

DIANE

Dashboard for the **I**nference and **A**nalysis of **N**etworks from **E**xpression data

Océane Cassan - oceane.cassan@supagro.fr
Supervised by Antoine Martin and Sophie Lèbre



NETBIO

December 8, 2020



b&pmp
& Montpellier



IMAG
INSTITUT MONTPELLERAIN
ALEXANDRE BRÉTHÉRON

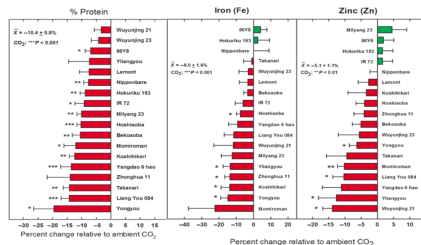


Thesis project

Statistical inference of the gene regulatory network in *Arabidopsis thaliana* under climate change

- Anthropic CO₂ emissions are expected to double until 2100
- Elevated CO₂ **depletes plants mineral status**
- Key **regulators** in this response are yet to be identified

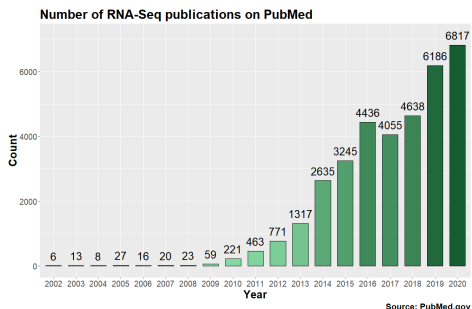
Genetic variation is found in mineral status response to elevated CO₂



Zhu 2018 Science Advances

Gene expression to understand regulatory mechanisms

- Gene expression reprogramming is studied to understand **development**, **adaptation** to environmental constraints in living organisms
- The rise of NGS techniques and **RNA-Seq** made available genome-wide transcripts quantification to researchers



Analyzing RNA-Seq data

To characterize new signalling pathways

- Which genes have their **expression significantly changed** by a perturbation?
- Can I distinguish **groups** of genes with similar behaviors?
- How to model **interactions and dependencies between genes**?

Analyzing RNA-Seq data

To characterize new signalling pathways

- Which genes have their **expression significantly changed** by a perturbation?
- Can I distinguish **groups** of genes with similar behaviors?
- How to model **interactions and dependencies between genes**?

Standard analysis pipelines require **programming** and **methodological** skills:

Bioinformatic pipeline

Quality control, fragments mapping, quantification

Analyzing RNA-Seq data

To characterize new signalling pathways

- Which genes have their **expression significantly changed** by a perturbation?
- Can I distinguish **groups** of genes with similar behaviors?
- How to model **interactions and dependencies between genes**?

Standard analysis pipelines require **programming** and **methodological** skills:

Bioinformatic pipeline

Quality control, fragments mapping, quantification

Statistical pipeline

Normalization, differential gene expression, ontology enrichment, clustering, co-expression or regulatory pathways reconstruction

Current user interfaces for facilitating RNA-seq analysis

Available satisfactory possibilities

- Normalization
- Exploration and visualization
- Differential Expression
- Gene Ontology Enrichment

See DEBrowser[Kucukural et al., 2019],
DEApp[Li and Andrade, 2017],
iGEAk[Choi and Ratner, 2019],
DEIVA[Harshbarger et al., 2017],
Shiny-Seq[Sundararajan et al., 2019],
IRIS-DEA[Monier et al., 2019], iDEP[Ge et al., 2018],
TCC-GUI[Su et al., 2019], ShinyBN [Chen et al., 2019],
or GeNeck [Zhang et al., 2019]

Current user interfaces for facilitating RNA-seq analysis

Available satisfactory possibilities

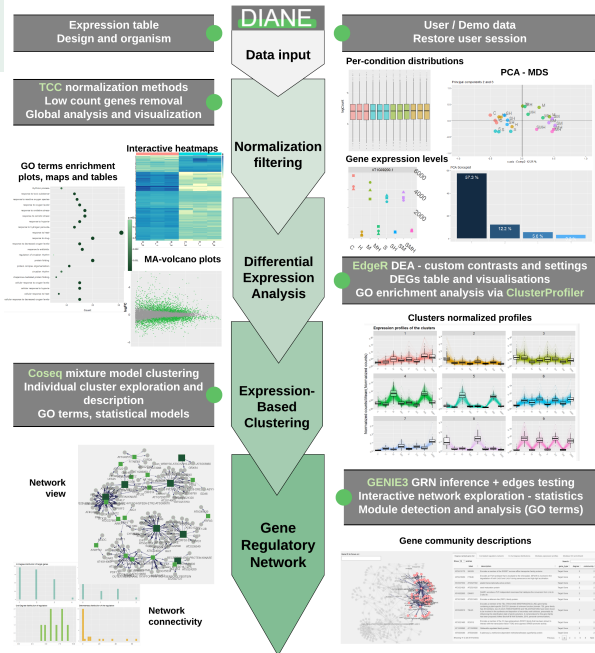
- Normalization
- Exploration and visualization
- Differential Expression
- Gene Ontology Enrichment

See DEBrowser[Kucukural et al., 2019], DEApp[Li and Andrade, 2017], iGEAk[Choi and Ratner, 2019], DEIVA[Harshbarger et al., 2017], Shiny-Seq[Sundararajan et al., 2019], IRIS-DEA[Monier et al., 2019], iDEP[Ge et al., 2018], TCC-GUI[Su et al., 2019], ShinyBN [Chen et al., 2019], or GeNeck [Zhang et al., 2019]

Absent or sub-optimal features

- Model based clustering
- Statistical models or Machine Learning for network inference
- Oriented toward regulation rather than co-expression
- Advanced exploration of gene clusters or inferred networks

Workflow



Multifactorial dataset of nutritional starvation and elevated CO₂

Complete 3 * 2 cross design :

- **CO₂** (400ppm - 900ppm)
- **Nitrate** (low - **high**)
- **Iron** (Starvation - **sufficiency**)

8 conditions with triplicates : 24 root transcriptomes



Arabidopsis thaliana

Normalization and low counts removal

Normalization

- EdgeR TMM strategy [Robinson and Oshlack, 2010]
- DESeq2 Median of ratios [Love et al., 2014]
- Prior removal of DEGs via TCC [Sun et al., 2013]

Filtering

Adjustable threshold for counts sum across samples

DIANE

▼ Dashboard
 Data import
 Normalization
 Exploratory analysis
 Differential Expression
 Expression based clustering
 Run a clustering
 Explore clusters
 Gene regulatory network
 Network inference
 Network analysis
 Ready to upload datasets

Data filtering and normalisation

Settings

Normalization

Prior removal of differentially expressed genes: OFF

Normalization method: tmm median none

NORMALIZE

Normalization needed

Low counts filtering

Minimum gene count sum across conditions: **FILTER**

Normalization needed

DOWNLOAD NORMALIZED COUNTS AS .CSV

DOWNLOAD NORMALIZED COUNTS AS .RDATA

GENERATE HTML REPORT

Per-condition expression distributions

logCount

sample

condition cNF1 cNF2 cNF3 cNF4 cNF5 cNF6 cNF7 cNF8 cNF9 cNF10 cNF11 cNF12 cNF13 cNF14 cNF15 cNF16 cNF17 cNF18 cNF19 cNF20 cNF21 cNF22 cNF23 cNF24

Exploratory transcriptome analysis

- Genes normalized expression levels
- PCA, Mutli-Dimensional Scaling plot



Differential Expression Analysis

Tests performed by EdgeR (Negative Binomial models) [McCarthy et al., 2012]

DIANE

- Context
- Data input
- Normalisation
- Exploratory analysis
 - Differential Expression
 - Expression based clustering
 - Run a clustering
 - Explore clusters
- Gene Regulatory Network
 - Network inference
 - Network analysis
- Ready to upload datasets

Differential expression analysis

Settings

Estimation of dispersion

Conditions to compare for differential analysis:

Reference	Perturbation
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input checked="" type="checkbox"/> cdf	<input type="checkbox"/> cdf
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input type="checkbox"/> CDF	<input checked="" type="checkbox"/> Cdf
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF
<input type="checkbox"/> CDF	<input type="checkbox"/> CDF

Adjusted p-value (FDR):

Minimum absolute log Fold Change:

DETECT DIFFERENTIALLY EXPRESSED GENES

See plots and tables for more details

303 up-regulated GENES
641 down-regulated GENES

Results table | MA - Volcano plots | Heatmap | Gene Ontology enrichment | Compare genes lists (Wires) | Results

MA

M-A plot

logFC

logCPM

DEG

Differential Expression Analysis

Over-representation tests are performed by ClusterProfiler to detect significantly enriched GO terms [Yu et al., 2012]

DIANE
Results table MA - Volcano plots Heatmap Gene ontology enrichment Compare genes lists (View)

Differential expression analysis

Settings

Estimation of dispersion

Conditions to compare for differential analysis:

Reference	Perturbation
<input type="checkbox"/> dNF	<input type="checkbox"/> dNF
<input checked="" type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> CNF	<input type="checkbox"/> CNF
<input type="checkbox"/> cNF	<input checked="" type="checkbox"/> cNF
<input type="checkbox"/> dNF	<input type="checkbox"/> CNF
<input type="checkbox"/> CNF	<input type="checkbox"/> CNF
<input type="checkbox"/> cNF	<input type="checkbox"/> CNF
<input type="checkbox"/> CNF	<input type="checkbox"/> CNF

Adjusted pvalue (FDR)

Minimum absolute log Fold Change:

DETECT DIFFERENTIALLY EXPRESSED GENES

[View](#)
See plots and tables for more details

505 ■
up regulated
GENES

641 ■
down regulated
GENES

↓ DOWNLOAD RESULT TABLE AS .CSV

↓ GENERATE HTML REPORT

START GO ENRICHMENT ANALYSIS

Dot plot
Enrichment map
Data table

Dot plot
Cellular component
Molecular function

Top number of GO terms to plot:

Enriched ontologies and their gene count

p.adjust

Differential Expression Analysis

Venn diagrams to compare lists of differentially expressed genes

DIANE

- ▼ Content
- ▣ Data import
- ▣ Normalisation
- ▣ Exploratory analysis
- ▣ Differential Expression
- Expression based clustering
- Run a clustering
- Explore clusters
- ▼ Gene Regulatory network
- Network inference
- Network analysis
- ▣ Ready to upload datasets

Differential expression analysis

Settings

Estimation of dispersion

Conditions to compare for differential analysis:

Reference	Perturbation
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input checked="" type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input type="checkbox"/> cNF
<input type="checkbox"/> cNF	<input checked="" type="checkbox"/> cNF

Adjusted p-value (fdr)

Minimum absolute log Fold change:

DETECT DIFFERENTIALLY EXPRESSED GENES

See plots and tables for more details

↑ 1070
 up-regulated
 GENES

↓ 104
 down-regulated
 GENES

DOWNLOAD RESULT TABLE AS .CSV

GENERATE HTML REPORT

Results table
MA - Volcano plots
Heatmap
Gene Ontology enrichment
Compare genes lists (Venn)
Results

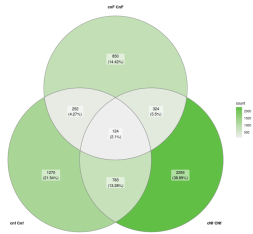
Once more than one differential expression analysis were performed, you can visualise and compare the different genes lists in a Venn diagram.

Please select between 2 and 4 lists of genes to show in the Venn diagram:

<input checked="" type="checkbox"/> cNF cNF	<input checked="" type="checkbox"/> cNF cNF	<input checked="" type="checkbox"/> cNF cNF
---	---	---

Differentially expressed genes to compare:

All
 Up-regulated
 Down-regulated



DOWNLOAD THE INTERSECTION OF ALL SETS (CENTRAL PART IN THE VENN DIAGRAM)

Genes specific to a comparison:

DOWNLOAD GENES SPECIFIC TO THE cNF cNF LIST

Co-expression clustering

Poisson or Gaussian Mixture Models allow gene profiles clustering via Coseq
[Rau et al., 2015, Rau and Maugis-Rabusseau, 2018]

DIANE

- Context
- Data import
- Normalisation
- Exploratory analysis
- Differential Expression
- Expression based clustering
- Run a clustering**
 - Explore clusters
 - Gene Regulatory Network
 - Network inference
 - Network analysis
 - Ready to upload datasets

Expression based clustering

Settings

Mixture models clustering

Input genes for clustering:
 chr1 chr1 1550 genes

Conditions to perform clustering on:

- GSE
- GPM
- GMP
- GCP
- GSI
- GTF
- GCI
- GCH

Min number of clusters:

Max number of clusters:

Network Model to use: Normal

Transformation prior to Gaussian Mixtures: arcsin

LAUNCH CLUSTERING

Done

See Coseq summary tab for more details

[GENERATE HTML REPORT](#)

Cluster profiles Clustering quality Coseq summary

Cluster	Clustering quality
1	0.95
2	0.55
3	0.55
4	0.15

Cluster	GSE (Green)	GPM (Purple)	Total
1	~20	~230	~250
2	~20	~130	~150
3	~20	~130	~150
4	~20	~130	~150

Clustering quality descriptors

The coseq package tests a range of different clusters in order to give the best fit to the data.

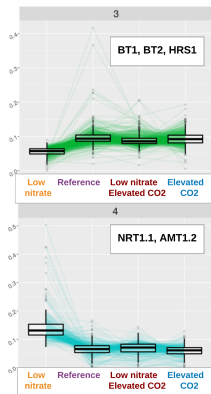
For each number of clusters, the ICL (Integrated Completed Likelihood) is computed. It combines two elements:

- The global **likelihood** of the clustering. It quantifies how accurate the clustering seems, regarding the posterior probability of each element to belong to its predicted cluster. It can be computed using the Poisson probability densities resulting from the proposed clustering, for all the genes.
- The **number of clusters**. As the likelihood tends to grow monotonously with the number of clusters, resulting in a very big number of groups, that would not be very informative for the user. Thus, the ICL penalises the clustering quality criteria with the number of clusters.

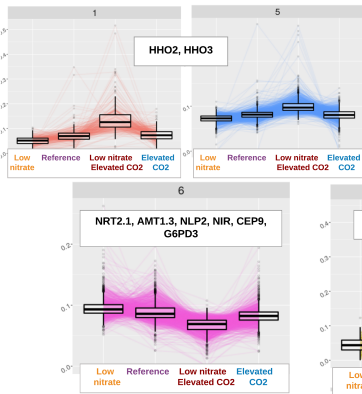
This is why the maximal value of ICL can be interpreted as an approximation of the ideal number of clusters.

Gene behaviors under CO₂ and nitrate perturbations

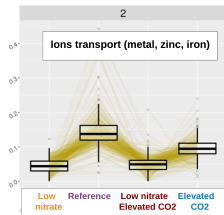
Elevated CO₂ cancels the regulation of nitrate related genes



Elevated CO₂ reverses the effect of low nitrate



Some genes stay downregulated under low nitrate and elevated CO₂

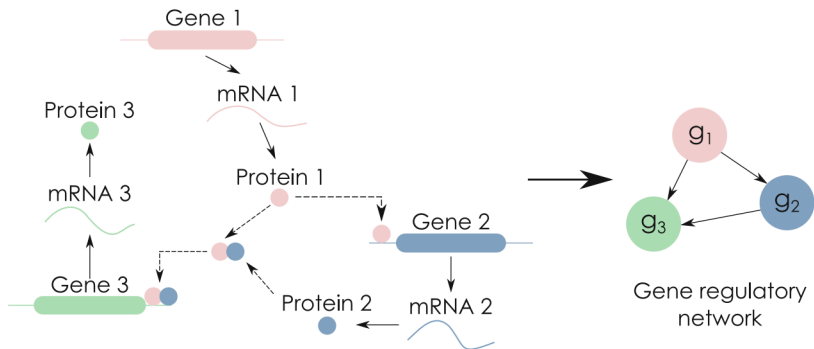


How does elevated CO₂ alter the response to nitrate starvation?

Either lessens, cancels or inverses the regulation observed in low nitrate only

Gene Regulatory Network Inference

- How to capture transcriptional dependencies and causality between genes ?



[Sanguinetti and Huynh-Thu, 2019]

Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

- **Non-linear** approach, captures regulators **cooperation** for gene regulation

Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

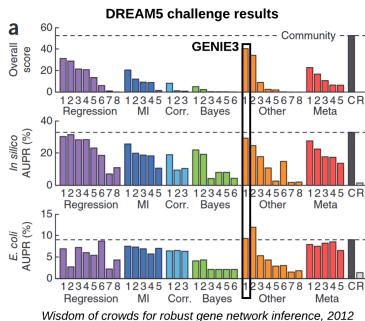
- **Non-linear** approach, captures regulators **cooperation** for gene regulation
- **Non-parametric** procedure, with very few parameters and no modelling assumptions

Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

- **Non-linear** approach, captures regulators **cooperation** for gene regulation
- **Non-parametric** procedure, with very few parameters and no modelling assumptions
- Easily **scalable** to large datasets

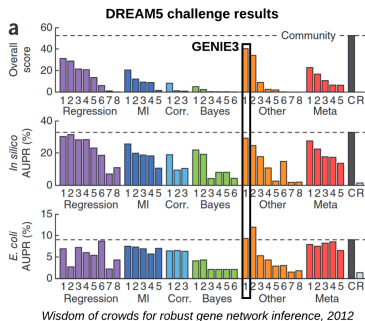
Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

- **Non-linear** approach, captures regulators **cooperation** for gene regulation
- **Non-parametric** procedure, with very few parameters and no modelling assumptions
- Easily **scalable** to large datasets
- Best performer on **DREAM4** and **DREAM5** challenges
[Greenfield et al., 2010, Marbach et al., 2016]



Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

- **Non-linear** approach, captures regulators **cooperation** for gene regulation
- **Non-parametric** procedure, with very few parameters and no modelling assumptions
- Easily **scalable** to large datasets
- Best performer on **DREAM4** and **DREAM5** challenges
[Greenfield et al., 2010, Marbach et al., 2016]



- **Oriented edges** from regulators to other genes

Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

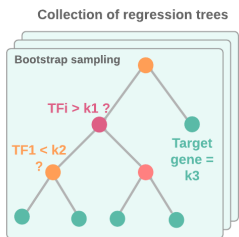
Ranking the regulators according to their **relevance for predicting** the other genes expression

	Response	Predictors		
	Target gene	TF 1	...	TF N
Experimental conditions				

Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

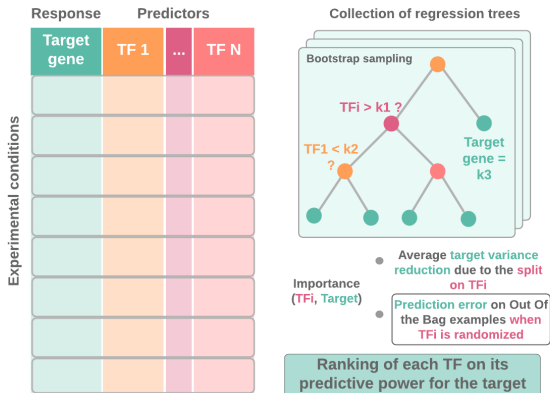
Ranking the regulators according to their **relevance for predicting** the other genes expression

	Response	Predictors		
	Target gene	TF 1	...	TF N
Experimental conditions				



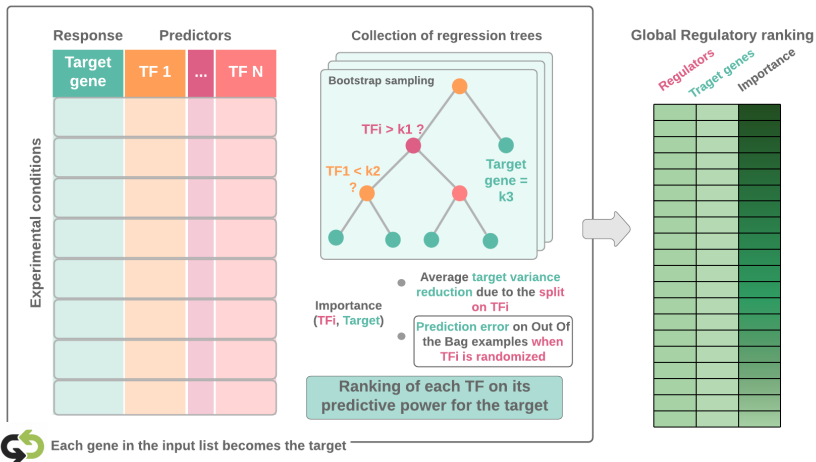
Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

Ranking the regulators according to their **relevance** for **predicting** the other genes expression



Random Forests inference with GENIE3 [Huynh-Thu et al., 2010]

Ranking the regulators according to their **relevance for predicting** the other genes expression



DIANE's interface for network inference

Network inference

Inference Settings

GENIE3 Gene regulatory network inference

Input genes for network inference:

Conditions used to infer network edges:

- OFF
- ONF
- ONP
- ONP
- ONF
- OFF
- ONF
- ONP

Use only specific clusters (from previous clustering job):

- 5
- 3
- 7
- 8
- 9
- 2
- 4
- 1

2102 REGULATORS PROVIDED FOR ARABIDOPSIS THALIANA

Or choose your own:

Upload custom CSV/TXT regulators list

Browse:

1550 INPUT GENES 173 REGULATORS AMONG THE INPUT GENES

Recommended: group regulators contained over (%)

Number of cores for multithreaded inference:

Number of trees for GENIE3 Random Forests:

Importance metric in random forests

Thresholding settings

Inference successfully completed
YOU CAN NOW PROCEED TO THRESHOLDING

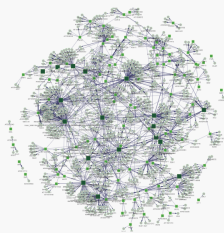
Without thresholding, we would obtain a fully connected weighted graph from GENIE3, with far too many links to be interpretable. In order build a meaningful network, this weighted adjacency matrix between regulators and targets has to be sparsified, and we have to determine the k higher regulatory weights that we consider significant.

As biological networks are known for their pronounced sparsity, considering all possible regulator-target pairs would be way off the intended. Studies found that the typical values of density (ratio between the number of edges, and the number of edges of the complete graph) in state of the art biological networks lie approximately between 0.1 and 0.001. We therefore propose to create a first graph, topologically consistent with biological network standards, which will be further refined by statistical testing. During the testing phase, the null distributions of regulatory weights are estimated by the permutations, thus providing an empirical p-value for the observed importance.

Network's connectivity density:

Edges: statistical testing

Your network is ready
YOU CAN EXPLORE IT IN THE NEXT TAB



Network exploration

- Gene connectivity ranking
- Individual nodes information, targets and regulators
- In and out degree distributions

DIANE

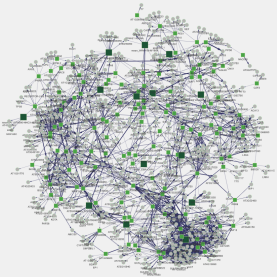
Network analysis

Nodes color: Gene type:

Cluster to explore: All 15 8 1 3 10 7 6 2 4 5 12 9 11 14

773 GENES 135 REGULATORS 383 EDGES

Gene ID to focus on:



Degree-ranked gene list: Correlated regulators network In-Out degree distributions Modules expression profiles Modules GO enrichment

Show entries

label	description	gene_type	degree	connectivity
AT4022680	HYB85	Regulator	71	2
AT3023350	HYB15	Regulator	64	2
AT3023730	EDP3	Regulator	58	2
AT3023650	WDX11	Regulator	52	2
AT3019580	ZF2	Regulator	48	2
AT4011880	AGL14	Regulator	47	2
HYB1_AT2021190-AT3046630	HYB1_WIKKY9-AT3046630	Grouped Regulators	42	2
AT5043175	AT5043175	Regulator	40	2
AT5067060	HEC1	Regulator	39	2
AT1072650	TR616	Regulator	33	7
AT1058100	TCF8	Regulator	32	2
AT4009390	AT4009390	Regulator	25	7
AT1057560	HYB50	Regulator	25	3
AT5066700	H6G3	Regulator	25	2
AT3022830	ZFN1	Regulator	25	3

Network exploration

- Gene connectivity ranking
- Individual nodes information, targets and regulators
- In and out degree distributions

DIANE

- Context
- Data report
- Network analysis
- Exploratory analysis
- Differential Expression
- Expression based clustering
- Gene co-expression
- Pathway analysis
- Gene Regulatory Network
- Network analysis
- Network analysis
- Ready to upload datasets

Network analysis

Subst. index: Gene type:

Cluster to explore:

Gene ID to focus on:

Gene ID to exclude:

Network visualization

Gene description

AGG | HMMG_AT2021900-AT3056490-AT3023200-AT3819580-AT1043175

Accession name: 1 mean_WBQY55-WBQY70-WBQY15-272-AT1043175

Description:

- AT2021900: member of WRKY Transcription Factor
- AT3056490: Member of WRKY Transcription Factor
- AT3023200: Member of the E2F53 factor gene family.
- AT3819580: Encodes zinc finger protein. mRNA levels are upregulated in response to ABA, high salt, and mild dehydration. The protein is localized to the nucleus and acts as a transcriptional repressor.
- AT1043175: basic helix-loop-helix (HLH) DNA-binding superfamily protein

Normalized expression levels

No regulators found

Targets:

Show: 10 entries

label	description	gene_type	degree	necessity
AT4002020	DWRN Encodes a polycomb group protein. Forms part of a larger protein complex that can include WDR5 (COH3/MEZ1/2), WDR4 (CORPUS2/MEZ1/2) and other polycomb group proteins. FERTILIZATION INDEPENDENT ENDODERMIS (FIE) and CORPUS LEAFY (CLF). This complex has a role in establishing FLC (FLOWERING LOCUS C) repression during vernalization. Performs a partially redundant role to FLC in controlling seed initiation by helping to suppress central cell nucleoskeletons proliferation within the FC.	Regulator	25	4
AT1017920	Single Myb Helixone (SH) gene family member. Contains terminal acidic-SAM1	Regulator	5	10

Network statistics

AG	AG	AG	AG	AG
36	38	1030	9260	16
REGULATORS	16	REGULATORS	1724	EDGES

In-out degree distributions: Modulus expression profiles: Modulus CC enrichment:

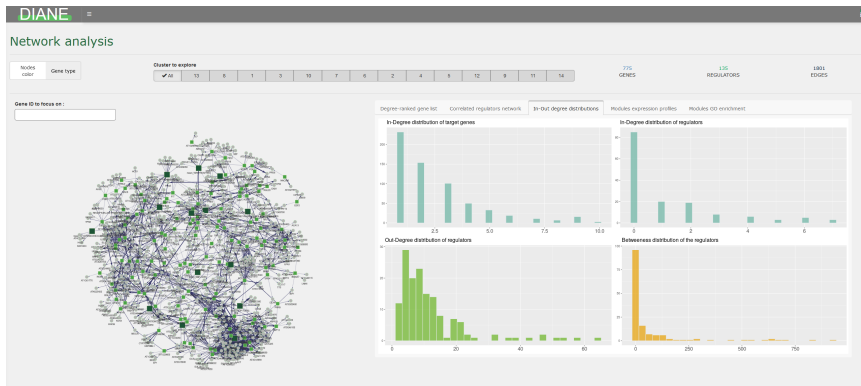
Search:

Description	gene_type	degree	necessity
Grouped Regulators	10	10	
Grouped Regulators	30	10	
encodes a MYB/CC1-related transcription gene family member. It is an active site in its transcription. Proteins in this family form a sequence of eight residues (TLPYRRE) downstream the transcription start site box.	Regulator	12	10
involved in gene silencing. Locus-specific regulator of 24S-RNA expression, works together with L2.0Y1-4 as the master pattern of assembly of P4-0-independent gene silencing.	Regulator	36	10
high DNA binding (DBP) gene family member. Contains terminal acidic-SAM1 domain.	Regulator	5	10
W-activating enzyme 17	Target Gene	5	10
involved in regulation of iron deficiency response genes, overexpression results in hyperaccumulation of Fe and Mn.	Target Gene	4	10
involved in regulation of iron deficiency response genes.	Target Gene	4	10
encodes glutathione transferase belonging to the tau class of the enzyme according to Ingole et al. (1995). Inhibition of the environmental pollutant 2,4-dinitrophenol. Antidotes plant over-expressing AtGT2.00 are more resistant to TNT, reduced their TNT toxic state and in basal media, and had reduced levels of glutathione when treated in the presence of TNT.	Target Gene	4	10
encodes a major leaf herbivore	Target Gene	4	10

Download edges as CSV table

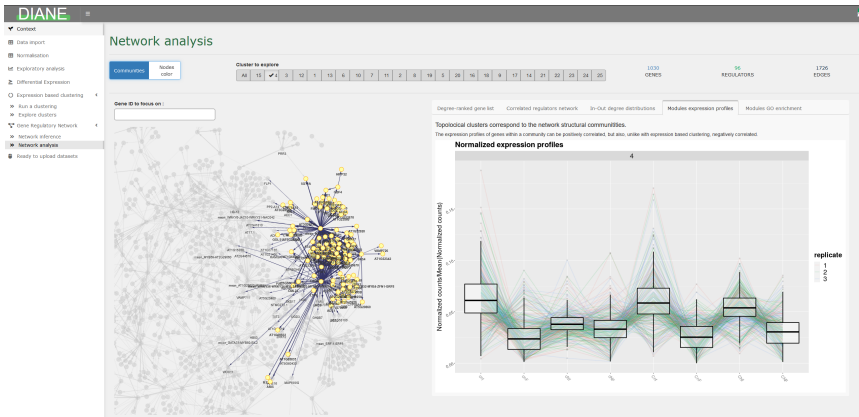
Network exploration

- Gene connectivity ranking
- Individual nodes information, targets and regulators
- **In and out degree distributions**



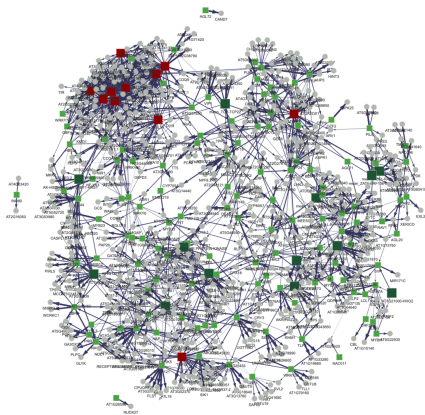
Communities discovery

- Louvain community detection [Blondel et al., 2008]
- Module expression profiles
- Module GO enrichment analysis



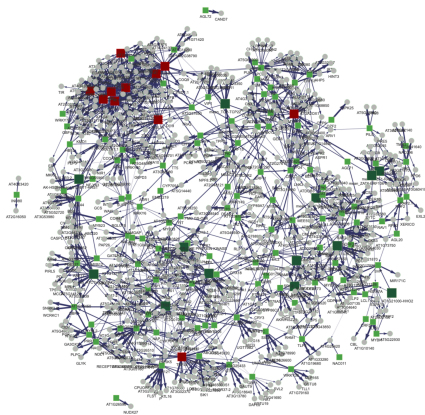
GRN under to elevated CO2 and low nitrate

1550 **differentially expressed genes** by elevated CO2 under low nitrate, 12 experimental samples



GRN under to elevated CO2 and low nitrate

1550 **differentially expressed genes** by elevated CO2 under low nitrate, 12 experimental samples

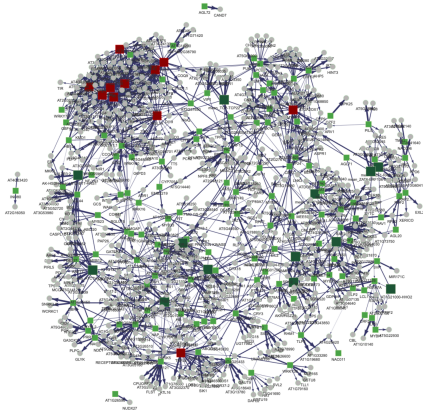


Validation of known genes interactions

- Confirmed associations between nitrate assimilation genes (BT1-BT2-HRS1, NRT2.1-NAR2.1...)

GRN under to elevated CO2 and low nitrate

1550 **differentially expressed genes** by elevated CO2 under low nitrate, 12 experimental samples



Validation of known genes interactions

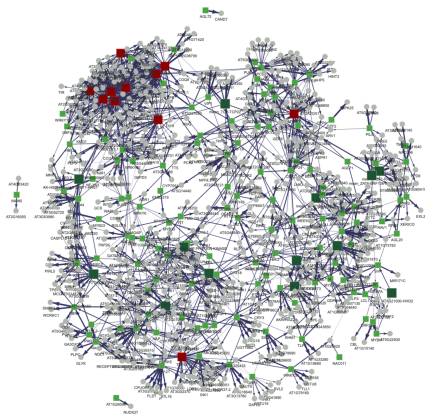
- Confirmed associations between nitrate assimilation genes (BT1-BT2-HRS1, NRT2.1-NAR2.1...)

Candidate genes discovery

- The most connected genes are potential **key regulators** of the response
- Mutant plants are programmed for **root system** and **shoots mineral status** phenotyping

GRN under to elevated CO2 and low nitrate

1550 **differentially expressed genes** by elevated CO2 under low nitrate, 12 experimental samples



Validation of known genes interactions

- Confirmed associations between nitrate assimilation genes (BT1-BT2-HRS1, NRT2.1-NAR2.1...)

Candidate genes discovery

- The most connected genes are potential **key regulators** of the response
- Mutant plants are programmed for **root system** and **shoots mineral status** phenotyping

Promising leads to explain the depletion of plants mineral status under elevated CO2

Conclusion

Highlights

- **Accessible** online or local user **interface**
- **Interactive**, efficient and **reproducible** analyses
- Strong methodological frameworks with parameterization help for clustering and network inference
- Annotations for **model organisms**, or **custom** files upload for other organism

Conclusion

Highlights

- **Accessible** online or local user **interface**
- **Interactive**, efficient and **reproducible** analyses
- Strong methodological frameworks with parameterization help for clustering and network inference
- Annotations for **model organisms**, or **custom** files upload for other organism

Perspectives

- Adapt inference methodology to **time series** RNA-Seq
- Integrate external databases or other -omics data (Chip-Seq, ATAC-Seq...)

Conclusion



Project links

Online version : <https://diane.bpmp.inrae.fr>

GitHub : <https://github.com/OceaneCsn/DIANE>

Documentation : <https://oceanecsn.github.io/DIANE/>

Conclusion



Project links

Online version : <https://diane.bpmp.inrae.fr>

GitHub : <https://github.com/OceaneCsn/DIANE>

Documentation : <https://oceanecsn.github.io/DIANE/>



Merci!

References I

- ▶ Archer, E. (2020).
rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics.
R package version 2.1.81.
- ▶ Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008).
Fast unfolding of communities in large networks.
Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008.
- ▶ Chen, J., Zhang, R., Dong, X., Lin, L., Zhu, Y., He, J., Christiani, D. C., Wei, Y., and Chen, F. (2019).
shinybn: an online application for interactive bayesian network inference and visualization.
BMC bioinformatics, 20(1):711.
- ▶ Choi, K. and Ratner, N. (2019).
IGEAK: An interactive gene expression analysis kit for seamless workflow using the R/shiny platform.
BMC Genomics, 20(1):177.
- ▶ Ge, S. X., Son, E. W., and Yao, R. (2018).
iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data.
BMC Bioinformatics, 19(1):1–24.

References II

- ▶ Greenfield, A., Madar, A., Ostrer, H., and Bonneau, R. (2010).
DREAM4: Combining genetic and dynamic information to identify biological networks and Dynamical Models.
PLoS ONE, 5(10).
- ▶ Harshbarger, J., Kratz, A., and Carninci, P. (2017).
DEIVA: A web application for interactive visual analysis of differential gene expression profiles.
BMC Genomics, 18(1):47.
- ▶ Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010).
Inferring Regulatory Networks from Expression Data Using Tree-Based Methods.
PLoS ONE, 5(9):e12776.
- ▶ Kucukural, A., Yukselen, O., Ozata, D. M., Moore, M. J., and Garber, M. (2019).
DEBrowser: Interactive differential expression analysis and visualization tool for count data
06 Biological Sciences 0604 Genetics 08 Information and Computing Sciences 0806
Information Systems.
BMC Genomics, 20(1):6.

References III

- ▶ Li, Y. and Andrade, J. (2017).
DEApp: An interactive web interface for differential expression analysis of next generation sequence data.
Source Code for Biology and Medicine, 12(1):10–13.
- ▶ Love, M. I., Huber, W., and Anders, S. (2014).
Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.
Genome Biology, 15(12):550.
- ▶ Marbach, D., Costello, J. C., Küffner, R., Vega, N., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., Stolovitzky, G., and Performed, G. S. M. (2016).
Wisdom of crowds for robust gene network inference the DREAM5 Consortium HHS Public Access.
Nat Methods, 9(8):796–804.
- ▶ McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012).
Differential expression analysis of multifactor rna-seq experiments with respect to biological variation.
Nucleic Acids Research, 40(10):4288–4297.

References IV

- ▶ Monier, B., McDermaid, A., Wang, C., Zhao, J., Miller, A., Fennell, A., and Ma, Q. (2019). IRIS-EDA: An integrated RNA-seq interpretation system for gene expression data analysis. *PLoS Computational Biology*, 15(2).
- ▶ Rau, A. and Maugis-Rabusseau, C. (2018). Transformation and model choice for RNA-seq co-expression analysis. *Briefings in Bioinformatics*, 19(3):425–436.
- ▶ Rau, A., Maugis-Rabusseau, C., Martin-Magniette, M.-L., and Celeux, G. (2015). Co-expression analysis of high-throughput transcriptome sequencing data with poisson mixture models. *Bioinformatics*, 31(9):1420–1427.
- ▶ Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3):R25.
- ▶ Sanguinetti, G. and Huynh-Thu, V. A. (2019). Gene regulatory networks.

References V

- ▶ Su, W., Sun, J., Shimizu, K., and Kadota, K. (2019).
TCC-GUI: A Shiny-based application for differential expression analysis of RNA-Seq count data.
BMC Research Notes, 12(1):133.
- ▶ Sun, J., Nishiyama, T., Shimizu, K., and Kadota, K. (2013).
TCC: An R package for comparing tag count data with robust normalization strategies.
BMC Bioinformatics, 14(1):219.
- ▶ Sundararajan, Z., Knoll, R., Hombach, P., Becker, M., Schultze, J. L., and Ulas, T. (2019).
Shiny-Seq: advanced guided transcriptome analysis.
BMC research notes, 12(1):432.
- ▶ Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012).
clusterprofiler: an r package for comparing biological themes among gene clusters.
Omics: a journal of integrative biology, 16(5):284–287.
- ▶ Zhang, M., Li, Q., Yu, D., Yao, B., Guo, W., Xie, Y., and Xiao, G. (2019).
Geneck: a web server for gene network construction and visualization.
BMC bioinformatics, 20(1):1–7.