



Analyze a count matrix using DIANE



Partner institutions



Non-exhaustive ecosystem of featured packages

Objectives

Know how you can import your data in DIANE

Perform an exploratory analysis of a count matrix.

Expected output

- Differentially expressed gene list
- GO enrichment of Differentially expressed genes list
- Diagnostic plots

Required files overview

Organism included in DIANE

Mandatory :

- Raw count matrix

Facultative :

- Experimental design file

Organism **not** included in DIANE

Mandatory :

- raw Count matrix

Facultative :

- Experimental design file
- Annotation file
- Gene ontology file

Gene	condition.one_1	condition.one_2	condition.two_1	condition.two_2
Gene.1	2	3	45	52
Gene.2	10	12	0	3
Gene.3	72	92	1020	1241
Gene.4	0	0	0	1

Format

- **Row** : Gene_ID
- **Values** : read count
- **Column** : Sample_name
 - Sample name (Without any “_” !!!)
 - —
 - Replicate

Gene	condition.one_1	condition.one_2	condition.two_1	condition.two_2
Gene.1	2	3	45	52
Gene.2	10	12	0	3
Gene.3	72	92	1020	1241
Gene.4	0	0	0	1

Note : the first column MUST be named “Gene”

Can be used to add more informations in some part of the analysis.

Example of experimental design file

Condition	Heat	Salt
Control	0	0
Heat	1	0
Salt	0	1
Heat.and.Salt	1	1

Can be used to work with custom organisms.

Example of annotation file

Gene	Label	Annotation
AT5G12110.1	AT5G12110	elongation factor 1-beta 1
AT5G35670.1	IQD33	Member of IQ67 (CaM binding) domain containing family.
AT4G02130.1	GATL6	Encodes a protein with putative galacturonosyltransferase activity.
AT3G54830.1	AT3G54830	Transmembrane amino acid transporter family protein

Can be used to perform GO enrichment analysis with custom organisms

Example of Gene ontology file

GeneID	GoID
AT1G01010	GO:0098542
AT1G01010	GO:0003700
AT1G01010	GO:0009414
AT1G01020	GO:0016125
AT1G01040	GO:0010267

Using DIANE

1. Import your data
2. Normalisation
3. Count matrix exploration
4. Differential expression analysis
5. Differentially expressed genes exploration

Download the guide / tutoriel / TP

hpc.cirad.fr/rnaseq_southgreen_2023/DIANE_rnaseq_TP.pdf

(Just type this in your web browser)

Used dataset

Source : Sewelam N, Brilhaus D, Bräutigam A, Alseekh S, Fernie AR, Maurino VG. **Molecular plant responses to combined abiotic stresses put a spotlight on unknown and abundant genes.** J Exp Bot. 2020 Aug 6;71(16):5098-5112. doi: 10.1093/jxb/eraa250. PMID: 32442250.

Dataset description

- Combination of 3 stresses : Heat, Manitol, Salt
- 8 conditions : C ; H ; M ; S ; HM ; HS ; MS ; HMS
- 3 replicates
- 24 samples
- Sequenced using Illumina Hiseq2500 with 100pb paired-end reads.

Download count matrix and the design

```
wget --no-check-certificate  
hpc.cirad.fr/rnaseq_southgreen_2023/data_DIANE_southgreen2023.zip
```

```
unzip data_DIANE_southgreen2023.zip
```

(Or you can just type the url inside of your web browser)

Launch DIANE

<http://shinyapps.southgreen.fr/app/dianeбетapca>

This is a beta version with some improvement. The official version is here :

diane.ipsim.inrae.fr/

all the different available analysis

The main window

DIANE

Context

Data Import

Normalisation

Exploratory analysis

Differential Expression

Expression based clustering

Gene Regulatory Network

Ready to upload datasets

Legal mentions

Software versions

Upload expression data and experimental design

Demo Arabidopsis data

Toggle to import your data

Expected gene IDs are in the form AT4G31550 FOR ARABIDOPSIS THALIANA

Separator :

Comma Semicolon Tab

Choose CSV/TXT expression file ?

Browse... No file selected

Separator :

Tab

Choose CSV/TXT gene information file (optional) ?

Browse... No file selected

27655 genes

8 conditions

24 samples

Alternative splicing aware
GENE IDENTIFIERS

Arabidopsis thaliana
ORGANISM DATABASE

Additional gene data available ✓
LABEL, DESCRIPTION

Seed ensuring reproducibility (optional, can be left as default value) :

101

CHANGE SEED

SET SEED

Preview of the expression matrix

Expression data preview

This might help you visualize the general aspect of the data and different sequencing depths of your conditions.

Design and gene information files

Separator :

Comma Semicolon Tab

Choose CSV/TXT design file (optional)

Browse... No file selected

Show 10 entries

Search:

	Salt	Mannitol	Heat
C	0	0	0
H	0	0	1
S	1	0	0
M	0	1	0
SM	1	1	0
SH	1	0	1
MH	0	1	1
SMH	1	1	1

Showing 1 to 8 of 8 entries

Previous 1 Next

Describe the levels of each factors for your conditions

Show 10 entries

Search:

	C_1	C_2	C_3	S_1	S_2	S_3	M_1	M_2	M_3	H_1	H_2	H_3	SM_1	SM_2	SM_3	SH_1	SH_2	SH_3	MH_1	MH_2	M
AT1G01010.1	127	67.9	65.5	94	88.1	95.9	65.1	100.3	126.8	95.4	135	117.2	96.7	104.4	98.1	94.7	96.1	101.3	82.8	107.4	

Upload expression data and experimental design

Expression file upload

Demo Arabidopsis data

Toggle to import your data

Expected gene IDs are in the form **AT5G10510** FOR ARABIDOPSIS THALIANA

Separator : ☒ Comma ☐ Semicolon ☐ Tab

Choose CSV/TXT expression file ?

Browse... No file selected

Separator : ☒ Tab

Choose CSV/TXT gene information file (optional) ?

Browse... No file selected

27655 genes

8 conditions

24 samples

Alternative splicing aware

Arabidopsis thaliana

Additional gene data available ✓

GENE IDENTIFIERS

ORGANISM DATABASE

LABEL, DESCRIPTION

Seed ensuring reproducibility (optional, can be left as default value) :

85

?

CHANGE SEED

SET SEED

Preview of the expression matrix

Expression data preview

This might help you visualize the general aspect of the data and different sequencing depths of your conditions.

Design and gene information files

Separator : ☒ Comma ☐ Semicolon ☐ Tab

Choose CSV/TXT design file (optional)

Browse... No file selected

Show entries

Search:

	Salt	Mannitol	Heat
C	0	0	0
H	0	0	1
S	1	0	0
M	0	1	0
SM	1	1	0
SH	1	0	1
MH	0	1	1
SMH	1	1	1

Showing 1 to 8 of 8 entries

Previous Next

Describe the levels of each factors for your conditions

1. Data import

Step 1* : disable demo data and choose an organism

Upload expression data and experimental design

Step 4 : Upload your experimental design

Expression file upload

Demo Arabidopsis data ☒ **Toggle to import your data** ☐

Expected gene IDs are in the form **AT5G10510** FOR ARABIDOPSIS THALIANA

Separator : ☒ Comma ☐ Semicolon ☐ Tab

Choose CSV/TXT expression file ?

Browse... No file selected

Separator : ☒ Tab

Choose CSV/TXT gene information file (optional) ?

Browse... No file selected

27655 genes **8** conditions **24** samples

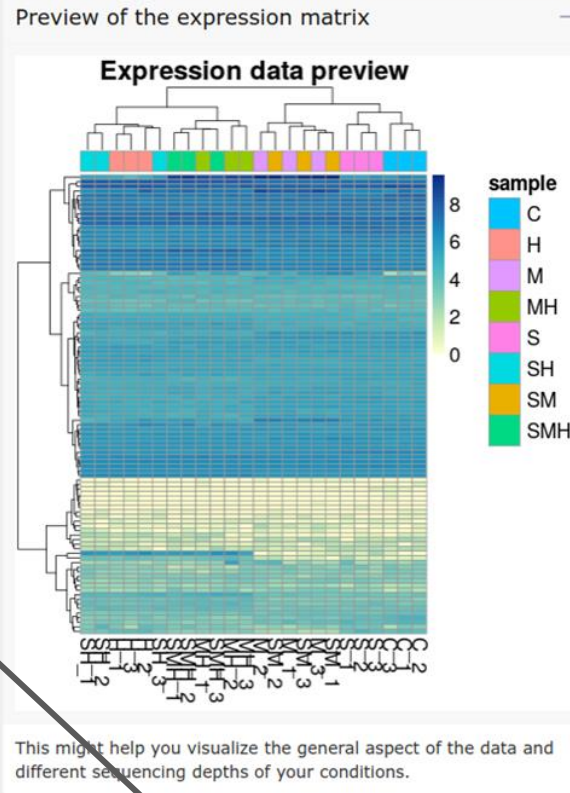
Alternative splicing aware
GENE IDENTIFIERS

Arabidopsis thaliana
ORGANISM DATABASE

Additional gene data available ✓
LABEL, DESCRIPTION

Seed ensuring reproducibility (optional, can be left as default value) :

85



Design and gene information files

Separator : ☒ Comma ☐ Semicolon ☐ Tab

Choose CSV/TXT design file (optional)

Browse... No file selected

Show **10** entries Search:

	Salt	Mannitol	Heat
C	0	0	0
H	0	0	1
S	1	0	0
M	0	1	0
SM	1	1	0
SH	1	0	1
MH	0	1	1
SMH	1	1	1

Showing 1 to 8 of 8 entries Previous **1** Next

Describe the levels of each factors for your conditions

Step 2* : Upload your count matrix here

Step 3 (if organism not included) : Upload your gene annotation here

Data filtering and normalisation

Settings

Normalization

Prior removal of differentially expressed genes: ☐ OFF

Normalisation method:

☒ tmm ☐ deseq2 ☐ none

NORMALIZE

Normalisation needed X

Low counts filtering

?

Minimal gene count sum across conditions :

240

FILTER

Normalisation needed X

Samples distributions

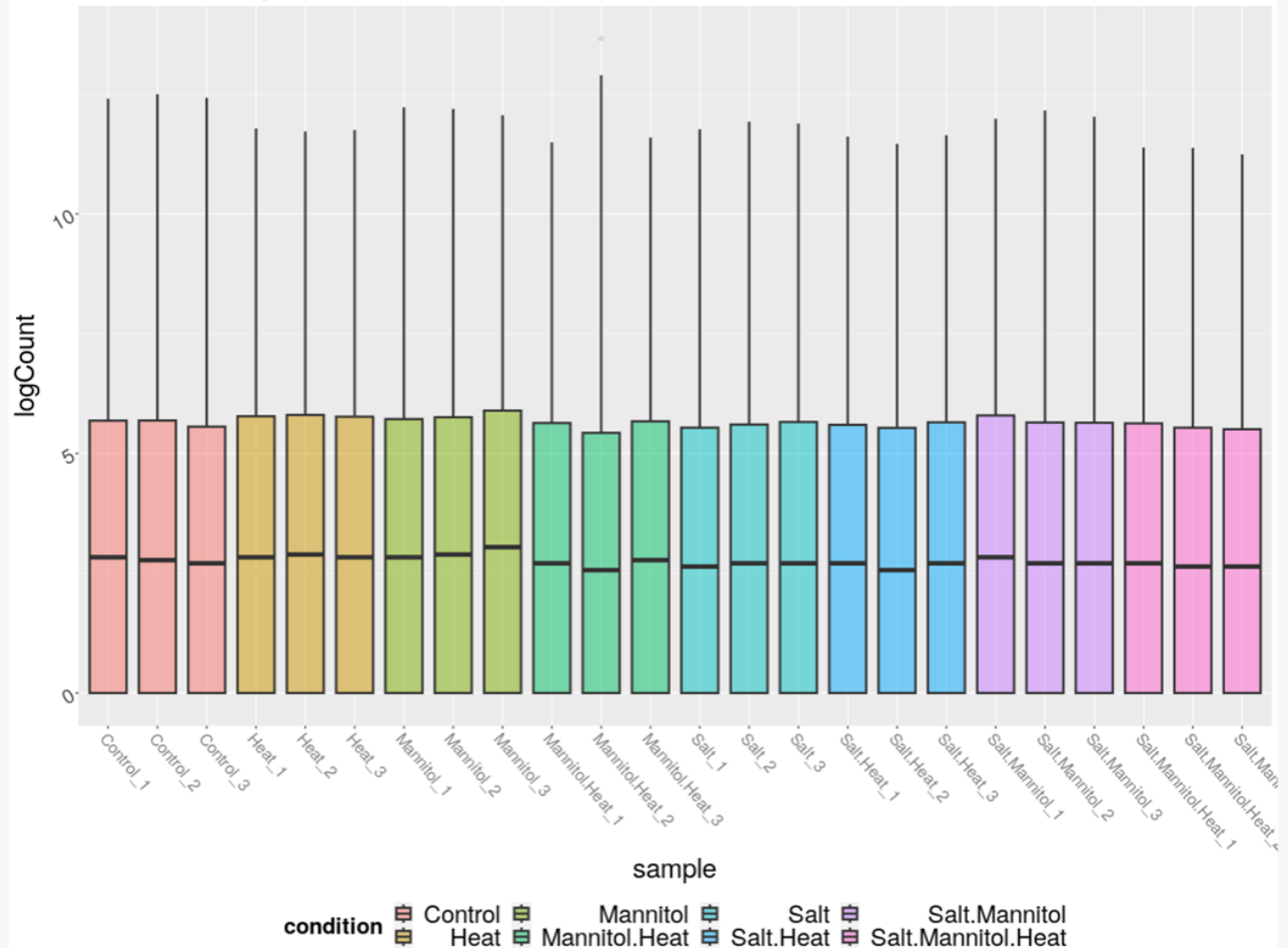
Summary

Data exploration

normalized

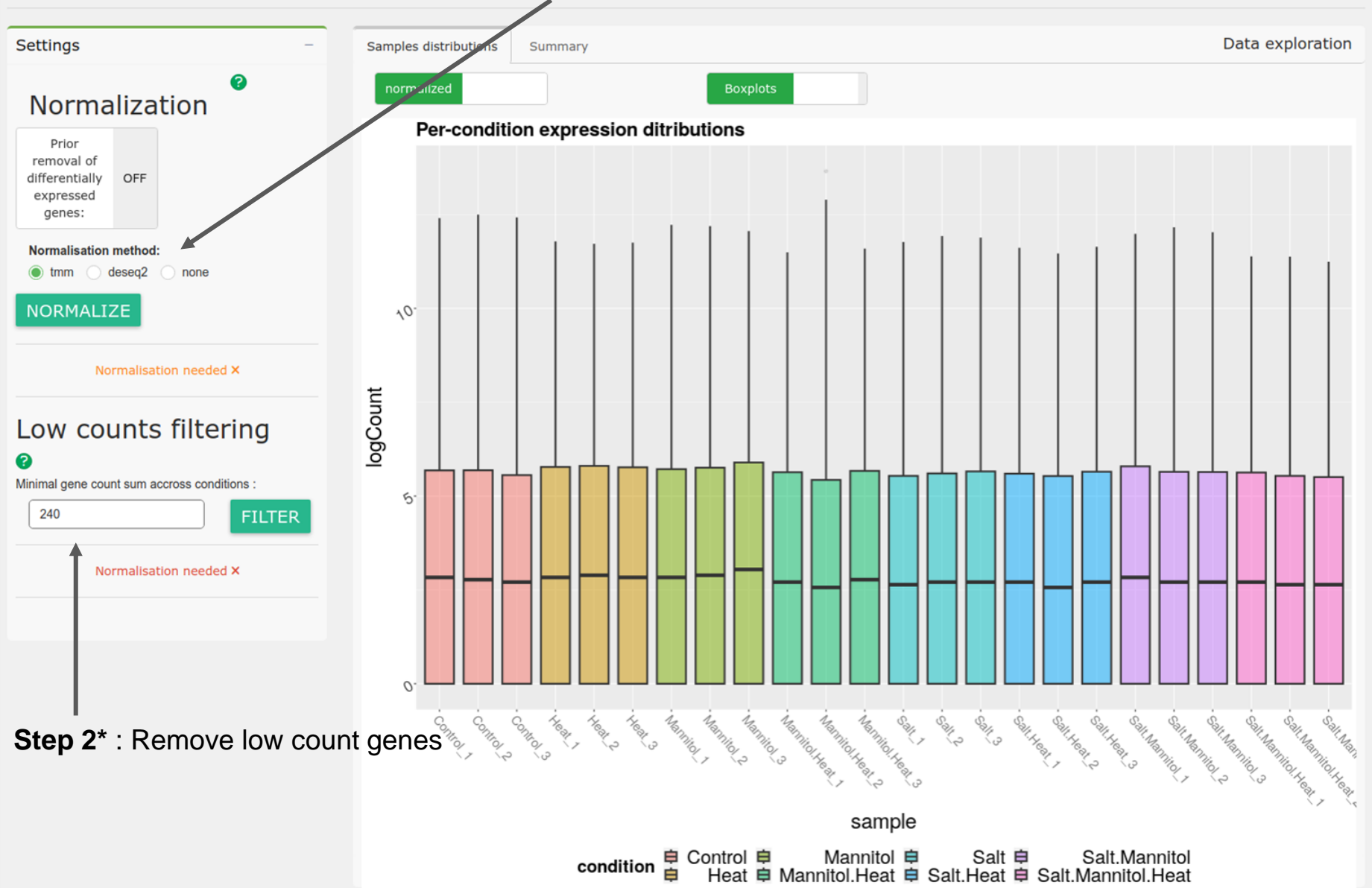
Boxplots

Per-condition expression distributions

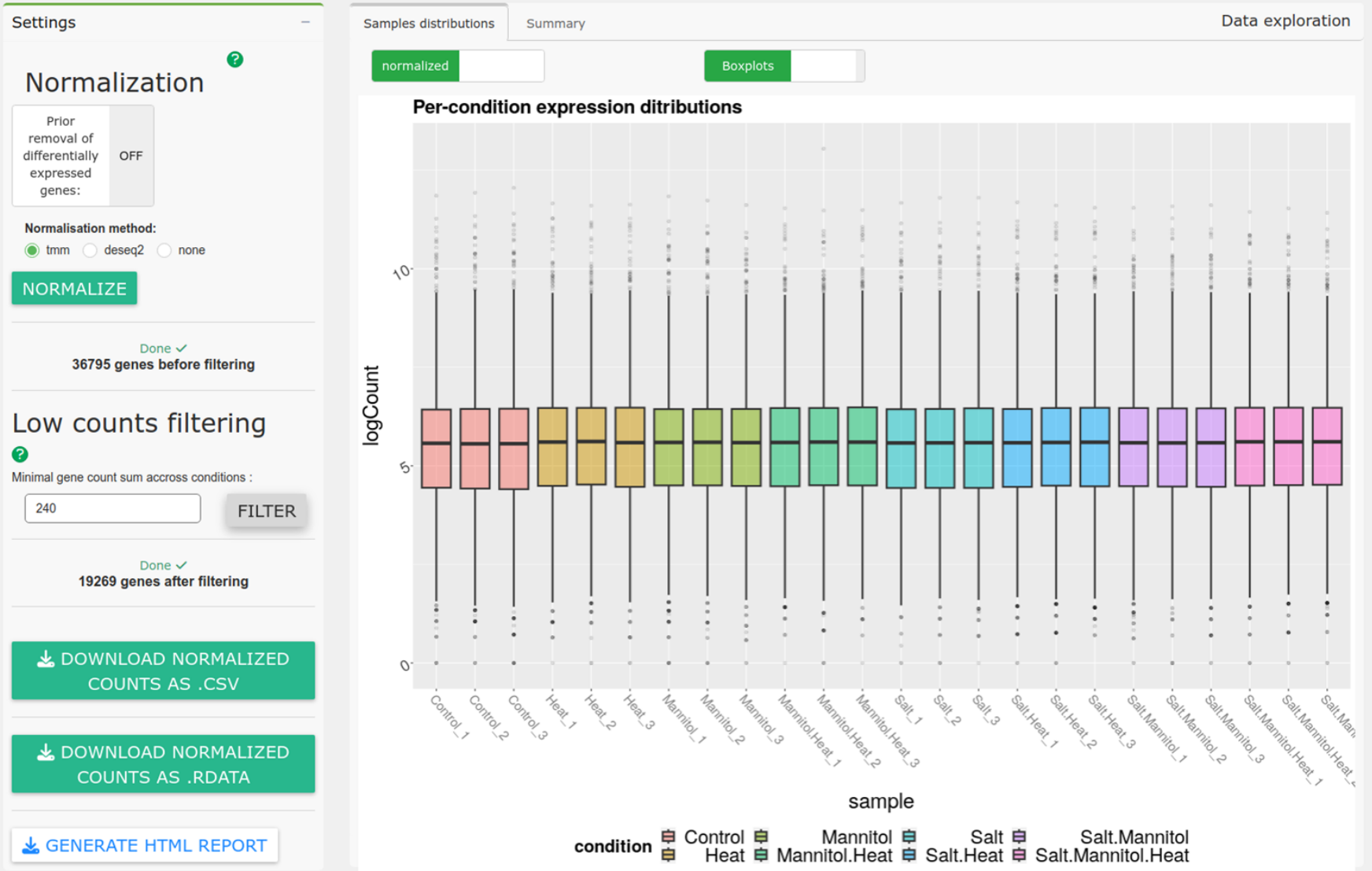


Data filtering and normalisation

Step 1* : normalize using any method



Data filtering and normalisation



Take a look at the distribution. Is it homogene ? Why did we remove low count genes ?

3. Count matrix exploration

Show the first 4 components for a quick overview

Explore normalized gene expression

PCA

Visualize gene expression levels

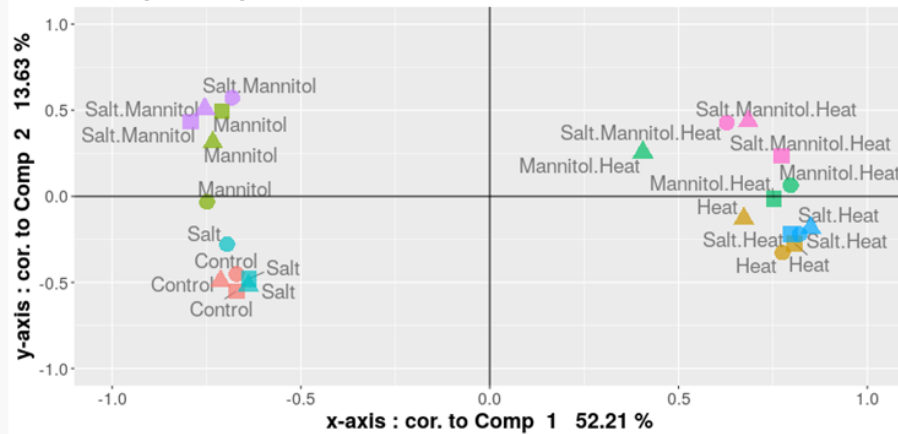
Explore normalized data

PCA Summary

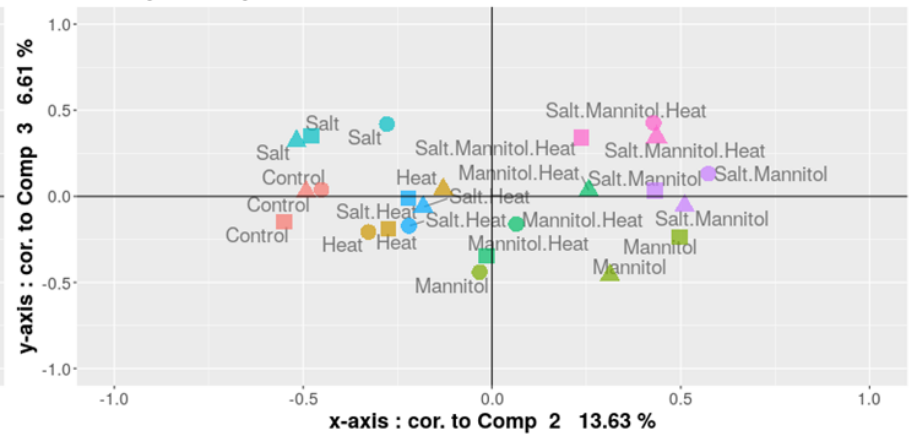
Specific PCA plot

PCA correlation plot

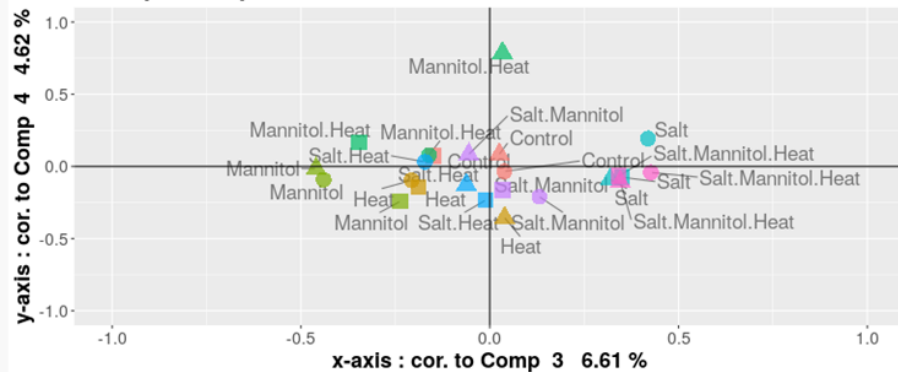
Principal components 1 and 2



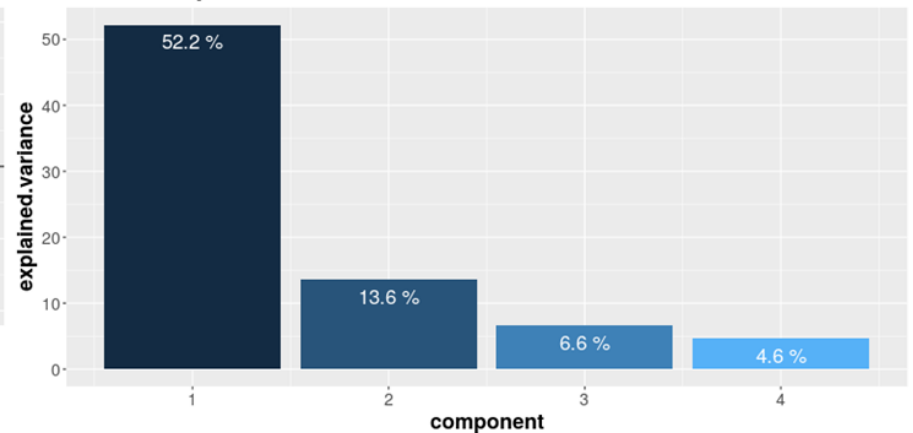
Principal components 2 and 3



Principal components 3 and 4



PCA Screeplot



replicate ● 1 ▲ 2 ■ 3
condition ● Control ● Mannitol ● Salt ● Salt.Mannitol
● Heat ● Mannitol.Heat ● Salt.Heat ● Salt.Mannitol.Heat

DOWNLOAD PCA PLOT

3. Count matrix exploration

Explore normalized gene expression

Show any component comparison



Which condition seems to have the most impact on component 1 ?

3. Count matrix exploration

Explore normalized gene expression



Which condition seems to have the most impact on component 1 ?

3. Count matrix exploration



Which condition seems to have the most impact on component 2 ?

3. Count matrix exploration

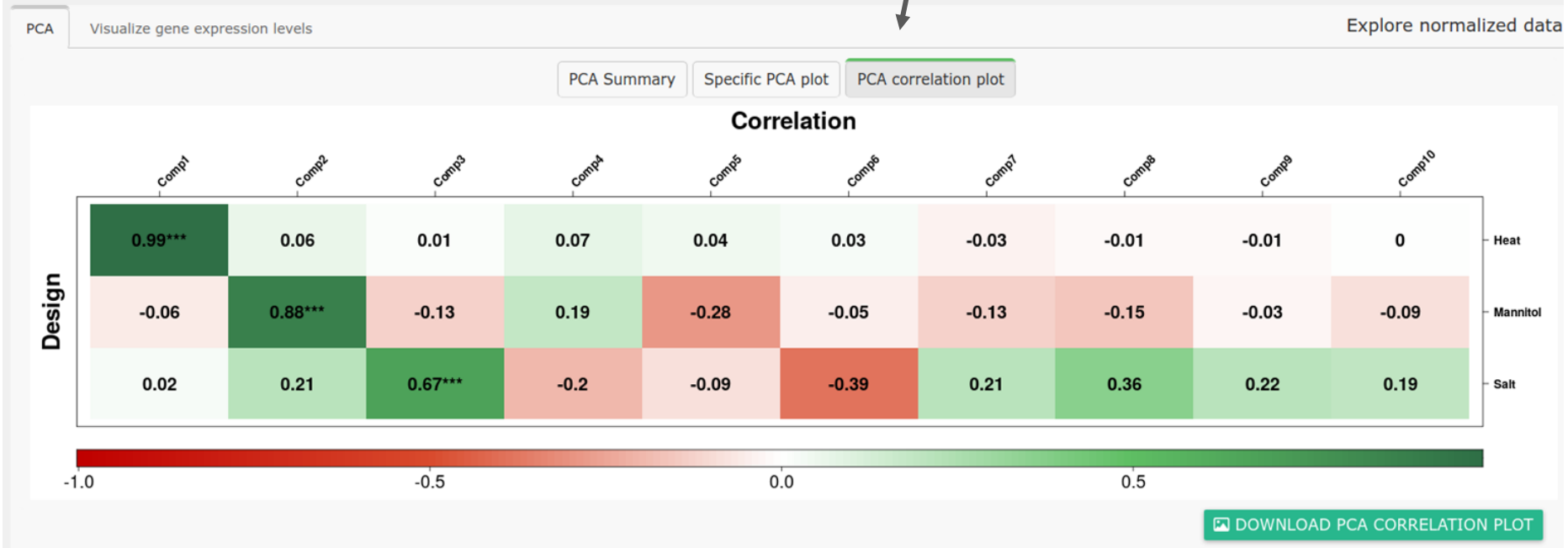


Which condition seems to have the most impact on component 2 ?

3. Count matrix exploration

Show correlation between components and conditions

Explore normalized gene expression



Does this agree with our observations ?

Differential expression analysis

Settings

Estimation of disperion



Conditions to compare for differential analysis :

Reference

✓ C
M
H
MH
SM
S
SH
SMH

Perturbation

C
✓ M
H
MH
SM
S
SH
SMH

Simple comparison

Reference

C

Perturbation

M

Adjusted pvalue (FDR)

0.05

Absolute Log Fold Change (Log_2 (Perturbation / Reference)) :

1

DETECT DIFFERENTIALLY EXPRESSED
GENES

Tests can be performed ✕

Results table

MA - Vulcano plots

Heatmap

Pvalues histogram

Gene Ontology enrichment

Compare genes lists (Venn)

Results

4. Differential expression

Differential expression analysis

1. Choose one or multiple conditions to compare

Settings

Estimation of dispersion



Conditions to compare for differential analysis

Reference

<input checked="" type="checkbox"/> C
<input type="checkbox"/> M
<input type="checkbox"/> H
<input type="checkbox"/> MH
<input type="checkbox"/> SM
<input type="checkbox"/> S
<input type="checkbox"/> SH
<input type="checkbox"/> SMH

Perturbation

<input type="checkbox"/> C
<input checked="" type="checkbox"/> M
<input type="checkbox"/> H
<input type="checkbox"/> MH
<input type="checkbox"/> SM
<input type="checkbox"/> S
<input type="checkbox"/> SH
<input type="checkbox"/> SMH

Simple comparison

Reference
C

Perturbation
M

Adjusted pvalue (FDR)

0.05

Absolute Log Fold Change (Log_2 (Perturbation / Reference)) :

1

**DETECT DIFFERENTIALLY EXPRESSED
GENES**

Tests can be performed ✕

Results table

MA - Vulcano plots

Heatmap

Pvalues histogram

Gene Ontology enrichment

Compare genes lists (Venn)

Results

2. Adjust the parameters

3. Run the analysis

4. Differential expression

Differentially expressed genes

Settings

Estimation of dispersion

?

Conditions to compare for differential analysis :

Reference

✓ C

M

H

MH

SM

S

SH

SMH

Perturbation

C

M

✓ H

MH

SM

S

SH

SMH

Simple comparison

Reference

C

Perturbation

H

Adjusted pvalue (FDR)

0.05

Absolute Log Fold Change (Log2 (Perturbation / Reference)) :

1

DETECT DIFFERENTIALLY EXPRESSED GENES

Done ✓

See plots and tables for more details

Current comparison

C VERSUS H

2061 ▲

up regulated GENES

1443 ▼

down-regulated GENES

Results table

MA - Vulcano plots

Heatmap

Pvalues histogram

Gene Ontology enrichment

Compare genes lists (Venn)

Results

Show 10 entries

Search:

	label	description	logFC	logCPM	FDR	Regulation
AT4G12400	HOP3	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins (Prasad 2010, Pubmed ID: 20856808) with potential to interact with Hsp90/Hsp70 as co-chaperones.	7.731	7.551	0.000	Up
AT5G48570	ROF2	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins (Prasad 2010, Pubmed ID: 20856808) with potential to interact with Hsp90/Hsp70 as co-chaperones.	7.068	8.845	0.000	Up
AT5G12110	AT5G12110	elongation factor 1-beta 1	5.946	7.789	0.000	Up
AT5G64510	TIN1	Encodes Tunicamycin Induced 1(TIN1), a plant-specific ER stress-inducible protein. TIN1 mutation affects pollen surface morphology. Transcriptionally induced by treatment with the N-linked glycosylation inhibitor tunicamycin.	5.854	6.094	0.000	Up
AT2G19310	HSP18.5	HSP20-like chaperones superfamily protein	4.196	7.648	1.266e-313	Up
AT4G25200	HSP23.6-MITO	AthHSP23.6-mito mRNA, nuclear gene encoding mitochondrial	9.387	7.767	2.334e-269	Up
AT1G07350	SR45A	Encodes a serine/arginine rich-like protein, SR45a. Involved in the regulation of stress-responsive alternative splicing.	3.522	7.493	7.150e-259	Up
AT2G29500	HSP17.6B	HSP20-like chaperones superfamily protein	7.714	9.476	2.560e-258	Up
AT1G74310	HSP101	Encodes ClpB1, which belongs to the Casein lytic proteinase/heat shock protein 100 (Clp/Hsp100) family. Involved in refolding of proteins which form aggregates under heat stress. Also known as AtHsp101. AtHsp101 is a cytosolic heat shock protein required for acclimation to high temperature.	4.483	7.190	1.941e-249	Up
AT3G12050	AT3G12050	Aha1 domain-containing protein	2.561	6.907	3.030e-242	Up

Showing 1 to 10 of 3,504 entries

Previous

1

2

3

4

5

...

351

Next

Settings

Estimation of disperion



Conditions to compare for differential analysis :

Reference

✓ C
M
H
MH
SM
S
SH
SMH

Perturbation

C
M
✓ H
MH
SM
S
SH
SMH

Simple comparison

Reference
C

Perturbation
H

Adjusted pvalue (FDR)

Absolute Log Fold Change (Log2 (Perturbation / Reference)) :

DETECT DIFFERENTIALLY EXPRESSED GENES

Done ✓

See plots and tables for more details

Current comparison
C VERSUS H

2061 ▲

up regulated
GENES

1443 ▼

down-regulated
GENES

Results table

MA - Vulcano plots

Heatmap

Pvalues histogram

Gene Ontology enrichment

Compare genes lists (Venn)

Results

Genes to study :

✓ All

Up-regulated

Down-regulated

**START GO ENRICHMENT
ANALYSIS**

✓ Dot plot

Enrichment map

Data table

✓ Biological process

Cellular component

Molecular function

5. DEG list GO enrichment

Settings

Estimation of disperion

Conditions to compare for differential analysis :

Reference	Perturbation
<input checked="" type="checkbox"/> C	<input type="checkbox"/> C
<input type="checkbox"/> M	<input type="checkbox"/> M
<input type="checkbox"/> H	<input checked="" type="checkbox"/> H
<input type="checkbox"/> MH	<input type="checkbox"/> MH
<input type="checkbox"/> SM	<input type="checkbox"/> SM
<input type="checkbox"/> S	<input type="checkbox"/> S
<input type="checkbox"/> SH	<input type="checkbox"/> SH
<input type="checkbox"/> SMH	<input type="checkbox"/> SMH

Simple comparison

Reference	Perturbation
C	H

Adjusted pvalue (FDR)

0.05

Absolute Log Fold Change (Log_2 (Perturbation / Reference)) :

1

DETECT DIFFERENTIALLY EXPRESSED GENES

Done ✓

See plots and tables for more details

Current comparison
C VERSUS H

2061 ▲
up regulated
GENES

1443 ▼
down-regulated
GENES

DOWNLOAD RESULT TABLE AS .TSV

Results table MA - Vulcano plots Heatmap Pvalues histogram **Gene Ontology enrichment** Compare genes lists (Venn)

Results

Genes to study :

All ☒ Up-regulated ☐ Down-regulated

START GO ENRICHMENT ANALYSIS

☒ Dot plot ☐ Biological process

Enrichment map ☐ Cellular component

Data table ☐ Molecular function

Top number of GO terms to plot :

57

Enriched ontologies and their gene count

GO terms (from top to bottom):

- U4 snRNA 3'-end processing-
- tRNA catabolic process-
- rRNA processing-
- rRNA methylation-
- rRNA metabolic process-
- rRNA catabolic process-
- RNA splicing, via transesterification reactions with bulged -
- RNA splicing, via transesterification reactions-
- RNA phosphodiester bond hydrolysis-
- RNA modification-
- RNA methylation-
- RNA 3'-end processing-
- ribosome biogenesis-
- ribosomal large subunit biogenesis-
- ribonucleoprotein complex biogenesis-
- response to unfolded protein-
- response to reactive oxygen species-
- response to oxygen levels-
- response to hypoxia-
- response to hydrogen peroxide-
- response to heat-
- response to endoplasmic reticulum stress-
- response to decreased oxygen levels-
- regulation of ATP-dependent activity-
- protein refolding-
- protein folding-
- protein complex oligomerization-
- polar nucleus fusion-
- nucleus organization-
- nuclear-transcribed mRNA catabolic process, exonucleolytic-
- nuclear RNA surveillance-
- nuclear polyadenylation-dependent tRNA catabolic process-
- negative regulation of signaling-
- negative regulation of signal transduction-
- negative regulation of cell communication-
- ncRNA processing-
- mRNA splicing, via spliceosome-
- mRNA metabolic process-
- mitochondrion organization-
- mitochondrial RNA metabolic process-
- mitochondrial mRNA modification-
- mitochondrial gene expression-
- karyogamy-
- exonucleolytic trimming to generate mature 3'-end of 5.8S rR-
- exonucleolytic trimming involved in rRNA processing-
- endoplasmic reticulum unfolded protein response-
- embryo sac development-
- embryo sac central cell differentiation-
- chaperone-mediated protein folding-
- chaperone cofactor-dependent protein refolding-
- cellular response to unfolded protein-
- cellular response to oxygen levels-
- cellular response to hypoxia-
- cellular response to heat-
- cellular response to decreased oxygen levels-
- 'de novo' protein folding-
- 'de novo' post-translational protein folding-

Count

p.adjust

0.03

0.02

0.01

Look at the GO term with the lowest pvalue. Is it coherent ?

5. DEG list GO enrichment

Settings

Estimation of disperion

?

Conditions to compare for differential analysis :

Reference

✓ C

M

H

MH

SM

S

SH

SMH

Perturbation

C

M

✓ H

MH

SM

S

SH

SMH

Simple comparison

Reference

C

Perturbation

H

Adjusted pvalue (FDR)

0.05

Absolute Log Fold Change (Log2 (Perturbation / Reference)) :

1

DETECT DIFFERENTIALLY EXPRESSED GENES

Done ✓

See plots and tables for more details

Current comparison

C VERSUS H

2061 ▲

up regulated

GENES

1443 ▼

down-regulated

GENES

DOWNLOAD RESULT TABLE AS .TSV

Results table

MA - Vulcano plots

Heatmap

Pvalues histogram

Gene Ontology enrichment

Compare genes lists (Venn)

Results

Genes to study :

All

✓ Up-regulated

Down-regulated

START GO ENRICHMENT ANALYSIS

✓ Dot plot

Enrichment map

Data table

✓ Biological process

Cellular component

Molecular function

Top number of GO terms to plot :

57

Enriched ontologies and their gene count

GO terms (from top to bottom):

- U4 snRNA 3'-end processing-
- tRNA catabolic process-
- rRNA processing-
- rRNA methylation-
- rRNA metabolic process-
- rRNA catabolic process-
- RNA splicing, via transesterification reactions with bulged -
- RNA splicing, via transesterification reactions-
- RNA phosphodiester bond hydrolysis-
- RNA modification-
- RNA methylation-
- RNA 3'-end processing-
- ribosome biogenesis-
- ribosomal large subunit biogenesis-
- ribonucleoprotein complex biogenesis-
- response to unfolded protein-
- response to reactive oxygen species-
- response to oxygen levels-
- response to hypoxia-
- response to hydrogen peroxide-
- response to heat-
- response to endoplasmic reticulum stress-
- response to decreased oxygen levels-
- regulation of ATP-dependent activity-
- protein refolding-
- protein folding-
- protein complex oligomerization-
- polar nucleus fusion-
- nucleus organization-
- nuclear-transcribed mRNA catabolic process, exonucleolytic-
- nuclear RNA surveillance-
- nuclear polyadenylation-dependent tRNA catabolic process-
- negative regulation of signaling-
- negative regulation of signal transduction-
- negative regulation of cell communication-
- ncRNA processing-
- mRNA splicing, via spliceosome-
- mRNA metabolic process-
- mitochondrion organization-
- mitochondrial RNA metabolic process-
- mitochondrial mRNA modification-
- mitochondrial gene expression-
- karyogamy-
- exonucleolytic trimming to generate mature 3'-end of 5.8S rR-
- exonucleolytic trimming involved in rRNA processing-
- endoplasmic reticulum unfolded protein response-
- embryo sac development-
- embryo sac central cell differentiation-
- chaperone-mediated protein folding-
- chaperone cofactor-dependent protein refolding-
- cellular response to unfolded protein-
- cellular response to oxygen levels-
- cellular response to hypoxia-
- cellular response to heat-
- cellular response to decreased oxygen levels-
- 'de novo' protein folding-
- 'de novo' post-translational protein folding-

Count: 25, 50, 75, 100

p.adjust: 0.03, 0.02, 0.01

Look at the GO term with the lowest pvalue. Is it coherent ?

Response to heat (what a surprise!)