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2017年 2-3月

RNA-seq 入門

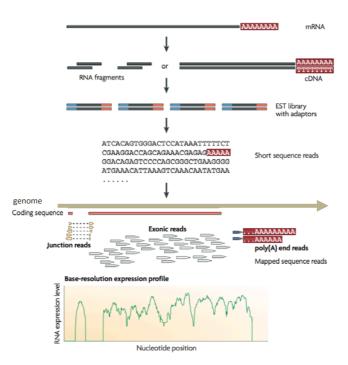
NGS の基礎から de novo 解析まで 準備編・実践編

March 9-10, 2017 @ NIBB (Okazaki)

重信秀治 / Shuji Shigenobu

• サポートWiki https://github.com/nibb-gitc/gitc2017mar-rnaseq/wiki

RNA-seq



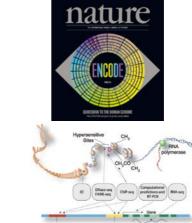
(Wang 2009 with modifications)

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

Beyond transcriptome

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





Two major goals

- Gene cataloguing
- Gene expression analysis

RNA-seq analysis pipeline for DE

<u>Differential</u> <u>Expression</u> analysis

Millions of short reads

pre-processing mapping

Reads aligned to reference

count by unit (gene, transcript, exon)

Table of counts

normalization
DE testing

List of DE gene

systems biology

- GO enrichment
- multivariate analysis
- network analysis

Biological insights

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
 - Genome-based: TopHat/Cufflinks
 - Transcriptome-based: Trinity/Bowtie/eXpress/edgeR

Biologist が身に付けるべき 6つの informatics スキル

- ・初級)UNIXの基礎
- 初級) 統計的な考え方と技術
- ・初級)業界標準のツール
- 初級) データ可視化
- ・中級) 初歩的なプログラミング
- ・中級) データベース