

## Micro-Analyzer: a tool for automatic pre-processing of multiple Affymetrix arrays

P.H. Guzzi<sup>✉</sup>, M. Cannataro

Department of Medical and Surgical Sciences, University "Magna Græcia" of Catanzaro, Catanzaro, Italy

### Motivations

A current trend in genomics is the investigation of cell mechanism using different technologies in order to explain the relationship among genes, molecular processes and diseases on a different scale. For instance, the combined use of expression arrays and SNP arrays has been demonstrated as an effective instrument in clinical practices [1,3,4]. Consequently, in a single experiment different kind of microarrays may be used, resulting in the production of different types of binary data (images and raw data). The analysis of microarray data requires an initial preprocessing phase of raw data that makes them suitable for use on existing platforms, such as the TIGR M4 Suite. An additional challenge to be faced by emerging data analysis platforms is the ability to treat in a combined way such different microarray data coupled with clinical data. In fact resulting integrated data may include both numerical and symbolic data (e.g. gene expression and SNPs regarding molecular data), as well as temporal data (e.g. the response to a drug, time to progression, survival rate, etc., regarding clinical data). Raw data preprocessing is a crucial step in analysis but is often performed in a manual and error prone way using different software tools. Thus novel, platform independent, and possibly open source tools enabling the semi-automatic preprocessing and annotation of different microarray data are needed.

### Methods

The paper presents Micro-Analyzer (Microarray Cel file Summarizer), a cross-platform tool for the automatic normalization, summarization and annotation of Affymetrix expression and SNP data binary data. It represents the evolution of the  $\mu$ -CS tool, extending the preprocessing to SNP arrays that were not allowed in  $\mu$ -CS [2]. Using the tools made available by Affymetrix (e.g. apt-summarize and apt-genotype), the user needs to download from the Affymetrix web site the right preprocessing and annotation libraries, then needs to manually invoke such tools to obtain

preprocessed data and then has to import them into an external data analysis tools, e.g. TMEV. This approach presents numerous drawbacks, among those the need to manually perform all these tasks and the possibility to use the wrong or older libraries, obtaining wrong results, finally, data must be manually imported into analysis tools. To reduce such drawbacks we propose Micro-Analyzer. Micro-Analyzer is based on a client-server architecture. The Micro-Analyzer client is provided as a Java standalone tool and enables users to read, preprocess and analyse binary microarray data (gene expression and SNPs). It avoids: (i) the manual invocation of external tools (e.g. the Affymetrix Power Tools), (ii) the manual loading of preprocessing libraries, and (iii) the management of intermediate files. The Micro-Analyzer server automatically updates the references to the summarization and annotation libraries, hiding to the user the location of libraries and automatizing the process of updating such libraries when new versions of the microarray are released. By using Micro-Analyzer the user may preprocess both data using a single tool, retaining the advantage of storing in a single way both preprocessing results and metadata.

### Results

Micro-Analyzer users can directly manage Affymetrix binary data without worrying about locating and invoking the proper preprocessing tools and chip-specific libraries. Moreover, users of the Micro-Analyzer tool can load the preprocessed data directly into the well-known TM4 platform, extending in such a way even the TM4 capabilities. Consequently, Micro Analyzer offers the following advantages: (i) it reduces possible errors in the preprocessing and further analysis phases, e.g. due to the incorrect choice of parameters or due to the use of old libraries, (ii) it enables the centralized pre-processing of different arrays, (iii) it may enhance the quality of further analysis by storing the information about the preprocessing steps.

**Availability**

<http://sourceforge.net/projects/microanalyzer/>

**References**

1. André Koschmieder, Karin Zimmermann, Silke Trüßl, Thomas Stoltmann, Ulf Leser Tools for managing and analyzing microarray data Briefings in Bioinformatics, Vol. 13, No. 1. (01 January 2012), pp. 46-60, doi:10.1093/bib/bbr010
2. Pietro Hiram Guzzi, Mario Cannataro: mu-CS: An extension of the TM4 platform to manage Affymetrix binary data. BMC Bioinformatics 11: 315 (2010)
3. [www.affymetrix.com](http://www.affymetrix.com)
4. Walker BA, Leone PE, Jenner MW, Li C, Gonzalez D, Johnson DC, Ross FM, Davies FE, Morgan GJ.