A Glance at DNA Microarray Technology and Applications

Amir Ata Saei¹ and Yadollah Omidi¹,²*

¹Research Center for Pharmaceutical Nanotechnology, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
²Ovarian Cancer Research Center, School of Medicine, University of Pennsylvania, Philadelphia, USA

Introduction

Proteins, the amazing molecules of nature are almost involved in any activity in the cells from production of energy and biosynthesis of all component macromolecules to the maintenance of cellular architecture, and the ability to act upon intra- and extracellular stimuli. Each cell within an organism contains DNA which is crucial to produce the entire repertoire of proteins to cover the needs of an organism. The human genome project has determined the sequences that make up the human genome (3 billion base pairs). The number of human genes is estimated to be 30,000 to 100,000. It is now well known that the complementary sequences of most mRNA molecules could be transcribed in any biological process. Only a portion of these genes are expressed and turned into functional proteins. However, some of the genes expressed in a single cell are likely to be present in all cells because they serve routine functions necessary for maintaining life in all cells and are called “housekeeping” genes. Other proteins serve specialized functions and are only required in particularly differentiated cell types for example, heart cells or neurons. Each cell’s function determines the genes that have to be expressed in that specific type of cell.

Activities of a cell are highly controlled by cellular networks or more clearly the protein concentration. When any kind of change is imposed to the cell system these cellular networks and regulatory mechanisms become active and thus can be more readily detected. Global knowledge or a fingerprint of the transcriptional state could provide a wealth of information useful to biologists. This knowledge can be used in prediction of unknown genes functions, identification of biomarkers, target discovery, accurate diagnostics, development of prognostic tests and disease sub-class determination. At the very least, comparison of gene expression patterns in normal and pathological cells could provide useful diagnostic information and help identify genes that would be reasonable targets for therapeutic intervention (Afshari et al. 1999; Bednar 2000; Chin and Kong 2002;
Saei and Omidi

Dixon 2002; Dudda-Subramanya et al. 2003). Schematic steps of DNA microarray technology is shown in Fig. 1.

Microarray technology is a recent hybridization-based technique (gets back to 1990s) that allows simultaneous analysis and consequently estimation of an abundance of many nucleic acid species. Microarray has perhaps been so far, the most important revolution in functional genomics.

As shown in Fig. 2, this technique involves robotic placement of individual, pure nucleic acid species on a glass surface. The entire complement of transcript mRNAs present in a particular cell type is extracted from cells and then a fluorophore-tagged cDNA representation of the extracted mRNAs is made in vitro by an enzymatic reaction termed reverse transcription. Then multiple fluorescently labeled nucleic acids are hybridized to the array, spots are detected and fluorophore-tagged hybrids are measured across the array with a scanning confocal microscope. The microarray technology is particularly useful for comparing the mRNAs from two cell types or two treatments.

Image capturing and analysis plus primary data extraction
Fluorophore-tagged representations of mRNA from two treatments, each tagged with a fluorophore emitting a different color light (usually green and red), are hybridized to the array of cDNAs and then fluorescence emission at the site of each immobilized cDNA is quantified and finally an image is produced. Measured fluorescent intensities ideally represent transcript levels in the sample. The main steps of the experimental approach of transcriptomic microarray are shown in Fig. 3 (panels A and B for wet and dry lab experiments, respectively).

![Fig. 1. Schematic steps of DNA microarray technology.](image)

![Fig. 2. Schematic illustration of spotted genes on a glass slide array. Glass slide arrays are produced by the robotic spotter that spots genes (e.g., PCR products, cDNAs, clone libraries or long oligonucleotides) onto coated glass slides. Each spot on the array represents a particular contiguous gene fragment, i.e. 40–70 nucleotides for oligonucleotide arrays, or several hundred nucleotides for PCR products.](image)

![Fig. 3. Main steps of the experimental approach of transcriptomics DNA microarray for wet (A) and dry (B) lab experiments.](image)
In single channel hybridization each slide is hybridized with a single biological sample labelled with a unique dye. Most new technologies follow this approach, e.g. Affymetrix, Agilent, Codelink. However in competitive hybridization each slide is hybridized with two biological samples each labelled with a different dye. Log ratios of the two color intensities ideally represent the relative abundance of the transcripts in one sample compared to the transcripts in the other one. The typical fluorescent images of hybridized cDNA microarray are shown in Fig. 4.

Most manufacturers of microarray scanners provide their own software for image processing (Korn et al. 2004). Image analysis level of the experiment includes scanning of the image, spot recognition, gridding, segmentation and intensity extraction (plus background subtraction) respectively. Gridding is finding the true place of spots on the array and matching them with their corresponding IDs (Giannakeas et al. 2006; Lonardi and Luo 2004; Zacharia and Maroulis 2008). Segmentation is where the spots can be separated from the background. It defines the shape of each spot. Many methods are included in softwares for segmentation such as Fixed Circle Segmentation, Adaptive Circle Segmentation, Adaptive Shape Segmentation and Histogram Segmentation, with their names indicating their function and mechanism. The selection of the best method is dependent on the quality of the produced images, dominant shape of the spots and personal experience (Ahmed et al. 2004; Katzzer et al. 2003; Lehmuussola et al. 2006). The intensity of a spot in microarray needs to be corrected for the background intensities to reduce biases. A simple method called global correction is to subtract a constant from all spot intensity values. Another method is local correction which subtracts different values depending upon the location of a spot. A problem with these methods is when the background intensity is larger than the spot intensity. This results in a negative number and makes further analysis inappropriate (e.g., log transform). To address this issue, more sophisticated background correction methods have been proposed, such as a two-dimensional locally weighted linear regression (LOWESS) smoothing. By subtraction of the background, the intensity of the spots on the array can be measured.

**Data mining**

Fig. 5 represents various steps of DNA microarray data mining and its translation into clinical applications.

**Normalization**

Many sources of errors and inconsistencies may be involved in image processing. These include but of course not limited to irregularities in the array surface, variations in the laboratory processes, different DNA strands having different hybridization properties, different platforms being used in the process, different scanner settings, different amount of mRNA used, dye effect (different dyes have different efficiencies after all), random noise, and background effects. These inequalities necessitate normalization. Normalization is the process that adjusts the individual hybridization intensities in a data matrix in order to balance them for the following data analysis. In other words, normalization is the process of correcting for bias within arrays and between arrays, prior to analysis (Kadanga et al. 2008). There are different normalization methods available. Global normalization, mean log centering, linear regression, LOWESS and rank invariant methods are mostly used (Chen 2003; Geller 2003; Quackenbush 2002). Among these methods, LOWESS analysis can remove intensity-dependent effects in the log2(ratio) values (Quackenbush 2002). LOWESS...
normalization can detect systematic deviations in the R-I (ratio-intensity) plot. In this method, systematic variations are corrected by (1) a locally weighted linear regression as a function of the log10 (intensity) and (2) compensating the experimentally observed ratio with the best-fit average log2 (ratio). Table I represents important considerations in different steps of microarray data management.

**Table 1. Considerations in different steps of microarray data management**

<table>
<thead>
<tr>
<th>Analysis step</th>
<th>Important considerations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental design and implementation</td>
<td>Number of the replicates must be determined carefully</td>
<td>(Bolstad 2004; Churchill 2002; Foster 2002; Kerr 2003; Simon 2003)</td>
</tr>
<tr>
<td>Image acquisition and analysis</td>
<td>Image should be scanned at appropriate resolution</td>
<td>(Istepanian 2003; Kadanga et al. 2008; Yang et al. 2002)</td>
</tr>
<tr>
<td>Data preprocessing and normalization</td>
<td>Poor quality spots and spots with intensity lower than the background plus two standard deviations should be discarded</td>
<td>(Cui 2003; Geller 2003; Quackenbush 2002)</td>
</tr>
<tr>
<td>Identification of differentially expressed genes</td>
<td>Use methods other than fixed threshold to infer significance</td>
<td>(Cui 2003; Gusnanto et al. 2007)</td>
</tr>
<tr>
<td>Dimension reduction</td>
<td>Use different methods to visualize the data from various perspectives</td>
<td>(Dai et al. 2006)</td>
</tr>
<tr>
<td>Supervised clustering or classification</td>
<td>Avoid over-training of the classifier</td>
<td>(Babjak 2004; Hawkins 2004; Jirapech-Umpai and Aitken 2005; Juan and Huang 2007; Khan et al. 2001)</td>
</tr>
</tbody>
</table>

After normalization the expression ratio can be calculated. The expression ratio is simply the normalized value of the expression level for a particular gene in the query sample divided by its normalized value for the control.

Then the ratio (T) for gene i can be written as:

$$T_i = \frac{R_i}{G_i}$$

where R and G represent the red (target) and green (reference) intensities.

The very basic preprocessing step is taking logarithm of each entry in gene expression data matrix in order to expand the dynamic range of gene expression signals. This is called log transformation.

$$T'_i = \log_2 \left( \frac{R_i}{G_i} \right)$$

**Dealing with missing values**

The gene expression data matrix may have missing values due to non-systematic inconsistencies such as pollution on the glass, image corruption during scanning, low resolution images, as well as systematic errors occurring in the microarray manufacturing process. Missing value estimation is important for at least two reasons. First, some popular analysis methods such as principal component analysis (PCA) require the complete data matrices to function. Second, most data mining methods can benefit from having accurate estimation of missing values. Model-based methods may be the most popular. Other common techniques include nearest neighbor methods, iterative analysis of variance methods, filling in least squares estimates (Bo et al. 2004; Kim et al. 2004), randomized inference, and likelihood-based approaches (Troyanskaya 2001). In the context of microarray, sometimes simple techniques such as replacing missing values with zeros or the average of the corresponding row or column are sufficient. However, these methods are not optimal because they do not consider problem-specific
information that may be useful for better estimation. More sophisticated approaches have been also proposed. For example, the KNImpute algorithm aims at minimizing data modelling assumptions and takes advantage of the correlation structure of the gene expression data by using genes with expression profiles similar to the gene of interest. As another example, the SVDImpute method exploits SVDs to estimate missing values by obtaining a set of mutually orthogonal expression patterns that can be linearly combined to approximate the expression of all genes in the data.

**Identification of differentially expressed genes**

All microarray experiments are carried out to find genes which are differentially expressed between two (or more) samples of cells (2000; Abiko et al. 2004; Acin et al. 2007; Caetano et al. 2004; De et al. 2004). This goal has two prerequisites. The first is to select a statistic which will rank the genes in order of evidence (significance) for differential expression, from strongest to weakest evidence. The second is to choose a critical-value for the ranking statistic above which any value is considered significant. Filtering unnecessary data has some advantages. First of all, not only elimination of the unchanged genes will help data mining procedures easier to handle but will also fasten them. The primary importance of ranking arises however, from the fact that only a limited number of genes can be followed up in a typical biological study due to limited resources. Methods used in finding differentially expressed genes are fixed cut-off threshold (usually 2 fold), unusual ratio, univariate statistical tests e.g. the t-test (Neely 2003) in case the samples are independent and F test or ANOVA (Pavlidis 2003) in case the number of conditions under study is more than two. All above methods assume that data follows normal distribution. When normal distribution criteria are not met, non-parametric tests like Kruskal-Wallis procedure (instead of one-way ANOVA) or Friedman procedure (instead of two-way ANOVA) are used.

With those tests that use \textit{P values} Bonferonni correction is used to reduce the number of false discovery rate (FDR) by reducing the significance cut-off. \textit{P value} is a popular cut-off and is defined to be the minimum false positive rate at which an observed statistic can be called significant. Genes with \textit{P values} smaller than the set threshold are more probably significant (Cheng and Pounds 2007; Grant et al. 2005; Gusnanto et al. 2007; Pawitan et al. 2005; Tsai et al. 2003).

**Higher level analysis of microarray data**

Once differentially expressed genes have successfully been distinguished, high level analyses or data mining of microarray data begins. Data Mining is all about automating the process of searching for patterns in the data. In other words, it is an iterative process of discovery.

**Dimension reduction**

The complexity of most data analysis algorithms depends on the number of input dimensions, so reducing the number of genes or experimental conditions in a microarray data set is helpful for efficient analysis, as long as the reduced data set maintains important information in the original data (Bura and Pfeiffer 2003; Dai et al. 2006).

Dimensionality reduction algorithms can be classified into feature selection and feature extraction. \textit{Feature selection} is to select \( k \) dimensions, out of the original \( d \) dimensions, that can best represent the original data set (Chen et al. 2007; Jirapech-Umpai and Aitken 2005). \textit{Feature extraction} is to find a new set of \( k \) dimensions that are some combinations of the original \( d \) dimensions. The most popular feature extraction algorithms may be the linear projection method such as \textit{principal component analysis} (PCA) for unsupervised learning (Li et al. 2008) and \textit{linear discriminant analysis} (LDA) for supervised learning (Shen et al. 2006). PCA is also called \textit{singular value decomposition} (SVD) depending on the context.

Other methods used in dimension reduction are Independent Component Analysis (Saidi et al. 2004; Zheng et al. 2008) (ICA), Correspondence Analysis (CA) (Fellenberg et al. 2001; Kishino and Waddell 2000; Tan et al. 2004) and Multidimensional Scaling (MDS).

**Clustering and classification**

When one has done multiple experiments, under different conditions -different patients, different time points, and etc- one can group the genes, which behave similarly and based on the pattern of the distinguishing genes, one can for example set boundaries between different subtypes of cancer. One can identify samples with similar expression level patterns or genes which are similar across samples. The main aim is to look for the most different features that should be the best at discriminating classes. Among different approaches used to pursue such aim, the “Euclidean distance clustering method” seems to be the commonest methodology. Fig. 6 represents the schematic illustration of Euclidean distance clustering method for expressed genes.

Supervised approaches are the analyses which are designed to determine the genes that fit a predetermined pattern. In the case of a supervised learning, one can use the annotation of either the gene or the sample, and create clusters of genes or samples in order to identify patterns that are characteristic for the cluster. In other words one can specify relationships among objects in supervised learning (Jirapech-Umpai and Aitken 2005). The main goal of supervised learning is data classification and subsequently prediction. Unlike supervised learning, unsupervised methods are used to characterize the components of a data set without the a
Reverse engineering of gene regulatory networks

Perhaps the most recent and the most important part in microarray data analysis is reverse engineering of gene regulatory networks for understanding the dynamics of gene expression. Pathway analysis towards functional enrichment can be fulfilled using two methods one of which is time-series data (Dewey 2002; Filkov et al. 2002; Klevecz et al. 2007; Maraziotis et al. 2007) and second one is steady-state data of gene knockouts (Rawool and Venkatesh 2007).

In the former approach the amount of expression of a certain gene at a certain time is a function of expression of the other genes at all previous time points. In the latter approach, the effects of deleting a certain gene on the expression of other genes are inspected and based on the regulation of the other genes; the function of that certain gene in regulation of the other genes is assessed. These methods still lack full applicability, because there is a need for more knowledge on sophisticated networks in the cells in order to identify the hidden role of different molecules in the circuitry of gene regulation.

Understanding the expression dynamics helps us infer innate complexities and phenomenological networks among genes. Defining the true place of the genes in cell networks is the main phase in our understanding of programming and functioning of living cells. Studying the regulation patterns of genes in groups, using clustering and classification methods helps us understand different pathways in the cell, their functions, regulations and the way one component in the system affects the other one. These networks can act as starting points for data mining and hypothesis generation, helping us reverse engineer.

So far various softwares have been used for image acquisition/processing and data mining. Table 2 represents some important softwares available for handling of microarray data. Of these softwares, some of them such as TM4 are freely available while some others such as ImaGen and GeneSight are commercially available. Among these tools, some deal with gene ontology which may help us towards better understanding of function genomics. For example, the Expression Analysis Systematic Explorer (EASE) developed by Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics team, is a customizable, standalone, Windows(c) desktop software application that facilitates the biological interpretation of gene lists derived from the results of microarray, proteomic, and SAGE experiments.
Table 2. Softwares available for preprocessing and/or data mining of microarray data

<table>
<thead>
<tr>
<th>Software</th>
<th>Methods available in the package</th>
<th>Advantages and/or features</th>
<th>Related URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImaGene</td>
<td>Automated grid-finding, spot-finding, spot &amp; array-level quality control, segmentation and normalization</td>
<td>ImaGene works with all scanners including Tecan, Agilent, GenePix (Axon), Perkin Elmer, Innopisy, and more; all kinds of arrays are supported (glass, filter, membrane, custom, or commercial).</td>
<td>Free trial version at <a href="http://www.biodiscovery.com/index/imagene">http://www.biodiscovery.com/index/imagene</a></td>
</tr>
<tr>
<td>Avadis</td>
<td>Data analysis and visualization</td>
<td>Avadis has a built-in GO browser to view ontology hierarchies. Prebuilt Affymetrix workflows. Avadis is highly tuned to work with Affymetrix GeneChip® data.</td>
<td><a href="http://avadis.strandgenomics.com/">http://avadis.strandgenomics.com/</a></td>
</tr>
<tr>
<td>(DAVID) Database for Annotation, Visualization and Integrated Discovery</td>
<td>Integrated solutions for the annotation and analysis of datasets</td>
<td>David can identify enriched biological themes, particularly GO terms, discover enriched functional-related gene groups, visualize genes on BioCarta &amp; KEGG pathway maps, list interacting proteins, explore gene names in batch, link gene-disease associations and highlight protein functional domains and motifs.</td>
<td><a href="http://david.abcc.ncifcrf.gov">http://david.abcc.ncifcrf.gov</a> /</td>
</tr>
<tr>
<td>EASE (Expression Analysis Systematic Explorer)</td>
<td>EASE, developed by DAVID Bioinformatics team, is a customizable, standalone, Windows(c) desktop software application that facilitates the biological interpretation of gene lists derived from the results of microarray, proteomic, and SAGE experiments.</td>
<td>EASE provides statistical methods for discovering enriched biological themes within gene lists, generates gene annotation tables, and enables automated linking to online analysis tools.</td>
<td></td>
</tr>
<tr>
<td>EGAN (Exploratory Gene Association Networks)</td>
<td>Visualizing and interpreting the results of high-throughput exploratory assays in an interactive hypergraph of genes, relationships (protein-protein interactions, literature co-occurrence, etc.) and meta-data (annotation, signaling pathways, etc.). EGAN provides comprehensive, automated enrichment analysis</td>
<td>Links to external web resources including more than 240,000 articles at PubMed, hypergeometric and GSEA-like enrichment statistics</td>
<td></td>
</tr>
<tr>
<td>FunCluster</td>
<td>Detecting co-regulated biological processes involving</td>
<td>FunCluster's functional analysis relies on GO and KEGG annotations and is currently available for three organisms: Homo sapiens, Mus musculus and Saccharomyces cerevisiae.</td>
<td>Software can be downloaded from the FunCluster website, or from the worldwide mirrors of CRAN. FunCluster is provided freely under the GNU General Public License 2.0.</td>
</tr>
<tr>
<td>GeneSifter</td>
<td>Statistical framework with 15 advanced options</td>
<td></td>
<td><a href="http://www.genesifter.net/web/">http://www.genesifter.net/web/</a></td>
</tr>
<tr>
<td>Expression Profiler</td>
<td>Clustering, pattern discovery, statistics (thru R), machine-learning algorithms and visualization.</td>
<td></td>
<td><a href="http://www.ebi.ac.uk/expressionprofiler/">http://www.ebi.ac.uk/expressionprofiler/</a></td>
</tr>
</tbody>
</table>

---

Copyright © 2011 by Tabriz University of Medical Sciences

BioImpacts, 2011, 1(2), 75-86 | 81
<table>
<thead>
<tr>
<th>Software</th>
<th>Methods available in the package</th>
<th>Advantages and/or features</th>
<th>Related URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenMAPP</td>
<td>Visualizing gene expression data on maps representing biological pathways and groupings of genes.</td>
<td>Integrated with GenMAPP are programs to perform a global analysis of gene expression or genomic data in the context of hundreds of pathway MAPPs and thousands of GO terms (MAPPFinder), import lists of genes/proteins to build new MAPPs (MAPPBuil</td>
<td><a href="http://www.genmapp.org/">www.genmapp.org/</a></td>
</tr>
<tr>
<td>Bioconductor</td>
<td>Bioconductor comes with many packages that cover the very parts of gene expression data mining.</td>
<td>Different packages are available in the Bioconductor website. When released BioC 2.5, consisted of 352 packages. For more information refer to <a href="http://www.bioconductor.org/">http://www.bioconductor.org/</a></td>
<td><a href="http://www.bioconductor.org/">http://www.bioconductor.org/</a></td>
</tr>
<tr>
<td>ReInet</td>
<td>Relevance Networks</td>
<td>The software is written in Java and it runs under any operating system. It dynamically determines the latest names, symbols, functions, and genome position for each gene and includes these in the relevance networks output.</td>
<td><a href="http://chip.org/reinnet/">http://chip.org/reinnet/</a></td>
</tr>
<tr>
<td>BASE (BioArray Software Environment)</td>
<td>Web-based microarray database and analysis platform</td>
<td></td>
<td><a href="http://base.thep.lu.se/">http://base.thep.lu.se/</a></td>
</tr>
<tr>
<td>Partek Genomics Suite</td>
<td>Advanced statistics and interactive data visualization specifically designed to extract biological signals from noisy data.</td>
<td></td>
<td><a href="http://www.partek.com">http://www.partek.com</a></td>
</tr>
<tr>
<td>TM4</td>
<td>The TM4 suite of tools consist of four major applications, Microarray Data Manager (MADAM), TIGR Spotfinder (image processing tool), Microarray Data Analysis System (MIDAS), and Multi-experiment Viewer (MeV), as well as a Minimal Information About a Microarray Experiment (MIAME)-compliant MySQL database.</td>
<td>MeV identifies patterns of gene expression and differentially expressed genes MADAM is a java-based application to load and retrieve microarray data to and from a database. TIGR Spotfinder is an image processing software. MIDAS is a microarray data quality filtering and normalization tool.</td>
<td><a href="http://www.tm4.org/">http://www.tm4.org/</a></td>
</tr>
<tr>
<td>BNAry</td>
<td>Constructing gene regulatory networks using Bayesian networks</td>
<td>BNAry can handle microarray datasets with missing data.</td>
<td><a href="http://www.cls.zju.edu.cn/binfo/BNAry/">http://www.cls.zju.edu.cn/binfo/BNAry/</a></td>
</tr>
<tr>
<td>ArrayPipe</td>
<td>Application features range from quality assessment of slides through various data visualizations to normalization and detection of differentially expressed genes.</td>
<td></td>
<td><a href="http://www.pathogenomics.ca/arpypipe/">http://www.pathogenomics.ca/arpypipe/</a></td>
</tr>
<tr>
<td>Engene</td>
<td>Visualizing, preprocessing and clustering</td>
<td>Clustering analysis algorithms include k-means, HAC, fuzzy c-means, kernel c-means, SOMs and PCA</td>
<td><a href="https://chirimoyo.ac.uma.es/engenet/">https://chirimoyo.ac.uma.es/engenet/</a></td>
</tr>
<tr>
<td>ExpressYourself</td>
<td>The software performs correction of the background array signal, normalization, scoring, combination of replicate experiments, filtering problematic regions of the array and quality assessment of hybridizations.</td>
<td>ExpressYourself investigates the quality of experiments by measuring hybridization consistency within single slides and across replicated experiments. The data quality step calculates the overall performance of experiments and highlights problematic array regions.</td>
<td>Freely available at <a href="http://bioinfo.mbb.yale.edu/expressyourself/">http://bioinfo.mbb.yale.edu/expressyourself/</a></td>
</tr>
<tr>
<td>FCluster</td>
<td>Fuzzy clustering of microarray data</td>
<td></td>
<td><a href="http://fuzzy.cs.uni-magdeburg.de/fcluster/">http://fuzzy.cs.uni-magdeburg.de/fcluster/</a></td>
</tr>
<tr>
<td>GEPAS (Gene Expression Pattern Analysis Suite)</td>
<td>Normalization, and preprocessing such as log transformation, replicate handling and missing value imputation. It supports hierarchical clustering and SOMs for data clustering.</td>
<td>On-line tutorials are available from main web server(<a href="http://bioinfo.cnio.es">http://bioinfo.cnio.es</a>).</td>
<td><a href="http://gepas.bioinfo.cnio.es">http://gepas.bioinfo.cnio.es</a></td>
</tr>
<tr>
<td>Genes@Work</td>
<td>Genes@Work is a pattern discovery and classification system.</td>
<td></td>
<td><a href="http://www.research.ibm.com/FunGen/FGGenesAtWor">http://www.research.ibm.com/FunGen/FGGenesAtWor</a></td>
</tr>
<tr>
<td>Spotfire</td>
<td>Spotfire allows users to interactively mine, visualize, and analyze large sets of technical, multidimensional data.</td>
<td></td>
<td><a href="http://spotfire.tibco.com/">http://spotfire.tibco.com/</a></td>
</tr>
</tbody>
</table>
Microarray data mining for novices

Ethical issues
None to be declared.

Conflict of interests
Authors declare no conflict of interest.

Acknowledgement
Authors are grateful to the Ministry of Health, Care and Medical Education for the financial support.

References


Barar J, Hamzeiy H, Mortazavi-Tabatabaei SA, Hashemi-Aghdam SE and Omidi Y. 2009. Genomic Signature and
Toxicogenomics Comparison of Polycationic Gene Delivery Nanosystems in Human Alveolar Epithelial A549 Cells. *Daru*, 17(3), 139-147.


