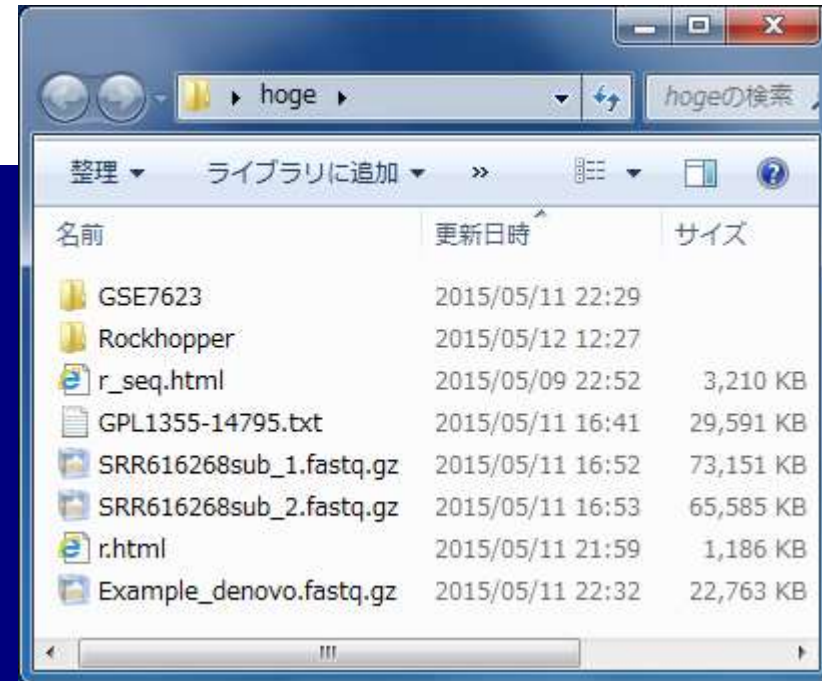


USBメモリ中のhogeフォルダをデスクトップにコピーしておいてください。

機能ゲノム学 第1回



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プログラム
門田幸二(かどた こうじ)
kadota@iu.a.u-tokyo.ac.jp
<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

講義予定

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。しかし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。本科目では、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。

■ 第1回(2015年5月12日)

- 原理、各種データベース、生データ取得
- 教科書の1.2節、2.2節周辺

■ 第2回(2015年5月19日)

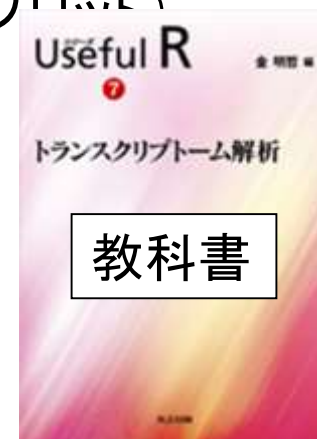
- 遺伝子発現行列作成(データ正規化)
- クラスタリング(データ変換や距離の定義など)、実験デザイン、分布
- 教科書の3.2節周辺

■ 第3回(2015年5月26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(2015年6月9日)

- 機能解析(Gene Ontology解析やパスウェイ解析)、分類など



数式の感覚を理解

ニュースレター中の教科書紹介記事(の一部)。数式が出ると混乱しがちなヒト向けに「重みつき平均」や「エントロピー」の具体的な計算手順を示しながら解説。

本書の主な存在意義は、講義や講習会の枠組みでは伝授が難しい数式の意味やアルゴリズム（問題解決のためのやり方や工夫）をじっくり学べる点であろう。なぜ二乗を含む数式にはないが三乗を含む数式には絶対値があるのか？なぜ数式の分母に非常に小さい数値が足されているのか？なぜ最大値で割るのか？なぜlogをとるのか？探索範囲を限定するとどういった結果になりうるのか？などバイ

オインフォーマティクス的なものの考え方や注意点を述べている箇所の理解がオススメポイントである。具体的には、2.2.3 (p45-62)や4.2.3 (p182-188)である。2.2.4 (p62-70)や3.2.2 (p107-111)で述べている内容も、中長期的には役立つ考え方であろう。全てが「RNA-seq」ではなく「マイクロアレイ」の項目であるが、RNA-seq部分のみ読むつもりであれば考えを改めるか購入しなくてもよいだろう。



COMPLEX ADAPTIVE TRAITS Newsletter Vol. 5 No. 7

発行：2015年3月30日

発行者：新学術研究領域「複合適応形質進化の遺伝子基盤解明」（領域代表者 長谷部光泰）

編集：COMPLEX ADAPTIVE TRAITS Newsletter 編集委員会（編集責任者 深津武馬）

領域URL：<http://staff.aist.go.jp/t-fukatsu/SGJHome.html>

Contents

- イントロダクション
 - トランスクリプトーム解析技術の原理や特徴
 - データ解析の全体像やスキル習得の意義
 - アセンブルプログラムのインストール
 - 発現データベース(DB)
 - Affymetrix GeneChip
- 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)

イントロダクション

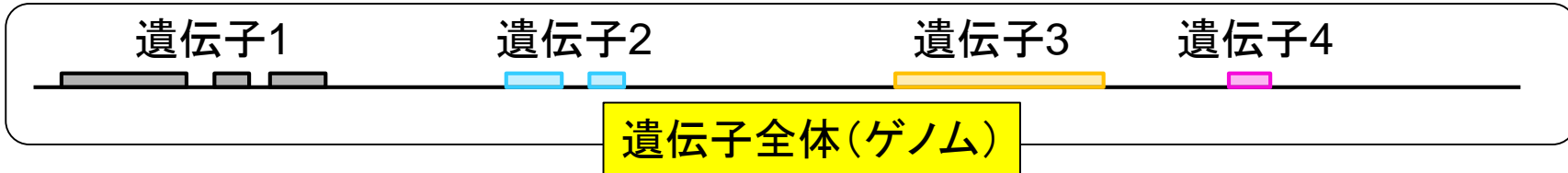
- トランスクリプトームとは
 - ある特定の状態の組織や細胞中に存在する全RNA(転写物、transcripts)の総体
- 様々なトランスクリプトーム解析技術
 - マイクロアレイ(配列既知の生物種)
 - Affymetrix GeneChipなど
 - 配列決定に基づく方法(配列未知でもよい)
 - EST、SAGE、CAGE、RNA-seqなど
 - 電気泳動に基づく方法
 - Differential Display、AFLP、HiCEPなど

調べたいサンプルでゲノム中のどの領域が、どういう時期に、どの程度転写されている(発現している)かを調べるのがトランスクリプトーム解析。「遺伝子発現解析」または単に「発現解析」は、トランスクリプトーム解析の一部。

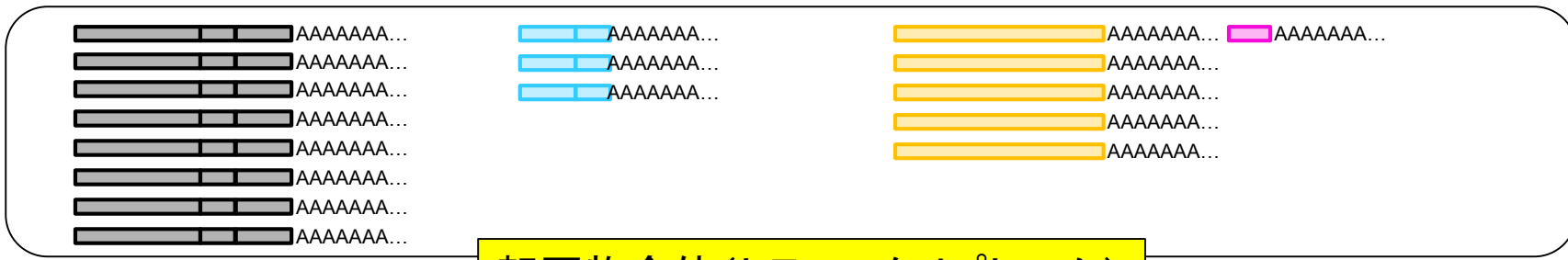
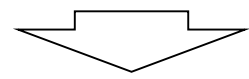
調べたい組織でどの遺伝子がどの程度発現しているのかを一度に観察

発現解析

- 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...)



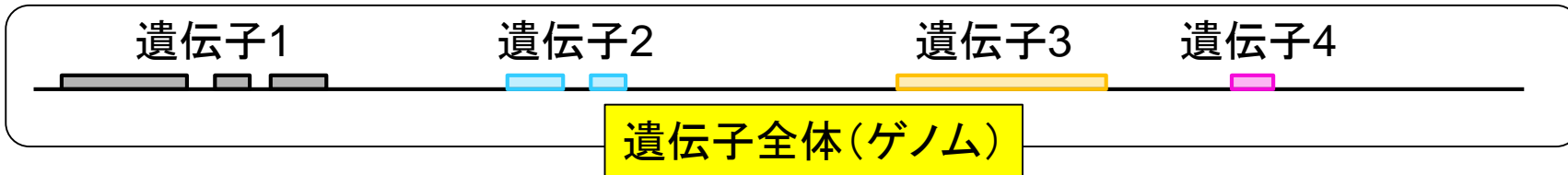
転写物全体(トランスクリプトーム)

- ・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
- ・...

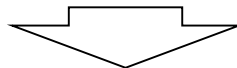
(光刺激前後など)異なる条件間での発現変動遺伝子 (Differentially Expressed Genes; DEGs)同定がよく行われる

発現解析

- 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...)



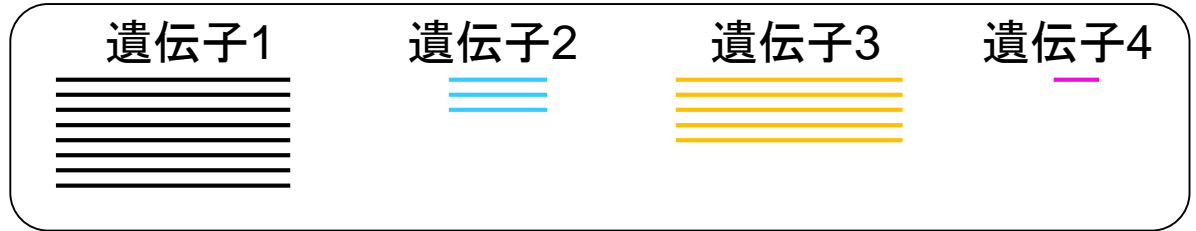
転写物全体(トランスクリプトーム)

- ・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

教科書p9の図1-8に示してあるように、実際には「遺伝子 = 転写物」とは限らない点に注意!

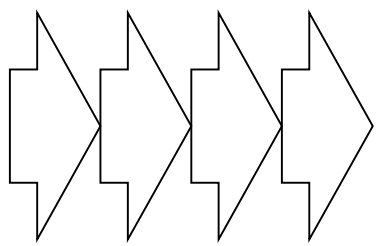
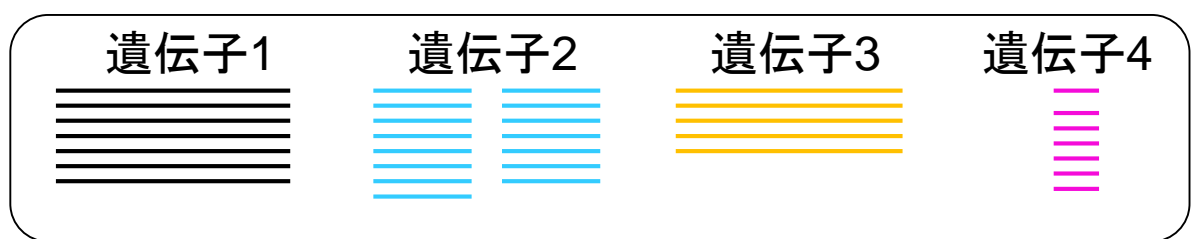
トランスクリプトーム取得

■ 光刺激前 (T1) 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」

■ 光刺激後 (T2) の目のトランスクリプトーム



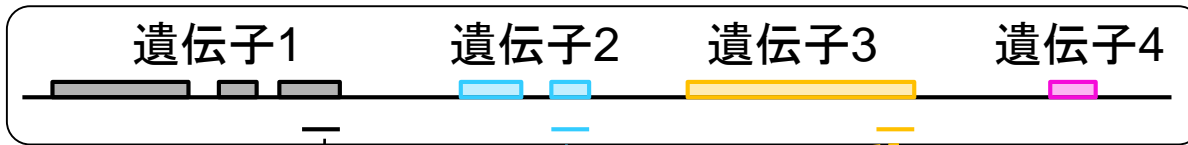
	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
...

・マイクロアレイ
・RNA-seq

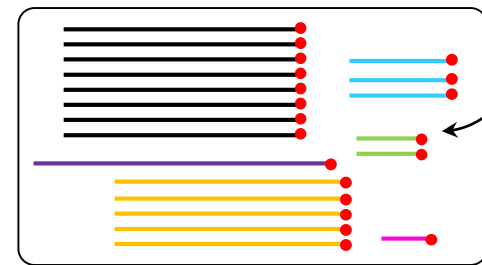
マイクロアレイ

搭載遺伝子数や種類はメーカー次第。
遺伝子4など、搭載されていない遺伝子
や未知遺伝子の発現情報は測定不可...

- 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

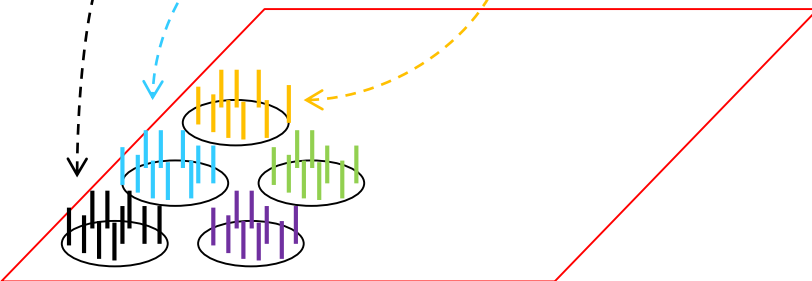
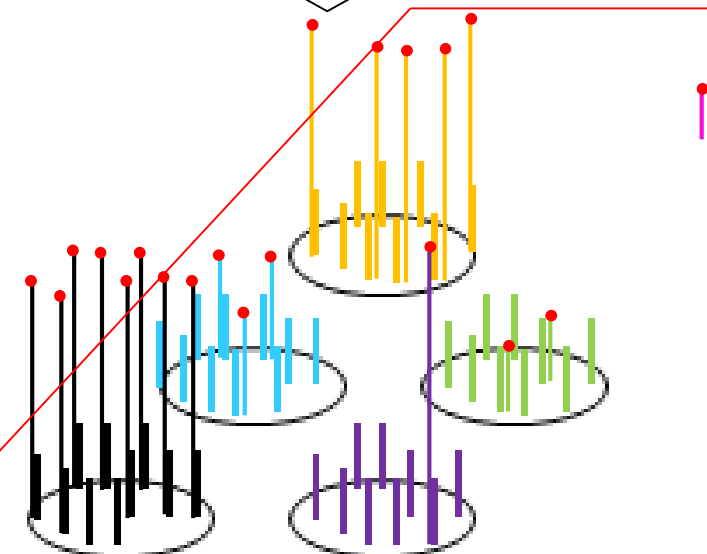
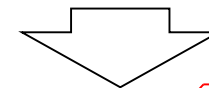


光刺激前(T1)の目の
トランスクリプトーム



蛍光
標識

ハイブリダイゼーション

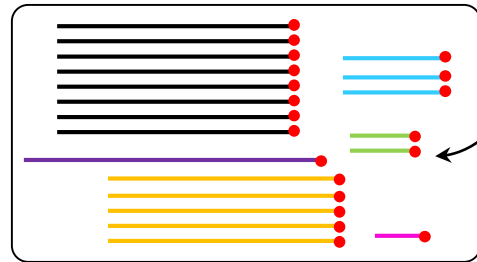


既知遺伝子(の配列の相補鎖)のプローブ
を搭載した”チップ”。12mm×12mm程度

マイクロアレイ

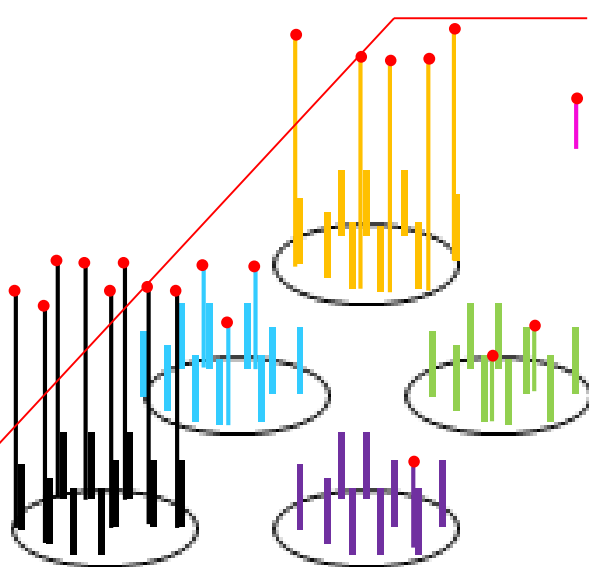
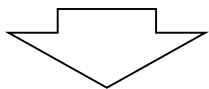
光刺激前(T1)と光刺激後(T2)の状態の数値データを比較して、サンプル(状態)間で発現変動遺伝子(DEG)を同定。

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

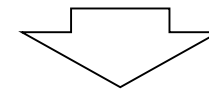
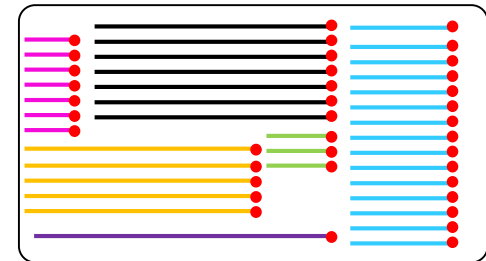


専用の検出器で各
遺伝子に対応する
領域の蛍光シグナ
ル強度を測定

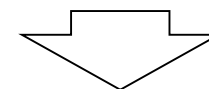


	T1
遺伝子1	8
遺伝子2	3
遺伝子3	5
遺伝子4	?
遺伝子5	...
...	...

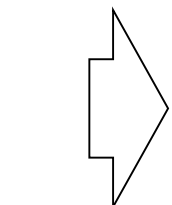
光刺激後(T2)の目の
トランスクリプトーム



ハイブリダイゼーション
とシグナル検出



T2
7
15
5
?
...
...



データ解析

光刺激前(T1)と光刺激後(T2)の状態の数値データを比較して、サンプル(状態)間で発現変動遺伝子(DEG)を同定。

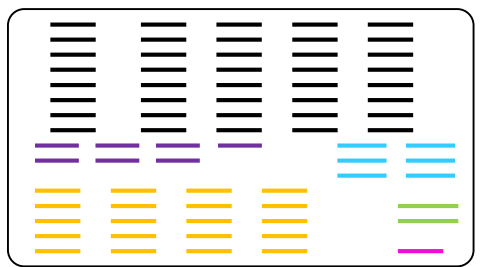
RNA-seq

■ 次世代シーケンサー(Illuminaの場合)

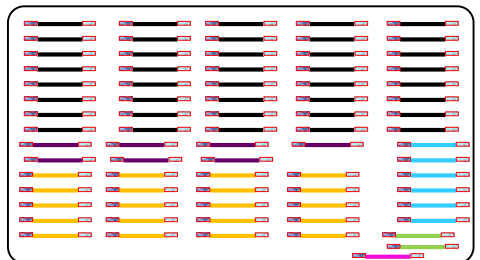
光刺激前(T1)の目のトランスクリプトーム



数百塩基程度に断片化



アダプター配列を両末端に付加



配列決定

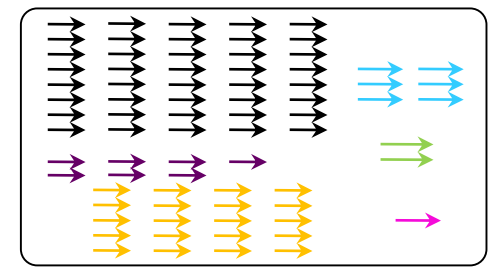
・ペアードエンド法
断片配列の両末端が数百塩基以内の対の2種類の配列が得られる



・シングルエンド法



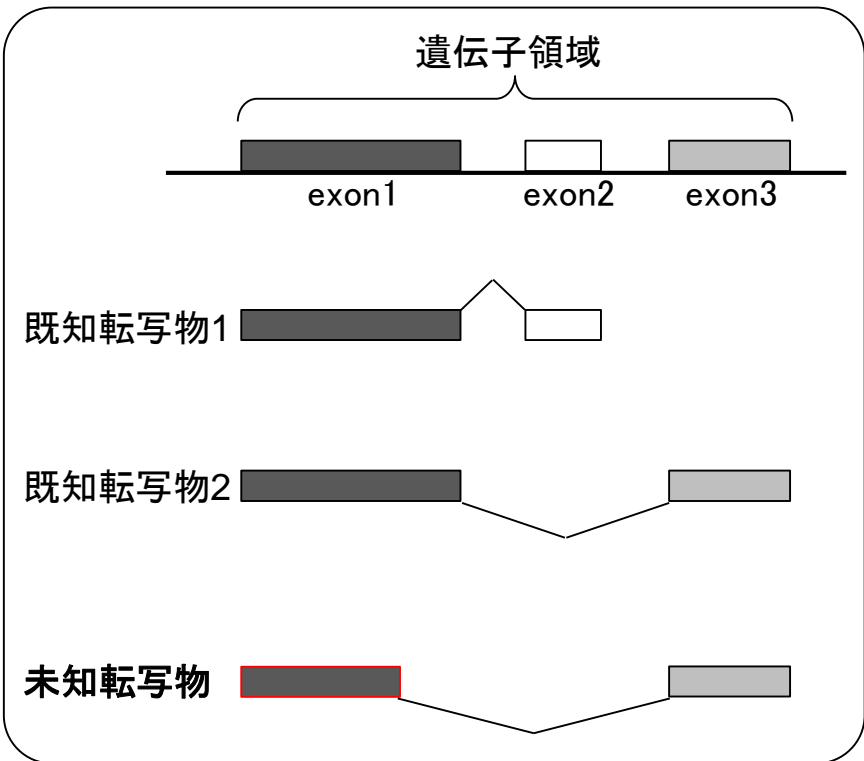
シングルエンド法の場合



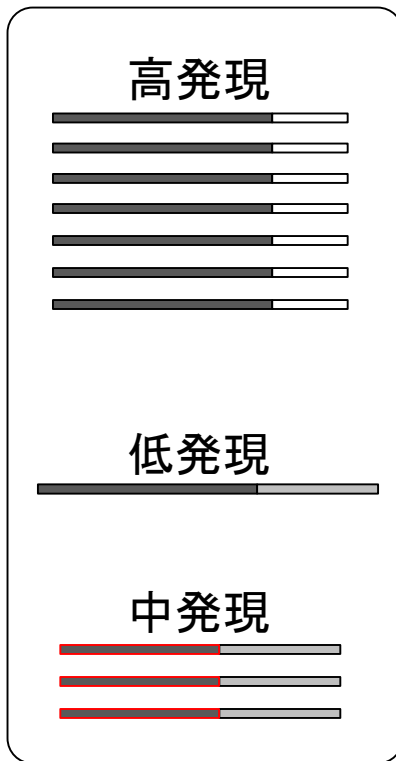
どの転写物由来か分からない塩基配列情報のみがRNA-seqによって得られる。これをもとに真の転写物の配列や発現情報を得るのがRNA-seq解析の目的。

基本イメージ

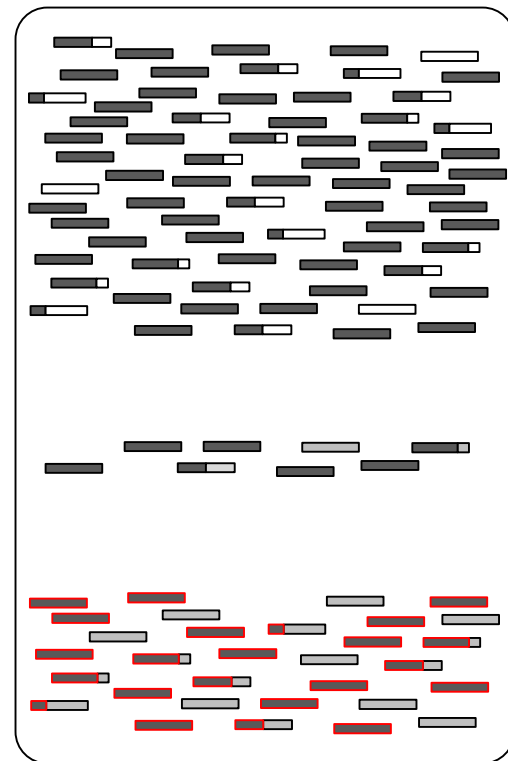
- 真の転写物情報: ある遺伝子領域中に既知転写物は2つ、未知転写物も1つ!
- 真の発現情報: 既知転写物1(高発現)、既知転写物2(低発現)、未知転写物(中発現)



真の転写物配列情報



真の発現情報

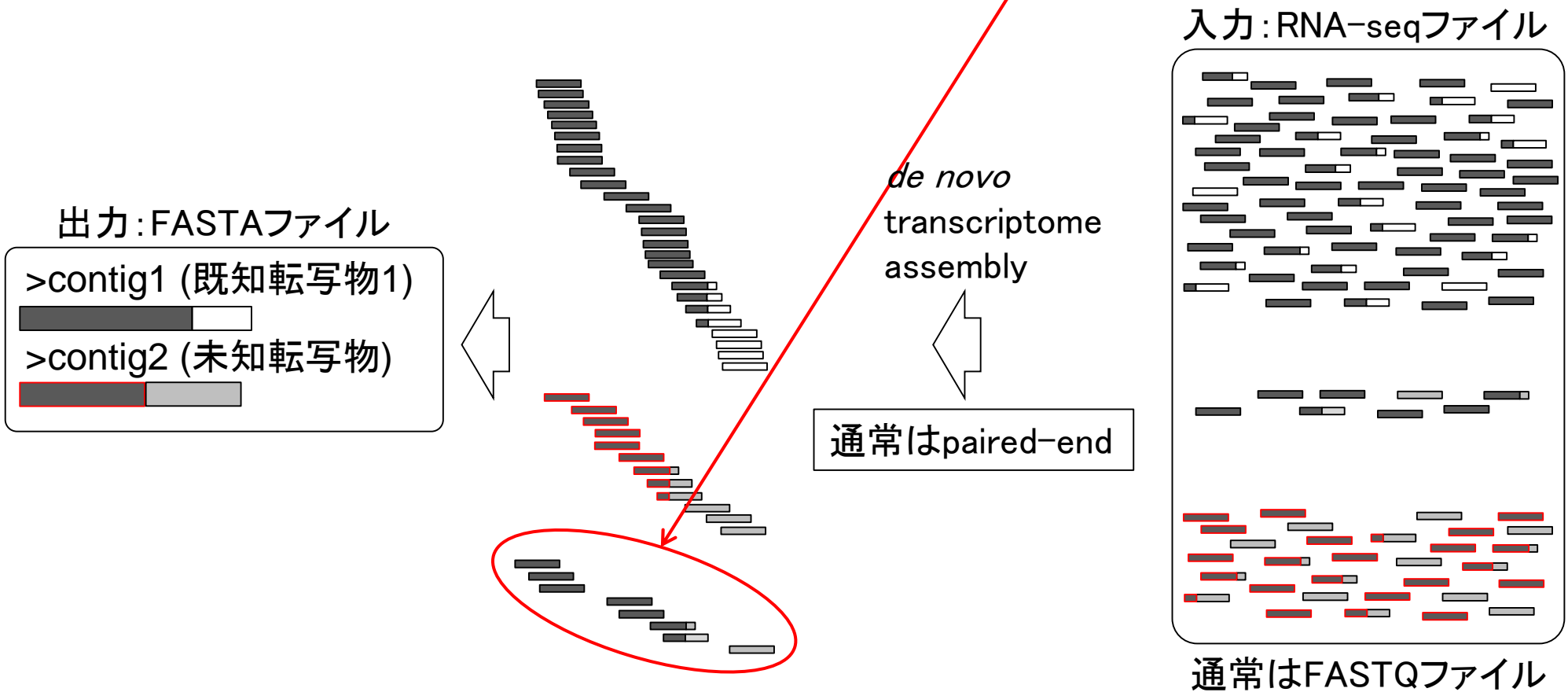


RNA-seqで得られるリード情報 (色は不明; single-endの場合)

基本イメージ

ターゲットサンプル中でそれほど発現していない転写物はアセンブルが原理的に困難。

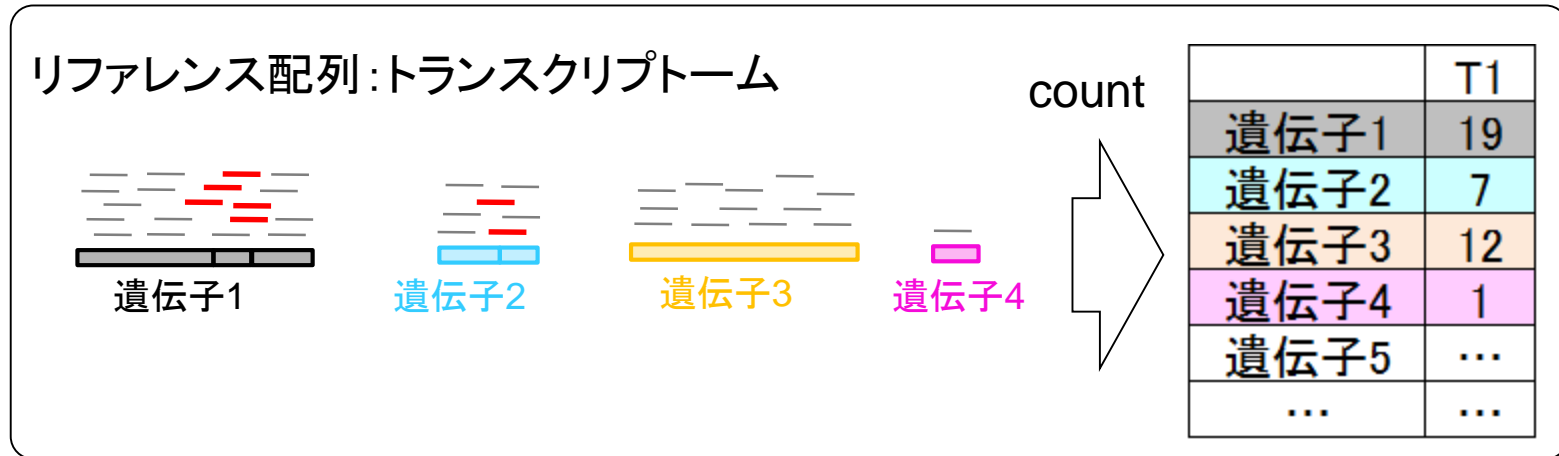
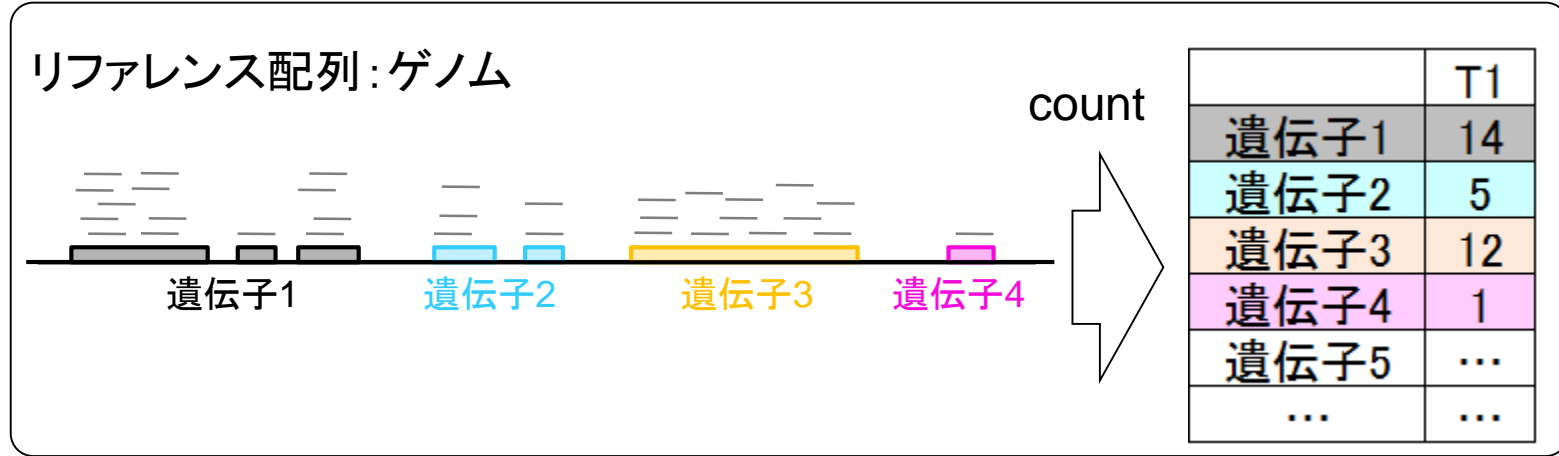
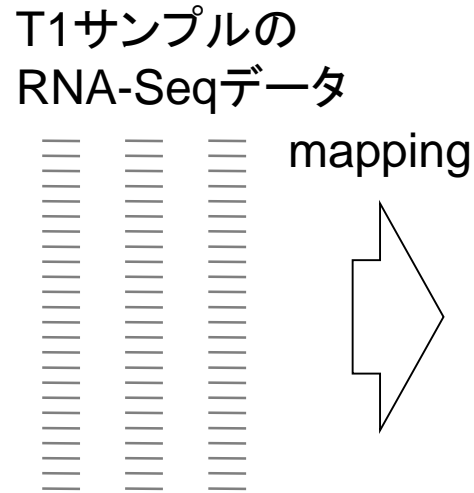
- RNA-seqデータのみしか手元にない場合：トランスクリプトーム配列取得



「マップされたリード数 = 発現量」ではないが、マップされたリード数のカウント情報は、発現量推定の基本情報です

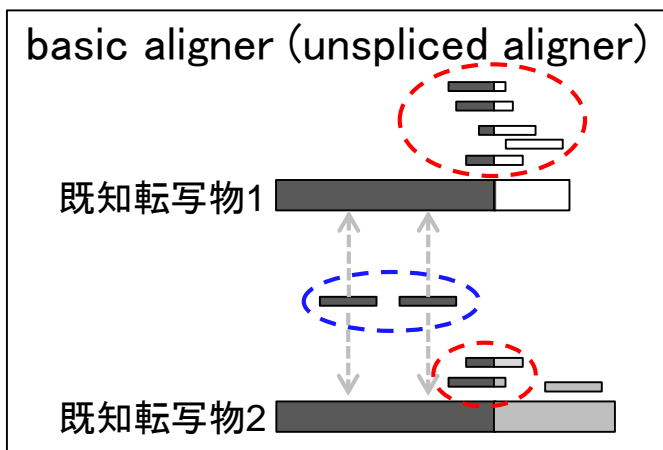
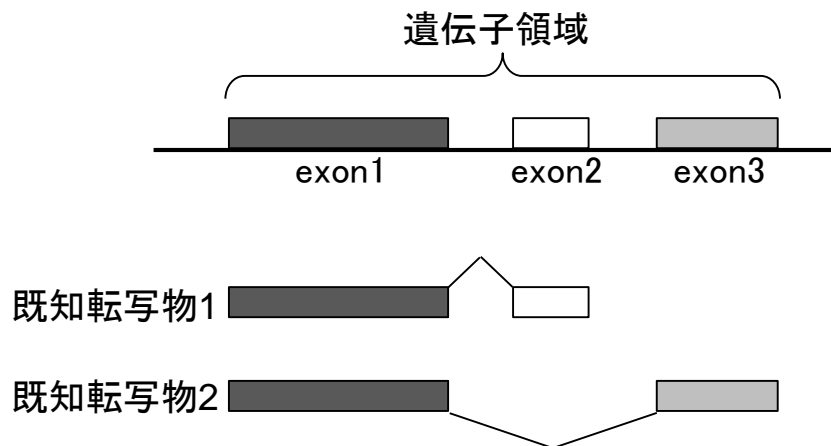
マッピングの基本イメージ

■ 基本的なマッピングプログラム (bowtieなどのbasic aligner) を用いた場合



現実には複雑...

ジャンクションリードのマッピング問題はリファレンス配列として転写物配列を採用すればいいように思えるが...。アセンブリは、キメラアセンブリ結果との戦い。転写物レベルの解像度で考えるとマップされたリードの割り振り問題など、遺伝子レベルの解像度で解析するときには考えなくてもよかった問題にぶち当たる。「高解像度の解析が原理的に可能」というポジティブなイメージの裏にはいろいろある。

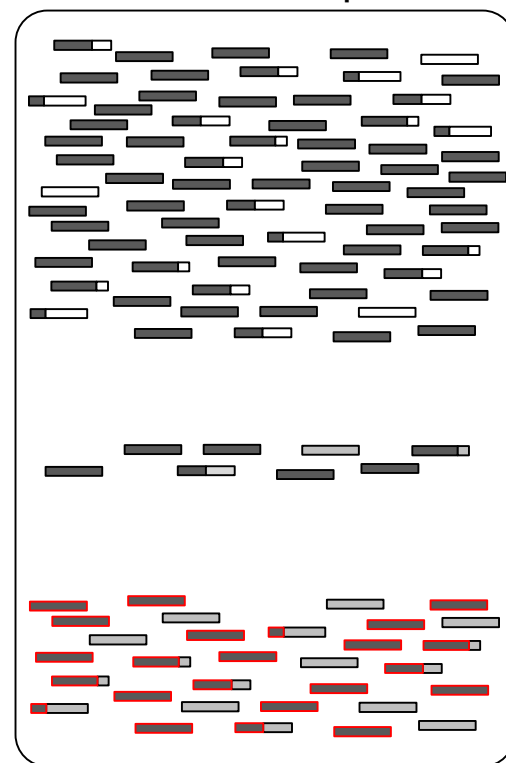


基本イメージ

リファレンス配列を利用することで
低発現転写物の遺伝子構造推定が
de novo assemblyに比べて容易に

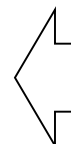
- リファレンスとしてゲノム配列が利用可能な場合：新規転写物の同定

入力1: RNA-seqファイル

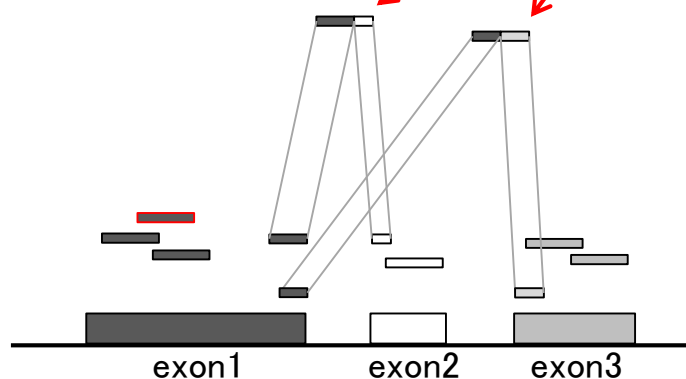


通常はFASTQファイル

マッピング

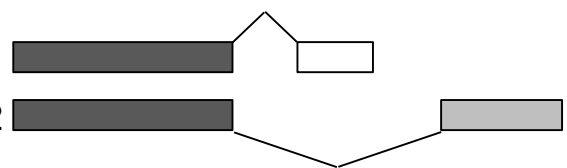


ジャンクションリード
もマッピング可能



入力2: ゲノム配列

既知転写物1
既知転写物2



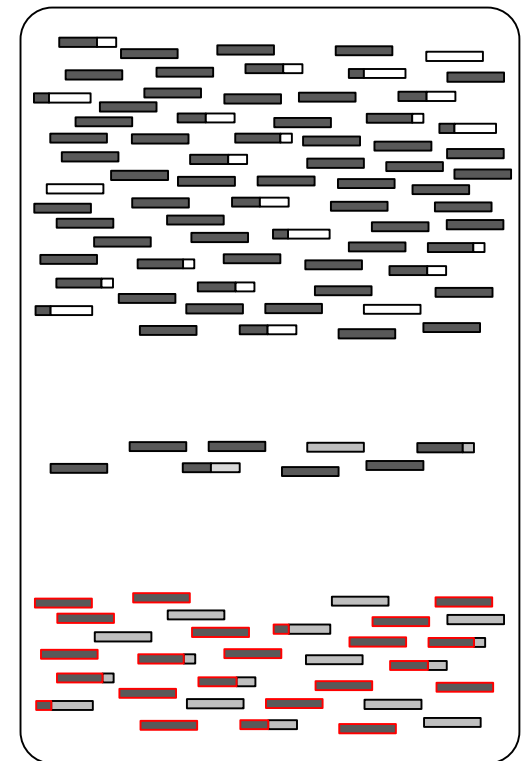
(入力3: アノテーション情報、
既知遺伝子座標情報)

基本イメージ

リファレンス配列を利用することで
低発現転写物の遺伝子構造推定が
de novo assemblyに比べて容易に

- リファレンスとしてゲノム配列が利用可能な場合：新規転写物の同定

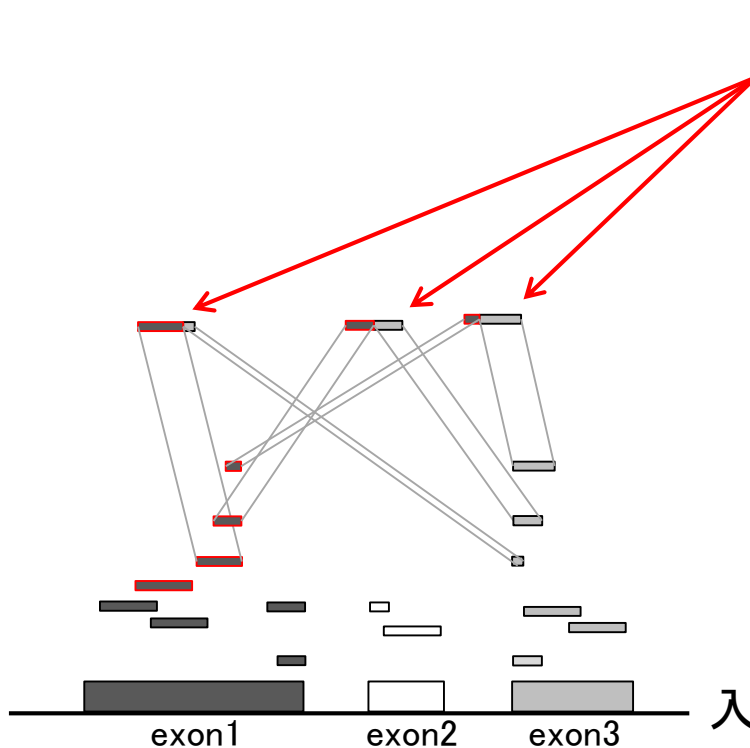
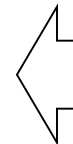
入力1: RNA-seqファイル



通常はFASTQファイル

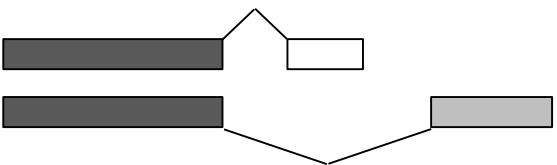
ジャンクションリード
もマッピング可能

マッピング



入力2: ゲノム配列

既知転写物1
既知転写物2

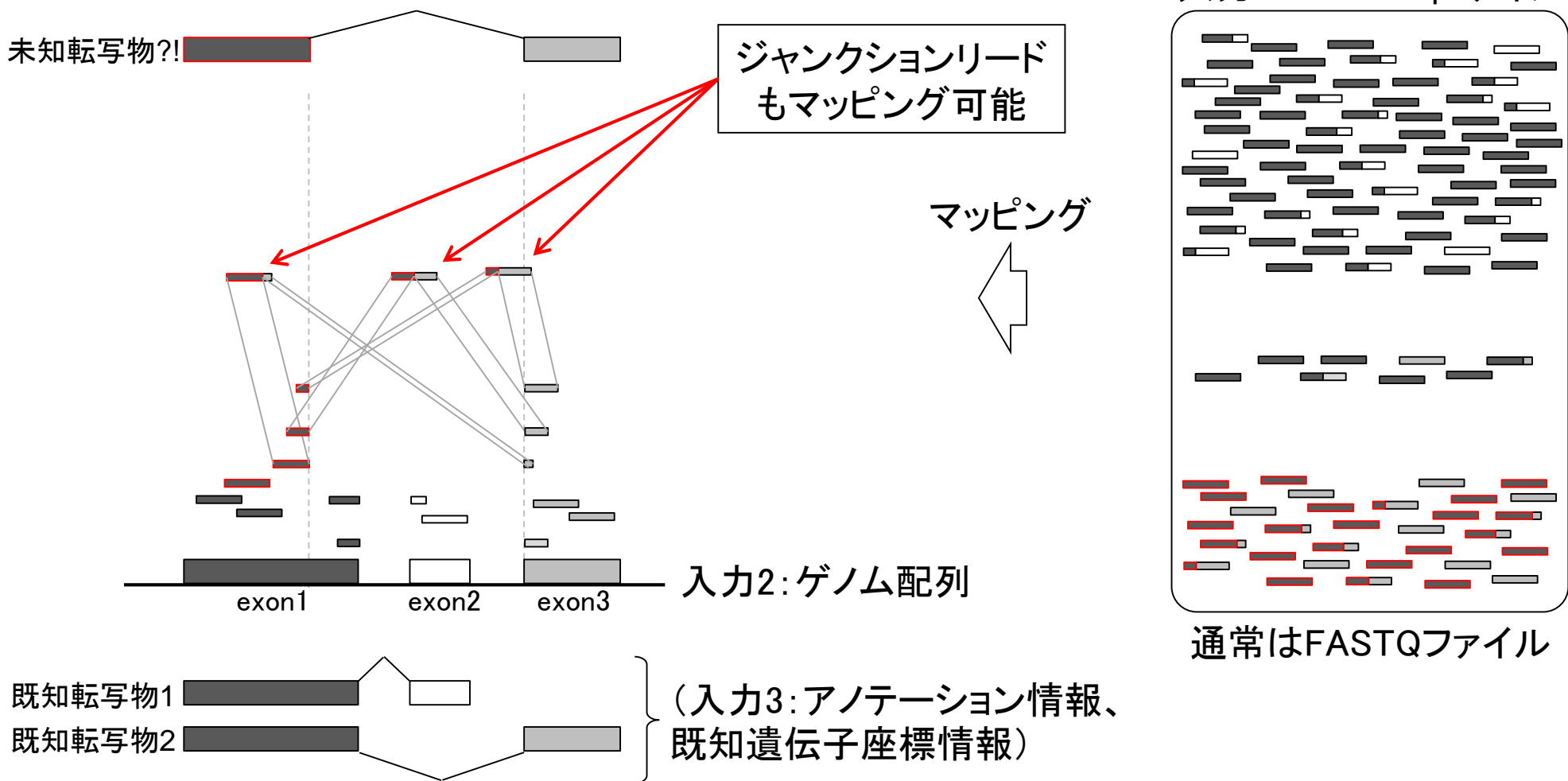


(入力3: アノテーション情報、
既知遺伝子座標情報)

基本イメージ

未知転写物同定が醍醐味。全般的な最新情報は2014年7月22日のイルミナウェビナー資料を参照のこと。

- リファレンスとしてゲノム配列が利用可能な場合：新規転写物の同定



様々な解析目的

非モデル生物を取扱う場合には、基本的にRNA-seqが有利な印象を受ける。しかし、ヒト・マウス・ラットなどのモデル生物で機能解析(Gene Ontology解析やパスウェイ解析)を行う場合には、マイクロアレイのほうがお手軽だと思います。

■ トランスクリプトーム配列取得

- RNA-seqを利用
- ゲノム配列既知の場合: 遺伝子構造推定、新規isoform同定など
- ゲノム配列未知の場合: トランスクリプトーム用アセンブラを実行

■ 遺伝子または転写物ごとの発現量の正確な推定

- 主にRNA-seq。ヒトやマウスなどのモデル生物はマイクロアレイも利用可能

■ 比較するサンプル間で発現変動している遺伝子または転写物の同定

- マイクロアレイ
 - 用いるアレイの種類(3' 発現解析用アレイ、エクソンアレイ、トランスクリプトームアレイなど)によって発現変動解析の解像度(遺伝子、exon、転写物レベルなど)が異なる。
 - アレイが提供されていない生物種の解析は不可能
- RNA-seq
 - 基本的に生物種非依存。任意のリファレンス配列(ゲノムまたはトランスクリプトーム)にリードをマップし、カウントデータ取得、統計解析。ゲノム配列がなくてもトランスクリプトーム配列をアセンブリで取得すればリファレンスとして利用可能。

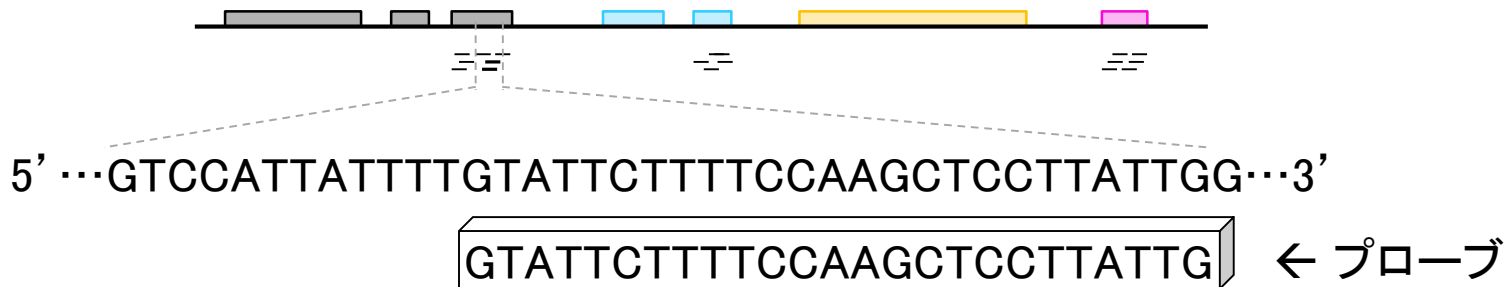
ステレオタイプなイメージ

■ 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 oxygenase-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。

解像度の高い解析が原理的に可能だが、とにかく取扱いづらい。

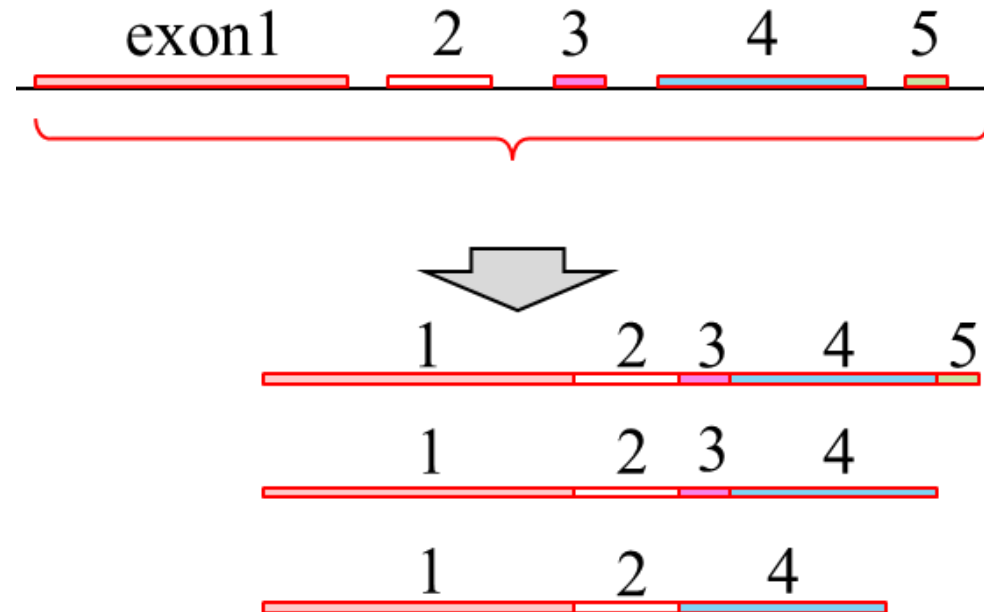
ステレオタイプなイメージ

■ RNA-seqの短所

- 取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扱い

■ RNA-seqの長所

- 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



機能解析のボトルネックは、利用する知識の解像度。RNA-seqの高解像度はオーバースペック。(個人の感想です)

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(GO)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」ではなく「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

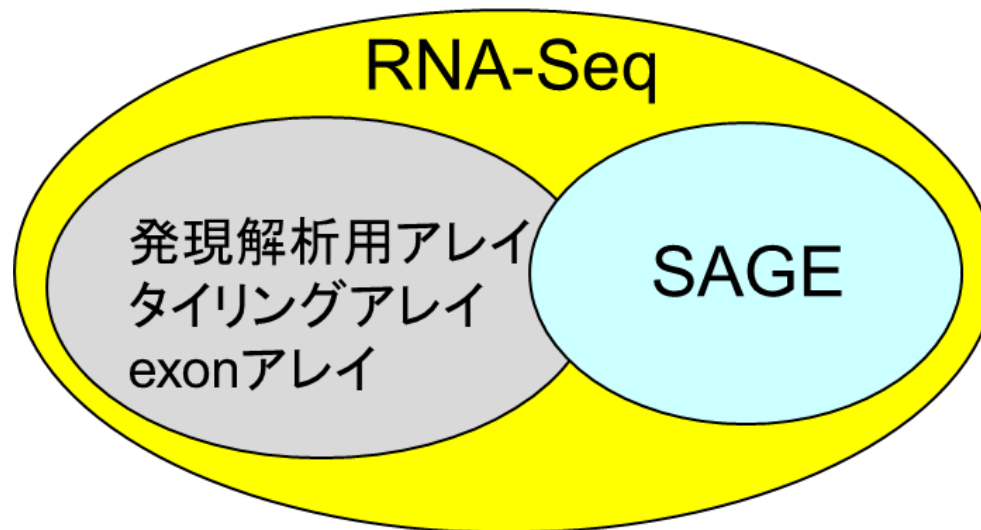
□ 異なるアレイであっても同一生物種であればマージ可能

- virtualArray (Heider and Alt, *BMC Bioinformatics*, 14:75, 2013)など

- ・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(Nutrigenomics)はアレイが未だ主流

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

Contents

- イントロダクション
 - トランスクリプトーム解析技術の原理や特徴
 - データ解析の全体像やスキル習得の意義
 - アセンブルプログラムのインストール
 - 発現データベース(DB)
 - Affymetrix GeneChip
- 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)

大まかにはこんな感じ。赤枠部分がNGS解析っぽいところ。

データ解析の全体像

	マイクロアレイ	RNA-seq
公共データ取得	GEO, ArrayExpress	GEO, ArrayExpress, NCBI SRA, EBI ENA, DDBJ SRA (DRA)
解析対象生物種	配列情報既知(アレイが提供されているもののみ)	モデル・非モデル問わず
生データ	プローブレベル数値データ	塩基配列(数億リード程度、数百塩基長)
		QC (Quality Control): クオリティチェック、フィルタリング、トリミング アセンブリでトランスクリプトーム配列取得(マッピング時のリファレンスとしても利用) マッピング(bowtie2, TopHat2など)でSAM/BAMファイル取得
発現行列作成	前処理法(MAS5, RMAなど)適用後に遺伝子発現行列を得る	アノテーションファイルを利用してカウントデータ、配列長補正後のRPKM/FPKM、転写物レベルの発現情報など取得
発現変動遺伝子(DEG)同定	基本Rを利用(limma, SAM, Rank productsなど)	基本Rを利用(cuffdiff2, edgeR, DESeq2, TCCなど)
機能解析	GSEA, GSA, Cytoscapeなど R/パッケージSeqGSEAなどを利用。	

様々な解析目的

歴史的に、塩基配列解析系プログラムのほとんどはLinux用のものが最初に提供されてきた。特にアセンブリ系はLinux用のみが基本。マッピングは(Windowsの)R環境でも数年前から可能になったが、内部的に用いるのは一世代前のLinux用プログラムの移植版で古い。

■ トランスクリプトーム配列取得

- RNA-seqを利用
- ゲノム配列既知の場合: 遺伝子構造推定、新規isoform同定など
- ゲノム配列未知の場合: トランスクリプトーム用アセンブラを実行

■ 遺伝子または転写物ごとの発現量の正確な推定

- 主にRNA-seq。ヒトやマウスなどのモデル生物はマイクロアレイも利用可能

■ 比較するサンプル間で発現変動している遺伝子または転写物の同定

- マイクロアレイ
 - 用いるアレイの種類(3' 発現解析用アレイ、エクソンアレイ、トランスクリプトームアレイなど)によって発現変動解析の解像度(遺伝子、exon、転写物レベルなど)が異なる。
 - アレイが提供されていない生物種の解析は不可能
- RNA-seq
 - 基本的に生物種非依存。任意のリファレンス配列(ゲノムまたはトランスクリプトーム)にリードをマップし、カウントデータ取得、統計解析。ゲノム配列がなくてもトランスクリプトーム配列をアセンブリで取得すればリファレンスとして利用可能。

比較的新しいプログラム(例:バクテリア用のアセンブラRockhopper 2)は、WindowsやMacintosh用も提供されている。


アセンブリ最新状況

(Rで)塩基配列解析

～NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モデル, バイオインフォマティクス～
(last modified 2015/05/09, since 2010)

What's new?


- このウェブページはフリーソフトRの[利用法](#)(Windows)の[書籍](#)もあります。
- パッケージの[簡所](#)を修正し、「アセンブル」の[平成27年度](#)(月)5月15日「[NGS速習](#)」(2015/04/21)の[参考資料](#)([日本語](#))
・[日本乳酸菌](#)

- 前処理 | トリミング | アダプター配列除去(応用) | [QuasR\(Gaidatzis 2014\)](#) (last modified 2014/06/13)
- 前処理 | トリミング | アダプター配列除去(応用) | [ShortRead\(Morgan 2009\)](#) (last modified 2014/06/18)推奨
- 前処理 | トリミング | [指定した末端塩基数だけ除去](#) (last modified 2013/06/15)
- [アセンブル](#) | [アセンブルについて](#) (last modified 2014/06/20)
- アセンブル | [ゲノム用](#) (last modified 2014/07/0)
- アセンブル | [トランスクリプトーム\(転写物\)用](#)  (last modified 2015/01/1)
- [マッピング](#) | [マッピングについて](#) (last modified 2015/01/1)
- マッピング | [basic aligner](#) (last modified 2014/08/08)
- マッピング | [splice-aware aligner](#) (last modified 2014/07/0)
- マッピング | [Bisulfite sequencing用](#) (last modified 2014/0)
- マッピング | [\(ESTレベルの長さの\)contig](#) (last modified 2013/06/19)
- マッピング | [基礎](#) (last modified 2013/06/19)
- マッピング | [single-end | ゲノム | basic aligner\(基礎\)](#) | [QuasR](#)
- マッピング | [single-end | ゲノム | basic aligner\(応用\)](#) | [QuasR](#)
- マッピング | [single-end | ゲノム | splice-aware aligner](#) | [QuasR](#)
- [マップ後](#) | [マップ後について](#) (last modified 2013/06/19)
- [マップ後](#) | [出力ファイル形式について](#) (last modified 2013/06/19)
- マップ後 | 出力ファイルの読み込み | [BAM形式](#) (last modified 2013/06/19)
- マップ後 | 出力ファイルの読み込み | [Bowtie形式](#) (last modified 2013/06/19)
- マップ後 | 出力ファイルの読み込み | [SOAP形式](#) (last modified 2013/06/19)

アセンブル | トランスクリプトーム(転写物)用 **NEW**

Rパッケージはおそらくありません。

プログラム:

- [Multiple-k: Surget-Groba and Montoya-Burgos, Genome Res., 2010](#)
- [Trans-ABYSS: Robertson et al., Nat Methods, 2010](#)
- [Rnnotator: Martin et al., BMC Genomics, 2010](#)
- [Trinity: Grabherr et al., Nat Biotechnol, 2011](#)
- [SOAPdenovo-trans: Luo et al., Gigascience, 2012](#)
- [Oases: Schulz et al., Bioinformatics, 2012](#)
- [EBARDenovo: Chu et al., Bioinformatics, 2013](#)
- [BRANCH: Bao et al., Bioinformatics, 2013](#)
- [IDBA-tran: Peng et al., Bioinformatics, 2013](#)
- [SOAPdenovo-Trans: Xie et al., Bioinformatics, 2014](#)
- [Rockhopper 2\(バクテリア用\): Tjaden B, Genome Biol., 2015](#) 
- [DETONATE\(RSEM-EVAL\): Li et al., Genome Biol., 2014](#)
- [Bridger: Chang et al., Genome Biol., 2015](#)
- [IFRAT: Mbandi et al., BMC Bioinformatics, 2015](#)

Rockhopper 2実行結果

①例題用ファイルを用いたde novo
アセンブル結果のサマリーファイル
を開いているところ。①206
transcriptsで、③平均長が338塩基。
④transcripts.txtがメイン。

```
summary.txt ← ①
↓
Assembling transcripts from reads in file: C:\Users\kadota\Downloads\Example_denovo.fastq.gz↓
↓
Aligning reads to assembled transcripts using file: C:\Users\kadota\Downloads\Example_denovo.fastq.gz↓
↓
Total reads in file: 360179↓
Perfectly aligned reads: 216274 60%↓ ②
↓
Total number of assembled transcripts: 203↓ ③
Average transcript length: 338↓
Median transcript length: 172↓
Total number of assembled bases: 68766↓
↓
Summary of results written to file: Rockhopper_Results/summary.txt↓ ①
Details of assembled transcripts written to file: Rockhopper_Results/transcripts.txt↓ ④
↓
FINISHED. ↓
```


Rockhopper 2実行結果

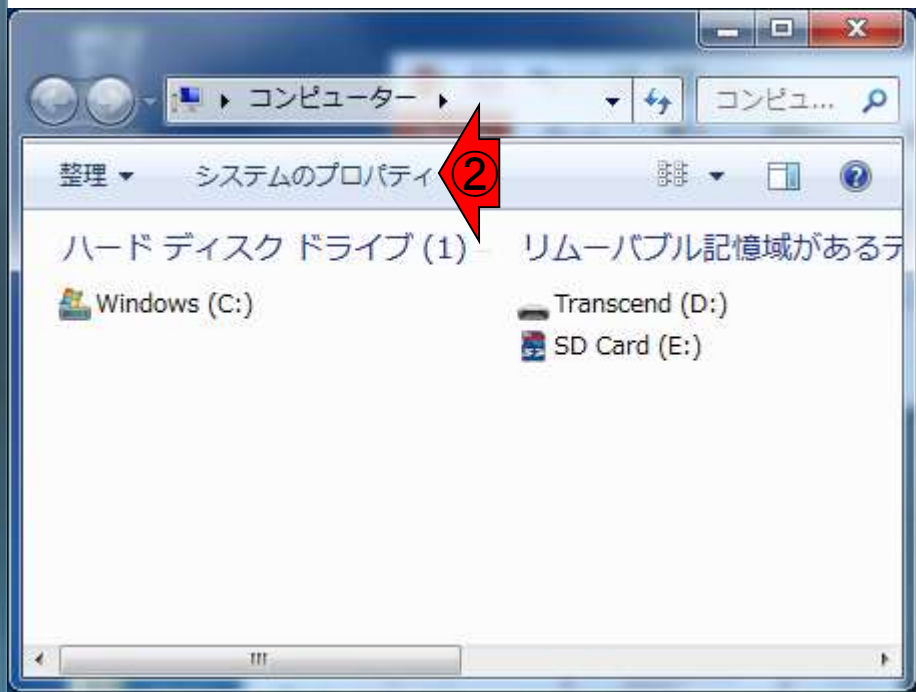
④のメインの出力結果ファイルを概観。(multi-)FASTA形式ではなく、タブ区切りテキストファイルで、⑤「塩基配列、配列長、マップされたリード数、*q*-value」の形式で表現されている。詳細はUser's Guideを。

⑤

```
transcripts.txt ×
Sequence      Length Expression 1
TAACCGCAGTGAGTGAGTCTGCAAAAAAATGAAATTGGGCAGTTGAAACCAGACGTTTTCGC
TGACCGAATATATAGTGGAGACGTTTAGATGGGTAAAATAATTGGTATCGACCTGGGTACT
GGCACCCTCCTCGCGTGTGGAGAACGCCGAAGGCGATCGCACCCAGCCTTCTATCATTGCCTATACCCAGGATGGTGAAACTCTAG
TTGGTCAGCCGGCTAAACGTCAGGCAGTGACGAACCCGCAAAACACTCTGTTTGCATTAAACGCCTGATTGGTCGCCGCTTCCAGGA
CGAAGAAGTACAGCGTGATGTTTCCATCATGCCGTTCAAAAATTATTGCTGCTGATAACGGCGACGCATGGGTGGAAGTTAAAGGCCAG
AAAAATGGCACCCGCCAGATTTCTGCTGAAGTGCTGAAAAAATGAAGAAAACCGCTGAAGATTACCTGGGTGAACCGGTA ACTGAAG
CTGTTATCACCGTACCGGCATACTTTAACGATGCTCAGCGTCAGGCAACCAGAGACGCAGGCCGTATCGCTGGTCTGGAAGTAAAACG
TATCATCAA
GGTGGTACT AAACCGTGGTTATGGGCGCAAGCTTCCGTAACATCGGCGAAATTCTGGAAGTGGCAGGCTGCGACCGTCTGACCATCGCACCCGGCACT
ACCTGGGGG GCTGAAAGAGCTGGCGGAGAGCGAAGGGGCTATCGAACGTA AACTGTCTTACACCGGCGAAGTGAAAGCGCGTCCGGCGCGTATCACT
TCCGCTGGG GAGTCCGAGTTCTGTGGCAGCACAAACCAGGATCCAATGGCAGTAGATAAACTGGCGGAAGGTATCCGTAAGTTTGCTATTGACCAGG
TACATCACT AAAAACTGGAAAAAATGATCGG 990 11485↓
ACCGTTCC TTCCATCTCTGATCGCGCATCCAGCGGCGTTTATCAGGATAAAGGCATCCCTGCGCTGGAAGAATGGCTGACATCGGCGCTAACCACG
CCGTTTGAAGTGGAAACCCGCTTAATCCC 117 151↓
GAGCAGGCGATCATCGAGCAAACGTTGTGTGAGCTGGTGGATGAAATGAGTTGCCATCTGGTGCTCACCACGGGCGGAACTGGCCCGG
CGCGTCGTGACGTAACGCCCGATGCGACGCTGGCAGTAGCGGACCGCGAGATGCCTGGCTTTGGTGAACAGATGCGCCAGATCAGCCT
GCATTTTGTACCAACTGCGATCCTTTCGCGTCAGGTGGGCGTGATTCGCAAACAGGCGCTGATCCTTAACTTACCCGGTCAGCCGAAG
TCTATTAAGAGACGCTGGAAGGTGTGAAGGACGCTGAGGGTAACGTTGTGGTACACGGTATTTTTGCCAGCGTACCGTACTGCATTC
AGTTGCTGGAAGGGCCATACGTTGAAACGGC 383 380↓
CAAAACAATTAGTGGGATTCACCAATCGGCAGAACGGTGCGACCAAACTGC 51 145↓
CCAGCAGGGCAAACAGCACGGTCAGGCTAAAGAAAACGAATTGCAGAACGCGTGCGCCTTTCAGCGTGCCGAAGAACATAAACAGCGT
AAATACGCCCCACAGACCCAG 109 323↓
```

自分のPC環境を適切に把握することが重要です。

課題1: インストール準備



課題1: インストール準備

このPCの場合は以下のように
解釈: ①OSはWindows 7。②メモ
リは8GB。③64bit マシン。

The screenshot shows the Windows 7 System Information window. The address bar indicates the path: コントロールパネル > システムとセキュリティ > システム. The left sidebar contains navigation links such as 'デバイス マネージャー', 'リモートの設定', 'システムの保護', and 'システムの詳細設定'. The main content area is titled 'コンピューターの基本的な情報の表示' and includes the following information:

- Windows Edition:** Windows 7 Professional (marked with a red arrow and circled '1'). Below it, it says 'Copyright © 2009 Microsoft Corporation. All rights reserved.' and 'Service Pack 1'. A link for 'Windows 7 の新しいエディションの追加機能の取得' is also present.
- システム:**
 - 評価:** 5.8 Windows エクスペリエンス インデックス
 - プロセッサ:** Intel(R) Core(TM) i7-4500U CPU @ 1.80GHz 2.40 GHz (marked with a red arrow and circled '2').
 - 実装メモリ (RAM):** 8.00 GB (marked with a red arrow and circled '2').
 - システムの種類:** 64 ビットオペレーティング システム (marked with a red arrow and circled '3').
 - ペンとタッチ:** タッチスクリーンでは、ペン入力とタッチ入力は サポート情報 利用できません (marked with a red arrow and circled '3').

The bottom of the window shows the title 'コンピューター名、ドメインおよびワークグループの設定'.

課題1: インストール準備

このMac PCの場合は以下のように解釈:
①OSはOS X Yosemite。②メモリは16GB。
③Windowsの「システムの種類」に相当するものはないようだが、メモリが8GB以上なら64bit マシンと解釈してよい。

Finder ファイル 編集 表示 移動 ウインドウ ヘルプ

この Mac について

システム環境設定...

App Store...

最近使った項目

Finder を強制終了

スリープ

再起動...

システム終了...

門田幸二 をログアウト.

概要 ディスプレイ ストレージ サポート サービス

OS X Yosemite

バージョン 10.10.2

MacBook Pro (Retina, 13-inch, Mid 2014)

プロセッサ 3 GHz Intel Core i7

メモリ 16 GB 1600 MHz DDR3

起動ディスク Macintosh HD

グラフィックス Intel Iris 1536 MB

シリアル番号 C02NP0HTG3QT

システムレポート...

ソフトウェア・アップデート...

™ and © 1983-2015 Apple Inc. All Rights Reserved. 使用許諾契約と保証

インストール

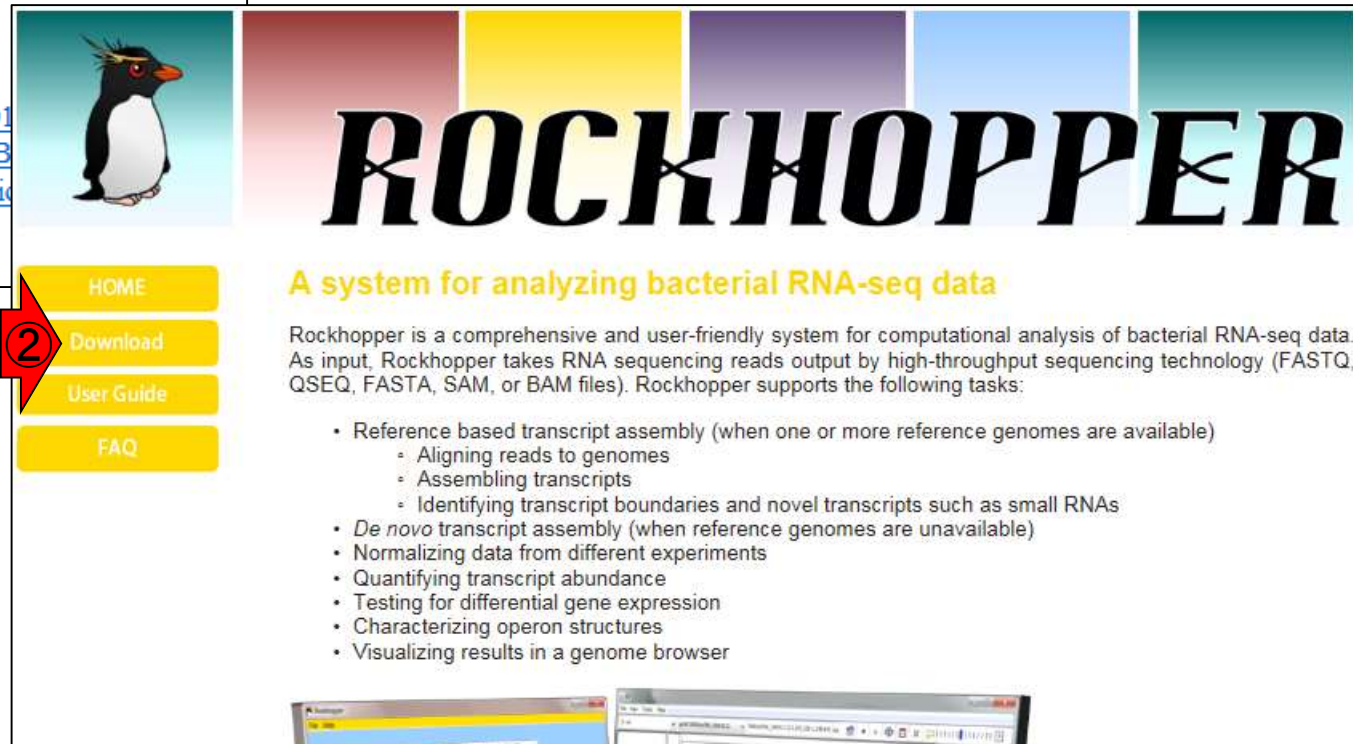
アセンブル | トランスクリプトーム(転写物)用 **NEW**

Rパッケージはおそらくありません。

プログラム:

- [Multiple-k](#): Surget-Groba and Montoya-Burgos, *Genome Res.*, 2010
- [Trans-ABYSS](#): Robertson et al., *Nat Methods*, 2010
- [Rnnotator](#): Martin et al., *BMC Genomics*, 2010
- [Trinity](#): Grabherr et al., *Nat Biotechnol.*, 2011
- [SOAPdenovo-trans](#): Luo et al., *Gigascience*, 2012
- [Oases](#): Schulz et al., *Bioinformatics*, 2012
- [EBARDenovo](#): Chu et al., *Bioinformatics*, 2013
- [BRANCH](#): Bao et al., *Bioinformatics*, 2013
- [IDBA-tran](#): Peng et al., *Bioinformatics*, 2013
- [SOAPdenovo-Trans](#): Xie et al., *Bioinformatics*, 2013
- ① [Rockhopper 2\(バクテリア用\)](#): Tjaden B, *Genome Biol.*
- [DETONATE\(RSEM-EVAL\)](#): Li et al., *Genome Biol.*
- [Bridger](#): Chang et al., *Genome Biol.*, 2015
- [IFRAT](#): Mbandi et al., *BMC Bioinformatics*, 2015

①Rockhopper 2をインストールしてみ
ましょう。②Downloadボタンを押す。



ROCKHOPPER

A system for analyzing bacterial RNA-seq data

Rockhopper is a comprehensive and user-friendly system for computational analysis of bacterial RNA-seq data. As input, Rockhopper takes RNA sequencing reads output by high-throughput sequencing technology (FASTQ, QSEQ, FASTA, SAM, or BAM files). Rockhopper supports the following tasks:

- Reference based transcript assembly (when one or more reference genomes are available)
 - Aligning reads to genomes
 - Assembling transcripts
 - Identifying transcript boundaries and novel transcripts such as small RNAs
- *De novo* transcript assembly (when reference genomes are unavailable)
- Normalizing data from different experiments
- Quantifying transcript abundance
- Testing for differential gene expression
- Characterizing operon structures
- Visualizing results in a genome browser

課題2: Javaバージョン調査

System Requirements

Rockhopper is implemented in Java, so you must have Java installed on your computer to use Rockhopper. To check if Java is installed on your computer, type "java -version" at any shell or terminal window or command prompt:

```
C:\>java -version
java version "1.7.0_11"
Java(TM) SE Runtime Environment (build 1.7.0_11-b21)
Java HotSpot(TM) 64-Bit Server VM (build 23.6-b04, mixed mode)
```

It is recommended that your computer has Java version 1.6 or later and your computer has at least 2 gigabytes of RAM. If you do not have Java or you need to update to a more recent version, you can do so by clicking the Java icon on the right:

Download Latest Release (Rockhopper version 2.0.3)

Rockhopper for Windows



Rockhopper for Mac



Rockhopper for any platform



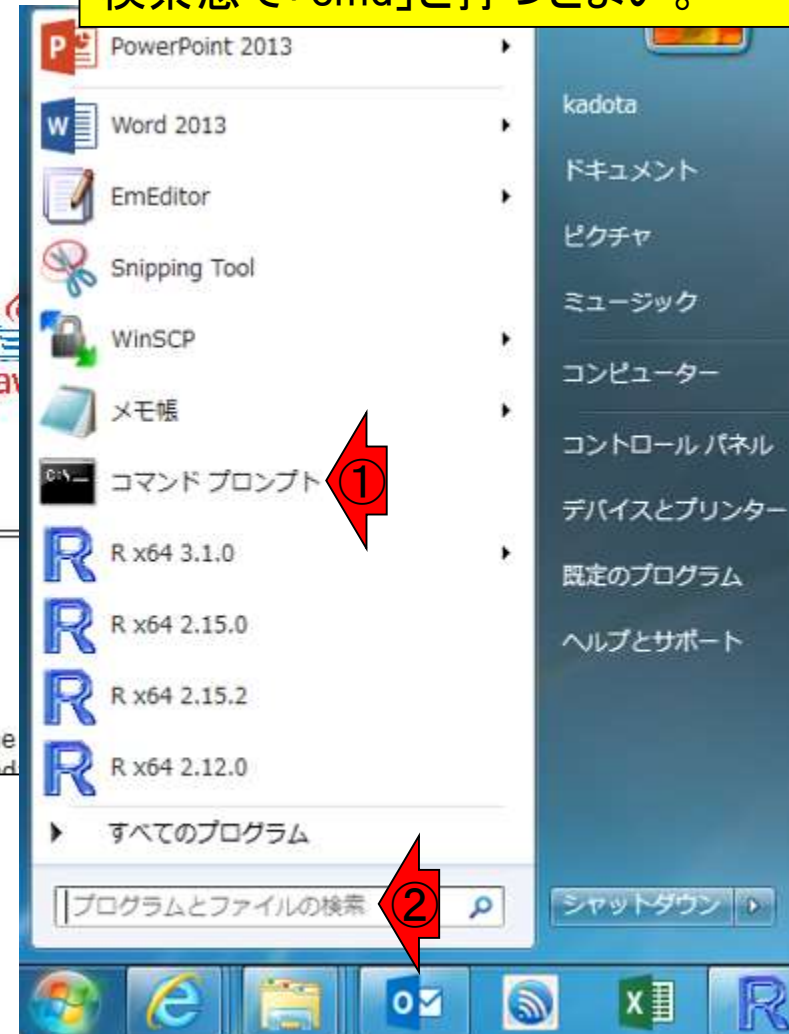
To execute the GUI

Rockhopper source code



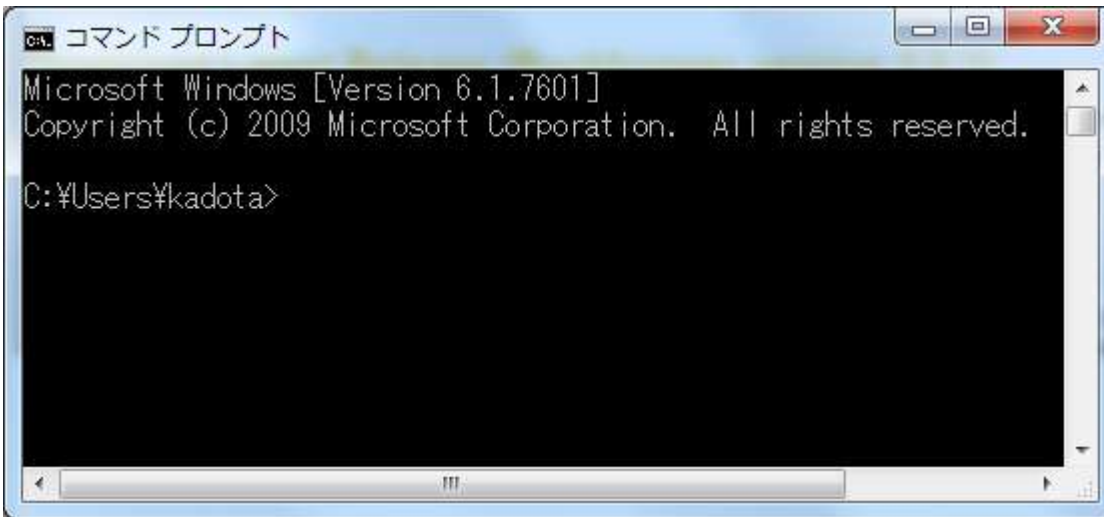
To extract the source code from the JAR file, use the following command

System Requirementとして、Java (ver. 1.6以上)が入っているかどうかを確認。Windowsの場合は、コマンドプロンプトを起動。①のようにアイコンをダブルクリックか、②検索窓で「cmd」と打つとよい。



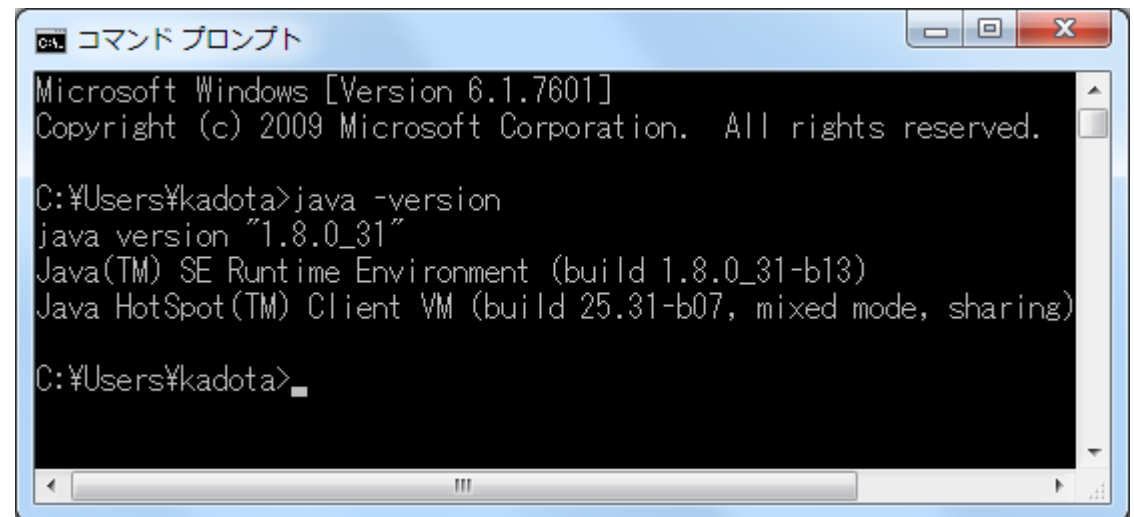
課題2: Javaバージョン調査

左上はコマンドプロンプト起動直後の状態。右下は「java -version」と打ち込んだ結果。このPCでは、「ver. 1.8.0_31」であることがわかる。



```
C:\ コマンド プロンプト
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\kadota>
```



```
C:\ コマンド プロンプト
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\kadota>java -version
java version "1.8.0_31"
Java(TM) SE Runtime Environment (build 1.8.0_31-b13)
Java HotSpot(TM) Client VM (build 25.31-b07, mixed mode, sharing)

C:\Users\kadota>
```


課題2: Javaバージョン調査

Macintoshの場合は、「アプリケーション - ①ユーティリティ - ②ターミナル.app」を起動。

The screenshot shows the 'Applications' folder in a Mac OS X environment. The left sidebar lists various categories like 'よく使う項目' (Frequently Used Items), 'デバイス' (Devices), and '共有' (Shared). The main pane displays a list of applications and folders. A red lightning bolt with the number '1' points to the 'ユーティリティ' (Utilities) folder, and another red lightning bolt with the number '2' points to the 'ターミナル.app' (Terminal.app) file.


名前	変更日	サイズ	種類
▼ ユーティリティ	2014年10月9日 23:11	--	フォルダ
AirMac ユーティリティ.app	2014年8月26日 3:17	68.8 MB	App
Audio MIDI 設定.app	2015年3月31日 19:46	7.5 MB	App
Bluetooth ファイル交換.app	2015年3月31日 19:46	1.3 MB	App
Boot Camp アシスタント.app	2015年3月31日 19:46	3.3 MB	App
ColorSync ユーティリティ.app	2015年3月31日 19:46	16.3 MB	App
Digital Color Meter.app	2014年9月10日 6:59	1 MB	App
Grapher.app	2015年3月31日 19:46	36.3 MB	App
VoiceOver ユーティリティ.app	2015年3月31日 19:46	8.3 MB	App
X11.app	2014年9月10日 6:34	181 KB	App
アクティビティモニタ.app	2015年3月31日 19:46	10.8 MB	App
キーチェーンアクセス.app	2015年3月31日 19:46	15.6 MB	App
クラブ.app	2014年6月10日 8:16	1.7 MB	App
コンソール.app	2015年3月31日 19:46	10.4 MB	App
システム情報.app	2015年3月31日 19:46	3.2 MB	App
スクリプトエディタ.app	2015年3月31日 19:46	8.4 MB	App
ターミナル.app	2015年3月31日 19:46	7.2 MB	App
ディスクユーティリティ.app	2015年3月31日 19:46	23.3 MB	App
移行アシスタント.app	2015年3月31日 19:46	2.5 MB	App

課題2: Javaバージョン調査

左上はターミナル起動直後の状態。右下は「java -version」と打ち込んだ結果。このPCには、まだJavaがインストールされていないことがわかる。

```
kadota — bash — 58x10
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$
```

```
kadota — bash — 67x12
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$ java -version
No Java runtime present, requesting install.
agribio-macbook:~ kadota$
```



“java”コマンドラインツールを使用するには、JDK をインストールする必要があります。

この Java Developer Kit のダウンロード Web サイトにアクセスするには、“詳しい情報...”をクリックしてください。


[詳しい情報...](#) [OK](#)

課題2: Javaバージョン調査

左上はターミナル起動直後の状態。右下は「java -version」と打ち込んだ結果。このPCには、まだJavaがインストールされていないことがわかる。

```
kadota — bash — 58x10
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$
```

```
kadota — bash — 67x12
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$ java -version
No Java runtime present, requesting install.
agribio-macbook:~ kadota$
```



“java”コマンドラインツールを使用するには、JDK をインストールする必要があります。

この Java Developer Kit のダウンロード Web サイトにアクセスするには、“詳しい情報...”をクリックしてください。

[詳しい情報...](#) [OK](#)

課題3: インストール本番

左側で①Windows。②Macintosh。これらを保存したものが右下の「hoge - Rockhopper」フォルダ中の実行ファイルたち。意味がわかるヒトは右下のものを実行するようにしてください。

System Requirements

Rockhopper is implemented in Java, so you must have Java installed on your computer to use Rockhopper. To check if Java is installed on your computer, type "java -version" at any shell or terminal window or command prompt:

```
C:\>java -version
java version "1.7.0_11"
Java(TM) SE Runtime Environment (build 1.7.0_11-b21)
Java HotSpot(TM) 64-Bit Server VM (build 23.6-b04, mixed mode)
```


It is recommended that your computer has Java version 1.6 or later and your computer has at least 2 gigabytes of RAM. If you do not have Java or you need to update to a more recent version, you can do so by clicking the Java icon on the right:



Download Latest Release (Rockhopper version 2.0)


①

Rockhopper for Windows




②

Rockhopper for Mac



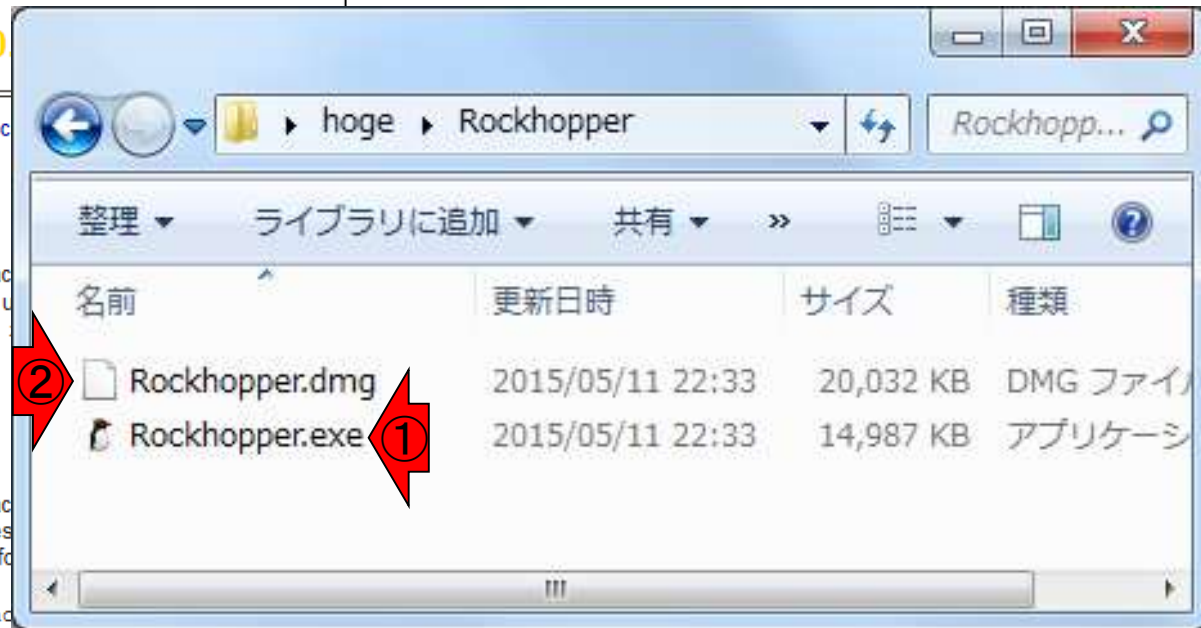
Error opening Rockhopper on Mac?

Rockhopper for any platform



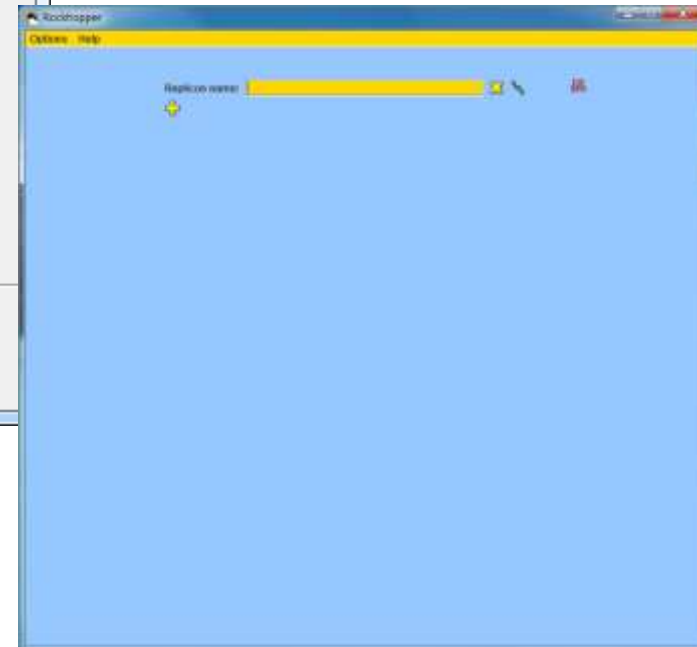
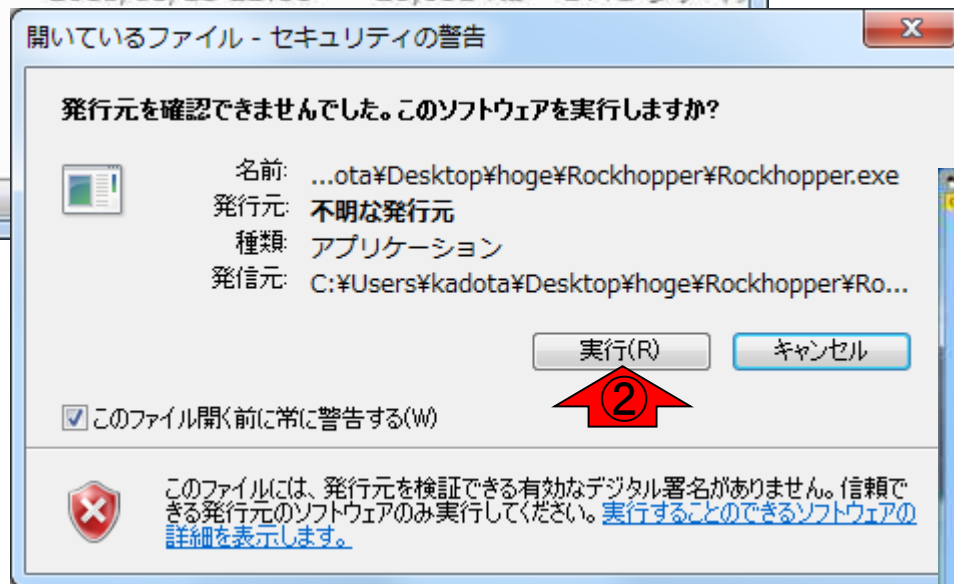
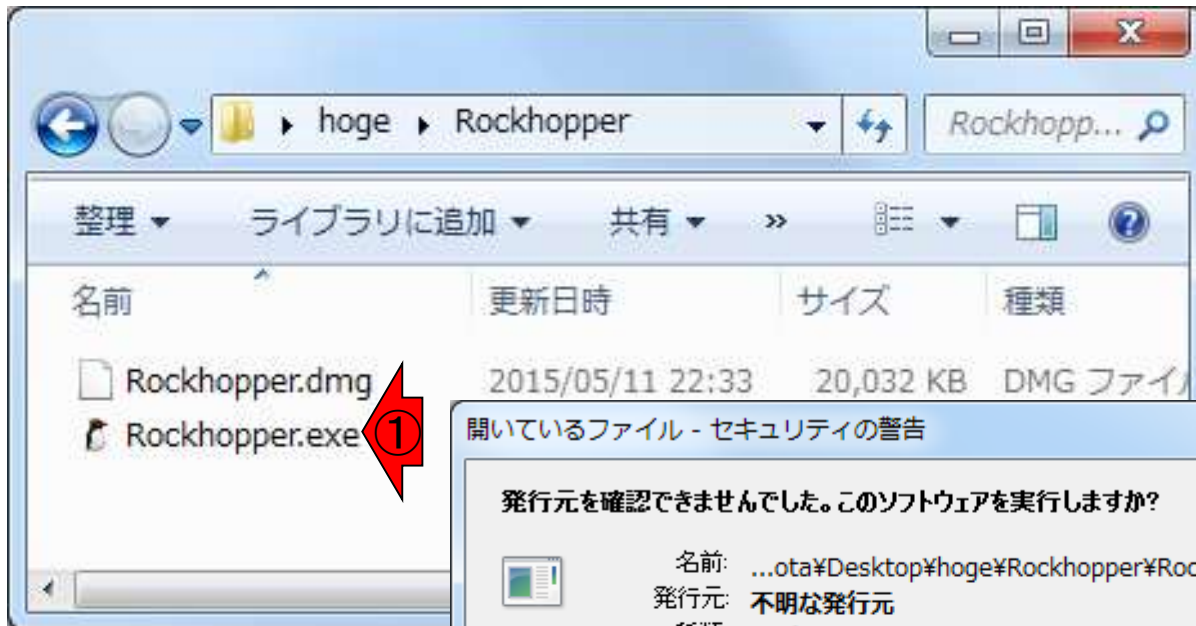
To execute the GUI version of Rockhopper, use the following command:
command:
java -Xmx1200m -jar Rockhopper.jar

To execute the command line version of Rockhopper, use the following command:
command:
java -Xmx1200m -cp Rockhopper.jar Rockhopper



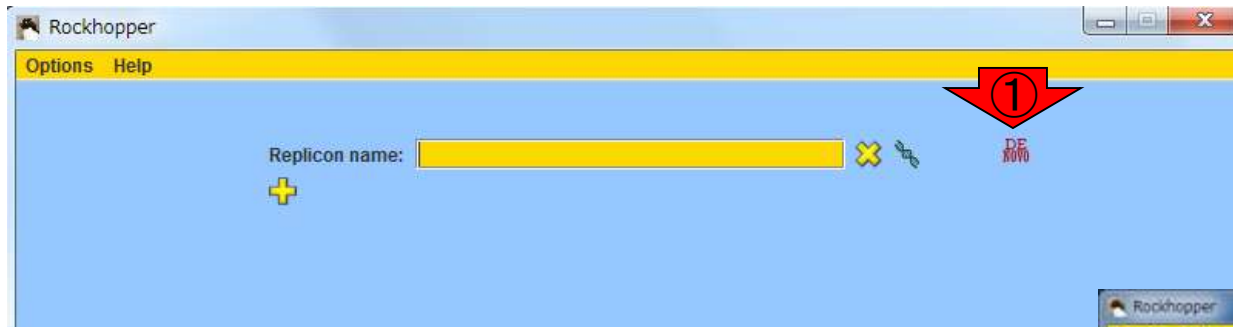
右下に示すようなGUI画面が起動するはず

課題3: インストール本番



課題3: インストール本番

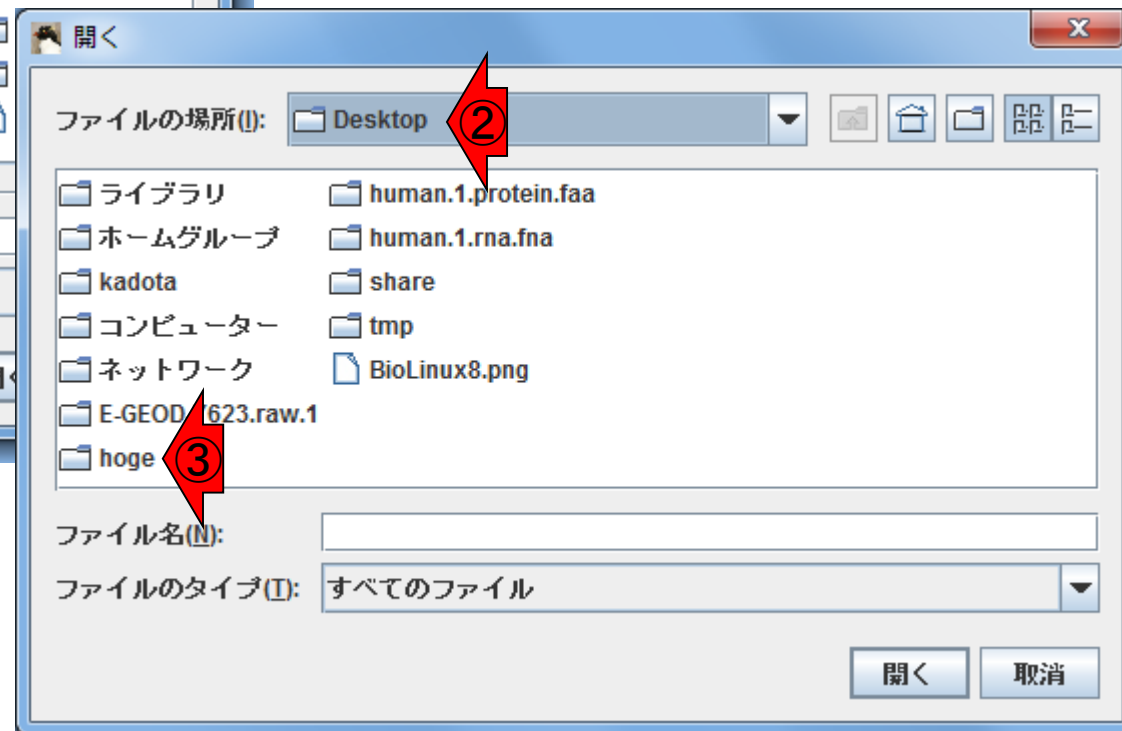
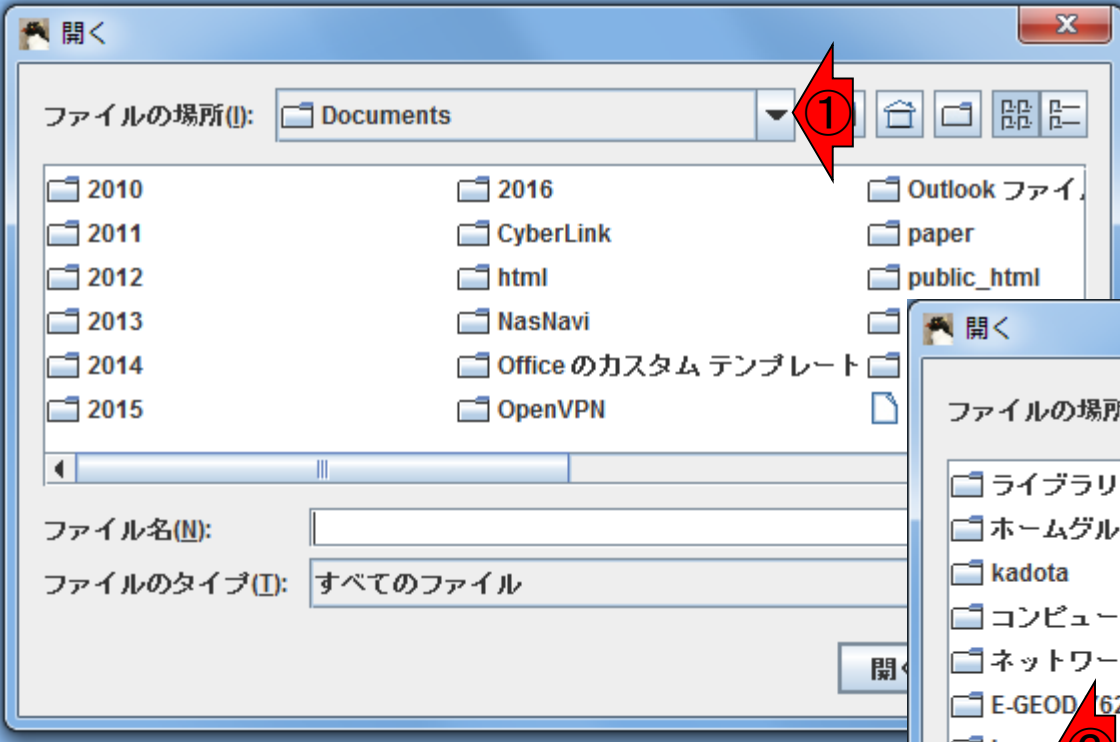
①(見づらいが)DE NOVOと赤字で書いている部分をクリック。②入力ファイルを聞かれているので例題用ファイル(Example_denovo.fastq.gz)保管場所を指定。基本は「デスクトップ - hoge」にあるはず。



課題3: インストール本番

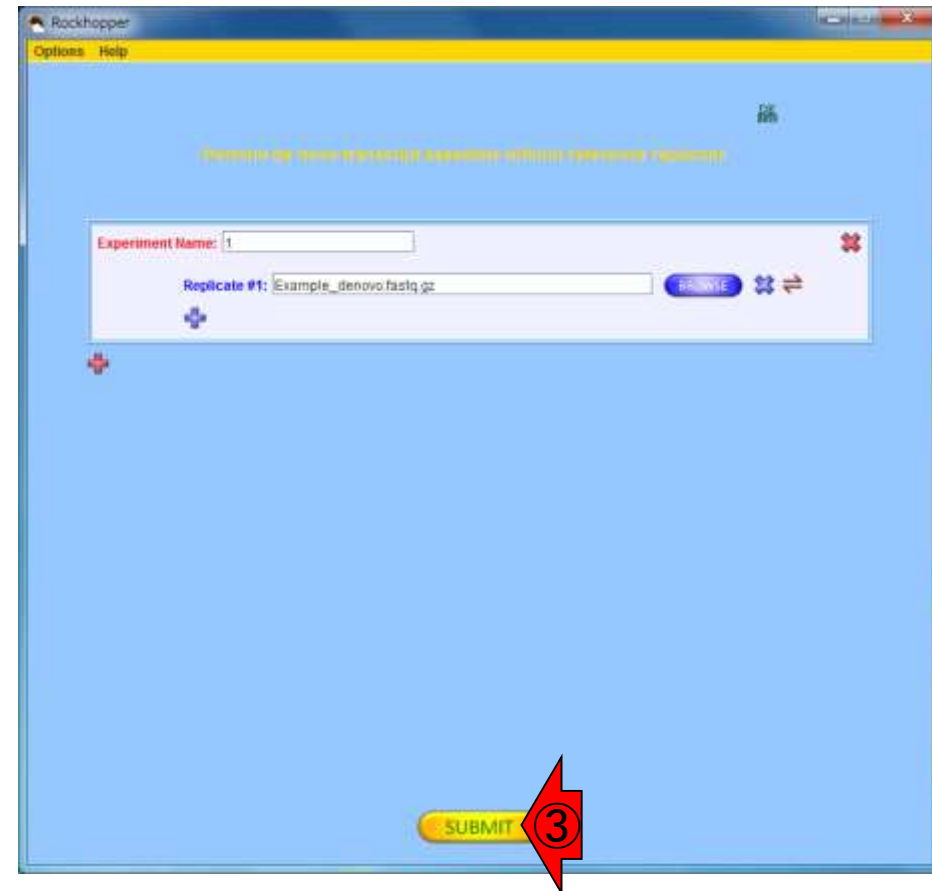
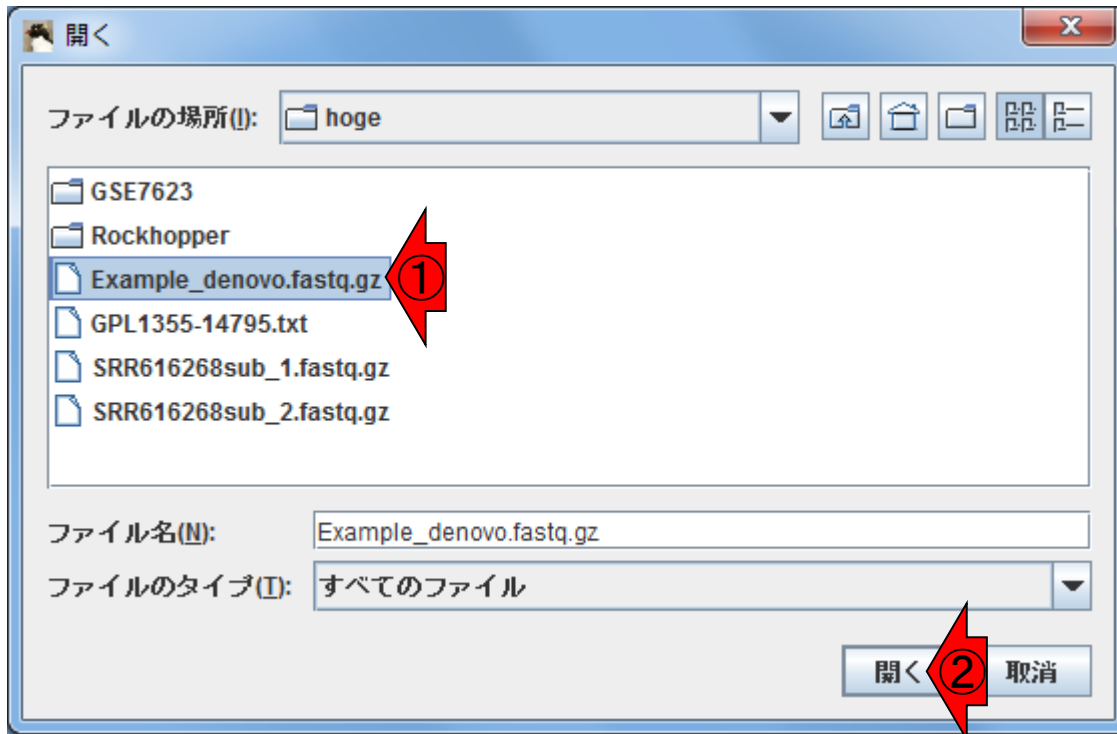
入力ファイルとして用いる例題用ファイル(Example_denovo.fastq.gz)保管場所を指定。デスクトップのhogeを指定するところまで。

Perform de novo transcript assembly without reference replicons



課題3: インストール本番

①入力ファイルとして用いる例題用ファイル(Example_denovo.fastq.gz)を選び、②開く、を押す。③SUBMITボタンを押す。



課題3: インストール本番

結果ファイルはRockhopperを起動したフォルダ(hoge - Rockhopper)中に作成される。Rockhopper_Resultsというフォルダ。

Progress

100%

Initializing RNAseq analysis...

Assembling transcripts from reads in file: C:\Users\kadota\Desktop\hoge\Example_den

Aligning reads to assembled transcripts using file: C:\Users\kadota\Desktop\hoge\Example_den

Total reads in file:	360179	
Perfectly aligned reads:	216274	60%

Total number of assembled transcripts: 203
Average transcript length: 338
Median transcript length: 172
Total number of assembled bases: 68766

Summary of results written to file: Rockhopper_Results/summary.txt
Details of assembled transcripts written to file: Rockhopper_Results/transcripts.txt

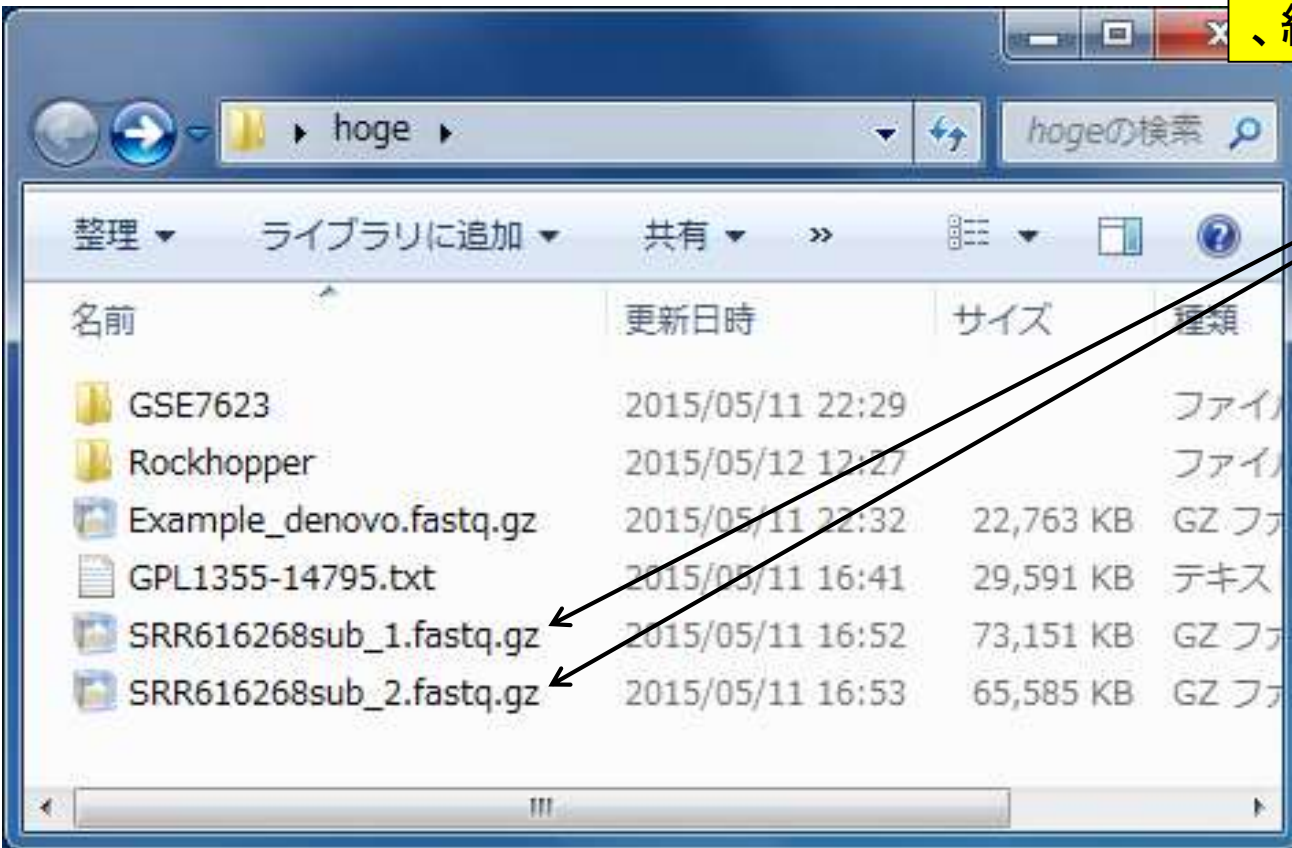
FINSIHED.

hoge > Rockhopper

名前	更新日時	サイズ	種類
Rockhopper_Results	2015/05/12 12:43		ファイル フォルダ
Rockhopper.dmg	2015/05/11 22:33	20,032 KB	DMG ファイル
Rockhopper.exe	2015/05/11 22:33	14,987 KB	アプリケーション

発展課題

乳酸菌RNA-seqデータ(100万リード)の paired-endファイルです。「Single-endのみ」と「paired-end」でアセンブルを行い、結果の違いについて考察せよ。



ファイル形式の変換

入出力形式はプログラムによって異なる。最新の(NGS解析用)プログラムを自在に使いこなしたい場合は、任意のフォーマットに自在に変換できるよう、得意なプログラミング言語のスキルを伸ばしておくとい。発展課題: 下記のような形式変換プログラムをRなど任意のプログラミング言語で作成せよ。

```
transcripts.txt x
AAACCGTGGTTATGGGCGCAAGCTTCCGTAACATCGGCGAAATTCTGGAAGTGC
GCTGAAAAGAGCTGGCGGAGAGCGAAGGGGCTATCGAACGTAACCTGTCTTACA
GAGTCCGAGTTCCTGTGGCAGCACAAACCAGGATCCAATGGCAGTAGATAAACTC
AAAAACTGGAAAAAATGATCGG 990 11485↓
TTCCATCTCTGATCGCGCATCCAGCGGCGTTTATCAGGATAAAGGCATCCCTGCGCTGGAAGAATGGCTGACATCGGCGCTAACCCAG
CCGTTTGAAGTGGAAACCCGCTTAATCCC 117 151↓
GAGCAGGCGATCATCGAGCAAACGTTGTGTGAGCTGGTGGATGAAATGAGTTGCCATCTGGTGCTCACCACGGGCGGAAGTGGCCCGG
CGCGTCGTGACGTAACGCCCGATGCGACGCTGGCAGTAGCGGACCGCGAGATGCCTGGCTTTGGTGAACAGATGCGCCAGATCAGCCT
GCATTTTGTACCAACTGCGATCCTTTTCGCGTCAGGTGGGCGTGATTCGCAAACAGGCGCTGATCCTTAACTTACCCGGTCAGCCGAAG
TCTATTAAGAGACGCTGGAAGGTGTGAAGGACGCTGAGGGTAACGTTGTGGTACACGGTATTTTGGCAGCGTACCGTACTGCATTC
AGTTGCTGGAAGGGCCATACGTTGAAACGGC 383 380↓
```

```
CAAAACAAT AAAAAGTGGAAAAAATGATCGG↓
CCAGCAGGG >ttranscript3:117:151↓
AAATACGCC TTCCATCTCTGATCGCGCATCCAGCGGCGTTTATCAGGATAAAGGCATCCCTGCGCTGGAAGAATGGCTGACATCGGCGCTAACCCAG
CCGTTTGAAGTGGAAACCCGCTTAATCCC↓
>ttranscript4:383:380↓
GAGCAGGCGATCATCGAGCAAACGTTGTGTGAGCTGGTGGATGAAATGAGTTGCCATCTGGTGCTCACCACGGGCGGAAGTGGCCCGG
CGCGTCGTGACGTAACGCCCGATGCGACGCTGGCAGTAGCGGACCGCGAGATGCCTGGCTTTGGTGAACAGATGCGCCAGATCAGCCT
GCATTTTGTACCAACTGCGATCCTTTTCGCGTCAGGTGGGCGTGATTCGCAAACAGGCGCTGATCCTTAACTTACCCGGTCAGCCGAAG
TCTATTAAGAGACGCTGGAAGGTGTGAAGGACGCTGAGGGTAACGTTGTGGTACACGGTATTTTGGCAGCGTACCGTACTGCATTC
AGTTGCTGGAAGGGCCATACGTTGAAACGGC↓
>ttranscript5:51:145↓
CAAAACAATTAGTGGGATTCACCAATCGGCAGAACGGTGGCACCAAACTGC↓
```

NGSハンズオン講習会

NGS講習会の一定割合をPerlやPythonのプログラミング言語が占める所以。ファイル形式変換プログラムでFASTQ → FASTAなど主要なものは、大抵どのプログラミング言語のものでも誰かが公開しているの、一から作成しなければいけない局面はほぼない。重要なのは、どこでどういう処理が行われているかの概要を把握し、必要に応じて変更するスキル。

- 7月22日(水): Bio-Linux 8とRのインストール状況確認
- 7月23日(木): Linux基礎。LinuxコマンドなどUNIXの基礎
- 7月24日(金): スクリプト言語。シェルスクリプト(アメリカエフ 服部恵美先生)
- 7月27日(月): スクリプト言語。Perl(アメリカエフ 服部恵美先生)
- 7月28日(火): スクリプト言語。**Python**(アメリカエフ 服部恵美先生)
- 7月29日(水): データ解析環境R(門田)
- 7月30日(木): データ解析環境R(門田)

- 8月3日(月): NGS解析。基礎(アメリカエフ 山口昌雄先生)
- 8月4日(火): NGS解析。ゲノムReseq、変異解析(アメリカエフ 山口昌雄先生)
- 8月5日(水): NGS解析。RNA-seq、統計解析(前半:アメリカエフ 山口昌雄先生、後半:門田)
- 8月6日(木): NGS解析。ChIP-seq(東京医科歯科大学 森岡勝樹先生)

- 8月26日(水): 予備日
- 8月27日(木): 予備日
- 8月28日(金): 予備日

NGSハンズオン講習会

多くのプログラムはLinux上で動作。Linuxの作法を一通り学び、路頭に迷うのを防ぐ。NGS解析用プログラム実行は数時間から数日レベル。シェルスクリプトで効率的に実行。

- 7月22日(水): Bio-Linux 8とRのインストール状況確認。主にPC持込者を対象。基本自習(門田)
- 7月23日(木): Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金): スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月): スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火): スクリプト言語。**Python**(アメリエフ 服部恵美先生)
- 7月29日(水): データ解析環境R(門田)
- 7月30日(木): データ解析環境R(門田)
- 8月3日(月): NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火): NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水): NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口昌雄先生、後半:門田)
- 8月6日(木): NGS解析。ChIP-seq(東京医科歯科大学 森岡勝樹先生)
- 8月26日(水): 予備日
- 8月27日(木): 予備日
- 8月28日(金): 予備日

NGSハンズオン講習会

RやJavaはプラットフォーム非依存(Win, Mac, Linuxどこでも動く)。Rは統計解析では必須。NGS利用目的の多くはRNA-seq。Rのみで一通り解析できていたマイクロアレイからの移行組も多い。

- 7月22日(水): Bio-Linux 8とRのインストール状況確認。
- 7月23日(木): Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金): スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月): スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火): スクリプト言語。Python(アメリエフ 服部恵美先生)
- 7月29日(水): データ解析環境R(門田)
- 7月30日(木): データ解析環境R(門田)
- 8月3日(月): NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火): NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水): NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口昌雄先生、後半:門田)
- 8月6日(木): NGS解析。ChIP-seq(東京医科歯科大学 森岡勝樹先生)
- 8月26日(水): 予備日
- 8月27日(木): 予備日
- 8月28日(金): 予備日

Contents

- イントロダクション
 - トランスクリプトーム解析技術の原理や特徴
 - データ解析の全体像やスキル習得の意義
 - アセンブルプログラムのインストール
 - 発現データベース(DB)
 - Affymetrix GeneChip
- 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)

発現DB



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](#)

[+ サイトマップ](#) [+ English](#)

- + 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。しかし、RNA-seqの解析にはマイクロアレイの知識を前提としています。本科目では、マイクロアレイの知識を前提として、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。

担当教員

門田幸二（東大・農・アグリバイオ / 特任准教授）

お知らせ

講義では、Rの様々なパッケージを利用します。持ち込み用PC

講義日程（平成27年度）

1. 平成27年05月12日
(Rで)塩基配列解析
(Rで)マイクロアレイデータ解析
2. 平成27年05月19日
3. 平成27年05月26日
4. 平成27年06月09日



発現DB

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

(Rで)マイクロアレイデータ解析

(last modified 2015/04/24, since 2005)

What's

・ 門田知見マイクロアレイです。(お知り料なし)

・ はじめ過去インインインインインイン

- ・ (削除予定)[Rのインストールと起動](#) (last modified 2014/05/14)
- ・ (削除予定)[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 2014/06/02)
- ・ [書籍 | について](#) (last modified 2014/05/12)
- ・ 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\(Affymetrix 3'発現アレイ\)](#) (last modified 2014/05/12)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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 2014/05/16)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ 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)
- ・ イントロ | [発現データ取得 | 公共DBから](#) (last modified 2014/05/11)
- ・ イントロ | [発現データ取得 | inSilicoDb\(Takemaru 2011\)](#) (last modified 2013/08/20)
- ・ イントロ | [発現データ取得 | ArrayExpress\(Kauffmann 2009\)](#) (last modified 2014/05/15) 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 | [発現データ取得 | 公共DB\(GEO\)から](#) (last modified 2013/08/18)



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

発現DB

イントロ | 発現データ取得 | 公共DBから

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](#)
 - [GSE7623](#)(ラット 24サンプル)
 - [GSE30533](#)(ラット 10サンプル)
 - [GSE2361](#)(ヒト 36サンプル)
 - [GSE10246](#)(マウス 182サンプル)
 - [GSE1133](#)(ヒトとマウス 438サンプル)
 - [GSE5364](#)(ヒト 341サンプル)
 - [GSE15998](#)(マウス 106サンプル)
- [ArrayExpress: Rustici et al., Nucleic Acids](#)
 - [GSE7623](#)(ラット 24サンプル)
 - [GSE30533](#)(ラット 10サンプル)
 - [GSE2361](#)(ヒト 36サンプル)
 - [GSE10246](#)(マウス 182サンプル)
 - [GSE1133](#)(リンク先なし): [Stachelscheid et al., Nucleic Acids](#)
 - [GSE5364](#)(ヒト 341サンプル)
 - [GSE15998](#)(マウス 106サンプル)

二次データベース

- [inSilico Db: Coletta et al., Genomics](#)
- [BioGPS: Wu et al., Nucleic Acids](#)
- [Expression Atlas: Petryszak et al., Nucleic Acids](#)
- [CellFinder: Stachelscheid et al., Nucleic Acids](#)

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.

Keyword or GEO Accession

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 3848
About GEO DataSets	Search GEO Documentation	Series: 57214
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 14265
About GEO2R Analysis	GEO BLAST	Samples: 1392679
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	

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

発現DB

イントロ | 発現データ取得 | 公共DBから

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](#)
 - [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](#)(ヒトとマウス 438サンプル, 1.7GB): [Su et al., Proc Natl Acad Sci U S A, 2004](#)
 - [GSE5364](#)(ヒト 341サンプル, 生データなし): [Yu et al., PLoS Genet., 2008](#)
 - [GSE15998](#)(マウス 106サンプル, 4.0GB): [原著論文はなし?!エクソナレイ](#)
- [ArrayExpress: Rustici et al., Nucleic Acids Res., 2013](#)



EMBL-EBI

Services Research Training About us

ArrayExpress

Examples: E-MEXP-31, cancer, p53, Geuvadis [Search](#) [Advanced](#)

Home Browse Submit Help About ArrayExpress [Feedback](#) [Login](#)

ArrayExpress – functional genomics data

ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

[Browse ArrayExpress](#)

Data Content
Updated today at 07:00

- 57519 experiments
- 1694733 assays
- 28.17 TB of archived data



◦ 57519 experiments
◦ 1694733 assays
◦ 28.17 TB of archived data

2つのDB間で用語の統一は
なされていないことがわかる

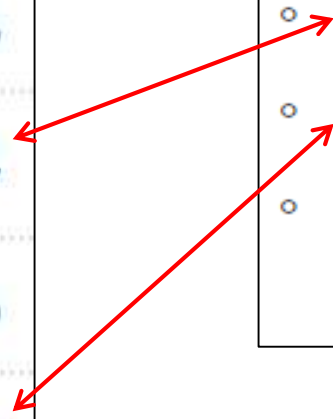
発現DB

NCBI GEO

DataSets:	3848
Series: 	57214
Platforms:	14265
Samples:	1392679

EMBL-EBI ArrayExpress

- 57519 experiments
- 1694733 assays
- 28.17 TB of archived data



データ解析の全体像

マイクロアレイ

NGS塩基配列データリポジトリがEMBL-EBI ENAという認識だが、ArrayExpressから利用可能だとする理由までは把握しきれてません。

RNA-seq

公共データ取得

GEO, ArrayExpress

GEO, ArrayExpress, NCBI SRA, EBI ENA, DDBJ SRA (DRA)

解析対象生物種

配列情報既知(アレイが提供されているもののみ)

モデル・非モデル問わず

EMBL-EBI

Services Research Training About us

生テ

ArrayExpress

Search

Examples: E-MEXP-31, cancer, p53, Geuvadis

Advanced

Home Browse Submit Help About ArrayExpress

Feedback Login

ArrayExpress – functional genomics data

ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

[Browse ArrayExpress](#)

Data Content

Updated today at 07:00

- 57519 experiments
- 1694733 assays
- 28.17 TB of archived data

Latest News

17 February 2015 - **RNA-seq expression data of many human cancer cell lines now available in ArrayExpress and Expression Atlas**

発現

発現
(DEC
機能

発現DB

イントロ | 発現データ取得 | 公共DBから

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](#)
 - [GSE7623](#)(ラット 24サンプル)
 - [GSE30533](#)(ラット 10サンプル)
 - [GSE2361](#)(ヒト 36サンプル)
 - [GSE10246](#)(マウス 182サンプル)
 - [GSE1133](#)(ヒトとマウス 438サンプル)
 - [GSE5364](#)(ヒト 341サンプル)
 - [GSE15998](#)(マウス 106サンプル)
- [ArrayExpress: Rustici et al., Nucleic Acids](#)
 - [GSE7623](#)(ラット 24サンプル)
 - [GSE30533](#)(ラット 10サンプル)
 - [GSE2361](#)(ヒト 36サンプル)
 - [GSE10246](#)(マウス 182サンプル)
 - [GSE1133](#)(リンク先なし): [Stachelscheid et al., Nucleic Acids](#)
 - [GSE5364](#)(ヒト 341サンプル)
 - [GSE15998](#)(マウス 106サンプル)

二次データベース

- [inSilico Db: Coletta et al., Genomics](#)
- [BioGPS: Wu et al., Nucleic Acids](#)
- [Expression Atlas: Petryszak et al., Nucleic Acids](#)
- [CellFinder: Stachelscheid et al., Nucleic Acids](#)

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.

Keyword or GEO Accession

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 3848
About GEO DataSets	Search GEO Documentation	Series: 57214
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 14265
About GEO2R Analysis	GEO BLAST	Samples: 1392679
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	

Platformsは、大まかにはアレイやNGS機器の種類数。

発現DB

NCBI GEO Overview

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 records are supplied by submitters

DataSets:	3848
Series:	57214
Platforms:	14265
Samples:	1392679

Platformsの例(2015年5月)

■ Affymetrix GeneChip

- Affymetrix Human Genome U133 Plus 2.0 Array: **GPL570**
 - 2003年11月リリース、54,675 probesets、105,000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、37,000枚以上
- Affymetrix Mouse Genome 430 2.0 Array: **GPL1261**
 - 2004年5月リリース、45,101 probesets、43,000枚以上
- Affymetrix Rat Genome 230 2.0 Array: **GPL1355**
 - 2004年6月リリース、31,099 probesets、17,000枚以上

■ Illumina BeadChip

- Illumina HumanHT-12 V4.0 expression beadchip: **GPL10558**
 - 2010年6月リリース、47,323 probes、33,000枚以上
- Illumina HumanHT-12 V3.0 expression beadchip: **GPL6947**
 - 2008年6月リリース、49,576 probes、20,000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、14,000枚以上

Platformsの例(2014年5月)

■ 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

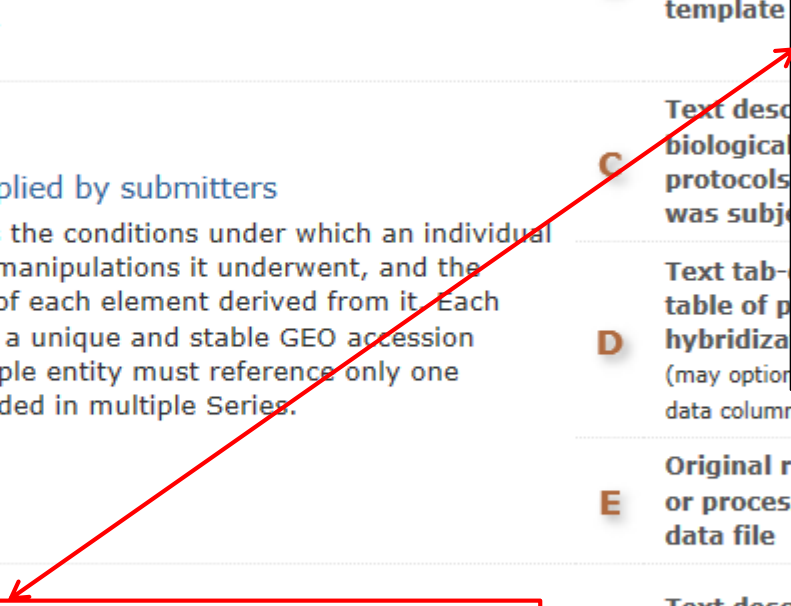
発現DB

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。

Platform	<p>Platform records are supplied by submitters</p> <p>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.</p> <p>Example Platform record »</p>
Sample	<p>Sample records are supplied by submitters</p> <p>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.</p> <p>Example Sample record »</p>
Series	<p>Series records are supplied by submitters</p> <p>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).</p> <p>Example Series record »</p>

A	Text description of the array or sequencer
B	Text table of the array template
C	Text description of the biological protocol used for the experiment
D	Text table of the probe hybridization data (may optionally include 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

DataSets:	3848
Series: 	57214
Platforms:	14265
Samples:	1392679



・ 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

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

発現DB

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

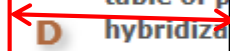
Sample records are supplied by submitters
 A Sample record describes the conditions under which an individual Sample was handled, the manipulations it underwent, and the abundance measurement of each element derived from it. Each Sample record is assigned a unique and stable GEO accession number (GSMxxx). 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 table of template
- C Text description of biological protocols was subjected to
- D Text table of hybridization (may optionally include 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

DataSets:	3848
Series: 	57214
Platforms:	14265
Samples:	1392679



Platforms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
- Affymetrix
- 2004
- Affymetrix
- 2004


■ Illumina Be

- Illumina
- 2010
- Illumina
- 2008

■ Agilent Mic

- Agilent-
- 2008

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.

Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 3848
About GEO DataSets	Search GEO Documentation	Series: 57215
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 14265
About GEO2R Analysis	GEO BLAST	Samples: 1392778
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	



Platformsの例

デフォルトはRelease dateになっている。用いられたサンプル数が多い順にソートして眺める場合は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)

Series	Samples	Platforms	DataSets	Summary	Advanced			
Search		14,265 platforms	Export	Page 1 of 714	Page size 20			
Accession	Title	Technology	Organism(s)	Data rows	Samples	Series	Contact	Release date
GPL10365	Agilent/IATS Sparus aurata oligonucleotide microarray 022077	spotted oligonucleotide	Sparus aurata	15,744	4	1	Josep A Calduch-Giner	May 09, 2015
GPL20167	Exiqon miRCURY LNA microRNA Array, 7th generation REV - hsa, mmu & rno (miRBase v18.0)	spotted oligonucleotide	Mus musculus	1,123	2	1	Yan Wang	May 09, 2015
GPL19113	Mesorhizobium array [MrNa1]	spotted oligonucleotide	Mesorhizobium sp. N33	5,774	42	1	Linda J Harris	May 08, 2015
GPL20113	CapitalBio Human LncRNA Microarray v2.0	spotted DNA/cDNA	Homo sapiens	16,839	12	1	Haixiang Wu	May 08, 2015
GPL15543	FGCVR_Arabidopsis_28K_v2.0	in situ oligonucleotide	Arabidopsis thaliana	27,834	12	1	Alberto Ferrarini	May 08, 2015
GPL20149	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [CDF: GATEExplorer_Ensembl v57]	in situ oligonucleotide	Homo sapiens	134,266	69	2	Paul Roy Heath	May 08, 2015
GPL20165	Human Extracellular Matrix and Adhesion Molecules PCR Array	RT-PCR	Homo sapiens	96	2	1	Hideo Kudo	May 08, 2015
GPL20168	Agilent-018073 Custom Brassica 60-mer two-color microarray [Probe Name version]	spotted oligonucleotide	Brassica	91,914	24	1	GEO admin	May 08, 2015
GPL20170	Illumina HiSeq 2000 (Blastomyces dermatitidis ATCC 26199)	high-throughput sequencing	Blastomyces dermatiti...				GEO	May 08, 2015
GPL20171	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [hugene10st_Hs_ENTREZG_19.0.0]	in situ oligonucleotide	Homo sapiens	19,700			GEO admin	May 08, 2015
GPL20172	Illumina MiSeq (Drosophila melanogaster; Homo sapiens)	high-throughput sequencing	Drosophila melanogas... Homo sapiens				GEO	May 08, 2015
GPL20173	Illumina HiSeq 2000 (Histoplasma capsulatum)	high-throughput sequencing	Histoplasma capsulatum				GEO	May 08, 2015
GPL17155	NimbleGen Rat RefSeq Promoter 720k	in situ oligonucleotide	Rattus norvegicus	713,735	2	2	Marcin Kruczyk	May 07, 2015



Platformsの例

(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。

Series	Samples	Platforms	DataSets	Summary	Adv			
Search		14,265 platforms	Export	Page 1 of 714	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	<i>Homo sapiens</i>	54,675	105665	3877	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	<i>Mus musculus</i>	45,101	43598	3330	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	<i>Homo sapiens</i>	22,283	37115	1024	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	47,323	33365	1000	Illumina Inc.	Jun 17, 2010
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	<i>Homo sapiens</i>	485,577	27581	407	Illumina Inc.	May 13, 2011
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Homo sapiens</i>	33,297	22489	1126	Affymetrix, Inc.	Dec 05, 2007
GPL13112	Illumina HiSeq 2000 (<i>Mus musculus</i>)	high-throughput sequencing	<i>Mus musculus</i>		21205	1467	GEO	Feb 02, 2011
GPL11154	Illumina HiSeq 2000 (<i>Homo sapiens</i>)	high-throughput sequencing	<i>Homo sapiens</i>		20519	1344	GEO	Nov 02, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	49,576	20154	432	Illumina Inc.	Jun 10, 2008
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Mus musculus</i>	35,557	17699	1432	Affymetrix, Inc.	Dec 05, 2007
GPL1355	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	<i>Rattus norvegicus</i>	31,099	17475	546	Affymetrix, Inc.	Jul 20, 2004
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	<i>Homo sapiens</i>	27,578	17009	300	Illumina Inc.	Apr 27, 2009
GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name)	in situ oligonucleotide	<i>Homo sapiens</i>	41,108	14671	619	Agilent Technologies	Feb 11, 2008

Platformsの例

①Illumina社のNGS機器であるHiSeq 2000に対して1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GPL IDが付与されているようだ。1年前は、Illumina HiSeq 2000 (*Mus musculus*)に対して、別のID (GPL18672)も割り当てられていた...

Series	Samples	Platforms	DataSets	Search	14,265 platforms			
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	<i>Homo sapiens</i>	54,675	105665	3877	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	<i>Mus musculus</i>	45,101	43598	3330	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	<i>Homo sapiens</i>	22,283	37115	1024	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	47,323	33365	1000	Illumina Inc.	Jun 17, 2010
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	<i>Homo sapiens</i>	485,577	27581	407	Illumina Inc.	May 13, 2011
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Homo sapiens</i>	33,297	22489	1126	Affymetrix, Inc.	Dec 05, 2007
GPL13112	Illumina HiSeq 2000 (<i>Mus musculus</i>)	high-throughput sequencing	<i>Mus musculus</i>		21205	1467	GEO	Feb 02, 2011
GPL11154	Illumina HiSeq 2000 (<i>Homo sapiens</i>)	high-throughput sequencing	<i>Homo sapiens</i>		20519	1344	GEO	Nov 02, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	49,576	20154	432	Illumina Inc.	Jun 10, 2008
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Mus musculus</i>	35,557	17699	1432	Affymetrix, Inc.	Dec 05, 2007
GPL1355	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	<i>Rattus norvegicus</i>	31,099	17475	546	Affymetrix, Inc.	Jul 20, 2004
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	<i>Homo sapiens</i>	27,578	17009	300	Illumina Inc.	Apr 27, 2009
GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name)	in situ oligonucleotide	<i>Homo sapiens</i>	41,108	14671	619	Agilent Technologies	Feb 11, 2008

Platformsの例

①Illumina社のNGS機器であるHiSeq 2000に対して1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GPL IDが付与されているようだ。1年前はGPL18672もIllumina HiSeq 2000 (Mus musculus)が割り当てられていた...

Series	Samples	Platforms	DataSets					
Search 14,265 platforms								
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	<i>Homo sapiens</i>	54,675	105665	3877	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	<i>Mus musculus</i>	45,101	43598	3330	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	<i>Homo sapiens</i>	22,283	37115	1024	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	47,323	33365	1000	Illumina Inc.	Jun 17, 2010
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	<i>Homo sapiens</i>	485,577	27581	407	Illumina Inc.	May 13, 2011
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Homo sapiens</i>	33,297	22489	1126	Affymetrix, Inc.	Dec 05, 2007
GPL13112	Illumina HiSeq 2000 (Mus musculus)	high-throughput sequencing	<i>Mus musculus</i>		21205	1467	GEO	Feb 02, 2011
GPL11154	Illumina HiSeq 2000 (Homo sapiens)	high-throughput sequencing	<i>Homo sapiens</i>		20519	1344	GEO	Nov 02, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	<i>Homo sapiens</i>	49,576	20154	432	Illumina Inc.	Jun 10, 2008
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Mus musculus</i>	35,557	17699	1432	Affymetrix, Inc.	Dec 05, 2007
GPL1355	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	<i>Rattus norvegicus</i>	31,099	17475	546	Affymetrix, Inc.	Jul 20, 2004
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	<i>Homo sapiens</i>	27,578	17009	300	Illumina Inc.	Apr 27, 2009
GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name	in situ oligonucleotide	<i>Homo sapiens</i>	41,108	14671	619	Agilent Technologies	Feb 11, 2008

Contents

- イントロダクション
 - トランスクリプトーム解析技術の原理や特徴
 - データ解析の全体像やスキル習得の意義
 - アセンブルプログラムのインストール
 - 発現データベース(DB)
 - Affymetrix GeneChip
- 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)

デバイスも進歩しているが、3' アレイで蓄積された過去のデータと比較しやすいため、3' アレイが今でも利用されます

Affymetrix GeneChip

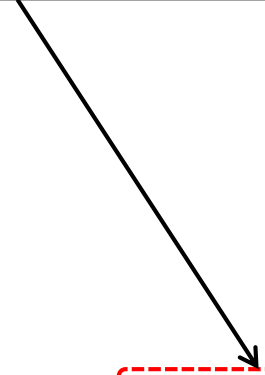
- **3'発現アレイ** → exon array → **transcriptome array**
 - Affymetrix Human Transcriptome Array (HTA 2.0)
 - Furney et al., *Cancer Discov.*, **3**: 1122-1129, 2013.
 - GPL17585(exon level)
 - GPL17586(gene level)



Affymetrix GeneChip

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

赤枠内ですすように、1つの遺伝子(転写物)の発現レベルを調べるのに、通常10個程度のプローブを利用。これをプローブセット(probeset)という。プローブごとに測定されたシグナル情報からなる数値ベクトルをスカラー値としてまとめる必要がある。プローブセット(≒遺伝子or転写物)の発現量算出に相当。



3'発現アレイの意味

Affymetrix GeneChipデータで、転写物の3'側に偏ったプローブ配列が利用されていることを確認してみる(わかるヒトは他のメーカーものでもよい)。Contact列のAffymetrix, Inc.のところをクリック。

Series	Samples	Platforms	DataSets	Summary	Advanced search			
Search		14,265 platforms	Export					
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	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	54,675	105665	3877	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Mus musculus</i>	45,101	43598	3330	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	22,283	37115	1024	Affymetrix, Inc.	Mar 11, 2002
GPL10558	Illumina HumanHT-12 V4.0 expression beadchip	oligonucleotide beads	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	47,323	33365	1000	Illumina Inc.	Jun 17, 2010
GPL13534	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	oligonucleotide beads	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	485,577	27581	407	Illumina Inc.	May 13, 2011
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	33,297	22489	1126	Affymetrix, Inc.	Dec 05, 2007
GPL13112	Illumina HiSeq 2000 (<i>Mus musculus</i>)	high-throughput sequencing	<input checked="" type="checkbox"/> <i>Mus musculus</i>		21205	1467	GEO	Feb 02, 2011
GPL11154	Illumina HiSeq 2000 (<i>Homo sapiens</i>)	high-throughput sequencing	<input checked="" type="checkbox"/> <i>Homo sapiens</i>		20519	1344	GEO	Nov 02, 2010
GPL6947	Illumina HumanHT-12 V3.0 expression beadchip	oligonucleotide beads	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	49,576	20154	432	Illumina Inc.	Jun 10, 2008
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Mus musculus</i>	35,557	17699	1432	Affymetrix, Inc.	Dec 05, 2007
GPL1355	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Rattus norvegicus</i>	31,099	17475	546	Affymetrix, Inc.	Jul 20, 2004
GPL8490	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	oligonucleotide beads	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	27,578	17009	300	Illumina Inc.	Apr 27, 2009
GPL6480	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name	in situ oligonucleotide	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	41,108	14671	619	Agilent Technologies	Feb 11, 2008



3'発現アレイの意味

①Affymetrix GeneChipのみだと、225種類存在することが分かる。赤矢印のものが3'発現アレイ。ここでは②ラットアレイ(Affymetrix Rat Genome 230 2.0 Array)に搭載されているプローブセットを眺める。

Series	Samples	Platforms	DataSets	Summary	Advanced search			
Search		225 platforms	①	Export				
Accession	Title	Technology	Organism(s)	Data rows	Samples	Series	Contact	Release date
Filter							Affymetrix, Inc.	
GPL570	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	in situ oligonucleotide	<i>Homo sapiens</i>	54,675	105676	3878	Affymetrix, Inc.	Nov 07, 2003
GPL1261	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array	in situ oligonucleotide	<i>Mus musculus</i>	45,101	43598	3330	Affymetrix, Inc.	May 25, 2004
GPL96	[HG-U133A] Affymetrix Human Genome U133A Array	in situ oligonucleotide	<i>Homo sapiens</i>	22,283	37115	1024	Affymetrix, Inc.	Mar 11, 2002
GPL6244	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Homo sapiens</i>	33,297	22489	1126	Affymetrix, Inc.	Dec 05, 2007
GPL6246	[MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Mus musculus</i>	35,557	17699	1432	Affymetrix, Inc.	Dec 05, 2007
GPL1355	② [Rat230_2] Affymetrix Rat Genome 230 2.0 Array	in situ oligonucleotide	<i>Rattus norvegicus</i>	31,099	17475	546	Affymetrix, Inc.	Jul 20, 2004
GPL6801	[GenomeWideSNP_6] Affymetrix Genome-Wide Human SNP 6.0 Array	in situ oligonucleotide	<i>Homo sapiens</i>	1,880,794	14465	255	Affymetrix, Inc.	Apr 30, 2008
GPL571	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	in situ oligonucleotide	<i>Homo sapiens</i>	22,277	12236	483	Affymetrix, Inc.	Nov 07, 2003
GPL198	[ATH1-121501] Affymetrix Arabidopsis ATH1 Genome Array	in situ oligonucleotide	<i>Arabidopsis thaliana</i>	22,810	11747	909	Affymetrix, Inc.	Jul 18, 2002
GPL5175	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]	in situ oligonucleotide	<i>Homo sapiens</i>	316,919	10440	293	Affymetrix, Inc.	May 21, 2007
GPL3718	[Mapping250K_Nsp] Affymetrix Mapping 250K Nsp SNP Array	in situ oligonucleotide	<i>Homo sapiens</i>	262,338	10275	165	Affymetrix, Inc.	May 13, 2006

3'発現アレイの意味

「Series (546)」は、このプラットフォーム (GPL1355; Affymetrix Rat Genome 230 2.0 Array)を用いて、大まかに546種類のデータセットの発現情報を利用可能ということ。

NCBI > GEO > **Accession Display**

Scope: **Self** Format: **HTML**

Platform GPL1355

Status	Public on Jul 20, 2004
Title	[Rat230_2] Affymetrix
Technology type	in situ oligonucleotide
Distribution	commercial
Organism	Rattus norvegicus
Manufacturer	Affymetrix
Manufacture protocol	see manufacturer's website

URL <http://www.affymetrix.com/index.affx>

Street address
City **Santa Clara**
State/province **CA**
ZIP/Postal code **95051**
Country **USA**

Samples (17475) [GSM42161](#), [GSM42162](#), [GSM42163](#), [GSM42164](#), [GSM42165](#), [GSM42166](#)
[# More...](#)

Series (546)
[# More...](#)

- [GSE2275](#) A Bioinformatic Analysis of Arginine-Sensitive Regulation of rat Hepatic Gene Expression
- [GSE2547](#) Olanzapine vs. Saline rat pfc
- [GSE2690](#) H_Alessio_05: sedentary and active lifestyles

Relations

- Alternative to [GPL6872](#) (Alternative CDF)
- Alternative to [GPL7598](#) (Alternative CDF)
- Alternative to [GPL8201](#) (Alternative CDF)
- Alternative to [GPL8597](#) (Alternative CDF)

3'発現アレイの意味

さらにページ下部に移動。Data tableでこのアレイに搭載されている①プローブセットのIDと対応するGenbank accession番号(GB_ACC)情報などを概観できる。②Download full tableを押すと、約30MBのアノテーションファイル(GPL1355-14795.txt)が得られる。

Data table

ID	GB_ACC	SPOT_ID	Species Scientific Name	Annotation Date	Sequence Type	Sequ Source
1367452_at	NM_133594		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank gb:NM_1
1367453_at	NM_053743		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank gb:NM_0
1367454_at	NM_021765		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank gb:NM_0
1367455_at	NM_053864		Rat			
1367456_at	NM_031237		Rat			
1367457_at	NM_053739		Rat			
1367458_at	NM_031342		Rat			
1367459_at	BE329198		Rat			
1367460_at	BM387347		Rat			
1367461_at	NM_080781		Rat			
1367462_at	U10861		Rat			
1367463_at	AI235510		Rat			
1367464_at	AI598391		Rat			
1367465_at	AI013627		Rat			
1367466_at	AI406518		Rat			
1367467_at	BG380638		Rat			
1367468_at	BM386654		Rat			
1367469_at	BG667162		Rat			

1367467_at	BG380638		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank	gb:BG38
1367468_at	BM386654		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank	gb:BM38
1367469_at	BG667162		Rattus norvegicus	Jun 9, 2011	Consensus sequence	GenBank	gb:BG66

Total number of rows: 31099

Table truncated, full table size: 29557 Kbytes.

Download full table...

Annotation SOFT table...

Download family

Format
SOFT ?
MINiML ?

Supplementary data files not provided

3'発現アレイの意味

ダウンロードしたアノテーションファイル(GPL1355-14795.txt)をエクセルで眺めているところ。慣れると、全体像から、(このファイルの場合は)最初の16行分がヘッダー行となっていることが分かる。17行目以降がメイン。

Excel screenshot showing the file 'GPL1355-14795.txt' with a formula bar containing '#ID = Affymetrix Probe Set ID'.

ID	GB_ACC	SPOT_ID	Species	Sc	Annotation	Sequence	Sequence	Target	Des	Representa	Gene Title	Gene Sym	ENTREZ_GI	RefSeq Tr	Gene
1367452_a	NM_133594		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_1335	NM_133594	SMT3 supp	Sumo2	690244	NM_133594	00164		
1367453_a	NM_053743		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0537	NM_053743	cell division	Cdc37	114562	NM_053743	00517		
1367454_a	NM_021765		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0217	NM_021765	coatomer p	Copb2	60384	NM_021765	00068		
1367455_a	NM_053864		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0538	NM_053864	valosin-co	Vcp	116643	NM_053864	00063		
1367456_a	NM_031237		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0312	NM_031237	ubiquitin-cl	Ube2d3	81920	NM_031237	00002		
1367457_a	NM_053739		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0537	NM_053739	beclin 1, au	Becn1	114558	NM_001034	00000		

The table shows the first 16 rows of the file, which are the header rows, and the first 6 rows of the main data. The data rows are highlighted with a red border in the original image.

3'発現アレイの意味

赤枠の列がGene Symbol列。例として、山中因子(Oct3/4, Sox2, Klf4, c-Myc)のプローブが搭載されているかどうかを調べてみる。

GPL1355-14795.txt - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 アドイン

A1 : #ID = Affymetrix Probe Set ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	#ID = Affymetrix Probe Set ID													
2	#GB_ACC = GenBank Accession Number													
3	#SPOT_ID = identifies controls													
4	#Species Scientific Name =													
5	#Annotation Date =													
6	#Sequence Type =													
7	#Sequence Source = The database from which the sequence used to design this probe set was taken.													
8	#Target Description = GenBank description associated with the representative public identifier. Blank for some probe sets.													
9	#Representative Public ID = The accession number of a representative sequence. Note that for consensus-based probe sets, the representative sequence													
10	#Gene Title = Title of Gene represented by the probe set.													
11	#Gene Symbol = A gene symbol, when one is available (from UniGene).													
12	#ENTREZ_GENE_ID = Entrez Gene Database UID													
13	#RefSeq Transcript ID = References to multiple sequences in RefSeq. The field contains the ID and Description for each entry, and there can be multiple													
14	#Gene Ontology Biological Process = Gene Ontology Consortium Biological Process derived from LocusLink. Each annotation consists of three parts:													
15	#Gene Ontology Cellular Component = Gene Ontology Consortium Cellular Component derived from LocusLink. Each annotation consists of three parts:													
16	#Gene Ontology Molecular Function = Gene Ontology Consortium Molecular Function derived from LocusLink. Each annotation consists of three parts:													
17	ID	GB_ACC	SPOT_ID	Species Sc	Annotation	Sequence	Sequence	Target Des	Representa	Gene Title	Gene Sym	ENTREZ_GI	RefSeq Tra	Gene
18	1367452_a	NM_133594		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_1335	NM_133594	SMT3 supp	Sumo2	690244	NM_133594	00164
19	1367453_a	NM_053743		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0537	NM_053743	cell division	Cdc37	114562	NM_053743	00517
20	1367454_a	NM_021765		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0217	NM_021765	coatome	Copb2	60384	NM_021765	00068
21	1367455_a	NM_053864		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0538	NM_053864	valosin-co	Vcp	116643	NM_053864	00063
22	1367456_a	NM_031237		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0312	NM_031237	ubiquitin-c	Ube2d3	81920	NM_031237	00002
23	1367457_a	NM_053739		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_0537	NM_053739	beclin 1, au	Becln1	114558	NM_001034	00000

準備完了

100%

3'発現アレイの意味

少なくともSox2 (1379942_at)と Klf4 (1387260_at)は、このアレイに搭載されていることがわかる。

GPL1355-14795.txt - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 アドイン 門田幸二

K12506 : *fx* Sox2

	A	B	C	D	E	F	G	H	I	J	K
12504	1379940_at	BE101968		Rattus nor	9-Jun-11	Consensus	GenBank	gb:BE101968	BE101968		
12505	1379941_at	BM384685		Rattus nor	9-Jun-11	Consensus	GenBank	gb:BM384685	BM384685		
12506	1379942_at	AA858564		Rattus nor	9-Jun-11	Consensus	GenBank	gb:AA858564	AA858564	SRY (sex d	Sox2
12507	1379943_at	AA943822		Rattus nor	9-Jun-11	Consensus	GenBank	gb:AA943822	AA943822		
12508	1379944_at	BE112610		Rattus nor	9-Jun-11	Consensus	GenBank	gb:BE112610	BE112610		

GPL1355-14795

GPL1355-14795.txt - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 アドイン 門田幸二

K19345 : *fx* Klf4

	A	B	C	D	E	F	G	H	I	J	K
19343	1387258_at	D11475		Rattus nor	9-Jun-11	Consensus	GenBank	gb:D11475	D11475	protein-L-	Pcmt1
19344	1387259_at	AF097593		Rattus nor	9-Jun-11	Consensus	GenBank	gb:AF097593	AF097593	cadherin 2	Cdh2
19345	1387260_at	NM_053713		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_053713	NM_053713	Kruppel-like	Klf4
19346	1387261_at	NM_017042		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_017042	NM_017042	protein phc	Ppp3cb
19347	1387262_at	NM_031119		Rattus nor	9-Jun-11	Consensus	GenBank	gb:NM_031119	NM_031119	Sjogren syn	Ssb

GPL1355-14795

準備完了

100%

GGRNAでSox2 (1379942_at)の転写物配列上でのプローブ位置を確認してみましょう。

3'発現アレイの意味

(Rで)マイクロアレイデータ解析

(last modified 2015/04/24, since 2005)

- (削除予定)Rのインストールと起動 (last modified 2014/05/14)
- (削除予定)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 2014/06/02)
- [書籍 | について](#) (last modified 2014/05/12)
- 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\(Affymetrix 3'発現アレイ\)](#) (last modified 2014/05/12)
- 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)
- 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)
- 書籍 | トランスクリプトーム解析 | [2.2.2 データの正規化](#)
- 書籍 | トランスクリプトーム解析 | [2.2.3 データの正規化](#)
- 書籍 | トランスクリプトーム解析 | [2.2.4 データの正規化](#)
- 書籍 | トランスクリプトーム解析 | [2.2.5 アノテーション](#)
- 書籍 | トランスクリプトーム解析 | [3.2.1 クラスタリング](#)
- 書籍 | トランスクリプトーム解析 | [3.2.2 実験デザイン](#)
- 書籍 | トランスクリプトーム解析 | [3.2.3 多重比較問題](#)
- 書籍 | トランスクリプトーム解析 | [3.2.4 各種プロット](#)
- 書籍 | トランスクリプトーム解析 | [4.2.1 2群間比較 \(last modified 2014/05/09\)](#)
- 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](#)
- 書籍 | トランスクリプトーム解析 | [4.2.3 多群間比較](#)
- インタロ | 発現データ取得 | [公共DBから](#) (last modified 2014/05/09)
- インタロ | 発現データ取得 | [inSilicoDb\(Taminau 2011\)](#)
- インタロ | 発現データ取得 | [ArrayExpress\(Kauffmann 2011\)](#)
- インタロ | 発現データ取得 | [GEOquery\(Davis 2007\)](#)
- インタロ | 発現データ取得 | [公共DB\(GEO\)](#)

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)

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](#)

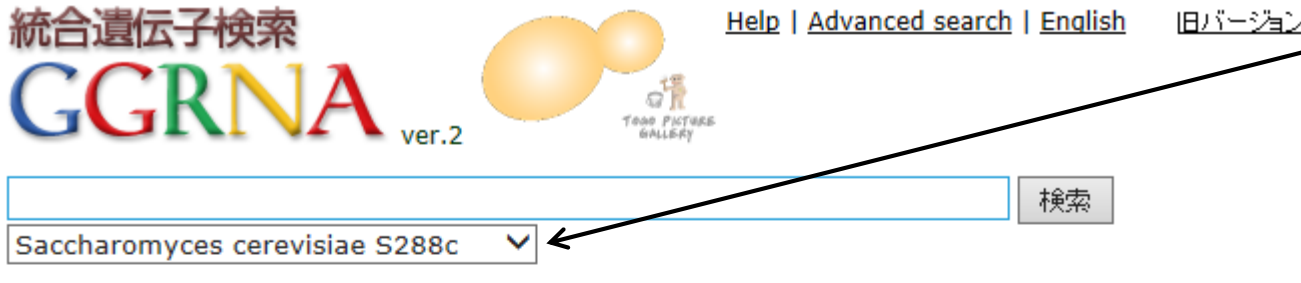
What's

• 門田知見
マイク
す。(C
お知
料な

• [はじめ](#)
• [過去](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)
• [イン](#)

3'発現アレイの意味

①Sox2のプローブセットID (1379942_at)を入力し、②デフォルトが酵母(Saccharomyces cerevisiae S288c)になっている場合は引っかけられないので、ここではZoo (All organisms in RefSeq)にしている。もちろんラット(Rattus norvegicus)にしてもよい。③検索ボタンを押す。



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

- 「homeobox」「claudin」..... フリーワード検索
- 「"RNA interference"」..... ダブルクオートで囲ってフレーズ検索
- 「Argonaute "PAZ domain"」..... Argonaute かつ "PAZ domain" のAND検索
- 「NM_001518」「10579」..... RefSeq IDやGene IDなど各種IDから検索
- 「symbol:VIM」..... 遺伝子名(symbolまたはsynonym)から検索
- 「ref:Naito」..... 文献情報のなかからフリーワード検索
- 「1552311_a_at」..... マイクロアレイのプローブIDから塩基配列を検索
- 「aa:KDEL」..... アミノ酸配列を検索
- 「caagaagaagattg」..... 塩基配列を検索
- 「comp:caagaagaagattg」..... 相補鎖を検索
- 「iub:aggtcannrtgacct」..... N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- 詳細な使い方

- 新着情報:
- 2015-01-13 HTTPSによる暗号化通信に対応 - <https://>
 - 2015-01-13 データベースをRefSeq rel. 69 (Jan, 2015)に更新
 - 2013-07-24 ソースを公開 - [GitHub](#)
 - 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを対応
 - 2012-05-29 下記論文の日本語による解説を「DBCLS」に掲載
 - 2012-05-29 GGRNAの論文がNucleic Acids Researchに掲載
 - 過去の新着情報

GGRNA ver.2 by @meso_cacase at DBCLS
This page is licensed under a [Creative Commons Attribution](#) license.



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

3'発現アレイの意味

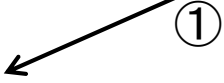
①Sox2のプローブセットID (1379942_at)は、全部で11個のプローブからなることが分かる。②転写物配列の長さは2,323 bpで、プローブのスタート位置は「1726, 1775, 1808, …, 2231」であり、これだけでも明らかに3'側に偏っていることが分かる。③全配列での位置を概観。

1379942_at
Zoo (All organisms in RefSeq) ▼

2015-05-11 18:36:33, GGRNA : RefSeq release 69 (Jan, 2015)

Summary:

- [seq:TTGGAACCAGGCTGCCGAGAATCCA \(2\)](#)
- [seq:TTAAGGACCGTTACAGACAAGGAAG \(2\)](#)
- [seq:TGAGAGCAAGTACTGGCAAGACCAT \(2\)](#)
- [seq:TAATAGCTTTTGTTCATCCACCC \(2\)](#)
- [seq:GGTTCTTGCTGGGTTTTGATTCTGC \(4\)](#)
- [seq:GGGCAAACGTTCTAGATTGTAATA \(2\)](#)
- [seq:CATTTTCGTGGTCTTGTITAGGGCA \(2\)](#)
- [seq:ATGTCCATTGTTTATAAGCTGAGAA \(36\)](#)
- [seq:ATGTCCATGCAGGTTGATATCGTIG \(5\)](#)
- [seq:ATACCATCCTTATAACAGTTACGTI \(2\)](#)
- [seq:AAAGATTTCGGCTGTTATTGGAACCA \(2\)](#)
- **INTERSECTION (2)**



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。



[Rattus norvegicus SRY \(sex determining region Y\)-box 2 \(Sox2\), mRNA. \(2323 bp\)](#)
 tgtacagtattatcgagataaacaatggcaatcaa**atgtccattg**ttataagctgagaa**ttgccaat**attttcgaggaaag**gg**
ttcttgctggg**ttt**gatt**ctgc**agcttaa**taaggaccg**ttacagacaaggaaggaattattcggattgaa**cg**tttagtt
 ttaa**att**gtacaaaagga**aaaca**tgagagcaag**tactgg**caagaccat**ttc**gtggt**ctt**g**tt**aggg**caa**acgt**ctag**
attgtactaa**at**ttta**act**tactgtta**agg**caaaaa**aaat**gtccat**gc**agg**tt**gat**at**cg**ttg**gta**att**ta**at**ag**ct**ttt
gttcaat**ccc**acc**ct**tttcatt**tt**gttcacata**aaa**at**at**ggaa**att**actgtg**tt**gaa**at**tttt**ct**tatg**gt**ttgta**at**att**ct**gta**at**
 tgtgat**at**tttaag**gt**tttt**cccc**ttttat**tt**ccgtag**tt**gtatt**ta**aa**ag**att**cg**g**ct**g**tt**att**gga**accagg**ct**g**cc**gagaa
tccatgtatattga**act**a**at**accat**cc**ttata**ac**ag**tt**ac**gt**ttcca**act**taag**tt**tactcca**at**g**ca**cag**tt**...

position 1726 1775 1808 1886 1908 1928 1985 2018 2173 2189 2231
Synonym: RGD1565646



3'発現アレイの意味

③全配列でのプローブ位置を概観した結果。3'発現アレイの意味がよく分かる。

1379942_at
Zoo (All organisms in RefSeq) ▾

検索

2015-05-11 18:36:33, GGRNA : RefSeq release 69 (Jan, 2015)

Summary:

- seq:TTGGAACCAGGCTGCCGAGAATCCA (2)
- seq:TTAAGGACCGTTACAGACRAGGAAG (2)
- seq:TGAGAGCAAGTACTGGCAAGACCAT (2)
- seq:TAATAGCTTTTGTTCATCCCACCC (2)
- seq:GGTTCTTGCTGGGTTTGTATTCTGC (4)
- seq:GGGCAACAGTTCTAGATTGACTAA (2)
- seq:CATTTTCGTGGTCTTGTITAGGGCA (2)
- seq:ATGTCCATGTTTATAAGCTGAGAA (36)
- seq:ATGTCCAIGCAGGTTGATATCGTTG (5)
- seq:ATACCATCCTTATAACAGTTACGTT (2)
- seq:AAAGATTCCGGCTGTTATTGGAACCA (2)
- INTERSECTION (2)

Results:

トップ50件を表示。検索語に色がつきます。重なりと色が濃く表示されます。



Rattus norvegicus SRY (sex determining region Y)-box 2 (Sox2), mRNA. (2323 bp)
 tgtacagtattatcgagataaacatggcaatcaaatgtccattgtttataagctgagaaattgccaatattttcgaggaaagg
 ttctgtcgggtttgattctgcagcttaaaataggaccgttacagacaaggaaggaattatccggattggaacgttttagtt
 taaaattgtacaaaaggaaaacatgagagcaagtactggcaagaccattttcgtggtctgttttagggcaaacgttctag
 attgactaaatttttaactactgttaaggcaaaaaaaatgtccatgcaggttgatattgtaattataatagctttt
 gttcaatcccaccccttttcaattttgtcacataaaaataggaaattactgtgttgaaatattttctatggtttgtaatttctgtaaat
 tgtgatatttaagggtttttcccttttattttccgtatgttatttaaaagattcggctgttattggaaccagggtgccggagaa
 tccatgtatataattgaaactaataccatccttataacagttacggttccaacttaagtttttactccattatgcacagttt...

position 1726 1775 1808 1886 1908 1928 1985 2018 2173 2189 2231
 Synonym: RGD1565646
 NM_001109181.1 - *Rattus norvegicus* (Norway rat) - NCBI - UCSC - RefEx(expression)

5'側

```
gtgtttgcaaaaagggaaaagtactttgctgcctctttaagactagggtcgggagaaagaag
aggagagaaaaagaaaggagagaagtggagcccgaggcttaagcctttccaaaaactaat
cacaacaatcgccggcgcccgaggaggagagcgactgtttttcatcccaattgcaactcgc
ccgtctcgagctccgcttcccccaactatttccgcccagatctccggcgaaggccgtgca
gccgacgaccccgcccgccgcccctgcatcccgcccccggcgccggcccccaggtcccgg
ccgggcccagggtcgccggccggccggccggccggccggccggccggccggccggccggcc
atgatggagacggagctgaagccggccggccctcagcaagcttcggggggcgccggcgagg
aggcaacgccaaggcgccggcgaaccggcgaaccagaagaacagcccgaccgctcaaga
ggcccatgaatgccttcatggtgtggtcccggggcagcggcgttaagatggcccaggagaac
cccaagatgcaactcggagatcagcaagcgcctggcgccgagtggaactttgtcgga
gaccgagaagcggccgttcatcgacgagggccaagcggctgcccgtctgacatgaaaggac
accggattataaataaccggccggcggaacccaagacgctcatgaagaaggataagtac
acgcttcccggaggcttgcctggccccggcggaacagcatggcgagcgggttgggggtgg
cgccggcctgggtgcccggcgtgaaccagcagatggacagctacggcagatgaacggctgga
gcaacggcagctacagatgatgcaggagcagctgggctaccgcagcaccggccctcaac
gctcacggcgccgacagatgcagccgatgcaccgctacgacgctcagcggccctgcagtaca
ctccatgaccagctcgcagacctacatgaacggctcggccacctacagcatgtcctactcgc
agcagggcacccccggataggcgtgggctccatgggctctgtggtcaagtccgaggccagt
tccagccccccggttacctcttctcccaactccagggcgccctgccaggccggggacct
ccgggacatgatcagcatgtacctcccggcggcggaggtgcccggagcccgctgcccagta
gactgcacatggcccagcactaccagagcggcccggtgcccggcaccggccatgaacggcaca
ctgcccctgtgcacatgtgagggcgggaccggcaactggagaaggggagagattttcaaa
aagatacaagggaattgggaggggtgcaaaaaggagagatgaagaaaaatctgaatgctcaaa
aggaaaaaaaatctcattaccgcagcaaaatgacagctgcccgaaaaaaacccaatcc
catccaaatgacgcaaaaaccgtgatgcccactagaaaactttatgagagatctggagga
aaaaaactacgcaaaaacttttttaaagtctagtggtacgttagggcgttcgcagggagt
tctcaaaagtcttaccagtaataattagaactagactccggcgatgaaaaagttttaat
atgtgcaagcaactttgtacagatattatcgagataaacatggcaatcaaatgtcoattgt
ttataagctgagaatttgccaatattttcgaggaaaagggttcttctgtgggttttgattctg
cagcttaaaataggaccgttacagacaaggaaggaatttattcggattgcaacgttttagt
ttaaaattgtacaaaaggaaaacatgagagcaagtactggcagcaagcaatcttctgtggtt
gttttagggcaaacgttctagattgtactaaatttttaactactgttaaggcaaaaaaaa
atgtcoatgcaaggttgatctggtgtaattataaagcttttggttcaatcccaccccttt
cattttgttcacataaaaaataggaaattactgtgtttgaaatattttctatggtttgtaa
tatttctgtaaatgtgatatttaagggtttttccccccttttattttccgtagttgtatt
taaaagattcggctgttattggaaccagggtgccgagaaatccatgtatataatttgaaactaa
tccatccttataacagttacggttccaacttaagtttttactccattatgcacagtttgaga
taaaaaatttttgaaataggacactga
```

3'側

課題4

山中因子 (Oct3/4, Sox2, Klf4, c-Myc) の1つであるKlf4 (1387260_at) のプローブ配列についてもSox2と同様に調べよ。

1387260_at Zoo (All organisms in RefSeq)

2015-05-11 20:20:16, GGRNA : RefSeq release 69 (Jan, 2015)

Summary:

- [seq:TTTTCTATATAGTTCCTTGCCTTAA \(83\)](#)
- [seq:TGCTCTTTGCCTTCTAAAAAGCCAT \(4\)](#)
- [seq:GTTTATTCCCAAGTATGCCTTTAAG \(6\)](#)
- [seq:GTGCTTGGTGAGTCGTGGTTCTAAA \(4\)](#)
- [seq:GAGACTGGATCTTCTATCATTCCAA \(4\)](#)
- [seq:GAAGACCAGAGTTCCTTGAATTGT \(3\)](#)
- [seq:ATAAGCCTGGTTTATTCTGTAAACA \(2\)](#)
- [seq:AGTCTGTTATGCACGTGGTTTCAG \(77\)](#)
- [seq:AATCTATTTTTGCTTCCGATCTAC \(50\)](#)
- [seq:AAAGTTCTCCAACGTGTCATACTT \(5\)](#)
- [seq:AAAGACCACCTTGTATGCTCTTTGC \(1\)](#)
- **INTERSECTION (1)**

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃くなります。

[Rattus norvegicus Kruppel-like factor 4 \(gut\) \(Klf4\), mRNA. \(2393 bp\)](#)
aatcacagaacagatggggtctgagactggatcttctatcattccaataccaatccgacttgaacaagactggactta
caaaatgccaaggggtgactggaagtttgatagatcagggtatacattaaatcagtgactggggggagggaagacc
agagttccctgaattgtgcttcaatgatgcaatatacatgaaagaccacctgtatgctctttgcttctaaaag
ccattatgacgtcagaggaagaggaagcaattcaggtacagaacgtgttctaatagcctaaacgatggtgcttggtag
tcgtggttctaaaggtaacaaacgggggagccaaagtctccaactgctgcatactttgacaaggaaaatctat
tgtctccgatctacattatgacctaagtcaggtaataagcctggttatttctgtaacattttatgacagacgtct
gtatgacactggtttcagatgtgcaataattgtacaatggttattccaagatgcctttaagcagaacaatgt
gttttctatatagttccttgccttaataaatatgtaataataat...
[position] 1791 1919 1968 1983 2071 2115 2150 2197 2236 2283 2322
Synonym: GKLF
NM_053713.1 - Rattus norvegicus (Norway rat) - [NCBI](#) - [UCSC](#) - [RefEx\(expression\)](#)

5'側

```

atcttcgttgacttcggggttgggtaccctctctctcttcggactccggaggac
ctctgggccccacattaatgaggcagccactggcgagtctgacatggctgtcag
cgacgctctgctcccgtccttctccacgttcgctcggcccggggaaggagaa
gacactgctccagcaggtgccccgactaacggttggogagaggaactctctcacat
gaagcgactcccccaactcccgccgccccctacgacctggcgcgacggtggccac
agacctggaagtggtagctggtgcagcttgcagcagtaacaacccggccctacc
ccggaggagaccgaggagtccaacgatctcctggacctagactttatccttccaa
ctcgctatcccaccaggaatcggtggccgcaacgctgaccacctggcgctcagctt
atcctcgtctccccagctagcagcgccctgccagcgccctccacctcagctt
cagctatccgatccgggccccgggtgaccggcggtggctgcgggcaacacaggtgg
agggctcctctacagccgagaatctgcccacctcccacggccccctcaacctggc
ggacatcaatgacgtgagccccctggggcggtctggtggtgagctcctgcgccgga
gttggaccagatatacatccgccaagcagcctcagccgacaggtggcggtgat
gggcaagtgtgtgctgaagggctctctgagcaccctggcagcaggtacaccagcc
ttcggtcatcagtgtagcaaggaagcccagacggcagccacctgtggtagtggc
gcctacagcggtggcccccgctatgtgccccaaagattaagcaagaggccgtcc
gtcctgacgggtcagccggtcctagaggccactgagcgctggacccccagctcag
caacggccaagcccaacacacacgacttccccctggggcgcgactccccacag
gactaccctacactgagtcggggaactgctgaacagcagggactgtcacctgg
cctgctctccccaggattccatccccatccggggcccagctacctccttct
gccagaccagatgcagtcgcaagtccccctctctccattatcaagagctcatgccacc
gggatcctgctgccagaggagcccaagccaaagagggggaagaaggtcttggccccg
gaaaagaacagccaccaactgtgactatgcaggtgtggcaaaaactatacga
gagttctcatctcaaggcacacctgcgaactcacacagggcgagaaccttaccactg
tgactgggacggctgtgggtggaattcgccccgctcagatgaaactgaccaggcacta
ccgcaaacacaccgggacccggcccttccagtcgcaaggtgcgacagggccttttc
caggtcggaccaccttgcttacacatgaagaggcacttttaattccactcgtgg
acatgaccacactgccaggagagagttcagtttttttttaacctttcacactgt
ctccccagggggaggagcccagctggcaagcgctacaatcatggtcaagttccca
gcaagtcagcttgtgaatggataatcaggagaaggaagagtcacaagggacaaaaga
aaagaaaagaaaaaataactaaaaaacaaaacaaaaaaacaaaagaaaa
aaatcacagaacagatggggtctgagactggatcttctatcattccaataccaatccgacttgaacaagactggactta
cgaactgaaacagactggacttacaacatgcaaggggtgactggaagttgtggat
atcagggtatacatataatcagtgacctggggggagggaagaccagagttcccttga
attgtgcttcaatgatgcaatatacatggaagaccacctgtatgctctttgctt
ctaaaaagccattatgacgtcagaggaaggaagcaacttcaggtacagactggtt
ctaatagcctaaacagatggtgcttgggtgagtggtttctaaaggtacaaaacgggg
gagccaaagtctccaaactgctgcaactttgacaaggaaaatctattttgtcttc
cgatctadatttatgacctaaagtcaggtaaataagcctggttattttctgtaacatt
tttatgacagactgtgtatgcaactggttccagatggtgcaataattgtacaa
tggtttatcccccaagtatgctttaaagcagaacaaatggttctttctatatagttcc
ttgcttaataaatatgtaataataattttaaaaaaaaaaaaaaaaaaaaaaaaaa

```

3'側

Contents

- イントロダクション
 - トランスクリプトーム解析技術の原理や特徴
 - データ解析の全体像やスキル習得の意義
 - アセンブルプログラムのインストール
 - 発現データベース(DB)
 - Affymetrix GeneChip
- 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)

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., *BBB.*, 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

発現DBからの生データ取得

- 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)
- 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)
- 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)
- 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)
- イントロ | 発現データ取得 | **①** [公共DBから](#) (last modified 2014/05/11)
- イントロ | 発現データ取得 | [inSilicoDb\(Taminau 2011\)](#) (last modified 2015/05/11) NEW
- イントロ | 発現データ取得 | [ArrayExpress\(Kauffmann 2009\)](#) (last modified 2014/05/15) 推奨
- イントロ | 発現データ取得 | [アノテーション](#)
- イントロ | アノテーション | [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。](#)
- イントロ | アノテーション | [一次データベース](#)
- イントロ | プローブ配列

イントロ | 発現データ取得 | 公共DBから

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](#)
- **②** [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)
- [GSE10246](#) (マウス)
- [GSE1133](#) (ヒトとマウス)
- [GSE5364](#) (ヒト 341 サンプル)
- [GSE15998](#) (マウス)
- [ArrayExpress](#): [Rustici et al.](#)
- [GSE7623](#) (ラット 24 サンプル)
- [GSE30533](#) (ラット 10 サンプル)
- [GSE2361](#) (ヒト 36 サンプル)
- [GSE10246](#) (マウス)
- [GSE1133](#) (リンク先)
- [GSE5364](#) (ヒト 341 サンプル)
- [GSE15998](#) (マウス)

Series GSE7623

[Query DataSets for GSE7623](#)

Status	Public on Jan 09, 2008
Title	24 h-fasting effects on the brown and white adipose tissue and liver
Organism	<i>Rattus norvegicus</i>
Experiment type	Expression profiling by array
Summary	The functional balance between brown adipose tissue (BAT) and white adipose tissue (WAT) is important for metabolic homeostasis. We compared the effects of fasting on the gene expression profiles in BAT, WAT and liver, using DNA microarray analysis. Tissues were obtained from rats that had been fed or fasted for 24 h. Taking the false discovery rate (FDR) into account, we extracted the top 1,000 genes that were expressed differentially between fed and fasted rats. In all three tissues, Gene Ontology analysis revealed marked changes in the expression of 'metabolism' category genes and a hypergeometric test demonstrated that within this category, lipid and protein biosynthesis-related genes were down-regulated. These findings indicate simultaneous down-regulation of genes involved in energy-consuming pathways in the BAT, WAT and liver of fasted rats. In the BAT of fasted rats, there was marked up-regulation of genes in the 'protein ubiquitination'

発現DBからの生データ取得

- ①全部で24サンプルのデータからなることが分かる(24枚のレイを使っている)。
- ②生データのダウンロードはここ。hogeフォルダ中にあり

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 25, 2007

Last update date Sep 27, 2013

Contact name Yuji Nakai

Organization name The University of Tokyo

Street address 1-1-1, Yayoi

City Bunkyo-ku

State/province Tokyo

ZIP/Postal code 113-8657

Country Japan

Platforms (1) [GPL1355](#) [Rat230_2] Affymetrix Rat Genome 230 2.0 Array

Samples (24) [GSM184414](#) Brown adipose tissue, fed #1
[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 ?

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



発現DBからの生データ取得

デスクトップ上でtarファイルを解凍しても、さらにgzip圧縮されたファイルの解凍を行う必要がある。さらに、事実上CELという拡張子のついたファイルしか利用しない。

ncbi.nlm.nih.gov から GSE7623_RAW.tar (61.7 MB) を開くか、または保存しますか?

ファイルを開く(O) 保存(S) キャンセル(C)

名前を付けて保存

デスクトップ

整理 新しいフォルダー

お気に入り

- SkyDrive
- ダウンロード
- デスクトップ
- 最近表示した場所

ライブラリ

- ドキュメント
- ピクチャ
- ビデオ
- ミュージック

ライブラリ

- ライブラリ
- システム フォルダー
- kadota
- システム フォルダー
- コンピューター
- システム フォルダー
- ネットワーク
- システム フォルダー
- E-GEOD-7623.raw.1
- ファイル フォルダー

ファイル名(N): GSE7623_RAW.tar

ファイルの種類(T): TAR ファイル (*.tar)

フォルダーの非表示

C:\Users\kadota\Desktop\GSE7623_RAW

整理 ライブラリに追加 共有 >>

名前	更新日時	種類	サイズ
GSM184414.CEL.gz	2007/04/28 2:39	GZ ファイル	2,494
GSM184414.CHP.gz	2007/04/28 2:40	GZ ファイル	169
GSM184414.EXP.gz	2007/04/28 2:39	GZ ファイル	1
GSM184415.CEL.gz	2007/04/28 2:39	GZ ファイル	2,503
GSM184415.CHP.gz	2007/04/28 2:40	GZ ファイル	169
GSM184415.EXP.gz	2007/04/28 2:39	GZ ファイル	1
GSM184416.CEL.gz	2007/04/28 2:39	GZ ファイル	2,404
GSM184416.CHP.gz	2007/04/28 2:40	GZ ファイル	170
GSM184418.CHP.gz	2007/04/28 2:40	GZ ファイル	170

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

R経由で生データ取得

デスクトップ上でtarファイルを解凍しても、さらにgzip圧縮されたファイルの解凍を行う必要がある。さらに、事実上CELという拡張子のついたファイルしか利用しない。

- ・ 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)
- ・ イントロ | 発現データ取得 | [公共DBから](#) (last modified 2014/05/11)
- ・ イントロ | 発現データ取得 | [inSilicoDb\(Taminau 2011\)](#) (last modified 2015/05/11) **NEW**
- ・ イントロ | 発現データ取得 | **1** [ArrayExpress\(Kauffmann 2009\)](#) (last modified 2014/05/15) **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 | 発現データ取得 | [ArrayExpress\(Kauffmann_2009\)](#) **NEW**

マイクロアレイデータベース [ArrayExpress](#) に登録されているデータを [ArrayExpress](#) というRパッケージで取得するやり方を示します。GEO IDでも検索可能であり、CELファイルデータも取得可能、任意の preprocessing法を適用可能、などの利点からこのパッケージ経由での利用をお勧めします。「ファイル」-「ディレクトリの変更」でファイルを保存したいディレクトリに移動し以下をコピー。

1. Affymetrixデータ [E-MEXP-1422 \(Bourgon et al., PNAS, 2010\)](#) のCELファイルを取得し、RMA法 ([Irizarry et al., Biostatistics, 2003](#)) を実行して得られた発現情報を取得したい場合:

以下の [ArrayExpress](#) 関数のオプションを `save=F` から `save=T` に変更すると、CELファイルなどを含む全データのダウンロードも同時に行ってくれます。が、そんなことをいちいちやらなくても `ReadAffy` 関数を用いて読み込んだ状態と同じなので直接RMA ([Irizarry et al., Biostatistics, 2003](#)) などの任意の正規化法を適用可能です。

2 3. Affymetrixデータ [GSE7623 \(Nakai et al., BBB, 2008\)](#) のCELファイルを取得したい場合:

```
out_f <- "E"
param <- "E"

#必要なパッケージをロード
library(ArrayExpress)
library(affy)
```

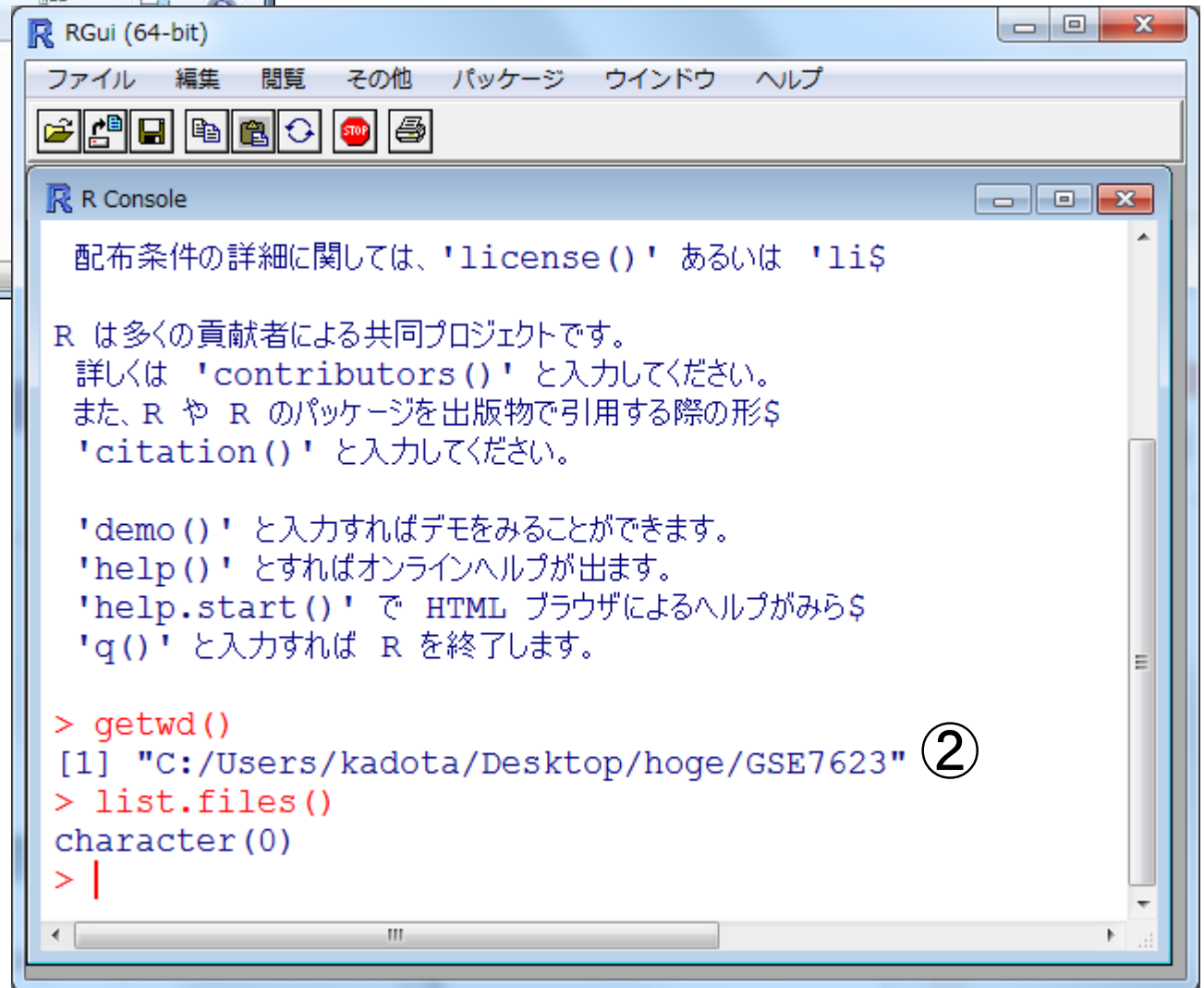
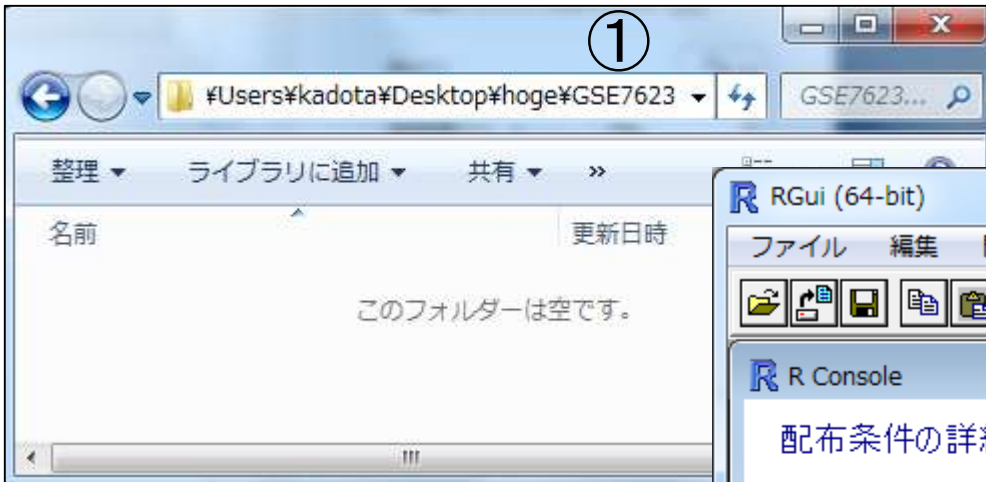
```
param <- "GSE7623" #入手したいIDを指定

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

#前処理(データ取得)
hoge <- getAE(param, type="raw", extract=F) #paramで指定したIDの生データをダウンロード
```

R経由で生データ取得

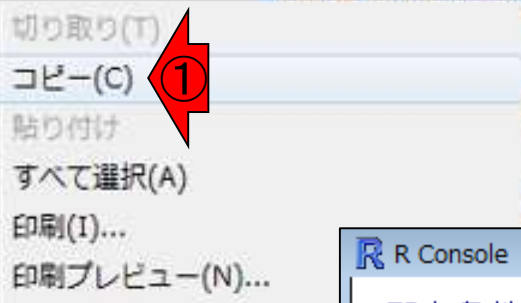
① デスクトップにhogeフォルダ、およびその中にGSE7623フォルダを作成する。② Rを起動し、作業ディレクトリをそこに変更しておく



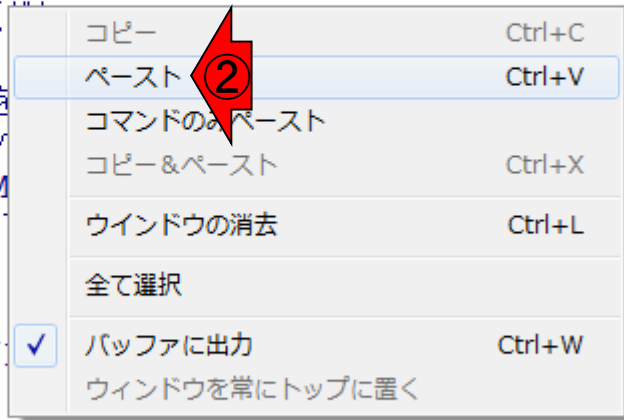
R経由で生データ取得

3. AffymetrixデータGSE7623 (Nakai et al., BBB, 2008)のCELファイルを取得したい場合:

```
param <- "GSE7623" #入手したいIDを指定
#必要なパッケージをロー
library(ArrayExpress)
#前処理(データ取得)
hoge <- getAE(param,
```



```
R Console
配布条件の詳細に関しては、'license()' あるいは 'li$
R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形$
'citation()' と入力してく
'demo()' と入力すればデモを
'help()' とすればオンライン
'help.start()' で HTML
'q()' と入力すれば R を終
> getwd()
[1] "C:/Users/kadota/
> list.files()
character(0)
> |
```



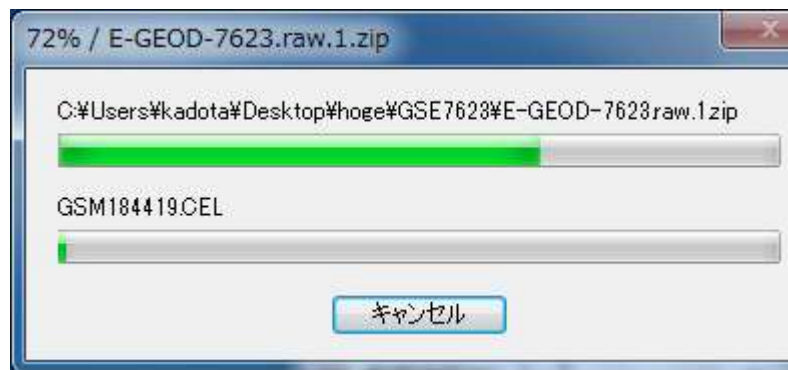
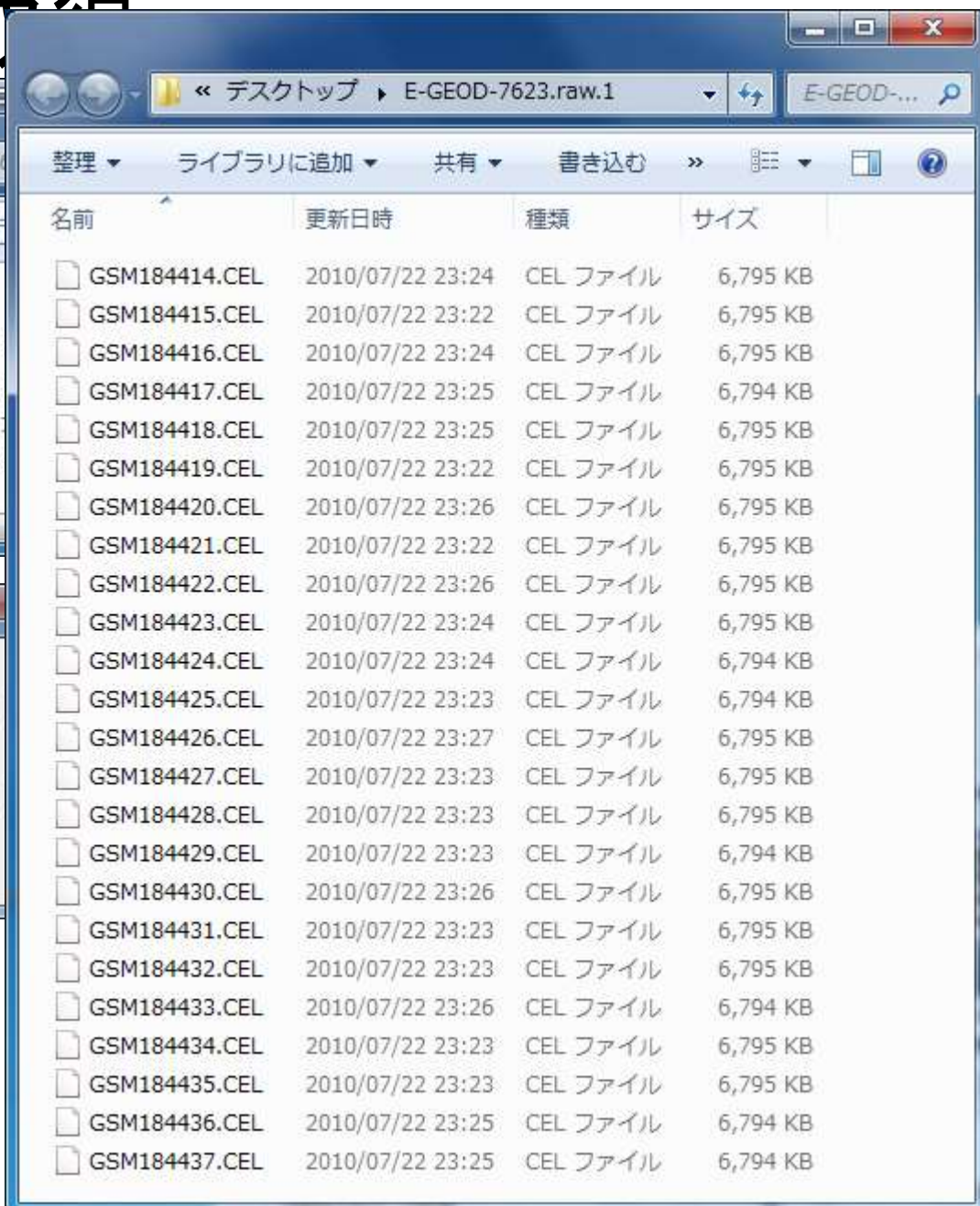
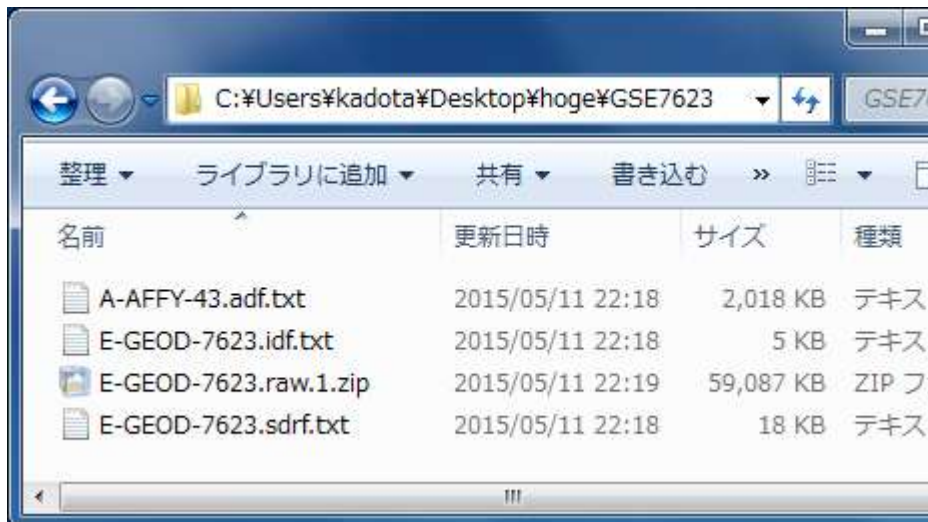
R経由で生データ取得

コピー実行後の状態。4つのファイルがダウンロードされている。このうち、zipファイル(の中身が目的物なので)を解凍。

```
R Console
URL 'http://www.ebi.ac.uk/arrayexpress/files/E-GEO$
Content type 'text/plain' length 18381 bytes (17 KB)
開かれた URL
5% downloaded
URL: ... i.ac.uk/arrayexpress/files/E-GEOD-7623/E-GEOD-7623.raw.1.zip
Copying raw data files
URL 'http://www.ebi.ac.uk/arrayexpress
Content type 'application/zip' length 6
開かれた URL
```

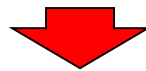
```
R Console
URL 'http://www.ebi.ac.uk/arrayexpress/files/E-GEO$
Content type 'text/plain' length 4484 bytes
開かれた URL
downloaded 4484 bytes
Copying raw data files
URL 'http://www.ebi.ac.uk/arrayexpress/files/E-GEO$
Content type 'application/zip' length 60504388 bytes$
開かれた URL
downloaded 57.7 MB
> list.files()
[1] "A-AFFY-43.adf.txt" "E-GEOD-7623.idf.txt"
[3] "E-GEOD-7623.raw.1.zip" "E-GEOD-7623.sdrf.txt"
> |
```

R経由で生データ取得



データ解析の全体像

	マイクロアレイ	RNA-seq
公共データ取得	GEO, ArrayExpress	GEO, ArrayExpress, NCBI SRA, EBI ENA, DDBJ SRA (DRA)
解析対象生物種	配列情報既知(アレイが提供されているもののみ)	モデル・非モデル問わず
生データ	プローブレベル数値データ	塩基配列(数億リード程度、数百塩基長) QC (Quality Control): クオリティチェック、フィルタリング、トリミング アセンブリでトランスクリプトーム配列取得(マッピング時のリファレンスとしても利用) マッピング(bowtie2, TopHat2など)でSAM/BAMファイル取得
発現行列作成	前処理法(MAS5, RMAなど)適用後に遺伝子発現行列を得る	アノテーションファイルを利用してカウントデータ、配列長補正後のRPKM/FPKM、転写物レベルの発現情報など取得
発現変動遺伝子(DEG)同定	基本Rを利用(limma, SAM, Rank productsなど)	基本Rを利用(cuffdiff2, edgeR, DESeq2, TCCなど)
機能解析	GSEA, GSA, Cytoscapeなど R/パッケージ SeqGSEAなどを利用。	




Javaインストール for Mac

以降のスライドは、門田のOS X Yosemite環境ではデフォルトでJavaが入っていなかったため、インストールしたときの奮闘記です。

```
kadota — bash — 58x10
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$
```

```
kadota — bash — 67x12
Last login: Tue May 12 11:54:03 on ttys000
agribio-macbook:~ kadota$ java -version
No Java runtime present, requesting install.
agribio-macbook:~ kadota$
```



“java”コマンドラインツールを使用するには、JDK をインストールする必要があります。

この Java Developer Kit のダウンロード Web サイトにアクセスするには、“詳しい情報...”をクリックしてください。

[詳しい情報...](#) [OK](#)

MacでのJavaインストール

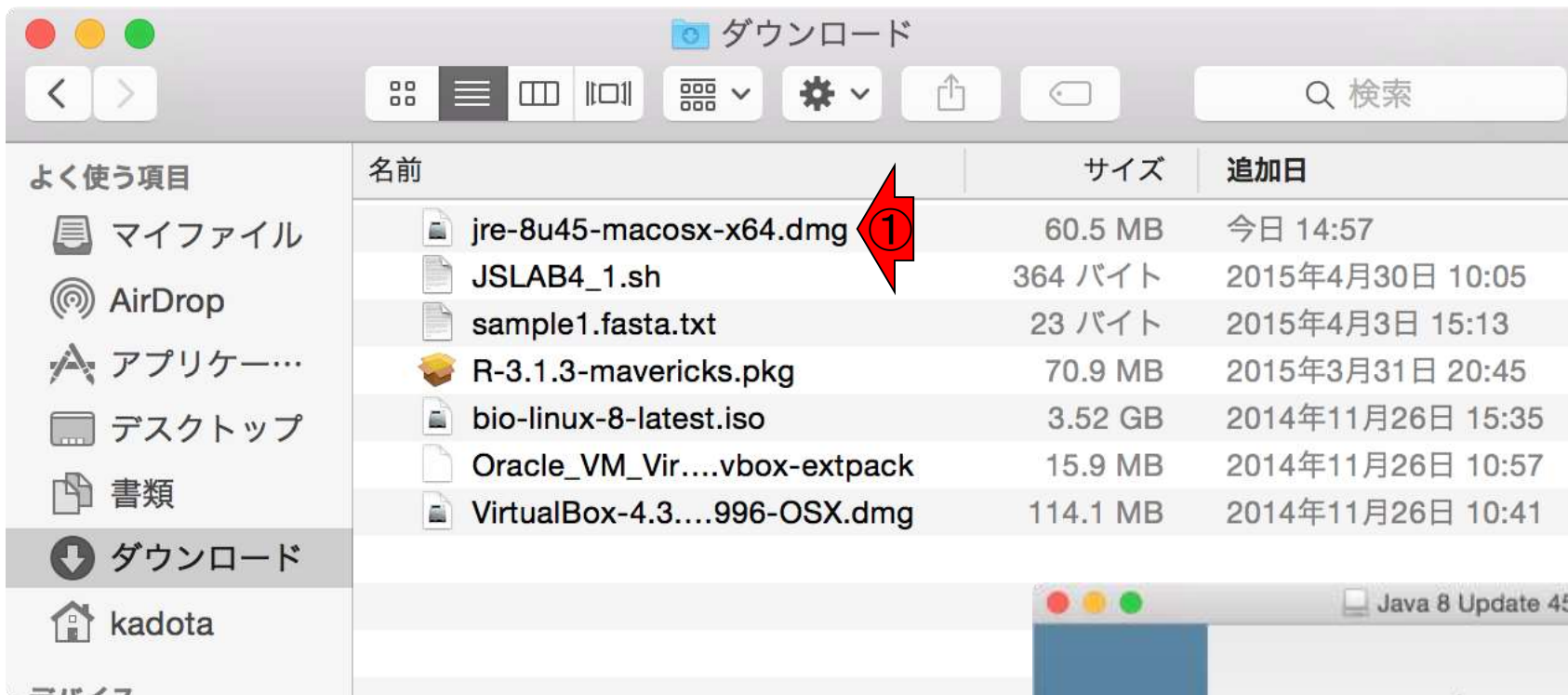
「Mac Java インストール」で検索して左記のサイトにたどり着いた。



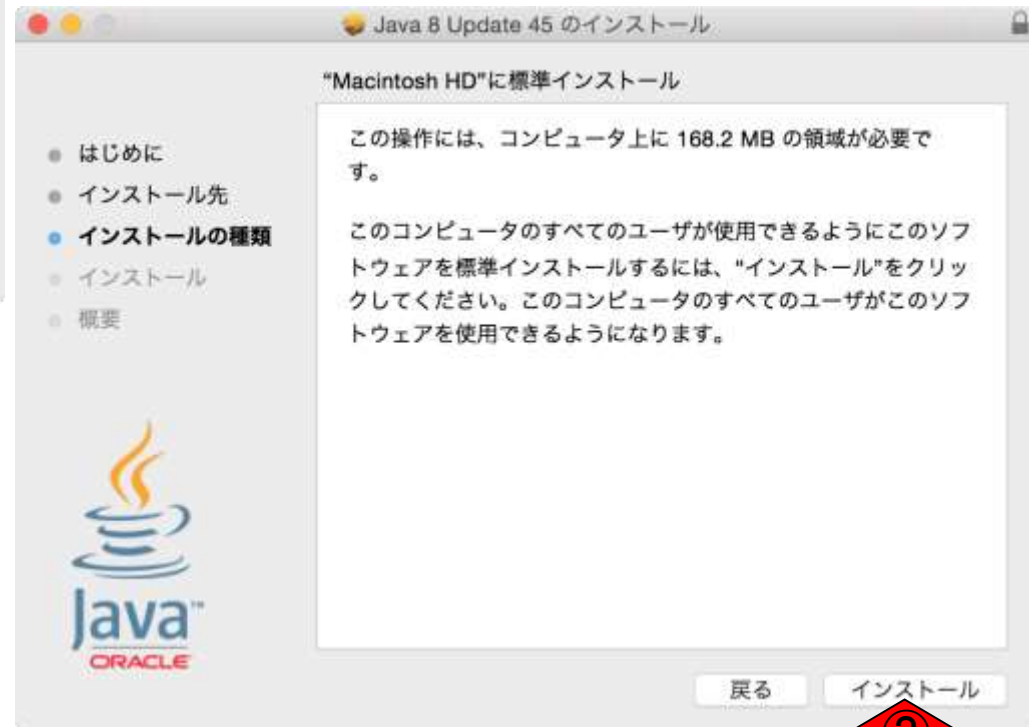
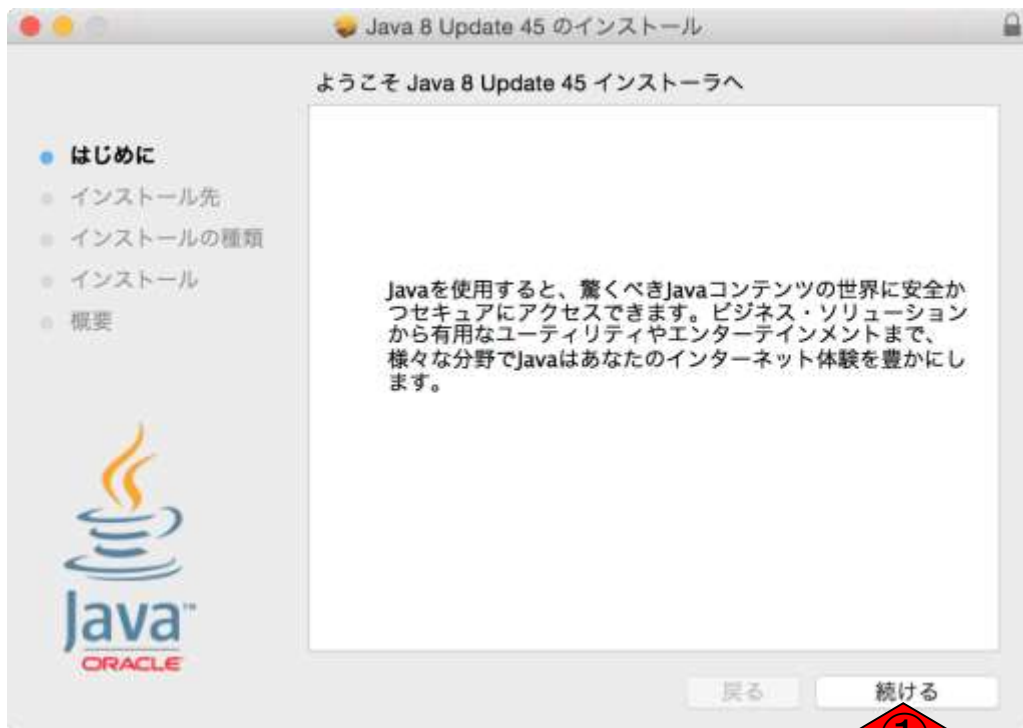
Javaのインストールが完了したら、お使いのブラウザでJavaを有効にするためにブラウザをリロード(Command+F)または終了(Command+Q)することが必要な場合があります。

MacでのJavaインストール

①「jre-...x64.dmg」ファイルをダブルクリック。②.pkgをダブルクリック



MacでのJavaインストール



MacでのJavaインストール

①ログインパスワードを入力して「ソフトウェアをインストール」。②閉じる。



インストーラが新しいソフトウェアをインストールしようとしています。これを許可するには、パスワードを入力してください。

ユーザ名： 門田幸二

パスワード： ●●●●●●

キャンセル

ソフトウェアをインストール



Java 8 Update 45 のインストール

インストールが完了しました。

- はじめに
- インストール先
- インストールの種類
- インストール
- 概要

Javaの更新が可能になると、プロンプトが表示されます。必ず更新をインストールして、パフォーマンスおよびセキュリティの最新の改善を反映させてください。
[更新設定についての詳細情報](#)

セキュリティ・レベルは最小限の値である「高」に設定されます。
[詳細情報](#)

ブラウザが開くので、Javaが動作していることを確認できます。



戻る

閉じる

MacでのJavaインストール

- ①「Javaのバージョンの確認」。
- ②「信頼」。



Java検出アプリケーションにより、実行の許可を求められます。「実行」をクリックすると、確認処理を続行できます。
システムの設定によっては、確認を続行するためにプロンプトへの応答が必要な場合があります。

Javaがインストールされていて確認でエラーが発生した場合は、構成(ブラウザ、Javaコントロール・パネ

MacでのJavaインストール

①「実行」して、「正常な設定です」となっていることを確認。

このアプリケーションを実行しますか。



名前: Java Detection

発行者: Oracle America, Inc.

場所: https://java.com

このアプリケーションは、コンピュータおよび個人情報を危険にさらす可能性がある無制限のアクセスで実行されます。上記の場所と発行者を信頼する場合にのみ、このアプリケーションを実行してください。

上記の発行者と場所からのアプリケーションについては、次回から表示しない(D)



詳細情報(M)

実行(R)

取消

①



ヘルプ・リソース

- Javaとは
- MacのFAQ

Mac OS X Chrome

MacのJava 7でChromeを使用できないのはなぜですか

すべてのJavaのダウンロード

他のコンピュータまたはオペレーティング・システム用のJavaをダウンロードする場合は、下のリンクをクリックしてください。
[すべてのJavaのダウンロー](#)

Javaのバージョンを確認しました



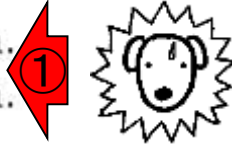
正常な設定です。

推奨バージョンのJavaがインストールされています (Version 8 Update 45).

MacでのJavaインストール

「正常な設定です」となっていることを確認したので、再度①ターミナル上で「java -version」とやってみると、なぜかまだエラーメッセージが！よくわからんがとりあえず②「詳しい情報」をクリックし、③たぶんこれだろうと思ったところをクリック。

```
kadota ~$ java -version
No Java runtime present, requesting install.
kadota ~$ java -version
No Java runtime present, requesting install.
kadota ~$
```



MacでのJavaインストール

左上のようなページに飛ばされ、よくわからないままページ下部に移動すると、OSごとにダウンロードできるものが存在する。①「Accept License Agreement」にチェックを入れる。

Oracle Technology Network > Java > Java SE > Downloads

Overview Downloads Documentation Community Technologies Training

Java SE Development Kit 8 Downloads

Thank you for downloading this release of the Java™ (JDK™). The JDK is a development environment for building and running applications using the Java programming language.

The JDK includes tools useful for developing and testing Java applications, compiling Java source code, and running Java programs on the Java platform.

See also:

- Java Developer Newsletter (tick the checkbox under News)
- Java Developer Day hands-on workshops (free and open to all)
- Java Magazine

JDK MD5 Checksum

JDK MD5 Checksum

Looking for JDK 8 on ARM?
JDK 8 for ARM downloads have moved to the JDK 8 for ARM download page.

Java SE Development Kit 8u45

You must accept the Oracle Binary Code License Agreement for Java SE to download this software.

Accept License Agreement Decline License Agreement

Product / File Description	File Size	Download
Linux x86	146.89 MB	jdk-8u45-linux-i586.rpm
Linux x86	166.88 MB	jdk-8u45-linux-i586.tar.gz
Linux x64	145.19 MB	jdk-8u45-linux-x64.rpm
Linux x64	165.24 MB	jdk-8u45-linux-x64.tar.gz
Mac OS X x64	221.98 MB	jdk-8u45-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	131.73 MB	jdk-8u45-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	92.9 MB	jdk-8u45-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	139.51 MB	jdk-8u45-solaris-x64.tar.Z
Solaris x64	95.88 MB	jdk-8u45-solaris-x64.tar.gz
Windows x86	175.98 MB	jdk-8u45-windows-i586.exe
Windows x64	180.44 MB	jdk-8u45-windows-x64.exe

Java SE Development Kit 8u45 Demos and Samples Downloads

You must accept the Oracle BSD License. to download this software.

Accept License Agreement Decline License Agreement

Product / File Description	File Size	Download
----------------------------	-----------	----------

MacでのJavaインストール

①Mac用のdmgファイルダウンロード

JDK MD5 Checksum

Looking for JDK 8 on ARM?
JDK 8 for ARM downloads have moved to the [JDK 8 for ARM download page](#).

Java SE Development Kit 8u45

You must accept the [Oracle Binary Code License Agreement for Java SE](#) to download this software.

Thank you for accepting the Oracle Binary Code License Agreement for Java SE; you may now download this software.

Product / File Description	File Size	Download
Linux x86	146.89 MB	jdk-8u45-linux-i586.rpm
Linux x86	166.88 MB	jdk-8u45-linux-i586.tar.gz
Linux x64	145.19 MB	jdk-8u45-linux-x64.rpm
Linux x64	165.24 MB	jdk-8u45-linux-x64.tar.gz
Mac OS X x64	221.98 MB	jdk-8u45-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	131.73 MB	jdk-8u45-solaris-sparcv9.tar.gz
Solaris SPARC 64-bit	92.9 MB	jdk-8u45-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	139.51 MB	jdk-8u45-solaris-x64.tar.Z
Solaris x64	95.88 MB	jdk-8u45-solaris-x64.tar.gz
Windows x86	175.98 MB	jdk-8u45-windows-i586.exe
Windows x64	180.44 MB	jdk-8u45-windows-x64.exe

Java SE Development Kit 8u45 Demos and Samples Downloads

You must accept the [Oracle BSD License](#). to download this software.

Accept License Agreement Decline License Agreement

MacでのJavaインストール

門田はこの段階で初めて、①ファイル名の先頭部分がjreとjdkという違いがあることに気づいた。直近でダウンロードしたのはjdkのほう。②実行。

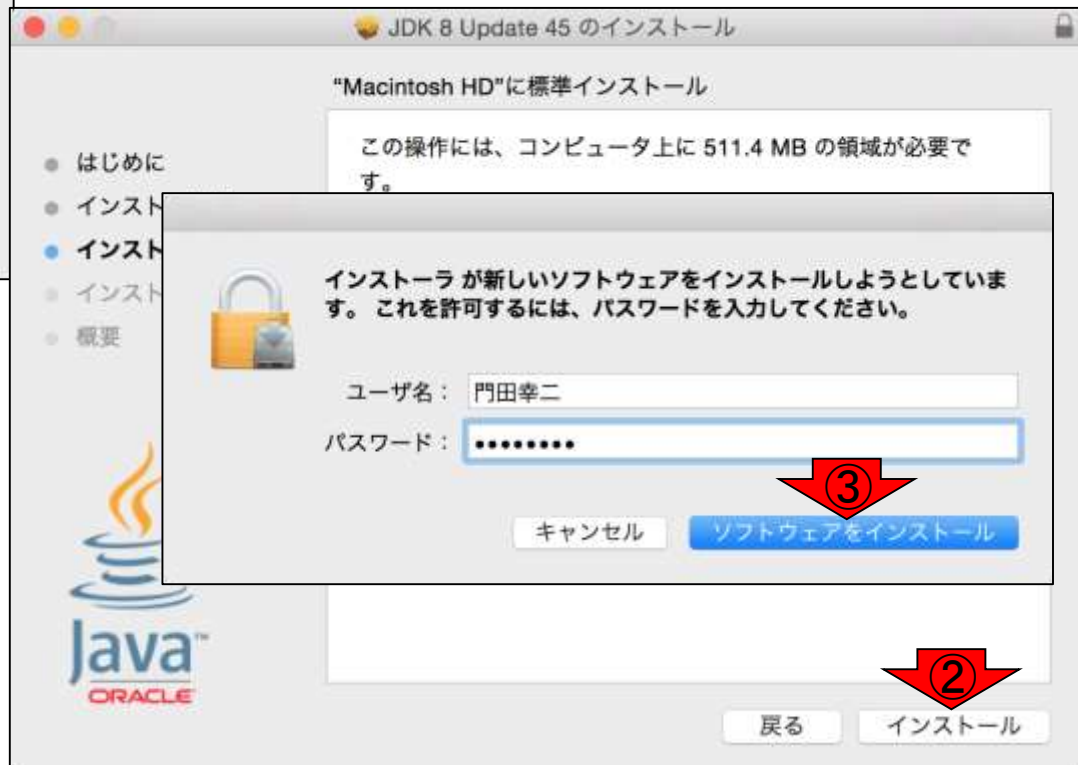
The image shows a Mac file manager window titled 'ダウンロード' (Downloads) with a table of files. A red arrow with the number '1' points to the file 'jdk-8u45-macosx-x64.dmg'. In the foreground, there is a window titled 'JDK 8 Update 45' showing the Java Development Kit installation interface. A red arrow with the number '2' points to the 'JDK 8 Update 45.pkg' icon.

名前	サイズ	追加日
jdk-8u45-macosx-x64.dmg	232.8 MB	今日 15:14
jre-8u45-macosx-x64.dmg	60.5 MB	今日 14:57
JSLAB4_1.sh	364 バイト	2015年4月30日 10:05
sample1.fasta.txt	23 バイト	2015年4月3日 15:13
R-3.1.3-mavericks...		
bio-linux-8-latest.is...		
Oracle_VM_Vir....v...		
VirtualBox-4.3....99...		

Java Development Kit
Double-click on icon to install
JDK 8 Update 45.pkg

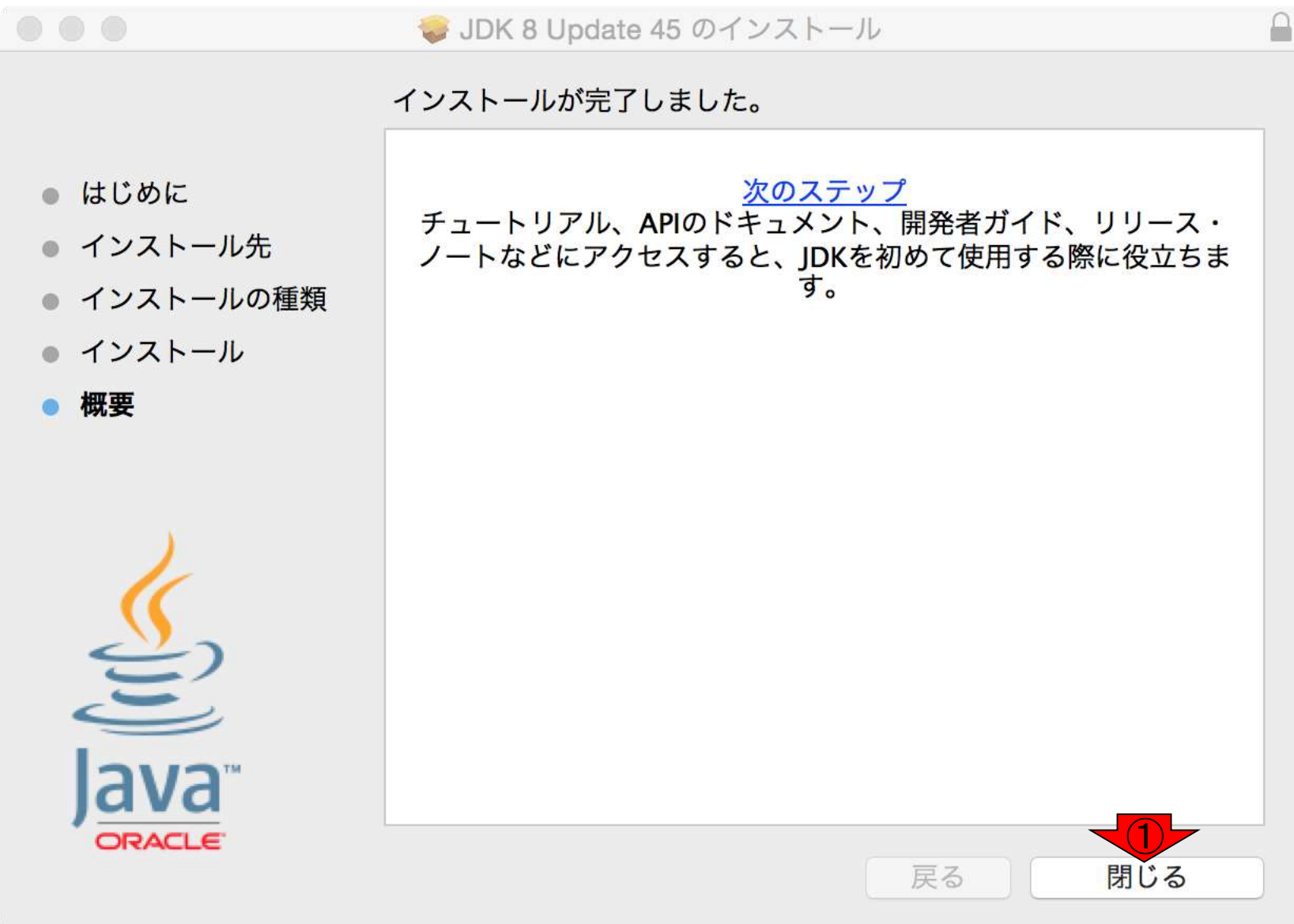
MacでのJavaインストール

①続ける、②インストール、③ログインパスワードを打ち込んで「ソフトウェアをインストール」。



インストールが完了しました、を確認して①閉じる。

MacでのJavaインストール



MacでのJavaインストール

再度①ターミナル上で「java -version」とやった結果。無事Java ver. 1.8.0_45がインストールされている。

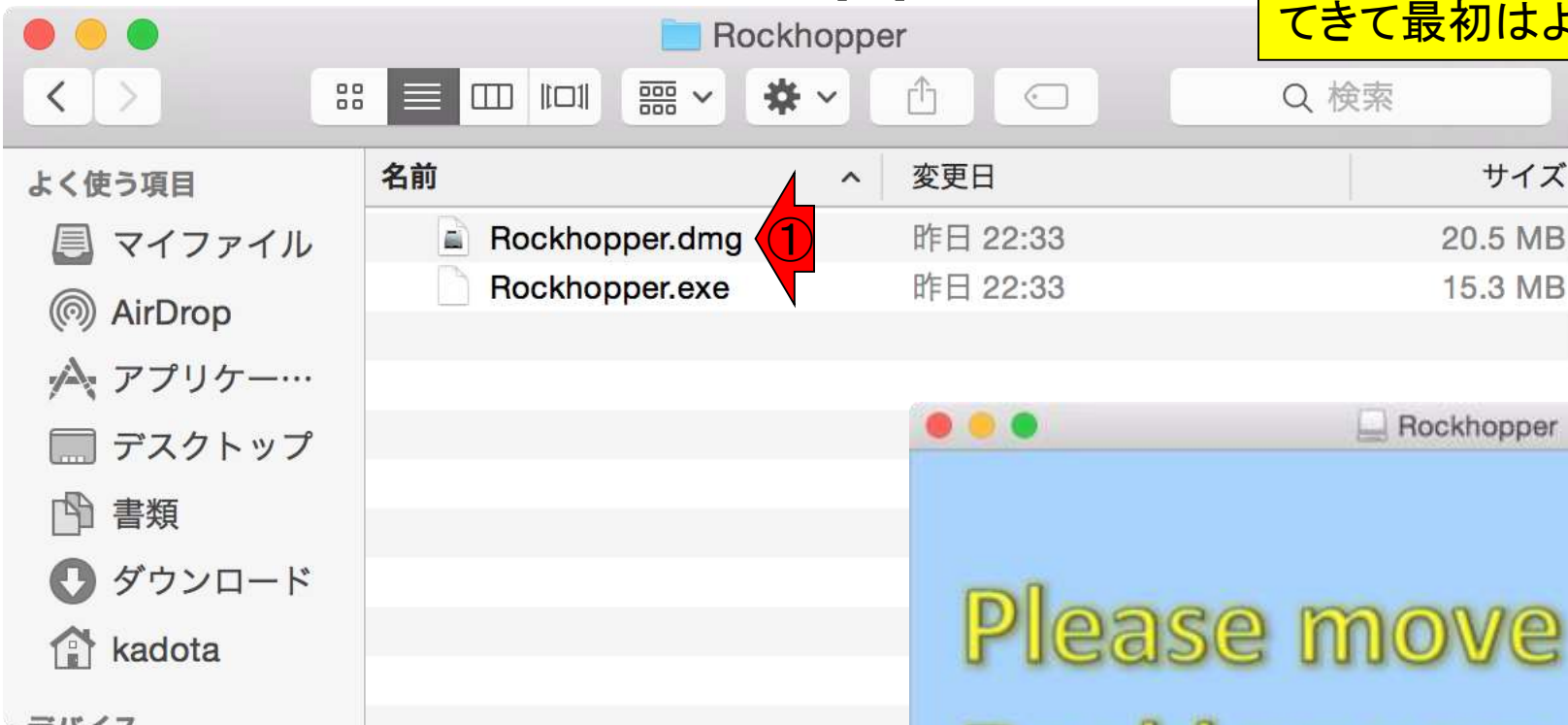


🏠 kadota — bash — 68x11

```
Last login: Tue May 12 15:02:35 on ttys000
agribio-macbook:~ kadota$ java -version
No Java runtime present, requesting install.
agribio-macbook:~ kadota$ java -version
No Java runtime present, requesting install.
agribio-macbook:~ kadota$ java -version ①
java version "1.8.0_45"
Java(TM) SE Runtime Environment (build 1.8.0_45-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.45-b02, mixed mode)
agribio-macbook:~ kadota$ █
```

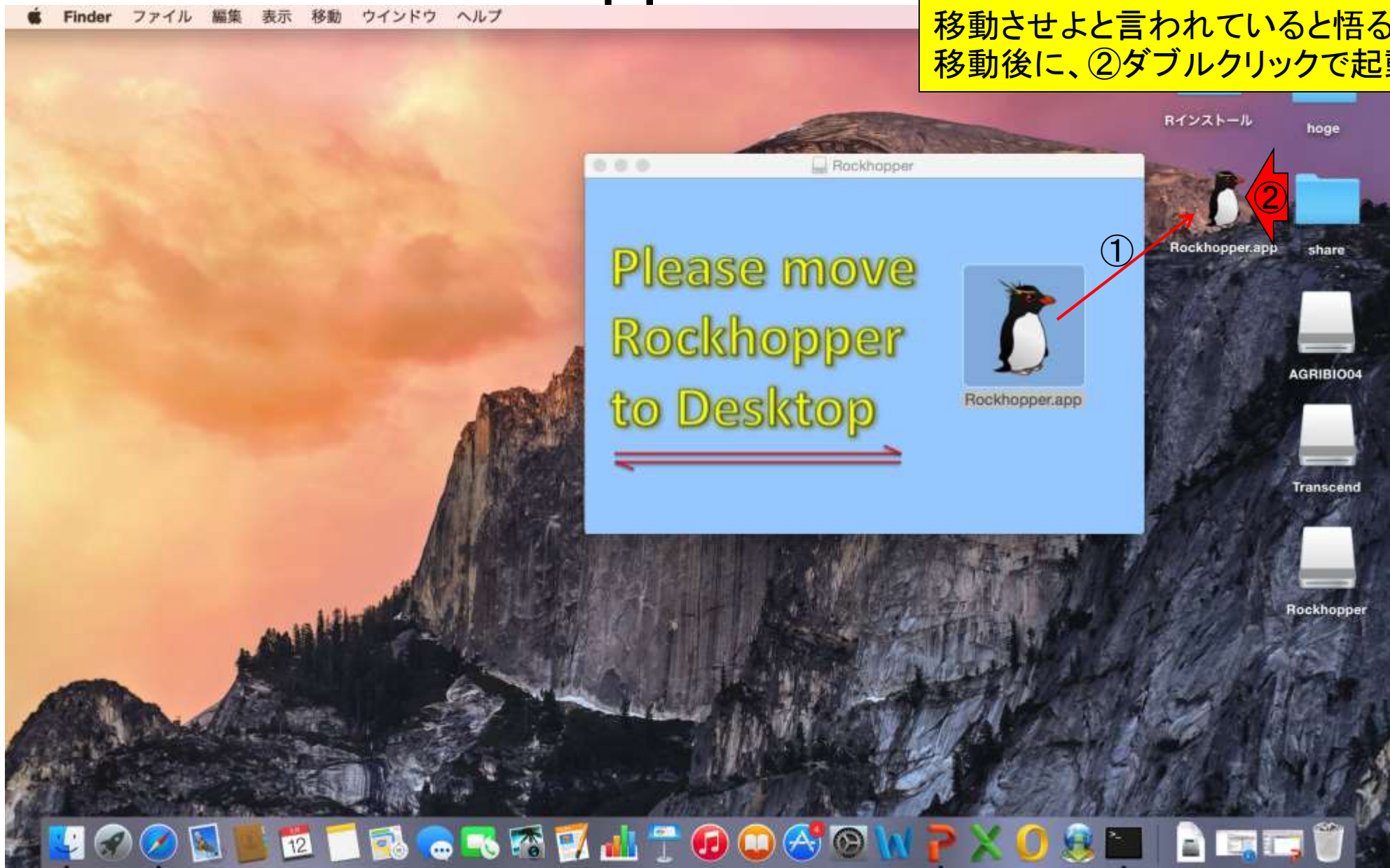
MacでのRockhopper実行

「hoge - Rockhopper」フォルダ中の
①Rockhopper.dmgファイルをダブル
クリック。②右下のような画面がで
てきて最初はよくわからなかった。



MacでのRockhopper実行

①何気なしにペンギン部分をマウスでいじって見たら、アイコンっぽくなっていたので、これをデスクトップに移動させよとされていると悟る。移動後に、②ダブルクリックで起動



MacでのRockhopper実行

起動せずに①のようなメッセージがでてしまう(爆)。気を取り直して②必要だと言われているものをダウンロード。



“Rockhopper.app”を開くには、以前の Java SE 6 ランタイムをインストールする必要があります。

以前の Java SE 6 のダウンロード Web サイトにアクセスするには、“詳しい情報...”をクリックしてください。

詳しい情報...

OK

①

Apple Inc.

Java SE Development Kit 8 - Downloads Java for OS X 2014-001

ストア Mac iPhone Watch iPad iPod iTunes サポ

 **Java for OS X 2014-001**

ダウンロード

②

Java for OS X 2014-001 ではインストールに関する改善が加えられています。また、Java for OS X の以前のバージョンの修正内容がすべて引き継がれています。このパッケージにより、Java for OS X 2013-005

MacでのRockhopper実行

(どうにかならんかと思いつつ)①
3つめのJava関連dmgファイルを
ダブルクリック。②pkgをダブルク
リック。

ダウンロード

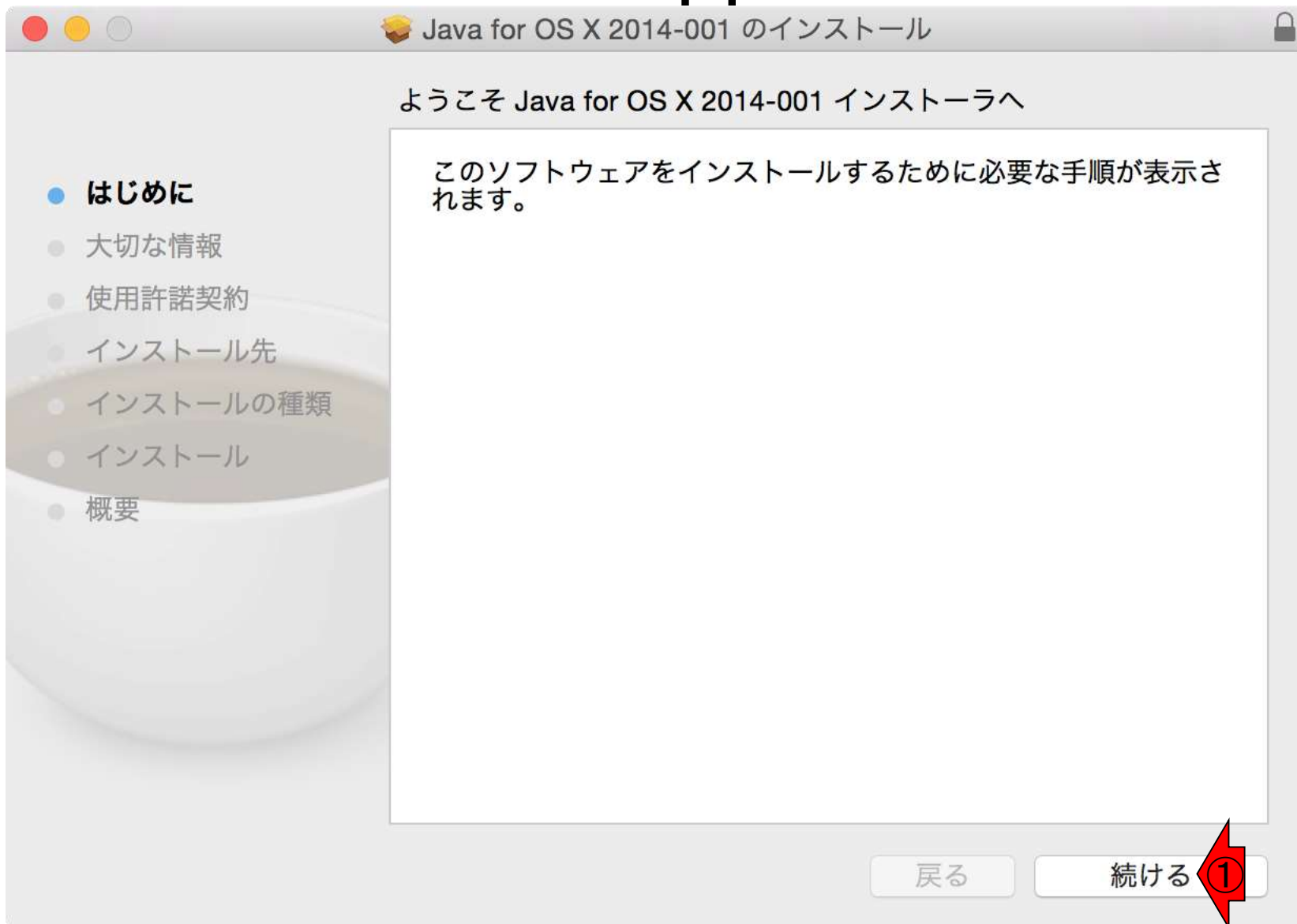
検索

よく使う項目	名前	サイズ	追加日	種類
マイファイル	JavaForOSX2014-001.dmg	66.9 MB	今日 15:30	ディスクイメージ
AirDrop	jdk-8u45-macosx-x64.dmg	232.8 MB	今日 15:14	ディスクイメージ
アプリケー...	jre-8u45-macosx-x64.dmg	60.5 MB	今日 14:57	ディスクイメージ
デスクトップ	JSLAB4_1.sh	364 バイト	2015年4月30日 10:05	シェルスクリプト
書類	sample1.fasta.txt	23 バイト	2015年4月3日 15:13	テキスト
ダウンロード	R-3.1.3-mavericks.pkg	70.9 MB	2015年3月31日 20:45	インストー...アー
kadota	bio-linux-8-latest.iso	3.52 GB	2014年11月26日 15:35	ISO ディスクイメ
デバイス	Oracle_VM_Vir....vbox-extpack	15.9 MB	2014年11月26日 10:57	VirtualBox.app 書
リモートデ...	VirtualBox-4.3....996-OSX.dmg	114.1 MB		
AGRIBIO04 ▲				

Java for OS X 2014-001

JavaForOSX.pkg

MacでのRockhopper実行



MacでのRockhopper実行

Java for OS X 2014-001 のインストール

大切な情報

Java for OS X 2014-001 ではインストールに関する改善が加えられています。また、Java for OS X の以前のバージョンの修正内容がすべて引き継がれています。このパッケージにより、Java for OS X 2013-005 に含まれるのと同じバージョンの Java 6 がインストールされます。

Java for OS X 2012-006 以降がインストールされていないシステムにこのアップデートを適用すると、Java SE 6 アプレットプラグインが無効になります。Web ページでアプレットを使用するには、“プラグインが見つかりません”ラベルの領域をクリックして、Oracle の提供する Java アプレットプラグインの最新バージョンをダウンロードしてください。

このアップデートをインストールする前に、Java アプリケーションを終了してください。

このアップデートについて詳しくは、次の Web サイトを参照してください：http://support.apple.com/kb/HT6133?viewlocale=ja_JP

このアップデートのセキュリティコンテンツについて詳しくは、次の Web サイトを参照してください：http://support.apple.com/kb/HT1222?viewlocale=ja_JP

プリント...

保存...

戻る

続ける

①

- はじめに
- **大切な情報**
- 使用許諾契約
- インストール先
- インストールの種類
- インストール
- 概要

MacでのRockhopper実行

Java for OS X 2014-001 のインストール

このソフトウェアのインストールを続けるには、ソフトウェア使用許諾契約の条件に同意する必要があります。

インストールを続けるには、“同意する”をクリックしてください。インストールをキャンセルしてインストーラを終了する場合は、“同意しない”をクリックしてください。

使用許諾契約を読む

同意しない

同意する

②

「約」
で、各
契約

の各条項に同意されない場合は、当該Appleソフトウェアを取得された場所へ返却の上、払い戻しを受けることができます。電子的にAppleソフトウェアにアクセスした場合は、「同意しません/承諾しません」ボタンをクリックしてください。お客様が購入されたハードウェアに含まれるAppleソフトウェアについて払い戻しを受けるためには、すべてのハードウェア及びソフトウェアを返却しなければなりません。

重要な通知： このソフトウェアは、マテリアルを複製するために使用する限りにおいて、著作権のないマテリアル、お客様が著作権を有する、または複製権を有するか複製を法的に認められたマテリアルの複製のためにのみ、お客様に対してライセンスが付与されるものです。マテリアルの複製権に関してご不明な点がございましたら、お客様の法律アドバイザーにご相談ください。

1. 総則

A. 本契約書が添付されているディスク、読み出し専用メモリー、その他の記録媒体またはその他あらゆる形態上の、Appleおよび任意のサードパーティのソフトウェア、書類および一切のフォント（以下「Appleソフトウェア」といいます）は、Apple Inc.（以下「Apple」といいます）が、お客様に対して、本契約条件に従う場合に限りライセンスを付与するものであり、販売するものではありません。Appleおよび/またはApple

プリント...

保存...

戻る

続ける

①

MacでのRockhopper実行

①インストール。②ログインパスワードを打ち込んで、「ソフトウェアのインストール」

Java for OS X 2014-001 のインストール

“Macintosh HD”に標準インストール

この操作には、コンピュータ上に 77 MB の領域が必要です。

ディ
ルス



インストーラが新しいソフトウェアをインストールしようとしています。これを許可するには、パスワードを入力してください。

ユーザ名： 門田幸二

パスワード： ●●●●●●

キャンセル

ソフトウェアをインストール

インストール先を変更...

戻る

インストール

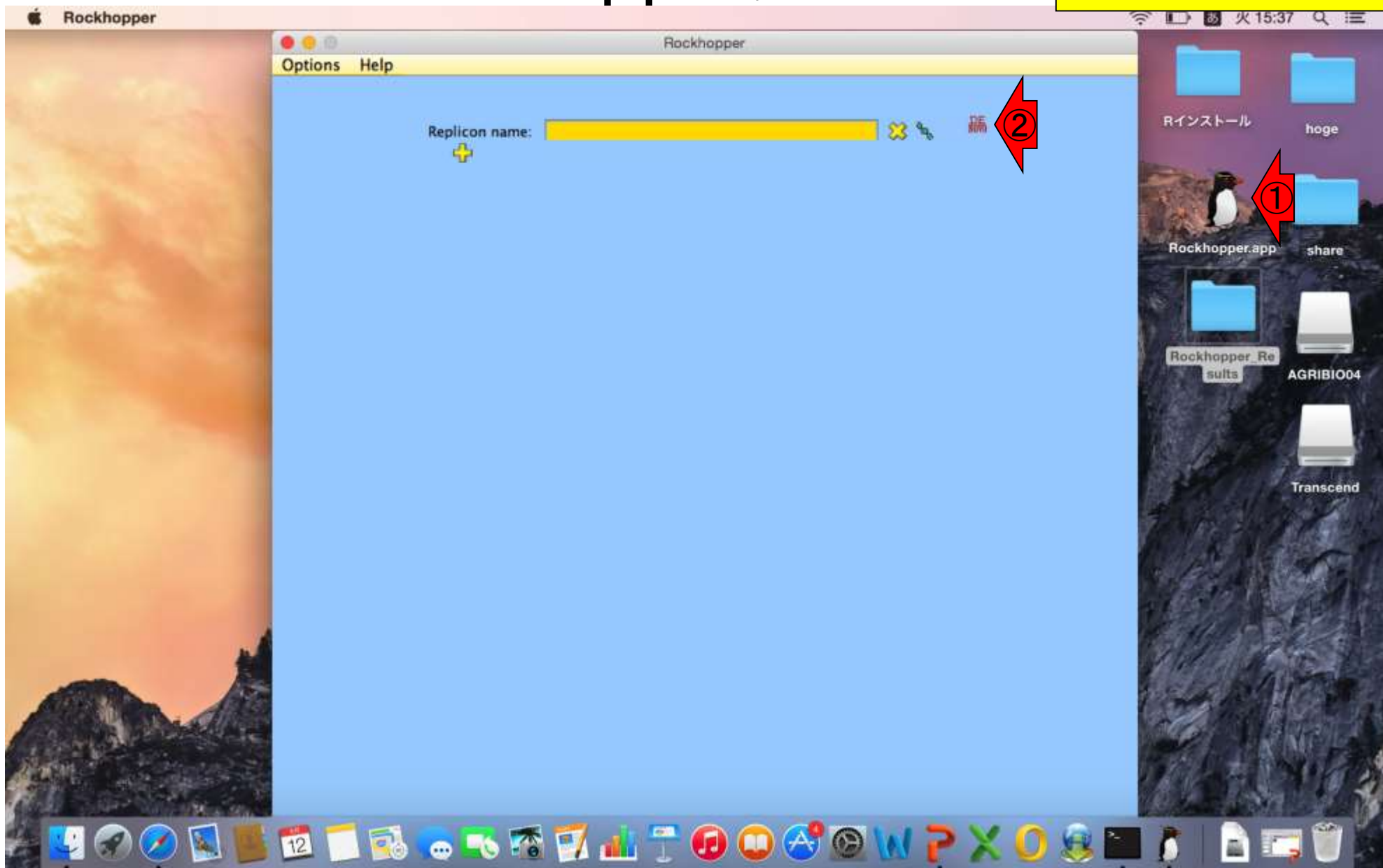
- はじめに
- 大切な情報
- 使用許諾契約
- インストール先
- **インストールの種類**
- インストール
- 概要

MacでのRockhopper実行



MacでのRockhopper実行

気を取り直して、①起動。
②赤い「DE NOVO」と書いてあるところをクリック。



MacでのRockhopper実行

例題用ファイル(Example_denovo.fastq.gz)を指定して④開く。

Rockhopper

Options Help

Experiment Name: 1

Replicate #1: [Yellow field] BROWSE

開く

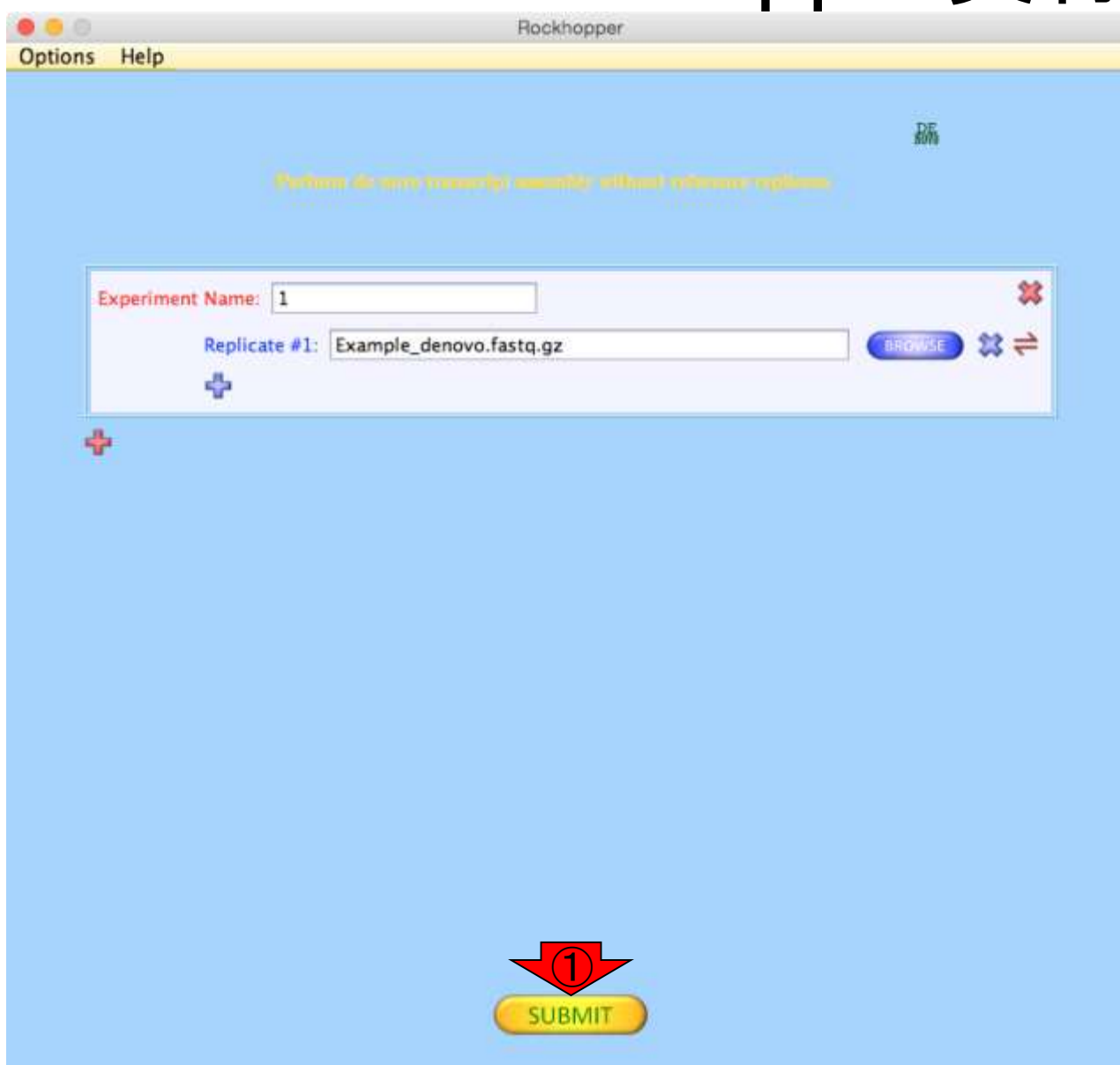
名前	変更日
20150512.pdf	2015年5月12日 13:45
Example_denovo.fastq.gz	2015年5月11日 22:32
GPL1355-14795.txt	2015年5月11日 16:40
GSE7623	2015年5月12日 14:20
r.html	2015年5月11日 21:59
r_seq.html	2015年5月9日 22:52
Rockhopper	2015年5月12日 14:20
SRR616268sub_1.fastq.gz	2015年5月11日 16:52
SRR616268sub_2.fastq.gz	2015年5月11日 16:53

ファイルフォーマット: すべてのファイル

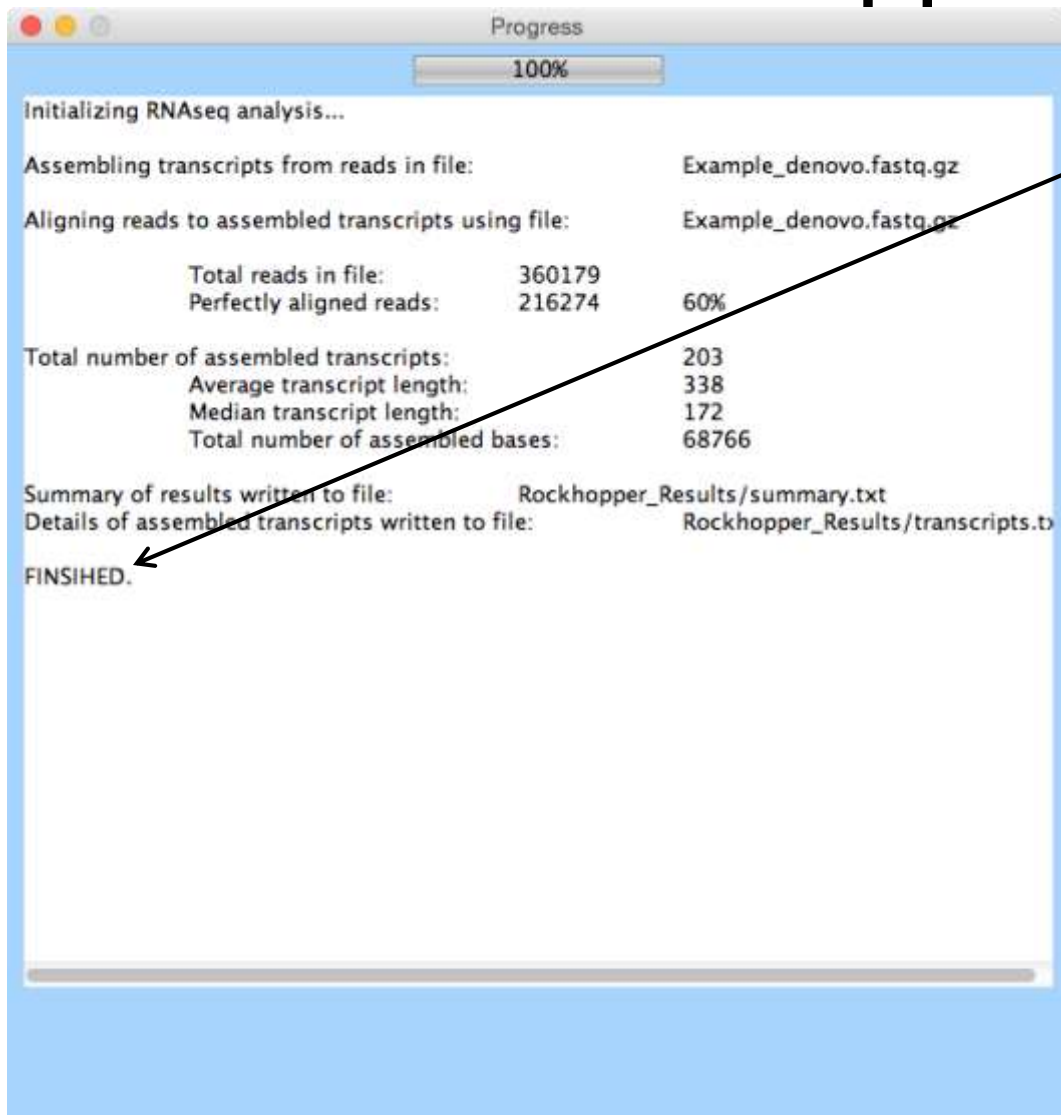
キャンセル 開く

SUBMIT

MacでのRockhopper実行



MacでのRockhopper実行



MacでのRockhopper実行

この場合、デスクトップにRockhopper.appがあるディレクトリ上に①Rockhopper_resultsというフォルダが自動生成され、その中にアセンブル結果ファイルができる。

