Omics Fusion – A Platform for Integrative Analysis of Omics Data

Benedikt G. Brink[1,2]**, Annica Seidel**[3]**, Nils Kleinbölting**[3]**, Tim W. Nattkemper**[1,2] **and Stefan P. Albaum**[1,3,*]

[1]International Research Training Group "Computational Methods for the Analysis of the Diversity and Dynamics of Genomes" GRK 1906, Bielefeld University, Germany

[2]Biodata Mining Group, Faculty of Technology, Bielefeld University, Germany

[3]Bioinformatics Resource Facility (BRF), Center for Biotechnology (CeBiTec), Bielefeld University, Germany

Summary

We present *Omics Fusion*, a new web-based platform for integrative analysis of *omics* data. *Omics Fusion* provides a collection of new and established tools and visualization methods to support researchers in exploring *omics* data, validating results or understanding how to adjust experiments in order to make new discoveries. It is easily extendible and new visualization methods are added continuously. It is available for free under:
`https://fusion.cebitec.uni-bielefeld.de/`

1 Introduction

With the advance of technology, generating data is no longer the limiting factor in biology. High-throughput experimental technologies transformed biological research from a relatively data-poor discipline to one that is data-rich. A key aspect of understanding and analyzing data is visualization. Analytical tools are very useful to solve a specific computational problem, whereas a powerful visualization can enable researchers to gain a mental model for their data and apply their biological knowledge.

Typically, molecular biology strives to understand and potentially optimize metabolic processes within a biological system such as a cell. Cells are living systems full of various functional molecules, which eventually determine the phenotype of the cells. Such molecules include mRNA transcribed from DNA, proteins translated from mRNA, and various metabolites generated by various enzymatic activities. Therefore, only analyzing the DNA sequences of genomes is not sufficient to obtain crucial information regarding the regulatory mechanisms involved in a cell's metabolism, e. g. responses to environmental factors and other stresses, or the production of metabolites. To understand the cell as one system, data from more than one *omics* discipline is needed [1].

---

[*]To whom correspondence should be addressed. Email: alu@cebitec.uni-bielefeld.de

2　Functionality

Omics Fusion is a platform for results of all kinds of data-rich high-throughput experiments, focusing on three classical fields in *omics*: transcriptomics, proteomics and metabolomics. It offers convenient data management, such as automated import of spreadsheets, along with connections to other platforms like EMMA [2], a system for the collaborative analysis and integration of microarray data, MeltDB [3], a software platform for the analysis and integration of metabolomics experiment data, or QuPE [4], a rich internet application for the analysis of mass spectrometry-based quantitative proteomics experiments. Here, we present the core functionalities of *Omics Fusion* in the order of a typical workflow.
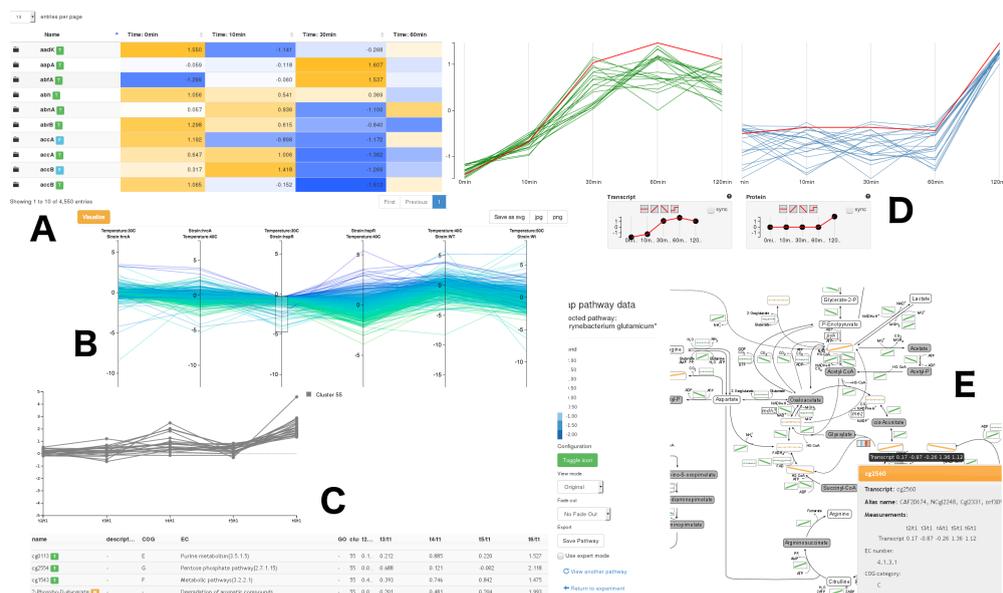

Figure 1: *Omics Fusion* **offers various visualization approaches for single- and polyomics data. This figure shows examples for: (A) Data management screen, (B) Parallel Coordinates, (C) Cluster Profiles, (D) Visual Profiling, (E) Pathway Map.**

Data manipulation

There are multiple tools available to manipulate data, ranging from simple but crucial normalization and filtering steps to transformation and missing value replacement. Data can also be enriched by querying other databases like KEGG, UniprotKb or NCBI/Entrez (Figure 1A).

Data analysis

Omics Fusion offers tools for descriptive statistics and distribution analysis to get an overview over the data, but also analysis of variance (ANOVA) for robust statistical testing. Furthermore, besides other classical methods like principial component analysis (PCA), *Omics Fusion* offers a hierarchical cluster analysis with automatic calculation of optimal cluster size and cluster grouping. This hierarchical clustering can be performed on data from multiple *omics* fields, grouping transcriptomic, proteomic and metabolomic data points with a similar signature. It facilitates the discovery of similar expression patterns throughout experiments from different *omics* fields (Figure 1C).

Visualization methods

Omics Fusion offers an increasing number of ways to explore and visualize your data. A few examples are box plots, scatter plots, parallel coordinates (Figure 1B) or parallel sets. Beyond that, users can choose from a number of custom visualizations that introduce new ways to look at data from different *omics* disciplines. An example for that is a method termed "visual profiling", which allows users to manually draw an arbitrary abundance profile and find all transcripts, proteins, or metabolites matching that prototype (Figure 1D).

Pathway map

The pathway viewer component implemented within the *Omics Fusion* framework enables the mapping of complete *polyomics* datasets on metabolic pathway images. Customized pathway maps can be imported as SVG-files and the interactive visualization provides different levels of highlighting important aspects of the data, including stylized icons for different expression patterns or a heatmap representation (Figure 1E).

3 Implementation

Omics Fusion is a web service based on the Spring MVC framework, written in Java and JavaScript, and is the successor to ProMeTra [5], an open source framework that provides visualization methods for polyomics datasets and uses static SVG graphics to generate pathway maps. *Omics Fusion* however focuses on interactivity. By using modern JavaScript techniques, new means for creating powerful visualizations are available, for example allowing users to click on objects, mark certain areas, zoom in or out, etc. This enables them to explore data without prior knowledge about it. Users can apply different filters, evaluate different clustering methods, search for patterns of co-regulated or differentially expressed transcripts, proteins and metabolites, or discover pathways that are affected by a certain condition.

To achieve this level of interactivity, the various visualizations are generated using the powerful *D3* library for Javascript [6]. This makes it easy for developers to extend the platform and add new visualizations in the future. It is also possible to call external scripts (Python, R, etc.), providing even more ways to customize a workflow.

Tasks that require lots of computing power, such as clustering algorithms, can be executed on the server side, and make use of the computational power of the Bioinformatics Resource Facility (BRF) in Bielefeld.

4 Conclusion

Omics Fusion is an extendible, web-based platform for the integrative analysis of *omics* data. It provides powerful analysis tools, including established methods for analyzing and visualizing single *omics* data, as well as new features for an integrative analysis of data from multiple *omics* disciplines. Furthermore, compared to other tools that offer a similar level of interactivity, *Omics Fusion* does not solely focus on networks and pathways (e. g. Cytoscape [7] or iPEAP [8]), nor is it limited to a specific organism.

New features are added continuously and will make *Omics Fusion* a comprehensive platform for *omics* data exploration, providing means to biologists for an easier and more efficient data analysis process. This can potentially provide new insights into biology, or at least simplify gathering of information and analyzing data from experiments with one or multiple *omics* approaches and hence accelerating the process of research.

Acknowledgements

References

[1] W. Zhang, F. Li and L. Nie. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156(Pt 2):287–301, 2010.

[2] M. Dondrup, S. P. Albaum, A. Goesmann et al. EMMA 2–a MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC bioinformatics*, 10:50, 2009.

[3] N. Kessler, H. Neuweger, A. Goesmann et al. Meltdb 2.0–advances of the metabolomics software system. *Bioinformatics*, 29(19):2452–2459, 2013.

[4] S. P. Albaum, H. Neuweger, A. Goesmann et al. Qupe–a Rich Internet Application to take a step forward in the analysis of mass spectrometry-based quantitative proteomics experiments. *Bioinformatics*, 25(23):3128–3134, 2009.

[5] H. Neuweger, M. Persicke, A. Goesmann et al. Visualizing post genomics data-sets on customized pathway maps by ProMeTra-aeration-dependent gene expression and metabolism of Corynebacterium glutamicum as an example. *BMC systems biology*, 3:82, 2009.

[6] M. Bostock, V. Ogievetsky and J. Heer. D^3: Data-Driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

[7] P. Shannon, A. Markiel, T. Ideker et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

[8] H. Sun, H. Wang, Q. Liu et al. iPEAP: integrating multiple omics and genetic data for pathway enrichment analysis. *Bioinformatics*, 30(5):737–739, 2014.