From gene expression to networks: R workflow development for analysis and visualization of RNA-seq data

LST&M LABORATOIRE DES SYMBIOSES TROPICALES & MÉDITERRANÉENNES

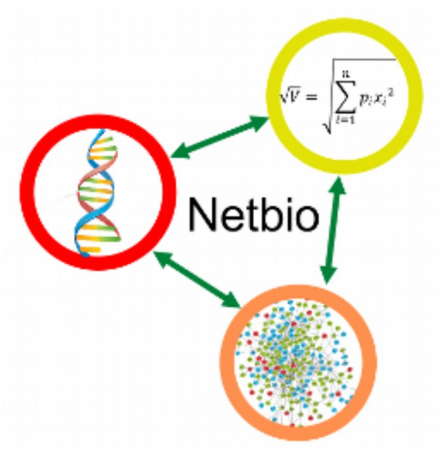
AGENCE NATIONALE DE LA RECHERCHE ANR PSYCHE

Ilana LAMBERT

PhD student

Medicago truncatula

Gene expression



RNA-seq data analysis

Networks

INRA SCIENCE & IMPACT

cirad LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT

IRD Institut de Recherche pour le Développement FRANCE

Ecole Doctorale GAIA Biodiversité Agriculture Alimentation Environnement Terre Eau

SupAgro Montpellier

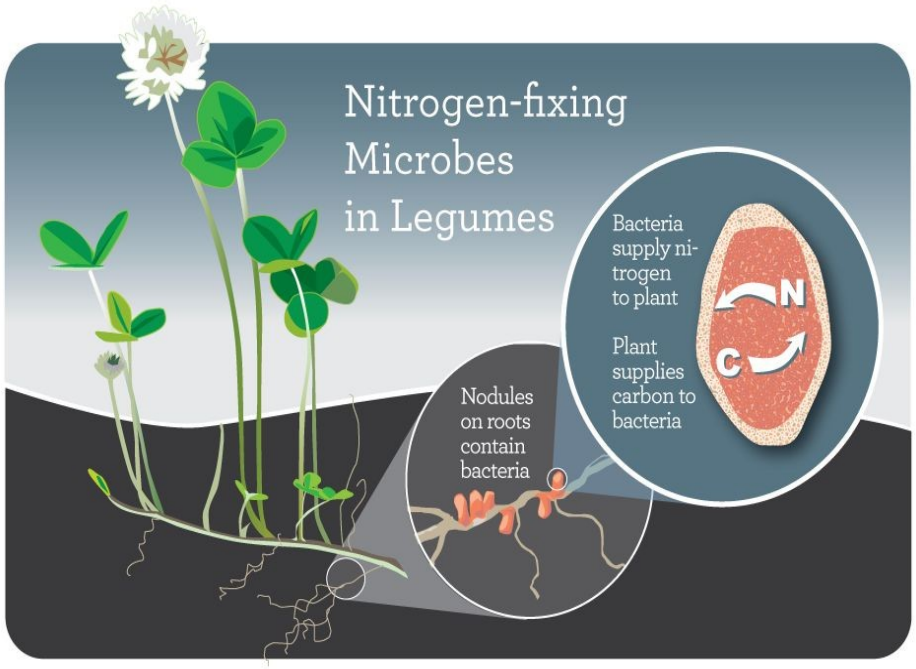
RSV Laboratoire de Recherche en Sciences Végétales

IPS2 Institute of Plant Sciences Paris - Saclay

Medicago truncatula

Legume plants :

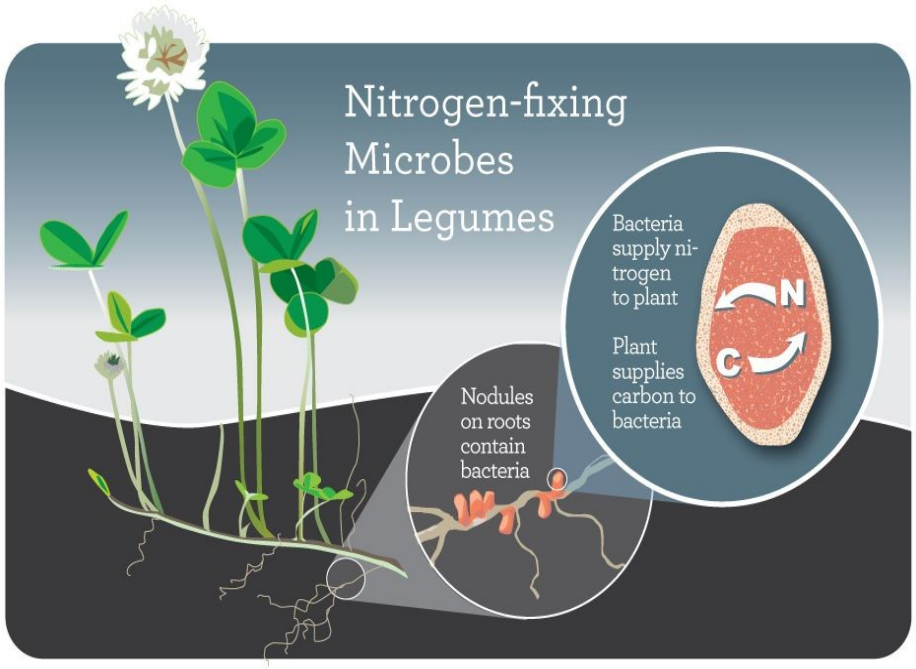
- Atmospheric N2 → a non-limiting source of nitrogen
- Symbiosis with Rhizobia
- Nodules
- Symbiotic nitrogen fixation (SNF)



Medicago truncatula

Legume plants :

- Atmospheric N2 → a non-limiting source of nitrogen
- Symbiosis with Rhizobia
- Nodules
- Symbiotic nitrogen fixation (SNF)



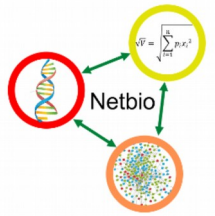
Development and function of root and nodule

→ sensitive to environmental **stresses**

Abiotic stresses

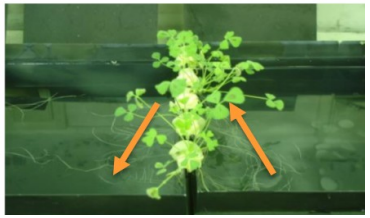
Biotic stresses

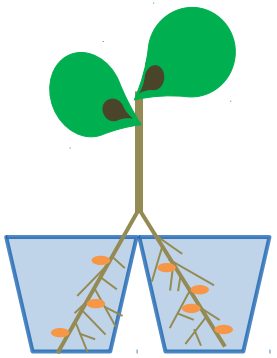
- Reduction of nitrogen fixation efficiency
- Unstable yields in legume crops



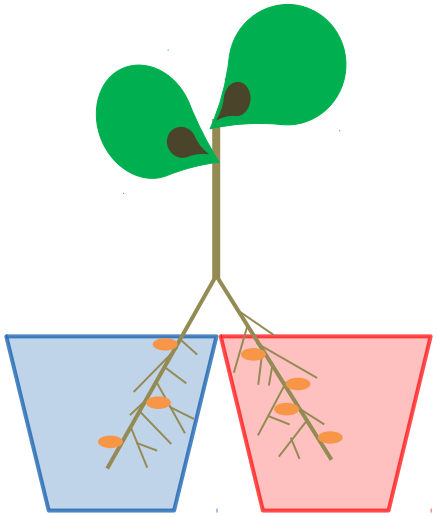
Systemic response

Split root experimental system



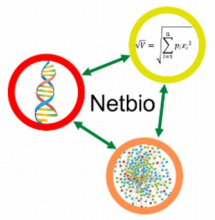


Control



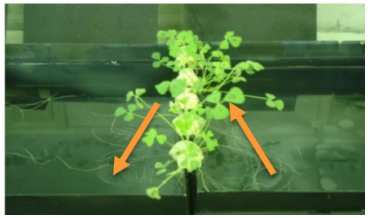
Favorable area

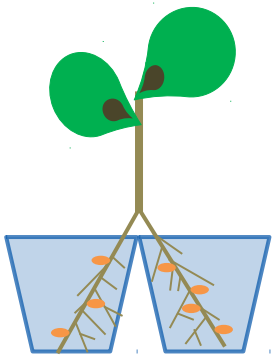
Stress area



Systemic response

Split root experimental system



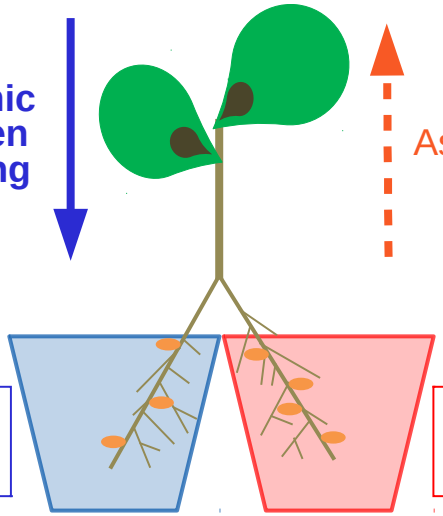


Control

Nitrogen demand « deficit »

Systemic nitrogen signaling

Assimilation



✔Simulation of nodule expansion
✔Simulation of nodule formation

Systemic nitrogen signaling

Compensatory response

Favorable area

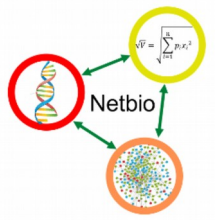
Stress area

Ex : nitrogen deficit

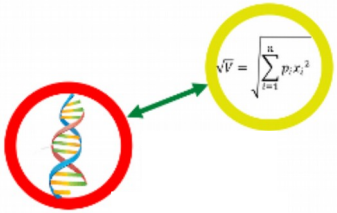
✔Arrest of nodule formation
✔Arrest of nodule development

Symbiotic nitrogen fixation inhibition

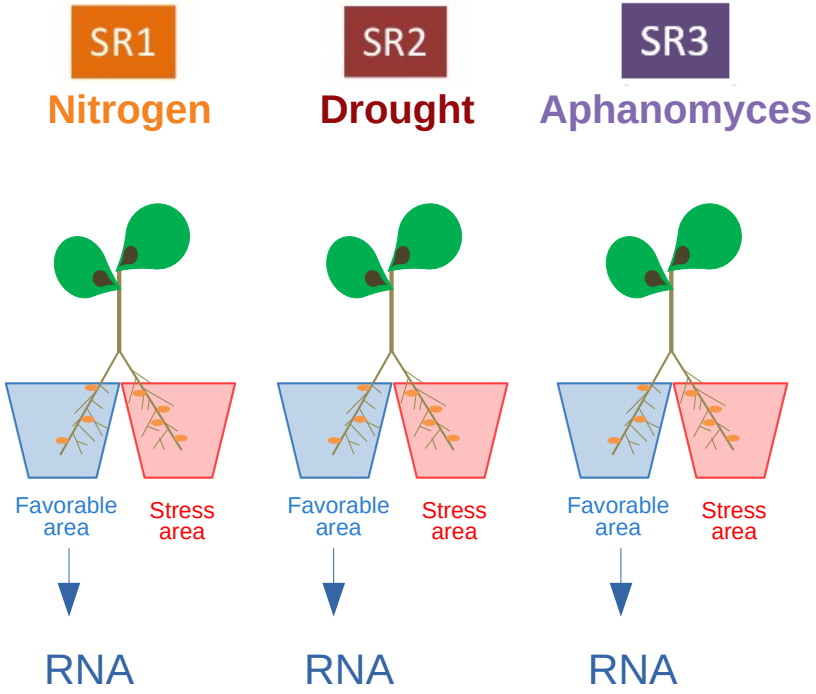
Local nitrogen signaling

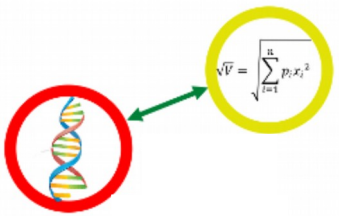


Ruffel et al, 2008 Plant Physiol.
Jeudy et al, 2010, New phytol.
Laguerre et al, 2012, New phytol.



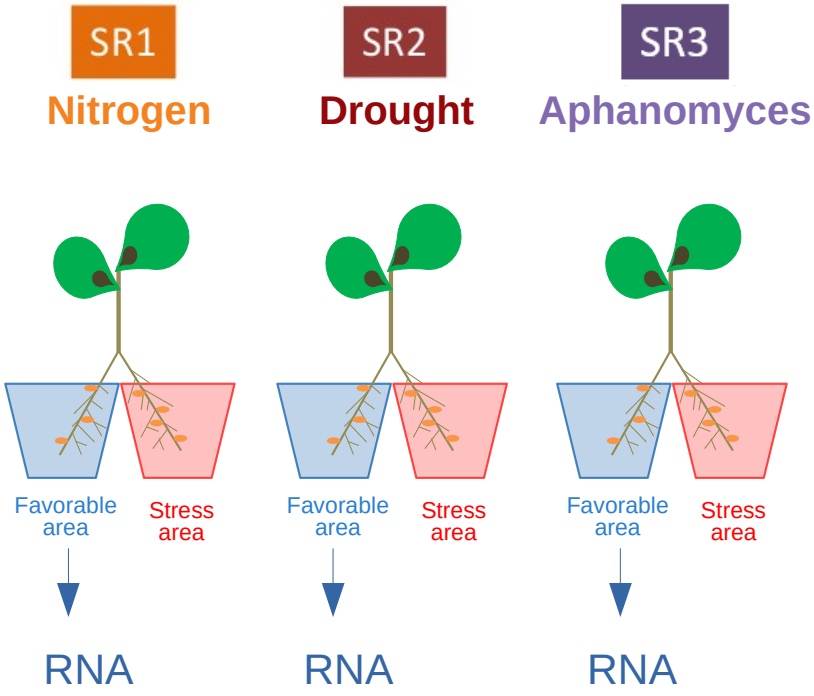
What are the molecular mechanisms underline the systemic signaling in response to abiotic and biotic stresses in *Medicago truncatula*?

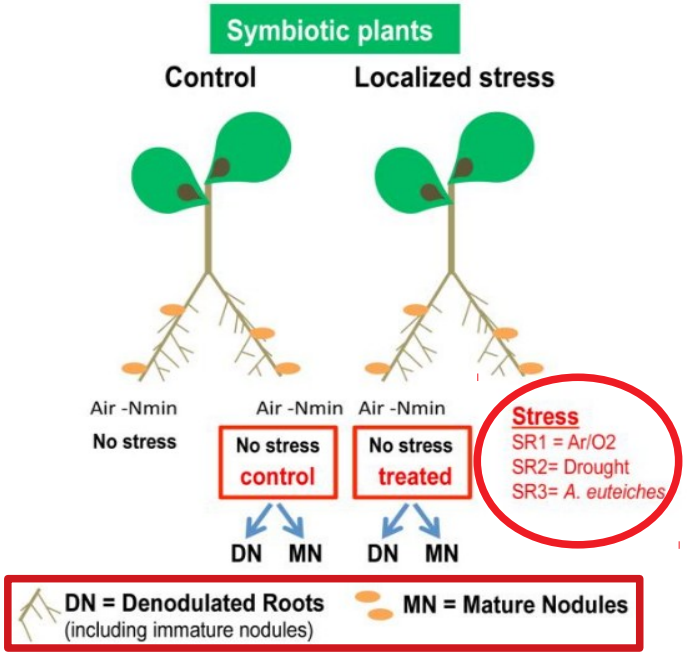




What are the molecular mechanisms underline the systemic signaling in response to abiotic and biotic stresses in *Medicago truncatula*?

PSYCHE



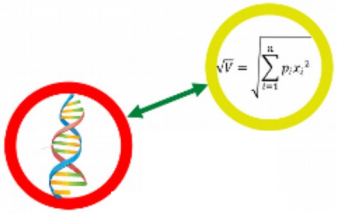


Stress factor

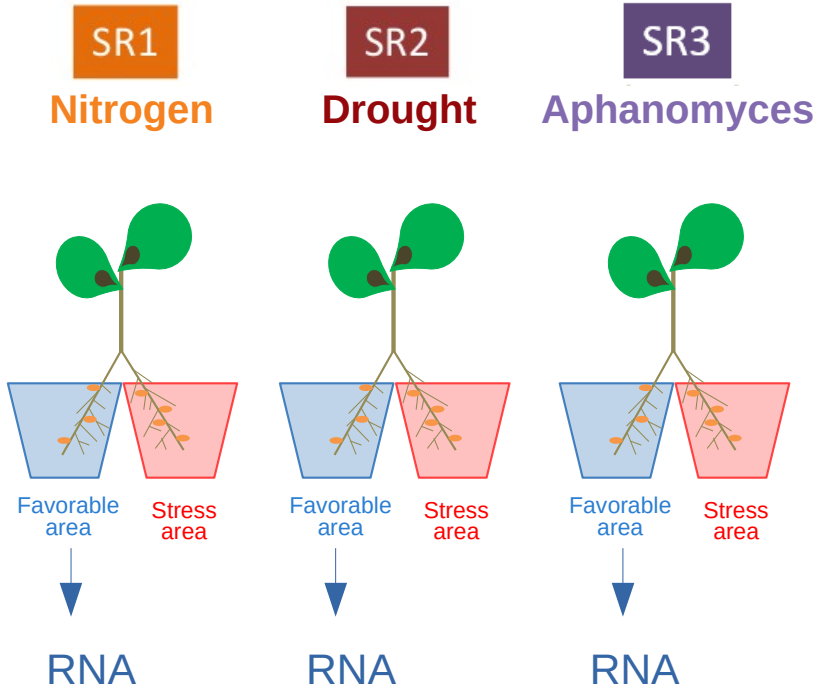
Tissu factor

Time-line treatment factor
Control – 6h – 1d – 3d – 5d

Replicate factor (3)



What are the molecular mechanisms underline the systemic signaling in response to abiotic and biotic stresses in *Medicago truncatula*?



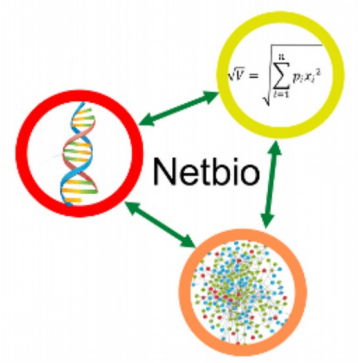
Functional annotation

Libraries preparation

Sequencing

Mapping

RNAseq data analysis



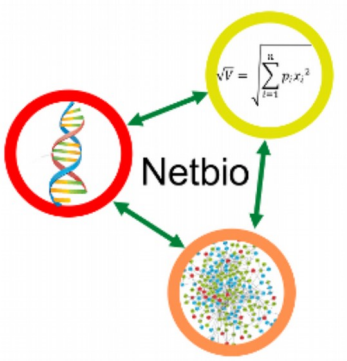
Are these mechanisms, common or specific to different stresses ?

SR1 **Nitrogen**

SR2 **Drought**

SR3 **Aphanomyces**

RNAseq data analysis



Are these mechanisms, common or specific to different stresses ?

Functional annotation

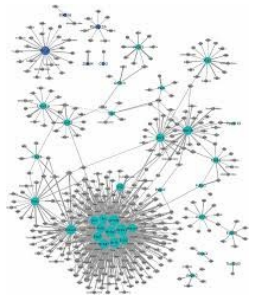
SR1 **Nitrogen**

SR2 **Drought**

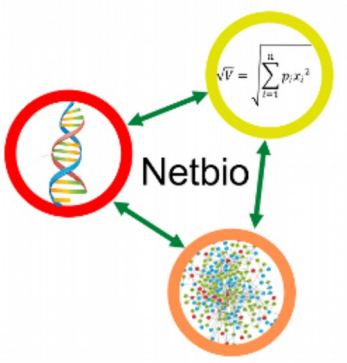
SR3 **Aphanomyces**

RNAseq data analysis

Comparisons and analysis of **gene expression networks** identified in response to different stresses



Thesis project



Are these mechanisms, common or specific to different stresses ?

Functional annotation

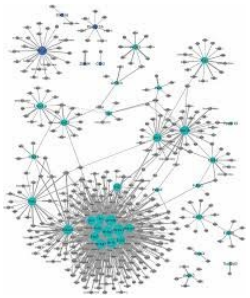
SR1 **Nitrogen**

SR2 **Drought**

SR3 **Aphanomyces**

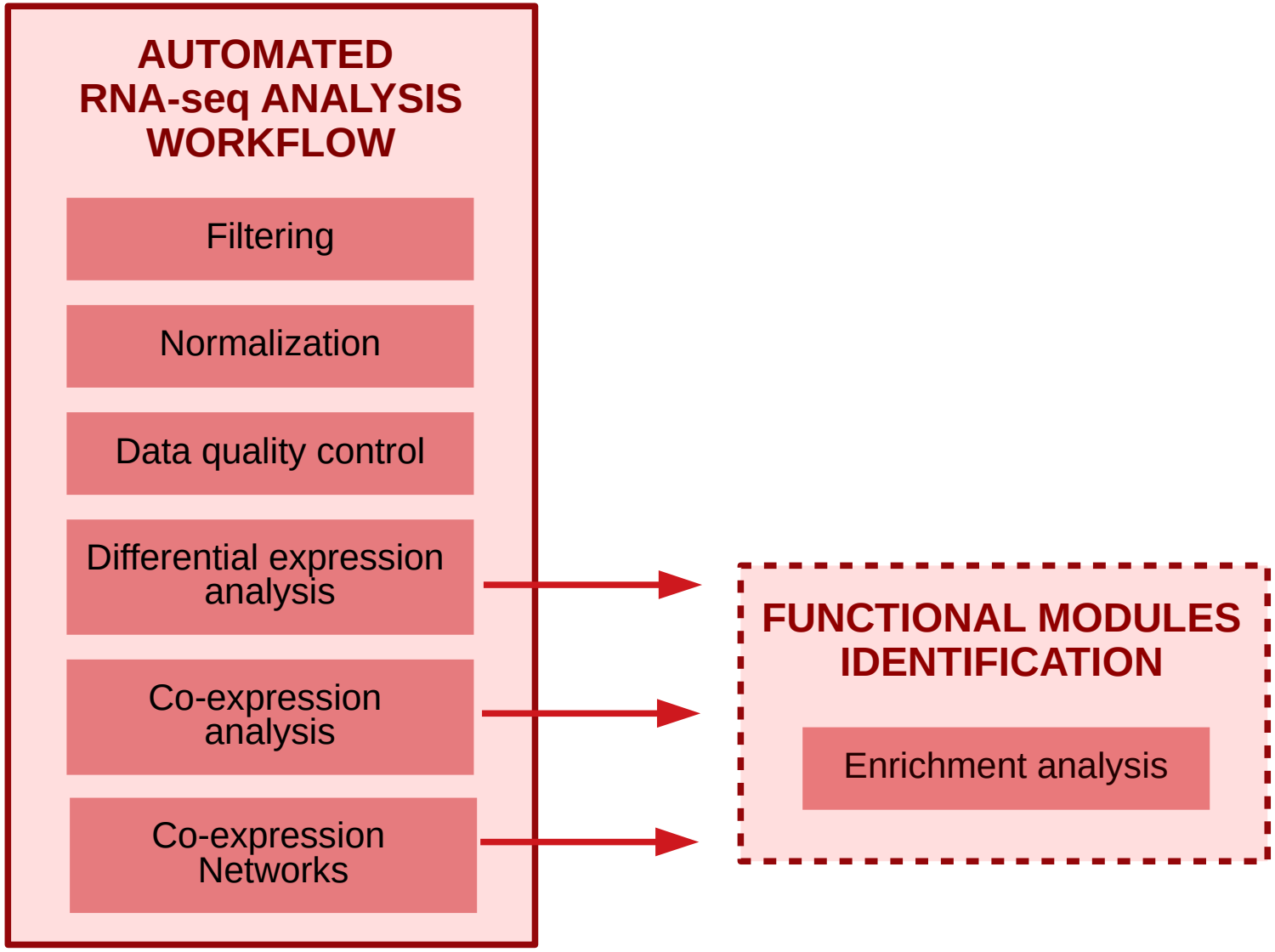
RNAseq data analysis

Comparisons and analysis of **gene expression networks** identified in response to different stresses



Automated R workflow for analysis and visualization of RNA-seq data

R workflow


AUTOMATED
RNA-seq ANALYSIS
WORKFLOW
Filtering
Normalization
Data quality control
Differential expression analysis
Co-expression analysis
Co-expression Networks
FUNCTIONAL MODULES IDENTIFICATION
Enrichment analysis

Filtering & Normalization

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Filtering

Method

Counts per million (CPM)

Tool

cpm function

edgeR package[1] (v.3.22.3)

Input

Table of counts

Outputs

Non expressed genes

Low expressed genes

Normalization[2]

Method

Trimmed Mean of M-values (TMM)

Tool

calcNormFactors function

edgeR package

Input

Filtered counts table

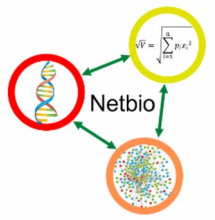
Outputs

Normalized counts

Mean;STD Normalized counts

Log2 Normalized counts

Log2 Mean;STD Normalized counts



1 Robinson *et al.*, 2010, Bioinformatics

2 Dillies et al., 2013, Briefing in Bioinformatics

Data quality control

AUTOMATED RNA-seq ANALYSIS WORKFLOW

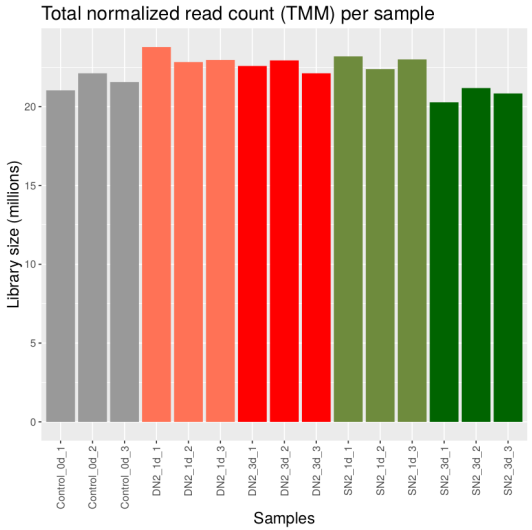
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

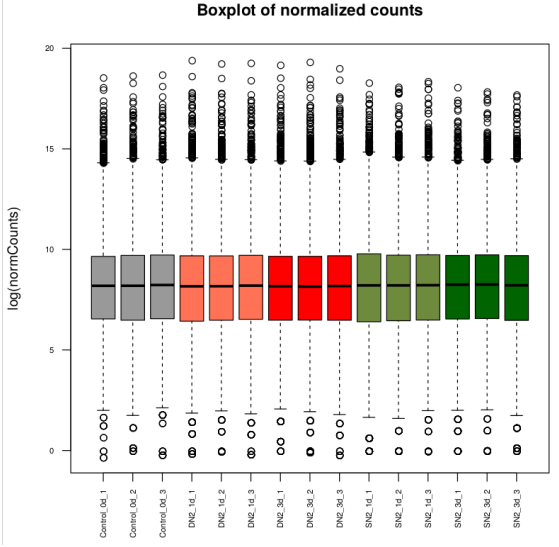
FUNCTIONAL MODULES IDENTIFICATION

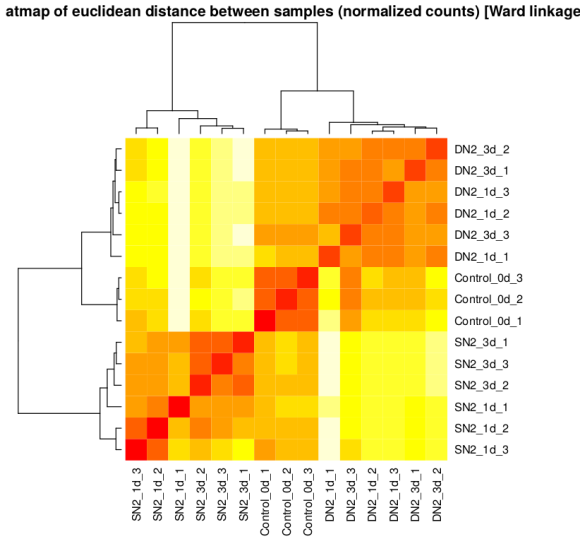
- Enrichment analysis

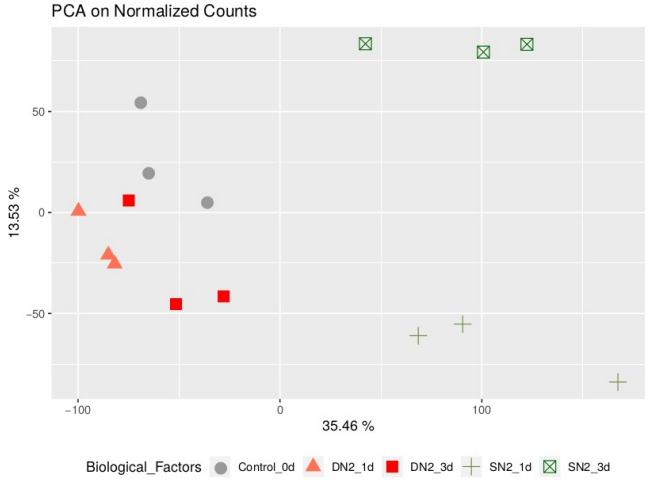
Tool

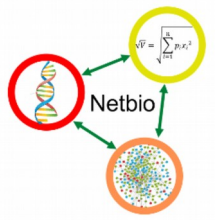
ggplot2, stats, FactoMineR packages











13/12/18

Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

Generalized linear models[1]

Tool

edgeR

For each gene:

Nitrogen factor $i \in \{DN2, SN2\}$

Time factor $j \in \{0d, 1d, 3d\}$

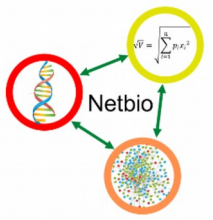
Replicate factor $h \in \{1, 2, 3\}$

Mean expression

$$\log(\lambda_{i,j,h}) = \mu + \alpha_i + \beta_j + \gamma_h + (\alpha\beta)_{i,j}$$

Interaction Nitrogen – Time

Example



[1] Rigaill et al., 2018, Briefings in Bioinformatics

Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

Generalized linear models[1]

Tool

edgeR

For each gene:

Nitrogen factor $i \in \{DN2,SN2\}$

Time factor $j \in \{0d,1d,3d\}$

Replicate factor $h \in \{1,2,3\}$

Mean expression

Example

$$\log(\lambda_{i,j,h}) = \mu + \alpha_i + \beta_j + \gamma_h + (\alpha\beta)_{i,j}$$

→ Interaction Nitrogen – Time

Biological question

Differential genes expressed between Deficit Nitrogen and Satiety Nitrogen conditions

Contrastes

(DN2) - (SN2)

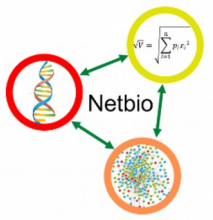
Equation

$$(\mu + \alpha_{DN2} + \beta_j + \gamma_h + (\alpha\beta)_{DN2,j}) - (\mu + \alpha_{SN2} + \beta_j + \gamma_h + (\alpha\beta)_{SN2,j})$$

$$(\alpha_{DN2} - \alpha_{SN2}) + ((\alpha\beta)_{DN2,j} - (\alpha\beta)_{SN2,j})$$

script

(DN2-SN2) <- c(0,-1,0,0,0,-(1/3),-(1/3))



13/12/18

[1] Rigaill et al., 2018, Briefings in Bioinformatics

Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

Generalized linear models[1]

Tool

edgeR

For each gene:

Nitrogen factor $i \in \{DN2,SN2\}$

Time factor $j \in \{0d,1d,3d\}$

Replicate factor $h \in \{1,2,3\}$

Example

Mean expression

$$\log(\lambda_{i,j,h}) = \mu + \alpha_i + \beta_j + \gamma_h + (\alpha\beta)_{i,j}$$

→ Interaction Nitrogen – Time

Biological question

Differential genes expressed between Deficit Nitrogen and Satiety Nitrogen conditions

Contrastes

(DN2) - (SN2)

Equation

$$(\mu + \alpha_{DN2} + \beta_j + \gamma_h + (\alpha\beta)_{DN2,j}) - (\mu + \alpha_{SN2} + \beta_j + \gamma_h + (\alpha\beta)_{SN2,j})$$

$$(\alpha_{DN2} - \alpha_{SN2}) + ((\alpha\beta)_{DN2,j} - (\alpha\beta)_{SN2,j})$$

script **(DN2-SN2) <- c(0,-1,0,0,0,-(1/3),-(1/3))**

Stress effect
(DN2 – SN2)

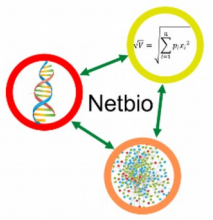
Time effect
(0d – 1d)
(0d – 3d)
(1d – 3d)

Interaction effect
(DN2_0d-DN2_1d)-(SN2_0d-SN2_1d)
(DN2_0d-DN2_3d)-(SN2_0d-SN2_3d)
(DN2_1d-DN2_3d)-(SN2_1d-SN2_3d)

16 contrastes

Stress effect for each time
(0d_DN2-0d_SN2)
(1d_DN2-1d_SN2)
(3d_DN2-3d_SN2)

Time effect for each stress
(DN2_0d-DN2_1d) (SN2_0d-SN2_1d)
(DN2_0d-DN2_3d) (SN2_0d-SN2_3d)
(DN2_1d-DN2_3d) (SN2_1d-SN2_3d)



Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

Generalized linear models[1]

Tool

edgeR

For each gene:

Mean expression

Nitrogen factor $i \in \{DN2,SN2\}$

Time factor $j \in \{0d,1d,3d\}$

Replicate factor $h \in \{1,2,3\}$

Example

$$\log(\lambda_{i,j,h}) = \mu + \alpha_i + \beta_j + \gamma_h + (\alpha\beta)_{i,j}$$

→ Interaction Nitrogen – Time

Biological question

Differential genes expressed between Deficit Nitrogen and Satiety Nitrogen conditions

Contrastes

(DN2) - (SN2)

Equation

$$(\mu + \alpha_{DN2} + \beta_j + \gamma_h + (\alpha\beta)_{DN2,j}) - (\mu + \alpha_{SN2} + \beta_j + \gamma_h + (\alpha\beta)_{SN2,j})$$

$$(\alpha_{DN2} - \alpha_{SN2}) + ((\alpha\beta)_{DN2,j} - (\alpha\beta)_{SN2,j})$$

script

(DN2-SN2) <- c(0,-1,0,0,0,-(1/3),-(1/3))

Stress effect
(DN2 – SN2)

Time effect
(0d – 1d)
(0d
(1d

Interaction effect
(DN2_0d-DN2_1d)-(SN2_0d-SN2_1d)
0d-DN2_3d)-(SN2_0d-SN2_3d)
1d-DN2_3d)-(SN2_1d-SN2_3d)

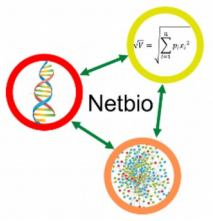
R function for automated writing all contrasts

16 contrastes

Stress effect for each time
(0d_DN2-0d_SN2)
(1d_DN2-1d_SN2)
(3d_DN2-3d_SN2)

(DN2_1d-DN2_3d)

(SN2_0d-SN2_1d)
(SN2_0d-SN2_3d)
(SN2_1d-SN2_3d)



Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Tool

edgeR

R function for automated writing all contrasts

Input

Normalized counts

Selection of contrasts of interest

Estimating dispersion

estimateGLMCommonDisp()
estimateGLMTrendedDisp()
estimateGLMTagwiseDisp()

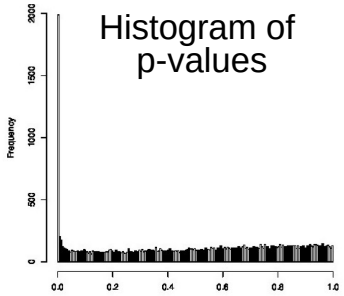
Inference (parameter estimation)

glmFit()

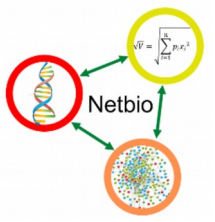
For each contrast

Likelihood ratio test

glmLRT()



Adjustement of the raw p-values
Benjamini-Hochberg procedure
Decision rule : FDR < 0.05
topTags()



Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Outputs

For each contrast

DEGs Results

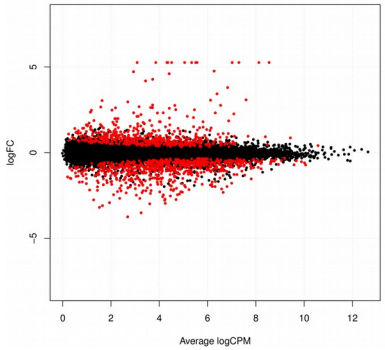
Gene_ID	mRNA_ID	Gene_Name
Medtr0007s0390	Medtr0007s0390.1	receptor-like Serine/Threonine-kinase plant
Medtr0184s0010	Medtr0184s0010.1	UDP-glucosyltransferase
Medtr0221s0020	Medtr0221s0020.2	transmembrane protein, putative
Medtr0221s0020	Medtr0221s0020.1	transmembrane protein, putative
Medtr0384s0020	Medtr0384s0020.1	hypothetical protein
Medtr0508s0010	Medtr0508s0010.1	ribosomal protein S8e family protein

DEGs Normalized counts

logFC	logCPM	LR	PValue	FDR
5.55391970508026	4.81041290555745	16.1086222041737	5.98111063390226e-05	0.0108642973940508
4.58308025736963	5.32108313676033	16.1233960404379	5.93463462776849e-05	0.0108642973940508
-4.58621127038228	6.11277538970362	16.1178308342364	5.95209903996132e-05	0.0108642973940508
-4.58621127038228	6.11277538970362	16.1178308342364	5.95209903996132e-05	0.0108642973940508
-5.45355905337383	5.690285243976	19.7243656112101	8.94535369589355e-06	0.00266942133362172
4.85866597399211	5.95949005692566	11.8634299454211	0.000572470120976213	0.0488095034982535

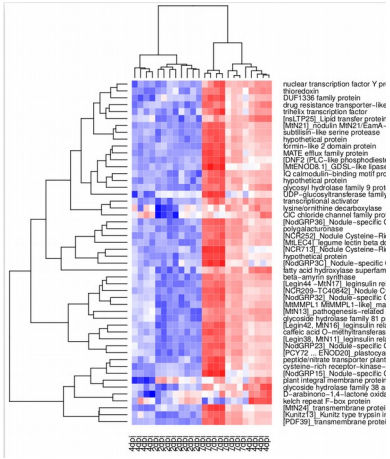
DEGs Log2 Normalized counts

MA-plot



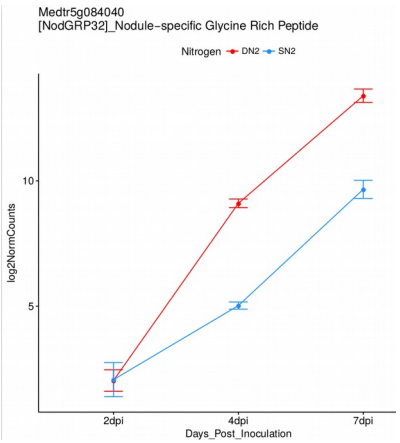
plotSmear function (edgeR package)

Gene clustering

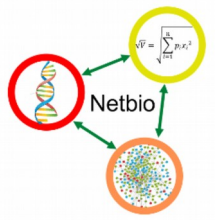


heatmap.2 function (gplots package) (v.3.0.1)

Single gene expression profile



ggline function (ggpubr package) (v.0.1.8)



Differential expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

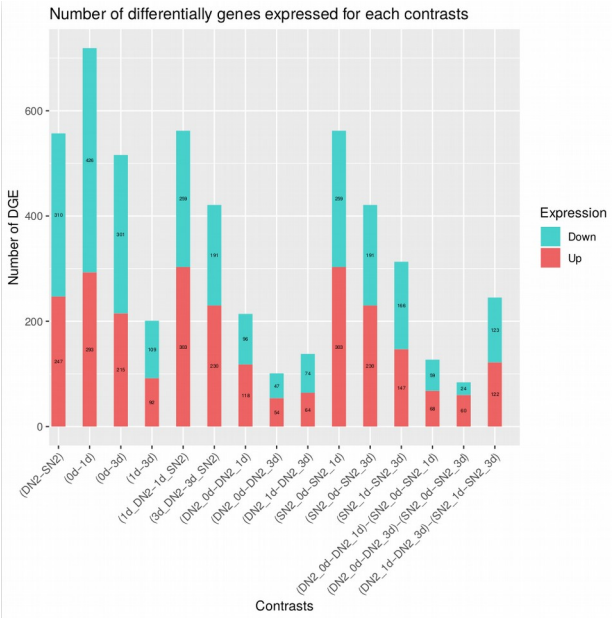
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Outputs

Summary results



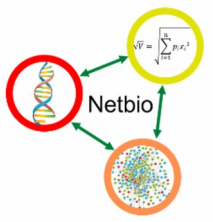
Histogram of Up/Down DGE
(ggplot2 package)

Comparison table

(data.table package)
(v.1.11.4)

Gene_ID	Contrast1	Contrast2	Contrast3	Contrast4	Contrast5	Contrast6
MtrunA17CPg0492331	0	0	0	1	0	0
MtrunA17Chr0c01g0488931	1	0	0	1	1	0
MtrunA17Chr0c01g0489041	1	0	0	1	1	0
MtrunA17Chr0c01g0489091	0	0	0	1	0	0
MtrunA17Chr0c01g0489181	0	0	0	0	0	0
MtrunA17Chr0c01g0489191	0	0	0	0	0	0

Venn diagram – Intersection – Union function



Co-expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

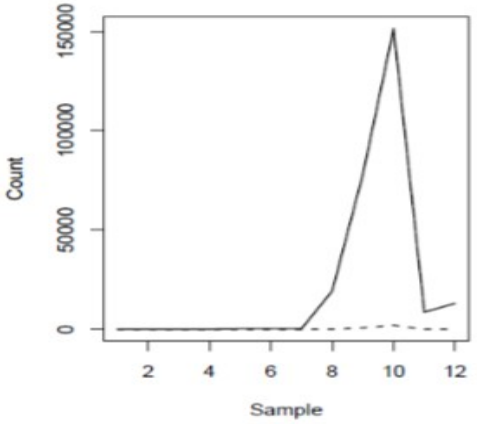
Method[1]

Tool

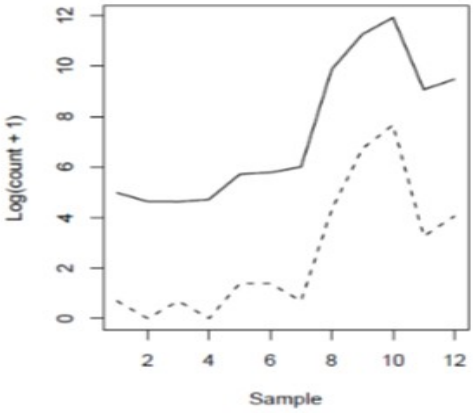
coseq package
(v.1.4.0)

Normalized expression profiles

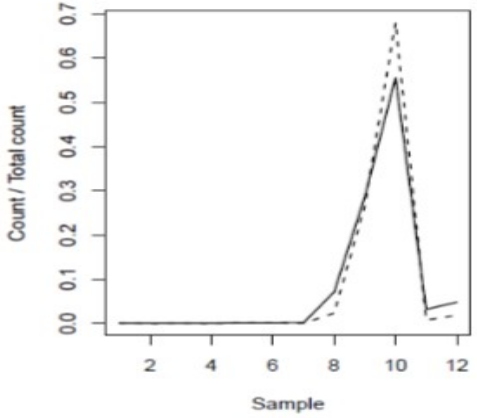
Raw counts



Log(counts+1)



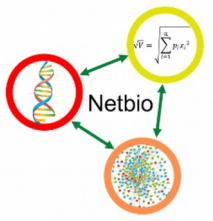
Normalized profiles



Normalized profile for gene i :

$$p_{ij} = \frac{y_{ij}/s_j + 1}{\sum_\ell y_{i\ell}/s_\ell + 1}$$

S_j : scaling normalization factors of raw counts



[1] Rau *et al.*, 2015, *Bioinformatics*

Co-expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

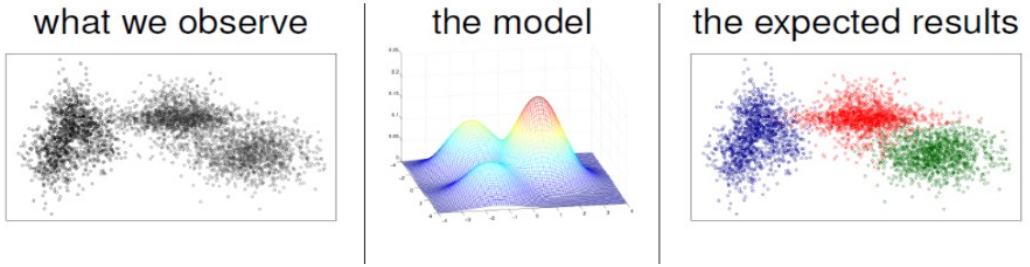
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method[1]

Gaussian mixture model



Transformation arcsin
Model Normal

Expectation–maximization (EM) algorithm

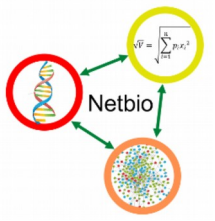
Tool

coseq package
(v.1.4.0)

Input

Raw counts table

List of DEGs



[1] Rau *et al.*, 2015, *Bioinformatics*

Co-expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

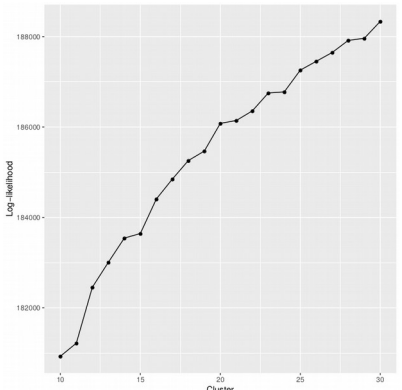
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

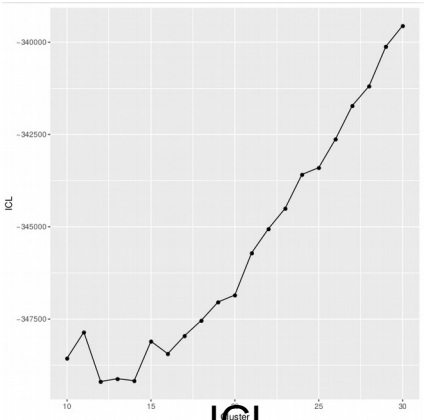
- Enrichment analysis

Outputs

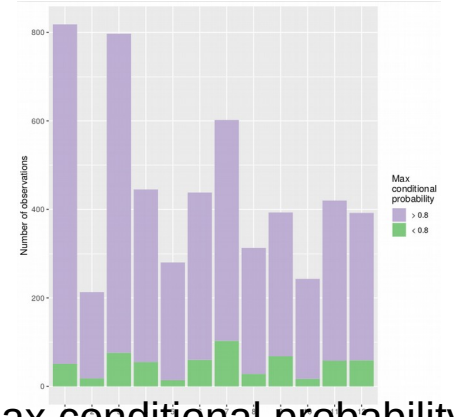
```
Summary coseq:
**********************************************
Model: Gaussian_pk_Lk_Ck
Transformation: arcsin
**********************************************
Clusters fit: 10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30
Clusters with errors: ---
Selected number of clusters via ICL: 12
ICL of selected model: -349190.3
**********************************************
Number of clusters = 12
ICL = -349190.3
**********************************************
Cluster sizes:
 Cluster 1  Cluster 2  Cluster 3  Cluster 4  Cluster 5  Cluster 6  Cluster 7  Cluster 8
       818        213        797        445        280        438        602        313
 Cluster 9 Cluster 10 Cluster 11 Cluster 12
       393        243        420        392

Number of observations with MAP > 0.90 (% of total):
4310 (80.5%)

Number of observations with MAP > 0.90 per cluster (% of total per cluster):
 Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5 Cluster 6 Cluster 7 Cluster 8 Cluster 9
 710       186       663       354       248       349       430       269       279
 (86.8%)   (87.32%)  (83.19%)  (79.55%)  (88.57%)  (79.68%)  (71.43%)  (85.94%)  (70.99%)
 Cluster 10 Cluster 11 Cluster 12
 201        327        294
 (82.72%)   (77.86%)   (75%)
```



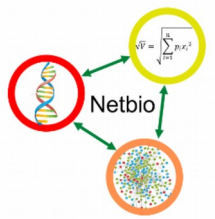
Log Likelihood



ICL



Max conditional probability



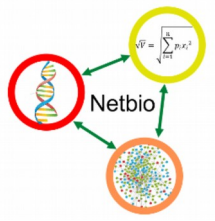
Co-expression analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

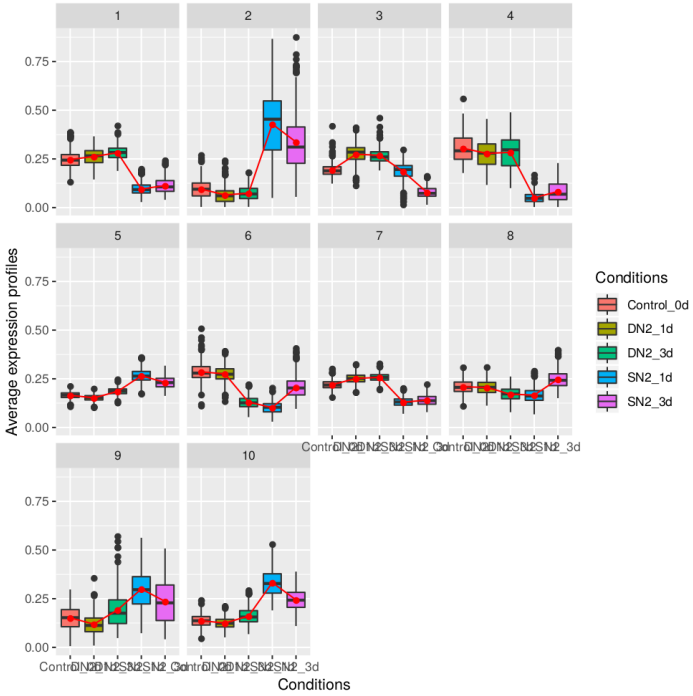
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

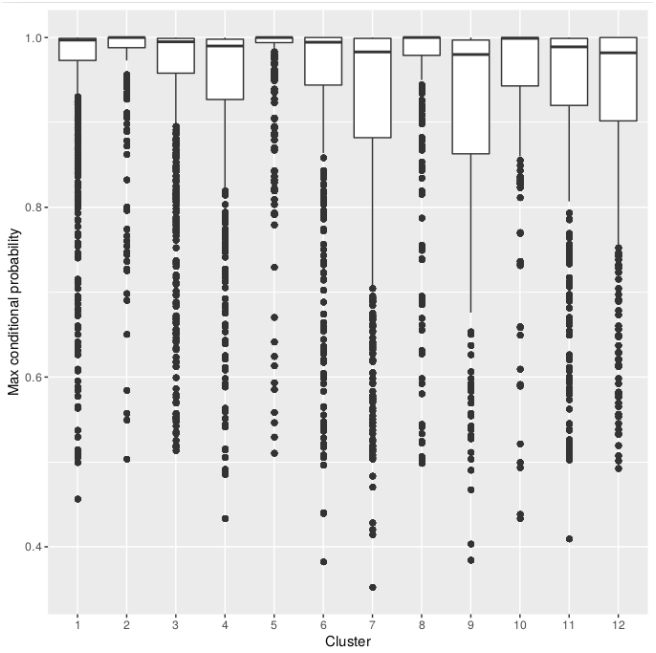
- Enrichment analysis



Outputs



Normalized expression profiles



Maximum conditional probability

Clusters table

Gene_ID	Clusters
Medtr0001s0660	2
Medtr0002s0020	1
Medtr0002s1060	10
Medtr0003s0560	3
Medtr0003s0580	4
Medtr0003s0590	3

Co-expression networks

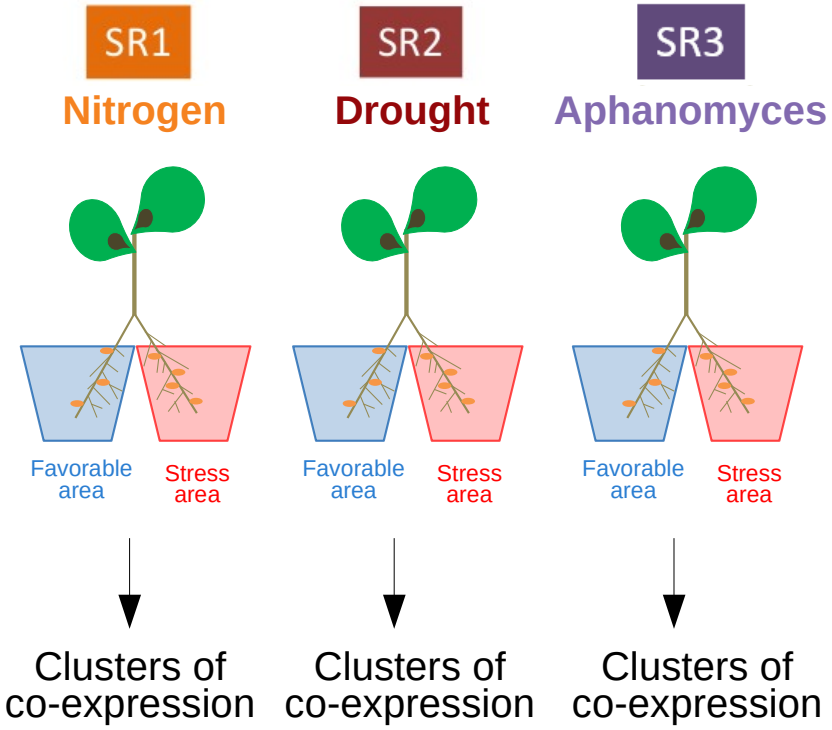
AUTOMATED RNA-seq ANALYSIS WORKFLOW

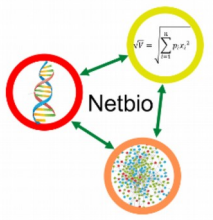
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method





Co-expression networks

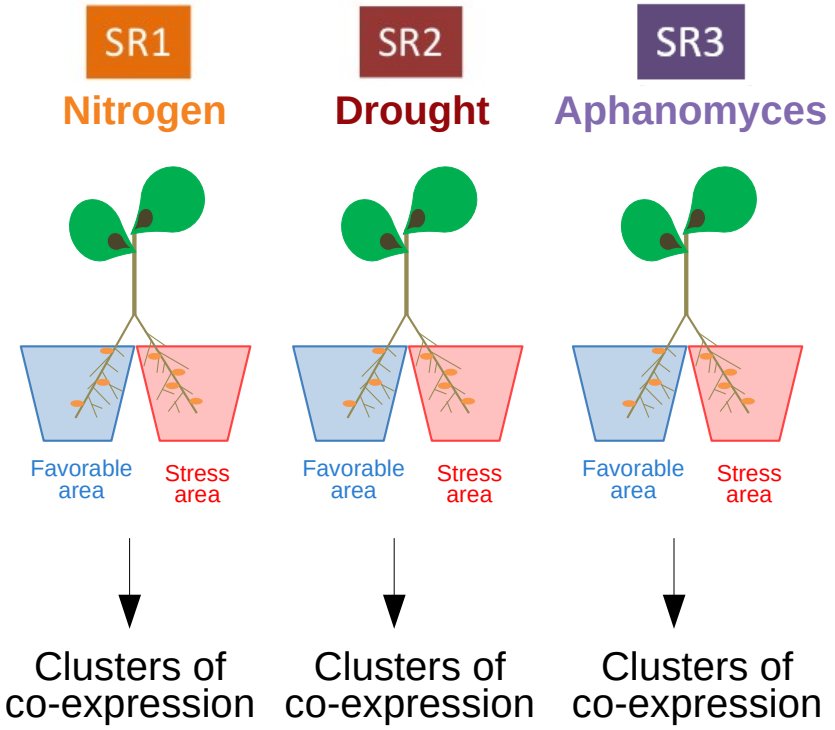
AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

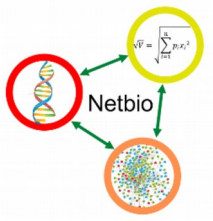
- Enrichment analysis

Method



Input

	SR1	SR2	SR3
Gene 1	1	2	8
Gene 2	1	2	9
Gene 3	1	1	10
Gene G	2	10	7



Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

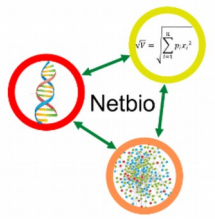
	SR1	SR2	SR3
Gene 1	1	2	8
Gene 2	1	2	9
Gene 3	1	1	10
Gene G	2	10	7

Table of occurrence

Calculation of score {0,1,2,3}

Co-expression networks

Tool



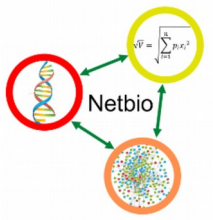
Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis



Method

	SR1	SR2	SR3
Gene 1	**1**	**2**	8
Gene 2	**1**	**2**	9
Gene 3	1	1	10
Gene G	2	10	7

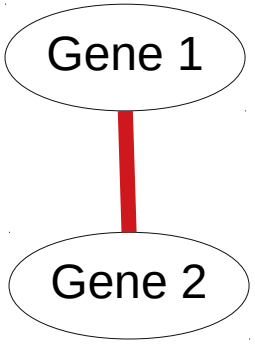
Table of occurrence

Calculation of score {0,1,2,3}

Ex:
(Gene 1 – Gene 2) = **2**

Tool

Co-expression networks



Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

	SR1	SR2	SR3
Gene 1	**1**	2	8
Gene 2	1	2	9
Gene 3	**1**	1	10
Gene G	2	10	7

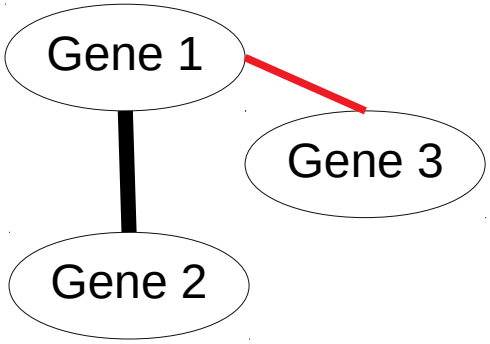
Table of occurrence

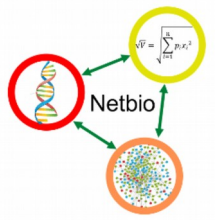
Calculation of score {0,1,2,3}

Ex:
(Gene 1 – Gene 2) = 2
(Gene 1 – Gene 3) = **1**

Tool

Co-expression networks





Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

	SR1	SR2	SR3
Gene 1	1	2	8
Gene 2	**1**	2	9
Gene 3	**1**	1	10
Gene G	2	10	7

Table of occurrence

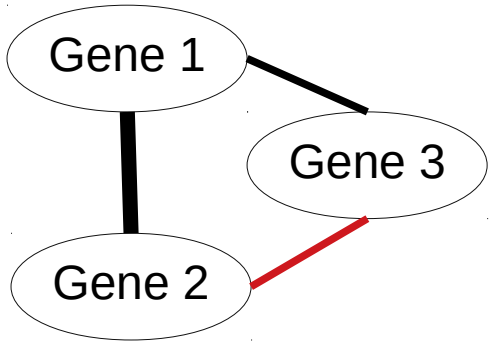
Calculation of score {0,1,2,3}

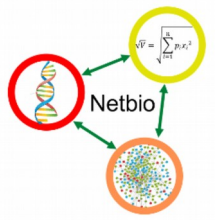
Ex:
(Gene 1 – Gene 2) = 2
(Gene 1 – Gene 3) = 1
(Gene 2 – Gene 3) = **1**

Tool

Cytoscape

Co-expression networks





Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

	SR1	SR2	SR3
Gene 1	1	2	8
Gene 2	1	2	9
Gene 3	1	1	10
Gene G	2	10	7

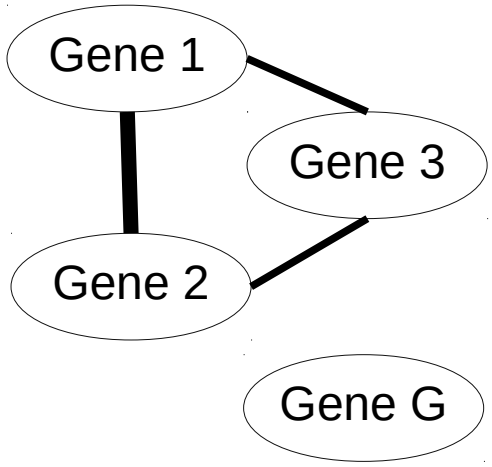
Table of occurrence

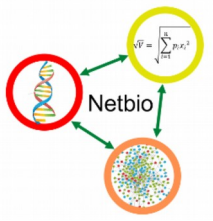
Calculation of score {0,1,2,3}

Ex:
(Gene 1 – Gene 2) = 2
(Gene 1 – Gene 3) = 1
(Gene 2 – Gene 3) = 1
(Gene 1 – Gene G) = 0
(Gene 2 – Gene G) = 0
(Gene 3 – Gene G) = 0

Tool

Co-expression networks





Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

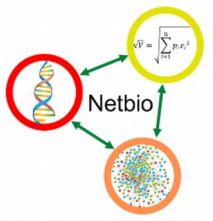
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Table of occurrence

Source_Gene	Edge_score	Target_Gene	SR1	SR2	SR3	Cluster_group
MtrunA17Chr1g0148241	2	MtrunA17Chr1g0209251	N	3	1	N31
MtrunA17Chr1g0149811	2	MtrunA17Chr1g0209691	N	7	3	N73
MtrunA17Chr4g0071991	1	MtrunA17Chr7g0240451	N	8	N	N8N
MtrunA17Chr5g0398081	3	MtrunA17Chr5g0444331	1	1	2	112



Co-expression networks

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

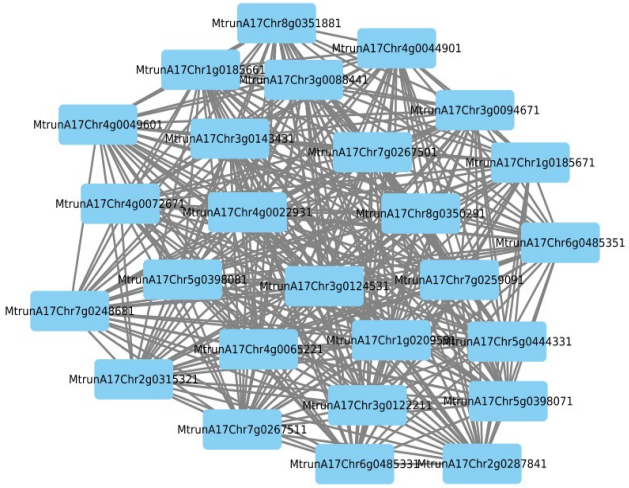
FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

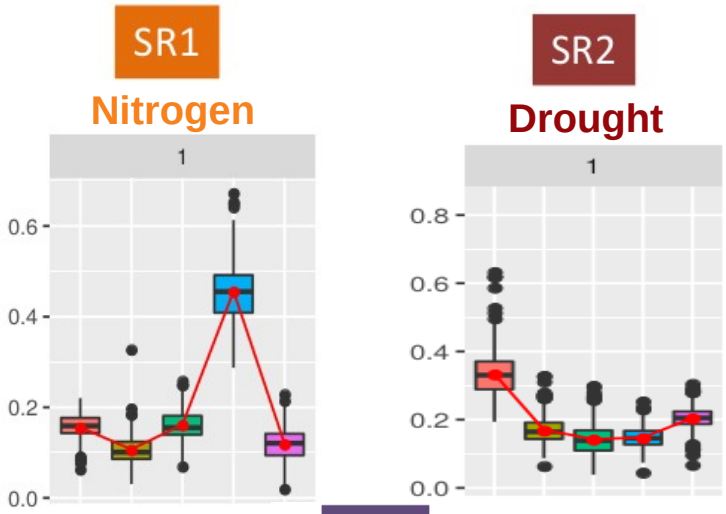
Table of occurrence

Source_Gene	Edge_score	Target_Gene	SR1	SR2	SR3	Cluster_group
MtrunA17Chr1g0148241	2	MtrunA17Chr1g0209251	N	3	1	N31
MtrunA17Chr1g0149811	2	MtrunA17Chr1g0209691	N	7	3	N73
MtrunA17Chr4g0071991	1	MtrunA17Chr7g0240451	N	8	N	N8N
MtrunA17Chr5g0398081	3	MtrunA17Chr5g0444331	1	1	2	112

Cluster_group = 112



Sub-Network of co-expression



Enrichment analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

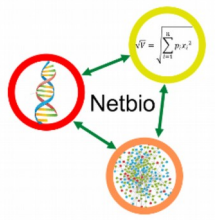
- Enrichment analysis

Method

Hypergeometric test

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

N : population size
K : number of success in the population
n : sample size
k : number of success in the sample



Enrichment analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

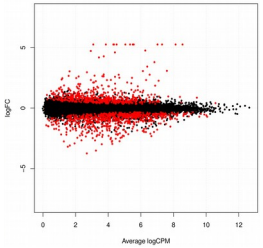
Hypergeometric test

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

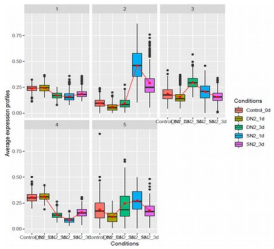
N : population size
K : number of success in the population
n : sample size
k : number of success in the sample

Inputs

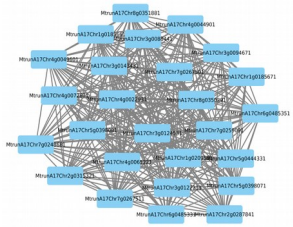
Differential expression analysis

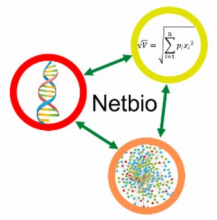


Co-expression analysis



Co-expression network





Enrichment analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis

Method

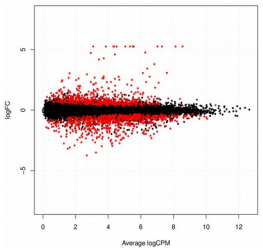
Hypergeometric test

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

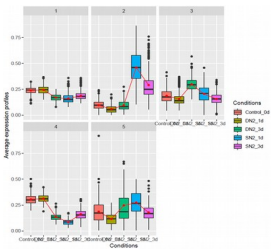
N : population size
K : number of success in the population
n : sample size
k : number of success in the sample

Inputs

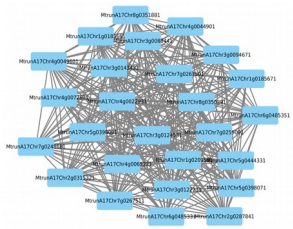
Differential expression analysis



Co-expression analysis



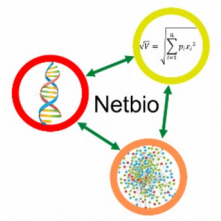
Co-expression network

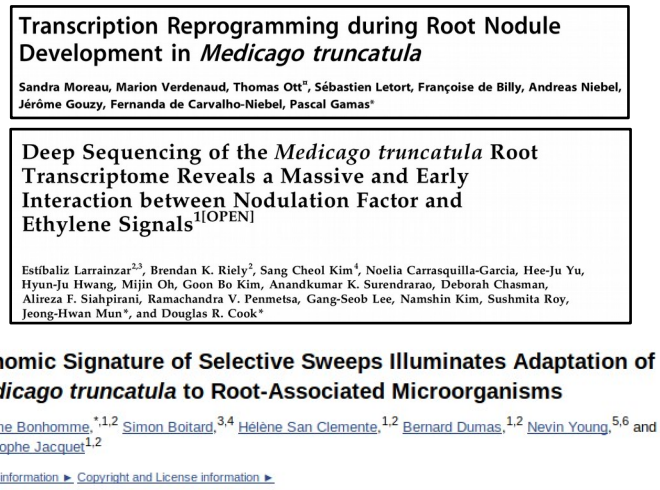


Ressources of annotation

Databases

Litterature





Transcription Reprogramming during Root Nodule Development in *Medicago truncatula*

Sandra Moreau, Marion Verdenaud, Thomas Ott, Sébastien Letort, Françoise de Billy, Andreas Niebel, Jérôme Gouzy, Fernanda de Carvalho-Niebel, Pascal Gamas*

Deep Sequencing of the *Medicago truncatula* Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals[1][OPEN]

Estíbaliz Larrainzar[2,3], Brendan K. Riely[2], Sang Cheol Kim[4], Noelia Carrasquilla-Garcia, Hee-Ju Yu, Hyun-Ju Hwang, Mijin Oh, Goon Bo Kim, Anandkumar K. Surendrarao, Deborah Chasman, Alireza F. Siahpirani, Ramachandra V. Penmetsa, Gang-Seob Lee, Namshin Kim, Sushmita Roy, Jeong-Hwan Mun*, and Douglas R. Cook*

Genomic Signature of Selective Sweeps Illuminates Adaptation of *Medicago truncatula* to Root-Associated Microorganisms

Maxime Bonhomme,[*,1,2] Simon Boitard,[3,4] Hélène San Clemente,[1,2] Bernard Dumas,[1,2] Nevin Young,[5,6] and Christophe Jacquet[1,2]

Author information ▸ Copyright and License information ▸

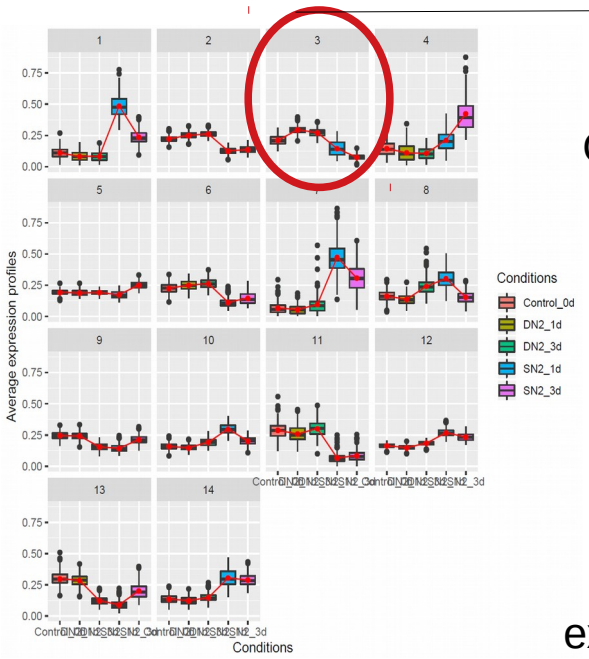
Enrichment analysis

AUTOMATED RNA-seq ANALYSIS WORKFLOW

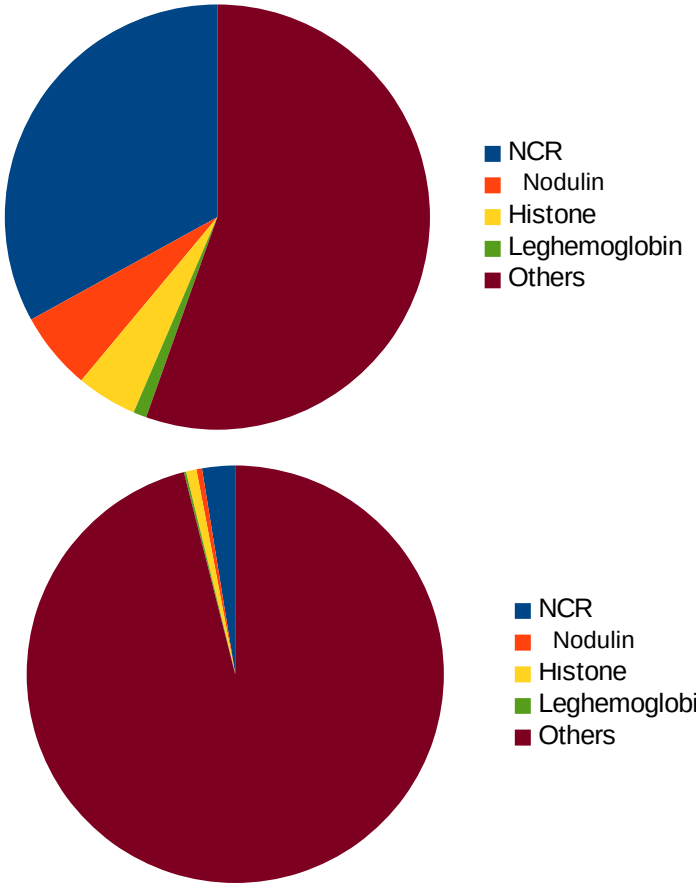
- Filtering
- Normalization
- Data quality control
- Differential expression analysis
- Co-expression analysis
- Co-expression Networks

FUNCTIONAL MODULES IDENTIFICATION

- Enrichment analysis



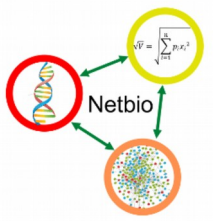
Cluster 3



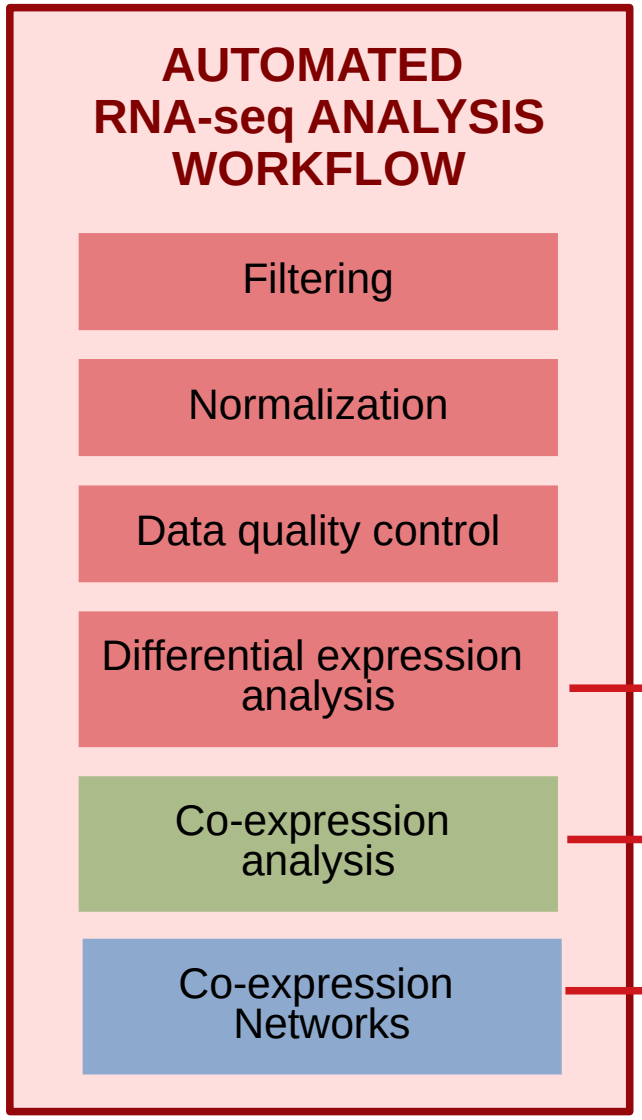
Total of expressed genes

P-value < 0,05

	Expected number of successes	FoldChange	Hypergeometric p-value	Conclusion
NCR	7.7547	12.9	4.1647^{e}-100	Over-enriched
NOD	1.3018	13.83	5.6779^{e}-19	Over-enriched
Histone	2.4905	5.62	6.3195^{e}-08	Over-enriched
Leghemoglobin	0.5094	5.89	0.0116	Over-enriched
Others	290.9433	1.73	2.0608	Under-enriched



Conclusion & Prospects



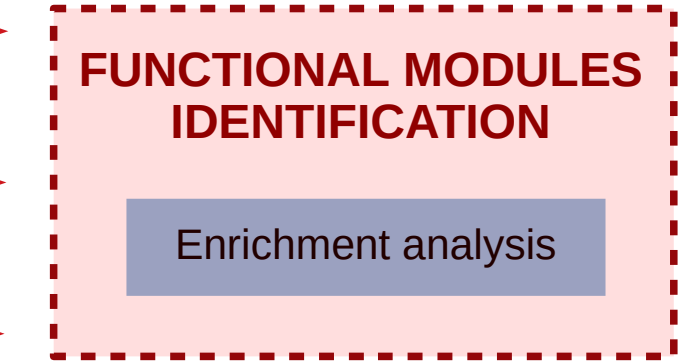
Relevant methods

Reproducible analysis

Comparable analysis

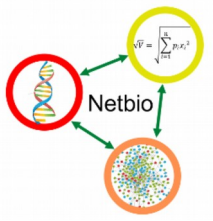
Compatible with all species

Relevant outputs for biologists (visualization, table, summary…)



Automation on going

Development on going



Acknowledgments

MASCE team

Brigitte Brunel
PR1, Montpellier SupAgro
Montpellier, France
Contact
Publications

Stefano Colella
CR1, Inra
Montpellier, France
Contact
Publications

Karine Heulin
IE, Inra
Montpellier, France
Contact

Marc Lepetit
DR2, Inra
Montpellier, France
Contact
Publications

Antoine Le Quéré
CR2, IRD
Rabat, Maroc
Contact
Publications

Marjorie Pervent
TR, Inra
Montpellier
Contact

Marc Tauzin
TR, Inra
Montpellier, France
Contact

Stephane BOVIN

Mathilde TANCELIN

Genomic networks team

Marie-Laure Martin-Magniette (DR, INRA)
Christine Paysant-Le-Roux (IE, INRA)
Véronique Brunaud (IR, INRA)
Nathalie Boudet (MdC, UEVE)
Cécile Guichard (IE, INRA)
Guillem Rigaill (CR, INRA)
Jean Philippe Tamby (IE, INRA)
Julien Rozière (IE, CDD)
Margot Correa (IE, CNRS)

Maxime BONHOMME

Thank you for your attention !