

# AJACS十勝2 遺伝子発現データの生物学的解釈を助けるデータベース・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

2019年9月25日(水) AJACS十勝2 @ 帯広畜産大学 情報処理センター1階 第1実習室 (北海道帯広市稲田町西2線11番地)

これは統合データベース講習会 AJACS十勝2「遺伝子発現データの生物学的解釈を助けるデータベース・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。© 2019 小野 浩雅, CC-BY-4.0

## 概要

本講習だけ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。続いて、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。

- 個々の遺伝子の発現プロファイルを調べる
  - RefEx
    - 【実習】 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】 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
- 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ますます繋がらなくなってしまいます。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)
    - AJ79 と入力
    - 質問をする (°д°)カタン
- 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)。
  - 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

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 18)(9/11暫定)

統合TVを知っていますか?	人数	割合
知らない	7 名	41 %
知っている	10 名	59 %
回答なし	1 名	- %
---		
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	5 名	29 %
これから実験をする・したい	4 名	24 %
公共データを活用する・したい	3 名	18 %
大規模発現解析の予定はない	5 名	29 %
回答なし	1 名	- %
---		

個々の遺伝子の発現プロファイルを調べる

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す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。

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.dbcis.jp/genelist.php?gene\_name[]=troponin&lang=ja&db=human

結果一覧 18 件中 1 - 10 件を表示

10 最初 < > 最後

ソート: GeneChip 発現量の高い順 リストをクリア リストを見る 0 ダウンロード

遺伝子名 3Dマップ 相対発現量

検索条件

遺伝子名

Include:

- troponin

組織

条件なし

オントロジー

条件なし

ファミリー

条件なし

その他のキー

条件なし

必ず含むデータセット

☐ ALL

☐ EST

☐ GeneChip

☐ CAGE

☐ RNA-seq

この条件で絞り込み

**troponin C type 1 (slow)**

リストに追加する リストから削除する

同義遺伝子名 Troponin C, slow skeletal and cardiac muscles, TNNC1, TNNC, TN-C, troponin C type 1 (slow), TNC

Refseq ID NM\_003280

GeneID 7134

Unigene ID Hs.118845

Probe set ID 209904\_at

EST

GeneChip

CAGE

RNA-seq

**troponin T type 1 (skeletal, slow)**

リストに追加する リストから削除する

同義遺伝子名 ANM, TnTs, sTnT, MGC104241, TNNT1, TNT, troponin T type 1 (skeletal, slow), Troponin T, slow skeletal muscle, Slow skeletal muscle troponin T

Refseq ID NM\_001126132

GeneID 7138

Unigene ID Hs.631558

Probe set ID 213201\_s\_at

EST

GeneChip

CAGE

RNA-seq

**troponin T type 1 (skeletal, slow)**

リストに追加する リストから削除する

同義遺伝子名 ANM, TnTs, sTnT, MGC104241, TNNT1, TNT, troponin T type 1 (skeletal, slow), Troponin T, slow skeletal muscle, Slow skeletal muscle troponin T

Refseq ID NM\_001126133

EST

GeneChip

refex.dbcis.jp/gene\_info.php?lang=ja&db=human&geneID=7138

troponin T type 2 (cardiac) 詳細情報を見る

同義遺伝子名 cTnT, TnTc, Cardiac muscle troponin T, TnTc, troponin T type 2 (cardiac), TNNT2, CMH2, Troponin T, cardiac muscle, MGC3889

発現データ

正常組織・臓器 40分類 細胞株・初代培養細胞・成体組織・胎児組織 (FANTOM5 CAGE)

発現マップ on BodyParts3D 組織40分類別データ Download

相対発現量を、人体 3D 画像にマップしたものです。  
Genechip 組織40分類 の発現パターンを使用しています。

EST

GeneChip

CAGE

RNA-seq

EST

GeneChip

CAGE

RNA-seq

皮膚・骨格筋を表示

アニメーション表示

詳細情報

IDs

遺伝子オントロジー (GO ID)

- 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。
  - 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をご覧ください。
    - <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)](<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](https://doi.org/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/togotv.2014.009](https://doi.org/10.7875/togotv.2014.009)

## 【実習】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  
ボディパーツ3Dで。  
▼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


臓器のアイコンをマウスオーバー

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

Advanced Search

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総合データベースセンターです。

reflex.dbcls.jp/geneist.php?lang=ja&db=human&roku\_valid=1&rk(31)=31&order\_key=score

- 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。)
- 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。
- 「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。
- 「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。
- 「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。
- 「Probe set ID」のリンク先をクリックすると、どのような情報が参照できるでしょうか。
- 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。



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

遺伝子詳細情報

正常組織・臓器 40分類

細胞株・初代培養細胞・成体組織・胎児組織 (FANTOM5 CAGE)

発現マップ on BodyParts3D

組織40分類別データ

相対発現量を、人体 3D 画像にマップしたものです。Genechip 組織40分類の発現パターンを使用しています。

EST

GeneChip

CAGE

RNA-seq

EST

GeneChip

CAGE

RNA-seq

詳細情報

IDs

遺伝子オントロジー (GO ID)

Refseq ID NM\_000508

Gene ID 2243

Unigene ID Hs.351593

Probe set ID 205649 s\_at [HG-U133\_Plus\_2]

Ensembl ID ENSG00000171360

オースログ対応遺伝子

マウス NM\_001111048, NM\_010196

ラット NM\_001008724, NM\_052797

染色体

Biological Process

Cellular Component

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

正常組織・臓器 40分類

細胞株・初代培養細胞・成体組織・胎児組織 (FANTOM5 CAGE)

サンプル名で検索

1/2 ↑ ↓ go

liver, adult

11.22:adult tissue

サンプル名

発現値: サンプルの分類

FANTOM5 CAGEのサンプル情報一覧はこちら

ドラッグしながら移動可能

## 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。

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  
ボディパーツ3Dで。  
▼もっと詳しく

検索結果一覧

結果一覧 470 件中 1 - 10 件を表示

ソート: Tissue Specificity, high

リストをクリア

リストを見る 0

ダウンロード

遺伝子名

3Dマップ

相対発現量

mannose-binding lectin (protein C) 2, soluble (opsonic defect)

リストに追加する

リストから削除する

同義遺伝子名 Mannan-binding protein, MBL, MGC116833, MBL2, MBP-C, Mannose-binding protein C precursor, MGC116832, Mannose-binding lectin, COLEC1, HSMBP, MBP1

Refseq ID NM\_000242

GeneID 4153

Unigene ID Hs.499674

Probe set ID 207256\_at

EST

GeneChip

CAGE

RNA-seq

fibrinogen alpha chain

リストに追加する

リストから削除する

同義遺伝子名 MGC119423, FGA, fibrinogen alpha chain, Fibrinogen alpha chain precursor, Fib2, MGC119422, MGC119425

Refseq ID NM\_000508

GeneID 2243

Unigene ID Hs.351593

Probe set ID 205649\_s\_at

EST

GeneChip

CAGE

RNA-seq

aldo-keto reductase family 1, member D1 (delta 4-3-ketosteroid-5-beta-reductase)

リストに追加する

リストから削除する

同義遺伝子名 3-oxo-5-beta-steroid 4-dehydrogenase, 3o5bred, SRD5B1, AKR1D1, Delta(4)-3-ketosteroid 5-beta-reductase, Aldo-keto reductase family 1 member D1

Refseq ID NM\_005989

GeneID 6718

Unigene ID Hs.201667

Probe set ID 207102\_at

EST

GeneChip

CAGE

RNA-seq

検索条件

遺伝子名

条件なし

組織

• 肝臓

オントロジー

条件なし

ファミリー

条件なし

その他のキー

条件なし

必ず含むデータセット

☐ ALL

☒ EST

☒ GeneChip

☐ CAGE

☐ RNA-seq

この条件で絞り込み

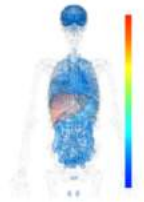
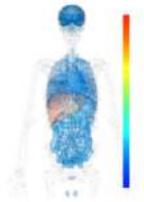

## 12. 様々な条件で検索結果を絞り込むことができます。絞り込み検索は左のバーから行えます。

- 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。



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	<div><div>EST</div><div>GeneChip</div><div>CAGE</div><div>RNA-seq</div></div> 	<div><div>EST</div><div>GeneChip</div><div>CAGE</div><div>RNA-seq</div></div> 	<div><div>EST</div><div>GeneChip</div><div>CAGE</div><div>RNA-seq</div></div> 	

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
遺伝子オントロジー Biological Process	blood coagulation	-	-
	platelet activation	-	-
	platelet degranulation	-	-
	protein polymerization	-	-
	response to calcium	-	-
	signal transduction	-	-
	bile acid biosynthesis	-	-
	bile acid catabolism	-	-
	bile acid metabolism	-	-
	C21-steroid hormone	-	-
	cholesterol catabolism	-	-
	digestion	-	-
	oxidation reduction	-	-
	acute-phase response	-	-
	complement activation	-	-
	complement activation	-	-
	complement activation	-	-
	defense response to	-	-
	defense response to	-	-
	innate immune response	-	-
	killing by host of	-	-
	negative regulation	-	-
	opsonization	-	-
	positive regulation	-	-
	response to oxidative stress	-	-

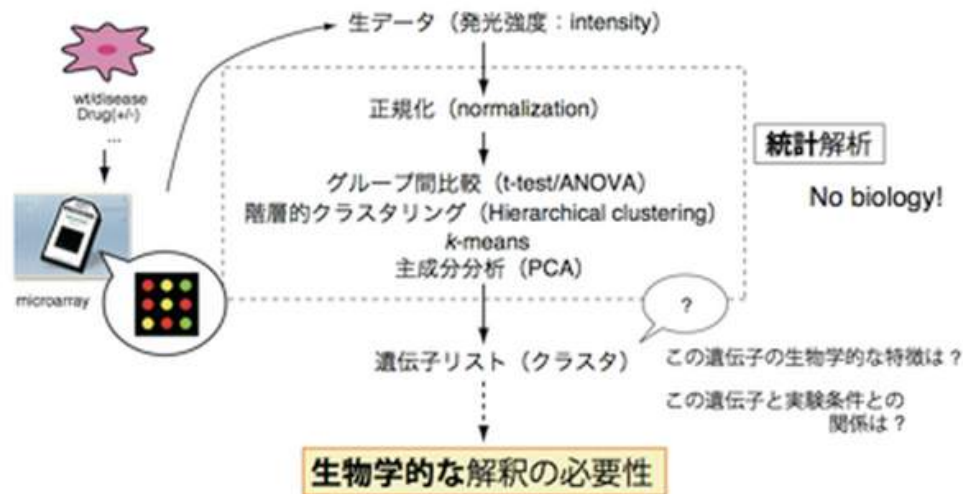
14. 自分の研究テーマに関連する、また興味のある遺伝子について検索してみましょう。

#### 関連するツール

- ExpressionAtlas
  - Expression Atlas は、EMBL-EBI (欧州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](#)
[Peak Browser](#)
[Target Genes](#)
[Colocalization](#)
[Enrichment Analysis](#)
[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 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](#)

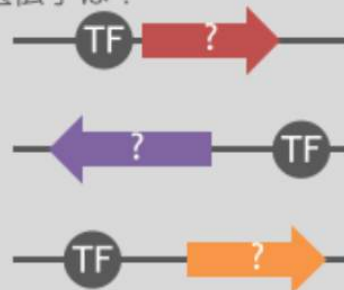
## ① Peak Browser

何がどこに結合？



## ② Target Genes

標的遺伝子は？



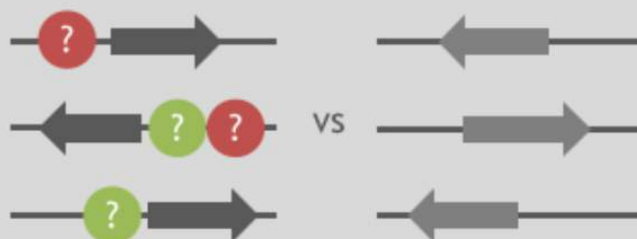
## ③ Colocalization

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、興味のある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、シーケンスモチーフを 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) (https://doi.org/10.15252/embr.201846255)

- Database
  - Oki, S; Ohta, T (2015): ChIP-Atlas. <http://dx.doi.org/10.18908/lsdba.nbdc01558-000>
- Publications citing ChIP-Atlas <http://chip-atlas.org/publications>

## 【実習】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
- 「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。

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.

Tutorial movie

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  $\pm$  TSS  $\pm$  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びます。(ので、講義中は見ただけにしてください)
4. submit したあとの画面


ChIP-AtlasPeak BrowserTarget GenesColocalizationEnrichment AnalysisDocumentationPublicationsFind 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_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:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&amp;format=html">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&amp;format=html</a>
Download TSV:	<a href="http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&amp;format=tsv">http://ddbj.nig.ac.jp/wabi/chipatlas/wabi_chipatlas_2018-1104-1735-21-275-473936?info=result&amp;format=tsv</a>



KYUSHU UNIVERSITY

IN COLLABORATION WITH



DBCLS

Databases Center for Life Science

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?
<a href="#">SRX100505</a>	TFs and others	HNFA4	Liver	Hep G2	21259	118/218	3453/18327	-30.0	-26.7	2.87	TRUE
<a href="#">SRX100544</a>	TFs and others	EP300	Liver	Hep G2	24334	122/218	4026/18327	-26.5	-23.3	2.55	TRUE
<a href="#">SRX100497</a>	TFs and others	RXRA	Liver	Hep G2	13022	92/218	2362/18327	-25.7	-22.5	3.27	TRUE
<a href="#">SRX100449</a>	TFs and others	HNFA4G	Liver	Hep G2	15919	102/218	3119/18327	-23.2	-20.1	2.75	TRUE
<a href="#">SRX100448</a>	TFs and others	FOXA2	Liver	Hep G2	45130	135/218	5342/18327	-22.6	-19.6	2.12	TRUE
<a href="#">SRX190234</a>	TFs and others	CEBPB	Liver	Hep G2	12483	84/218	2308/18327	-21.1	-18.2	3.06	TRUE
<a href="#">SRX100506</a>	TFs and others	FOXA1	Liver	Hep G2	50941	141/218	6086/18327	-20.4	-17.5	1.95	TRUE
<a href="#">SRX100477</a>	TFs and others	FOXA1	Liver	Hep G2	40732	127/218	5256/18327	-18.8	-15.9	2.03	TRUE
<a href="#">SRX100552</a>	TFs and others	SP1	Liver	Hep G2	19032	129/218	5594/18327	-17.4	-14.7	1.94	TRUE
<a href="#">SRX150698</a>	TFs and others	HNFA4	Liver	Hep G2	10069	76/218	2336/18327	-16.0	-13.3	2.74	TRUE
<a href="#">SRX150701</a>	TFs and others	CEBPB	Liver	Hep G2	18637	105/218	4173/18327	-15.5	-12.8	2.12	TRUE
<a href="#">SRX018625</a>	TFs and others	HNFA4	Liver	Hep G2	2654	33/218	516/18327	-13.9	-11.3	5.38	TRUE
<a href="#">SRX1531773</a>	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

Previous12345...223Next

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\\_2018-1104-1735-21-275-473936?info=result&format=html](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

▪ → 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。
- https://raw.githubusercontent.com/AJACS-training/AJACS79/master/03\_hono/README.md

14/19

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を使ってマイクロアレイデータを解析する 2012](#)
- [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もある。

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

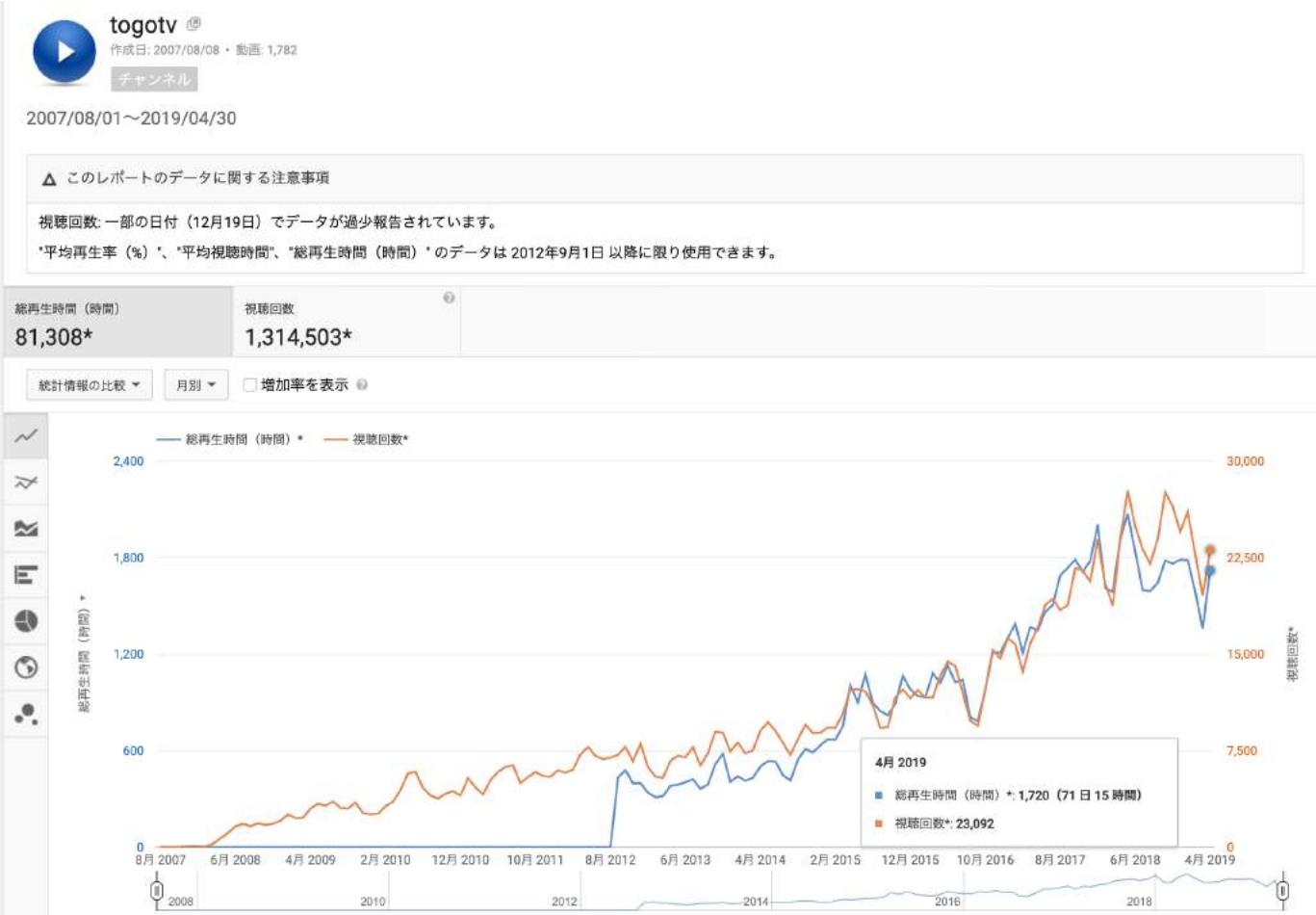
統合TV

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/>



- 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)
- 1600本を超える動画が公開されており、[YouTube版](#)だけで のべ 1,300,000回以上 再生されています。(2019年4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



共焦点レーザー走査型顕微鏡で見た結果です

YouTube版を視聴できない方はオリジナル版ファイル(mov形式)をダウンロードして、ご覧ください。

MGI (Mouse Genome Informatics; 「エムジーアイ」と発音します) は米国メーン州にあるThe Jackson Laboratory (ジャクソン研究所) が提供しているマウスに関する遺伝子、ゲノム、生物学的な情報を提供する統合データベースです。Tnt2 (troponin T2, cardiac) を例にして、MGIに登録されているさまざまな実験条件で得られたマウスの遺伝子発現情報を閲覧する方法を紹介します。

▶ 見どころダイジェスト

1. MGIにアクセスし、目的の遺伝子情報を検索する。(0:11)
2. 発現情報のデータを実験方法別に調べる。(1:11)
3. Western blotの結果を見る。(1:40)
4. RT-PCRの結果を見る。(2:29)
5. RNA in situの結果を見る。(3:29)
6. Northern blotの結果を見る。(4:08)
7. Immunohistochemistryの結果を見る。(5:02)
8. 組織、細胞ごとの発現量を調べる。(6:16)
9. 各組織での発現情報を確認する。(6:45)
10. 実験結果の図から発現量を調べる。(7:00)

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

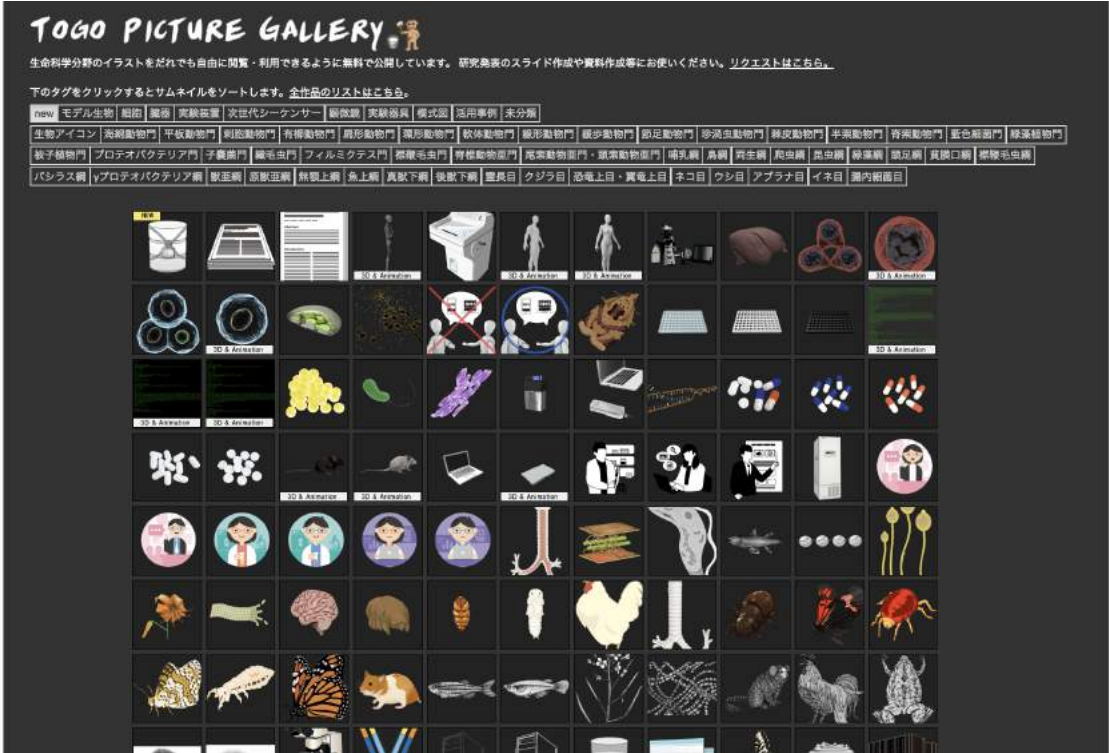
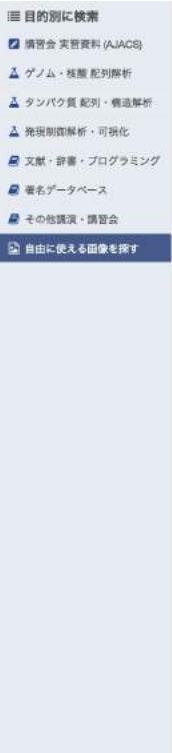
- 2019-04-03 NCBI GEOのデータセットブラウザを使って公共データの遺伝子発現解析を行う 2019
- 2019-01-05 ChIP-Atlasを使って興味ある遺伝子リストを制御する可能性の高い転写因子を調べる ~Enrichment Analysisの使い方~
- 2018-12-22 公共のChIP-seqデータをフル活用できるChIP-Atlas @ 第41回日本分子生物学会年会
- 2018-10-21 Dataset2Tools でオミックスデータとその解析事例、計算ツールを検索し、再現性の高い再解析を行う
- 2018-09-18 Bgee を使って、複数の生物種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、オフ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。
  - 本講義中も、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。

## 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。





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でコンテンツを作り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。
  - 「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.medsico.jp/books/products/detail.php?product\\_id=3665](https://www.medsico.jp/books/products/detail.php?product_id=3665)

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
  - [文献情報を用いたサービスを活用する @ AJACS越後](#)
- 塩基配列解析に関わる基礎知識(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@総合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
  - 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](#)
- 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ーションに関連する情報を調べる

---

## まとめ

---

- つまみ食いのではありますが、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)」なので使ってナンボです。
  - 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で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、「総合TV」でまず探して・見てみる
  - 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](#)
-