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** False (by default)
     - **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
     - **3' Adapters → Enter custom 3' adapter 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** False (by default)
  - **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
  - **Add new Dataset → Dataset 1** Dark.sample. read1. [...] .paired.fastq

- Step 2b: Concatenate datasets
  - **Concatenate Dataset** BlueLight.sample. read2. [...] .paired.fastq
  - **Add new Dataset → Dataset 1** Dark.sample. read2. [...] .paired.fastq

- Step 3:
  - Rename your datasets. Ex:
    all.read1.cleaned.paired.fastq
    all.read2.cleaned.paired.fastq

- Step 4: normalize_by_kmer_coverage
  - **JM: 1G**
  - **max_cov** 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
  - **Process paired reads by averaging stats between pairs and retaining linking info** Yes

- Step 5:
  - Rename your datasets. Ex:
    all.read1.cleaned.paired.fastq.normalized_left_reads.fastq
    all.read2.cleaned.paired.fastq.normalized_right_reads.fastq

- Step 6: Trinity
  - **JM** 1G (by default)
  - **CPU** 2 (by default)
  - **Paired or Single-end data** Paired (by default)
  - **Left/Forward strand reads**
    all.read1.cleaned.paired.fastq.normalized_left_reads.fastq
  - **Right/Reverse strand reads**

all.read2.cleaned.paired.fastq.normalized_right_reads.fastq

- **Strand-specific Library type** None
- **Group pairs distance** 500
- **Path reinforcement distance** 75 (by default)

- 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
  - **Trinity assembly** Trinity_assembly.fasta
  - **Paired or single-end data** paired
  - **Left/Forward strand reads** all.read1.cleaned.paired.fastq
  - **Right/Reverse strand reads** all.read2.cleaned.paired.fastq
  - **Transcripts source** Trinity

- Step 10:
  - Rename your assembly file. Ex:
    RSEM.isoforms.results
  - Datatype: Roadmaps → tabular

- 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 differentielle

Lancement des outils pas à pas

- Step 1a: RSEM Align and Estimate
  - **Trinity assembly** Trinity_assembly.filtered.fasta
  - **Paired or single-end data** paired
  - **Left/Forward strand reads** BlueLight.sample. read1. [...] .paired.fastq
  - **Right/Reverse strand reads** BlueLight.sample. read2. [...] .paired.fastq
  - **Transcripts source** Trinity

- Step 1b: RSEM Align and Estimate
  - **Trinity assembly** Trinity_assembly.filtered.fasta
  - **Paired or single-end data** paired
  - **Left/Forward strand reads** Dark.sample. read1. [...] .paired.fastq
  - **Right/Reverse strand reads** Dark.sample. read2. [...] .paired.fastq
  - **Transcripts source** Trinity

- 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 1a: Trinity transcripts to candidate peptides (transcriptsToOrfs)
  - **Trinity assembly** Trinity_assembly.fasta
  - **Minimum peptide length** 50 (100 by default)
  - **Retain all ORFs found that are of minimum length in nucleotdes** 200 (900 by default)
  - **Number of top longest ORFs to train Markov Model** 500(by default)
  - **Strand specific type** NOT strand specific, examine both strands (by default)
  - **Genetic Code** use default (universal) (by default)
  - **Search PFAM database** NO (by default)

- Step 1b: Extracting likely coding regions from Trinity transcripts (Transdecoder)
  - **Trinity transcript** Trinity_assembly.fasta
  - **Minimum Protein Length** 50 (100 by default)
  - **Genetic code** universal (by default)
  - **Search Pfam** Yes (by default)
  - **Verbose** Yes (by default)
  - **Strand specific** No (yes by default)
  - **Top longest ORFs to train Markov Model** 500 (by default)
  - **Retain long ORFs** 200 (900 by default)
  -

•