

TECHNICAL NOTE

Open Access

FARO server: Meta-analysis of gene expression by matching gene expression signatures to a compendium of public gene expression data

Mieszko P Manijak and Henrik B Nielsen*

Abstract

Background: Although, systematic analysis of gene annotation is a powerful tool for interpreting gene expression data, it sometimes is blurred by incomplete gene annotation, missing expression response of key genes and secondary gene expression responses. These shortcomings may be partially circumvented by instead matching gene expression signatures to signatures of other experiments.

Findings: To facilitate this we present the Functional Association Response by Overlap (FARO) server, that match input signatures to a compendium of 242 gene expression signatures, extracted from more than 1700 *Arabidopsis* microarray experiments.

Conclusions: Hereby we present a publicly available tool for robust characterization of *Arabidopsis* gene expression experiments which can point to similar experimental factors in other experiments. The server is available at <http://www.cbs.dtu.dk/services/faro/>.

Findings

Often gene expression studies identify more differentially expressed genes than can readily be functionally analyzed in follow up experiments. Fortunately, some of these genes typically are annotated either directly or through sequence similarity to other annotated genes, helping the scientist to interpret the observed transcriptional response. In many cases the transcripts can even be annotated with controlled vocabularies like the Gene Ontology [1] or Kyoto Encyclopedia of Genes and Genomes [2], facilitating systematic annotation analysis. Numerous successful examples of this type of analysis are found in the literature [3]. However, this type of analysis depends on a high coverage of annotated genes that respond transcriptionally to stimulus. Alternatively, meta-analysis of gene expression data can identify experimental conditions that result in similar transcriptional responses. This type of analysis has been done in a series of organisms, for example in yeast [4], in Human cell lines [5] and in *Arabidopsis thaliana* [6]. A

key utilization of this type of analysis is in mutant, disease and drug characterization and matching.

Algorithm

Here we present a web-based implementation of the Functional Association by Response Overlap (FARO) [6] approach allowing comparison of a user provided gene expression signature, against a pre-compiled compendium. The approach matches the transcriptional response based on the identity and the response direction (over or under expressed) of the differentially expressed genes, ignoring the magnitude of the response. Previously, we demonstrated that this simplistic approach largely overcomes experimental biases and allows reliable comparison between experiments conducted under varying conditions in different laboratories and at the same time is simple enough to allow human interpretations of the results [6]. The approach gains most of its robustness from avoiding direct comparison between the measurements in different experiments and instead comparing outcomes of comparisons between contrasts contained within a experimental design. Hence, the between experiment similarity measure is the number of intersecting genes between lists of

* Correspondence: hbjorn@cbs.dtu.dk
Center for Biological Sequence Analysis, Department of Systems Biology,
Technical University of Denmark, Kemitorvet, Building 208, DK-2800 Lyngby,
Denmark

differentially expressed genes from two experiments. In addition, congruence of the gene expression response direction adds important insight into the nature of the signature comparison.

Testing

Since the web server implementation of the FARO approach based on the script prepared and tested in the original experiment [6] testing was restricted to processing data submitted via the web site by a user. The server was tested with data files containing mixed Affymetrix ATH1 probe identifiers and AGI locus identifiers as well as unknown identifiers and empty lines. Tests proved that the implementation handles all mentioned cases correctly.

Implementation

The FARO server allows the user to compare an expression signature against a compendium of signatures. The latter consists of 242 experimental signatures defined by the top 1209 differentially expressed genes. 1209 is the median number genes being significant across the compendium at significance level of 0.05. The experimental factors were extracted from more than 1700 public microarray experiments. The experimental factors represent various conditions and perturbations that are described in details on the server webpage. The server accepts a table containing at least 50 identifiers of either Affymetrix ATH1 probe set or AGI locus identifiers and compares the query gene list to the FARO compendium at the probe set level. For full functionality the input table must contain two columns containing identifiers and response direction as indicated by a signed number (possibly the log fold change), respectively. Optionally, the response direction may be indicated by "+" or "-", or alternatively left out entirely. In the latter case the congruence analysis is omitted.

The comparison returns a list of associated experimental factors that are filtered according to user specified thresholds. Two options are available for setting the threshold:

1. The Overlap percentage threshold returns associated factors that have overlap with the query list that equals or exceeds the indicated percentage. Here the percentage means the percentage of the query length.

2. The Rank threshold returns the r factors with the strongest overlap to the query list. Where the user specifies the rank (r) threshold.

The server output is given as a table of functional associated compendium factors. For each factor a list of overlapping TAIR annotated gene identifiers is given. The table of associated factors furthermore contains statistics on the overlap including the number of overlapping genes, the p-value (Fisher's exact test), the

response direction similarity (congruence) and the significance of the congruence estimated by binomial statistics. Furthermore, the FARO server visualizes the query experiment in the context of a dynamic graph displaying the association network of all compendium factors (as seen in Figure 1D). The dynamic graph enables the user to navigate through the association space to deepen the understanding of the individual factors meaning.

Availability and Requirements

Project name: FARO server

Project home page: <http://www.cbs.dtu.dk/services/faro/>

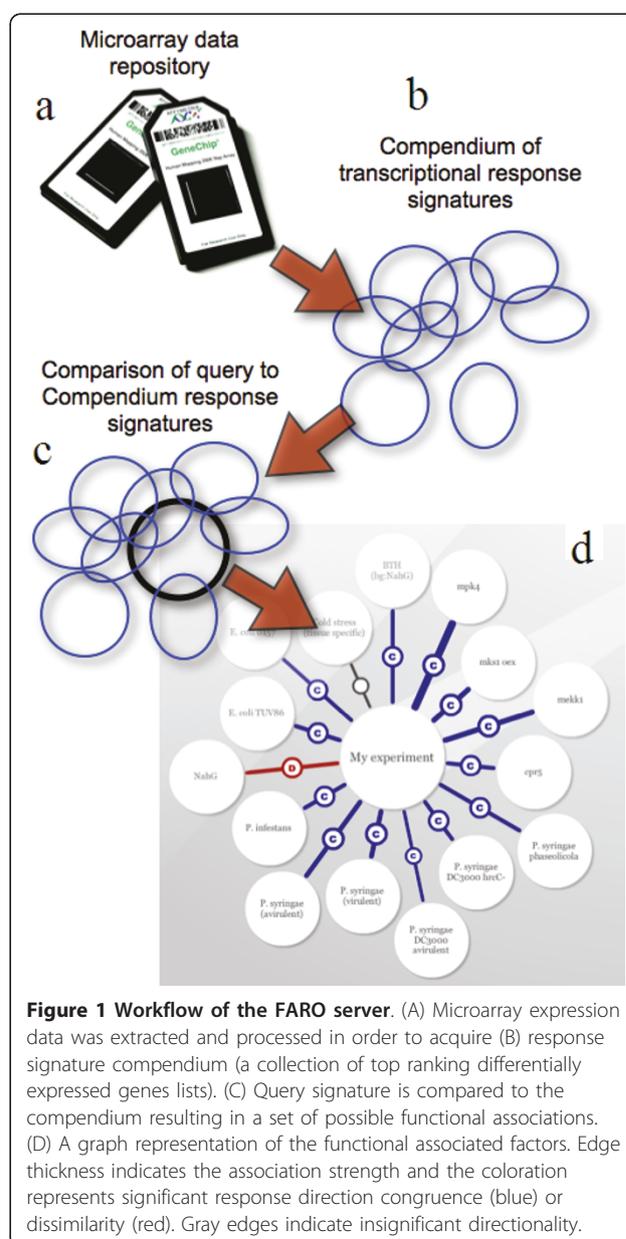


Figure 1 Workflow of the FARO server. (A) Microarray expression data was extracted and processed in order to acquire (B) response signature compendium (a collection of top ranking differentially expressed genes lists). (C) Query signature is compared to the compendium resulting in a set of possible functional associations. (D) A graph representation of the functional associated factors. Edge thickness indicates the association strength and the coloration represents significant response direction congruence (blue) or dissimilarity (red). Gray edges indicate insignificant directionality.

Operating system(s): Platform independent

Programming language: Perl

Other requirements: None

License: GNU GPL.

Any restrictions to use by non-academics: none

Acknowledgements

This work is supported by a grant from The Danish Agricultural and Veterinary Research Council (Multistress, SJVF)

Authors' contributions

HBN conceived and implemented the original algorithm, MPM adapted the original implementation for web server purposes and implemented the web server and performed tests. Both authors wrote and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 20 November 2010 Accepted: 11 June 2011

Published: 11 June 2011

References

1. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nature Genetics* 2000, **25**(1):25-29.
2. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Research* 2002, **32** Database: D277-280.
3. Rhee SY, Wood V, Dolinski K, Draghici S: **Use and misuse of the gene ontology annotations.** *Nature Reviews* 2008 **7**:509-515, Genetics 9.
4. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburty K, Simon J, Bard M, Friend SH: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102**(1):109-126.
5. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR: **The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.** *Science* 2006, **313**(5795):1929-1935, New York, N.Y.
6. Nielsen HB, Mundy J, Willenbrock H: **Functional Associations by Response Overlap (FARO), a functional genomics approach matching gene expression phenotypes.** *PLoS One* 2007, **7**:e676.

doi:10.1186/1756-0500-4-181

Cite this article as: Manijak and Nielsen: FARO server: Meta-analysis of gene expression by matching gene expression signatures to a compendium of public gene expression data. *BMC Research Notes* 2011 **4**:181.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

