

# AJACS米子 遺伝子発現データベース・解析ツールの使い方

---

情報・システム研究機構  
ライフサイエンス統合データベースセンター  
小野 浩雅 (<https://sites.google.com/a/dbcls.rois.ac.jp/hono/>) hono@dbcls.rois.ac.jp  
2015年8月4日(火) AJACS米子@鳥取大学米子キャンパス

---

これは統合データベース講習会 AJACS米子「遺伝子発現データベース・解析ツールの使い方」の講習資料です。  
この内容の続編として、AJACS御茶ノ水(2015年5月)における応用・実践編 (<https://github.com/AJACS-training/AJACS53/blob/master/yoki/>) がありますので、こちらもあわせてご利用ください。  
講習会全体のプログラムはこちら (<http://events.biosciencedbc.jp/training/ajacs55/>) 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データベース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遺伝子発現データ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。  
また、自ら行なった大規模発現解析の(あるいは公共データベースから取得・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、コンピュータを使って以下の内容について説明します。

- 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
  - 個々の遺伝子の発現プロファイルを調べる
    - RefEx
      - 【実習1】 RefExを使って、組織特異的遺伝子を検索する
  - 数十～数千の遺伝子群の生物学的解釈
    - DAVID
      - 【実習2】 DAVIDを用いて、発現データの結果を生物学的に解釈する
      - 【実習3】 DAVIDを用いて、発現データの結果を生物学的に解釈する その2
- 

### 講習に際しての注意とお願い

- 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。おらかな気持ちで臨みましょう。
  - わからないことがあっ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!)。おいてけぼりは楽しくありません。
- 

### いざ講習、その前に

- 本講習内容をスムーズに理解するために押さえておくとい基礎知識
    - 「遺伝子のDB・ウェブツールの基礎」(2012年8月AJACS筑波2) (<http://motdb.dbcls.jp/?AJACS33%2Fmeso#e3b0f070>)
  - 非モデル生物のデータをモデル生物のデータに見立てるためのID対応表づくりについて
    - 「コマンドラインで遺伝子配列を解析する」(2012年7月) (<http://motdb.dbcls.jp/?AJACS32%2Fbono>)
  - 次世代シーケンス解析について
    - 「次世代シーケンサー (NGS) と関連するデータベース・ツール」(2014年7月AJACS信州) (<http://motdb.dbcls.jp/?AJACS48%2Fnakazato>)
    - 「galaxyによるNGS解析」(2015年1月AJACS三島) (<http://togotv.dbcls.jp/20150203.html>)
  - さらにもっと基礎から次世代シーケンス解析について知りたい方向け
    - 「パイオインフォマティクス人材育成カリキュラム(次世代シーケンサ)速習コース」の動画 ([https://www.youtube.com/playlist?list=PL0uaKHgcG00abmj1Nzs1SUhqKLjf\\_PFBb](https://www.youtube.com/playlist?list=PL0uaKHgcG00abmj1Nzs1SUhqKLjf_PFBb))
- 

## 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/ja/> (<http://togotv.dbcls.jp/ja/>)

TOGO TV  
CURATED

生命科学系DB・ツール使い倒し系チャンネル

はじめての方へ 番組ランキングほかの便利な方法よくある質問スタッフ番組リクエスト・お問い合わせ

使い倒し系チャンネル  
**統合TV**  
旧 統合TVはこちらから!

DBCLS  
Database Center  
for Life Science

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!

検索窓にキーワードを入れると、入力したごとに即座に候補の番組が絞り込まれます。  
先頭のタイトル行をクリックすると、昇順・降順で並び替えができます。  
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!!

10 エントリを表示  
検索: 0

目的別に検索!

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  
文献検索・辞書情報収集PC環境構築  
DBCLSサービス講演・講習動画

番組タイトル (YOUTUBEへのリンク)

番組の概要 (画像をクリックすると番組の再生ページへ移動します。)

Colilを使って論文の引用情報を検索する  
150205版

Colilは、生命科学分野の文献間の引用関係を検索・閲覧できるサービスです。  
ある論文について、他の論文が本文中でどのように引用し記述しているかを、効率的に調べることができます。  
検索例やデータの詳細については、Colil Data Portalにまとめられています。

Tutorial movies for TargetMine  
~ Keyword・Template Search ~

(<http://gyazo.com/457d7a15c537adc10d9b99596447508f>)

- YouTube版もあります <http://www.youtube.com/user/togotv/videos> (<http://www.youtube.com/user/togotv/videos>)

TOGO TV  
YouTube

togotv.dbcls.jp/

チャンネル登録 779

ホーム 動画 再生リスト チャンネル フリートーク 概要

統合TV プロモーションムービー

視聴回数 1,913 回 1 年前

本家はこちら <http://togotv.dbcls.jp/2013...>  
統合TVは、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です。数多くの解説動画の情報をうまく整理し、ユーザーが真に必要な情報を素早く容易にたどりつけるようにまとめています。生命科学研究に頻繁に使われるサイトを知り、その使い倒し術を知る手段として、日々の実験や研究の進展のためぜひご利用ください。

人気のアップロード

パワーポイントの図形描画機能でイラストをつくる方法  
3 年前・視聴回数 23,286 回  
本家はこちら <http://togotv.dbcls.jp/20110509.html#p01>  
スライドやポスター等のプレゼンテーション用資料に「イメージ通りのイラストがみつかる...

(<http://gyazo.com/4c4ffa07ba8a0ea1846a2e76be02284e>)

- 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版を見ればおよその内容がわかるようになっています。
  - 今回の講習に関連する内容の多くは、統合TV の発現制御解析 カテゴリー (<http://togotv.dbcls.jp/ja/contents/category/expression>) にあります。
  - 過去の講習会の内容はほとんどが統合TVに収録 (<http://togotv.dbcls.jp/ja/contents/category/dbcls#%E7%B5%B1%E5%90%88%E3%83%87%E3%83%BC%E3%82%BF%E3%83%99%E3%83%BC%E3%82%B9%E8%AC%9B%E7%BF%92%E5%8B%95%E7%94%BB>) 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 ([https://docs.google.com/forms/d/15pxJHnsV\\_Cu8B55Xy0F3jg5S9FXupkbBqONrZsyUE7k/viewform](https://docs.google.com/forms/d/15pxJHnsV_Cu8B55Xy0F3jg5S9FXupkbBqONrZsyUE7k/viewform)) へどうぞ!!
- 統合TVを作りたい方、募集中です。

---

## 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/> (<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
  - 最新新たに、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 (<http://refex.dbcls.jp/about.php?lang=ja>)
- 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 IDH3 $\alpha$  expression promotes malignant tumor growth by inducing HIF-1-mediated metabolic reprogramming and angiogenesis, Oncogene, (22 December 2014) | doi:10.1038/onc.2014.411 (<http://www.nature.com/ncj/journal/vaop/ncurrent/full/ncj2014411a.html>)
- 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チで検索可能
  - 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1】RefExを使って、組織特異的遺伝子を検索する

- 【復習用】RefExの使い方 (<http://youtu.be/GC2gLzvF6t0>)
- <http://refex.dbcls.jp/> (<http://refex.dbcls.jp/>) を開きます。
- 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出るので、調べたい臓器（例として肝臓）をクリックします。



キーワードで検索

検索

ex) troponin, ALB

組織特異的に発現する  
遺伝子を見る



臓器のアイコンをマウスオーバー

遺伝子オントロジー  
Gene Ontology

- cellular process
- biological regulation
- metabolic process

- multicellular organismal process
- response to stimulus
- developmental process

他のオントロジーを選ぶ

遺伝子ファミリー  
InterPro

- RNA recognition motif, RNP-1
- Pleckstrin homology
- Krüppel-associated box

- Protein kinase-like domain
- Zinc finger, C2H2-like
- GPCR, rhodopsin-like superfamily

他のファミリーを選ぶ

染色体

染色体領域を選ぶ

Advanced Search

Advanced Search

ページ上部に戻る

RefExについて

RefExの使い方

ダウンロード



RefEx はCreative Commons 表示 2.1 日本 Licenseの下でライセンスされています。  
原著作者はライフサイエンス統合データベースセンターです。



(<http://gyazo.com/fab7f0ba81afbce32061692c344bf03f>)

- 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。)
- 各遺伝子の青字の部分 (例 fibrinogen alpha chain ([http://refex.dbcls.jp/gene\\_info.php?lang=ja&db=human&geneID=2243&refseq=NIM\\_000508&unigene=Hs.351593&probe=205649\\_s\\_at](http://refex.dbcls.jp/gene_info.php?lang=ja&db=human&geneID=2243&refseq=NIM_000508&unigene=Hs.351593&probe=205649_s_at)))をクリックすると詳細情報を閲覧できます。

refex.dbcls.jp

結果一覧 1265 件中 1 - 10 件を表示

10 最初 < > 最後

ソート: Tissue Specificity, high

リストをクリア

リストを見る 0

ダウンロード

遺伝子名 3Dマップ 相対発現量

**検索条件**

遺伝子名  
条件なし

組織  
• 肝臓

オントロジー  
条件なし

ファミリー  
条件なし

その他のキー  
条件なし

必ず含むデータセット

☐ ALL

☐ EST

☒ GeneChip

☐ CAGE

☐ RNA-seq

この条件で絞り込み

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

リストに追加する リストから削除する

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

EST

GeneChip

CAGE

RNA-seq

**fibrinogen alpha chain**

リストに追加する リストから削除する

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

EST

GeneChip

CAGE

RNA-seq

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

リストに追加する リストから削除する

同義遺伝子名 3-oxo-5-beta-steroid 4-dehydrogenase, 3o5bred, 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

EST

GeneChip

CAGE

RNA-seq

**complement component 9**

リストに追加する リストから削除する

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

EST

GeneChip

CAGE

RNA-seq

nodata

検索結果一覧

(http://gyazo.com/2a250c033aac172fae84b89033e1b225)

- 「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。
- 「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。



## fibrinogen alpha chain

詳細情報を見る

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

### 発現データ

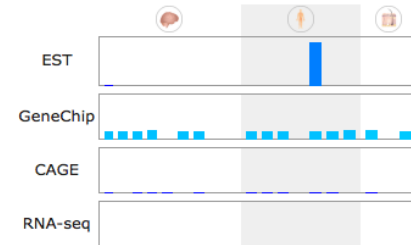
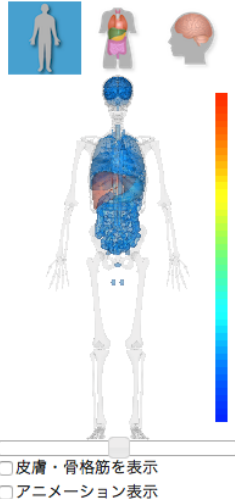
正常組織・臓器 40分類

正常組織 (発生期含む)・初代培養細胞 (FANTOM5 CAGE)

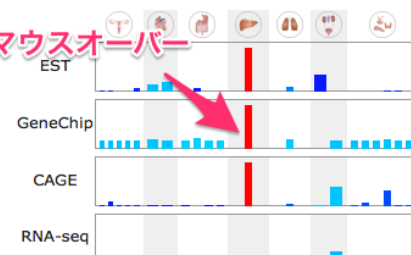
発現マップ on BodyParts3D

組織40分類別データ [Download](#)

相対発現量を、人体 3D 画像にマップしたものです。  
Genechip 組織40分類 の発現パターンを使用しています。



バーをマウスオーバー



### 詳細情報

IDs

Refseq ID [NM\\_000508](#)  
Gene ID [2243](#)  
Unigene ID [Hs.351593](#)  
Probe set ID [205649\\_s\\_at](#) [HG-U133\_Plus\_2]  
Ensembl ID [ENSG00000171560](#)

オーソログ対応遺伝子

マウス[1] [NM\\_001111048](#), [NM\\_010196](#)  
ラット[3] [NM\\_001008724](#), [NM\\_052797](#)

染色体

4q31.3 [155,504,278](#) - [155,511,918](#)  
(<http://gyazo.com/b60518629c6dd0fe8163776cc7824a3c>)

遺伝子オントロジー (GO ID)

#### Biological Process

[GO:0007596](#) [blood coagulation](#)  
[GO:0030168](#) [platelet activation](#)  
[GO:0002576](#) [platelet degranulation](#)  
[GO:0051258](#) [protein polymerization](#)  
[GO:0051592](#) [response to calcium ion](#)  
[GO:0007165](#) [signal transduction](#)

#### Cellular Component

[GO:0005938](#) [cell cortex](#)  
[GO:0009986](#) [cell surface](#)  
[GO:0009897](#) [external side of plasma membrane](#)  
[GO:0005576](#) [extracellular region](#)  
[GO:0005615](#) [extracellular space](#)  
[GO:0005577](#) [fibrinogen complex](#)

詳細情報ページ

- 「Download」をクリックすると、表示中の遺伝子の組織40分類別の発現データがタブ区切り形式でダウンロードできます。
- 「Probe set ID」のリンク先をクリックすると、どういう情報が参照できるでしょうか。



## fibrinogen alpha chain

詳細情報を見る

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

### 発現データ



(<http://gyazo.com/78a17e8253cb9ed64f6becf96b5a1e03>)

- 遺伝子オントロジー(Gene Ontology (<http://array.cell-innovator.com/?p=1085>) :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 (<http://refex.dbcls.jp/genelist.php?lang=ja&db=human&idkind=1&ids=GO:0007596>) blood coagulation をクリックしてみましょう。
- 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力す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 (<http://bit.ly/fantom5cagehuman>) も閲覧可能です。
- 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。
- 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「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。
- 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 ([http://refex.dbcls.jp/genelist.php?lang=ja&db=human&roku\\_valid=1&rk\[31\]=31&order\\_key=score](http://refex.dbcls.jp/genelist.php?lang=ja&db=human&roku_valid=1&rk[31]=31&order_key=score)) 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」はなんなのでしょうか。



<http://gyazo.com/f832aab525efcbd99854b8c920be0fcf>



refex.dbcls.jp

RefEx - 一覧 RefEx - Gene List

遺伝子を並べて表示

Refseq ID	NM_000508	NM_005989	NM_000242
Gene ID	2243	6718	4153
Unigene ID	Hs.351593	Hs.201667	Hs.499674
probe set ID	205649_s_at	207102_at	207256_at
Ensembl ID	ENSG00000171560	ENSG00000122787	ENSG00000165471
染色体	4.q31.3 [155504278 - 155511918]	7.q33 [137687070 - 137802732]	10.q21.1 [54525140 - 54531460] LRG_154, [5001 - 11321]
遺伝子ファミリー (Interpro ID)	-	Aldo/keto reductase Aldo/keto reductase	- C-type lectin C-type lectin fold C-type lectin-like
遺伝子オントロジー Biological Process	blood coagulation platelet activation platelet degranulation protein polymerization response to calcium signal transduction	androgen metabolic process bile acid biosynthesis bile acid catabolic process bile acid metabolic process C21-steroid hormone biosynthesis cholesterol catabolic process digestion oxidation reduction	acute-phase response complement activation complement activation complement activation defense response to defense response to innate immune response killing by host of negative regulation opsonization positive regulation response to oxidative stress

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

Refseq ID NM\_000242

EST

GeneChip

(http://gyazo.com/0c604ddee80bf4adf14ce52876a5744)

- 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。

## BioGPS

- ヒト、マウス、ラットのさまざまな組織や細胞(株)における遺伝子発現プロファイルのデータベース
- BioGPS (<http://biogps.org/>) は Affymetrix 社製のマイクロアレイである GeneChip を用いたさまざまな組織や細胞(株)遺伝子発現プロファイルのデータベース。
- 検索した遺伝子に対して、種々の外部データベースを横断検索することができるだけでなく、それらの設定を保存したり、表示方法を自由にカスタマイズすることができる「Gene annotation portal」。
- 外部データベースには、Wikipedia(Gene Wiki)、著名な試薬会社の検索窓へのリンク集、pathway、Nature系DB、モデル生物DB、文献DBなど多種多様
- マウスのエクソアレイのデータから遺伝子のスプライシングバリエーション(Splicing variant)の発現状況も調べることが可能。最近ではCircadian関係のデータも。
- さらに最近のアップデートで、NCBI Gene Expression Omnibus (GEO)中から選抜されたデータセットに切り替えて発現状況を調べることが可能に。

### 【実習(skip)】 BioGPSを使ってある遺伝子の発現プロファイルを調べる

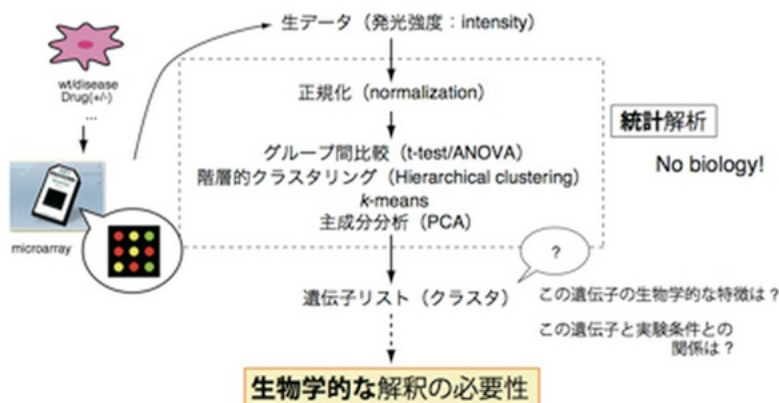
- 【復習用】 遺伝子発現プロファイルデータベースBioGPSを使い倒す2012 ([http://www.youtube.com/watch?v=X\\_exdocRIVQ](http://www.youtube.com/watch?v=X_exdocRIVQ))
- 【以前の講習会動画】 遺伝子発現データベースの活用法 (<http://togotv.dbcls.jp/20100829.html#p01>)

- <http://biogps.org/> (<http://biogps.org/>) を開きます。2. 骨格筋の分化決定遺伝子であるMyogenic differentiation 1(Myod)の発現プロファイルを調べてみましょう。中央の検索窓に「myod」と入力し、「search」を押します。
- 表示された検索結果の中から「ID 4654」をクリックします。
- 最初はヒトのマイクロアレイデータが表示されます。
- 画面左側の"Current Gene List"は右上の<<アイコンをクリックすると非表示にできます。非表示にすることで画面を広く使うことができます。
- ページ内のウィンドウは通常のウィンドウと同じようにドラッグによる移動やサイズの変更などを行うことができます。 歯車マークのメニューから"Open in browser" を選択すると、新しいタブで表示できます。
- "Search" と書かれた窓に単語(組織名など)を入力すると、その単語の含まれた部分が赤くハイライト表示されます。今回は "Muscle" と入力してみます。
- "Zoom" のバーを用いることで、グラフの表示範囲を調整することが出来ます。
- 発現量を示すバーをクリックすると発現強度の値が表示されます。
- マイクロアレイデータ右上の"Species: Hs"をクリックするとマウスやラットを選択できるので、"M. musculus (Mouse)"をクリックしてマウスのデータを表示できます。
- MyoDはどの組織、細胞で強く発現しているでしょうか？
- 場合によっては"Probeset"のプルダウンメニューから複数の項目を選択できる場合があります。これはどのようなケースが考えられるでしょうか？
- "Static Image" をクリックすると、ズームや検索機能などのついていない、画像だけのグラフで表示されます。低スペックなマシンでは、こちらの方が軽快に動作するでしょう。
- "Correlation"タブをクリックして検索すると、発現パターンが似ている他の遺伝子を検索できますが、どのような遺伝子が出てくるでしょうか？
- "Downloads" をクリックすると現在表示している遺伝子の発現データをCSV形式でダウンロードできます。
- "Dataset"の右にある"change"をクリックすると、デフォルトで用意されているデータセットやNCBI GEO中のデータセットを検索でき、それらのデータに表示を切り替えることができます。"Species: Hs"に切り替えてから、"change"をクリックしたあと、"Default Datasets"から"Barcode on normal tissues (262 samples)"を選択します。どのようにデータが変わったでしょうか。
- さらに"Search"からキーワード検索で、GEOのデータを検索してみましょう。"C2C12"と検索するとどのようなデータが選択できるでしょうか。
- 右上の「default layout」をクリックすると、検索した遺伝子に関して種々の外部データベースを横断検索できますが、どのようなデータが閲覧できるのか調べてみましょう。
- 左上の「Search」タグをクリックして検索画面にもどり、自分の興味ある遺伝子について同様に検索してみましょう。すぐに自分の興味ある遺伝子が浮かばない場合は、著名なiPS細胞 (<http://ja.wikipedia.org/wiki/%E4%BA%BA%E5%B7%A5%E5%A4%9A%E8%83%BD%E6%80%A7%E5%B9%B9%E7%B4%B0%E8%83%9E>) を作るために必要な4因子 (Oct3/4・Sox2・Klf4・c-Myc) がどの組織で発現しているか、またデータを切り替えて検索してみましょう。
- 【余談】 BioGPSのiPhoneアプリ (<http://biogps.org/iphone/>) が無料で公開されていますので、「あの遺伝子はどの組織で発現しているのかな？」とふと調べたいときにお手持ちのiPhoneで遺伝子発現を調べられます。

## 数十～数千の遺伝子群の生物学的解釈

### DAVID: The Database for Annotation, Visualization and Integrated Discovery

- マイクロアレイデータの生物学的な解釈
- マイクロアレイ実験の一般的な目的は、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。



(<http://gyazo.com/52cb4c40b1313a52f8ded6923bdd8ef0>)

- 今回は、その方法の一つとして、マイクロアレイの結果にGene Ontology (<http://array.cell-innovator.com/?p=1085>) の用語を付与することで、生物学的な解釈を行います。

#### マイクロアレイデータの準備

- サンプルデータとして、NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>) から取得した公共の遺伝子発現データを用います。このデータは、ある実験の前後の2群間で有意に発現減少した遺伝子群のリストです。

→ マル秘遺伝子リスト ([https://raw.githubusercontent.com/AJACS-training/AJACS53/master/hono/secret\\_list.txt](https://raw.githubusercontent.com/AJACS-training/AJACS53/master/hono/secret_list.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2】 DAVID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】 DAVIDを使ってマイクロアレイデータを解析する 2012 ([http://youtu.be/Wife7Z\\_Mvxg](http://youtu.be/Wife7Z_Mvxg))
- 【復習用】 DAVIDの使い方 実践編 (<http://youtu.be/4ft1ma9lRc>)
- 上部メニューの「Start Analysis」をクリックします。

**DAVID Bioinformatics Resources 6.7**  
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

**Shortcut to DAVID Tools**

**Functional Annotation**  
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)

**Gene Functional Classification**  
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)

Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.7

2003 - 2014

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an [update to the sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for

[What's Important in DAVID?](#)

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)

(<http://gyazo.com/f976f39aeb060a96a790f0e5b281aabe>)

- 画面左側バーで、probe IDリストをコピー or ファイルを指定します。
- リストのIDの種類タイプを選択します。... 今回は、「AFFY\_ID」と「Gene List」
- Submit List をクリックするとリストが読み込まれます。

**Analysis Wizard**

[Tell us how you like the tool](#)  
[Contact us for questions](#)

← Step 1. Submit your gene list through left panel.

An example:

Copy/paste IDs to "**box A**" -> Select Identifier as "**Affy\_ID**" -> List Type as "**Gene List**" -> Click "**Submit**" button

1007\_s\_at  
1053\_at  
117\_at  
121\_at  
1255\_g\_at  
1294\_at  
1316\_at  
1320\_at  
1405\_i\_at  
1431\_at  
1438\_at  
1487\_at  
1494\_f\_at

**Upload List Background**

**Upload Gene List**

[Demolist 1](#) [Demolist 2](#)  
[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

Or

B: Choose From a File

ファイルを選択 選択されていません

Multi-List File ?

Step 2: Select Identifier

AFFYMETRIX\_3PRIME\_IVT\_ID

Step 3: List Type

Gene List ☐  
Background ☐

Step 4: Submit List

Submit List

(<http://gyazo.com/e8275cf9dbb203b3d8577307b462c783>)

- アップロードしたリストは、左側バーの「List Manager」で「Uploaded List\_1」として保存されています。削除やrenameもできます。



Upload

List

Background

Gene List Manager

Select to limit annotations by one or more species

[Help](#)

- Use All Species -

Arabidopsis thaliana(2928)

Unknown(1)

Select Species

List Manager

Help

List\_1

Select List to:

Use

Rename

Remove

Combine

Show Gene List

[View Unmapped Ids](#)

Analysis Wizard

[Tell us how you like the tool](#)
[Contact us for questions](#)

Step 1. Successfully submitted gene list

Current Gene List: List\_1

Current Background: Arabidopsis thaliana

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

Functional Annotation Tool

Functional Annotation Clustering

Functional Annotation Chart

Functional Annotation Table

(<http://gyazo.com/e8270d82a68decba0249daa49914fba9>)

- 解析を続けます。真ん中の「Functional Annotation Tool」をクリックします。
- 「Gene Ontology」をクリックすると、Gene Ontologyを用いた解析の細かいメニューが表示されます。

Upload

List

Background

Gene List Manager

Select to limit annotations by one or more species

[Help](#)

- Use All Species -

Arabidopsis thaliana(2928)

Unknown(1)

Select Species

List Manager

Help

List\_1

Select List to:

Use

Rename

Remove

Combine

Show Gene List

[View Unmapped Ids](#)

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: List\_1

Current Background: Arabidopsis thaliana

2916 DAVID IDs

Check Defaults ☒

Clear All

Functional Categories (3 selected)

Gene\_Ontology (3 selected)

<input type="checkbox"/>	GOTERM_BP_1	66.6%	1943	Chart	
<input type="checkbox"/>	GOTERM_BP_2	64.8%	1891	Chart	
<input type="checkbox"/>	GOTERM_BP_3	61.9%	1806	Chart	
<input type="checkbox"/>	GOTERM_BP_4	57.2%	1668	Chart	
<input type="checkbox"/>	GOTERM_BP_5	51.5%	1501	Chart	
<input type="checkbox"/>	GOTERM_BP_ALL	66.7%	1944	Chart	
<input checked="" type="checkbox"/>	GOTERM_BP_FAT ?	61.6%	1795	Chart	
<input type="checkbox"/>	GOTERM_CC_1	77.7%	2267	Chart	
<input type="checkbox"/>	GOTERM_CC_2	77.6%	2264	Chart	
<input type="checkbox"/>	GOTERM_CC_3	77.2%	2251	Chart	
<input type="checkbox"/>	GOTERM_CC_4	67.9%	1979	Chart	
<input type="checkbox"/>	GOTERM_CC_5	64.0%	1867	Chart	
<input type="checkbox"/>	GOTERM_CC_ALL	77.7%	2267	Chart	
<input checked="" type="checkbox"/>	GOTERM_CC_FAT ?	68.5%	1997	Chart	
<input type="checkbox"/>	GOTERM_MF_1	70.8%	2065	Chart	
<input type="checkbox"/>	GOTERM_MF_2	68.1%	1985	Chart	
<input type="checkbox"/>	GOTERM_MF_3	61.5%	1792	Chart	
<input type="checkbox"/>	GOTERM_MF_4	56.7%	1653	Chart	
<input type="checkbox"/>	GOTERM_MF_5	45.5%	1326	Chart	
<input type="checkbox"/>	GOTERM_MF_ALL	70.9%	2066	Chart	
<input checked="" type="checkbox"/>	GOTERM_MF_FAT ?	62.7%	1828	Chart	

General Annotations (0 selected)

Literature (0 selected)

Main\_Accessions (0 selected)

Pathways (1 selected)

Protein\_Domains (3 selected)

Protein\_Interactions (0 selected)

Tissue\_Expression (0 selected)

(<http://gyazo.com/38905ceb16d6b702059667e4fb404531>)

- 今回は、GOTERM\_BP\_FAT (BP = Biological Process)に注目します。その右の「Chart」をクリックすると結果がポップアップされます。

12 of 14

## Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List\_1












Current Background: Arabidopsis thaliana

2916 DAVID IDs

 Options

301 chart records

 [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">photosynthesis</a>	RT		102	3.5	1.5E-45	2.7E-42
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">photosynthesis, light reaction</a>	RT		56	1.9	6.2E-29	5.7E-26
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">pigment metabolic process</a>	RT		53	1.8	5.7E-21	3.5E-18
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">pigment biosynthetic process</a>	RT		47	1.6	2.6E-19	1.2E-16
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">carboxylic acid biosynthetic process</a>	RT		117	4.0	9.0E-18	3.3E-15
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">organic acid biosynthetic process</a>	RT		117	4.0	9.0E-18	3.3E-15
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">generation of precursor metabolites and energy</a>	RT		111	3.8	2.8E-15	8.4E-13
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">response to abiotic stimulus</a>	RT		241	8.3	1.1E-13	2.8E-11
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">plastid organization</a>	RT		41	1.4	2.9E-13	6.7E-11
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">nitrogen compound biosynthetic process</a>	RT		124	4.3	5.1E-13	1.0E-10
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">heterocycle biosynthetic process</a>	RT		50	1.7	9.6E-13	1.8E-10

(<http://gyazo.com/78301700c3d952957dd599bbb83c785f>)

- タイトル行をクリックするとソートできます。
- さらに、GOTERM\_CC\_FAT や GOTERM\_MF\_FAT を見て、上位にリストされたGOTermにどのような共通点・相違点があるでしょうか。
  - CC = Cellular Component

## Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List\_1












Current Background: Arabidopsis thaliana

2916 DAVID IDs

 Options

68 chart records

 [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">plastid part</a>	RT		515	17.7	9.0E-190	2.9E-187
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">chloroplast part</a>	RT		504	17.3	1.1E-187	1.8E-185
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">plastid</a>	RT		987	33.8	1.4E-165	1.5E-163
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">chloroplast</a>	RT		973	33.4	1.3E-164	1.1E-162
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">plastid thylakoid</a>	RT		230	7.9	7.0E-110	4.6E-108
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">chloroplast thylakoid</a>	RT		230	7.9	7.0E-110	4.6E-108
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">organelle subcompartment</a>	RT		230	7.9	4.5E-109	2.4E-107
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">thylakoid</a>	RT		278	9.5	8.0E-109	3.7E-107
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">thylakoid part</a>	RT		228	7.8	6.0E-105	2.5E-103
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">chloroplast thylakoid membrane</a>	RT		200	6.9	4.4E-100	1.6E-98
<input type="checkbox"/>	GOTERM_CC_FAT	<a href="#">plastid thylakoid membrane</a>	RT		200	6.9	4.4E-100	1.6E-98

(<http://gyazo.com/117720204dfb06a3f3605f4aedec2dba>)

- MF = Molecular Function

## Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List\_1

Current Background: Arabidopsis thaliana

2916 DAVID IDs











 Options

Rerun Using Options

Create Sublist

151 chart records

 [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">cofactor binding</a>	<a href="#">RT</a>		109	3.7	3.8E-8	4.4E-5
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">chlorophyll binding</a>	<a href="#">RT</a>		14	0.5	1.5E-5	8.4E-3
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">vitamin B6 binding</a>	<a href="#">RT</a>		32	1.1	3.4E-5	1.3E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">pyridoxal phosphate binding</a>	<a href="#">RT</a>		32	1.1	3.4E-5	1.3E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">poly(U) RNA binding</a>	<a href="#">RT</a>		11	0.4	4.6E-5	1.3E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">poly-pyrimidine tract binding</a>	<a href="#">RT</a>		11	0.4	4.6E-5	1.3E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">ATP-dependent peptidase activity</a>	<a href="#">RT</a>		12	0.4	8.3E-5	1.9E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">rRNA binding</a>	<a href="#">RT</a>		23	0.8	8.5E-5	1.6E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">coenzyme binding</a>	<a href="#">RT</a>		72	2.5	9.7E-5	1.6E-2
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">vitamin binding</a>	<a href="#">RT</a>		39	1.3	1.2E-4	1.7E-2

(<http://gyazo.com/6feb8e34beab45769e2d3e66c3c5d570>)

- Pathways > KEGG\_PATHWAY や Tissue Expression > UP\_TISSUE などとも見てみましょう。
- DAVID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。
  - マル秘遺伝子リストは「ある実験の前後の2群間で有意に発現減少した遺伝子群のリスト」
  - 生物種はArabidopsis thaliana (シロイヌナズナ)

答え合わせ (<https://github.com/AJACS-training/AJACS55/blob/master/hono/answer.md>)

### 【実習2】DAVIDを用いて、発現データの結果を生物学的に解釈する その2

- DAVID の使い方に慣れてきたところで、実戦的な生物学的解釈に挑戦してみましょう。
- 今回は「正解」はありません。情報分析力と想像力が問われます。
- 例題は、GSE28619 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28619>) をつかいます。
  - 健常者 vs アルコール性肝炎患者 の2群比較です。
  - 多重比較法 (Benjamini & Hochberg) を指定して、有意水準1%未満かつ2倍以上発現差のあった遺伝子群のリストをあらかじめ用意しました。
    - 「健常者>AH患者\_遺伝子リスト」GEO2R\_Ctrl.txt ([https://raw.githubusercontent.com/AJACS-training/AJACS55/master/hono/GEO2R\\_Ctrl.txt](https://raw.githubusercontent.com/AJACS-training/AJACS55/master/hono/GEO2R_Ctrl.txt))
    - 「AH患者>健常者\_遺伝子リスト」GEO2R\_AH.txt ([https://raw.githubusercontent.com/AJACS-training/AJACS55/master/hono/GEO2R\\_AH.txt](https://raw.githubusercontent.com/AJACS-training/AJACS55/master/hono/GEO2R_AH.txt))
    - (この遺伝子リストの作り方は、AJACS御茶ノ水の回 (<https://github.com/AJACS-training/AJACS53/blob/master/yoki/>) で解説しています。)
  - このデータを使った論文 (Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis. (<http://www.ncbi.nlm.nih.gov/pubmed/22637703>)) も出ていますが、答え合わせをするもよし、著者らが見逃している着眼点はないか探すもよし、です。