

平成 29 年 8 月 29 日

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

国立遺伝学研究所

**【報道発表資料】**

本件の取り扱いについては、下記の解禁時刻以降でお願い申し上げます。

**日本時間 平成 29 年 8 月 30 日午後 2 時**

## 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」を開発

ー最大 556 個の臓器・細胞における遺伝子発現状況をワンタッチでー

- ▶ **生命科学の公共データベースには多種多様なデータ（遺伝子発現等）が大量に蓄積されているが、利用が困難**
- ▶ **生物種や臓器別の各遺伝子の遺伝子発現量を簡単に検索、閲覧できるウェブツール「RefEx」は、遺伝子解析研究をサポートし、生命科学や医学研究への幅広い貢献が期待**
- ▶ **今後も世界的な大規模研究プロジェクトの成果を取り込み、さらに機能を強化**

### 発表概要

現在、生命科学分野においては誰でも利用可能なデータが公共データベースとして多数存在しているものの、実際にそれらを自らの研究に利用しようとしたときに、どれを使ったらよいかわからないといった問題があります。とくに遺伝子発現データは、DNA マイクロアレイの発明によってゲノム規模の測定が可能となってから、さまざまな研究グループによって異なる測定手法を用いて産生されたデータが指数関数的に蓄積していました。大量の遺伝子発現データの中から、まずどれを選び、調べればよいのかの指針になりうる代表的な遺伝子発現量データセットあるいはリファレンス（参照）データが必要とされている状況でした。

このたび、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（DBCLS）の小野 浩雅 特任助教、坊農 秀雅 特任准教授、情報・システム研究機構 国立遺伝学研究所の小笠原 理 特任准教授、大久保 公策 教授の研究グルー

ブは、遺伝子発現解析の基準となる各遺伝子の遺伝子発現量を簡単に検索、閲覧できるウェブツール「RefEx」(<http://refex.dbcls.jp>)を開発しました。複数の遺伝子発現計測手法によって得られた哺乳類の正常組織、細胞等における遺伝子発現データを収集し並列に表現することによって、各組織における遺伝子発現状況を計測手法間の差異とともに直感的に比較できることが特長です。RefEx を用いることで、生命現象の解明や医薬品の開発等につながる研究成果の解釈などが容易になり、生命科学研究の進展に大いに寄与することが期待されます。

## 発表内容

### 1. 背景

生命科学分野では、1990 年代後半から、急速に進展したゲノムプロジェクトやオミクスプロジェクトにより大量のデータが産み出されるようになり、その結果、世界中で数千をこえる多様なデータベースが研究成果として公開されています。自らが生み出すデータに加えて、既存の公開データをうまく使いこなすことが研究の効率的な進展に不可欠になっていますが、「必要なデータベースが見つからない」「使い方がよくわからない」「データを組み合わせてより高度な解析ができないか」など不便を訴える研究者の意見も多く、データベースを効率よく利用するための環境整備は充分ではないという現状があります。

そこで、ライフサイエンス統合データベースセンター（DBCLS）では、2007 年の発足以来、データベースの再利用性を高めるための情報技術の研究開発を進めてきました。とくに遺伝子発現データについては、仮説の構築や研究計画の立案、実験データの解釈などさまざまな状況で幅広い分野の研究者に利用される汎用的なデータであることから、再利用に資する代表的なデータを選び出して整理し、それらを並べて閲覧できるウェブインターフェースを整備してきました。2011 年よりそのウェブインターフェースを RefEx（Reference Expression dataset）と名付け、遺伝子発現量を手軽に可視化できるウェブツールとして開発を進めてきました。

### 2. 研究の成果


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（DBCLS）の小野浩雅特任助教、坊農秀雅特任准教授、情報・システム研究機構 国立遺伝学研究所の小笠原理特任准教授、大久保公策教授の研究グループは、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」(<http://refex.dbcls.jp>)を開発しました（ 1）。



図 1 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」

RefEx は、トップページ（検索フォーム）、検索結果一覧、個別の遺伝子の詳細情報、の 3 つを柱として構成されています。ヒト、マウス、ラットの 3 種の生物種に対応しており、その切り替えは、トップページ左上部のアイコンをクリックして行います。もっとも基本的なキーワード・遺伝子名検索では文字を入力する度に検索語の候補が提示されるので、それらから選択することで容易にキーワード入力を行うことができます。また、「転写因子」や「G タンパク質共役受容体」、「2 番染色体」などのように、ある分類に属する遺伝子群についてまとめて検索・比較できるよう整理されています。さらに、さまざまな実験における比較対照などに用いられる『組織特異的遺伝子』を測定データから独自に算出し、組織ごとに一覧することができます。Advanced search では、複雑な検索条件を一度に指定することが可能であり、あらかじめ ID 情報などが手元にある場合には、目的とするデータに簡単に行き着くことができます。

検索結果一覧ページでは、項目別ソートおよび絞り込み検索が可能で、検索条件を柔軟に入れ替えながら検索結果を閲覧・比較することができます。検索結果一覧および個別の遺伝子の詳細情報ページでは、組織間の比較と測定手法間（EST（注 1）、GeneChip（注 2）、CAGE（注 3）、RNA-seq（注 4））の比較を両立させた相対発現量が棒グラフで示されるとともに人体の 3D モデルに発現量を反映させたヒートマップが表示されます。またリスト機能を使えば、検索結果の個別の遺伝子について一時的に保存しておくことができます。リストに追加した遺伝子は、最大でその 3 つについて、40 分類の組織・臓器における発現データを比較しながら、遺伝子に付与された機能に関する注釈情報（Gene Ontology（注 5）他）を見比べることができます。これらの機能は、新たな知識発見あるいは仮説の構築をサポートします。詳細情報ページに記載された種々の ID には、それぞれ RefEx の内部リンクやオリジナルのデータベ

スサイトへの外部リンクが貼られており、同じ分類に属する遺伝子を再検索したり、RefEx 自体を遺伝子検索の起点とすることもできます。

検索結果一覧や詳細情報ページのデータはいずれもダウンロードすることが可能で、手元のデータと参照することも、それらを使った再解析も自由にできます。

さらに最近、理化学研究所の FANTOM (注 6) プロジェクト 5 (FANTOM5) によって大量の遺伝子発現データが公開され、RefEx に掲載されました。(FANTOM5 の成果公開については、理化学研究所から 8 月 30 日に以下のタイトルで公表されています。「網羅的な RNA アトラスの基盤データを詳細に記述 – FANTOM5 によるオープンサイエンスを加速 –」) これらの FANTOM5 データは、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得られ、それらを閲覧できるウェブサイトも公開されていますが、多くの生命科学研究者にとってはその規模の大きさと複雑さから再利用が難しいものでした。RefEx を通じて、これらの高精度かつ広範囲な組織や臓器 (ヒトで 556 種) における遺伝子発現データについても可視化および比較を簡単に行えるようになりました。

なお、RefEx の使い方は、統合 TV (注 7) のチュートリアル動画としても紹介されています。

(<http://doi.org/10.7875/togotv.2014.009>)

### 3. 波及効果

RefEx が提供するすべてのデータは、クリエイティブ・コモンズライセンス (注 8) のもとで、オープンデータとして自由にダウンロードおよび再利用することができます。外部の研究データレポジトリ「figshare」(注 9) にも全てのデータが DOI (注 10) 付きで公開されています (<https://doi.org/10.6084/m9.figshare.c.3812815>)。RefEx は生命科学データの共有および再利用の活用例のひとつであり、データ駆動型研究のためのデータセット、ウェブツールとしてだれでも自由に使うことができます。

RefEx を利用することで、研究者は研究対象とする遺伝子が平常時にどの組織、細胞でどの程度発現しているのかについて、自ら実験をすることなく確認することができます。また、研究者がしばしば遭遇する馴染みのない遺伝子について、一般的には個別の研究論文における実験データや記述などからそれらの生物学的特徴を類推したりしますが、RefEx では実験デザインに左右されない大規模かつ網羅的な測定データから研究者自身の目でそれらを簡単に確認することができます。さらに、研究者の用意した複数の遺伝子 ID について一括で検索できる機能を備えているほか、リスト機能を用いて遺伝子の詳細データを並列に比較することができるため、遺伝子発現解析などで見出された遺伝子群の関係性を知るためのツールとしても有用です。このような活用法によって、RefEx は遺伝子発現解析のための強力なウェブツールとして生命科学および医学研究に幅広く貢献することが期待されます。

一般社会においても、新聞の見出しなどで“やせる遺伝子、発見”のような表現が見かけられるようになりましたが、一つ一つの遺伝子がどのような働きをもつのかについて、科学研究に裏打ちされた正確な情報源が求められています。将来的には、研究者だけでなく、一般の人も遺伝子について検索することが日常的になったときに、その第一選択肢として使われることを目指しています。

#### 4. 今後の予定

今後は、世界各地で進められている遺伝子発現に関する大規模研究プロジェクト（FANTOM、GTEx、Human Cell Atlas など）を中心に、高精度かつ広範囲な遺伝子発現データを収集し、統合することによって、より有用性の高い参照データの作成を進める予定です。

また、それらの参照データを簡単に検索したり、発現データ同士を詳細に比較したりすることを可能にする直感的なウェブインターフェースの開発を進めます。

※本ツールは国立研究開発法人科学技術振興機構（JST） が実施するライフサイエンスデータベース統合推進事業の一環として、JST バイオサイエンスデータベースセンターと DBCLS の共同研究により開発されたものです。

## 掲載論文

RefEx, a reference gene expression dataset as a web tool for the functional analysis of genes  
Ono H, Ogasawara O, Okubo K, Bono H  
Scientific Data, 4:170105 DOI: 10.1038/sdata.2017.105

## 用語解説

### 注 1 EST

Expressed Sequence Tag の略。 mRNA 断片配列の数を塩基配列解読することで数え、その遺伝子発現量を測定する手法。

### 注 2 GeneChip

Affymetrix 社（現 ThermoFisher Scientific 社）のマイクロアレイ。公共データベース中の多くのデータは現状この GeneChip によって測定された遺伝子発現データである。

### 注 3 CAGE

Cap Analysis of Gene Expression の略。 mRNA の 5'末端を選択的に配列解読する方法で、転写開始点の特定に用いられるほか、その末端に付与した印の数を数えることで遺伝子発現量の測定も可能となる。

#### **注 4 RNA-seq**

次世代シーケンサーによって塩基配列解読することで遺伝子発現量を測定する方法。

#### **注 5 Gene Ontology**

遺伝子機能情報をコンピュータ上で扱うための、遺伝子に関する制限語彙とその分類分けのこと。

#### **注 6 FANTOM**

Functional ANnotation of Mammalian genome の略で、理化学研究所が中心となって行われてきた、転写産物の機能解析やアノテーションなどを行う国際的な共同研究プロジェクト。

#### **注 7 統合 TV**

DBCLS で作成している生命科学分野のデータベースやウェブツールの使い方を動画で配信するサービス。

#### **注 8 クリエイティブ・コモンズライセンス**

法律や技術に関する専門的な知識がなくても、作成者の希望する条件を組み合わせることで、インターネットを通じて著作物を世界に発信することができるライセンスシステムである。利用者にとっては、作者が指定した条件を守る限り著作権侵害を心配する必要がない。

#### **注 9 figshare**

研究者が研究の過程で得た図表やその他の付随するデータをクラウド上で公開するシステム。

#### **注 10 DOI**

Digital Object Identifier の略で、インターネット上のドキュメントに恒久的に与えられる識別子のこと。

### **問い合わせ先**

#### **【研究内容に関すること】**

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

特任准教授 坊農 秀雅（ぼうのう ひでまさ） e-mail : bono@dbcls.rois.ac.jp

特任助教 小野 浩雅（おの ひろまさ） e-mail : hono@dbcls.rois.ac.jp

#### **【広報に関すること】**

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

広報担当 特任准教授 箕輪 真理（みのわ まり）

Tel : 04-7135-5508 FAX : 04-7135-5534 e-mail : public\_relations@dbcls.rois.ac.jp