**DBComposer**: An R package for integrative analysis and management of gene expression microarray data

Lingjia Kong, Kaisa-Leena Aho, Kirsi Granberg, Christophe Roos, Reija Autio

Abstract—**DBComposer** is an R package with a graphical user interface (GUI) to analyze and integrate gene expression microarray data. With **DBComposer**, the data can be easily annotated, preprocessed and analyzed in several ways. **DBComposer** can also serve as a personal expression microarray database allowing users to store multiple datasets together for later retrieval or data analysis. It takes advantage of many R packages for statistics and visualizations, and provides a flexible framework to implement custom workflows to extend the data analysis capabilities.

As the high-throughput gene expression measurement technologies, such as microarrays, have become widely available and more affordable, researchers today have more choices in their research and experimental design. However, the speed at which this data is translated into knowledge is slower than the rate of data generation. Meanwhile a wealth of advanced software for statistical analysis of microarrays has been developed as R [1] packages in Bioconductor [2]. Often these packages have only command-line interfaces, which build a barrier for non-computational scientists to analyze the data. Therefore, we have developed the R package - **DBComposer**, which provides an advanced set of microarray analysis algorithms within an easy-to-use GUI, and helps users discover valuable information from microarray experiments without spending too much time learning R.

**DBComposer** was developed in R and Bioconductor, while the GUI was implemented with the RGtk2 toolkit [3]. The operations of **DBComposer** are grouped into three major panels: **Import**, **Select** and **Analyze** (Fig 1). Each panel has several utility functions. In the **Import** panel, raw microarray data from Affymetrix, Illumina and one channel data from Agilent can be read in, preprocessed, and normalized using the methods summarized in Table I. The preprocessed data can be then stored, which activates the **Select** panel where users can browse and select the data for the **Analyze** panel. Any user-selected samples within a jointly normalized set can be analyzed with a selection of computational tools in the **Analyze** panel. By repeating **Import** and **Store**, users can create a personal microarray database. It is then possible to test hypotheses using all the collected data with the analysis tools from the **Analyze** panel. For example, datasets from public databases or the user’s own datasets can both be incorporated into **DBComposer**, and the expression of a set of interesting genes can be queried across multiple experiments.

One of the main objectives of developing **DBComposer** is maximizing the re-usability of microarray data by simplifying the data integration and analysis. The package offers a simple point-and-click interface to utilize state-of-art statistical methods to preprocess and analyze microarray data. Users can thus easily manage their own microarray database in **DBComposer**, allowing the seamless integration of their own data with data from public databases.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Preprocessing and Normalization</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affymetrix</td>
<td>Robust Multi-array Average (RMA) [4]</td>
<td>Affymetrix default CDF</td>
</tr>
<tr>
<td>Illumina</td>
<td>background corrected using lumi package [6] and quantile normalization</td>
<td>lumiHumaAll.db</td>
</tr>
<tr>
<td>Agilent</td>
<td>background corrected with the normexp method [7] and quantile normalization</td>
<td>biomaRt [8]</td>
</tr>
</tbody>
</table>

One of the main objectives of developing **DBComposer** is maximizing the re-usability of microarray data by simplifying the data integration and analysis. The package offers a simple point-and-click interface to utilize state-of-art statistical methods to preprocess and analyze microarray data. Users can thus easily manage their own microarray database in **DBComposer**, allowing the seamless integration of their own data with data from public databases.

### References