

国際塩基配列データベース見聞録

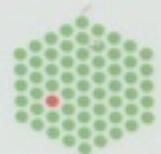
大学共同利用機関法人 情報・システム研究機構
データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター

坊農 秀雅

bono@dbcls.rois.ac.jp



EMBL-EBI



本務: エンドユーザ向け公共データベース統合利用環境の構築ならびにその利用普及活動

- DBCLS勤続もうすぐ10年目
 - 三島歴3年目
 - 1年契約の研究員
 - 非公式なリーダー, not ボス
 - 各地での講習会講師
- DRY解析教本出しました
- 低酸素・酸化ストレスの共同研究
 - パーキンソン病モデルカイコ



NMRはじめました



平成28年5月10日
北海道大学
科学技術振興機構(JST)
慶應義塾大学

がんになりにくい長寿ハダカデバネズミから初めてiPS細胞作製に成功 ～二重の防御で腫瘍を作らないことを発見～

ポイント

- 長寿命で、がんになりにくい体質のハダカデバネズミからiPS細胞を作製することに初めて成功した。
- マウスやヒトのiPS細胞は、未分化な細胞が混入すると腫瘍(奇形腫)を形成する問題があるが、ハダカデバネズミのiPS細胞は未分化な状態で移植しても腫瘍を形成しなかった。
- ハダカデバネズミだけが持つ、初期化やがん化を二重に防ぐメカニズムを明らかにした。将来は、人間の健康長寿やがんの予防に役立つと期待される。

北海道大学 遺伝子病制御研究所の三浦 恭子 講師、慶應義塾大学 医学部 生理学教室の岡野 栄之 教授らの研究グループは、約30年の長寿命で、がんになりにくい体質の「[ハダカデバネズミ^{注1\)}](#)」から*IPS細胞^{注2)}*を作製することに世界で初めて成功しました。

非モデル生物(ハダカデバネズミ: Naked Mole Rat)における遺伝子機能解析(mRNA-seq)に貢献

朝日新聞
DIGITAL

トップニュース スポーツ カルチャー 特集・連載 [HEW](#) オピニオン 写真

新着 社会 政治 経済・マネー 國際 テック&サイエンス 教育 環境・エネルギー 医療・健康

トピックス オバマ米大統領 「パナマ文書」 熊本地震 どすこいタイムズ ビル&メリンダ・ゲイツと考

朝日新聞デジタル > [記事](#)

研究・開発 医療・病気

apital 医療ニュース ニュース・フォーカス

シリーズ: その他

がん化しにくいiPS細胞、ネズミから作成 北大など

竹石涼子 2016年5月11日06時15分

シェア 0 ツイート [list](#) 1 ブックマーク 1 メール 印刷



ハダカデバネズミ=北海道大提供

アフリカ原産で、極めてがんになりにくいとされるハダカデバネズミから iPS細胞 を作ることに、 北海道 大や 慶應大 などのチームが成功した。10日付の英科学誌ネイチャー・コミュニケーションズ(電子版)に論文を発表した。がん化しにくいのが特徴で、 iPS細胞 を使った 再生医療 の安全性向上につながる可能性があるという。

北大の三浦恭子講師らのチームは、ハダカデバネズミの皮膚から iPS細胞 を作り、性質を調べた。

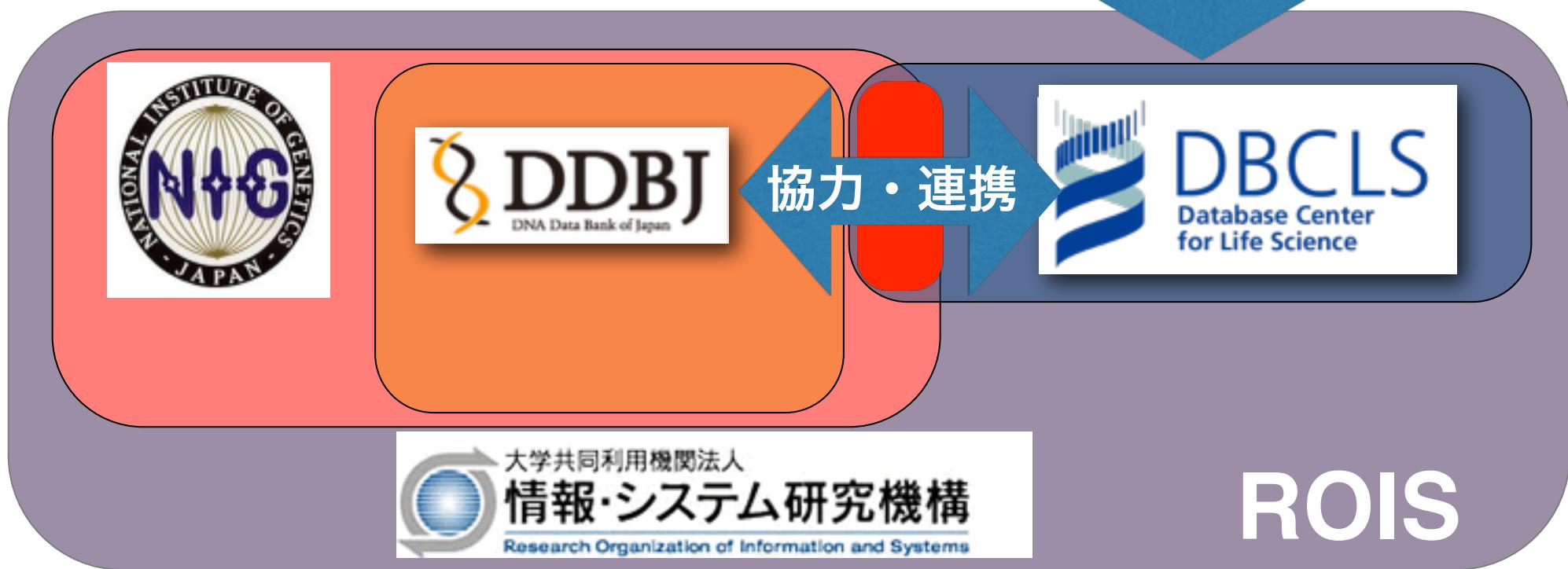
[PR]

その結果、ヒトやマウスの細胞から作った iPS細胞 ではそのまま移植するとがんができるのに対し、ハダカデバネズミの iPS細胞 はがんを作らなかった。

遺伝子を解析したところ、ハダカデバネズミの iPS細胞 では、がん化を防ぐ特定の遺伝子が働いていたほか、細胞のがん化を引き起こす [がん遺伝子](#) も変化して働かないようになっていた。また、 iPS細胞 になる過程で、がん化しにくいものだけが生き残ることもわかった。

三浦講師は「今回の成果を応用することで、より安全に iPS細胞 を作る方法の開発につながる可能性がある」と話している。

赤の部分がDBCLS三島。
平成26年4月から、DDBJ
と同じ建物で活動、協力・
連携を進めている。



WELCOME TO THE

EMBL-EBI
SOUTH
BUILDING



HOME OF THE

elixir
ELIXIR HUB







EMBL-EBI

EUROPEAN
BIOINFORMATICS
INSTITUTE





OUR MISSION

To provide **freely available data** and bioinformatics services to all facets of the scientific community in ways that **promote scientific progress**



Bigger and Bigger data

Keep the data flowing

Modern research requires rapid and efficient access to shared 'big data'. But storing, moving and computing massive datasets present overwhelming challenges to off-the-shelf informatics, for hardware as well as software.

EMBL-EBI leads the DNA sequencing community in its quest to increase the fluidity of data movement, storage and access for large-

scale computation. Raw DNA sequences are flowing into the European Nucleotide Archive (ENA) and European Genome-phenome Archive (EGA) at exponential rates. To keep this flood of data under control, sophisticated data compression techniques have been urgently needed.

Nucleotide Sequences



CRAM

If you've ever worked with raw sequence data, you may be familiar with 'video streaming'. This is a data-compression method that describes which video frames to retransmit, and which can be buffered. For example, you will want to see the faces of actors you probably won't miss if some of the data background is obscured.

CRAM is similar to this form of video, in that it compresses data using a method called CRAM allows DNA sequence data sets to be moved between and between sites of smaller and more fastidious packages. The method is compressed using new network friendly protocols, which allow for the efficient and secure transfer of compressed data across physical networks like GIGANT and JNET.

CRAM is an innovative way to store and use huge quantities of human and other genome data, improving its accessibility and reducing costs. As an effort involving the DNA sequencing community, it will ensure that future big data can be stored and retrieved efficiently and securely over the coming decades.



Reminder to deposit DNA sequences

nature.com : Publications A-Z index : Browse by subject

Cart : Login : Register



Search

Go

► Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 533 > Issue 7602 > Correspondence > Article

ARTICLE PREVIEW

[view full access options >](#)

NATURE | CORRESPONDENCE



Databases: Reminder to deposit DNA sequences

Steven L. Salzberg

Nature 533, 179 (12 May 2016) | doi:10.1038/533179a

Published online 11 May 2016

Citation

Reprints

Rights & permissions

Article metrics

Editors' pick

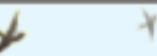


The truth about reproducibility: A survey of over 1,500 researchers shows that 70% have failed to reproduce another's experiment >

Science jobs

Science events

naturejobs.com



Research Engineer / Research Scientist in Renewable Energy

King Fahd University of Petroleum & Minerals

Faculty Positions at IDG / McGovern Institute for Brain Research at Peking University

As members of the Advisory Committee to the International Nucleotide Sequence Database Collaboration (INSDC), which includes the DNA Data Bank of Japan (DDBJ), European Nucleotide Archive (ENA) and GenBank databases, we wish to remind the research community of the importance of depositing complete DNA-sequence data in these databases on publication...

Reminder to deposit DNA sequences

Science

Home News Journals Topics Careers

Search



Log in • My account • Contact Us

Subscribe Renew my subscription • Sign up for alerts • Librarians



19



0

LETTERS

Reminder to deposit DNA sequences

Mark Blaxter¹, Antoine Danchin², Babis Savakis³, Kaoru Fukami-Kobayashi⁴, Ken Kurokawa⁵, Sumio Sugano⁶, Richard J. Roberts⁷, Steven L. Salzberg^{8,*}, Chung-I Wu^{9,10}

+ Author Affiliations

*Corresponding author. Email: salzberg@jhu.edu

Science 13 May 2016:
Vol. 352, Issue 6287, pp. 780
DOI: 10.1126/science.aaf7672

Article

Figures & Data

Info & Metrics

eLetters

PDF

You do not have access to the full text of this article, the first page of the PDF of this article appears below.

INSIGHTS | LETTERS

LETTERS

Edited by Jennifer Sills

Reminder to deposit



for the article's inappropriate placement and to henceforth refrain from publishing articles that suggest Crimea to be under Russian jurisdiction.

The News story invokes a picture that is very different from the reality in Crimea: Occupation authorities prosecute everyone who dares express an opinion opposing the



Science

Vol 352, Issue 6287
13 May 2016

Table of Contents
Print Table of Contents
Advertising (PDF)
Classified (PDF)
Masthead (PDF)

ARTICLE TOOLS

Email

Download Powerpoint

Print

Save to my folders

Alerts

Request Permissions

Citation tools

Share

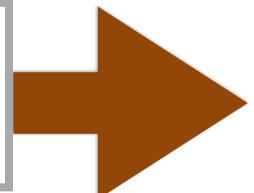
SIMILAR ARTICLES IN:

- PubMed
- Google Scholar

International Nucleotide Sequence Database Collaboration (INSDC)

	DDBJ	EMBL-EBI	NCBI
Next Generation reads	Sequence Read Archive		Sequence Read Archive
Capillary reads	Trace Archive		Trace Archive
Annotated sequences	DDBJ	European Nucleotide Archive (ENA)	DDBJ
Samples	BioSample		BioSample
Studies	BioProject		BioProject

アーカイブされているが、
検索や再利用し辛い



DBCLS SRA

Yellow pages for Sequence Read Archive(SRA)

<http://SRA.dbcls.jp/>





Trends & Search SRA data

→ [for more detail](#)

Species

- █ Homo sapiens
- █ Mus musculus
- █ human gut metagenome
- █ Oryza sativa
- █ Streptococcus pneumoniae

314068


 Search

Study Type

- █ Whole Genome Sequencing
- █ Other
- █ Transcriptome Analysis
- █ Metagenomics
- █ Population Genomics

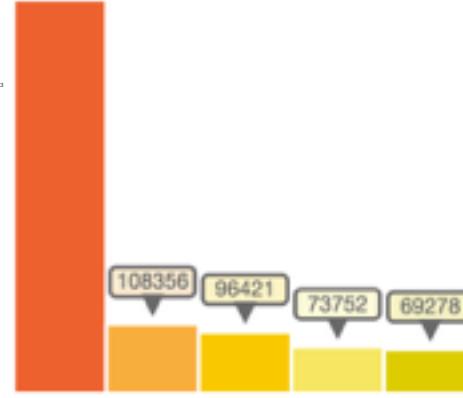
28081


 Select your option

Platform

- █ Illumina HiSeq 2000
- █ Illumina MiSeq
- █ 454 GS FLX Titanium
- █ Illumina HiSeq 2500
- █ Illumina Genome Analyzer II

616085


 Search

Usage

- i Click bars or bubbles and you can see more details of selected data.
- i Input keywords and you can view ratio of feature selection.
- i You can see result of combined search criteria too.

Search Conditions

Free Keyword

 Free Keyword

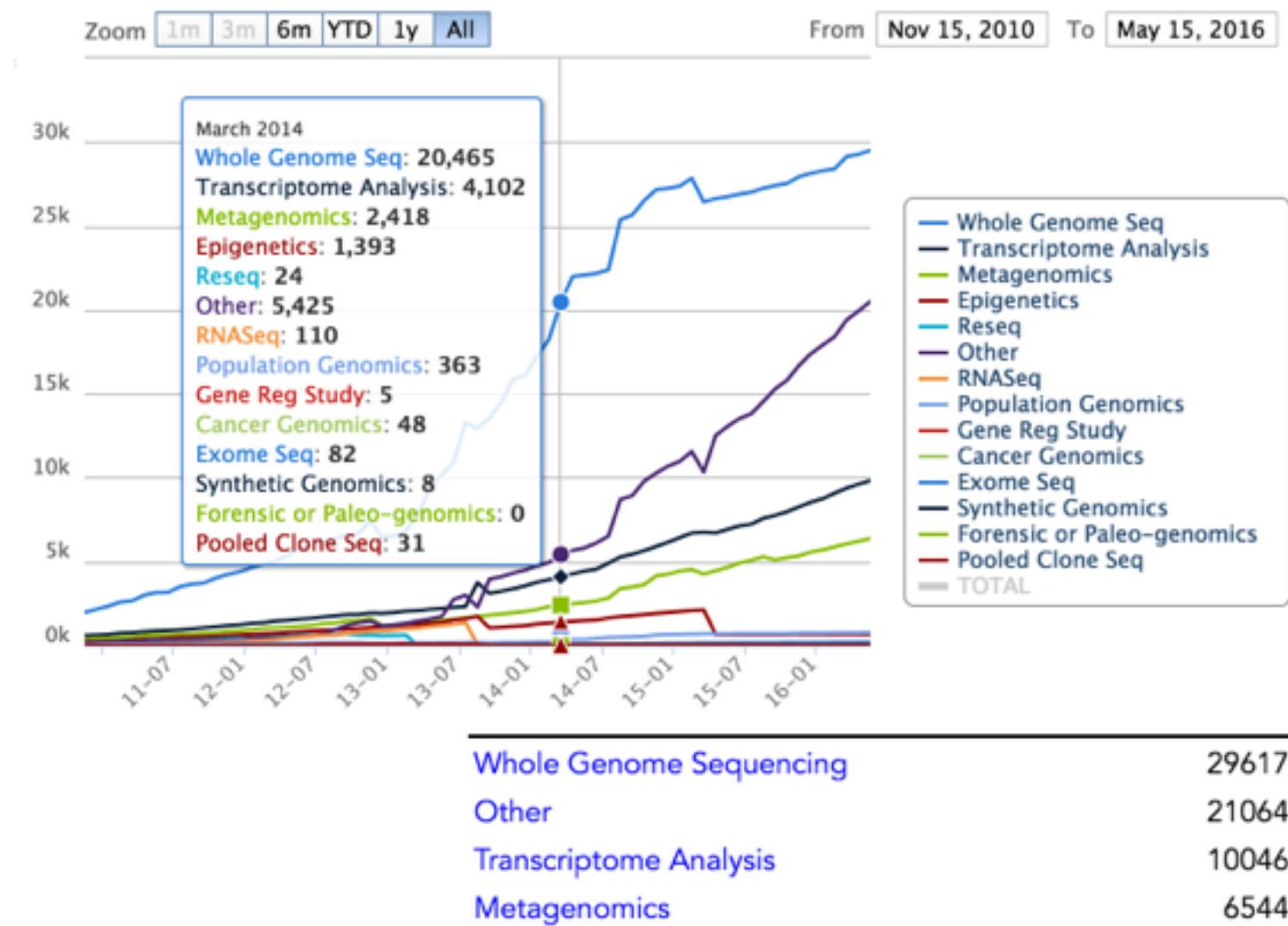


DBCLS SRA

text

Trends in SRA data

By study types



Licensed under CC-BY 4.0 ©2016 HIDEMASA BONO(DBCLS)

NBDCヒトデータベースについて

ヒトに関するデータは、次世代シークエンサーをはじめとした解析技術の発達に伴って膨大な量が産生されつつあり、それらを整理・格納して、生命科学の進展のために有効に活用するためのルールや仕組みが必要です。

国立研究開発法人科学技術振興機構(JST)バイオサイエンスデータベースセンター(NBDC)では、個人情報の保護に配慮しつつヒトに関するデータの共有や利用を推進するために、ヒトに関する様々なデータを共有するためのプラットフォーム『NBDCヒトデータベース』を設立するとともに、国立遺伝学研究所DNA Data Bank of Japan(DDBJ)と協力して、ヒトに関するデータを公開しています。

本Webサイトを通じて、ヒトに関する[データの利用](#)及びヒトに関する[データの提供](#)を行なうことができます。データ共有についての概要は[こちら](#)をご参照下さい。

新着情報

2016/03/29

制限公開データ(Type I)1件
(hum0045.v1)が公開されました

2016/03/09

オープンデータ1件(hum0029.v1)が公開されました

2016/03/02

制限公開データ(Type I)1件
(hum0042.v1)が公開されました

[▶ ニュース一覧へ](#)

利用可能な研究データ一覧

データ利用方法は[こちら](#)をご覧下さい。

- JGA(Japanese Genotype-phenotype Archive)
- EGA(European Genome-phenome Archive)
- dbGaP(The database of Genotypes and Phenotypes)

§ JGA Meta Viewer

Study List

Studies

Menu

[Studies](#)[Datasets](#)[JGA Home](#)

Accession	Title
JGAS000000000001	Identification and Characterization of Cancer Mutations in Japanese LungAdenocarcinoma without Sequencing of Normal Tissue Counterparts
JGAS000000000002	Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing.
JGAS000000000004	Mutational Analysis Reveals the Origin and Therapy-driven Evolution of Recurrent Glioma
JGAS000000000005	Gene expression of human Th17 cells before and after activation
JGAS000000000006	DNA methylation dynamics during early human development
JGAS000000000008	Research for candidate genes of splenic epidermoid cyst
JGAS000000000009	Whole exome sequence analysis in multiple system atrophy



INSDC

1. Next generation reads(SRA)
2. Capillary reads
3. **Annotated sequences**
4. Samples(BioSample)
5. Studies(BioProject)

Refseq

RefSeqのデータを
検索したい

統合遺伝子検索
GGRNA

BLASTで不可能な短い
断片配列で検索したい

Ultrafast sequence search
GGGenome

GGRNA,GGGenome

<http://ggrna.dbcls.jp/>

<http://gggenome.dbcls.jp/>

遺伝子をGoogleのように検索できるサイトです。 [NCBI RefSeq](#) の transcript を全文検索します。

検索例:

- 「[homeobox](#)」 「[claudin](#)」 フリーワード検索
- 「["RNA interference"](#)」 ダブルクオートで囲ってフレーズ検索
- 「[Argonaute "PAZ domain"](#)」 Argonaute かつ "PAZ domain" のAND検索
- 「[NM_001518](#)」 「[10579](#)」 RefSeq IDやGene IDなど各種IDから検索
- 「[symbol:VIM](#)」 遺伝子名(symbolまたはsynonym)から検索
- 「[ref:Naito](#)」 文献情報のなかからフリーワード検索
- 「[1552311_a_at](#)」 マイクロアレイのプローブIDから塩基配列を検索
- 「[aa:KDEL](#)」 アミノ酸配列を検索
- 「[caagaqaqagattg](#)」 塩基配列を検索
- 「[comp:caagaqaqagattg](#)」 相補鎖を検索
- 「[lub:aggcannrtgacct](#)」 N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- [詳細な使い方](#)
- ゲノム配列を検索したいときは **GGRNA** 《ゲゲゲノム》へ
 - ヒトゲノムの検索例: <http://GGGenome.dbcls.jp/TTCATTGACAACATT>

新着情報:

- 2016-03-22 データベースをRefSeq rel. 75 (Mar, 2016)に更新。
- 2015-01-13 HTTPSによる暗号化通信に対応 - <https://GGRNA.dbcls.jp/>
- 2013-07-24 ソースを公開 - [GitHub](#)
- 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを検索できます。
- 2012-05-29 下記論文の日本語による解説を「DBCLSからの成果発信」に掲載。
- 2012-05-29 GGRNAの論文が*Nucleic Acids Research*に掲載されました。
- [過去の新着情報](#)



INSDC

1. Next generation reads(SRA)
2. Capillary reads
3. Annotated sequences
4. Samples(BioSample)
5. Studies(BioProject)

RNAseq
ChIPseq

GEO ArrayExpress

microarray
(GeneChip,
Oligoarray)

NGS登場によってバラ
バラにアーカイブされ、
検索や再利用し辛い



AOE(All Of gene Expression)
Graph shortcut for expression data
 <http://AOE.dbcls.jp/>





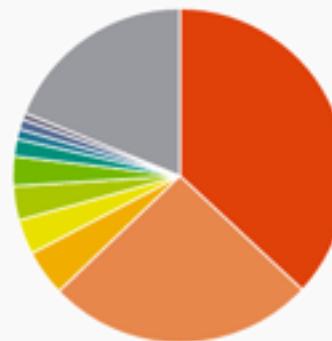
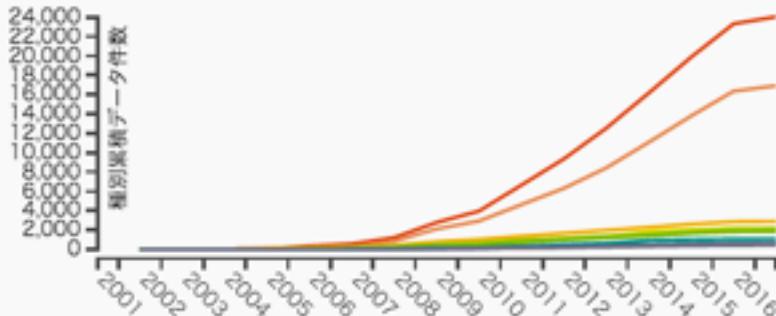
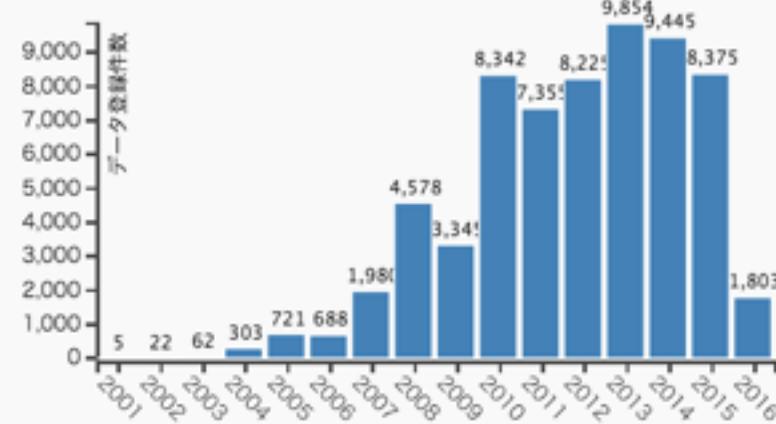
登録データランキング

データリスト

使い方

生物種別登録データランキング

選択した項目でリストを表示



1	Homo sapiens	24,040	
2	Mus musculus	16,902	
3	Arabidopsis thaliana	2,893	
4	Drosophila melanogaster	2,204	
5	Rattus norvegicus	2,145	
6	Saccharomyces cerevisiae	1,867	
7	Caenorhabditis elegans	1,078	
8	Escherichia coli	634	
9	Danio rerio	622	
10	Sus scrofa	432	
	others	12,286	

Togo picture gallery by DBCLS is licensed under a Creative Commons Attribution 2.1 Japan license (c)

手法別登録データランキング

選択した項目でリストを表示

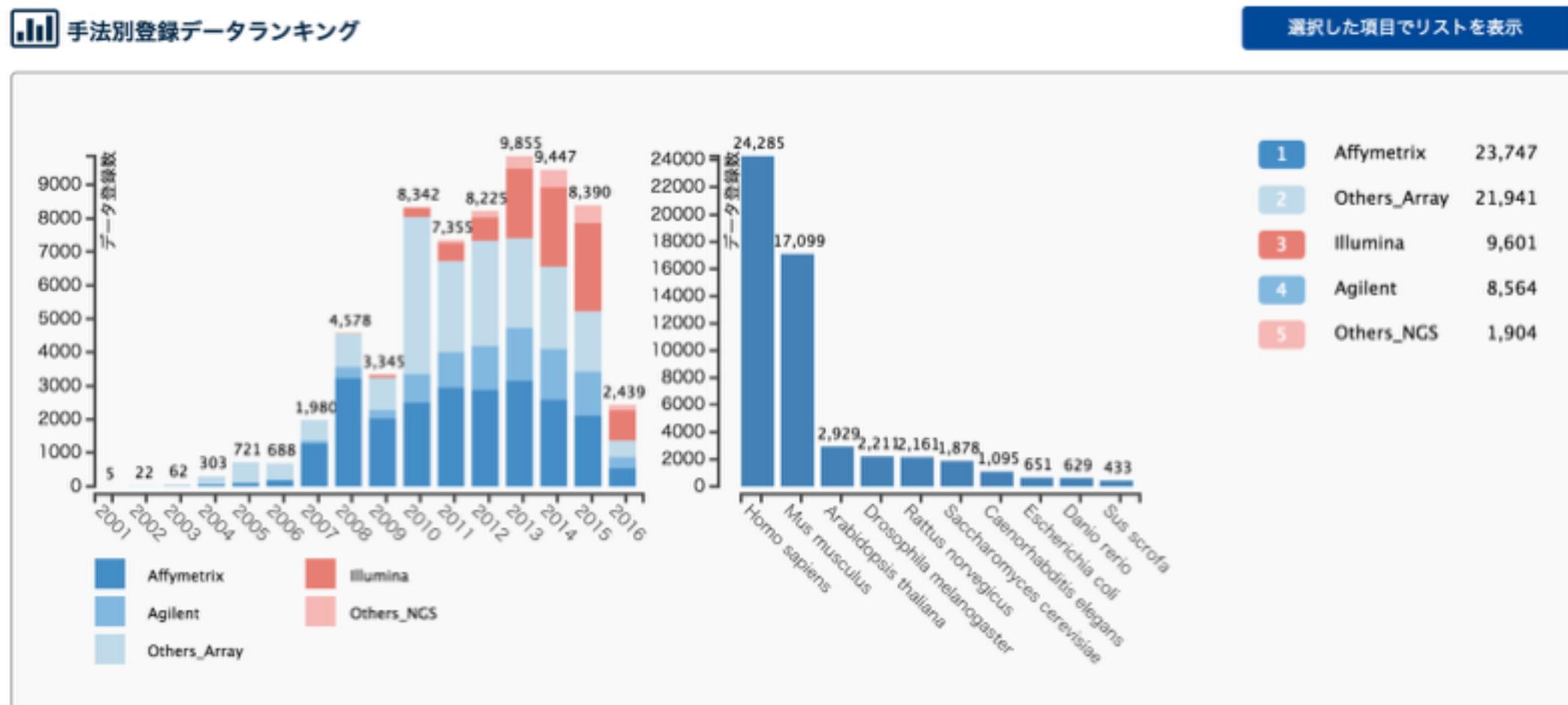


1	Affymetrix	23,615
2	Others Array	21,820

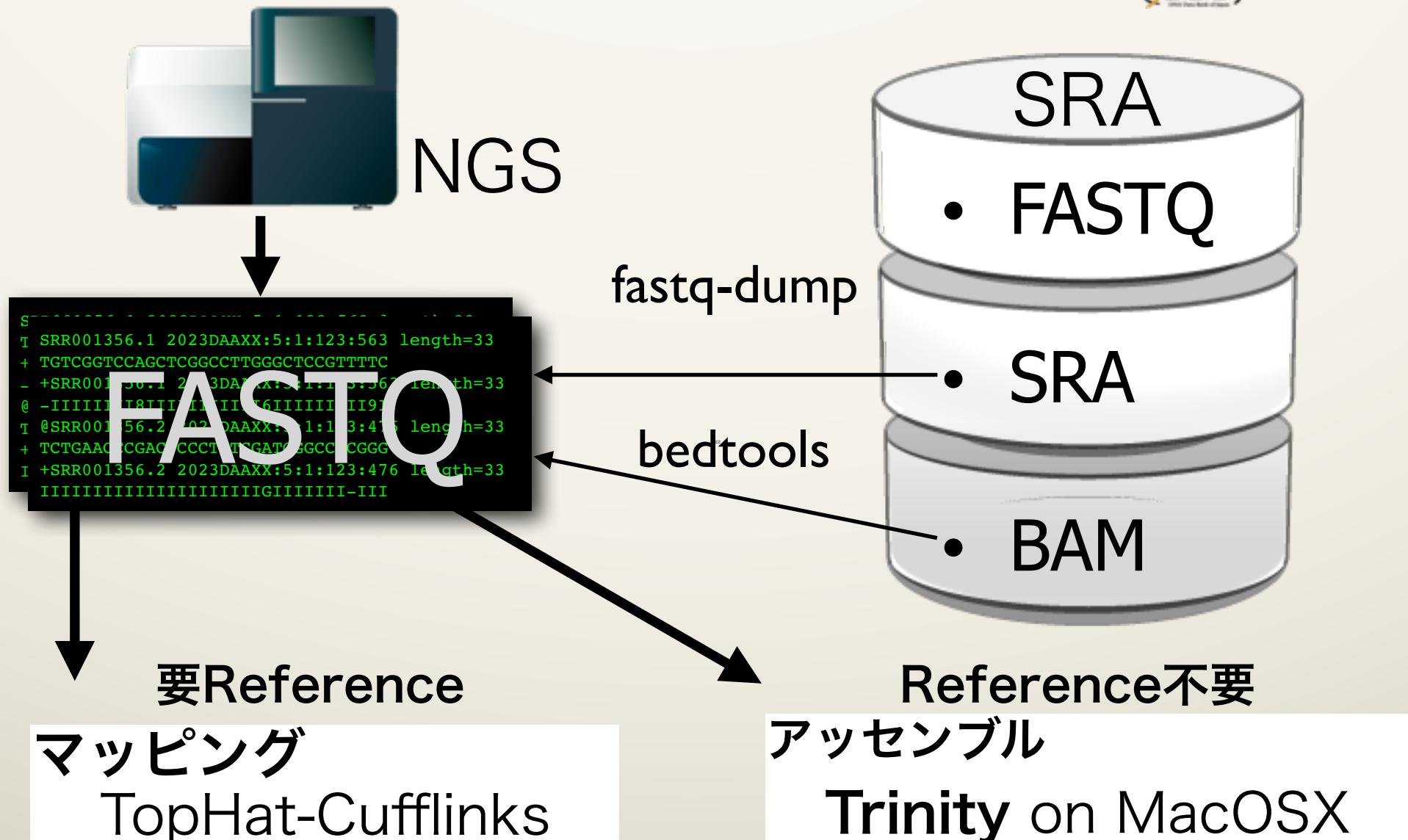
RNA-seqは本当に使われるようになった?

データレコード数の変化: AOEで見てみましょう

<http://aoe.dbcls.jp/>



RNAseq

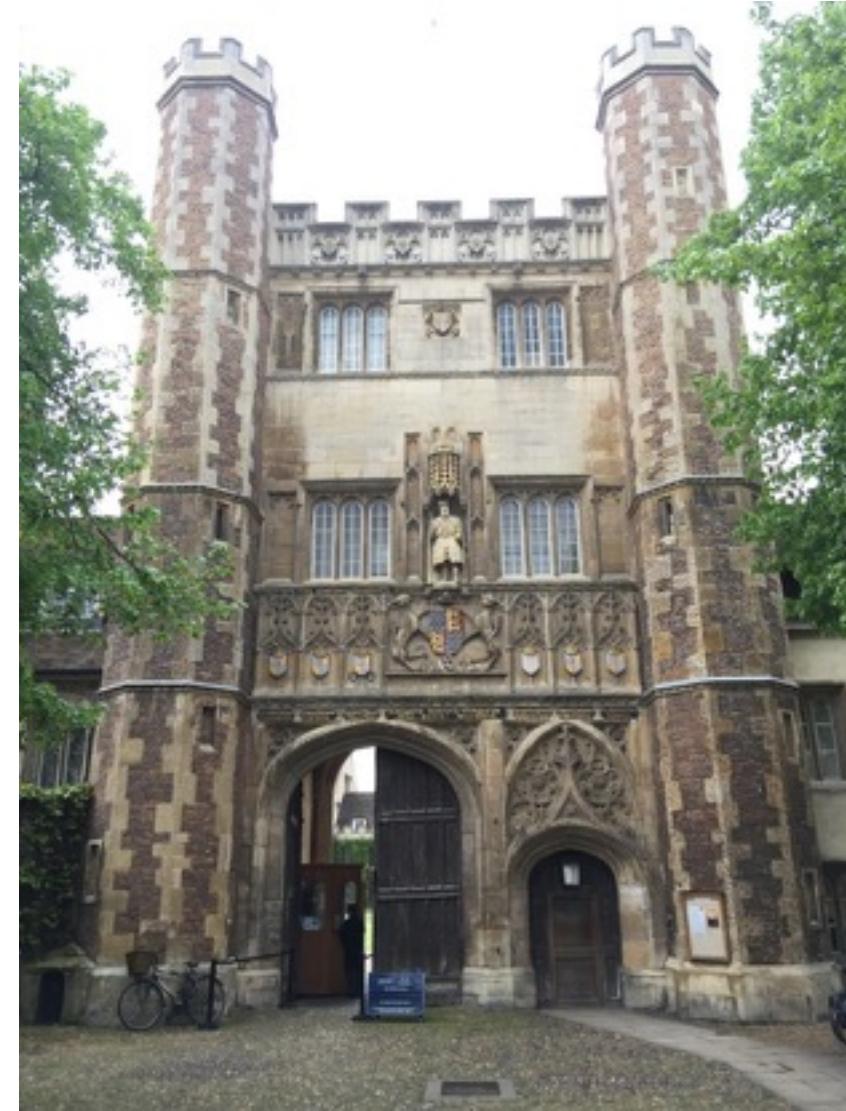


- transcriptome assemblyのツールとして有名
 - 転写単位ごとにクラスタリング
 - 入力:FASTQ、出力: FASTA
 - <http://trinityrnaseq.github.io/>
- 新規モデル生物で特に重用

ごく最近、 homebrewのTrinityが
バージョンアップ
→MacOSXで簡単によく動くに!

<http://bonohu.jp/blog/2016/05/09/trinity-on-macosx/>

<http://bonohu.jp/blog/2016/05/27/trinity-transcript-quantification/>





DNA Double Helix 1953

"The secret of life"

For decades the Eagle was the local
pub for scientists from the nearby
Cavendish Laboratory.

It was here on February 28th 1953 that
Francis Crick and James Watson first
announced their discovery of how
DNA carries genetic information.

Unveiled by James Watson
25th April 2003

EAGLE



THE EAGLE

GREENE KING

