OposSOM-2.0.1-biomart

Please work through the documentation and add your comments to the bottom of this page, or email comments to support@cyverse.org. Thank you.

What is OposSOM?

Rationale and background

High-throughput technologies such as whole genome transcriptional profiling have revolutionized molecular biology and provide an incredible amount of data. On the other hand, these techniques pose methodological challenges simply by the huge and ever-increasing amount of data they produce. Researchers need adequate tools to extract the information content of the data in an effective and intelligent way. This includes algorithmic tasks such as data compression and filtering, feature selection, linkage to functional context, and proper visualization. In particular, this latter task is important because an intuitive visualization of massive amounts of data promotes quality control, the discovery of the data's intrinsic structure, functional data mining and finally the generation of hypotheses.

Introduction and overview

OposSOM bundles a series of sophisticated analysis methods with intuitive visualization options to study high-dimensional data with the special focus on gene-centered expression data. The algorithm transforms whole genome expression pattern of genes into a SOM coordinate system, which allows intuitive visualization of transcriptional activity of each sample in terms of mosaic portraits. This approach simultaneously searches for features which are differentially expressed and correlated in their profiles in the set of samples studied.

Loffler-Wirth H, Kalcher M, Binder H (2015). "oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on Bioconductor." Bioinformatics.

Wirth H, Loffler M, von Bergen M, Binder H (2011). "Expression cartography of human tissues using self organizing maps" BMC Bioinformatics.

A brief overview of the oposSOM workflow:

- Single gene expression values are clustered to metagenes using a self-organizing map.
- Based on the identified metagenes, visualizations (e.g. expression portraits), downstream sample similarity analyses (e.g. hierarchical clustering, ICA) and functional enrichment analyses are performed.
- The results are given within a separate folder and can be browsed using the summary HTML file.

Prerequisites

1. A CyVerse account. (Register for a CyVerse account at https://user.cyverse.org/.)
2. An up-to-date Java-enabled web browser. (Firefox recommended. If you wish to work with your own large datasets and upload them using icommands, Chrome is not suitable due to its issues in utilizing 64-bit Java.)

Mandatory requirements

Input file:

1. Normalized RNA-seq counts file - This is the text file that contains normalized counts that are generated using edgeR/DESeq2
2. Sample metadata file - This is the text file that contains the information of labels and colors of your samples.

Biomart parameters:

1. Dataset name (character): Name of the dataset. Used to name results folder and environment image (default: "Unnamed").
2. Species (Select your species from the list): Select your species from the list of species. If your species is not present in the species, we are currently on a different app where the user can provide their custom file

Optional parameters

1. Log transformation (Check if you want the app to log-transform your data before running opossom)

Starting an OposSOM-2.0.1-biomart job in the DE (Test run)

⚠ The example data for the test run is located at: /iplant/home/shared/iplantcollaborative/example_data/OposSOM

1. In the DE Apps window, search for and open OposSOM-2.0.1-biomart.
2. In the Analysis Name field:
   a. Change the name for your analysis (optional).
   b. Enter any comments (optional).
   c. In the Select output folder field, click Browse and navigate to the folder of your choice, or leave the default name of iplant/home/username/analyses.
   d. To retain copies of the input files in your analysis results output folder, click this check box (not recommended).
3. Click to open the Input files panel:
   a. For the Normalized RNA-seq counts file, click Browse and navigate to your normalized RNA-seq counts file. For the test run use - **Atha_normalized_counts.txt**
   b. For the Sample metadata file, click Browse and navigate to the metadata file for samples. For the test run use - **Atha_label_color.txt**
4. Click to open the Biomart parameters panel:
   a. Species (Select your species from the list):
5. Optional parameters:
   a. Log transformation (check this box)
6. Click Launch Analysis.
7. After successful completion of running of the app, the following files and figures are generated from the run.

Results Overview

The results comprise a variety of PDF documents with plots and images addressing the input data, supplementary descriptions of the SOM generated, the metadata obtained by the SOM algorithm, the sample similarity structures and also functional annotations.

```
+ /CSV Sheets
+ /Data Overview
+ /Geneset Analysis
+ /Sample Similarity Analysis
+ /Summary Sheets - Groups
+ /Summary Sheets - PATs
+ /Summary Sheets - Samples
+ /Supporting Maps&Profiles
+ Expression Potraits - alternative scales.pdf
+ Expression Potraits.pdf
+ Summary.html
```