

# FlexArray

Statistical Data Analysis Software for Gene Expression Microarrays, Made With Life Scientists in Mind

## Introduction

One of the greatest sources of confusion in the analysis of microarray experiments is that data analysis of differential expression is non-standard and the variation between methods can be non-trivial. Unfortunately, it has been difficult for most life scientists to run, compare and contrast methods, in part because they are available in different formats (e.g. Microsoft Excel add-ins, stand-alone packages, R language). The Bioconductor R package, which contains the most comprehensive set of methods, has not resolved this problem because it has proven challenging to use for many end-users. This lack of an easy-to-use and extensible software package with a broad set of methods has also slowed down the transfer of knowledge between algorithm developers and algorithm users.

The FlexArray software is our proposed answer to these concerns. It is a Windows software package that has been created to simplify data analysis of expression microarrays without sacrificing the comprehensiveness. The target audience of FlexArray is life scientists. The software currently supports Affymetrix GeneChips and Illumina BeadChips expression arrays. It is well suited for small- to medium-size projects. This software can be found at <http://genomequebec.mcgill.ca/FlexArray>.

FlexArray provides answers to the following questions:

- How do I assess the quality of the data?
- What is a typical analysis workflow?
- What analysis methods can I use?
- How do I go about using the methods I have selected?
- How does my choice of methods impact the outcome of the analysis?
- How do I make sure the analysis is being kept track of?
- How do I repeat an analysis for another experiment?
- How do I share the methods I used with my colleagues?

FlexArray is a gene list generating tool, not a data mining tool. It is also an exploration tool for analysis methods and algorithms.

In a sense, FlexArray is a Bioconductor “wrapper”. It taps directly into the R programming language, especially the renowned and well-respected Bioconductor package for normalization algorithms, statistical tests, and other complex data-processing tasks.

FlexArray is easy to use:

- It has a modern, intuitive, and highly interactive graphical user interface.
- It contains many useful plots for the visualization of data.
- The analysis is driven by context-sensitive menus.
- It is comprehensive, including the most popular methods of analysis and visualization, and several promising less-known methods.

FlexArray addresses a number of common analysis workflow problems:

- Data import is straightforward and flexible.
- Annotations can later be imported from text files.
- The entire analysis is visualized in the form of a “pipeline” to indicate the data flow.
- Methods of the same type (e.g. normalization, testing) can be applied in parallel, and their results stored in a single project file. This way the outputs of different methods can be compared and contrasted.
- Detailed analysis history is maintained, making it possible not only to retrieve the specifics of the way a method was applied in the past, but even to repeat the exact same analysis should the need arise.
- It is possible to save a particular analysis workflow to disk as a “protocol” and to apply it in the future to analyze different datasets.

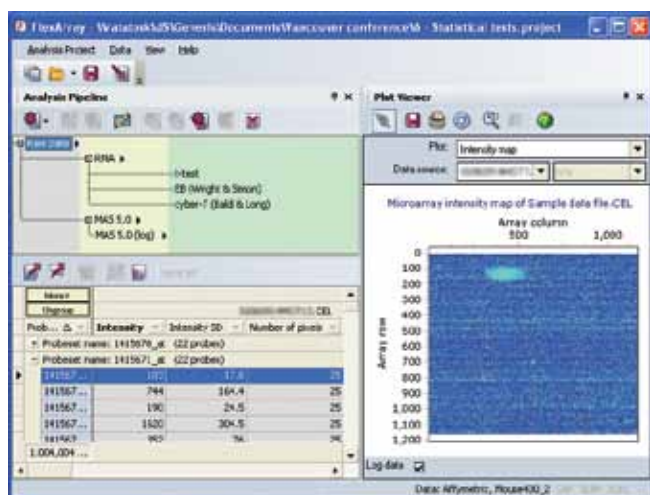
FlexArray can easily be extended by end-users to include new analysis methods and plots. This feature could be used by bioinformaticians to integrate their own algorithms and data processing methods directly within FlexArray.

Currently, FlexArray has over 350 registered users, 290 of whom have downloaded the software. FlexArray has 95 users at McGill University, close to 140 users elsewhere in Canada, and 120 international users.

FlexArray is a project funded by Génome Québec. The originator and long-time supervisor of the project was Professor Robert Nadon from McGill University. FlexArray is free for academic and government researchers.

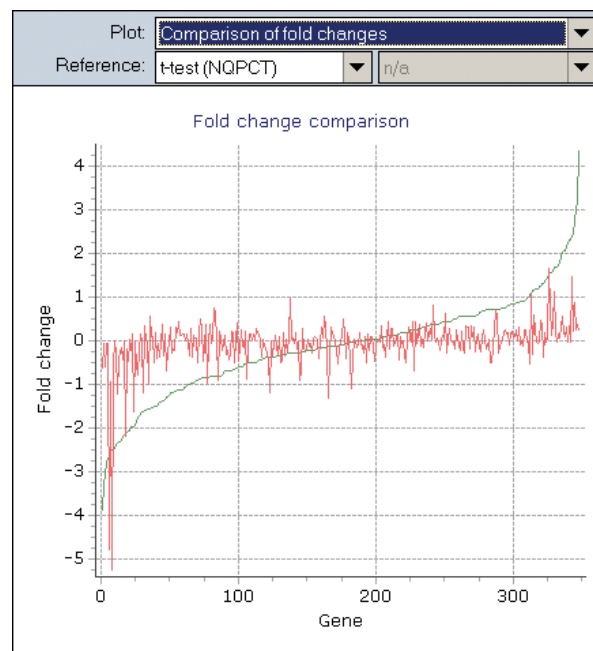
# FlexArray

Statistical Data Analysis Software for Gene Expression Microarrays, Made With Life Scientists in Mind



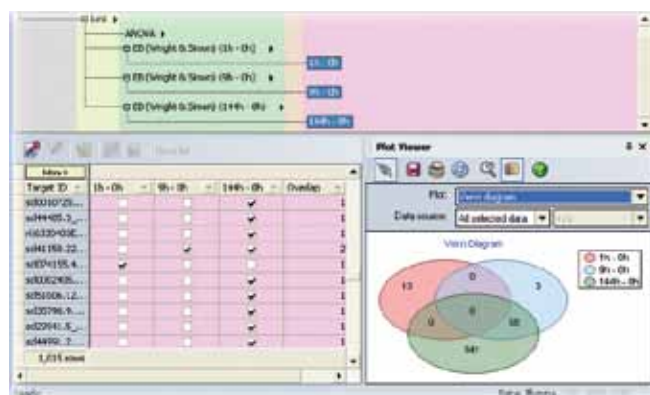
**Figure 1. Overview of FlexArray's user interface.**

The analysis pipeline shows the methods that have been applied to the raw data. Color-coded vertical bands group data by category (e.g. raw, normalized, statistical test output). The data table displays data from the currently selected analysis pipeline node, in this case – raw Affymetrix data, grouped by probe set. The plot viewer displays a raw intensity map of one of the arrays, with a smudged area.



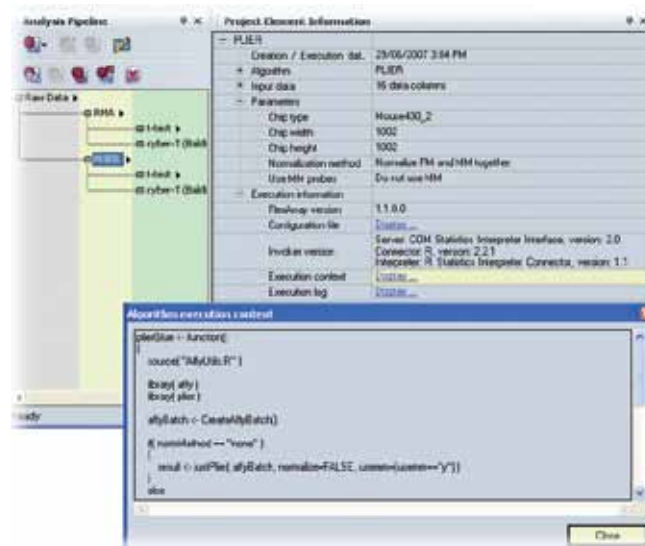
**Figure 3. Comparing data from different technologies in FlexArray.**

The green line represents fold changes obtained from a t-test performed on protein counts (NQPCT data), and the red line represents fold changes for corresponding genes obtained from a t-test performed on gene expression values.



**Figure 2. Overview of FlexArray's user interface.**

Managing and comparing gene lists in FlexArray. The data comes from a multiple-point time series experiment, with 3 replicates per time point. The analysis pipeline shows that three gene lists (selected nodes in the pink band) were generated from comparisons between three separate time points and the zero time point. The statistical test used was the Wright and Simon's implementation of the Empirical Bayes method. The Venn diagram offers a rough visual comparison of the three lists, while the data table allows for a more detailed analysis (a check box indicates that a particular target ID is present on a given gene list).



**Figure 4. Analysis history in FlexArray contains information about a method executed on the data, date and time of execution, method's parameters, version of FlexArray and R, the R code used to execute the method, a full execution log, and more.**

