PLAZA workshop

4A - Comparing ChIP-Seq and DE genes using the Workbench

September 14-15, 2017

# Topics Covered

In this series of exercises you will learn how to combine different experimental gene sets using the PLAZA platform in order to answer specific basic biological questions related to gene regulation.

# Exercises

## Part 1: Use the PLAZA 4.0 Monocots Workbench to upload and compare different experimental gene sets

In this exercise, we will study which biological processes are controlled by the E2F transcription factor (TF), through the integration of ChIP-Seq and Differential Expression (DE) genes.

1. Using the PLAZA workbench, upload two experimental datasets which were generated for the Arabidopsis E2Fa TF (both files can be found in the PLAZA FTP folder <ftp://ftp.psb.ugent.be/pub/plaza/workshop/Elixir2017>)
   * E2Fa up-regulated genes from Naouar et al., 2008 (Quantitative RNA expression analysis with Affymetrix Tiling 1.0R arrays identifies new E2F target genes) – Table S1. Copy-paste the GeneIDs in a new workbench experiment “E2Fup” after using the Excel function Data – Filter:
     + Tiling3 log2 FC≥1 (column D)
     + Tiling3 qvalue<0.01 (column G)
   * E2F bound genes from Verkest et al., 2014 (A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification.) – Table S2. Copy-paste the GeneIDs (column I) in a new workbench experiment “E2Fchip” after using the Excel function Data – Filter
     + Method = TChAP
2. Phylogenetic profiling
   * For the “E2Fup” experiment, how many gene families are present having at least 10 genes present in this experiment? [View associated gene families – Sort table using header ‘#associated genes’]
   * What is the largest gene family present in this experiment which is species-specific for *A. thaliana*? What is its function? [View associated gene families – Sort table using header ‘ath’ and explore column ‘species’]
   * What is the largest gene family present in this experiment which is present in all plant species? What is its function and how many genes part of this gene family are up-regulated by E2F?
3. GO enrichment
   * Determine which Biological Processes are up-regulated by the E2F TF? Use a pvalue<0.001 cutoff to focus on the most important GO terms [View the GO enrichment – Explore both the “View Biological process enrichment graph.” as well as the “GO enrichment data table”]
   * How many genes have GO Molecular Function DNA helicase activity? [GO enrichment data table – Click subset ratio]
   * To how many gene families do these genes belong? [Toolbox Subset options - Create new experiment from this subset or add subset to existing experiment + View associated gene families]
   * For which gene family involved in DNA helicase activity are all 8 gene family members upregulated? What is the function of this gene family? [for the newly generated experiment containing the DNA helicase activity genes , run View associated gene families]
4. Overlap DE and ChIP genes to identify genes being both bound and regulated by E2F
   * How many genes are both bound and regulated by E2F? [Starting from E2Fup experiment – Compare with other PLAZA workbench experiments, select E2Fchip as second experiment – Compare genes]
   * Save these genes in a new Workbench experiments called “E2Fup\_chip”.
   * Perform GO enrichment and determine how many genes with DNA helicase activity are both bound and regulated by E2F.
5. Functional analysis E2Fup\_chip genes
   * Using “View the associated functional annotation”, determine how many genes are annotated with the GO term ‘transcription factor activity, sequence-specific DNA binding’. [Click num\_genes]
   * Save these genes in a new experiment called “E2Fup\_chipTF” [Click num\_genes and Create new Workbench experiment]
   * Which InterPro domains are most frequent in this dataset?