

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2021年 9月

RNA-seq 入門

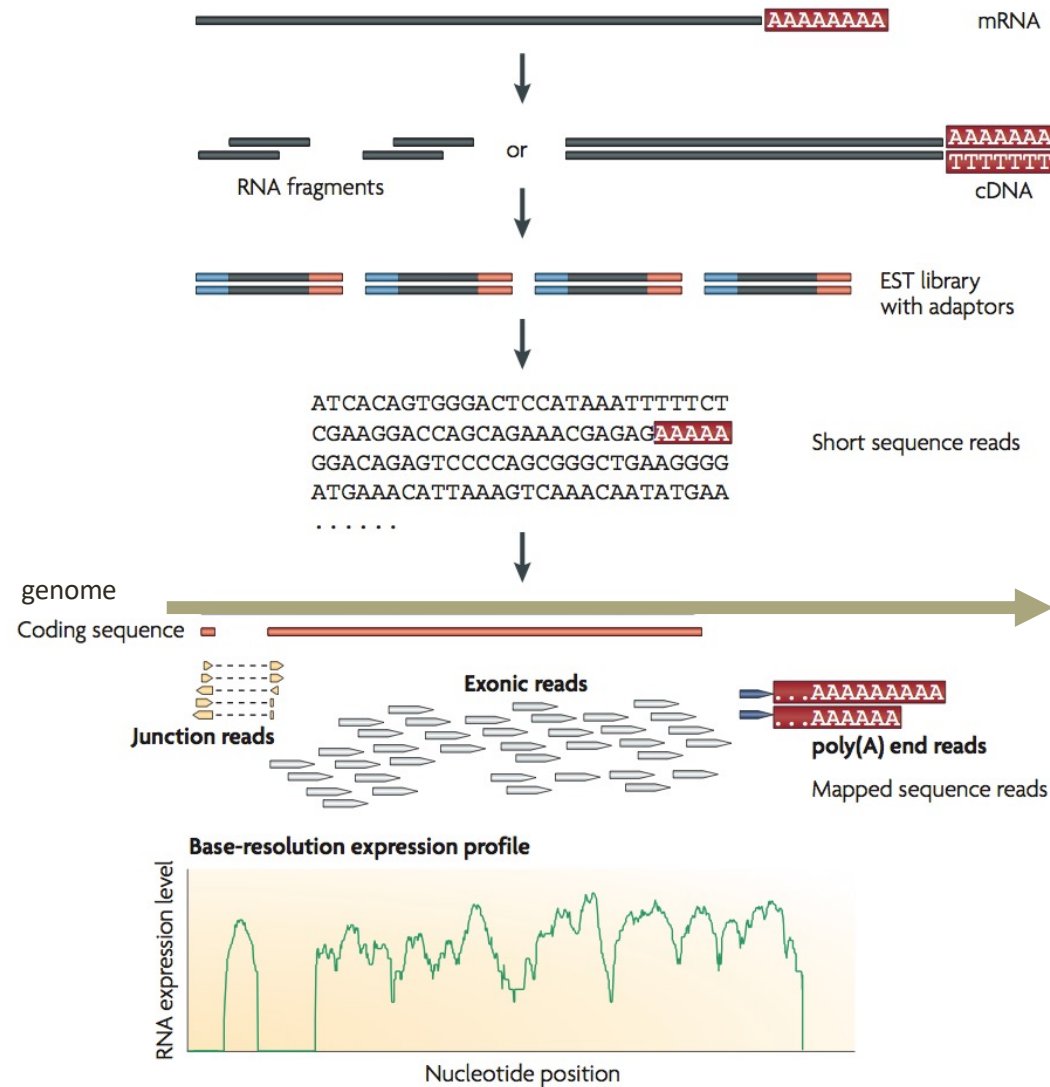
RNA-seq 解析パイプライン

September 15-16, 2021 online

重信秀治 / Shuji Shigenobu

<https://github.com/nibb-gitc/gitc2021sep-rnaseq/wiki>

RNA-seq



(Wang 2009 with modifications)

Two major goals

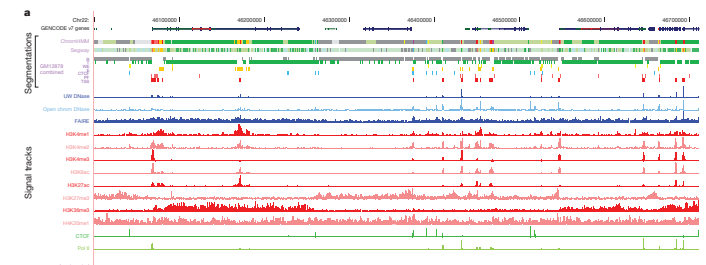
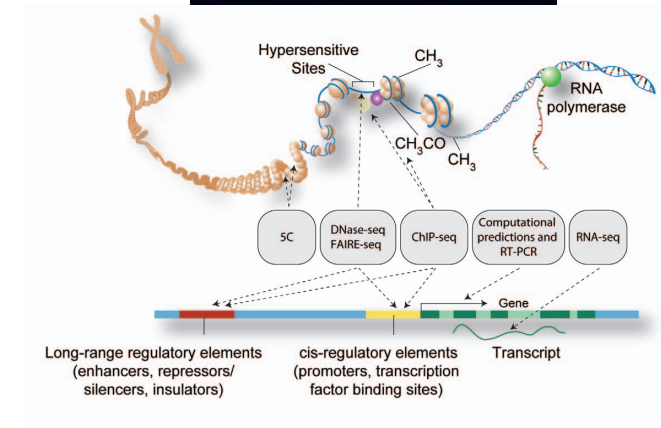
- Gene cataloguing
- Gene expression analysis

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
- Spatial transcriptome

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

RNA-seq 解析パイプライン

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

NGS解析入門（2021年8月）

- UNIX基本コマンド
- R入門
- 統計学入門
- NGS基本データフォーマット
- NGS基本ツール
- エディタとスクリプト
- テキストファイル処理

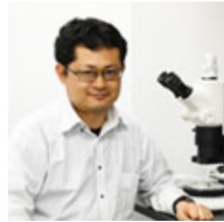
RNA-seq解析パイプライン（今回）

- 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
- 実践演習

RNA-seq解析パイプラインの位置づけ・目標

- 「NGS解析入門」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】



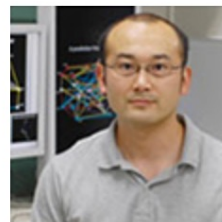
内山 郁夫

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】



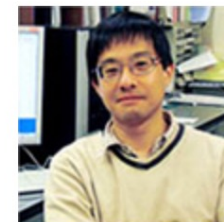
西出 浩世

基礎生物学研究所
生物機能解析センター
技術主任



佐藤 昌直

北海道大学大学院
農学研究院
助教



山口 勝司

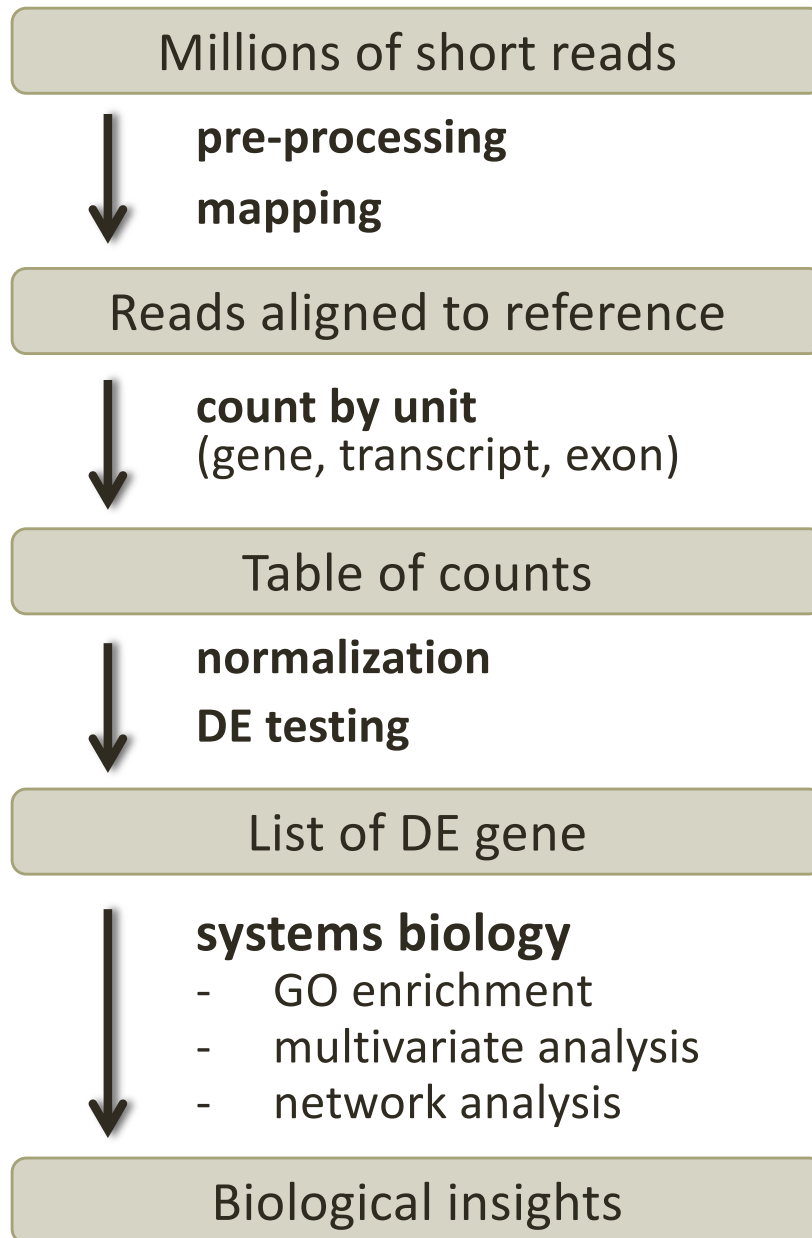
基礎生物学研究所
生物機能解析センター
技術主任

[フロアサポート、事務局]

- 杉浦、森、中村、森、牧野

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多く開発されてきたが、近年成熟の域に入ってきたように見受けられる。

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

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、Slackで受講生どうし教えあう！。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。

質問力を向上しよう

- 研究に関する質問は通常の研究活動のそれと同じ。
- コンピュータに関する質問に際しては、質問者にも工夫と心構えが必要。
 - 自分が実現しなかったこと（目的・ゴール）を明示。
 - 自分が実施したことを正確に伝える。例えば実行したコマンド、パラメータ詳細。問題の起きた環境（OSやソフトウェアのバージョンなど）。
 - 何が起きたかを正確に伝える。エラーメッセージは情報の宝庫。
 - 自分なりの原因の予想や、解決策案を提示するとなお良い。自分なりに試行錯誤したことがあればその情報も。
- コマンドヘルプやmanページを積極的に活用する。
- 質問が上手な人は、成長が早いです。

講義資料@wiki

- サポート Wiki

<https://github.com/nibb-gitc/gitc2021sep-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 **Slack**
 - チャンネル #講義質問_questions

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(IGVなど)やRはローカルPCで実行します。
 - Rはbias5にもインストールされています。

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: (通知済み)
- 演習用データの場所

bias5: ~/gitc/data/
- 演習用データは以下からもダウンロード可能。
 - <https://www.nibb.ac.jp/cproom/gitc/gitc202108.html>
 - Webブラウザ上でダウンロード。もしくはコマンドラインから、`wget`や`curl`などのコマンドを利用。