
8		TYMP	1.12E-6	7.1E-5	3.70	3.70	2,834.66	765.69	3.58E-5	8.62E-4	2.36
9		USP18	2.57E-6	1.44E-4	3.75	3.75	3,702.36	986.13	4.71E-4	4.67E-3	1.98
10		SEPT5	4.46E-6	2.07E-4	884,867.82	884,867.82	88.49	1E-4	6.78E-6	4.29E-4	526,756.67
11		MICALL1	5.04E-6	2.07E-4	2.42	2.42	10,255.41	4,235.15	4.52E-3	0.03	1.42
12		MAFF	5.18E-6	2.07E-4	1.62	1.62	4,087.83	2,524.26	3.98E-3	0.02	1.21
13		CDC42EP1	5.46E-6	2.07E-4	0.42	-2.38	3,869.22	9,204.17	3.98E-5	9.15E-4	0.51
14		RNF185	6.1E-6	2.07E-4	1.95	1.95	2,939.25	1,508.12	0.09	0.20	1.14
15		SDF2L1	6.14E-6	2.07E-4	2.15	2.15	1,655.69	771.57	6.69E-3	0.03	1.33
16		GRK3	8.04E-6	2.42E-4	0.37	-2.69	1,486.04	4,002.06	2.29E-5	6.46E-4	0.42
17		TRIOBP	8.13E-6	2.42E-4	0.63	-1.60	5,670.86	9,048.19	2.3E-5	6.46E-4	0.66
18		TCN2	8.94E-6	2.51E-4	2.58	2.58	1,596.62	619.72	0.02	0.06	1.38
19		H1F0	9.58E-6	2.55E-4	0.50	-2.00	17,826.30	35,720.17	2.01E-7	3.66E-5	0.33
20		APOL2	1.06E-5	2.69E-4	1.99	1.99	3,366.18	1,694.70	2.21E-5	6.46E-4	1.87
21		EMID1	1.3E-5	3.12E-4	347,980.82	347,980.82	34.80	1E-4	1.4E-5	5.92E-4	316,231.97

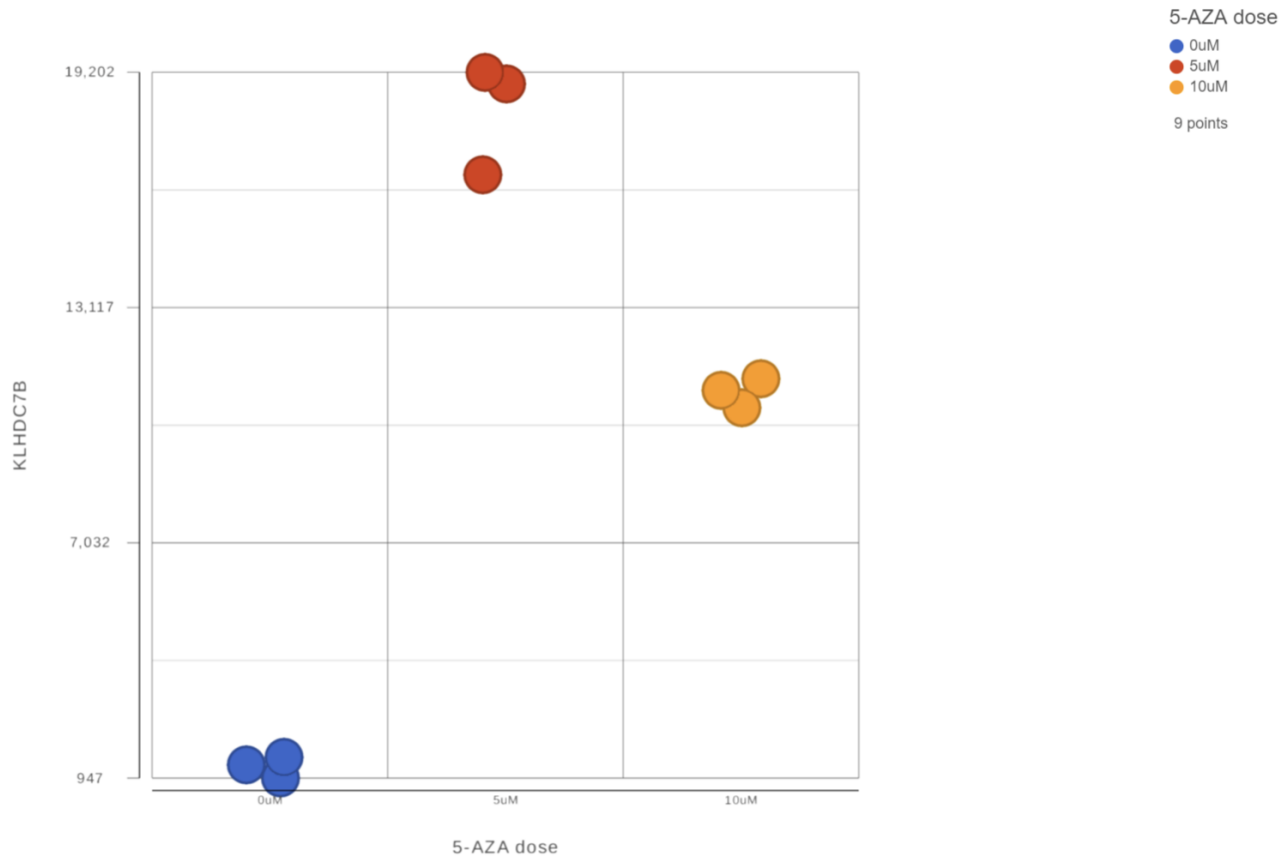
Volcano Plot for All Genes

- Click  next to the 5uM vs. 0uM comparison



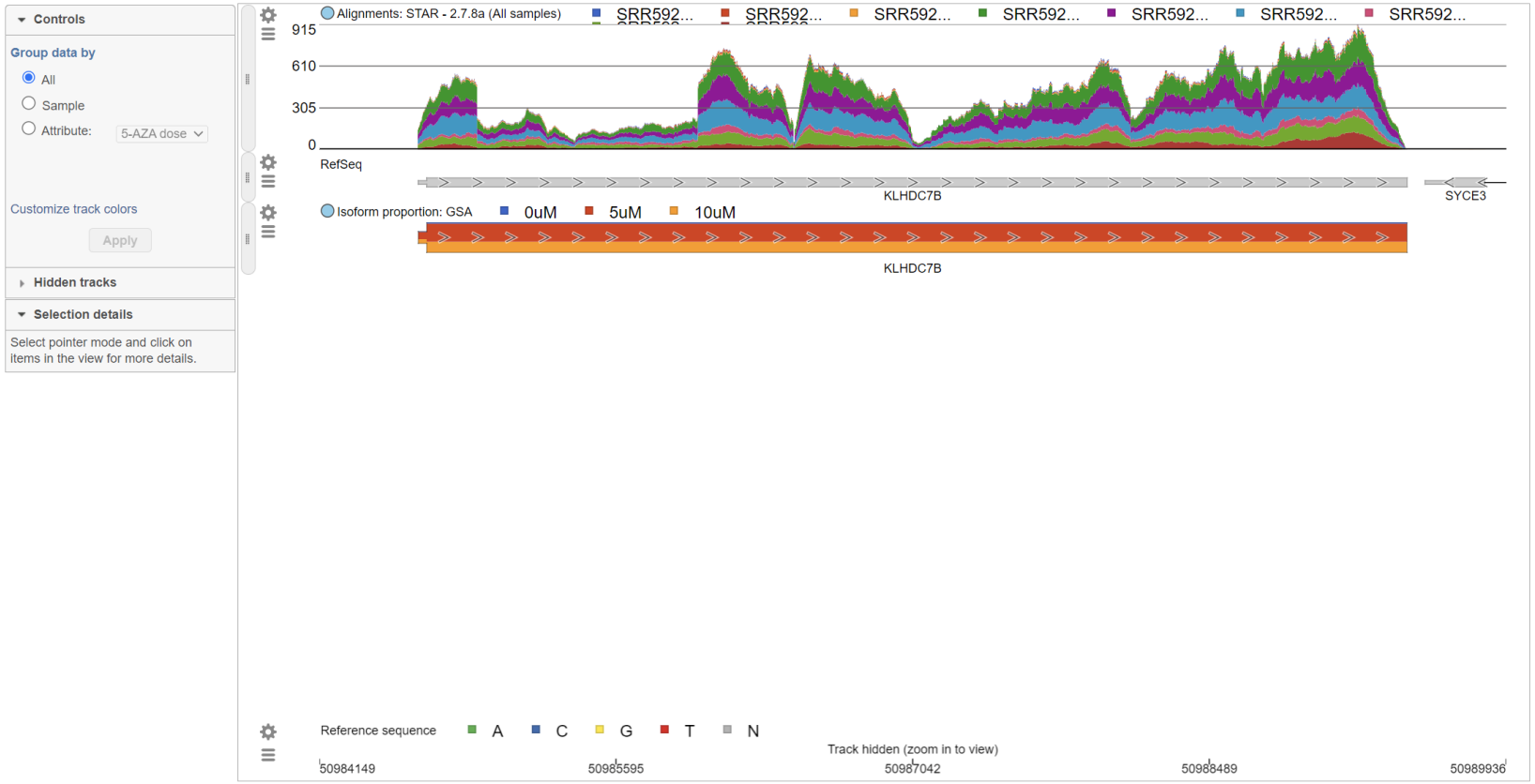
Dot Plot for One Gene

- Select  next to a gene symbol to open a dot plot for the gene



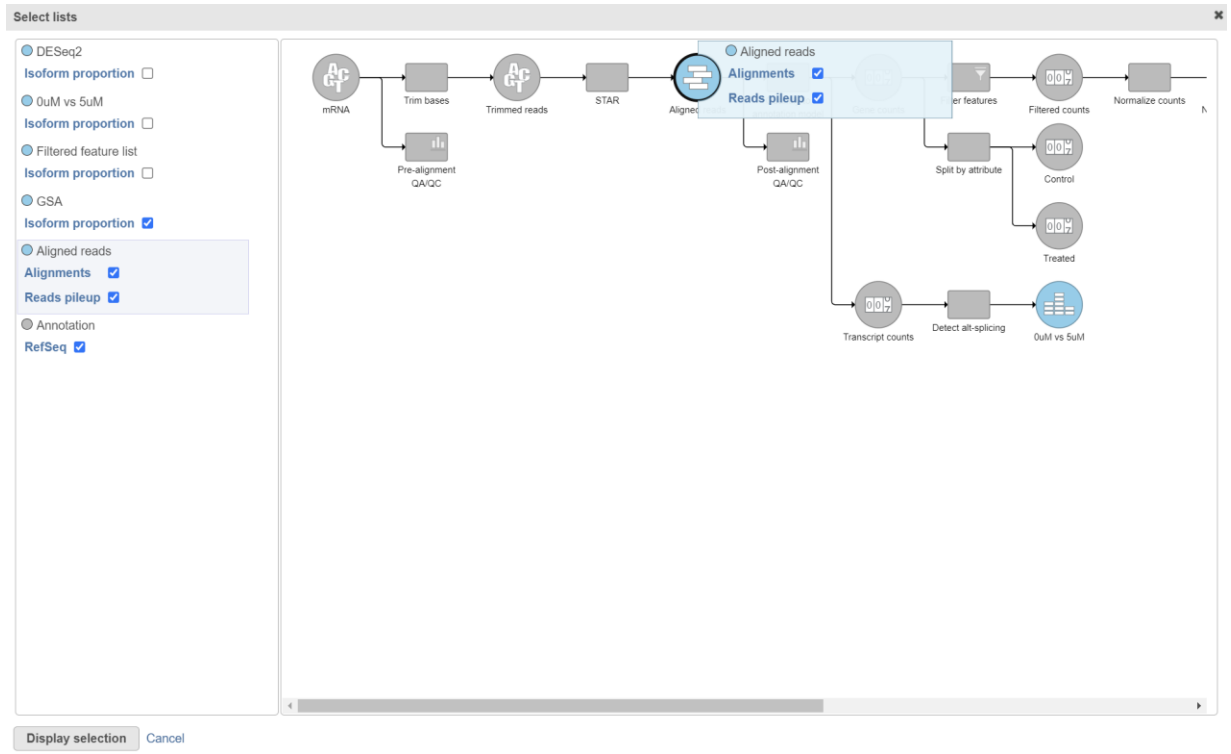
Browse Gene in Chromosome View

- Click  next to a gene symbol to open Chromosome view



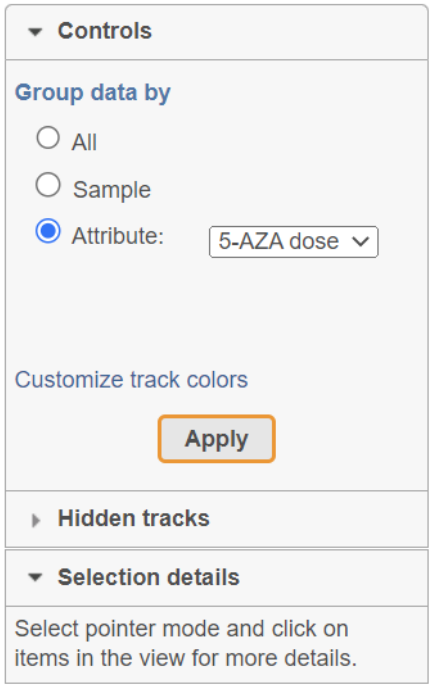
Browse Gene in Chromosome View

- Click **Select tracks**
- Choose **Aligned Reads** data node
- Click **Reads pileup** under Aligned reads
- Click **Display selection** to make the change



Browse Gene in Chromosome View

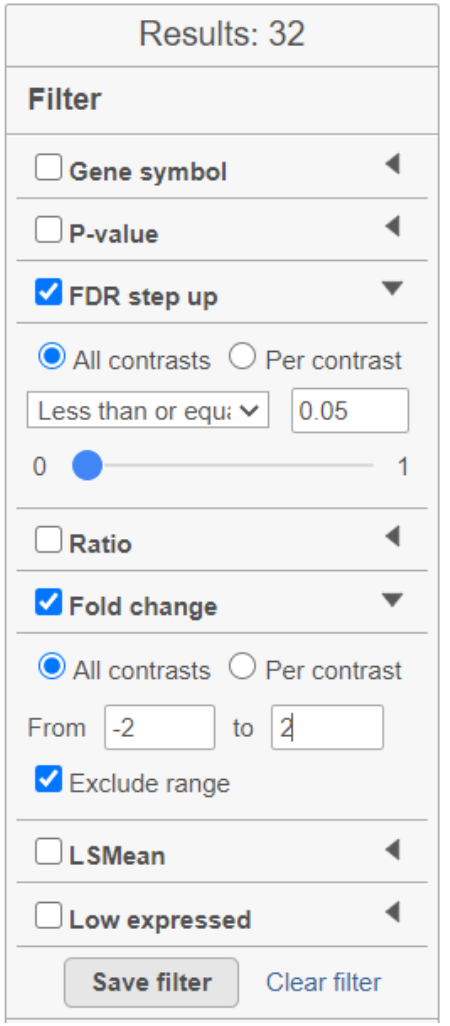
- Group data by **Attribute** and choose **5-AZA**
- Click **Apply**





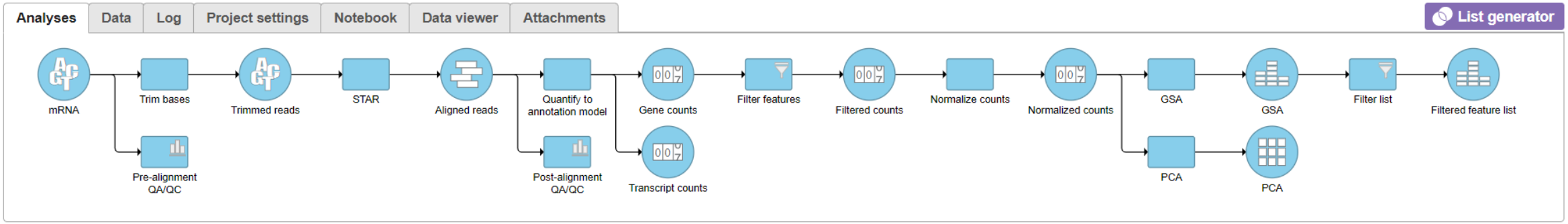
Identify DEG

- Switch to the browser tab showing the GSA report
- Click **FDR step up**
- Click the triangle next to FDR step up
- Set the cutoff value to **0.05**
- Hit Enter
- Click **Fold-change**
- Click the triangle next to Fold-change
- Set to **From -2 to 2** with **Exclude range** selected
- Hit Enter



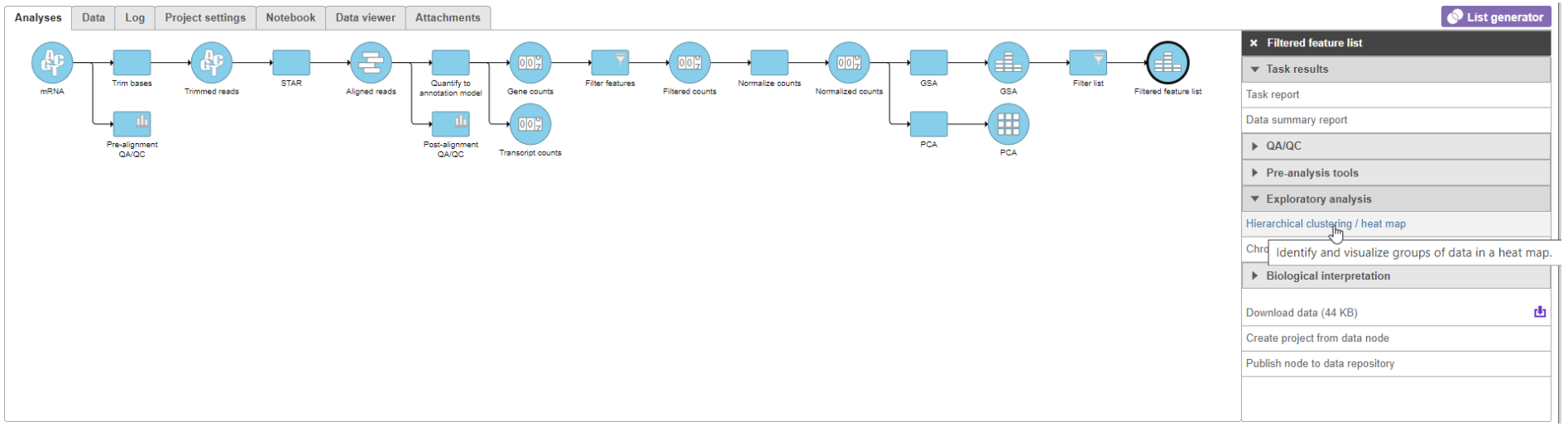
Identify DEG

- Click **Generate filtered node** to create a data node with only the genes that pass the filter



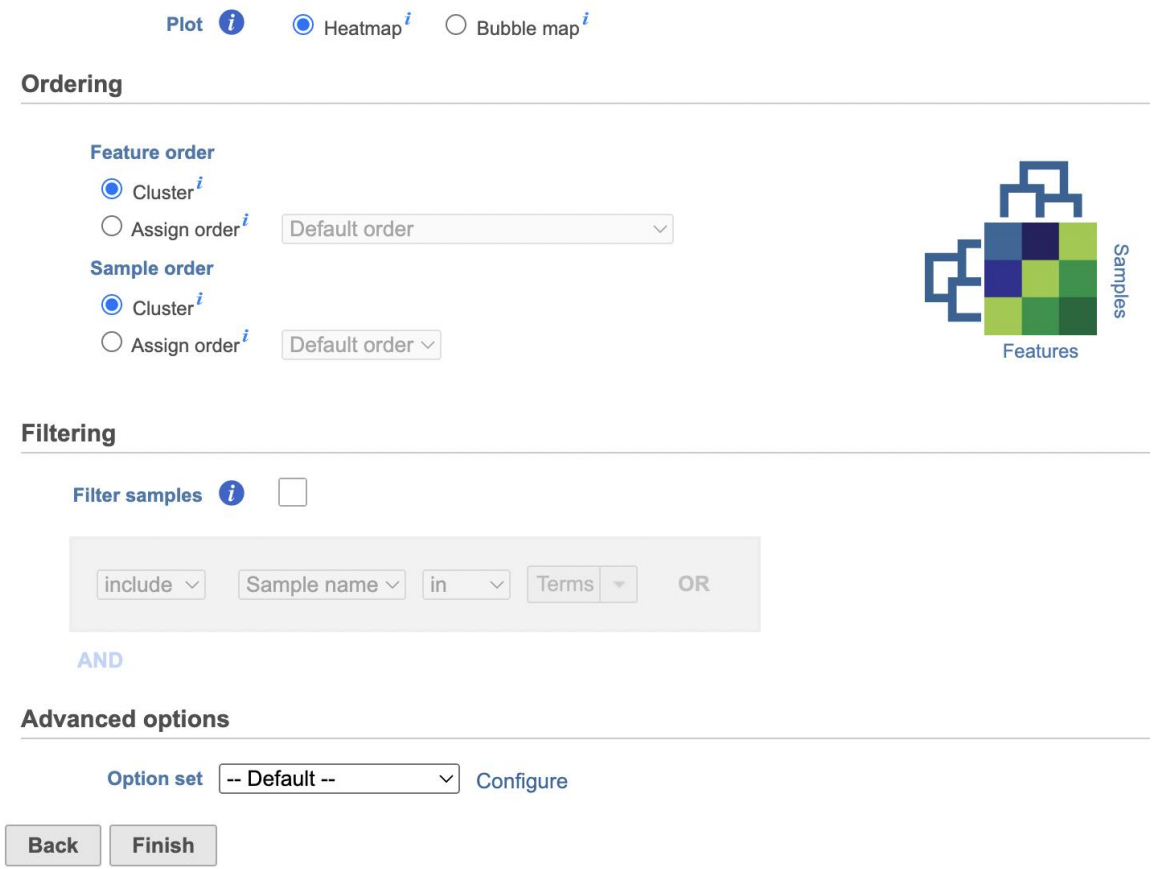
Hierarchical Clustering

- Click **Feature list** data node
- Click **Exploratory analysis** in the task menu
- Click **Hierarchical clustering / heatmap**



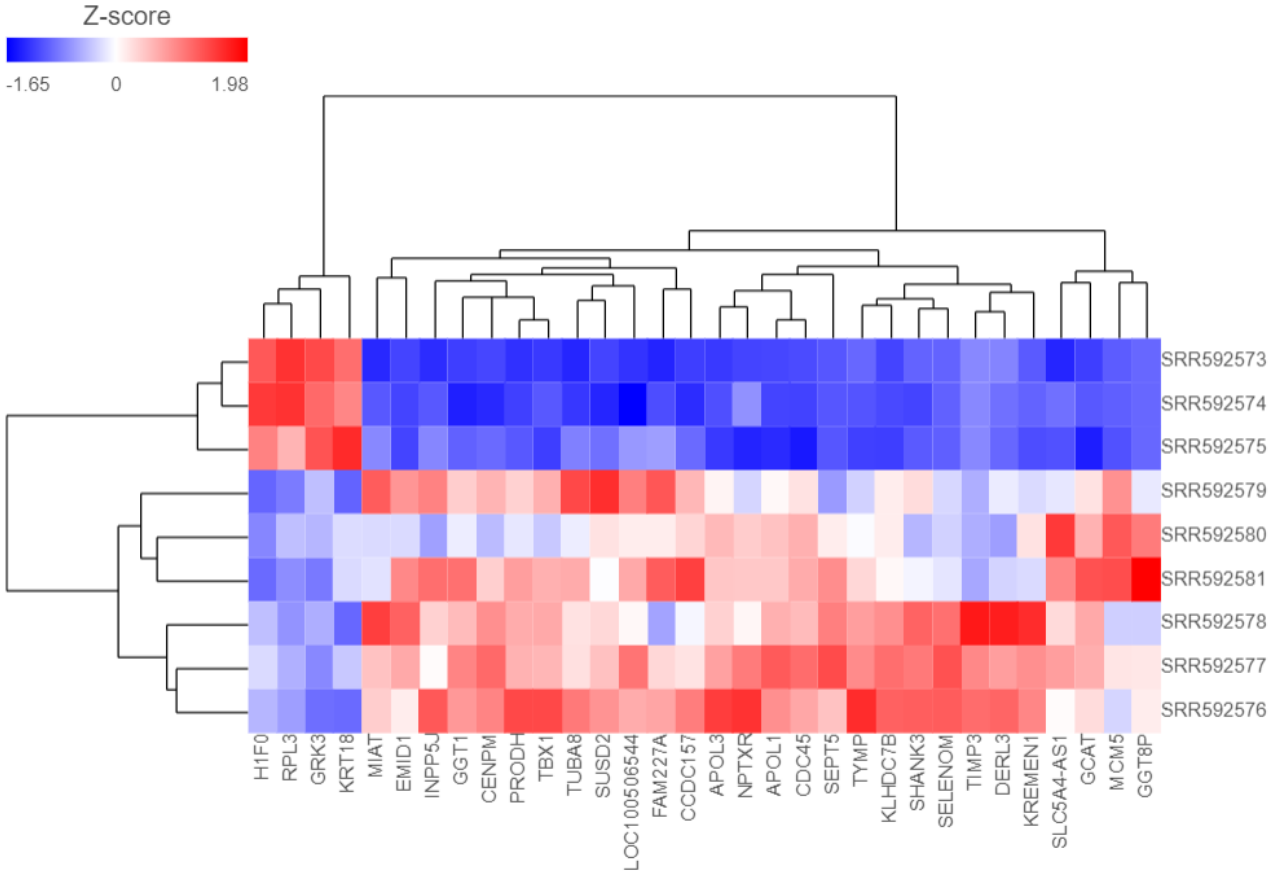
Hierarchical Clustering & Heatmap

- Click **Finish** to run with default settings



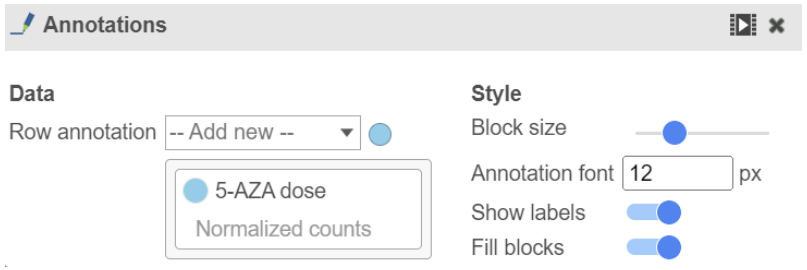
Hierarchical Clustering & Heatmap

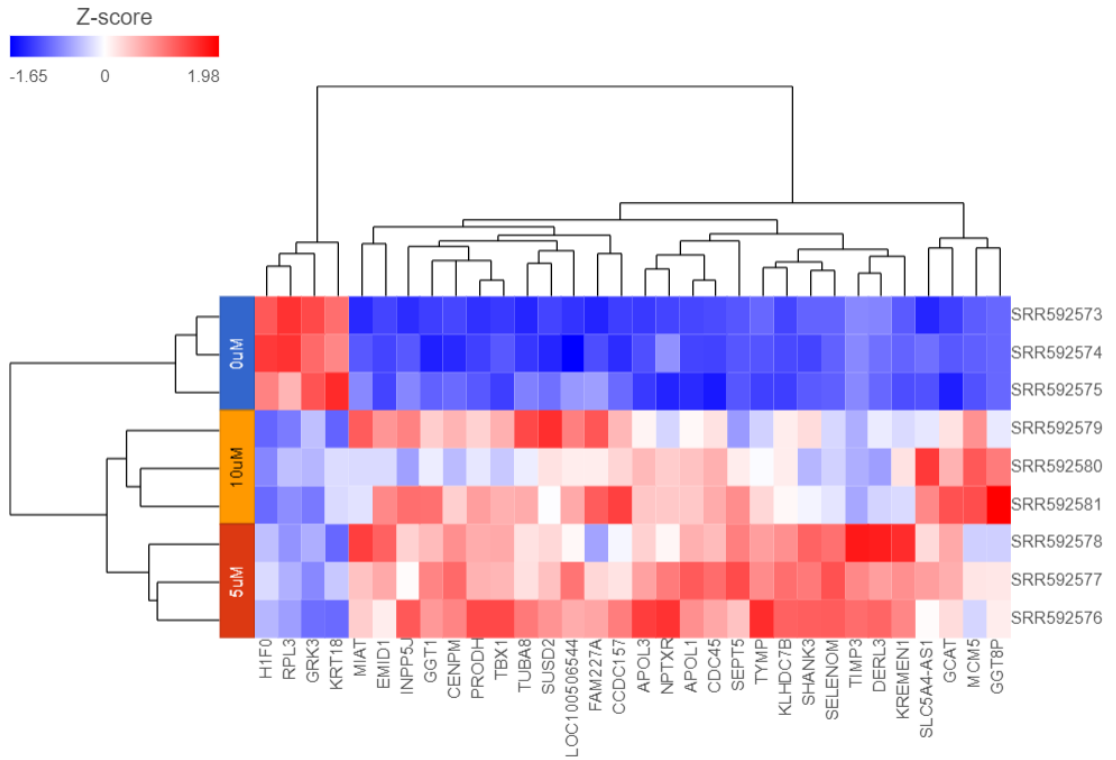
- Double-click the **Hierarchical clustering / heatmap** task node



Hierarchical Clustering & Heatmap

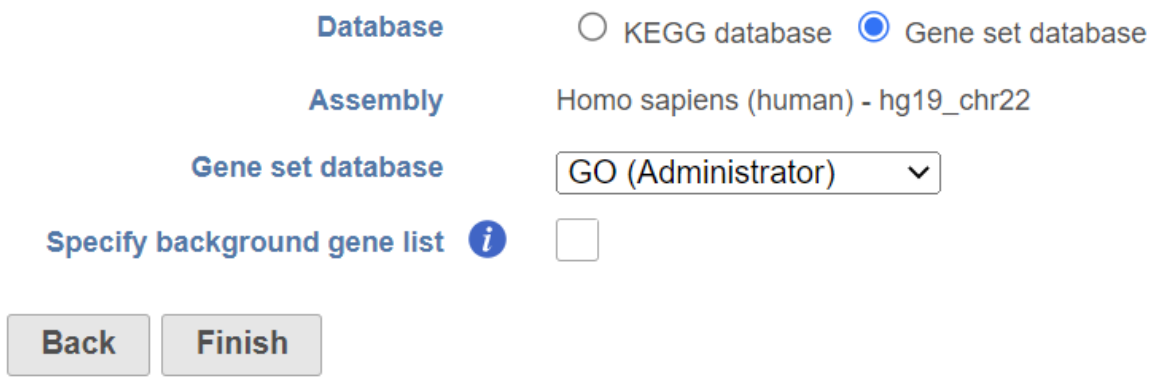
- Expand the **Annotations** card
- Set the Row annotation to **5-AZA Dose**





Enrichment Analysis

- Click the project name to return to the Analyses tab
- Click **Filtered feature list** data node
- Click **Biological interpretation** in the task menu
- Click **Gene set enrichment**
- Select **Gene set database** to perform GO enrichment analysis



Enrichment Analysis

- Double-click **Gene set enrichment** task node to open the task report

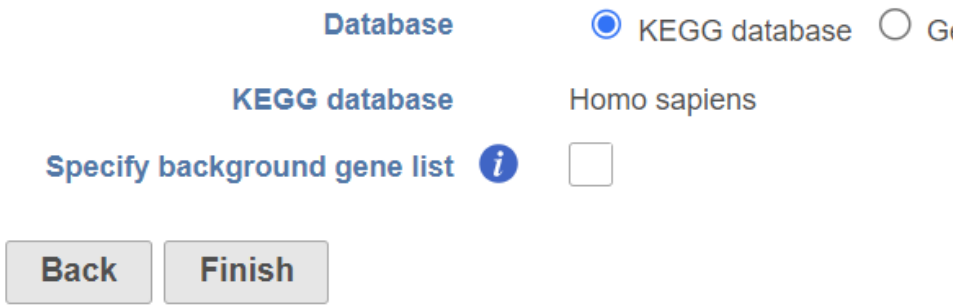
Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
GO:0031644	regulation of nervous system process	6.28	1.87E-3	3	2	
GO:0043648	dicarboxylic acid metabolic process	5.63	3.6E-3	2	0	
GO:0097110	scaffold protein binding	5.63	3.6E-3	2	0	
GO:1901606	alpha-amino acid catabolic process	5.63	3.6E-3	2	0	
GO:0000724	double-strand break repair via homologous recombination	5.63	3.6E-3	2	0	
GO:0000725	recombinational repair	5.63	3.6E-3	2	0	
GO:0000727	double-strand break repair via break-induced replication	5.63	3.6E-3	2	0	
GO:0003688	DNA replication origin binding	5.63	3.6E-3	2	0	
GO:0006260	DNA replication	5.63	3.6E-3	2	0	
GO:0006270	DNA replication initiation	5.63	3.6E-3	2	0	
GO:0006536	glutamate metabolic process	5.63	3.6E-3	2	0	
GO:0071824	protein-DNA complex subunit organization	5.11	6.03E-3	3	4	
GO:1901605	alpha-amino acid metabolic process	5.11	6.03E-3	3	4	
GO:0065004	protein-DNA complex assembly	5.11	6.03E-3	3	4	
GO:0006520	cellular amino acid metabolic process	4.68	9.25E-3	3	5	
GO:0017124	SH3 domain binding	4.57	0.01	2	1	
GO:0044843	cell cycle G1/S phase transition	4.57	0.01	2	1	
GO:0000082	G1/S transition of mitotic cell cycle	4.57	0.01	2	1	
GO:0002020	protease binding	4.57	0.01	2	1	
GO:0009064	glutamine family amino acid metabolic process	4.57	0.01	2	1	
GO:0044282	small molecule catabolic process	4.42	0.01	4	12	
GO:0042177	negative regulation of protein catabolic process	3.91	0.02	2	2	
GO:0099601	regulation of neurotransmitter receptor activity	3.91	0.02	2	2	
GO:0051703	intraspecies interaction between organisms	3.91	0.02	2	2	
GO:0003697	single-stranded DNA binding	3.91	0.02	2	2	

Rows per page 25 (1 of 51) Download

Operators such as >= or <= can be used in the numeric columns' filters.

KEGG Enrichment Analysis

- Click the project name to return to the Analyses tab
- Click **Filtered feature list** data node
- Click **Biological interpretation** in the task menu
- Click **Gene set enrichment**
- Select **KEGG database**



KEGG Enrichment Analysis

- Double-click **Pathway enrichment** task node to open the task report

Gene set	Description	Enrichment score	P-value	Genes in list	Genes not in list	
path:hsa04724	Glutamatergic synapse	3.18	0.04	2	2	
path:hsa05130	Pathogenic Escherichia coli infection	2.73	0.07	2	3	
path:hsa03030	DNA replication	2.40	0.09	1	0	
path:hsa04740	Olfactory transduction	2.40	0.09	1	0	
path:hsa05143	African trypanosomiasis	2.40	0.09	1	0	
path:hsa00240	Pyrimidine metabolism	2.40	0.09	1	0	
path:hsa00260	Glycine, serine and threonine metabolism	2.40	0.09	1	0	
path:hsa00330	Arginine and proline metabolism	2.40	0.09	1	0	
path:hsa00430	Taurine and hypotaurine metabolism	2.40	0.09	1	0	
path:hsa04110	Cell cycle	1.86	0.16	2	6	
path:hsa05032	Morphine addiction	1.75	0.17	1	1	
path:hsa00590	Arachidonic acid metabolism	1.75	0.17	1	1	
path:hsa04340	Hedgehog signaling pathway	1.39	0.25	1	2	
path:hsa05219	Bladder cancer	1.39	0.25	1	2	
path:hsa04070	Phosphatidylinositol signaling system	1.15	0.32	1	3	
path:hsa04144	Endocytosis	1.15	0.32	1	3	
path:hsa04145	Phagosome	1.15	0.32	1	3	
path:hsa04540	Gap junction	1.15	0.32	1	3	
path:hsa00480	Glutathione metabolism	1.15	0.32	1	3	
path:hsa00983	Drug metabolism - other enzymes	1.15	0.32	1	3	
path:hsa01100	Metabolic pathways	1.14	0.32	5	37	
path:hsa04915	Estrogen signaling pathway	0.97	0.38	1	4	
path:hsa00562	Inositol phosphate metabolism	0.97	0.38	1	4	
path:hsa04062	Chemokine signaling pathway	0.82	0.44	1	5	
path:hsa04141	Protein processing in endoplasmic reticulum	0.71	0.49	1	6	

Rows per page 25 (1 of 2) Download

Operators such as >= or <= can be used in the numeric columns' filters.