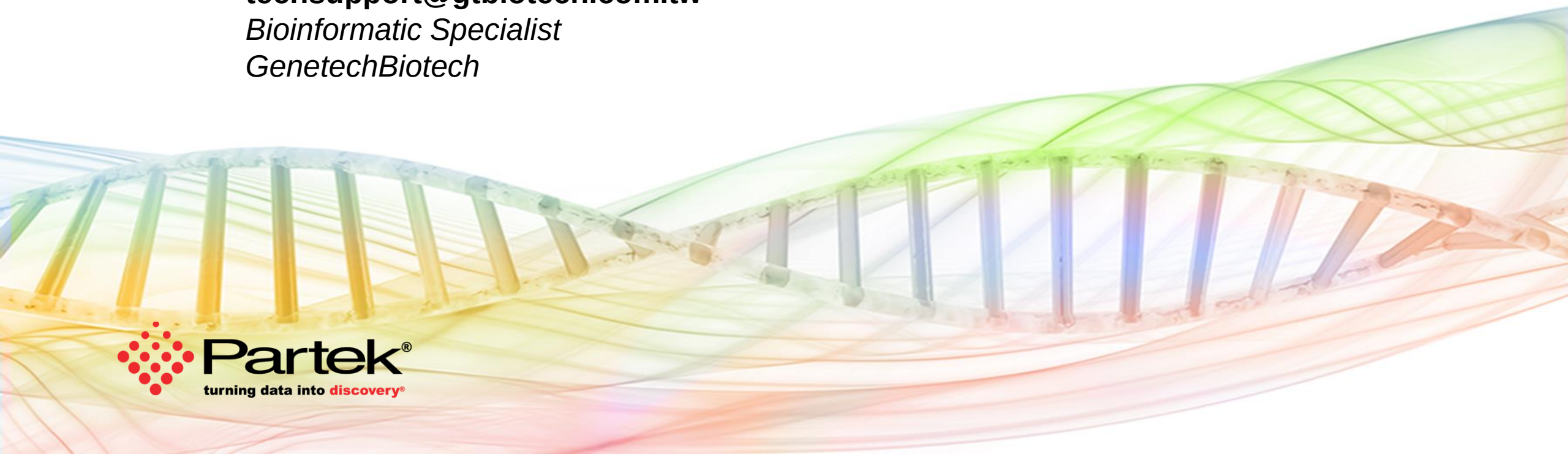
From Raw Data to Pathways: Easy Genomics Analysis with Partek Flow

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



Partek®
turning data into discovery®

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

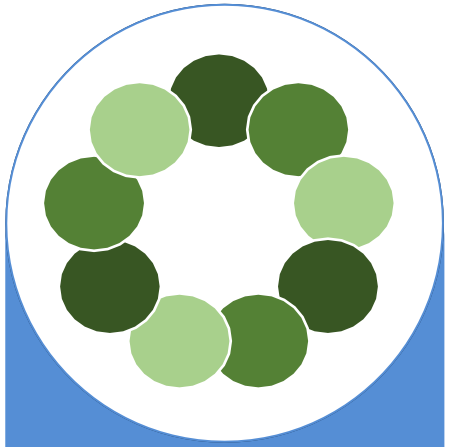
Partek®

Partek Flow: Start-to-Finish Bioinformatics Solution

Interactive Visualizations

Collaborative Web Environment

Powerful Statistics



Comprehensive Application Support

Partek™ an Illumina company

Publicly Available Statistical Algorithms and Tools

Alignment
- Bowtie
- BWA
- Isaac
- TopHat
- TMAP
- Bowtie
- GSNAP
- STAR
- HISAT

Differential analysis
- Limma
- DESeq2
- Poisson
- Negative binomial
- Non-parametric ANOVA

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

QA/QC reports
- Pre-alignment
- Post-alignment
- ERCC spike-in
- Single cell quality

Clustering
- Hierarchical
- K-means
- Graph-based

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

Variant calling
- Samtools
- LoFreq
- CNVkit
- FreeBayes
- Strelka
- GATK

Variant annotation
- SnpEff
- dbSNP
- VEP
- Custom databases

Peak calling
- MACS2
- TSS plot
- Motif detection

Quantification
- Partek E/M
- HTSeq
- Cufflinks

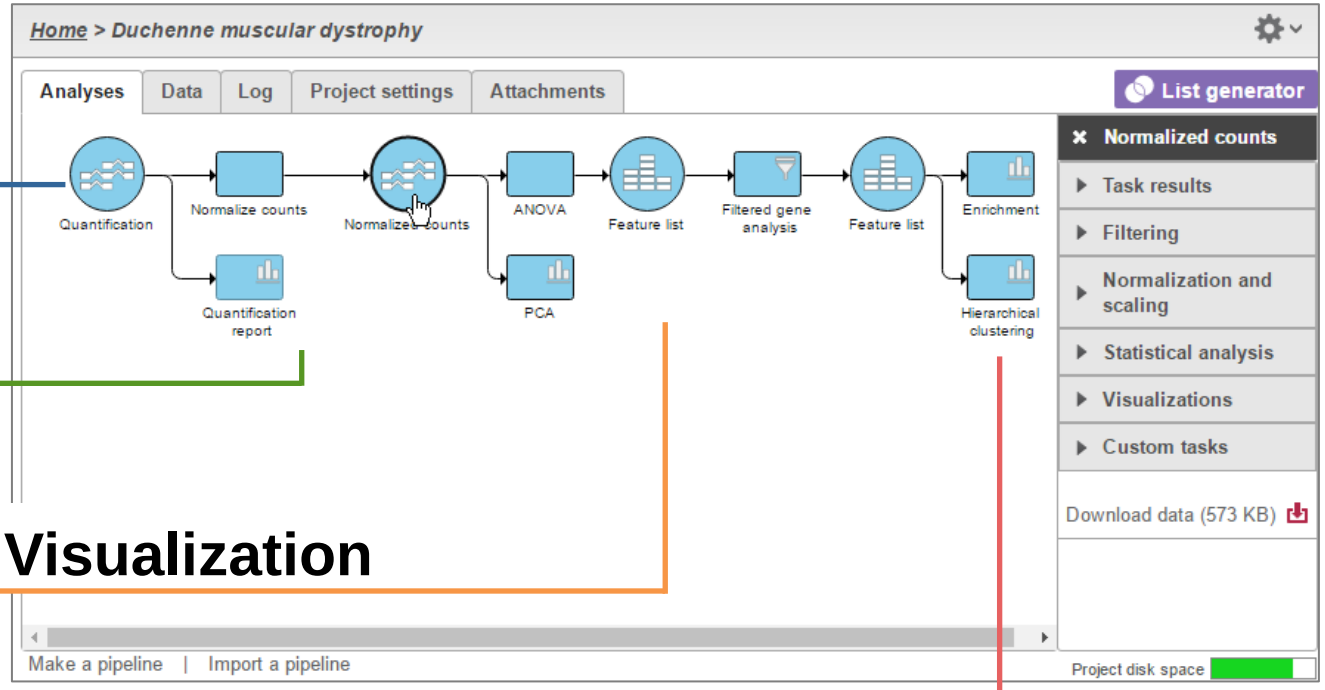
Visual Analysis Process

1 **Import Data**

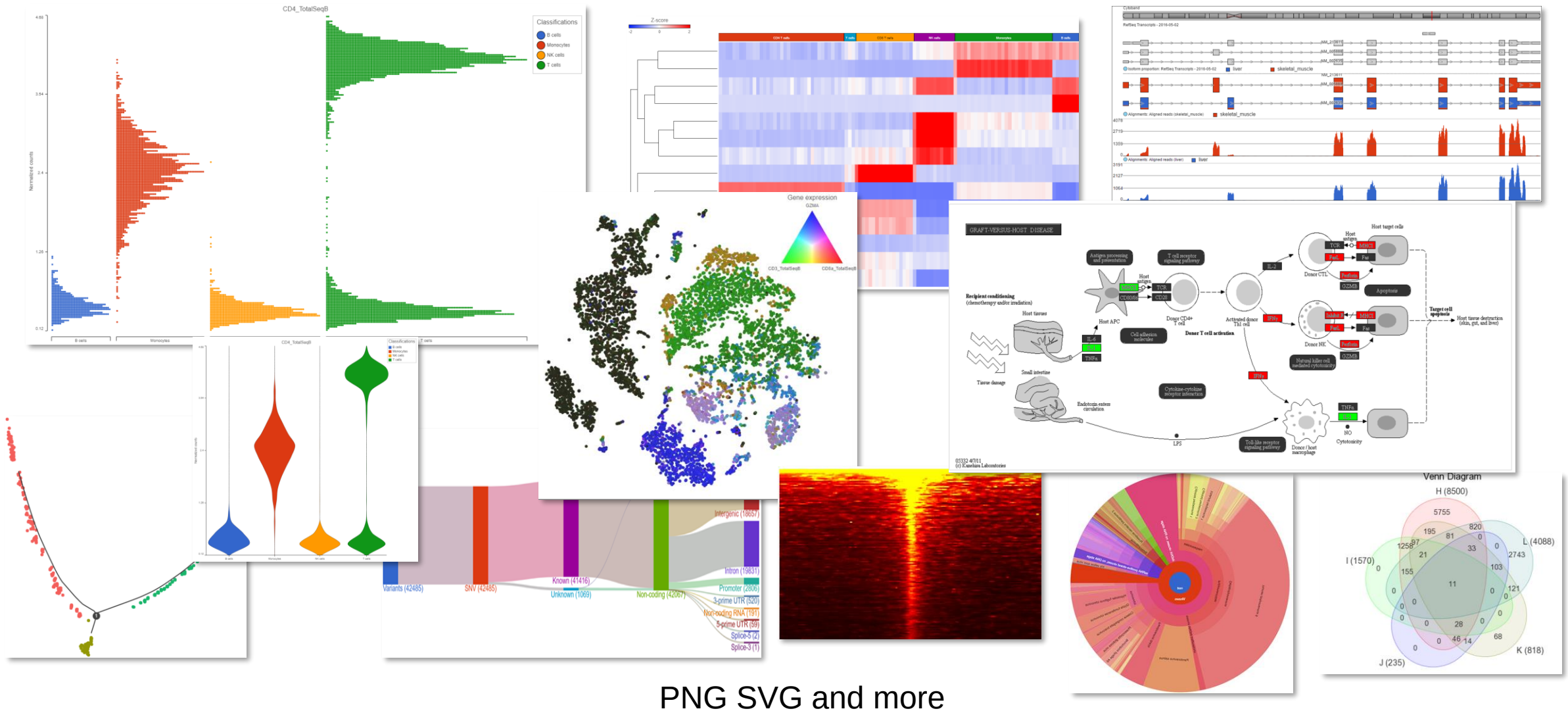
2 **QA/QC**

3 **Powerful Statistics & Visualization**

4 **Biological Interpretation**



Compelling and Publishable Visualizations

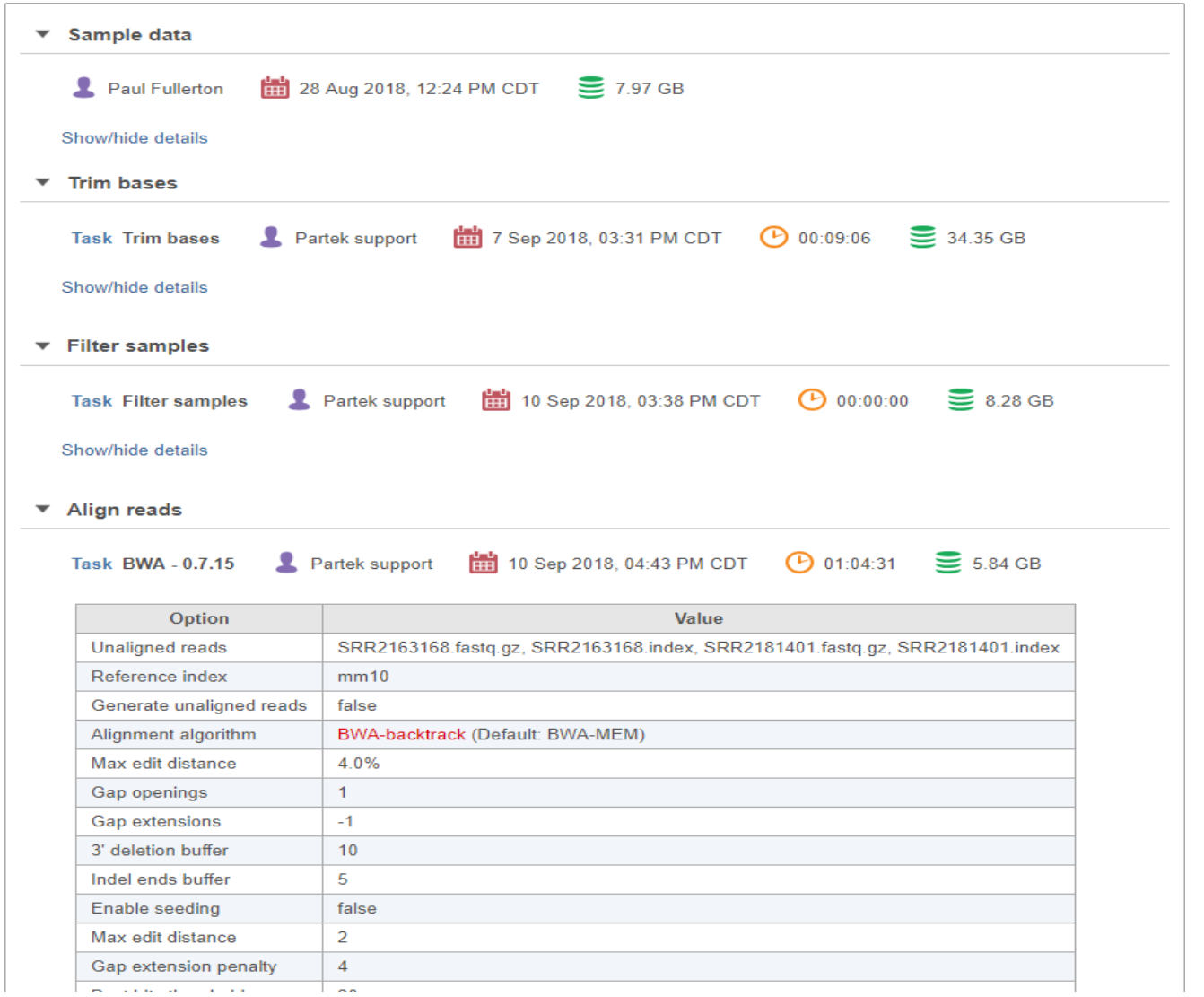


PNG SVG and more

Partek®

Summary Report

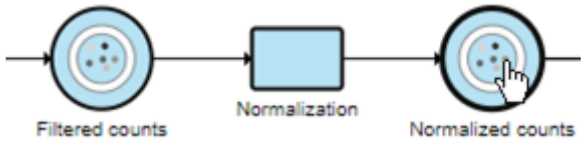
- Who
- When
- What
- How long
- How much

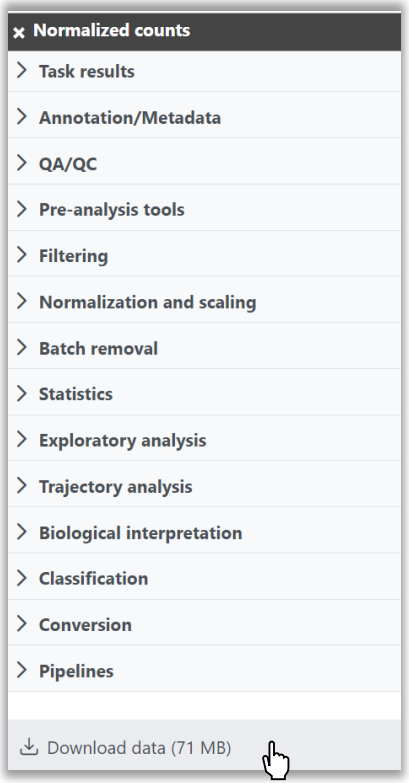


Option	Value
Unaligned reads	SRR2163168.fastq.gz, SRR2163168.index, SRR2181401.fastq.gz, SRR2181401.index
Reference index	mm10
Generate unaligned reads	false
Alignment algorithm	BWA-backtrack (Default: BWA-MEM)
Max edit distance	4.0%
Gap openings	1
Gap extensions	-1
3' deletion buffer	10
Indel ends buffer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

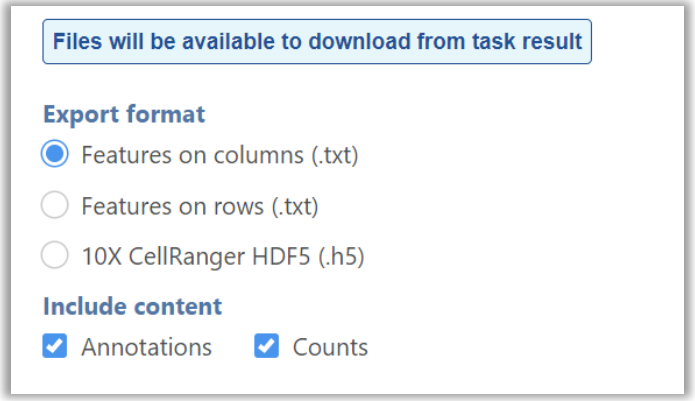
Export Data

Choose Any Data



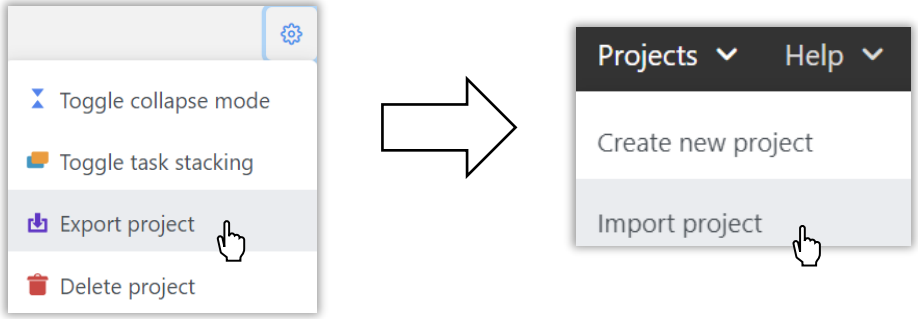


Download in Industry Standard Formats



FASTQ, BAM, TXT, and more

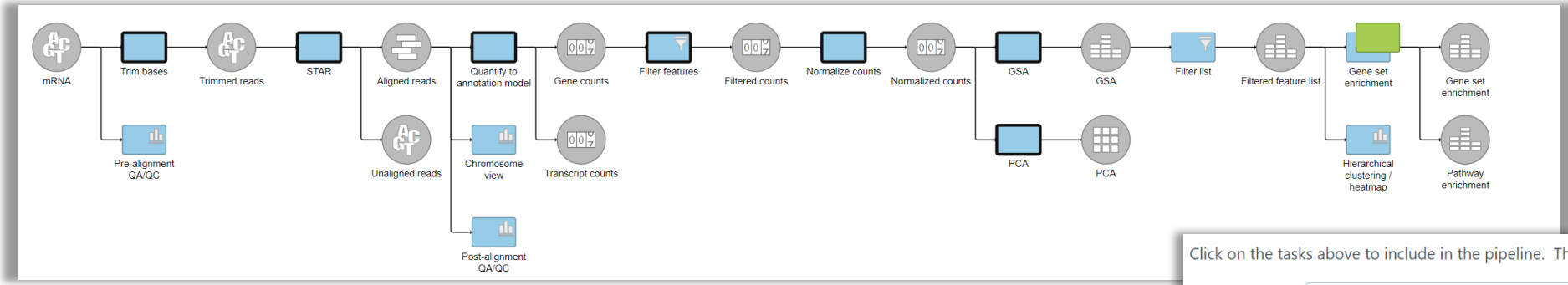
Export and Import Analysis Projects

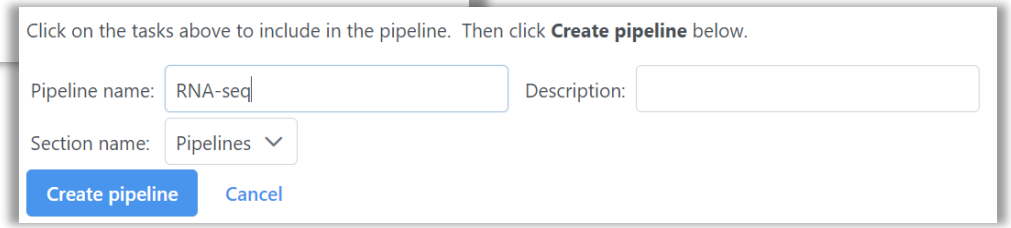


Partek™ an Illumina company

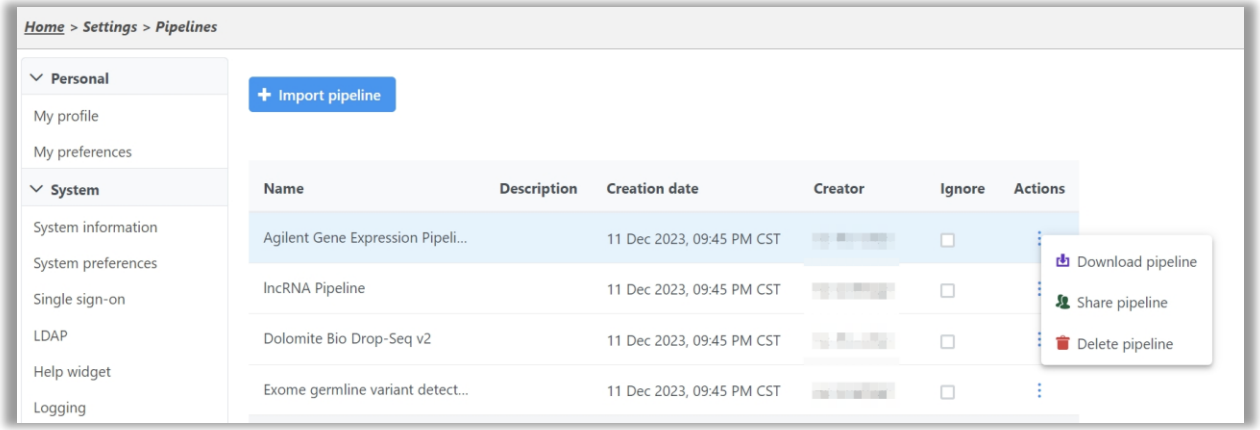
Build, Reuse, and Share Analysis Pipelines

Build Analysis Pipelines





Save, Share, and Manage



Partek™ an Illumina company

Compatible with All Major Genomics Formats and Assays

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

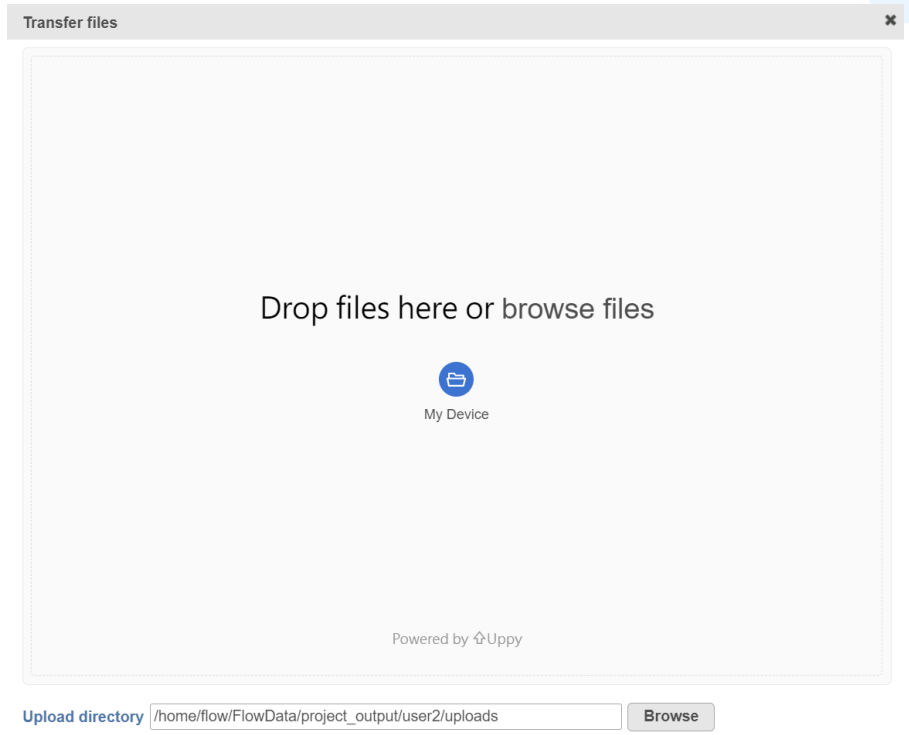
Microarray | Next Generation Sequencing | qPCR

Partek Flow
RNA-Seq Analysis

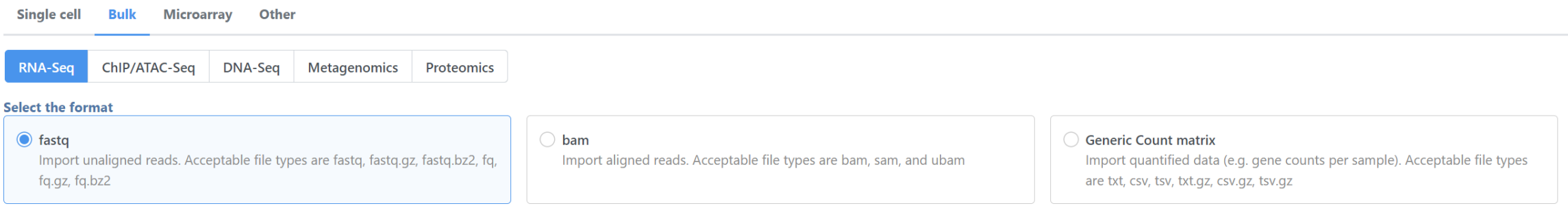
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Transfer Files

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



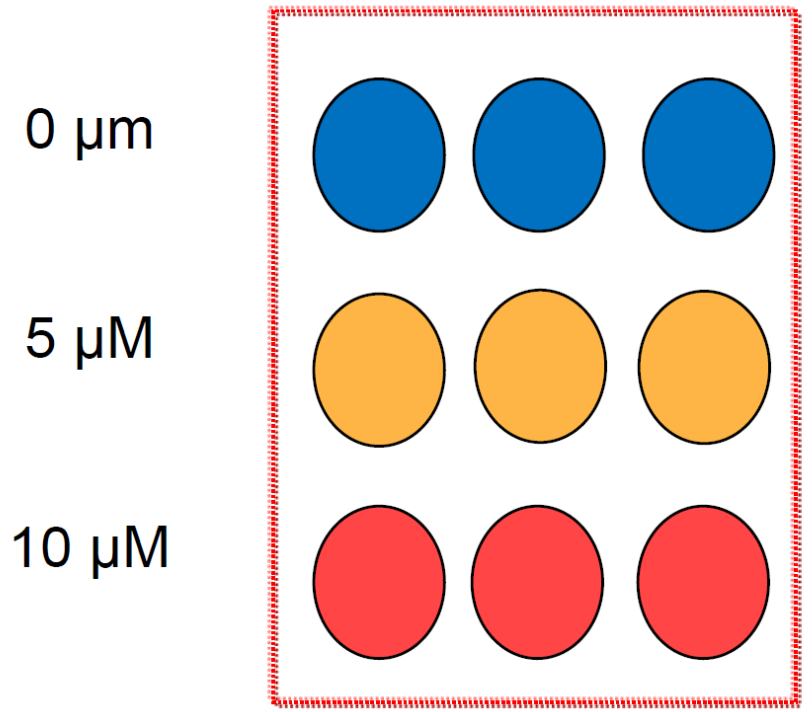
Import Data



- 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 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)

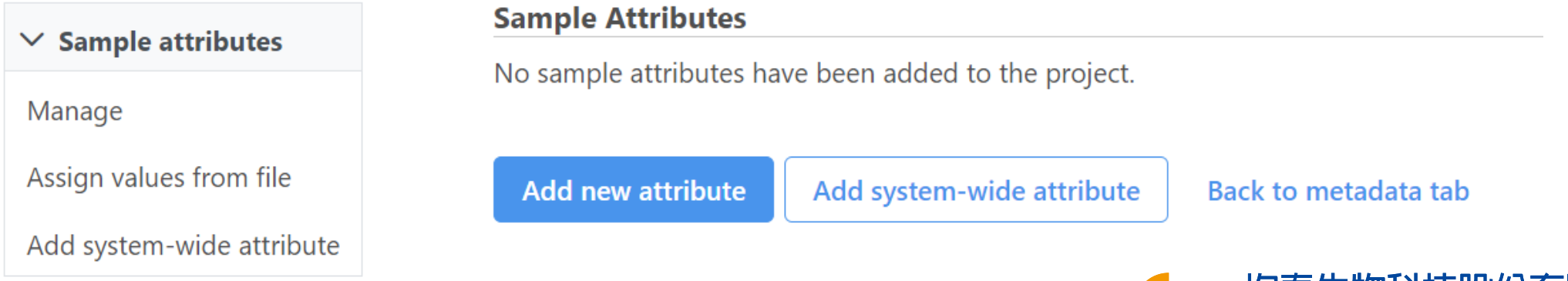


均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Sample Attribute Assignment

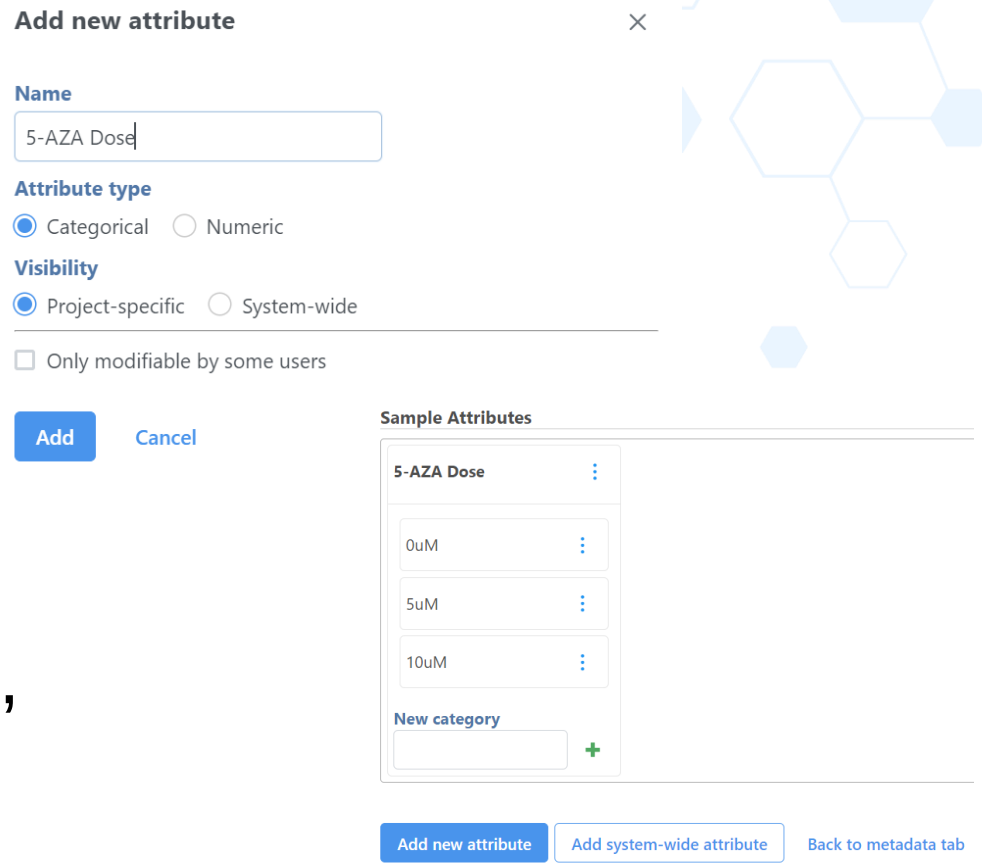
Once the download completes, the sample table will appear in the **Metadata** tab.

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



Sample Attribute Assignment

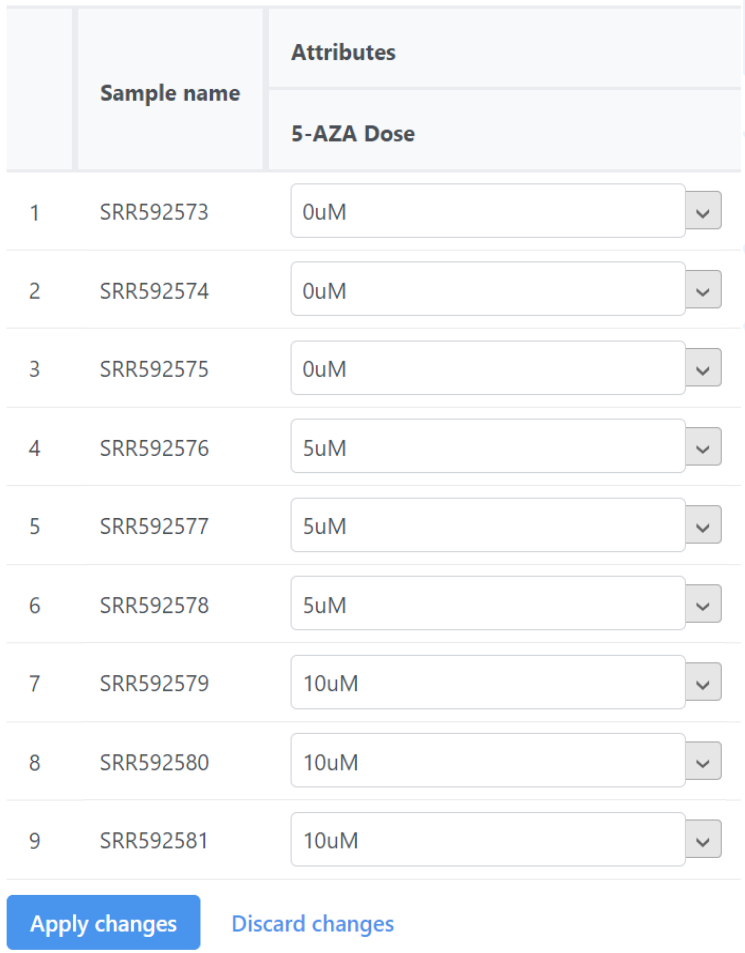
- Name the attribute **5-AZA Dose**
- Click **Categorical** and **Project-specific**
- Click **Add**
- Name the first New category **0uM**
- Click **Add**
- Repeat for two additional categories, **5uM** and **10uM**



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**

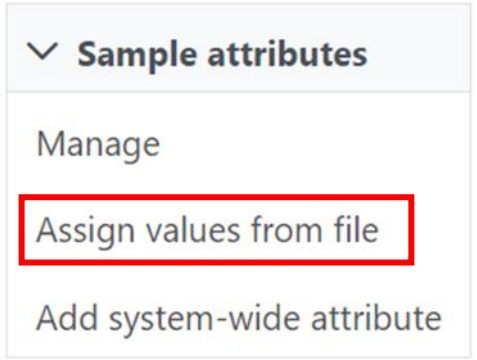


	Sample name	Attributes
		5-AZA Dose
1	SRR592573	0uM
2	SRR592574	0uM
3	SRR592575	0uM
4	SRR592576	5uM
5	SRR592577	5uM
6	SRR592578	5uM
7	SRR592579	10uM
8	SRR592580	10uM
9	SRR592581	10uM

Apply changes Discard changes

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

Assign attribute from file



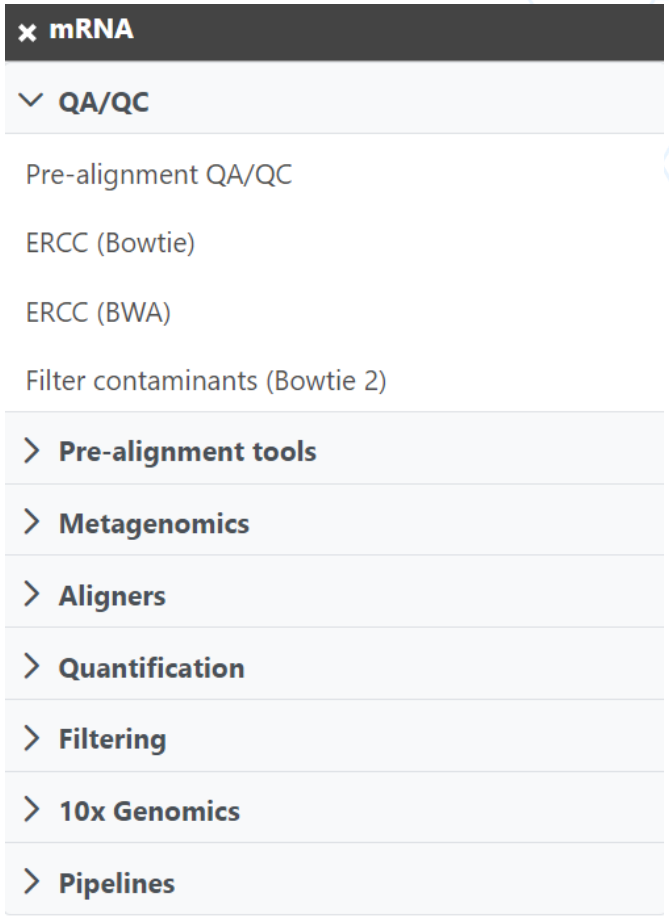
RNA_attribute.txt

D: > GT > 教育訓練 > Partekflow >

	sample name	Treatment
1		
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

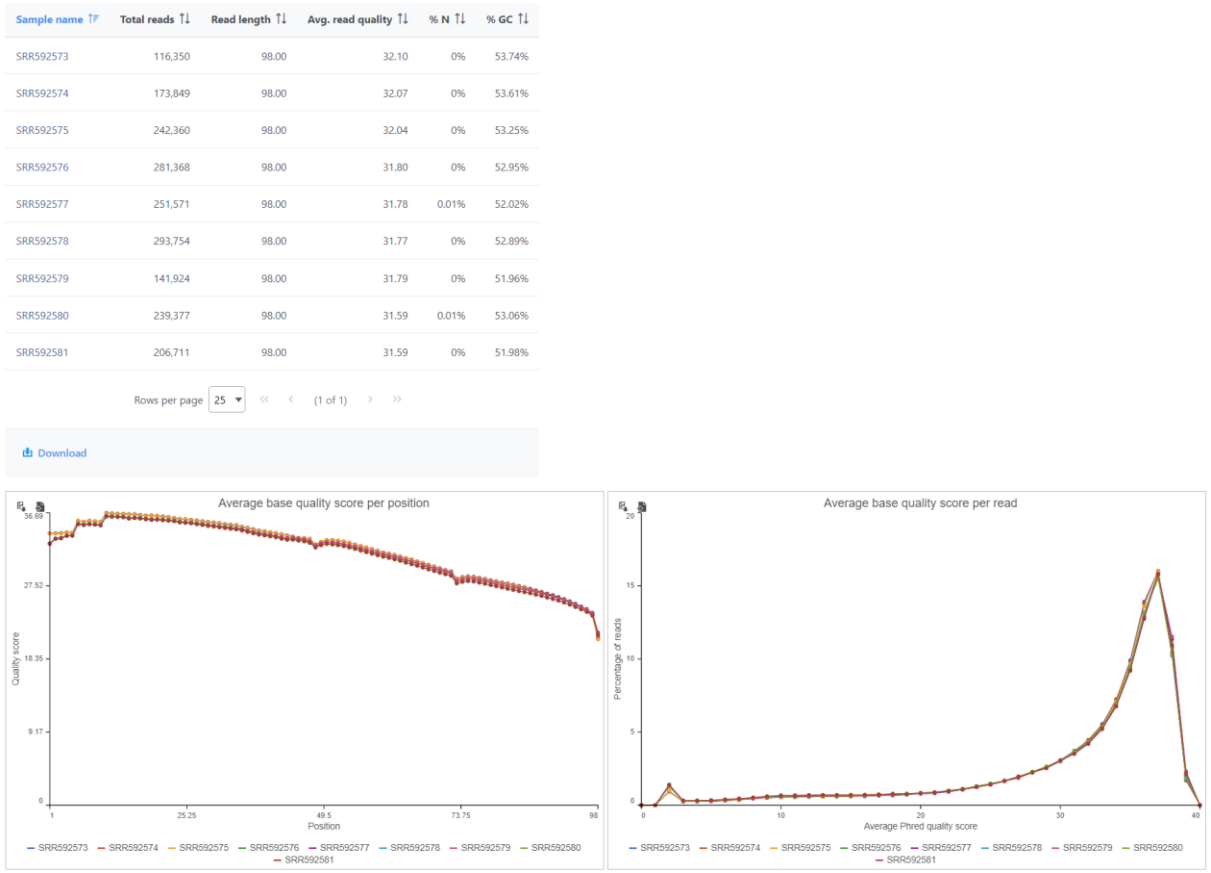
Pre-alignment QA/QC

- 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%

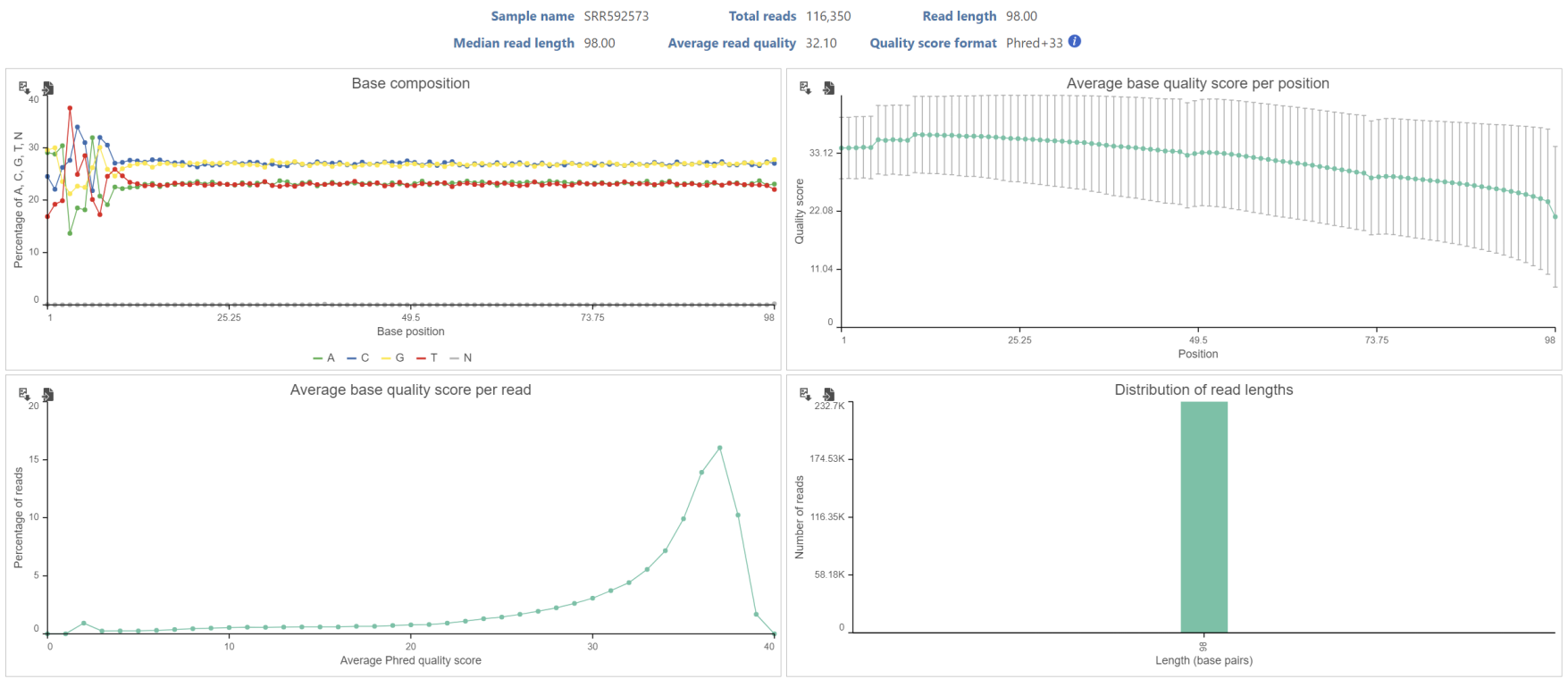
Rows per page 25 (1 of 1)

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

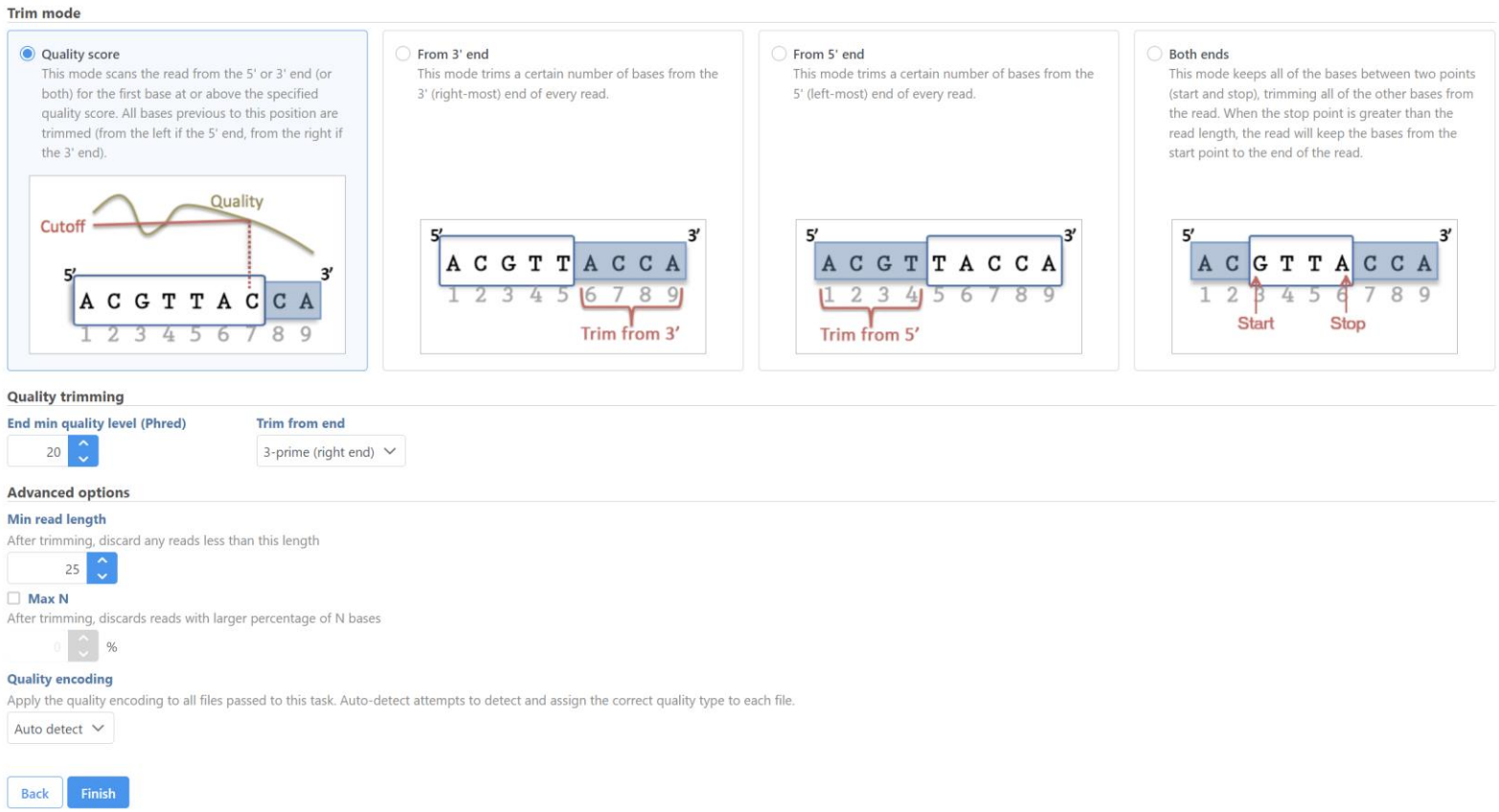


Trim Bases

- 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**



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

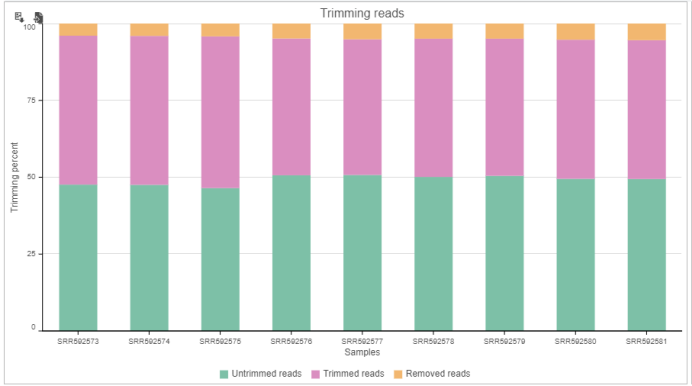
Trim Bases

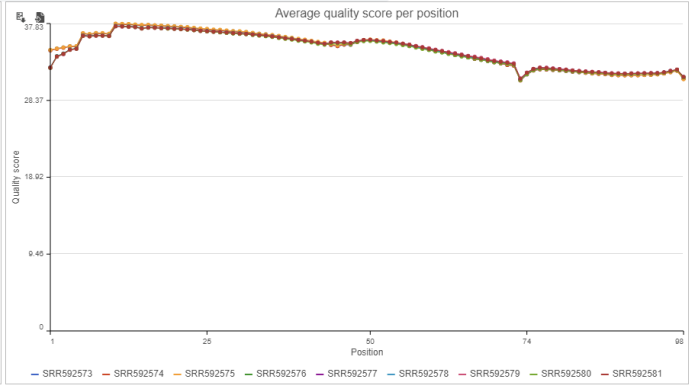
- Double-click **Trimmed reads** data node to open the task report

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)

Download

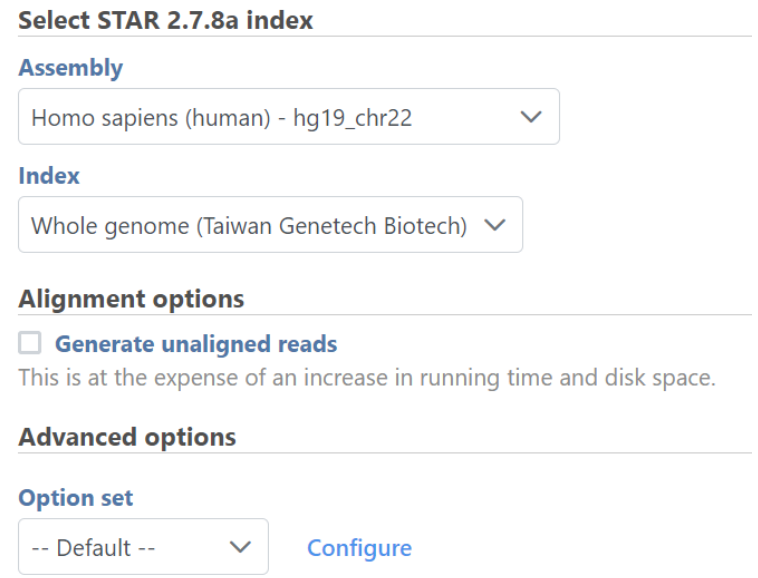




均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Alignment

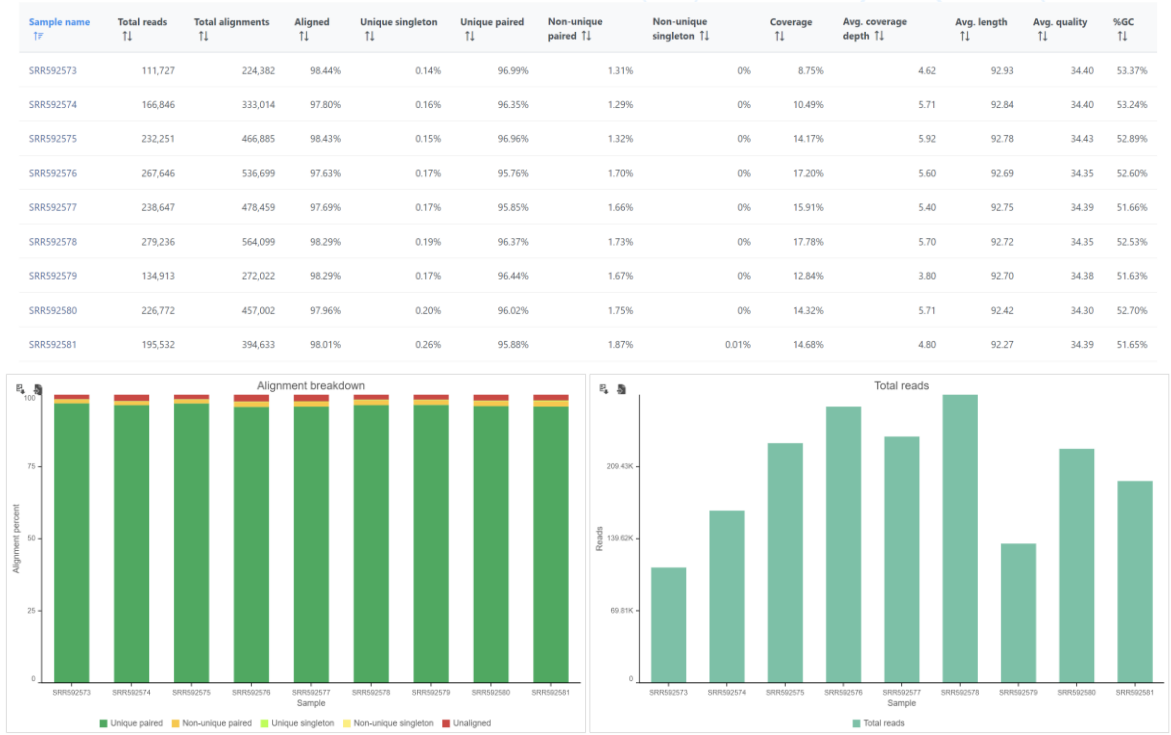
- Click the project name to return to the Analyses tab
- Click **Trimmed reads**
- Click **STAR** from Aligners
- Choose **Homo sapiens - hg19_chr22** for Assembly and **Whole genome** for Index
- Click **Finish**



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%

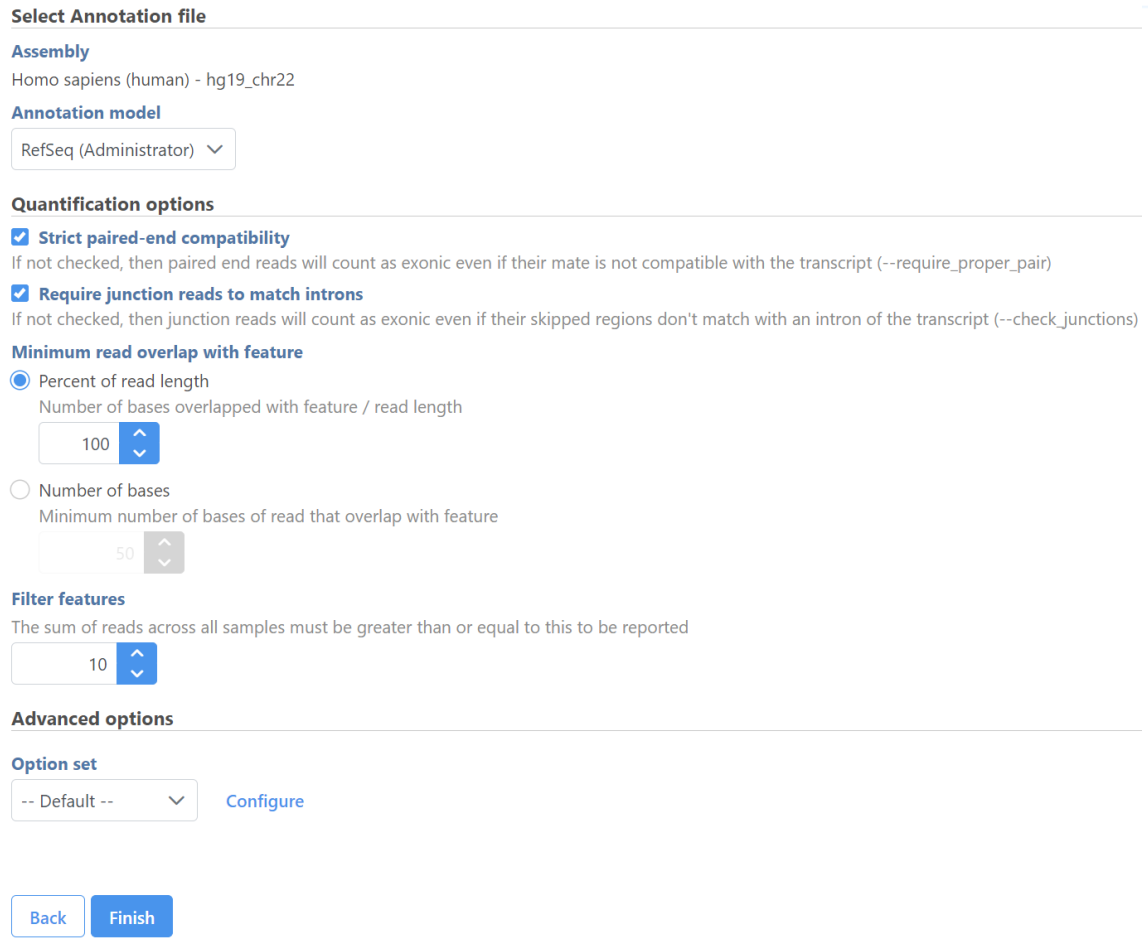
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

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)**

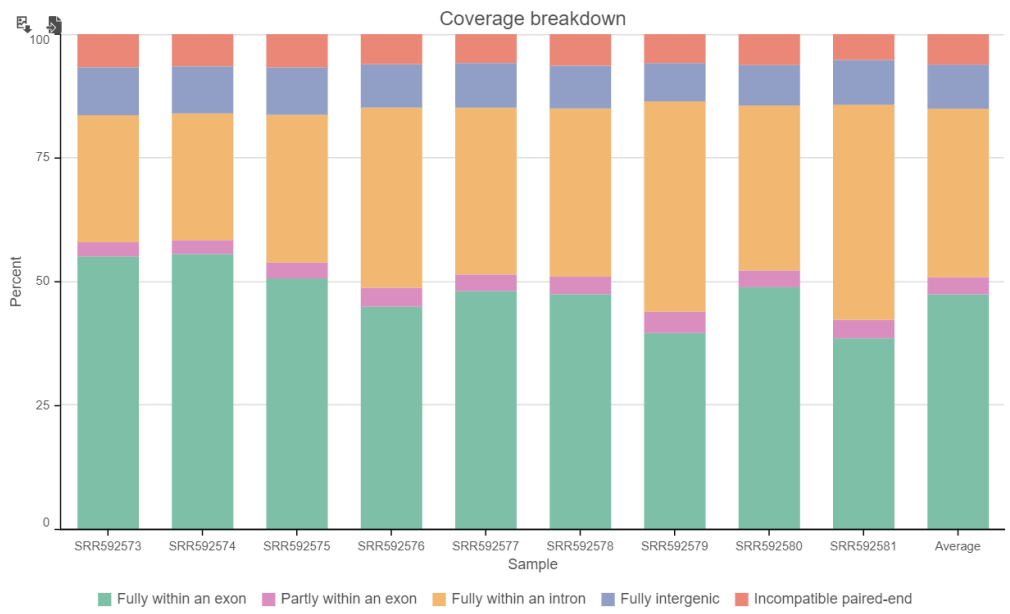
Quantification

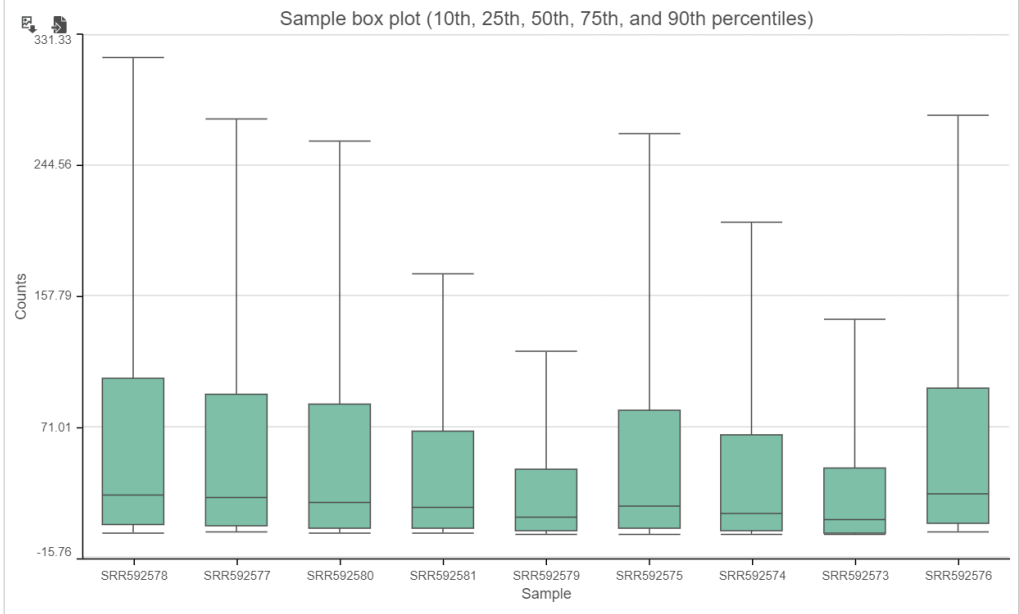
- Choose the **RefSeq** for Annotation model
  - You may need to download it first, via Library File Management
- Click **Finish**



Quantification

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





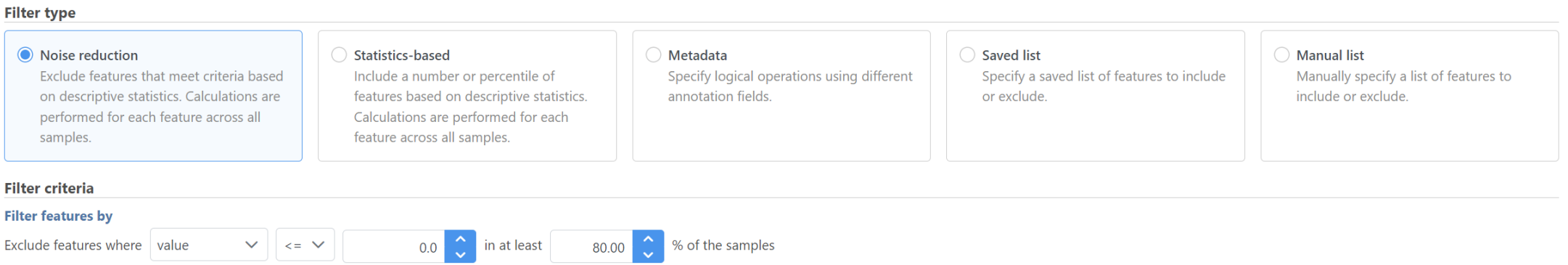
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

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**

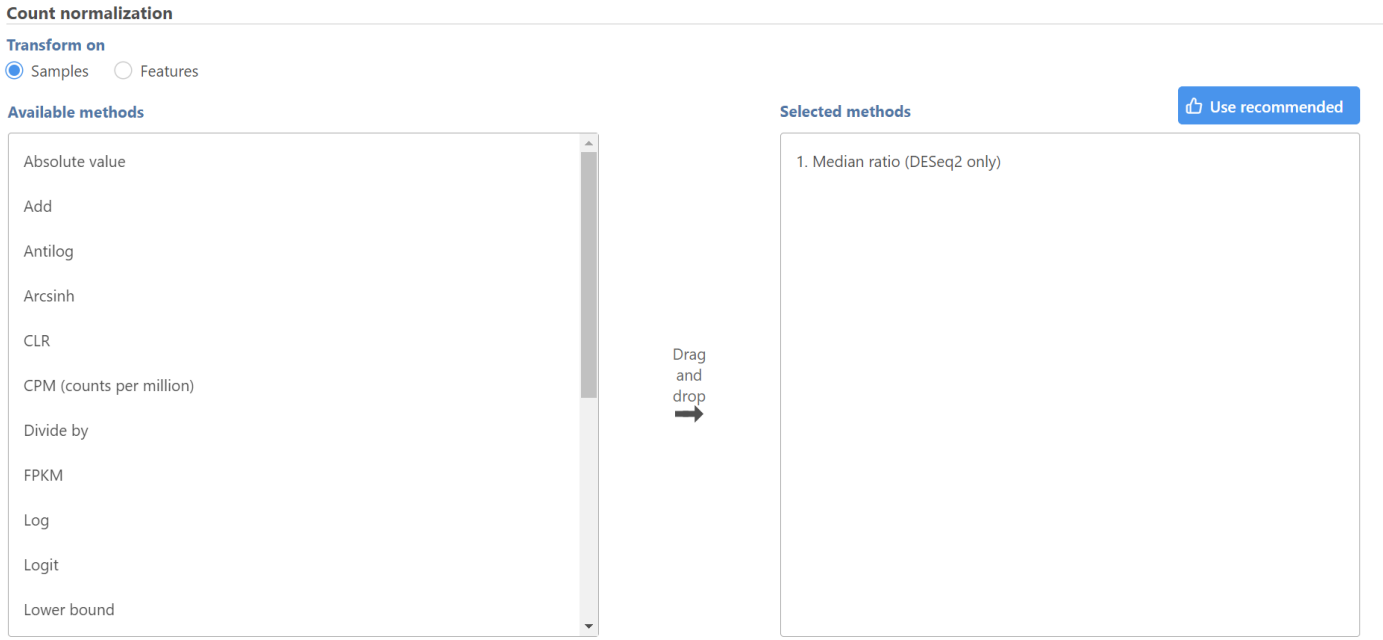
Filter Features

- Click the **Noise reduction filter** checkbox
- Set the filter to **Exclude features where value <= 0 in 80% 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**
- Select **Recommended** if no preferred methods
- Click **Finish**



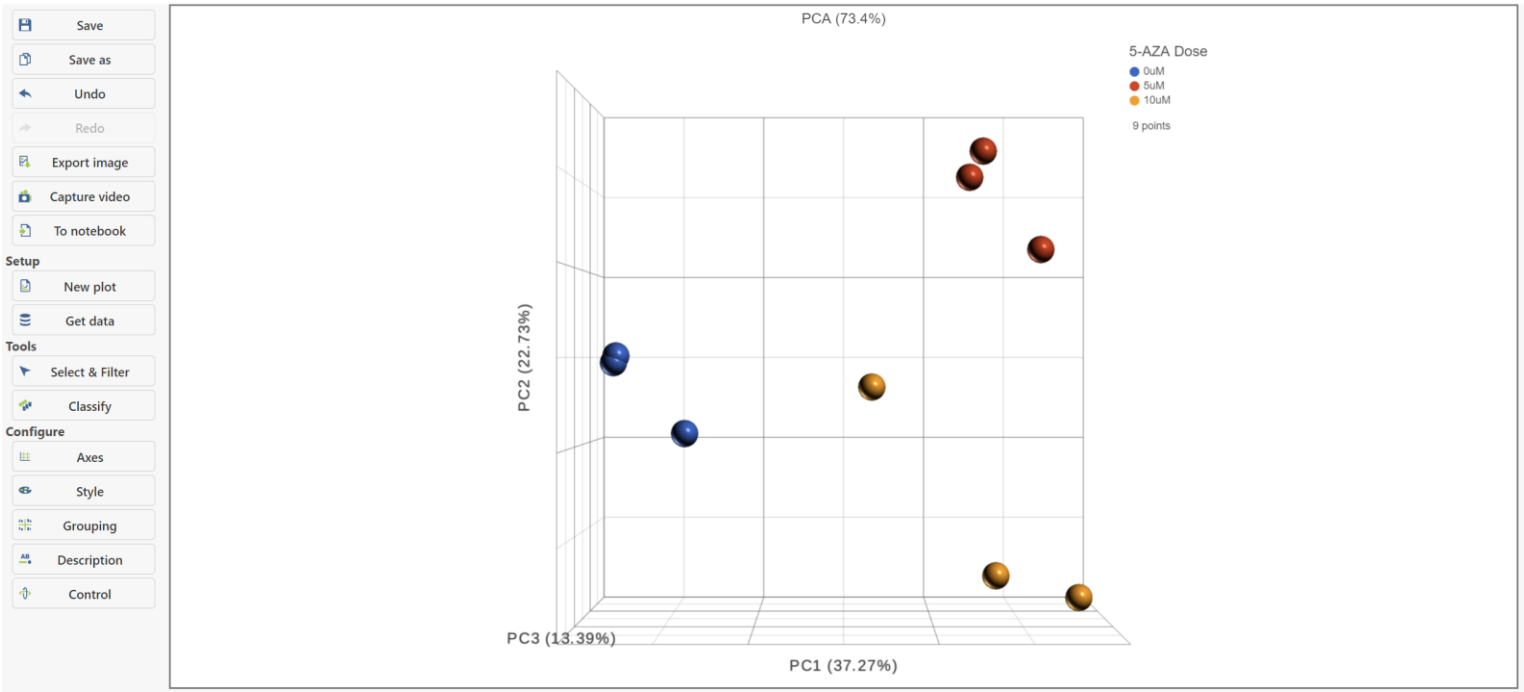
Principal Components Analysis (PCA)

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

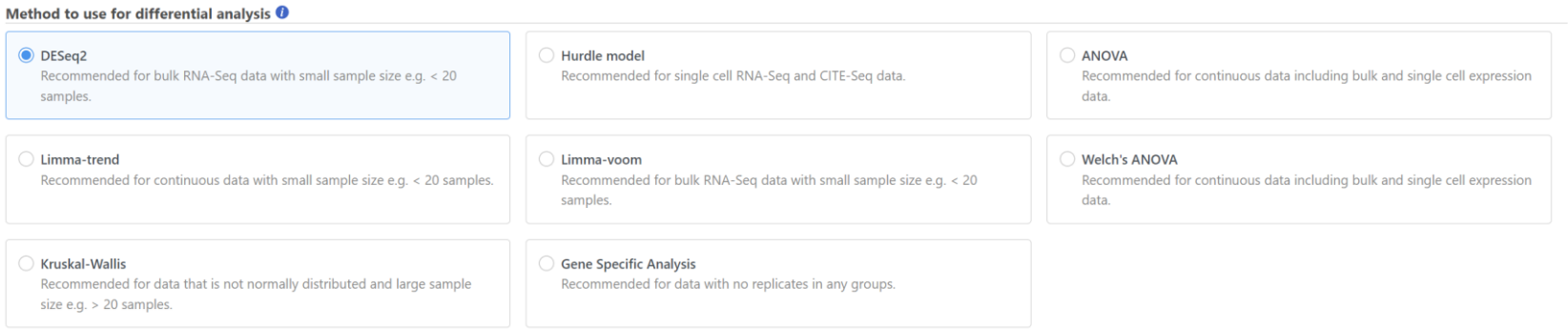
PCA

- 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**



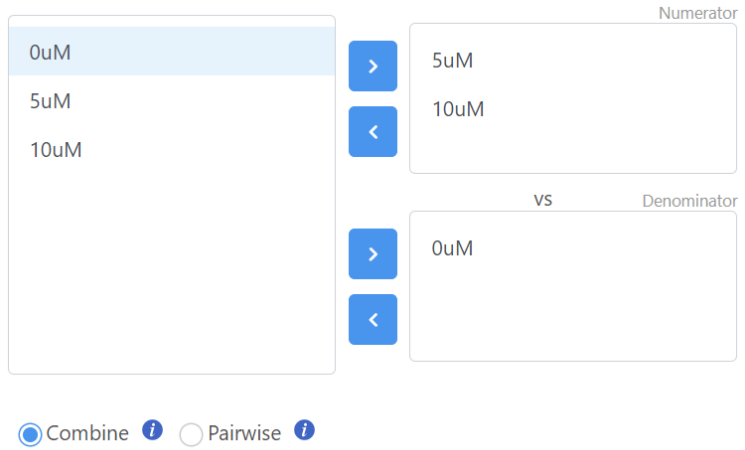
Differential Analysis

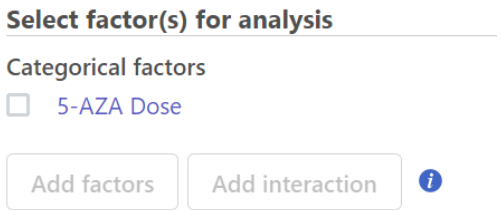
- Click **Normalized counts** node
- Click **Statistics** in the task menu
- Click **Differential analysis** in the task menu
- Select a differential analysis method
- In this tutorial we are going to use **DESeq2**

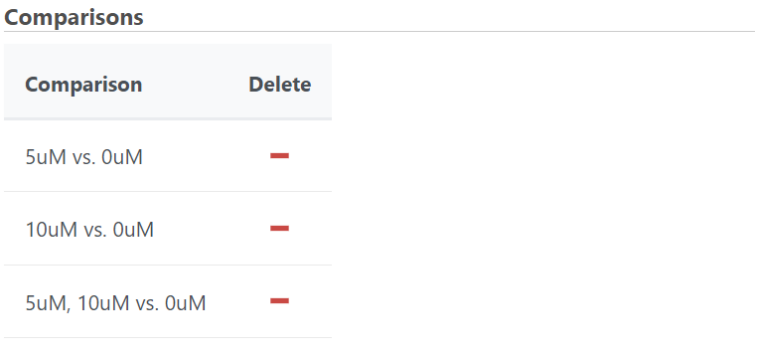


Differential Analysis

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



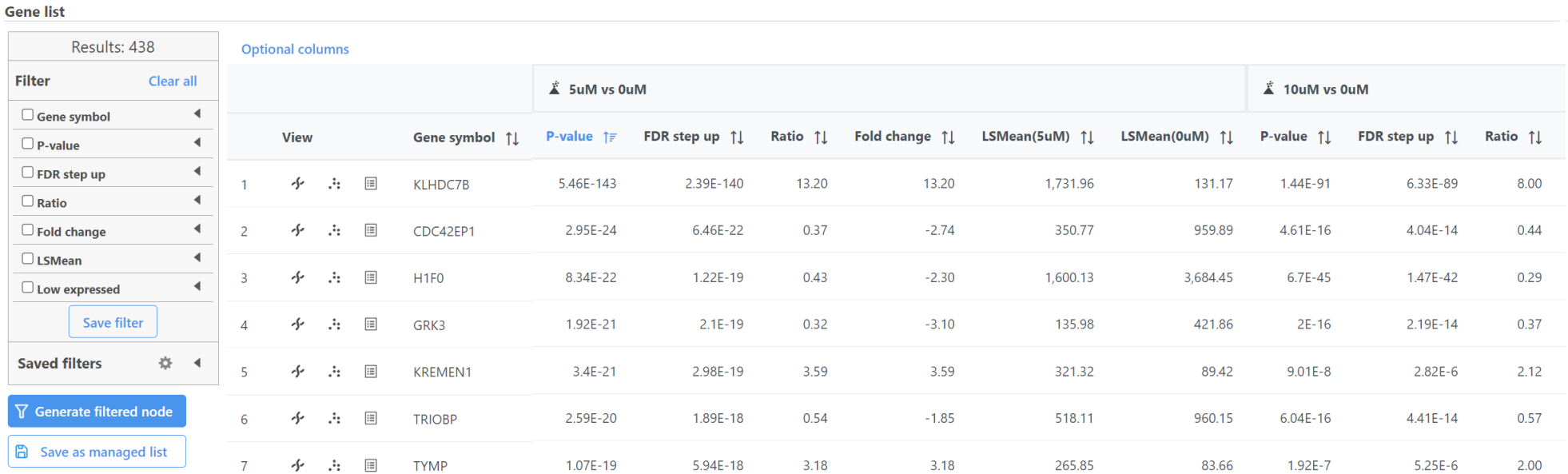




均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Differential Analysis

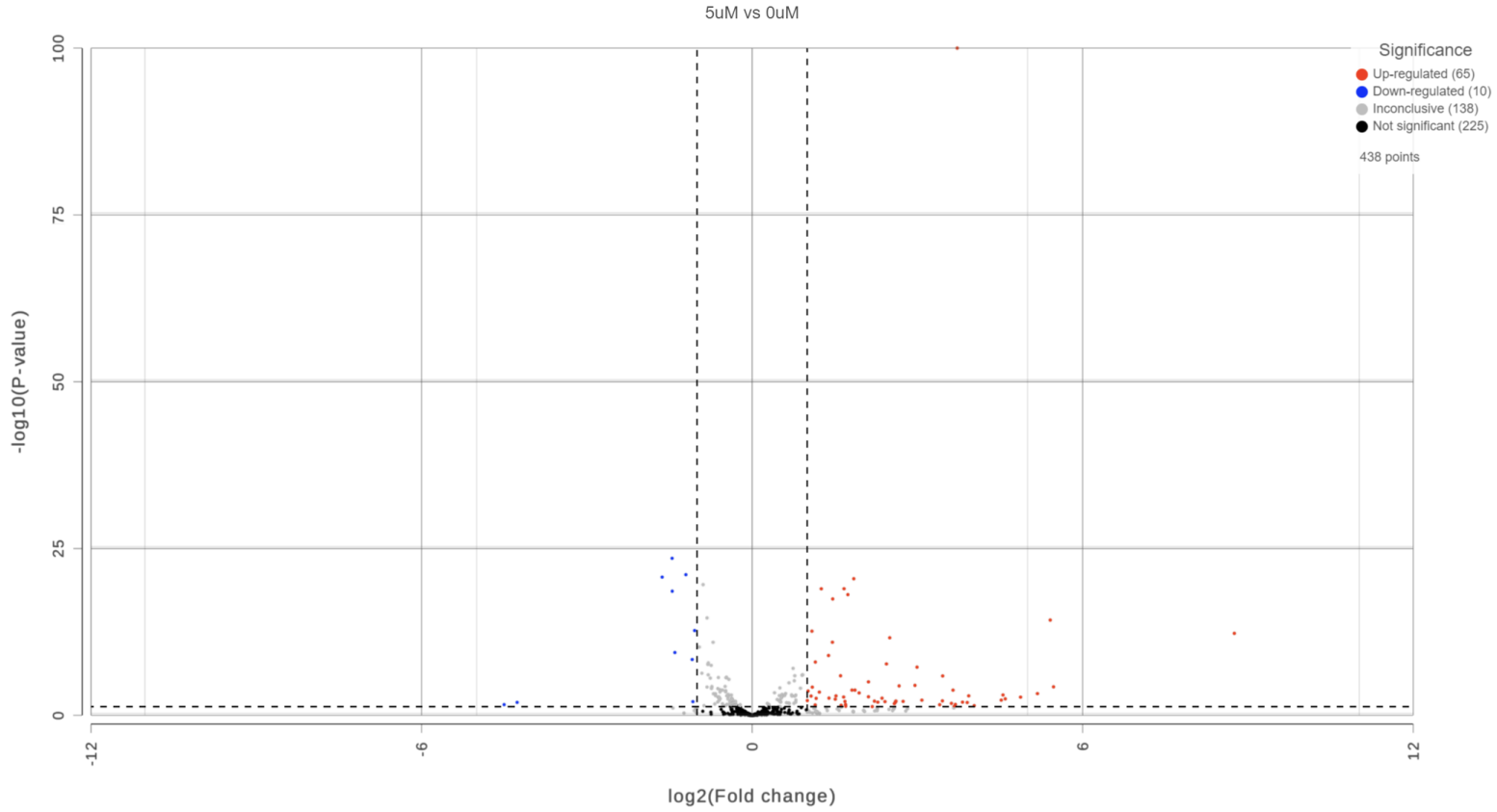
- Double click **DESeq2** data node to open the task report



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

Volcano Plot for All Genes

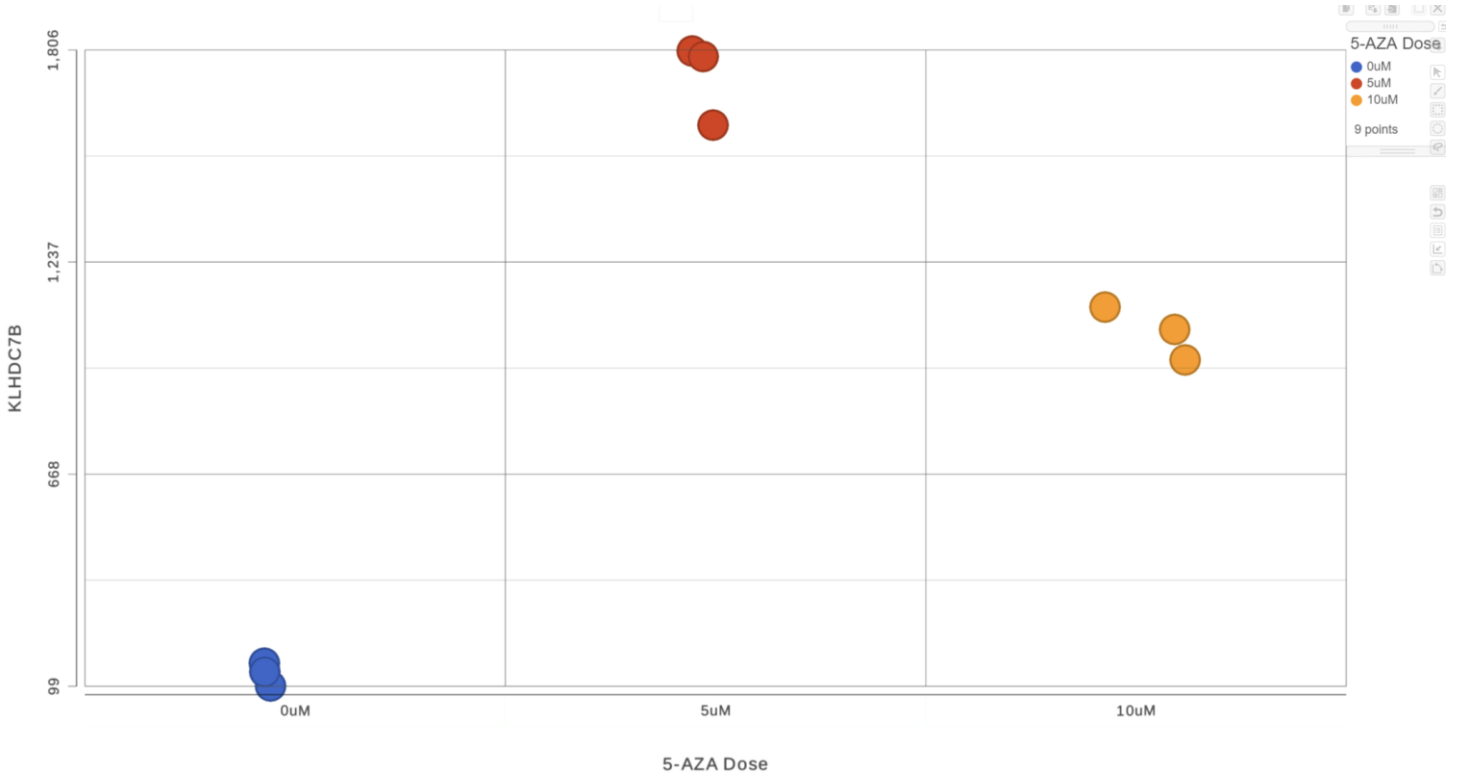
- Click  next to the 5uM vs. 0uM comparison



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

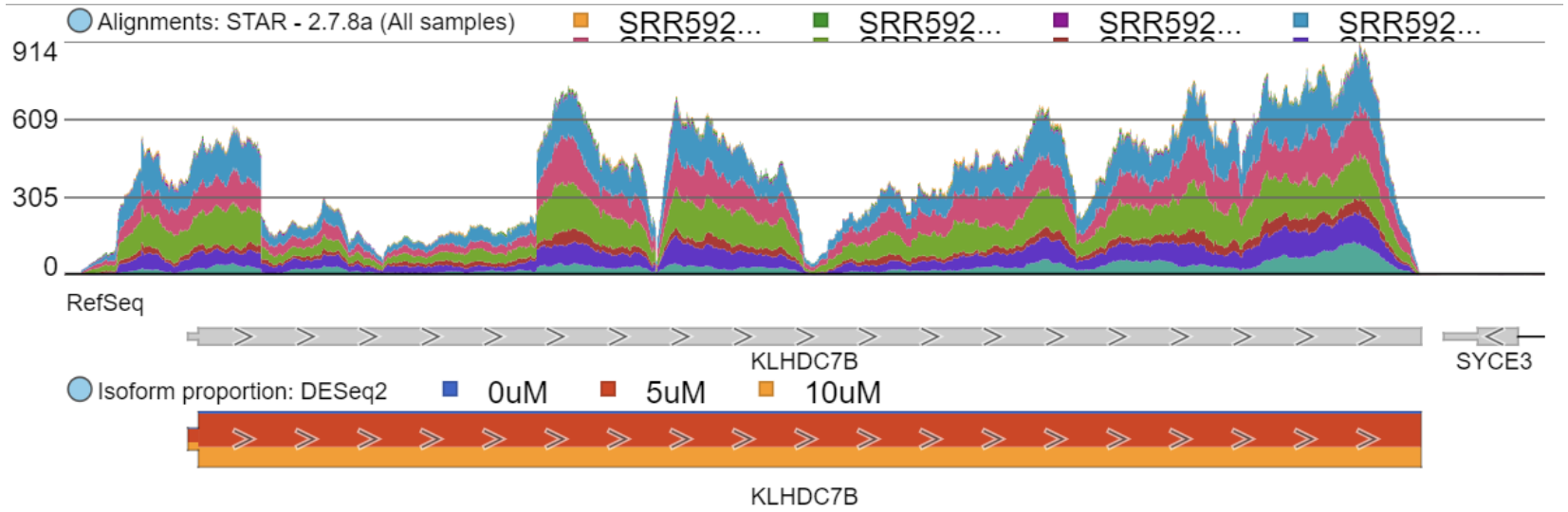
Dot Plot for One Gene

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



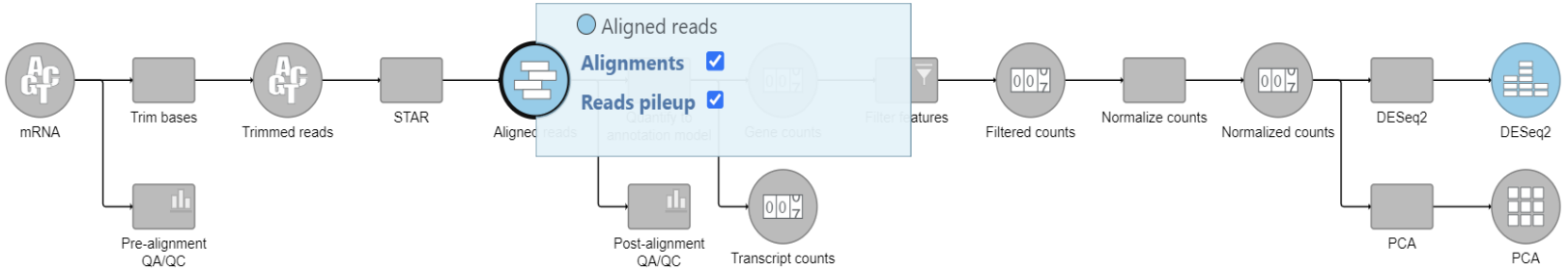
Browse Gene in Chromosome View

- Click  next to a gene symbol to open Chromosome view



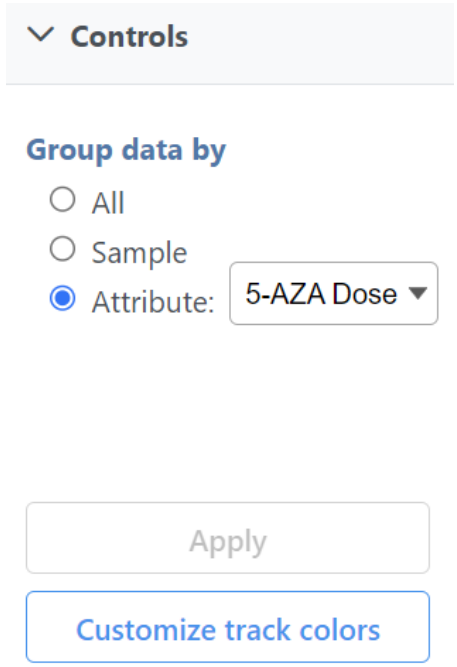
Browse Gene in Chromosome View

- 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

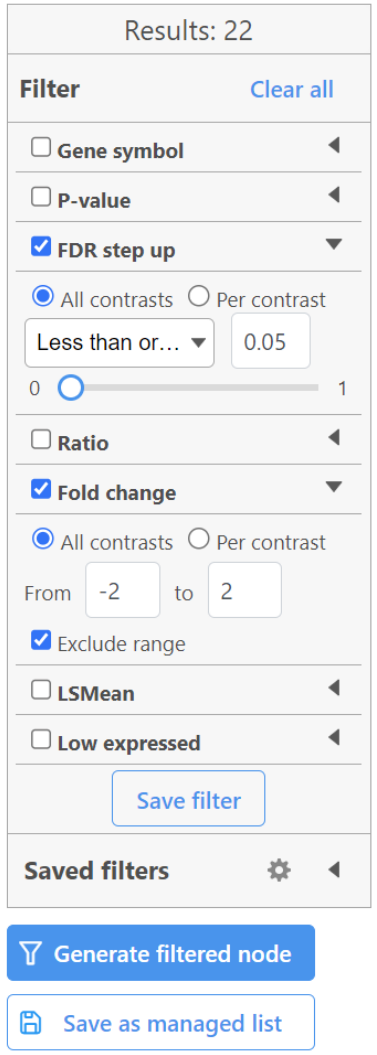
- Group data by **Attribute** and choose **5-AZA**
- Click **Apply**



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



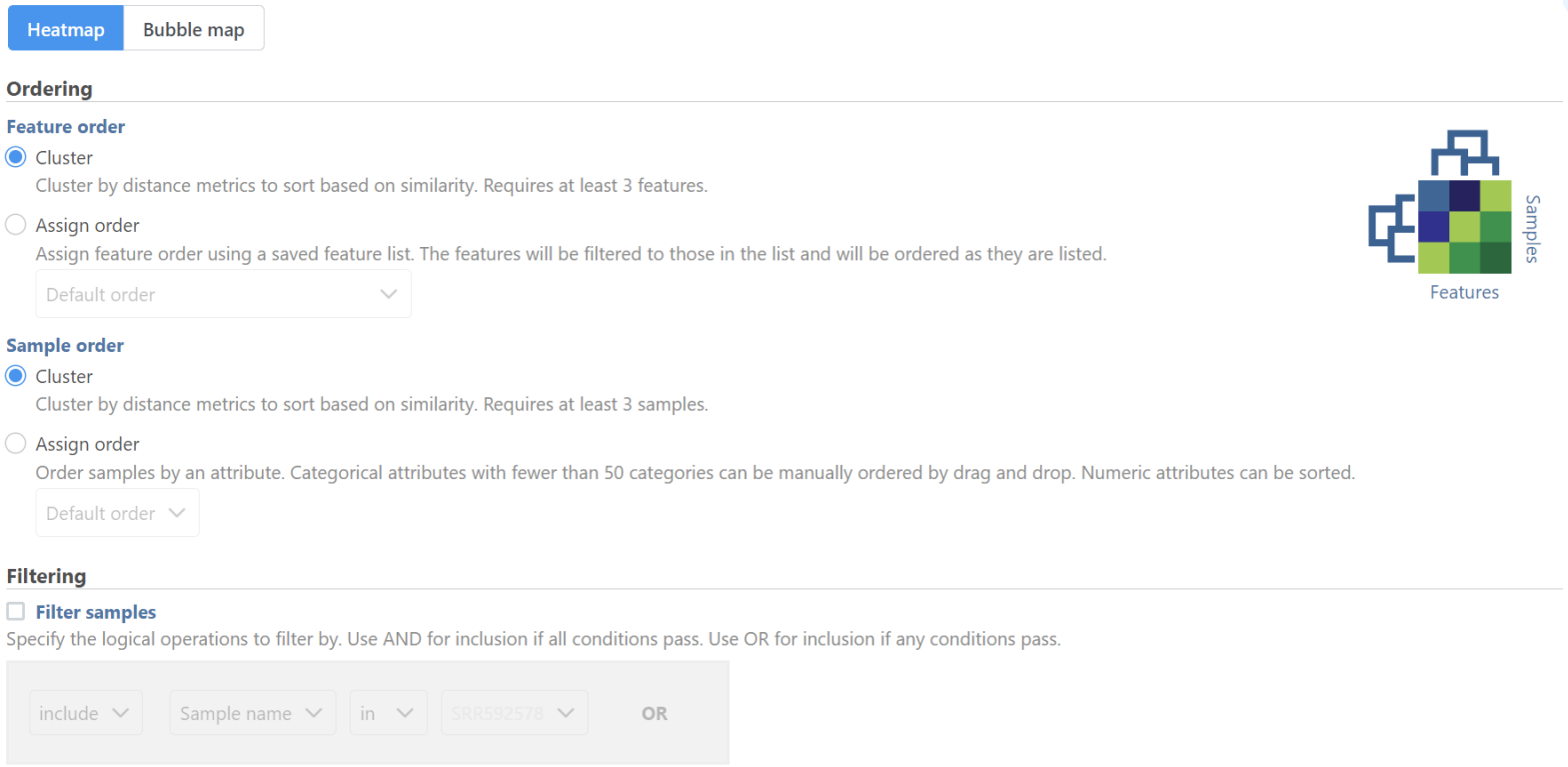
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

Hierarchical Clustering

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

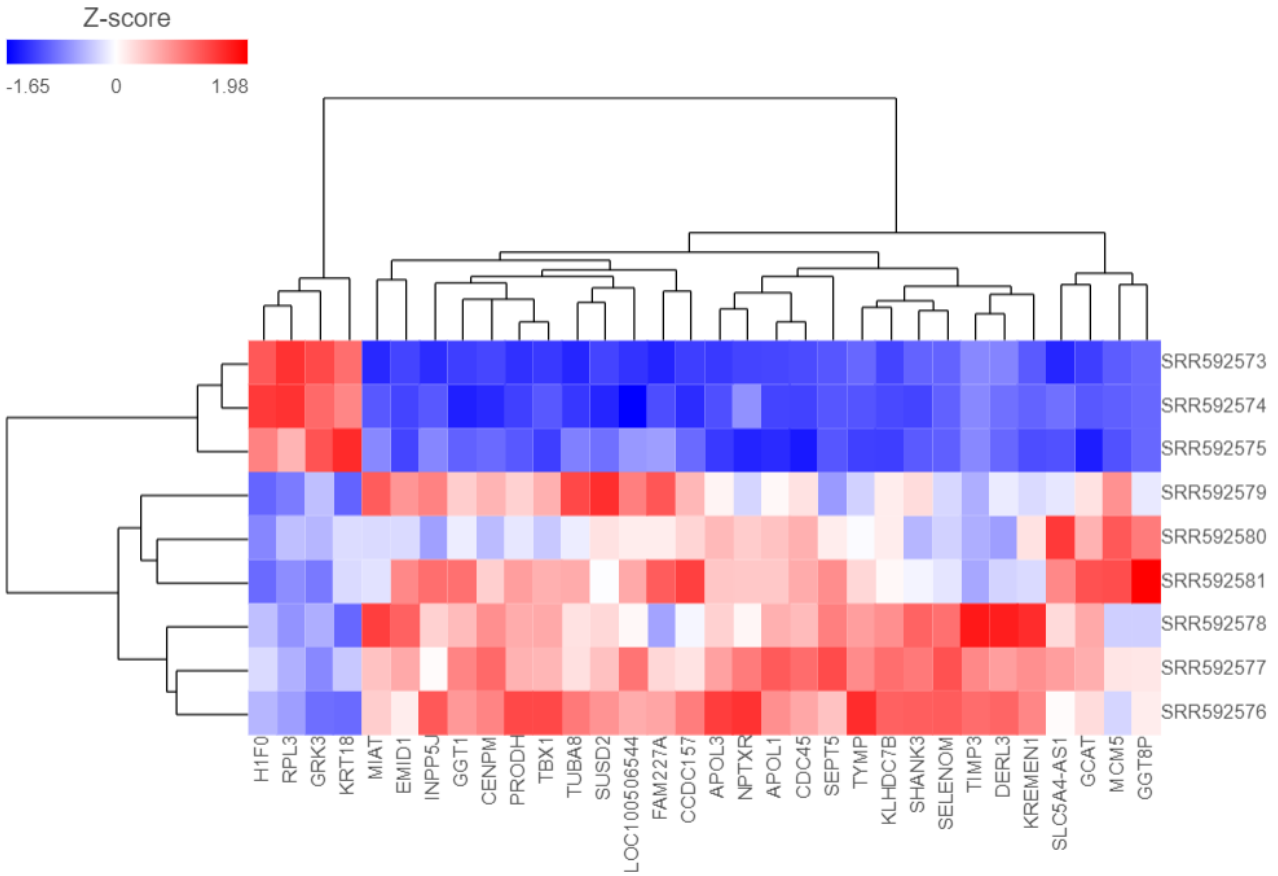
Hierarchical Clustering & Heatmap

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



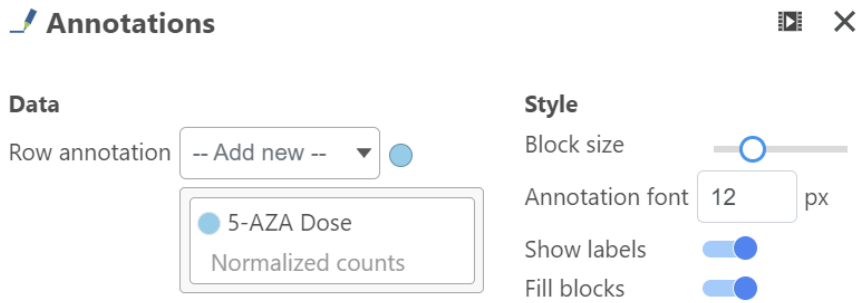
Hierarchical Clustering & Heatmap

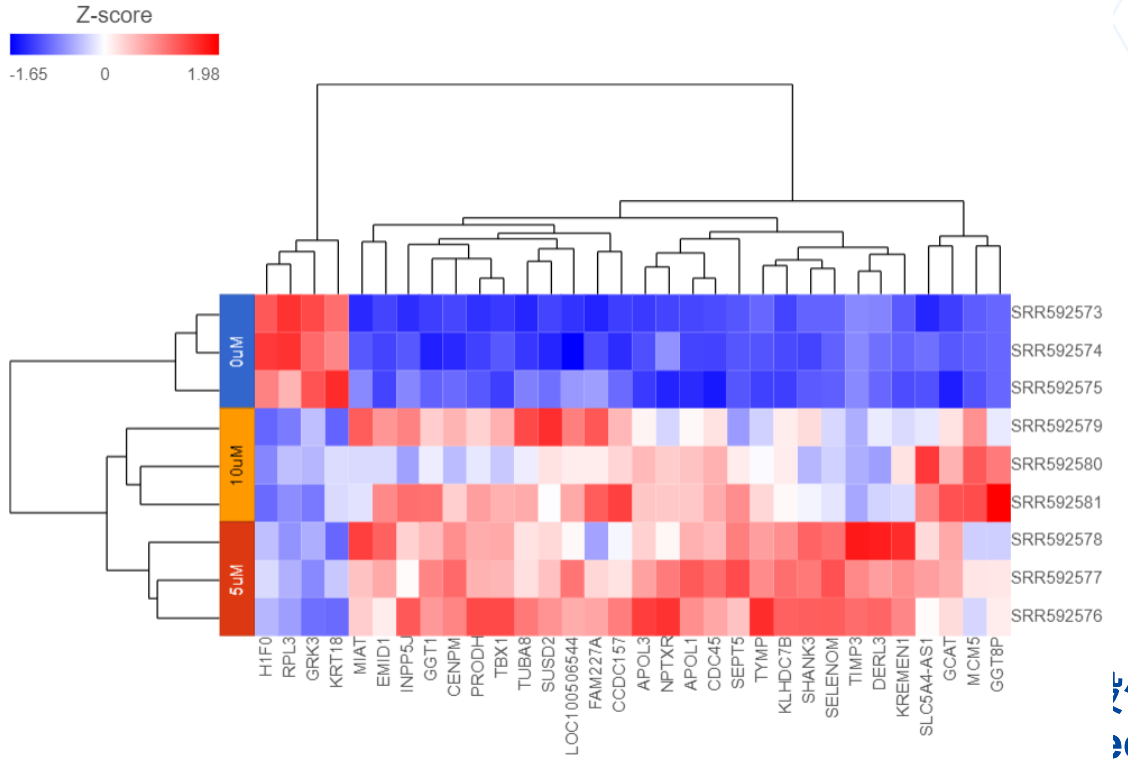
- Double-click the **Hierarchical clustering / heatmap** task node



Hierarchical Clustering & Heatmap

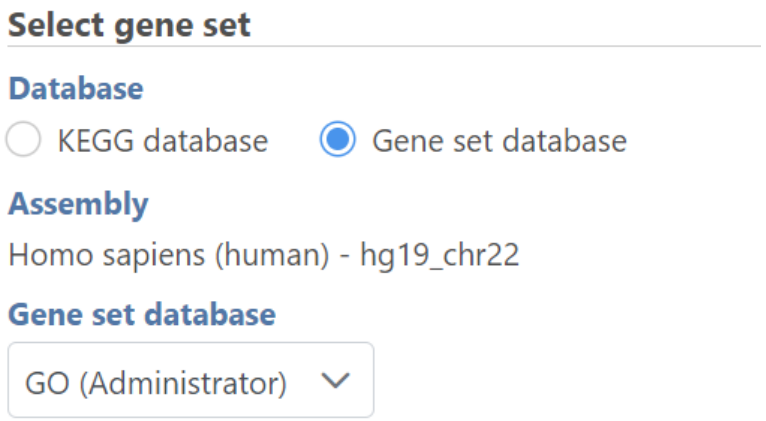
- Expand the **Annotations** card
- Set the Row annotation to **5-AZA Dose**





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



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

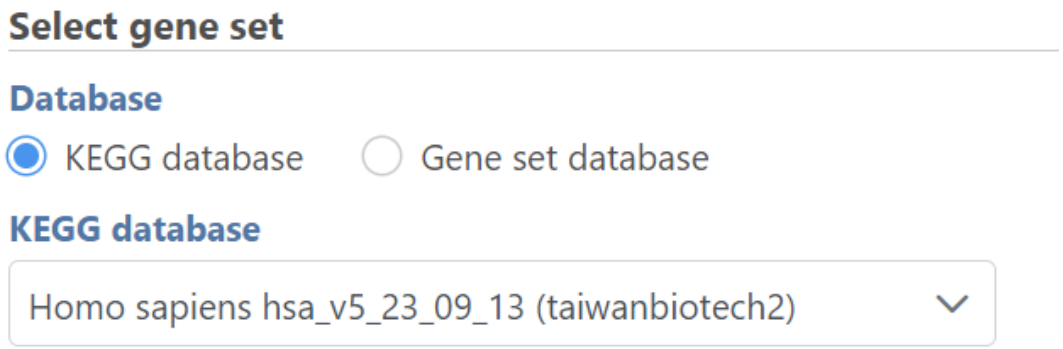
Enrichment Analysis

- 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**



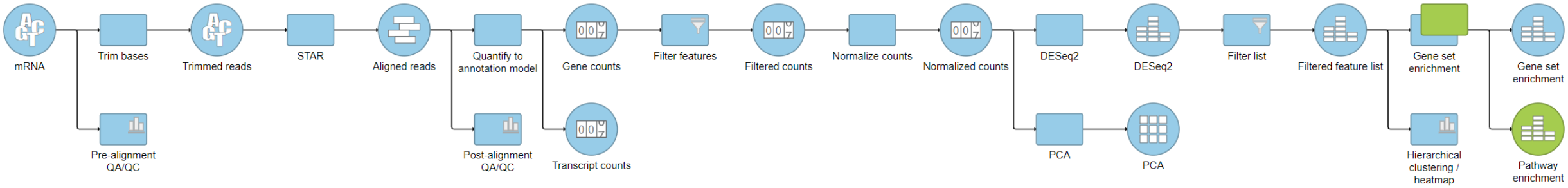
均泰生物科技股份有限公司
Genetech Biotech Co., Ltd.

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



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