ÓMICAS

El futuro
es de todos

Gobierno
de Colombia

COLOMBIA
CIENTÍFICA

Identification of stress-responsive genes in differential co-expression networks

Camila Riccio Rengifo

Advisors:

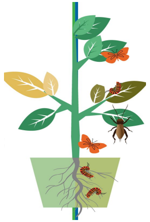
Camilo Rocha
Jorge Finke

In-silico

Quantification of the stress response in plants

Abiotic stress

- Temperature
- Drought/flood
- Salinity
- Aluminum

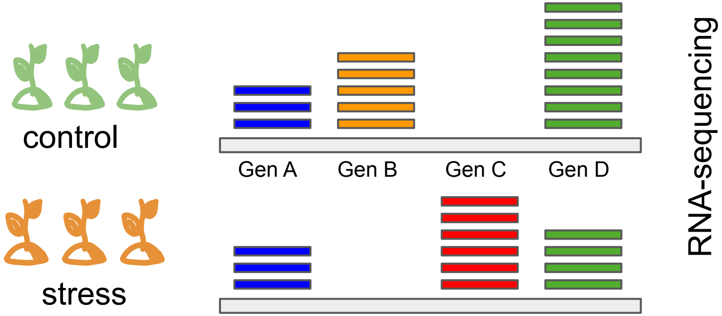


- Pathogens
- Insects
- Herbivores

Biotic stress

Response

Transcriptomic

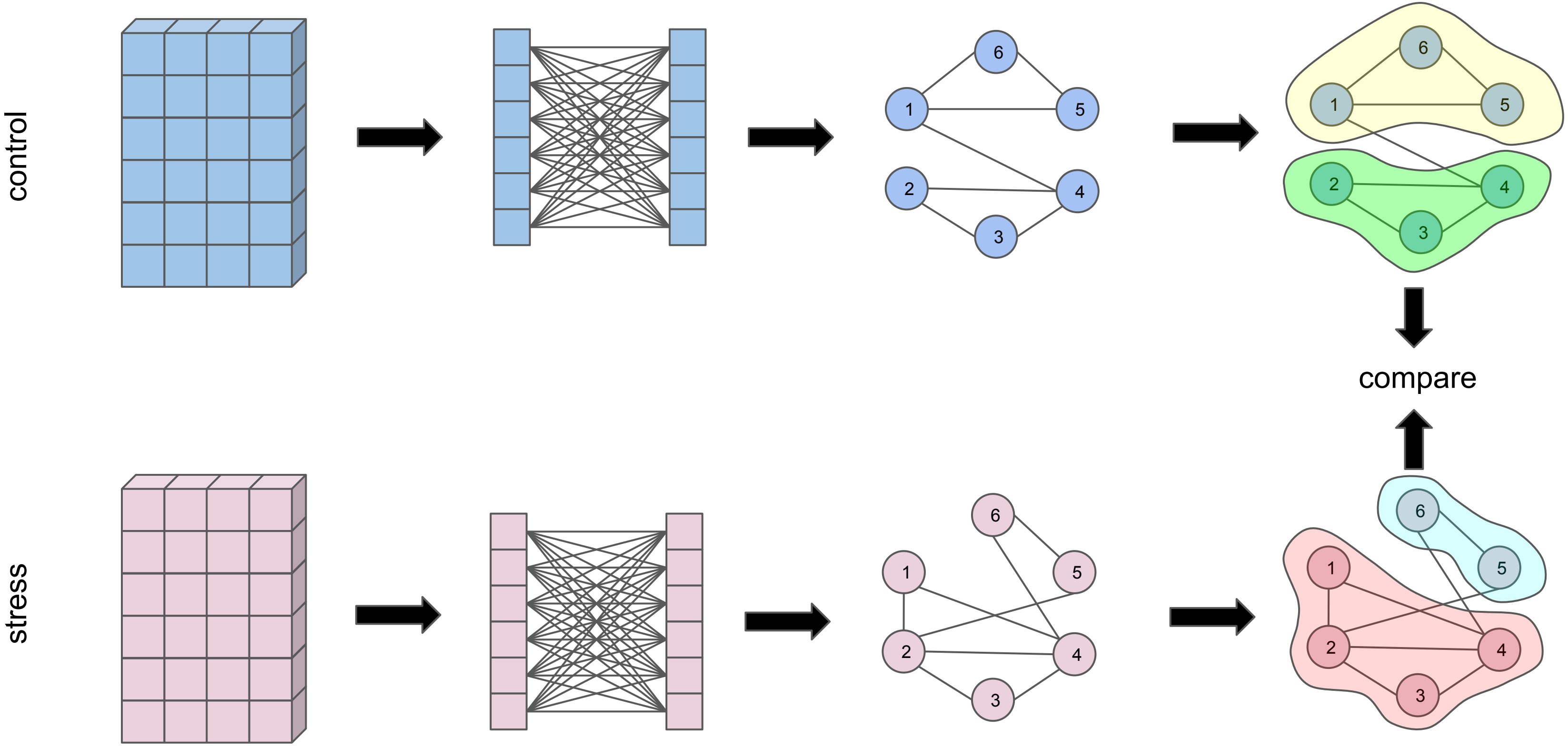


Phenomic

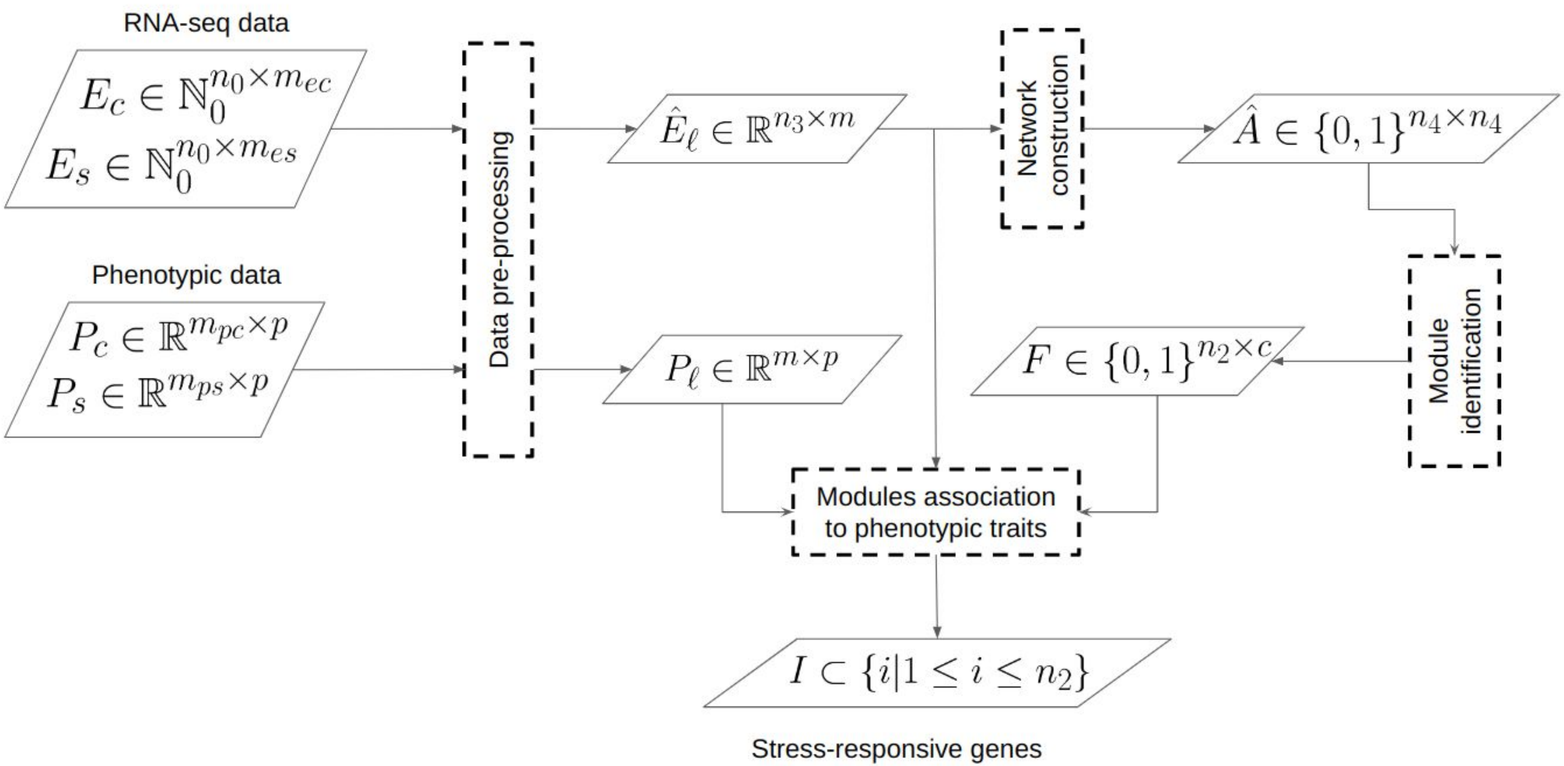
Morphological data
- Plant height
- Primary/secondary branch thickness
- Number of primary/secondary branches
- Root length

Ion concentration
- Na+, K+
- Al3+, Ca2+, Mg2+

Traditional approaches



Workflow for Differential Co-expression Analysis



RNA-seq data

$E_c \in \mathbb{N}_0^{n_0 \times m_{ec}}$

$E_s \in \mathbb{N}_0^{n_0 \times m_{es}}$

Phenotypic data

$P_c \in \mathbb{R}^{m_{pc} \times p}$

$P_s \in \mathbb{R}^{m_{ps} \times p}$

Data pre-processing

$\hat{E}_\ell \in \mathbb{R}^{n_3 \times m}$

$P_\ell \in \mathbb{R}^{m \times p}$

Network construction

$\hat{A} \in \{0,1\}^{n_4 \times n_4}$

Module identification

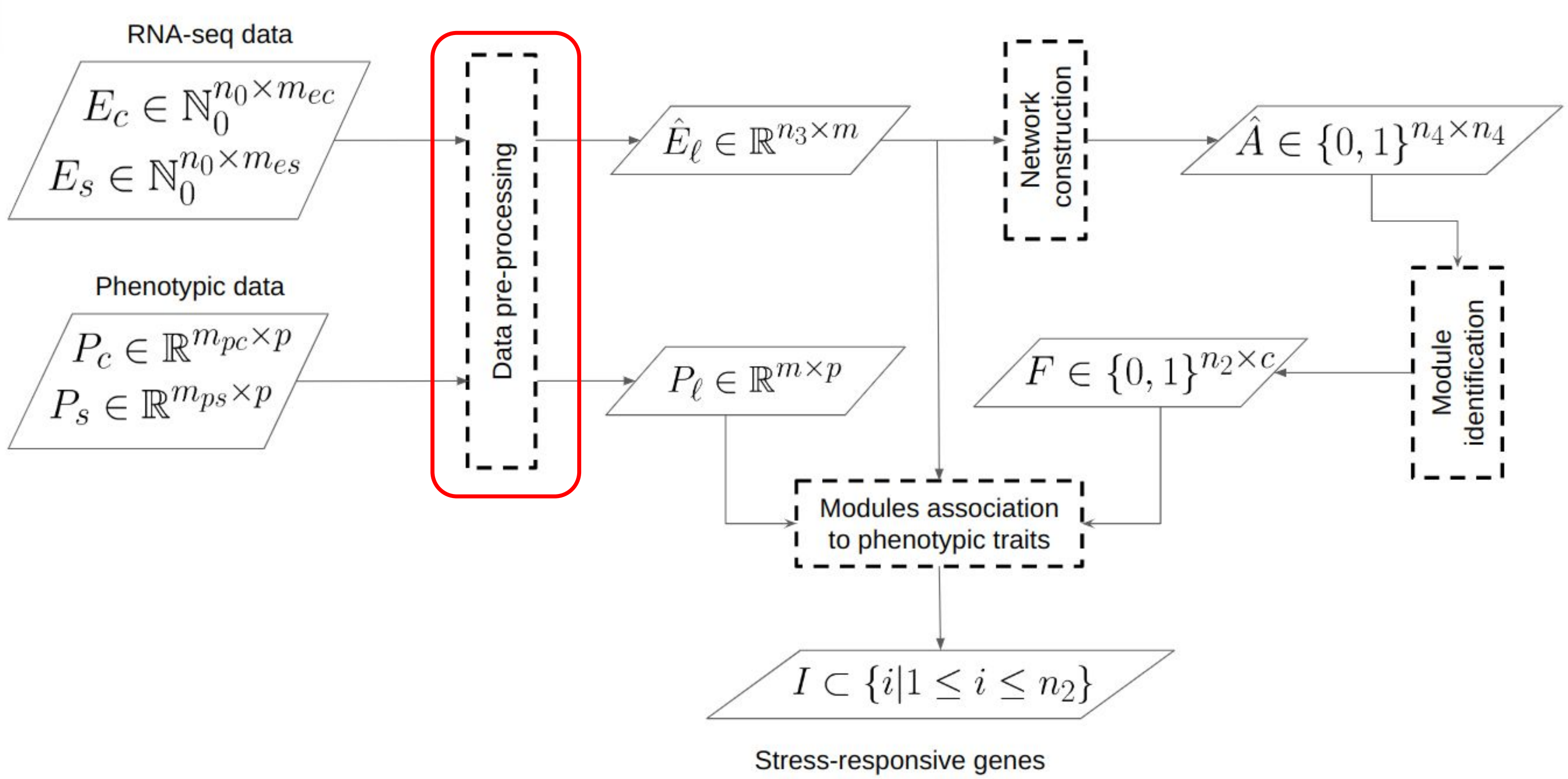
$F \in \{0,1\}^{n_2 \times c}$

Modules association to phenotypic traits

$I \subset \{i | 1 \leq i \leq n_2\}$

Stress-responsive genes

Workflow



RNA-seq data

$E_c \in \mathbb{N}_0^{n_0 \times m_{ec}}$

$E_s \in \mathbb{N}_0^{n_0 \times m_{es}}$

Phenotypic data

$P_c \in \mathbb{R}^{m_{pc} \times p}$

$P_s \in \mathbb{R}^{m_{ps} \times p}$

Data pre-processing

$\hat{E}_\ell \in \mathbb{R}^{n_3 \times m}$

$P_\ell \in \mathbb{R}^{m \times p}$

Network construction

$\hat{A} \in \{0,1\}^{n_4 \times n_4}$

Module identification

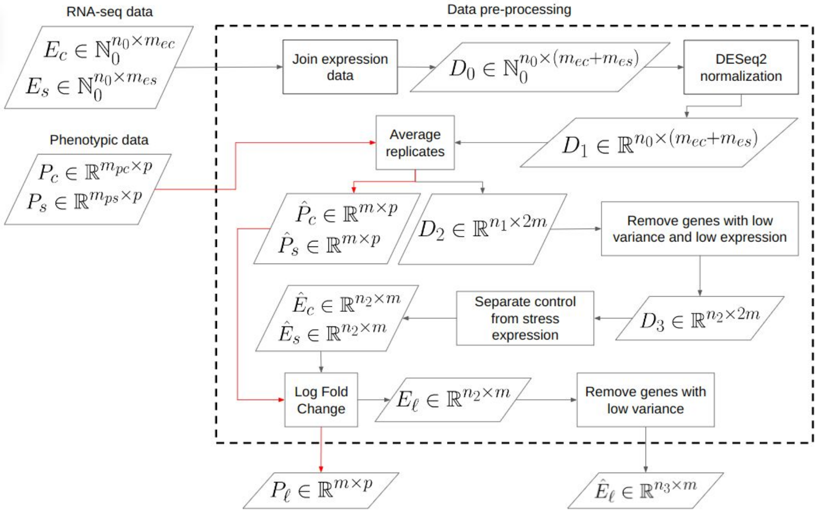
$F \in \{0,1\}^{n_2 \times c}$

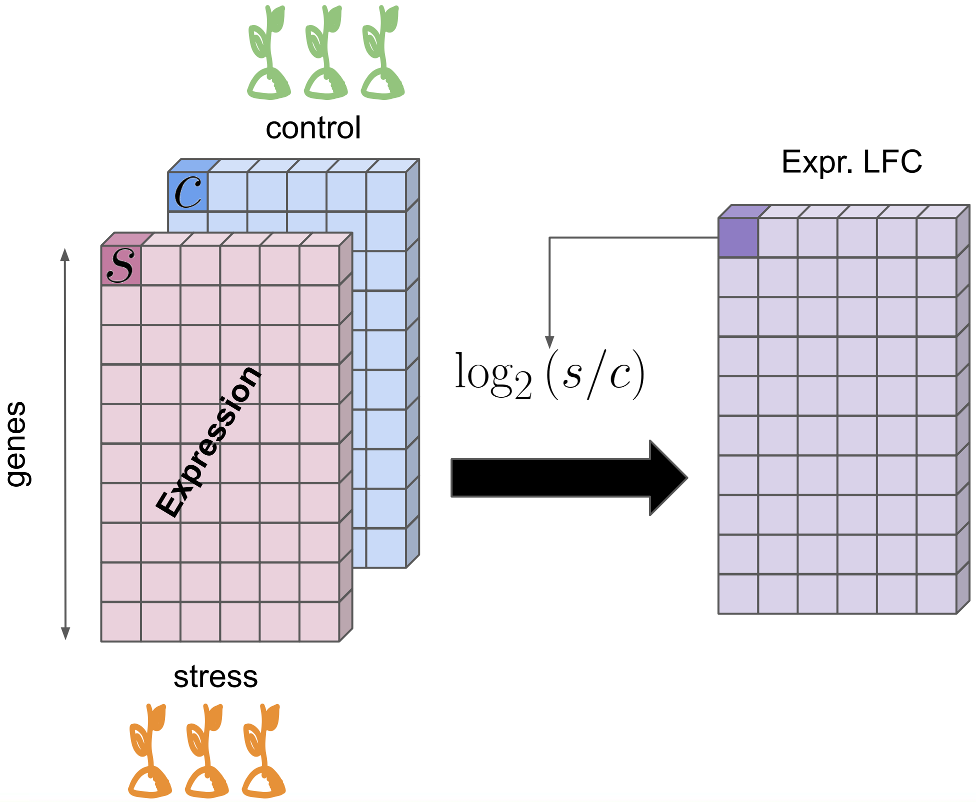
Modules association to phenotypic traits

$I \subset \{i | 1 \leq i \leq n_2\}$

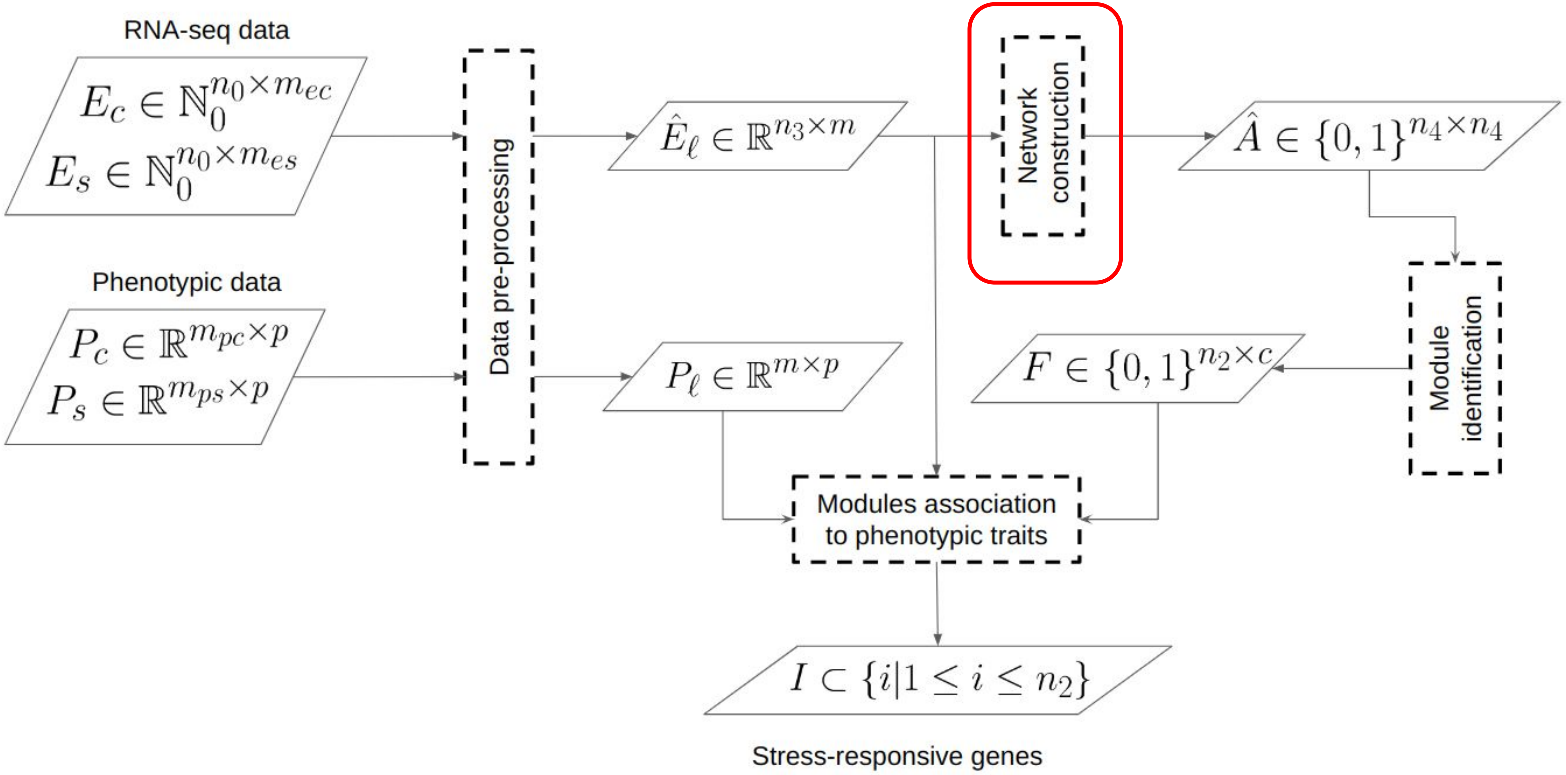
Stress-responsive genes

Data pre-processing

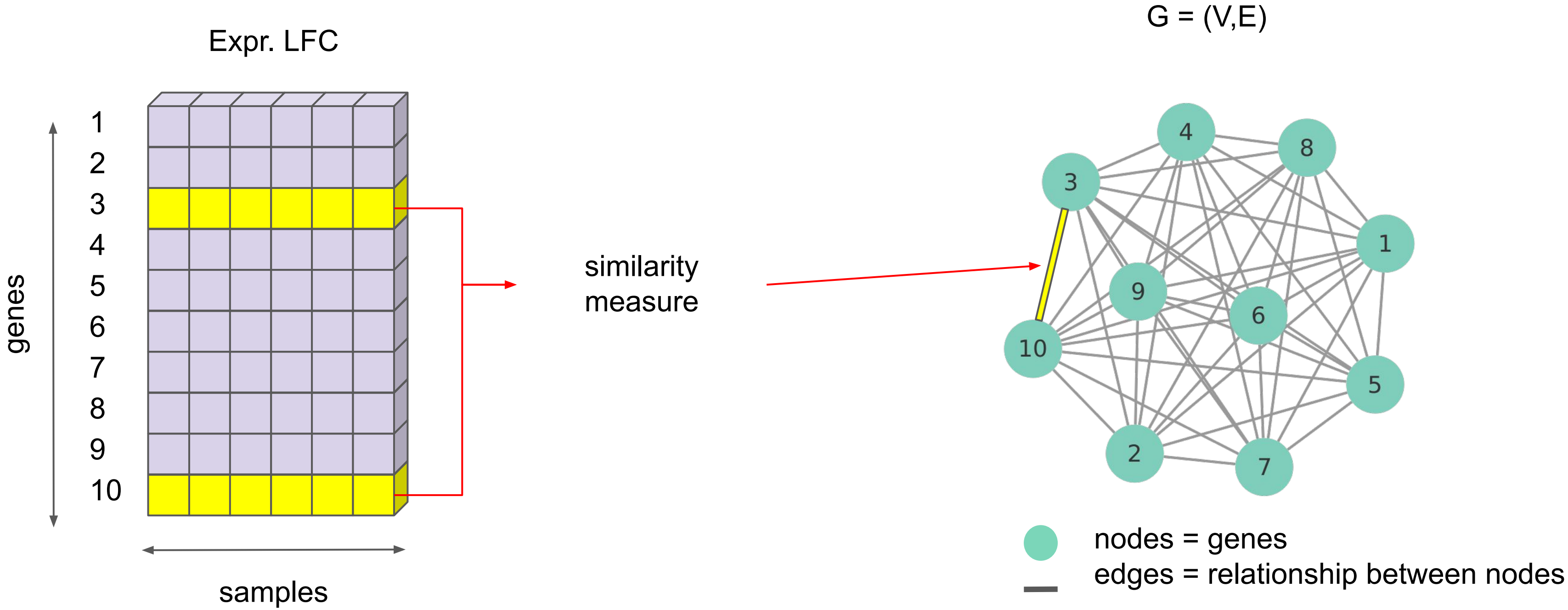




Workflow


RNA-seq data
$E_c \in \mathbb{N}_0^{n_0 \times m_{ec}}$
$E_s \in \mathbb{N}_0^{n_0 \times m_{es}}$
Phenotypic data
$P_c \in \mathbb{R}^{m_{pc} \times p}$
$P_s \in \mathbb{R}^{m_{ps} \times p}$
Data pre-processing
$\hat{E}_\ell \in \mathbb{R}^{n_3 \times m}$
$P_\ell \in \mathbb{R}^{m \times p}$
Network construction
$\hat{A} \in \{0,1\}^{n_4 \times n_4}$
Module identification
$F \in \{0,1\}^{n_2 \times c}$
Modules association to phenotypic traits
$I \subset \{i | 1 \leq i \leq n_2\}$
Stress-responsive genes

Network construction

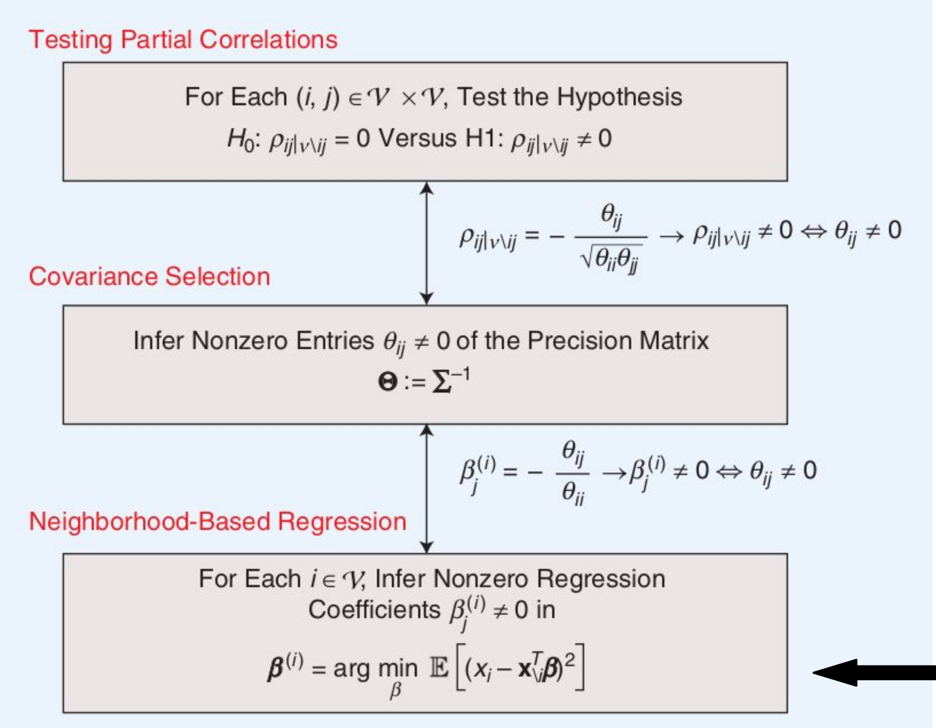


similarity measure

● nodes = genes

▬ edges = relationship between nodes

Network construction



Testing Partial Correlations

For Each $(i, j) \in \mathcal{V} \times \mathcal{V}$, Test the Hypothesis
H_0: $\rho_{ij|\mathcal{V}\setminus ij} = 0$ Versus H1: $\rho_{ij|\mathcal{V}\setminus ij} \neq 0$

$$\rho_{ij|\mathcal{V}\setminus ij} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}} \rightarrow \rho_{ij|\mathcal{V}\setminus ij} \neq 0 \Leftrightarrow \theta_{ij} \neq 0$$

Covariance Selection

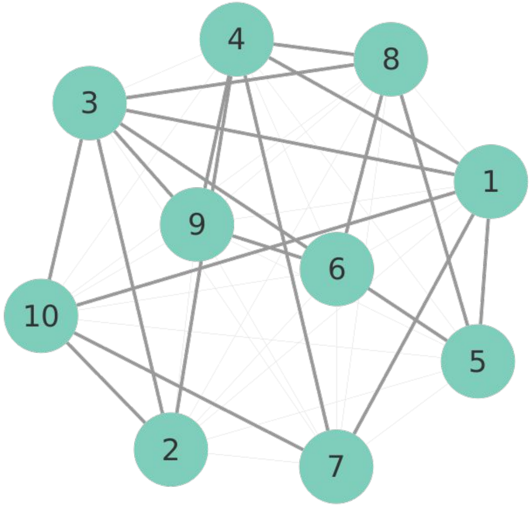
Infer Nonzero Entries $\theta_{ij} \neq 0$ of the Precision Matrix
$\boldsymbol{\Theta} := \boldsymbol{\Sigma}^{-1}$

$$\beta_j^{(i)} = -\frac{\theta_{ij}}{\theta_{ii}} \rightarrow \beta_j^{(i)} \neq 0 \Leftrightarrow \theta_{ij} \neq 0$$

Neighborhood-Based Regression

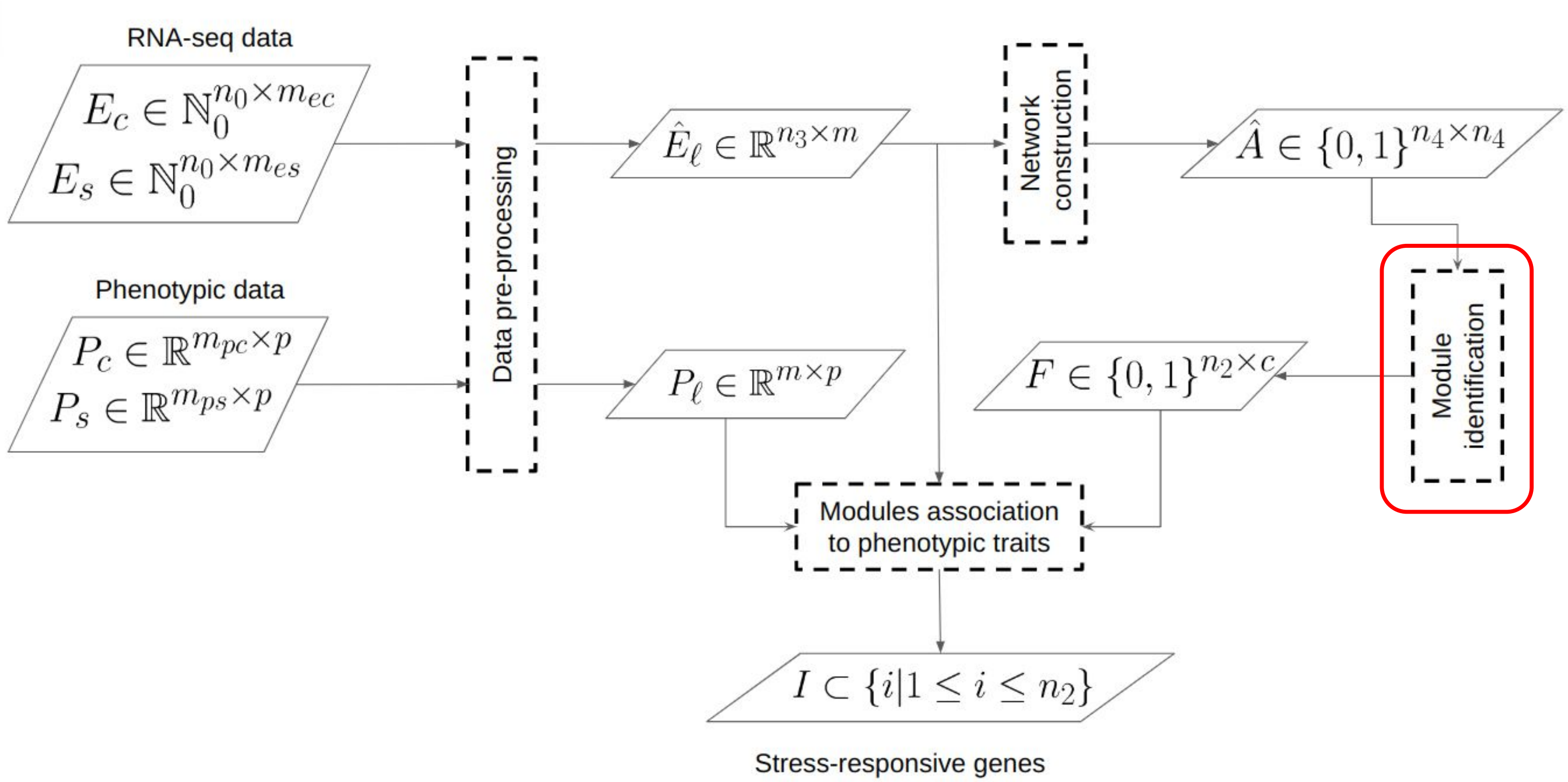
For Each $i \in \mathcal{V}$, Infer Nonzero Regression Coefficients $\beta_j^{(i)} \neq 0$ in

$$\boldsymbol{\beta}^{(i)} = \arg\min_{\beta} \mathbb{E}\left[(x_i - \mathbf{x}_{\setminus i}^T \boldsymbol{\beta})^2\right]$$



less dense networks

Workflow



RNA-seq data

$E_c \in \mathbb{N}_0^{n_0 \times m_{ec}}$

$E_s \in \mathbb{N}_0^{n_0 \times m_{es}}$

Phenotypic data

$P_c \in \mathbb{R}^{m_{pc} \times p}$

$P_s \in \mathbb{R}^{m_{ps} \times p}$

Data pre-processing

$\hat{E}_\ell \in \mathbb{R}^{n_3 \times m}$

$P_\ell \in \mathbb{R}^{m \times p}$

Network construction

$\hat{A} \in \{0,1\}^{n_4 \times n_4}$

Module identification

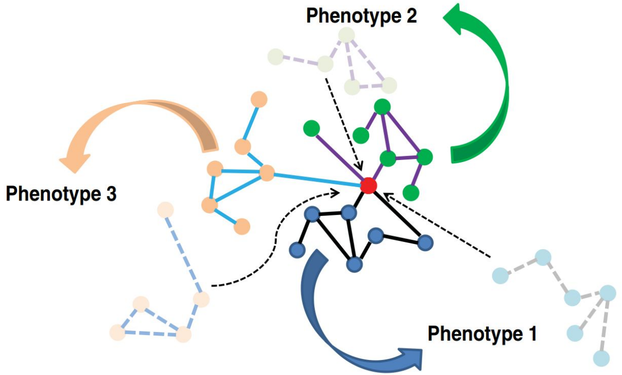
$F \in \{0,1\}^{n_2 \times c}$

Modules association to phenotypic traits

$I \subset \{i | 1 \leq i \leq n_2\}$

Stress-responsive genes

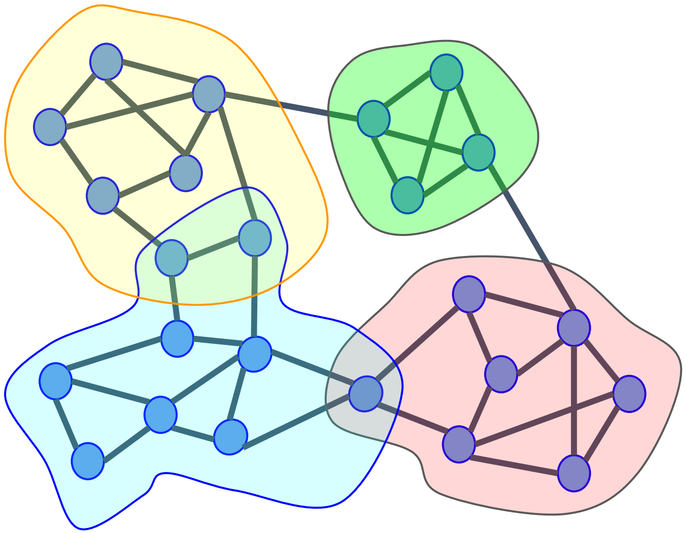
Module identification



Gene co-expression modules are likely to overlap due to the multiple regulatory domains a gene can be part of.

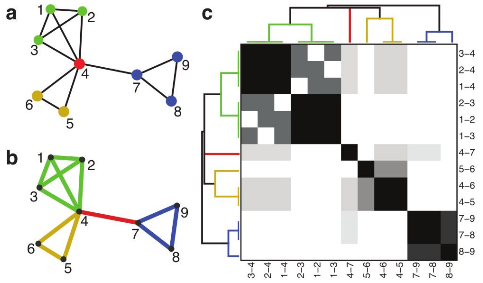
Pleiotropy is the phenomenon of a single gene controlling or influencing multiple (and possibly unrelated) phenotypic traits.

Clustering

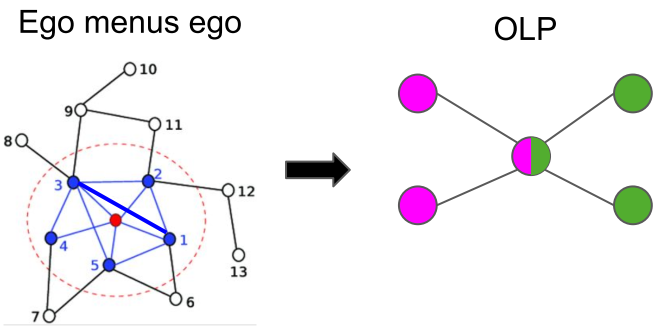


Clustering with overlapping nodes

Hierarchical Link Clustering (HLC)

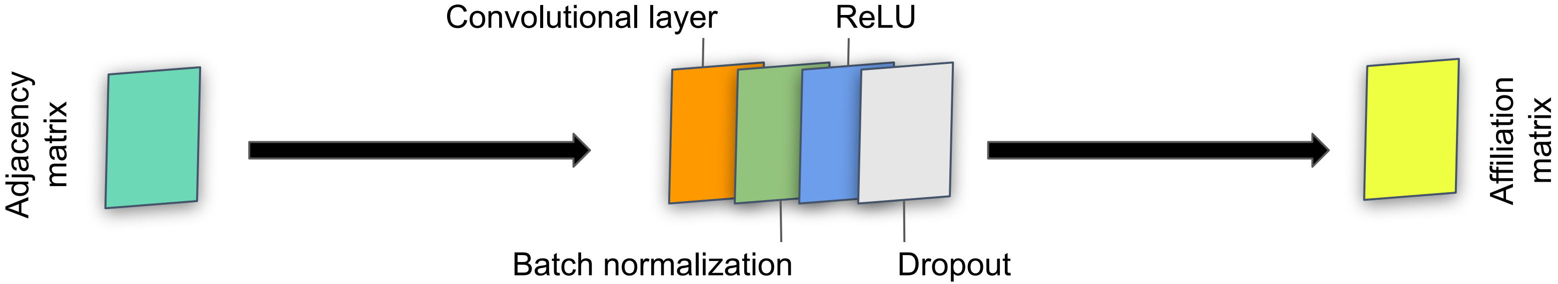


ANGEL


Ego menus ego

OLP

Neural Overlapping Community Detection (NOCD)


Adjacency matrix

Convolutional layer

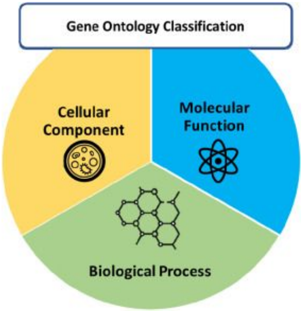
ReLU

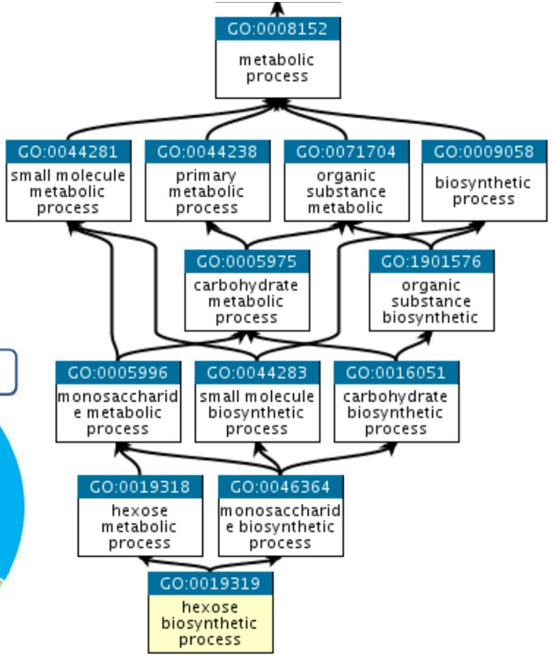
Batch normalization

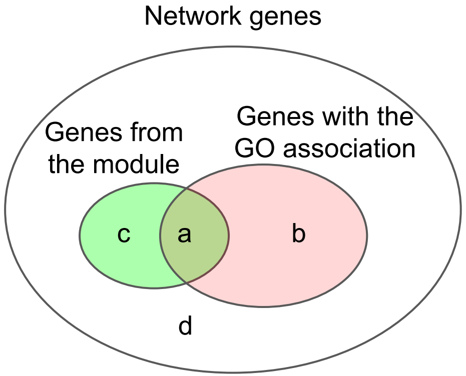
Dropout

Affiliation matrix

Module enrichment





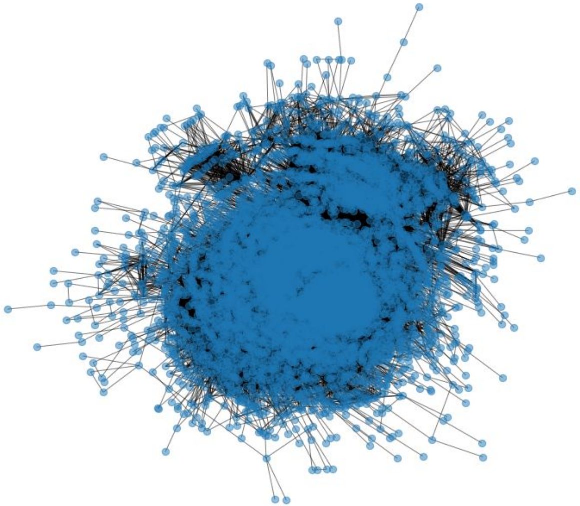


Fisher test:

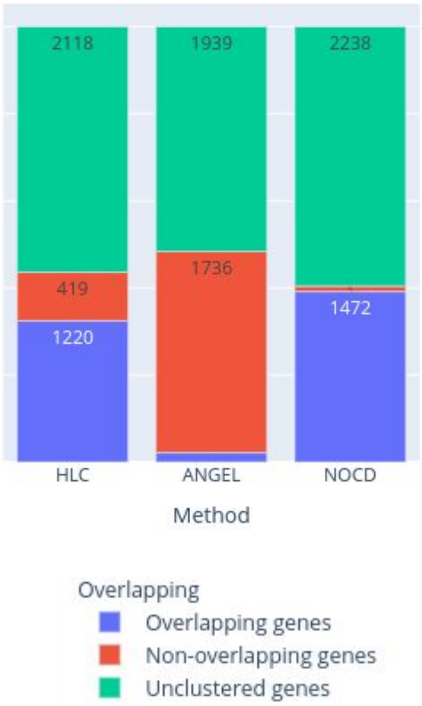
H0: The module is a random sample from the network.
H1: The module has more genes annotated with the GO term than expected by chance.

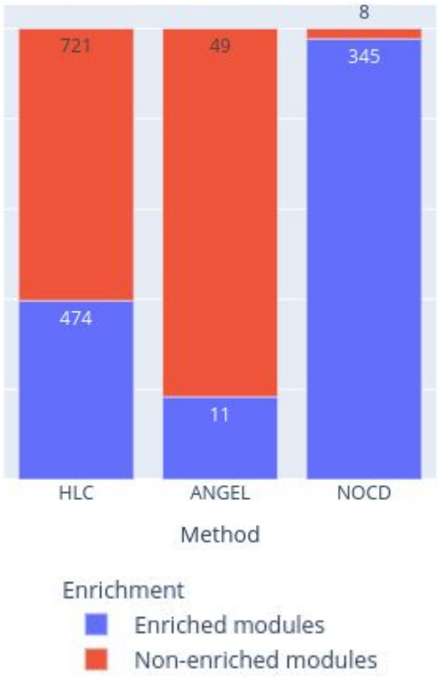
$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b}\binom{c+d}{d}}{\binom{n}{b+d}}$$

Module enrichment

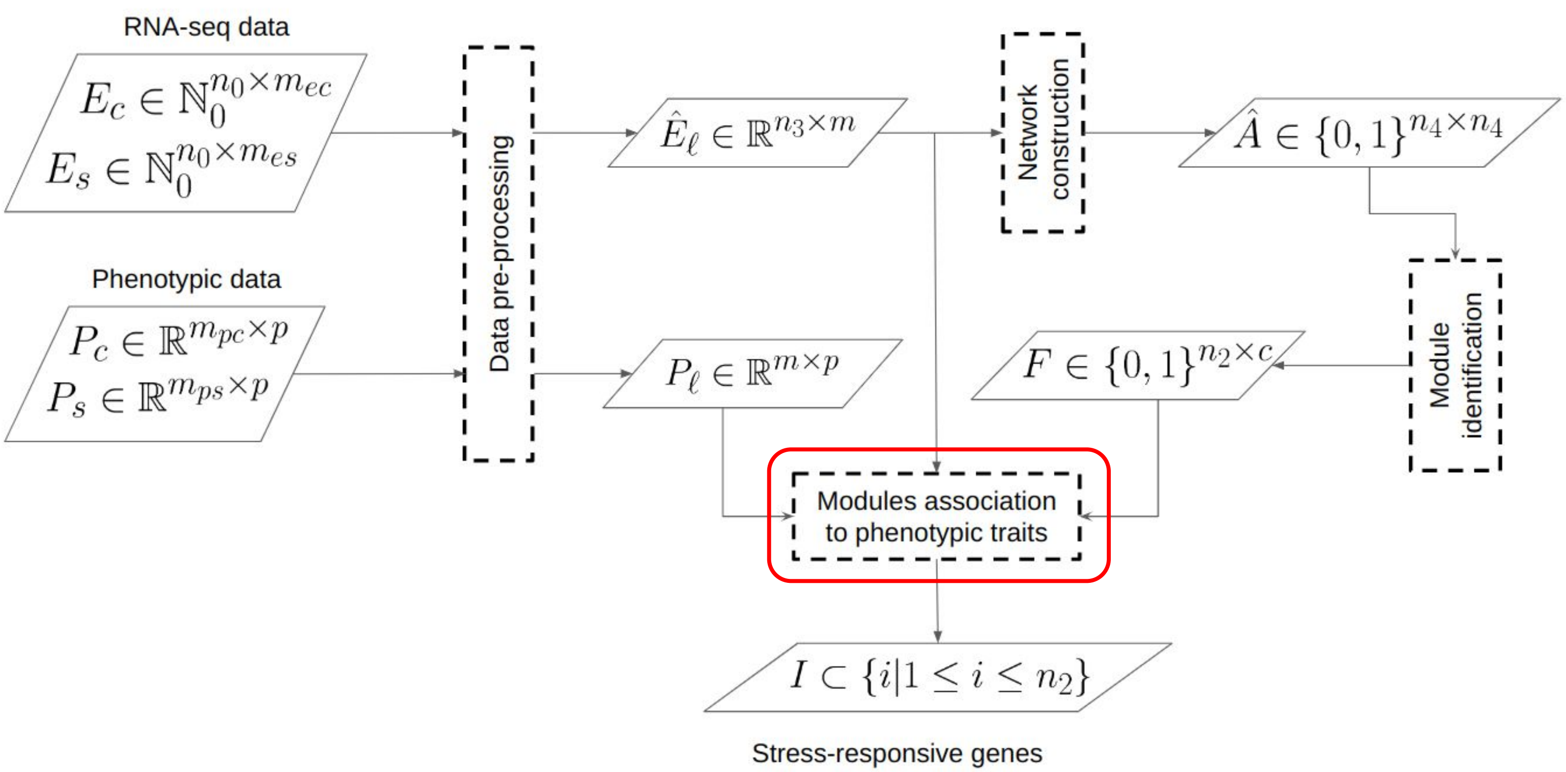


|V| = 3,757
|E| = 246,360





Workflow



RNA-seq data

$E_c \in \mathbb{N}_0^{n_0 \times m_{ec}}$

$E_s \in \mathbb{N}_0^{n_0 \times m_{es}}$

Phenotypic data

$P_c \in \mathbb{R}^{m_{pc} \times p}$

$P_s \in \mathbb{R}^{m_{ps} \times p}$

Data pre-processing

$\hat{E}_\ell \in \mathbb{R}^{n_3 \times m}$

$P_\ell \in \mathbb{R}^{m \times p}$

Network construction

$\hat{A} \in \{0,1\}^{n_4 \times n_4}$

Module identification

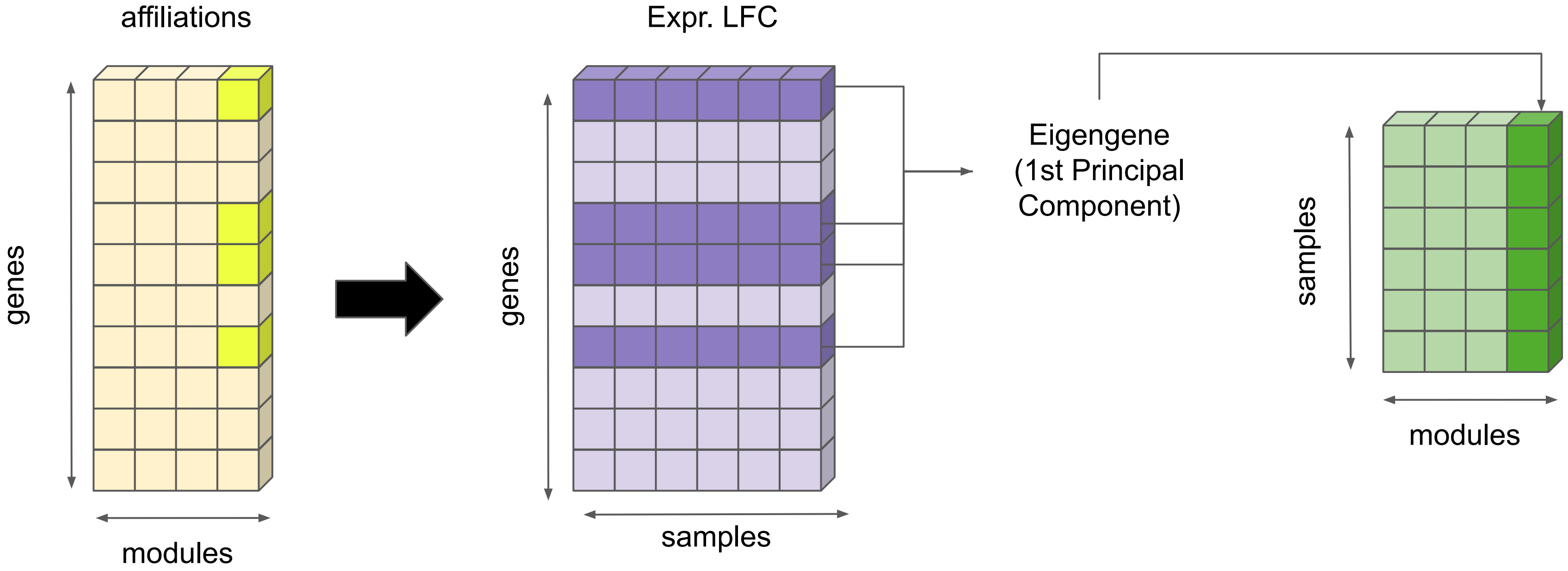
$F \in \{0,1\}^{n_2 \times c}$

Modules association to phenotypic traits

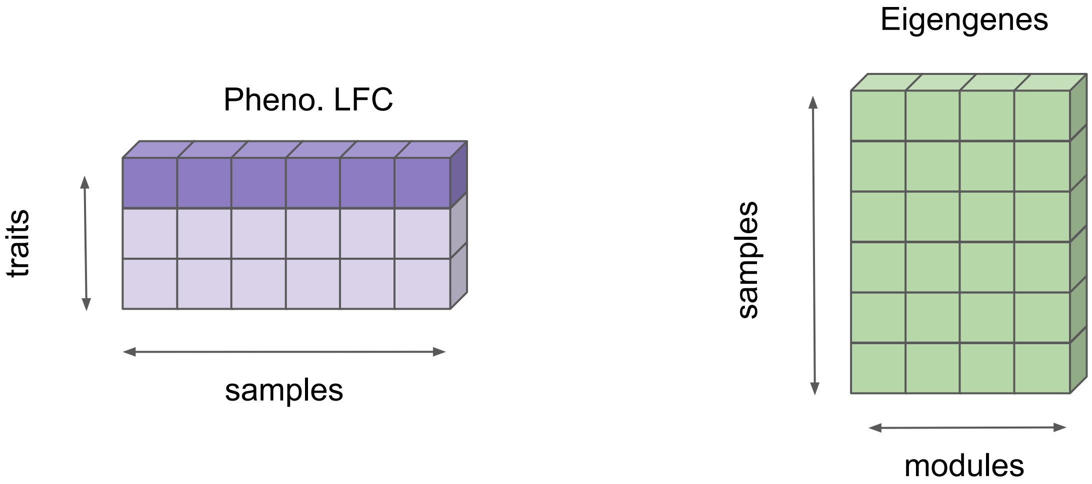
$I \subset \{i | 1 \leq i \leq n_2\}$

Stress-responsive genes

Module association to phenotypic traits



Module association to phenotypic traits



🟪 Y_i

🟩 X

→

LASSO

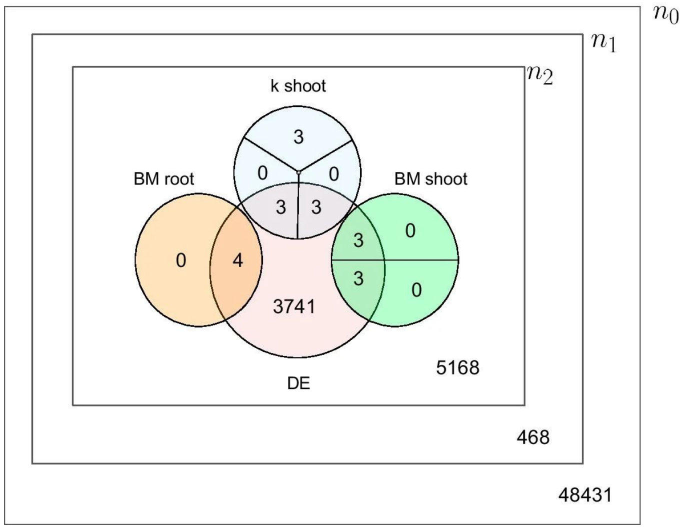
$$\underset{\beta_i}{\text{minimize}} \left\|Y_i - X_i\beta_i\right\|_2^2 + \lambda \left\|\beta_i\right\|_1$$

→

subset of modules

Module association to phenotypic traits

Case study: *Oryza sativa* (rice) under salt stress



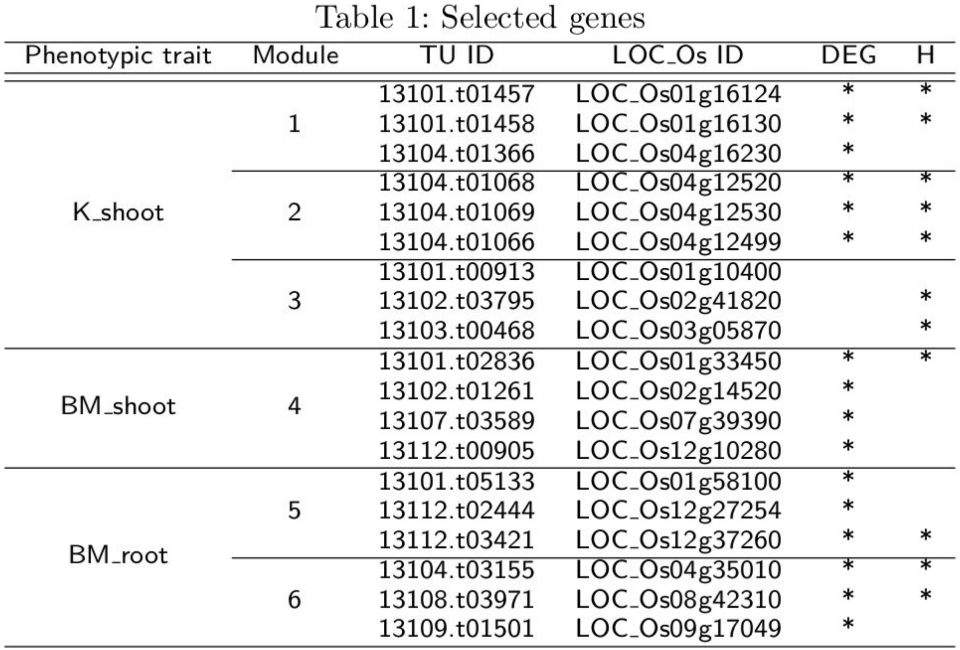


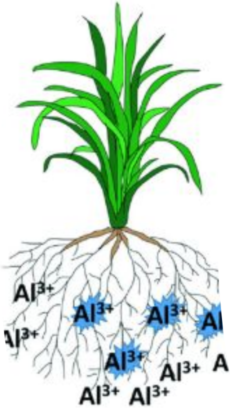
Table 1: Selected genes

Phenotypic trait	Module	TU ID	LOC_Os ID	DEG	H
K_shoot	1	13101.t01457	LOC_Os01g16124	*	*
		13101.t01458	LOC_Os01g16130	*	*
		13104.t01366	LOC_Os04g16230	*	
	2	13104.t01068	LOC_Os04g12520	*	*
		13104.t01069	LOC_Os04g12530	*	*
		13104.t01066	LOC_Os04g12499	*	*
	3	13101.t00913	LOC_Os01g10400		
		13102.t03795	LOC_Os02g41820		*
		13103.t00468	LOC_Os03g05870		*
BM_shoot	4	13101.t02836	LOC_Os01g33450	*	*
		13102.t01261	LOC_Os02g14520	*	
		13107.t03589	LOC_Os07g39390	*	
		13112.t00905	LOC_Os12g10280	*	
BM_root	5	13101.t05133	LOC_Os01g58100	*	
		13112.t02444	LOC_Os12g27254	*	
		13112.t03421	LOC_Os12g37260	*	*
	6	13104.t03155	LOC_Os04g35010	*	*
		13108.t03971	LOC_Os08g42310	*	*
		13109.t01501	LOC_Os09g17049	*	

What's next?

What's next?

- Systematic comparison between the results obtained with the different algorithms applied in each step of the workflow.
- Comparison of results with standard techniques (e.g., WGCNA and DiffCoExp)
- Application to OMICAS own data: Rice under aluminum stress.
- Application in cenicaña data.
- NMTF algorithm as a clustering method to incorporate a priori information on gene interactions.



cenicaña

Thank you very much

Questions?

ómicas

Aliados

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

cenicaña

FEDEARROZ
FEDERACIÓN NACIONAL DE ARROCEROS

NIAB

intelecto
Soluciones y Tecnología

HI TECH
AUTOMATIZACION S.A.

Caltech

GHENT UNIVERSITY

東京大学
THE UNIVERSITY OF TOKYO

ILLINOIS

Pontificia Universidad JAVERIANA Cali

IES Ancla

UNIVERSIDAD DEL QUINDÍO

Universidad de Ibagué

UNILLANOS
UNIVERSIDAD DE LOS LLANOS

Universidad de los Andes Colombia

UNIVERSIDAD ICESI

Pontificia Universidad JAVERIANA Bogotá

Apoyan

El futuro es de todos
Gobierno de Colombia

COLOMBIA CIENTÍFICA