

## RNA-seq de novo ABiMS

### *Cleaning*

1. **import des données d'entrée depuis Data Libraries :**
  - Shared Data → Data Libraries → RNA-seq de-novo
  
2. **lancement des programmes de nettoyage pas à pas**
  - **FastQC:Read QC**
    - **Short read data from your current history**  
BlueLight.sample.read1.fastq
  
  - **Step1 : prinseq\_lite**
    - **reads fastq file** BlueLight.sample.read1.fastq
    - **phread64** No
    - **trim\_ns\_left** 1
    - **trim\_ns\_right** 1
    - **ns\_max\_n** 0
    - **trim\_qual\_right** 20
    - **min\_qual\_mean** 25
    - **min\_len** 50
    - **noniupac** True
  
  - **FastQC:Read QC**
    - **Short read data from your current history**  
BlueLight.sample.read1.fastq\_good.fastq
  
  - **Step2 : Cutadapt**
    - **Fastq file to trim**  
BlueLight.sample.read1.fastq\_good.fastq
    - **5' or 3' (Anywhere) Adapters → Source: Enter custom sequence**  
AGATCGGAAGAGCACACGTCTGAACTCCAG
    - **Output filtering options → set filters → Minimum length** 50

## Cleaning (suite)

- **FastQC:Read QC**
  - **Short read data from your current history**  
BlueLight.sample.read1.fastq\_good.fastq.cutadapt.fastq
- **Step3 : prinseq\_lite**
  - **reads fastq file** BlueLight.sample.read1.fastq\_good.fastq.cutadapt.fastq
  - **phread64** No
  - **min\_len** 50
  - **nonlupac** True (by default)
  - **trim\_tail\_left** 5
  - **trim\_tail\_right** 5
  - **lc\_method** entropy
  - **lc\_threshold** 70
- **FastQC:Read QC**
  - **Short read data from your current history**  
BlueLight.sample.read1.fastq\_good.fastq.cutadapt.fastq\_good.fastq
- **Step4 : riboPicker**
  - **from**  
BlueLight.sample.read1.fastq\_good.fastq.cutadapt.fastq\_good.fastq
  - **Reference Database**  
Non-redundant Ribosomal RNA Database (rrnadb)
- **FastQC:Read QC**
  - **Short read data from your current history**  
BlueLight.sample.read1. [...] cutadapt.fastq\_good.fastq.nonrrna.fastq

## 2. création du workflow via l'historique

-  → **Extract Workflow** : RNAseq – cleanning

## 3. visualisation du workflow

- Workflow → [RNAseq – cleanning] → Edit

## *Cleaning (suite 2)*

### 4. lancement du workflow sur les 3 jeux restants

- Workflow → [RNAseq – cleanning] → Run
- paramétrages
  - **Step 1: Input dataset**
    - **Condition A - Read 1** BlueLight.sample.read1.fastq
    - or BlueLight.sample.read2.fastq
    - or Dark.sample.read1.fastq
    - or Dark.sample.read2.fastq
  - **Step3: prinseq\_lite**
  - **Step5: Cutadapt**
  - **Step7: prinseq\_lite**
  - **Step9: riboPicker**

# Assemblage

## Lancement des outils pas à pas

- **Step 1a: Get pairs**
  - **left reads fastq file** BlueLight.sample.read1. [...] .nonrrna.fastq
  - **right reads fastq file** BlueLight.sample.read2. [...] .nonrrna.fastq
- **Step 1b: Get pairs**
  - **left reads fastq file** Dark.sample.read1. [...] .nonrrna.fastq
  - **right reads fastq file** Dark.sample.read2. [...] .nonrrna.fastq
- **Step 2a: Concatenate datasets**
  - **Concatenate Dataset** BlueLight.sample.read1. [...] .paired.fastq
  - **Insert Dataset** Dark.sample.read1. [...] .paired.fastq
- **Step 2b: Concatenate datasets**
  - **Concatenate Dataset** BlueLight.sample.read2. [...] .paired.fastq
  - **Insert Dataset** Dark.sample.read2. [...] .paired.fastq
- **Step 3:**
  - Rename your datasets. Ex:  
all.read1.cleaned.paired.fastq  
all.read2.cleaned.paired.fastq
- **Step 4: Insilico Read Normalization**
  - **Maximum coverage** 30
  - **Paired or single-end data** paired
  - **Left/Forwards strand reads** all.read1.cleaned.paired.fastq
  - **Right/Reverse strand reads** all.read2.cleaned.paired.fastq
- **Step 6: Trinity**
  - **Paired or Single-end data** Paired
  - **Left/Forward strand reads**  
Insilico read normalization on data ... and data ...: Left reads
  - **Right/Reverse strand reads**  
Insilico read normalization on data ... and data ...: Right reads
  - **Strand-specific Library type** None
  - **Group pairs distance** 500
  - **Path reinforcement distance** 75
- **Step 7**
  - Rename your assembly file. Ex:  
Trinity\_assembly.fasta

- **Step 8: basic stats of the assembly**
  - **Trinity assembly** Trinity\_assembly.afsta
- **Step 9: RSEM Align and Estimate**
  - **Transcripts** Trinity\_assembly.fasta
  - **Abundance estimation method** RSEM
  - **Paired or single-end data** paired
  - **Left reads** all.read1.cleaned.paired.fastq
  - **Right reads** all.read2.cleaned.paired.fastq
  - **seqType** fastq
- **Step 10:**
  - Rename your assembly file. Ex:  
RSEM.isoforms.results
- **Step 11: Filter fasta by rsem values**
  - **Trinity Fasta File** Trinity\_assembly.fasta
  - **RSEM output** RSEM.isoforms.results
  - **FPKM cutoff** 1
  - **Isopct cutoff** 1
- **Step 12 :**
  - Rename your filtered assembly file. Ex:  
Trinity\_assembly.filtered.fasta

## Analyse différentielle

### Lancement des outils pas à pas

- **Step 1a: RSEM Align and Estimate**
  - **Transcripts** Trinity\_assembly.filtered.fasta
  - **Abundance estimation method** RSEM
  - **Paired or single-end data** paired
  - **Left reads** BlueLight.sample.read1.[...] .paired.fastq
  - **Right reads** BlueLight.sample.read2.[...] .paired.fastq
  - **seqType** fastq
- **Step 1b: RSEM Align and Estimate**
  - **Transcripts** Trinity\_assembly.filtered.fasta
  - **Abundance estimation method** RSEM
  - **Paired or single-end data** paired
  - **Left reads** Dark.sample.read1.[...] .paired.fastq
  - **Right reads** Dark.sample.read2.[...] .paired.fastq
  - **seqType** fastq
- **Step 2:**
  - Rename your datasets. Ex:
    - RSEM.BlueLight.isoforms.results
    - RSEM.BlueLight.genes.results
    - RSEM.Dark.isoforms.results
    - RSEM.Dark.genes.results
- **Step 3: Merging Tabular**
  - **With header** True
  - **Data column number** 5
  - **Tabular file** RSEM.BlueLight.isoforms.results
  - **Sample name** BlueLight
  - **Tabular file** RSEM.Dark.isoforms.results
  - **Sample name** Dark
- **Step 4: Trinity run DE analysis**
  - **Merge output file** Tabular merge
  - **Method** edgeR
  - **Replicate** No
  - **Min row sum counts** 10

## ***Annotation***

- **Step 1b: Extracting likely coding regions from Trinity transcripts (Transdecoder)**
  - **Trinity transcript** Trinity\_assembly.fasta
  - **Minimum Protein Length** 50
  - **Genetic code** universal
  - **Search Pfam** Yes
  - **Verbose** Yes
  - **Strand specific** No
  - **Top longest ORFs to train Markov Model** 500
  - **Retain long ORFs** 900

•