

# AJACS下総 実験データの生物学的解釈をするための遺伝子発現DB・ウェブツールの使い方

大学共同利用機関法人 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

小野 浩雅

[hono@dbcls.rois.ac.jp](mailto:hono@dbcls.rois.ac.jp)

2017年12月19日(火) AJACS下総 @ 千葉大学みのはな同窓会館多目的ホール

これは統合データベース講習会 AJACS下総「実験データの生物学的解釈をするための遺伝子発現DB・ウェブツールの使い方」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](#)がありますので、こちらもあわせてご活用ください。

講習会全体のプログラムは[こちら](#)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】これまで学んだことを踏まえて、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。おおらかな気持ちで臨みましょう。

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。
- 実験的な試みとしてWeb上で匿名で質問・コメントできるフォームを用意してみました。
  - [goo.gl/slides/bza77j](http://goo.gl/slides/bza77j)

受講前アンケートにご協力いただき、ありがとうございます (回答数 35)

統合TVを知っていますか?	人数	割合
知らない	16 名	46 %
聞いたことがある	3 名	9 %
知っている	4 名	11 %
使ったことがある	4 名	11 %
使っている	3 名	9 %
回答なし	5 名	14 %

自分で実験して得た、数十～数千の遺伝子からなる 「遺伝子リスト」(例: 発現差のあった遺伝子など)を持っていますか?	人数	割合
これから実験をする・したい	10 名	29 %
公共データを活用する・したい	10 名	29 %
既に持っている	3 名	9 %
大規模発現解析の予定はない	7 名	20 %
回答なし	5 名	14 %

## 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/>

「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。

### 目的別に検索

- 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
- 自由に使える画像を探す

### 関連するタグから検索

- [ゲノム \(251\)](#) [遺伝子 \(397\)](#)
- [タンパク質 \(197\)](#)
- [配列解析 \(227\)](#)
- [発現解析 \(300\)](#) [NGS \(235\)](#)
- [文献検索 \(234\)](#)
- [情報収集 \(121\)](#)
- [環境設定 \(122\)](#)
- [DBCLS \(181\)](#) [English \(185\)](#)
- [ウェブツール \(200\)](#)
- [ソフトウェア \(71\)](#)
- [データベース \(319\)](#)
- [講演 \(563\)](#) [実習 \(335\)](#)
- [NCBI \(61\)](#) [GEO \(18\)](#)
- [UCSC \(22\)](#) [EBI \(12\)](#)
- [Ensembl \(23\)](#) [KEGG \(17\)](#)
- [PDB \(24\)](#) [DDBJ \(42\)](#)
- [PDBj \(11\)](#) [NBDC \(53\)](#)
- [転写因子 \(16\)](#) [化合物 \(19\)](#)
- [メタボローム \(25\)](#) [画像 \(8\)](#)
- [植物 \(11\)](#) [微生物 \(16\)](#)
- [メタゲノム \(10\)](#) [変異解析 \(6\)](#)
- [CRISPR \(6\)](#) [多型 \(29\)](#)
- [塩基配列 \(158\)](#) [オーソログ \(2\)](#)
- [アミノ酸 \(68\)](#) [構造解析 \(52\)](#)

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

## Q 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 1308 件)

番組のタイトルや画像をクリックすると番組の再生ページへ移動します。[リクエストはこちら。](#)

表示件数を選ぶ ▾

検索窓にキーワードを入れると、入力の度ごとに即座に候補の番組が絞り込まれます



### [NGSハンズオン2017] NGS解析(初～中級) Hi-C解析

本日の統合TVは、2017年8月28日(月)～9月2日(金)に開催された、**バイオインフォマティクス人材育成カリキュラム 平成29年度 NGSハンズオン講習会** から、**国立遺伝学研究所 東 光一氏**による「NGS解析(初～中級) Hi-C解析」をお送りします。約4時間15分です。

本講習では、主にChromosome Conformation Capture 実験で生成されたデータの情報解析について講習を行なう予定です。ヒトゲノムを対象としたHi-Cデータを用いて、マッピング、フィルタリング、コンタクトマップ生成、正規化処理、染色体三次元構造再構成など、各種ツールやPythonスクリプトなどを用いて行なう方法を中心にハンズオン形式で学びます。

講義資料 (PDF: 約12MB)

この動画と講習資料が同時に見られる「講習会 実習資料(AJACS)」ページは[こちら](#)です。

講習会の一連の動画はYouTubeの再生リストからもご覧いただけます。



### [NGSハンズオン2017] NGS解析(初～中級) メタゲノム解析

本日の統合TVは、2017年8月28日(月)～9月2日(金)に開催された、**バイオインフォマティクス人材育成カリキュラム 平成29年度 NGSハンズオン講習会** から、**国立遺伝学研究所 森 宙史氏**による「NGS解析(初～中級) メタゲノム解析」をお送りします。約5時間22分です。

本講習では、主にilluminaのメタゲノム配列データから、配列のクオリティフィルタリング、系統マーカー遺伝子を用いた配列相同性検索による系統組成推定、アセンブル、遺伝子予測、遺伝子機能組成推定等を自作のPerlスクリプトと各種ツール (IDBA-UD, Bowtie2, BLAST等) を用いて行なう方法を中心にハンズオン形式で学びます。

講義資料 (PDF: 約5MB)

この動画と講習資料が同時に見られる「講習会 実習資料(AJACS)」ページは[こちら](#)です。

講習会の一連の動画はYouTubeの再生リストからもご覧いただけます。



### [NGSハンズオン2017] NBDC・DBCLSの各種サービス 公共NGSデータの検索と登録

本日の統合TVは、2017年8月28日(月)～9月2日(金)に開催された、**バイオインフォマティクス人材育成カリキュラム 平成29年度 NGSハンズオン講習会** から、**ライフサイエンス統合データベースセンター (DBCLS)** 仲里 猛留 による「公共NGSデータの検索と登録」をお送りします。約50分です。

本講習では、主にNBDC事業およびNBDCが共同研究機関とともに開発・提供するサービスを紹介します。公共NGSデータの検索と登録を中心にハンズオン形式で学びます。

講義資料 (PDF: 約9MB)

この動画と講習資料が同時に見られる「講習会 実習資料(AJACS)」ページは[こちら](#)です。

講習会の一連の動画はYouTubeの再生リストからもご覧いただけます。



### [NGSハンズオン2017] NBDC・DBCLSの各種サービス 今日から使える便利な生命科学系公共データベース in DBCLS

本日の統合TVは、2017年8月28日(月)～9月2日(金)に開催された、**バイオインフォマティクス人材育成カリキュラム 平成29年度 NGSハンズオン講習会** から、**ライフサイエンス統合データベースセンター (DBCLS)** 小野 浩雅 による「NBDC・DBCLSの各種サービス 今日から使える便利な生命科学系公共データベース in DBCLS」をお送りします。約40分です。

本講習では、主にNBDC事業およびNBDCが共同研究機関とともに開発・提供するサービスを紹介します。今日から使える便利な生命科学系公共データベース in DBCLSを中心にハンズオン形式で学びます。

講義資料 (PDF: 約3MB)

この動画と講習資料が同時に見られる「講習会 実習資料(AJACS)」ページは[こちら](#)です。

講習会の一連の動画はYouTubeの再生リストからもご覧いただけます。



### [NGSハンズオン2017] NBDC・DBCLSの各種サービス NBDCの紹介～NGS関連サービスを中心に

本日の統合TVは、2017年8月28日(月)～9月2日(金)に開催された、**バイオインフォマティクス人材育成カリキュラム 平成29年度 NGSハンズオン講習会** から、**バイオサイエンスデータベースセンター (NBDC)** / **ライフサイエンス統合データベースセンター (DBCLS)** 箕輪 真理 による「NBDC・DBCLSの各種サービス NBDCの紹介～NGS関連サービスを中心に」をお送りします。約41分です。

本講習では、主にNBDC事業およびNBDCが共同研究機関とともに開発・提供するサービスを紹介します。NBDCの紹介～NGS関連サービス



YouTube

ホーム 急上昇

BEST OF YOUTUBE 音楽 スポーツ ゲーム 映画 テレビ番組 ニュース ライブ スポットライト 360 度の全方位動画

ログインする

チャンネル登録 1,031

人気の動画 グリッド

togotv

ホーム 動画 再生リスト チャンネル フリートーク 概要

アップロード済み

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 視聴回数 33,329 回

Gmailの使い方(基本編) 視聴回数 11,194 回

PCRプライマー設計ツール Primer3の使い方 視聴回数 22,128 回

インストールが完了すると、デスクトップに「R 2.8.1」のアイコンが作成され! 視聴回数 9,652 回

統計解析ソフト「R」の使い方～導入編～ 視聴回数 9,652 回

PCRプライマー設計ツール Primer3の使い方 2011 視聴回数 11,334 回

オントモグラフィーを使い倒す 視聴回数 12,076 回

Primer BLASTの使い方 2010 視聴回数 7,756 回

PyMOLを使い倒す 視聴回数 7,616 回

PowerPointの図形描画機能でイラストをつくる方法 その3 視聴回数 7,106 回

blastの使い方 視聴回数 5,809 回

ImageJを利用して画像を処理・解析する 視聴回数 5,816 回

MEGAを使って配列アライメントおよび系統解析をする 視聴回数 5,558 回

PubMedの使い方～基本編～ 視聴回数 5,359 回

配列のオンラインメント作成ツール ClustalWを使い倒す 視聴回数 5,346 回

統計解析ソフト「R」の使い方～正規化編～ 視聴回数 5,280 回

PyMOLを使ってタンパク質の構造を見る 視聴回数 5,259 回

遺伝子発現情報データベース NCBI Gene Expression... 視聴回数 4,912 回

CLUSTALWで配列のオンラインメントを作成する 視聴回数 3,884 回

NCBI BLASTの使い方～基本編～ 2010 視聴回数 3,837 回

ImageJの画像処理パッケージ Fiji を使って画像を三次元的に解析... 視聴回数 3,825 回

高速アライメントツール BLAT をプライマー設計支援ツールと... 視聴回数 3,797 回

DBTSSを使って遺伝子の発現制御領域（プロモーター領域）を... 視聴回数 3,574 回

統計解析ソフト「R」の使い方～導入編～ 視聴回数 3,571 回

【総合TV】How to use Gene Expression Omnibus (GEO)... 視聴回数 3,475 回

第35回日本分子生物学年会オンライン抄録集の使い方 視聴回数 3,201 回

ApE(A plasmid Editor)を利用してプラスミドを設計する 視聴回数 3,243 回

transeqで塩基配列をアミノ酸配列に変換する 視聴回数 3,154 回

統合TVをiTunesで見る 視聴回数 3,016 回

もっと読み込む

YouTube

言語: 日本語

国: 指定なし

制限付きモード: オフ

履歴

ヘルプ

YouTubeについて プレスルーム 著作権センター クリエイター向け 広告掲載 開発者向け +YouTube

利用規約 プライバシー ポリシーとセキュリティ フィードバックの送信 新機能を試してみませんか

- YouTubeのチャンネル登録をすると更新情報がメールで届きます。
- ウェブサイトへのアクセスの仕方から結果の解釈まで、操作の一挙手一投足がわかります。
- 1300本を超える動画が公開されており、YouTube版だけでのべ 900,000回以上 再生されています。(2017年11月末現在)



2007/08/08～2017/11/30

## △ このレポートのデータに関する注意事項

"平均再生率（%）"、"平均視聴時間"、"総再生時間（時間）" のデータは 2012年9月1日 以降に限り使用できます。



動画 地域 日付 もっと見る ▾

動画	総再生時間（時間）*	視聴回数 ? ↓	平均視聴時間* ?	平均再生率* ? (%) *
パワーポイントの図形描画機能でイラストを...	2,716 (5.2%)	44,526 (4.9%)	3:56	32%
PCRプライマー設計ツール Primer3の使い方	293 (0.6%)	24,039 (2.6%)	3:20	34%
Gmailの使い方(基本編)	889 (1.7%)	19,930 (2.2%)	2:40	36%
パワーポイントの図形描画機能でイラストを...	1,287 (2.5%)	18,813 (2.1%)	4:28	39%
PCRプライマー設計ツール Primer3の使い方 2...	1,099 (2.1%)	17,588 (1.9%)	4:53	48%

- 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を見ればお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データベースやウェブツールは、統合TV の「発現解析」タグから検索できます。
- 統合TVに掲載されているコンテンツについてご引用いただく際に、恒久的な URL として DOI (Digital Object Identifier) を使用することができます。
- 2014年8月以降に開催された過去の講習会の資料・テキストと動画が「講習会 実習資料 (AJACS)」で閲覧できるようになり、受講生の復習のみならず、初学者の学習教材としてご活用いただけます。
- 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!!
- 研究発表のスライド作成や資料作成、論文の図表等でどなたでも自由にお使いいただける画像は、自由に使える画像を探す

- 統合TVを作つてみたい方、募集中です。(オンラインでの作成環境を整備しており、遠隔地でもOKです)

---

習熟度ややりたいこと別にご参考ください

- 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よい基礎知識
  - 「塩基配列解析のためのデータベース・ウェブツール」(2015年9月AJACS伊予)
- 遺伝子発現データを公共DBで検索・取得・解析する方法について
  - 「遺伝子発現DB・ウェブツールの使い方 応用・実践編」(2015年5月AJACS御茶ノ水)
  - 「遺伝子発現DBを含む公共オミックスDBの使い方」(2017年8月AJACS河内)
- 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)
- 次世代シーケンス(NGS)データの解析について
  - 「次世代シーケンサー (NGS) と関連するデータベース・ツール」(2015年9月AJACS伊予)
  - 「次世代シーケンサー(NGS)データから遺伝子発現を見るためのホップ&ステップ」(2015年9月AJACS伊予)
- NGS解析について、さらにもっと基礎から応用までを深く学びたい方向け(それぞれ約50時間程度)
  - 「バ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(2014年8月)」のYouTubeリスト
  - 「バ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(NGS)ハンズオン講習会(2015年8月)」のYouTubeリスト
  - 上記の動画+講習会資料のまとめページ@統合TV

---

個々の遺伝子の発現プロファイルを調べる

## RefEx (Reference Expression dataset)

- 遺伝子発現解析の基準となるデータを快適に検索できるウェブツール
  - <http://refex.dbcls.jp/>
- 公共DBにある正常組織や細胞株における遺伝子発現データを再利用・整理
- 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
- 論文出ました!(2017年8月)
  - Ono H, Ogasawara O, Okubo K, Bono H
  - RefEx, a reference gene expression dataset as a web tool for the functional analysis of genes
  - Scientific Data, 4:170105
  - DOI: [10.1038/sdata.2017.105](https://doi.org/10.1038/sdata.2017.105)
  - この論文に関する日本語のレビュー記事 @ ライフサイエンス新着論文レビュー
  - Nature ダイジェスト著者インタビュー記事「FANTOM5データを誰でも活用できる形に」
- このツールでできること

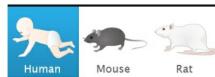
- 正常組織における遺伝子発現データを調べる
  - 測定手法による遺伝子発現量の差異を比較する
  - 組織特異的遺伝子をワンタッチで検索可能
  - 遺伝子発現解析などで見出された不詳な遺伝子群の機能および関係性を調べる
- 
- RefExで掲載されているデータはすべて再利用可能
    - オリジナルデータの再処理方法の詳細は[GitHub](#)に
    - 再処理済みの発現データやサンプルアノテーション等のすべてのデータは[figshare](#)に
    - 「The RefEx analysis」として論文に引用していただいた活用例
      - [Aberrant IDH3 \$\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 @ Figure 6](#)
      - がん研究者が、発現解析実験で見出した数百個の治療標的・候補遺伝子の絞込みに使えないか検討した。
      - これらの候補遺伝子の正常組織における発現量が低ければ、治療標的とした場合に悪影響・副作用が小さくなると仮説した。
      - 実際に、これらの遺伝子の発現量をRefExで確認し、追加確認実験の優先順位付けを効率的に行うことができた。

---

### 【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](#)

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](#)
- [Krueppel-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



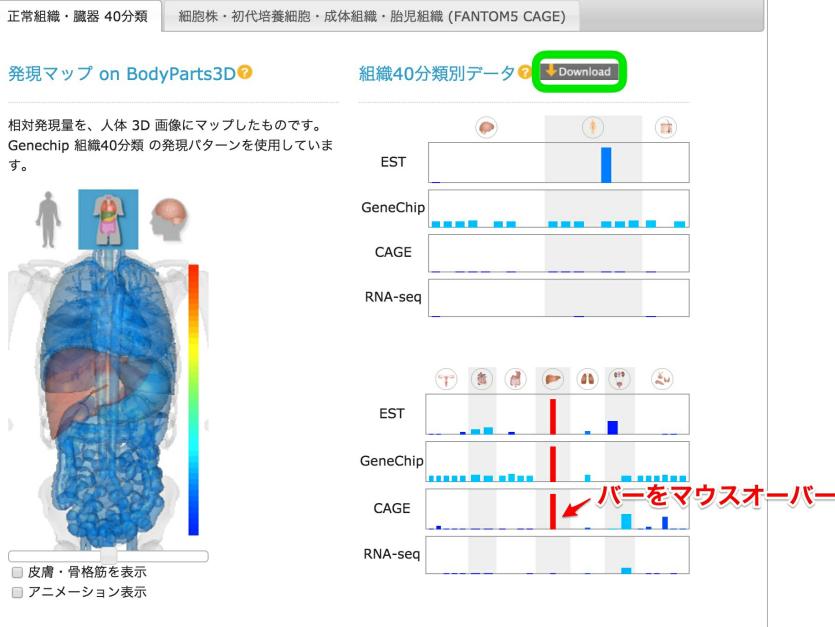
RefEx はCreative Commons 表示 2.1 日本 License の下でライセンスされています。  
原著者はライフサイエンス統合データベースセンターです。[CC BY]

[refex.dbcls.jp/genelist.php?lang=ja&db=human&roku\\_valid=1&r\[31\]=31&order\\_key=score](http://refex.dbcls.jp/genelist.php?lang=ja&db=human&roku_valid=1&r[31]=31&order_key=score)

- 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](#))をクリックすると詳細情報を閲覧できます。
  - 「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。
  - 「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。
  - 遺伝子オントロジー(Gene Ontology: 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。

## 発現データ

## 遺伝子詳細情報



## 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[2] NM\_001111048, NM\_010196  
 ラット[6] NM\_001008724, NM\_052797

## 染色体 ?

## 遺伝子オントロジー (GO ID) ?

## Biological Process

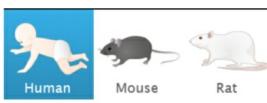
GO:0007596 blood coagulation  
 GO:0030168 platelet activation  
 GO:0002576 platelet degranulation  
 GO:0051258 protein polymerization  
 GO:0051592 response to calcium ion  
 GO:0007165 signal transduction

## Cellular Component

GO:0005938 cell cortex  
 GO:000986 cell surface  
 GO:0009897 external side of plasma membrane  
 GO:0005576 extracellular region  
 GO:0005615 extracellular space

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。

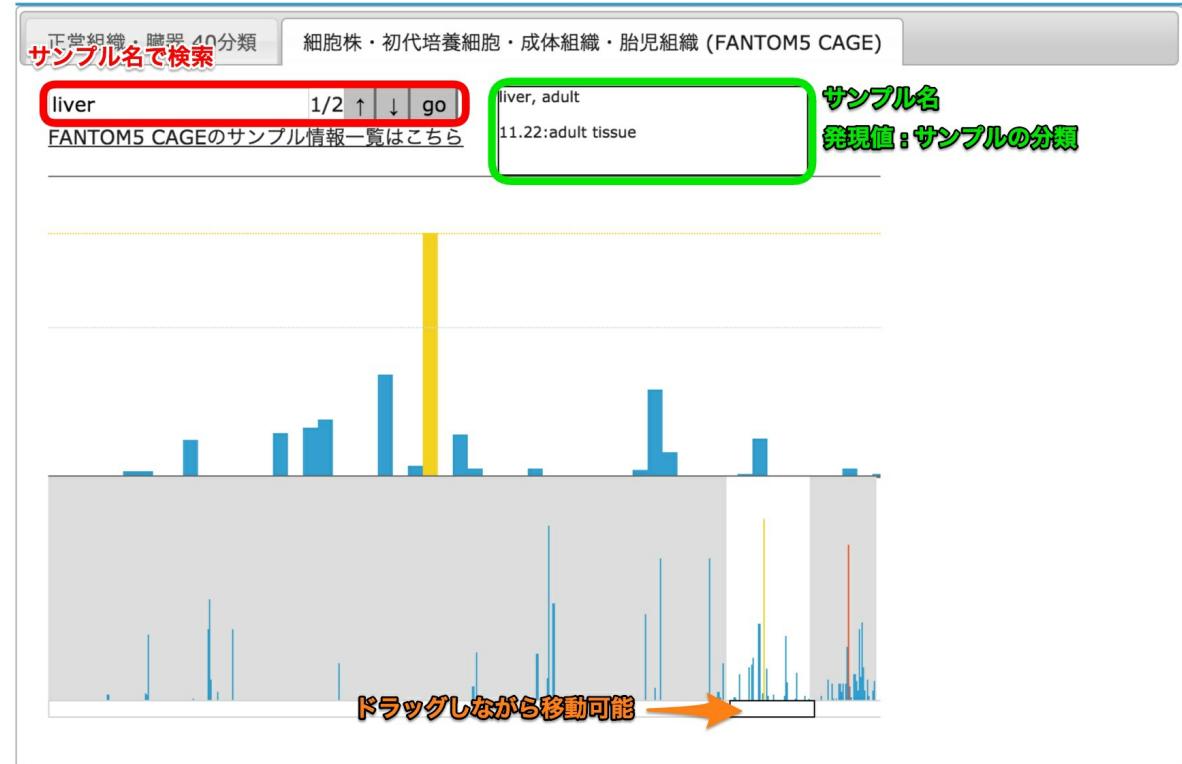


## fibrinogen alpha chain

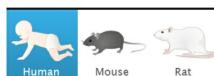
[詳細情報を見る](#)

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

### 発現データ



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。



結果一覧 470 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

**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**

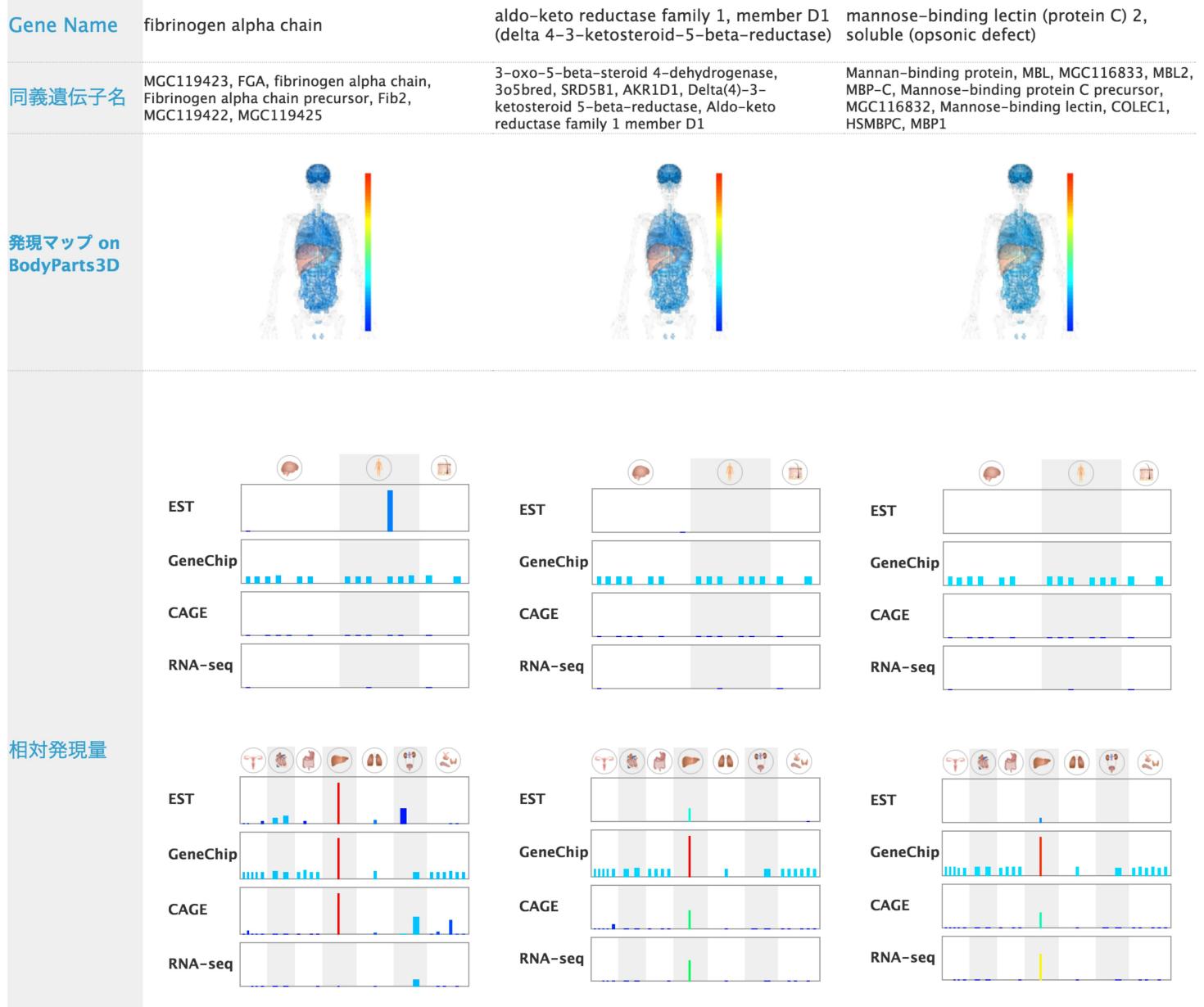


12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- ・ 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。
- ・ 遺伝子発現データやGeneOntology情報を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。その違いからどういうことが推測できるでしょうか。



Refseq ID	<a href="#">NM_000508</a>	<a href="#">NM_005989</a>	<a href="#">NM_000242</a>
Gene ID	<a href="#">2243</a>	<a href="#">6718</a>	<a href="#">4153</a>
Unigene ID	<a href="#">Hs.351593</a>	<a href="#">Hs.201667</a>	<a href="#">Hs.499674</a>
probe set ID	<a href="#">205649_s_at</a>	<a href="#">207102_at</a>	<a href="#">207256_at</a>
Ensembl ID	<a href="#">ENSG00000171560</a>	<a href="#">ENSG00000122787</a>	<a href="#">ENSG00000165471</a>
染色体	<a href="#">4.q31.3 [155504278 – 155511918]</a>	<a href="#">7.q33 [137687070 – 137802732]</a>	<a href="#">10.q21.1 [54525140 – 54531460] LRG_154. [5001 – 11321]</a>
遺伝子ファミリー (Interpro ID)	- - - - - -  blood coagulation platelet activation platelet degranulati ... protein polymerizati ... response to calcium ... signal transduction	- - - - - -  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	差分が明確に		

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。

## 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**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
- マウスのエキソンアレイのデータから遺伝子のスプライシングバリエント(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。

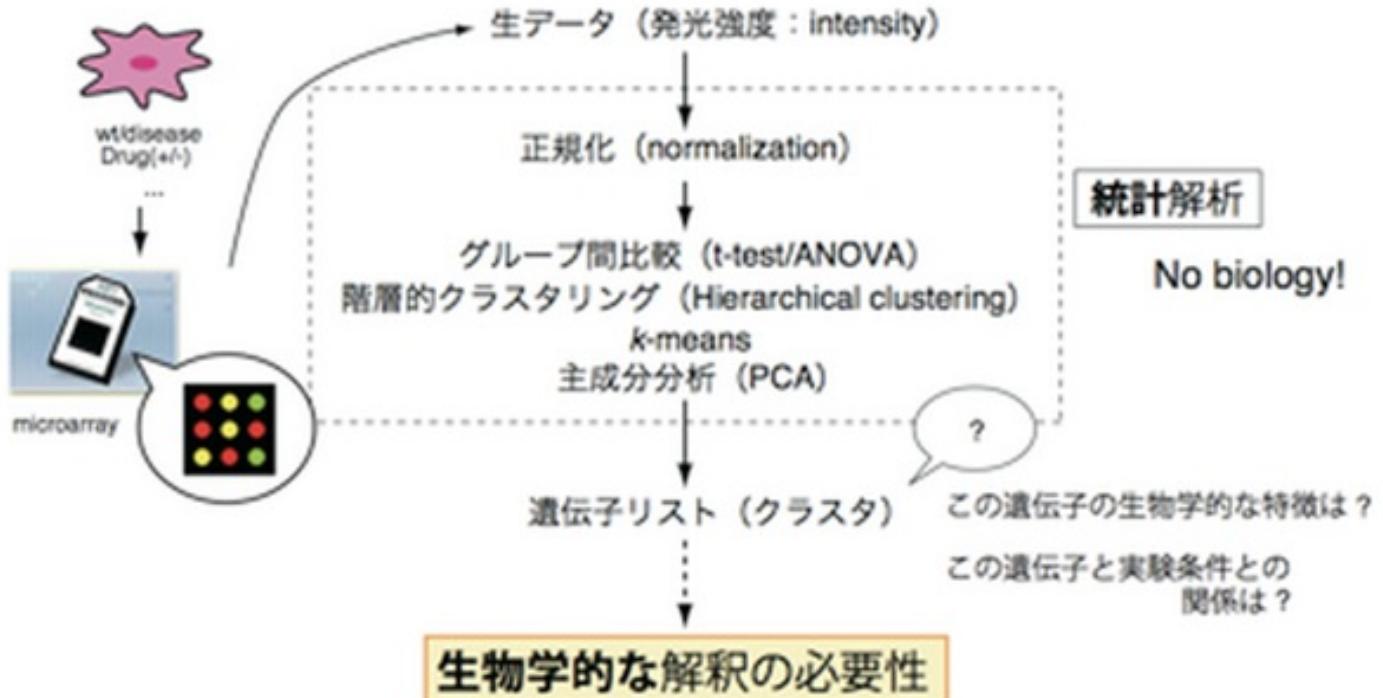
## 【実習(skip)】 BioGPSを使ってある遺伝子の発現プロファイルを調べる

- 【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
- 【以前の講習会動画】 遺伝子発現データベースの活用法

1. <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。
  2. 表示された検索結果の中から「ID 4654」をクリックします。
  3. 最初はヒトのマイクロアレイデータが表示されます。
  4. 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。
  5. ページ内のウインドウは通常のウイ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。
  6. "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。
  7. "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。
  8. 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。
  9. 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。
  10. MyoDはどの組織、細胞で強く発現しているでしょうか？
  11. 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？
  12. "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。
  13. "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？
  14. "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。
  15. "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。
  16. 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？
  17. 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。
  18. 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

- マイクロアレイやNGS実験を行うと大量の発現変動遺伝子 (Differentially Expressed Genes: DEGs)が得られます。
- 一般的な遺伝子発現解析の第一歩は、実験条件によって得られた数十～数千のDEGs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、Gene Ontology (GO) の用語を使って、マイクロアレイ実験で得られたDEGsのもつ機能に、どのような特徴があるのか(転写因子活性に関する遺伝子が多いのか、細胞周期に関する遺伝子が多いのか?、Wntパスウェイに関する遺伝子が多いのか?、など)を解析することで、生物学的解釈をしてみましょう。

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

- アメリカ国立アレルギー・感染症研究所が開発・運用
- 原著論文 PMID: 19131956
- 遺伝子リストのコピペで簡単にエンリッチメント解析 ( GO、KEGG など )
- 対応生物種・遺伝子IDが豊富。ID変換ツールもある
- IDリストしか投げられない(発現量込みやタイムコースデータは不可)
- 2010年以来データ更新が止まっていたが、最近、アップデートされた。DAVID 6.8 (current beta release) May. 2016

### 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 (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
- 【復習用】 DAVIDの使い方 実践編

- DAVIDにアクセスし、上部メニューの「Start Analysis」をクリックします。

## 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

Search

### 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](#)

2. 画面左側バーで、probe IDリストをコピペ or ファイルを指定します。
3. 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」
4. Submit List をクリックするとリストが読み込まれます。

**Upload List Background**

**Upload Gene List**

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

[Upload Help](#)

**Step 1: Enter Gene List**

A: Paste a list

[Clear](#)

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**

## 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

5. 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。

## 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:

[Show Gene List](#)

[View Unmapped Ids](#)

[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](#)

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

7. 「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。

## 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:

[Show Gene List](#)

[View Unmapped Ids](#)

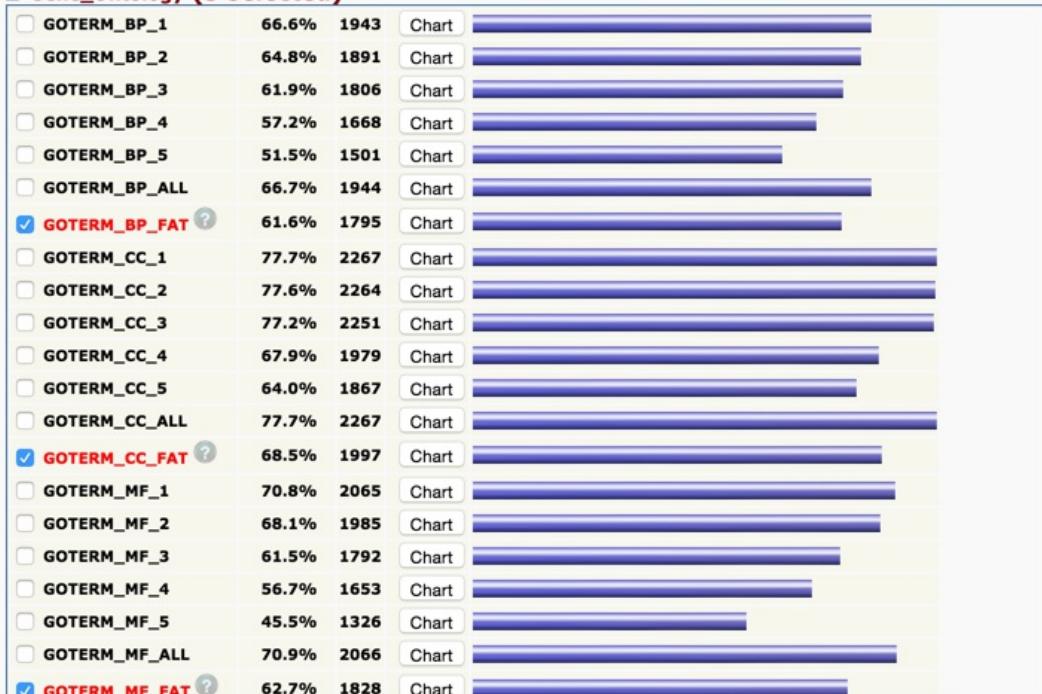
## Annotation Summary Results

[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) 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)

8. 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。

## Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List\_1

Current Background: Arabidopsis thaliana

2916 DAVID IDs

Options

[Rerun Using Options](#) [Create Sublist](#)

301 chart records

[Download File](#)

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

9. タイトル行をクリックするとソートできます。

10. 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

[Rerun Using Options](#) [Create Sublist](#)

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

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

11. Pathways > KEGG\_PATHWAY や Tissue Expression > UP\_TISSUE なども見てみましょう。

12. 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

【実習3】これまで学んだことを踏まえて、発現データの結果を生物学的に解釈する

- 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](#) 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](#)
    - 「AH患者>健常者\_遺伝子リスト」 [GEO2R\\_AH.txt](#)
    - (この遺伝子リストの作り方は、AJACS御茶ノ水の回 で解説しています。)
- DAVID 以外のツールを使ってみる
  - [GeneTrail2](#)
    - 2016年1月公開。ザールラント大(独)が開発・運用。原著論文 [PMID: 26787660](#)
    - トランスクリプトームのほか、プロテオーム、miRNA、SNP にも対応
    - GSE 番号 の入力だけで、GEOから直接データ取得が可能
    - IDリストのほか発現量込みリスト、タイムコースデータなども使用可能
    - 主要なモデル生物種に対応
    - 実験系に適した統計解析の選択肢が豊富
    - 同じ生物種間であれば、別の解析結果同士を比較することも可能

- 統合TV あります → [GeneTrail2を使って、エンリッチメント解析を行う](#)
  - 解析結果セットはダウンロード可能だがアップロードして再表示はできない
  - データセットによってエラーが出て解析できない...(バグ?)
- [Metascape](#)
  - 2015年10月スタート。原著論文 [PMID: 26651948](#)
  - 「なぜ、DAVIDはもはや使うべきではないのか」提言 → metascape 使おう
  - 対応ID: Entrez Gene ID, RefSeq, Gene Symbol, Ensembl, UCSC, UniProt.
  - 生物種は、ヒト、マウス、ラットのみ
  - IDリストのほかタイムコースなどの複数リストデータも使用可能
  - 複数リスト間のアノテーションについて差分表示が可能
  - GOのエンリッチメント解析で階層的クラスタリングもできる
  - 統合TV あります → [Metascapeを使って、遺伝子リストの生物学的解釈をする](#)
  - 講習動画・資料あります 「遺伝子発現DBを含む公共オミックスDBの使い方」(2017年8月AJACS河内)
- [GeneSetDB](#)
  - 九州大学 荒木さんが開発。オークランド大学バイオインフォマティクス研究所が運用。原著論文 [PMID: 23650583](#)
  - 医学・薬学分野に特化したデータベースを解析対象にすることができる
  - 統合TV あります → [GeneSetDBで遺伝子解析とエンリッチメント解析を行う](#)
    - [2:50~ エンリッチメント解析を行う](#)
- 一応ひとつの答え
  - このデータを使った論文があります。
    - [Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis.](#)
    - [Gut. 2013 Mar;62\(3\):452-60. doi: 10.1136/gutjnl-2011-301146.](#)
  - 似たような結論が導かれましたか? あるいは、著者らが見逃している(かもしれない)着眼点や新たな着想が得られましたか?

## まとめ

- つまり食い的ではありますが通り一遍の大規模発現データに対する生物学的解釈の方法を学びました。
- 「道具」を知って使い方が分かれば、あとは情報分析力と想像力の勝負。
- ぜひご自身のデータ、あるいはご自身のテーマに関連する公共データの生物学的解釈をしてみましょう。
- 実戦実践あるのみ