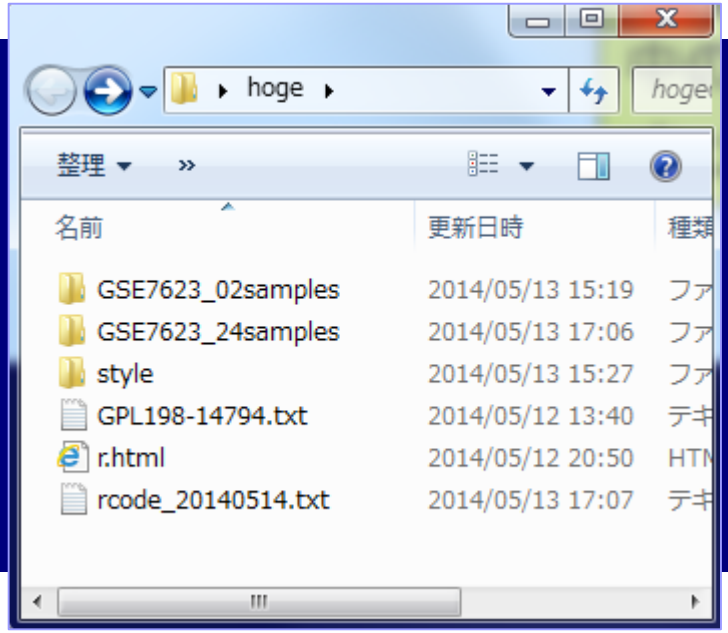
講義室後ろにあるUSBメモリ中のhogeフォルダをデスクトップにコピーしておいてください。



機能ゲノム学 第1回

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
kadota@iu.a.u-tokyo.ac.jp

前回（4/30）のhogeフォルダがデスクトップに残っているかもしれないのでご注意ください。

NGS速習コース開催（9/1~12@東大農）

実施日	実施時間	大項目	項目番号および項目	習得技術	レベル	形式	担当講師(敬称略)
9月1日	10:40-12:00	1. コンピュータリテラシーとサーバー設計	1-1. OS、ハード構成	コンピューターの基本の理解	初級	講義	中村保一(DDBJ)
	13:15-14:45		1-2. ネットワーク基礎	インターネット、セキュリティの基本の理解	初級	講義	中村保一(DDBJ)
	15:00-16:30		1-3. UNIX I	UNIXの基礎の理解 Linux導入	初級	実習	仲里猛留(DBCLS)
	16:45-18:15						仲里猛留(DBCLS)
9月2日	10:30-12:00				中級	実習	仲里猛留(DBCLS)
	13:15-14:45						仲里猛留(DBCLS)
	15:00-16:30						仲里猛留(DBCLS)
	16:45-18:15						仲里猛留(DBCLS)
9月3日	10:30-12:00		1-4. スクリプト言語	Perl シェルスクリプト	中級	実習	山口昌雄(アメリエフ)
	13:15-14:45						山口昌雄(アメリエフ)
	15:00-16:30						山口昌雄(アメリエフ)
	16:45-18:15						山口昌雄(アメリエフ)
9月4日	10:30-12:00						山口昌雄(アメリエフ)
	13:15-14:45						山口昌雄(アメリエフ)
	15:00-16:30						山口昌雄(アメリエフ)
	16:45-18:15						山口昌雄(アメリエフ)
9月5日	10:30-12:00	2. 配列インフォマティクス	2-1. 配列解析基礎	配列、ゲノムデータ記述のフォーマット、アラインメント(DP)、 データベース検索(BLAST、BLAT)等の基礎的な配列比較解析の原理と実習	初級	実習	坊農秀雅(DBCLS)
	13:15-14:45						坊農秀雅(DBCLS)
	15:00-16:30		2-2. バイオ系データベース概論	基本的な各種バイオ系データベースの理解、統合DBの利用法	初級	実習	小野浩雅(DBCLS)
	16:45-18:15						小野浩雅(DBCLS)
9月8日	10:30-12:00	3. データ解析基礎	3-1. R基礎1	R言語の基礎(インストールから利用まで)	初級	実習	門田幸二(東京大学)
	13:15-14:45		3-2. R基礎2	ファイルの読み込み、行列演算の基本	初級	実習	門田幸二(東京大学)
	15:00-16:30		3-3. R各種パッケージ	Rの各種パッケージのインストール法と代表的なパッケージの利用法	中級	実習	門田幸二(東京大学)
	16:45-18:15						門田幸二(東京大学)
9月9日	10:30-12:00		3-4. R bioconductor I	Bioconductorの利用法	中級	実習	門田幸二(東京大学)
	13:15-14:45						門田幸二(東京大学)
	15:00-16:30		3-5. R bioconductor II	FASTAandFASTQ形式ファイルの読み込み ファイル形式の変換(FASTQ->FASTA)、クオリティチェック、 リード配列長分布、フィルタリングやトリミング、GC含量計算など	中級	実習	門田幸二(東京大学)
	16:45-18:15						門田幸二(東京大学)
9月10日	10:30-12:00	4. 次世代シークエンサ	4-1. 次世代シークエンサ基礎I	原理の理解	初級	講義	倉田哲也(NAIST)
	13:15-14:45		4-2. 次世代シークエンサ基礎II	応用分野とそのための計測技術の理解 (RNA-seq、ChIP-seq、がんゲノム、個人ゲノム、環境ゲノム、Hi-C)	初級	講義	倉田哲也(NAIST)
	15:00-16:30		4-3. 次世代シークエンサ実習I	ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習	山口昌雄(アメリエフ)
	16:45-18:15						山口昌雄(アメリエフ)
9月11日	10:30-12:00		4-4. 次世代シークエンサ実習II	代表的なパイプラインについての実習：多型解析(IGV)	初級	実習	山口昌雄(アメリエフ)
	13:15-14:45						山口昌雄(アメリエフ)
	15:00-16:30			代表的なパイプラインについての実習：RNA-seq			山口昌雄(アメリエフ)
	16:45-18:15						山口昌雄(アメリエフ)
9月12日	10:30-12:00			代表的なパイプラインについての実習：ChIP-seq、及び統合解析			河岡慎平(ATR)
	13:15-14:45						河岡慎平(ATR)
	15:00-16:30	6. 分子生命科学	6-1. 分子生命科学概論	複製、転写、翻訳、代謝、シグナル伝達などの基礎知識	初級	講義	河岡慎平(ATR)
			6-2. オミクス概論	ゲノム以外のオミクスデータの基礎知識			
			6-3. 遺伝/進化概論	ゲノムデータを扱う上での遺伝学、進化学の基礎知識			
	16:45-18:15	5. ゲノム関連の倫理・法律	5-1. ゲノム情報倫理概論	ゲノム情報を扱う上で、プライバシー保護などの必要な倫理的問題、法的問題の国内外の状況を理解し、ゲノム情報を適切に利用できるようにする。匿名化、暗号化、情報セキュリティ概要			笠輪真理(NBDC) 川嶋実苗(NBDC)

申込み受付は6/2の11:00-（おそらくそれで枠がいっぱいになるので）**2週間全て参加可能な本当にやる気のあるヒトのみ**募集。
全国から募集（アグリバイオの講義とは無関係）

講義予定

- 第1回（2014年5月14日）
  - 原理、各種データベース、生データ取得、遺伝子発現行列作成（データ正規化）
  - 教科書の1.2節、2.2節周辺
- 第2回（2014年5月21日）
  - クラスタリング（データ変換や距離の定義など）、実験デザイン、分布
  - 教科書の3.2節周辺
- 第3回（2014年5月28日）
  - 発現変動解析（多重比較問題）、各種プロット（M-A plotや平均-分散プロット）
  - 教科書の3.2節と4.2節周辺
- 第4回（2014年6月4日）
  - 機能解析（Gene Ontology解析やパスウェイ解析）、分類など

授業の目標・概要

細胞中で発現している全転写物（トランスクリプトーム）の解析技術は、マイクロアレイから次世代シーケンサ（RNA-seq）に移行しつつあります。RNA-seqデータ解析の多くは、マイクロアレイの知識を前提としています。また、ニュートリゲノミクス（食品系）分野では、マイクロアレイは現在でも主流派です。**マイクロアレイデータを主な例として**、各種トランスクリプトーム解析手法について解説します。

Contents（第1回）

- イントロダクション
  - マイクロアレイの原理や特徴（長所・短所）
  - データ解析例とバイオインフォマティクス要素技術
  - 発現データベース（DB）
  - Affymetrix GeneChipの用語：CELファイル、プローブセット、summarization･･･
- 発現DBからのプローブレベルデータ取得
  - GEOウェブサイト経由
  - R経由（教科書の§2.2.1）
- 前処理法の適用（プローブレベルデータ → 発現行列データ）
  - MAS法、RMA法、RMX法（RobLoxBioC）、IRON法（教科書の§2.2.2～2.2.4）
  - データの正規化（グローバル正規化、quantile正規化）

トランスクリプトームとは

- ある特定の状態の組織や細胞中に存在する全RNA（転写物、transcripts）の総体
- 様々なトランスクリプトーム解析技術
  - マイクロアレイ
    - cDNAマイクロアレイ、Affymetrix GeneChip、タイリングアレイなど
  - 配列決定に基づく方法
    - EST、SAGE、CAGE、次世代シーケンサ(RNA-seq)など
  - （電気泳動に基づく方法）
    - Differential Display、AFLP、HiCEPなど

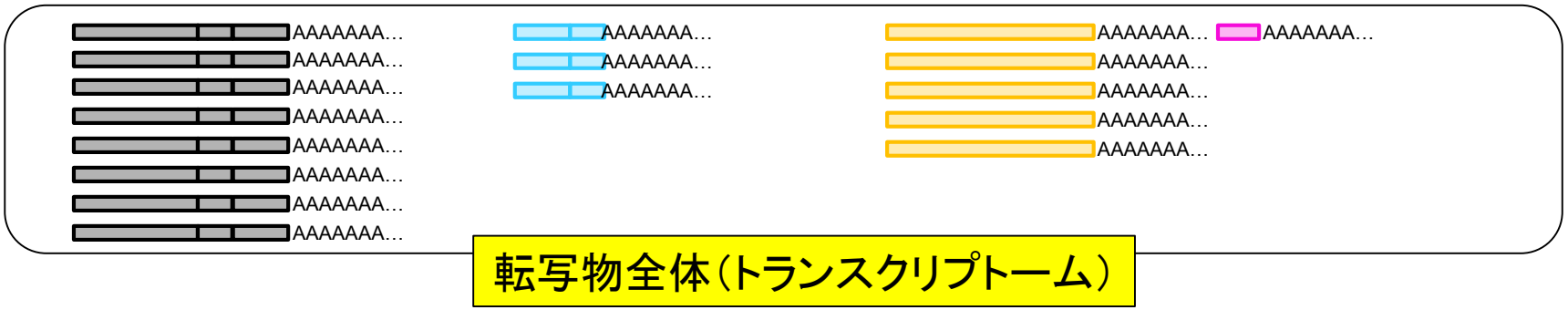
調べたい組織でどの遺伝子がどの程度発現しているのかを一度に観察

トランスクリプトームとは

- ある状態のあるサンプル（例：目）のあるゲノムの領域

ヒト

・どの染色体上のどの領域にどの遺伝子があるかは調べる個体（例：ヒト）が同じなら不変（目だろうが心臓だろうが…）

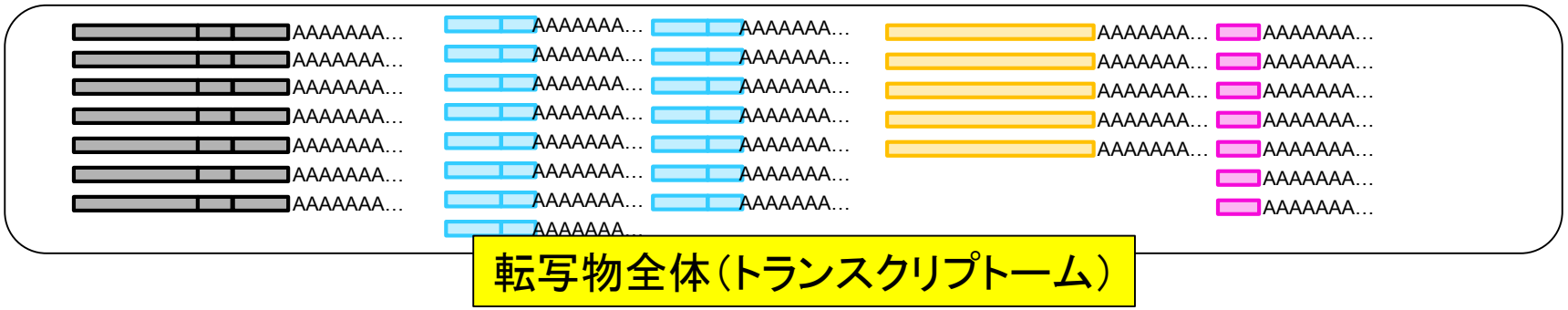


・遺伝子1は沢山転写されている（発現している）
・遺伝子4はごくわずかしか転写されてない
・…

トランスクリプトームとは

- ある状態のあるサンプル（例：目）のあるゲノムの領域

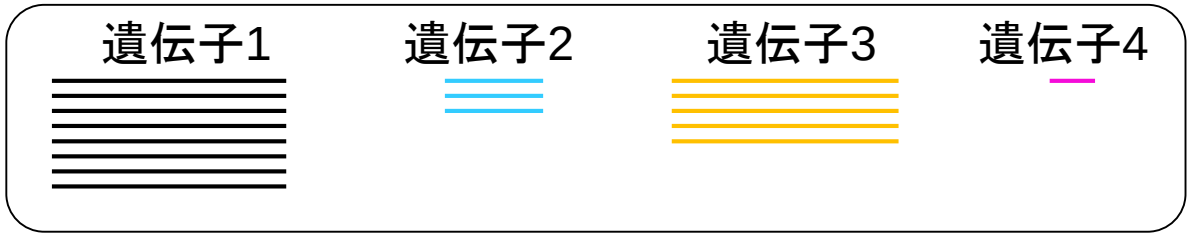
・どの染色体上のどの領域にどの遺伝子があるかは調べる個体（例：ヒト）が同じなら不変（目だろうが心臓だろうが…）



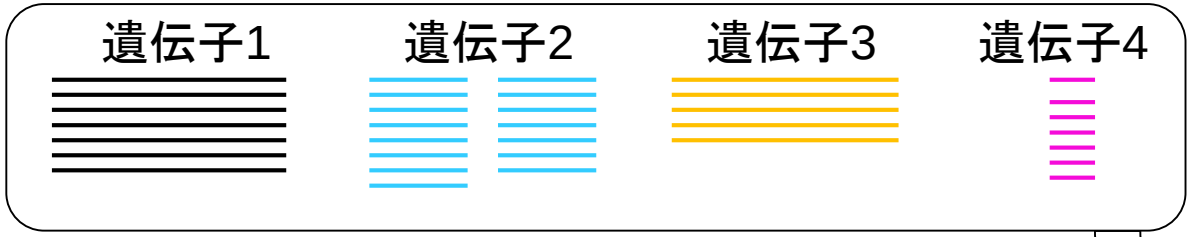
・遺伝子2は光刺激に応答して発現亢進
・遺伝子4も光刺激に応答して発現亢進

トランスクリプトーム情報を得る手段

- 光刺激前（T1）の目のトランスクリプトーム



- 光刺激後（T2）の目のトランスクリプトーム



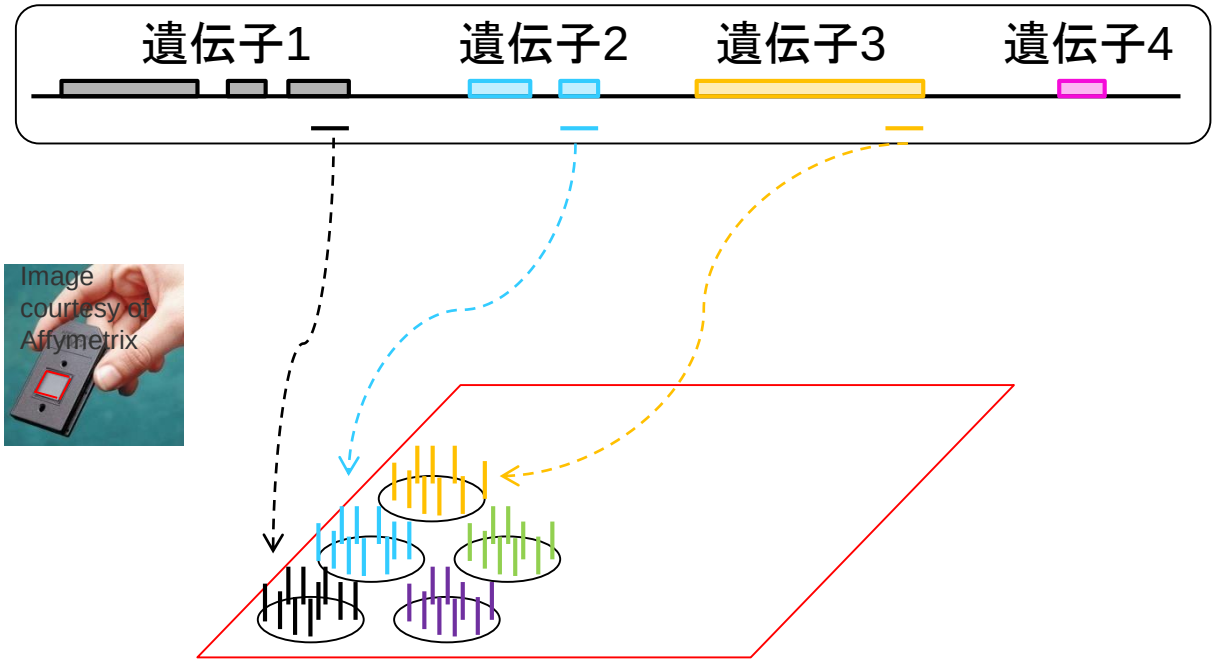
これがいわゆる
「遺伝子発現行列」

	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
…	…	…

・マイクロアレイ
・RNA-seq

トランスクリプトーム取得（マイクロアレイ）

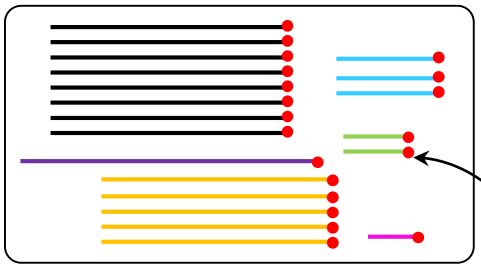
- よく研究されている生き物は多数の遺伝子（の配列情報）がわかっている



わかっている遺伝子（の配列の相補鎖）を搭載した”チップ”

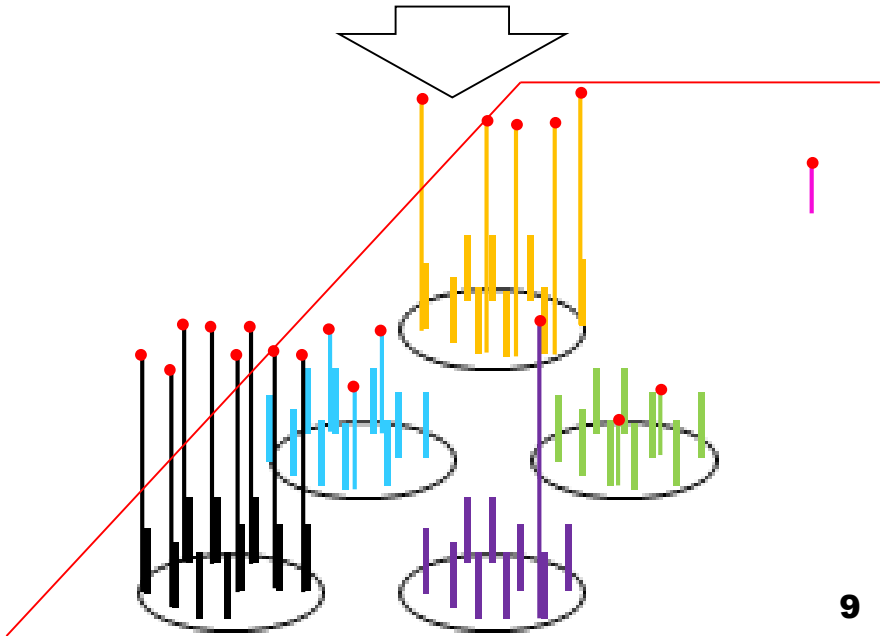
・メーカーによって搭載されている遺伝子の種類が異なる
→搭載されていない遺伝子（未知遺伝子含む、例：遺伝子4）の発現情報は測定不可…

光刺激前（T1）の目のトランスクリプトーム



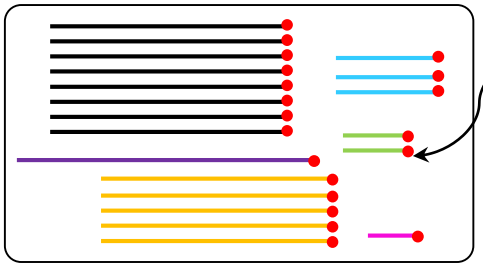
蛍光標識

ハイブリダイゼーション
（二本鎖形成）

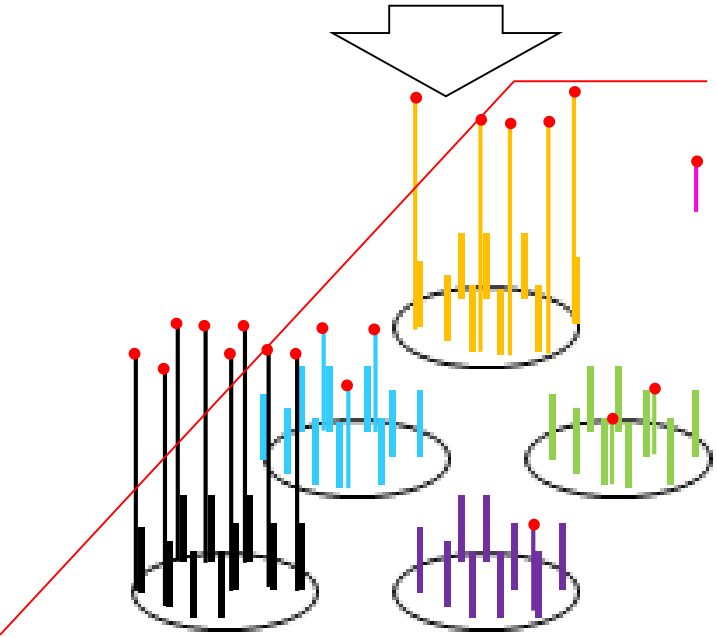


マイクロアレイデータ → 遺伝子発現行列

- 光刺激前（T1）の目のトランスクリプトーム

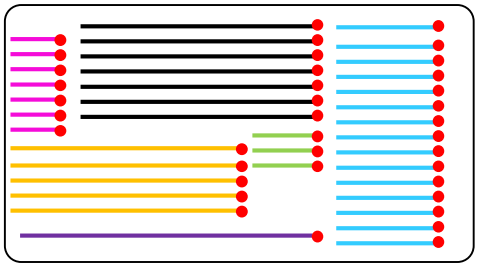


ハイブリダイゼーション
（二本鎖形成）



専用の検出器で各
遺伝子に対応する
領域の蛍光シグナ
ル強度を測定

光刺激後（T2）の目の
トランスクリプトーム



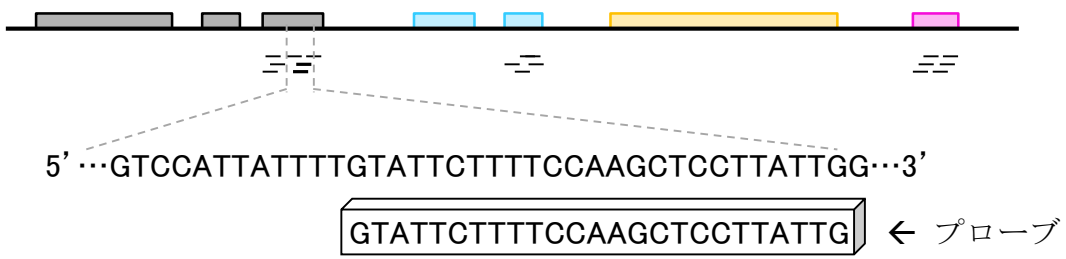
ハイブリダイゼーション
と
シグナル検出

	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	?	?
遺伝子5	…	…
…	…	…

正規化

ステレオタイプなイメージ

- マイクロアレイの長所
  - 取り扱いやすいデータ量（~100Mb程度）
  - 長年の実績：解析手法がほぼ確立。（Windows Rのみで解析可能）
  - 検査用チップが利用可能（MammaPrintなど）
- マイクロアレイの短所
  - 解析可能範囲が搭載転写物に限定
  - プローブが3’末端に偏っている（3’発現解析用アレイ）
  - ダイナミックレンジが狭い



マイクロアレイの実用例

- MammaPrint：乳癌予後予測検査サービス（2008年）
  - 乳癌手術を受けた患者の転移・再発の可能性に関する情報提供
  - 70遺伝子の活性を測定
  - 不必要な補助化学療法などを避けることが可能（ローリスク群）
- 安全に登山をするための新たなバイオマーカー、heme oxigenase-1（HO-1）の発見
  - **背景**：エベレスト頂上は酸素濃度が薄く、通常は10秒程度で意識を失うらしい。三浦雄一郎氏（登山家）がなぜ酸素ボンベなしでエベレスト頂上で数十秒もコメントできるのか？
  - **実験**：低酸素室滞在**前後**の白血球の遺伝子発現変化をマイクロアレイで調査した結果、エベレスト登頂経験者はHO-1が低酸素刺激で特異的に変化（発現上昇）
  - **結論**：HO-1は、生体が低酸素に曝されたときに血管を拡張したり、低酸素による酸化ストレス傷害に対して抗酸化作用を示したりする。エベレスト登頂経験者はHO-1濃度を高めて高度への順化を行うために、低酸素時にも血管が拡張しやすく末梢循環が保たれているのだろう。

ただし、HO-1が搭載されていないマイクロアレイでは測定不可能！

http://www.jmac.or.jp/ja/home-ja/news/133-news20131017.html 参考

JMAC 特定非営利活動法人 バイオチップコンソーシアム
Japan Micro Array Consortium

HOME | JMACとは | バイオチップとは | 活動内容 | 入会申し込み | リンク

2013/10/17 [プレスリリース] 世界初、マイクロアレイ関連技術の国際標準化を達成

世界初、マイクロアレイ関連技術の国際標準化を達成
～ ISO最終国際規格案、各国投票で承認 ～

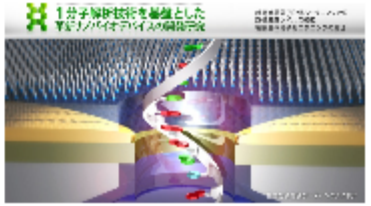
このたび、特定非営利活動法人バイオチップコンソーシアム（以下JMAC）は、当コンソーシアムが中心となって、提案・手続きを進めてまいりましたマイクロアレイに関わる世界初の規格案が、国際標準として承認されましたことをお知らせいたします。

記

マイクロアレイとは、遺伝子検査に主に用いられるバイオチップ[*1]の一種で、基板上に一群のDNAプローブ（特定の遺伝子を検出するためのDNA断片）を高密度で付着させて配置し、直接あるいは間接的に、大量の生物由来物質を高スループットで分析するものです。例えばヒトの全遺伝子のような、多種類の合成DNAを基板上に配置して、遺伝子の活動を示す発現量を測定するマイクロアレイが研究分野で使われています。2000年にヒトゲノムの配列が解明されてから、ヒトの遺伝子配列が理論的には全て得られるため、これまではバイオの研究現場で、全遺伝子の活動を調べる用途などに使われて来ました。

近年マイクロアレイは、食品分野をはじめ、環境や医療などの分野における遺伝子検査に利用される動きが活発になり、これまでの研究用途から産業用途へと、市場がシフトしてきました。これはマイクロアレイに関する技術が成熟度を増し、分析精度や再現性が大幅に向上したことが原因と考えられます。しかし、産業利用のためには、開発各社で解決できない問題もありました。その1つが、測定精度の管理です。研究に用いるのであれば、1つの解析対象に対し、複数のチップを用いて結果を取得した後で統計解析を行うことも可能です。しかし産業利用のためには、主にコストの面から、1度の解析により期待される精度で結果を得る必要があります。このような背景から、まずマイクロアレイの測定精度に関わる用語の定義を統一し、それに加えて精度保証のための手段を提供するような標準が必要となっていました。

最先端研究開発支援プログラム
（FIRSTプログラム）



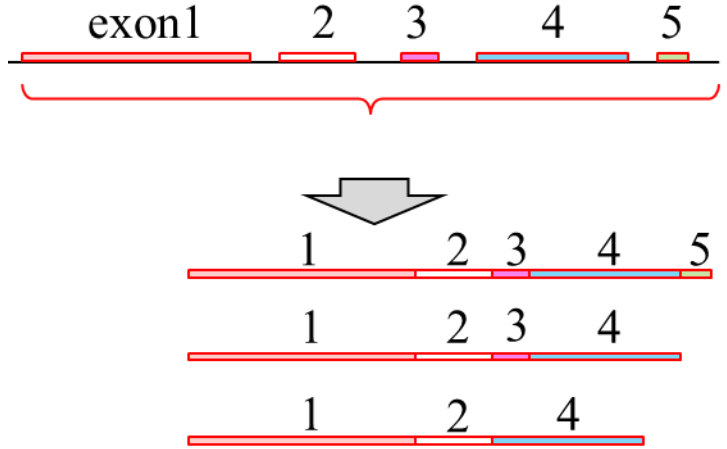
サイト内検索
検索...
クエリ送信

ログインフォーム
ユーザ名
パスワ
自動ログイン
ログイン
パスワードを忘れましたか？
ユーザ名を忘れましたか？

主に産業界の活動

ステレオタイプなイメージ

- RNA-seqの短所
  - 取り扱いづらいデータ量（数百Gb?!）
  - Windows userは自力解析が困難（ほとんどがLinux用）
  - ダイナミックレンジが広いがために?!変な結果に遭遇。
  - ゼロカウントデータの取り扱い
- RNA-seqの長所
  - （多少のoff-targetは含むが）全発現転写物の解析が可能
  - 解像度：遺伝子レベル → 転写物レベル
  - ダイナミックレンジが広い

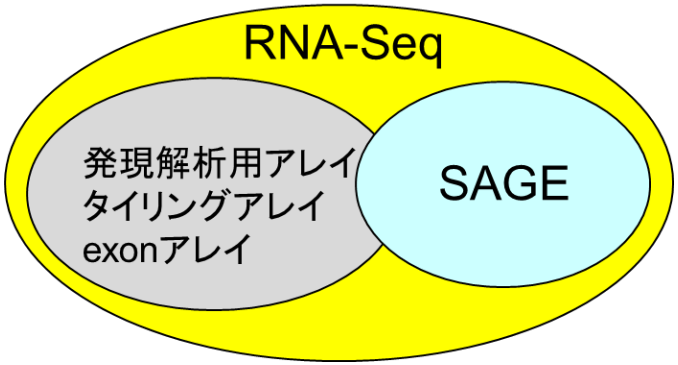


マイクロアレイ

- 機能（遺伝子セット）解析が主目的の場合にはまだ主役
  - Gene Ontology解析やパスウェイ解析
    - 実績のある市販アレイに搭載されている遺伝子のみでも「この栄養素はこのパスウェイに効いている」的な新規知見が得られればよい、という思想
    - 「個別の遺伝子の変動解析」というよりは「遺伝子セットの変動解析」
  - 同一アレイを用いている限り全体的な情報量が豊富
    - 公共データベース（GEO, ArrayExpressなど）
    - 3’発現解析用アレイが未だに使われる所以
  - 異なるアレイであっても同一生物種であればマージ可能
    - virtualArray (Heider and Alt, *BMC Bioinformatics*, **14**:75, 2013)など

意義

- 全体的にはマイクロアレイ → RNA-seq
  - 転写物全体の配列情報を取得可能（RefSeqのようなmulti-fasta形式のファイルをゲットできるイメージ）
  - 選択的スプライシングの全体像の理解
  - 発現変動exonや転写物の同定

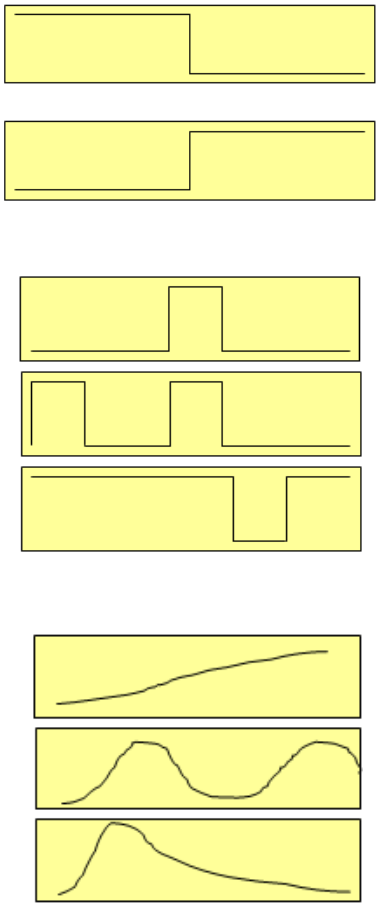


発行日：2013年9月30日

・データ解析の基本的な考え方はマイクロアレイと同じ
・食品系の研究（Neutrigenomics）はアレイが未だ主流

データ解析もいろいろ

発現変動遺伝子同定



遺伝子発現行列

二群間比較用

	A群 A1	A2	…	B群 B1	B2	…
gene *1*	$x^A_{1,1}$	$x^A_{1,2}$		$x^B_{1,2}$	$x^B_{1,2}$	
gene *2*	$x^A_{2,1}$	$x^A_{2,2}$		$x^B_{2,2}$	$x^B_{2,2}$	
…	…	…	…	…	…	…
gene *i*	$x^A_{i,1}$	$x^A_{i,2}$		$x^B_{i,2}$	$x^B_{i,2}$	
…	…	…	…	…	…	…
gene *n*	$x^A_{n,1}$	$x^A_{n,2}$		$x^B_{n,2}$	$x^B_{n,2}$	

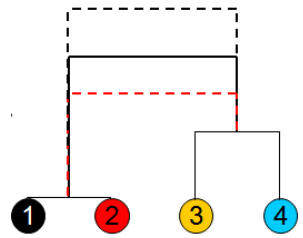
様々な組織（条件）

	S1	S2	S3	S4	…
gene *1*	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	…
gene *2*	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	…
…	…	…	…	…	…
gene *i*	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	
…	…	…	…	…	…
gene *n*	$x_{n,1}$	$x_{n,2}$	$x_{n,3}$	$x_{n,4}$	…

時系列データ

	T1	T2	T3	T4	…
gene *1*	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	…
gene *2*	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	…
…	…	…	…	…	…
gene *i*	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	
…	…	…	…	…	…
gene *n*	$x_{n,1}$	$x_{n,2}$	$x_{n,3}$	$x_{n,4}$	…

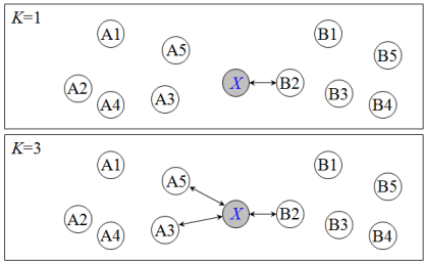
クラスタリング



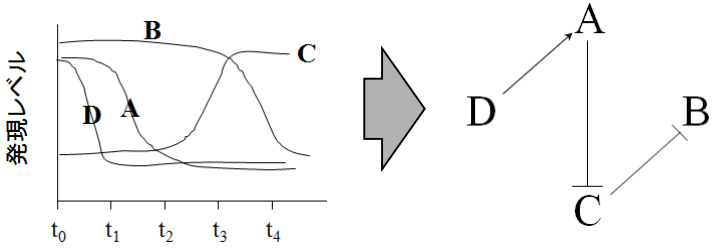
機能解析

- Gene Ontology（GO）
- パスウェイ解析

分類（診断）



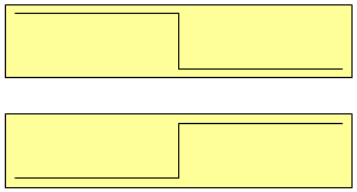
遺伝子ネットワーク推定



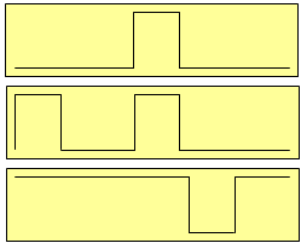
バイオインフォマティクス要素技術

- 相関係数やエントロピーなどの応用例を紹介

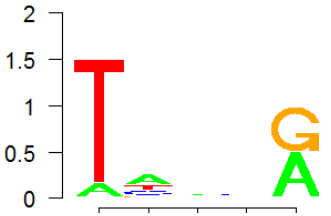
二群間比較



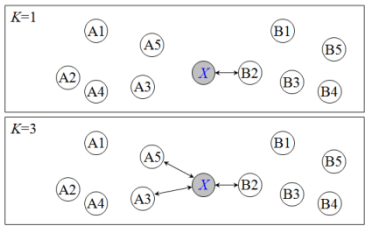
組織特異的遺伝子



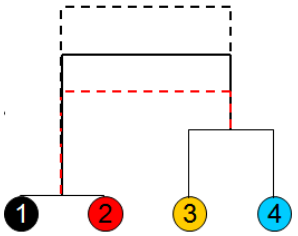
Sequence logo



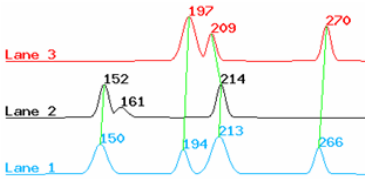
分類（診断）



クラスタリング

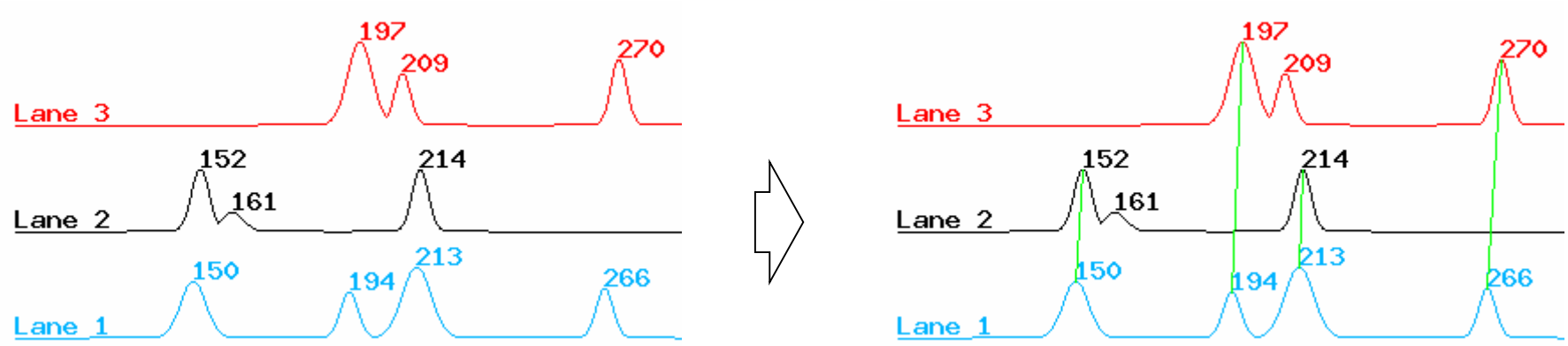


同一ピーク同定



基本スキルのみでいろいろなことができます

クラスタリングの考えを同一ピーク認識に応用



②ピーク間距離を計算

②'クラスター間距離が最短のものをマージ

Lane	M.W.
1	150
1	194
1	213
1	266
2	152
2	161
2	214
3	197
3	209
3	270

①分子量でソート

Lane	M.W.	
1	150	2
2	152	9
2	161	33
1	194	3
3	197	12
3	209	4
1	213	1
2	214	52
1	266	4
3	270	

c.	TDF
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

Contents（第1回）

- イントロダクション
  - マイクロアレイの原理や特徴（長所・短所）
  - データ解析例とバイオインフォマティクス要素技術
  - 発現データベース（DB）
  - Affymetrix GeneChipの用語：CELファイル、プローブセット、summarization･･･
- 発現DBからのプローブレベルデータ取得
  - GEOウェブサイト経由
  - R経由（教科書の§2.2.1）
- 前処理法の適用（プローブレベルデータ → 発現行列データ）
  - MAS法、RMA法、RMX法（RobLoxBioC）、IRON法（教科書の§2.2.2～2.2.4）
  - データの正規化（グローバル正規化、quantile正規化）

発現データベース（DB）


+サイトマップ +English
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
Agricultural Bioinformatics Research Unit
受講生の方へ
研究者の方へ
+ホーム
+本ユニットについて
+メンバー
+教育プログラム
+研究フォーラム
+イベント
+お問い合わせ
+リンク
+モバイルサイト
ホーム > 教育プログラム > 各講義のページ > 9.機能ゲノム学
9.機能ゲノム学
授業の目標・概要
細胞中で発現している全転写物（トランスクリプトーム）の解析技術は、マイクロアレイから次世代シーケンサ（RNA-seq）に移行しつつあります。RNA-seqデータ解析の多くは、マイクロアレイの知識を前提としています。また、ニュートリゲノミクス（食品系）分野では、マイクロアレイは現在でも主流派で
ム解析手法について解説し
担当教員
門田幸二 （東大・農・アク
お知らせ
講義では、Rの様々なパッケ
起動を参考にして必要なパ
参考図書
受講者は以下の教科書を用
門田幸二 著 (金明哲 編)、
2014年4月。ISBN:978-4
東京大学
THE UNIVERSITY OF TOKYO
参考図書
受講者は以下の教科書を用意してください。
門田幸二 著 (金明哲 編)、「シリーズ Useful R ⑦ トランスクリプトーム解析」、共立出版、
2014年4月。ISBN:978-4-320-12370-0
講義日程（平成26年度）
1. 平成26年05月14日
(Rで)マイクロアレイデータ解析
2. 平成26年05月21日
3. 平成26年05月28日
4. 平成26年06月04日
(Rで)マイクロアレイデータ解析です

http://www.iu.a.u-tokyo.ac.jp/~kadota/r.html (Rで)マイクロアレイデー...

(Rで)マイクロアレイデータ解析

(last modified 2014/05/09, since 2005)

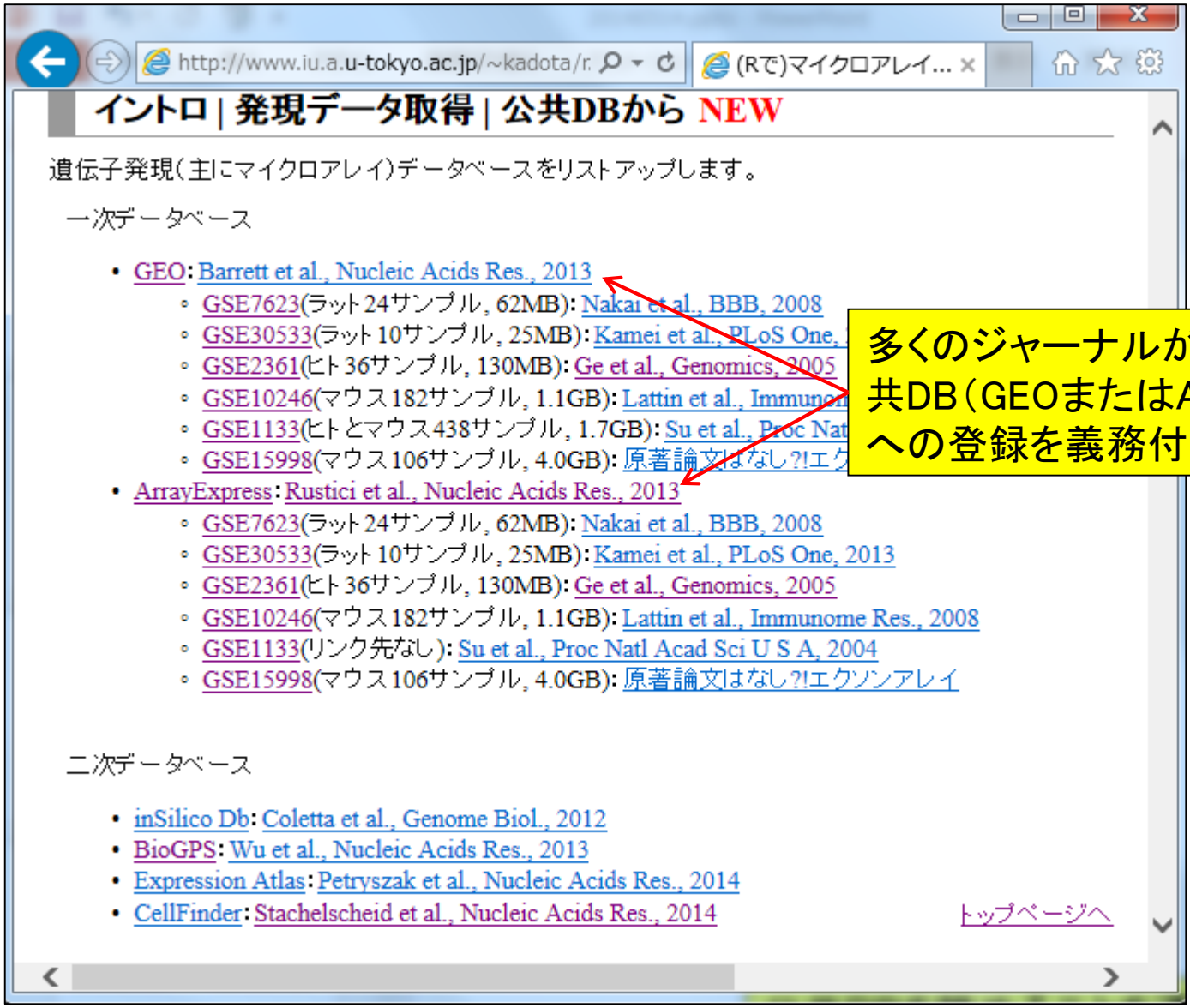
What's new?

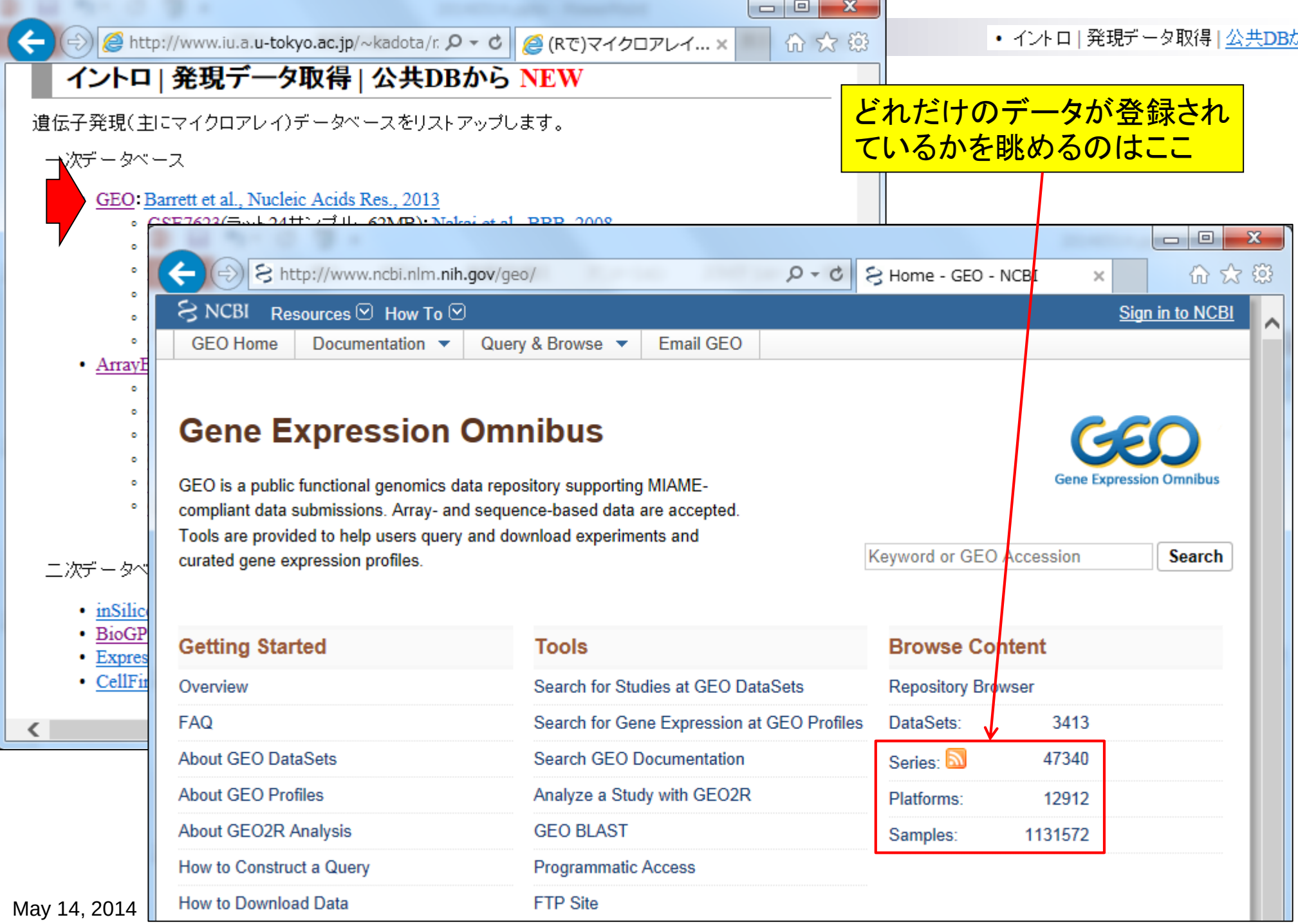
- 門田幸二 著シリーズ Useful R 第7巻トランスクリプトーム解析刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法 (Kadota et al., 2006)、WAD法 (Kadota et al., 2008)などについての解説も含んでいます。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍 | トランスクリプトーム解析 | ...」に掲載してあります。(2014/04/27) NEW
- お知らせは主に(Rで)塩基配列解析で行っておりますのでそちらをご覧ください。(2014/03/05)

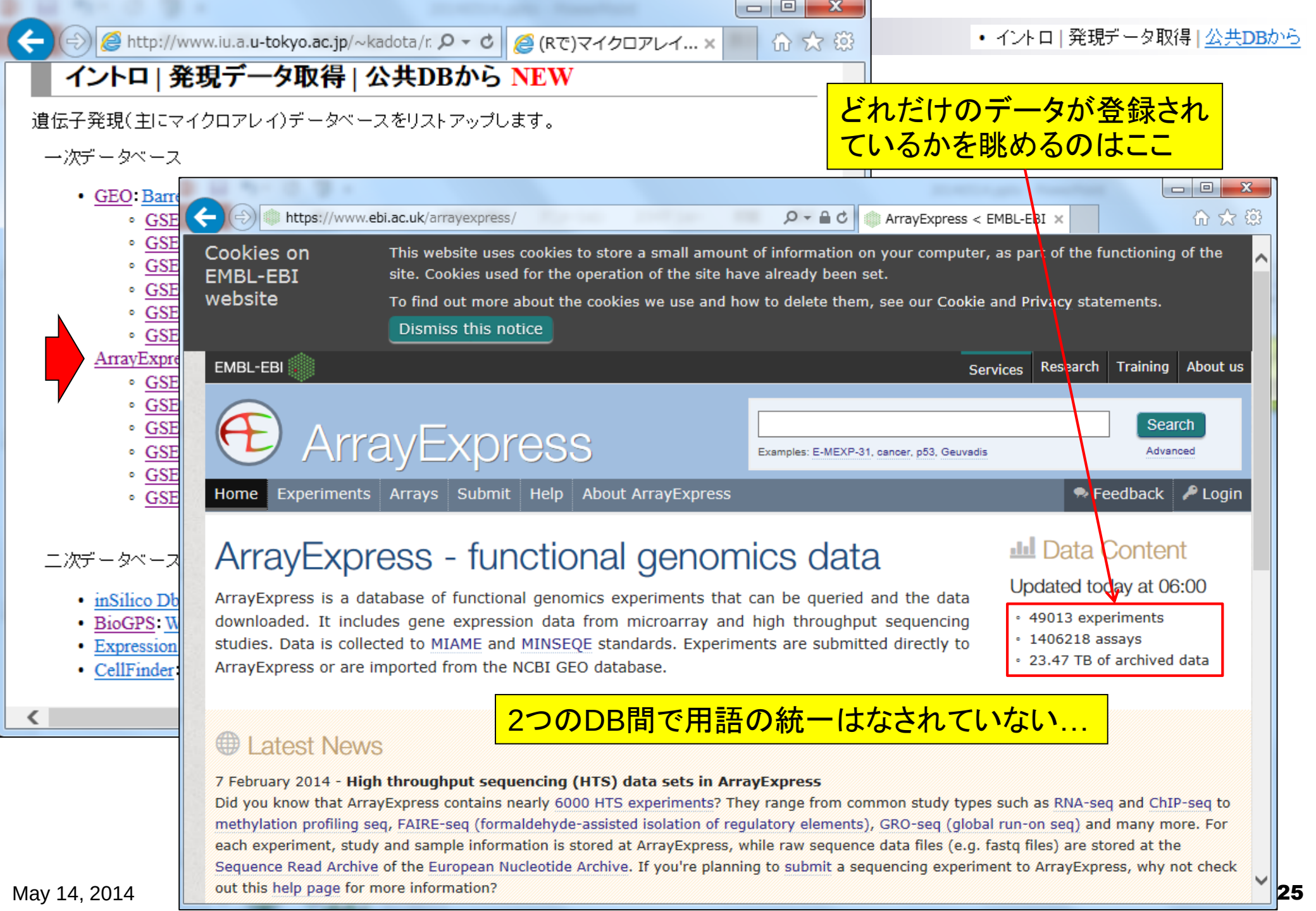
- はじめに(last modified 2014/01/21)
- 過去のお知らせ(last modified 2014/03/03)
- Rのインストールと起動(last modified 2014/04/17) NEW
- Rの昔のバージョンのインストール(last modified 2012/04/07)
- 使用例(初心者向け)(last modified 2011/09/15)
- サンプルデータ (last modified 2013/11/25)
- 書籍 |について (last modified 2014/04/17) NEW
- 書籍 | トランスクリプトーム解析 | 1.1 はじめに (last modified 2014/05/09) NEW
- 書籍 | トランスクリプトーム解析 | 1.2.1 原理(Affymetrix 3'発現アレイ) (last modified 2014/05/09) NEW
- 書籍 | トランスクリプトーム解析 | 1.2.2 最近の知見 (last modified 2014/05/09) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.1 生データ(プローブレベルデータ)取得 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.2 データの正規化(基礎) (last modified 2014/04/17) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.3 データの正規化(計算例) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.4 データの正規化(その他) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.5 アノテーション情報 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.1 クラスタリング(データ変換や距離の定義など) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.2 実験デザイン, データ分布, 統計解析との関係 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.3 多重比較問題 (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.4 各種プロット(M-A plotや平均-分散プロットなど) (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.1 2群間比較 (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.2 他の実験デザイン(paired, multi-factor, 3群間) (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.3 多[illegible]比較(特異的発現パターン) (last modified 2014/04/20) NEW
- イントロ | 発現データ取得 | 公共DBから[illegible]odified 2014/05/10) NEW
- イントロ | 発現データ取得 | inSilicoDb(Tar[illegible]au 2011) (last modified 2013/08/20)
- イントロ | 発現データ取得 | ArrayExpress(Kauffmann 2009) (last modified 2013/08/29)推奨
- イントロ | 発現データ取得 | GEOquery(Davis 2007) (last modified 2013/08/20)

公共DBを眺めることを通じて、**3'発現アレイ**といわれる所以を知ろう

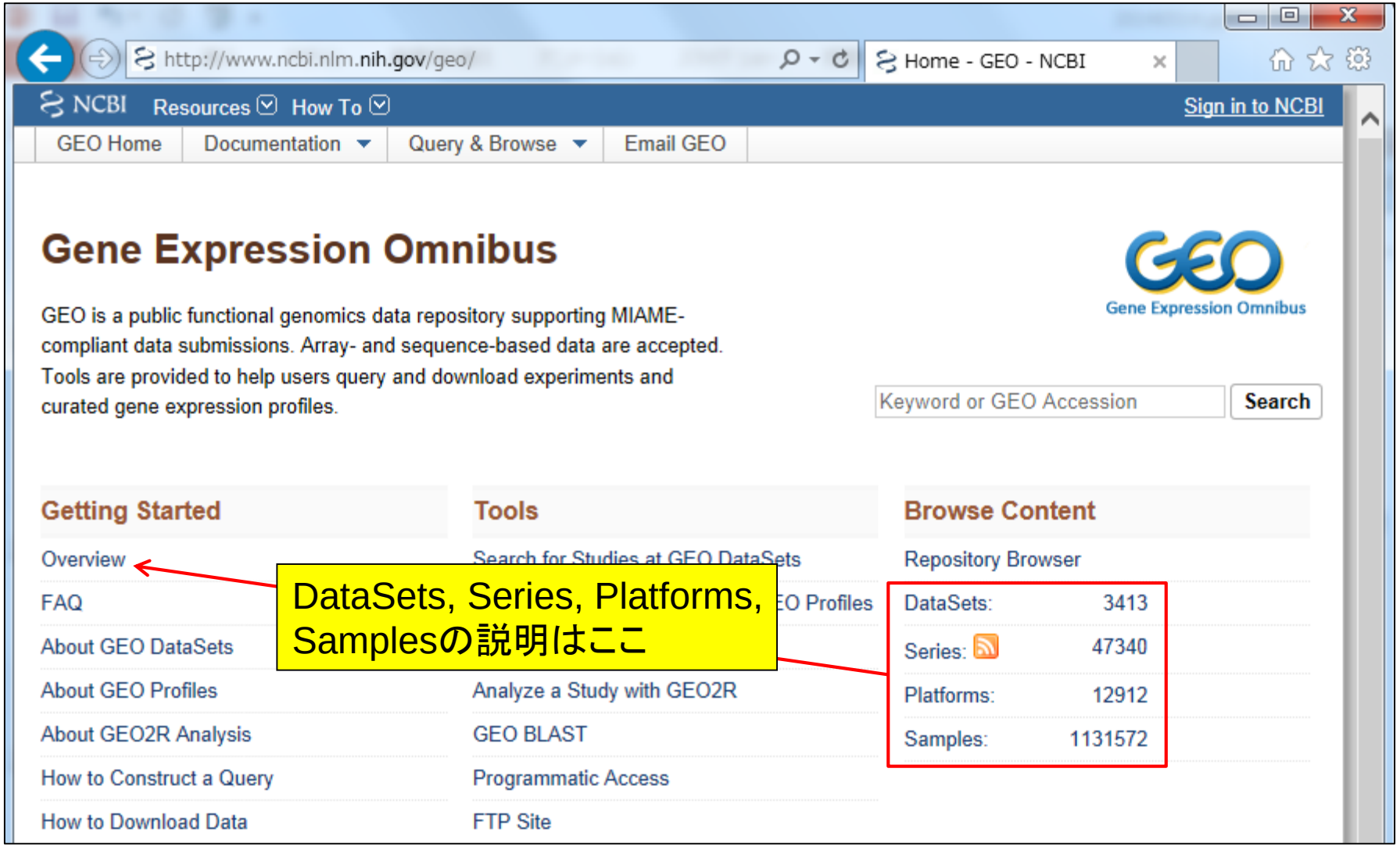
発現DB

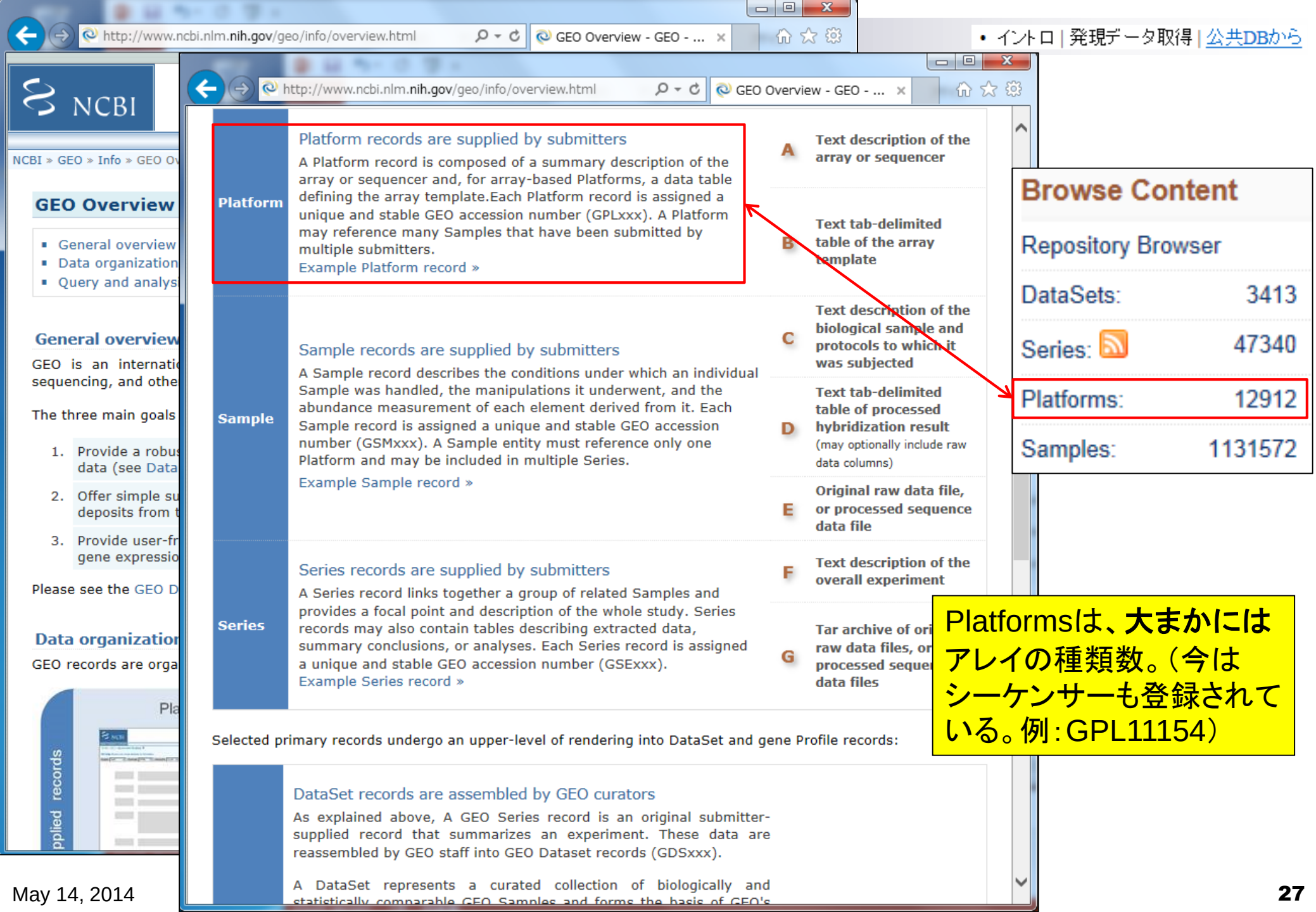

http://www.iu.a.u-tokyo.ac.jp/~kadota/r
(Rで)マイクロアレイ...
イントロ | 発現データ取得 | 公共DBから NEW
遺伝子発現(主にマイクロアレイ)データベースをリストアップします。
一次データベース
GEO: Barrett et al., Nucleic Acids Res., 2013
GSE7623(ラット24サンプル, 62MB): Nakai et al., BBB, 2008
GSE30533(ラット10サンプル, 25MB): Kamei et al., PLoS One,
GSE2361(ヒト36サンプル, 130MB): Ge et al., Genomics, 2005
GSE10246(マウス182サンプル, 1.1GB): Lattin et al., Immuno
GSE1133(ヒトとマウス438サンプル, 1.7GB): Su et al., Proc Nat
GSE15998(マウス106サンプル, 4.0GB): 原著論文はなし?!エク
ArrayExpress: Rustici et al., Nucleic Acids Res., 2013
GSE7623(ラット24サンプル, 62MB): Nakai et al., BBB, 2008
GSE30533(ラット10サンプル, 25MB): Kamei et al., PLoS One, 2013
GSE2361(ヒト36サンプル, 130MB): Ge et al., Genomics, 2005
GSE10246(マウス182サンプル, 1.1GB): Lattin et al., Immunome Res., 2008
GSE1133(リンク先なし): Su et al., Proc Natl Acad Sci U S A, 2004
GSE15998(マウス106サンプル, 4.0GB): 原著論文はなし?!エクソンアレイ
二次データベース
inSilico Db: Coletta et al., Genome Biol., 2012
BioGPS: Wu et al., Nucleic Acids Res., 2013
Expression Atlas: Petryszak et al., Nucleic Acids Res., 2014
CellFinder: Stachelscheid et al., Nucleic Acids Res., 2014
トップページへ
多くのジャーナルが生データの公共DB(GEOまたはArrayExpress)への登録を義務付けている。


http://www.iu.a.u-tokyo.ac.jp/~kadota/r
(Rで)マイクロアレイ...
イントロ | 発現データ取得 | 公共DBから NEW
遺伝子発現(主にマイクロアレイ)データベースをリストアップします。
一次データベース
GEO: Barrett et al., Nucleic Acids Res., 2013
二次データベ
イントロ | 発現データ取得 | 公共DBから
どれだけのデータが登録されているかを眺めるのはここ
http://www.ncbi.nlm.nih.gov/geo/
Home - GEO - NCBI
NCBI Resources How To
Sign in to NCBI
GEO Home Documentation Query & Browse Email GEO
Gene Expression Omnibus
GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.
Gene Expression Omnibus
Keyword or GEO Accession
Search
Getting Started
Overview
FAQ
About GEO DataSets
About GEO Profiles
About GEO2R Analysis
How to Construct a Query
How to Download Data
Tools
Search for Studies at GEO DataSets
Search for Gene Expression at GEO Profiles
Search GEO Documentation
Analyze a Study with GEO2R
GEO BLAST
Programmatic Access
FTP Site
Browse Content
Repository Browser
DataSets: 3413
Series: 47340
Platforms: 12912
Samples: 1131572


イントロ | 発現データ取得 | 公共DBから NEW
遺伝子発現(主にマイクロアレイ)データベースをリストアップします。
一次データベース
GEO: Barr
GSE
GSE
GSE
GSE
GSE
GSE
ArrayExpre
GSE
GSE
GSE
GSE
GSE
GSE
二次データベース
inSilico Db
BioGPS: W
Expression
CellFinder
イントロ | 発現データ取得 | 公共DBから
どれだけのデータが登録されているかを眺めるのはここ
https://www.ebi.ac.uk/arrayexpress/
ArrayExpress < EMBL-EBI
Cookies on EMBL-EBI website
This website uses cookies to store a small amount of information on your computer, as part of the functioning of the site. Cookies used for the operation of the site have already been set.
To find out more about the cookies we use and how to delete them, see our Cookie and Privacy statements.
Dismiss this notice
EMBL-EBI
Services
Research
Training
About us
ArrayExpress
Search
Examples: E-MEXP-31, cancer, p53, Geuvadis
Advanced
Home
Experiments
Arrays
Submit
Help
About ArrayExpress
Feedback
Login
ArrayExpress - functional genomics data
ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to MIAME and MINSEQE standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.
Data Content
Updated today at 06:00
49013 experiments
1406218 assays
23.47 TB of archived data
2つのDB間で用語の統一はなされていない…
Latest News
7 February 2014 - High throughput sequencing (HTS) data sets in ArrayExpress
Did you know that ArrayExpress contains nearly 6000 HTS experiments? They range from common study types such as RNA-seq and ChIP-seq to methylation profiling seq, FAIRE-seq (formaldehyde-assisted isolation of regulatory elements), GRO-seq (global run-on seq) and many more. For each experiment, study and sample information is stored at ArrayExpress, while raw sequence data files (e.g. fastq files) are stored at the Sequence Read Archive of the European Nucleotide Archive. If you're planning to submit a sequencing experiment to ArrayExpress, why not check out this help page for more information?

発現DB


http://www.ncbi.nlm.nih.gov/geo/
Home - GEO - NCBI
NCBI
Resources
How To
Sign in to NCBI
GEO Home
Documentation
Query & Browse
Email GEO
Gene Expression Omnibus
GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.
Gene Expression Omnibus
Keyword or GEO Accession
Search
Getting Started
Overview
FAQ
About GEO DataSets
About GEO Profiles
About GEO2R Analysis
How to Construct a Query
How to Download Data
Tools
Search for Studies at GEO DataSets
EO Profiles
Analyze a Study with GEO2R
GEO BLAST
Programmatic Access
FTP Site
Browse Content
Repository Browser
DataSets: 3413
Series: 47340
Platforms: 12912
Samples: 1131572
DataSets, Series, Platforms, Samplesの説明はここ



Platformsは、**大まかには**アレイの種類数。（今はシーケンサーも登録されている。例：GPL11154）

Platformの例

- Affymetrix GeneChip
  - Affymetrix Human Genome U133 Plus 2.0 Array: **GPL570**
    - 2003年11月リリース、54,675 probesets、94,000枚以上の利用実績
  - Affymetrix Human Genome U133A Array: **GPL96**
    - 2002年3月リリース、22,283 probesets、34,000枚以上
  - Affymetrix Mouse Genome 430 2.0 Array: **GPL1261**
    - 2004年5月リリース、45,101 probesets、38,000枚以上
  - Affymetrix Rat Genome 230 2.0 Array: **GPL1355**
    - 2004年6月リリース、31,099 probesets、10,000枚以上
- Illumina BeadChip
  - Illumina HumanHT-12 V4.0 expression beadchip: **GPL10558**
    - 2010年6月リリース、47,323 probes、18,000枚以上
  - Illumina HumanHT-12 V3.0 expression beadchip: **GPL6947**
    - 2008年6月リリース、49,576 probes、18,000枚以上
- Agilent Microarray
  - Agilent-014850 Whole Human Genome Microarray 4x44K G4112F: **GPL6480**
    - 2008年2月リリース、41,108 probes、11,000枚以上

同じメーカー、同じ生物種でも様々なバージョンのアレイが存在する

http://www.ncbi.nlm.nih.gov/geo/info/overview.html

GEO Overview - GEO - ...

NCBI

NCBI » GEO » Info » GEO Ov

GEO Overview

- General overview
- Data organization
- Query and analys

General overview

GEO is an internati
sequencing, and othe

The three main goals

1. Provide a robus
data (see Data
2. Offer simple su
deposits from t
3. Provide user-fr
gene expressio

Please see the GEO D

Data organizatio

GEO records are orga

Platform

Platform records are supplied by submitters

A Platform record is composed of a summary description of the array or sequencer and, for array-based Platforms, a data table defining the array template.Each Platform record is assigned a unique and stable GEO accession number (GPLxxx). A Platform may reference many Samples that have been submitted by multiple submitters.

Example Platform record »

Sample

Sample records are supplied by submitters

A Sample record describes the conditions under which an individual Sample was handled, the manipulations it underwent, and the abundance measurement of each element derived from it. Each Sample record is assigned a unique and stable GEO accession number (GSMxxx). A Sample entity must reference only one Platform and may be included in multiple Series.

Example Sample record »

Series

Series records are supplied by submitters

A Series record links together a group of related Samples and provides a focal point and description of the whole study. Series records may also contain tables describing extracted data, summary conclusions, or analyses. Each Series record is assigned a unique and stable GEO accession number (GSExxx).

Example Series record »

A	Text description of the array or sequencer
B	Text tab-delimited table of the array template
C	Text description of the biological sample and protocols to which it was subjected
D	Text tab-delimited table of processed hybridization result (may optionally include raw data columns)
E	Original raw data file, or processed sequence data file
F	Text description of the overall experiment
G	Tar archi raw data processe data files

Selected primary records undergo an upper-level of rendering into DataSet and gene Profile recor

DataSet records are assembled by GEO curators

As explained above, A GEO Series record is an original submitter-supplied record that summarizes an experiment. These data are reassembled by GEO staff into GEO Dataset records (GDSxxx).

A DataSet represents a curated collection of biologically and statistically comparable GEO Samples and forms the basis of GEO's

Browse Content

Repository Browser

DataSets:	3413
Series:	47340
Platforms:	12912
Samples:	1131572

Seriesは、一つの研究プロジェクトなどで用いた複数サンプルからなるグループをまとめたもの。大まかには論文ごとのIDという理解でよい。

Seriesの例

- Affymetrix GeneChip
  - Ge et al., *Genomics*, 86: 127–141, 2005
    - **GSE2361**、ヒト36サンプル、GPL96を利用
  - Nakai et al., *Biosci Biotechnol Biochem.*, 72: 139–148, 2008
    - **GSE7623**、ラット24サンプル、GPL1355を利用
  - Kamei et al., *PLoS One*, 8: e65732, 2013
    - **GSE30533**、ラット10サンプル、GPL1355を利用
- Illumina BeadChip
  - Sharma et al., *Cancer Cell*, 23: 35–47, 2013
    - **GSE28680**、ヒト24サンプル、GPL10558を利用
- NGSデータも…
  - Neyret-Kahn et al., *Genome Res.*, 23: 1563–1579, 2013
    - **GSE42213**、ヒト26サンプル、GPL10999とGPL11154を利用
      - **GSE42211**、ヒト20サンプル、GPL10999とGPL11154を利用（ChIP-seq）
      - **GSE42212**、ヒト6サンプル、GPL10999を利用（RNA-seq）
  - Huang et al., *Development*, 139: 2161–2169, 2012
    - **GSE36469**、シロイヌナズナ8サンプル、GPL13222を利用

・NGSデータも登録されている
・1論文1 GSE IDとは限らない
・1 GSE ID 1 GPL IDとは限らない

http://www.ncbi.nlm.nih.gov/geo/info/overview.html

GEO Overview - GEO - ...

NCBI

NCBI » GEO » Info » GEO Ov

GEO Overview

- General overview
- Data organization
- Query and analys

General overview

GEO is an internati
sequencing, and othe

The three main goals

1. Provide a robus
data (see Data
2. Offer simple su
deposits from t
3. Provide user-fr
gene expressio

Please see the GEO D

Data organizatio

GEO records are orga

http://www.ncbi.nlm.nih.gov/geo/info/overview.html

GEO Overview - GEO - ...

Platform

Platform records are supplied by submitters

A Platform record is composed of a summary description of the array or sequencer and, for array-based Platforms, a data table defining the array template.Each Platform record is assigned a unique and stable GEO accession number (GPLxxx). A Platform may reference many Samples that have been submitted by multiple submitters.

Example Platform record »

Sample

Sample records are supplied by submitters

A Sample record describes the conditions under which an individual Sample was handled, the manipulations it underwent, and the abundance measurement of each element derived from it. Each Sample record is assigned a unique and stable GEO accession number (GSMxxx). A Sample entity must reference only one Platform and may be included in multiple Series.

Example Sample record »

Series

Series records are supplied by submitters

A Series record links together a group of related Samples and provides a focal point and description of the whole study. Series records may also contain tables describing extracted data, summary conclusions, or analyses. Each Series record is assigned a unique and stable GEO accession number (GSExxx).

Example Series record »

A	**Text description of the array or sequencer**
B	**Text tab-delimited table of the array template**
C	**Text description of the biological sample and protocols to which it was subjected**
D	**Text tab-delimited table of processed hybridization result** (may optionally include raw data columns)
E	**Original raw data file, or processed sequence data file**
F	**Text description of the overall experiment**
G	**Tar archive of original raw data files, or processed sequence data files**

Selected primary records undergo an upper-level of rendering into DataSet and gene Profile records:

DataSet records are assembled by GEO curators

As explained above, A GEO Series record is an original submitter-supplied record that summarizes an experiment. These data are reassembled by GEO staff into GEO Dataset records (GDSxxx).

A DataSet represents a curated collection of biologically and statistically comparable GEO Samples and forms the basis of GEO's

Browse Content

Repository Browser

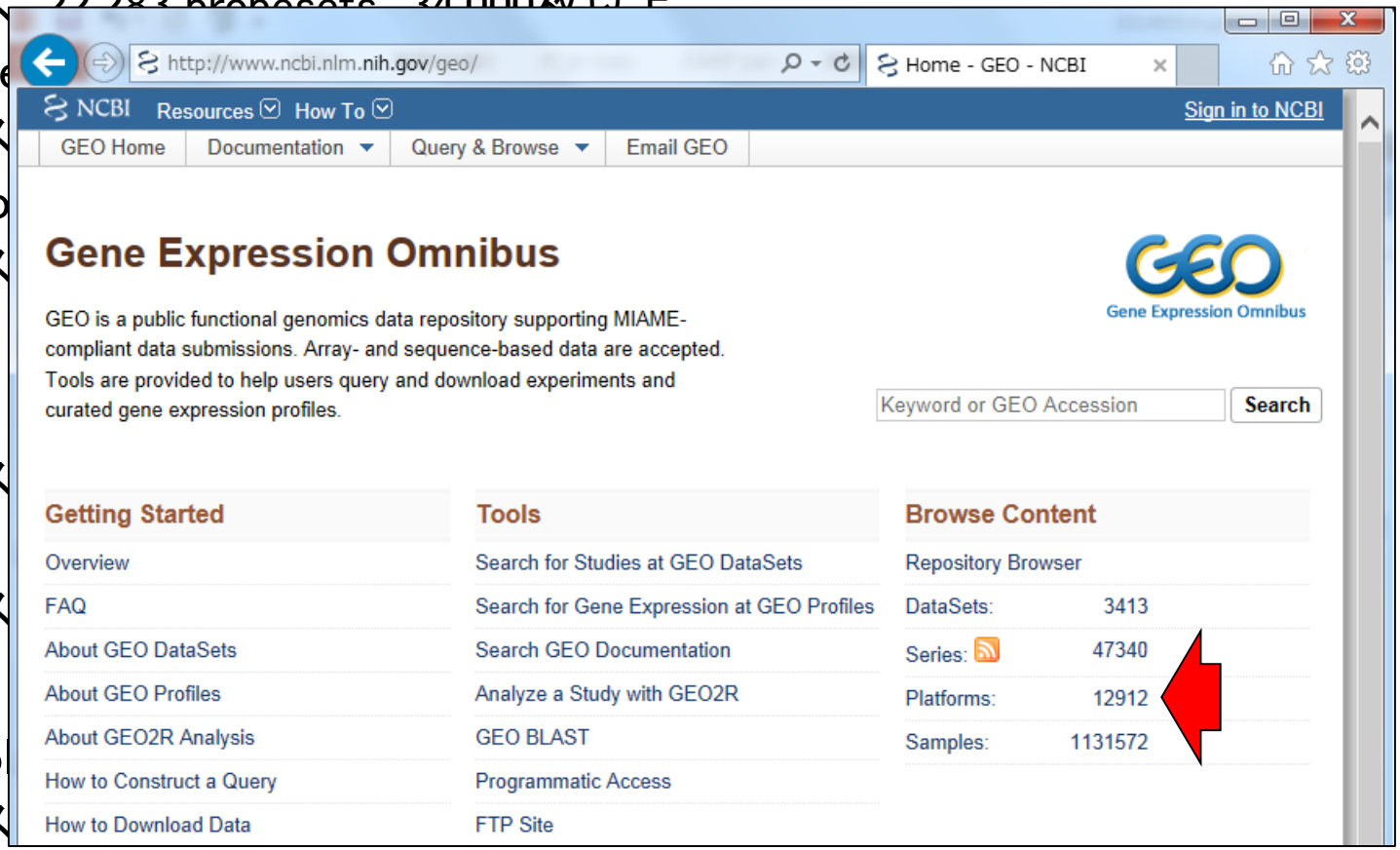
DataSets:	3413
Series:	47340
Platforms:	12912
Samples:	1131572

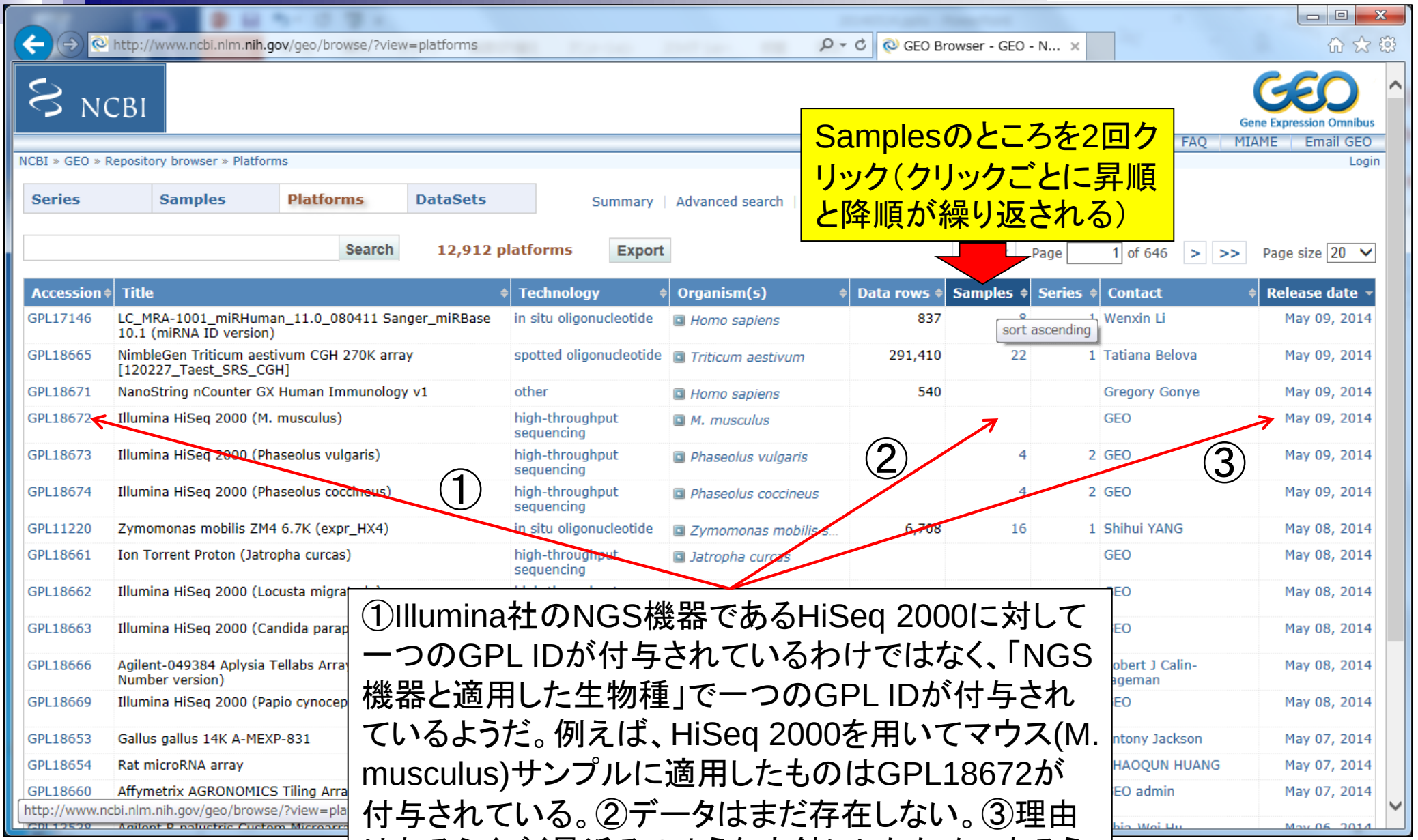
Samplesは、登録されているサンプル数

Platformの例

- Affymetrix GeneChip
  - Affymetrix Human Genome U133 Plus 2.0 Array: **GPL570**
    - 2003年11月リリース、54,675 probesets、94,000枚以上の利用実績
  - Affymetrix Human Genome U133A Array: **GPL96**
    - 2002年3月リリース、22,283 probesets、34,000枚以上
  - Affymetrix Mouse Ge
    - 2004年5月リリース
  - Affymetrix Rat Geno
    - 2004年6月リリース
- Illumina BeadChip
  - Illumina HumanHT-1
    - 2010年6月リリース
  - Illumina HumanHT-1
    - 2008年6月リリース
- Agilent Microarray
  - Agilent-014850 Who
    - 2008年2月リリース

どうやって調べたのか?




http://www.ncbi.nlm.nih.gov/geo/browse/?view=platforms
GEO Browser - GEO - N...
NCBI
Gene Expression Omnibus
FAQ | MIAME | Email GEO
NCBI » GEO » Repository browser » Platforms
Login
Series | Samples | Platforms | DataSets
Summary | Advanced search
Search
12,912 platforms
Export
Page 1 of 646 > >> Page size 20
Samplesのところを2回クリック（クリックごとに昇順と降順が繰り返される）
sort ascending
Accession | Title | Technology | Organism(s) | Data rows | Samples | Series | Contact | Release date
GPL17146 | LC_MRA-1001_miRHuman_11.0_080411 Sanger_miRBase 10.1 (miRNA ID version) | in situ oligonucleotide | Homo sapiens | 837 | | | Wenxin Li | May 09, 2014
GPL18665 | NimbleGen Triticum aestivum CGH 270K array [120227_Taest_SRS_CGH] | spotted oligonucleotide | Triticum aestivum | 291,410 | 22 | 1 | Tatiana Belova | May 09, 2014
GPL18671 | NanoString nCounter GX Human Immunology v1 | other | Homo sapiens | 540 | | | Gregory Gonye | May 09, 2014
GPL18672 | Illumina HiSeq 2000 (M. musculus) | high-throughput sequencing | M. musculus | | | | GEO | May 09, 2014
GPL18673 | Illumina HiSeq 2000 (Phaseolus vulgaris) | high-throughput sequencing | Phaseolus vulgaris | | 4 | 2 | GEO | May 09, 2014
GPL18674 | Illumina HiSeq 2000 (Phaseolus coccineus) | high-throughput sequencing | Phaseolus coccineus | | 4 | 2 | GEO | May 09, 2014
GPL11220 | Zymomonas mobilis ZM4 6.7K (expr_HX4) | in situ oligonucleotide | Zymomonas mobilis s... | 6,708 | 16 | 1 | Shihui YANG | May 08, 2014
GPL18661 | Ion Torrent Proton (Jatropha curcas) | high-throughput sequencing | Jatropha curcas | | | | GEO | May 08, 2014
GPL18662 | Illumina HiSeq 2000 (Locusta migra | | | | | | EO | May 08, 2014
GPL18663 | Illumina HiSeq 2000 (Candida parap | | | | | | EO | May 08, 2014
GPL18666 | Agilent-049384 Aplysia Tellabs Arra Number version) | | | | | | obert J Calin-ageman | May 08, 2014
GPL18669 | Illumina HiSeq 2000 (Papio cynocep | | | | | | EO | May 08, 2014
GPL18653 | Gallus gallus 14K A-MEXP-831 | | | | | | ntony Jackson | May 07, 2014
GPL18654 | Rat microRNA array | | | | | | HAOQUN HUANG | May 07, 2014
GPL18660 | Affymetrix AGRONOMICS Tiling Arra | | | | | | EO admin | May 07, 2014
①
②
③
①Illumina社のNGS機器であるHiSeq 2000に対して一つのGPL IDが付与されているわけではなく、「NGS機器と適用した生物種」で一つのGPL IDが付与されているようだ。例えば、HiSeq 2000を用いてマウス(M. musculus)サンプルに適用したものはGPL18672が付与されている。②データはまだ存在しない。③理由はおそらくごく最近そのような方針にしたためであろう

http://www.ncbi.nlm.nih.gov/geo/browse/?view=platforms&display=20&zsort=samples

GEO Browser - GEO - N...

NCBI

Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » Repository browser » Platforms

Login

Series | Samples | Platforms | DataSets

Summary | Advanced search | Find platform

Search

12,912 platforms

Export

<< < Page 1 of 646 > >> Page size 20

Accession	Title	Technology	Organism(s)	Data rows	Samples	Series	Contact	Release date
GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	in situ oligonucleotide	*Homo sapiens*	54,675	94089	3451	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	*Mus musculus*	45,101	38555	2955	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	*Homo sapiens*	22,283	34783	987	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	*Homo sapiens*	47,323	18387	609	Illumina Inc.	Jun 17, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	*Homo sapiens*	49,576	18312	388	Illumina Inc.	Jun 10, 2008
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	*Homo sapiens*	33,297	17742	878	Affymetrix, Inc.	Dec 05, 2007
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	*Homo sapiens*	27,578	14800	263	Illumina Inc.	Apr 27, 2009
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	*Mus musculus*	35,557	13817	1134	Affymetrix, Inc.	Dec 05, 2007
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	*Homo sapiens*	485,577	12217	199	Illumina Inc.	May 13, 2011
GPL6801	[GenomeWideSNP_6] Affymetrix Genome-Wide Human SNP 6.0 Array	in situ oligonucleotide	*Homo sapiens*	1,880,794	12077	207	Affymetrix, Inc.	Apr 30, 2008
GPL6480	Agilent-014850 Whole Human Genome M G4112F (Probe Name version)			08	11621	509	Agilent Technologies	Feb 11, 2008
GPL571	[HG-U133A_2] Affymetrix Human Genom			77	11036	445	Affymetrix, Inc.	Nov 07, 2003
GPL198	[ATH1-121501] Affymetrix Arabidopsis A			10	10682	820	Affymetrix, Inc.	Jul 18, 2002
GPL1355	[Rat230_2] Affymetrix Rat Genome 230			99	10482	493	Affymetrix, Inc.	Jul 20, 2004
GPL4133	Agilent-014850 Whole Human Genome M G4112F (Feature Number version)			20	10183	528	Agilent Technologies	Aug 17, 2006
GPL3718	[Mapping250K_Nsp] Affymetrix Mapping Array			38	9823	156	Affymetrix, Inc.	May 13, 2006

（NGS機器も含まれるため、もはや正確な言い回しではないが…）赤枠部分がアレイごとに用いられたサンプル数。トップ3はいずれもAffymetrix **3’発現アレイ**と呼ばれるもの。

様々なDNAマイクロアレイ（DNAチップ）

1. スポット型（Stanford大学）
   - 搭載DNA：cDNA（またはoligonucleotide）
   - 解析法：2色法（比較したい2サンプルを同時に分析）
2. プリント型（Agilent社）
   - 搭載DNA：oligonucleotide（60mer）
   - 解析法： 2色法または1色法
3. **合成オリゴ型（Affymetrix社）**
   - **搭載DNA：oligonucleotide（25mer）**
   - **解析法： 1色法（調べたい1サンプルを分析）**

Stanford型

Affymetrix型

得られる遺伝子発現データのイメージ

- 二色法の場合

	目的試料	対照試料	目的/対照	log(比)
遺伝子1	100	100	1	0
遺伝子2	4000	1000	4	2
遺伝子3	7000	7000	1	0
遺伝子4	2000	8000	0.25	−2
…	…	…	…	…

目的試料中の遺伝子2の発現レベルは対照試料に比べて4倍高い

- 一色法の場合

	目的試料
遺伝子1	100
遺伝子2	4000
遺伝子3	7000
遺伝子4	2000
…	…

目的試料中で遺伝子3は沢山発現している

Affymetrix型マイクロアレイ（GeneChip®）

- 目的試料の発現情報を直接「**シグナル強度**」として得る

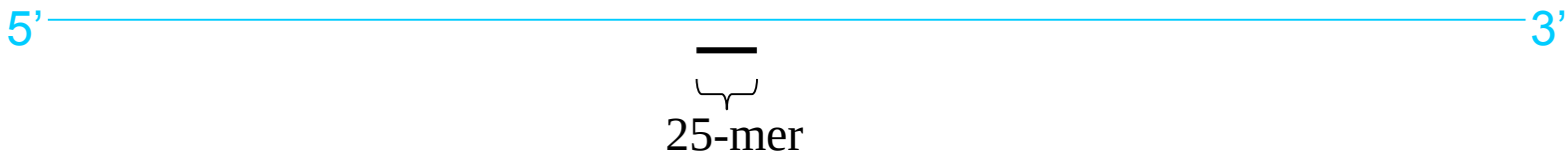
Affymetrix GeneChipデータ解析

- 25-mer程度では
  - 本当に目的遺伝子の発現を調べられているのか？！

ヒト→ 3Gbp(=3×10^9 bp) < 4^25 (=1×10^15 bp)

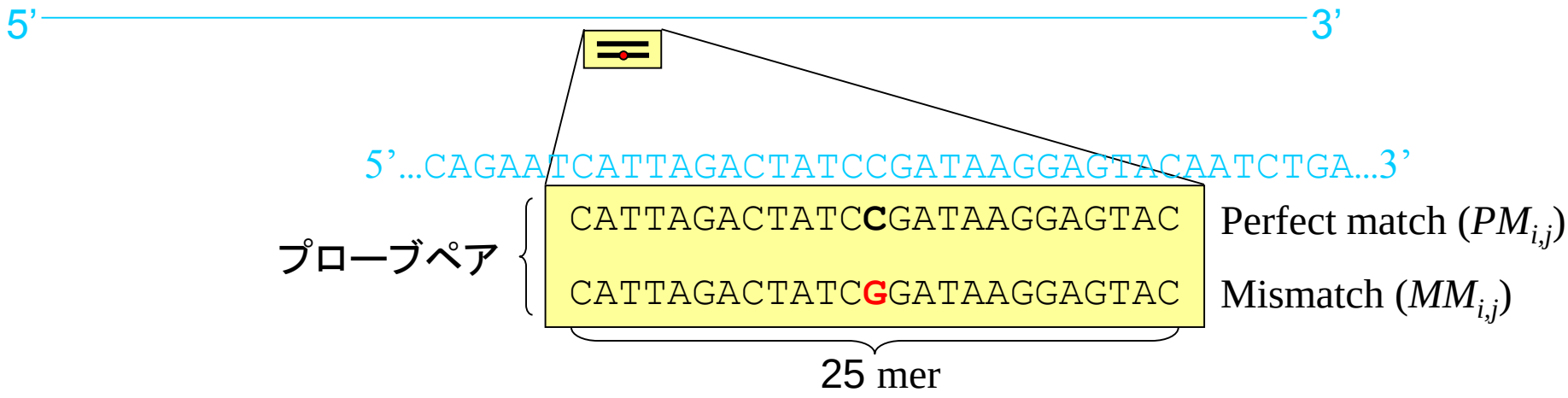
理論上は25merで充分…

  - 発現量を正確に定量できるのか？



Affymetrix GeneChipデータ解析

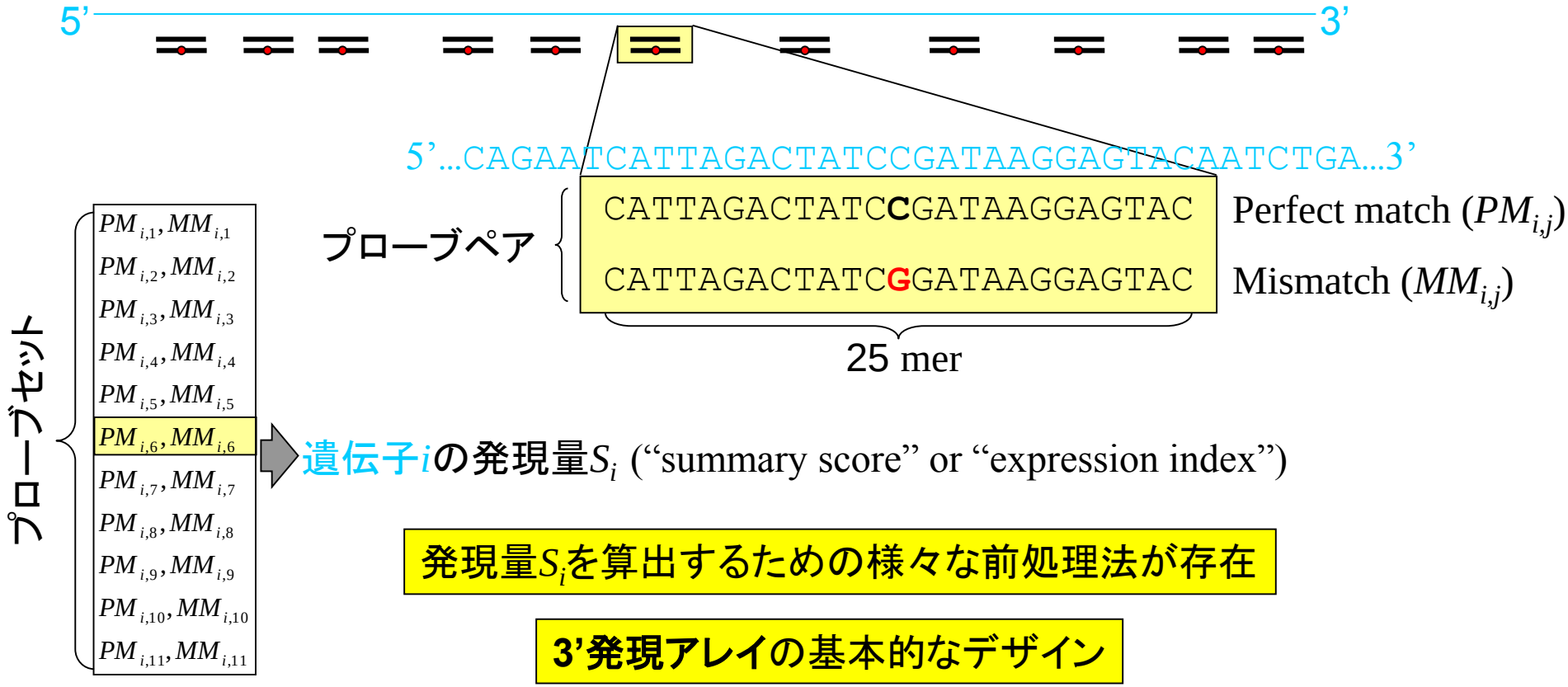
- 遺伝子iの発現量S_iを正確に知るために
  - PM/MMプローブ戦略（ユニークな配列選択と最適T_m）



特異的なハイブリダイゼーションと非特異的なハイブリダイゼーションを区別すべく、目的遺伝子配列に対してPMと一塩基MMがペアになっているのが特徴的

Affymetrix GeneChipデータ解析

- 遺伝子iの発現量S_iをn_i(n_i=11~20)種類のプローブペアのシグナル強度をもとに計算

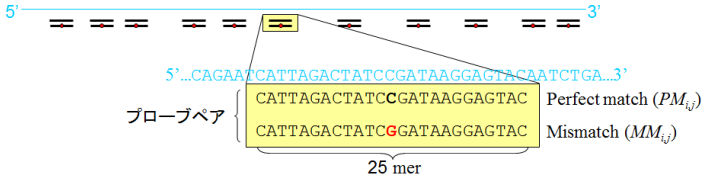


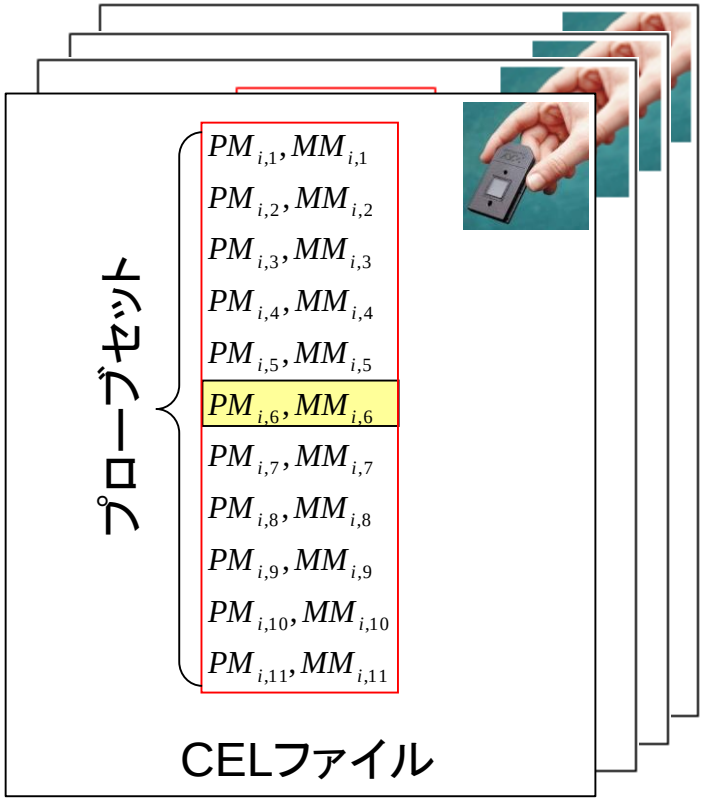
発現量S_iを算出するための様々な前処理法が存在

3’発現アレイの基本的なデザイン

全体的なイメージ

－ CELファイル －
チップ上に搭載されている全遺伝子のプローブのシグナル強度情報を含むファイル





前処理法

遺伝子発現行列

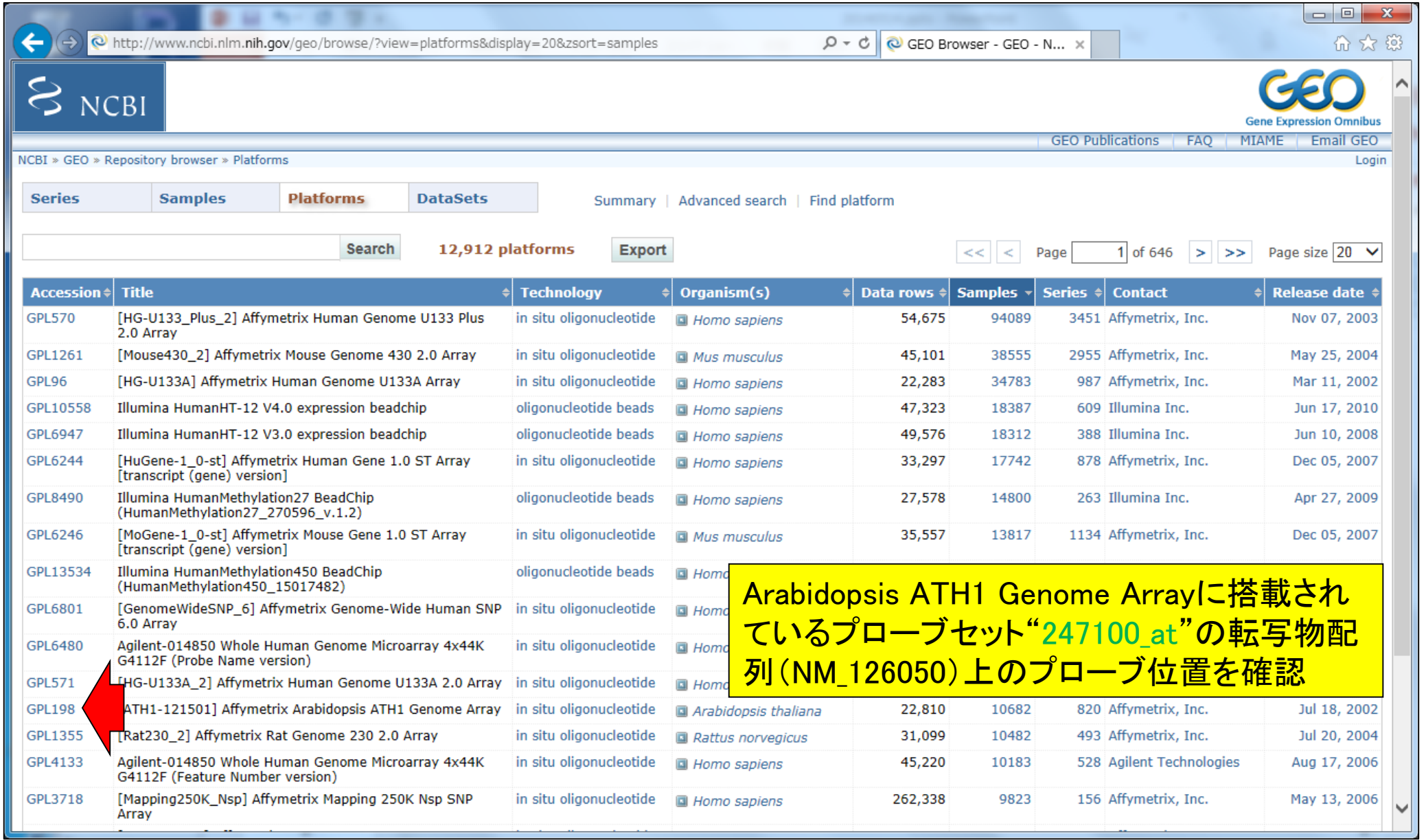
	S1	S2	S3	S4	…
gene *1*	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	…
gene *2*	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	…
…	…	…	…	…	…
gene *i*	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	
…	…	…	…		
gene *n*	$x_{n,1}$	$x_{n,2}$			

hoge_q3.txt

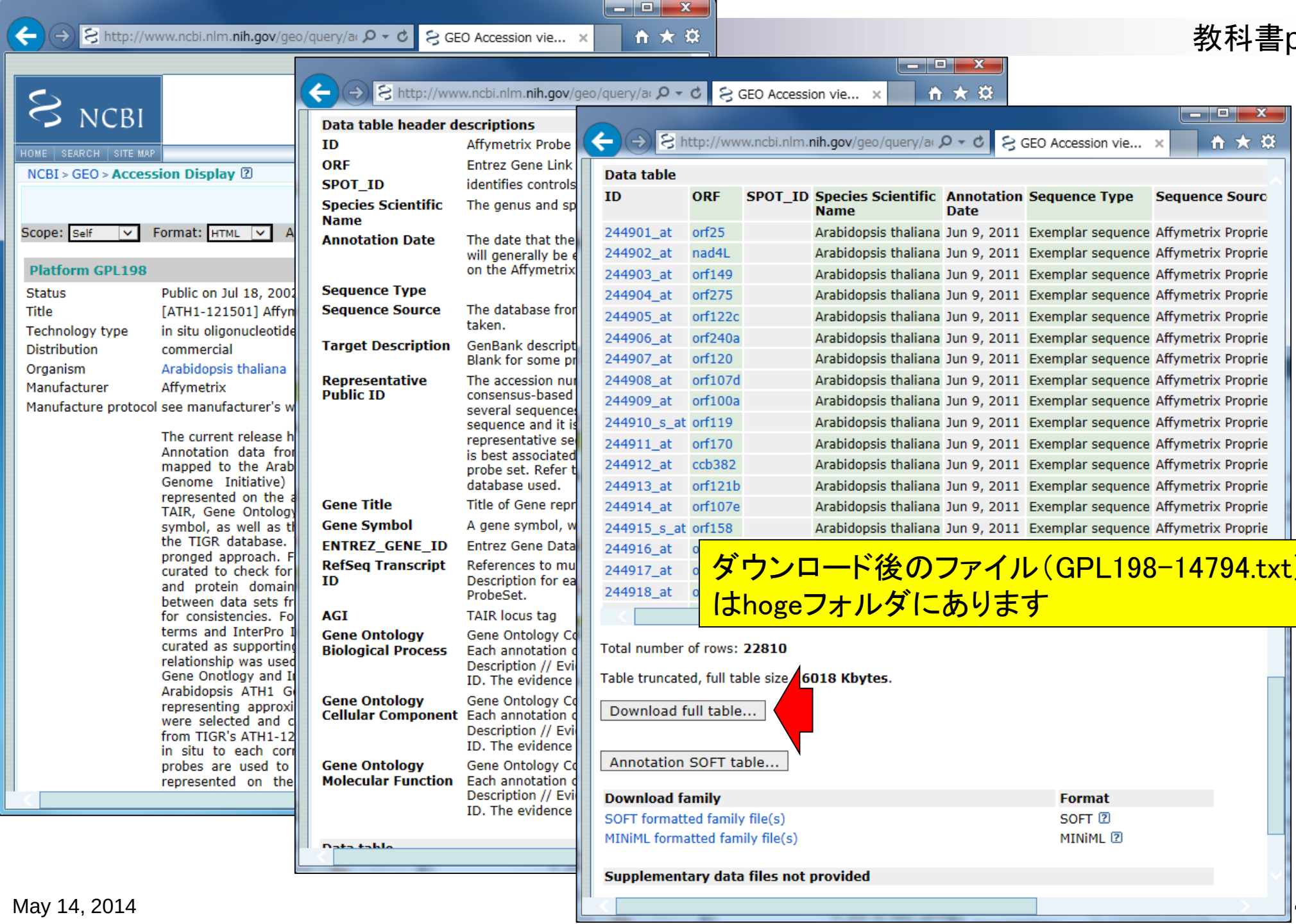
	A	B	C	D
1		sample1	sample2	sample3
2	gene1	10.87	12.13	8.87
3	gene2	7.03	7.17	12.13
4	gene3	8.50	8.50	10.87
5	gene4	12.13	10.87	7.17
6	gene5	5.77	6.30	5.77
7	gene6	8.87	8.87	8.50
8	gene7	6.77	7.03	6.30
9	gene8	6.30	6.77	6.77
10	gene9	7.17	5.77	7.03
11	gene10	4.80	4.80	4.80

適用した前処理法の数だけ**遺伝子発現行列**が存在

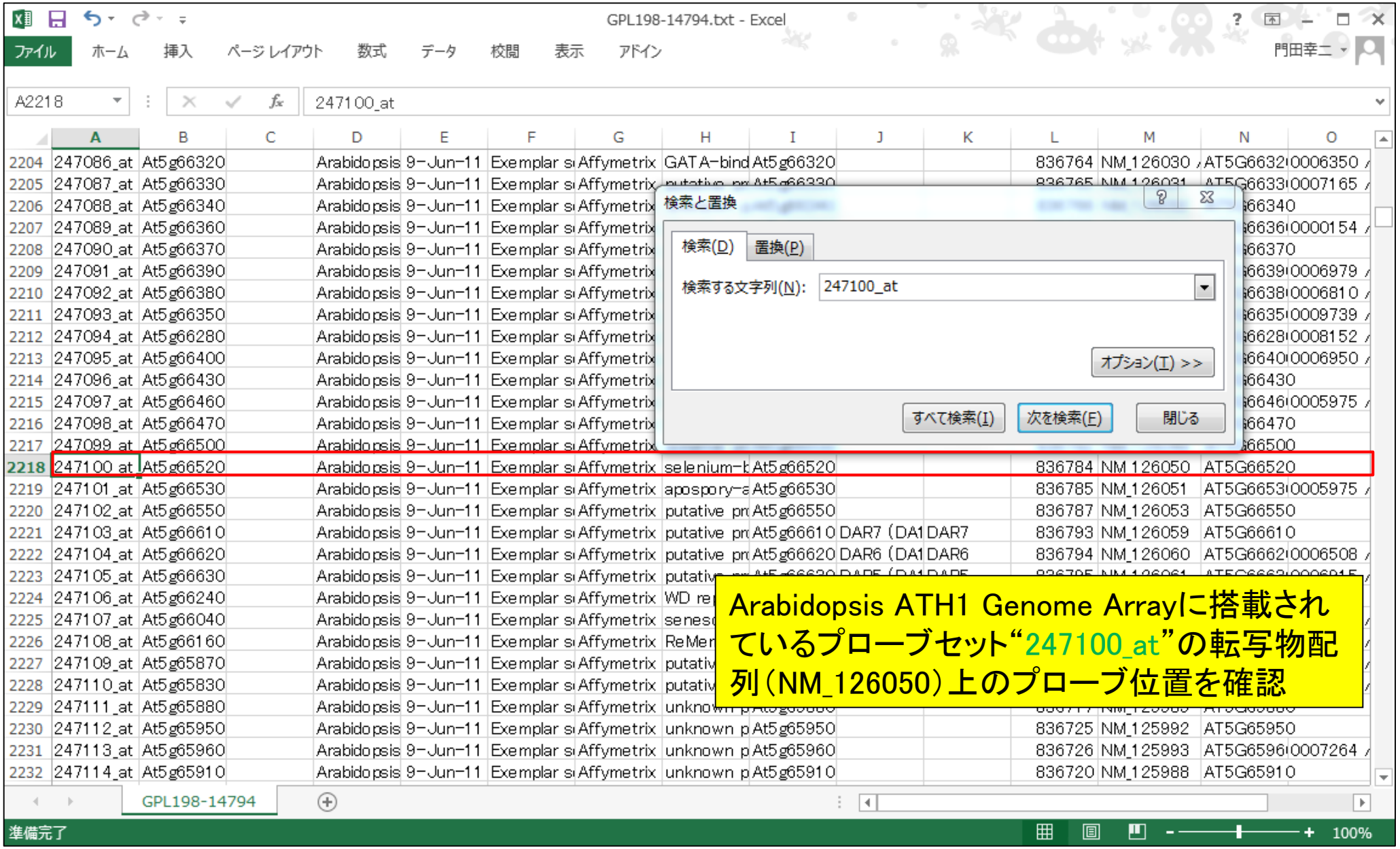
3’発現アレイの意味を確認



Accession	Title	Technology	Organism(s)	Data rows	Samples	Series	Contact	Release date
GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	in situ oligonucleotide	*Homo sapiens*	54,675	94089	3451	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	*Mus musculus*	45,101	38555	2955	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	*Homo sapiens*	22,283	34783	987	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	*Homo sapiens*	47,323	18387	609	Illumina Inc.	Jun 17, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	*Homo sapiens*	49,576	18312	388	Illumina Inc.	Jun 10, 2008
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	*Homo sapiens*	33,297	17742	878	Affymetrix, Inc.	Dec 05, 2007
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	*Homo sapiens*	27,578	14800	263	Illumina Inc.	Apr 27, 2009
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	*Mus musculus*	35,557	13817	1134	Affymetrix, Inc.	Dec 05, 2007
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	*Homo*					
GPL6801	[GenomeWideSNP_6] Affymetrix Genome-Wide Human SNP 6.0 Array	in situ oligonucleotide	*Homo*					
GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	in situ oligonucleotide	*Homo*					
GPL571	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	in situ oligonucleotide	*Homo*					
GPL198	[ATH1-121501] Affymetrix Arabidopsis ATH1 Genome Array	in situ oligonucleotide	*Arabidopsis thaliana*	22,810	10682	820	Affymetrix, Inc.	Jul 18, 2002
GPL1355	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	*Rattus norvegicus*	31,099	10482	493	Affymetrix, Inc.	Jul 20, 2004
GPL4133	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version)	in situ oligonucleotide	*Homo sapiens*	45,220	10183	528	Agilent Technologies	Aug 17, 2006
GPL3718	[Mapping250K_Nsp] Affymetrix Mapping 250K Nsp SNP Array	in situ oligonucleotide	*Homo sapiens*	262,338	9823	156	Affymetrix, Inc.	May 13, 2006


NCBI
HOME SEARCH SITE MAP
NCBI > GEO > Accession Display
Scope: Self Format: HTML
Platform GPL198
Status Public on Jul 18, 200
Title [ATH1-121501] Affyn
Technology type in situ oligonucleotide
Distribution commercial
Organism Arabidopsis thaliana
Manufacturer Affymetrix
Manufacture protocol see manufacturer's w
Data table header descriptions
ID Affymetrix Probe
ORF Entrez Gene Link
SPOT_ID identifies controls
Species Scientific Name The genus and sp
Annotation Date The date that the
Sequence Type
Sequence Source The database fror
Target Description GenBank descript
Representative Public ID The accession nu
Gene Title Title of Gene repr
Gene Symbol A gene symbol, w
ENTREZ_GENE_ID Entrez Gene Data
RefSeq Transcript ID References to mu
AGI TAIR locus tag
Gene Ontology Biological Process Gene Ontology Co
Gene Ontology Cellular Component Gene Ontology Co
Gene Ontology Molecular Function Gene Ontology Co
Data table
ID ORF SPOT_ID Species Scientific Name Annotation Date Sequence Type Sequence Sourc
244901_at orf25 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244902_at nad4L Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244903_at orf149 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244904_at orf275 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244905_at orf122c Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244906_at orf240a Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244907_at orf120 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244908_at orf107d Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244909_at orf100a Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244910_s_at orf119 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244911_at orf170 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244912_at ccb382 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244913_at orf121b Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244914_at orf107e Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
244915_s_at orf158 Arabidopsis thaliana Jun 9, 2011 Exemplar sequence Affymetrix Proprie
ダウンロード後のファイル（GPL198-14794.txt）はhogeフォルダにあります
Total number of rows: 22810
Table truncated, full table size 6018 Kbytes.
Download full table...
Annotation SOFT table...
Download family
Format
SOFT formatted family file(s)
SOFT
MINiML formatted family file(s)
MINiML
Supplementary data files not provided

3’発現アレイの意味を確認



	A	B	D	E	F	G	H	I	J	K	L	M	N	O
2204	247086_at	At5g66320	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	GATA-bind	At5g66320			836764	NM_126030	AT5G6632	0006350
2205	247087_at	At5g66330	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putative pr	At5g66330			836765	NM_126031	AT5G6633	0007165
2206	247088_at	At5g66340	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							66340	
2207	247089_at	At5g66360	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6636	0000154
2208	247090_at	At5g66370	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							66370	
2209	247091_at	At5g66390	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6639	0006979
2210	247092_at	At5g66380	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6638	0006810
2211	247093_at	At5g66350	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6635	0009739
2212	247094_at	At5g66280	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6628	0008152
2213	247095_at	At5g66400	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6640	0006950
2214	247096_at	At5g66430	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							66430	
2215	247097_at	At5g66460	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							6646	0005975
2216	247098_at	At5g66470	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							66470	
2217	247099_at	At5g66500	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix							66500	
2218	247100_at	At5g66520	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	selenium-b	At5g66520			836784	NM_126050	AT5G66520	
2219	247101_at	At5g66530	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	apospory-a	At5g66530			836785	NM_126051	AT5G6653	0005975
2220	247102_at	At5g66550	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putative pr	At5g66550			836787	NM_126053	AT5G66550	
2221	247103_at	At5g66610	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putative pr	At5g66610	DAR7 (DA1	DAR7	836793	NM_126059	AT5G66610	
2222	247104_at	At5g66620	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putative pr	At5g66620	DAR6 (DA1	DAR6	836794	NM_126060	AT5G6662	0006508
2223	247105_at	At5g66630	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putativ							
2224	247106_at	At5g66240	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	WD re							
2225	247107_at	At5g66040	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	senes							
2226	247108_at	At5g66160	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	ReMer							
2227	247109_at	At5g65870	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putativ							
2228	247110_at	At5g65830	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	putativ							
2229	247111_at	At5g65880	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	unknown p							
2230	247112_at	At5g65950	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	unknown p	At5g65950			836725	NM_125992	AT5G65950	
2231	247113_at	At5g65960	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	unknown p	At5g65960			836726	NM_125993	AT5G6596	0007264
2232	247114_at	At5g65910	Arabidopsis	9-Jun-11	Exemplar s	Affymetrix	unknown p	At5g65910			836720	NM_125988	AT5G65910	

Arabidopsis ATH1 Genome Arrayに搭載されているプローブセット“247100_at”の転写物配列（NM_126050）上のプローブ位置を確認

教科書p5

(Rで)マイクロアレイデータ解析

(last modified 2014/05/12, since 2005)

What's new?

- 『田幸二 著シリーズ Useful R 第7巻トランスクリプトーム解析刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法 (Kadota et al., 2006)、WAD法 (Kadota et al., 2008)などについての解説も含んでいます。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍 | トランスクリプトーム解析 | ...」に掲載してあります。(2014/04/27) NEW
- お知らせは主に(Rで)塩基配列解析で行っておりますのでそちらをご覧ください。(2014/03/05)

- はじめに(last modified 2014/01/21)
- 過去のお知らせ(last modified 2014/03/03)
- Rのインストールと起動(last modified 2014/04/17) NEW
- Rの昔のバージョンのインストール(last modified 2012/04/07)
- 使用例(初心者向け)(last modified 2011/09/15)
- サンプルデータ (last modified 2013/11/25)
- 書籍 |について (last modified 2014/04/17) NEW
- 書籍 |トランスクリプトーム解析 | 1.1 はじめに (last modified 2014/05/09
- 書籍 |トランスクリプトーム解析 | 1.2.1 原理(Affymetrix 3'発現アレイ) (
- 書籍 |トランスクリプトーム解析 | 1.2.2 最近の知見 (last modified 2014
- 書籍 |トランスクリプトーム解析 | 2.2.1 生データ(プローブレベルデータ
- 書籍 |トランスクリプトーム解析 | 2.2.2 データの正規化(基礎) (last mod
- 書籍 |トランスクリプトーム解析 | 2.2.3 データの正規化(計算例) (last m
- 書籍 |トランスクリプトーム解析 | 2.2.4 データの正規化(その他) (last m
- 書籍 |トランスクリプトーム解析 | 2.2.5 アノテーション情報 (last modifie
- 書籍 |トランスクリプトーム解析 | 3.2.1 クラスタリング(データ変換や距離
- 書籍 |トランスクリプトーム解析 | 3.2.2 実験デザイン、データ分布、統計
- 書籍 |トランスクリプトーム解析 | 3.2.3 多重比較問題 (last modified 20
- 書籍 |トランスクリプトーム解析 | 3.2.4 各種プロット(M-A plotや平均-分
- 書籍 |トランスクリプトーム解析 | 4.2.1 2群間比較 (last modified 2014/0
- 書籍 |トランスクリプトーム解析 | 4.2.2 他の実験デザイン(paired, multi-
- 書籍 |トランスクリプトーム解析 | 4.2.3 多群間比較(特異的発現パター
- イントロ | 発現データ取得 | 公共DBから (last modified 2014/05/11) NE
- イントロ | 発現データ取得 | inSilicoDb(Taminau_2011) (last modified 20
- イントロ | 発現データ取得 | ArrayExpress(Kauffmann_2009) (last modif

Arabidopsis ATH1 Genome Arrayに搭載されているプローブセット“247100_at”の転写物配列(NM_126050)上のプローブ位置を確認

書籍 | トランスクリプトーム解析 | 1.2.1 原理(Affymetrix 3'発現アレイ) NEW

シリーズ Useful R 第7巻トランスクリプトーム解析のp3-5です。

p3:

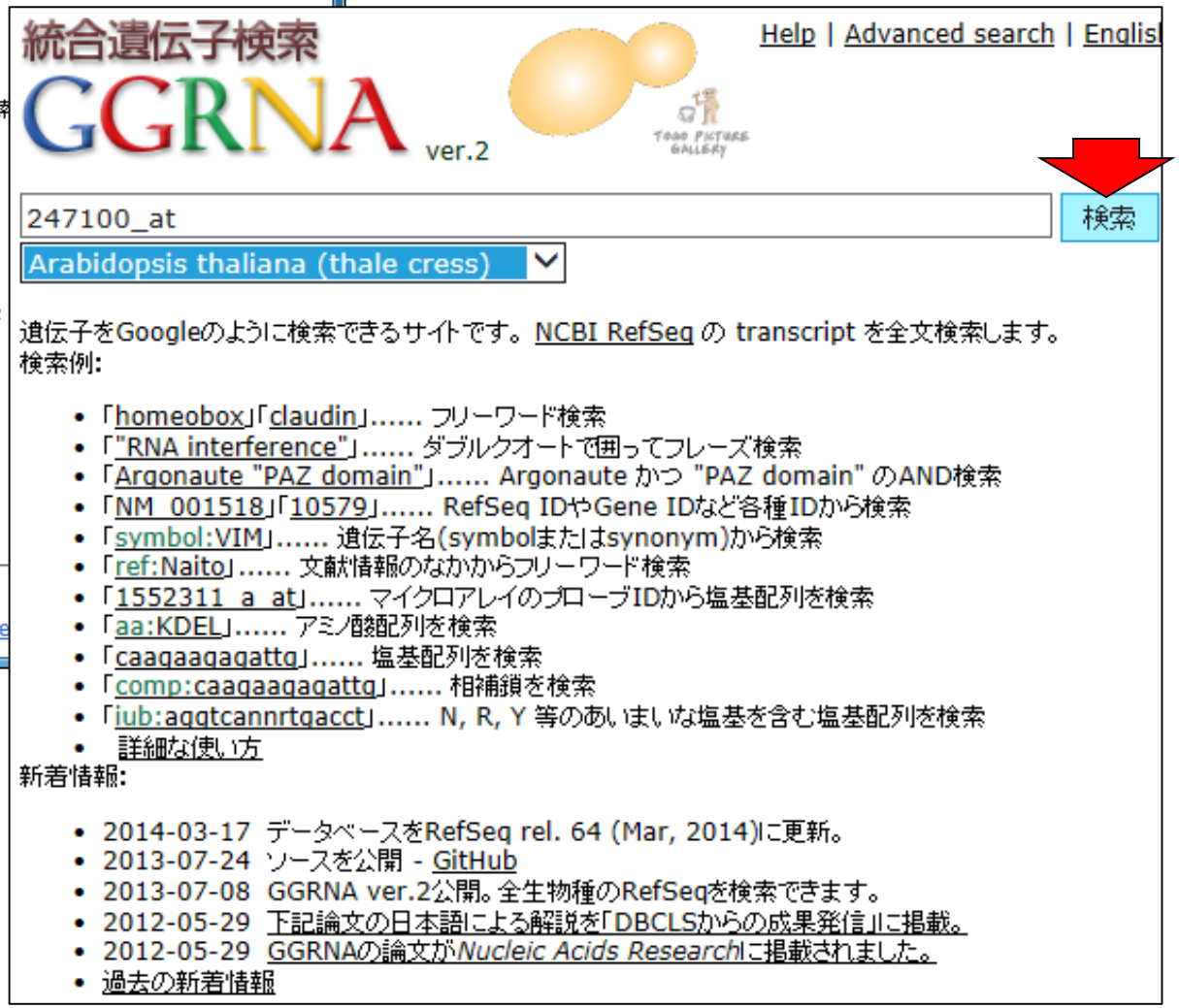
- GEO: Barrett et al., Nucleic Acids Res., 2013
- ヒト用: GPL570(Affymetrix Human Genome U133 Plus 2.0 Array)
- マウス用: GPL1261(Affymetrix Mouse Genome 430 2.0 Array)
- ラット用: GPL1355(Affymetrix Rat Genome 230 2.0 Array)

p5:

- アノテーションファイル(GPL570-13270.txt)は、GPL570の「Download full table...」ボタンを押すことでダウンロード可能。
- GGRNA: Naito and Bono, Nucleic Acids Res., 2012
- Probe Search
- GGRNAを用いたHumo sapiens(human)中の1552263_atの検索結果

トップページへ

Arabidopsis ATH1 Genome Arrayに搭載されているプローブセット“247100_at”の転写物配列（NM_126050）上のプローブ位置を確認



http://ggrna.dbcls.jp/ja/at/247100_at　GGRNA | Results

統合遺伝子検索
GGRNA ver.2

Help | Advanced search | English

247100_at　検索

Arabidopsis thaliana (thale cress)

2014-05-12 15:35:19, GGRNA : RefSeq release 64 (Mar, 2014)

Summary:

- seq:TTGCTCAAAGCCTGTCGAATTCACA (1)
- seq:TTAGCGGGAGACCGATCACACCCAG (1)
- seq:GGGACCACACACGAGTTTTTAGCGG (1)
- seq:GGGAAATGCAGTTGTGGGGACTACT (1)
- seq:GATGCTGCTTGATCTAGTTGATGAT (1)
- seq:GAGAGGCTATTGTGCATCAGCATAG (1)
- seq:GACTAAACCAGGAACCATTATTCGT (1)
- seq:GACCGATCACACCCAGAGATAGAGA (1)
- seq:GAAATTGGCTATTACATACGGGTTA (1)
- seq:AGATAGGACAAGGTTCCATCATTTC (1)
- seq:AAGTAACGAAGCTCATCTCTAAGAT (1)
- **INTERSECTION (1)**

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

Arabidopsis thaliana pentatricopeptide repeat-containing protein mRNA, complete cds. (1863 bp)

gcgttgctcaaagcctgtcgaattcacaaaaacatcgaattgggagaggaaattggagaaatcttaattgcaatagatccata
tcatggcggaagatatgttcataaggcaaatattcatgctatggataaaaagtgggacaaagcagctgaaacaagaagattgatg
aaagaacaaggagtagcaaaagttccaggatgtagtacaattagcttggaagggaccacacacgagtttttagcgggaga
ccgatcacacccagagatagagaaaattcaatctaaatggagaatcatgagaaggaaacttgaggaaaacgggtacgtacc
agagttagaagagatgctgcttgatctagttgatgatgatgaaagagaggctattgtgcatcagcatagcgagaaattg
gctattacatacgggttaatcaagactaaaccaggaaccattattcgtataatgaagaatctccgagtatgcaaagattgtc
acaaagtaacgaagctcatctctaagatatacaagagggatattgtaatgagagataggacaaggttccatcatttcaga
gatgggaaatgcagttgtggggactactgg...

position 1264 1480 1498 1507 1602 1634 1662 1692 1754 1803 1834

Synonym: K1F13.18; K1F13_18

NM_126050.1 - Arabidopsis thaliana (thale cress) - NCBI - TAIR

Data Export:

下記より最大10000件まで検索結果を取得できます。

- タブ区切りテキスト → 表示 | ダウンロード
  エクセル等の表計算ソフトに直接コピペできます。
- JSON形式 → リンク | ダウンロード

5'側

atgaatgtgatctcatgctccttctcattggaacataatctctacgagacaatgtcttgtc
tccagagatgctcaaagcaagaagaactgaagcaaatccacgctcgcatgctgaaaactgg
cttgatgcaggattcttatgcaatcacaaagtttctttctttctgcatttcttcaacgtct
tccgacttttgccttatgcccagattgtgtttgacgggtttgatcgaccagatactttct
tgtggaacctaatgatcagagggttctcgtgctcagacgaacccgagaggtctcttctcct
gtatcaacgtatgctctgttcttcagctcctcataacgcgtatacttttccgtctcttctc
aaagcttgttcgaacctatctgcatttgaagaaacaacgcaaattcacgcacagatcacga
aacttggatatgaaaatgatgtctatgcagtgaattctctgattaattcatatgctgtgac
cgggaatttcaagctagctcaccttctctttgacagaatccccgaacctgatgatgtctcg
tggaactctgtgatcaaaggttatgtaaaagctggaaaaatggatattgcattaacgttat
tcaggaaaatggcagaaaagaatgctatatcatggactacgatgatctctgggtatgttca
agcagacatgaacaaggaagctctgcaattgtttcatgaaatgcagaattcagatgttgag
cctgataatgtttccctagctaatgctctctcagcttgtgctcagctcggagcactcgagc
aaggga
ttgtgt
aagaat
atggccatggaagagaagctattagcaaattcatggagatgcaaaagatgggaattaagcc
aaatgtgatcactttcactgcggttctcacggcttgcagctacacaggactagttgaagaa
ggaaagttgatattctacagcatggagagagattacaacctgaaaccgaccatcgagcatt
atggctgtattgttgatttactcggtcgagctggattgcttgatgaagcaaaacgtttcat
tcaggagatgccattgaagccaaatgctgtgatatggggtgcgttgctcaaagcctgtcga
attcacaaaaacatcgaattgggagaggaaattggagaaatcttaattgcaatagatccat
atcatggcggaagatatgttcataaggcaaatattcatgctatggataaaaagtgggacaa
agcagctgaaacaagaagattgatgaaagaacaaggagtagcaaaagttccaggatgtagt
acaattagcttggaagggaccacacacgagtttttagcgggagaccgatcacacccagaga
tagagaaaattcaatctaaatggagaatcatgagaaggaaacttgaggaaaacgggtacgt
accagagttagaagagatgctgcttgatctagttgatgatgatgaaagagaggctattgtg
catcagcatagcgagaaattggctattacatacgggttaatcaagactaaaccaggaacca
ttattcgtataatgaagaatctccgagtatgcaaagattgtcacaaagtaacgaagctcat
ctctaagatatacaagagggatattgtaatgagagataggacaaggttccatcatttcaga
gatgggaaatgcagttgtggggactactggtaa

3'発現アレイの意味がよく分かります

3'側

マイクロアレイ（デバイスの進歩）

- **3’発現アレイ** → exon array → transcriptome array
  - Affymetrix Human Transcriptome Array (HTA 2.0)
  - Furney et al., *Cancer Discov*., **3**: 1122-1129, 2013.
  - GPL17585(exon level)
  - GPL17586(gene level)

転写物数は有限であるため、RNA-seqによる網羅的な同定後は**トランスクリプトームアレイ**に移行するほうがお手軽かもしれない

プローブ配列をテキストファイルで取り扱うこともできます。ここで示されているのはPerfect Match (PM)プローブ配列のみ

http://ggrna.dbcls.jp/ja/at/247100_at GGRNA | Results

統合遺伝子検索
GGRNA ver.2

Help | Advanced search | English

247100_at 検索

Arabidopsis thaliana (thale cress)

2014-05-12 15:35:19, GGRNA : RefSeq release 64 (Mar, 2014)

Summary:

- seq:TTGCTCAAAGCCTGTCGAATTCACA (1)
- seq:TTAGCGGGAGACCGATCACACCCAG (1)
- seq:GGGACCACACACGAGTTTTTAGCGG (1)
- seq:GGGAAATGCAGTTGTGGGGACTACT (1)
- seq:GATGCTGCTTGATCTAGTTGATGAT (1)
- seq:GAGAGGCTATTGTGCATCAGCATAG (1)
- seq:GACTAAACCAGGAACCATTATTCGT (1)
- seq:GACCGATCACACCCAGAGATAGAGA (1)
- seq:GAAATTGGCTATTACATACGGGTTA (1)
- seq:AGATAGGACAAGGTTCCATCATTTC (1)
- seq:AAGTAACGAAGCTCATCTCTAAGAT (1)
- **INTERSECTION (1)**

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

Arabidopsis thaliana pentatricopeptide repeat-containing pro
cds. (1863 bp)
gcgttgctcaaagcctgtcgaattcacaaaaacatcgaattgggagaggaaattggaga
tcatggcggaagatatgttcataaggcaaatattcatgctatggataaaaagtgggacaaag
aaagaacaaggagtagcaaaagttccaggatgtagtacaattagcttggaagggaccaca
ccgatcacacccagagatagagaaaattcaatctaaatggagaatcatgagaaggaaa
agagttagaagagatgctgcttgatctagttgatgatgatgaaagagaggctattgtg
gctattacatacgggttaatcaagactaaaccaggaaccattattcgtataatgaaga
acaagtaacgaagctcatctctaagatatacaagagggatattgtaatgagagatag
gatgggaaatgcagttgtggggactactgg...
position 1264 1480 1498 1507 1602 1634 1662 1692 1754 1803 18
Synonym: K1F13.18; K1F13_18
NM_126050.1 - Arabidopsis thaliana (thale cress) - NCBI - TAIR

Data Export:

下記より最大10000件まで取得できます。

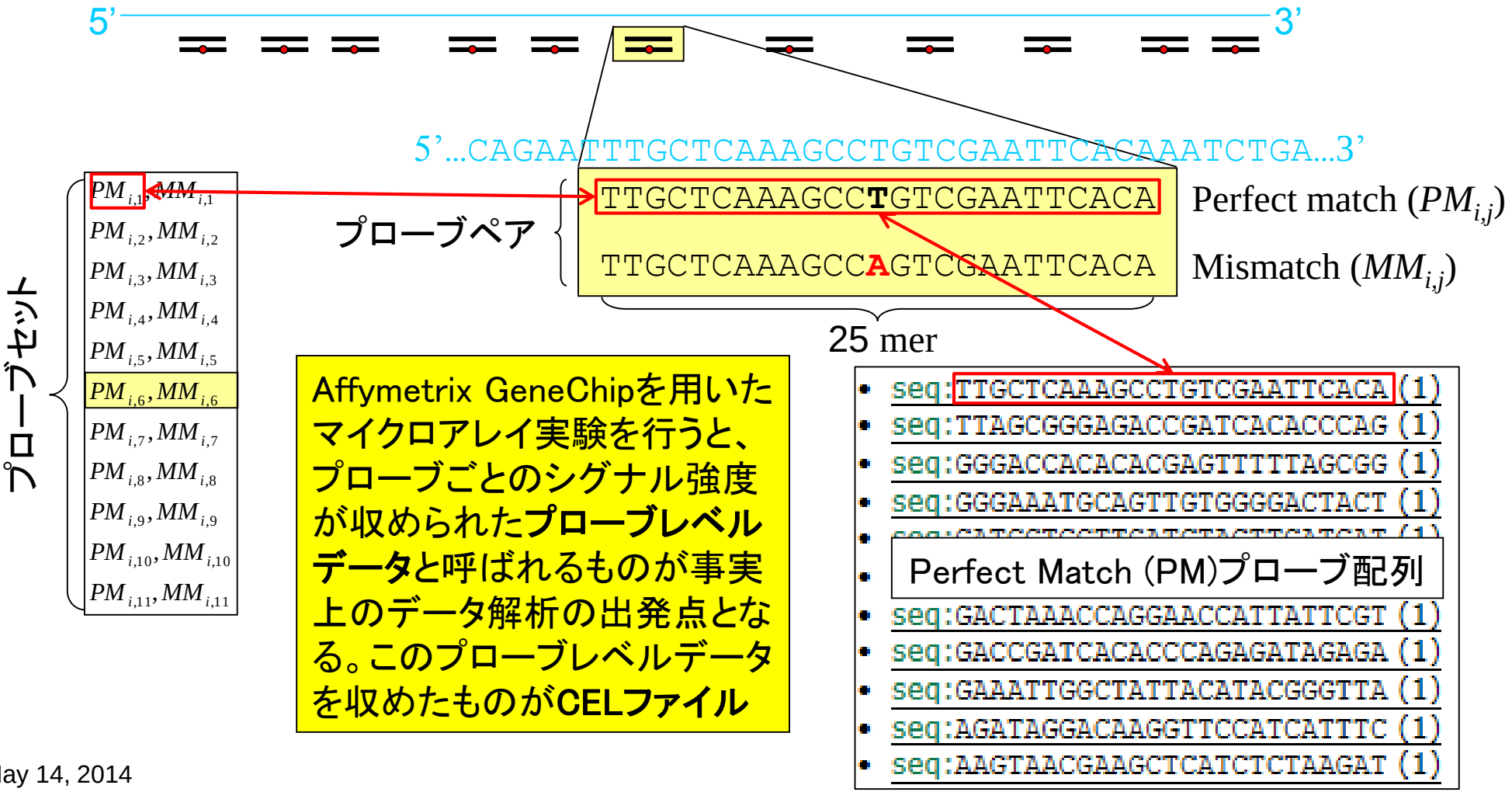
- タブ区切りテキスト → 表示 | ダウンロード
  エクセル等の表計算ソフトに直接コピペできます。
- JSON形式 → リンク | ダウンロード

http://ggrna.dbcls.jp/at/247100_at.txt GGRNA... ggrn...

```
# [ GGRNA.v2 | 2014-05-12 16:10:02 ]
#
# seq:TTGCTCAAAGCCTGTCGAATTCACA 1
# seq:TTAGCGGGAGACCGATCACACCCAG 1
# seq:GGGACCACACACGAGTTTTTAGCGG 1
# seq:GGGAAATGCAGTTGTGGGGACTACT 1
# seq:GATGCTGCTTGATCTAGTTGATGAT 1
# seq:GAGAGGCTATTGTGCATCAGCATAG 1
# seq:GACTAAACCAGGAACCATTATTCGT 1
# seq:GACCGATCACACCCAGAGATAGAGA 1
# seq:GAAATTGGCTATTACATACGGGTTA 1
# seq:AGATAGGACAAGGTTCCATCATTTC 1
# seq:AAGTAACGAAGCTCATCTCTAAGAT 1
# [INTERSECTION]        1
#
# accession     version gi      length  symbol  synonym geneid  division        sou
NM_126050       NM_126050.1     18425083        1863            K1F13.18; K1F13_18
```

Affymetrix GeneChipデータ解析

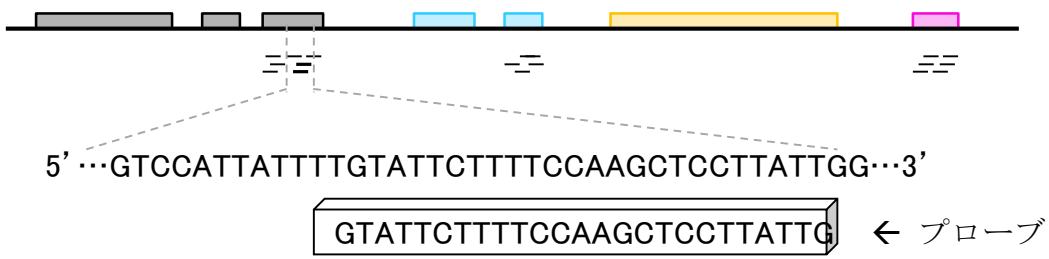
- 遺伝子iの発現量S_iをn_i(n_i=11~20)種類のプローブペアのシグナル強度をもとに計算



Affymetrix GeneChipを用いたマイクロアレイ実験を行うと、プローブごとのシグナル強度が収められた**プローブレベルデータ**と呼ばれるものが事実上のデータ解析の出発点となる。このプローブレベルデータを収めたものが**CELファイル**

ステレオタイプなイメージ

- マイクロアレイの長所
  - 取り扱いやすいデータ量（~100Mb程度）
  - 長年の実績：解析手法がほぼ確立。（Windows Rのみで解析可能）
  - 検査用チップが利用可能（MammaPrintなど）
- マイクロアレイの短所
  - 解析可能範囲が搭載転写物に限定
  - プローブが3’末端に偏っている（3’発現解析用アレイ）
  - **ダイナミックレンジが狭い**



ダイナミックレンジ周辺の雑感

- 既知濃度のspike-inデータとシグナル強度との直線性
- Hekstra et al., *Nucleic Acids Res*., **31**: 1962-1968, 2003
  - マイクロアレイはシグナル強度が高発現側で飽和し、低発現側では実際の濃度よりも高めに見積もられる（Fig. 4B）

プローブレベルのハイブリダイゼーションはLangmuir-adsorption modelに従う

ダイナミックレンジ周辺の雑感

- Langmuir-adsorption modelによる直線性向上の取り組み
  - 非特異的結合（non-specific binding; NSB）の理解
  - 総説(Harrison et al., *Nucleic Acids Res*., **41**: 2779-2796, 2013)
    - Gが4つ以上連続するプローブは外れ値になりやすい(Upton et al., 2008)
    - 4G signatureを持つプローブ同士がGカルテットを形成(Langdon et al., 2009)
    - …
  - 方法
    - Hook法 (Binder et al., *Algorithms Mol. Biol*., **3**: 11, 2008)
    - Inverse Langmuir法(Mulders et al., *BMC Bioinformatics*, **10**: 64, 2009)
    - MSNS model (Furusawa et al., *Bioinformatics*, **25**: 36-41, 2009)

ダイナミックレンジ向上を目指した方法は存在する

ダイナミックレンジ周辺の雑感

- 既知濃度のspike-inデータとシグナル強度との直線性
- “昔の方法”で数値化したアレイデータとの比較が多い
  - Nookaew et al., *Nucleic Acids Res.*, **40**: 10084-10097, 2012
    - PLIER(2004年ごろ)とcubic spline法(Workman et al., 2002)
  - Xu et al., *BMC Bioinformatics*, **14 Suppl 9**: S1, 2013
    - RMA (Irizarry et al., *Biostatistics*, **4**: 249-264, 2003)
  - Raghavachari et al., *BMC Med. Genomics*, **5**: 28, 2012
    - RMA (Irizarry et al., *Biostatistics*, **4**: 249-264, 2003)
  - Mortazavi et al., *Nat. Methods*, **5**: 621-628, 2008
    - MAS5 (Hubbell et al., *Bioinformatics*, **18**: 1585-1592, 2002)

比較的最近の方法との評価をすべきではある

Contents（第1回）

- イントロダクション
  - マイクロアレイの原理や特徴（長所・短所）
  - データ解析例とバイオインフォマティクス要素技術
  - 発現データベース（DB）
  - Affymetrix GeneChipの用語：CELファイル、プローブセット、summarization･･･
- 発現DBからのプローブレベルデータ取得
  - GEOウェブサイト経由
  - R経由（教科書の§2.2.1）
- 前処理法の適用（プローブレベルデータ → 発現行列データ）
  - MAS法、RMA法、RMX法（RobLoxBioC）、IRON法（教科書の§2.2.2～2.2.4）
  - データの正規化（グローバル正規化、quantile正規化）

発現DBからの生データ取得

- Affymetrix GeneChip
  - Ge et al., *Genomics*, 86: 127-141, 2005
    - **GSE2361**、ヒト36サンプル、GPL96を利用
  - Nakai et al., *Biosci Biotechnol Biochem.*, 72: 139-148, 2008
    - **GSE7623**、ラット24サンプル、GPL1355を利用
  - Kamei et al., *PLoS One*, 8: e65732, 2013
    - **GSE30533**、ラット10サンプル、GPL1355を利用
- Illumina BeadChip
  - Sharma et al., *Cancer Cell*, 23: 35-47, 2013
    - **GSE28680**、ヒト24サンプル、GPL10558を利用
- NGSデータも…
  - Neyret-Kahn et al., *Genome Res.*, 23: 1563-1579, 2013
    - **GSE42213**、ヒト26サンプル、GPL10999とGPL11154を利用
      - **GSE42211**、ヒト20サンプル、GPL10999とGPL11154を利用(ChIP-seq)
      - **GSE42212**、ヒト6サンプル、GPL10999を利用(RNA-seq)
  - Huang et al., *Development*, 139: 2161-2169, 2012
    - **GSE36469**、シロイヌナズナ8サンプル、GPL13222を利用

GSE7623のプローブレベルデータ取得(つまりCELファイル取得)を行います。

- 書籍 | トランスクリプトーム解析 | 4.2.1 2群間比較 (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.2 他の実験デザイン(paired, multi-factor, 3群間) (last mo
- 書籍 | トランスクリプト ム解析 | 4.2.3 多群間比較(特異的発現パターン) (last modified 2014
- イントロ | 発現データ取 公共DBから (last modified 2014/05/11) NEW
- イントロ | 発現データ取 | inSilicoDb(Taminau_2011) (last modified 2013/08/20)
- イントロ | 発現データ取得 | ArrayExpress(Kauffmann_2009) (last modified 2013/08/29)推奨
- イントロ | 発現デ タ取得
- イントロ | アノテー
- イントロ | アノテー
- イントロ | アノテー
- イントロ | プローブ
- イントロ | トランス
- イントロ | トランス
- イントロ | Affymet
- 正規化(cDNA or
- 正規化 | Stanford

- イントロ | 発現データ取得 | 公共DBから

イントロ | 発現データ取得 | 公共DBから NEW

遺伝子発現(主にマイクロアレイ)データベースをリストアップします。

一次データベース

- GEO: Barrett et al., Nucleic Acids Res., 2013
  - GSE7623(ラット24サンプル, 62MB): Nakai et al., BBB, 2008
  - GSE30533(ラット10サンプル, 25MB): Kamei et al., PLoS One, 2013
- Array

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Gene Expression Omnibus

GSE7623 Search

Getting Started

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- GEO BLAST
- Programmatic Access
- FTP Site

Browse Content

- Repository Browser
- DataSets: 3413
- Series: 47349
- Platforms: 12911
- Samples: 1132039

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI › GEO › **Accession Display**

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE7

Series GSE7623 Query DataSets

Status	Public on Jan 09, 2008
Title	24 h-fasting effects on the brown and white adipose tissue an
Organism	Rattus norvegicus
Experiment type	Expression profiling by array
Summary	The functional balance between brown adipose tissue (BAT adipose tissue (WAT) is important for metabolic home compared the effects of fasting on the gene expression pr WAT and liver, using DNA microarray analysis. Tissues were o rats that had been fed or fasted for 24 h. Taking the false d (FDR) into account, we extracted the top 1,000 genes that we differentially between fed and fasted rats. In all three ti Ontology analysis revealed marked changes in the ex 'metabolism' category genes and a hypergeometric test d otein biosynthesis-related cate simultaneous down- pathways in the BAT, WA s, there was marked up- on' category, suggestin ved in saving energy as a
Overall design	Rats (7-week-old) were given a commercial diet between 10:0 for 7 days. At 10:00 on day 8, they were divided into two comprised animals of similar body weights. One group continu as described above (fed group, n=4 for array analysis), where group were not fed (fasted group, n=4 for array analysis). received water ad libitum. At 16:00 on day 8, the liver, inter and perinephrial WAT were excised, and analyzed for fasting e
Contributor(s)	Nakai Y, Hashida H, Kadota K, Minami M, Shimizu K, Matsumoto Abe K
Citation(s)	Nakai Y, Hashida H, Kadota K, Minami M et al. Up-regulation of related to the ubiquitin-proteasome system in the brown adipo 24-h-fasted rats. *Biosci Biotechnol Biochem* 2008 Jan;72(1):139 PMID: 18175912

全部で24サンプルのデータからなることが分かる（24枚のアレイを使っている）

Contributor(s)	Nakai Y, Hashida H, Kadota K, Minami M, Shimizu K, Matsumoto I, Kato H, Abe K
Citation(s)	Nakai Y, Hashida H, Kadota K, Minami M et al. Up-regulation of genes related to the ubiquitin-proteasome system in the brown adipose tissue of 24-h-fasted rats. *Biosci Biotechnol Biochem* 2008 Jan;72(1):139-48. PMID: 18175912
Submission date	Apr 2
Last update date	Sep 2
Contact name	Yuji M
Organization name	The U
Street address	1-1-1, Yayoi
City	Bunkyo-ku
State/province	Tokyo
ZIP/Postal code	113-8657
Country	Japan

ラットゲノムからプローブを設計した、**ラット用のチップ**を用いて発現データを得ている

Platforms (1)	GPL1355 [Rat230_2] Affymetrix Rat Genome 230 2.0 Array
Samples (24)	GSM184414 Brown adipose tissue, fed #1
⊞ More...	GSM184415 Brown adipose tissue, fed #2
	GSM184416 Brown adipose tissue, fed #4

Relations

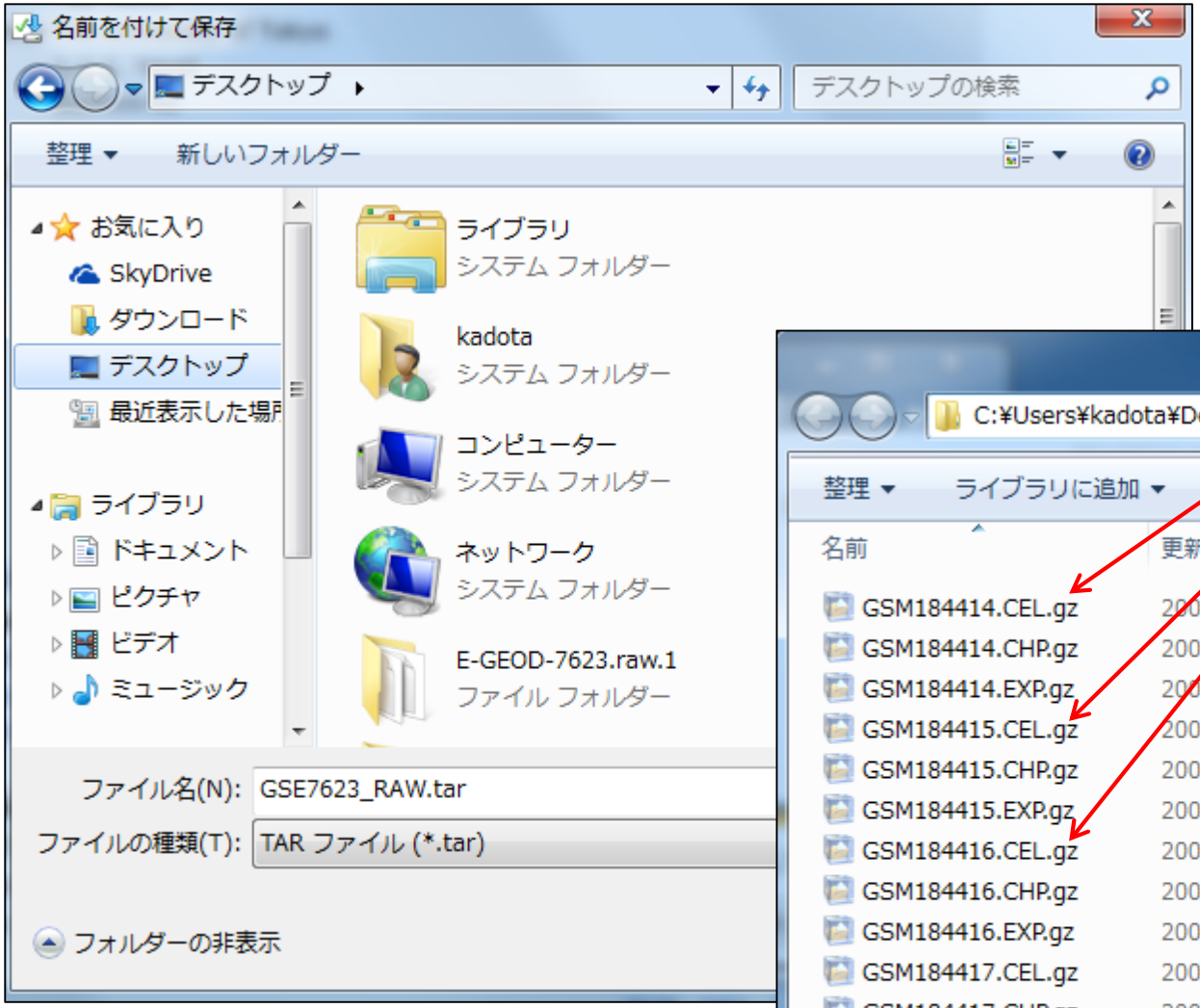
BioProject PRJNA100245

Analyze with GEO2R

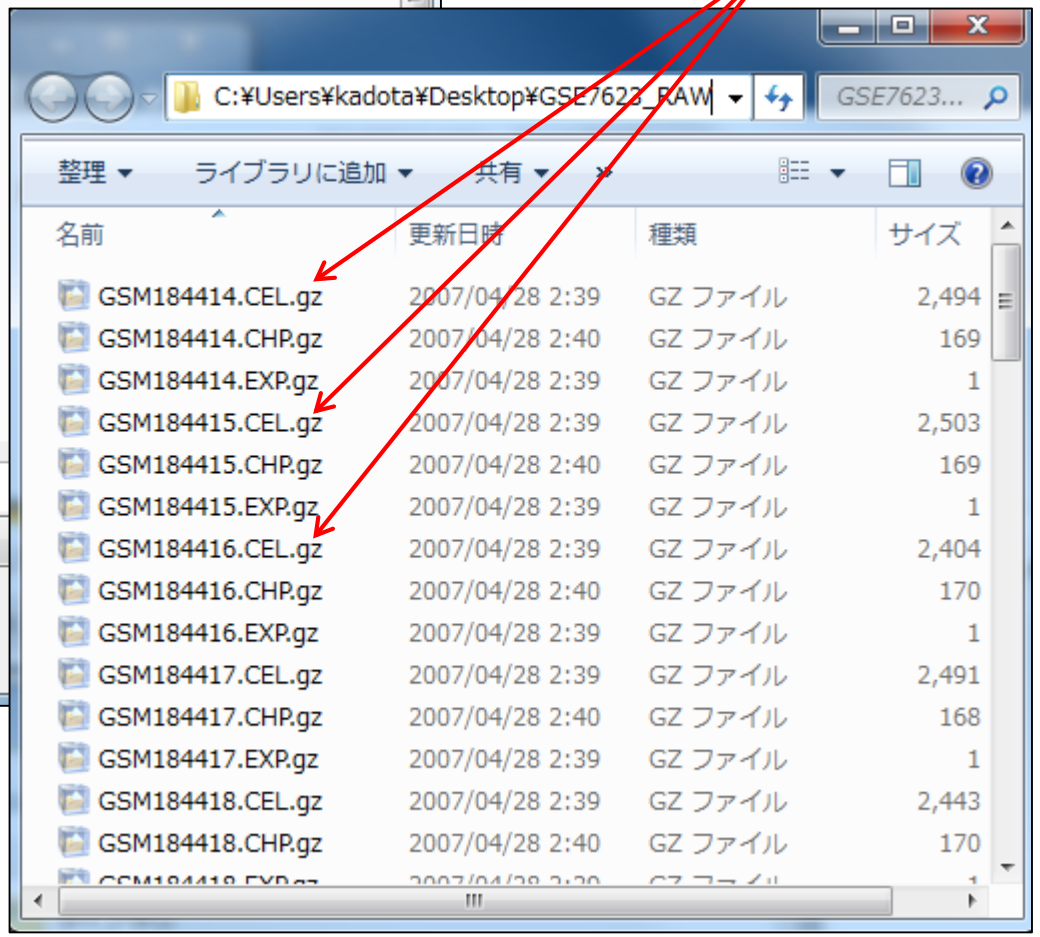
Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

生データのダウンロードはここ。hogeフォルダ中にあり

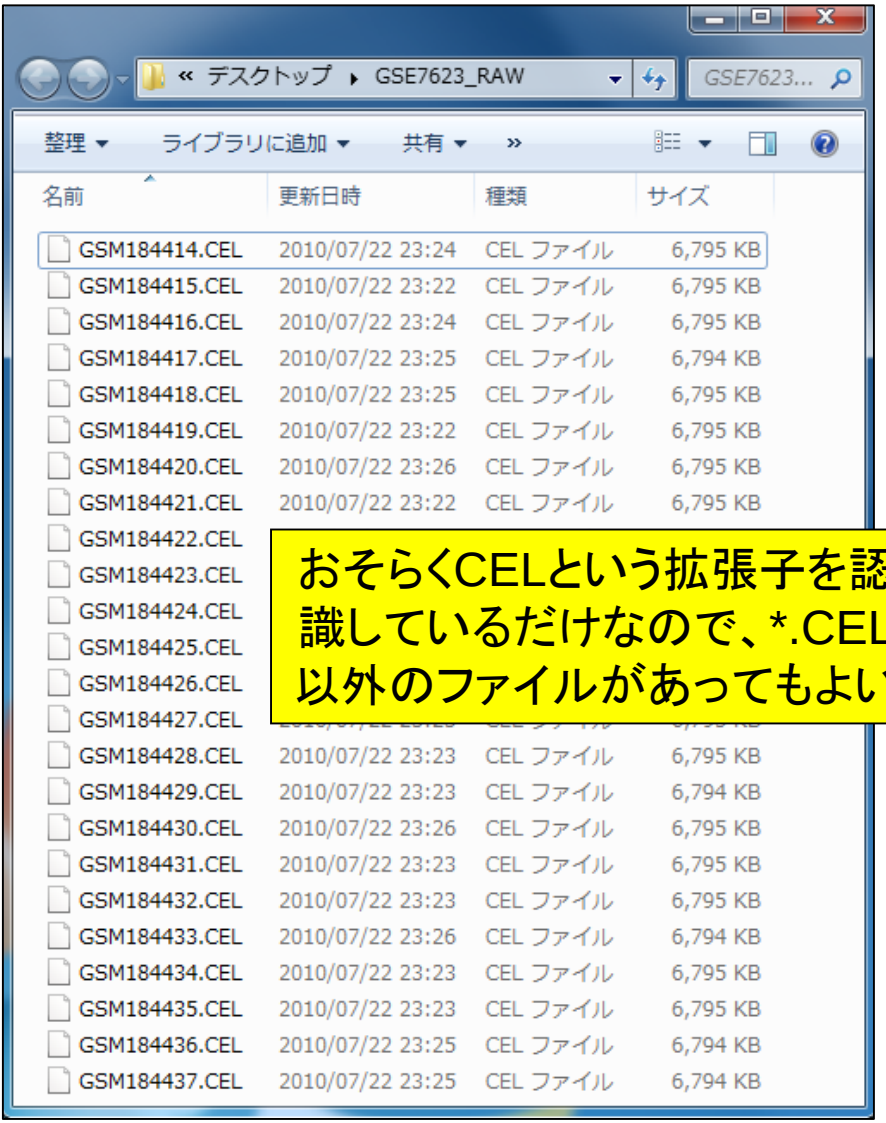
Supplementary file	Size	Download	File type/resource
GSE7623_RAW.tar	61.7 Mb	(http)(custom)	TAR (of CEL, CHP, EXP)
Raw data provided as supplementary file			
Processed data provided as supplementary file			



デスクトップ上でtarファイルを解凍しても、さらにgzip圧縮されたCELファイルの解凍を行う必要がある

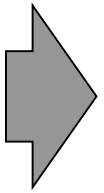


前処理法適用前の状態



おそらくCELという拡張子を認識しているだけなので、*.CEL以外のファイルがあってもよい

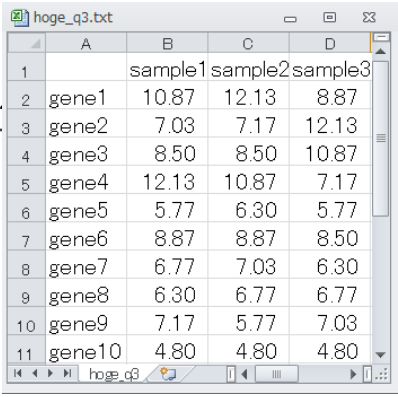
前処理法



－ CELファイル －
チップ上に搭載されている全遺伝子のプローブのシグナル強度情報を含むファイル

遺伝子発現行列

	S1	S2	S3	S4	…
gene *1*	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	…
gene *2*	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	…
…	…	…	…	…	…
gene *i*	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	
…	…	…	…		
gene *n*	$x_{n,1}$	$x_{n,2}$			


hoge_q3.txt

	A	B	C	D
1		sample1	sample2	sample3
2	gene1	10.87	12.13	8.87
3	gene2	7.03	7.17	12.13
4	gene3	8.50	8.50	10.87
5	gene4	12.13	10.87	7.17
6	gene5	5.77	6.30	5.77
7	gene6	8.87	8.87	8.50
8	gene7	6.77	7.03	6.30
9	gene8	6.30	6.77	6.77
10	gene9	7.17	5.77	7.03
11	gene10	4.80	4.80	4.80

R経由で生データ取得(教科書の§2.2.1)

- サンプルデータ (last modified 2013/11/25)
- 書籍|について (last modified 2014/04/17) NEW
- 書籍|トランスクリプトーム解析|1.1 はじめに (last modified 2014/05/09) NEW
- 書籍|トランスクリプトーム解析|1.2.1 原理(Affymetrix 3'発現アレイ) (last modified 2014/05/12) NEW
- 書籍|トランスクリプトーム解析|1.2.2 最近の知見 (last modified 2014/05/09) NEW
- 書籍|トランスクリプトーム解析|2.2.1 生データ(プローブレベルデータ)取得 (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|2.2.2 データの正規化(基礎) (last modified 2014/04/17) NEW
- 書籍|トランスクリプトーム解析|2.2.3 データの正規化(計算例) (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|2.2.4 データの正規化(その他) (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|2.2.5 アノテーション情報 (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|3.2.1 クラスタリング(データ変換や距離の定義など) (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|3.2.2 実験デザイン、データ分布、統計解析との関係 (last modified 2014/04/18) NEW
- 書籍|トランスクリプトーム解析|3.2.3 多重比較問題 (last modified 2014/04/19) NEW
- 書籍|トランスクリプトーム解析|3.2.4 各種プロット(M-A plotや
- 書籍|トランスクリプトーム解析|4.2.1 2群間比較 (last modified
- 書籍|トランスクリプトーム解析|4.2.2 他の実験デザイン(paired
- 書籍|トランスクリプトーム解析|4.2.3 多群間比較(特異的発現
- イントロ|発現データ取得|公共DBから (last modified 2014/05/
- イントロ|発現データ取得|inSilicoDb(Taminau 2011) (last mod
- イントロ|発現データ取得|ArrayExpress(Kauffmann 2009) (las
- イントロ|発現データ取得|GEOquery(Davis 2007) (last modifi
- イントロ|アノテーション情報取得|公共DB(GEO)から (last mod
- イントロ|アノテーション情報取得|GEOquery(Davis 2007) (last
- イントロ|アノテーション情報取得|Rのパッケージ*.dbから (last

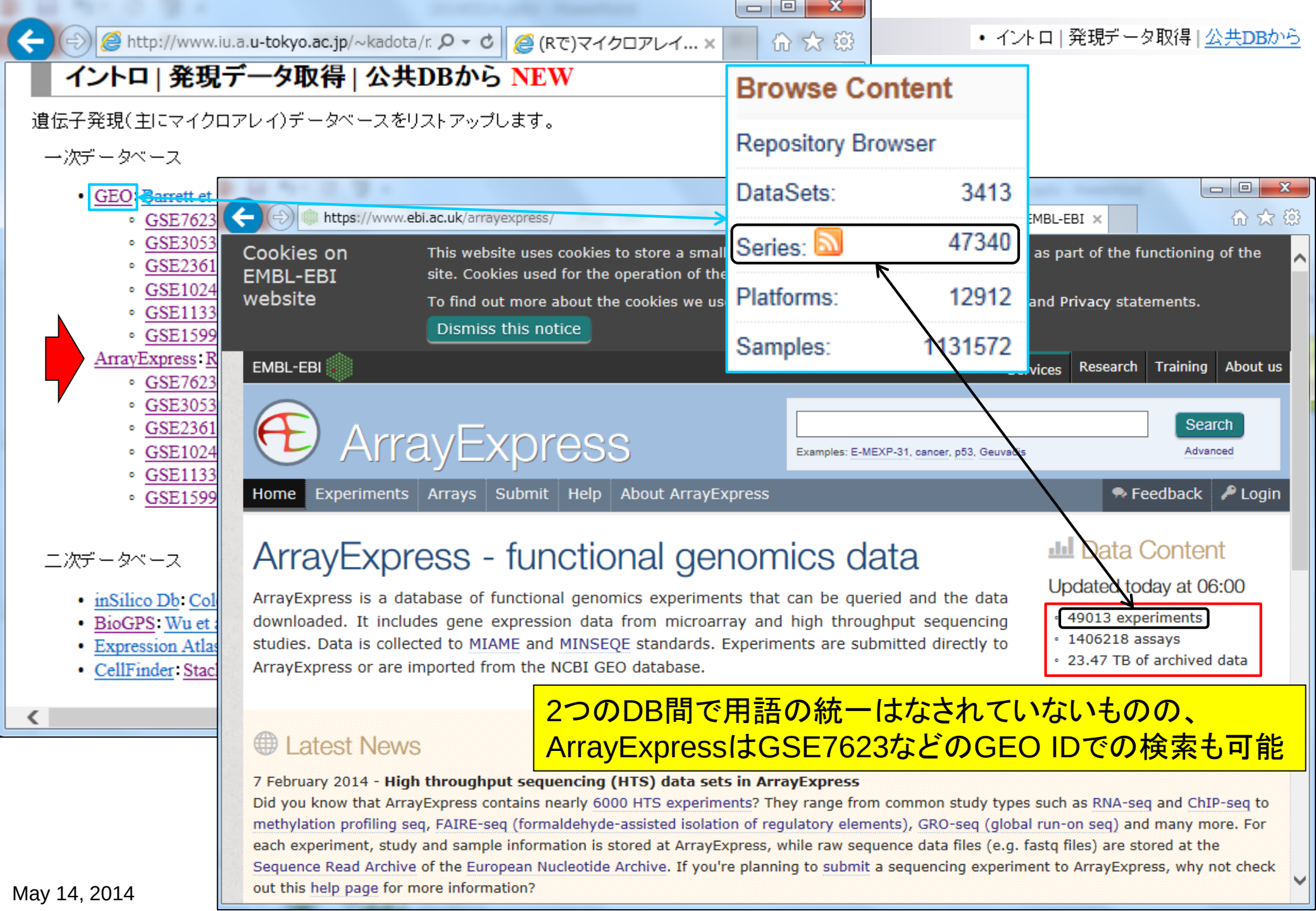
経験上、R経由でArrayExpressからのダウンロードのほうが簡単

書籍|について NEW

2014年4月に(Rで)塩基配列解析および(Rで)マイクロアレイデータ解析を体系的にまとめた以下の書籍が出版されました。ここでは(Rで)マイクロアレイデータ解析に関連したマイクロアレイ部分の書籍中のRコードを章・節・項ごとに示します。

なるべく書籍中の記述形式に準拠しますが、例えばp72中の最初のsetwd("C:/Users/kadota/Desktop")というコマンドを忠実に実行してもエラーが出るだけです。これはkadotaというヒトのPC上でのみ成立するコマンドだからです。ここで利用しているsetwd関数は作業ディレクトリの変更に相当し、「ファイル」-「ディレクトリの変更」でデスクトップに移動することと同義です。したがって、このページ全体で統一的に使っているように、ディレクトリの変更作業自体はR Gui画面左上の「ファイル」メニューを利用することとし、setwd関数部分の記述は省略します。

- 門田幸二著(金明哲 編), シリーズ Useful R 第7巻トランスクリプトーム解析, 共立出版, 2014年4月. ISBN: 978-4-320-12370-0


イントロ | 発現データ取得 | 公共DBから NEW
遺伝子発現(主にマイクロアレイ)データベースをリストアップします。
一次データベース
GEO: Barrett et
GSE7623
GSE3053
GSE2361
GSE1024
GSE1133
GSE1599
ArrayExpress: R
GSE7623
GSE3053
GSE2361
GSE1024
GSE1133
GSE1599
二次データベース
inSilico Db: Col
BioGPS: Wu et
Expression Atlas
CellFinder: Stac
イントロ | 発現データ取得 | 公共DBから
Browse Content
Repository Browser
DataSets: 3413
Series: 47340
Platforms: 12912
Samples: 1131572
https://www.ebi.ac.uk/arrayexpress/
EMBL-EBI
Cookies on EMBL-EBI website
This website uses cookies to store a small
site. Cookies used for the operation of the
To find out more about the cookies we us
Dismiss this notice
as part of the functioning of the
and Privacy statements.
EMBL-EBI
Research
Training
About us
ArrayExpress
Search
Examples: E-MEXP-31, cancer, p53, Geuvadis
Advanced
Home
Experiments
Arrays
Submit
Help
About ArrayExpress
Feedback
Login
ArrayExpress - functional genomics data
ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to MIAME and MINSEQE standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.
Data Content
Updated today at 06:00
49013 experiments
1406218 assays
23.47 TB of archived data
2つのDB間で用語の統一はなされていないものの、
ArrayExpressはGSE7623などのGEO IDでの検索も可能
Latest News
7 February 2014 - High throughput sequencing (HTS) data sets in ArrayExpress
Did you know that ArrayExpress contains nearly 6000 HTS experiments? They range from common study types such as RNA-seq and ChIP-seq to methylation profiling seq, FAIRE-seq (formaldehyde-assisted isolation of regulatory elements), GRO-seq (global run-on seq) and many more. For each experiment, study and sample information is stored at ArrayExpress, while raw sequence data files (e.g. fastq files) are stored at the Sequence Read Archive of the European Nucleotide Archive. If you're planning to submit a sequencing experiment to ArrayExpress, why not check out this help page for more information?

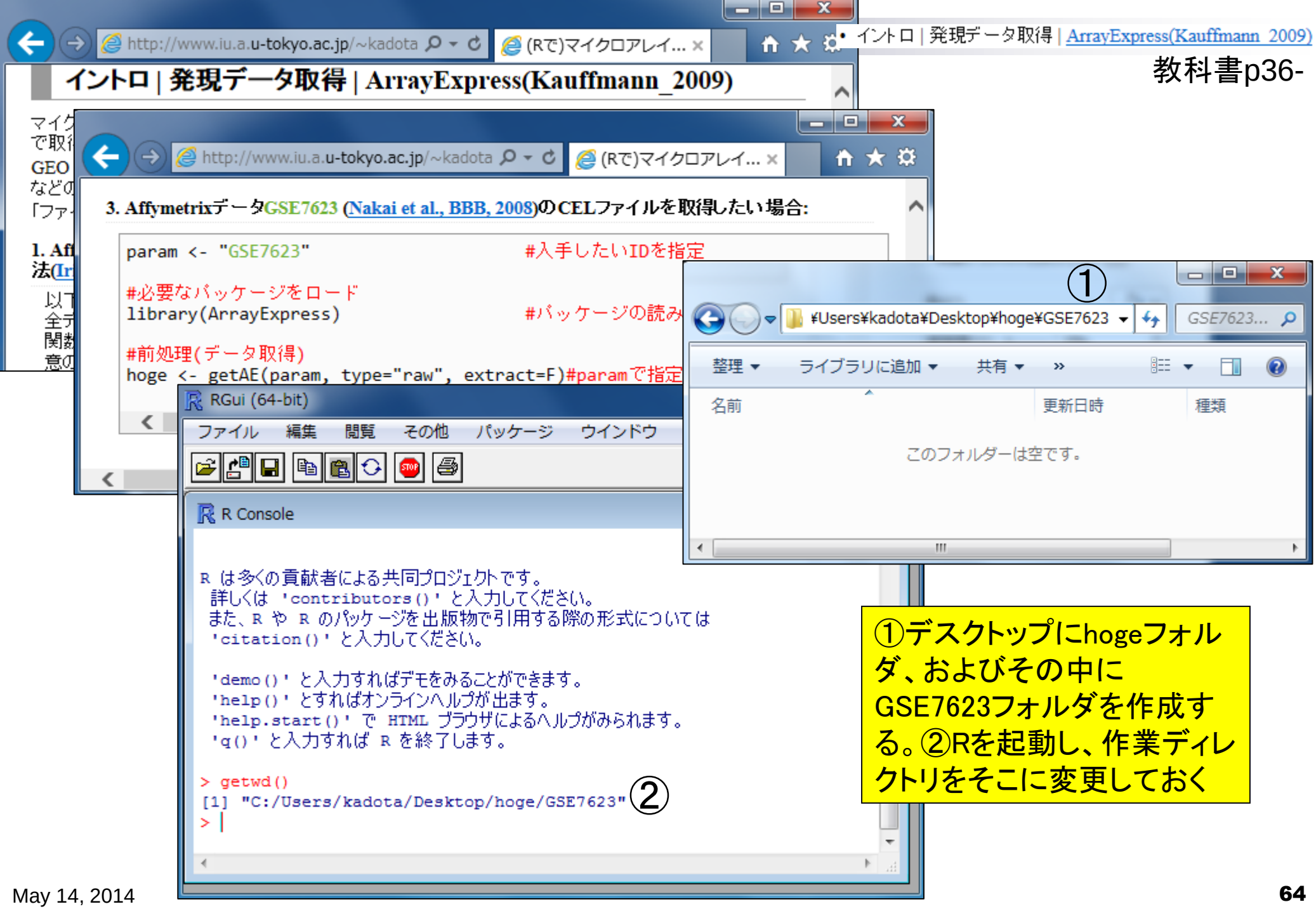
R経由で生データ取得（教科書の§2.2.1）

- サンプルデータ (last modified 2013/11/25)
- 書籍 | について (last modified 2014/04/17) NEW
- 書籍 | トランスクリプトーム解析 | 1.1 はじめに (last modified 2014/05/09) NEW
- 書籍 | トランスクリプトーム解析 | 1.2.1 原理(Affymetrix 3'発現アレイ) (last modified 2014/05/12) NEW
- 書籍 | トランスクリプトーム解析 | 1.2.2 最近の知見 (last modified 2014/05/09) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.1 生データ(プローブレベルデータ)取得 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.2 データの正規化(基礎) (last modified 2014/04/17) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.3 データの正規化(計算例) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.4 データの正規化(その他) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 2.2.5 アノテーション情報 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.1 クラスタリング(データ変換や距離の定義など) (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.2 実験デザイン、データ分布、統計解析との関係 (last modified 2014/04/18) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.3 多重比較問題 (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 3.2.4 各種プロット(M-A plotや平均-分散プロットなど) (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.1 2群間比較 (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.2 他の実験デザイン(paired, multi-factor, 3群間) (last modified 2014/04/19) NEW
- 書籍 | トランスクリプトーム解析 | 4.2.3 多群間比較(特異的発現パターン) (last modified 2014/04/20) NEW
- イントロ | 発現データ取得 | 公共DBから (last modified 2014/05/11) NEW
- イントロ | 発現データ取得 | inSilicoDb(Taminau 2011) (last modified 2013/08/20)
- イントロ | 発現データ取得 | ArrayExpress(Kauffmann 2009) (last modified 2013/08/29)推奨
- イントロ | 発現データ取得 | GEOquery(Davis 2007) (last modified 2013/08/20)
- イントロ | アノテーション情報取得 | 公共DB(GEO)から (last modified 2013/08/18)
- イントロ | アノテーション情報取得 | GEOquery(Davis 2007) (last modified 2013/09/18)推奨
- イントロ | アノテーション情報取得 | Rのパッケージ*.dbから (last modified 2013/08/18)

教科書中のRコードはこちら

GSE7623のプローブレベルデータ取得（つまりCELファイル取得）をR経由で行います。

教科書p36-



①デスクトップにhogeフォルダ、およびその中にGSE7623フォルダを作成する。②Rを起動し、作業ディレクトリをそこに変更しておく

教科書p36-

R Console

```
 URL 'http://www.ebi.ac.uk/arrayexpress/files/E-GEOD-7623/E-GEOD-$
Content type 'text/plain' length 4484 bytes
 開かれた URL
downloaded 4484 bytes

Copying raw data files

 URL 'http://www.ebi.ac.uk/arrayexpress/files/E-GEOD-7623/E-GEOD-$
Content
 開かれた
downloa

>
```

コピペ実行後

Users¥kadota¥Desktop¥hoge¥GSE7623

整理 ライブラリに追加 共有

名前

A-AFFY-43.adf.txt

E-GEOD-7623.idf.txt

E-GEOD-7623.raw.1.zip 2014/05/13 12:25 ZIP ファイル

E-GEOD-7623.sdrf.txt 2014/05/13 12:22 テキスト ドキ

4つのファイルが作成されるので、zipファイルを解凍

37% / E-GEOD-7623.raw.1.zip

C:¥Users¥kadota¥Desktop¥hoge¥GSE7623¥E-GEOD-7623raw.1zip

GSM184431.CEL

キャンセル

C:¥Users¥kadota¥Desktop¥E-GEOD-7623.raw.1

整理 ライブラリに追加 共有 書き込む

名前	更新日時	種類	サイズ
GSM184414.CEL	2010/07/22 23:24	CEL ファイル	6,795 KB
GSM184415.CEL	2010/07/22 23:22	CEL ファイル	6,795 KB
GSM184416.CEL	2010/07/22 23:24	CEL ファイル	6,795 KB
GSM184417.CEL	2010/07/22 23:25	CEL ファイル	6,794 KB
GSM184418.CEL	2010/07/22 23:25	CEL ファイル	6,795 KB
GSM184419.CEL	2010/07/22 23:22	CEL ファイル	6,795 KB
GSM184420.CEL	2010/07/22 23:26	CEL ファイル	6,795 KB
GSM184421.CEL	2010/07/22 23:22	CEL ファイル	6,795 KB
GSM184422.CEL	2010/07		KB
GSM184423.CEL	2010/07		KB
GSM184424.CEL	2010/07		KB
GSM184425.CEL	2010/07/22 23:23	CEL ファイル	6,794 KB
GSM184426.CEL	2010/07/22 23:27	CEL ファイル	6,795 KB
GSM184427.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184428.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184429.CEL	2010/07/22 23:23	CEL ファイル	6,794 KB
GSM184430.CEL	2010/07/22 23:26	CEL ファイル	6,795 KB
GSM184431.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184432.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184433.CEL	2010/07/22 23:26	CEL ファイル	6,794 KB
GSM184434.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184435.CEL	2010/07/22 23:23	CEL ファイル	6,795 KB
GSM184436.CEL	2010/07/22 23:25	CEL ファイル	6,794 KB
GSM184437.CEL	2010/07/22 23:25	CEL ファイル	6,794 KB

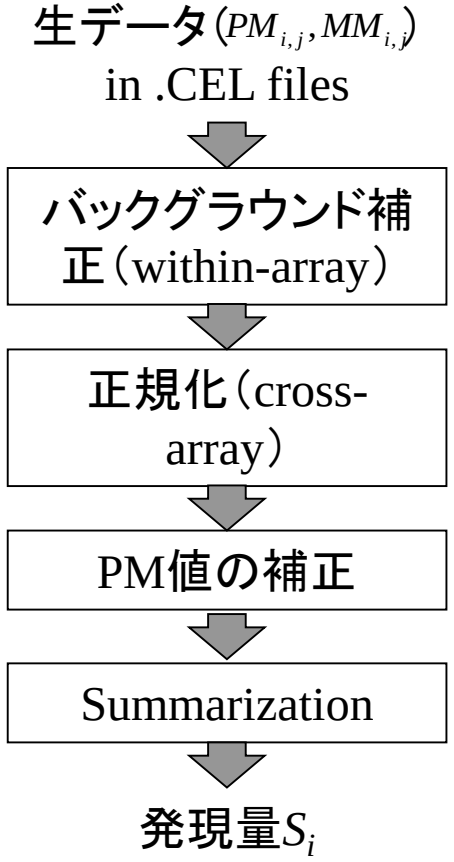
CELファイルのみからなるので便利

Contents（第1回）

- イントロダクション
  - マイクロアレイの原理や特徴（長所・短所）
  - データ解析例とバイオインフォマティクス要素技術
  - 発現データベース（DB）
  - Affymetrix GeneChipの用語：CELファイル、プローブセット、summarization･･･
- 発現DBからのプローブレベルデータ取得
  - GEOウェブサイト経由
  - R経由（教科書の§2.2.1）
- 前処理法の適用（プローブレベルデータ → 発現行列データ）
  - MAS法、RMA法、RMX法（RobLoxBioC）、IRON法（教科書の§2.2.2～2.2.4）
  - データの正規化（グローバル正規化、quantile正規化）

様々な前処理法

- MBEI (Li and Wong, *PNAS*, **98**: 31-36, 2001)
- VSN (Huber et al., *Bioinformatics*, **18** Suppl 1: S96-104, 2002)
- MAS5 (Hubbell et al., *Bioinformatics*, **18**: 1585-92, 2002)
- RMA (Irizarry et al., *Biostatistics*, **4**: 249-64, 2003)
- GCRMA (Wu et al., *Tech. Rep., John Hopkins Univ.*, 2003)
- PDNN (Zhang et al., *Nat. Biotechnol.*, **21**: 818-21, 2003)
- PLIER (Affymetrix, 2004)
- SuperNorm (Konishi, T., *BMC Bioinformatics*, **5**: 5, 2004)
- multi-mgMOS (Liu et al., *Bioinformatics*, **21**: 3637-3644, 2005)
- GLA (Zhou and Rocke, *Bioinformatics*, **21**: 3983-3989, 2005)
- FARMS (Hochreiter et al., *Bioinformatics*, **22**: 943-949, 2006)
- DFW (Chen et al., *Bioinformatics*, **23**: 321-327, 2007)
- Hook (Binder et al., *AMB*, **3**: 11, 2008)
- GRSN (Pelz et al., *BMC Bioinformatics*, **9**: 520, 2008)
- RMX (Kohl et al., *BMC Bioinformatics*, **11**: 583, 2010)
- KDL and KDQ (Hsieh et al., *BMC Bioinformatics*, **12**: 222, 2011)
- IRON (Welsh et al., *BMC Bioinformatics*, **14**: 153, 2013)



様々な前処理法

- **MAS5** (Hubbell et al., *Bioinformatics*, **18**: 1585-92, 2002)
  - 特徴：アレイごとに独立して前処理を実行（per-array basis）
  - 正規化：**グローバル正規化**
- **RMA** (Irizarry et al., *Biostatistics*, **4**: 249-64, 2003)
  - 特徴：読み込んだ複数サンプル（複数アレイ）の情報を用いて前処理を実行（multi-array basis）
  - 正規化：**quantile正規化**（プローブレベルデータに対して実行）

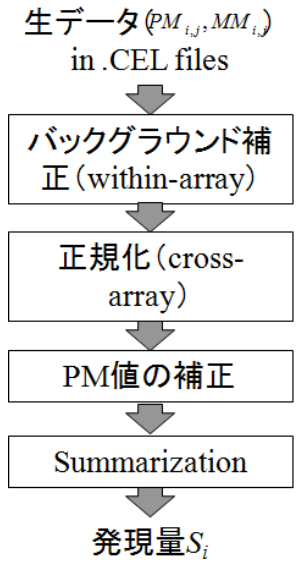


Table 1: Frequency of preprocessing algorithms used during 2003 – 2008

	2003	2004	2005	2006	2007	2008
MAS (2002)	8	34	53	42	47	16
RMA (2003)		8	15	29	20	9
MBEI (2001)	0	3	7	16	8	3
GCRMA (2004)			0	5	8	4
VSN (2002)	0	0	0	4	0	2

Our investigation was performed for 394 different papers with analyses performed using the Affymetrix HG-U133A array (Gene Expression Omnibus (GEO) ID: GPL96) [32]. These results were

よく使われているのはMAS5とRMAです

要素技術(グローバル正規化)

- 「各サンプルから測定されたmRNAの全体量は一定」と仮定
  - マイクロアレイ上の遺伝子数が少ない場合は非現実的だが、数千～数万種類の遺伝子が搭載されているので妥当(だろう)

	sample1	sample2
gene1	10.5	12.4
gene2	6.4	7.1
gene3	8.0	8.5
gene4	10.8	11.4
gene5	5.6	6.7
gene6	8.4	8.9
gene7	6.2	7.0
gene8	6.1	6.8
gene9	6.6	6.5
gene10	5.1	5.8
平均値	7.4	8.1

正規化

	sample1	sample2
gene1	14.2	15.3
gene2	8.7	8.8
gene3	10.9	10.5
gene4	14.7	14.1
gene5	7.6	8.3
gene6	11.4	11.0
gene7	8.4	8.6
gene8	8.3	8.4
gene9	9.0	8.0
gene10	6.9	7.2
平均値	10.0	10.0

チップごとに独立して正規化(per-array basis)

要素技術（Quantile正規化）

- 「シグナル強度の順位が同じなら値も同じ」と仮定

正規化前

s1	s2
10.5	12.4
6.4	7.1
8.0	8.5
10.8	11.4
5.6	6.7
8.4	8.9
6.2	7.0
6.1	6.8
6.6	6.5
5.1	5.8

data19.txt

列ごとにソート →

s1	s2
5.1	5.8
5.6	6.5
6.1	6.7
6.2	6.8
6.4	7.0
6.6	7.1
8.0	8.5
8.4	8.9
10.5	11.4
10.8	12.4

行ごとの平均を算出 →

Ave.
5.5
6.1
6.4
6.5
6.7
6.9
8.3
8.7
11.0
11.6

対応する行の要素の元の位置に平均値を代入 →

正規化後

s1	s2
11.0	11.6
6.7	6.9
8.3	8.3
11.6	11.0
6.1	6.4
8.7	8.7
6.5	6.7
6.4	6.5
6.9	6.1
5.5	5.5

データセット中のサンプル数が変わると結果が変わる（multi-array basis）

要素技術（Quantile正規化）

- 「シグナル強度の順位が同じなら値も同じ」と仮定

正規化前

s1	s2	s3
10.5	12.4	9.3
6.4	7.1	13.2
8.0	8.5	10.7
10.8	11.4	7.8
5.6	6.7	5.2
8.4	8.9	9.0
6.2	7.0	6.1
6.1	6.8	7.3
6.6	6.5	7.7
5.1	5.8	3.5

列ごとにソート →

s1	s2	s3
5.1	5.8	3.5
5.6	6.5	5.2
6.1	6.7	6.1
6.2	6.8	7.3
6.4	7.0	7.7
6.6	7.1	7.8
8.0	8.5	9.0
8.4	8.9	9.3
10.5	11.4	10.7
10.8	12.4	13.2

行ごとの平均を算出 →

Ave.
4.8
5.8
6.3
6.8
7.0
7.2
8.5
8.9
10.9
12.1

対応する行の要素の元の位置に平均値を代入 →

正規化後

s1	s2	s3
10.9	12.1	8.9
7.0	7.2	12.1
8.5	8.5	10.9
12.1	10.9	7.2
5.8	6.3	5.8
8.9	8.9	8.5
6.8	7.0	6.3
6.3	6.8	6.8
7.2	5.8	7.0
4.8	4.8	4.8

data19_plus1.txt

データセット中のサンプル数が変わると結果が変わる（multi-array basis）

BMC Bioinformatics. 2013 Apr 11;14:124. doi: 10.1186/1471-2105-14-124.

The impact of quantile and rank normalization procedures o power of gene differential expression analysis.

Qiu X, Wu H, Hu R.

Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Rochester, New York 14642, USA. huruizg@hotmail.com.

Abstract

BACKGROUND: Quantile and rank normalizations are two widely used pre-pr designed to remove technological noise presented in genomic data. Subseque such as gene differential expression analysis is usually based on normalized e study, we find that these normalization procedures can have a profound impac expression analysis, especially in terms of testing power.

RESULTS: We conduct theoretical derivations to show that the testing power o expression analysis based on quantile or rank normalized gene expressions ca with fixed sample size no matter how strong the gene differentiation effects are extensive simulation analyses and find the results corroborate theoretical predi

CONCLUSIONS: Our finding may explain why genes with well documented str not always detected in microarray analysis. It provides new insights in microarr design and will help practitioners in selecting proper normalization procedures.

マイクロアレイ

RNA−seq

正規化はRNA−seq
でも議論されている

RNA. 2012 Jun;18(6):1279-88. doi: 10.1261/rna.030916.111. Epub 2012 Apr 24.

Evaluation of normalization methods in mammalian microRNA-Seq data.

Garmire LX, Subramaniam S.

Department of Bioengineering, Jacobs School of Engineering, University of California at San Diego, La Jolla, California 92093-0412, USA. lgarmire@gmail.com

Abstract

Simple total tag count normalization is inadequate for microRNA sequencing data generated from the next generation sequencing technology. However, so far systematic evaluation of normalization methods on microRNA sequencing data is lacking. We comprehensively evaluate seven commonly used normalization methods including global normalization, Lowess normalization, Trimmed Mean Method (TMM), quantile normalization, scaling normalization, variance stabilization, and invariant method. We assess these methods on two individual experimental data sets with the empirical statistical metrics of mean square error (MSE) and Kolmogorov-Smirnov (K-S) statistic. Additionally, we evaluate the methods with results from quantitative PCR validation. Our results consistently show that Lowess normalization and quantile normalization perform the best, whereas TMM, a method applied to the RNA-Sequencing normalization, performs the worst. The poor performance of TMM normalization is further evidenced by abnormal results from the test of differential expression (DE) of microRNA-Seq data. Comparing with the models used for DE, the choice of normalization method is the primary factor that affects the results of DE. In summary, Lowess normalization and quantile normalization are recommended for normalizing microRNA-Seq data, whereas the TMM method should be used with caution.

Brief Bioinform. 2012 Sep 17. [Epub ahead of print]

A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Guedj M, Jaffrézic F; on behalf of The French StatOmique Consortium.

Abstract

During the last 3 years, a number of approaches for the normalization of RNA sequencing data have emerged in the literature, differing both in the type of bias adjustment and in the statistical strategy adopted. However, as data continue to accumulate, there has been no clear consensus on the appropriate normalization method to be used or the impact of a chosen method on the downstream analysis. In this work, we focus on a comprehensive comparison of seven recently proposed normalization methods for the differential analysis of RNA-seq data, with an emphasis on the use of varied real and simulated datasets involving different species and experimental designs to represent data characteristics commonly observed in practice. Based on this comparison study, we propose practical recommendations on the appropriate normalization method to be used and its impact on the differential analysis of RNA-seq data.

前処理法の違いを実感してみよう

- MAS5 (Hubbell et al., *Bioinformatics*, **18**: 1585–92, 2002)
  - 特徴：アレイごとに独立して前処理を実行（per-array basis）
  - 正規化：**グローバル正規化**
- RMA (Irizarry et al., *Biostatistics*, **4**: 249–64, 2003)
  - 特徴：読み込んだ複数サンプル（複数アレイ）の情報を用いて前処理を実行（multi-array basis）
  - 正規化：**quantile正規化**（プローブレベルデータに対して実行）
- RMX (Kohl et al., *BMC Bioinformatics*, **11**: 583, 2010)
  - 教科書中のRobLoxBioCと同じ方法

- 正規化 | Affymetrix GeneChip | について (last modified 2013/09/02)
- 正規化 | Affymetrix GeneChip | frma (McCall_2010) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | rmx (Kohl_2010) (last modified 2013/11/19)推奨
- 正規化 | Affymetrix GeneChip | GRSN (Pelz_2008) (last modified 2013/05/27)
- 正規化 | Affymetrix GeneChip | Hook (Binder_2008) (last modified 2013/05/30)
- 正規化 | Affymetrix GeneChip | DFW (Chen_2007) (last modified 2013/08/20)
- 正規化 | Affymetrix GeneChip | FARMS (Hochreiter_2006) (last modified 2013/08/20)
- 正規化 | Affymetrix GeneChip | multi-mgMOS (Liu_2005) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | GCRMA (Wu_2004) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | PLIER (Affymetrix_2004) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | VSN (Huber_2002) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | RMA (Irizarry_2003) (last modified 2013/08/21)
- 正規化 | Affymetrix GeneChip | MAS5.0 (Hubbell_2002) (last modified 2013/11/25)
- 正規化 | Affymetrix GeneChip | MBEI (Li_2001) (last modified 2013/08/21)

3つの前処理法をやります

正規化 | Affymetrix GeneChip | RMA (Irizarry_2003)

Affymetrix chip (GeneChipTM)を用いて得られた*.CELファイルを元に、RMA(Irizarry et al., Biostatistics, 2003)アルゴリズムを用いてSummary scoreを算出。
「ファイル」-「ディレクトリの変更

1. (CELファイルがあるディレ

```
out_f <- "hoge1.txt"

#必要なパッケージをロー
library(affy)

#データファイルの読み込
hoge <- ReadAffy()

#本番
eset <- rma(hoge)

#ファイルに保存
write.exprs(eset, fil
```

正規化 | Affymetrix GeneChip | MAS5.0 (Hubbell_2002)

Affymetrix chip (GeneChipTM)を用いて得られた*.CELファイルを元に、MAS5.0 (Hubbell et al., Bioinformatics, 2002)アルゴリズムを用いてSummary scoreを算出するやり方を示します。低発現領域で
のばらつきが大きいことが指摘さ
をすれば決して悪い方法ではな
アレイごとに独立して正規化を行
の利点があります。
「ファイル」-「ディレクトリの変更」

1. (CELファイルがあるディレクト

```
out_f <- "hoge1.txt"

#必要なパッケージをロード
library(affy)

#データファイルの読み込み
hoge <- ReadAffy()

#本番
eset <- mas5(hoge)

#対数変換
summary(exprs(eset))
exprs(eset)[exprs(eset)
summary(exprs(eset))
exprs(eset) <- log(expr

#ファイルに保存
write.exprs(eset, file=
```

正規化 | Affymetrix GeneChip | rmx (Kohl_2010)

RobLoxBioCというRパッケージ中に実装されているrob
summarization法です。論文中にMAS5の拡張版と書い
score計算時に(MAS5で採用されている)Tukey's biweig
ポイントのようです。サンプルごとに独立して正規化を
前のデータになっているので、robloxbioc関数をかけた
換したものを出力しています。
「ファイル」-「ディレクトリの変更」で適切なディレクトリに

1. (CELファイルがあるディレクトリ上で)手元にあるCEL

```
out_f <- "hoge1.txt"                    #出力ファイル名を指定してout_fに格納

#必要なパッケージをロード
library(RobLoxBioC)                     #パッケージの読み込み

#データファイルの読み込み(*.CELファイル)
hoge <- ReadAffy()                      #*.CELファイルの読み込み

#本番
eset <- robloxbioc(hoge)                #rmxを実行し、結果をesetに保存

#対数変換
summary(exprs(eset))                    #得られたesetの遺伝子発現行列のシグナル強度
exprs(eset)[exprs(eset) < 1] <- 1       #対数変換（log2）できるようにシグナル強度が
summary(exprs(eset))                    #上記処理後のシグナル強度分布を再び表示させ
exprs(eset) <- log(exprs(eset), 2)      #底を2として対数変換

#ファイルに保存
write.exprs(eset, file=out_f)           #結果を指定したファイル名で保存
```

トップページへ

hoge – GSE7623_24samples フォルダ中には、実行後のファイルがある。実際にやるのはGSE7623_02samplesのみ

hoge – GSE7623_24samples フォルダにディレクトリ変更して前処理法を実行。テンプレートスクリプトは出力ファイル名が同じことに注意

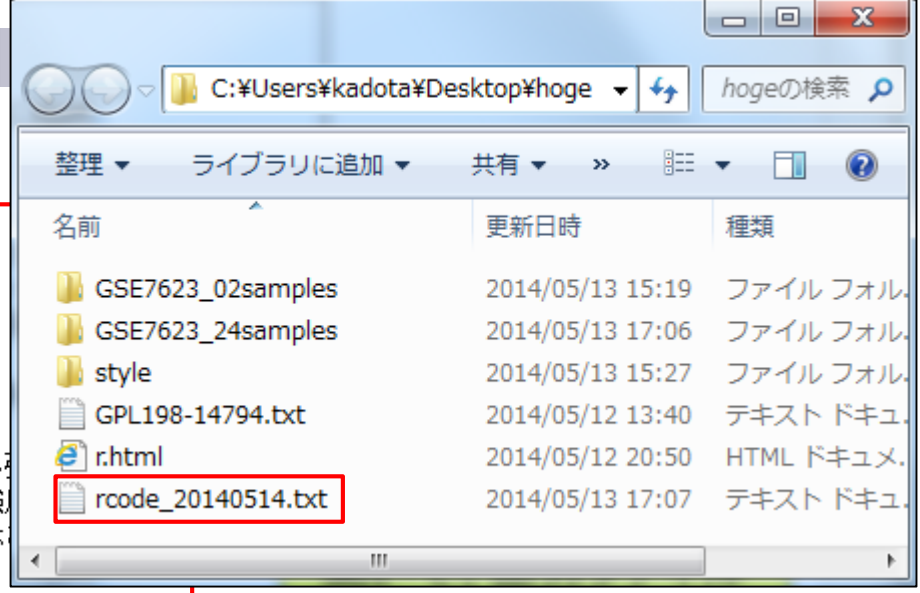
3つのコードの主な違いは、前処理法の違いを表す関数名とパッケージ名部分

門田のやり方

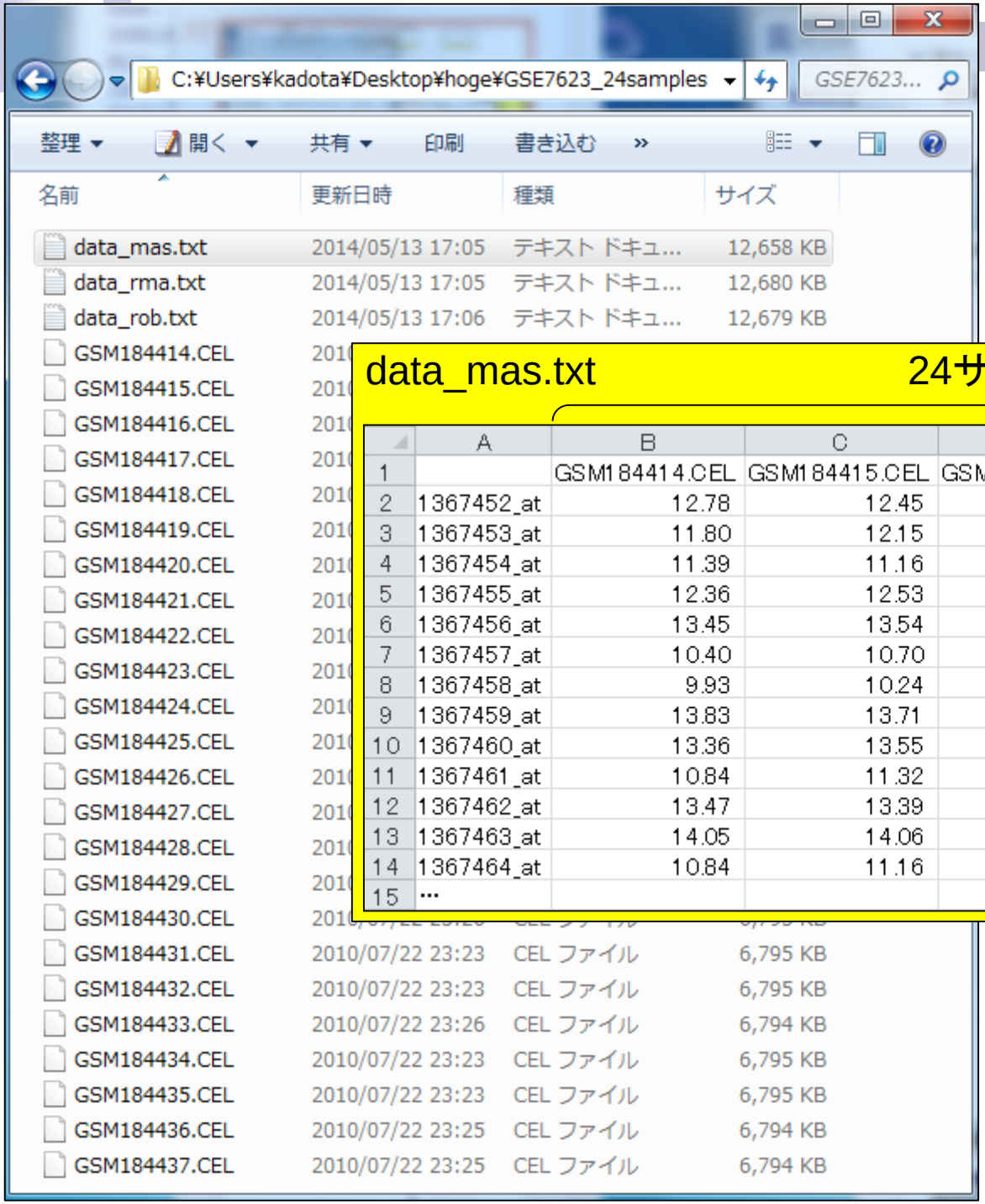
```
###############
###  MAS5  ###
###############
out_f <- "data_mas.txt"              #出力ファイル名を指定してout_fに格納
library(affy)                        #パッケージの読み込み
hoge <- ReadAffy()                   #*.CELファイルの読み込み
eset <- mas5(hoge)                   #MASを実行し、結果をesetに保存
summary(exprs(eset))                 #得られたesetの遺伝子発現行列のシグナル
exprs(eset)[exprs(eset) < 1] <- 1    #対数変換（log2）できるようにシグナル強
summary(exprs(eset))                 #上記処理後のシグナル強度分布を再び表示
exprs(eset) <- log(exprs(eset), 2)   #底を2として対数変換
write.exprs(eset, file=out_f)        #結果を指定したファイル名で保存

###############
###  RMA  ###
###############
out_f <- "data_rma.txt"              #出力ファイル名を指定してout_fに格納
library(affy)                        #パッケージの読み込み
hoge <- ReadAffy()                   #*.CELファイルの読み込み
eset <- rma(hoge)                    #RMAを実行し、結果をesetに保存
write.exprs(eset, file=out_f)        #結果を指定したファイル名で保存

############################
###  RMX (RobLoxBioC)  ###
############################
out_f <- "data_rob.txt"              #出力ファイル名を指定してout_fに格納
library(RobLoxBioC)                  #パッケージの読み込み
hoge <- ReadAffy()                   #*.CELファイルの読み込み
eset <- robloxbioc(hoge)             #rmxを実行し、結果をesetに保存
summary(exprs(eset))                 #得られたesetの遺伝子発現行列のシグナル強度分布を表
exprs(eset)[exprs(eset) < 1] <- 1    #対数変換（log2）できるようにシグナル強度が1未満のも
summary(exprs(eset))                 #上記処理後のシグナル強度分布を再び表示させて確認
exprs(eset) <- log(exprs(eset), 2)   #底を2として対数変換
write.exprs(eset, file=out_f)        #結果を指定したファイル名で保存
```



メモ帳やワードパッドなどのテキストエディタを開いて、出力ファイル名などを適宜変更した一連のコードをファイルとして保存しています。プログラムの実行時間は7~8分程度。



data_mas.txt

24サンプル（24列）

	A	B	C	D	E	F	G
1		GSM184414.CEL	GSM184415.CEL	GSM184416.CEL	GSM184417.CEL	GSM184418.CEL	…
2	1367452_at	12.78	12.45	12.81	12.30	12.59	
3	1367453_at	11.80	12.15	11.94	11.97	11.85	
4	1367454_at	11.39	11.16	11.15	11.21	11.54	
5	1367455_at	12.36	12.53	12.43	12.60	12.44	
6	1367456_at	13.45	13.54	13.55	13.63	13.37	
7	1367457_at	10.40	10.70	10.48	10.46	10.14	
8	1367458_at	9.93	10.24	9.97	9.96	8.70	
9	1367459_at	13.83	13.71	13.95	13.70	13.77	
10	1367460_at	13.36	13.55	13.48	13.43	13.54	
11	1367461_at	10.84	11.32	10.98	11.05	10.40	
12	1367462_at	13.47	13.39	13.44	13.43	13.37	
13	1367463_at	14.05	14.06	14.10	13.86	13.79	
14	1367464_at	10.84	11.16	11.09	11.34	11.46	
15	…						

31099 probesetIDs（31099行）

MAS正規化法同士の結果を比較

GSE7623_24samples　　24サンプル（24列）

	A	B	C	D	E	F	G
1		GSM184414.CEL	GSM184415.CEL	GSM184416.CEL	GSM184417.CEL		
2	1367452_at	12.78	12.45	12.81	12.30		
3	1367453_at	11.80	12.15	11.94	11.97		
4	1367454_at	11.39	11.16	11.15	11.21		
5	1367455_at	12.36	12.53	12.43	12.60		
6	1367456_at	13.45	13.54	13.55	13.63		
7	1367457_at	10.40	10.70	10.48	10.46		
8	1367458_at	9.93	10.24	9.97	9.96		
9	1367459_at	13.83	13.71	13.95	13.70		
10	1367460_at	13.36	13.55	13.48	13.43		
11	1367461_at	10.84	11.32	10.98	11.05		
12	1367462_at	13.47	13.39	13.44	13.43		
13	1367463_at	14.05	14.06	14.10	13.86		
14	1367464_at	10.84	11.16	11.09	11.34		
15	…						

GSE7623_02samples　2サンプル（2列）

	A	B	C
1		GSM184414.CEL	GSM184415.CEL
2	1367452_at	12.78	12.45
3	1367453_at	11.80	12.15
4	1367454_at	11.39	11.16
5	1367455_at	12.36	12.53
6	1367456_at	13.45	13.54
7	1367457_at	10.40	10.70
8	1367458_at	9.93	10.24
9	1367459_at	13.83	13.71
10	1367460_at	13.36	13.55
11	1367461_at	10.84	11.32
12	1367462_at	13.47	13.39
13	1367463_at	14.05	14.06
14	1367464_at	10.84	11.16
15	…		

MASはアレイごとに独立して前処理を実行するので（サンプル数の増減にかかわらず）同一サンプル間で得られる数値情報は不変。

RMA正規化法同士の結果を比較

GSE7623_24samples　　24サンプル（24列）

	A	B	C	D	E	F	G
1		GSM184414.CEL	GSM184415.CEL	GSM184416.CEL	GSM184417.CEL		
2	1367452_at	10.52	10.23	10.35	10.11		
3	1367453_at	9.66	10.05	9.90	9.82		
4	1367454_at	9.65	9.40	9.44	9.45		
5	1367455_at	10.76	11.10	10.82	11.03		
6	1367456_at	11.71	11.60	11.59	11.49		
7	1367457_at	8.96	8.77	8.74	8.74		
8	1367458_at	8.28	8.55	8.66	8.43		
9	1367459_at	11.81	11.66	11.70	11.52		
10	1367460_at	11.63	11.62	11.48	11.51		
11	1367461_at	9.38	9.47	9.41	9.33		
12	1367462_at	11.94	11.75	11.89	11.66		
13	1367463_at	12.38	12.08	12.34	12.02		
14	1367464_at	9.48	9.61	9.63	9.58		
15	…						

GSE7623_02samples　2サンプル（2列）

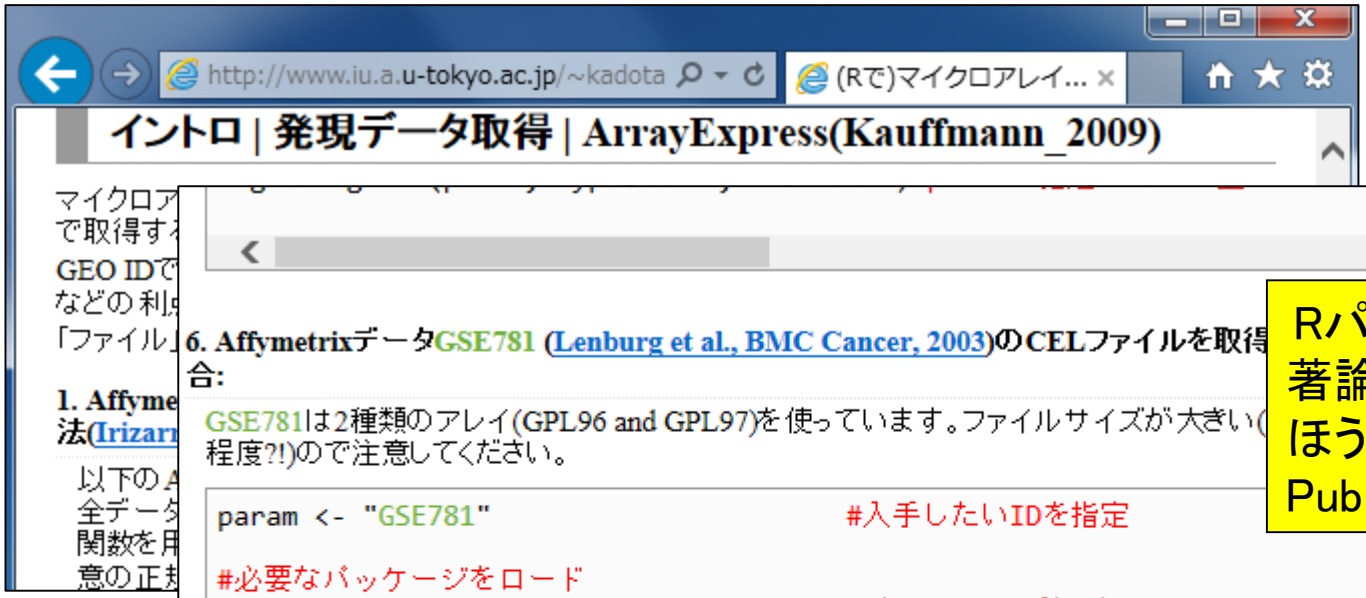
	A	B	C
1		GSM184414.CEL	GSM184415.CEL
2	1367452_at	10.53	10.22
3	1367453_at	9.91	10.17
4	1367454_at	9.68	9.54
5	1367455_at	10.69	10.85
6	1367456_at	11.60	11.51
7	1367457_at	8.89	8.93
8	1367458_at	8.17	8.47
9	1367459_at	11.98	11.73
10	1367460_at	11.60	11.76
11	1367461_at	9.02	9.11
12	1367462_at	11.88	11.68
13	1367463_at	12.41	12.20
14	1367464_at	9.43	9.62
15	…		

課題1：RMAは同一サンプル間で得られる数値が異なっていることがわかる。この理由を簡潔に説明せよ。
課題2：RMX（RobLoxBioC）についても同様の比較を行い、正規化の特徴について簡潔に述べよ（per-array basis or multi-array basis）。ヒントは教科書p39の表2-1。

教科書§2-2-2～§2-2-4について

- §2-2-2 データの正規化(基礎)
  - 行列データへのアクセスの基本をおさらい。列名変更。
  - summary関数やapply関数。箱ひげ図をpng形式で保存。
- §2-2-3 データの正規化(計算例)
  - MAS5前処理法を例として、警告メッセージへの対応やサブセットでの実行、プローブごとのシグナル強度の抽出、プローブ配列情報取得（GGRNAと同じような機能）。
  - 折れ線グラフの作成手順などを折りまぜながら、数式の解読が苦手なヒト向けに、重みつき平均の一種であるTukey's biweight estimator計算手順の解説を通じて、重みをつけるという概念の具現化や用いるパラメータの意味合いや感覚を述べている。また、一連の作業を繰り返して、より頑健な値を得るというひらめきやその具体的事例としてRobLoxBioCの計算例を示している。**本書の醍醐味的部分!**
- §2-2-4 データの正規化(その他)
  - RMAの改良版開発に至る背景（quantile正規化時にサンプル数の増減で結果が変わること）、およびプローブ効果、バッチ効果、トレーニングセット、リファレンス分布の例や基本的な考え方を述べている。また、refRMA, frozen RMA, IRON, frmaTools周辺の比較的最近提唱された方法の特徴についても述べている。

原著論文の引用はお忘れなく



Rパッケージやプログラムの多くは原著論文が存在する。各項目の最後のほうにRパッケージとその原著論文のPubMedへのリンクを張ってあります。