

Poster presentation

Open Access

Using a literature-based NMF model for discovering gene functional relationships

Elina Tjioe¹, Michael Berry^{*2}, Ramin Homayouni³ and Kevin Heinrich⁴

Address: ¹Genome Science and Technology Graduate School, University of Tennessee, Knoxville, TN 37996, USA, ²Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA, ³Bioinformatics Program, Department of Biology, University of Memphis, Memphis, TN 38152, USA and ⁴Computable Genomix LLC, Bartlett, TN 38133, USA

Email: Michael Berry* - berry@eecs.utk.edu

* Corresponding author

from UT-ORNL-KBRIN Bioinformatics Summit 2008
Cadiz, KY, USA. 28–30 March 2008

Published: 8 July 2008

BMC Bioinformatics 2008, 9(Suppl 7):P1 doi:10.1186/1471-2105-9-S7-P1

This abstract is available from: <http://www.biomedcentral.com/1471-2105/9/S7/P1>

© 2008 Tjioe et al; licensee BioMed Central Ltd.

Background

The rapid growth of the biomedical literature and genomic information present a major challenge for determining the functional relationships among genes. Several bioinformatics tools have been developed to extract and identify gene relationships from various biological databases. In this study, we develop a Web-based bioinformatics tool called Feature Annotation Using Nonnegative matrix factorization (FAUN) to facilitate both the discovery and classification of functional relationships among genes. The algorithms of nonnegative matrix factorization (NMF) described in [1] are used. Both the computational complexity and parameterization of NMF for processing gene sets are discussed. FAUN is first tested on a small manually constructed 50-gene (50TG) collection that we, as well as others, have previously used [2]. The screenshots of FAUN feature classification and gene-to-gene correlation for the 50TG collection are shown in Figures 1 and 2. We then apply FAUN to analyze several microarray-derived gene sets obtained from studies of the developing cerebellum in normal and mutant mice. FAUN provides utilities for collaborative knowledge discovery and identification of new gene relationships from text streams and repositories (e.g. MEDLINE). It is particularly useful for the validation and analysis of gene associations suggested by microarray experimentation. FAUN tool is publicly available at <https://shad.eecs.utk.edu/faun>.

Discussion

For a preliminary assessment of FAUN feature classification, each gene in the 50TG collection was classified based on its most dominant annotated feature or based on some feature weight threshold. The FAUN classification using the strongest feature (per gene) yielded 90% accuracy. A FAUN-based analysis of a new cerebellum gene set has revealed new knowledge – the gene set contains a large component of transcription factors.

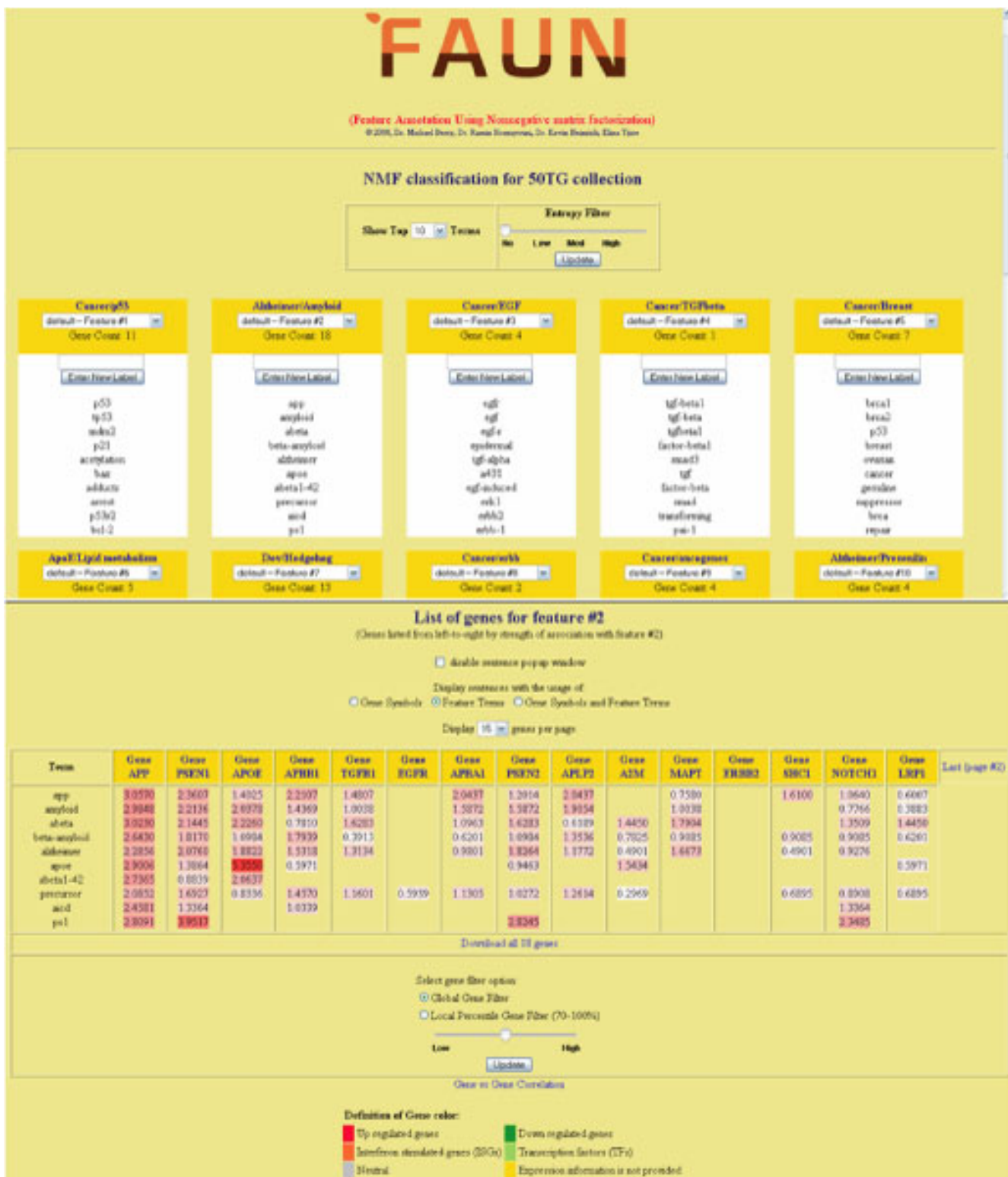


Figure 1
FAUN screenshot I. The upper window shows the classification features for 50TG collection and their top associated terms; the lower window shows use of dominant terms across genes highly associated with the user-selected feature.

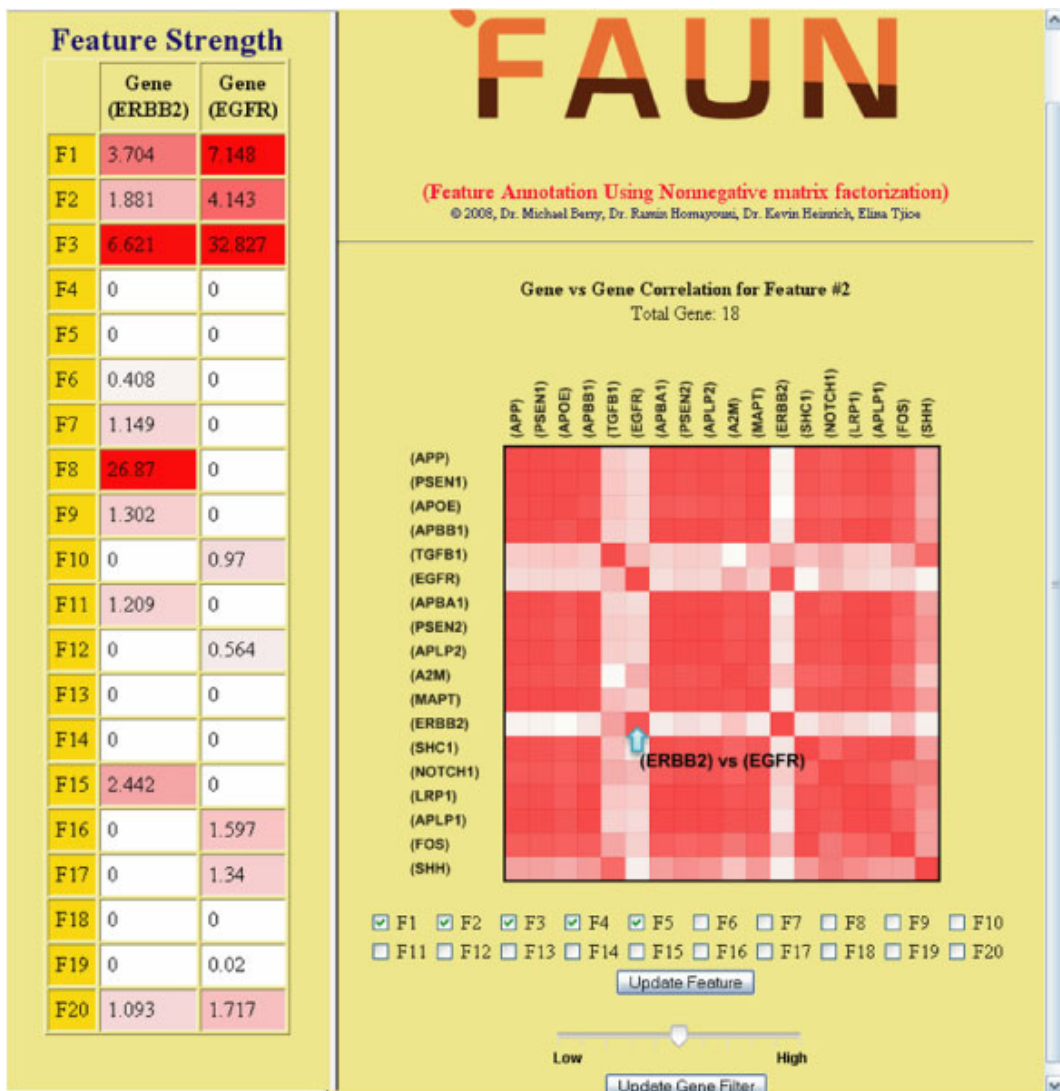


Figure 2
FAUN screenshot 2. The right window shows the correlation between genes highly associated with the user-selected feature; the left window shows the feature strength for the genes from the user-selected correlation cell (pointed by arrow).

Acknowledgements

This work is supported by an NIH-subcontract (HD052472) involving the University of Tennessee, University of Memphis, Oak Ridge National Laboratory, and the University of British Columbia.

References

- Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ: **Algorithms and Applications for Approximate Nonnegative Matrix Factorization.** *Computational Statistics & Data Analysis* 2007, **52(1)**:155-173.
- Homayouni R, Heinrich K, Wei L, Berry MW: **Gene clustering by latent semantic indexing of MEDLINE abstracts.** *Bioinformatics* 2005, **21(1)**:104-115.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."
 Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp