

基礎生物学研究所
ゲノムインフォマティクス・トレーニングコース2020年 6月

RNA-seq 入門

NGS の基礎から *de novo* 解析まで
準備編・実践編

June 4-5, 2020 @ NIBB (Okazaki)

重信秀治 / Shuji Shigenobu

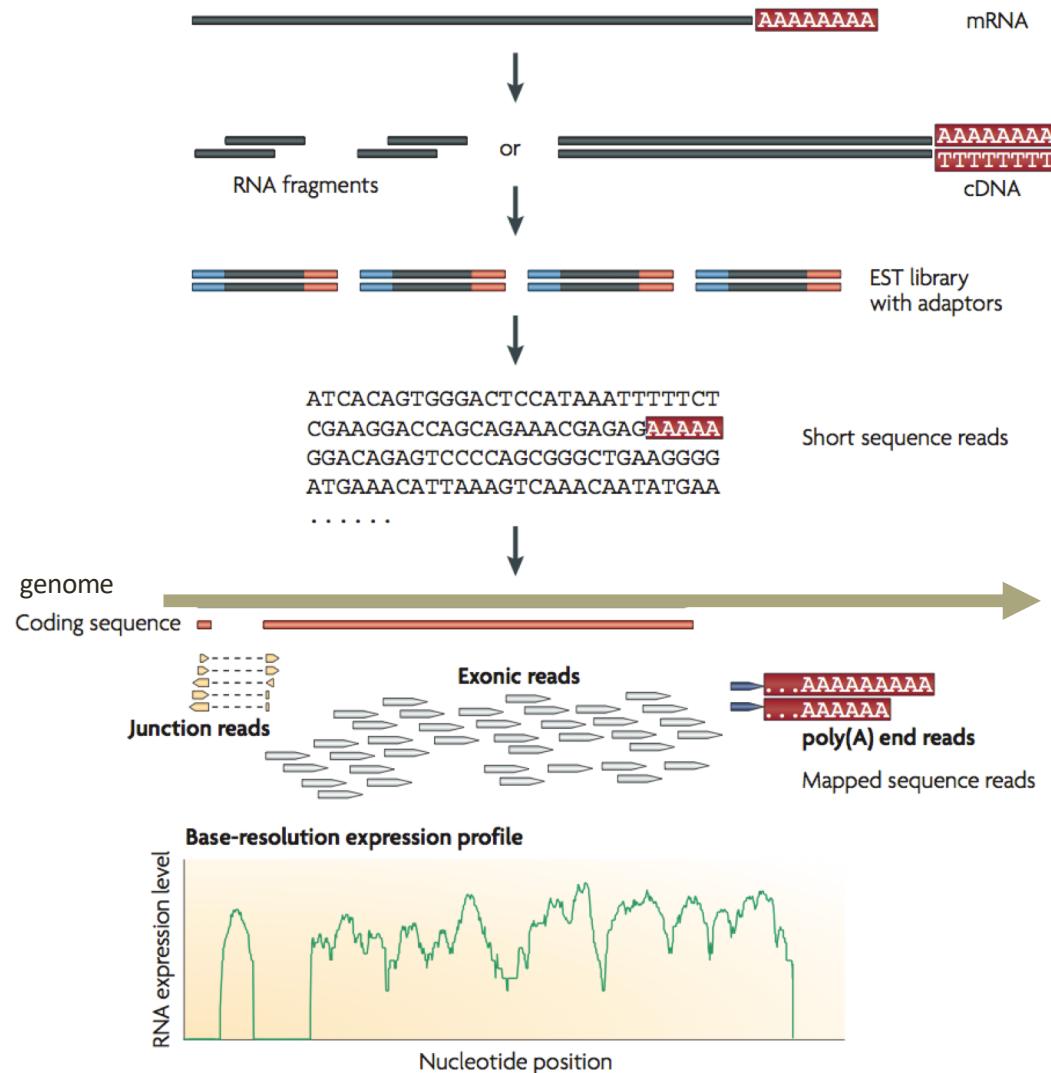
- サポートWiki

<https://github.com/nibb-gitc/gitc2020jun-rnaseq/wiki>

オンライン講義のための準備

- Lecture with **Zoom**
 - 接続の方法
 - 質問の方法
- Support with **Discord**
 - テキストチャンネル #general
 - ボイスチャンネル

RNA-seq



(Wang 2009 with modifications)



All aTwitter over an
Internet study p. 683

The extragalactic background's
uneven glow pp. 694 & 732

A cellular target for human
norovirus pp. 700 & 755

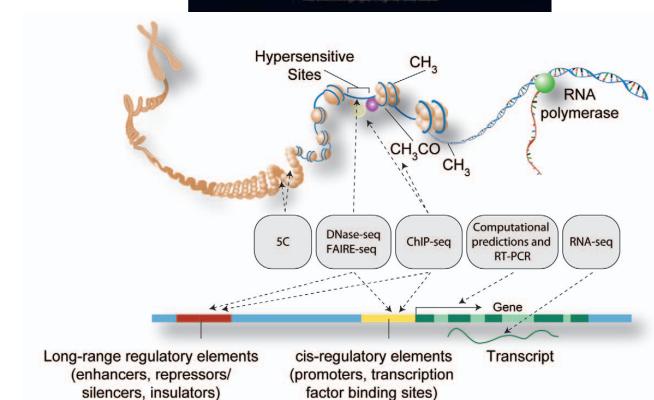


Two major goals

- Gene cataloguing
- Gene expression analysis

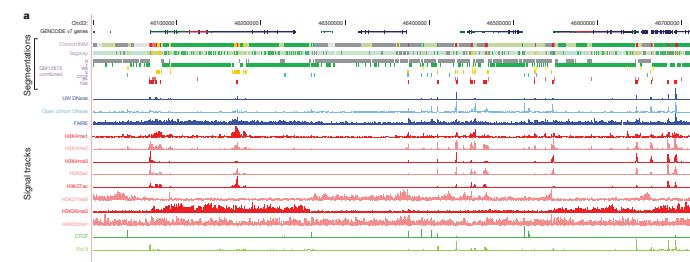
RNA-seq is unraveling complexities of eukaryotic transcriptomes in model and non-model organisms

- Gene expression analysis
- Novel gene discovery (model org.)
 - Coding and non-coding genes
- Gene cataloguing (non-model org.)
- Anti-sense transcripts
- RNA editing
- Novel splicing variants & fusion genes
- Allele-specific expression
- Single-cell RNA-seq for cell population analysis



Beyond transcriptome

- DB for proteome analysis
- SNP finding
- *and more ...*



- Problems in RNA-seq informatics
 - Too difficult for ordinary biologists
 - Why difficult?
 - No single standard pipeline established. Not well-documented.
 - Too big.
 - Requires fundamental skills of bioinformatics.

基礎生物学研究所

ゲノムインフォマティクス・トレーニングコース

RNA-seq 入門

NGS の基礎から *de novo* 解析まで
準備編・実践編

生物情報学を必ずしも専門としない生物研究者が、RNA-seqで得られる網羅的な遺伝子発現データを解析し、生物学的な情報を抽出するための、基礎的技術と考え方を身に付けることを目的としたコース。

<理念>

- For biologists
- 基礎力と即戦力をバランス良く

RNA-seq 入門

準備編

- UNIX基礎
- シェルスクリプト
- R入門
- NGS基本データフォーマット
- NGS基本ツール
- テキストファイル処理
- 統計学入門

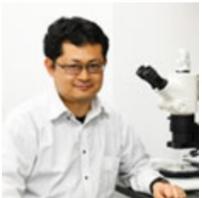
実践編

- NGS 基本フォーマット・ツール
復習
- NGS データ可視化ツール
- RNA-seq パイプライン：基礎
- RNA-seq パイプライン：トランスクリプトベース
- RNA-seq パイプライン：ゲノムベース
- RNA-seq パイプライン：de novo
- 多変量解析
- 機能アノテーションとGO解析
- 演習

「実践編」の位置づけ・目標

- 「準備編」で身に付けたUNIX, R, NGS等の基本スキルを基盤に、RNA-seqデータ解析の一連の基本パイプラインを自分自身で実行できるスキルを身に付ける。
- RNA-seq基本パイプラインを自分のプロジェクト向けにアレンジできる程度の応用力を身に付ける。

講師



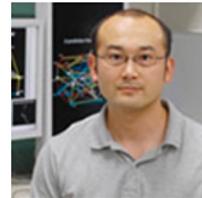
重信 秀治

基礎生物学研究所
生物機能解析センター
教授
【オーガナイザー】



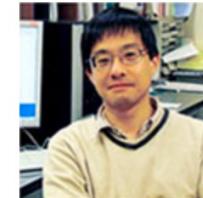
内山 郁夫

基礎生物学研究所
生物機能解析センター
准教授
【オーガナイザー】



佐藤 昌直

北海道大学大学院
農学研究院
助教



山口 勝司

基礎生物学研究所
生物機能解析センター
技術主任

[フロアサポート、事務局]

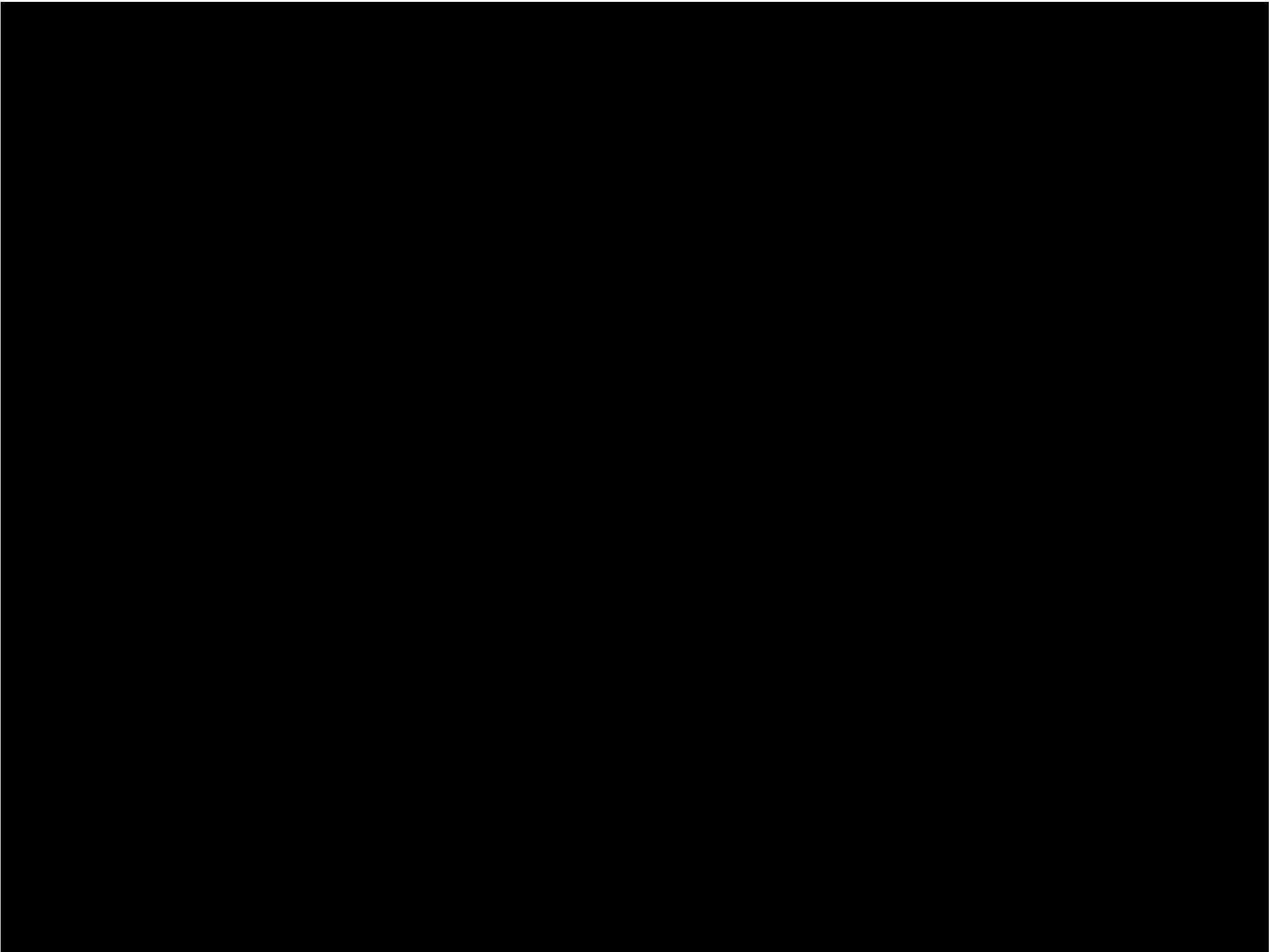
- ・ 杉浦、中村、西出、頼本、森、牧野

自己紹介

- 専門: ゲノム進化学、共生生物学
- 実験生物学とバイオインフォマティクス
- トランскriプトーム: ESTの時代より大規模遺伝子発現解析に従事。EST, microarray, RNA-seq。ショウジョウバエから非モデル生物まで。
- 基生研・生物機能解析センターにおいて年間約50件のNGS関連の共同利用研究を実施。
- NGS関連原著論文：多数
- Editorial Board Member of *Scientific Reports* (Genetics and Genomics section)
- NGS, RNA-seq論文のハンドリング、peer review を多く手がける。

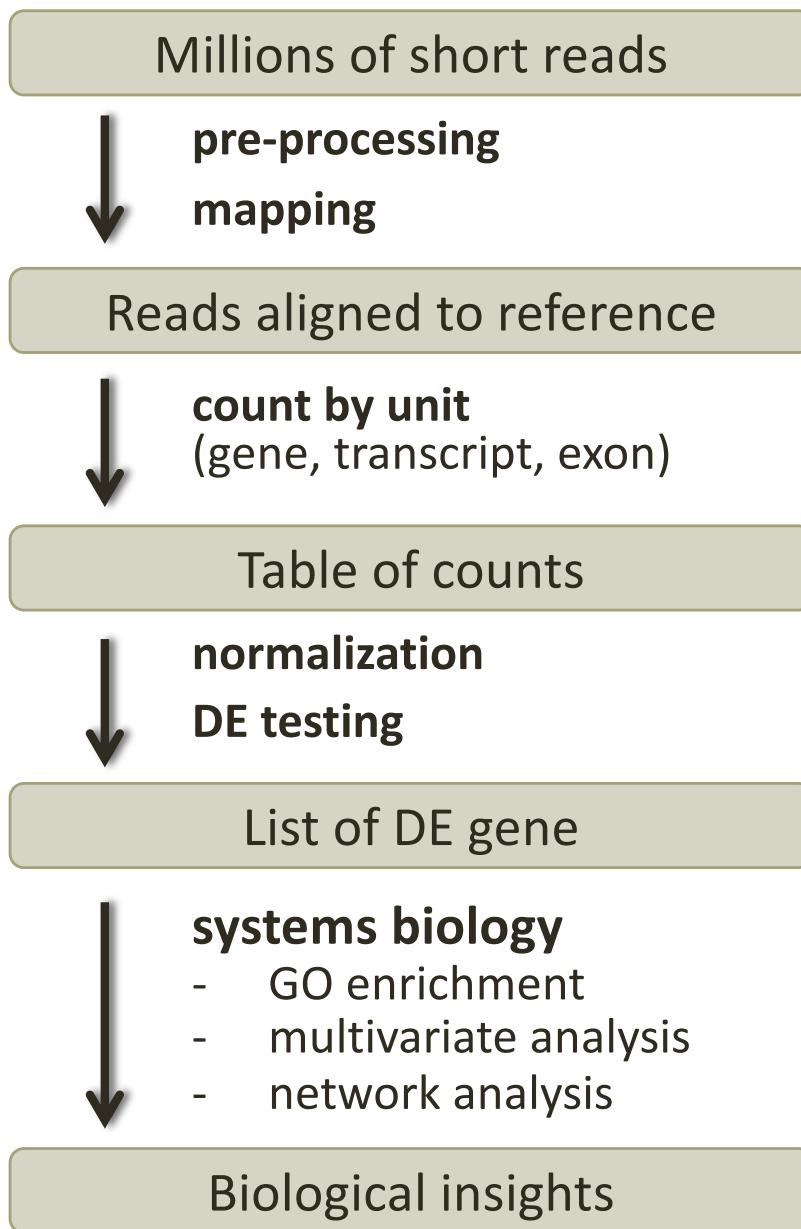


nature publishing group npg



RNA-seq analysis pipeline for DE

Differential Expression analysis



解析ツールの現状: RNA-seq

- ・全てのプロセスをこなせる万能ツールはない。
- ・それぞれのステップに特化したツール群が次々に登場している。

基本戦略

- ・各ステップに最適なツールをチョイス、組み合わせた、解析パイプラインの構築。

Pipeline

- ・本コースで学ぶオススメの2つのパイプライン
 - ・Transcriptome-based: Salmon/edgeR
 - ・Genome-based: HISAT2/StringTie/edgeR

RNA-seq 入門

準備編

- UNIX基礎
- シェルスクリプト
- R入門
- NGS基本データフォーマット
- NGS基本ツール
- テキストファイル処理
- 統計学入門

実践編

- NGS 基本フォーマット・ツール
復習
- NGS データ可視化ツール
- RNA-seq パイプライン：基礎
- RNA-seq パイプライン：トランスクリプトベース
- RNA-seq パイプライン：ゲノムベース
- RNA-seq パイプライン：de novo
- 多変量解析
- 機能アノテーションとGO解析
- 演習

オススメ勉強法

- ・ コマンドやプログラムは自分で試してみる。
copy & pasteでなくタイピングすること。
(熊楠メソッド)
- ・ 気軽に質問する。講師はもちろん、(隣や前後の受講生にも)。その一方で、ヘルプやマニュアルドキュメントをうまく活用する。
- ・ 自分の研究との接点を常に意識する。自分の研究に応用する。

- サポートWiki

<https://github.com/nibb-gitc/gitc2020jun-rnaseq/wiki>

- テキスト
- 練習問題
- 参考情報
- コース終了後も公開

オンライン講義のための準備

- Lecture with **Zoom**
 - 接続の方法
 - 質問の方法
- Support with **Discord**
 - テキストチャンネル #general
 - ボイスチャンネル

環境構築・接続

NIBBのサーバ（リモート環境）とお手元のPC（ローカル環境）を併用して実習を行います。

- UNIX系のソフトウェアは、NIBBの大型計算サーバbias5にログインして、Linux環境で実行します。
 - ターミナルソフトウェアをローカルPCに準備
 - ssh接続、scpコマンド（VPN接続）
 - UNIXコマンドの基礎知識
- GUIソフトウェアはローカルPCで実行します。
 - IGV, ErmineJ, 表計算ソフトウェアなど
 - RもローカルPCでの実行を基本とします。

アカウント・演習データ

- bias5アカウント情報
 - account : courseXX (XXはお一人ずつ異なります)
 - password: (通知ズミ)
- 演習用データの場所
bias5: ~/data/
- 演習用データを皆さんのお一カ月環境にもコピーしましょう

<お一カ月環境のターミナルから>

```
$ scp -r courseXX@bias5.nibb.ac.jp:data ./
```

RNA-seq 入門

さあ、始めましょう

準備編

- UNIX基礎
- シェルスクリプト
- R入門
- NGS基本データフォーマット
- NGS基本ツール
- テキストファイル処理
- 統計学入門

実践編

- NGS 基本フォーマット・ツール
復習
- NGS データ可視化ツール
- RNA-seq パイプライン：基礎
- RNA-seq パイプライン：トランスクリプトベース
- RNA-seq パイプライン：ゲノムベース
- RNA-seq パイプライン：de novo
- 多変量解析
- 機能アノテーションとGO解析
- 演習

