

AJACS名古屋2 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

2018年12月5日(水) AJACS名古屋2 @ 愛知県がんセンター 国際医学交流センター 大会議室

これは統合データベース講習会 AJACS名古屋2 「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Enrichment Analysis 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。
 - <https://app2.sli.do/event/zl5dsf7c/questions>

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 68)

統合TVを知っていますか?	人数	割合
知らない	27 名	40 %
聞いたことがある	6 名	9 %
知っている	5 名	7 %
使ったことがある	13 名	19 %
使っている	10 名	15 %
回答なし	7 名	10 %

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	24 %
公共データを活用する・したい	11 名	16 %
既に持っている	25 名	37 %
大規模発現解析の予定はない	8 名	12 %
回答なし	8 名	12 %

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

- <https://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 [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

関連するタグから検索

- ゲノム (289)
- 遺伝子 (445)
- タンパク質 (219)
- 配列解析 (251)
- 発現解析 (338)
- NGS (254)
- 文献検索 (270)
- 情報収集 (137)

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/>

The screenshot shows the YouTube channel page for 'togotv'. At the top, there's a search bar and a login button. The channel's logo features a brown blocky character from Minecraft. Below the logo, the channel name 'togotv' and its subscriber count 'チャンネル登録者数 2,151 人' are displayed. A red button on the right says 'チャンネル登録'. The main navigation menu includes 'ホーム', '動画' (selected), '再生リスト', 'チャンネル', 'フリートーク', and '概要'. Below the menu, a message says 'アップロード済み ▾ すべて再生' and a '並べ替え' button. The video grid displays numerous thumbnail images of various educational and technical videos, such as 'PCRプライマー設計ツール Primer3の使い方' and 'Pythonが遅く明るい未来'.

- 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)

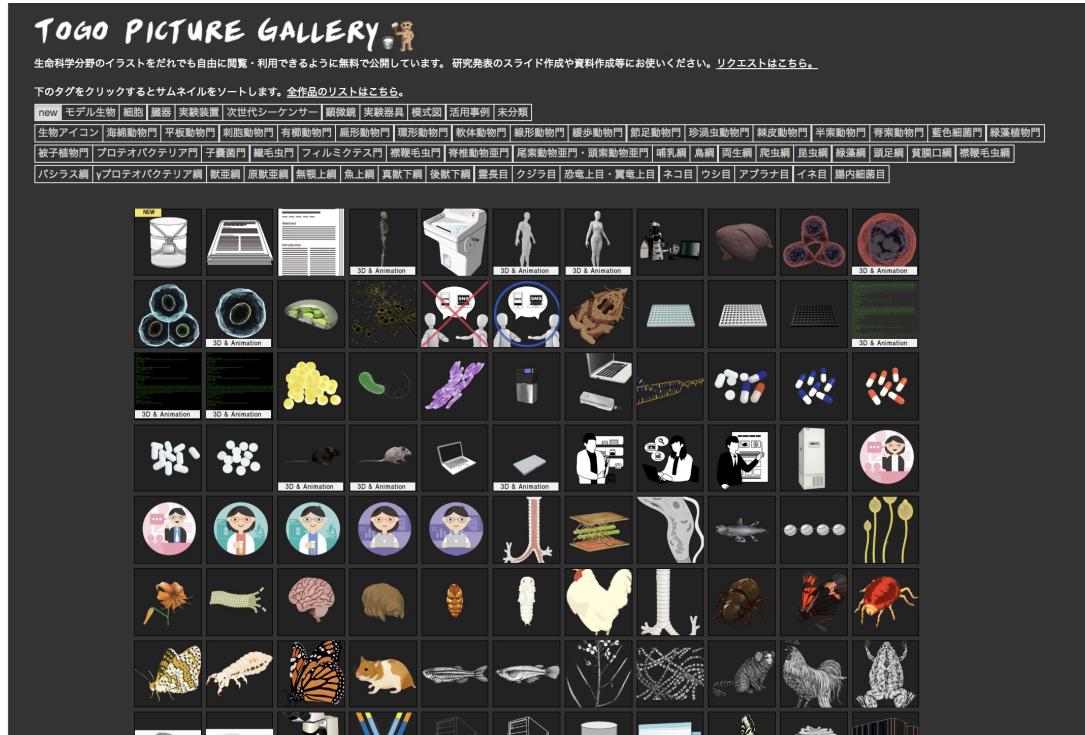


- 講義・講習などの参考資料や後輩指導の教材として利用できます。
 - 本講義中、本家サイトが繋がらない時は、統合TVを見ればおおよその内容がわかるようになっています。
 - DBCLSの提供する便利な各種サービスをレビューする内容もあります。
 - [【NGSハンズオン2017】NBDC・DBCLSの各種サービス](#) 今日から使える便利な生命科学系公共データベース in DBCLS
- 統合TVに掲載されているコンテンツについてご引用いただく際に、恒久的な URL として DOI (Digital Object Identifier) を使用することができます。
- 2014年8月以降に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、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・塩基配列解析
- 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をやってみたい方、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。)
- 統合TVを効果的に利用するためのガイドブックが出版されました。
 - 生命科学データベース・ウェブツール－図解と動画で使い方がわかる！ 研究がはかどる定番18選－
 - 「定番」として何がよく使われているのかを知り、その使い方を学び、どう使うと便利なのかについて、体系的にまとめて俯瞰的に捉えられるように編集されています。
 - 「この順で動画をみていくと、こういうスキルを獲得できる」というような体系的な教材です。
 - <https://www.amazon.co.jp/dp/4815701431/>
 - https://www.medsi.co.jp/books/products/detail.php?product_id=3665

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
 - [塩基配列解析およびゲノム編集のためのデータベース・ウェブツール](#)
- 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」\(2015年5月AJACS御茶ノ水\)](#)
- PubMed検索のプロによる文献検索のイロハを学ぶ
 - [文献情報を用いたサービスを活用する @ AJACS越後](#)
- 次世代シーケンス(NGS)データの解析について
 - [NGSデータから新たな知識を導出するためのデータ解析リテラシー @ AJACS浜松](#)
- 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)
 - 「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)
 - 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
- ビッグデータ時代の疾患ゲノム解析で使いこなしたいデータベース
 - The Cancer Genome Atlas (TCGA) を使って各癌種の公開データを検索・ダウンロードする
 - The Cancer Genome Atlas (TCGA) を使って各癌種の公開データを解析する
 - COSMICでがん遺伝子の体細胞変異について調べる
 - COSMIC-3Dを使ってがん遺伝子のコードするタンパク質の立体構造をがんで見られる変異の情報と重ねて見る
 - 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）
 - GWAS Catalog を使って、GWASで見つかった形質と多型の関連について検索する
 - ClinVarで疾患に関連する変異を検索し、Variation Viewerで視覚的に変異を探す
 - gnomADを使ってヒトのエキソームやゲノムのデータから変異を探す
 - GTEx Portalを使ってヒトの各組織での遺伝子発現量や影響するeQTLを調べる
 - TogoVarでヒトゲノムに存在するバリエントに関する情報を調べる

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
 - <https://refex.dbcls.jp/>

refex.dbcls.jp

Reference
Expression
Dataset
[English](#) | [日本語](#)


RefEx

RefExの論文が出版されました。あなたの研究に役立つたらぜひ引用を!!
▼もっと詳しく


Human Mouse Rat

キーワードで検索 ex) troponin, ALB

組織特異的に発現する遺伝子を見る



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





- 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 IDH3 α 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
- 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
- 統合TV 「RefExの使い方」 DOI: 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. <https://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

肝臓

遺伝子オントロジー Gene Ontology

- cellular process
- biological regulation
- metabolic process
- multicellular organismal process
- response to stimulus
- developmental process

他のオントロジーを選ぶ

遺伝子ファミリー InterPro

- RNA recognition motif, RNP-1
- Pleckstrin homology
- Krüppel-associated box
- Protein kinase-like domain
- Zinc finger, C2H2-like
- GPCR, rhodopsin-like superfamily

他のファミリーを選ぶ

染色体

染色体領域を選ぶ

Advanced Search Advanced Search

ページ上部に戻る

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



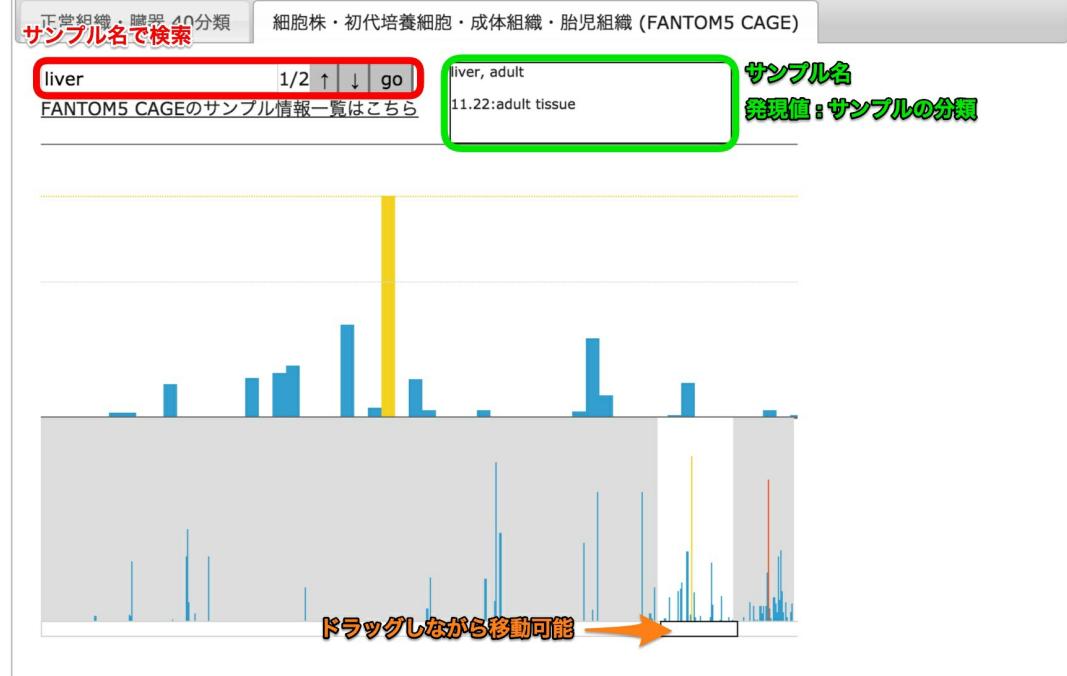
Rat

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

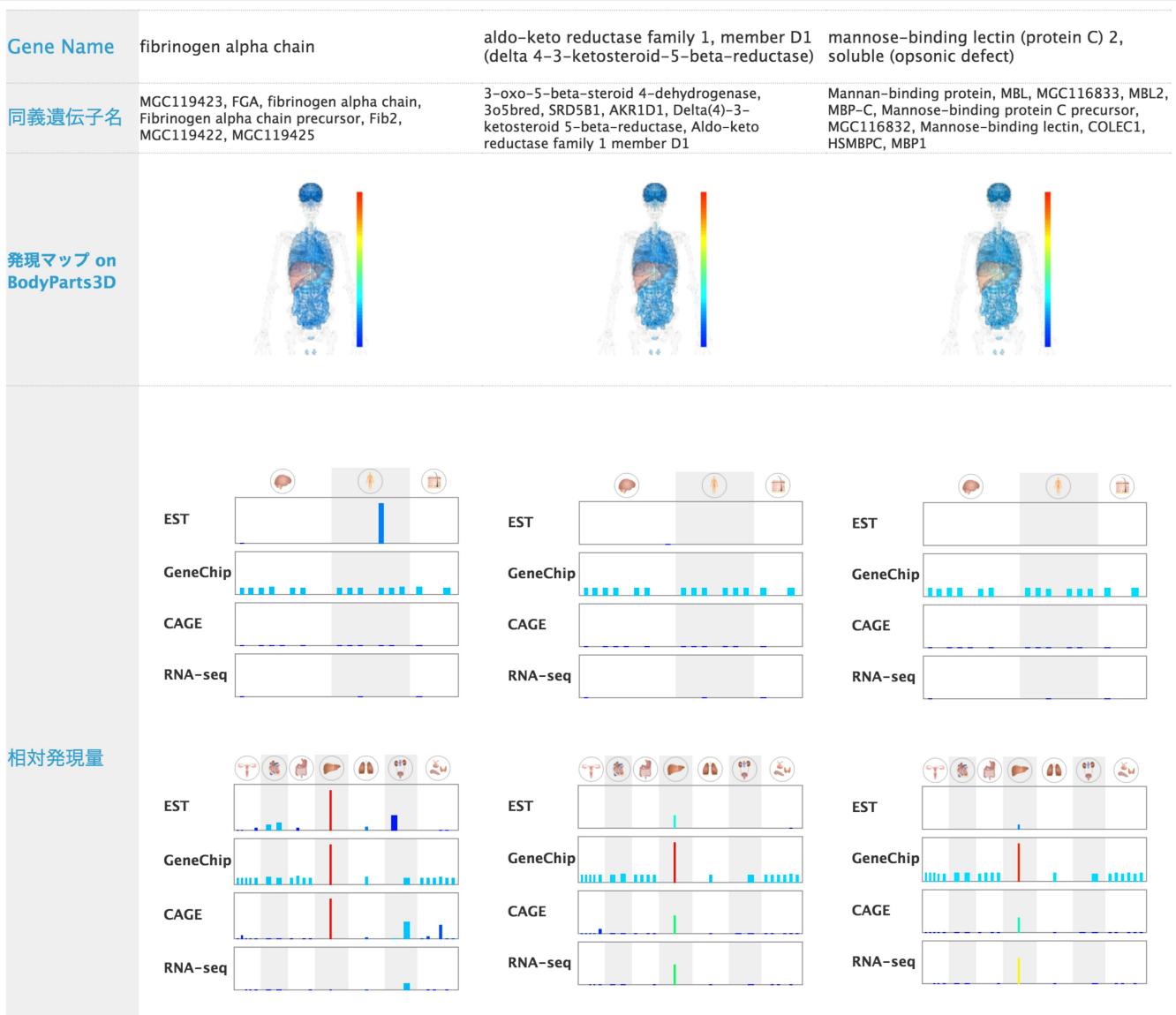


12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- 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ィンドウに「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。

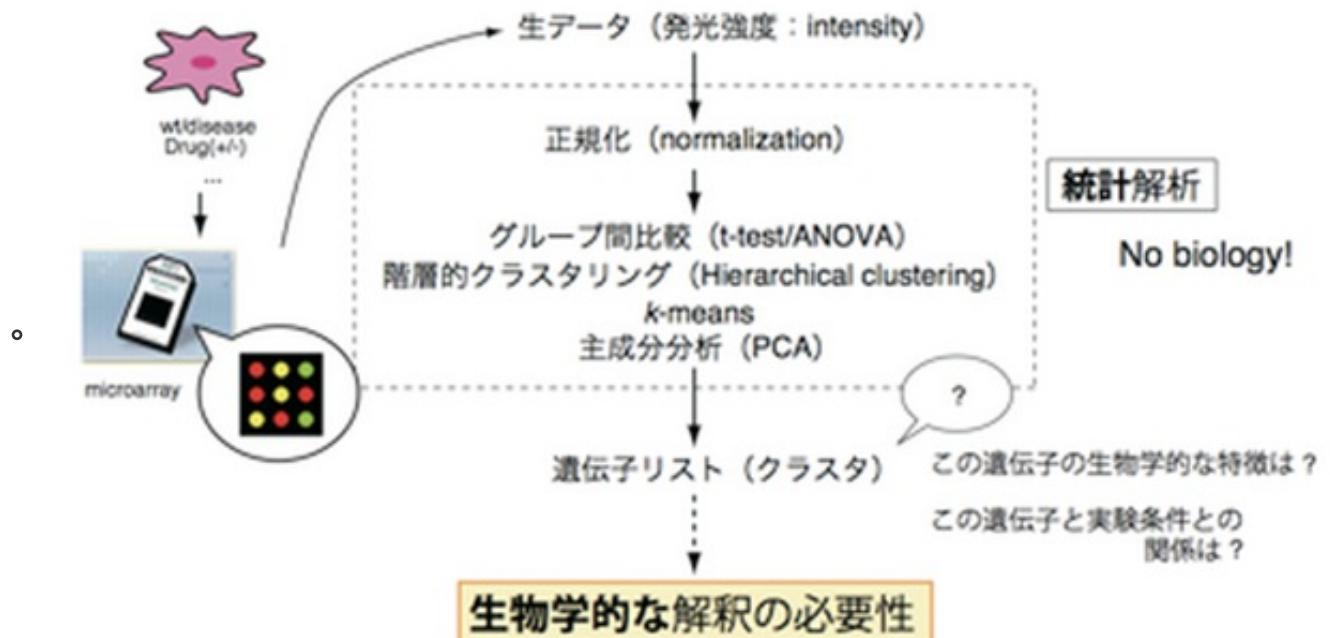


Refseq ID	NM_000508	NM_005989	NM_000242
Gene ID	2243	6718	4153
Unigene ID	Hs.351593	Hs.201667	Hs.499674
probe set ID	205649_s_at	207102_at	207256_at
Ensembl ID	ENSG00000171560	ENSG00000122787	ENSG00000165471
染色体	4.q31.3 [155504278 – 155511918]	7.q33 [137687070 – 137802732]	10.q21.1 [54525140 – 54531460] 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's ... negative regulation ... opsonization positive regulation ... response to oxidativ ...

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。

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

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 78,000 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](#)

Enrichment Analysis

predicts proteins bound to given genomic loci and genes (formerly known as *in silico* ChIP).

[Watch Movie](#)



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

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

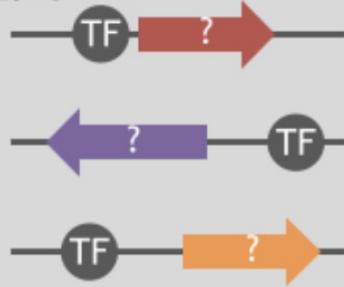
① Peak Browser

何がどこに結合？



② Target Genes

標的遺伝子は？



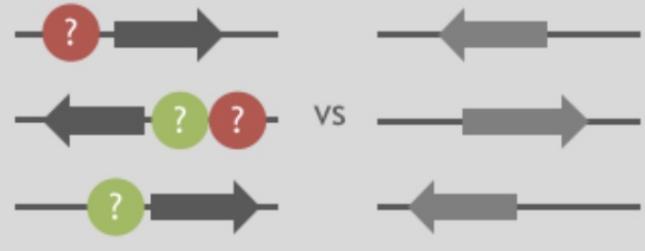
③ Colocalization

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~」

Enrichment Analysis

- 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、シーケンスマチーフを 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/lstdc01558-000>
- Publications citing ChIP-Atlas <http://chip-atlas.org/publications>

【実習2】 ChIP-AtlasのEnrichment Analysis 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
 - [181205_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 - Enrichment Analysis](#)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 - Enrichment Analysis

[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 (34390)
 DNase-seq (1511)
 Histone (8982)
 RNA polymerase (1104)
TFs and others (8368)
 Input control (4002)
 Unclassified (6194)
 No description (4229)

2. Cell type Class

All cell types (34390)
 Adipocyte (292)
 Blood (9096)
 Bone (655)
 Breast (3699)
 Cardiovascular (1080)
 Digestive tract (2447)
 Epidermis (1089)

3. Threshold for Significance ⓘ

50
100
 200
 500

4. Select your data

Genomic regions (BED) or sequence motif ⓘ
 Gene list (Gene symbols) ⓘ

NR1I3
 hmgcL
 serpinC1
 ALDH2
 SHMT2
 SULT1A3
 SULT1A4
 Armc6

ファイルを選択 ファイル未選択

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: 5 mins

3. submit すると遺伝研スパコンへクエリが飛ぶ(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 - Enrichment Analysis

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_2018-1104-1735-21-275-473936
Submitted at:	17:35:18 (Nov-04-2018)
Estimated finishing time:	17:48:18 (Nov-04-2018)
Current time:	17:43:34 (Nov-04-2018)
Status	running
Result URL:	http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&format=html
Download TSV:	http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&format=tsv



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 / Enrichment Analysis

Search for proteins significantly bound to your data.

Show [100] entries

Search:

My project

ID	Antigen class	Antigen	Cell class	Cell	Num of peaks	Overlaps / My data	Overlaps / Control	Log P- val	Log Q- val	Fold Enrichment	FE > 1?
SRX100505	TFs and others	HNF4A	Liver	Hep G2	21259	118/218	3453/18327	-30.0	-26.7	2.87	TRUE
SRX100544	TFs and others	EP300	Liver	Hep G2	24334	122/218	4026/18327	-26.5	-23.3	2.55	TRUE
SRX100497	TFs and others	RXRA	Liver	Hep G2	13022	92/218	2362/18327	-25.7	-22.5	3.27	TRUE
SRX100449	TFs and others	HNF4G	Liver	Hep G2	15919	102/218	3119/18327	-23.2	-20.1	2.75	TRUE
SRX100448	TFs and others	FOXA2	Liver	Hep G2	45130	135/218	5342/18327	-22.6	-19.6	2.12	TRUE
SRX190234	TFs and others	CEBPB	Liver	Hep G2	12483	84/218	2308/18327	-21.1	-18.2	3.06	TRUE
SRX100506	TFs and others	FOXA1	Liver	Hep G2	50941	141/218	6086/18327	-20.4	-17.5	1.95	TRUE
SRX100477	TFs and others	FOXA1	Liver	Hep G2	40732	127/218	5256/18327	-18.8	-15.9	2.03	TRUE
SRX100552	TFs and others	SP1	Liver	Hep G2	19032	129/218	5594/18327	-17.4	-14.7	1.94	TRUE
SRX150698	TFs and others	HNF4A	Liver	Hep G2	10069	76/218	2336/18327	-16.0	-13.3	2.74	TRUE
SRX150701	TFs and others	CEBPB	Liver	Hep G2	18637	105/218	4173/18327	-15.5	-12.8	2.12	TRUE
SRX018625	TFs and others	HNF4A	Liver	Hep G2	2654	33/218	516/18327	-13.9	-11.3	5.38	TRUE
SRX1531773	TFs and others	MED1	Liver	Hep G2	3984	32/218	507/18327	-13.4	-10.8	5.31	TRUE

Showing 1 to 100 of 22,299 entries

Previous 1 2 3 4 5 ... 223 Next

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_2018-1104-1735-21-275-473936?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

さらに学びたい方向け

データ可視化・解析ツール

- Tableau Public** を使って 行列データを直感的に可視化し解析する
- Morpheus** を使って、様々な行列データをヒートマップで表示しウェブブラウザでデータ解析をする
- Dataset2Tools** でオミックスデータとその解析事例、計算ツールを検索し、再現性の高い再解析を行う
 - Enrichr

まとめ

- つまみ食い的ではありますが、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を学びました。
- これらは顕微鏡 や 実験試薬 などと同じ「道具(ツール)」です。
- 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!!