1. Background

- During the environmental stimuli, plants switch on a complex gene expression modulation to cope with stress periods.
- Gene expression is controlled mainly at the transcriptional level where the interaction between transcription factor (TF) and cis-regulatory sequences in the upstream regions of genes is the core code.
- While much is known about TFs, cis regulatory elements (motifs) remain poorly understood.

2. Motivation

- Computational motif finding is a daunting task, however the emergence of high-throughput data, improved the accuracy of de Novo motif discovery.
- We used various algorithms in order to detect short DNA-motifs that are statistically over-represented upstream regions of co-regulated genes.

3. Materials and Methods

RNA-seq Workflow

- 8 RNA-Seq data
- Trimmomatic
  - Quality control
- Kallisto
  - Pseudo-alignment
- Sleuth
  - Gene expression profiling
- Clustering Analysis
  - Degs
  - Hierarchical
- De Novo motif Discovery
  - Input
    - Genes IDs / Modules
    - Promoters
    - Motifs discovery
    - Oligos / dyads
    - Motifs clustering
    - Grouping motifs sharing common DNA-Binding-Domain
    - Motifs comparison
      - Logo alignment and identification

4. Results

4.1 Clustering analysis

- Figure 1. Module analysis of the differentially expressed genes (DEGs). (A): Gene dendrogram based on dissimilarity measure (1-TOM). Each color is associated to one module. The colors are assigned based on the module size. (B): heatmap of DEGs pooled into 45 modules. Columns represent the conditions and rows correspond to modules with log2(1/TPM mean). Red and blue squares represent relative increase and decrease in TPM abundance.

4.2 Cis regulatory sequence prediction

- Figure 2. Over-represented DNA-motifs in two upstream regions from 150 DEGs of module 16. Black bars indicate the significance of the predicted motifs when compared to negative controls (gray bars).

5. References


6. Funding

N. Ksouri received a PhD fellowship awarded by the Government of Aragón. This work was supported by the MINECO and the Government of Aragón with projects AGL2017-83358-R and A99_17R, co-financed with FEDER and ESF, respectively.

PRESENTED AT XIV SYMPOSIUM ON BIOINFORMATICS, GRANADA 14TH-16TH NOVEMBER, JBI 2018