Adaptable integration, analysis and data-sharing of diverse multi-omic and low-throughput data: computational platforms from European consortia

To cite this version:

HAL Id: hal-02808978
https://hal.inrae.fr/hal-02808978
Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Adaptable integration, analysis and data-sharing of diverse multi-omic and low-throughput data: computational platforms from European consortia.

Distributed research consortia need:
- Shared access to versioned, structured and validated data
- Ability to query across all data and meta-data
- Flexibility to rapidly encode new questions

Data Overview
High-throughput data-sets include:
- Transcriptome (AT1H and AGRONOMICIST IIling arrays)
- Proteome (TRAD and label free MS-MS)
- Primary, secondary and lipid metabolites (GC-MS/MS)
- Epigenome methylation states

Additionally, a diverse range of low-throughput data are generated including:
- Visible phenotypes
- Enzyme activities
- Metabolite quantifications

Data Analysis:
- Growth, expansion and diurnal time in two environmental conditions
- Water deficit

Example data processing pipeline using database stored routines

Integrating metabolite quantifications (from spectrophotometry and MS) with enzyme activities

Having multiple integrated datasets available to all project partners through a project data warehouse enables and encourages a variety of data mining strategies. For example, metabolites and enzymes activities were analyzed using a variety of methods including profile plots, ANOVA, PCA, correlation matrices and with a variety of clustering methods. Availability of consistent summary statistics (biological replicate values/variances) together with individual sample measurements for data-sets is essential to allow the comparison of results based on a simple version of the data.

Integration of proteome, transcriptome and phenotype
Integration of quantitative protein and transcript abundances together with phenotype data enables a systems understanding of leaf development through growth, expansion and diurnal time in two environmental conditions.

Conclusions
- Molecular systems biology analysis of Arabidopsis leaf growth reveals adaptation to water deficit
- Systems-based analysis of growing Arabidopsis leaf reveals adaptation to water deficit
- Available at: www.agronomics.ethz.ch

Molecular Systems Biology (2012) 8

Published by Molecular Systems Biology