Bioinformatics Summer School 2014

Konstantin Okonechnikov
Max Planck Insitute For Infection Biology

Practical RNA-seq for biologists

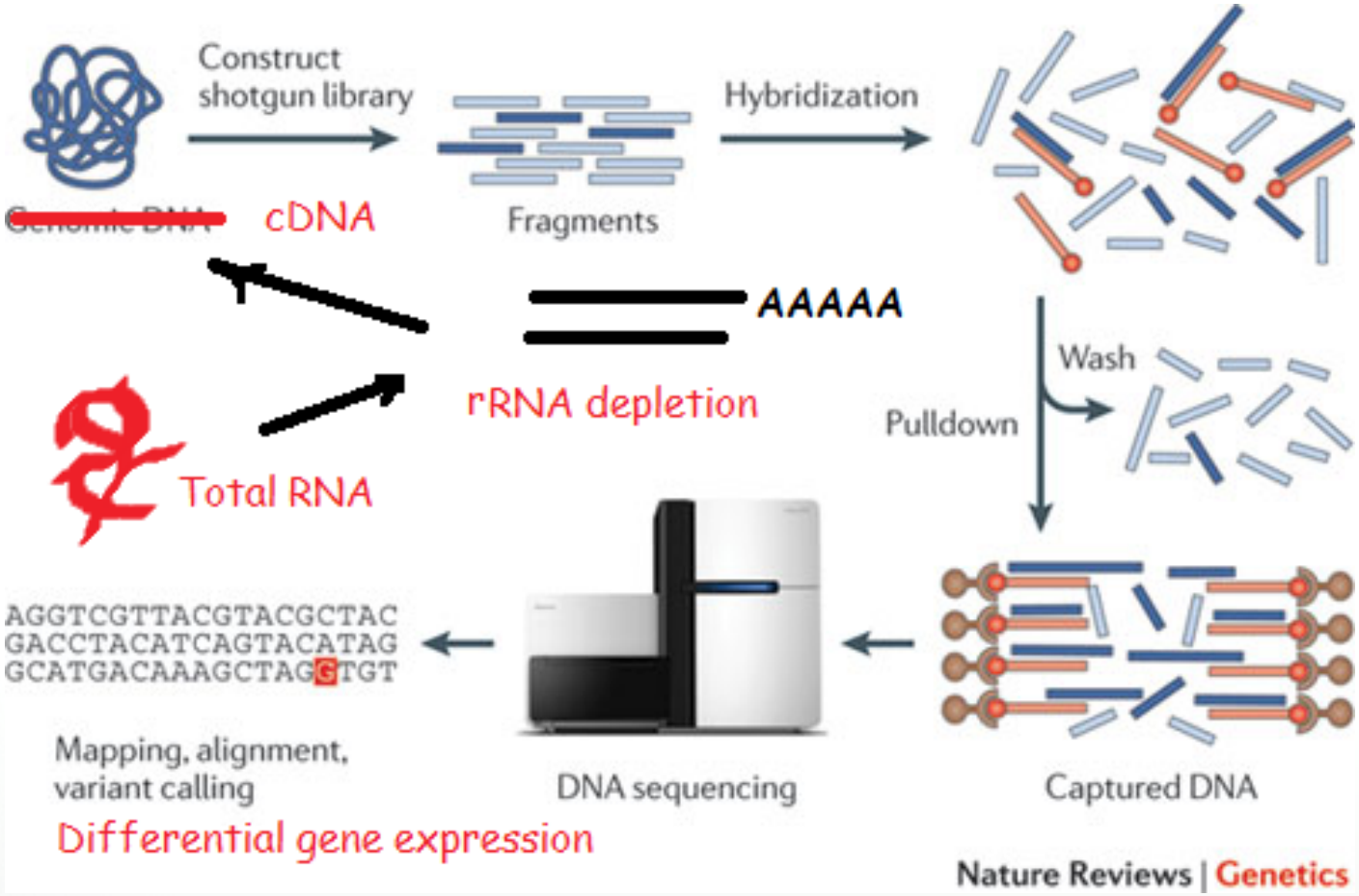
Летняя Школа по Биоинформатике 2014

Seminar goals

- Learn how to perform RNA-seq data analysis to get answers to your biological questions
- Get familiar with bionformatics tools for RNA-seq
- The following topics with be covered:
  - Quality control
  - Mapping of reads
  - Differential gene expression
  - Transcript assembly (reference-guided)
  - Fusion genes detection (bonus)

What **is not** covered: RNA-seq for non-reference organisms

RNA-seq overview



RNA-seq experiments

Differential gene expression (DE)

Other types of analysis:

- Novel genes discovery
- Alternative splicing
- Transcriptome reconstruction
- Fusion gene detection
- Small RNA studies

Things to keep in mind

Transcriptome analysis is more complex:

- alternative splicing
- pseudogenes
- non-coding transcripts

- Expression levels vary singnificantly: biological replicates are required
-

Popular pipelines for DE: **Tuxedo**

Protocol: www.nature.com/nprot/journal/v7/n3/fig_tab/nprot.2012.016_F2.html

- Differential expression analysis
- Transcript reconstruction (reference guided)
- Alternative splicing
- Requirements:
  - Unix command line knowledge
  - GUI wrappers available: Galaxy, Unipro UGENE

Popular pipelines for DE: **DESeq/edgR**

Protocol: http://www.nature.com/nprot/journal/v8/n9/full/nprot.2013.099.html

- **Robust** differential expression analysis
- Multilevel fact design
- Alternative splicing (exon based, DEXSeq)
- Requirements:
  - Unix command line, R language

Common pipeline steps

- Quality control
  - FastQC
- Spliced aware mapping
  - Tophat, STAR
- More quality controls
  - Qualimap, RNA-seq QC

Tuxedo steps

- Run Cufflinks to reconstruct the transcripts (cuffmerge, cuffdiff)
- Post-analysis using R-cumme Bound and other tools

Deseq/edgeR steps

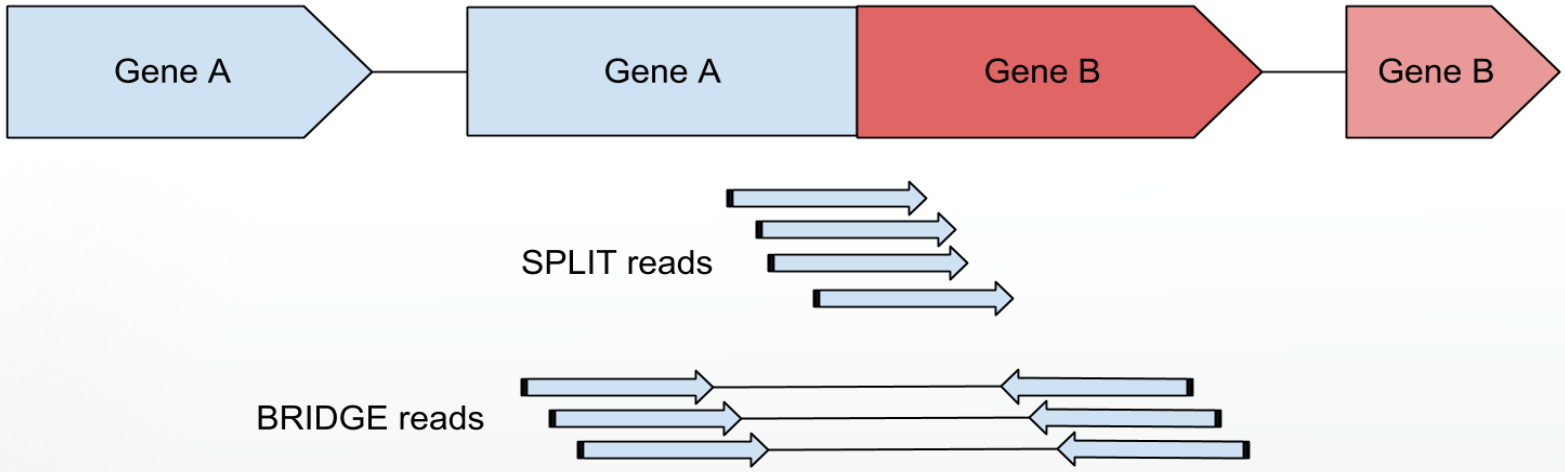
- Read counting
- EDA
- Gene expression
  - Normalization
  - Dispersion estimation
  - Statistical testing

Example case: long-living nematodes

- Nematodes fed with Glycin. It increases their long-livity. Which genes change their expression?
- Experiment design: 3 Glcn+ vs 3 Glcn- RNA-seq
- See additional files: **analysis.txt, diffExprAnalysis.R**

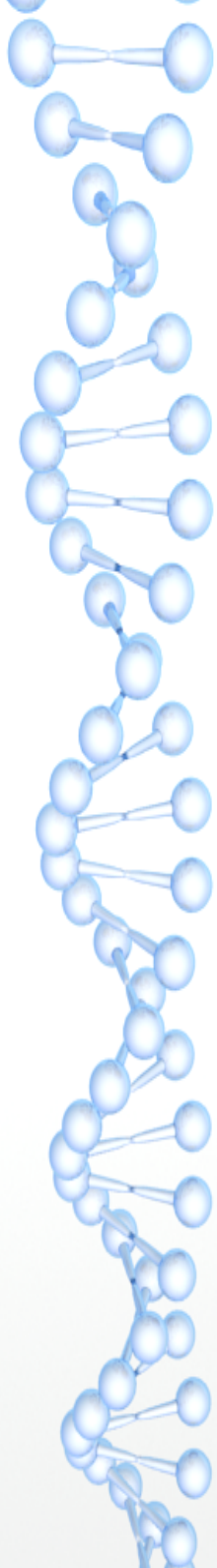
Fusion genes search

- Fusion genes consists of parts of different genes
- Can be detected from RNA-seq using mapping or assembly approach



Fusion genes search

- Existing tools based on mapping:
  - DeFuse
  - ChimeraScan
  - TophatFusion
  - InFusion



Thanks for the attention!