

# AJACS番町1 実験データの生物学的解釈を助ける遺伝子発現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)

2018年8月29日(水) AJACS番町1 @ JST東京本部（サイエンスプラザ）B1大会議室

これは統合データベース講習会 AJACS番町1「実験データの生物学的解釈を助ける遺伝子発現DB・ウェブツールの使い方」の講習資料です。

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
  - ChIP-Atlas
    - 【実習2】ChIP-Atlasのin silico ChIPを使って、興味ある遺伝子リストを制御する可能性の高い転写因子を調べる

## 講義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。
- こんなことは知ってて当たり前だと他の人に思われる質問を歓迎します。質問することのハードルを下げます。
  - 知っている人は講師を助けてください。サポート大歓迎です。
  - あなたが疑問に思ったことは、実は、隣の人やその隣の人もそう思っていることが多いです。
  - 当たり前に感じる質問や一見関係なさそうな質問がでると、「そういう質問をしてもよいのだ」という空気になり、この講義から得られる情報が増え、皆さんの受講満足度が上がります(たぶん)。
  - でも講師も知らないことは(多々)あります。(以下ループ)
- 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。
  - 講師用
  - [受講者質問用フォーム](#)(右クリックから「新しいタブで開く」推奨)

---

#### 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 87)

統合TVを知っていますか?	人数	割合
知らない	25名	29 %
聞いたことがある	8名	9 %
知っている	15名	17 %
使ったことがある	22名	25 %
使っている	13名	15 %
回答なし	4名	5 %

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	16名	18 %
公共データを活用する・したい	28名	32 %
既に持っている	21名	24 %
大規模発現解析の予定はない	15名	17 %
回答なし	7名	8 %

# 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/>

The screenshot shows the TOGO TV homepage with a search bar at the top. Below it, a search result for 'depmap' is displayed. The result title is 'depmapを使ってがん細胞が依存する遺伝子の情報を調べる (がんの治療標的となる遺伝子を発見する)' (Discover genes that depend on cancer cells using depmap). To the right of the main content, there are two small video thumbnail previews: one for 'depmapを使ってがん細胞が依存する遺伝子の情報を調べる (がんの治療標的となる遺伝子を発見する)' and another for 'Google スライドの图形描画機能で科学イラストを作る'.

TOGO TV 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 再生数ランキング お問い合わせ・番組をリクエスト▼

『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

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 1443 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。番組リクエストやお問い合わせは[こちらからどうぞ!](#)

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

depmapを使ってがん細胞が依存する遺伝子の情報を調べる (がんの治療標的となる遺伝子を発見する)

depmap (the Cancer Dependency Map) は、米国ブロード研究所 (Broad Institute)と英国サンガー研究所 (Wellcome Sanger Institute)が中心となって提供する、がん細胞の持っている遺伝子変異や発現量変化などの特徴と、がんが生存し成長するために依存している遺伝子の情報を関連付け、治療の標的を見つけるのを助け治療法の開発を促すことを目的としたデータベースです (原著論文: "Defining a Cancer Dependency Map" DOI: 10.1016/j.cell.2017.06.010)。500 以上の細胞株について、遺伝子依存やオミクスデータ、薬剤感受性のデータがまとめられています。

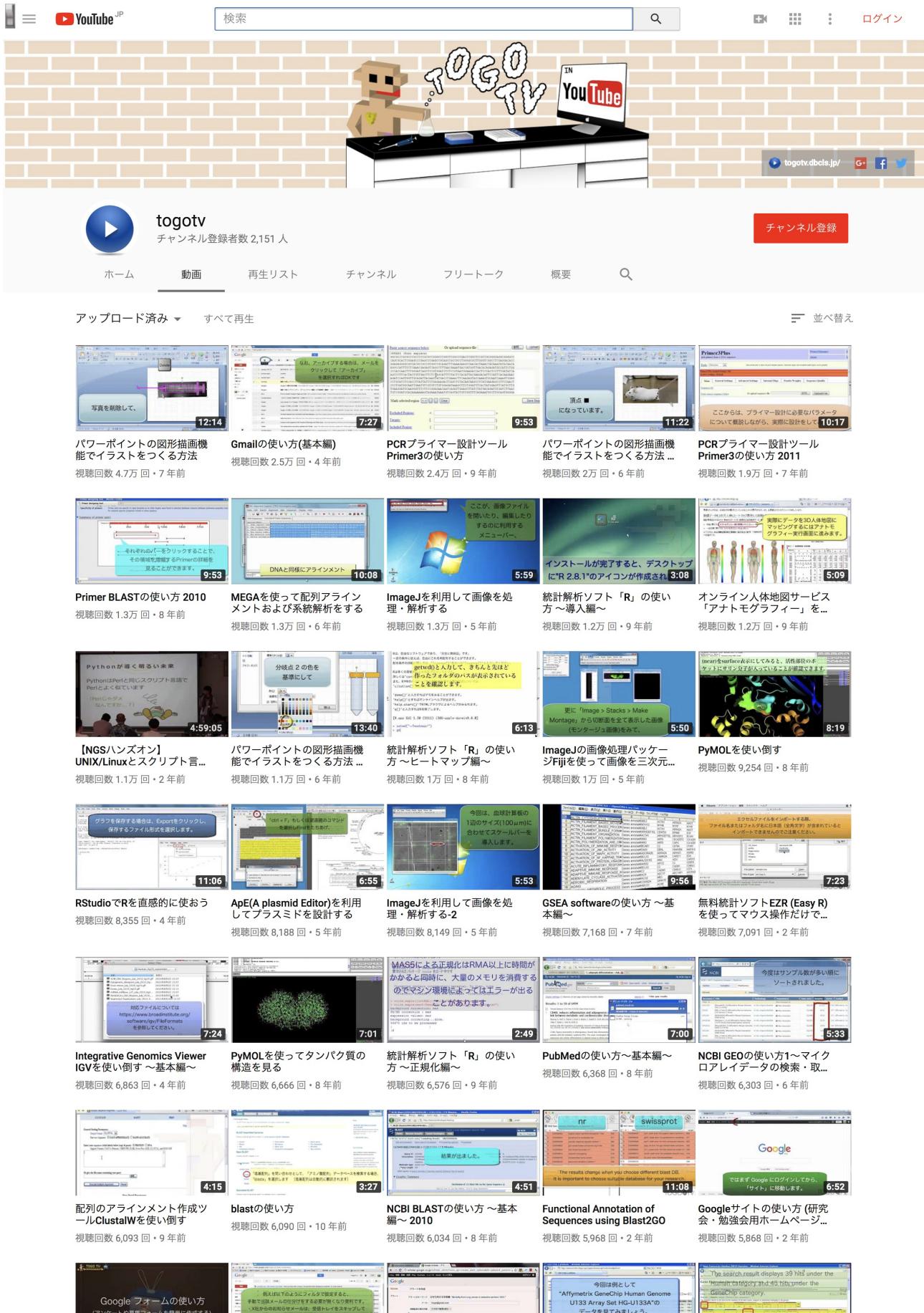
今回は、がん原遺伝子として知られているBRAF (B-Raf proto-oncogene, serine/threonine kinase)を例に、depmapの基本的な使い方からdepmapに収載されているさまざまなデータの見方について紹介します。

Google スライドの图形描画機能で科学イラストを作る 

Googleスライド (Google Slide)は、Google 社が提供するオンライン プレゼンテーション アプリケーションです。Googleアカウントがあれば無料で利用でき、他のユーザーとの共同編集が簡単にできることや、データの自動保存機能などWebアプリケーションならではの特長があります。パワーポイント (Microsoft PowerPoint)との互換性もあり、読み込み・書き出しに対応しています。

Googleスライドにも图形描画機能（オートシェイプ）があり、パワーポイント同様に簡単なマウス操作で图形を作成することができます。色や形も自由に設定することができます。今回はパワーポイントの图形描画機能で科学イラストを作る～遠沈管編～2018でも取り上げた遠沈管のイラスト作成を例にGoogleスライドの图形描画機能の使い方やイラスト作成におけるtipsを紹介します。

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



- 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。

- 1400本を超える動画が公開されており、YouTube版だけで のべ 1,000,000回以上 再生されています。(2018年5月末現在)

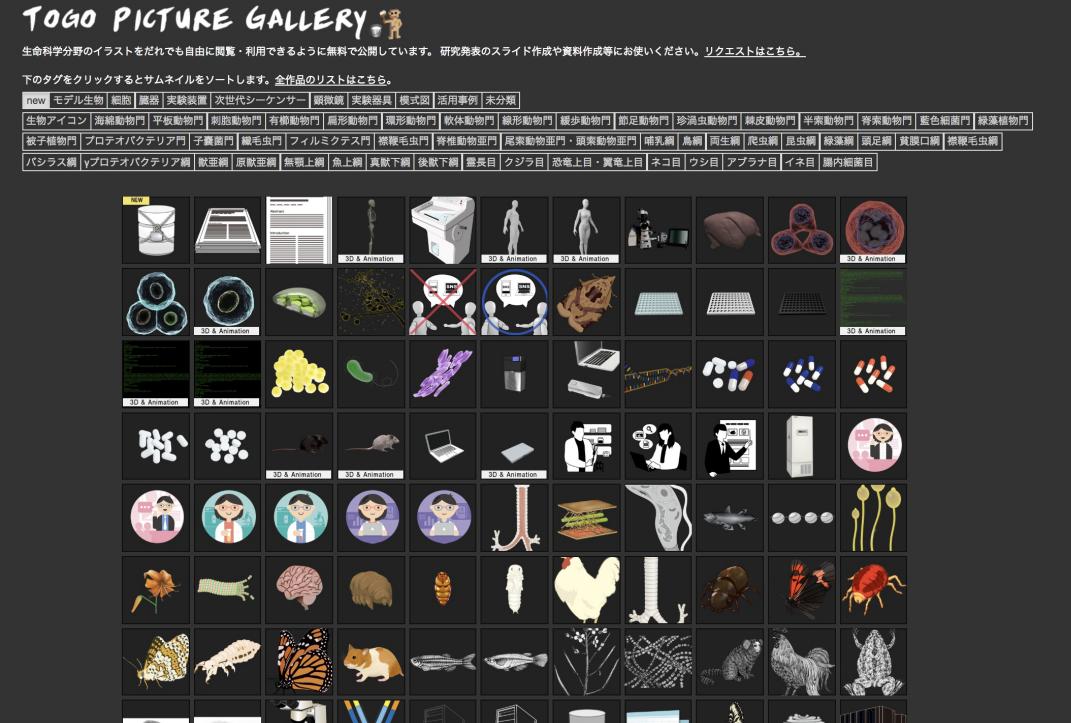


- 講義・講習などの参考資料や後輩指導の教材として利用できます。
  - 本講義中、本家サイトが繋がらない時は、統合TVを見ればおおよその内容がわかるようになっています。
  - DBCLSの提供する便利な各種サービスをレビューする内容もあります。
    - [【NGSハンズオン2017】NBDC・DBCLSの各種サービス 今日から使える便利な生命科学系公共データベース in DBCLS](#)
- 統合TVに掲載されているコンテンツについてご引用いただく際に、恒久的な URL として DOI (Digital Object Identifier) を使用することができます。
- 2014年8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ます、受講生の復習のみならず、初学者の学習教材として活用できます。
- 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500点以上を一覧できます。
  - 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。

■ 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



- 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を作成するための方、募集中です。(オンラインで完結する作成環境を整備しており、遠隔地でもOKです。謝金あり。)

習熟度ややりたいこと別にご参考ください

- 塩基配列解析に関わる基礎知識について
  - 「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
  - 【NGS】に関する動画・講習会資料・新着論文レビュー
- 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)
  - 「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)」
  - NGSハンズオン講習会2016
  - NGSハンズオン講習会2017
  - 上記の動画+講習会資料のまとめページ@統合TV
- 疾患と多型、変異に関するデータベース (COSMIC, TCGA 等)

- 【多型】に関する動画・講習会資料・新着論文レビュー

## 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/>

The screenshot shows the homepage of the RefEx website. At the top, there is a navigation bar with icons for Human, Mouse, and Rat. Below the navigation bar, there is a search bar with the placeholder "キーワードで検索" (Search by keyword) and a yellow "検索" (Search) button. To the right of the search bar, there is a link "ex) troponin, ALB".

On the left side, there is a sidebar with the following sections:

- Reference Expression Dataset**: Includes links for English and Japanese.
- Human**, **Mouse**, **Rat** icons.
- 遺伝子オントロジー Gene Ontology**: Includes lists of cellular process, biological regulation, metabolic process, multicellular organismal process, response to stimulus, and developmental process.
- 遺伝子ファミリー InterPro**: Includes lists of RNA recognition motif, RNP-1, Pleckstrin homology, Krueppel-associated box, Protein kinase-like domain, Zinc finger, C2H2-like, and GPCR, rhodopsin-like superfamily.
- 染色体**: Includes a "染色体領域を選ぶ" (Select chromosome region) button.

In the center, there is a section titled "組織特異的に発現する遺伝子を見る" (View genes expressed specifically in tissues) with icons representing various organs: brain, heart, muscle, uterus, liver, intestines, lungs, kidneys, and skin.

On the right side, there is a message: "RefExの論文が出版されました。あなたの研究に役立ったらぜひ引用を!!" (The RefEx paper has been published. Please cite it if it is useful for your research!) and a link "▼もっと詳しく" (More details).





- 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](#)
- 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 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 @ 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。
- その他RefExを引用した論文の一覧はこちらでご覧いただけます。
    - <https://dbcls.rois.ac.jp/references.html#RefEx>

## 参考文献

- Hiromasa Ono, Osamu Ogasawara, Kosaku Okubo, Hidemasa Bono **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)
- 川路 英哉、粕川 雄也、坊農 秀雅、小野 浩雅 「FANTOM5データを誰でも活用できる形に」 Scientific Data誌著者インタビュー(平成29年8月29日) <https://www.natureasia.com/ja-jp/scientificdata/papers-from-japan/fantom5>
- 小野 浩雅・坊農 秀雅 「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RefEx」 ライフサイエンス新着論文レビュー(平成29年9月5日) DOI: [10.7875/first.author.2017.093](https://doi.org/10.7875/first.author.2017.093)
- 統合TV 「RefExの使い方」 DOI: [10.7875/togotv.2014.009](https://doi.org/10.7875/togotv.2014.009)

---

## 【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。



Human Mouse Rat 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

他のオントロジーを選ぶ

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

他のファミリーを選ぶ

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

染色体 染色体領域を選ぶ

Advanced Search Advanced Search ページ上部に戻る

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。



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。



Human



Mouse



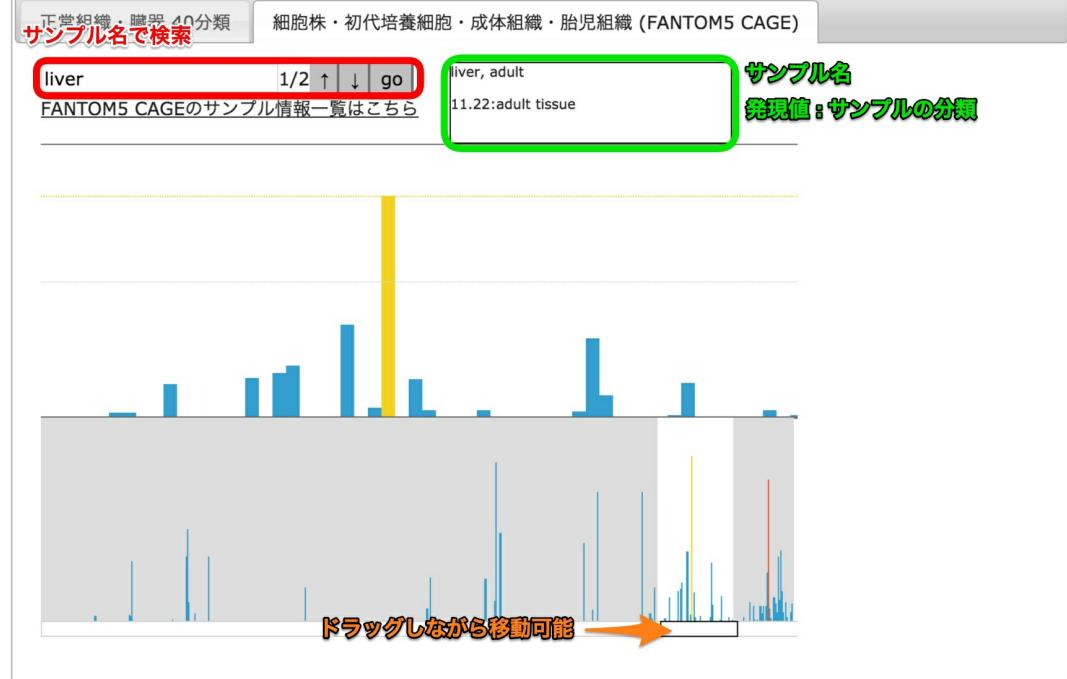
Rat

## 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。

## 検索結果一覧

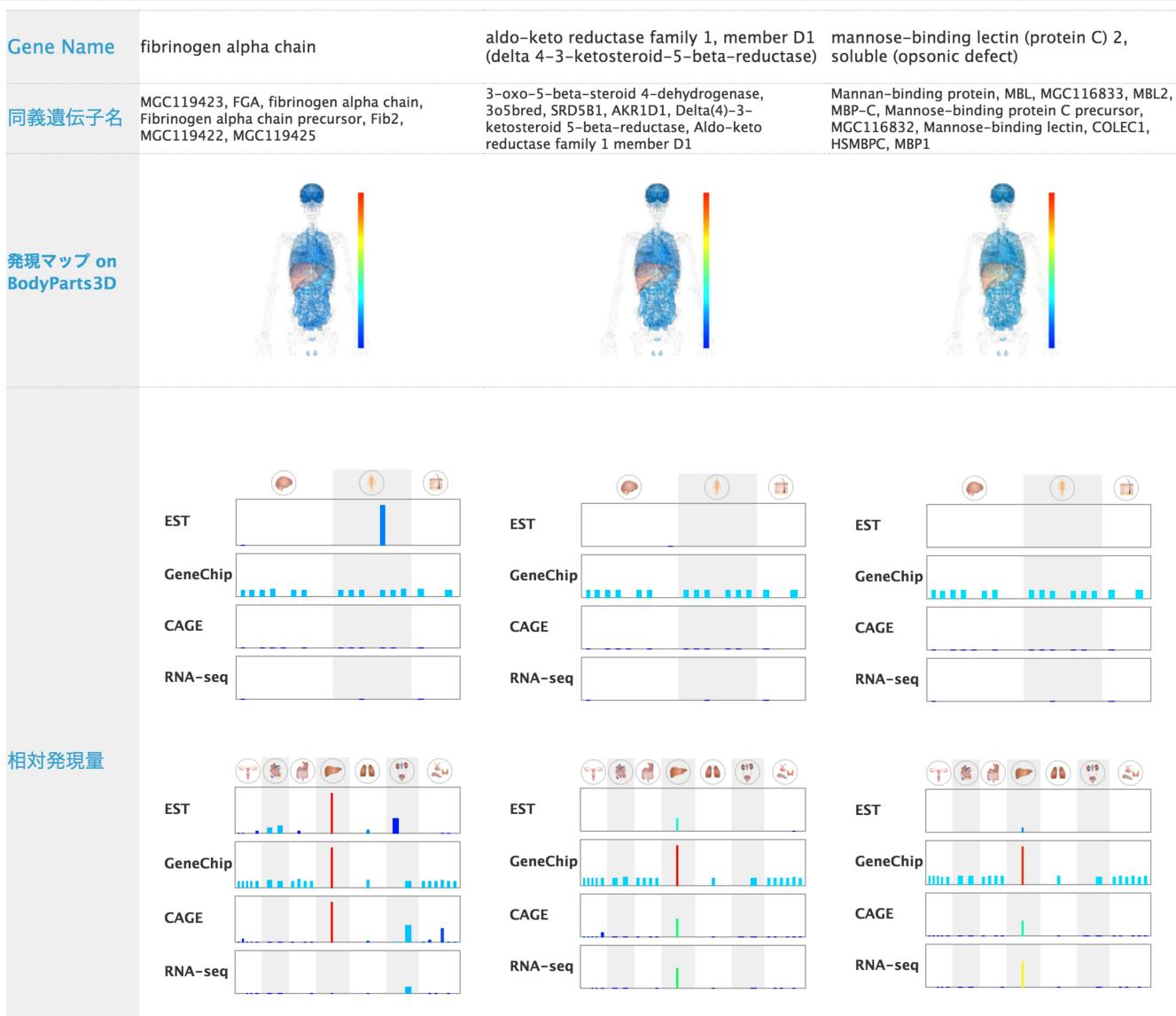


12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- 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。

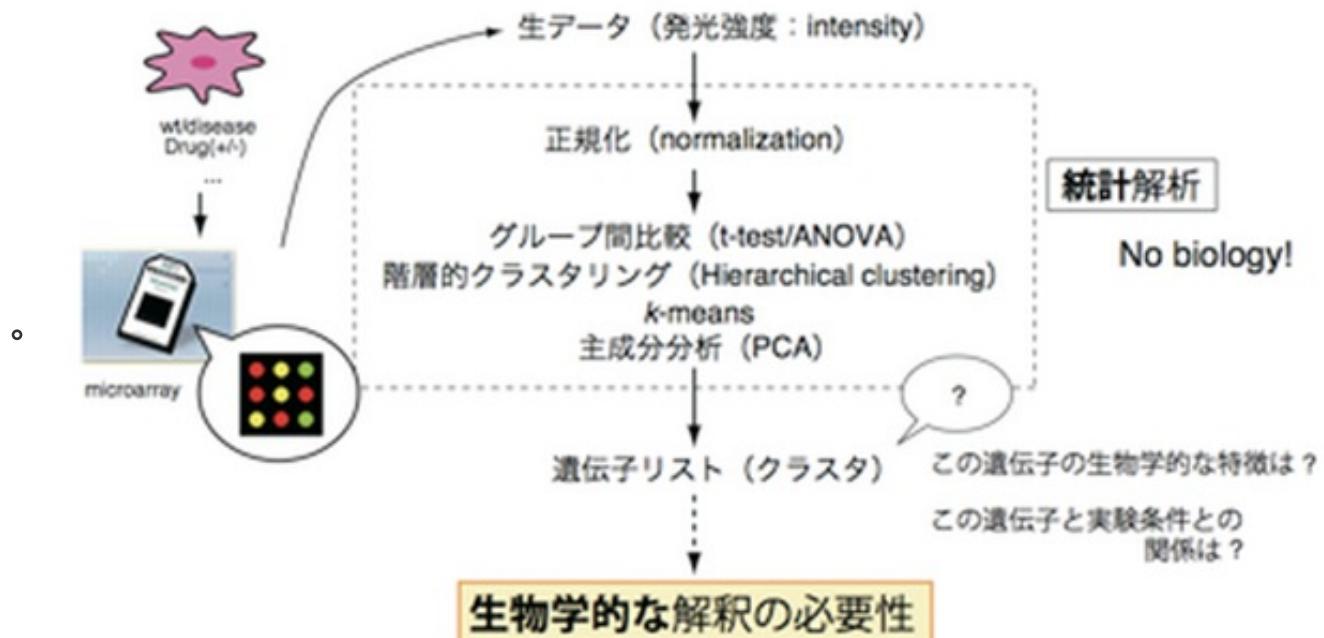


Refseq ID	<a href="#">NM_000508</a>	<a href="#">NM_005989</a>	<a href="#">NM_000242</a>
Gene ID	<a href="#">2243</a>	<a href="#">6718</a>	<a href="#">4153</a>
Unigene ID	<a href="#">Hs.351593</a>	<a href="#">Hs.201667</a>	<a href="#">Hs.499674</a>
probe set ID	<a href="#">205649_s_at</a>	<a href="#">207102_at</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)	- - - - - - <a href="#">Aldo/keto reductase</a> <a href="#">Aldo/keto reductase</a>	- - - - - - <a href="#">差分が明確に</a>	- - <a href="#">C-type lectin</a> <a href="#">C-type lectin fold</a> <a href="#">C-type lectin-like</a>
遺伝子オントロジー Biological Process	<a href="#">blood coagulation</a> <a href="#">platelet activation</a> <a href="#">platelet degranulati ...</a> <a href="#">protein polymerizati ...</a> <a href="#">response to calcium ...</a> <a href="#">signal transduction</a>	- - - - - -	- - - - - -
		<a href="#">bile acid biosynthet ...</a> <a href="#">bile acid catabolic ...</a> <a href="#">bile acid metabolic ...</a> <a href="#">C21-steroid hormone ...</a> <a href="#">cholesterol cataboli ...</a> <a href="#">digestion</a> <a href="#">oxidation reduction</a>	- - - - - - -
			<a href="#">acute-phase response</a> <a href="#">complement activatio ...</a> <a href="#">complement activatio ...</a> <a href="#">complement activatio ...</a> <a href="#">defense response to ...</a> <a href="#">defense response to ...</a> <a href="#">innate immune respon ...</a> <a href="#">killing by host's ...</a> <a href="#">negative regulation ...</a> <a href="#">opsonization</a> <a href="#">positive regulation ...</a> <a href="#">response to oxidativ ...</a>

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。

## 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。



## ChIP-Atlas

ChIP-Atlasは、論文などで報告された ChIP-seq データを閲覧し、利活用するためのウェブサービスです。データ処理の知識やスキルがない方でも簡単に利用できます。データソースは、公開 NGS データレポジトリ (NCBI, EMBL-EBI, DDBJ) に登録されたほぼ全ての ChIP-seq データです。ChIP-Atlas は、九州大学大学院医学研究院 発生再生学分野 (<http://www.dev.med.kyushu-u.ac.jp>) と DBCLS が共同で開発しています。

(<http://chip-atlas.org/>)

ChIP-Atlas Peak Browser Target Genes Colocalization *in silico* ChIP Documentation Publications

Find an experiment ▾

# ChIP-Atlas

ChIP-Atlas is an integrative and comprehensive database for visualizing and making use of public ChIP-seq data. ChIP-Atlas covers almost all public ChIP-seq data submitted to the SRA (Sequence Read Archives) in NCBI, DDBJ, or ENA, and is based on over 69,000 experiments.

[Watch movie introduction](#)

The four main features of ChIP-Atlas are:

### Peak Browser

graphically visualizes protein binding on given genomic loci with genome browser (IGV).

[Watch Movie](#)

### Target Genes

predicts target genes bound by given transcription factors.

[Watch Movie](#)

### Colocalization

predicts partner proteins colocalizing with given transcription factors.

[Watch Movie](#)

### *in silico* ChIP

predicts proteins bound to given genomic loci and genes.

[Watch Movie](#)

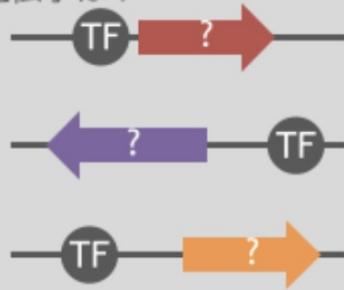
## ① Peak Browser

何がどこに結合？



## ② Target Genes

標的遺伝子は？



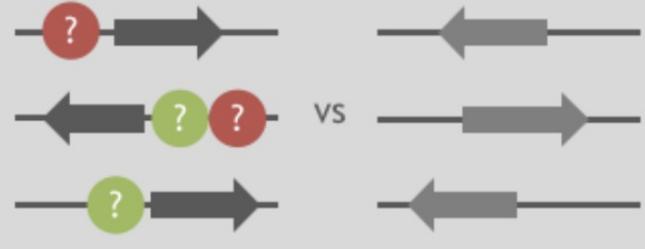
## ③ Colocalization

共局在パートナーは？



## ④ in silico ChIP

Enrichment 解析など



## ChIP-Atlasの機能

### Peak Browser

- 既報の ChIP-seq データをまとめて閲覧し、何がどこに結合しているかが一目でわかります。Integrative Genomics Viewer (IGV) によりスムーズなブラウジングが可能で、興味の遺伝子のシス調節領域を予測したり、それを制御する転写因子の予測ができます。
  - [ChIP-Atlasを使って既報のChIP-seqデータをまとめて閲覧する ~Peak Browserの使い方~](#)

### Target Genes

- 興味のある転写因子を選択し、その標的遺伝子候補を検索できます。
  - 統合TV 「[ChIP-Atlasを使って興味のある転写因子を選択しその標的遺伝子候補を検索する ~Target Genes の使い方~](#)」

### Colocalization

- 興味のある転写因子を選択し、それとゲノム上で共局在する転写因子候補を検索できます。
  - 統合TV 「[ChIP-Atlasを使って共局在タンパク質を探す ~Colocalizationの使い方~](#)」

### in silico ChIP

- ユーザデータを受け付け、既存データとの比較解析をおこないます。たとえば、興味のある遺伝子リストを submit すると、それらをまとめて制御する転写因子候補が返されます。ほかにも BED 形式のファイルや、シークエンスマチーフを submit すると、それらに enrichment する転写因子群が返されます。

## 利用例

- 論文として発表された ChIP-Seq データを閲覧したい

- 興味のあるゲノム領域における、転写因子や修飾ヒストンの分布を知りたい
- 興味のある転写因子の下流遺伝子や、複合体形成パートナーを知りたい 自身の研究データと公開 ChIP-seq データを用いて比較解析をおこないたい

## 参考文献

- Source code and documentation
  - <https://github.com/inutano/chip-atlas>
- Preprint
  - Shinya Oki, Tazro Ohta, et al. Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions. bioRxiv 262899; doi: <https://doi.org/10.1101/262899>
- Database
  - Oki, S; Ohta, T (2015): ChIP-Atlas. <http://dx.doi.org/10.18908/lMDBc01558-000>
- Publications citing ChIP-Atlas <http://chip-atlas.org/publications>

## 【実習2】 ChIP-Atlasのin silico ChIP を使って、興味ある遺伝子リストを制御する可能性の高い転写因子を調べる

- 「発現差のあった遺伝子リスト」を持っている想定で、それらの遺伝子に結合しうる、あるいは上流でそれらの遺伝子の発現を制御する可能性がある転写因子を検索する
  - 使用するデータ
    - [180627\\_List\\_of\\_GeneSymbol\\_txt](#)
      - ある「興味ある遺伝子リスト」をGeneSymbolにID変換したデータ。
      - これを使って、もともとどういう遺伝子リストだったかを考察します。
    - ChIP-Atlas では、遺伝子IDとしてGeneSymbolのみを受け付けているので、それ以外のIDで遺伝子リストを持っている場合は、適宜変換が必要です。
      - ID変換はいろいろなツールがあるが、今回は[HGNC BioMart](#)を利用する。
        - HGNC(The HUGO Gene Nomenclature Committee)はヒトのGeneSymbolを認定・管理している機関。
        - [DAVID\(Database for Annotation, Visualization and Integrated Discovery\)](#) のGene ID Conversion Toolも便利。(使い方動画)
1. [ChIP-Atlas - in silico ChIP](#)にアクセスする
  2. 下図のようにオプションを設定する

# ChIP-Atlas - *in silico* ChIP

[Tutorial movie ▾](#)

Analyze your data with public ChIP-seq data.

[H. sapiens](#)
[M. musculus](#)
[R. norvegicus](#)
[D. melanogaster](#)
[C. elegans](#)
[S. cerevisiae](#)

## 1. Antigen Class

All antigens (31081)  
 DNase-seq (1440)  
 Histone (7674)  
 RNA polymerase (1011)  
**TFs and others (7353)**  
 Input control (3388)  
 Unclassified (6329)  
 No description (3886)

## 2. Cell type Class

All cell types (31081)  
 Adipocyte (275)  
 Blood (8600)  
 Bone (466)  
 Breast (2889)  
 Cardiovascular (785)  
 Digestive tract (2349)  
 Epidermis (964)

## 3. Threshold for Significance

50  
**100**  
 200  
 500

## 4. Select your data

- Genomic regions (BED) or sequence motif [ⓘ](#)  
 Gene list (Gene symbols) [ⓘ](#)

F2  
 CFB  
 FGA  
 Mpst  
 C8G  
 CSAD  
 Cyp26a1  
 F11r

ファイルを選択 | 180612\_List\_...eSymbol.txt  
 Choose local file [Try with example](#)

## 5. Select dataset to be compared

- Refseq coding genes (excluding user data) [ⓘ](#)  
 Gene list (Gene symbols) [ⓘ](#)

## 6. Describe datasets

User data title [ⓘ](#)

My data

Compared data title [ⓘ](#)

Control

Project title [ⓘ](#)

My project

Distance range from TSS [ⓘ](#)

- 5000 bp ≤ TSS ≤ + 5000 bp

**submit**

Estimated run time: 4 mins

3. submit すると遺伝研スパコンヘクエリが飛ぶ(ので、講義中は見てるだけにしてください)

4. submit したあとの画面

# ChIP-Atlas - *in silico* ChIP

Analyze your data with public ChIP-seq data.

Result page URL will be available for a week from the time when 'status' is 'finished'.

Project title	My project
Request ID	wabi_chipatlas_2018-0606-1701-36-865-010614
Submitted at:	17:01:37 (Jun-06-2018)
Estimated finishing time:	17:05:37 (Jun-06-2018)
Current time:	17:07:02 (Jun-06-2018)
Status	running
Result URL:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=html">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=html</a>
Download TSV:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=tsv">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=tsv</a>



THIS WORK IS SUPPORTED BY NIG SUPERCOMPUTER SYSTEM AND NATIONAL BIOSCIENCE DATABASE CENTER.

NEED HELP? CREATE AN ISSUE ON [GITHUB](#) OR [CONTACT US](#)

## 5. 計算が終わるまで待つ

## ChIP-Atlas - *in silico* ChIP

Analyze your data with public ChIP-seq data.

Result page URL will be available for a week from the time when 'status' is 'finished'.

Project title	My project
Request ID	wabi_chipatlas_2018-0606-1701-36-865-010614
Submitted at:	17:01:37 (Jun-06-2018)
Estimated finishing time:	17:05:37 (Jun-06-2018)
Current time:	17:10:01 (Jun-06-2018)
Status	finished
Result URL:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=html">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=html</a>
Download TSV:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=tsv">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&amp;format=tsv</a>



THIS WORK IS SUPPORTED BY NIG SUPERCOMPUTER SYSTEM AND NATIONAL BIOSCIENCE DATABASE CENTER.  
NEED HELP? CREATE AN ISSUE ON [GITHUB](#) OR CONTACT US

## 6. 計算が終わると、「Result URL」が有効になる

- 今回の例では、[http://ddbj.nig.ac.jp/wabi/chipatlas/wabi\\_chipatlas\\_2018-0606-1701-36-865-010614?info=result&format=html](http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-0606-1701-36-865-010614?info=result&format=html)

## 7. 結果の解釈をする

- 今回は、どういう「興味ある遺伝子リスト」をクエリとしたか考察してみましょう。
- 結果の見方としては、「p-valueが低く、Overlaps/My dataが多く、Fold Enrichmentが高い」転写因子がたくさんヒットしてくると入力した遺伝子群をまとめて制御する、マスター転写因子を抽出できている確度が高い

## 8. 答え合わせ

# まとめ

- つまり食い的ではありますが、通り一遍の今日から使える便利な生命科学系公共データベース・ウェブツールを学びました。
- 顕微鏡 や 実験試薬 などと同じ「道具(ツール)」
- 便利な「道具」を知って、その使い方が分かれば、あとは情報分析力と想像力の勝負。
- 仮説構築から始まり、実験計画・検証、データ解析、そして論文執筆(以下ループ)という研究サイクルを加速化・効率化していきましょう。
- 困ったら、統合TV!! (\*宣伝)
- 研究に役立ったら、ぜひ引用・クレジットを!
- DBCLSの提供するサービス(あるいはそれ以外でも)が、あなたの研究に役立ったら、どんなに些細な事でもぜひ引用(論文、URL等)してください。DBCLSの活動は、提供するサービスがどのくらい活用されたかについて主に引用数などで評価されており、利用者の方の積極的なサポートが必要不可欠です!!

## もし時間が余れば

# 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。

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](#)

**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

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

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](#)

- 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。
- 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。

## 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](#)

← 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。

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

## 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](#)

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:

[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 \(0 selected\)](#)

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>		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	

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	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

- 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

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