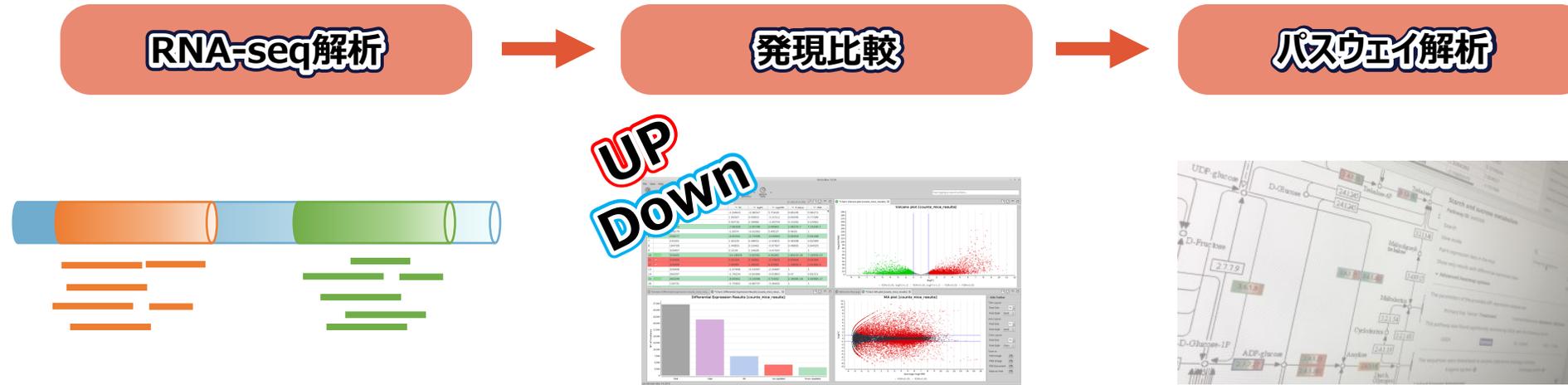




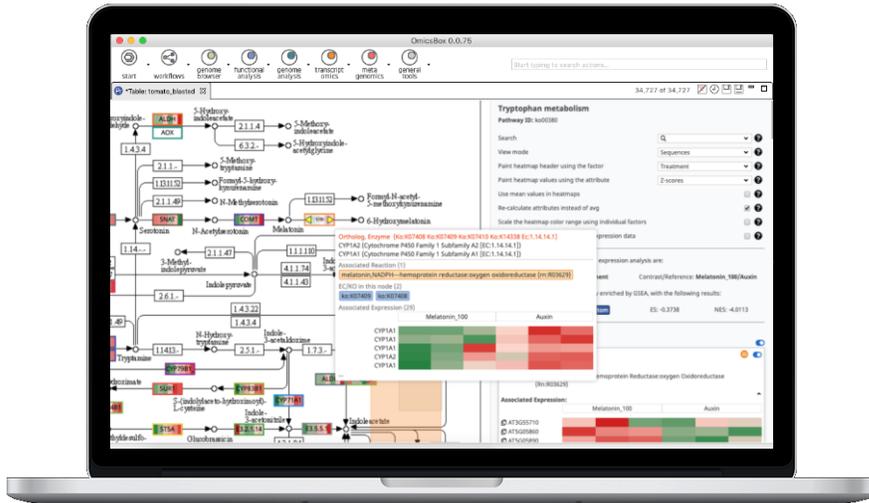
非モデル生物のパスウェイ解析

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

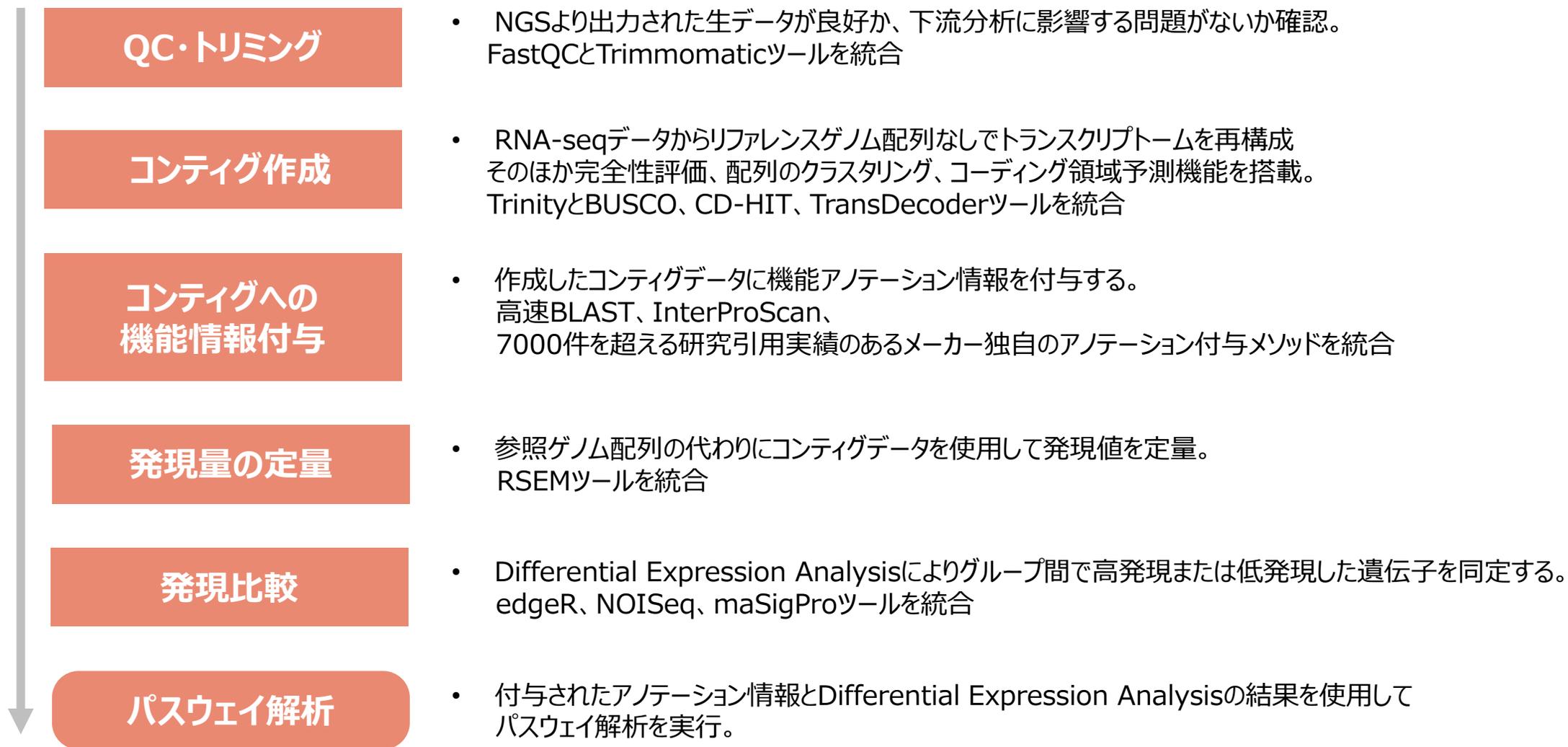


- 一般的なRNA-seq解析のワークフローは発現値を定量し、発現比較により統計的に発現変動を示した遺伝子を抽出する。
- パスウェイ解析は、ほとんどのオミクス研究における実験結果の最終的な生物学的解釈における重要なステップである。データに含まれる生物学的メカニズムの概要を簡単に把握し、結果の解釈を大幅に強化し情報を要約することができる。

OmicsBoxのパスウェイ解析機能



- ❑ 2つのメジャーなパブリックパスウェイデータベース（Reactome・KEGG）
- ❑ シーケンスに関連付けられたパスウェイのテーブル表示
- ❑ 統計的有意性によるソート
- ❑ パスウェイ図の表示
- ❑ データベースごとにリンクされたパスウェイやシーケンスの数などの結果を含むレポート
- ❑ フィッシャーの統計的手法を使用してパスウェイエンリッチメント解析
- ❑ Gene Setエンリッチメント解析
- ❑ パスウェイ内に含まれる情報を検索
- ❑ 高度なヒートマップオプション





・データが良好か、下流分析に影響する問題がないか確認

Welcome Message | FASTQ Quality Check (Dataset) | FASTQ Quality Check (ERR1948631_1.fastq) | FASTQ Quality Check (clean_ERR1948631_1.fq) | Chart: Adapter Content

FASTQ Quality Check

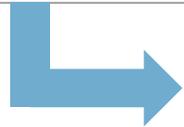
Name: Dataset

Overall Results

Name	Per Base Sequence Quality	Per Sequence Quality Scores	Per Base Sequence Content	Per Sequence GC Content	Per Base N Content
ERR1948631_1.fastq	PASS	PASS	FAIL	PASS	PASS
clean_ERR1948631_1.fq	PASS	PASS	FAIL	PASS	PASS

Name	Sequence Length Distribution	Adapter Content	Overrepresented Sequences	Sequence Duplication Levels	Report
ERR1948631_1.fastq	PASS	FAIL	WARNING	FAIL	🔍
clean_ERR1948631_1.fq	WARNING	PASS	WARNING	FAIL	🔍

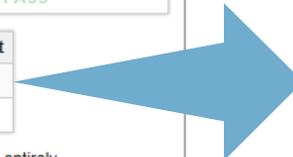
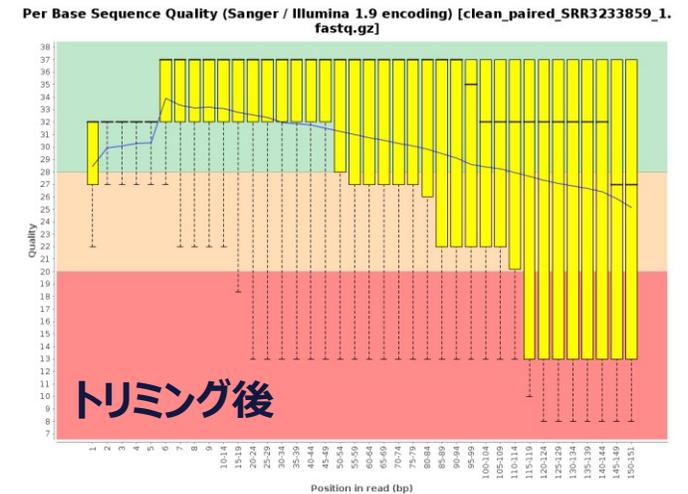
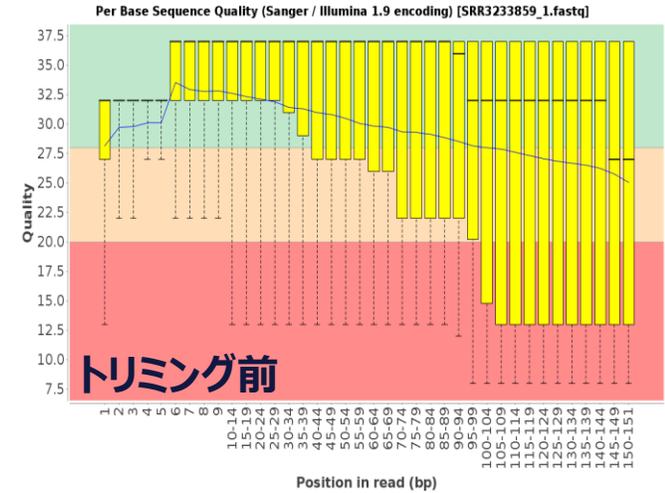
The FASTQ quality check task is performed by nine analysis modules. The table above provides a quick evaluation of whether the results of each module seem entirely normal (pass), slightly abnormal (warning) or very unusual (fail). Note that these evaluations must be taken in the context of what is expected from the library. For example, some experiments may be expected to produce libraries which are biased in particular ways. Therefore, the summary evaluations should be treated as pointers that guide the preprocessing of the libraries.



✓解析が終了するとレポートが作成

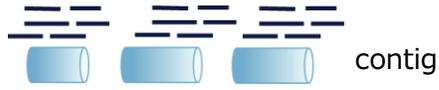
正常 (PASS)
わずかに異常 (WARNING)
異常 (FAIL)

シーケンスデータの品質をすばやく評価



レポートのアイコンをクリック→さらに詳細な結果を見ることが可能

RNA-Seq de novo assembly



完全性評価



配列のクラスタリング



コーディング領域予測



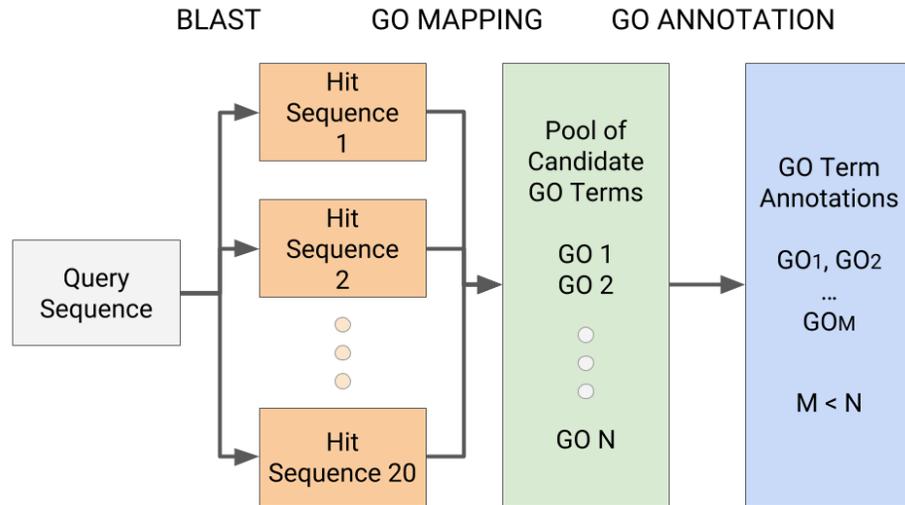
タンパク質配列
CDS配列
ORF座標

リストが作成される

- RNA-seqデータからリファレンスゲノム配列なしでトランスクリプトームを再構成
- 長い連続したシーケンス（コンティグ）を作成する
- 計算負荷のかかる解析だがOmicsBoxではメーカサーバ上で安定かつ高速解析される
- コンティグデータの完全性を評価
- 90%を超える種でsingle-copy orthologを持つグループに対してアセンブルされた配列の中に存在するか確認
- de novo assemblyで作成されたコンティグデータから類似する配列を取り除く
- 重複したコンティグを除くことで冗長性を減らす
- 転写産物配列内のコーディング領域を検出
- 信頼の高いORFが抽出され、機能を予測するために使用される。

fasta形式で保存→発現量の定量で使用する

メーカー開発のBlast2GOアノテーションにより信頼性の高いGO情報を割り当てる



- BLAST : クエリセットに類似した配列を検索
特にBLASTの解析は解析時間が必要なケースが多く中には数日かかることも
OmicsBoxではメーカークラウド上で高速検索することが可能*

- GO Mapping : GO termを取得

- GO Mappingの結果に対しBLAST2GO Annotationを実行



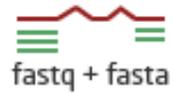
* Computation Unitsは、CloudBlast解析を行うごとにUnitが消費されます。
(InterProScanのCloud解析でも消費) [詳細は弊社HP](#)

- InterProScan : タンパク質のドメイン構造やモチーフを検索

InterProScanの結果をアノテーション結果にマージして最適な機能アノテーション情報を付与できる。

.box形式で保存→パスウェイ解析で使用する

● Transcript-level Quantification



The transcript-level quantification tool estimates gene and isoform expression from RNA-Seq reads (FASTQ). It is based on the RSEM software package, which allocates multi-mapping reads among transcripts using an expectation maximization approach. This tool requires a set of reference transcript sequences (FASTA), such as one produced by a de novo transcriptome assembler. It is executed via the BioBam Bioinformatics Cloud Platform.

- OmicsBoxはリファレンスゲノムを必要としないトランスクリプト定量パッケージを搭載
- 参照ゲノム配列の代わりにトランスクリプトシーケンスのセット（コンティグデータ）を使用

Nr	Name	SRR4044961	SRR4044963
1	TRINITY_DN801_c0_g1_i3.p1	22	5
2	TRINITY_DN10550_c0_g1_i1.p2	7	0
3	TRINITY_DN10550_c0_g1_i1.p1	52	31
4	TRINITY_DN412_c0_g1_i30.p1	1027	1044
5	TRINITY_DN19849_c0_g1_i1.p1	16	8
6	TRINITY_DN37230_c0_g1_i1.p1	6	9

- 発現レベルに関するテーブルが作成される
- OmicsBoxではCount Tableとして保存する。

前項で作成した定量データを使用して、差次的に発現する遺伝子群を特定。

3つの戦略でデータを比較することが可能



- Pairwise Differential Expression Analysis
 - ペアワイズ差次的発現解析により、さまざまな実験条件を考慮して差次的に発現する遺伝子を同定。



- Pairwise Differential Expression Analysis (Without Replicates)
 - ペアワイズ差次的発現分析 (Replicateなし) は、どの実験条件にもReplicateがない場合に適している。



- Time Course Expression Analysis
 - 経時的RNA-Seq実験で有意な発現プロファイルの違いがある遺伝子を検出する。

Experimental Design

Sample	Genotype
SRR4044961	wild_type
SRR4044962	wild_type
SRR4044963	7B1_mutant
SRR4044964	7B1_mutant

```
experimental_design - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ^J
Sample Genotype
SRR4044961 wild type
SRR4044962 wild type
SRR4044963 7B1 mutant
SRR4044964 7B1 mutant
```

別途実験条件ファイルを用意
(タブ区切りテキストファイル)

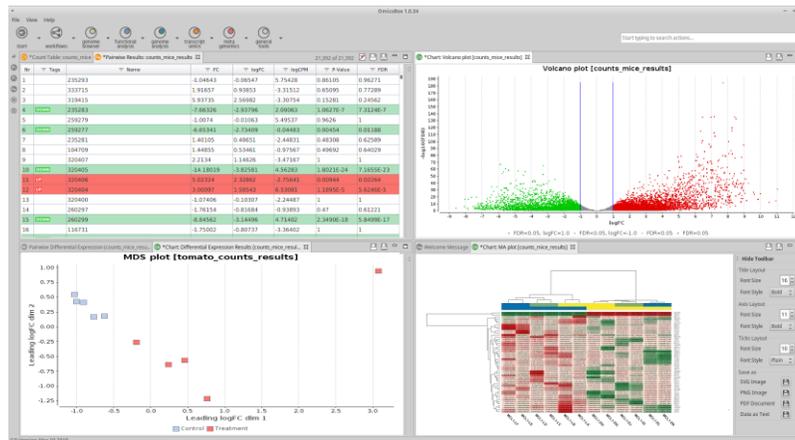
発現比較

Nr	Tags	Name	FC	logFC	logCPM	P-Value	FDR
1		235293	-1.04643	-0.06547	5.75428	0.86105	0.96271
2		333715	1.91657	0.93853	-3.31512	0.65095	0.77289
3		319415	5.93735	2.56982	-3.30754	0.15281	0.24562
4	Down	235283	-7.66326	-2.93796	2.09063	1.0627E-7	7.3124E-7
5		259279	-1.0074	-0.01063	5.49537	0.9626	1
6	Down	259277	-6.65341	-2.73409	-0.04483	0.00454	0.01188
7		235281	1.40105	0.48651	-2.44831	0.48308	0.62589
8		104709	1.44855	0.53461	-0.97567	0.49692	0.64029
9		320407	2.2134	1.14626	-3.47167	1	1
10	Down	320405	-14.18019	-3.82581	4.56283	1.8021E-24	7.1655E-23
11	UP	320406	5.02324	2.32862	-2.75641	0.00944	0.02264
12	UP	320404	3.00097	1.58543	6.53081	1.1895E-5	5.6246E-5
13		320400	-1.07406	-0.10307	-2.24487	1	1
14		260297	-1.76154	-0.81684	-0.93893	0.47	0.61221
15	Down	260299	-8.84562	-3.14496	4.71402	2.3490E-18	5.8499E-17
16		116731	-1.75002	-0.80737	-3.36402	1	1
17	Down	104732	-3.47919	-1.79875	3.25074	2.2662E-5	1.0070E-4
18		103406	1.29735	0.37557	2.48284	0.36875	0.50263
19	UP	116701	18.51428	4.21057	6.60578	1.3707E-17	3.1294E-16
20		104718	2.49063	1.31651	5.74437	0.07328	0.13301
21		547127	-1.80941	-0.85552	3.01372	7.1112E-4	0.00227
22		547109	-1.3601	-0.44372	-2.26423	0.71616	0.83479
23		104721	-1.50558	-0.59032	6.55243	0.00155	0.00455
24		67071	2.43207	1.23266	3.26233	0.26439	0.63081

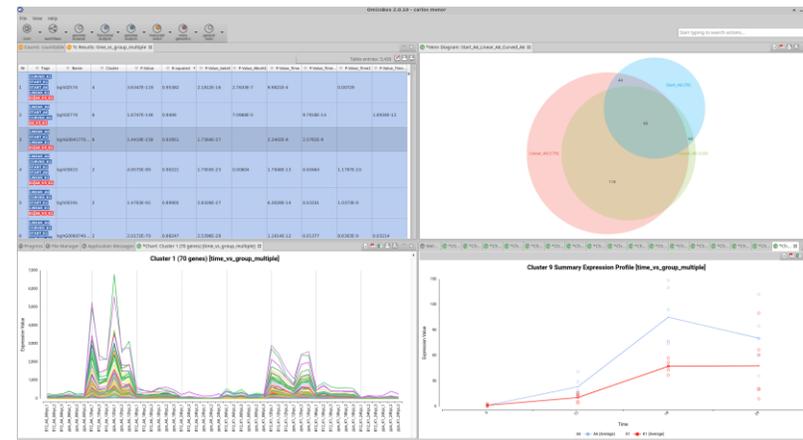
【結果】

アップレギュレート、ダウンレギュレートのタグ付けされたテーブルが作成される。

.box形式で保存→パスウェイ解析で使用する



Pairwise Differential Expression Analysis



Time Course Expression Analysis

Combined Analysis (KEGG, Reactome)

Input

The analysis extracts and combines pathway information using Gene Ontology terms, Enzyme codes and sequence data as input.
This information can be provided in various formats:

- FASTA file with nucleotide or protein sequences. Linking method: BLAST against Reactome and EggNog search for KEGG.
- Annotation file (.annot) which contains gene ontology and/or enzyme codes. Linking method: ID matching.
- OmicsBox project file (.box) which can contain both, sequence information and annotations (recommended).

Sequences (.box, .fasta, .annot) Browse...

C:\Users\Marta\Downloads\testing\fasta_file.box

Pathway information will be more meaningful when combined with expression data. Differential expression has to be provided as .box file. Pairwise comparisons as well as time series are supported. These files can easily be generated from any count table from within OmicsBox.

Differential Expression Data (.box) Browse...

Pathway Enrichment Analysis

Gene Set Enrichment Analysis

Fisher's Exact Test

Default < Back **Next >** Run Cancel

これまで作成したファイルを指定。

- シーケンス情報 (.fasta、.box) 、GO / ECアノテーション (.annot、.box) 、またはその両方 (.box) を含むファイルを選択。
= コンティグデータの (.box) ファイル
- 前項で作成したペアワイズやタイムコースなどの Differential Expression オブジェクト (.box) を選択
パスウェイオブジェクトに有用な情報が追加される。
 - 実験計画情報
 - カウント値
 - 差次的に発現する特徴
- エンリッチメント解析を有効にして、リンクされた配列の数、発現変動した配列、あるいはグループの中でパスウェイがより頻繁に関連しているかソートすることができる。

Combined Analysis (KEGG, Reactome)

Configuration Reactome Pathway Analysis

Reactome is free, manually curated and peer-reviewed database of over 2500 pathways. Pathway consists of reactions. Reactions are defined as any event in biology that changes the state of a biological molecule like e.g. binding, activation, translocation or degradation. Reactome includes inferred orthologous reactions of + 15 non-human species.

Run Reactome Pathway Analysis

Pathway Linking Options

Run Blast to link via Protein IDs

Link with GeneOntology Terms

Filtering Options

Keep Most Specific Pathways

Give Priority to Taxon

Top Priority Taxon Homo sapiens

Blast Expectation Value 1.0E-3

Include Categories Disease, Gene expression (Transcription), Protein localiza

Please Cite:
 Fabregat A et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic acids research*, 46(D1), D649-D655.

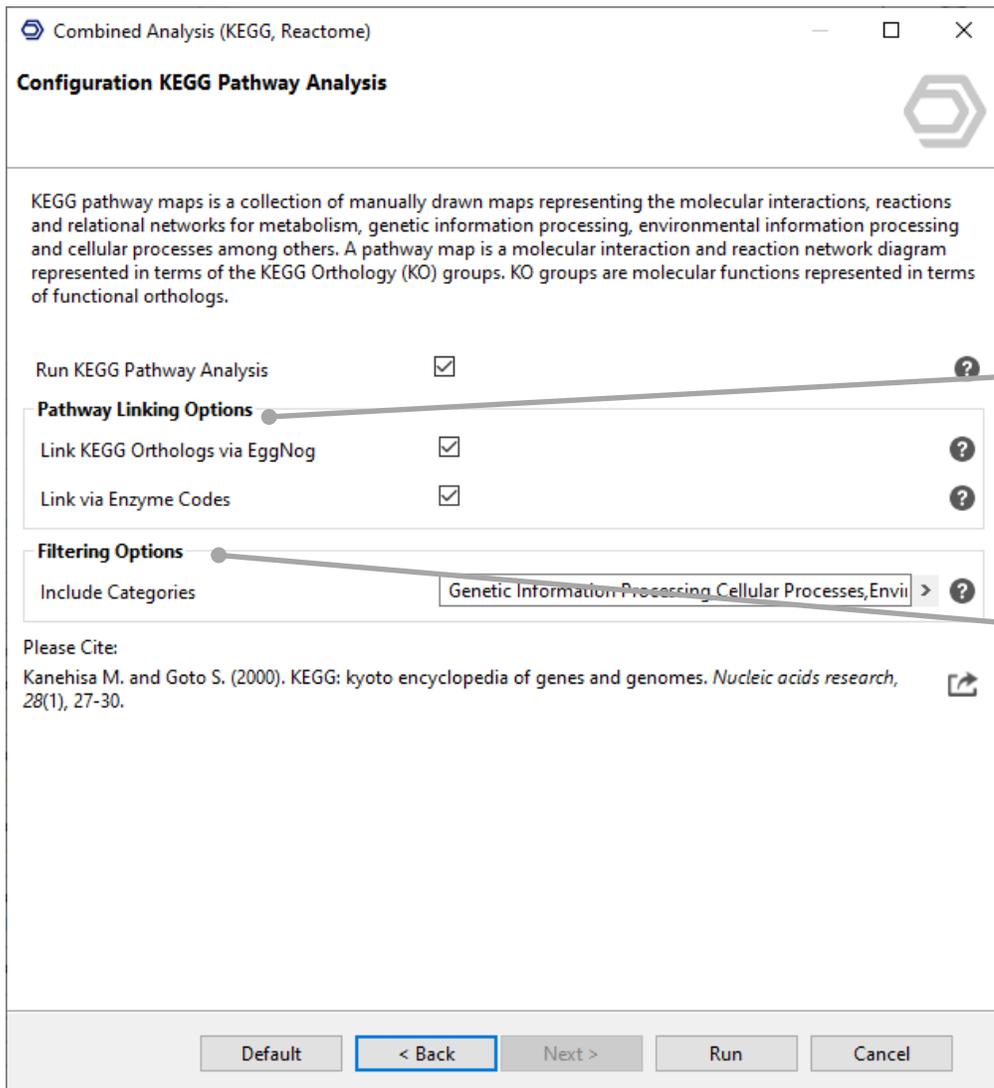
Default < Back **Next >** Run Cancel

Reactomeデータベースに関する設定

ヒトの生物学におけるパスウェイと反応の精査されたデータベースだが、他の15の非ヒト種の推定されるオルソログを含んでいる。

- Reactomeのパスウェイに関連するすべてのUniprotタンパク質の配列を含むカスタムデータベースに対してBLASTを実行。^{*} 入力したファイルにGO termsが含まれる場合はその情報を使用し、データベースの情報と紐づけることができる。
- 同じパスウェイだが種固有の異なるバージョンが含まれる。パスウェイは階層的に構成されている。
 → 結果として出力テーブルに、興味のないエントリが多すぎる可能性がある。この設定で可能な限り類似したエントリーを破棄する。

^{*} CloudBlastで高速解析されるためComputation Unitsが消費されます。



KEGGデータベースに関する設定*

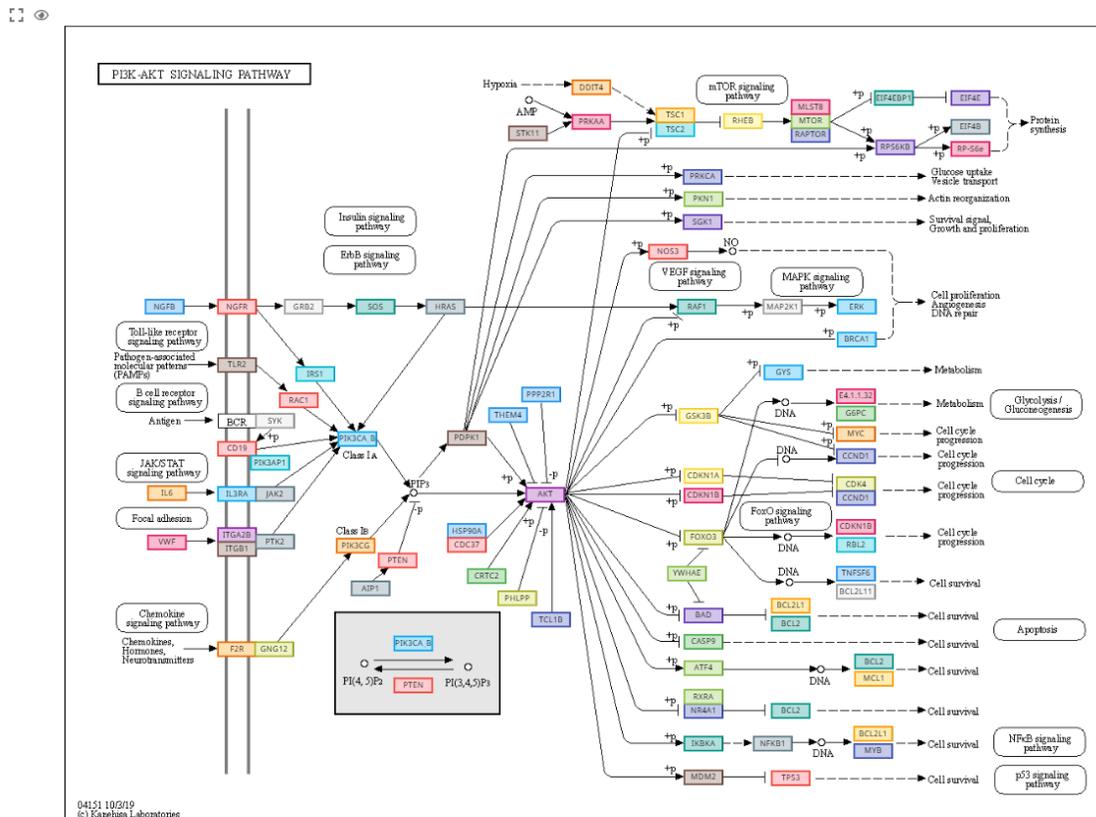
代謝、遺伝子情報処理、細胞プロセス、生物系、病気、および医薬品開発のための分子相互作用、反応、および関係ネットワークの知識を含むパスウェイマップのコレクション。

- 配列 (fastaファイル) が与えられた場合EggNOG アノテーションファイル(.annot or .box)が与えられた場合Enzyme Codesが KEGGパスウェイを検索するために使用される。
- 対象のカテゴリのみを選択することにより、検出されるパスウェイ数が減る。これは、エンリッチメント解析で実行される多重検定の修正にプラスの影響を与える可能性がある。

* 本機能はアカデミックユーザー限定機能です。

Database	Pathway	ID	Species	#Diff. Expr Seqs	GSEA Tag	GSEA NES	Fisher Tag	
KEGG	Huntington disease	ko05016	None	2	TOP	2.7651854	UNDER	0.003
KEGG	Amyotrophic lateral sclerosis	ko05014	None	3	TOP	3.1286435	UNDER	0.003
KEGG	Coronavirus disease - COVID-19	ko05171	None	3	TOP	5.471398	UNDER	0.003
KEGG	Alzheimer disease	ko05010	None	4	TOP	2.3306258	UNDER	0.045
KEGG	Styrene degradation	ko00643	None	5		.	OVER	0.003
KEGG	Aminobenzoate degradation	ko00627	None	5		.	OVER	0.001
KEGG	Pathways of neurodegeneration - multiple diseases	ko05022	None	6	TOP	2.3956614	UNDER	0.022
Reactome	Trehalose biosynthesis	R-MTU-868688	Mycobacteriu...	7		.	OVER	0.036
KEGG	Anthocyanin biosynthesis	ko00140	None	7		.	OVER	0.000
KEGG	Flavonoid biosynthesis	ko00940	None	7	BOTTOM	-3.1903944	OVER	0.000
Reactome	Cellular hexose transport	R-CEL-10000	None	8		.	OVER	0.036
Reactome	PPARA activates gene expression	R-PPA-10000	None	8	BOTTOM	-2.1536667	OVER	0.036
KEGG	Insulin resistance	ko00940	None	8		.	OVER	0.010
KEGG	Galactose metabolism	ko00940	None	8	BOTTOM	-1.8304862	OVER	0.008
KEGG	Longevity regulating pathway - multiple species	ko00940	None	8		.	OVER	0.005
KEGG	Cyanocamino acid metabolism	ko00940	None	8	BOTTOM	-2.5056581	OVER	0.008
KEGG	Nitrogen metabolism	ko00940	None	8	BOTTOM	-1.8590983	OVER	0.000
KEGG	Antigen processing and presentation	ko00940	None	8	BOTTOM	-2.9050708	OVER	0.006
KEGG	Tryptophan metabolism	ko00940	None	8	BOTTOM	-4.0113273	OVER	0.004
KEGG	Starch and sucrose metabolism	ko00500	None	21	BOTTOM	-2.0570998	OVER	0.008
KEGG	Phenylpropanoid biosynthesis	ko00940	None	22	BOTTOM	-4.4628797	OVER	0.006
KEGG	MAPK signaling pathway - plant	ko04016	None	24	BOTTOM	-2.7887363	OVER	0.004
KEGG	Plant hormone signal transduction	ko04075	None	28	BOTTOM	-3.00218	OVER	0.045

- シーケンスに関連付けられたパスウェイが特定されテーブルが開く。
- シーケンスの総数よりも、エンリッチメント解析による統計的有意性によるソートが優先される。
- 重要な結果を見つけるため、結果テーブルのフィルタリングと並べ替えを組み合わせる。最も差次的に発現する配列を持つパスウェイをソートし、エンリッチメント解析のされたタグTOPまたはBOTTOMのみをフィルタリングする。これにより、多くの差次的に発現する遺伝子を含む統計的にエンリッチメントされたパスウェイを見つけることが可能。



PI3K-Akt signaling pathway
Pathway ID: ko04151

Search:

View mode:

Paint expression data in the map:

Show only results with differential expression data:

Advanced heatmap options:

The parameters of the provided diff. expression analysis are:
Primary Exp. Factor: Stage Contrast/Reference: Lactate/Pregnant

This pathway was found significantly enriched by GSEA, with the following results:

GSEA	bottom	ES: -0.0750	NES: -1.6239
------	--------	-------------	--------------

Linked KO Sequences:

15 sequences with:
KO IFNA (Ko:K05414)
No Reaction Interferon Alpha (IFNA)

Associated Expression:

	Lactate	Virgin	Pregnant
15964	Red	Green	Green
15974	Red	Green	Green
230398	Red	Green	Green
319146	Red	Green	Green
435791	Red	Green	Green
545645	Red	Green	Green
545646	Red	Green	Green
545648	Red	Green	Green
545649	Red	Green	Green
545650	Red	Green	Green
545651	Red	Green	Green
545652	Red	Green	Green
545654	Red	Green	Green
545655	Red	Green	Green
668208	Red	Green	Green

12 sequences with:
KO ...

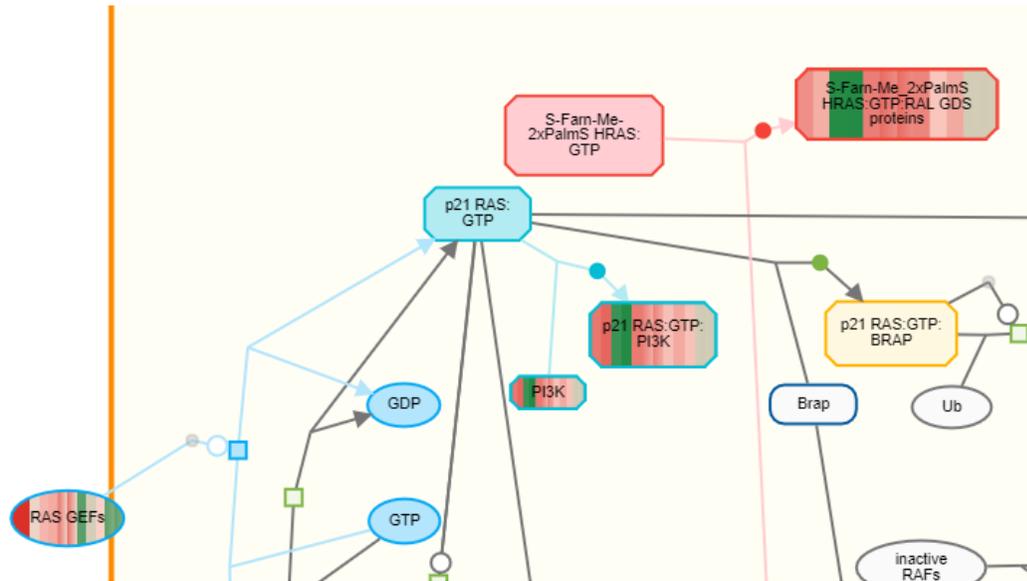
パスウェイマップ

- デフォルトのマップビューモードでは、背景色が各用語または反応に割り当てられた単色で表示される。
- サイドバーパネルからパスウェイの表示情報を変更
- 情報パネルからパスウェイの詳細情報や発現データと関連する配列のヒートマップ、表示情報のON、OFFが可能

Paint expression data in the map



Show only results with differential expression data



表示を切り替えることで・・・

- パスウェイ中にサンプル間のヒートマップを反映
- 発現変動を示した遺伝子のみを表示

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: biosupport@filgen.jp