

**RNA-Seq based analysis of differential gene expression
Using RobiNA – a quick guide**

Step 1: Data import

RobiNA - The transcriptomics data preprocessor. Version 1.2.1

Welcome to RobiNA!


The first step of analysing RNA-Seq data in Robin is to import the raw sequence data (in FastQ format). Based on the input data, RobiNA will try to guess the Illumina pipeline version automatically. Setting the pipeline version manually is possible but not recommended unless you know exactly what you're doing.

Illumina pipeline override manually

Alternative data input

Importing SAM/BAM format alignments will skip the quality checking, read trimming and reference mapping steps and skip forward to library configuration.

In case you already have a counts table in a simple tab-separated text file, you can import it into Robin and skip forward to the statistics step directly



Raw read data files

- /Users/marc/Desktop/RobiNA_testset/SRR392118.fastq.sample.fastq
- /Users/marc/Desktop/RobiNA_testset/SRR392119.fastq.sample.fastq
- /Users/marc/Desktop/RobiNA_testset/SRR392120.fastq.sample.fastq
- /Users/marc/Desktop/RobiNA_testset/SRR392121.fastq.sample.fastq
- /Users/marc/Desktop/RobiNA_testset/SRR392122.fastq.sample.fastq
- /Users/marc/Desktop/RobiNA_testset/SRR392123.fastq.sample.fastq**

Welcome to RobiNA ○ idle

After selecting „alternative input“ you can load SAM/BAM alignments or precomputed counts tables

click the question mark for details on the required counts table format

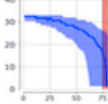
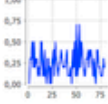
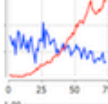
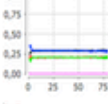

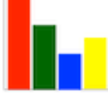
1 Click "Add" to import raw FASTQ files....

Step 2: Quality checking

RobiNA – The transcriptomics data preprocessor. Version 1.2.1

Input quality assessment

RobiNA offers a choice of analyses that give an overview of the quality of the input RNA-Seq reads. You can freely choose and combine the listed methods. Depending on the amount of input reads and number of methods chosen the runtime may vary considerably.

-  **Base call quality** –Shows summary plots of the base call qualities for each input file Include? [?](#) more
-  **Consecutive homopolymers** –Identifies read positions that contain homopolymers across reads Include? [?](#) more
-  **Kmer frequency** –Identifies the kmer frequencies across reads Include? [?](#) more
-  **Base call frequencies** –Plots the nucleotide base call across all reads Include? [?](#) more
-  **Overrepresented sequences** –Extracts sequences that occur more often than expected. Include? [?](#) more
-  **Basic statistics** –Computes some basic overview figures such as nucleotide composition, number of reads, lengths etc. Include? [?](#) more

check all

1
Freely combine quality check modules by selecting „include“ ...

2 ...and optimize the settings according to your needs. Starting more parallel processes will require a multi CPU machine with sufficient memory

General File settings Kmer

free/total RAM: 171.1/4,096MB (4.2%)
free/total Java heap: 232.1/2,043.9MB (11.4%)
No. CPUs: 2

Parallel processes: 2
Illumina pipeline: 1.8+
 override manually

idle

Check the memory information to make sure there's enough free resources. Sticking to the defaults might result in longer runtime but is also more stable

Step 3: Raw read trimming to remove low quality data

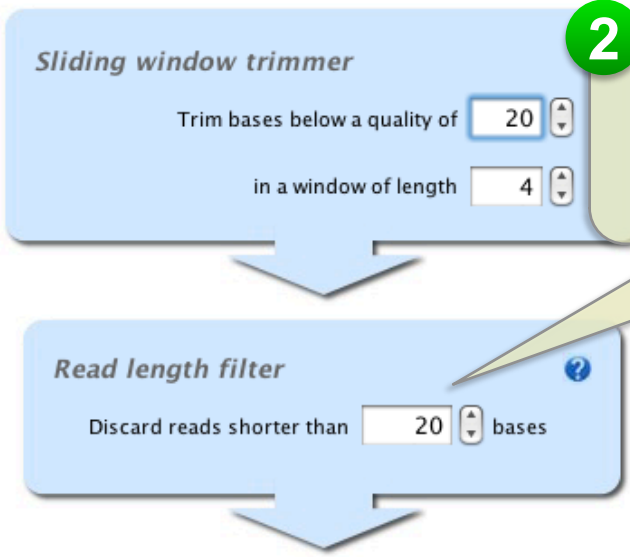
Raw read preprocessing

Raw sequencing reads may contain unwanted data to different extent and should be pre-processed before commencing the main analysis. Possible issues are low quality reads, contamination by adapter sequences etc. The Trimmomatic tool offers various processing modules that can be combined into a

preprocessin pipeline that cleans up the data. In Addition, data originating from a multi-plex sequencing run needs to be pre-processed. For the barcode sequences you can use Trimmomatic modules from the list below simply by drag & drop. To change the order of the modules please click "clear".

1 Drag & drop trimmer modules from the list to the open space on the right. Trimmers will be executed in the order you choose

- Adapter clipper
- Leading quality trimmer
- Read length cropper
- Trailing quality trimmer



2 Settings for each individual trimmer module can be adapted to your needs

Click „Clear“ to start again when not satisfied with the or order of modules

Step 4: Experiment layout

RobiNA - The transcriptomics data preprocessor. Version 1.2.1

Experiment layout

The first step to set up your experiment will be to define the different **conditions** that were applied (e.g. "wildtype" vs "mutant" or "stressed" vs "control" etc.). Please define all conditions that were used. In the second step you have to define **samples** (at least one) for each condition. Each sample is going to be treated as a **true biological replicate**. Please arrange your samples they were generated from. (Samples can contain more than one file)

1 Enter conditions

control
stress
mutant

Add Remove

2 Add samples

Trimmed input files

- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392119.fastq.sample.fastq
- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392120.fastq.sample.fastq
- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392121.fastq.sample.fastq
- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392122.fastq.sample.fastq
- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392123.fastq.sample.fastq

Condition: control Add Remove

Samples

control_1 198281 reads

Files in sample

- /Users/marc/Desktop/TESTESTES/input/TRIMMED_SRR392118.fastq.sample.fastq

1 First enter each experimental condition used in your experiment – e.g. „control“, stress“ etc.

2 Then define the samples: E.g. file SRR392118.fastq contains the reads of the first replicate of the control sample

Previous Next

Welcome to RobiNA idle

? Select the sample and click „Remove“ if you made a mistake and want to add the file to another sample

2 Then define the samples: E.g. file SRR392118.fastq contains the reads of the first replicate of the control sample

1 First enter each experimental condition used in your experiment – e.g. „control“, stress“ etc.

Step 5: Mapping of the reads to a reference -> Generate counts table

RobiNA - The transcriptomics data preprocessor. Version 1.2.1

Mapping of reads

To extract the transcription profile, the filtered reads need to be mapped to a reference transcriptome or genome. The Robin package bundles a choice of tools that map short reads to a reference. Depending on which tool you choose, the reference sequence(s) can be supplied in different formats.

1 Choose reference type

- Transcriptome**
Choose this option if you want to supply a transcriptome in the form of a multiple FASTA file with entries for each transcript. Reads that are not mapping to any transcript in the reference will not be taken into the analysis.
- Genome**
Choosing this option requires both a genome sequence file (in FASTA format) plus an annotation file (GFF3 format).

2 Choose mapping tool

Select **Bowtie**

Bowtie settings

Presets: no mismatches, unique alignments only
 up to 2 mismatches, unique alignments only
 custom

seed mismatch: []
seed length: [28]
mismatch quality: [70]
more args: []

3 Specify reference data

Reference transcriptome

TAIR10_cds_201...

Sequences:
Avg. length:
N50:
N content:
Ambiguous Seqs.:

4 Start mapping!

Time elapsed: [] Stop mapping

Mapping progress log

Mapping result overview

Reset Previous Next

Welcome to RobiNA

idle

1 Select the reference type. Both transcriptomes and genomes are possible

2 Choose the settings for bowtie and conform your choice by clicking "Use these settings"

Bowtie indices only have to be built once. The drop box shows a list of already available indices

3 Start the mapping process

Step 5: Mapping of the reads to a reference -> Generate counts table

RobiNA - The transcriptomics data preprocessor. Version 1.2.1

Mapping of reads

To extract the transcription profile, the filtered reads need to be mapped to a reference transcriptome or genome. The Robin package bundles a choice of tools that map short reads to a reference. Depending on which tool you choose, the reference sequence(s) can be supplied in different formats.

1 Choose reference type

- Transcriptome
- Genome

2 Choose mapping tool

Select **Bowtie**

Bowtie settings

Presets: no mismat...

seed mismatch: 0

seed length: 28

mismatch quality: 70

more args:

Use these settings

4 Start mapping!

Time elapsed: 00:01:45 Stop mapping

Mapping progress log

```
##### Mapping reads of sample control_2
# reads processed: 98281
# reads with at least one reported alignment: 40445 (41.15%)
# reads that failed to align: 50841 (51.73%)
# reads with alignments suppressed due to -m: 6995 (7.12%)
Reported 40445 alignments to 1 output stream(s)
```

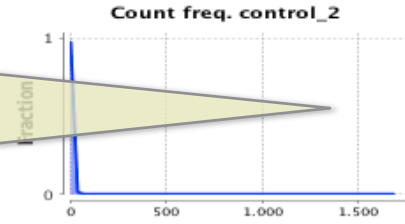
All mapping tasks finished.

The reads could be mapped to a total of 14488 different transcripts.

Mapping result overview

Sample: control_2

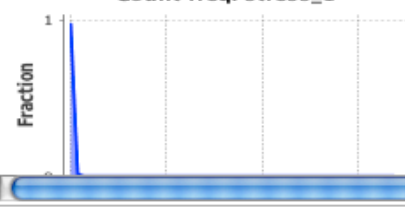
Count freq. control_2



reads processed: 97856
reads with at least one reported alignment: 41945 (42.86%)
reads that failed to align: 50841 (51.76%)
reads with alignments suppressed due to -m: 5258 (5.37%)
Reported 41945 alignments to 1 output stream(s)

Sample: stress_1

Count freq. stress_1



reads processed: 99394
reads with at least one reported alignment: 41702 (41.96%)
reads that failed to align: 52841 (52.89%)
reads with alignments suppressed due to -m: 5125 (5.16%)
Reported 41702 alignments to 1 output stream(s)

Reset Previous Next

Welcome to RobiNA idle

1 After the mapping process is finished, a summary graph of the gene count distribution and mapping details are shown for each sample that was defined in the experiment layout step (4)

Step 6: Analysis setup

RobiNA - The transcriptomics data preprocessor. Version 1.2.1

Design your experiment

1 To define a comparison of interest, simply hold down the CTRL key and click-drag to draw an arrow between two boxes. The arrow shown here defines the comparison: „control“ minus „stress“

Each box represents all replicate samples taken for the respective condition

Right-click and choose „delete“ to remove accidentally drawn arrows

2 Modify the statistics method and settings according to your need or use the robust defaults

Analysis Method

- edgeR
- DESeq

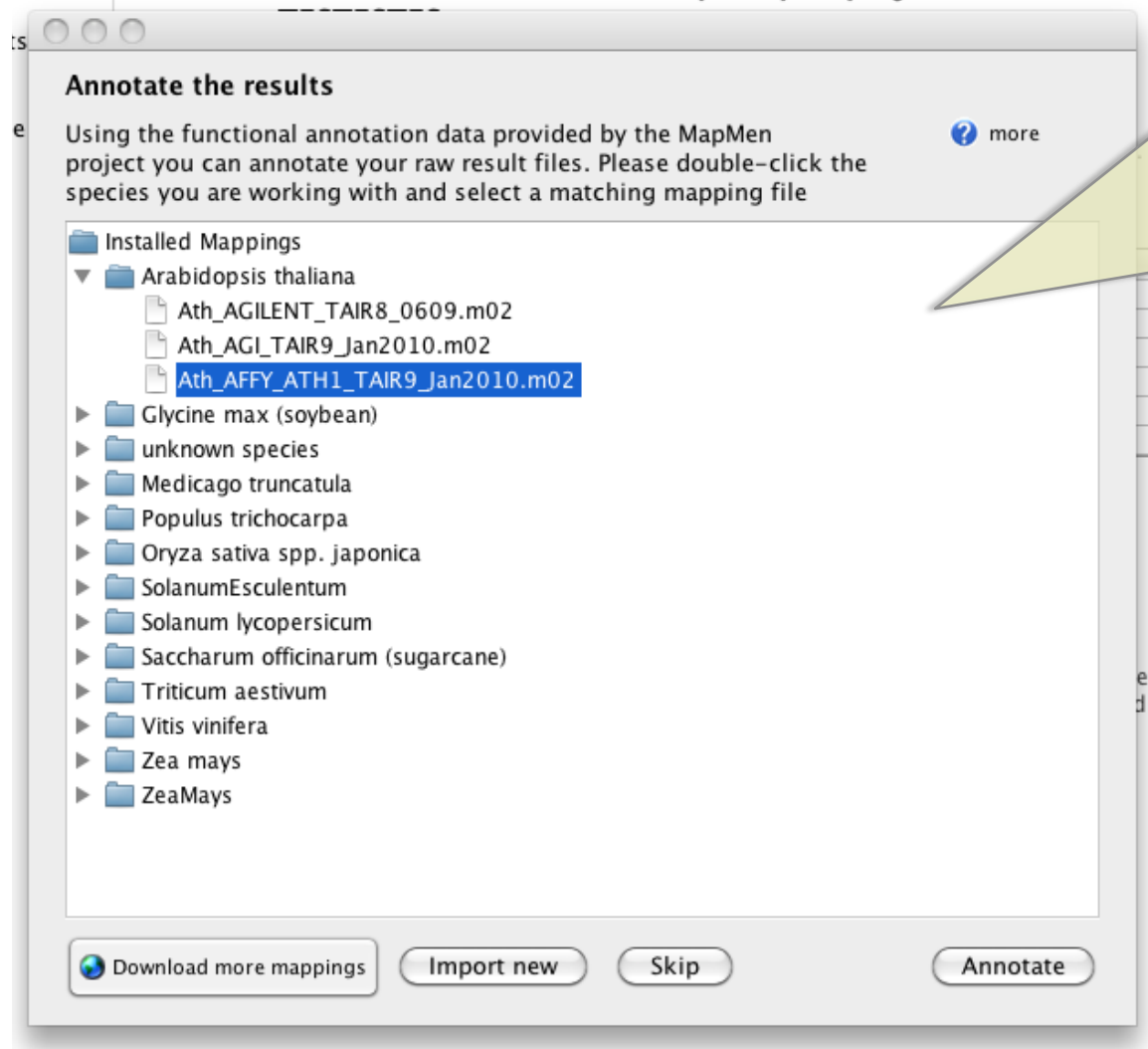
Settings

- p-value cutoff: 0.05
- P value correction: BH
- Correct GC bias within lanes: full
- Between lanes: full
- Log-fold change min=1
- estimate RPKM expression values
- Dispersion: auto

← Previous Next →

Welcome to RobiNA idle

Optional step 6.2: Annotate the results with functional MapMan Bins



In case you are working with an organism for which a MapMan mapping is available you can choose to create an additional result table in which all genes are annotated with functional terms (named „Bins“).

Step 7: Analysis finished - browse results

Browse your results

You can now browse the results of the analysis.

All results will be written to the project folder when you close Robin.

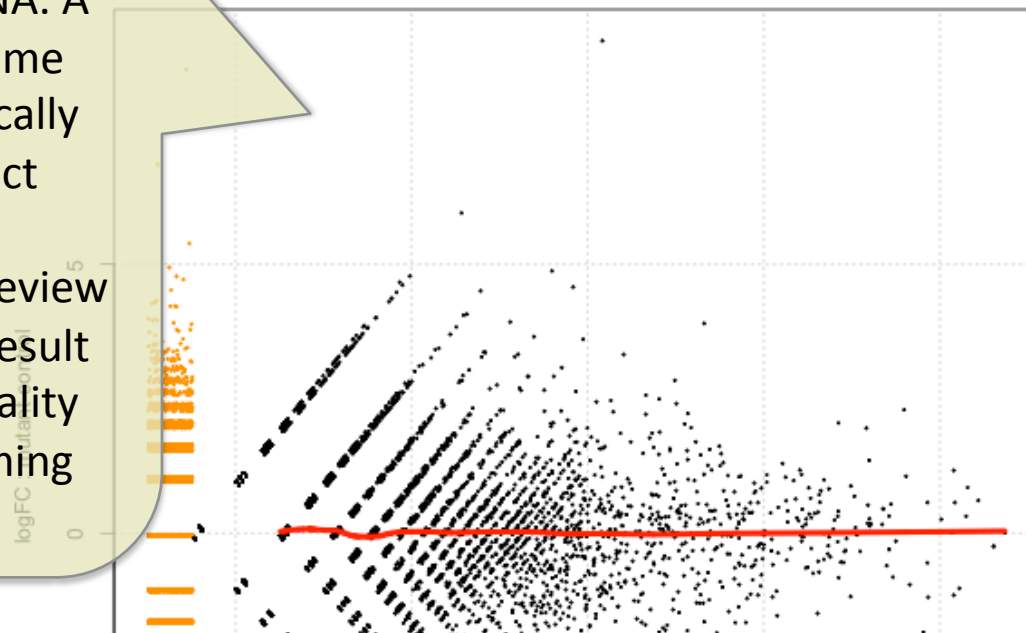
The analysis is done – You can browse the results directly within RobiNA. A PDF file with the same content is automatically saved in the project directory for documentation and review (plus, of course, all result tables and plots, quality check results, trimming details etc.)

MA plots of each comparison

The MA plots show the log₂ fold change (M; logFC) plotted versus the average expression strength (A; LogConc) for each of the comparisons that was computed. Usually, these scatter plots show a trumpet-like shape which is attributed to the fact that genes with a lower expression signal strength are more strongly affected by noise than strongly expressed genes.

According to the assumption that under most experimental conditions the bulk of the genes of an organism are not responding differentially, the cloud of points should be centered around a log fold change of 0. Genes that were called significantly differentially expressed are shown in red.

MA plot of contrast control-stress



← Previous × Close