

AJACS-training / AJACS66

 Watch 2  Star 0  Fork 2

 Code

 Issues 0

 Pull requests 0

 Projects 0

Insights ▾

Branch: master ▾

AJACS66 / 03_bono /

Create new file Find file History

 bonohu committed on GitHub Create README.md

Latest commit fb61d2a 7 days ago

..

 README.md

Create README.md

7 days ago

 README.md

遺伝子発現DBを含む公共オミックスDBの使い方 at AJACS河内

情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター

坊農 秀雅 bono@dbcls.rois.ac.jp

2017年8月24日

これは統合データベース講習会AJACS河内「遺伝子発現DBを含む公共オミックスDBの使い方」の資料です。

概要

本講習は、だれでも自由に使うことができる公共DBやウェブツールを活用して、研究のさまざまな場面で調べることの多い個々のオミックスDBを簡単に調べるための方法と基礎知識について学びます。とくに、需要の増している公共DBから遺伝子発現データを検索し取得してくる方法について詳しく説明、実習します。また、自ら行なった大規模発現解析の(あるいは公共DBから取得・解析した)結果として得られた数百～数千におよぶ遺伝子セットについて、生物学的な解釈をする方法とその結果の考察を実践します。

講習の流れ

今回の講習では、以下の内容について順番に説明します。【実習】の部分はみんなでやる実習、【応用】は早くできてしまった人用の応用課題です。

- 研究現場で頻繁に使われるDBやツールを知る: 統合TV
- 公共オミックスDBとは
- 公共オミックスDBの使い方
 - EBI ArrayExpress
 - 【実習1】ArrayExpressを使って、興味ある実験データセットを検索する
 - NCBI Gene Expression Omnibus (GEO)
 - AOE (All of gene expression)
 - 【実習2】AOEを使って、興味ある実験データセットを絞り込む
- 個々の遺伝子の発現プロファイルなどを調べる
 - 【実習3】RefExを使って、組織特異的遺伝子を検索する
 - 【応用1】ChIP-Atlasを使って注目の遺伝子がコードされたゲノム領域にあるChIP-seqピークを検索する
- 数十～数千の遺伝子群の生物学的解釈
 - 【実習4】metascapeを用いて、発現データの結果を生物学的に解釈する
 - 【応用2】複数のエンリッチメント解析ツールを用いて、発現データの結果を生物学的に解釈する

講習に際しての注意とお願ひ

- みんなで同時にアクセスするとサイトにつながりにくくなることが予想されます。
 - 資料を見ながら自力で進められそうな方はどんどん先に、そうでない方は講師と一緒にすすめていきましょう。
 - サイトの反応が悪い時はタイミングをずらして実行してみてください。
 - 反応が無いからと言って何度もクリックするとますます繋がらなくなってしまいます。おおらかな気持ちで臨みましょう。
- わからないことがあつたら挙手にてスタッフにお知らせください。
 - 遠慮は無用です(そのための講習会です!)。おいてければ楽しくありません。

研究現場で頻繁に使われるデータベースやツールを知る

統合TV

- 生命科学分野の有用なデータベースやツールの使い方を動画で紹介するウェブサイト
 - <http://togotv.dbcls.jp/>

The screenshot shows the TOGO-TV website interface. At the top, there's a navigation bar with links for 'はじめての方へ' (For beginners), '再生数ランキング' (Ranking by views), and 'お問い合わせ・番組をリクエスト' (Contact us). Below the navigation, a search bar contains the query '全番組のリストから、調べたいDBやウェブツールに関するキーワードで検索! (全 1266 件)'. To the left, there's a sidebar with a '目的別に検索' (Search by purpose) section containing categories like '講習会 実習資料 (AJACS)', 'ゲノム・核酸 配列解析', 'タンパク質 配列・構造解析', '発現制御解析・可視化', '文献・辞書・プログラミング', '著名データベース', 'その他講演・講習会', and '自由に使える画像を探す'. Below that is a '関連するタグから検索' (Search by related tags) section with tags such as 'ゲノム (216)', '遺伝子 (341)', 'タンパク質 (170)', '配列解析 (190)', '発現解析 (270)', 'NGS (202)', and '文献検索 (196)'.

- YouTube版もあります <http://youtube.com/togotv/>
- ウェブサイトへのアクセスから結果の見方まで、操作の一挙手一投足がわかります。
 - 講義・講習などの参考資料や後輩指導の教材として利用できます。
 - 本講習中、本家サイトが繋がらない時は、統合TVのYouTube版を見ればおおよその内容がわかるようになります。
 - 今回の講習に関連する内容の多くは、「発現解析」タグのついた動画にあります。
 - 過去の講習会の内容はそのほとんどが統合TVに収録されており、いつでもどこでも繰り返し復習できるようになっています。
- お探しの動画が見つからない or 統合TV未掲載の場合は、[統合TV番組リクエストフォーム](#)へどうぞ!
- 統合TVを作ってくれる方、募集中!!

公共オミックスDBとは

mind the gap

遺伝子発現解析などのオミックス解析→DB(というか、公共アーカイブ)に登録

- 日本では論文投稿前が現在一般的
- 欧米では研究費の条件で多くの場合データを出したらわりとすぐ

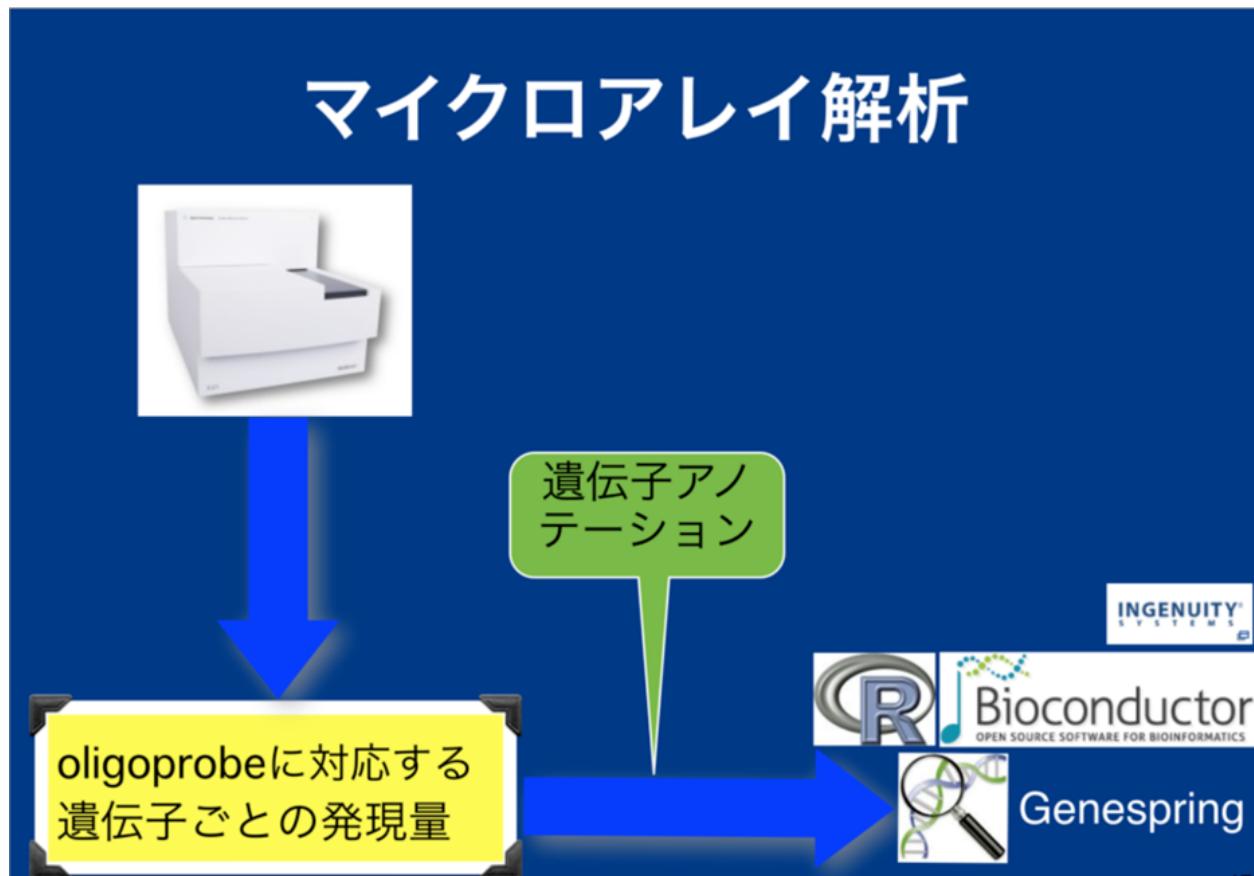
その歴史

- マイクロアレイの発明→網羅的遺伝子発現定量が可能に→遺伝子発現DB

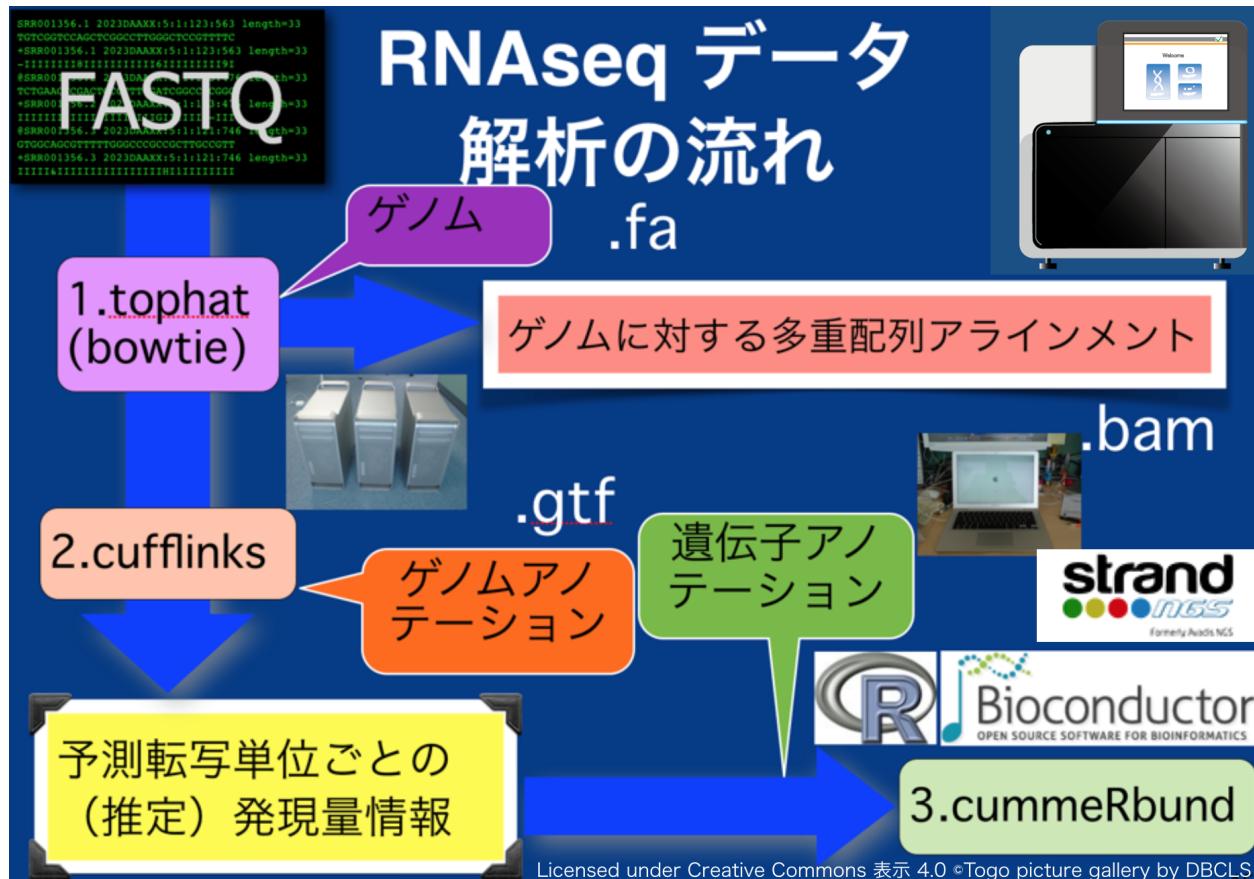
- アレイのイメージデータ(CELファイルなど)
- 定量データ(Series Matrix File)
 - DDBJ CIBEX (更新停止)
 - NCBI Gene Expression Omnibus (GEO)
 - EBI ArrayExpress
 - ArrayExpressはGEOのデータを定期的に取り込み続けている→こちらの使用を推奨
- 次世代シーケンサー(NGS: Next Generation Sequencer)の発明→個々のサンプルでのtranscript sequencing (RNA-seq)が現実的に
 - データはNGS配列DB(SRA: Sequence Read Archive) and/or 遺伝子発現DBへ?
 - NCBI, EBIでは遺伝子発現DB(それぞれGEO, ArrayExpress)に登録すれば、配列データ(FASTQファイル)がSRAにも登録される状況
 - DDBJ(日本)もその仕組みを現在構築中
 - 現状、どこかの公共DBに登録されていればOK

発現定量のステップ

マイクロアレイ



NGS(RNA-seq)



- 詳しい手順は、次世代シーケンサーDRY解析教本「Level2② 発現解析」p97-141などを参考に



公共オミックスDBの使い方

EBI ArrayExpress

- EBIが提供・維持管理している遺伝子発現情報のデータベース
 - <http://www.ebi.ac.uk/arrayexpress/>
 - NCBI GEOのデータも取り込んでいて、こちらのほうがデータ数が多い
 - インターフェースが使いやすい

【実習1】 ArrayExpressを使って、興味あるマイクロアレイの実験データセットを検索する

1. ArrayExpressのサイトにアクセス、検索窓に'cancer'と入力

The screenshot shows the ArrayExpress homepage. At the top, there's a navigation bar with links for Services, Research, Training, and About us. Below the navigation is a search bar containing the word 'cancer'. A tooltip below the search bar lists examples: E-MEXP-31, cancer, p53, Geuvadis. To the right of the search bar are buttons for Search and Advanced. The main title 'ArrayExpress – functional genomics data' is centered above a brief description of the archive. Below the description is a 'Browse ArrayExpress' link.

2. 候補となるキーワードがいろいろ出てきました。cancerの左にある'+'をクリックして解いてみましょう

This screenshot shows the same ArrayExpress homepage as the previous one, but with a dropdown menu expanded over the search bar. The dropdown contains a list of cancer-related terms, starting with 'cancer' and including 'carcinoma', 'central nervous system cancer', 'childhood cancer', 'colorectal cancer', 'eccrine sweat gland cancer', 'gum cancer', 'lymphoid neoplasm', 'mesothelioma', 'oral cavity cancer', 'pharynx cancer', 'sarcoma', and 'tumour of cranial and spinal nerves'. At the bottom of the dropdown is a link labeled 'cancer aggressiveness measurement'.

3. 今回はそのまま'Search'ボタンを押して、cancerで検索してみます

4. 黄色でハイライトされた'cancer'以外に、オレンジ色でcancer関係のキーワードもハイライトされているのを確認しましょう

Page 1 2 3 4 5 6 .. 608		Showing 1 - 25 of 15189 experiments				Page size 25 50 100 250 500			
Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEO-71472	Effect of DNMT3A R882H mutation or WT expression on epigenetic landscapes of hematopoietic stem/progenitor cells with NRAS G12D co-transduction (ChIP-seq)	ChIP-seq	Mus musculus	13	23/06/2016		-	-	-
E-GEO-71439	Expression profiling of murine leukemia stem cell (LSC) lines established ex vivo by coexpression of R882H-mutated DNMT3A and NRAS-G12D post treatment with Dot1l inhibitor (Microarray)	transcription profiling by array	Mus musculus	10	23/06/2016			-	-
E-GEO-71437	Effect of DNMT3A R882H mutation or WT on gene expression in hematopoietic stem/progenitor cells with NRAS G12D co-transduction (Microarray)	transcription profiling by array	Mus musculus	12	23/06/2016			-	-
E-GEOD-39292	Oncogenic EGFR Represses the TET1 DNA Demethylase to Induce Silencing of Tumor Suppressors in Cancer Cells	transcription profiling by array	Homo sapiens	12	23/06/2016		-	-	-
E-MTAB-2672	Functional and genetic heterogeneity of distinctive leukaemic stem cell populations in CD34- human Acute Myeloid Leukaemia	RNA-seq of coding RNA	Homo sapiens	124	22/06/2016	-		23	-
E-GEOD-83607	Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at P ^c G target genes in <i>Arabidopsis</i> , and relocation of H3K27 trimethylation to transposons [RNA-seq]	RNA-seq of coding RNA	Arabidopsis thaliana	2	22/06/2016		-	-	-
E-GEOD-83606	Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at P ^c G target genes in <i>Arabidopsis</i> , and relocation of H3K27 trimethylation to	ChIP-chip by tiling array	Arabidopsis thaliana	6	22/06/2016			-	-

5. 単にcancerではヒット数が多いので、'oral cancer'で検索してみましょう

EMBL-EBI

ArrayExpress

Search results for **oral cancer**

Showing 1 - 25 of 231 experiments

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEOID-73521	NUPR1 a new target in liver cancer : implication in controlling cell growth, migration, invasion and sorafenib resistance	transcription profiling by array	Homo sapiens	2	15/06/2016			2	-
E-MTAB-4863	Copy number Analysis of Head and Neck cancer cases from two high incidence regions	genotyping by array	Homo sapiens	225	14/06/2016	-		-	-
E-GEOID-75540	A minimal DNA 5-methylcytosine signature of oral tongue squamous cell carcinoma	transcription profiling by array, methylation profiling by array	Homo sapiens	151	01/06/2016		-	29	-
E-GEOID-75539	A minimal DNA 5-methylcytosine signature of oral tongue squamous cell carcinoma [HumanHT-12 V4.0]	transcription profiling by array	Homo sapiens	15	01/06/2016		-	10	-
E-GEOID-75538	A minimal DNA 5-methylcytosine signature of oral tongue squamous cell carcinoma [HumanHT-12 WG-DASL V4.0]	transcription profiling by array	Homo sapiens	28	01/06/2016		-	9	-
E-GEOID-75537	A minimal DNA 5-methylcytosine signature of oral tongue squamous cell carcinoma [methylation]	methylation profiling by array	Homo sapiens	108	01/06/2016		-	11	-

6. デフォルトでは'Released'(データの公開日)が新しいものから古いもの順にならんでいます。'Views'(閲覧数)をクリックして並べ替えてみましょう

EMBL-EBI

ArrayExpress

Search results for **oral cancer**

Showing 1 - 25 of 231 experiments

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-62	Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines	transcription profiling by array	Homo sapiens, Mus musculus	5372	04/08/2010			41982	-
E-MTAB-2706	RNA-seq of 675 commonly used human cancer cell lines	RNA-seq of coding RNA	Homo sapiens	675	25/11/2014	-		12209	
E-TABM-185	Transcription profiling by array of integrated human experiments involving the hgU133a platform to investigate a global map of human gene expression	transcription profiling by array	Homo sapiens	5896	20/11/2007			9474	-
E-MTAB-3732	A comprehensive human expression map	transcription profiling by array	Homo sapiens	27871	23/07/2015		-	9205	-
E-GEOID-7307	Transcription profiling by array of human samples with various different diseases	transcription profiling by array	Homo sapiens	677	20/06/2008			7249	-
E-TABM-302	Transcription profiling of human head and neck squamous cell carcinomas to identify metastasis predictors and descriptors, and differentiation classifiers	transcription profiling by array	Homo sapiens	81	23/08/2007			1777	-

7. 詳細を見るには'Accession'のリンクをクリックします。番号の意味は[こちらを参照](#)

Home | Browse | Submit | Help | About ArrayExpress | Feedback | Login

ArrayExpress > Search results for "oral cancer" > E-MTAB-62

E-MTAB-62 - Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines

Status	<i>Released on 4 August 2010, last updated on 2 May 2014</i>
Organism	Homo sapiens, Mus musculus
Samples (5372)	Click for detailed sample information and links to data └ found inside: 600MPE breast cancer, a498 renal cell carcinoma, A549 lung adenocarcinoma, A673 Ewing tumor, ABC-1 lung adenocarcinoma, acute lymphoblastic leukemia, acute megakaryoblastic leukaemia, acute monoblastic and monocytic leukemia, acute myeloid leukemia, acute myelomonocytic leukemia, acute promyelocytic leukemia, alveolar rhabdomyosarcoma, AMO-1 myeloma, AU565 breast carcinoma, B-cell lymphoma, breast cancer, leukemia, nervous system neoplasm, non breast carcinoma, sarcoma, ...
Array (1)	A-AFFY-33 - Affymetrix GeneChip Human Genome HG-U133A [HG-U133A]
Protocols (1)	Click for detailed protocol information
Description	The experiment contains systematically annotated and consistently normalized human gene expression data matrix of 5372 samples integrated from 206 public experiments of a HG-U133A array platform. The dataset is a subset of a larger pool of 9004 samples gathered from ArrayExpress and GEO websites and checked for quality assessment and suitability for data co-analysis as described by Bolstad et al 2005. The sample annotations have been subject to semiautomatic curation and manual generalization of 369 biological groups which have been additionally organized into blood/non-blood, 14 and 15 meta-classes.
Experiment types	transcription profiling by array, co-expression, data integration
Contact	Margus Lukk <margus.lukk@cruk.cam.ac.uk>
Citations	Multiple receptor tyrosine kinases promote the <i>in vitro</i> phenotype of metastatic human osteosarcoma cell lines. Rettew AN, Young ED, Lev DC, Kleinerman ES, Abdul-Karim FW, Getty PJ, Greenfield EM. <i>Oncogenesis</i> 1:e34 (2012), Europe PMC 23552467 Coevolution reveals a network of human proteins originating with multicellularity. Bezginov A, Clark GW, Charlebois RL, Dar VU, Tillier ER. <i>Mol Biol Evol</i> 30(2):332-346 (2013), Europe PMC 22977115

8. ブラウザのバックボタンで戻ります。'Title'には研究内容のタイトル、'Type'には実験の種類、'Organism'には生物種が書かれています

EMBL-EBI

Services | Research | Training | About us

ArrayExpress

Search results for oral cancer

Examples: E-MEXP-31, cancer, p53, Geuvadis | Advanced

Feedback | Login

Filter search results

Show more data from EMBL-EBI

Page	1 2 3 4 5 6 .. 10	Showing 1 - 25 of 231 experiments					Page size	25 50 100 250 500	
Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-62	Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines	transcription profiling by array	Homo sapiens, Mus musculus	5372	04/08/2010			41982	-
E-MTAB-2706	RNA-seq of 675 commonly used human cancer cell lines	RNA-seq of coding RNA	Homo sapiens	675	25/11/2014	-		12209	
E-TABM-185	Transcription profiling by array of integrated human experiments involving the hg133a platform to investigate a global map of human gene expression	transcription profiling by array	Homo sapiens	5896	20/11/2007			9474	-
E-MTAB-3732	A comprehensive human expression map	transcription profiling by array	Homo sapiens	27871	23/07/2015		-	9205	-
E-GEOID-7307	Transcription profiling by array of human samples with various different diseases	transcription profiling by array	Homo sapiens	677	20/06/2008			7249	-
E-TABM-302	Transcription profiling of human head and neck squamous cell carcinomas to identify metastasis predictors and descriptors, and differentiation classifiers	transcription profiling by array	Homo sapiens	81	23/08/2007			1777	-

9. 'Processed'にアイコンがあるものは解析済みデータがあることを、'Raw'にアイコンがあるものは生データがダウンロード可能であることを示しています

10. アイコンが下向き矢印のものは直接ダウンロードが始まり、クリップのものはダウンロードのできるサイトにジャンプします

EMBL-EBI

ArrayExpress

Home | Browse | Submit | Help | About ArrayExpress | Feedback | Login

Examples: E-MEXP-31, cancer, p53, Geuvadis | Advanced

Search

ArrayExpress > Browse > E-MTAB-62 > Files > Raw data

E-MTAB-62 - Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines

	E-MTAB-62.raw.1.zip	154.7 MB	31 January 2012, 14:35
	E-MTAB-62.raw.2.zip	148.6 MB	31 January 2012, 14:36
	E-MTAB-62.raw.3.zip	149.7 MB	31 January 2012, 14:38
	E-MTAB-62.raw.4.zip	149.6 MB	31 January 2012, 14:39
	E-MTAB-62.raw.5.zip	151.0 MB	31 January 2012, 14:40
	E-MTAB-62.raw.6.zip	150.5 MB	31 January 2012, 14:42
	E-MTAB-62.raw.7.zip	150.9 MB	31 January 2012, 14:43
	E-MTAB-62.raw.8.zip	153.3 MB	31 January 2012, 14:45
	E-MTAB-62.raw.9.zip	150.5 MB	31 January 2012, 14:46
	E-MTAB-62.raw.10.zip	150.3 MB	31 January 2012, 14:31
	E-MTAB-62.raw.11.zip	150.1 MB	31 January 2012, 14:33
	E-MTAB-62.raw.12.zip	181.5 MB	31 January 2012, 14:33
	E-MTAB-62.raw.13.zip	177.4 MB	31 January 2012, 14:34
	E-MTAB-62.raw.14.zip	154.3 MB	31 January 2012, 14:34
	E-MTAB-62.raw.15.zip	152.3 MB	31 January 2012, 14:34
	E-MTAB-62.raw.16.zip	152.4 MB	31 January 2012, 14:34
	E-MTAB-62.raw.17.zip	152.0 MB	31 January 2012, 14:34

11. ブラウザのバックボタンで戻ります。'Atlas'にアイコンがあるものはExpression Atlasにデータが収録されていることを示しています。

EMBL-EBI

Expression Atlas

Enter gene query... Examples: ASPM, Apoptosis, ENSMUSG00000021789, "zinc finger"

Services | Research | Training | About us | Feedback

Home | Release notes | FAQs | Download | Help | Licence | About

RNA-seq of 675 commonly used human cancer cell lines
Organism(s): Homo sapiens
Reference(s): 25485619 (Filter by genes in paper)
Raw Data Provider: Genentech

Gene query ? Filtered by ? Cell line ? Expression level cutoff ?

protein_coding × Disease: (any cell lines) 0.5
 Exact match
glioma
 Organism part: brain
 Specific ?
 Change filters >

Apply | Reset

Ensembl Genome Browser Open

Please select a cell line and a gene from the table

Showing 50 of 14487 genes found: (show by gene set) Download all results 4 144 ?

Display levels

Gene	G111	G112	G118	G120	G121	G122	G124	G130	G140	G141	G142	G22	G28T	G44	G59	G61	G84	G96	Hs 683	LN-18	LN-229	SF268	ST539
ZPLD1																							
PNMAS																							
KLRB1																							
GPR84																							
OGN																							
MYCN																							

12. 自分の研究テーマに近い、また興味のあるマイクロアレイデータが利用可能か検索してみましょう。

NCBI Gene Expression Omnibus (GEO)

- NCBIが提供・維持管理している遺伝子発現情報（主にマイクロアレイ）のデータベース
 - <https://www.ncbi.nlm.nih.gov/geo/>
- 自分の興味のある発現データセットや遺伝子プロファイルを検索することができるだけでなく、それらの生データを自由にダウンロードすることができます。

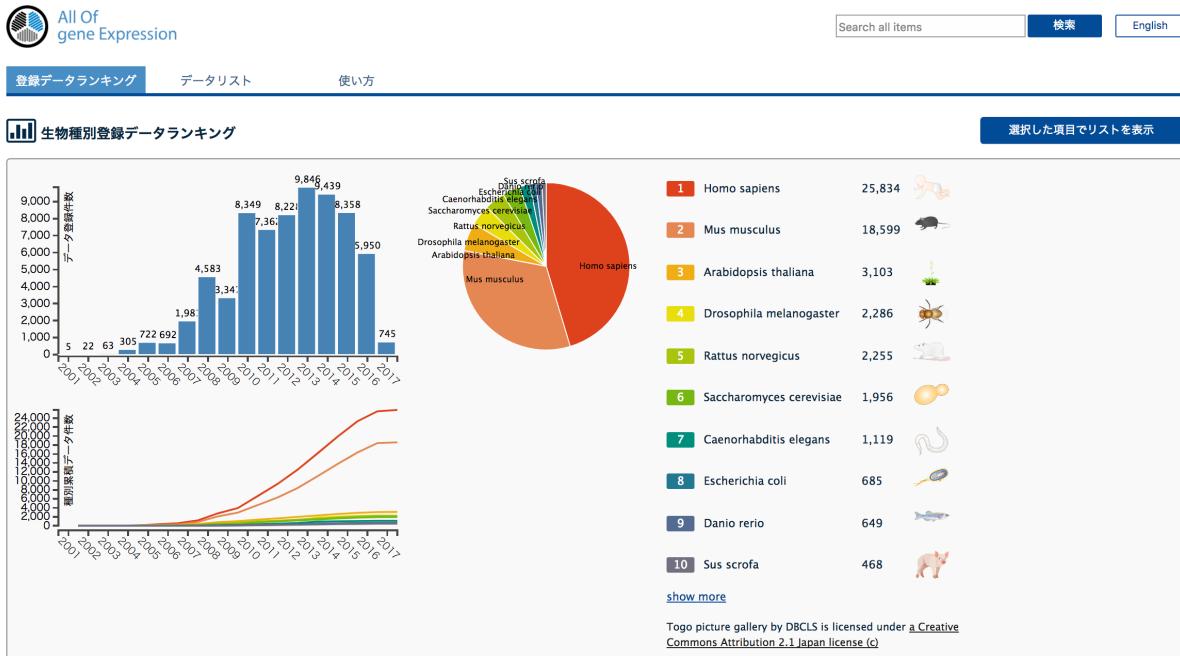
- GEOのエントリについて (GEO ID番号の最初の3文字の意味)
 - GPL: Platform – マイクロアレイチップの種類
 - GSM: Sample – 1枚のマイクロアレイチップから得られたサンプルデータ
 - GSE: Series – 1つの実験で得られたGSMのセット
 - GDS: DataSet – NCBIのスタッフが解析可能なデータを集めて再編成したGSMのセット
- 【参考】NCBI GEOの使い方1～マイクロアレイデータの検索・取得～2017

AOE(All of gene Expression)

- 遺伝子発現用のデータ目次
 - 年ごと、生物種ごとにヒストグラム表示
- <http://AOE.dbcls.jp/>
- マイクロアレイ(Affymetrix, Agilent, それ以外) + RNA-seq(Illumina, それ以外)
- キーワード検索も

【実習2】AOEを使って、低酸素関連遺伝子発現データを検索する

1. <http://AOE.dbcls.jp/> にアクセスします



2. 「生物種別登録データランキング」で生物種のところ(8. Escherichia coli)をクリックすると、その生物種だけのデータ登録数に絞りこめます



登録データランキング

データリスト

使い方

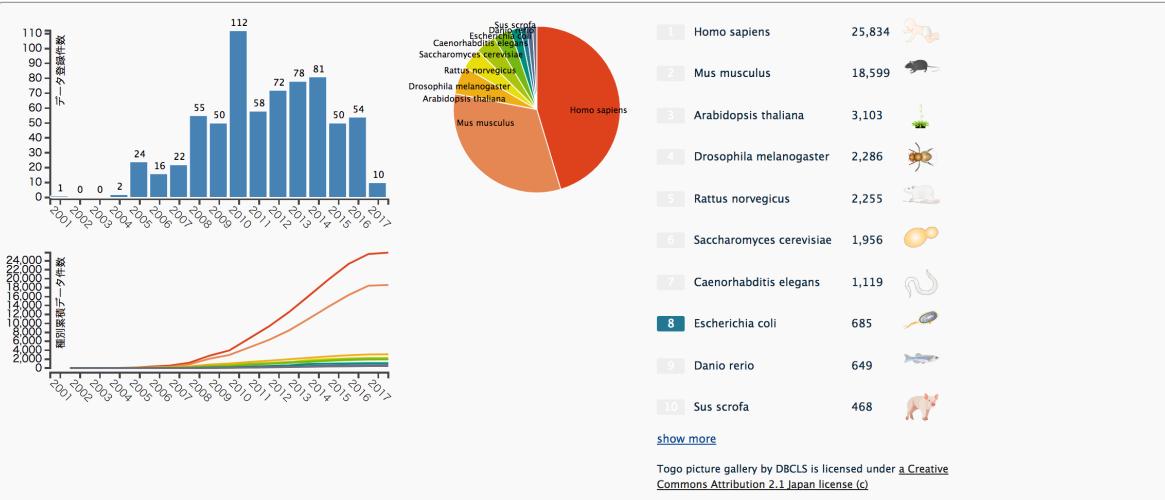
Search all items

検索

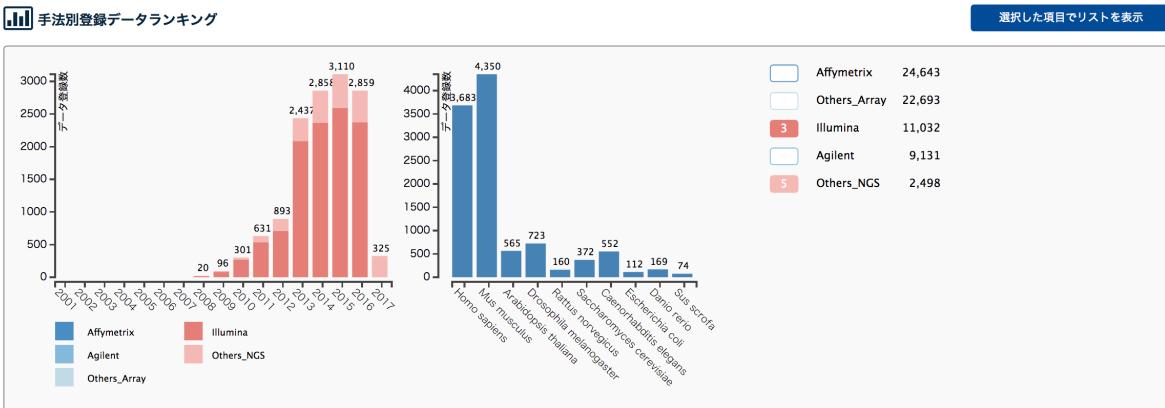
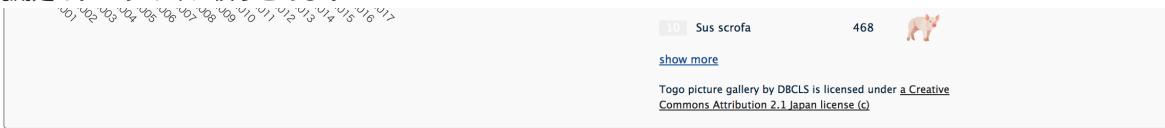
English

生物種別登録データランキング

選択した項目でリストを表示



3. 「手法別登録データランキング」で赤系の色(4. illuminaと5. Other_NGS)だけクリックして残すと、NGSによる遺伝子発現測定のデータのみに絞りこめます



4. 上部の検索窓でキーワード(例えば、 hypoxia)を入れて検索ボタンを押すと、そのキーワードを実験タイトルに含むエントリだけに絞りこめます



hypoxia

検索

English



【参考】AOEを使って遺伝子発現データベースの統計を見ながら検索する

個々の遺伝子の発現プロファイルを調べる

RefEx (Reference Expression dataset)

- ヒト、マウス、ラットの遺伝子発現情報リファレンスデータセット
 - <http://refex.dbcls.jp/>
- 4つの異なる実験手法 (EST、GeneChip、CAGE、RNA-seq) によって得られた正常組織、初代培養細胞、細胞株における遺伝子発現データを検索、閲覧可能
 - 最近新たに、FANTOM5 CAGEデータが追加(ヒト556種、マウス286種)

- 掲載しているデータやオリジナルデータなどの詳細については、[RefExについて](#)
- RefExで掲載されているデータはすべて再利用可能
 - 「RefEx analysis」として論文に引用していただいたケースも
 - [Aberrant IDH3α expression promotes malignant tumor growth by inducing HIF-1-mediated metabolic reprogramming and angiogenesis, Oncogene, \(22 December 2014\) | doi:10.1038/onc.2014.411](#)
- このツールでできること
 - 正常組織における遺伝子発現データを調べる
 - 測定手法による遺伝子発現量の差異を比較する
 - 組織特異的遺伝子をワンタップで検索可能
 - 遺伝子発現解析などで見出された不詳な遺伝子群の機能および関係性を調べる

【実習3】RefExを使って、組織特異的遺伝子を検索する

- 【復習用 統合TV】RefExの使い方

1. <http://refex.dbcls.jp/> を開きます。

2. 画面中央の「組織特異的に発現する遺伝子を見る」の臓器アイコンにカーソルを合わせると、更に詳細な部位のアイコンがるので、調べたい臓器（例として肝臓）をクリックします。

The screenshot shows the RefEx homepage. At the top, there are links for Reference, Expression, Dataset, and Language selection (English | 日本語). The main header features the RefEx logo with a circular icon containing three horizontal bars. To the right, there's a sidebar with text about comparing gene expression across organs, experimental methods, and 3D organ models, along with a link to 'もっと詳しく'.

In the center, there's a search bar with the placeholder 'キーワードで検索' and a yellow '検索' button. Below it, a section titled '組織特異的に発現する遺伝子を見る' (View tissue-specific genes) shows icons for various organs: Human, Mouse, Rat, Brain, Body, Skin, Uterus, Heart, Intestine, Liver, Kidney, Lung, Pancreas, and Bladder. A red arrow points to the 'Liver' icon, which is highlighted with a blue background. To the left of this section, there's a '遺伝子オントロジー Gene Ontology' section with '遺伝子ファミリー InterPro' below it. To the right, there are lists of biological processes and domain families, each with a '他のオントロジーを選ぶ' or '他のファミリーを選ぶ' button.

At the bottom, there are sections for '染色体' (Chromosome) and 'Advanced Search' (with a '染色体領域を選ぶ' button). A 'ページ上部に戻る' (Return to top of page) button is located at the very bottom right.

3. 検索結果一覧が表示されます。検索結果一覧では、「ソート項目の切り替え」や「絞り込み検索」、「リストへの追加」ができます。(手順11以降で解説します。)

4. 各遺伝子の青字の部分（例 *fibrinogen alpha chain*）をクリックすると詳細情報を閲覧できます。

The screenshot shows the search results page for the query "mannose-binding lectin (protein C) 2, soluble (opsonic defect)". The search interface includes filters for species (Human, Mouse, Rat), tissue (肝臓), and data sources (GeneChip). A green box highlights the "リストを見る" (View list) button, which is currently empty. Below the search bar, there are tabs for "3Dマップ" (3D Map), "相対発現量" (Relative Expression), and a set of icons representing various tissues.

mannose-binding lectin (protein C) 2, soluble (opsonic defect)

同義遺伝子名: Mannan-binding protein, MBL, MGC116833, MBL2, MBP-C, Mannose-binding protein C precursor, MGC116832, Mannose-binding lectin, COLEC1, HSMBPC, MBP1

Refseq ID: NM_000242
GeneID: 4153
Unigene ID: Hs.499674
Probe set ID: 207256_at

fibrinogen alpha chain

同義遺伝子名: MGC119423, FGA, fibrinogen alpha chain, Fibrinogen alpha chain precursor, Fib2, MGC119422, MGC119425

Refseq ID: NM_000508
GeneID: 2243
Unigene ID: Hs.351593
Probe set ID: 205649_s_at

aldo-keto reductase family 1, member D1 (delta 4-3-ketosteroid-5-beta-reductase)

同義遺伝子名: 3-oxo-5-beta-steroid 4-dehydrogenase, 3 β s-hydroxysteroid SRD5B1, AKR1D1, Delta(4)-3-ketosteroid 5-beta-reductase, Aldo-keto reductase family 1 member D1

Refseq ID: NM_005989
GeneID: 6718
Unigene ID: Hs.201667
Probe set ID: 207102_at

complement component 9

同義遺伝子名: C9, complement component 9, Complement component C9 precursor

Refseq ID: NM_001737
GeneID: 735
Unigene ID: -
Probe set ID: 206727_at

検索結果一覧

Each gene entry includes a 3D human body model showing expression levels across different tissues, and a bar chart for four data sources: EST, GeneChip, CAGE, and RNA-seq. The RNA-seq chart for the first gene shows a prominent red bar for the liver, while the others show mostly blue bars. The other genes show similar patterns.

5. 「ヒートマップ on Bodyparts3D」では、表示する部位の切り替え（全身・体幹部・頭部）ができます。「皮膚・骨格筋を表示」もしくは「アニメーション表示」にチェックを入れるとどのように表示されるでしょうか。

6. 「組織40分類別データ」では、バーの上にマウスオーバーすると測定部位と発現値が表示されます。



7. 「Download」をクリックすると、表示中の遺伝子の組織40分類別の発見データがタブ区切り形式でダウンロードできます。

8. 「Probe set ID」のリンク先をクリックすると、どう情報が参照できるでしょうか。



9. 遺伝子オントロジー(GO ID)をクリックすると、そのGO termを持つ他の遺伝子を一括で検索できます。

- 例として、[GO:0007596 blood coagulation](#)をクリックしてみましょう。

10. 右側のFANTOM5 CAGEのタブをクリックすると、FANTOM5 CAGEデータのビューアに切り替わります。

- ビューアは上部が拡大図で、下部が全体表示になっています。
- 検索窓にキーワードを入れるとサンプル名を検索できます。ヒットしたサンプルはオレンジ色で強調されます。
- 右側に、サンプル名と発現値、サンプル分類が表示されます。
- [RefEx用に整理したサンプル情報一覧](#)も閲覧可能です。

11. 検索結果一覧に戻ります。ソート項目を切り替えて、どのように結果が変わるでしょうか。

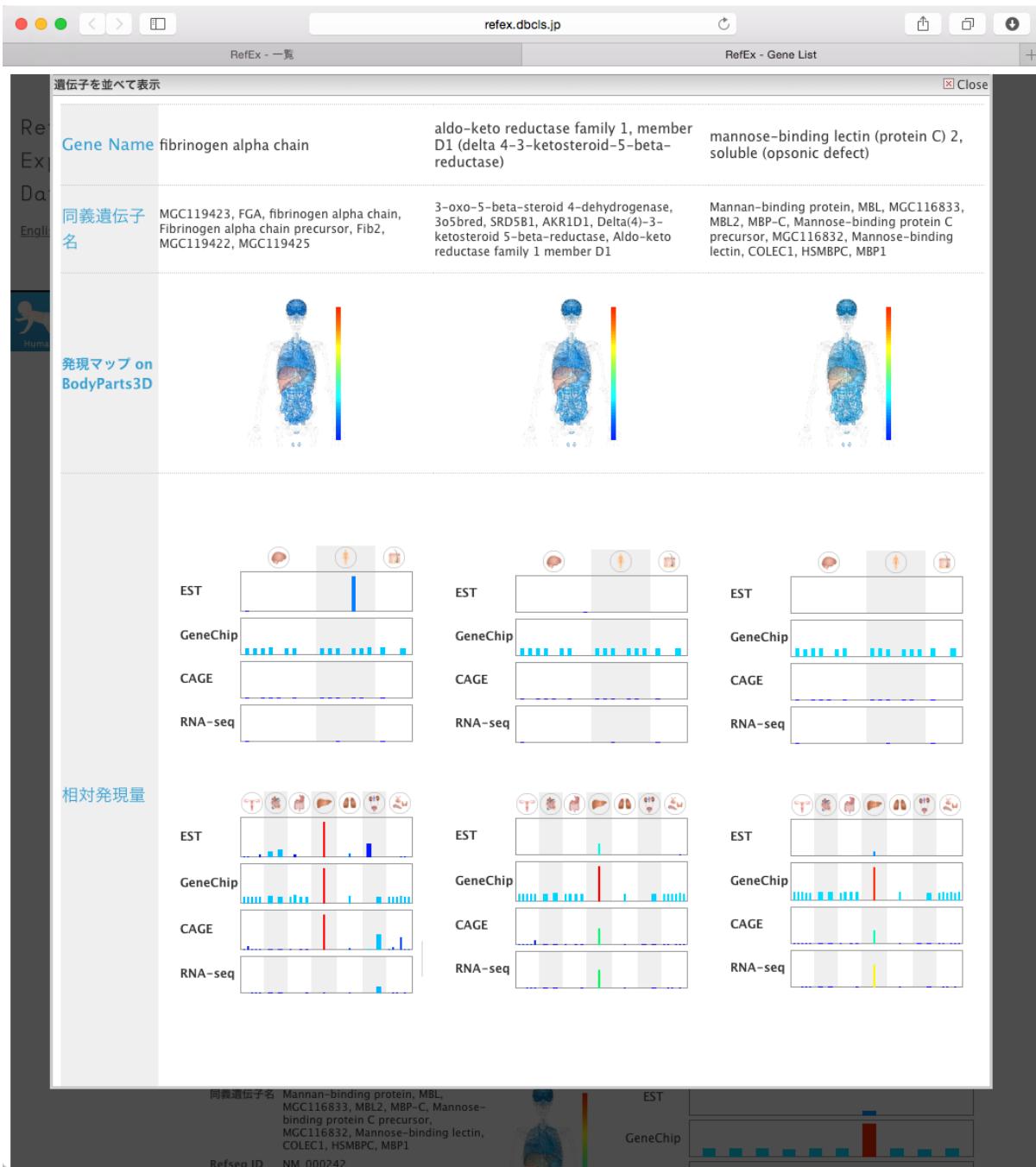
12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- 遺伝子名に「liver」を含むデータは何件あるでしょうか。
- 「遺伝子名」の下の「条件なし」をクリックして表示されるウインドウに「liver」と入力し、「Include」をクリックし、「この条件で絞り込み」を押します。
- 「遺伝子名」の項目で「Exclude」に「solute」を加えると、検索結果はどう変わるでしょうか。
- 「組織」の項目で、データ元をRNA-seqに変更したり、臓器の指定を追加すると検索結果はどう変わるでしょうか。
- 「必ず含むデータセット」の「ALL」にチェックを入れると、検索結果はどう変わるでしょうか。

13. 個々の遺伝子の詳細情報は、リストに追加することで並列に比較することができます。

- [肝臓特異的遺伝子の検索結果一覧](#)に移動して、3つの遺伝子を「リストに追加」してみましょう。
- 追加した件数は「リストを見る」の横に表示されます。
- 「リストを見る」をクリックするとリストに移動します。
- 『並べて表示する』にチェックを入れて、「遺伝子を並べて表示」をクリックします。

- 並列に比較することで見えてくる「違い」はなんでしょうか。





14. 自分の研究テーマに関連する、また興味のある遺伝子について検索してみましょう。

【応用1】 ChIP Atlasを使って注目の遺伝子がコードされたゲノム領域にあるChIP-seqピークを検索する

ChIP Atlasは公共のChIP-seqデータを再解析することで簡単に公共データを再利用できる仕組みです。

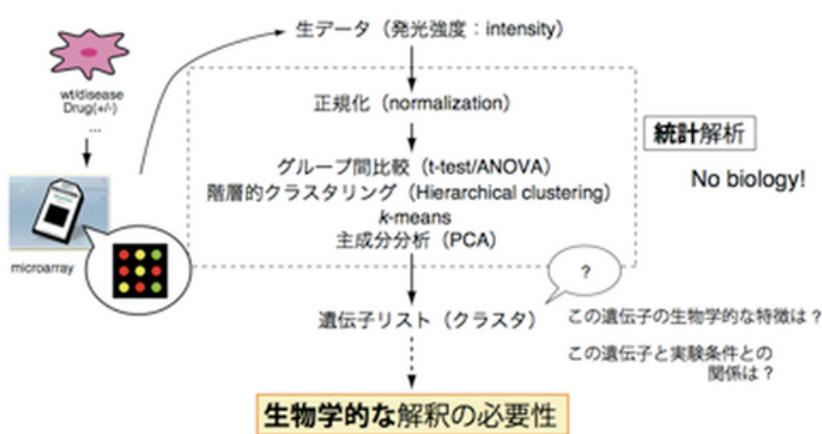
- Peak Browser
- Target genes

の機能に関して簡単に説明します。

数十～数千の遺伝子群の生物学的解釈

metascape

- マイクロアレイやRNA-seqデータの生物学的な解釈
- マイクロアレイやRNA-seq、すなわち遺伝子発現解析の一般的な目的は、実験条件によって得られた数十～数千の遺伝子群の発現が生物学的にどういう意味を持つかを考えることです。



- 今回は、その方法の一つとして、遺伝子リストにある遺伝子群に機能アノテーション(Gene Ontologyなど)を付与し、生物学的な解釈を行います。

遺伝子リストの準備

- サンプルデータとして、公共遺伝子発現データベースからメタ解析してえた遺伝子リストを用います。この遺伝子リスト `affy10.txt` は、ある刺激前後の2群間で発現増加した実験が10回以上あった遺伝子群のリストです。
→ [affy10.txt](#) (右クリックして「新しいタブで開く」もしくは「名前を付けてリンク先を保存」してください。)
- このデータは、どのような実験から得られたデータなのか、どのように解釈できるのかをDAVIDを使って考察してみましょう！

【実習4】metascapeを用いて、発現データの結果を生物学的に解釈する

- 【復習用 統合TV】Metascapeを使って、遺伝子リストの生物学的解釈をする

1. metascapeのウェブサイトにアクセスします。



The screenshot shows the "Step 1" interface of the Metascape web application. It includes the following elements:

- Input Options:** A checkbox for "Multiple Gene List" and a "Select File..." button for uploading files in .xls, .xlsx, .csv, or .txt formats.
- Alternative Input:** A text input field for "Or paste a gene list" with placeholder text: "Accept Gene ID/Symbol/RefSeq/Ensembl/UniProt/UCSC".
- File Formats:** A section titled "Upload File Format" with options for "Single List" (accepting .xls, .xlsx, .csv, .txt) and "Multiple List" (accepting .xls, .xlsx, .csv, .txt).
- Test Upload:** Buttons for "single list" and "3 gene lists".
- Test Identifiers:** Buttons for "Gene Symbol", "RefSeq", and "Entrez Gene ID".
- Next Steps:** Buttons for "Express Analysis" and "Custom Analysis" under "Step 2".

2. probe IDリストをコピペ or ファイルを指定し、`Submit` をクリックします。

Step 1

Or paste a gene list

```
1555881_s_at
1554290_at
1554153_a_at
1554015_a_at
1553589_a_at
```

Step 2

Step 3 Express Analysis Custom Analysis

Upload File Format
Single List: .xls/xlsx .csv .txt
Multiple List: .xls/xlsx .csv .txt

Test Upload
[single list](#)
[3 gene lists](#)

Test Identifiers
[Gene Symbol](#)
[RefSeq](#)
[Entrez Gene ID](#)

3. リストのIDの種類タイプが自動的に推定されます。... 今回は、`Affymetrix ID` と推定されました。

Step 1

Or paste a gene list

```
1555881_s_at
1554290_at
1554153_a_at
1554015_a_at
1553589_a_at
```

Step 2 Optional if you only consider human species in your study.
Input as species: Any Species ▾
Analysis as species: H. sapiens (629) ▾

Step 3 Express Analysis Custom Analysis

Upload File Format
Single List: .xls/xlsx .csv .txt
Multiple List: .xls/xlsx .csv .txt

Test Upload
[single list](#)
[3 gene lists](#)

Test Identifiers
[Gene Symbol](#)
[RefSeq](#)
[Entrez Gene ID](#)

4. `Express Analysis` をクリックするとデータ解析が実行されます。

5. しばらく待つと解析が終了し `Analysis Report Page` が現れます。それをクリックします。

Step 3 Express Analysis Custom Analysis

Please click the button below to see reports

Analysis Report Page

100

6. 解析を続けます。真ん中の「Functional Annotation Tool」をクリックします。

7. Enrichment解析の結果が表示されます。

[Gene List Report Excel Sheets](#)

[Gene List Report PPT file](#)

[All in One Zip File](#)

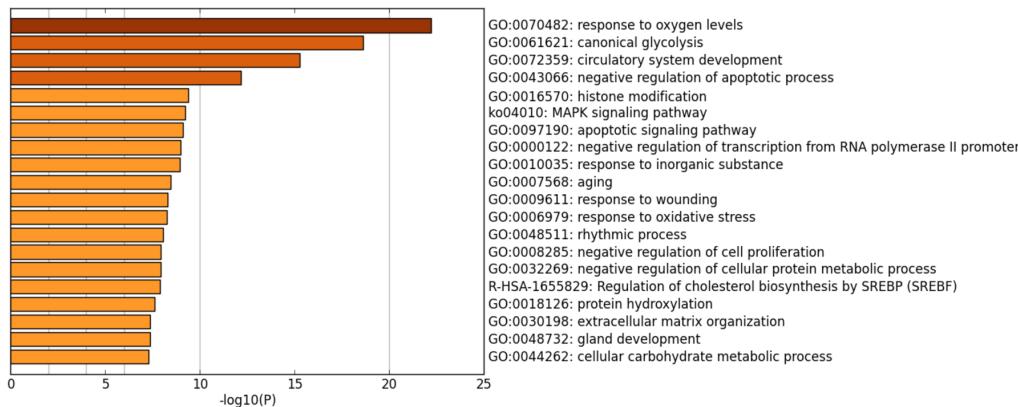
Warning: this report page can be accessed within the next 72 hours; all data will be deleted afterwards.

Metascape Gene List Analysis Report

[metascape.org¹](#)

Enrichment Summary

Figure 1. Heatmap of enriched terms across input gene lists, colored by p-values.



8. 下部のPDFアイコンをクリックすると、Enrichment解析のヒストグラムがPDF形式でダウンロードできます。

9. Gene List Report Excel Sheets や Gene List Report PPT file をクリックするとそれぞれExcelとPowerPointのファイル形式で結果が得られます。

10. このページの下部には、Enrichment解析の詳細な結果などが表示されていますので見てみましょう。

【応用2】複数のエンリッチメント解析ツールを用いて、発現データの結果を生物学的に解釈する

1. かつてはDAVIDというツールでこのエンリッチメント解析は行われていました。同様の結果が得られるか、試してみましょう。

- 【統合TV】 DAVIDを使ってマイクロアレイデータを解析する 2012
- 【統合TV】 DAVIDの使い方 実践編

2. また、特に医学・薬学分野に関連した情報を解析対象にできるのが特長のGeneSetDBというツールもあります。これも試してみましょう。

- 【統合TV】 GeneSetDBで遺伝子解析とエンリッチメント解析を行う 2012

3. 複数のツールで得られた結果を踏まえ、「ある実験」とはどのような実験であったか考察してみましょう。

- affy10.txt は「ある刺激前後の2群間で発現が1.5倍以上上昇した実験が10回以上あった遺伝子群のリスト」

4. 早く終わった方は、さらに20回以上発現上昇があった遺伝子群のリスト、affy20.txt（右クリックして「新しいタブで開く」もしくは「名前を付けてリンク先を保存」してください）で同じデータ解析をやってみましょう。10回以上のリストと比べてどういった違いがみられるでしょうか？

- このようにいろいろな閾値を試して、結果を見て考察し、最終的な閾値としております(例: Table 4. Gene Set Enrichment Analysis of up/down-regulated genes after UV irradiation. DOI: 10.1371/journal.pone.0116007.t004)