

# Redes de co-expresión en el mejoramiento de cultivos

Camila Riccio Rengifo

Candidata a Doctora  
Doctorado en Ingeniería y Ciencias Aplicadas  
Pontificia Universidad Javeriana – Cali, Colombia



Pontificia Universidad  
**JAVERIANA**  
Cali



# Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles



Aliados



Apoyan



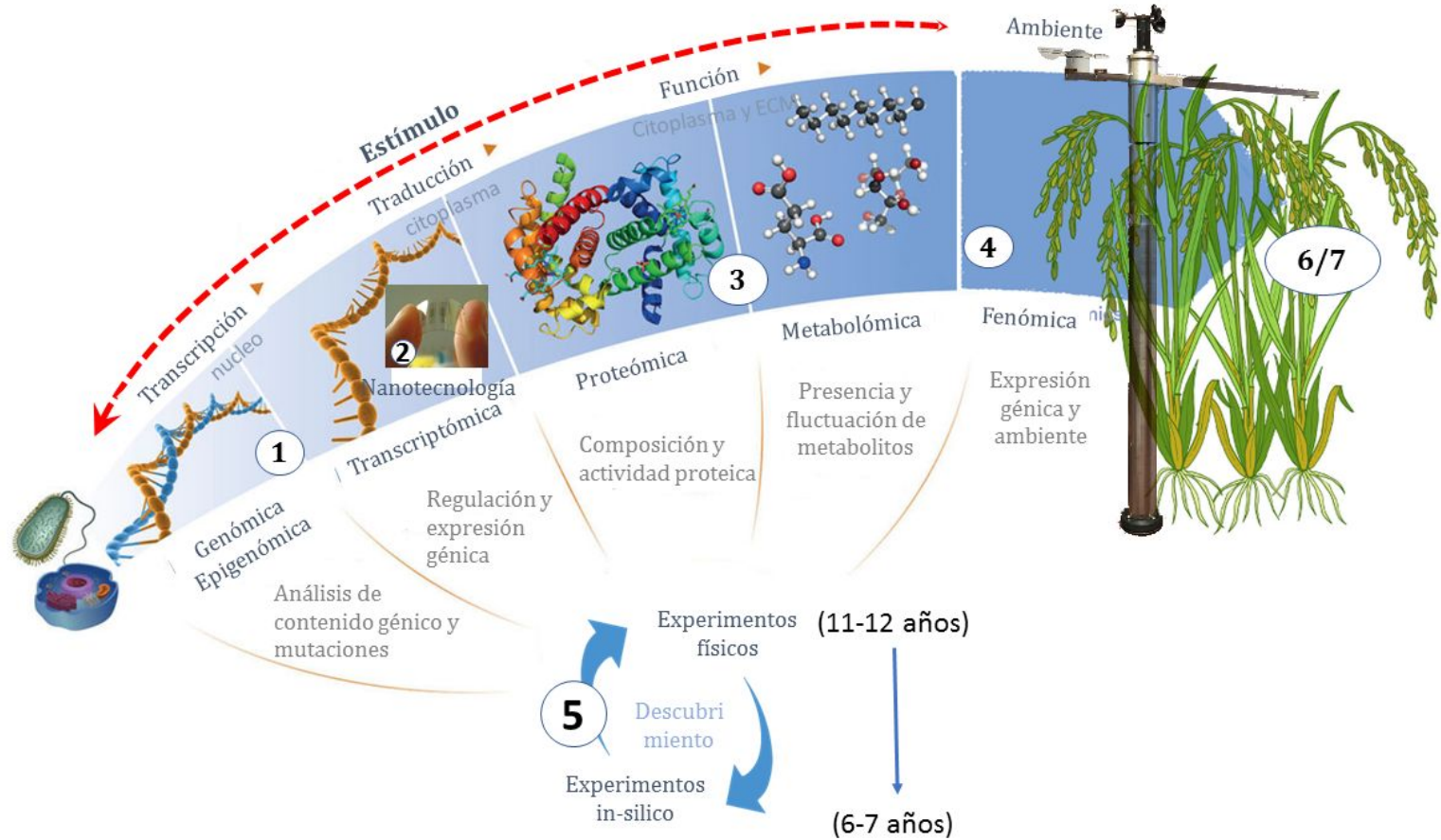
El futuro  
es de todos

Gobierno  
de Colombia



COLOMBIA  
CIENTIFICA

# ÓMICAS



# Cuantificación de la respuesta a estrés en plantas

## Estrés Abiótico

- Calor/Frío
- Sequía/Inundación
- Salinidad
- Aluminio



- Patógenos
- Insectos
- Herbívoros

## Estrés Biótico

Respuesta

Transcriptómica

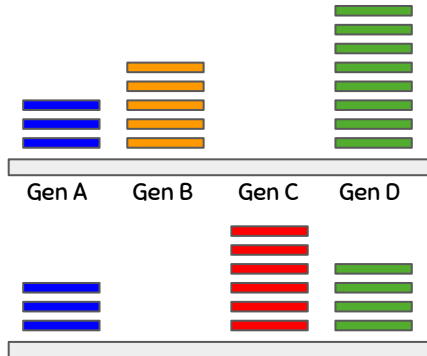
Fenómica



control



estrés



RNA-seq

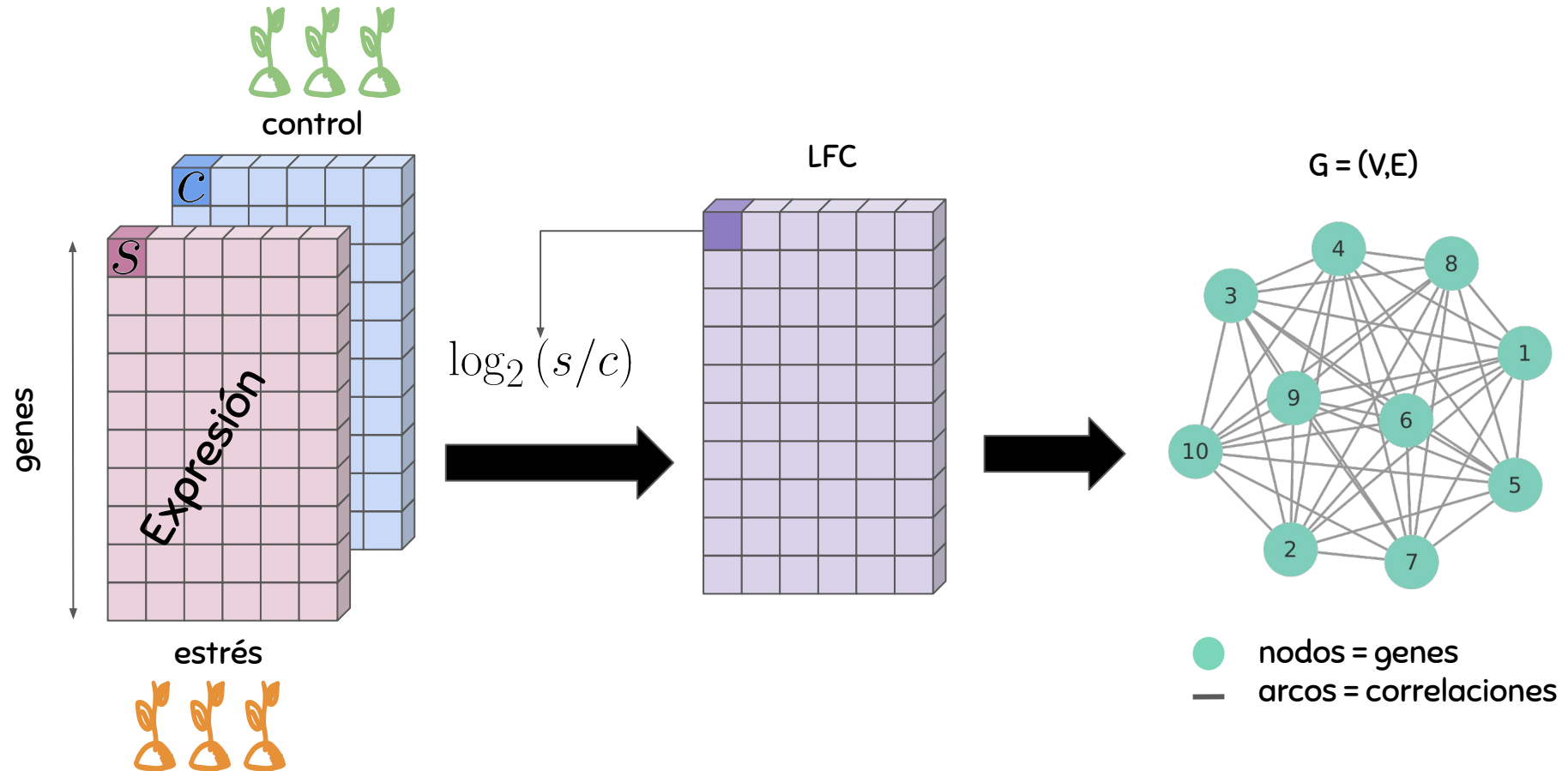
## Datos morfológicos

- Altura de la planta
- Grosor ramas primarias/secundarias
- Número de ramas primarias/secundarias
- Longitud raíces

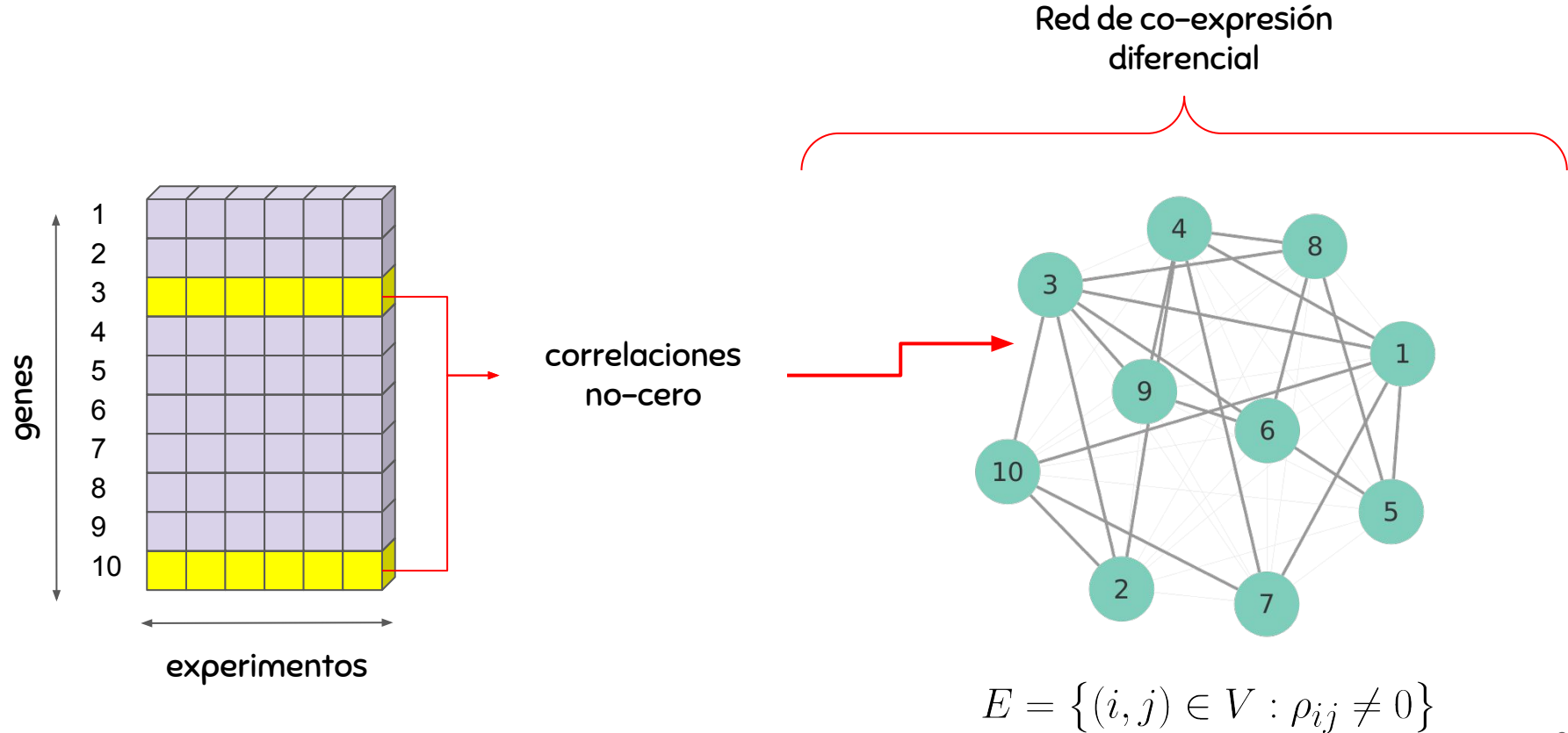
## Concentración de iones

- $\text{Na}^+$ ,  $\text{K}^+$
- $\text{Al}^{3+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$

# Red de co-expresión diferencial



# Inferencia de correlaciones no-cero



# Inferencia de correlaciones no-cero

## Test correlaciones parciales

Para cada  $(i, j) \in V \times V$ , probar la hipótesis

$$H_0 : \rho_{ij|V \setminus ij} = 0 \quad \text{vs} \quad H_1 : \rho_{ij|V \setminus ij} \neq 0$$

## Selección de covarianza

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

Inferir entradas no-cero  $\theta_{ij} \neq 0$  de la matriz de precisión

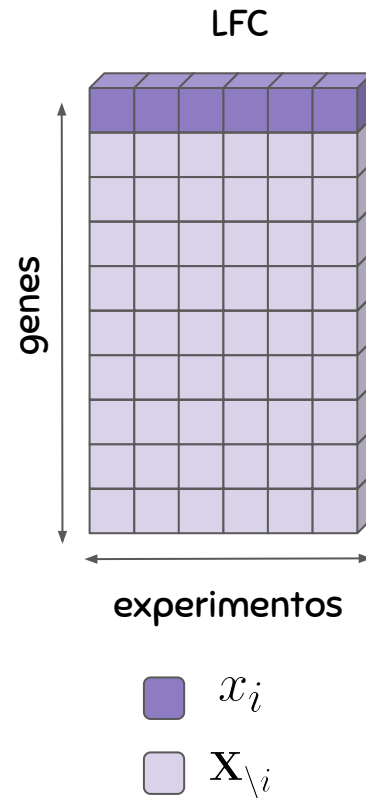
$$\Theta := \Sigma^{-1}$$

## Regresión basada en vecinos

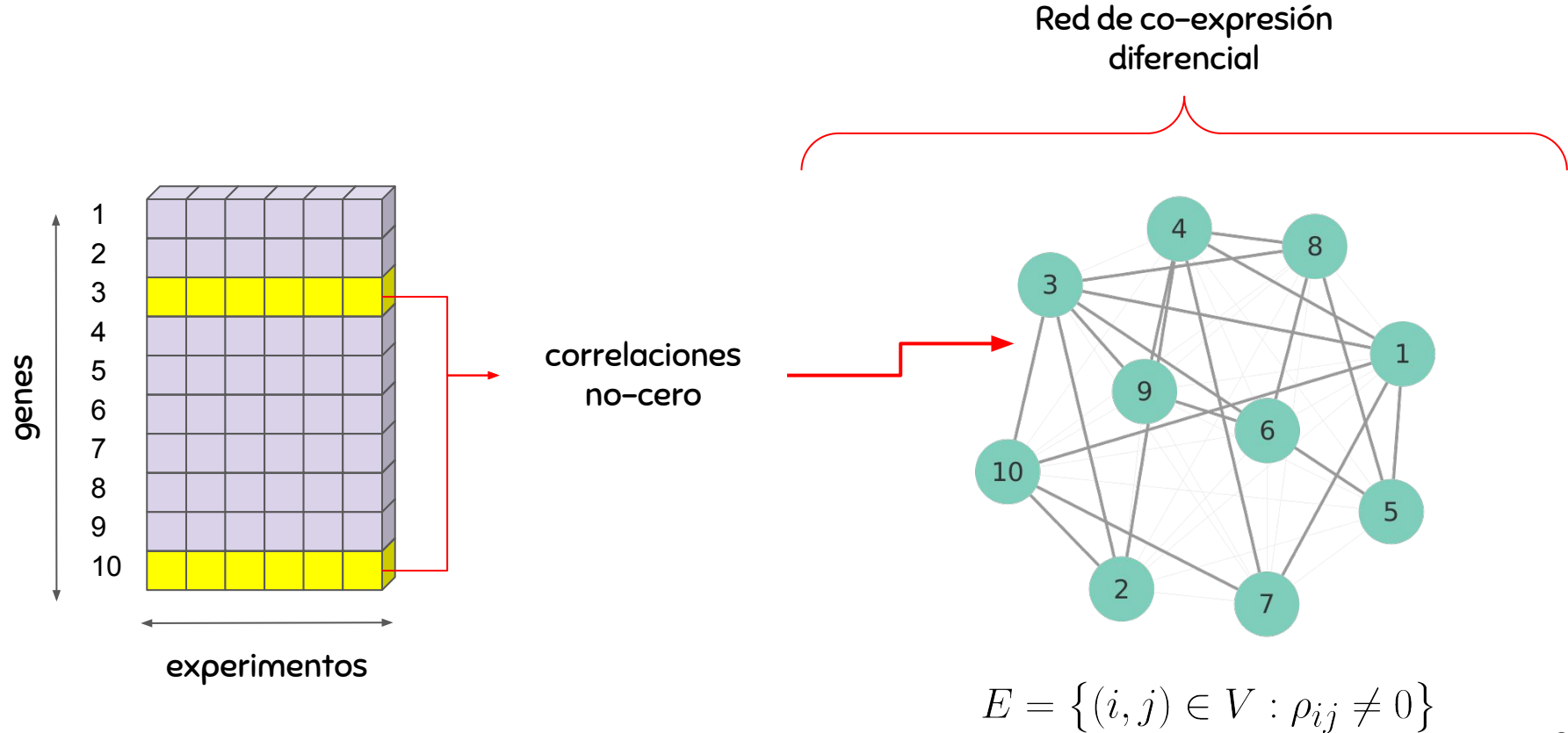
$$\updownarrow \quad \beta_j^{(i)} = -\frac{\theta_{ij}}{\theta_{ii}} \rightarrow \beta_j^{(i)} \neq 0 \Leftrightarrow \theta_{ij} \neq 0$$

Para cada  $i \in V$ , inferir coeficientes de regresión no-cero  $\beta_j^{(i)} \neq 0$  en

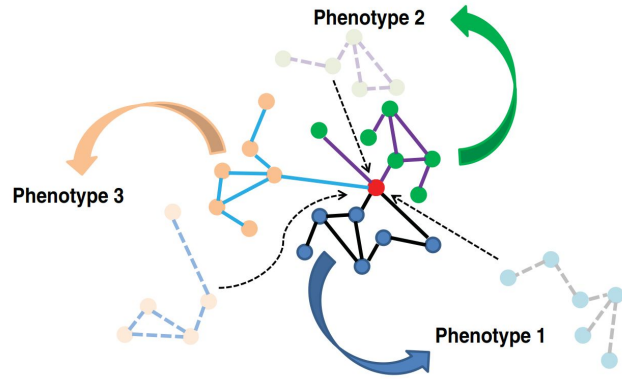
$$\beta^{(i)} = \arg \min_{\beta} \mathbb{E} \left[ \left( x_i - \mathbf{x}_{\setminus i}^T \beta \right)^2 \right]$$



# Inferencia de correlaciones no-cero

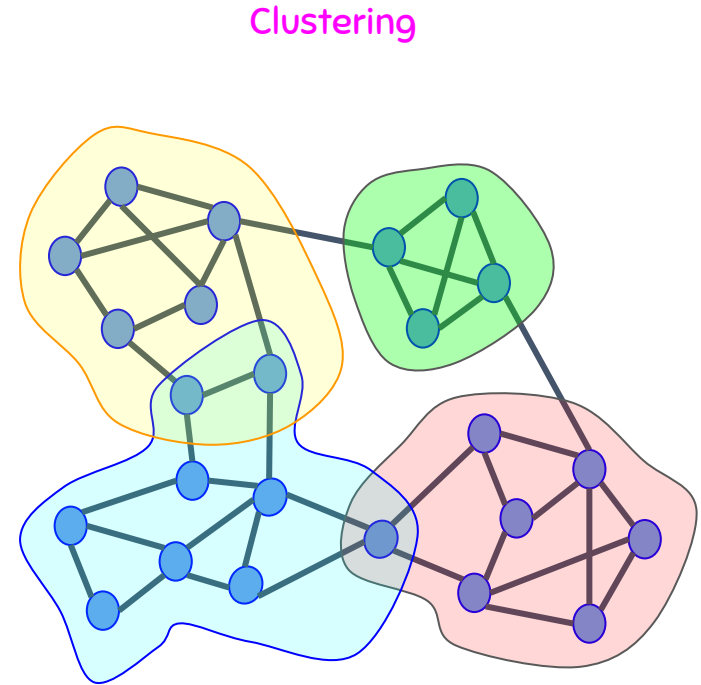


# Detección de módulos con superposición



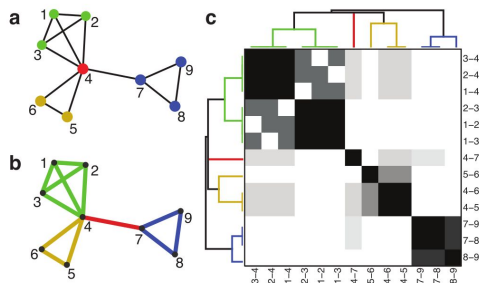
Los módulos de genes co-expresados tienden a **superponerse** debido a los múltiples dominios regulatorios de los que puede hacer parte un gen.

**Pleiotropía** es el fenómeno de un solo gen controlando o influenciando múltiples (y posiblemente no relacionados) rasgos fenotípicos.



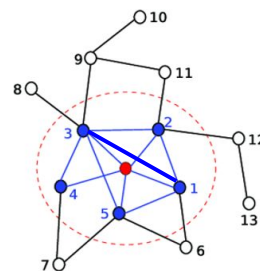
# Detección de módulos con superposición

## Hierarchical Link Clustering (HLC)

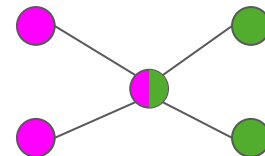


## ANGEL

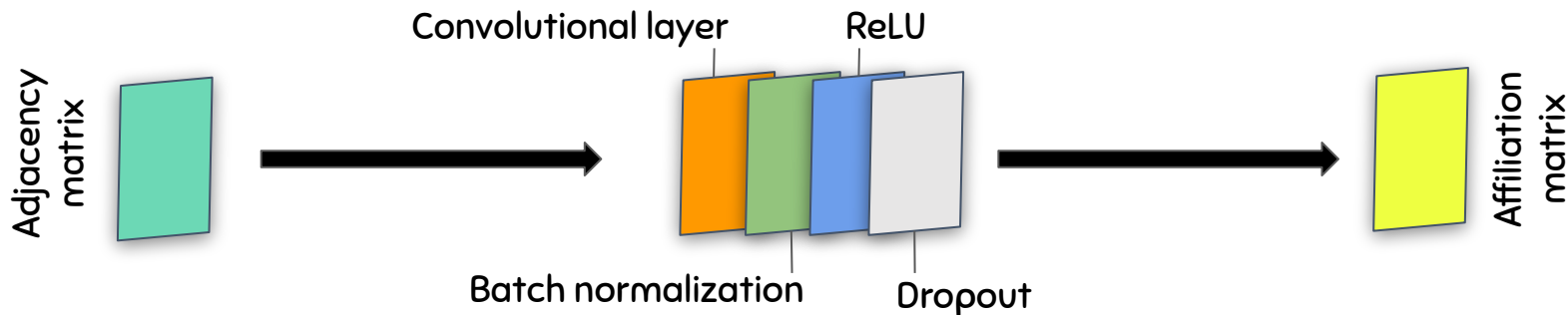
Ego menos ego



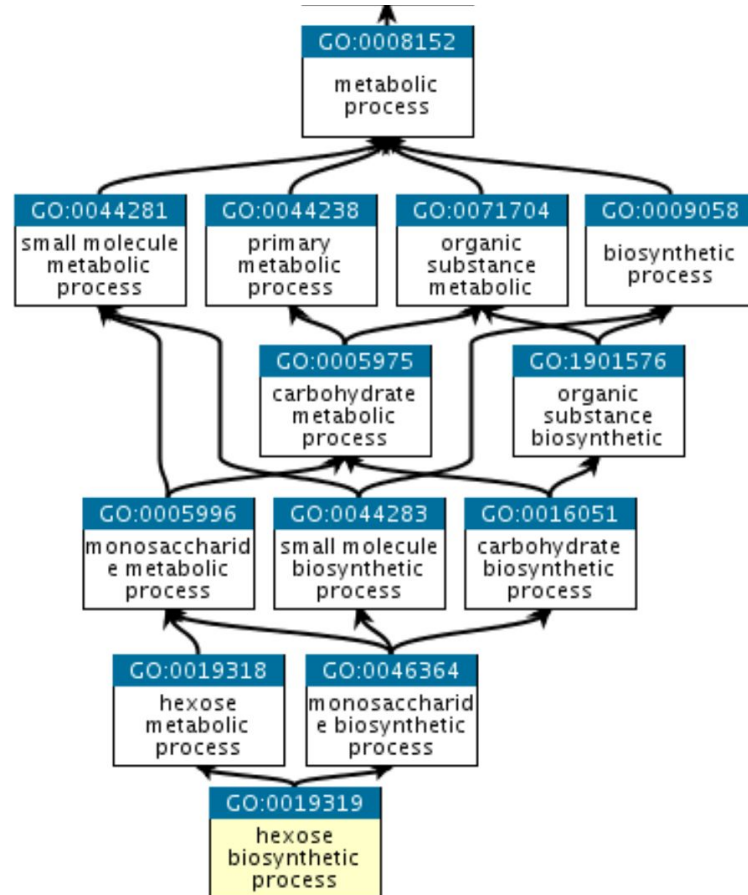
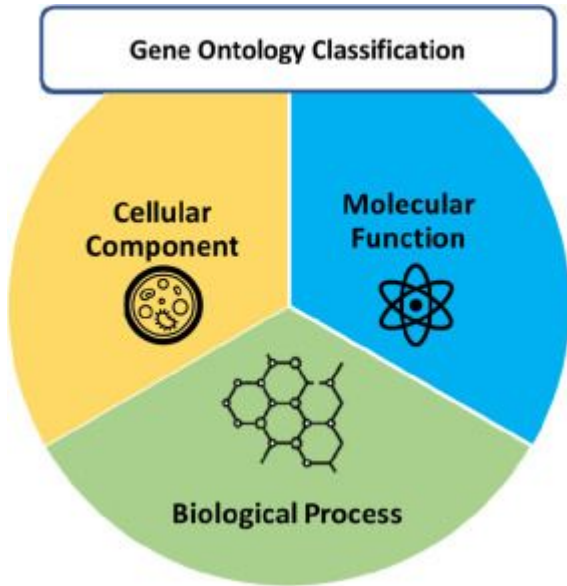
OLP



## Neural Overlapping Community Detection (NOCOD)

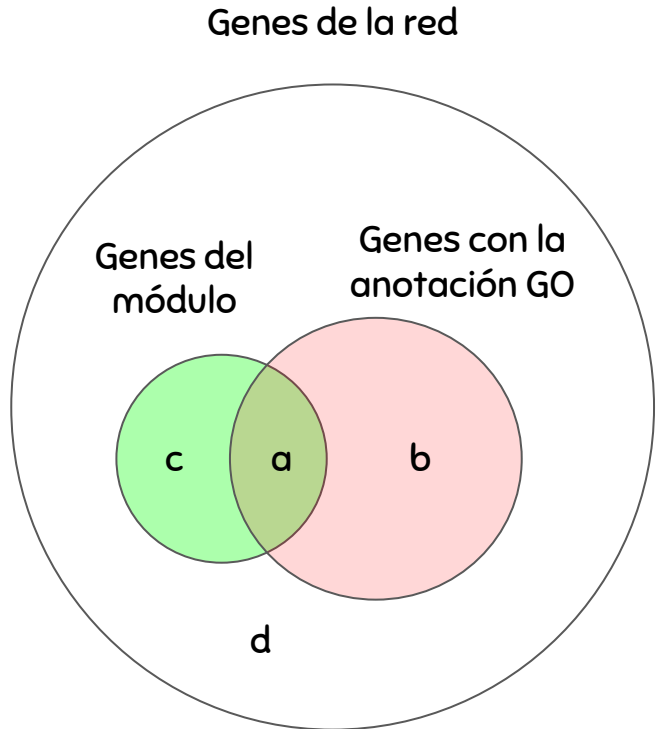


# Enriquecimiento de módulos



# Enriquecimiento de módulos

Para cada módulo evaluar:



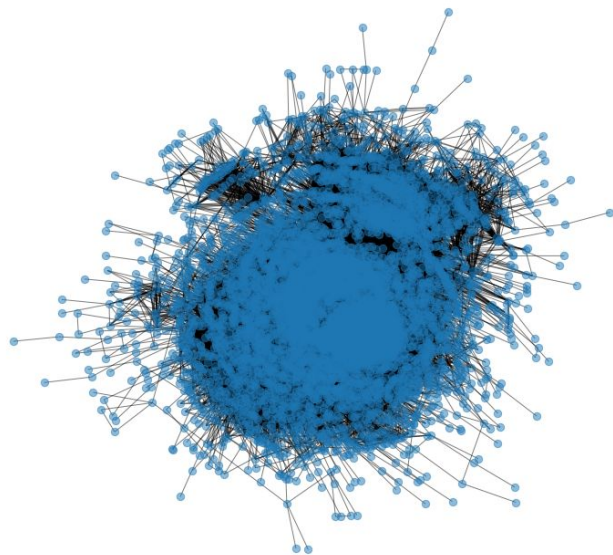
Test de Fisher:

**H0:** El módulo es una muestra aleatoria de la red.

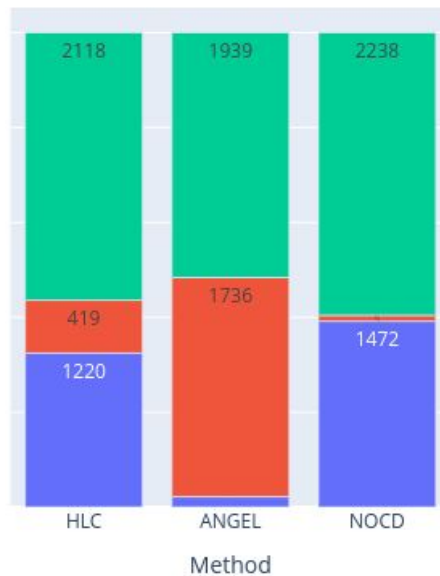
**H1:** El módulo tiene más genes anotados con el término GO que el esperado por azar.

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

# Enriquecimiento de módulos

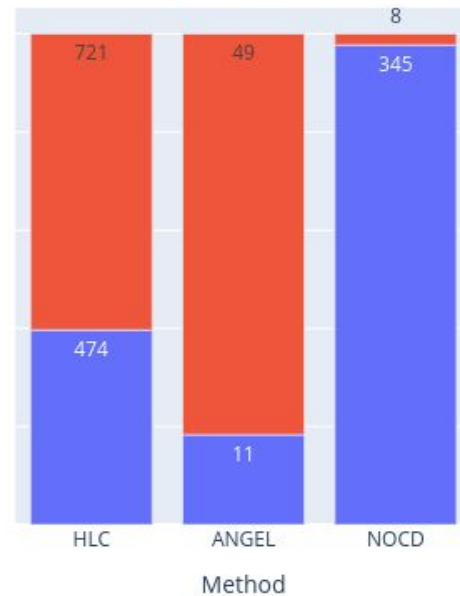


$|V| = 3,757$   
 $|E| = 246,360$



Overlapping

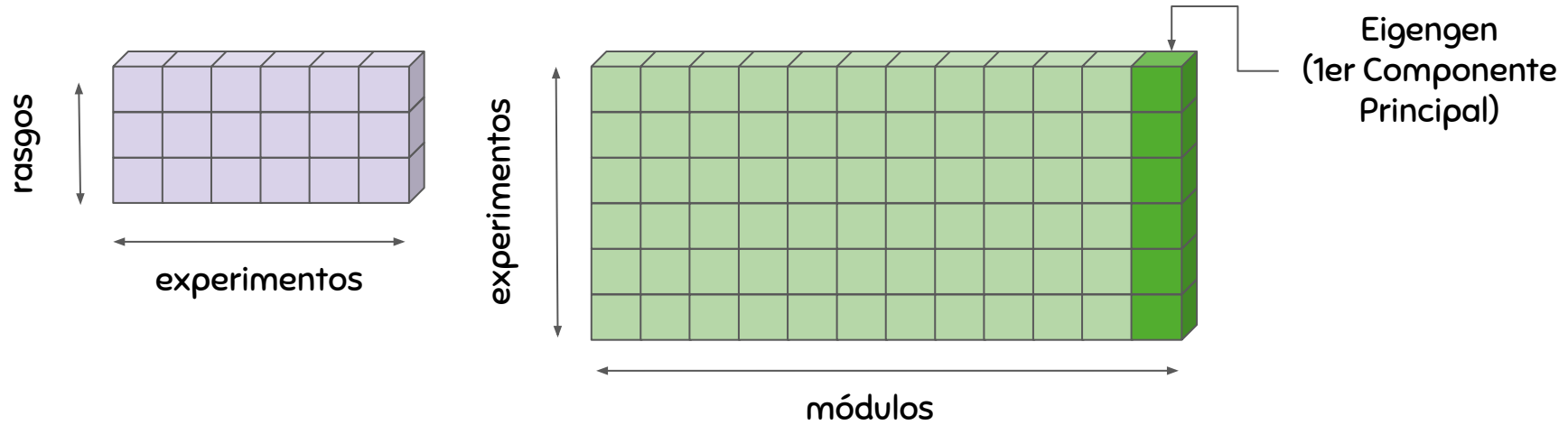
- Overlapping genes
- Non-overlapping genes
- Unclustered genes



Enrichment

- Enriched modules
- Non-enriched modules

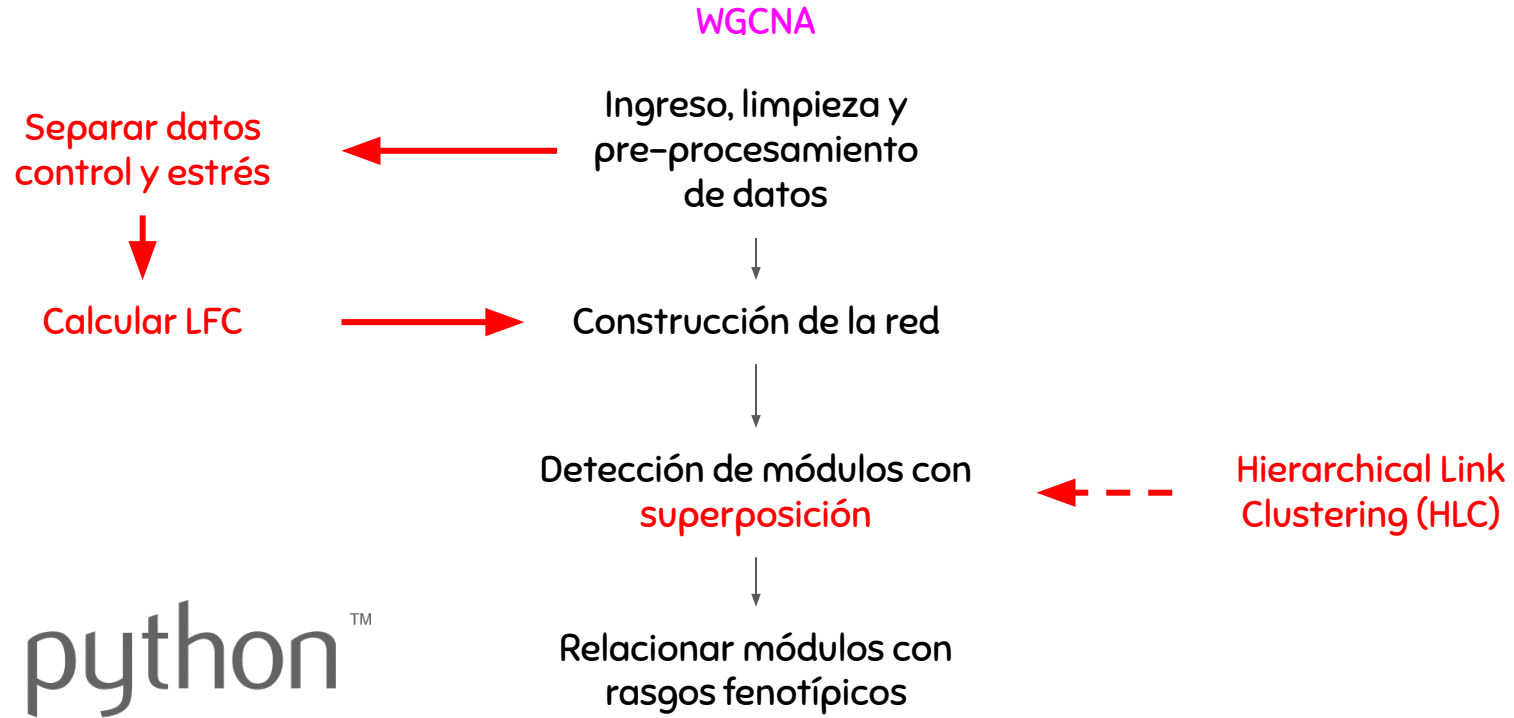
# Selección de módulos relevantes respecto a un rasgo fenómico



LASSO

$$\sum_{i=1}^m \left( y_i - \sum_{j=1}^c \alpha_j x_{ij} \right)^2 + \lambda \sum_{j=1}^c |\alpha_j|$$

# Resumen de la metodología



## METHODOLOGY ARTICLE

## Open Access

# Identifying stress responsive genes using overlapping communities in co-expression networks

Camila Riccio-Rengifo<sup>1\*</sup>, Jorge Finke<sup>2</sup> and Camilo Rocha<sup>2</sup>

\*Correspondence: camila.riccio@javerianacali.edu.co  
<sup>1</sup> Department of Natural Sciences and Mathematics, Pontificia Universidad Javeriana, Cali, Colombia  
Full list of author information is available at the end of the article

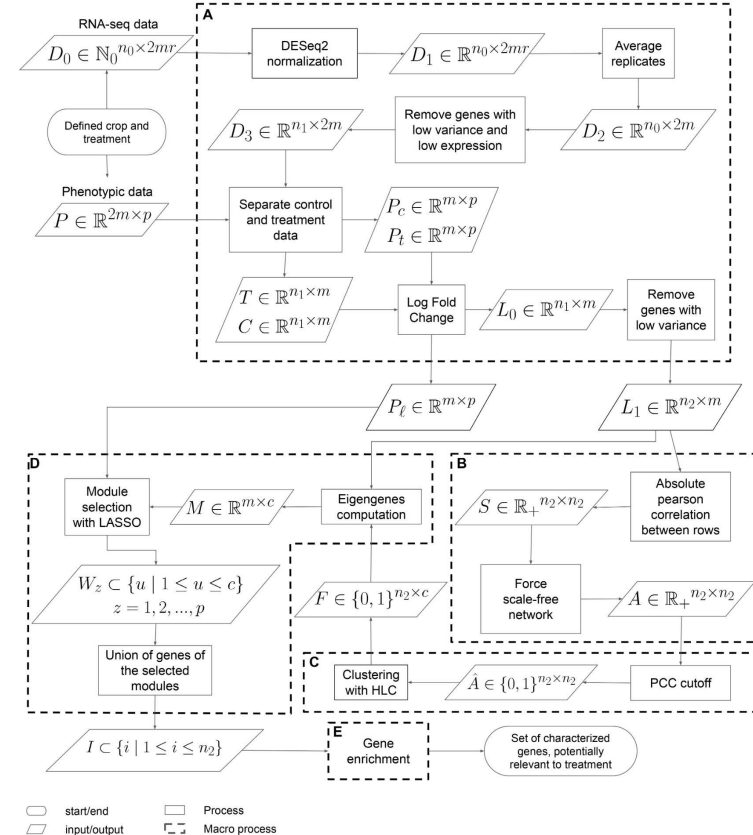
### Abstract

**Background:** This paper proposes a workflow to identify genes that respond to specific treatments in plants. The workflow takes as input the RNA sequencing read counts and phenotypic data of different genotypes, measured under control and treatment conditions. It outputs a reduced group of genes marked as relevant for treatment response. Technically, the proposed approach is both a generalization and an extension of WGCNA. It aims to identify specific modules of overlapping communities underlying the co-expression network of genes. Module detection is achieved by using Hierarchical Link Clustering. The overlapping nature of the systems' regulatory domains that generate co-expression can be identified by such modules. LASSO regression is employed to analyze phenotypic responses of modules to treatment.

**Results:** The workflow is applied to rice (*Oryza sativa*), a major food source known to be highly sensitive to salt stress. The workflow identifies 19 rice genes that seem relevant in the response to salt stress. They are distributed across 6 modules: 3 modules, each grouping together 3 genes, are associated to shoot K content; 2 modules of 3 genes are associated to shoot biomass; and 1 module of 4 genes is associated to root biomass. These genes represent target genes for the improvement of salinity tolerance in rice.

**Conclusions:** A more effective framework to reduce the search-space for target genes that respond to a specific treatment is introduced. It facilitates experimental validation by restraining efforts to a smaller subset of genes of high potential relevance.

**Keywords:** Stress-responsive genes, Co-expression network, Overlapping communities, Phenotypic traits, LASSO, Salinity, Rice, *Oryza sativa*



# Caso de estudio en arroz (*Oryza sativa*) bajo estrés salino

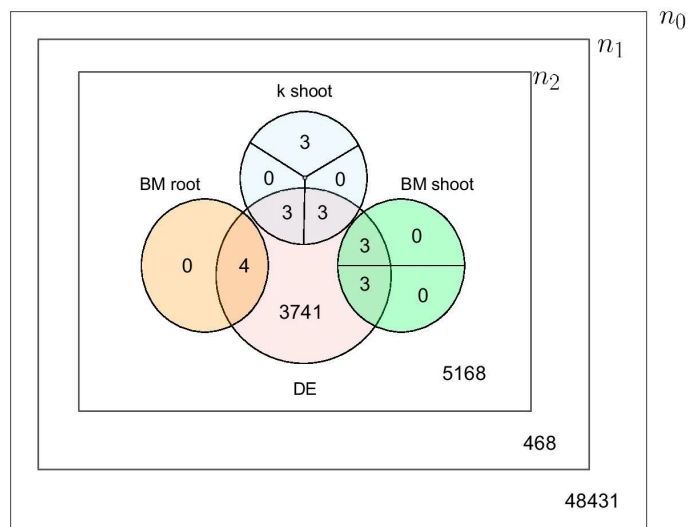


Table 1: Selected genes

| 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 | *   |   |
|                  |        | 13101.t05133 | LOC_Os01g58100 | *   |   |
| BM_root          | 5      | 13112.t02444 | LOC_Os12g27254 | *   |   |
|                  |        | 13112.t03421 | LOC_Os12g37260 | *   | * |
|                  |        | 13104.t03155 | LOC_Os04g35010 | *   | * |
|                  | 6      | 13108.t03971 | LOC_Os08g42310 | *   | * |
|                  |        | 13109.t01501 | LOC_Os09g17049 | *   |   |

genes reportados anteriormente por otros autores como relevantes en la respuesta a estrés salino

Muchas gracias!!

¿Preguntas?