

AJACS薩摩 生物学的解釈をするための遺伝子発現DB・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

2016年1月26日(火) AJACS薩摩@鹿児島大学医学部 共通教育棟6階 マルチメディア情報演習室(桜ヶ丘キャンパス)

これは統合データベース講習会 AJACS薩摩「生物学的解釈をするための遺伝子発現DB・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

<http://doi.org/10.7875/ajacs.2015.007> 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 <http://events.biosciencedbc.jp/training/ajacs58> 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。

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 37)

- 統合TVを知っていますか?
 - 知らない **12名 (32%)**
 - 聞いたことがある **7名 (19%)**
 - 知っている **4名 (11%)**
 - 使ったことがある **8名 (22%)**
 - 使っている **6名 (16%)**
- 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。
 - これから実験をする・したい **7名 (19%)**
 - 公共データを活用する・したい **10名 (27%)**
 - 既に持っている **18名 (49%)**
 - 大規模発現解析の予定はない **2名 (5%)**

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 <http://togotv.dbcls.jp/ja/>

- 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/>

TOGO-TV 生命科学系DB・ツール使い倒し系チャンネル

『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

- AJACS講習会資料
- 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
- 学会講演・講習会

関連するタグから検索

- AJACS (276) BioHackathon (74)
- Bioconductor (18) CAGE (6)
- CRISPR (3) ChIP-seq (10)
- DBCLS (174) DDBJ (37)
- DDBJing (15) EBI (10)
- EMBOSS (21) English (108)
- Ensembl (19) FANTOM (4)
- GEO (17) Galaxy (31) Google (11)
- KEGG (14) Linux (12) NBDC (30)
- NCBI (51) NGS (100)
- NGS速報・ハンズオン (52)
- PDB (20) PDBj (8) Perl (8)
- Python (2) R (24) RDF (24)
- RNA-seq (15) Reseq (4) UCSC (16)
- UNIX (25) アミノ酸 (68)
- アライメント (34)
- ウェブツール (109) オーソログ (2)
- クリエイティブ・コモンズ (10)
- ゲノム (123)
- ゲノムリガラシ講座 (18)
- シェルスクリプト (4) スパコン (9)
- ソフトウェア (51) タンパク質 (75)
- データベース (233)
- トーゴーの日 (73)
- バイオライン (23)
- パスウェイ解析 (39)
- パワーポイント (3) フェノーム (1)
- プログラミング (36)
- プロテオーム (5) ポータル (37)
- メタゲノム (4) メタボローム (14)
- メール (2) ワークフロー (9)
- 倫理 (2) 創業 (4) 化合物 (17)
- 可視化 (146) 壁基配列 (134)
- 変異解析 (2) 多型 (25) 実習 (213)
- 家系図 (1) 微生物 (11)
- 情報収集 (44) 抗体 (7)
- 文献検索 (72) 文章執筆 (16)
- 植物 (8) 構造解析 (45) 特許 (9)
- 環境設定 (81) 生命活動 (3) 画像 (7)
- 疾患情報 (11) 発現解析 (161)
- 登録方法 (6) 糖鎖 (4) 系統樹 (14)
- 表現型 (1) 講演 (401) 転写因子 (14)
- 辞書 (24) 遺伝子 (201)
- 配列解析 (113)

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 1029 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。

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

生物多様性情報分野からのインプット @Annotathon2015

本日の統合TVは、2015年11月12-13日(木・金)に行われた2015年度 国立遺伝学研究所 研究会 Annotathon2015 (生命科学データベースの利用価値向上のためのアノテーションマラソン) から 国立科学博物館 神保 宇嗣 研究員による「生物多様性情報分野からのインプット」をお送りします。約22分です。
研究会の一連の動画はYouTubeの再生リストからもご覧いただけます。

微生物ゲノムアノテーションの現状と問題点@Annotathon2015

本日の統合TVは、2015年11月12-13日(木・金)に行われた2015年度 国立遺伝学研究所 研究会 Annotathon2015 (生命科学データベースの利用価値向上のためのアノテーションマラソン) から 東京工業大学大学院 生命理工学研究科 生命情報専攻 森 宙史 助教による「微生物ゲノムアノテーションの現状と問題点」をお送りします。約35分です。
研究会の一連の動画はYouTubeの再生リストからもご覧いただけます。

ChIP-seqデータベースのためのメタ情報アノテーション@Annotathon2015

本日の統合TVは、2015年11月12-13日(木・金)に行われた2015年度 国立遺伝学研究所 研究会 Annotathon2015 (生命科学データベースの利用価値向上のためのアノテーションマラソン) から 九州大学大学院 医學研究院 発生再生医学分野 沖 真弥 助教による「ChIP-seqデータベースのためのメタ情報アノテーション」をお送りします。約15分です。
既報のChIP-Seqデータを簡単な操作で自動的に可視化するSeqTailorや、既報のChIP-Seqデータの可視化と解析を行うChIP-Atlasを開発する上で必要となるメタデータのアノテーションの現状と対策についてのお話です。
研究会の一連の動画はYouTubeの再生リストからもご覧いただけます。

ゲノムアノテーションそもそも@Annotathon2015

本日の統合TVは、2015年11月12-13日(木・金)に行われた2015年度 国立遺伝学研究所 研究会 Annotathon2015 (生命科学データベースの利用価値向上のためのアノテーションマラソン) から 国立遺伝学研究所 大量遺伝情報研究室 中村 保一 教授による「ゲノムアノテーションそもそも」をお送りします。約9分です。
研究会の一連の動画はYouTubeの再生リストからもご覧いただけます。

Annotathonのこれまでとこれから@Annotathon2015

本日の統合TVは、2015年11月12-13日(木・金)に行われた2015年度 国立遺伝学研究所 研究会 Annotathon2015 (生命科学データベースの利用価値向上のためのアノテーションマラソン) から ライフサイエンス統合データベースセンター 仲里 猛留による「Annotathonのこれまでとこれから」をお送りします。約9分です。
研究会の一連の動画はYouTubeの再生リストからもご覧いただけます。

番組メニュー

- はじめての方へ
- 統合TVの特徴
- 統合TVの使い方
- 統合TVの歴史
- スタッフ
- サイトポリシー
- ロゴについて
- AJACS講習会資料
- 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
- 学会講演・講習会
- 活用事例
- 再生数ランキング

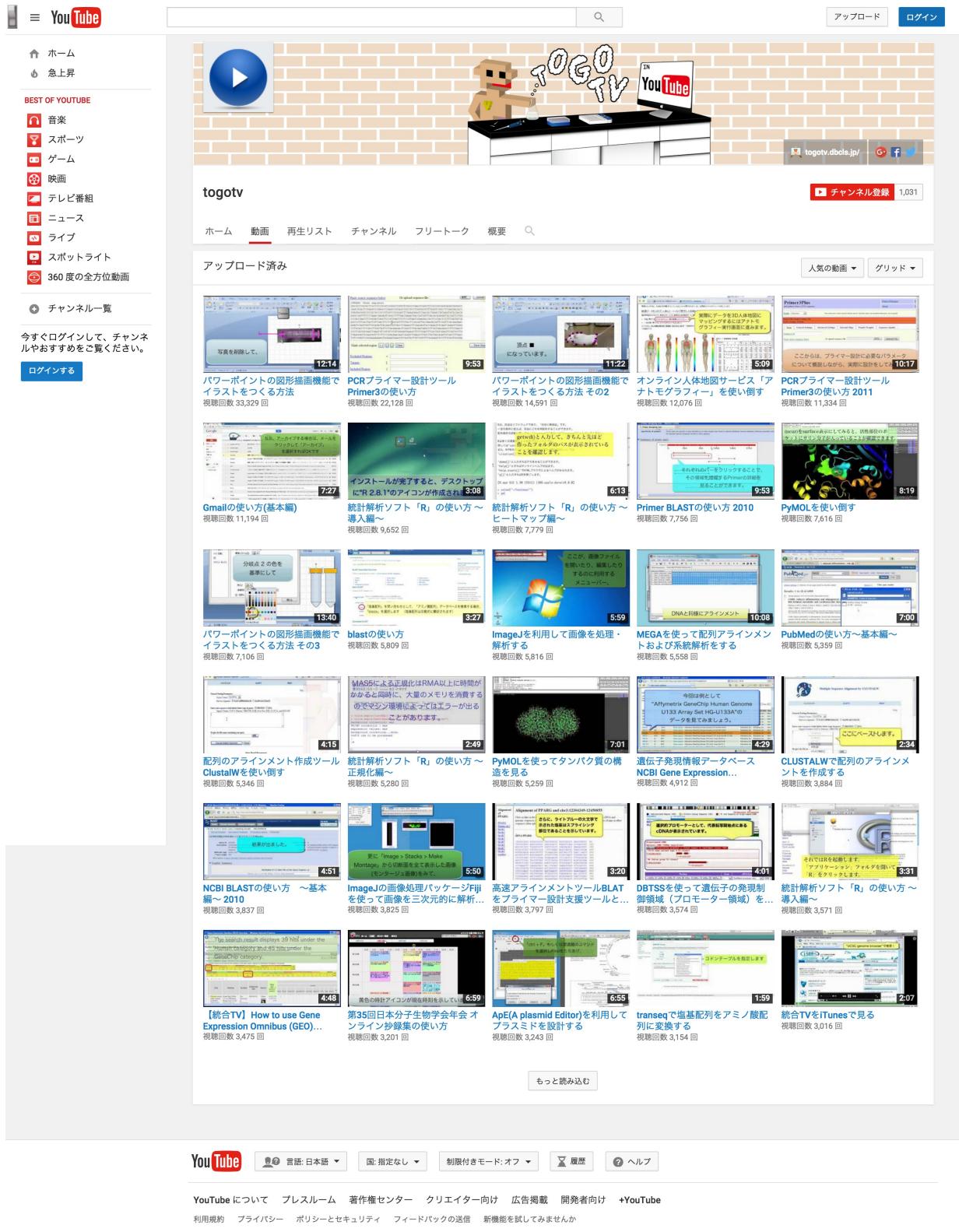
質問

- よくある質問
- 番組リクエスト

最新情報

- @togotv
- RDF

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



- 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。
 - 1000本を超える動画が公開されており、YouTube版だけで のべ 560,000回以上再生されています。(2015年12月末現在)



<https://gyazo.com/35ed4ae06281d5c36b332c264285fbc7>

- 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 <http://tогotv.dbcls.jp/ja/tags.html?tag=%E7%99%BA%E7%8F%BE%E8%A7%A3%E6%9E%90>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
<https://ja.wikipedia.org/wiki/%E3%83%87%E3%82%B8%E3%82%BF%E3%83%AB%E3%82%AA%E3%83%96%E3%82%B8%E3%82%A7%E3%82%AF%E3%83%88%E8%AD%98%E5%88%A5%E5%AD%90>(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「AJACS講習会資料 http://togotv.dbcls.jp/ja/ajacs_text.html」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
https://docs.google.com/forms/d/15pxJHnsV_Cu8B55Xy0F3jg5S9FXupkbBqONrZsyUE7k/viewformへどうぞ!!
- 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)
<http://togotv.dbcls.jp/ja/ajacs2015033.html>
- 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) <http://togotv.dbcls.jp/ja/ajacs2015007.html>
- 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(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) <http://togotv.dbcls.jp/ja/ajacs2015031.html>
 - 「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) <http://togotv.dbcls.jp/ja/ajacs2015030.html>
- 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 https://www.youtube.com/playlist?list=PL0uaKHgcG00abmj1Nzs1SUhqKLjf_PFBB
 - 「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
<https://www.youtube.com/playlist?list=PL0uaKHgcG00Yo0Cn0rcF23xof5hqCzGQb>

- 上記の動画+講習会資料のまとめページ@統合TV
[http://tgotv.dbcls.jp/ja/tags.html?
tag=NGS%E9%80%9F%E7%BF%92%E3%83%BB%E3%83%8F%E3%83%B3%E3%82%BA%E3%82%AA%E3%83%B3](http://tgotv.dbcls.jp/ja/tags.html?tag=NGS%E9%80%9F%E7%BF%92%E3%83%BB%E3%83%8F%E3%83%B3%E3%82%BA%E3%82%AA%E3%83%B3)

個々の遺伝子の発現プロファイルを調べる

RefEx (Reference Expression dataset) <http://refex.dbcls.jp/>

- 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
<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 IDH3a 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/onc/journal/vaop/nccurrent/full/onc2014411a.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://doi.org/10.7875/togotv.2014.009> DOI:
10.7875/togotv.2014.009 <http://doi.org/10.7875/togotv.2014.009>
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。

The screenshot shows the RefEx search interface. At the top, there are links for Reference, Expression, Dataset, English | 日本語, and a logo with the text "RefEx". To the right, there is a section for "Organ-specific discovery comparison" with "4 experimental methods" and a link to "View more". Below this is a search bar with placeholder text "ex) troponin, ALB". To the left of the search bar are icons for Human, Mouse, and Rat. A "Keyword search" input field is also present. In the center, there is a grid of icons representing different organs. One of the icons, representing the liver, is highlighted with a red arrow and a hand cursor, indicating it has been clicked. Below this grid, there are sections for "Organ-specific genes" (with a "Gene Ontology" link), "Genetic family" (with an "InterPro" link), and "Chromosome" (with a "Chromosome region selection" button). At the bottom, there are links for "Advanced Search" and "Advanced Search" (disabled), and a "Return to top" button.

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 http://refex.dbcls.jp/gene_info.php?lang=ja&db=human&genelID=2243&refseq=NM_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。
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分類別の発現データがタ

ブ区切り形式でダウンロードできます。

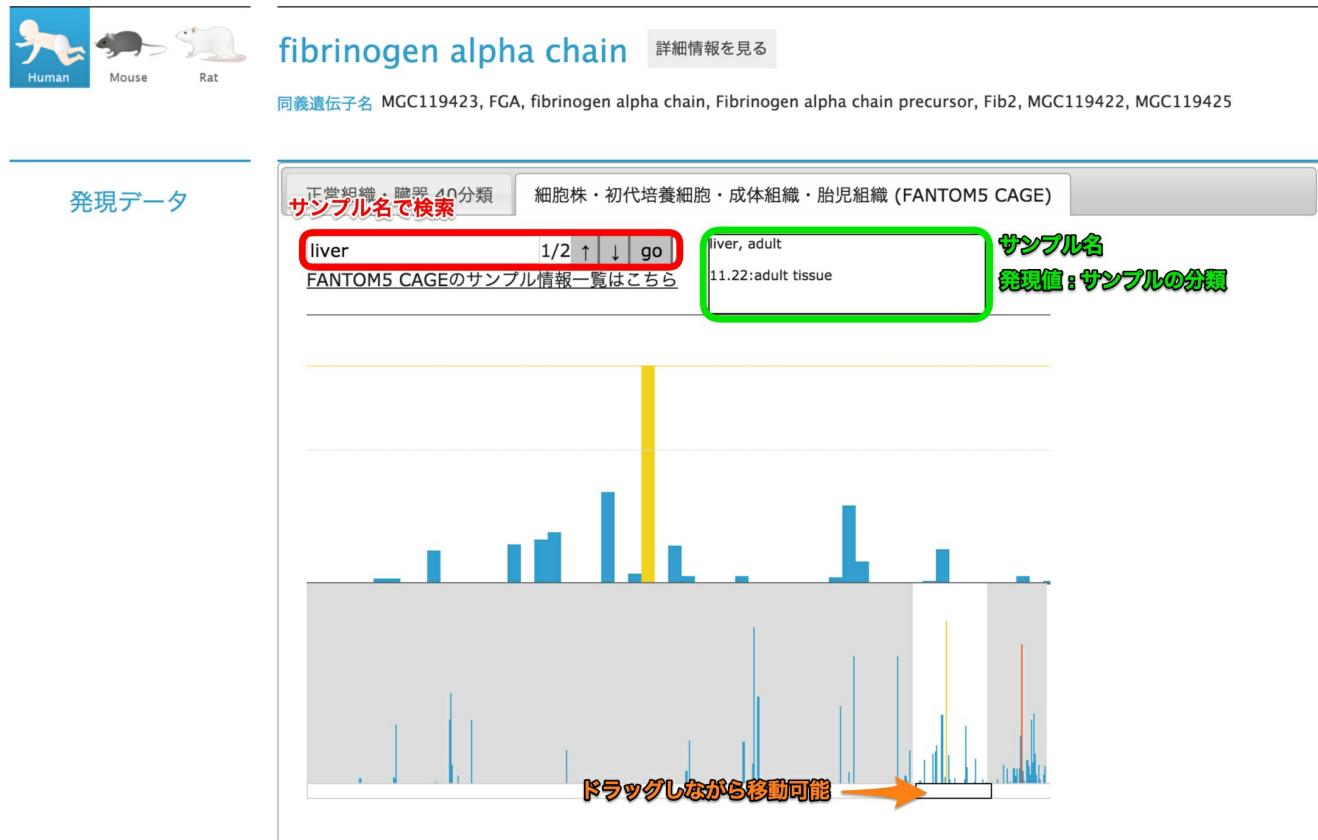
8. 「Probe set ID」のリンク先をクリックすると、どういう情報が参照できるでしょうか。
9. 遺伝子オントロジー(Gene Ontology <http://array.cell-innovator.com/?p=1085>:GO ID)をクリックすると、そのGO termを持つ他の遺伝子を一括で検索できます。
- 例として、GO:0007596 <http://refex.dbcls.jp/genelist.php?lang=ja&db=human&idkind=1&ids=GO:0007596> blood coagulation をクリックしてみましょう。



<https://gyazo.com/3cca3d33c17e845ad5dae8749d7fdb15>

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 <http://bit.ly/fantom5cagehuman>も閲覧可能です。



<https://gyazo.com/51d0b3db07efe64a6fdafb054cd4217f>

- 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。



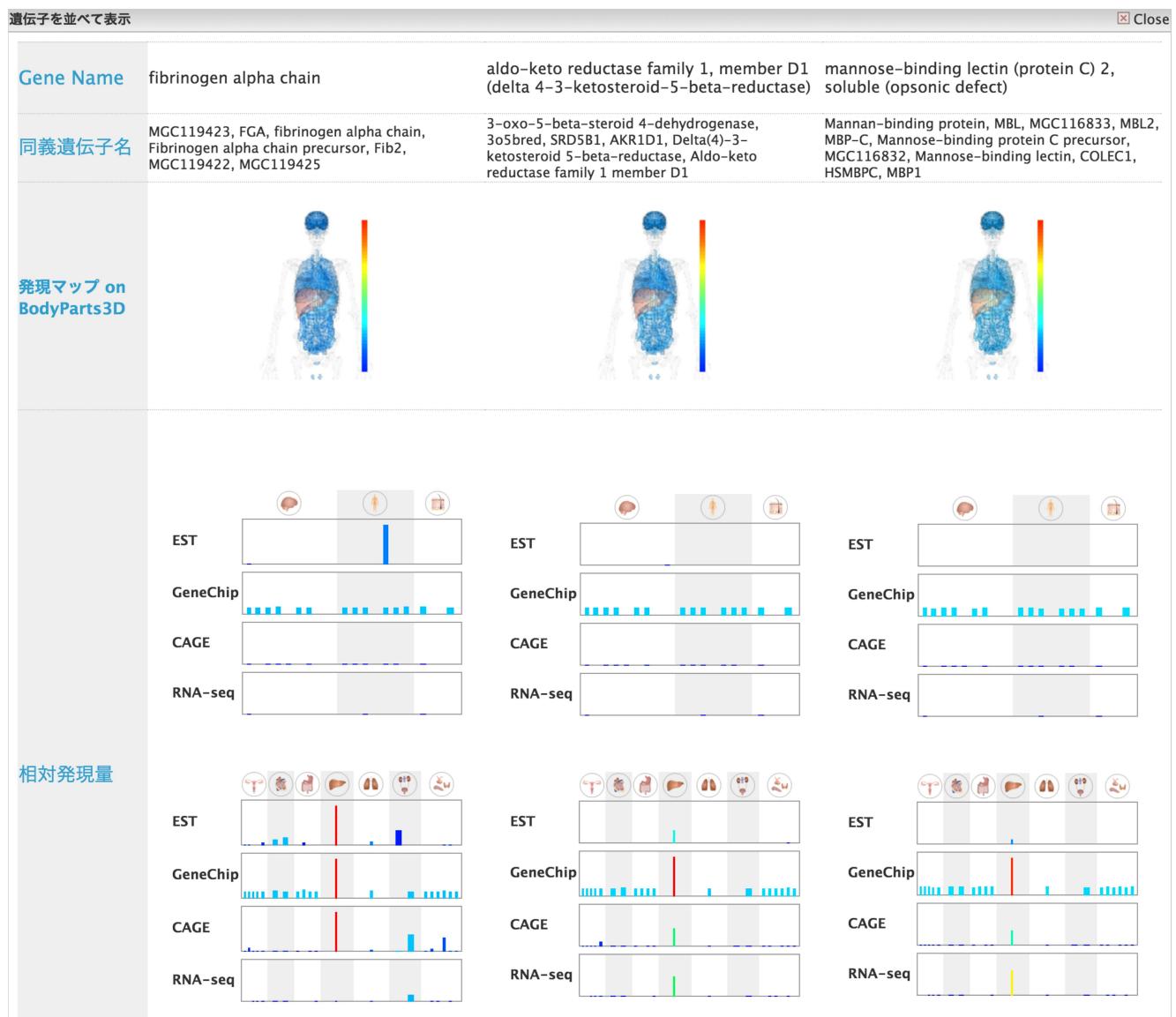
<https://gyazo.com/5e87d20d9b31711e797ef17677be7263>

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 http://refex.dbcls.jp/genelist.php?lang=ja&db=human&roku_valid=1&rk%5B31%5D=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」はなんでしょうか。



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 - - - 差分が明確に	- - 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 degranulati ... protein polymerizati ... response to calcium ... signal transduction	- - - - - -	- - - - - -
		bile acid biosynthet ... bile acid catabolic ... bile acid metabolic ... C21-steroid hormone ... cholesterol cataboli ... digestion oxidation reduction	- - - - - - -
			acute-phase response complement activatio ... complement activatio ... complement activatio ... defense response to ... defense response to ... innate immune respon ... killing by host of s ... negative regulation ... opsonization positive regulation ... response to oxidativ ...

<https://gyazo.com/3f2f9e5cbf135a2c298ae164d6360302>

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 <http://biogps.org/>

- 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

- マウスのエキソソ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

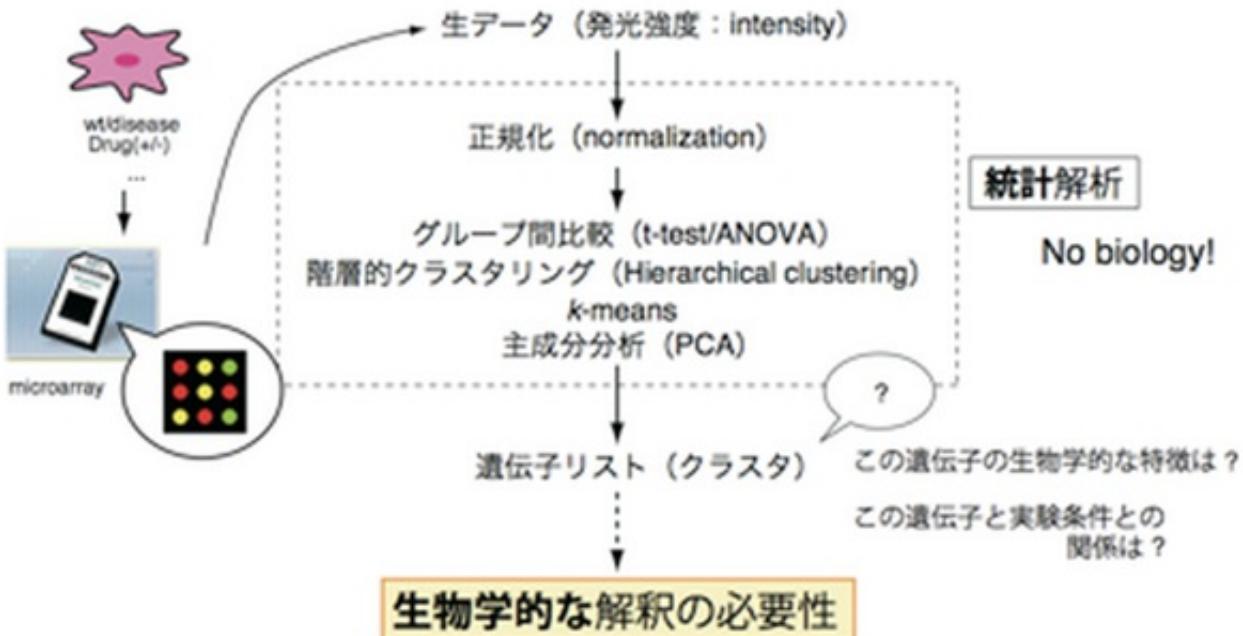
- 【復習用】遺伝子発現プロファイルデータベースBioGPSを使い倒す2012
<http://doi.org/10.7875/togotv.2012.075> DOI: 10.7875/togotv.2012.075
<http://doi.org/10.7875/togotv.2012.075>
 - 【以前の講習会動画】遺伝子発現データベースの活用法
<http://togotv.dbcls.jp/ja/20100829.html#p01>
1. <http://biogps.org/> <http://biogps.org/>を開きます。2.骨格筋の分化決定遺伝子である Myogenic 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。
 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できます。
 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
<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 <http://david.abcc.ncifcrf.gov/>

- 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。

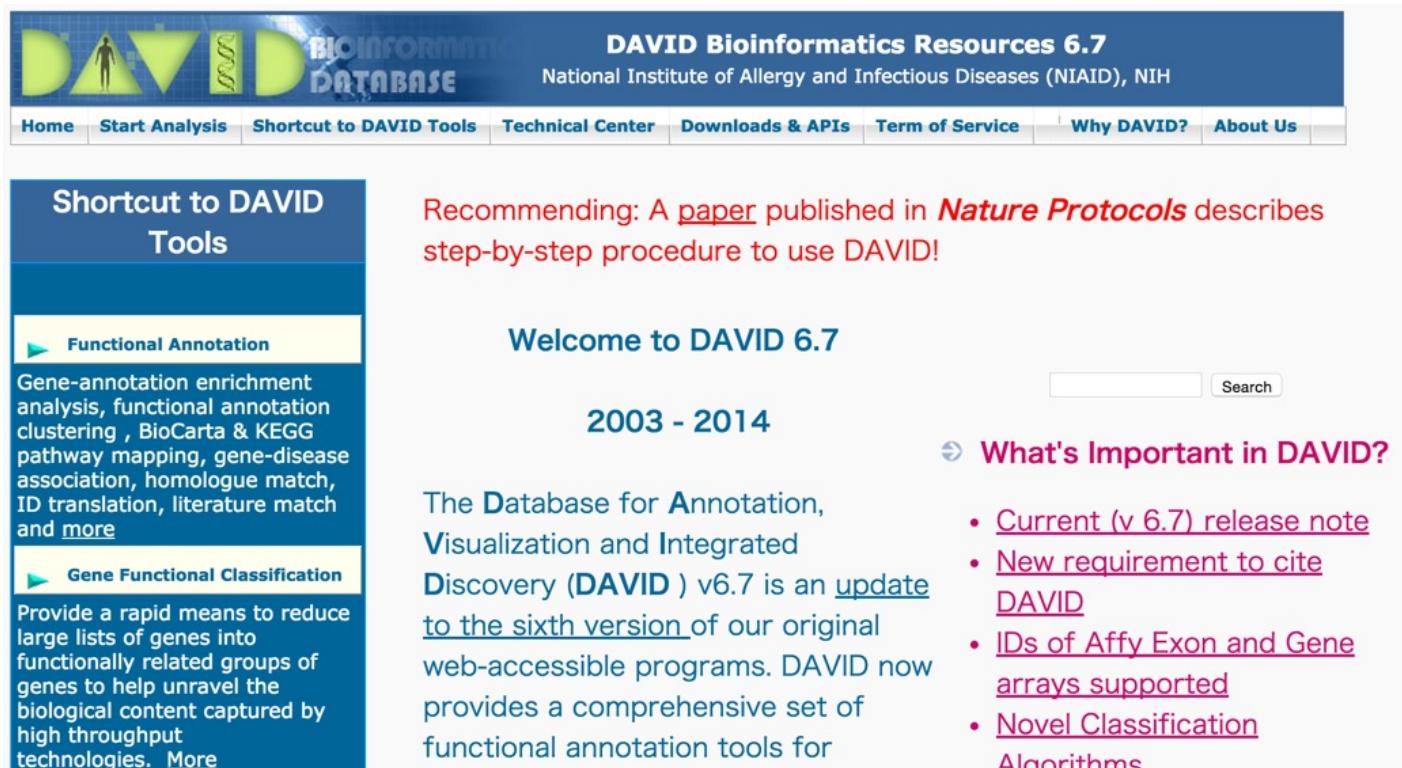
マイクロアレイデータの準備

- 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 (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://doi.org/10.7875/togotv.2012.079> DOI: 10.7875/togotv.2012.079
<http://doi.org/10.7875/togotv.2012.079>
 - 【復習用】 DAVIDの使い方 実践編 <http://doi.org/10.7875/togotv.2013.033> DOI: 10.7875/togotv.2013.033 <http://doi.org/10.7875/togotv.2013.033>
1. <http://david.abcc.ncifcrf.gov/> <http://david.abcc.ncifcrf.gov/>にアクセスし、上部メ

ニューの「Start Analysis」をクリックします。



The screenshot shows the DAVID Bioinformatics Resources 6.7 homepage. At the top, there's a navigation bar with links for Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us. The main content area has a blue header "Shortcut to DAVID Tools". Below it, a red text box says: "Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!". To the left, there are two boxes: "Functional Annotation" (describing gene-annotation enrichment analysis) and "Gene Functional Classification" (describing how it reduces gene lists into functional groups). The central part features a "Welcome to DAVID 6.7" section with a timeline from 2003 to 2014, followed by a detailed description of the database's features. On the right, there's a sidebar titled "What's Important in DAVID?" with a bulleted list of updates.

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

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>

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 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

[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

<http://gyazo.com/e8275cf9dbb203b3d8577307b462c783>

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:

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

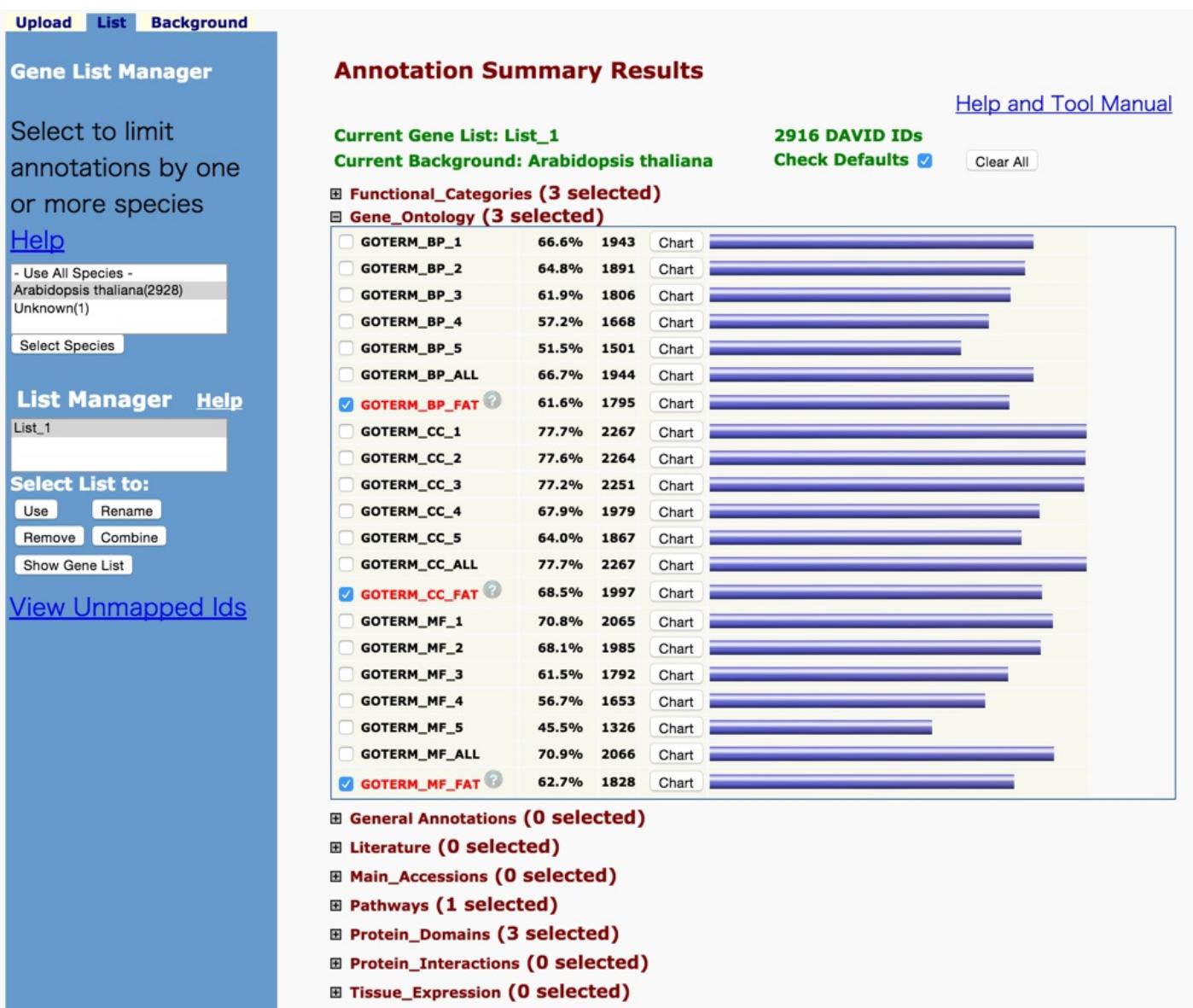
[Which DAVID tools to use?](#)

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。



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

[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	plastid part	RT		515	17.7	9.0E-190	2.9E-187
<input type="checkbox"/>	GOTERM_CC_FAT	chloroplast part	RT		504	17.3	1.1E-187	1.8E-185
<input type="checkbox"/>	GOTERM_CC_FAT	plastid	RT		987	33.8	1.4E-165	1.5E-163
<input type="checkbox"/>	GOTERM_CC_FAT	chloroplast	RT		973	33.4	1.3E-164	1.1E-162
<input type="checkbox"/>	GOTERM_CC_FAT	plastid thylakoid	RT		230	7.9	7.0E-110	4.6E-108
<input type="checkbox"/>	GOTERM_CC_FAT	chloroplast thylakoid	RT		230	7.9	7.0E-110	4.6E-108
<input type="checkbox"/>	GOTERM_CC_FAT	organelle subcompartment	RT		230	7.9	4.5E-109	2.4E-107
<input type="checkbox"/>	GOTERM_CC_FAT	thylakoid	RT		278	9.5	8.0E-109	3.7E-107
<input type="checkbox"/>	GOTERM_CC_FAT	thylakoid part	RT		228	7.8	6.0E-105	2.5E-103
<input type="checkbox"/>	GOTERM_CC_FAT	chloroplast thylakoid membrane	RT		200	6.9	4.4E-100	1.6E-98
<input type="checkbox"/>	GOTERM_CC_FAT	plastid thylakoid membrane	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	cofactor binding	RT	109	3.7	3.8E-8	4.4E-5	
<input type="checkbox"/>	GOTERM_MF_FAT	chlorophyll binding	RT	14	0.5	1.5E-5	8.4E-3	
<input type="checkbox"/>	GOTERM_MF_FAT	vitamin B6 binding	RT	32	1.1	3.4E-5	1.3E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	pyridoxal phosphate binding	RT	32	1.1	3.4E-5	1.3E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	poly(U) RNA binding	RT	11	0.4	4.6E-5	1.3E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	poly-pyrimidine tract binding	RT	11	0.4	4.6E-5	1.3E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	ATP-dependent peptidase activity	RT	12	0.4	8.3E-5	1.9E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	rRNA binding	RT	23	0.8	8.5E-5	1.6E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	coenzyme binding	RT	72	2.5	9.7E-5	1.6E-2	
<input type="checkbox"/>	GOTERM_MF_FAT	vitamin binding	RT	39	1.3	1.2E-4	1.7E-2	

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

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 <https://github.com/AJACS-training/AJACS55/blob/master/hono/answer.md>

【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 <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
- (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。)

- DAVID 以外のツールを使ってみる

- DAVIDでは主にGeneOntologyを見ていましたが、医学・薬学分野に特化した情報を探査対象にしたGeneSetDB <http://genesetdb.auckland.ac.nz/haeremai.html> を使ってみるという手もあります。
- 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 <http://tогotv.dbcls.jp/ja/20160102.html>
- 2:50~ エンリッチメント解析を行う <https://www.youtube.com/watch?v=qqF19PaURsA&feature=youtu.be&t=2m50s>

- 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.
<http://www.ncbi.nlm.nih.gov/pubmed/22637703>
 - 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。
- 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