

# From Raw Data to Pathways: Easy Genomics Analysis with Partek Flow

---

Eric Chen  
techsupport@gtbiotech.com.tw  
*Bioinformatic Specialist*  
*GenetechBiotech*

# Who is Partek

## Mission

To empower scientists to make scientific breakthroughs in human genetics, disease relationships, drug discoveries, diagnoses, and disease treatments.



Founded in

1993

for data mining and artificial  
intelligence

Over

8,500

peer-reviewed citations

More than

40,000

researcher questions answered

Customers in over

40

countries

# Partek Flow: Start-to-Finish Bioinformatics Solution



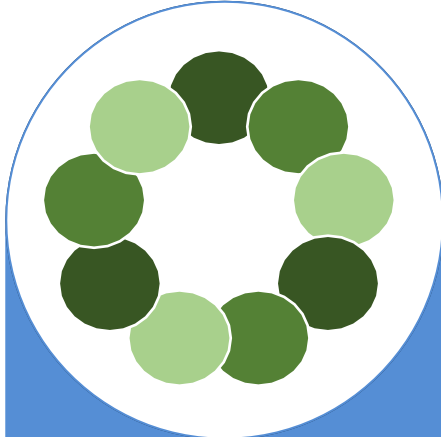
**Interactive  
Visualizations**



**Collaborative  
Web  
Environment**



**Powerful  
Statistics**



**Comprehensive  
Application  
Support**

# Publicly Available Statistical Algorithms and Tools

## Alignment

Bowtie	Bowtie
BWA	GSNAP
Isaac	STAR
TopHat	HISAT
TMAP	



## QA/QC reports

Pre-alignment  
Post-alignment  
ERCC spike-in  
Single cell quality



## Variant calling

Samtools	FreeBayes
LoFreq	Strelka
CNVkit	GATK



## Differential analysis

Limma	Negative binomial
DESeq2	Non-parametric ANOVA
Poisson	



## Clustering

Hierarchical  
K-means  
Graph-based



## Variant annotation

Snpeff	VEP
dbSNP	Custom databases



## Metagenomics

Kraken  
Alpha and beta diversity  
Quantification at taxonomic levels  
Differential analysis at taxonomic levels



## Data exploration

PCA	Heat map
t-SNE	Violin plot
Dot plot	Histograms
Box plot	Chromosome view
Pathway	2D & 3D Scatter Plot
Bar chart	Pie chart
Bubble map	UMAP



## Peak calling

MACS2	Motif detection
TSS plot	



## Quantification

Partek E/M	Cufflinks
HTSeq	



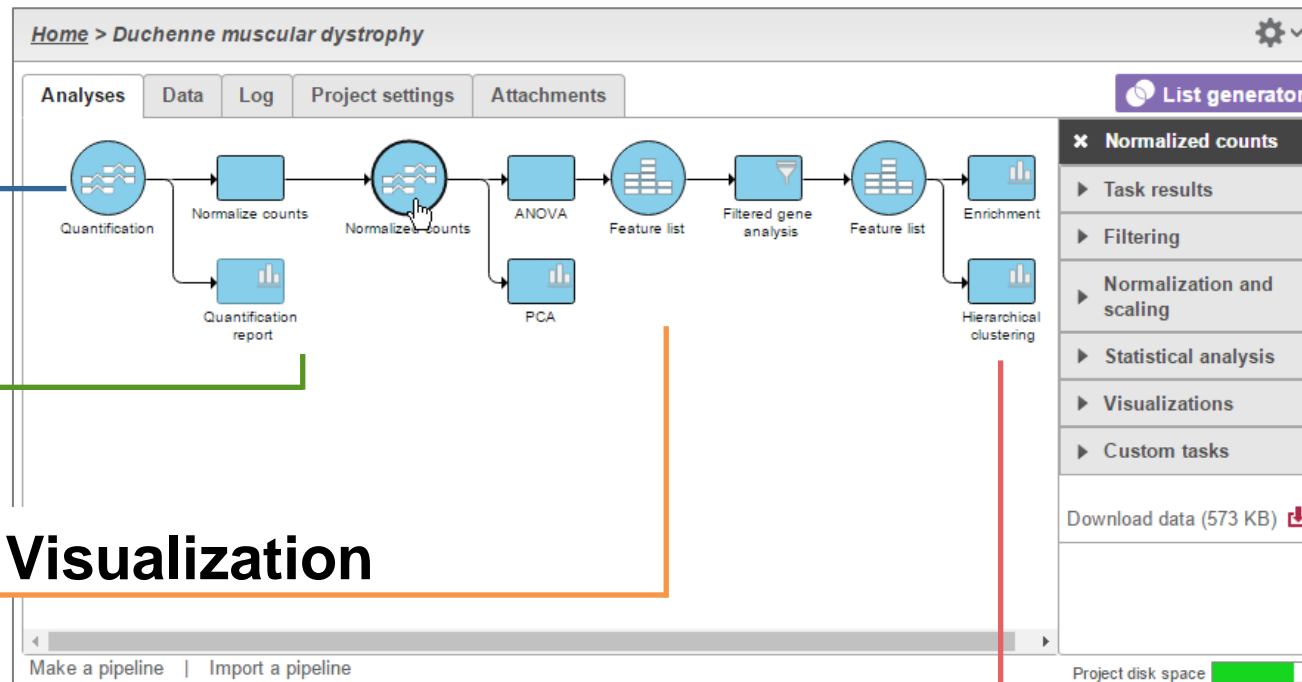
# Visual Analysis Process

## 1 Import Data

## 2 QA/QC

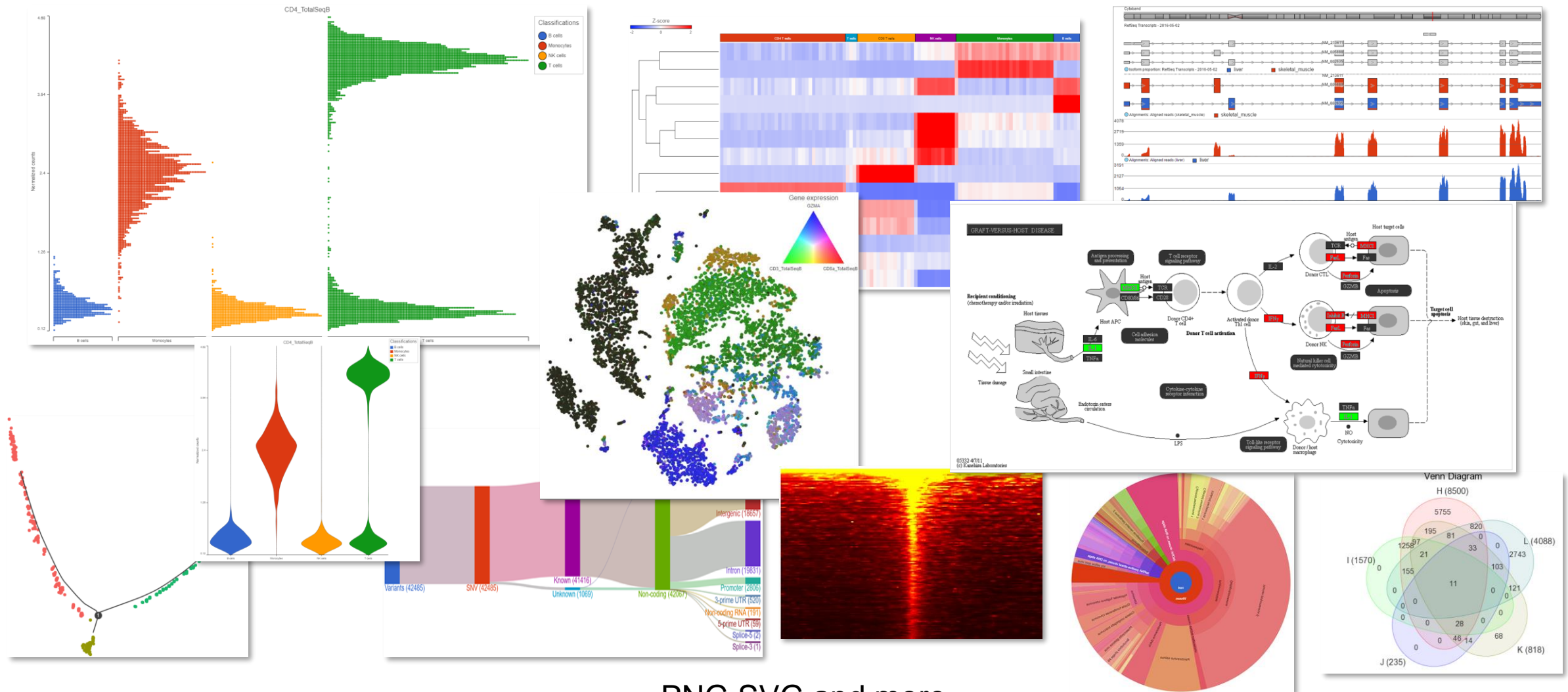
## 3 Powerful Statistics & Visualization

## 4 Biological Interpretation





# Compelling and Publishable Visualizations






PNG SVG and more

# Summary Report

- Who
- When
- What
- How long
- How much





**▼ Sample data**

 Paul Fullerton  28 Aug 2018, 12:24 PM CDT  7.97 GB

[Show/hide details](#)

---





**▼ Trim bases**

**Task** Trim bases  Partek support  7 Sep 2018, 03:31 PM CDT  00:09:06  34.35 GB

[Show/hide details](#)

---





**▼ Filter samples**

**Task** Filter samples  Partek support  10 Sep 2018, 03:38 PM CDT  00:00:00  8.28 GB

[Show/hide details](#)

---

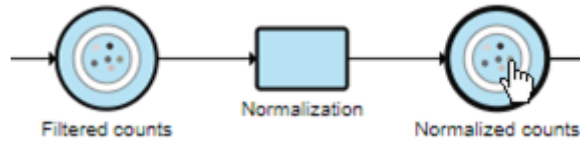
**▼ Align reads**

**Task** BWA - 0.7.15  Partek support  10 Sep 2018, 04:43 PM CDT  01:04:31  5.84 GB

Option	Value
Unaligned reads	SRR2163168.fastq.gz, SRR2163168.index, SRR2181401.fastq.gz, SRR2181401.index
Reference index	mm10
Generate unaligned reads	false
Alignment algorithm	<b>BWA-backtrack</b> (Default: BWA-MEM)
Max edit distance	4.0%
Gap openings	1
Gap extensions	-1
3' deletion buffer	10
Indel ends buffer	5
Enable seeding	false
Max edit distance	2
Gap extension penalty	4

# Export Data

## Choose Any Data



A screenshot of a software interface showing a list of data categories under the heading 'Normalized counts'. The categories are: Task results, Annotation/Metadata, QA/QC, Pre-analysis tools, Filtering, Normalization and scaling, Batch removal, Statistics, Exploratory analysis, Trajectory analysis, Biological interpretation, Classification, Conversion, Pipelines, and Download data (71 MB). A hand cursor is pointing at the 'Download data (71 MB)' option at the bottom.

## Download in Industry Standard Formats

A screenshot of a dialog box titled 'Files will be available to download from task result'. It contains the following options:  
**Export format**  
 Features on columns (.txt)  
 Features on rows (.txt)  
 10X CellRanger HDF5 (.h5)  
**Include content**  
 Annotations  Counts

FASTQ, BAM, TXT, and more

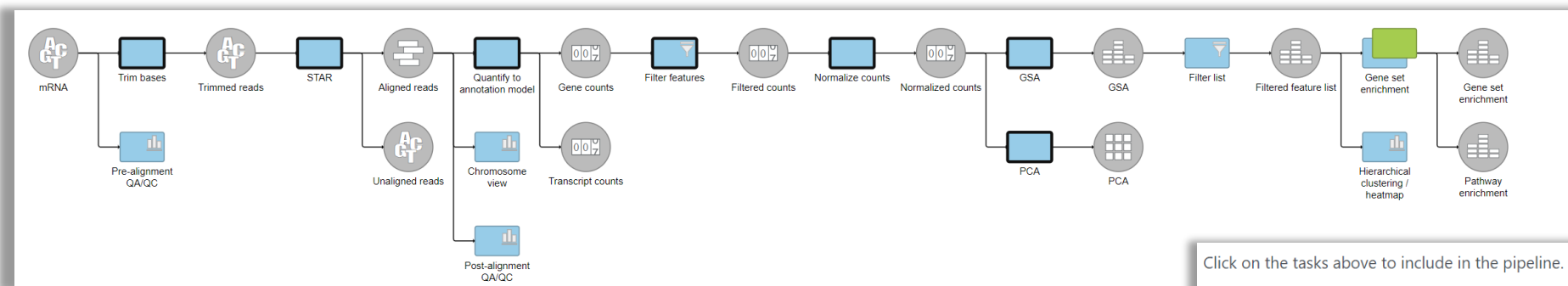
## Export and Import Analysis Projects

A screenshot of a software interface showing a menu with the following options: Toggle collapse mode, Toggle task stacking, Export project, and Delete project. A hand cursor is pointing at the 'Export project' option. An arrow points to the right, where another screenshot shows a 'Projects' dropdown menu with 'Create new project' and 'Import project' options. A hand cursor is pointing at the 'Import project' option.



# Build, Reuse, and Share Analysis Pipelines

## Build Analysis Pipelines



Click on the tasks above to include in the pipeline. Then click **Create pipeline** below.

Pipeline name:

Description:

Section name:

[Create pipeline](#)

[Cancel](#)

## Save, Share, and Manage

Home > Settings > Pipelines

[+ Import pipeline](#)

Name	Description	Creation date	Creator	Ignore	Actions
Agilent Gene Expression Pipeli...		11 Dec 2023, 09:45 PM CST		<input type="checkbox"/>	<a href="#">Download pipeline</a>
IncrNA Pipeline		11 Dec 2023, 09:45 PM CST		<input type="checkbox"/>	<a href="#">Share pipeline</a>
Dolomite Bio Drop-Seq v2		11 Dec 2023, 09:45 PM CST		<input type="checkbox"/>	<a href="#">Delete pipeline</a>
Exome germline variant detect...		11 Dec 2023, 09:45 PM CST		<input type="checkbox"/>	

# Compatible with All Major Genomics Formats and Assays

RNA | Noncoding RNA | SC RNA | DNA | ChIP | ATAC | Metagenomics



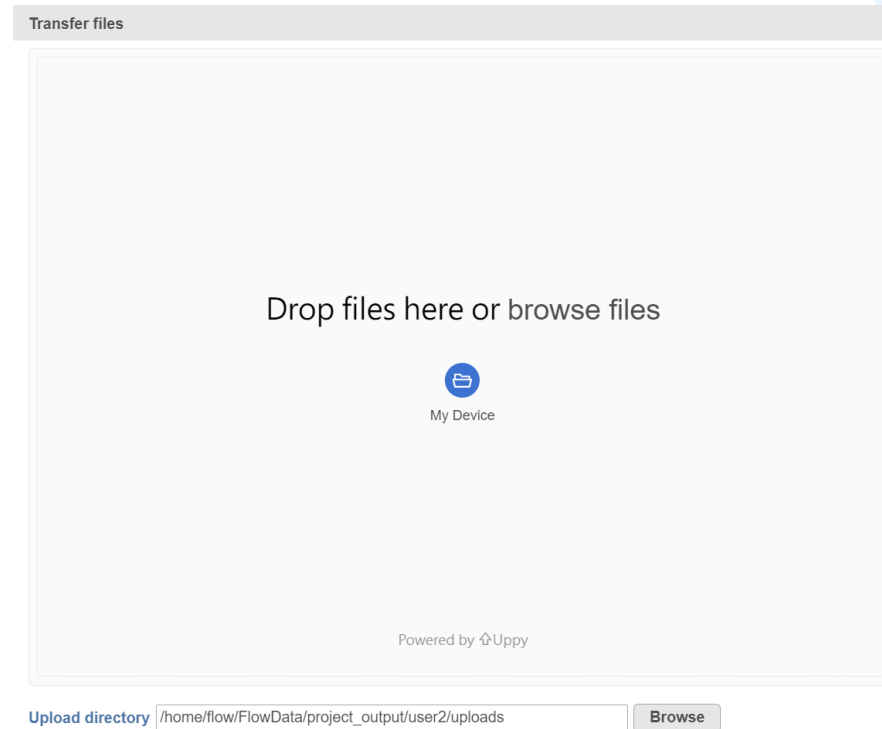
Microarray | Next Generation Sequencing | qPCR

# Partek Flow RNA-Seq Analysis



# Transfer Files

- To move files from your local computer to the Partek server, please **Transfer files** first



# Import Data



Single cell **Bulk** Microarray Other

RNA-Seq **ChIP/ATAC-Seq** DNA-Seq Metagenomics Proteomics

## Select the format

**fastq**  
Import unaligned reads. Acceptable file types are fastq, fastq.gz, fastq.bz2, fq, fq.gz, fq.bz2

**bam**  
Import aligned reads. Acceptable file types are bam, sam, and ubam

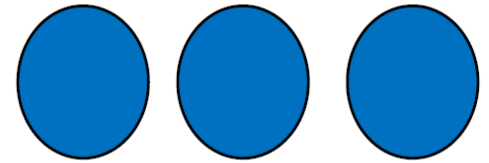
**Generic Count matrix**  
Import quantified data (e.g. gene counts per sample). Acceptable file types are txt, csv, tsv, txt.gz, csv.gz, tsv.gz

- Click **+New project** and enter a project name
- Click **Add data**
- Select the format

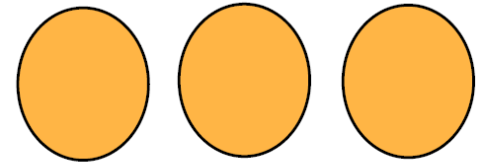
# Experiment Description

- HT29 colon cancer cells exposed to 5-aza drug with 3 different doses
  - 0  $\mu\text{M}$  (Control)
  - 5  $\mu\text{M}$
  - 10  $\mu\text{M}$
- 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)

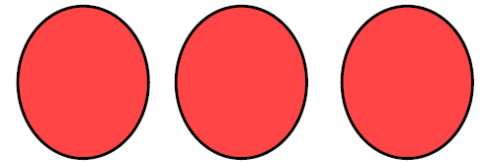
0  $\mu\text{M}$



5  $\mu\text{M}$



10  $\mu\text{M}$

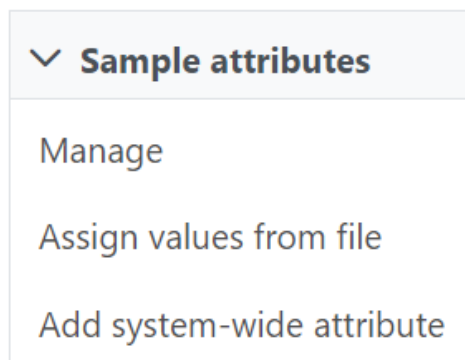




# Sample Attribute Assignment

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

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



## Sample Attributes

No sample attributes have been added to the project.

Add new attribute

Add system-wide attribute

Back to metadata tab

# 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**

## Add new attribute

### Name

### Attribute type

Categorical  Numeric

### Visibility

Project-specific  System-wide

Only modifiable by some users

Add

Cancel

## Sample Attributes

5-AZA Dose

0uM

5uM

10uM

New category

+

Add new attribute

Add system-wide attribute

Back to metadata tab



均泰生物科技股份有限公司  
Genetech Biotech Co., Ltd.

# Sample Attribute Assignment

- Click **Back to metadata 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**

	Sample name	Attributes
		5-AZA Dose
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

[Apply changes](#) [Discard changes](#)



# Assign attribute from file



▼ **Sample attributes**

Manage

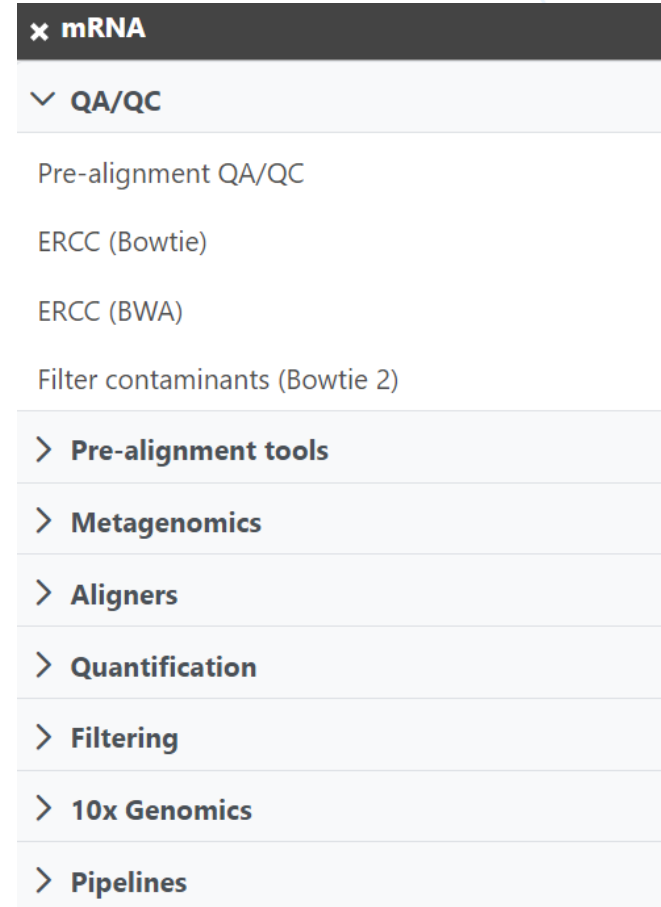
Assign values from file

Add system-wide attribute

```
RNA_attribute.txt × Training
D: > GT > 教育訓練 > Partekflow >
1 sample name Treatment
2 SRR592573 0uM
3 SRR592574 0uM
4 SRR592575 0uM
5 SRR592576 5uM
6 SRR592577 5uM
7 SRR592578 5uM
8 SRR592579 10uM
9 SRR592580 10uM
10 SRR592581 10uM
```

# 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



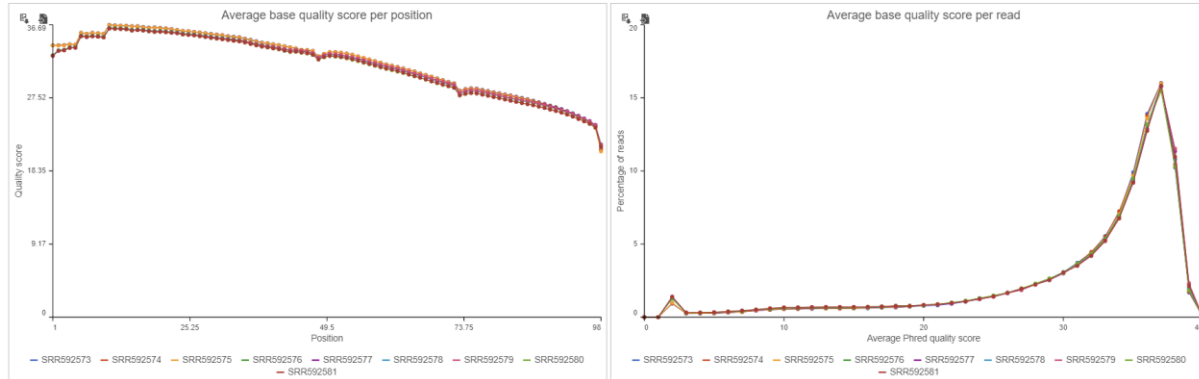
# Pre-alignment QA/QC

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

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

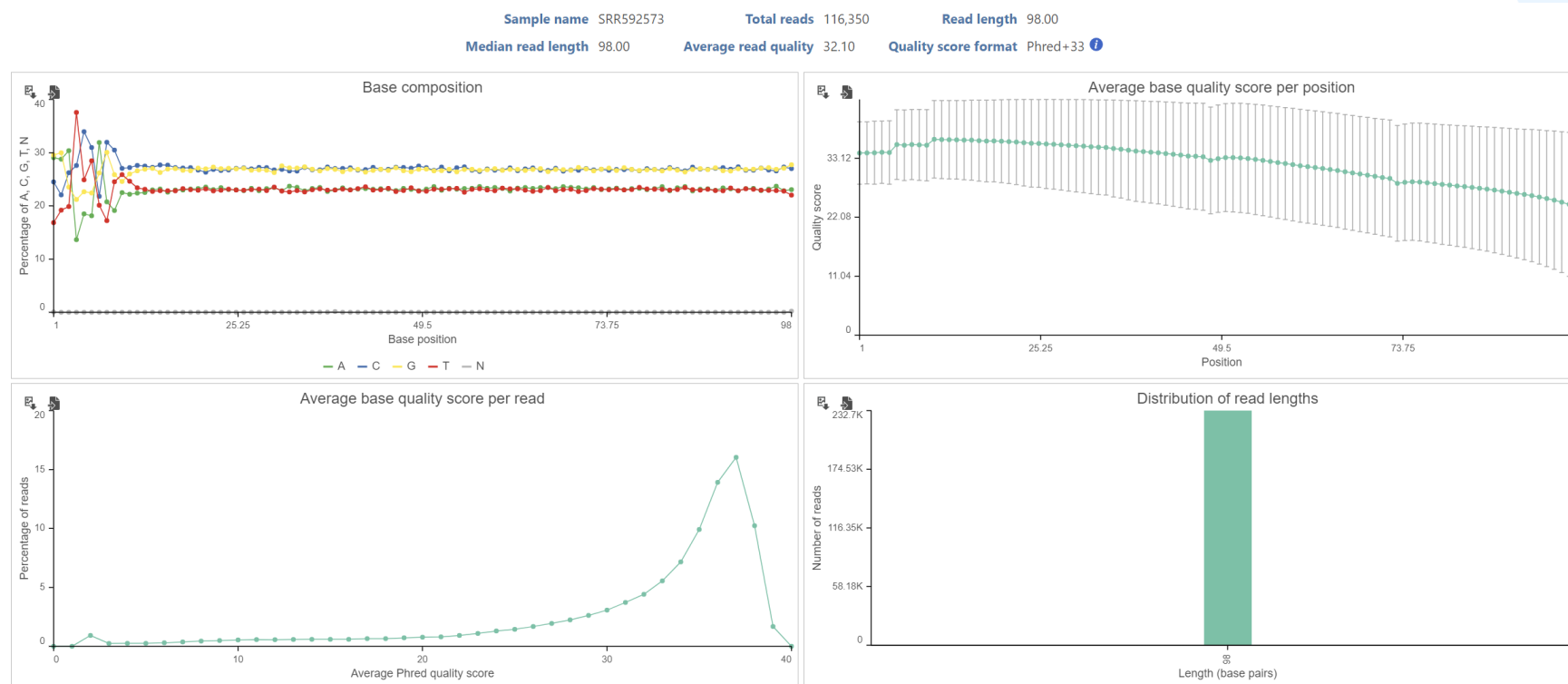


均泰生物科技股份有限公司  
Genetech Biotech Co., Ltd.



# 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

- Choose **Quality score** as the trim mode
- Edit End min quality level (Phred) to **20**
- Click **Finish**

**Trim mode**

- Quality score**  
This mode scans the read from the 5' or 3' end (or both) for the first base at or above the specified quality score. All bases previous to this position are trimmed (from the left if the 5' end, from the right if the 3' end).
- From 3' end**  
This mode trims a certain number of bases from the 3' (right-most) end of every read.
- From 5' end**  
This mode trims a certain number of bases from the 5' (left-most) end of every read.
- Both ends**  
This mode keeps all of the bases between two points (start and stop), trimming all of the other bases from the read. When the stop point is greater than the read length, the read will keep the bases from the start point to the end of the read.

**Quality trimming**

End min quality level (Phred):    
Trim from end:

**Advanced options**

**Min read length**  
After trimming, discard any reads less than this length

**Max N**  
After trimming, discards reads with larger percentage of N bases  
  %

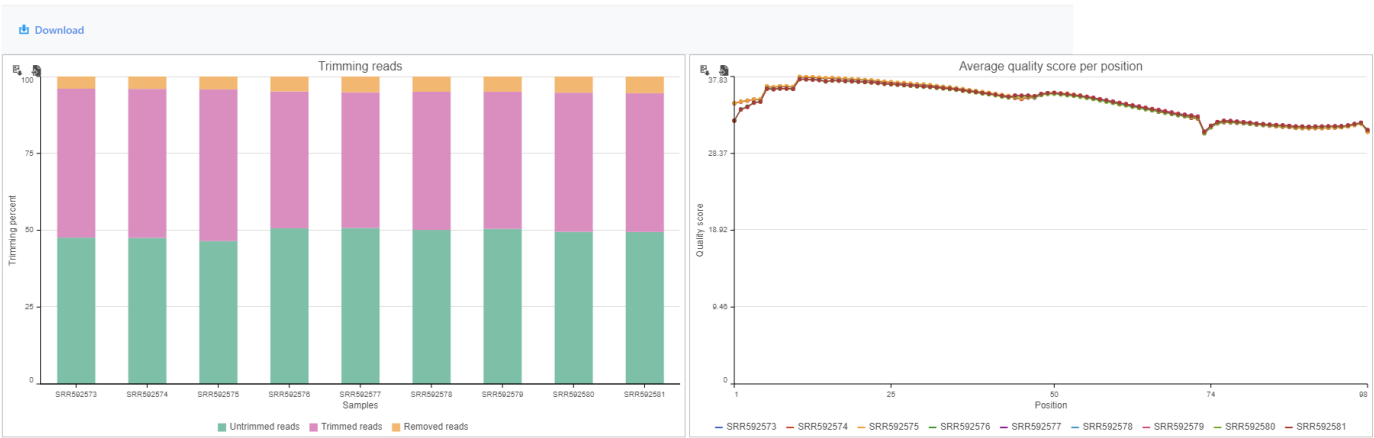
**Quality encoding**  
Apply the quality encoding to all files passed to this task. Auto-detect attempts to detect and assign the correct quality type to each file.

# Trim Bases

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

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) > >>



# Alignment

- Click the project name to return to the Analyses tab
- Click **Trimmed reads**
- Click **STAR** from Aligners
- Choose **Homo sapiens - hg19\_chr22** for Assembly and **Whole genome** for Index
- Click **Finish**



## Select STAR 2.7.8a index

### Assembly

Homo sapiens (human) - hg19\_chr22

### Index

Whole genome (Taiwan Genetech Biotech)

## Alignment options

**Generate unaligned reads**

This is at the expense of an increase in running time and disk space.

## Advanced options

### Option set

-- Default --

[Configure](#)



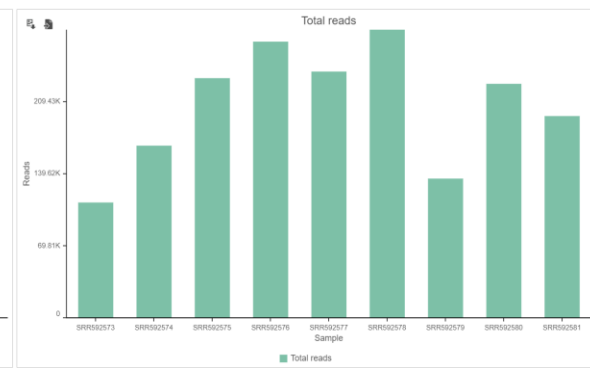
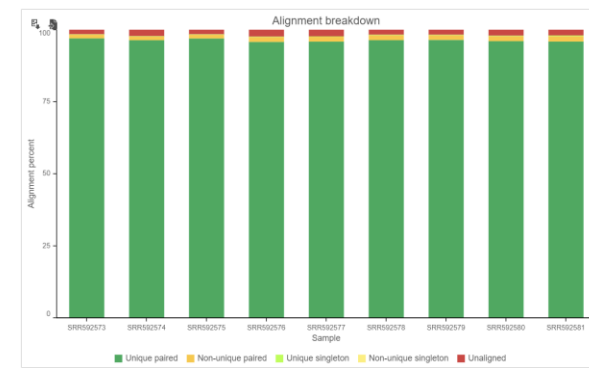
均泰生物科技股份有限公司  
Genetech Biotech Co., Ltd.

# Post-alignment QA/QC

- Click **Aligned reads** data node
- Click **QA/QC** in the task menu
- Click **Post-alignment QA/QC**
- Double-click the **Post-alignment QA/QC** task node to view the task report



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	224,382	98.44%	0.14%	96.99%	1.31%	0%	8.75%	4.62	92.93	34.40	53.37%
SRR592574	166,846	333,014	97.80%	0.16%	96.35%	1.29%	0%	10.49%	5.71	92.84	34.40	53.24%
SRR592575	232,251	466,885	98.43%	0.15%	96.96%	1.32%	0%	14.17%	5.92	92.78	34.43	52.89%
SRR592576	267,646	536,699	97.63%	0.17%	95.76%	1.70%	0%	17.20%	5.60	92.69	34.35	52.60%
SRR592577	238,647	478,459	97.69%	0.17%	95.85%	1.66%	0%	15.91%	5.40	92.75	34.39	51.66%
SRR592578	279,236	564,099	98.29%	0.19%	96.37%	1.73%	0%	17.78%	5.70	92.72	34.35	52.53%
SRR592579	134,913	272,022	98.29%	0.17%	96.44%	1.67%	0%	12.84%	3.80	92.70	34.38	51.63%
SRR592580	226,772	457,002	97.96%	0.20%	96.02%	1.75%	0%	14.32%	5.71	92.42	34.30	52.70%
SRR592581	195,532	394,633	98.01%	0.26%	95.88%	1.87%	0.01%	14.68%	4.80	92.27	34.39	51.65%





# 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**

**Select Annotation file**

---

**Assembly**  
Homo sapiens (human) - hg19\_chr22

**Annotation model**  
RefSeq (Administrator) ▾

---

**Quantification options**

**Strict paired-end compatibility**  
If not checked, then paired end reads will count as exonic even if their mate is not compatible with the transcript (--require\_proper\_pair)

**Require junction reads to match introns**  
If not checked, then junction reads will count as exonic even if their skipped regions don't match with an intron of the transcript (--check\_junctions)

**Minimum read overlap with feature**

Percent of read length  
Number of bases overlapped with feature / read length  
100 ▾

Number of bases  
Minimum number of bases of read that overlap with feature  
50 ▾

**Filter features**  
The sum of reads across all samples must be greater than or equal to this to be reported  
10 ▾

---

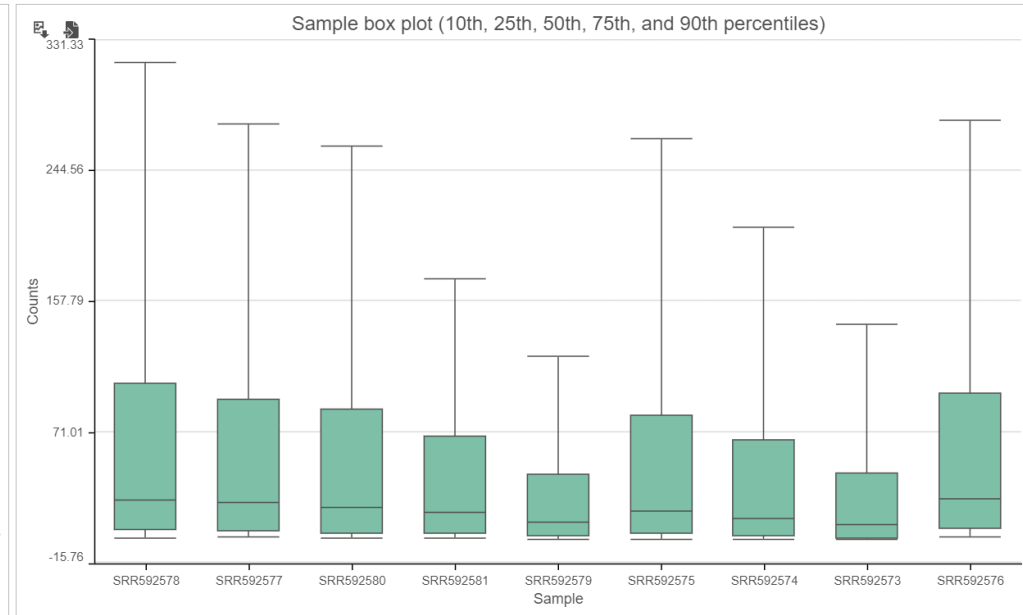
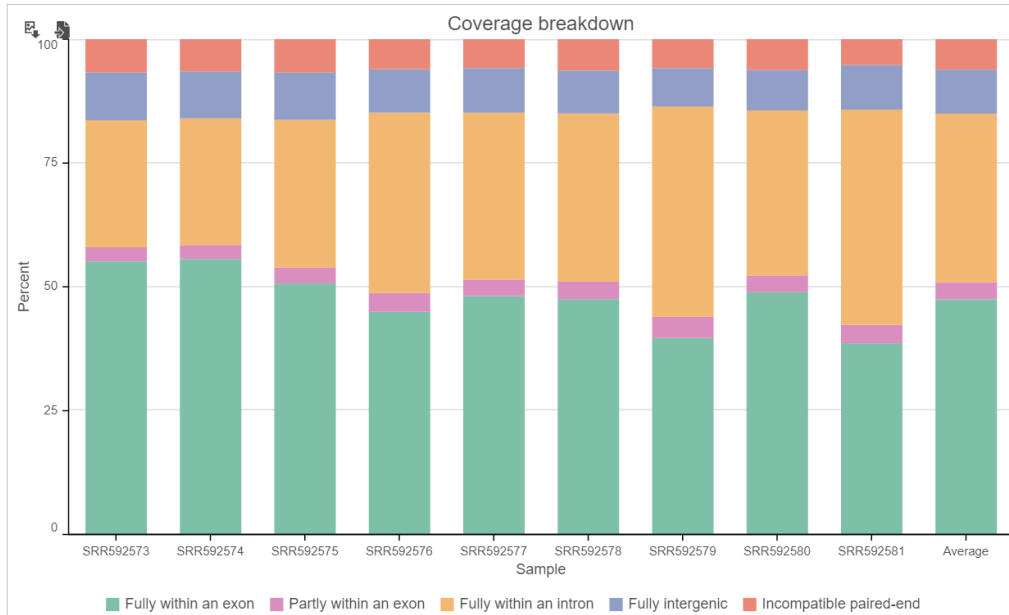
**Advanced options**

**Option set**  
-- Default -- ▾ [Configure](#)

[Back](#) [Finish](#)

# Quantification

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



# Filter 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  $\leq 0$  in 80% of cells** using the drop-down menus and text boxes
- Click **Finish** to apply the filter

## Filter type

**Noise reduction**

Exclude features that meet criteria based on descriptive statistics. Calculations are performed for each feature across all samples.

**Statistics-based**

Include a number or percentile of features based on descriptive statistics. Calculations are performed for each feature across all samples.

**Metadata**

Specify logical operations using different annotation fields.

**Saved list**

Specify a saved list of features to include or exclude.

**Manual list**

Manually specify a list of features to include or exclude.

## Filter criteria

### Filter features by

Exclude features where    in at least  % of the samples



# Normalization

- Click **Filtered counts** data node
- Click **Normalization and scaling** in the task menu
- Click **Normalization**
- Select **Recommended** if no preferred methods
- Click **Finish**



Count normalization

Transform on  
 Samples  Features

Available methods

- Absolute value
- Add
- Antilog
- Arcsinh
- CLR
- CPM (counts per million)
- Divide by
- FPKM
- Log
- Logit
- Lower bound

Drag and drop →

Selected methods Use recommended

1. Median ratio (DESeq2 only)



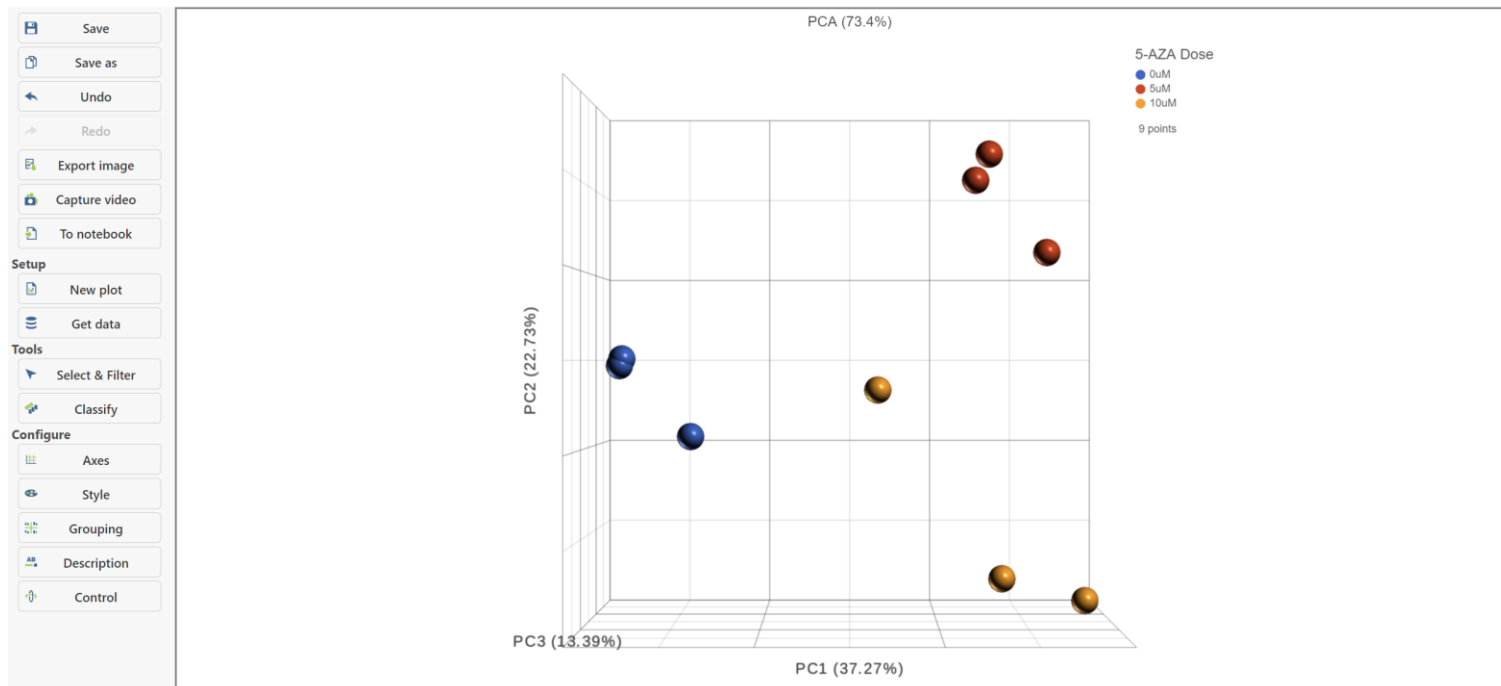
# Principal Components Analysis (PCA)

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
- Select a differential analysis method
- In this tutorial we are going to use **DESeq2**

Method to use for differential analysis ⓘ

<input checked="" type="radio"/> <b>DESeq2</b> Recommended for bulk RNA-Seq data with small sample size e.g. < 20 samples.	<input type="radio"/> <b>Hurdle model</b> Recommended for single cell RNA-Seq and CITE-Seq data.	<input type="radio"/> <b>ANOVA</b> Recommended for continuous data including bulk and single cell expression data.
<input type="radio"/> <b>Limma-trend</b> Recommended for continuous data with small sample size e.g. < 20 samples.	<input type="radio"/> <b>Limma-voom</b> Recommended for bulk RNA-Seq data with small sample size e.g. < 20 samples.	<input type="radio"/> <b>Welch's ANOVA</b> Recommended for continuous data including bulk and single cell expression data.
<input type="radio"/> <b>Kruskal-Wallis</b> Recommended for data that is not normally distributed and large sample size e.g. > 20 samples.	<input type="radio"/> <b>Gene Specific Analysis</b> Recommended for data with no replicates in any groups.	

# Differential Analysis

- Select the attribute **5-AZA Dose**
- Click **Next**
- Select comparison pairs and click **Add comparison**
- Click **Finish**

Select factor(s) for analysis


Categorical factors

5-AZA Dose

Add factors Add interaction *i*

Selected factor(s)

Factor	Delete
5-AZA Dose	—



0uM 5uM 10uM

Numerator

5uM 10uM

vs

Denominator

0uM

Combine *i*  Pairwise *i*

Add comparison

## Comparisons

Comparison	Delete
5uM vs. 0uM	—
10uM vs. 0uM	—
5uM, 10uM vs. 0uM	—



# Differential Analysis

- Double click **DESeq2** data node to open the task report

Gene list

Results: 438

Filter [Clear all](#)

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

[Save filter](#)

Saved filters

[Generate filtered node](#)

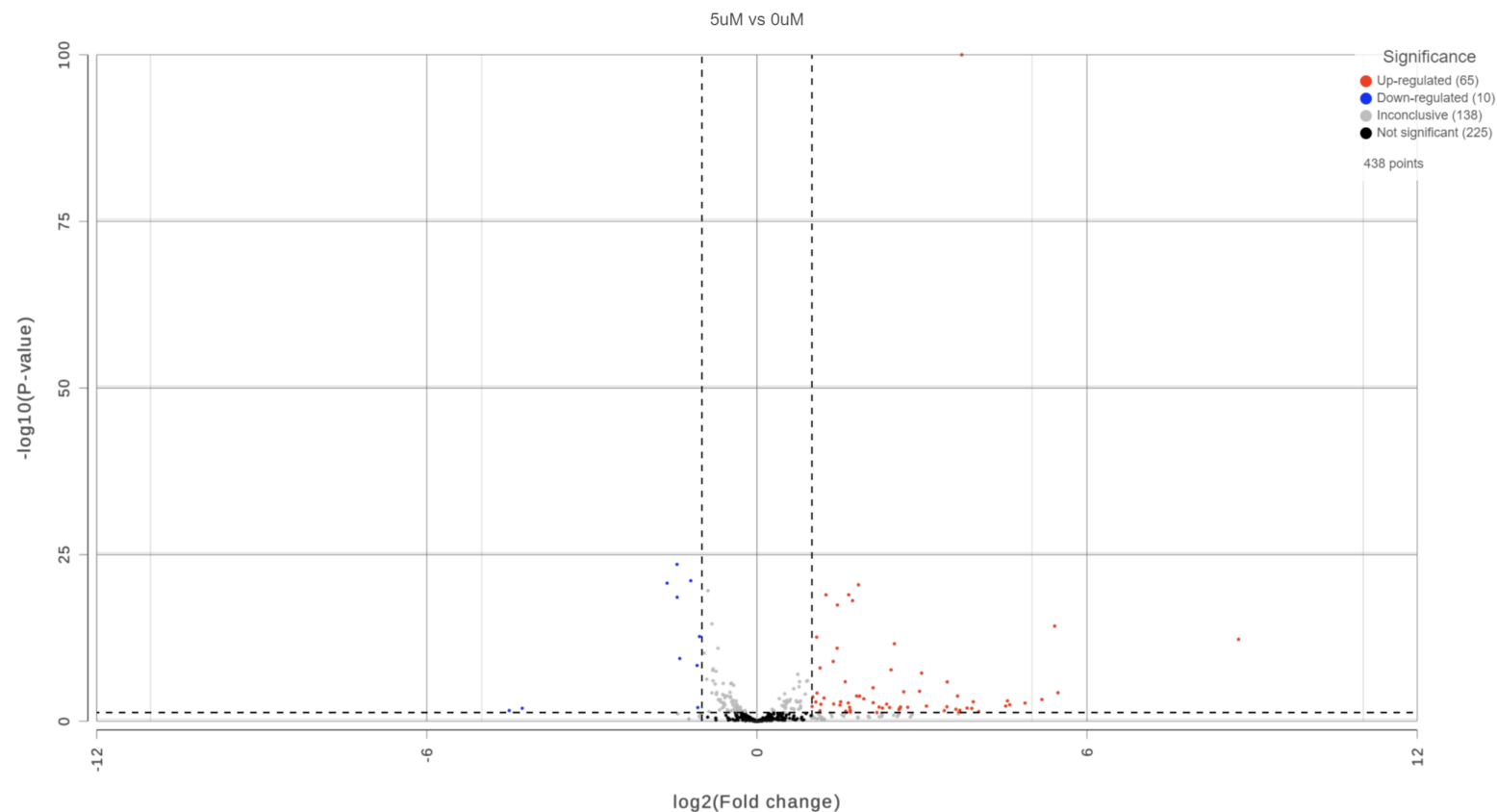
[Save as managed list](#)

Optional columns

		5uM vs 0uM						10uM vs 0uM			
View	Gene symbol	P-value	FDR step up	Ratio	Fold change	LSMean(5uM)	LSMean(0uM)	P-value	FDR step up	Ratio	
1	KLHDC7B	5.46E-143	2.39E-140	13.20	13.20	1,731.96	131.17	1.44E-91	6.33E-89	8.00	
2	CDC42EP1	2.95E-24	6.46E-22	0.37	-2.74	350.77	959.89	4.61E-16	4.04E-14	0.44	
3	H1FO	8.34E-22	1.22E-19	0.43	-2.30	1,600.13	3,684.45	6.7E-45	1.47E-42	0.29	
4	GRK3	1.92E-21	2.1E-19	0.32	-3.10	135.98	421.86	2E-16	2.19E-14	0.37	
5	KREMEN1	3.4E-21	2.98E-19	3.59	3.59	321.32	89.42	9.01E-8	2.82E-6	2.12	
6	TRIOBP	2.59E-20	1.89E-18	0.54	-1.85	518.11	960.15	6.04E-16	4.41E-14	0.57	
7	TYMP	1.07E-19	5.94E-18	3.18	3.18	265.85	83.66	1.92E-7	5.25E-6	2.00	
8	APOL1	1.09E-19	5.94E-18	2.39	2.39	581.53	243.57	4.89E-12	2.68E-10	1.97	
9	RPL3	2.5E-19	1.22E-17	0.37	-2.73	592.53	1,618.44	3.68E-20	5.37E-18	0.35	

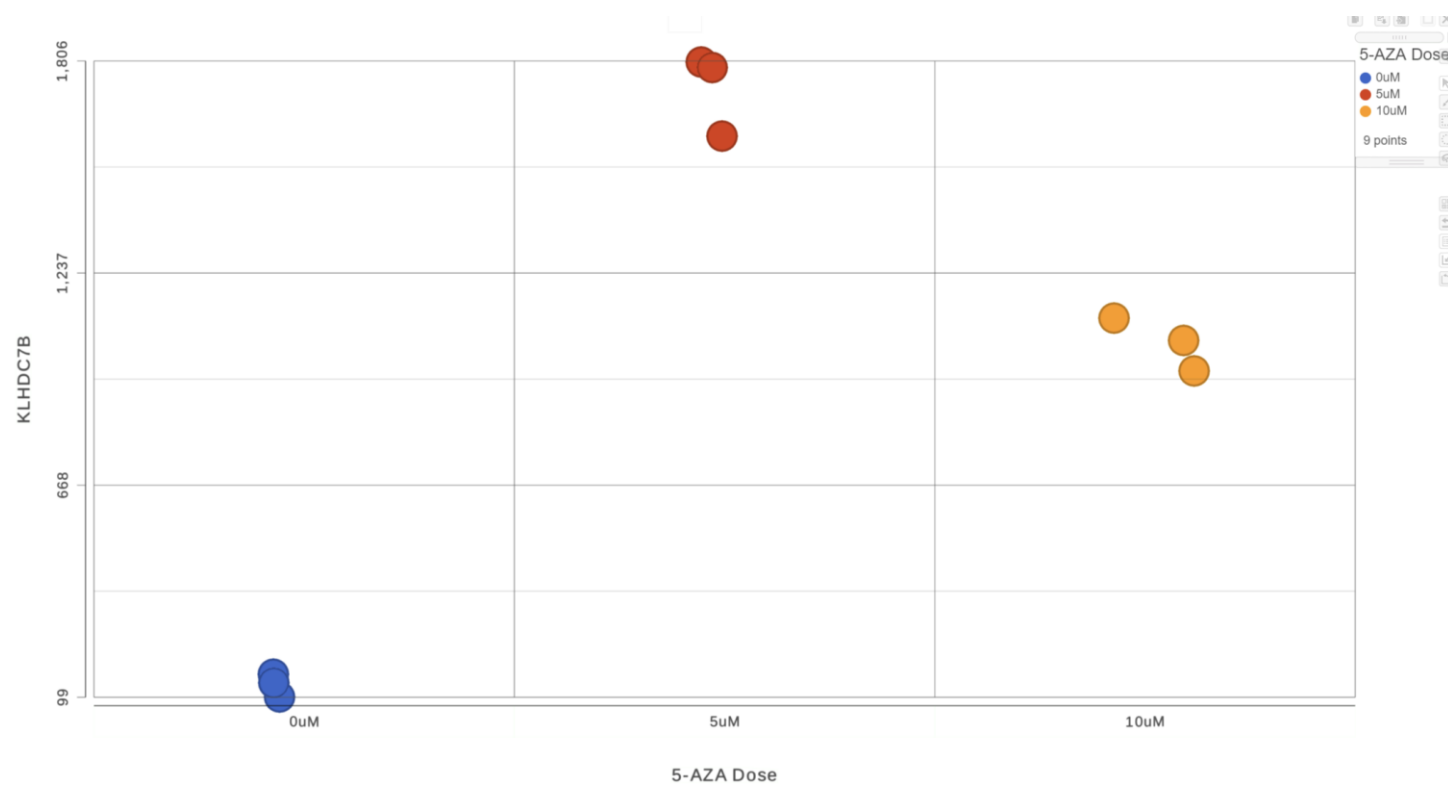
# Volcano Plot for All Genes

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



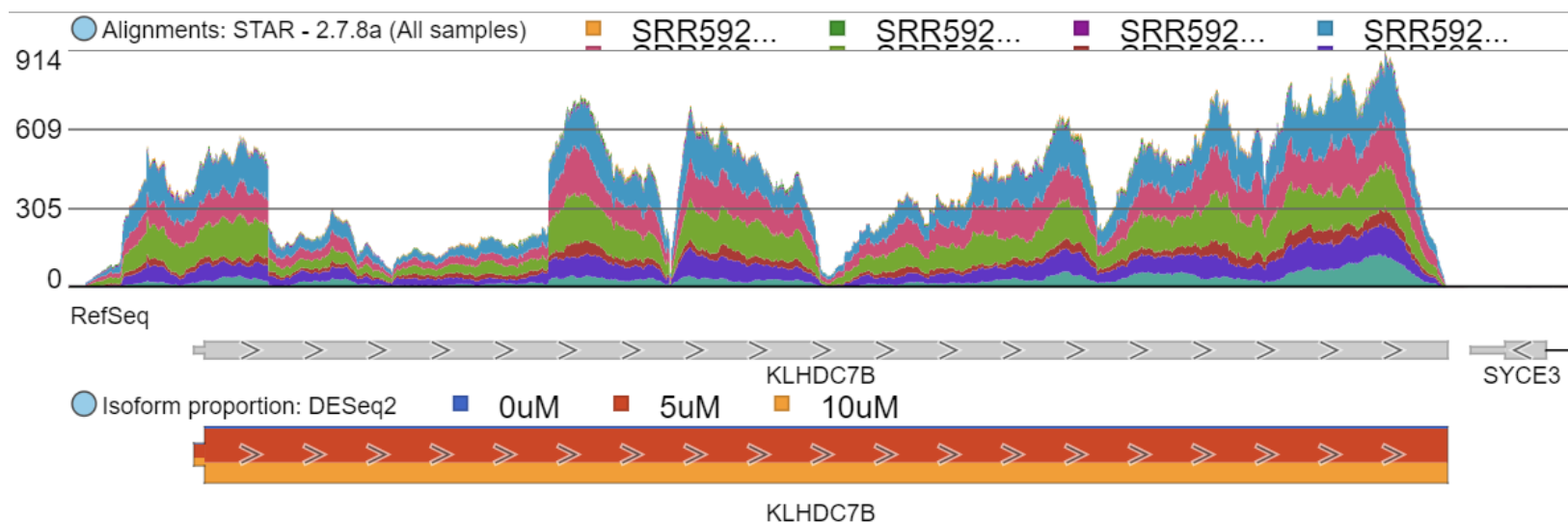
# Dot Plot for One Gene

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



# Browse Gene in Chromosome View

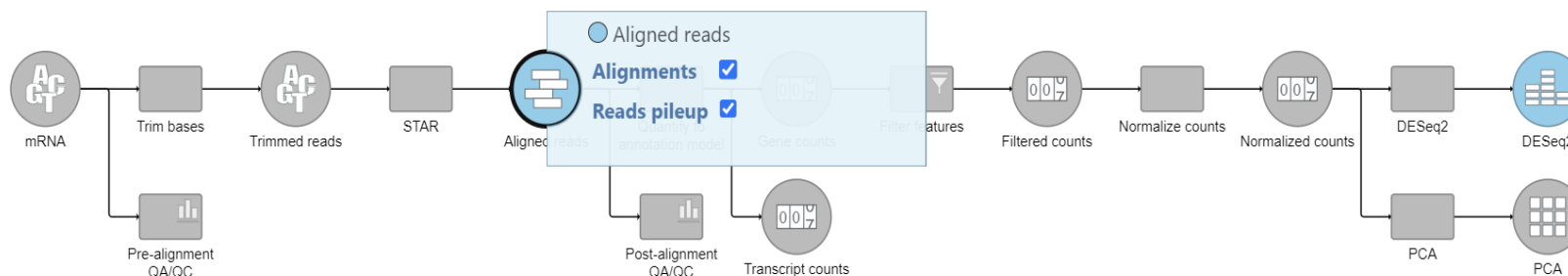
- Click  next to a gene symbol to open Chromosome view





# 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**

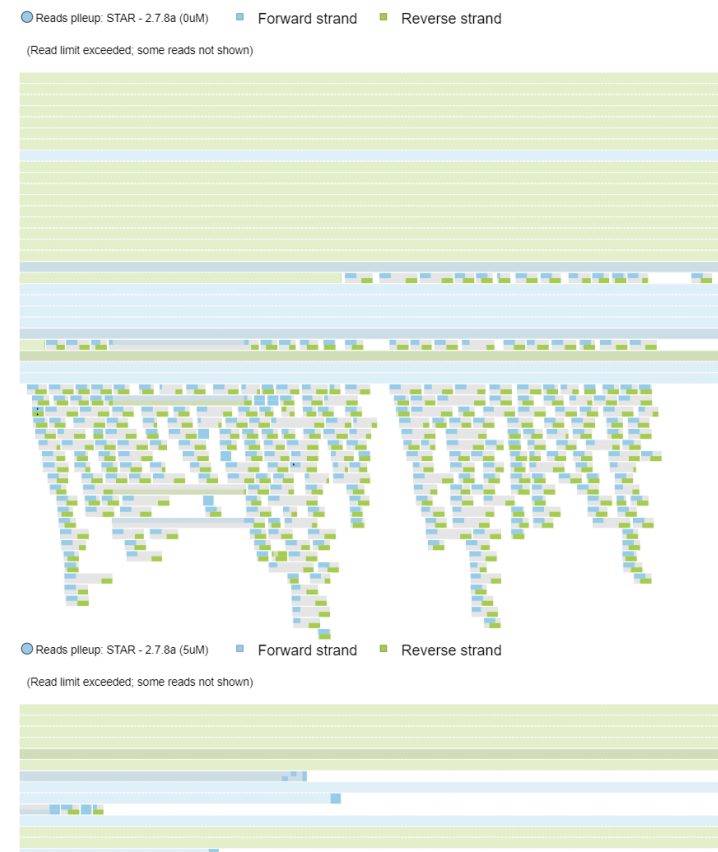
Controls

Group data by

- All
- Sample
- Attribute: 5-AZA Dose ▼

Apply

Customize track colors



均泰生物科技股份有限公司  
Genetech Biotech Co., Ltd.

# Identify DEG

- Switch to the browser tab showing the DESeq2 report
- Click **FDR step up**
- Set the cutoff value to **0.05**
- Click **Fold-change**
- Set to **From -2 to 2** with **Exclude range** selected
- Click **Generate filtered node** to create a data node with only the genes that pass the filter

The screenshot shows a filter configuration panel for a DESeq2 report. At the top, it displays "Results: 22" and a "Clear all" link. The filter settings are as follows:

- Filter** (Clear all)
- Gene symbol
- P-value
- FDR step up**
  - All contrasts  Per contrast
  - Less than or... 0.05
  - Slider: 0 to 1
- Ratio
- Fold change**
  - All contrasts  Per contrast
  - From -2 to 2
  - Exclude range
- LSMean
- Low expressed

Buttons at the bottom: "Save filter", "Generate filtered node", and "Save as managed list".



# 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

Heatmap  Bubble map

**Ordering**

**Feature order**

Cluster  
Cluster by distance metrics to sort based on similarity. Requires at least 3 features.

Assign order  
Assign feature order using a saved feature list. The features will be filtered to those in the list and will be ordered as they are listed.

Default order

**Sample order**

Cluster  
Cluster by distance metrics to sort based on similarity. Requires at least 3 samples.

Assign order  
Order samples by an attribute. Categorical attributes with fewer than 50 categories can be manually ordered by drag and drop. Numeric attributes can be sorted.

Default order

**Filtering**

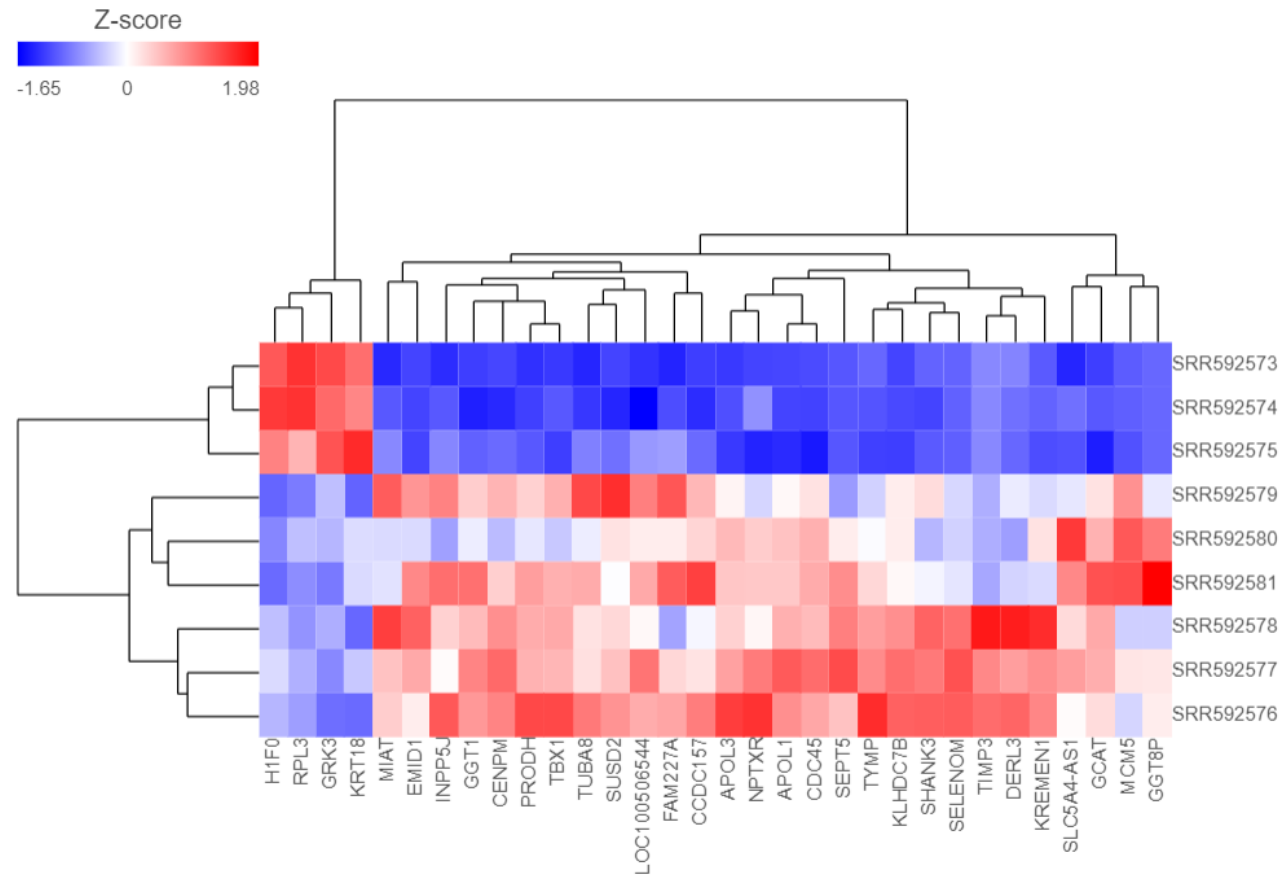
**Filter samples**  
Specify the logical operations to filter by. Use AND for inclusion if all conditions pass. Use OR for inclusion if any conditions pass.

include  Sample name  in  SRR592578  OR



# 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**

## Annotations

### Data

Row annotation

-- Add new --

5-AZA Dose

Normalized counts

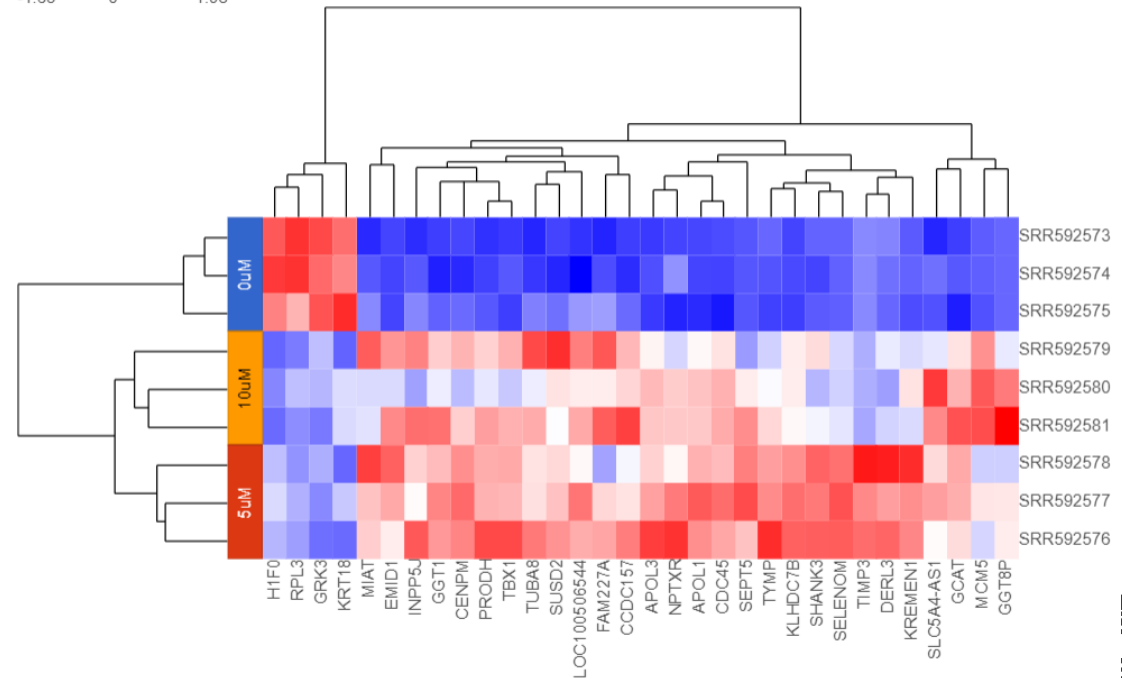
### Style

Block size

Annotation font 12 px

Show labels

Fill blocks



# 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

## Select gene set

### Database

KEGG database  Gene set database

### Assembly

Homo sapiens (human) - hg19\_chr22

### Gene set database

GO (Administrator) ▼





# Enrichment Analysis

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

Gene set ↑↓	Description ↑↓	Type ↑↓	Enrichment score ↑↓	P-value ↑↓	FDR step up ↑↓	Rich factor ↑↓	Genes in set ↑↓	Genes in list ↑↓	Genes not in list ↑↓	Genes in list, not in set ↑↓	Genes not in list, not in set ↑↓	
GO:1901605	alpha-amino acid metabolic process	biological process	9.30	9.18E-5	0.43	1.00	3	3	0	13	317	☰ 📄
GO:0005198	structural molecule activity	molecular function	7.36	6.36E-4	0.68	0.40	10	4	6	12	311	☰ 📄
GO:0006520	cellular amino acid metabolic process	biological process	6.39	1.68E-3	0.68	0.50	6	3	3	13	314	☰ 📄
GO:0042219	cellular modified amino acid catabolic process	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0043648	dicarboxylic acid metabolic process	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0044843	cell cycle G1/S phase transition	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0022616	DNA strand elongation	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:1901606	alpha-amino acid catabolic process	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0000082	G1/S transition of mitotic cell cycle	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0003688	DNA replication origin binding	molecular function	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄
GO:0006270	DNA replication initiation	biological process	6.13	2.17E-3	0.68	1.00	2	2	0	14	317	☰ 📄

# 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**

## Select gene set

### Database

KEGG database  Gene set database

### KEGG database

Homo sapiens hsa\_v5\_23\_09\_13 (taiwanbiotech2) ▼



# KEGG Enrichment Analysis

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

Gene set ↑↓	Description ↑↓	Enrichment score ↑↓	P-value ↑↓	FDR step up ↑↓	Rich factor ↑↓	Genes in set ↑↓	Genes in list ↑↓	Genes not in list ↑↓	Genes in list, not in set ↑↓	Genes not in list, not in set ↑↓	
path:hsa04724	Glutamatergic synapse	4.03	0.02	1.00	0.67	3	2	1	10	133	☰☱
path:hsa05032	Morphine addiction	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa00430	Taurine and hypotaurine metabolism	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa04740	Olfactory transduction	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa03030	DNA replication	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa03010	Ribosome	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa00260	Glycine, serine and threonine metabolism	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa00330	Arginine and proline metabolism	2.50	0.08	1.00	1.00	1	1	0	11	134	☰☱
path:hsa04110	Cell cycle	2.27	0.10	1.00	0.29	7	2	5	10	129	☰☱
path:hsa00590	Arachidonic acid metabolism	1.84	0.16	1.00	0.50	2	1	1	11	133	☰☱
path:hsa04145	Phagosome	1.84	0.16	1.00	0.50	2	1	1	11	133	☰☱
path:hsa05012	Parkinson disease	1.82	0.16	1.00	0.22	9	2	7	10	127	☰☱



# Pipeline Overview

