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

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
2016年11月8日(火) バイオインフォマティクス講義@広尾学園中学校・高等学校

このウェブサイトの短縮URL

http://bit.ly/hiroo2016

これは広尾学園中学校・高等学校 のバイオインフォマティクス講義「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 がありますので、こちらもあわせてご活用ください。

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。

© 2016 DBCLS, Licensed under CC-BY 4.0

- 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

統合TVを知っていますか?	人数	割合
知らない		
聞いたことがある		
知っている		
使ったことがある		
使っている		
回答なし		

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		
公共データを活用する・したい		
既に持っている		
大規模発現解析の予定はない		
回答なし		

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/

© 2016 DBCLS, Licensed under CC-BY 4.0

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
- 自由に使える画像を探す

関連するタグから検索

ゲノム (158) 遺伝子 (276) タンパク質 (122) 配列解析 (150) 発現解析 (214) NGS (133) 文献検索 (131) 情報収集 (74) 環境設定 (101)

DBCLS (177) English (152) ウェブツール (171) ソフトウェア (62) データベース (273) 講演 (469) 実習 (267)

NCBI (51) GEO (17) UCSC (16) EBI (10) Ensembl (19) KEGG (15) PDB (23) DDBJ (41) PDBj (9) NBDC (39) 転写因子 (13) 化合物 (19) メタボローム (18) 画像 (7) 植物 (9) 微生物 (13) メタゲノム (4) 変異解析 (2) CRISPR (5) 多型 (28) 塩基配列 (139) オーソログ (2) アミノ酸 (60) 構造解析 (48) 抗体 (7) プロテオーム (6) アラインメント (32) 系統樹 (15) EMBOSS (21) パスウェイ解析 (49) FANTOM (4) RNA-seq (21) ChIP-seq (16) Reseq (4) CAGE (7) パイプライン (27) ワークフロー (16) 辞書 (27) 特許 (9) 文章執筆 (19) 可視化 (173) 家系図 (2) クリエイティブ・コモンズ (10) 疾患情報 (13) 倫理 (2) 創薬 (4) Galaxy (47) UNIX (25) Linux (12) R (29) Bioconductor (19) プログラミング (36) シェルスクリプト (4) Perl (8) Python (2) スパコン (10) RDF (51) パワーポイント (3) メール (2) 登録方法 (7) Google (11) BioHackathon (114) トーゴーの日 (73) ゲノムリテラシー講座 (18) All-in-one合同講習会 (13) AJACS (330) NGS速習・ハンズオン (52) DDBJing (15)

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 1142 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

Metascapeを使って、遺伝子リストの生物学的解釈をする

マイクロアレイ実験や次世代シーケンシング(NGS)などで得られた大量な遺伝子を扱う際に、どういう機能を持つ遺伝子が変化する傾向にあるかを知るエンリッチメント解析は非常に有用な解析手法です。
Metascapeは、ウェブ上でヒト・マウス・ラットのエンリッチメント解析ができるツールです。リストに含まれる遺伝子にどのような機能を持つものが多いか、クラスタリングなどによって直感的に分析することができます。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で、複数リスト間のアノテーションについて差分を表示することもできます。
本動画ではMetascapeの基本的な使い方について、説明します。

Cytoscape を使ったデータの可視化 @ AJACS京都2

本日の統合TVは、2016年9月1-2日に開催された統合データベース講習会：AJACS京都2から、バイオサイエンスデータベースセンター (NBDC) 櫛田 達矢 による「Cytoscapeを使ったデータの可視化」をお送りします。約1時間22分です。
本講習では、Cytoscapeの特徴、機能について概説したのち、ライフサイエンスデータの可視化方法について、実際の操作方法をハンズオン形式で学びます。 講習会で使用したテキスト・資料はこちらからご覧いただけます。
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データから新たな知識を導出するための高次解析 @ AJACS京都2

本日の統合TVは、2016年9月1-2日に開催された統合データベース講習会：AJACS京都2から、理化学研究所 情報基盤センター バイオインフォマティクス研究開発ユニット 尾崎 遼 基礎科学特別研究員 による「NGSデータから新たな知識を導出するための高次解析」をお送りします。約2時間35分です。
本講習では、自ら取得した(または公共データベースから取得した)NGS低次解析済みデータの高次解析の方法を学ぶことを目的として、NGSデータの形式を理解すること、NGSデータ高次解析の基礎知識を理解すること、自分でコマンドを打って高次解析を体験すること、自身の問題意識に応じて解析の方法を調べられるようになることを目標にしています。
講習会で使用したテキスト・資料はこちらからご覧いただけます。
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。

次世代シーケンスデータベースの使い方 @ AJACS京都2

本日の統合TVは、2016年9月1-2日に開催された統合データベース講習会：AJACS京都2から、ライフサイエンス統合データベースセンター (DBCLS) 大田 達郎 による「次世代シーケンスデータベースの使い方」をお送りします。約1時間35分です。
本講習では、主に新型DNAシーケンサー (High-Throughput Sequencer, HTS) から得られる塩基配列データと、それに基づく生物学的データを公開しているデータベースの概要、およびデータベースからのデータの取得の手順を学びます。
講習会で使用したテキスト・資料はこちらからご覧いただけます。
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。

パスウェイデータベース @ AJACS京都2

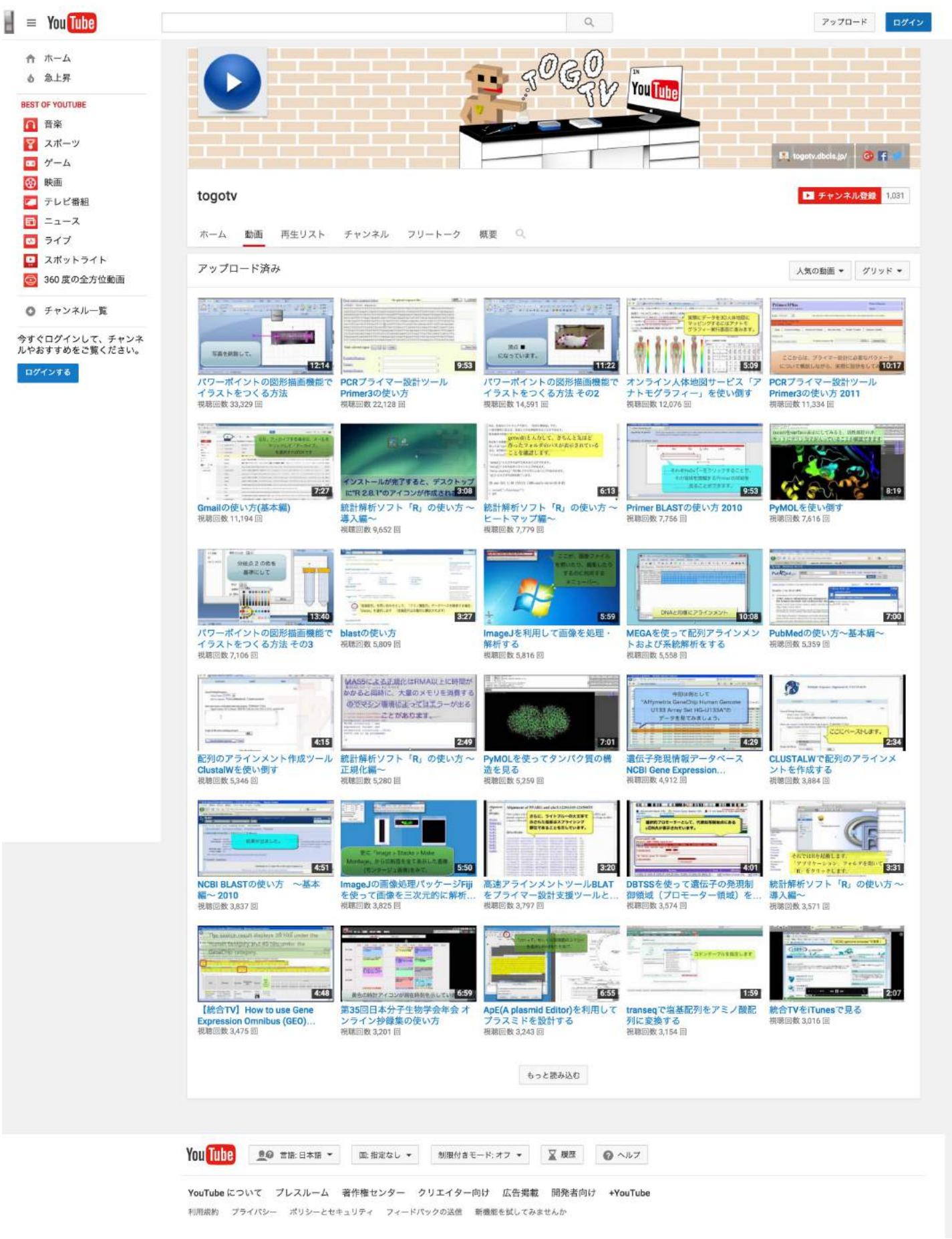
本日の統合TVは、2016年9月1-2日に開催された統合データベース講習会：AJACS京都2から、京都大学 化学研究所 五斗 進 准教授 による「パスウェイデータベース」をお送りします。約1時間30分です。
本講習では、パスウェイデータベースの概要とデータ形式などを紹介するとともに、主にBioCyc、Reactome、KEGGの使って、手持ちのオミックスデータをパスウェイにマッピングする方法について実習形式で紹介しています。
講習会で使用したテキスト・資料はこちらからご覧いただけます。
講習会の一連の動画はYouTubeの再生リストからもご覧いただけます。

first < 1 2 3 4 5 > last

© 2016 DBCLS, Licensed under CC-BY 4.0

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

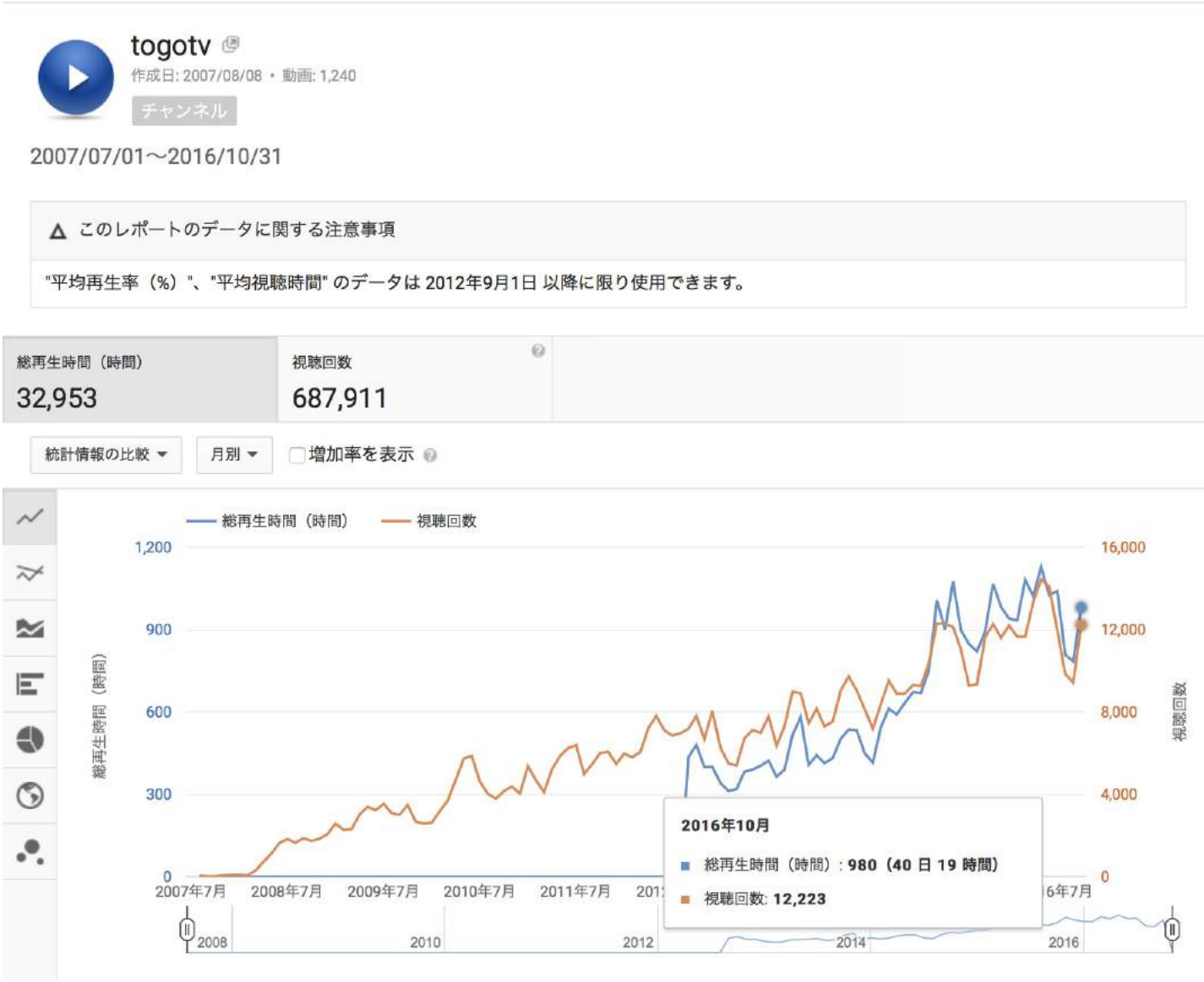
© 2016 DBCLS, Licensed under CC-BY 4.0



- 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。

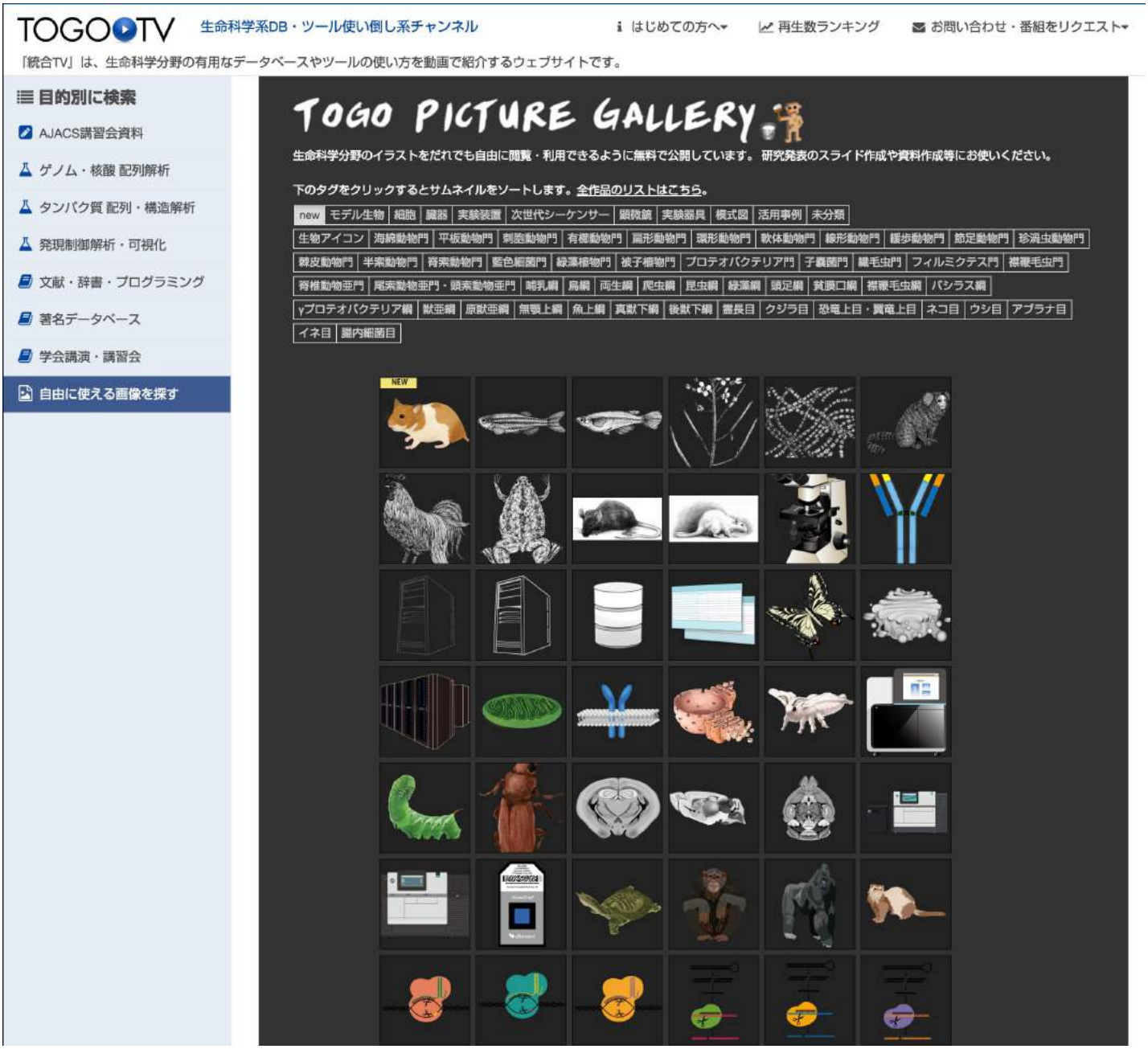
- ウェブサイトへのアクセスの仕方から結果の解釈まで、操作の一挙手一投足がわかります。
  - 1100本を超える動画が公開されており、YouTube版だけで のべ 680,000回以上 再生されています。(2016年10月末現在)

© 2016 DBCLS, Licensed under CC-BY 4.0



- 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「AJACS講習会資料」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。
- 誰でも自由に利用可能なライフサイエンス分野のイラストが、統合TVから閲覧、利用することができるようになりました。「自由に使える画像を探す」
  - Togo picture galleryと生物アイコンの全画像460点を一覧できます。
  - 研究発表のスライド作成や資料作成等に、ぜひお使いください。

© 2016 DBCLS, Licensed under CC-BY 4.0



- 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・ウェブツールとCRISPRガイドRNA設計 @ AJACSこまち」(2016年8月)
- 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)
- 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)

© 2016 DBCLS, Licensed under CC-BY 4.0

- 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（次世代シークエンサ）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 次世代シークエンサ(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)

- 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

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。

© 2016 DBCLS, Licensed under CC-BY 4.0

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。
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: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。

© 2016 DBCLS, Licensed under CC-BY 4.0

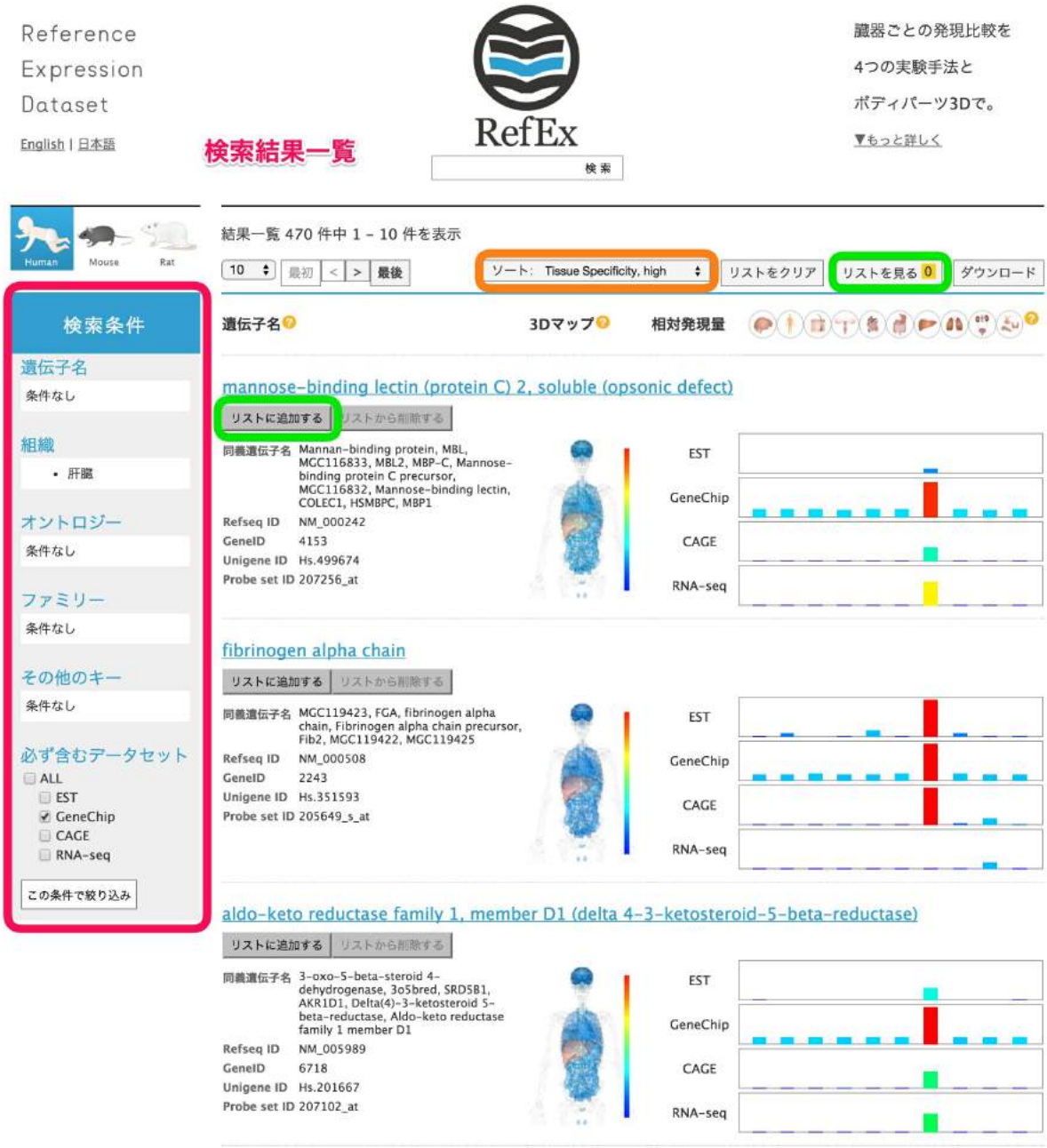
10. 右側のFANTOM5 CAGEのタブをクリックすると、FANTOM5 CAGEデータのビューアに切り替わります。

- 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。

© 2016 DBCLS, Licensed under CC-BY 4.0

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。

© 2016 DBCLS, Licensed under CC-BY 4.0



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。

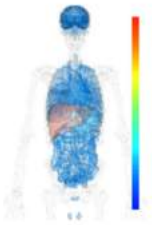
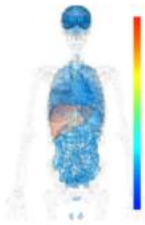
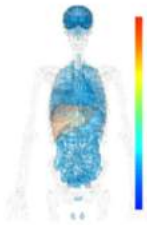
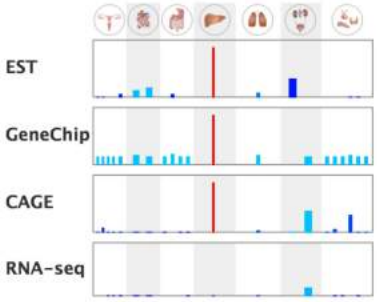
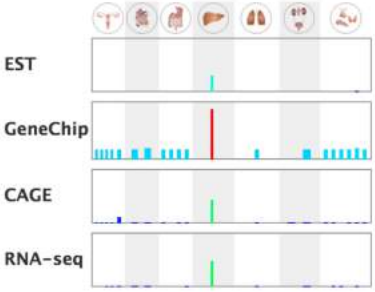
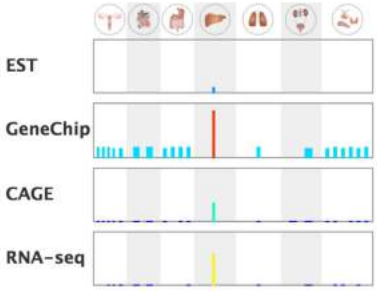
- 遺伝子名に「liver」を含むデータは何件あるでしょうか。
- 「遺伝子名」の下の「条件なし」をクリックして表示されるウインドウに「liver」と入力し、「Include」をクリックし、「この条件で絞り込み」を押します。
- 「遺伝子名」の項目で「Exclude」に「solute」を加えると、検索結果はどう変わるでしょうか。
- 「組織」の項目で、データ元をRNA-seqに変更したり、臓器の指定を追加すると検索結果はどう変わるでしょうか。
- 「必ず含むデータセット」の「ALL」にチェックを入れると、検索結果はどう変わるでしょうか。

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。

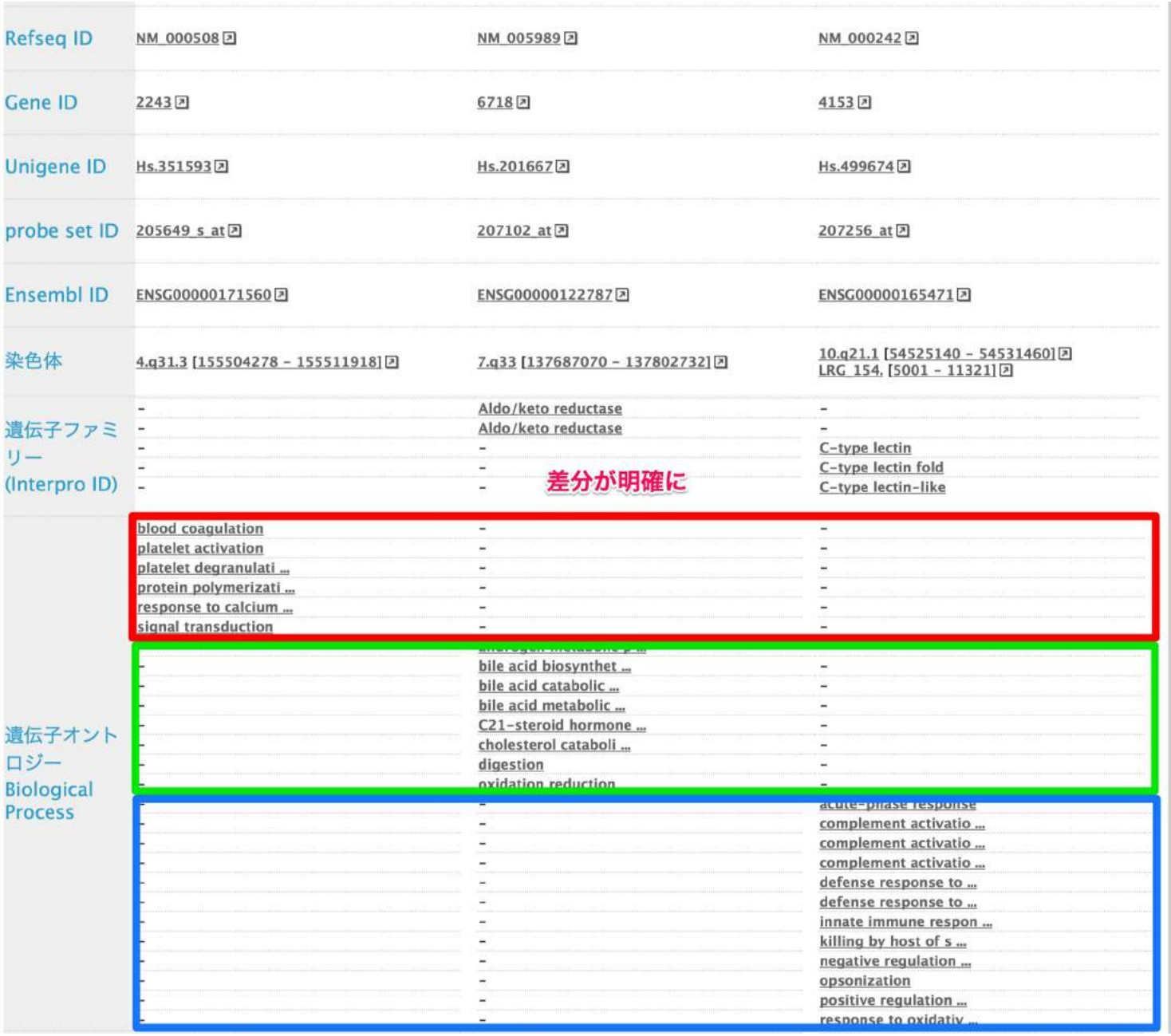
- 肝臓特異的遺伝子の検索結果一覧に移動して、3つの遺伝子を「リストに追加」してみましょう。
- 追加した件数は「リストを見る」の横に表示されます。
- 「リストを見る」をクリックするとリストに移動します。
- 『並べて表示する』にチェックを入れて、「遺伝子を並べて表示」をクリックします。
- 遺伝子発現データやGeneOntology情報を並列に比較することで見えてくる「違い」はなんでしょうか。その違いからどういうことが推測できるでしょうか。

© 2016 DBCLS, Licensed under CC-BY 4.0

遺伝子を並べて表示 Close

Gene Name	fibrinogen alpha chain	aldo-keto reductase family 1, member D1 (delta 4-3-ketosteroid-5-beta-reductase)	mannose-binding lectin (protein C) 2, soluble (opsonic defect)
同義遺伝子名	MGC119423, FGA, fibrinogen alpha chain, Fibrinogen alpha chain precursor, Fib2, MGC119422, MGC119425	3-oxo-5-beta-steroid 4-dehydrogenase, 3o5bred, SRD5B1, AKR1D1, Delta(4)-3-ketosteroid 5-beta-reductase, Aldo-keto reductase family 1 member D1	Mannan-binding protein, MBL, MGC116833, MBL2, MBP-C, Mannose-binding protein C precursor, MGC116832, Mannose-binding lectin, COLEC1, HSMBPC, MBP1
発現マップ on BodyParts3D			
相対発現量			
			

© 2016 DBCLS, Licensed under CC-BY 4.0



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アント(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。

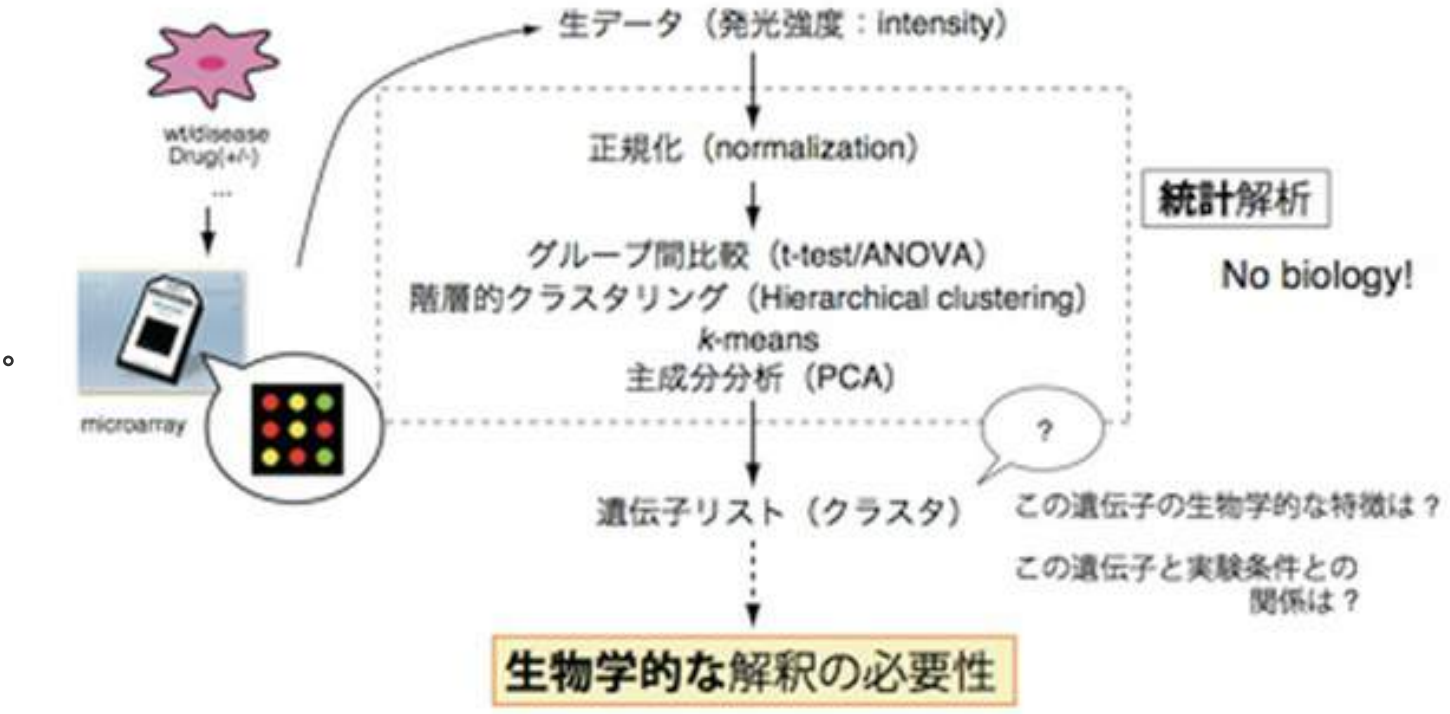
© 2016 DBCLS, Licensed under CC-BY 4.0

【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

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。

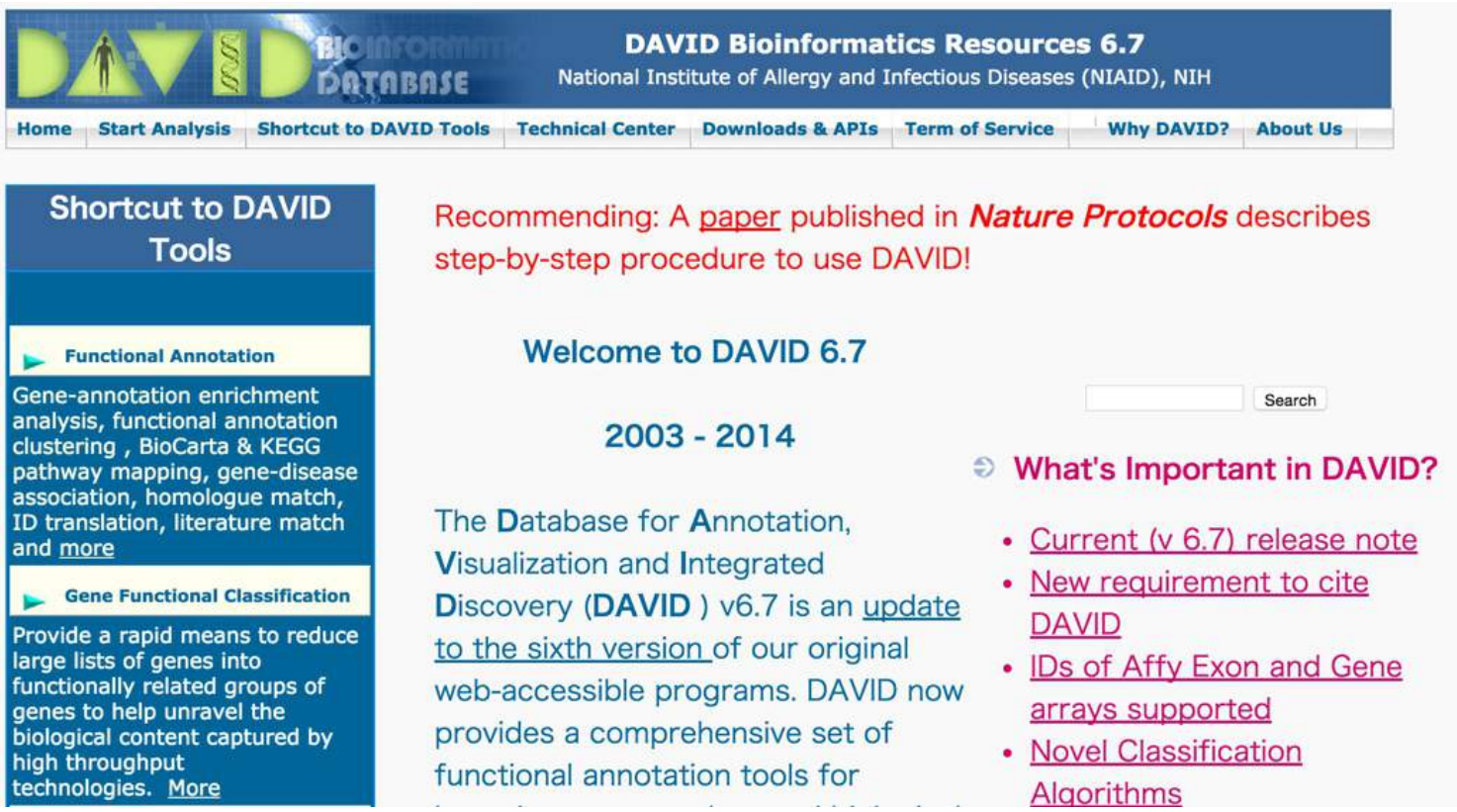
  → マル秘遺伝子リスト （右クリックして「新しいタブで開く」もしくは「名前を付けてリンク先を保存」してください。）

- 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！

© 2016 DBCLS, Licensed under CC-BY 4.0

【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

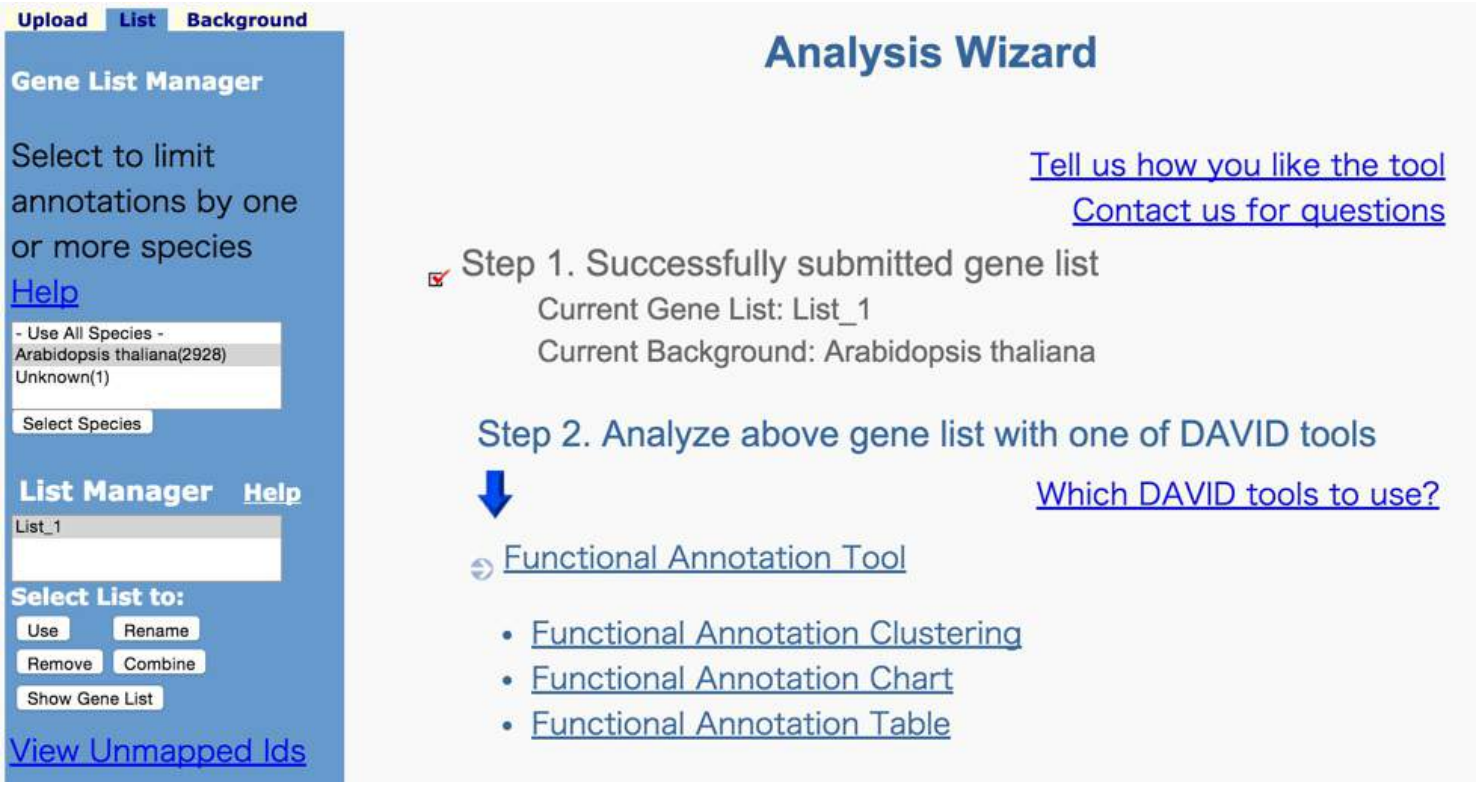
- 【復習用】DAVIDを使ってマイクロアレイデータを解析する 2012
- 【復習用】DAVIDの使い方 実践編

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


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。

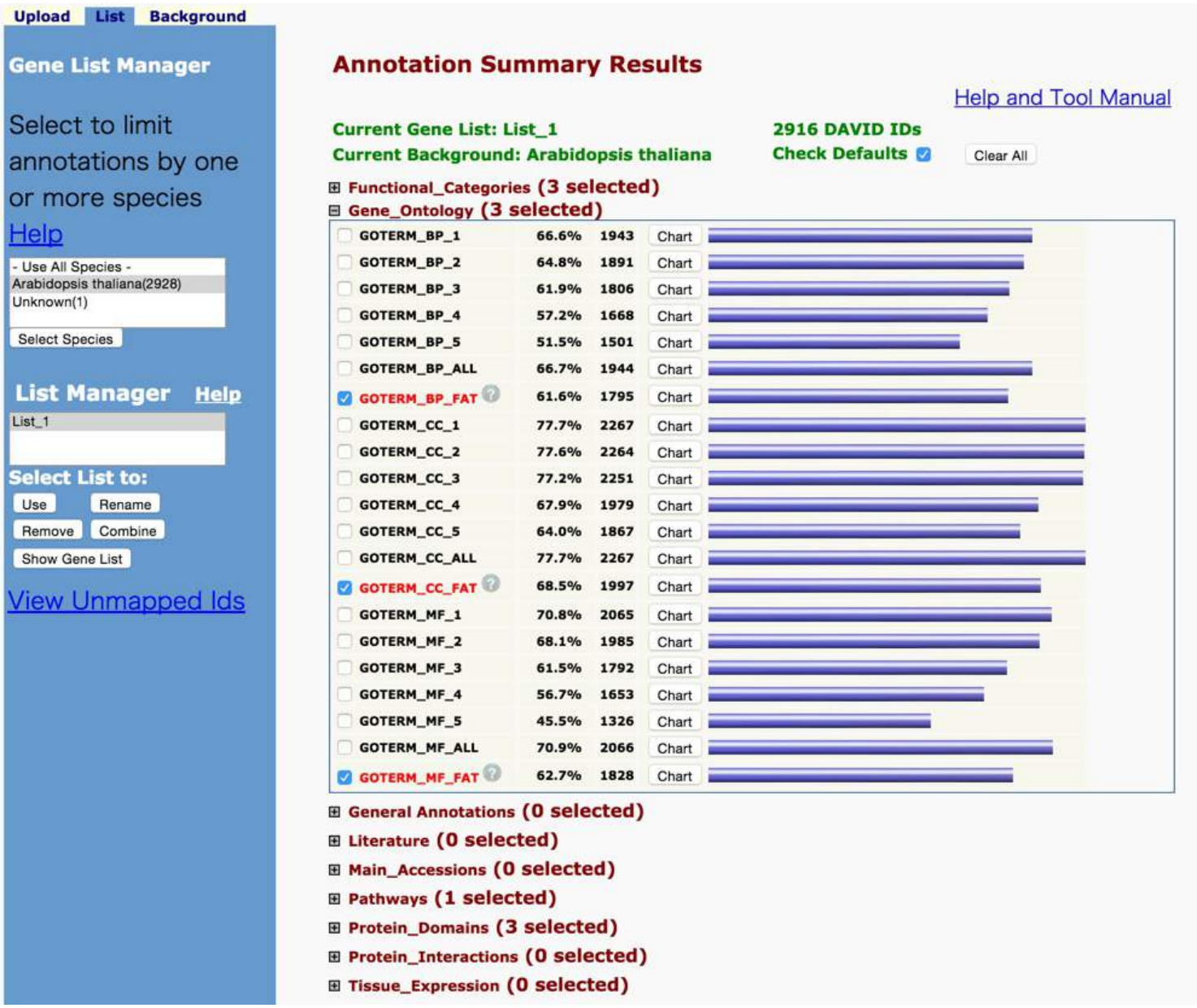
© 2016 DBCLS, Licensed under CC-BY 4.0

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。

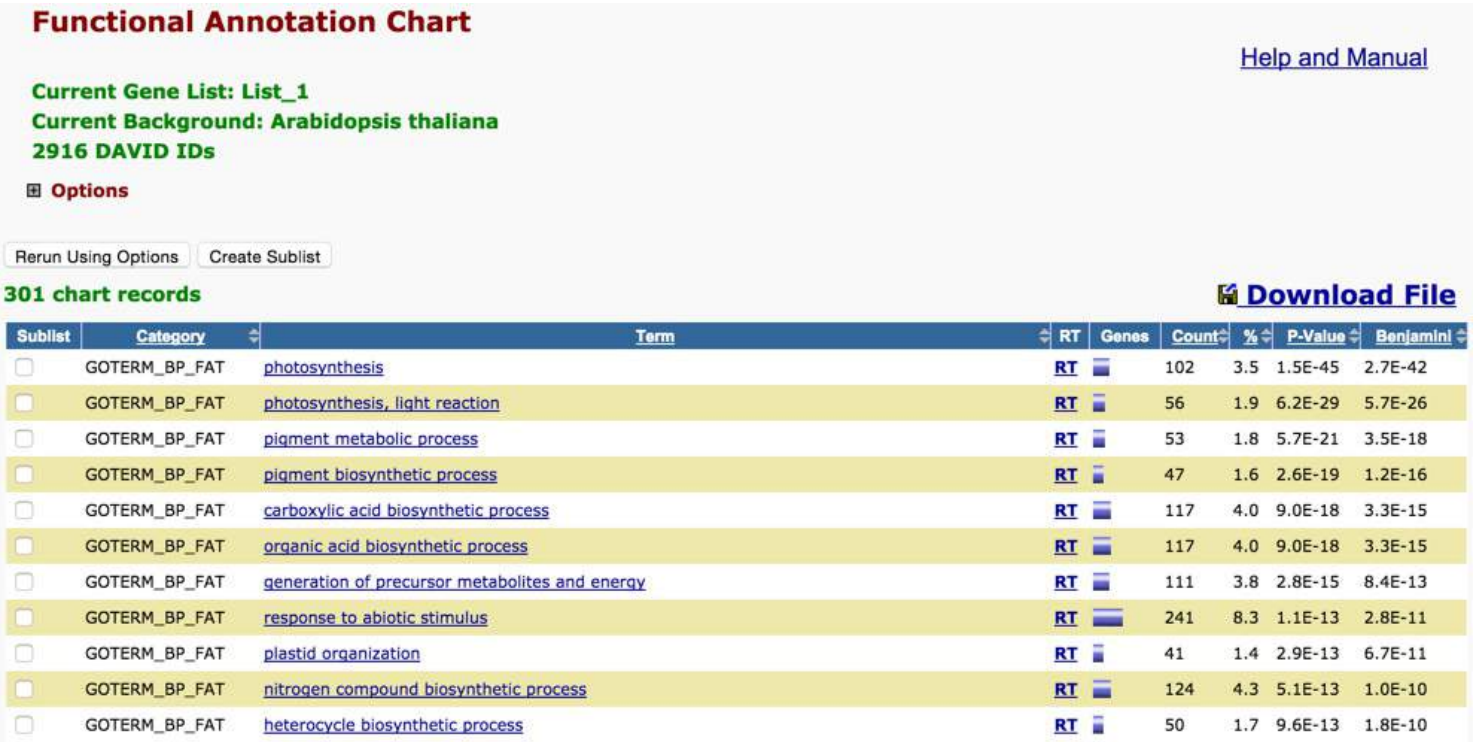


6. 解析を続けます。真ん中の「Functional Annotation Tool」をクリックします。
7. 「Gene Ontology」をクリックすると、Gene Ontologyを用いた解析の細かいメニューが表示されます。

© 2016 DBCLS, Licensed under CC-BY 4.0



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。



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

© 2016 DBCLS, Licensed under CC-BY 4.0

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

© 2016 DBCLS, Licensed under CC-BY 4.0

【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
      - まだシステムが不安定(?)で大量クエリ投げると結果が帰ってこない場合もある
      - Chromeだとjavascript周りでエラーが出て使えない(?)こともある

  - 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

- つまみ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。

© 2016 DBCLS, Licensed under CC-BY 4.0

- 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

© 2016 DBCLS, Licensed under CC-BY 4.0