

Poster B-14
Contextual Analysis and
Visualization of High Throughput
Experimental Data



Authors:

Nenad Bartonicek (*Bioinformatics Group, Faculty of Science, Zagreb University, Horvatovac 102a, 10000 Zagreb, Croatia.*)

Igor Segota (*Department of Physics, Faculty of Science, Zagreb University, Bijenicka cesta 32, 10000 Zagreb.*)

Kristian Vlahovicek (*Bioinformatics Group, Faculty of Science, Zagreb University, Horvatovac 102a, 10000 Zagreb, Croatia.*)

Short Abstract: MADNet (MicroArray Database Network) is a user-friendly data mining tool for rapid analysis of diverse biological data such as microarray, phage display or even metagenomes. It presents biological information in the context of metabolic and signaling pathways, gene orthologs, transcription factors and drugs through minimal user input.

Long Abstract:

High throughput technologies allow scientists to simultaneously study expression of thousands of genes, or even entire genomes. Use of microarrays started as an appealing, but unreliable and expensive tool for snapshots of cell conditions, accompanied by a rigorous statistical and biochemical follow-up. In the last decade most of the problems concerning precision, accuracy and reproducibility have been accounted for through technological and algorithmic improvements. This resulted in a widespread presence of microarrays in research or even diagnostic laboratories. Still, the final step in microarray analysis, the biological interpretation of data, remains a lingering issue. A researcher with insufficient statistical or bioinformatic knowledge may find himself/herself in the forest of information too large to grasp or manipulate, with every gene descriptor leading to many different knowledge databases.

We have developed MADNet, MicroArray Database Network, an easy to use web-based contextual analysis and visualization tool for data mining. MADNet provides the user with several advantages, which minimize the knowledge, effort and time required to gain insight into a large amount of data.

First, user does not need to have any prior knowledge of the existent microarray formats and gene nomenclature. An emphasis was placed on intelligent pattern recognition and screening of user's data input through multiple databases. MADNet will attempt to spare the user of unnecessary input, while recognizing file format, species of the tested organism and gene annotation type. Due to data limitations and the unavoidable possibility of ambiguous detection, MADNet allows user interventions in the automatic detection process.

Second, MADNet provides a systems biology approach to complex research problems in a user friendly interface. The tool integrates several types of biological information from existent databases; NCBI nucleotide and protein databases, gene ortholog databases, metabolic and signalling pathway databases, as well as transcription regulation databases. After data input, MADNet finds the metabolic and signalling pathways with the most significant expression changes, according to various significance scores (p-value of binomial distribution, Z-score). Information from other integrated databases is then presented in the

context of the maps of metabolic and regulatory pathways in a clear and organized manner, also available in a graphical and tabular format.

Furthermore, special attention has been given to the processivity of the MADNet analysis. Even though MADNet can analyze 50.000 genes in a couple of minutes (mostly depending on the upload time), user's query is stored in the form of sessions, which further diminishes time required for repetitive calculations. User can also perform simultaneous analysis of several different pathways, or same pathways from different experiments. In addition, MADNet provides the user with the automatically generated complete reports of the analyzed data in excel and PDF formats, which is especially convenient for high throughput analysis of experiments.

Finally, MADNet is not confined only to microarrays and existent integrated databases, but can be used to analyze gene information from different experiment types and database resources. Due to its modular structure, MADNet can mine information from diverse origins such as tissue and protein microarrays, SAGE and phage display, the only requirement being the data columns with gene identification numbers, change descriptors and optional statistical significance. Since it is also highly processive, it can also analyze whole metagenomes, which is especially interesting in the context of recent progress of whole environment sequencing. Furthermore, the research team is ready to promptly enter database of user's interest and make it available through the MADNet interface, in order to answer to personalized needs of different research communities.

MADNet is located on www.bioinfo.hr/madnet and is freely available for testing and usage in non-commercial purposes.