

# CLC Genomics Workbenchを用いた RNA-seq解析

フィルジェン株式会社 バイオインフォマティクス部  
( support@filgen.jp )



シーケンスデータ・メタデータの  
インポート

Import

クオリティチェック

QC for Sequencing Reads  
Trim Reads

発現量計算

RNA-seq Analysis

サンプル間比較・図の作成

Differential Expression for RNA-seq  
Create Venn Diagram for RNA-seq  
PCA for RNA-seq など

シーケンスデータ・メタデータの  
インポート

Import

クオリティチェック

QC for Sequencing Reads  
Trim Reads

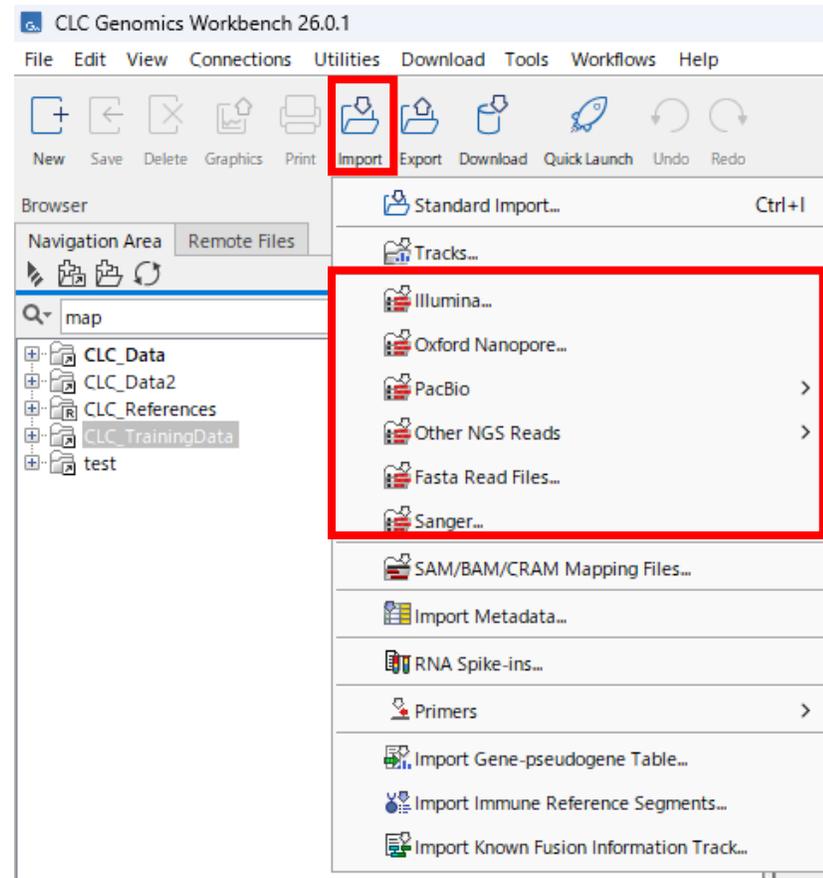
発現量計算

RNA-seq Analysis

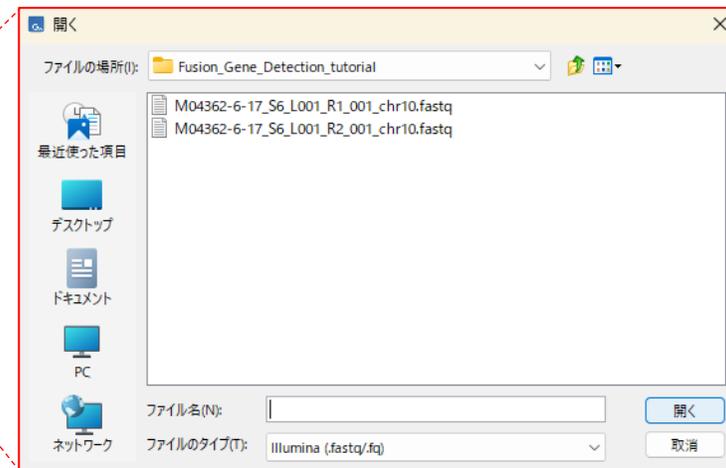
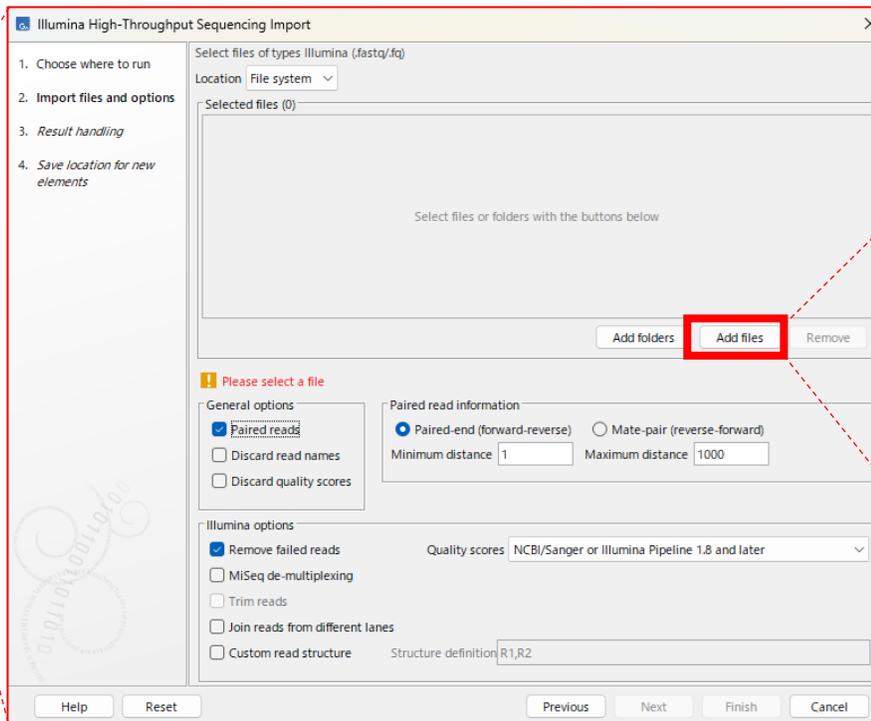
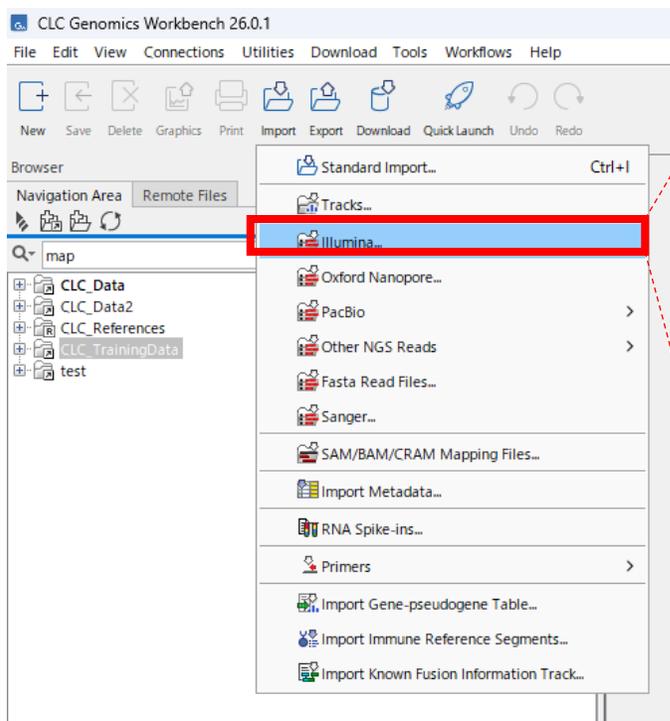
サンプル間比較・図の作成

Differential Expression for RNA-seq  
Create Venn Diagram for RNA-seq  
PCA for RNA-seq など

- ✓ CLC Genomics Workbenchでは、シークエンサー機種やに合わせたインポートメニューを利用可能
- ✓ ToolbarのImportアイコンから表示されるインポーターから選択して、インポートを実行



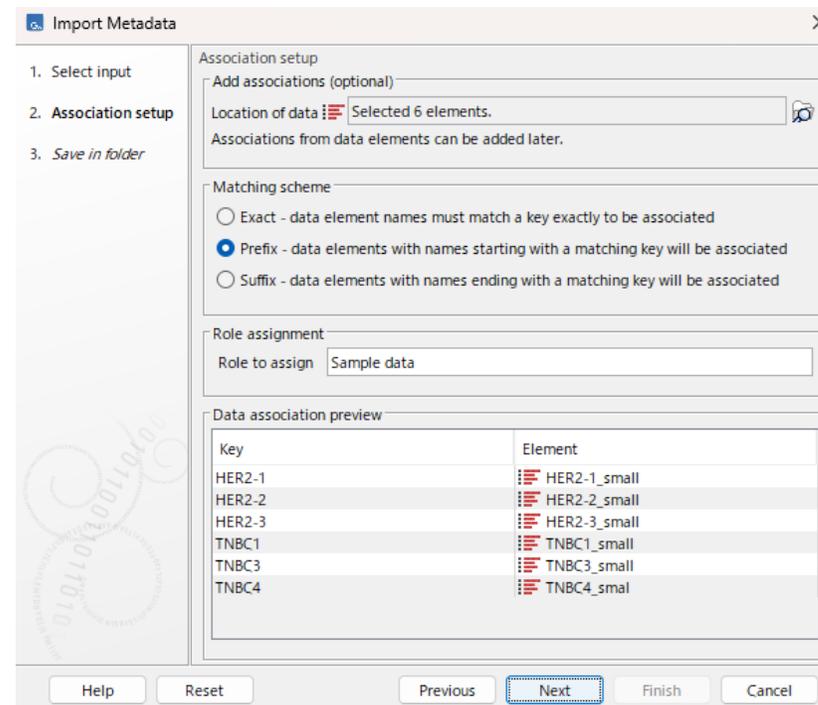
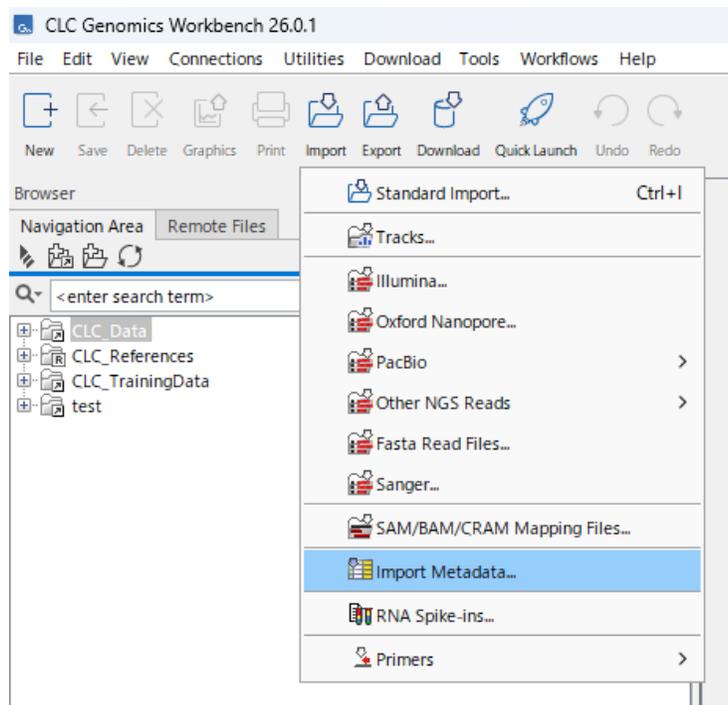
- ✓ シーケンサー機種などに合わせてインポーターを選択し、シークエンスデータファイルを選択
- ✓ ペアエンドシークエンスデータのインポートにも対応



- ✓ シークエンスデータがインポートされ、各種解析に使用できるようになる
- ✓ 各リードの塩基配列やクオリティスコアを確認できる

The screenshot displays the CLC Genomics Workbench 26.0.1 interface. The main window shows a list of sequence reads with their corresponding base sequences and quality scores. The reads are organized into a grid, with each row representing a different read. The quality scores are visualized as a bar chart below each sequence. The interface includes a navigation pane on the left, a toolbar at the top, and a settings panel on the right. The settings panel is currently open, showing options for sequence layout, alignment, and annotation. The sequence layout settings are set to 'No spacing', 'Numbers on sequences' is checked, and 'Numbers on plus strand' is also checked. The alignment settings are set to 'Align labels' and 'Sequence label' is set to 'Name'. The annotation settings are set to 'Show annotations' and 'Use gradients' is checked. The sequence label is set to 'Name' and the position is set to 'Next to sequence'. The offset is set to 'Little offset' and the label is set to 'Stacked'. The restriction sites, motifs, residue coloring, nucleotide info, positional stats, find, and text format options are all unchecked.

- ✓ 後にサンプル間比較用ツールを使用する場合、サンプルのグループ分類やグラフ表示に用いる属性情報などを、Excelファイルなどにメタデータとしてまとめておく必要がある。
- ✓ 作成したメタデータファイルは、先にインポートしておいた各サンプルのシークエンスデータと関連付けてソフトウェアにインポートすることで、シークエンスデータおよびそこから派生する各種解析データに情報が付加され、後の解析に使用できるようになる。



# メタデータのインポート

	A	B	C	D
1	Sample Name	Factor 1 (sample)	Factor 2 (environment)	Factor 3 (RNA prep)
2	HER2-1	HER2		2 Centrifuge
3	HER2-2	HER2		3 Centrifuge
4	HER2-3	HER2		2 Vacuum
5	TNBC1	TNBC		2 Centrifuge
6	TNBC3	TNBC		3 Vacuum
7	TNBC4	TNBC		3 Vacuum

- HER2-1\_small
- HER2-2\_small
- HER2-3\_small
- TNBC1\_small
- TNBC3\_small
- TNBC4\_small

 Import Metadata...

 Metadata

Sample Name	Factor 1 (sample)	Factor 2 (environment)	Factor 3 (RNA prep)
HER2-1	HER2		2 Centrifuge
HER2-2	HER2		3 Centrifuge
HER2-3	HER2		2 Vacuum
TNBC1	TNBC		2 Centrifuge
TNBC3	TNBC		3 Vacuum
TNBC4	TNBC		3 Vacuum

Fixed Fields

- ▼ Name [Edit](#)  
HER2-1\_small
- ▼ Description [Edit](#)
- ▼ Metadata [Refresh](#)
  - ▼ From 'Metadata' [Find](#) [Refresh](#) [Delete](#) [Edit](#)  
This Sequence List is 'Sample data' for:  
Sample Name : HER2-1  
Factor 1 (sample) : HER2  
Factor 2 (environment) : 2  
Factor 3 (RNA prep) : Centrifuge
- ▼ File info [Refresh](#)  
File size: 41.5 MB  
Compression enabled: No
- ▼ Read group [Edit](#)  
ID: 33570caf-1283-4d23-beb3-13afc94bb9e6  
Sample: SRR1027183\_1  
Platform: Illumina
- ▼ Paired status [Edit](#)  
Paired sequences  
Minimum distance: 0, Maximum distance: 557  
Read orientation: Forward Reverse

シーケンスデータ・メタデータの  
インポート

Import

クオリティチェック

QC for Sequencing Reads  
Trim Reads

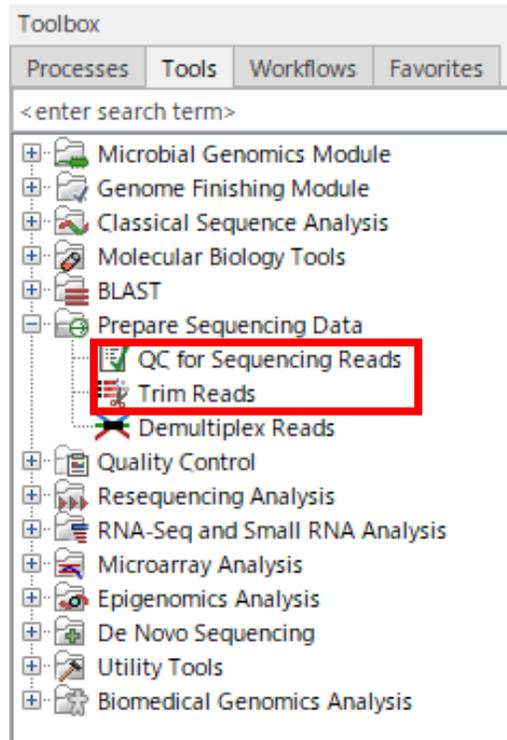
発現量計算

RNA-seq Analysis

サンプル間比較・図の作成

Differential Expression for RNA-seq  
Create Venn Diagram for RNA-seq  
PCA for RNA-seq など

- ✓ インポートしたシーケンスデータに対して、クオリティチェックレポートの作成や、低クオリティリードの除去などを行う
- ✓ その他、重複リードの除去や、マルチプレックスシーケンス時のサンプルバーコードのソートなどの、各種データ前処理用ツールなども利用が可能



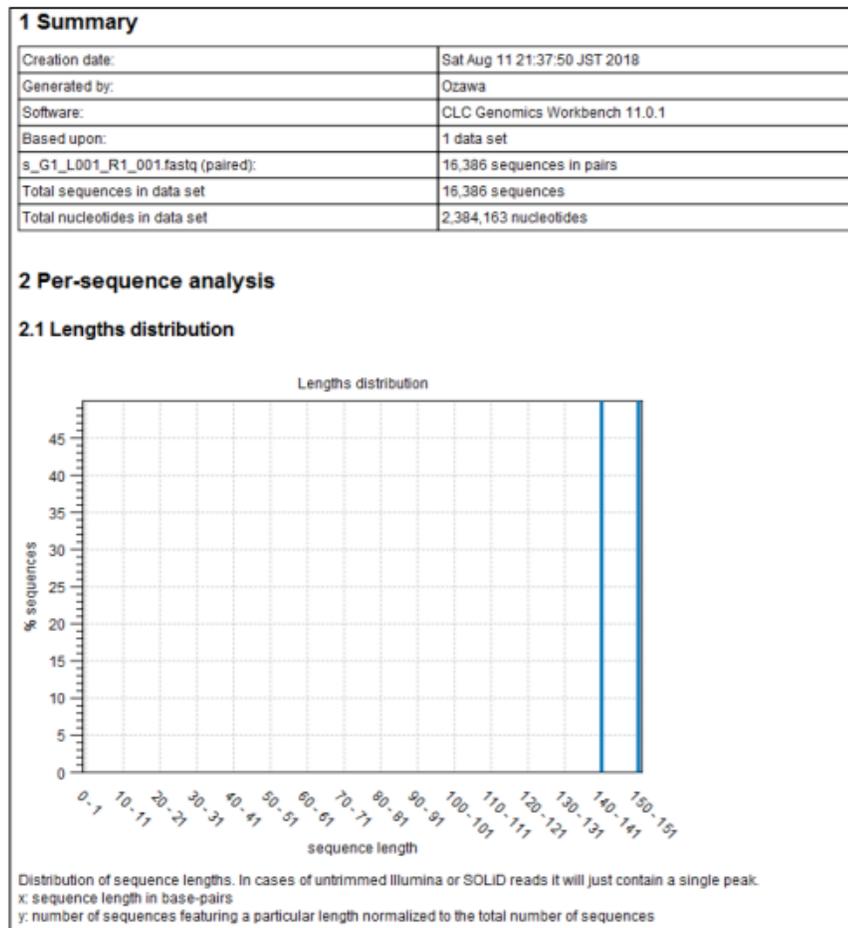
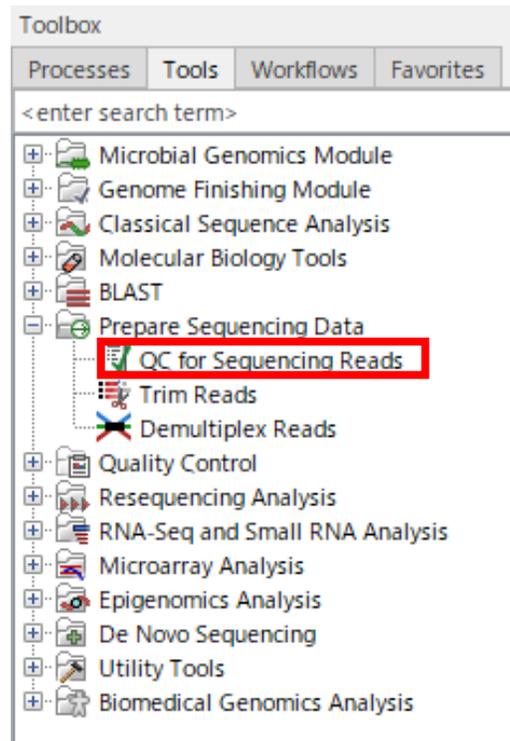
## QC for Sequencing Reads

インポートしたシーケンスデータのクオリティやPCR Duplicateの状況を確認するためのレポートを作成

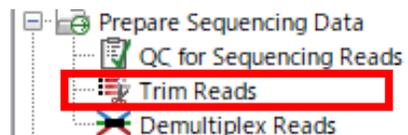
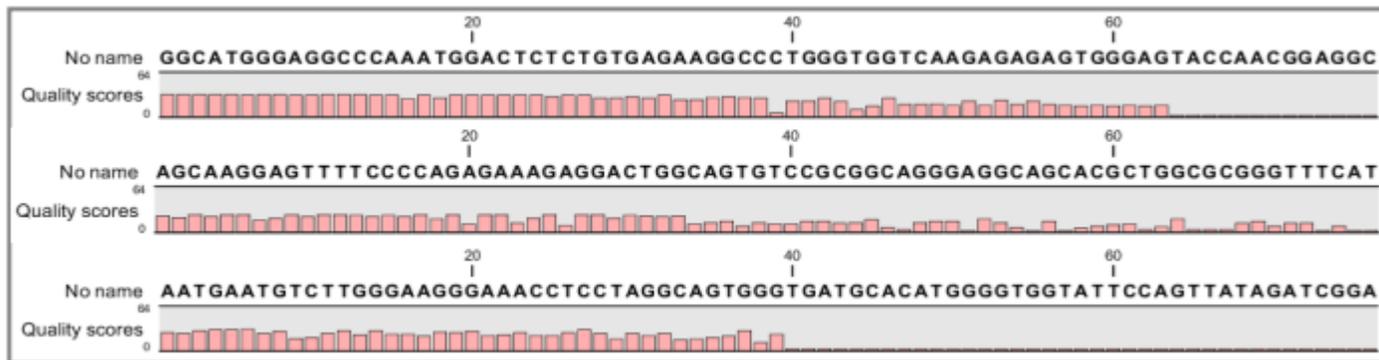
## Trim Reads

アダプターの除去、クオリティスコアによる除去、長さを指定した除去などを選択・組み合わせて、リードのトリミングを実行

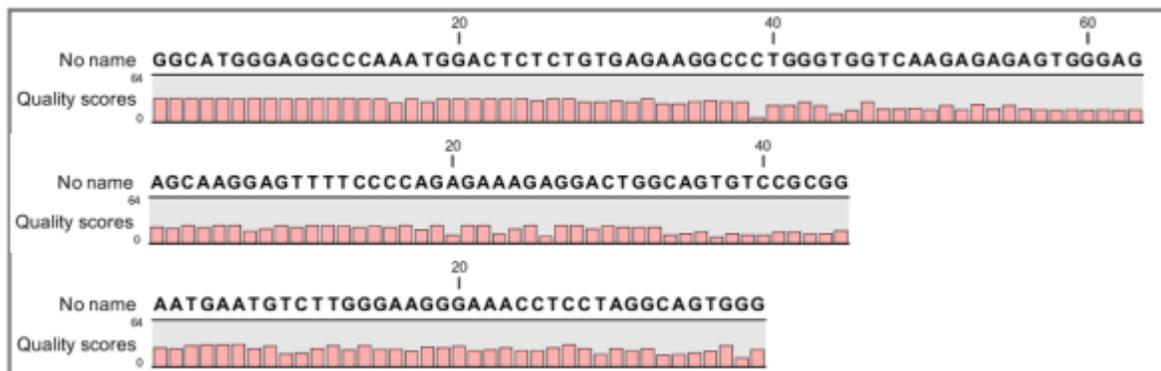
- ✓ QC for Sequencing Readsでは、シーケンスデータのクオリティ情報をまとめたレポートが作成される
- ✓ GC含量やクオリティスコア分布などのグラフデータや数値データを確認が可能



# リードのQCとトリミング



- ✓ Trim Readsの使用により、各リードの低クオリティ部分がカットされる
- ✓ アダプター配列の除去なども可能



シーケンスデータ・メタデータの  
インポート

Import

クオリティチェック

QC for Sequencing Reads  
Trim Reads

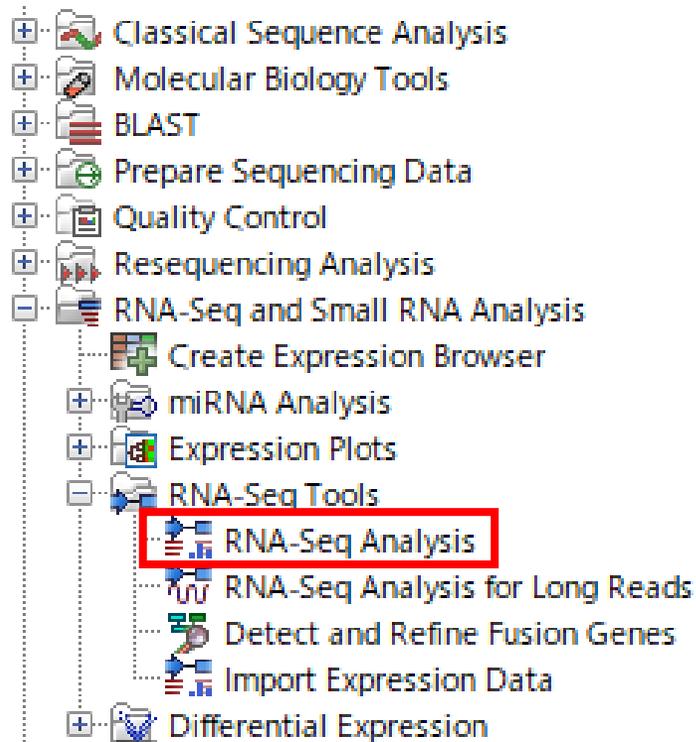
発現量計算

RNA-seq Analysis

サンプル間比較・図の作成

Differential Expression for RNA-seq  
Create Venn Diagram for RNA-seq  
PCA for RNA-seq など

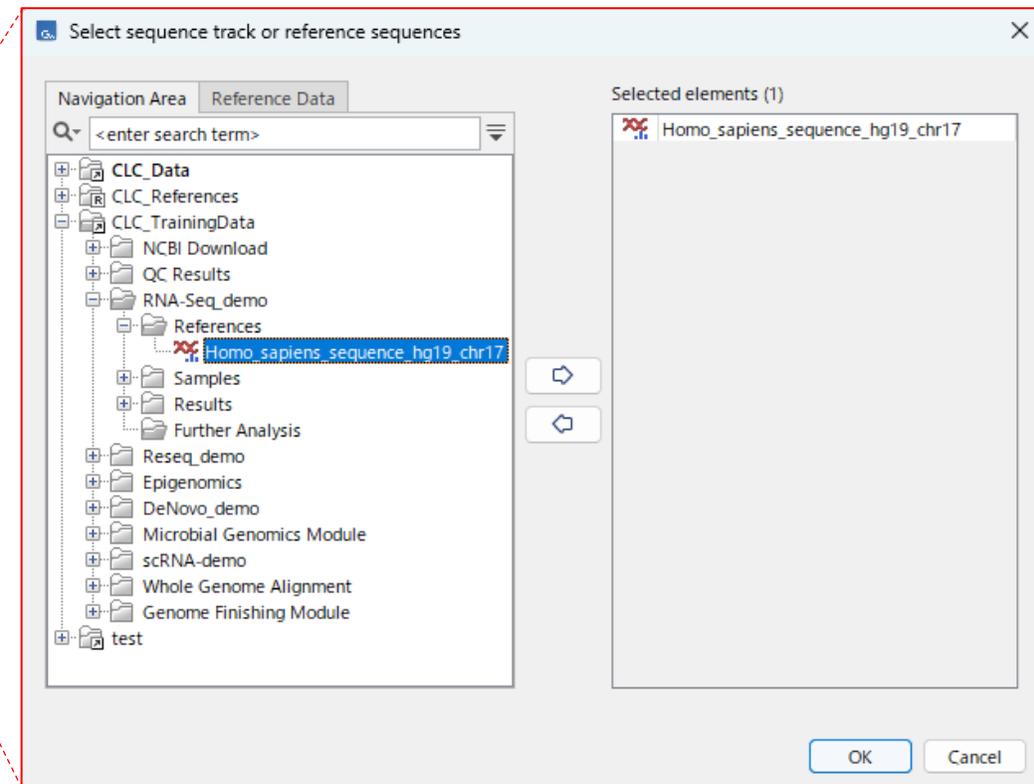
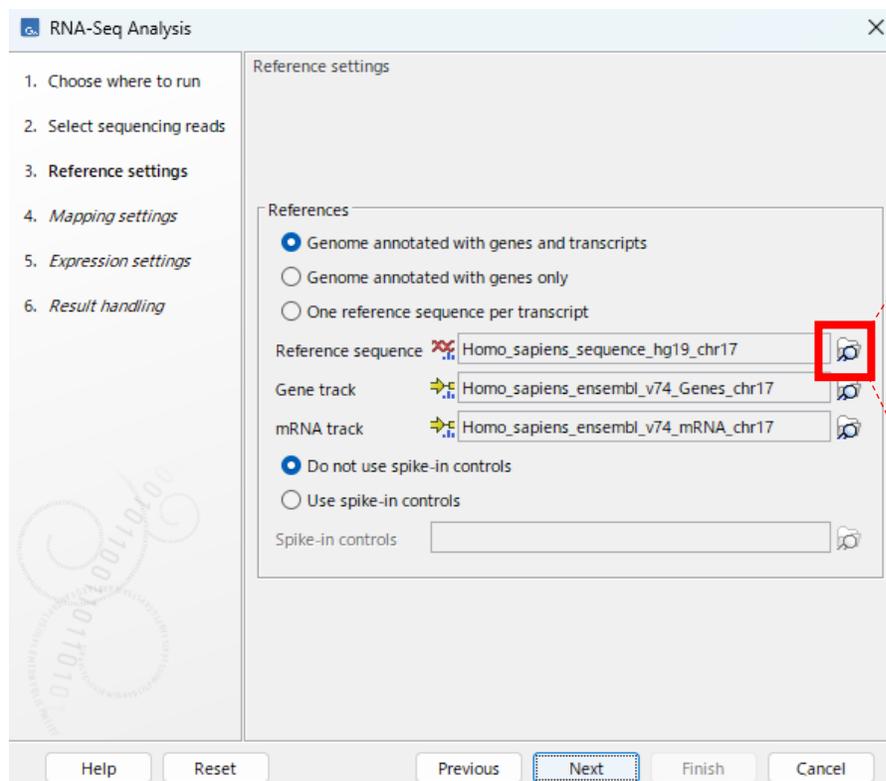
- ✓ サンプルのシーケンスデータを参照ゲノムあるいは遺伝子配列にマッピングを行い、遺伝子ごとの発現量のカウントを行う
- ✓ 遺伝子ごとあるいは転写物ごとの発現量データの外、マッピングデータや融合遺伝子候補のデータも得ることができる



## RNA-seq Analysis

任意の参照ゲノムや遺伝子配列に対して、シーケンスデータのマッピングを行い、同時に遺伝子発現量の測定を行う

- ✓ RNA-Seq Analysisでは、実行時のオプションパラメータで、任意の参照ゲノム配列および付随するアノテーションデータを選択が可能
- ✓ ヒト、マウス、ラットなどのモデル生物の参照ゲノムデータは、ソフトウェア標準搭載のダウンロードツールから取得でき、その他NCBIに登録されている参照ゲノムデータや、ユーザーカスタム作成の遺伝子配列データを使用することも可能

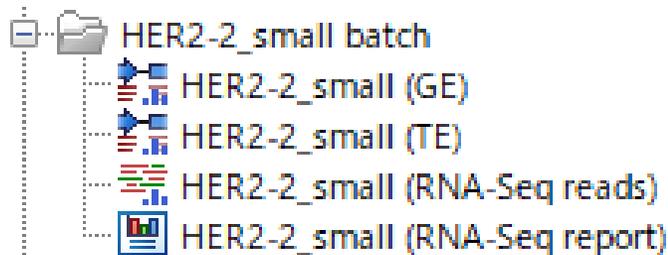


## Gene-Level Expression data

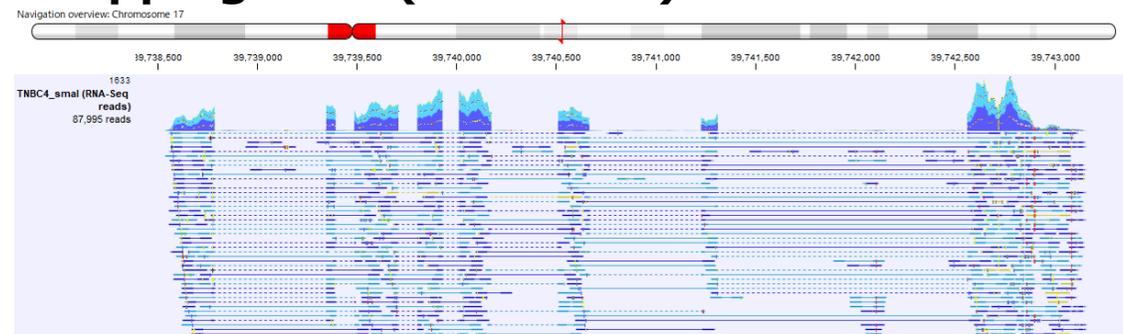
Name	TPM	Exons	Gene ID	ENSEMBL
ACACA	38.33	67	ENSG00000132142	<a href="#">ENSG00000132142</a>
TADA2A	1,261.94	50	ENSG00000108264	<a href="#">ENSG00000108264</a>
RP11-378E13.4	0.00	2	ENSG00000267613	<a href="#">ENSG00000267613</a>
CTC-421K24.1	0.00	2	ENSG00000267562	<a href="#">ENSG00000267562</a>
DUSP14	1,354.15	6	ENSG00000161326	<a href="#">ENSG00000161326</a>
SYNRG	2,747.23	56	ENSG00000006114	<a href="#">ENSG00000006114</a>

## Transcript-Level Expression data

Name	TPM	Gene name	Transcript length	Exons	Gene ID
ACACA_1	0.00	ACACA	9962	56	ENSG00000132142
ACACA_2	0.00	ACACA	1469	10	ENSG00000132142
ACACA_3	59.49	ACACA	1064	5	ENSG00000132142
ACACA_4	35.70	ACACA	1064	5	ENSG00000132142
ACACA_5	0.00	ACACA	1627	8	ENSG00000132142
ACACA_6	52.97	ACACA	717	2	ENSG00000132142
ACACA_7	0.00	ACACA	918	2	ENSG00000132142
TADA2A_1	220.28	TADA2A	4253	16	ENSG00000108264



## Mapping data (オプション)



シーケンスデータ・メタデータの  
インポート

Import

クオリティチェック

QC for Sequencing Reads  
Trim Reads

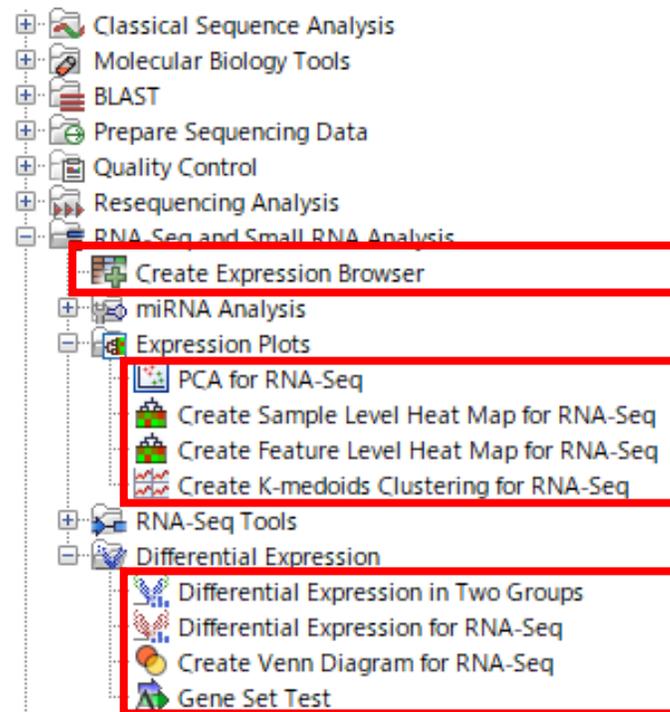
発現量計算

RNA-seq Analysis

サンプル間比較・図の作成

Differential Expression for RNA-seq  
Create Venn Diagram for RNA-seq  
PCA for RNA-seq など

- ✓ 各サンプルごとの遺伝子発現量データを取得した後は、統計処理によりサンプル間比較を行い、発現変動遺伝子の探索や図表の作成を行う
- ✓ サンプルのグループ情報や属性情報をまとめたメタデータが必要となるツールもある



## Create Expression Browser

各遺伝子の発現量や、サンプル間の変動量を集計した表を作成

## PCA for RNA-seq

主成分分析 (Principal Component Analysis)

## Create Feature Level Heat Map for RNA-seq

二次元階層クラスタリング解析とヒートマップ作成

## Create K-medoids Clustering for RNA-seq

K-medoidsクラスタリング

## Differential Expression for RNA-seq

発現変動遺伝子の解析

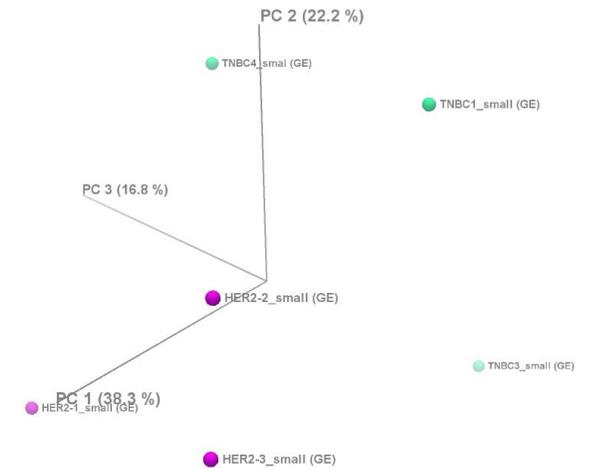
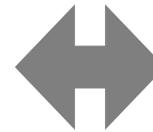
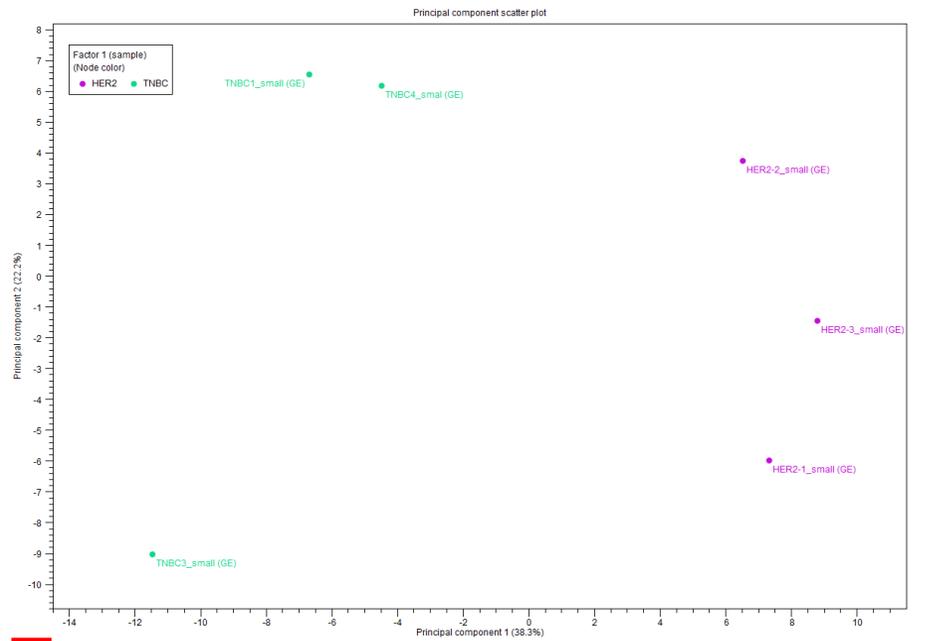
## Create Venn Diagram for RNA-seq

発現変動遺伝子リストのベン図作成

## Gene Set Test

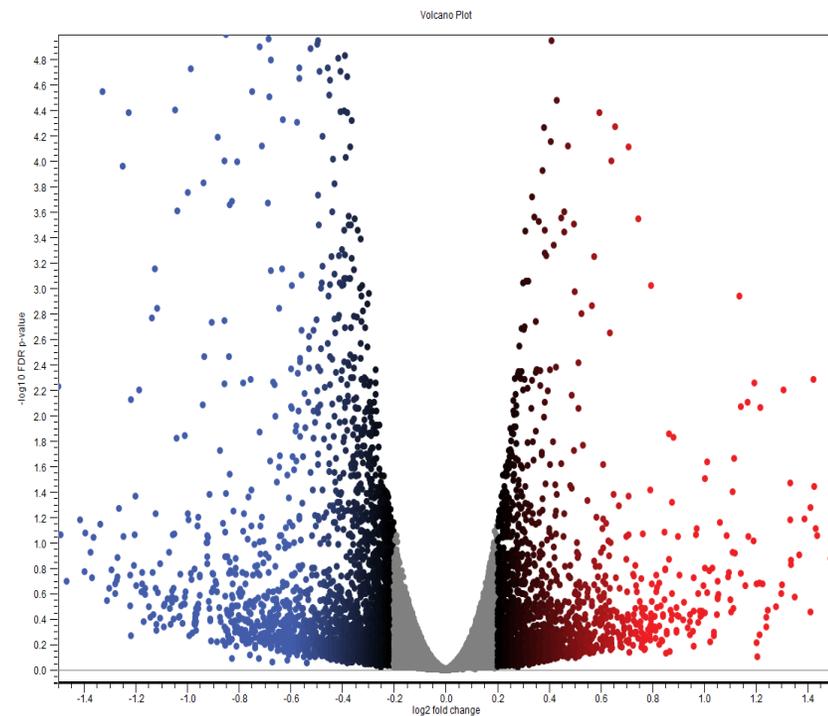
発現変動遺伝子の機能解析

- ✓ 2D表示と3D表示に対応
- ✓ あらかじめ関連付けておいたメタデータのグループ分類情報に基づき、プロットを色分け可能
- ✓ 各遺伝子が各主成分にどの程度寄与しているかも表示可能



# Differential Expression for RNA-seq

- ✓ サンプル間の発現変動を示すFold ChangeとP値が計算され、発現変動遺伝子のフィルタリングやボルケーノプロットによる表示が可能
- ✓ ボルケーノプロットでは、発現の上昇と低下でプロットの色を変えたり、Fold Changeの量に応じてグラデーションをかけたりすることも可能



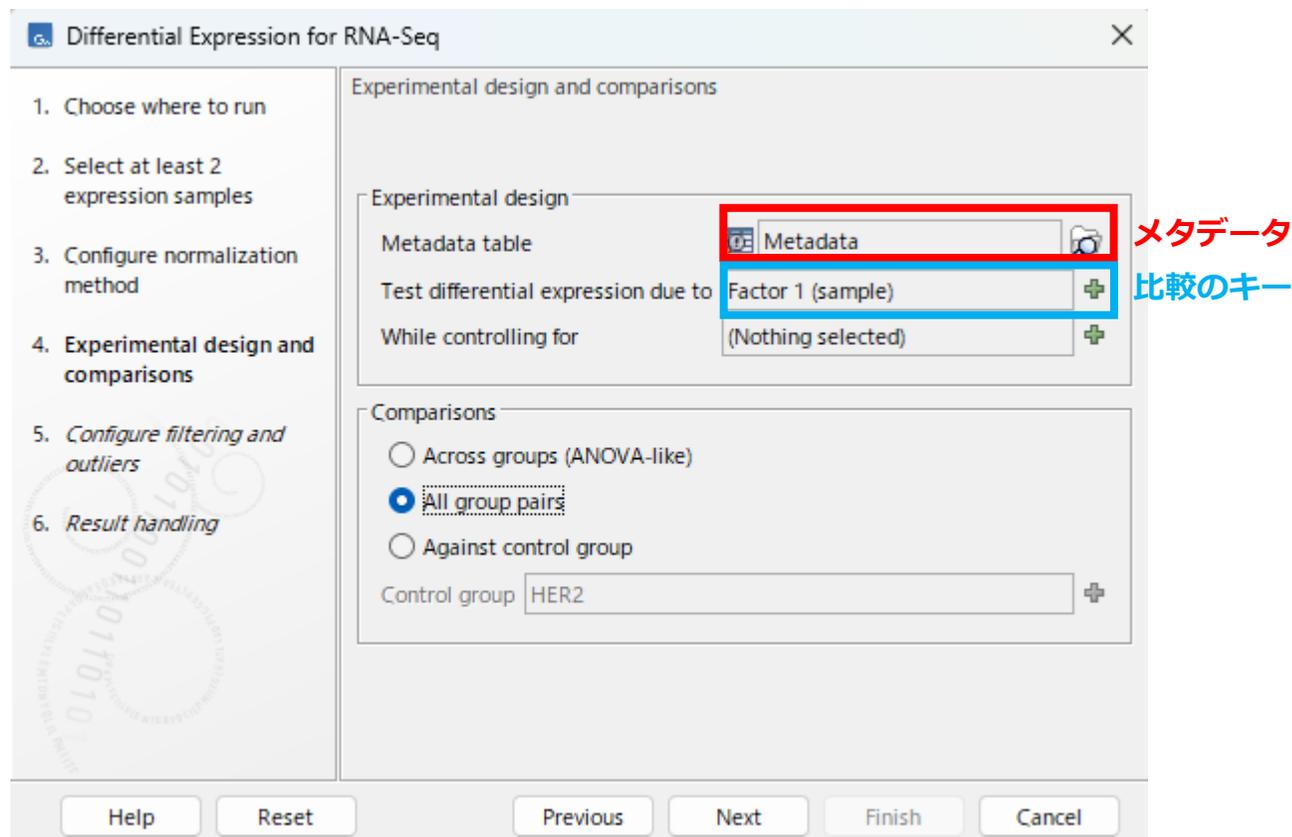
Match all  Match any

Log<sub>2</sub> fold change > 2

P-value < 0.05

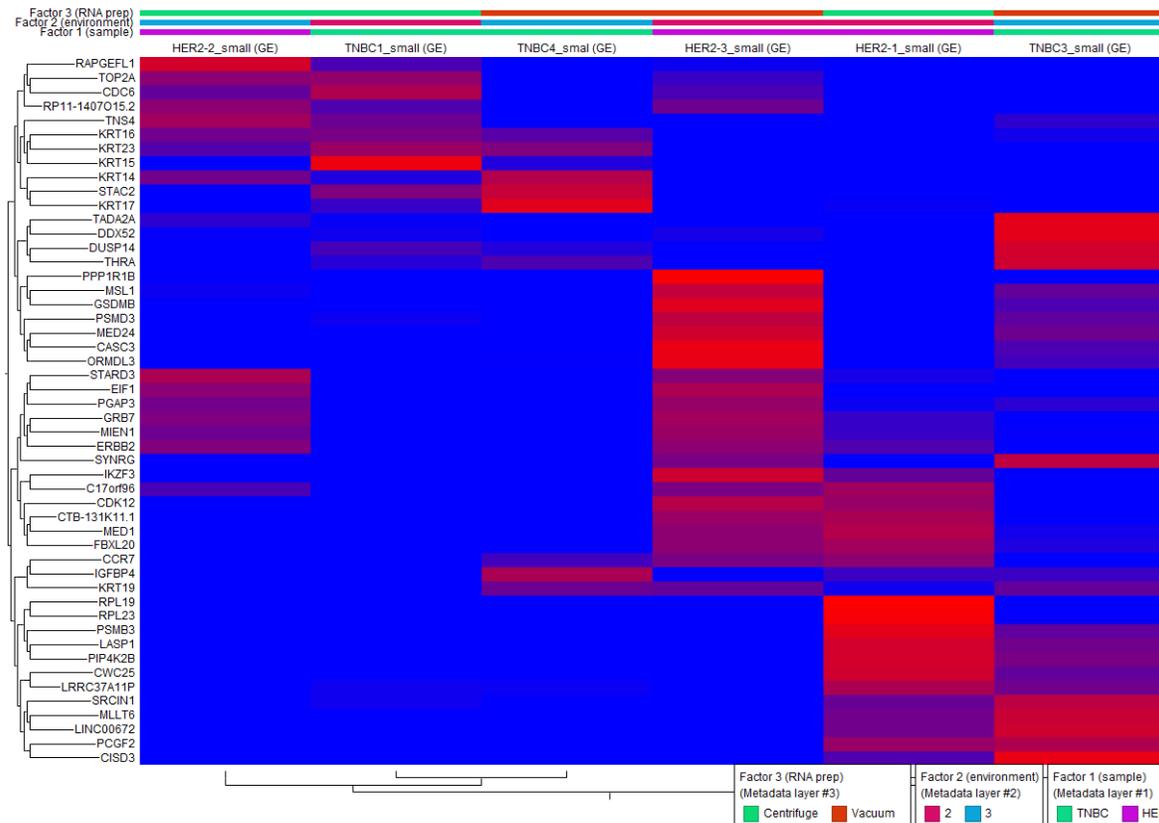
Name	Chromosome	Region	Max group mean	Log <sub>2</sub> fold change	Fold change	P-value	FDR p-value	Bonferroni
ACACA	17	complement(35441923..35766909)	99.84	-0.43	-1.35	0.66	0.83	1.00
TADA2A	17	35766965..35839835	4,474.22	-0.49	-1.41	0.59	0.77	1.00
RP11-378E13.4	17	complement(35768023..35768681)	0.00	NaN	NaN	NaN	NaN	NaN
CTC-421K24.1	17	35814273..35815033	250.65	-0.34	-1.27	1.00	1.00	1.00
DUSP14	17	35849937..35873603	9,039.49	-1.37	-2.58	0.07	0.21	1.00
SYNRG	17	complement(35874900..35969544)	5,360.85	0.34	1.27	0.61	0.79	1.00
RP11-697E22.1	17	35969787..35973886	397.89	0.25	1.19	0.75	0.86	1.00
DDX52	17	complement(35969787..36003493)	5,595.17	-0.44	-1.36	0.50	0.71	1.00
RP11-697E22.2	17	36002976..36044255	584.78	-1.15	-2.21	0.31	0.53	1.00
RP11-697E22.3	17	36009162..36010922	197.79	-2.96	-7.79	0.09	0.24	1.00
HNF1B	17	complement(36046435..36105237)	151.25	-4.12	-17.34	3.88E-4	8.15E-3	0.06
RP11-115K3.1	17	36105206..36108563	0.00	NaN	NaN	NaN	NaN	NaN
AC091199.1	17	complement(36158869..36162759)	0.00	NaN	NaN	NaN	NaN	NaN
RP11-115K3.2	17	complement(36187925..36244363)	364.41	-0.29	-1.22	0.77	0.87	1.00
RNU6-489P	17	36267501..36267607	0.00	NaN	NaN	NaN	NaN	NaN
TBC1D3F	17	36283971..36294915	1,153.21	-0.58	-1.49	0.21	0.40	1.00
RP11-1407O15.3	17	36333639..36333776	0.00	NaN	NaN	NaN	NaN	NaN
TBC1D3	17	complement(36337711..36358166)	610.71	-1.41	-2.65	0.41	0.64	1.00
RP11-1407O15.2	17	complement(36337717..36421858)	2,636.93	0.71	1.63	0.40	0.62	1.00
Y_RNA_1	17	complement(36387075..36387176)	0.00	NaN	NaN	NaN	NaN	NaN
MRPL45	17	36452989..36479101	8,198.95	0.29	1.22	0.57	0.76	1.00
GPR179	17	complement(36481413..36499730)	318.95	-2.55	-5.85	0.05	0.16	1.00

- ✓ 本ツールを使用する際は、あらかじめインポートしておいたメタデータテーブルと、比較のキーとなる情報を含んだ列のヘッダーを指定する
- ✓ グループ間の総当たり、ANOVAなど比較方法を選択できる



# Create Feature Level Heat Map for RNA-seq

- ✓ 発現変動を示す遺伝子や、任意の遺伝子リストに含まれる遺伝子のみを使用して、解析を実行することが可能
- ✓ あらかじめ関連付けておいた、サンプルメタデータのグループ分類情報に基づき、サンプルを色分けしたラベル表示が可能



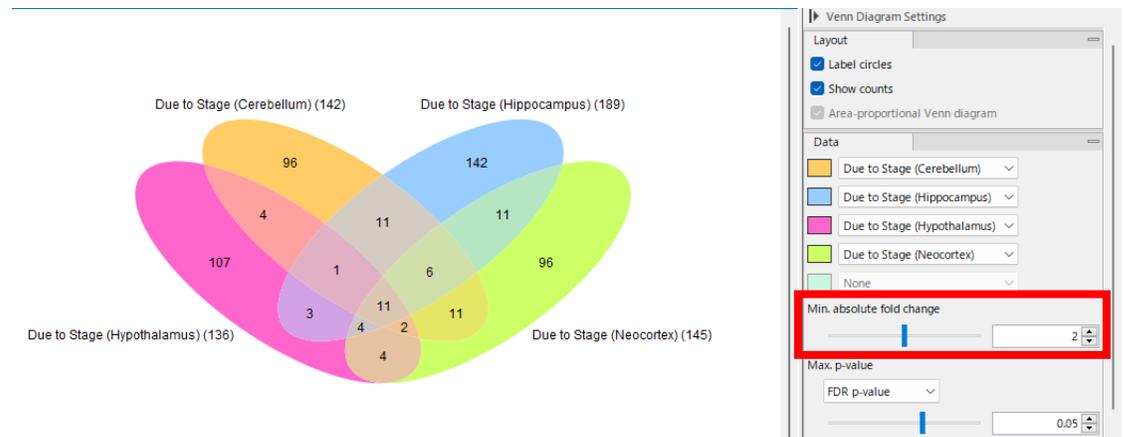
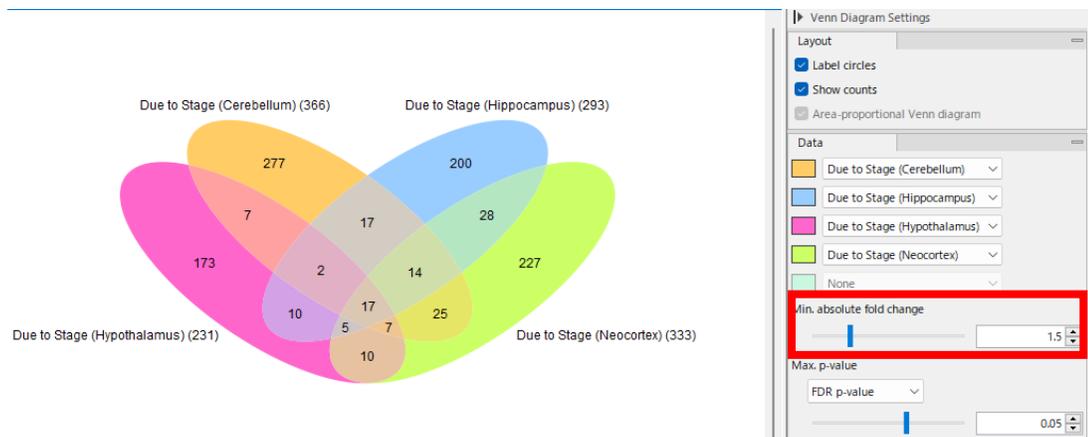
- ✓ パラメータとして、クラスタリングのアルゴリズムやデータのフィルタリング設定を行う

The screenshot shows the 'Create Feature Level Heat Map for RNA-Seq' dialog box at step 3, 'Distances'. The left sidebar lists the steps: 1. Choose where to run, 2. Select at least 2 expression samples, 3. Distances, 4. Feature filters, and 5. Result handling. The main area is divided into two sections: 'Distance measure' and 'Cluster linkage'. Under 'Distance measure', 'Euclidean distance' is selected with a radio button, while 'Manhattan distance' and '1 - Pearson correlation' are unselected. Under 'Cluster linkage', 'Complete linkage' is selected with a radio button, while 'Single linkage' and 'Average linkage' are unselected. At the bottom, there are buttons for 'Help', 'Reset', 'Previous' (highlighted with a dashed border), 'Next', 'Finish', and 'Cancel'.

The screenshot shows the 'Create Feature Level Heat Map for RNA-Seq' dialog box at step 4, 'Feature filters'. The left sidebar is the same as in the previous screenshot. The main area is divided into three sections: 'Filter settings', 'Fixed number of features', and 'Filter features by statistics'. 'Filter settings' has a dropdown menu set to 'Fixed number of features'. 'Fixed number of features' has two input fields: 'Number of features' set to 25 and 'Minimum counts in at least one sample' set to 10. 'Filter features by statistics' has a 'Statistical comparison' dropdown, 'Minimum absolute fold change' set to 1.5, and radio buttons for 'P-value', 'FDR p-value' (selected), and 'Bonferroni'. A 'Threshold' input field is set to 0.05. At the bottom, there are buttons for 'Help', 'Reset', 'Previous', 'Next' (highlighted with a dashed border), 'Finish', and 'Cancel'.

# Create Venn Diagram for RNA-seq

- ✓ Differential Expression for RNA-Seqツール で作成した発現変動解析データを複数セット用いて、データ間の遺伝子の重複などを表すベン図の作成を行うツール
- ✓ Fold ChangeやP値の閾値を変更し、変更結果をリアルタイムにベン図に反映が可能
- ✓ ベン図上の任意のエリアを選択することで、該当する遺伝子データを容易に取得可能

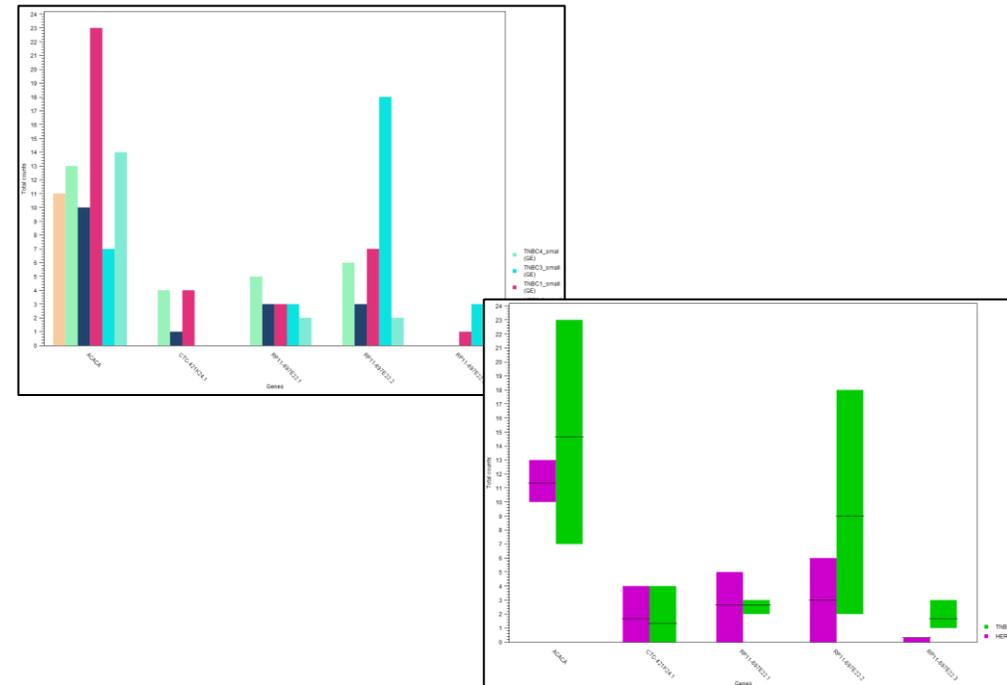


Fold Changeの条件を1.5から2に変更

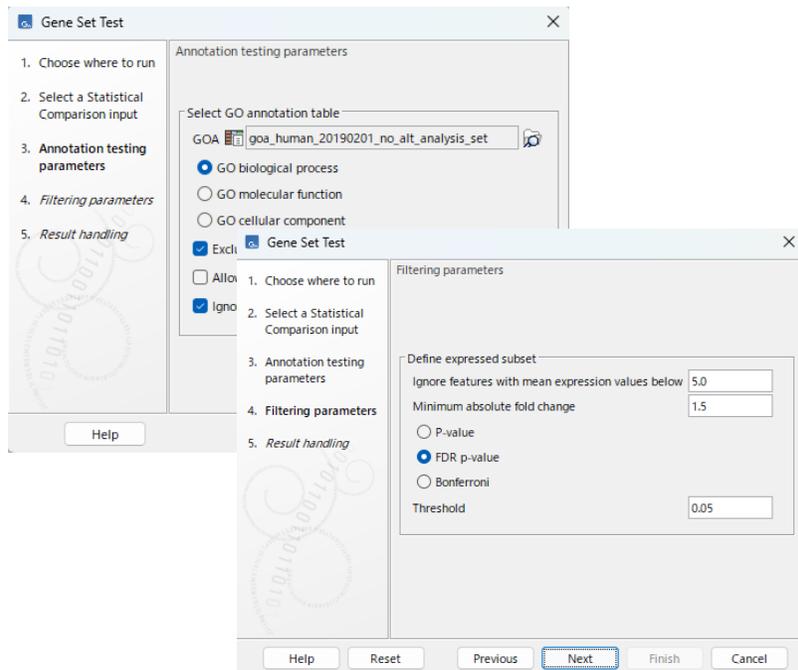
- ✓ Differential Expression for RNA-Seqツールで作成した発現変動解析データと、RNA-Seqの発現データを統合したリストを作成するツール
- ✓ Gene Ontologyなどの遺伝子機能アノテーションデータも同時に表示させ、外部データベースへのリンクも使用可能になる
- ✓ 注目したい遺伝子を選択し、プロットを切り替えることで、サンプルごともしくはグループごとの発現量を示すグラフの作成も可能

Name	Identifier	Due to Factor 1 (sample)						Database	E
		Max group means	Fold change	Log fold chan...	P-value	FDR p-value	Bonferroni		
ACACA	<a href="#">ENSG000001132142</a>	99.84	-1.35	-0.43	0.66	0.83	1.00	UniProtKB	C
TADA2A	<a href="#">ENSG00000108264</a>	4,474.22	-1.41	-0.49	0.59	0.77	1.00	UniProtKB	O
RP11-378E13.4	<a href="#">ENSG00000267613</a>	0.00	NaN	NaN	NaN	NaN	NaN		
CTC-421K24.1	<a href="#">ENSG00000267662</a>	250.65	-1.27	-0.34	1.00	1.00	1.00		
DUSP14	<a href="#">ENSG00000161326</a>	9,039.49	-2.58	-1.37	0.07	0.21	1.00	UniProtKB	O
SYNRG	<a href="#">ENSG00000006114</a>	5,360.85	1.27	0.34	0.61	0.79	1.00	UniProtKB	Q
RP11-697E22.1	<a href="#">ENSG00000267642</a>	397.89	1.19	0.25	0.75	0.86	1.00		
DDX52	<a href="#">ENSG00000141141</a>	5,595.17	-1.36	-0.44	0.50	0.71	1.00	UniProtKB	Q
RP11-697E22.2	<a href="#">ENSG00000267668</a>	584.78	-2.21	-1.15	0.31	0.53	1.00		
RP11-697E22.3	<a href="#">ENSG00000267789</a>	197.79	-7.79	-2.96	0.09	0.24	1.00		
HNF1B	<a href="#">ENSG00000108753</a>	151.25	-17.34	-4.12	3.88E-4	8.15E-3	0.06	UniProtKB	P

Select Genes in Other Views Copy Gene Names to Clipboard



- ✓ Differential Expression for RNA-Seqツールで作成した発現変動解析データを用いて、GOレベルの変動解析を行うツール
- ✓ ツールの使用には、Gene Ontology情報を含んだ遺伝子機能アノテーションデータが必要
- ✓ パラメータとして、Gene Ontology情報や遺伝子のフィルタリング設定を行う



GO term	Description	Detected Genes	DE Genes	P-values
<a href="#">0031424</a>	keratinization	58	21	0.03
<a href="#">0044707</a>	single-multicellular organism process	63	22	0.05
<a href="#">0038127</a>	ERBB signaling pathway	2	2	0.08
<a href="#">0038128</a>	ERBB2 signaling pathway	2	2	0.08
<a href="#">0043405</a>	regulation of MAP kinase activity	2	2	0.08
<a href="#">0046777</a>	protein autophosphorylation	2	2	0.08
<a href="#">0060260</a>	regulation of transcription initiation from RNA polymerase II promoter	2	2	0.08
<a href="#">1905475</a>	regulation of protein localization to membrane	2	2	0.08
<a href="#">1905477</a>	positive regulation of protein localization to membrane	2	2	0.08
<a href="#">2000142</a>	regulation of DNA-templated transcription, initiation	2	2	0.08
<a href="#">0032501</a>	multicellular organismal process	65	22	0.08

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～17 : 00)

FAX: 052-624-4389

E-mail: [support@filgen.jp](mailto:support@filgen.jp)