

# AJACSオンライン3 遺伝子発現データ解析（RefEx、ChIP-Atlas、DAVID、metascape、GEO2Rなど）

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)

2020年11月11日(水) AJACSオンライン3

これは統合データベース講習会 AJACSオンライン3 「遺伝子発現データ解析（RefEx、ChIP-Atlas、DAVID、metascape、GEO2Rなど）」の講習資料です。

講習会全体のプログラムは[こちら](#)です。 © 2020 小野 浩雅, CC-BY-4.0

## 自己紹介

- 小野 浩雅
  - RefExの開発者、統合TVの運営・編集者
  - Twitter

## 概要

本講習では、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、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ための各種ツールとその使い方を学びます。さらに、公共データベース上にある生データを解析するためのツールを紹介します。

## 講習の流れ

今回の講習では、以下の内容について説明します。

- 個々の遺伝子の発現プロファイルを調べる
  - RefEx
    - 【デモ】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
    - 【デモ】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
  - DAVID
  - metascape
  - Enrichr
- 公共データベース上にある遺伝子発現データを解析する

- GEO
  - GEO2R
  - GREIN
  - iDEP
  - BioJupies
  - 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
- 

## 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。
  - Web上で質問・コメントできるフォームを用意していますので、ご活用ください。
    - <https://www.sli.do> (「新しいタブで開く」とよいです)
      - ウェブブラウザで sli.do と入力(スマホでもアクセス可能)
      - AJ84 と入力
      - 質問をする (°д°)カンタン
  - こんなことは知ってて当たり前だと他の人に思われるような「シロウト」質問を歓迎します。
    - 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、皆さん受講満足度が上がります(たぶん)。
    - あなたがいま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。
    - 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
    - でも講師も知らないことは(多々)あります。(以下ループ)
- 

**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 158)(11/2時点)**

統合TVを知っていますか?	人数	割合
知らない	21 名	13 %
知っている	137 名	87 %

---

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
既に持っている	61 名	40.7 %

---

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: 発現差のある 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	29 名	19.3 %
公共データを活用する・したい	47 名	31.3 %
大規模発現解析の予定はない	13 名	8.7 %

## 誰でも使える公共の遺伝子発現データ

- さまざまな実験で得られたデータは、論文投稿時などに公共データベース上に登録し、その後誰でも参照可能になるようになることが義務付けられていることが多いですが、それは遺伝子発現データについても同様です。公共データベースには多種多様な遺伝子発現データが日々大量に登録、蓄積され続けています。
  - これらのデータをうまく活用すれば、例えば、予備実験をせずに済む、自分の実験結果を支持する知見が得られる、仮説立案の新たな切り口が得られるなど、多くのメリットがあります。
  - 一方で、いざ使おうと思ったときに、その膨大さから、どのデータを選び、詳しく調べればよいのか分かりにくく、利用することが困難で諦めてしまうことが多いです。
- これらの困難さを軽減するために、いくつかの二次的なデータベースやツールが開発されています。これらをうまく使うことで、自分のデータと組み合わせられる遺伝子発現データをすばやく検索、発見し、活用できるようになるでしょう。

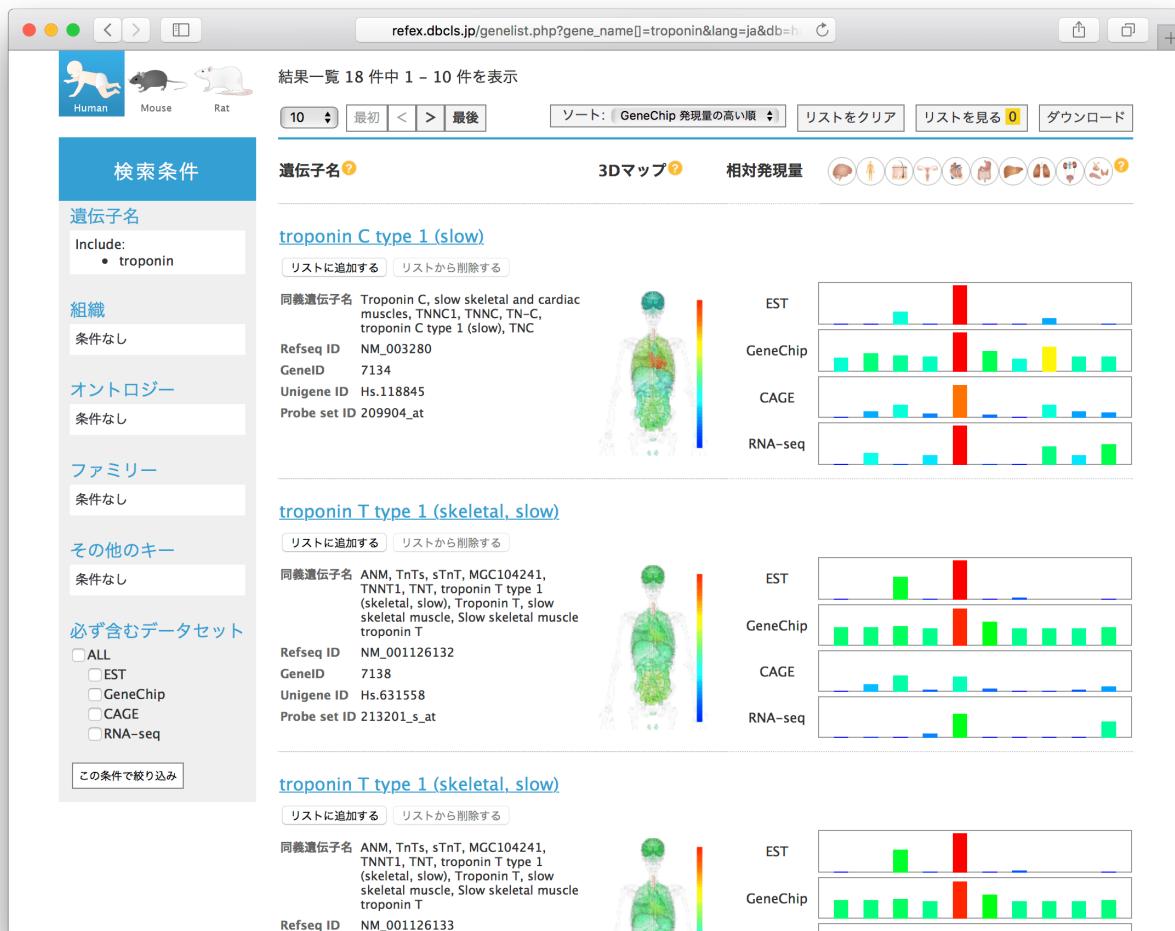
## 個々の遺伝子の発現プロファイルを調べる

### 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/>

The screenshot shows the RefEx website interface. At the top, there are links for Reference, Expression, and Dataset, along with language options (English | 日本語). The RefEx logo is prominently displayed. A message on the right side reads: "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!!" and "▼もっと詳しく". Below the header, there are icons for Human, Mouse, and Rat. A search bar contains the placeholder "キーワードで検索" and a yellow "検索" button. To the right of the search bar is an example: "ex) troponin, ALB". 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) features icons for various organs: brain, human figure, heart, uterus, liver, intestines, lungs, kidneys, and pancreas. Below this, there are sections for "遺伝子オントロジー Gene Ontology" and "遺伝子ファミリー InterPro". The GO section lists processes like cellular process, biological regulation, metabolic process, multicellular organismal process, response to stimulus, and developmental process. The InterPro section lists motifs like RNA recognition motif, RNP-1, Pleckstrin homology, Krueppel-associated box, Protein kinase-like domain, Zinc finger, C2H2-like, and GPCR, rhodopsin-like superfamily. A "染色体" (Chromosome) section includes a "染色体領域を選ぶ" (Select chromosome region) button. A "他のオントロジーを選ぶ" (Select other ontologies) button is also present.





- 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
  - FANTOM5 CAGEデータ(phase1)も閲覧可能です(ヒト556種、マウス286種)
    - FANTOM5 CAGEデータ
      - 理化学研究所のFANTOM プロジェクト5 ([FANTOM5](#)) では、ゲノムにコードされているプロモーターと転写因子制御ネットワークを明らかにすることを目的として膨大なCAGEデータを測定し、それらを閲覧できるウェブサイトも公開されている。
      - 現在、RefExに載っているデータで、一番、データ精度が高く、また、サンプル数も多いため、オススメです。
      - 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 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 @ 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・利用した論文の一覧はこちらでご覧いただけます。(2020年10月末で32報)
  - <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] (<http://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] (<http://doi.org/10.7875/first.author.2017.093>)

- 統合TV 「RefExの使い方」 DOI: 10.7875/togtv.2014.009

## 【使用例】 RefExを使って、組織特異的遺伝子を検索する

- 【統合TV】 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。

Reference  
Expression  
Dataset  
English | 日本語



臓器ごとの発現比較を  
4つの実験手法と  
ボディーパーツ3Dで。  
▼もっと詳しく



キーワードで検索

検索

ex) troponin, ALB

組織特異的に発現する  
遺伝子を見る遺伝子オントロジー  
Gene Ontology

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

他のオントロジーを選ぶ

遺伝子ファミリー  
InterPro

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

他のファミリーを選ぶ

染色体

染色体領域を選ぶ

Advanced Search

Advanced Search

ページ上部に戻る

RefExについて

RefExの使い方

ダウンロード



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

refex.dbcls.jp/genelist.php?lang=ja&db=human&roku\_valid=1&rkr[31]=31&order\_key=score

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。

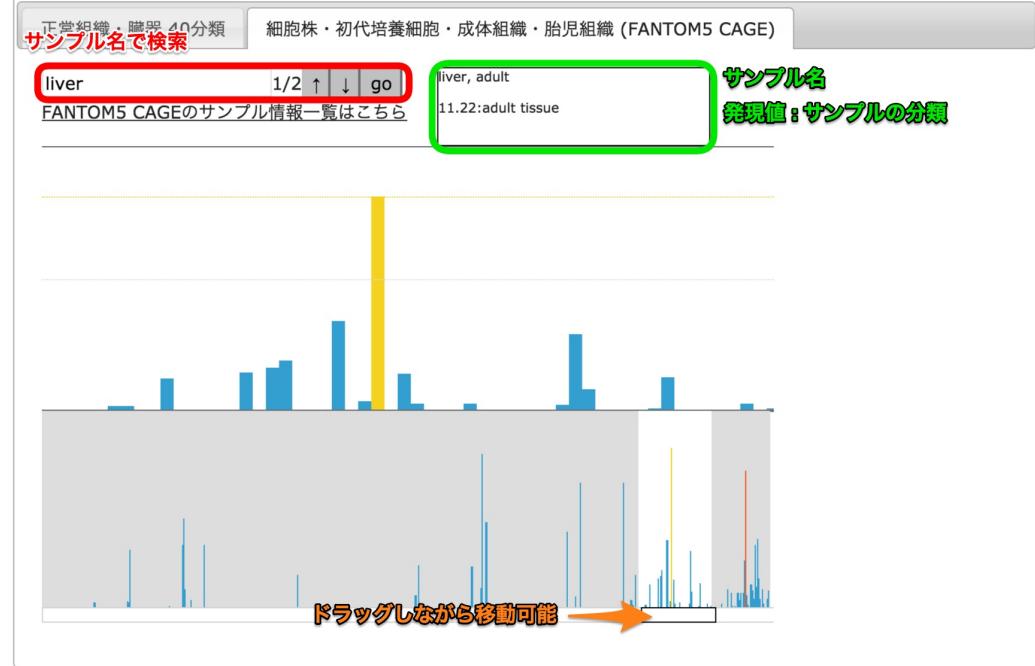


## fibrinogen alpha chain

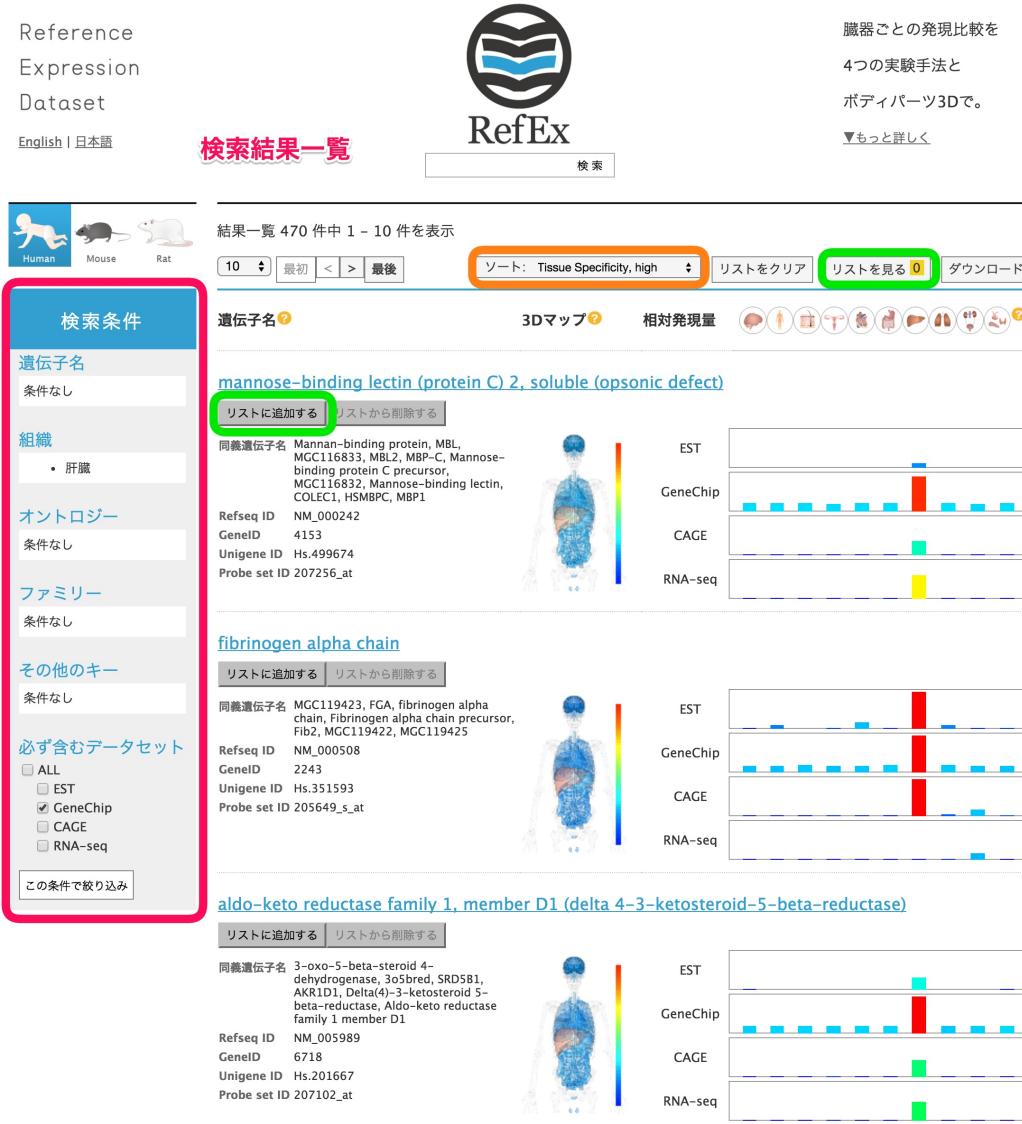
[詳細情報を見る](#)

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。

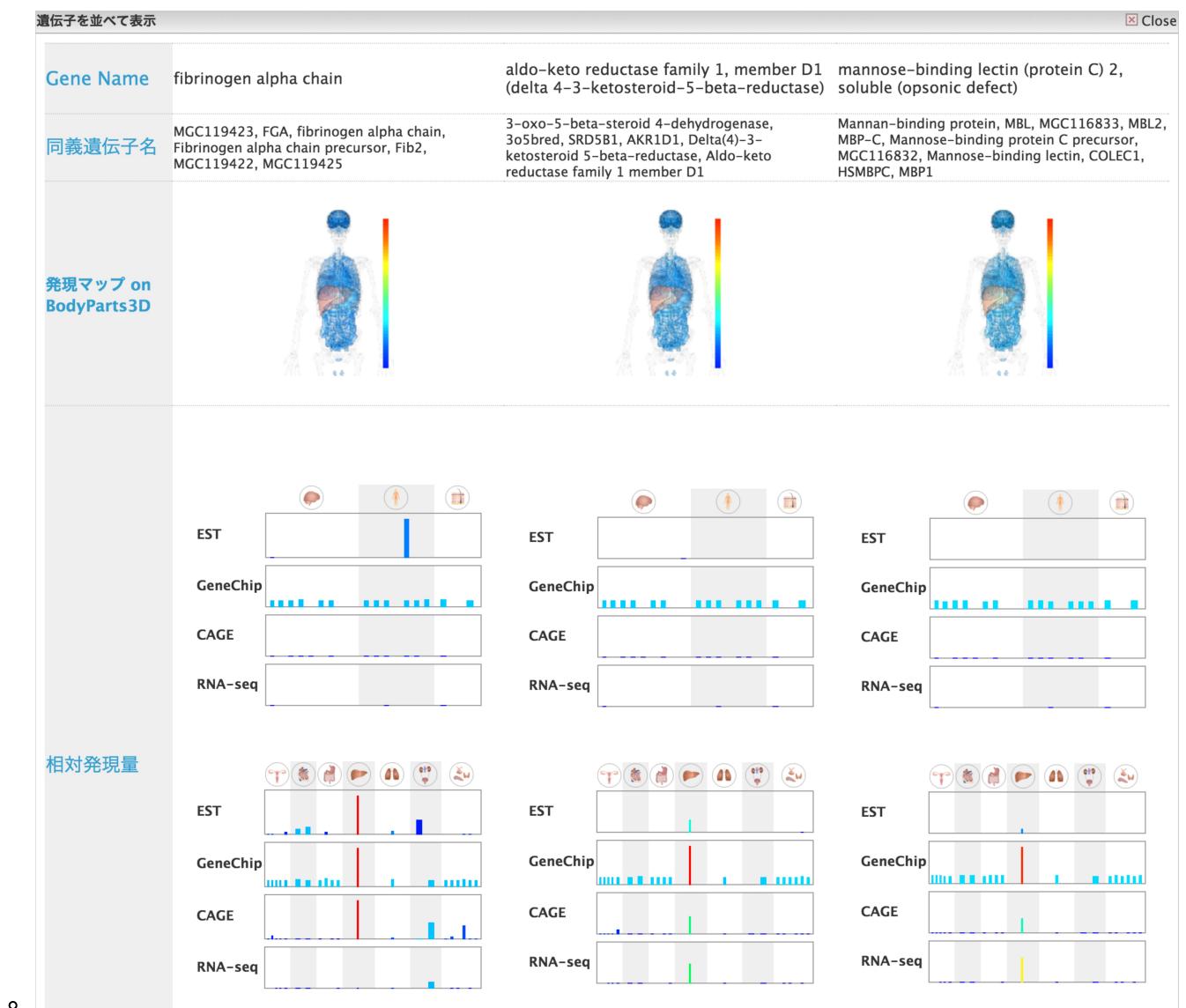


## 12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- 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	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 - - -	- - C-type lectin C-type lectin fold C-type lectin-like
差分が明確に			
遺伝子オントロジー Biological Process	blood coagulation	-	-
	platelet activation	-	-
	platelet degranulati ...	-	-
	protein polymerizati ...	-	-
	response to calcium ...	-	-
	signal transduction	-	-
	metabolic process	-	-
	bile acid biosynthet ...	-	-
	bile acid catabolic ...	-	-
	bile acid metabolic ...	-	-
	C21-steroid hormone ...	-	-
	cholesterol cataboli ...	-	-
	digestion	-	-
	oxidation reduction	-	-
	innate immune response	-	-
	acute-phase response	-	-
	complement activatio ...	-	-
	complement activatio ...	-	-
	complement activatio ...	-	-
	defense response to ...	-	-
	innate immune respon ...	-	-
	killing by host of s ...	-	-
	negative regulation ...	-	-
	opsonization	-	-
	positive regulation ...	-	-
	response to oxidativ ...	-	-
	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。

## 関連するツール

- ExpressionAtlas

- Expression Atlas は、EMBL-EBI (欧洲バイオインフォマティクス研究所) が運営する、様々な生物種における遺伝子発現情報を提供するウェブサイトである。40 の生物種について、3000 以上の実験に基づき、組織や細胞種、発生段階、疾患の有無などの条件別に、遺伝子の発現情報をまとめられています。
- 【統合TV】 Expression Atlas で 様々な生物種の組織や疾患などにおける遺伝子発現の情報を調べる

- Bgee

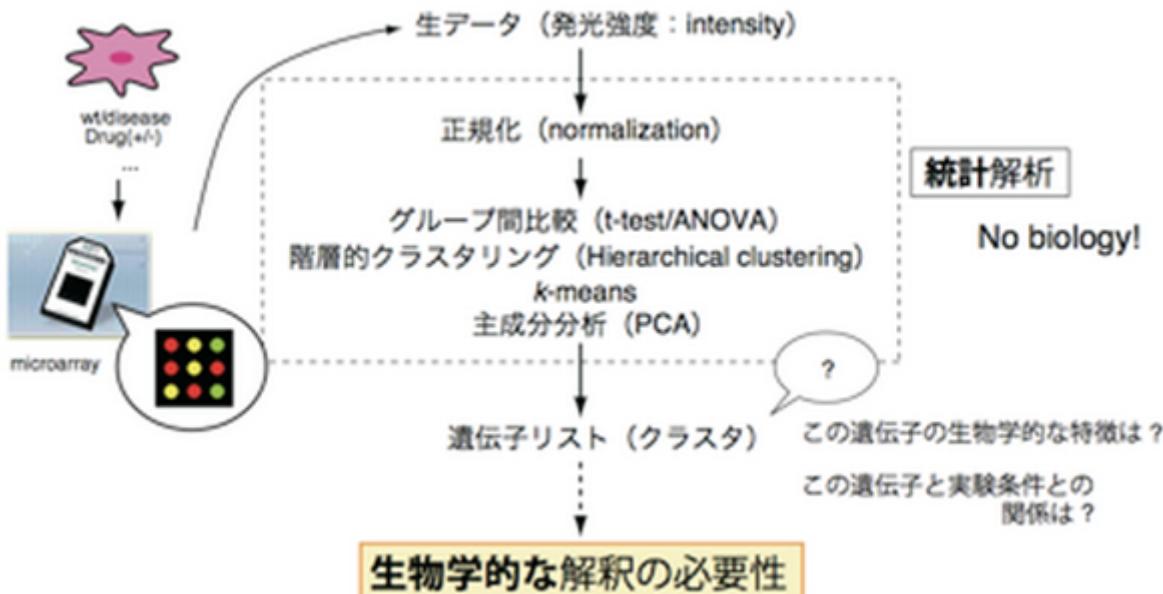
- Bgee (a DataBase for Gene Expression Evolution: ビージーと発音します) は、SIB Swiss Institute of Bioinformatics/ローザンヌ大学の Robinson-Rechavi Group - Evolutionary Bioinformaticsが提供するデータベースで、これはRNA-seqやマイクロアレイなどから得られた遺伝子発現パターンについて、多くの生物種間で比較することができます。
- 【統合TV】 Bgee を使って、複数の生物種の正常組織における遺伝子発現データを検索、比較、取得する

- BioGPS

- AffymetrixのマイクロアレイであるGeneChipを用いたヒト、マウス、ラットを始め全9生物種における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です。
- 【統合TV】 遺伝子発現プロファイルデータベースBioGPSを使い倒す 2012

# 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](#)

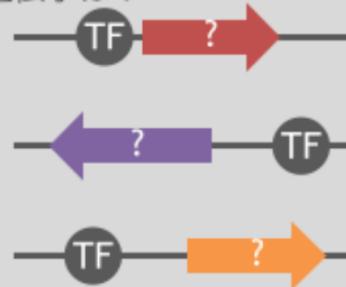
### ① Peak Browser

何がどこに結合？



### ② Target Genes

標的遺伝子は？



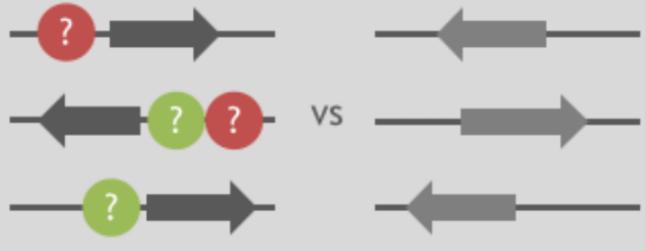
### ③ Colocalization

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。
  - 【統合TV】 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。
  - 【統合TV】 ChIP-Atlas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 ~ Enrichment Analysisの使い方~

### 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>
- Paper
  - Shinya Oki, Tazro Ohta, Go Shioi, Hideki Hatanaka, Osamu Ogasawara, Yoshihiro Okuda, Hideya Kawaji, Ryo Nakaki, Jun Sese, and Chikara Meno. **ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data.**

EMBO Reports

[<https://doi.org/10.15252/embr.201846255>] (<https://doi.org/10.15252/embr.201846255>)

- Database
  - Oki, S; Ohta, T (2015): ChIP-Atlas. <http://dx.doi.org/10.18908/lMDBA.nBDC01558-000>
- Publications citing ChIP-Atlas <http://chip-atlas.org/publications>

### 【使用例】 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

- 【統合TV】 ChIP-Atlas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 ~Enrichment Analysisの使い方~

- 【統合TV】 ChIP-Atlas: 既報のChIP-seqデータをフル活用できる @ AJACS町田
  - 開発者の沖さん直々の講義・講習
  - 「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

- 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りますが、今回はHGNC BioMartを利用します。
    - HGNC(The HUGO Gene Nomenclature Committee)はヒトのGeneSymbolを認定・管理している機関。
    - 【統合TV】 HGNCを使ってヒト遺伝子の正式略称(GeneSymbol)を検索する(+ヒトとマウスの遺伝子IDを変換する)
    - 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にアクセスします。

2. 下図のようにオプションを設定します。

The screenshot shows the ChIP-Atlas Enrichment Analysis interface. At the top, there is a navigation bar with links: ChIP-Atlas, Peak Browser, Target Genes, Colocalization, Enrichment Analysis (which is highlighted in blue), Documentation, Publications, and a search bar labeled "Find an experiment".

The main area is titled "ChIP-Atlas - Enrichment Analysis" and contains the following sections:

- 1. Antigen Class:** A dropdown menu showing categories like "All antigens (34390)", "DNase-seq (1511)", "Histone (8982)", "RNA polymerase (1104)", "TFs and others (8368)" (which is selected), "Input control (4002)", "Unclassified (6194)", and "No description (4229)".
- 2. Cell type Class:** A dropdown menu showing categories like "All cell types (34390)", "Adipocyte (292)", "Blood (9096)", "Bone (655)", "Breast (3699)", "Cardiovascular (1080)", "Digestive tract (2447)", and "Epidermis (1089)".
- 3. Threshold for Significance ⓘ:** A dropdown menu showing values 50, 100 (which is selected), 200, and 500.
- 4. Select your data:** Options for selecting genomic regions or a gene list. It shows a list of genes: NR1I3, hmgcL, serpinc1, ALDH2, SHMT2, SULT1A3, SULT1A4, Armc6. Below this are buttons for "ファイルを選択" (File selection), "Choose local file", and "Try with example".
- 5. Select dataset to be compared:** Options for Refseq coding genes (excluding user data) (selected) and Gene list (Gene symbols).
- 6. Describe datasets:** Fields for User data title (My data), Compared data title (Control), Project title (My project), and Distance range from TSS (- 5000 bp ≤ TSS ≤ + 5000 bp). A "submit" button is at the bottom.

3. submit すると遺伝研スパコンへクエリが飛びます。(ので、講義中は見てるだけにしてください)

#### 4. submit したあの画面

ChIP-Atlas Peak Browser Target Genes Colocalization Enrichment Analysis Documentation Publications Find an experiment ▾

## 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_2020-1105-0855-56-916-954292
Submitted at:	08:55:57 (Nov-05-2020)
Estimated finishing time:	09:02:57 (Nov-05-2020)
Current time:	08:56:37 (Nov-05-2020)
Status	queued
Result URL:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2020-1105-0855-56-916-954292?info=result&amp;format=html">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2020-1105-0855-56-916-954292?info=result&amp;format=html</a>
Download TSV:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2020-1105-0855-56-916-954292?info=result&amp;format=tsv">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2020-1105-0855-56-916-954292?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ちます

## 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	21265	115/214	3442/18213	-28.8	-24.6	2.84	TRUE
SRX3230395	TFs and others	PAXIP1	Liver	Hep G2	9446	78/214	1605/18213	-27.4	-23.6	4.14	TRUE
SRX3230397	TFs and others	PROX1	Liver	Hep G2	22274	102/214	2822/18213	-27.2	-23.5	3.08	TRUE
SRX3230385	TFs and others	FOXP1	Liver	Hep G2	15626	81/214	1836/18213	-25.9	-22.3	3.75	TRUE
SRX100544	TFs and others	EP300	Liver	Hep G2	24336	119/214	4010/18213	-25.4	-22.0	2.53	TRUE
SRX100497	TFs and others	RXRA	Liver	Hep G2	13022	90/214	2356/18213	-24.9	-21.5	3.25	TRUE
SRX100449	TFs and others	HNF4G	Liver	Hep G2	15922	101/214	3109/18213	-23.3	-19.9	2.76	TRUE
SRX100448	TFs and others	FOXA2	Liver	Hep G2	45142	131/214	5322/18213	-21.1	-17.9	2.09	TRUE
SRX190234	TFs and others	CEBPB	Liver	Hep G2	12489	83/214	2297/18213	-21.0	-17.8	3.08	TRUE
SRX100506	TFs and others	FOXA1	Liver	Hep G2	50949	137/214	6063/18213	-18.9	-15.7	1.92	TRUE
SRX100477	TFs and others	FOXA1	Liver	Hep G2	40732	123/214	5231/18213	-17.5	-14.3	2.00	TRUE
SRX150698	TFs and others	HNF4A	Liver	Hep G2	10069	76/214	2329/18213	-16.5	-13.4	2.78	TRUE
SRX3230382	TFs and others	FOXA2	Liver	Hep G2	43313	138/214	6588/18213	-16.0	-13.0	1.78	TRUE
SRX100552	TFs and others	SP1	Liver	Hep G2	19079	125/214	5595/18213	-15.9	-12.9	1.90	TRUE
SRX150701	TFs and others	CEBPB	Liver	Hep G2	18637	104/214	4152/18213	-15.6	-12.6	2.13	TRUE
SRX018625	TFs and others	HNF4A	Liver	Hep G2	2654	33/214	514/18213	-14.2	-11.2	5.46	TRUE
SRX1531773	TFs and others	MED1	Liver	Hep G2	3984	32/214	504/18213	-13.6	-10.7	5.40	TRUE
SRX100538	TFs and others	HDAC2	Liver	Hep G2	16079	98/214	4328/18213	-11.6	-8.7	1.93	TRUE
SRX190266	TFs and others	NR2F2	Liver	Hep G2	18212	107/214	5012/18213	-11.3	-8.4	1.82	TRUE

Showing 1 to 100 of 14,630 entries

Previous 1 2 3 4 5 ... 147 Next

#### 6. 計算が終わると、「Result URL」が有効になります。

- 今回の例では、[http://ddbj.nig.ac.jp/wabi/chipatlas/wabi\\_chipatlas\\_2020-1105-0855-56-916-954292?info=result&format=html](http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2020-1105-0855-56-916-954292?info=result&format=html)

## 7. 結果の解釈をします。

- 今回は、どういう「興味ある遺伝子リスト」を使ったか、について上記の結果から考察してみましょう。
  - p-valueが小さい=統計的に有意にEnrichしている
  - Overlaps/My dataが多い=自分の遺伝子リスト中でピークがオーバーラップしている数
  - Fold Enrichmentが高い=(Overlaps/Control)/(Overlaps/My data)の割合が高い
  - これらを満たす転写因子がたくさんヒットしている → 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可能性が高いと解釈できる。

## 8. 答え合わせ

### 関連するツール

- [DAVID](#)
  - DAVIDはマイクロアレイ実験から得られたデータを解析するツールです。このツールを使うことで発現変動のあった遺伝子群の特徴を可視化し、直感的に分析することができます。DAVIDという名前はThe Database for Annotation, Visualization and Integrated Discoveryの頭文字に由来しています。
  - 【統合TV】 DAVIDの使い方 実践編
  - 【統合TV】 実験データの生物学的解釈をするための遺伝子発現DB・ウェブツールの使い方@AJACS下総
- [Metascape](#)
  - 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。
  - 【統合TV】 Metascapeを使って、遺伝子リストの生物学的解釈をする
- [Enrichr](#)
  - 【統合TV】 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
  - (6:00~) Enrichr による RNA-seq データのエンリッチメント解析
  - 特定のモデル生物(ショウジョウバエ、酵母、線虫、ゼブラフィッシュ)に特化したEnrichrもあります。

## 公共データベース上にある遺伝子発現データを解析する

### GEO

- NCBIが提供・維持管理している世界最大の遺伝子発現情報のデータベース
- 自分の興味のある発現データセットや遺伝子プロファイルを検索することができるだけでなく、それらの生データを自由にダウンロードすることができます。
  - GEOのエントリについて (GEO ID番号の最初の3文字の意味)
    - GPL: Platform — 実験手法の種類別に振られたID
    - GSM: Sample — 1つのサンプルから得られたデータごとに振られたID
    - GSE: Series — 1つの実験で得られたGSMのセットごとに振られたID
    - GDS: DataSet — NCBIのスタッフが解析可能なデータを集めて再編成したGSMのセット
  - 【統合TV】 NCBI GEOのデータセットブラウザを使って公共データの遺伝子発現解析を行う 2019

## GEO2R

- NCBIが提供している遺伝子発現データの解析ツール
  - 統計解析ソフト [R](#) をベースに解析します。
  - GEOに登録されているデータセットの中から、それぞれのサンプルデータを比較したいグループに分け、統計解析することによって発現量に差がある遺伝子群のリストを取得できます。
- マイクロアレイの生データの検索から簡易的な統計解析までを収録したハンズオン動画をご参照ください。
  - 【統合TV】遺伝子発現DB・ウェブツールの使い方 応用・実践編

### 関連するツール

#### **GREIN (GEO RNA-Seq Experiments Interactive Navigator)**

- GEOに登録されているRNA-Seqデータを再利用するための処理済みデータと各種解析ツールを提供しているウェブツール
  - GREINで提供されている処理済みデータセットはヒト・マウス・ラットを対象に6,500以上(40万サンプル相当)
  - 各データは、[GREP2: GEO RNA-seq Experiments Processing Pipeline](#)によって処理されています。
  - データセットのクオリティコントロールや発現パターンの視覚化、サンプルサイズとそれに基づく発現変動遺伝子の検出分析などのRNA-seqデータの再利用に必要な典型的な調査が容易に実行できます。
  - 発現変動遺伝子の検出やクラスタリング、主成分分析(PCA)、t-SNE、エンリッチメント解析およびネットワーク解析についてもインタラクティブに実行することができます。
- 【統合TV】GREINを使ってNCBI GEOのRNA-seqデータを分析する

#### **iDEP (integrated Differential Expression and Pathway analysis)**

- ウェブブラウザ上でRNA-seqデータ解析を行うことができるウェブツール
  - RNA-seqやマイクロアレイ、ChIP-seq実験等で得られた遺伝子発現データ（リードカウントまたはFPKM、正規化済みの発現値、Fold-changeなど）を入力
  - ヒートマップやPCA、発現差解析、パスウェイ解析、エンリッチメント解析、バイクラスタリング法および共発現ネットワーク解析などの一連のデータ解析をインタラクティブに実行することができます。
- 【統合TV】iDEPを使ってウェブブラウザ上でRNA-seqデータ解析を行う

#### **BioJupies**

- ウェブブラウザ上で、RNA-seqデータ解析の[Jupyter Notebook](#)を作成することができるツール
  - データ解析ツールも多数用意されており、オプションを自由に指定することや作成された図を動かしながらデータの解釈をすることができます。
  - 作成したJupyter Notebookは個別にURLも発行されるので、データの共有や第三者による再現が可能です。
    - Jupyter Notebook (読み方は「ジュピター・ノートブック」)とは、ノートブックと呼ばれる形式で作成したプログラムを実行し、実行結果を記録しながら、データの分析作業を進めるためのツールのこと
- 【統合TV】BioJupiesを使ってウェブブラウザ上でRNA-seqデータ解析を行う

## 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://tgotv.dbcls.jp/>

「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。11月に正式公開を予定している新インターフェースのプレビュー版は[こちら](#)

### 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 (363)  遺伝子 (539)
- タンパク質 (252)
- 配列解析 (288)

## 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 1835 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



### 文献検索／論文執筆支援 (PubMed、Allie、Colil、inMeXesなど) @ AJACSオンライン1

本日の統合TVは、2020年9月16日に開催された統合データベース講習会:AJACSオンライン1から、[ライフサイエンス統合データベースセンター \(DBCLS\)](#) 山本 泰智による「文献検索／論文執筆支援 (PubMed、Allie、Colil、inMeXesなど)」をお送りします。約56分です。

本講習では、主に2020年にリニューアルされたPubMed (バズメド) の効率的な検索方法を紹介するとともに、DBCLSの提供する文献情報活用サービスであるAllieやinMeXes、Colilの使い方について紹介しています。

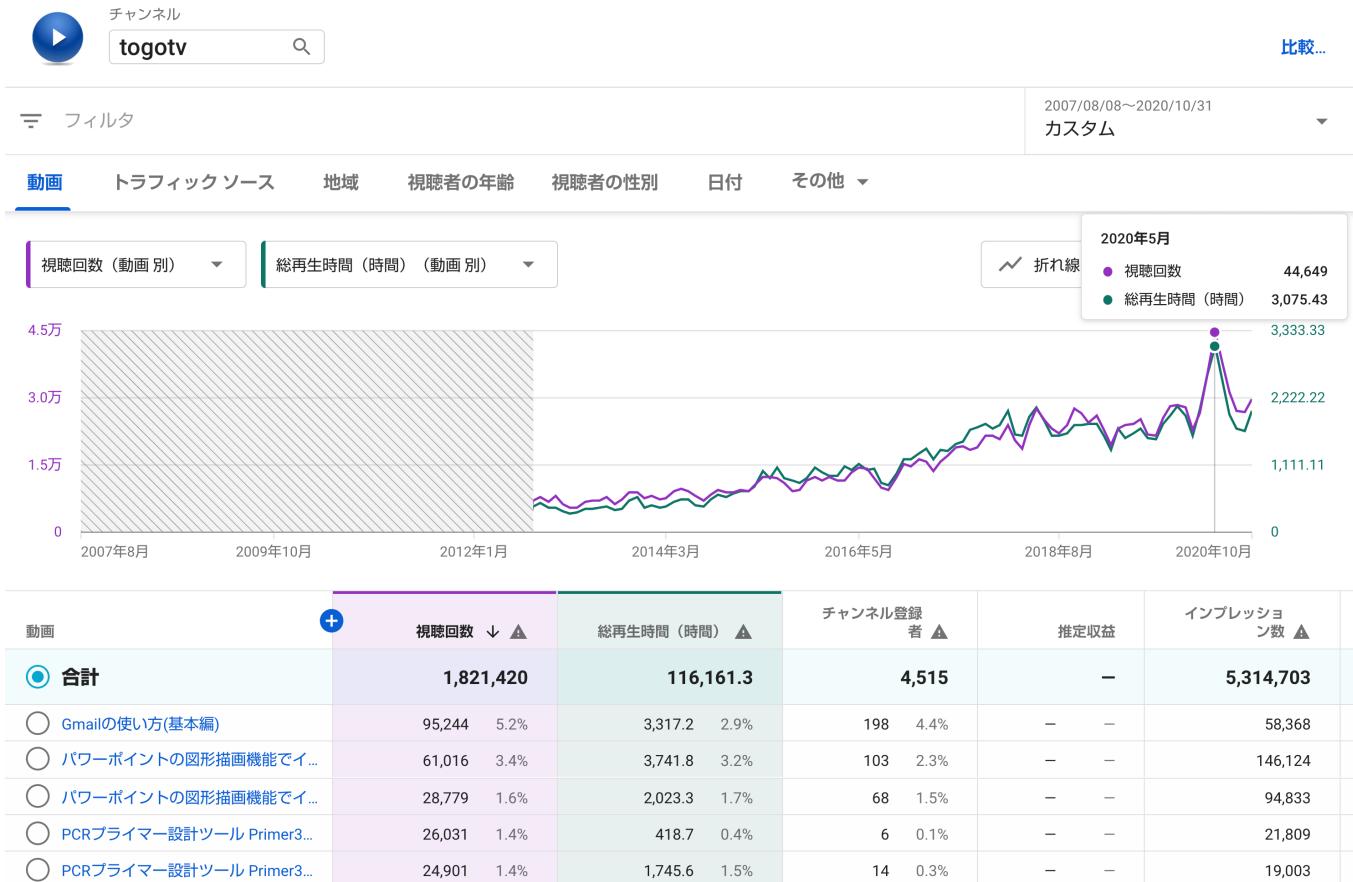


### iMETHYL を使って多型やeQTL情報と紐づいたメチル化情報を活用する

iMETHYLは、いわて東北メディカル・メガバンク機構(Imm)が管理するデータベースです。主にIMMが集計した、日本人のDNAメチル化データが収録されており、DNAメチル化情報やQTL(量的形質遺伝子座) 解析のデータを、変異情報とともにゲノムブラウザで閲覧することができます。2020年8月現在の最新バージョンでは、血球系細胞のデータが公開されています。今回は、乳がんへの強い関連が示されているBRCA1遺伝子と、溶菌作用を持つ酵素であるリゾチームをコードするLYZ遺伝子を例に、その使い方を紹介します。



- 各動画はYouTubeに上がっており、おなじみの再生画面で快適にご覧いただけます。(環境に応じた解像度、倍速表示等)
- 1800本を超える動画が公開されており、YouTube版だけで のべ 1,800,000回以上 再生されています。(2020年10月末現在)



## 2019-05-07 MGIを使ってマウスの遺伝子発現情報を調べる 2019

この動画を引用する際はDOIをご利用ください。DOI: 10.7875/togtv.2019.105

この記事は、クリエイティブ・コモンズ表示 ライセンスの下で利用可能です。利用にあたっては、下記のクレジットを必ず表示してください。  
© 2016 DBCLS TogoTV / CC-BY-4.0

### こちらの番組もオススメ

- 2019-04-03 NCBI GEOのデータセットプラウザを使って公共データの遺伝子発現解析を行う 2019
- 2019-01-05 ChIP-Atlas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～Enrichment Analysisの使い方～
- 2018-12-22 公共のChIP-seqデータをフル活用できるChIP-Atlas @ 第41回日本分子生物学会年会
- 2018-10-21 DatasetTools 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
- 2018-09-18 Bgee を使って、複数の生物種の正常組織における遺伝子発現データを検索、比較、取得する

- 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。
- 動画ファイルのみの一部および一括ダウンロード(NBDC生命科学系データベースアーカイブ)も可能で、オンライン視聴もできます。
- 動画の概要を示すダイジェスト見出しで視聴すべき箇所がわかりやすくなっています。
- 各動画には、恒久的な URL として利用されている DOI (Digital Object Identifier) が付与されています(引用可能)。
- 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。

### 動画以外のコンテンツも拡充中

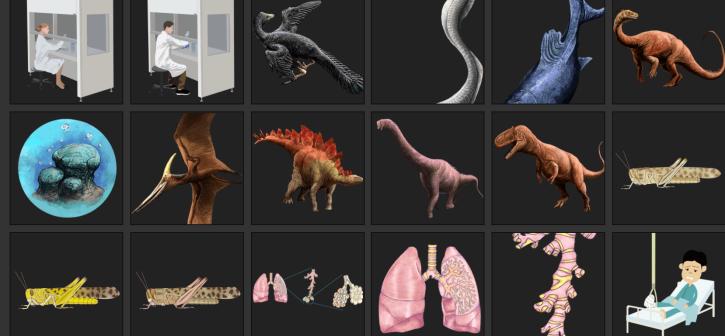
- AJACS講習会資料
  - 2014年8月以降に開催された過去の講習会の資料・テキストと動画が同時閲覧できます。
  - 受講生の復習のみならず、初学者の学習教材として活用できます。
- Togo Picture Gallery(静止画)
  - 誰でも自由に利用可能なライフサイエンス分野のイラストが、統合TVから閲覧、利用することができるようになりました。「自由に使える画像を探す」
  - Togo picture galleryと生物アイコンの全画像700点以上を一覧できます。
  - 研究発表のスライド作成や資料作成等に、ぜひお使いください。

**TOGO PICTURE GALLERY** 

生命科学分野のイラストをだれでも自由に閲覧・利用できるように無料で公開しています。研究発表のスライド作成や資料作成等にお使いください。リクエストは[こちら](#)。

下のタグをクリックするとサムネイルをソートします。全作品のリストは[こちら](#)。

new	モデル生物	細胞	臓器	実験装置	次世代シーケンサー	顕微鏡	実験器具	模式図	活用事例	未分類
生物アイコン	海綿動物門	平板動物門	刺胞動物門	有殻動物門	扁形動物門	環形動物門	軟體動物門	線形動物門	緩歩動物門	節足動物門
珍渦虫動物門	棘皮動物門	半索動物門	脊索動物門	藍色細菌門	綠藻植物門	被子植物門	プロテオバクテリア門	子糞菌門	纖毛虫門	
フィルミクテス門	襟翼毛虫門	脊椎動物亜門	尾索動物亜門	頭索動物亜門	哺乳綱	鳥綱	両生綱	爬虫綱	昆虫綱	緑藻綱
貧膜口綱	襟翼毛虫綱	バシラス綱	γプロテオバクテリア綱	獸亞綱	原獸亞綱	無頸上綱	魚上綱	真黙下綱	後黙下綱	靈長目
恐竜上目	翼竜上目	ネコ目	ウシ目	アブラナ目	イネ目	腸内細菌目				クジラ目
恐竜上目	翼竜上目	ネコ目	ウシ目	アブラナ目	イネ目	腸内細菌目				



## 募集と宣伝

- お探しの動画・静止画が見つからない場合は、[統合TV番組リクエストフォーム](#)でお気軽にリクエストを。
- 統合TVでコンテンツを作成してみたい方も募集中です。
  - 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。(2018年12月)
  - [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。[\(目次\)](#)
  - 「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。

## ウェブサイトリニューアル(11月中を予定)

- プレビューサイトでご覧いただけます。



# TOGO TV

生命科学分野における有用な情報を紹介するウェブサイトです。  
だれでも自由に閲覧し再利用することができます。?

- ・データベースやツールの動画マニュアル
- ・講演や講習会動画、資料
- ・イラスト

キーワードから動画を探す

検索例: [PubMed](#), [BLAST](#), [NGSのハンズオン講習](#)



## スキル別コースから探す



## NEW 新着動画



## 習熟度ややりたいこと別に、とりあえず押さえておきたい統合TVオススメ番組

- DBCLSの提供する便利な各種サービスをレビュー
  - 【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
- PubMed検索のプロによる文献検索のイロハを学ぶ
  - 文献検索／論文執筆支援 (PubMed、Allie、Colil、inMeXesなど) @ AJACSオンライン1
  - PubMedの新インターフェースの紹介もあります。
- 塩基配列解析に関する基礎知識(遺伝子IDとそのデータベース)とゲノム編集について
  - 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
- 次世代シーケンス(NGS)データ解析に必要な基礎知識とリテラシーを学ぶ
  - 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
- 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
  - Morpheus を使って 行列データをアノテーションに基づきわかりやすく表示する
- 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

## まとめ

- つまり食い的ではありますが、遺伝子発現データの生物学的解釈を助ける(ごく一部の)データベース・ウェブツールの使い方を紹介しました。
- データベースやウェブツールは、顕微鏡 や 実験試薬 などと同じ「道具(ツール)」なので使ってナシボです。
- 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です。
  - 正面からしか見られなかったものが横や後ろやナナメから見ることができて初めて気づくことがあるかもしれません。
- 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。
- データベースやウェブツールで困ったら、「統合TV」でまず探して・見てみる・使ってみる
- 研究に役立ったら、ぜひ引用・クレジットを!
  - NBDC/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。NBDC/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やアクセス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!!
  - NBDC関連サービスの活用に関する情報提供フォーム