食品機能解析研究とバイオインフォマティクス

東京大学大学院農学生命科学研究科
アグリバイオインフォマティクス教育研究ユニット
門田 幸二（かどた　こうじ）
http://www.iu.a.u-tokyo.ac.jp/~kadota/
kadota@iu.a.u-tokyo.ac.jp

Google 門田　東大

講演資料はホームページから取得可能です

門田　幸二のホームページ

- **名前**
  門田　幸二(かどた　こうじ)
- **所属**
  東京大学 大学院農学生命科学研究科 アグリバイオインフォマティクス教育研究ユニット
- **身分**
  特任准教授
- **研究分野**
  バイオインフォマティク
- **所属学会**
  日本バイオインフォマテ
  日本分子生物学会
- **研究テーマ(last mod**
  トランスクリプトーム解
  サーなどによって得ら
  命科学への応用を目
  ーとしています。これま
  ム解析」でひとまとめに
  レイデータ解析」と「(R
  す。

- **講演など(上記講義以外)(last modified: 2013.03.22)** NEW
  23. 題目:「食品機能解析研究とバイオインフォマティクス」, 日本農芸化学会2013年度大会・シンポジウム4SY08, 東北大学(宮城), 2013.03.27
  22. 題目:「Rでトランスクリプトーム解析」, HPCI チュートリアルセミナー, 生命情報工学研究センター(東京), 2013.03.07
  21. 題目:「Rでトランスクリプトーム解析」, HPCI チュートリアルセミナー, 生命情報工学研究センター(東京), 2012.03.09
  20. 題目:「Rによるトランスクリプトーム解析～NGS由来塩基配列データを自在に解析する～」, Rでつなぐ次世代オミックス情報統合解析研究会, 理化学研究所横浜研究所(神奈川), 2012.02.22
  19. 題目:「RNA-Seqデータ解析リテラシー」, Illumina Webinar Series・RNAシーケンスを始めよう・セッション3:データ解析, イルミナ株式会社(東京), 2011.11.17
  18. 題目:「農業生物のトランスクリプトーム解析における情報処理」, 東京大学大学院農学生命科学研究科第91回アグリバイオインフォマティクスセミナー, 東京大学(東京), 2011.11.11
  17. 題目:「RNA-Seqデータ解析における正規化法の選択:RPKM値でサンプル間比較は危険?!」, イルミナ株式会社 バイオインフォマティクス講習会【中級】, 富士ソフトウェア　アキバプラザ(東京), 2011.9.29

自己紹介（バイオインフォマティクスな人）

学歴

- 1995年3月
  - 高知工業高等専門学校・工業化学科　卒業
- 1997年3月
  - 東京農工大学・工学部・物質生物工学科　卒業
- 1999年3月
  - 東京農工大学・大学院工学研究科・物質生物工学専攻　修士課程修了
- 2002年3月
  - 東京大学・大学院農学生命科学研究科・応用生命工学専攻　博士課程修了
  - 学位論文：「cDNAマイクロアレイを用いた遺伝子発現解析手法の開発」
    （指導教官：清水謙多郎教授）

博士課程の頃一度だけ農芸化学会年会に出たことがありますが…

職歴

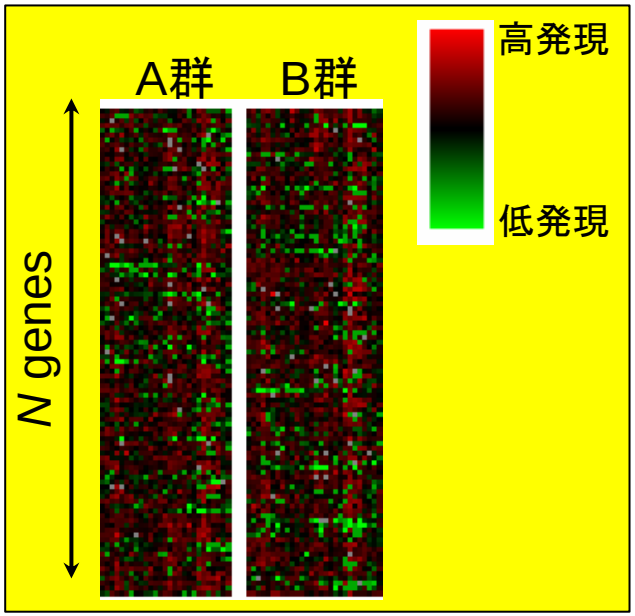
- 2002/4/1~
  - 産総研・生命情報科学研究センター（CBRC）
- 2003/11/1~
  - 放医研・先端遺伝子発現研究センター
- 2005/2/16~
  - 東京大学・大学院農学生命科学研究科

私の知っている食品機能解析研究…

- 遺伝子発現（トランスクリプトーム）データの解析
  - 「機能性食品（A群） vs. 非機能性食品（B群）」の二群間比較

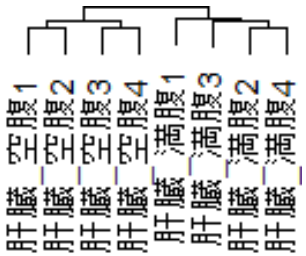
マイクロアレイ

RNA-seq

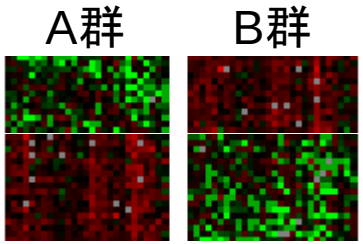


遺伝子発現データ

クラスタリング



発現変動遺伝子（DEG）検出



機能解析

・Gene Ontology（GO）解析
・パスウェイ解析
…

…

http://www.illuminakk.co.jp/systems/hiseq_systems/hiseq_2500_1500.ilmn

Contents

マイクロアレイ

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - 機能解析（GO解析など）結果に影響
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」

  …

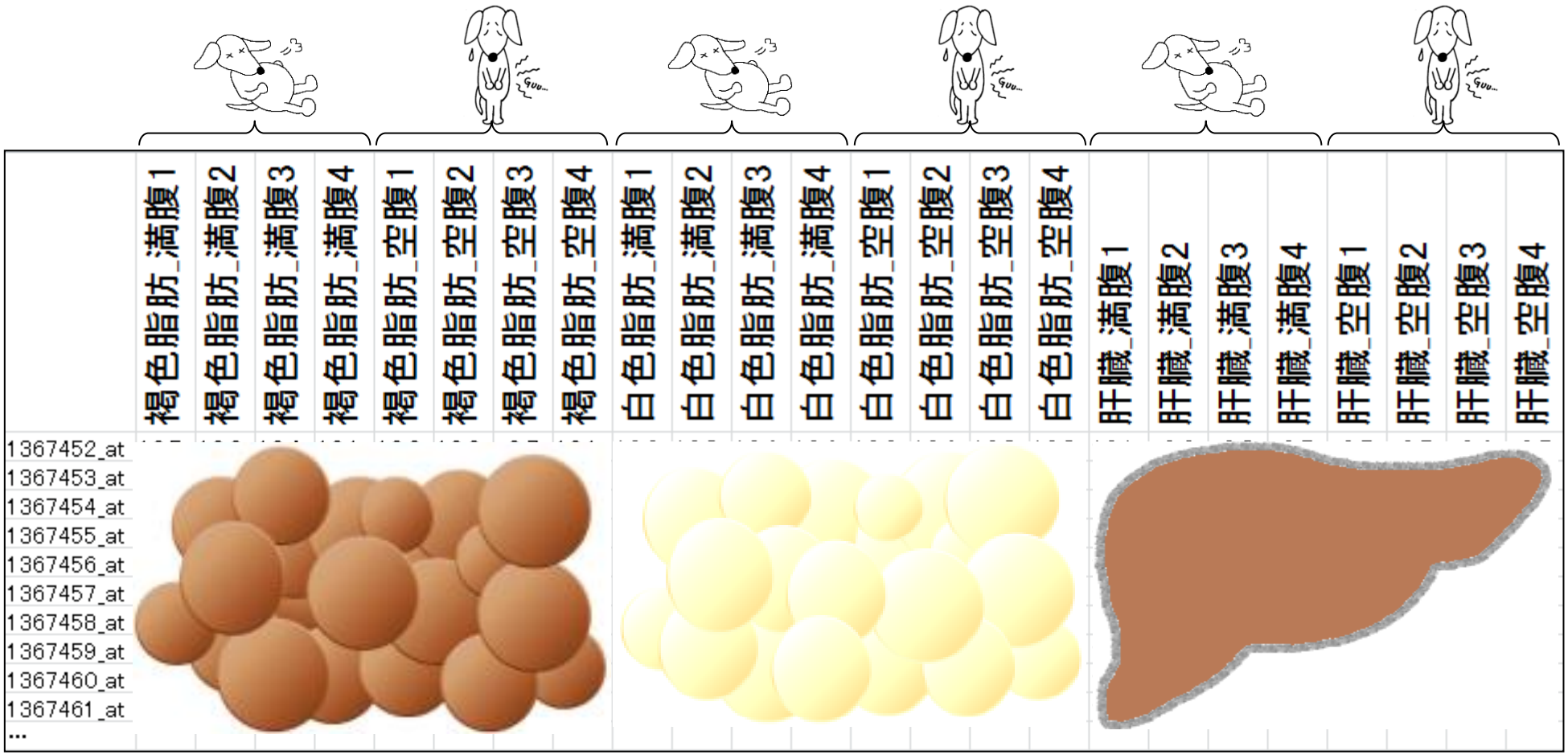
Contents

マイクロアレイ

- **（サンプル間）クラスタリング**
  - データの大まかな特徴を把握可能
  - **発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見**
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - 機能解析（GO解析など）結果に影響
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」

  …

サンプル間クラスタリング



31,099 行（probesets） × 24 列（samples）のマイクロアレイデータ

クラスタリング結果

クラスタリング結果を眺めれば発現変動遺伝子（DEG）に関するおおよその見当がつきます
→ クラスタリングって重要

4通りの二群間比較(A群 vs. B群)を行う

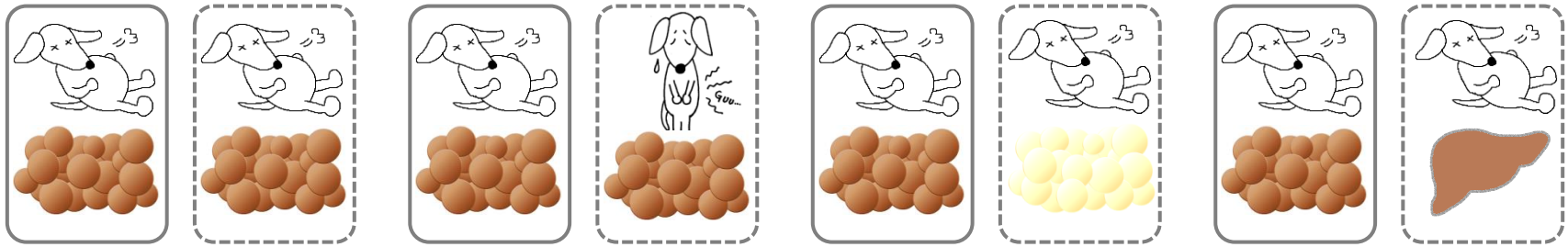
	褐色脂肪_満腹1	褐色脂肪_満腹2	褐色脂肪_満腹3	褐色脂肪_満腹4	褐色脂肪_空腹1	褐色脂肪_空腹2	褐色脂肪_空腹3	褐色脂肪_空腹4	白色脂肪_満腹1	白色脂肪_満腹2	白色脂肪_満腹3	白色脂肪_満腹4	白色脂肪_空腹1	白色脂肪_空腹2	白色脂肪_空腹3	白色脂肪_空腹4	肝臓_満腹1	肝臓_満腹2	肝臓_満腹3	肝臓_満腹4	肝臓_空腹1	肝臓_空腹2	肝臓_空腹3	肝臓_空腹4
解析1	A1	A2	B1	B2																				
解析2	A1	A2			B1	B2																		
解析3	A1	A2							B1	B2														
解析4	A1	A2															B1	B2						

解析1の予想:DEG**なし**

解析4の予想:DEG**あり**(多め)

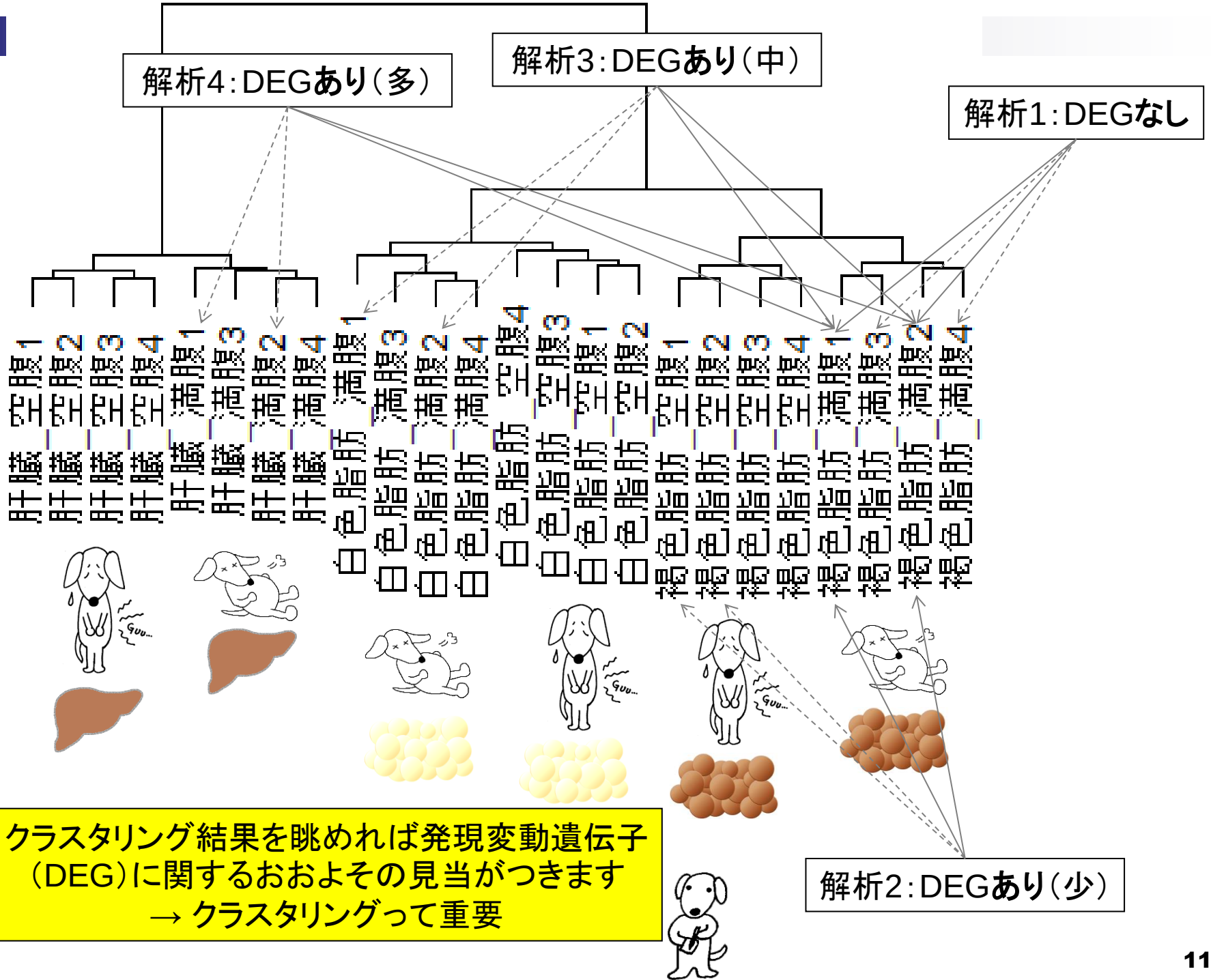
二群間比較結果は予想通り

遺伝子数	解析1 A群：褐色脂肪_満腹 B群：褐色脂肪_満腹	解析2 A群：褐色脂肪_満腹 B群：褐色脂肪_空腹	解析3 A群：褐色脂肪_満腹 B群：白色脂肪_満腹	解析4 A群：褐色脂肪_満腹 B群：肝臓_満腹
FDR < 0.05	0	0	0	1061
FDR < 0.10	0	4	0	4287
FDR < 0.15	0	135	1	6963
FDR < 0.20	0	689	505	9357
FDR < 0.25	0	2071	3519	11397
FDR < 0.30	0	3546	6780	13033



DEGなし

DEG数少なめ ← DEG**あり** → DEG数多め


解析4：DEGあり（多）
解析3：DEGあり（中）
解析1：DEGなし
肝臓1
肝臓2
肝臓3
肝臓4
褐色脂肪1
褐色脂肪3
褐色脂肪2
褐色脂肪4
白色脂肪4
白色脂肪3
白色脂肪1
白色脂肪2
褐色脂肪1
褐色脂肪2
褐色脂肪3
褐色脂肪4
褐色脂肪1
褐色脂肪3
褐色脂肪2
褐色脂肪4
クラスタリング結果を眺めれば発現変動遺伝子（DEG）に関するおおよその見当がつきます
→ クラスタリングって重要
解析2：DEGあり（少）

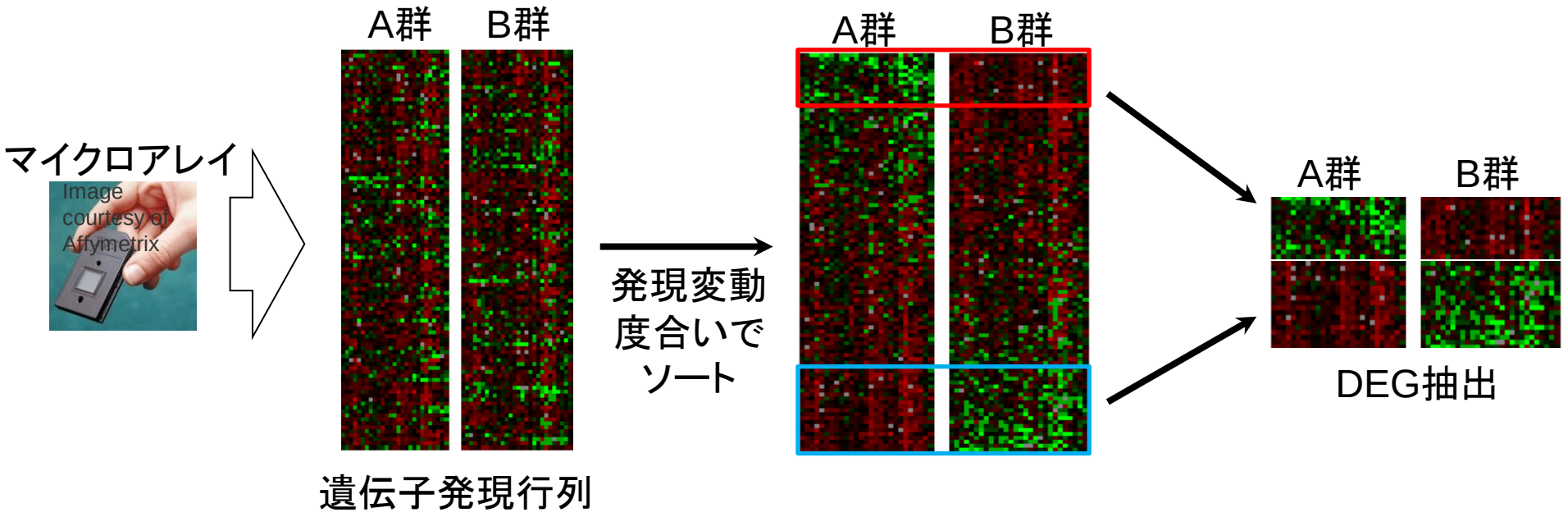
Contents

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- **発現変動遺伝子（DEG）検出**
  - 機能解析（GO解析など）結果に影響
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」

  …

発現変動遺伝子（DEG）検出

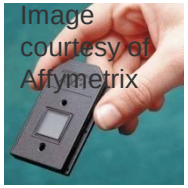
- 例：「機能性食品（A群） vs. 非機能性食品（B群）」



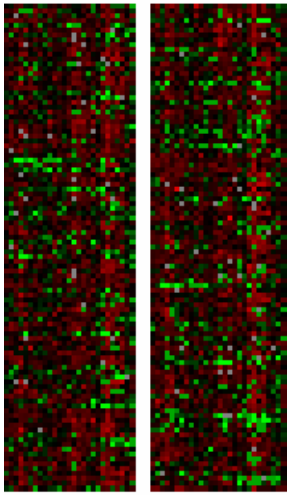
用いるDEG検出法によって得られる結果がかなり違う…

用いるDEG検出法によって結果が異なる…

マイクロアレイ

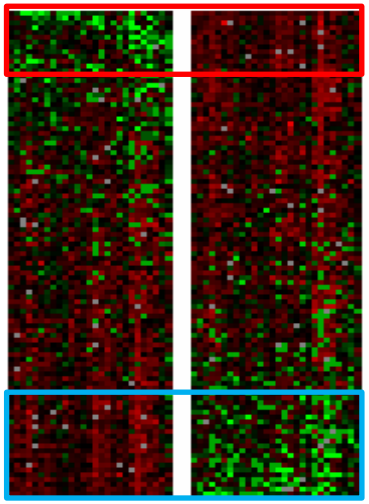


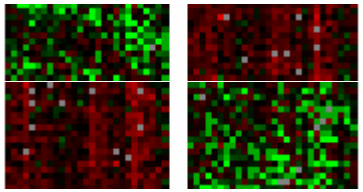
A群　B群



遺伝子発現行列

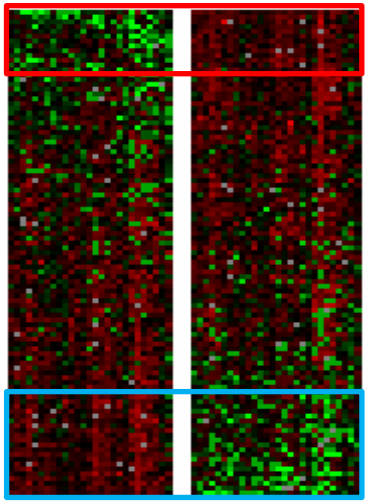
log(B/A)でソート

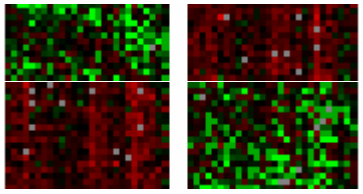




DEGセット

*t*統計量でソート





DEGセット

系統の異なる検出法から得られるDEGセット間の一致度は低い
→ 機能解析結果が異なる

用いるDEG検出法によって結果が異なる…

- DEGセットに占める共通遺伝子の割合(%)

「倍率変化(FC)系の方法」間の一致度は高い

「*t*検定系の方法」間の一致度は高い

	AD	WAD	FC	RP	modT	samT	shrinkT
WAD	62						
FC	79	50					
RP	85	60	72				
modT	35	32	32	37			
samT	36	33	33	38	88		
shrinkT	34	31	31	35	93	88	
ibmT	38	33	35	39	88	84	85

「**系統の異なる**検出法」間の一致度は低い

先行研究で用いられたDEG検出法とは異なる系統のDEG検出法を採用すると、異なる結果が導かれうる

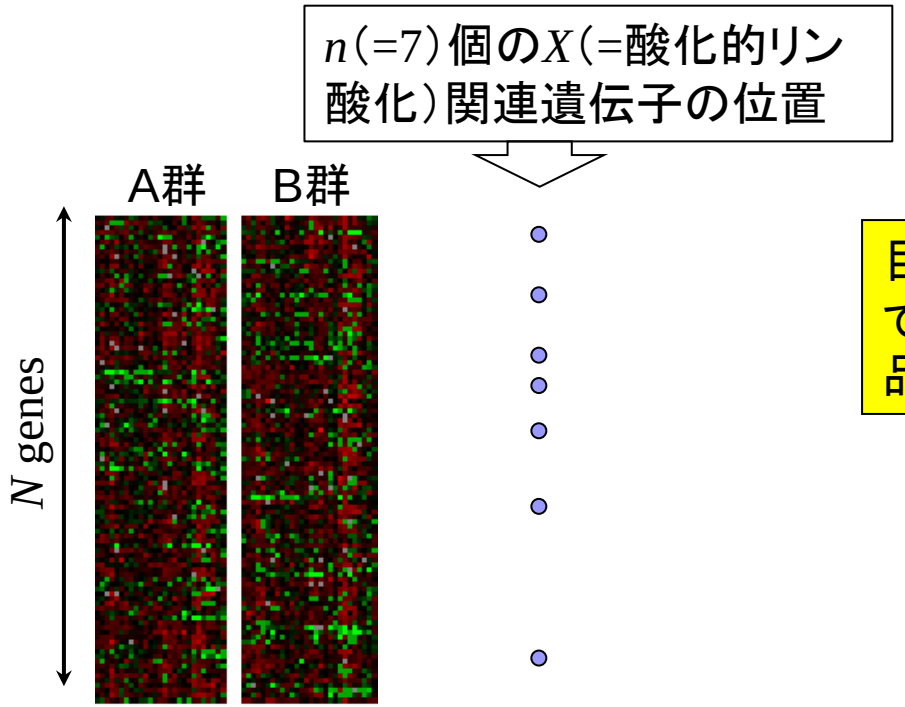
Contents

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - **機能解析（GO解析など）結果に影響**
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」

  …

（食品）機能解析

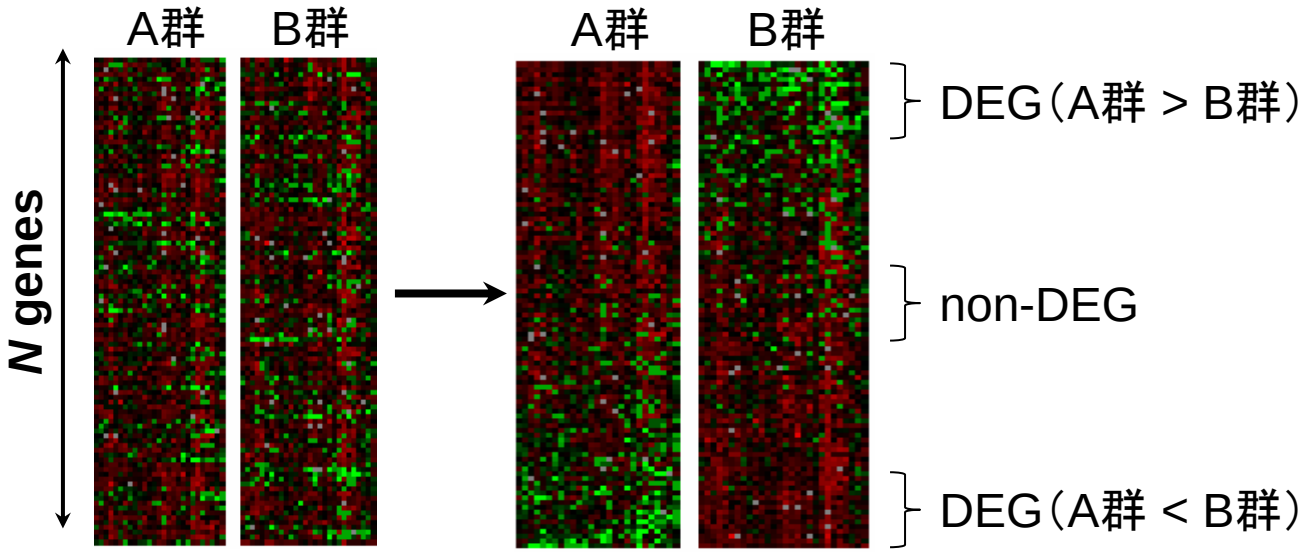
- 機能解析（GO解析など）≒ 発現変動遺伝子**セット**解析
  - 「GO Biological Process由来遺伝子セット群」の解析 → 発現変動遺伝子**セット**を同定
  - 「KEGG Pathway由来遺伝子セット群」の解析 → 発現変動遺伝子**セット**を同定
  - 「ある機能*X*に関連した遺伝子**セット**」が比較するサンプル間で変動したかどうかを評価



目的：「*X*関連遺伝子**セット**」が変動しているかどうかを調べたい（機能性食品が作用しているところを知りたい）

機能解析（遺伝子セット解析）

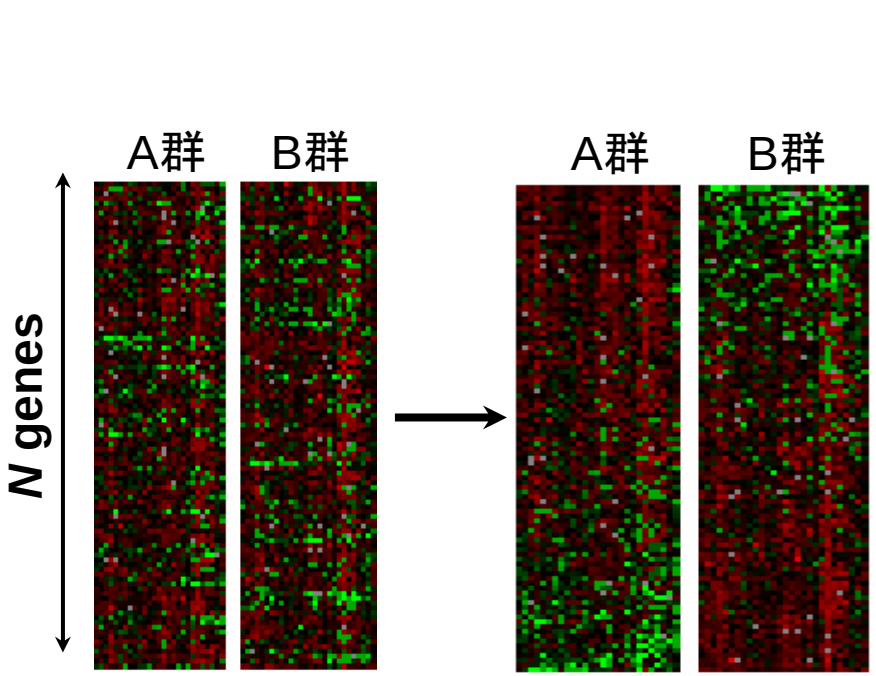
- 遺伝子ごとの統計量を算出（発現変動の度合いを数値化）
  - *t*-統計量、log（B/A）、相関係数、SAM、WAD、…

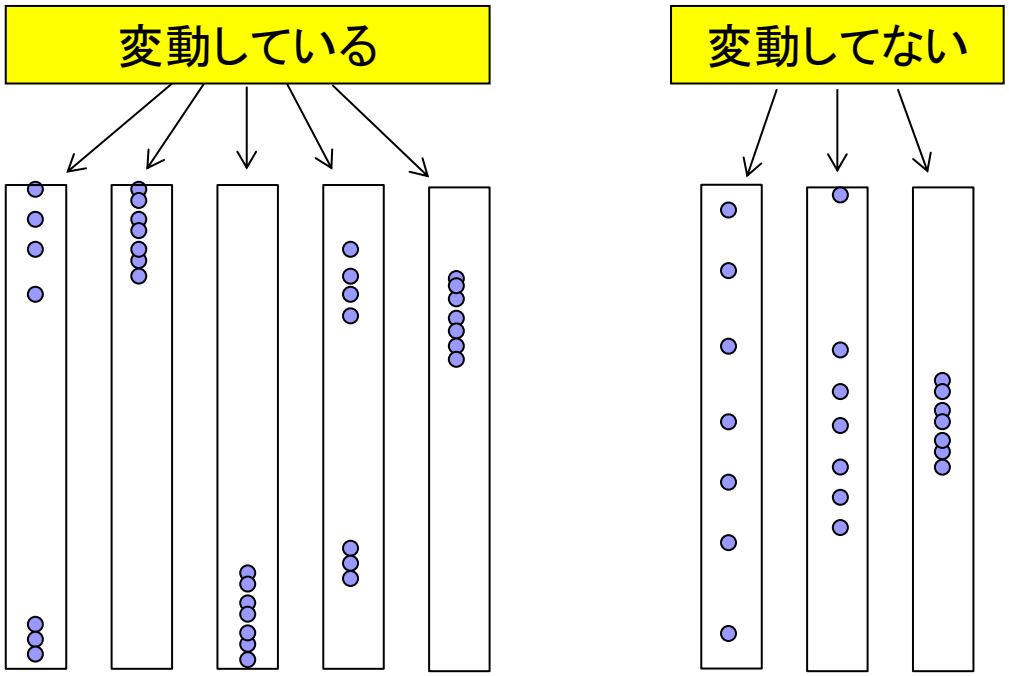


DEG検出（ランキング）手順を内部的に行っている

機能解析（遺伝子セット解析）

- 発現変動順にソート後の「ある機能Xに関連した遺伝子**セット**」のステレオタイプな分布

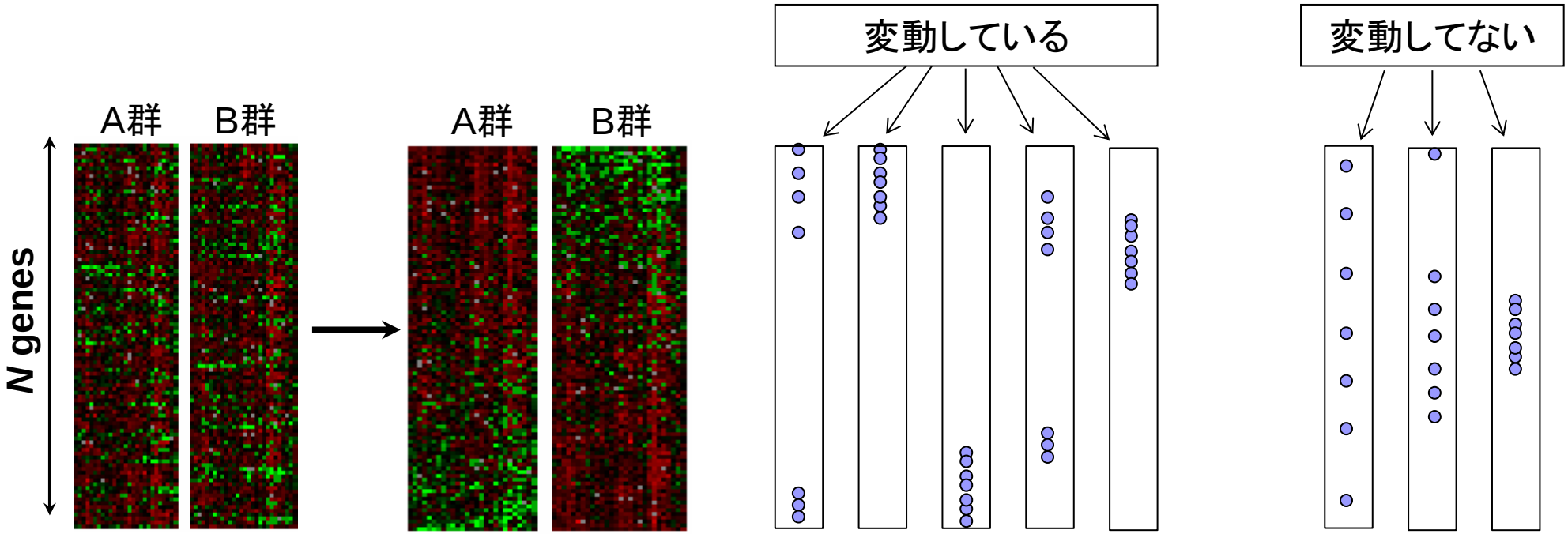




全体的な「偏り」の度合を評価しています

機能解析結果はDEG検出法次第…

- 発現変動順にソート後の「ある機能Xに関連した遺伝子**セット**」のステレオタイプな分布



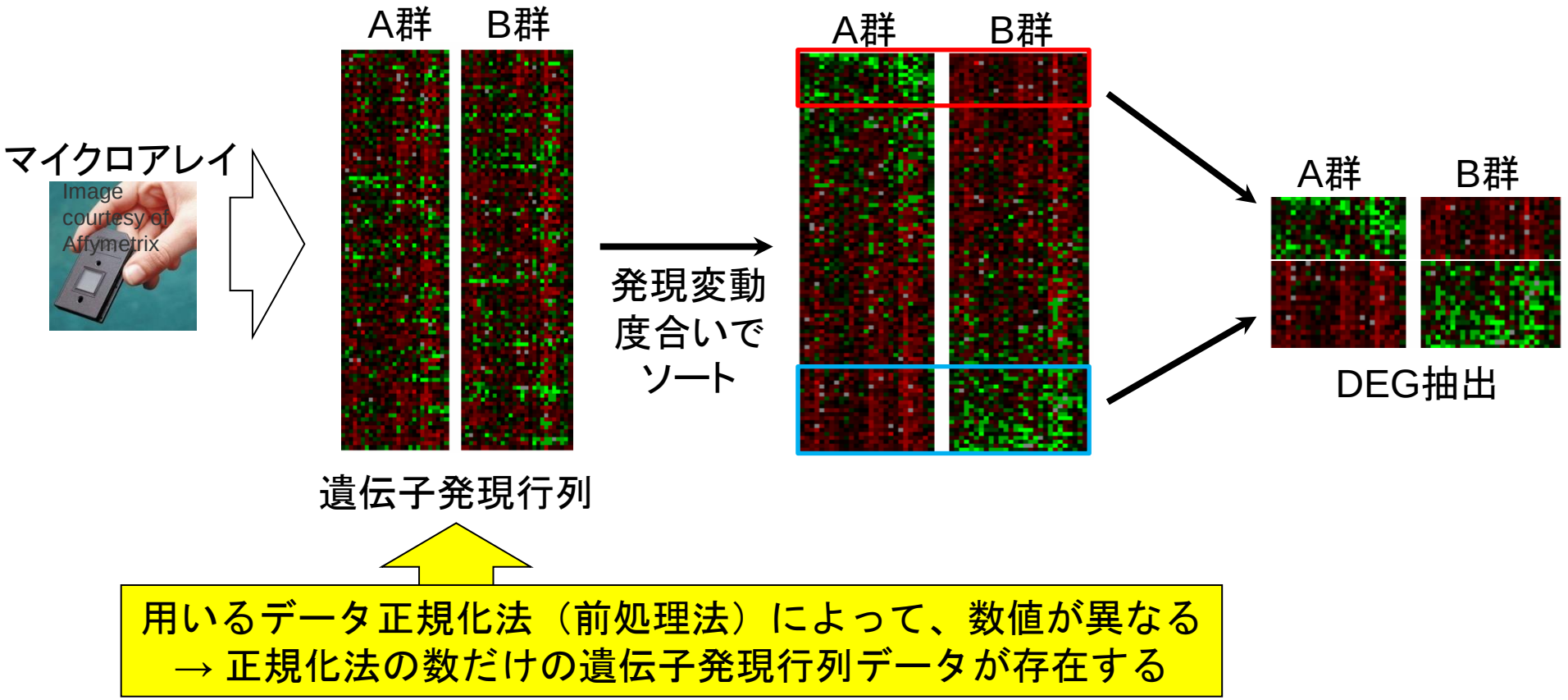
用いるDEG検出（ランキング）法が異なれば「発現変動遺伝子セットの上位リスト」が当然違ってくる

Contents

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - 機能解析（GO解析など）結果に影響
  - **一般的な解析手順：①データ正規化 → ②DEG検出**
    - **現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要**
  - **DEG同定法は基本的に「倍率変化に基づく方法」がお勧め**
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」
  - …

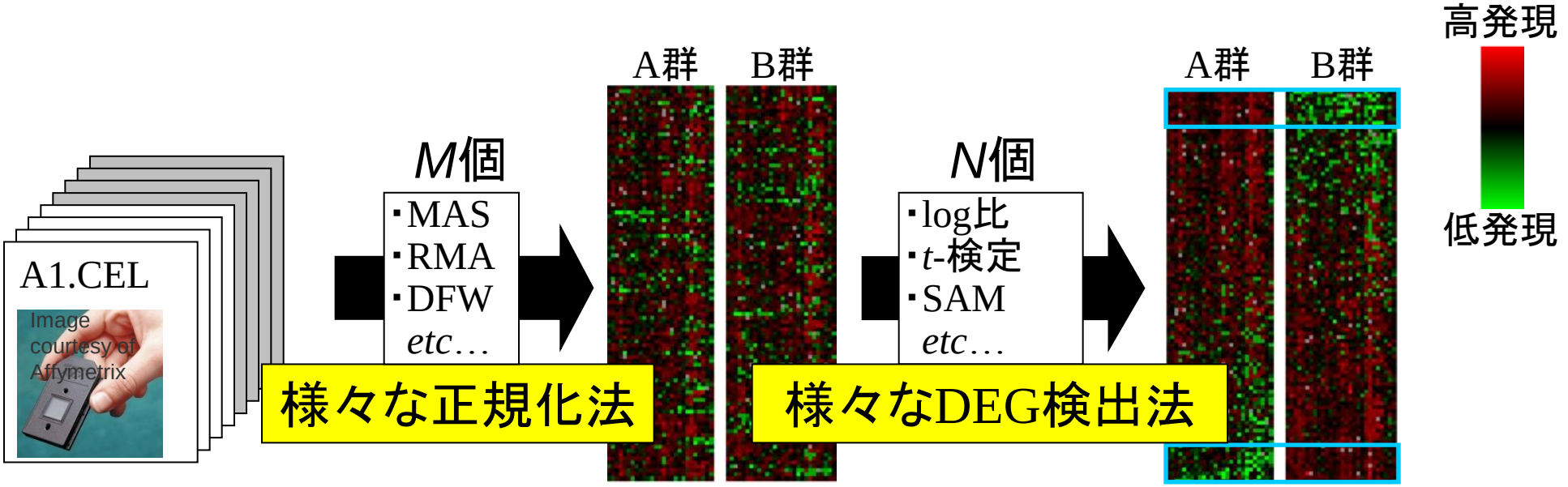
一般的な解析手順（DEG検出）

- **①データ正規化 → ②DEG検出**



一般的な解析手順（DEG検出）

- **①データ正規化 → ②DEG検出**



手法選択ガイドライン

- 評価基準が**感度・特異度**の場合
  - MAS5（正規化法） → WAD（DEG検出法）
  - RMA（正規化法） → Rank products（DEG検出法）
  - …
- 評価基準が**再現性**の場合
  - …（正規化法） → WAD（DEG検出法）

いずれもDEG検出法は
「倍率変化に基づく方法」

- 考察
  - 従来、感度・特異度の高い方法は複製実験データを利用してバラツキを見積もる**「統計的検出法」**だと言われてきた（例：t検定、SAMなど）…。
  - RMAやDFWなどの比較的最近提案された正規化法は、全サンプルデータを読み込んでバラツキをコントロールした結果（遺伝子発現行列）を出力する。
    → DEG検出時にデータのバラツキを見積もる必要のない状態

DEG検出時にバラツキを考慮しない**「倍率変化に基づく方法」**で充分（むしろ適している）

イメージ

- 昔のガイドライン：統計的検出法（t-検定など）

Gene ID	A1	A2	A3	A4	…	B1	B2	B3	B4	…
Gene1										
Gene2										
Gene3										
Gene4										
Gene5										
Gene6										
Gene7										
…										

バラツキをコントロールしていない正規化法（log変換後のデータ）

$$t\text{統計量} = \frac{B\text{の平均} - A\text{の平均}}{A\text{のバラツキ} + B\text{のバラツキ}}$$

- 今のガイドライン：**倍率変化に基づく方法**（log ratioなど）

Gene ID	A1	A2	A3	A4	…	B1	B2	B3	B4	…
Gene1										
Gene2										
Gene3										
Gene4										
Gene5										
Gene6										
Gene7										
…										

バラツキをコントロールした正規化法

$$\log \text{ratio} = B\text{の平均} - A\text{の平均}$$

バラツキのコントロールは正規化時に実行済み

Contents

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - 機能解析（GO解析など）結果に影響
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」
  - …

NGS（RNA-seq）データ解析の場合は？

- ある特定のサンプル内での**遺伝子間**の発現量の大小関係を知りたい場合
  - 「配列長」由来bias：長いほど沢山sequenceされる
  - 「GC含量」由来bias：カウント数の分布がGC含量依存的である
- **サンプル間**比較（sample A vs. Bなど）で、発現変動遺伝子（DEG）を調べたい場合
  - 「sequence depthの違い」：総リード数がx倍違うと全体的にx倍変動…
  - 「組成の違い」：サンプル特異的高発現遺伝子の存在で比較困難に…
  - RPM（CPM）正規化 → TMM正規化 → TbT正規化 → iDEGES正規化

RPM（CPM）正規化	TMM正規化	TbT正規化	iDEGES正規化
総リード数を揃えるだけ	DEGを（正確には見積もらないので）多めにトリム	正規化の手順の中で同定したDEGをトリムすることでより頑健に	律速であったDEG同定部分の改良により、より頑健且つ高速に

配列長を考慮した発現量推定のイメージ

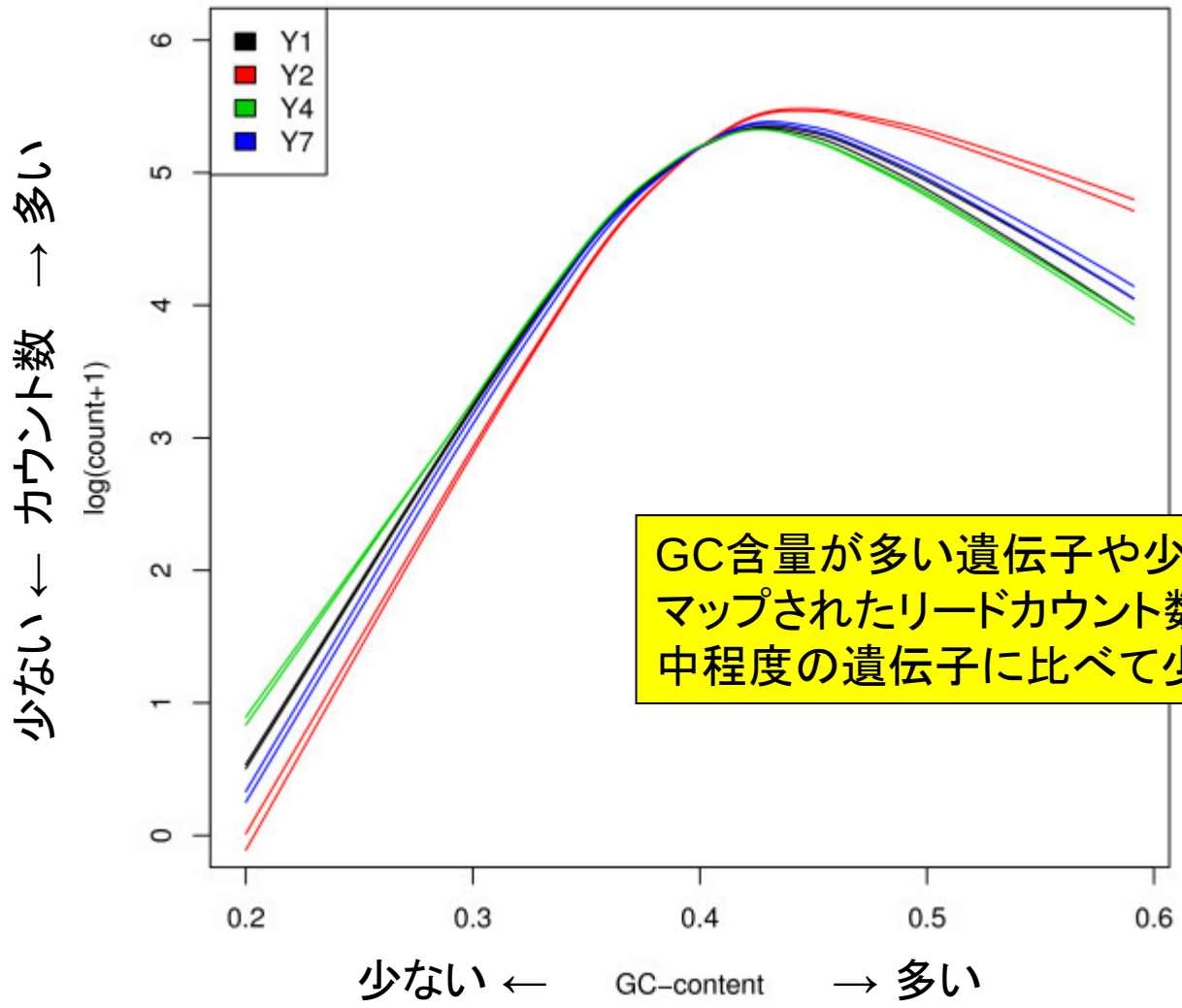
- gene1: 3 exons (middle length), 14 reads mapped (**low** coverage)
- gene2: 3 exons (middle length), 56 reads mapped (**high** coverage)
- gene3: 2 exons (**short** length), 12 reads mapped (middle coverage)
- gene4: 2 exons (**long** length), 31 reads mapped (middle coverage)

マップされたリード分布　　生リードカウント結果　　補正度の発現量

・長さが同じならリード数の多い方が発現量高い（gene 1 vs. 2）
・長いほどマップされるリード数が多くなる効果を補正する必要がある（gene 3 vs. 4）

一つのサンプル内で転写物（遺伝子）間の発現レベルの大小を比較したい場合には配列長を考慮すべきである

GC biasの実例



GC含量が多い遺伝子や少ない遺伝子上にマップされたリードカウント数は、GC含量が中程度の遺伝子に比べて少ない傾向にある

もう少し詳細を知りたい方は…

講義を受講して下さい。東大生以外の学生や社会人（ポスドク含む）も例年2割程度は受講しています。4/5(金),17:15-ガイダンス@東大農

まとめ

- （サンプル間）クラスタリング
  - データの大まかな特徴を把握可能
  - 発現変動遺伝子（Differentially Expressed Genes; DEGs）に関する知見
    - 「DEGの有無」や「DEG数の大小」
    - DEG検出法（対応あり or なし）の選択
- 発現変動遺伝子（DEG）検出
  - 機能解析（GO解析など）結果に影響
  - 一般的な解析手順：①データ正規化 → ②DEG検出
    - 現実には多数の方法が存在…。「正規化法」と「DEG検出法」の組み合わせが重要
  - DEG同定法は基本的に「倍率変化に基づく方法」がお勧め
- NGS（RNA-seq）データ解析の場合は？
  - 大枠としては同じだが…DEG検出法は「統計的な方法」がお勧め
  - 「広いダイナミックレンジ」→「正規化がより困難」
  - 「より詳細な転写物レベルの情報」→「機能解析可能なアノテーション情報が乏しい」

  …

謝辞

共同研究者

清水 謙多郎　先生（東京大学・大学院農学生命科学研究科）

西山 智明　先生（金沢大学・学際科学実験センター）

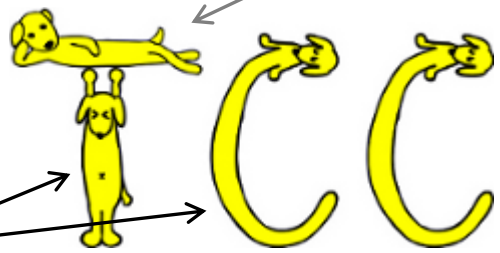
孫 建強　氏（東京大学・大学院農学生命科学研究科・修士課程1年）

グラント

- 基盤研究(C)（H24-26年度）:「シークエンスに基づく比較トランスクリプトーム解析のためのガイドライン構築」(代表)
- 新学術領域研究(研究領域提案型)(H22年度-):「非モデル生物におけるゲノム解析法の確立」(分担;研究代表者:西山智明)

（妻の）門田 雅世さま作

挿絵やTCCのロゴなど



（有能な秘書の）三浦 文さま作

http://www.iu.a.u-tokyo.ac.jp/~kadota/TCC/

RNA-seqデータ解析用Rパッケージ

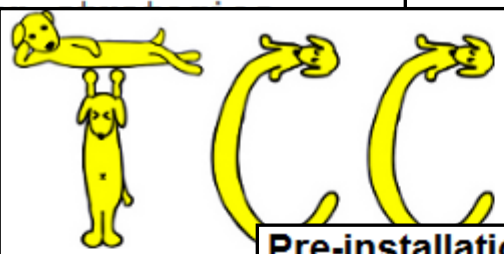
TCC: TCC: Differential expression analysis for tag count data with robust normalizatio

This package provides functions for performing differential expr
expressed gene elimination strategy. A simple unified interface i
functions to calculate normalization factors and estimate differen
edgeR, baySeq, and DESeq packages. The appropriate combi
more robust and accurate estimation performed easily than dire
Functions to produce simulation data under various conditions a
provided.

Version:	1.0.0
Depends:	R (≥ 2.15), edgeR, baySeq, DESeq, ROC, m
Published:	2013-01-10
Author:	Sun Jianqiang, Tomoaki Nishiyama, Kentaro S
Maintainer:	Tomoaki Nishiyama <tomoakin at staff.kanaza
License:	GPL-2
Copyright:	Authors listed above
URL:	http://www.iu.a.u-tokyo.ac.jp/~kadota/TCC
Citation:	TCC citation info
CRAN checks:	TCC results

Downloads:

Package source:	TCC_1.0.0.tar.gz
MacOS X binary:	not available, see check log.
Windows binary:	TCC_1.0.0.zip
Reference manual:	TCC.pdf
News/ChangeLog:	NEWS



TCC: an R p
with robust

The R package, T
to perform differen
expression analys
throughput sequer
gene expression.
method (TbT; Kad
replicates. The str
expressed genes
demonstrated that
normalization is es
DEGs are top-ran
integrated analysi
compared with oth

Pre-installation

The main functionalities in **TCC** consist of combinations of functions from the above three R/Bioconductor packages (edgeR, DESeq, and baySeq). If those packages have not been installed in your R environment, you need to enter the following commands after starting R:

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("edgeR", "baySeq", "DESeq", "ROC"))
```

Installation

This package is available at the CRAN repository. To install the package, visit the repository for TCC or enter the following command after starting R:

```
install.packages("TCC", type = "source")
```

Bioconductor likeなUser's Guide (Vignette)もあります

Documentation

User's Guide　　R script
Reference Manual

理想的な実験デザイン（二群間比較）

- サンプルA vs. Bの比較（Kidney vs. Liver；腎臓 vs. 肝臓）

Gene ID	A1	A2	A3	A4	…	B1	B2	B3	B4	…
Gene1										
Gene2										
Gene3										
Gene4										
Gene5										
Gene6										
Gene7										
…										

A1：ある生物の**腎臓**
A2：同じ生物種の別個体の**腎臓**
A3：同じ生物種のさらに別個体の**腎臓**
…
B1：ある生物の**肝臓**
B2：同じ生物種の別個体の**肝臓**
…

Biological replicatesのデータ
生物学的なばらつき（個体間の違い）を考慮すべし

倍率変化がだめな理由をデモ

- 例題：Marioni et al., *Genome Res.*, **18**: 1509-1517, 2008のデータ

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
EnsemblGeneID	R1 L1 Kidney	R1 L3 Kidney	R1 L7 Kidney	R2 L2 Kidney	R2 L6 Kidney	R1 L2 Liver	R1 L4 Liver	R1 L6 Liver	R1 L8 Liver	R2 L3 Liver
ENSG00000146556	0	0	0	0	0	0	0	0	0	0
ENSG00000197194	0	0	0	0	0	0	0	0	0	0
ENSG00000197490	0	0	0	0	0	0	0	0	0	0
ENSG00000205292	0	0	0	0	0	0	0	0	0	0
ENSG00000177693	0	0	0	0	0	0	0	0	0	0
ENSG00000209338	0	0	0	0	0	0	0	0	0	0
ENSG00000196573	0	0	0	0	0	0	0	0	0	0
ENSG00000177799	0	0	0	0	0	0	0	0	0	0
ENSG00000209341	0	0	0	0	0	0	0	0	0	0
ENSG00000209342	0	0	2	4	3	0	0	0	1	0
ENSG00000209343	0	0	0	0	0	0	0	0	0	0
ENSG00000209344	0	0								
ENSG00000209346	0	0								
ENSG00000209349	0	0	0	0	0	0	0	0	0	0
ENSG00000209350	4	7	3	6	7	35	32	31	29	34
ENSG00000209351	0	0	0	0	0	0	0	0	0	0
ENSG00000209352	0	0	1	1	0	2	0	0	0	0
ENSG00000212679	110	131	149	112	118	177	135	141	148	145
ENSG00000212678	12685	13204	12403	13031	13268	9246	9312	8746	8496	9070
ENSG00000185097	0	0	0	0	0	0	0	0	0	0

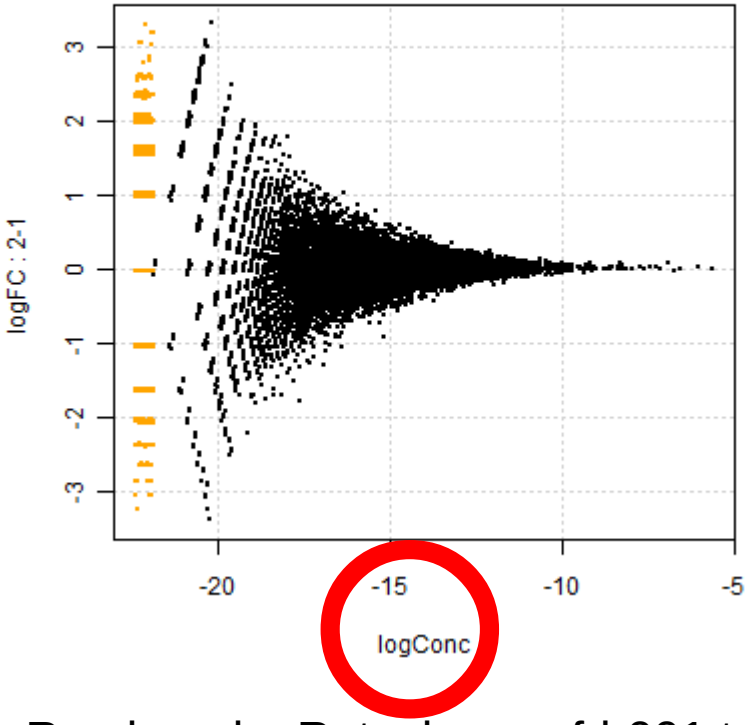
発現変動遺伝子がないデータで二群間比較をしてみる

A群　　B群

倍率変化がだめな理由をデモ

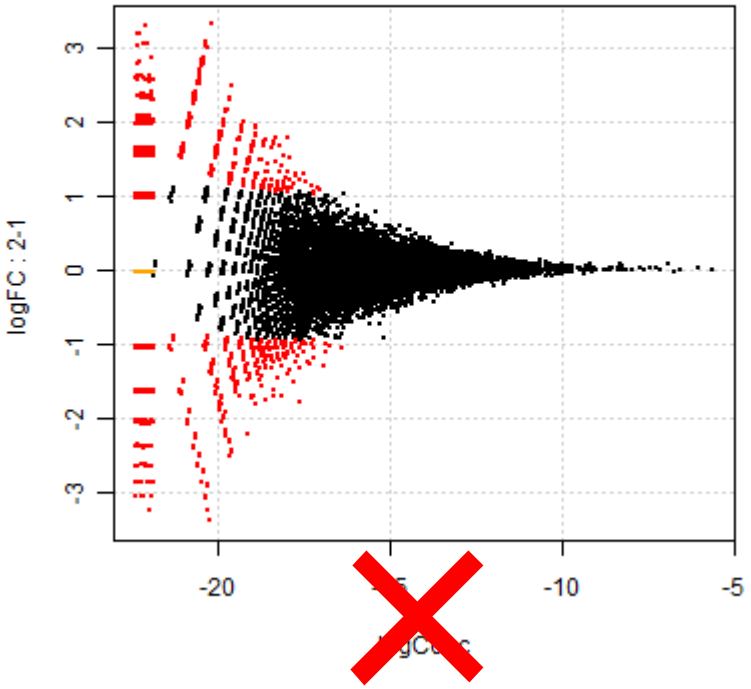
- 例題：Marioni et al., *Genome Res.*, **18**: 1509-1517, 2008のデータ（の一部）
  - (A1, A2) vs. (A3, A4)の二群間比較結果

*edgeR*でFDR < 0.01を満たすものは0個



Rcode_edgeR_tech_rep_fdr001.txt

（*edgeR*で）2倍以上発現変動しているものは3814個



Rcode_edgeR_tech_rep_fc2.txt

低発現領域でlog比が大きくなる現象をうまくモデル化することが重要