RおよびBioconductorを使ったバイオデータ解析

坊農 秀雅(ぼうのう ひでまさ)

大学共同利用機関法人 情報・システム研究機構

ライフサイエンス統合データベースセンター

(DBCLS)

坊農秀雅







- 読み: ぼうのうひでまさ
- 京大化研→理研GSC→埼玉医大→DBCLS
- 専門:
  - –バイオインフォマティクス
  - –ゲノム生物学(微生物→マウス、最近は昆虫)
- ドメイン: bonohu.jp

twitter可(むしろ推奨)



bonohu bono@dbcls.jp

© 2014 DBCLS Licensed under CC BY 2.1JAPAN

学会年会でよろずデータ解析相談



主に内製でツールやDB開発・維持

塩基配列



DBCLS Galaxy

DBCLS SRA

CRISPRdirect

統合遺伝子検索
GGRNA

統合遺伝子検索GGRNA

超絶高速ゲノム配列検索
GGGenome

高速塩基配列検索GGGenome

Gendoo

遺伝子発現

RefEx

AOE (All Of gene Expression)

http://dbcls.rois.ac.jp/services

© 2014 DBCLS Licensed under CC BY 2.1JAPAN

Lizio *et al. Genome Biology* (2015) 16:22
DOI 10.1186/s13059-014-0560-6

Genome Biology

SOFTWARE Open Access

Gateways to the FANTOM5 promoter level mammalian expression atlas

Marina Lizio[1,2], Jayson Harshbarger[1,2], Hisashi Shimoji[1,2], Jessica Severin[1,2], Takeya Kasukawa[2], Serkan Sahin[1,2], Imad Abugessaisa[2], Shiro Fukuda[1], Fumi Hori[1,2], Sachi Ishikawa-Kato[1,2], Christopher J Mungall[5], Erik Arner[1,2], J Kenneth Baillie[7], Nicolas Bertin[1,2,19], Hidemasa Bono[10], Michiel de Hoon[1,2], Alexander D Diehl[13], Emmanuel Dimont[12], Tom C Freeman[7], Kaori Fujieda[10], Winston Hide[12,17], Rajaram Kaliyaperumal[8], Toshiaki Katayama[15], Timo Lassmann[1,2,18], Terrence F Meehan[6], Koro Nishikata[16], Hiromasa Ono[10], Michael Rehli[9], Albin Sandelin[11], Erik A Schultes[8,14], Peter AC 't Hoen[8], Zuotian Tatum[8], Mark Thompson[8], Tetsuro Toyoda[16], Derek W Wright[7], Carsten O Daub[1], Masayoshi Itoh[1,2,3], Piero Carninci[1,2], Yoshihide Hayashizaki[1,3], Alistair RR Forrest[1,2*], Hideya Kawaji[1,2,3,4*] and the FANTOM consortium

Big project (FANTOM5)の

データ再利用性向上にも加担

© 2014 DBCLS Licensed under CC BY 2.1JAPAN

作ったぞ、さあ

使え、では

使われないので

© 2014 DBCLS Licensed under CC BY 2.1JAPAN

OPEN ACCESS Freely available online

回転

PLOS ONE

Identification of Key Uric Acid Synthesis Pathway in a Unique Mutant Silkworm *Bombyx mori* Model of Parkinson's Disease

Hiroko Tabunoki[1]*¤, Hiromasa Ono[2], Hiroaki Ode[1], Kazuhiro Ishikawa[1], Natsuki Kawana[1], Yutaka Banno[3], Toru Shimada[4], Yuki Nakamura[5], Kimiko Yamamoto[5], Jun-ichi Satoh[1], Hidemasa Bono[2]*

1 Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Tokyo, Japan, **2** Database Center for Life Science (DBCLS), Research Organization of Information and Systems (ROIS), Tokyo, Japan, **3** The Center of Genetic Resources, Kyushu University, Fukuoka, Japan, **4** Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan, **5** Insect Genome Research Unit, National Institute of Agrobiological Sciences, Tsukuba, Japan

PLOS ONE

実際に使いたい研究者と共に

RESEARCH ARTICLE

Superoxide Dismutases, SOD1 and SOD2, Play a Distinct Role in the Fat Body during Pupation in Silkworm *Bombyx mori*

Yosui Nojima[1,2], Katsuhiko Ito[1], Hiromasa Ono[3], Takeru Nakazato[3], Hidemasa Bono[3], Takeshi Yokoyama[1], Ryoichi Sato[4], Yoshitaka Suetsugu[5], Yuki Nakamura[5], Kimiko Yamamoto[5], Jun-ichi Satoh[2], Hiroko Tabunoki[1]*, Hajime Fugo[1]

坊農秀雅 曰く

「これからは、生物学者自身がデータ解析する時代なんです」

データ解析ぐらいは
自分できるようにならないと…

© 2014 DBCLS Licensed under CC BY 2.1JAPAN

そこで

- R
  - –アルファベットでSの前の文字
    - ・S言語のフリー版として長らく開発されてきた
    - ・余談: フリーなSPSSとしてPSPP
  - – 「統計解析と可視化のための環境」
- BioConductor(Bioc)
  - –大規模ゲノムデータの解析と理解のためのライブラリー群
  - –頻繁に更新していくRの最新版に追随

統計解析環境R

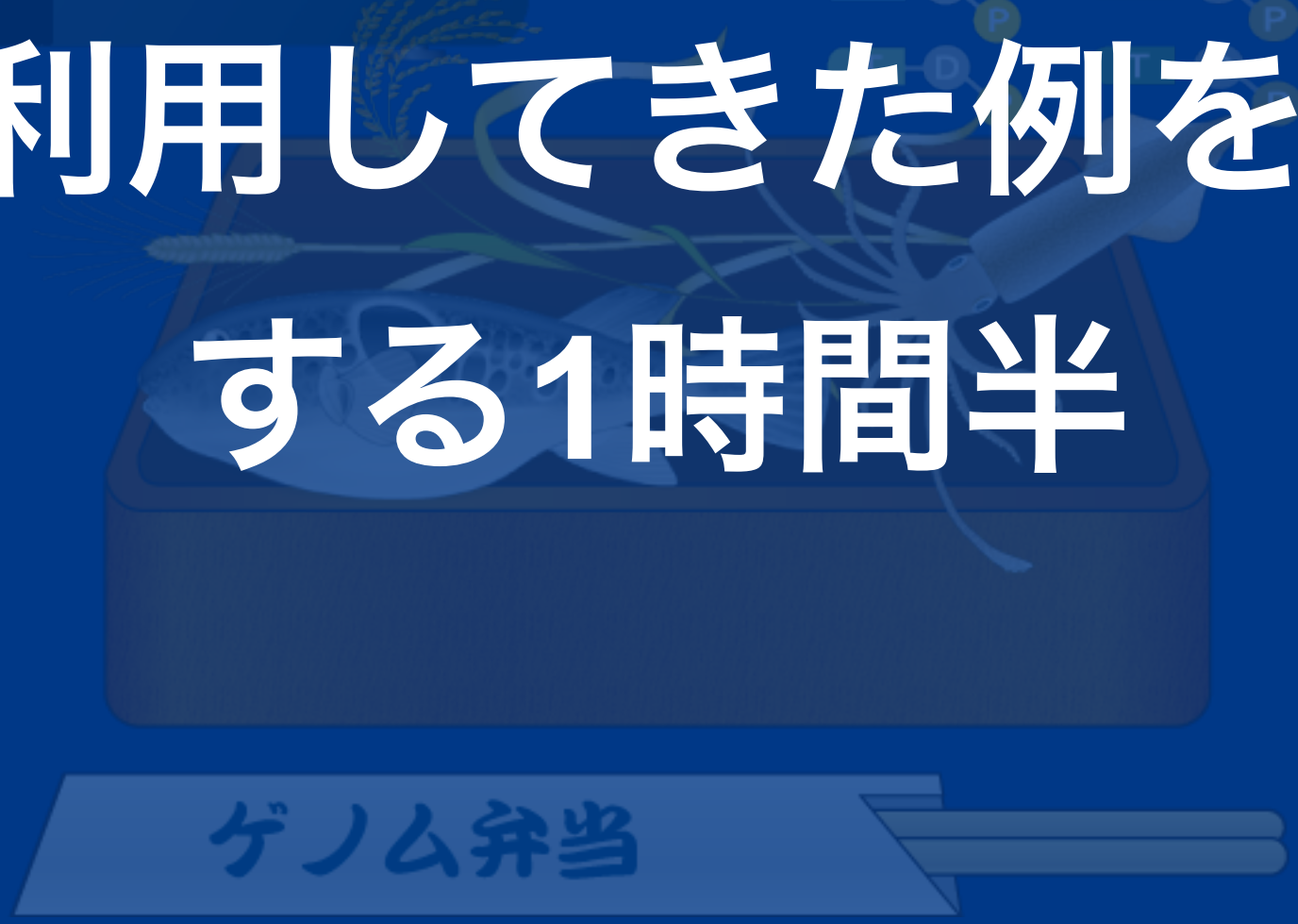
- Rを使ったトランスクリプトーム解析
  - –(Rで)マイクロアレイデータ解析
    - http://www.iu.a.u-tokyo.ac.jp/~kadota/r.html
  - –(Rで)塩基配列解析
    - http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html



- トランスクリプトーム解析 by 門田幸二 from 共立出版
  - –http://www.kyoritsu-pub.co.jp/bookdetail/9784320123700

DBCLS
Database Center for Life Science

一生命科学研究者の坊農が研究を遂行するにあたって Rを利用してきた例を紹介する1時間半



すべてはMotDBに

- MotDB: Master of the database
  - 「データベースの達人」という名の講習会資料アーカイブのこと
  - これまでの約50回分の講習会資料がここに
- 今の講習のページはこちら↓

http://motdb.dbcls.jp/?AJACS52