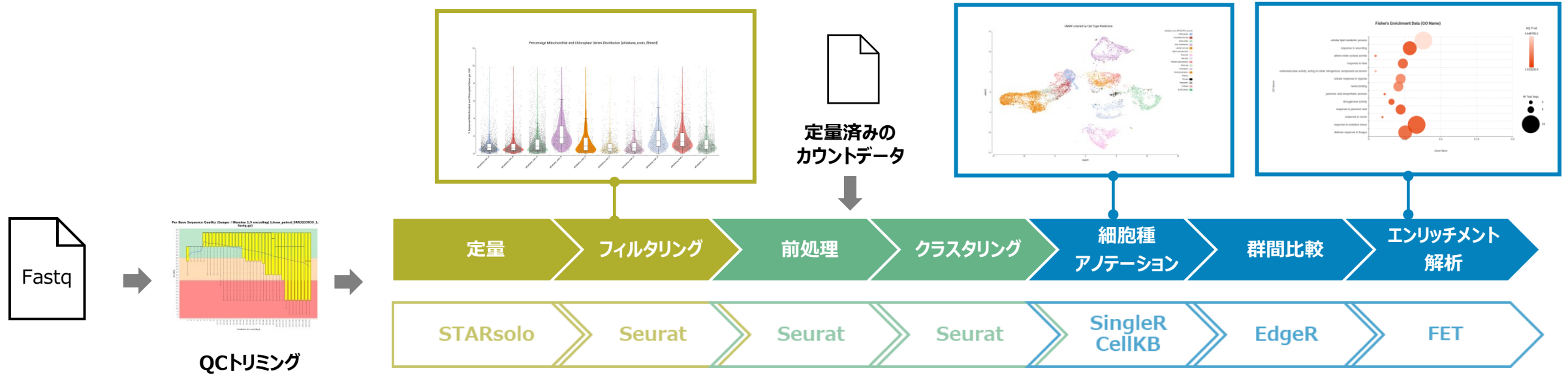




植物のシングルセルRNA seqデータ解析

フィルジェン株式会社
バイオインフォマティクス部

シングルセルRNA-Seqデータ解析



生データや定量済みファイルから解析可能

マウス操作

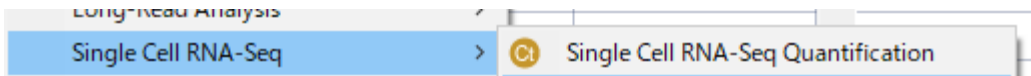
高スペックPCは不要

- *Arabidopsis thaliana* (シロイヌナズナ)
- 通常の培地 (7サンプル) 高スクロース培地 (3サンプル)
- シングルセルRNA-SeqライブラリはDrop-seqバージョン3.1を使用

合計 10 のサンプル



scRNA-Seq Quantification



Single Cell RNA-Seq Quantification Options

Cell Barcoded Technologies

Supported technologies: 10x Chromium, Drop-seq, CEL-Seq, etc.
With these library construction technologies, the produced reads contain Cell Barcodes and UMIs. Each FASTQ contains reads from multiple cells. These technologies are also known as 3'-Enrichment technologies.

Full-length Technologies

Supported technologies: SMART-Seq, SMARTer.
With these library construction technologies, the produced reads do not contain barcodes. Each pair of FASTQ files belong to a different cell. These technologies are also known as Full-length technologies

Default < Back **Next >** Cancel Run

テクノロジーの種類を選択します。

Input

Note: This tool makes use of free cloud computation resources. This is an introductory offer and may change in a future release depending on the overall resource consumption of this feature.

Reference Genome

Reference genome Browse...

Annotation GFF/GTF

Annotation File Browse...

Exon Feature

Overhang

FASTQ

Single-cell RNA-Seq Reads 2 Files Paired-End Clear Add Files

Upstream Files Pattern

Downstream Files Pattern

Strandness

Default < Back **Next >** Run Cancel

リファレンスゲノムを選択

リファレンスアノテーション
GFFファイルを選択

シーケンスデータを選択
複数サンプルの場合ある場合でも
1サンプルずつ解析する必要があります。

scRNA-Seq Quantification

Configuration. Reads Configuration.

Preset

Library Technology: Drop-seq

Read Configuration

Barcode Mate: 10x Chromium 3' v1

Cell Barcode Start: Drop-seq

Cell Barcode Length: 12

UMI Start: 13

UMI Length: 8

Clip from 5' end:

5' Number of Bases: 00

Clip from 3' end:

3' Number of Bases: 00

Cell Barcode Detection

Add Cell Barcode Whitelist:

Cell Barcodes Whitelist: Browse...

Cell Barcode Match Type: 1MM Multi

Default < Back Next > Run Cancel

プラットフォームを選択します。

Configuration. Alignment Parameters.

2-pass Mapping:

Min. Intron Length: 20

Max. Intron Length: 1000000

Max. # of Mismatches: 999

Max. # of Multiple Alignments: 20

Include Chimeric Alignments:

Max. Distance Between Mates: 1000000

Configuration. Counting Configuration.

UMIs

UMI Collapsing: UMI Tools

UMI Filtering: Multi Mapping UMIs

Features

Multimapping Reads: Rescue

Feature Counting: Exons+Introns

Cells

Cell Filtering: Empty Drops

Num. Expected Cells: 3000

その他解析条件に合わせた設定を行います。

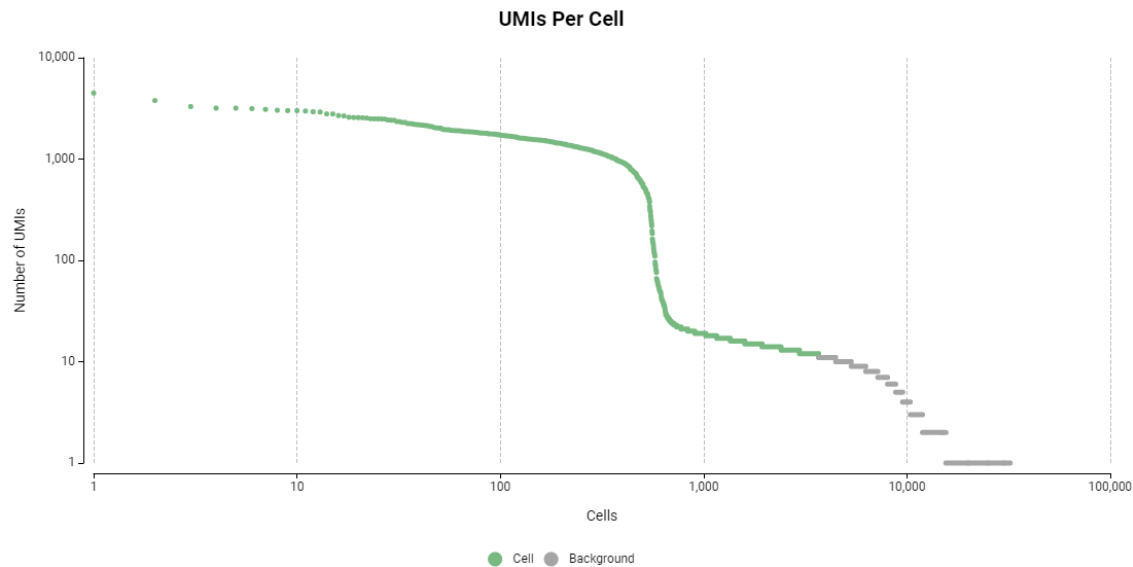
scRNA-Seq Quantification

Cell	#Features	Counts	Feature ID	Name	#Cells	Counts
AAACCTGAGCGTTTAC	2285	6068	ENSCAFG00845015183	ENSCAFG00845015183	0	0
AAACCTGAGGAGTCTG	1386	3725	ENSCAFG00845015195	ENSCAFG00845015195	0	0
AAACCTGCACCAGGCT	1216	2645	ENSCAFG00845015208	ENSCAFG00845015208	0	0
AAACCTGCACGGTAAG	894	2050	ENSCAFG00845015217	ENSCAFG00845015217	0	0
AAACCTGCATACGCCG	691	1070	ENSCAFG00845015230	ENSCAFG00845015230	0	0
AAACGGGAGACTTTCG	1108	2297	ENSCAFG00845015240	ENSCAFG00845015240	0	0
AAACGGGGTCCCTACT	1307	2720	ENSCAFG00845015261	MAGEB4	0	0
AAACGGGTCAAACCTG	1253	2960	ENSCAFG00845015275	ENSCAFG00845015275	0	0
AAACGGGTCTAACTTC	1889	5122	ENSCAFG00845015297	ENSCAFG00845015297	0	0
AAACGGGTCTTGATC	1395	3528	ENSCAFG00845015303	ENSCAFG00845015303	0	0
AAAGATGAGCTTTGGT		70	ENSCAFG00845015316	EN		0
AAAGATGAGTCCAGGA		70	ENSCAFG00845015329	U4		0
AAAGATGCATTAACCG	1267	2772	ENSCAFG00845015345	ENSCAFG00845015345	0	0
AAAGATGCATTGACA	837	1645	ENSCAFG00845015364	U6	0	0
AAAGATGGTCGACTAT	1550	3591	ENSCAFG00845015457	GK	172	182
AAAGATGTCGGTTAAC	1243	2746	ENSCAFG00845015511	ENSCAFG00845015511	0	0
AAAGATGTCTTAGCCC	1728	4503	ENSCAFG00845015550	ENSCAFG00845015550	12	12
AAAGATGTCTTGGGTA	1126	2511	ENSCAFG00845015575	U6	0	0
AAAGCAAAGGAGCGTT	763	2459	ENSCAFG00845015589	ENSCAFG00845015589	42	42
AAAGCAACAGGAATCG	1319	3540	ENSCAFG00845015610	ENSCAFG00845015610	0	0
AAAGCAAGTACCGGCT	1305	2873	ENSCAFG00845015615	ENSCAFG00845015615	0	0
AAAGCAAGTCTTCAAG	1520	4153	ENSCAFG00845015631	ENSCAFG00845015631	0	0
AAAGCAATCCCTCAGT	1245	2797	ENSCAFG00845015637	ENSCAFG00845015637	0	0

解析が完了するとマトリクスデータが得られます。

クラスタリングなどの解析はこちらのサイドパネルから実行します。

マトリクスデータの外、定量化プロセスの概要を記載したレポートとUMI Per Cell プロットが得られます。



Y軸にはUMIの総数、X軸にはデータセットで検出されたすべてのCellバーコードを表しています。灰色の部分はバックグラウンドノイズとみなし緑色の部分の細胞がマトリクスデータに保持されます。

Single Cell RNA-Seq Quantification Results

Summary

Statistic	Value
Number of Reads	70758378
Sequencing Saturation	0.643287
Q30 Bases in RNA read	0
Reads Mapped to Genome: Unique+Multiple	0.679928
Reads Mapped to Genome: Unique	0.641817
Reads Mapped to GeneFull: Unique+Multiple GeneFull	0.644255
Reads Mapped to GeneFull: Unique GeneFull	0.5965
Estimated Number of Cells	1113
Fraction of Unique Reads in Cells	0.622961
Mean Reads per Cell	23624
Median Reads per Cell	14845
UMIs in Cells	8027126
Mean UMI per Cell	7212
Median UMI per Cell	5144
Mean GeneFull per Cell	2400
Median GeneFull per Cell	2120
Total GeneFull Detected	22013

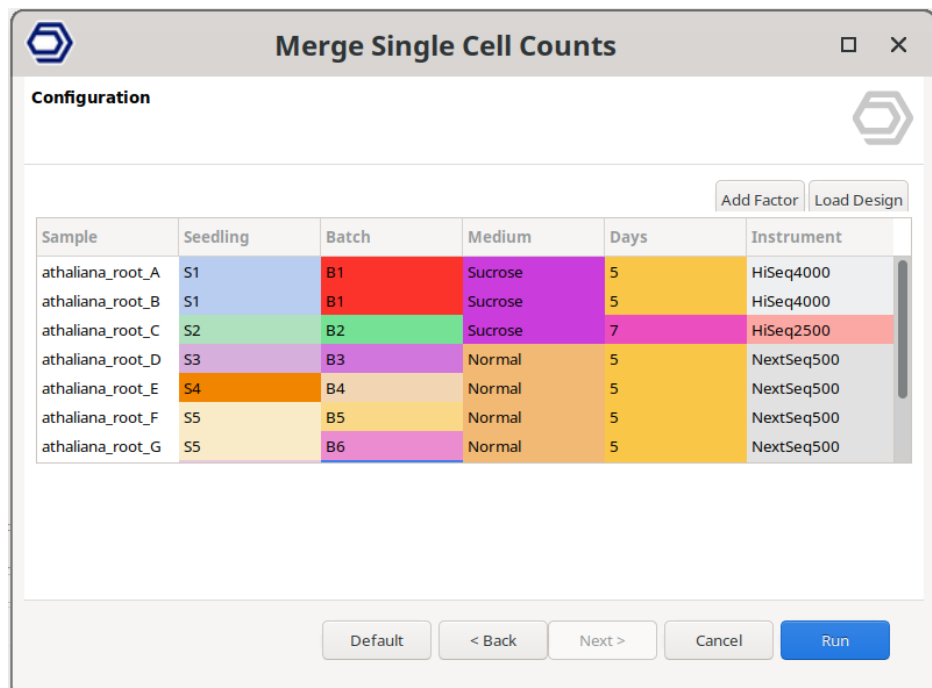
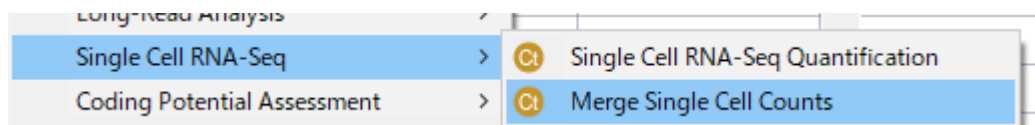
Feature Stats

This section provides statistics regarding the feature reads, that is, the reads containing transcript sequences.

The following statistics refer to reads that have been discarded for the mentioned reasons.

Statistic	Value
Unmapped	22116074

複数サンプル使用時は各マトリクスデータを作成後、マージを行います。



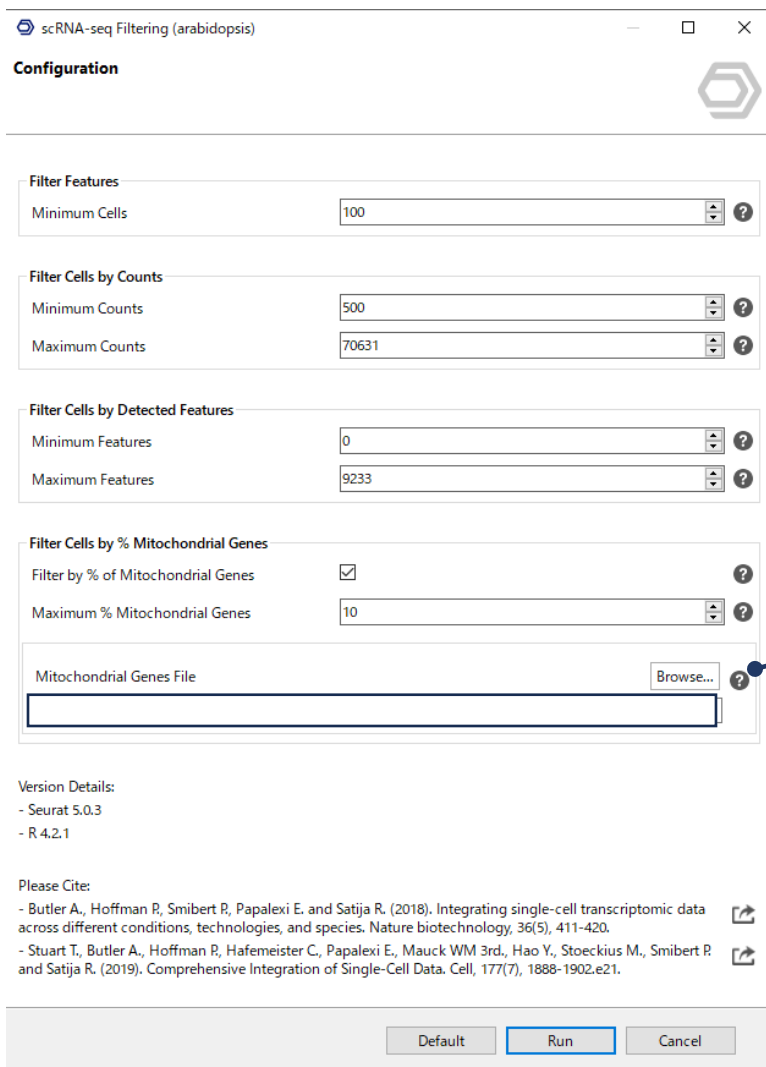
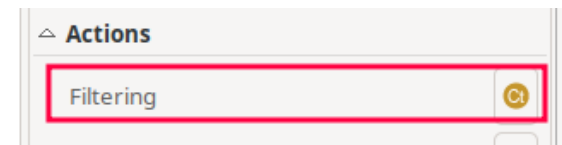
*Single Cell Counts: Merge X

Cell	#Features	Counts	Barcode
arabidopsis_A-TTTTGCCGTA	4317	13181	AT1G01010
arabidopsis_A-TTTTTGCTTTAT	1107	1624	AT1G01020
arabidopsis_A-TTTTTGTGCGAT	572	984	AT1G01030
arabidopsis_A-TTTTTTCGCACC	1117	2611	AT1G01040
arabidopsis_A-TTTTTTAGGAG	609	1380	AT1G01046
arabidopsis_B-AAAACGCCAGC	1137	1621	AT1G01050
arabidopsis_B-AAAACATCGTGG	718	1301	AT1G01060
arabidopsis_B-AAAACGGCTTCT	639	1293	AT1G01070
arabidopsis_B-AAAACGGGTAGA	593	769	AT1G01080
arabidopsis_B-AAAACACTACGCTC	580	1114	AT1G01090
arabidopsis_B-AAAACATATCA	378	660	AT1G01100
arabidopsis_B-AAAAGCTTCGCG	1107	1837	AT1G01110

設定画面では、マージするデータの選択の他、各データがどのような特徴を持つかメタデータを作成します。

解析後、新たにマトリクスデータが作成され各バーコードの接頭辞にサンプル名が追加されます。

定量プロセスでもフィルタリングが行われますが、さらにここではノイズとなるデータをフィルタリングします。



scRNA-seq Filtering (arabidopsis)

Configuration

Filter Features

Minimum Cells: 100

Filter Cells by Counts

Minimum Counts: 500

Maximum Counts: 70631

Filter Cells by Detected Features

Minimum Features: 0

Maximum Features: 9233

Filter Cells by % Mitochondrial Genes

Filter by % of Mitochondrial Genes:

Maximum % Mitochondrial Genes: 10

Mitochondrial Genes File: Browse...

Version Details:

- Seurat 5.0.3
- R 4.2.1

Please Cite:

- Butler A., Hoffman P., Smibert P., Papalexi E. and Satija R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology, 36(5), 411-420.
- Stuart T., Butler A., Hoffman P., Hafemeister C., Papalexi E., Mauck WM 3rd, Hao Y., Stoeckius M., Smibert P. and Satija R. (2019). Comprehensive Integration of Single-Cell Data. Cell, 177(7), 1888-1902.e21.

Default Run Cancel

最小セル数、最小・最大カウント、最小・最大特徴数を任意に指定することができます。

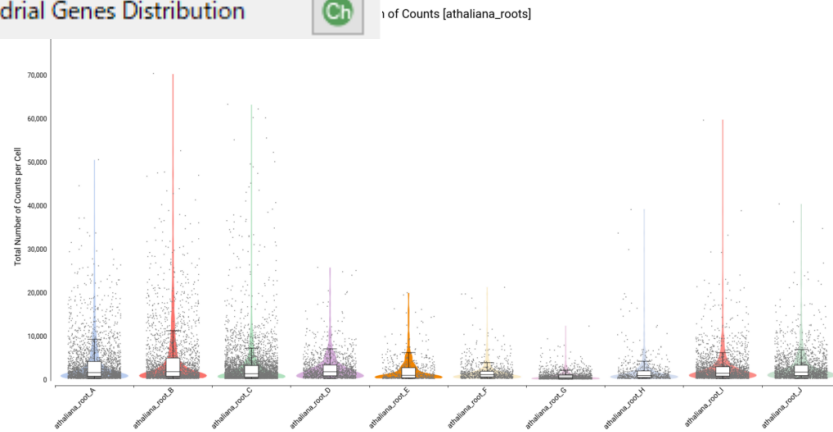
ミトコンドリア遺伝子、葉緑体遺伝子のリストを指定することで、任意のパーセンテージを超えるミトコンドリア遺伝子・葉緑体遺伝子を持つ細胞を破棄できます。

フィルタリング

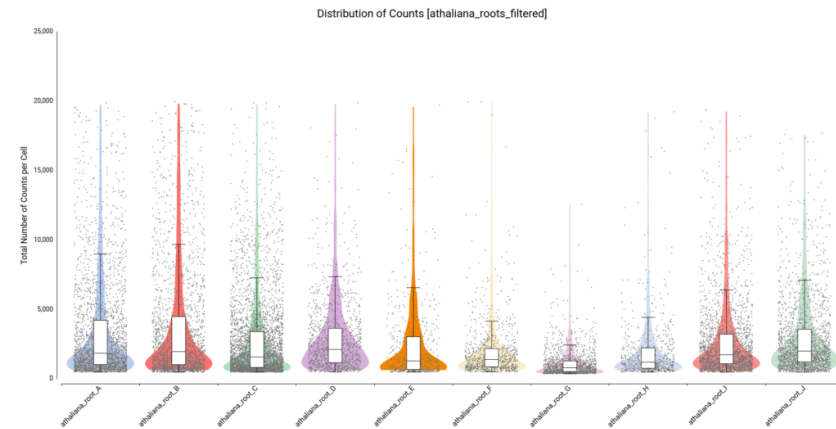
Charts

- Total Counts Distribution Ch
- Expressed Genes Distribution Ch
- % Mitochondrial Genes Distribution Ch

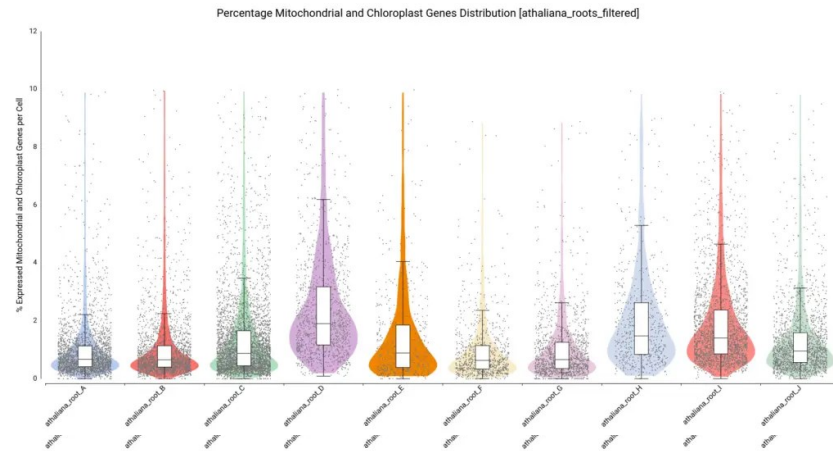
フィルタリング後、新たにマトリクスデータが作成されます。
フィルタリング前後のバイオリンプロットを作成し、結果を比較できます。



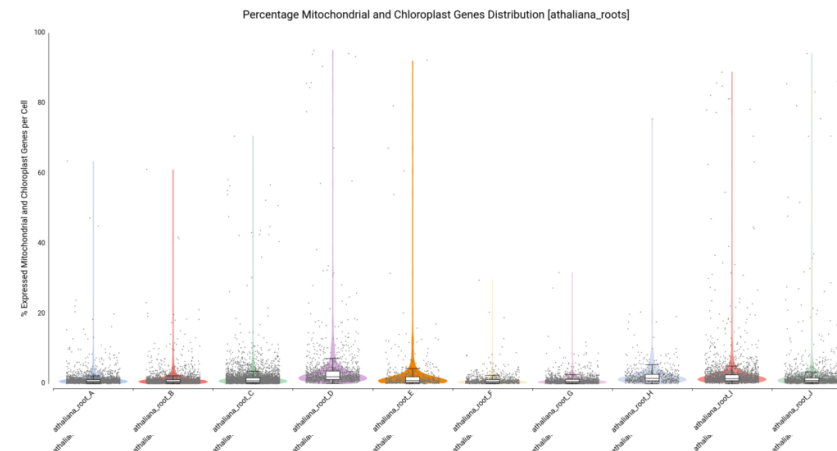
フィルタリング前のセルあたりのカウントの分布



フィルタリング後のセルあたりのカウントの分布



フィルタリング前のミトコンドリア遺伝子と葉緑体遺伝子の割合の分布



フィルタリング後のミトコンドリア遺伝子と葉緑体遺伝子の割合の分布

scRNA-Seq Clustering (arabidopsis_filtered)

Configuration: Preprocessing

This tool is designed to perform the clustering of cells coming from single-cell RNA sequencing (scRNA-seq) data. Prior to the clustering, this tool allows the preprocessing the data in order to make it suitable for the clustering algorithm. This application is based on the widely-used Seurat package.

Normalization

Normalize Data

Normalization Method Log Normalization

Data Adjustment

High Variable Features 3000

Scale Data

Center Data

Data Correction

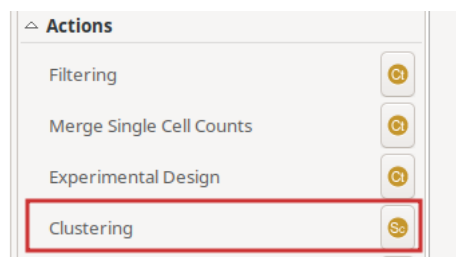
Mitochondrial Genes File Browse...

Regress Out Cell Cycle Genes Browse...

Dimensional Reduction

Principal Components 50

Default < Back **Next >** Run Cancel



フィルタリング後、クラスタリングに進みます。
クラスタリングの最初のウィザードでは、前処理の
設定を行います。

正規化の実施の有無

ミトコンドリア遺伝子・葉緑体遺伝子のリストに基づく修正

scRNA-Seq Clustering (arabidopsis_filtered)

Configuration: Multi-sample Data Integration

The Multi-sample Data Integration step aims to integrate scRNA-seq datasets coming from different samples or conditions. As a result, this step enables the identification of shared cell types across datasets and thus further downstream comparative analyses.

Integration Factor: medium (2) ?

Integration Method: Seurat-CCA ?
Harmony ?
Seurat-CCA ?
Seurat-Joint PCA ?
Seurat-RPCA ?

Seurat Options

N. Dimensions for Integration: 10 ?

K Anchor: 30 ?

K Score: 100 ?

K Weight: 100 ?

Harmony Options

Theta: 2 ?

Lambda: 1 ?

Tau: 0 ?

N° Clusters: 5 ?

Epsilon: 4E-4 ?

Default < Back Next > Run Cancel

scRNA-Seq Clustering (arabidopsis_filtered)

Configuration: Clustering

This step performs the actual cell clustering. It groups cells with similar expression patterns, which should correspond to the same cell type.

Clustering

Define Dimensions by: Manual ?

Number of Dimensions: 20 ?

k-value: 20 ?

Resolution: 0.6 ?

UMAP Configuration

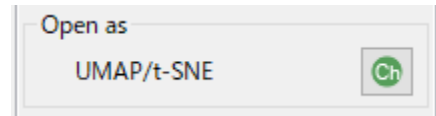
Point's Minimum Distance: 0.3 ?

Point's Spread: 1 ?

Default < Back Next > Run Cancel

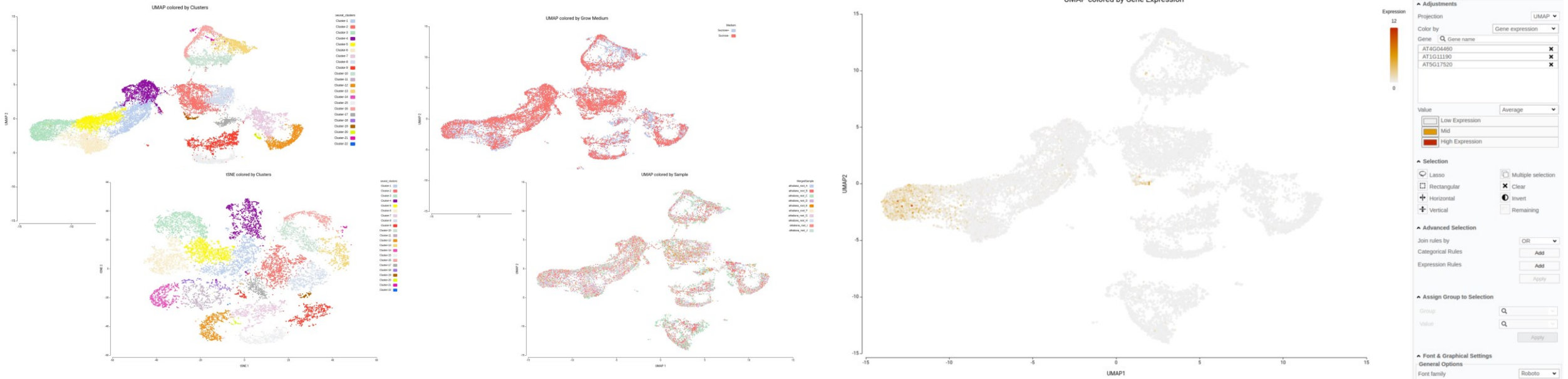
クラスタリングを行うパラメータを設定します。
Dimensionsは、クラスタの分離と分解能に影響を及ぼします。
Resolutionを増やすとクラスター数が増えます。

データの統合に用いられる手法を任意に指定します。
このオプションは複数サンプルの場合行い異なるバッチや条件のデータ
セットを統合する際に使用されます。



結果としてマトリクスデータが作成されます。
サイドパネルよりUMAP/t-SNEを表示できます。

UMAP/t-SNEの切り替え、メタデータの切り替えの他、
新規アノテーションの作成、遺伝子発現の視覚化を行うことができます。



Seuratで取得したクラスターによって
色分けされたUMAPとt-SNE

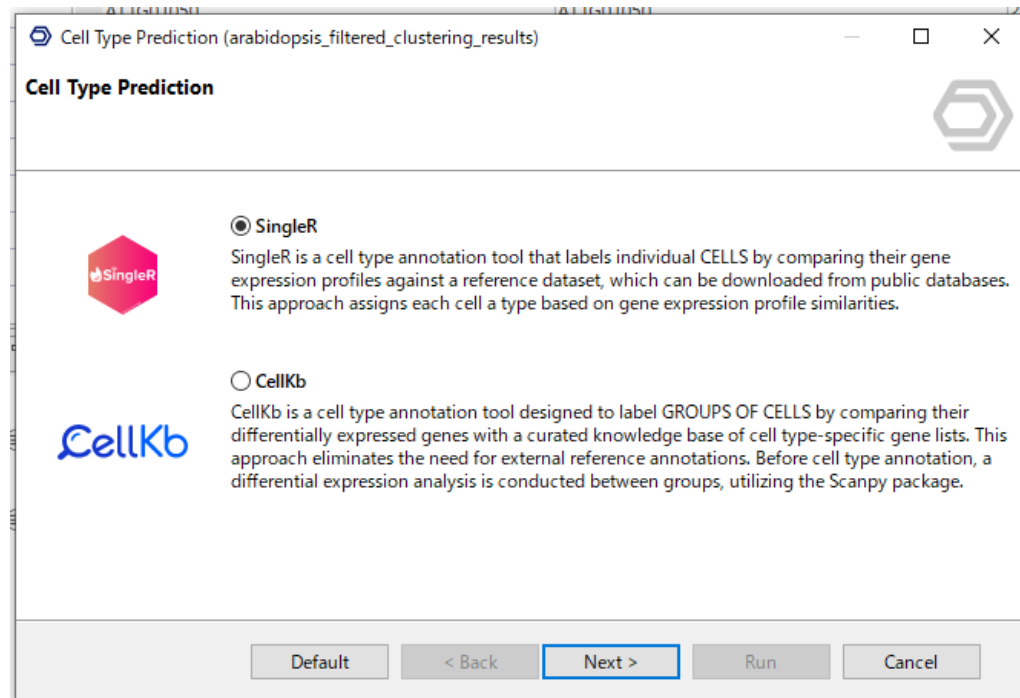
条件（通常の培地と高シヨ糖培
地）サンプル（10サンプル）によって
色分けされたUMAP

既知のLateral Root Cap遺伝子マーカーの平均
発現によって色分けされた細胞を含むUMAP

Cell Type Prediction



クラスタリング後、サイドパネルからCell type Predictionを行います。



SingleR

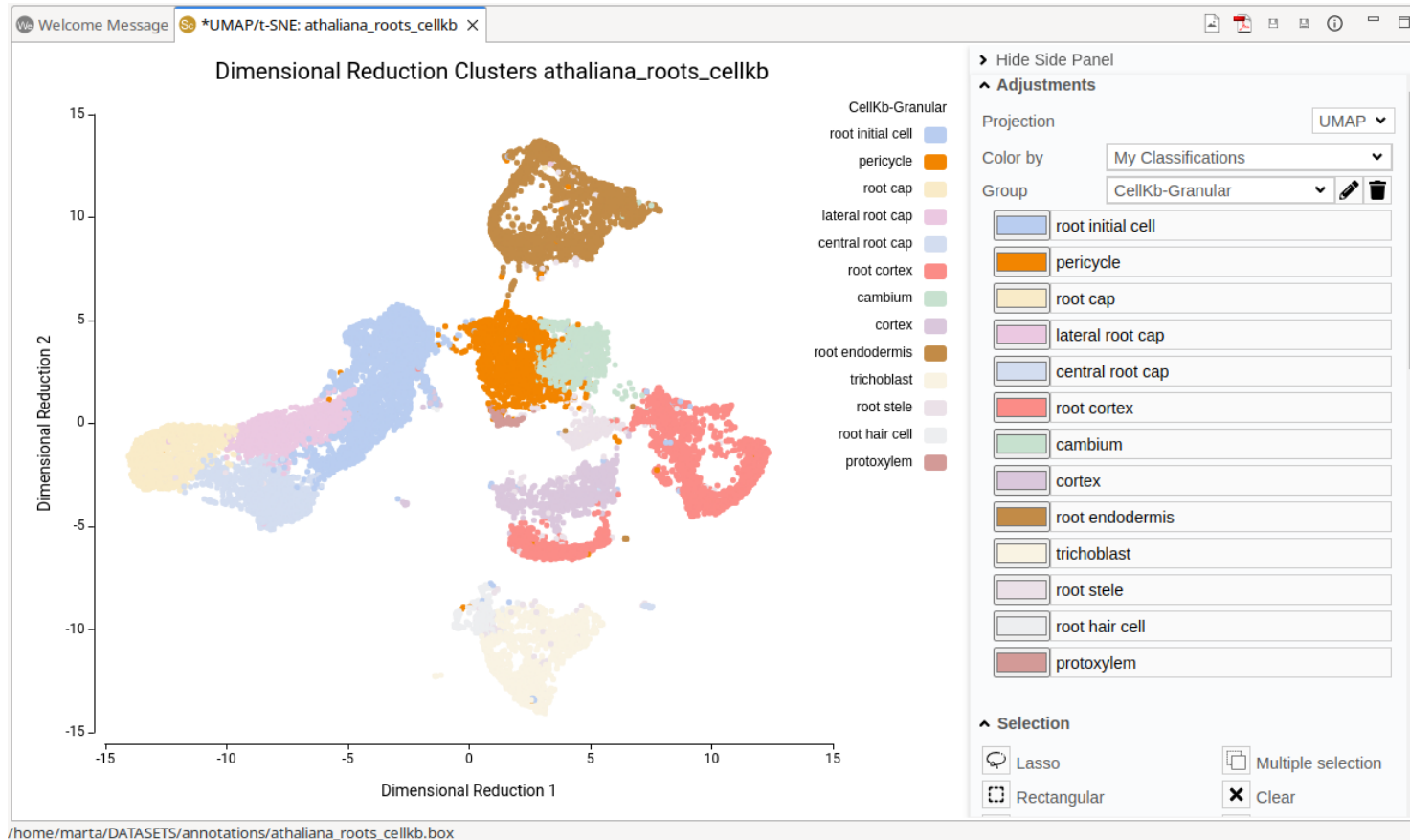
公開データベースからダウンロードできる参照データセットと遺伝子発現プロファイルと比較して、個々の細胞にラベルを付けます。このアプローチでは、遺伝子発現プロファイルの類似性に基づいて各細胞にタイプを割り当てます。

CellKb

キュレーションされた知識ベースに対して、発現の異なる遺伝子と比較することで、細胞グループにラベルを付けます。このツールでは、外部参照アノテーションは必要ありません。

CellKbデータベースのみの販売も行っております。

[データベースの詳細について](#)



どちらのプログラムを使用した場合でも、
UMAP/t-SNEのメタデータ内に細胞種に関する
情報が追加されます。

Differential Expression Analysis

Differential Expression



細胞種アノテーション後、サイドパネルからDifferential Expression Analysisを行います。

Configuration 1: Filtering and Normalization

Genes Filtering

Counts Per Million

CPM Filter: 1

Cells Reaching CPM Filter: 1

Raw Counts

Minimum Sample Count: 10

Minimum Total Count: 15

Normalization

Normalization Method: TMM with Zero Pairing

Buttons: Default, < Back, Next >, Cancel, Run

カウント数の少ない遺伝子を分析から除外するオプション。最小CPMと遺伝子のCPMがフィルターレベルを超えるセルの最小数を設定します。

Configuration 3: Design

Design

Simple Design

Multifactorial Design

Biological Replicates: Sample

Blocking Factor: medium

Primary Target

Primary Factor: Seurat Clusters

Primary Contrast Conditions: Cluster-1

Primary Reference Conditions: Cluster-2, Cluster-3, Cluster-4, Cluster-5, Cluster-6, Cluster-

Test Contrasts Separately:

Secondary Target

Secondary Factor: SRP267870_1

Secondary Contrast Conditions: Columella root cap

Secondary Reference Conditions:

Buttons: Default, < Back, Next >, Run, Cancel

Simple Design

1つの実験要因のみを考慮して比較します。
例) クラスタ1と残りのクラスタの比較、
クラスタ1、2、3 vs クラスタ4、5、6の比較など

Multiple Design

2つの要因を考慮して比較します。
例)
健康グループのクラスタ1 vs 疾患グループのクラスタ1
など

レプリケート情報を含む要素

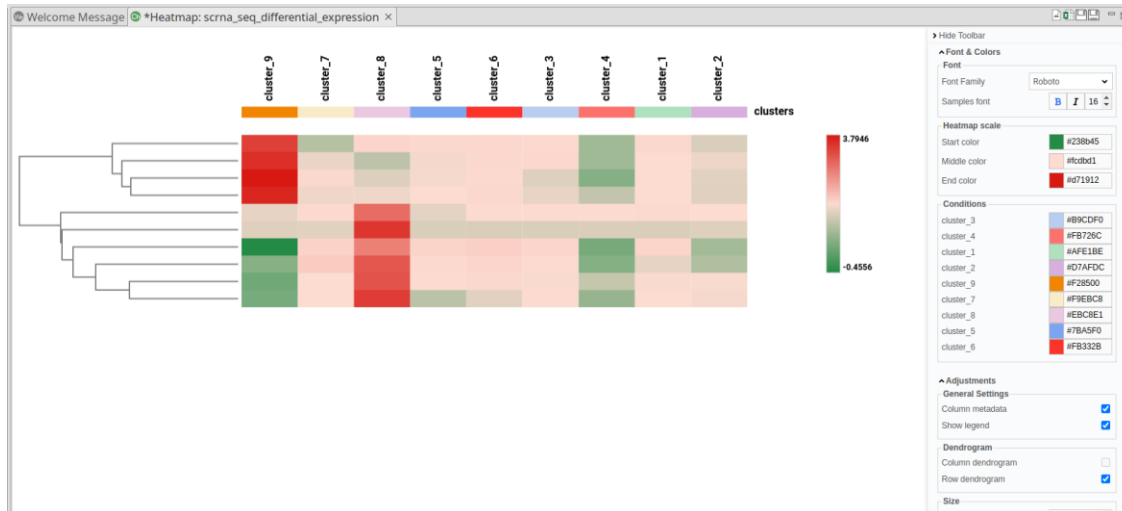
結果に干渉する可能性のある別のファクターがある場合指定

チェックした場合、コントラストとして選択された各クラスタが、リファレンスクラスタに対してテストされる。

Differential Expression Analysis

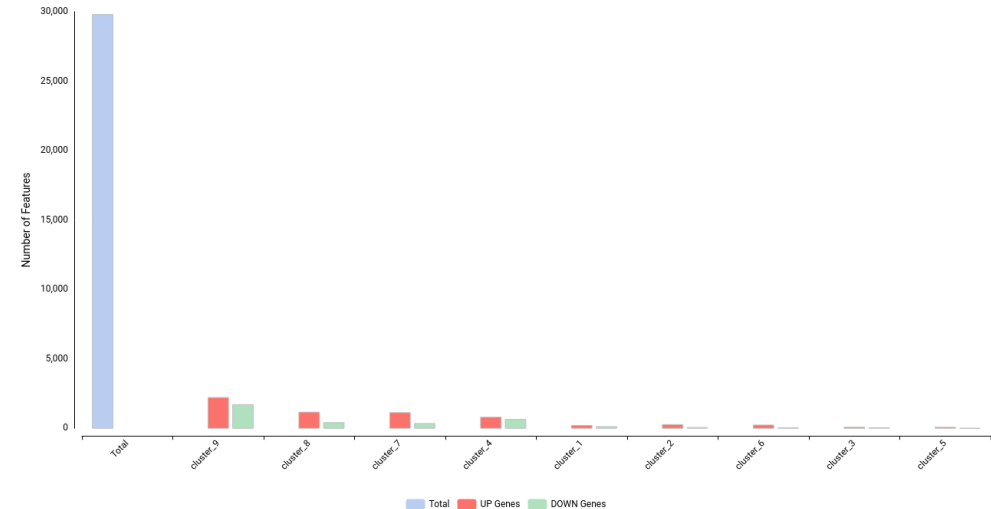
Tags	Contrast	Reference	Feature	FDR	logCPM	logFC	LR	Pvalue
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	-0.29017	5.60482	73.73546	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	-1.07299	3.80291	58.86691	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	-0.51338	7.66682	56.25231	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	2.16898	3.76491	55.21582	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	4.6159	2.90735	52.33218	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	4.45271	3.19027	50.61802	0
DOWN	cluster_1	cluster_9,clus...	ENSG0000002...	0	3.07451	-11.05655	42.8974	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	1.92687	4.44785	42.11914	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0	-1.50958	8.83064	41.25226	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00001	4.23193	6.29306	35.75649	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00001	3.49885	3.05985	35.63237	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00001	-2.03253	8.16088	35.30693	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00001	-0.78736	6.0574	34.84093	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00001	-2.67069	3.4582	34.26514	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00002	-3.04191	4.60302	32.71138	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00002	4.70804	4.27552	32.63564	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00002	-1.57393	2.34859	32.26896	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00005	4.03306	2.39889	30.69504	0
UP	cluster_1	cluster_9,clus...	ENSG0000001...	0.00006	2.33964	3.32793	30.0679	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00007	-2.67157	4.55383	29.83111	0
UP	cluster_1	cluster_9,clus...	ENSG0000002...	0.00008	-1.35783	2.73788	29.51971	0
DOWN	cluster_1	cluster_9,clus...	ENSG0000001...	0.00008	3.05292	-10.60946	29.35155	0

解析結果として発現変動遺伝子のリストや各種チャートが作成されます。



ヒートマップ

DE Results Overview



DE 結果の概要チャート

Fisher's Exact Test



Differential Expression Analysis結果のサイドパネルからエンリッチメント解析を実行できます。

*Functional Analysisモジュールの機能

低発現か高発現遺伝子を指定

テストする条件を指定

リファレンスアノテーションを指定
解析結果の遺伝子名と一致する参照アノテーションデータを別途ダウンロードしておく必要があります。

*今回はOmicsBoxに搭載されたBioMartからのデータダウンロード機能を使用し得たファイルを使用 (functional analysis > Load > Load Data from BioMart)。

OmicsBox 2.1.197 - Marta Benegas

File View Help

general tools genome analysis transcript omics functional analysis meta genomics workflows

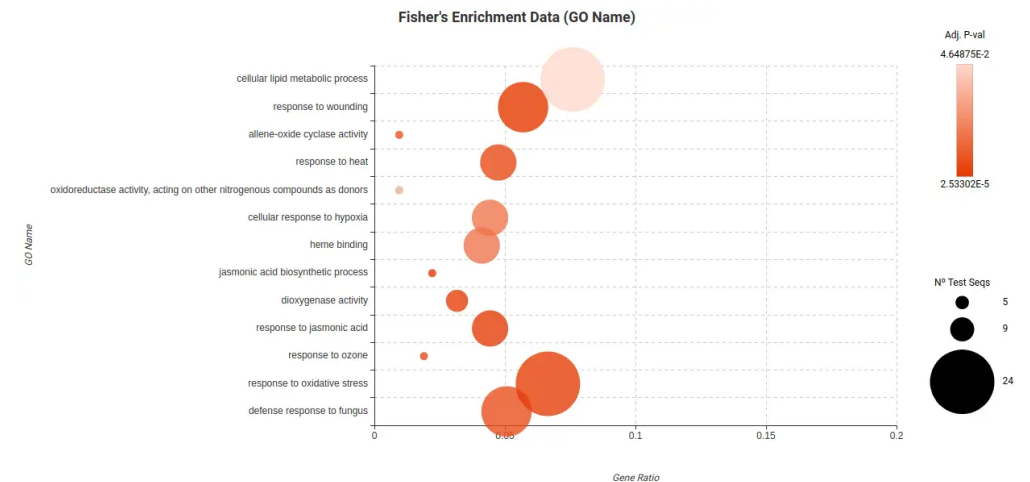
Start typing to search actions...

Sc Results: counts_islets_clustering_results *Sc Pairwise DE Results: scRNA-seq Differential Expression *Table: GO IDs Fisher Result

Table entries: 22,667

Tag	GO Term	GO Name	GO Category	Adj. P-value	P-value	Nr Test	Nr Reference	Not Annot Test	Not Annot Ref
OVER	GO:0045202	synapse	CELLULAR_CO...	1.106967E-17	4.883606E-22	121	1261	869	27101
OVER	GO:0098916	anterograde tr...	BIOLOGICAL_P...	1.537048E-15	2.034299E-19	78	650	912	27712
OVER	GO:0007268	chemical synap...	BIOLOGICAL_P...	1.537048E-15	2.034299E-19	78	650	912	27712
OVER	GO:0099537	trans-synaptic ...	BIOLOGICAL_P...	2.383404E-15	4.205946E-19	78	659	912	27703
OVER	GO:0099536	synaptic signal...	BIOLOGICAL_P...	9.874912E-15	2.178257E-18	78	680	912	27682
OVER	GO:0043005	neuron projection	CELLULAR_CO...	2.950718E-14	7.810609E-18	112	1273	878	27089
OVER	GO:0030424	axon	CELLULAR_CO...	1.014304E-13	3.132363E-17	70	592	920	27770
OVER	GO:0007267	cell-cell signaling	BIOLOGICAL_P...	5.097505E-13	1.799093E-16	124	1561	866	26801
OVER	GO:0030054	cell junction	CELLULAR_CO...	8.884793E-13	3.527734E-16	149	2072	841	26290
OVER	GO:0098793	presynapse	CELLULAR_CO...	1.991231E-11	8.784714E-15	57	467	933	27895
OVER	GO:0042995	cell projection	CELLULAR_CO...	1.356576E-10	6.977503E-14	150	2240	840	26122
OVER	GO:0007399	nervous syste...	BIOLOGICAL_P...	1.356576E-10	7.181769E-14	155	2347	835	26015
OVER	GO:0098794	postsynapse	CELLULAR_CO...	1.639222E-10	9.532343E-14	63	591	927	27771
OVER	GO:0050804	modulation of c...	BIOLOGICAL_P...	1.639222E-10	1.012446E-13	51	408	939	27954
OVER	GO:0099177	regulation of tr...	BIOLOGICAL_P...	1.666247E-10	1.102647E-13	51	409	939	27953
OVER	GO:0120025	plasma membr...	CELLULAR_CO...	1.72215E-10	1.215617E-13	144	2130	846	26232
OVER	GO:0036477	somatodendriti...	CELLULAR_CO...	2.87306E-10	2.154763E-13	74	787	916	27575
OVER	GO:0048812	neuron project...	BIOLOGICAL_P...	1.873406E-9	1.487683E-12	59	567	931	27795
OVER	GO:0120039	plasma membr...	BIOLOGICAL_P...	4.744147E-9	3.976653E-12	59	582	931	27780
OVER	GO:0048858	cell projection ...	BIOLOGICAL_P...	6.204978E-9	5.4749E-12	59	587	931	27775
OVER	GO:0032990	cell part morph...	BIOLOGICAL_P...	6.453879E-9	5.979241E-12	60	605	930	27757
OVER	GO:0097060	synaptic memb...	CELLULAR_CO...	7.377292E-9	7.485671E-12	43	342	947	28020
OVER	GO:0048667	cell morphogen...	BIOLOGICAL_P...	7.377292E-9	7.336465E-12	54	510	936	27852
OVER	GO:0043025	neuronal cell b...	CELLULAR_CO...	9.372837E-9	9.924034E-12	50	451	940	27911
OVER	GO:0044797	cell body	CELLULAR_CO...	1.22163E-8	1.347366E-11	54	519	936	27843

GO Version: Jul 1 2022



エンリッチメント解析の結果のテーブルやチャートを作成できます。
特定の条件（例：クラスターなど）にどのような機能が含まれているか調べることができます。

OmicsBox のシングルセルRNA-seq

- 生データ、あるいは定量済みデータを使用
- Differential Expression Analysisやエンリッチメント解析などの下流分析
- 手動でキュレーションされたCellKBを使用した細胞種アノテーション
- 初心者でも解析できるインターフェース
- 7日間無料のデモライセンス → [詳細\(PDF\)](#)



お問い合わせ先：フィルジェン株式会社

TEL 052-624-4388 (9:00～17 : 00)

FAX 052-624-4389

E-mail: support@filgen.jp