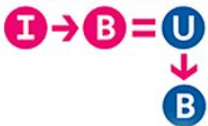


Studying the transcriptome using RNA-seq

Cecilia Coimbra Klein

IBUB
Institut de Biomedicina
de la Universitat de Barcelona



UNIVERSITAT DE
BARCELONA



UVIC UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA



Master in Omics
Data Analysis

Outline

Outline

1. Introduction
 - 1.1. Goal
 - 1.2. Content overview
 - 1.3. Hands-on data
2. Basic concepts
3. Short-read RNA-seq data processing
4. Gene level RNA-seq data analysis
5. Isoform level RNA-seq analyses
6. Regulation of gene expression

Goal

Goal

- Study the transcriptome using command line
 - Review concepts and analysis of RNA-seq
 - Basic linux commands, AWK programming, Git and Docker, R wrappers
 - Main focus on command line hands-on

Content overview

Content overview

1. Introduction
2. Basic concepts
3. Short-read RNA-seq data processing
4. Gene level RNA-seq data analysis
5. Isoform level RNA-seq analyses
6. Regulation of gene expression

Content overview

1. Introduction

- 1.1. Goal
- 1.2. Content overview
- 1.3. Hands-on data

2. Basic concepts

- 2.1. Hands-on:
 - 2.1.1. Basic Linux Commands
 - 2.1.2. Git
 - 2.1.3. Docker
- 2.2. RNA-seq:
 - 2.2.1. RNA biology
 - 2.2.2. NGS technologies
 - 2.2.3. RNA-seq experimental design
 - 2.2.4. Reference gene annotation
 - 2.2.5. Data formats

Content overview

3. Short-read RNA-seq data processing
 - 3.1. Quality control
 - 3.2. Read mapping
 - 3.3. Visualization of gene expression signal
 - 3.4. Gene expression quantification and normalization
4. Gene level RNA-seq data analysis
 - 4.1. Sample clustering based on gene expression
 - 4.2. Differential gene expression
 - 4.3. Gene ontology (GO) term enrichment
5. Isoform level RNA-seq analysis
 - 5.1. AS events from genomic annotation
 - 5.2. PSI values
 - 5.3. Differential splicing analysis
 - 5.4. Functional analysis

Content overview

- 6. Regulation of gene and isoform expression
 - 6.1. ChIP-seq data processing
 - 6.1.1. Mapping
 - 6.1.2. Peak calling
 - 6.1.3. Visualisation of signal
 - 6.2. ChIP-seq data analysis
 - 6.2.1. Genomic locations
 - 6.2.2. Differential peaks per tissue
 - 6.2.3. BED files in UCSC browser
 - 6.3. Multi-omics data analysis
 - 6.3.1. Promoter regions of differentially expressed genes
 - 6.3.2. ATAC-seq signal in the UCSC genome browser
 - 6.3.3. Promoter regions of differentially spliced genes
 - 6.3.4. Omics portals

Hands-on

Hands-on content overview

- **Basic concepts**
 - Basic Linux Commands / AWK programming / Git / Docker
 - Data exploration / Data formats / data subsetting
- **Short-read RNA-seq data processing**
 - Quality control - FastQC
 - Mapping - STAR / Quantification - RSEM
- **Gene level RNA-seq data analysis**
 - R wrappers
- **Isoform level RNA-seq data analysis**
 - Isoform usage - SUPPA
- **Regulation of gene and isoform expression**
 - ChIP-seq data processing
 - ChIP-seq data analysis
 - Multi-omics data analysis



Data Analysis Hands-on

Hands-on data

- Forebrain, heart and liver of 12.5 days mouse embryos
 - 2 bio replicates
 - RNA-seq, CHIP-seq and ATAC-seq
- References:
 - mouse genome – mm10 assembly
 - gene annotation – gencode vM4
- Processing:
 - References: a small sample of the genome and annotation (21 chromosomes, 1Mb long)
 - Data: one sample only (100,000 alignment-based pre-filtered reads)
- Analysis:
 - all samples

<https://public-docs.crg.es/rguigo/Data/cklein/courses/UVIC/handsOn/>

Hands-on

Introduction 1

https://public-docs.crg.es/rguigo/Data/cklein/courses/UVIC/handsOn/#_introduction